Automatic Image Labeling System for Outdoor Scene

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

Abstract

Automatic image annotation is important to both video and image retrieval and computer object recognition. To ensure easy sharing of an effective searching over a huge and fast growing number of online images, real-time automatic annotation by words is an imperative but highly challenging task. The system tends to partition the image plane to a set of non-overlapping regions with coherent image appearance. In the proposed system, the efficient segmentation algorithm is developed by combining the N-cut segmentation algorithm to partition the image region and color component analysis to extract center region. In the system, there are six number of segmented parts are used. Five segmented parts are retrieved by using N-cut and center image is retrieved using color component analysis method.

Keywords: segmentation, image matching, image annotation

1. Introduction

Automatic Image Annotation (AIA) is on the frontier of different fields: image analysis, machine learning, media understanding and information retrieval. The aim of annotation methods is to attach textual labels to un-annotated images or the unlabeled images, as the descriptions of the content or objects in the images. The process of assigning automatically some text description to an unknown image (or partly unknown) is called automatic image annotation. The annotation system needs many images that are very similar to the image to annotate to perform a correct annotation.

Automated annotation depends on some advances in image segmentation result; the system might be to recognize categories of objects. To achieve the better annotated result, image segmentation is important step. In the proposed system, efficient automatic image annotation system emerged as a solution to the time–consuming work of annotating large datasets such as nature scenes with the RGB color value. The main aim of the proposed system is to create a model able to assign visual terms to an image in order to successfully describe it.

To achieve the precise segmented parts, the efficient segmentation algorithm is developed by combining the N-cut segmentation algorithm and color component analysis to extract center region. After segmentation, calculate the color values of each segmented images and examine the associated word by matching with color database. The main objects in the input image are annotated automatically with their associated text. Input images are nature scene images.

2.Related Work

The method of a color image segmentation system performs color, clustering in a color space followed by color region segmentation in the image domain. The region segmentation algorithm merges clusters in the image domain based on color similarity and spatial adjacency is present in Color Image Segmentation in the Color and Spatial Domains.Recently, Gould et al. [9] proposed a model for segmentation and labeling where new region hypotheses were generated through a sequential procedure, where uniform label swaps for all the pixels contained inside individual segment proposals are accepted if they reduce the value of a global energy function.Boykov and Jolly [10] introduce a technique for segmentation using a graph to represent the image, and then using a Min-Cut/Max-Flow algorithm to segment the graph. Pixels in the image are represented by nodes on the graph. The edge weights on the graph are defined by a cost function, which is defined by region and boundary information from the image.S. Hamid Amiri, Mansour Jamzad[11] propose an annotation approach which follows the ALIPR structure. To describe the image contents, authors proposed an approach which extracts two discrete distributions as signatures for color and texture.

3. System Overview

In the proposed system design, there are six main steps: image preprocessing, image enhancement, color image segmentation, image feature extraction, image matching, and image labeling. These steps are shown in Figure1. Image resizing, noise filtering, de-blurring processes are the main process in the pre-processing step. The input image is firstly cropping to retrieve center image by using color component analysis method. And then the remaining image is segmented by using N-cut segmentation algorithm.

After segmenting, the system compares the segmented parts with the stored images in database. There are two steps in image matching.In the early perception stages of human beings, similar colors are grouped together to fulfill the tasks of identification.

