Image Annotation Refinement Using Dynamic Weighted Voting Based on Mutual Information

Haiyu Song
Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education,
Jilin University, Changchun, China;
Institute of Computer Graphics and Image Processing, Dalian Nationalities University, Dalian, China
Email: hysong07@mails.jlu.edu.cn

Xiongfei Li*
Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education,
Jilin University, Changchun, China;
Email: jlulxf@gmail.com

Pengjie Wang
State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou, China
Institute of Computer Graphics and Image Processing, Dalian Nationalities University, Dalian, China
Email: wpj@dlnu.edu.cn

***Abstract*—Automatic image annotation is a promising solution to narrow the semantic gap between low-level content and high-level semantic concept, which has been an active research area in the fields of image retrieval, pattern recognition, and machine learning. However, even the most dedicated annotation algorithms are often unsatisfactory. Image annotation refinement has attracted much more attention recently. In this paper, a novel refinement algorithm using dynamic voting based on mutual information is proposed. Unlike the traditional refinement algorithm, the proposed algorithm adopts dynamic weighted voting to measure the dependence between the candidate annotations, which not only permits that the annotations with higher probabilities deny the annotations with lower probabilities, but also permits that the annotations with lower probabilities deny the annotations with higher probabilities. The proposed refinement algorithm adopts progressive method instead of iterative, which can significantly decrease the time cost of refining annotations. In order to further improve efficiency without sacrificing precision, we propose the block-based normalized cut algorithm to segment image. Experiments conducted on standard Washington Ground Truth Image Database demonstrate the effectiveness and efficiency of our proposed approach for refining image annotations.**

***Index Terms*—image annotation refinement, image retrieval, mutual information, normalized cut, relevance model**

I. INTRODUCTION

Image retrieval has been an extremely active area of research in the fields of computer vision and pattern recognition for almost 20 years [1]. There are many representative content based image retrieval (CBIR) systems such as QBIC, Columbia VisualSEEK, MIT Photobook, whose "content" is some kind of objective statistic character of images that couldn't be understood by human beings directly. Moreover, for most users, articulating a content-based query using these low-level features can be non-intuitive and difficult. Many users prefer using keyword to retrieve image such as Google image. As a result, some query-by-keyword systems extract textual words from webpage associated with image, and the retrieval task will be simplified into a typical textual retrieval. But a large number of images created by digital camera haven't any textual information available. Moreover, sometimes the associated textural words haven't any semantic information. Some researches advise manually annotating images. This approach encounters the problem of inconsistency and subjectivity among different annotators, and the process is time-consuming as well. Especially, with the explosive increase of images available, manually annotating all images is impractical [2].

Some researches have strived to apply object recognition technology to improve the semantic level of image unassociated with any textual words. Decades of research have shown that designing a generic computer algorithm that can learn concepts from images to linguistic terms is highly difficult. Much success has been achieved in recognizing a relatively small set of objects or concepts within specific domains [3], but there is no generic algorithm for object recognition. As a result, the approaches of object recognition can't be used as generic algorithm of image understanding.

Automatically annotating images with keywords is a solution to the aforementioned problems, and many researches have devoted to develop automatic image

* Xiongfei Li is the corresponding author.

Project number: No.60675008

doi:10.4304/jsw.6.11.2239-2246

annotation system. Automatic annotation of image has been a highly challenging problem for computer scientists and electrical and electronic scientists over past four decades. The effective and efficient annotation approach can lead to breakthroughs in a wide range of applications including Web image retrieval, multimedia management and so on.

II. RELATED WORKS

A. Annotation Models

In recent years, many annotation algorithms have been proposed, which can be classified into four categories, i.e. classification method, co-occurrence method, graph-based method, and probabilistic modeling method.

The classification method always is based on machine learning technology. Firstly, the user manually assigns each training image to one of K pre-defined labels. Then, train the K classifiers by machine learning algorithm through training image dataset, which is called model generation. Finally, the classifier is able to classify the unlabeled image into the learned class label. The label name is the annotation keyword of corresponding images because every keyword corresponds to one classifier. After model generation, the mission of this annotation is to propagate the learned labels or label names to unlabeled images. The algorithms of classification can be classified into two models: generative model, discriminative model. The generative model is a classifier based on probabilistic density estimation such as Gaussian density and Bayesian network, while the discriminative model is a classifier based on feature space partitioning such as ANN and SVM. The representative works using generative model include MLP neural network [4][5][6], SVM method [7][8], K-NN[9][10], Bayes method[11][12][13], 2-D Hidden Markov Model[14]. All the classification methods for automatic image annotation are specified in a small number of categories and are unscalable for huge amount of images with infinite semantics keywords. They can't be applied into the problem of general images annotation.

