Scene Retrieval of Natural Images

J.F. Serrano[1], J.H. Sossa[1], C. Avilés[2], R. Barrón[1], G. Olague[3], and J. Villegas[1]

[1] Centro de Investigación en Computación-Instituto Politécnico Nacional (CIC- IPN)
UPLM – Zacatenco, Av. Juan de Dios Bátiz y Othón de Mendizábal s/n Col. Lindavista
C.P. 07738 México, D. F. Tel.: 57296000 ext. 56588
[2] Universidad Autónoma Metropolitana-Azcapotzalco. Departamento de Electrónica,
Av. San Pablo 180, Col Reynosa, México D.F. C.P. 02200, Tel.: 53189550 ext 1026
[3] Centro de Investigación Científica y de Educación Superior de Ensenada, BC.
Km 107 Carretera Ensenada-Tijuana, No. 3918, Zona Playitas, C.P. 22860, México
jfserranotal@gmail.com, hsossa@cic.ipn.mx,
caviles@correo.azc.uam.mx, rbarron@cic.ipn.mx,
gustavo.olague@gmail.com, jvillegas@gmail.com

Abstract. Feature extraction is a key issue in Content Based Image Retrieval (CBIR). In the past, a number of describing features have been proposed in literature for this goal. In this work a feature extraction and classification methodology for the retrieval of natural images is described. The proposal combines fixed and random extracted points for feature extraction. The describing features are the mean, the standard deviation and the homogeneity (form the co-occurrence) of a sub-image extracted from the three channels: H, S and I. A *K*-MEANS algorithm and a 1-NN classifier are used to build an indexed database of 300 images. One of the advantages of the proposal is that we do not need to manually label the images for their retrieval. After performing our experimental results, we have observed that in average image retrieval using images not belonging to the training set is of 80.71% of accuracy. A comparison with two similar works is also presented. We show that our proposal performs better in both cases.

1 Introduction

Nowadays, due the availability of large storage spaces a huge number of images can be found in the Internet. With this huge distributed and heterogeneous image database, people want to search and make use of the images there contained. A great challenge emerges: finding out accurate ways of searching images. Basically, images can be retrieved in two ways, firstly, text based and secondly, content-based or query by example based. Text-based retrieval approaches are very well-known and widely used. In this case users are provided with a text area to enter the key words (usually the image file name) on the basis of which image searching is done. It is widely used in Google web based image searching technique.

The concept CBIR has a main drawback: The images in the database are manually annotated using key words. This is known to be a very time consuming process for any large database [1], [2]. Also retrieval depends on the human perception based text annotation.

E. Bayro-Corrochano and J.-O. Eklundh (Eds.): CIARP 2009, LNCS 5856, pp. 774–781, 2009.

To avoid the above mentioned problems, a second approach, Content-Based Image Retrieval (CBIR) has been proposed by researchers. The term CBIR seems to have originated in the earlier 90´s [1], [4], [5], [6], [10], [12], [14] and [15]).

CBIR includes research on: Automatic Feature Extraction ([2], [3]), Automatic Feature Extraction with a Semantic Content ([4], [5], [6], [9], [10] and [11]) and data representation ([7]). CBIR techniques use low-level features such as texture, color and shape to represent images and retrieves images relevant to the query image from the image database. Among those low level image features, texture features has been shown very effective and subjective [15].

In this paper we describe a CBIR based methodology. In the next section we describe each of the steps composing the proposed approach.

2 Methodology

In this section we describe each of the stages of the proposed methodology for the retrieval of natural images into a database. It involves two basic stages as follows:

Training stage. This stage is divided into two main phases as shown in Fig. 1(a). During the first phase a set of 300 images in RGB format is first read. Each image is converted to HSI format. To each image, 300 pixels are uniformly selected at random (see Fig. 2(a)). Taking each of the 300 points as the center we open a squared window of size of 10×10 pixels around it.

