Fusion of Region and Image-based Techniques for Automatic Image Annotation

Yang Xiao[1] Tat-Seng Chua[1] Chin-Hui Lee[2]

[1] School of Computing, National University of Singapore, Singapore, 117543
[2] School of Electrical and Computer Engineering Georgia Institute of Technology,
Atlanta, GA. 30332, USA
xiaoy@comp.nus.edu.sg chuats@comp.nus.edu.sg chl@ece.gatech.edu

ABSTRACT. We propose a concept-centered approach that combines region- and image-level analysis for automatic image annotation (AIA). At the region level, we group regions into separate concept groups and perform concept-centered region clustering separately. The key idea is that we make use of the inter- and intra-concept region distribution to eliminate unreliable region clusters and identify the main region clusters for each concept. We then derive the correspondence between the image region clusters and concepts. To further enhance the accuracy of AIA task, we employ a multi-stage kNN classification using the global features at the image level. Finally, we perform fusion of region- and image-level analysis to obtain the final annotations. Our results have been found to improve the performance significantly, with gains of 18.5% in recall and 8.3% in “number of concepts detected”, as compared to the best reported AIA results for the Corel image data set.

Keywords: Automatic Image Annotation, multi-stage kNN, Kullback-Leibler divergence

1. INTRODUCTION

Conventional content-based image retrieval (CBIR) systems require users to retrieve images based on low-level content attributes. Ideally, the users would prefer to query an image database by issuing text-based semantic queries. To facilitate text-based retrieval of images, the images must be annotated with a set of concepts. The automatic image annotation (AIA) involves the analysis of low-level content features of images at the regions/blocks or image level to infer the presence of semantic concepts.

AIA has received extensive attention recently. Starting from a training set of annotated images, many statistical learning models have been proposed in the literature to associate low-level visual features with semantic concepts [1,3,5,17,18,19]. The methods can be divided into two groups: the image-based vs. the region-based methods. The image-based methods [1] attempt to directly label images with concepts based on the selection of low level global features. These methods result in low-cost frameworks for feature extraction and image classification. But using only global visual properties limit their effectiveness to mostly scene-type concepts and are not effective for object-type concepts. The second group is the region-based methods [3,5,9,10,11,12,17,18,19] that are based on the idea of first dividing the images into regions or fixed-sized blocks. A statistical model is then learnt from the annotated training images to link image regions directly to concepts and use this as the basis to annotate testing images. Most existing region-based methods adopt the discrete approach by tackling the problem in two steps: (1) clustering all image regions to region clusters; and (2) finding joint probability of region clusters and concepts. The performance of region-based methods is strongly influenced by the quality of clustering and consequently the linking of region clusters and concepts, both of which are unsatisfactory.

One of the problems of current AIA systems is that the analysis is carried out at the region or image level. The region level analysis is limited by the accuracy of clustering, and is able to capture mostly object level information. On the other hand, image level analysis is simple but is able to capture only global scene level contents. To overcome the problems of both techniques and to enhance the overall AIA performance, we need to analyze image semantics at multiple levels, the content (region) and concept (image) levels. Thus in this research, we propose a novel concept-centered framework to facilitate effective multi-level annotation of images at region and image levels. The main techniques and contributions of our work include: (1) We propose a novel concept-centered region-based clustering method to tackle the correspondence between the concepts and regions. The process utilizes intra- and inter-concept region distributions to automatically identify the main region clusters (blobs) for each concept, obtain the representative region clusters and typical features for each concept, and use the information to annotate the regions. (2) We perform multi-level annotation by fusing the results of region-level and image-level annotations.

The rest of the paper is organized as follows. Section 2 presents a brief overview of the design of the system. Section 3 discusses the region-based concept-centered technique. Section 4 describes the image-based multi-stage kNN classifier. In Section 5, the image- and region-level results are fused in two stages to produce the multi-level semantics for the testing images, along with results and discussions. Finally Section 6 concludes the paper.

2. SYSTEM DESIGN

To address the limitations of current AIA systems, our concept-centered AIA system aims to solve the correspondence between image regions and concepts at region-level analysis, and then combine region- and image-level analysis for automatic image annotation, which produces multi-level (both concept level and content level) semantics of images. The overall system consists of 4 main modules as shown in Figure 1.

