

УДК 519.92

## **Устойчивый метод поиска изображений в визуальных базах данных**

А.В Шаронов, Н.А. Максимов, Д.П. Синча

### **Аннотация**

Значительный интерес к беспилотным летательным аппаратам в последнее время усилил необходимость создания эффективной системы распознавания изображений [1]. В отличие от других систем распознавания образов (систем биометрической идентификации, систем распознавания лиц, систем распознавания документов и т.д.), непосредственным объектом, подлежащим классификации (распознаванию) в данном случае, является само изображение, которое, естественно, может быть искажено различными факторами. В основе любой системы распознавания изображений лежат, главным образом, два алгоритма: алгоритм формирования дескриптора (индекса) изображения (т.е. организованной совокупности признаков изображения) и алгоритм оценки степени сходства двух изображений по их дескрипторам. Достаточно перспективный вариант этих алгоритмов представлен в данной статье.

### **Ключевые слова**

распознавание образов, изображение, признаки, степень сходства, Поиск изображений в базах данных, искажение изображений, классификация изображений.

### **Введение**

Существующие методы распознавания поиска изображений в визуальных базах данных узко специализированы и используют, как правило, различные эвристические методы, соответствующие специфике практического применения. [2,3]. В большинстве случаев для индексирования изображений используют цветовые характеристики изображений, в частности цветовые гистограммы, форму объектов и некоторые другие признаки [3,4]. Однако такого рода признаки становятся малоэффективными, если искомое и эталонное изображение достаточно сильно отличаются из-за помех, изменения масштаба, кадрирования или изменения яркости. В этой связи встает задача создания эффективных алгоритмов индексирования цветных (полутоновых) изображений и алгоритмов сравнения эталонных и предъявляемых изображений.

### Постановка задачи

Поставим задачу следующим образом. Имеется коллекция (произвольного объема) растровых неповторяющихся изображений произвольной тематики. Иными словами, имеется база данных изображений-эталонов. Каждое изображение может иметь произвольное пространственное разрешение (желательно не ниже 100x100 пикселей) и произвольную глубину цвета/битность (желательно не ниже 8 бит/пиксель). Каждое изображение коллекции может быть как цветным, так и не цветным. Изображения не содержат никакой сопутствующей информации (никаких ключевых слов, фраз на естественном языке или каких-либо иных метаданных). На изображения не нанесены водяные знаки. Каждое изображение базы формирует отдельный класс и является единственным представителем данного класса. Необходимо создать систему, на вход которой подается удовлетворяющее вышеприведенным ограничениям изображение, которое может, как принадлежать, так и не принадлежать коллекции. Подающееся на вход системы изображение из коллекции может быть искажено одним или несколькими следующими факторами:

- групповые помехи произвольной формы и размера,
- поворот,
- изменение масштаба,
- нелинейное изменение яркости,
- изменение разрешения,
- сдвиг,
- кадрирование/ декадрирование,
- зеркальное отображение.

Примеры некоторых из перечисленных искажений показаны на рисунке 1.

Система должна правильно классифицировать входное изображение, т.е., либо отнести его к одному из классов (естественно, верному) (результат работы – представитель этого класса), либо резюмировать, что входное изображение не принадлежит ни к одному классу.

Иными словами, ставится задача создания системы распознавания изображений, которая принципиально сводится к созданию двух упомянутых выше алгоритмов.

