A New Method for Automatic Image Annotation using Region Features and WordNet Semantic Similarity

Gongwen XU[a,*], Zhijun ZHANG[b], Xiaomei LI[c], Lina XU[b]

[a]*Business School,* Shandong Jianzhu University, Jinan, Shandong, 250101, China
[b]*School of Computer Science and Technology,* Shandong Jianzhu University, Jinan, Shandong, 250101, China
[c]*Cancer Center of the Second Hospital*, Shandong University, Jinan, , Shandong, 250013, China

***Abstract* — Image retrieval technology has always been an active research field due to the ever-increasing number of pictures taken worldwide. Image annotation technology can manage and process the large number of images effectively. With image annotation technology, users can retrieve useful information from the image database. Recently automatic image annotation has emerged as an important research direction. The image can be automatically annotated by a method based on region features, but the retrieval precision is not high. In this paper, a new automatic image annotation method is proposed based on region features and WordNet semantic similarity. Firstly, the image is labeled according to the region similarity, and then the WordNet semantic similarity is used to improve the annotation results. Our experimental results show that the method proposed in this paper can annotate the images effectively and accurately.**

Keywords- automatic image annotation; region features; WordNet; semantic similarity

I. INTRODUCTION

With the popularization of digital camera and other digital facilities, the number of images on the network has been increasing in geometric growth mode. And image retrieval technique has become an active research field. Image retrieval methods can be classified into two types based on different retrieval methods. One is text-based image retrieval technique[1], the other is content-based image retrieval technique[2]. In the text-based image retrieval technique, the images were annotated manually, and then the image retrieval was transformed to the related key words' query and matching. The advantages of the text-based retrieval are convenient and high-speed. The only thing for the users to do is inputting the keywords to inquire and get related results. But the text-based image retrieval technique is time-consuming. This method need people to annotate the image manually first. The content-based image retrieval technique is designed to search the similar images in the database depending on the visual features. Although content-based image retrieval technique has achieved many research results, but in this method the images are annotated using the images' low-level features. The inquiring keywords are visual features and not the natural languages, so there contains semantic gap between the images and the annotations. Many researchers begin to combine the semantic information to improve the effect of content-based image retrieval technique. The semantic information usually is the textual keyword which describes the image semantic property. Because annotation semantic information manually is quite time-consuming and energy-consuming, the automatic image annotation technique comes into being [3].

There are two main methods about image annotation. The first one is manual annotation based on supervised learning methods[4], which annotates the images using the probability information. These supervised learning methods can ensure the high accuracy but they also need a lot of time and manual labor. The second one is the unsupervised automatic annotation method based on images' low-level features[5]. The automatic image annotation means that the computer can label the images automatically based on the images' content. Usually the automatic image annotation technology is used to organize the images in image retrieval system and search the images related to the users' inquiring keywords.

The automatic image annotation can be divided into two classes. The first one is based on classification, and the other one is based on retrieval. The classification-based automatic image annotation method regards the image annotation problem as the classification problem[6], to train classifier through manual annotation training set and then use classifier to annotate the images. Jeon etc. put forward a probability-based learning model, which annotates the images according to similarity between the keywords and the images[7]. In the retrieval-based automatic image annotation method, the unlabeled image is taken as the querying image, and then the related images which are labeled in advanced in the annotation image database are selected. The unlabeled image will be annotated with the keywords of the selected images. Although in this method the annotation model doesn't need to be created beforehand, a large amount of annotated images are needed to build the annotation database.

In the automatic image annotation method, the images are annotated through building the relationship between the image segmentation area and semantic information. Firstly, the visual features of the image areas are extracted, and then the relevance between the areas and semantic information is calculated according to the annotated images. At last, the unlabeled images are annotated based on the above prior

DOI 10.5013/ IJSSST.a.17.27.38

ISSN: 1473-804x online, 1473-8031 print

knowledge. As that automatic image annotation method only considers visual features, the annotation accuracy is not high. In this paper, an improved automatic image annotation method which combines the region features and semantic similarity is proposed. At first the images are annotated according to region similarity, and then the WordNet semantic similarity is considered to improve the annotation effect.

II. IMAGE ANNOTATION

There are two methods to annotate the semantic information to the images. The first one is called weak annotation, which annotate the keywords to the whole image, not the image area. The second one is the area level annotation, which annotate the keywords to the area in the image. In this paper the image annotation refers to the second one.

A. Image Segmentation

In the automatic image annotation based on region, the image is segmented into multiple homogenous regions using segmentation technology[8,9]. Each area correspond one subject, and each homogenous region has the simple semantic so the semantic content can be described accurately. Many segmentation algorithms have been carried out, in which the N-cut algorithm[10] and JSEG algorithm[11] are representative.

