International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.12 (2015), pp.59-66
http://dx.doi.org/10.14257/ijsip.2015.8.12.07

Content-Based Image Retrieval Improved by Incorporating Semantic Annotation via Query Expansion

Guoqing Xu, Jian Li, Chunyu Xu and Qi Wang

School Of Computer and Information Engineering
Nanyang Institute of Technology, Nanyang China
xu@mail.nyist.edu.cn

Abstract

Automatic image annotation (AIA) is expected to be a promising way to improve the performance of content-based image retrieval (CBIR). However, current image annotation results are always incomplete and noisy, and far from practical usage. In this paper, we incorporate semantic annotations into CBIR via query expansion scheme to improve retrieval accuracy. In the proposed method, semantic annotations of test images are obtained using a visual nearest-neighbor-based annotation model. And both visual features and annotation keywords are used to represent images. The similarity between two images is determined by their visual similarity and semantic similarity. The method is evaluated on the well-known Pascal VOC 2007 dataset using standard performance evaluation metric. The experimental results indicate that the performance of CBIR can be improved by incorporating semantic annotation via query expansion.

Keywords: *AIA, CBIR, query expansion, semantic similarity*

1. Introduction

Image retrieval has received much attention from researchers in the area of image understanding [1]. In the last decade significant progress has been made in the development of image retrieval. Generally, image retrieval studies can be classified into three categories: traditional text based annotation, CBIR and AIA [2]. CBIR can index and retrieve images automatically using visual features. But CBIR always suffers from the problem of so-called semantic gap, and fail to achieve satisfactory retrieval results [3]. AIA was proposed to overcome the shortcoming. The aim of AIA is to automatically assign a few relevant keywords to new images base on their visual contents [4]. Then image retrieval can be formulated as text retrieval problem. Unfortunately, AIA is a challenging task, and the annotation results are usually incomplete and contained wrong concepts [5]. Obviously the performance of retrieval is seriously affected if these annotations are used directly.

The limitation of image retrieval algorithms motivated researchers to develop new frameworks to retrieve desired images. In order to improve retrieval performance, automatic query expansion method was introduced in image retrieval. Zhou *et. al.*, [6] used a soft query expansion approach based on keyword similarity matrix learned through user relevance feedback information, to infer keywords related to the query. However, there are not always enough feedback information to make the methods perform effectively. A context expansion approach was explored by expanding the key regions of the queries using highly correlated environmental regions in [7], to take advantages of the correlations between image key regions and environmental regions. Recently, an retrieval framework based on automatic query expansion in a concept feature space was presented by generalizing the vector space model of information retrieval in [8], where images are modeled as ‘bag of concepts’. Most of these methods used visual features with semantic information.

ISSN: 2005-4254 IJSIP

In this paper, we incorporate semantic annotation into CBIR via query expansion approach. The image retrieval process is showed in Figure 1.

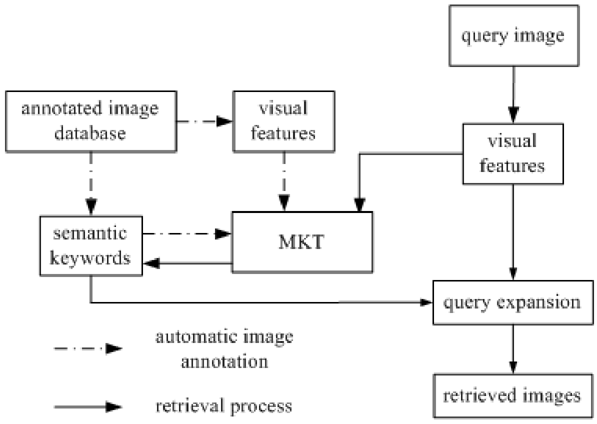


Figure 1. Framework of Query Expansion for Image Retrieval Incorporated Semantic Annotations

In this method, visual features are extracted first of all and automatic image annotation is performed using Modified keywords transfer mechanism (MKT) for the test images. Then for a query image, its visual features and annotated keywords are matched separately with test images. And images are ranked by sorting their match scores in descending order. We detail our approach in the following.

