

Поиск изображений по визуальному содержанию в графических базах данных и сети Интернет

И.Е. Десятников

Аннотация. В статье представлена система поиска изображений по визуальному содержанию. Архитектуру баз данных для задачи поиска предлагается построить с помощью алгебры полных и замкнутых групп – элементов теории активного восприятия, разработанной на кафедре «Вычислительные системы и технологии» НГТУ им Р.Е. Алексеева. Разработаны алгоритмы поиска не только исходных, но также зашумлённых, отредактированных и испорченных изображений, разработан поиск похожих изображений, который по достоверности и производительности может конкурировать с существующими методами поиска изображений в сети Интернет. Приведены экспериментальные результаты, полученные для базы данных произвольных изображений размером около 10 млн. объектов.

Ключевые слова: теория активного восприятия, CBIR, распознавание, компьютерное зрение.

Введение

За последние годы объем мультимедиа-данных вырос в несколько раз. Все больше людей используют поиск изображений в сети Интернет для оформления своих презентаций, публикаций и просто для удовлетворения любопытства. При этом классифицированной является лишь незначительная часть этих изображений (в основном, это касается коммерческих баз данных).

По некоторым подсчетам, количество изображений в сети Интернет на сегодняшний день составляет более 10^{12} и с каждым годом увеличивается. Однако вся эта информация бесполезна без точного, удобного и быстрого поиска по ней.

Задача поиска изображений в последние годы активно развивается и привлекает всё больше исследователей. Для проверки гипотезы о росте количества публикаций автор выполнил простое упражнение. Проводился поиск публикаций, содержащих в своей теме фразу «Image

Retrieval» с помощью Google Scholar [1]. Результат исследования показан на Рис. 1.

Активное развитие исследований в области поиска изображений связано, в первую очередь, с резко возросшим объемом обрабатываемой информации, с которой человек уже не справляется. Поэтому и повышается роль автоматизированных систем для решения задачи поиска.

Любая система поиска изображений строится с позиций теории распознавания образов. Проблема распознавания (классификации) является весьма общей, хотя она и возникла в связи с решением частных задач – задач распознавания фигур (цифр, букв, изображений), звуков (речи, шума), диагностики заболеваний или неисправностей и т.п. Распознавание представляет первую и важную ступень обработки информации, получаемой нами при помощи органов чувств и приборов.

Распознавание образов – это отнесение исходных данных к определенному классу путем выделения существенных признаков или свойств, характеризующих эти данные (из общей массы

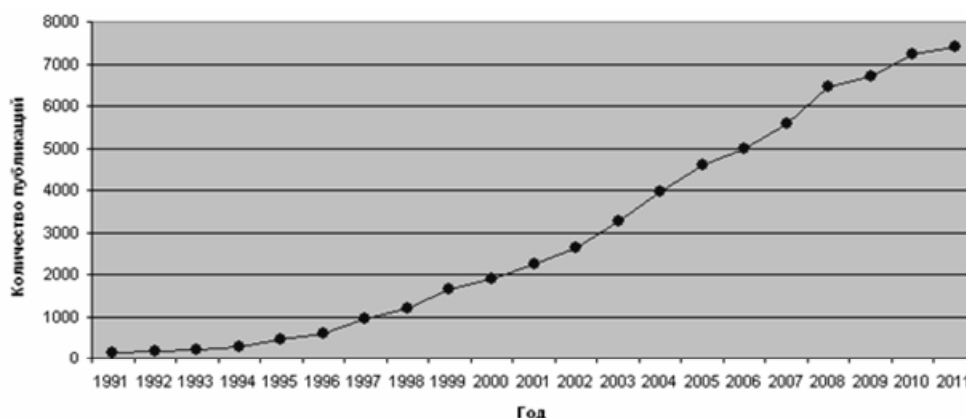


Рис. 1. Количество публикаций по теме поиска изображений

несущественных деталей). Под классом образов понимается некоторое подмножество, определяемое рядом свойств, общих для всех элементов этого класса. То общее, что объединяет объекты в класс, называют эталоном.