Another category is co-occurrence model. Mori proposed a method for annotating image grids using co-occurrences [15], which annotates image using the co-occurrence of words and image regions (blocks). Based on co-occurrence model, Duygulu [16] proposed a novel translation model using machine translation method. In translation model, Duygulu proposed use a vocabulary of blobs to describe image so that the annotation can be considered as the task of translating a vocabulary of blobs into a vocabulary of annotation keywords. The co-occurrence model and translation model tend to assign the frequent words to all blobs.

The graph-based model proposed by Pan [17] treats the image, annotation, and regions as three types of nodes in graph. The advantage of the model is domain independent, but it is a NP-problem.

The CMRM [18] is another probabilistic-based method, which is inspired by relevance language model. The CMRM consider the visual feature such as color, texture, and shape information as another language just like textual language. It uses probabilistic method to predict the probability of generating a word given the blobs in an image. In fact, the isolated pixels or even blocks (regions) in an image are often hard to interpret. The CMRM don't assume there is a one-to-one correspondence between blobs and words, but only assume that a set of keywords is related to the set of blobs, which can capture the latent semantic information. Compared with classification and co-occurrence models, the probabilistic-based model such as CMRM is the best model, but it still can't meet the user's need.

B. Refinement Approaches

Many automatic image annotation algorithms have been used for image retrieval system. However, the accuracy of retrieval is unsatisfactory because none of these algorithms can ideally bridge the semantic-gap between low-level features and high-level semantic meanings. Many researches have strived to invent novel algorithms of automatic annotation to improve the quality of annotation. Automatic annotation may not attain extremely high accuracy with the present state of the computer vision and image processing technology. Despite many new algorithms were proposed, the result of automatic annotation is still unsatisfactory. More and more researches tend to refine the current annotation results instead of inventing new algorithms because many of the annotation keywords are inappropriate for image content.

Jin et al. [19] have done pioneering work on annotation refinement based on knowledge technology, which remove the irrelevant annotations according to the relationship between annotations based on knowledge of WordNet[20]. Subsequently, Jin et al.[21] proposed graph-partitioning approach to refine the annotations. Wang et al. proposed a novel refinement approach using Random Walk with Restarts [22]. Most of refinement algorithms only consider the relationship between candidate annotations, which means the two refined annotations are same only if their candidate annotations are same. The content-based annotation refinement proposed by Wang [23] considered the query image content, which means the two refined annotations may be different due to their image content. But in fact, aforementioned algorithms are assumed "majority should win" only based on relevant keywords number, no adequately considering their weight. Our previous work based on relevance model has considered their significance while refining annotations, but significance of every annotation is static. In this paper, we propose a new annotation refinement algorithm based on mutual information of candidate annotations. In the new algorithm, the weight of each annotation is dynamic and the number of refined annotations is adaptive.

III. PROPOSED APPROACH

The humans perceive and understand object of vision of image depending on context, priori knowledge, imagination, and specific details instead of individual

segmentation or isolated region. Motivated by image perception and understanding of humans, we propose the new algorithm of image annotation refinement using dynamic weighted voting based on mutual information.

A. Annotation Approach

We adopt improved relevance model to annotate image with specified keywords. Unlike the traditional relevance model, we propose that the weights of each blob and candidate annotation keyword in an image are computed to apply into the relevance model for generating annotation keywords. The weight of Blob is computed similar to the TF*IDF model in traditional textual information retrieval system [24], while the weight of candidate annotation is calculated by the probability of producing the candidate annotation using relevance model [18].

The proposed algorithm can be described as follows:

a) Segment image into some regions.

b) Determine the maximum inscribed rectangle of segmented region.

c) Compute the feature vector of inscribed rectangle in every images.

d) Learn the blob from the feature vector based on machine learning approach.

e) Let the $freq_{i,j}$ be the raw frequency of blob ID I in the image ID J.

f) The normalized frequency $f_{i,j}$ of blob I in the image J is given by $f_{i,j}=\frac{freq_{i,j}}{\max_l freq_{l,j}}$.

g) The weight of blob I in the image J is give by the best known term-weighting schemes $w_{i,j}=f_{i,j}*\log\frac{N}{n_i}$, where N is the size of the training image dataset, and n_i is the number of images in which the blob I appears.

h) Compute the candidate annotation according to the relevance model.

i) Assign the probability of each candidate annotation keyword of vocabulary to the weight of annotation.

j) The top N keywords are the candidate annotations.