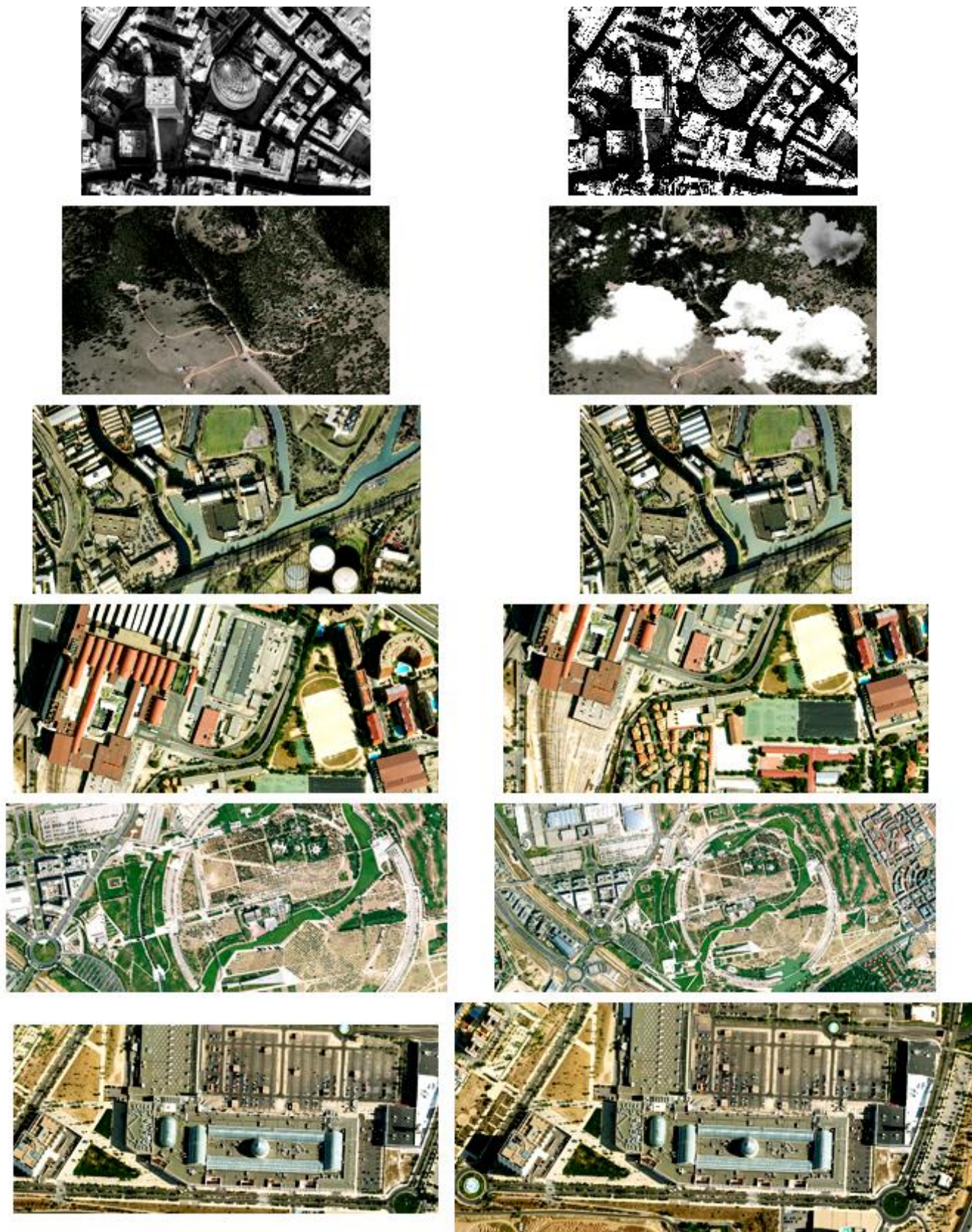


Рисунок 1 – Примеры изображений-эталонов (левый столбец) и их искаженных версий (правый столбец). Соответствующие искажающие факторы (сверху-вниз): нелинейное изменение яркости, групповые помехи, кадрирование, сдвиг, изменение масштаба, декадрирование

### Алгоритм формирования дескриптора изображения

Вследствие наличия искажающих факторов необходимо разрабатывать такие признаки изображения, каждый из которых будет малочувствителен хотя бы к одному из рассматриваемых факторов. По этой же причине разрабатываемые признаки изображения должны, по крайней мере, слабо коррелировать, а лучше, если они будут иметь слабую статистическую взаимосвязь. Силу корреляционной взаимосвязи можно оценить посредством выборочной оценки коэффициента корреляции Пирсона [5]. Силу статистической взаимосвязи можно оценить посредством величины средней взаимной информации [6]. Соответственно, дескриптор изображения в нашем случае – это просто вектор признаков.

