Построение соответствий между низкоуровневыми характеристиками и семантикой статических изображений

© Н. Васильева

Санкт-Петербургский
Государственный Университет
Natalia.Vassilieva@soft-werke.com

© Б. Новиков

Санкт-Петербургский
Государственный Университет
borisnov@acm.org

Аннотация

Качество систем поиска изображений по содержанию, использующих низкоуровневые характеристики изображения, до сих пор нельзя признать удовлетворительным. Многие исследователи видят проблему в «семантическом разрыве» между низкоуровневым содержанием изображения, которым оперирует система, и семантикой изображения, необходимой пользователю. Данная работа предлагает подход к построению соответствий между низкоуровневыми характеристиками и семантикой изображений.

1 Введение

Задача поиска статических изображений продолжает оставаться крайне актуальной на протяжении последних десятилетий. Несмотря на неослабевающий интерес исследователей и большое количество работ в данной области есть еще много открытых вопросов и нерешенных задач. В связи с все возрастающими объемами мультимедиа коллекций основное внимание исследователей уделяется алгоритмам поиска изображений по содержанию (Content Based Image Retrieval, CBIR), позволяющим полностью автоматизировать процесс построения индекса. В условиях отсутствия какой-либо дополнительной информации об изображении для построения индекса используются низкоуровневые характеристики самого изображения, такие как цвет, яркость, текстура. В последние годы было предложено большое количество различных моделей описания данных характеристик, функций расстояния и основанных на них оценок схожести изображений. Тем не менее, качество работы систем поиска изображений по содержанию все еще нельзя признать удовлетворительным. Человек, сравнивая изображения, сравнивает их семантику, смысловое наполнение, в то время как система производит оценку схожести изображений на основе низкоуровневых характеристик. Одним из приоритетных направлений исследований в данной области является разработка алгоритмов, направленных на уменьшение «семантического разрыва» между результатами анализа изображений системой и визуальным восприятием пользователя.

Данная работа описывает подход к решению задачи построения соответствий между низкоуровневыми характеристиками изображений и их семантикой. Результатом работы является таблица соответствий низкоуровневых и лексических характеристик изображений. Использование данной таблицы в процессе поиска изображений позволит наладить диалог между пользователем и системой. В частности, пользователь получит возможность формулировать запрос к системе на естественном языке.

Работа является одним из этапов построения системы поиска изображений по содержанию, ориентированной на пользователя. Описание общей архитектуры данной системы можно найти в [8].

2 Обзор существующих подходов

Вопросам индексирования и поиска изображений по содержанию посвящено огромное количество работ, указывающее на неослабевающую активность исследователей в данной области. Хороший обзор основных направлений исследований в CBIR представлен в работах [5, 7].

Системы поиска изображений по содержанию используют различные алгоритмы обработки цифрового сигнала для построения сигнатур – представлений изображения в виде ограниченного набора параметров. В большинстве предлагаемых подходов одна сигнатура описывает одну из ха-

Труды 7ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2005, Ярославль, Россия, 2005.

рактеристик изображения (цвет, текстура, и т.п.). Иными словами, одному и тому же изображению может соответствовать несколько сигнатур, каждая из которых описывает одну из характеристик.

При поиске в качестве запроса чаще всего выступает изображение-образец. Результатом поиска является набор изображений, близких к изображению-запросу по человеческим впечатлениям. Некоторые системы предлагают пользователю составить эскиз того изображения, которое он хочет найти [4]. Однако оба подхода заставляют пользователя формулировать свой запрос на языке, понятном системе, но не столь удобном для пользователя. Использование изображения-образца в качестве запроса предполагает наличие у пользователя такого образца, что далеко не всегда так. Составление эскиза искомого изображения заставляет пользователя переформулировать запрос на языке цвета и форм. Пользователю же намного проще формулировать свой запрос на естественном языке, описывая словами то, что он хочет видеть на искомом изображении. Однако на сегодняшний день не известны системы, позволяющие строить индекс автоматически на основе содержания изображения и предоставляющие возможность составления запроса на естественном языке.

