

16. Swail C., Jennings S. Enhanced and synthetic vision system concept for application to search and rescue missions // Symposium on «Sensor Data Fusion and Integration of the Human Element». Ottawa, 1998. P. 15-1–15-6.
17. Young S., Kakarlapudi S., Uijt de Haag M. A Shadow detection and extraction algorithm using digital elevation models and X-Band weather radar measurements // International Journal of Remote Sensing. 2005. V. 26. N 8. P. 1531–1549.
18. Honeywell's primus epic integrated avionics system provides advanced flight deck functionality. SmartView synthetic vision system [Электронный ресурс]. Режим доступа: [http://www.honeywellbusinessaviation.com/primus\\_epic/performance/smartview](http://www.honeywellbusinessaviation.com/primus_epic/performance/smartview), свободный. Яз. англ. (дата обращения 28.11.2013).

- Костишин Максим Олегович** – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; инженер, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, job.max@me.com
- Жаринов Игорь Олегович** – доктор технических наук, доцент, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; руководитель учебно-научного центра, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, igor\_rabota@pisem.net
- Жаринов Олег Олегович** – кандидат технических наук, доцент, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, Россия, zharinov73@hotmail.ru
- Нечаев Владимир Анатольевич** – доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; руководитель научно-исследовательского центра, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, nil-12@mail.ru
- Суслов Владимир Дмитриевич** – доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; руководитель экспертного совета, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, postmaster@elavt.spb.ru
- Maxim Kostishin** – postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; engineer, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov, Saint Petersburg, Russia, job.max@me.com
- Igor Zharinov** – D.Sc., Associate professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Head of learning scientific center, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov, Saint Petersburg, Russia, igor\_rabota@pisem.net
- Oleg Zharinov** – PhD, Associate professor, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia, zharinov73@hotmail.ru
- Vladimir Nechaev** – Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Head of scientific research center, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov Saint Petersburg, Russia, nil-12@mail.ru
- Vladimir Suslov** – Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Head of expert council, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov, Saint Petersburg, Russia, postmaster@elavt.spb.ru

УДК 007.51

## ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ БАКТЕРИЙ В ПРОДУКТАХ ПИТАНИЯ

А.П. Саенко<sup>а</sup>, В.М. Мусалимов<sup>а</sup>, Ш. Лерм<sup>б</sup>, Г. Линц<sup>б</sup>

<sup>а</sup>Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, musvm@yandex.ru

<sup>б</sup>Технический университет Ильменау, Ильменау, Германия, steffen.lerm@tu-ilmenau.de

Рассматривается один из способов решения актуальной проблемы обеспечения контроля качества продуктов питания с использованием методов машинного обучения. Существующие в настоящее время методы анализа требуют специального лабораторного оборудования, значительного времени и сильно зависят от квалификации и некоторых физиологических особенностей эксперта, в то время как предлагаемый метод позволяет существенно снизить затраты за счет автоматизации процесса. Рассмотрено устройство, реализующее данный метод, принцип действия которого основан на флуоресцентной микроскопии. Для набора обучающих данных решена задача классификации объектов на

классы «бактерия» и «сторонний артефакт», проведен сравнительный анализ различных алгоритмов классификации (метод опорных векторов, случайный лес, деревья решений C4.5, метод  $k$  ближайших соседей, метод Байеса), который показал, что наилучшей эффективностью обладают метод опорных векторов и случайный лес.

Данная работа выполнена на кафедре мехатроники Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики и кафедре контроля качества и промышленной обработки изображений Технического университета г. Ильменау (Германия) в рамках совместной программы «Михаил Ломоносов» Министерства образования и науки Российской Федерации и Германской службы академических обменов.