В качестве эталона выбирают образ, обладающий минимальной изменчивостью значимых признаков на множестве возможных представлений данного образа. Предполагается, что образу (эталону) соответствует компактное множество точек в пространстве признаков (гипотеза компактности), образующих разделимые классы. Однако шумы, помехи (организованные и неорганизованные), структурные вариации одного и того же представителя класса приводят к значительному увеличению указанного объема и, как следствие, перекрытию классов, а значит, к снижению достоверности классификации. В результате проблема принятия решения замыкается на проблеме формирования системы признаков, позволяющей сформировать такой эталон, который обеспечивал бы условия компактности.

В общем случае, распознавание изображений – это трех этапный информационный процесс нахождения системы признаков объекта исследования и принятия классификационного решения методами теории принятия решения [2] (Рис. 2).

«Узким» местом всех существующих методов распознавания является задача выбора признаков. Кроме того, сложность задач выбора признаков и классификации резко возрастает и становится проблематичной в условиях априорной неопределенности. Для многих методов

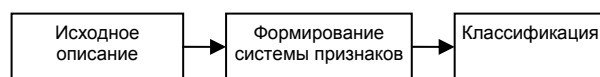


Рис. 2. Схема информационных преобразований в системах распознавания

процесс разделения входного множества изображений на классы представляется трудоемким либо неоптимальным. Эта задача является очень важной, ведь если мы сможем точно (в виде конкретных признаков) сформулировать то общее, что объединяет объекты в класс, то задача распознавания сведется к сравнению признаков предъявляемых объектов с заранее известными. Таким образом, трудно формализовать процесс нахождения эталона класса. Эта задача усложняется, когда изображения подвергаются воздействию шумов и помех.

Со всеми вышеперечисленными проблемами предполагается справиться, используя теорию активного восприятия [3-5].

1. Алгоритм выделения классов изображений

В [5] был предложен метод вычисления признаков изображений. Кратко данный метод состоит из двух этапов (Рис. 3):

1. проводится дихотомия изображения как множества M на n равных частей и подсчитывается визуальная масса каждой из них – получаем матрицу $||m_{ij}||$ изображения;
2. на матрицу $||m_{ij}||$ накладываются фильтры пространственного дифференцирования – получаем 15-мерный вектор μ .

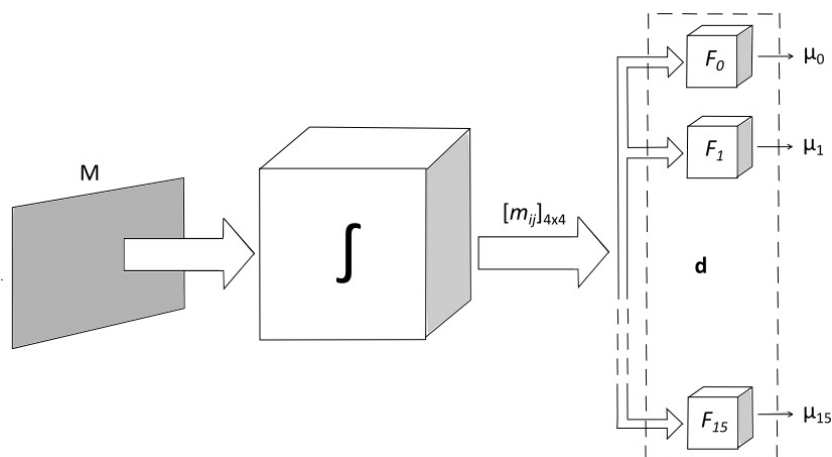


Рис. 3. Схема информационных преобразований

Полученный вектор и есть характеристика исходного изображения.

