УДК 004.932

Вестник СибГАУ
2014. № 3(55). С. 120–125

АВТОМАТИЧЕСКОЕ АННОТИРОВАНИЕ ЛАНДШАФТНЫХ ИЗОБРАЖЕНИЙ

А. В. Проскурин

Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева
Российская Федерация, 660014, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31
E-mail: Proskurin.AV.WOF@gmail.com

Поиск изображений в сети Интернет и специализированных базах является актуальной задачей. Для такого поиска целесообразно применять системы автоматического аннотирования изображений на основе низкоуровневых характеристик. Проведен анализ существующих методов автоматического аннотирования изображений, а также алгоритмов автоматической сегментации. Приведены описания модели машинного перевода и алгоритма сегментации JSEG. Предложен набор визуальных признаков для описания областей изображений, включающий статистические признаки второго порядка и фрактальные признаки. Разработан алгоритм AALI на основе модели машинного перевода. Предложенный алгоритм аннотирует с точностью до 88 % и применим для аннотирования изображений в специализированных базах и сети Интернет.

Ключевые слова: ландшафтные изображения, автоматическое аннотирование, алгоритм сегментации JSEG, текстурные признаки.

Vestnik SibGAU
2014, No. 3(55), P. 120–125

AUTOMATIC LANDSCAPE IMAGE ANNOTATION

A. V. Proskurin

Siberian State Aerospace University named after academician M. F. Reshetnev
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660014, Russian Federation
E-mail: Proskurin.AV.WOF@gmail.com

The image retrieval in the Internet and specialized datasets is the important task. For such retrieval is expedient to apply the systems of automatic image annotation (AIA) based on low-level features. Due to wide variety of images, it's sometimes useful to categorize images and to customize methods of AIA according these categories. In this article, the automatic landscape image annotation (ALIA) is discussed. Natural objects (rocks, clouds and etc.) on the landscape images often include just one texture. Because of that, for ALIA enough use of the machine translation model. In this model, the process of image annotation is analogous to the translation of one form of representation (image regions) to another form (keywords). Firstly, a segmentation algorithm is used to segment images into object-shaped regions. Then, cauterization is applied to the feature descriptors that are extracted from all the regions, to build visual words (clusters of visually similar image regions). Finally, a machine translation model is applied to build a translation table containing the probability estimations of the translation between image regions and keywords. An unseen image is annotated by choosing the most likely word for each of its regions. The ALIA algorithm was developed using machine translation model, wherein on the step of segmentation is applied algorithm of color-texture segmentation JSEG. Additionally, before segmentation the image is reduced in order to prevent appearance of small regions and to reduce computational cost, and after segmentation the resulting segmentation map is increased to the size of original image. The region descriptor including second-order statistical features and fractal features was proposed to describe the received segments. The extracted feature vectors are clustered using algorithm k-means. The proposed algorithm annotates the landscape images with 88 % precision and can be applied to annotate the images from specialized image datasets and the Internet.

Keywords: landscape images, automatic annotation, algorithm JSEG, texture features.

Введение. В последние два десятилетия развитие цифровых технологий привело к резкому росту количества изображений в сети Интернет. В связи с этим возникла необходимость в эффективной системе поиска. Существуют три подхода к поиску визуальной информации: составление текстовых аннотаций, анализ содержания, автоматическое аннотирование. В поисковых системах, основанных на первом подходе,

изображениям вручную присваиваются текстовые описания, после чего поиск осуществляется как с текстовыми документами. Использование текстовых запросов удобно для пользователей, однако аннотирование большого количества изображений вручную непрактично, а аннотации часто субъективны. Системы поиска изображений по содержанию лишены этих проблем – поиск производится на основе анализа и сравнения низкоуровневых признаков изображения, таких как цвет или текстура. Однако при этом наблюдается проблема семантического разрыва – отсутствие связи между низкоуровневыми признаками изображения и его интерпретацией человеком. Третий подход предполагает автоматическое аннотирование изображений (ААИ) на основе их низкоуровневых признаков и, таким образом, имеет преимущества двух первых подходов. В связи с большим разнообразием изображений иногда полезно выделять отдельные категории и в соответствии с ними настраивать методы ААИ. В данной статье рассматриваются вопросы автоматического аннотирования ландшафтных изображений (ААЛИ), которые являются широко распространенными объектами поиска.

