Semantic-based image retrieval in the VQ compressed domain using image annotation statistical models

Masoumeh Shariat , Amir-Masoud Eftekhari-Moghadam*

Department of Electrical, Computer and IT engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran

Received 12 December 2010; revised 25 April 2011; accepted 10 May 2011

Abstract

Since most of visual data is stored in the compressed form, investigating semantic retrieval techniques with the description capability of image semantics in the image compression domain is highly desirable. Regardless of the fact that content based image retrieval (CBIR) based on the Vector Quantization (VQ) compression method is more accurate than the other methods, it is expected that semantic retrieval can also be effective. Thus, the goal of this study is to develop a novel automatic image annotation method in the compressed domain. To this end, firstly the images are compressed using the VQ compression method and then are segmented into equal rectangular regions. Each region in the labelled image will be assigned a visual weight that will be calculated. In the annotation process, the relevance model which is a joint probability distribution of the word annotations and the image regional and global features vector is computed through the training set. Therefore, the unlabelled images are annotated. Finally, the image is retrieved on the basis of its semantic concepts. The experiments over 5k Corel images have shown that the retrieval performance of the method suggested here is higher than that of other methods in the uncompressed domain.

Keywords: Semantic-based image retrieval, Image annotation, Vector quantization, Compressed domain.

1. Introduction

Image retrieval has been a very active research area since the 1970s. The text-based image retrieval can be traced back to the late 1970s. A very popular framework of image retrieval then was to first annotate the images by text and use text-based database management systems (DBMS) to perform image retrieval. In this way, users employed the keywords they had in mind to access the related images [1]. This was called the text-based image retrieval. Users of text-based DBMS faced several major problems. First, annotating images required a lot of time and effort, and it also greatly depended on how the operator interpreted the images. The second major problem was that different users would see different concepts in the same image, and therefore the annotations added to the images did not cover the whole scope of the inquiry. As a result, text–based inquiries were not sufficiently thorough or expressive [1, 2].

In the early 1990s, with the increase in the volume of images in the databases such as the internet and also for the purpose of overcoming the difficulties faced in using text-based systems, the extension of CBIR systems was proposed, the task of which was the automatic feature extraction of images through the employment of visual concepts. The user would select one or more of the visual features and determine ranges for their values. Then the retrieval process was carried out on the basis of the provided information. Most CBIR systems perform feature extraction as a preprocessing step. Once obtained, visual features act as inputs to subsequent image analysis tasks such as similarity estimation [1, 3].

The initial CBIR techniques usually use a color histogram to represent each image in the database. The first commercial system that used this technique was QBIC. One problem with such color-based retrieval techniques is that they consider the color distribution of pixels while ignore spatial relationships among pixels. This leads to lower retrieval performance because very different images can have similar color histograms [1].

* Corresponding Author. Email: Eftekhari@qiau.ac.ir

Most of CBIR methods are introduced in the uncompressed domain, but the internet and the wireless systems of communication, storage, processing, and transmitting multi-media data (such as texts, speech, voice, and videos) have greatly progressed. The storage space required and the time spent for transmitting data depend on the size of the data. Techniques of compressing data developed for creating information require little space, thus memory and transmission costs are reduced while quality is retained as much as possible [5-7].

The VQ is a simple and efficient program for compressing images. This method is inherently an indexing technique, therefore it has turned into a promising method for combining image compression and indexing techniques. As compressed data obtained from the VQ can be directly mapped using pixel patterns, indexing and retrieval based on compressed VQ images can capture the meanings and features of the images. Several methods based on different techniques of VQ compression have been introduced, for example, in [5, 9], two indexing and image retrieval programs based on the histogram of the indices obtained from the VQ are presented, and in [5, 6] a method based on the indices generated from IC-VQ is introduced [8]. The proposed scheme extracts image features based on the relationship between the indices of IC-VQ compressed images. The features extracted represent the content of images (and this vector is resistant to scale, rotation, transmission, and image size.) The reason for the effectiveness of this method is that the inter block and intra block relationships among image blocks are used to create a diagram of the continuous region. In [9], a codebook is used for the image retrieval and indexing of colored images. In conclusion, it was found out that the VQ technique is robust for compressed as well as uncompressed images.

