# Быстродействующий алгоритм семантической классификации JPEG–изображений

Дорогов А.Ю., Курбанов Р.Г., Разин В.В.

Санкт-Петербургский государственный электротехнический университет (СПбГЭТУ) "ЛЭТИ", dorogov@lens.spb.ru

***Аннотация.*** В работе обсуждается алгоритм семантической классификации JPEG изображений и результаты его моделирования в программной среде МАТЛАБ. Предлагаемый алгоритм основан на трех основные идеях: 1) для классификации изображений используется спектральное признаковое пространство, формируемое стандартной процедурой блочного кодирования JPEG-формата, что позволяет производить классификацию без восстановления изображения; 2) семантика полного изображения является производной от семантики сегментов изображения, что позволяет реализовать экономную иерархическую процедуру классификации; 3) исключаются какие-либо априорные предположения о конфигурации семантического класса в пространстве признаков, классификация выполняется по достоверным прецедентам базы данных.

В контексте семантической классификации представлены новые алгоритмы адаптивной сегментации изображений, алгоритмы информативной оценки системы первичных признаков и формирования сложных вторичных признаков, алгоритмы нечеткой метрической классификации сегментов изображения, алгоритмы нечеткой иерархической классификации изображений по результатам сегментной классификации.

## 1. Введение и обзор ключевых работ

**1.1. Постановка задачи.** По статистическим оценкам в мультимедийных базах данных до 80 процентов изображений представлены в JPEG формате или производных от него (JFIF, SPIFF, JBIG, JPEG-EXIF, MPEG). Формат JPEG представляет собой один из лучших методов сжатия с потерями, в котором обобщен полувековой опыт исследований инженеров и ученых, работающих в компьютерной, телевизионной и других областях, связанных с человеческим зрением и компьютерной графикой. В мультимедиа и Интернет технологиях этот формат широко используется для хранения, обработки и передачи изображений по каналам связи.

Типичной задачей обработки мультимедиа изображений является их классификация. Среди различных видов классификаций по уровню значимости и уровню сложности выделяется задача семантической классификации зрительных образов [1,2], позволяющая получить содержательный для человека ответ на вопрос: «Что изображено на картинке?». Прогресс в распознавании семантики образов, оказывает непосредственное воздействие на развитие систем компьютерного зрения, робототехнических систем, поисковых систем Интернет технологий, специализированных баз данных для мультимедиа систем. Время, затраченное на классификацию образов, имеет решающее значение как для баз данных большого объема, так и для систем, работающих в реальном масштабе времени. Высокое быстродействие систем классификации может быть достиг-

нуто за счет сокращения непродуктивных преобразований видеообраза и использования быстрых алгоритмов обработки данных.

**1.2. Типичная схема семантической классификации.** Анализ семантики является вершиной иерархической процедуры обработки изображений. В основании пирамиды лежат методы формирования первичной системы информативных признаков. Основное требование на данном этапе – обеспечить максимально-возможную инвариантность признаков к топологическим преобразованиям и высокое быстродействие в получении первичной информации. Первичные признаки представлены в пространстве высокой размерности и имеют значительные различия по уровню информативности, эти обстоятельства препятствуют их непосредственному использованию в задаче классификации.

На следующем уровне иерархии формируется система вторичных признаков с примерно одинаковым уровнем значимости. Исходной информацией при этом служит анализ накопленной базы данных, который производится всякий раз при добавлении новых данных. Главное требование к процедуре анализа состоит в том, чтобы подобрать оптимальную систему признаков и сократить размерность признакового пространства. В режиме обучения классифицирующая система, взаимодействуя с оператором, накапливает банк семантических понятий. Временные затраты на анализ базы данных на этом этапе не имеют решающего значения, поскольку не влияют на быстродействие системы в рабочем режиме.

На верхнем уровне пирамиды, параметрическими или непараметрическими методами решается задача семантической классификации образов. Ответ может быть многозначным, поэтому необходимо ранжировать полученные решения, используя ту или иную оценку уровня значимости.

**1.3. Обзор существующих методов решения**. Известные методы решения задачи семантической классификации изображения, сохраняя в целом типичную схему, имеют большое разнообразие в способах и стратегиях реализации этапов.

Проблема распознавания семантики изображения (в зарубежной литературе используется аббревиатура CBIR – Content-based image retrieval) имеет более чем 20-летнюю историю. Первые работы относятся к 80-м годам прошлого века [3]. Однако наиболее существенное развитие это направление получило в последние несколько лет. Многие из выполненных исследований посвящены развитию стратегий последовательного уточнения запроса и оптимизаций поисковых процедур для изображений в конкретных базах данных большого объема [3-7]. Хорошо известным продуктом является система анализа семантики MARS, разработанная в университете Illinois [4]. К другим средствам относятся системы анализа PicToSeek [6], DrawSearch [7] и Viper [8]. Общая стратегия в разработке таких систем семантического поиска состоит в создании нового запроса, который оптимизируется в процессе диалога с пользователем. Однако для эффективного использования эта стратегия требует сложной трансформации базы данных в лингвистическую модель с взвешенным набором весов для терминологических переменных. Кроме того, в некоторых системах (Viper [8]) генерации подобной модели приводит к очень большой размерности признакового пространства. Процедура рафинирования поискового запроса использует

следующие методы: многомерные индексные структуры [9], набор признаков, выделяемых из примеров, представляющих интерес для пользователя [10], и взвешенные средние для позитивных и негативных примеров [11]. Интерактивный диалог с пользователем в CBIR системах трактуется как парадигма супервизорного обучения, стимулирующая механизм человеческого восприятия образов. В качестве средства формирования весов используются супервизорные нейронные сети [12], методы вероятностной классификации [13] и взвешенные евклидовы расстояния [14] в многомерном пространстве.

В процессе диалога производится назначение весов терминологическим понятиям. В результате строится некоторая функция подобия. Однако разделяющая мощность такой функции сильно ограничена, поскольку из-за проблем вычислительной сложности для построения функций подобия используется, как правило, не более чем квадратичная форма. Нейросетевые модели хорошо подходят для решения данной задачи, но требуют наличия большого объема обучающих данных для каждого нового запроса.

Альтернативный подход не супервизорного обучения, основанный на использовании самоорганизующихся карт Кохонена, рассматривается в работе [15]. Достоинство этого метода в том, что в процессе обучения системы не требуется взаимодействия с человеком-оператором. Но вследствие ограниченности правил формирования карт это, с другой стороны, приводит к снижению разделяющей мощности системы распознавания. Кроме того, для восприятия результата классификации человеком необходимо выполнить трактовку автоматически вырабатываемых концептуальных понятий, что не всегда возможно.

Задача семантического распознавания пересекается с задачей создания искусственного разума [16,17]. Это направление развивается в нашей стране в Институте проблем управления РАН им. В.А. Трапезникова. Сущность подхода состоит в трансформации количественной информации любого вида к «текстовой форме» (точнее к номинальной шкале). В процессе не супервизорного обучения реализуется выделение повторяющихся текстовых цепочек, которые трактуются как концептуальные понятия. Главная проблема использования данного подхода связана с адекватной трансформацией изображения к «текстовому» образу.

