Automatic image captioning system using integration of N-cut and color-based segmentation method

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

Abstract

Automatic image annotation is the task of automatically assigning words to an image that describes the content of the image. In this paper, an automatic image annotation and retrieval model is developed base on the intensity invariant approach. The system applied various preprocessing steps such as Gray scale converting, Noise filtering for image enhancing. In image segmentation, the system segment particular image according to their color intensity. In feature extraction, it extracts individual objects and assign associated words for each object to form training images. After feature extraction, each object is compared with annotated dataset to assign words and images respectively. The fabric and un-fabric images are applied for training and testing. The top words are described for annotated image in result.

Keywords*: image annotation, segmentation, fabric image*

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.

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. A computer system automatically assigns metadata in the form of text description or keywords to a digital image are called automatic image annotation.

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, N-Cut 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[1].

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. 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. Hugo Jair Escalante Balderas [6] presents methods for image annotation and retrieval that are based on the semantic cohesion among terms. This proposed hierarchy is composed of six branches to annotate.

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 rest of this paper is organized as follows. Section 2 express the types of image that used in

system, Section 3 represent the system overview and Section4 represents the segmentation process and Section5 describes the experimental result and Section6 is conclusion of this paper.

2. Fabric Image

The various types of fabric images are used for annotation. Two or more things are included in the background of fabric image. And it is difficult to extract background and foreground features from image. Fabric refers to any material made through weaving, knitting, spreading, crocheting, or bonding that may be used in production of further goods. There are many types of fabric image. They are as follows:

Primary fabric — a fabric created during the original formation of the rock.

Shape fabric — a fabric that is defined by the preferred orientation of inequant elements within the rock.

S-fabric — a planar fabric such as cleavage or foliation

L-fabric — a linear fabric such as mineral stretching lineation

Penetrative fabric — a fabric that is present throughout the rock

Some images for annotation are shown in figure 1.

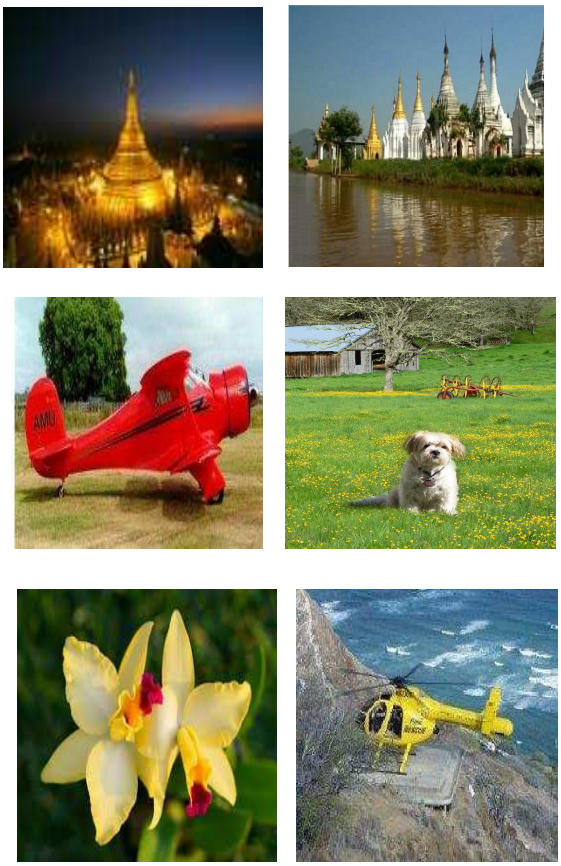


Figure1. Some images

3. System Overview

The system overview of the proposed automatic image annotation system is shown in figure 2. 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, the integration of two segmentation methods is proposed. 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. 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.

