УДК 004.932.2

М.Н. ФАВОРСКАЯ, А.В. ПРОСКУРИН

КАТЕГОРИЗАЦИЯ СЦЕН НА ОСНОВЕ РАСШИРЕННЫХ ЦВЕТОВЫХ ДЕСКРИПТОРОВ

Фаворская М.Н., Проскурин А.В. **Категоризация сцен на основе расширенных цветовых дескрипторов.**

Аннотация. Категоризация сцен при автоматическом аннотировании изображений предполагает обязательный этап извлечения дескрипторов для построения гистограмм визуальных слов. Изучено семейство новых цветовых дескрипторов на основе точечных особенностей, инвариантных не только к геометрическим преобразованиям, но к изменениям освещенности. Особенностью дальнейшего алгоритма является предварительная цветовая и текстурная сегментация на основе алгоритма J-SEG с ранжированием полученных регионов по площади. Для построения визуальных слов и категоризации по методу опорных векторов используются расширенные цветовые дескрипторы, рассчитанные в 5–7 регионах с наибольшей площадью. Представлены сравнительные результаты экспериментальных оценок точности категоризации изображений из тестового набора 2688 изображений с применением расширенных цветовых дескрипторов.

Ключевые слова: автоматическое аннотирование изображений, категоризация сцен, метод опорных векторов, цветовые дескрипторы.

Favorskaya M.N., Proskurin A.V. **Scene Categorization Based on Extended Color Descriptors.**

Abstract. In automatic annotation systems, a scene categorization involves the compulsory stage of descriptor extraction in order to build a histogram of visual words. A family of new color descriptors based on point features, which are invariant not only to geometric transforms but also light changing, is investigated. In following, the algorithm executes a preliminary color and texture segmentation based on J-SEG algorithm. The received regions are ranked by areas. The extended color descriptors computing in 5–7 large area regions are applied for visual word construction. Then images are categorized by support vector machine. The comparative results of experimental estimators present the precision values of image categorization by use a test dataset containing 2,688 images.

Keywords: automatic image annotation, scene categorization, support vector machine, color descriptors.

1. Введение. Активное распространение цифровых устройств со встроенными видеокамерами привело к экспоненциальному увеличению количества изображений, доступных пользователям в сети Интернет. В связи с этим возникла проблема их эффективного поиска. Возможное решение заключается в автоматическом аннотировании изображений (ААИ) и последующем использовании хорошо известных методов текстового поиска. При этом под ААИ подразумевается автоматическая генерация текстового описания изображения на основе анализа его содержания [1]. Также активно предпринимаются попытки реализации систем ААИ на мобильных платформах, имеющих ограниченные вычислительные ресурсы [2]. Данная тенденция способствует

разработке быстрых алгоритмов ААИ. При этом наиболее рациональным представляется подход, когда вначале выполняется категоризация изображений по типу сцены (например, внутри / снаружи помещения и т. д.), после чего в каждой категории используется дерево решений для определения ключевых слов.

В общем случае системы категоризации сцен используют машинное обучение на основе признакового описания изображений. Однако в изображениях одной и той же категории возможны существенные различия в ракурсе съемки (рисунок 1, а), условиях освещения (рисунок 1, б) и наличии дополнительных объектов, не принадлежащих к категории (рисунок 1, в). Изображения, представленные на рисунке 1, взяты из тестового набора OT8 [3].

Рис. 1. Пример различий: (а) в ракурсе съемки, (б) условиях освещения, (в) наличии небольших объектов

Указанные артефакты приводят к снижению точности категоризации. Для решения этих проблем в данной статье предложено семейство локальных дескрипторов, инвариантных к повороту и масштабированию объектов, сдвигу и масштабированию цветовой интенсивности, а также метод категоризации сцен, использующий вычисление признаков только для наибольших по площади регионов изображения.

2. Локальные дескрипторы. В настоящее время известны различные дескрипторы, однако наиболее распространенными являются Scale-Invariant Feature Transform (SIFT) [4] и Speeded-Up Robust Features (SURF) [5]. В данной работе рассматривается дескриптор SURF и его модификации, т. к. экспериментально показано, что применение SURF дескрипторов требует на порядок меньших вычислительных ресурсов, обеспечивая примерно одинаковые результаты по сравнению с дескриптором SIFT [6, 7].

Базовый алгоритм SURF состоит из двух частей: обнаружение точек интереса и построение дескриптора. Обнаружение точек интереса осуществляется с помощью матрицы Гессе $\mathbf{H}(\mathbf{p},\sigma)$:

$$\mathbf{H}(\mathbf{p};\sigma)=\begin{bmatrix} L_{xx}(\mathbf{p},\sigma) & L_{xy}(\mathbf{p},\sigma) \\ L_{xy}(\mathbf{p},\sigma) & L_{yy}(\mathbf{p},\sigma) \end{bmatrix}, \quad (1)$$

где $\mathbf{p}(x,y)$ — точка в изображении I, σ — масштаб фильтра, $L_{xx}(\mathbf{p},\sigma)$ — свертка части изображения $I(\mathbf{p})$ в точке $\mathbf{p}$ со второй производной Гауссиана $g(\sigma)$:

$$L_{xx}(\mathbf{p},\sigma)=I(\mathbf{p})*\frac{\partial^2}{\partial x^2}g(\sigma). \quad (2)$$

Значения $L_{xy}(\mathbf{p},\sigma)$ и $L_{yy}(\mathbf{p},\sigma)$ вычисляются аналогично выражению (2). Определитель матрицы Гессе (гессиан) обладает инвариантностью относительно вращения, однако чувствителен к изменению масштаба. В связи с этим гессианы вычисляются для нескольких масштабов изображения, тем самым формируя пирамиду карт отклика. В качестве точек интереса выбираются локальные максимумы Гессианов, соответствующие локальным максимумам изменения градиента яркости (пятна, углы и края линий и т. п.).