Приведенный обзор показывает, что в настоящее время научно-техническое направление семантической классификации изображений активно развивается, но до полного решения еще далеко, каждый из известных методов имеет как достоинства, так и определенные недостатки. Наилучшие результаты следует ожидать при решении конкретных задач.

**1.4. Конкретизация целей исследования.** В нашей работе предлагается ограничиться задачей семантической классификации JPEG-изображений. Мотивы такого подхода обусловлены тем обстоятельствам, что JPEG-сжатие основано на использовании априорных данных о свойствах человеческого зрения, поэтому можно ожидать, что кодированный образ будет представлен в признаковом пространстве с достаточно высокой степенью информативности.

В качестве первичной системы признаков предлагается использовать спектральные коэффициенты блочного кодирования. Размерность пространства первичных признаков существенно ограничена за счет используемой в JPEG схемы

адаптивного сжатия информации. Таким образом, снимается проблема формирования первичной системы признаков и устраняется необходимость их прореживания по уровню информативности, это обеспечивает принципиальную основу для построения быстрых процедур семантической классификации. Главной целью настоящего исследования является разработка методов и алгоритмов семантической классификации JPEG-изображений.

## 2. Идея исследования

В JPEG формате цветные изображения представляются в виде яркостной ( $Y$ ) и двух цветоразностных компонент ( $Cb, Cr$ ). Яркостная и цветоразностные компоненты образуют трехмерное цветовое пространство, подобное цветовому пространству, используемому в телевизионном вещании по системе СЕКАМ. Формулы преобразования цветового пространства RGB («Красный, Зеленый, Синий») к цветоразностному пространству $YCbCr$ приведены в рекомендации JFIF [21].

Алгоритм JPEG-сжатия включает в себя два этапа. На первом этапе выполнятся разделение изображения на блоки размеров 8×8 пикселей, которые подвергаются двумерному ортогональному косинусному преобразованию. В результате получаются матрицы спектральных коэффициентов размером 8×8. Было доказано, что косинусное преобразование является оптимальным для случайных процессов марковского типа. Случайные процессы этого класса в статистическом смысле хорошо аппроксимируют случайное множество изображений. Это означает, что, обеспечивая в среднем достаточно хорошее качество восстановленного изображения, можно ограничиться только несколькими спектральными коэффициентами и отбросить ряд высокочастотных компонент. В результате объем передаваемой информации значительно сокращается, причем оставшиеся коэффициенты несут наиболее существенную информацию об изображении.

Прореживание коэффициентов реализуется за счет процедуры нелинейного квантования при переходе к 8 (или 12) битному представлению данных. Кроме того, при кодировании учитывается также, что человеческий глаз обладает более низким пространственным разрешением для цветовых перепадов по сравнению с разрешением для перепадов яркости. Поэтому цветоразностные компоненты кодируются с вдвое меньшим пространственным разрешением, что в два раза сокращает объем информации для каждой компоненты.

На втором этапе производится сжатие данных без потерь за счет использования кода Хаффмана. В процедуре Хаффмана числа кодируются неравномерным самосинхронизирующимся бинарным кодом. Сжатие достигается за счет того, что статистически наиболее вероятные значения кодируются короткими кодовыми словами, а наименее вероятные - длинными. В стандарте JPEG предусмотрена передача кодирующих таблиц кода Хаффмана вместе с изображением, однако, в большинстве случаев используются рекомендуемые в стандарте [21] примеры кодирующих таблиц, полученные разработчиками стандарта при статистическом исследовании множества изображений. На выходе кодировщика Хаффмана формируется непрерывный битовый поток данных.

При восстановлении изображения, после этапа декодирования Хаффмана, восстанавливаются 8-ми (или 12-ти) битные коды переданных спектральных коэффициентов по яркостной и каждой цветоразностной компонентам. Эти коды можно использовать как первичные информативные признаки, которые уже не требуют какого-либо цензурирования по уровню значимости. Таким образом, для выполнения классификации изображения нет необходимости в его полном восстановлении, поскольку признаковое пространство уже на промежуточном этапе подготовлено процедурой JPEG–кодирования. Исключение этапа полного восстановления изображения обеспечивает значительную экономию времени при решении задачи семантической классификации. Общая схема кодирования/декодирования JPEG изображений на уровне блоков представлена на рис. 1. Вертикальная серая стрелка указывает позицию отбора первичных информативных признаков.

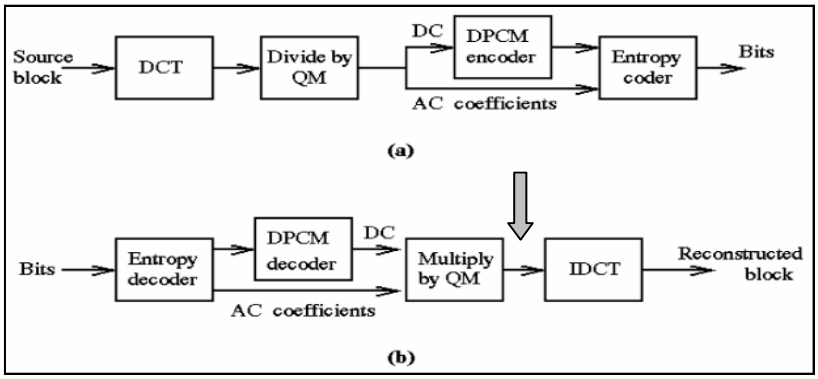



Рис. 1. a) Схема JPEG – кодирования, b) Схема декодирования JPEG-изображений. DCT, IDCT – прямое и инверсное косинусное преобразование, DC – нулевой спектральный коэффициент, AC – множество высокочастотных коэффициентов, QM – процедуры квантования/ деквантования спектральных коэффициентов, Entropy coder/decoder – кодирование и декодирование Хаффмана, Bits –битовый поток в канале связи.

В процедуре JPEG–кодирования для каждого блока изображения вычисляются нулевые спектральные коэффициенты (DC), которые несут информацию о среднем уровне яркостной и цветоразностных компонент. В предлагаемом подходе эта тройка коэффициентов используется для блочной пространственной сегментации изображения. Сегментация реализуется процедурой метрической кластеризации, с метрикой Хэмминга, межклассовое расстояние оценивается по принципу «максимально удаленных соседей». В зависимости от уровня сложности изображение может быть разделено на 1-7 пространственных сегментов. В каждом сегменте в свою очередь выделяются непрерывные компоненты, число которых не ограничивается. Сегменты описываются в 100–мерном признаковом пространстве (принципы формирования вторичных признаков представлены в разделе 3).

В процессе обучения в базе данных накапливаются 100–мерные вектора, каждый из которых соответствует одному сегменту изображения. При формиро-

вании базы данных оператор присваивает сегменту концептуальное понятие (*concept*), которое может быть снабжено определением (*modifier*). Тестовые комбинации *modifier/concept* определяют семантику сегмента. Таким же образом присваивается семантика всему изображению. В базе данных выявляется частотное отношение между семантикой изображения и семантикой сегментов. На основе энтропийной меры вторичные признаки в базе данных оцениваются по уровню информативности. Полученные оценки используются для агрегации вторичных признаков с целью повышения уровня их информативности. В результате на уровне сегментов образуются 20-25 сложных признаков (конкретное значение заведомо не определено и зависит от накопленной базы данных). Все сложные признаки имеют примерно одинаковый уровень информативности. Значение уровня информативности определяет степень доверия к сложному признаку и используется впоследствии для вычисления функций принадлежности к семантическому образу.

