

Поиск изображений. Выбор оптимальных весов для слияния метрик по цвету и текстуре в зависимости от запроса-образца §

© Илья Марков
Санкт-Петербургский
Государственный
Университет
lya.Markov@gmail.com

© Наталья Васильева
Санкт-Петербургский
Государственный
Университет
Natalia@ntc-it.ru

© Александр Яремчук
Санкт-Петербургский
Государственный
Университет
AYaremchuk@softdev.spb.ru

Аннотация

В информационном поиске изображений по содержанию с использованием низкоуровневых характеристик общепринятой практикой является независимая оценка изображений по цвету и текстуре с дальнейшим смешиванием этих оценок без учета особенностей изображения-запроса. В данной работе выдвигается и проверяется гипотеза о том, что можно определить оптимальные коэффициенты для смешивания оценок по цвету и текстуре в зависимости от характеристик запроса, использование которых улучшит качество поиска.

1 Введение

Основными характеристиками при оценке естественных изображений (фотографий) разнородного содержания являются цвет и текстура. Под разнородным содержанием мы подразумеваем, что рассматриваемая коллекция изображений не принадлежит одной определенной тематике. (В случае, если коллекция содержит фотографии с заранее известной темой, то, в зависимости от нее, может оказаться возможным подобрать другие характеристики, поиск по которым будет эффективнее).

Общепринятой практикой является независимая оценка изображений по цвету и текстуре с дальнейшим смешиванием этих значений для получения единой общей оценки. Однако наряду с большим количеством исследований, посвященных различным способам анализа тех или иных

низкоуровневых характеристик, лишь некоторые работы рассматривают задачу синтеза независимых оценок по различным характеристикам изображения. Большинство работ, посвященных данному вопросу занимаются подбором оптимального алгоритма смешивания вне зависимости от изображения-запроса. Для определения наиболее подходящих коэффициентов смешивания во время поиска иногда используется механизм обратной связи.

Мы выдвигаем гипотезу, что результат поиска можно улучшить, если выделить зависимости между характеристиками запроса и оптимальными для него коэффициентами для смешивания цветовой и текстурной оценок. В качестве алгоритма смешивания используются линейные комбинации соответствующих метрик (смешанные метрики).

Другими словами, мы строим адаптивный алгоритм слияния без использования механизма обратной связи. Возможно, применение механизма обратной связи дает лучший результат, но в-первых, не все поисковые системы оснащены таким механизмом, а во-вторых, те, которые им пользуются, возможно, смогут быстрее осуществлять поиск, если уже на первом шаге до получения первой оценки будут использовать оптимальные для заданного запроса коэффициенты.

2 Существующие подходы

Стандартным способом представления цветовой характеристики изображения являются цветовые гистограммы ([17]): пространство всех цветов разбивается на подмножества с тем, чтобы похожие цвета, по возможности, попали в один элемент разбиения (цветовая область), и для каждой цветовой области подсчитывается количество пикселей, чей цвет принадлежит этой области. Для сравнения изображений на основании цветовых гистограмм в качестве метрики используется сумма модулей разностей значений гистограммы для

* Работа частично поддержана РФФИ (грант 07-07-00268а).

Труды 9^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2007, Переславль-Залесский, Россия, 2007.

каждой цветовой области. В [15], в качестве развития этого направления, предлагаются кумулятивные гистограммы, в которых значение для каждой цветовой ячейки складывается из количества пикселей данного цвета и значения, соответствующего предыдущей ячейке.

В той же работе предлагается альтернативный подход к представлению цветовой характеристики: сохранение трех первых моментов статистического распределения каждого цветового канала. В этом случае, для описания цветовой характеристики требуется всего 9 чисел. В качестве метрики применяется взвешенная сумма разностей между соответствующими моментами изображений. При сравнении с кумулятивными гистограммами этот метод показывает лучшие результаты. В [16] цветовые каналы рассматриваются не как независимые распределения, а как части трехмерного распределения: помимо средних значений по каждому каналу для характеристики берутся также попарные ковариации каналов.