Далее возле найденной точки интереса выбирается квадратный регион с размером сторон $20s$, где s – масштаб. Полученный регион интереса разбивается на квадратные блоки размером 4×4 элементов. В каждом блоке для 5×5 равномерно распределенных точек вычисля-

ются отклики вейвлета Хаара по горизонтальному L_x и вертикальному L_y направлениям. При этом полученные значения взвешиваются с помощью фильтра Гаусса, центрированного на точке интереса для подавления шумов. На следующем шаге для каждого блока формируется вектор $\mathbf{VD}_{SURF} = \left(\sum L_x, \sum L_y, \sum |L_x|, \sum |L_y|\right)$, образуя часть дескриптора. Конечный SURF дескриптор представляет собой объединение всех 16 векторов и, таким образом, имеет размерность 64.

SURF дескриптор инвариантен к повороту и масштабированию, однако в связи с использованием фильтра Гаусса происходит размытие краев и деталей изображения, что снижает точность описания. Для решения этой проблемы в работе [8] было предложено семейство дескрипторов Gauge SURF (G-SURF), основанных на использовании калибровочных координат. При построении этих дескрипторов в каждой точке интереса рассчитываются вектор градиента **w** и перпендикулярный к нему вектор **v**:

$$\mathbf{w} = \left(\frac{\partial L(\mathbf{p},\sigma)}{\partial x}, \frac{\partial L(\mathbf{p},\sigma)}{\partial y} \right) = \frac{1}{\sqrt{L_x^2(\mathbf{p},\sigma) + L_y^2(\mathbf{p},\sigma)}} \cdot \left(L_x(\mathbf{p},\sigma), L_y(\mathbf{p},\sigma) \right), \quad (3)$$

$$\begin{aligned} \mathbf{v} &= \left(\frac{\partial L(\mathbf{p},\sigma)}{\partial y}, -\frac{\partial L(\mathbf{p},\sigma)}{\partial x} \right) = \\ &= \frac{1}{\sqrt{L_x^2(\mathbf{p},\sigma) + L_y^2(\mathbf{p},\sigma)}} \cdot \left(L_y(\mathbf{p},\sigma), -L_x(\mathbf{p},\sigma) \right) \end{aligned}. \quad (4)$$

Наибольший интерес представляют производные второго порядка выражений (3), (4), использующие матрицы Гессе (выражение (1)) и обозначенные как $L_{ww}(\mathbf{p},\sigma)$ и $L_{vv}(\mathbf{p},\sigma)$:

$$\begin{aligned} L_{ww}(\mathbf{p},\sigma) = \frac{1}{L_x^2(\mathbf{p},\sigma) + L_y^2(\mathbf{p},\sigma)} \left(L_x(\mathbf{p},\sigma) \;\; L_y(\mathbf{p},\sigma) \right) \cdot \\ \cdot \begin{pmatrix} L_{xx}(\mathbf{p},\sigma) & L_{xy}(\mathbf{p},\sigma) \\ L_{yx}(\mathbf{p},\sigma) & L_{yy}(\mathbf{p},\sigma) \end{pmatrix} \begin{pmatrix} L_x(\mathbf{p},\sigma) \\ L_y(\mathbf{p},\sigma) \end{pmatrix} \end{aligned}, \quad (5)$$

$$\begin{aligned} L_{vv}(\mathbf{p},\sigma) = \frac{1}{L_x^2(\mathbf{p},\sigma) + L_y^2(\mathbf{p},\sigma)} \left(L_y(\mathbf{p},\sigma) \;\; -L_x(\mathbf{p},\sigma) \right) \cdot \\ \cdot \begin{pmatrix} L_{xx}(\mathbf{p},\sigma) & L_{xy}(\mathbf{p},\sigma) \\ L_{yx}(\mathbf{p},\sigma) & L_{yy}(\mathbf{p},\sigma) \end{pmatrix} \begin{pmatrix} L_y(\mathbf{p},\sigma) \\ -L_x(\mathbf{p},\sigma) \end{pmatrix} \end{aligned}. \quad (6)$$

Выражение (6) для расчета $L_{vv}(\mathbf{p},\sigma)$ часто используется как детектор «хребтов» («хребет» – это протяженный регион с приблизительно постоянной шириной и интенсивностью, точки которого являются локальными максимумами). Выражение (5), вычисляющее $L_{ww}(\mathbf{p},\sigma)$, содержит информацию об изменении градиента в направлении градиента. Тем самым, дескриптор G-SURF не размывает края на изображении, в то же время оказывает эффект размытия на текстуру, что является положительным фактором для снижения шумов.

В общем виде схема вычисления дескриптора G-SURF совпадает с базовым алгоритмом построения SURF дескриптора, однако для описания блоков региона интереса используется вектор $\mathbf{VD}_{G-SURF} = \left(\sum L_{ww}, \sum L_{vv}, \sum |L_{ww}|, \sum |L_{vv}|\right)$. При этом фильтр Гаусса не оказывает эффекта размытия всего изображения, что позволяет повысить точность описания. Дескриптор G-SURF также как и дескриптор SURF инвариантен к повороту и масштабированию объектов. Однако оба дескриптора SURF и G-SURF вычисляются только на изображениях в оттенках серого и не учитывают цветовую информацию, полезную при категоризации сцен.

3. Цветовые дескрипторы. Для описания цветовой информации были предложены дескрипторы, инвариантные к изменениям цветовой интенсивности, среди которых следует отметить следующие хорошо зарекомендовавшие себя дескрипторы:

- rg-гистограмма основана на нормализованной цветовой модели RGB, в которой хроматические компоненты *r* и *g* описывают цветовую информацию [9]. При этом компонент *b* является избыточным, поскольку $r+g+b=1$:

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \dfrac{R}{R+G+B} \\ \dfrac{G}{R+G+B} \\ \dfrac{B}{R+G+B} \end{pmatrix}. \qquad (7)$$

Благодаря нормализации компоненты *r* и *g* инвариантны к масштабированию интенсивности.