За последние пятнадцать лет был предложен ряд подходов к автоматическому аннотированию изображений [1]. Большинство из них представляют аннотирование как процесс присоединения к цифровому изображению метаданных в виде ключевых слов. В контексте аннотирования ландшафтных изображений, в которых природные объекты (скалы, облака и др.) часто состоят из одной текстуры либо одна текстура занимает значительную их часть, достаточным является использование модели машинного перевода [2].

Модель машинного перевода. В модели машинного перевода процесс аннотирования рассматривается как перевод из одной формы представления (область изображения) в другую форму (ключевое слово). С этой целью используется таблица перевода. Для ее создания обучающие изображения, имеющие глобальные аннотации, предварительно сегментируются на области. Из областей, размер которых больше определенного порога, извлекаются векторы признаков, при этом с каждым вектором ассоциируется весь набор ключевых слов изображения. Все полученные векторы группируются в кластеры визуально похожих областей изображений, называемых визуальные слова. Сформированным визуальным словам c_j присваивается по одному ключевому слову w_i, вероятность принадлежности этому визуальному слову $P(w_i|c_j)$ которого наибольшая:

$$P\left(w_i \mid c_j\right) = \frac{m_{j,i}}{\sum_{k=1}^{W} m_{j,k}}, \qquad (1)$$

где $m_{j,i}$ – количество векторов признаков в визуальном слове c_j, помеченных ключевым словом w_i; W – количество всех ключевых слов, используемых для аннотирования изображений.

При аннотировании нового изображения оно также разбивается на сегменты, из которых извлекаются векторы признаков. После этого определяется принадлежность каждого вектора тому или иному визуальному слову. Векторам-запросам присваиваются ключевые слова, ассоциированные с визуальными словами, в которые они были определены. Таким образом, в модели машинного перевода каждой области изображения присваивается одно ключевое слово. Рассмотрим подробнее основные моменты этой модели: сегментацию изображения и извлечение низкоуровневых признаков областей.

Алгоритм сегментации JSEG. При решении задачи ААИ целесообразно применение следующих подходов к сегментации изображений [3]: разбиение на прямоугольные плитки; алгоритмы на основе свойств пикселов; методы выделения границ; алгоритмы, основанные на свойствах областей. При сегментировании ландшафтных изображений, в которых текстуры занимают значительную часть изображения, наиболее лучший результат показывают алгоритмы из последнего подхода. Среди них одним из перспективных является алгоритм JSEG, реализующий цветотекстурную сегментацию [4]. Рассмотрим его работу подробнее.

На первом этапе алгоритма JSEG происходит квантование пикселов по цвету без значительного ухудшения качества изображения. После квантования пикселам назначаются метки, уникальные для каждого цвета. Новое созданное изображение меток называется картой классов. На ее основе строится так называемое J-изображение [4], в котором значения пикселов – это J-значения, вычисленные по локальным окнам, центрированным на этих пикселах. Рассмотрим метод вычисления J-значений.

Пусть Z – это набор N точек в карте классов, попавших в локальное окно, а $\mathbf{z} = (x, y)$ – отдельная точка в карте классов ($\mathbf{z} \in Z$). Тогда общая дисперсия координат точек карты классов S_T определяется следующей формулой:

$$S_T = \sum_{z \in Z} \left|\mathbf{z} - \mathbf{m}\right|^2, \qquad (2)$$

где $\mathbf{m}$ – среднее значение координат точек карты классов, попавших в локальное окно.

