

А. С. ПОТАПОВ, И. П. ГУРОВ, А. Н. АВЕРКИН

## АНАЛИЗ МОДЕЛИ КЛЕТОК ЗРИТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ ПРИНЦИПА МИНИМИЗАЦИИ РЕПРЕЗЕНТАЦИОННОЙ ДЛИНЫ ОПИСАНИЯ ИЗОБРАЖЕНИЙ

Представлен теоретико-информационный анализ модели клеток зрительной коры, селективных по отношению к пространственным паттернам. Показано, что методы текстурной сегментации, основанные на этой модели, уменьшают длину описания изображения. Явное использование длины описания в качестве целевой функции позволяет уточнить данную модель.

*Ключевые слова:* зрительная кора, модель нейронов, минимальная длина описания, текстурная сегментация.

**Введение.** Сегментация изображений является одной из фундаментальных проблем компьютерного зрения. В этом направлении проведено большое количество исследований и получены значимые результаты, однако остается ряд важных нерешенных вопросов, среди которых выделяется проблема сегментации сложнотекстурированных изображений, когда текстуры изображения априори не известны.

Модели текстур должны создаваться в соответствии с эмпирическими данными, полученными либо посредством анализа статистик естественных сцен [1, 2], либо на основе нейрофизиологических исследований зрительных систем [3, 4]. Отклики нейронов зрительной коры хорошо согласуются со статистиками естественных сцен [5]. Нейрофизиологические данные позволяют получить дополнительную информацию о том, как выполняется сам процесс сегментации. В связи с этим многие методы текстурной сегментации изображений, например предложенные в работе [6], основаны именно на нейрофизиологических данных.

Между зрительными системами даже близких видов млекопитающих имеются определенные различия, однако существует также и большое количество общих черт: в частности, в представлении пространственной информации, которое практически у всех позвоночных осуществляется с использованием вейвлет-подобных преобразований [5], что свидетельствует об их фундаментальной значимости при обработке визуальной информации и, в особенности, при выполнении ее текстурного анализа.

Детальные нейрофизиологические исследования выявили конкретные типы клеток, участвующих в вейвлет-подобных разложениях. В частности, функционирование простых клеток зрительной коры хорошо моделируется с помощью двумерных функций Габора [7]. Современные биологически мотивированные методы сегментации имитируют наиболее поздно открытые клетки, имеющие рецептивные поля, которые обеспечивают селективность по отношению к пространственно-периодическим паттернам.

Несмотря на то что такой имитационный подход весьма эффективен, он, по мнению авторов, не является достаточным, поскольку нейрофизиологические данные описывают

зрительную систему лишь частично и приближенно, а при моделировании не учитывается все разнообразие известных типов клеток и взаимодействий между ними. В связи с этим актуальной является задача строгого описания моделей текстурной сегментации, имитирующих работу клеток зрительной коры.

В настоящей статье представлен анализ функционирования методов сегментации, использующих функции Габора, и модели селективных по отношению к пространственным паттернам клеток на основе принципа минимизации репрезентационной длины описания. Выбор этого принципа обусловлен возможностью с его помощью объективной оценки качества представлений изображений [8]. При этом задача текстурной сегментации трактуется как задача построения оптимального (имеющего минимальную длину) описания изображения в рамках данного представления. Показано, что если выбор представления производится в соответствии с нейрофизиологическими данными, то функционирование имитационных моделей может быть объяснено как минимизация длины описания изображения.

**Модели клеток, селективных по отношению к периодическим паттернам.** Отклик простых клеток зрительной коры на пространственные стимулы часто описывается с помощью семейства функций Габора [6]:

$$g(x, y | \mathbf{w}) = \tilde{g}_{\theta, \sigma, \lambda, \varphi}(x - x_0, y - y_0), \quad \mathbf{w} = (x_0, y_0, \theta, \sigma, \lambda, \varphi); \quad (1)$$

$$\tilde{g}_{\theta, \sigma, \lambda, \varphi}(x, y) = e^{-(\tilde{x}^2 + \gamma^2 \tilde{y}^2)/2\sigma^2} \cos\left(2\pi \frac{\tilde{x}}{\lambda} + \varphi\right), \quad \tilde{x} = x \cos \theta + y \sin \theta, \quad \tilde{y} = -x \sin \theta + y \cos \theta,$$

где  $x, y$  — координаты клетки в плоскости изображения;  $x_0, y_0$  — координаты центра рецептивного поля на изображении;  $\gamma = 0,5$  — константа, определяющая „вытянутость“ рецептивного поля клетки;  $\theta, \sigma, \lambda, \varphi$  — свободные параметры, определяющие конкретную форму рецептивного поля (его ориентацию, размеры, а также частоту и фазу периодического компонента); иногда используется также фиксированное отношение  $\sigma/\lambda = 0,56$ , которое, как полагается в работе [6], характерно для нейронов зрительной коры.

