УДК 004.932.72'1

А.С. Мельниченко

АВТОМАТИЧЕСКАЯ АННОТАЦИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ГЛОБАЛЬНЫХ ПРИЗНАКОВ*

В данной работе рассматривается проблема автоматического аннотирования изображений с целью описания изображений набором ключевых слов, позволяющих осуществлять поиск по текстовому запросу в больших коллекциях изображений. Построенная в результате анализа, сравнения и модификации эффективных современных методов представления и поиска изображений модель аннотирования алгоритмизована и реализована программно, найдены оптимальные наборы параметров.

Автоматическая аннотация изображений; поиск изображений; признаки изображений; вероятностные модели; языковые модели.

A.S. Melnichenko

AUTOMATIC IMAGE ANNOTATION BASED ON GLOBAL IMAGE FEATURES

The problem of automated image annotation is considered in this work. The task is introduced and solved under the assumption of further application of its results to the problem of retrieval large collections of images. Existing methods are reviewed and analyzed their advantages and disadvantages. The task is splitted into stages and the most effective method through the existing and improved solutions is proposed for each stage. The program implementation have been done for all major stages of the concidered methods of image annotation.

Automated image annotation; image retrieval; image processing; image features; probability models; language models; language smoothing.

Автоматическая аннотация изображений – это процесс автоматического присвоения системой метаданных в форме заголовка или ключевых слов цифровому изображению на основании только визуальной информации, содержащейся в изображении. Автоматическая аннотация изображений имеет своей целью описание с помощью слов визуальной информации содержащейся на изображении (как низкоуровневой информации, связанной с цветами, текстурами изображения, так и высокоуровневой, связанной с семантикой изображения).

Автоматическая аннотация изображений является очень востребованной в последние годы областью компьютерного зрения в связи с ее потенциальным влиянием на понимание изображений и веб-поиск.

Предпосылки и проблемы автоматической аннотации изображений

Сегодня, в связи с развитием мультимедийных технологий, у пользователей и организаций появились огромные базы данных изображений. Общее же количество изображений в Интернете оценивается величиной 10^{10} и с каждым годом удваивается. В этих условиях весьма актуальным становится вопрос эффективного управления большими коллекциями изображений, семантической классификации и быстрого и точного поиска нужных пользователю изображений.

* Работа выполнена при финансовой поддержке РФФИ, проект №08-07-00129, №07-07-00067.

На сегодняшний день существует два основных подхода к поиску изображений: *поиск на основе текстового запроса* и *поиск на основе визуального образца*. В первом случае в качестве запроса выступает набор ключевых слов, и система должна вернуть наиболее удовлетворяющие им изображения. Во втором случае в качестве запроса выступает изображение-образец (например, нарисованное от руки, сканированное или изображение низкого качества), а система находит все подобные ему изображения. Главными областями применения подхода поиска на основе визуального образца являются те, где важнее найти визуальное сходство изображений: поиск в медицинских коллекциях, например, среди рентгеновских снимков; поиск в дизайнерских коллекциях, когда дизайнер ищет некоторое подходящее по цветовой гамме и текстуре изображение; поиск в архивах правоохранительных органов интересующих криминалистов лиц или объектов. Поиск по текстовому запросу более удобен для пользователя и обычно применяется там, где более важна семантика искомого изображения. Во многих случаях поиск по текстовому запросу является единственным приемлемым для пользователя вариантом поиска (например, когда у пользователя нет образца изображения, похожего на искомое, искомое понятие сложно выразить визуальным образцом или когда пользователь хочет найти изображения, не обладающие визуальным сходством). Трудностью поиска по текстовому запросу является то, что для его проведения все изображения в коллекции должны иметь текстовые описания: ключевые слова, тэги. Но ручное аннотирование столь больших объемов изображений – слишком трудоемкий процесс. В связи с этим с конца 90-х гг. началась активная разработка методов автоматического аннотирования изображений. С появлением методов автоматического аннотирования ручное аннотирование множества изображений перестает быть необходимым, что позволяет включить в поиск большее количество картинок и улучшить его качество. В настоящее время также производятся попытки решить проблему субъективности аннотаций, зависимости от языка путем введения единого словаря понятий и применением различных лингвистических технологий. Наиболее важная область применения такого поиска – это поиск в Интернете, поэтому разработка методов автоматического аннотирования является весьма актуальной задачей.