Вне зависимости от сегментного деления на уровне образа изображения формируются еще 18 признаков, которые образуют параллельную независимую систему информативных признаков. Эти признаки также адаптивно агрегируются в сложные признаки. В результате образуется 5-7 сложных признаков «уровня образа». Таким образом, семантика образа распознается по двум системам признаков.

На уровне образов это прямое распознавание, которое производится по системе сложных признаков «уровня образа» методами метрической классификации. Результаты накапливаются в базе данных, и их частотность определяет степень доверия к результатам распознавания.

На уровне сегментов распознавание семантики образа выполняется в два этапа. На первом этапе распознается семантика сегментов изображения подобно тому, как это делается для признаков «уровня образа». А на втором этапе на основе найденного ранее по базе данных статистического отношения между семантиками сегментов и образов определяются функции принадлежности для семантики образов.

Результаты по обеим системам признаков объединяются, образуя окончательный результат семантического распознавания. В целом система распознавания неявно реализует нечеткую нейронную сеть, в которой выходными лингвистическими переменными являются семантики образов.

## 3. Реализация алгоритмов и результаты экспериментов

Алгоритм декодирования JPEG-изображения реализован в соответствии с процедурами установленными рекомендациями T.81, T.83, T.84 [18-20] международного агентства в области коммуникаций ITU-T. Из допустимого множества профилей в рамках проведенного исследования в полном объеме реализован наиболее распространенный профиль *Baseline*, поддержанный рекомендацией JFIF [21]. Процедура сжатия данных основана на блочном кодировании изображений реализуемого на основе дискретного косинусного преобразования (DCT) над блоком размером 8×8 (см. рис. 2).

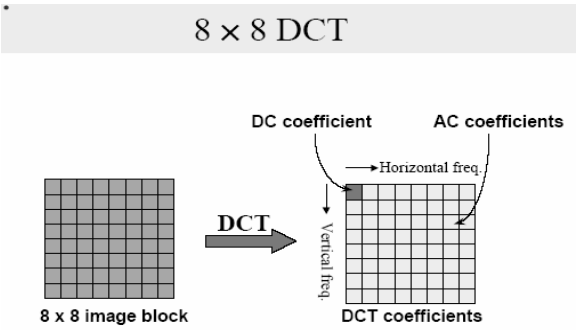



Рис. 2. Блочное кодирование. DCT – дискретное косинусное преобразование, DC – нулевой спектральный коэффициент, AC – множество высокочастотных коэффициентов.

**3.1. Формирование первичных информативных признаков**. На выходе декодирующей процедуры блок изображения представлен тремя матрицами размером 8×8. Каждая матрица содержит косинусный двумерный спектр $C_k(i,j)$ ($k$=1,2,3) в цветовом пространстве $YCbCr$. Двумерное косинусное преобразование определяется выражениями:

$$C_k(i,j) = \frac{1}{4}\sum_{x=0}^{7}\sum_{y=0}^{7} f_k(x,y)\cos\frac{(2x+1)i\cdot\pi}{16}\cos\frac{(2y+1)j\cdot\pi}{16}, \qquad i,j>0,$$

$$C_k(0,0) = \frac{1}{8}\sum_{x=0}^{7}\sum_{y=0}^{7} f_k(x,y), \quad i,j=0,$$

где $x, y$ координаты пикселей компоненты изображения $f_k$.

Из матриц косинусного спектра формируется четыре векторных признака:

- **DCf[3]** – цветовой фон блока – трехмерный вектор, составленный из значений постоянных составляющих (DC-составляющих) текущего блока. $DCf(k) = C_k(0,0)$.

- **ACf[3]** – вариабельность цветности – трехмерный вектор, составленный из эффективных значений переменных составляющих цветности для трех компонент блока изображения:

$$ACf(k) = \sqrt{\sum_{i=1}^{7}\sum_{j=1}^{7} C_k(i,j)^2},$$

где $C_k(i,j)$ – спектральные коэффициенты блока для $k$-й компоненты блока изображения.

- **Cont[3]** – цветовой контраст – трехмерный вектор, координаты которого определяются правилом:

$$Cont(k) = \frac{1}{\sqrt{2}}\frac{ACf(k)}{DCf(k)}.$$

Если все пиксели блока изображения имеют одинаковую цветность, то векторные признаки $ACf$ и $Cont$ равны нулевому вектору.

**- Angle[3]** – доминирующая угловая ориентация градиента цветности. Для вычисления признаковых координат для каждой компоненты вычисляются оценки градиента по трем направлениям:

Горизонтальное направление $G_h(k) = \sqrt{\sum_{j=1}^{7} C_k(0,j)^2}$ .

Вертикальное направление $G_v(k) = \sqrt{\sum_{i=1}^{7} C_k(i,0)^2}$ .

Наклонное направление (под углом π/4) $G_d(k) = \sqrt{\sum_{i=1}^{7} C_k(i,i)^2}$ .

Из трех оценок определяется максимальная, и координате вектора признака присваивается положительное значение угла, соответствующее максимальной оценке. Таким образом, для каждой координаты вектора признаков возможны три значения $0, \pi/4, \pi/2$ .

**3.2. Алгоритм сегментации изображения.** DC-коэффициенты блоков изображения по яркостной и двум цветоразностным компонентам образуют множество векторов размерности 3. Множество кластеризуется метрическим методом, расстояние между векторами определяется выражением:

$$L = \sum_{i=1}^{3} |x_i - y_i| / D_i ,$$

где $x_i, y_i$ – координаты векторов, $D_i$ – диапазон вариаций значений по координате. Расстояние между классами оценивается по принципу «максимально-удаленных соседей».

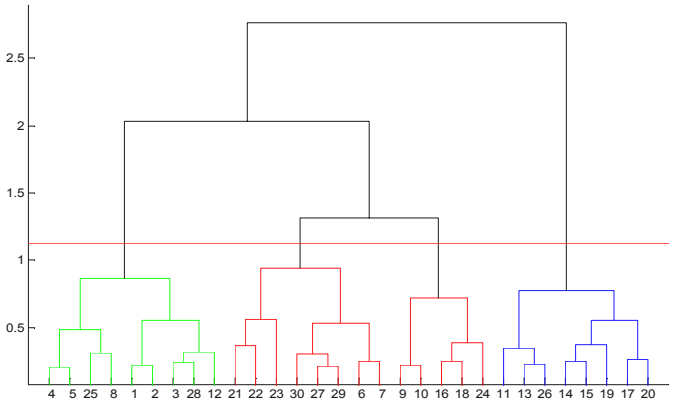



Рис. 3. Дендрограмма сегментации изображения. Горизонтальная линия соответствует адаптивному уровню разделения.