2. Proposed Approach

2.1. Visual Feature Extraction and Automatic Image Annotation

Commonly used visual features in CBIR include color, texture and shape. It has been shown that employing multiple kinds of features to describe an image is a good choice for image annotation and retrieval. We choose to extract six visual features to comprehensively represent images, including Color Layout descriptor, Scalable Color descriptor, Grid Color Moment, Edge Histogram descriptor, Homogeneous Texture descriptor and Wavelet Moment.

Then images can be annotated. MKT is used here as the annotation model due to its good performance [9]. Please refer to [9] for more information regarding MKT.

2.2. Query Expansion Using Annotations for CBIR

When a query image is given, the corresponding visual features and semantic annotations are obtained by applying the steps above, and both of them are used as the query terms. For a query image, its visual features and annotated keywords are matched separately with the annotated images.

The visual distance between two images is the sum of normalized distance between their six visual features. Semantic similarity is another determining factor. Several techniques are proposed to compute semantic similarity between keywords, such as information content based and graph-based similarity measure. They are proposed for finely annotated keywords. However, keywords obtained through AIA are incomplete and noisy. In addition, there are usually more than one semantic keywords assigned to one image. Hence, we just make use of shared keywords of two images. For simplicity, the number of shared keywords is considered as the measurement of semantic similarity. As the number of shared keywords of two images increase, the similarity grows linearly.

Then for each pair of images, the visual similarity and semantic similarity are normalized respectively, and the sum of the two determines the total similarity between two images. For each query image, retrieved images are ranked by sorting the similarity scores in descending order.

All printed material, including text, illustrations, and charts, must be kept within the parameters of the 8 15/16-inch (53.75 picas) column length and 5 15/16-inch (36 picas) column width. Please do not write or print outside of the column parameters. Margins are 1 5/16 of an inch on the sides (8 picas), 7/8 of an inch on the top (5.5 picas), and 1 3/16 of an inch on the bottom (7 picas).

3. Experiments and Discussion

3.1. Experimental Data and Performance Metrics

To test the performance of the proposed approach, automatic image annotation and image retrieval experiments are conducted on the PascalVOC07 (Pascal Visual Object Classes Challenge 2007) database [10]. This database is created for recognizing objects from a number of visual object classes in realistic scenes. There are 20 object classes used in this database. And there are 9963 annotated images in total, and each image is associated with 2.47 keywords on average.

For the automatic image annotation experiment, 8967 images with annotations are taken as the ground truth and the rest is used for testing. The test-train ratio is consistent with typical annotation methods, such as the label transfer mechanism [11]. The effectiveness of annotation performance is measured using precision, recall and F1, following previous studies.

For image retrieval experiment, images annotated using MKT are used to test the performance of retrieval incorporated semantic annotations. That is to say, the annotated 996 images are used as the testing database. Each image in testing database is taken as a query, and its annotation keywords are taken as expansion. All images in testing database including query image are ranked according to their similarities to the query image. And one retrieved image is considered to be relevant to the query if they share at least one keyword within the ground truth annotations, as indicated in [5]. The effectiveness of retrieval performance is measured using average Precision@K. For one query, Precision@K is the proportion of the relevant images among the top-K retrieved images.

3.2. Results and Comparisons

Each image is assigned with several keywords ranging from 1 to 6. The precision, recall and F1 values under different number of keywords are shown in Table 1.

Table 1. Annotation Performance using MKT (%) [9]

No. of annotated keywords	1	2	3	4	5	6
precision	30.15	27.81	21.45	19.20	16.83	15.04
recall	16.53	31.23	41.75	52.12	59.70	65.94
F1	21.35	29.42	28.34	28.06	26.26	24.49

For each query image in testing database, all the annotated keywords are used as expansion. Since each image can be assigned with one to six keywords, we test the performance with same number of keywords expansion (referred to as QEIS), which can be seen in Figure 2-7. We count Precision@K with K varying from 5 to 30 with an interval of five.