Разработанный дескриптор изображения состоит из трех групп признаков. Первая группа признаков получается следующим образом. Исходное изображение преобразуется в полутоновое, если изначально таковым не являлось. Далее, используя метод выделения локальных двоичных микроструктур (метод LBP), полученное полутоновое изображение приводится к LBP-изображению, т.е. к такому изображению, в котором яркость каждого пикселя суть LBP-код (рисунок 2).



Рисунок 2 – Исходное 24-битное изображение (слева) и соответствующее ему 8-битное LBP-изображение (справа)

Подробное описание метода LBP может быть найдено в [7]. Суть метода заключается в следующем (рисунок 3). Вокруг каждого пикселя изображения ( $g_c$ ) описывается окружность радиуса  $R$  (в нашем случае  $R=1$ ). Далее определяются яркости ( $I$ )  $P$  точек (у нас  $P=8$ ), равномерно распределенных по окружности. Если точка совпадает с пикселем изображения (точки  $g_0, g_2, g_4$  и  $g_6$ ), то яркость этой точки есть яркость пикселя. В противном случае (точки  $g_1, g_3, g_5$  и  $g_7$ ) яркость точки получают посредством интерполяции по 4 соседям. Выбрав одну из  $P$  точек в качестве начальной ( $g_0$ ) и задав направление обхода (оба действия выполняются *всегда* единообразно), вычисляется величина, называемая LBP-кодом:

$$LBP(x, y) = \sum_{p=0}^{P-1} s(I(g_p) - I(g_c)) \cdot 2^p$$

$$s(x) = \begin{cases} 1 & \text{при } x \geq 0 \\ 0 & \text{при } x < 0 \end{cases}$$

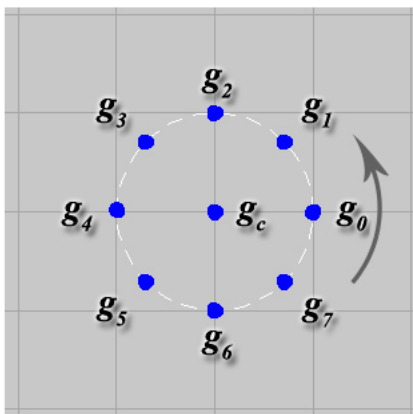


Рисунок 3 – Иллюстрация процесса получения LBP-кода для пикселя  $g_c$ . Узлы сетки соответствуют пикселям изображения

Предположим, что в качестве дескриптора исходного изображения выступает нормированная гистограмма соответствующего ему LBP-изображения, т.е. вектор размерности  $256 \times 1$ , каждый элемент которого есть частота появления соответствующего LBP-кода на LBP-изображении. Однако такой дескриптор не будет эффективным при решении поставленной задачи по крайней мере, из-за двух причин.

*Первая причина* состоит в слишком большой размерности пространства признаков. Пусть некоторому пикселю  $g_i$  изображения поставлен в соответствие LBP-код  $LBP(g_i) = 31$ . Повернув изображение на  $90^\circ$  против часовой стрелки, получим  $LBP(g_i) = 124$ , а на  $90^\circ$  по часовой стрелке –  $LBP(g_i) = 199$ . Получить устойчивость (но не инвариантность) к повороту изображения можно, если сгруппировать LBP-коды и вычислять не частоты появления отдельных LBP-кодов, а суммы частот появления LBP-кодов каждой группы. Для приведенного примера группа будет следующей:  $\{31, 62, 124, 143, 199, 227, 241, 248\}$ . Таким образом, 256 признаков исходного изображения “сжимаются” до 36. Однако и такой дескриптор изображения недостаточно эффективен вследствие *второй причины*, которая сводится к учету редко появляющихся LBP-кодов. Продемонстрируем это на следующем примере. Сформируем объемную ( $V = 40000$  изображений) коллекцию растровых неповторяющихся изображений, применив к каждому из них метод LBP, и, построив в итоге совокупную нормированную гистограмму LBP-кодов (рисунок 4), мы увидим, что большинство LBP-кодов появляются достаточно редко. Необходимо объединить все “редкие” коды в одну группу, что ведет к “сжатию” 36 признаков исходного изображения до 16. Итак, пока мы характеризуем исходное изображение 16 признаками – это предварительная