**Ключевые слова:** машинное обучение, обнаружение бактерий.

## DETECTION OF BACTERIA IN FOODSTUFF BY MACHINE LEARNING METHODS

A. Saenko<sup>c</sup>, V. Musalimov<sup>c</sup>, S. Lerm<sup>d</sup>, G. Linss<sup>d</sup>

<sup>c</sup> Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, alexey.saenko@gmail.com

<sup>d</sup> Ilmenau University of Technology, Ilmenau, Germany, steffen.lerm@tu-ilmenau.de

The paper deals with an actual problem of ensuring the control of foodstuff quality by means of machine learning methods. Existing analysis methods require special laboratory environment, significant time and depend on the qualification and some physiological characteristics of an expert while the suggested method gives the possibility to decrease significantly the costs due to automatization. The mobile analysis platform performing this method is based on the fluorescence microscopy. The problem of the object classification as either “bacterium” or “third-party artifact” was solved for the test data with some classification algorithms as support vector machine, random forest, decision tree C4.5,  $k$ -nearest neighbors, Bayes method. The analysis showed that the most effective algorithms are support vector machine and random forest. This research is performed on the Mechatronics Department of Saint Petersburg National Research University of Information Technologies, Mechanics and Optics and the Quality Assurance and Industrial Image Processing Department of Ilmenau University of Technology with the support of the program “Mikhail Lomonosov” of the Ministry of Education and Science of Russia and the German Academic Exchange Service.

**Keywords:** machine learning, bacteria detection.

### Введение

В последние годы методы машинного обучения все чаще применяются для решения большого количества задач в различных отраслях науки и техники [1–7]. С другой стороны, одной из важнейших проблем современного мира является обеспечение контроля качества продуктов питания, что требует решения комплексной задачи, включающей в себя, в том числе, мероприятия по обнаружению бактерий в жидких образцах мясного сока. Оценка риска заражения мяса заключается в определении вида и количества бактерий в пробе. Существующие в настоящее время методы анализа требуют специального лабораторного оборудования, значительного времени и сильно зависят от квалификации и некоторых физиологических особенностей эксперта, обеспечивающего визуальный контроль микроскопических изображений исследуемых проб. В частности, международный стандарт [8] определяет процедуру контроля качества продуктов, состоящую из этапов, показанных на рис. 1. Выполнение всех этапов осуществляется в течение 3–5 дней.

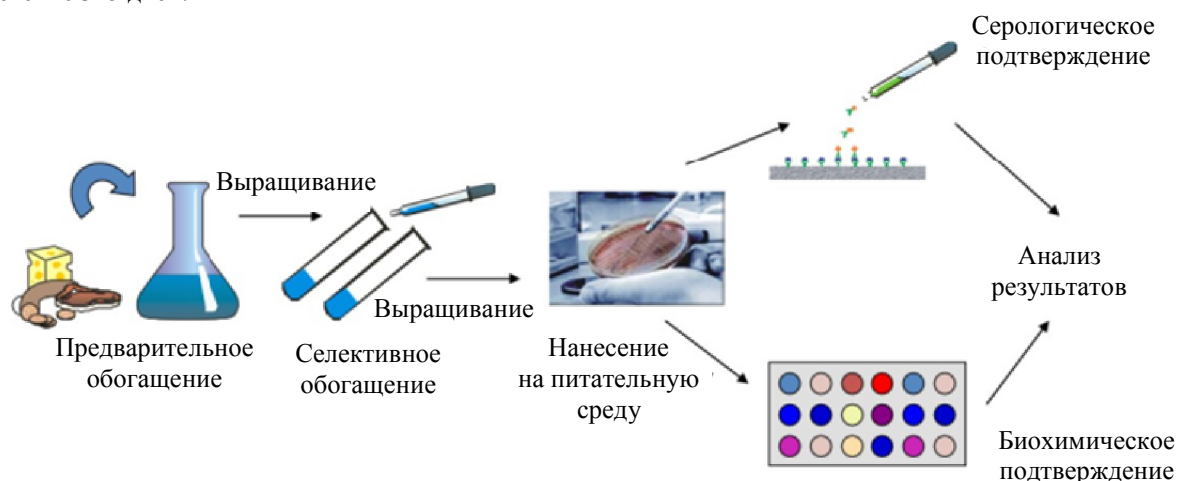


Рис. 1. Процедура контроля качества мясных продуктов питания, определенная стандартом [8] (источник: презентация о проекте LabChip Ю на ежегодном собрании CoOPTICS)

Указанные недостатки являются основной предпосылкой для создания нового более совершенного способа контроля качества продуктов питания, способного обеспечивать своевременное предоставление результатов до момента доставки продуктов потребителям, а также до момента порчи продуктов. В каче-

стве варианта решения данной проблемы предлагается использование мобильной платформы для анализа продуктов питания (мясного сока) на наличие бактерий, описанной в работе [9] и изображенной на рис. 2.