For comparison, the performance of CBIR where only the visual features are used is given as the baseline method (referred to as visual-baseline). Furthermore, the retrieval performances where query images are represented using only keywords are also shown for comparison (referred to as semantic-baseline).

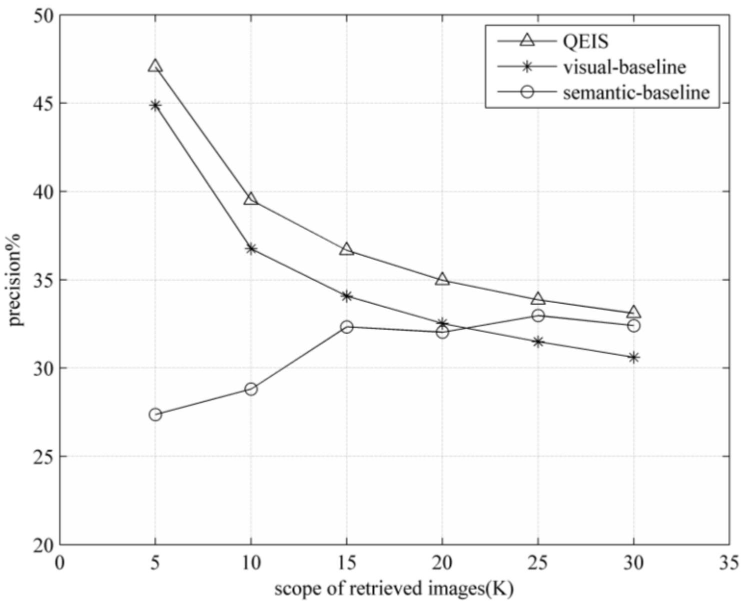


Figure 2. Precisions@K with One Keyword for Query Expansion

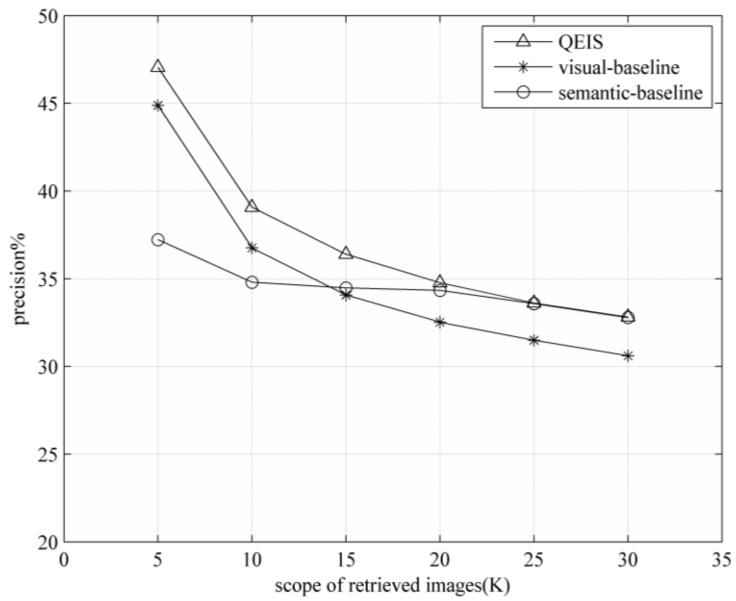


Figure 3. Precisions@K with Two Keywords for Query Expansion

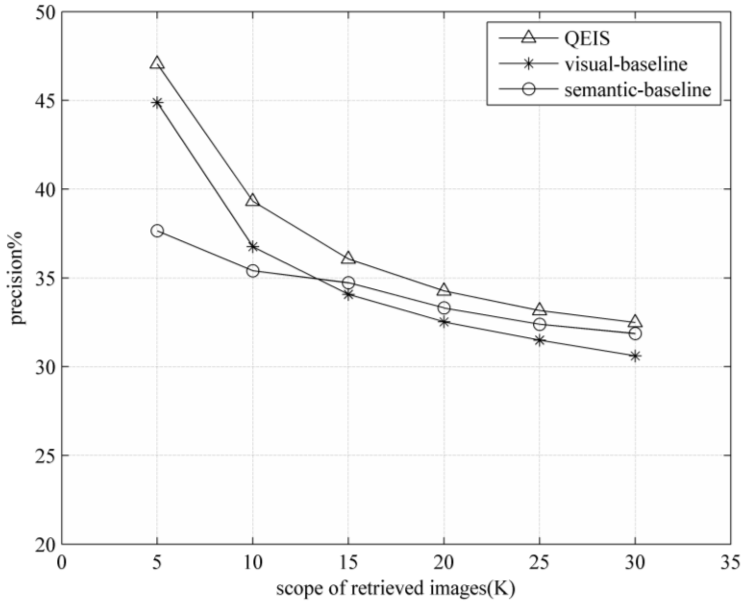


Figure 4. Precisions@K with Three Keywords for Query Expansion

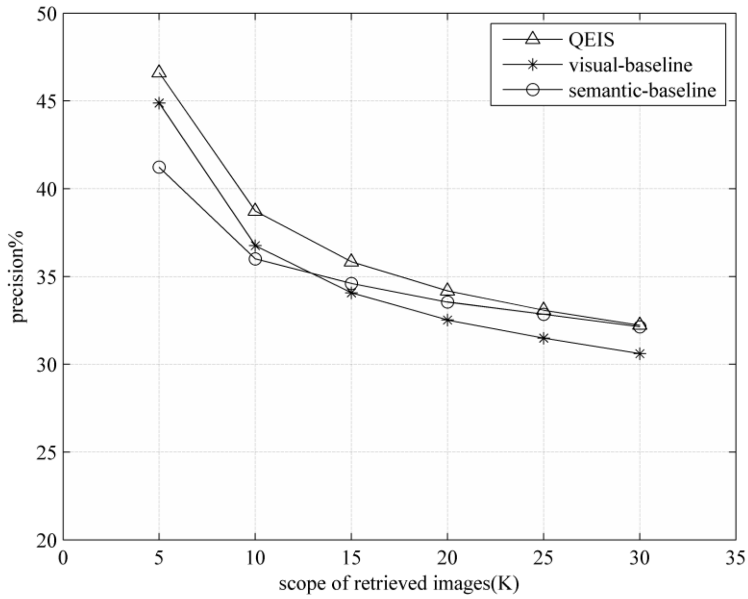