По данному алгоритму для всех изображений вычисляется вектор и формируется база данных, содержащая изображения и их вектора. Для каждого изображения достаточно вычислить его вектор всего один раз – когда изображение попадает в коллекцию. Но в случае больших баз изображений (несколько миллионов и больше) задача поиска в такой коллекции становится весьма сложной. Необходимо разделить входное множество изображений на подмножества (классы изображений).

Предлагается сделать центрами классов изображений «полные» и «замкнутые» группы [3, 4]:

- полные (алгебраические) группы – P_{ni} , образованные на тройках операторов (V_i, V_j, V_k) , для которых справедливы соотношения: $V_i + V_j + V_k \equiv e_1$ – единица; $V_i V_j V_k$ – образ (на операции умножения) и описание группы P_{ni} ;

- замкнутые (алгебраические) группы – P_{si} , образованные на четверке операторов (V_i, V_j, V_p, V_m) , где $(V_i, V_j, V_k) \in P_{ni}$, $(V_n, V_m, V_k) \in P_{nj}$, с описанием $V_i V_j + V_p \bar{V}_m$ (необходимое число инверсий операторов нечетно) и единицей – $V_i + V_j + V_p + \bar{V}_m \equiv e_1$.

Множества P_{ni} и P_{si} конечны и имеют мощности 35 и 105 соответственно (отметим, что 36-м элементом множества полных групп является оператор V_0). Семейства этих групп допускают их использование на этапах принятия решений и понимания анализируемого изображения.

В соответствии с изложенным пространство классов – это фактически пространство эталонных изображений, представленное в 15-мерном признаковом пространстве. В этом пространстве подмножества эталонных изображений в силу гипотезы компактности образуют классы близких эталонных изображений – точек подмножества, для которых близость необходимо должна определяться некоторым евклидовым расстоянием

$$r = \sqrt{\sum_i (x_i - x_i^0)^2},$$

где $\{x_i\}$ – координаты наблюдаемой точки – эталонного изображения; $\{x_i^0\}$ – координаты некоторого эталона для данного класса эталонных изображений.

Такой подход естественен, правомерен и является стандартным подходом в теории распознавания образов. Однако он требует знания значения r и априорного знания того, что есть эталон класса. В качестве последнего должно выступать некоторое обобщенное изображение на подмножестве эталонных изображений, позволяющее успешно решить проблему классификации: «свой – чужой». Такое обобщенное изображение называется **выработанным (обобщенным) эталоном** класса эталонных изображений [4].

В такой постановке естественен вопрос о процедуре формирования выработанного эталона, сходимости процедуры (т.е. ее конечности и скорости сходимости), достоверности подмножества эталонных изображений, участ-

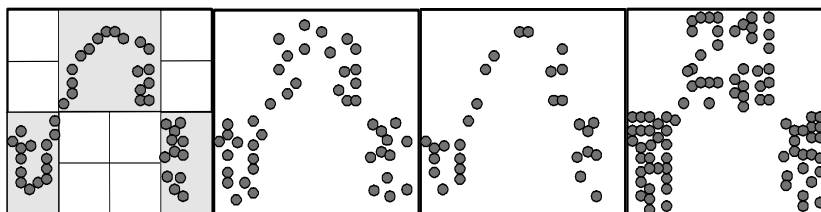
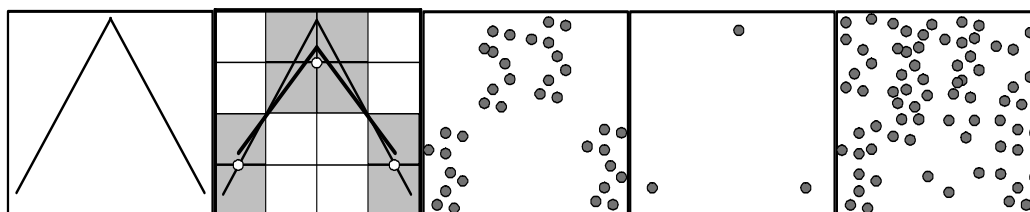
Рис. 4. Подмножество изображений точечного класса врождённого эталона V_6 

Рис. 5. Подмножество изображений класса эталона

вующих в формировании выработанного эталона. В теории распознавания образов в роли процедуры формирования выработанного эталона выступает процесс обучения, который необходимо должен быть конечен, а его результаты необходимо должны сходиться на некотором множестве подмножеств классов.