As with most research in AIA, we consider the case where the concepts are annotated at the image level. Hence each segmented region within an image will inherit all the concepts annotated for that image. As only one or two concepts are likely to be relevant to a particular region, the problem then becomes one of identifying the main concept associated with each region, while eliminating the rest of co-occurring concepts at the image level. To tackle the problem, Module 1 performs concept-centered region clustering to identify the main region clusters for each concept by taking into consideration the inter- and intra-region distributions. The main region clusters for each concept are used later as the basis to associate regions to concepts. To incorporate image level AIA, Module 2 performs multi-stage kNN classification at the image level to deduce the most similar images. This is based on the assumption that images with same semantic contents usually share some common low-level features. The kNN of similar images are used for refining region-level candidates and performing image-level annotation. Next, we perform the fusion of region- and image-level results in two stages. Module 3 performs an essentially region-based AIA that uses multi-stage kNN to constrain the results. We expect the outcome to be high precision annotation at the region-level. Module 4 fuses the AIA results of image-level and region- level method using Bayesian method. We expect the eventual results at image-level to have high recall while maintaining the precision of the region based method.

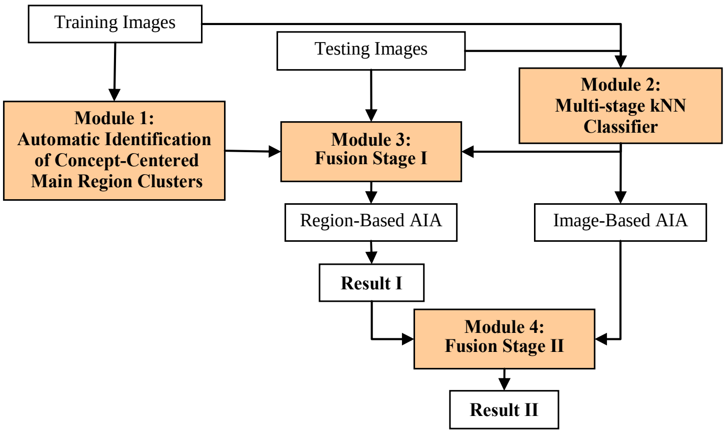

Training Images
Testing Images
Module 1: Automatic Identification of Concept-Centered Main Region Clusters
Module 2: Multi-stage kNN Classifier
Module 3: Fusion Stage I
Region-Based AIA
Image-Based AIA
Result I
Module 4: Fusion Stage II
Result II

Figure 1. Concept-centered AIA system workflow

3. CONCEPT-CENTERED REGION CLUSTERING

3.1 Overview of Concept-Centered Region-Based Clustering

At the region level, we first perform the segmentation of training images into regions and merge the smaller regions into modified regions using the k-Means method. As we do not know which specific concept is relevant to which region, we simply associate all annotation concepts for the training image to all its regions. The existing methods treat an image as consisting of a set of region clusters and analyze the semantic concept of each region cluster to build a vocabulary of concepts to represent the whole image. Two difficulties arising from this approach are: (1) how to generate the region clusters of the whole image set; and (2) how to analyze the semantic contents of each region cluster with respect to a set of pre-defined concepts. To overcome the first problem, instead of performing clustering of all regions across all concepts as is done in most current approaches, we group regions into separate concept groups based on the concepts that they have inherited. By specifically focus on the regions that have the possibility of representing this concept, we hope to minimize the noise resulting from clustering of heterogeneous regions across all concepts using low-level features. At the concept level, we perform clustering of the regions from different images using the k-Means clustering and Davies-Bouldin validation method to group similar regions to clusters. Optimal k for k-Means is decided by the following steps: We run the k-Means on the given dataset multiple times for different k, and the best of these is selected based on sum of squared errors. Finally, the Davies-Bouldin index

$$DB = \frac{1}{k}\sum_{i=1}^{k}\max_{i\neq j}\left\{\frac{\Delta(X_i)+\Delta(X_j)}{\delta(X_i,X_j)}\right\} \qquad \textbf{(1)}$$

is calculated for each clustering [15], where $\delta(X_i,\ X_j)$ defines the intercluster distance between clusters X_i and X_j; $D(X_i)$ represents the intracluster distance of cluster X_i, and k is the number of clusters. Small index values correspond to good clusters, that is to say, the clusters are compact and their centers are far away from each other. Therefore, $argmin_k(DB)$ is chosen as the optimal number of clusters, k. Consequently, we obtain several clusters under each concept.