- Opponent-гистограмма вычисляется для изображений в цветовом пространстве Opponent, включающего два цветовых канала O_1, O_2 и компоненту интенсивности O_3 [10]:

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \dfrac{R-G}{\sqrt{2}} \\ \dfrac{R+G-2B}{\sqrt{6}} \\ \dfrac{R+G+B}{\sqrt{3}} \end{pmatrix}. \quad (8)$$

Как отмечают авторы работы [10], цветовые компонент O_1 и O_2 инвариантны к сдвигу интенсивности, в то время как канал интенсивности O_3 не обладает такими инвариантными свойствами.

- Hue-гистограмма. Для перевода изображения из цветового пространства Opponent в цветовую модель HSI (Hue, Saturation, Intensity) используется следующее выражение:

$$\begin{pmatrix} h \\ s \\ i \end{pmatrix} = \begin{pmatrix} \operatorname{actan}\left(O_1/O_2\right) \\ \sqrt{O_1^2+O_2^2} \\ O_3 \end{pmatrix}. \quad (9)$$

При этом компонент h (оттенок) обладает нестабильностью вблизи серого цвета. В работе [10] было выяснено, что определенность оттенка обратно пропорциональна насыщенности (компонент s). Таким образом, hue-гистограмма становится более устойчивой при умножении каждого значения оттенка на соответствующее значение насыщенности. Hue-гистограмма инвариантна к сдвигу и масштабированию цветовой интенсивности.

- Нормализованная RGB-гистограмма. RGB-гистограмма не является инвариантной к изменениям в условиях освещения. Однако инвариантность к сдвигу и масштабированию интенсивности может быть достигнута нормализацией распределения значений пикселов [11]:

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \dfrac{R-\mu_R}{\sigma_R} \\ \dfrac{G-\mu_G}{\sigma_G} \\ \dfrac{B-\mu_B}{\sigma_B} \end{pmatrix}, \quad (10)$$

где μ_i — среднее значение в i-ом канале, σ_i — среднеквадратичное отклонение в i-ом канале. Среднее значение μ_i и среднеквадратичное

отклонение σ_i вычисляются по выбранной области (блок или все изображение).

4. Семейство цветовых G-SURF дескрипторов. На основе G-SURF и приведенных выше цветовых дескрипторов (выражения (7) – (10)) разработано семейство дескрипторов, инвариантных к повороту и масштабированию объектов, сдвигу / масштабированию интенсивности:

- rgG-SURF. Для обеих компонент *r* и *g* вычисляются G-SURF дескрипторы, после чего они соединяются в один итоговый, размерностью 2×64. Такой дескриптор обладает инвариантностью к масштабированию интенсивности.
- OppG-SURF описывает все каналы цветового пространства Opponent с помощью G-SURF дескриптора. Благодаря компонентам O_1 и O_2 этот дескриптор инвариантен к сдвигу интенсивности.
- HueG-SURF. G-SURF вычисляется для взвешенного канала *hue* в цветовом пространстве HSI. Этот дескриптор обладает инвариантностью к сдвигу и масштабированию интенсивности. Следует отметить, что в работе [10] предложен дескриптор, в котором SIFT дескриптор, вычисленный на изображении в оттенках серого, объединен с *hue*-гистограммой. Подобным образом был создан дескриптор HHG-SURF, в котором вместо SIFT использовался G-SURF дескриптор. Проведенные эксперименты, не включенные в эту работу, показали, что HueG-SURF и HHG-SURF имеют близкие по значениям результаты категоризации, однако вычисление и последующая обработка HHG-SURF требует больших вычислительных затрат в связи с большей размерностью дескриптора (100 для HHG-SURF).
- RGBG-SURF вычисляется для всех нормализованных каналов цветового пространства RGB. Полученный дескриптор инвариантен к сдвигу и масштабированию интенсивности, а также к изменению интенсивности цветов в отдельных каналах.

5. Алгоритм категоризации сцен. Представим задачу категоризации сцен в следующем виде. Пусть $\mathbf{C}=\{c_1,c_2\ldots,c_c\}$ — множество категорий сцен, полученных на этапе обучения, а $\mathbf{D}=\{\mathbf{d}_1,\mathbf{d}_2,\ldots,\mathbf{d}_k\}$ — множество глобальных дескрипторов изображений. Тогда категоризация изображения заключается в поиске отображения $f:\mathbf{d}_i\to c_j$, которое однозначно ассоциирует дескриптор *i*-го изображения $\mathbf{D}_i$ с категорией c_j. Таким образом, категоризация сцен требует два инструмента: метод извлечения глобальных дескрипторов и их классификатор.

Для описания изображения широко используется метод Bag-of-Visual-Words (BoVWs) [12], состоящий из трех этапов:

– извлечение из изображений локальных дескрипторов (в данном случае вычисление цветовых G-SURF дескрипторов, указанных в п. 4);

– кластеризация полученных дескрипторов и создание из центров кластеров словаря визуальных слов $\mathbf{V} = \{v_1, v_2 \ldots, v_n\}$;

– формирование на основе набора локальных дескрипторов изображения I глобального дескриптора $\mathbf{VD}_i$ как гистограммы визуальных слов.

Полученные глобальные дескрипторы вместе с метками категорий затем используются для обучения классификатора. Часто для этого применяется машина опорных векторов (Support Vector Machine – SVM) [13]. Рассмотрим данный метод подробнее.

Пусть имеется набор данных для двух категорий $S = \{(\mathbf{D}, \mathbf{L})\}$, где $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k\}$ — множество глобальных дескрипторов изображений, а $\mathbf{L} = \{l_1, l_2 \ldots, l_m\}$ — множество соответствующих изображениям меток категорий, принимающих значения ± 1. В случае линейной разделимости классов метод опорных векторов вычисляет гиперплоскость следующего вида:

$$f(\mathbf{D}) = \operatorname{sign}\left(\left\langle \mathbf{z}^{\mathrm{T}}, \mathbf{D} \right\rangle + b\right),$$

где $\mathbf{z}$ — вектор, перпендикулярный к гиперплоскости; b — параметр, характеризующий расстояние от начала координат до гиперплоскости.