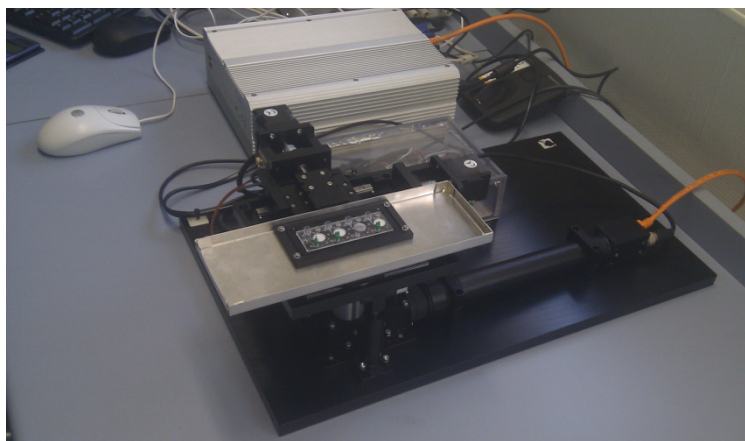
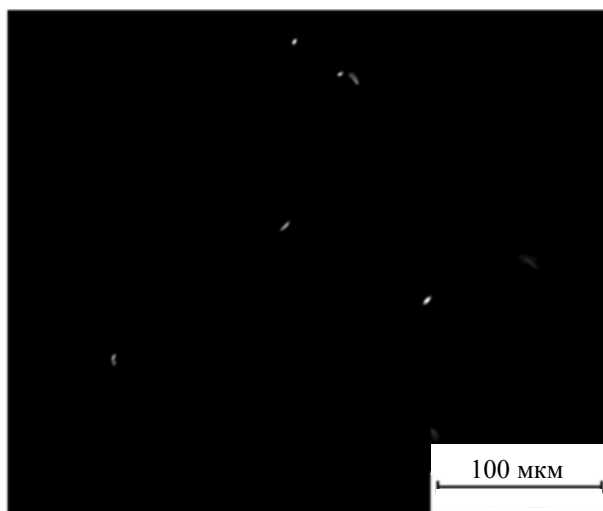


Рис. 2. Мобильная аналитическая платформа



а

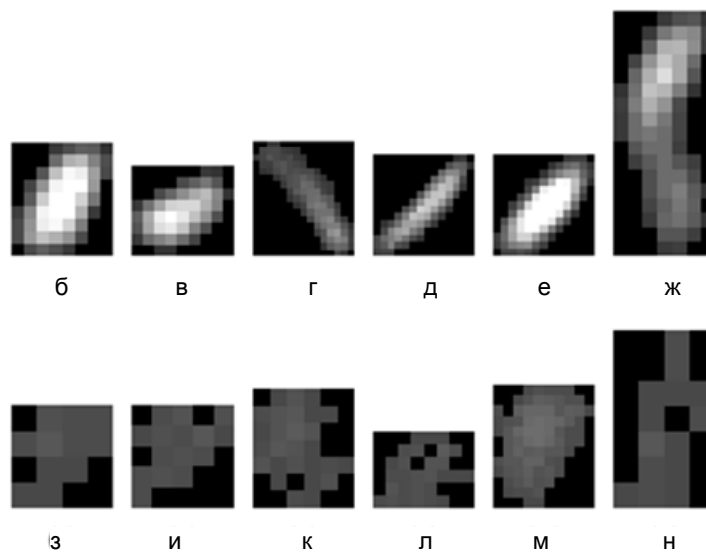


Рис. 3. Пример исходного изображения (а) и выделенных на нем объектов: бактерий (б–ж) и сторонних артефактов (з–н)

Все изложенное дает основание утверждать, что применение методов обработки изображений и машинного обучения с целью выявления бактерий в продуктах питания является актуальной задачей, решение которой позволит значительно ускорить и упростить процесс контроля. В соответствии с этой целью настоящей работы является решение поставленной задачи путем классификации объектов, а также сравнительный анализ эффективности различных алгоритмов.