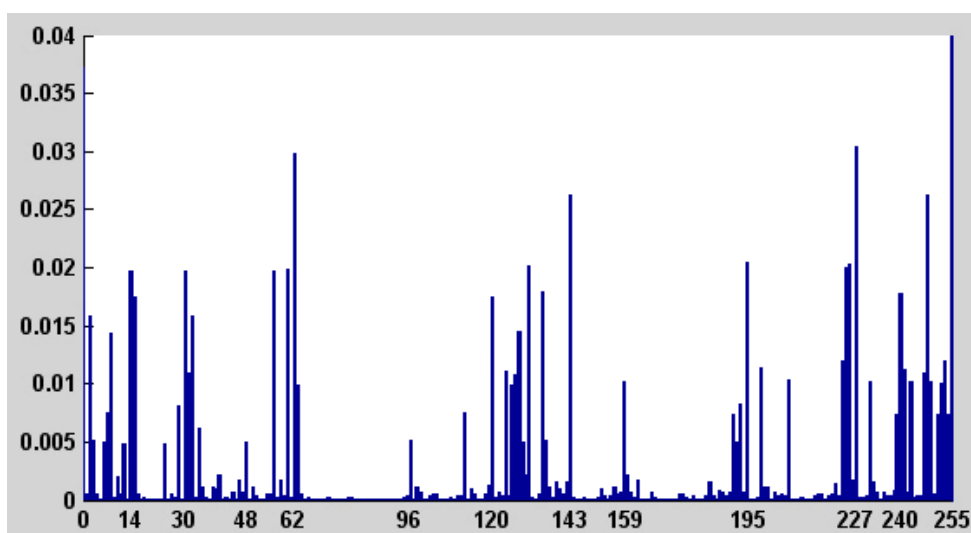


Рисунок 4 – Совокупная нормированная гистограмма LBP-кодов

первая группа искомых признаков. Также формируются предварительные вторая и третья группы искомых признаков (о них – далее). Окончательные группы искомых признаков получаются следующим образом. Пусть исходное изображение характеризуется  $N = N_1 + N_2 + N_3$  признаками.

Вычислив для каждого изображения вышеупомянутой коллекции ( $V = 40000$ ) эти  $N$  признаков, получим матрицу объект-признак [8] размером  $40000 \times N$ . Выбрав любые два столбца этой матрицы, и, воспользовавшись выборочным коэффициентом корреляции, можно оценить силу корреляционной взаимосвязи между соответствующими признаками. Если вместо выборочного коэффициента корреляции использовать среднюю взаимную информацию, то можно оценить силу статистической взаимосвязи между данными признаками. Построим две матрицы признак-признак размерами  $N \times N$ . Каждый элемент первой матрицы – величина средней взаимной информации для соответствующих признаков. Каждый элемент второй матрицы – величина выборочного коэффициента корреляции для соответствующих признаков (т.н. корреляционная матрица). Совместный анализ обеих матриц определяет размерность и состав дескриптора исходного изображения ( $N' = N'_1 + N'_2 + N'_3$ ,  $N' < N$ ).

Предварительные вторая и третья группы искомых признаков получаются посредством дальнейшей обработки LBP-изображения (смотри рисунок 2). При этом используется концепция линейно-симметрического изображения [9], введенная Джозефом Бигуном. На рисунке 5 приведены примеры линейно-симметрических (л-с) изображений. Упрощенно (точное определение смотри в [9]) можно определить л-с изображение, как “полосатое”, т.е. каждое такое изображение выглядит, как группа параллельных друг другу “прямых” полос в общем случае различных ширин и яркостей. Яркость в пределах полосы неизменна. Количество полос произвольно. Единичный вектор перпендикулярный всем полосам л-с изображения задает направление линейной симметрии этого изображения. Невырожденное л-с изображение имеет два и только два (противоположных друг другу) направления линейной симметрии.