В теории активного восприятия пространством врожденных эталонов являются:

1. пространство операторов $V = \{V_i\}$; поскольку множество V имеет мощность, равную пятнадцати, то для полного представления необходимо евклидово пространство E^{15} ;

2. пространство полных групп $P_n = \{P_{ni}\}$; поскольку множество P_n имеет мощность 35 (или 36, если учитывать единственную полную группу (удовлетворяющую соответствующему определению) – оператор V_0), для полного представления необходимо евклидово пространство E^{35} ;

3. пространство замкнутых групп $P_s = \{P_{si}\}$; поскольку множество P_s имеет мощность 105, то для полного представления необходимо евклидово пространство E^{105} .

Таким образом, имеем три типа пространств эталонов.

Пусть в базе данных имеется эталон, например, образ оператора V_6 , который относится к множеству врожденных эталонов. Тогда на любых вариантах наблюдаемых изображений $\{M_k\}$ (Рис.4), для которых выполняется $\mu_6 = \mu_0$, решение однозначно, т.е. достоверно.

Характерная особенность такого класса изображений отражается в составе полутоновой матрицы $\|m_{ij}\|$, непосредственно отображающей структуру врожденного эталона (вариант изображения без помех). Такое подмножество изображений образует **точечный класс** в пространстве классов. Естественно, что если множества $\{V_i\}$, $\{P_{ni}\}$, $\{P_{si}\}$ – врожденные эталоны, то получаем базу врожденных точечных классов, обеспечивающую одномоментное узнавание изображения.

Пусть в качестве эталона выступает эталонное изображение, например, буква «Л» (Рис. 5). Описанием этого изображения в базисе $\{V_i\}$ является оператор V_6 . Тогда любое изображение (с помехами либо без них), удовлетворяющее условию $\mu_6 \gg \mu_i \quad \forall i \neq 0$ (близко к μ_0 по сравнению с остальными), относится к классу эталонного изображения. Такое подмножество изображений образует «облако» в пространстве классов, называется **размытым классом** (классом эталона) в пространстве классов и составлено из подмножества точечных классов, где к точечному классу эталона относится эталонное изображение, представленное в базисе врожденных эталонов.

Рассмотрим вопрос перехода от одного точечного класса врожденного эталона к другому точечному классу при воздействии на изображение равномерных шумов и помех – такая ситуация довольно часто встречается в реальном