Figure 5. Precisions@K with Four Keywords for Query Expansion

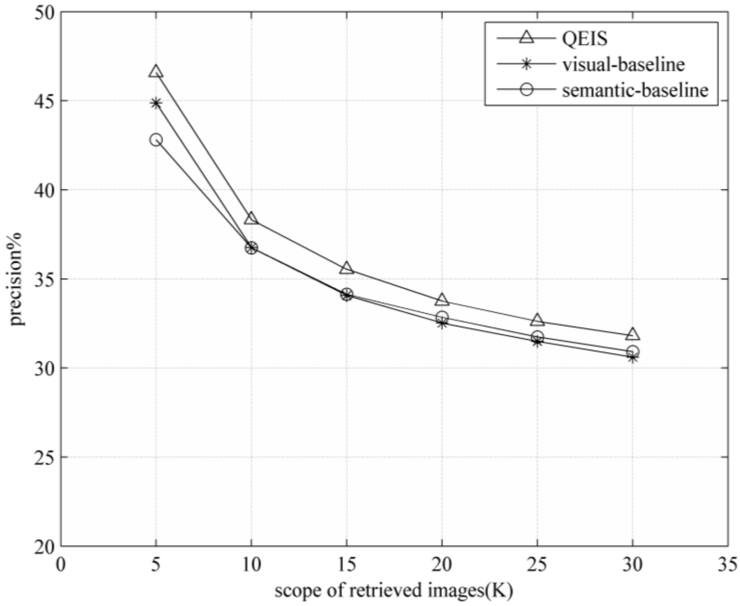


Figure 6. Precisions@K with Five Keywords for Query Expansion

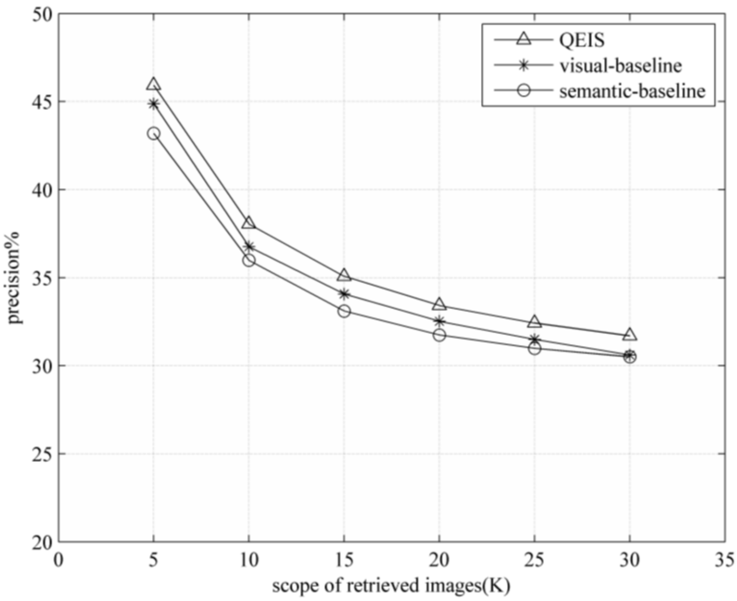


Figure 7. Precisions@K with Six Keywords for Query Expansion

From Figure 2-7, we can observe some interesting results. Firstly, although the annotation keywords are incomplete and noisy, the retrieval performance is improved when incorporating them via query expansion. Let's take Figure 2 for instance, where one keyword is used as expansion. For visual_baseline and semantic_baseline methods, the Precision@10 are 36.7% and 28.8% respectively, and it is increased to 39.5% using query expansion. Secondly, regardless of the number of keywords used as expansion, retrieval performance is improved by query expansion. For example, the Precision@10 is 39.1% when two keywords are incorporated and 38.3% when five keywords are incorporated, as shown in fig3 and fig6 respectively. Both of them are higher than that of visual_baseline. Thirdly, for different number of retrieved images, precisions are increased in most cases. As shown in Figure 7, Precisions@K with K= {5, 10, 15, 20, 25, 30} of QEIS are higher than that of visual_baseline and semantic_baseline methods.

Figure 8 shows some examples of retrieved image with one keyword used as expansion.

