

УДК 004.932.2:004.93'1

М.М. Гридин<sup>1</sup>, В.А. Гороховатский<sup>2</sup><sup>1</sup> Харьковский национальный университет радиоэлектроники, Харьков<sup>2</sup> Харьковский учебно-научный институт ГВУЗ «Университет банковского дела», Харьков

## АНАЛИЗ ЛИНГВИСТИЧЕСКИХ ХАРАКТЕРИСТИК СТРУКТУРНЫХ ОПИСАНИЙ В РАСПОЗНАВАНИИ ВИЗУАЛЬНЫХ ОБЪЕКТОВ

Обсуждаются вопросы сопоставления описаний при распознавании визуальных объектов в системах компьютерного зрения. Предложена технология структурного распознавания, которая базируется на методе *tf-idf*, применяемом в математической лингвистике и информационном поиске. В результате выполнено преобразование описания из множества характерных признаков изображения к вектору лингвистических характеристик, что значительно снижает вычислительные затраты. Представлены результаты экспериментов, подтверждающих работоспособность и результативность метода.

**Ключевые слова:** компьютерное зрение, база изображений, структурное распознавание изображений, характерные признаки, схема взвешивания *tf-idf*, векторное описание, мера подобия, релевантность.

### Введение

Структурные описания изображений в системах компьютерного зрения создаются в виде множества дескрипторов – числовых векторов из пространства  $R^n$ . Описание  $Z \subset R^n$  визуального объекта представляют конечным мультимножеством дескрипторов – характерных признаков (ХП). Применяемый в обработке изображений метод SURF формирует описание в виде подмножества  $R_1^n = \{z \mid z \in R^n, \|z\| \approx 1\}$ , где  $R_1^n \subset R^n$  – подмножество  $n$ -мерных вещественных векторов с близкой к единице евклидовой нормой:  $\|z\| = \sqrt{\sum_{i=1}^n z_i^2} \approx 1$  [1, 2].

Поскольку число классов объектов в базе визуальных данных может быть значительным (тысячи и более эталонов или видео кадров), время, необходимое для поиска соответствия нового образа путем сравнения с каждым из эталонов базы данных может стать непомерно высоким. В таком случае вместо сравнения описаний изображений как множеств целесообразно перейти к сопоставлению векторов, характеризующих эти множества. Для этого нужно, например, путем кластеризации множества ХП сформировать векторное описание [1, 3].

Эффективным методом, применяемым в математической лингвистике для быстрого поиска документов по их содержанию, есть употребление показателей частоты встречаемости определенных слов из документа поиска и словаря, на котором построено содержание базы документов [3 – 6]. К таким методам относят *tf-idf*, который представлен рядом модификаций [4]. Распознавание по структурным признакам аналогично лингвистическому подходу не требует учета последовательности формирования дескрипторов в описании объекта. Может оказаться плодотворной идея построить структурное распознавание так, чтобы дескрипторы изображения представить в виде «визуальных слов» в сформированном для базы изображений «словаре», а лингвистические характеристики конкретного объекта – в виде вектора. В таком случае эталон и распознаваемый объект будут представлены векторным описанием, что значительно снижает затраты на распознавание.

Анализ лингвистических характеристик дает также возможность оценить, насколько часто встречается конкретное слово (в структурном описании – дескриптор) и сосредоточить поиск или обработку на редких и, следовательно, более информативных признаках-словах. При этом предварительно из описания можно отфильтровать часто встречающиеся признаки, не всегда полезные для принятия решений.

Целью работы есть изучение возможности и эффективности применения лингвистической схемы взвешивания *tf-idf* на множестве характерных признаков базы изображений в плане построения векторного представления описания объектов для структурного распознавания. Построение лингвистических характеристик позволяет перейти к сжатому векторному виду эталонов и объекта, что значительно сокращает объем вычислительных затрат и улучшает быстродействие распознавания.