Отклик  $r_{\theta, \sigma, \lambda, \varphi}(x_0, y_0)$  простой клетки, помещенной в точку  $(x_0, y_0)$  изображения  $f(x, y)$  и обладающей рецептивным полем  $\tilde{g}_{\theta, \sigma, \lambda, \varphi}(x - x_0, y - y_0)$ , определяется сверткой

$$r_{\theta, \sigma, \lambda, \varphi}(x_0, y_0) = (f \otimes \tilde{g}_{\theta, \sigma, \lambda, \varphi})(x_0, y_0).$$

Вид функций Габора (1) обеспечивает селективность простых клеток по отношению к периодическим паттернам, однако для выполнения сегментации необходимо объединить информацию об откликах соседних простых клеток.

Сложные клетки комбинируют отклики двух простых клеток. Отклик сложной клетки моделируется как вычисление габоровской энергии:

$$E_{\theta, \sigma, \lambda}(x_0, y_0) = \sqrt{r_{\theta, \sigma, \lambda, 0}^2(x_0, y_0) + r_{\theta, \sigma, \lambda, -\pi/2}^2(x_0, y_0)}.$$

Дальнейшее усложнение моделей осуществляется путем введения латеральных связей между клетками следующего уровня, который представляет собой рекуррентную нейронную сеть. Нейрон этого уровня имеет связи со всеми нейронами в кольце, внешний и внутренний радиусы ( $r_1$  и  $r_2$ ) которого определяются параметром  $\sigma$ , задающим размер рецептивного поля соответствующей клетки. Характер связи между нейронами с параметрами  $(x_0, y_0, \theta_0, \sigma_0, \lambda_0)$  и  $(x, y, \theta, \sigma, \lambda)$  обусловлен сходством параметров габоровской функции: если значения параметров  $(\theta_0, \sigma_0, \lambda_0)$  и  $(\theta, \sigma, \lambda)$  близки, то связи между соответствующими нейронами возбуждающие, в противном случае — тормозящие. Релаксация такой сети приводит к образованию областей в пространстве  $(x, y)$  с близкими по значению параметрами габоровской функции  $(\theta, \sigma, \lambda)$ , при которых сложные клетки в данной области имеют в среднем большой отклик (клетки с не характерными для области параметрами подавляются). Эти

области соответствуют областям однородных текстур, состоящих из элементов с определенными пространственными характеристиками. Здесь, однако, имеется большой произвол в том, как именно задавать структуру латеральных связей. Выбор этих связей неявно задает представление, в рамках которого описывается изображение. Для строгого подхода к построению представлений изображений следует воспользоваться недавно разработанным принципом минимизации репрезентационной длины их описания, изложенным в работе [8].

**Принцип минимизации длины описания.** Принцип минимизации длины описания (МДО) формулируется следующим образом: моделью  $m^*$ , наилучшим образом описывающей некоторый набор данных  $f$ , является модель, которая позволяет минимизировать сумму:

- длины описания модели  $L(m)$ ;
- длины описания данных в рамках модели  $L(f | m)$ .

В качестве длин описания в теоретических работах выступает алгоритмическая сложность строк символов, соответствующих данным и модели. При этом полагается, что весь объем данных доступен одновременно.

При решении задач анализа изображений требуется их независимая интерпретация. В этих условиях принцип МДО необходимо уточнить как принцип репрезентационной МДО (РМДО), состоящий из двух частей:

1) моделью  $m^*$ , наилучшим образом описывающей некоторый набор данных  $f$  в рамках представления  $S$ , является модель, минимизирующая сумму:

$$L_S(f, m) = L_S(m) + L_S(f | m), \quad (2)$$

где  $L_S$  — длина описания в рамках данного представления;

2) лучшим представлением  $S$  для данной выборки изображений  $F = \{f_1, \dots, f_n\}$  является представление, для которого минимизируется сумма:

- длины представления  $L(S)$ ;
- суммы длин описаний изображений  $\sum_{i=1}^n K(f_i)$ , где  $K(f_i) = \min_m L_S(f_i, m)$ .

