

УДК 535.4

ОСОБЕННОСТИ ПРОЕКТИРОВАНИЯ ДИФРАКЦИОННЫХ РЕШЕТОК

В.Н. Васильев, Д.Н. Воронцов, А.И. Драгунов, И.Л. Лившиц, Е.А. Соколова

Исследуются особенности автоматизированного проектирования различных типов дифракционных решеток с использованием программы ZEMAX. Показано, что выбор моделей задания дифракционных решеток в большинстве случаев определяется технологией их изготовления. Представлены рекомендации по заданию пяти типов решеток – эллиптической решетки первого и второго типа, голограммы, бинарной поверхности первого типа и оптически сформированной голограммы. Показана связь способа моделирования дифракционных решеток с технологией их изготовления. Даны рекомендации по выбору оптимальной модели для нарезных и голограммных дифракционных решеток.

Ключевые слова: дифракционная решетка, спектрометр, голограмма, фазовые коэффициенты.

Введение

Дифракционные оптические элементы находят все более широкое применение в различных оптических системах, однако процесс их практического проектирования вызывает сложности в большинстве работ, что связано в основном с ограниченным количеством публикаций по этой актуальной тематике.

Программы автоматического расчета оптических систем существенно экономят время разработчика. Исторически такие программы создавались в первую очередь для проектирования линзовых объективов и уже позже в них добавлялись возможности моделирования других оптических систем, например, таких как спектрометры с дифракционными решетками. Тип оптической поверхности под названием «Дифракционная решетка» присутствует в программе ZEMAX. Однако этот тип поверхности позволяет моделировать только дифракционную решетку с прямыми и равноотстоящими штрихами, что уже недостаточно при современном развитии технологии изготовления дифракционных решеток, а также теории aberrаций дифракционных решеток, имеющих неплоскую поверхность, сложную форму штрихов и неравномерное распределение штрихов по поверхности решетки. Программа содержит также еще ряд оптических поверхностей, которые могут быть использованы для моделирования различных дифракционных решеток. Однако в описании программы [1] отсутствует информация о том, что эти типы поверхностей могут быть использованы для моделирования дифракционных решеток. Вопрос о моделировании различных типов дифракционных решеток не рассматривается в учебных курсах, читаемых разработчиками программы для пользователей. Более того, инструкторы, читающие эти курсы, чаще всего затрудняются ответить на вопросы, связанные с моделированием сложных типов решеток, поскольку не достаточно знакомы с технологией их изготовления.

В настоящей работе на простом примере моделирования и оптимизации спектрометра с вогнутой сферической дифракционной решеткой и плоским приемником последовательно показано, какие поверхности могут быть использованы для такого моделирования, какой уровень компенсации aberrаций может быть достигнут при использовании различных моделей, как параметры оптимизации связаны с технологическими параметрами изготовления дифракционных решеток. Целью работы является объединение достижений в автоматическом проектировании, теории aberrаций и технологии изготовления дифракционных решеток в одну систему, что должно позволить более широкому кругу расчетчиков оптических систем успешно проектировать спектрометры.

Поверхность «Дифракционная решетка»

Такая решетка имеет постоянный шаг и прямолинейные штрихи. При фиксированном радиусе кривизны параметры для оптимизации самой решетки отсутствуют.

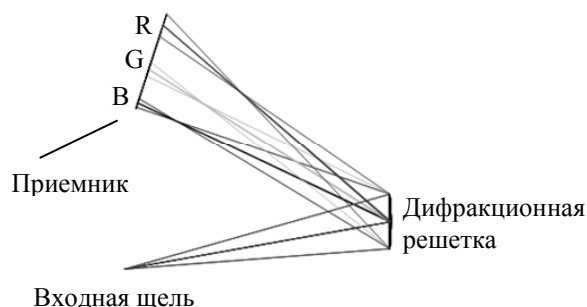


Рис. 1. Ход лучей в спектрометре с решеткой, имеющей прямые равноотстоящие штрихи

Можно оптимизировать только параметры схемы спектрометра. У такой решетки фокальные кривые в двух взаимно-перпендикулярных плоскостях не совпадают: в меридиональной плоскости – это круг Роуланда, в сагиттальной – прямая Сиркса, т.е. касательная к кругу Роуланда, перпендикулярная нормали решетки. Спектральные изображения точки на входной щели являются астигматическими и расфокусированными. Ход лучей в спектрометре с такой решеткой показан на рис. 1.

Эллиптическая решетка первого типа

Такая поверхность в общем случае может иметь эллипсоидальную форму. Однако для большинства спектрометров целесообразно использовать сферическую поверхность как частный случай эллипсоида. В отличие от предыдущего типа поверхности, эллиптическая решетка первого типа позволяет вводить и оптимизировать коэффициенты неравномерности шага α и β , которые являются параметрами оптимизации и могут быть технологически реализованы при изготовлении решетки на делительной машине. Шаг эллиптической решетки изменяется по закону

$$d = d_0 + \alpha y + \beta y^2,$$

где d_0 – расстояние между штрихами в центре решетки; d – расстояние между штрихами в точке на поверхности решетки с координатой y .

Шаг решетки с переменным шагом, задаваемым коэффициентами μ и ν , изготовленной на делительной машине, изменяется по закону

$$d = d_0(1 + \mu y + \nu y^2).$$

Таким образом, оптимизированные при помощи программы ZEMAX коэффициенты неравномерности шага решетки перед изготовлением должны быть пересчитаны по формулам

$$\mu(\text{мм}^{-1}) = \frac{\alpha}{d_0}(\text{мкм}),$$

$$\nu(\text{мм}^{-1}) = \frac{\beta}{d_0}(\text{мкм}).$$

Эллиптическая решетка второго типа

Эллиптическая решетка второго типа имеет равноотстоящие, но криволинейные штрихи. Дополнительным параметром оптимизации в этом случае является кривизна штриха, задаваемая параметром θ , который также является тангенсом угла наклона оси качения резца делительной машины при изготовлении дифракционной решетки. Пересчет параметров перед изготовлением в данном случае не требуется. Однако при оптимизации надо иметь в виду, что большинство конструкций делительных машин не позволяет наклонять ось качения резца на угол, превышающий 30° . На делительной машине также могут быть изготовлены дифракционные решетки, имеющие и переменный шаг, и криволинейные штрихи. Однако программа ZEMAX не имеет поверхности, при помощи которой можно непосредственно моделировать такие решетки. Для моделирования таких решеток может быть использована бинарная поверхность первого типа, свойства которой рассмотрены ниже.

Голограмма

Поверхность, обозначенная в программе ZEMAX как голограмма, подходит для моделирования голограммной решетки, записанной при помощи двух точечных источников (точечных диафрагм, в которые сфокусированы записывающие пучки). Дополнительными параметрами оптимизации в данном случае являются координаты точечных источников, используемых для записи решетки. В качестве исходных параметров можно выбрать источники, расположенные симметрично на круге Роуланда. Например, для записи решетки, имеющей 1200 штрихов на миллиметр (расстояние между штрихами в центре решетки равно 833,333 нм) на сферической поверхности с радиусом кривизны R , равным 100 мм, при помощи лазера с длиной волны 457 нм, углы падения излучения двух записывающих пучков на сферическую поверхность, γ и $-\gamma$, могут быть рассчитаны по формуле

$$\gamma = \arcsin[(1/d)/2] = \arcsin[(457/833,333)/2] = 15,9144^\circ.$$

Исходные координаты точечных источников $Y1$, $Z1$, $Y2$ и $Z2$ в этом случае рассчитываются по формулам

$$Y1 = R \sin \gamma \cos \gamma = 26,3691 \text{ мм} \quad Z1 = -R \cos 2\gamma = -92,4814 \text{ мм},$$

$$Y2 = -R \sin \gamma \cos \gamma = -26,3691 \text{ мм} \quad Z2 = -R \cos 2\gamma = -92,4814 \text{ мм}.$$

При оптимизации необходимо контролировать сохранение заданного числа штрихов на миллиметр (постоянной решетки), а также реализуемость схемы записи – ограничение углов падения излучения на сферическую поверхность.

Оптически сформированная голограмма

Поверхность, названная в программе ZEMAX оптически сформированной голограммой, может быть использована для моделирования голограммной решетки, записанной в более сложной оптической схеме, когда после точечных диафрагм на пути записывающих пучков используются дополнительные оптические элементы, например, линзы или сферические зеркала. Параметры этих элементов являются дополнительными параметрами оптимизации. В случае, когда оптимизация схемы записи при помощи двух точечных источников не дает желаемого результата, в нее вводят дополнительные оптические элементы. При этом каждый записывающий пучок моделируется в отдельном файле, и при оптимизации изменяются параметры как основного файла, в котором смоделирована оптическая схема спектрометра, так и двух вспомогательных файлов, соответствующих двум записывающим пучкам со всеми оптическими элементами, формирующими падающие на заготовку решетки волновые фронты. Перед введением в схему записи дополнительных оптических элементов координаты точечных источников можно пересчитать в параметры, используемые во вспомогательных файлах, по следующим формулам:

$$\text{Thickness} = (Y^2 + Z^2)^{1/2},$$

$$\text{TiltaboutX} = \arctg Y/Z.$$

Бинарная поверхность первого типа

Дифракционная решетка с любой формой штрихов и любым их расположением может быть описана полиномиальным уравнением, которое связывает координаты точки на его поверхности с номером штриха и набором коэффициентов. Ограничивая число членов разложения определенными степенями координат, мы ограничиваем число рассматриваемых аберраций. В программе ZEMAX такая решетка моделируется как бинарный элемент, заданный набором фазовых коэффициентов, однозначно связанных с коэффициентами разложения в степенной ряд функции оптического пути, а значит, и с аберрациями. Распределение штрихов на поверхности решетки при этом описывается формулой

$$f(x, y) = \frac{1}{2\pi} \sum_{i=0}^{n/2} \sum_{k=0}^{n/2i} X_i Y_k x^{2i} y^k,$$

где $X_0 Y_0 = 0$; $X_0 Y_1 = 2\pi N$, ближайшее целое значение $f(x, y)$ является номером штриха, который может быть как положительным, так и отрицательным (штрих номер 0 проходит через центр решетки), N – число штрихов на мм в центре решетки. Связь фазовых коэффициентов бинарной поверхности (при нормирующем радиусе, равном единице) с коэффициентами разложения в ряд функции оптического пути выражается формулами [2]

$$C_{\text{ZEMAX}} = 2\pi C_n$$

$$C_1 = H_{010} = 0, \quad C_2 = H_{100}, \quad C_3 = H_{020} / 2,$$

$$C_4 = H_{110} = 0, \quad C_5 = H_{200} / 2,$$

$$C_6 = H_{030} = 0, \quad C_7 = -H_{120} / 2, \quad C_8 = H_{210} = 0,$$

$$C_9 = -H_{300} / 6,$$

$$C_{10} = H_{040} / 24, \quad C_{11} = H_{130} = 0, \quad C_{12} = H_{220} / 6,$$

$$C_{13} = H_{310} = 0, \quad C_{14} = H_{400} / 24.$$

Каждый коэффициент H_{ijk} влияет на определенную аберрацию дифракционной решетки. Например, C_2 связан с расстоянием между штрихами решетки. Он равен отношению длины волны записи к расстоянию между штрихами в вершине решетки. C_3 связан с кривизной центрального штриха. Если C_3 равен нулю, проекция центрального штриха на плоскость, касательную к сферической поверхности решетки в ее центре, является прямой линией. Этот коэффициент определяет фокусировку в сагиттальной плоскости и используется для компенсации астигматизма первого порядка. C_5 характеризует линейное изменение шага решетки. Он показывает, насколько меридиональная фокальная кривая отличается от круга Роуланда, для которого $C_5=0$. C_9 характеризует меридиональную кому. Если $C_9=0$, меридиональная кома отсутствует на круге Роуланда. C_7 связан с сагиттальной комой, C_{14} определяет сферическую аберрацию, и т.д. Эти коэффициенты также однозначно связаны с коэффициентами разложения в ряд функции оптического пути вычисления наклона оси качения резца, являющимися технологическими параметрами для изготовления решетки на делительной машине:

$$H_{ijk} = \frac{\partial^j n(y, z)}{\partial y^i \partial z^j}, \quad H_{010} = 0, \quad H_{100} = \frac{1}{d_0}, \quad H_{020} = \frac{\text{tg}\gamma}{d_0 R}, \quad H_{110} = 0, \quad H_{200} = \frac{\text{tg}\gamma - \mu R}{d_0 R},$$

$$H_{030} = 0, \quad H_{120} = -\frac{\mu \text{tg}\gamma}{d_0 R}, \quad H_{210} = 0, \quad H_{300} = \frac{2R(\mu^2 - \nu) - 3\mu \text{tg}\gamma}{d_0 R},$$

$$H_{040} = \frac{3\text{tg}\gamma(\mu R\text{tg}\gamma - 1)}{d_0 R^2}, H_{130} = 0, H_{220} = \frac{2\text{tg}\gamma(1 + 2\mu^2 R^2 - 2R^2\nu - \mu R\text{tg}\gamma)}{d_0 R^3},$$

$$H_{130} = 0, H_{400} = \frac{3\left(-2\mu^2 + 4\mu\nu + \frac{(1 + 4\mu^2 R^2 - 4R^2\nu)\text{tg}\gamma}{R^3} - \frac{\mu\text{tg}2\gamma}{R^2}\right)}{d_0}.$$

Они также связаны с параметрами записи голограммной решетки, когда для записи используются два точечных источника:

$$H_{010} = 0, H_{110} = 0, H_{030} = 0, H_{210} = 0, H_{130} = 0, H_{230} = 0,$$

$$H_{120} = \frac{\sin\gamma}{r_c^2} - \frac{\cos\gamma\sin\gamma}{Rr_c} - \frac{\sin\delta}{r_d^2} + \frac{\cos\delta\sin\delta}{Rr_d},$$

$$H_{200} = -\frac{\cos\gamma}{R} + \frac{\cos^2\gamma}{r_c} + \frac{\cos\delta}{R} - \frac{\cos^2\delta}{r_d},$$

$$H_{120} = \frac{\sin\gamma}{r_c^2} - \frac{\cos\gamma\sin\gamma}{Rr_c} - \frac{\sin\delta}{r_d^2} + \frac{\cos\delta\sin\delta}{Rr_d},$$

$$H_{200} = \frac{3\cos\gamma\sin\gamma(R\cos\gamma - r_c)}{Rr_c^2} - \frac{3\cos\delta\sin\delta(R\cos\delta - r_d)}{Rr_d^2},$$

$$H_{040} = \frac{3}{4} \left[\frac{4(R - r_c\cos\gamma)}{R^2 r_c} - \frac{\left(2 - \frac{2r_c\cos\gamma}{R}\right)^2}{r_c^2} - \frac{4(R - r_d\cos\delta)}{R^2 r_d} - \frac{\left(2 - \frac{2r_d\cos\delta}{R}\right)^2}{r_d^2} \right],$$

$$H_{220} = \frac{1}{4} \left[\frac{4(R - r_c\cos\gamma)}{R^2 r_c} - \frac{\left(2 - \frac{2r_c\cos\gamma}{R}\right)^2}{r_c^2} - \frac{4(R - r_d\cos\delta)}{R^2 r_d} + \frac{\left(2 - \frac{2r_d\cos\delta}{R}\right)^2}{r_d^2} + \right.$$

$$\left. + \frac{6\left(2 - \frac{2r_c\cos\gamma}{R}\right)\sin^2\gamma}{r_c^2} - \frac{\left(2 - \frac{2r_d\cos\delta}{R}\right)\sin^2\gamma}{r_d^2} \right],$$

$$H_{440} = \frac{3}{4} \left\{ \frac{4(R - r_c\cos\gamma)}{R^2 r_c} - \frac{\left(2 - \frac{2r_c\cos\gamma}{R}\right)^2}{r_c^2} - \frac{4(R - r_d\cos\delta)}{R^2 r_d} + \frac{\left(2 - \frac{2r_d\cos\delta}{R}\right)^2}{r_d^2} + \right.$$

$$\left. + \frac{12\left(2 - \frac{2r_c\cos\gamma}{R}\right)\sin^2\gamma}{r_c^2} - \frac{20\sin^4\gamma}{r_c^2} - \frac{12\left(2 - \frac{2r_d\cos\delta}{R}\right)\sin^2\delta}{r_d^2} + \frac{20\sin^4\delta}{r_d^2} \right\}.$$

После нахождения оптимальных фазовых коэффициентов необходимо сконструировать такую схему записи, которая обеспечит реализацию этих коэффициентов. Изменить значения фазовых коэффициентов можно при помощи введения в схему записи дополнительных элементов – линз, зеркал и дифракционных решеток. Пример такой схемы записи показан на рис. 2.

Моделируется такая система пятью сферическими поверхностями, две пропускающие, одна отражающая и снова две пропускающие в обратном порядке. Последняя пропускающая поверхность преобразуется в отражательную поверхность, описываемую теми же фазовыми коэффициентами, что и решетка, которую требуется записать. При этом знак расстояния от этой поверхности до поверхности изображения (которое совпадает с объектом – исходной точкой) не меняется на противоположный, как положено после отражения, а остается прежним. Знак порядка дифракции меняется на противоположный по

сравнению с тем, в котором решетка должна работать в спектрометре. Изображение точки в данной модели оптимизируется до идеального качества. Параметрами оптимизации являются: расстояние от точечного источника до заготовки; радиус кривизны задней поверхности заготовки; толщина заготовки; расстояние от заготовки до отражательного элемента; угол наклона отражательного элемента; радиус кривизны отражательного элемента и другие его параметры, если он является дифракционным. При оптимизации не должно нарушаться условие встречи каждого луча с его собственным отражением на поверхности фоторезиста.

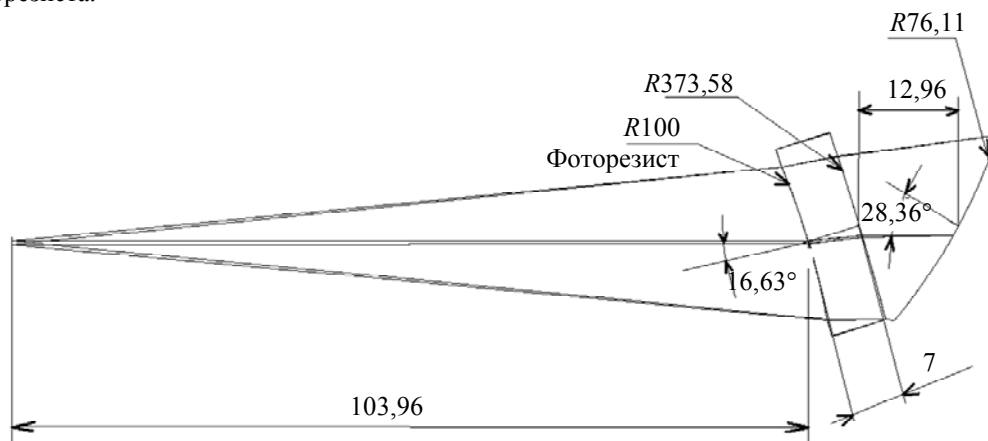


Рис. 2. Модель схемы записи решетки при помощи падающего и отраженного пучка

Следует отметить, что дифракционные решетки, записанные при помощи падающего и отраженного пучков, в отличие от решеток, записанных в попутных пучках, имеют треугольную форму штрихов, т.е. являются решетками с блеском. Длина волны блеска дифракционных решеток, записанных при помощи установок, построенных по схемам, подобным оптической схеме, показанной на рис. 2, определяется расстоянием между пучностями стоячих волн в слое фоторезиста, которое, в свою очередь, зависит от длины волны лазера и показателя преломления фоторезиста. Согласно результатам работ [3–6], она приблизительно равна половине длины волны лазерного излучения. Таким образом, в случае использования излучения аргонового лазера с длиной волны 457,9 нм длина волны блеска приблизительно равна 228,95 нм. Более подробный расчет эффективности дифракционных решеток может быть выполнен, например, при помощи программы PCGrate [7], алгоритм которой разработан на основе электромагнитной теории [8]. Программы оптического дизайна не позволяют проводить энергетические расчеты, выходящие за рамки скалярной теории.

Заключение

В работе показано, как при помощи программы ZEMAX могут быть построены модели спектрометров с дифракционными решетками, изготовленными с использованием различных технологий. Приведены разъяснения того, как параметры оптимизации дифракционной решетки связаны с технологическими параметрами ее изготовления, а также формулы, описывающие эту связь. Выявлены ограничения, которые необходимо учитывать при оптимизации оптических схем спектрометров, чтобы получать технологически реализуемые результаты. Работа объединяет вопросы оптического дизайна спектрометра и технологию изготовления дифракционных решеток, показывает их однозначную связь.

Возможные варианты моделирования дифракционной решетки в зависимости от технологии ее изготовления представлены в таблице.

	Дифракционная решетка	Эллиптическая решетка	Голограмма	Оптическая и записанная голограмма	Бинарная поверхность
Делительная машина	+	+			+
Голограммная запись			+	+	+
Прямая запись лазерным или электронным пучком	+				+

Таблица. Возможные варианты моделирования дифракционной решетки в зависимости от технологии ее изготовления

Работа проводилась при финансовой поддержке Министерства образования и науки Российской Федерации в рамках государственного контракта 11.519.11.6014.

Литература

1. ZEMAX 12 OpticalDesignProgramUser'sManual, May 10, 2012 [Электронный ресурс]. – Режим доступа: <http://www.radiantzemax.com/en/support/downloads/zemax.aspx>, свободный. Яз. англ. (дата обращения 18.10.2012).
2. Sokolova E. Simulation of Mechanically Ruled Concave Diffraction Gratings by Use of an Original Geometric Theory // Appl. Opt. – 2004. – V. 43. – № 1. – P. 20–28.
3. Sokolova E., Kruizinga B., Valkenburg T., Schaarsberg J. Recording of concave diffraction gratings in counterpropagating beams using meniscus blanks // J. Mod. Opt. – 2002. – V. 49. – № 11. – P. 1907–1917.
4. Hutley M.C. Diffraction Gratings. – London: Academic Press, 1982. – 330 p.
5. Hutley M.C. Concave diffraction gratings. – NPL report MOM 77. – Oct. 1985. – 28 p.
6. Пейсахсон И.В. Приближенная оценка эффективности отражательных дифракционных решеток // Оптический журнал. – 1997. – Т. 64. – № 11. – С. 53–57.
7. Kotov A., Golubenko I. PC Grate Series, ver. 3.0 // User's Manual, St. Petersburg, Russia, 1999. – P. 5–11.
8. Голубенко И.В. Численный анализ свойств голограммных отражательных дифракционных решеток: Дис. ... канд. физ.-мат. наук. – Гос. оптич. ин-т. им. С.И. Вавилова, 1987. – 183 с.

- Васильев Владимир Николаевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, член-корр. РАН, ректор, Vasilev@mail.ifmo.ru
- Воронцов Дмитрий Николаевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, инженер, dimitrion@hotmail.ru
- Драгунов Александр Иванович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, AlexDragunov27@yandex.ru
- Лившиц Ирина Леонидовна** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, ст. научный сотрудник, зав. лабораторией, Ecenter-optica10@yandex.ru
- Соколова Елена Алексеевна** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, ст. научный сотрудник, elena@scarlet.nl

УДК 535.317

АБЕРРАЦИОННЫЙ АНАЛИЗ КОМПОЗИЦИИ ТОНКОГО ОПТИЧЕСКОГО КОМПОНЕНТА С КОНЦЕНТРИЧЕСКИМ МЕНИСКОМ

В.В. Ежова, В.А. Зверев, Т.В. Точилина

Показано, что в изображении, образованном оптической системой, состоящей из тонкой линзы и концентричного входному зрачку мениска конечной толщины, принципиально можно достичь плананастигматической коррекции аберраций. Однако при этом возникает проблема выбора материала линз. Подобный анализ коррекционных возможностей оптической системы выполнен и в том случае, когда тонкий компонент состоит из двух тонких линз, оптическая сила которых имеет разный знак.

Ключевые слова: изображение, оптическая система, тонкий компонент, аберрация, концентрическая система, входной зрачок.

Введение

Теоретическую базу композиции оптических систем, удовлетворяющих требованиям современных оптических устройств, составляют результаты исследования аберрационных свойств оптических поверхностей, отдельных линз и их сочетаний. Эти исследования определяют суть научной школы вычислительной оптики в СПб НИУ ИТМО, основы которой были заложены трудами профессора М.М. Русинова и его учеников [1–4]. Предлагаемая работа посвящена исследованию аберрационных свойств тонкого компонента с концентрическим мениском конечной толщины, цель которого определяется потребностью в развитии теории композиции оптических систем соответствующего типа и в решении задач оптимизации их параметров.

Постановка задачи

Если расстояния между поверхностями сколь угодно сложной системы не являются коррекционными параметрами, в первом приближении их можно принять равными нулю. Такую систему будем называть тонким компонентом. При этом равным нулю будет и расстояние между главными плоскостями системы. Тонкий компонент можно считать простейшей структурной единицей при построении любой

оптической системы. В параксиальной области тонкий компонент будем характеризовать его оптической силой ϕ . Оптическая сила тонкого компонента равна $\phi_k = \sum_{i=0}^m \phi_i$, где ϕ_i – оптическая сила i -ой линзы

тонкого компонента. Параметры тонкого компонента позволяют получить апланатическую коррекцию aberrаций в образованном изображении, т.е. коррекцию aberrаций широких пучков лучей. Для коррекции aberrаций узких пучков лучей необходим дополнительный компенсатор, в качестве которого можно применить второй тонкий компонент, расположенный на конечном расстоянии от первого. Однако известно, что главные плоскости концентрического мениска совмещены и проходят через центр кривизны поверхностей мениска. Таким образом, при конечном расстоянии между главными плоскостями тонкий компонент и концентрический мениск могут располагаться в непосредственной близости друг к другу. По этой причине важно выяснить возможность коррекции осевых и полевых aberrаций в изображении, образованном такой достаточно компактной системой.

Анализ возможной коррекции aberrаций в изображении, образованном оптической системой, состоящей из тонкого компонента и концентрического мениска конечной толщины

При нормировке величин $n' = 1$, $\alpha' = 1$, $\beta_1 = 1$ выражения, определяющие коэффициенты сферической aberrации, комы и астигматизма третьего порядка изображения, образованного тонким компонентом, можно представить в следующем виде:

$$S_I = f'_k P^\infty, \quad S_{II} = z_p P^\infty + f'_k W^\infty, \quad (1)$$

$$S_{III} = \left(\frac{1}{f'_k} \right) z_p^2 P^\infty + 2z_p W^\infty + f'_k, \quad (2)$$

где P^∞ и W^∞ – основные параметры тонкого компонента; z_p – расстояние от осевой точки тонкого компонента до осевой точки входного зрачка.

Положив $z_p = 0$, получаем $S_I = f'_k P^\infty$; $S_{II} = f'_k W^\infty$; $S_{III} = f'_k$. При $P^\infty = 0$ и $W^\infty = 0$ независимо от положения входного зрачка имеем $S_I = 0$; $S_{II} = 0$; $S_{III} = f'_k$. Пусть $S_{II} = 0$ и $S_{III} = 0$. Тогда из соотношений (1) и (2) находим, что

$$P^\infty = \left(\frac{f'_k}{z_p} \right)^2 > 0, \quad W^\infty = -\frac{f'_k}{z_p}.$$

Вполне очевидно, что чем меньше абсолютная величина параметров P^∞ и W^∞ , тем больше должна быть абсолютная величина отрезка z_p , необходимая для компенсации остаточных комы и астигматизма изображения. В общем случае выражение (2) можно рассматривать как квадратное уравнение относительно переменной величины z_p . Решая это уравнение, получаем

$$z_p = -\frac{W^\infty}{P^\infty} f'_k \left[1 \pm \sqrt{1 - \frac{P^\infty}{W^{\infty 2}} \left(1 - \frac{S_{III}}{f'_k} \right)} \right]. \quad (3)$$

Это уравнение имеет вещественное решение при соблюдении условия $1 - \frac{P^\infty}{W^{\infty 2}} \left(1 - \frac{S_{III}}{f'_k} \right) \geq 0$. При $S_{III} = 0$ это условие принимает вид: $W^{\infty 2} - P^\infty \geq 0$. Основные параметры P^∞ и W^∞ одиночной тонкой линзы в воздухе взаимосвязаны соотношением [5]:

$$P^\infty = P_0^\infty + \left[1 - \frac{1}{(n+1)^2} \right] \left[W^\infty - \frac{1}{2(2+n)} \right]^2, \quad (4)$$

где $P_0^\infty = \frac{n(4n-1)}{4(2+n)(n-1)^2}$. Это соотношение можно преобразовать к виду

$$W^{\infty 2} - P^\infty = \frac{1}{(n+1)^2} \left(W^\infty + \frac{1}{2}n \right)^2 - \frac{n^2}{4(n-1)^2}. \quad \text{Условию } W^{\infty 2} - P^\infty \geq 0 \text{ соответствует выражение}$$

$$\left(W^\infty + \frac{1}{2}n \right)^2 \geq \frac{1}{4} \left(\frac{n+1}{n-1} \right)^2 n^2. \quad \text{Отсюда следует, что величина параметра } W^\infty \text{ должна удовлетворять усло-}$$

виям $W_1^\infty \geq \frac{n}{n-1}$; $W_2^\infty \geq -\frac{n^2}{n-1}$. При значениях параметра W^∞ , удовлетворяющих этим условиям, и при расстоянии до входного зрачка, определяемом формулой (3), в изображении, образованном тонкой линзой в воздухе, будут отсутствовать кома и астигматизм третьего порядка. Однако при этом параметр $P^\infty \neq 0$; следовательно, не будет равен нулю и коэффициент S_I , определяющий остаточную сферическую aberrацию изображения.

Дополним тонкий компонент мениском, поверхности которого concentричны центру входного зрачка, расположенному в переднем фокусе тонкого компонента, как показано на рисунке. Главные плоскости concentрического мениска совмещены и проходят через центр кривизны его поверхностей. В результате получаем оптическую систему, у которой оптическая сила и задний фокальный отрезок равны

$$\varphi = \varphi_M + \varphi_K - \varphi_M \varphi_K d;$$

$$s'_{F'} = \frac{(1 - \varphi_M d)}{\varphi},$$

где d – расстояние от задней главной точки мениска до осевой точки тонкого компонента.

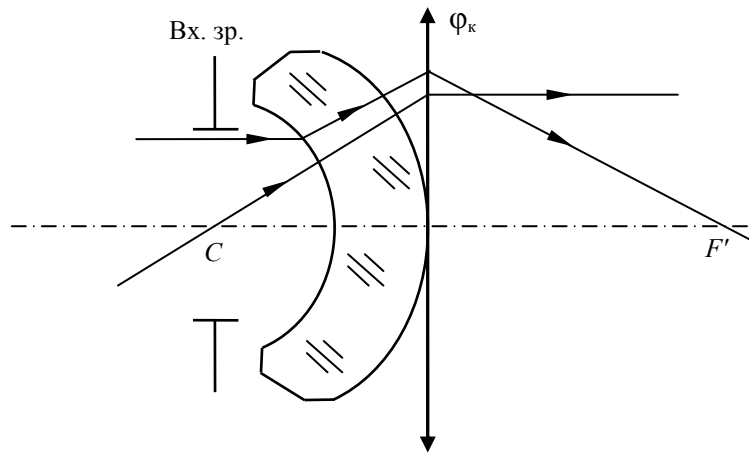


Рисунок. Схема оптики объектива

При $d = f'_K$ оптическая сила системы $\varphi = \varphi_M + \varphi_K - \varphi_M = \varphi_K$, а задний фокальный отрезок $s'_{F'} = \frac{(\varphi - \varphi_M)}{\varphi^2} = f'(1 - f'\varphi_M)$. При $n' = 1$, $\alpha' = 1$, $\beta_1 = 1$ и $z_p = -f'_K = -f'$ имеем соотношения

$$S_I = S_{IM} + h_K P_K,$$

$$S_{II} = S_{IIM} + H_K P_K - J W_K = S_{IIM} - f'(P_K - W_K),$$

$$S_{III} = S_{IIIM} + \frac{f'^2}{h_K} P_K - 2 \frac{f'^2}{h_K} W + f',$$

где $h_K = s'_{F'} = f'(1 - f'\varphi_M)$. В рассматриваемом случае выражения, определяющие коэффициенты S_{IIM} и S_{IIIM} , удобно представить в виде

$$S_{IIM} = \sum_{i=1}^2 h_i P_i \frac{\beta_{i+1} - \beta_i}{\alpha_{i+1} - \alpha_i};$$

$$S_{IIIM} = \sum_{i=1}^2 h_i P_i \left(\frac{\beta_{i+1} - \beta_i}{\alpha_{i+1} - \alpha_i} \right)^2.$$

При входном зрачке, расположенном в центре кривизны поверхностей concentрического мениска, углы $\beta_1 = \beta_2 = \beta_3$. Вполне очевидно, что при этом $S_{IIM} = S_{IIIM} = 0$. Тогда

$$S_I = S_{IM} + h_K P_K, \quad (5)$$

$$S_{II} = -f'(P_K - W_K), \quad (6)$$

$$S_{III} = \frac{f'^2}{h_K} (P_K - 2W_K) + f'. \quad (7)$$

Из выражений (6) и (7) следует, что при $S_{II} = 0$ параметр $P_K = W_K$, а коэффициент

$$S_{III} = f' \left(1 - \left(\frac{f'}{h_K} \right) W_K \right) = f' \left(1 - \left(\frac{f'}{h_K} \right) P_K \right).$$

Пусть $S_I = 0$. Тогда $P_K = - \left(\frac{S_{IM}}{h_K} \right)$. При этом коэффициент

$$S_{III} = f' \left(1 + \left(\frac{f'}{h_K^2} \right) S_{IM} \right). \quad (8)$$

Концентрический мениск определим углами осевого виртуального [6] (нулевого) луча с оптической осью в виде

$$\begin{aligned} \alpha_1 &= 0 & n_1 &= 1 \\ \alpha_2 &= \alpha_M & d &= d_M & n_2 &= n_M \\ \alpha_3 &= \alpha'_M & n_3 &= 1. \end{aligned}$$

Толщина мениска $d_M = r_1 - r_2$. Приближенно можно принять $r_2 \cong f' = -f'$. Тогда $r_1 = d_M - f'$.

Применив формулу $n_{i+1}\alpha_{i+1} - n_i\alpha_i = h_i \frac{n_{i+1} - n_i}{r_i}$, находим значение угла α_M :

$$\alpha_M = h_1 \frac{n_M - 1}{n_M r_1} = \frac{n_M - 1}{n_M} \frac{1}{d_M \varphi - 1}. \quad (9)$$

Угол $\alpha_M = h_1 \varphi_M = f' \varphi_M$.

Оптическая сила мениска

$$\varphi_M = (n_M - 1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right) + \frac{(n_M - 1)^2 d_M}{n_M r_1 r_2} = \frac{n_M - 1}{n_M} \frac{d_M \varphi^2}{d_M \varphi - 1}. \quad (10)$$

При этом

$$\alpha'_M = \frac{n_M - 1}{n_M} \frac{d_M \varphi}{d_M \varphi - 1} = \frac{n_M - 1}{n_M} + \alpha_M.$$

Коэффициент S_{IM} определим выражением

$$S_{IM} = h_1 P_1 + h_2 P_2 = f' P_1 + h_K P_2, \quad (11)$$

где $P = \left(\frac{\Delta \alpha}{\Delta v} \right)^2 \Delta \alpha v$;

$$h_K = s'_{f'} = f' (1 - f' \varphi_M) = f' \frac{n_M - \varphi d_M}{n_M (1 - \varphi d_M)}. \quad (12)$$

Полученные соотношения позволяют выражение (11) преобразовать к виду

$$S_{IM} = f' \frac{n_M - 1}{n_M^3} \frac{\varphi d_M}{(\varphi d_M - 1)^3} \left[n_M \varphi^2 d_M^2 - \frac{n_M^3 - 1}{n_M - 1} (\varphi d_M - 1) \right]. \quad (13)$$

Вполне очевидно, что при $d_M = 0$ коэффициент $S_{IM} = 0$. При этом в соответствии с формулой (8) коэффициент $S_{III} = f'$, что и следовало ожидать. Кривизна поверхности изображения, образованного системой тонкого компонента с концентрическим мениском, определяется коэффициентом S_{IV} , равным $S_{IV} = S_{IVM} + S_{IVK}$. В рассматриваемом случае коэффициенты S_{IVM} и S_{IVK} удобно определить формулами

$S_{IVM} = \sum_{i=1}^2 \frac{n_{i+1} - n_i}{n_{i+1} n_i r_i}$; $S_{IVK} = \sum_{i=1}^m \frac{\varphi_i}{n_i}$. Используя полученные соотношения, находим, что $S_{IVM} = \varphi_M$. При этом

$$S_{IV} = \varphi_M + \sum_{i=1}^m \frac{\varphi_i}{n_i}.$$

Приближенно принимаем $\sum_{i=1}^m \frac{\varphi_i}{n_i} \approx \frac{\varphi_K}{n_K} = \frac{\varphi}{n_K}$. Тогда $S_{IV} = \varphi_M + \frac{\varphi_M}{n_K}$. Применив формулу (10) и выра-

зив линейные величины в масштабе фокусного расстояния системы, т.е. при $\varphi = 1$, при $\sum_{i=1}^m \frac{\varphi_i}{n_i} = \pi$ получаем

$$d_M = \frac{n_M (S_{IV} - \pi)}{1 + n_M (S_{IV} - \pi - 1)}. \quad (14)$$

Полученные соотношения позволяют представить выражения (9), (12) и (13) в виде

$$\alpha_M = S_{IV} - \pi - \frac{(n_M - 1)}{n_M}; \quad (15)$$

$$h_K = -S_{IV} + \pi + 1; \quad (16)$$

$$S_{IM} = \frac{n_M}{(n_M - 1)^2} (S_{IV} - \pi)^3 - \frac{n_M^3 - 1}{n_M} \frac{(S_{IV} - \pi)^2}{(n_M - 1)^2} + \frac{n_M^3 - 1}{n_M^2} \frac{S_{IV} - \pi}{n_M - 1}, \quad (17)$$

при этом в соответствии с формулой (8)

$$S_{III} = 1 + \frac{S_{IM}}{(1 + \pi - S_{IV})^2}. \quad (18)$$

Из выражения (18) следует, что коэффициент $S_{III} = 0$ при условии

$$(1 + \pi - S_{IV})^2 + S_{IM} = 0.$$

Для тонкого компонента в виде одиночной линзы параметр $\pi = \frac{1}{n_K}$. Пусть $n_K = n_M$. При этом выражения (14)–(18) принимают вид

$$d_M = \frac{n_M S_{IV} - 1}{n_M (S_{IV} - 1)};$$

$$\alpha_M = S_{IV} - 1;$$

$$h_K = \frac{1 - n_M (S_{IV} - 1)}{n_M};$$

$$S_{IM} = \frac{n_M}{(n_M - 1)^2} (S_{IV} - 1)^3 - \frac{n_M - 1}{n_M} S_{IV}^2 + \frac{n_M^2 - 1}{n_M^2} S_{IV};$$

$$S_{III} = 1 + \frac{n_M^3}{(n_M - 1)^2} \frac{(S_{IV} - 1)^3}{[1 - n_M (S_{IV} - 1)]^2} + \frac{n_M - 1}{1 - n_M (S_{IV} - 1)} S_{IV}.$$

Легко убедиться, что при $S_{IV} = 0$ коэффициент $S_{III} = 0$ при $n_M \approx 1,905$.

В табл. 1 для ряда значений коэффициента S_{IV} приведены численные значения показателя преломления материала мениска и тонкой линзы в воздухе при $S_{III} = 0$.

S_{IV}	0	0,05	0,1	0,15	0,2	0,25
n_M	1,9052	1,8313	1,7606	1,6932	1,6290	1,5678

Таблица 1. Зависимость $n_M = n_M(S_{IV})$ при $n_K = n_M$

Из табл. 1 следует, что для построения оптической системы объектива рассматриваемой конструкции при плананастигматической коррекции aberrаций в видимой области спектра необходимо решить проблему выбора материала линз. При неполной коррекции кривизны поверхности изображения построение такой системы вполне возможно.

В соответствии с формулой (5) коэффициент $S_I = 0$ при значении параметра

$$P_K = -\frac{S_{IM}}{h_K} = -\frac{S_{IM}^2}{1 + \pi - S_{IV}}. \quad (19)$$

Из формулы (6) следует, что коэффициент $S_{II} = 0$ при $P_K = W_K$. В общем случае взаимосвязь этих параметров с основными параметрами тонкого компонента определяется выражениями [5]

$$P_K = (\alpha' - \alpha)^3 P_K^\infty + 4\alpha(\alpha' - \alpha)^2 W_K^\infty + \alpha(\alpha' - \alpha)[2\alpha(2 + \pi) - \alpha'], \quad (20)$$

$$W_K = (\alpha' - \alpha)^2 W_K^\infty + \alpha(\alpha' - \alpha)(2 + \pi). \quad (21)$$

Здесь угол $\alpha' = 1$, а угол $\alpha' = \alpha'_M$. Параметр P_K^∞ тонкой линзы в воздухе определяется выражением (4), которое можно представить в виде

$$P_k^\infty = P_{0k}^\infty + a(W_k^\infty - W_{0k}^\infty)^2, \quad (22)$$

$$\text{где } P_{0k}^\infty = \frac{n_k(4n_k-1)}{4(2+n_k)(n_k-1)^2}, \quad W_{0k}^\infty = \frac{1}{2(2+n_k)}, \quad a = \frac{n_k(2+n_k)}{(n_k+1)^2}.$$

Заменяя параметр P_k^∞ в выражении (20) выражением (22), при $P_k = -\frac{n_k S_{IV}}{1+n_k(S_{IV}-1)}$, где $n_k = n_m$,

получаем квадратное уравнение относительно величины параметра W_k^∞ . Подставив значение параметра W_k^∞ , полученное в результате решения этого уравнения, в выражение (21), найдем значение параметра W_k . В общем случае $W_k \neq P_k$. Тогда, дополнив рассматриваемую систему такой же, получим симметричную систему, формирующую изображение с поперечным увеличением $V = -1^\times$. Для перехода от симметричной системы к системе, формирующей изображение бесконечно удаленного предмета, можно применить метод сохранения углов излома луча осевого пучка, предложенный в [7]. В результате получим значения конструктивных параметров оптической системы объектива типа «Планар», которые можно рассматривать в качестве исходных для последующей оптимизации по критерию качества изображения.

Если одиночную положительную линзу тонкого компонента дополнить тонкой отрицательной линзой, то жесткая взаимосвязь основных параметров тонкого компонента P_k^∞ и W_k^∞ нарушается. В этом случае коррекция первичной комы изображения, сформированного образованной системой, вполне возможна.

Подставив выражение (17) в формулу (18), получаем

$$S_{III} = 1 - \frac{n_m}{(n_m-1)^2} \frac{(\pi - S_{IV})^3}{(1 + \pi - S_{IV})^2} + \frac{n_m^3 - 1}{n_m^2} \frac{\pi - S_{IV}}{(n_m-1)^2} \frac{1 - n_m(1 + \pi - S_{IV})}{(1 + \pi - S_{IV})^2}. \quad (23)$$

Для системы двух тонких линз из кронового и флинтowego стекол, имеющих средние значения показателя преломления и коэффициента дисперсии, в общем случае в первом приближении величину π можно принять равной 0,7. При этом в соответствии с выражением (23) при $S_{IV} = 0$ коэффициент $S_{III} = 0$ при $n_m \approx 2,37$. При применении кроновых стекол марок СТК и флинтowych стекол марок ТФ величину π можно принять равной 0,55. В этом случае при $S_{IV} = 0$ коэффициент $S_{III} = 0$ при $n_m \approx 1,98$. В оптических системах, предназначенных для работы в инфракрасной области спектра, применяют материалы, показатели преломления которых могут принимать значения от $n = 1,4$ (например, флюорит) до $n = 4$ (например, германий). При этом параметр π может оказаться равным примерно 0,4. В табл. 2 при трех значениях параметра π для ряда значений коэффициента S_{IV} приведены численные значения показателя преломления материала мениска при $S_{III} = 0$.

π	S_{IV}					
	0	0,05	0,10	0,15	0,20	0,25
0,70	2,380	2,240	2,100	1,970	1,841	1,716
0,55	1,970	1,841	1,716	1,596	1,482	1,370
0,40	1,596	1,480	1,370	1,274	1,185	1,110

Таблица 2. Зависимость $n_m = n_m(S_{IV}, \pi)$

Приведенные в табл. 2 величины могут служить ориентиром для выбора материала концентрического мениска по показателю преломления n_m при соответствующих значениях параметра π и коэффициента S_{IV} . Выполнив подстановки в соотношение (14), находим толщину d_m мениска. Формула (17) позволяет вычислить значение коэффициента S_{IV} . Применив формулу (19), можем определить значение параметра P_k , а следовательно, и параметра $W_k = P_k$. Решив систему уравнений (20) и (21), находим значения основных параметров P_k^∞ и W_k^∞ , которые позволяют оценить требуемую сложность конструкции тонкого компонента и вычислить его конструктивные параметры.

Заключение

Выполненные исследования позволили показать возможность коррекции аберраций в изображении, образованном оптической системой, состоящей из тонкого компонента и концентрического мениска конечной толщины, а числовые исследования подтвердили такую возможность. Применяя полученные аналитические соотношения, можно определить параметры рассматриваемой системы. Получена оптическая система объектива с вынесенным входным зрачком при телецентрическом ходе главных лучей в пространстве изображений. Показано, что при $S_I = 0$ и $S_{II} = 0$ значение коэффициента S_{III} не зависит от положения входного зрачка. В связи с этим в частном случае входной зрачок можно совместить, например, с первой поверхностью мениска.

Литература

1. Русинов М.М. Композиция оптических систем. – Л.: Машиностроение, Ленингр. отделение, 1989. – 383 с.
2. Зверев В.А. Идеи композиции как принцип построения рациональной конструкции оптической системы // Научно-технический вестник СПб ГИТМО (ТУ). – 2002. – Вып. 5. – С. 56–71.
3. Грамматин А.П., Демидова Е.А., Зверев В.А., Романова Г.Э. Аберрационные свойства оптической системы из двух отражающих поверхностей сферической формы с компенсатором // Оптический журнал. – 2004. – № 4. – С. 11–15.
4. Бронштейн И.Г., Лившиц И.Л., Kim Young-Gi, Kim Tae-Young, Jung Phil-No. Выбор оптической схемы и расчет малогабаритных объективов для мобильных телефонов // Оптический журнал. – 2009. – Т. 76. – № 5. – С. 25–31.
5. Слюсарев Г.Г. Методы расчета оптических систем. – Л.: Машиностроение, Ленингр. отделение, 1969. – 672 с.
6. Зверев В.А. Основы геометрической оптики. – СПб: СПб ГИТМО (ТУ), 2002. – 218 с.
7. Русинов М.М. Техническая оптика: Учебное пособие для вузов. – Л.: Машиностроение, Ленингр. отделение, 1979. – 488 с.

- Ежова Василиса Викторовна** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, evv_foist@mail.ru
- Зверев Виктор Алексеевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, post_vaz@rambler.ru
- Точилина Татьяна Вячеславовна** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, tvtochilina@mail.ru

УДК 535.421

МОДЕРНИЗИРОВАННЫЙ КОМПЛЕКС СИНТЕЗА И ВОССТАНОВЛЕНИЯ ГОЛОГРАММ-ПРОЕКТОРОВ ФРЕНЕЛЯ

С.Н. Корешев, А.Д. Громов, О.В. Никаноров

Представлен модернизированный вариант специализированного программного комплекса синтеза и восстановления голограмм-проекторов Френеля. Примененный алгоритм основан на новом методе синтеза голограмм-проекторов, заключающемся в разбиении объекта на типовые элементы его структуры. Достоинством метода является обеспечиваемая им возможность существенного сокращения объема требуемых вычислений. Проведена оценка производительности предлагаемого метода. Приведены результаты экспериментальной проверки работоспособности модернизированного комплекса.

Ключевые слова: голографическая проекционная фотолитография, голограмма-проектор, синтез голограмм, цифровое восстановление голограмм, алгоритм, программный комплекс.

Введение

В современной литературе имеются сведения о разработках голографических фотолитографических систем на основе практически всех известных видов голограмм, в частности, голограмм Френеля [1, 2], голограмм сфокусированного изображения [3, 4]. Особый интерес вызывает возможность использования синтезированных голограмм [5, 6]. Наиболее перспективными представляются системы, основанные на использовании отражательной рельефно-фазовой синтезированной голограммы-проектора Френеля (ГПФ), рассчитываемой с помощью компьютера и отображаемой на носителе с помощью лазерного или электронно-лучевого генератора изображения.

Основные трудности при изготовлении синтезированных ГПФ вызывают процессы синтеза и отображения на носителе рассчитанной голограммной структуры. В процессе синтеза объект представляется в виде работающего на пропускание бинарного двумерного дискретного транспаранта. Сложность синтеза ГПФ обуславливается, в первую очередь, необходимостью предельно точного математического описания процесса формирования голографического поля, не допускающего использования приближения Фраунгофера и Френеля, так как при выборе оптимальных параметров схемы предполагается использовать малое расстояние между плоскостью объекта и голограммы [7, 8] в сочетании с необходимостью минимизации влияния дискретного представления двумерного объекта и синтезированной голограммы, присущего цифровым голограммам, на качество восстанавливаемых с их помощью изображений [7].

Из доступной литературы, посвященной синтезу и отображению на носителе ГПФ, предназначенных для использования в фотолитографическом процессе, наибольший интерес вызывают работы [5, 6, 9]. Отметим, что демонстрация фотолитографического процесса, реализованного с помощью синтезированной ГПФ, была осуществлена в СПб НИУ ИТМО [9] с помощью разработанного авторами специализированного комплекса синтеза и восстановления ГПФ [10]. Реализованные в этом комплексе методы синтеза и восстановления основывались на принципе Гюйгенса–Френеля. В основу метода положено математическое описание физических процессов записи и восстановления ГПФ.

Соответствующий ему алгоритм синтеза моделирует физический процесс формирования голографического поля. При этом считается, что объект, представляющий собой работающий на пропускание бинарный двумерный транспарант, установленный параллельно плоскости синтеза голограммы на расстоянии h от нее, освещается параллельным пучком когерентного излучения, нормально падающего на поверхность транспаранта. В качестве опорной волны в рассматриваемом методе используется плоская волна, падающая под углом θ на плоскость синтеза ГПФ. Расчет распределения амплитуды объектной волны в плоскости синтеза голограммы в этом методе сводится к векторному суммированию комплексных амплитуд полей, формируемых точками объекта (рис. 1).

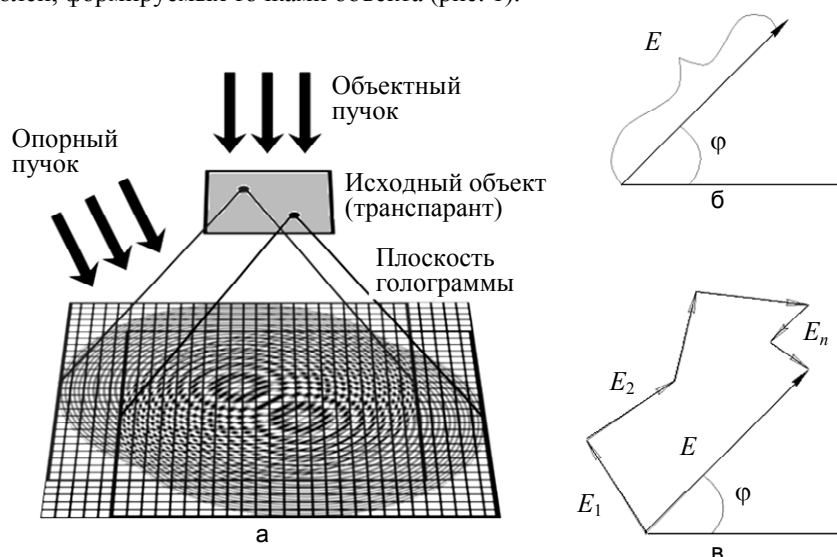


Рис. 1. Метод синтеза ГПФ на основе вычисления комплексного вектора электромагнитного поля, состоящий в суммировании комплексных амплитуд полей, формируемых точками объекта (а), в представлении комплексной амплитуды в виде вектора (б) и в суммировании комплексных амплитуд (в)

Амплитуда объектной волны, рассчитанная для каждой точки плоскости синтеза голограммы, складывается с соответствующей амплитудой опорной волны. Полученная сумма амплитуд возводится по модулю в квадрат, тем самым, определяется распределение интенсивности голографического поля, соответствующего синтезируемой ГПФ.

Описанный метод синтеза ГПФ обладает несомненным достоинством, заключающимся в возможности достижения высокой точности расчета структуры ГПФ. Вместе с тем он не свободен от недостатков. Основным из них, на взгляд авторов, является большой объем требуемых вычислений, особенно при сложной структуре объекта, например, при разводке жидкокристаллических дисплеев.

Метод синтеза ГПФ, основанный на использовании типовых элементов структуры объекта

Необходимость повышения производительности процесса синтеза ГПФ привела к разработке нового метода. В его основу положено разбиение исходного объекта на типовые элементы его структуры с последующим использованием в процессе синтеза голограммы заранее рассчитанных амплитуд волновых полей, сформированных этими типовыми элементами объекта. Суть алгоритма состоит в том, что

одинаковые элементы объекта, например точки, формируют одинаковые картины распределения комплексных амплитуд в плоскости синтеза голограммы до момента наложения опорной волны.

Таким образом, представляя объект в виде набора типовых элементов, при расчете амплитуды объектной волны в процессе синтеза голограммы можно ограничиться лишь расчетом локальной области распределения комплексной амплитуды для каждого типа элементов из этого набора. Это допустимо из-за дискретизации голограммы, приводящей к ограничению максимальной частоты, которая может быть записана [7]. Описываемую картину распределения комплексной амплитуды, полученную от типового элемента, условно назовем «штампом». Также стоит отметить, что для исключения несовмещения фрагментов штампов необходимо, чтобы отношение периода дискретизации объекта к периоду дискретизации голограммы было целым числом.

Впоследствии необходимо последовательно накладывать штампы в местах голограммы, соответствующих одинаковым элементам, на которые разбит объект; условно назовем этот метод «штамповкой». При пересечении штампов производится сложение значений комплексных амплитуд.

Голограммы, полученные с использованием нового алгоритма, представлены на рис. 2.

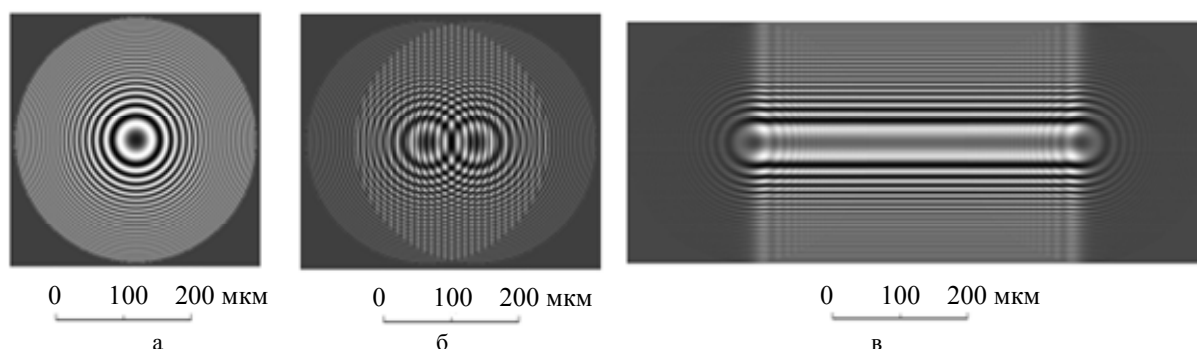


Рис. 2. Голограммы точки (а), двух точек (б) и линии (в), синтезированные с помощью предлагаемого метода

Оценка производительности предлагаемого метода

При сравнительной оценке производительности описываемого метода следует выделить два наиболее ресурсоемких показателя синтеза голограммы одной светящейся точки: P_e – количество машинных операций, необходимое для расчета распределения комплексной амплитуды от одной светящейся точки (расчет штампа); P_{sum} – количество операций, необходимое для записи штампа в матрицу, описывающую плоскость голограммы. Количество операций, необходимое для расчета голограммы из N элементов, при использовании предыдущего метода составляет

$$P_1 = (P_e + P_{sum})N,$$

а для предлагаемого –

$$P_2 = P_e + P_{sum}N.$$

Тогда снижение ресурсоемкости синтеза можно оценить как

$$k = \frac{(P_e + P_{sum})N}{P_e + P_{sum}N}.$$

При расчете голограмм особо сложных объектов величиной P_e в знаменателе можно пренебречь, тогда

$$k \approx \frac{(P_e + P_{sum})N}{P_{sum}N} = \frac{P_e}{P_{sum}} + 1.$$

Необходимо учитывать, что из-за отсутствия сложных операций (возведение в степень, извлечение квадратного корня и т.д.) P_{sum} намного меньше P_e .

При этом целесообразно проводить разбивку объекта на типовые элементы, обладающие различными видами симметрии, в результате чего возможно дополнительное сокращение времени, затрачиваемого на расчет штампов, так как распределение комплексной амплитуды у симметричных объектов также симметрично.

На рис. 3 проиллюстрирован типовой алгоритм расчета картины распределения комплексной амплитуды для элемента с центральной симметрией на примере вещественной части штампа. В данном случае необходимо провести расчет только одной «четверти» распределения. Полученную в итоге «четверть» необходимо скопировать с поворотом на остальные 3 «четверти». Это действие не приведет к дополнительному боковому смещению фрагментов при рекомендованном выше выборе соотношения дискретизации голограммы и объекта и повороте на угол, кратный 90° .

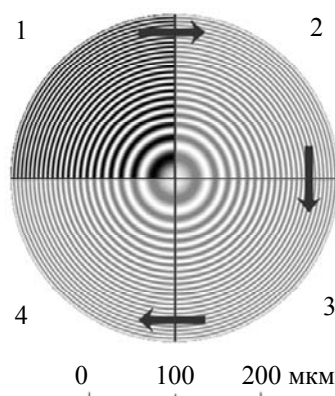


Рис. 3. Синтез картины распределения комплексной амплитуды элемента с центральной симметрией (на примере вещественной части штампа)

При практической оценке производительности с помощью комплексов, в которых реализованы сравниваемые методы, были синтезированы голограммы размером 1560×1560 пикселей с периодом дискретизации 1 мкм. В качестве объекта использовалось бинарное изображение размером 80×80 пикселей с периодом дискретизации 4 мкм. Площадь структуры составляла 20% от общей площади. Длина волны излучения λ была выбрана равной 488 нм, расстояние между плоскостью объекта и плоскостью регистрации голограммы h составило 5,062 мм, а угол падения опорного пучка θ был принят равным $10,5^\circ$. Выбор таких параметров предопределен результатами работы [7]. При оценке результатов было установлено, что использование модернизированного комплекса позволяет сократить время расчета тестовой голограммы с 1 часа 13 минут 40 секунд до 14 минут 26 секунд.

Проверка работоспособности модернизированного программного комплекса

Работоспособность предлагаемого метода синтеза была проверена экспериментально. С помощью программных комплексов, реализующих прежний и модернизированный методы, и при использовании параметров, выбранных в соответствии с результатами работы [7], синтезированы две отражательные рельефно-фазовые бинарные ГПФ, предназначенные для работы с длиной волны излучения $\lambda = 13,5$ нм. В качестве тестового использовалось бинарное изображение реперных знаков размером 23×23 пикселя (рис. 4) с периодом дискретизации $a = 80$ нм.

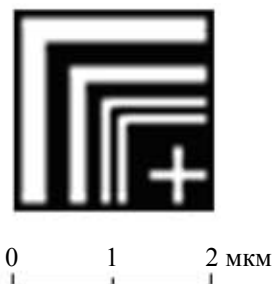


Рис. 4. Изображение исходного объекта

Размер синтезированных ГПФ составлял 435×435 пикселей с периодом дискретизации $d = 20$ нм, расстояние между плоскостью объекта и плоскостью регистрации голограммы h было принято равным 20,3 мкм, а угол падения опорного пучка θ составил $14,7^\circ$.

На рис. 5 представлены изображения бинаризированных голограмм, синтезированных с помощью сравниваемых программных комплексов.

Восстановление синтезированных бинарных рельефно-фазовых отражательных голограмм осуществлялось в виртуальном пространстве с помощью программного комплекса в соответствии с параметрами их синтеза. Восстановленные изображения (рис. 6) имеют схожий интервал уровней порога, имитирующего реакцию фоторезиста на засветку актиничным излучением. По величине этого интервала можно судить о сложности подбора экспозиции при физическом восстановлении, а значит, данная величина может быть выбрана критерием качества восстанавливаемого изображения [10].

Схожий интервал уровней порога (62 и 72 градации серого из 256) и идентичность структур после пороговой обработки свидетельствует о работоспособности модернизированного программного ком-

плекса, т.е. существует значение экспозиции, при которой можно получить изображение идентичное исходному объекту [10].

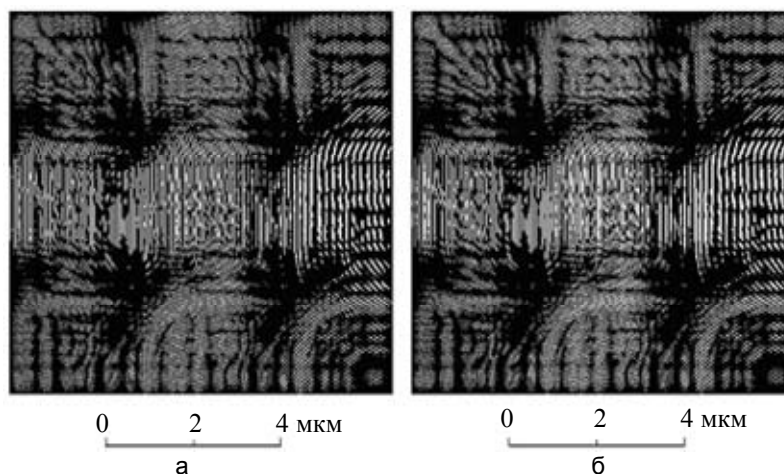


Рис. 5. Бинарные голограммы, синтезированные с помощью немодернизированного (а) и модернизированного программного комплекса (б)

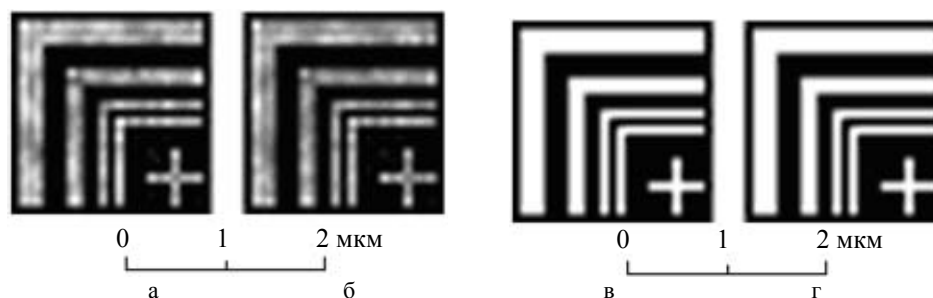


Рис. 6. Изображения, полученные при восстановлении голограммы, синтезированной немодернизированным комплексом (а) и модернизированным комплексом (б); те же изображения после пороговой обработки: немодернизированным комплексом (в) и модернизированным комплексом (г)

Заключение

Проведена модернизация программного комплекса синтеза голограммы-проектора Френеля. Применен метод «штамповки», в основу которого положено разбиение исходного объекта на типовые элементы его структуры и последующее использование в процессе синтеза голограммы заранее рассчитанных амплитуд волновых полей, сформированных этими типовыми элементами. Это позволило существенно сократить время синтеза голограммы-проектора Френеля. Представлена теоретическая и экспериментальная оценка производительности предлагаемого метода синтеза. Работоспособность комплекса подтверждена экспериментально в ходе сравнения результатов синтеза и восстановления голограмм, полученных при использовании исходного и модернизированного комплексов.

Литература

1. Корешев С.Н., Ратушный В.П. Использование метода голографии для получения изображений двумерных объектов при решении задач фотолитографии высокого разрешения // Оптический журнал. – 2004. – Т. 71. – № 10. – С. 32–39.
2. Koreshev S.N., Ratushnyi V.P. Holographic method for obtaining images with limiting high resolution for extreme shot-wave lithography problems // Proc. SPIE. – 2004. – V. 5290. – P. 221–232.
3. Clube F., Gray S., Struchen D., Tisserand J., Malfoy S., Darbellay Y. Holographic microlithography // Opt. Eng. – 1995. – V. 34. – № 9. – P. 2724–2730.
4. Корешев С.Н., Ратушный В.П. Голограммы сфокусированного изображения в задаче высокоразрешающей проекционной голографической фотолитографии // Оптика и спектроскопия. – 2006. – Т. 101. – № 6. – С. 1038–1042.
5. Jacobsen C., Howells M. Projection x-ray lithography using computer-generated holograms: A study of compatibility with proximity lithography // J. Appl. Phys. – 1992. – V. 71. – P. 2993–3001.

6. Naullenau P.P., Salmassi F., Cullikson E.M., Liddle J.A. Design and fabrication of a high-efficiency extreme-ultraviolet binary phase-only computer-generated hologram // Appl. Optics. – 2007. – V. 46. – № 14. – P. 258–2585.
7. Корешев С.Н., Никаноров О.В., Козулин И.А. Выбор параметров синтеза голограмм-проекторов для фотолинтографии // Оптический журнал. – 2008. – Т. 75. – № 9. – С. 29–34.
8. Домненко В.М., Бурсов М.В., Иванова Т.В. Моделирование формирования оптического изображения: Учебное пособие. – СПб: НИУ ИТМО, 2011. – 141 с.
9. Корешев С.Н., Никаноров О.В., Ратушный В.П. Восстановление синтезированных голограмм-проекторов Френеля при углах падения восстанавливающей волны, превышающих угол падения опорной волны при синтезе голограммы // Оптика и спектроскопия. – 2011. – Т. 111. – № 1. – С. 156–161.
10. Корешев С.Н., Никаноров О.В., Иванов Ю.А., Козулин И.А. Программный комплекс для синтеза и цифрового восстановления голограмм-проекторов: влияние параметров синтеза на качество восстановленного изображения // Оптический журнал. – 2010. – Т. 77. – № 1. – С. 42–48.

- Корешев Сергей Николаевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, Koreshev@list.ru
- Громов Александр Дмитриевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Warhammer@bk.ru
- Никаноров Олег Викторович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, инженер-программист, Nikanorovov@gmail.com

УДК 53.083

МЕТОД ОПРЕДЕЛЕНИЯ ПРОСТРАНСТВЕННОГО ПОЛОЖЕНИЯ ОБЪЕКТА ВБЛИЗИ ВОДНОЙ ПОВЕРХНОСТИ

Д.Ю. Крысин

Предложен бесконтактный метод измерения пространственного положения объектов вблизи зеркальной водной поверхности (балльность волнения 0 баллов), основанный на использовании технологий технического зрения. Особенностью метода является возможность одновременного измерения высоты и угловой ориентации объекта. Рассмотрен алгоритм работы измерительной системы. Приведено описание разработанной установки, предназначенной для исследования предложенного метода. При проведении экспериментов достигнутая точность измерения высоты в диапазоне 500–1000 мм составила 2–3%. Погрешность измерения угловой ориентации при этом не превышала 10%. Рассмотрены источники погрешностей и способы повышения точности измерений.

Ключевые слова: измерение высоты, водная поверхность, система технического зрения, пространственное положение, экспериментальная установка.

Введение

Управление гидросамолетами имеет ряд особенностей по сравнению с управлением сухопутными самолетами и требует от пилота особой подготовки и навыков. Одним из специфических и наиболее сложных элементов пилотирования гидросамолетов является посадка на гладкую невзволнованную водную поверхность (балльность волнения 0 баллов). Визуальная оценка истинной высоты полета в таких условиях является крайне проблематичной. При посадке гидросамолета на зеркальную водную поверхность необходимо руководствоваться показаниями приборов, а также надводными и наземными ориентирами.

В соответствии с действующими Федеральными авиационными правилами полетов в воздушном пространстве Российской Федерации (пункт 134 главы XXIII «Полеты гидросамолетов»), «...посадка гидросамолета в штиль при зеркальной поверхности воды производится на гидроаэродром со специально оборудованным стартом. При отсутствии на гидроаэродроме такого старта посадка разрешается гидросамолетам, оснащенным аппаратурой, обеспечивающей посадку на зеркальную поверхность» [1]. Данное правило продиктовано, прежде всего, необходимостью обеспечения безопасности полетов. Очевидно, основным блоком указанной аппаратуры должен быть высотомер малых высот. Однако нужно отметить, что большинство небольших гидросамолетов в базовой комплектации не оснащаются высотомерами малых высот. Это объясняется двумя основными причинами.

Первой и самой важной причиной является отсутствие высотомеров малых высот в списке обязательного авиационного оборудования для данного класса летательных аппаратов. Вторая причина состоит в относительно высокой стоимости данного оборудования. На больших гидросамолетах широко используются радиовысотомеры [2]. Но на малых гидросамолетах их использование в большинстве случаев нецелесообразно с экономической точки зрения. Отсутствие точной приборной информации о высоте полета вблизи зеркальной поверхности может стать (и иногда становится) причиной летных происшеств-

вий. Рассмотренные факты позволяют сделать вывод об актуальности поиска альтернативных способов измерения истинной высоты подвижных объектов вблизи невзволнованной водной поверхности.

Постановка задачи

Автором было выполнено исследование особенностей основных методов бесконтактного измерения дальности, используемых в авиации и промышленности. Рассматривались следующие системы: барометрические высотомеры, радиотехнические высотомеры, емкостные высотомеры, измерители пространственного положения на базе приемников сигналов глобальных спутниковых навигационных систем GLONASS и GPS, гамма-лучевые дальномеры, ультразвуковые дальномеры, лазерные дальномеры, системы технического зрения (СТЗ) [3]. В большей части случаев проведенный анализ выявил невозможность прямого использования перечисленных систем для измерения истинной высоты полета малых гидросамолетов вблизи зеркальной водной поверхности. В то же время технологии технического зрения являются более молодой областью и развиваются очень интенсивно. Одной из задач, успешно решаемых с помощью СТЗ, является бесконтактное измерение пространственного положения и формы объектов.

В предлагаемой работе рассматриваются возможности бесконтактной системы измерения истинной высоты объекта вблизи невзволнованной водной поверхности на базе технологий технического зрения. Основные требования к системе: диапазон измеряемых высот 0,2–5 м, погрешность измерения – 2–3% от измеряемой высоты, постоянная времени 0,5 с [4–6].

Особенности применения СТЗ

В настоящий момент нет готовых решений задачи измерения высоты объекта вблизи водной поверхности с использованием технологий технического зрения. Существующие пассивные стереоскопические СТЗ, 3D-сканеры на базе структурированного света, времяпролетные камеры (tof-камеры) имеют общую особенность, существенно ограничивающую их функциональность. Она состоит в том, что для их нормальной работы желательно, чтобы объекты, попадающие в поле зрения камер, отражали свет диффузно. Без существенного изменения программно-алгоритмического и аппаратного обеспечения данные системы не позволяют корректно проводить измерения в случаях, когда объекты сцены имеют зеркальные поверхности относительно большой площади. Этот факт объясняется тем, что при съемке зеркальных поверхностей на снимках фиксируются не только объекты, но и зеркальные изображения объектов, попадающие в поле зрения этих камер [7]. Фотоснимки являются двумерными изображениями трехмерной сцены. Зеркала же позволяют наблюдать зеркальные изображения объектов (мнимые или действительные в зависимости от формы зеркала). При фотографировании зеркал на снимки попадают изображения объектов, которых на самом деле может не быть в поле зрения камер. По этой причине можно утверждать, что при фотографировании зеркал фиксируются фотоизображения зеркальных изображений объектов. Рассмотренное явление при определенных обстоятельствах может привести к невозможности функционирования СТЗ. Наличие прозрачных и частично прозрачных поверхностей в поле зрения камер также может оказать негативное влияние на работоспособность и функциональность СТЗ. Использование рассмотренных СТЗ для измерения относительного пространственного положения водной поверхности также сопряжено с описанными сложностями. Это объясняется тем, что водная поверхность является частично прозрачной для световых лучей, а при отсутствии сильных возмущений обладает свойством зеркального отражения [8].

В результате экспериментальных и теоретических исследований автором был предложен метод определения пространственного положения объекта вблизи зеркальной водной поверхности, в основе которого лежит использование теории технического зрения и свойства зеркального отражения водной поверхности.

Структурная схема СТЗ

Структурная схема предложенной СТЗ представлена на рис. 1. СТЗ устанавливается на объект, пространственное положение которого необходимо контролировать. Блок подсветки 1 в общем случае состоит из массива опорных световых элементов, излучающих рассеянный свет в направлении зеркальной водной поверхности. Отразившись от поверхности, часть света попадает в объективы системы цифровых камер 4. Камеры фиксируют стереоскопические изображения зеркальной водной поверхности. Одновременно на снимках фиксируется зеркальное отражение объекта вместе с опорными световыми элементами СТЗ. Блок сопряжения 3 необходим для организации обмена информацией с вычислительным блоком 2. Вычислительный блок выполняет обработку полученных цифровых стереоскопических снимков, распознавание опорных световых элементов СТЗ и вычисление пространственного положения объекта относительно зеркальной водной поверхности. Здесь и далее под пространственным положением объекта понимается его высота и угловая ориентация относительно водной поверхности. Возможность настройки камер и блока подсветки придает измерительной системе свойство адаптивности. Адаптивность необходима для реакции на изменяющиеся внешние условия и положение объекта в пространстве.



Рис. 1. Структурная схема СТЗ

Описание метода измерения

Иллюстрация метода измерения представлена на рис. 2. Рисунок содержит три ряда изображений.

В верхнем ряду показано вертикальное перемещение некоторого объекта относительно невзволнованной водной поверхности (истинная высота увеличивается слева направо). На нижней плоскости объекта установлена стереоскопическая СТЗ, структурная схема которой приведена на рис. 1. Для наглядности блок подсветки представляет собой единственный световой элемент, расположенный между двумя камерами. Световой элемент представляет собой источник рассеянного света со сферической линзой. Было теоретически и экспериментально доказано, что минимальное количество световых элементов, необходимое для получения оценок высоты и угловой ориентации объекта, равно одному.

В среднем ряду рисунка представлены фотоизображения водной поверхности, фиксируемые левой от наблюдателя камерой СТЗ. На снимках за счет свойства зеркального отражения света, которым обладает водная поверхность, фиксируется зеркальное отражение нижней поверхности объекта. При этом фиксируются и видимые элементы СТЗ – световой элемент и две камеры. Изображение нижней плоскости объекта имеет форму прямоугольника. Изображение светового элемента находится в центре и имеет форму светлого круга. Радиально от него расположены две камеры, имеющие форму черных квадратов.

В нижнем ряду рисунка представлены изображения, попадающие в поле зрения и зафиксированные правой камерой. Изображения подобны тем, что были зафиксированы левой камерой.

В качестве дополнительной иллюстрации на рис. 3 приведена упрощенная схема съемки, поясняющая работу СТЗ. Она содержит следующие элементы: S_L и S_R – центры фотографирования левой и правой камер соответственно; S_L, X_L, Y_L, Z_L и S_R, X_R, Y_R, Z_R – прямоугольные фотограмметрические системы координат (ФСК) камер; L – точечный световой элемент; W – плоская поверхность, зеркально отражающая свет; I – мнимое изображение светового элемента L ; H – высота светового элемента относительно зеркальной поверхности; 1 и 2 – световые лучи; R_L и R_R – точки отражения лучей 1 и 2 от зеркальной поверхности W ; F_L и F_R – плоскости снимков камер; P_L и P_R – фотоизображения светового элемента на снимках, фиксируемых камерами.

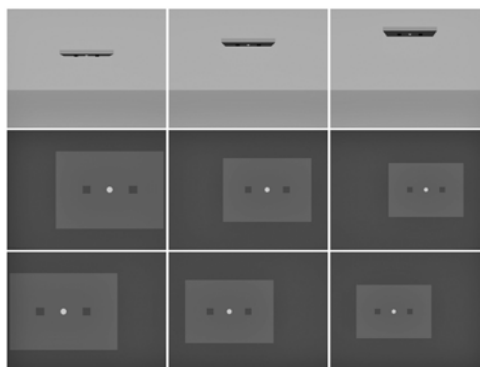


Рис. 2. Перемещение объекта и соответствующие стереоизображения: верхний ряд – положение объекта относительно зеркальной поверхности; средний ряд – соответствующие снимки левой камеры стереосистемы; нижний ряд – соответствующие снимки правой камеры стереосистемы

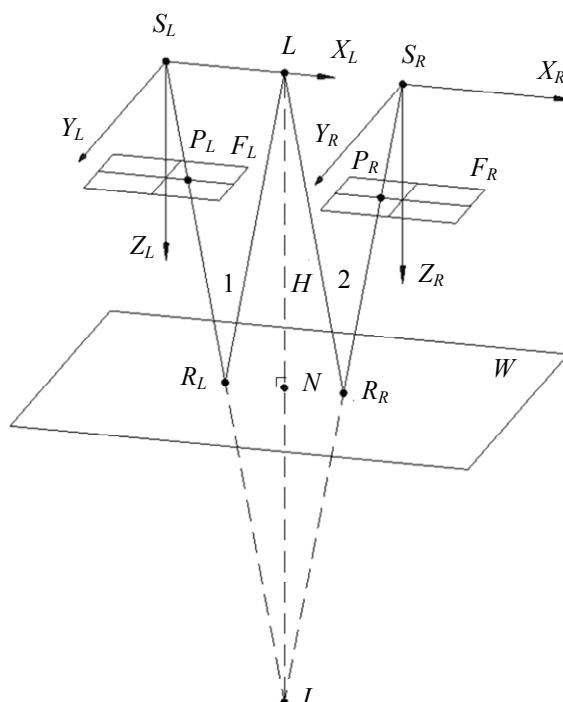


Рис. 3. Упрощенная схема съемки

Рис. 2, 3 показывают, что масштаб изображений, фиксируемых стереоскопическими камерами, изменяется в зависимости от высоты. При изменении угловой ориентации объекта изображения также будут изменяться. Суть предложенного метода измерения заключается в том, что с помощью обработки стереоскопических изображений можно вычислить пространственное положение объекта относительно зеркальной поверхности. Отдельно следует отметить, что предложенный метод обеспечивает измерение не только истинной высоты, но и угловой ориентации объекта относительно водной поверхности. Априорными данными при этом являются элементы внутреннего и внешнего ориентирования камер стереосистемы и их пространственное расположение относительно световых элементов.

Задача измерения положения объекта в пространстве сводится к задаче измерения положения СТЗ, точнее, к задаче измерения положения ФСК камер СТЗ. Зная положение ФСК относительно зеркальной поверхности и способ установки СТЗ на объекте, можно вычислить искомое пространственное положение объекта относительно водной поверхности.

В соответствии с предложенным методом пространственное положение объекта относительно невозмущенной водной поверхности может быть определено, если выполняются следующие условия:

- мнимое зеркальное изображение, по крайней мере, одного светового элемента блока подсветки попадает в область перекрытия камер системы;
- фотограмметрические координаты светового элемента известны;
- фотоизображение светового элемента на стереоскопических снимках поддается распознаванию.

Алгоритмическое обеспечение

Обобщенная блок-схема предложенного алгоритма вычисления пространственного положения объекта представлена на рис. 4. В соответствии с приведенной блок-схемой алгоритм разбивается на 6 основных этапов.

Этап 1. Получение стереоскопических изображений водной поверхности с помощью цифровых камер СТЗ.

Этап 2. Предварительная обработка изображений, сегментация изображений, распознавание и определение плоских координат фотоизображений опорных элементов объекта. Сегментация является наиболее ответственной процедурой. При проведении экспериментов положительные результаты были получены с помощью цветовой сегментации в векторном пространстве RGB на базе вычисления расстояния Махаланобиса [9]

$$D_M(z, m) = \sqrt{(z - m)^T C^{-1} (z - m)},$$

где C^{-1} – матрица, обратная к ковариационной матрице репрезентативной выборки векторов RGB; m – математическое ожидание репрезентативной выборки векторов RGB; z – координаты произвольной точки пространства RGB.



Рис. 4. Обобщенная блок-схема алгоритма вычисления пространственного положения объекта

Этап 3. Ассоциирование фотоизображений световых элементов с их оригиналами. В качестве опорных могут быть использованы любые контрастные элементы объекта, попадающие на стереоизображения. Источники света имеют преимущество по причине простоты распознавания и обеспечения работы системы в условиях недостаточного естественного освещения.

Этап 4. Вычисление фотограмметрических координат мнимых зеркальных изображений световых элементов. Данная задача сводится к определению пространственных координат точки по координатам ее фотоизображений на стереоснимках [10]. В теории фотограмметрии эта операция называется прямой пространственной фотограмметрической засечкой. Исходными данными являются координаты световых элементов на полученных снимках, а также элементы внутреннего и внешнего ориентирования стереосистемы камер.

Этап 5. Вычисление пространственного положения объекта относительно водной поверхности. Если предположить, что L_X, L_Y, L_Z – координаты светового элемента в ФСК одной из камер (рис. 3), а I_X, I_Y, I_Z – координаты его мнимого изображения в той же системе координат, полученные на этапе 4, то высота H СТЗ, равная половине длины перпендикуляра, проведенного из центра светового элемента к зеркальной поверхности, может быть вычислена по формуле [7, 11]

$$H = \frac{1}{2} \cdot \sqrt{(I_X - L_X)^2 + (I_Y - L_Y)^2 + (I_Z - L_Z)^2}.$$

Угловая ориентация ФСК относительно зеркальной поверхности вычисляется с использованием свойств скалярного произведения векторов. Выражение для вычисления угла наклона одной из осей ФСК относительно зеркальной плоскости имеет вид

$$\alpha = \arccos \frac{(I_X - L_X)}{\sqrt{(I_X - L_X)^2 + (I_Y - L_Y)^2 + (I_Z - L_Z)^2}} - \frac{\pi}{2}.$$

Для остальных осей углы наклона вычисляются аналогично. Углы крена и тангажа могут быть вычислены с использованием элементов матрицы направляющих косинусов.

Этап 6. Визуальный вывод информации о высоте и угловой ориентации объекта. При необходимости возможно формирование оповещений о незапланированном режиме движения объекта.

Экспериментальная установка

Для проверки алгоритмического и программного обеспечения была разработана и изготовлена экспериментальная установка (рис. 5). Установка включает следующие основные части: СТЗ (две цифровые USB-камеры, блок подсветки и источник питания); персональный компьютер, используемый в качестве вычислительного блока; штатив; резервуар с водой. В качестве световых элементов блока подсветки используются полупроводниковые светоизлучающие диоды со сферической линзой и матовой рассеивающей поверхностью. Большая часть опытов проводилась при использовании одного светодиода, расположенного между камерами. Как было сказано выше, одного светового элемента и системы из двух камер достаточно для измерения и высоты, и угловой ориентации объекта относительно зеркальной водной поверхности.

Программно-алгоритмическое обеспечение было реализовано в средах MATLAB и Simulink. Его тестирование проводилось на стереоизображениях, полученных экспериментально с помощью разработанной установки.

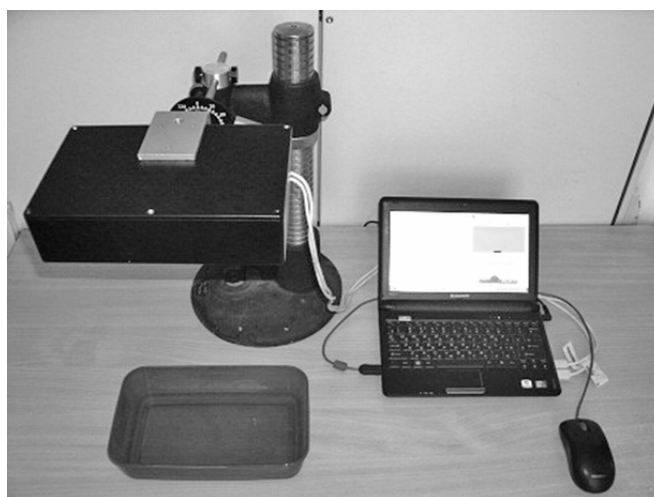


Рис. 5. Экспериментальная установка

Экспериментальные результаты

Заключение

Эксперименты по оценке точности измерений проводились в диапазоне высот 300–1000 мм. Углы наклона СТЗ относительно плоскости горизонта находились в пределах $\pm 10^\circ$. Достигнутая точность измерения высоты относительно зеркальной водной поверхности составила 8–15 мм (2–3% от измеряемой величины). Погрешность измерения углов наклона не превышала 1° . Период дискретизации измерений составлял 600–700 мс.

Проанализированы факторы, влияющие на точность измерений. Выявлена значительная зависимость погрешности от точности определения пространственного положения опорных световых элементов в ФСК. После монтажа СТЗ необходимые координаты известны только приблизительно, что обусловлено производственной неточностью установки световых элементов и камер относительно несущей конструкции СТЗ, а также отсутствием точной информации о положении оптических центров фотографирования камер. По этой причине была разработана процедура калибровки, предназначенная для определения фотограмметрических координат световых элементов. При калибровке вместо водной поверхности в качестве вспомогательного элемента используется плоское зеркало. Оно устанавливается в заранее известные пространственные положения, в которых фиксируются стереоснимки. С помощью обработки полученных стереоснимков вычисляются искомые координаты световых элементов. По существу, разработанная процедура является способом определения пространственного положения объектов, находящихся вне поля зрения стереосистемы камер. После ее выполнения точность вычисления высоты относительно зеркальной поверхности была повышена на 3–5%. Предполагается, что данный результат может быть улучшен с помощью решения задач оптимизации процесса калибровки.

Дальнейшее повышение точности измерений возможно за счет использования камер с лучшими техническими характеристиками, повышения точности калибровки камер, увеличения числа камер и световых элементов, повышения точности определения фотограмметрических координат световых элементов, применения теории фильтрации при обработке снимков и вычислении пространственного положения, использования дополнительной информации о динамике объекта.