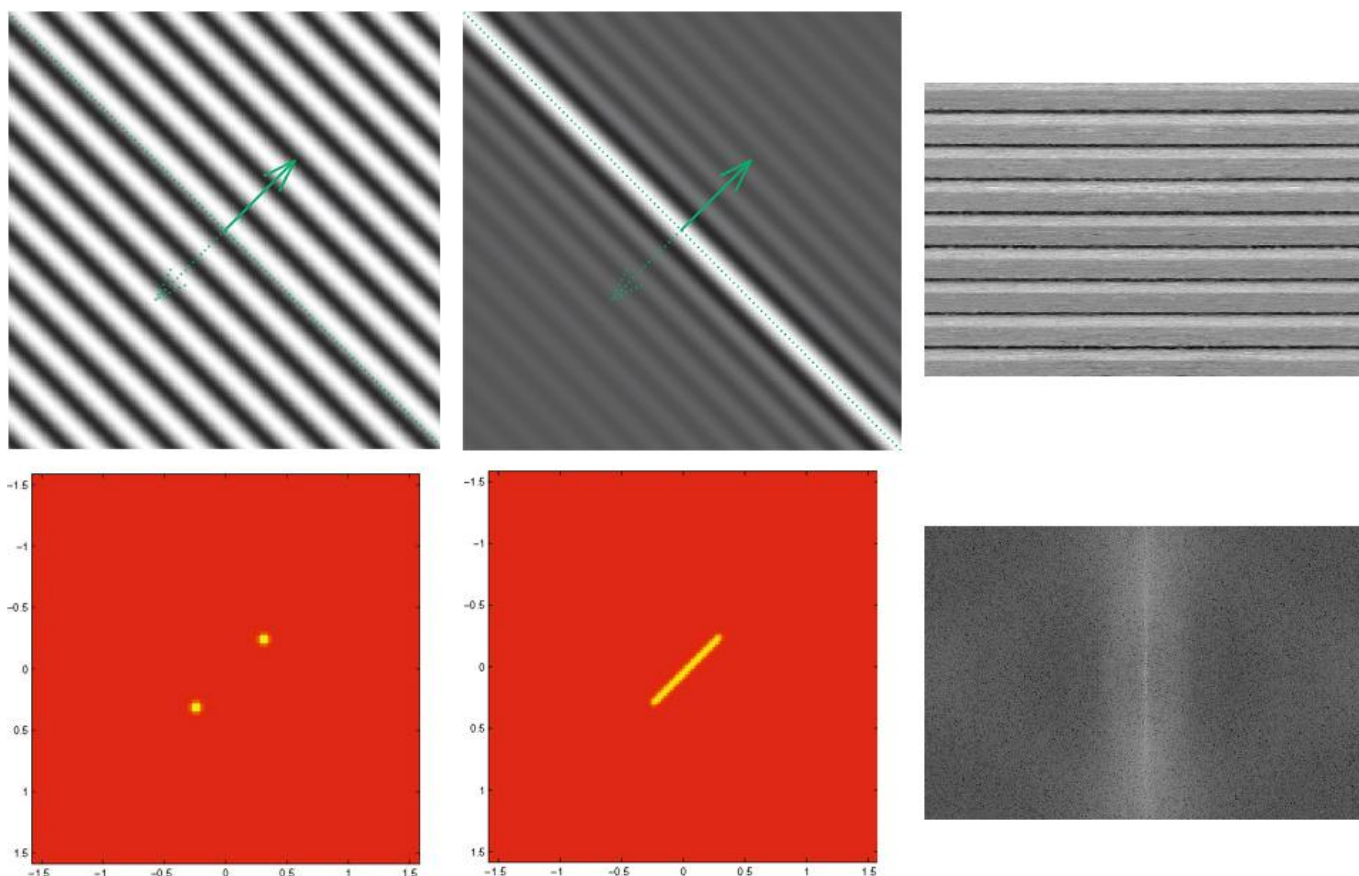


Рисунок 5 – (Верхняя строка) Два линейно-симметрических изображения (искусственные) (слева и в центре) и одно почти линейно-симметрическое изображение (реальное) (справа); Зеленые вектора задают направления линейной симметрии изображений.