B. Refinement Algorithm

In probability theory and classical information theory, the mutual information of two random variables X and Y is quantity that measures the mutual dependence of the two variables, which can be defined as:

$$I(X;Y)=\sum_{y\in Y}\sum_{x\in X}p(x,y)\log(\frac{p(x,y)}{p_1(x)p_2(y)})dxdy \quad (1)$$

where p(x,y) is the joint probability distribution function of X and Y, and $p_1(x)$ and $p_2(y)$ are the marginal probability distribution functions of X and Y respectively. Mutual information is shown in Fig. 1, which can be equivalently expressed as:

$$I(X; Y) = H(X)-H(X|Y) \quad (2)$$

where H(X) is the entropy of the random variable X, and H(X|Y) is the conditional entropy. As mutual information is symmetric, it can be written as:

$$I(X; Y) = H(Y)-H(Y|X) \quad (3)$$

The words with high relevance in semantic must have high dependency. We can use mutual information to measure the relevance of annotations in semantics. Namely, we can use mutual information to refine the annotations produced by automatic image annotation algorithms.

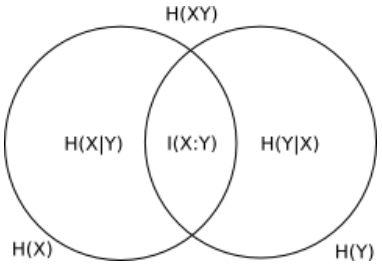


Figure 1. Mutual Information of X and Y.

The refinement algorithm based on WordNet proposed by Jin et al. [19] can be described by Markov process as follow:

$$P(w_i|I_q)=\alpha\sum_{j=1}^{N}P(w_i \mid w_j) \quad (4)$$

Where $p(w_i|w_j)$ is calculated by the semantic similarity of w_i and w_j in WordNet. α is a normalization constant[23]. The annotation keywords with smaller probability value will be removed according to the above formula. The refinement algorithm has three shortcomings. First, the WordNet is independent of specific image dataset. Second, the WordNet can't include all annotation keywords, which means that it can't deal with some annotations. Third, the refinement process is independent of the query image, which means the same candidate annotations will produce the same refined annotations. In order to solve the above problems, Wang [23] proposed a novel algorithm called content-based image annotation refinement, which adapt an iterative process as follows:

$$P^{(t+1)}(w_i \mid I_q)=\sum_{j=1}^{N+1}p(w_i \mid w_j, I_q).p^{(t)}(w_j \mid I_q) \quad (5)$$

Where $P^{(t+1)}(w_i \mid I_q)$ is the probability that the annotation is in the state w_j at time t. $P(w_i \mid I_q)$ is defined as the stationary probability of state w_i in the Markov chain, which is given the following formula [23]:

$$P(w_i|I_q)=\lim_{t->\infty}p^{(t)}(w_i \mid I_q) \quad (6)$$

The Wang's algorithm can achieve better performance, but it still has two difficulties which lead to high time

complexity. The first shortcoming is that the evaluation of any candidate annotation word involves all the other candidate annotations including irrelevant or noisy words, which will result in low performance. The second shortcoming is iterative refinement process with low convergence speed.

We have proposed refinement algorithm based on Boolean Model in our prior work [25]. The refinement algorithm can be described as follows:

a) Compute the probability of each candidate keyword of vocabulary according to relevance model.

b) Construct candidate keywords list $L_{candidate}=\{P_1, \ldots, P_m\}$.

c) Remove the largest element P_i from Lcandidate , and create the keywords list $L_{keyword}=\{P_i\}$.

d) Pick out the current maximum element P_x from $L_{candidate}$, and compute the comprehensive probability $P_{x\text{-}word\text{-}occurrence}$ of its co-occurrence with all the element of $L_{keyword}$.

e) If the comprehensive co-occurrence probability $P_{x\text{-}word\text{-}occurrence}$ is larger than threshold value $T_{co\text{-}occur}$ set by user, store P_x into $L_{keyword}$.

f) Repeat the step d-f until the list $L_{candidate}$ is empty or the size of $L_{keyword}$ is larger than the user-specific threshold.

g) The elements of keywords list $L_{keyword}$ are refined annotation with textual keywords.

