

В работе представлен алгоритм для определения статистических характеристик ансамбля частиц, образующих фотонные кристаллы либо диэлектрические метаматериалы. Следует отметить, что исходное соотношение алгоритма  $F(x, y, a_1, a_2, \dots, a_n) = 0$ , т.е. функция, задающая на плоскости  $(x, y)$  семейство кривых с параметрами  $a_1, a_2, \dots, a_n$ , совсем не обязательно должна сводиться к уравнению окружности  $(x - x_0)^2 + (y - y_0)^2 - R_0^2 = 0$ . Алгоритм позволяет распознавать объекты любой формы, однако количество параметров и, следовательно, время вычислений и возможные погрешности вычислений существенно возрастают. Мы подробно рассмотрели случай окружности в качестве примера для демонстрации метода обработки изображений широко распространенных структур, состоящих из цилиндров или сфер. Такие объекты имеют границы с радиальной симметрией, и задача  $F(x, y, x_0, y_0, R_0) = 0$  сводится к поиску массива трех параметров  $(x_0, y_0, R_0)$ . Полученные результаты, представленные на рис. 5, демонстрируют высокую эффективность данного алгоритма.

Авторы благодарят П.А. Белова и Ю.С. Кившаря за обсуждение результатов работы. Работа выполнена в НИУ ИТМО при финансовой поддержке Министерства образования и науки Российской Федерации (соглашение № 14.В37.21.1964) и РФФИ (проект № 11-02-00865).

#### Литература

1. Inoue K. and Ohtaka K. Photonic Crystals: Physics, Fabrication and Applications. – Springer, 2004. – 332 p.
2. Limonov M.F. and De La Rue R.M. Optical properties of photonic structures: interplay of order and disorder. – CRC Press, Taylor & Francis Group 2012. – 514 p.
3. O'Brien S. and Pendry J.B. Photonic band-gap effects and magnetic activity in dielectric composites // J. Phys.: Cond. Matt. – 2002. – V. 14. – P. 4035.
4. Hosseinzadeh A. and Semouchkina E. Effect of permittivity on energy band diagrams of dielectric metamaterial arrays // MOTL. – 2013. – V. 55. – P. 134–137.
5. Rybin M.V., Sinev I.S., Samusev A.K., Samusev K.B., Trofimova E.Yu., Kurdyukov D.A., Golubev V.G. and Limonov M.F. Dimensionality effects on the optical diffraction from opal-based photonic structures // Phys. Rev. B. – 2013. – V. 87. – P. 125131.
6. Дырнаев А.В., Потапов А.С. Комбинированный метод подсчета эритроцитов на изображениях мазков крови // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 19–23.
7. Palacios-Lidón E., Juárez B.H., Castillo-Martínez E. and López C. Optical and morphological study of disorder in opals // J. Appl. Phys. – 2005. – V. 97. – P. 63502.
8. Rengarajan R., Mittleman D., Rich C. and Colvin V. Effect of disorder on the optical properties of colloidal crystals // Phys. Rev. E. – 2005. – V. 71. – P. 16615.
9. Gonzalez R.C. Woods R.E. Digital Image Processing. – Prentice Hall, 2002. – Chapter 10. – 793 p.
10. Sobel I. An isotropic image gradient operator. Machine Vision for Three-Dimensional Scenes. – N.Y.: Academic Press. – 1990. – P. 376–379.
11. Canny J. A computational approach to edge detection // IEEE Transactions on pattern analysis and machine intelligence. – 1986. – V. 8. – P. 679.

**Самусев Кирилл Борисович** – Россия, Санкт-Петербург, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, ст. научный сотрудник, k.samusev@phoi.ifmo.ru

**Рыбин Михаил Валерьевич** – Россия, Санкт-Петербург, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, ст. научный сотрудник, m.rybin@phoi.ifmo.ru

**Лимонов Михаил Феликсович** – Россия, Санкт-Петербург, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, ведущий научный сотрудник, m.limonov@phoi.ifmo.ru

УДК 004.932

## МЕТОДИКА СРАВНЕНИЯ АЛГОРИТМОВ СТЕРЕОЗРЕНИЯ ПРИ ВОССТАНОВЛЕНИИ ТРЕХМЕРНОЙ МОДЕЛИ ЛИЦА ЧЕЛОВЕКА

С.В. Пономарев

Представлены результаты сравнительного анализа эффективности алгоритмов стереозрения в рамках решения задачи восстановления трехмерной модели лица человека. Рассмотрены методы стереозрения, относящиеся к трем основным классам – локальные, глобальные и полуглобальные методы. В качестве критериев количественной оценки работы алгоритмов предложена мера расхождения восстановленной карты дальности и эталонной карты дальности, полученной с использованием камеры со структурированной инфракрасной подсветкой. При сравнении осуществлялся переход в трехмерное пространство, проводилось вычисление характеристик облака точек, учитывалось каче-