In the image segmentation methods based on graph theory, an image is considered as undirected weighted graph, $G=\{V,E,W\}$. In the formula, *V* represents the nodes set, in which the image pixel is regarded as the node. *E* represents the connection between the nodes. W_{ij} represents the weight between the nodes. The weight can be calculated by the distance between the pixels, brightness or other information. For example, one image can be segmented into two parts, A and B: $A\bigcup B=V, A\bigcap B=\varnothing$.The similarity between the two subsets can be calculated by the following formula.

$$cut(A,B)=\sum_{i\in A,\,j\in B} W(i,j) \qquad (1)$$

Shi and Malik have proposed the Normalized-cut to describe the divergence between two classed and an N-cut metric is achieved to measure the divergence[12].

$$Ncut(A,B)=\frac{cut(A,B)}{assoc(A,V)}+\frac{cut(B,A)}{assoc(B,V)} \qquad (2)$$

In the above formula, *assoc(A,V)* represents the sum of weight between node A and all other nodes. The best segmentation method can be gained to minimize the object function as formula 3.

$$Min\ \ Ncut(A,B) \qquad (3)$$

B. Image Clustering

After the image segmentation, the homogenous areas can be clustered under some rules. The cluster algorithm is often used to classify the low-level features before the image annotation. The most commonly used clustering algorithms are K-means Algorithm[13], Expectation Maximization (EM) Algorithm[14] and Discrete Distribution (D2) Clustering Algorithm[15]. Because the K-means Algorithm can segment the image based on the image color features, it is widely used in the image low-level features cluster.

The K-means Cluster Algorithm process is shown below.

1) K documents are selected from N documents as the centroid;

2) The remainder documents are measured the distance to each centroid, and then they are classified to the nearest centroid.

3) After the cluster, the new centroid of each class is recalculated.

4) Repeating the 2) ~ 3) steps until the new centroid is equal to the previous one, or the distance between them is smaller than a designated threshold. The algorithm stops.

C. Features Extraction

The main task of features extraction is to extract the feature information which can represent the image visual content. In this paper, the image color feature and texture feature are extracted and the values of them are calculated via the certain algorithm.

The color feature is easily extracted as there is large amount of color information in the image, so the color feature is commonly considered as the visual content feature in the image retrieval and annotation[16]. The color information can be represented in different space, and the most common space is RGB color space, whose color space comprises 3 base colors, red, green and blue. In this color space, other colors can be represented by these three colors lineally. Besides RGB color space, HSV[17] is also the common color space to describe image color features. Because HSV color space is visual perceived and accords with the human sense, HSV color space is used in this paper. HSV color space consists of hue, saturation, value and can be transformed from RGB color space under the following formula.

$$H=\begin{cases}\arccos\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & B\le G\\[2ex] 2\pi-\arccos\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & B>G\end{cases} \qquad (4)$$

$$S=\frac{\max(R,G,B)-\min(R,G,B)}{\max(R,G,B)} \qquad (5)$$

$$V=\begin{cases}0,V\in[0,0.2]\\ 1,V\in[0.2,0.7]\\ 2,V\in[0.7,1]\end{cases} \qquad (6)$$

The ranges of the values are shown below, $R,G,B \in [0,1,\cdots,255]$, $H \in [0,1,\cdots 360]$, $S,V \in [0,1]$.

The HSV color space values need to be quantified as its dimensions are very high. The quantification methods refer to[18]. After quantification, the suitable image color features are ready for the experiment.

The image texture is decided by the physical properties such as the surface roughness of the object. The different texture can be easily achieved so the visual information can be extracted from the texture. The texture information is a very important visual content features and the texture features are extracted from the image using gray level co-occurrence matrix method. The texture feature information based on gray level co-occurrence matrix has 14 parameters, among which 4 types of parameters that have strong describing ability are selected. They are angular second order moment, inertia moment, entropy, locally stationary parameters.

Now the segmented areas are clustered under K-means method. The representative areas in each class are selected to be annotated manually. The average of the color feature value and texture feature value of the selected areas are used to quantize the annotation keywords.

$$K_i = \frac{1}{l}\sum_{j=1}^{l} X_{ij} \tag{7}$$

In the formula, K_i is a multidimensional vector which is the *i*th annotation keyword numerical value. *l* is the number of areas that are annotated with K_i. X_{ij} represents the eigenvalues vector of the *j*th area with the annotation K_i.