На рис.3 представлена примерная дендрограмма, где ветви дерева отражают кластерную иерархию множества из 30 точек. На горизонтальной оси дендрограммы представлено множество классифицируемых точек, по вертикальной оси отложены расстояния между точками в цветовом пространстве. Горизонтальные дуги объединяют близкие точки в кластеры. В МАТЛАБе для построения дендрограммы используется встроенная функция *dendrogram* из расширения *Statistic Toolbox*. Разделение на классы производится адаптивно, линия кластерного раздела на дендрограмме кластерных расстояний (см. рис. 3) проводится, когда отношение двух смежных уровней дендрограммы падает ниже 85%. По визуальным оценкам этот порог обеспечивает результаты, близкие к субъективным решениям человека.

На рис. 4 показан результат сегментации изображения («оранжевая роза»), соответствующий выше приведенной дендрограмме. Блоки 8×8 помечены точечными маркерами различного вида, каждый из которых отвечает одному из четырех образованных сегментов изображения.

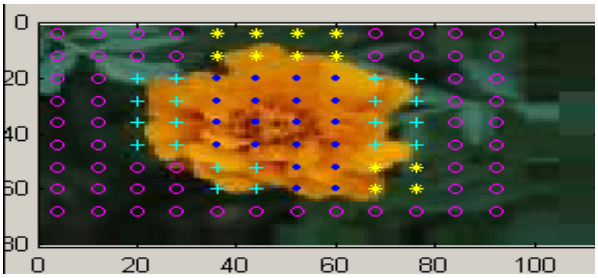



Рис. 4. Сегментированное изображение

**3.3. Формирование вторичных информативных признаков.** Система вторичных признаков строится на основе первичных информативных признаков. Характерная особенность вторичных признаков состоит в том, что все они имеют определенный физический смысл, и в той или иной мере отражают логическую информацию, которую человек использует при анализе семантики изображения.

**3.3.1. *Признаковое пространство сегмента.*** Каждый блок сегмента интерпретируется как точка в дискретной плоскости блоков. Координаты $(r,c)$ точки отсчитываются от левой и верхней грани изображения и выражаются в числе блоков. Площадь сегмента (далее обозначается через $S$) определяется числом входящих в него точек плоскости блоков. Обозначим через $R$ и $C$ размеры изображения в блоках, по вертикали и горизонтали соответственно. Признаковое пространство для сегмента включает характеристики, представленные в табл. 1.

**3.3.2. *Признаковое пространство полигона.*** Известно, что при восприятии изображения человеческий глаз последовательно фиксирует взгляд на ряде характерных точек [22]. Точка фиксации взгляда – это область изображения, ми-

нимально достаточная для выделения человеческим глазом ее отличительных особенностей. При анализе JPEG-изображения естественно принять, что размер точки фиксации взгляда определяется размером одного блока изображения. Число точек фиксации взгляда зависит от уровня сложности изображения, однако в любом случае это число значительно меньше, чем число блоков изображения. Множество точек фиксации взгляда для изображения в целом или его фрагмента далее называется полигоном. При моделировании данной функции человеческого восприятия была выбрана упрощенная схема, закрепляющая число точек полигона в зависимости от представительности объекта. Принято, что полигон всего изображения состоит из 21 точки, полигон сегмента состоит из 14 точек, полигон непрерывной области сегмента состоит из 7 точек. На рис. 5 показаны точки полигона для яркостной компоненты сегмента граничных областей. Точки условно соединены ребрами графа.

**Табл. 1. Информативные признаки сегмента**

Param.potential = $S/R \cdot C$ – потенциал сегмента равен относительной площади, занимаемой сегментом в поле изображения.

Param.YCbCr_DC=CL(k) – массив из трех элементов; содержит яркостную и цветоразностные компоненты Y,Cb,Cr для доминирующего фона сегмента.

Param.YCbCr_AC=aYCbCr – вариабельность цветности сегмента, трехмерный вектор, координаты которого равны среднеквадратичным значениям цветовой вариабельности блоков сегмента, для вычисления значений используется следующее правило:

$$aYCbCr(i) = \frac{1}{S}\sqrt{\sum_{s} ACf(i)^2} \ ,$$

Param.LPosition=LPosition – структура, содержащая логические позиции сегмента в поле изображения.

LPosition.top=1 – сегмент касается верхней границы изображения

LPosition.bottom=1 – сегмент касается нижней границы изображения

LPosition.left=1 – сегмент касается левой границы изображения

LPosition.right=1 – сегмент касается правой границы изображения

LPosition.center=1 – сегмент имеет блоки, принадлежащие области центра

LPosition.quadrant1=1 – сегмент имеет блоки, принадлежащие первому квадранту

LPosition.quadrant2=1 – сегмент имеет блоки, принадлежащие второму квадранту

LPosition.quadrant3=1 – сегмент имеет блоки, принадлежащие третьему квадранту

LPosition.quadrant4=1 – сегмент имеет блоки, принадлежащие четвертому квадранту

Param.Narea – число изолированных областей в сегменте

Координаты (x,y) точек полигона отсчитываются от левой и верхней грани изображения и выражаются в пикселях с точностью до размера блока. Точки полигона выделяются по значению модулей векторов либо цветовой вариабельности $|ACf|$, либо цветового контраста $|Cont|$, либо угловой ориентацией градиента. Выбор критерия в исследовательской моделирующей программе осуществляет оператор. При закрепленном числе точек полигона пороговый уровень выделения точек также является информативным признаком.

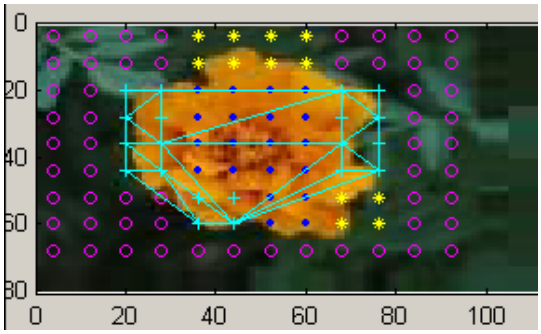



Рис. 5. Полигон сегмента граничных областей.

По полигону строится 12 векторных вторичных признаков, характеризующих его форму. Набор признаков, описывающих полигон, представлен в табл. 2.

