

АЛГОРИТМЫ ПОИСКА ИЗОБРАЖЕНИЙ В БАЗАХ ВИДЕОДАНЫХ

Десятников И.Е., Утробин В.А.

Нижегородский государственный технический университет им. Р.Е. Алексеева

Аннотация

В статье рассмотрены существующие на сегодняшний день методы поиска изображений в базах видеоданных. Алгоритмы поиска изображений предлагается построить с позиции теории активного восприятия, разработанной на кафедре «Вычислительные системы и технологии» НГТУ им Р.Е. Алексеева. Достоинство данного подхода заключается в низкой вычислительной сложности, а значит, в высоком быстродействии. Каждое изображение в разработанной системе поиска представляется в виде 15-мерного вектора. Разработаны алгоритмы поиска не только исходных изображений, но также зашумлённых, отредактированных, испорченных, изображений с любым углом поворота от исходного, разработан поиск похожих изображений, который по достоверности и производительности может конкурировать с существующими методами поиска изображений в сети Интернет.

Ключевые слова: поиск изображений, теория активного восприятия.

Введение

С каждым годом объём обрабатываемой информации неуклонно растёт, что требует формализации и последующей алгоритмизации процессов, ранее выполнявшихся вручную. Одним из ключевых понятий в автоматической обработке информации является поиск объектов определённого класса, в частности, поиск изображений в больших базах данных.

По некоторым подсчётам, количество изображений в сети Интернет на сегодняшний день составляет более 10^{12} и с каждым годом увеличивается в геометрической прогрессии. Выполнять поиск вручную на таком количестве не представляется возможным.

Можно уверенно прогнозировать, что по мере расширения доступа к электронным архивам изображений и видео будет возрастать значимость поиска изображения по его содержанию. Для обоснования этого прогноза достаточно привести тот факт, что значительная часть информации, поступающей из окружающего мира, воспринимается нами именно в зрительной форме.

1. Обзор методов поиска изображений

Существующие методы поиска изображений в базах данных можно разделить на два вида:

- 1) поиск по текстовому описанию;
- 2) поиск по визуальному содержанию.

До недавнего времени традиционным считался поиск визуальной информации, опирающийся на индексирование текстовых описаний, ассоциированных с изображением. Однако поиск по текстовой информации имеет ряд недостатков.

Во-первых, необходимо, чтобы оператор вручную пометил все изображения ключевыми словами. Каждый день в глобальной сети Интернет появляется огромное количество новых изображений, и, чтобы их проиндексировать, нужны громадные человеческие ресурсы.

Во-вторых, неоднозначность соответствия между визуальным содержанием и текстовым описанием снижает показатели точности и полноты поиска.

Наконец, ещё один недостаток заключается в существовании изображений, которые вообще трудно описать словами (очевидный пример – абстрактные картины).

Совершенно иным методом решения задачи поиска в базе изображений является поиск по визуальному подобию *Content Based Image Retrieval (CBIR)*. Данный термин впервые был использован в 1992 г. для обозначения поиска изображения в базе данных на основе анализа таких характеристик, как цвета, текстура и очертания его элементов. Со временем содержание термина расширилось, и теперь с его помощью обозначают достаточно широкий спектр технологий поиска изображений, основанных на анализе содержимого изображений.

В самом общем виде *CBIR*-система работает подобно любой другой поисковой системе – в два этапа. На первом этапе индексирования каждое изображение описывается и заносится в базу данных. Только в этом случае систему интересуют не ключевые слова или имена файлов, а определённые параметры самого изображения, анализируемые с помощью специальных алгоритмов. Обычно это уже названные выше параметры цвета, текстуры и очертаний. Полученные данные сохраняются в индексной базе. После этого можно вести поиск по определённым значениям таких параметров, сравнивать их между собой или с представленным системе изображением. Это уже второй этап – нахождение в базе изображений с близкими признаками – другими словами, визуально похожих. На этапе поиска свойства одной картинке сравниваются с аналогичными данными других изображений, хранящихся в индексной базе.