Figure 2(b) shows several examples. To each of the 300 windows the following features are extracted: the mean, the standard deviation [13] and the homogeneity obtained from the co-occurrence matrix [8]. This is done for the corresponding

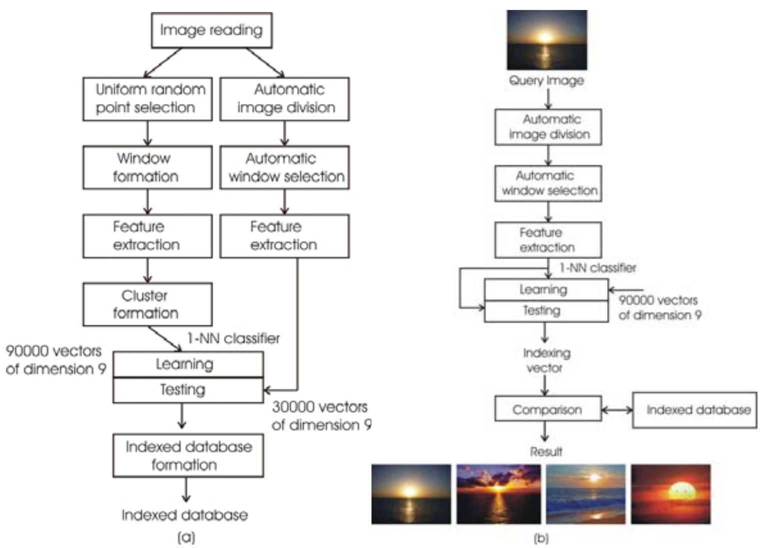

Fig. 1. (a) Flow Diagram for the training stage. (b) Flow diagram of the testing stage.

sub-image channel: hue (H), saturation (S) and brightness (I) of an image. The corresponding describing vector for each window of the image has thus nine components, three for H channel, three for S channel and three for I channel.

We take the resulting 90,000 describing vectors (300 for each of the 300 images) and a *K*-MEANS algorithm is applied to obtain how many of these 90,000 features are divided into six classes. For the 300 images chosen in this paper for training, Table 1 shows how many vectors fall into class one, how many vectors fall into class two, and so on until class six. This gives somehow the probability that a given class belongs to the 300 images.

Fig. 2. (a) For sub-image description 300 image pixels are automatically and uniformly selected at random. (b) For automatically image segmentation around each of the 300 pixels a square window of $M \times N$ is opened. In this figure only 20 points are shown as an example.

Table 1. Distribution of the 90,000 features into the 6 chosen classes

Class number	Number of features per class
1	14,647
2	16,106
3	7,104
4	19,155
5	11,848
6	21,140
	Total 90,000

During the second phase, to the same set of 300 images an automatic partition is performed as shown in Fig. 3(a). As shown in this figure each image is divided into 10×10 regions of 72×48 pixels per region. For each of these 100 sub-images we take a window of 10×10 pixels as shown in Fig. 3(b). To each of the resulting 100 windows, again the same: mean, standard deviation and the homogeneity are computed in the three same channels. Each window is described again in the form of vector of nine components. As a result we have 30,000 vectors (100 for each of the 300 images).

To create the indexed database of the 300 images used for training we proceed as follows. We take the 90,000 describing vectors obtained in the first phase of training and the 30,000 describing vectors obtained in the second phase of training and input them to a 1-NN classifier. As a result we obtain an indexed database containing the following information as shown in Fig. 4.

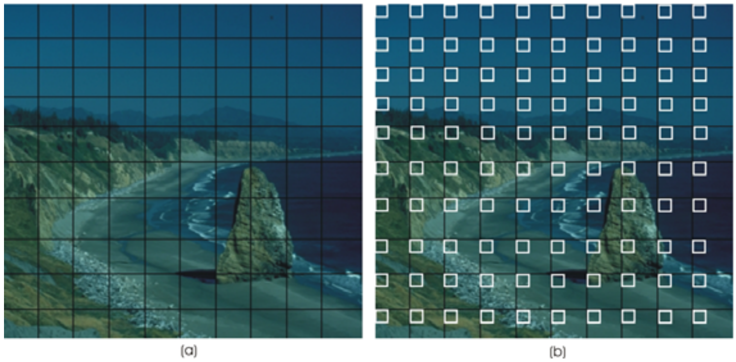

Fig. 3. (a) An image is uniformly divided into 100 sub-images to get 100 describing features. (b) For each sub-images, a window of 10×10 pixels is selected to compute the corresponding describing vector.