Результаты исследования предложенного в работе метода позволяют сделать вывод о возможности применения технологий технического зрения для измерения пространственного положения объектов вблизи зеркальной водной поверхности. Особенностью метода является то, что он обеспечивает измерение не только истинной высоты, но и угловой ориентации объекта. Предложено алгоритмическое обеспечение. Рассмотрена экспериментальная установка, разработанная для исследования предложенного метода. Приводятся полученные результаты оценки его эффективности. Проанализированы источники погрешностей и способы повышения точности измерений. Планируется, что дальнейшее исследование будет направлено на расширение диапазона условий эксплуатации измерительной системы путем оптимизации ее конструкции и программно-алгоритмического обеспечения.

Литература

1. Приложение 6 к конвенции о международной гражданской авиации. Эксплуатация воздушных судов. Часть I. Международный коммерческий воздушный транспорт. Самолеты. – 9-е изд. – Международная организация гражданской авиации, 2010. – 258 с.
2. Радиовысотомер А-037. Руководство по технической эксплуатации. – ГУ1.000.051 РЭ, 1983. – 287 с.
3. Клюев Г.И. и др. Авиационные приборы и системы: Учебное пособие. – Ульяновск: УЛГТУ, 2010. – 343 с.
4. Nebylov A.V. et al. Sea Wave Parameters, Small Altitudes and Distances Measurers Design for Motion Control Systems of Ships, Wing-in-Surface Effect Craft and Seaplanes // AGARD-NATO Conference Proceedings NO 556. Dual Usage of Military and Commercial Technology on Guidance and Control. – AGARD-NATO, Neuilly-sur-Seine, France, 1995. – 12 p.
5. Nebylov A.V., Wilson P. Ekranoplane – Controlled Flight close to the Sea. Monograph. – WIT-Press, Southampton, UK, 2002. – 312 p.
6. Небылов А.В. Измерение параметров полета вблизи морской поверхности. – СПб: СПбГААП, 1994. – 308 с.
7. Шрёдер Г., Трайбер Х. Техническая оптика. – М.: Техносфера, 2006. – 423 с.
8. Шифрин К.С. Введение в оптику океана. – Л.: Гидрометеиздат, 1983. – 280 с.
9. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
10. Назаров А.С. Фотограмметрия. – Минск: ТетраСистемс, 2006. – 368 с.
11. Немченко К.Э. Аналитическая геометрия. – М.: Эксмо, 2007. – 352 с.
12. George M., Sukkarieh S. Tightly Coupled INS/GPS with Bias Estimation for UAV Applications // Proceedings of Australian Conference on Robotics and Automation. – 2005. – 7 p.

Крысин Дмитрий Юрьевич

– Санкт-Петербургский государственный университет аэрокосмического приборостроения, аспирант, wtxt@ya.ru

УДК 530.145

ВРАЩАТЕЛЬНАЯ ЗАВИСИМОСТЬ РЕЗОНАНСОВ ТИПА КОРИОЛИСА И ФЕРМИ В МАЛЫХ ЛИНЕЙНЫХ МОЛЕКУЛАХ

М.А. Смирнов

Исследована зависимость эффективных параметров в модельном гамильтониане изучаемой проблемы от молекулярных констант с точки зрения применяемой при построении гамильтониана схемы упорядочения колебательно-вращательных взаимодействий. Сформулированы соотношения для генераторов преобразования в различных схемах упорядочения колебательно-вращательных взаимодействий при помощи техники «распутывания» экспоненциальных операторов. При помощи техники проекционных операторов выделен оператор резонансного взаимодействия из оператора общего вида в эффективном вращательном гамильтониане. Показано, что при сильном резонансном взаимодействии зависимость от упорядочения колебательно-вращательных взаимодействий более слабая, нежели при среднем и слабом резонансах.

Ключевые слова: колебательно-вращательные взаимодействия, резонансы Ферми и Кориолиса, схемы упорядочения колебательно-вращательных взаимодействий.

Введение

Изучение вращательной зависимости ангармонических колебательных и колебательно-вращательных (КВ) случайных резонансов представляет значительный интерес в анализе динамики колебания и вращения молекул. Одной из болезненных точек в описании молекулярных спектров остается область сильных случайных резонансов, КВ анализ в которой дает худшие статистические показатели по сравнению с нерезонансными областями или областями слабых резонансов. Другой важной проблемой анализа энергетического спектра молекул является предсказание поведения КВ уровней при больших энергиях вращения молекулы. Рассмотрение этих двух проблем с позиций новых моделей для описания вращательной зависимости случайных резонансов Ферми- и Кориолисова типов для трехатомных линейных молекул и является предметом обсуждения в этой работе.

В линейных молекулах типа CO_2 (симметрии $D_{\infty h}$) и HCN (симметрии $C_{\infty v}$) гармоническое силовое поле имеет особенность, заключающуюся в близости резонанса частоты валентного колебания типа $\Sigma(+, g)$ и двойной частоты изгибного колебания типа Π . Кроме того, существует также резонанс частот валентных колебаний, Σ_g^+ и Σ_u^- . Какие резонансные операторные члены в матрице энергии будут связывать соответствующие этим частотам невозмущенные диагональные матричные элементы и каково их место в иерархии операторов в эффективном гамильтониане (ЭГ) – предмет обсуждения в настоящей работе.

На важность вращательной зависимости чисто колебательного резонанса было впервые указано Аматом и Пимбертом [1] при изучении Ферми-резонанса в CO_2 . Недиagonalный матричный элемент, связывающий колебательные состояния $|V_1, V_2, I_2, V_3\rangle$ и $|V_1 - 1, V_2 + 2, I_2, V_3\rangle$, может быть представлен в виде [1]

$$W = \frac{1}{2} \left[-K_{122} / \sqrt{2} + \delta J(J+1) \right] v_1^{1/2} \left[(V_2 + 2)^2 - I_2^2 \right]^{1/2}.$$

Параметр δ во вращательной зависимости резонанса трактовался в ранних работах как варьируемый параметр и был определен из анализа спектров для ряда молекул HCN , ClCN [2]. Было отмечено [2, 3] влияние этого резонанса на вращательные постоянные колебательных состояний, участвующих в резонансе. Аномалия во вращательной постоянной B_v колебательных состояний $(0, 2^0, 0)$ и $(1, 0^0, 0)$ HCN была успешно объяснена Ферми-резонансом между этими уровнями [2]. Ванг и др. в работе [3] указали на то, что знание δ может быть успешно использовано для оценки ангармонических постоянных K_{223} и K_{2212} из пересечения B и W кривых в случае резонанса между уровнями $(1, 0^0, 0)$, $(0, 2^2, 0)$ и $(0, 0^0, 1)$. Это утверждение не совсем верно: как будет показано ниже, если уровни $(1, 0^0, 0)$ и $(0, 2^0, 0)$ связаны резонансом Ферми, то уровни $(1, 0^0, 0)$, $(0, 2^2, 0)$ и $(0, 0^0, 1)$ связаны резонансом типа Кориолиса, а вращательные зависимости этих резонансов имеют различную природу и аналитические выражения для параметров δ (Ферми-тип) и γ (Кориолисов тип).

Дэвис и Оверенд [4] были первыми, кто пытался объяснить происхождение параметра δ , используя технику метода возмущений. И если выражения для δ , полученные ими, были неполными, тем не менее, ими было достигнуто удовлетворительное воспроизведение порядков величин экспериментально определенных δ . Наиболее основательный подход к теоретическому определению δ был предпринят в работе [5], опираясь на метод построения контактными преобразованиями ЭГ, развитый Аматом, Нильсеном, Голдсмитом [6]. Авторы работы [5] правильно указали, что происхождение δ может быть просле-

жено из тех операторов колебательно-вращательного гамильтониана (КВГ), чьи матричные элементы имеют зависимость от квантовых чисел вида $-v_1^{1/2}[(V_2+2)^2 - I_2^2]^{1/2} J(J+1)/2$, т.е. операторы вида $J^2 q_1 q_2^2$ будут давать вклады в δ .

В этой работе была исследована вращательная зависимость двух случайных резонансов в молекулах типа CO_2 и HCN , $v_1(\Sigma_g^+) \approx 2v_2$ и $v_1(\Sigma_g^+) + v_2(\Pi_u) \approx v_3(\Sigma_u^-)$. Если первый резонанс относится к Ферми типу резонансов, который связывает колебательные уровни одного типа симметрии, то второй относится к Кориолисову типу резонансов, который связывает вращательные подуровни колебательных уровней разных типов симметрии. В предыдущей работе [7] на основе концепции связанных схем упорядочения КВ взаимодействий в молекуле [8] для линейных молекул был развит новый подход к описанию вращательной зависимости случайных резонансов ранга 3 [9]. Был построен методом контактных преобразований (КП) ЭГ для изучаемой проблемы в виде бесконечного тейлоровского ряда по степеням углового момента $J_\perp^2$. Это представление ЭГ в виде ряда по $J_\perp^2$ стало возможным благодаря применению одной из предельных схем упорядочения, соответствующей модели сверхбыстрого ротатора [8]. Было отмечено, что этот ряд можно трактовать как разложение в ряд Лоррана некоторой аналитической функции углового момента.

Схемы упорядочения колебательно-вращательных взаимодействий в молекулах

Прежде чем применять последовательные КП к КВ гамильтониану, описывающему вращательную зависимость тройного межмодового случайного резонанса, необходимо выяснить:

1. к какому порядку по λ отнести операторный член H_{mn} в $H^{(2N)}$ или, по-другому, как соотносить формальное разложение $H_{VR} = H^{(0)} + \sum_n \lambda^n H^{(n)}$ и разложение в форме $H = H_{20} + \sum_{mn} H_{mn}$ КВ гамильтониана квазижесткой молекулы;
2. ввиду того, что $S^{(N)}$ операторы зависят от колебательных и от вращательных операторов, возникает вопрос, к какому порядку по λ отнести колебательные и вращательные коммутаторы, возникающие из общего коммутатора,

$$[S, h] = [S_V S_R, h_V h_R] = [S_V, h_V] \frac{1}{2} [S_R, h_R]_+ + [S_R, h_R] \frac{1}{2} [S_V, h_V]_+,$$

где $S_V(h_V)$ и $S_R(h_R)$ – колебательные и вращательные множители в $S(h)$, $[A, B]_+ = AB + BA$.

Исходя из этого, приходим к проблеме упорядочения возмущений H_{mn} в H_{VR} . Таким образом, в теории КВ спектров квазижестких молекул необходимо решать проблему соответствия $H^{(N)} \Rightarrow H_{mn}$ формального разложения гамильтониана в теории возмущений и фактического разложения КВГ.

Разложение КВГ (обратного тензора инерции и потенциальной функции) в ряд по ядерным смещениям в системе обозначений, предложенных Ватсоном, имеет вид

$$H_{VR} = \sum_{mn} H_{mn} = H_{vib} + H_{cor} + H_{rot},$$

где H_{mn} – группа членов степени m по колебательным операторам (q_l или p_l) и степени n по вращательным операторам (J_α). Коэффициенты в H_{mn} имеют порядок величины

$$\chi^{m-2-2n} \omega_{vib}, \quad (1)$$

где χ – параметр Борна–Оппенгеймера $(m_e/m_n)^{1/4} \approx 1/10$.

Для малых значений квантовых чисел гамильтониан совокупности гармонических осцилляторов H_{20} дает доминирующий вклад в матрицу КВ энергии. Разработанный в работах Михайлова [8, 10] и Ватсона [9] подход основан на концепции упорядочения КВ взаимодействий в квазижестких молекулах в зависимости от порядков величин колебательных и вращательных операторов. В операторной формулировке метода возмущений, в частности, метода КП, удается найти точные аналитические соотношения операторов в ЭГ $\tilde{H}_{VR} = \sum \tilde{H}_{mn}$, построенных в разных схемах упорядочения КВ взаимодействий. Физические и математические принципы теории связанных схем упорядочения КВ взаимодействий развиты в работе [11].

Если « λ -порядок» операторов H_{mn} определить в виде $\alpha m + \beta n$, где α и β – рациональные числа, то « λ -порядки» в схемах упорядочения, применявшихся в теории спектров молекул, определяются следующим образом [11]:

$$\langle \lambda \text{-порядок} \rangle H_{mn}(S_{mn}) = \begin{cases} m + \varepsilon n & - & \text{Ватсон (W)} \\ m + n & - & \text{Амат – Нильсен (A – N)} \\ m + 2n & - & \text{Борн – Оппенгеймер – Ока (B – O – O)} \\ \varepsilon m + 2n & - & \text{Михайлов (M)} \end{cases}.$$

Условия, накладываемые на порядки величин колебательных и вращательных операторов и коммутаторов, для схем упорядочения, сформулированных выше в уравнениях (1), приведены в таблице.

	(W)	(B-O-O)	(A-N)	(M)
R	1	1	1	$\chi^{-1+\varepsilon}$
J	$\chi^{-2+\varepsilon}$	1	χ^{-1}	1
$[p, q]_V$	$-i$	$-i$	$-i$	$-i\chi^{-2+2\varepsilon}$
$[J_\alpha, J_\beta]_R$	$-i\chi^{-2+\varepsilon} J_\gamma$	$-iJ_\gamma$	$-i\chi J_\gamma$	$-iJ_\gamma$

Таблица. Порядки величин колебательных и вращательных операторов в различных схемах упорядочения KB взаимодействий

Определение случайных резонансов в спектре

Определим случайные резонансы в молекулярном энергетическом спектре с помощью функции от гармонических частот [11]

$$\Phi(\omega) = \sum_{i=1}^P m_i \sigma_i \omega_i, \quad (2)$$

где ω_i – гармонические частоты колебаний; m_i – натуральные числа; $\sigma_i = \pm 1$ – знаковые переменные; P – полиада колебательных состояний. Случайный резонанс в молекулярном энергетическом спектре определим условием на $\Phi(\omega)$ функцию

$$\Phi(\omega) \approx 0 \quad \Theta(\omega).$$

Эквивалентная форма записи этого условия, используемая в спектроскопической литературе, такова:

$$\sum_{j=1}^r m_i \omega_i \approx \sum_{k=r+1}^P m_{r+j} \omega_{r+j}. \quad (3)$$

Техникой проекционных операторов может быть выделен оператор резонансного взаимодействия из оператора общего вида в эффективном вращательном гамильтониане

$$V = \sum_{q=\sum_{i=1}^P m_i} C_{n_1 n_2 \dots n_q}^{\sigma_{n_1} \sigma_{n_2} \dots \sigma_{n_q}} a_{n_1}^{\sigma_{n_1}} a_{n_2}^{\sigma_{n_2}} \dots a_{n_q}^{\sigma_{n_q}},$$

где $a_{n_i}^{\sigma_{n_i}} = q_i - i\partial/\partial q_i$ – лестничные операторы; q_i – нормальные координаты; $C_{n_i}^{\sigma_{n_i}}$ – параметры, а суммирование проводится по полиадам. Общий вид проекционного оператора дан в [10]. Тогда нерезонансная часть оператора V , равная $V^{(*)}$, получается из (3) введением $(1-\Delta)$ символа, т.е.

$$V^{(*)} = V - V(Res) = V(1-\Delta).$$

Для тройного ($\omega_a \approx \omega_b + \omega_c$ или $\omega_a \approx 2\omega_b$) случайного резонанса $\Delta^{(P)}$ символ определяется выражением

$$\Delta^{(3)} = \delta_{\sigma_a \sigma_{n_1}} \delta_{\sigma_b \sigma_{n_2}} \delta_{\sigma_c \sigma_{n_3}} \delta_{\sigma_a, -\sigma_b},$$

где n_1, n_2, n_3 – индексы суммирования в операторе взаимодействия V .

Как было установлено в [10], определенные операторы в ЭГ могут быть представлены в виде

$$\hat{H}_{mn}^{(g)} = \hat{H}_{mn}^{(g')} + h_{mn}^{(g \rightarrow g')}.$$

Величины $h_{mn}^{(g \rightarrow g')}$ являются, по сути, теоретической неопределенностью упорядочения операторных элементов ЭГ. Эта величина обращается в нуль в приближении изолированного колебательного состояния и зависит в явном виде от функции $\tilde{\Phi}(\omega)$ (2). На основании определения коэффициентов величины

$$\hat{\Phi}(\omega) = \left[h_{mn}^{(M \rightarrow W)} \right]_{n_1 \dots n_m}^{\sigma_1 \dots \sigma_m} = h_{n_1 \dots n_m}^{\sigma_1 \dots \sigma_m; \alpha_1 \dots \alpha_n} \cdot \prod_{i=1}^m a_{m_i}^{\sigma_i} \prod_{j=1}^n J \alpha_j,$$

где $\sigma_i = \pm 1$; $\alpha_i = x, y, z$ (или $0, \pm 1$),

$$\hat{\Phi}(\omega) = {}_{(M \rightarrow W)} h_{n_1 \dots n_m}^{\sigma_1 \dots \sigma_m; \alpha_1 \dots \alpha_n}$$

может быть открыта следующая классификация случайных резонансных взаимодействий по типам: сильное $\Phi(\omega) = 0$, среднее $\Phi(\omega) \approx O(\omega)$, слабое $\Phi(\omega) \approx \omega$.

Рассмотрим тройной межмодовый резонанс на примере резонанса $\omega_1 + \omega_2 \approx \omega_3$ для $\tilde{H}_{31}$ оператора линейных молекул типа XYZ:

$$\tilde{\Phi}(\omega)_{\text{triad}} = {}_{(M \rightarrow W)} h_{123}^{\sigma\sigma-\sigma} = \sum_{m, \sigma_m} \frac{\Phi_{12m}(-R_m^3 + \gamma R_m^3)(\omega_1 + \omega_2 - \omega_3)\sigma_m}{(-\omega_3 + \gamma\omega_m)(\omega_1 + \omega_3 - \gamma\omega_m)}. \quad (4)$$

Из (4) следует результат, заключающийся в том, что при строгом равенстве $\omega_1 + \omega_2 = \omega_3$

$$\tilde{\Phi}(\omega)_{\text{triad}} = {}_{(M \rightarrow W)} h_{123}^{\sigma\sigma-\sigma} = 0,$$

можно показать [10], что при сильном резонансе зависимость от упорядочения КВ взаимодействий более слабая, нежели при среднем и слабом резонансах.

Эффективный гамильтониан для Ферми- и Кориолисова типов случайных резонансов в линейных молекулах типа CO₂ и HCN

Для рассматриваемого типа трехатомных линейных молекул гамильтониан гармонического осциллятора определяется уравнением [7]

$$H_{02} = \frac{1}{4} \sum_{n=1,3} \omega_n a_n^{\sigma} a_n^{-\sigma} + \frac{1}{8} \omega_2 \sum_{\sigma=\pm 1} a_2^{\sigma} a_2^{-\sigma-\tau},$$

где лестничные операторы даются выражениями $a_n^{\sigma} = q_n - i\sigma p_n$; $a_2^{\sigma} = a_{2x}^{\sigma} + i\tau a_{2y}^{\sigma}$ и выполняются коммутационные соотношения $[a_n^{\sigma}, H_{20}] = -\sigma \omega_n a_n^{\sigma}$, $[a_i^{\sigma}, a_j^{\sigma'}] = 1/4 \delta_{ij}(\sigma - \sigma')(\tau\tau' - 1)$.

Используя трансформационные свойства введенных лестничных операторов [11], для рассматриваемого типа линейных молекул в схеме упорядочения Ватсона может быть построен ЭГ, описывающий вращательную зависимость Ферми- и Кориолисова типов резонансов в этих молекулах.

Для Ферми резонанса $\omega_1 \approx \omega_2$ теоретико-групповой анализ операторных членов H_{mn} в группировке Ватсона приводит к следующему виду для гамильтониана:

$$H_{(Fermi)}^{(a)} = H_{30} + \sum_{n=0}^{\infty} \tilde{H}_{32n}^{(a)} = \left\{ \Phi_{221} + {}_{(JJ)} h_{221}^{(a)} J_{\perp}^2 + {}_{(JJJ)} h_{221}^{(a)} J_{\perp}^4 + \dots \right\} a_2^{\sigma} a_2^{-\sigma-\tau} a_1^{-\tau},$$

где ${}_{(JJ)} h_{221}^{(a)} J_{\perp}^2, \dots$ – коэффициенты в H_{mn} и $J_{\perp}^2 = J^2 - J_z^2 = 1/2 \sum_{\tau} J_{\tau} J_{-\tau}$.

Кроме того Ферми-резонанс описывается в гамильтониане операторными членами типа вращательного l -резонанса, изученного Аматом, Нильсеном [12]. Соответствующий вклад в ЭГ от этого типа членов может быть представлен в виде

$$H_{(Fermi)}^{(e)} = \tilde{H}_{32} + \sum_{n=1}^{\infty} \tilde{H}_{32n}^{(e)} = \left\{ {}_{(JJ)} h_{221}^{(e)} + {}_{(JJJ)} h_{221}^{(e)} J_{\perp}^2 + \dots \right\} (a_2^{\sigma})^2 a_1^{-\sigma} a_1^{-\sigma} J_{-\tau}^2.$$

Кориолисов резонанс в молекулах такого типа, $\omega_1 + \omega_2 \approx \omega_3$, может быть исследован таким же способом, как и Ферми-резонанс, и соответствующий вклад в гамильтониан имеет вид

$$H_{(Coriolis)} = \tilde{H}_{31} + \sum_{n=1}^{\infty} \tilde{H}_{32n+1} = \left\{ {}_{(J)} h_{123} + {}_{(JJ)} h_{123} J_{\perp}^2 + \dots \right\} a_1^{\sigma} a_2^{\sigma} a_3^{-\sigma} J_{-\tau}.$$

В ЭГ, описывающем вращательную зависимость Ферми- и Кориолисова типов резонансов, в рассматриваемых трехатомных линейных молекулах определяющими в группировке (W), являются только три типа операторов, которые можно представить в виде $a_1^{-\sigma} a_2^{\sigma} a_2^{-\sigma-\tau} f(J_{\perp}^2)$, $a_1^{\sigma} a_2^{\sigma} a_3^{-\sigma} \phi(J_{\perp}^2)$, $a_1^{-\sigma} (a_2^{\sigma})^2 \rho(J_{\perp}^2)$. Функции f , ϕ и ρ от $J_{\perp}^2$ по построению методом возмущений определяются соответствующими выражениями, приведенными выше в круглых скобках в ЭГ, $H_{\text{eff}} = H^{(a)}(F) + H^{(e)}(F) + H(C)$.

В этой модели для рассматриваемых случайных резонансов в трехатомных линейных молекулах, построенной на основе предельной схемы упорядочения (W), присутствуют только три типа матричных элементов по квантовому числу проекции колебательного момента двукратно вырожденной моды:

1. $H^{(a)}(F) \rightarrow \langle l | H_{(a)}(F) | l \rangle$;
2. $H^{(e)}(F) \rightarrow \langle l_k | H_{(e)}(F) | l \pm 2, k \pm 2 \rangle$;
3. $H(C) \rightarrow \langle l_k | H_{(e)}(F) | l \pm 1, k \pm 1 \rangle$.

Эти три типа членов, описывающие всю картину вращательной зависимости Ферми и Кориолисова типов резонансов в рассматриваемых молекулах, удовлетворяют следующим правилам отбора по квантовому числу $G = k - l$, $\Delta(k - l) = \Delta G = 0$, которые следуют из инвариантности гамильтониана относительно операций эрмитовости, обращения времени и преобразований элементов групп симметрии $D_{\infty h}$ и $C_{\infty v}$ [11].

Эффективный гамильтониан для возбужденных состояний

Как было отмечено во введении, можно показать, что возбужденные уровни $(1, 0^0, 0)$ и $(0, 2^0, 0)$ связаны резонансом Ферми, а уровни $(1, 0^0, 0)$, $(0, 2^2, 0)$ и $(0, 0^0, 1)$ связаны резонансом Кориолиса. Действительно, используя трансформационные операторы повышения и понижения a_2^{σ} , a_1^{σ} , a_3^{σ} [11] и тот факт,

что моды v_2 и v_3 относятся к типу П, а моды v_1 и v_3 соответственно к типам симметрии Σ_g^+ и Σ_u^- , причем индексы «+» и «-» имеют место для HCN типа молекул, а индексы «g» и «u» – для CO₂ типа линейных молекул, можно найти отличные от нуля матричные элементы от операторов $\tilde{H}_{mn}$ в ЭГ для указанных состояний. В доминантном приближении отличны от нуля следующие матричные элементы $\tilde{H}_{mn}$ операторов: $\langle 10^0 0 | H_{30} | 02^0 0 \rangle$, $\langle 00^0 1 | H_{30} | 02^0 0 \rangle$, $\langle 10^0 0 | \tilde{H}_{32} | 02^0 0 \rangle$, $\langle 00^0 1 | \tilde{H}_{32} | 02^0 0 \rangle$, $\langle 10^0 0 | \tilde{H}_{21} | 00^0 1 \rangle$. В молекулах типа HCN и CO₂ существуют тройные межмодовые резонансы $\langle 00^0 1 | \tilde{H}_{31} | 11^0 0 \rangle$, связывающие состояния $(1, 1^1, 0)$ и $(0, 0^0, 1)$.

Проведем анализ оператора $\tilde{H}_{31}^{(g)}$ с точки зрения теории связанных схем упорядочения КВ взаимодействий. Применим метод КП для построения $\tilde{H}_{31}$. Коммутаторные выражения $\tilde{H}_{31}^{(g)}$ в предельных группировках (M) и (W) и группировке (A–N) могут быть представлены в виде

$$\tilde{H}_{31}^{(g)} = \tilde{H}_{31} + \tilde{h}_{31}^{(группировка)} \in ,$$

где

$$\tilde{H}_{31} = H_{31} + i[S_{30}, \tilde{H}_{21}] .$$

В последнем уравнении S_{30} – колебательный генератор преобразования метода КП, $H_{21} = \tilde{H}_2 + (H_{21})$ – оператор кориолисова взаимодействия, записанный в виде суммы диагональной H_{21} и недиагональной $\tilde{H}_{21}$ частей в базисе H_{20} , $[\tilde{H}_{21}, H_{20}] = \tilde{H}_{21}H_{20} - H_{20}\tilde{H}_{21} = 0$.

Для линейных молекул $\tilde{H}_{21} = 0$ и H_{31} оператор в разложении H_{VR} имеет вид [13]

$$H_{31} = \frac{1}{8B_e} \left(\sum_{S\sigma} B_S a_S^\sigma \right) H_{21} = \frac{1}{8B_e} \sum_{S\sigma} \zeta_{S\tau} B_S \omega_S \frac{(\sigma' \omega_{S'} - \sigma'' \omega_{S'}) \tau}{\sqrt{\omega_{S'} \omega_{S'}}} a_S^\sigma a_{S'}^{\sigma'} a_{S''}^{\sigma''} J_\tau ,$$

где $\zeta_{S\tau}$ – кориолисовы постоянные; B_e – вращательные постоянные; ω_i – гармонические частоты колебаний; a_S^σ , $J_\tau = J_x + i\tau J_y$ – лестничные операторы.

Используя технику «распутывания» экспоненциальных операторов [11, 14] с учетом соотношений $H_{30} = i[S_{30}, H_{20}]$, $\tilde{H}_{21} = H_{21} + i[S_{21}, H_{20}]$ и тождеств Якоби для генераторов S_{21} , S_{30} , $S_{31}^{(g)}$, можно установить следующие соотношения для генераторов преобразования $S_{31}^{(g)}$ и $H_{31}^{(g)}$ операторов в различных группировках КВ взаимодействий:

$$S_{31}^{(g)} = S_{31}^{(g')} + S_{31}^{(g \rightarrow g')} ; \quad (5)$$

$$S_{31}^{(g \rightarrow g')} = C_{31}^{(g \rightarrow g')} [S_{21}, S_{30}] , \quad (6)$$

где коэффициенты $C_{31}^{(g \rightarrow g')}$ равны

$$C_{31}^{(g \rightarrow g')} = \begin{cases} 1 & \text{для } (M \rightarrow W) \\ \frac{1}{2} & \text{для } (M \rightarrow AN) \\ -\frac{1}{2} & \text{для } (W \rightarrow AN) \end{cases} \quad (7)$$

Исходя из (5)–(7), можно получить, что для генераторов $S_{31}^{(g)}$ в группировке Амата–Нильсена и группировках Ватсона и Михайлова имеет место уравнение

$$S_{31}^{(AN)} = \frac{1}{2} \{ S_{31}^{(M)} + S_{31}^{(W)} \} .$$

Операторы $\tilde{H}_{31}^{(g)}$ в предельных группировках связаны соотношением

$$\tilde{H}_{31}^{(W)} = [[S_{21}, S_{30}], H_{20}] + \tilde{H}_{31}^{(M)} = h_{31}^{(W \rightarrow M)} + \tilde{H}_{31}^{(M)} .$$

Детальное выражение для $h_{31}^{(W \rightarrow M)}$ через молекулярные постоянные для любой молекулы приведено в работе [8].

Выражение оператора $\tilde{H}_{32}$ в группировке (W) через коммутаторы S_{mn} и H_{mn} приведено в работе [9], соотношение для генераторов $S_{32}^{(g)}$ в предельных группировках, полученное техникой «распутывания» экспоненциальных операторов, дано в [11].

Можно показать, исходя из точных аналитических выражений для $\tilde{H}_{31}^{(g)}$ и $H_{31}^{(M \rightarrow W)}$, что для молекул типа CO_2 указанные операторные члены равны нулю для резонанса $\omega_1 \approx 2\omega_2$, однако они отличны от нуля для молекул типа HCN при записи их для резонанса типа $\omega_1 + \omega_2 \approx \omega_3$. Приведем точные аналитические выражения для параметров в $H_{31}^{(W)}$ и $\tilde{H}_{31}^{(M \rightarrow W)}$ через молекулярные постоянные для рассматриваемого тройного резонанса $\omega_1 + \omega_2 \approx \omega_3$ для трехатомных линейных молекул типа XY_2 и XYZ . Наиболее простое выражение для $\tilde{H}_{31}^{(g)}$ получается в группировке (W), так как оно не содержит ангармонических частотных знаменателей. Для молекул типа XYZ в случае $\omega_1 + \omega_2 \approx \omega_3$ резонанса оно имеет вид

$$\tilde{H}_{31} = \sum_{\sigma\tau} h_{123}^{\sigma\tau-\sigma} (\sigma\tau) a_1^\sigma a_2^{\sigma\tau} a_3^{-\sigma} J_{-\tau} = h_{123} q_1 q_3 (q_{2x} J_y + q_{2y} J_x),$$

$$^{(W)}h_{123}^{\sigma\sigma\tau-\sigma} = \frac{1}{8} \frac{B_e \zeta_{32}}{\sqrt{\omega_3 \omega_2 (\omega_3 - \omega_2^2)}} \left[\Phi_{123} (\omega_3^2 + \omega_2^2) + \Phi_{122} (4\omega_3 \omega_2) \right] + \frac{\zeta_{32} B_1 (\omega_3 + \omega_2)}{8 B_e \sqrt{\omega_3 \omega_2}}$$

и

$$^{(W)}h_{123}^{\sigma\sigma\tau-\sigma} = h_{123}^x = h_{123}^y.$$

Выражение упрощается для симметричных трехатомных молекул. В этом случае для кориолисовых постоянных и вращательных производных имеют место следующие равенства [13]:

$$\zeta_{32} = 0, \quad \zeta_{31} = 1, \quad c_1 = \left(\frac{2B_e}{\omega_1} \right)^{3/2}, \quad c_3 = 0.$$

Здесь введено альтернативное определение c_1 и c_2 для вращательных производных [12]:

$$B_1 = c_1 \omega = \sqrt{\frac{(2B_e)^3}{\omega_1}}. \quad (8)$$

С учетом соотношений (8) выражение $^{(W)}h_{123}$ для молекул типа CO_2 принимает вид

$$^{(W)}h_{123} = \frac{1}{16} \frac{(2B_e)}{\sqrt{\omega_2 \omega_3 (\omega_3 - \omega_2^2)}} \left(\Phi_{123} (\omega_3^2 + \omega_2^2) + \Phi_{122} 4\omega_3 \omega_2 \right) + \frac{1}{4} \sqrt{\frac{2B_e}{\omega_1}} \frac{\omega_3 + \omega_2}{\sqrt{\omega_3 \omega_2}}.$$

Заключение

В работе автор попытался проанализировать вращательную зависимость ангармонических колебательных и колебательно-вращательных случайных резонансов с точки зрения динамики колебания и вращения молекул. На основе введенной классификации случайных резонансов исследована зависимость эффективных параметров в модельном гамильтониане от молекулярных констант для резонансов Ферми и Кориолиса с точки зрения применяемой при построении гамильтониана схемы упорядочения колебательно-вращательных взаимодействий. Сформулированы соотношения для генераторов преобразования в различных схемах упорядочения колебательно-вращательных взаимодействий при помощи техники «распутывания» экспоненциальных операторов. Выделен оператор резонансного взаимодействия из оператора общего вида в эффективном вращательном гамильтониане. Показано, что при сильном резонансном взаимодействии, зависимость от упорядочения колебательно-вращательных взаимодействий более слабая, нежели при среднем и слабом резонансах.

Литература

1. Amat G., Pimbert M. On Fermi resonance in carbon dioxide // J. Mol. Spectrosc. – 1965. – V. 16. – P. 278.
2. Wang V.K., Overend J. The general quartic force field of HCN // Spectrochim. Acta. – 1976. – V. 32. – P. 1043.
3. Wang V.K., Goplen T.G., Overend J. Determination of anharmonic potential constants in linear XYZ molecules // J. Mol. Spectrosc. – 1975. – V. 46. – P. 509.
4. Davis K.A., Overend J. The rotational dependence of purely vibrational anharmonic resonances // Spectrochim. Acta. – 1976. – V. 32. – P. 1571.
5. Mishra K.C., Mohanty B.S. Rotational dependence of Fermi resonance in HCN and ClCN // J. Chem. Phys. – 1978. – V. 69. – P. 2064.
6. Amat G., Goldsmith M., Nielsen H.H. Higher order rotation-vibration energies of polyatomic molecules // J. Chem. Phys. – 1957. – V. 27. – P. 838.
7. Mikhailov V.M., Smirnov M.A. On the Rotational Dependence of Fermi Type Resonance Interactions in Molecules // SPIE. – 1996. – V. 3090. – P. 135–142.
8. Михайлов В.М. Микроволновая спектроскопия и ее применения // Научный совет по спектроскопии. – М.: АН СССР, 1985. – С. 235–328.

9. Aliev M.R., Watson J.K.G. *Molecular Spectroscopy: Modern Research* // Academic Press. – 1985. – V. 2. – P. 2–67.
10. Михайлов В.М. Схемы упорядочения (группировки) колебательно-вращательных возмущений в квазижестких молекулах // *Опт. атмосф. и океана*. – 2001. – Т. 14. – № 1. – С. 20–33.
11. Алиев М.Р., Михайлов В.М. Колебательно-вращательные спектры молекул // *Научный совет по спектроскопии*. – М.: АН СССР, 1987. – С. 120.
12. Amat G., Nielsen H.H. Rotational distortion in linear molecules arising from *l*-type resonance // *J. Mol. Spectrosc.* – 1958. – V. 2. – P. 163–172.
13. Watson J.K.G. Higher-order *l*-doubling of linear molecules // *J. Mol. Spectrosc.* – 1983. – V. 101. – P. 83–93.
14. Papousek D., Aliev M.R. *Molecular Vibration-Rotational Spectra*. – Elsevier, 1982. – 320 p.

Смирнов Максим Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, доцент, windy_hub@mail.ru

УДК 535.42; 535.417

ИТЕРАЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ ФАЗОВОЙ ПРОБЛЕМЫ В ОПТИКЕ И ИХ ОСОБЕННОСТИ

С.С. Налегает, Н.В. Петров, В.Г. Беспалов

Представлена общая концепция решения фазовой проблемы с использованием дополнительных наборов данных, которыми являются распределения интенсивности в плоскости формирования изображения, полученные при изменении одного или нескольких параметров установки. Предложенная концепция позволяет легко модифицировать итерационные методы восстановления волнового фронта в соответствии со спецификой решаемых задач. Ключевыми положениями являются внесение существенных различий в регистрируемые распределения интенсивности путем изменения варьируемого параметра в процессе их записи и использование математической модели, точно описывающей процесс распространения излучения через установку. С этой точки зрения проанализированы возможности использования нелинейных оптических эффектов для восстановления фазы волнового фронта.

Ключевые слова: фазовая проблема, восстановление фазы волнового фронта, нелинейность, нелинейные оптические эффекты.

Введение

При регистрации изображения в виде пространственного распределения интенсивности его светового поля происходит потеря важной информации о фазе рассеянной объектом волны. Возникает так называемая фазовая проблема в оптике – проблема восстановления формы волнового фронта объекта из измеренных распределений интенсивности [1]. В настоящее время существует множество методов ее решения, равно как и классификаций этих методов – по типу исследуемых объектов, наложенных ограничений, использования априорной информации об объекте и др. Различают детерминированный подход [2, 3], когда решение может быть получено аналитически, и итерационный подход [4, 5], когда информация о фазе восстанавливается в ходе последовательных приближений. В рамках этой работы ограничимся рассмотрением итерационных методов, проведя сравнительный анализ с точки зрения использования в них дополнительных массивов данных, позволяющих восстановить фазовую информацию. Отметим лишь некоторые недавние работы [6, 7], использующие детерминированный подход.

Результаты расчетов волновых полей с использованием методов решения фазовой проблемы находят широкое применение в различных областях науки и техники: в астрономии [5], рентгенографии [8] и электронной микроскопии [9]. Их используют для решения задач регистрации [10] и коррекции [11] аберраций волновых фронтов, измерения форм и деформаций объектов [12], в фазовой оптической микроскопии [13, 14] для задач биологии и медицины.

В настоящее время возможны подходы к решению фазовой проблемы, использующие нелинейные эффекты в процессе восстановления фазы волнового фронта. Это открывает возможности по разработке методов микроскопии с превосходящим дифракционным пределом разрешения [15, 16] за счет восстановления динамики поля, претерпевающего нелинейные превращения, и реконструкции пропущенных пространственных частот [16]. Однако реализовать это в полной мере мешает модуляционная неустойчивость и шум, поэтому определение разрешения, достижимого на практике, является открытой задачей.

Современные схемные решения с использованием итерационного алгоритма для восстановления волнового фронта

Ключевой идеей современных методов восстановления фазы волнового фронта оптического излучения является запись не двух распределений интенсивностей, как в алгоритмах Гершберга–Сакстона–Фиенапа [4, 5], а целого набора, характеризуемого различиями определенных параметров. Схемные реше-

ния с использованием итерационного алгоритма восстановления фазы можно представить блок-схемой (рис. 1). Основой является система излучения 1. Здесь могут быть использованы источники различного спектрального диапазона (от терагерцового диапазона частот до УФ); монохроматические [10–14] или мультиспектральные [17–20] (в этом случае система излучения может содержать компоненты, осуществляющие перестройку по длинам волн). Объект исследования 2 может быть самым разнообразным: пропускающим [21] и отражающим [22]; амплитудным при наличии фазовых шумов; чисто фазовым, амплитудно-фазовым или самосветящимся [5].

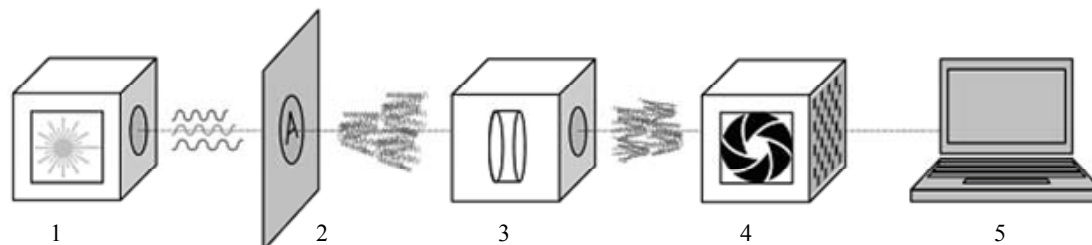


Рис. 1. Блок-схема систем итерационного восстановления волнового фронта: 1 – система излучения; 2 – исследуемый объект; 3 – фазовый анализатор; 4 – система регистрации; 5 – персональный компьютер

Рассеянное объектом излучение попадает на фазовый анализатор (фазовый транспарант) 3. В общем случае здесь подразумевается элемент или система, ответственные за изменение фазы волнового фронта в процессе его распространения. В методах восстановления фазы в качестве такого анализатора может выступать линза, сложная оптическая система формирования изображения [23], различные элементы, выполняющие линейные преобразования [24], случайный амплитудный или фазовый экран [10, 25], наконец, свободное пространство [21, 22]. Кроме того, здесь могут быть использованы пространственные модуляторы света или различные дифракционные оптические элементы, например, формирующие оптические вихри (спиральные пучки). Использование оптических вихрей также позволяет избежать стагнации итерационного алгоритма [26]. В методах восстановления фазы с нелинейными оптическими средами в качестве анализатора 3 используются элементы, выполняющие различного рода нелинейные преобразования, например, нелинейный пространственный фильтр (фазовый экран) в виде системы из двух линз и тонкого слоя среды, обладающей керровской нелинейностью, помещенный в общую фокальную плоскость линз [25].

Важную роль играет и система регистрации 4, в которой обычно присутствуют затвор и матричный фотоприемник. Для случая фазового анализатора в виде свободного пространства система регистрации должна располагаться на моторизированной подвижке, обеспечивающей возможность регистрации в различных сечениях пространства. Все элементы связаны с персональным компьютером 5, с которого осуществляется управление всей системой и обработка записанных массивов данных. Итерационный алгоритм, используемый для извлечения полной информации о волновом фронте из распределений интенсивности, также может быть различным. В зависимости от математической модели многократно повторяющейся процедурой может быть последовательный расчет распространения волнового фронта от объекта до экрана и назад [17–20, 27], расчет распространения назад и вперед только в отдельной части установки (например, в системе регистрации [12, 13, 21, 22]), независимый расчет с различными значениями варьируемого параметра с последующим усреднением [24] и т.п.

В соответствии с условиями задачи можно комбинировать различные элементы. Например, для наблюдения прозрачных объектов, проводя регистрацию в различных плоскостях, целесообразно использовать фазовый анализатор в виде рассеивающего диффузора, поскольку в этом случае возникает большая засветка, и начинают доминировать дифракционные эффекты, возникающие на краях апертур, а интенсивность высоких пространственных частот, непосредственно переносящих информацию о фазе волнового фронта и отвечающих в изображении за мелкие детали, может быть плохо различимой.

Роль варьируемых параметров при восстановлении фазы волнового фронта

Для эффективного восстановления фазы волнового фронта, рассеянного объектом, необходимо наличие математической модели, описывающей процесс прохождения оптического излучения через всю систему с достаточной точностью. При этом большое значение приобретает минимизация погрешностей измерений, учет имеющихся на практике отклонений в значениях параметров установки от используемых в численном расчете. Для минимизации погрешностей на стадии отладки установки могут быть использованы дополнительные алгоритмы численной обработки.

Проиллюстрируем сказанное на следующем примере. Пусть имеется установка, в которой варьируемыми параметрами являются расстояние от объекта до экрана и длина волны. Изменение параметров реализуется за счет перемещения системы регистрации и переключений в системе излучения, а в качестве фазового анализатора выступает свободное пространство [17–20]. Предположим, существуют некото-

рые отклонения в значениях вводимых в расчет параметров, например, с недостаточной точностью определен размер пикселя регистрирующей матрицы, неточно определено расстояние от объекта до плоскости регистрации и др. Тогда алгоритм обработки может содержать цикл, на каждом шаге которого производится несколько итераций для оценки восстанавливаемого объекта, при этом исследуемая на предмет ошибки величина на каждом шаге меняется. В результате отслеживается изменение качества восстановления (например, при помощи расчета нормированной среднеквадратической ошибки – в ручном режиме, по графику, или в автоматическом) от значения исследуемой величины, на котором находится экстремум, отвечающий за максимальное качество, и определяется значение исследовавшейся величины, при котором производятся все дальнейшие расчеты.

На рис. 2 представлено изменение величины нормированного среднеквадратического отклонения E , возникающее при определении точного расстояния до объекта. E характеризует величину ошибки, возникающей при восстановлении информации о фазе, и вычисляется как нормированная разность энергий восстановленного и зарегистрированного изображений:

$$E = \sqrt{\frac{\sum |V - \sqrt{I}|^2}{\sum |\sqrt{I}|^2}},$$

где V – амплитуда рассчитанного волнового фронта; I – зарегистрированное распределение интенсивности. Как видно из рис. 2, для физического эксперимента величина ошибки примерно в 10 раз превышает соответствующее ей значение для моделирования. Это связано с тем, что помимо исследованного в модели параметра, в реальные условия эксперимента вносят свой вклад и другие величины, влияющие на величину ошибки.

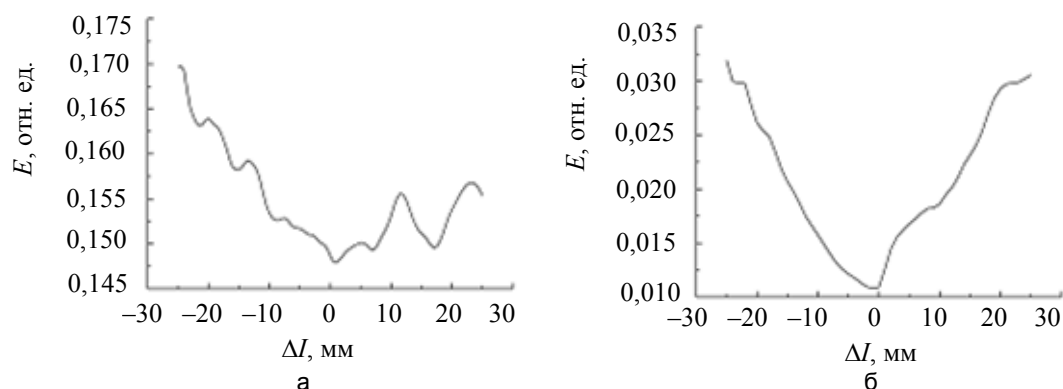


Рис. 2. Зависимость нормированного среднеквадратического отклонения E от использованного в расчете расстояния до объекта: экспериментальные данные (а); численное моделирование (б)

Немаловажным является также наличие существенных различий между зарегистрированными распределениями интенсивностей. Авторами было проведено исследование [27], в ходе которого был определен критерий оптимальной записи данных для рассматриваемой в данном примере конфигурации.

Для более наглядной иллюстрации влияния указанных факторов на качество восстановления можно привести два изображения (рис. 3). На первом из них представлен результат, полученный без использования корректирующих процедур фокусировки, а на втором – после корректировки измеренного расстояния до объекта и погрешности в размерах пикселя регистрирующей матрицы. На рис. 2, б, прослеживается положительный эффект использования корректирующих процедур: надпись отображена заметно лучше, более четко проявилось изображение апертуры.

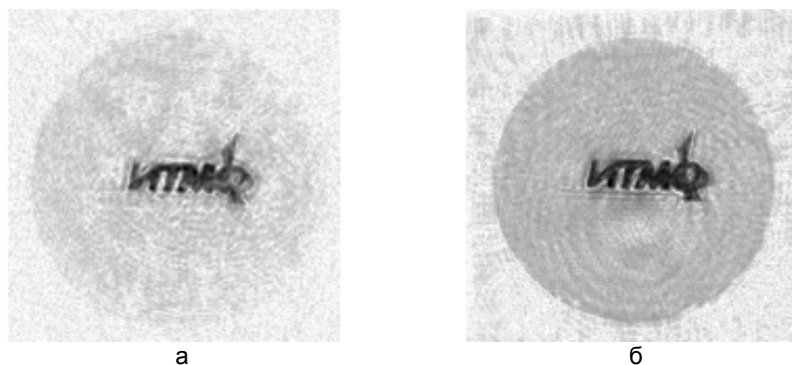


Рис. 3. Восстановленное изображение объекта без применения корректирующих процедур (а) и с их применением (б)

Использование нелинейных оптических эффектов

С точки зрения внесения существенных изменений в регистрируемые распределения интенсивности большие перспективы может представить использование оптических нелинейных сред. Например, в работе [28] алгоритм Гершберга–Сакстона был расширен для возможности использования нелинейных фоторефрактивных сред в качестве фазового анализатора. На основе результатов восстановления распределения фазы было выявлено увеличение контраста у полученного на выходе изображения, а также уменьшение ошибок восстановления фазы.

Более конкретно можно рассмотреть случай использования линейного фазового анализатора в виде линзы в сравнении с нелинейной фоторефрактивной средой [25]. Наиболее значительное отличие нелинейного фильтра заключается в его нечувствительности к распространенным искажениям волнового фронта (в стандартном фазово-контрастном методе искажения волнового фронта ведут к нарушению функции анализатора). Кроме того, в градиентных линейных методах (в том числе при использовании линзы в качестве фазового анализатора) начальное приближение выбирается более или менее произвольно, так как не существует универсальных правил при выборе начальной оценки фазы. В результате сходимость алгоритма может резко падать, что, в свою очередь, приводит к снижению точности восстановления фазы. Но в случае использования нелинейного фильтра существует эффективный способ выбора начальной оценки, причем проблема извлечения фазы формулируется так же, как и для случая с анализатором в виде линзы: фаза может быть получена из распределений интенсивности во входной и выходной плоскостях фильтра, которые легко поддаются измерению. Можно также использовать метод минимизации функционала ошибки для расчета последовательности приближений фазы.

Один из главных недостатков линзы в качестве фазового анализатора проявляется в том, что измерения распределений интенсивности проводятся, как правило, в фокальной плоскости линзы или вблизи нее, пространственная область распределения интенсивности, таким образом, обычно менее 1 мм в диаметре. По этой причине задача получения достаточного числа пикселей (разрешения) становится весьма трудной при измерениях интенсивности на столь небольшой площади. Увеличение полученного изображения (распределения интенсивности) телескопическими системами также не спасает положения из-за неизбежно возрастающих аберраций вследствие несовершенства оптических элементов.

Стоит отметить, что как для успешной реконструкции фазы, так и для расчета прямого или обратного распространения [16] необходимо знать нелинейную функцию распространения, которая в скалярном паракиальном приближении может быть описана нелинейным уравнением Шредингера:

$$\frac{\delta\psi}{\delta z} = \left[i \frac{1}{2k} \nabla_{\perp}^2 + i\Delta n(\psi) \right] \psi \equiv [D + N(\psi)]\psi,$$

где $k = 2\pi/\lambda$; $\Delta n(\psi)$ – нелинейная «добавка» к показателю преломления; D и N – линейный и нелинейный операторы соответственно. Зная начальное или конечное пространственное распределение фазы волнового фронта, можно рассчитать прямое или обратное распространение $\psi(x, y, z)$, используя метод расщепления по физическим процессам [29]. В нем линейные и нелинейные операторы действуют независимо с каждым новым единичным приращением для дистанции распространения dz :

$$\psi(x, y, z \pm dz) \approx e^{\pm dz \cdot D} e^{\pm dz \cdot N(\psi)} \psi(x, y, z).$$

Исследуемый объект 2 (рис. 1) или материал, из которого он состоит, также могут обладать какими-либо свойствами оптической нелинейности. При этом линейные методы исследования не всегда могут быть использованы [16, 30], так как они дают корректный результат, только если пиковая интенсивность облучающего нелинейную среду излучения много меньше порога самофокусировки, и нелинейностью среды можно пренебречь [31]. Исходя из этого, для работы с нелинейными средами разработка новых изначально нелинейных методов (учитывающих природу нелинейности таких сред) либо значительная доработка линейных [16, 30] могут быть востребованными и иметь определенные преимущества.

Для задач обработки и формирования изображений, как и для восстановления фазовой информации от интенсивной световой волны, прошедшей через нелинейную среду, требуется как можно более точная математическая модель, описывающая ее изменения в пространстве и времени [16] в процессе распространения внутри среды. Больше информации о динамике распространяющегося светового пучка в нелинейной дефокусирующей среде в любой ее произвольной точке может дать метод расчета обратного или прямого распространения с использованием цифровой голографии [16]. Причем для учета нелинейности может быть использован расчет приведенного выше нелинейного уравнения Шредингера.

Использование нелинейно-оптических явлений для источника излучения

В итерационных методах восстановления фазовой информации на основе измеренных распределений интенсивности одним из изменяемых параметров может служить длина волны излучения, освещающего исследуемый объект. Посредством последовательного изменения рабочей длины волны в системе излучения и последующего измерения соответствующих распределений интенсивности возможен

расчет информации о фазе. При этом генерация дополнительных спектральных компонент может быть реализована посредством использования каких-либо нелинейных преобразований первичного излучения – например, на основе эффекта вынужденного комбинационного рассеяния [32] или параметрической генерации [33].

Возможность использования нескольких спектральных компонент суперконтинуума в системе излучения для восстановления волновых фронтов была продемонстрирована нами ранее в [18]. Явление вынужденного комбинационного рассеяния (ВКР) так же может быть использовано для генерации дополнительных спектральных компонент, что было показано, например, в [34]. В предшествующей нашей статье [17] также обсуждались возможности использования явления ВКР и генерации дополнительных спектральных компонент суперконтинуума (при использовании фотонно-кристаллического оптического волокна), что впоследствии было успешно реализовано [19]. При этом численная модель восстановления фазы основывалась на итерационном алгоритме, в котором для расчета использовался весь спектр длин волн суперконтинуума.

Таким образом, в задачах восстановления волнового фронта нелинейные эффекты могут находить применение как в системе фазового анализатора, так и в системе излучения. Тем не менее, преимущества нелинейных итерационных методов восстановления фазы основываются на использовании в них нелинейного фазового анализатора, в то время как нелинейность системы излучения не приносит соизмеримой с первым случаем выгоды.

Заключение

Для итерационных методов в рамках общей концепции восстановления волнового фронта при отсутствии фазовой информации по дополнительным наборам данных, полученных при изменении одного или нескольких параметров, было проведено обобщение непосредственных схем их реализации на практике, представлены основные составляющие элементы и общие схемные решения.

Продemonстрировано, что отклонения от математической модели, используемой в итерационном расчете, значительно ухудшают качество восстановления волнового фронта. В связи с этим для минимизации погрешностей предложены дополнительные алгоритмы численной обработки.

Была рассмотрена роль варьируемых параметров при восстановлении волнового фронта. Главной их задачей является внесение существенных различий в регистрируемые распределения интенсивности. В этой связи перспективными при решении фазовой проблемы в оптике являются нелинейные оптические среды, поскольку они могут обеспечивать значительные изменения данных. Наиболее интересные возможности открываются при использовании нелинейных оптических элементов в роли фазового анализатора. Например, для сред, обладающих керровской нелинейностью, это позволяет увеличить контраст выходного расчетного изображения [24], повысить устойчивость метода восстановления к искажениям волнового фронта [17].

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы, соглашение № 14.B37.21.1561.

Литература

1. Кузнецова Т.И. О фазовой проблеме в оптике // УФН. – 1988. – 154. – С. 677–690.
2. Teague M.R. Deterministic phase retrieval: a Green's function solution // J. Opt. Soc. Am. – 1983. – V. 73. – P. 1434–1441.
3. Kolenovic E. Correlation between intensity and phase in monochromatic light // J. Opt. Soc. Am. A. – 2005. – V. 22. – P. 899–906.
4. Gerchberg R.W., Saxton W.O. A practical algorithm for the determination of phase from image and diffraction plane pictures // Optik. – 1972. – V. 35. – P. 237–246.
5. Fienup J. Phase retrieval algorithms: a comparison // Appl. Opt. – 1982. – V. 21. – P. 2758–2769.
6. Honigstein D., Weinroth J., Werman M., Lewis A. Applying heavy atom concepts with scanning probe technology to achieve a non-iterative exact solution to the phase problem in optical imaging // ACS Nano. – 2012. – V. 6. – P. 220–226.
7. Sheludko D.V., McCulloch A.J., Jasperse M., Quiney H.M., Scholten R.E. Non-iterative imaging of inhomogeneous cold atom clouds using phase retrieval from a single diffraction measurement // Opt. Express. – 2010. – V. 18. – P. 1586–1599.
8. Miao J., Charalambous P., Kirz J., Sayre D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens // Nature. – 1999. – V. 400. – P. 342–344.
9. Zuo J., Vartanyants I., Gao M. et al. Atomic resolution imaging of a carbon nanotube from diffraction intensities // Science. – 2003. – V. 300. – P. 1419.
10. Anand A., Pedrini G., Osten W., Almoró P. Wavefront sensing with random amplitude mask and phase retrieval // Opt. Lett. – 2007. – V. 32. – P. 1584–1586.

11. Almoró P.F., Gundu P., Hanson S. Numerical correction of aberrations via phase retrieval with speckle illumination // *Opt. Lett.* – 2009. – V. 34. – P. 521–523.
12. Anand A., Chhaniwal V.K., Almoró P. et al. Shape and deformation measurements of 3D objects using volume speckle field and phase retrieval // *Opt. Lett.* – 2009. – V. 34. – P. 1522–1524.
13. Almoró P.F., Pedrini G., Gundu P.N. et al. Phase microscopy of technical and biological samples through random phase modulation with a diffuser // *Opt. Lett.* – 2010. – V. 35. – P. 1028–1030.
14. Camacho L., Mico V., Zalevsky Z., Garcia J. Quantitative phase microscopy using defocusing by means of a spatial light modulator // *Opt. Express.* – 2010. – V. 18. – P. 6755–6766.
15. Jia S., Wan W., Fleischer J.W. Forward four-wave mixing with defocusing nonlinearity // *Opt. Lett.* – 2007. – V. 32. – P. 1668–1670.
16. Barsi C., Wan W., Fleischer J.W. Imaging through nonlinear media using digital holography // *Nature Photonics.* – 2009. – V. 3. – P. 211–215.
17. Petrov N.V., Bepalov V.G., Gorodetsky A.A. Phase retrieval method for multiple wavelength speckle patterns // *Speckle 2010: Optical Metrology. Proc. of SPIE.* – 2010. – V. 7387. – P. 7387 1T.
18. Petrov N.V., Volkov M.V., Gorodetsky A.A., Bepalov V.G. Image reconstruction using measurements in volume speckle fields formed by different wavelengths // *Biomedical Applications of Light Scattering V. Proc. of SPIE.* – 2011. – V. 7907. – P. 790718-1.
19. Петров Н.В., Беспалов В.Г., Волков М.В. Восстановление волнового фронта при безопорной цифровой ПЗС-регистрации мультиспектральных спекл-картин // *Наносистемы: физика, химия, математика.* – 2011. – Т. 2. – № 1. – С. 82–90.
20. Petrov N.V., Bepalov V.G., Volkov M.V. Phase retrieval of THz radiation using set of 2D spatial intensity measurements with different wavelengths // *Proceedings of SPIE.* – 2012. – V. 8281. – P. 82810J.
21. Pedrini G., Osten W., Zhang Y. Wave-front reconstruction from a sequence of interferograms recorded at different planes // *Opt. Lett.* – 2005. – V. 30. – P. 833–835.
22. Almoró P.F., Pedrini G., Osten W. Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field // *Appl. Opt.* – 2006. – V. 45. – P. 8596–8605.
23. Schiske P. Image reconstruction by means of focus series // *Journal of Microscopy.* – 2002. – V. 207. – P. 154–154.
24. Rodrigo J.A., Duadi H., Alieva T., Zalevsky Z. Multi-stage phase retrieval algorithm based upon the gyrator transform // *Opt. Express.* – 2010. – V. 18. – P. 1510–1520.
25. Ivanov V.Yu., Sivokon V.P., Vorontsov M.A. Phase retrieval from a set of intensity measurements: theory and experiment // *J. Opt. Soc. Am. A.* – 1992. – V. 9. – P. 1515–1524.
26. Larkin K.G., Bone D.J., Oldfield M.A. Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform // *The Journal of the Optical Society of America A.* – 2001. – V. 18. – № 8. – P. 1862–1870.
27. Petrov N.V., Galiaskarov A.N., Nikolaeva T.Yu. Bepalov V.G. The features of optimization of a phase retrieval technique in THz frequency range // *Speckle 2012: Proc. of SPIE.* – 2012. – V. 8413. – P. 84131T.
28. Lu C.-H., Barsi C., Fleischer J. Phase Retrieval through nonlinear Media // *FIO / LS Technical Digest. OSA.* – 2011. – P. 1–4.
29. Агравал Г.П. Нелинейная волоконная оптика. – М.: Мир, 1996. – 328 с.
30. Goy A., Psaltis D. Digital reverse propagation in focusing Kerr media // *Phys. Rev.* – 2011. – A 83. – 031802(R). – P. 031802-1.
31. Abdollahpour D., Papazoglou D.G., Tzortzakis S. Four-dimensional visualization of single and multiple laser filaments using in-line holographic microscopy // *Phys. Rev.* – 2011. – V. A 84. – P. 053809.
32. Петров Н.В., Беспалов В.Г., Жевлаков А.П., Солдатов Ю.И. Исследования двухдлинноволновой цифровой спекл-фотографии для анализа фазовых неоднородностей в гидросфере // *Сб. трудов VI международной конференции молодых ученых и специалистов «Оптика-2009».* – 2009. – С. 279–282
33. Sutherland R.L. *Handbook of Nonlinear Optics* (Second edition). – N.Y., Basel: Marcel Dekker, 2003. – 946 p.
34. Sokolovskaya A.I., Brekhovskikh G.L., Kudryavtseva A.D. Wavefront reconstruction (conjugation) by Stimulated Scatterings: A review // *IEEE J. Quantum Electron.* – 1988. – V. 23. – P. 1332–1343.

- Налегает Сергей Сергеевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, fioi@inbox.ru, s.nalegaev@gmail.com
- Петров Николай Владимирович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, мл. научный сотрудник, Nickolai.petrov@gmail.com
- Беспалов Виктор Георгиевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, victorbespaloff@gmail.com

УДК 535.33; 535.372; 535.354

ЛЮМИНЕСЦЕНЦИЯ МАРГАНЦА ВО ФТОРФОСФАТНЫХ СТЕКЛАХ

В.А. Асеев, Е.В. Колобкова, Я.А. Некрасова, Н.В. Никоноров, А.С. Рохмин

Разработаны и синтезированы фторфосфатные стекла, содержащие MnF_2 в концентрациях 2–20 мол.%. Исследована зависимость спектрально-люминесцентных свойств стекол от концентрации ионов-активаторов. Обсуждаются перспективы применения такой среды в качестве красного люминофора в диодах белого света.

Ключевые слова: белые светодиоды, красный люминофор, стекла, активированные марганцем, фторфосфатные стекла.

Введение

На сегодняшний день с внедрением светодиодов связаны перспективы развития целого ряда направлений – от сигнальных световых приборов и оборудования для световой индикации и рекламы до специальных приложений в производстве мобильных устройств и дисплеев. Основной привлекательной чертой светодиодной технологии является принципиально более высокий по сравнению с альтернативными технологическими решениями уровень светоотдачи, что способно привести к значительным экономическим и социальным эффектам. Важнейшим из них является радикальное сокращение затрат электроэнергии на освещение, составляющих, по различным оценкам, до 18–20% всех затрат произведенной электроэнергии [1].

Наиболее распространенной технологией получения белого света является нанесение желтого люминофора – нанокристаллов алюмоиттриевого граната, активированных ионами церия, на полупроводниковую основу – синий диод на базе InGaN . Однако такие светодиоды обладают низким коэффициентом цветопередачи и так называемым «холодным» белым светом, так как их излучение не охватывает всего видимого диапазона. Для получения нейтрального или теплого белого света (более близкого к солнечному) в современных диодах такого типа необходимо расширять спектр свечения в красную область. Другими словами, к уже существующим диодам «холодного» белого света (цветовая температура ~7000 K) необходимо добавить компонент, вносящий в спектр красную и оранжевую составляющие (580–650 нм). Таким образом, изменяя интенсивность красной компоненты (варьируя толщину слоя красного люминофора или концентрацию активатора) возможно создавать белые светодиоды со свечением с различной цветовой температурой (3000–7000 K). Наибольший интерес представляют диоды белого света с нейтральным (6000–6500 K) и теплым белым свечением (5000 K).

В основном для создания красных люминофоров используются сульфидная, силикатная, нитридная или фторидная матрицы стекла. Выбор фторфосфатной матрицы обусловлен ее высокой химической и термической стабильностью, а также возможностью введения высоких концентраций ионов-активаторов [2, 3]. Сделанные оценки указывают на перспективность использования ионов как редкоземельных, так и переходных металлов, а также их комбинации во фторидных стеклах и наностеклокерамиках на их основе. Следует отметить, что материалы, активированные ионами переходных металлов, обладают широкими полосами поглощения и люминесценции и относительно дешевы. Например, ионы Mn^{2+} обладают интенсивной полосой люминесценции с максимумом ~620 нм [4]. Также возможно введение высоких концентраций марганца с сохранением его валентного состояния. Однако положение полос возбуждения и люминесценции такого активатора зависит от степени его окисления и координации [5].

Целью настоящей работы является исследование концентрационной зависимости спектрально-люминесцентных свойств фторфосфатных стекол, активированных ионами Mn^{2+} .

Эксперимент

В работе исследовались оксифторидные стекла системы 5 мол.% $\text{Ba}(\text{PO}_3)_2$ –95 мол.% $\text{MgPbCa}(\text{Ba})\text{SrAl}_2\text{F}_{14}$, в качестве активатора вводили MnF_2 . Концентрация MnF_2 составила 2–20 мол.% при одновременном введении EuF_3 в количестве 0,2 мол.%. Европий вводился в качестве сенситизатора для марганца при возбуждении на 465 нм. Для синтеза стекол применялись материалы марки с чистотой 99,99%, выпускаемые отечественной химической промышленностью. Варка стекол производилась в электрической лабораторной печи с силитовыми нагревателями, обеспечивающими нагрев рабочей зоны до 1450°C. Варка стекол осуществлялась в открытых корундовых тиглях в атмосфере воздуха при температуре 900–1050°C в течение 30 мин.

Спектры поглощения измерялись на спектрофотометре Lambda 900 (Perkin Elmer) в диапазоне 300–500 нм с шагом 0,1 нм, время интеграции 0,2 с. Для возбуждения люминесценции использовалось излучение импульсного лазера LS-2131M фирмы Lotis ТП с приставкой-преобразователем излучения HG-T, длина волны лазерного излучения составляла 355 нм, энергия возбуждения – 10 мДж. Регистрация спектров проводилась в видимом диапазоне длин волн (400–800 нм) с использованием монохроматора (Model Acton-300, Acton Research Corp.), фотоэлектронного умножителя (Model Hamamatsu R928) и цифрового синхронного усилителя (Model SR850, Stanford Research Systems). Все измерения проводились при комнатной температуре и были нормированы на кривую спектральной чувствительности приемника.

Результаты и обсуждения

В ходе работы были исследованы спектрально-люминесцентные свойства стекол, активированных ионами двухвалентного марганца в концентрациях MnF_2 : 2, 5, 10, 15 и 20 мол.%. На рис. 1 представлен спектр поглощения образца, активированного ионами Mn^{2+} и Eu^{3+} , отмечены основные энергетические переходы.

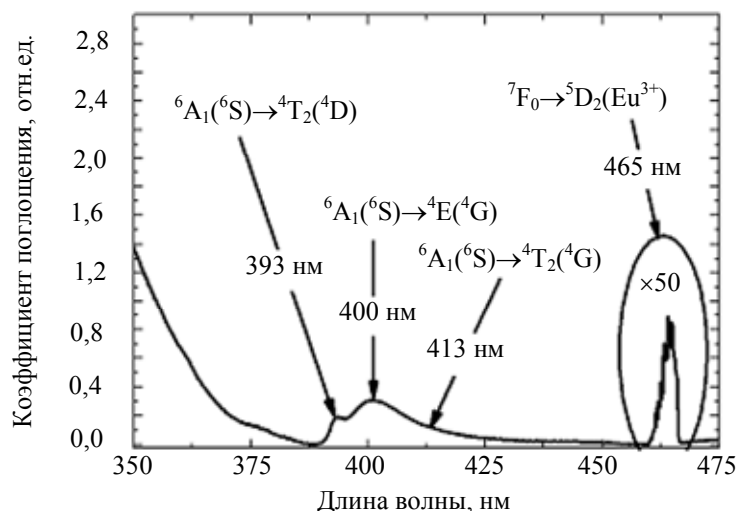


Рис. 1. Спектр поглощения ионов Mn^{2+} и Eu^{3+} в оксифторидном стекле

По литературным данным [6], полосы поглощения на 393, 400 и 413 нм соответствуют следующим энергетическим переходам ионов марганца ${}^6\text{A}_1({}^6\text{S}) \rightarrow {}^4\text{T}_2({}^4\text{D})$, ${}^6\text{A}_1({}^6\text{S}) \rightarrow {}^4\text{E}({}^4\text{G})$ и ${}^6\text{A}_1({}^6\text{S}) \rightarrow {}^4\text{T}_2({}^4\text{G})$. Поглощение Mn^{2+} в области 450–465 нм отсутствует, поэтому в систему вводят ионы Eu^{3+} с полосой на 465 нм (${}^7\text{F}_0 \rightarrow {}^5\text{D}_2$). Накачка такой системы на длине волны в области 450–465 нм становится возможной за счет передачи энергии с уровня ${}^5\text{D}_2$ европия на уровень ${}^4\text{T}_2({}^4\text{G})$ марганца.

Также представлены спектры поглощения для исследуемого концентрационного ряда (рис. 2).

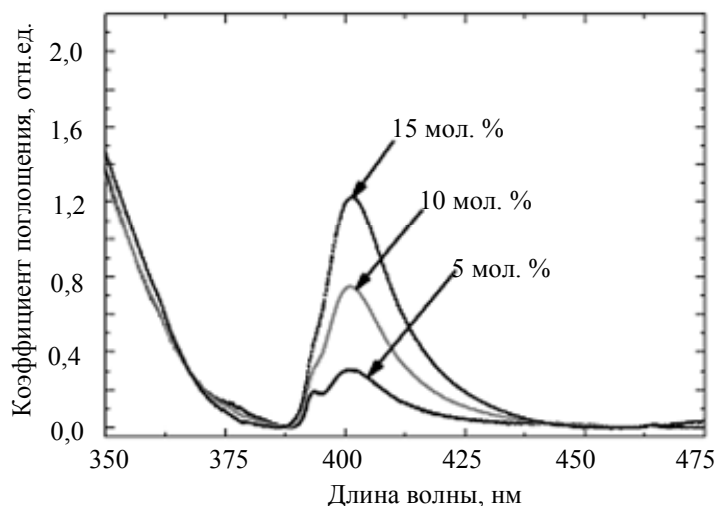


Рис. 2. Спектры поглощения оксифторидного стекла с разной концентрацией MnF_2

При изменении концентрации с 5 до 20 мол.% наблюдается рост значения коэффициента поглощения марганца на длине волны 400 нм, что позволяет более эффективно поглощать возбуждения диода накачки.

На спектре люминесценции (рис. 3) наблюдаются полосы, связанные с излучением как ионов двухвалентного марганца (${}^4\text{T}_1({}^4\text{G}) \rightarrow {}^6\text{A}_1({}^6\text{S})$), так и трехвалентного европия (${}^7\text{F}_1 \rightarrow {}^5\text{D}_0$ (595 нм), ${}^7\text{F}_2 \rightarrow {}^5\text{D}_0$ (618 нм) и ${}^7\text{F}_4 \rightarrow {}^5\text{D}_0$ (700 нм)). При увеличении концентрации марганца до 20 мол.% выделить полосы европия становится практически невозможно, поскольку они перекрыты широкой полосой марганца. Увеличение концентрации ионов-активаторов приводит к смещению максимума широкого пика люминесценции в более длинноволновую область. Это может быть связано с изменением действия поля лигандов на ионы Mn^{2+} при изменении ближайшего окружения этих ионов.

Распад люминесценции происходит по экспоненциальному закону. В результате концентрационного тушения при увеличении содержания ионов Mn^{2+} в составе оксифторидных стекол с 5 до 20 мол.% наблюдается уменьшение времени жизни люминесценции с 17,7 до 7,3 мс.

На рис. 4 представлен спектр диода при добавлении люминофора, активированного марганцем, при возбуждении белым диодом с температурой 9500 К. Введение красной компоненты позволило сместить суммарный спектр излучения в сторону больших значений длин волн. Так, добавление люминофора на основе MnF_2 , позволило сместить максимум полосы с 550 нм до 610 нм. По спектрам излучения при возбуждении светоизлучающего диода (СИД) с цветовой температурой 9500 К были оценены значения цветовой температуры и индекса цветопередачи для образца с концентрацией 2 мол.% (таблица).

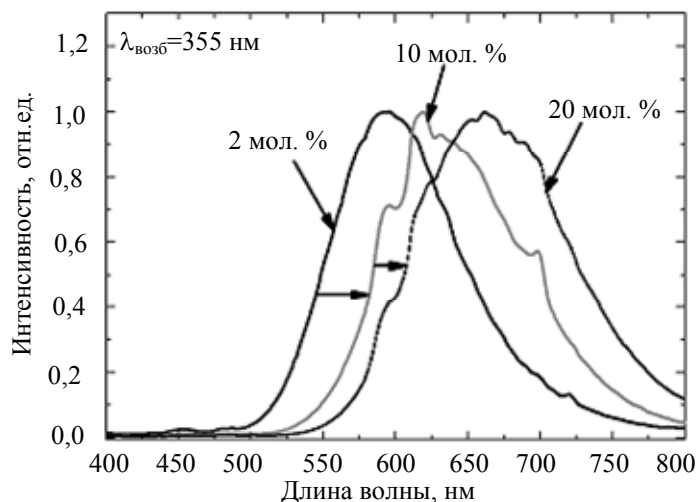


Рис. 3. Спектры люминесценции оксифторидного стекла с разной концентрацией MnF_2

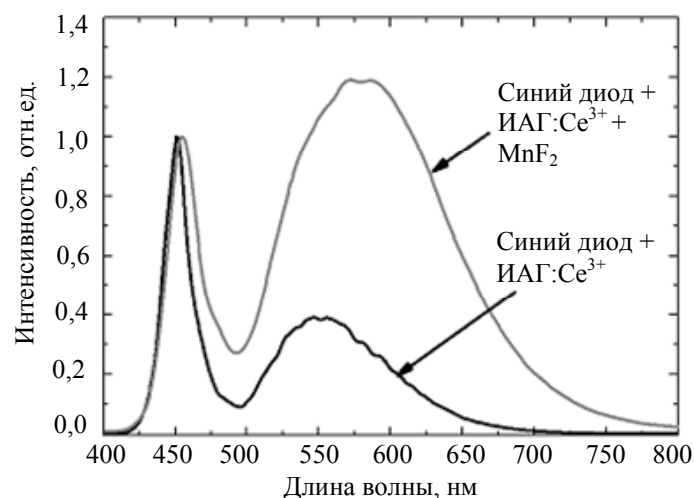


Рис. 4. Спектры люминесценции образцов с MnF_2

СИД	CIE x, y	CCT, К	CRI	λ_{max} , нм
Синий диод	0,1633; 0,0436	—	—	452
ИАГ: Ce^{3+}	0,2792; 0,2984	9540	65	553
MnF_2	0,3918; 0,3466	3450	81	610

Таблица. Оптические характеристики полученных светодиодов (CIE – координаты цветового пространства; CCT – цветовая температура; CRI – индекс цветопередачи)

По полученным данным можно сделать вывод, что введение красного люминофора на основе оксифторидного стекла, активированного марганцем, позволило уменьшить значение цветовой температуры с 9500 К до 3500 К, а также увеличить индекс цветопередачи с 65 до 81.

Заключение

Синтезирован новый тип люминофора, основанный на прозрачном оксифторидном стекле, легированном ионами марганца и европия. По результатам проведенных исследований концентрационной зависимости спектрально-люминесцентных свойств полученных образцов сделаны следующие выводы:

- введение красного люминофора позволило сместить максимум полосы в красно-оранжевой области спектра в сторону больших длин волн, тем самым уменьшив значение цветовой температуры излучения с 9500 К до 3500 К. Индекс цветопередачи составил примерно 81;
- на основе нового красного люминофора возможно создание долгоживущих (до 100 000 ч свечения) энергоэффективных источников белого света с улучшенным индексом цветопередачи и спектром, максимально приближенным к солнечному. Такие источники могут найти применение в системах освещения помещений, улиц, подсветке зданий, а также в качестве подсветки жидкокристаллических дисплеев и мониторов.

Авторы работы выражают благодарность Олегу Алексеевичу Усову (Физико-технический институт им. А.Ф. Иоффе РАН) за помощь в проведении эксперимента.

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (Соглашение № 14.В37.21.0169, Минобрнауки РФ).

Литература

1. Дорожная карта «Использование нанотехнологий в производстве светодиодов». – М.: ГК Роснотех, 2010. – 205 с.
2. Справочник. Светодиодное освещение: Принципы работы, преимущества и области применения / Под ред. Дж. Вейнерт, Ч. Сполдинг. – Philips Solid-State Lighting Solutions, Inc., 2010. – 156 p.
3. Aseev V., Kolobkova E., Nikonov N. New nanoglassceramics doped with rare earth ions and their photonic applications // Chapter in book «Advanced Photonic Sciences» / Ed. by M. Fadhal. – InTech. – 2012. – P. 105–131.
4. Xinguo Zhang, Menglian Gong. A new red-emitting Ce^{3+} , Mn^{2+} -doped barium lithium silicate phosphor for NUV LED application // Materials Letters. – 2011. – V. 65. – P. 1756–1758.
5. Reisfeld R., Kisilev A. Luminescence of manganese (II) in 24-phosphate glasses // Chemical Physics Letters. – 1984. – V. 111. – № 1, 2. – P. 19–24.
6. Шамшури А.В., Маскалюк Л.Г., Репин А.В. Люминофоры на основе твердых растворов фосфатов цинка и магния, активированные ионами марганца // Труды Одесского политехнического института. – 1999. – В. 3. – С. 230–232.

Асеев Владимир Анатольевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, Aseev@oi.ifmo.ru

Колобкова Елена Вячеславовна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор химических наук, профессор, Kolobok106@rambler.ru

Некрасова Яна Андреевна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Nekrasova@oi.ifmo.ru

Никоноров Николай Валентинович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, зав. кафедрой, Nikonov@oi.ifmo.ru

Рохмин Алексей Сергеевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, Rokhmin@oi.ifmo.ru

УДК 681.5.11

АНАЛИЗ РОБАСТНОСТИ НЕАДАПТИВНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ЭЛЕКТРОПРИВОДА С ВАРИАЦИЯМИ СТРУКТУРЫ И ПАРАМЕТРОВ

А.А. Абдуллин, В.Н. Дроздов

Рассматриваются численные методы анализа робастности свойства устойчивости линейной неадаптивной системы управления электроприводов при заданных количественных оценках вариаций модели объекта. Описывается алгоритм процедуры оценки, иллюстрированный примером. Предложенный метод анализа робастности гарантирует работоспособность реальной системы.

Ключевые слова: электропривод, неадаптивная система управления, робастность, свойство устойчивости.

Введение

В практике электропривода адаптивное управление не получило широкого распространения. Как правило, используются системы управления с фиксированными параметрами и структурой, настройка которых если и производится, то эпизодически усилиями операторов [1–3]. В качестве исходных данных для проектирования регулятора с фиксированными параметрами и структурой используются математическая модель объекта управления (ОУ), модель внешних воздействий и требования определенного качества системы. Математические модели физических объектов, естественно, отличаются от идеальных моделей, применяемых в качестве исходных данных для синтеза регулятора. Возникает вопрос о сохранении некоторых желаемых свойств, прежде всего, свойства устойчивости, при вариациях модели ОУ в системе с рассчитанным регулятором [4]. В развитие положений работы [4] в [5] определяется робастность системы как способность сохранять некоторое свойство системы при известных количественных оценках вариаций математической модели. В практике проектирования регуляторов систем управления электроприводом обычно пренебрегают известными малыми постоянными времени. Пределы изменения параметров физических объектов, как правило, также известны. В связи с этим вполне правомерно ставить вопрос об исследовании робастности систем управления электроприводом. В настоящее время нет единой теории исследования робастности систем [6], по этой причине, на взгляд авторов, можно рассматривать любые подходы к проблеме. В настоящей работе предлагаются легко реализуемые численные методы исследования робастности систем управления электроприводом с фиксированным регулятором.

Постановка задачи

Линейная модель состояния электропривода с нагрузкой, в том числе упругой, имеет стандартный вид:

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{q})\mathbf{x} + \mathbf{B}(\mathbf{q})\mathbf{u}, \quad (1)$$

здесь $\mathbf{x} \in R^n$ – вектор состояния, включающий в общем случае переменные состояния модели внешнего воздействия при реализации изодромного (грубого) управления; $\mathbf{u}$ – скалярное управляющее воздействие; $\mathbf{q} \in R^r$, $\mathbf{q} = \mathbf{q}_0 \pm \Delta\mathbf{q}$ – вектор параметров, претерпевающий вариацию $\Delta\mathbf{q}$; $\mathbf{A}(\mathbf{q})$ – матрица состояния $n \times n$; $\mathbf{B}(\mathbf{q})$ – матрица управления $n \times 1$.

Выделим медленные, $\mathbf{x}_M \in R^{n_1}$, и быстрые, $\mathbf{x}_B \in R^{n_2}$, переменные вектора состояния $\mathbf{x}$, $n_1 + n_2 = n$, и преобразуем (1) к виду

$$\begin{aligned} \dot{\mathbf{x}}_M &= \mathbf{A}_M(\mathbf{q})\mathbf{x}_M + \mathbf{A}_{12}(\mathbf{q})\mathbf{x}_B + \mathbf{B}_M(\mathbf{q})\mathbf{u}, \\ \dot{\mathbf{x}}_B &= \mathbf{A}_{21}(\mathbf{q})\mathbf{x}_M + \mathbf{A}_B(\mathbf{q})\mathbf{x}_B + \mathbf{B}_B(\mathbf{q})\mathbf{u}. \end{aligned} \quad (2)$$

Медленная и быстрая составляющие вектора $\mathbf{x}$ представляются в виде

$$\mathbf{x}_M = \mathbf{C}_M\mathbf{x}, \quad \mathbf{x}_B = \mathbf{C}_B\mathbf{x}, \quad (3)$$

где матрица $\mathbf{C}_M$ имеет размерность $n_1 \times n$, а матрица $\mathbf{C}_B$ – размерность $n_2 \times n$. В общем случае управляющее воздействие может воздействовать как на быстрые, так и на медленные составляющие, как это отмечено в (2). Такая ситуация характерна для векторного управляющего воздействия.

Множество собственных чисел $\{\lambda_{Bi}, i = \overline{1, n_2}\}$ матрицы $\mathbf{A}_B(\mathbf{q})$ лежит значительно левее множества собственных чисел $\{\lambda_{Mj}, j = \overline{1, n_1}\}$ матрицы $\mathbf{A}_M(\mathbf{q})$ на комплексной плоскости, так что выполняется неравенство

$$\min(\text{mod}(\text{Re}(\lambda_{Bi}))) \geq 10 \max(\text{mod}(\text{Re}(\lambda_{Mj}))). \quad (4)$$

Коэффициент 10 в последнем неравенстве принят из следующих соображений. В инженерной практике принято считать, что если значения каких-либо величин различаются не менее чем на порядок, то это различие существенное.

В этом случае быстрые движения $\mathbf{x}_B$ после любого возмущения затухают за короткие промежутки времени, так что в течение основных промежутков времени можно считать $\dot{\mathbf{x}}_B = 0$, при этом второе уравнение в (2) преобразуется к виду

$$0 = \mathbf{A}_{21}(\mathbf{q})\mathbf{x}_M + \mathbf{A}_B(\mathbf{q})\mathbf{x}_B + \mathbf{B}_B(\mathbf{q})u. \quad (5)$$

Матрица $\mathbf{A}_B(\mathbf{q})$ является гурвицевой, поэтому она обратима, тогда из (5) находим

$$\mathbf{x}_B = -\mathbf{A}_B^{-1}(\mathbf{q})(\mathbf{A}_{21}(\mathbf{q})\mathbf{x}_M + \mathbf{B}_B(\mathbf{q})u). \quad (6)$$

Подставив (6) в первое уравнение (2), получим редуцированную модель ОУ в виде

$$\dot{\mathbf{x}}_M = \mathbf{A}_R(\mathbf{q})\mathbf{x}_M + \mathbf{B}_R(\mathbf{q})u,$$

где

$$\mathbf{A}_R(\mathbf{q}) = \mathbf{A}_M(\mathbf{q}) - \mathbf{A}_{12}(\mathbf{q})\mathbf{A}_B^{-1}(\mathbf{q})\mathbf{A}_{21}(\mathbf{q}), \mathbf{B}_R(\mathbf{q}) = \mathbf{B}_M(\mathbf{q}) - \mathbf{A}_{12}(\mathbf{q})\mathbf{A}_B^{-1}(\mathbf{q})\mathbf{B}_B(\mathbf{q}).$$

В пользу редуцирования моделей объектов можно привести следующие соображения [5, 6]. Во-первых, в распоряжении проектировщика никогда нет модели, абсолютно точно описывающей поведение объекта. При получении модели всегда, вольно или невольно, приходится принимать какие-то допущения, в том числе, пренебрегать быстрыми, с точки зрения разработчика, процессами в объекте, и надо быть уверенным, что подобные допущения не приведут к значительным неприятностям. Синтез регулятора для редуцированной модели с последующей проверкой свойств системы с полной моделью дает уверенность в том, что спроектированный регулятор обеспечит желаемые свойства системы при управлении реальным объектом. Вторая причина заключается в том, что параметры регулятора, синтезированного для модели, с существенно различающимися темпами собственных движений могут отличаться на несколько порядков. Цифровая реализация таких регуляторов может встретить определенные трудности. Кроме того, с упрощенными моделями работать значительно легче, чем с моделями высокого порядка.

Положив в редуцированной модели номинальное (медианное) значение вектора параметров $\mathbf{q} = \mathbf{q}_0$, получим идеализированную расчетную модель ОУ

$$\dot{\mathbf{x}}_M = \mathbf{A}_R\mathbf{x}_M + \mathbf{B}_Ru, \quad (7)$$

где $\mathbf{A}_R = \mathbf{A}_R(\mathbf{q})|_{\mathbf{q}=\mathbf{q}_0}$; $\mathbf{B}_R = \mathbf{B}_R(\mathbf{q})|_{\mathbf{q}=\mathbf{q}_0}$.