Для решения задачи поиска предложено множество различных методов выделения признаков изображений. В качестве признаков предлагается брать цветовые характеристики [1-2], текстурные характеристики [3-4], дескрипторы формы [5]. В качестве меры сходства или отличия чаще всего предлагается брать различные метрики (сравнение мер рас-

стояния, часто применяемых для поиска изображений, проведено в [6]).

Из обзора существующих методов можно сделать вывод, что задача поиска изображений в последние годы активно развивается и привлекает всё больше исследователей. Это связано, в первую очередь, с резко возросшим объёмом обрабатываемой информации, с которой человек уже не справляется. Поэтому и повышается роль автоматизированных систем, которые будут выполнять задачу поиска.

Заметим, что большинство существующих алгоритмов ориентировано для поиска незашумлённых изображений, т.е. поиск выполняется только для исходных изображений. В реальности же изображения могут быть зашумлены, немного отредактированы, повернуты на какой-то угол, часть изображения может быть просто утеряна или закрыта пятном.

Цель работы: разработать алгоритмы для выполнения поиска исходных, а также зашумлённых и отредактированных изображений. Итогом должна стать поисковая программная система для выполнения поиска в больших базах данных (от 10^4 объектов).

Алгоритмы поиска объекта в базе данных предлагается разработать с позиции теории активного восприятия, разработанной на кафедре «Вычислительные системы и технологии» НГТУ им. Р.Е. Алексеева [7 - 8].

2. Элементы теории активного восприятия с позиций построения поисковой системы изображений

Проблематике раскрытия априорной неопределённости изображения как объекта исследования посвящена теория активного восприятия. Выделим необходимые для дальнейшего изложения положения данной технологии. Во-первых, определим объект исследования.

Изображение M – это множество пространственно-распределённых и координатно-упорядоченных элементов (пикселей), каждый из которых в фиксированный момент времени имеет неотрицательное целое значение в замкнутой области определения G (поле зрения):

$$M = \begin{cases} \mu(x, y), & \text{если } (x, y) \in G \subset R^2, \\ 0, & \text{если } (x, y) \notin G, \end{cases}$$

где $\mu(x, y)$ – входная функция изображения.

Поскольку объект исследования ограничен полем зрения и принадлежит множеству положительных действительных чисел, то всё изображение в целом и любая его подобласть определения G_i допускают преобразование проектирования (в теории активного восприятия данное преобразование определено Q -преобразованием) (рис. 1):

$$m(G_i) = \iint_{(x,y) \in G_i} \mu(x, y) dx dy \quad \forall i, \tag{1}$$

или в дискретном варианте:

$$m(G_i) = \sum_{(k,n) \in G_i} \mu[k, n] \quad \forall i.$$

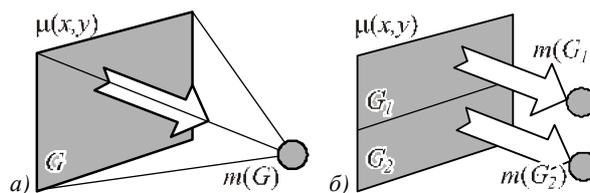


Рис. 1. Реализация Q -преобразования по всей области изображения (а) и по двум её подобластям (б)

Преобразование (1) наделено следующими свойствами:

1. Применимо к любому объекту исследования типа изображений в любом диапазоне частот.
2. С позиций математики реализует отображение в действительное абсолютно гладкое пространство (многообразие), к которому можно применять бесконечное число дифференцирований, позволяющих выявлять топологию этого пространства (структуру анализируемого изображения).
3. С позиций физики имеет смысл массы (в теории активного восприятия определено визуальной массой).
4. Наделено свойством минимально возможной вычислительной сложности, поскольку для оцифрованных изображений это единственная операция суммирования.

Поскольку (1) – преобразование интегрирования, то существует обратное ему – преобразование внешнего дифференцирования. Вместе они образуют композицию, раскрывающую неопределённость объекта исследования – изображения M в поле зрения. При этом сначала реализуется преобразование интегрирования, к результатам которого применяется операция дифференцирования по всей области определения либо любой его подобласти без реализации операции свёртки, используемой в известных системах обработки изображений.