In our Boolean model based on refinement approach, the convergence speed of refinement is high, but it will leads to low precision for some images. The reason is that the refinement algorithm is assumed that the candidate annotation with the maximum probability is correct. If the condition is false, the refined process and all the results are unreliable. In the refinement process, every judgment only considers new coming candidate annotation keyword (Px) while it is absolutely impossible to remove any keywords in $L_{keyword}$. The fixed or state sequence of removing noisy keywords is not reasonable. Intuitively, some keywords confirmed in the former steps should permit to remove in the later step in the context of more corresponding information. In fact, because the probabilities of all candidate annotations are automatically generated by the annotation algorithm without any manual participation, the candidate annotation keyword with high probability only means the possibility of being refined or final annotation is high. Any algorithm completely depending on a static probability is unreliable.

In this paper, we propose a dynamic approach in contrast to the prior works, which is not completely depended on the maximum fixed probability originally produced by annotation algorithm. Every step of refining annotation not only considers the new coming candidate annotation but also the refined annotations, which means the former can deny the latter, and vice versa. In priori algorithms, the sequence has great significance in refining next candidate annotation. In our new algorithm, it is the combination of weighted correspondence that determines whether the candidate annotation is refined/final annotation or not. We use mutual information to measure the dependence between two keywords, and probabilities of the two candidate annotation to present the weight of the correspondence significance. Moreover, our new proposed algorithm will produce adaptive numbers of refined annotation keywords. In the existing annotation refinement approaches, the relevance of annotation keywords is measured by Boolean model and pure co-occurrence approach. In our new approach, the measure is based on mutual information between candidate annotations. The new algorithm can be described as follows:

a) Compute the probability of each candidate keyword of vocabulary according to relevance model.

b) Construct candidate keywords list $L_{candidate}=\{w_1, \ldots, w_m\}$,where $P_1>P_2>P_3>\ldots.>P_m$, and P_i is the probability of w_i.

c) Remove the first element w_1 from $L_{candidate}$, and create the original keywords list $L_{keyword}=\{w_1\}$.

d) Pick out the front (largest) element w_{m+1} from current $L_{candidate}$, where the m is the member number of list $L_{keyword}$.

e) Compute the comprehensive significance $S=\sum_{i=1}^{m} I(w_i; w_{m+1}) \times K_i$, where the $I(w_i, w_{m+1})$ is the mutual information between w_i and w_{m+1}, and K_i represent the weight ($K_i = P_i$).

f) If S is larger than threshold T specified by user, append W_{m+1} to $L_{keyword}$. Goto step h.

g) If $S - I(w_t; w_{m+1}) > \sum_{i=1}^{m+1} (w_i, w_t)$ *(where t<m),* append W_{m+1} *to* $L_{keyword}$ and remove the w_t from $L_{keyword}$.

h) Repeat the step d-h until the list $L_{candidate}$ is empty or the size of $L_{keyword}$ is larger than the user-specific threshold.

i) The elements of keywords list $L_{keyword}$ are refined or final annotations with textual keywords.

IV. FEATURE EXTRACTION

A. Image Segmentation

The global visual feature can't narrow the semantic gap between the semantic concept and image visual content. As a result, the global feature is inappropriate for image annotation. The local visual feature is too sensitive to noise, so it isn't suitable for image annotation. The visual feature of region is a better choice for image annotation. Furthermore, the relevance model of image annotation requires the feature vector based on image block or region. Image segmentation is a important but challenging step to extract region information in the field of image analysis and understanding.

A lot of algorithms and technologies have been proposed and developed for image segmentation, among which the normalized cut (N-cut) [26] has been a promising method and active research area due to better

quality. But the N-cut can't process medium and large image dataset online due to incredible memory and time complexity. Moreover, the N-cut algorithm can't utilize texture information since its basic unit of clustering algorithm is pixel. To overcome the drawback and improve the description power, we have proposed a block-based N-cut algorithm for image segmentation in prior work [27], whose basic unit of clustering is image block instead of individual pixel. We adopt the proposed block-based N-cut algorithm to segment image, which can be described as follows:

a) Segment image into several 7*7 pixels blocks;

b) Construct color feature vector for image block with 72-dimension gray feature vector;

c) Construct texture feature vector of image block, including average, variance, maximum, minimum, and weighted centroid pixel. Weighted centroid pixel is calculated by convolution product of pixels value distribution and normal distribution;

d) Construct graph G, whose vertex is corresponding to image block;

e) Compute weighted adjacency matrix W, whose edge weight w_{ij} is similarity and distance metric of node i and j corresponding to block i and j respectively;

f) Compute the unnormalized Laplacian L;

g) Compute the first m eigenvectors $v_1, \ldots, v_m$ of the generalized eigenproblem $Lv = \lambda Dv$;

h) Let $V \in R^{n \times m}$ be the matrix containing the vectors $v_1, \ldots, v_m$ as columns;

i) For i = 1, . . . , n, let $y_i \in R^m$ be the vector corresponding to the i-th row of V;

j) Select the optimal parameters k and m;

k) Cluster the points $(yi)_{i=1,\ldots,n}$ in R^m with the k-means algorithm into clusters $C_1, \ldots, C_k$.