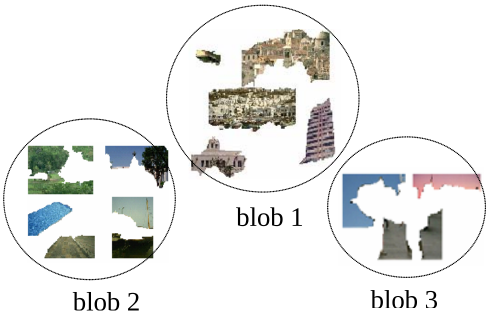


Figure 2. Example of region clusters for the concept “Building”

Figure 2 shows an example of the region clusters generated for the concept “Building”. As can be seen, blob 1 (or region cluster 1) is composed of the representative regions for the “Building” concept, while the others blobs may include regions for co-occurring concepts or a mixture of them. We call blob 1 the “main blob” of concept “Building”. In this research, we intend to automatically identify the main blobs of an individual concept. The main blobs found can then be used as the basis for region annotation, image annotation, and even image retrieval. The identification process involves two stages. First we eliminate the unreliable clusters, which are those that clearly do not represent the current (base) concept. Their elimination reduces the possible clusters for main cluster identification. Second, we identify the main blobs, which are the most representative of the base concept. The following sections describe the details of these two identification processes.

3.2 Unreliable Blob Identification

We aim to utilize the concept co-occurrence and the relationship of intra-concept region clusters to find the most unreliable region blobs, i^{un} , under the base concept *T*. Let *W(T)* represent the related (co-occurring) concepts with *T*, including *T* itself. The algorithm is as follows:

First, we cluster regions of training image set *I(T)* into *L* blobs $R(I(T)_i)$, $i=1,\ldots,L$.

Second, given an training image set *I(G)* where $G \in W(T)\backslash T$, we remove part of the images in *I(G)* that correspond to any concepts in $W(T)\backslash G$. The remaining image set is:

$$S_G = I(G) \backslash \bigcup_{x \in W(T)\backslash G} \left(I(x) \cap I(G) \right) \tag{2}$$

Here, *G* is the only shared concept between *I(T)* and S_G. This means that we have eliminated the probabilities that images in *I(T)* would be similar with images in S_G due to other concepts beside *G*.

Third, we cluster the regions of S_G into optimal number of clusters $R(S_{Gj})$, $j=1,\ldots,J$, and compute the Euclidean distance of intra-clusters:

$$\Lambda(i,j) = dist(R(I(T)_i), R(S_{Gj})) \ . \tag{3}$$

At $\arg\min_i(\Lambda)$, that region blob *i* under *I(T)* is most similar to certain blob under S_G. We increment V_i at $\arg\min_i(\Lambda)$, where V_i measures the degree of unreliability of blob *i*.

Fourth, we repeat the second and third steps on all related concepts $W(T)\backslash T$. The result

$$i^{un} = \arg\max_i (V_i) \tag{4}$$

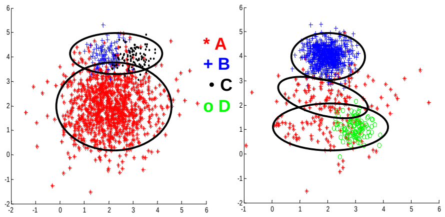
is the most unreliable blob for the base concept *T*.

3.3 Main Blob Identification

Next we aim to identify the main blob i^* , which best represent the semantic meaning within the blobs of the base concept:

$$i^* = \arg\max_i P(concept \mid blob_i) \text{, } i=1,\ldots,L. \tag{5}$$

First we investigate two properties of the distributions of regions in blobs under the base concept and all related concepts: (1) the representative regions are compactly-clustered under the base concept; and (2) they are dispersed under other related concepts. Figure 3 presents an example of region data projected in 2-D space to explain these properties. We assume that there are four kinds of region data, shown in different symbols, representing concepts A, B, C and D. Figures 3(a) and 3(b) respectively show the region distributions under concepts A and B. The ellipses represent the region blobs. It is observed that the representative regions for concept A are compactly-clustered under concept A, while dispersed under concept B, and vice versa. Also the representative regions for concept B in the blob of concept A are only part of the regions in the main blob under B. From the 1st property, regions of main blob under the base concept are distributed to more region blobs under related concepts than the non-representative regions. From the 2nd property, regions of non-main region blobs under the base concept are distributed to only one or few region blobs under their correspondent concepts.