**Табл. 2. Информативные признаки полигона**

| |
|---|
| Polygon.Potential=S/(X*Y) – матрица потенциалов полигонов<br>Polygon.Diameter=D/sqrt(X^2+Y^2) – матрица относительных диаметров полигонов (относительно диагонали отображения)<br>Polygon.FormFactor =S/D^2 – фактор формы<br>Polygon.DAngle – углы ориентаций максимального диаметра для полигонов<br>Polygon.EAngle=EAngle – углы ориентаций вектора главной компоненты полигонов.<br>Polygon.Radius=Radius – максимальные относительные радиусы для полигонов<br>Polygon.EFactor=EFactor – отношение минимального диаметра аппроксимирующего эллипса полигона к максимальному диаметру.<br>Polygon.Mass_center=Mass_center./[X,Y] – матрица относительных координат центра масс точек полигона.<br>Polygon.Polygon_center=Polygon_center./[X,Y] – матрица относительных координат геометрического центра полигона (координаты центра масс выпуклой оболочки полигона).<br>Polygon.Level=Level – пороговые уровни выделения точек полигонов.<br>Polygon.MaxValue – максимальные значения уровней выделения для точек полигона.<br>Polygon.Points=Points – массив ячеек, содержащих координаты точек полигона<br>Polygon.Graph – массив ячеек, содержащий матрицы смежностей графа Делоне для точек полигона |
| S – площадь полигона в пикселях<br>X*Y – размеры изображения в пикселях<br>D – длина максимальной диагонали изображения |

Каждое поле структуры *Polygon* имеет три компоненты. Координаты центра масс и центра масс выпуклой оболочки полигона представляют собой матрицы размером 3×2. Все координаты представлены в относительных единицах, нормировка выполняется по отношению к фактическим размерам изображения. Линейные размеры нормируются по отношению к максимальной диагонали изображения. Следующие математические выражения используются для формирования признакового пространство полигона:

- *Относительный центр масс полигона* – определятся его координатами: $x_c = \frac{1}{X \cdot N}\sum_{i=1}^{N} x_i, \qquad y_c = \frac{1}{Y \cdot N}\sum_{i=1}^{N} y_i$

где $X, Y$ размеры изображения, выраженные в пикселях, $N$ – число точек полигона.

- *Относительный диаметр полигона* – равен отношению максимального диаметра полигона к диагонали изображения, где диагональ изображения вычисляется выражением $Diagonal = \sqrt{X^2 + Y^2}$ .

- *Доминирующее направление полигона* – выраженный в радианах угол наклона диаметра полигона к положительному направлению горизонтальной оси изображения.

- *Относительный радиус полигона* – максимальное расстояние точки полигона от его центра масс.

- *Коэффициент формы* – отношение площади полигона к квадрату его диаметра.

- *Коэффициент деформации* – отношение длины минимального диаметра эллипса (аппроксимирующего выпуклую оболочку полигона) к максимальному диаметру эллипса. Аппроксимирующий эллипс определяется главными компонентами множества точек выпуклой оболочки полигона. Выпуклая оболочка включает в себя только те точки полигона, которые образуют максимальную выпуклую поверхность.

- *Геометрический центр* – центр масс выпуклой оболочки полигона.

- *Угол наклона аппроксимирующего эллипса* – выраженный в радианах угол наклона максимального диаметра аппроксимирующего эллипса к положительному направлению горизонтальной оси изображения.

Для выделения выпуклой оболочки и формирования параметров полигона используются встроенные МАТЛАБ функции: *convhull, pdist, mean, princomp*. Полигоны строятся для сегмента в целом, каждой непрерывной компоненты сегмента и для всего изображения в целом. На рис. 6 показан полигон уровня изображения для яркостной компоненты.

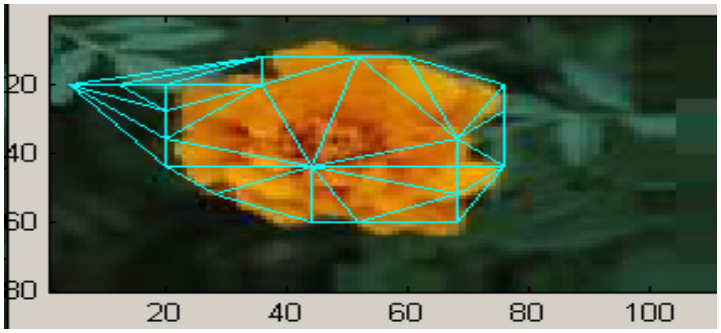



Рис. 6. Полигон уровня изображения

В пространство признаков сегмента включается только максимальная непрерывная компонента, матрица смежностей не используется и зарезервирована для будущих версий. Всего для количественной характеристики сегмента используется 100 вторичных признаков в покоординатном исчислении. По полигону уровня изображений формируется 39 вторичных признаков. Процедура формирования вторичных информативных признаков используется как на этапе накопления базы данных, так и на этапе распознавания.

На этапе накопления базы данных, для каждого сегмента и каждого образа оператор вводит концептуальные понятия (*concept*) и определения к ним (*modifier*) в виде текстовых слов. Эта пара слов считается семантикой изображения или его сегмента. Вместе с вектором признаком эта информация сохраняется в базе данных. По базе данных строится нечеткое отношение *Transit* между семантикой сегмента и семантикой образа. Значения матрицы нечеткого семантического отношения определяется числом совместного проявления в базе данных семантики сегмента с определенной семантикой образа.

**3.4. Алгоритм формирования сложных признаков**. Сложные признаки – это третий уровень в иерархии пространств информативных признаков. Этот уровень необходим, для того чтобы перед выполнением процедуры классификации сформировать признаковое пространство, анизотропное по информативности.

**3.4.1.** ***Оценка информативности.*** Формирование сложных признаков выполняется на основе анализа информативности признаков в накопленной базе данных. Для анализа информативности признаков используется мера взаимной информации признака [23] с верифицированной семантикой сегмента и/или изображения в целом. Предварительно количественные признаки квантуются с точностью 5% от диапазона изменения, а семантика сегментов и образов кодируется числовыми кодами. В итоге поле признаков приводится к номинальной шкале, где каждое значение можно считать буквой. Для количественных признаков число букв в алфавите равно 20. Для семантики число букв определяется числом различных семантических понятий в базе данных. Оценка информативности признаков выполняется на основе информационной матрицы, по одной координате которой представлены коды признака, а по другой коды семантического образа, пример подобной матрицы показан ниже.

| $x/y$ | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | $m_{00}$ | $m_{01}$ | $m_{02}$ | $m_{03}$ | $m_{04}$ | $m_{05}$ | $m_{06}$ | $m_{07}$ | $m_0$ |
| 1 | $m_{10}$ | $m_{11}$ | … | | | | | … | $m_1$ |
| 2 | … | … | | | | | | | $m_2$ |
| 3 | … | … | | | | | | | $m_3$ |
| 4 | $m_{40}$ | $m_{41}$ | … | | | | | … | $m_4$ |
| | $n_0$ | $n_1$ | $n_2$ | $n_3$ | $n_4$ | $n_5$ | $n_6$ | $n_7$ | $M$ |

Данную матрицу можно рассматривать как накопленную по базе данных статистику работы канала передачи информации. Каждый элемент матрицы указыва-

ет на число переходов $i$-ой буквы алфавита признака $x$ в $j$-ю букву алфавита признака $y$. Числа $m_0 \cdots m_4$ – (суммы по строкам) определяют композицию признака $x$. Числа $n_0 \ldots n_7$ – (суммы по столбцам) определяют композицию признака $y$. Величина $M = \sum_j n_j = \sum_i m_i$ равна числу переданных символов. Энтропии признаков определяются выражениями:

$H_x = -\sum_i \frac{m_i}{M} \log \frac{m_i}{M}$ – энтропия входа,

$H_y = -\sum_j \frac{n_j}{M} \log \frac{n_j}{M}$ – энтропия выхода,

$H_{x \otimes y} = -\sum_i \sum_j \frac{m_{ij}}{M} \log \frac{m_{ij}}{M}$ – энтропия канала,

$H_{x:y} = H_x + H_y - H_{x \otimes y}$ – взаимная энтропия.