B. Feature Vector Extraction

Image visual feature can be color, texture, and shape information. Color is the most used visual feature for image retrieval due to its efficiency. Shape is one of the most important features for describing the object, which means that it suitable for object recognition. Texture is an important element of images for surface, object identification, and region distinctions, which means the texture has more discriminative ability for region feature. The shape information aims to determine and identify different region. We don't use the edge information as shape signature. We have compared many texture feature including co-occurrence matrix, texture structure, frequency spectral method, and gray histogram by experiments. The results show that the gray histogram is a better choice due to its better performance and suitable to large-scale data-processing need. We use color histogram as color feature. The feature vector of region is a combination of texture and color feature of corresponding region. The blobs are generated by clustering algorithm such as k-means.

V. Experimental Results

A. Experimental Design

We use the Ground Truth Image Database provided by University of Washington [28] as training and test image dataset, which have been manually annotated with an average of 5 keywords per image. The total collection consists of 17 categories, each containing about 58 broadly similar images. We divided the dataset into three parts--with 60% training set image, 20% evaluation set images and 20% testing set images. The images of training dataset are annotated with keywords manually. The system automatically annotated the other 40% dataset. The annotation keywords, blobs information, region-based feature vector of all the collection are stored in database.

If a user submits a query image, the annotation module of the system will assign certain annotations to the query image. The process is as follow. Firstly, the system extracts the region-based feature vector from the image. Secondly, translate the feature vectors into blob IDs. Thirdly, generate the annotation for query image using relevance model. Fourthly, if the size of the keywords vocabulary is n, then the query feature vector q is represented as $q=\{w_{1,q}, w_{2,q}, \ldots, w_{n,q}\}$, and the feature vector of vector space model for image j is represented as $I_j=\{ w_{1,j}, w_{2,j}, \ldots, w_{n,j} \}$. Fifthly, the system retrieves the candidate image with respect to the region feature vectors of the query image. Sixthly, the system filters the candidate retrieval images using annotations of query. Finally, the top N similar retrieval images are considered as final retrieval images.

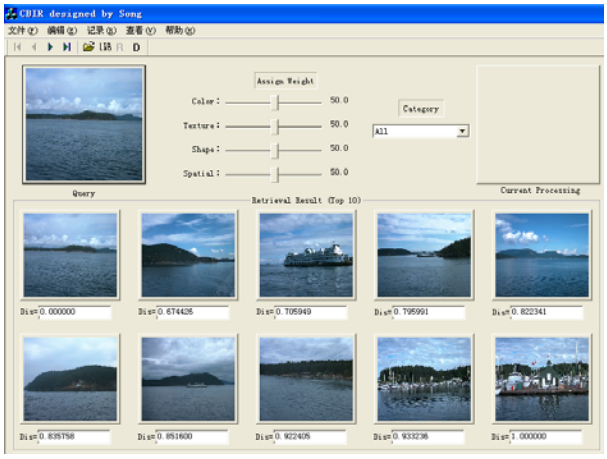
Although all the images have been segmented and features have been extracted from every region when generating blobs in reposition module, the region-based feature vector is only to generate blob and only a few will be used in every query, so the system doesn't store the region-based feature vector. After refining the result images according to blobs similarity, the system segments each of refined result images into regions, and extracts region-based feature vector online. Apply k-means to adaptively generate k clusters with k depending on the image being processed. The collection of pixels in the same region forms a relatively homogeneous region of color and texture. Supposed the z_l is the feature vector of region l, the feature vector of image can be represented as $\{(z_1, f_1), (z_2, f_2), .., (z_k, f_k)\}$, a weight set of vectors, where f_l is the percentage of pixels falling into region l. We use f_l to represent the significance of region l from the view of visual content.