animal tiger snow natural winter *landscape castle lake forest tree*

Рис. 1. Пример автоматического присвоения ключевых слов изображениям

Обработка изображений, кроме технических особенностей (трудностей хранения и обработки больших объемов данных), обладает своей спецификой, коренным образом отличающей ее от обработки текстовой информации. Поэтому разработка эффективных методов навигации и поиска в больших базах данных изображений, представляет собой весьма сложную и не нашедшую на сегодняшний день

удовлетворительного решения задачу. Основной проблемой, затрудняющей эффективное и однозначное решение проблемы поиска в коллекции изображений, является так называемая проблема «*семантического разрыва*» — отсутствия однозначной связи между низкоуровневыми характеристиками и семантикой изображения. Критерий сходства, вводимый на основании числовых значений пикселов изображений или полученных на их основе производных характеристик, не в состоянии уловить семантику запроса, подразумеваемую пользователем. Подобные проблемы, связанные со смысловой неопределенностью, возникают также и в системах автоматизированной смысловой обработки текстов. Проблема же семантического разрыва в области обработки изображений является намного более сложной, так как она должна быть решена не в рамках естественного языка, а в рамках весьма неоднозначной связи языка и визуальных характеристик изображений. Попытки преодолеть проблему семантического разрыва влекут существование множества разнообразных представлений визуальных характеристик изображений. Одной из проблем автоматического аннотирования также является выбор представления изображений, позволяющего эффективно различать визуальные особенности изображений, важные для описания их словами.

Ещё одной трудностью, связанной с созданием баз изображений, является отсутствие общих, универсальных методов, подходящих для любых коллекций изображений. Разные коллекции и разные задачи требуют своих методов обработки и поиска. Кроме того, методы автоматического аннотирования изображений обычно требуют обучения на коллекции изображений, имеющих аннотации. В связи с этим возникают некоторые дополнительные проблемы. Различные коллекции изображений аннотируются разными наборами слов с использованием различных правил. Из-за больших различий в специфике обучающей коллекции и той коллекции, на которой будет применена построенная модель, ее работа может быть некорректна. В дополнение ко всему, пользователь, которому направлены результаты аннотации, например для поиска среди коллекции, не всегда знает словарь ключевых слов. Поэтому построение подходящего словаря ключевых слов, задействованного в модели автоматического аннотирования, также является важной задачей.

Этапы решения задачи автоматического аннотирования изображений

Задачу автоматической аннотации изображений можно разбить на три этапа:

1. Разработка представления визуальных признаков изображений, составление словаря этих признаков: $B = \{b_1, b_2, \ldots, b_N\}$, где b_i, $i = 1, \ldots, N$ – один из выбранных признаков изображения.
2. Составление обучающей выборки $Q = \{I_1, I_2, \ldots, I_K\}$ аннотированных изображений, а также словаря ключевых слов этой выборки $W = \{w_1, \ldots, w_M\}$.
3. Условимся также в дальнейшем обозначать $I \in Q$ – изображение из обучающей выборки, а $J \notin Q$ — новое неаннотированное изображение не из обучающей выборки, для которого мы хотим построить аннотацию.
4. Построение математической вероятностной модели $M_{B,W}$, описывающей взаимосвязи слов из W с визуальными признаками из B путем нахождения совместного распределения $P(w, I)$ для каждого слова w из словаря ключевых слов и каждого изображения I из обучающей выборки, оценка параметров модели по обучающей выборке Q.

Все эти три этапа очень важны для эффективной работы алгоритма.