(a) concept A (b) concept B

Figure 3. The distribution of regions under concepts A and B

Given the base concept T, after the elimination of unreliable blob i^{un} , the remaining L' region blobs are $R(I(T)_i)$, $i = 1,\ldots,L'$. Also, for every related concept of T, $B \in W(T) \backslash T$, we group and cluster the regions under related concept B into J region blobs $R(I(B)_j)$, $j=1,\ldots,J$. Then we build two functions, f and g, which focus on the relationship of region distribution by exploiting the above two properties to decide the main blobs. $f(i)$ makes use of Kullback-Leibler (K-L) divergence [13] to measure how well the distribution in blob set of related concepts matches the distribution in certain blob i of the base concept. On the other hand, $g(i)$ uses the distribution factor to measure the degree of distribution diversity of the image regions from blob i of base concept to the blobs of related concepts.

In probability theory and information theory, the K-L divergence is a natural distance measure from a "true" probability distribution p to an "arbitrary" probability distribution q. $f(i)$ is defined by the sum of all the related concepts on the mean K-L divergence between a certain blob i in the base concept T and the blob set $blobs(B)$ in a related concept B:

$$f(i) = \sum_{B \in W(T) \backslash T} \left(\frac{1}{\| blobs(B) \|} * \sum_{j \in blobs(B)} D_{KL}(p_j \| q_i) \right), \quad i = 1,\ldots,L'. \qquad \textbf{(6)}$$

where q_i is the distribution of $R(I(T)_i)$, p_j is the distribution of $R(I(B)_j)$, and $\|blobs(B)\|$ is the number of blobs in concept B.

For probability distributions p and q of a discrete variable the K-L divergence between p and q with respect to p is defined to be:

$$D_{KL}(p \| q) = \sum_{k} p(k) \log\left(\frac{p(k)}{q(k)}\right) . \qquad \textbf{(7)}$$

The K-L divergence is the expected amount of information that a sample from p gives of the fact that the sample is not from distribution q. From the above distribution property, the regions in the main blob of base concept, comparing with the regions in the other blobs, should be distributed more universally in the blobs of all the related concepts. So the main blob should get the minimization of *f(i)*, which means:

$$i^* = \arg\min_{i} f(i) , i = 1,\ldots,L' . \qquad \textbf{(8)}$$

On the other hand, for every other concept *B*, we record how the shared regions between *T* and *B* are distributed under each concept. To do this, we first compute *V(i,j)*, which is set to one if the region cluster *j* of concept *B* has share regions with region cluster *i* of base concept *T*. Otherwise *V(i,j)* is set to zero. We then compute the distribution parameter $N_{T,i,B}$, which is the number of region clusters in *B* that has shared regions with cluster *i* of base concept *T*, as follows:

$$N_{T,i,B} = \sum_{j \in blobs(B)} V(i, j) . \qquad \textbf{(9)}$$

After analyzing all the related concepts, the region clusters that achieve the maximum of that sum of $N_{T,i,B}$ on all related concepts *B* will be considered as the main blob of the base concept *T*:

$$i^* = \arg\max_{i} g(i) = \arg\max_{i} \left(\sum_{B \in W(T) \setminus T} N_{T,i,B} \right), \; i = 1,\ldots,L' . \qquad \textbf{(10)}$$

Finally, we fuse the results for main blob derived from the two functions:

$$i^* = F\left(\arg\min_{i} f(i), \arg\max_{i} g(i) \right), \; i = 1,\ldots,L' . \qquad \textbf{(11)}$$

where *F(·)* is simply an union operation in our test.

After we obtain the representative regions and typical features from the main blobs for each concept, we could use the information to annotate the regions and images. It will be discussed in Section 5.