Известно [23], что $H_{x:y} \le H_x$ и $H_{x:y} \le H_y$. Взаимная энтропия характеризует степень похожести признаков $x$ и $y$. Удобно использовать нормированную величину $S_{x:y} = H_{x:y} * 2 \big/ \left( H_x + H_y \right)$, которая изменяется в диапазоне $[0,1]$. Если $S_{x:y} = 1$, то признаки $x$ и $y$ совпадают (с точностью до замены букв). Если $S_{x:y} = 0$, то признаки $x$ и $y$ сильно различаются. Величина $S_{x:y}$ (когда $x$ является измеряемым информативным признаком, а $y$ – семантикой сегмента) рассматривается как степень надежности признака $x$ и используется далее для вычисления функции принадлежности для лингвистических переменных. Данный метод позволяет выполнить оценку информативности одиночных признаков по отношению к семантике сегмента или образа и отбросить слабые признаки.

**3.4.2.** ***Агрегирование признаков.*** Взаимная энтропия используется также для оценки связи между измеряемыми информативными признаками. Признаки независимы, если их взаимная энтропия равна нулю. Агрегация независимых информативных признаков в сложный признак приводит к усилению разделяющей способности по отношению к семантике образа [23]. Поэтому на этапе классификации все признаки подвергаются агрегации. При равном уровне информативности эффект от агрегации тем выше, чем выше степень взаимной независимости признаков. Агрегация выполняется как адаптивная кластеризующая процедура, реализующая последовательное объединение наиболее независимых признаков, при этом используется тот же механизм адаптации, что и при сегментировании изображения. Мерой расстояния между признаками $x$ и $y$ является величина $1 - S_{x:y}$. В табл. 3 демонстрируется эффект агрегации на примере 5 признаков полигона образа. Числовые значения в таблице равны степени информативности $S_{x:y}$ одиночных и сложных признаков по отношению к семантике образа. В контексте нечеткой классификации информативность трактуется как степень достоверности признака.

Структура сложных признаков переопределяется после каждого изменения базы данных.

**Табл. 3. Информативность признаков**

| Признаки полигона | Одиночный признак | Сложный признак |
|---|---|---|
| FormFactor | 0.659 | 0.8719 |
| EFactor | 0.677 | |
| Potential | 0.673 | 0.9222 |
| Radius | 0.651 | |
| Mass_center (y) | 0.637 | |
| Polygon_center (y) | 0.643 | |

**3.5. Семантическая классификация**. Семантическая классификация выполняется в признаковом пространстве третьего уровня. Предъявляемое изображение разделяется на сегменты, и по каждому сегменту вычисляются значения сложных признаков, которые представляют собой векторы небольшой размерности. Размерность признакового пространства переопределяется заново при пополнении базы данных.

**3.5.1.** ***Нечеткая классификация сегментов.*** Каждый сегмент изображения обрабатывается по той же схеме, что и при формировании базы данных. Текущее значение сложного признака сравнивается со значениями одноименного признака всех верифицированных сегментов базы данных (прецедентами). В результате сравнения определяются стандартные евклидовы расстояния между текущими и верифицированными значениями. Уровень достоверности классификации по данному сложному признаку определяется величиной:

$$\xi_i = 1 - \min_k \left( d_{ki} \right) \Big/ \max_k \left( d_{ki} \right),$$

где $d_{ki}$ – расстояние текущего значения $i$ -го признака до $k$ -го прецедента. Уровень достоверности максимален и равен 1, когда в базе данных существует прецедент, совпадающий с текущим значением признака. Функция принадлежности к прецеденту вычисляется по правилу:

$$\eta_k = \frac{1}{p} \sum_{i=1}^{p} \min \left( \xi_i, \chi_i \right),$$

где $p$ – число сложных признаков на уровне сегментов, $\chi_i$ – уровень достоверности сложного признака. Функция принадлежности к семантике $s$ определяется выражением:

$$\mu_s = \max_{\eta_k \in K_s} \left( \eta_k \right),$$

где $K_s$ – множество прецедентов с семантикой $s$. Значения функции $\mu_s$ определяют степени принадлежности анализируемого сегмента термам лингвистической переменной «семантика сегмента». На заключительном этапе выделяется терм, которому соответствует максимальное значение функции принадлежности. Семантика этого терма рассматривается как окончательный результат классификации сегмента, а соответствующее значение функции принадлежности определяет уровень его достоверности.

***3.5.2. Нечеткая классификация образов по семантике сегментов.*** Семантическая классификация образов выполняется на основе матрицы нечеткого отношения *Transit* между семантикой сегментов и семантикой образов. Просматриваются семантики классифицированных сегментов, и выполняется вычисление функции принадлежности к семантике образов, по следующему правилу:

$$\mu_{\mathrm{Im}}(i) = \frac{1}{n \cdot m(i)} \sum_{k=1}^{n} \mu_s(k)\, Transit(i,k),$$

где $n$ – число сегментов в образе, $\mu_s(k)$ – уровень достоверности семантики $k$-го сегмента, $m(i) = \max_k \left( Transit(i,k) \right)$ – нормирующий множитель.

***3.5.3. Нечеткая классификация образов по параметрам полигона образа.*** По каждому сложному признаку полигона образа выполняется метрическая классификация по всем прецедентам базы данных. Текущее значение сложного признака сравнивается со значениями одноименного признака всех прецедентов. В результате сравнения определяются стандартные евклидовы расстояния между текущими и верифицированными значениями. Уровень достоверности классификации по каждому сложному признаку определяется величиной:

$$\xi_i = 1 - \min_k (d_{ki}) \Big/ \max_k (d_{ki}),$$

где $d_{ki}$ – расстояние текущего значения $i$-го признака до $k$-го прецедента. Достоверность максимальна и равна 1, когда в базе данных существует прецедент, совпадающий с текущим значением признака. Функция принадлежности к прецеденту вычисляется по правилу:

$$\eta_k = \frac{1}{g} \sum_{i=1}^{g} \min(\xi_i, \chi_i),$$

где $g$ – число сложных признаков на уровне сегментов, $\chi_i$ – уровень достоверности сложного признака. Функция принадлежности к семантике образов $p$ определяется выражением

$$\mu_s = \max_{\eta_k \in K_s} (\eta_k),$$

где $K_p$ – множество прецедентов с семантикой $s$. Значения функции $\mu_p$ определяют степени принадлежности анализируемого образа термам лингвистической переменной «семантика образа».

***3.5.4. Комплексирование результатов.*** Результаты семантической классификации, полученные по двум параллельным системам признаков, конъюнктивно объединяются. Результирующая функция принадлежности определяется правилом:

$$\mu = \min\left( \mu_{\mathrm{Im}}, \mu_p \right).$$

Если множество диагнозов пусто или максимальное значение функции принадлежности меньше заданного порога, то констатируется отказ от классификации. Данная ситуация возникает, когда база данных недостаточно представительна.