The annotation keyword K_i is regard as the seed set. With regard to the testing images, firstly the value vector *T* of the color feature and texture feature of the segmented areas is calculated. Then the Euclidean Distance between the area feature vector *T* and keyword vector K_i. If the Euclidean Distance is lower than a certain threshold, the keyword K_i can annotate the image area.

III. ANNOTATION EFFECT IMPROVEMENT BASED ON WUP

A. WUP Method

WordNet[19] is one of the famous database which can be used to calculate the semantic similarity between two concepts (or word sense) with six methods to count similarity and three methods to count concepts correlation. All of these methods are based on the WordNet lexical database. In this paper, WUP method was used to measure the semantic similarity of the annotation words. The WUP[20] is a method used for measuring similarity based on path structure which was proposed by Wu and Palmer. It takes into account the path association information among the concept nodes, parent nodes and root nodes. The similarity between nodes S_1 and S_2 are computed as following.

$$sim_{wup}(s_1, s_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3} \tag{8}$$

In the above formula, S_3 is the most upper parent node of S_1 and S_2 , N_1 is the number of nodes in the path from S_1 to S_3, N_2 is the number of nodes in the path from S_2 to S_3, N_3 is the number of nodes in the path from S_3 to root path .

B. Improving the Annotation Precision

After numbering the keywords, the semantic similarity between keywords are calculated using WUP similarity. A symmetrical correlation matrix *T* is achieved through the calculation of correlation between keywords.

$$T = \begin{bmatrix} w_{11}, w_{12} \ldots . w_{1n} \\ w_{21}, w_{22} \ldots . w_{2n} \\ : \qquad : \\ w_{n1}, w_{n2} \ldots . w_{nn} \end{bmatrix} \tag{9}$$

In this matrix, *n* is the number of annotation words in the dataset. W_{ij} is the correlation degree between the *i*th and the *j*th annotation words. A vector *X* can be achieved for a certain keyword *i* by descending correlation degree order, X=[w_{il},w_{im},……] (*l,m* is the sequence number for the annotation words) . For an image, *y* correlation annotation keywords can be gained via the annotation method based on region. By computing the semantic similarity between the annotation words database and this image's annotation keywords, the *y* correlations are produced. Then, if there are repetitions of annotation word number in the *y* correlation vector where the correlation is larger than a designed threshold, the annotation words with repeating number will annotate this image complementarily.

IV. THE EXPERIMENT RESULT ANALYSIS

In this section, the experimental dataset will be introduced and the experiment will be analyzed to show the validity and the effect of the method proposed in this paper.

A. Dataset

To test and verify the method in this paper, Corel5K image database is used as the training and testing dataset. There are 5000 images in this database and are classified 50 types, each of which includes 100 images. The images contain city, mountain, sky, sun, sea, building, horses and so on. The experimental simulation platform is MATLAB2012a.

Each type of image set has 100 images, so 70 images in the class were selected as the training set. Each image was divided into 1-5 parts through N-Cut image segmentation algorithm, and then k-means algorithm was used to cluster the segmented region. The representative area of each category was selected to be labeled manually. Then every image was labeled with 1-4 words. After training with 70 images, the remaining 30 images were used as the testing set.

B. Metrics

After the experiment, the recall ratio, precision ratio and F1 value were used to measure the performance of experimental results. For a given annotation keyword *K*, the number of image which contains *K* in testing set is K_t. The number of images which contain keywords *K* in the retrieval result after using the annotation model is K_s, and the number of the correct result is K_r.

The following formula is the Recall ratio, which means the fraction of images that are relevant to the query that are successfully retrieved.

$$Recall = \frac{K_r}{K_t} \tag{10}$$

The following formula is the Precision ratio, which means the fraction of retrieved images that are relevant to the query.

$$Precision = \frac{K_r}{K_s} \tag{11}$$

F1 value is the harmonic mean of Recall and Precision, and it is a comprehensive index.

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{12}$$

C. Results Analysis

The threshold value of the correlation degree between the keyword in database and the labeled words of the image need to be determined when using WUP to improve the annotation accuracy. The relationship between the threshold value and the Precision ratio, Recall ratio and F1 value is drawn as figure 1.