Текстура изображения рассматривается в [1]. С помощью анализа независимых компонент (Independent Component Analysis – ICA) выделяются фильтры, которые призваны отражать основные направления текстуры для той базы изображений, на основе которой они строятся. В качестве характеристического вектора изображения берется набор его сверток с каждым из фильтров, а в качестве метрики используется дивергенция Kullback-Leibler. Особенностью фильтров, полученных методом ICA, является их естественный характер: они отражают основные направления текстуры для набора изображений, по которому были построены. В этом их преимущество по сравнению с широко применяемыми фильтрами Габора (Gabor), которые строятся на основе различных сечений функции Габора, то есть носят искусственный характер.

Сравнительно небольшое количество исследований посвящено задаче синтеза оценок, полученных независимо по различным низкоуровневым характеристикам. Для решения данной проблемы представляется возможным использовать общие алгоритмы синтеза данных (data fusion), такие как CombSUM, CombMNZ и другие ([12], [10]). В некоторых работах эта задача решается применительно именно к характеристикам изображений. Так, авторы [3] предлагают алгоритмы, основанные на алгебре нечетких множеств. Однако указанные подходы рассматривают все входные источники данных как равнозначные и никак не учитывают особенности запроса.

3 Постановка задачи

Как уже говорилось, мы предполагаем, что результат поиска можно улучшить, если использовать оптимальные коэффициенты для смешивания оценок по цвету и текстуре для

заданного образца. Итоговую оценку схожести изображений с образцом мы будем вычислять с помощью смешанных метрик, которые выводятся из цветовой и текстурной по следующей формуле: $a \cdot C + (1 - a) \cdot T$, где C – метрика по цветовой характеристике, T – метрика по текстурной характеристике, a – параметр из отрезка $[0, 1]$.

Цель работы заключается в том, чтобы проверить следующую гипотезу. Мы предполагаем, что для схожих изображений-запросов оптимальное значение параметра a будет одинаковым. А значит можно выделить классы семантически близких изображений, определить оптимальное значение данного параметра для каждого класса и определить общие характеристики изображений, принадлежащих одному классу. Далее, во время поиска по запросу определять, к какому из выделенных классов относится изображение-образец, и использовать оптимальное для данного класса значение параметра a для вычисления смешанной метрики.

Таким образом, задача сводится к проверке существования и определению единого оптимального значения параметра a для схожих изображений. Возможно, не для каждого из классов семантически близких изображений удастся определить это значение и вычислить общие характеристики. В таком случае для изображения-образца, который невозможно отнести ни к одному из определенных классов с известным значением параметра a , во время поиска следует использовать некоторое среднее значение этого параметра.

Вследствие того, что статистический метод представления цветовой характеристики дает лучшие результаты по сравнению с гистограммным как по релевантности результатов, так и по производительности ([15]), было решено представлять эту характеристику в виде трех первых моментов цветových каналов и попарных их ковариаций, а в качестве цветовой метрики была выбрана метрика из работы [16].

Текстурная характеристика представляется в виде набора сверток изображения с фильтрами, полученными с помощью ICA, а текстурная метрика, взятая из [1], является дивергенцией Kullback-Leibler. Как говорилось в предыдущем разделе, получение фильтров с помощью ICA более естественно (по сравнению, например, с фильтрами Gabor), что и повлияло на наш выбор.

4 Описание эксперимента

Экспериментальная база данных составлена из 285 фотографий COREL различной тематики: в основном, это природа, городские пейзажи, фотографии людей и животных. Для каждого изображения посчитаны цветовая и текстурная характеристики, и для каждой пары вычислены расстояния по обеим характеристикам с использованием соответствующих метрик. Метрики

нормированы таким образом, чтобы каждая из функций принимала значения от 0 до 1.

Для эксперимента выбрано 5 значений параметра a : 0, 0.25, 0.5, 0.75 и 1, что дает 5 различных смешанных метрик (первая и последняя совпадают, соответственно, с текстурной и цветовой).

Так как поисковые системы, в конечном счете, ориентированы на пользователей, экспериментальный стенд представляет собой интерактивную систему, цель работы которой – выяснить субъективное мнение людей о схожести изображений. Внешний вид основного окна стенда представлен на рис. 1.