4 IMAGE-BASED MULTI-STAGE KNN CLASSIFIER

Beside the region-level analysis, we perform image-level analysis using a multi-stage kNN technique. Since images with same semantic meaning usually share some common low-level feature, the multi-stage kNN can be used to perform image matching for annotation at the image level.

As illustrated in Figure 4, the multi-stage system can be viewed as a series of classifiers, each of which provides increased accuracy on a correspondingly smaller set of entities, at a constant classification cost per stage. It can exceed the performance of any of its individual layers only if the classifiers appearing at different layers employ different feature spaces [7]. For effectiveness of multi-stage kNN classifier, we arrange the features in the order that make the classifier at the 1st stage to have high sensitivity (few false negatives), while the classifier at the 2nd stage to have high specificity (few false positives) but less sensitivity. As compared to color histogram, the auto-correlogram is more stable to changes in color, large appearance, contrast and brightness. It thus serves as a good 1st stage feature to avoid removing too many false negatives, paving the way for the use of simple edge and color histogram features in the 2nd stage. So for the 1st stage, we adopt HSV auto-correlogram; while for 2nd stage, we use the HSV histogram combining with edge histogram. More specifically, we select the top 100 kNN images for the 1st stage and 4 nearest images for the 2nd stage.

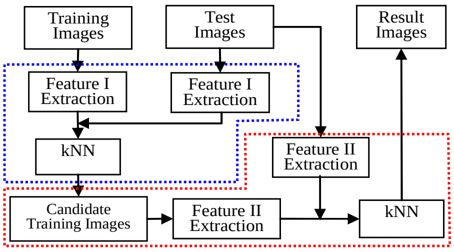


Figure 4. Image-based multi-stage kNN classifier

5 TWO-STAGE MULTI-LEVEL FUSION AND RESULTS

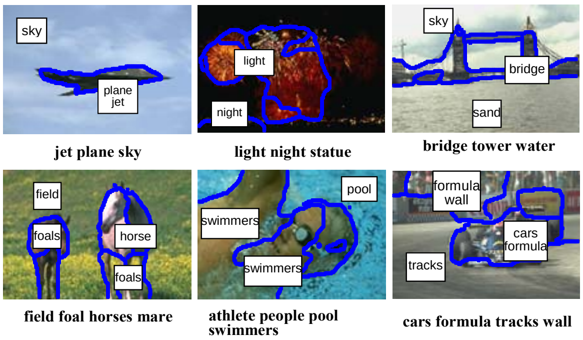
We fuse the region- and image-based results in two stages to perform automatic image annotation.

5.1 Fusion Stage I − Fusion of Region-Based Methods with Multi-stage kNN

The main objective of region-level analysis is to enhance the ability of capturing as well as representing the focus of user's perceptions to local image content. We have obtained the main region blobs of each concept for the training images in Section 3.3. The explicit concept of each training region can be determined from which main region blobs that it belongs to. During testing, in order to refine the possible concept range of the test images, we first apply the multi-stage kNN classifier as described in Section 4 to find several most similar training images for each test image. After that, for each region in the test image, kNN is again applied at the local region feature level to find the nearest 2 regions from among the regions of the most similar training images. The concepts of these two nearest training regions are assigned as annotated concepts of the test image region.

One advantage of region-based method is that it provides annotation at the region level. It allows us to pin-point the location of region representing each concept. It thus provides information beyond what is provided by most image-level annotation methods. The use of kNN to narrow the search range further enhances the precision. We thus expect the overall fusion to have good precision.

As with all the other experiments [3,5,6], we use the Corel data set that has 374 concepts with 4,500 images for training and 500 for testing. Images are segmented using the Normalized Cut [16] and each region is represented as a 30 dimensional vector, including region color, average orientation energy and so on [3]. The results are presented based on the 260 concepts which appear in both the training and test sets. Annotation results for several test images are showed in Figure 5. The concepts shown in the rectangles are the results of region annotation. Ground truth is shown under each image for comparison. The results show that our region-based technique could provide correct annotation at the region level in most cases.



jet plane sky

light night statue

bridge tower water

field foal horses mare

athlete people pool swimmers

cars formula tracks wall

Figure 5. Region annotation for images

Measuring the performance of methods that predict a specific correspondence between regions and concepts is difficult. One solution, applicable to a large number of images, is to predict the concepts for the entire images and use the image level annotation performance as a proxy. For each test image, we derive its annotation concepts by combining the concepts of each region that it contains and use this as the basis to compute the precision and recall. The number of concepts for each image is not restricted here. Table 1 shows the results of image level annotation in terms of average precision (P), recall (R), and F_1 over all the concepts, and the number of concepts detected (# of Det.), i.e. concepts with recall > 0. The results show that our region-based techniques could achieve an average F_1 measure of 0.20, with 114 detected concepts that have at least one correct answer.