ство трехмерной поверхности, восстановленной с помощью карты дальности, и данные о производительности алгоритмов. На основе количественной и качественной оценки работы алгоритмов произведен выбор наилучшего алгоритма. Результаты работы могут быть использованы для повышения эффективности применения методов стереозрения в целях идентификации личности человека.

**Ключевые слова:** карта дальности, алгоритмы стереозрения, трехмерная модель лица, биометрические системы.

### Введение

В настоящее время актуальна разработка биометрических систем, позволяющих обеспечить автоматическую идентификацию личности человека с высокой степенью надежности. Одним из перспективных направлений в данной области является восстановление трехмерной информации о форме лица. Использование трехмерной модели лица человека в качестве дополнительного идентификатора личности позволяет повысить надежность и качество распознавания, ускорить процесс идентификации.

Для решения этой задачи применяются различные программные и аппаратные средства, в частности, одним из возможных вариантов является использование трехмерных сканеров, которые, обладая высокой точностью, имеют ряд ограничений в применении к живым объектам. Ограничения связаны с требованием статичности сканируемого объекта в течение всего времени сканирования, чего трудно достичь в случае человека, а также с тем, что сканирование является активным методом исследования, связанным с освещением анализируемого объекта, что накладывает значительные ограничения на характер излучения с точки зрения безопасности здоровья человека. Помимо 3D-сканеров, для восстановления трехмерной формы объектов может применяться также метод восстановления глубины из фокусировки [1], но он не подходит для динамичных объектов.

В то же время использование пассивных методов, работающих в условиях естественного освещения, потенциально имеет более широкое применение. В этой связи перспективны методы стереоскопического зрения [2]. Несмотря на то, что было разработано большое число алгоритмов, позволяющих решить проблему стереозрения, лишь небольшое число работ посвящено сравнению существующих алгоритмов и оценке их производительности. В [3] представлена одна из наиболее полных классификаций алгоритмов стереозрения, однако оценка характеристик алгоритмов проводилась для наиболее общего случая без учета специфики различных предметных областей.

С другой стороны, известны работы, в которых исследовалось решение проблемы стереозрения непосредственно для восстановления трехмерной модели лица человека [2, 4], но на момент их публикации некоторые алгоритмы, которые сейчас являются распространенными, в частности, полуглобальный алгоритм стереозрения [5], еще не были представлены. Разработанная в настоящей работе методика количественной и качественной оценки характеристик алгоритмов стереозрения позволяет выявить наиболее эффективные алгоритмы для данной предметной области при рассмотрении широко используемых в настоящее время алгоритмов.

### Классификация алгоритмов стереозрения

Существуют различные варианты классификации алгоритмов сопоставления двух изображений для создания системы стереозрения. Один из наиболее распространенных вариантов состоит в разделении на локальные и глобальные алгоритмы [3]. Для локальных алгоритмов диспаратность вычисляется для каждой точки изображения и зависит только от сравнения локальных областей – окон вокруг каждой точки и соответствующей точки на сопоставляемом изображении. В глобальных алгоритмах диспаратность находится для всех пикселей сопоставляемых изображений, при этом вычисляется минимум функционала энергии

$$E(f) = E_d(f) + \lambda E_s(f), \quad (1)$$

где  $E_d(f)$  – унарный потенциал, который характеризует степень согласованности данной конфигурации с парой входных изображений и вычисляется с помощью пиксельных стоимостей;  $E_s(f)$  – парный потенциал, который явным образом реализует предположение о гладкости, т.е. накапливает величину штрафа, налагаемого на данную конфигурацию при нарушении непрерывности диспаратностей.

Реализацию большинства алгоритмов, относящихся как к локальному, так и к глобальному подходу, можно условно представить в виде четырех этапов:

- подсчет стоимости соответствий;
- суммирование найденных стоимостей;
- вычисление и оптимизация диспаратностей;
- уточнение диспаратностей.

В локальных алгоритмах для того, чтобы определить совпадение точек на двух изображениях, используется подсчет стоимости соответствия для каждого пикселя в строке; самым вероятным кандидатом является точка с наименьшей стоимостью.