Слева отображается случайным образом выбранный образец, а основную часть экрана занимает результат поиска, полученный с

применением смешанной метрики со случайно выбранным параметром a (из 5-ти возможных значений). Ассессору предлагается оценить схожесть образца и каждого из полученных изображений по трехбалльной шкале: "совсем не похожи", "немного похожи", "похожи". В качестве результата оценки сохраняется пара сравниваемых изображений и сама оценка (соответственно, 0, 0.5, 1).

Для разделения результатов, полученных от разных ассессоров, введена предварительная регистрация. Каждый участник может пройти любое количество итераций сравнения, на каждой из которых случайным образом выбирается изображение-образец и смешанная метрика для поиска похожих изображений.



Рис. 1. Внешний вид основного окна экспериментального стенда.

5 Анализ результатов

Статистика, собранная в результате участия в эксперименте порядка 160 добровольцев, обрабатывается следующим образом. Для каждой пары изображений вычисляется средняя оценка по всем участникам эксперимента, оценившим данную пару:

$$sim_{av}(img_1, img_2) = \frac{\sum_{users} sim_{user}(img_1, img_2)}{|users|},$$

где $sim_{user}(img_1, img_2) \in \{0, 0.5, 1\}$.

На основании полученной информации изображения разбиваются на кластеры схожих изображений. Причем в процессе кластеризации участвуют только те пары, которые оценили не менее $user_thesh$ участников. Кластер строится таким образом, чтобы для любого изображения,

входящего в него, существовало хотя бы одно изображение из того же кластера такое, что средняя пользовательская оценка для этой пары ($sim_{user}(img_1, img_2)$) не меньше заданного порогового значения кластеризации $cluster_thresh$:

$$sim_{user}(img_1, img_2) \geq cluster_thresh.$$

Среди полученных кластеров рассматриваются только те, размер которых не меньше $image_thresh$.

При анализе результатов использовались следующие значения параметров:

$$user_thesh \in \{3, 4, 5\},$$

$$cluster_thresh \in \{0.5, 0.6, 0.7\},$$

$$image_thresh = 5.$$

При $user_thesh = 3$ кластеры строятся на основе информации, полученной от трех пользователей, которая не всегда может быть достоверной. По этой

причине некоторые кластеры получаются неоднородными и не представляют ценности для дальнейшего изучения и определения общей смешанной метрики. Наоборот, при $user_thesh = 5$ довольно много информации отсеивается, поэтому некоторые кластеры получаются неполными и могут при дальнейшей обработке дать не вполне адекватные результаты. Кроме того, оказалось, что при $user_thesh = 4$ часть кластеров равна аналогичным для $user_thesh \in \{3,5\}$ (поэтому значения 3 и 5 в ряде случаев можно вообще не рассматривать), остальные же, на наш взгляд, представляют собой "золотую середину" относительно значений 3 и 5 параметра $user_thesh$. Поэтому в дальнейшем рассматриваются, в основном, кластеры, полученные при $user_thesh = 4$ с оглядкой на остальные результаты.

Каждое изображение кластера используется в качестве образца для поиска по каждой из смешанных метрик. От результата, полученного с использованием смешанной метрики, отбираются первые $|cluster|$ элементов и для них подсчитывается полнота: $\frac{|result \cap cluster|}{|result|}$

(отношение количества изображений из результата, входящих в кластер, к размеру кластера). Для каждой из смешанных метрик эта информация суммируется по всем изображениям кластера, и таким образом получается средняя полнота. В качестве наилучшей метрики выбирается метрика с максимальной суммой.

В результате обработки получилось 8 кластеров (описание см. в табл. 1). Зависимости средней полноты от используемой смешанной метрики (выраженной значением параметра a) для каждого из них представлены на рис. 2 и 3. Из рис. 2 видно, что для Кластера 1 цветовая метрика дает существенно лучший результат, чем остальные. Для Кластера 2 полнота растет равномерно, увеличиваясь в сторону цвета. Текстурная метрика проявляет себя значительно хуже остальных для

Кластера 3, в то время как остальные смешанные метрики дают схожий результат с тяготением к цвету. Кластер 7 и Кластер 8, представленные на рис. 3, наглядно показывают, что в некоторых случаях значительно эффективнее применять комбинацию метрик, нежели одну из них (полнота для чистых метрик – цветовой и текстурной – минимальная среди всех смешанных).