Table 1. Result of region-based AIA

P	**R**	**F_1**	**# of Det.**
0.19	0.21	0.20	114

In comparison with the state-of-the-arts systems listed in Table 3, the performance of the region-based method is better than most except the top two systems. It should be noted that our region-based method provides annotation at region level as shown in Figure 5 instead of just at image level without location information. To enhance the annotation performance at the expense of location, we explore an image-based AIA approach in next Section.

5.2 Fusion Stage II − Fusion of Region-Based AIA and Image-Based AIA

At the image level, we first perform the multi-stage kNN to obtain several nearest training images for each test image. We sum up the concepts of these training images to arrive at a frequency measure for each available concept. To annotate the test image, we choose the highest frequency concepts until the desired number of concepts is reached. For those concepts with equal frequency, we give priority to those belonging to the annotation of the nearer image.

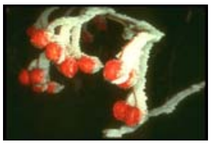
Test Image			
Ground-truth	deer forest tree white-tailed	caribou grass herd tundra	birds fly nest
Image-based	zebra herd plane grass **tree**	**grass** flowers	**birds** flowers
Region-based	water **deer white-tailed** giraffe	**tundra** flowers **herd**	tree **birds** flowers **nest fly**
Fusion	zebra herd plane grass **tree** water **deer white-tailed**	**grass** flowers **tundra herd**	**birds** flowers tree **nest fly**
Test Image			
Ground-truth	buildings clothes shops street	frost fruit ice	food market people shops
Image-based	**clothes street** museum fountain	**frost ice** spider	buildings tree farms
Region-based	**buildings shops** **street** people	water **ice fruit**	clouds **people** house **shops**
Fusion	**clothes street** museum **buildings shops** people	**frost ice** spider water **fruit**	buildings tree farms clouds **people** house **shops**

Figure 6. Annotation results of image- and region-based methods

To illustrate the results of image-based method against that of the region-based method obtained in Section 5.1, Figure 6 shows some automatic annotation results of both methods. Under each image, the ground truth are shown at the top line, followed by the annotation results of the image-based method in the middle line, with the results of region-based method at the third line. Concepts in bold correspond to correct matches. It can be seen that global feature-based results at image level are more concerned with abstract background and frequently occurring concepts, while local region based results are more concerned with specific object-type concepts. It is clear that both methods produce different results, and we should be able to improve the results further by combining both.

Thus, in order to improve the recalls of the overall performance, we employ Bayesian fusion method [4] to perform the fusion. We expect the final results to have better recall while maintaining high precision.

We use the same Corel data set as described in Section 5.1. Table 2 shows the results of AIA for region-based (R_B), image-based (I_B), and fusion (R+I) methods. The desired number of concepts for each test image is set to 8. We can see from Table 2 that the fusion improves the overall performance, with the F_1 measure improve steadily from 0.20 (region-based method) to 0.24 (image-based method) and then to 0.26 (fusion of both). The number of detected concepts reaches 144 for the fusion approach. It is clear that fusion improves the performance for either the region-based AIA or image-based AIA. Figure 6 gives examples of the concepts annotated using the fusion approach (shown in line 4 under each image). It can be seen from the examples that our proposed method is able to infer more correct annotations.

Table 2. Results of fusing region and image-based AIA

	P	**R**	**F_1**	**# of Det.**
R_B	0.19	0.21	0.20	114
I_B	0.23	0.26	0.24	122
R+I	0.23	0.32	0.26	144

Comparison with published results for same data set is listed in Table 3. The results show that our proposed method outperforms the continuous relevance model and other models on the Corel data set. It achieves the best average recall and best number of detected concepts. At the same time, our precision is not too bad. Overall, it improves the performance significantly by 18.5% in recall and 8.3% in the “number of concepts detected”, as compared to the best result that has been reported.