Искомая гиперплоскость должна разделять пространство признаков таким образом, чтобы в одном полупространстве оставались объекты только одного класса. Если получено несколько таких гиперплоскостей, то выбирают гиперплоскость с максимальной шириной разделяющей полосы (расстоянием от гиперплоскости до объектов классов). Для этого решается следующая оптимизационная задача:

$$\begin{cases} \left\langle \mathbf{z}, \mathbf{z} \right\rangle \to \min \\ l_i\left(\left\langle \mathbf{z}^{\mathrm{T}}, \mathbf{D}_i \right\rangle + b\right) \geq 1, \quad i = 1, \ldots, k \end{cases}.$$

После вычисления необходимых параметров классификатор может быть использован для категоризации новых данных. Однако на практике данные редко разделяются линейно. Для решения этой проблемы используется два способа. Первый заключается в том, что алгоритму позволяется допускать ошибки на обучающей выборке. С этой целью вводятся дополнительные переменные $\xi_i \geq 0$, характеризующие

величину ошибки для объектов $\mathbf{D}_i$, $i = 1, \ldots, k$. В этом случае оптимизационная задача принимает следующий вид:

$$\begin{cases} \dfrac{1}{2}\langle \mathbf{z}, \mathbf{z} \rangle + \alpha \sum\limits_{i=1}^{N} \xi_i \to \min \\ l_i \left(\langle \mathbf{z}^{\mathrm{T}}, \mathbf{D}_i \rangle + b \right) \geq 1, \quad i = 1, \ldots, k, \\ \xi_i \geq 0, \quad i = 1, \ldots, k \end{cases}$$

где α — параметр, позволяющий регулировать отношение между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки.

Другой способ решения проблемы линейной неразделимости классов основан на переходе от исходного пространства признаков $\mathbf{R}^z$ к новому пространству с более высокой размерностью $\mathbf{H}$ с помощью преобразования $\varphi : \mathbf{R}^z \to \mathbf{H}$. При этом отображение φ выбирается таким образом, чтобы в пространстве $\mathbf{H}$ данные были линейно разделимы. В полученном пространстве $\mathbf{H}$ построение SVM проводится точно также, как и ранее, однако скалярное произведение $\langle \mathbf{D}, \mathbf{D}' \rangle$ в пространстве $\mathbf{R}^z$ заменяется на ядро:

$$K(\mathbf{D}, \mathbf{D}') = \langle \varphi(\mathbf{D}), \varphi(\mathbf{D}') \rangle .$$

До сих пор не существует общих методов выбора ядра, поэтому на практике чаще всего используют следующие виды:

– полиномиальное ядро:

$$K(\mathbf{D}, \mathbf{D}') = \left(\gamma \langle \mathbf{D}, \mathbf{D}' \rangle + \alpha \right)^{\beta},$$

где γ, α, β — настраиваемые коэффициенты;

– радиальная базисная функция (РБФ):

$$K(\mathbf{D}, \mathbf{D}') = \exp\left(-\gamma \left\| \mathbf{D} - \mathbf{D}' \right\|^2 \right),$$

– сигмоид:

$$K(\mathbf{D}, \mathbf{D}') = \tanh\left(\gamma \langle \mathbf{D}, \mathbf{D}' \rangle + \alpha \right).$$

Следует отметить, что базовый алгоритм SVM разработан для классификации данных на два класса. Для решения задач классификации нескольких классов применяется метод «один против всех», в ре-

зультате которого обучается m SVM-классификаторов – по одному для каждой категории c_j.

Несмотря на простоту и эффективность метода опорных векторов, у него есть существенный недостаток – SVM неустойчив к шуму. Одним из способов решения этой проблемы является повышение эффективности описания изображений. Исходя из предположения, что изображения сцен состоят из наборов однородных регионов и небольших нетипичных объектов, в данной статье предлагается предварительно сегментировать изображения, после чего извлекать признаки только из 5–7 крупных регионов. Таким образом, алгоритм категоризации сцен имеет следующий вид:

– Шаг 1. Сегментация всех изображений (в работе применен алгоритм цвето-текстурной сегментации J-SEG [14]).

– Шаг 2. Выбор 5–7 наибольших регионов в каждом изображении.

– Шаг 3. Извлечение локальных дескрипторов из выбранных регионов.

– Шаг 4. Создание словаря визуальных слов (используется алгоритм k-средних).

– Шаг 5. Формирование BoVWs-дескрипторов изображений.

– Шаг 6. Обучение SVM-классификаторов.

В предлагаемом алгоритме существенное место занимают локальные дескрипторы, от которых требуется высокая устойчивость к различным изменениям.

6. Результаты экспериментальных исследований. Для экспериментов использовался набор из 8 категорий сцен (далее OT8) [3]. OT8 состоит из 2688 изображений, разделенных на 8 категорий: coast, mountain, forest, open country, street, inside city, tall buildings и highways. Размер каждого изображения 256×256 пикселов. Для обучения из каждой категории случайным образом выбиралось по 100 изображений, остальные использовались для тестирования. На рисунке 2 представлены примеры оригинальных изображений из набора OT8, их сегментированные прототипы (с применением J-SEG алгоритма) с закрашенными белым цветом небольшими регионами и изображения с найденными точками интереса.

Для формирования словаря визуальных слов из обучающей выборки случайным образом выбиралось 200 000 дескрипторов, которые кластеризовались с помощью алгоритма k-средних. В этой работе количество кластеров (визуальных слов) равно 400. С помощью словаря каждому изображению присваивалось BoVWs-описание. В качестве классификатора использовалась реализованная в библиотеке LibSVM [15] машина опорных векторов с ядром в виде радиальной

базисной функции. Все вычисления повторялись 5 раз, после чего точность усреднялась.