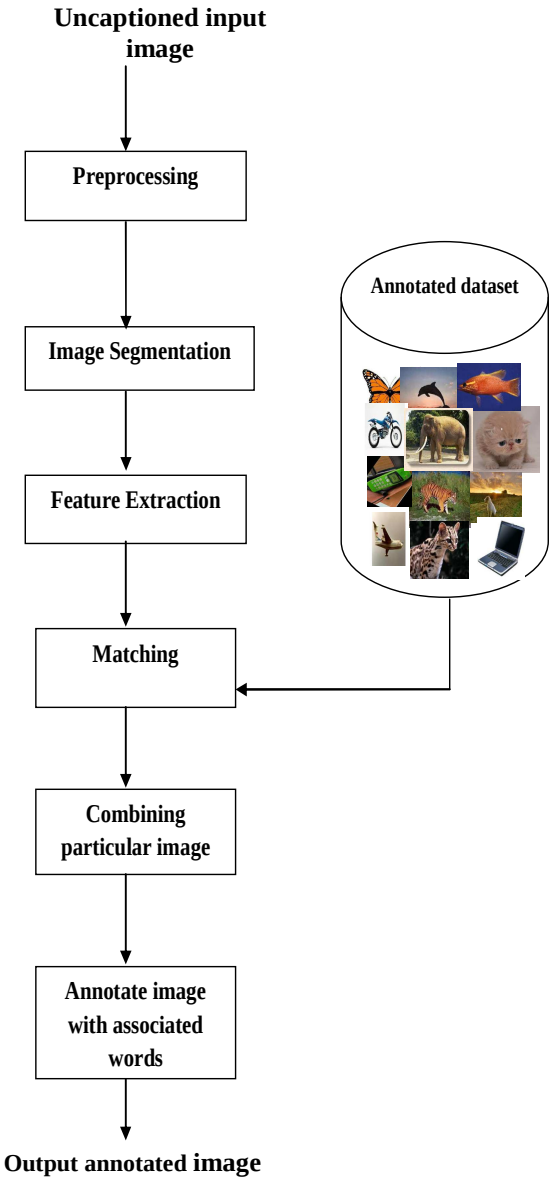


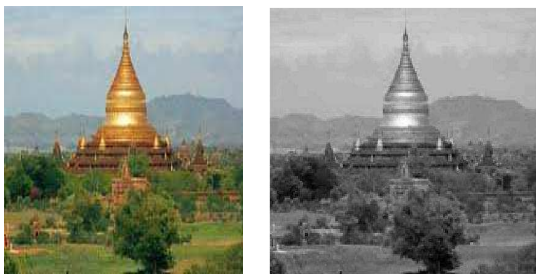
Figure2. System overview of automatic image annotation

Then individual segmented images are combined to form output annotated image. Finally the system annotates the image with their associated words.

4. Segmentation

4.1 Pre-processing for annotation

This section discusses the pre-processing stage for automatic image annotation system. It is needed to enhance the image and reduce the artifacts that appear during the acquisition process. Noise filtering, grayscale converting and histogram equalization are performed in this stage. Figure 3 illustrated the result image of enhancing process.



(a) **(b)**

Figure 3. Enhanced image

.

4.2 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. A clustering method based on graph partitioning is applied for segmentation.

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.

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

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, then rewrite N*cut* (A,B) as:

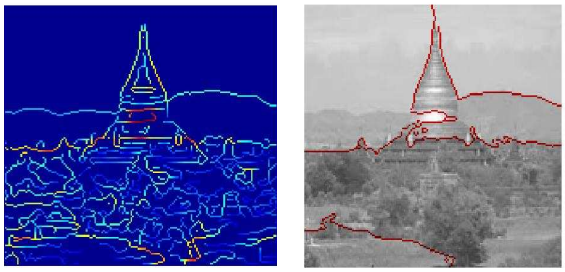


Figure.4 Edge image

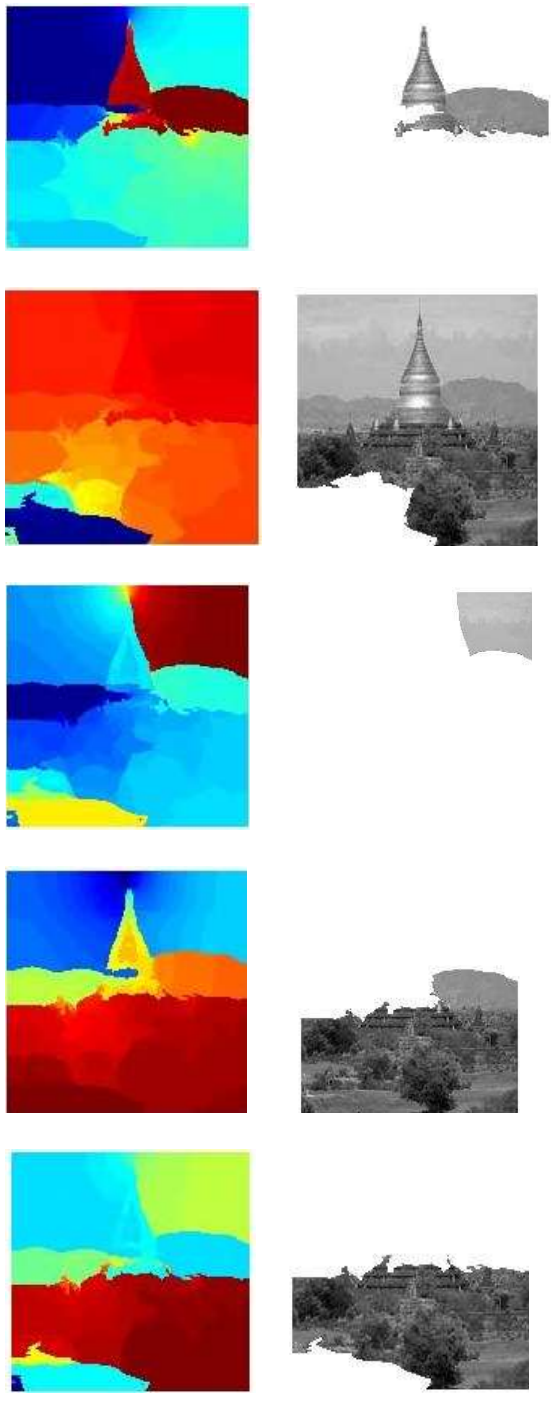


Figure5. Segmented parts of an image

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

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

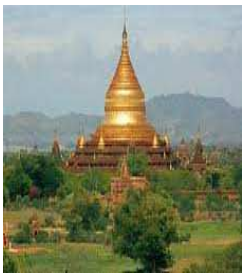
The advancement in color technology facilitated the achievement of meaningful color segmentation. This stage converts the input color image to segmented image. The edge image and segmented parts are described in figure 4 and figure 5, respectively.

4.3 Feature Extraction

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.

5. Experimental Result

Both fabric image and un-fabric images are used for image annotation. Five segmented parts are used to extract feature from an image for development. After feature extraction process, the segmented images are compared with the image in the annotated dataset. Matching in this stage, the system produces the predicted annotated words for image. At combining stage, individual images are combined to form the output annotated image .And then the system annotates suitable words for input image. Finally, the system combined the individual images are combined to form output annotated image. Experiment in the Corel dataset is performed to validate our system and the results are promising.



Pagoda	Sky	Mountain	Bushes

Figure6. Annotated output image

6. 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, fabric and un-fabric images are used for image annotation, 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. This system presents a simple framework for automatic image annotation based on image segmentation and expresses the segmentation time in seconds. Since generally it is tedious and time-consuming for humans to manually annotate the keywords in the object/region level for data collection. This system is developed in segmentation stage by many segmentation parts of an image. So, features are accurately extracted for annotation. Therefore, segmentation and features extraction stages are more mentioned for image annotation system.

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] Hugo Jair Escalanre Balderas,Manuel Montes, Enrique Sucar, “Semantic Cohension for Image Annotation and Retrieval” ,2010

[7] Jia-Yu Pan, Hyung-Jeong Yang, Pinar Duygulu and Christos Faloutsos,”Automatic image captioning”, proceedings of the 2004 IEEE (ICME’04)