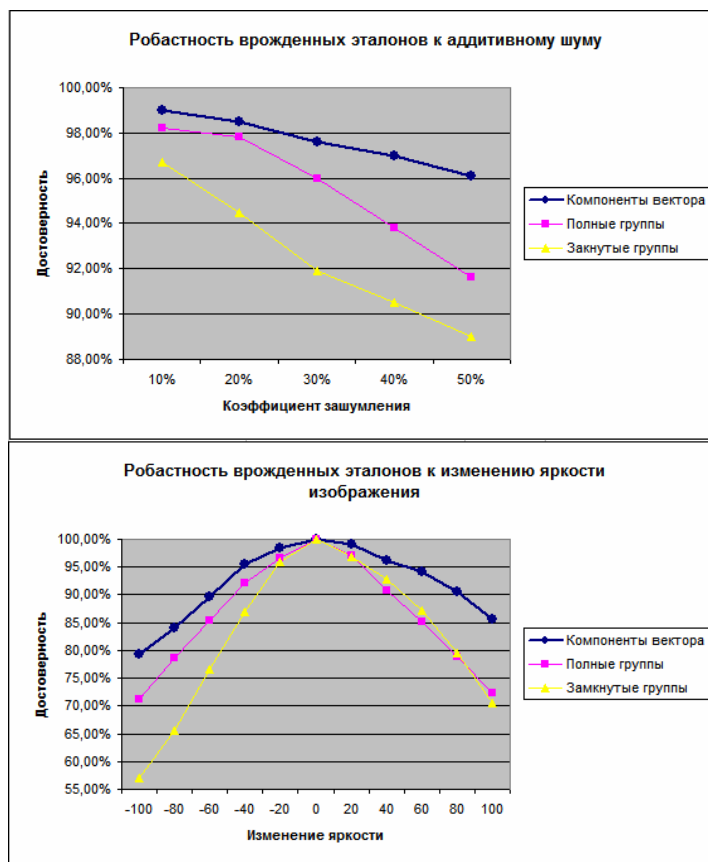


Рис. 6. Робастность врожденных эталонов к воздействию аддитивных шумов и изменению яркости изображения

мире, и разрабатываемая поисковая система должна быть робастна к такого рода помехам.

В рамках данной работы было проведено исследование робастности множеств врожденных эталонов $\{V_i\}$, $\{P_{ni}\}$, $\{P_{si}\}$ к воздействию равномерных шумов на изображение, а также к изменению яркости изображения. На Рис. 6 видно, что наиболее робастным к данному рода изменениям является множество врожденных эталонов $\{V_i\}$.

Исходя из вышеизложенного материала, можно сделать вывод, что поисковая система должна состоять из иерархии классов (Рис. 7):

- на первом уровне иерархии множество врожденных эталонов $\{V_i\}$ как самое робастное к воздействию шумов и помех;
- на втором уровне иерархии множество врожденных эталонов $\{P_{ni}\}$;
- на третьем уровне иерархии множество врожденных эталонов $\{P_{si}\}$.

Такая архитектура базы данных дает возможность хранить и эффективно выполнять поиск более чем на 10^{12} объектах.

2. Тестирование разработанных алгоритмов поиска

Описанные в данной статье алгоритмы были реализованы в виде законченного программного продукта. Ниже представлены результаты разработанных алгоритмов для конфигурации ПК AMD Phenom II X4 945 3 ГГц, 2Гб ОЗУ и размера базы данных около 10^7 объектов. База данных содержит изображения произвольного размера и содержания.

Поиск незашумленных изображений выполняется в среднем за 100 мс. Отметим, что скорость распознавания простого изображения мозгом человека составляет по разным оценкам 40-250 мс [6]. Это время необходимо кратковременной памяти человека, чтобы сравнить