Experimental results of content-based systems in the compressed and uncompressed domains indicate that low-level concepts are often unable to describe the high-level semantic concepts of the human brain. Therefore, a retrieval system was introduced which was called the semantic retrieval method employing a combination of text based systems and content based systems. Most methods used in the semantic retrieval make use of the strategy of automatic or semi-automatic image annotation. The usual reason for annotating images through the use of keywords is to facilitate access to the images. In recent years, many algorithms have been introduced for AIA [10].

The main idea of AIA techniques is to automatically learn semantic concept models from a large number of train images, and use the concept models to label test images. As with text document retrieval, images annotated with semantic labels can be retrieved by keywords [10-12]. We can classify them into 4 categories: vector space models, classification methods, graph-based methods, and statistical models [13].

Statistical techniques comprise the most popular method. As the basic principle behind them is to estimate the probability of the words related to the query made by the user, many statistical methods related to the automatic image annotation have been introduced in recent years. The Co-occurrence method introduced by Mori et al. in 1999 [14] was the first statistical method of annotating images. In this method, images are divided into equal rectangular segments. Then following the extraction and clustering of the texture and color features of each segment, the probability of having one word for each cluster is obtained, and the images are annotated using the probabilities. Duygulu et al. (2002) also proposed the machine translation model which is a region-based method [13, 14].

Afterwards, Jeon et al. (2003) improved the results of Duygulu et al. (2002) by introducing a generative language model for image annotation, referred to as the cross-media relevance model (CMRM) [15]. Yet, criticizing the last two models mentioned above, Lavrenko et al. (2003) argued that the process of quantization from continuous image features into discrete blobs, which is done in the machine translation model and the CMRM model, causes the loss of useful information in image regions. In another study, Monay and Gatica-Perez (2003) applied the probabilistic latent semantic analysis (PLSA) (Hofmann, 1999) to the image auto-annotation. Both text words and image features were treated as terms [10]. Moreover, Pan et al. (2004) proposed a series of auto-annotation methods which capture the association between words and blobs (Duygulu et al., 2002) through their pattern of occurrence over the entire training set [10]. Carbonetto (2004) [16] proposed a method which considered the spatial relationships between the regions of the image, and Careneiro (2007) introduced an approach which estimated the distributed semantic categories. Finally, in 2009 an extended CMRM method was put forward which combined the global and the regional features together with the subjective ones in order to estimate the probability of relating them for the purpose of describing the meaning of images [17].

This study also aims to introduce a new method for automatic image annotation in the VQ compressed domain. In this method, images are first compressed using the VQ compression method. Then in the training stage, the images are segmented into equal rectangular regions, and the regional features are extracted and weighted. In the annotation process, the joint probability distribution of the word annotations and the image regional and global features vector are computed using the training set. Afterwards, the unlabelled images are annotated. At the end, the image is retrieved on the basis of its semantic concepts.

The organization of the paper is as follows. In Section 2, the VQ compression method is briefly described. The proposed algorithm is explained in Section 3. In Sections 4 and 5, the experimental results and the concluding remarks are presented, respectively.

2. Vector Quantization

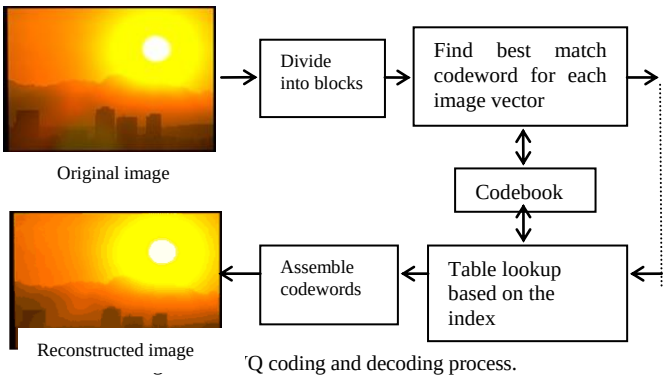
VQ has been used for image compression for many years. In this section, we will briefly review the basic concepts of the VQ image compression [8].