### Принцип действия мобильной аналитической платформы

Принцип действия рассматриваемой оптической измерительной системы основан на флуоресцентной микроскопии. Измеряемым объектом является мембрана с биологическими образцами, представляющая собой дискообразную полость диаметром около 10 мм и толщиной около 2 мм. Сама мембрана при этом располагается на подвижной платформе, которая осуществляет прецизионные перемещения по горизонтали и вертикали. Источник света, облучая мембрану светом определенной длины волны, возбуждает флуоресценцию, изображения которой записываются камерой. Испытанные образцы содержали бактерии вида *Legionella pneumophila* – палочковидного возбудителя легионеллеза длиной около 2 мкм и шириной 0,3–0,9 мкм. Каждое полученное изображение покрывает площадь  $0,84 \times 7,07 \text{ мм}^2$ , что существенно меньше общей области образца, поэтому для каждой мембраны осуществляется создание набора изображений для каждого положения по горизонтали и вертикали. Возможные способы оптимизации маршрута обхода образца оптической измерительной системой описаны в работах [9, 10].

Эмпирически установлено, что разница между информативностью цветных и монохромных изображений в рамках данной задачи незначительна, поэтому предпочтение отдано монохромным как менее объемным и более простым в обработке.

### Обнаружение бактерий

Задача обнаружения бактерий сводится к задаче классификации на два непересекающихся класса («бактерия» и «сторонний артефакт») и, по сути, является частным случаем задачи машинного обучения, которая в общем виде заключается в необходимости при конечном множестве классов  $Y = 1, 2, \dots, l$  построить алгоритм, который по объекту  $x$  определяет точное или достаточно точное значение  $y(x)$  [3]. В качестве исходных данных принимается пространство допустимых объектов  $X$ , пространство меток  $Y$ , а также целевая функция  $y(x)$ , заданная в конечном множестве точек обучающей выборки  $y(x_1), y(x_2), \dots, y(x_m)$ . Обучающая выборка представляет собой матрицу с описанием объектов  $\mathbf{X}$  и вектор меток  $\mathbf{Y}$ :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},$$

где  $m$  – количество объектов, а  $n$  – количество признаков этих объектов. Таким образом, каждый ряд матрицы  $\mathbf{X}$  соответствует одному объекту  $x_i$ , представленному в виде вектора признаков, а каждый элемент  $y_i \in \{0, 1\}$  определяет класс  $i$ -го объекта.

Функция потерь  $L(A(x), y(x))$  показывает, насколько ответ  $A(x)$  соответствует верному ответу  $y(x)$  и определяется как

$$L(A(x), y(x)) = \begin{cases} 1, & A(x) \neq y(x) \\ 0, & A(x) = y(x) \end{cases}.$$

Соответственно, алгоритмы машинного обучения должны решать задачу оптимизации в виде

$$\frac{1}{m} \sum_{i=1}^m L(A(x), y(x)) \rightarrow \min.$$

Во время подготовки обучающей выборки для классификации объектов после получения снимков они были обработаны с целью сегментации объектов, которые затем классифицировались экспертами – сотрудниками микробиологической лаборатории с помощью программной системы бинаризации и сегментации изображений [11]. В результате была получена база, состоящая из более 200 изображений (60 бактерий и более 140 сторонних артефактов), которая впоследствии сократилась до 120 изображений (по 60 представителей каждого класса, рис. 3). Затем для каждого изображения извлекались 30 признаков, таких как длина и диаметр области, моменты области, большая и малая полуоси эллипса, описанного вокруг области, округлость и др. [12–15]. Гистограммы всех признаков (рис. 4) позволили выбрать релевантные, на основе которых осуществлялась классификация.