Таким образом, выбор оптимального представления изображений — эмпирическая задача, решение которой определяется средней, по некоторой выборке, длиной описания. В этом случае рассматривается вопрос о выборе представления для описания текстур изображений с двух позиций: компоненты представления изображений заимствуются из имитационных моделей, и определяется, как эти компоненты позволяют в среднем уменьшать длину описания изображений. Сам же процесс (имеющий также нейрофизиологические корреляты) текстурного анализа изображения рассматривается как процесс минимизации длины его описания в рамках данного представления.

**Анализ модели текстурной сегментации.** В соответствии с принципом РМДО необходимо найти такое представление, в рамках которого длины описаний изображений в среднем минимизируются. Выбор самих представлений должен базироваться на эмпирических данных. Приведем вначале ранее разработанное упрощенное представление [9], а затем попытаемся его расширить с учетом моделей текстуры.

В рамках этого представления, которое обозначим через  $S_0$ , изображение  $f(x, y)$ , заданное в некоторой области  $G$ , т.е.  $f: G \rightarrow R$ , описывается как совокупность непересекающихся областей:  $G = \bigcup_i G_i$ . Содержимое каждой из областей  $G_i$  (сужение функции  $f$  на область  $G_i$ , т.е.  $f_i = f|_{G_i}$ ), в свою очередь, описывается индивидуальной стохастической моделью. Подобное представление обосновано тем, что наблюдаемые сцены состоят из видимых

поверхностей с собственным распределением коэффициентов отражения. В простейшем случае яркости ( $I$ ) пикселей в области  $G_i$  описываются как статистически независимые и одинаково распределенные (функции распределения различны для разных областей). Тогда длина описания  $L(f_i)$  содержимого каждой области изображения будет равна энтропии  $H(f_i)$  значений яркости ее пикселей, умноженной на их число  $\|G_i\|$ . Величина  $H(f_i)$  является оценкой математического ожидания длины описания значений яркости в области  $G_i$ , представленным кодом Хаффмана. Области, приводящие к минимизации длины описания изображения, действительно соответствуют его разным областям. На рис. 1, а приведен фрагмент исходного изображения, а на рис. 1, б, в — две выделенные на нем области; соответствующие гистограммы распределения яркостей, представленные на рис. 1, г—е (где  $N$  — число пикселей), демонстрируют уменьшение длины описания изображения: энтропия яркостей исходного изображения составляет 4,375, а в областях — 4,297 и 3,032, что соответствует уменьшению длины описания с 59 845 до 51 581 бит.

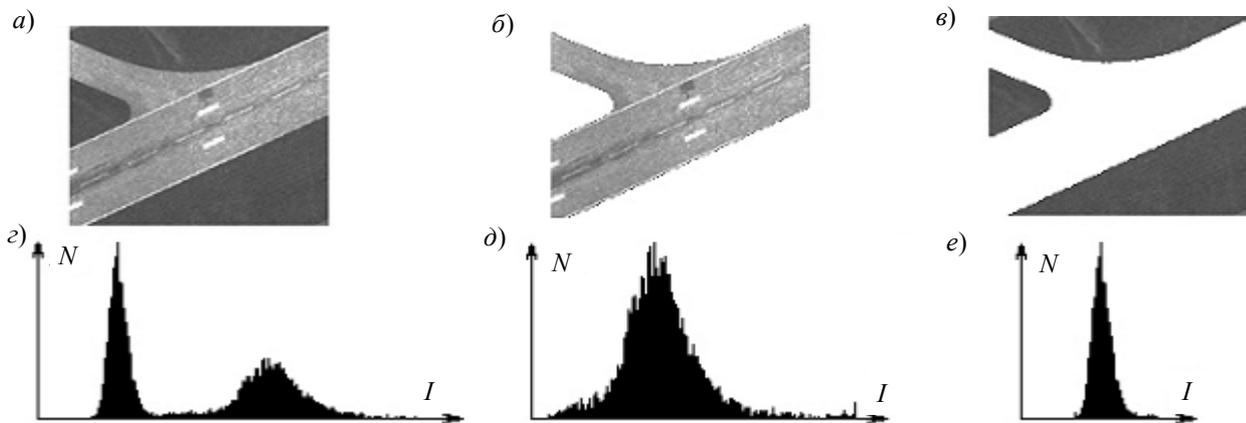


Рис. 1

Помимо яркости пикселей внутри каждой области, также необходимо описать положение этих областей, т.е. их границы, которые могут быть представлены в виде структурных элементов. Однако здесь этот компонент описания не представляет интереса (как и сам алгоритм сегментации, см. [9, 10]), поскольку рассматриваются отдельные области изображения.