A vector quantization can be defined as a mapping Q of K- dimensional Euclidean space R^k into a finite subset Y of R^k, that is:

$$Q: R^K \rightarrow Y \quad (1)$$

Therefore, a VQ maps a block of the pixels (normally 4×4 pixels in size) of an image on a numerical value. This increases the efficiency of the method. In (1), Y is the set of vectors which are reproduced. There is a single codebook in the encoder and the decoder which is built by using the LBG algorithm. Its inputs include combinations of the pixels present in the block [18-20].

In the encoder, each input vector x belonging to R^k is adapted to a code word in the codebook and then the index of this code word is transferred to the data vector. To find the best code word adapted to a data vector, a distance metric such as the Euclidean or the Manhattan distance can be used. In the decoder, the index of the code word is employed to represent the original vector of the data. By replacing an index for a feature vector, the image describes the features of the blocks of the image in the form of a file of indices [18].



'Q coding and decoding process.

This method can express the relationship between the pixels of each block. Therefore, it can be used to extend effective projects of indexing and retrieving images [11-15]. Figure 1 illustrates the process of generating an index for an image block (vector) [8].

3. The proposed scheme

Equations in the statistical models use the labelled training set to show the relation between the visual features and the keywords. In these methods, the common probability distribution of the image regions features and the keywords in the training images are obtained in the training step and then the training images are annotated. In the testing step, an unlabelled image is given to the model, and the distribution of common probability among the visual features and each of the keywords in the word set is obtained using the knowledge developed by the system in the previous stage. In the end, the test images are automatically annotated. In general, several statistical methods have been introduced which are based on the image regions or the entire image [10, 13].

In this paper, the characteristic of the statistical methods of annotating images is used for annotating and retrieving compressed images. In this method, first the training images for building the codebook VQ are selected, and then the codebook is built on the basis of the LBG algorithm. For the process of annotation, the images in the database are divided into the two groups of labelled training images J and unlabelled test images Q. The training images are compressed using the codebook, and then the compressed training images are divided into equal rectangular segments. After that, the indices equivalents to each region are extracted as features. Finally, the features extracted from the compressed images are weighted as explained below.

3.1. Calculation of the Features Weight

To calculate the weights of the indices, first the probability of the occurrence of each index in each region is obtained by using the following formula:

$$IF_{ir} = \frac{N_{ir}}{\sum_{i=1}^{n} N_{ir}} \quad (2)$$

where N^{ir} is the number of the i^{th} index in the r^{th} region, and the denominator of the fraction is the sum of all the indices of that region. The probability of the occurrence of the index in all the regions is calculated as follows:

$$RF_i = \frac{R_i}{|R|} \quad (3)$$

where |R| is the total number of trained regions and R_i is the number of the regions where the i^{th} index is located. Results obtained from these two formulas are multiplied by each other, and thus the weight of each index in each region is determined:

$$weight_{ir} = IF_{ir} \times RF_i \quad (4)$$

3.2. Automatic Image Annotation

At the stage of the automatic annotation, we can calculate the joint probability distribution of the word annotations and the image regional and global features vector using the training set, and then combine all the concepts obtained to describe the unlabelled image.

We assume J is the set of labelled training images and Q the set of unlabelled test images. Each training image is labelled with one set of rectangular regions $I_J = R_I = \{r_1, r_2, ..., r_n\}$. We also have the set of words corresponding to I_J: $W_M = \{w_1, w_2, ..., w_m\}$. A priori

probability of the words is obtained by using the following formula:

$$P(w_m) = \frac{E_w}{\sum_{m=1}^{M} E_{wm}} \quad (5)$$

where E_w is the number of the words W_m in the training set, and the denominator of the fraction is the total number of training words [5].

Then we take T to be a test image. It will be turned into the equal rectangular segments $T_Q = R_T = \{ r_1', r_2', ..., r_t' \}$ after it is compressed, and the regions obtained will be weighted as previously explained. To combine the regional and global features in the proposed method, the histogram of the VQ indices of all test (H_T) and training images (H_I) are calculated .