Предположим, что множество Z отображается в виде C классов Z_i, $i = 1, \ldots, C$. Обозначим сумму дисперсий координат отдельных классов как S_M. Зная общую дисперсию карты классов и сумму дисперсий классов, можно выразить J-значение:

$$J = (S_T - S_M)/S_T. \qquad (3)$$

Более низкое J-значение указывает на то, что пиксел находится внутри однородной области, более высокое значение – пиксел находится на границе областей. Таким образом, J-изображение можно представить в виде 3D-карты поверхности, которая хорошо сегментируется с помощью метода выращивания областей. Выращивание областей заключается в определении набора «семян», имеющих самые низкие локальные J-значения, к которым присоединяются близлежащие пикселы. Процесс сегментации

итеративен: вначале алгоритм сегментирует все изображение с использованием большого локального окна, после чего сегментация повторяется с меньшим масштабом для областей, полученных на предыдущем шаге. После выращивания областей излишне сегментированные области объединяют на основе схожести их цветов. Полученные сегменты в дальнейшем описываются с помощью низкоуровневых текстурных признаков.

Текстурные признаки. Ландшафтные изображения можно интерпретировать как совокупность текстурных фрагментов естественного происхождения. Текстура в широком смысле этого слова – это некоторым образом организованный локальный участок изображения, обладающий свойством однородности видеоданных. Известно три основных подхода к анализу текстур [5]: структурный, статистический и фрактальный. При описании ландшафтных текстур наиболее информативными являются статистические признаки второго порядка и фрактальные признаки. Рассмотрим их подробнее.

Для вычисления статистических признаков второго порядка используются матрицы смежности, элементы $p(i, j)$ которых содержат относительные частоты наличия на изображении двух соседних пикселов с яркостями $i, j \in G$, где $G = \{1, 2, \ldots, N_g\}$ – множество квантованных значений яркости от 1 до N_g. Обычно матрицы смежности вычисляют для горизонтального, вертикального и диагонального направлений. Хараликом было предложено 14 признаков [6], среди которых наиболее информативными являются:

– момент обратной разности – отражает степень разброса элементов матрицы смежности вокруг главной диагонали:

$$T_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+(i-j)^2}; \quad (4)$$

– корреляция как мера линейности регрессионной зависимости яркости на изображении – вычисляется по следующей формуле:

$$T_2 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{(i-\mu_x)\cdot(j-\mu_y)\cdot p(i,j)}{\sigma_x\cdot\sigma_y}, \quad (5)$$

где μ_x, μ_y, σ_x, σ_y – средние значения и среднеквадратичные отклонения элементов матрицы смежности по строкам и столбцам соответственно;

– энтропия – выражает неравномерность распределения яркостных свойств элементов изображения:

$$T_3 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\cdot\log\big(p(i,j)\big); \quad (6)$$

– информационная мера корреляции – рассчитывается как

$$T_4 = \sqrt{1-e^{-2\cdot(HXY-T_3)}}, \quad (7)$$

$$HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)\cdot p_y(j)\cdot\log\big(p_x(i)\cdot p_y(j)\big), \quad (8)$$

где $p_x(i)$, $p_y(j)$ – сумма значений элементов матрицы смежности по строке i и столбцу j соответственно;

– однородность – определяется по формуле

$$T_5 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+|i-j|}. \quad (9)$$

В результате пять статистических признаков (T_1, …, T_5) используются как RD_1, …, RD_5 элементы вектора признаков, описывающих области.

Для определения естественных текстур также часто используют измерение фрактальной размерности D, являющейся уникальной характеристикой природных объектов. Для вычисления фрактальной размерности обычно используется метод покрытия кубами [7]. Представим связную область A на множестве R^n. Предположим, что область A можно покрыть n-мерным кубом размером $L_{\max}$. Если область A является уменьшенной копией с коэффициентом r, то существует $N = r^{-D}$ подобластей. Поэтому число кубов размером $L = rL_{\max}$, необходимых для покрытия всей области, задается выражением:

$$N(L) = 1/r^D = [L_{\max}/L]^D. \quad (10)$$

Простым способом определения величины D из соотношения (10) является покрытие пространства размерностью n сеткой из кубов с длиной стороны L и подсчет количества непустых кубов K. Вычисление параметра $N(L)$ для нескольких значений L позволяет определить размерность D по наклону линии, проходящей через последовательность заполненных ячеек наименьшего размера, расположенных вдоль линии $\{\log L; -\log N(L)\}$.