**3.6. Результаты эксперимента.** Для проведения экспериментальных исследований была разработана программа классификации JPEG-изображений в среде МАТЛАБ. Программа позволяет также формировать базу данных и выполнять оценки информативных признаков. Рабочее окно программы показано на рис. 7. При декодировании изображений в правом информационном поле отображаются обнаруженные маркеры JPEG-формата. Основным управляющими элементами интерфейса являются управляющие кнопки ***Decode JPEG-file*** – запуск процедуры декодирования файла, ***Dialog*** – вызов окна управления базой данных, ***Classify*** – запуск процедуры классификации. Диалоговое окно, представленное в нижнем правом углу скриншота, реализует ввод семантических понятий и управление базой данных.

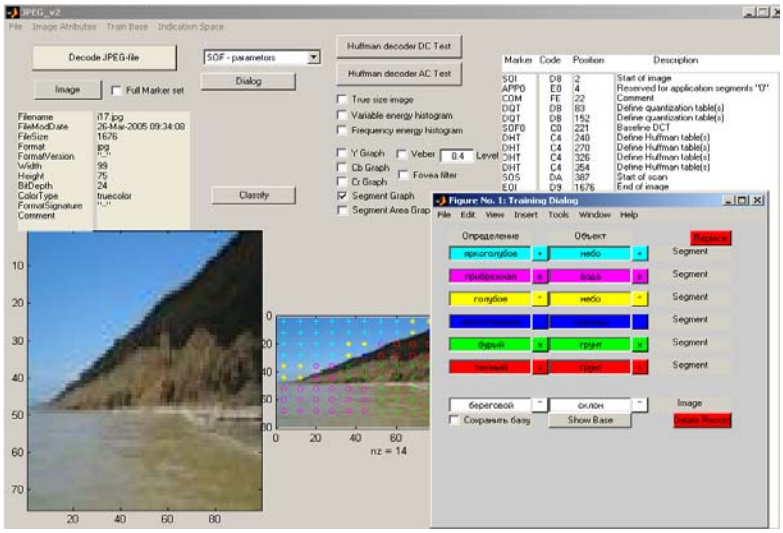



Рис. 7. Скриншот рабочего окна

Для проведения экспериментов с помощью данной программы была создана база данных, содержащая информацию по 100 изображениям. Характеристики базы данных представлены в табл. 4.

**Табл. 4. Характеристика базы данных**

| |
|---|
| Число примеров в базе данных=100 |
| Число сегментов в базе данных=535 |
| Число концептуальных понятий в базе данных=86 |
| Число модификаторов понятий в базе данных=154 |
| Число сочетаний *Concept/Modifier* на уровне сегментов=214 |
| Число сочетаний *Concept/Modifier* на уровне образов=67 |
| Средний размер изображений 120*120 пикселов |

Из таблицы видно, что представительность базы данных в среднем не высокая; на каждое семантическое понятие *Concept/Modifier* на уровне образов приходится 1.5 примера, а для уровня сегментов 2.5 примера. Однако некоторые концепты представлены достаточно полно. В табл. 5 показана частотность наиболее представительных концептов на уровне сегментов. Результаты экспериментов показали, что изображения, которые содержат наиболее представительные концепты базы данных, достаточно уверенно распознаются по семантике.

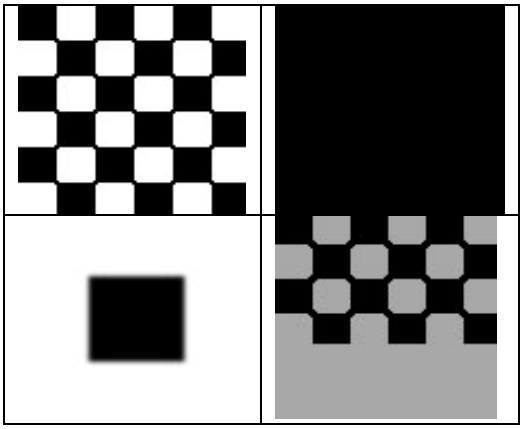
**Табл. 5. Представительность базы данных**

| **Частотность Concept/Modifier на уровне сегментов** | **Частотность Concept на уровне сегментов** |
|---|---|
| цветотеневая граница 54 | небо 111 |
| текстурный фрагмент 20 | граница 81 |
| голубое небо 20 | листва 43 |
| белесое небо 17 | вода 24 |
| облачное небо 16 | грунт 23 |
| светло-голубое небо 16 | берег 21 |
| зеленая листва 12 | фрагмент 20 |
| темно-зеленая листва 12 | пятно 17 |

Ниже приведены два характерных примера работы алгоритма семантической классификации.

*1. Текстурный фрагмент.* В первом, втором и третьем квадранте табл. 6 приведены образы, входящие в базу данных, в четвертом квадранте (правый нижний угол) представлен тестовый образ.

**Табл. 6. Текстурный фрагмент**



На рис. 8 показан скриншот результата классификации. Левое окно – результат классификации по признакам полигона, среднее окно – результат классификации по признакам сегментов, правое окно – комплексный результат.

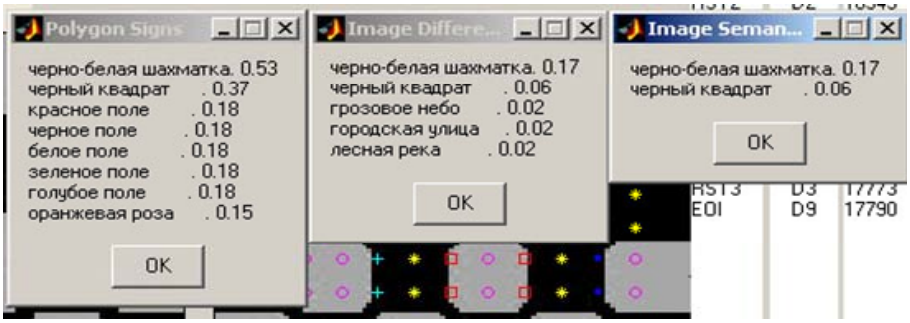



Рис. 8. Результаты классификации

Числа определяют значения функции принадлежности лингвистическому терму. При выводе на экран результаты семантической классификации образов упорядочиваются по убыванию функции принадлежности.

2. *Лесная река*. В первом, втором и третьем квадрантах табл. 7 выборочно показаны примеры, входящие в базу данных, в четвертом квадранте представлено тестовое изображение.

**Табл. 7. Лесная река**

Результаты семантической классификации тестового изображения показаны на рис. 9. Правое окно соответствует комплексному результату. Первую позицию с наибольшим значением достоверности занимает семантика «лесная река», что соответствует фактическому содержанию тестируемого изображения. Интересно отметить, что остальные результаты оказались по смыслу близкими к достоверному семантическому значению.