Afterwards, the probability of the presence of all the words W_M in the image is calculated as shown below, and the image is labelled to the word which has the highest probability to be present in the image [5, 6].

$$P(W, R_T, H_T|I) = \sum_{I \in J} P(I)' P(W|I)' P(H_T|I) \prod_{i=I}^{t} P(r_i'|I) \quad (6)$$

The prior probabilities $P(I)$ can be kept uniform over all images in I, then $P(W|I)$ the probability of drawing the word from the model of image I is 0 or 1. To calculate the probability of any similarity between the training and the test images, with global feature $P(H_T|I)$ and regional feature $P(R_T|I)$, we can do as follows:

$$P(H_T|I) = D_{KL}(H_T|H_I) = \sum_{i=I}^{n} P(a_i) \log\left(\frac{P(a_i)}{P(b_i)}\right) \quad (7)$$

$$P(r'|I) = P(r'|R_I) = \underset{i=1}{\overset{n}{arg\,max}}\, CosSimilarity(r'|r_i) \quad (8)$$

$$CosSimilarity_{r_a, r_b} = \frac{r_a . r_b}{|r_a|' |r_b|} \quad (9)$$

We now use the following equation to obtain the words which have greater probabilities:

$$m^* = \underset{m=1}{\overset{m}{arg\,max}} (P(W_m, R_T, H_T|I)' P(W_m)) \quad (10)$$

In this equation, $P(W_m)$ is the probability of the m^{th} word in the test images obtained in the previous stage. At the end of the annotating stage, the words with the greatest probabilities are assigned to the related image in order to describe it.

In the stage of retrieving the image by using the key word, the probabilities of the word we have in mind are calculated for all the images. Then, the images obtained on the basis of these probabilities are arranged in the descending order. Finally, the images having the greatest probabilities are retrieved.

4. Experimental results

One of the greatest challenges in assessing the efficiency of image retrieval systems is the lack of a standard benchmark needed for this task. Although many attempts have been made in recent years to overcome this challenge, the problem has not been solved due to the great variety in application areas and in the needs of the users of these systems. Therefore, we have tried to use the known principles and references to build the database used in testing our method [1, 4].

The image database used in this section includes 5000 images of the known set of images in the COREL set provided by Duygulu et al. (2002) which is already separated into a training set with 4500 images and a test set with 500 images. It is made up of 50 main groups, each with 100 images. This set of images comprises different themes, including out of town areas, natural scenery, and objects and animals. Most of the images have 4 word annotations while a few have 1, 2, 3 or 5. The vocabulary size of the whole set is 374 and that of the test set is 263 [21].