Сводная информация по всем кластерам приведена в табл. 1.

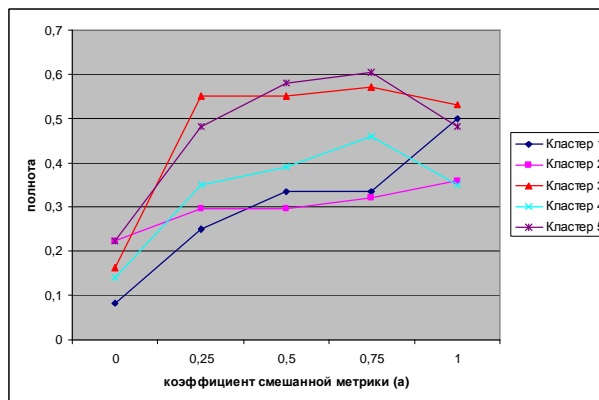


Рис. 2. Кластеры с преобладанием цветовой характеристики.

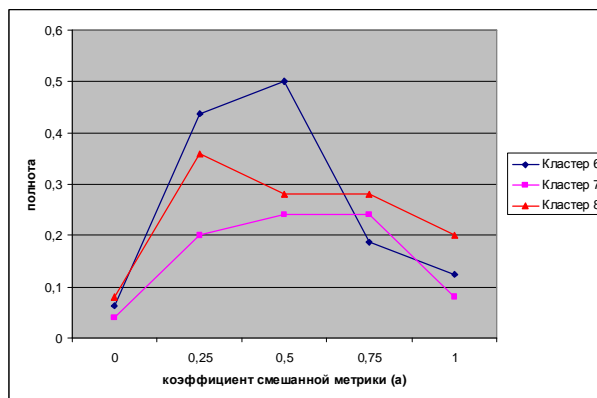


Рис. 3. Кластеры со смещением соотношения характеристик в сторону текстуры.

| Номер кластера | Описание | Размер | Смешанная метрика | $user_thesh$ | $cluster_thresh$ |
|----------------|-------------------------------|--------|-------------------------------|---------------|-------------------|
| 1 | Небо (рис. 4) | 7 | $1 \cdot C + 0 \cdot T$ | 4 - 5 | 0.7 |
| 2 | Животные с пятнистой окраской | 10 | $1 \cdot C + 0 \cdot T$ | 4 | 0.6 |
| 3 | Облака | 8 | $0.75 \cdot C + 0.25 \cdot T$ | 4 - 5 | 0.7 |
| 4 | Лесные озера | 9 | $0.75 \cdot C + 0.25 \cdot T$ | 4 | 0.6 |
| 5 | Поля, луга, равнины | 10 | $0.75 \cdot C + 0.25 \cdot T$ | 4 | 0.7 |
| 6 | Листва (рис. 5) | 5 | $0.5 \cdot C + 0.5 \cdot T$ | 4 | 0.6 |
| 7 | Небоскребы | 6 | $0.5 \cdot C + 0.5 \cdot T$ | 4 | 0.5 |
| 8 | Группы людей (рис. 6) | 5 | $0.25 \cdot C + 0.75 \cdot T$ | 4 | 0.6 |

Табл. 1. Кластеры, полученные в результате эксперимента.



Рис. 4. Класс с преобладанием цветовой характеристики.



Рис. 5. Изображения, для которых цветовая и текстурная характеристики равнозначны.



Рис. 6. Класс с преобладанием текстурной характеристики.

6 Заключение и дальнейшая работа

С помощью экспериментов мы показали, что для разных семантических классов изображений можно определить относительную важность цветовой и текстурной характеристик при осуществлении поиска по элементам класса. В целом, как и следовало ожидать, цветовая характеристика оказывается более важной для большинства изображений. Но в некоторых случаях эффективней использовать комбинацию метрик по цвету и текстуре, нежели одну из них.