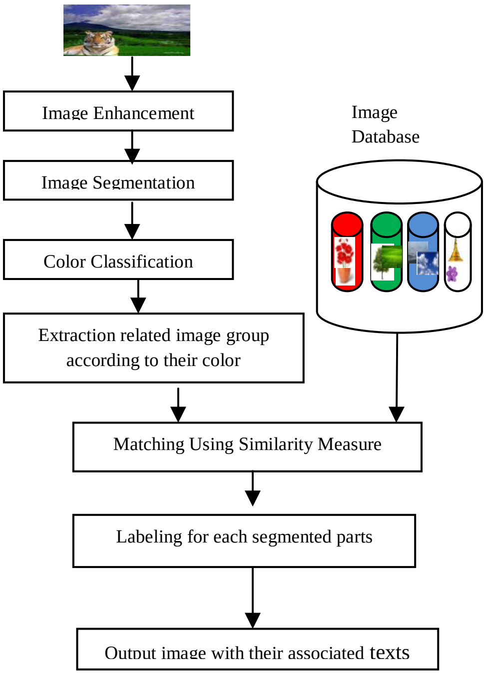

Figure1. The proposed System Design

Based on this assumption, the systems presents a scheme in this section to group similar colors of an image and extract the characteristic colors as the features for identification. The proposed scheme for extracting the color features is similar to the design of the color palette for an image.If the color of center image is red, the system loads the red database and matches the image with the segmented images in this red database. If the color is green, the system matches with the green database. Image matching aims at establishing the relation between two images. Image matching is also applied when features, others than points. The similarity measure has to measure how well the two images match with each other.

So, the test image is compared with only in the blue database, not in other database. According to this process, the matching with the particular database is faster than matching with the whole database.The system can classify each pixel in the image by calculating the Euclidean distance between that pixel and each color marker. Traditionally, Euclidean distances are used to measure the similarity between the query and the images in the database.The smallest distance will tell that the pixel most closely matches that color marker.

4. Image Enhancement

To obtain the high quality image, some pre-processing steps are acquired. In the proposed system, input image is firstly converted RGB to gray. It converts RGB values to gray-scale values by forming a weighted sum of the *R*, *G*, and *B* components are shown in equation (1):

$$0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

Digital images are prone to a variety of types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene[4]. Some noises in images are Gaussian noise, Rayleigh noise, Erlang (Gamma) noise, Exponential noise, Uniform noise and Impulse (salt and pepper) noise.

5. Image Segmentation

In the task of segmentation of some complex pictures, such as outdoor and natural images, it is often difficult to obtain satisfactory results using only one approach to image segmentation. Segmentation is usually the first task of any image analysis process, and thus, subsequent tasks rely heavily on the quality of segmentation. For this reason, a considerable care is taken to improve the probability of a successful segmentation.

5.1. Center Image Segmentation

Center image segmentation is the important step in this system, because center image is focused while everybody takes the photo. Center image is a main image in the whole photo. People who took the photo focus the main object in the center of the image. So, the center image is very important over the whole image. The center image gives the accurate text for annotation. Cropping an image is the really important task as it may be needed to extract some specific part of an image or to change aspect ratio of an image. In the proposed system, connected component analysis method is developed for center image segmentation shown in Figure2.

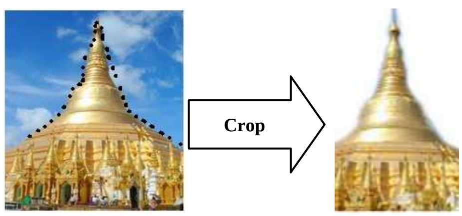

Figure2. Center Image Segmentation

5.2. Color Image Segmentation

The color has great importance in content-based image retrieval systems, which is stored in the intensity vectors of image pixels, and this information can be retrieved simply. However the color can be represented in different color spaces, and that is not irrelevant which color space is used in an application[3]. The color information of an image represented in an arbitrary color space can be stored on several ways, but that is strongly application dependent which representation method is the most efficient at a determined search.

One of the major problems of color segmentation is how to employ color information as a whole for each pixel. When color is projected onto three components, color information is so scattered that color images become simply multispectral images and the kind of information that humans can perceive is partially lost.RGB space is a widely used color space for image display shown in Figure3.

Figure3. RGB Color Image

A normalized cut (N-cut) is one that seeks to create a cut, which divides the input graph into "object-like" regions. Such a cut minimizes the capacity across the cut, while maximizing the capacity within the regions created by the cut. A graph G = (V,E) can be partitioned into two disjoint sets, $A, B, A \cup B = V, A \cap B = 0$ 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. The weight on the graph edge connecting node u and v is set to be $w(u,v) = e^{-(d(u,v)/\sigma_x)}$,where d(u,v) is the Euclidean distance between the two nodes, and σ_x controls the scale of the spatial proximity measure. σ_x is set to be 2.0. The theoretic language it is called the cut:

$$cut(A,B) = \sum_{u \in A, v \in B} w(u,v) \qquad (2)$$

In the N-cut segmentation algorithm, the numbers of segmented parts are started from 2 and above. Figure4 is represented for the value of n=2, n=3,n=4 and n=5. The features of image are not sure to annotate. The value of n=5 is the most suitable cut for this system because parts of the image are certainly segmented according to their region values.