Очевидным упрощением данного представления является предположение о статистической независимости яркостей пикселей внутри областей. Расширим данное представление с учетом рассмотренной выше модели клеток, селективных по отношению к периодическим паттернам, обозначив новое представление через  $S_T$ . Предположим, что содержимое  $f_i$  каждой области описывается взвешенной суммой некоторого количества габоровских функций:  $\sum_j A_{i,j} g(x, y | \mathbf{w}_{i,j})$ , где  $\mathbf{w}_{i,j}$  — вектор параметров  $j$ -й функции в  $i$ -й области,  $A_{i,j}$  — вес этой функции. Если векторы параметров известны, то для описания содержимого области достаточно знать значения невязок:

$$\varepsilon_i(x, y) = f_i(x, y) - \sum_{j=1}^{N_i} A_{i,j} g(x, y | \mathbf{w}_{i,j}).$$

В предположении независимости невязок в соответствии с принципом РМДО (2) можно получить следующее выражение:

$$L_{S_T}(f_i, \{\mathbf{w}_{i,j}\}_{j=1}^{N_i}) = H(\varepsilon_i) \|G_i\| + L_{S_T}(\{\mathbf{w}_{i,j}\}_{j=1}^{N_i}),$$

где второе слагаемое — длина описания параметров модели (т.е. совокупности функций Габора).

В предположении, что невязки распределены нормально, поиск оптимальных параметров для каждой из функций можно вести посредством минимизации дисперсии невязок. Предположим, что параметры  $(x_0, y_0, \theta, \sigma, \lambda)$  зафиксированы. Оптимальное значение параметра  $\varphi$  определяется путем минимизации ошибки:

$$\varepsilon_{\Sigma}(A_i, \varphi) = \sum_{x,y} \varepsilon_i^2(x, y) = \sum_{x,y} \left[ f_i(x, y) - \bar{f}_i - A_i g_{\theta, \sigma, \lambda, \varphi}(x - x_0, y - y_0) \right]^2. \quad (3)$$

Очевидно, это соответствует минимизации выражения

$$\varepsilon_{\Sigma}(C_i, S_i) = \sum_{x,y} \left[ f_i(x, y) - \bar{f}_i - C_i g_{\theta, \sigma, \lambda, 0}(x - x_0, y - y_0) - S_i g_{\theta, \sigma, \lambda, -\pi/2}(x - x_0, y - y_0) \right]^2, \quad (4)$$

причем  $C_i = A_i \cos \varphi$ ,  $S_i = A_i \sin \varphi$ . Отсюда следует, что  $A_i$  — это отклик сложной клетки, работу которой можно трактовать как минимизацию локальной ошибки описания изображения.

На рис. 2, а, б приведен пример габоровской функции при  $\varphi=0$  и  $\varphi=\pi/2$ , а на рис. 2, в, г представлены соответственно исходное изображение и изображение как результат вычитания функции Габора с параметрами, минимизирующими дисперсию невязок (3).