Задачами исследования есть изучение особенностей построения вектора-признака *tf-idf* применительно к множествам дескрипторов структурных признаков изображений, а также оценивание качества и эффективности употребления векторного сравнения для прикладных примеров баз изображений.

Задачами исследования есть изучение особенностей построения вектора-признака *tf-idf* применительно к множествам дескрипторов структурных признаков изображений, а также оценивание качества и эффективности употребления векторного сравнения для прикладных примеров баз изображений.

### Построение схемы взвешивания *tf-idf*

Метод *tf-idf* применяют в задачах информационного поиска при сопоставлении и оценке содержания текстовых документов, составленных из ко-

нечного числа терминов – словаря [4]. Для этого вычисляют числовой вектор, компонентами которого есть вес отдельного термина.

В результате каждый документ получает описание в виде числового вектора с размерностью установленного объема словаря. Этот способ позволяет сравнивать между собой разнообразные по содержанию и размеру документы и получать значения релевантности (подобия) для пар документов.

Суть схемы tf-idf состоит в следующем. Вначале необходимо сформировать словарный состав – конечную базу терминов, в рамках которой реализуется лингвистический анализ.

Затем для каждого термина словаря вычисляется частота термина в документе (tf – term frequency) и обратная документная частота (idf – inverse document frequency).

Модель данных в методе tf-idf называют мешком слов (bag of words model). В рамках этой модели порядок следования терминов (контекст) в документе игнорируется, а основное внимание уделяется количеству вхождений каждого термина. Интуитивно ясно, что два документа с одинаковыми «мешками слов» по содержанию сходны [6]. Использование обратной документной частоты idf необходимо для ослабления воздействия часто встречающихся терминов, чтобы ограничить их влияние на оценку релевантности документов. Вес tf-idf термина определяется как произведение

$$g_i = \frac{a_i}{c} \cdot \log_2(b/d_i), \quad (1)$$

где  $a_i$  – число встречаемости  $i$ -го термина в документе общим объемом из  $c$  терминов,  $b$  – количество документов в коллекции,  $d_i$  – число документов из их общего фиксированного числа  $b$ , в которых встречается  $i$ -й термин,  $\log_2$  – логарифм числа при основании два [4].

Функция tf-idf реализует интуитивное представление о том, что чем чаще терм встречается в документе, тем точнее он описывает его содержание, и чем в большем числе документов встречается терм, тем менее различительным признаком он является.

В литературе [4, 6] предложено и исследовано огромное разнообразие вариантов метода tf-idf для решения прикладных задач. Основные идеи модификаций сводятся к построению эвристических правил для ранжирования и пороговой обработки значений компонент вектора tf-idf в целях управления весами терминов по критерию частотности, а также обеспечения необходимой скорости поиска данных в объемных базах документов.

Полученные по схеме tf-idf веса  $g_i$  нормализуют к 1 путем деления на норму вектора:

$$g_i = g_i / \sqrt{\sum_{i=1}^h g_i^2}, \quad (2)$$

где  $h$  – размер словаря.

Нормализацию (2) рекомендуется проводить, т.к. небольшие документы имеют тенденцию быть представленными короткими векторами термов, в то время как объемные документы – более длинными векторами.

Данная ситуация приводит к тому, что большие документы имеют более высокую вероятность быть выбранными по сравнению с документами незначительного объема.

В результате вычисления (1) – (2) получаем числовой вектор  $\{g_i\}_{i=1}^h$ . Каждый документ теперь будет интерпретироваться нормированным вектором (2), состоящим из компонент, соответствующих множеству терминов словаря.

Отметим следующие важные для распознавания множеств дескрипторов изображений свойства компоненты  $g_i$  вектора tf-idf [4]:

1)  $g_i$  достигает максимального значения, если термин встречается много раз в небольшом количестве документов (усиливая их отличие от других документов);

2)  $g_i$  уменьшается, если термин встречается в каком-то документе лишь несколько раз или встречается во многих документах (тем самым формируя менее выраженный сигнал о релевантности документа);

3)  $g_i$  достигает минимального значения, если термин встречается в подавляющем числе документов.