By minimizing this cut value, one can optimally bi-partition the graph and achieve good segmentation.

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \quad (3)$$

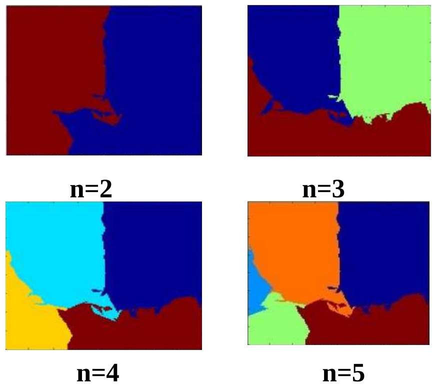

Figure4. Color Segmented Parts using Various N-cut values for Segmentation

5.3 Image Segmentation with Proposed Modified N-cut Algorithm

The modified N-cut segmentation algorithm is used to extract the center image by using connected component labeling. In addition to the extracted segments of the N-cut algorithm, the center image extraction is performed to accurately annotate the main object. The N-cut algorithm extracts the five segments of the images and the modified N-cut algorithm extracts the six segments of the image including the center image. The segmentation time of the proposed segmentation algorithm is not a big deal. The better annotation result is achieved because of the accurate segmentation of the main object.

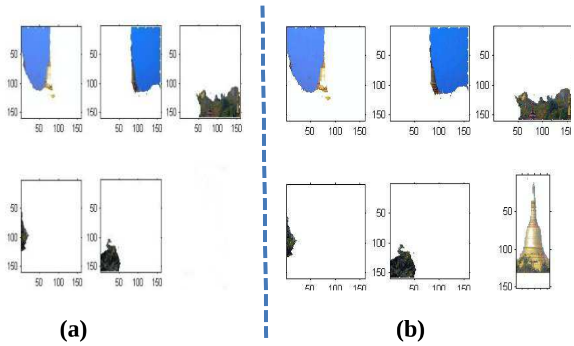

Figure5. Segmented Result using (a)Original N-cut and (b) Modified N-cut Algorithm

6.Image Matching

Image matching aims at establishing the relation between two images[8]. Image matching is also applied when features, others than points. Two types of image matching techniques are used in this paper. They are image matching using color and image matching using similarity measure.

6.1 Image Matching using Color

Color features are among the most important features used in image database retrieval. Due to its compact representation and low complexity, direct histogram comparison is the most commonly used technique in measuring color similarity of images. However, it has many serious drawbacks, including a high degree of dependency on color codebook design, sensitivity to quantization boundaries, and inefficiency in representing images with few dominant colors[5].

Color matching quantifies which colors and how much of each color exists in a region of an image and uses this information to check if another image contains the same colors in the same ratio. Use color matching to compare the color content of an image or regions within an image to reference color information [7].

The numbers of images used in the proposed system are described in Table 1 according to their associated color database (RED, GREEN, BLUE, and YELLOW). In the entire database, 1000 images are used in RED database, 1100 images are used in GREEN database, 900 images in BLUE database and 1000 images are used in YELLOW database.

In RED database, there are 150 images of fruits,150 of flowers,100 building, 100 of cars, 100 of airplane, 50 of motorbike, 50 of towers, 150 of pagoda, 150 of animals. In GREEN database,there are150 images of fruits,150 of flowers,100 building, 100 of cars, 100 of airplane, 50 of motorbike, 100 of leaf, 100 of tree, 100 of grass and 150 of animal. In BLUE,there are 150 images of fruits, 150 of flowers,100 building, 100 of cars, 100 of airplane, 50 of motorbike,50 of sky, 50 of water,150 of animal.In YELLOW database,there are150 images of fruits,150 of flowers,100 building, 100 of cars, 100 of airplane, 50 of motorbike, 200 of pagoda and 150 of animal.