Операция дифференцирования – это отношение между двумя визуальными массами сравниваемых подобластей поля зрения:

$$\Delta m(G) \equiv \mu_i = m(G_1) - m(G_2), \tag{2}$$

где $G = G_1 \cup G_2$ – дихотомия области G на пару непересекающихся подобластей, для которых значения $m(G_i)$ определены по (1) (рис. 1).

Реализовать преобразование (2) можно с помощью маски F как покрытия изображения M . Результатом являются следующие возможные значения:

$$\mu_i = \begin{cases} > 0, & \text{если } m(G_1) > m(G_2), \\ < 0, & \text{если } m(G_1) < m(G_2), \\ = 0, & \text{если } m(G_1) = m(G_2). \end{cases}$$

Отношение (2) позволяет выявить асимметрию и симметрию двух подобластей анализируемого изображения по одному возможному направлению (на рис. 1 выявляется отношение по направлению y декартовой системы координат). Для того, чтобы выявить полное множество булевых функций (т.е. бинарных отношений по (2) на множестве вариантов

дихотомий области G) в соответствии с положениями булевой алгебры, необходимо 2^{2^n} дихотомий, где n – число переменных. Для двумерного изображения $n=2$, поэтому число дихотомий равно 16 (включая нулевую). В результате имеем 15 независимых направлений в многомерном векторном пространстве. Каждому направлению соответствует компонента μ_i , а множеству направлений – 15-мерный вектор $\vec{\mu} = (\mu_1 e_1, \mu_2 e_2, \dots, \mu_{15} e_{15})$, описывающий изображение в поле зрения или любой его подобласти. Компонента μ_0 , соответствующая нулевому направлению, является визуальной массой всего изображения.

Конструктивно множество масок, покрывающих область определения изображения G , определено на системе базисных функций Уолша системы Хармута и упорядочено на двумерной решётке [7-8]. Маски-покрытия (их нумерация условна) построены на квадратном (либо прямоугольном) клеточном пространстве (рис. 2).

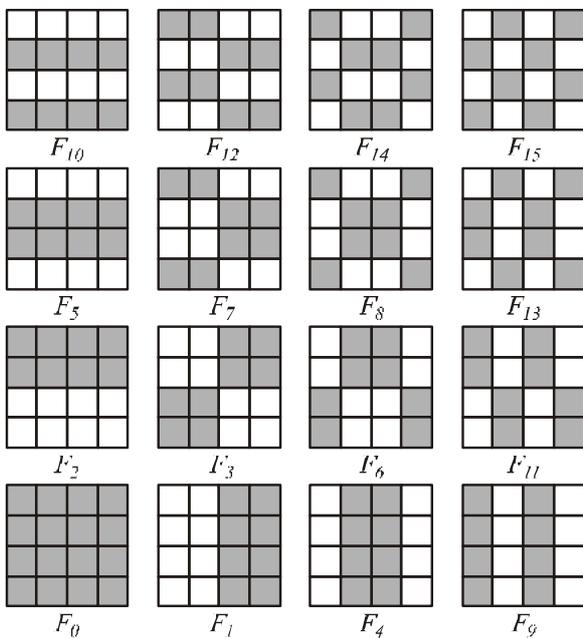


Рис. 2. Маски-покрытия $\{F_i\}$

Использование функций Уолша в силу их бинарности обеспечивает простоту анализа отношений на множестве $m(G_i)$, упорядоченном на матрице $[m_{ij}]_{4 \times 4}$, так как используются простейшие операции – сложение (1) и вычитание (2). Эта система преобразований имеет минимально возможную вычислительную сложность, в отличие от стандартных преобразований, требующих реализации свёртки, а на уровне весовых коэффициентов – операции арифметического умножения. Схема информационных преобразований на операциях (1), (2) приведена на рис. 3 и может использоваться для любой подобласти изображения. Например, на рис. 4 число уровней пирамиды переменного разрешения определяется необходимой точностью решения, а вариант выделения необходимой подобласти – стратегией восприятия.