C1	C2	C3	C4	C5	C6	→	Name of Image
40	16	23	20	1	0	→	Image 1.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	19	9	9	15	20	→	Image k.jpg
							⋮
7	23	7	32	19	12	→	Image 300.jpg

Fig. 4. Form of the indexed database

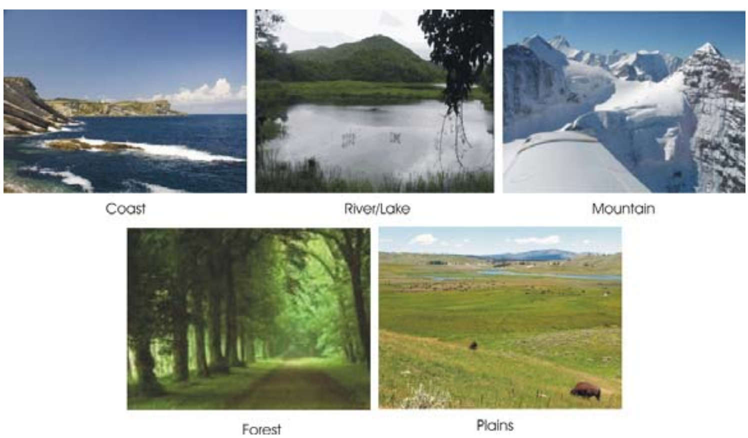

Fig. 5. Five different image classes have been manually chosen for image retrieval purposes

Retrieval stage. This stage is divided into the phases shown in Fig. 1(b). As shown, a query image is presented to the system. To this image the same feature extraction phases used during training are applied. As a result we get 100 describing vectors, These 100 vectors are presented to already trained 1-NN classifier. As a result we just get one indexing vector. This vector contains the probability that each one of the six classes C1, C2, C3, C4, C5 and C6 is contained in the query image. This vector is compared with the 300 vectors saved in the indexed database. To reduce the computing time and to get better retrieval results, we just take into account the two higher components of the six classes. As a distance we use the Euclidean distance. For retrieval purposes we have chosen five different kinds of images as show in Fig. 5. These five different types of images were manually selected.

Note. For testing our proposal we have chosen 300 natural images of the Corel Image Database (720×480). These images were provided by J. Vogel [4], [5], [6] and [15]. The 300 images used for training were grouped into the 5 types of images as follows: 54 mountains images, 54 lakes images, 54 coastal images, 54 forest images, 54 prairies images, and 30 clouds images.

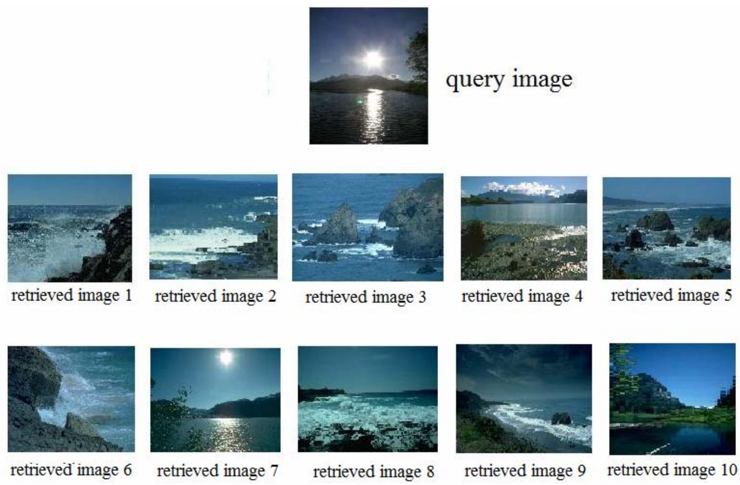

Fig. 6. Images retrieved given a query image of a sunset

3 Experimental Results

In this section we present the experimental results that validate our proposal. For this we have selected from Internet 221 images. These images are different from those used for training. We presented each of these 221 images to the system and asked it to show us the most 10 similar images from the indexed database. Figure 6 shows a query example. From Fig. 6 we can see for example that the system retrieves correctly 9 images and retrieves incorrectly 1 image (image 10). This gives a 90% of efficiency for this retrieval (full test can be shown in figures 7, 8 and 9). To test the efficiency of the proposal we have used the following two measures:

$$P = \frac{\text{No. of relevant images retrived}}{\text{Total no. of images retrieved}} \tag{1}$$

$$R = \frac{\text{No. of relevant images retrived}}{\text{Total no. of relevant images in database}} \tag{2}$$

The first measure represents the number of relevant images retrieved with respect to the total number of images asked to be retrieved. The second measure represents the relevant images retrieved with respect to the total number of images used for training for a given class.