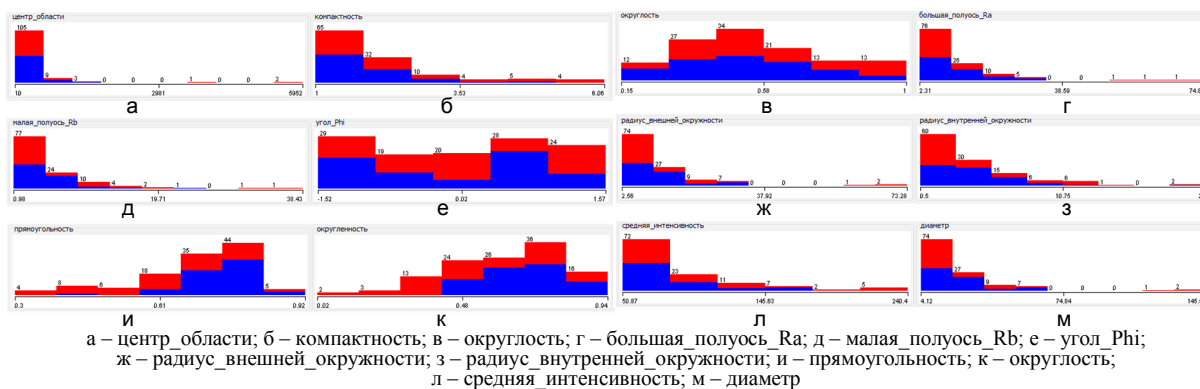


Рис. 4. Гистограммы признаков (красным цветом обозначены значения для бактерий, синим – для сторонних артефактов)

### Оценка эффективности классификаторов

Обычно оценка эффективности алгоритмов классификации производится экспериментально, из-за существенной неформальности большинства задач распознавания, и выражается в способности выбранных алгоритмов принимать верные решения, что характеризуется уровнем ошибок первого («ложный пропуск» – событие ложно не обнаруживается) и второго рода («ложное обнаружение» – событие ошибочно считается произошедшим).

Если количество объектов для каждого класса в тестовом наборе  $N = Np + Nn$ , где  $N$  – общее количество объектов,  $Np$  – количество бактерий и  $Nn$  – количество сторонних артефактов, а количество ложных пропусков и ложных обнаружений равно  $FN$  и  $FP$  соответственно, то количество верных пропусков и верных обнаружений определяется как

$$TP = Np - FN,$$

$$TN = Nn - FP.$$

При этом уровни ошибок выражаются следующим образом:

$$nFN = \frac{FN}{Np} \cdot 100\%,$$

$$nFP = \frac{FP}{Nn} \cdot 100\%,$$

$$nTN = \frac{TN}{Nn} \cdot 100\%,$$

$$nTP = \frac{TP}{Np} \cdot 100\%.$$

Одним из способов оценки эффективности алгоритмов классификации в заданных условиях является мера расстояния до точки  $(0, 1)$  на ROC-диаграмме (Receiver Operating Characteristic – операционная характеристика приемника), которая вычисляется как  $E = \sqrt{FP_{rate}^2 + (1 - TP_{rate})^2}$  [1, 5, 6]. При этом минимальное возможное значение 0 соответствует наилучшей эффективности ( $FP_{rate} = 0$ , а  $TP_{rate} = 1$ , т.е. все бактерии верно определены как экземпляры класса «бактерия» и ни один сторонний артефакт не определен как экземпляр класса «бактерия»). Максимальное значение  $\sqrt{2}$  отображает наихудшую эффективность при  $FP_{rate} = 1$  и  $TP_{rate} = 0$ . Таким образом, для процентного выражения эффективности классификатора, принимая за 100% максимальную эффективность, получаем  $E_{\%} = \left(1 - \frac{E}{\sqrt{2}}\right) \cdot 100\%$ .

Проведенный анализ эффективности в работе [16] показал, что наилучшей эффективностью, около 82%, обладают метод опорных векторов и случайный лес. Для сравнения: деревья решений C4.5 и метод  $k$  ближайших соседей показали эффективность около 72%, а метод Байеса – 38%.

### Заключение

В работе продемонстрирована эффективность решения проблемы обнаружения бактерий для контроля качества продуктов питания методами обработки изображений и машинного обучения, описан принцип действия мобильной аналитической платформы для мгновенного анализа образцов, предложены

пути решения проблемы путем классификации объектов. Кроме этого, приведен сравнительный анализ эффективности различных классификаторов.