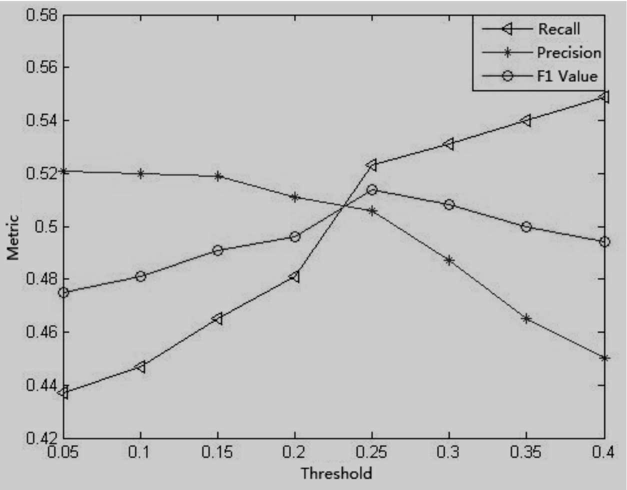


Figure 1. The Relationship Between the Threshold and The Metrics

Through the figure 1, it can be seen that the greater of the threshold value, the greater of the recall ratio, and the smaller of precision ratio. When the threshold value is set to 0.25, the experimental effect is best, where precision ratio is 0.523 and recall is 0.506 and the F1 value is 0.514.

When annotating the images, two methods were used for comparison. The first method only used region similarity and the second one used the combination of the first method and the annotation words correlation. When calculating the correlation of the labeling words, the threshold is set to 0.25.

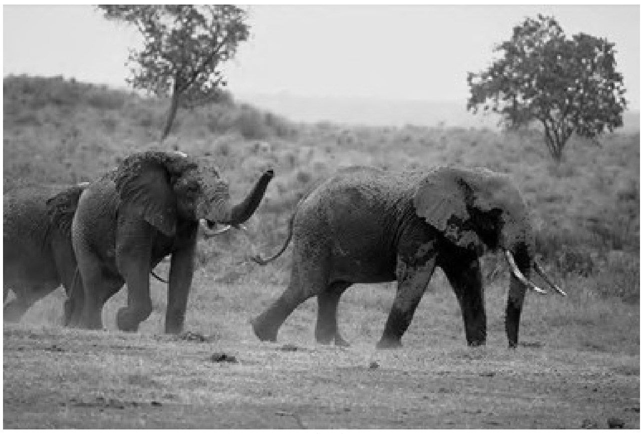
The experimental results are shown as table 1.

TABLE 1. EXPERIMENTAL COMPARISON

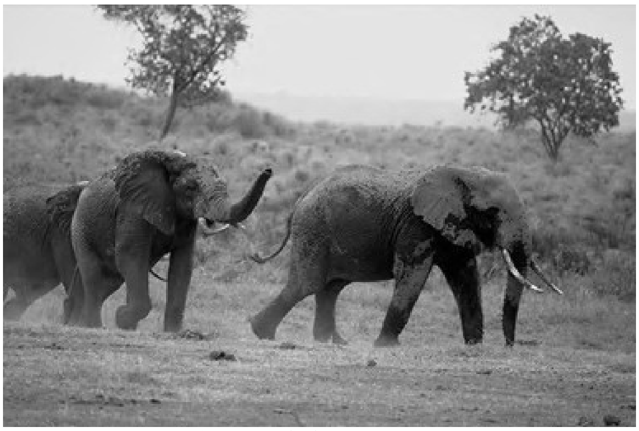
	Precision Ratio	Recall Ratio	F1 Value
Using Region Similarity only	0.461	0.484	0.472
Combining Method	0.523	0.506	0.514

In the table 1, it can be seen that, when only using the region similarity to annotate the image, the precision ratio is 0.461, the recall ratio is 0.484, and F1 value is 0.472. When combining the keywords correlation and the region similarity method, the precision ratio is 0.523, increased by 13.45%; the recall ration is 0.506, increased by 4.55%; the F1 value is 0.514, increased by 8.90%. The experimental results showed that it is effective to label the image when combined with the region similarity and the keywords correlation method.

When using the combination method to annotate one image in the dataset, the experimental results are shown in figure 2.



(1) elephant; tree.



(2) elephant; tree; grass.

Figure 2. The Annotation Results.

When using region similarity method, the image was labeled with words ‘elephant’ and ‘tree’. When using the combining method, the image was labeled with another word, ‘grass’, which is added by calculating the semantic similarity between the words in dataset and ‘elephant’, ‘tree’.

V. CONCLUSIONS

In this paper, a method using annotation correlation was proposed to improve the image annotation effect. Firstly the N-Cut method was used to segment the images into some homogenous regions. Then K-Means method was used to cluster the segmented areas. The unlabeled images were annotated according to the semantic similarity between the regions of the labeled images and the annotation words. At last the WUP method was used to calculate the semantic similarity among the annotation words to improve the image annotation effect. The experimental showed that this method can effectively annotate the images automatically.