Использование признаков, описывающих характерные особенности визуального содержания изображений, выделяющих важные для восприятия человеком характеристики изображений и позволяющих определять сходство и различие изображений, является необходимым условием эффективной работы методов обработки изображений и, в частности, методов автоматической аннотации. Принято различать региональное (основанное на локальных характеристиках отдельных регионов изображений) и глобальное (вычисляемые для всего изображения в целом) представления изображений. Так, региональное представление используется в таких моделях, как Co-occurrence Model[1], Translation Model[2], Cross-Media Relevance Model[3, 4], а глобальное в Global image features and nonparametric density estimation[5] и Automatic linguistic indexing model[6]. При этом в первых трех моделях не ставилась целью разработка признаков изображений, а использовалось готовое представление. Автоматическое же получение региональных представлений представляет собой весьма трудную задачу. В то же время в последних двух моделях, нашедших практическое применение в поисковых системах Behold[1] и ALIPR[2], а также в моделях глобального аннотирования[7], с успехом применяются глобальные признаки изображений. В работах [8, 9] также показывается связь глобальных признаков с человеческим восприятием изображений, важность цветовых и текстурных характеристик. Исходя из всего этого использование глобальных признаков можно считать перспективным. Поэтому в данной работе в качестве первого этапа решения поставленной задачи рассматривается разработка методов выделения глобальных признаков изображений с использованием актуальных техник машинного зрения.

Выборка аннотированных изображений, используемая для обучения, также имеет большое значение для методов автоматической аннотации. Необходимо, чтобы аннотации изображений в обучающей выборке были полными, для каждого изображения было определено несколько ключевых слов. Множество слов, встречающихся в аннотациях изображений обучающей выборки, составляют словарь ключевых слов W. Необходимо, чтобы слова словаря были представлены в обучающей выборке достаточным для оценки некоторой выбранной модели количеством и чтобы размер словаря ключевых слов не превышал размер словаря визуальных признаков. В данной работе была использована выборка, составленная на основе базы аннотированных изображений mirflickr[3], состоящей из изображений пользователей он-лайн сервиса www.flickr.com и содержащей пользовательские аннотации на 7 различных языках. В результате составления удовлетворяющей описанным выше требованиям обучающей выборки размер полученной выборки составил $|Q|=2500$, размер словаря ключевых слов $|W|=50$ (в связи со спецификой коллекции использовались ключевые слова на английском языке).

И, наконец, третьей важной задачей построения методов автоматической аннотации является выбор модели, которая будет ставить в соответствие визуальным признакам изображений слова. Вероятностные методы применительно к автоматическому аннотированию изображений представляют собой расширение моделей для поиска текста [10] с учетом особенностей изображений. Большинство используемых для автоматического аннотирования изображений методов являются

[1] http://behold.cc – он-лайн система для поиска изображений по текстовому запросу.
[2] http://alipr.com – он-лайн система для аннотирования изображений и поиска по визуальному подобию.
[3] http://mirflickr08.com

вероятностными, а Cross-Media Relevance Model (CMRM) является также расширением класса языковых моделей, осуществляющих многоязычный текстовый поиск, на случай аннотирования изображений, т.е поиск на двух «языках»: естественном языке и языке низкоуровневых характеристик изображений. Так как языковые модели являются сравнительно простым в реализации и в то же время хорошо зарекомендовавшим себя в информационном поиске классом моделей, в данной дипломной работе предлагается использовать модель CMRM в качестве модели для оценки распределения слов и визуальных признаков изображений.

Все этапы поставленной задачи были адгоритмизованны и реализованы программно с целью проведения общего вычислительного эксперимента, направленного на оценку параметров модели и проверку ее работоспособности для реальной коллекции изображений.

Построение словаря визуальных признаков на основе глобальных характеристик изображений

Цветовые гистограммы изображений

Распределение цветов на изображении – очень важная для зрительного восприятия человека характеристика. Эксперименты по изучению психофизиологических особенностей человеческого зрения [8, 9] показали, что распределения цветов на изображении бывает достаточно для различения типов сцен и многих типов объектов.