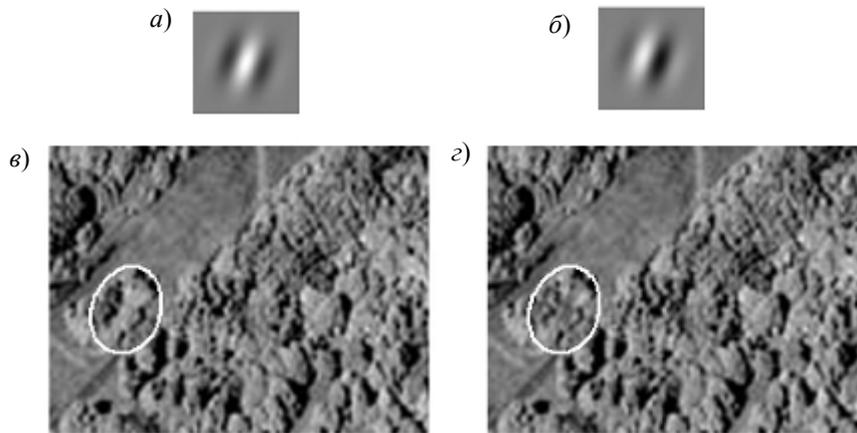


Рис. 2

Работу клеток следующего уровня можно интерпретировать как минимизацию второго слагаемого длины описания  $L_{S_T}(\{\mathbf{w}_{i,j}\}_{j=1}^{N_i})$ . Действительно, если изображение разделено на области, то параметры функций Габора в каждой области можно описывать совместно. Чем ближе значения этих параметров, тем меньшей будет длина описания. Поскольку значения параметров габоровских функций выбираются дискретно, то нетрудно получить оценку длины описания параметров через энтропии гистограмм внутри области:

$$L_{S_T}(\{\mathbf{w}_{i,j}\}_{j=1}^{N_i}) = N_i \left[ H(\{\theta_{i,j}\}_{j=1}^{N_i}) + H(\{\sigma_{i,j}\}_{j=1}^{N_i}) + \dots \right].$$

Таким образом, положительные и отрицательные связи между нейронами в рассмотренной выше модели можно интерпретировать как уменьшение длины совместного описания параметров габоровских функций, совокупность которых аппроксимирует значения яркостей изображения в данной области.

На основе принципа РМДО можно строго определить глобальный критерий качества описания, что позволяет установить наилучший способ введения латеральных связей в данной нейронной сети. Предложенное представление также предполагает явное группирование габоровских функций и определение областей на изображении. При этом некоторая габоровская функция описывает изображение лишь в той области, к которой она относится. И только в этой области должны вычисляться невязки. В нейросетевой модели в рецептивные поля

простых клеток могут попадать пиксели разных областей, что искажает отклик клеток и препятствует установлению точных границ областей в процессе сегментации. Вычисление невязок только внутри области предполагает управление рецептивными полями простых клеток с верхних уровней нейронной сети, на которых производится сегментация. Более детальный анализ перечисленных возможностей и разработка соответствующего алгоритма сегментации на основе предложенного представления изображений выходят за рамки данной статьи.

**Экспериментальные результаты.** В ходе эксперимента было произведено последовательное определение функции Габора с такими параметрами и координатами, при которых после ее вычитания длина описания содержимого области максимально уменьшалась. При этом были выбраны параметры  $\theta, \sigma, \lambda$  (один из 32 вариантов), описаны параметры разложения  $C_i$  и  $S_i$  (см. уравнение (4)) и координаты  $(x_0, y_0)$ . Длина описания одной функции Габора при этих условиях оценивалась равной 20 битам.

Сравнение длин описаний изображений проводилось в рамках двух представлений:  $S_0$  и  $S_T$ . Эксперименты показали, что применение функций Габора позволяет дополнительно уменьшить длину описания изображений, содержащих естественные текстуры, в среднем на 5,4 %. В соответствии с принципом РМДО это означает, что представление  $S_T$  (включающее аппроксимацию изображения функциями Габора) эффективнее представления  $S_0$  при описании изображений рассмотренного типа.

На рис. 3, а, б представлены исходное изображение и изображение как результат вычитания 114 функций Габора, а на рис. 3, в, г — соответствующие гистограммы распределения яркостей. Изменение гистограммы соответствует снижению энтропии с 7,3 до 7,1 и (с учетом сложности модели) уменьшению длины описания на 529 бит. Уменьшение длины описания наблюдалось для большинства (99 %) естественных текстур, но было сравнительно небольшим: это связано, по-видимому, с тем, что в эксперименте рассматривалась совокупность функций Габора с ограниченным набором параметров.