Существует несколько различных подходов, основанных на вычислении стоимости значений интенсивности пикселей. Можно выделить такие способы, как нахождение абсолютной разности интенсивностей (Absolute Intensity Differences, AD) и квадрата разности интенсивностей (Squared Intensity

Differences, SD). Абстрагируясь от функции стоимости, можно записать выражение для вычисления стоимости соответствия точек в виде

$$d_p = \arg \min \sum_{q \in W_p} c(q, q-d), \quad 0 \leq d \leq d_{\max}$$

где  $d_p$  – искомая диспаратность в пикселе  $p$ ;  $c(p, q)$  – функция стоимости перехода от пикселя  $p$  к пикселю  $q$ ;  $W_p$  – окно вокруг пикселя  $p$ ;  $d_{\max}$  – максимальное значение диспаратности.

Для квадрата разностей интенсивностей это выражение примет вид

$$SDC(p, d) = \sum_{p \in W} |I_{bp} - I_{mq}|^2, \quad q = p - d.$$

Для абсолютной разности интенсивностей выражение будет следующим:

$$ABC(p, d) = \sum_{p \in W} |I_{bp} - I_{mq}|, \quad q = p - d.$$

Эффективность обоих подходов весьма высока, но из-за предположения о совпадении интенсивностей у соответствующих точек стереопары сильное влияние на конечный результат оказывают особенности съемки (экспозиция, случайные блики), так и шумы.

Целью глобального подхода является нахождение наилучшей карты диспаратности для всего изображения. Типичным способом решения проблемы соответствия для глобальных алгоритмов является поиск минимума функционала энергии.

В связи с этим существует также разделение алгоритмов данной группы по способу минимизации энергии. Чаще всего используется либо динамическое программирование, либо нахождение минимального разреза графов (graph cuts). Алгоритмы разреза графов также получили название двумерных. Их производительность ниже, но результаты имеют более высокую точность. В ходе использования одномерных алгоритмов строки изображений обрабатываются независимо друг от друга. Вследствие этого их скорость выше, но возникает эффект «гребенки» – рассогласование значений диспаратности на уровне отдельных строк.

Рассмотрим поиск наилучшей конфигурации  $f$  для минимизации функционала глобальной энергии  $E$ , представленного в уравнении (1). Первое слагаемое, унарный потенциал, характеризует степень согласованности данной конфигурации с парой входных изображений и подсчитывается с помощью пиксельных стоимостей:

$$E_d(f) = \sum C(x, y, f(x, y)) = \sum_{p \in I} m(p, p - d_p).$$

Здесь  $C(x, y, f(x, y))$  – функция пиксельной стоимости в общем виде;  $m(p, q)$  – мера цветового сходства между пикселем  $p$  одного изображения и пикселем  $q$  другого изображения.

Второе слагаемое, парный потенциал, явным образом реализует предположение о гладкости, т.е. накапливает величину штрафа, налагаемого на данную конфигурацию при нарушении непрерывности диспаратностей. Обычно при вычислении парного потенциала учитываются соседние пиксели:

$$E_s(f) = \sum_{\langle p, q \rangle \in N} s(d_p, d_q),$$

где  $N$  – множество всех пар соседних пикселей;  $s(d_p, d_q)$  – функция штрафа за нарушение гладкости.

Существует ряд решений для нахождения минимума функционала энергии, среди которых известны два наиболее распространенных подхода – динамическое программирование и использование случайных полей Маркова (нахождение минимального разреза графа, алгоритм распространения доверия).

Алгоритм полуглобального стереозрения Semi-Global Stereo Matching (SGSM) представляет собой совмещение принципов локального и глобального подходов. С одной стороны, поиск глобального минимума энергии является NP-полной задачей и не может быть осуществлен за приемлемое время. С другой стороны, можно выполнить очень быстрый поиск в одном измерении (построчно), в частности, с использованием динамического программирования. Однако возникает проблема рассогласованности отдельных строк, связанная с полной проверкой одного направления (построчного) и отсутствием проверок в остальных (например, по столбцам). Для устранения недостатков локального и глобального подходов был предложен способ агрегации стоимости соответствия точек – одновременное суммирование стоимости путей по всем направлениям [5].

### Сравнительный анализ эффективности алгоритмов стереозрения

Было проведено сравнение рассмотренных выше алгоритмов с целью выявления наиболее подходящего для реконструкции карт дальности для поверхности лиц. Рассматривались следующие методы:

- корреляционный метод (локальный алгоритм с фиксированным размером окна) – CORR;
- глобальный алгоритм, основанный на динамическом программировании с применением DSI – DP;
- алгоритм разрезания графов (Graph Cuts) – GC;
- алгоритм распространения доверия (Belief Propagation) – BP;

– Semi-Global Stereo Matching – SGSM.