Для оценивания эффективности предложенных дескрипторов в тестовый набор были искусственно добавлены изменения: поворот изображений (значения углов: $\pm 2{,}5°, \pm 5{,}0°, \pm 7{,}5°, \pm 10°$), масштабирование интенсивности (значения множителей: $1{,}1^{\pm 1}, 1{,}25^{\pm 1}, 2{,}0^{\pm 1}$) и сдвиг интенсивности (значения: $\pm 5, \pm 10, \pm 15, \pm 20$). В таблицах 1, 2 и 3 представлены данные, полученные при категоризации измененных изображений.

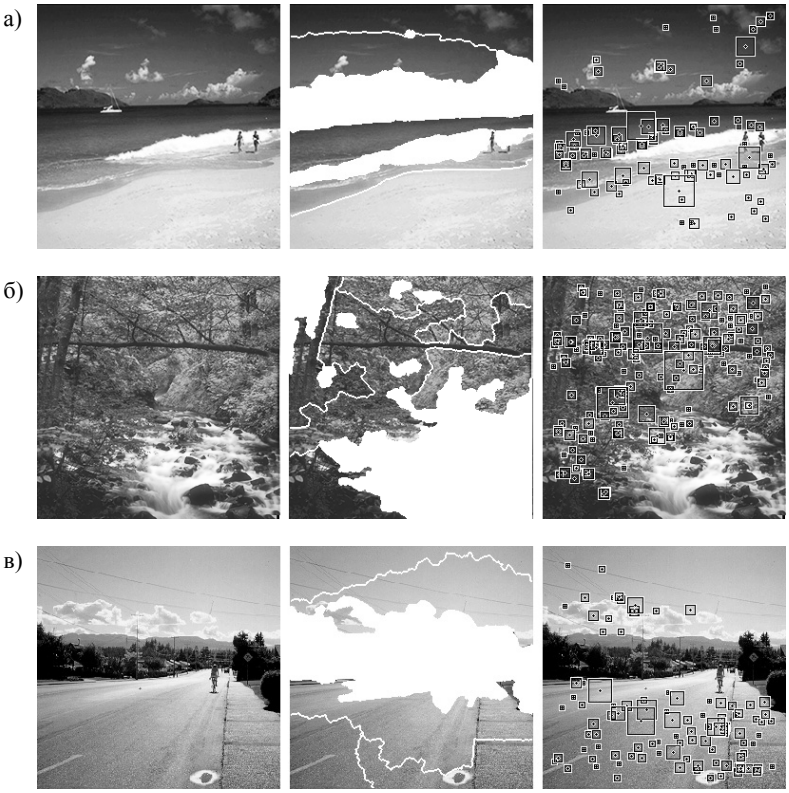

Рис. 2. Пример исходных изображений, их сегментированных прототипов и найденных точек интереса из набора изображений OT8: (а) coast_bea3; (б) forest_land810; (в) highway_urb471; (г) insidecity_hous50; (д) mountain_sharp48; (е) opencountry_open55; (ж) street_par203; (з) tallbuilding_urban1210

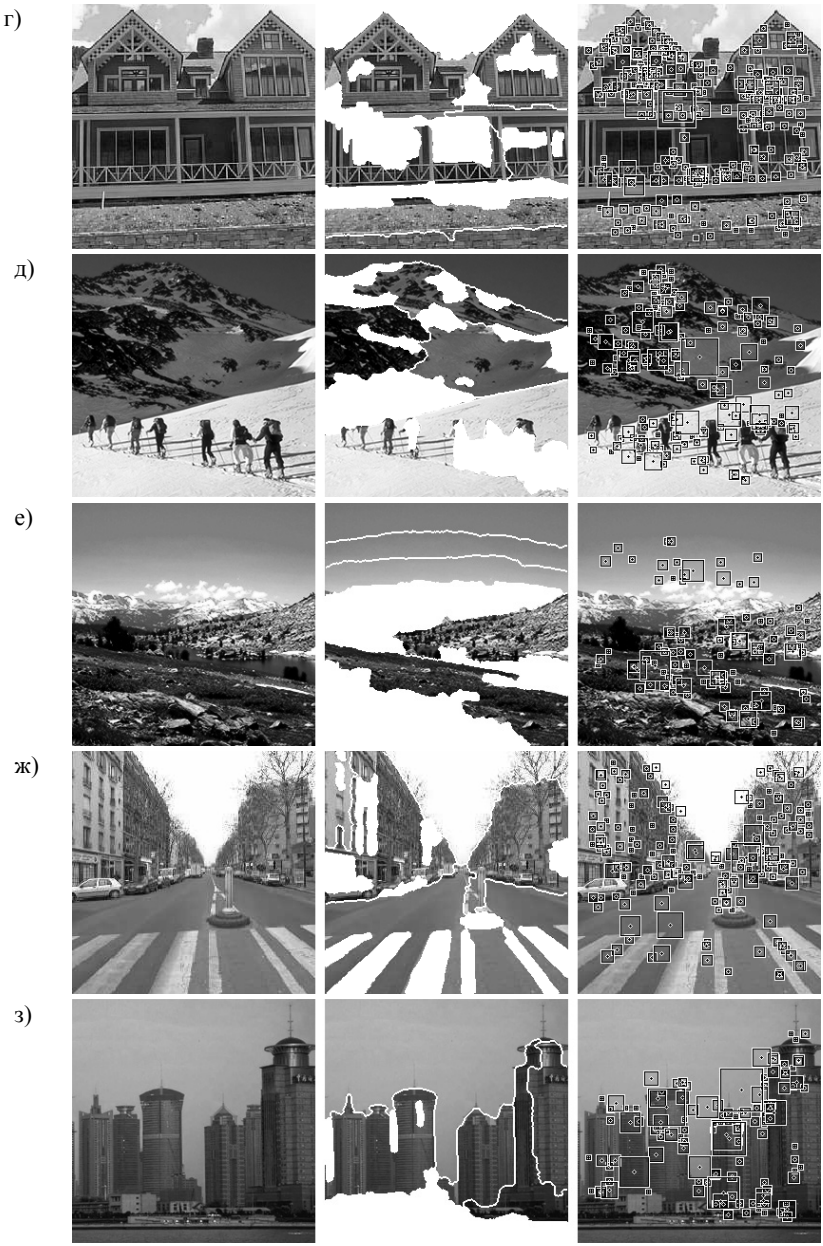

Рис. 2. (продолжение)

Таблица 1. Результаты точности категоризации набора OT8 при поворотах изображений (%)

Угол поворота, °	–10	–7,5	–5,0	–2,5	0,0	2,5	5,0	7,5	10
SURF	78,4	78,8	79,3	79,7	80,1	79,6	79,1	78,9	78,6
G-SURF	82,9	83,4	83,9	84,3	84,4	84,1	83,8	83,5	83,1

По результатам, представленным в таблицы 1, отметим, что устойчивость дескрипторов SURF и G-SURF к поворотам приблизительно одинаковая, однако точность категоризации с использованием дескриптора G-SURF на 4–5 % выше.