Однако существует ряд фракталов, имеющих схожие фрактальные размерности, но резко отличающиеся текстуры. Для их описания дополнительно рассчитывается лакунарность (заполнение), которую можно описать следующей формулой:

$$\lambda = (\sigma/\mu)^2, \quad (11)$$

где σ – среднеквадратическое отклонение значений пикселов изображения фрактала; μ – среднее значение пикселов изображения фрактала.

Заполнение мало для плотной текстуры и велико, когда текстура зернистая, что позволяет определить разнородность фрактальных образований. Таким образом, фрактальные признаки D и λ используются как RD_6 и RD_7 элементы вектора признаков, описывающих области. Сформированный вектор признаков **RD** = $\{RD_1, \ldots, RD_7\}$ извлекается из всех областей, полученных после сегментации изображений. Перед дальнейшим их использованием все значения элементов предварительно приводятся к диапазону изменения значений [0; 1].

Алгоритм автоматического аннотирования ландшафтных изображений. На основе модели машинного перевода был разработан метод автоматического

аннотирования ландшафтных изображений. В этом методе основным является алгоритм заполнения таблицы перевода (рис. 1). Рассмотрим его подробнее.

Алгоритм заполнения таблицы перевода выполняется для набора цветных изображений, каждое из которых заранее описано с помощью нескольких ключевых слов. На первом этапе осуществляется предобработка изображений с целью удаления шумов и выравнивания освещенности (рис. 1, блок 2). Для этого применяется фильтр Гаусса [8] и алгоритм «Серый мир» [8] соответственно.

Полученные изображения переводятся в оттенки серого (рис. 1, блок 3), а также используются для формирования карт сегментации, в которых пикселам отдельных сегментов присваиваются разные метки (рис. 1, блоки 4–6). Для этого изображения после предобработки уменьшаются до размеров 320×240 пикселов с целью предотвращения появления небольших областей и уменьшения вычислительных затрат. После этого осуществляется сегментация с помощью алгоритма JSEG [4]. Полученные карты сегментации для уменьшенных изображений приводятся к размеру оригинальных изображений. На следующем шаге карты сегментации накладываются на изображения в оттенках серого (рис. 1, блок 7).

После формирования набора сегментов в оттенках серого из каждого сегмента извлекается вектор признаков **RD** (рис. 1, блок 8). При этом с каждым вектором ассоциируется весь набор ключевых слов, принадлежавших исходному изображению. Извлеченные векторы признаков затем кластеризуются с помощью алгоритма *k*-средних (рис. 1, блок 9). Для сформированных кластеров вычисляются условные вероятности принадлежности им ключевых слов по формуле (1). Ключевые слова с наибольшей вероятностью и центры кластеров (эталоны) заносятся в таблицу перевода (рис. 1, блок 10).

Для аннотирования новых изображений на основе созданной таблицы перевода используется алгоритм, представленный на рис. 1, за исключением блоков кластеризации признаков и заполнения таблицы перевода. Вместо них происходит сравнение извлеченных векторов с сохраненными в таблице эталонами с помощью косинусной метрики. Вектор-запросу присваивается ключевое слово наиболее близкого вектор-эталона. Далее изображение аннотируется ключевыми словами своих вектор-признаков.

Экспериментальные результаты. Для обучения таблицы перевода, а также тестирования алгоритма аннотирования изображений было сформировано 10 тематических наборов ландшафтных изображений (горы, облака, вода и т. д.), содержащих от 50 до 150 изображений (общее количество изображений – 1030), при этом 60 % изображений (620 штук) было включено в обучающую выборку, а остальные (410 штук) – в тестовую. Тематические наборы были составлены выбором ландшафтных изображений из специализированной базы тестовых изображений IAPR TC-12 Benchmark [9]. Пример использованных изображений представлен на рис. 2. В результате ААЛИ могут быть найдены верные ключевые слова (подчеркнутые слова на рис. 2) и ошибочно присвоенные слова.