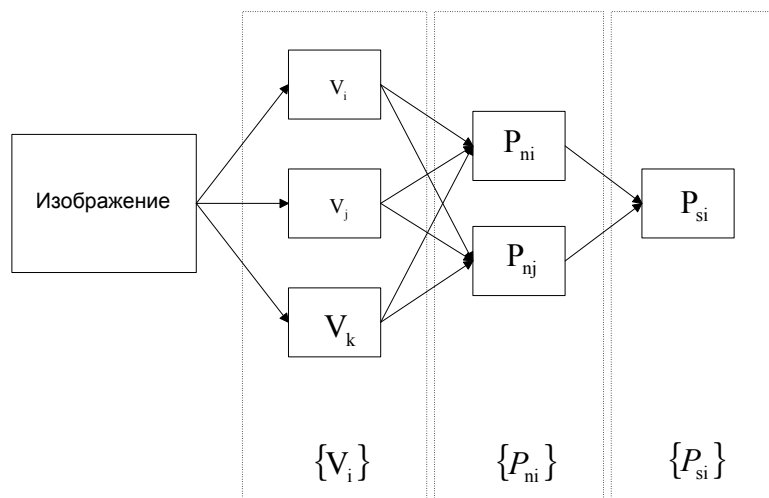


Рис. 7. Иерархия операторов и групп в поисковой системе

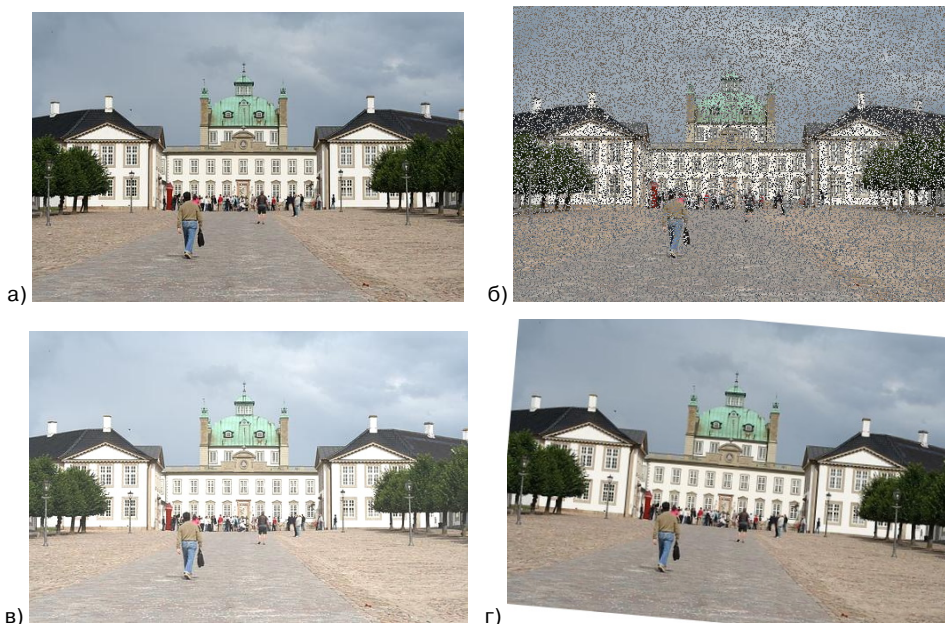


Рис. 8. Примеры зашумлённых и отредактированных изображений (а) исходное изображение, (б) аддитивный шум с вероятностью 30%, (в) яркость изображения +30, (г) поворот изображения на 5 градусов

свое содержимое с запасами долговременной. Как видим, разработанный алгоритм даже превосходит возможности человеческого мозга.

Для поиска зашумлённых изображений, а также изображений с измененной яркостью поворотом на небольшой угол (Рис. 8), требуется более широкая выборка изображений из базы. Поэтому данный вид поиска характеризуется средней скоростью поиска (среднее время поиска): 500 мс. Результаты работы алгоритма для разных типов помех и искажений представлены на Рис. 9.

Также имеется возможность выполнять поиск не зашумлённых, но поврежденных изображений, например, часть изображения потеряна или стерта. При этом может быть стерто до 75% исходного изображения. Пример такого изображения представлен на Рис. 10.

В данном случае, чтобы успешно выполнить задачу поиска, необходимо перейти на следующий уровень пирамиды разрешения, т.к. верхний уровень поврежден. Время поиска больше, чем в случае «грубого» поиска, и составляет 250 мс.