Поэтому большое значение для вычисления цветовых характеристик имеет выбор цветового пространства. Примером цветового пространства, цвета в котором хорошо согласуются с человеческим восприятием цвета, является пространство *CIELab*. При разработке *CIELab* преследовалась цель создания цветового пространства, изменение цвета в котором будет линейным с точки зрения человеческого восприятия, то есть одинаковое изменение значений координат цвета в разных областях цветового пространства будет производить одинаковое ощущение изменения цвета. Таким образом математически корректировалась бы нелинейность восприятия цвета человеком.

Смысл цветовых компонентов точки в пространстве *CIELab* состоит в следующем. Координата L (lightness) задает светлоту цвета, хроматическая составляющая – двумя полярными координатами a и b. Первая обозначает положение цвета в диапазоне от зеленого до пурпурного, вторая – от синего до желтого. Преобразование из промежуточного пространства *CIE XYZ[]* в *CIELab* осуществляется следующим образом:

$$L = 116 f(Y/Y_n) - 16,$$
$$a = 500[f(X/X_n) - f(Y/Y_n)],$$
$$b = 200[f(Y/Y_n) - f(Z/Z_n)],$$

где X_n, Y_n и Z_n – это координаты белой точки в значениях *CIE XYZ*, для стандартного осветителя D50 они равны:

$$X_n = 96.42,\ Y_n = 100,\ Z_n = 82.49.$$

$$f(t) = \begin{cases} t^{1/3}, & t > (6/29)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & t \text{„} (6/29)^3 \end{cases}$$

На основе пространства *CIELab* вводится пространство *CIELCh*, в котором компонента L остается такой же, как и в *CIELab*, а компоненты C (Chroma) и h (hue) рассчитываются следующим образом:

$$C = (a^2 + b^2)^{1/2}, \quad h = \frac{b}{a}.$$

Характеристика однородности фона изображений

Для различения типов сцен также может быть полезен признак, характеризующий однородность изображения и его фон. Наличие так называемого фона – областей компактно расположенных пикселей, обладающих примерно одинаковыми цветовыми признаками, является важной характеристикой, сильно влияющей на восприятие изображения. Для вычисления признака содержания фона (*Bg*) предлагается использовать комбинацию методов, предложенных в статьях [11, 12].

Признаком отсутствия у изображения фона можно считать равномерность его гистограммы. Рассмотрим цветное изображение I размером $X_I \times Y_I$, имеющее три цветовые компоненты I_{ij}^R, I_{ij}^G, I_{ij}^B, $i = 1,\ldots,X_I$, $j = 1,\ldots,Y_I$ для *R, G, B* соответственно. Элементы этих компонент принимают значения из диапазона $\{0,1,\ldots,M\}$. В качестве численной оценке степени равномерности гистограмм цветовых компонент H_i^R, H_i^G, H_i^B, $i = 0,\ldots,M$ можно использовать энтропию как меру информативности каждого из столбцов гистограммы:

$$E^x = \sum_{i=0}^{M} \delta_{i,x}, \qquad x = R,G,B, i = 0,\ldots,M,$$

где

$$\delta_{i,x} = \begin{cases} 0, & \text{если } \mathrm{H_i^x} = 0, \\ -\dfrac{H_i^x}{X_I Y_I} \log\left(\dfrac{H_i^x}{X_I Y_I}\right), & \text{если } \mathrm{H_i^x} \neq 0. \end{cases}$$

На основании полученного значения энтропии определяется признак содержания фона в каждой из цветовых компонент изображения:

$$B^x = 1 - \frac{E^x}{\log M}, \qquad x = R,G,B.$$

а также для всего изображения: $B = B^R B^G B^B$.

В [11] показано, что построенный таким образом признак будет инвариантным к конкретному цвету и интенсивности фона, а также к контрасту информативной части изображения по отношению к окружающему ее фону и будет изменяться в пределах от 0 до 1. Для достижения монотонного изменения признака относительно процентного отношения пикселей, принадлежащих фону, для изображений, содержащих шум или подвергшихся сжатию, в процессе вычислительных экспериментов была использована предварительная обработка изображений медианным фильтром. Полученное в ходе эксперимента типичное значение признака для фотографий реальных сцен лежит в пределах 0,08 – 0,15.