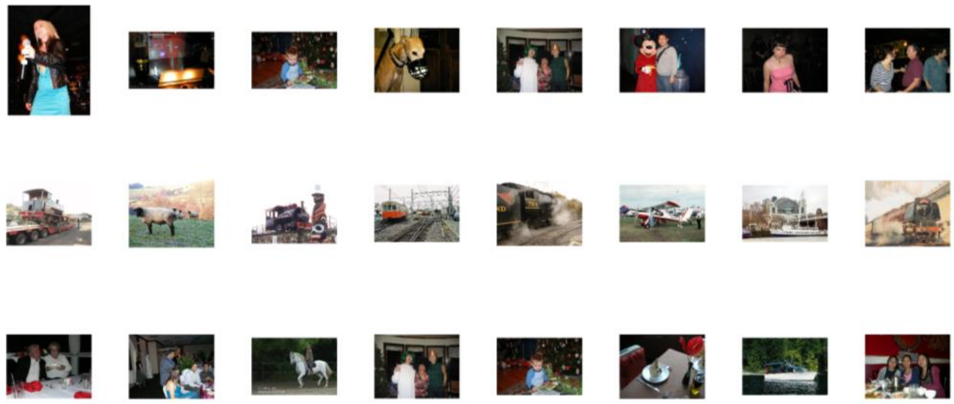


Figure 8. Illustration of Image Retrieval Results using Query Expansion. The First Column Shows Three Query Images, and the Top Eight Retrieved Images are Given in the Corresponding Row

The three images in the first column are taken as query separately, and the top eight retrieved images (covering query image) are listed in the corresponding row.

4. Conclusions

We have proposed a query expansion method which incorporates semantic annotations to improve the performance of content-based image retrieval. Though annotation keywords are incomplete and noisy, better performance can be obtained when they are incorporated into content-based image retrieval. The evaluation on the well-known Pascal VOC 2007 dataset indicates the effectiveness of the proposed method.

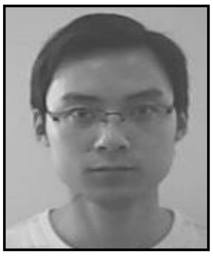
Acknowledgment

This work is supported by the Scientific Research Staring Foundation of Nanyang Institute of Technology and Key scientific research project for He'nan Higher Education.

References

[1] R. Datta, D. Joshi, J. Li and J. Z. Wang, ACM Computing Surveys, vol. 40, no. 2, **(2008)**.
[2] D. Zhang, M.M. Islam, G. Lu, Pattern Recogn., vol. 45, no. 1, **(2012)**.
[3] C. Lin, X. Dong, I.W. Tsang and L. Jiebo, IEEE T Multimedia, vol. 14, no. 4, **(2012)**.
[4] Z. Shaoting, H. Junzhou, L. Hongsheng and D. Metaxas, IEEE T SYST MAN CY B, vol. 42, no. 3, **(2012)**.
[5] L. Wu, R. Jin and A.K. Jain, IEEE T PATTERN ANAL, vol. 35, no. 3, **(2013)**.
[6] X.S. Zhou and T.S. Huang, IEEE MultiMedia, vol. 9, no. 2, **(2003)**.
[7] X.-J. Wang, W.-Y. Ma and X. Li, MULTIMEDIA SYST, vol. 11, no. 4, **(2006)**.
[8] M. M. Rahman, S. K. Antani and G.R. Thoma, INFORM PROCESS MANAG, vol. 47, no. 5, **(2011)**.
[9] G.Q. Xu and Z.Z. Mu, Int J Comput Sci. Iss., vol. 10, no. 2, **(2013)**.
[10] M. Everingham, L. Gool, C.K. Williams, John Winn and Andrew Zisserman, Int J Comput Vis., vol. 88, no. 2, **(2010)**.
[11] A. Makadia, V. Pavlovic and S. Kumar, Int J Comput Vis., vol. 90, no. 1, **(2010)**.

Authors



Guoqing Xu, male, received the Ph.D. in Control Science and Control Engineering from the University of Science and Technology Beijing, China, in 2014. He is a lecturer in Nanyang Institute of Technology. His research interests include content-based image retrieval, automatic image annotation, machine learning, and pattern recognition.

Jian Li, male, is an associate professor in Nanyang Institute of Technology. He got his doctor's degree from the University of Science and Technology Beijing, China, in 2013. His research interests include digital signal processing and content-based image retrieval.

Chunyu Xu, female, is a lecturer in Nanyang Institute of Technology. Her research interests include digital signal processing and Image processing.

Qi Wang, female, received the Master's Degree in Communication and Information system from the Northwest University, China, in 2012. She is a lecturer in Nanyang Institute of Technology. Her research interests include digital signal processing, Image processing and pattern recongnition.