(Нижняя строка) - Амплитудные спектры приведенных изображений

Л-с изображения обладают следующим замечательным свойством. Если и только если изображение является линейно-симметрическим, его Фурье-спектр (а также амплитудный и энергетический спектры) сконцентрирован в прямую линию, проходящую через начало координат (т.н. центральную прямую) и задающую направления линейной симметрии этого изображения.

Можно получить два интересных признака изображения, аппроксимировав его л-с изображением. Процесс аппроксимации произвольного изображения л-с изображением эквивалентен процессу аппроксимации Фурье-спектра (или амплитудного спектра, или энергетического спектра) этого изображения центральной прямой. Задача аппроксимации – поиск такой ориентации центральной прямой, при которой ошибка аппроксимации будет минимальной (мы определяем ошибку посредством метода ортогональной регрессии [10]). **Ориентация оптимальной (дающей наименьшую ошибку) центральной прямой (угол  $\alpha$ ) и обеспечиваемый ею показатель качества аппроксимации ( $C$ ) есть искомые (интересные) признаки изображения.**



Каждый пиксель LBP-изображения рассматривается вместе со своей окрестностью, как отдельное (маленькое) изображение, для которого вычисляются признаки  $\alpha$  и  $C$ . Нормированные гистограммы признаков  $\alpha$  и  $C$  анализируются подобно тому, как было рассмотрено ранее, и “сжимаются” в предварительные вторую и третью группы искомым признаков.

#### Алгоритм оценки степени сходства двух изображений по их дескрипторам

Степень сходства  $\rho$  изображений  $Im_1$  и  $Im_2$ , характеризуемых векторами признаков  $D_1$  и  $D_2$  размерностью  $N'$ , вычисляется следующим образом.

$$\rho(Im_1, Im_2) = \frac{\sum_{j=1}^{N'} s_j}{N'}$$

$$s_j = \begin{cases} \frac{\Delta_j - ad_j}{\Delta_j}, & \text{если } \Delta_j - ad_j \geq 0 \\ 0, & \text{иначе} \end{cases}$$

$$AD = |D_1 - D_2|$$

Здесь  $\Delta$  – вектор допусков, определяемых экспериментально.

Низкая сложность ( $O(n)$ ) рассмотренных алгоритмов, их естественная распараллеливаемость и обеспечиваемая ими высокая точность распознавания определяют эффективность построенной на их основании системы распознавания изображений “VARdi”[1], основные характеристики которой приведены ниже.

#### Конфигурации тестовых компьютеров:

PC 1: AMD Athlon 64 3000+(1.7 GHz) + 1GB RAM + Windows XP 32-bit

PC 2: Intel Q6600 (2.4GHz) (использовалось одно ядро) + 4GB RAM + Windows XP 64-bit

Таблица 1 – Время формирования дескриптора изображения в системе “VARdi”

Размер полноцветного (24-bit) изображения	Время формирования дескриптора изображения (сек)	
	PC 1	PC 2
768x512 (1.125 Mb)	1.34	0.69
2048x1536 (9 Mb)	10.75	5.5
4096x2304 (27 Mb)	32.26	16.5

Таблица 2 – Точность распознавания системы “VARdi” при различных искажающих факторах