Текстурные признаки изображений

Текстура является очень важным с точки зрения восприятия и распознавания объектов человеком свойством изображений [13]. На многих изображениях она позволяет определять свойства областей, критически важные для корректного выполнения анализа содержания изображения. Свойства, отвечающие человеческому восприятию, имеют текстурный признак *Tamura*, введенный Hideyuki Tamura в [14]. Введенный текстурный признак включает такие характеристики, как грубость (coarseness), контраст (contrast), направленность (directionality) текстуры.

Признак, характеризующий *грубость*, неявно связан с размером примитивных элементов, формирующих текстуру. Он рассчитывается как усредненный по всем точкам (x, y) изображения размер $S_{best}(x, y)$ окна, в котором абсолютные разности между парами средних интенсивностей, вычисленных в горизонтальном и вертикальном направлениях относительно текущей точки, максимальны:

$$F_{crs} = \sum_{i=1}^{M}\sum_{j=1}^{N} S_{best}(i, j).$$

Признак, характеризующий *контраст*, является показателем того, как уровни серого $q : q = 1, 2, \ldots, q_{max}$ варьируются в пределах изображения I и в какой степени их распределение смещено к белому или черному. Для вычисления контраста используются центральные моменты второго σ и четвертого μ_4 порядков гистограммы уровней серого изображения, и значение признака определяется как:

$$F_{con} = \frac{\sigma}{\gamma^{1/4}},\ \gamma = \frac{\mu_4}{\sigma^4},$$

Ещё одним признаком текстуры согласно Tamura является *направленность*. Этот признак вычисляется с использованием значений величин и направлений градиента интенсивности в точках изображения,

Строится гистограмма направлений градиента $H_{dir}(\theta)$, и в качестве признака затем используется либо сама эта гистограмма, либо вычисляется степень направленности, связанная с величинами пиков гистограммы:

$$F_{dir} = 1 - rn_{peaks} \sum_{p=1}^{n_{peaks}} \sum_{\theta \in w_p} \left(\theta - \theta_p\right)^2 H_{dir}(\theta),$$

где n_{peaks} – число пиков гистограммы, больших некоторого порога, θ_p – соответствующее p-му пику значение угла θ, w_p – некоторая окрестность p-го пика, r – нормализующий множитель.

Признак F_{dir} лучше использовать для изображений, содержащих одну текстуру, фотографии же различных сцен, для которых нужно получить текстурные характеристики, содержат, как правило, несколько типов текстур. Поэтому в качестве признака направленности для них было решено использовать вместо одного признака F_{dir} гистограмму H_{dir}. В экспериментах была использована гистограмма с 16 ячейками.

Гистограммы ориентаций градиентов

В рассматриваемом виде эти дескрипторы (Histogram of oriented gradients – *HoG*) были впервые предложены и описаны исследователями Navneet Dalal и Bill Triggs в работах [15, 16]. Основанием рассматриваемого метода построения дескрипторов на основе HoG является то, что наличие и форма локальных объектов на изображении могут быть описаны распределением интенсивности градиентов. Изображение делится на малые регионы — «клетки», для каждой клетки составляется гистограмма направлений градиентов точек этой клетки. Комбинация этих гистограмм и представляет собой дескриптор. Существует несколько типов дескрипторов, наиболее часто используемым является прямоугольный дескриптор – R-HoG (рис. 2). Использованный в эксперименте дескриптор имеет 3×3 клеток, 4 блока, включающих 4 клетки и 9 ячеек гистограммы.