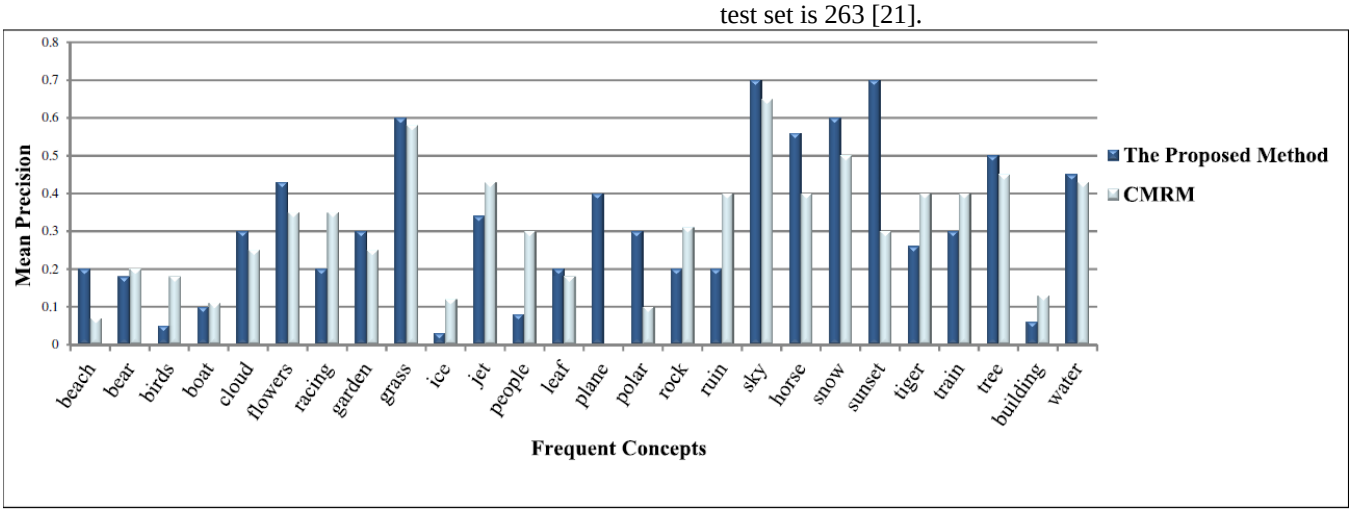


Fig. 2. The precision obtained for some of the words

One hundred different images of the test set were used to find the most suitable sizes of codebooks and codewords to be employed for the process of image compression. After conducting several tests using codebooks and codewords of different sizes, the codebook having the size 1024 and the codewords with the size 4 × 4 were selected. Moreover, the Hue Saturation Value (HSV) color space was employed for the compression process. After the compression process, each image was divided into six equal segments and then the indices in each segment were obtained through the VQ compression method. Finally, the images were annotated and retrieved on the basis of the proposed algorithm.

Now, three measures of mean precision, recall, and F1 are used to evaluate the annotation method:

$$Precision = \frac{c}{n} \tag{11}$$

$$Recall = \frac{c}{N} \tag{12}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{13}$$

In these equations, *c* is the number of words which are correctly annotated with the inquiry word using the proposed method, *n* the number of images annotated with the query word using the proposed method, and *N* the number of images annotated with the query word by the human user [10, 13].

Table1
Comparison of the mean results obtained in the proposed method with those produced in other methods

Methods	**Precision**	**Recall**	**F1-Measure**
The Proposed Method	0.23	0.25	0.239
CRM	0.16	0.19	0.181
CMRM	0.12	0.09	0.111
Co-occurrence	0.02	0.03	0.024

Table 1 reports the measures of precision and recall, and the F1 measure obtained by using the method suggested in the present paper and those produced by other annotation methods of Co-occurrence, CMRM and CRM which are employed in the uncompressed domain. In the proposed method, the obtained mean precision is 0.23, the mean recall 0.25, and the F1 measure 0.239. Thus the proposed method is more efficient than the other methods mentioned above.

Figure 2 illustrates the precision obtained for some of the words in different groups. The results indicate that the proposed method performs better for some words such as sky, water, sun, tree, grass, and in general for natural or outdoor images, but the precision of the method is low for word groups like people, buildings, the marketplace, home appliances, and indoor images. The reason for this low level of precision is the great variety in the color and texture of this category of images. Consequently, the indices obtained after the compression of the images will also be very scattered and different. In turn, the indices of the regions related to these images will have a low, and usually the same, weight. What all these add up to is that recognition of these concepts will be difficult for this system, and that the precision of annotation and retrieval of this category of images will be low. Our goal in future research is to address this shortcoming.

Table 2 shows some test images together with their true, CMRM and predicted annotations, and Figure 3 depicts some of the retrieval results using the keywords "*horse*", "*race",* "*jet",* "*bear*" and *"tiger*" as a single word query.

5. Conclusion

In this paper, a novel semantic automatic image annotation method based on the VQ image compression is devised. The suggested approach extracts both regional and global features from the compressed images. The experimental results show that the method is a good choice for automatic image annotation in the compressed domain. In order to evaluate the efficiency of the proposed method, 5000 images from the Corel database were used. Furthermore, the method was compared with the Co-occurrence, CMRM and CRM methods which are used in the uncompressed domain. Results obtained from this comparison show that the method introduced in this study is more efficient than the other methods.

Fig.3. Semantic retrieval on Corel. Each row shows the top five matches to a semantic query. From the top to the bottom: *"horse", "race", "jet", "bear"* and "*tiger*".

