Automatic Image Annotation and Retrieval

May The` Yu , Myint Myint Sein
University of Computer Studies, Yangon
maytheyu@gmail.com, myintucsy@gmail.com

Abstract

In this paper, an automatic image annotation and retrieval model is developed base on the intensity invariant approach. The given uncaptioned image is divided into background and foreground images and segmented into regions, which are classified into region types using a variety of features. Firstly, preprocessing stages such as gray-scale converting, noise filtering for image enhancing is processed. After segmentation, calculate the eigenvectors of images and examined the associated word by using database. The various types of images are applied for training and testing. The top words are described for annotated image in result. Manual image annotation is time-consuming, laborious and expensive; so, there has been a large amount of research done on automatic image annotation and retrieval technologies are combined to improve the performance.

Keywords

Automatic image annotation, image retrieval

1. Introduction

Nowadays, digital photography is a common technology for capturing and archiving images due to the decreased costs for multimedia recording and storage devices, high transmission rates, and improved compression techniques, the digital image collections have grown rapidly in recent years. Annotation is considered as a pre-stage of the retrieval process. Image annotation, also known as image tagging, is a process by which labels or tags are associated with images, either manually, automatically or semi-automatically. A computer system automatically assigns metadata in the form of text description or keywords to a digital image are called automatic image annotation or image annotation propagation.

There exist some methods that cluster image representations and text to produce a representation of a joint distribution linking images and words. Image segmentation into regions may help to find out the semantic relation between words and objects contained in image. To segment an input image, we use N-Cut that measures both the total dissimilarity between the different groups as well as the total similarity within the groups. The objective of N-Cut is to use the low-level coherence of brightness, color, texture or attributes to sequentially come up with hierarchical partitions. N-Cut is an unbiased measure of disassociation between subgroups of a graph and it has the nice property that minimizing normalized cut leads directly to maximizing the normalized association, which is an unbiased measure for total association within the subgroups. The computational time of N-Cut is also high. [1]

The main challenge in automated image annotation is to create a model able to assign visual terms to an image in order to successfully describe it. The starting point is a training set of images that have already been annotated by humans. Image analysis techniques are used to extract features from the images such as color, texture, and shape, in order to model the distribution of a term being present in the image. Features can be obtained from the blobs, which are segmented parts of the image. To extract the same feature information from an unseen image, the system compares it with all the previously created models.

The main stages of automatic image annotation are as follow:

1. Image processing consists of extracting image data (i.e. region segmentation) from the image.

2. Semantic learning consists of two processes. First, similar extracted image data will be grouped into clusters. Second, modeling the correlation between the visterm and the textual information is performed.

3. Annotation scheme consists of an image processing (segmentation and clustering) and label the words for new image. [2]

The rest of this paper is organized as follows. Section 2 consist related work of this paper, Section 3 explain the overview of the system, Section 4 represent the image segmentation process of the system and Section 5 is conclusion of this paper.

2. Related work

Different machine learning methods for image annotation model the association between words and images or image regions. Isolated pixels or even regions in an image are often hard to interpret.

Jin et al. [3] propose a new framework for automated image annotation that estimates the probability for a language model to be use for annotation an image. They use a word-to-word correlation which is taken into account through the Expectation Maximization (EM) algorithm for finding optimal language model for the given image.

J. Jeon al. at [4] who use as annotation framework the Cross-Media Relevance Model (CMRM) and apply the Automatic Local Analysis, a method for performing query expansion in Information Retrieval.

Luis Garcia Ugarriza, Eli Saber[6] addresses the problems of automatic image annotation (AIA) for the purpose of image indexing & retrieval in an Annotation Based Image Retrieval (ABIR) system.

J.Shi and J. Malik. [1] developed a grouping algorithm based on the view that perceptual grouping should be a process that aims to extract global impressions of a scene and provides a hierarchical description of it and proposed the normalized cut criteria for segmenting the graph.

Changhu Wang [5] presents a multi-label sparse coding framework for feature extraction and classification within the context of automatic image annotation.

Trong-Tˆon Pham [2] presents an automatic image annotation system using a fusion of region-based and saliency-based models .In this paper, he proposed three main stages of an automatic image annotation system. There are image processing, semantic learning, and annotation scheme. In image processing, they extract image data using region segmentation and saliency point detection.

Jia-Yu [7] presents correlations between image features and keywords, so they can automatically find good keywords for a new image. In this paper, they proposed 4 methods to estimate a translation table.