ACKNOWLEDGMENTS

The authors would like to thank the Project of Shandong Province Higher Educational Science and Technology Programs (J14LN59, J16LN18), the Development Projects of Science and Technology of Shandong Province (2014GGX101011, 2015GGX101018, 2016GGX101035), the Second Hospital of Shandong University Seed Fund (S2015010004) for financial support.

REFERENCES

[1] G. W. Xu, X. M. Li, J. Lei, “An image retrieval and semantic mapping method based on region of interest”. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, vol. 6, pp. 88-95, 2015.

[2] T. Kato, “Database architecture for content-based image retrieval”, In Proc. Of SPIE Int. Conf.on Image Storage and Retrieval System, San Jose, CA, USA, pp. 112-123, 1992.

[3] G. W. Xu, Z. J. Zhang, W. J. Qi, M. H. Liao, L. N. Xu, H. L. Zhao, “Image Automatic Annotation Based on the Similarity of Regions”, Journal of Computational Information Systems vol.10, pp. 9397–9404, 2014.

[4] G. Carneiro, A.B. Chan, P.J. Moreno, N. Vasconcelos, “Supervised Learning of Semantic Classes for Image Annotation and Retrieval” , IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 394-410, 2007.

[5] G. W. Xu, L. N. Xu, X. M. Li, W. J. Qi, “The Image Retrieval Method Based on the Homolographic Block Color Histogram”, Chemical Engineering Transactions. Vol. 51,pp. 403-408, 2016.

[6] H.J. Escalante, C.A. Hernández, J.A. Gonzalez, A. López-López, etc., “The segmented and annotated IAPR TC-12 benchmark” , Computer Vision and Image Understanding, vol. 114, pp. 419-428, 2010.

[7] J. Jeon, V. Lavrenko, R. Manmatha, “Automatic Image Annotation and Retrieval using Cross-Media Relevance Models”. In: 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.119-126, 2003.

[8] B. C. Ko, H.Byun, “Frip: a region-based image retrieval tool using automatic image segmentation and stepwise Boolean and matching”, IEEE Trans. On multimedia, vol. 7, pp. 105-113, 2005.

[9] Y. Liu, D. Zhang, G. Lu, “Region-based image retrieval with high-level semantics using decision tree learning”, Pattern Recognition, vol. 8, pp.2554-2570, 2008.

[10] S. Jianbo, M.Jitendra, “Normalized cuts and image segmentation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.8 , pp.888-905, 2000.

[11] T. Liu, K. Muramatsu, M. Daigo, et al. “Region Segmentation of Multi-spectral Remote Sensing Images Based on JSEG Algorithm”, Doshisha University World Wide Business Review, vol.10, pp.26-34, 2009.

[12] J. Shi, J. Malik, “Normalized cuts and image sementation”, IEEE Computer Vision & Pattern Recognition, vol. 8, pp. 888-905,1997.

[13] N. A. M. Isa, S. Salamah, U. K. Ngah, “Adaptive fuzzy moving K-means clustering algorithm for image segmentation”, IEEE Transactions on Consumer Electronics, vol. 55, pp. 2145-2153, 2009.

[14] M. Fakheri, T. Sedghi. Color image retrieval technique based on EM segmentation algorithm. Telecommunications (IST), 2010 5th International Symposium on Telecommunications, pp. 793-795, 2010.

[15] Y. Zhang, J. Z. Wang, J. Li, “Parallel Massive Clustering of Discrete Distributions”, ACM Transactions on Multimedia Computing Communications & Applications, vol. 11, pp. 1-24, 2015.

[16] G. W. XU, L. N. Xu, X. M. LI, X. B. TANG, X. Y. LI, C. X. XU, “An Image Retrieval Method based on Visual Dictionary and Saliency Region”, Int. J. Signal Process. Image Process. Pattern Recogn. Vol.9, pp.263-274, 2016.

[17] J. Wang, B. Kong, J. Q. Li, “Color-Based Image Retrieval”. Computer Systems & Applications, vol. 7, pp. 160-164, 2011.

[18] K. Mikolajczyk, S. Cordelia, “A performance evaluation of local descriptors”. IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 10, pp. 1615-1630, 2005.

[19] T. Deselaers, V. Ferrari, “Visual and Semantic Similarity in ImageNet”. In: CVPR 2011. vol. 32, pp. 1777-1784, 2011.

[20] Z. Wu, M. Palmer, “Verb Semantics and Lexical Selection”. In: ACL 1994 Proceedings of the 32nd annual meeting on Association for Computational Linguistics, pp.133-138, 1994.