Для линейного стационарного объекта (7) синтезируется регулятор, например, реализующий алгоритм модального управления

$$u = -\mathbf{K}\mathbf{x}_M. \quad (8)$$

Матрица $\mathbf{K}$ выбирается таким образом, чтобы обеспечить желаемые собственные числа матрицы состояния $\mathbf{F}$ замкнутой системы:

$$\dot{\mathbf{x}}_M = \mathbf{F}_M\mathbf{x}_M, \quad (9)$$

где $\mathbf{F}_M = \mathbf{A}_M - \mathbf{B}_M\mathbf{K}$.

Желаемые собственные числа задаются обычно в виде корней характеристического уравнения какого-либо эталонного фильтра. В качестве такого фильтра часто выбирают фильтр Бесселя соответствующего порядка с полосой пропускания ω . Заметим, что полоса пропускания эталонного фильтра определяет быстродействие системы (9).

В действительности закон управления (8) используется для управления объектом (1), отличающимся от объекта (7), для которого проектировался этот закон. Возникает задача оценки отличия свойств объекта (1), замкнутого регулятором (8), от свойств объекта (7), замкнутого регулятором (8). Насколько оправдываются наши надежды на сохранение свойств в первой системе, заложенные при проектировании второй системы?

Исследование робастности замкнутой системы

Замкнем объект (1) регулятором (8), в результате получим

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{q})\mathbf{x} - \mathbf{B}(\mathbf{q})\mathbf{K}\mathbf{x}_M$$

или, учитывая (3),

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{q})\mathbf{x} - \mathbf{B}(\mathbf{q})\mathbf{K}\mathbf{C}_M\mathbf{x}. \quad (10)$$

Обозначим

$$\mathbf{F}(\mathbf{q}) = \mathbf{A}(\mathbf{q}) - \mathbf{B}(\mathbf{q})\mathbf{K}\mathbf{C}_M.$$

Ожидается, что спектр $\sigma\{\mathbf{F}(\mathbf{q}_0)\} = \{\lambda_i(\mathbf{q}_0) : \det[\lambda_i(\mathbf{q}_0)\mathbf{I} - \mathbf{F}(\mathbf{q}_0)] = 0, i = \overline{1, n}\}$ собственных значений матрицы $\mathbf{F}(\mathbf{q}_0)$ является объединением спектров матриц $\mathbf{A}_B(\mathbf{q}_0)$ и $\mathbf{F}_M$

$$\sigma\{\mathbf{F}(\mathbf{q}_0)\} = \sigma\{\mathbf{A}_B(\mathbf{q}_0)\} \cup \sigma\{\mathbf{F}_M\}. \quad (11)$$

В действительности эти ожидания не оправдываются, и возникают два вопроса. Во-первых, оценить степень искажения (11) и выявить условия, при которых сохраняется допущение о разделении движений в системе (10) на быстрые и медленные. Во-вторых, выяснить, сохраняется ли свойство устойчивости системы (10) при изменении параметров $\mathbf{q}$ в заданных пределах, и оценить изменение степени устойчивости в этих условиях.

Предлагается следующая процедура исследования робастности свойства устойчивости в системе (10). Строятся траектории собственных чисел матрицы $\mathbf{F}_M$ системы (9) при изменении полосы пропускания ω эталонного фильтра, с ростом ω собственные числа перемещаются влево на комплексной плоскости, удаляясь от границы устойчивости. При тех же условиях строятся траектории собственных чисел матрицы $\mathbf{F}(\mathbf{q}_0)$ системы (10). Очевидно, подмножества $\{\lambda_{Mj}, j = \overline{1, n_1}\}$ и $\{\lambda_{Bi}, i = \overline{1, n_2}\}$ множества собственных чисел матрицы $\mathbf{F}(\mathbf{q}_0)$ с ростом ω сближаются. Определяется такое максимальное значение $\omega_{\max}$, при котором еще выполняется условие (4), т.е. остается справедливым разделение переменных состояния системы на быстрые и медленные. Наложение траекторий собственных чисел матрицы $\mathbf{F}_M$ на траектории подмножества $\{\lambda_{Mj}, j = \overline{1, n_1}\}$ множества собственных чисел матрицы $\mathbf{F}(\mathbf{q}_0)$ позволяет оценить эволюцию степени устойчивости сингулярно возмущенной системы (10) при изменениях полосы пропускания эталонного фильтра, задающего динамические свойства проектируемой системы.

После установления значения $\omega_{\max}$ выбирается полоса пропускания ω_0 эталонного фильтра, например, из условий обеспечения заданного быстродействия системы, при выполнении неравенства

$$\omega_0 \leq \omega_{\max}.$$

В этом случае обеспечивается робастность относительно свойства устойчивости сингулярно возмущенной системы (10) с номинальными значениями параметров.

Теперь можно обратиться к исследованию робастности относительно свойства устойчивости системы (10) при изменении параметров $\mathbf{q}$ в известных пределах. Здесь также можно воспользоваться построением траекторий собственных чисел матрицы $\mathbf{F}(\mathbf{q})$ при изменении отдельных переменных вектора $\mathbf{q}$. Однако быстрый и надежный результат дает применение теоремы Харитоновой [6, 7]. Согласно этой теореме, в пространстве R^r параметров строится прямоугольный параллелепипед, ограничивающий область изменения параметров. Координаты центра параллелепипеда задаются номинальными значениями параметров, координаты углов параллелепипеда определяются заданными пределами изменения параметров [6, 7]. Координаты углов этого параллелепипеда называются угловой реализацией вектора параметров. Вычисляются собственные числа матрицы $\mathbf{F}(\mathbf{q})$ при всех угловых реализациях вектора параметров. Если все эти собственные числа лежат в левой полуплоскости плоскости корней, то система (10) робастна относительно свойства устойчивости. Нанеся собственные числа матрицы $\mathbf{F}(\mathbf{q})$ на комплексную плоскость при всех угловых реализациях вектора параметров, можем оценить степень устойчивости системы (10) при вариациях ее параметров.

В случае невыполнения условий теоремы Харитоновой необходимо рассчитывать новые параметры закона управления (9). Единственным переменным параметром в исходных данных на расчет матрицы $\mathbf{K}$ в рассматриваемой постановке задачи является полоса пропускания эталонного фильтра.

Закон управления (8) предполагает наличие датчиков для всех медленных переменных вектора состояния объекта. Использование стационарного наблюдателя неизмеряемых переменных в законе управления не вносит принципиальных изменений в процедуру анализа робастности параметрически и сингулярно возмущенного объекта.

Пример

Анализ робастности неадаптивной системы управления электропривода рассмотрим на примере электромеханической системы (ЭМС), которая функционально содержит три основных узла: механизм (Мх), представленный двухмассовой расчетной схемой; электромеханический преобразователь энергии (ЭМП или электрическая машина); управляемый преобразователь (УП) электрической энергии. Структурная схема такой системы представлена на рис. 1.

Вектор состояния ЭМС, представленной на рис. 1, имеет вид

$$\mathbf{x} = [\Omega_0 \quad M \quad \Omega_1 \quad M_{12} \quad \Omega_2 \quad \alpha_2].$$

Учитывая значения моментов инерции механизма, можно предположить, что постоянная времени УП и электрическая постоянная времени ЭМП незначительны по отношению к постоянной времени механизма. Следовательно, векторы быстрых и медленных переменных состояния принимают вид

$$\mathbf{x}_B = [\Omega_0 \quad M],$$

$$\mathbf{x}_M = [\Omega_1 \quad M_{12} \quad \Omega_2 \quad \alpha_2].$$

Полагая равными нулю малые постоянные времени, получим упрощенную ЭМС, структурная схема которой представлена на рис. 2.

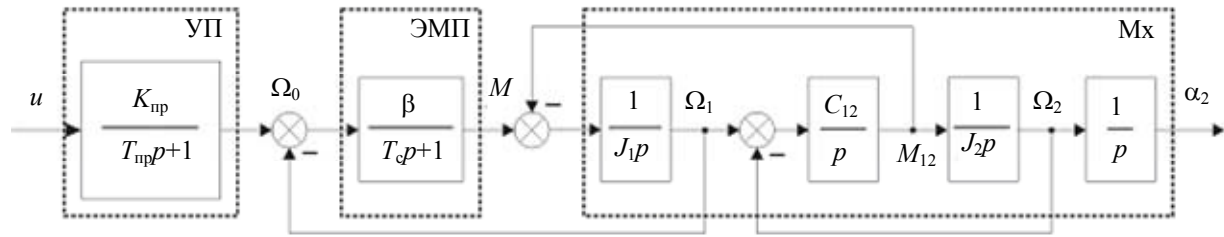


Рис. 1. Структурная схема ЭМС: $J_1=6600 \text{ кг}\cdot\text{м}^2$ – момент инерции первой массы; $J_2=197300 \text{ кг}\cdot\text{м}^2$ – момент инерции второй массы; $C_{12}=8,62\cdot 10^8 \text{ Н}\cdot\text{м/рад}$ – коэффициент упругости; $\beta=2,9\cdot 10^4 \text{ Н}\cdot\text{м}\cdot\text{с/рад}$ – жесткость механической характеристики ЭМП; $T_e=1,6\cdot 10^{-4} \text{ с}$ – электрическая постоянная времени фазной обмотки ЭМП; $K_{пр}=0,026 \text{ рад/(В}\cdot\text{с)}$ – пропорциональный коэффициент УП; $T_{пр}=2\cdot 10^{-4} \text{ с}$ – постоянная времени УП

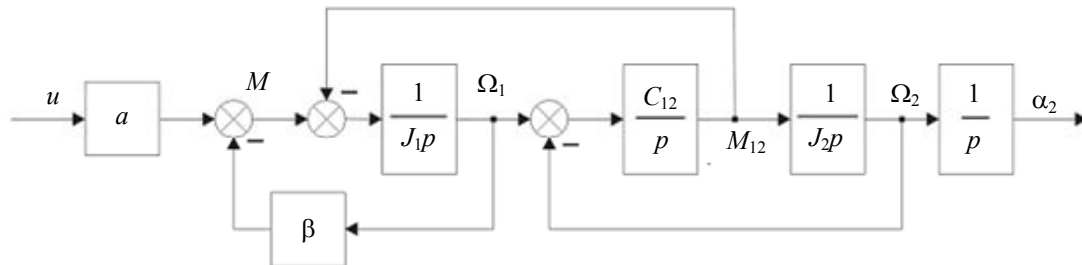


Рис. 2. Структурная схема упрощенной ЭМС

При синтезе регулятора, реализующего алгоритм модального управления (8) упрощенной ЭМС, для расчета коэффициентов матрицы **K** использовался эталонный фильтр Бесселя четвертого порядка.

На рис. 3 представлены корни матрицы состояния упрощенной и исходной замкнутых систем.

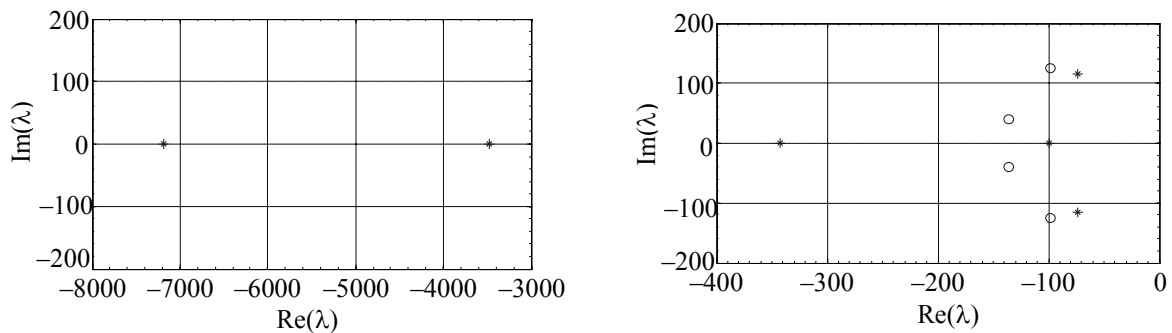


Рис. 3. Корни матрицы состояния упрощенной (о) и исходной (*) замкнутых систем

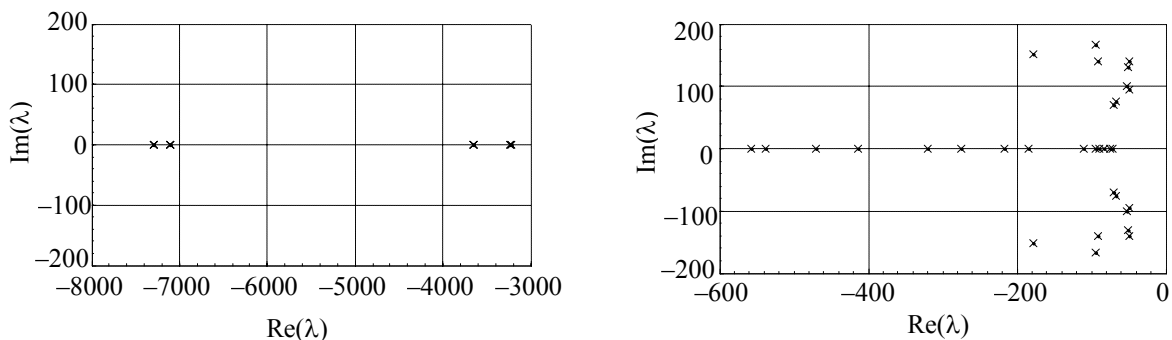


Рис. 4. Корни матрицы состояния замкнутой системы с измененными параметрами

Здесь полоса пропускания эталонного фильтра Бесселя принимает значение 150 рад/с, при этом выполняется условие (4). Дальнейшее увеличение полосы пропускания нарушает это условие. Из рис. 3 можно заключить, что упрощение структуры ОУ за счет принятых допущений не сказывается на работоспособности исходной (реальной) системы. При этом значение степени устойчивости составляет $\eta=74 \text{ с}^{-1}$.

При анализе параметрической чувствительности значения моментов инерции J_1, J_2 и коэффициента упругости C_{12} отклонялись на величину $\pm 15\%$. Следовательно, имеем восемь угловых состояний прямоугольного параллелепипеда, ограничивающего область изменения параметров. На рис. 4 представлены корни матрицы состояния замкнутой системы в восьми угловых состояниях, т.е. 48 корней для системы шестого порядка. По корням, представленным на рис. 4, можно сделать вывод, что система робастна относительно свойства устойчивости, при этом значение степени устойчивости составляет $\eta=48 \text{ с}^{-1}$.

Заключение

Предлагаемая процедура исследования робастности неадаптивных систем управления электроприводом, без затруднений реализуемая с использованием современных средств вычислительной математики, доступна, на взгляд авторов, широкому кругу инженеров.

Вне поля зрения в работе остался вопрос исследования робастности системы при функциональных возмущениях, основной составной частью которых в системах управления электроприводом является момент сухого трения. Этот вопрос требует отдельного рассмотрения.

Литература

1. Балковой А.П., Цаценкин В.К. Прецизионный электропривод с вентильными двигателями. – М.: МЭИ, 2010. – 322 с.
2. Толмачев В.А., Антипова И.В., Фомин С.Г. Математическая модель следящего электропривода оси опорно-поворотного устройства // Научно-технический вестник СПбГУ ИТМО. – 2007. – № 44. – С. 142–146.
3. Терехов В.М., Осипов О.И. Системы управления электроприводов. – М.: Академия, 2005. – 300 с.
4. Красовский Н.Н. Некоторые задачи теории устойчивости движения. – М.: Физматгиз, 1959. – 211 с.
5. Мирошник И.В., Никифоров В.О., Фрадков А.Л. Нелинейное и адаптивное управление сложными динамическими системами. – СПб: Наука, 2000. – 549 с.
6. Никифоров В.О., Ушаков А.В. Управление в условиях неопределенности: чувствительность, адаптация, робастность. – СПб: СПб ГИТМО (ТУ), 2002. – 232 с.
7. Харитонов В.Л. Устойчивость вложенных семейств полиномов // Автоматика и телемеханика. – 1995. № 5. – С. 170–178.

Абдуллин Артур Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, artur.abdullin@gmail.com

Дроздов Валентин Нилович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, професор, drozdovuprint@rambler.ru

УДК 004.93 + 62.51 + 625.03

СПЕКТРАЛЬНЫЙ АНАЛИЗ ДАННЫХ С АКСЕЛЕРОМЕТРОВ ДЛЯ ЗАДАЧ ОБНАРУЖЕНИЯ И ИДЕНТИФИКАЦИИ ТРАНСПОРТНЫХ СРЕДСТВ

Д.Е. Обертов

Рассматриваются проблемы частотного и временного анализа вибраций дорожного покрытия, вызванных проезжающими транспортными средствами. Вибрации измеряются акселерометрами, устанавливаемыми на дорожную разметку. Реализуется предварительная обработка данных перед дальнейшим использованием их для анализа параметров движения транспортного потока. Приведены результаты обработки данных с помощью спектрального анализа и фильтрации.

Ключевые слова: акселерометр, обнаружение транспортных средств, идентификация классов транспортных средств, интеллектуальные транспортные системы, спектры, спектральная обработка, спектральный диапазон, быстрое преобразование Фурье, короткооконное преобразование Фурье.

Введение

В последнее время большое внимание уделяется проектированию и разработке интеллектуальных транспортных систем, в том числе интеллектуальных дорог, которые снабжены средствами, позволяющими обеспечивать взаимодействие и связь между транспортной инфраструктурой и транспортными средствами

(ТС) [1–5]. Идентификация классов ТС является важным компонентом безопасности дорожного движения. В связи с этим создание системы и алгоритма, который с большой производительностью и надежностью мог бы работать в различных погодных условиях, в любое время суток, является актуальной проблемой. Необходимо, по возможности, обеспечить бесконтактность измерений, что позволит продлить время эксплуатации системы, уменьшить количество переустановок оборудования и вмешательство людей в дорожное движение, приводящее к нарушению движения транспорта во время установки.

ТС могут быть классифицированы по различным критериям в зависимости от задач. Наиболее очевидным является применение идентификации классов ТС для мониторинга и статистики дорожного движения [5, 6]. Эта информация может быть использована в дальнейшем для планирования ремонта дорог, организации других эксплуатационных мероприятий и принятия мер по разгрузке транспортной сети. В настоящей работе исследуется возможность использования акселерометров для обнаружения и идентификации классов ТС. Для предварительной обработки сигнала предлагается применить спектральный и временной анализ данных с акселерометров, встроенных в полотно дороги. Характеристики используемого акселерометра представлены в [7]. Представленные в данной работе материалы частотного и временного анализа данных с акселерометров являются развитием работ, изложенных в [8].

Методика эксперимента

Исходные данные получены автором в университете г. Лулеа (Швеция, Lulea Technical University) с февраля по май 2011 г. в рамках проекта iRoad совместно с Wolfgang Birk (координатор проекта), Roland Hostettler, а также Владимиром Бардовым. Эксперименты проводились на дороге вблизи г. Лулеа, имеющей 4 полосы движения (по две в каждом направлении), движение регулярное. Разрешенная максимальная скорость 110 км/ч. Условия эксперимента: зимний день, температура воздуха -5°C . На дороге не было гололеда. Сенсорный узел был установлен на дорожной разметке и состоял из акселерометра (частота дискретизации 22000 Гц) и магнитометра (частота дискретизации 200 Гц). Данные предварительно усиливались с помощью усилителя У (рис. 1), затем передавались на компьютер с помощью USB-кабеля, записывались с помощью специальной программы и обрабатывались в пакете MATLAB.

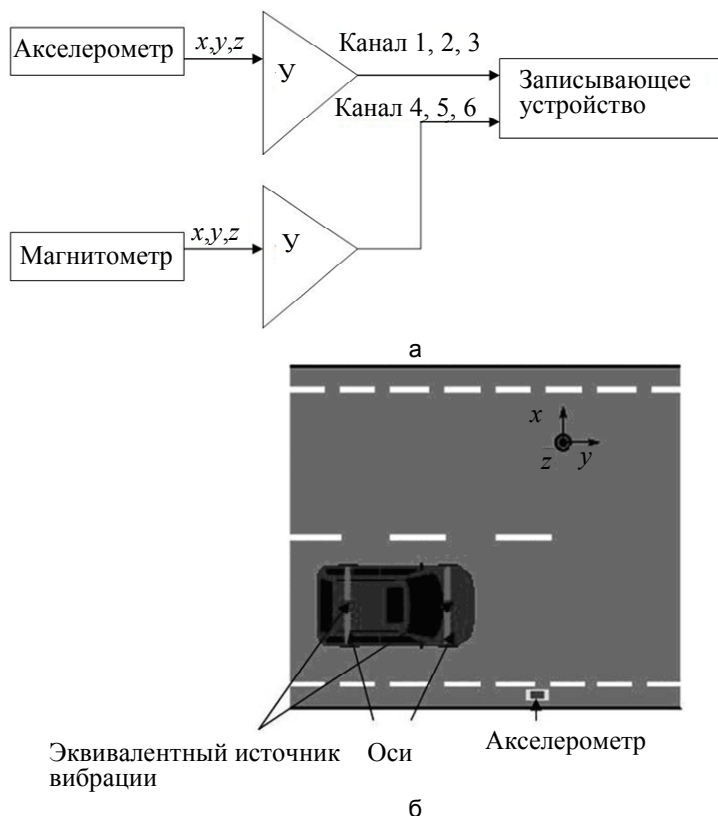


Рис. 1. Схема регистрации: функциональная схема измерительного блока (а); схема расположения датчика и система координат (б)

Следует отметить, что система предполагает проведение измерений дистанционно, с использованием модуля Wi-Fi, что делает их достаточно мобильными и привлекательными для потенциальных пользователей. На рис. 1 показана функциональная схема измерительной установки (рис. 1, а) и схема расположения сенсорного узла (рис. 1, б), а на рис. 2 – фотография сенсорного узла. В эксперименте сен-

сорный узел крепился к дорожной разметке с помощью клейкого материала. При успешной реализации проекта предусмотрена установка этого узла в специальных отверстиях в местах дорожной разметки.



Рис. 2. Сенсорный узел измерительной установки

Временной анализ

Во временной области проанализированы форма сигнала, его продолжительность и уровень. Сигнал представлен в виде временных рядов, которые дают некоторую первичную информацию [4–6], такую, к примеру, как повышенная сейсмическая активность.

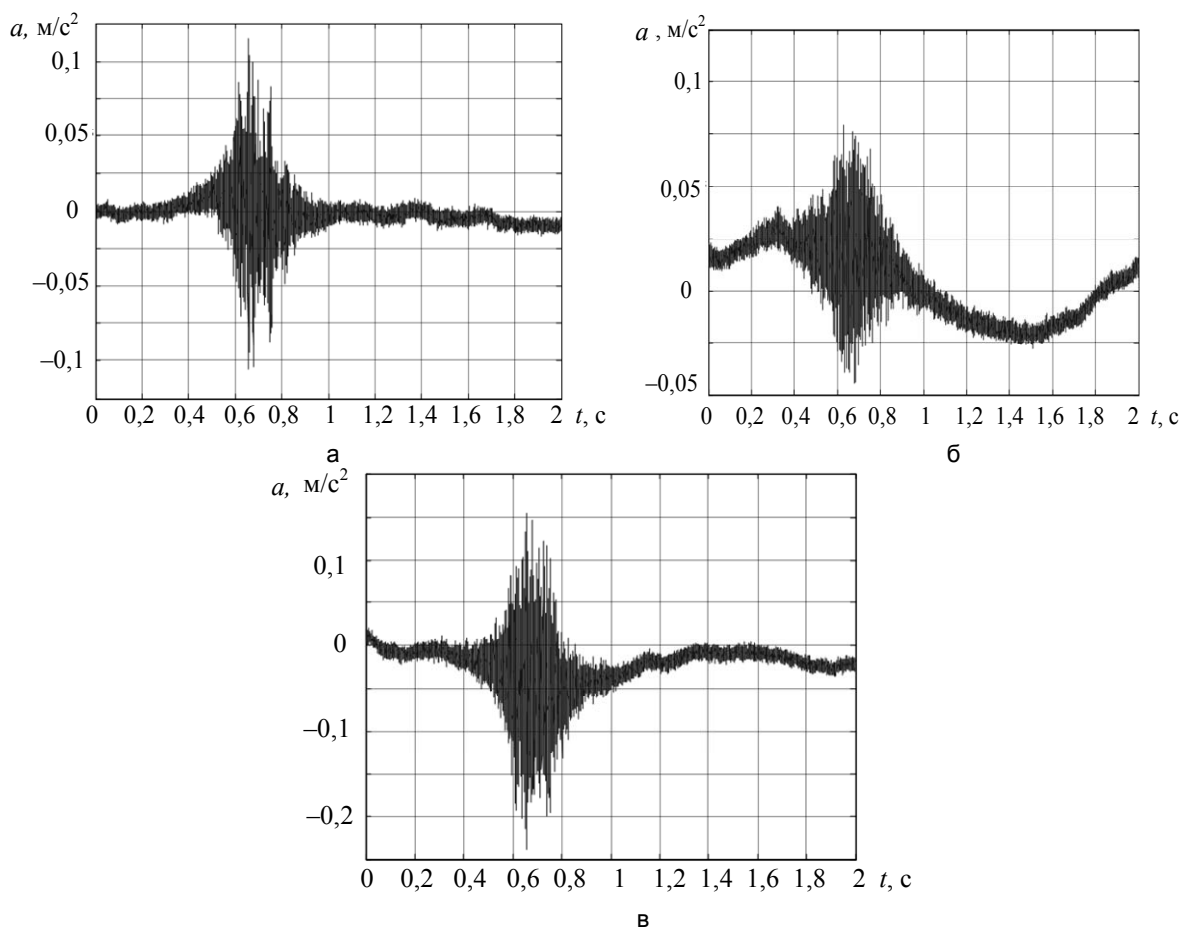


Рис. 3. Вибрации дороги в трех перпендикулярных направлениях: (а – $0x$; б – $0y$; в – $0z$). Легковой автомобиль, скорость $V=22$ м/с

Рассмотрим компоненты вибрации по трем осям $0x$, $0y$, $0z$ для легкового автомобиля (рис. 3). Ось $0y$ направлена вдоль полотна в направлении движения, ось $0x$ – вдоль полотна перпендикулярно движению, и ось $0z$ – перпендикулярно полотну дороги. Сейсмические волны, вызванные любой формой возбуждения, распространяются как объемные (P - и S -волны) и поверхностные (волны Лява и Рэлея). В за-

висимости от способа распространения волны имеют различную степень затухания. Поверхностные волны испытывают ослабление $1/\sqrt{r}$ в однородной среде, где r – расстояние от источника волны до рассматриваемой точки. Объемные волны ослабляются пропорционально $1/r$ в однородной среде [4]. Р-волны и волны Лява являются основной причиной горизонтальных вибраций в x - и y -направлениях. Волны Рэлея являются главной причиной вертикальных вибраций в z -направлении. Кроме того, в [4–6] было показано, что колебания в близких к поверхности слоях несут около $2/3$ энергии в поверхностной волне, и только $1/3$ распространяется в виде объемных волн.

Такая же закономерность сохраняется и для других типов ТС. В частности, авторами получены подобные сигналы и для грузового транспорта. По этой причине дальнейший спектральный анализ был ограничен осью Oz .

Спектральный анализ

При частотном анализе были применены быстрое преобразование Фурье (БПФ) и короткооконное преобразование Фурье (КОПФ).

На рис. 4, а, представлен результат БПФ сигнала, полученного от легкового автомобиля, движущегося со скоростью 22 м/с; на рис. 4, б – сигнала от грузовика с прицепом, движущегося со скоростью 25 м/с. На рис. 5 показаны результаты КОПФ для тех же сигналов. Затемненные области на рис. 5 соответствуют наибольшим амплитудам сигнала (более высокоэнергетическим компонентам). На рис. 6 для наглядности показаны в трехмерном виде совмещенные изображения двух видов преобразования.

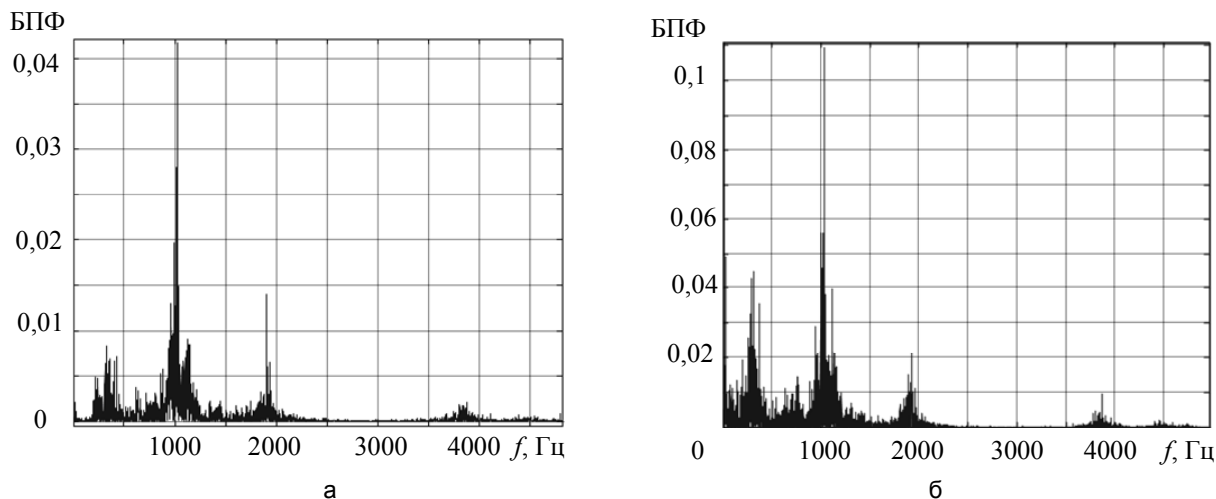


Рис. 4. Результат БПФ для сигнала от легкового автомобиля $V=22$ м/с (а) и грузовика с прицепом $V=25$ м/с (б)

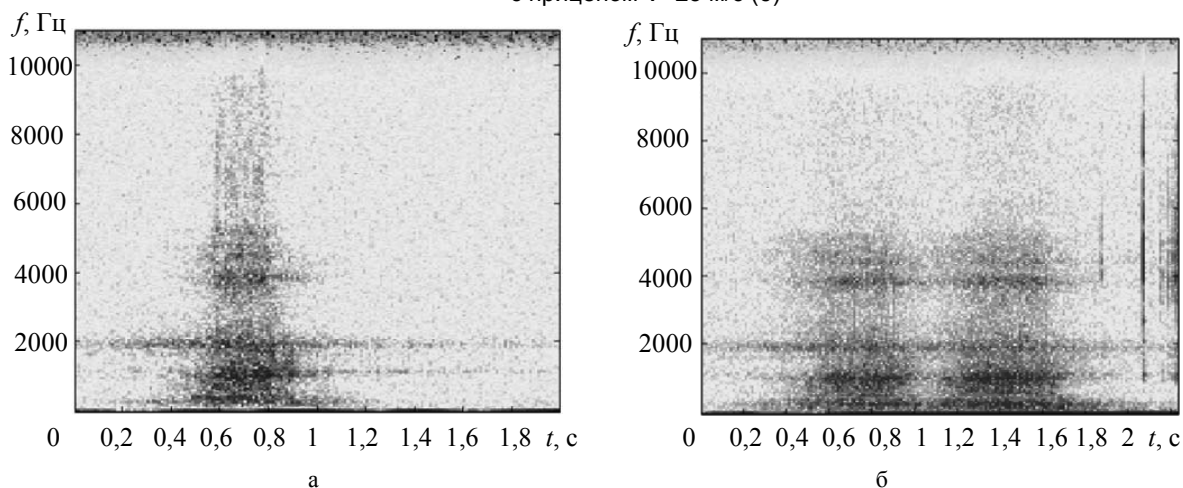


Рис. 5. Результат КОПФ для сигнала от легкового автомобиля $V=22$ м/с (а) и грузовика с прицепом $V=25$ м/с (б)

Частотные характеристики сигналов, полученных от разных видов транспорта, имеют схожие компоненты. В общем случае можно выделить две характерные полосы частот: до 1750 Гц и от 2 кГц до 3 кГц. Отчетливо видно, что наиболее информативной является полоса частот от 850 Гц до 1750 Гц. Однако сигнал вибрации, вызванной проездом грузовика, содержит высокоэнергетические компоненты на

более высоких частотах, от 3,5 кГц до 4 кГц. Это свойство сигнала может быть использовано для идентификации транспортных средств.

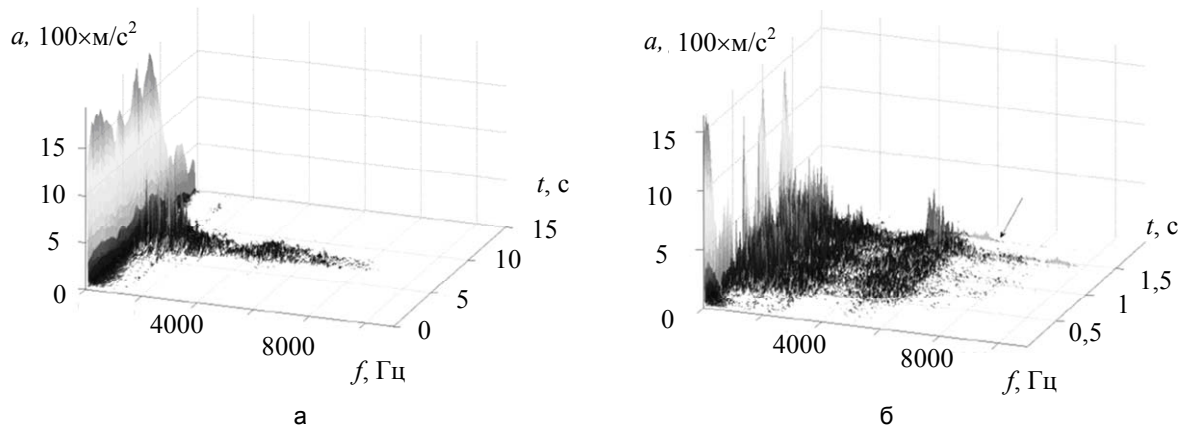


Рис. 6. Совмещение короткооконного и быстрого преобразований Фурье: легковая машина $V=22$ м/с (а); грузовик с прицепом $V=25$ м/с (б)

Вместе с тем анализ спектральных характеристик (рис. 4–6) показывает, что максимум энергии сигнала приходится на частотный диапазон 850–1750 Гц. Однако он содержит значительное число как высокочастотных, так и низкочастотных помех, вызванных присутствием на дороге посторонних источников шумов, к примеру, связанных с проведением ремонтных работ. Дальнейшая обработка сигнала целесообразна после отфильтровывания сигналов помех полосовым фильтром. В этой работе авторами был предложен эллиптический фильтр, обеспечивающий максимальное затухание в полосе задерживания по сравнению с другими фильтрами [9].

Фильтрация с применением эллиптического полосового фильтра

Как отмечено выше, предварительная обработка сигнала с помощью полосового фильтра позволяет снизить уровень шума и помех для более качественной обработки сигнала. Исходя из полученных ранее результатов спектрального анализа, предложены следующие параметры полосового фильтра:

- нижняя частота среза $f_l=850$ Гц;
- верхняя частота среза $f_u=1750$ Гц;
- максимальная пульсация полосы пропускания $R_p=1$ дБ;
- минимальное ослабление в полосе задерживания $R_s=50$ дБ.

Расчетные логарифмическая амплитудно-частотная (ЛАЧХ) и фазо-частотная (ЛФЧХ) характеристики такого фильтра показаны на рис. 7. Вид исходного сигнала (рис. 6) после фильтрации показан на рис. 8.

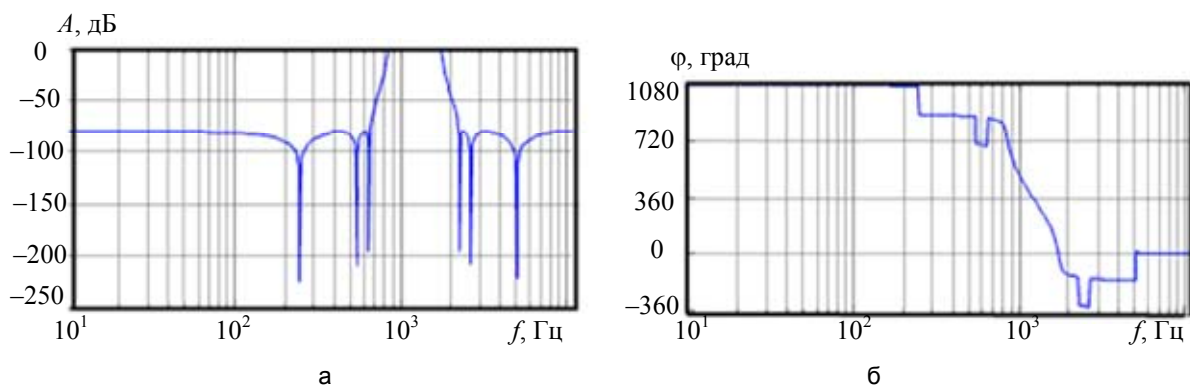


Рис. 7. ЛАЧХ (а) и ЛФЧХ (б) полосового фильтра

Последующая обработка сигнала от движущихся транспортных средств включает устранение отрицательных значений путем возведения в квадрат, сглаживание с помощью скользящего окна или фильтра нижних частот и нахождение амплитудного порога для обнаружения транспортного средства и временного порога, позволяющего разделять транспортные средства, движущиеся в сравнительной близости. Последовательность обработки сигнала, полученного от движущегося легкового автомобиля (рис. 3, в), и его форма на различных этапах обработки представлены на рис. 8 (а–г). Отметим, что данным методом можно определять скорости транспортных средства с использованием двух акселерометров.

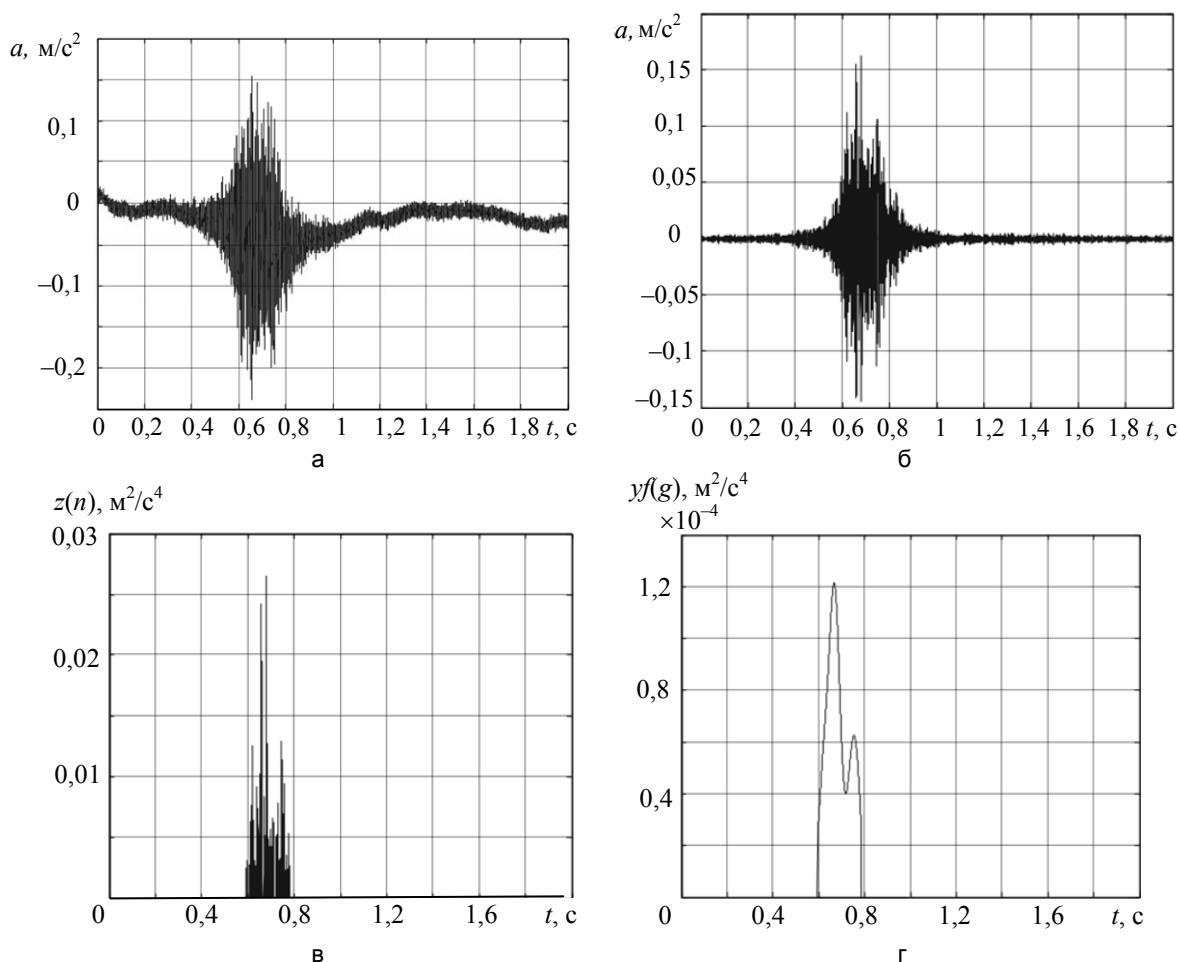


Рис. 8. Сигнал датчика на разных этапах обработки: исходный сигнал с акселерометра (а); сигнал после полосового фильтра (рис. 7) (б); сигнал после возведения в квадрат и сравнения с пороговым значением (в); сигнал $yf(g)$ после сглаживания скользящим окном, фильтром низких частот и удалением помех (г)

Заключение

Полученные результаты могут быть применены в реальных устройствах организации дорожного движения для построения интеллектуальных систем с целью обнаружения транспортных средств, идентификации их классов, подсчета количества ТС и мониторинга их скорости.

Концепция разрабатываемой системы состоит в следующем. Формируется сеть сенсорных узлов, установленных на дорожной разметке на расстоянии 50 м друг от друга. Датчики крепятся к дорожному покрытию в районе дорожной разметки с помощью специального клея или монтируются в дорожное покрытие. Возможно их питание от солнечных батарей. Они обмениваются информацией между собой и серверами за пределами дороги.

Датчик может менять цвет или подавать иные световые сигналы (к примеру, мигать) за несколько сотен метров до начала полосы гололеда или какой-то другой критической ситуации. Датчики могут передавать сигнал ближайшему дорожному знаку, который также подает предупреждающие сигналы о необходимости уменьшить скорость, сменить полосу движения и т.д. Иначе говоря, сенсоры передают всю полезную информацию транспортной инфраструктуре, которая затем использует ее для обеспечения безопасности и повышения качества дорожного движения. В данный момент для тестирования развернута сеть из 400 датчиков, которые выполняют вышеуказанные функции. Таким образом, реализуется концепция «умных дорог».

Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (соглашение № 14.В37.21.0659).

Литература

1. Birk W., Osipov E., Eliasson J., iRoad - cooperative road infrastructure systems for driver support // 16th ITS World Congress. – Stockholm, Sweden, 21–25 September 2009. – P. 1–8.
2. Birk W., Osipov E. On the design of cooperative road infrastructure systems // Reglermote 2008. – June 2008. – P. 349–355.

3. Birk W., Eliasson J., Lindgren P., Osipov E., Riliskis L. Road Surface Networks Technology Enablers for Enhanced ITS // 2nd IEEE Vehicular Networking Conf., VNC 2010. – Jersey City, NJ, 2010. – 13–15 Dec. – P. 152–159.
4. Hostettler R., Birk W., Lundberg Nordenvaad M. Surface mounted vehicle property sensing for cooperative vehicle infrastructure systems // Proceedings of the 69th IEEE Vehicular Technology Conference. – Barcelona, Spain, 26–29 April 2009. – P. 1–5.
5. Hostettler R. Traffic Counting Using Measurements of Road Surface Vibrations // Master's thesis. – Lulea University of Technology, 2009. – P. 1–48.
6. Hostettler R., Birk W., Nordenvaad L.M. On the Feasibility of Road Vibrations-based Vehicle Property Sensing // IET Intelligent Transport Systems. – 2010. – V. 4. – № 4. – P. 356–364.
7. Datasheet Model 832 accelerometer [Электронный ресурс]. – Режим доступа: http://www.meas-spec.com/product/t_product.aspx?id=5593, свободный. Яз. англ. (дата обращения 10.03.2011).
8. Обертов Д.Е., Бардов В.М. Алгоритм идентификации классов транспортных средств с помощью акселерометров // Информационно-управляющие системы. – 2012. – № 5 (60). – С. 15–18.
9. Лем Г. Аналоговые и цифровые фильтры. Расчет и реализация: Пер. с англ. – М.: Мир, 1982. – 592 с.

Обертов Дмитрий Евгеньевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, dimondking@mail.ru

УДК 681.51

АНАЛИЗ ЛИНЕЙНЫХ СИСТЕМ С ПЕРЕМЕННЫМИ ПАРАМЕТРАМИ ДЛЯ СИНТЕЗА ИНТЕРВАЛЬНЫХ НАБЛЮДАТЕЛЕЙ¹

С.Г. Чеботарев, А.С. Кремлев

Анализируются условия синтеза интервальных наблюдателей для линейных систем с неизвестными нестационарными параметрами. В англоязычной литературе этот класс систем получил наименование Linear Parameter-Varying Systems (LPV systems). Рассматриваются возможные варианты соблюдения свойств кооперативности системы, проводится анализ условий, при которых для данной комбинации свойств системы возможно построение интервального наблюдателя, позволяющего получить область оценок переменных состояния системы, гарантированно содержащую фактическое значение состояния в данный момент времени.

Ключевые слова: интервальные наблюдатели, кооперативность, мецлеровость, параметрическая неопределенность.

Введение

Проблема оценки неизмеряемого вектора состояния очень актуальна, и ее решение требуется во многих случаях [1–3]. В некоторых ситуациях, в связи с наличием неопределенности (параметрической и/или сигнальной), построение обычного наблюдателя, сходящегося в случае отсутствия шума к идеальному значению состояния, невозможно. Тем не менее, интервальная оценка остается возможной. Под интервальной оценкой здесь понимается работа интервального наблюдателя, который гарантирует оценивание множества допустимых значений для вектора состояний системы. Размер рассчитанного множества должен быть пропорционален неопределенности модели объекта. Неопределенности рассматриваются как детерминированные, но неизвестные функции времени. С этими ограничениями на параметры можно оценить границы ненаблюдаемых переменных: это так называемый «робастный» метод. Обычно интервальные наблюдатели генерируют два вектора оценок – минимальных и максимальных значений для каждого элемента вектора состояний объекта.

Есть несколько подходов к построению интервальных наблюдателей [4, 5–7]. В настоящей работе рассматриваются ситуация построения интервального наблюдателя, основанного на теории монотонных систем [4, 7]. Этот подход был расширен в работе [8] на случай нелинейных систем с использованием LPV-представления с известными минорной и мажорной матрицами, а в работе [9] – для наблюдаемых нелинейных систем. Одно из самых сложных допущений для построения интервального наблюдателя, связанное с кооперативностью динамики ошибки интервальной оценки, было решено в [9, 10]. Известно, что построение интервального наблюдателя возможно, если динамика ошибки оценки обладает свойствами кооперативности и устойчивости. Было показано, что при некоторых мягких условиях, применяя подобное преобразование координат, гурвицеву матрицу можно преобразовать в гурвицеву и мецлерову матрицу (кооперативную матрицу). Матрица преобразования – это решение уравнения Сильвестра, и конструктивный порядок решения этого уравнения был приведен в [9].

¹ Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (соглашение № 14.B37.21.1928).

Работа организована следующим образом. Некоторые основные факты из теории интервального оценивания приведены в разделе «Общие сведения». Анализ систем и условия построения интервального наблюдателя для различных комбинаций их свойств описаны в разделе «Основной результат».

Общие сведения

Для матрицы $A \in R^{m \times n}$ определим $\bar{A} = \max\{0, A\}$, $\underline{A} = \bar{A} - A$. Запись $A \in M$ означает, что матрица A – мецлерова, т.е. имеет неотрицательные элементы вне главной диагонали.

Лемма 1. Пусть $x \in R^n$ будет вектором переменных, $\underline{x} \leq x \leq \bar{x}$ для некоторых $\underline{x}, \bar{x} \in R^n$, и $A \in R^{m \times n}$ будет постоянной матрицей, тогда

$$\bar{A}\underline{x} - \underline{A}\bar{x} \leq Ax \leq \bar{A}\bar{x} - \underline{A}\underline{x}.$$

Доказательство. Отметим, что $Ax = (\bar{A} - \underline{A})x$, что для $\underline{x} \leq x \leq \bar{x}$ дает необходимые оценки.

Напомним, что матрица $A \in R^{m \times n}$ называется гурвицевой, если все ее собственные значения имеют отрицательную вещественную часть. Любое решение линейной системы

$$\dot{x} = Ax + \omega(t), \omega: R_+ \rightarrow R^n,$$

с $x \in R^n$ и мецлеровой матрицей A поэлементно неотрицательно для всех $t \geq 0$ при условии, что $x(0) \geq 0$ [11]. Такие динамические системы называются кооперативными (монотонными) [11].

Лемма 2. Даны матрицы $A \in R^{m \times n}$, $R \in R^{m \times n}$ и $C \in R^{p \times n}$. Если существует матрица $L \in R^{n \times p}$ такая, что матрицы $A - LC$ и R имеют одинаковые собственные значения, тогда $R = S^{-1}(A - LC)S$, где матрица $S \in R^{n \times n}$ при условии, что пары $(A - LC, e_1)$ и (R, e_2) наблюдаемы для некоторых $e_1 \in R^{1 \times n}$, $e_2 \in R^{1 \times n}$. Этот результат был использован в [9] для построения интервальных наблюдателей для линейных стационарных систем с мецлеровой матрицей R (основная трудность состоит в доказательстве существования вещественной матрицы S и обеспечении конструктивного метода ее расчета).

Основной результат

Рассматривается следующий вид линейной системы, имеющей зависимость от неизвестных нестационарных параметров $p \in P$

$$\begin{cases} \dot{x} = A(p, t)x + d(t, y, u), \\ y = Cx, \end{cases} \quad (1)$$

где $x \in R^n$, $u \in R^m$, $y \in R^p$ – это состояние, вход и выход системы (1) соответственно, $x \in P \in R^q$ – это вектор неизвестных сигналов или параметров. В настоящей работе рассматривается случай без шума в канале измерения, предлагаемый результат может быть расширен на случай с шумом, это расширение опущено для краткости изложения.

Допущение 1. Состояние x и вход u ограничены – стандартные допущения в теории оценивания.

Допущение 2. Если даны границы $[\underline{x}, \bar{x}]$ состояния x , то значения функции $d(t, y, u)$ заключены в интервале $[d(t, y, u), \bar{d}(t, y, u)]$.

Допущение 3. Существует коэффициент усиления наблюдателя L , который обеспечивает устойчивость нестационарной матрицы динамики наблюдателя с матрицей $P(t)$ функции Ляпунова V ; это допущение определяет условия устойчивости динамики оценки:

$$V = \underline{e}^T P(t) \underline{e} + \bar{e}^T P(t) \bar{e}.$$

Множество работ по построению интервального наблюдателя [4, 7–9] посвящено случаю с постоянной матрицей A (или в результате некоторых преобразований ошибка оценки может быть представлена в форме с постоянной матрицей A), и далее может быть найден такой коэффициент усиления наблюдателя L , что $A - LC$ является гурвицевой и мецлеровой. В дальнейших работах [12] снимается предположение о том, что существует коэффициент усиления наблюдателя, который делает динамику ошибки оценки устойчивой и кооперативной. Коэффициент усиления наблюдателя должен, как обычно, обеспечить устойчивость ошибки наблюдения, а далее предлагается статическое преобразование координат, которое обеспечивает требуемое свойство кооперативности.

Рассматриваемый метод построения интервального наблюдателя для линейных систем с переменными параметрами тесно связан с теорией дифференциальных неравенств и кооперативными системами. Основное условие построения подобного устройства оценки, работающего в условиях интервальности и неизвестного характера изменения параметров, от которых зависит система, связано с кооперативностью динамики ошибки интервальной оценки. Далее рассматриваются возможные варианты выполнения данного ограничения, на основе чего строится дальнейший анализ.

1. $(t) \geq 0; \mathbf{A}(\mathbf{p}, t) \in M$.
2. $\mathbf{x}(t) \geq 0; \mathbf{A}(\mathbf{p}, t) \notin M; \exists \mathbf{L}: (\mathbf{A}(\mathbf{p}, t) - \mathbf{LC}) \in M$.
3. $\mathbf{x}(t) \geq 0; \mathbf{A}(\mathbf{p}, t) \notin M; \exists \mathbf{L}: (\mathbf{A}(\mathbf{p}, t) - \mathbf{LC}) \in M; \exists \mathbf{S}: \mathbf{R} = \mathbf{S}^{-1}(\mathbf{A}(\mathbf{p}, t) - \mathbf{LC})\mathbf{S} \in M$.
4. $\mathbf{x}(t) \in \mathbb{R}^n; \mathbf{A}(\mathbf{p}, t) \in M; \exists \mathbf{X}: \mathbf{x}^* = \mathbf{x} - \mathbf{X} \geq 0$.
5. $\mathbf{x}(t) \in \mathbb{R}^n; \mathbf{A}(\mathbf{p}, t) = \mathbf{A} \in M; \exists \Delta \mathbf{A}: -\Delta \mathbf{A} \bar{\mathbf{x}} \leq d(t, \mathbf{x}) \leq \Delta \mathbf{A} \bar{\mathbf{x}}$.
6. $\mathbf{x}(t) \in \mathbb{R}^n; \mathbf{A}(\mathbf{p}, t) \notin M; \exists \mathbf{L}: (\mathbf{A}(\mathbf{p}, t) - \mathbf{LC}) \in M; \exists \mathbf{X}: \mathbf{x}^* = \mathbf{x} - \mathbf{X} \geq 0$.
7. $\mathbf{x}(t) \in \mathbb{R}^n; \mathbf{A}(\mathbf{p}, t) = \mathbf{A} \notin M; \exists \mathbf{L}: (\mathbf{A}(\mathbf{p}, t) - \mathbf{LC}) \in M; \exists \Delta \mathbf{A}: -\Delta \mathbf{A} \bar{\mathbf{x}} \leq d(t, \mathbf{x}) \leq \Delta \mathbf{A} \bar{\mathbf{x}}$.
8. $\mathbf{x}(t) \in \mathbb{R}^n; \mathbf{A}(\mathbf{p}, t) \notin M; \exists \mathbf{L}: (\mathbf{A}(\mathbf{p}, t) - \mathbf{LC}) \in M; \exists \mathbf{S}: \mathbf{R} = \mathbf{S}^{-1}(\mathbf{A}(\mathbf{p}, t) - \mathbf{LC})\mathbf{S} \in M$.
 $\psi = \mathbf{S}^{-1}\mathbf{x}; \exists \Phi: \psi^* = \psi - \Phi \geq 0$.
9. $\mathbf{x}(t) \in \mathbb{R}^n; \mathbf{A}(\mathbf{p}, t) \notin M; \exists \mathbf{L}: (\mathbf{A}(\mathbf{p}, t) - \mathbf{LC}) \in M; \exists \mathbf{S}: \mathbf{R} = \mathbf{S}^{-1}(\mathbf{A}(\mathbf{p}, t) - \mathbf{LC})\mathbf{S} \in M$;
 $\mathbf{R} - \text{const}; \psi = \mathbf{S}^{-1}\mathbf{x}; \exists (\bar{\mathbf{S}}^{-1} - \underline{\mathbf{S}}^{-1})\Delta \mathbf{A}(\bar{\mathbf{S}} + \underline{\mathbf{S}}) = \Delta \mathbf{A}^*: -\Delta \mathbf{A}^* \bar{\psi} \leq d(t, \psi) \leq \Delta \mathbf{A}^* \bar{\psi}$.

Отдельно рассматривается ситуация, когда состояния системы неотрицательны (п.п. 1–3), и ситуация, когда состояния знакопеременны (п.п. 4–9). Такое разделение обусловлено различным характером поведения границ полученного интервала оценок состояния. Это видно, если записать уравнение ошибки для рассматриваемой системы.

Пусть $\underline{\mathbf{x}}, \bar{\mathbf{x}}$ обозначают оценки состояния системы для нижней и верхней границ интервала соответственно. Тогда производные оценок состояния системы для каждой из двух границ примут вид

$$\dot{\underline{\mathbf{x}}} = \underline{\mathbf{A}}\underline{\mathbf{x}} + \underline{d}(t, y, u) + \underline{\mathbf{L}}\mathbf{C}(\underline{\mathbf{x}} - \underline{\mathbf{x}}),$$

$$\dot{\bar{\mathbf{x}}} = \bar{\mathbf{A}}\bar{\mathbf{x}} + \bar{d}(t, y, u) + \bar{\mathbf{L}}\mathbf{C}(\bar{\mathbf{x}} - \bar{\mathbf{x}}).$$

Запишем уравнение ошибки для нижней границы интервала как разность между состоянием и его оценкой:

$$\underline{\mathbf{e}} = \underline{\mathbf{x}} - \underline{\mathbf{x}}, \quad \dot{\underline{\mathbf{e}}} = \dot{\underline{\mathbf{x}}} - \dot{\underline{\mathbf{x}}} = \mathbf{A}(\mathbf{p}, t)\underline{\mathbf{x}} + d(t, y, u) - \underline{\mathbf{A}}\underline{\mathbf{x}} - \underline{d}(t, y, u) - \underline{\mathbf{L}}\mathbf{C}(\underline{\mathbf{x}} - \underline{\mathbf{x}}).$$

Для верхней границы интервала уравнение ошибки записывается аналогично.

После преобразований динамика ошибки принимает следующий вид:

$$\dot{\underline{\mathbf{e}}} = (\underline{\mathbf{A}} - \underline{\mathbf{L}}\mathbf{C})\underline{\mathbf{e}} + (\mathbf{A}(\mathbf{p}, t) - \underline{\mathbf{A}})\underline{\mathbf{x}}.$$

Исходя из вида уравнения, можно сделать вывод о том, что второе слагаемое оказывает влияние на значение ошибки, следствием чего является возможность смены относительной ориентации границ интервала оценок состояния системы, если состояние меняет знак. Следовательно, системы с неотрицательными и знакопеременными значениями состояния стоит рассматривать отдельно и во втором случае предпринимать действия по исключению влияния данного фактора. Это можно осуществить двумя способами – сдвигом координат в область с положительными значениями состояний (п.п. 4, 6, 8) или записью уравнений, описывающих динамику наблюдателей, в другом виде, с использованием постоянной матрицы $\mathbf{A}$ и матрицы $\Delta \mathbf{A}$, состоящей из неотрицательных элементов (п.п. 5, 7, 9).

Как уже отмечалось, необходимым условием построения интервального наблюдателя является особый вид матрицы, описывающей динамику объекта и наблюдателей. Они должны обладать свойством мецлеровости. Кроме того, так как на данном этапе рассматриваются устойчивые линейные системы, эти матрицы также должны удовлетворять условиям гурвицевости. Здесь можно отметить три варианта:

1. система сразу обладает необходимыми свойствами мецлеровости и гурвицевости (п.п. 1, 4, 5); подобный вариант встречается крайне редко;
2. может быть найден такой коэффициент усиления наблюдателя $\mathbf{L}$, что $\mathbf{A} - \mathbf{LC}$ является гурвицевой и мецлеровой (п.п. 2, 6, 7);
3. коэффициент усиления наблюдателя $\mathbf{L}$ должен обеспечить лишь устойчивость ошибки наблюдения, а далее предлагается статическое преобразование координат, которое обеспечивает требуемое свойство кооперативности (п.п. 3, 8, 9).

Таким образом, при анализе условий синтеза интервальных наблюдателей для линейных систем с неизвестными нестационарными параметрами необходимо принять во внимание знакопеременность значений состояний рассматриваемой системы и способ получения матрицы, описывающей динамику наблюдателей, в мецлеровом виде. Предполагается также устойчивость системы и обеспечение устойчивости траекторий наблюдателей с помощью выбора коэффициентов усиления $\mathbf{L}$. Все возможные комбинации указанных условий приведены в п.п. 1–9. Следовательно, для любой системы из рассматриваемого класса можно построить интервальный наблюдатель, позволяющий получить область оценок переменных состояния системы, гарантированно содержащую фактическое значение состояния в данный момент времени, при соблюдении определенных условий, которые были проанализированы в данной работе.

Заключение

Показано, что интервальный наблюдатель позволяет получить область оценок переменных состояния системы, гарантированно содержащих фактическое значение состояния в данный момент времени. Получены условия построения подобного устройства оценки для рассматриваемого класса систем при различных комбинациях свойств системы.

Доказательство ограниченности и поиск условий ограниченности траекторий наблюдателя для каждого из представленных случаев является направлением будущих исследований.

Литература

1. Meurer T., Graichen K., Gilles E.-D. (Eds). Control and Observer Design for Nonlinear Finite and Infinite Dimensional Systems // Lecture Notes in Control and Information Sciences. – Springer, 2005. – V. 322. – P. 422.
2. Fossen T.I., Nijmeijer H. New Directions in Nonlinear Observer Design. – Springer, 1999. – 525 p.
3. Nonlinear Observers and Applications. Lecture Notes in Control and Information Sciences. – Besançon G. (Ed.), Springer, 2007. – V. 363. – 224 p.
4. Bernard O., Gouzé J.L. Closed loop observers bundle for uncertain biotechnological models // J. Process Control. – 2004. – V. 14. – P. 765–774.
5. Jaulin L. Nonlinear bounded-error state estimation of continuous time systems // Automatica. – 2002. – V. 38. – № 2. – P. 1079–1082.
6. Kieffer M., Walter E. Guaranteed nonlinear state estimator for cooperative systems // Numerical Algorithms. – 2004. – V. 37. – P. 187–198.
7. Moisan M., Bernard O., Gouzé J.L. Near optimal interval observers bundle for uncertain bio-reactors // Automatica. – 2009. – V. 45. – № 1. – P. 291–295.
8. Raïssi T., Videau G., Zolghadri A. Interval observers design for consistency checks of nonlinear continuous-time systems // Automatica. – 2010. – V. 46 – № 3. – P. 518–527.
9. Raïssi T., Efimov D., Zolghadri A. Interval state estimation for a class of nonlinear systems // IEEE Trans. on Automatic Control. – 2012. – V. 57. – № 1. – P. 260–265.
10. Mazenc F., Bernard O. Interval observers for linear time-invariant systems with disturbances // Automatica. – 2011. – V. 47. – № 1. – P. 140–147.
11. Smith H.L. Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems of Surveys and Monographs // AMS. – Providence, 1995. – V. 41. – 174 p.
12. Efimov D., Raïssi T., Chebotarev S., Zolghadri A. Interval State Observer for Nonlinear Time Varying Systems // Automatica. – 2012 [Электронный ресурс]. – Режим доступа: <http://www.sciencedirect.com/science/article/pii/S0005109812004761>, своб., Яз. англ. (дата обращения 02.11.2012).

Чеботарев Станислав Геннадьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, freest5@gmail.com

Кремлев Артем Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, kremlev_artem@mail.ru

УДК 621.371

МОДЕЛИРОВАНИЕ ПОЛЯ DVB-T В УСЛОВИЯХ ГОРОДСКОЙ ЗАСТРОЙКИ

А.Ф. Ломакин, С.И. Школьный

Результаты расчетов напряженности поля DVB-T по модели Лонгли–Райса сопоставлены с измерениями, выполненными на местности. Прогноз и серия натурных измерений напряженности поля выполнены для различных типов трасс распространения волн в условиях сильно пересеченной местности и плотной городской застройки. Обсуждаются результаты сравнений модельных и экспериментальных данных.

Ключевые слова: DVB-T, распространение радиоволн, модель Лонгли–Райса.

Введение

В процессе внедрения цифрового эфирного телевидения стандарта DVB-T в городах Приморского края поставлена задача обеспечения полного покрытия зоны обслуживания с заданными показателями качества (95% по времени и 95% на местности) при условии допустимого значения уровня помехи на телевизионные приемники при портативном приеме. Важным этапом при проектировании любой сети эфирного вещания является расчет гарантированной зоны обслуживания, в связи с этим адекватное моделирование распространения радиоволн в городской среде в настоящее время является достаточно актуальной задачей. С точки зрения распространения радиоволн современный город представляет собой сложную неоднородную электродинамическую структуру, в которой рельеф подстилающей поверхности оказывает определяющее влияние на формирование поля в точке приема. Для расчета зон покрытия раз-

работано множество моделей, различающихся по своим методам, способам параметризации характеристик окружающей среды, точности, скорости и применимости для различных условий. На текущий момент зона покрытия телевизионного передатчика рассчитывается в подавляющем большинстве случаев на основе статистических методов, которые длительное время являются основными и по-прежнему актуальны. В диапазоне частот телевизионных каналов (I–IV диапазоны) для расчета напряженности поля применяются модели Хаты, Окамуры, Парсона, Сакагами–Кубои, Булингтона, COST-231, ITU-R.1546 и др. Однако эти модели малоприменимы для условий Владивостока с его хаотичной застройкой, сложным рельефом и наличием участков распространения радиоволн над водной поверхностью. Среди статистических моделей, на взгляд авторов, для условий Владивостока более подходит модель Лонгли–Райса, учитывающая неоднородности рельефа и некоторые недостатки отмеченных выше моделей. К особенностям данной модели стоит отнести относительно малую чувствительность к частоте сигнала, позволяющую вести расчет для достаточно широкополосного сигнала, каким является сигнал DVB-T, сводящийся к расчету для центральной частоты.

Цель данной работы – предварительная оценка границ применимости модели Лонгли–Райса при расчете зон покрытия цифрового телевизионного передатчика для различных трасс распространения сигнала.

Моделирование поля DVB-T

В настоящее время в Приморском крае завершается строительство сети DVB-T. На завершающем этапе требуется ее оптимизация на самом сложном участке – в г. Владивосток и на прилегающих к нему территориях. Город расположен на полуострове Муравьева-Амурского и с трех сторон окружен морем. Рельеф центральной части города определяют несколько сопков высотой до 250 м и залив Золотой рог, который глубоко вдается в полуостров и как бы разрезает город на две части. Городская телебашня расположена на сопке Орлиное гнездо высотой 192 м и имеет собственную высоту 195 м. Из-за географического положения и сложного рельефа во Владивостоке наблюдается дефицит пригодной для строительства земли. По плотности застройки центральной части город может сравниться с мегаполисами.

Как уже отмечалось, традиционно применяемые в России статистические методы прогнозирования распространения радиоволн [1, 2] малоприменимы для условий Владивостока. Хотя в них и включены некоторые инструменты, позволяющие учесть особенности рельефа, такие модели обеспечивают достоверный прогноз только при достаточно однородной подстилающей поверхности. С другой стороны, относительно простые и точные методы, основанные на приближении геометрической оптики [3], также не могут быть использованы, так как они требуют подробной трехмерной карты города с высоким разрешением. Кроме этого, при проведении подобных расчетов для достаточно крупной территории предъявляются высокие требования к ресурсам компьютера, а сами расчеты занимают много времени.

Анализ литературных источников по использованию различных моделей прогнозирования распространения цифрового сигнала показал, что для расчетов зоны покрытия телевизионного передатчика на территориях со сложным рельефом и плотной городской застройкой более всего подходит модель Лонгли–Райса [4]. Данная модель не слишком известна в России, однако широко используется за рубежом. Хорошая достоверность модели Лонгли–Райса подтверждена практическими исследованиями [5]. Модель позволяет производить вычисление медианного значения ослабления для сложных трасс в диапазоне частот от 20 МГц до 10 ГГц и предназначена для определения медианных значений ослаблений для длинных трасс в условиях пересеченной местности. Модель включает в себя два варианта расчета: «точка–точка» и «точка–область». Первый предполагает наличие цифровой карты местности и предназначен для более точных расчетов, в то время как в варианте «точка–область» оценка производится по общим параметрам подстилающей поверхности. В данном исследовании расчеты проводились в режиме «точка–точка».

Для проведения расчетов был использован вариант реализации модели Лонгли–Райса в программном комплексе Terrain Analysis Package (TAP™, версия 6) фирмы Softwright [6] и цифровая карта территории с разрешением 3". Последовательно проводились расчеты и анализировались зоны покрытия территории г. Владивосток для мощностей передатчика 200 Вт, 500 Вт, 1 кВт, 2 кВт, 5 кВт. Расчеты выполнены при следующих параметрах сети DVB-T и окружающей среды:

- частота вещания – 37 телевизионный канал (598–606 МГц);
- высота подвеса передающей антенны 180 м;
- передающая антенна – панели Kathrein с усилением 14,1 дБи;
- поверхностная рефракция 301;
- удельная электропроводность почвы 0,005;
- относительная диэлектрическая проницаемость почвы 15;
- тип климата – континентальный умеренный.

Расчеты проводились для условий портативного наружного приема (заданные показатели качества приема – 95% по месту, 95% по времени). Высота приемной антенны в расчетах принималась равной 2 м.

Анализ результатов моделирования показал, что оптимальная мощность телевизионного передатчика для Владивостока составляет 2 кВт, при этом необходимый расчетный уровень сигнала в 63 дБмкВ/м [7] отмечается на удалении от точки вещания до 15 км, что обеспечивает необходимую зону обслуживания. Установлено, что внутри зоны покрытия существует несколько участков тени, обусловленных дифракционными потерями, вызванными сложным рельефом города. На этих участках прием ухудшен, а в некоторых случаях и невозможен. Такими «зонами тени» являются густонаселенные районы города. Кроме этого, практически вся пригородная зона города попадает в зону с недостаточным уровнем сигнала. При увеличении мощности передатчика до 5 кВт картина не меняется.

Таким образом, результаты модельных расчетов показывают, что при планировании цифрового телевидения DVB-T в режиме портативного приема нет возможности покрыть территорию г. Владивостока вещанием с одного передатчика.

Оценка результатов моделирования

Сопоставление экспериментальных данных и результатов моделирования проводилось на основе статистических расчетов среднего расхождения и среднеквадратичного отклонения между измеренными и рассчитанными значениями потерь в точке расположения приемной антенны. Натурные измерения напряженности электромагнитного поля, создаваемого цифровым телевизионным передатчиком мощностью 2 кВт, проводились с использованием анализатора спектра ETL Rohde&Schwarz и профессиональной измерительной антенны HE300 с коэффициентом усиления 10 дБи. Измерения производились на высоте 2 м над уровнем земли, что соответствует условиям портативного наружного приема. По результатам модельных расчетов были выделены пять характерных зон радиотени, в которых и проводились измерения в строгом соответствии с рекомендациями Международного союза электросвязи [7].

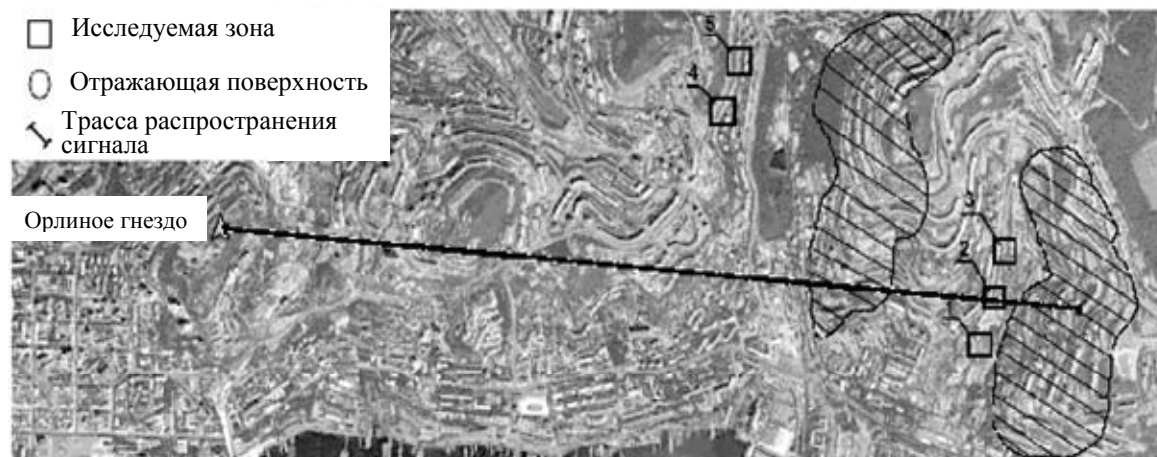


Рис. 1. Пространственное расположение зон радиотени и отражающих поверхностей

На рис. 1 представлено пространственное расположение характерных зон радиотени. Выделено 50 экспериментальных областей, из которых 17 относятся к открытому типу трассы, а 33 – к областям радиотени. Измерения уровня сигнала DVB-T проводились в каждой области не менее чем в 15 пространственных точках, общее число измерений составило 873. Измерения выполнялись по стандартной процедуре – в четырех точках приема, расположенных на расстоянии 40 длин волн исследуемого ТВ канала (в данном случае – 602 МГц, длина волны 50 см). Измерения уровня сигнала по четырем точкам осреднялись. Полученная медиана относилась к конкретной пространственной точке области, а рассчитанные медианы вновь осреднялись по областям. В пределах конкретной области все измерения проводились в течение одного дня.

На рис. 2 представлены предварительные результаты сравнения экспериментальных данных и данных, полученных в результате расчетов по модели Лонгли–Райса.

Видно, что модель позволяет получить достаточно точный прогноз напряженности поля DVB-T для открытых трасс (правая часть рис. 2), а в случае закрытых и полужакрытых трасс расчетные значения оказались сильно заниженными. Так, при проведении измерений было обнаружено, что в некоторых районах города уровень отраженного сигнала превышает 63 дБмкВ/м на уровне 2 м от земли, что позволяет вести уверенный прием сигнала на внешнюю портативную антенну. В то же время расчетные значения прямого сигнала для этих точек, полученные по модели Лонгли–Райса, составляют 25–30 дБмкВ/м. Известно, что технология DVB-T позволяет вести прием не только прямого сигнала, но и отраженного от объектов рельефа и зданий. Теоретические исследования [8] и натурные измерения [9] показывают, что при углах падения, близких к нормали, наблюдается зеркальное отражение электромагнитной волны от железобетонных поверхностей с коэффициентом отражения близким к 1. На практике при расчетах зоны

покрытия эфирных сетей для целей проектирования и оптимизации не учитывается возможность приема отраженного сигнала. В условиях сложного рельефа, в сочетании с плотной городской застройкой, данный фактор может оказывать влияние на размер зоны обслуживания.

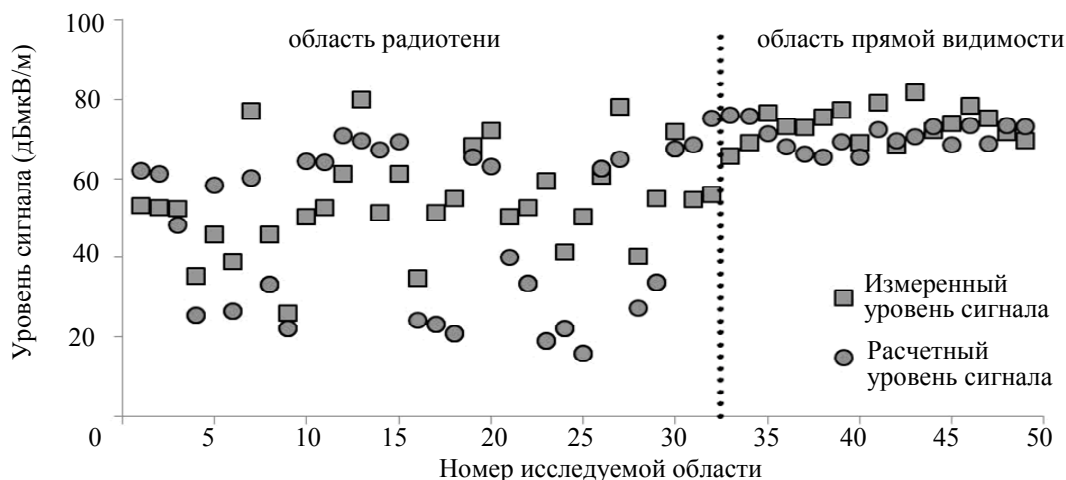


Рис. 2. Измеренные и расчетные значения напряженности поля DVB-T

В ходе исследования была сделана попытка численно оценить напряженность поля в зонах радиотени в предположении, что сигнал зеркально отражается от препятствий, расположенных перпендикулярно к направлению распространения. Плотная застройка Владивостока со зданиями, выполненными преимущественно из железобетонных конструкций, позволяет говорить о высокой степени достоверности такого предположения. Коэффициент отражения от железобетона принят авторами равным 0,91. Ослабление отраженного сигнала (рис. 3) рассчитывается как сумма ослабления прямого сигнала (линия 1) от телебашни «Орлиное гнездо» до отражающей поверхности (точка 2), потерь на отражение и ослабления сигнала (линия 3) от отражающей поверхности до исследуемой точки (точка 4).

Расчет ослабления на первом участке трассы производился по модели Лонгли–Райса в нескольких точках, полученные значения напряженности осреднялись. Особенностью второго участка трассы распространения сигнала являлась его небольшая протяженность (не превышает значений 1 км). Применение достаточно сложной модели Лонгли–Райса на этом отрезке пути нерационально, так как между отражающей поверхностью и оцениваемой точкой нет препятствий, попадающих в существенную зону распространения радиоволн. Известные формулы Введенского [10] и Окамура–Хата на расстояниях менее 1 км от точки излучения дают высокую погрешность. В настоящей работе использовалась достаточно простая формула, заимствованная из [11]:

$$E_{\text{дБмкВ/м}} = 64,5 + 20 \lg(H \cdot h) - 40 \lg R - \sqrt{\frac{H}{h}} \cdot 10 \lg\left(1 + \frac{R}{R_m}\right) - 5 \lg f, \quad (1)$$

где H и h – высота подъема передающей антенны и высота подвеса приемной антенны соответственно, м; f – частота, МГц; R – дальность от вещательного центра до точки приема, в которой осуществляется измерение напряженности поля E , км; $R_m = 4,12(\sqrt{H} + \sqrt{h})$ – дальность радиовидимости. Приведенная зависимость выражает собой аппроксимацию кривых, представленных в рекомендациях R.1546 [2] и может использоваться для коротких трасс. Результаты расчетов приведены в таблице.

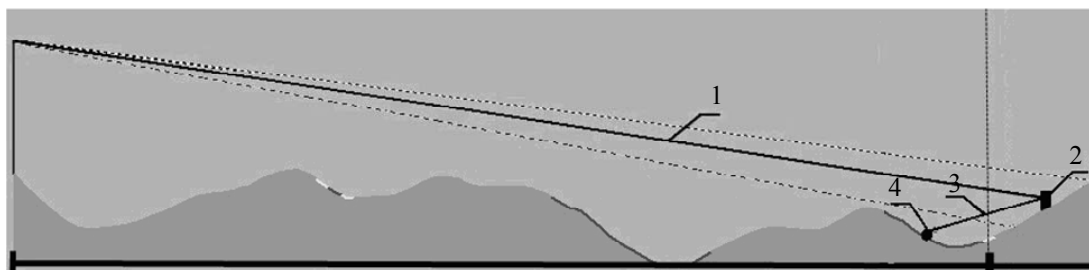


Рис. 3. Реальный профиль трассы распространения сигнала в зоне радиотени:

- 1 – участок трассы передатчик-отражающая поверхность, рассчитанный по модели Лонгли–Райса;
- 2 – точка отражения; 3 – участок трассы отражающая поверхность-точка приема, рассчитанный по формуле (1); 4 – точка приема

Номер зоны	Среднее расчетное значение напряженности поля на отражающем объекте $E_{отр}$, дБмкВ/м	Расчетное значение отраженного сигнала $E_{пр}$, дБмкВ/м	Среднее измеренное значение отраженного сигнала, дБмкВ/м	Различие между расчетным и измеренным значениями
1	75,9	53,4	52,6	-0,8
2	81,3	56,7	59,5	2,8
3	80,1	52,0	50,3	-1,7
4	79,0	48,4	51,3	2,9
5	83,3	52,4	55,1	2,7

Таблица. Сравнение медианных расчетных и измеренных значений напряженности поля DVB-T в зонах радиотени

Различия между расчетными и измеренными уровнями сигнала DVB-T невелики и в среднем не превышают 3 дБмкВ/м, что соизмеримо с точностью измерений. Среднеквадратичное отклонение между измеренными и рассчитанными значениями потерь по всем зонам не превышает 5,5 дБмкВ/м. Полученные предварительные результаты позволяют говорить о практической важности оценки уровня отраженного сигнала в рамках существующих параметрических моделей. Даже простой способ учета отраженного сигнала при прогнозировании поля DVB-T позволяет повысить точность модельных расчетов. Конечно, разработка параметрической модели, учитывающей отраженный сигнал, является сложной задачей, требующей большого количества экспериментальных данных и дополнительных исследований.

Заключение

Прогнозирование распространения радиоволн при помощи параметрической модели Лонгли–Райса показало, что плотная застройка и сложный рельеф приводят к появлению зон радиотени даже на небольших расстояниях от передающей антенны, что существенно затрудняет оптимизацию сети DVB-T. Измеренные и расчетные значения напряженности электромагнитного поля хорошо согласуются для открытых трасс, при этом прогноз напряженности поля несколько завышен, примерно на 3 дБмкВ/м по отношению к измеренным уровням сигнала. В случае закрытых и полужакрытых трасс расчетные значения оказались сильно заниженными. При экспериментальном исследовании проблемных зон распространения сигнала DVB-T на территории Владивостока установлено наличие многочисленных участков, где отраженный сигнал распространяется на значительные расстояния с хорошими качественными характеристиками, достаточными для уверенного приема на портативную антенну. Полученные предварительные результаты могут быть объяснены существенным упрощением модели Лонгли–Райса при учете некоторых особенностей распространения сигнала на реальных трассах. В то же время даже простые способы учета отраженных сигналов в модели показывают довольно хорошее соответствие расчетных и измеренных данных. Очевидно, что учет возможности приема отраженного сигнала мог бы позволить отказать от части дополнительных передатчиков и ретрансляторов при проектировании сети либо уменьшить их мощность.

Литература

- Okumura Y. Field strength and its variability in the VHF and UHF land mobile radio service // Review Elec. Commun. Lab. – 1968. – V. 16. – С. 825–843.
- Рекомендация МСЭ R.1546-3 (Критерии планирования для услуг наземного цифрового телевидения в диапазонах ОВЧ/УВЧ) [Электронный ресурс]. – Режим доступа: <http://www.itu.int/rec/R-REC-P.1546/>, свободный. Яз. рус. (дата обращения 25.08.2012).
- Карпов А.А. Модель городского многолучевого радиоканала с предварительной обработкой данных о городской застройке // Журнал радиоэлектроники. – 2008. – № 8. – С. 63–68.
- Hufford G., Longley A., Kissick W. Guide to the use of the ITS irregular terrain model in the area prediction mode, NTIA. – Tech. Rep. 1982. – P. 82–100.
- Perez-Fontan F., Hernando-Rabanos J.M. Comparison of irregular terrain propagation models for use in digital terrain data based radiocommunication system planning tools // IEEE Transactions on Broadcasting. – 1995. – V. 41. – № 2. – P. 63–68.
- Notes on Longley-Rice Propagation [Электронный ресурс]. – Режим доступа: http://www.softwright.com/faq/engineering/prop_longley_rice.html, свободный. Яз. англ. (дата обращения 25.08.2012).
- Рекомендация МСЭ-R BT.1368-8. Критерии планирования для услуг наземного цифрового телевидения в диапазонах ОВЧ/УВЧ.

8. Dalke R.A., Holloway C.L., McKenna P., Johansson M., Ali A.S. Effects of reinforced concrete structures on RF communications // IEEE Trans. Electromagn. Compat. – 2000. – V. 42. – № 4. – P. 486–496.
9. Kakar F., Sani K.A., Elahi F. Essential factors influencing building penetration loss // 11th IEEE International Conference on Communication Technology (ICCT). – Hangzhou, China, 2008. – 10–12 November. – P. 1–4.
10. Введенский Б.А. Распространение ультракоротких радиоволн. – М.: Наука, 1973. – 408 с.
11. Песков С.Н. Аналитические методы расчета напряженности поля, создаваемой передатчиком // Теле-Спутник, 2008. – № 10. – С. 94–97.

Ломакин Александр Федорович – Дальневосточный федеральный университет, кандидат географических наук, доцент, lomakin11@land.ru
Школьный Станислав Игоревич – Дальневосточный федеральный университет, аспирант, nimagt@gmail.com

УДК 553.087

ПРИМЕНЕНИЕ ГЛУБИННЫХ ТРЕХКООРДИНАТНЫХ ДАТЧИКОВ НАКЛОНА ПРИ РЕШЕНИИ ЗАДАЧ ГЕОТЕХНИЧЕСКОГО МОНИТОРИНГА

Т.В. Падерина, В.В. Яковлев, Ю.М. Гордеев

Анализируются результаты применения первой отечественной измерительной системы с глубинными трехкоординатными датчиками наклона на базе микромеханических акселерометров при решении задачи геотехнического мониторинга горных участков олимпийской трассы Адлер – Красная поляна. Показано, что рассматриваемая система позволяет реализовать оперативный контроль горизонтальных смещений массивов грунта в течение длительного времени, обеспечивая тем самым своевременное предупреждение возможных неблагоприятных ситуаций в процессе строительства и эксплуатации объекта. Достигнутый уровень точности, составляющий несколько миллиметров, и относительно невысокая стоимость делают данную систему привлекательной при организации непрерывного контроля особо протяженных геологических объектов.

Ключевые слова: геотехнический мониторинг, микромеханический акселерометр, трехкоординатный датчик наклона.

Введение

Геологическое пространство (горные породы и грунты, которые служат основанием построек, строительных объектов) не является застывшей формацией. Происходящие в нем процессы – изменение уровней подземных вод, изменение химического состава последних, развитие карста – и целый ряд других факторов кардинально меняют инженерно-геологические условия в недрах, требуя проведения их постоянного контроля и изучения.

Геотехнический мониторинг – это система наблюдений и контроля за состоянием и изменением грунтовых, природных и техногенных условий в процессе строительства и эксплуатации объекта.

Арсенал технических средств геотехнического мониторинга в мире насчитывает несколько десятков наименований, среди которых – экстенсометры, датчики нагрузки, тензометрические датчики различного типа, измерители трещин и стыков, приборы для гидрогеологического контроля, оптические и спутниковые геодезические приборы, лазерные сканеры, георадары и т.д. Тем не менее, несмотря на такое разнообразие средств, постоянный мониторинг различных промышленных и геологических объектов – плотин, мостов, оползневых склонов, карьеров, территорий над горными выработками и т.п., как в России (вплоть до последнего времени), так и за рубежом (несколькими годами ранее) представлял собой достаточно трудную, а главное, чрезвычайно дорогостоящую задачу.