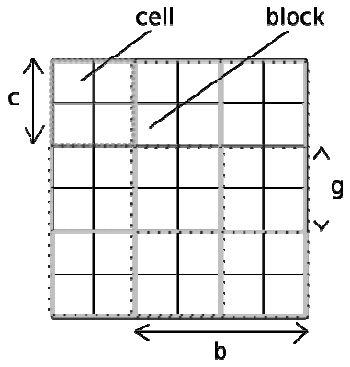

Рис. 2. Прямоугольный дескриптор R-HoG

Для улучшения качества детекции локальные гистограммы могут быть нормализованы по контрасту путем вычисления меры интенсивности по большему региону изображения, т.н. блоку, и использования этого значения для нормализации гистограмм всех клеток в блоке. Такая нормализация ведет к лучшей инвариантности относительно изменений освещенности и теням. Авторы, предлагающие дескрипторы HoG, подчеркивают преимущество этого похожа по сравнению с подобными подходами edge orientation histograms, scale-invariant feature transform descriptors, shape contexts.

Составление словаря визуальных признаков

Описанные выше признаки изображений были использованы автором данной работы для поиска по визуальному подобию в [17, 18], что позволило на порядок повысить точность поиска по сравнению с другими ранее используемыми признаками.

Для построения модели аннотирования нам необходимо составить «словарь», из используемых признаков с помощью некоторого алгоритма кластеризации (в ходе эксперимента использовались известные алгоритмы k-means и k-medoids). Будем строить словарь признаков размером N, в котором присутствует некоторое количество N_k признаков каждого выбранного типа k. Состав одного из словарей, использовавшегося в вычислительном эксперименте, показан а табл. 1. Обозначение каждого признака состоит из его названия и длины получаемого вектора.

Таблица 1

Состав словаря визуальных признаков

Название признака	Длина вектора	Признаков в словаре
Bg - 1	1	10
HoG - 144	144	20
RGB - 256	256	20
CIELCh - 60	60	20
Tamura - 18	18	20
Всего	90 признаков	

Вероятностная модель автоматического аннотирования

Перейдем теперь к рассмотрению вероятностной модели. Пусть $J = \{b_1,\ldots,b_s\}$ – представление нового неаннотированного изображения с помощью признаков из словаря признаков $b_i \in B$. Мы хотим выбрать набор ключевых слов $\{w_1,\ldots,w_t\}$, $w_i \in W$, которые наиболее адекватно описывают содержание изображения.

Для построения вероятностной модели аннотирования можно перенести подход, использующийся при построении вероятностных языковых моделей текстового поиска, на случай изображений. Аналогично тому, как в языковых моделях для каждого документа d строится модель M_d, предположим, что для каждого изображения I можно построить модель M_I, так называемую модель релевантности для I. Модель M_I представляет собой вероятностное распределение $P(\cdot \mid I)$, которое содержит все возможные слова и признаки из словаря признаков B и словаря слов W обучающей выборки. Модель релевантности M_d будем рассматривать как «черный ящик», содержащий все возможные признаки, которые могут появиться у изображения I, и все возможные слова, которые могут появиться в его аннотации. Представление изображения в виде набора признаков $\{b_1,\ldots,b_s\}$ тогда можно рассматривать как случайную выборку из этой модели.

Для присвоения изображению J ключевых слов $\{w_1,\ldots,w_t\}$ согласно принципу ранжирования по вероятности, мы должны оценить вероятности $P(w \mid J)$ для каждого ключевого слова w из словаря, т. е. оценить вероятности:

$$P(w \mid I) \approx P(w \mid b_1,\ldots,b_s).$$

Следуя идее моделировать каждое изображение как «черный ящик», предположим, что появление w и $b_1,\ldots,b_s$ на изображении – события взаимно независимые, поэтому мы можем оценить совместное распределение для каждого слова w и признаков $b_1,\ldots,b_s$ как:

$$P(w,b_1,\ldots,b_s) = \sum_{I \in Q} P(I)P(w \mid I)\prod_{i=1}^{s} P(b_i \mid I).$$

Считая распределение изображений в обучающей выборке равномерным, отбросим $P(I)$.