## Реализация лингвистического анализа для структурных описаний

Рассмотрим базу изображений в виде конечного множества  $Z = \{Z^i\}_{i=1}^J$  из  $J$  эталонных описаний. Обозначим мощности множеств  $s = \text{card } Z$ ,  $s_i = \text{card } Z^i$ . При этом  $s = \sum_{i=1}^J s_i$ . Каждое описание  $Z^i$  представляет собой конечное множество из  $s_i$  векторов-дескрипторов SURF размерностью 64 [2]. Определим словарь  $H$  как конечное подмножество  $H \subseteq Z$  множества  $Z$ . Построение словаря в базе  $Z = \{Z^i\}_{i=1}^J$  реализуем последовательностью действий.

1. Для  $z \in Z$  ищем подобные ему  $z^* \in Z^{\wedge}$

$$\rho(z, z^*) \leq \varepsilon, \quad (3)$$

где  $\varepsilon$  – некоторый порог подобия;  $\rho$  – расстояние между дескрипторами в векторном пространстве (например, евклидово).

2. Исключаем из  $Z$  элементы, удовлетворяющие (3), и вносим элемент  $Z$  в словарь  $H$ , фиксируя число его повторений для вычисления (1).

3. Для остальных элементов  $Z$  выполняем шаги 1, 2 до тех пор, пока множество  $Z$  не окажется пустым.

В результате такой последовательности действий, зависящей от параметра  $\varepsilon$ , получим словарь  $H \subseteq Z$  объемом  $h = \text{card } H$  элементов. Используя словарь  $H$  для обработки каждого из эталонных описаний  $Z^i$ , в соответствии с методом tf-idf вычислим вектор

$$g^i[Z^i] = (g_1, g_2, \dots, g_k, \dots, g_h), \quad (4)$$

где  $g_k$  определяются по формулам (1), (2).

### Распознавание на основе сопоставления векторов tf-idf

В результате применения лингвистического подхода для множества структурных признаков получена новая форма описания, где каждый эталон  $Z^i$  представляет собой вектор (4) из  $h$  элементов:  $Z^i \rightarrow g^i$ . На этапе распознавания объекта, изображенного структурным описанием в виде вектора весов tf-idf, выполняем следующие действия. Вычисляем векторное представление  $g$  для распознаваемого изображения и последовательно находим его расстояние до каждого из эталонных векторов. В результате получим вектор  $\beta_j = \rho(g, g^j)$ ,  $j = \overline{1, J}$  значений расстояний. Объект отнесем к классу  $j$  в соответствии с правилом

$$j = \arg \min_{d=1, \dots, J} \beta_d, \quad (5)$$

т.е. выбирается минимум среди компонентов  $\beta_j$ .

Схема процесса распознавания изображена на рис. 1.

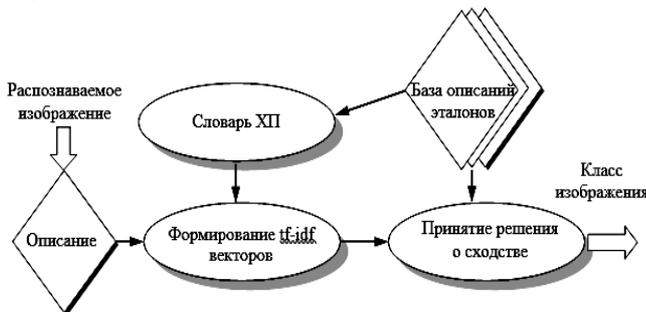


Рис. 1. Схема распознавания

Мера различия векторов-дескрипторов SURF, используемая в (3) при построении словаря  $H$  и при определении встречаемости термина в ходе вычисления (1) имеет вид евклидова расстояния

$$\rho(x, y) = \sqrt{\sum_{i=1}^{64} (x_i - y_i)^2}, \quad (6)$$

где  $x, y$  – значения дескриптора SURF для сопоставляемых элементов описания.