Многие исследователи выделяют несколько уровней содержания изображений [1]. Цвет и яркость относят к низкоуровневому содержанию, физические объекты (такие как человек, машина, дерево) - к содержанию высокого уровня, текстуру часто называют содержанием среднего уровня. В литературе представлено большое количество различных подходов к моделированию низкоуровневых характеристик изображения и большинство систем CBIR построено на алгоритмах обработки низкоуровневого содержания. В последние годы исследователи пришли к пониманию недостаточности таких алгоритмов. Необходимо принимать во внимание особенности визуального восприятия человека, его ориентированность на семантику изображения.

Одним из распространенных подходов, призванных уменьшить разрыв между представлением изображения системой и восприятием пользователя являются системы обратной связи (relevance feedback) [6]. Основанные на итеративном процессе поиска, такие системы собирают статистику по реакции пользователя, которая в следующий раз будет учитываться при формировании ответа системы. Примером практической реализации такой системы может служить система, разработанная в лаборатории CLIPS [3]. Большое внимание в данной работе уделено также вопросам интерфейса для построения эффективного диалога с пользователем.

В работе [1] представлен ряд экспериментов, показавший возможность использования слов английского языка в качестве основы индекса изображений. Такой индекс хорошо согласуется с семантикой изображения. Однако авторы статьи не предлагают метода автоматического построения подобного индекса, а проиндексировать подобным образом большие коллекции изображений вручную не представляется возможным.

3. Предлагаемое решение

Задачей проводимого исследования является построение системы поиска изображений по содержанию, способной обрабатывать запрос на естественном языке. В качестве множеств индексируемых изображений мы рассматриваем коллекции любительских фотографий и исходим из предположения, что никакой дополнительной информации о семантике изображений не доступно. Для того чтобы система могла отвечать на текстовый запрос, мы предполагаем использовать таблицу соответствий между низкоуровневыми и лексическими характеристиками изображений, способ построения которой описывается в данной статье.

Авторами работы [1] экспериментальным путем было показано, что «лексические базисные функции», представляющие собой набор слов естественного языка (авторы использовали английский язык) являются хорошим инструментом для описания семантики изображения. В данной работе для описания смыслового наполнения изображения также используются наборы ключевых слов – лексические характеристики. Для описания содержания изображения на данный момент используется только цвет. Цветовая характеристика является основной для естественных изображений для визуального восприятия человека.

Для построения таблицы соответствий сначала были вычислены базисные цветовые характеристики на основе отобранного обучающего набора изображений. Данные характеристики описывают закономерности распределения цвета для групп схожих изображений. Далее для каждой из групп были отобраны лексические характеристики (ключевые слова), наиболее четко описывающие семантику изображений группы. Сопоставив базисные цветовые и лексические характеристики групп схожих изображений, мы получили таблицу, описывающую соответствие словесных описаний изображений их цветовому содержанию.

3.1 Построение базисных цветовых характеристик

Для построения базисных цветовых характеристик было отобрано 300 любительских фотографий с пейзажами и видами городов. Такое ограничение на содержание изображений было сделано ввиду использования только цветовых характеристик для анализа изображений. Эти