Table 2
Automatic image annotation results

					
Ground Truth	Flower, Leaf, Petals, Stems	Frost, Ice, Sky, Tree	Flower, Garden, House, Window	Field, Foals, Horse, Mare	Castle, Shrubs, Sky, Water
CMRM	Flower, Grass, Field, Tree, Building	Sea, Water, Tree, Grass, Sky	Flower, Grass, Sky, Field, People	Field, Horse, Foals	Sky, Water, Grass, Tree, Stone
The Proposed Approach	Flower, Field, Grass	Sky, Ice, Grass	Flower, Grass, Garden, Tree,	Grass, Horse, Foals, Field	Sky, Water, Tree

References

[1] R.Datta, Djoshi, Jia Li,J.Z. Wang, " Image retrieval: ideas, influences, and trends of the new age" ACM Computing Surveys, vol. 40, no. 2, Article 5, April .2008.

[2] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based image retrieval at the end of the early years", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 , pp. 1349-1380,2000.

[3] Ying Liua, Dengsheng Zhanga, Guojun Lua,Wei-Ying Mab " A survey of content-based image retrieval with high-level semantics " Pattern Recognition Society ,vol.40,pp.262-282, January 2007.

[4] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos,"Supervised learning of semantic classes for image annotation and retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no.3, , pp. 394-410. Mar. 2007

[5] F. Idris, S. Panchanathan, "Storage and retrieval of compressed images", IEEE Trans. Comput. Electron. 41 (3) (1995).937–941.

[6] F. Idris, S. Panchanathan," Algorithms for indexing of compressed images", Proceedings of Visual'96: International Conference on Visual Systems, Melbourne, 1996.

[7] M.K. Mandal, F. Idris, S. Panchanathan, "Image and video indexing in the compressed domain", Proc of SPIE: Multimedia Storage and Archiving Systems II, Vol. 3229, pp. 2–13, Dallas, TX, Nov. 1997.

[8] M. Eftekhari-Moghadam, J.Shanbehzadeh, F.Mahmoudi and H.Soltanian-Zadeh, "Image retrieval based on index compressed vector quantization"Pattern Recognition, vol. 36, pp. 2635- 2647, 2003.

[9] S.Teng, G.Lu, " Image Indexing And Retrieval Based On Vector Quantization" Pattern Recognition ,Vol.40 , pp.3299 – 3316,2007

[10] T. Sumathi, C.Lakshmi Devasena, and M.Hemalatha, " An Overview of Automated Image Annotation Approaches" International Journal of Research and Reviews in Information Sciences,vol. 1, No. 1, March 2011

[11] N. Bassiou , C.Kotropoulos, " RPLSA: A novel updating scheme for probabilistic latent semantic analysis"computer speech and language,vol.25,pp.741-760,2011.

[12] F.Shi, J.Wang, Z.Wang, " Region-based supervised annotation for semantic image retrieval", International Journal of Electronics and Communications (AEÜ),Article in press, March 2011

[13] A. Hanbury, " A survey of methods for image annotation" Journal of Visual Languages and Computing,vol.19,pp.617-627, January 2008.

[14] 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, 1999.

[15] J. Jeon, V. Lavrenko and R. Manmatha. "Automatic image annotation and retrieval using cross-media relevance models" in Proc. SIGIR conf,2003,paper 26 , p.119

[16] M.E. Elalami. "Unsupervised image retrieval framework based on rule base system" Expert Systems with Applications, vol.38, pp.3539-3549, 2011.

[17] Y.Wang, T. Shaogang Gong, X.Sheng Hua," Combining global, regional and contextual features for automatic image annotation" Pattern Recognition,vol. 42,pp.259-266,2009

[18] G. Boopathy, S. Arockiasamy "Implementation of vector quantization for image compression - A Survey "Global Journal Of Computer Science And Technology, Vol. 10 , pp.22-28,April 2010

[19] Daniele Cerra and Mihai Datcu. " Image retrieval using compression-based techniques" in Proc.SCC conf, 2010,paper 19, p.18-21

[20] Mei-Lei Lv, Bei-Bei Liu, Zhe-Ming Lu, " Real-time image semantic retrieval based on VQ"in Proc.PCSPA conf,2010,paper 75,p.281

[21] P.Duygulu, K.Barnard, N.Freitas, and D.Forsyth,"Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary",In The Seventh European Conference on Computer Vision, pp.97–112, Copenhagen, Denmark, 2002.