Наряду с (6) применимы также нормированный коэффициент корреляции, квадрат евклидова расстояния, взвешенное евклидово расстояние, хеммингово расстояние, расстояние Чебышева и др. [1].

### Результаты компьютерного моделирования

Для проведения экспериментов разработан программный модуль на языке Java. Вычисление ХП методом SURF осуществлено с применением библиотеки компьютерного зрения OpenCV [5] для базы из  $b=5$  изображений имен брендов фирм (рис. 2). На рис. 3 в виде квадратиков показаны координаты множества ХП для изображения рис. 2, д, а на рис. 4 – пример распознаваемого изображения, которое получено увеличением масштаба и поворотом эталонного изображения бренда Lenovo. Табл. 1 содержит количества ХП в эталонных описаниях изображений рис. 2.

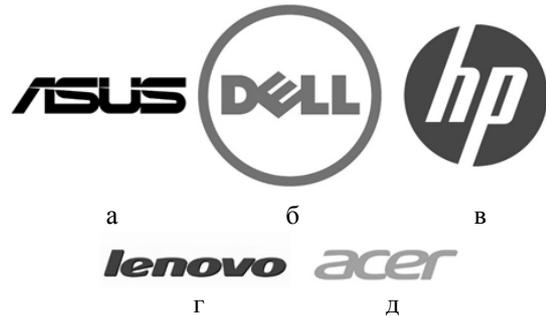


Рис. 2. Изображения брендов из базы (а-д)

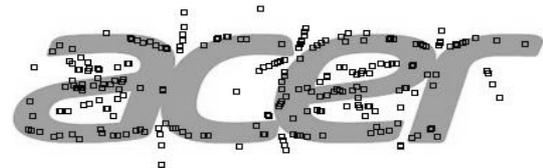


Рис. 3. Координаты ХП для изображения рис. 2, д



Рис. 4. Расознаваемое изображение

Таблица 1  
Количества ХП в описаниях базы изображений рис. 2

Бренды	Acer	Asus	Dell	HP	Lenovo
Число ХП	237	352	638	405	367

Общее количество ХП рассмотренной базы равно 1999. Для формирования словаря в (3) выбран порог  $\varepsilon=0,6$ . Величина порога определена из соображения, чтобы распределение ХП по частоте встречаемости было приблизительно одинаковым, что исключает значительное преобладание низкочастотных или высокочастотных ХП. За счет этих действий размер словаря будет не слишком большим, что приводит к замедлению вычислений, и в то же время не слишком маленьким, что может ухудшить качество распознавания.

В результате обработки базы мы получили словарь объемом 42 ХП, применяя который вычислили и нормализовали вектора tf-idf для эталонов базы,

матрица евклидовых расстояний между эталонами как векторами tf-idf представлена в табл. 2.

Таблица 2

Матрица расстояний между эталонами

Бренды	Acer	Asus	Dell	HP	Lenovo
Acer	0	0,8579	0,6952	1,3414	0,7991
Asus	0,8579	0	1,0965	1,3514	0,8126
Dell	0,6752	1,0965	0	1,3961	1,0315
HP	1,3414	1,3514	1,3961	0	1,3448
Lenovo	0,7991	0,8126	1,0315	1,3448	0

Как видим из табл. 2, в пространстве tf-idf наблюдаются существенные различия между структурными описаниями разных изображений брендов, что свидетельствует о высокой степени различимости эталонов базы. Используя построенный словарь ХП, вычисляем расстояния между распознаваемым изображением и множеством эталонов (табл. 3).

Таблица 3

Расстояния между изображением и эталонами

Бренды	Acer	Asus	Dell	HP	Lenovo
Расст.	0,837	0,808	1,061	1,215	0,297

Как видим из табл. 3, минимальное значение евклидового расстояния соответствует эталонному изображению бренда Lenovo. Его значение 0,297 значительно меньше, чем ближайшее по величине расстояние 0,8087, что подтверждает надежное распознавание с применением разработанного подхода.