Ситуация кардинально изменилась с внедрением MEMS-технологий (Microelectromechanical systems), а точнее, с появлением на рынке недорогих микромеханических акселерометров (ММА) и созданием на их базе новой системы геотехнического контроля, представляющей собой сборку независимых триад ММА, установленных в гибкой водонепроницаемой трубе с целью измерения трехмерной деформации грунта на глубину до 100 м с интервалом 0,5–2,0 м. Такие системы, названные Shape Acceleration Array (SAA), впервые были применены компаниями «Geokon» (США), «Soil Instruments Ltd.» (Англия), «Measurand Inc.» (Канада) примерно в 2007–2009 г.г. [1].

Настоящая работа посвящена анализу результатов применения первых отечественных систем, подобных SAA, разработанных российской компанией ЗАО «Специальное конструкторское бюро приборов подземной навигации» («СКБ ПН») (г. Санкт-Петербург) для проведения геотехнического мониторинга горных участков олимпийской трассы Адлер – Красная поляна.

Контрольные скважины с глубинными трехкоординатными датчиками наклона

В свете современных подходов к вопросам безопасности сооружений и новой концепции строительных технологий система мониторинга должна разрабатываться уже на стадии проектирования объ-

екта, и создание такой системы целесообразно начинать с исследования современного состояния горных пород, оценки инженерно-геологических особенностей подземного пространства, возможных геологических, техногенных, технических причин зарождения деформации.

Для контроля геологического пространства строительного объекта – тоннельного комплекса, расположенного на горных участках олимпийской трассы Адлер – Красная поляна (рис. 1), наиболее эффективным и экономичным признан подход, основанный на периодической фиксации относительных смещений пород, формирующих эти склоны.

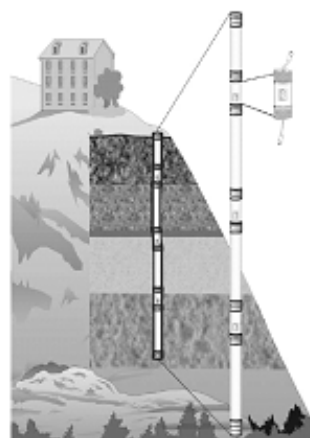


Рис. 1. Район северных порталов тоннельного комплекса № 3, июнь 2011

С этой целью используется система вертикально разбуренных контрольных скважин (рис. 2, а), обсаженных пластиковыми трубами, в которые погружаются гибкие «косы» с трехкоординатными датчиками наклона (ТДН) на базе ММА (рис. 2, б). При смещении пласта горных пород в направлении, перпендикулярном оси ствола, происходит деформация участков контрольной скважины и, как следствие, соответствующих секций косы с датчиками. На рис. 3, а, наглядно показана деформация (изгиб) участка косы из-за смещения грунта. Данный изгиб, характеризующийся углом θ , можно определить по показаниям акселерометров, установленных в косе. В свою очередь, угол θ и известное значение длины L между датчиками позволяют вычислить параметр d , который и является искомым значением горизонтального смещения грунта на данной глубине.



а



б

Рис. 2. Бурение вертикальной скважины в районе строящегося тоннеля (а); пример косы, установленной в вертикальную скважину (б)

Принципиальным условием применения данного метода является необходимость установки нижнего участка косы в массив грунта, который не подвергается деформации. Именно благодаря наличию такого «опорного» положения появляется возможность измерить все последующие смещения вышележащих пластов.

Предлагаемый метод обладает рядом существенных преимуществ по сравнению с другими, ранее применявшимися при геотехническом мониторинге. К основным его достоинствам можно отнести следующие:

- исключительно протяженная зона непрерывного контроля;
- возможность повторного применения косы даже при ее значительной деформации;
- малый диаметр косы и, в силу этого, возможность ее установки практически в любом месте контролируемого геологического пространства без нарушения его структурной целостности.

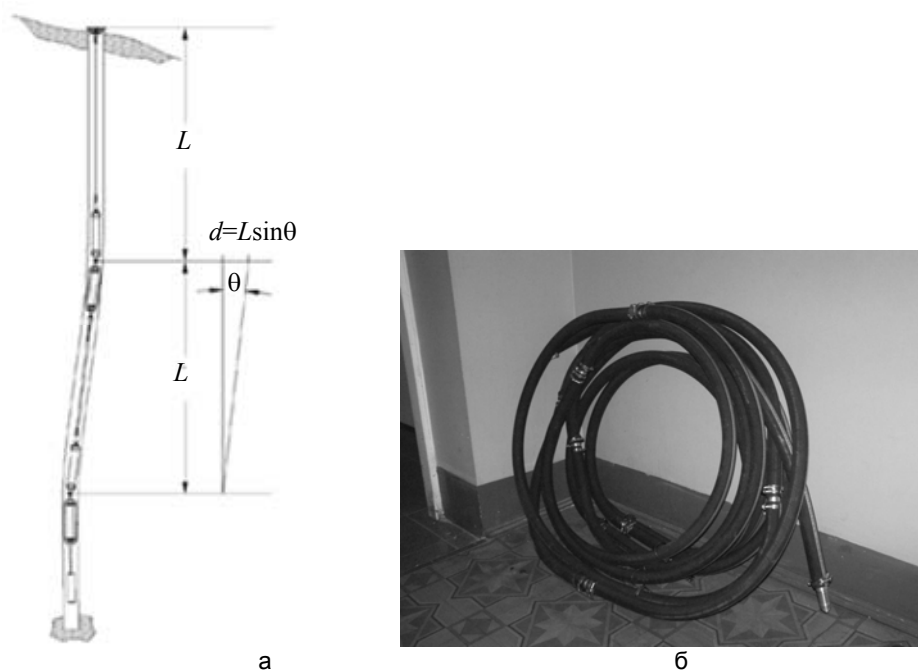


Рис. 3. Коса с ТДН: пример деформации косы, обусловленной смещением грунта (а); фрагмент косы (20м×10 ТДН), разработанной ЗАО «СКБ ПН» (б)

Специалистами компании ЗАО «СКБ ПН», в том числе и авторами данной работы, была разработана система мониторинга с использованием ТДН на базе акселерометров ADXL103 (Analog Devices). Каждый ТДН вместе с контроллером схемы питания и интерфейса, а также датчиком температуры, располагается на индивидуальном жестком шасси длиной 25 см и диаметром 27 мм. ТДН вместе с присоединенным к нему кабелем помещается в гибкую трубу-проставку, длина которой в зависимости от решаемой задачи может составлять 0,5; 1; 2 м. Объединение труб-проставок позволяет легко собрать косу нужной длины для погружения в скважину (рис. 3, б). Материал трубы-проставки выбирается специальным образом, чтобы коса могла работать только на изгиб (без скручивания). Все ТДН соединяются общим кабелем, по которому подается питание и производится считывание информации. В состав системы мониторинга также входит персональный компьютер, устройство сопряжения для подключения косы к компьютеру, программное обеспечение, осуществляющее сбор, обработку и визуальную интерпретацию поступающих данных.

Все ТДН проходят калибровку на стенде, в процессе которой определяются их масштабные коэффициенты, погрешности смещения нуля, «геометрические» параметры, коэффициенты температурного влияния. Полученный массив коэффициентов модели погрешностей каждого ТДН записывается в память соответствующего датчика и используется для коррекции его выходных показаний.

Разработке данной системы предшествовало проведение долговременных стендовых испытаний стабильности показаний применяемых акселерометров при изменении температуры окружающей среды в заданном диапазоне. Эти испытания подтвердили высокую надежность ТДН и возможность их работы с требуемыми характеристиками (чувствительность на уровне 1 угл. мин) в течение длительного времени (не менее года).

Результаты проведения геотехнического мониторинга горных склонов в ходе строительства трассы Адлер – Красная поляна

Ниже представлены результаты опытной эксплуатации рассматриваемой системы геомониторинга при строительстве горной трассы Адлер – Красная поляна в 2011 г.

Один из строительных объектов северного портала тоннельного комплекса расположен в зоне, где возможен сход оползня. В связи с этим для его мониторинга на горных склонах, непосредственно прилегающих к строительству, были установлены несколько контрольных скважин с глубинными ТДН.

Длина контрольного ствола всегда выбирается исходя из условий мониторинга конкретного геологического пространства. Для рассматриваемого строительного объекта протяженность участка горных пород, подверженных возможному смещению, составляет около 32 м, глубже располагаются более прочные грунты (мергель). Поскольку при установке контрольного ствола, как уже отмечалось выше, обязательным условием является закрепление его нижней части в прочных породах, было выбрано значение общей длины скважинного комплекса (контрольного ствола) – 38 м, при этом шаг установки ТДН L составлял 2 м.

В качестве иллюстраций на рис. 4, 5 показаны результаты измерений горизонтальных смещений массивов грунта в скважине № 1, мониторинг которых проводился в следующих числах 2011 г. – 02.09; 21.09; 05.10; 31.10; 09.11; 09.12.

Искомые смещения определялись с помощью вычисления следующих разностей:

1. $S_E^{21.09} - S_E^{02.09}$, $S_N^{21.09} - S_N^{02.09}$;
2. $S_E^{05.10} - S_E^{02.09}$, $S_N^{05.10} - S_N^{02.09}$;
3. $S_E^{31.10} - S_E^{02.09}$, $S_N^{31.10} - S_N^{02.09}$;
4. $S_E^{09.11} - S_E^{02.09}$, $S_N^{09.11} - S_N^{02.09}$;
5. $S_E^{09.12} - S_E^{02.09}$, $S_N^{09.12} - S_N^{02.09}$,

где S_E ; S_N – значения восточной и северной координаты контрольной скважины, рассчитанные в соответствующий день 2011 г. (верхний индекс) [2]. Полученные таким образом горизонтальные смещения массивов грунта отображены на рис. 4, 5. Анализ результатов мониторинга показал, что общий устойчивый сдвиг массивов грунта в скважине № 1 наблюдался до глубины 27 м, при этом наибольшее смещение, как и ожидалось, претерпели пласты, расположенные вблизи поверхности горного склона. Устье скважины № 1 за весь интервал наблюдения сдвинулось в направлении востока на 172 мм (рис. 4), в направлении севера (рис. 5) – на 210 мм.

Сравнение значений горизонтальных смещений устьев контрольных скважин, рассчитанных приведенным выше способом, с аналогичными параметрами, полученными с помощью используемой на территории строительства трассы Адлер – Красная поляна системы GPS-мониторинга, показало их хорошую корреляцию (невязка показаний составляет около 1%).

Надо отметить, что в последние годы метод спутниковой геодезии все чаще стал применяться при строительстве тоннелей, сооружении мостов, прокладке магистральных трубопроводов и т.д. Специфика этого метода для решения геодезических задач высокоточного измерения координат наземных объектов состоит в использовании относительных измерений, при которых дальности до спутников определяются фазовым методом и по ним вычисляются приращения координат между станциями с установленными спутниковыми приемниками [3].

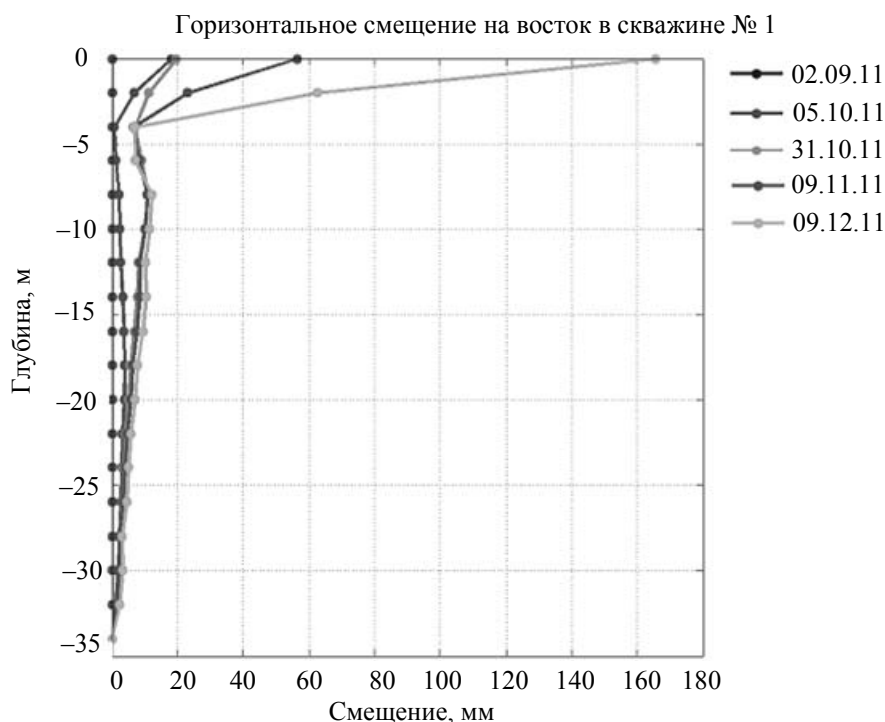


Рис. 4. Смещение массивов грунта по линии запад-восток в скважине №1

Многочисленный опыт работ с использованием спутниковых методов показывает, что при благоприятном расположении спутников и при исключении многолучевого распространения сигнала возможно достижение точности измерения координат на уровне нескольких миллиметров ($\pm(3-5)$ мм) [4]. В связи с этим с большой уверенностью можно считать, что точность представленной в данной работе системы геомониторинга на базе ТДН находится в этих же пределах.

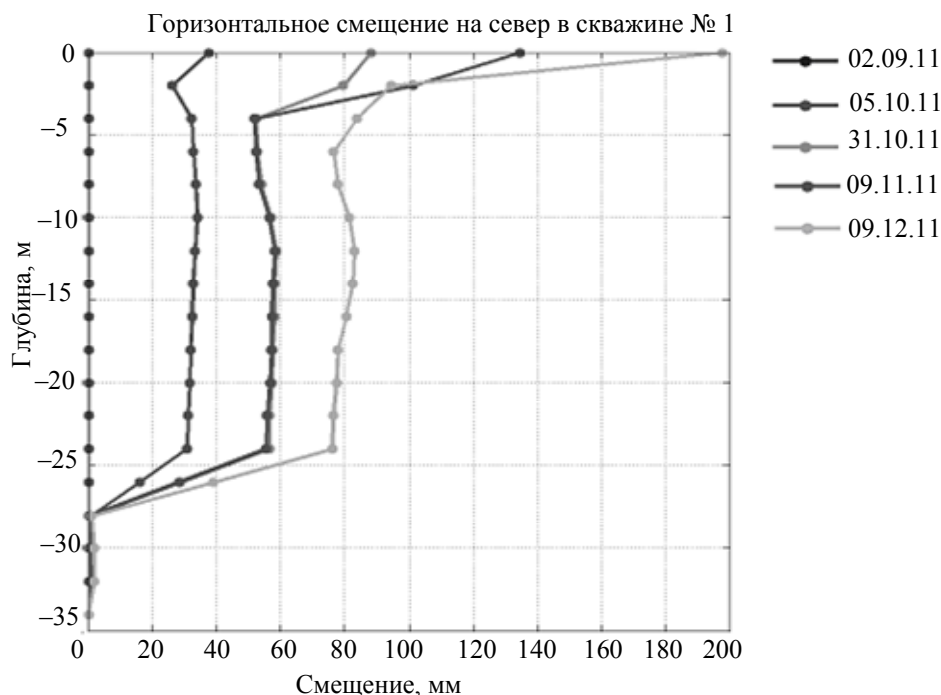


Рис. 5. Смещение массивов грунта по линии север–юг в скважине №1

Заключение

Результаты применения первой отечественной измерительной системы с глубинными трехкоординатными датчиками наклона при решении задачи геотехнического мониторинга горных участков олимпийской трассы Адлер – Красная поляна подтвердили ее высокую надежность и способность обеспечить контроль горизонтальных смещений и деформаций массивов грунта с высокой точностью в течение длительного времени. Данная система позволяет реализовать практически круглосуточные геотехнические измерения и наблюдения, без которых невозможно своевременное предупреждение различных неблагоприятных ситуаций в процессе строительства и эксплуатации промышленных объектов.

Литература

1. Bennett V., Zeghal M., Abdoun T. Wireless Shape-Acceleration Array System for Local Identification of Soil and Soil Structure Systems // Transportation Research Record: Journal of the Transportation Research Board. – 2007. – V. 12. – P. 60–66.
2. Биндер Я.И., Падерина Т.В., Лысенко А.С., Федорович А.Н. Об использовании различных схем гироинклинометров для непрерывной съемки скважин произвольной ориентации. Проблемы и решения // Гироскопия и навигация. – 2010. – № 4. – С. 92–110.
3. Трехо Сото Мануэль. Математический анализ спутниковых геодезических сетей при изучении деформаций инженерных сооружений // Геодезия и аэрофотосъемка. – 2007. – № 1. – С. 67–75.
4. Трехо Сото Мануэль. Оценка точности топоцентрических прямоугольных координат при изучении деформаций крупных инженерных сооружений спутниковыми методами // Изв. вузов. Геодезия и аэрофотосъемка. – 2006. – № 6. – С. 75–86.

Падерина Татьяна Владимировна – ОАО « Концерн «ЦНИИ «Электроприбор», кандидат технических наук, ст. научный сотрудник, paderinata@rambler.ru

Яковлев Владимир Викторович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, v-v-y@ya.ru

Гордеев Юрий Михайлович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, 133880@gmail.com

УДК 004.4'2

**ПРИМЕНЕНИЕ ЗАВИСИМЫХ СИСТЕМ ТИПОВ СО СТРУКТУРНОЙ
ИНДУКЦИЕЙ ДЛЯ ВЕРИФИКАЦИИ РЕАКТИВНЫХ ПРОГРАММ**

Я.М. Малаховски, Г.А. Корнеев

Предлагается конструктивное доказательство (в виде программы на языке Agda) того, что решение задачи о выполнимости булевой формулы при помощи исчисления секвенций и верификация с использованием темпоральных спецификаций на языке CTL могут быть представлены в разрешимом подмножестве зависимой системы типов.

Ключевые слова: задача о выполнимости булевой формулы, верификация, функциональное программирование, системы типов, зависимые типы, Agda.

Введение

Одной из современных тенденций при разработке программного обеспечения является так называемое доказательство по построению (proof by construction), когда нежелательные состояния в программе (например, некорректные состояния ресурсов) исключаются не при помощи проверок, осуществляемых внешними средствами, а при помощи свойств, которыми обладают используемые языки программирования. Одним из самых известных таких методов является RAII [1]. Также все большую популярность набирают средства аннотирования программ, такие как, например, nullness annotations, позволяющие не отделять спецификацию программы от ее кода. Это позволяет избавиться от целого класса ошибок, связанных с рассинхронизацией программы и ее спецификации в процессе разработки.

В функциональном программировании [2, 3] для решения задач доказательства программ по построению и их аннотирования применяют системы типов. В соответствии с изоморфизмом Кэрри–Ховарда различным системам типов функциональных языков программирования соответствуют различные логические системы [4]. Например, система типов простого типизированного лямбда-исчисления изоморфна минимальной логике высказываний, простые зависимые типы изоморфны логике первого порядка, а зависимые типы высших порядков изоморфны логике высших порядков [4–6]. При этом в таких системах тип терма (типовая аннотация терма) является утверждением, а сам терм – доказательством этого утверждения. Отмеченное свойство позволяет использовать некоторые функциональные языки программирования (имеющие в качестве системы типов зависимую или более мощную) для проверки правильности некоторых классов математических доказательств (утверждения, для доказательства которых достаточно интуиционистской логики), а также для доказательства практически произвольных утверждений о самих программах. Однако большинство распространенных функциональных языков программирования (например, Haskell [7]) не обладают такими возможностями. Это связано с тем, что эти языки сохраняют возможность автоматической реконструкции типовых аннотаций в программе (автоматический вывод типов всех термов), а потому их системы типов не являются Тьюринг-полными [6]. В свою очередь, использование Тьюринг-полной зависимой системы типов делает неразрешимой задачу проверки эквивалентности типов и может привести к бесконечно долгой компиляции программы [4, 5].

Интересным подходом к решению данной проблемы обладают языки Agda [8] и Coq [9]. Все программы на этих языках завершаются благодаря возможности использования в программе только структурной индукции (но не неограниченной рекурсии), а система типов совпадает по вычислительной мощности с остальной частью языка. Оказывается, что на этих языках можно выражать и проверять многие полезные свойства программ, не теряя при этом свойств разрешимости.

Другими словами, возвращаясь к задачам доказательства программ по построению и их аннотирования, можно утверждать, что в некоторых функциональных языках программирования аннотирование термов программы практически произвольными утверждениями (типами) и написание программ, корректных по построению (написание термов, имеющих необходимые типы), являются встроенными возможностями.

Таким образом, свойства программ (например, что программа удовлетворяет некоторой темпоральной спецификации [10]), обычно проверяемые при помощи сторонних утилит (например, построением модели программы и ее верификации при помощи средств проверки моделей), в языках Agda и Coq могут быть проверены средствами самих языков, если их удастся закодировать в виде утверждений в системах типов этих языков. Иначе говоря, требуемые свойства программ требуется переписать (формализовать) в утверждения, разрешимые в зависимой системе типов со структурной индукцией. Это не всегда является тривиальной задачей, поскольку, кроме кодирования самого утверждения, требуется привести и доказательство этого утверждения (также закодированное в виде терма программы).

Целью настоящей работы является конструктивное доказательство (в виде программы на языке Agda) того, что решение задачи о выполнимости булевой формулы при помощи исчисления секвенций и верификация с использованием темпоральных спецификаций на языке CTL [10] могут быть представлены в разрешимом подмножестве зависимой системы типов.

Результаты данной работы были использованы [11] при создании встроенных предметно-ориентированных языков (embedded Domain Specific Language, eDSL), предназначенных для описания реактивных программ с использованием автоматного программирования [12].

Все используемые свойства языка Agda, выходящие за рамки простого типизированного лямбда-исчисления, явно обозначены. Все функции из листингов, определения которых не приводятся, а отсутствие определения никак не комментируется, являются функциями стандартной библиотеки.

Особенности используемого языка программирования

Язык программирования Agda является чистым ленивым функциональным языком программирования со структурной индукцией [13]. Однако в качестве системы типов используется разрешимое подмножество зависимой системы типов. При этом и сам язык, и система типов совпадают по вычислительной мощности и не являются Тьюринг-полными. Классы типов [7] в языке отсутствуют, однако их можно эмулировать при помощи встроенных средств. Язык частично сохраняет возможность реконструкции типовых аннотаций (только тривиальные аннотации для термов non-ground типов).

Структурно-рекурсивное решение задачи о выполнимости булевой формулы

Классический метод поиска контрпримера к логическому утверждению, рассматриваемый в работе [14], не может быть непосредственно закодирован на языке Agda, поскольку перенос формулы из одного множества в другое не позволяет построить из последовательности получаемых секвенций явную убывающую последовательность. Для решения данной проблемы следует «транспонировать» формулировку алгоритма поиска решения.

Поскольку валидация при помощи исчисления секвенций использует базис И-ИЛИ-НЕ, то необходимо явно его определить:

```
data Logic (V : Set) : Set where
  Latom : V -> Logic V           -- Вхождение переменной.
  Ltrue Lfalse : Logic V         -- Тожественно истинное и
                                -- ложное утверждения.
  Lnot : Logic V -> Logic V      -- Отрицание.
  _L^_ _LV_ : Logic V -> Logic V -- Конъюнкция и
                                -- дизъюнкция.
```

Приведем пример применения предложенного логического базиса:

```
data Var : Set where           -- Логические
  X1 X2 X3 : Var              -- переменные

example : Logic Var
example = ((Latom X1) L^ (Latom X2)) LV (Lnot (Latom X3))
```

Интерпретатор таких формул (назовем его `compute`), принимающий в качестве входных параметров формулу и функцию из имени логической переменной в ее значение (типа `Bool`) и возвращающий результат вычисления формулы, реализуется очевидным образом при помощи сопоставления с образцом (pattern matching, оператор `case`).

«Транспонирование» классического алгоритма заключается в замене в нем двух множеств секвенции на список пар вида (сторона, формула), где сторона задается элементом типа `Bool` (`true` – левая, `false` – правая). При этом множество выполняющих наборов для формулы будет списком списков таких пар.

После этого, используя то наблюдение, что множество выполняющих наборов является булевой алгеброй, введем на нем стандартные логические операции. При этом пустой список будет являться нулевым элементом, а список, содержащий пустой список – нейтральным элементом:

```
Seq1 : Set -> Set
Seq1 V = List (List V) -- Множество выполняющих наборов

zero : ∀ {V} -> Seq1 V
zero = []

one : ∀ {V} -> Seq1 V
one = [ [] ]
```



```

single : ∀ {V} -> V -> Seq1 V
single v = [ [ v ] ]

_sor_ : ∀ {V} -> Seq1 V -> Seq1 V -> Seq1 V
x sor y = x ++ y

_sand_ : ∀ {V} -> Seq1 V -> Seq1 V -> Seq1 V
x sand y = concatMap (λ xel -> map (λ yel -> (xel ++ yel)) y) x

```

После этого функция (назовем ее *solve*), принимающая имя столбца (напомним, обозначаемое элементом типа *Bool*) и логическую формулу и возвращающая список ее выполняющих наборов, реализуется следующим образом:

```

solve : ∀ {V} -> Bool -> Logic V -> Seq1 (Bool × V)
solve any (Latom x) = single (any , x)
solve true  Ltrue = one
solve false Ltrue = zero
solve true  Lfalse = zero
solve false Lfalse = one
solve any (Lnot x) = solve (not any) x
solve true  (x L∧ y) = (solve true x) sand (solve true y)
solve false (x L∧ y) = (solve false x) sor (solve false y)
solve true  (x L∨ y) = (solve true x) sor (solve true y)
solve false (x L∨ y) = (solve false x) sand (solve false y)

```

Нетрудно видеть, что в этом терме все рекурсивные вызовы производятся с меньшими аргументами, что позволяет данному коду успешно пройти тест на завершаемость, выполняемый компилятором языка Agda. Само же исчисление секвенций реализуется фильтрацией из списка выполняющих наборов невыполнимых (пустых или правый и левый столбцы которых пересекаются).

Подробная реализация всех упомянутых функций приводится в работе [11].

Грамматика языка CTL и представление верифицируемой программы

Аналогичным логическому базису образом определяется и синтаксис выражений языка CTL:

```

data CTL (A : Set) : Set where
  |atom : A -> CTL A
  |true |false : CTL A
  |not : CTL A -> CTL A
  |_|∧_|∨_| : CTL A -> CTL A -> CTL A
  |X∧ X∃ : CTL A -> CTL A
  |_U∨_|U∃_| : CTL A -> CTL A -> CTL A

```

Для реализации CTL-верификатора была выбрана функциональная версия классического алгоритма [15], представленная в работе [16]. Основная идея данной интерпретации заключается в том, что вместо вычисления подформулы с использованием динамического программирования можно непосредственно вычислять верифицируемую формулу, кешируя (мемоизируя) промежуточные результаты. При этом появляется возможность отказаться от поиска компонент сильной связности, поскольку каждый путь в модели Крипке представляет собой либо некоторую последовательность состояний без циклов, либо некоторую (возможно, пустую) последовательность состояний без циклов, завершающуюся циклом. В такой формулировке формула языка CTL, вычисление которой закикливается (т.е. ее выполнимость требует ее же выполнимости), считается невыполнимой.

Передать знание о структуре путей в модели Крипке компилятору не представляется возможным. Однако можно заметить, что путь максимальной длины для вычисления CTL-формулы в модели Крипке – гамильтонов цикл. В связи с этим для доказательства завершаемости алгоритма предлагается добавить размер модели Крипке в качестве параметра верификатора CTL-формулы и уменьшать это значение при рекурсивном вычислении той же подформулы в новом состоянии. Реализация верификатора в настоящей работе не приводится вследствие ее большого объема и необходимости рассмотреть несколько нетривиальных приемов, за подробностями направляем читателя к работе [11].

Исчисление секвенций и верификация в системе типов

Подчеркнем, что выше были описаны функции, проверяющие выполнимость булевой формулы и соответствие заданной модели Крипке некоторой темпоральной спецификации на языке программирования CTL, тотальность (totality) и завершаемость которых проходит проверку компилятором языка Agda. Поскольку в языке Agda система типов не отличается по вычислительной мощности от остальной части

языка, то для представления свойств валидации и верификации в системе типов необходимо построить соответствующие предикаты, использующие функции, разрешающие проверяемые свойства. В системе типов предикатом над данными является функция из значения какого-то типа в другой тип. При этом типом-результатом в общем случае может являться произвольный тип, описывающий свойства рассматриваемых данных.

При разработке автоматного eDSL [11] свойства, описывающие валидность и соответствие автомата некоторой темпоральной спецификации, являются внешними по отношению к функциональной программе, на которой их выражают. По этой причине информацию о них можно только стереть из типов. Таким образом, предикаты для валидации автомата (проверки полноты и непротиворечивости исходящих из состояния дуг, над которыми надписаны логические выражения) и верификации в системе типов принимают следующий вид:

```
-- S = state, V = variable, O = output action
-- Последний аргумент - список дуг переходов состояния автомата
isValid :  $\forall \{S\ V\ O\} \rightarrow \text{List } (\text{Logic } V \times (S \times \text{List } O)) \rightarrow \text{Set}$ 
isValid g = if valid g then Top else Bot

-- Последний аргумент - описание автомата
isVerified :  $\forall \{V\} \rightarrow \text{CTL } V \rightarrow \text{Auto} \rightarrow \text{Set}$ 
isVerified f a = if verified f a then Top else Bot
```

В приведенном листинге Top – тип с одним элементом, а Bot – тип без элементов. Другими словами, функции isValid и isVerified проверяют соответствующие свойства и стирают из типа всю информацию о проверке, за исключением ее результата. Кроме того, использование этих типов позволяет компилятору автоматически выводить значения аргументов для них, что позволяет сделать аргументы этого типа неявными аргументами функции.

После этого написание функций, проверяющих требуемые свойства автоматов на этапе компиляции, заключается в простом добавлении к ним неявного аргумента вида

```
{v : isValid l},
```

что является стандартным приемом.

Заключение

В работе показано, что решение задачи о выполнимости булевой формулы и проверка соответствия модели Крипке заданной темпоральной спецификации на языке CTL представимы в разрешимом подмножестве зависимой системы типов. Данный факт был использован в работе [11] для создания системы, поддерживающей программирование в автоматном стиле [12] на функциональных языках программирования. Описанные функции развивают методы, предложенные в работах [17, 18], но, в отличие от последних, позволяют производить валидацию и верификацию автоматов на этапе компиляции.

Таким образом, разработанные в работе термы не только являются важным теоретическим результатом в области границ применимости структурно-рекурсивных функций, но и позволяют производить проверку соответствия программы темпоральной спецификации в инструменте, предназначенном для автоматного программирования. При этом такой инструмент не требует никаких внешних средств для своей работы, а его (частичная) корректность может быть доказана на самом языке Agda.

Литература

1. Страуструп Б. Дизайн и эволюция языка C++. – СПб: Питер, 2006. – 448 с.
2. Абельсон Х., Сассман Д. Структура и интерпретация компьютерных программ. – М.: Добросвет, 2006. – 608 с.
3. Bird R. Introduction to Functional Programming Using Haskell. – Prentice-Hall, 1998. – 448 p.
4. Sørensen M.H.B., Urzyczyn P. Lectures on the Curry-Howard Isomorphism. – Elsevier, 2006. – 442 p.
5. Thompson S. Type Theory and Functional Programming. – Addison-Wesley, 1991. – 365 p.
6. Пирс Б. Типы в языках программирования [Электронный ресурс]. – Режим доступа: <http://newstar.rinet.ru/~goga/tapl/>, свободный. Яз. рус. (дата обращения 29.02.12).
7. Худак П., Петерсон Дж., Фасел Дж. Мягкое введение в Хаскелл // RSDN Magazine. – 2006. – № 4. – С. 35–67; 2007. – № 1. – С. 3–16.
8. The Agda Wiki [Электронный ресурс]. – Режим доступа: <http://wiki.portal.chalmers.se/agda/>, свободный. Яз. англ. (дата обращения 29.02.12).
9. The Coq Proof Assistant [Электронный ресурс]. – Режим доступа: <http://coq.inria.fr/>, свободный. Яз. англ. (дата обращения 29.02.12).
10. Вельдер С.Э., Лукин М.А., Шалыто А.А., Яминов Б.Р. Верификация автоматных программ. – СПбГУ ИТМО, 2011. – 246 с.

11. Малаховски Я.М. Применение систем типов для валидации и верификации автоматных программ. Магистерская диссертация. НИУ ИТМО, 2011 [Электронный ресурс]. – Режим доступа: <http://is.ifmo.ru/papers/2011-master-malakhovski/>, свободный. Яз. рус. (дата обращения 07.03.2012).
12. Поликарпова Н.И., Шалыто А.А. Автоматное программирование. – СПб: Питер. 2009. – 176 с.
13. Norrel U. Dependently Typed Programming in Agda // *Advanced Functional Programming*. – V. 5832. – 2009. – P. 230–266.
14. Верещагин Н.К., Шень А. Лекции по математической логике и теории алгоритмов. Ч. 2. Языки и исчисления. – М.: МЦНМО, 2002. – 288 с.
15. Clarke E.M., Emerson E.A., Sistla A.P. Automatic verification of finite-state concurrent systems using temporal logic specifications // *ACM Transactions on Programming Languages and System*. – 1986. – № 8. – P. 244–263.
16. CTL Model Checking in Haskell: A Classic Algorithm Explained as Memoization [Электронный ресурс]. – Режим доступа: <http://www.kennknowles.com/blog/2008/05/07/ctl-model-checking-in-haskell-a-classic-algorithm-explained-as-memoization/>, свободный. Яз. англ. (дата обращения 29.02.12).
17. Малаховски Я.М., Шалыто А.А. Конечные автоматы в чистых функциональных языках программирования. Автоматы и Haskell // *RSDN Magazine*. – 2009. – № 3. – С. 20–26.
18. Малаховски Я.М., Корнеев Г.А. Валидация автоматов с переменными на функциональных языках программирования // *Научно-технический вестник СПбГУ ИТМО*. – 2010. – № 6 (70). – С. 73–77.

Малаховски Ян Михайлович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, trojan@rain.ifmo.ru

Корнеев Георгий Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, kgeorgiy@rain.ifmo.ru

УДК 621.391.037.372

АЛГОРИТМ ОЦЕНКИ ОТНОШЕНИЯ СИГНАЛ/ШУМ РЕЧЕВЫХ СИГНАЛОВ

М.Б. Столбов

Предложен алгоритм оценки интегрального значения отношения сигнал/шум и его значений в частотных полосах для определения качества фонограмм в системе верификации дикторов. Особенность разработанного алгоритма состоит в робастности по отношению к большой вариативности условий записи и качества фонограмм, а также возможности выполнения оценки в режиме реального времени, т.е. в темпе поступления речевого сигнала. В основу алгоритма положены новые способы оценки спектра шума и детектирования речи. Эксперименты показали достаточную для практических применений достоверность оценок отношения сигнал/шум в диапазоне от 6 до 26 дБ на записях длительностью от 10 с и более.

Ключевые слова: отношение сигнал/шум, детектирование речевых кадров, оценка спектра шума.

Введение

Предварительная оценка качества речевого материала является важной в задаче идентификации личностей по голосу (идентификации диктора). Одним из основных показателей, влияющих на качество идентификации диктора, является отношение уровня исходного речевого сигнала к уровню присутствующего в фонограмме шума. Данное отношение может меняться на протяжении фонограммы вследствие вариаций уровня фонового шума и параметров речи диктора. В качестве интегральной меры качества фонограммы целесообразно использовать среднее по фонограмме отношение сигнал/шум (ОСШ).

Средняя величина ОСШ может использоваться для оценки тестовых и обучающих данных, идентификации канала записи фонограммы, выбора рабочего диапазона частот и т.д. Для решения данных задач могут быть использованы оценки ОСШ в частотных полосах или интегральное по частотам значение ОСШ. Мерой качества речевого сигнала в фонограмме является средняя по всем речевым фрагментам фонограммы оценка ОСШ – так называемое сегментное ОСШ.

Закрепленные в стандартах алгоритмы оценки ОСШ предполагают, что известны как шум, так и полезный сигнал [1, 2]. Однако эти алгоритмы не могут быть использованы в ситуации, когда имеется единственная фонограмма с зашумленным речевым сигналом. В этом случае необходимо решить задачу слепой оценки ОСШ. Предложено несколько групп методов слепой оценки ОСШ (см., например, список литературы в работе [3]):

- распознавание участков речевого сигнала и шума с применением детектора речи, по которым вычисляются оценки спектра шума и речи;
- оценка гистограммы амплитуд огибающих спектра, из которой отдельно определяются распределения речи и шума;

- оценка текущих спектров шума (например, на основе отслеживания минимумов огибающих спектра сигнала);
- оценка статистических параметров распределений спектральных амплитуд (например, статистик высокого порядка).

Последние два направления получили к настоящему моменту наибольшее распространение. Однако задача слепой оценки ОСШ по-прежнему далека от своего решения [4]. Непосредственное использование алгоритмов сторонних разработчиков представляется затруднительным по ряду причин:

- разнообразие требований к алгоритмам в зависимости от области их практического применения;
- использование в алгоритмах параметров, зависящих от типа сигнала (частота дискретизации и пр.), которые в конечном итоге подбираются эмпирически;
- отсутствие общей методики сравнения различных алгоритмов (базы данных, критерии оценки и пр.).

Перечисленные обстоятельства обусловили разработку собственного алгоритма оценки ОСШ, удовлетворяющего следующим требованиям:

- робастность оценки ОСШ для шумов различных типов (в том числе нестационарных);
- работоспособность в интервале значений ОСШ (динамическом диапазоне) фонограмм от 6 до 24 дБ;
- возможность оценки ОСШ на коротких фонограммах (на общей длительности речи не менее 10 с);
- возможность вычисления оценки ОСШ в режиме реального времени;
- устойчивость работы алгоритма к различным типам помех;
- инвариантность к нормировке и частоте дискретизации речевого сигнала;
- низкие вычислительные затраты.

Цель работы – описание практической реализации алгоритма слепой оценки ОСШ, пригодного для работы в широком диапазоне условий (по типам шумов, каналам записи, изменчивости акустической обстановки и пр.).

Алгоритм оценки ОСШ

В основу алгоритма оценки ОСШ положен метод, основанный на вычислении оценок текущего спектра шума. Пусть обрабатываемый сигнал $x(i)$ представляет собой сумму речевого сигнала $s(i)$ и шума $n(i)$:

$$x(i) = s(i) + n(i).$$

При этом принимается, что речевой сигнал и шум статистически независимы. Во многих практических случаях данное условие выполняется. Тогда теоретические кратковременные спектры мощности можно записать как

$$Px(k, m) = Ps(k, m) + Pn(k, m),$$

где k и m – индексы частоты и кадра соответственно; $Px(k, m)$, $Ps(k, m)$, $Pn(k, m)$ – спектры мощности зашумленного сигнала, речи и шума соответственно. Тогда ОСШ на кадре данных в частотных полосах запишется как:

$$SNR(k, m) = Ps(k, m)/Pn(k, m) = (Px(k, m) - Pn(k, m))/Pn(k, m) = Px(k, m)/Pn(k, m) - 1.$$

Интегральное по частоте значение ОСШ на кадре данных выражается следующей формулой:

$$SNR(m) = Ps(m)/Pn(m),$$

где $Ps(m)$, $Pn(m)$ – мощности речи и шума на кадре m .

Интегральной характеристикой качества фонограммы в целом является среднее по всему файлу отношение сигнал/шум. Более представительной характеристикой качества речевого сигнала является сегментное ОСШ, вычисляемое как среднее значение поккадровых оценок ОСШ на речевых сегментах сигнала:

$$SSNR(k) [\text{дБ}] = 10 \log_{10} \langle SNR(k, m) \rangle_{SP},$$

$$SSNR [\text{дБ}] = 10 \log_{10} \langle SNR(m) \rangle_{SP},$$

где $\langle \rangle_{SP}$ обозначает усреднение по кадрам речи.

На практике теоретические значения спектров мощности сигнала $Px(k, m)$ и шума $Pn(k, m)$ неизвестны, поэтому используются их оценки. Оценка ОСШ в частотных полосах выражается следующим образом:

$$SNR(k, m) = \max \{ \delta, |Y(k, m)|^2 / \tilde{N}(k, m)^2 - 1 \},$$

где $\tilde{N}(k, m)^2$ – оценки спектральной плотности мощности шума (СПМ) на кадре с индексом m ; δ – минимальное значение оценки ОСШ; $|Y(k, m)|^2$ – сглаженная по времени оценка СПМ сигнала на кадре с индексом m :

$$|Y(k, m)| = \alpha |Y(k, m-1)| + (1 - \alpha) |X(k, m)|,$$

где $X(k, m)$ – кратковременный спектр сигнала $x(i)$; α – коэффициент забывания (от 0,75 до 0,8).

Таким образом, центральными элементами оценки ОСШ являются алгоритм оценки текущего спектра шума и детектор речевого сигнала. Рассмотрим вкратце предложенные нами алгоритмы.

В основе базового алгоритма оценки спектра шума лежит итеративный алгоритм оценки амплитудного спектра, построенный на идее управляемого порога, вычисляемого по отношению спектральных амплитуд зашумленного сигнала и шума [5]. Недостаток базового алгоритма состоит в том, что при появлении резких всплесков энергии сигнала он останавливает подстройку спектра шума. На практике такие ситуации (например, увеличение коэффициента усиления записывающего устройства и др.) могут привести к полной остановке обновления оценки спектра шума.

Вместо управляемого порога предлагается применить управляемые коэффициенты сглаживания. В этом случае оценка амплитуды спектра шума осуществляется рекурсивно, кадр за кадром, без поиска пауз речи, с использованием экспоненциального сглаживания с коэффициентами $\beta(k, m)$, управляемыми в каждой спектральной полосе индивидуально:

$$\tilde{N}(k, m) = \beta(k, m) \tilde{N}(k, m-1) + (1 - \beta(k, m)) |Y(k, m)|.$$

Коэффициенты сглаживания меняются в зависимости от отношений $\tilde{N}(k, m-1)/|Y(k, m)|$. Управление коэффициентами учитывает тот факт, что основная часть спектральных амплитуд $|Y(k, m)|$ на шуме распределена в интервале $0,5N(k, m) - 2N(k, m)$. При выходе амплитуд $|Y(k, m)|$ за верхнюю границу этого интервала величина коэффициента $\beta(k, m)$ уменьшается, а в случае выхода за нижнюю границу увеличивается. Алгоритм является эффективным в вычислительном отношении и позволяет получать оценки ОСШ в режиме реального времени.

Предложенный алгоритм оценки спектра шума продемонстрировал работоспособность на шумах различных типов (стационарных и нестационарных) и в условиях различных видов помех. В качестве критерия точности алгоритма оценки ОСШ использована интегральная по частоте относительная ошибка [6],

$$\text{LogErr}(m) = \langle 20 \log_{10} |N(k, m)| / \tilde{N}(k, m) \rangle,$$

где $\langle \rangle$ обозначает усреднение по кадрам сигнала.

На рис. 1 приведен пример зависимости относительной ошибки оценки ОСШ от времени для предложенного алгоритма для речи, зашумленной реальным шумом кондиционера с отношениями сигнал/шум 34 дБ (кривая 1), 16 дБ (кривая 2), 4 дБ (кривая 3), и шума без речи (кривая 4).

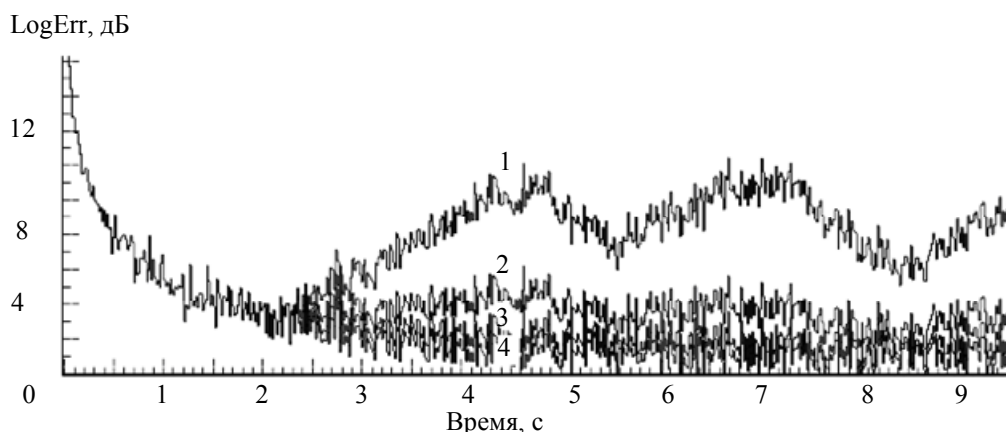


Рис. 1. Зависимость логарифмических ошибок от времени для речевого сигнала с разными значениями ОСШ

Из рис. 1 видно, что на начальном участке настройки оценки спектра шума ошибка оценки уменьшается. На участке, содержащем речь (начиная со 2-й секунды), ошибки оценки возрастают по мере увеличения величины ОСШ сигнала.

Сопоставление уровней ошибки оценки ОСШ для предложенного алгоритма и других известных алгоритмов [6] показывает, что предложенный алгоритм дает меньшие ошибки, вплоть до ОСШ, равного +28 дБ. Аналогичные результаты были получены при исследовании зависимости ошибки от используемой полосы частот.

Оценка интегрального по частоте значения ОСШ на кадре рассчитывается как отношение мощностей речевого сигнала и шума в полосе частот 300–3300 Гц:

$$\text{SNR}(m) = \max \left\{ \delta, \frac{\sum_{k=Kb}^{k=Ke} |Y(k, m)|^2}{\sum_{k=Kb}^{k=Ke} \tilde{N}(k, m)^2} - 1 \right\},$$

где Kb, Ke – значения индексов частоты, соответствующие 300 Гц и 3300 Гц; δ – минимальное значение оценки ОСШ.

Вторым важным элементом алгоритма оценки сегментного ОСШ является детектирование кадров с речью. Для детектирования речевых кадров была применена интегральная мера

$$T(m) = \frac{1}{(Ke - Kb) \sum_{k=Kb}^{k=Ke} \frac{|Y(k, m)|^2}{\tilde{N}(k, m)^2}},$$

где Kb , Ke – значения индексов частоты, соответствующие 300 Гц и 3300 Гц.

В случаях, когда значение $T(m)$ на кадре больше 1,3, кадр классифицируется как речевой. Отличие предложенного детектора от традиционных энергетических детекторов заключается в его устойчивости к наиболее распространенным тональным помехам. Действительно, в случае большого значения отдельной тональной компоненты спектра отношение $Y(k, m)/\tilde{N}(k, m)$ будет близким к единице и не внесет значительного вклада в величину $T(m)$. Величины всех порогов и сглаживающих констант были определены экспериментальным путем.

Эксперименты

Целью экспериментов было исследование диапазона работоспособности реализованного алгоритма оценки ОСШ. Алгоритм исследовался на фонограммах с известными значениями ОСШ.

В качестве иллюстрации на рис. 2 показаны зависимости оценок ОСШ сигнала в частотных полосах для зашумленной речи при ОСШ, равном 4 дБ, 10 дБ, 16 дБ, 22 дБ, 28 дБ.

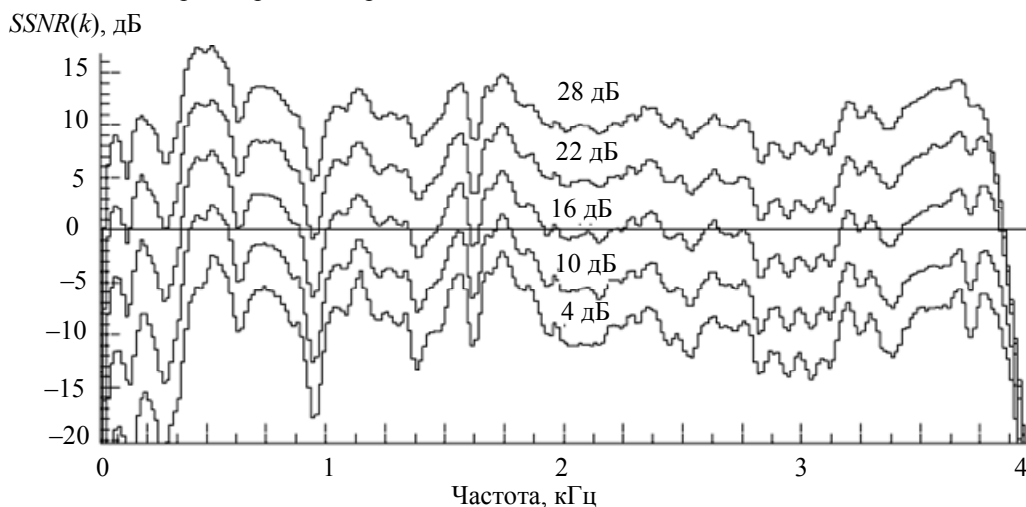


Рис. 2. Оценки $SSNR(k)$ [дБ] для зашумленной речи при разных значениях ОСШ сигнала

Графики, приведенные на рис. 2, демонстрируют изменение оценки сегментного ОСШ, соответствующего величине ОСШ сигнала.

Проведенные эксперименты выявили следующие характеристики алгоритма:

- по мере уменьшения величины ОСШ количество распознанных речевых кадров уменьшается, поскольку в качестве речевых сегментов распознаются лишь кадры с относительно большими локальными значениями ОСШ. Вследствие этого для сигналов с малыми значениями ОСШ алгоритм показал завышенные оценки $SSNR$.
- для сигналов с большими значениями ОСШ (больше 30 дБ) алгоритм показал заниженные оценки $SSNR$, что связано с тем, что алгоритм оценки спектра шума дает завышенные оценки спектра шума на участках речевого сигнала с большими величинами ОСШ.

Для расширения рабочего диапазона алгоритма оценки значения $SSNR$ корректировались путем следующего нелинейного преобразования:

$$SSNR_c [\text{дБ}] = 1,4 (SSNR [\text{дБ}] - 1).$$

Пример скорректированной оценки интегрального значения ОСШ для различных типов шумов и значений ОСШ приведен в таблице.

Экспериментально определенные значения времени настройки алгоритма показали, что оценки ОСШ приближались к устойчивым значениям на участке речи длительностью 3 с и становились достоверными на участке речи длительностью около 10 с. Время настройки алгоритма в значительной мере определяется параметром скорости адаптации в алгоритме оценки спектра шума. В данном варианте постоянная времени адаптации была задана равной 1 с. Поскольку для некоторых применений (например, текстозависимой верификации диктора) предъявляется требование к длительности сигнала не более 5 с, задача сокращения времени настройки остается актуальной.

Эксперименты с реальными зашумленными фонограммами показали хорошее соответствие оценок ОСШ ожидаемым значениям.

	$SSNR_c$							
	-2 дБ	4 дБ	10 дБ	16 дБ	22 дБ	28 дБ	34 дБ	40 дБ
Шум 1	-1,7	2,9	8,9	15,6	22,5	29,4	36,6	38,7
Шум 2	-1,35	3,11	9,11	15,9	22,9	30,2	37,3	38,6
Шум 3	-1,16	3,52	9,86	16,5	23,5	30,5	37,6	39,0
Шум 4	0,03	5,26	11,87	18,8	26,0	33,1	38,4	39,7
Среднее по шумам	-1,05	3,95	9,93	16,7	23,7	30,1	37,5	39,0

Таблица. Оценки $SSNR_c$ для различных уровней и типов шумов и уровней ОСШ, дБ

Применение алгоритма в системе верификации диктора

Пример графического экрана для процедуры оценки ОСШ представлен на рис. 3.

Работа алгоритма в системе происходит в два этапа. Сначала на кадрах речи вычисляется матрица ОСШ, $SNR(k, m)$. По ней в качестве выходных параметров вычисляются средние по всей фонограмме оценки ОСШ в частотных полосах и интегральное значение ОСШ, по которым принимается решение о качестве фонограммы. Затем, в случае пригодности фонограммы, производится более детальный анализ, и отбрасываются «плохие» кадры с малым значением ОСШ.

Основным результатом работы является практическое внедрение разработанного алгоритма в систему идентификации диктора. К настоящему моменту алгоритм прошел практическую проверку на больших объемах данных.

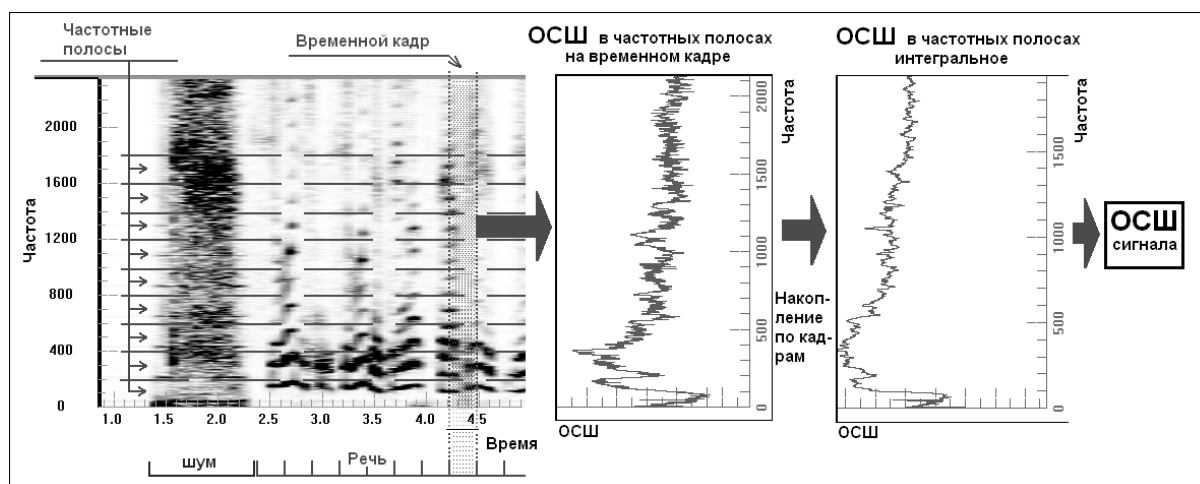


Рис. 3. Иллюстрация процедуры оценки ОСШ в частотных полосах

Заключение

Предложен алгоритм автоматической оценки интегральных и спектральных (в частотных полосах) значений ОСШ на фонограммах с зашумленной речью, использующий оценки текущих значений амплитудного спектра шума.

Основными компонентами реализованной схемы оценки ОСШ являются рекурсивный алгоритм оценки амплитудного спектра шума и детектор речи. Алгоритм оценки амплитудного спектра шума не требует наличия пауз речи в сигнале и устойчив к различным помехам. Детектор речи устойчив к присутствию в сигнале мощных тональных помех.

Предложенный алгоритм продемонстрировал свою работоспособность как на тестовых, так и на реальных записях речи. В настоящее время алгоритм используется в ряде продуктов ООО «ЦРТ».

Разработанный алгоритм оценки ОСШ удовлетворяет предъявленным требованиям, в частности: достоверная оценка ОСШ в интервале от 6 до 24 дБ на фонограммах, содержащих речь длительностью от 10 с.

Основными задачами дальнейшей работы являются обеспечение достоверности оценок при значениях ОСШ менее +6 дБ и на длительностях речи не менее 10 с.

Литература

1. ITU-T Rec. P. 56. Objective measurement of active speech level. – 1993. – Approved in Dec. 2011. – Printed in Switzerland, Geneva, 2012. – 17 p.
2. ITU-T G. 160. Objective measures for the characterization of the basic functioning of noise reduction algorithms. – 2008. – Approved in Nov. 2009. – Printed in Switzerland, Geneva, 2010. – 14 p.

3. Kim C., Stern R.M. Robust Signal-to-Noise Ratio Estimation Based on Waveform Amplitude Distribution Analysis // Proc. INTERSPEECH-2008. – Brisbane, Australia, 2008. – P. 2598–2601.
4. Nemer E., Goubiran R., Mahmoud S. SNR Estimation of Speech signals Using Subbands and Fourth-Order Statistics // IEEE Signal Processing Letters. – 1999. – V. 6. – № 7. – P. 171–174.
5. Hirsch H.G., Ehrlicher C. Noise estimation techniques for robust speech recognition // Proc. ICASSP. – Detroit, Michigan, 1995. – V. 1. – P. 153–156.
6. Hergolz C., Jeub M., Nelke C., Beaugeant C., Vary P. Evaluation of Single- and Dual-channel Noise Power Spectral Density Estimation Algorithms for Mobil Phones // Proc. Konferenz Elektronische Sprachsignalverarbeitung (EESV). – Aachen, Germany, 2011. – P. 1–10.

Столбов Михаил Борисович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ООО «ЦРТ-инновации», кандидат технических наук, ст. научный сотрудник, доцент, stolbov@speechpro.com

УДК 004.4'242

МЕТОД ПОСТРОЕНИЯ УПРАВЛЯЮЩИХ АВТОМАТОВ НА ОСНОВЕ МУРАВЬИНЫХ АЛГОРИТМОВ

Д.С. Чивилихин, В.И. Ульянов

Предлагается метод построения управляющих конечных автоматов по выбранной функции приспособленности, основанный на применении муравьиных алгоритмов. Проводится его апробация на задаче об «Умном муравье». Показано, что метод обладает большей производительностью по сравнению с традиционным подходом к построению управляющих автоматов на основе генетических алгоритмов. Предлагаемый метод не требует реализации таких нетривиальных операторов, как, например, скрещивание или мутация.

Ключевые слова: конечные автоматы, муравьиные алгоритмы.

Введение

В последние годы для решения разнообразных задач все чаще применяется автоматное программирование [1]. В рамках этого подхода поведение программ описывается с помощью детерминированных конечных автоматов. При этом часто эвристическое построение автоматов затруднено, поэтому для этой цели применяются различные эволюционные алгоритмы, такие как генетические алгоритмы [2–5] и эволюционные стратегии [2, 6].

При построении автоматов с помощью эволюционных алгоритмов вводится функция, называемая функцией приспособленности, отражающая то, насколько автомат близок к решению задачи. Таким образом, построение автомата для той или иной задачи сводится к построению автомата с наибольшим или наименьшим значением функции приспособленности. Отдельно отметим, что введение функции приспособленности не предполагает наличие эталонного автомата. Примерами задач, для которых можно найти решение, представленное управляющим автоматом, с помощью эволюционных алгоритмов, являются:

- задача об «Умном муравье» [1, 3, 7];
- задача о завоевании ресурсов [2];
- задача о построении автомата на основе тестовых примеров [3, 8].

В начале работы эволюционный алгоритм генерирует некую начальную популяцию особей (в данном случае – конечных автоматов). Обычно особи начальной популяции генерируются случайным образом. Далее, пока не будет выполнено условие останова, выполняются операции мутации, скрещивания и селекции. Операции мутации и скрещивания по некоторым правилам изменяют особи в текущей популяции. Операция селекции определяет, какие особи перейдут в следующую популяцию.

При построении конечных автоматов с помощью эволюционных алгоритмов требуется разрабатывать операции мутации и скрещивания. В настоящей работе предлагается новый метод построения управляющих автоматов, основанный на муравьином алгоритме [9], не требующий реализации обозначенных операций.

Вопрос сходимости муравьиных алгоритмов изучен на данный момент сравнительно слабо. Доказательство сходимости некоторых классов муравьиных алгоритмов можно найти в [10]. Вопрос о сходимости других типов муравьиных алгоритмов остается на данный момент открытым. Сходимость предложенного алгоритма построения автоматов в настоящей работе не изучается.

Муравьиные алгоритмы

Муравьиные алгоритмы – группа алгоритмов оптимизации на графах, принцип работы которых основан на поведении муравьев, ищущих путь от муравейника до источника пищи. Примером задачи, которая может быть решена с помощью муравьиного алгоритма, является задача о коммивояжере [11].

В муравьиных алгоритмах решения строятся набором агентов-муравьев, использующих при выборе пути в графе некоторую стохастическую стратегию. Решения могут быть представлены как путями в графе, так и отдельными его вершинами. Каждому ребру графа (u, v) (здесь u и v – вершины графа) сопоставлено некоторое действительное число τ_{uv} , называемое значением феромона, и (необязательно) некоторое эвристическое расстояние η_{uv} . Значения феромона модифицируются муравьями в процессе поиска, в то время как эвристические расстояния задаются перед началом работы алгоритма и далее не изменяются.

Можно выделить три основных этапа работы муравьиного алгоритма, которые повторяются, пока не будет найдено допустимое решение или пока не выполнится условие останова. На первом этапе, назовем его **ConstructSolutions**, каждый муравей совершает переходы по ребрам графа, запоминая свой путь – последовательность ребер. Находясь в определенной вершине, каждый муравей определяет следующее ребро, исходя из эвристических расстояний и значений феромона на ребрах, инцидентных данной вершине. Определив ребро, муравей добавляет его к своему пути и переходит по этому ребру к следующей вершине.

На втором этапе – **UpdatePheromones** – значения феромона на всех ребрах графа изменяются. Значение феромона на ребре может увеличиться, если по этому ребру прошел муравей, или уменьшиться вследствие испарения. При испарении значения феромона на всех ребрах графа изменяются в заданное число раз. На третьем (необязательном) этапе – **DaemonActions** – применяется некоторая локальная оптимизация или другие действия, которые не могут быть выполнены отдельными муравьями.

Управляющие конечные автоматы

Будем называть управляющим детерминированный конечный автомат (ДКА), на каждом переходе которого записаны событие и последовательность выходных воздействий. События поступают на вход автомата из внешней среды. При поступлении события автомат переходит в новое состояние и передает выходные воздействия объекту управления. Пример управляющего конечного автомата приведен на рис. 1.

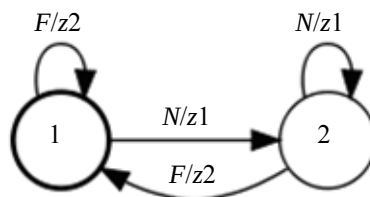


Рис. 1. Пример управляющего конечного автомата

Для данного автомата множество входных событий равно $\{F, N\}$, а множество выходных воздействий – $\{z1, z2\}$. Стартовое состояние автомата на рис. 1 и на всех других рисунках в данной работе отмечено жирной линией.

Предлагаемый метод построения управляющих автоматов на основе муравьиных алгоритмов

Рассмотрим задачу о построении управляющего конечного автомата [1] при фиксированном множестве входных событий, выходных воздействий и введенной функции приспособленности. Пусть задано число состояний N , множество входных событий $Events$ и множество выходных воздействий $Actions$ автомата. Задача состоит в поиске автомата с наибольшим значением функции приспособленности.

Рассмотрим представление конечного автомата в виде графа переходов. При этом подходе состояния автомата соответствуют вершинам графа, а переходы – его дугам. Рассмотрим множество всех автоматов с параметрами $(N, Events, Actions)$. Любой автомат из выбранного множества как граф является подграфом полного недетерминированного конечного автомата (НКА) с теми же параметрами. Под полным НКА понимается автомат, в котором из каждого из N состояний по каждому входному событию из $Events$ существует переход в каждое состояние с каждым выходным воздействием из $Actions$. Пример такого автомата с двумя состояниями приведен на рис. 2.

Задача оптимизации ДКА для муравьиного алгоритма может быть сформулирована следующим образом. Задано множество входных событий $Events$, множество выходных воздействий $Actions$ и число состояний N . Также задана функция приспособленности, определенная на множестве всех ДКА с указанными параметрами, сопоставляющая автомату действительное число. На полном НКА с теми же параметрами требуется найти такой путь, чтобы ДКА, составленный из ребер и вершин найденного пути, имел наибольшее значение функции приспособленности.

В начале работы алгоритма всем ребрам полного НКА приписывается некоторый одинаковый ненулевой вес – значение феромона в терминах муравьиных алгоритмов. В каждую вершину полного НКА помещается по муравью. Каждый муравей выбирает переходы из своего состояния по каждому входному

символу, исходя из значений феромона на переходах. Выбор производится с помощью метода «рулетки» [11]. Заметим, что, в отличие от классического муравьиного алгоритма, описанного в разделе «Муравьиные алгоритмы», каждый муравей строит переходы только из своего состояния, не перемещаясь в другие состояния (вершины).

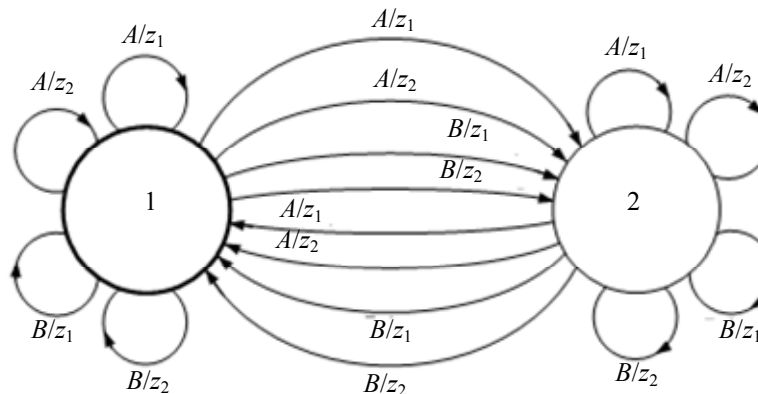


Рис. 2. Пример полного недетерминированного конечного автомата

По выбранным муравьями состояниям и переходам строится ДКА. Далее для ДКА вычисляется выбранная функция приспособленности. При этом запоминаются переходы, которые совершал автомат. К весам посещенных при вычислении функции приспособленности переходов НКА добавляются дополнительные веса, пропорциональные вычисленному значению. Отметим, что веса тех переходов, которые не были посещены при вычислении функции приспособленности, не изменяются, а сами переходы удаляются из ДКА. Также из ДКА удаляются непосещенные состояния.

На следующем этапе со всех ребер НКА «испаряется» феромон – веса всех ребер уменьшаются в одинаковое число раз.

Наконец, проверяются условия сходимости (достижения требуемого значения функции приспособленности) и стагнации. Под стагнацией понимается ситуация, при которой значение функции приспособленности лучшего автомата не увеличивается в течение некоторого фиксированного промежутка времени (числа шагов алгоритма). В таком случае алгоритм перезапускается. При достижении требуемого значения функции приспособленности алгоритм прекращает свою работу, выдавая лучший из построенных автоматов.

Экспериментальное исследование

В рамках экспериментального исследования предлагаемый метод был протестирован на задаче об «Умном муравье» [1, 4, 7]. Для вычислений использовался персональный компьютер с процессором Intel Core i7 2600 с тактовой частотой 3,4 ГГц.

Опишем задачу об «Умном муравье». Задано квадратное тороидальное поле размером 32×32 клетки. На поле вдоль заданной ломаной расположено 89 «яблок» – единиц еды, причем яблоки располагаются не в каждой клетке этой ломаной. В начале пути муравей находится в левой верхней клетке поля и «смотрит» на восток. Поле, начальное расположение муравья и расположение еды на поле показано на рис. 3. Белые клетки пусты, в черных клетках расположена еда, а серым цветом отмечены клетки ломаной, не содержащие еды.

Муравей видит одну клетку перед собой: он может определить, есть еда в следующей клетке или нет. У муравья есть 200 ходов. На каждом ходу он может выполнить одно из трех действий: повернуть налево, повернуть направо, перейти на одну клетку вперед. При этом если клетка, в которую перешел муравей, содержит еду, муравей съедает эту еду.

В данной задаче рассматриваются два входных события – **N** (в следующей клетке нет еды) и **F** (в следующей клетке есть еда) – а также три выходных воздействия: **L** (повернуть налево), **R** (повернуть направо) и **M** (перейти на одну клетку вперед).

Задача состоит в том, чтобы построить систему управления муравьем, которая позволит ему съесть всю еду за минимальное число шагов. В работе [3] предлагается решение данной задачи путем построения конечного автомата, управляющего муравьем, с помощью генетического программирования. В указанной работе был найден автомат, имеющий семь состояний и позволяющий муравью съесть всю еду за отведенное число шагов. Число вычислений функции приспособленности, потребовавшихся для построения такого автомата – порядка сотен миллионов ($160 \cdot 10^6$ и $250 \cdot 10^6$ для двух разных экспериментов).

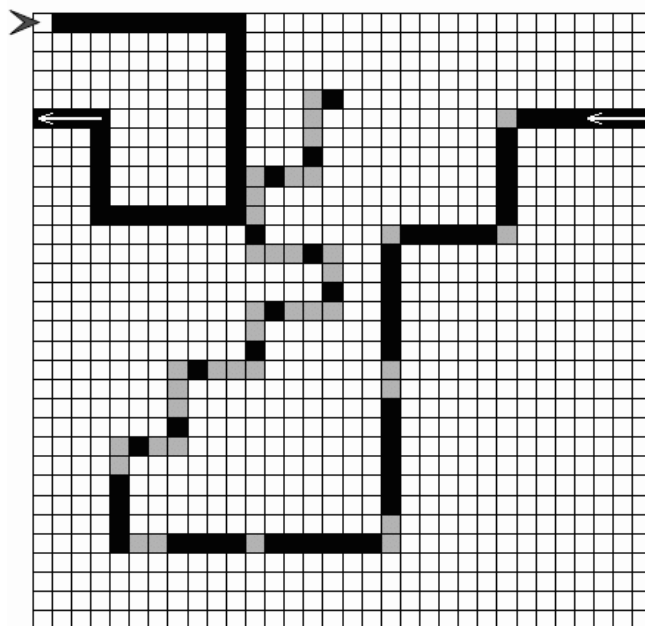


Рис. 3. Поле, начальное расположение муравья и еды в задаче об «Умном муравье»

Традиционно функция приспособленности автомата, управляющего умным муравьем, имеет вид [1]

$$f(fsm) = n + \frac{200 - numberOfSteps}{200}, \quad (1)$$

где fsm – конечный автомат; n – число съеденных яблок, а $numberOfSteps$ – номер хода, на котором было съедено последнее яблоко.

При использовании функции приспособленности (1) предлагаемый муравьиный алгоритм не нашел оптимального решения задачи об «Умном муравье». В настоящей работе авторами предлагается использовать модифицированную функцию приспособленности, учитывающую число состояний автомата:

$$f(fsm) = n + \frac{200 - numberOfSteps}{200} + 0,1 \cdot (M - N), \quad (2)$$

где M – число состояний НКА, в котором проводится поиск с помощью муравьиного алгоритма, а N – число состояний ДКА, посещенных при вычислении функции приспособленности. При использовании выражения (2) оптимизация проводится одновременно по числу съеденных яблок, числу потребовавшихся для этого шагов, а также по числу использованных состояний автомата.

Авторами было проведено 15 экспериментов по построению оптимального автомата. В каждом эксперименте были использованы следующие параметры алгоритма:

- скорость испарения феромона – 0,5;
- стагнационный параметр – 10^6 ;
- число состояний полного НКА – 12.

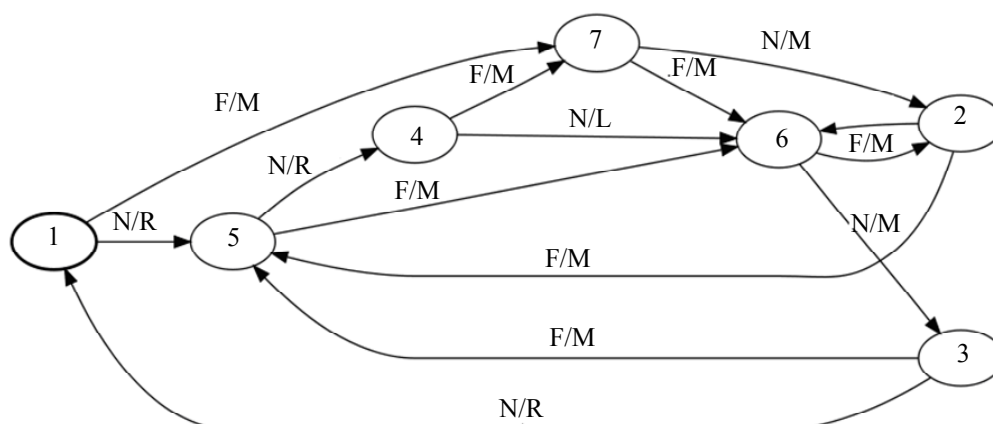


Рис. 4. Управляющий автомат для задачи об «Умном муравье»

В каждом из экспериментов был получен автомат, имеющий семь состояний, который съедает всю еду за 189 ходов. Минимальное число вычислений функции приспособленности, потребовавшееся для

построения оптимального автомата $-1,65 \cdot 10^6$, максимальное $-60,16 \cdot 10^6$, среднее значение $-26,49 \cdot 10^6$ (стандартное отклонение $-17,45 \cdot 10^6$). При этом среднее время построения оптимального автомата составило примерно 32 мин. Диаграмма переходов одного из построенных автоматов приведена на рис. 4. При этом отметим, что для построения такого автомата с помощью генетического программирования в работе [3] потребовалось в среднем $220 \cdot 10^6$ вычислений функции приспособленности. Таким образом, с помощью предложенного алгоритма удалось найти решение задачи об «Умном муравье» в 8,3 раза быстрее, чем при использовании генетического алгоритма.

Заключение

Разработан метод построения управляющих конечных автоматов, основанный на муравьиных алгоритмах. Разработано программное средство на языке Java, реализующее предлагаемый метод. Метод был протестирован на задаче об «Умном муравье». При использовании модифицированной функции приспособленности, учитывающей число состояний автомата, удалось найти решение этой задачи в 8,3 раза быстрее, чем с помощью генетического алгоритма.

Преимущество предлагаемого подхода перед генетическим алгоритмом также состоит в том, что для решения каждой конкретной задачи не требуется определять функции мутации и скрещивания, что обычно является нетривиальной задачей.

Исследование поддержано в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

Литература

1. Поликарпова Н.И., Шалыто А.А. Автоматное программирование. – СПб: Питер, 2009. – 176 с.
2. Spears W.M., Gordon D.F. Evolving finite-state machine strategies for protecting resources // Proceedings of the International Symposium of Methodologies for Intelligent Systems 2000. ACM Special Interest Group on Artificial Intelligence. – Springer-Verlag, 2000. – P. 166–175.
3. Царев Ф.Н. Метод построения автоматов управления системами со сложным поведением на основе тестов с помощью генетического программирования // Материалы Международной научной конференции «Компьютерные науки и информационные технологии». – Саратов: СГУ, 2009. – С. 216–219.
4. Царев Ф.Н., Шалыто А.А. Применение генетического программирования для генерации автомата в задаче об «Умном муравье» // Сборник трудов IV-ой Международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте». – М.: ФИЗМАТЛИТ, 2007. – Т. 2. – С. 590–597.
5. Александров А.В., Казаков С.В., Сергушичев А.А., Царев Ф.Н., Шалыто А.А. Генерация конечных автоматов для управления моделью беспилотного самолета // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2 (72). – С. 3–11.
6. Lucas S.M. Evolving finite state transducers: Some initial explorations. Genetic Programming // Lecture Notes in Computer Science. – Berlin: Springer, Heidelberg, 2003. – V. 2610. – P. 241–257.
7. Лобанов П.Г. Использование генетических алгоритмов для генерации конечных автоматов: Дис. ... канд. тех. наук: 05.13.11. – СПбГУ ИТМО, 2008. – 114 с.
8. Егоров К.В., Царев Ф.Н. Совместное применение генетического программирования и верификации моделей для построения автоматов управления системами со сложным поведением // Сборник докладов конференции молодых ученых и специалистов «Информационные технологии и системы». – М.: ИППИ РАН, 2009. – С. 77–82.
9. Dorigo M. Optimization, Learning and Natural Algorithms // Ph. D. thesis. – Politecnico di Milano. Italy, 1992. – 140 p.
10. Dorigo M., Stützle T. Ant Colony Optimization. – MIT Press, 2004. – 305 p.
11. Dorigo M., Gambardella L.M. Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem // IEEE Transactions on Evolutionary Computation. – 1997. – V. 1. – № 1. – P. 53–66.
12. Baker J.E. Reducing Bias and Inefficiency in the Selection Algorithm // Proceedings of the Second International Conference on Genetic Algorithms and their Application. – USA: Lawrence Erlbaum Associates, 1987. – P. 14–21.

Чивилихин Даниил Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, chivilikhin.daniil@gmail.com

Ульянцев Владимир Игоревич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, ulyantsev@rain.ifmo.ru

УДК 621.397

ПРОГРАММИРОВАНИЕ ЗАДАЧ ВОССТАНОВЛЕНИЯ ИСКАЖЕННЫХ ИЗОБРАЖЕНИЙ НА C/C++ В СИГНАЛЬНЫХ МИКРОПРОЦЕССОРАХ ФИРМЫ TEXAS INSTRUMENTS

К.А. Кирьянов, В.С. Сизиков

Рассматривается инструментальная реализация алгоритмов восстановления искаженных (смазанных, дефокусированных и (или) зашумленных) изображений. Рассмотрены особенности программирования на языке C/C++ ранее разработанных и опубликованных алгоритмов для прямой, а также более сложной и важной для практики обратной задачи.

Ключевые слова: искажение изображения, прямая и обратная задачи, язык C/C++, сигнальный микропроцессор, персональный компьютер.

Введение. Реализуемые методы и алгоритмы

Известные задачи восстановления (реконструкции, реставрации) искаженных (смазанных, дефокусированных и (или) зашумленных) изображений сформулированы в работах [1–12] и др. Они обычно описываются набором одномерных интегральных уравнений (ИУ) Фредгольма I рода типа свертки,

$$Aw \equiv \int_{-\infty}^{\infty} h(x - \xi) w_y(\xi) d\xi = g_y(x) + \delta g, \quad a \leq x \leq b, \quad c \leq y \leq d, \quad (1)$$

или одним двумерным ИУ Фредгольма I рода типа свертки,

$$Aw \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x - \xi, y - \eta) w(\xi, \eta) d\xi d\eta = g(x, y) + \delta g, \quad a \leq x \leq b, \quad c \leq y \leq d, \quad (2)$$

где h – функция рассеяния точки (ФРТ), в большинстве случаев пространственно-инвариантная (разностная); A – оператор; w и g – распределение интенсивности по истинному и искаженному изображениям соответственно; δg – помеха. В (1) ось x направлена вдоль смаза, а y играет роль параметра. Набор уравнений (1) часто используется в задаче смазывания, а (2) – в задаче дефокусирования изображения [2–11].

Задача решения ИУ (1) и (2) является некорректной (существенно неустойчивой), поэтому ИУ (1) при каждом y часто решают методом квадратур или методом одномерного преобразования Фурье (ПФ) с регуляризацией Тихонова [7, 9–11], а ИУ (2) – методом двумерного ПФ с регуляризацией Тихонова [2–7, 10, 11]. Для решения ИУ (1) и (2) используют также метод параметрической фильтрации Винера [4, 5, 7, 9–11], алгоритм Люси–Ричардсона [5] и метод «слепой» деконволюции [5]. Для решения ИУ (1) или (2) можно использовать также метод итеративной регуляризации Фридмана [10, 11, 13] (о других методах см. [4, 5, 13]). Эти методы могут быть реализованы не только на универсальных ЭВМ или персональных компьютерах (ПК) [3, 5, 10, 11], но и на спецпроцессорах, в частности, сигнальных микропроцессорах или систолических процессорах [8, 12, 14–16]. Использование спецпроцессоров ведет к уменьшению габаритов обрабатывающих устройств и позволяет повысить качество прибора наблюдения (фотоаппарата, телескопа, микроскопа, спектрометра и т.д.) за счет встроенных алгоритмов обработки изображения.

В работе рассматриваются особенности задачи реализации методов и алгоритмов решения ИУ (1) и (2) применительно к восстановлению искаженных изображений с помощью цифровых сигнальных процессоров (ЦСП) и программируемых логических интегральных схем (ПЛИС) [8, 12, 14–16]. В работе [12] уже рассмотрена специфика реализации на спецпроцессорах метода квадратур и метода ПФ с регуляризацией Тихонова применительно к ИУ (1). В настоящей работе остановимся на методе итеративной регуляризации Фридмана решения ИУ (2).

Метод итеративной регуляризации Фридмана

Полагаем, что ядро h уравнения (2) симметрично, т.е. $h(x, y) = h(|x|, |y|)$. Тогда метод Фридмана для решения ИУ (2) описывается следующим соотношением [10, 11, 13]:

$$w_k(x, y) = w_{k-1}(x, y) + v \left[g(x, y) - \int_a^b \int_c^d h(x - \xi, y - \eta) w_{k-1}(\xi, \eta) d\xi d\eta \right], \quad (3)$$

где $w_{k-1}(x, y)$ и $w_k(x, y)$ – решения, полученные в $(k-1)$ -й и k -й итерациях (приближениях) соответственно, при этом $w_0(x, y)$ – начальное приближение; $k = 1, 2, 3, \dots$ – номер итерации; v – параметр, удовлетворяющий условию $0 \leq v \leq 2/\|A\|$, где $\|A\| = \left\{ \int_a^b \int_c^d h^2(x, y) dx dy \right\}^{1/2}$ – норма ФРТ.

Для задач дефокусирования $w_k(x, y)$ – изображение, получаемое в k -й итерации, а $g(x, y)$ – правая часть уравнения (искаженное изображение). Метод Фридмана весьма чувствителен к начальному при-

ближению, которое при реализации обычно полагается равным $w_0(x, y) = 0$. Задавшись каким-то (не обязательно нулевым) начальным приближением и выполняя итерационный процесс (3), в принципе можно получить приближение к искомому решению (истинному изображению). Но, поскольку задача некорректна, а правая часть обычно зашумлена, то до некоторого номера итераций k^* процесс обычно сходится к решению, после чего он начинает расходиться [10, рис. 2.70; 13, рис. 21]. Поэтому нужно ввести критерий останова процесса, т.е. выбора числа итераций k^* , которое играет роль параметра регуляризации [13, с. 272–275].

Особенности программной реализации алгоритмов на PC и DSP

Оценим объем памяти, требуемый для размещения одного изображения. Для примера рассмотрим серое изображение размером $512 \times 512 = 262\,144$ пикселей, записываемое с разрядностью 8 бит на пиксель. Очевидно, что для хранения всего изображения необходимо приблизительно 2 Мбит памяти. Для реализации метода Фридмана требуется хранить 3 изображения: g , w_{k-1} и w_k , что составляет 6 Мбит памяти. Эта величина не критична для ПК, но сопоставима с общим объемом памяти современных высокопроизводительных ЦСП. Таким образом, ключевым моментом рассматриваемой задачи для ЦСП является оптимизация объема используемой оперативной памяти.

Реализация алгоритмов на ПК. На ПК в системе Windows на одну точку изображения отводится 32 бита (4 Б). Из них первый байт (A) отвечает за прозрачность изображения, а три других байта (R , G и B) содержат значения интенсивностей по соответствующим цветовым каналам ($R_{исх}$, $G_{исх}$, $B_{исх}$). В данной работе мы рассматриваем обработку серых изображений, поэтому полагаем по умолчанию $A = 255$ и $R = G = B = (R_{исх} + G_{исх} + B_{исх})/3$. Данные о каждой точке имеют тип BYTE (или unsigned char), значения которого лежат в пределах 0–255, в то время как при решении прямой и обратной задач нужны данные типа double. Поэтому при программной реализации алгоритмов требуется преобразовывать исходные данные в формат double.

При решении прямой задачи (моделировании искажений изображения) часто используются искусственные граничные условия [4, 5], в частности, дополнение исходного изображения нулями вне границ. Однако можно использовать более естественный способ – усечение изображения [7, 9–11]. Кроме того, для уменьшения эффекта Гиббса при решении обратной задачи (задачи восстановления изображения) можно использовать размытие краев изображения [7, 9–11]. Если восстановление осуществляется с помощью алгоритма быстрого преобразования Фурье (БПФ), что особенно характерно для задачи дефокусирования (решения уравнения (2)), то может потребоваться дополнение исходного изображения нулями до размера $2^M \times 2^N$. Чтобы данные не выходили за значение 255, вводится нормировка ФРТ, после чего вычисляется свертка с нормированной ФРТ (прямая задача). Для вывода на экран полученного искаженного изображения осуществляется обратное преобразование типа double в тип BYTE. Дробная часть данных при этом теряется, поэтому числовой массив типа double необходимо сохранять для решения обратной задачи.



Рис. 1. Дефокусированное и зашумленное изображение с размытием краев (а); изображение, восстановленное методом итеративной регуляризации Фридмана ($\nu = 1/\|A\|$, число итераций $k^* = 200$) (б)

Перейдем к анализу обратных задач. В обратной задаче смазывания решается уравнение (1) методом квадратур или ПФ с регуляризацией Тихонова, а также методом параметрической фильтрации Винера, итераций Фридмана и т.д. Особенности методов изложены в работах [7–12]. В обратной задаче дефокусирования, в случае пространственно-инвариантной ФРТ [2–12] решается уравнение (2) методом двумерного ПФ с регуляризацией Тихонова (или другим методом). В конце восстановления изображения путем вычисления обратного ПФ от регуляризованного спектра [7–12] из результата берется только действительная часть, а мнимая часть обратного ПФ отбрасывается. Если же ФРТ является пространствен-

но-зависимой, то решение целесообразно осуществить рассмотренным выше методом итеративной регуляризации Фридмана [10–13]. После восстановления изображения необходимо снова преобразовать данные в формат BYTE и отобразить восстановленное изображение. Пример восстановления дефокусированного (и зашумленного) изображения методом Фридмана приведен на рис. 1.

Реализация алгоритмов на ЦСП. Большинство задач цифровой обработки основано на численных методах работы с матрицами, а также на методах реализации свертки и дискретных ортогональных преобразований [15–20]. Как отмечалось в [12], основная сложность в программировании алгоритмов с помощью ЦСП – это чрезвычайно ограниченные размеры оперативной памяти. В работах [8, 12] был обоснован выбор элементной базы для решения рассматриваемых задач, в том числе, микропроцессора TMS320C6457, который имеет не более 2 МБ оперативной памяти. Как отмечалось в [12], для методов, основанных на дискретных ортогональных преобразованиях, в частности, на ПФ, одним из оптимальных по быстродействию и по распределению памяти является БПФ. Для двумерного БПФ при получении двумерного Фурье-спектра из исходного изображения нужен исходный двумерный массив (изображение) и еще два одномерных массива для временных данных по столбцам. В C/C++ двумерный массив – это длинный одномерный массив с расположением элементов по строкам. В связи с этим для вычисления БПФ по столбцам необходимо брать элемент текущего столбца и записывать его во временный массив, длина которого равна степени 2.