Рис. 1. Блок-схема алгоритма заполнения таблицы перевода

Рис. 2. Примеры тестовых изображений и их автоматического аннотирования

Точность описания изображений Precision определялась с помощью формулы

$$\text{Precision} = \frac{1}{C} \cdot \sum_{i=1}^{C} \frac{B_i}{A_i}, \qquad (12)$$

где C – количество изображений в наборе; A_i – общее количество автоматически присвоенных изображению i ключевых слов; B_i – количество правильно присвоенных изображению i ключевых слов.

Результаты представлены на рис. 3.

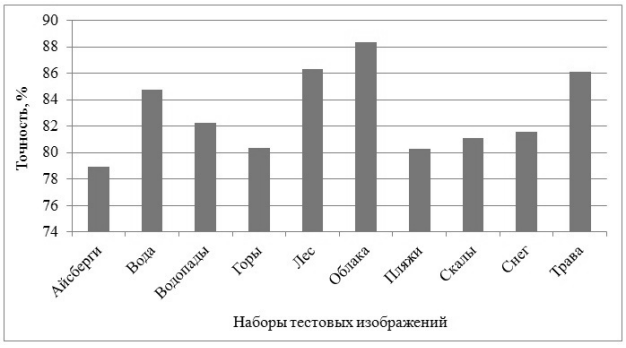

Рис. 3. Диаграмма, отображающая точность автоматического аннотирования

Из представленной диаграммы видно, что процент правильно присвоенных ключевых слов изменяется от 78,9 до 88,3 % в зависимости от преобладающей текстуры на изображении. Относительно низкий процент для гор и скал можно объяснить большим разнообразием текстур, обозначаемых одним и тем же ключевым словом. Также было замечено, что довольно часто происходит аннотирование песка ключевыми словами «снег» и «лед» и наоборот. Это связано с тем, что при определенном масштабе текстуры песка и снега имеют близкие характеристики. Избежать подобных ошибок можно, добавив в вектор признаков цветовые характеристики сегментов.

Для тестирования быстродействия работы алгоритмов было создано 4 набора по 100 изображений, имеющих различные варианты разрешения. Тестирование быстродействия системы производилось для двух алгоритмов системы: алгоритма предварительной обработки и сегментации изображения, алгоритма извлечения признаков и аннотирования изображения (рис. 4). Использовалась системная плата Acer JM50-HR; процессор Intel Core i5-2430M 2,4 ГГц; память (ОЗУ) Kingston 1333 МГц (PC3-10700) DDR3 4 ГБ; видеокарта NVIDIA GeForce GT 540M 1 ГБ; основной жесткий диск Seagate 5400 rpm 500 ГБ. Вычисления производились с использованием одного процессорного ядра.

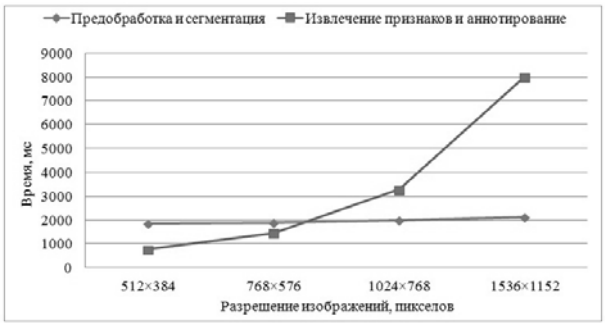

Рис. 4. График зависимости времени выполнения алгоритмов от разрешения изображения

Из графика видно, что время, затрачиваемое на предварительную обработку и сегментацию, возрастает практически линейно с увеличением разрешения изображений. Это связанно с тем, что предварительная обработка в большей степени имеет линейную зависимость, а сегментация выполняется для изображения уменьшенного до одного и того же разрешения (320×240 пикселов). Экспоненциальный рост затрачиваемого времени на извлечение признаков и аннотирование изображения обусловлено особенностью вычисления статистических признаков, требующих большого количества проходов по изображению.