Fig. 7 shows the performance of our proposal against the method described in [14]. As we can appreciate our proposal performed a little better.

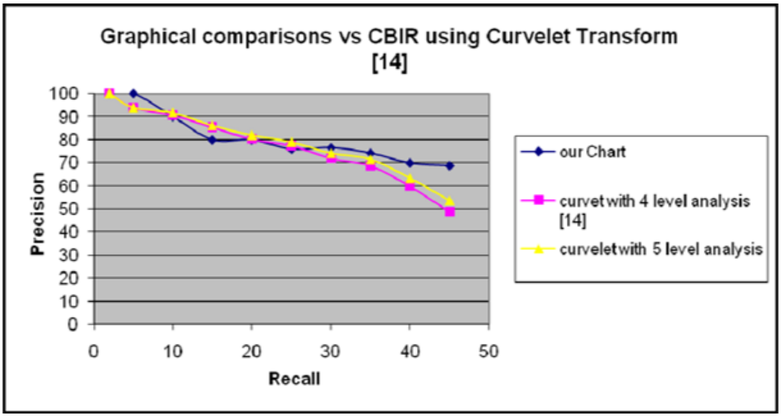

Fig. 7. Performance of our proposal against the method described in [14]. We get a 79.05%, while in [14] they get a 77.71% of efficiency when using as a query the coastal image shown in Fig. 6.

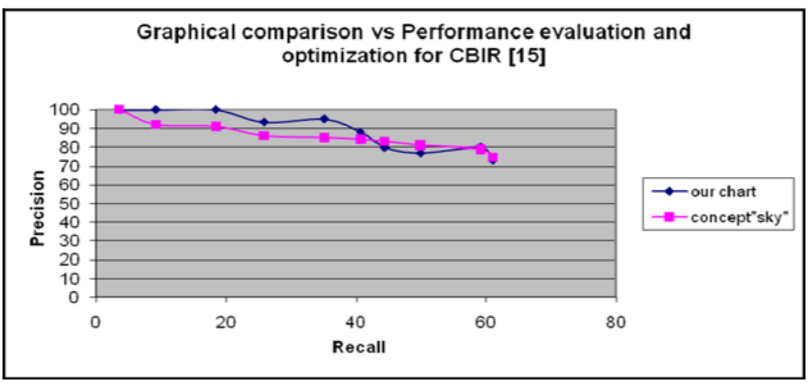

Fig. 8. Performance of our proposal against the method described in [15]. We get a 85.93%, while in [15] they get a 85.61% of efficiency when using as a query a red sunset image.

In Figures 8 and 9, we compare our proposal against the method reported in [15]. As can be seen the performance of our proposal is just a little better than the one reported in [15].

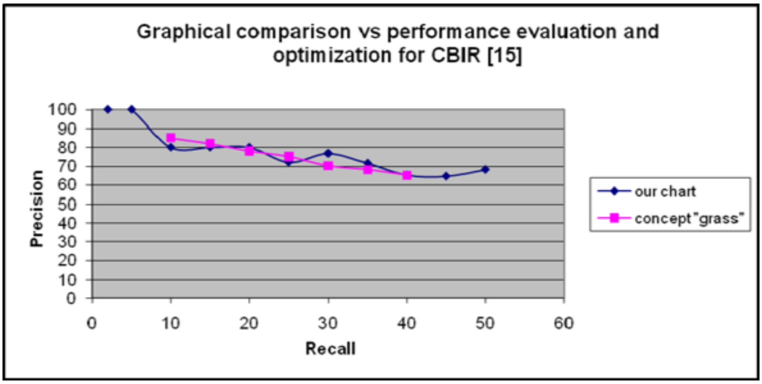

Fig. 9. Performance of our proposal against the method described in [15]. We get a 77.16%, while in [15] they get a 74.17% of efficiency when using as a query the forest image.