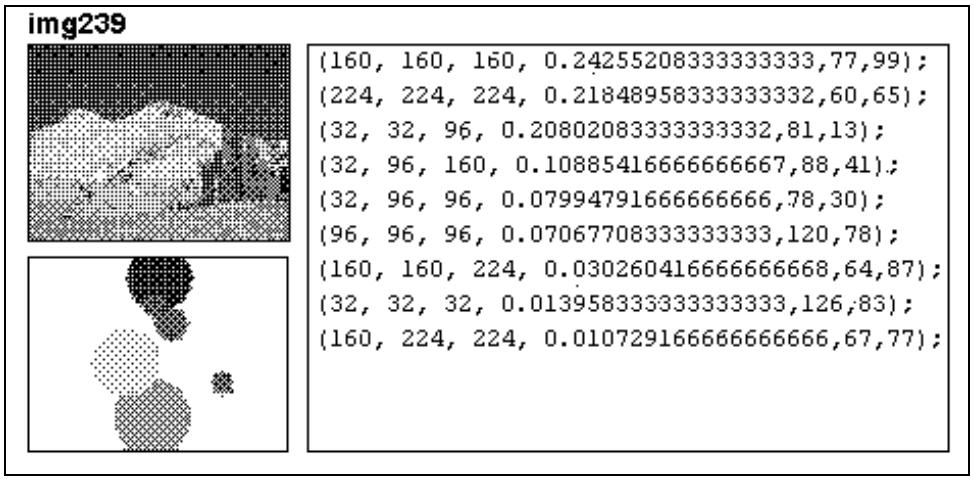


Рис. 1 Изображение и его цветовая характеристика

виды изображений наиболее полно описываются на основе одной информации о цвете, в то время как для описания других групп изображений (сцены интерьера, изображения людей) необходим также анализ текстуры и форм.

Отобранные изображения были разбиты на кластеры на основе их цветовых характеристик. Для построения кластеров была выбрана относительно простая модель представления цветовой характеристики изображения. Кластеризация является более простой задачей, чем поиск схожих изображений по изображению-запросу, поэтому для кластеризации представляется возможным использование такой модели изображения.

Цветовое пространство RGB было разбито на 64 куба – цветовых промежутка. Каждая цветовая ось (r, g, b) была разбита на 4 равных отрезка. Для каждого цветового промежутка было вычислено количество пикселей изображения, попадающих в промежуток, а также определено пространственное расположение соответствующего цветового пятна.

Для каждого изображения из обучающего набора была вычислена его цветовая характеристика, задаваемая набором векторов вида:

(r, g, b, p, x, y), где

r, g, b – характеристики цвета цветового промежутка;

p – отношение количества пикселей, принадлежащих данному цветовому промежутку, к общему числу пикселей в изображении;

x, y – координаты центра цветового пятна.

Один вектор соответствует одному цветовому промежутку, цвета которого присутствуют на изображении.

На рис. 1 приведен пример цветовой характеристики изображения. Диаграмма, расположенная под изображением, показывает вычисленные цветовые пятна. Каждому вектору характеристики соответствует одна окружность на диаграмме. Размер, цвет и расположение окружностей соответствует компонентам полученных векторов.

Далее для каждой пары изображений из обучающего набора была вычислена оценка степени схожести изображений с помощью функции:

где

$$Eucl((x_{Ik}, y_{Ik}), (x_{Jk}, y_{Jk})) = \begin{cases} \sqrt{(x_{Ik} - x_{Jk})^2 + (y_{Ik} - y_{Jk})^2}, & p_{Ik} > 0 \quad p_{Jk} > 0 \\ const, & p_{Ik} = 0 \quad p_{Jk} = 0 \end{cases},$$

N - количество цветовых промежутков.

Кластеры были построены с помощью пакета CLUTO [2] с использованием алгоритма повторяемых разбиений. Примеры полученных кластеров представлены на рис. 2, 3.

Базисные цветовые характеристики представляют собой характеристики полученных кластеров и строятся как среднее на основе характеристик изображений кластера. Один вектор базисной характеристики соответствует цветовому промежутку, цвета которого присутствуют на всех изображениях кластера, по которому строилась базисная характеристика. Таким образом, базисная цветовая характеристика представляет собой набор векторов вида

(R, G, B, P,X, Y), где

R, G, B – характеристики цветового промежутка (цвета данного цветового промежутка присутствуют на всех изображениях соответствующего кластера);

P – среднее отношение количества пикселей, принадлежащих данному цветовому проме-

жутку, к общему числу пикселей в изображениях кластера:

$\overline{\overline{\quad}}$

X, Y – средние координаты центра цветового пятна на изображениях кластера: $\sum_{i=1}^{N} p_i$