ЦСП TMS320C6457 – это процессор с фиксированной точкой, в то время как необходимо работать с данными типа double и выполнять операции с плавающей точкой. Как отмечено в [12], ядро микропроцессора имеет 32 разряда, а данные типа double имеют 64 разряда. Для преодоления этого противоречия фирмой Texas Instruments [21] разработана библиотека IQmath [22] для работы с виртуальной точкой, с помощью которой ЦСП с фиксированной точкой (процессоры с ядром C64x+) реализуют арифметику с плавающей точкой. Процессоры с плавающей точкой (C67x+) работают на более низких частотах. Если процессор C6457 работает на частоте 1 ГГц, то тактовая частота любого из процессоров C67x+ не превышает 400 МГц. Также ощутима разница в размерах оперативной памяти: если у процессора C6457 ее размер составляет 2 МБ (кэш L2 внутри кристалла), то у большинства процессоров серии C67x+ он не превышает 300 КБ.

Во всех микропроцессорах старший разряд – всегда знаковый. На рис. 2 показано перемножение двух чисел с точкой. Для отображения результатов перемножения двух четырехразрядных чисел с полной точностью требуется 8 разрядов. Но точка перемещается также в результате перемножения. Для обеспечения необходимой точности достаточно правильно извлечь 4 разряда, как это показано на рис. 2.

Множимое и множитель имеют по 4 разряда: один до точки и три – после нее. То же самое относится и к результату. Следовательно, процессор должен иметь возможность работать с данными типа double, которые имеют 64 разряда, несмотря на то, что ядро процессора имеет только 32 разряда. Микропроцессоры серии C64x+ способны выполнять за 1 такт две операции умножения. С помощью библиотеки IQMath можно реализовать программу так, чтобы извлекать только нужные разряды из результата. Заметим, что в принципе это можно сделать и без использования упомянутой библиотеки, но тогда код становится менее читаемым и более громоздким.

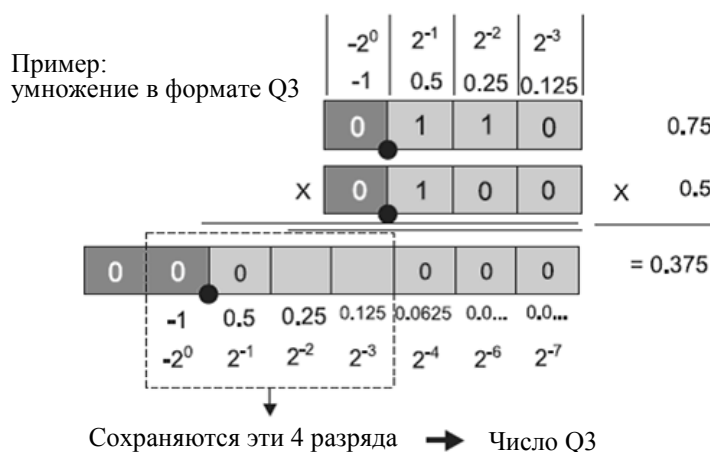


Рис. 2. Перемножение двух чисел с точкой

На рис. 3 показаны три примера кода: первый записан для операций с плавающей точкой, второй реализован только на языке C, третий (Q-вариант) реализован с помощью библиотеки IQMath. Следует заметить, что 32-разрядные данные для DSP C64x+ представлены типом данных int, а 64-разрядные данные – типом long long. После умножения происходит сдвиг на Q разрядов, после чего следует операция суммирования (накопления). Рисунок наглядно показывает, насколько проще и нагляднее код с использованием библиотеки IQMath по сравнению с кодом при реализации только на языке C для ЦСП.

Для выполнения свертки и различных ортогональных преобразований для микропроцессоров TI DSP C64x+ существует библиотека dsplib, в которой имеются написанные библиотеки функций БПФ, оптимизированные библиотеки векторно-матричных перемножений и различных операций с матрицами. Это существенно упрощает работу по написанию кода по сравнению с ситуацией двадцатилетней давности, когда еще не существовали С-компиляторы для ЦСП и разработчики были вынуждены работать исключительно на языке Ассемблер [23].

Плавающая точка	<pre>float Y, M, X, B; Y = M * X + B;</pre>
Q-вариант, реализованный только на языке C	<pre>int Y, M, X, B; Y = ((long long) M * (long long) X) >> Q + B;</pre>
Q-вариант, реализованный с помощью библиотеки IQMath	<pre>_iq Y, M, X, B; Y = _IQmpy(M, X) + B;</pre>

Рис. 3. Три примера кода

Отметим, что производителем Einfochips Limited поставляется отладочный модуль TMDSEVM6457L на базе микропроцессора TMS320C6457 фирмы Texas Instruments. Кроме памяти, расположенной непосредственно внутри кристалла микропроцессора, на борту модуля имеется две микросхемы памяти общим размером 256 МБ и JTAG-интерфейс для пошаговой отладки программ, а также HPI-порт для обмена данными, который может служить для подключения оценочных модулей АЦП и ЦАП через плату сопряжения.

Заключение

В работе рассмотрены сравнительные особенности инструментальной реализации методов и алгоритмов восстановления искаженных (смазанных, дефокусированных, зашумленных) изображений на персональных ЭВМ (PC) и цифровых сигнальных процессорах (DSP). Анализ выполнен применительно к решению прямой и обратной задач смазывания и дефокусирования изображений, описываемых одномерными и двумерными интегральными уравнениями, решение которых осуществляется устойчивыми методами (в работе основное внимание уделено методу итеративной регуляризации Фридмана).

Особенности программирования изложены применительно к языку C/C++. Установлено, что наиболее быстрая реализация алгоритмов достигается в случае разработки кода на языке C/C++ с использованием специализированных библиотек фирмы Texas Instruments под каждое ядро процессора. Для создания быстродействующих алгоритмов необходимо использовать Ассемблер, который позволяет наиболее полно адаптировать код к типу процессора и периферийной схмотехники. Однако в этом случае время написания кода и трудоемкость работы велики. На практике для решения задач чаще всего используется код, написанный на языке C/C++, в некоторых случаях с использованием подпрограмм на языке Ассемблер.

Конвейерная архитектура, реализованная в процессорах C64x+, позволяет выполнять несколько операций одновременно, но требования к алгоритму при этом таковы, что данные выполняемых операций не должны зависеть друг от друга не только в момент выполнения операций, но и для следующего набора команд в конвейер. В этой связи алгоритмы для DSP необходимо реализовывать таким образом, чтобы его оптимизация на уровне конвейера производилась наиболее эффективным образом.

Выявлено, что в среде Code Composer Studio v4.2.0 имеется хороший инструментарий для Фурье-анализа входных последовательностей, а также хорошая графическая оболочка для построения и анализа временных зависимостей, что существенно упрощает работу с необходимым математическим аппаратом.

Проанализированы возможности ПК и ЦСП по быстродействию и памяти, а также по разрядности и форматам чисел. В качестве элементной базы выбран микропроцессор TMS320C6457 – процессор с фиксированной точкой, а вычисления, связанные с обработкой изображений, нужно вести с плавающей точкой. Показано, что для перехода к плавающей (виртуальной) точке целесообразно использовать разработанную фирмой Texas Instruments библиотеку IQmath.

Литература

1. Бейтс Р., Мак-Доннелл М. Восстановление и реконструкция изображений. – М.: Мир, 1989. – 336 с.
2. Бакушинский А.Б., Гончарский А.В. Некорректные задачи. Численные методы и приложения. – М.: Изд-во МГУ, 1989. – 199 с.

3. Сизиков В.С. Математические методы обработки результатов измерений. – СПб: Политехника, 2001. – 240 с.
4. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2006. – 1072 с.
5. Гонсалес Р., Вудс Р., Эддинс С. Цифровая обработка изображений в среде MATLAB. – М.: Техносфера, 2006. – 616 с.
6. Воскобойников Ю.Е. Комбинированный нелинейный алгоритм восстановления контрастных изображений при неточно заданной аппаратной функции // Автометрия. – 2007. – Т. 43. – № 6. – С. 3–18.
7. Сизиков В.С., Римских М.В., Мирджамолов Р.К. Реконструкция смазанных и зашумленных изображений без использования граничных условий // Оптический журнал. – 2009. – Т. 76. – № 5. – С. 38–46.
8. Кирьянов К.А. Инструментальная реализация алгоритмов реконструкции искаженных изображений // Труды 20-й Междунар. конф. «GraphiCon-2010». – СПб: СПбГУ ИТМО, 2010. – С. 188–191.
9. Сизиков В.С. Прием «усечение–размытие–поворот» для восстановления искаженных изображений // Оптический журнал. – 2011. – Т. 78. – № 5. – С. 18–26.
10. Сизиков В.С. Обратные прикладные задачи и MatLab. – СПб: Лань, 2011. – 256 с.
11. Сизиков В.С. Интегральные уравнения и MatLab в задачах томографии, иконики и спектроскопии. – Saarbrücken : LAP (LAMBERT Academic Publishing), 2011. – 252 с.
12. Кирьянов К.А., Сизиков В.С. Применение сигнальных микропроцессоров в задачах реконструкции искаженных изображений // Изв. вузов. Приборостроение. – 2011. – Т. 54. – № 7. – С. 20–26.
13. Верлань А.Ф., Сизиков В.С. Интегральные уравнения: методы, алгоритмы, программы. – Киев: Наук. думка, 1986. – 544 с.
14. Кухарев Г.А., Тропченко А.Ю., Шмерко В.П. Системные процессоры для обработки сигналов. – Минск: Беларусь, 1988. – 127 с.
15. Кухарев Г.А., Тропченко А.Ю. Системный процессор для обращения матриц // Изв. вузов. Приборостроение. – 1990. – Т. 33. – № 11. – С. 23–27.
16. Тропченко А.Ю. Аппаратные средства для цифровой обработки сигналов: Уч.-методич. пособие по дисциплине «Методы обработки сигналов и изображений». – СПб: СПбГУ ИТМО, 2005. – 138 с.
17. Рабинер Л., Гоулд В. Теория и применение цифровой обработки сигналов. – М.: Мир, 1978. – 848 с.
18. Царев А.П. Алгоритмические модели и структуры высокопроизводительных процессоров цифровой обработки сигналов. – Szczecin: Informa, 2000. – 237 с.
19. Поршнев С.В. Вычислительная математика. Курс лекций. – СПб: БХВ-Петербург, 2004. – 304 с.
20. Киреев В.И., Пантелеев А.В. Численные методы в примерах и задачах. – М.: Высшая школа, 2006. – 480 с.
21. Texas Instruments. Digital Signal Processors [Электронный ресурс]. – Режим доступа: <http://www.ti.com/dsp>, свободный. Яз. англ. (дата обращения 18.10.2012).
22. Фингер Р. Библиотека функций IQMath для C64x+ с плавающей точкой в ЦСП с фиксированной точкой. Компоненты TI // Бюллетень научно-технич. информации. – 2011. – Вып. 2(30).
23. Blonstein S., Katorgi M. eXpressDSP For Dummies (eXpressDSP для «чайников»): Пер. с англ. – Новосибирск: 2004. – 101 с.

Кирьянов Константин – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, kiryancon@front.ru
Александрович
Сизиков Валерий Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, sizikov2000@mail.ru

УДК 004.5

КОНЕЧНО-АВТОМАТНАЯ РЕАЛИЗАЦИЯ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА ВСТРАИВАЕМОЙ СИСТЕМЫ НА ОСНОВЕ АСПЕКТНО-ОРИЕНТИРОВАННОГО ОПИСАНИЯ

В.М. Васюков

Предлагается методика перехода от описания пользовательского интерфейса встраиваемой системы с использованием аспектно-ориентированного подхода к описанию в виде конечного автомата. Применимость предлагаемого способа демонстрируется на примере устройства индикации какой-либо физической величины. Производится описание последовательных переходов от первоначального описания пользовательского интерфейса в терминах аспектно-ориентированного подхода к ориентированному графу, описывающему возможные варианты изменения состояний пользовательского интерфейса, и от ориентированного графа к детерминированному конечному автомату, математическое описание которого дает возможность использовать его как каркас будущего прототипа пользовательского интерфейса. Рассмотренный в работе порядок перехода от описания пользовательского интерфейса встраиваемой системы в терминах аспектно-ориентированного подхода к описанию пользовательского интерфейса встраиваемой

системы в виде конечного автомата демонстрирует применимость аспектно-ориентированного подхода для разработки и прототипирования пользовательских интерфейсов встраиваемых систем.

Ключевые слова: аспектно-ориентированный подход, интерфейс, конечный автомат.

Введение

В аспектно-ориентированном подходе, предложенном широкой общественности в 1998 г. [1] как новая парадигма программирования, предлагается проводить декомпозицию предметной области по сквозной функциональности. В качестве примеров использования данного подхода часто приводятся задачи ведения журналов выполнения программы или контроля предоставления доступа к ресурсам, отмечая, что данный вариант декомпозиции позволяет уменьшить сложность реализации решаемой задачи. При рассмотрении пользовательских интерфейсов встраиваемых систем можно выделить сквозную функциональность, связанную с различной реакцией элементов пользовательского интерфейса в зависимости от состояния системы. В настоящей работе рассматривается возможность описания пользовательского интерфейса в терминах аспектно-ориентированного подхода целиком, с последующим переходом к конечному автомату, отслеживающему изменение режимов работы пользовательского интерфейса.

Описание пользовательского интерфейса в терминах аспектно-ориентированного подхода

Рассмотрим воздействие пользователя на элемент интерфейса как единицу сквозной функциональности a , а текущее состояние системы – как точку среза p . Для аспекта характерно наличие нескольких точек среза, т.е. какого-то непустого их множества $a = \{p_1, p_2, \dots, p_n\} = \{P\}$. При таком подходе поль-

зовательский интерфейс можно представить как сумму аспектов $A = \sum_{i=1}^m a_i$, множество срезов которой

может быть получено аналогичным способом: $P_{\text{системы}} = \{P_1 \cup P_2 \cup \dots \cup P_m\}$. Полное множество точек срезов $P_{\text{системы}}$ включает в себя подмножества $p_1^1, p_1^2, \dots, p_m^n$ точек срезов конкретных аспектов $a_1, a_2, \dots, a_m$ и позволяет получить полное описание состояний интерфейса встраиваемой системы. Если какой-либо аспект a_i не имеет в своем подмножестве $p_1, p_2, \dots, p_n$ точек среза точку из множества $P_{\text{системы}}$ точек срезов, его поведение в данной точке описывается идентично одной из имеющихся $p_1, p_2, \dots, p_n$ точек среза.

В качестве примера опишем как сумму аспектов A устройство индикации какой-либо физической величины с возможностью задания трех диапазонов (норма, предупреждение, авария). Описываемое устройство имеет три кнопки и четырехразрядный индикатор (рис. 1). Воздействие пользователя на систему возможно через нажатие одной из трех имеющихся кнопок. Описание аспектов, соответствующих нажатию кнопок, представлено в табл. 1. В зависимости от режима системы будет меняться и выводимая на индикатор величина. Описание аспекта «вывод» представлено в табл. 2.

Аспект		Точка среза (состояние системы / режим)	Поведение элемента
Нажатие кнопки «1»	p_1^1	Изменение порога предупреждения	Уменьшение значения порога предупреждения
	p_1^2	Изменение порога аварии	Уменьшение значения порога аварии
	p_1^3	Выбор для изменения порога предупреждения	Переход к p_{11}
	p_1^4	Выбор для изменения порога аварии	Переход к p_{12}
Нажатие кнопки «2»	p_2^1	Изменение порога предупреждения	Увеличение значения порога предупреждения
	p_2^2	Изменение порога аварии	Увеличение значения порога аварии
Нажатие кнопки «3»	p_3^1	Изменение порога предупреждения	Внесение изменений и переход к p_{35}
	p_3^2	Изменение порога аварии	Внесение изменений и переход к p_{35}
	p_3^3	Выбор для изменения порога предупреждения	Переход к p_{34}
	p_3^4	Выбор для изменения порога аварии	Переход к p_{35}
	p_3^5	Индикация значения измеренной величины	Переход к p_{33}

Таблица 1. Описание нажатий кнопок

Аспект		Точка среза (состояние системы / режим)	Поведение элемента
Вывод	p_4^1	Изменение порога предупреждения	Вывод порога предупреждения
	p_4^2	Изменение порога аварии	Вывод порога аварии
	p_4^3	Выбор для изменения порога предупреждения	Вывод порога предупреждения
	p_4^4	Выбор для изменения порога аварии	Вывод порога аварии
	p_4^5	Индикация значения измеренной величины	Вывод измеренной величины

Таблица 2. Описание работы индикатора

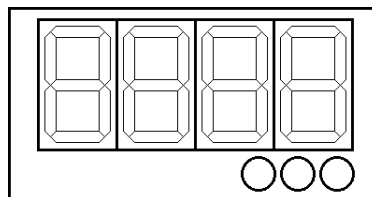


Рис. 1. Внешний вид устройства

В табл. 1, 2 индексами для элементов p_i^j служат i – номер аспекта; j – номер состояния по порядку. Для различных аспектов i состояния j с одинаковыми номерами тождественны. Получим множество срезов:

$$p_1^1 \cup p_1^2 \cup p_1^3 \cup p_1^4 \cup p_2^1 \cup p_2^2 \cup p_3^1 \cup p_3^2 \cup p_3^3 \cup p_3^4 \cup p_3^5 \cup p_4^1 \cup p_4^2 \cup p_4^3 \cup p_4^4 \cup p_4^5 = \\ \{p_1^1 \cup p_1^2 \cup p_1^3 \cup p_1^4 \cup p_3^5\} = P_{\text{системы}}.$$

Описание ориентированного графа

Полученное описание позволяет представить порядок изменения режимов интерфейса в виде орграфа. Формальное определение орграфа [2] выглядит как $D = (V, E)$, где E есть множество упорядоченных пар вершин $v \in V$. Так как интерфейс встраиваемой системы изменяет свое состояние (режим) в определенном порядке, то можно представить переходы между элементами множества $P_{\text{системы}}$ как матрицу смежности M с размерностью, определяемой кардинальным числом [3], $|P_{\text{системы}}|$. Используя матрицу смежности M , можно построить множество E , а множеством V будет являться $P_{\text{системы}}$.

Создадим матрицу смежности M , размерностью $|P_{\text{системы}}| = 5$ для нашего примера (табл. 3), в качестве ребер ориентированного графа будем использовать информацию о переходах между состояниями из табл. 1:

	p_1^1	p_1^2	p_1^3	p_1^4	p_3^5
p_1^1	–	–	1	–	–
p_1^2	–	–	–	1	–
p_1^3	–	–	–	–	1
p_1^4	–	–	1	–	–
p_3^5	1	1	–	1	–

Таблица 3. Матрица смежности M

Таким образом, рассматривая пользовательский интерфейс встраиваемой системы как сумму аспектов A , можно построить ориентированный граф состояний системы (рис. 2).

Описание конечного автомата

Опишем пользовательский интерфейс как детерминированный конечный автомат [4] с помощью диаграммы состояний (рис. 3):

- вершины – состояния конечного автомата, в данном примере они определены на множестве $P_{\text{системы}}$ как состояния системы p_i^j .
- ребра – переходы из одного состояния p_k^m в другое p_s^l , определены как матрица смежности M размерностью $|P_{\text{системы}}| = 5$;
- нагрузка – символы, при которых осуществляется данный переход, в случае если переход из состояния p_m в состояние p_k возможен при появлении одного из нескольких символов, то над дугой диаграммы (ветвью графа) должны быть написаны они все; нагрузка описывается множеством воздействий, осуществляемых пользователем, в данном примере это нажатия на кнопки, описываемые как аспекты $a1, a2, a3$, эти же символы будут использоваться для обозначения осуществления переходов.

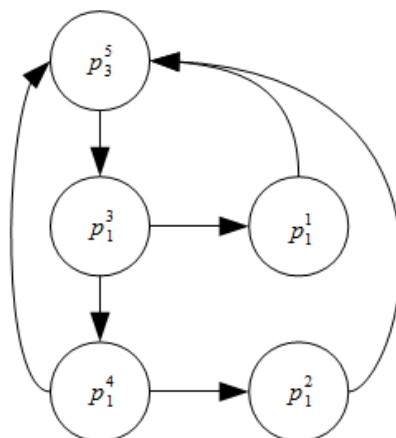


Рис. 2. Ориентированный граф

Полученное математическое описание конечного автомата дает возможность использовать его либо как каркас будущего прототипа пользовательского интерфейса, либо, при сохранении в каком-либо виде, как описание поведения пользовательского интерфейса для интерпретатора. Применение интерпретатора дает возможность использовать единый вид описания конечного автомата как для различных операционных систем, так и для различных микроконтроллеров, что позволяет применять эволюционное прототипирование, с каждой итерацией разработки приближая прототип к реальной системе.

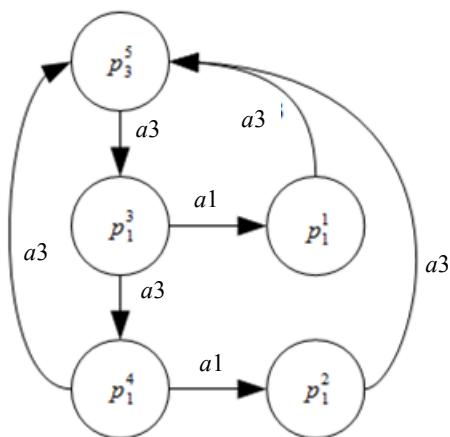


Рис. 3. Диаграмма состояний конечного автомата

Заключение

Рассмотренная в данной работе методика перехода от описания пользовательского интерфейса встраиваемой системы в терминах аспектно-ориентированного подхода к описанию пользовательского интерфейса встраиваемой системы в виде конечного автомата демонстрирует применимость данной методики для разработки и прототипирования пользовательских интерфейсов встраиваемых систем. Использование аспектно-ориентированного подхода в рамках встраиваемых систем, характеризующихся большим числом состояний, позволяет разработчику сконцентрироваться на описании поведения элементов пользовательского интерфейса в конкретных состояниях, скрывая от него механизм контроля состояний, что снижает трудоемкость реализации прототипа.

Литература

1. Kiczales G., Lamping J., Mendhekar A., etc. Aspect-oriented programming // Proceedings of the European Conference on Object-Oriented Programming (ECOOP). – Finland, Springer-Verlag LNCS 1241. – June 1997. – P. 220–242.
2. Кристофидес Н. Теория графов. Алгоритмический подход. – М.: Мир, 1978. – 432 с.
3. Куратовский К., Мостовский А. Теория множеств. – М.: Мир, 1970. – 416 с.
4. Хопкрофт Дж., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений. – М.: Вильямс, 2002. – 528 с.

Васюков Василий Михайлович – Балтийский государственный технический университет «Военмех» им. Д.Ф. Устинова, аспирант, witcher237@rambler.ru

УДК 621.391.037.372

ГИСТОГРАММНАЯ НОРМАЛИЗАЦИЯ РЕЧЕВЫХ ПРИЗНАКОВ В ЗАДАЧЕ ВЕРИФИКАЦИИ ДИКТОРОВ

Ю.Н. Матвеев, А.К. Шулипа

Содержится краткое описание алгоритма гистограммной нормализации речевых признаков применительно к задаче верификации дикторов. Приведены результаты верификационных тестов при различных параметрах и режимах нормализации. На основании полученных данных сделаны выводы об эффективности использования нормализации речевых признаков для улучшения качества верификации дикторов и найдены оптимальные условия использования алгоритма нормализации.

Ключевые слова: верификация дикторов, речевые признаки, гистограммная нормализация.

Введение

Как отмечено в работе [1], успех в распознавании личности по голосу (распознавании диктора) в гораздо большей степени зависит от метода компенсации канала, чем от выбора признаков. В качестве одного из путей решения проблемы канала в работе [2] было предложено использовать методы нормализации признаков путем их трансформирования (feature warping). Данный метод с 80-х годов прошлого века широко используется в обработке изображений [3–5], а для обработки речевых сигналов он получил распространение только в начале настоящего века [6].

Процедура нормализации речевых признаков выполняется на стадии предобработки речевого сигнала и заключается в приведении формы распределения признаков к заданному виду. В качестве заданного обычно используется нормальное (гауссово) распределение [6] и такую процедуру нормализации часто называют «гауссинизацией» [7]. Целью нормализации является уменьшение влияния вариативности внешних факторов на оценки параметров статистической модели голоса диктора и компенсация расхождения каналов (условий тестирования и обучения). Таким образом, система верификации дикторов, использующая нормализованные признаки становится более робастной.

В случае дискретных сигналов нормализация основывается на модификации гистограмм распределений и поэтому называется гистограммной [8].

Известные методы гистограммной нормализации

На практике нормализация может быть выполнена одновременно на всем произнесении или последовательно, при смещении скользящего окна по речевому сигналу. В первом случае каждый элемент входных данных преобразуется с учетом его ранга, который определяется на всей последовательности векторов признаков. Во втором варианте нормализации производится трансформирование элемента, находящегося в центре окна, согласно его рангу среди других элементов из интервала, ограниченного окном. Алгоритм гистограммной нормализации речевых признаков с использованием скользящего окна позволяет выделять сигнал в зоне сильной нелинейности каналов связи, телефонных трубок.

В качестве речевых признаков для систем верификации наиболее часто рассматриваются вектора, которые состоят из набора мэл-частотных кепстральных коэффициентов (mel frequency cepstral coefficients, MFCC), а также их производных. Алгоритм нормализации признаков, предложенный в [2], предполагает нормализацию только MFCC-коэффициентов, на основе которых затем вычисляются оставшиеся компоненты вектора признаков (первые и вторые производные). Как показали результаты экспериментов, при таком исполнении алгоритм нормализации работает неэффективно.

Целью настоящей работы является разработка модификации алгоритма гистограммной нормализации, устраняющей недостатки известных решений.

Описание алгоритма гистограммной нормализации

В алгоритме гистограммной нормализации производится трансформирование вектора признаков в центре скользящего окна размерности N . Пусть известен входной набор векторов признаков в пределах окна $\{x_1, x_2, \dots, x_N\}$. Все векторы во входном наборе имеют $\dim(x)$ компонент. Далее необходимо вычислить относительный ранг для каждого i -го ($i \in \{1, \dots, \dim(x)\}$) компонента вектора $x_{\frac{N}{2}}^{i/}$ в центре скользящего окна и найти для него соответствующее трансформированное значение $x_{\frac{N}{2}}^{i//}$. Скользящее окно последовательно сдвигается на вектор по всему произнесению и, таким образом, проводится трансформирование всех входных признаков, попавших в интервал $\left[\frac{N}{2} : 1 - \frac{N}{2}\right]$. Для произвольного положения скользящего окна ранг центрального вектора определяется согласно выражению

$$R\left(x_{\frac{N}{2}}^i\right) = \frac{1}{N} \sum_{k=1}^N l\left(x_k^i\right) - \frac{1}{2N}$$

при

$$l\left(x_k^i\right) = \begin{cases} 1, & x_k^i \leq x_{\frac{N}{2}}^i; \\ 0, & x_k^i > x_{\frac{N}{2}}^i, \end{cases}$$

где $i \in \{1, \dots, \dim(x)\}$ – индекс компоненты вектора признаков; $\dim(x)$ – число компонент вектора признаков x ; N – число векторов признаков в скользящем окне. Новое (трансформированное) значение признака из центра окна $x_{\frac{N}{2}}^{i//}$ находится по значению «старого» значения признака $x_{\frac{N}{2}}^{i/}$ из уравнения

$$R\left(x_{\frac{N}{2}}^{i/}\right) = Y\left(x_{\frac{N}{2}}^{i//}\right),$$

где

$$Y(x) = \int_{-\infty}^x \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt.$$

Область значений величины R ограничена интервалом $R \in \left[\frac{1}{2N} : 1 - \frac{1}{2N}\right]$.

На практике, чтобы при каждом смещении окна избежать численного решения интегрального уравнения

$$Y_j = \int_{-\infty}^{x_j} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt,$$

целесообразно предварительно рассчитать таблицу значений X_j для каждого Y_j следующим образом:

$$Y_j = \frac{1}{2N}(2j+1), \quad j \in [0 \dots N-1].$$

Точность вычисления интеграла для табличных значений Y_j должна быть не хуже $\frac{1}{2N}$.

Как было отмечено выше, нормализация выполняется только для векторов признаков, находящихся в центре скользящего окна. При этом часть векторов признаков, находящихся в интервалах длительностью $\frac{N}{2}$ в начале и конце произнесения, остается нетрансформированной. Для решения данной задачи необходимо рассчитать ранг этих векторов при начальном и конечном положении скользящего окна и выполнить для них трансформирование к нормальному распределению.

Рассмотренный алгоритм, аналогично известным алгоритмам нормализации кепстрального среднего и дисперсии [9], также дает набор центрированных кепстральных признаков MFCC с нулевым средним и единичной дисперсией.

Экспериментальные исследования предложенного метода

Эксперименты выполнялись для системы верификации дикторов на основе GMM-UBM-моделей голосов дикторов с использованием смесей гауссовых распределений (Gaussian mixture model, GMM) и универсальной фоновой модели (Universal Background Model, UBM) [1, 10]. Фонограммы для тестовых произнесений и эталонов выбирались из базы телефонных разговоров NIST-2008 [11]. Обучение проводилось на фонограммах телефонных баз NIST2005, NIST2006 [12, 13]. Все фонограммы, как для обучения, так и для тестирования, содержат речь только на английском языке. Условия обучения и тестирования при изменении параметров нормализации оставались постоянными. При обучении UBM использовались фонограммы 207 мужских и 275 женских голосов дикторов (3571 файлов). Множество обучения матрицы собственных каналов состояло из 190 мужских и 180 женских фонограмм (2583 файлов). Число компонент гауссовой смеси UBM $M = 512$, число собственных каналов $R = 50$. В качестве речевых признаков рассматривались MFCC-коэффициенты, их первая и вторая производные, объединенные в вектор размерности $L = 39$ (13mfcc + 13 delta mfcc + 13 delta delta mfcc). Выделение речевых сегментов проводилось с помощью детектора основного тона. Для построения GMM-эталона и компенсации влияния каналов использовалась схема Фогта [14]:

$$\mu = \mathbf{m} + \mathbf{D}\mathbf{z} + \mathbf{U}\mathbf{x},$$

$$\mu_0 = \mathbf{m} + \mathbf{D}\mathbf{z},$$

где $\mathbf{m}$ – супервектор (объединение векторов средних значений гауссоид смеси) GMM-UBM-размерности $P=M \cdot L$; $\mathbf{D}$ – диагональная матрица размерности $P \times P$; $\mathbf{U}$ – прямоугольная матрица собственных каналов размерности $R \times P$; μ – супервектор GMM эталона размерности P ; μ_0 – супервектор GMM эталона с учетом компенсации влияния эффектов канала размерности P ; $\mathbf{z}$ – вектор факторов внутридикторской вариативности размерности P ; $\mathbf{x}$ – вектор факторов вариативности, обусловленных эффектами канала размерности R .

Тестирование проводилось для фонограмм как с мужскими, так и с женскими голосами, причем пол дикторов на тестовых и эталонных фонограммах всегда совпадал. Число тестовых попыток для каждого типа сравнения представлено в табл. 1.

Тип сравнения	Женщины	Мужчины
Свой–свой	960	1274
Свой–чужой	13776	12672

Табл. 1. Число тестовых попыток при различных типах сравнения

Первоначально было исследовано влияние способов нормализации на качество верификации. Рассматривались два варианта нормализации: нормализация признаков путем вычитания кепстрального среднего (CMS) и нормализация трансформированием признаков (Feature Warping). При этом нормализация трансформированием признаков в одном случае выполнялась отдельно для каждой компоненты вектора признаков FW1, в другом – проводилась нормализация только для кепстральных коэффициентов вектора признаков и уже по ним рассчитывались производные первого и второго порядка FW2 [2]. Размер скользящего окна нормализации для FW1 и FW2 составлял 300 векторов. Результаты тестов верификации представлены в табл. 2. Для оценки качества верификации использовалось значение равновероятной ошибки (Equal Error Rate, EER) пропуска чужого и отклонения своего диктора.

	EER, %		
	CMS	FW1	FW2
Женщины	6,0	4,3	5,3
Мужчины	4,3	3,8	4,8

Табл. 2. Результаты верификации при использовании различных способов нормализации входных признаков

Далее было исследовано влияние размера скользящего окна нормализации на ошибку верификации дикторов. Результаты тестирования были получены при нормализации всех компонент вектора признаков как для MFCC-коэффициентов, так и для их производных (табл. 3).

Пол дикторов	Размер скользящего окна				
	100	200	300	400	500
Женщины	5,3	4,4	4,3	4,5	4,9
Мужчины	3,8	3,8	3,8	3,4	3,7

Табл. 3. Влияние размера скользящего окна нормализации на ошибку верификации дикторов (EER, %)

Заключение

В работе предложен алгоритм гистограммной нормализации кепстральных признаков со скользящим окном. Нормализация осуществляется независимо для всех компонент вектора кепстральных признаков – как для MFCC-коэффициентов, так и их первых и вторых производных.

Проведенные эксперименты показали, что использование предложенного алгоритма гистограммной нормализации существенно улучшает результаты верификации дикторов по сравнению с нормализацией методом вычитания кепстрального среднего. Это объясняется тем, что приведение формы распределения каждой компоненты вектора речевых признаков к заданному виду (нормальному распределению) позволяет устранить эффекты, связанные с рассогласованием условий обучения и тестирования.

Как показали эксперименты, нормализация отдельно каждой компоненты вектора кепстральных признаков является оптимальной и дает относительное снижение ошибки верификации более чем на 20% для фонограмм как с мужскими, так и с женскими голосами, при размере скользящего окна нормализации 400 векторов (4 с). Повышение робастности метода на более коротких произнесениях (1–4 с) является предметом дальнейших исследований.

Литература

1. Reynolds D., Rose R. Robust text-independent speaker identification using Gaussian mixture speaker models // IEEE Trans. Speech Audio Process. – 1995. – V. 3. – P. 72–83.
2. Pelecanos J., Sridharan S. Feature warping for robust speaker verification // Proc. A Speaker Odyssey. The Speaker Recognition Workshop, 2001. – P. 243–248.
3. Матвеев Ю.Н., Очин Е.Ф. Нелинейное преобразование видеосигнала на основе алгоритма скользящей эквализации гистограмм // Изв. вузов СССР. Радиоэлектроника. – 1985. – № 1. – С. 81–82.
4. Матвеев Ю.Н., Очин Е.Ф. Выполнение операции скользящего выравнивания гистограммы в матричном процессоре // Автометрия. – 1988. – № 1. – С. 14–17.
5. Кучеренко К.И., Матвеев Ю.Н., Очин Е.Ф. Устройство для скользящей эквализации гистограмм. Авт. свид. СССР, кл. G 06 F 15/36, 15/62. 1989.
6. Benesty J., Sondhi M.M., Huang Y. Springer Handbook of Speech Processing. – Springer, 2007. – P. 657–660.
7. Ramaswamy G.N., Gopinath R.A. Short-Time Gaussianization For Robust Speaker Verification // Acoustics, Speech, and Signal Processing. – 2002. – V. 1. – P. 681–684.
8. Матвеев Ю.Н., Очин Е.Ф. Структура устройства модификации гистограмм изображений // Тезисы докладов II Всесоюзной конференции «Методы и средства обработки сложной графической информации». – Горький: ГГУ, 1985. – 320 с.
9. Recent Advances in Robust Speech Recognition Technology / Javier Ramírez and Juan Manuel Górriz, eds. – Bentham Science Publishers, 2011. – 223 p.
10. Kenny P., Ouellet P., Dehak N., Gupta V., Dumouchel P. A Study of Inter-Speaker Variability in Speaker Verification // IEEE Trans. Audio Speech and Language Processing. – 2008. – V. 16. – № 5. – P. 980–988.
11. 2008 NIST Speaker Recognition Evaluation Test Set. [Электронный ресурс]. – Режим доступа: <http://www ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2011S08>, свободный. Яз. англ. (дата обращения 18.10.2012).
12. 2005 NIST Speaker Recognition Evaluation Training Data [Электронный ресурс]. – Режим доступа: <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2011S01>, свободный. Яз. англ. (дата обращения 18.10.2012).
13. 2006 NIST Speaker Recognition Evaluation Training Set [Электронный ресурс]. – Режим доступа: <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2011S09>, свободный. Яз. англ. (дата обращения 18.10.2012).
14. Vogt R., Sridharan S. Explicit Modelling of Session Variability for Speaker Verification // Computer Speech & Language, 2008. – V. 22. – № 1. – P. 17–38.

Матвеев Юрий Николаевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ООО «ЦРТ-инновации», доктор технических наук, профессор, главный научный сотрудник, matveev@speechpro.com

Шулина Андрей Константинович

– ООО «ЦРТ-инновации», научный сотрудник, shulipa@speechpro.com

УДК 621.391.037.372

СИСТЕМА ИДЕНТИФИКАЦИИ ВОЗРАСТНОЙ ГРУППЫ ГОВОРЯЩЕГО ПО ЗАПИСЯМ СПОНТАННОЙ РЕЧИ

К.К. Симончик

Предлагается использовать популярный в текстонезависимой идентификации диктора метод выделения i -векторов для решения задачи идентификации возрастной группы говорящего. Исследуется две реализации системы идентификации возрастной группы говорящего: предложен подход на базе машины опорных векторов, а также подход на основе линейной регрессионной модели. В обоих случаях была достигнута хорошая надежность детектирования возрастной группы диктора по записям фонограмм устной речи. Средний процент правильной идентификации возрастной группы диктора составил 61% и 65% соответственно на речевой базе NIST SRE 2008.

Ключевые слова: возраст, i -вектор, SVM, линейная регрессия.

Введение

В современном мире широко развиваются речевые технологии различного применения – синтеза, идентификации, распознавания речи. Последние две технологии позволяют выделить из речи информацию различного типа – «кто» говорит на фонограмме и «что» именно говорится. Однако по записи речи можно получить и иную информацию – пол диктора, его эмоциональное состояние и т.д. Работа посвящена проблеме определения возраста диктора по записям его устной речи.

Информация о возрасте диктора важна для применения в различных областях. Например, в зависимости от возраста говорящего могут применяться соответствующие акустические модели для распознавания речи. Контроль возраста ребенка позволит разграничить доступ к информационному контенту Интернета. Автоматизированная сегментация голосовых запросов в колл-центр по возрасту позволит сделать их обработку эффективнее и т.д.

На сегодняшний день наиболее популярными подходами в определении возрастной группы говорящего являются модели на основе использования смесей гауссовых распределений (СГР) и машин опорных векторов (support vector machine, SVM) [1–4], а также скрытых марковских моделей [4]. В настоящей работе для решения задачи идентификации возрастной группы предложено использовать широко применяемый в текстонезависимой идентификации по голосу подход, основанный на представлении модели речи диктора в виде многомерного вектора, который используется в качестве входных данных для классификатора возрастной группы. Проводится исследование двух методов классификации: с использованием SVM и на основе линейной регрессионной модели.

Описание схемы выделения признаков

Схема выделения i -векторов в качестве признаков подразумевает анализ всего речевого материала, присутствующего на фонограмме, и его последовательное преобразование в многомерный вектор (рис. 1). Более подробно об этом можно узнать в [6].

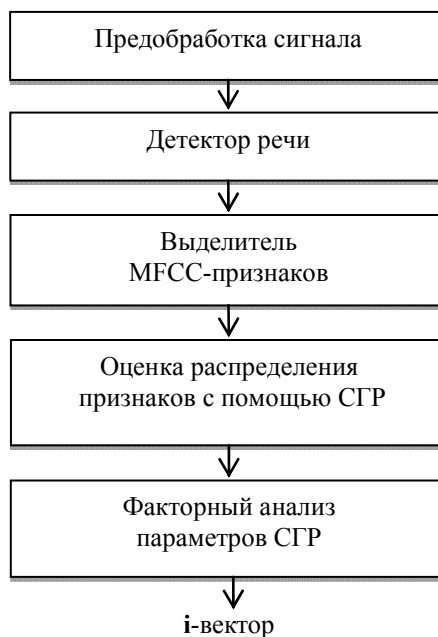


Рис. 1. Схема выделителя признаков

Перед модулем выделения речи использовались специальные алгоритмы предобработки всего сигнала [7–9]. Так как на реальных записях, сделанных в обычных офисных или бытовых условиях, часто присутствуют посторонние сигналы, такие как импульсные, мультитональные помехи, а также музыка и перегруженные участки речи, предварительная отбраковка непригодных для анализа участков фонограммы позволяет повысить эффективность дальнейшей обработки детектором речи.

Далее использовался детектор речи, основанный на оценке энергии сигнала в речевом диапазоне частот.

В качестве речевых признаков использовались мел-частотные кепстральные коэффициенты (Mel frequency cepstral coefficients, MFCC) размерностью 39 [10, 11]. Длина временного окна установлена в 20 мс с шагом 10 мс. Распределение MFCC-признаков моделировалось с помощью СГР [10]. В качестве универсальной фоновой модели (Universal background model, UBM) была использована СГР размерностью 512 смесей. База обучения UBM состояла из порядка 60000 фонограмм (более 4000 дикторов), взятых из речевых баз NIST SRE 98, 99, 2000, 2003, 2004, 2005, 2006, 2008 (мужской и женский гендеры). Далее на основе подхода, описанного в [1], производился факторный анализ параметров СГР с целью понижения размерности данных и выделялся скрытый $\mathbf{i}$ -вектор:

$$\boldsymbol{\mu} = \mathbf{m} + \mathbf{T}\boldsymbol{\omega} + \boldsymbol{\varepsilon},$$

где $\boldsymbol{\mu}$ – супервектор параметров СГР модели диктора; $\mathbf{m}$ – супервектор параметров UBM; $\mathbf{T}$ – матрица, задающая базис в редуцированном пространстве признаков; $\boldsymbol{\omega}$ – $\mathbf{i}$ -вектор в редуцированном пространстве признаков, $\boldsymbol{\omega} \in N(0,1)$; $\boldsymbol{\varepsilon}$ – вектор ошибки.

Использование классификатора на основе SVM

SVM является классификатором, впервые предложенным Вапником [11]. Идея SVM заключается в поиске оптимальной разделяющей гиперплоскости между двумя классами. Формула, задающая решающую функцию SVM, приведена ниже:

$$f(\mathbf{x}) = \sum_{i=1}^N w_i K(\mathbf{x}, \mathbf{x}_i) + w_0,$$

где N – число опорных векторов; $\mathbf{x}_i$ – i -й опорный вектор; $K(.,.)$ – ядро SVM. Так как SVM предназначен для разделения множества входных данных на 2 класса, а возрастных групп в исследуемой задаче было более двух, то была использована стратегия «турнир на выбывание» [12], которая предполагала обучение $M(M-1)/2$ классификаторов. Каждый SVM-классификатор был обучен на разделение только двух групп, а окончательное решение о принадлежности $\mathbf{i}$ -вектора к той или иной группе выносилось по анализу решений всех классификаторов.

Для принятия решения о принадлежности объекта возрастной группе использовалась следующая логика. На каждом шаге распознавание $\mathbf{i}$ -вектора осуществлял только один классификатор – «победившая» группа продолжала борьбу и определяла следующий используемый классификатор. Процесс осуществлялся до тех пор, пока не оставалась одна победившая группа, которая считалась результатом распознавания.

Таким образом, схема детектора возраста с использованием SVM-классификатора выглядела, как показано на рис. 2. Было выбрано линейное ядро SVM как достаточно простое и наиболее робастное.

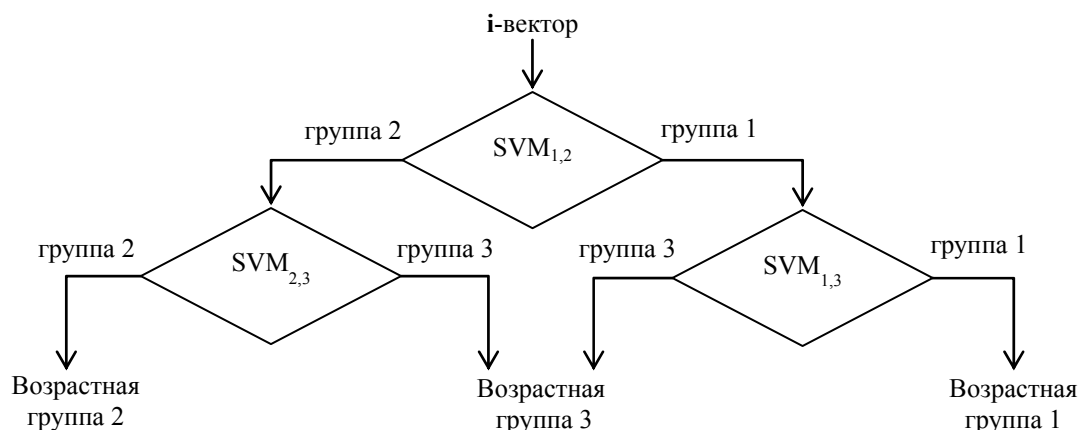


Рис. 2. Блок-схема детектора возраста на основе SVM

Классификатор на основе линейной регрессионной модели

Модель линейной регрессии позволяет определить линейную зависимость между скаляром y и несколькими наблюдаемыми переменными [14]. Линейная регрессия задается следующим уравнением:

$$y = \mathbf{A}\boldsymbol{\omega} + b,$$

где $\boldsymbol{\omega}$ – наблюдаемый вектор признаков (i -вектор); $\mathbf{A}$ – матрица обобщенной линейной модели; b – смещение; y – выходная переменная, связанная с возрастом диктора. Оптимизация параметров модели производилась с использованием минимизации функции квадратичного отклонения:

$$\sum_{i=1}^M (y - \hat{y})^2 \rightarrow \min,$$

где M – объем выборки. Далее производилось разбиение выходного скаляра y на возрастные классы с использованием простейшего порогового классификатора (рис. 3).



Рис. 3. Блок-схема детектора возраста на основе линейной регрессии

Эксперименты

Для экспериментов была взята речевая база NIST SRE 2008. Из нее были отобраны фонограммы тех дикторов, возраст которых можно было рассчитать. Оценка возраста производилась как разница между датой записи и годом рождения диктора. Всего было отобрано 16077 фонограмм 439 дикторов мужчин и 727 дикторов женщин возраста от 17 до 84 лет, говорящих на английском, китайском, арабском, русском и других языках (всего более 10 языков). Каналы связи были представлены микрофоном и телефоном. Речевая база была разбита на два равных множества, Train (обучающее) и Test (тестовое), таким образом, чтобы дикторы из разных множеств не пересекались.

Фонограммы дикторов были поделены на 3 возрастные группы:

1. от 16 до 25 лет – Young Adult (Молодые люди);
2. от 26 до 70 лет – Adult (Взрослые);
3. старше 70 лет – Senior (Пожилые).

На обучающем материале Train было рассчитано 3 SVM-классификатора для разделения попарно следующих возрастных групп: Young Adult/Adult, Young Adult/Senior и Adult/Senior. На множестве Train также был обучен второй классификатор, базирующийся на модели линейной регрессии. На рис. 4 показана зависимость выходного значения y от реального возраста диктора, записанного на фонограмме.

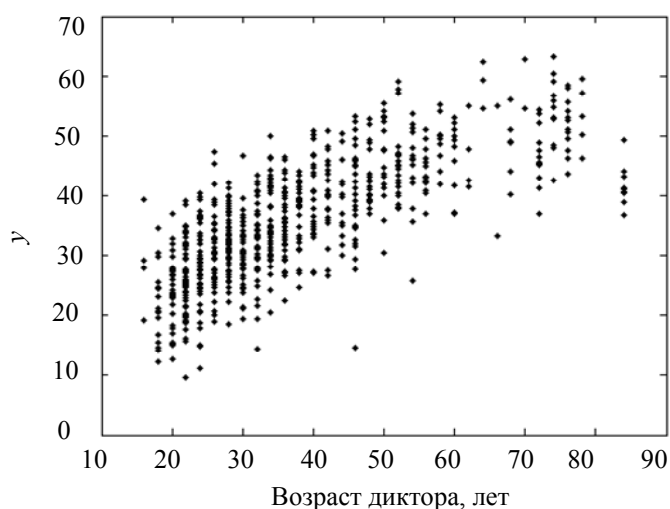


Рис. 4. Зависимость параметра y от возраста диктора на фонограмме

На рисунке видно, что существует достаточно высокая корреляция между параметром y регрессионной модели и возрастом диктора. Тем не менее, зависимость нельзя назвать строго линейной, что делает уместным использование индивидуальных пороговых значений y для каждой возрастной группы.

Далее для каждой из двух систем было проведено тестирование на множестве Test. В табл. 1, 2 показаны матрицы результатов классификации с помощью SVM-классификатора и классификатора на базе

линейной регрессионной модели. На рис. 5 показаны распределения выхода второго классификатора для различных возрастных групп. Стоит отметить, что распределения групп, не являющихся соседними, Young Adult и Senior, почти не пересекаются.

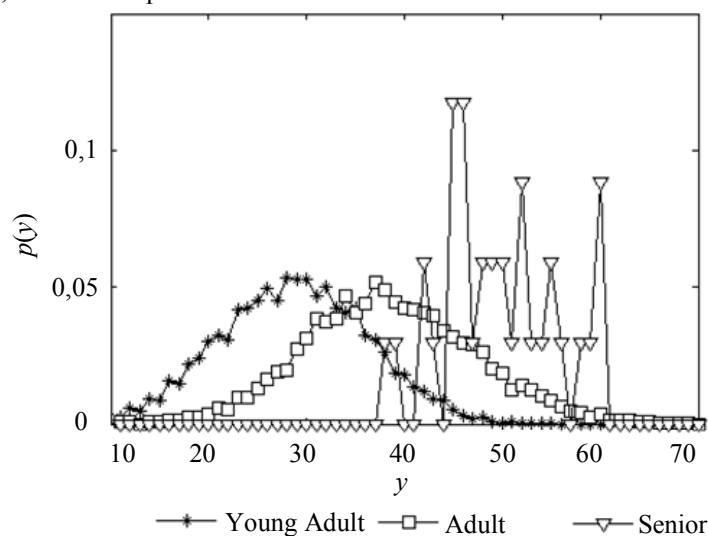


Рис. 5. Распределение параметра y регрессионной модели для каждой возрастной группы

Тест	Результаты классификации		
	Young Adult	Adult	Senior
Young Adult	72%	27%	1%
Adult	26%	72%	2%
Senior	0%	62%	38%

Таблица 1. Результаты экспериментов системы на основе SVM классификатора

Тест	Результаты классификации		
	Young Adult	Adult	Senior
Young Adult	64%	35%	1%
Adult	21%	58%	21%
Senior	0%	26%	74%

Таблица 2. Результаты экспериментов системы на основе линейной регрессионной модели

Заключение

В работе представлено два подхода к реализации детектора возраста говорящего на основе представления модели голоса в виде i -вектора.

В первую очередь необходимо заметить, что основная ошибка детектирования наблюдается только между смежными возрастными группами (Young Adult/Adult and Adult/Senior), тогда как ошибка между группами Young Adult и Senior не превышает 1%.

SVM-классификатор показывает хорошую надежность классификации для групп Young Adult и Adult, однако для группы Senior его точность в 2 раза ниже, чем у подхода на основе линейной регрессионной модели. Этот результат может быть объяснен недостаточным объемом обучающего множества группы Senior для SVM-классификатора. В отличие от SVM, линейная регрессионная модель демонстрирует более стабильный результат по всем возрастным группам и, в конечном итоге, показывает более высокую среднюю надежность детектирования возраста (65% против 61% у SVM).

Линейная регрессионная модель, обеспечивающая проекцию параметра y , одинаковую для всех групп, обеспечивает лучшую надежность и робастность идентификации возраста для набора данных, приведенных в данной работе.

В целом можно констатировать, что средняя надежность идентификации указанных возрастных групп по обоим представленным методам оказалась в 1,5 раза выше, чем у подхода, описанного в работе [4].

Литература

1. Feld M., Burkhardt F., Muller C. Automatic Speaker Age and Gender Recognition in the Car for Tailoring Dialog and Mobile Services // INTERSPEECH-2010. – 2010. – P. 2834–2837.

2. Bocklet T., Maier A., Bauer J.G., Burkhardt F., Noth E. Age and Gender Recognition for Telephone Applications Based on GMM Supervectors and Support Vector Machines // Proceedings of ICASSP 2008. – Las Vegas, 2008. – P. 1605–1608.
3. Dobry G., Hecht R.M., Avigal M., Zigel Y. Dimension Reduction Approaches for SVM based Speaker Age Estimation // INTERSPEECH-2009. – Brighton, UK, 2009. – P. 2031–2034.
4. Porat R., Lange D., Zigel Y. Age recognition based on speech signals using weights supervector // INTERSPEECH-2010. – 2010. – P. 2814–2817.
5. Metze F., Ajmera J., Englert R., Bub U., Burkhardt F., Stegmann J., Müller C., Huber R., Andrassy B., Bauer J.G., Little B. Comparison of Four Approaches to Age and Gender Recognition for Telephone Applications // Proceedings of ICASSP 2007. – Honolulu, 2007. – V. 4. – P. 1089–1092.
6. Dehak N., Dehak R., Kenny P., Brummer N., Ouellet P., Dumouchel P. Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification // INTERSPEECH-2009. – Brighton, UK, 2009. – P. 1559–1562.
7. Алейник С.В., Матвеев Ю.Н., Раев А.Н. Метод оценки уровня клиппирования речевого сигнала // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 79–83.
8. Козлов А.В., Лоханова А.И., Симончик К.К. Алгоритм детектирования музыкальных фрагментов в задачах речевой обработки // Научно-технические ведомости СПбГПУ. – 2010. – № 4 (103). – С. 7–11.
9. Симончик К.К., Галинина О.С., Капустин А.И. Алгоритм обнаружения речевой активности на основе статистик основного тона в задаче распознавания диктора // Научно-технические ведомости СПбГПУ. – 2010. – № 4 (103). – С. 18–23.
10. Белых И.Н., Капустин А.И., Козлов А.В., Лоханова А.И., Матвеев Ю.Н., Пеховский Т.С., Симончик К.К., Шулипа А.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // Информатика и ее применения. – 2012. – Т. 6. – Вып. 1. – С. 24–31.
11. Spiegel W., Stemmer G., Lasarczyk E., Kolhatkar V., Cassidy A., Potard B., Shum S., Song Y.C., Xu P., Beyelerlein P., Harnsberger J., Nöthm E. Analyzing Features for Automatic Age Estimation on Cross-Sectional Data // INTERSPEECH-2009. – Brighton, UK. – 2009. – P. 2923–2926.
12. Vapnik V.N. Statistical Learning Theory. – New York: Wiley, 1998. – 732 p.
13. Platt J., Cristianini N., Shawe-Taylor J. Large Margin DAGS for Multiclass Classification // Advances in Neural Information Processing Systems. – MIT Press, 2000. – P. 547–553.
14. Rawlings J., Pantula S. Dickey D. eds. Applied Regression Analysis. – Springer Texts in Statistics, 1998. – 736 p.

Симончик Константин – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ООО «ЦРТ-инновации», кандидат технических наук, доцент, руководитель отдела, simonchik@speechpro.com

УДК 004.021

МЕТОД СБОРКИ КОНТИГОВ ГЕНОМНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ СОВМЕСТНОГО ПРИМЕНЕНИЯ ГРАФОВ ДЕ БРЮИНА И ГРАФОВ ПЕРЕКРЫТИЙ

А.В. Александров, С.В. Казаков, С.В. Мельников, А.А. Сергушичев, Ф.Н. Царев

Предлагается метод сборки контигов геномных последовательностей. Особенностью этого метода является разбиение процесса сборки контигов на два этапа – сборка квазиконтигов из чтений и сборка контигов из квазиконтигов. На первом из этапов используется граф де Брюина, на втором – граф перекрытий. Описываются результаты экспериментального исследования разработанного метода на чтениях генома рыбы *Maylandia zebra*, размер генома которой составляет примерно миллиард нуклеотидов. Преимущество разработанного метода состоит в том, что для его работы требуется существенно меньше оперативной памяти по сравнению с существующими программными средствами для сборки генома.

Ключевые слова: сборка генома, контиги, граф де Брюина, граф перекрытий.

Введение

Многие современные задачи биологии и медицины требуют знания геномов живых организмов, которые состоят из нескольких нуклеотидных последовательностей молекул дезоксирибонуклеиновой кислоты (ДНК). В связи с этим возникает необходимость в дешевом и быстром методе определения последовательности нуклеотидов в образце ДНК. Существующие устройства для чтения ДНК не позволяют считать за один раз всю молекулу ДНК. Вместо этого они позволяют читать фрагменты генома небольшой длины. Длина фрагмента может быть разной, она является важным параметром секвенирования – от нее напрямую зависит стоимость секвенирования и время, затрачиваемое на чтение одного фрагмента: чем больше длина считываемого фрагмента, тем выше стоимость чтения и тем дольше это чтение проис-

ходит. В связи с этим в настоящее время распространение получил следующий дешевый и эффективный подход: сначала выделяется случайно расположенный в геноме фрагмент длиной около 500 нуклеотидов, а затем считываются его префикс и суффикс (длиной порядка 80–120 нуклеотидов каждый). Эти префикс и суффикс называются *парными чтениями*. Описанный процесс повторяется такое число раз, чтобы обеспечить достаточно большое покрытие генома чтениями. Указанным образом работают, например, секвенаторы компании Illumina [1].

Отметим, что описанные выше префикс и суффикс читаются с разных нитей ДНК: один – с прямой, другой – с обратно-комплементарной, причем неизвестно, какой откуда. По этой причине удобно рассматривать геном и чтения, дополненные своими обратно-комплементарными копиями.

Задачей сборки генома является восстановление последовательности ДНК (ее длина составляет от миллионов до миллиардов нуклеотидов у разных живых существ) на основании информации, полученной в результате секвенирования. Этот процесс делится, как правило, на следующие этапы:

1. исправление ошибок в данных секвенирования;
2. сборка квазиконтигов – фрагментов, префиксы и суффиксы которых были получены на этапе секвенирования;
3. сборка контигов – максимальных непрерывных последовательностей нуклеотидов, которые удалось восстановить;
4. построение скэффолдов – последовательностей контигов, разделенных промежутками, для длин которых известны верхние и нижние оценки.

Одной из наиболее часто используемых при сборке генома математических моделей является граф де Брюина [2]. На его использовании основаны следующие программные средства сборки генома: Velvet [3], ALLPATHS [4], ABySS [5], SOAPdenovo [6], EULER [7].

Одним из недостатков, которым обладают перечисленные программные средства, является большой объем оперативной памяти, необходимый для сборки генома размером в миллиард нуклеотидов. Так, например, SOAPdenovo необходимо порядка 140 Гб оперативной памяти, а ABySS – 21 компьютер с 16 Гб оперативной памяти каждый (всего – 336 Гб). Такие затраты памяти обусловлены наличием ошибок секвенирования в исходных данных, которые ведут к увеличению размера графа де Брюина, а также неоптимальным методом хранения этого графа.

Целью настоящей работы является разработка алгоритма сборки контигов геномной последовательности, использующего меньший объем оперативной памяти по сравнению с существующими. Входные данные для алгоритма представляют собой набор парных чтений, полученных на секвенаторе, а выходные данные – набор контигов.

Для исправления ошибок в данных секвенирования в настоящей работе используется алгоритм, описанный в работе [8]. Метод сборки контигов базируется на методе, предложенном в работе [9], для сборки бактериальных геномов размером в несколько миллионов нуклеотидов. Отличие предлагаемого метода от известного состоит в том, что предлагаемый метод рассчитан на сборку геномов в несколько миллиардов нуклеотидов. Построение скэффолдов в настоящей работе не рассматривается.

Архитектура метода сборки контигов

Сборка контигов в предлагаемом методе выполняется в два этапа.

1. Сборка квазиконтигов из чтений геномной последовательности.
2. Сборка контигов из квазиконтигов. Осуществляется с использованием графа перекрытий и метода Overlap–Layout–Consensus [10].

Сборка квазиконтигов из чтений геномной последовательности. Для сборки квазиконтигов осуществляется построение графа де Брюина, в котором множество ребер состоит только из «надежных» $(k+1)$ -меров – тех, которые встречаются в чтениях достаточно большое число раз, не меньшее некоторого порогового значения, для того чтобы их можно было с очень большой вероятностью считать входящими в геном. Рассмотрим работу алгоритма на примере 10 пар чтений генома из 25 символов (рис. 1).

Если взять порог частоты для вхождения в граф равным единице, а $k=3$, то получится граф де Брюина, изображенный на рис. 2 (для простоты обратно-комплементарные ребра на рисунках не показаны).

Одним из важных свойств графа де Брюина является наличие в нем пути, соответствующего геному, при условии достаточного покрытия чтениями. В частности, это означает наличие пути для каждого из фрагментов, из которых были получены парные чтения. Предлагаемый метод сборки квазиконтигов основан на поиске таких путей.

Из всех путей, начало и конец которых совпадают с парой чтений, вызывают интерес только те, которые укладываются в априорные границы длин фрагментов, поэтому слишком короткие и слишком длинные пути можно отбросить. Оставшиеся пути – «хорошие» кандидаты на роль пути, соответствующего фрагменту в действительности. Если такой путь (рис. 3) – единственный, то можно с очень большой уверенностью сказать, что он соответствует реальной подстроке геномной последовательности, поэтому этот фрагмент считается восстановленным, а найденный путь выводится как квазиконтиг.

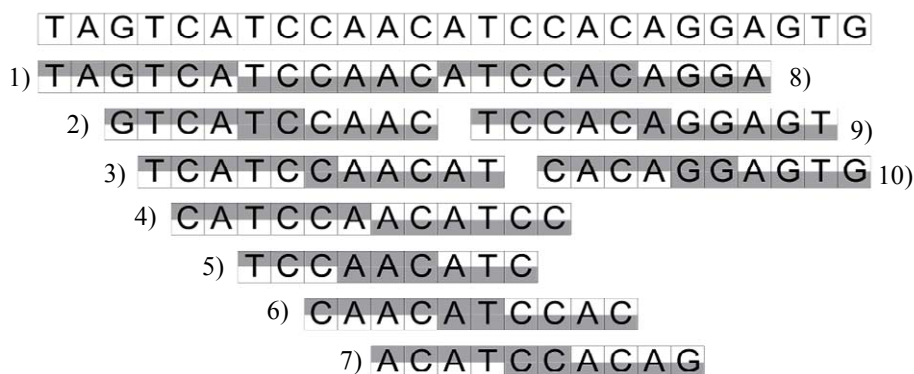


Рис. 1. Геном и его парные чтения

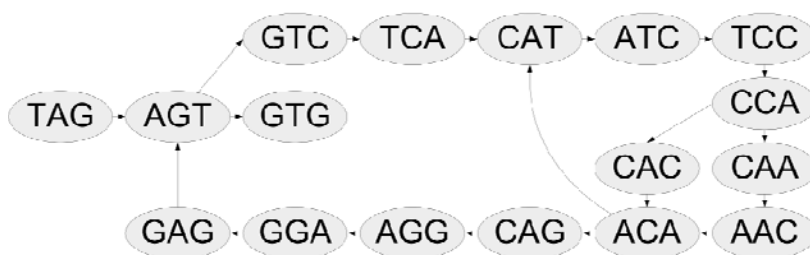


Рис. 2. Граф де Брюина при $k=3$

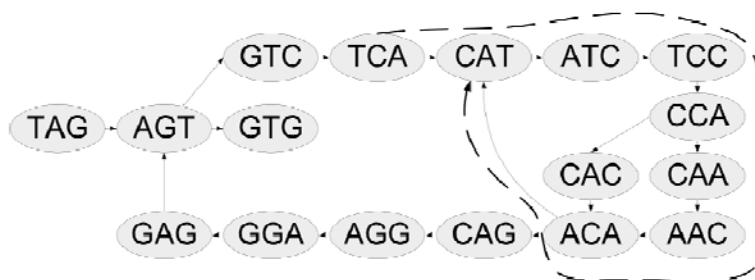


Рис. 3. Путь в графе де Брюина, соответствующий паре чтений номер 3

Если паре чтений в графе де Брюина соответствуют несколько путей (рис. 4), то из такой пары чтений квазиконтинг не генерируется.

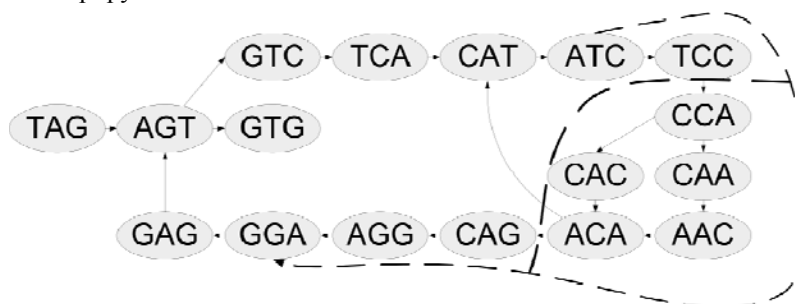


Рис. 4. Пути в графе де Брюина, соответствующие паре чтений номер 8

Приведем формальное описание алгоритма. Пусть путь p_1 длины l_1 ведет из вершины v_1 в вершину v_2 , а путь p_2 длины l_2 ведет из вершины v_2 в вершину v_3 . Будем обозначать конкатенацию этих путей, т.е. путь длины $l_1 + l_2$, соединяющий вершины v_1 и v_3 , проходящий сначала по пути p_1 , а затем – по p_2 , как $p_1 \cdot p_2$. Рассмотрим множества путей P_1 и P_2 .

С помощью $P_1 \cdot P_2$ будем обозначать все пути, которые можно получить конкатенацией путей p_1 и p_2 из P_1 и P_2 соответственно, т.е. $P_1 \cdot P_2 = \{p_1 \cdot p_2 \mid p_1 \in P_1, p_2 \in P_2\}$. Заметим, что конкатенировать можно только те пары путей p_1 и p_2 , у которых последняя вершина p_1 совпадает с первой вершиной p_2 . Например, если P_1 состоит из одного пути $v_1 \rightarrow v_2$, а множество P_2 состоит из путей $v_2 \rightarrow v_3$ и $v_4 \rightarrow v_5$, то множество $P_1 \cdot P_2$ будет состоять из одного пути $v_1 \rightarrow v_3$.

Задача, решаемая алгоритмом, состоит в поиске всех путей, соединяющих две заданные вершины v_1 и v_2 , длины которых лежат в промежутке $[l_{\min}; l_{\max}]$. Будем обозначать множество всех путей из v_1 в v_2 длины l как P^l , тогда искомое множество всех путей из v_1 в v_2 будет получаться объединением множеств

$$P^l: P = \bigcup_{l=l_{\min}}^{l_{\max}} P^l.$$

Для выявления таких путей будем применять двунаправленный поиск [11] – одновременный поиск путей, ведущих из первой вершины, и путей, ведущих во вторую вершину. Это позволяет сократить время работы с $O(d^{l_{\max}})$ до $O(d^{l_{\max}/2})$, где d – средняя исходящая степень вершины графа (для графов де Брюина, встречающихся на практике, эта величина несущественно превышает единицу).

Применимость двунаправленного поиска объясняется тем, что любой путь p длины l из v_1 в v_2 можно разбить на два более коротких пути p_1 и p_2 длиной l_1 и l_2 так, чтобы выполнялись равенства $p = p_1 p_2$, $l = l_1 + l_2$.

Для реализации двунаправленного поиска параллельно запустим два обхода в ширину: из вершины v_1 по прямым ребрам и из v_2 – по обратным. Тогда на каждом шаге l можно поддерживать следующий инвариант: для первой вершины будем хранить множество $P_1^{l_1}$ всех исходящих из нее путей длины l_1 , а для второй – множество $P_2^{l_2}$ всех входящих путей длины l_2 , причем $l_1 + l_2 = l$. Таким образом, на l -ом шаге можно получить все пути длины l из v_1 в v_2 путем конкатенацией путей из множеств $P_1^{l_1}$ и $P_2^{l_2}$: $P^l = P_1^{l_1} \cdot P_2^{l_2}$.

На начальном шаге этого алгоритма l , l_1 и l_2 равны нулю, а P_1^0 и P_2^0 содержат по одному пути нулевой длины (эту пути состоят соответственно из вершин v_1 и v_2).

Если E – это множество всех ребер графа, то шаг в первом обходе осуществляется по формуле $P_1^{l_1+1} = P_1^{l_1} \cdot E$, а шаг во втором обходе по формуле $P_2^{l_2+1} = E \cdot P_2^{l_2}$.

Для того чтобы сократить потребление памяти при применении предлагаемого метода, необходимо иметь компактное представление используемого подграфа графа де Брюина. Для этого достаточно хранить только множество его ребер, что можно эффективно делать, используя, например, хеш-таблицу с открытой адресацией [12]. Преимуществами такого подхода хранения перед другими являются его простота в реализации и достаточно высокое быстродействие.

Также для уменьшения требуемого объема оперативной памяти используется то, что каждый $(k+1)$ -мер входит в граф вместе с обратным-комплементарным. Тогда вместо пары $(k+1)$ -меров s и s^c можно хранить только один, определяемый по некоторому правилу – например, таким правилом может быть выбор лексикографически минимального $(k+1)$ -мера. В этом случае необходимый объем памяти уменьшается примерно в 2 раза для четных k и ровно в 2 раза – для нечетных (только для четных k существуют обратные-комплементарные себе $(k+1)$ -меры).

Сборка контигов из квазиконтигов. Сборка контигов из квазиконтигов основана на подходе Overlap-Layout-Consensus и состоит из нескольких этапов:

1. построение графа перекрытий между квазиконтигами;
2. уточнение графа перекрытий;
3. поиск контигов в графе перекрытий.

Для поиска перекрытий построим строку вида $C_1 C_2 C_3 \dots C_n$, где C_i – i -й квазиконтиг, а n – число квазиконтигов. После этого необходимо построить суффиксный массив [13] для этой строки – отсортированный массив всех суффиксов строки. С его помощью можно найти все квазиконтиги, в которых встречается заданная подстрока. Это можно сделать, например, с помощью бинарного поиска суффиксов в суффиксном массиве, которые начинаются с заданной подстроки. Они будут располагаться рядом за счет сортировки.

Зафиксируем квазиконтиг, все перекрытия с которым требуется найти. Будем рассматривать его префиксы в порядке увеличения длины, начиная с минимального порога. Для каждого префикса будем проверять, входит ли такая подстрока с добавленным в конце \$ в суффиксный массив. Для того чтобы учесть еще и неточные перекрытия, проверяются не только сами префиксы, но и префиксы, в которые внесены небольшие изменения.

На следующем этапе происходят анализ найденных перекрытий, а также добавление и удаление перекрытий. Добавление происходит в случае, если квазиконтиг A перекрывается с квазиконтигами B и C , причем квазиконтиги B и C должны перекрываться, но такое перекрытие найдено не было. Удаление происходит, если квазиконтиг A перекрывается с квазиконтигом B , но B не похож на большинство квазиконтигов, с которыми перекрывается A .

Для поиска контигов выполняется поиск в ширину [12] в графе перекрытий. Он прерывается, если после текущего квазиконтига нет консенсуса – это означает, что квазиконтиги, перекрывающиеся с ним, различаются в большом числе позиций.

Экспериментальное исследование

Экспериментальные исследования разработанного метода проводились в рамках проекта Assemblathon 2, организованного университетом Калифорнии Санта-Круз [14]. Одним из наборов данных, который был подготовлен организаторами, являлся набор чтений рыбы *Maylandia zebra*. Размер генома этой рыбы оценивается примерно в 1 млрд нуклеотидов. Чтения геномной последовательности были разбиты на несколько библиотек, характеристика которых приведена в таблице.

Средний размер фрагмента в библиотеке	Покрытие
180	60
2500	62
5000	14
7000	16
9000	15
11000	12
40000	2,5

Таблица. Характеристика чтений генома рыбы *Maylandia zebra*

Общий объем исходных данных составлял 140 Гб. Для сборки контигов использовалась только одна из библиотек чтений – со средним размером фрагмента в 180 нуклеотидов и 60-кратным покрытием. Алгоритмы сборки генома были реализованы на языке программирования Java. Для запуска программ использовался компьютер с 32 Гб оперативной памяти и двумя 4-ядерными процессорами. Суммарное время работы всех трех этапов – исправления ошибок, сборки квазиконтигов и сборки контигов – составило 5 суток. Опишем подробнее результаты каждого из этапов.

Перед исправлением ошибок чтения были обрезаны таким образом, чтобы вероятность отдельной ошибки в каждом нуклеотиде не превышала 10%. После этого длина всех чтений в среднем уменьшилась на 20%. Исправление ошибок работало в течение 42 ч. В результате было найдено 150 млн исправлений. Всего было 600 млн. чтений, поэтому было исправлено в среднем каждое четвертое чтение.

Сборка квазиконтигов заняла 38 ч. Квазиконтиги были получены из 60% парных чтений.

Сборка контигов выполнялась за 26 ч. В результате было получено 734165 контигов, суммарный размер которых составляет 680×10^6 нуклеотидов. Длина максимального контига составляет 23514 нуклеотидов, средняя длина – 927, значение метрики N50 – 1799.

Отметим, что при использовании программного средства SOAPdenovo необходимый для сборки этого генома объем оперативной памяти составил 140 Гб, а при использовании программного средства ABySS – 336 Гб. Это позволяет говорить о том, что у разработанного метода требования к объему оперативной памяти существенно меньше.

Заключение

Предложен метод сборки контигов геномных последовательностей, основанный на совместном использовании графа де Брюина и графа перекрытий. Экспериментальное исследование этого метода проведено в рамках проекта Assemblathon 2. Это исследование показало, что разработанный метод обладает существенно меньшими требованиями к объему оперативной памяти, чем существующие.

Исследования выполнялись в рамках Федеральных целевых программ «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» (государственный контракт № 07.514.11.4010) и «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы» (государственный контракт № 16.740.11.0495).

Литература

1. Illumina, Inc. [Электронный ресурс]. – Режим доступа: <http://www.illumina.com/>, свободный. Яз. англ. (дата обращения 21.03.2012).
2. Pevzner P.A. 1-Tuple DNA sequencing: computer analysis // J. Biomol. Struct. Dyn. – 1989. – V. 7. – P. 63–73.
3. Zerbino D.R., Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs // Genome Research. – 2008. – V. 18. – P. 821–829.
4. Butler J., MacCallum I., Kleber M., Shlyakhter I.A., Belmonte M.K., Lander E.S., Nusbaum C., Jaffe D.B. ALLPATHS: De novo assembly of wholegenome shotgun microreads // Genome Research. – 2008. – V. 18. – P. 810–820.
5. Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J., Birol I. ABySS: A parallel assembler for short read sequence data // Genome Research. – 2009. – V. 19. – P. 1117–1123.

6. Li R., Zhu H., Ruan J., Qian W., Fang X., Shi Z., Li Y., Li S., Shan G., Kristiansen K. et al. De novo assembly of human genomes with massively parallel short read sequencing // *Genome Research*. – 2010. – V. 20. – P. 265–272.
7. Pevzner P.A., Tang H., Waterman M.S. EULER: An Eulerian path approach to DNA fragment assembly // *Proc. Natl. Acad. Sci.* – 2001. – V. 98. – P. 9748–9753.
8. Александров А.В., Казаков С.В., Мельников С.В., Сергушичев А.А., Царев Ф.Н., Шалыто А.А. Метод исправления ошибок в наборе чтений нуклеотидной последовательности // *Научно-технический вестник СПбГУ ИТМО*. – 2011. – № 5 (75). – С. 81–84.
9. Исенбаев В.В. Разработка системы секвенирования ДНК с использованием paired-end данных. Бакалаврская работа. СПбГУ ИТМО, 2010 [Электронный ресурс]. – Режим доступа: http://is.ifmo.ru/genom/_isenbaev_thesis.pdf, свободный. Яз. англ. (дата обращения 21.03.2012).
10. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome // *Nature*. – 2001. – V. 409. – № 6822. – P. 860–921.
11. Рассел С., Норвиг П. Искусственный интеллект: современный подход: Пер. с англ. – 2-е изд. – М.: Вильямс. 2006. – 1408 с.
12. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. – М.: Вильямс, 2011. – 1296 с.
13. Гасфилд Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология. – СПб: Невский диалект, 2003. – 654 с.
14. Проект Assemblathon 2 [Электронный ресурс]. – Режим доступа: www.assemblathon.org, свободный. Яз. англ. (дата обращения 21.03.2012).

<i>Александров Антон Вячеславович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, alantbox@gmail.com
<i>Казаков Сергей Владимирович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, kazakov_sergey_v@mail.ru
<i>Мельников Сергей Вячеславович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, melnikov@rain.ifmo.ru
<i>Сергушичев Алексей Александрович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, alsergbox@gmail.com
<i>Царев Федор Николаевич</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, fedor.tsarev@gmail.com

УДК 539.264; 535.372

ИССЛЕДОВАНИЕ ИТТРИЕВЫХ СВИНЦОВО-ФТОРИДНЫХ НАНОСТЕКЛОКЕРАМИК, АКТИВИРОВАННЫХ ИОНАМИ НЕОДИМА

А.Ю. Бибик, А.О. Трофимов, Р.К. Нуриев, В.А. Асеев, Е.В. Колобкова, Н.В. Никоноров

Исследованы физико-химические и люминесцентные свойства иттриевых свинцово-фторидных стекол, активированных неодимом, и наностеклокерамик на их основе. Изучено выделение кристаллических фаз после термообработки стекол, рассчитаны размеры кристаллов и определены параметры элементарной ячейки.

Ключевые слова: активированная стеклокерамика, рентгенофазовый анализ, кристаллическая фаза, люминесценция.

Введение

Одним из перспективных материалов является наностеклокерамика, представляющая собой стекло с распределенными в его объеме нанокристаллами. Особый интерес представляет разработка и исследование стеклокристаллических материалов на основе фторсиликатных стекол, активированных редкоземельными ионами. Такие материалы сочетают в себе лучшие свойства низкофононных фторидных кристаллов, а также имеют высокую механическую прочность и химическую устойчивость [1]. Наностеклокерамики, активированные ионами неодима, представляют большой интерес в качестве лазерных материалов, работающих в ближнем ИК-диапазоне (1,06 и 1,3 мкм) на переходах ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$ и ${}^4F_{3/2} \rightarrow {}^4I_{13/2}$ соответственно. Настоящая работа посвящена исследованию структурных и люминесцентных свойств иттрий-свинцово-оксифторидных наностеклокерамик, активированных ионами неодима.

Экспериментальная часть

В работе были исследованы стекла состава $30\text{SiO}_2\text{-}15\text{AlO}_{3/2}\text{-}29\text{CdF}_2\text{-}18\text{PbF}_2\text{-}5\text{ZnF}_2\text{-}x\text{NdF}_3\text{-(}3\text{-}x\text{)YF}_3$, где $x=3,0; 2,9; 2,5; 1; 0,5; 0,2; 0,1; 0$, а также наностеклокерамики на их основе [2]. Синтез проводился в течение 30 мин. при температуре $T=1050^\circ\text{C}$ в открытых корундовых тиглях в атмосфере воздуха. Далее стекла подвергали термической обработке в течение $t_{\text{терм.}}=30, 60, 120$ мин. при температуре начала кристаллизации $T_{\text{н.к.}}=500^\circ\text{C}$ для получения наностеклокерамик. Температура начала кристаллизации определялась методом дифференциальной сканирующей калориметрии. На рис. 1, а, представлена рентгенограмма исходного и термообработанного стекла.

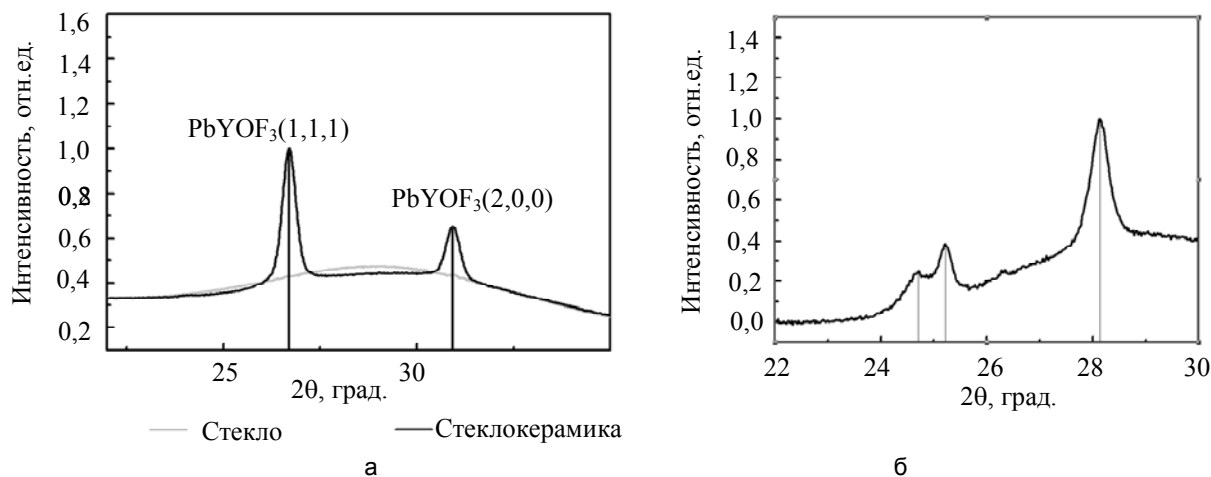


Рис. 1. Дифрактограммы стекла и наностеклокерамики, термообработанной в течение 60 мин, концентрация NdF_3 – 0 мол. % (а); концентрация NdF_3 – 3 мол. % (б)

В исходном стекле кристаллические фазы отсутствуют, а в результате вторичной термообработки стекла с 3 мол. % YF_3 образуется кристаллическая фаза, в кристаллографическом отношении соответствующая иттриевому оксифториду свинца PbYOF_3 [2]. Выделяющаяся фаза – флюоритоподобная кубическая гранецентрированная элементарная ячейка с размером $5,74 \text{ \AA}$.

При концентрации ионов неодима 3 мол. %, т.е. полном замещении ионов иттрия, формируется гексагональная кристаллическая фаза. В кристаллографическом отношении структура выделяющейся фазы соответствует кристаллу NdF_3 (рис. 1, б). Размер элементарной ячейки равен $5,84 \text{ \AA}$.

В ходе исследования был проведен рентгенофазовый анализ стекол и наностеклокерамик на их основе. На основе интенсивности и полуширины дифракционных отражений был сделан вывод, что процесс объемной кристаллизации в неодимовых наностеклокерамиках завершается после двух часов термообработки. Для всех исследованных концентраций фторида неодима объем кристаллической фазы в результате термообработки не зависит от концентрации ионов-активаторов.

При концентрациях фторида неодима от 0,1 до 2,9 мол.% в результате термообработки формируется кубическая гранецентрированная элементарная ячейка $\text{PbY}_{(1-x)}\text{Nd}_x\text{OF}_3$. Это связано с тем, что ионы неодима встраиваются в кристалл, замещая иттрий, поскольку иттрий и неодим имеют близкие по значению размеры ионного радиуса. Размер элементарной ячейки зависит от концентрации фторида неодима и колеблется в диапазоне 5,74–5,83 Å.

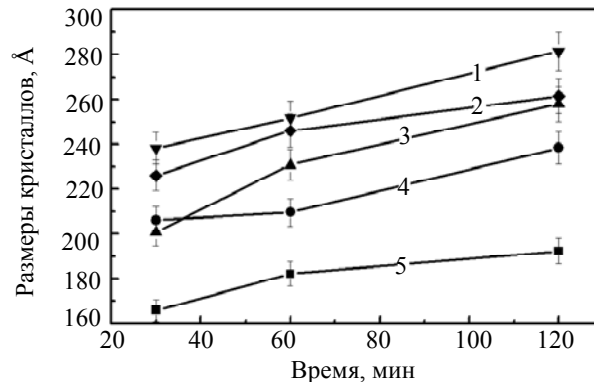


Рис. 2. Зависимость размера кристаллов от времени и концентрации NdF_3 для стеклокерамик, термообработанных в течение 30, 60 и 120 мин. NdF_3 мол. %: 0,2 (кривая 1); 0,5 (кривая 2); 1,0 (кривая 3); 2,5 (кривая 4); 3 (кривая 5)

Размеры кристаллов рассчитывались по методу Шеррера. При увеличении времени термообработки размер кристаллов увеличивается, но, как видно из графика (рис. 2), изменения незначительны. Например, для концентрации 0,2 мол.% NdF_3 размер кристаллов изменяется с 238 Å до 282 Å.

Вхождение иона неодима в кристаллическую фазу влияет на его спектрально-люминесцентные свойства (рис. 3). Так, изменяется форма спектра люминесценции и относительные интенсивности полос для различных времен термообработок. При термообработке также проявляется штарковская структура, обусловленная вхождением неодима в кристаллическую фазу $\text{PbY}_x\text{Nd}_{(1-x)}\text{OF}_3$.

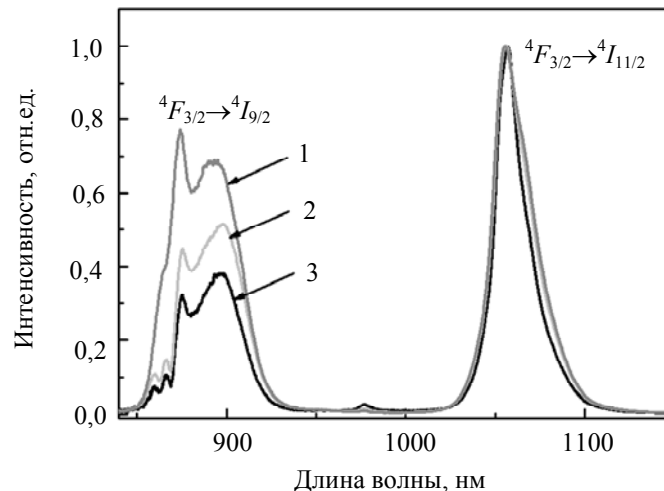


Рис. 3. Спектры люминесценции для образцов с концентрацией NdF_3 1 мол.%; $\lambda_{\text{накачки}} = 808$ нм. Время термообработки: исходное (кривая 1); 30 мин (кривая 2); 120 мин (кривая 3)

Выводы

В работе были исследованы люминесцентные свойства иттрий-свинцово-оксифторидных наностеклокерамик, активированных ионами неодима. В результате вторичной термообработки выделяются кристаллические фазы, зависящие от концентрации фторида неодима и иттрия. Размеры элементарной ячейки выделяющихся кристаллов составили от 5,74 Å до 5,84 Å в зависимости от содержания фторида неодима в исходном стекле.