Для применения результатов данной работы на практике мы планируем, в первую очередь, выделить общие численные характеристики полученных классов, чтобы в дальнейшем осуществлять классификацию изображений-запросов не на основе содержимого имеющихся классов, а на основе их средних характеристик. Количество сравнений, которое при этом потребуется, будет равно количеству классов.

Необходимо, однако, чтобы полученные общие характеристики существенно отличались для различных классов. Невыполнение этого условия сделает невыполнимой задачу классификации запроса-образца по выделенным классам и, как

следствие, определение для него оптимального значения коэффициента a .

Литература

- [1] H. Borgne, A. Guerin-Dugue, A. Antoniadis. Representation of images for classification with independent features. In *Pattern Recognition Letters*, volume 25, pages 141-154, 2004.
- [2] E.L. Broek, E.M. Rikxoort, T.E. Schouten. Human-Centered Object-Based Image Retrieval. <http://www.ita.cs.ru.nl/publications/papers/th/Broe05b/Broe05b.html>, 2005.
- [3] P.J. Deer, P.W. Eklund. On the Fusion of Image Features. <http://citeseer.ist.psu.edu/162546.html>.
- [4] S. Fan. Shape Representation and Retrieval Using Distance Histograms. Technical Report, 2001.
- [5] R.S. Gray. Content-based Image Retrieval: Color and Edges. Technical report, 1995.
- [6] A. Guerin-Dugue, S. Ayache, C. Berrut. Image retrieval: a first step for a human centered approach. In *Fourth Pacific-Rim Conference on Multimedia*, 2003.
- [7] A. Hanbury, J. Serra. Mathematical Morphology in the $L^*a^*b^*$ Colour Space. <http://citeseer.ist.psu.edu/646818.html>, 2001.
- [8] C.G. Healey, J.T. Enns. A Perceptual Colour Segmentation Algorithm. Technical Report, 1996.

- [9] J. Hu, A. Mojsilovic. Optimal Color Composition Matching of Images. In *Proceedings of ICPR*, volume 4, pages 47-51, 2000.
- [10] J.H. Lee. Analyses of Multiple Evidence Combination. In *Proceedings of the 20th Annual International ACM-SIGIR Conference*, pages 267-276, 1997
- [11] W.K. Leow, R. Li. The Analysis and Applications of Adaptive-Binning Color Histograms. <http://citeseer.ist.psu.edu/667678.html>.
- [12] D. Lilis, F. Toolan, A. Mur, L. Peng, R. Colier, J. Dunnion. Probability-Based Fusion of Information Retrieval Result Sets. In *Proceedings of the 16th AICS Conference*, 2005.
- [13] A. Mojsilovic. A method for color naming and description of color composition in images. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 789-792, 2002.
- [14] M. Stricker, M. Orengo. Similarity of Color Images. In *Proceedings of the SPIE Conference*, volume 2420, pages 381-392, 1995.
- [15] M. Stricker, A. Dimai. Color Indexing with Weak Spatial Constraints. In *Proceedings of the SPIE Conference*, 1996.
- [16] M. Stricker, A. Dimai. Spectral Covariance and Fuzzy Regions for Image Indexing. In *Machine Vision and Applications*, volume 10, pages 66-73, 1997.
- [17] M.J. Swain, D.H. Ballard. Color Indexing. In *International Journal of Computer Vision*, volume 7(1), pages 11-32, 1991.

Image retrieval. Optimal weights for color and texture fusion based on query object.

I. Markov, N. Vassilieva, A. Yaremchuk

It is a common way in CBIR to process different image features independently to estimate image similarity. Color and texture are common features which are used for searching in natural images. This paper proposes the hypothesis that it is possible to mark out optimal weights for fusing color and texture-based estimations in accordance with query image features. Linear combination of color and texture metrics is considered as a mixed-metrics. Clusters of images with common features and optimal weights for them are presented based on experimental results. The results of the paper can be used to determine the best weights for particular query and thus improve image retrieval.