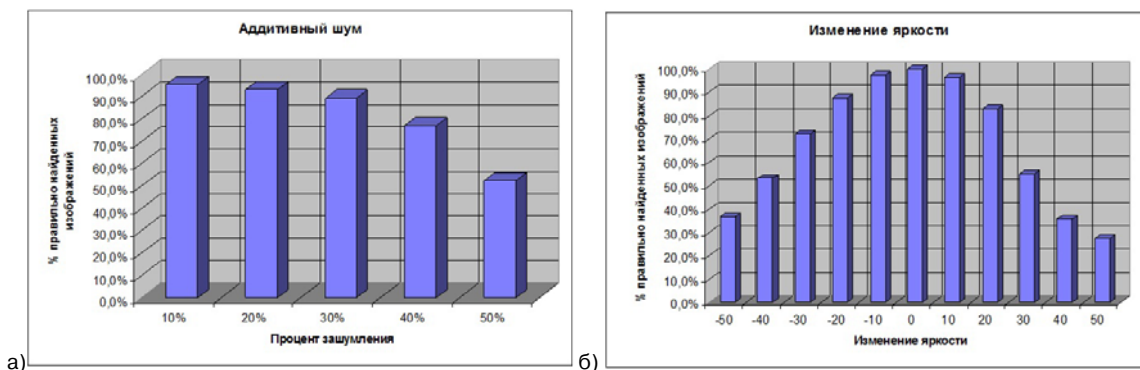


Рис. 9. Результаты по алгоритму «Помехоустойчивый поиск изображения»

- а) тестирование с воздействием аддитивного шума
- б) тестирование с изменением яркости исходного изображения



Рис. 10. Пример поврежденного изображения

Поиск похожих изображений предназначен для поиска набора похожих изображений. На выходе алгоритма получаем некоторое количество изображений, похожих на искомый объект, причем изображения упорядочены по степени сходства (критерием служит Евклидово расстояние между векторами). Среднее время поиска 1 с. Пример поиска показан на Рис. 11.

Заключение

В данной работе представлена информационная модель системы поиска изображений по визуальному содержанию. Разработан алгоритм формирования классов изображений на основе множеств выработанных эталонов $\{V_i\}$, $\{P_{ni}\}$, $\{P_{si}\}$. Проведено исследование устойчивости полных и замкнутых групп к воздействию равномерных шумов и изменению яркости изображения. Обоснована возможность их применения для формирования центров классов изображений.

На основе данной модели можно строить базы видеоданных с общим количеством изображений более 10^{12} объектов. В отличие от ранее опубликованных алгоритмов, обеспечивается более низкая вычислительная сложность и возможность работать с зашумлёнными изображениями.

Разработанные методы поиска показывают отличные результаты по производительности и достоверности, несмотря на то, что тесты проводились на вполне обычном персональном компьютере. Анализ результатов позволяет сделать вывод, что разработанная поисковая система успешно справляется с поставленной задачей поиска не только исходных, но и зашумлённых и отредактированных изображений.

К сожалению, из-за отсутствия единой базы для тестирования невозможно сравнить результаты с возможностями других систем. К тому же большинство систем имеют коммерческий характер и обычно имеют узкую специализацию. Автор считает, что полученные результаты



Рис. 11. Пример поиска изображений

а) исходное изображение

б) результаты выполнения поиска похожих изображений

соответствуют реалиям сегодняшнего дня, и данный метод может применяться для построения систем контроля и поиска изображений в базах данных, в том числе и в сети Интернет.

Литература

1. Google Scholar [Электронный ресурс]. – Режим доступа: <http://scholar.google.ru/> (дата обращения: 19.10.2012).
2. Утробин В.А. Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений: учеб. пособие; НГТУ. – Нижний Новгород. – 2001. – 234 с.
3. Утробин В.А. Физические интерпретации элементов алгебры изображений // Успехи физических наук (УФН). – 2004. – Т. 174, № 10. – С. 1089-1104.
4. Утробин, В.А. Компьютерная обработка изображений. Принятие решений в пространстве эталонов: учеб. пособие; НГТУ. – Нижний Новгород. – 2004. – 221 с.
5. Десятников И.Е., Утробин В.А. Алгоритмы поиска изображений в базах видеоданных // Компьютерная оптика. – 2011. – Т. 35, № 3. – С. 416-422.
6. Демидов В.Е. Как мы видим то, что видим. М.: Знание. – 1987. – 240 с.

Десятников Игорь Евгеньевич. Аспирант, руководитель группы разработки в компании ООО «МФИ Софт». Окончил Нижегородский государственный технический университет им. Р.Е. Алексеева в 2010 году. Автор двух печатных работ. Область научных интересов: компьютерное зрение, обработка изображений, поиск изображений. Email: digore@bk.ru