Размер кристаллов изменяется в диапазоне 200–300 Å в зависимости от времени термообработки. Показано, что увеличение времени термообработки для наностеклокерамик приводит к изменению формы спектра и относительных интенсивностей люминесценции, что говорит о вхождении редкоземельного иона в кристалл.

Заключение

Таким образом, прозрачные иттриевые свинцово-фторидные наностеклокерамики, активированные ионами неодима, представляют интерес в качестве лазерных материалов и других перспективных приложений фотоники.

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы» (Соглашение № 14.В37.21.0169, Минобрнауки РФ).

Литература

1. Колобкова Е.В., Мелехин В.Г., Пенигин А.Н. Оптическая стеклокерамика на основе фторсодержащих силикатных стекол, активированных редкоземельными ионами // Физика и химия стекла. – 2007. – Т. 33. – № 1. – С. 12–19.
2. Асеев В.А., Голубков В.В., Колобкова Е.В., Никоноров Н.В. Лантаноидные оксифториды свинца в стеклообразной матрице // Физика и химия стекла. – 2012. – Т. 38. – № 2. – С. 238–246.

<i>Бибик Анастасия Юрьевна</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Anastasiya.bibik@list.ru
<i>Трофимов Александр Олегович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Eherptional777@mail.ru
<i>Нуриев Рустам Какабаевич</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Nuryev@oi.ifmo.ru
<i>Асеев Владимир Анатольевич</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, Aseev@oi.ifmo.ru
<i>Колобкова Елена Вячеславовна</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор химических наук, профессор, Kolobok106@rambler.ru
<i>Никоноров Николай Валентинович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, зав. кафедрой, Nikonorov@oi.ifmo.ru

УДК 543.456+52-17

ПРОГРАММНЫЙ МОДУЛЬ ДЛЯ ОЦЕНКИ ПАРАМЕТРОВ ДИСПЕРСНЫХ СИСТЕМ ПО ИХ ИЗОБРАЖЕНИЯМ ДЛЯ КОЛИЧЕСТВЕННОГО ТРЕХМЕРНОГО СТЕРЕОАНАЛИЗА

А.П. Калинин, В.В. Манойлов, О.А. Приходько

Рассмотрена возможность применения среды программирования MATLAB для статистического стереоанализа дисперсных систем по их двумерным изображениям. Описаны функции и результаты использования программного модуля, подготовленного для построения вероятностной, статистически обусловленной 3D-модели, которая соответствует более точному описанию структуры и прогнозированию свойств дисперсных систем.

Ключевые слова: среда программирования, статистический анализ, дисперсные системы, микроструктура.

Введение

Для микроструктурного анализа материалов в настоящее время существуют программные пакеты, в частности, ImageExpert Pro 3, Spectr-Met, Microkon Met [1–3] и др. Недостатком данных пакетов является отсутствие функций по построению вероятностных 3D-моделей внутренней структуры материалов на основе статистической обработки двумерных изображений плоских сечений с применением соотношений, известных из стереометрической металлографии [4] и предназначенных для стереоанализа структур. Например, в третьей версии ImageExpert Pro 3 реализована трехмерная визуализация с использованием технологии OpenGL, позволяющая получить реалистичное объемное изображение поверхностного рельефа, синтезированного на основе послойной оптической микроскопии. Однако отсутствует построение вероятностных трехмерных структур по следам их элементов на двумерных сечениях и проекциях, которое дает наиболее объективное представление о внутренней структуре материалов.

Цель настоящей работы заключалась в создании программного модуля в универсальной программной среде MATLAB для реализации возможностей стереоанализа, который позволял бы определять трехмерные геометрические параметры структуры материалов по их статистически обусловленным вероятностным 3D-моделям, построенным на основе анализа элементов двумерных изображений структур.

В качестве контрольных объектов, использованных для разработки и настройки программного модуля, были взяты типичные двумерные изображения структур дисперсных систем (ДС) с разными типами входящих в них фаз: обе фазы твердые (Т/Т) – сплав железа и углерода; одна фаза твердая, другая газообразная (Т/Г) – порошок абразива и пенополистирол; одна фаза твердая, другая жидкая (Т/Ж) – золь.

Знание статистических параметров трехмерной структуры таких сплавов, как ДС типа Т/Т, позволит:

- получить наиболее достоверные зависимости между свойствами и структурой, а также между структурой, составом и режимами технологической обработки сплавов;
- более объективно выявить физическую природу процессов, протекающих в сплавах при их обработке и эксплуатации;
- более точно выбрать оптимальный состав, наилучшую технологию получения и обработки, обеспечивающие создание нужной структуры и, следовательно, требуемых свойств сплавов.

Знание статистических трехмерных параметров таких ДС, как порошки абразива (Т/Г) и золи (Т/Ж), даст возможность более объективно оценить степень их дисперсности, однородность дисперсной фазы с учетом формы частиц.

Создание в среде MATLAB программного модуля для построения трехмерных моделей по изображениям двумерных сечений и проекций разбито на два этапа:

1. разработка, отладка и тестирование модуля статистической обработки и анализа двумерных изображений;
2. дополнение разработанного модуля функциями для построения и анализа 3D-моделей ДС с использованием статистических параметров двумерных изображений.

Ограничения, присутствующие при анализе структуры ДС по двумерным изображениям

В случае анализа структуры ДС типа Т/Т методом оптической металлографии выявляют и изучают двумерные изображения микроструктур в сечениях образца плоскостью шлифа (рис. 1).

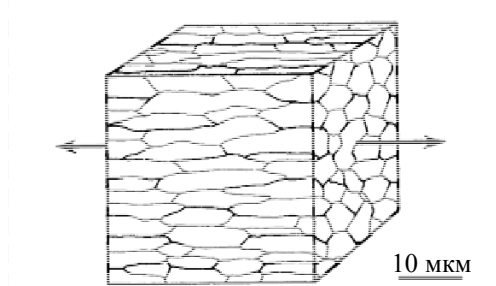


Рис. 1. Двумерные изображения структуры условного образца при различной ориентации плоскости шлифа. Стрелками показано направление текстуры поликристалла

Однако рассмотрение отдельных двумерных изображений создает лишь приблизительное представление о действительном строении трехмерной структуры [4], так как:

1. плоскость шлифа проходит через элементы структуры по случайному сечению, не отражающему объективно их размеры;
2. вероятность пересечения крупных зерен поликристалла и включений дисперсных фаз больше, чем мелких, поэтому доля крупных зерен и включений, определенная по двумерным изображениям, оказывается больше, чем в реальной трехмерной структуре;
3. форма зерен и других структурных элементов может быть несимметричной, что вызывает затруднения при анализе структуры по изображениям отдельных случайных сечений.

В случае дисперсных порошков и золь (рис. 2) наблюдается проекция частиц на плоскость изображения, однако полученная информация будет необъективной, так как частицы могут быть вытянуты вглубь слоя и иметь сложную форму. Неизбежно также заслонение частиц друг другом.

Несмотря на отмеченные ограничения, параметры двумерных изображений применяют для изучения структуры и свойств ДС, однако более объективную оценку структуры и свойств можно получить, используя сведения о трехмерной структуре этих систем. Без разрушения образцов такие сведения можно получить при изучении вероятностной, статистически обусловленной 3D-модели ДС, построенной на основе соотношений, используемых в стереометрической металлографии. Так как методы стереометрической металлографии являются статистическими, то анализируемые ДС должны представлять собой

статистически значимую совокупность элементов. В настоящей работе представлены начальные результаты разработки и выявления возможностей программного модуля, который на основе статистической оценки параметров двумерных изображений ДС позволяет изучать трехмерную структуру материалов.

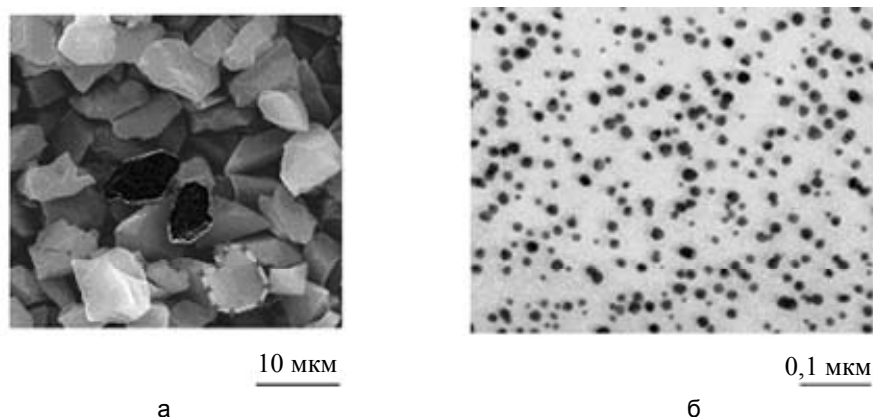


Рис. 2. Изображение двумерной проекции порошка (а) и слоя золя (б). На рис. 2, а, в полуавтоматическом режиме точками и контуром обведены нечеткие границы частицы

Программный модуль статистической обработки и анализа двумерных изображений ДС и построения 3D-моделей в среде MATLAB

Модуль реализует следующие основные функции:

- ввод и масштабирование исходных двумерных изображений различных сечений и проекций;
- пространственная и частотная фильтрация изображений для повышения четкости, контрастности и яркости отдельных элементов структуры;
- пороговая обработка изображений для преобразования в бинарный вид;
- морфологическая обработка изображений для выявления границ элементов структуры;
- определение координат, количества и геометрических параметров элементов структуры;
- построение эмпирических распределений геометрических параметров элементов структуры с оптимизацией размера интервалов гистограмм;
- вычисление основных эмпирических моментов распределения геометрических параметров элементов структуры;
- выявление законов распределения геометрических параметров элементов структуры с применением известных критериев согласия [5];
- построение вероятностной 3D-модели структуры ДС по двумерным изображениям с использованием соотношений стереоанализа.

Модуль способен в автоматическом и полуавтоматическом режимах выполнять обработку изображений элементов структуры, имеющих нечеткие границы или границы с разрывами. Для восстановления границ элементов структуры в автоматическом режиме изображение переводится в однобитовый формат с последующим морфологическим анализом [6, 7], позволяющим выявить контуры элементов. В полуавтоматическом режиме выделение неотчетливых границ производится при вмешательстве оператора. В частности, на изображении порошков (рис. 2) показан контур, замыкаемый программой через точки, нанесенные оператором на границу изображения частицы.

Процесс построения программным модулем изображения 3D-модели слоя золя бромида серебра по изображению (рис. 2, б) включает следующие этапы:

- обработка изображения с переводом в однобитовый формат (рис. 3, а);
- размещение в трехкоординатном пространстве (рис. 3, б) центров частиц дисперсной фазы в соответствии с их координатами на изображении проекции и принятым законом распределения частиц по толщине слоя;
- расчет для каждой частицы диаметра эквивалентной по объему сферы;
- размещение каждой эквивалентной сферы в начале координат (рис. 3, в);
- деформация каждой эквивалентной сферы до получения эллипсоида вращения (рис. 3, г) с параметрами, максимально соответствующими данной реальной частице, и поворот большой оси эллипсоида в соответствии с принятым законом распределения частиц по направлениям;
- перемещение центров объемов эллипсоидов в точки с соответствующими координатами (рис. 3, д).

Для достоверного построения вероятностной трехмерной модели ДС типа Т/Т используются статистические параметры, полученные при анализе двумерных изображений структуры трех взаимно перпендикулярных сечений образца ДС. Кроме описанных выше процедур, вводится расчет диаметров эк-

вивалентных сфер с учетом их случайного рассеяния плоскостью, подчиняющийся закономерности, описанной в [4]. Пример построения 3D-модели для образца полистирола (рис. 4) показывает перспективность дальнейшего развития разработанного модуля с ориентацией на большую детализацию и соответствие полученных изображений реальной структуре.

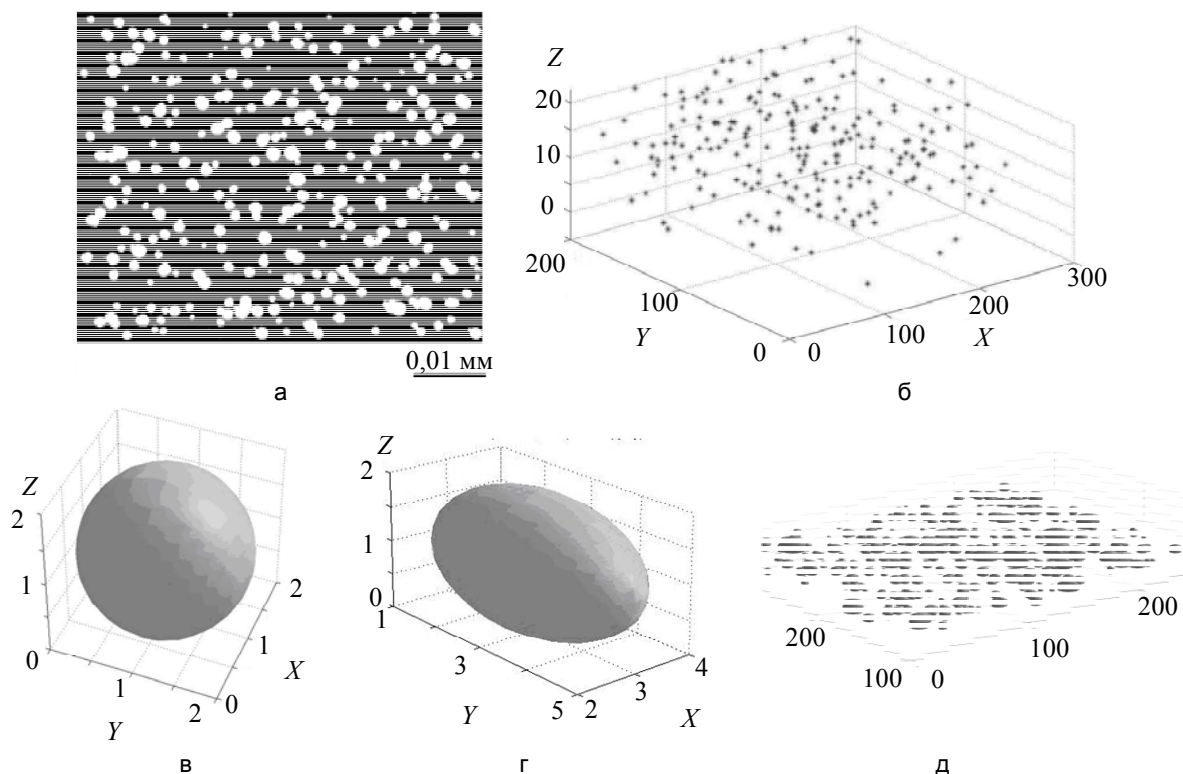


Рис. 3. Результаты этапов создания трехмерной модели слоя золя бромида серебра: изображение двумерной проекции слоя золя в однобитовом формате (а); предполагаемое положение (*) центров частиц золя в трехкоординатном пространстве (б); сфера (в) и эллипсоид (г), эквивалентные объему конкретной частицы золя (шкалы в нанометрах); изображение вероятностной 3D-модели слоя золя (шкалы в нанометрах) (д)

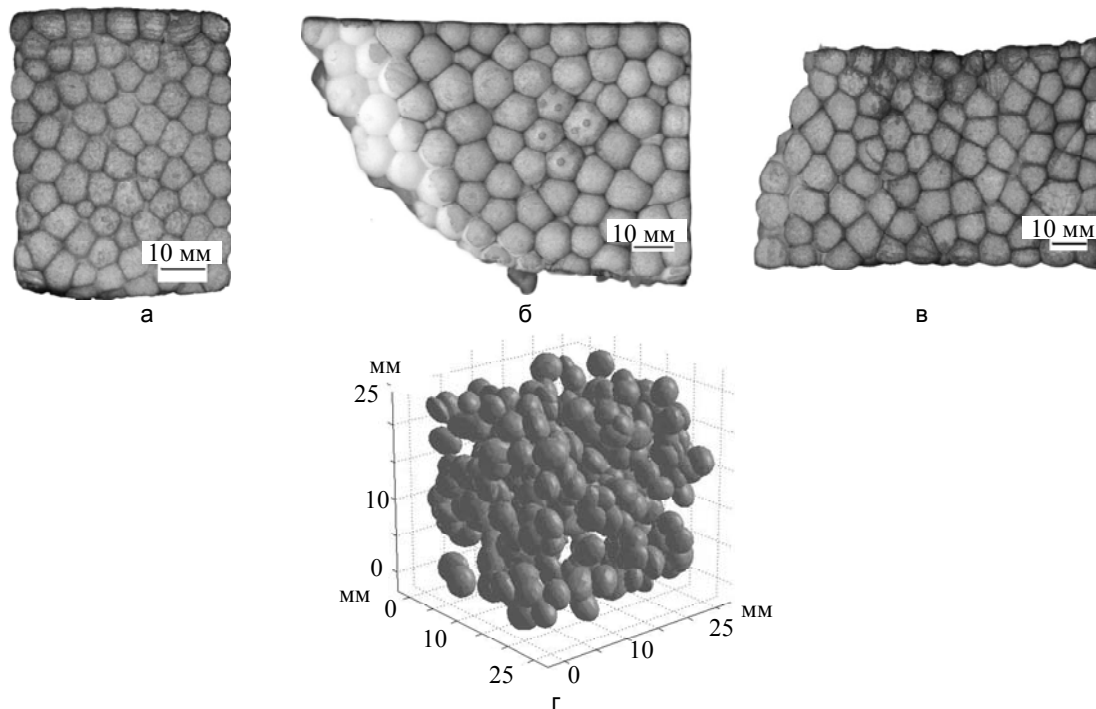


Рис. 4. Изображение вероятностной трехмерной модели структуры пенополистирола (г), созданной с использованием трех плоских, взаимно перпендикулярных сечений: X–Y (а); X–Z (б); Y–Z (в) (шкалы в миллиметрах)

Результаты тестирования модуля

С целью тестирования программного модуля на предмет выявления параметров двумерной структуры ДС и корректности их статистической обработки было выбрано двумерное изображение стандартной феррито-перлитной микроструктуры стали марки 15 (рис. 5, а), по которой оценивается балл зерна поликристалла. После перевода в однобитовый формат и фильтрации (рис. 5, б) к полученному бинарному изображению применялся морфологический анализ. При этом определялись координаты элементов структуры и их границы для последующей оценки геометрических параметров (площадей, периметров, формы). После построения гистограмм, в частности, для корня квадратного из площади зерен феррита ($S^{0.5}$), отражающего их средний размер (рис. 5, в), подбирался теоретический закон распределения с использованием критериев согласия Пирсона и Колмогорова–Смирнова.

Полученные числовые значения параметров структуры стали марки 15 и образцов других сталей из базы данных [8] показали, что применение модуля:

- продемонстрировало (с доверительной вероятностью 0,9) совпадение искомых параметров с их стандартными значениями;
- выявило соответствие вычисленных номеров зерна эталонам [9];
- позволило определить теоретический закон распределения размеров элементов структуры (рис. 5, в), соответствующий логарифмически нормальному распределению, которое является типичным для ДС [10, 11] и дает основу для построения 3D-модели;
- не приводит к потере и искажению информации при обработке изображений.

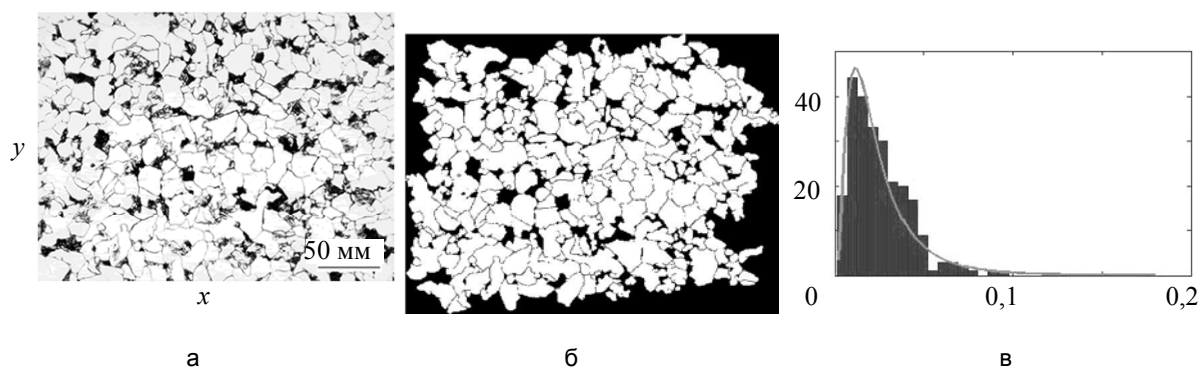


Рис. 5. Стандартная феррито-перлитная структура стали марки 15: до обработки изображения (а); после обработки изображения (б); эмпирическое и теоретическое распределение $S^{0.5}$ (S – площадь зерен феррита), соответствующее логнормальному закону распределения (горизонтальная ось – $S^{0.5}$ в мм; вертикальная ось – количество зерен) (в)

Заключение

Проведенные исследования по разработке программного модуля и анализу результатов его работы позволяет сделать следующие выводы и предположения.

1. На основе универсальной программной среды MATLAB создан доступный, модифицируемый программный модуль стереоанализа структуры дисперсных систем, имеющий все необходимые функции для эффективной обработки двумерных изображений и статистического анализа их геометрических параметров с использованием соотношений стереометрической металлографии для построения вероятностных 3D-моделей дисперсных систем.
2. Описаны основные функции модуля и продемонстрирована корректность его работы на примере анализа структур дисперсных систем различного типа.
3. Новизна работы заключается в разработке на основе программной среды MATLAB модуля, включающего совокупность методов обработки двумерных изображений дисперсных систем для построения 3D-моделей структур на основе вероятностных законов распределения геометрических параметров их элементов.
4. Практическая значимость работы заключается в получении инструмента для построения и изучения изображений трехмерной структуры дисперсных систем на основе вероятностной, статистически обусловленной 3D-модели, позволяющей получить более объективную информацию о структуре дисперсных систем, чем при изучении только двумерных изображений.
5. Перспектива развития данного модуля состоит в дальнейшей детализации воспроизведения формы и размеров элементов структуры при построении вероятностных 3D-моделей и их связи со свойствами дисперсных систем.

Литература

1. Программный пакет SPECTR MET 4.5.5 [Электронный ресурс]. – Режим доступа: http://www.niir.ru/catalog/show.php?node_id=218, свободный. Яз. рус. (дата обращения 21.09.2012).
2. Программа ImageExpert Pro3 для количественного анализа изображений [Электронный ресурс]. – Режим доступа: <http://www.nexsys.ru/iepro3x.htm>, свободный. Яз. рус. (дата обращения 21.09.2012).
3. Программный пакет обработки и анализа изображений металлографии MICROKON MET [Электронный ресурс]. – Режим доступа: http://www.microkon.ru/program_mmet.html, свободный. Яз. рус. (дата обращения 21.09.2012).
4. Салтыков С.А. Стереометрическая металлография. – М.: Металлургия, 1976. – 270 с.
5. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. – 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2004. – 573 с.
6. Дьяконов В. MATLAB. Основы обработки сигналов и изображений. – СПб: Питер, 2002. – 608 с.
7. Консультационный центр MATLAB [Электронный ресурс]. – Режим доступа: <http://matlab.exponenta.ru>, свободный. Яз. рус. (дата обращения 21.09.2012).
8. База данных микроструктур металлов и сплавов [Электронный ресурс]. – Режим доступа: <http://www.microstructure.ru>, свободный. Яз. рус. (дата обращения 21.09.2012).
9. ГОСТ 5639-82. Стали и сплавы. Методы выявления и определения величины зерна. – М.: Изд-во стандартов, 1983. – Введ. 01.01.83. – 21 с.
10. Лемешко Б.Ю., Чимитова Е.В. О выборе числа интервалов в критериях согласия типа χ^2 // Заводская лаборатория. Диагностика материалов. – 2003. – Т. 69. – № 1. – С. 61–67.
11. Колмогоров А.Н. О логарифмически нормальном законе распределения размеров частиц при дроблении // Доклады АН СССР. – 1941. – Т. 31. – № 2. – С. 99–101.

- Калинин Антон Павлович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, formant2004@mail.ru
- Манойлов Владимир Владимирович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, доцент, профессор, manoilov_vv@mail.ru
- Приходько Олег Анатольевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, доцент, oar10@yandex.ru

УДК 678.666.01

ВЛИЯНИЕ МОДИФИКАЦИИ ПОВЕРХНОСТИ НАПОЛНИТЕЛЯ НА СВОЙСТВА ЛАТЕКСНЫХ ПЛЕНОК

Е.А. Рюткянен, Н.В. Сиротинкин, М.В. Успенская

Представлены результаты исследования свойств полимерных пленок хлоропренового латекса, наполненных модифицированными микрошариками, где в качестве модификатора используется тонкое полимерное покрытие на основе поливинилхлоридной пленки. Выявлены закономерности влияния представленных наполнителей на свойства композиционных материалов.

Ключевые слова: наполнитель, стеклянные микрошарики, адгезия, хлоропреновый латекс, модификация.

Введение

Современные технологии позволяют производить стеклосыпучие порошки, представляющие собой стеклянные частицы сферической формы диаметром 10–200 мкм. Такие характеристики наполнителей, как сферическая форма, достаточная прочность, высокая адгезия к большинству известных полимеров, предопределяют их широкое использование [1].

Стеклянные микрошарики (СМШ) – один из видов наполнителей, получающий все более широкое применение в различных областях техники. В связи с интенсивным развитием инновационного подхода в бизнесе возросли и требования к созданию новых композиционных материалов, способных к длительной эксплуатации в жестких условиях, например, под действием высоких температур, больших и разнообразных механических нагрузок, химически активных сред, излучений и т.д. С целью улучшения физико-механических и эксплуатационных характеристик полимеров и композитов на их основе, применяемых в промышленности чаще всего используют модификацию поверхности наполнителя [2].

Известно, что при введении частиц наполнителя в полимерную матрицу между полимером и наполнителем возможно адсорбционное, а в некоторых случаях и химическое взаимодействие на границе раздела фаз. Это взаимодействие тем сильнее, чем больше поверхность контакта полимера с наполнителем, т.е. чем меньше размер частиц наполнителя и, соответственно, больше его удельная поверхность, которая характеризует размеры внутренних полостей пористого тела или частиц раздробленной фазы

дисперсной системы. Адсорбционное и химическое взаимодействия на границе раздела фаз зависят от многих параметров, например, от природы полимера и наполнителя, условий смешения и приготовления композита, наличия модификаторов на поверхности частиц наполнителя и т.д. [3]. В качестве модификаторов поверхности наполнителей перспективно использование различных полиуретановых покрытий. Так при создании тонкопленочных термоизоляционных высокопрочных и высокотехнологичных материалов, в том числе многослойных, используют СМШ и микросферы, модифицированные полиуретановым наноструктурным покрытием, что приводит к упрочнению материала в среднем в 2–6 раз [4].

Наибольший интерес представляют работы по аппретации стекла и стеклянных изделий высокомолекулярными соединениями. Эти работы основываются на выборе типа полимера, его микроструктуры, разветвленности, полярности, варьировании толщины наносимой пленки и т.п., что позволяет регулировать уровень защиты пленок от внешних воздействий, светопоглощение в необходимой области спектра, адгезионную способность материала и т.п.

В настоящее время в промышленности известны СМШ, модифицированные гидрофобными полиуретановыми наноструктурными покрытиями с толщиной до 10–30 нанометров [4, 5]. В качестве наполнителей полимерной матрицы все более широкое распространение получают алюмосиликатные зольные микросферы (ЗМ), извлекаемые из золы-уноса, образующейся при сжигании угля на тепловых электростанциях. Для обеспечения прочного сцепления наполнителя с полимерной матрицей проводят модификацию микросфер различными аппретами. Роль аппрета в композициях для обработки стеклянных наполнителей (таких как стекловолокно, СМШ, алюмосиликатные микросферы (АСМ)) сводится к образованию на поверхности наполнителя соединений, которые, прочно удерживаясь на ней, повышают активность поверхности по отношению к полимеру, так как содержат функциональные группы, по которым может идти химическое взаимодействие с молекулами связующего [6]. Универсальные возможности аппретирующих агентов подтверждены различными областями их применения, такими как нефтепереработка, лакокрасочная промышленность, литейное, текстильное и химическое производства, а также при производстве каучука и эластомеров, герметиков и т.п. В качестве модификаторов поверхности ЗМ используют полимеры производных (мет)акриловой кислоты, такие как амиды, метиловые эфиры различной молекулярной массы, а также производные силана, например, γ -аминопропилтриэтоксисилан [7, 8].

Недостаточная изученность влияния модифицированных наполнителей на физико-механические и реологические характеристики полимерных матриц и взаимосвязи между составом и физико-химическими и механическими свойствами полимерных материалов препятствуют расширению возможностей их применения, что делает этот вопрос крайне актуальным. Таким образом, целями настоящей работы являются исследование возможности модификации поверхности СМШ хлоропреновым латексом и изучение влияния наполнителя на эксплуатационные характеристики получаемых композиционных материалов.

Экспериментальная часть

Латексные пленки изготавливались согласно ГОСТ 24920-81 [9] из хлоропренового латекса марки LD-750 (табл. 1), в качестве отвердителя использовалась мочевиная. СМШ дисперсностью 0–200 мкм (табл. 2) были модифицированы полимерной поливинилхлоридной (ПВХ) нанопленкой согласно методике, описанной в работе [4], где в качестве растворителя использовали дихлорметан, в качестве осадителя – воду.

Параметр	Значение
Внешний вид	молочно-белая жидкость
Массовая доля сухого вещества, %	50
pH	8
Минимальная температура пленкообразования, °C	–40
Внешний вид пленки	однородная, прозрачная

Таблица 1. Свойства латекса марки LD-750

Параметр	Значение
Внешний вид	серо-зеленый порошок
Теплопроводность, Вт/мК	0,05–0,13
Плавучесть, вес %	99,3
Истинная плотность, г/см ³	2,5
Морозоустойчивость	более 20 циклов
Температура начала размягчения, °C	1400
Гигроскопичность, %	до 0,15

Таблица 2. Свойства СМШ

Наличие полимерной пленки на поверхности СМШ доказывали с помощью исследовательского микроскопа ПМТ-3. Испытания прочности и относительного удлинения полимерных пленок при разрыве проводились на разрывной машине РМИ-5 согласно ГОСТ 270-75 [10], водопоглощение определялось в соответствии с ГОСТ 20869-75 [11].

Результаты исследования

Как известно, хлоропеновые каучуки, например «Найрит», обладают высокой прочностью на разрыв, которая составляет 25–32 МПа вследствие высокой регулярности строения вулканизатов [12]. Минеральные наполнители и сажи снижают сопротивление разрыву, но поиск эффективных наполнителей, повышающих модуль упругости, еще не окончен. Наполнение хлоропеновых каучуков недорогими и доступными стеклянными микросферами и микрошариками приводит к монотонному уменьшению значений прочности материала с ростом концентрации наполнителя.

На рис. 1, 2 показано влияние СМШ и АСМ на прочность и эластичность полимерных пленок на основе хлоропенового латекса. Как видно из графиков, наполнение хлоропеновых каучуков недорогими и доступными СМШ и АСМ монотонно снижает прочность материала с ростом концентрации наполнителя. Можно заметить и схожий вид кривых при наполнении латекса LD-750 стеклянными микрошариками и латекса «Найрита» – алюмосиликатными микросферами [2]. Уменьшение значений прочности и эластичности является следствием химической инертности наполнителя и появления открытых пор, количество которых возрастает с увеличением концентрации частиц.

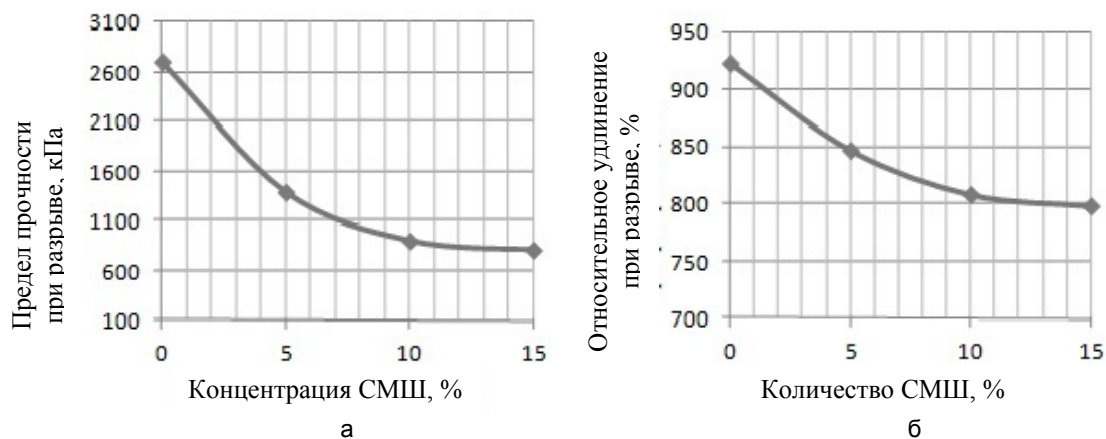


Рис. 1. Зависимость предела прочности (а) и относительного удлинения при разрыве (б) латексных пленок на основе хлоропенового латекса LD-750, наполненного СМШ

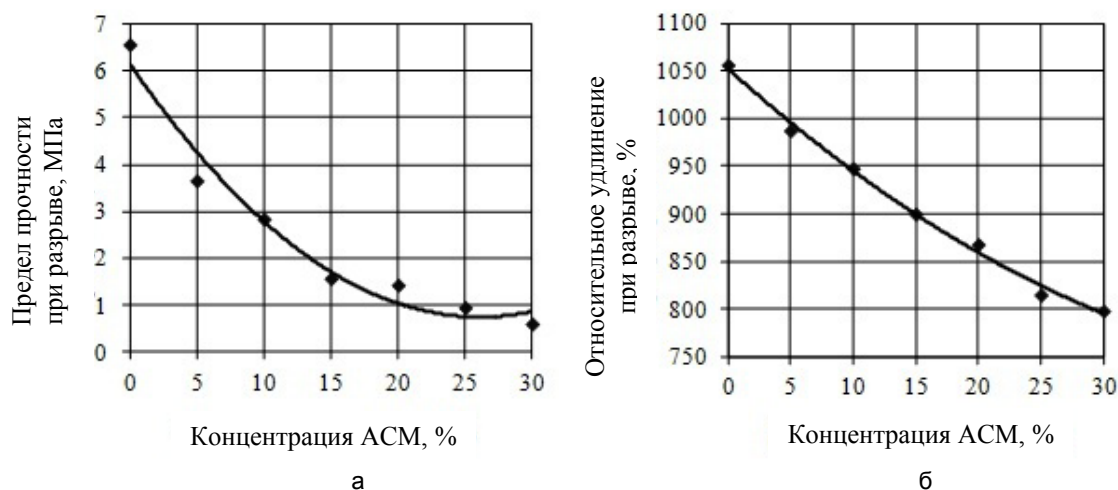


Рис. 2. Зависимость предела прочности (а) и относительного удлинения при разрыве (б) латексных пленок на основе хлоропенового латекса «Найрит», наполненного АСМ

Возникновение открытых пор обусловлено состоянием поверхности наполнителя, например, его неровностями и шероховатостями. Для устранения этих дефектов целесообразно наносить на поверхность частиц полимерный слой, совместимый со связующей композицией – полимерной матрицей. К числу таких модификаторов относится ПВХ. Наличие полимера на поверхности стеклянных шариков видно на рис. 3.

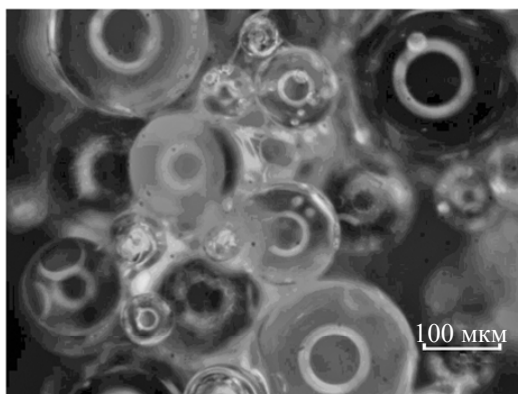


Рис. 3. Микрофотографии модифицированных СМШ

При среднем диаметре стеклянных микрошариков $d_{\text{СМШ}} = 200 \text{ мкм}$, плотности ПВХ $\rho_{\text{ПВХ}} = 1,39 \text{ г/см}^3$, насыпной плотности СМШ $\rho_{\text{СМШ}} = 2 \text{ г/см}^3$ средняя расчетная толщина полимерных пленок составляет около 1,5 нм. Модифицированный наполнитель (обозначенный на рис. 4, 5 как СМШ+ПВХ) уменьшает водопоглощение и увеличивает плотность всей полимерной композиции, что свидетельствует об уменьшении пористости (рис. 4), а также значительно увеличивает эластичность композиции при небольших концентрациях (на 300% при 5%-ном наполнении модифицированным наполнителем) и не существенно влияет на модуль прочности (рис. 5). На рис. 4, 5 содержание наполнителя 0% относится к исходной, ненаполненной латексной пленке.

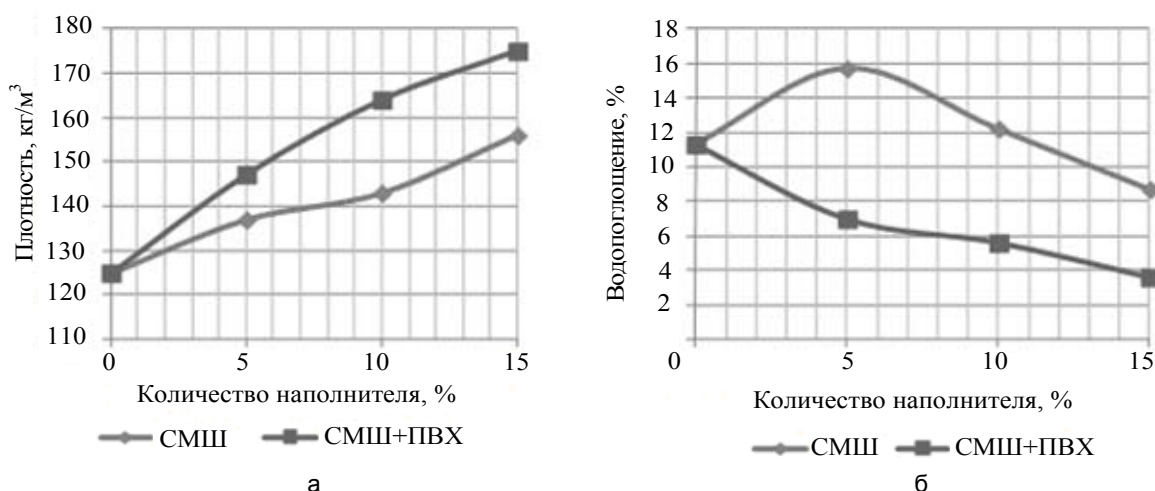


Рис. 4. Зависимость плотности (а) и водопоглощения (б) наполненных латексных пленок от концентрации наполнителей

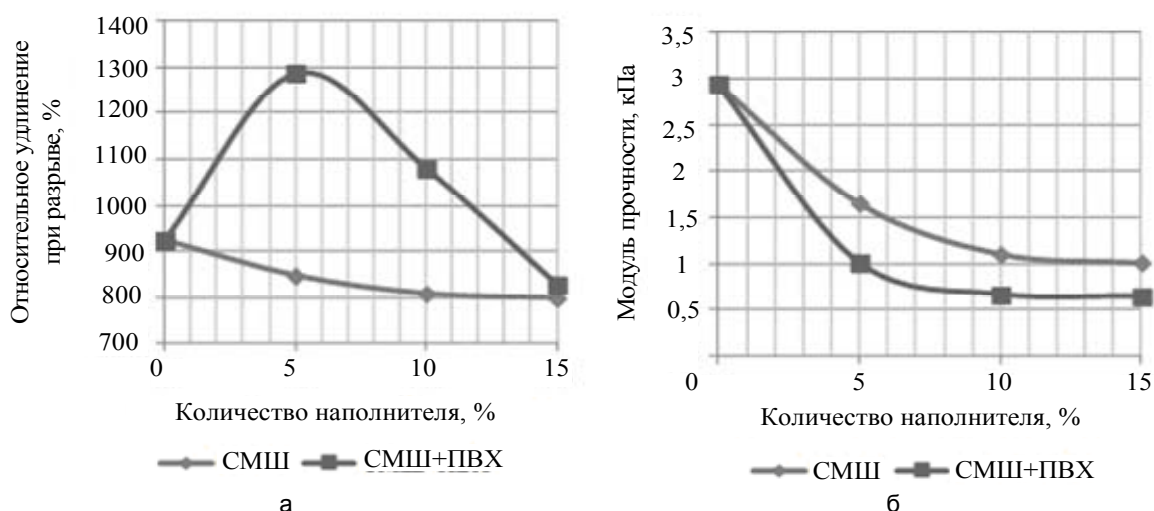


Рис. 5. Зависимость эластичности (а) и модуля прочности (б) латексных пленок от количества наполнителя

Спектральный анализ (ИК-спектроскопия) показывает, что между поверхностью СМШ и ПВХ не образуются ковалентные и водородные связи, а между полимером и стеклом, по-видимому, действует механическая адгезия, вызванная заполнением дефектов поверхности полимером.

Учитывая в целом общую химическую природу промежуточного слоя толщиной 1,5 нм между наполнителем и полимерной матрицей, становится понятной инертность полимера к воздействию модифицированного наполнителя и возникновение максимума на кривой «относительное удлинение–концентрация модифицированного СМШ».

По сравнению с исходными микрошариками поверхность модифицированных СМШ не является шероховатой (рис. 3): дефекты заполнены полимером, следовательно, устраняются причины механической адгезии на поверхности раздела и возникает когезионное взаимодействие двух хлорсодержащих полимеров. Как следствие, увеличивается относительное удлинение при разрыве (рис. 5).

Заключение

В результате проведенных исследований выявлено влияние полимерной пленки ПВХ, модифицирующей поверхность частиц наполнителя – СМШ, на физико-механические свойства наполненных латексных пленок. Продemonстрировано уменьшение пористости материала и водопоглощения за счет устранения шероховатостей на поверхности микрошариков. Показано, что малая модифицирующая добавка полимера (1,5 нм) на поверхности стеклянных микрошариков приводит к значительному возрастанию эластичности при небольших концентрациях наполнителя.

Литература

1. Асланова М.С., Стеценко В.Я., Шустрое А.Ф. Полые неорганические микросферы // Химическая промышленность за рубежом: обзор информ. – НИИТЭХИМ, 1981. – С. 33–51.
2. Давудов М.Г. Трудногорючий эластичный полимерный изоляционный материал: Дис.... канд. техн. наук: 05.17.06. – СПб, 2011. – 128 с.
3. Иванова В.Н., Алеушин В.Н. Технология резиновых изделий. – Л.: Химия, 1988. – 288 с.
4. Сиротинкин Н.В., Бестужева В.В., Бондарева Е.А., Омельчук Ю.В., Давудов М.Г. Поверхностная модификация стеклянных микрошариков полиуретанами // Каучук и резина. – 2010. – № 6. – С. 26–30.
5. Способ получения гидрофобных покрытий на поверхности стекла: Пат. 2065417 Рос. Федерация: МКИ⁶С03 С 17/30, Колосов Г.Г. – № 2065417; Заявл. 23.03.1993; Опубл. 20.08.1996.
6. Демина Н.М. Современные тенденции развития в области кремнийорганических аппретов для стекловолокон // Стекло и керамика. – 1999. – № 7. – С. 18–21.
7. Варламова Л.П., Извозчикова В.А., Рябов С.А. и др. Влияние алюмосиликатных микросфер на физико-механические и реологические свойства жестких пенополиуретанов // ЖПХ. – 2008. – Т. 81. – № 3. – С. 502–504.
8. Варламова Л.П., Черкасов В.К., Семенов Н.М. и др. Влияние модификации поверхности алюмосиликатных зольных микросфер на физико-механические свойства пенополиуретана // ЖПХ. – 2009. – Т. 82. – № 6. – С. 1040–1042.
9. ГОСТ 24920-81. Латексы синтетические. Правила приемки, отбор и подготовка проб. – Введ. 01.01.83. – М.: Изд-во стандартов, 1983. – 18 с.
10. ГОСТ 270-75. Резина. Метод определения упруго-прочностных свойств при растяжении. – Введ. 01.01.78. – М.: Изд-во стандартов, 1975. – 11 с.
11. ГОСТ 20869-75. Пластмассы ячеистые жесткие. Метод определения водопоглощения. – Введ. 01.07.76. – М.: Изд-во стандартов, 2000. – 4 с.
12. Кирпичников П.А., Аверко-Антонович Л.А., Аверко-Антонович Ю.О. Химия и технология синтетического каучука. – М.: Химия, 1975. – С. 380–387.

<i>Рюткянен Евгений Александровна</i>	– Санкт-Петербургский государственный технологический институт (технический университет), аспирант, ryutkyanen.evgen@mail.ru
<i>Сиротинкин Николай Васильевич</i>	– Санкт-Петербургский государственный технологический институт (технический университет), доктор химических наук, профессор, декан, biotechnology_faculty@technolog.edu.ru
<i>Успенская Майя Валерьевна</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, доцент, зав. кафедрой, mv_uspenskaya@mail.ru

УДК 338.45

**ПРИНЦИПЫ ПОСТРОЕНИЯ ОТРАСЛЕВОЙ СИСТЕМЫ
АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ В АВИАЦИОННОМ
ПРИБОРОСТРОЕНИИ**

П.П. Парамонов, Ю.А. Гатчин, И.О. Жаринов, О.О. Жаринов, М.С. Дейко

Рассматриваются принципы построения отраслевой системы автоматизированного проектирования бортового авиационного оборудования. Приведены маршруты проектирования изделий, функциональная схема программного обеспечения отраслевой системы проектирования. Описаны аппаратные решения реализации системы автоматизированного проектирования на базе локальной вычислительной сети приборостроительного предприятия.

Ключевые слова: отраслевая система автоматизированного проектирования, авиационное оборудование.

Введение

Современные тенденции развития приборостроительной отрасли в области проектирования бортового радиоэлектронного оборудования (БРЭО) неразрывно связаны с применением систем автоматизированного проектирования (САПР). Разработчики САПР предлагают различные универсальные программные инструменты, предназначенные для автоматизации отдельных этапов проектирования БРЭО. Известны, например, пакеты прикладных программ:

- трехмерного проектирования конструкций аппаратуры БРЭО – AutoCAD, Solidworks, UniGraphics;
- инженерных расчетов и моделирования тепловых полей, создаваемых аппаратурой БРЭО: BETAsoft, Sauna, Асоника-Т;
- моделирования электромагнитных полей, создаваемых аппаратурой БРЭО – OrCAD Family Release, GENESYS;
- спектрального анализа радиочастотных сигналов, генерируемых и принимаемых аппаратурой БРЭО – TESLA;
- моделирования прочностных и резонансных характеристик конструкций авиационных изделий – Samcef, Mecano, Boss Quattro;
- автоматической трассировки печатных плат с учетом трехмерной компоновки элементов монтажа – ACCEL EDA, PCAD;
- конструирования и разводки жгутов в трехмерном пространстве корпусов аппаратуры БРЭО – UG/Wiring;
- моделирования гидродинамических процессов в системах охлаждения авиационных изделий с учетом тепловыделения электрорадиоэлементов и физических процессов теплопереноса – Fine/Turbo;
- электронного моделирования радиотехнических сигналов и цепей – Cadence Design, Mentor Graphics, XILINX Foundation, ALTERA, MicroCap;
- проектирования программного обеспечения C/C++/C#, ADA и др.

Данные программные инструменты составляют основу автоматизации проектных технологий и являются законченными программными решениями. Специфичность форматов хранения и представления данных результатов проектирования в таких пакетах программ существенно влияет на уровень автоматизации процесса проектирования БРЭО в целом. Это связано в первую очередь с тем, что формат представления данных результатов проектирования какой-либо составной части изделия, полученного с применением САПР одного вида, оказывается несогласованным с форматом хранения и представления данных в САПР другого вида, применяемого на последующем, более позднем, этапе проектирования.

В таких случаях возникает потребность введения в технологические маршруты проектирования изделий «ручных» процедур подготовки (переформатирования) отдельных электронных документов, что существенно снижает эффект применения САПР и повышает суммарную трудоемкость разработки конструкторской, технологической и программной документации на изделие в целом.

Для исключения или минимизации объемов «ручных» работ в процессе проектирования все применяемые при разработке БРЭО средства САПР должны интегрироваться в единую отраслевую САПР [1], обладающую совместимыми форматами представления, хранения и передачи данных и поддерживающую все этапы жизненного цикла БРЭО – от маркетинговых исследований целесообразности разработки новых видов изделий до их последующей утилизации. Соответствующая функциональная схема взаимодействия разнородных средств САПР в составе единой отраслевой САПР поддержки жизненного цикла БРЭО приведена на рис. 1.

Как следует из рис. 1, интеграция разнородных средств САПР предполагает объединение технологических операций, реализуемых каждой из них в отдельности, в единую проектную среду «проектирование–производство–эксплуатация» [2]. Единая проектная среда обеспечит автоматизированную поддержку этапов жизненного цикла БРЭО и позволит исследовать и разрабатывать «сквозные» проектные технологии создания изделий, входящих в БРЭО, до их реализации в виде готовых физических объектов.



Рис. 1. Функциональная схема взаимодействия разнородных средств САПР в составе единой отраслевой САПР поддержки жизненного цикла БРЭО (CAE – Computer Aided Engineering; CAD – Computer Aided Design; CAM – Computer Aided Manufacturing; ECAD – Electronic Computer Aided Design; PDM – Product Data Management; ERP – Enterprise Resource Platting; MRP – Manufacturing Requirement Platting; MES – Manufacturing Executing System; SCM – Supply Chain Management; CRM – Customer Relationship Management; SCADA – Supervisory Control and Data Acquisition; CNC – Computer Numerical Control; S&SM – Sales and Service Management; CPC – Collaborative Product Commerce; CASE – Computer Aided Software Engineering)

Технологический цикл разработки, производства и испытаний БРЭО на основе отраслевой САПР

Технологический цикл разработки, производства и испытаний аппаратуры БРЭО на основе отраслевой САПР авиационного приборостроения представлен на рис. 2.

Как следует из рис. 2, основу технологического цикла разработки, производства и испытаний аппаратуры БРЭО составляют проектные процедуры структурного проектирования отдельных изделий, реализации их аппаратных и программных компонентов, поддержки задач технологической подготовки производства, изготовления и испытаний изделия. Существенной составляющей проектирования БРЭО является также этап комплексирования аппаратуры [3].

Основная задача комплексирования аппаратуры заключается в объединении методами структуризации разнородных приборов и устройств (изделие 1, изделие 2 и т.д.) в единый комплекс бортового оборудования. Объединение осуществляется на аппаратном и программном уровне через анализ информационных потоков аэродинамических, пилотажно-навигационных и др. параметров движения летательного аппарата и синтез структур БРЭО. Результатом этапа комплексирования аппаратуры является комплект документации на комплекс БРЭО. В соответствии с принятой терминологией комплекс БРЭО представляет собой систему более высокого порядка по отношению к отдельным ее компонентам.

Таким образом, задачу автоматизации этапов «проектирование–производство–эксплуатация» БРЭО на основе единой проектной среды отраслевой САПР авиационного приборостроения следует рассматривать как задачу создания такой интегрированной САПР, в составе которой:

- автоматизировались бы все основные проектные процедуры проектирования БРЭО;

- взаимодействие (обмен данными между приложениями) разнородных средств САПР осуществлялось бы на основе стандартизованных протоколов передачи информации, например, по стандартам группы STEP (Standard for the Exchange of Product).

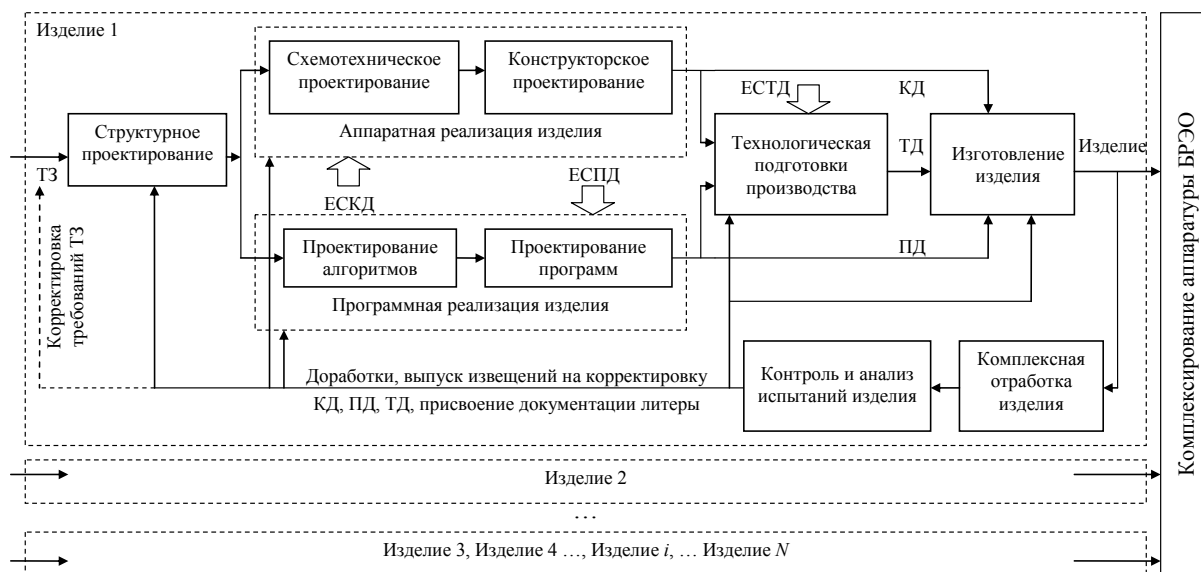


Рис. 2. Технологический цикл разработки, производства и испытаний аппаратуры БРЭО (КД – конструкторская документация; ТД – технологическая документация; ПД – программная документация; ЕСКД – единая система конструкторской документации; ЕСТД – единая система технологической документации; ЕСПД – единая система программной документации; ТЗ – техническое задание)



Рис. 3. Схема взаимодействия основных функциональных компонентов, автоматизирующих проектные процедуры создания аппаратуры комплексов БРЭО

Схема взаимодействия основных функциональных компонентов, автоматизирующих проектные процедуры создания БРЭО на основе отраслевой САПР, представлена на рис. 3.

Как следует из рис. 3, автоматизация этапа структурного проектирования изделия осуществляется средствами автоматизации процедур выбора оптимальной (по заранее введенному критерию) структуры изделия и процедур распределения функций между аппаратными и программными компонентами изделия. Методы выбора оптимальной структуры изделия в классе бортового приборного оборудования рассмотрены в работе [4], решение задачи «о назначении» функциональных задач на доступные вычислительные ресурсы БРЭО рассмотрено в [5]. Автоматизация этапа разработки аппаратной реализации изде-

лия осуществляется последовательно на основе автоматизации процедур схмотехнического проектирования и процедур разработки конструкции изделия. Методы автоматизации процедур схмотехнического проектирования и генерации проектных решений рассмотрены в работах [6] и [7] соответственно.

Автоматизация этапа разработки программной реализации изделия осуществляется последовательно на основе автоматизации процедур проектирования алгоритмического обеспечения и процедур автоматизации программирования. Методы автоматизации процедур проектирования алгоритмов и программ рассмотрены в работе [8].

Результатом автоматизации этапов разработки аппаратной и программной реализации изделия являются комплекты программной и конструкторской документации, представляющие собой основу (входные данные) для этапа технологической подготовки производства. Автоматизация технологической подготовки производства реализуется на основе принципов стандартизации и унификации технологических операций и технологических процессов с применением универсального станочного оборудования. При этом комплект КД на изделие трансформируется в формат 3-Dimension (3D)-моделей.

3D-модели служат основой для изготовления механических деталей изделия автоматизированным способом на станках с числовым программным управлением. Сборочное производство (поверхностный монтаж электрорадиоэлементов, изготовление многослойных печатных плат) также реализуется автоматизированным способом на основе маршрутных карт, преобразующих конструкторские документы в управляющие программы для тепловых печей и автоматизированных установщиков компонентов.

Автоматизация процедур контроля и испытаний изделия осуществляется вычислительными средствами диагностики, разрабатываемыми автоматизированным способом. В состав средств диагностики входят программные компоненты диагностических тестов исправности аппаратуры, входящие в состав комплекта программного обеспечения (ПО) изделия, аппаратные компоненты автоматизированных рабочих мест (АРМ) по проверке, настройке и регулировке аппаратуры и испытательное оборудование предприятия.

Таким образом, автоматизация основных процедур проектирования, изготовления и испытаний аппаратуры может быть осуществлена на основе интеграции специализированных компонентов САПР, взаимодействующих между собой посредством стандартизованных протоколов передачи информации.

Очевидным преимуществом создания отраслевой САПР является возможность ее обучения в процессе эксплуатации. В результате однократного применения формируются отраслевые библиотеки элементов автоматизации, пригодные для повторного использования на предприятиях приборостроительного профиля, уточняются принципы стандартизации и унификации в разработке, производстве и эксплуатации авиационного приборного оборудования, разрабатываются средства, обеспечивающие автоматизацию разработки самих средств автоматизации проектирования.

Функциональная схема программного обеспечения отраслевой САПР и особенности ее применения

Функциональная схема ПО типовой САПР АРМ разработчика БРЭО с различными формами взаимодействия «САПР_i – файл результата проектирования» приведена на рис. 4.

В состав ПО отраслевой САПР входят компоненты взаимодействия ПО САПР с операционной системой инструментальной ЭВМ, на основе которой организовано АРМ разработчика аппаратуры БРЭО, и специализированные компоненты. Специализированные компоненты представлены совокупностью:

- инструментальной составляющей, осуществляющей математическую поддержку решения основных проектных задач и реализации основных проектных операций;
- рабочей составляющей, осуществляющей диалоговое взаимодействие разработчика аппаратуры и рабочей среды проектирования;
- компонентов общего назначения, реализующих невидимые для пользователя операции;
- программы-монитора, взаимодействующей с компонентами операционной системы инструментальной ЭВМ для функционирования универсальных процедур и функций ПО САПР.

Файл результатов проектирования отраслевой САПР формируется в рабочей оболочке САПР и доступен для передачи по локальной сети предприятия всем специализированным САПР_i. Как следует из рис. 4, возможны два различных режима работы ПО САПР с файлами результатов проектирования: непосредственный, когда каждое прикладное ПО САПР_i может обрабатывать (читать, редактировать) только файлы, разработанные в данной среде проектирования, и универсальный, когда файлы результатов проектирования, подготовленные в различных САПР, доступны для обработки прикладному ПО любой САПР на основе унифицированных методов доступа. Для построения отраслевой САПР БРЭО, очевидно, пригоден второй (универсальный) вариант.

Схема взаимодействия разнородных САПР_i в составе единой отраслевой САПР авиационного приборостроения приведена на рис. 5.

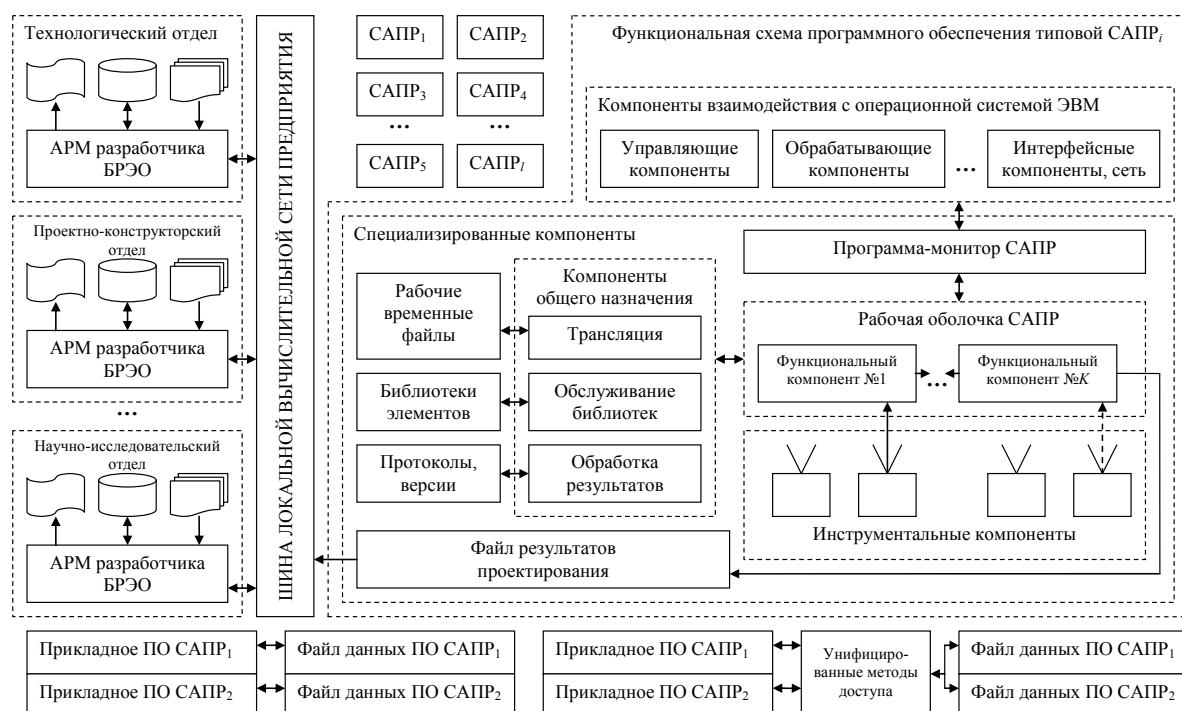


Рис. 4. Функциональная схема ПО типовой САПР автоматизированного рабочего места разработчика БРЭО с различными формами взаимодействия «САПР_i – файл результата проектирования»

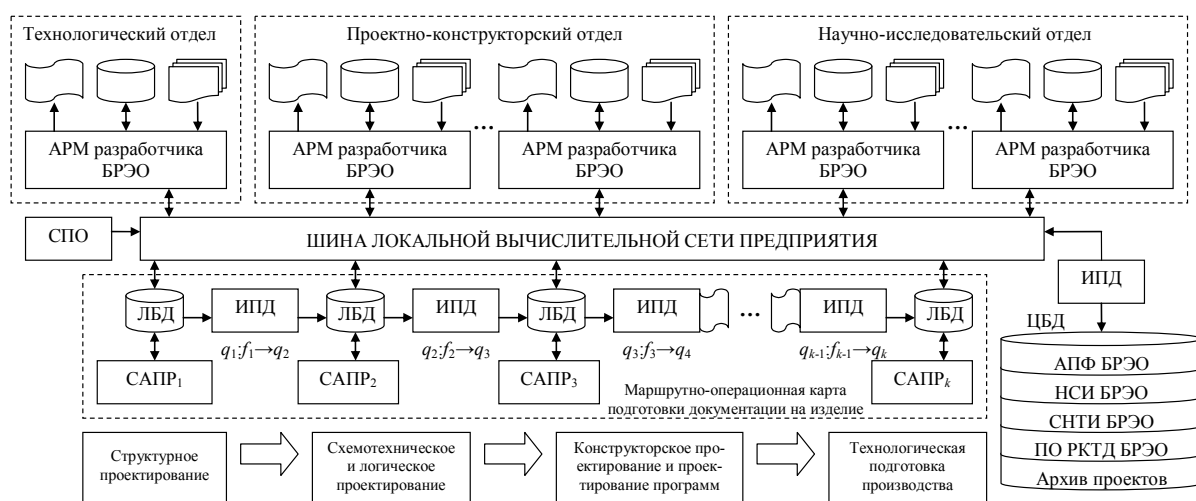


Рис. 5. Схема организации процесса сквозного проектирования изделий БРЭО на базе локальной вычислительной сети предприятия и центрального банка данных проектов предприятия

Взаимодействие разнородных САПР_i в составе единой отраслевой САПР БРЭО осуществляется программным способом на основе интерфейсов преобразования данных (ИПД). Функция интерфейса заключается в нестандартизованных преобразованиях результатов проектирования САПР_i, хранящихся в локальном банке данных (ЛБД) этой САПР, в вид, пригодный для ввода этих данных в САПР_{i+1} и для последующей обработки. Функцию преобразования (технологический переход) данных q_i в q_{i+1} осуществляет оператор f_i (оператор преобразования форматов). Физический обмен данными между разнородными САПР осуществляется через технические средства шины локальной вычислительной сети предприятия и специализированное сетевое программное обеспечение (СПО).

Как следует из рис. 5, все технические отделы, принимающие участие в разработке конструкторской, технологической и программной документации, имеют доступ к локальной вычислительной сети предприятия. Доступ осуществляется через подключенные унифицированные АРМ разработчиков БРЭО.

Результаты проектирования бортового авиационного оборудования с использованием отраслевой САПР хранятся в центральном банке данных (ЦБД) предприятия, доступ к которому также осуществляется с АРМ разработчиков через ИПД. В ЦБД хранятся разрабатываемые предприятием алгоритмы и программы функционирования (АПФ) БРЭО; документация на изделия; нормативно-справочная инфор-

мация (НСИ) – описание изделий, описание материалов, описание технологических процессов; справочная научно-техническая информация (СНТИ) – описание авторских изобретений, патентов, результатов выполнения научно-исследовательских и опытно-конструкторских работ (НИОКР); программное обеспечение разработки конструкторской и технологической документации (ПО РКTD).

Заключение

Исследования проблемы построения отраслевых САПР авиационного приборостроения проводятся сегодня как в нашей стране, так и за рубежом. В частности, известны работы [9] в области разработки технологий сквозного проектирования авиационного оборудования в США, которые получили название АТИР (Avionics Technology Integrated and Prototyping). Технология АТИР была впервые использована в качестве элемента САПР исследовательских стадий проектирования компаниями Boeing и Lockheed Martin в процессе выполнения НИОКР AVSEP (Avionics Virtual Systems Engineering and Prototyping) при разработке архитектуры БРЭО многоцелевого ударного истребителя-бомбардировщика F35 по программе Joint Strike Fighter [10]. Проведение НИОКР AVSEP позволило американским разработчикам БРЭО определить основные процедуры итерационного процесса проектирования, посредством которого проектные организации должны осуществлять переход от утвержденной концепции построения БРЭО к конкретному составу блоков (модулей), являющихся материализацией технического облика комплекса.

В нашей стране, в современных рыночных условиях проблема автоматизации решения проектных задач имеет ключевое значение для успешного функционирования предприятия приборостроительного профиля. Создание отраслевой САПР авиационного приборостроения имеет стратегическую направленность и способствует техническому и программному переоснащению проектного и производственного оборудования промышленных предприятий. Применение отраслевой САПР авиационного приборостроения жестко не связано со структурой предприятия и возможно практически на любом предприятии авиационной отрасли.

Рассмотренные элементы отечественной отраслевой САПР были апробированы на практике. Экспериментально полученные оценки длительности проведения НИОКР с использованием элементов отраслевой САПР и без них представлены на рис. 6. Нетрудно видеть, что длительность этапов проектирования комплекса БРЭО с использованием средств САПР составляет 7 лет (кривая 1), без использования средств САПР – 9 лет (кривая 2), с применением САПР и принципов преемственности, унификации и стандартизации разработок – 4,5 года (кривая 3). Таким образом, эффект от автоматизации проектных процедур выражается в сокращении продолжительности этапов проектирования БРЭО на 2–4,5 года.

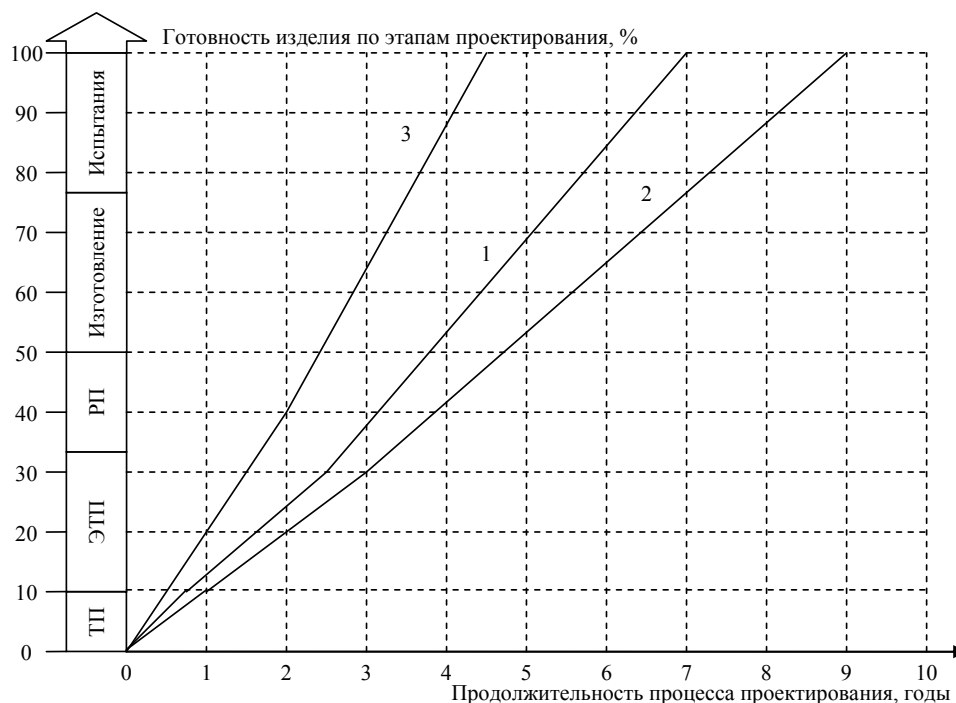


Рис. 6. Экспериментальные оценки длительности проведения НИОКР по разработке, производству и испытаниям комплексов БРЭО: а) с использованием средств САПР (кривая 1); б) без использования средств САПР (кривая 2); в) с использованием принципов унификации и стандартизации аппаратно-программных разработок предприятия (кривая 3): ТП – техническое предложение; ЭТП – эскизно-технический проект (объединение этапов эскизного и технического проектирования); РП – рабочее проектирование) на основе отраслевой САПР

Литература

1. Гатчин Ю.А., Жаринов И.О., Жаринов О.О. Архитектура программного обеспечения автоматизированного рабочего места разработчика бортового авиационного оборудования // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 2 (78). – С. 140–141.
2. Гатчин И.Ю., Жаринов И.О., Жаринов О.О., Косенков П.А. Реализация жизненного цикла «проектирование–производство–эксплуатация» бортового оборудования на предприятиях авиационной промышленности // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 2 (78). – С. 141–143.
3. Андреев Л.В., Богословский С.В., Видин Б.В., Жаринов И.О., Жаринов О.О., Парамонов П.П., Сабо Ю.И. Формализация вектора наблюдений измерительного комплекса беспилотных летательных аппаратов // Изв. вузов. Приборостроение. – 2009. – Т. 52. – № 11. – С. 23–27.
4. Сабо Ю.И., Жаринов И.О. Критерий подобия проектных решений требованиям технического задания в авионике // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 3 (67). – С. 57–63.
5. Видин Б.В., Жаринов И.О., Жаринов О.О. Декомпозиционные методы в задачах распределения вычислительных ресурсов многомашинных комплексов бортовой авионики // Информационно-управляющие системы. – 2010. – № 1. – С. 2–5.
6. Гатчин Ю.А., Видин Б.В., Жаринов И.О., Жаринов О.О. Метод автоматизированного проектирования аппаратных средств бортового оборудования // Изв. вузов. Приборостроение. – 2010. – Т. 53. – № 5. – С. 5–10.
7. Шек-Иовсепянц Р.А., Жаринов И.О. Генерация проектных решений бортового оборудования с использованием аппарата генетических алгоритмов // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 3 (67). – С. 67–70.
8. Суслов В.Д., Шек-Иовсепянц Р.А., Видин Б.В., Жаринов И.О., Немолочнов О.Ф. К вопросу об унификации бортовых алгоритмов комплексной обработки информации // Изв. вузов. Приборостроение. – 2006. – Т. 49. – № 6. – С. 39–40.
9. Алакоз Г.М., Котов А.В., Курак М.В. и др. Вычислительные наноструктуры: в 2 ч. Ч. 1. Задачи, модели, структуры. – М.: Интернет-Университет Информационных Технологий. БИНОМ. Лаборатория знаний, 2010. – 487 с.
10. Craig E. Steidle. The Joint Strike Fighter Program // Johns Hopkins Apl Technical Digest. – 1997. – № 1. – V. 18. – P. 6–20.

Парамонов Павел Павлович	– ФГУП «СПб ОКБ «Электроавтоматика» имени П. А. Ефимова», доктор технических наук, профессор, директор, postmaster@elavt.spb.ru
Гатчин Юрий Арменакович	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, gatchin@mail.ifmo.ru
Жаринов Олег Олегович	– Санкт-Петербургский государственный университет аэрокосмического приборостроения, кандидат технических наук, доцент, zharinov@hotmail.ru
Жаринов Игорь Олегович	– ФГУП «СПб ОКБ «Электроавтоматика» имени П. А. Ефимова», доктор технических наук, доцент, руководитель учебно-научного центра, igor_rabota@pisem.net
Дейко Михаил Сергеевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, postmaster@elavt.spb.ru

УДК 004.421.2

ВЫЧИСЛЕНИЕ МИНИМАЛЬНОГО ПО ДЛИНЕ ПУТИ ПРОВОДНИКА В ТОПОЛОГИЧЕСКОЙ ТРАССИРОВКЕ ПЕЧАТНОГО МОНТАЖА

Ю.И. Попов, С.И. Попов

В топологическом подходе к трассировке печатных плат возникает задача вычисления пути (формы) проводника по его топологическому пути. Предложен алгоритм, вычисляющий путь минимальной длины с соблюдением конструктивно-технологических ограничений. Алгоритм используется при расчете формы проводников в системе автоматизированного проектирования TороR.

Ключевые слова: САПР, автоматическая трассировка, топологическая трассировка.

Введение

Существует два основных подхода к автоматической трассировке печатных плат: геометрический и топологический. Первый предлагает на протяжении процесса трассировки оперировать метрикой коммутационного пространства, т.е. такими понятиями, как *координаты, длина, ширина, расстояние*. Вторым предлагает использовать только такие понятия, как *лежать слева, справа, между; обход по* или *против часовой стрелки* [1].

В 1970-х г.г. Р.П. Базилевич [2] предложил во втором подходе разделить процесс трассировки на два этапа: топологический и геометрический. На первом этапе проводники назначаются в широкие области, в которых они не фиксированы жестко, а лишь задано их относительное положение. На втором этапе – метризации – определяется конкретное положение сегментов проводника с соблюдением конструктивно-технологических ограничений.

В отличие от геометрического подхода, когда на конфигурацию проводника оказывают влияние только соседние элементы топологии, положение которых, как правило, фиксировано, при топологическом подходе на путь (форму) проводника может влиять элемент, находящийся от него на произвольном расстоянии. Проблема состоит в точном учете влияния элементов топологии при вычислении минимального по длине пути проводника.

На форму отдельно взятого сегмента проводника, как правило, влияет лишь небольшая часть элементов топологии. По этой причине для повышения эффективности представляет интерес заранее определить элементы, которые могут оказать влияния на форму сегмента, и не рассматривать остальные при ее вычислении. Предложенный алгоритм реализует эту идею и вычисляет минимальные по длине пути проводников с соблюдением конструктивно-технологических ограничений.

Минимальный по длине путь проводника

Для представления минимального по длине пути проводника рассмотрим следующий пример из [1]. Пусть в некотором зале расставлено множество круглых колонн разного диаметра. Путник, петляя между колоннами, прошел из точки A зала в точку B длинным маршрутом, ни разу не пересекая собственный путь. По дороге он разматывал клубок ниток. Существует бесконечное множество путей, топологически эквивалентных маршруту путника, т.е. проходящих между теми же парами колонн. Но среди этих путей существует кратчайший. Этот путь можно узнать, натянув нить. Нить примет форму единственной кривой, и длина этой кривой будет минимально возможной. Натянутая нить будет огибать колонны по дугам окружностей с соответствующими радиусами, а с одной окружности на другую будет переходить по прямой, касательной к обеим окружностям (рис. 1).



Рис. 1. Натянутая нить огибает колонны по дугам окружностей: нить путника (а); натянутая нить путника (б)

Печатная плата, в общем случае, имеет несколько слоев. Каждый из них может быть рассмотрен независимо. Под проводником в работе понимается фрагмент трассы, ограниченный точками ветвления, переходными отверстиями, контактными площадками компонентов, областями металлизации. Проводник прокладывается с соблюдением конструктивно-технологических ограничений (ширины проводников, размеры объектов печатного монтажа, зазоры между ними). Например, если путь проводника w проходит мимо контактной площадки C , то следует учесть ее размеры, другие проводники (с их ширинами), проходящие между w и C , а также требуемые зазоры между ними.

История вопроса

Первый подход к решению задачи вычисления минимального по длине пути предложил в 1980 г. М. Томпа (M. Tompa) [3]. Однако алгоритм имеет ряд существенных ограничений: узлы компонентов упорядочены и находятся на параллельных сторонах канала, ширины проводников нулевые, минимальные зазоры между проводниками одинаковы.

В 1988 г. Ш. Гао (S. Gao) с соавторами опубликовал работу [4]. В ней предложен более сложный алгоритм, однако также с существенными ограничениями на входные данные: одинаковые зазоры между проводниками, нулевые ширины, наличие лишь точечных контактных площадок компонентов.

В 1999 г. Т. Хама (T. Hama) и Х. Этох (H. Etoh) предложили новый подход к решению задачи [5]. Они используют триангуляцию Делоне для разбиения монтажного пространства. Условие Делоне (окружность, описанная вокруг треугольной грани не содержит других вершин триангуляции) они используют как ограничение для расстояния, на котором учитывается влияние контактных площадок на форму проводника. На реальных платах зазор между контактной площадкой и проводником может быть произвольно большим, и соответственно расчет будет некорректным. Кроме того, в предложенном подходе рассматриваются только круглые контактные площадки.

В системе автоматизированного проектирования (САПР) FreeStyle Router в 1990-х г.г., а затем и в САПР ТороR (вплоть до версии 5.2) был реализован алгоритм, позволяющий вычислять пути проводников [1]. Однако алгоритм не учитывал влияние на путь проводника контактных площадок, расположенных на достаточном удалении от него. Недочет любого фактора приводит к получению конфигураций с нарушениями конструктивно-технологических ограничений.

Представление объектов печатного монтажа набором круглых примитивов

При работе с объектами печатного монтажа удобно использовать разбиение монтажного пространства, например, в виде триангуляции Делоне или квазитриангуляции [1]. Объекты произвольной формы в разбиении представлены не одной вершиной, а набором вершин, соответствующих вершинам многоугольника, а в ряде случаев и дополнительными вершинами, например, расположенными вдоль линейно протяженных объектов. На рис. 2 приведено возможное расположение вершин на границе прямоугольного объекта печатного монтажа.

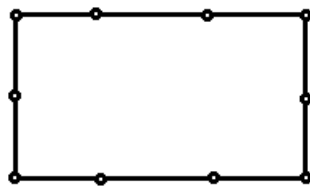


Рис. 2. Расположение вершин разбиения на границе прямоугольного объекта

С этими вершинами можно связать круглые примитивы, радиус которых зависит от выбранной модели. Между проводником, проходящим мимо объекта, и примитивом должен быть соблюден зазор, заданный конструктивно-технологическими ограничениями. Чтобы проводник не «провис» между двумя соседними огибаемыми примитивами одного объекта под действием других примитивов (возможно, другого объекта), следует учитывать расстояние до ребра, соединяющего вершины этих примитивов.

Таким образом, «некруглость» объектов просто увеличивает число вершин разбиения, поэтому в дальнейшем изложении без потери общности принимаем, что имеются только круглые контактные площадки, к ним приравняем переходные отверстия, точки ветвления и другие круглые примитивы, разбивающие более сложные объекты. Разбиение монтажного пространства – триангуляция.

Алгоритм вычисления минимального по длине пути

После топологического этапа пути проводников могут быть представлены последовательностью ребер, которые они пересекают. Рассмотрим топологический путь проводника как последовательность отрезков, проходящих, например, через центры пересекаемых ребер.

В общем случае каждая контактная площадка на печатной плате может повлиять на длину пути. Рассмотрим расстояние d между центром круглой контактной площадки C и средней линией сегмента проводника w . Оно должно быть не меньше расстояния, складывающегося из конструктивно-технологических ограничений:

- радиус контактной площадки C ;
- ширины проводников, расположенных между w и C ;
- половина ширины проводника w ;
- зазоры между ними.

Для каждого отрезка топологического пути проводника определяются дуги окружностей, центры которых расположены в центрах контактных площадок, а радиусы равны расстояниям d . Путь проводника не может пересекать дуги, так как они построены по конструктивно-технологическим ограничениям.

Алгоритм вычисления минимального по длине пути состоит из двух этапов.

1. Построение вышеописанных окружностей.
2. Вычисление минимального по длине пути, огибающего эти окружности.

Рассмотрим сначала вычисление пути.

Алгоритм вычисления пути, огибающего окружности

Основная идея алгоритма состоит в следующем [1, 3–5]. Пусть дано множество окружностей, упорядоченных вдоль топологического пути. Если задано направление пути, то окружности можно разделить на левые (D , C) и правые (A , B) относительно пути (рис. 3). К окружностям в порядке очередности строятся касательные.

Сначала следует построить касательные к первым левой (D) и правой (A) окружностям, проходящие через начальную точку S . В результате образуется воронка, в пределах которой должен располагаться путь. Если построить касательную из точки S к следующей правой окружности B , то окружность A можно исключить из рассмотрения, так как она находится правее. Касательная из точки S к окружности

C пересекает окружность D , поэтому последняя становится частью воронки (рис. 3, а). Далее аналогично строятся касательные к следующим левым и правым окружностям.

Если правая касательная пересекает левую окружность (рис. 3, б), то левая касательная (ее отрезок) из точки S к этой окружности становится геометрически частью вычисляемого пути (рис. 3, в). В этом случае необходимо далее строить общие касательные с окружностью D (вместо проходящих через точку S). Для левой касательной построение аналогично.

Если два отрезка касательных добавлены в путь, то дуга окружности между ними также становится частью пути.

Полученная последовательность отрезков касательных и дуг окружностей составляет путь минимальной длины.

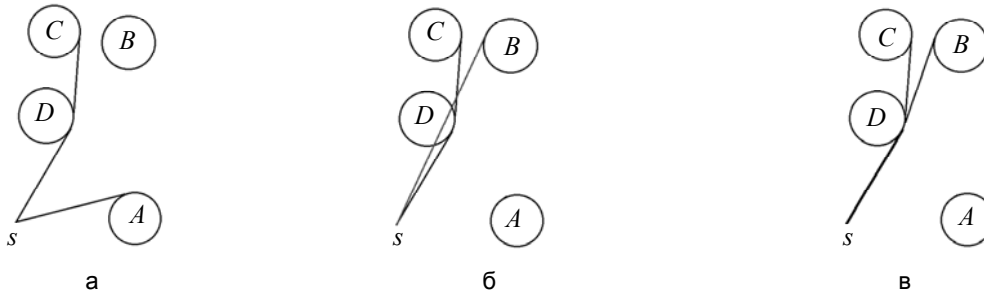


Рис. 3. Вычисление пути: построение касательных (а); пересечение касательной и окружности (б); отрезок касательной добавлен в путь (в)

Построение окружностей

Предлагается строить окружности в два этапа. На первом этапе для каждой контактной площадки следует построить множество окружностей. Процесс их построения можно представить как волну, распространяющуюся от центра контактной площадки, поглощая встречающиеся проводники и разделяясь при встрече других контактных площадок. Эти окружности не предназначены для непосредственного вычисления путей проводников, они помогут определить контактные площадки, которые влияют на форму сегментов. На втором этапе, используя эту информацию, следует построить окружности для каждого проводника в отдельности. Опишем этапы по порядку.

На первом этапе для каждой вершины v (соответствующей контактной площадке) рассматривается каждая грань, которой принадлежит вершина v , и выполняются следующие шаги.

1. Из центра контактной площадки A построить две касательные к окружностям контактных площадок B и C (рис. 4). (Если контактные площадки пересекаются, то перейти к следующей грани.) Рассматриваются касательные, пересекающие ребро b . Они вырезают в плоскости сектор, внутри которого дуга окружности может влиять на минимальный по длине путь проводника. Этот сектор позволяет пространственно ограничить направления, в которых контактная площадка влияет на проводники (сектор в процессе выполнения алгоритма может только уменьшаться).

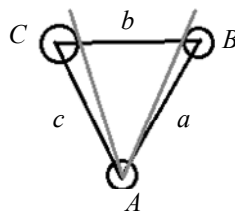


Рис. 4. Из центра A к окружностям контактных площадок строятся две касательные

2. Вычислить радиусы окружностей с учетом путей проводников, пересекающих грань. На рис. 5 изображена контактная площадка A и три проводника: $w1$, $w2$ и $w3$. Здесь известно только, в каком порядке расположены пути проводников относительно друг друга и контактных площадок. Вычислим последовательно радиусы окружностей для каждого из проводников, необходимые для соблюдения конструктивно-технологических ограничений. Первый радиус (для проводника $w1$) – это сумма радиуса контактной площадки A , зазора до проводника $w1$ и половины ширины проводника $w1$. Второй радиус (для $w2$) – это сумма первого радиуса, зазора между проводниками $w1$ и $w2$, а также половины ширины проводника $w2$. Последний радиус может влиять на проводник $w3$ за ребром b .

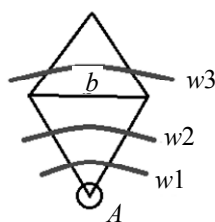


Рис. 5. Вычисление радиусов с учетом контактной площадки А и проводников w1, w2 и w3

3. Если последняя окружность с вычисленным радиусом пересекает ребро b , то запомнить информацию о ней в ассоциации с ребром (она будет использована на шаге 2 второго этапа). За ребром b начальной грани находятся грань и два ребра (рис. 6). Через них следует продолжить распространение вычисления радиусов. При этом необходимо построить касательные к окружности контактной площадки за ребром b , так как они могут уменьшить сектор. На рис. 6 за ребром b находятся ребра d и e . Проводник $w2$ не влияет на радиус окружности для ребра e (не будет учтен при проверке на пересечение). Для ребра d ситуация сложнее. Если бы проводник $w2$ пересекал ребро e , то можно было бы однозначно указать на необходимость его учета. В приведенном случае проводник учитывается условно. Если за ребром d на расстоянии, учитывающем ширину проводника $w2$ и соответствующие зазоры, нет других проводников, то процесс распространения в данном направлении будет остановлен. Продолжать вычисление радиусов (за ребра e и d) следует, пока зазоры и ширины проводников могут его увеличить, до пересечения очередного ребра.