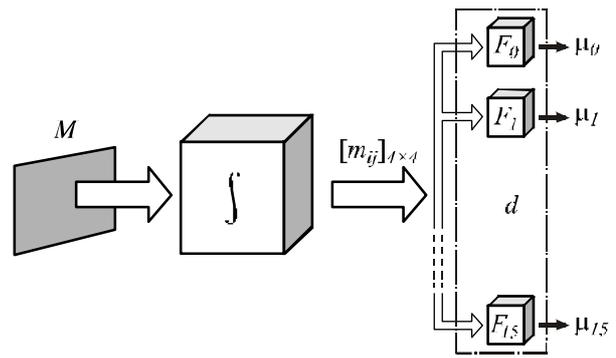


Рис. 3. Схема информационных преобразований

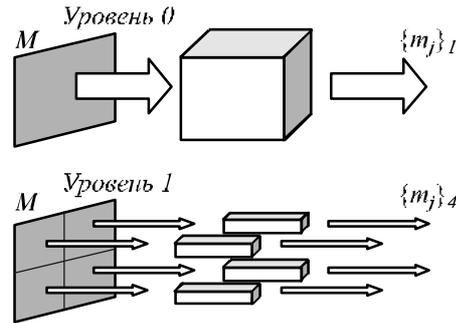


Рис. 4. Пирамида переменного разрешения

Преобразования (1) и (2) – это операции с целыми числами, и поэтому исчезает проблема накопления ошибки при округлениях.

Начиная анализ с верхних уровней пирамиды (рис. 4), где разрешающая способность нижних уровней мала, мы учитываем только низкочастотные компоненты изображения, которые и есть объекты (например, дома). В этом случае детали мелки, а шумы и помехи как бы исчезают.

Таким образом, схема последовательности информационных преобразований состоит из двух этапов:

1) проводим дихотомию изображения как множества M на 16 равных частей и подсчитываем визуальную массу каждой из них – получаем матрицу $\|m_{ij}\|$ изображения;

2) накладываем на матрицу $\|m_{ij}\|$ маски пространственного дифференцирования – получаем 15-мерный вектор. Полученный вектор и есть характеристика исходного изображения.

По данному алгоритму для всех изображений вычисляется вектор и формируется база данных, содержащая изображения и их вектора.

На рис. 5а представлен пример изображения, на рис. 5б – 15-мерный вектор-описание данного изображения.

3. Алгоритмы поиска изображений

Пусть имеется база видеоданных N изображений (порядка 10^4 - 10^5 объектов) и изображение, которое нужно найти. В процессе анализа возможных вариантов решения данной задачи был выбран двухэтапный подход поиска. Его суть заключается в постепенном сужении области поиска: на первом этапе отбираются изображения в количестве n (выборка

производится по каждой составляющей вектора исходного изображения), наиболее близкие к искомому в признаковом пространстве. На втором этапе все n отобранных портретов проходят более тщательное сравнение с исходным изображением. Критерием близости при этом является Евклидово расстояние между векторами.

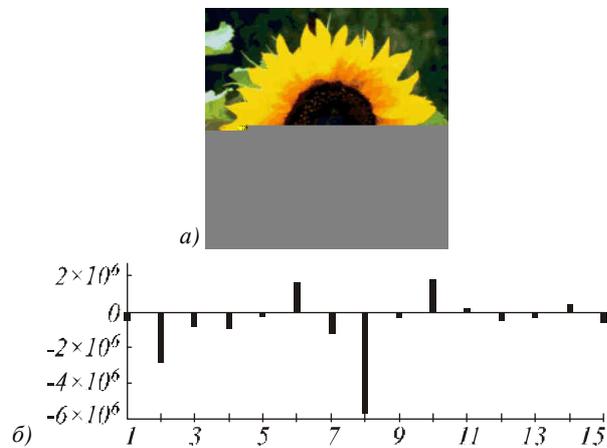


Рис. 5. Пример изображения (а) и его вектор-описание (б)

Описанный в данной статье алгоритм был реализован в виде законченного программного продукта. Описываемая ниже программа предназначена, в первую очередь, для экспериментальной проверки эффективности предложенного метода поиска, отладки и демонстрации их результатов.

Ниже представлены результаты разработанных алгоритмов для конфигурации ПК AMD Turion 64 2,19 ГГц, 1Гб ОЗУ и размера базы данных 20000 объектов. База данных содержит изображения произвольного размера и содержания.