На рисунке отображены карты дальности, построенные данными алгоритмами для стереопар с изображениями человеческого лица.

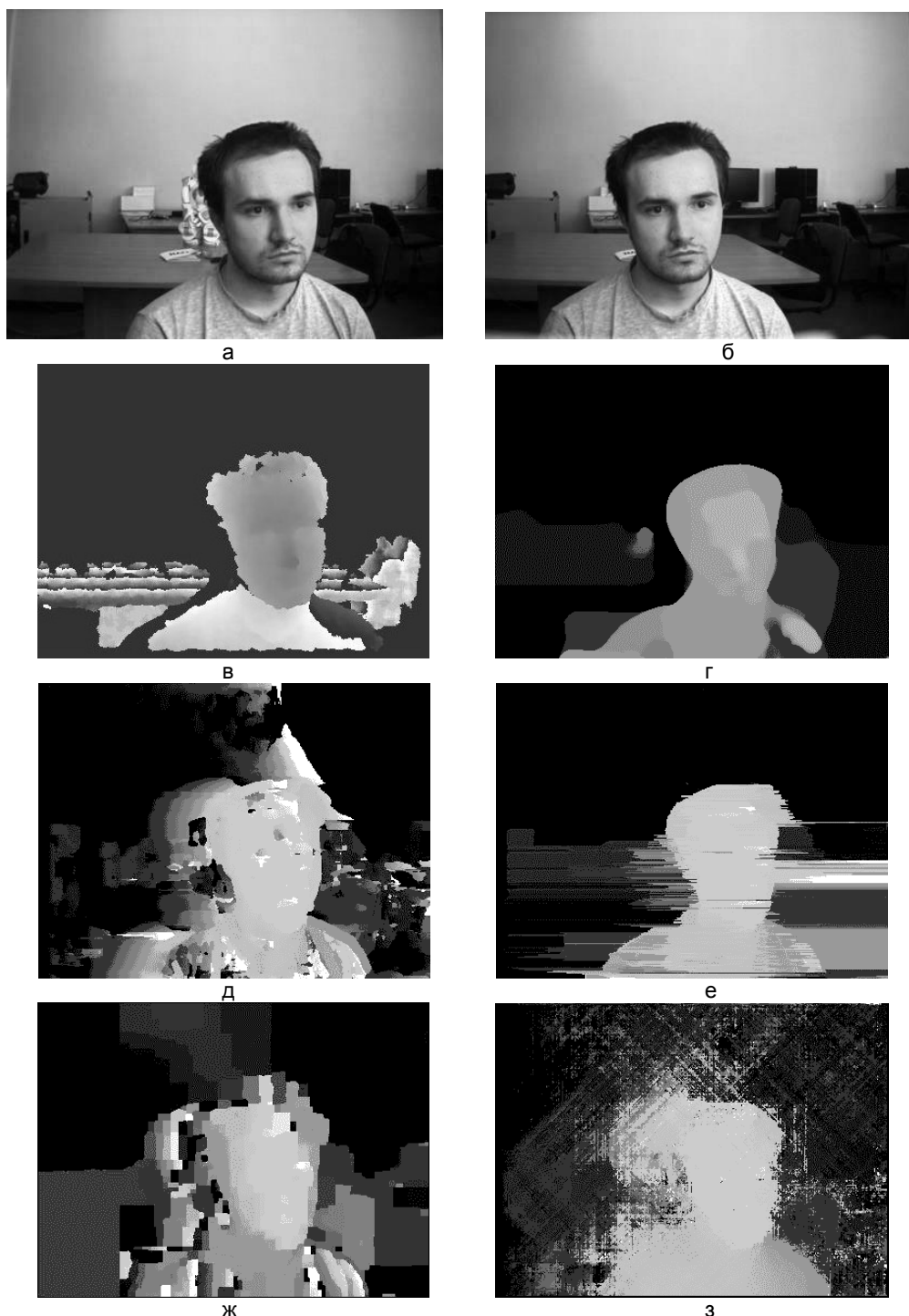


Рисунок. Результаты работы алгоритмов: (а)–(б) – исходная пара изображений; (в) – эталонная карта дальности; (г) – GC; (д) – CORR; (е) – DP; (ж) – BF; (з) – SGSM

Для количественной оценки эффективности представленных алгоритмов в качестве эталонной карты дальности была использована карта дальности, получаемая с камеры Kinect, работающей с использованием активной инфракрасной подсветки. Карты дальности, полученные в результате работы исследуемых алгоритмов стереозрения, сравнивались с эталонной картой следующим образом.