Таблица 2. Результаты точности категоризации набора OT8 при масштабировании интенсивности (%)

Множитель	2^{-1}	$1{,}5^{-1}$	$1{,}25^{-1}$	$1{,}1^{-1}$	1	1,1	1,25	1,5	2
G-SURF	82,4	83,7	84,4	85,1	85,4	83,6	80,6	74,3	59,9
rgG-SURF	76,6	76,8	77,0	77,4	77,6	76,9	76,9	74,6	68,7
OppG-SURF	84,2	84,4	84,7	84,9	85,3	84,5	83,3	79,7	70,1
HueG-SURF	66,8	70,0	70,1	71,3	71,6	71,5	70,4	67,9	62,7
RGBG-SURF	85,2	85,3	85,3	85,6	85,7	84,2	83,2	77,2	61,0

Из таблицы 2 видно, что дескриптор HueG-SURF не обладает устойчивостью к масштабированию интенсивности, а дескриптор G-SURF обладает лишь частичной устойчивостью. При этом применение дескрипторов rgG-SURF и HueG-SURF дает низкую точность категоризации. Резкое падение точности в двух последних столбцах (при значениях множителей 1,5 и 2) связано с тем, что значения большей части пикселов на изображениях превысили 255 и были обрезаны.

Таблица 3. Результаты точности категоризации набора OT8 при сдвиге интенсивности (%)

Сдвиг	–20	–15	–10	–5	0	5	10	15	20
G-SURF	82,2	83,4	83,9	84,7	85,4	84,9	84,2	83,4	82,6
rgG-SURF	76,4	76,7	77,2	77,5	77,6	77,6	77,4	77,4	77,1
OppG-SURF	85,3	85,6	85,8	85,5	85,3	85,2	85,1	84,6	84,3
HueG-SURF	70,6	70,8	70,9	71,6	71,6	71,2	71,0	70,9	70,9
RGBG-SURF	84,4	84,9	85,1	85,3	85,7	84,9	84,8	84,8	84,6

Как следует из таблицы 3, дескриптор G-SURF обладает частичной устойчивостью к сдвигу значений интенсивности цветов (имеется в виду сдвиг интенсивности относительно белого цвета). При этом все четыре предложенных дескриптора устойчивы к сдвигу, однако дескрипторы rgG-SURF и HueG-SURF по-прежнему показывают низкую точность категоризации.

Были проведены дополнительные эксперименты для оценки точности определения отдельных категорий изображений. Полученные результаты для набора OT8 представлены в таблице 4.

Как видно из таблицы 4, разные дескрипторы показывают хорошие результаты только для определенных категорий изображений. Таким образом, общую точность категоризации можно повысить, вычисляя дескрипторы на разных цветовых каналах с автоматическим определением весов для каждой категории.

Таблица 4. Результаты точности определения отдельных категорий изображений в наборе OT8 (%)

Категория	Coast	Forest	Highway	Inside city	Mountain	Open country	Street	Tall building
G-SURF	78,3	93,9	80,4	77,9	88,2	*77,8*	91,1	*95,9*
rgG-SURF	84,1	91,8	77,2	69,7	71,4	61,6	77,6	87,3
OppG-SURF	83,4	*96,4*	79,9	78,0	85,5	75,9	*91,5*	92,0
HueG-SURF	77,9	83,9	66,0	59,2	69,9	51,3	79,1	85,8
RGBG-SURF	78,3	95,7	82,9	79,4	*89,3*	*77,8*	88,5	93,9
Oliva A., Torralba A. [3]	79,0	91,0	*87,0*	*90,0*	81,0	71,0	89,0	82,0
Battiato S. и др. [16]	*85,0*	93,0	82,0	87,0	85,0	74,0	89,0	88,0
Gazolli K., Salles E. [17]	84,0	89,0	85,0	80,0	78,0	73,0	86,0	73,0

Следует отметить, что точность категоризации может быть повышена путем создания визуальных слов адаптивным алгоритмом кластеризации Enhanced Self-Organizing Incremental Neural Network [18]. Для формирования BoVWs-описания используя множественную ассоциацию (Multi-Assignment) [19] (ассоциация локального дескриптора не с одним визуальным словом, а с несколькими визуальными словами). Также метод сопоставления пространственных пирамид (Spatial Pyramid Matching) [20] позволит улучшить результаты категоризации.

7. Заключение. В статье представлен метод категоризации набора изображений, использующий расчет локальных дескрипторов только в больших по площади регионах изображений и выполняющий кластеризацию на основе машины опорных векторов. Разработано семейство новых цветовых дескрипторов, инвариантных к повороту и масштабированию объектов, а также к сдвигу и масштабированию цветовой интенсивности. Проведенные экспериментальные исследования показали, что точность категоризации достигает 78–96 % в зависимости от категории изображений, что на 4–5 % выше по сравнению

с применением известных дескрипторов. Приведены методы повышения точности категоризации, которые целесообразно рассмотреть в последующих исследованиях.