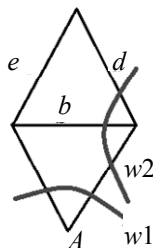


Рис. 6. Учет проводника w2 при вычислении радиусов окружностей

На втором этапе по очереди рассматриваются топологические пути проводников (последовательности отрезков). Пересекая грань, топологический путь через одно ребро входит, а через другое выходит из нее. Существуют не более двух граней, в которых путь пересекает только одно ребро, – грани, в вершинах которых путь начинается и заканчивается. Для каждого топологического пути выполняются следующие шаги.

1. Каждое пересекаемое топологическим путем ребро инцидентно двум вершинам. Между контактной площадкой вершины и проводником w находятся другие проводники. Для каждой из вершин необходимо построить окружности с радиусом, равным сумме радиуса контактной площадки, ширин проводников и зазоров между ними. Последним учтенным в радиусе будет рассматриваемый отрезок топологического пути.
2. Для каждого отрезка топологического пути по триангуляции определить окружности, построенные на первом этапе. Эти окружности могут повлиять на минимальный по длине путь, но их радиусы следует изменить таким образом, чтобы последним учтенным в радиусе стал рассматриваемый отрезок топологического пути.

Дуги построенных окружностей, ограниченные секторами, учитывают все необходимые конструктивно-технологические ограничения, поэтому можно применить алгоритм вычисления пути, огибающего окружности.

Заключение

В описанном подходе, в отличие от предложенных ранее, отсутствует ограничение на форму канала, учитываются ширины всех проводников и необходимые зазоры, нет ограничений на их величины. Объекты произвольной формы моделируются набором круглых примитивов с обработкой возможного «провисания» проводников. При этом учет перечисленных факторов не сопровождается снижением производительности.

Асимптотические оценки алгоритма: $O(B \cdot F^2 \cdot \log(B \cdot F))$ – время; $O(B \cdot F^2)$ – память, где B – общее число изломов путей проводников, состоящих из отрезков между ребрами; F – число контактных площадок на плате – совпадают с оценками алгоритма из [4]. Но в [4] рассматривается влияние каждой контактной площадки на каждый проводник посредством простого перебора и упорядочивания. При

этом строятся заранее все возможные окружности для всех проводников. Предложенный алгоритм строит окружности итеративно, заранее определяется, что в продолжении итераций нет необходимости. Таким образом, предложенный метод не только свободен от существенных недостатков предыдущих версий, но и превосходит их в эффективности. Алгоритм реализован и используется в последней версии системы автоматизированного проектирования ТороR.

Литература

1. Лузин С.Ю., Лячек Ю.Т., Петросян Г.С., Полубасов О.Б. Модели и алгоритмы автоматизированного проектирования радиоэлектронной и электронно-вычислительной аппаратуры. – СПб: БХВ-Петербург, 2010. – 224 с.
2. Базилевич Р.П. Декомпозиционные и топологические методы автоматизированного конструирования электронных устройств. – Львов: Вища школа, 1981. – 168 с.
3. Tompa M. An Optimal Solution to a Wire-Routing Problem // Journal of Computer and System Sciences 23. – 1980. – P. 127–150.
4. Gao S., Jerrum M., Kaufmann M., Mehlhorn K., Rulling W., Storb C. On continuous homotopic one layer routing // in Proc. 4th Annu. ACM Symp. Computational Geometry. – 1988. – P. 392–402.
5. Hama T., Etoh H. Curvilinear Detailed Routing with Simultaneous Wire-Spreading and Wire-Fattening // IEEE Transactions on computer-aided design of integrated circuits and systems. – 1999. – V. 18. – № 11. – P. 1646–1653.

Попов Юрий Игоревич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, uig-porov@rambler.ru

Попов Сергей Игоревич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, sergey.popov@yandex.ru

УДК 004.912

**АЛГОРИТМ ОПРЕДЕЛЕНИЯ СПАМНОСТИ ДОКУМЕНТОВ НА ОСНОВЕ
ФЕЙЕРОВСКИХ ОТОБРАЖЕНИЙ**

А.Г. Коробейников, С.Ю. Блинов, А.В. Лейман, Г.Л. Маркина, И.М. Кутузов

Предложен алгоритм определения спамности почтовых документов на базе метода опорных векторов. Модификация метода заключается в построении разделяющей гиперплоскости на основе фейеровских отображений, что позволяет работать с нестационарными данными, характерными для задач классификации документов.

Ключевые слова: классификация информации, спам, задачи сильной отделимости, гильбертово пространство.

Введение

Информационные технологии породили все увеличивающийся поток разнородной информации. Основной задачей поисковых систем (поисковых машин) является предоставление качественных результатов, т.е. наиболее важных релевантных страниц. Для этого необходимо решать задачу классификации. В связи с этим теория, методы и алгоритмы классификации информации являются бурно развивающимся научным направлением.

Одной из важнейших проблем, встающей практически перед каждым пользователем Интернет, является борьба со спамом, т.е. задача фильтрации (классификации) поступающей информации. В настоящее время разработан ряд технологий построения фильтров – сервисов для отсеивания нежелательной информации. Эти технологии можно разделить на настраиваемые вручную и автоматизированные. Настраиваемые вручную фильтры основываются на списках доступа: пользователь непосредственно выбирает либо нежелательные адреса (при политике пропуска по «черному списку»), либо разрешенные адреса (при политике пропуска по «белому списку»). Однако ручные способы фильтрации нежелательных сообщений малоэффективны и требуют постоянного обновления списков доступа, создавая дополнительную нагрузку на пользователя. Кроме того, ручная категоризация неприменима, если необходимо классифицировать большой объем информации за ограниченное время.

Применение автоматизированных технологий фильтрации основано на использовании методов распознавания образов, искусственного интеллекта, математической статистики и т.д. [1, 2].

Одним из самых популярных на сегодняшний день является фильтр, основанный на байесовском подходе (наивный байесовский классификатор), в котором предполагается, что различные термины сообщения независимы друг от друга. Но для повышения эффективности таких фильтров необходимо учитывать семантические связи между терминами, что требует привлечения методов семантического анализа, тем самым существенно повышая нагрузку на систему и увеличивая время работы самого фильтра при незначительном повышении эффективности фильтрации.

Таким образом, существует потребность в разработке новых методов и алгоритмов классификации информации для решения задачи фильтрации нежелательных сообщений, что делает актуальной тему настоящей работы. В работе предложен алгоритм построения классификатора, базирующийся на построении разделяющей гиперплоскости в гильбертовом пространстве на основе фейеровского отображения [3, 4].

Постановка задачи классификации

Формально постановка задачи классификации выглядит следующим образом.

Пусть дано конечное множество категорий (классов) $C = \{c_1, c_2, \dots, c_{|C|}\}$ и конечное множество документов $D = \{d_1, d_2, \dots, d_{|D|}\}$. Целевая функция (функционал, классификатор) $\Phi : D \times C \rightarrow \{-1, 1\}$, определяющая для каждой пары <документ, категория> соответствие их друг другу, неизвестна. Требуется найти классификатор Φ' , т.е. функцию, максимально близкую к функции Φ [4]. Если пересечение двух категорий пусто, то имеет место бинарная классификация, которая часто используется в фильтрации спама. Если имеются образцы из каждой категории (объекты) и заранее известно, к какой категории они принадлежат, то такие задачи называются обучением с учителем, а известные данные – обучающей выборкой. В предлагаемом методе используется именно этот подход.

Машинное обучение предполагает наличие обучающей и контрольной выборки. Дана начальная коллекция документов $\Omega = \{d_1, d_2, \dots, d_{|\Omega|}\} \subset D$, где значения целевой функции Φ известны для $\forall (d_i, c_j) \in \Omega \times C$. Эта коллекция Ω разбивается на два непересекающихся множества: $\Omega = \Omega_1 \cup \Omega_2$ и $\Omega_1 \cap \Omega_2 = \emptyset$ (Ω_1 – спам, Ω_2 – не спам). Формирование этих подмножеств может происходить вручную, т.е. эксперты после работы с множеством документов мощности $|\Omega|$ выносят решение: если документ d_i является спамом, то он принадлежит множеству $d_i \in \Omega_1$; наоборот, если документ не спам, то $d_i \in \Omega_2$.

Обозначим обучающее множество $Q = \{(d_i, y_i)\}$, где $d_i \in \Omega$, $y_i \in C$. Классификатор Φ обучается индуктивно на основе выявленных характеристик документов [5].

Анализ метода построения классификатора

Построение классификатора будем проводить на основе метода опорных векторов (SVM – support vector machines) [6]. SVM – это набор схожих алгоритмов на основе обучения с учителем, применяющийся для анализа данных и распознавания образов в задачах классификации и регрессионном анализе. SVM является линейным классификатором. На основе обучающей выборки алгоритм помогает предсказать, в какую категорию из двух заранее заданных попадает элемент, подлежащий классификации.

Особым свойством SVM является непрерывное уменьшение эмпирической ошибки классификации и увеличение зазора. По этой причине этот метод также известен как метод классификатора с максимальным зазором. SVM считается очень перспективным для решения задач классификации (кластеризации) с использованием методов теории искусственного интеллекта.

Основная идея, используемая в SVM, – построение гиперплоскости или набора гиперплоскостей в пространстве более высокой размерности и максимизация расстояния между построенной гиперплоскостью и классами обучающей выборки. Но в настоящее время еще не разработаны общие методы построения таких гиперплоскостей (спрямляющих пространств или ядер классификатора), наиболее подходящих для конкретной задачи. Построение адекватного ядра является искусством и, как правило, опирается на априорные знания о предметной области. На практике «вполне разумные» ядра классификатора, выведенные из содержательных соображений, далеко не всегда оказываются положительно определенными, что, естественно, не может не сказаться на качестве решения. Именно построению такой гиперплоскости посвящена настоящая работа.

Разработка алгоритма решения задачи защиты от спама на базе SVM

В задаче классификации информации для задачи защиты от спама необходимо определять, является данный документ спамом или нет. В предлагаемом подходе для решения этой задачи вводится система метрик. Пользователь разделяет свои письма и отмечает те, которые считает спамом, т.е. строит обучающее множество Q . На основании этой информации строятся выпуклые оболочки в виде систем линейных неравенств. Первая система задает множество точек-документов, определяемых как спам, вторая система – множество точек-документов, определяемых как не спам. Построив слой наибольшей толщины, разделяющий два многогранника, можно разработать процедуру, позволяющую автоматически разделять (классифицировать) документы на «хорошие» и «плохие». Получив новый документ и прочитав его характеристики, получаем точку в рассматриваемом пространстве. Если данная точка попадает в «плохое» полупространство, делается предположение, что это спам; если в «хорошее» – не спам; если точка попадает внутрь слоя, письмо доставляется пользователю с пометкой «возможно, спам».

Для решения задачи сильной отделимости обычно применяют итерационный процесс, использующий операцию проектирования. Однако на практике применение этого метода существенно ограничено тем фактом, что далеко не всегда удается построить конструктивную формулу для вычисления проекции точки на выпуклое множество. В связи с этим целесообразно произвести замену операции проектирования последовательностью фейеровских отображений [7].

Кроме того, алгоритмы разделения многогранников на основе фейеровских отображений обладают тем преимуществом по сравнению с проективными и другими известными методами, что они применимы к нестационарным задачам, т.е. к задачам, в которых исходные данные могут меняться в процессе решения задачи. Такой нестационарной задачей является, например, задача о спам-филтре.

Рассмотрим алгоритмы для решения задачи сильной отделимости.

Пусть даны два выпуклых непересекающихся многогранника $M \subset R^n$ и $N \subset R^n$, заданные системами линейных неравенств

$$M = \{x | Ax \leq b\} \neq \emptyset; N = \{x | Bx \leq d\} \neq \emptyset. \quad (1)$$

Задача сильной отделимости заключается в нахождении слоя наибольшей толщины, разделяющего M и N . Эта задача эквивалентна задаче отыскания расстояния между M и N в смысле метрики:

$$\rho(M, N) = \min \{\|x - y\| \mid x \in M, y \in N\}. \quad (2)$$

Если $x_1 \in M$ и $y_1 \in N$ являются arg-точками задачи (2), т.е. $\rho(M, N) = \|x_1 - y_1\|$, то слоем наибольшей толщины, разделяющим множества M и N , является

$$P = \{x | x \in P_1 \cap P_2\},$$

где $\langle, \rangle$ – скалярное произведение двух векторов; P_1 и P_2 – полупространства, задаваемые линейными неравенствами

$$\langle x - x_1, x_1 - y_1 \rangle \leq 0 \text{ и } \langle y - y_1, x_1 - y_1 \rangle \geq 0.$$

Следовательно, задачу сильной отделимости можно сформулировать так:

$$\{x_1, y_1\} = \arg \min \{\|x - y\| \mid x \in M, y \in N\}. \quad (3)$$

Задачу (3) можно решить при помощи известного алгоритма последовательного проектирования.

Алгоритм решения задачи сильной отделимости ($\mathfrak{R}$). Даны два выпуклых непересекающихся многогранника $\mathbf{M} \subset \mathbf{R}^n$ и $\mathbf{N} \subset \mathbf{R}^n$, заданные системами линейных неравенств (1). Обозначим отображение (проектирование) точки на $\mathbf{M}$ через $\pi_{\mathbf{M}}$, а на $\mathbf{N}$ – $\pi_{\mathbf{N}}$. Зададим произвольное начальное приближение $w_0 \in \mathbf{R}^n$. Выберем фиксированное положительное вещественное число ε . Тогда алгоритм решения задачи сильной отделимости будет состоять из следующих шагов.

Шаг 0. $k := 0$.

Шаг 1. $x_{k+1} := \pi_{\mathbf{M}}(w_k)$.

Шаг 2. $y_{k+1} := \pi_{\mathbf{N}}(w_k)$.

Шаг 3. $w_{k+1} := (x_{k+1} + y_{k+1})/2$.

Шаг 4. $k := k + 1$.

Шаг 5. Если $\min\{\|x_{k+1} - x_k\|, \|y_{k+1} - y_k\|\} \geq \varepsilon$, то перейти к шагу 1.

Шаг 6. Конец.

Если множества $\mathbf{M}$ и $\mathbf{N}$ достаточно просты в смысле реализации операции проектирования точек на них, то представленный алгоритм $\mathfrak{R}$ может быть использован на практике. Но если $\mathbf{M}$ и $\mathbf{N}$ – произвольные многогранники, то $\mathfrak{R}$ не может быть признан эффективным, так как неизвестен универсальный конструктивный метод построения проекции точки на многогранник. В этом случае решить задачу можно, если вместо операции проектирования использовать фейеровские отображения.

Дадим определение фейеровского отображения. Пусть $\varphi: \mathbf{R}^n \rightarrow \mathbf{R}^n$. Отображение φ называется $\mathbf{M}$ -фейеровским, если выполняются следующие два условия:

$$\varphi(y) = y, \forall y \in \mathbf{M}; \|\varphi(x) - y\| < \|x - y\|, \forall x \notin \mathbf{M}, y \in \mathbf{M}.$$

Класс $\mathbf{M}$ -фейеровских отображений обозначим через $\mathbf{F}_{\mathbf{M}}$.

Под фейеровским процессом, порождаемым однозначным $\mathbf{M}$ -фейеровским отображением $\varphi \in \mathbf{F}_{\mathbf{M}}$ при произвольном начальном приближении $x_0 \in \mathbf{R}^n$, будем понимать последовательность $\{\varphi^k(x_0)\}_{k=0}^{+\infty}$.

Известно [8], что в случае, когда однозначное $\mathbf{M}$ -фейеровское отображение φ является непрерывным, фейеровский процесс сходится к точке, принадлежащей множеству

$$\mathbf{M}: \{\varphi^k(x_0)\}_{k=0}^{+\infty} \rightarrow \bar{x} \in \mathbf{M}.$$

Это означает, что для любого вещественного $\varepsilon > 0$ существует целое положительное число K , такое, что для всех $k > K$ имеем $\|x_k - \bar{x}\| < \varepsilon$.

Спроектируем $\mathbf{M}$ -фейеровское отображение, согласно [8]. Представим систему линейных неравенств, задающих многогранник $\mathbf{M}$, в виде

$$\mathbf{M} = \{x | Ax \leq b : l_j(x) = \langle a_j, x \rangle - b \leq 0, j = 1, \dots, m\},$$

где $\langle a_j, x \rangle \neq 0$ для любого j . Зададим $l_j^{plus}(x)$ следующим образом:

$$l_j^{plus}(x) = \max\{l_j(x), 0\}, j = 1, \dots, m.$$

Тогда отображение вида

$$\varphi(x) = x - \sum_{j=1}^m (a_j)^2 \lambda_j \frac{l_j^{plus}(x)}{\|a_j\|^2}$$

будет $\mathbf{M}$ -фейеровским для любой системы положительных коэффициентов $\{a_j > 0\}, j = 1, \dots, m$, таких, что $\sum_{j=1}^m a_j = 1$, и коэффициентов релаксации $0 < \lambda_j < 2$. Аналогичным образом сконструируем $\mathbf{N}$ -фейеровское отображение ψ . Используя отображения φ и ψ , можно построить алгоритм, решающий задачу сильной отделимости с использованием фейеровских отображений.

Определение. Пусть задано однозначное непрерывное отображение $\varphi \in \mathbf{F}_{\mathbf{M}}$. Под φ -проектированием (псевдопроектированием) точки $x \in \mathbf{R}^n$ на множество $\mathbf{M}$ будем понимать отображение $\pi_{\mathbf{M}}^{\varphi}: \mathbf{R}^n \rightarrow \mathbf{M}$, задаваемое соотношением $\pi_{\mathbf{M}}^{\varphi}(x) = \lim_{k \rightarrow \infty} \varphi^k(x)$. Точку $\pi_{\mathbf{M}}^{\varphi}(x)$ будем называть псевдо-проекцией точки x на множество $\mathbf{M}$.

Предположим, что в контексте решения задачи (3) существуют два однозначных непрерывных фейеровских отображения $\varphi \in \mathbf{F}_{\mathbf{M}}$ и $\psi \in \mathbf{F}_{\mathbf{N}}$. Используя операции φ - и ψ -проектирования, построим алгоритм Ξ , решающий задачу сильной отделимости с использованием фейеровских отображений.

Алгоритм Ξ . Зададим произвольное начальное приближение $w_0 \in \mathbf{R}^n$. Выберем фиксированное положительное вещественное число ε . Тогда алгоритм решения задачи сильной отделимости с использованием фейеровских отображений будет состоять из следующих шагов.

Шаг 0. $k := 0$.

Шаг 1. $x_{k+1} := \pi_M^{\Phi}(w_k)$.

Шаг 2. $y_{k+1} := \pi_N^{\Phi}(w_k)$.

Шаг 3. $w_{k+1} := (x_{k+1} + y_{k+1})/2$.

Шаг 4. $k := k + 1$.

Шаг 5. Если $\|w_{k+1} - w_k\| \geq \varepsilon$, то перейти к шагу 1.

Шаг 6. Конец.

Основной вопрос при обосновании применимости алгоритма Ξ к задаче классификации информации заключается в доказательстве сходимости этого алгоритма к требуемому решению. Обычно для обоснования сходимости алгоритма Ξ необходимо более сильное свойство, чем просто фейеровость отображения Φ , а именно локально сильное фейеровское отображение [9]. В этом случае сходимость алгоритма строго доказана [9]. Очевидно, что в алгоритме Ξ ресурсоемкими являются шаги 1 и 2. На каждом из этих шагов реализуется последовательный фейеровский процесс, в результате которого получаем псевдопроекцию точки на многогранник. Многогранник, задаваемый системой линейных неравенств, всегда является выпуклым замкнутым множеством, т.е. всегда существует решение.

Рассмотрим алгоритм создания обучающей выборки.

Будем производить действия по следующему алгоритму:

1. Берем обучающее множество Ω и считываем его термы. Получаем множество термов $T = \{t_1, t_2, \dots, t_p\}$, $p = 1, \dots, |T|$, $T = \bigcup d_i$, $i = 1, \dots, |D|$.
2. Проводим лексикографическое упорядочивание множества (преобразуем T в T_{base}): $T \rightarrow T_{base}$, $|T_{base}| \leq |T|$, $L_{base} = |T_{base}|$.
3. На базе обучающего множества Ω формируем частотный словарь слов (термов) Datatable, в котором каждому $t_i \in T_{base}$, $i = 1, \dots, |L_{base}|$ соответствуют числа: v_{1i} – частота встречаемости в Ω_1 , v_{2i} – частота встречаемости в Ω_2 ; $Datatable = \{t_1, v_{11}, v_{12}, t_2, v_{21}, v_{22}, \dots, t_r, v_{r1}, v_{r2}\}$, $r = L_{base}$.
4. Берем последовательно все документы $d_i \in \Omega_1$, $i = 1, \dots, |\Omega_1|$.
5. Считываем все термы из документа d_i : $TT = \{tt_1, \dots, tt_k\}$, $k = |d_i|$.
6. Проводим лексикографическое упорядочивание множества TT (преобразуем TT в TT_{base}): $TT \rightarrow TT_{base}$, $|TT_{base}| \leq |TT|$, $l_{base} = |TT_{base}|$.
7. На базе множеств T_{base} , TT_{base} , частотного словаря и вектора признаков формируем вектор $x_i = \{v_{11}, v_{12}, v_{21}, v_{22}, \dots, v_{p1}, v_{p2}, sv_1, \dots, sv_n\}$, где $p = L_{base}$, и $v_{ij} = 0$, если $tt_i \in TT \cap T_{base} = \emptyset$. Обозначим множество векторов x_i через $X = \{x_i\}$, $i = 1, \dots, |\Omega_1|$. Получаем множество, соответствующее спаму.
8. Повторяем пункты 4–7, но только для множества Ω . Обозначим множество, полученное на шаге 8, через Y . Получаем множество, соответствующее не спаму.
9. Строим по алгоритму Ξ разделяющую гиперплоскость между X и Y . Получаем нормальный вектор к гиперплоскости w и параметр b (порог классификации).
10. Строим классифицирующую функцию Φ' .

Окончательно алгоритм определения спамности документа выглядит так.

1. Берем документ $d_i \in D \setminus \Omega$, $i = 1, \dots, |D \setminus \Omega|$.
2. Считываем все термы из документа d_i : $TT = \{tt_1, \dots, tt_k\}$, $k = |d_i|$.
3. Проводим лексикографическое упорядочивание множества TT (преобразуем TT в TT_{base}): $TT \rightarrow TT_{base}$, $|TT_{base}| \leq |TT|$, $l_{base} = |TT_{base}|$.
4. Определяем множество признаков для данного документа. Если данный признак не определен, то присваиваем $sv_i = 0$.
5. На базе множеств T_{base} , TT_{base} , частотного словаря и множества признаков формируем вектор $x = \{v_{11}, v_{12}, v_{21}, v_{22}, \dots, v_{p1}, v_{p2}, sv_1, \dots, sv_n\}$, где $p = L_{base}$, $v_{ij} = 0$, если $tt_i \in TT \cap T_{base} = \emptyset$.
6. Определяем спамность документа d_i при помощи классификатора $\Phi'(x) = \text{sign}(\langle w, x \rangle + b)$.

Заключение

Разработанный алгоритм определения спамности документа, или классификации, конечно же, имеет свои недостатки. Попытки построить для решения задачи классификации один алгоритм, удовлетворяющий всех, заранее обречены на провал. По этой причине имеет смысл объединить несколько алгоритмов в композицию, в надежде на то, что погрешности различных алгоритмов взаимно компенсируются. Но в этом случае возникает масса вопросов. Например, при каких условиях качество композиции окажется лучше, чем у отдельных базовых алгоритмов? Как настраивать базовые алгоритмы, учитывая,

что они будут работать в составе композиции? Возможно ли приспособить для их настройки стандартные методы обучения? Как обойтись минимальным числом базовых алгоритмов?

Формально эти вопросы можно записать следующим образом. Пусть имеется задача обучения по прецедентам $\{X, Y, y^*, X^t, Y^t\}$, где X – множество объектов; Y – множество ответов; $y^* : X \rightarrow Y$ – отображение (неизвестная целевая зависимость); $X^t = (x_1, \dots, x_l)$ – обучающая выборка; $Y^t = (y_1, \dots, y_l)$ – вектор ответов на обучающих объектах, $y_i = y^*(x_i)$. Требуется построить алгоритм $alg: X \rightarrow Y$, аппроксимирующий целевую зависимость y^* на всем множестве X .

Введем множество R , называемое пространством оценок [10]. Рассмотрим алгоритмы, имеющие вид суперпозиции

$$alg(x) = \Pi(b(x)),$$

где $b : X \rightarrow R$ – алгоритмический оператор; $\Pi : R \rightarrow Y$ – решающее правило.

Многие алгоритмы классификации имеют именно такую структуру: сначала вычисляются оценки принадлежности объекта категориям (классам), а затем решающее правило переводит эти оценки в номер класса. Значение оценки, как правило, характеризует степень уверенности классификации. В одних алгоритмах это вероятность принадлежности объекта заданному классу, в других – расстояние от объекта до разделяющей поверхности. Возможны и другие интерпретации оценок. В работе был использован вариант

$$Y = \{-1, +1\}, \Pi(z) = \text{sign}(z), alg(x) = \Phi'(x) = \text{sign}(\langle w, x \rangle + b).$$

Дальнейшие научные исследования будут посвящены теоретической разработке построения

$$alg(x) = \Pi(b(x)),$$

которые позволят повысить эффективность фильтрации спама.

Литература

1. Сидоркина И.Г., Коробейников А.Г., Кудрин П.А. Алгоритм распознавания трехмерных изображений с высокой детализацией // Вестник Марийского государственного технического университета. – 2010. – № 2 (9). – С. 91–99.
2. Гришенцев А.Ю., Коробейников А.Г. Декомпозиция N-мерных цифровых сигналов по базису прямоугольных всплесков // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 4 (80). – С. 75–79.
3. Халмош П. Гильбертово пространство в задачах: Пер. с англ. И.Д. Новикова и Т.В. Соколовской / Под ред. Р.А. Минлоса. – М.: Мир, 1970. – 352 с.
4. Еремин И.И. Фейеровские методы сильной отделимости выпуклых полиэдральных множеств // Изв. вузов. Сер. математика. – 2006. – № 12. – С. 33–43.
5. Лифшиц Ю. Классификация текстов. Алгоритмы для Интернета. – 2005 [Электронный ресурс]. – URL: <http://yury.name/internet>, свободный. Яз. рус. (дата обращения 24.09.2012).
6. Fan R.-E., Chen P.-H., Lin C.-J. Working set selection using second order information for training SVM // Journal of Machine Learning Research. – 2005. – V. 6. – P. 1889–1918.
7. Ерёмин И.И., Мазуров В.Д. Нестационарные процессы математического программирования. – М.: Наука, 1979. – 288 с.
8. Ерёмин И.И. Теория линейной оптимизации. – Екатеринбург: УрО РАН, 1999. – 312 с.
9. Нурминский Е.А. Использование дополнительных малых воздействий в фейеровских моделях итеративных алгоритмов // Ж. вычисл. матем. и матем. физ. – 2008. – Т. 48. – № 12. – С. 2121–2128.
10. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. – 1978. – Т. 33. – С. 5–68 [Электронный ресурс]. – Режим доступа: <http://www.ccas.ru/frc/papers/zhuravlev78prob33.pdf>, свободный. Яз. рус. (дата обращения 24.09.2012).

Коробейников Анатолий Григорьевич

– Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова Российской академии наук (СПбФ ИЗМИРАН), доктор технических наук, профессор, зам. директора, Korobeynikov_A_G@mail.ru

Блинов Станислав Юрьевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, stasblino@yandex.ru

Лейман Альберт Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, allxxl@yandex.ru

Маркина Галина Леонидовна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, начальник отдела, lina@mail.ifmo.ru

Кутузов Илья Михайлович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, formalizator@gmail.com

УДК 004.931

АНАЛИЗ СПОСОБОВ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ В СЕТИ ИНТЕРНЕТ

Е.Е. Бессонова, И.А. Зикратов, В.Ю. Росков

Рассматриваются механизмы идентификации пользователей в сети Интернет. Предложен сравнительный анализ способов идентификации на основе регрессионного анализа и энтропийного подхода. Для проверки полученных результатов проведен вычислительный эксперимент.

Ключевые слова: идентификация, информативность, признак, кортеж, пользователь.

Введение

Одной из основных задач современной теории и техники автоматического управления является задача идентификации систем, т.е. определение структуры и параметров систем по наблюдениям [1].

В частности, в теории защиты информации актуальной является вопрос идентификации пользователя в сети Интернет. Актуальность данной темы обусловлена целесообразностью идентификации субъектов Сети при построении системы защиты информации для выявления нарушителей.

Целью данной работы является сравнительный анализ способов идентификации пользователей.

Для современных информационных систем применяются способы идентификации, основанные на хранении IP-адресов компьютеров посетителей и записи на компьютер пользователя данных Cookie. Однако оба способа не позволяют в ряде случаев достичь требуемой степени достоверности идентификации [2]. В работе [3] показан способ идентификации, основанный на регрессионном анализе, что позволяет осуществить рациональный выбор признаков, необходимых для повышения степени достоверности идентификации пользователя в сети Интернет. В процессе работы на основании сформированного признакового пространства с помощью регрессионного анализа были выбраны наиболее информативные признаки. Под информативностью признаков понимается степень влияния признака в кортеже идентифицируемого объекта на результат отождествления с имеющимися профилями пользователей.

Проведен сравнительный анализ метода идентификации пользователя по его кортежу путем сравнения его с накопленной базой. Результатом работы является кортеж, состоящий из восьми наиболее информативных идентификаторов: ETag, Supercookie, Cookie, MAC, IP, шрифты через Flash, плагины, шрифты через ActiveX. По сравнению с Cookie, такой кортеж обеспечивает в 6,3 раза большую информативность (4,35 против 0,69).

Кроме способа, основанного на методе регрессионного анализа, авторами исследован также метод формирования признакового пространства, основанный на вычислении количества энтропии [4].

Энтропия – это количество информации, приходящейся на одно элементарное сообщение источника, вырабатывающего статистически независимые сообщения. Для расчета энтропии Шеннон предложил уравнение

$$H(X) = - \sum_{i=0}^N P \log P(X_i),$$

где X – дискретная случайная величина с диапазоном изменчивости N ; $P(X_i)$ – вероятность i -го уровня X [5]. Количество энтропии, которое содержит в себе признак, можно вычислить по формуле

$$\Delta S = \log_2 P_{x \in X}(x)$$

где $P(x)$ – вероятность появления значения x признака X .

Понятие энтропии может быть использовано для оценки признаков профиля пользователя. Значения энтропии отдельных признаков, которые приводит исследование, сделанное Electronic Frontier Foundation, отображены в таблице [6].

Для сравнительного анализа методов, основанных на регрессионном анализе и энтропийном подходе, был проведен эксперимент.

Наименование признака	Энтропия, бит
Заголовок Http User-Agent	10,0
Список установленных плагинов	15,4
Список установленных шрифтов	13,9
Поддержка supercookies	2,12
Заголовок Http Accept	6,09
Временная зона	3,04
Включенность cookies	0,353

Таблица. Количество энтропии информативных признаков (по данным Electronic Frontier Foundation)

Проведение эксперимента

Авторами был сделан сравнительный анализ двух кортежей, полученных в указанных выше работах, по степени их достоверности и скорости идентификации пользователей. Для этого был проведен эксперимент с целью вычисления степени достоверности идентификации и времени работы. В качестве результатов эксперимента получены временные характеристики работы двух кортежей, а также зависимость количества идентифицированных пользователей от уровня шума для обоих кортежей.

В качестве входных данных были использованы: учетные записи, выбранные в случайном порядке (эталон); статистика учетных записей пользователей, заходящих на тестовый сайт не менее двух раз; признаки, упорядоченные по возрастанию информативности.

Целью эксперимента являлось определение степени достоверности работы двух кортежей и их скорости обработки данных. Результаты эксперимента представлены на графиках (рис. 1).

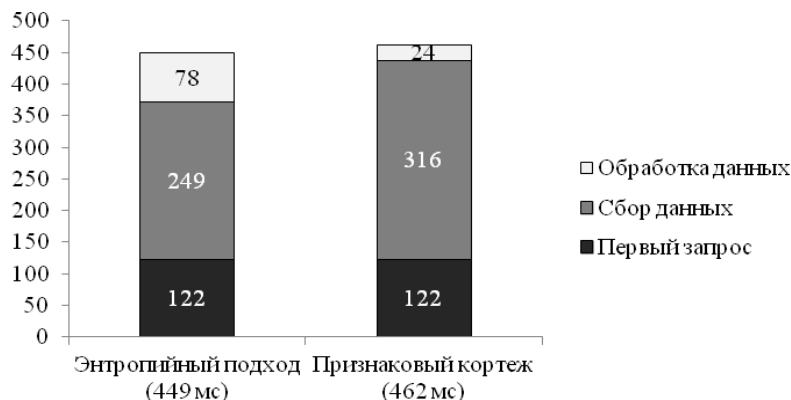


Рис. 1. Сравнение временных характеристик процессов обработки кортежей

В представленных на графике результатах можно выделить три временных интервала:

1. первый запрос – сколько времени занимает загрузка страницы, к которой подключен скрипт с данными;
2. сбор данных – время от начала работы скрипта до момента отправки их на сервер;
3. обработка данных – время с момента отправки данных на сервер до получения результата идентификации.

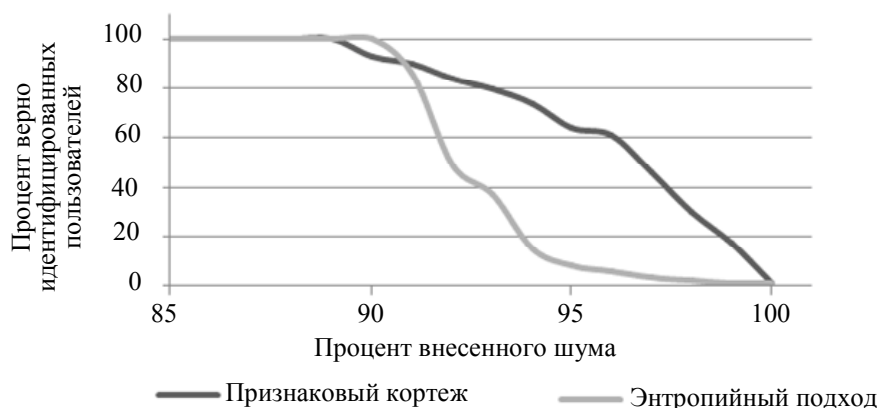


Рис. 2. Сравнение достоверности идентификации для различных способов

Данные получены при усреднении 10 000 запросов. Очевидно, что длительность первого запроса к странице во всех случаях одинакова. Так как у энтропийного кортежа меньше признаков, сбор данных происходит быстрее. Подсчет энтропии занимает больше времени, чем предложенный авторами метод идентификации. Таким образом, экспериментально полученная скорость работы обоих методов отличается незначительно. При использовании признакового кортежа время увеличивается на 2,9%.

Эксперимент показал, что при внесенном шуме, составляющем 89%, оба подхода демонстрируют одинаковую эффективность. Однако при внесенном шуме более 90% энтропийный подход резко ухудшает свои результаты и при 95% шума может идентифицировать менее 10% пользователей, тогда как метод идентификации, предложенный в работе [3], позволяет идентифицировать в 6 раз больше пользователей (рис. 2).

Заключение

Таким образом, для агрессивной среды более подходящим для использования является предложенный авторами метод, так как он показал более высокую степень достоверности. Для неагрессивной среды представляется возможным использовать энтропийный подход, так как он имеет более высокое быстроедействие. Однако следует учитывать, что разность в быстроедействии процессов идентификации при реализации рассматриваемых методов незначительна.

Литература

1. Цыпкин Я.З. Информационная теория идентификации. – М.: Наука. ФИЗМАТЛИТ, 1995. – 336 с.
2. McKinkley K. Cleaning Up After Cookies. iSec Partners White Paper [Электронный ресурс]. – Режим доступа: http://www.isecpartners.com/storage/white-papers/iSEC_Cleaning_Up_After_Cookies.pdf, свободный. Яз. англ. (дата обращения 15.10.2011).
3. Бессонова Е.Е., Зикратов И.А., Колесников Ю.Л., Росков В.Ю. Способ идентификации пользователя в сети Интернет // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 133–137.
4. Eckersley P. How Unique Is Your Web Browser? – Electronic Frontier Foundation, 2010. – 19 p.
5. Шеннон К. Работы по теории информации и кибернетики. – М.: Издательство иностранной литературы, 1963. – 830 с.
6. Eckersley P. A Primer on Information Theory and Privacy. – Electronic Frontier Foundation. – 2010. – 25 p.

Бессонова Екатерина Евгеньевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, bessonova@cit.ifmo.ru

Зикратов Игорь Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, zikratov@cit.ifmo.ru

Росков Владислав Юрьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, vos@vos.uz

УДК 004.056

АЛГОРИТМ ДЕПЕРСОНАЛИЗАЦИИ ПЕРСОНАЛЬНЫХ ДАННЫХ

А.С. Куракин

Рассматривается проблема обеспечения безопасности информационных систем персональных данных на стадиях разработки и оптимизации. Для решения данной проблемы предлагается использовать алгоритм деперсонализации для понижения класса информационных систем персональных данных, что, в свою очередь, снимает основную часть требований по применению организационных мер и технических средств защиты информации.

Ключевые слова: обезличивание, алгоритм деперсонализации, защита информации, персональные данные, информационные системы персональных данных

Введение

Требования Федерального закона «О персональных данных» [1] и постановления Правительства Российской Федерации «Об утверждении Положения об обеспечении безопасности персональных данных при их обработке в информационных системах персональных данных» [2] обязывает организации, обрабатывающие персональные данные (далее – операторы), обеспечивать соответствующую информационную безопасность персональных данных (ПДн). В свою очередь, выполнение данных требований влечет за собой значительные материальные затраты на внедрение дополнительных средств защиты информации, что зачастую не предусмотрено бюджетом операторов. Альтернативным законным способом решения данной проблемы является обезличивание персональных данных, так как оно позволяет снизить требования к уровню защищенности данных, что влечет за собой соответствующее сокращение расходов на обеспечение их информационной безопасности. Под обезличиванием персональных данных, как правило, понимают алгоритмы, в результате выполнения которых невозможно определить принадлежность персональных данных их владельцу.

Наиболее часто на практике применяются следующие способы обезличивания персональных данных [3]:

1. уменьшение перечня сведений, подлежащих автоматизированной обработке; в результате получают перечень данных, который соответствует оптимальному объему персональных данных о каждом субъекте, необходимому для хранения в информационных системах персональных данных (ИСПДн);
2. замена идентификаторами части сведений, что позволяет понизить класс ИСПДн, так как система оперирует не с персональными данными субъектов напрямую, а с идентификационной информацией, лишенной содержательного контекста;

3. уменьшение детализации некоторых сведений; способ направлен на то, чтобы сделать персональные данные субъектов менее точными. Это может достигаться, в том числе, путем группирования общих или непрерывных характеристик;
4. замена чисел минимальным, средним или максимальным значением; так как не всегда есть необходимость обрабатывать часть персональных данных каждого субъекта, то можно заменять данные минимальным, средним или максимальным значением по всей выборке или отдельным ее частям;
5. обработка групп сведений в разных информационных системах, для чего ИСПДн разделяются на взаимодействующие участки системы. Данный способ можно применять для оптимизации набора средств защиты информации (СЗИ), используемых в каждом сегменте. Это, с одной стороны, приводит к снижению стоимости защиты, а с другой – снижает избыточность СЗИ в тех случаях, когда защищаемые данные расположены неравномерно по системе.

Для всех перечисленных способов критерием качества деперсонализации является вероятность получения персональных данных на основании утечки обезличенных данных конкретного человека. При этом предполагается, что злоумышленнику известен контекст обработки, а также дополнительная информация из других источников.

Основным недостатком указанных способов является то, что они не гарантируют отсутствие возможности получения персональной информации посредством контекстного анализа открытой информации, в том числе получаемой из смежных систем.

Целью данной работы является разработка стойкого алгоритма обезличивания персональных данных (далее по тексту – алгоритма деперсонализации), основанного на применении перестановок и позволяющего проводить обезличивание больших массивов персональных данных при минимальных объемах служебной информации.

Алгоритм деперсонализации

Перспективным способом решения поставленной задачи является перестановка персональных данных, относящихся к различным субъектам. Данный способ обладает следующими преимуществами: персональные данные хранятся в одной ИСПДн, значительно снижается вероятность успеха контекстного анализа.

Предлагаемый алгоритм деперсонализации построен на следующих принципах:

1. разбиение исходного множества данных на подмножества [4], что позволяет сократить размерность и упростить его практическую реализацию;
2. использование циклических перестановок [4], что реализует собственно перемешивание данных.

В качестве исходных данных возьмем таблицу персональных данных $D(d_1, d_2, \dots, d_N)$, где N – число атрибутов, а M – число строк таблицы.

Далее рассмотрим множество данных, относящееся к одному атрибуту – $d_i (i=1, 2, \dots, N)$. Это множество атрибута $d_i - A_i$, содержит M элементов. Все элементы каждого множества A_i пронумерованы от 1 до M , и в таблице $D(d_1, d_2, \dots, d_N)$ совокупность элементов множеств разных атрибутов с одинаковыми номерами будем называть записью с соответствующим номером. При этом в исходной таблице каждая запись имеет определенный смысл, связанный с конкретным субъектом (физическим лицом), т.е. содержит персональные данные конкретного лица, определенного в этой же записи.

Алгоритм обеспечивает перестановку данных каждого множества атрибутов исходной таблицы пошагово. На каждом шаге используется принцип циклических перестановок.

Проведем разбиение множества A_i на $A_{ij} (M > A_i > 1)$ непересекающихся подмножеств A_{ij} , где число элементов подмножества A_{ij} равно $M_{ij} (M > M_{ij} > 1), j=1, 2, \dots, K_i$. Все элементы каждого подмножества A_{ij} считаем занумерованными от 1 до M_{ij} , эти номера будем называть внутренними номерами элементов подмножества. Внешний номер элемента в подмножестве A_{ij} , имеющего внутренний номер k , обозначим – $m_{ijk} (1 \leq m_{ijk} \leq M)$, где m_{ijk} – это порядковый номер элемента на множестве A_i , соответствующий элементу с внутренним номером k . Разбиение каждого множества должно обладать следующими свойствами:

1. $A_i = \bigcup_{j=1}^{K_i} A_{ij}$ – подмножества разбиения включают все элементы множества A_i ;
2. $A_{ij} \neq \emptyset, A_{ij} \cap A_{im} = \emptyset$ для всех $j, m=1, 2, \dots, K_i$ – каждое подмножество не пусто, а пересечение любых двух подмножеств пусто;
3. $m_{ij1} = m_{i(j-1)M_{i(j-1)}} + 1$ для всех $j=2, \dots, K_i$ – для любых двух подмножеств A_{ij} и $A_{i(j-1)}$ элемент с первым внутренним номером подмножества A_{ij} имеет на единицу больший внешний номер, чем внешний номер элемента с наибольшим внутренним номером подмножества $A_{i(j-1)}$;
4. если $k_1 > k_2$, то $m_{ijk_1} > m_{ijk_2}$ для всех $i=1, 2, \dots, N; j=1, 2, \dots, K_i$ – упорядоченность внешней и внутренней нумераций для всех множеств и подмножеств их разбиения совпадают;

5. $M = \sum_{j=1}^{K_i} M_{ij}$ – суммарное число элементов всех подмножеств A_{ij} равно общему числу элементов множества A_i .

Для каждого подмножества A_{ij} определим циклическую перестановку (подстановку) $p_{ij}(r_{ij})$, задаваемую следующим образом:

$$p_{ij}(r_{ij}) = \begin{pmatrix} 1 & 2 & 3 & \dots & (M_{ij}-1) & M_{ij} \\ (M_{ij}-r_{ij}+1) & (M_{ij}-r_{ij}+2) & (M_{ij}-r_{ij}+3) & \dots & (M_{ij}-r_{ij}-1) & (M_{ij}-r_{ij}) \end{pmatrix}.$$

Здесь элементы первой строки матрицы, стоящей в правой части равенства, соответствуют внутренним номерам элементов подмножества A_{ij} до перестановки (в исходной таблице), а элементы, стоящие во второй строке, соответствуют внутренними номерами элементов подмножества A_{ij} , стоящим на местах с номерами, определенными в верхней строке, после перестановки.

Таким образом, в перестановке (подстановке) $p_{ij}(r_{ij})$ производится циклический сдвиг всех элементов подмножества на число r_{ij} ($1 \leq r_{ij} \leq M_{ij}-1$). Будем называть величину r_{ij} параметром перестановки $p_{ij}(r_{ij})$. Данный параметр задается генератором случайных чисел (ГСЧ) в интервале $[1; M_{ij}-1]$. Теперь все перестановки для всех подмножеств множества A_i можно задать набором (вектором) параметров $r_i = (r_{i1}, r_{i2}, \dots, r_{iK_i})$. Вектор параметров перестановок r_i задает первый уровень алгоритма деперсонализации, т.е. перестановки первого уровня.

Рассмотрим теперь множество $a_i = (a_{i1}, a_{i2}, \dots, a_{iK_i})$, состоящее из K_i элементов. Здесь элемент a_{ij} соответствует подмножеству A_{ij} , $j=2, 3, \dots, K_i$. Для этого множества определим циклическую перестановку $p_{0j}(r_{0j})$:

$$p_{0j}(r_{0j}) = \begin{pmatrix} 1 & 2 & 3 & \dots & (K_i-1) & K_i \\ (K_i-r_{0i}+1) & (K_i-r_{0i}+2) & (K_i-r_{0i}+3) & \dots & (K_i-r_{0i}-1) & (K_i-r_{0i}) \end{pmatrix},$$

где элементы верхней строки матрицы перестановки соответствуют исходным номерам элементов множества a_i (подмножеств A_{ij}), а элементы нижней строки матрицы соответствуют номерам элементов множества a_i , стоящим на местах с номерами, определенными в верхней строке, после перестановки.

Таким образом, в перестановке $p_{0j}(r_{0j})$ производится циклический сдвиг элементов множества a_i (подмножеств множества A_i) на число r_{0i} ($1 \leq r_{0i} \leq K_i-1$) – параметр перестановки. Данный параметр r_{0i} задается ГСЧ в интервале $[1; K_i-1]$. Эту перестановку будем называть перестановкой второго уровня.

В результате последовательного проведения перестановок первого и второго уровней получается перемешивание элементов множества A_i так, что меняется нумерация этих элементов по отношению к исходной нумерации.

Определим теперь нумерацию элементов множества A_i после проведения всех перестановок. С учетом правил перемножения перестановок имеем следующую результирующую перестановку:

$$p_i(r_{0i}, r_i) = \begin{pmatrix} [1 \dots M_{i(K_i-r_{0i}+1)}] & [(M_{i(K_i-r_{0i}+1)}+1) \dots (M_{i(K_i-r_{0i}+1)}+M_{i(K_i-r_{0i}+2)})] \\ [m_{i(K_i-r_{0i}+1)} \dots m_{i(K_i-r_{0i}+1)}M_{i(K_i-r_{0i}+1)}] & [m_{i(K_i-r_{0i}+2)} \dots m_{i(K_i-r_{0i}+2)}M_{i(K_i-r_{0i}+2)}] \\ \dots & \dots \\ [M-M_{i(K_i-r_{0i}+1)} \dots M] & \dots \\ \dots & [m_{i(K_i-r_{0i})} \dots m_{i(K_i-r_{0i})}M_{i(K_i-r_{0i})}] \end{pmatrix}.$$

Здесь верхняя строка матрицы содержит порядковые номера элементов множества атрибута i в соответствии с их размещением в столбце после перестановок, а нижняя строка – внешние номера элементов множества этого атрибута, соответствующие их размещению в исходной таблице.

Примеры реализации алгоритма

Пример 1. Пусть $M=15$, $K_i=4$ и $M_{i1}=4$, $M_{i2}=4$, $M_{i3}=4$, $M_{i4}=3$, при этом $d_i=(b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}, b_{11}, b_{12}, b_{13}, b_{14}, b_{15})$, $A_{i1}=(b_1, b_2, b_3, b_4)$, $A_{i2}=(b_5, b_6, b_7, b_8)$, $A_{i3}=(b_9, b_{10}, b_{11}, b_{12})$, $A_{i4}=(b_{13}, b_{14}, b_{15})$, $r_i=(2, 1, 2, 1)$, $r_{0i}=2$.

После применения перестановок первого уровня имеем

$$p_{i1}(2) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ b_3 & b_4 & b_1 & b_2 \end{pmatrix}, \quad p_{i1}(1) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ b_6 & b_7 & b_8 & b_5 \end{pmatrix},$$

$$p_{i2}(2) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ b_{11} & b_{12} & b_9 & b_{10} \end{pmatrix}, \quad p_{i4}(1) = \begin{pmatrix} 1 & 2 & 3 \\ b_{14} & b_{15} & b_{13} \end{pmatrix}.$$

Результирующая перестановка имеет вид

$$p(2, (2, 1, 2, 1)) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ b_{11} & b_{12} & b_9 & b_{10} & b_{14} & b_{15} & b_{13} & b_8 & b_4 & b_1 & b_2 & b_6 & b_7 & b_9 & b_5 \end{pmatrix}.$$

Теперь представим, что алгоритм перестановки, определенный для множества, соответствующего одному атрибуту, применяется ко всем множествам атрибутов исходной таблицы. В этом случае полный алгоритм перестановки задается следующим набором параметров:

1. $(K_1, K_2, \dots, K_N)$ – множество, определяющее количество подмножеств для множества каждого атрибута, которое определяет подмножества элементов $(A_{11}, A_{12}, \dots, A_{1K_1}), (A_{21}, A_{22}, \dots, A_{2K_2}), \dots, (A_{N1}, A_{N2}, \dots, A_{NK_N})$.
2. $(M_{11}, M_{12}, \dots, M_{1K_1}), (M_{21}, M_{22}, \dots, M_{2K_2}), \dots, (M_{N1}, M_{N2}, \dots, M_{NK_N})$ – множество, определяющее число элементов в подмножествах для множества каждого атрибута;
3. $((r_{01}, r_1), (r_{02}, r_2), \dots, (r_{0N}, r_N))$ – множество параметров перестановок для множества каждого атрибута.

Этот набор задает параметры алгоритма деперсонализации для исходной таблицы $D(d_1, d_2, \dots, d_N)$.

В результате применения процедуры вместо исходной таблицы $D(d_1, d_2, \dots, d_N)$ получается таблица обезличенных данных $\tilde{D}(d_1, d_2, \dots, d_N)$. Набор параметров

$$C(D(d_1, d_2, \dots, d_N)) = \left\{ (K_1, K_2, \dots, K_N), ((M_{11}, M_{12}, \dots, M_{1K_1}), (M_{21}, M_{22}, \dots, M_{2K_2}), \dots, (M_{N1}, M_{N2}, \dots, M_{NK_N})), ((r_{01}, r_1), (r_{02}, r_2), \dots, (r_{0N}, r_N)) \right\}.$$

полностью и однозначно задает алгоритм деперсонализации для исходной таблицы $D(d_1, d_2, \dots, d_N)$.

Пример 2. Пусть исходная таблица $D(d_1, d_2, \dots, d_N)$ имеет вид, представленный в табл. 1.

Атрибут d_1	Атрибут d_2	Атрибут d_3	Атрибут d_4	Атрибут d_5	Атрибут d_6
q_1	r_1	s_1	t_1	u_1	v_1
q_2	r_2	s_2	t_2	u_2	v_2
q_3	r_3	s_3	t_3	u_3	v_3
q_4	r_4	s_4	t_4	u_4	v_4
q_5	r_5	s_5	t_5	u_5	v_5
q_6	r_6	s_6	t_6	u_6	v_6
q_7	r_7	s_7	t_7	u_7	v_7
q_8	r_8	s_8	t_8	u_8	v_8
q_9	r_9	s_9	t_9	u_9	v_9
q_{10}	r_{10}	s_{10}	t_{10}	u_{10}	v_{10}

Таблица 1. Исходная таблица данных

Для этой таблицы заданы следующие параметры алгоритма деперсонализации:

$$C(D(d_1, d_2, d_3, d_4, d_5, d_6)) = \left\{ (3, 2, 4, 3, 3, 2), ((3, 3, 4), (6, 4), (2, 3, 2, 3), (3, 4, 3), (5, 2, 3), (3, 7)), ((2, (1, 2, 3)), (1, (3, 1)), (3, (1, 2, 1, 1)), (2, (2, 1, 2)), (2, (4, 1, 1)), (1, (1, 4))) \right\}.$$

После выполнения алгоритма деперсонализации получаем таблицу $\tilde{D}(d_1, d_2, \dots, d_N)$ (табл. 2).

Как видно из примера, в результате применения алгоритма деперсонализации получена преобразованная таблица, в которой записи не соответствуют записям в исходной таблице, что обеспечивает достаточно высокую сложность восстановления исходной таблицы при отсутствии сведений о параметрах алгоритма деперсонализации. Доступность персональных данных (получение достоверных персональных сведений при легитимном обращении к ним) обеспечивается посредством решения обратного алгоритма деперсонализации, результатом чего является формирование исходной таблицы.

Атрибут d_1	Атрибут d_2	Атрибут d_3	Атрибут d_4	Атрибут d_5	Атрибут d_6
q_{10}	r_8	s_9	t_{10}	u_9	v_8
q_7	r_9	s_{10}	t_8	u_{10}	v_9
q_8	r_{10}	s_8	t_9	u_8	v_{10}
q_9	r_7	s_2	t_3	u_5	v_4

q_2	r_4	s_1	t_1	u_1	v_5
q_3	r_5	s_5	t_2	u_2	v_6
q_1	r_6	s_3	t_5	u_3	v_7
q_6	r_1	s_4	t_6	u_4	v_2
q_4	r_2	s_7	t_7	u_7	v_3
q_5	r_3	s_6	t_4	u_6	v_1

Таблица 2. Результат второго шага

Реализация обратного алгоритма деперсонализации

Пусть в столбце атрибута d_i таблицы $\tilde{D}(d_1, d_2, \dots, d_N)$ (табл. 2) выбран элемент номер n_i , тогда из матрицы результирующей перестановки $p_{0j}(r_{0j})$ можно получить номер этого элемента в исходной таблице – $d_i(n_i)$, который находится как элемент второй строки столбца номер n_i . Далее в каждом столбце атрибута d_j в соответствии с матрицей результирующей перестановки $p_{0j}(r_{0j})$ находится элемент, номер которого равен номеру столбца, во второй строке которого стоит число $m_i(n_i)$ (номер элемента в исходной таблице). Таким образом, после просмотра всех столбцов таблицы $\tilde{D}(d_1, d_2, \dots, d_N)$ будет построена запись, соответствующая элементу номер n_i из множества атрибута d_i и записи номер $m_i(n_i)$ в таблице $D(d_1, d_2, \dots, d_N)$. (табл. 1).

Практическая реализация алгоритма

При практической реализации алгоритма деперсонализации контроль целостности данных в файле обеспечивается путем проверки текущей контрольной суммы всего файла (сформированной при сохранении (модификации) файла) и контрольной суммы, рассчитываемой при последующем открытии файла. Как правило, данные алгоритмы контроля целостности хранимой в постоянном запоминающем устройстве (ПЗУ) информации (файлов) реализованы в механизмах защиты операционной системы, аппаратно-программном модуле доверенной загрузки (АПМДЗ) или специализированном программном обеспечении – СЗИ от несанкционированного доступа (НСД).

Перспективным развитием алгоритма деперсонализации в части обеспечения контроля целостности информации (персональных данных) является интеграция механизма формирования имитовставки, что обеспечит, помимо контроля целостности, более высокую степень защиты от НСД.

В качестве программной реализации алгоритма деперсонализации ПДн на языке C# предлагается экспериментальный образец программного обеспечения (ЭО ПО) «Depersonalization». Данное ПО работает в двух режимах:

- режим 1: программа производит деперсонализацию исходных данных;
- режим 2: программа осуществляет обратный алгоритм деперсонализации, приводя данные к исходному виду.

Оценка защищенности алгоритма деперсонализации

Для оценки защищенности предложенного алгоритма деперсонализации используем такую характеристику, как число вариантов деперсонализации, получаемых при применении данного алгоритма.

Число возможных различных вариантов разбиения множества из M элементов на K_i подмножеств, удовлетворяющих условиям разбиения, приведенным выше, при заданном наборе $(M_{i1}, M_{i2}, \dots, M_{iK_N})$ равно $(K_i)!$ (при условии, что все подмножества имеют различное число элементов). Максимальное число возможных вариантов для заданного набора разбиений N множеств атрибутов равно

$$V((K_1, K_2, \dots, K_N), (M_{11}, M_{12}, \dots, M_{1K_1}), \dots, (M_{i1}, M_{i2}, \dots, M_{iK_N})) = \prod_{i=1}^N (K_i)! (K_i - 1) (M_{i1} - 1) (M_{i2} - 1) \dots (M_{iK_i} - 1).$$

При большом количестве записей число вариантов получается очень большим, что обеспечивает очень малую вероятность подбора параметров и соответственно хорошую защиту обезличенных данных.

Для числовой оценки рассчитаем число возможных вариантов разбиения на примере ИСПДн, в которой одновременно обрабатываются паспортные данные 100 субъектов в пределах конкретной организации:

- фамилия;

- имя;
- отчество;
- серия и номер паспорта;
- дата рождения;
- пол;
- адрес.

В данном примере для простоты вычислений зададим одинаковые параметры разбиений для всех множеств атрибутов: $K_i = 10, i = \overline{1, N}, N = 7$. Мощности подмножеств распределим следующим образом: $M_1=5, M_2=6, M_3=7, M_4=8, M_5=9, M_6=11, M_7=12, M_8=13, M_9=14, M_{10}=15$.

В результате вычислений по вышеуказанной формуле получаем максимальное число возможных вариантов для заданного набора разбиений, равное $1,13 \times 10^{117}$. Таким образом, в случае, когда злоумышленник, имея деперсонализированные персональные сведения, применяет для получения достоверной информации метод «грубой силы» (полного перебора) [5], ему требуется более $1,19 \times 10^{108}$ лет при вычислительных ресурсах, обеспечивающих скорость перебора 3000000 вариантов в секунду.

Заключение

Предложенный алгоритм деперсонализации является перспективным и оптимальным решением задач по обеспечению информационной безопасности персональных данных, обрабатываемых в информационных системах персональных данных.

Данный алгоритм обладает следующими преимуществами:

- обеспечивает защиту персональных сведений от несанкционированного доступа, в том числе от компрометации информации при ее утечке по техническим каналам;
- обеспечивает гарантированный доступ к персональным данным при легитимном обращении;
- персональные сведения хранятся в одной таблице;
- получение персональных сведений посредством контекстного анализа или путем перебора весьма трудоемко, а зачастую практически невозможно;
- параметры перестановки задаются при помощи генератора случайных чисел, что увеличивает стойкость алгоритма к взлому.

Наибольшая эффективность при применении данного алгоритма проявляется в случае, когда в информационных системах персональных данных содержатся большое количество персональных данных субъектов, что обеспечивает наибольшую защиту информационной системе.

Литература

1. Федеральный закон Российской Федерации от 27 июля 2006 г. № 152-ФЗ. О персональных данных. Принят Государственной Думой Федерального Собрания Российской Федерации 8 июля 2006 г.; одобрен Советом Федерации Федерального Собрания Российской Федерации 14 июля 2006 г. // Российская газета. – 2006. – 29 июля.
2. Об утверждении Положения об обеспечении безопасности персональных данных при их обработке в информационных системах персональных данных: Постановление Правительства Российской Федерации от 17 ноября 2007 г. № 781 г. Москва // Российская газета. – 2007. – 21 ноября.
3. Царев Е. Информационная безопасность по-русски // День 4. Обезличивание персональных данных. 2009 – [Электронный ресурс]. – URL: <http://www.tsarev.biz/news/den-4-obezlichivanie-personalnykh-dannykh>, свободный. Яз. рус. (дата обращения: 29.03.12).
4. Стенли Р. Перечислительная комбинаторика. – М.: Мир, 1990. – 440 с.
5. Полный перебор // Википедия – свободная энциклопедия – [Электронный ресурс]. – URL: <http://ru.wikipedia.org/wiki>, свободный. Яз. рус. (дата обращения: 15.04.2012).

Куракин Александр Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, nirtit@gmail.com

УДК 621.3

**АЛГОРИТМ ИДЕНТИФИКАЦИИ РЕКВИЗИТОВ ФИЗИЧЕСКИХ ЛИЦ
В БАЗАХ ДАННЫХ НА ОСНОВЕ МЕТРИКИ ЛЕВЕНШТЕЙНА****Н.И. Лиманова, М.Н. Седов**

При передаче данных от одного учреждения к другому возникает проблема персональной идентификации физических лиц, у которых частично или полностью не совпадают реквизиты. В работе предложен оптимальный алгоритм идентификации на основе метрики Левенштейна, позволяющий выполнять поиск физических лиц в базе данных на основе нечеткого сравнения. Алгоритм реализован на языке PL-SQL в СУБД Oracle 11g.

Ключевые слова: межведомственный информационный обмен, идентификация, нечеткое сравнение, поиск реквизитов физических лиц, функция интеллектуального сравнения, персональный идентификационный номер.

Введение

В процессе межведомственного информационного обмена возникает проблема согласования основных реквизитов (ФИО, даты рождения, адреса, паспортных данных и т.п.) физических лиц в базах данных различных ведомств, обменивающихся информацией. Проблема персональной идентификации приобретает наибольшую актуальность для физических лиц, у которых частично или полностью не совпадают реквизиты.

Для удобства обработки данных каждому набору реквизитов в базах данных присваивается так называемый персональный идентификационный номер (ПИН). В случае обработки или передачи данных о физическом лице вся привязка осуществляется именно к этому ПИНу. В России, к сожалению, пока нет единой базы с реквизитами всех жителей, поэтому в разных ведомствах ведется свой отдельный реестр физических лиц и заводятся свои ПИНЫ. Проблема возникает при осуществлении обмена информацией о жителях между организациями, так как необходимо выполнить привязку входящих реквизитов к уже имеющимся. Для однозначной привязки необходимо выполнить интеллектуальный поиск физического лица в базе-приемнике, который должен учитывать множество факторов: и потенциальные ошибки при ручном вводе, и отсутствующие или устаревшие реквизиты, и т.п. Естественным предположить, что подобный поиск целесообразно реализовать в виде специализированного программного обеспечения [1].

Традиционно данная проблема решается путем анализа тождественности основных реквизитов физического лица. Таких реквизитов несколько: фамилия, имя, отчество, дата рождения, серия, номер паспорта и адрес. Однозначно определив совпадение существующих и новых реквизитов, можно выполнить идентификацию физического лица в базе данных. Данный метод поиска выполняется вручную только в том случае, когда объем передаваемой информации невелик (количество физических лиц не более 30). При больших объемах передаваемых данных используется автоматизированное сравнение тождественности реквизитов. Такой подход позволяет определить в среднем 50–60% от общего числа идентифицируемых физических лиц. Оставшиеся 40–50% представляют собой персональные данные, в которых частично или полностью не совпадают реквизиты. Такую информацию вручную обрабатывать еще сложнее.

Неправильная идентификация может привести также к большому количеству данных в отчете для ручной отработки, к присвоению ПИНа не тому человеку и к добавлению излишних данных. Последствия таких ошибок в худшем случае могут полностью парализовать работу учреждения на неопределенное время, в лучшем – отнять более 10% рабочего времени специалистов на исправление ошибок. Анализ существующего программного обеспечения показал, что единого идентификатора нет, универсальный алгоритм идентификации также отсутствует. Так как большинство реквизитов физических лиц имеет строковый тип, то естественно предположить, что необходимый алгоритм должен анализировать именно строковые значения.

Математическая модель

Известно несколько видов метрик, отражающих интуитивное понятие схожести строк. Наиболее распространены расстояния Хемминга, метрика Левенштейна и расстояние редактирования [2–4].

Расстояние Хемминга определяется для строк одинаковой длины и задается как число позиций, в которой символы не совпадают. Фактически расстояние Хемминга рассчитывается как минимальная цена преобразования одной строки в другую, когда возможна только одна операция редактирования строк – замена.

В случае, когда требуется произвести сравнение строк разной длины, используются метрика Левенштейна или расстояние редактирования. Эти две метрики очень похожи по построению и фактически являются одной и той же метрикой, несколько модифицированной для каждого случая. Так, например, метрика Левенштейна определяется как минимальная цена преобразования одной строки в другую с использованием трех операций – вставки, замены и удаления символа, причем все три операции имеют одинаковый вес.

Расстояние редактирования является модификацией метрики Левенштейна на случай, когда разрешены всего две операции – вставки и удаления.

В связи с изложенным была выбрана именно общая метрика Левенштейна, которая поддерживает все три операции со строкой. Для дальнейшей работы была построена лингвистическая переменная «схожесть строк». Решено выделить следующие термы: «строки совпадают», «строки почти совпадают», «строки похожи», «строки похожи и не похожи одновременно», «строки не похожи».

В результате анализа функций принадлежности лингвистических термов возникла необходимость модификации метода вычисления метрики Левенштейна. Потребовалось модифицировать метрику таким образом, чтобы расстояние между строками зависело, в том числе, и от длины сравниваемых строк.

Теорема. Обозначим величиной $p(s_1, s_2)$ метрику Левенштейна, а величиной $\|s_i\|$ – длину строки s_i . Тогда функция

$$r(s_1, s_2) = \frac{p(s_1, s_2)}{\max \{\|s_1\|, \|s_2\|\}}, \quad (1)$$

является метрикой.

Доказательство. Поскольку $p(s_1, s_2)$ – метрика, то имеем

$$p(s_1, s_2) \geq 0,$$

$$p(s_1, s_2) = p(s_2, s_1),$$

$$p(s_1, s_2) + p(s_2, s_3) \geq p(s_1, s_3)$$

для любых строк s_1, s_2 и s_3 .

Учитывая эти соотношения и равенство (1), приходим к выводу, что $r(s_1, s_2)$ удовлетворяет первым двум аксиомам, определяющим метрику. Остается доказать, что для любых строк s_1, s_2 и s_3 функция $r(s_1, s_2)$ удовлетворяет неравенству треугольника, т.е.

$$r(s_1, s_2) + r(s_2, s_3) \geq r(s_1, s_3).$$

Запишем это неравенство в виде:

$$\frac{p(s_1, s_2)}{\max \{\|s_1\|, \|s_2\|\}} + \frac{p(s_2, s_3)}{\max \{\|s_2\|, \|s_3\|\}} - \frac{p(s_1, s_3)}{\max \{\|s_1\|, \|s_3\|\}} \geq 0.$$

Возможны следующие случаи:

1. $\|s_1\| \leq \|s_2\| \leq \|s_3\|$;
2. $\|s_2\| \leq \|s_3\| \leq \|s_1\|$;
3. $\|s_3\| \leq \|s_1\| \leq \|s_2\|$;
4. $\|s_2\| \leq \|s_1\| \leq \|s_3\|$;
5. $\|s_1\| \leq \|s_3\| \leq \|s_2\|$;
6. $\|s_3\| \leq \|s_2\| \leq \|s_1\|$.

Рассмотрим первый случай. Имеем:

$$\begin{aligned} & \frac{p(s_1, s_2)}{\max \{\|s_1\|, \|s_2\|\}} + \frac{p(s_2, s_3)}{\max \{\|s_2\|, \|s_3\|\}} - \frac{p(s_1, s_3)}{\max \{\|s_1\|, \|s_3\|\}} = \frac{p(s_1, s_2)}{\|s_2\|} + \frac{p(s_2, s_3)}{\|s_3\|} - \frac{p(s_1, s_3)}{\|s_3\|} \geq \\ & \geq \frac{1}{\|s_3\|} (p(s_1, s_2) + p(s_2, s_3) - p(s_1, s_3)) \geq 0. \end{aligned}$$

Таким образом, для первого случая неравенство треугольника выполняется. Поскольку второй случай аналогичен первому, на основании подобных выкладок делаем вывод, что для второго случая неравенство треугольника также выполняется.

Перейдем к рассмотрению третьего случая, где

$$r(s_1, s_2) + r(s_2, s_3) - r(s_1, s_3) = \frac{1}{\|s_2\|} (r(s_1, s_2) + r(s_2, s_3)) - \frac{1}{\|s_1\|} r(s_1, s_3). \quad (2)$$

Рассмотрим вопрос о том, когда достигается минимум функции, находящейся в правой части этого равенства. Понятно, что если выражение $r(s_1, s_2) + r(s_2, s_3)$ достигает минимума, а $r(s_1, s_3)$ максимума, то значение всего выражения будет минимальным. Указанные два условия могут выполняться одновременно, если одновременно выполняются два следующих утверждения:

1. строки s_1 и s_3 не имеют общих символов;
2. строки s_1 и s_3 входят в качестве подстрок в s_2 .

Тогда

$$r(s_1, s_3) = \max \{\|s_1\|, \|s_3\|\} = \|s_1\|,$$

$$r(s_1, s_2) = \|s_3\| + \|C\|, \quad r(s_2, s_3) = \|s_1\| + \|C\|,$$

и, таким образом, минимальное значение выражения (2) запишется в виде

$$\frac{\|s_3\| + \|C\| + \|s_1\| + \|C\|}{\|s_3\| + \|s_1\| + \|C\|} - \frac{\|s_1\|}{\|s_1\|} = \frac{\|C\|}{\|s_3\| + \|s_1\| + \|C\|} \geq 0.$$

Следовательно, в третьем случае для функции $r(s_1, s_3)$ также выполняется неравенство треугольника. Остальные случаи аналогичны уже рассмотренным. Таким образом, функция $r(s_1, s_2)$ является метрикой, заданной на множестве строк. Теорема доказана.

Замечание. Функция $r(s_1, s_2)$ принадлежит отрезку $[0, 1]$ для любых строк s_1 и s_2 .

В предложенном алгоритме данная метрика применяется для работы со строковыми реквизитами физических лиц, к которым относятся ФИО, адрес, документ и т.д. В связи с этим построенная с использованием данной метрики лингвистическая переменная позволяет обрабатывать запросы поиска для человека, похожего на другого человека по реквизитам. Приняв от пользователя такой запрос, мы фактически получаем два значения – значение искомого реквизита и радиус поиска.

Алгоритм идентификации реквизитов физических лиц

Укрупненная блок-схема разработанного алгоритма идентификации реквизитов физических лиц в базах данных представлена на рисунке.

В реализации алгоритма на языке PL-SQL СУБД Oracle 11g за предварительную выборку всех записей, отдаленно похожих на искомую, отвечает блок «Запрос количества идентичных людей в базе данных». Этот блок работает по алгоритму прямого частичного сравнения разных наборов реквизитов, например, имени, отчества и даты рождения, формируя тем самым рабочий набор данных для рассматриваемого алгоритма идентификации. Затем в работу вступает «Блок сравнения реквизитов», ключевые функции которого отводятся логически выделенным процедурам COMPARISON_STRING и COMPARISON_NUMBER, созданным на основе модифицированного метода вычисления метрики Левенштейна, которые позволяют проводить интеллектуальное сравнение двух похожих строк или чисел с учетом возможных неточностей или ошибок ввода. С помощью указанных процедур программа формирует набор совпадений и по результатам обработки предлагаемой и искомой записи выносит решение об идентичности. Например, у человека совпадает имя, отчество, дата рождения и номер паспорта, а в фамилии допущена ошибка в одну букву. В данном случае программа однозначно идентифицирует реквизиты. Данные процедуры могут применяться не только для идентификации реквизитов, но также везде, где требуется полнотекстовый поиск с нечетко заданными входными данными.

Алгоритм идентификации аккумулирует так называемый «опыт прошлых идентификаций» и записывает его в специально отведенное место в базе данных для использования в последующих идентификациях. Это позволяет сохранить не только результаты автоматической работы программы, но и решения операторов после отработки ими оставшихся не найденных реквизитов.

Для сравнительного анализа разработанного алгоритма рассмотрим технологию идентификации на основе прямого сравнения. При использовании данной технологии предпочтение отдается скорости обработки записей, а не качеству принятия решения системой. В итоге, после окончания работы процедуры на основе прямого сравнения, остается много данных (около 20–30% от общего количества строк), не связанных с исходными, которые необходимо обрабатывать вручную, что крайне затруднительно при больших объемах обрабатываемых данных.

В качестве тестовой среды были выбраны сводные базы данных населения города с количеством записей примерно 800 000, СУБД Oracle 11g, сервер HP ProLiant DL160 G6.

При сравнении рабочих показателей двух алгоритмов выявлены особенности алгоритмов:

- алгоритм прямого сравнения
 - скорость обработки данных – примерно 100 000 строк в час;
 - точность идентификации (вероятность точного поиска реквизитов) – примерно 80%;
- алгоритм идентификации на основе нечеткого сравнения
 - скорость обработки данных – примерно 80 000 строк в час;
 - точность идентификации (вероятность точного поиска реквизитов) – примерно 99%.

Таким образом, разработанный алгоритм, хотя и демонстрирует несколько меньшую скорость обработки, но за счет интеллектуальной системы принятия решений минимизирует ручную работу оператора по отработке результатов, чего не может предложить алгоритм прямого сравнения.

Заключение

Рассмотренный метод идентификации персональных данных на основе нечеткого сравнения позволяет быстро определять людей, используя данные ранее проведенного поиска. Встроенная система приоритета реквизитов позволяет идентифицировать человека в таких случаях, как смена фамилии, имени, переезд, ошибки при ручном вводе данных, а также при частично отсутствующих реквизитах.

Рассмотренную в работе процедуру идентификации можно рассматривать как часть системы поддержки принятия решений. Процедура не требует вмешательства оператора, накапливает опыт в процессе работы, позволяя тем самым полностью освободить специалистов от низкопрофильной, неэффективной, ручной работы напрямую с наборами реквизитов физических лиц, хранящимися в базах данных.

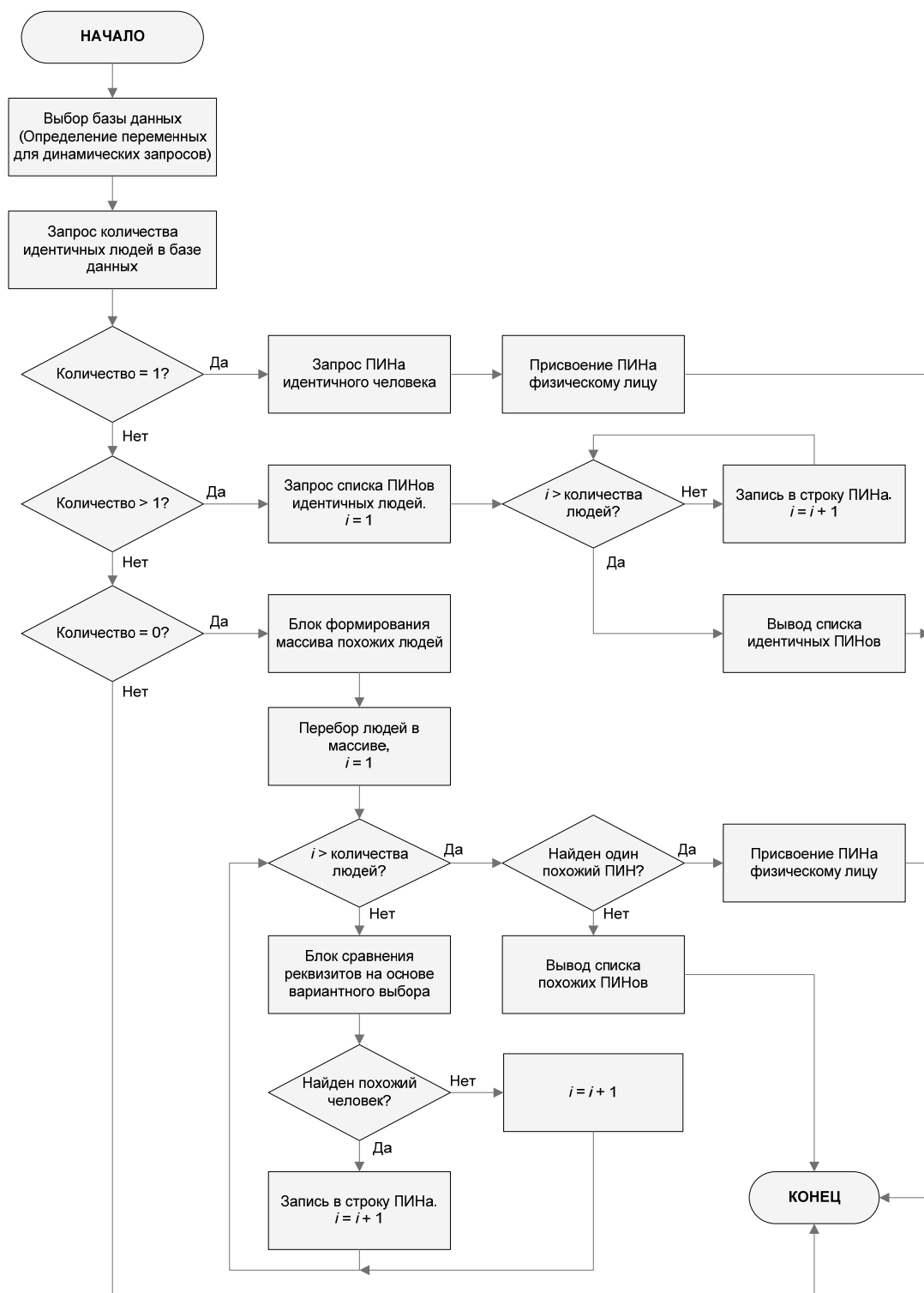


Рисунок. Укрупненная блок-схема алгоритма идентификации реквизитов физических лиц в базах данных

В перспективе данный алгоритм обладает возможностью успешного внедрения в системы глобального объединения хранилищ государственных или коммерческих организаций для ведения единой базы данных населения любой страны мира. Логическая структура разработанного алгоритма позволяет реализовать его на любом популярном языке программирования. Масштабируемость алгоритма позволяет применять программные процедуры на его основе как в малых организациях, так и в крупных корпорациях – везде, где ведется и актуализируется реестр данных физических лиц. Возможные примеры использования – портал госуслуг, медицинские электронные системы, кадровые и бухгалтерские системы учета служащих, банковские системы хранения данных о клиентах и т.п.

Алгоритм реализован на языке PL-SQL системы управления базами данных Oracle 11g. Разработанное программное обеспечение, реализующее метод автоматизированного поиска персональных дан-

ных на основе нечеткого сравнения, внедрено и успешно функционирует с 2007 г. в муниципальном учреждении «Городской информационный центр» г. Тольятти Самарской области.

Литература

1. Международный фонд автоматической идентификации. Технологии автоматической идентификации [Электронный ресурс]. – Режим доступа: <http://www.fond-ai.ru/art1/art223.html>, свободный. Яз. рус. (дата обращения 16.06.2012).
2. Хемминг Р.В. Теория кодирования и теория информации: Пер. с англ. / Под ред. Б.С. Цыбакова. – М.: Радио и связь, 1985. – 176 с.
3. Левенштейн В.И. Двоичные коды с исправлением выпадений, вставок и замещений символов // ДАН СССР. – 1965. – Т. 163. – № 4. – С. 845–848.
4. Бойцов Л.М. Анализ строк [Электронный ресурс]. – Режим доступа: http://itman.narod.ru/articles/infoscope/string_search.1-3.html, свободный. Яз. рус. (дата обращения 16.06.2012).

Лиманова Наталья Игоревна – Тольяттинский государственный университет, доктор технических наук, профессор. Natalya-Igorevna@yandex.ru
Седов Максим Николаевич – Мэрия городского округа Тольятти, инженер-программист, SedovMN@inbox.ru

УДК 004.822

**РАЗРАБОТКА ОНТОЛОГИИ ОТКРЫТЫХ ГОСУДАРСТВЕННЫХ ДАННЫХ
НА ПРИМЕРЕ РАСХОДОВ БЮДЖЕТА САНКТ-ПЕТЕРБУРГА****О.В. Пархимович, В.В. Власов, Д.И. Муромцев**

Представлено описание методологии и процесса разработки онтологии открытых государственных данных на примере ведомственной структуры расходов бюджета Санкт-Петербурга. Выполнено обоснование необходимости в создании онтологии как с точки зрения формирования единого терминологического словаря для пользователей и программных агентов, так и для интеграции информационных ресурсов и систем. Приведена общая структура онтологии, описан процесс ее реализации. Кроме того, предложенная схема разработки онтологии позволяет осуществлять ее дальнейшую модификацию и дополнение при изменении исходных данных или нормативных документов.

Ключевые слова: открытые государственные данные, онтологии, семантические сети.

Введение

В последнее время проблематика создания и публикации открытых государственных данных становится более актуальной: например, со слов экс-министра экономического развития Российской Федерации (РФ) Эльвиры Набиуллиной, «Минэкономразвития России считает необходимым развивать тематику открытых государственных данных и планирует проводить с 2012 года ряд работ в рамках государственной программы «Информационное общество (2011–2020 годы)» [1].

Под открытыми государственными данными в мировой практике понимают публичную государственную информацию, предоставляемую в цифровом виде посредством сети Интернет в форме, допускающей последующий анализ и ее повторное использование. В результате становится возможным создание сервисов и проектов, решающих отдельные проблемы граждан, появление новых возможностей для коммерческой деятельности, подкрепляющей экономику государства, повышение эффективности работы государственных органов и увеличение прозрачности управления государством.

В работе отражены результаты разработки онтологии открытых государственных данных ведомственной структуры расходов бюджета (ВСРБ) Санкт-Петербурга, которая является одним из Приложений Закона Санкт-Петербурга о бюджете [2], содержащим распределение бюджетных ассигнований, предусмотренных Законом о бюджете на соответствующий финансовый год главным распорядителем бюджетных средств по разделам, подразделам, целевым статьям и видам расходов бюджетной классификации РФ. Выбор именно этих массивов данных обусловлен их доступностью (данные публикуются на сайте администрации Санкт-Петербурга), достоверностью, актуальностью и наличием четкой структуры данных, что значительно упрощает проведение экспериментальной и аналитической части работы.

В настоящее время Приложения к Закону о бюджете Санкт-Петербурга, содержащие данные о бюджете, публикуются в формате PDF [3], который не содержит структуры информации и не может быть использован для анализа и машинной обработки (например, для автоматизированной визуализации данных). Важность публикации в структурированном формате информации о бюджете городов подтверждается открытием в октябре 2011 г. портала «Открытый бюджет» города Москвы [4], на котором можно не только получить всю интересующую информацию в простой и доступной форме (как текстовой, так и визуальной), но и скачать массивы данных в структурированном формате, а также предложить свои идеи по изменению структуры бюджета.

В ВСРБ Санкт-Петербурга содержится важная информация о распределении доходов города. Значительная часть доходов формируется из налоговых отчислений граждан, поэтому они должны иметь возможность изучения и анализа рассматриваемого документа в удобных для них форматах. Данная работа позволяет решить указанную проблему, предоставляя указанные данные в формате онтологии, пригодном для машинной обработки, что делает анализ данных менее трудоемким.

Постановка задачи

Онтология расходов бюджета Санкт-Петербурга предназначена для предоставления информации в структурированных форматах, которые могут быть проанализированы пользователями или использованы разработчиками при создании приложений.

Структура разработки онтологии представлена на рис. 1 и заключается в следующем. Массив данных в формате PDF скачивается с официального источника. После этого исследуется его структура, на основании которой проектируется иерархия классов онтологии. Полученная иерархия классов создается в системе Protégé и заполняется экземплярами классов. Полученный проект системы Protégé экспортируется в формат RDFs в виде двух RDFs-файлов: файла иерархии классов и файла экземпляров.

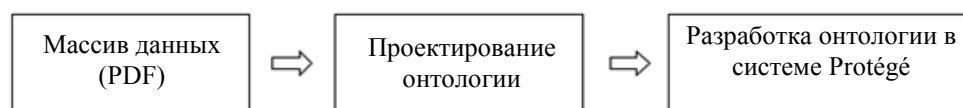


Рис. 1. Схема разработки онтологии

Исследование структуры данных ВСРБ Санкт-Петербурга

Термин «ведомственная структура расходов бюджета» в Бюджетном кодексе РФ определен как распределение бюджетных ассигнований, предусмотренных законом (решением) о бюджете на соответствующий финансовый год главным распорядителем бюджетных средств, по разделам, подразделам, целевым статьям и видам расходов бюджетной классификации РФ [5].

ВСРБ Санкт-Петербурга представляет собой таблицу, состоящую из сумм и статей расходов и классификаторов, обеспечивающих международную сопоставимость данных. Коды содержат основную информацию, необходимую для анализа бюджета. Чтобы перевести данные в структурированный формат, необходимо сгенерировать справочники классификаторов, благодаря которым станет возможным автоматизированный анализ информации. Согласно Указанию «О порядке применения бюджетной классификации Российской Федерации», утвержденным Приказом Министерства финансов РФ от 28.12.2010 № 190н, «классификация расходов бюджетов представляет собой группировку расходов бюджетов всех уровней и отражает направление бюджетных средств на выполнение единицами сектора государственного управления и местного самоуправления основных функций, решение социально-экономических задач» [5]. В структуре расходов используются следующие классификаторы.

- Коды главных администраторов доходов бюджета. Главными администраторами доходов являются органы государственной власти Санкт-Петербурга, которые имеют право распределять бюджетные средства между подведомственными распорядителями и получателями бюджетных средств.
- Код раздела и подраздела. В классификации расходов бюджета содержатся 14 разделов, отражающих направление финансовых ресурсов на выполнение основных функций государства. Каждый раздел детализирован подразделами, конкретизирующими направление бюджетных средств.
- Код целевой статьи. Целевые статьи обеспечивают привязку бюджетных ассигнований к конкретным направлениям деятельности субъектов бюджетного планирования и участников бюджетного процесса в пределах подразделов классификации расходов бюджетов. Код целевой статьи состоит из 7 знаков: первые три знака отражают статью расходов, следующие два – код программы (принадлежность расходов к соответствующему закону, иному нормативно-правовому акту, устанавливающему выплату), и последние – код подпрограммы (виды выплат в рамках закодированного на уровне программы закона, иного нормативно-правового акта). Перечни целевых статей, используемых в бюджетах субъектов РФ, формируются соответствующим финансовым органом субъекта РФ.
- Код вида расходов. Виды расходов детализируют направление финансирования расходов бюджетов как по целевым статьям, так и по целевым программам расходов. Они утверждаются федеральным законом на соответствующий год.
- Код операций сектора государственного управления (ОСГУ). Классификация ОСГУ является группировкой операций в зависимости от их экономического содержания. В рамках данной классификации ОСГУ разделены на текущие (доходы и расходы), инвестиционные (операции с нефинансовыми активами) и финансовые (операции с финансовыми активами и обязательствами). Группы детализируются статьями и подстатьями.