Table 3. Comparison with other results

Method	**P**	**R**	**# of Det.**
TM [3]	0.06	0.04	49
CMRM [17]	0.10	0.09	66
ME [18]	0.09	0.12	N.A.
CRM [19]	0.16	0.19	107
MBRM [5]	0.24	0.25	122
MFoM [6]	0.25	0.27	133
Proposed	**0.23**	**0.32**	**144**

6 CONCLUSION

In this paper, we proposed a novel concept-centered region-based approach for correlating the image regions with the concepts, and combining region- and image-level analysis for multi-level image annotation. At the region level, we employed a novel region-based AIA framework that centers on regions under a specific concept to derive region semantics. Our system aims for automatic identification of the main region blob under each concept by using inter- and intra-concept region distribution. The main region blobs found are then used to determine the explicit correspondence of region to concept. At the image level, we applied a multi-stage kNN classifier based on global features to help region-level AIA. Finally, we performed the fusion of region- and image-based AIA. The results have been found to outperform previously reported AIA results for the Corel dataset.

For future work, we plan to further explore the integration of region- and image-based techniques for image/video classification and retrieval.

REFERENCES

1. A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42:145-175, 2001.
2. A.Yavlinsky, E. Schofield and S. Rüger. Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation. *Int'l Conf on Image and Video Retrieval* (CIVR), 507-517, Springer LNCS 3568, Singapore, July 2005.
3. P. Duygulu, K. Barnard, N. de Fretias, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the European Conference on Computer Vision*, 97-112, 2002.
4. Richard O. Duda, Peter E. Hart, David G. Stork. Pattern Classification, 2nd Edition. Nov. 2000.
5. S. Feng, R. Manmatha, and V. Lavrenko. Multiple bernoulli relevance models for image and video annotation. In *the Proceedings of the International Conference on Pattern Recognition* (CVPR), volume 2, 1002-1009, 2004.
6. S. Gao, D.-H. Wang, and C.-H. Lee. Automatic Image Annotation through Multi-Topic Text Categorization. To appear in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), Toulouse, France, May 14-19 2006.
7. T. E. Senator. Multi-Stage Classification. In *Proceedings of the Fifth IEEE International Conference on Data Mining* (ICDM'05), 2005.
8. T. -S. Chua, S.-Y. Neo, H.-K. Goh, M. Zhao, Y. Xiao, G. Wang. TRECVID 2005 by NUS PRIS in TRECVID 2005, NIST, Gaithersburg, Maryland, USA, Nov 14-15 2005.
9. K. Barnard and D. Forsyth. Learning the semantics of words and pictures. In *International Conference on Computer Vision*, 2: 408-415, 2001.
10. P. Brown, S. D. Pietra, V. D. Pietra, and R. Mercer. The mathematics of statistical machine translation: Parameter estimation. In *Computational Linguistics*, 19(2):263-311, 1993.
11. Y. Mori, H. Takahashi, and R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In *MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, 1999.
12. D. Blei and M. I. Jordan. Modeling annotated data. In *26th Annual International ACM SIGIR Conference*, 127-134, Toronto, Canada, July 28-August 1 2003.
13. S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22, 1951.
14. J.-Y. Pan, H.-J. Yang, C. Faloutsos, and P. Duygulu. Automatic multimedia cross-modal correlation discovery. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD2004), Seattle, WA. August 22-25, 2004.
15. D.L.Davies, D.W.Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 1, No. 2, pp. 224-227, 1979.
16. J. Shi and J. Malik, Normalized cuts and image segmentation. *Proc. of IEEE CVPR 97*, 1997.
17. J. Jeon, V. Lavrenko and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models, In *Proc. of 26th Intl. ACM SIGIR Conf.*, pp. 119–126, 2003.
18. J. Jeon and R. Manmatha. Using maximum entropy for automatic image annotation. In *Proc. of the Int'l Conf on Image and Video Retrieval (CIVR 2004),* 24-32. 2004.
19. V. Lavrenko, R. Manmatha and J. Jeon, A model for learning the semantics of pictures, *Proc. of NIPS'03*, 2003.