Литература

1. *Zhang D., Islam Md.M., Lu G.* A Review on Automatic Image Annotation Techniques // Pattern Recognition. 2012. vol. 45. no. 1. pp. 346–362.
2. *Qin C., Bao X., Choudhury R.R., Nelakuditi S.* TagSense: a Smartphone-based Approach to Automatic Image Tagging // Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys'11). 2011. pp. 1–14.
3. Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. URL: http://people.csail.mit.edu/torralba/code/spatialenvelope (дата обращения: 11.05.2015).
4. *Lowe D.G.* Distinctive Image Features from Scale-Invariant Keypoints // International Journal of Computer Vision. 2004. vol. 60. no. 2. pp. 91–110.
5. *Bay H., Ess A., Tuytelaars T., Gool L.V.* Speeded-Up Robust Features (SURF) // Computer Vision and Image Understanding. 2008. vol. 110. no. 3. pp. 346–359.
6. *Favorskaya M.*, *Jain L.C.*, *Buryachenko V.* Digital Video Stabilization in Static and Dynamic Scenes. Computer Vision in Control Systems-1 // ISRL. Springer Cham Heidelberg New York Dordrecht London: Springer International Publishing Switzerland. 2015. vol. 73. pp. 261–309.
7. *Jain L.C.*, *Favorskaya M.*, *Novikov D.* Panorama Construction from Multi-view Cameras in Outdoor Scenes. Computer Vision in Control Systems-2 // ISRL. Springer Cham Heidelberg New York Dordrecht London: Springer International Publishing Switzerland. 2015. vol. 75. pp. 71–108.
8. *Alcantarilla P.F., Bergasa L.M., Davison A.J.* Gauge-SURF Descriptors // Image and Vision Computing. 2013. vol. 31. no. 1. pp. 103–116.
9. *Gevers T., van de Weijer J., Stokman H.* Color Feature Detection: an Overview. Color Image Processing: Methods and Applications / edited by R. Lukac, K.N. Plataniotis. // University of Toronto. Ontario, Canada: CRC Press. 2006. pp. 203–226.
10. *Van de Weijer J., Gevers T., Bagdanov A.* Boosting Color Saliency in Image Feature Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006. vol. 28. no. 1. pp. 150–156.
11. *Van de Sande K.E.A., Gevers T., Snoek C.G.M.* Evaluating Color Descriptors for Object and Scene Recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009. vol. 32. no. 9. pp. 1582–1596.
12. *Csurka G., Dance C.R., Fan L., Willamowski J., Bray C.* Visual Categorization with Bags of Keypoints // Proceedings of Workshop on Statistical Learning in Computer Vision (ECCV'2004). 2004. pp. 1–22.
13. *Vapnik V.N.* Statistical Learning Theory // New York: Wiley. 1998. 768 p.
14. *Deng Y., Manjunath B. S.* Unsupervised Segmentation of Color-Texture Regions in Images and Videos // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001. vol. 23. no. 8. pp. 800–810.
15. LIBSVM – a Library for Support Vector Machines. URL: http://www.csie.ntu.edu.tw/~cjlin/libsvm (дата обращения: 11.05.2015).
16. *Battiato S., Farinella G.M., Guarnera M., Ravi D., Tomaselli V.* Instant Scene Recognition on Mobile Platform // Computer Vision (ECCV 2012). Workshops and Demonstrations. Lecture Notes in Computer Science. 2012. vol. 7585. pp. 655–658.
17. *Gazolli K., Salles E.* A Contextual Image Descriptor for Scene Classification // Trends in Innovative Computing. 2012. pp. 66–71.

18. *Проскурин А. В.* Формирование визуальных слов для автоматического аннотирования изображений на основе самоорганизующейся нейронной сети // Цифровая обработка сигналов и ее применение (ДСПА'2014). Сб. научн. тр. 16-й Международной конференции. Москва: ИПУ РАН. 2014. Т. 2. С. 487–491.
19. *Jiang Y.G., Ngo C., Yang J.* Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval // Proceedings of International Conference on Image and Video Retrieval (CIVR '2007). 2007. pp. 494–501.
20. *Lazebnik S., Schmid C., Ponce J.* Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2006. vol. 2. pp. 2169–2178.

References

1. Zhang D., Islam Md.M., Lu G. A Review on Automatic Image Annotation Techniques. *Pattern Recognition*. 2012. vol. 45. no. 1. pp. 346–362.
2. Qin C., Bao X., Choudhury R.R., Nelakuditi S. TagSense: a Smartphone-based Approach to Automatic Image Tagging. Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys'11). 2011. pp. 1–14.
3. Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. Available at: http://people.csail.mit.edu/torralba/code/spatialenvelope (accessed 11.05.2015).
4. Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. 2004. vol. 60. no. 2. pp. 91–110.
5. Bay H., Ess A., Tuytelaars T., Gool L.V. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*. 2008. vol. 110. no. 3. pp. 346–359.
6. Favorskaya M., Jain L.C., Buryachenko V. Digital Video Stabilization in Static and Dynamic Scenes. Computer Vision in Control Systems-1. *ISRL*. Springer Cham Heidelberg New York Dordrecht London: Springer International Publishing Switzerland. 2015. vol. 73. pp. 261–309.
7. Jain L.C., Favorskaya M., Novikov D. Panorama Construction from Multi-view Cameras in Outdoor Scenes. Computer Vision in Control Systems-2. *ISRL*. Springer Cham Heidelberg New York Dordrecht London: Springer International Publishing Switzerland. 2015. vol. 75. pp. 71–108.
8. Alcantarilla P.F., Bergasa L.M., Davison A.J. Gauge-SURF Descriptors. *Image and Vision Computing*. 2013. vol. 31. no. 1. pp. 103–116.
9. Gevers T., van de Weijer J., Stokman H. Color Feature Detection: an Overview. Color Image Processing: Methods and Applications. Edited by R. Lukac, K.N. Plataniotis. University of Toronto. Ontario. Canada: CRC Press. 2006. pp. 203–226.
10. Van de Weijer J., Gevers T., Bagdanov *A.* Boosting Color Saliency in Image Feature Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006. vol. 28. no. 1. pp. 150–156.
11. Van de Sande K.E.A., Gevers T., Snoek C.G.M. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009. vol. 32. no. 9. pp. 1582–1596.
12. Csurka G., Dance C.R., Fan L., Willamowski J., Bray C. Visual Categorization with Bags of Keypoints. Proceedings of Workshop on Statistical Learning in Computer Vision (ECCV'2004). 2004. pp. 1–22.
13. Vapnik V.N. Statistical Learning Theory. New York: Wiley. 1998. 768 p.
14. Deng Y., Manjunath B. S. Unsupervised Segmentation of Color-Texture Regions in Images and Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2001. vol. 23. no. 8. pp. 800–810.
15. LIBSVM – a Library for Support Vector Machines. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed 11.05.2015).