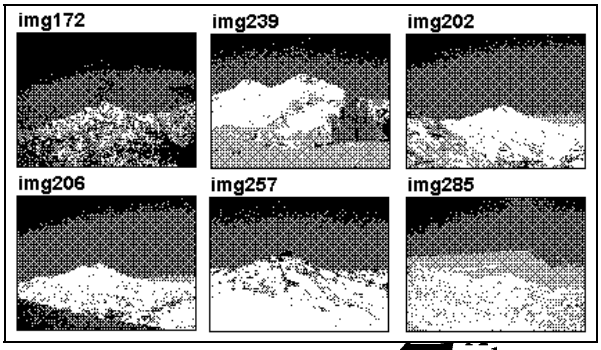
$\overline{\overline{P}} = \overline{\overline{\quad}}$, где N

$\overline{\overline{\quad}}$ $\overline{\overline{\quad}}$

N - количество изо

p - компонента век

соответсвующе G, B).

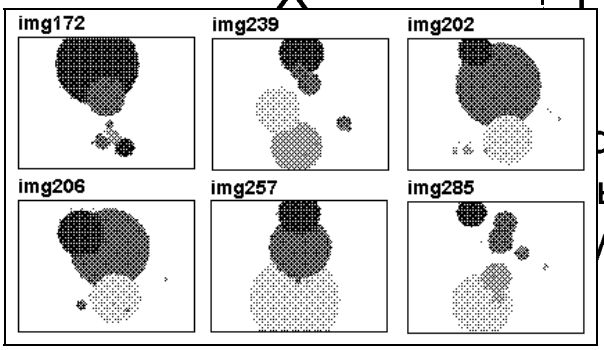


а) Изображения кластера

$\sum_{i=1}$ $\overline{\overline{\quad}}$

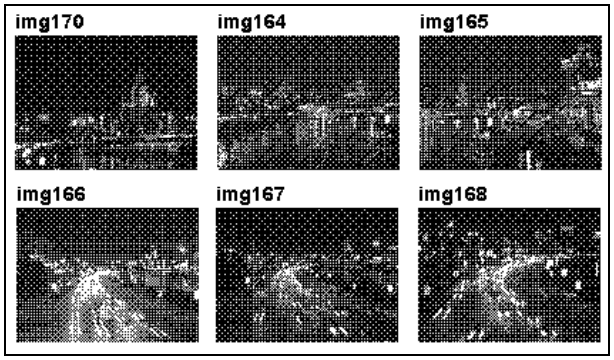
$X =$ $: Y =$ где

N

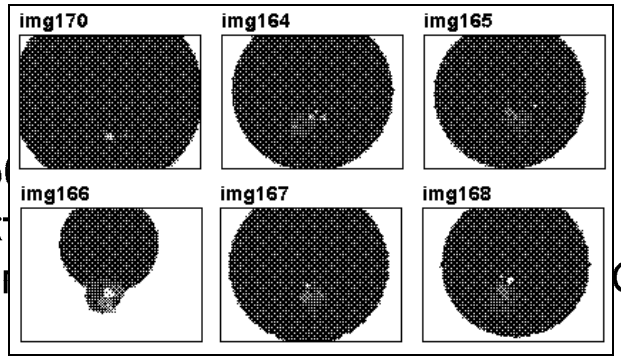


б) Диаграммы цветовых характеристик

Рис. 2 Набор изображений, попавших в один кластер (кластер А).



а) Изображения кластера



б) Диаграммы цветовых характеристик

Рис. 3 Набор изображений, попавших в один кластер (кластер Б).

3.2 Построение базисных лексических характеристик

В качестве базисных лексических характеристик предлагается использовать слова русского языка, которым соответствует определенный зрительный образ. К примеру, такие абстрактные понятия как «воля», «разум» не имеют четких ассоциаций со зрительным образом, в то время как слова «лес», «небо», «трава» имеют сопоставимые зрительные образы. Помимо слов, не имеющих сопоставимого зрительного образа, не могут быть использованы в качестве базисных характеристик и слова, являющиеся обобщением группы более конкретных понятий, имеющих зрительные образы. Это такие слова как «животное», «мебель».