Table 1.The Number of Images in Color Databases

Database	Number of Images
RED	1000
GREEN	1100
BLUE	900
YELLOW	1000

In the Figure7, the color value of segmented part is (118,166,238), so the system determines to compare the color database is blue. Furthermore, since color alone has little semantic meaning, color features tend to be more domain independent compared to other features such as shape or even texture[6]. In this system, color matching is not sufficient to be annotated the object is. Therefore, image matching using similarity measure is second step in image matching.

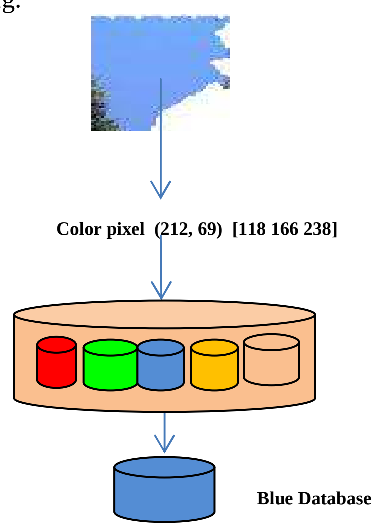

Figure7. Matching with Color Value

6.2 Image Matching Using Similarity Measure

The similarity measure has to measure how well the two images match with each other. This is usually based on the normalized cross correlation or on a function of the intensity differences of the values themselves, actually measuring the distance between the images.The basis of many measures of similarity and dissimilarity is Euclidean distance. The distance between vectors X and Y is defined ineq(4).

$$d(x,y)=\sqrt{\sum_{i}^{n}(x_i-y_i)^2} \quad (4)$$

In other words, Euclidean distance is the square root of the sum of squared differences between corresponding elements of the two vectors.Euclidean distance is only appropriate for data measured on the same scale[1]. In Figure 8, input images is segmented using n=5. But the two blue images and 3 green images are segmented using N-cut segmentation method. According to the proposed system, center image is one of the segmented parts, so the final segmented parts are 6 parts. The same images are compared once with their associated database, so the system can reduce the computation time.

Notice that for this kind of data, the variables are the columns. A variable records the results of a measurement. For our purposes, in fact, it is useful to think of the variable as the measuring device itself. This means that it has its own scale, which determines the size and type of numbers it can have. Direct or indirect connections between images and keywords can be used to extend relevant image set.

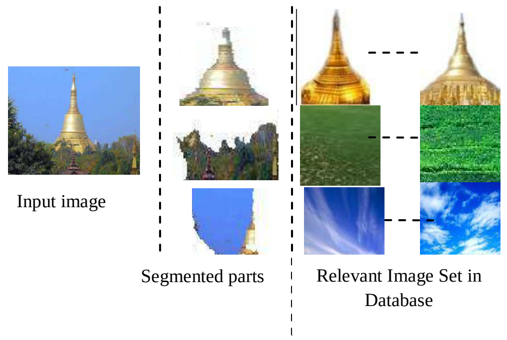

Figure8.Relevant Image Set with the Segmented Parts of an Image

7. Experimental Results

Image is classified into one or more classes, and finally annotated by propagation of the corresponding class words [2].The final result of automatic image annotation is described in Figure 9.

Figure 9. Annotated Output Result

In the Figure10, image annotation performance is evaluated by comparing the captions automatically generated for the test set, with the human-produced ground-truth. Images in column 1 and column 2 are identical with the manual annotation while image in column 3 differs in one word. In the proposed system, the general terms are used for annotation (e.g., animal" instead of "Tiger").

Images			
Original Annotation	Sky Tree Pagoda	Sky Grass Tower	Sky Grass Tiger
Automatic Annotation	Pagoda Sky Tree	Tower Sky Grass	Animal Sky Grass

Figure 10.Automatic Annotation (best three words) Compared with the Original Manual Annotations

In Table 2, all 9 kinds of images, namely the pagoda, tower, building, animals, sky, grass, water, tree, and fruits have been selected from the database to form annotation for performance assessment. The results are shown in the Table 2 and Figure 11. The precision rates of all 9 kinds of annotated samples are 92.85% for pagoda, 90% for tower, 91.25% for building, 91.66% for animals, 80% for grass and tree, 70% for sky and water, and 87.5% for fruits respectively.