Искажающий фактор	Дополнительная информация	Точность распознавания
Отсутствует	—	1
Групповые помехи	Искажение вплоть до 25% площади изображения	Не менее 0,93
Групповые помехи	Искажение вплоть до 40% площади изображения	Не менее 0,75
Изменение масштаба	Вплоть до 2.0 раз	Не менее 0,9
Кадрирование	Отсечение вплоть до 50% площади изображения	Не менее 0,95
Кадрирование	Отсечение вплоть до 70% площади изображения	Не менее 0,88
Декадрирование	Увеличение площади изображения не более чем на 50%	Не менее 0,97
Нелинейное изменение яркости	Функции изменения яркости изображения - гладкие и неубывающие	0,99
Сдвиг	Искажение вплоть до 25% площади изображения	Не менее 0,98
Сдвиг	Искажение вплоть до 50% площади изображения	Не менее 0,77
Изменение разрешения	Вплоть до 3.2 раз	Не менее 0,9
Зеркальное отображение	Относительно горизонтальной, вертикальной или обеих осей	1
Поворот	Угол поворота изображения произволен	Не менее 0,84

### Выводы и область применения

Работоспособность программы “VARdi” проверялась на различных классах объектов - аэрофотоснимкам, снятым в разных ракурсах, с разных высот, при различном освещении, а также на портретах людей, снимках медицинского характера изображения произвольного характера. Были получены хорошие результаты, во многих случаях даже превосходящие данные, приведенные в таблицах 1-3.

Универсальность алгоритма, построенного на строгих математических принципах, позволяет применять, по мнению авторов, данный подход в следующих областях:

- отслеживание перемещения объектов на видеоизображениях;
- поиск изображений любого характера в Интернете или любых других больших коллекциях изображений;
- автоматическое формирование панорамных снимков из отдельных изображений;
- автоматическая «сшивка» аэрофото- и космических снимков, снятых в разное время, с разных высот и разных ракурсов в единую фотокарту.

В настоящее время исследуется возможность использования системы распознавания изображений “VARdi” в качестве подсистемы системы позиционирования беспилотного летательного аппарата.

Работа выполнена при финансовой поддержке РФФИ проектам 09-08-00689-а, 10-08-00397-а.

### **Библиографический список**

1. Синча Д.П. Принципы построения универсальной системы распознавания искаженных изображений // Труды XXVIII Межрегиональной научно-технической конференции, г. Серпухов, 2009 г., Серпуховский ВИ РВ, Сборник №3, с.149-151.
2. Н.С. Байгарова, Ю.А. Бухштаб, А.А. Воробьев, А.А. Горный Организация управления базами визуальных данных Препринт Института прикладной математики им. М.В. Келдыша РАН, 2000, N 6
3. Jain, R. and Gupta, A., Computer Vision and Visual Information Retrieval, 1996  
<http://vision.ucsd.edu/papers/rosenfeld/>
4. Н.С. Байгарова, Ю.А. Бухштаб, А.А. Горный Методы индексирования и поиска визуальных данных. Препринт Института прикладной математики им. М.В. Келдыша РАН, 2000, N 7.
5. Кобзарь А.И. Прикладная математическая статистика // М:Физматлит, 2006, с.607.
6. Потапов А.С. Распознавание образов и машинное восприятие // СПб:Политехника, 2007, с.45.
7. Ojala T., Pietikainen M., Maenpaa T. Multiresolution gray scale and rotation invariant texture analysis with local binary patterns // IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, 2002.
8. Большаков А.А., Каримов Р.Н. Методы обработки многомерных данных и временных рядов // М:Горячая линия – Телеком, 2007, с.20.
9. Bigun J. Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision // Springer, 2006.
10. Дрейпер Н., Смит Г. Прикладной регрессионный анализ // М:Диалектика, 2007.

**Сведения об авторах**

Шаронов Анатолий Васильевич, профессор, заведующий кафедрой Московского авиационного института (национального исследовательского университета), д. т.н., тел.: +7 (495) 656-44-85, e-mail: [sharonov@mai.ru](mailto:sharonov@mai.ru)

Максимов Николай Анатольевич, доцент Московского авиационного института (национального исследовательского университета), к.т.н., тел.: +7 (495) e-mail: [n-a-maximov47@Yandex.ru/](mailto:n-a-maximov47@Yandex.ru/).

Синча Дмитрий Петрович, аспирант Московского авиационного института (национального исследовательского университета), тел.: + 7 (499) 504-14-97, e-mail: [atsglance@list.ru](mailto:atsglance@list.ru)