«Грубый» поиск

«Грубый» поиск по базе предназначен для поиска изображений, зашумлённых не более чем на 5% (аддитивный шум). Характеризуется большой скоростью поиска (за счёт малой выборки объектов на первом этапе обработки). Результаты работы алгоритма представлены в табл. 1.

Таблица 1. Результаты по алгоритму «Грубый» поиск изображений»

Аддитивный шум	Достоверность поиска
–	100%
5%	100%
10%	50% (50% – ошибки I рода)

Поиск изображения в данном режиме выполняется в среднем за 75 мс. Авторы считают этот результат приемлемым, поскольку скорость распознавания простого изображения мозгом человека составляет по разным оценкам 40 - 250 мс [9]. Это время необходимо кратковременной памяти человека, чтобы сравнить своё содержимое с запасами долговременной. Как видим, разработанный алгоритм успешно укладывается в указанные временные рамки.

Помехоустойчивый поиск

Помехоустойчивый поиск предназначен для поиска зашумлённых изображений, а также изображений с изменённой яркостью, поворотом на небольшой угол (до 10°). Примеры таких изображений представлены на рис. 6.

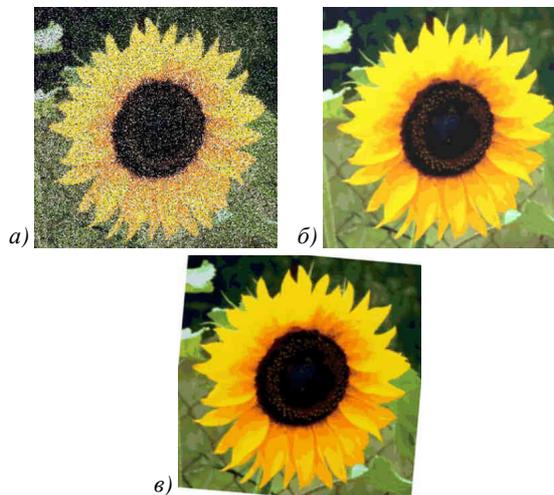


Рис. 6. Примеры зашумлённых и отредактированных изображений: аддитивный шум вероятностью 30% (а), яркость изображения +30 (б), поворот изображения 5 градусов (в)

Данный вид поиска характеризуется средней скоростью поиска (за счёт средней выборки объектов на первом этапе обработки). Производительность (среднее время поиска): 250 мс. Результаты работы алгоритма для разных типов помех и искажений представлены в табл. 2-4.

Таблица 2. Результаты по алгоритму «Помехоустойчивый поиск изображения»

Аддитивный шум	Достоверность поиска
10%	100%
20%	100%
30%	100%
40%	96% (2% – ошибки I рода, 2% – ошибки II рода)
50%	90% (2% – ошибки I рода, 8% – ошибки II рода)

«Многоуровневый» поиск

Данный режим предназначен для поиска не зашумлённых, но повреждённых изображений, например, часть изображения потеряна или стёрта. При этом может быть стёрто до 95% исходного изображения. Пример такого изображения представлен на рис. 7.

В данном случае, чтобы успешно выполнить задачу поиска, необходимо перейти на следующий уровень пирамиды разрешения, т.к. верхний уровень повреждён.

Таблица 3. Результаты по алгоритму
«Помехоустойчивый поиск изображения».
Тестирование с поворотом исходного изображения

Поворот (в градусах)	Достоверность поиска
2	80% (10% – ошибки I рода, 10% – ошибки II рода)
4	60% (10% – ошибки I рода, 30% – ошибки II рода)
6	40% (10% – ошибки I рода, 50% – ошибки II рода)
8	20% (10% – ошибки I рода, 70% – ошибки II рода)

Таблица 4. Результаты по алгоритму
«Помехоустойчивый поиск изображения».
Тестирование с изменением яркости
исходного изображения

Яркость	Достоверность поиска
±10	100%
±20	85% (15% – ошибки II рода)
±30	70% (5% – ошибки I рода, 25% – ошибки II рода)
±40	40% (15% – ошибки I рода, 45% – ошибки II рода)
±50	20% (20% – ошибки I рода, 60% – ошибки II рода)



Рис. 7. Часть изображения испорчена

Время поиска немного больше, чем в случае «грубого» поиска. Объясняется это тем, что в данном случае необходимо сравнить не 1 вектор, а сразу несколько, т.к. операции производятся на втором уровне пирамиды разрешения. Тем не менее, 110 мс – отличный показатель для поиска испорченных изображений. Аналогов данному виду поиска и алгоритму нет.