In distance or similarity metrics, the measurement always uses one-to-one mapping formula, such as Euclidean distance, Minkowski, which can only be applied to measure similarity between two vectors with the same size. But for different size of comparing vector, the traditional methods are unsuitable. In this paper, we use Earth Movers Distance (EMD) as similarity formula to measure the similarity of two images described by a weight set of vectors with different size. Let $P = \{(p_1, w_{p1}), \ldots, (p_m, w_{pm})\}$ be the first signature with m regions,

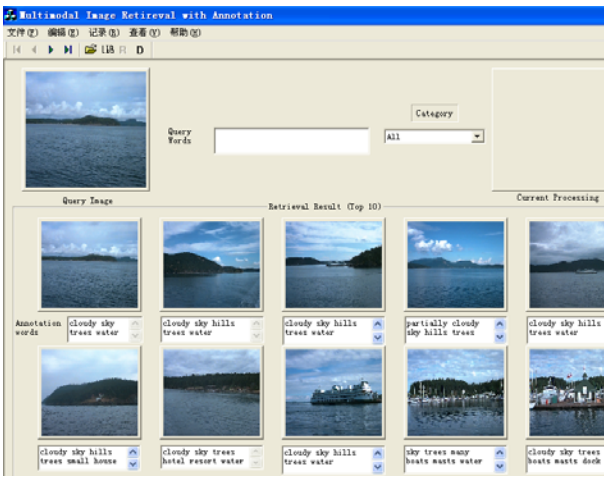
where p_i is the region representative and w_{pi} is the weight of the region; $Q=\{(q_1, w_{q1}), \ldots, (q_n, w_{qn})\}$ be the second signature with n regions; and $D=[d_{ij}]$ be the ground distance matrix where d_{ij} is the ground distance between regions p_i and q_j. The main task of similarity computation is to find a flow $F=[f_{ij}]$, with f_{ij} the flow between p_i and q_j.

B. Performance Evaluation

There is no direct evaluation metrics for automatic image annotation. As the images of Washington dataset have been manually annotated, we can compare the automatic annotation with manual annotation so as to evaluate the annotation performance. We can measure the performance of annotation according to the multimodal image retrieval, which combines annotation keyword-based image retrieval into the content-based image retrieval. The performance of retrieval is evaluated by F, which is defined as 2*MAP*Recall/(MAP+Recall). MAP (Mean average Precision) is the arithmetic mean of average precision, while recall is percentage of all the relevant images in the search database which are retrieved.



(A) Retrieval Results in CBIR using region feature.



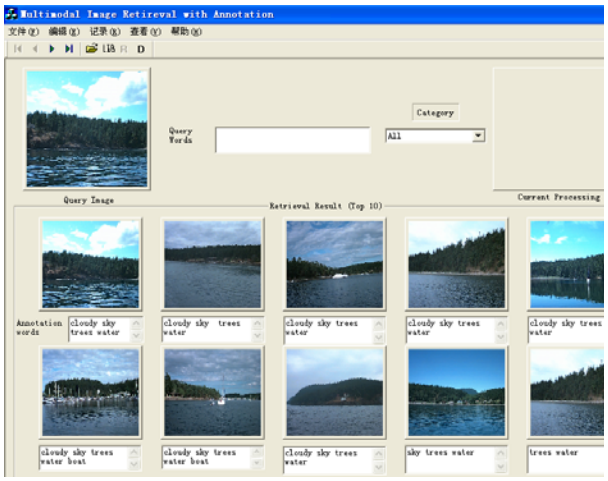
(B) Multimodal Image Retrieval with Refined Annotations.

Figure 2. Retrieval performance comparison between our proposed approach and region-based image retrieval

C. Experimental Results

To evaluate the proposed algorithm, we use the multimodal image retrieval system, which retrieves images based on annotation keywords and region feature. The Fig.2 (A) and Fig.3 (A) are the results of content-based image retrieval system, and the Fig.2 (B) and Fig.3 (B) are the results of multimodal image retrieval system respectively. As shown in Fig.2, the multimodal image retrieval system using our proposed image annotation refinement algorithm outperforms the traditional content-based image retrieval using region feature. When the image feature vector is global feature in CBIR, the multimodal retrieval using proposed image annotation refinement algorithm is far superior to the CBIR.

(A) Retrieval Results in CBIR using global feature



(B) Multimodal Retrieval Image with Refined Annotations.