J Jeon, V Lavrenko, and R Manmatha[8] present an automatic approach to annotating and retrieving images based on a training set of images.

Jin *et al*. [9] have done pioneering work on annotation refinement using a generic knowledge-based WordNet. From a small candidate annotation set obtained by an annotation method, irrelevant annotations are pruned using WordNet.

3. System Overview

In this system, first the user chooses the desired image to annotate. When the uncaptioned image is entered, the system made pre-processing steps such as gray-scale converting, noise filtering and so on. After converting gray-scale image, the system process segmentation step. In segmentation step, we propose the integration of two segmentation methods. The two segmentation methods are color-based segmentation method and N-cut segmentation methods. This method segments incoming input image by computing their color intensity. To reduce time consuming, the system divides the 3*3 image. Before comparing with all images in datasets, the upper regions are compare with sky, the lower regions are compare with grass and middle image are compared with others according to their color intensity. In order to this stage, processing time is faster than other methods. In feature extraction, the system extracts individual images from output of segmentation process. And then individual image are compared with annotated dataset to caption their associated words. Then individual segmented images are combined to form output annotated image. Finally the system annotates the image with their associated words. The overview of the porposed system is shown in figure1.

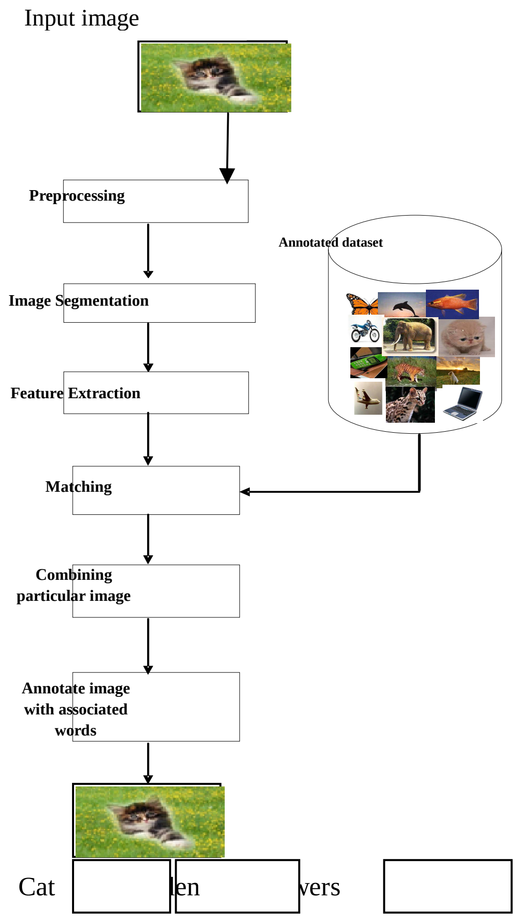

Figure1. System overview of automatic image annotation

4. Image Segmentation

Image segmentation is a fundamental task in many computer vision applications and it can be defined as the classification of all the pictures elements or pixels in an image into different clusters that exhibit similar features. The advancement in color technology facilitated the achievement of meaningful color segmentation. This stage converts the input color image to segmented image [8]. This stage is the main stage for automatic image annotation system. The original image, converted gray-scale image, edge image and segmented image are shown in figure2, figure3, figure4 and figure5 respectively.

Figure2. Original input image

Figure3. Gray scale image

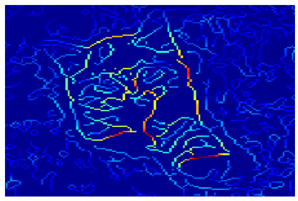

Figure4. Edge of image

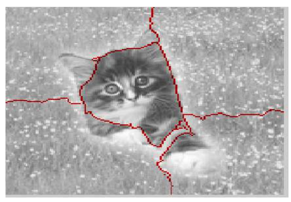

Figure5.Segmented image

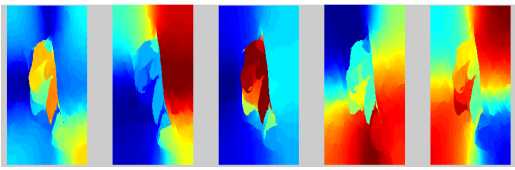

Figure6. Segmented parts of an image

A graph $G = (V, E)$ can be partitioning into two disjoints sets, A,B, $AUB=V$, $A \cap B = \Phi$, by simply removing edges connecting the two parts. The degree of dissimilarity between these two pieces can be computed as total weight of the edges that have been removed. In graph theoretic language, it is called the cut:

The optimal bi-partitioning of a graph is the one that minimizes this cut value.

$$cut(A,B) = \sum_{u \in A, v \in B} w(u,v) \qquad (1)$$

Given a partition of nodes of a graph, V, into two sets A and B, let x be an $N=|V|$ dimensional indicator vector, $x_i = 1$ if node i is in A and -1, otherwise .Let $d(i)=\Sigma_j w(i,j)$ be the total connection from node I to all other nodes. With the definitions x and d, we can rewrite N*cut* (A,B) as:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,B)} + \frac{cut(B,A)}{assoc(B,V)}$$

$$= \frac{\sum_{(x_i>0, x_j<0)} -w_{ij}x_i x_j}{\sum_{x_i>0} d_i} + \frac{\sum_{(x_i<0, x_j>0)} -w_{ij}x_i x_j}{\sum_{x_i<0} d_i} \qquad (2)$$

The advancement in color technology facilitated the achievement of meaningful color segmentation. This stage converts the input color image to segmented image [6]. A feature is defined as an "interesting" part of an image, and features are used as a starting point for many computer vision algorithms. Since features are used as the starting point and main primitives for subsequent algorithms, the overall algorithm will often only be as good as its feature detector. In this system, feature extraction is based on their segmentation image. Individual image are extracted from input image by using color intensity value of their eigenvectors.

This is the overall processes of the automatic image annotation. And then the system annotates suitable words for input image. Experiment in the Corel dataset is performed to validate our system and the results are promising. Figure 6 shows the segmented parts of the input image by N-cut eigenvectors and figure7 is the final annotated image.

Cat Garden ers

Figure7.Annotated output image

5. Conclusion

The main aim of automatic image annotation is to create a model able to assign visual terms to an image in order to successfully describe it. In this system, we use dataset, where all annotations are approximately equal in length and words reflect prominence of objects in the image. There is a problem of very large collections of digital images without annotations continue to grow. Automatic image annotation emerged as a solution to the time-consuming work of annotating large datasets and as an intermediate step in the retrieval process.

An image retrieval system is a computer system for browsing, searching, and retrieving images from a large database of digital images. We have presented a simple framework for automatic image annotation based on image segmentation. Since generally it is tedious and time-consuming for humans to manually annotate the keywords in the object/region level for data collection. The annotation system needs many images that are very similar to the image to annotate to perform a correct annotation [7]. In our experiment, the integration of color-based segmentation and N-Cut algorithm are used to segment the image. In our system, first segment an unknown image using color-based segmentation algorithm and then to retrieve more accurate segments, the system used N-Cut algorithm.

References

[1] J.Shi and J. Malik. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):888–905, 2000.

[2] Trong-Ton Pham. Automatic Image Annotation: Towards a Fusion of Region-based and Saliency-based Models : September 2006

[3] Jin,R., Chai, J.Y., Si, L.: Effective automatic image annotation via a coherent language model and active learning. In: Proceedings of the 12th International ACM Conference on Multimedia (MM), New York,NY, USA, ACM(2004)

[4] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In Proceedings of the 26th annual international ACMSIGIR conference on Research and development in information retrieval, pages 119–126. ACM Press, 2003

[5] Changhu Wang, Shuicheng Yan, Lei Zhang, Hong-Jiang Zhang : Multi-Label Sparse Coding for Automatic Image Annotation, 2008

[6] Luis Garcia Ugarriza, Eli Saber, Aurtomatic Image Segmentation by dynamic region growth and multi-resolution merging, IEEE Transactions on image processing, Vol.18, No.10,October 2009

[7] Jia-Yu Pan, Hyung-Jeong Yang, Pinar Duygulu and Christos Faloutsos,"Automatic image captioning", proceedings of the 2004 IEEE International conference on multimedia and expo(ICME'04)

[8] J Jeon, V Lavrenko, and R Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, 2003.

[9] Y. Jin, L. Khan, L. Wang and M. Awad. Image annotations by combining multiple evidence & wordNet. In Proceedings of the 13th Annual ACM international Conference on Multimedia (Hilton, Singapore, November 06 - 11, 2005).

[10] Y. Mori, H. Takahashi, and R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999.

[11] Kobus Barnard et al., Matching Words and pictures, Journal of Machine Learning Research, Vol 3, pp 1107-1135. 2003