Заключение. В статье представлены результаты исследований автоматического аннотирования ландшафтных изображений. Был применен метод машинного перевода, в котором на этапе сегментации используется алгоритм цветотекстурной сегментации JSEG. Для описания ландшафтных изображений предложен набор текстурных признаков, включающий статистические признаки второго порядка и фрактальные признаки. Разработанный алгоритм позволяет автоматически аннотировать ландшафтные изображения с достоверностью от 78,9 до 88,3 %. Он может быть использован для аннотирования специализированных баз ландшафтных изображений и изображений в сети Интернет. Для повышения точности аннотирования планируется расширение вектора признаков, описывающих текстурные фрагменты, а также применение адаптивных методов кластеризации.

Библиографические ссылки

1. Zhang D., Islam Md. M., Lu G. A review on automatic image annotation techniques // Pattern Recognition. 2012. Vol. 45(1). P. 346–362.

2. Duygulu P., Barnard K., Freitas N., Forsyth D. Object recognition as machine translation // In The Seventh European Conference on Computer Vision. Part IV. 2002. P. 97–112.

3. Dey V., Zhang Y., Zhong M. A review on image segmentation techniques with remote sensing perspective // Proceedings of the International Society for Photogrammetry and Remote Sensing Symposium (ISPRS10). Vol. XXXVIII, Part 7A. 2010. P. 31–42.

4. Deng Y., Manjunath B. S. Unsupervised Segmentation of Color-Texture Regions in Images and Video // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001. Vol. 23(8). P 800–810.

5. Фисенко В. Т., Фисенко Т. Ю. Компьютерная обработка и распознавание изображений : учеб. пособие. СПб. : Изд-во ИТМО, 2008. 195 с.

6. Haralick R. M., Shanmugam K., Dinstein I. H. Textural Features for Image Classification // IEEE Trans. on Systems, Man and Cybernetics. 1973. Vol. 3(6). P. 610–621.

7. Favorskaya M. N., Petukhov N. Y. Recognition of natural objects on air photographs using neural networks // Optoelectronics, Instrumentation and Data Processing. 2011. Vol. 47(3). P. 233–238.

8. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М. : Изд. дом «Техносфера», 2008. 1072 с.

9. IAPR TC-12 Benchmark [Электронный ресурс]. URL: http://www-i6.informatik.rwth-aachen.de/imageclef/resources/iaprtc12.tgz (дата обращения: 20.11.2013).

References

1. Zhang D., Islam Md. M., Lu G. A review on automatic image annotation techniques. *Pattern Recognition*. 2012. Vol. 45(1). P. 346–362.

2. Duygulu P., Barnard K., Freitas N., Forsyth D. Object recognition as machine translation. *In The Seventh European Conference on Computer Vision*. Part IV, 2002, p. 97–112.

3. Dey V., Zhang Y., Zhong M. A review on image segmentation techniques with remote sensing perspective. *Proceedings of the International Society for Photogrammetry and Remote Sensing Symposium (ISPRS10)*. Vol. XXXVIII, Part 7A, 2010, p. 31–42.

4. Deng Y., Manjunath B. S. Unsupervised Segmentation of Color-Texture Regions in Images and Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2001, Vol. 23(8), p. 800–810.

5. Fisenko V. T., Fisenko T. Yu. *Komp'yuternaya obrabotka i raspoznavanie izobrazheniy* [Computer image processing and recognition]. St. Petersburg, ITMO Publ., 2008, p. 195.

6. Haralick R. M., Shanmugam K., Dinstein I. H. Textural Features for Image Classification. *IEEE Trans. on Systems, Man and Cybernetics*. 1973, Vol. 3(6), p. 610–621.

7. Favorskaya M. N., Petukhov N. Y. Recognition of natural objects on air photographs using neural networks. *Optoelectronics, Instrumentation and Data Processing*. 2011, Vol. 47(3), p. 233–238.

8. Gonzalez R. C., Woods R. E. Digital image processing (3rd edition). *Prentice-Hall*, Inc. Upper Saddle River, NJ, USA, 2006, p. 976.

9. IAPR TC-12 Benchmark. Available at: http://www-i6.informatik.rwth-aachen.de/imageclef/ resources/iaprtc12.tgz (accessed 20.11.2013).

© Проскурин А. В., 2014