Figure 3. Retrieval performance comparison between our proposed approach and global feature-based image retrieval

VI. Conclusions and Future Works

In this paper, we propose a novel approach based on mutual information to automatically refine image annotation. In contrast to previous algorithms, our algorithm not only permits that the annotations with higher probabilities deny the annotations with lower probabilities, but also permits that the annotations with lower probabilities deny the annotations with higher probabilities. When computing the dependency between annotations, we introduce weighted combination assigned

by original corresponding probability. Experimental results demonstrate the effectiveness and efficiency of our proposed approach for refining image annotations.

We believe that image annotation refinement will be research focus in the future. However, to make image annotation ideal performance still has a long way to go. To really advance the image annotation and refinement, it has been increasingly accepted that other research domains achievements are need. In the future work, researchers may combine image analysis and understanding with context information, prior knowledge, pattern recognition, machine learning, and more domain knowledge to improve refinement algorithm.

ACKNOWLEDGMENT

We would like to express our grateful thanks to the faculty and students in Institute of Computer Graphics and Image Processing for their helping in experiment and evaluation. These include Xiaoli Zhang, Huojie Li, Lintong Huang, Fangfang Han, and Yixi Zhou. This work was supported in part by National Natural Science Foundation of China under grants 60675008.

REFERENCES

[1] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, Ramesh Jain, “Content-Based Image Retrieval at the End of the Early Years”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22(12) , pp.1349 – 1380, December 2000.

[2] Jing Liu, Mingjing Li, Wei-Ying Ma, Qingshan Liu, Hanqing Lu, “An adaptive graph model for automatic image annotation”, Proceedings of the 8th ACM international workshop on Multimedia information retrieval, pp.61-70, New York :ACM Press ,2006.

[3] Jia Li, James Z. Wang, “Automatic linguistic indexing of pictures by a statistical modeling approach”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25(9), pp1075-1088, September 2003.

[4] Lim J-H, Tian Q, Mulhem P, “Home photo content modeling for personalized event-based retrieval”, IEEE Multimedia, vol.10(4), pp.28-37, 2003.

[5] Park SB, Lee JW, Kim SK, “Content-based image classificatio using a neural network”, Pattern Recognition Letters, vol. 25, pp.287-300, 2004.

[6] Kim, S. Park, S., Kim, M., “Image classification into object/non-object classes” Proceedings of the International Conference on Image and Video Retrieval, Lecture Notes in Computer Science 3115 Springer 2004, Berlin ; New York: Springer Press, pp.393-400, 2004.

[7] Chapelle O, Haffner P, Vapnik VN., “Support vector machines for histogram-based image classification”, IEEE Trans Neural Networks, vol.10(5), pp. 1055-1064, 1999

[8] Brank J. Using image segmentation as a basis for categorization. Informatica, 2002; 26: 4.

[9] Qbal Q, Aggarwal JK., “Retrieval by classification of images containing large manmade objects using perceptual grouping”, Pattern Recognition, vol.35(7), pp.1463-1479. 2002.

[10] Cheng Y-C, Chen S-Y., “Image classification using color, texture and regions”, Image and Vision Computing, vol. 21(9), pp. 759-776, 2003.

[11] Vailaya A, Figueiredo AT, Jain AK, Zhang H-J., “Image classification for content-based indexing”, IEEE Trans Image Proc, vol. 10(1), pp.117-130, 2001.

[12] Luo J., Savakis A., “Indoor vs outdoor classification of consumer photographs using low-level and semantic features”, Proceedings of the IEEE International Conference on Image Processing, IEEE Press, pp. 745-748, 2001.

[13] Jeon J., Manmatha R. “Using maximum entropy for automatic image annotation”, Proceedings of the International Conference on Image and Video Retrieval, Berlin ; New York : Springer Press, pp.24-32, 2004.

[14] Wang JZ, Li J., “Learning-based linguistic indexing of pictures with 2-D MHMMs”, Proceedings of the ACM International Conference on Multimedia, New York :ACM Press, pp. 436-445, 2002.

[15] Y. Mori, H. Takahashi, and R. Oka., “Image-to-word transformation based on dividing and vector quantizing images with words”, In First International Workshop on Multimedia Intelligent Storage and Retrieval Management, New York: ACM Press, 1999.

[16] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. , “Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary”, In Seventh European Conference on Computer Vision, Berlin ; New York : Springer Press, pp.97-112, 2002.

[17] Pan, J.Y., Yang, H.J. and Pinar, D., “Automatic multimedia cross-modal correlation discovery”, The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York : ACM Press, pp.653-658, August 2004.

[18] J. Jeon, V. Lavrenko, and R. Manmatha, “Automatic image annotation and retrieval using cross-media relevance models”, In Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in information Retrieval, New York : ACM Press, 2003.