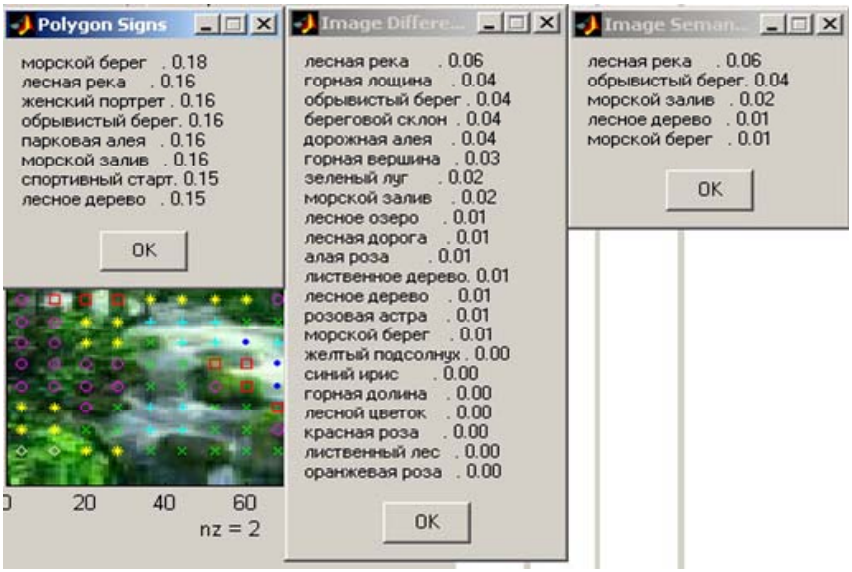



Рис. 9. Результаты классификации образа «Лесная река»

## 4. Выводы и обсуждение результатов

Проведенные исследования доказали принципиальную возможность реализации процедуры семантической классификации изображений в признаковом пространстве JPEG-формата. Исключение процедуры полного восстановления изображения является принципиальным решением, которое дает возможность обеспечить высокое быстродействие классифицирующих подсистем. Для экспериментальных исследований в проекте использовалась программная среда MATLAB, что было обусловлено богатством ее функциональных возможностей в программном и пользовательском интерфейсе и высоким уровнем оперативности при отладке программ. Однако данная среда является интерпретирующей, поэтому скорость выполнения программ существенно ниже, чем в любой компилирующей среде.

Примерные оценки быстродействия дают следующие цифры. Программная процедура, написанная на языке MATLAB, выполняла полное декодирование и сегментацию JPEG-изображения размером 100*150 пикселей за 10 сек (на процессоре Intel 2,4 ГГц). Семантическая классификация изображения с отображением результата выполнялась за 7 сек. Встроенная функция MATLAB, также реализующая полное JPEG-декодирование, выполняется за 0.1 сек. Поэтому можно ожидать, что время выполнения процедуры семантической классификации в компилирующей программной среде для данного размера изображения будет также на уровне 0.1 сек.

Представительность базы данных должна быть существенно выше, чем в проведенных экспериментах. Удовлетворительная классификация может быть

обеспечена при среднем уровне 15-20 образов на семантический класс. Накопление базы данных – процесс достаточно длительный, поскольку связан с действиями оператора. На описание одного изображения (в сегментном представлении) оператором тратится 2-4 минуты. Кроме того, семантическое описание подвержено субъективным взглядам оператора. Субъективность накопленных данных можно уменьшить двумя путями: 1) объединением баз данных созданных различными операторами, 2) добавлением уровня онтологий, определяющего отношения между семантиками. Уровень онтологий может быть выстроен в автоматическом режиме по совпадениям концептов или модификаторов. Этот путь является более перспективным, поскольку позволяет без расширения базы данных добавить еще один уровень в иерархию семантической классификации, увеличив тем самым достоверность ее работы.

## 5. Литература

1. Smith J.R., S.F.Chang Visualseek: A fully automated content based image query system. In: Proc. ACM Multimedia, Boston, MA, Nov, 1996.

2. Milind R.N., T.S.Huang Extracting Semantics from Audiovisual Content: The Final Frontier in Multimedia Retrieval.

3. Salton G., McGill M.J. Introduction to Modern Information Retrieval. New York: McGraw-Hill, 1983.

4. Rui Y., Hyang T.S. and Mehrota S. Content-based image retrieval with relevance feedback in MARS. In: Proc.IEEE Int. Conf. Image Processing, Santa Barbara, CA, 1997, pp. 815-818.

5. Celentano A., Sciasico E.D. Feature integration and relevance feedback analysis in image similarity evalution. J.Electron. Imaging, vol. 7, no.2, pp. 308-317, 1998.

6. Geveres T., Smeulders A.W.M. PieToSeek: Combining color and shape invariant features for image retrieval. IEEE Trans. Image Processing, vol. 9, pp. 102-119, 2000.

7. Sciascio E.Di., Mongiello. DrawSearch: A tool for interactive content-based image retrieval over the net. Proc. SPIE, vol. 3656, pp. 561-572, 1999.

8. Můller H., Můller S., Marcand-Maillet, and Squire D. McG. Strategies for positive and negative relevance feedback in image retrieval. In: Proc. Int. Conf. Pattern Recognition, Barcelona, Spain, 2000.

9. Porkaew S., Mehrotra S., Ortega M. and Chakrabarti K. Similarity search using multiple examples in MARS. In: Visual Information and Information Systems. New York: Springer-Verlag, 1999, pp. 68-75.

10. Ciocca G., Schettini R. Using a relevance feedback mechanism to improve content-based image retrieval. In: Visual Information and Information Systems, New York: Springer-Verlag, 1999, pp. 105-114.

11. Muneesawwang P., Guan L. Anonliner RBF model for interactive content-based image retrieval. In: Proc. 1st IEEE Pasific-Rim Conf. Multimedia, Syney, Australia, Dec. 2000, pp. 188-191.

12. Ei-Nada, Wernick M.N., Yang. Y., Galatsonos N.P. Image retrieval based on similarity learning. In: Proc. IEEE. Conf. Image Processing, vol. 3. Vancouver, BC, Canada, 2000, pp.772-775.

13. Peng J., Bhanu B., Oing S. Probabilistic feature relevance learning for content-based image retrieval. Comput. Vision Image Understanding, vol. 75, no. ½ pp. 150-164, 1999.

14. Sclaroff S., Taycher L., Cascia M.L. ImageRover: A content-based image browser for the world wide web. In: Proc. IEEE Workshop Content-Based Access Image Video Libraries, Puerto Rico, June 1997, pp. 2-9.

15. Muneesawwang P., Guan L. Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture. IEEE Trans. on Neural Networks, vol. 13, no 4, july 2002.

16. Бодякин В.И. "Куда идешь, Человек? Основы эволюциологии. (информационный подход), М.: СИНТЕГ, 1998, 332с.

17. Бодякин В.И. "Исследование структурных моделей открытых динамических систем", специальность: 05.13.01 (Управление в технических системах), автореферат диссертации и диссертация на соискание ученой степени к.ф.-м.н., Москва – 1999г.

18. T.81 - Information technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines

19. T.83 - Information technology – Digital compression and coding of continuous-tone still images: Compliance testing.

20. T.84 - Information technology – Digital compression and coding of continuous-tone still images: Extensions.

21. JFIF 1.02 (JPEG File Interchange Format).

22. Ю.К. Гаврилей, А.И. Самарин, М.А. Шевченко Активный анализ изображений в системах с фовеальным восприятием. – Нейрокомпьютеры в системах обработки изображений, №7-8, 2002.- С.34-46.

23. Гоппа В.Д. Введение в алгебраическую теорию информации. М.: Наука. Физматлит, 1995. 112с.

Статья поступила 31 мая 2006 г.
После доработки 29 августа 2006 г.