Таким образом, благодаря сделанным предположениям о независимости, нахождение модели релевантности M_I сводится к нахождению оценок $P(w|I)$ и $P(b|I)$ для термов словарей ключевых слов и визуальных признаков соответственно. В качестве моделей, соответствующих распределениям признаков и слов, можно использовать языковые модели. Соответствующие языковым моделям линейные интерполяционные оценки вероятностей имеют вид:

$$P_\lambda(w|I) = (1-\alpha_I)\frac{N(w,I)}{|I|} + \alpha_I \frac{N(w,Q)}{|Q|},$$

$$P_\lambda(b|I) = (1-\beta_I)\frac{N(b,I)}{|I|} + \beta_I \frac{N(b,Q)}{|Q|},$$

где $N(w,I)$ обозначает число раз, которое слово w встречается в аннотации изображения I, $N(w,Q)$ – сколько раз оно появляется во всех аннотациях обучающей выборки, $|I|$ – общее количество слов и признаков у изображения I, $|Q|$ – общее количество изображений в обучающей выборке. $N(b,I)$ и $N(b,Q)$ обозначают то же самое для признаков соответственно.

Параметры сглаживания $0 < \alpha_I < 1$ и $0 < \beta_I < 1$ для слов и признаков изображений соответственно должны выбираться для каждого изображения $I \in Q$ отдельно. Используются разные параметры для модели слов и модели признаков, потому что вероятностные распределения слов и термов в общем случае отличаются. Распределение слов подчинено закону Зипфа, распределение же признаков в словаре признаков зависит от метода кластеризации, применяющегося для построения этого словаря, и является более равномерным.

Возникает проблема выбора параметров α, β. Обе модели можно представить как языковые модели, поэтому для нахождения параметров можно воспользоваться. В [19] предлагается выбирать для подобных языковых моделей параметр сглаживания в диапазоне от 0,6 до0,8. Но, так как распределения слов и признаков различны и имеют специфику обучающей выборки, отличающую их от обычных слов в тексте, моделирующихся языковыми моделями, в данном случае лучшим вариантом будет найти эти параметры на основе обучающей выборки. Для этого в данной работе предлагается модифицированный метод максимизации энтропии сглаженного распределения слов:

$$-\frac{1}{|Q|}\sum_{j=1,\ldots,|Q|} P_\lambda(t|I_j)\log P_\lambda(t|I_j) \to \max_\lambda,$$

$|Q|$ – размер обучающей выборки, $I_j \in Q, \quad j = 1,\ldots,|Q|$ – входящие в эту выборку изображения, t – оцениваемый терм (слово, признак), P_λ зависит от конкретной модели. Вместе с использованием группировки термов по их значениям частоты появления в обучающей выборке такой подход позволяет значительно сократить время вычисления оптимальных параметров. Для нахождения параметров согласно этому подходу необходимо для каждого значения параметра из некоторого диапазона вычислить величину энтропии и взять в качестве оптимального значение параметра, соответствующего максимальной из таких величин.

Таким образом, построив по описанным выше формулам аннотации для всех изображений большой коллекции, мы можем производить поиск по запросу из ключевых слов, возвращая в качестве результата изображения с наиболее пересекающимися с запросом аннотациями.

Выводы

water, sun, landscape, outdoor, lake tree, landscape, night, city, sky, cloud, nature