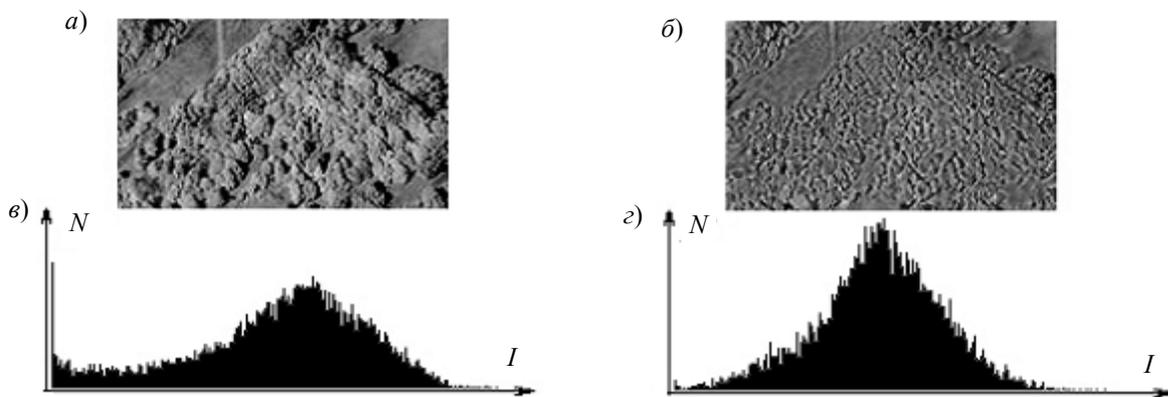


Рис. 3

Было установлено также, что уменьшение количества различных параметров функций Габора для некоторой области может приводить к уменьшению длины описания, однако дополнительное ограничение области значений возможных параметров, как правило, ведет к ухудшению результатов. Для разных текстур предпочтительный набор параметров оказался различным, что может быть использовано для их разделения при сегментации.

Анализ искусственных текстур показал, что использование функций Габора приводит к уменьшению длины описания только в 28 % случаев, что свидетельствует о необходимости расширения представления изображений с помощью других средств описания их текстур.

**Заключение.** Исследование модели нейронов, селективных по отношению к пространственным паттернам, и анализ функционирования этой модели на основе принципа РМДО

позволили расширить разработанное ранее иерархическое представление изображений, включив в него модель текстурной сегментации. Как показали результаты экспериментов, в рамках расширенного представления длина описания изображений, содержащих естественные текстуры, в среднем уменьшилась, что, согласно принципу РМДО, означает повышение качества представления. Разработка и реализация автоматических методов текстурной сегментации изображений, воплощающих рассмотренное представление, является предметом дальнейших исследований.

## СПИСОК ЛИТЕРАТУРЫ

1. *Kersten D.* Predictability and redundancy of natural images // *J. Optical Soc. Amer. A.* 1987. Vol. 4, N 12. P. 2395—2400.
2. *Huang J. G., Mumford D. B.* Statistics of natural images and models // *Proc. of IEEE Intern. Conf. on Computer Vision and Pattern Recognition.* 1999. P. 541—547.
3. *Barlow H. B.* What is the computational goal of the neocortex? // *Large-scale neuronal theories of the brain / Eds.: C. Koch, J. L. Davis.* Cambridge, MA: MIT Press, 1994. P. 1—22.
4. *Atick J. J.* Entropy minimization: A design principle for sensory perception? // *Intern. J. of Neural Systems.* 1992. Vol. 3. P. 81—90.
5. *Field D. J.* Wavelets, vision and the statistics of natural scenes // *Phil. Trans. R. Soc. Lond. A.* 1999. Vol. 357. P. 2527—2542.
6. *Grigorescu C., Petkov N., Westenberg M.A.* Contour detection based on nonclassical receptive field inhibition // *IEEE Trans. on Image Processing.* 2003. Vol. 12, N 7. P. 729—738.
7. *Daugman J. G.* Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters // *J. Optical Soc. Amer. A.* 1985. Vol. 2, N 7. P. 1160—1169.
8. *Потапов А. С.* Выбор представлений изображений на основе минимизации репрезентационной длины их описания // *Изв. вузов. Приборостроение.* 2008. Т. 51, № 7. С. 3—7.
9. *Potapov A. S.* Image matching with the use of the minimum description length approach // *Proc. SPIE.* 2004. Vol. 5426. P. 164—175.
10. *Potapov A. S., Gamayunova O. S.* Information criterion for constructing the hierarchical structural representations of images // *Proc. SPIE.* 2005. Vol. 5807. P. 443—454.

## Сведения об авторах

- Алексей Сергеевич Потапов** — д-р техн. наук; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра компьютерной фотоники; E-mail: pas.aicv@gmail.com
- Игорь Петрович Гуров** — д-р техн. наук, профессор; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра компьютерной фотоники
- Антон Нилович Аверкин** — студент; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра компьютерной фотоники; E-mail: ant-averkin@rambler.ru

Рекомендована кафедрой  
компьютерной фотоники

Поступила в редакцию  
19.02.08 г.