Table 2.Precision for Images of the 9 Kinds of Samples Images

Image Class	Test Images	Correct Images	Precision (%)
Pagoda	350	325	92.85%
Tower	50	45	90%
Building	400	365	91.25%
Animals	600	550	91.66%
Grass	100	80	80%
Tree	100	80	80%
Sky	50	35	70%
Water	50	35	70%
Fruits	600	525	87.5%

Figure 11. The Precision Rates of Images of the 9 Kinds of Samples Images

8. Conclusion

The images that user wants to be annotated can be labeled by using the proposed system. The advantage of automatic image annotation is that queries can be more naturally specified by the user. Generally speaking image tagging is a form of supervised classification of pictorial data.

In this work, the effective and efficient image segmentation algorithm is developed for automatic image annotation system. The goal of segmentation is to simply or change the representation of an image into something that is more meaningful and easier to analyze. This approach is most related to the graph theoretic formulation of grouping. The given uncaptioned image is segmented into regions, which are classified into region types using a variety of features. In the proposed system, the efficient segmentation algorithm is developed by combining the N-cut segmentation algorithm and color component analysis to extract center region.The multiple segmented parts of an image are extracted by the modified N-cut segmentation algorithm. The annotations of each segmented part are performed according to the result of color matching. This system will provide not only the annotation of context of image but also the retrieval of the name of the place.

Automated image annotation system generates a set of keywords that helps to understand the scene represented in the image. The annotation system needs many images that are very similar to the image to annotate to perform a correct annotation. In the proposed system, segmented parts of an image are added in database by classifying their color intensities (Red, Green, Blue, and other). The input image is not necessary to match with the whole database. The images are only matched in their associated color databases. Therefore, the proposed automatic image annotation system can speed up in retrieval process.

References

[1] G. Babbar, Punam Bajaj, AnuChawla, and Monika Gogna,“ A comparative study of image matching algorithms” International Journal of Information, Technology and Knowledge Management, July-December. 2(2), pp.337-339, 2010.

[2] M. Bieliková, E. Kuric, “Automatic Image Annotation Using Global and Local Features”, In Proceeding of Semantic Media Adaptation and Personalization, SMAP 2011, Los Alamitos: IEEE Computer Society, 2011, pp. 33 – 38.

[3] J. Carreira, SminchisescuC.Cpmc, “ Automatic object segmentation using constrained parametric min-cuts”, IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7),2012, pp.1312–1328.

[4] Z. Chen, H. Fu, Z. Chi, and D. Feng, “A Neural Network Model with Adaptive Structure for Image Annotation”, Proceedings, the 11th International Conference on Control, Automation, Robotics and Vision (ICARCV 2010), 7-10 December 2010, Singapore, pp. 1865-1870.

[5] V. K. Dehariya, S. K. Shrivastava, R. C. Jain, “Clustering of Image Data Set Using K-Means and Fuzzy K-Means Algorithms”, International conference on CICN, 2010, pp. 386-391.

[6] S. Fei, Fangfang Yang, Jiajun Wang, “Supervised Semantic Image Annotation Using Region Relevance”, International Conference on Medical Physics and Biomedical Engineering 2012 , Volume 33, 2012, pp.912–918.

[7] A. Fraser, I. Jazayeri, and S.Cronk.2010, “A feature-based matching strategy for automated 3D model reconstruction in multi-Image Close-Range Photogrammetry”, ASPRS 2010 Annual Conference, April 26-30, 2010.

[8] A. Golnaz, Murat Birinci, Fernando Diaz-de-Maria, MoncefGabbouj, Edward J.Delp, “A region-dependent image matching method for image and video annotation”, 2011 , In Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop , 13-15 June 2011,pp. 121 - 126 .

[9] S. Gould, Fulton R, Koller D (2009a), “ Decomposing a scene into geometric and semantically consistent regions”, In: IEEE International Conference on Computer Vision,2009.

[10] Y. Boykov, and Jolly, M.-P. ,“ Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images”, In Proceedings of the International Conference on Computer Vision (ICCV), volume I, pp.105–112, 2001.

[11] A.Hamid, Mansour Jamzad, “Large-Scal Image Annotation using Prototype-based Models”,7th International Symposium on Image and Signal Processing and Analysis(ISPA2011)September4-6, 2011.