Рис. 3. Пример работы алгоритма

В данной работе предложена улучшенная вероятностная модель аннотирования изображений, построенная на основе наиболее дискриминативных глобальных признаков изображений, которые позволяют эффективно различать не только типы сцен, но и локальные детали изображений. Предложен модифицированный метод нахождения параметров сглаживания модели. Проведен ряд вычислительных экспериментов, свидетельствующих о справедливости сделанных в работе выводов и доказывающих вычислительную эффективность предложенных методов. Примеры результатов работы предложенного метода приведены на рис. 3.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Mori Y.* Image-to-word transformation based on dividing and vector quantizing images with words / Y. Mori and H. Takahashi and R. Oka // CMU CS Technical Report . 1999.
2. *Duygulu P.* Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary / P. Duygulu, K Barnard, N. de Fretias, D. Forsyth // Proceedings of the European Conference on Computer Vision .– 2002. – P. 97-112.
3. *Jeon J.* Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. / *J. Jeon, V. Lavrenko, R. Manmatha* // International Conference of SIGIR - 2003. – P. 679-714.
4. *Lavrenko V.* A model for learning the semantics of pictures / *V. Lavrenko, R. Manmatha, J. Jeon* // Proceedings of the 16th Conference on Advances in Neural Information Processing Systems NIPS .– 2003. – P. 85-92.
5. *Alexei Yavlinsky.* Image indexing and retrieval using automated annotation // PhD thesis, University of London, Imperial College of Science, Technology and Medicine; Department of Computing, 2007.
6. *Li J.* Real-time Computerized Annotation of Pictures / Jia Li, James Z. Wang // IEEE Transactions on Pattern Analysis and Machine Intelligence – 2008. Vol. 30(6). – P. 985–1002.
7. *Oliva A.* Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope / Aude Oliva, Antonio Torralba // International Journal of Computer Vision. – 2001. Vol. 42. – P. 145-175.
8. *Hancock P.* The principal components of natural images / *P.J. Hancock, R.J. Baddeley, L.S. Smith* // Network. – 1992. Vol. 3. – P. 61–70.

9. *Henderson J.* High level scene perception / J.M. Henderson, A. Hollingworth // Annual Review of Psychology. – 1999. Vol. 50. – P. 243–271.
10. *Manning C.* Introduction to Information Retrieval / Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze // Cambridge University Press. – 2008, 525 p.
11. *Абрамов С.* Мера содержания фона на основе энтропии для поиска и сортировки изображений в базах данных / С.К. Абрамов, В.В. Лукин, Н.Н. Пономаренко // Радиоэлектронные и компьютерные системы. – 2007. Vol. 2(21). – С. 24-28.
12. *Пономаренко Н.* Устойчивый поиск изображений по полному и тематическому подобию с использованием многопараметровой классификации / Н.Н. Пономаренко and В.В. Лукин and С.К. Абрамов // Интернет-математика – 2007. – С. 171-180.
13. *Rao A.* Identifying high level features of texture perception / A.R. Rao, G.L. Lohse // Graphical Models and Image Processing – 1993. Vol. 55. – P. 218–233.
14. *Tamura H.* Texture features corresponding to visual perception / Hideyuki Tamura, Shunji Mori, Takashi Yamawaki // IEEE Trans. On Sys. Man, and Cyb. Vol. 8(6). – P. 460–473.
15. *Navneet Dalal.* Finding people in images and videos // PhD thesis, Institut National Polytechnique de Grenoble. 2006.
16. *Dalal N.* Histograms of Oriented Gradients for Human Detection / Navneet Dalal, Bill Triggs // International Conference on Computer Vision & Pattern Recognition – 2005. – P. 886-893.
17. *Goncharov A.* Pseudometric Approach to Content Based Image Retrieval and Near Duplicates Detection / A. Goncharov, A. Melnichenko // Труды Российского семинара по Оценке Методов Информационного Поиска. – 2008. – P. 120-135.
18. *Melnichenko A.S.* Content-Based Search of Visually Similar Images using Wavelet-Transformation // Тезисы докладов 9-й Международной научной конференции «Распознавание Образов и анализ изображений: новые информационные технологии» - 2008. – С. 26-29.
19. *Zhai C.* The Dual Role of Smoothing in the Language Modeling Approach / Chengxiang Zhai, John Lafferty // Proceedings of the Workshop on Language Models for Information Retrieval – 2001. – P. 31-36.

Мельниченко Александра Сергеевна
Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.
E-mail: alexandramelnichenko@gmail.com.
347928, г. Таганрог, пер. Некрасовский, 44.
Тел.: 8(8634) 371-606.
Кафедра высшей математики; аспирантка.

Melnichenko Alexandra Sergeevna
Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education “Southern Federal University”.
E-mail: alexandramelnichenko@gmail.com.
44, Nekrasovskiy, Taganrog, 347928, Russia.
Phone: 8(8634)371-606.
The Department of Higher Mathematics; post-graduate student.