Проектирование онтологии

Онтологии необходимы для определения словаря терминов, совместно используемых специалистами. Они состоят из машинно-интерпретируемых формулировок основных понятий и отношений между ними, определяя общий словарь для ученых, которым необходимо использовать информацию о предметной области. Онтологии обычно используются другими приложениями, например, онтология ВСРБ Санкт-Петербурга может применяться для визуализации содержащихся в ней данных по выбранным пользователем параметрам. Одной из основных задач онтологий является совместное использование людьми или программными агентами общего понимания структуры информации.

В рассматриваемом случае онтология – формальное явное описание понятий в рассматриваемой предметной области (классов), слотов (свойств каждого понятия) и наложенных на них ограничений [6]. Класс является центральным элементом онтологии и описывает понятия предметной области. Например, класс главных администраторов доходов представляет всех администраторов, а конкретные администраторы, например, «Администрация Фрунзенского района» являются экземплярами этого класса. У класса могут быть подклассы, представляющие более конкретные понятия. Например, класс администраторов доходов можно разделить на подкласс администраций районов Санкт-Петербурга и подкласс комитетов

Санкт-Петербурга. Слоты описывают свойства классов и экземпляров: у класса администраторов доходов могут быть слоты «код» и «руководитель».

Перед проектированием онтологии необходимо определить, какую область она будет охватывать и ответы на какие типы вопросов в ней должны содержаться. В данном случае областью онтологии является представление структуры ВСПБ Санкт-Петербурга и сумм расходов. Используя приложения, основанные на данной онтологии, пользователь сможет получать информацию по интересующим его статьям расходов, сравнивать расходы различных распорядителей бюджетных средств или суммы по операциям сектора государственного управления. Разработку онтологии можно разбить на три этапа: определение классов и их иерархии, определение слотов и описание допустимых значений, создание экземпляров классов и заполнение значений слотов экземпляров.

В иерархии классов ВСПБ Санкт-Петербурга необходимо выделить классы используемых в бюджете классификаторов (распорядители бюджетных средств, операции сектора государственного управления, разделы и др.) и класс расходов, содержащий значения сумм расходов с их классификаторами.

После определения классов создается внутренняя структура понятий, которая описывается с помощью свойств классов. Свойства могут быть внутренними (например, количество сотрудников распорядителей бюджетных средств), внешними (например, государственный орган, контролирующий расходы распорядителей бюджетных средств), частью целого (например, функциональные подразделения распорядителей бюджетных средств), отношениями с другими индивидуальными концептами. Слоты могут иметь различные ограничения, описывающие тип значения, разрешенные значения, число значений (мощность) и другие свойства значений. Например, значение слота «название» в классе «распорядители бюджетных средств» – слот с типом значения «строка». Мощность слота определяет количество значений, которые может иметь слот. Например, слот «сумма» класса «Расходы» может иметь только одно значение, т.е. единичную мощность.

Последним шагом в разработке онтологии является создание экземпляров классов. Для определения отдельного экземпляра необходимо выбрать класс, создать отдельный экземпляр этого класса и ввести значения слотов. Например, можно создать отдельный экземпляр класса «Операции сектора государственного управления», значением слота «Наименование» которого будет строка «Оплата труда и начисления на выплаты по оплате труда», а значением слота «Код ОСГУ» будет число «210». Благодаря созданию классов онтологии создается структура расходов бюджета Санкт-Петербурга, а в результате создания отдельных экземпляров эта структура наполняется конкретными значениями.

На основании результатов исследования структуры данных ВСПБ Санкт-Петербурга была разработана онтология, структура которой представлена на рис. 2. В данной онтологии содержатся 6 классов с классификаторами, применяемыми в бюджете Санкт-Петербурга (классы «Главный распорядитель бюджетных средств», «Вид расходов», «Операции сектора государственного управления», «Подраздел», «Раздел», «Целевая статья») и класс «Расходы», содержащий все строки с суммами расходов ВСПБ Санкт-Петербурга.

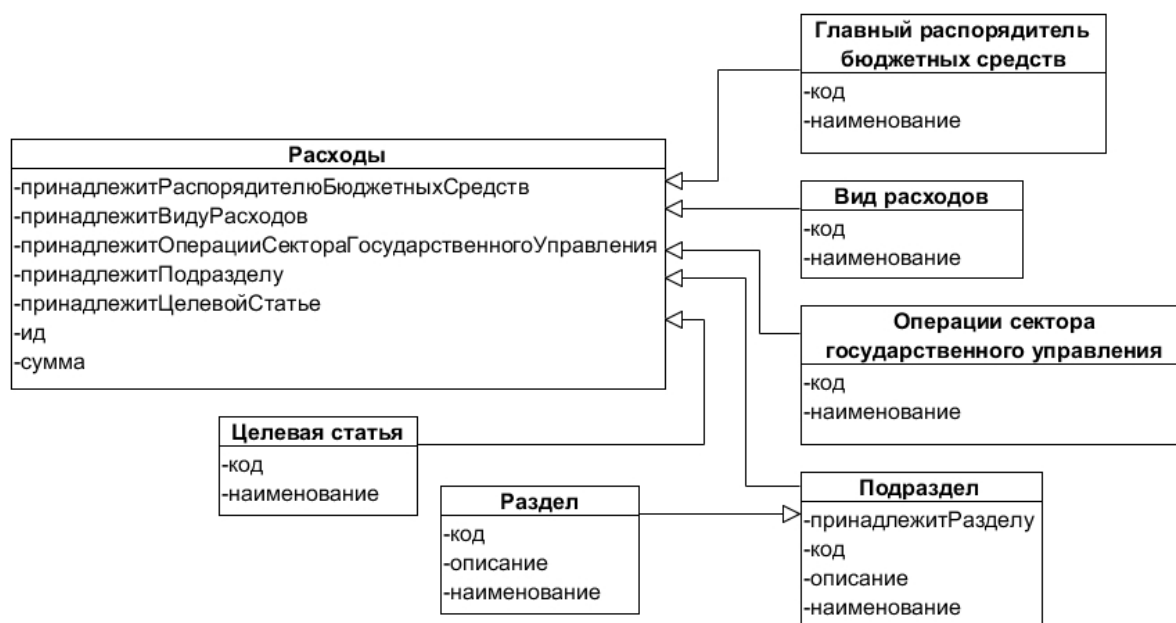


Рис. 2. Структура онтологии ВСПБ Санкт-Петербурга

Реализация онтологии в системе Protégé

На основе структуры классов для ВСРБ Санкт-Петербурга разработана онтология с помощью программного обеспечения Protégé [7]. Выбор системы Protégé для реализации онтологии объясняется тем, что данная программа является свободным, открытым редактором онтологий и фреймворком для построения баз знаний. Онтологии, разработанные с помощью данной системы, могут быть экспортированы в форматы RDF (RDF Schema), OWL и XML-схема. В работе использован редактор Protégé-OWL, позволяющий пользователям строить онтологии для семантической паутины, в частности на OWL [8]. С помощью такой онтологии формальная семантика OWL определяет, как получать логические следствия (факторы, которые не присутствуют непосредственно в онтологии, но могут быть выведены из существующих посредством семантики). Данные выводы могут быть основаны на одном документе или на множестве распределенных документов, которые объединяются с использованием механизмов OWL.

При создании нового класса (например, класса «Вид расходов») ему по умолчанию присваивается стандартное имя (основанное на названии проекта), для изменения которого необходимо изменить значение поля «Name». В системе Protégé приняты правила наименования, согласно которым первая буква в каждом слове в имени класса пишется в верхнем регистре, а остальные буквы – в нижнем. Согласно данным правилам, название для класса «Вид расходов» будет записано как ClassOfExpenditures. В поле Description (описание) сохраняется русское название класса. В системе Protégé классы могут быть как конкретными (Concrete), на основе которых система может создавать экземпляры, так и абстрактными, т.е. классами, у которых экземпляров нет. По умолчанию создается класс «конкретного» типа, изменить который можно в поле Role. Рассматриваемый класс может иметь экземпляры, поэтому тип класса изменять не нужно. Для добавления нового слота необходимо открыть окно создания слота и заполнить все необходимые поля – наименование слота, тип значения, мощность слота и др. У рассматриваемого класса есть два слота – слот codeClassOfExpenditures (код «Вид расходов») и слот nameClassOfExpenditures (наименование «Вид расходов»); в данном случае оба слота имеют единичную мощность и тип данных «строка». Аналогично описанному выше способу создаются все классы из структуры онтологии, представленной на рис. 2.

Для дальнейшего использования полученной онтологии без системы Protégé возможно ее экспортирование в формат RDFs. Фрагмент кода полученного файла представлен в Листинге 1. Тег `<rdfs:Class rdf:about="&rdf_;ClassOfExpenditures">` декларирует некий класс, который описывается в некотором пространстве имен `<rdf>` как ClassOfExpenditures. Указание различных пространств имен дает возможность повторно использовать другие онтологии, уточнять и расширять их, объявляя новые подклассы. Данная возможность позволяет осуществлять интеграцию узконаправленных информационных ресурсов, объединенных широкой предметной областью.

```
<rdfs:Class rdf:about="&rdf_;ClassOfExpenditures"
  rdfs:comment="Вид расходов"
  rdfs:label="ClassOfExpenditures">
  <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="&rdf_;codeClassOfExpenditures"
  rdfs:label="codeClassOfExpenditures">
  <rdfs:domain rdf:resource="&rdf_;ClassOfExpenditures"/>
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

Листинг 1. Структура классов в RDFs-файле

Структура кода класса «ClassOfExpenditures» («Вид расходов») заключается в следующем: тег `<rdf:about>` обеспечивает название или ссылку на онтологию, которой в данном случае является ссылка `<http://protege.stanford.edu/rdf/ClassOfExpenditures>`; тег `<rdfs:comment>` обеспечивает возможность аннотировать онтологию, т.е. добавить комментарий о том, что данный класс обозначает «Вид расходов»; тег `<rdfs:label>` обеспечивает удобное для чтения пользователем название класса, тег `<rdfs:subClassOf>` связывает рассматриваемый класс с его надклассом, которым в данном случае является класс Resource. У рассматриваемого класса есть два свойства: codeClassOfExpenditures и nameClassOfExpenditures, которые в синтаксисе RDFs-файлов прописываются с тегом `<rdf:property>`. Тег `<rdfs:domain>` показывает, что данное свойство относится к свойствам класса ClassOfExpenditures, а тег `<rdfs:range>` обозначает диапазон свойств, ограничивающих экземпляры класса (в данном случае они должны быть типа Literal).

После создания иерархии классов необходимо создать экземпляры, т.е. непосредственно данные базы знаний. Перед этим необходимо еще раз проверить структуру классов, так как при необходимости изменения ее в дальнейшем, возможна потеря уже введенной информации.

Источником данных для класса CEOIncome («Главный распорядитель бюджетных средств») является документ «Перечень и коды главных администраторов доходов бюджета Санкт-Петербурга, которые являются органами государственной власти Санкт-Петербурга, и закрепляемые за ними виды доходов бюджета Санкт-Петербурга». Значениями поля codeCEOIncome являются значения столбца «Код главного администратора бюджетной классификации РФ», значениями поля nameCEOIncome – значения столбца «Наименование».

Источником данных для классов Section («Раздел») и Subsection («Подраздел») являются «Правила отнесения расходов всех бюджетов бюджетной системы РФ на соответствующие классификации расходов»: значением свойства codeSection является номер раздела, nameSection – название раздела, descriptionSection – описание раздела. Аналогично – для класса codeSubsection: свойство codeSubsection – номер подраздела, nameSubsection – название подраздела, descriptionSubsection – описание подраздела, значением свойства belongToSection является экземпляр класса Section, подразделом которого является рассматриваемый экземпляр класса.

Источником значений для классов ClassOfExpenditures («Вид расходов») и TargetArticle («Целевая статья») является Ведомственная структура расходов бюджета Санкт-Петербурга на 2012 год. Для класса ClassOfExpenditures значениями свойства codeClassOfExpenditures являются значения столбца «Код целевой статьи», значения свойства nameClassOfExpenditures – содержание столбца «Наименование» в строках с нумерацией третьего уровня. Аналогично для класса TargetArticle значениями свойства codeTargetArticle являются значения столбца «Код целевой статьи», а значениями свойства nameTargetArticle – содержание столбца «Наименование» в строках с нумерацией второго уровня.

Источником значений для классов OperationOfGovernment («ОСГУ») и Costs («Расходы») является ведомственная структура сводной бюджетной росписи бюджета Санкт-Петербурга на 2012 год. Для класса OperationOfGovernment значениями свойства codeOperationOfGovernment являются значения столбца «Код ОСГУ», для свойства nameOperationOfGovernment – значения столбца «Наименования» для строк с нумерацией третьего уровня. Для класса Costs значениями свойств belongToCEOIncome, belongToClassOfExpenditures, belongToOperationOfGovernment, belongToSubsection, belongToTargetArticle являются соответственно экземпляры классов CEOIncome, ClassOfExpenditures, OperationOfGovernment, Subsection, TargetArticle, у которых значения кодов соответствуют значениям соответствующих столбцов ведомственной структуры сводной бюджетной росписи; значением свойства Sum является значение столбца «Сумма»; значением свойства id является 16-значная строка, состоящая из кодов подраздела, целевой статьи, вида расходов и операции сектора государственного управления.

Создание экземпляров всех классов, кроме экземпляров класса Costs, происходит с помощью интерфейса программы Protégé. После этого полученный проект экспортируется в формат RDFs, получая на выходе два RDFs-файла: файл с иерархией классов (фрагменты файла приводились в Листинге 1) и файл с экземплярами классов (фрагмент файла представлен в Листинге 2). Его структура аналогична структуре файла с иерархией классов, за исключением того, что вместо тегов <rdfs:Class> или <rdf:Property> указывается тег <rdf:ClassName>, где ClassName – имя класса, экземпляр которого создается в данном теге.

```
<rdf_:ClassOfExpenditures rdf:about="&rdf_;BudgetSPb2_Class115"
  rdf_:codeClassOfExpenditures="012"
  rdf_:nameClassOfExpenditures="Выполнение функций государственными органами"
  rdfs:label="012"/>
```

Листинг 2. Структура экземпляров в RDFs-файле

Схема автоматизированного создания экземпляров класса Costs представлена на рис. 3. Из первоначального файла ведомственной структуры сводной бюджетной росписи бюджета Санкт-Петербурга на 2012 год (формат PDF) данные копируются в файл формата TXT. После этого с помощью регулярных выражений удаляются все текстовые данные и нумерация строк из столбца «Номер» ведомственной структуры сводной бюджетной росписи. В результате данных действий получен файл, в котором содержатся данные для свойств экземпляра класса Costs: belongToCEOIncome, belongToSubsection, belongToTargetArticle, belongToClassOfExpenditures, belongToOperationOfGovernment и sum, разделенные символом пробела. Данный файл является структурированным, поэтому для его преобразования можно написан парсер на языке программирования Java.

Принцип работы парсера следующий: на входе парсер получает два файла: RDFs-файл с созданными экземплярами класса (source.txt) и полученный txt-файл с данными расходов бюджета Санкт-Петербурга (data.txt). В парсере открываются два потока для чтения данных из файлов и поток для записи результатов

работы программы в файл (result.txt). После считывания всех данных в массив парсер построчно сохраняет значение кодов из файла data.txt в переменные, одноименные будущим свойствам экземпляров класса Costs, и записывает их в выходной файл с соблюдением синтаксиса RDFs-файла. Полученный в результате работы программы файл с экземплярами класса Costs копируется в файл с экземплярами классов онтологии. Таким образом, в результате получается онтология расходов бюджета Санкт-Петербурга на 2012 г., состоящая из двух RDFs-файлов, пригодная для дальнейшей машинной обработки.

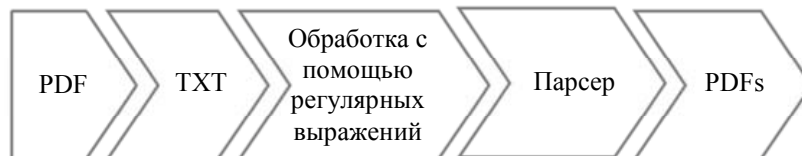


Рис. 3. Схема автоматизированного создания экземпляров класса Costs

Заключение

В работе представлены этапы разработки онтологии ведомственной структуры расходов бюджета Санкт-Петербурга – постановка задачи, исследование структуры данных, проектирование онтологии и ее реализация в системе Protégé. Результатом работы является описание методики создания онтологии открытых государственных данных и разработанный пример в формате RDFs-онтология, состоящий из 7 базовых классов и около 8 тыс. экземпляров. Полученную онтологию можно использовать в качестве источника данных для онлайн-сервисов, позволяющих осуществлять автоматизированную обработку и анализ открытых государственных данных. Примером такого сервиса может являться прототип системы публикации открытых государственных данных расходов бюджета, также полученный в рамках представленной работы. К реализованным функциям прототипа относятся загрузка массива данных (онтологии) и просмотр метаинформации о нем, выбор данных для построения таблиц и графиков и их отображение. На данный момент разработка находится на стадии оптимизации работы прототипа.

Литература

1. Доклад «Об обеспечении доступа населения к информации о деятельности государственных органов и органов местного самоуправления» [Электронный ресурс]. – Режим доступа: <http://ivan.begtin.name/wp-content/uploads/2011/09/doklad.pdf>, свободный. Яз. рус. (дата обращения: 17.05.2012).
2. Закон Санкт-Петербурга от 26.10.2011 № 658-120 «О бюджете Санкт-Петербурга на 2012 год и на плановый период 2013–2014 годов».
3. Законы о бюджете Санкт-Петербурга // Комитет финансов Санкт-Петербурга [Электронный ресурс]. – Режим доступа: <http://www.fincom.spb.ru/comfin/budget/laws.htm>, свободный. Яз. рус. (дата обращения: 01.10.2012).
4. Портал открытого бюджета города Москвы [Электронный ресурс]. – Режим доступа: <http://budget.mos.ru>, свободный. Яз. рус. (дата обращения: 01.10.2012).
5. Указания о порядке применения бюджетной классификации РФ (утверждены Приказом Министерства финансов РФ от 28.12.2010 № 190н) [Электронная версия]. – Режим доступа: <http://www1.minfin.ru/ru/budget/classandaccounting/classification/>, свободный. Яз. рус. (дата обращения: 01.10.2012).
6. Муромцев Д.И. Онтологический инжиниринг знаний в системе Protégé. Методическое пособие, Санкт-Петербург, 2007 г. [Электронная версия]. – Режим доступа: <http://books.ifmo.ru/book/pdf/243.pdf>, свободный. Яз. рус. (дата обращения: 01.10.2012).
7. Сайт ПО Protégé Университет Стэнфорда [Электронный ресурс]. – Режим доступа: <http://protege.stanford.edu/>, свободный. Яз. рус. (дата обращения: 01.10.2012).
8. OWL Web Ontology Language Guide [Электронная версия]. – Режим доступа: <http://www.w3.org/TR/owl-guide/>, свободный. Яз. рус. (дата обращения: 01.10.2012).

Пархимович Ольга Владимировна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, olya.parkhimovich@gmail.com

Власов Виталий Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, inxhaos@gmail.com

Муромцев Дмитрий Ильич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, mouromtsev@mail.ifmo.ru

УДК 548.73, 539.264

РЕНТГЕНОФАЗОВЫЙ АНАЛИЗ ИТТЕРБИЙ-ЭРБИЕВЫХ ОКСИФТОРИДНЫХ НАНОСТЕКЛОКЕРАМИК

А.О. Трофимов, А.Ю. Бирик, Р.К. Нуриев, Н.В. Никоноров, Е.В. Колобкова, В.А. Асеев

Исследованы кристаллические фазы наностеклокерамик, активированных ионами иттербия и эрбия, методом рентгенофазового анализа. Определены состав кристаллической фазы, размер кристаллов и параметры элементарной ячейки.

Ключевые слова: рентгеновская дифракция, кристаллическая фаза, наностеклокерамика.

В настоящее время прозрачные фторсодержащие наностеклокерамики представляют большой интерес в качестве активных сред для лазеров, усилителей и конверторов света [1]. При разработке таких материалов необходимо контролировать состав и размер кристаллической фазы. Такой контроль необходим для управления степени вхождения активатора в кристаллическую фазу и уменьшения уровня светорассеяния на границе раздела фаз. Настоящая работа посвящена рентгенофазовому исследованию свойств свинцово-фторидных наностеклокерамик, активированных ионами иттербия и эрбия.

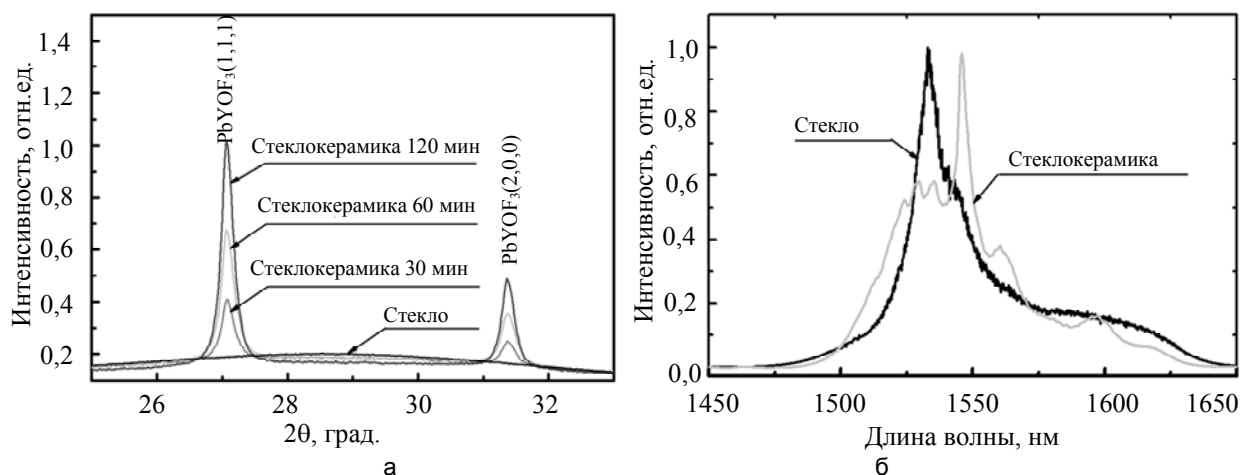


Рисунок. Дифрактограммы стекла и стеклокерамик, активированных эрбием (0,1 мол.%) (а); спектры поглощения стекла и стеклокерамики (б)

В работе исследованы иттербий-эрбиевые стекла состава $30\text{SiO}_2\text{-}15\text{AlO}_{3/2}\text{-}29\text{CdF}_2\text{-}18\text{PbF}_2\text{-}5\text{ZnF}_2\text{-}x\text{ErF}_3\text{-(}3\text{-}x\text{)YbF}_3$, где $x=0; 0,05; 0,1; 0,2; 0,5$. Для получения прозрачной наностеклокерамики исходные стекла подвергались вторичной термообработке при температуре 500°C . Температура вторичной термообработки выбиралась, исходя из данных, полученных методом дифференциальной сканирующей калориметрии. Время термообработки составило 30, 60 и 120 мин. Рентгенофазовый анализ проводился на длине волны $\lambda=1,5418 \text{ \AA}$ ($\text{CuK}\alpha$).

Исследования стеклокерамик с замещенными на иттрий ионами редкой земли показали, что в результате термообработки формируется кубическая гранцентрированная элементарная ячейка, в кристаллографическом отношении соответствующая кристаллу PbYOF_3 [2]. Постоянная решетки равна $5,74 \text{ \AA}$. При активации иттербием и эрбием формируется аналогичная флюоритоподобная кристаллическая фаза. Однако постоянная решетки в этом случае изменяется и равна $5,79 \text{ \AA}$. Размеры кристалла рассчитаны по формуле Шеррера и достигают 40 нм . Увеличение времени термообработки приводит к увеличению интенсивности дифракционных пиков (рисунок, а). Однако ширина пиков по 2θ на полувысоте не меняется. Это означает, что увеличение времени термообработки не приводит к значительному изменению размера кристаллов, а изменение интенсивности пиков связано с увеличением объема кристаллической фазы. При этом процесс объемной кристаллизации полностью проходит за 2 часа. Увеличение постоянной решетки свидетельствует о том, что редкоземельные ионы переходят в кристалл, замещая иттрий, т.е. формируется кристалл вида $\text{PbYb}_{(1-x)}\text{Er}_x\text{OF}_3$. Это подтверждается спектрами поглощения для данных наностеклокерамик (рисунок, б). Исходя из этого, можно заключить, что присутствие фторидов редкоземельных ионов полностью обуславливает выделение флюоритоподобной кристаллической фазы, а их концентрация в исходном стекле полностью определяет объем выделившейся фазы. Таким образом, прозрачные оксифторидные наностеклокерамики, легированные иттербием и эрбием, являются перспек-

тивными кандидатами для различных приложений фотоники как активные среды для лазеров, усилителей и конверторов света.

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (Соглашение № 14.В37.21.0169, Минобрнауки РФ).

1. Асеев В.А., Колобкова Е.В., Некрасова Я.А. Низкотемпературные измерения апконверсионной люминесценции в наноструктурированных активированных стеклокерамиках // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 3 (73). – С. 22–25.
2. Асеев В.А., Голубков В.В., Колобкова Е.В., Никоноров Н.В. Лантаноидные оксифториды свинца в стеклообразной матрице // Физика и химия стекла. – 2012. – Т. 38. – № 2. – С. 238–246.

Трофимов Александр Олегович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Eceptional777@mail.ru

Бибик Анастасия Юрьевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Anastasiya.bibik@list.ru

Нурьев Рустам Ккакбаевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Nuryev@oi.ifmo.ru

Никоноров Николай Валентинович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, зав. кафедрой, Nikonov@oi.ifmo.ru

Колобкова Елена Вячеславовна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор химических наук, профессор, Kolobok106@rambler.ru

Асеев Владимир Анатольевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, Aseev@oi.ifmo.ru

УДК 621.396

СПОСОБ РАСШИРЕНИЯ ФУНКЦИОНАЛЬНЫХ ВОЗМОЖНОСТЕЙ МОБИЛЬНОГО ТЕРМИНАЛА МОНИТОРИНГА ПОДВИЖНЫХ ОБЪЕКТОВ

А.А. Тихомолов, К.О. Ткачев, Н.С. Кармановский

Представлена структурная схема модульного мобильного терминала мониторинга подвижных объектов с применением двух однокристальных микроконтроллеров, объединенных одной системной шиной. Предложенное решение существенно расширяет функциональные возможности терминала и делает его легко адаптируемым под требования конкретных заказчиков. Схема реализована в конструкции блока, прошедшего сертификацию.

Ключевые слова: мобильный терминал, микроконтроллер, модульный принцип.

Мобильный терминал мониторинга подвижных объектов – устройство, устанавливаемое на транспортное средство и обеспечивающее автоматическое определение координат по спутниковым навигационным радиосигналам, передачу в центр мониторинга координатной и служебной информации о состоянии его исполнительных устройств. В случае необходимости мобильный терминал обеспечивает голосовую связь с центром мониторинга, прием сигналов из центра мониторинга и их обработку исполнительными устройствами. Типовая структурная схема мобильного терминала включает однокристальный микропроцессор и подключенные к нему навигационный приемник, радиоканальный модем передачи данных, энергонезависимую память, исполнительные устройства и блок управления питанием. Функциональные возможности мобильного терминала определяются составом входящих в него компонентов и программным обеспечением обработки сигналов.

При создании мобильных терминалов встают такие проблемы, как разнообразие и неравномерное распределение применяемых каналов передачи данных (используемые средства связи в конкретном регионе), большое число видов подвижных объектов контроля (автомобильный, железнодорожный, водный транспорт); большое количество и разнообразие используемых датчиков и исполнительных устройств (аналоговые, цифровые датчики, устройства вывода визуальной информации, устройства ввода информации); разнообразие программного обеспечения и протоколов, принимаемых мониторинговыми центрами. Следует также учитывать наличие в мире большого числа производителей навигационных приемников.

Традиционно мобильные терминалы создаются по принципу «все в одном», когда все устройства реализованы на одной плате и в одном корпусе («Титан-10» системы «Алмаз», «А3-М2 ГЛОНАСС» системы «Аркан», «ОБ1-ГЛОНАСС системы «Апро-Страж», Cyber GLX системы «BusinessNavigator» и др.) [1]. Такой вариант предпочтителен с точки зрения габаритных размеров, надежности и потребляемой мощности.

Функциональные возможности терминала, реализованного по модульному принципу, в конечном итоге определяются мощностью и архитектурой выбранного однокристального микроконтроллера. При попытке выполнения технического задания заказчика часто требуется доработка существующих вариантов устройства с изменением количества внешних интерфейсов и функций (если это позволяет используемый однокристальный микроконтроллер) либо разработка нового с применением более производительного однокристального микроконтроллера.

Авторами предложен и реализован вариант мобильного терминала с применением двух однокристальных микроконтроллеров. Первый является стандартным, а второй выбирается, исходя из конкретных задач расширения функций, поставленных заказчиком (рис. 1). Основной (базовый) модуль выполняется на микроконтроллере с минимальным набором функций. При расширении по требованию заказчика количества функций мобильного терминала, на последовательную системную шину устанавливается дополнительный микроконтроллер (дополнительный модуль), берущий на себя недостающие функции обработки сигналов и позволяющий увеличить число внешних интерфейсов.

Основной особенностью работы предлагаемой схемы является то, что оба однокристальных микроконтроллера постоянно выполняют обработку результатов измерений различных параметров системы и данных, полученных со встроенных аналого-цифровых преобразователей. В результате этого система, построенная на основе двух контроллеров, потребляет в два раза больше энергии, чем основанная на одном контроллере. Если в активном режиме работы увеличение потребляемой устройством мощности за счет дополнительного модуля практически не заметно, так как основными энергопотребителями являются радиомодем и навигационный приемник, то в режиме пониженного потребления (спящий режим) это приводит к почти двукратному повышению потребляемой мощности.

По этой причине для сохранения параметров устройств в режиме пониженного энергосбережения был реализован программный алгоритм изменения тактовой частоты работы микроконтроллеров. Снижение энергопотребления достигается снижением тактовой частоты работы обоих микроконтроллеров, когда каждый обрабатывает информацию независимо от другого. При приходе внешних событий, тревог или запросов, получивший их микроконтроллер переводит всю систему на более высокую тактовую частоту работы.



Рис. 1. Структурная схема разработанного мобильного терминала

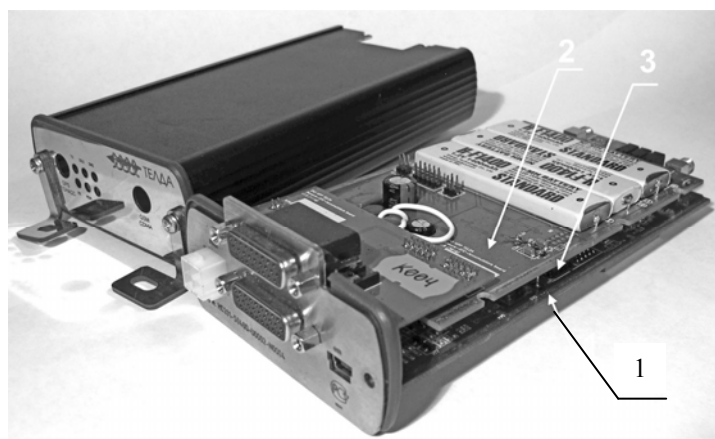


Рис. 2. Внешний вид устройства, извлеченного из корпуса: 1 – основной модуль; 2 – дополнительный модуль; 3 – разъем системной шины последовательной передачи данных

В реализованном проекте к основным функциям исходного микроконтроллера относятся обработка географических координат, полученных с навигационного приемника; получение информации с четырех цифровых входов; управление четырьмя внешними цифровыми выходами; управление громкой свя-

зью, питанием; передача информации через GSM- или CDMA-радиомодем (в зависимости от модификации).

Применение схемы с двумя микроконтроллерами позволило добавить в мобильный терминал 7 аналоговых линий для подключения внешних датчиков, расширить перепрограммируемую память до 8 МБ, создать интерфейс управления внешней УКВ радиостанцией, и подключить клавиатуру ввода данных. При этом габаритные размеры и потребляемая мощность остались практически прежними по сравнению с устройством, выполненном на одном микроконтроллере.

Схема и конструкция дополнительного модуля легко адаптируются под иные требования заказчика. Мобильный терминал мониторинга подвижных объектов GSM/CDMA-GPS/ГЛОНАСС реализован в виде модуля «Радай» (рис. 2) ЗАО «НПП ТЕЛДА». Модуль сертифицирован Федеральным агентством по техническому регулированию и метрологии, а также имеет сертификат СССТ МВД РФ.

Л. Методические рекомендации по применению спутниковых навигационно-мониторинговых систем на основе радионавигационной системы ГЛОНАСС в интересах органов внутренних дел. – М.: ГУ НПО «СТиС» МВД России, 2009. – 112 с.

Тихомолов Анатолий Алексеевич – ЗАО «НПП ТЕЛДА», начальник проектно-конструкторского отдела, a.tikhomolov@telda.ru

Ткачев Константин Олегович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ст. преподаватель, kottok1@yandex.ru

Кармановский Николай Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, karmanov50@mail.ru

УДК 681.786.4

ИССЛЕДОВАНИЕ АВТОКОЛЛИМАЦИОННОЙ СИСТЕМЫ ИЗМЕРЕНИЯ УГЛОВЫХ ДЕФОРМАЦИЙ КРУПНОГАБАРИТНЫХ ОБЪЕКТОВ

Т.В. Тургалиева, И.А. Коняхин

В лабораторных условиях исследованы характеристики системы определения угловых деформаций крупногабаритных объектов типа основного зеркала радиотелескопов. Результаты исследований подтвердили эффективность предложенной системы.

Ключевые слова: автоколлиматор, тетраэдрический отражатель, угол скручивания, коллимационные углы, деформации элементов радиотелескопа.

Для определения пространственного углового положения крупногабаритных объектов относительно некоторой базы широко используются автоколлимационные системы. В частности, исправление негомологических деформаций основного зеркала радиотелескопа миллиметрового диапазона длин волн подстройкой всей поверхности основного зеркала к положению идеальной параболы можно осуществить после измерения автоколлимационным методом отклонения нормали к поверхности зеркала от номинального положения в контролируемых точках. При этом затрагивается задача измерения взаимного положения объектов, под которыми понимаются два объекта, один из которых принимается за базовый (неподвижный), а второй – контролируемый – перемещается относительно него.

Для решения указанной задачи оптико-электронными измерительными средствами предлагается использовать трехкоординатный автоколлиматор с увеличенной чувствительностью измерения угла скручивания [Л], в котором в качестве контрольного элемента используется тетраэдрический отражатель (ТО), с заданными отступлениями δ от 90° двух двугранных углов между отражающими гранями. При отражении такой ТО разделяет падающий по оси пучок автоколлиматора на две пары пучков, одна из которых может использоваться для измерения угла скручивания. Каждый из пучков пары для измерения угла скручивания Θ_3 , составляет с осью объекта автоколлиматора угол Δ , который численно равен коэффиценту передачи между углом поворота ТО на угол скручивания и отклонением отраженного пучка от первоначального направления и определяет чувствительность измерения скручивания. Увеличение чувствительности Δ при использовании типового автоколлиматора с малым угловым полем обеспечивается переотражением пучка, реализуемым дополнительным плоским зеркалом. Коллимационные углы измеряются по части пучка, отраженной от фронтальной грани ТО, как от автоколлимационного зеркала.

В лаборатории кафедры оптико-электронных приборов и систем НИУ ИТМО по схеме трехкоординатного автоколлиматора [Л] изготовлен макет системы контроля угловых деформаций крупногабаритных объектов и собран экспериментальный стенд.

В состав стенда входят (рисунок): автоколлиматор 1; ТО, установленный на поворотный столик (ПС) 2 (ПС позволяет задавать повороты относительно коллимационных осей ОХ, ОУ на углы Θ_1 , Θ_2 и оси скручивания ОZ на угол Θ_3); персональный компьютер 3; блок питания источника излучения автоколлиматора 4; миллиамперметр 5; оптическая скамья ОСК-2. 6; визуальный автоколлиматор АКТ-15 с отражателем в виде плоского зеркала 7 для контроля углового положения ПС; блок питания источника излучения 8 автоколлиматора АКТ-15.

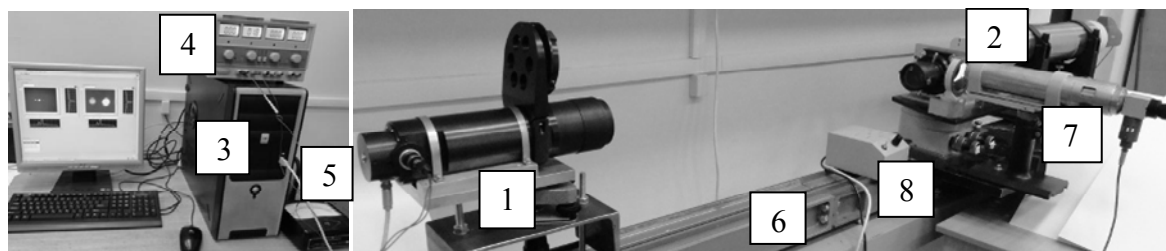


Рисунок. Стенд для испытания автоколлимационной системы измерения угловых деформаций крупногабаритных объектов

В ходе экспериментальных исследований доказана линейность статической характеристики автоколлимационной системы, при этом оценка среднеквадратического значения погрешности σ составила $20''$ при измерении угла скручивания в диапазоне $\pm 25'$ и $3''$ и при измерении коллимационных углов в диапазоне $\pm 10'$. Полученные метрологические характеристики соответствуют требуемым для систем измерения угловых деформаций элементов конструкции крупногабаритных объектов, в частности, зеркал современных радиотелескопов.

Работа проводилась при финансовой поддержке Министерства образования и науки Российской Федерации в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы.

[Л]. Коняхин И.А., Копылова Т.В. Трехкоординатный оптико-электронный автоколлиматор с увеличенной чувствительностью измерения угла скручивания // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 6 (70). – С. 9–11.

Тургулиева Татьяна Валерьевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, kopylova_tv@mail.ru

Коняхин Игорь Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, igor@grv.ifmo.ru

УДК 681.786.4, 681.7.08

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ОПТИКО-ЭЛЕКТРОННОЙ СИСТЕМЫ ИЗМЕРЕНИЯ ДЕФОРМАЦИЙ ЭЛЕМЕНТОВ КОНСТРУКЦИИ РАДИОТЕЛЕСКОПА РТ70 (СУФФА)

А.А. Усик

Предложена методика исследований многоматричной системы измерения деформаций зеркальных компонентов полноповоротных радиотелескопов. Приводятся результаты экспериментов с макетом системы.

Ключевые слова: радиотелескоп РТ70, оптико-электронная система, измерение деформаций элементов конструкции.

Для обеспечения работы современных полноповоротных радиотелескопов типа РТ70 в миллиметровом диапазоне длин волн необходима компенсация деформаций отражающих элементов с погрешностью не более $0,1$ мм. Применение оптико-электронных систем в качестве первичных измерительных преобразователей комплекса компенсации деформаций позволяет обеспечить требуемую точность измерения смещений элементов.

Оптико-электронная система измерения деформаций отражающих элементов радиотелескопа включает две подсистемы, расположенные на опорном кольце вблизи вершины главного зеркала [1].

Первая подсистема определяет координаты визирных целей (инфракрасных излучающих диодов), расположенных в контрольных точках поверхности главного зеркала, и является первичным звеном системы компенсации деформаций поверхности. Вторая подсистема измеряет положение визирных целей на контррефлекторе, что позволяет определить его смещение относительно номинального положения. Каждая подсистема состоит из однотипных базовых измерительных блоков (БИБ), измеряющих деформацию в нескольких (до 40) контрольных точках [2].

БИБ включает единый объектив (оптическую систему), секционный анализатор в виде группы матричных приемников оптического излучения (по количеству визирных целей) и устройство обработки видеокладов. Формируемое поле анализа блока образовано отдельными дискретными матричными приемниками оптического излучения (МПОИ), причем на каждом МПОИ формируется изображение только одной визирной цели (ВЦ), что определяет отсутствие перекрытия угловых полей поля, соответствующих отдельным МПОИ. Использование многоматричного секционного анализатора обусловлено широким диапазоном дистанций до ВЦ (3–40 м), определяющим значительное несовпадение плоскостей их изображений.

Для оценки эффективности разработанной концепции БИБ с секционным анализатором был создан макет, конструктивно состоящий из трех несущих пластин, соединенных наборными стойками. На первой пластине закреплен объектив РФ-5 ($F=450,36$ мм), на средней – пять фотоприемных модулей на базе

КМОП-матрицы OmniVision OV5620, на последней – устройство обработки. Для предотвращения фоновой засветки макет был закреплен на пластину-основание и закрыт прямоугольным кожухом.

Эксперименты проводились в полигоне – комнате размерами 18×6×4 м, в которой на имитаторе опорного кольца (массивном основании) был установлен макет БИБ. На стенах и консольных элементах полигона в соответствии с расположением контрольных точек главного зеркала радиотелескопа на различных дистанциях и высотах от БИБ закреплены четыре макета ВЦ в виде инфракрасных излучающих диодов Kingbright L-34SF4C с диафрагмами диаметром 0,3 мм. Диапазон расстояний от БИБ до ВЦ по линии визирования составляет 3–15 м. Две из четырех ВЦ расположены на моторизованных линейных подвижках Standa 8MT30-50DCE с разрешением 0,014 мкм. Перемещение ВЦ в заданном диапазоне имитирует смещение контрольных точек главного зеркала или контррефлектора вследствие деформаций.

При снятии статической характеристики макета БИБ по командам с управляющего компьютера подвижные ВЦ перемещались в заданном диапазоне с определенным шагом, при этом в каждой точке диапазона с МПОИ секционного анализатора проводился съем серии из 100 кадров с последующей их обработкой. Обработка результатов измерений включала следующие этапы: исключение известных систематических погрешностей, проверка гипотезы о принадлежности результатов наблюдений принятой модели распределения, удаление промахов, определение функции преобразования базового измерительного блока для каждой секции анализатора методами регрессионного анализа.

В результате установлено, что функции преобразования описываются линейной моделью с одной независимой переменной, при этом оценка среднеквадратичной погрешности измерения не превышает 0,025 мм на дистанции 15 м. Таким образом, для практических дистанций в 30 м, соответствующих размерам зеркальных элементов современных радиотелескопов миллиметрового диапазона, погрешность измерения оптико-электронной системы не превысит допустимой величины в 0,1 мм.

Исследования выполнены при финансовой поддержке Министерства образования и науки Российской Федерации в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы по государственному контракту № П684.

1. Артеменко Ю.Н., Коняхин И.А., Панков Э.Д., Тимофеев А.Н. Оптико-электронные системы измерения деформаций элементов конструкции радиотелескопа миллиметрового диапазона РТ-70 (Суффа) // Изв. вузов. Приборостроение. – 2008. – Т. 51. – № 9. – С. 5–10.
2. Михеев С.В., Усик А.А., Кулешова Е.Н. Многоканальная оптико-электронная система контроля деформаций сооружений // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 4 (74). – С. 159–160.

Усик Александр Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, usik.aa@gmail.com

УДК 338.45

ОРГАНИЗАЦИЯ ПЛАНИРОВАНИЯ ДЕЯТЕЛЬНОСТИ ПРИБОРОСТРОИТЕЛЬНОГО ПРЕДПРИЯТИЯ ПРИ ПРОВЕДЕНИИ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ И ОПЫТНО-КОНСТРУКТОРСКИХ РАБОТ В АВИАЦИОННОЙ ОТРАСЛИ П.П. Парамонов, Ю.А. Гатчин, И.О. Жаринов, О.О. Жаринов, Р.А. Шек-Иовсепянц

Рассматриваются формы планирования деятельности приборостроительного предприятия авиационной отрасли. Приводятся особенности стратегического, тактического и оперативного планирования.

Ключевые слова: планирование, научно-исследовательская работа, приборостроение.

Инновационная деятельность современного приборостроительного предприятия авиационной отрасли сопряжена с выполнением проектов в рамках реализации научно-исследовательских и опытно-конструкторских работ (НИОКР). Залогом успеха реализации НИОКР является эффективная система планирования проведения НИОКР. Как показано на рисунке, система планирования проведения НИОКР состоит из методов и средств реализации процедур:

- стратегического планирования;
- тактического планирования;
- оперативного планирования.

Стратегическое планирование проведения НИОКР основано на анализе научно-технического потенциала сотрудников предприятия, имеющихся у предприятия производственных мощностей и состоянии и перспектив развития авиационной отрасли на современном этапе. В значительной степени стратегический план деятельности предприятия определяется развиваемым в Российской Федерации институтом Федеральных целевых государственных программ, диктующим на рынке разработок авиационных приборов и систем динамику спроса и предложений.

Стратегический план деятельности предприятия является результатом творческой деятельности руководства предприятия и определяет производственно-исследовательскую политику предприятия на период 5–10 лет. Стратегический план деятельности предприятия составляется в рамках проведения мар-

кетинговых исследований, свидетельствующих о наличии в отрасли спроса на определенный вид продукции (авиационные приборы и системы). Результатом маркетинговых исследований является перечень перспективных к освоению изделий, востребованных на рынке. При составлении стратегического плана используются математические методы прогнозирования в эконометрии, статистические методы вычисления оценок финансовых показателей.

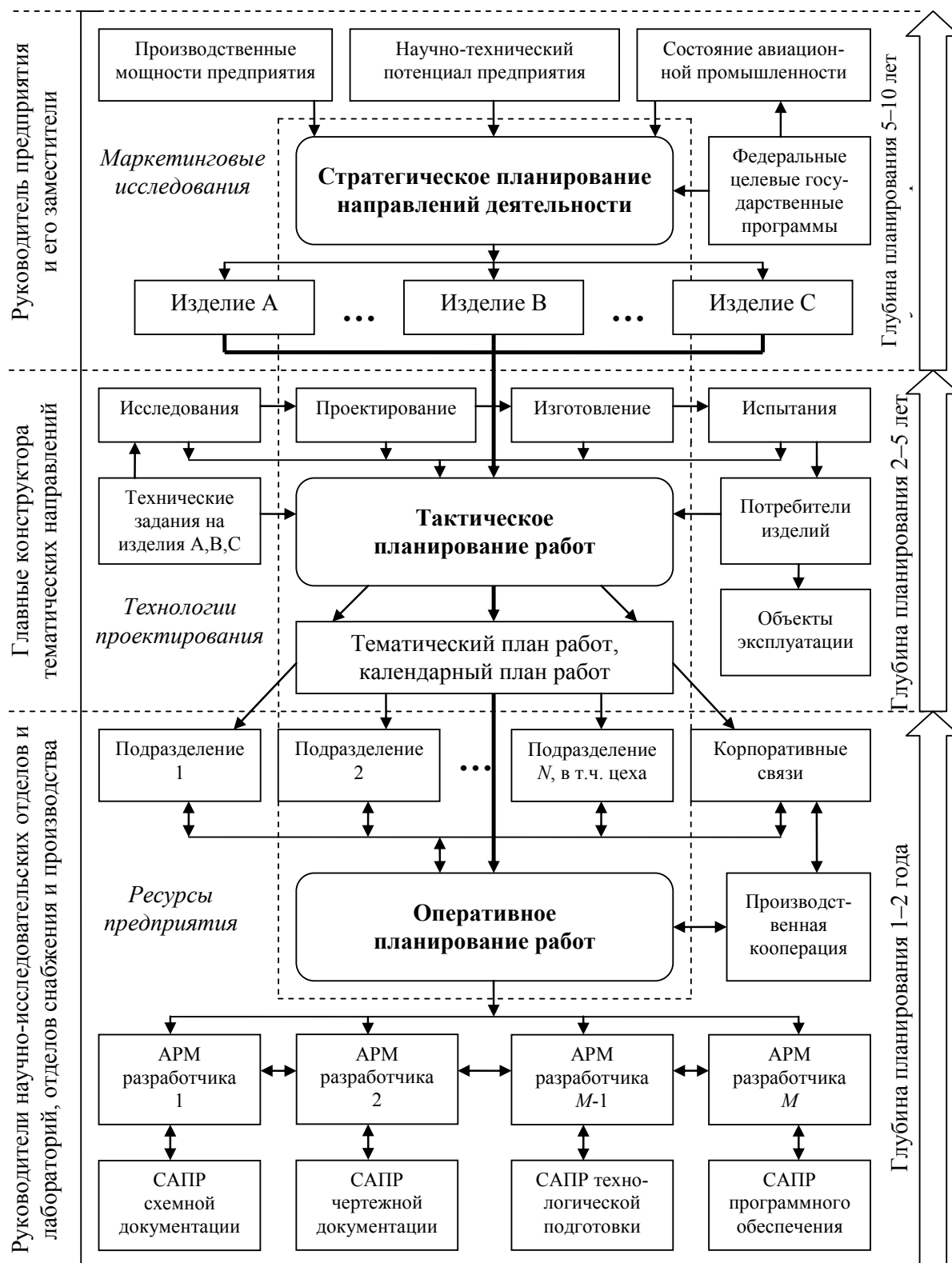


Рисунок. Схема организации и обеспечения на предприятии процессов планирования при выполнении НИОКР по разработке авиационного приборного оборудования

Тактическое планирование проведения НИОКР является результатом деятельности главных кон-

структуров тематических направлений предприятия и составляет основу календарного и тематического планов работ подразделений предприятия на период 2–5 лет. Тактическое планирование проведения НИОКР является следствием анализа:

- тактико-технических требований технического задания на вновь разрабатываемые изделия;
- необходимости проведения всех этапов проектирования изделий (техническое предложение, эскизное проектирование, технический проект и т.д.);
- трудоемкости проведения проектных и производственных работ, испытаний изделий;
- освоенных на предприятии технологий проектирования и производства.

При составлении календарного и тематического планов деятельности производственных подразделений предприятия принимаются во внимание сроки проектирования и изготовления продукции, а также желаемые сроки получения продукции потребителями – авиационными летно-испытательными базами, самолетостроительными компаниями, войсковыми частями и т.д. При составлении тактического плана используются математические методы маршрутизации и масштабирования сетей, сетевые графики.

Оперативное планирование проведения НИОКР является результатом деятельности руководителей научно-исследовательских и проектно-конструкторских подразделений, отделов снабжения и производства. Оперативное планирование осуществляется на период 1–2 года и является следствием анализа:

- имеющихся в распоряжении начальников отделов материально-технических ресурсов предприятия (автоматизированные рабочие места проектировщиков (АРМ), локальные вычислительные сети, специализированное сетевое программное обеспечение поддержки разработок и т.д.);
- обеспеченности АРМ системами автоматизации проектирования (САПР), реализующими единый технологический цикл «проектирование–производство–эксплуатация» авиационной продукции [Л];
- сложившихся в отрасли корпоративных связей между предприятиями, посредством которых реализуется система производственной кооперации и снабжения при разработке и изготовлении изделий.

При составлении оперативных планов используются математические методы составления расписаний, решаются задачи «о назначении», задача «коммивояжера».

Таким образом, совокупность форм стратегического, тактического и оперативного планирования проведения НИОКР является системообразующим фактором в реализации этапов НИОКР, объединяющим все виды имеющихся у предприятия ресурсов, учитывающим отраслевую роль и место предприятия в приборостроении, а также соответствующим интересам государства, выраженным в форматах федеральных целевых государственных программ, утверждаемых Правительством Российской Федерации.

Л. Гатчин И.Ю., Жаринов И.О., Жаринов О.О., Косенков П.А. Реализация жизненного цикла «проектирование–производство–эксплуатация» бортового оборудования на предприятиях авиационной промышленности // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 2 (78). – С. 141–143.

Парамонов Павел Павлович – ФГУП «СПб ОКБ «Электроавтоматика» имени П. А. Ефимова», доктор технических наук, профессор, директор, postmaster@elavt.spb.ru

Гатчин Юрий Арменакович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, gatchin@mail.ifmo.ru

Жаринов Игорь Олегович – ФГУП «СПб ОКБ «Электроавтоматика» имени П. А. Ефимова», доктор технических наук, доцент, руководитель учебно-научного центра, igor_rabota@pisem.net

Жаринов Олег Олегович – Санкт-Петербургский государственный университет аэрокосмического приборостроения, кандидат технических наук, доцент, zharinov@hotmail.ru

Шек-Иовсепяц Рубен Ашотович – ФГУП «СПб ОКБ «Электроавтоматика» имени П. А. Ефимова», доктор технических наук, профессор, главный конструктор, postmaster@elavt.spb.ru

УДК 681.786.4

СТЕРЕОСКОПИЧЕСКАЯ ОПТИКО-ЭЛЕКТРОННАЯ СИСТЕМА ДЕФОРМАЦИОННОГО МОНИТОРИНГА КРУПНОГАБАРИТНЫХ КУПОЛЬНЫХ КОНСТРУКЦИЙ К.Г. Араканцев, С.В. Михеев

Приведено обоснование возможности реализации и рассмотрены особенности применения стереоскопического метода измерения пространственных координат объектов в оптико-электронной системе контроля деформаций крупногабаритных купольных конструкций.

Ключевые слова: система стереоскопическая, контроль деформаций, погрешность, конструкция купольная.

В настоящее время одним из направлений повышения эффективности строительного производства является широкое применение легких пространственных конструкций, в том числе сетчатых оболочек и куполов [1]. Купольные конструкции широко используются при строительстве крупных спортивных, торговых, развлекательных площадок и прочих вместительных сооружений.

Купольные крупногабаритные сооружения различного назначения, как и многие другие объекты техногенной сферы, являются потенциально опасными в плане угрозы жизни людей со стороны возмож-

ных техногенных катастроф. Причины опасных нарушений функционирования объектов могут заключаться в ошибках, допущенных на стадии конструирования, при строительстве и неправильной эксплуатации, а также вследствие влияния неблагоприятных природных факторов. Поэтому сегодня чрезвычайную актуальность приобретают измерительные системы, позволяющие в режиме реального времени осуществлять мониторинг технического состояния сооружений купольного типа, своевременно определять их опасное состояние и предотвращать тем самым большие человеческие жертвы, экономические потери и угрозы окружающей среде.

При решении таких задач эффективны оптико-электронные системы, реализующие стереоскопический метод измерений, который заключается в определении пространственного положения объекта по координатам его изображений, построенных объективами на матричных приемниках двух измерительных оптико-электронных преобразователей (ИОЭП). Важной особенностью конструкций типа «купол» является наличие несущего конструктивного элемента в виде опорного кольца, расположенного вблизи вершины купола. Опорное кольцо может использоваться в качестве базового элемента, по периметру которого с равным угловым интервалом расположены ИОЭП, составляющие систему мониторинга деформаций. Контрольные элементы (КЭ), выполненные в виде полупроводниковых излучающих диодов, закрепляются в отдельных точках несущей конструкции, как показано на рисунке.



Рисунок. Купольная конструкция с КЭ и ИОЭП по периметру опорного кольца

В случае купольной конструкции, показанной на рисунке, целесообразно использовать радиальную схему размещения КЭ. При условии, что в поле зрения каждого ИОЭП целиком попадает два радиальных сегмента конструкции, а размещение ИОЭП на опорном кольце выполнено так, что каждый сегмент оказывается общим для двух смежных ИОЭП (см. область перекрытия полей зрения ИОЭП на рисунке), становится возможна реализация стереоскопического метода измерения координат КЭ. Стереоскопический метод непрерывного деформационного мониторинга купольных конструкций обеспечивает СКО случайной погрешности не более 1 мм [2], что позволяет стереоскопическим оптико-электронным системам составить серьезную конкуренцию (в том числе по стоимости и степени автоматизации измерительного процесса) серийно выпускаемым лазерным системам сканирующего типа.

Исследование проведено в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы».

1. Тур В.И. Купольные конструкции: формообразование, расчет, конструирование, повышение эффективности: Учебное пособие. - М.: Издательство АСВ, 2004. – 96 с.
2. Михеев С.В., Араканцев К.Г., Копылова Т.В. Оптико-электронные системы контроля конструкций крупных промышленных сооружений по положению их элементов // Научно-технический вестник СПбГУ ИТМО, 2010, 4 (68), С. 119

Араканцев Константин Геннадьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, kostya3312@gmail.com

Михеев Сергей Васильевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, msv@grv.ifmo.ru

SUMMARY

OPTICAL AND OPTICAL ELECTRONIC SYSTEMS. OPTICAL TECHNOLOGIES CAD FEATURES FOR DIFFRACTION GRATINGS MODELING

V. Vasiliev, D. Vorontsov, A. Dragunov, I. Livshits, E. Sokolova

Computer-aided design features for various types of diffraction gratings using ZEMAX software are investigated. It is shown that the model set choice of diffraction gratings in the most cases is determined by technology of their manufacture. Guidelines for five types of gratings are provided: an elliptical one of the first and second type, a hologram, a binary surface of the first type and optically generated hologram. Connection between the modeling way of diffraction gratings and their manufacturing technology is shown. Recommendations are given for selecting the optimal model for threaded and holographic diffraction gratings.

Keywords: diffraction grating, spectrometer, hologram, phase coefficients.

Vladimir Vasilyev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Rector, D.Sc., Professor, corresponding member of RAS, Vasilev@mail.ifmo.ru

Dmitriy Vorontsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, engineer, dimitrion@hotmail.ru

Alexander Dragunov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, AlexDragunov27@yandex.ru

Irina Livshits – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, laboratory head, PhD, senior research scientist, Ecenter-optica10@yandex.ru

Elena Sokolova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior scientific researcher, D.Sc., elena@scarlet.nl

ABERRATION ANALYSIS OF THE FINE OPTICAL COMPONENT COMPOSITION WITH CONCENTRIC MENISCUS

V. Yezhova, V. Zverev, T. Tochilina

It is shown that in the image, formed by optical system, consisting of a fine lens and concentric to the final thickness meniscus input pupil of an eye, plananastigmatic aberration correction is possible to achieve. However, therewith the problem of the lenses material choice appears. Such analysis of correction possibilities of the optical system is executed in case, when fine component consists of two fine lenses with optical power of the different sign.

Keywords: image, optical system, fine component, aberration, concentric system, input pupil of an eye.

Vasilisa Yezhova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, engineer, evv_foist@mail.ru

Viktor Zverev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, post_vaz@rambler.ru

Tatyana Tochilina – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Assistant professor, PhD, tvtochilina@mail.ru

MODERNIZED SOFTWARE COMPLEX FOR SYNTHESIS AND RECONSTRUCTION OF FRESNEL HOLOGRAMS-PROJECTORS

S. Koreshev, A. Gromov, O. Nikanorov

A modernized version of specialized software complex for synthesis and reconstruction of Fresnel holograms-projectors is presented. The algorithm used in the complex is based on the new method of hologram-projectors synthesis - the object partitioning to the standard elements of its structure. The method advantage is the essential reduction of the required calculations. The performance estimation of the new method is given. Results of the successful experimental verification of modernized complex efficiency are given.

Keywords: holographic-projection photolithography, hologram-projector, synthesis of hologram-projectors, digital reconstruction of holograms, algorithm, software complex.

Sergey Koreshev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, Koreshev@list.ru

Alexander Gromov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Warhammer@bk.ru

Oleg Nikanorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, leading programming engineer, PhD, Nikanorovov@gmail.com

SPATIAL POSITION DETERMINATION METHOD FOR THE OBJECT NEAR WATER SURFACE**D. Krysin**

A non-contact method for spatial position measuring of the objects near the smooth water surface (sea roughness code 0 points) is proposed, based on machine vision technology application. The feature of this method is the possibility of simultaneous measurement of the object height and angular orientation. The algorithm of the measurement system operation is considered. Developed installation description is given intended for the proposed method study. The measurement accuracy of height in the range of 500-1000 mm was 2-3% during the experiment. Measurement error of the angular orientation did not exceed 10%. The error sources and ways of the measurement accuracy improvement are analyzed.

Keywords: altitude measurement, water surface, machine vision system, spatial position, experimental installation.

Dmitry Krysin – Saint Petersburg State University of Aerospace Instrumentation, Assistant, wtxt@ya.ru

PHOTONICS AND OPTICAL INFORMATICS**ROTATIONAL DEPENDENCE OF CORIOLIS AND FERMI TYPE RESONANCES IN SMALL LINEAR MOLECULES****M. Smirnov**

The dependence of effective parameters in model Hamiltonian is investigated from the point of view of applied ordering scheme of oscillatory rotational interactions. Relations for transformation generators in different ordering schemes of oscillatory rotational interactions are formulated by “disentanglement” technique. The operator of resonance interaction is selected from general operator in effective rotational Hamiltonian on the base of projection operator technique. At strong resonance interaction the dependence from ordering of oscillatory rotational interactions is shown to be weaker than for the cases of middle and weak resonances.

Keywords: oscillatory rotational interactions, Fermi and Coriolis type resonances, ordering schemes of oscillatory rotational interactions.

Maxim Smirnov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associated professor, PhD, windy_hub@mail.ru

SPECIAL FEATURES OF ITERATION METHODS FOR PHASE PROBLEM IN OPTICS**S. Nalegaev, N. Petrov, V. Bespalov**

The article deals with general conception for phase problem solution by means of extra data sets which are intensity distributions at image formation plane received at variation of one or several parameters of experimental setup. Offered conception allows easy modifying of iterative methods of wave front reconstruction in accordance with specificity of the current tasks. Insertion of essential differences in registering intensity distributions by change of varying parameter in the process of their recording and presence of mathematical model, accurately describing the process of radiation propagation through optical scheme are the key aspects. From this point of view possibilities of nonlinear optical effects application for wave front phase reconstruction were analyzed.

Keywords: phase problem, wave front phase reconstruction, nonlinearity, nonlinear optical effects.

Sergey Nalegaev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, fioi@inbox.ru, s.nalegaev@gmail.com

Nikolai Petrov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, junior scientific researcher, PhD, Nickolai.petrov@gmail.com

Victor Bespalov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, victorbespaloff@gmail.com

MANGANESE LUMINESCENCE IN FLUORINE-PHOSPHATE GLASSES**V. Aseev, E. Kolobkova, Ya. Nekrasova, N. Nikonorov, A. Rokhmin**

The fluorine-phosphate glasses doped with 2 - 20 mol% MnF_2 have been developed and synthesized. Active ions concentration dependence on the spectral and luminescent properties of glasses has been investigated. Application of red phosphor based on fluorine-phosphate glasses doped with MnF_2 for white light diodes was discussed.

Keywords: white light-emitting diodes, red phosphor, glasses doped with manganese, fluorine-phosphate glasses.

Vladimir Aseev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Assistant, Aseev@oi.ifmo.ru

Elena Kolobkova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Kolobok106@rambler.ru

Yana Nekrasova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Nekrasova@oi.ifmo.ru

Nicolai Nikonorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, Nikonorov@oi.ifmo.ru

Alexei Rokhmin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Assistant, Rokhmin@oi.ifmo.ru

ANALYSIS AND SYNTHESIS OF COMPLEX SYSTEMS ROBUSTNESS ANALYSIS OF THE ELECTRIC DRIVE NON-ADAPTIVE CONTROL SYSTEM WITH STRUCTURE AND PARAMETERS VARIATIONS

A. Abdullin, V. Drozdov

The article deals with the numerical methods of the electric drive linear non-adaptive control system stability robustness under the given quantitative estimates of variations in the object model.

Keywords: electric drive, non-adaptive control system, robustness, stability.

Artur Abdullin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, artur.abdullin@gmail.com

Valentine Drozdov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, drozdovuprint@rambler.ru

ACCELEROMETER DATA SPECTRAL ANALYSIS FOR VEHICLES DETECTION AND IDENTIFICATION

D. Obertov

The article deals with the problems of frequency and time analysis of road surface vibrations caused by passing vehicles. The vibrations are measured by accelerometers installed in the traffic lane markings. Data preprocessing is implemented before using them to analyze the parameters of the traffic flow. The results of spectral analysis and filtering of data are presented.

Keywords: accelerometer, vehicles detection, identification of vehicle classes, intelligent transport systems, spectra, spectral processing, spectral range, fast Fourier transform (FFT), short-windowed Fourier transform.

Dmitry Obertov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, dimondking@mail.ru

ANALYSIS OF LINEAR SYSTEMS WITH VARIABLE PARAMETERS FOR INTERVAL OBSERVER SYNTHESIS

S. Chebotarev, A. Kremlev

Synthesis conditions of interval observers for linear systems with unknown time-varying parameters are analyzed. Possible cases for compliance with the properties of the cooperative system are considered. The analysis of conditions is held making it possible to construct an interval observer for a given combination of the system properties. Such observer allows getting estimates of the system state variables which contain with assurance the actual value of the state at a given time.

Keywords: interval observers, cooperativeness, Metzler, parametric uncertainty.

Stanislav Chebotarev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, freest5@gmail.com

Artem Kremlev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, kremlev_artem@mail.ru

DVB-T FIELD MODELING IN URBAN AREAS

A. Lomakin, S. Shkolniy

Calculation results of DVB-T field modeling intensity made by Longley-Rice model are compared with country measurements. Prognoses and a series of intensity field country measurements are made for different types of wave propagation paths in the conditions of rugged country and dense urban building environment. Comparison results of model and experimental data are discussed.

Keywords: DVB-T, radio wave propagation, Longley-Rice model.

Alexander Lomakin – Far Eastern Federal University, lecturer, PhD, Associate professor, lomakin11@land.ru

Stanislav Shkolniy – Far Eastern Federal University, Postgraduate, nimagt@gmail.com

DEEP-WATER THREE-AXIS TILT SENSORS APPLICATION IN SOLVING GEOTECHNICAL MONITORING TASKS

T. Paderina, V. Yakovlev, Y. Gordeev

The article deals with application results of the first national measurement system with deep-water three-axis tilt sensors (DTT) based on micromechanical accelerometers (MMA) in the solution of highlands geotechnical monitoring on Olympic speedway Adler - Krasnaya Polyana. It is shown that this system provides on-line monitoring of horizontal displacement array of the ground during a long period of time, thereby providing the timely warning of possible adverse situations during construction and operation. The achieved accuracy level

equal to a few millimeters and relatively low cost makes this system more attractive, especially in the continuous monitoring organization of lengthy geological objects.

Keywords: geotechnical monitoring, micromechanical accelerometer, three-coordinate tilt sensor.

Tatiana Paderina – Concern CSRI Elektropribor, JSC, senior staff scientist, PhD, paderinata@rambler.ru

Vladimir Yakovlev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, v-v-y@ya.ru

Yuri Gordeev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, 133880@gmail.com

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES VERIFICATION OF REACTIVE PROGRAMS BY DEPENDENT TYPE SYSTEMS WITH STRUCTURAL INDUCTION

Ya. Malakhovskiy, G. Korneev

The article is devoted to a constructive proof (in Agda programming language) of the fact that Boolean formula satisfiability problem by means of sequent calculus and verification with CTL specifications can be implemented as structurally recursive dependently typed functions.

Keywords: Boolean formula satisfiability problem, verification, functional programming, type systems, dependent types, Agda.

Yan Malakhovskiy – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, trojan@rain.ifmo.ru

Georgiy Korneev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, kgeorgiy@rain.ifmo.ru

SIGNAL-TO-NOISE RATIO ESTIMATION ALGORITHM FOR SPEECH SIGNALS

M. Stolbov

We propose an estimation algorithm for integral value of signal-to-noise ratio (SNR) and its values in frequency domains for sound records quality testing in the speakers' verification system. The algorithm is characterized by robustness with respect to large variability in recording conditions and sound record quality, as well as the possibility to perform SNR estimation in real time, i.e. at the rate of the incoming signal. The algorithm is based on new procedures of noise spectrum estimation and speech frames detection. The algorithm permits to evaluate SNR in the range from +6dB to +26dB for recordings with a duration of 10s or longer.

Keywords: signal-to-noise ratio, speech frames detection, estimation of noise spectrum.

Mikhail Stolbov – "STC - Innovation", Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior research fellow, Associated professor, PhD, stolbov@speechpro.com

ANT COLONY OPTIMIZATION-BASED METHOD OF FINITE-STATE MACHINE INDUCTION

D. Chivilikhin, V. Ul'yantsev

The article deals with a method for inducing finite-state machines for a given fitness function, based on an ant colony optimization algorithm. An experimental evaluation of the proposed method is presented on the problem of "Intelligent ant". The results of experimental evaluation show that the proposed method outperforms genetic algorithms on this problem. Furthermore, the proposed method doesn't require an implementation of some non-trivial operators for such problem (i.e. recombination and mutation operators in genetic algorithms).

Keywords: finite-state machines, ant colony optimization algorithms.

Daniil Chivilikhin - Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, chivilikhin.daniil@gmail.com

Vladimir Ul'yantsev - Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, ulyantsev@rain.ifmo.ru

C/C++ PROGRAMMING OF DISTORTED IMAGE RESTORATION PROBLEMS IN TEXAS INSTRUMENTS SIGNAL MICROPROCESSORS

K. Kir'yanov, V. Sizikov

An instrumental realization of restoration algorithms for distorted (smeared, defocused and (or) noisy) images is considered. Specific features of programming in C/C++ language of previously developed and published algorithms for direct and, more complicated and important for practice, inverse problem are considered.

Keywords: image distortion, direct and inverse problems, C/C++ language, signal microprocessor, personal computer.

Konstantin Kir'yanov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Postgraduate, kiryancon@front.ru

Valery Sizikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, sizikov2000@mail.ru

FINITE-STATE MACHINE IMPLEMENTATION OF EMBEDDED SYSTEM USER INTERFACE BASED ON ASPECT-ORIENTED DESCRIPTION

V. Vasyukov

This article describes the technique of transition from user interface definition of embedded system by aspect-oriented approach to definition in the form of a finite-state machine. The applicability of the proposed method is shown on the example of some physical value indicating device. The method describes gradual transition from the initial description of the user interface in terms of the aspect-oriented approach to a directed graph, describing the possible state changes of the user interface, and from the directed graph to a deterministic finite automaton, the mathematical description of which makes it possible to use it as a framework of the future user interface prototype. Examined method of transition from the user interface description of embedded system in terms of the aspect-oriented approach to the user interface description of embedded system as a finite state automaton shows the feasibility of aspect-oriented approach for the development and prototyping of user interfaces of embedded systems.

Keywords: aspect-oriented approach, interface, finite-state machine.

Vasily Vasyukov - Baltic State Technical University "Voenmekh" named after D.F. Ustinov, postgraduate, witcher237@rambler.ru

HISTOGRAM NORMALIZATION OF SPEECH FEATURES FOR SPEAKERS' VERIFICATION

Yu. Matveev, A. Shulipa

The article provides a brief description of the algorithm for histogram normalization of speech features for speakers' verification. Results of verification tests with different parameters and modes of normalization are presented. Obtained data make it possible to conclude about the effectiveness of the speech features normalization for increasing the performance of speakers' verification. Optimal parameters of the normalization algorithm were found.

Keywords: speakers' verification, speech features, histogram normalization.

Yuri Matveev – "STC - Innovation", LLC, chief researcher; Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., matveev@speechpro.com

Andrei Shulipa – "STC -Innovation", LLC, scientific researcher, shulipa@speechpro.com

IDENTIFICATION SYSTEM OF THE SPEAKER'S AGE GROUP BY SPONTANEOUS SPEECH RECORD

K. Simonchik

The paper proposes to use the popular text independent speaker identification method based on i-vectors extraction for the solution of the speaker's age group identification. Two realizations of speaker's age group detector based on using i-vectors are researched: an approach based on a support vector machine classifier and another one based on a multiple linear regression. Both methods provide a high accuracy rate of speaker's age group detection by spontaneous speech recordings. The average rate of correct speaker's age group identification was 61% and 65% respectively for the NIST SRE 2008 speech database.

Keywords: age, i-vector, SVM, linear regression.

Konstantin Simonchik – «Speech Technology Center» Ltd., Head of section; Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associated Professor, PhD, simonchik@speechpro.com

GENOME CONTIGS ASSEMBLY METHOD BASED ON DE BRUJIN GRAPHS AND OVERLAP GRAPHS

A. Alexandrov, S. Kazakov, S. Melnikov, A. Sergushichev, F. Tsarev

A method for genome contigs assembly is proposed. This method consists of two stages – quasicontigs assembly and contigs assembly. De Brujin graphs are used on the first stage and overlap graphs on the second one. The method was tested on *Maylandia zebra* fish genome containing nearly 1 billion nucleotides. Combining de Brujin graphs and overlap graphs makes this method much more space-efficient than the others.

Keywords: genome assembly, contigs, de Brujin graph, overlap graph.

Anton Alexandrov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, alantbox@gmail.com

Sergey Kazakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, kazakov_sergey_v@mail.ru

Sergey Melnikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, melnikov@rain.ifmo.ru

Alexei Sergushichev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, alsergbox@gmail.com

Fedor Tsarev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, fedor.tsarev@gmail.com

MATERIALS TECHNOLOGY AND NANOTECHNOLOGIES

LEAD-FLUORINE NANOGLOSS CERAMIC INVESTIGATION DOPED WITH YTTRIUM AND NEODYMIUM IONS

A. Bibik, A. Trofimov, R. Nuryev, V. Aseev, E. Kolobkova, N. Nikonorov

The spectral-luminescent and crystallization properties of lead-fluorine glass and nanoglass ceramic doped with yttrium and neodymium have been investigated. The precipitation of crystalline phases after heat treatment has been observed. The crystal structure, lattice constant and size of the crystalline phase have been determined.

Keywords: activated glass ceramic, x-ray phase analysis, crystalline phase, luminescence.

Anastasiya Bibik – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Anastasiya.bibik@list.ru

Alexander Trofimov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Exeptional777@mail.ru

Rustam Nuryev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Nuryev@oi.ifmo.ru

Vladimir Aseev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Assistant, Aseev@oi.ifmo.ru

Elena Kolobkova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Kolobok106@rambler.ru

Nicolai Nikonorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, Nikonorov@oi.ifmo.ru

SOFTWARE MODULE FOR PARAMETERS ESTIMATING OF DISPERSE SYSTEMS BY THEIR IMAGES FOR QUANTITATIVE THREE-DIMENSIONAL STEREO ANALYSIS

A. Kalinin, V. Manoilov, O. Prikhod'ko

The article deals with the possibility of using MatLab programming environment for statistical stereo analysis of disperse systems by their two-dimensional images. We describe features and results of the software module application prepared for creation of a probabilistic, statistically specific 3D-model, which corresponds to a more precise description of the structure and prediction for disperse systems properties.

Keywords: programming environment, statistical analysis, disperse systems, microstructure.

Anton Kalinin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, formant2004@mail.ru

Vladimir Manoilov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Associate professor, manoilov_vv@mail.ru

Oleg Prikhod'ko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, oap10@yandex.ru

INFLUENCE OF FILLER SURFACE MODIFICATION ON LATEX FILMS PROPERTIES

E. Ryutkyan, N. Sirotinkin, M. Uspenskaya

The article presents research results of the polymeric films properties based on chloroprene rubber latex, filled with modified glass microspheres, where a thin polymer film coating is used as a modifier. Influence regularities of these fillers on the composite materials properties are given.

Keywords: filler, glass microspheres, adhesion, chloroprene latex, modification.

Evgenia Ryutkyan – Saint Petersburg State Technological Institute (Technical University), postgraduate, ryutkyan.evgen@mail.ru

Nikolai Sirotinkin – Saint Petersburg State Technological Institute (Technical University), Dean, D.Sc., Professor, biotechnology_faculty@technolog.edu.ru

Maya Uspenskaya – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Associate professor, mv_ushenskaya@mail.ru

COMPUTER-AIDED DESIGN SYSTEMS

PRINCIPLES OF BRANCH SYSTEM CREATION FOR THE AUTOMATED DESIGN IN AVIATION INSTRUMENTATION

P. Paramonov, Yu. Gatchin, I. Zharinov, O. Zharinov, M. Deyko

Principles of branch system creation for the automated design of the onboard aviation equipment are considered. Products design routes and software functional diagram of the branch design system are given. Hardware realization solutions of the automated design system based on the local computer network of the instrument-making enterprise are described.

Keywords: branch automated design system, aviation equipment.

Pavel Paramonov – Saint Petersburg Scientific Design Bureau “Electroavtomatika” named after P. A. Efimov, Director, D.Sc., Professor, postmaster@elavt.spb.ru

Yuri Gatchin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, gatchin@mail.ifmo.ru

Igor Zharinov – Saint Petersburg Scientific Design Bureau “Electroavtomatika” named after P. A. Efimov, Head of Educational and Scientific Centre, D.Sc., Associate professor, igor_rabota@pisem.net

Oleg Zharinov – Saint Petersburg State University of Aerospace Instrumentation, Associate professor, PhD, Associate professor, zharinov@hotmail.ru

Mikhail Deyko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, postmaster@elavt.spb.ru

CALCULATION OF THE MINIMUM PATH LENGTH WIRE IN THE PRINTED WIRING TOPOLOGICAL ROUTING

Yu. Popov, S. Popov

There is a problem to calculate the path (shape) of a wire by its topological path in the topological approach to the wire routing. An algorithm which calculates the minimum length path of the wire considering design rules is proposed. The algorithm is used for wire shapes calculation in TopoR electronic design automation.

Keywords: EDA, automatic routing, topological routing.

Yuri Popov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, yurpopov@rambler.ru

Sergey Popov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, sergey.popove@yandex.ru

INFORMATION SECURITY METHODS AND SYSTEMS ALGORITHM OF SPAM-BELONGING DEFINITION FOR DOCUMENTS ON THE BASIS OF FEJER DISPLAYS

A. Korobeynikov, S. Blinov, A. Leyman, G. Markina, I. Kutuzov

The algorithm for definition of e-mail documents belonging to spam on the basis of support vector method is offered. Method modification consists of a dividing hyperplane creation, based on Fejer displays, that allows working with non-stationary data, characteristic for document classification problems.

Keywords: information classification, spam, problems of strong separability, Hilbert space.

Anatoly Korobeynikov – Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences Saint Petersburg Branch (SPbF IZMIRAN), Deputy director for science, D.Sc., Professor, Korobeynikov_A_G@mail.ru

Stanislav Blinov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, stasblino@yandex.ru

Albert Leyman – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, allxxl@yandex.ru

Galina Markina – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Section head, lina@mail.ifmo.ru

Il'ya Kutuzov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, formalizator@gmail.com

ANALYSIS OF INTERNET USER IDENTIFICATION METHODS

E. Bessonova, I. Zikratov, V. Roskov

The article deals with mechanisms for user identification in the Internet. A comparative analysis of two ways for user identification based on entropy and regression analysis is proposed. A computational experiment was conducted for results verification.

Keywords: identification, self-descriptiveness, feature, tuple, user.

Ekaterina Bessonova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, bessonova@cit.ifmo.ru

Igor Zikratov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, zikratov@cit.ifmo.ru

Vladislav Roskov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, vos@vos.uz

DEPERSONALIZATION ALGORITHM OF PERSONAL INFORMATION

A. Kurakin

The article deals with the problem of information protection for personal data information systems at development and optimization stages. For solution of this problem depersonalization algorithm is offered to decrease a class of personal data information systems that, in turn, removes the main part of requirements for application of organizational measures and means of information protection.

Keywords: depersonalization, depersonalization algorithm, information protection, personal data, personal data information systems.

Alexander Kurakin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, nirtit@gmail.com

ALGORITHM OF INDIVIDUALS' DETAILS IDENTIFICATION IN THE DATABASES BY LEVENSTEIN METRIC

N. Limanova, M. Sedov

A problem of individuals' personal identification with partially or fully conflicting details occurs when transferring data from one institution to another. An optimal identification algorithm is proposed based on Levenshtein metric that makes it possible to search individuals in the database by fuzzy comparison. The algorithm is implemented in PL-SQL language in Oracle 11g database.

Keywords: interdepartmental information exchange, identification, fuzzy comparison, search of individuals' details identification, function of intellectual comparison, personal identification number (PIN).

Nataliya Limanova – Togliatti State University, Professor, D.Sc., Professor, Natalya-Igorevna@yandex.ru

Maksim Sedov – Togliatti Mayor's Office, engineer-programmer of the 1st category, SedovMN@inbox.ru

ECONOMICS AND FINANCES. MANAGEMENT ONTOLOGY DEVELOPMENT OF OPEN GOVERNMENT DATA ON EXAMPLE OF ST. PETERSBURG BUDGET EXPENDITURES

O. Parkhimovich, V. Vlasov, D. Mouromtsev

This issue examines development stages of open government data publication system for St. Petersburg budget departmental expenditures. The reasons for ontology creation are given from the point of view unified terminological dictionary formation for users and for integration of information resources and systems. The common ontology structure and the process of ontology developing for this system are described. Suggested design method makes it possible to modify and update developed ontology with the change in input data or regulatory papers.

Keywords: open government data, ontology, semantic web.

Olga Parkhimovich – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, olya.parkhimovich@gmail.com

Vitaly Vlasov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Assistant, inxaoc@gmail.com

Dmitry Mouromtsev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, mouromtsev@mail.ifmo.ru

BRIEF REPORTS

X-RAY DIFFRACTION ANALYSIS OF OXIFLUORIDE NANOGLOSS CERAMIC DOPED WITH YTTERBIUM AND ERBIUM

A. Trofimov, A. Bibik, R. Nuryev, N. Nikonorov, E. Kolobkova, V. Aseev

The precipitating of the crystalline phase in oxifluoride nanogloss ceramic doped with ytterbium and erbium has been observed. The crystal structure, lattice constant and size of the particles have been determined.

Keywords: X-ray diffraction, crystalline phase, nanogloss ceramic.

Alexander Trofimov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Exeptional777@mail.ru

Anastasia Bibik – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Anastasiya.bibik@list.ru

Rustam Nuryev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Nuryev@oi.ifmo.ru

Nicolai Nikonorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, Nikonorov@oi.ifmo.ru

Elena Kolobkova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Kolobok106@rambler.ru

Vladimir Aseev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Assistant, Aseev@oi.ifmo.ru

METHOD OF FUNCTIONALITY EXPANSION OF MOBILE TERMINAL FOR MOBILE OBJECTS MONITORING

A. Tikhomolov, K. Tkachev, N. Karmanovsky

The article deals with a scheme of module mobile terminal for monitoring of mobile objects (GLONASS/GPS tracker) using two single-chip microcontrollers, connected through a common system bus. This method provides

a large functionality expansion of the tracker and makes it easily adapting according to the demands of customers. The scheme is implemented in the certified tracker construction.

Keywords: mobile terminal, microcontroller, modular approach.

Anatoly Tikhomolov – JSC «NPP TELDA», Head of R&D department, a.tikhomolov@telda.ru

Konstantin Tkachev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior lecturer, kottok1@yandex.ru

Nicolay Karmanovsky – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, karmanov50@mail.ru

RESEARCH OF AUTOCOLLIMATING ANGULAR DEFORMATION MEASUREMENT SYSTEM FOR LARGE-SIZE OBJECTS

T. Turgalieva, I. Konyakhin

Characteristics of the system were studied in laboratory conditions. The research confirmed effectiveness of the proposed angular deformation measurement system for large-size objects such as the primary mirror of a radio telescope.

Keywords: autocollimator, tetrahedral reflector, roll angle, pitch and yaw angles, radio telescope units deformation.

Tatiana Turgalieva – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, kopylova_tv@mail.ru

Igor Konyakhin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, igor@grv.ifmo.ru

EXPERIMENTAL RESEARCH OF THE OPTICAL ELECTRONIC SYSTEM FOR DEFORMATION MEASUREMENT OF STRUCTURAL ELEMENTS OF THE RT70 (SUFFA) RADIO TELESCOPE

A. Usik

Research method of multi matrix system for deformation measuring of mirror elements of rotary radio telescopes is suggested. Results of the model experiment are presented.

Keywords: RT70 radio telescope, optical electronic system, structural elements deformation measurement.

Alexander Usik – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, usik.aa@gmail.com

ORGANIZATION AND PLANNING OF THE INSTRUMENT-MAKING ENTERPRISE ACTIVITIES AT CARRYING OUT OF SCIENTIFIC-RESEARCH AND EXPERIMENTAL-DESIGN WORKS IN AVIATION

P. Paramonov, Yu. Gatchin, I. Zharinov, O. Zharinov, R. Shek-Iovsepjantz

Forms of planning for the activities of the instrument-making enterprise in aviation are considered. The features of the strategic, tactical and operational planning are given.

Keywords: planning, scientific-research work, instrument-making.

Pavel Paramonov – Saint Petersburg Scientific Design Bureau “Electroavtomatica” named after P. A. Efimov, Director, D.Sc., Professor, postmaster@elavt.spb.ru

Yuri Gatchin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, gatchin@mail.ifmo.ru

Igor Zharinov – Saint Petersburg Scientific Design Bureau “Electroavtomatica” named after P. A. Efimov, Head of Educational and Scientific Centre, D.Sc., Associate professor, igor_rabota@pisem.net

Oleg Zharinov – Saint Petersburg State University of Aerospace Instrumentation, Associate professor, PhD, Associate professor, zharinov@hotmail.ru

Ruben Shek-Iovsepjantz – Saint Petersburg Scientific Design Bureau “Electroavtomatica” named after P. A. Efimov, Chief constructor, D.Sc., Professor, postmaster@elavt.spb.ru

STEREOSCOPIC OPTICAL ELECTRONIC SYSTEM FOR DEFORMATION MONITORING OF LARGE-SIZE DOME STRUCTURES

K. Arakantsev, S. Mikhhev

The article deals with implementation possibility reasoning and application features of stereoscopic method for objects coordinates measuring in optical electronic system for deformation monitoring of large-size dome structures.

Keywords: stereoscopic system, deformation monitoring, measurement error, dome structure.

Konstantin Arakantsev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, kostya3312@gmail.com

Sergei Mikhhev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, msv@grv.ifmo.ru