Таблица 5. Результаты по алгоритму
«Многоуровневый» поиск изображения»

Аддитивный шум	Достоверность поиска
–	100%
5%	100%
10%	50% (50% – ошибки I рода)

Поиск похожих изображений

Данный режим предназначен для поиска набора похожих изображений. На выходе алгоритма получаем некоторое количество изображений, похожих на искомый объект, причём изображения упорядочены по степени сходства (критерием служит Евклидово расстояние между векторами). Среднее время поиска – 800 мс.

Разработанные методы поиска изображений показывают отличные результаты по производительности и достоверности, несмотря на то, что тесты проводились на относительно устаревшем (по меркам сегодняшнего дня) 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.

Заключение

В данной работе разработаны алгоритмы поиска изображений в базах видеоданных. Эти алгоритмы, в отличие от ранее опубликованных в литературе, обеспечивают более низкую вычислительную сложность и позволяют работать с зашумлёнными изображениями. Разработан алгоритм поиска испорченных изображений. В работе показано, что с помощью данного метода можно выполнять поиск изображений, большая часть которых изменена или стёрта. При небольшом усовершенствовании алгоритма, а именно: при переходе на самые нижние уровни пирамиды разрешения имеется возможность выполнять поиск лишь по небольшой части изображения. Разработан новый алгоритм поиска похожих изображений, который может успешно конкурировать по производительности и достоверности с существующими алгоритмами поиска изображений в сети Интернет.

К сожалению, из-за отсутствия единой базы для тестирования невозможно сравнить результаты с другими системами. К тому же большинство систем имеют коммерческий характер и обычно узкую специализацию. Авторы считают, что полученные результаты 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. **Васильева, Н.** Построение соответствий между низкоуровневыми характеристиками и семантической статистических изображений / Н. Васильева, Б. Новиков. – Труды 7-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – Ярославль, 2005.
2. **Swain, M.J.** Color Indexing / M.J. Swain, D.H. Ballard // International Journal of Computer Vision. – 1991. – Vol. 7(1). – P. 11-32.
3. **Tamura, H.** Textural features corresponding to visual perception / H. Tamura, S. Mori, T. Yamawaki // IEEE