### Литература

1. Bigus J.P. Data Mining with Neural Networks. McGraw-Hill, 1996. P. 220.
2. The handbook of data mining / Ed. N.Ye. Lawrence Erlbaum Associates, 2003. 689 p.
3. Дьяконов А.Г. Анализ данных, обучение по прецедентам, логические игры, системы WEKA, RapidMiner и MatLab (Практикум на ЭВМ кафедры математических методов прогнозирования): Учеб. пособие. М.: Издательский отдел факультета ВМК МГУ им. М.В. Ломоносова, 2010. 278 с.
4. Anding K. Automatisierte Qualitätssicherung von Getreide mit überwachten Lernverfahren in der Bildverarbeitung: Dissertation zur Erlangung der akademischen Grades Doktoringenieur (Dr.-Ing.). Ilmenau, Germany: Technische Universität Ilmenau, 2010. 235 p.
5. Witten I.H., Frank E., Hall M.A. Data Mining: Practical Machine Learning Tools and Techniques. 3<sup>rd</sup> ed. Morgan Kaufmann, 2011. 629 p.
6. Bramer M. Principles of data mining. 2<sup>nd</sup> ed. Springer, 2013. 440 p.
7. Латышев С.М., Воронин А.А., Андинг К., Линц Э., Курицын П.А. Оптико-электронные методы и средства идентификации веществ и материалов // Изв. вузов. Приборостроение. 2013. Т. 56. № 10. С. 81–87.
8. ISO 6579:2002. Microbiology of food and animal feeding stuffs – Horizontal method for the detection of Salmonella spp. 08.08.2002. 32 p.
9. Lerm S., Holder S., Gopfert A., Futterer R., Linss G. Concepts of a scanning hardware platform for high-resolution image processing with Lab-on-a-chip analysis // Proc. of the 15<sup>th</sup> International Symposium «MECHATRONIKA». Prague, 2012. P. 1–4.
10. Lerm S. Objektsegmentierung von kompakten Schüttgut für die technische Erkennung: Dissertation zur Erlangung der akademischen Grades Doktoringenieur (Dr.-Ing.). Ilmenau, Germany: Technische Universität Ilmenau, 2012. 235 p.
11. Саенко А.П. Программная система бинаризации и сегментации изображений: Свидетельство о государственной регистрации программы для ЭВМ 2013613134. Заявл. 2013.
12. Shapiro L., Stockman G. Computer Vision. Prentice Hall PTR, 2001. 580 p.
13. Gonzalez R.C., Woods R.E. Digital Image Processing. 2<sup>nd</sup> ed. Prentice Hall, 2002. 793 p.
14. Erhardt A. Einführung in die Digitale Bildverarbeitung. Vieweg+Teubner Verlag, 2008. 248 p.
15. HALCON Version 11.0.1 – HALCON / HDevelop Reference Manual. MVTec Software GmbH, 2012. 2352 p.
16. Саенко А.П. Оценка эффективности обнаружения бактерий методами обработки цифровых изображений и интеллектуального анализа данных // Сборник научных трудов Одиннадцатой сессии международной научной школы «Фундаментальные и прикладные проблемы надежности и диагностики машин и механизмов». СПб: Институт проблем машиноведения РАН, 2013. С. 318–321.

<b>Саенко Алексей Петрович</b>	– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, alexey.saenko@gmail.com
<b>Мусалимов Виктор Михайлович</b>	– доктор технических наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, musvm@yandex.ru
<b>Лерм Штеффен</b>	– аспирант, Технический университет Ильменау, Ильменау, Германия, steffen.lerm@tu-ilmenau.de
<b>Линц Герхард</b>	– доктор технических наук, профессор, зав. кафедрой, Технический университет Ильменау, Ильменау, Германия, gerhard.linss@tu-ilmenau.de
<b>Alexey Saenko</b>	– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, alexey.saenko@gmail.com
<b>Victor Musalimov</b>	– D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, musvm@yandex.ru
<b>Steffen Lerm</b>	– postgraduate, Ilmenau University of Technology, Ilmenau, Germany, steffen.lerm@tu-ilmenau.de
<b>Gerhard Linss</b>	– D.Sc., Professor, Department head, Ilmenau University of Technology, Ilmenau, Germany, gerhard.linss@tu-ilmenau.de