3.3 Сопоставление низкоуровневых и лексических характеристик

Базисным низкоуровневым характеристикам сопоставляются лексические характеристики, лучшим образом описывающие изображения соответствующего кластера. Таблица соответствий строится на основе статистики, полученной в ходе экспериментов по сопоставлению кластеров изображений и базисных лексических характеристик с участием различных пользователей. В качестве примера в таблице А приведены лексиче-

ские характеристики, соответствующие кластерам, изображенным на рис. 2, 3.

Кластер	Лексические характеристики
Кластер А	город, ночь, река, шоссе
Кластер Б	снег, зима, небо, гора

Таблица А Лексические характеристики кластеров А, Б

4 Заключение и дальнейшая работа

В данной работе представлен подход к построению соответствий между низкоуровневыми и лексическими характеристиками изображения. Лексические характеристики легко сопоставимы с семантикой изображения, что дает основания говорить о построении соответствия между низкоуровневым содержанием изображения и его семантикой. Построение таблицы соответствий не является полностью автоматическим процессом – требуется участие пользователя для сопоставления групп изображений и словесных описаний. Однако данная таблица, построенная единожды на основе обучающего набора, может быть использована в последствии для поиска по различным коллекциям изображений с использованием запроса на естественном языке.

Данная работа является одним из первых этапов построения системы поиска изображений по содержанию, описание которой можно найти в [8].

Литература

[1] Black, K Kahol, G Fahmy, P Kuchi, S Panchanathan. Characterizing the high-level content of natural images using lexical basis functions. *Human Vision and Electronic Imaging Conference SPIE 2003, Santa Clara*.

[2] CLUTO: Software Package for Clustering High-Dimensional Datasets *http://www-users.cs.umn.edu/~karypis/cluto/*

[3] Guérin-Dugué A., Ayache S., Berrut C., Image retrieval : a first step for a human centered approach. *Fourth Pacific-Rim Conference on Multimedia (ICICS-PCM 2003)*, Singapore, *15-18 December 2003, 2003.*

[4] Jacobs C. E., Finkelstein A., and Salesin D. H. Fast multiresolution image querying. ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), pp 277–286, Los Angeles, CA, August 1995.

[5] Rubner Y. Perceptual Metrics for Image Database Navigation. *PhD thesis, Stanford University. May 1999.*

[6] Rui Yong, Huang Tomas, Ortega Michael, Mehrotra Sharad. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Video Technology*. 1998.

[7] Smeulders A., Worring M., Santini S., Gupta A. and Jain R., Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on PAMI, vol. 22, n° 12, pp. 1349-1380, 2000.*

[8] Vassilieva N., Novikov B. A Similarity Retrieval Algorithm for Natural Images. *Proc. of the Baltic DB&IS'2004, Riga, Latvia, June 2004.*

Establishing a correspondence between low-level features and semantics of fixed images.

N. Vassilieva, B. Novikov

The performance of content-based image retrieval systems is still judged to be unsatisfactory. Many researches see the main reason for that in so-called “semantic gap” between low-level features and high-level content of the image. Bridging this semantic gap is known as a difficult task.

This paper describes a novel way to establish a correspondence between low-level color-features and high-level content of images. We consider a database of natural images where no additional semantic information about images is available. To build a correspondence table, *basis color features* were built as a first step. These basis color features describes a regularity of color distribution for groups of similar images. A training set of open-scene images was defined. The training set was clustered based on color data. Basis color features are defined as a color features mean of the same cluster’s images. As a second step, a list of primitive words (which we called *lexical basis features*) was selected from a lexicon of the Russian language. Then lexical basis features were used to describe a semantic content of one cluster’s images. Putting together basis color features and basis lexical features for each cluster we set up a correspondence between semantic and color content of natural fixed images.