- Производилась локализация области лица на эталонной карте и карте сравнения, вычислялась средняя квадратичная ошибка непосредственно для двух изображений.
- Осуществлялся переход от двумерного к трехмерному представлению, т.е. карты дальности преобразовывались в облака точек с использованием внутренних параметров стереокамеры и камеры Kinect.

Облака точек выравнивались друг относительно друга, и вычислялась средняя квадратичная ошибка в трехмерном пространстве.

Для каждого алгоритма оценивалось время выполнения. Сводные результаты для группы исследуемых алгоритмов представлены в таблице.

Метод стереозрения	Среднее время работы, с	Средняя квадратичная ошибка, отн. ед.	MD, отн. ед.	DEV, отн. ед.
CORR	>0,1	34,63	0,23	0,21
SGSM	2,6	33,06	0,17	0,15
DP	4,6	33,32	0,18	0,14
BP	33,1	29,29	0,15	0,19
GC	550,0	26,63	0,14	0,14

Таблица. Количественная оценка эффективности алгоритмов  
(MD – среднее расстояние между облаком точек, полученным для данного алгоритма и облаком точек для эталонной карты дальности; DEV – средняя квадратичная ошибка для среднего расстояния)

Опираясь на полученные результаты, можно провести качественную оценку работы алгоритмов. Учитывалось также и время выполнения методов. Самым быстрым, как и ожидалось, оказался локальный алгоритм (время порядка десятых долей секунды), но искажения и размытие контуров объектов не предполагают его использование для реконструкции объектов с большим числом мелких деталей. Глобальные методы позволяют восстановить карту диспаратности с обеспечением большей устойчивости, однако у алгоритма, основанного на динамическом программировании, наблюдается эффект «гребенки», устранение которого требует дополнительных мер постобработки в рамках последующего моделирования. Алгоритм Graph Cuts выполняется за большое время (порядка 10 мин), что делает невозможным его применение в реальных условиях (кроме того, получаемая карта недостаточно детальна по глубине). У алгоритма Belief Propagation наблюдаются ярко выраженные дефекты в фоновой области, по краям изображения резко снижается плавность переходов. Средняя квадратичная ошибка как в 2,5D, так и в трехмерном представлении, имеет меньшее значение для глобальных алгоритмов; для Belief Propagation характерно более высокое значение средней квадратичной ошибки для среднего расстояния между облаками точек, что подтверждает качественные оценки.

Метод SGSM, во-первых, оказался сравнимым по быстродействию с локальными алгоритмами, во-вторых, по сравнению с Graph cuts и Belief Propagation обеспечивает большую плавность для различных значений диспаратности. Его количественные характеристики сопоставимы с алгоритмом на основе динамического программирования. Рассогласованности наблюдаются только на уровне отдельных точек, тем самым упрощается процесс постобработки. В связи с этим эффективность данного метода в рамках решения задачи восстановления трехмерной модели лица выше.

В результате анализа быстродействия, качественной и количественной оценки полученных карт дальностей наилучшим был признан метод SGSM.

### Заключение

В ходе работы был осуществлен сравнительный анализ методов стереозрения, относящихся к трем основным классам: локальные методы, глобальные методы и полуглобальные методы. Количественные и качественные оценки показывают, что наиболее предпочтительным для работы в исследуемой предметной области является алгоритм полуглобального стереозрения в силу возможности восстановления с его помощью плавных переходов дальности по изображениям лиц в масштабе почти реального времени.

Следует учитывать, что все рассмотренные алгоритмы стереозрения являются алгоритмами общего назначения, и результирующие карты дальности не являются оптимальными, так как содержат различные дефекты и не учитывают априорно известную предметно-специфическую информацию. В связи с этим представляется необходимой разработка методик, которые позволят на этапах постобработки и работы самого алгоритма повысить качество карт дальностей, используя ключевые особенности лица человека.

Работа выполнена при поддержке Министерства образования и науки Российской Федерации, грант МД-2040.2010.9.