## Выводы

**Основным результатом** исследований настоящей статьи есть построение и результативное применение лингвистических характеристик на основе схемы tf-idf в задачах структурного распознавания изображений. **Научная новизна** проведенного исследования состоит в разработке метода структурного распознавания изображений путем применения схемы лингвистического анализа (час-

тоты встречаемости структурных признаков tf-idf) для формирования и сопоставления сжатого векторного представления множества дескрипторов в структурном описании изображения. Переход к векторному представлению существенно увеличивает быстродействие распознавания. **Практическая ценность работы** – получение прикладной программной модели для структурного распознавания на основе лингвистического подхода и проверка её работоспособности в конкретных примерах баз изображений. **Перспективой применения** лингвистического анализа на множестве описаний в виде характерных признаков изображений может быть использование разнообразия вариантов построения и обработки моделей весов tf-idf, а также выбор оптимального способа построения словаря признаков для обеспечения высокого качества распознавания в произвольной базе изображений.

## Список литературы

1. Гороховатский В.А. Структурный анализ и интеллектуальная обработка данных в компьютерном зрении: монография / В.А. Гороховатский. – Х.: Компания СМИТ, 2014. – 316 с.
2. Bay H. Surf: Speeded up robust features / H. Bay, T. Tuytelaars, L. Van Gool // European Conference on Computer Vision. – 2006. – P. 404-417.
3. Szeliski R. Computer Vision: Algorithms and Applications / R. Szeliski. – London: Springer, 2010. – 979 p.
4. Маннинг К.Д. Введение в информ. поиск / К.Д. Маннинг, П. Рагхаван, Х. Шютце. – М.: Вильямс, 2011. – 528 с.
5. Using SIFT/SURF for Object Recognition in OpenCV Java [Электронный ресурс] / Dummy's codes. – Режим доступа : <http://dummyscodes.blogspot.com/2015/12/using-siftsurf-for-object-recognition.html/>.
6. Солтон Дж. Динамические библиотечно-информационные системы / Дж. Солтон. – М.: Мир, 1979. – 558 с.

Поступила в редколлегию 24.02.2016

**Рецензент:** д-р техн. наук, проф. Е.П. Пуятин, Харьковский национальный университет радиоэлектроники, Харьков.

## АНАЛІЗ ЛІНГВІСТИЧНИХ ХАРАКТЕРИСТИК СТРУКТУРНИХ ОПИСІВ У РОЗПІЗНАВАННІ ВІЗУАЛЬНИХ ОБ'ЄКТІВ

М.М. Грідін, В.О. Гороховатський

Обговорюються питання зіставлення описів при розпізнаванні візуальних об'єктів у системах комп'ютерного зору. Запропоновано технологію структурного розпізнавання, яка базується на методі tf-idf, що застосовується в математичній лінгвістиці та інформаційному пошуку. У результаті виконано перетворення опису з множини характерних ознак зображення до вектору лінгвістичних характеристик, що значно знижує обчислювальні витрати. Представлені результати експериментів, що підтверджують працездатність і результативність методу.

**Ключові слова:** комп'ютерний зір, база зображень, структурне розпізнавання зображень, характерні ознаки, схема зважування tf-idf, векторний опис, міра подібності, релевантність.

## ANALYSIS OF THE LINGUISTIC CHARACTERISTICS STRUCTURAL DESCRIPTION IN RECOGNIZING VISUAL OBJECTS

M.M. Gridin, V.A. Gorokhovatsky

The questions of matching the descriptions at recognizing visual objects in computer vision systems are discussed. The technology of structural recognition which based on the method of tf-idf which used in mathematical linguistics and information retrieval is offered. As a result of perform the conversion of descriptions from variety of the characteristic features of image to the vector of linguistic characteristics that significantly reduces the computational cost. The results of experiments confirming the efficiency and effectiveness of the method are presented.

**Keywords:** computer vision, image database, structural image recognition, features, weighting scheme tf-idf, vector description, similarity measure, relevancy.