16. Battiato S., Farinella G.M., Guarnera M., Ravi D., Tomaselli V. Instant Scene Recognition on Mobile Platform. Computer Vision (ECCV 2012). Workshops and Demonstrations. Lecture Notes in Computer Science. 2012. vol. 7585. pp. 655–658.
17. Gazolli K., Salles E. A Contextual Image Descriptor for Scene Classification. *Trends in Innovative Computing*. 2012. pp. 66–71.
18. Proskurin A.V. [Creating Visual Words for Automatic Image Annotation Based on Self-Organizing Incremental Neural Network]. *Tsifrovaia obrabotka signalov I ee primenenie (DSPA-2014)*: *Sb. naychn. tr. 16-i Mezhdunarodnoi konferentsii* [Proceedings of the 16th International Conference "Digital Signal Processing and its Applications"]. Moscow: ISC RAS. 2014. vol. 2. pp. 487–491 (In Russ.).
19. Jiang Y.G., Ngo C., Yang J. Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval. Proceedings of International Conference on Image and Video Retrieval (CIVR '2007). 2007. pp. 494–501.
20. Lazebnik S., Schmid C., Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2006. vol. 2. pp. 2169–2178.

Фаворская Маргарита Николаевна — д-р техн. наук, профессор, заведующий кафедрой информатики и вычислительной техники, Институт информатики и телекоммуникаций Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева (СибГАУ). Область научных интересов: цифровая обработка изображений и видеопоследовательностей, распознавание образов, кластеризация, информационные технологии. Число научных публикаций — 170. favorskaya@sibsau.ru; пр. им. газ. "Красноярский рабочий", 31, Красноярск, 660014; р.т.: +7 391 291 9240, Факс: +7 391-291-91-47.

Favorskaya Margarita Nikolaevna — Ph.D., Dr. Sci., professor, head of informatics and computer techniques department, Institute of Informatics and Telecommunications of Siberian State Aerospace University named after academician M.F. Reshetnev (SibSAU). Research interests: digital image and videos processing, pattern recognition, fractal image processing, artificial intelligence, information technologies, remote sensing. The number of publications — 170. favorskaya@sibsau.ru; 31, Krasnoyarsky Rabochy av., Krasnoyarsk, 660014; office phone: +7 391 291 9240, Fax: +7 391 291 9147.

Проскурин Александр Викторович — аспирант кафедры информатики и вычислительной техники, Институт информатики и телекоммуникаций Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева (СибГАУ). Область научных интересов: цифровая обработка изображений, распознавание образов. Число научных публикаций — 17. Proskurin.AV.WOF@gmail.com; пр. им. газ. "Красноярский рабочий", 31, Красноярск, 660014; р.т.: +7(391)291-9241, Факс: +7(391)291-9147.

Proskurin Alexander Viktorovich — Ph.D. student of informatics and computer science department, Institute of Informatics and Telecommunications, Siberian State Aerospace University named after academician M.F. Reshetnev (SibSAU). Research interests: digital image processing, pattern recognition. The number of publications — 17. Proskurin.AV.WOF@gmail.com; 31, Krasnoyarsky Rabochy av., Krasnoyarsk, 660014; office phone: +7(391)291-9241, Fax: +7(391)291-9147.

РЕФЕРАТ

Фаворская М.Н., Проскурин А.В. **Категоризация сцен на основе расширенных цветовых дескрипторов.**

Значительный рост количества изображений в сети Интернет и необходимость их поиска предполагает разработку систем автоматической категоризации изображений. Однако задача автоматической категоризации не является тривиальной, поскольку между изображениями часто наблюдаются существенные различия в ракурсе съемки, условиях освещения и наличии объектов, не принадлежащих категории. Для решения этих проблем предложен алгоритм категоризации сцен на основе описания изображений как гистограмм визуальных слов и машины опорных векторов. Разработано семейство новых цветовых дескрипторов на основе локального дескриптора Gauge Speeded-Up Robust Features, инвариантного к повороту и масштабированию. Предложенные дескрипторы дополнительно являются инвариантными к изменениям цветовой интенсивности. Они применяются для описания 5–7 больших по площади регионов изображения после предварительной цвето-текстурной сегментации нс основе J-SEG алгоритма. Проведенные экспериментальные исследования показали, что точность категоризации достигает 78–96 % в зависимости от категории изображений, что на 4–5 % выше по сравнению с применением известных дескрипторов. Приведены методы повышения точности категоризации, которые целесообразно рассмотреть в последующих исследованиях.

SUMMARY

Favorskaya M.N., Proskurin A.V. **Scene Categorization Based on Extended Color Descriptors.**

Huge volume of images in WWW and necessity of their retrieval assume the development of systems for automatic images categorization. However, the task of automatic image categorization is not trivial due to essential differences between images in viewpoint shooting, light intensities, and additional objects in images, which do not concern to some category. The algorithm of scene categorization based on image description as a histogram of visual words and support vector machine is designed in order to compensate such image artifacts. Family of novel color descriptors based on local Gauge Speeded-Up Robust Features, which is invariant to rotation and scaling, has been developed. Additionally, the proposed descriptors are invariant to light intensity. They are used for description of 5–7 large area regions in image after preliminary color and texture segmentation based on J-SEG algorithm. Experimental researches show that values of categorization precision achieve 78–96 % in dependence on image category. These values exceed results received by use of conventional descriptors on 4–5 %. The improved categorization methods are mentioned as future investigation.