[19] 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 , New York: ACM Press, 2005.

[20] S. L. Feng, R. Manmatha, and V. Lavrenko, “Multiple Bernoulli relevance models for image and video annotation”, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol.2, pp.1002-1009, IEEE CS Press, 2004.

[21] Yohan Jin, Latifur Khan, B.Prabhakaran, "To be Annotated or not?: the Randomized Approximation Graph Algorithm for Image Annotation Refinement Problem",Â Proceedings of the 24th International Conference on Data Engineering, IEEE Press, April 2008.

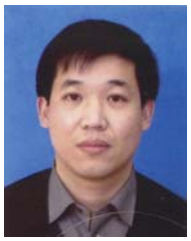
[22] C. Wang, F. Jing, L. Zhang, and H. J. Zhang, “Image Annotation Refinement using Random Walk with Restarts“, In Proceedings of the 14th Annual ACM international Conference on Multimedia, New York: ACM Press, 2006.

[23] Ricaro Baeza-Yates, Berthier Ribeiro-Neto, Modern Information Retrieval. New York: ACM Press，pp27-31, 1999.

[24] Changhu Wang, Feng Jing, Lei Zhang, Hong-Jiang Zhang. “Content-based Image Annotation Refinement”, IEEE Conference on In Computer Vision and Pattern Recognition, IEEE CS Press, pp.1-8, 2007.

[25] Haiyu Song, Xiongfei Li, Pengjie Wang, “Automatic Image Annotation based on Improved Relevance Model”, 2009 Asia-Pacific Conference on Information Processing, IEEE CS Press, pp.59-63, 2009.

[26] Jianbo Shi and Jitendra Malik, "Normalized Cuts and Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI), vol22(8), 2000
[27] Haiy Song, Xiongfei Li, Pengjie Wang, Jingrun Chen. Block-based Normalized-cut Algorithm for Image Segmentation, Lecture Notes in Electrical Engineering, Springer, January 2010, in press
[28] www.cs.washington.edu/research/imagedatabase/groundtruth/



Haiyu Song received the BS degree in computer & application in 1996, the MSc degree in computer software and theory in 2003, all from Jilin University, china. Now he is a Ph.D. Candidate in computer software and theory at Jilin University.

During 1996-2000, he was a software engineer in ACBC Co., Ltd., china, where the team chaired by him has developed a data acquisition and management system for PHILIPS gauge, rejected product automatic recognition system, and other pattern recognition software. Since 2003, he has been a member of the faculty of the computer science and engineering at Dalian Nationalities University, Dalian, China. He has authored or coauthored more than 20 research papers, such as "A General Multi-step Algorithm for Voxel Traversing Along a Line" (*Computer Graphics Form*, vol27(1), 2008), "Study on Model and Method of Image Data Mining" (*Journal of Jilin University*, vol32(1), 2002). His research interests include image analysis and understanding, image retrieval, data mining, and computer graphics.

Xiongfei Li received the BS degree in computer software in 1985 from Nanjin University, the MSc degree in computer software in 1988 from the Chinese academy of sciences, the PhD degree in communication and information system in 2002 from Jilin University.

Since 1988, he has been a member of the faculty of the computer science and technology at Jilin University, Changchun, China. He is a professor of computer software and theory at Jilin University. He has authored more than 40 research papers, including "A Data Mining Algorithm Based on Calculating Multi-Segment Support"(*Chinese Journal of Computer*, vol24(6), 2001), "An Algorithm for Discovering Association Rules Based on the Attributes of Items"(*Chinese Journal of Computer*, vol25(12), 2002). His research interests include data mining, intelligent network, and multimedia.

Prof. Li is a member of the IEEE. He is also an expert of minister of education evaluating invite tender for electromechanical product.

Pengjie Wang received the BS degree in computer science and technology in 2001, the MSc degree in computer software and theory in 2004, all from Jilin University, china. Now he is a Ph.D. Candidate in computer software and theory at Zhejiang University.

Since 2004, he has been a member of the faculty of the computer science and engineering at Dalian Nationalities University, Dalian, China. He has authored or coauthored more than 20 research papers, including "Improved point-by-point algorithm for generating nonparametric curves" (*Journal of Computational Information Systems*, vol3 (2), 2007), "Lossless Compression of Point-based Models"(*Journal of Information and Computational Science*, vol5(6)， 2008). His research interests include computer graphics, and digital entertainment.