- Transactions on Systems, Man, and Cybernetics. – 1978. – Vol. 8(6). – P. 460-473.
4. **Manjunath, B.S.** Texture features for browsing and retrieval of image data / B.S. Manjunath, W.Y. Ma // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1996 – Vol. 18 (8). – P. 837-842.
 5. **Zhang, D.** Content-Based Shape Retrieval Using Different Shape Descriptors / D. Zhang, G. Lu // A Comparative Study, In IEEE International Conference on Multimedia and Expo. – 2001. – P. 289-293.
 6. **Rubner, Y.** A Metric for Distributions with Applications to Image Databases / Y. Rubner, C. Tomasi // In Proceedings of the Sixth International Conference on Computer Vision, IEEE Computer Society. – 1998. – P. 59.
 7. **Утробин, В.А.** Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений / В.А. Утробин. – Н. Новгород: НГТУ, 2001. – 234 с.
 8. **Утробин, В.А.** Физические интерпретации элементов алгебры изображений / В.А. Утробин // Успехи физических наук (УФН). – 2004. – Т. 174, № 10. – С. 1089-1104.
 9. **Демидов, В.Е.** Как мы видим то, что видим / В.Е. Демидов. – М.: Знание, 1987. – 240 с.
- References**
1. **Vasil'eva, N.** Establishing a correspondence between low-level features and semantics of static images / N. Vasil'eva, B. Novikov. – Proceedings of the 7th Russian Scientific Conference “Digital Libraries: Advanced methods and Technologies, Digital Collections”. – Yaroslavl, 2005. – (in Russian).
 2. **Swain, M.J.** Color Indexing / M.J. Swain, D.H. Ballard // International Journal of Computer Vision. – 1991. – Vol. 7(1). – P. 11-32.
 3. **Tamura, H.** Textural features corresponding to visual perception / H. Tamura, S. Mori, T. Yamawaki // IEEE Transactions on Systems, Man, and Cybernetics. – 1978. – Vol. 8(6). – P. 460-473.
 4. **Manjunath, B.S.** Texture features for browsing and retrieval of image data / B.S. Manjunath, W.Y. Ma // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1996 – Vol. 18 (8). – P. 837-842.
 5. **Zhang, D.** Content-Based Shape Retrieval Using Different Shape Descriptors / D. Zhang, G. Lu // A Comparative Study, In IEEE International Conference on Multimedia and Expo. – 2001. – P. 289-293.
 6. **Rubner, Y.** A Metric for Distributions with Applications to Image Databases / Y. Rubner, C. Tomasi // In Proceedings of the Sixth International Conference on Computer Vision, IEEE Computer Society. – 1998. – P. 59.
 7. **Utrobina, V.A.** Information models of visual perception for the problems of computer image processing / V.A. Utrobina. – Nizhny Novgorod: “NSTU” Publisher, 2001. – 234 p. – (in Russian).
 8. **Utrobina, V.A.** Physical interpretation of the elements of the images / V.A. Utrobina // Successes of Physical Sciences. – 2004. – V. 174, N 10. – P. 1089-1104. – (in Russian).
 9. **Demidov, V.E.** As we can see what we see / V.E. Demidov. – Moscow: “Knowledge” Publisher, 1987. – 240 p. – (in Russian).

THE SEARCH ALGORITHM FOR IMAGE IN DATABASES

Desyatnikov I.E., Utrobina V.A.

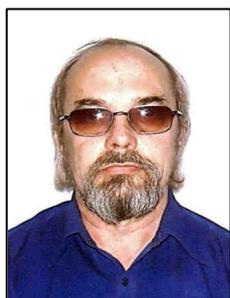
Nizhny Novgorod State Technical University R.E. Alekseeva

Abstract

This paper briefly discusses the existing methods for image lookup in video databases. A new method of image lookup is suggested; it is based upon active perception theory which is being developed at the Computing Systems and Technologies department of NSTU. Its key advantage is its low computing cost and thus high performance. According to the method, every image is represented as a 15-dimensional vector. Not only the original images are matched; the algorithm also handles noisy, edited, corrupted, rotated images. An algorithm for matching similar images is proposed as well. Its reliability and performance allow it to compete with existing commercially available image lookup technologies.

Key words: image retrieval, active perception theory.

Сведения об авторах



Утробин Владимир Александрович – выпускник радиотехнического факультета Горьковского политехнического института, доктор технических наук, профессор, заместитель заведующего кафедрой «Вычислительные системы и технологии» НГТУ, автор более 100 научных работ в области теории обработки многомерных сигналов, распознавания образов, моделирования биосистем.

E-mail: utrobina-va@yandex.ru.

Vladimir Aleksandrovich Utrobina – graduated from Radio Engineering faculty of Gorky Polytechnic Institute. He is a Doctor of Technical Science, Professor, Vice Chairman of the Computing Systems and Technologies department of NSTU. He authored more than 100 papers on multidimensional signal processing, pattern recognition, biosystems modeling.



Десятников Игорь Евгеньевич, 1987 года рождения, в 2010 году с отличием окончил Нижегородский государственный технический университет им. Р.Е. Алексеева по специальности «Вычислительные машины, системы, комплексы, сети». В настоящее время является аспирантом по направлению «Теоретические основы информатики». Область научных интересов: компьютерная оптика, обработка изображений, поиск изображений.

E-mail: digore@bk.ru.

Igor Evgenjevich Desyatnikov, (b. 1987) graduated with honours (2010) from Nizhny Novgorod State Technical University R.E. Alekseeva, majoring in Computer Science, systems, facilities, networks. He is currently a postgraduate student in "Theoretical Foundations of Informatics". His areas of research are computer optics, image processing, image search.

Поступила в редакцию 4 апреля 2011 г.