4 Conclusions

In this paper we have described a methodology that allows to automatically retrieving natural images from a database. During learning the proposal takes as input a set of images divided into five classes: coasts, lake/rivers, mountains, forests and plains. It extracts from them describing features from sets of points randomly and automatically selected. A *K*-means classifier is used to form six different clusters from the describing features obtained from the randomly and automatically chosen points. A 1-NN classifier is used to build an indexed database from the combination of all the describing vectors.

During retrieval the already trained 1-NN classifier is used to retrieve from the indexed database the most similar images given a query image. The experimental results show that our proposal performs better than two reported method in the literature. For this we have used the precision/recall measure.

Nowadays we are testing the proposal with more images and with more types of image classes and with more cluster regions. Also we are trying to use interest point detectors to select the points from which the describing vectors are going to be computed. We are also going to test with other describing features and other classifiers.

Acknowledgments. This work has been supported by the National Polytechnic Institute of Mexico (SIP-IPN), under grants 20090620 and 20091421, by the Mexican Science and Technology National Council (CONACyT), under grant 182938 and the European Commission and CONACYT under grant FONCICYT 93829.

References

[1] Long, F., Zhang, H.J., Feng, D.D.: Fundamentals of Content Image retrieval. In: Feng, D. (ed.) Multimedia Information Retrieval and Management. Springer, Heidelberg (2003)

[2] del Bimbo, A.: A Perspective View on Visual Information Retrieval Systems. In: Workshop on Content Based Access of Image and Video Libraries, vol. 21, pp. 108–109. IEEE, Los Alamitos (1998)

[3] Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys 40(2), paper 5 (April 2008)

[4] Vogel, J., Schiele, B.: Semantic Modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision 72(2), 133–157 (2007)

[5] Vogel, J., Schiele, B.: Semantic Modelling of Natural Scenes for Content-Based Image Retrieval. Int. J. of CV (2006), doi:10.1007/s11263-006-8614-1

[6] Vogel, J., Schwaninger, A., Wallraven, C., Bülthoff, H.H.: Categorization of natural scenes: local vs. global information. In: Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV 2006), June 2006, pp. 33–40. ACM Press, New York (2006)

[7] Gonzalez Garcia, A.C.: Image retrieval based on the contents. PhD Thesis. Center for Research in Computing (CIC)-IPN, Mexico DF (September 2007)

[8] Presutti, M.: Co-currency Matrix in Multispectral Classification: Tutorial for Educators textural measures. In: The 4th day Educacao em Sensoriamento Remote Ambito not do Mercosul, Sao Leopoldo RS. Brazil, Augest 11-13 (2004)

[9] Rui, Y., Huang, Th.S., Chang, Sh.F.: Image Retrieval: Currente techniquies, Promissing Directions, and open Issues. Journal of Visual Communication and Image Representation 10, 39–62 (1999)

[10] Liu, Y., Zhang, D., et al.: A survey of Content-Based Image Retrieval with high-level semantics. Pattern Recognition 40, 262–282 (2007)

[11] Li, J., Wang, J.Z.: Real-Time Computerized Annotation of Pictures. In: Proceedings of the 14th annual ACM international conference on Multimedia, pp. 911–920 (2006)

[12] Hiremath, P.S., Pujari, J.: Content Based Image Retrieval using Color, Texture and Shape features. In: 15th International Conference on Advanced Computing and Communications, pp. 780–784 (2007)

[13] Fukunaga, K.: Introduction to statistical Pattern Recognition. Academic Press, New York (1990)

[14] Sumana, I.J., Islam, M.M., Zhang, D., Lu., G.: Content Based Image Retrieval using curvelet transform. In: 2008 IEEE 10th Workshop on Multimedia Signal Processing, October 8-10 2008, pp. 11–16 (2008)

[15] Vogel, J., Schiele, B.: Performance evaluation and optimization for Content-Based Image Retrieval. Pattern Recognition 39(5), 897–909 (2006)