### Литература

1. Аверкин А.Н., Потапов А.С., Рожков А.С. Формирование и визуализация 3D-изображений микрообъектов по серии видеокадров с изменяемой фокусировкой // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 6 (76). – С. 12–16.
2. Leclercq P., Liu J., Woodward A., Delmas P. Which Stereo Matching Algorithm for Accurate 3D Face Creation? // IWCIA. – 2004. – P. 690–704.

3. Scharstein D., Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // *Int. Journal of Computer Vision*. – 2002. – V. 47. – P. 7–42.
4. Chan M., Chen C.Y., Barton G., Delmas P., Gimelfarb G., Leclercq P., Fisher T.A. Strategy for 3D Face Analysis and Synthesis // *Image and Vision Computing New Zealand Conference*. – 2003. – P. 384–389.
5. Hirschmuller H. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2008. – V. 30. – № 2. – P. 328–341

*Пономарев Святослав Владимирович* – Россия, Санкт-Петербург, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, инженер, sv.v.ponomarev@gmail.com

УДК 004.932.2

## **АНАЛИЗ ОСОБЕННОСТЕЙ ИЕРАРХИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ В СОВРЕМЕННЫХ СИСТЕМАХ КОМПЬЮТЕРНОГО ЗРЕНИЯ**

**В.Н. Васильев, И.П. Гуров, А.С. Потапов**

Рассмотрены широко применяющиеся в прикладных задачах методы компьютерного зрения. Установлено, что эти методы осуществляют анализ изображений преимущественно на одном уровне – пиксельном, контурном, структурном и т.д. Определены недостатки одноуровневых представлений изображений, ограничивающие инвариантные свойства использующих эти представления методов, что, в частности, выражается при решении задачи распознавания в снижении возможности по различению объектов разных классов. Обоснована возможность преодоления этих ограничений с помощью иерархических представлений. Предложены пути развития теории синтеза оптимальных иерархических систем анализа изображений, в которых снижение вероятности выбора наилучшей гипотезы верхнего уровня из-за принятия гипотез промежуточных уровней минимально. Минимизацию ошибки аппроксимации распределения апостериорных вероятностей гипотез верхних уровней путем учета их зависимости только от лучших гипотез нижних уровней предложено осуществлять с помощью введения обратных связей между уровнями и адаптивного выбора гипотез всех уровней с максимизацией их совместной вероятности.

**Ключевые слова:** иерархические представления, компьютерное зрение, анализ изображений.

### **Введение**

Область компьютерного зрения приобретает все большее прикладное значение. При этом некоторые методы оказываются успешными при решении широкого круга задач, благодаря чему приобретают популярность и начинают использоваться многими разработчиками при создании прикладных систем компьютерного зрения, особенно после реализации этих методов в таких широко используемых при прототипировании программных библиотеках, как OpenCV. В качестве примера можно привести методы выделения ключевых точек на изображениях с последующим построением локальных инвариантных дескрипторов (наиболее известными являются SIFT [1] и SURF [2]), каскады Хаара, а также нейросетевые методы [3], используемые при выделении объектов с устойчивой формой, например, лиц людей, гистограммы ориентированных градиентов [4] (используемые, например, при детектировании пешеходов на изображениях), мешки визуальных слов [5] (популярные в задачах извлечения изображений из баз) и т.д.

Существуют также методы для решения таких задач, как построение плотных карт стереодиспаратности (например, методом отсечения графов или полуглобальным методом стереозрения [6]), и распространены готовые технические решения, наподобие устройства Kinect, весьма популярного у разработчиков систем компьютерного зрения при восстановлении трехмерной формы и распознавании поз и жестов людей.

Известны общепринятые методы обработки, например, метод Кенни для детектирования контуров или преобразование Хафа для построения геометрических элементов (тоже, кстати, реализованные в OpenCV и применяемые до сих пор [7]), но такие методы не давали законченного решения конкретных задач и не имели стандартной реализации. Сейчас для решения многих частных прикладных задач целесообразно воспользоваться имеющимся инструментарием. У специалистов, использующих компьютерное зрение при решении практических задач, вполне может сложиться впечатление, что знания имеющегося инструментария и умения его применять вполне достаточно для работы в этой области.

С учетом того, что появление упомянутых эффективных на практике методов вряд ли можно считать результатом какого-то научного прорыва, возникает естественный вопрос о теоретическом прогрессе в данной области, его направленности и востребованности. Здесь необходимо отметить, что существует множество очень сложных задач, которые не могут быть решены с помощью имеющихся технологий. Хотя такие технические достижения, как беспилотный автомобиль, управляемый, в том числе с использованием методов компьютерного зрения, оказываются возможными, проблема семантической пропасти