Image Retrieval Using Efficient Region-based Matching

*Mann-Jung Hsiao
Dept. of Computer Science and Eng.
Tatung University
hsiao@mis.knjc.edu.tw

Tienwei Tsai
Dept. of Info. Management
Chihlee Institute of Tech.
twt@mail.chihlee.edu.tw

Te-Wei Chiang
Dept. of Accounting Info. System
Chihlee Institute of Tech.
ctw@mail.chihlee.edu.tw

Yo-Ping Huang
Dept. of Electrical Eng.
National Taipei Univ. of Tech.
yphuang@ntut.edu.tw

***Abstract*—Identifying regions of interest (ROI) plays a vital role for humans to find desired images in content-based image retrieval (CBIR). To enhance the performance of CBIR systems, we propose a simple but generally effective model to express ROI rather than pursuing sophisticated image analysis techniques. Our approach partitions an image into a number of regions with fixed absolute locations. User can formulate a query by selecting the interesting regions in the image. Candidate images are then analyzed, by inspecting each region in turn, to find the best matching region with the query region. Experimental results show that the presented model is generally effective and particularly suitable for images with regions having features which significantly differ from the global image features.**

***Index Terms*—Content-based image retrieval, regions of interest, region-based image retrieval, discrete cosine transform.**

I. INTRODUCTION

The emergence of multimedia, the availability of large digital archives, and the rapid growth of the world wide web (www) have recently attracted research efforts in providing tools for effective retrieval of image data based on their content. Such retrieval is known as content-based image retrieval (CBIR). CBIR is a complex and challenging problem spanning diverse algorithms all over the retrieval processes including color space selection, feature extraction, similarity measurement, retrieval strategies, relevance feedback, etc. Some general reviews of CBIR literature can be found in [2][3][10][14][15]. Smeulder et al. reviewed more than 200 references in this field [14]. Datta et al. studied 120 of recent approaches [2]. Veltkamp et al. gave an overview of 43 content-based image retrieval systems [15]. Deselaers et al. presented an experimental comparison for a large number of different features [3]. Liu et al. provided a comprehensive survey of the recent technical achievements in high-level semantic-based image retrieval [10]. Although various CBIR techniques have been established and good performance results were demonstrated, there are still many problems not satisfactorily solved.

One of the general open problems is the gap between the low-level visual features and human semantic interpretation of an image. To narrow down this gap, recently some approaches have been proposed to bridge the semantic gap. In general, these methods can be classified into three categories: 1) relevance feedback, 2) high-level semantic features, and 3) region-based image retrieval (RBIR) [5]. Relevance feedback is used to learn users' intentions, which has been proved effective in some cases [4]. Yet, the feedback information is typically used only to re-weight the features used within a global similarity measure. Semantic features are used to capture high-level concepts from low-level features. It often requires the semantics of the image database be pre-defined by domain experts [13]. RBIR tends to search the interested regions that closed to the query target, instead of the whole images [8]. However, the segmentation algorithms are

complex and computation intensive and the segmentation results are often not correct. To solve this, some approaches break images into a fixed number of regular rectangular regions. Rudinac et al. partitioned images into 4x4 non-overlapped regions and 3x3 overlapped regions [12]. It is expectable that using more regions better results may be produced but the execution speed becomes unsatisfactory slow for a large database. Amir et al. divided images at a coarser granularity level in a video retrieval system, using a fixed 5-region layout (4 equal corner regions and an overlapping center region of the same size) [1]. Our former study has proposed a region-based approach to solve the above problems by dividing each image into five regions, which is similar to Amir's layout [7]. However, the relevance between a query image and candidate images is evaluated by comparing the regions of the same position. To further improve the retrieval performance, this approach first lets the user select the region of interest (ROI) in the query image to express his/her intentions. Candidate images are then analyzed, by inspecting each region in turn, to find the best matching region with the query region. In other words, the distance between the query image and a candidate image is the smallest distance between the query region and five regions within the candidate image.

The feature extraction method for each region is another important issue. Being an elementary process, the feature extraction will be invoked very frequently; therefore, it should be time-efficient and accurate. To reduce the processing time, we employed Discrete Cosine Transform (DCT) to extract the main features of regions. The DCT has been proved successful at de-correlating and concentrating the energy of image data. It has brought on the proliferation of visual data stored in the JPEG and MPEG compressed formats. This has made some significance influence on the image retrieval research and application [6]. In our approach, an image is first converted to YUV color space and then transformed into DCT coefficients for each region. A block size of 4x4 DCT coefficients in the upper-left corner constitutes the feature vector of a region. Note that the feature vector is further categorized into four groups to express its average grayness and three directional texture characteristics: vertical, horizontal and diagonal. In our approach, a friendly user interface is employed for user to express his/her personal view of perceptual texture properties for the ROI. The experimental system shows that this approach is generally effective and particularly suited for images with regions having features which significantly differ from the global image features.

This paper is organized as follows. The next section introduces region of interest and segmentation. Section III illustrates the feature extraction method. The similarity measurement is presented in Section IV. Section V presents experimental results. Finally, conclusions are drawn in Section VI.

II. REGION OF INTEREST AND SEGMENTATION

Let us consider the following example. You may want to look for images containing a similar region/object in any position as in the query image, which is defined as region of interest (ROI) in CBIR. This leads to a number of solutions that do not treat the image as a whole, but rather deal with regions within an image [9][11]. The problem is discussed as follows.

A. Region of interest

Existing CBIR can be categorized into two major classes, namely, global methods and localized methods [9]. Global methods exploit features from the whole image and compute the similarity between images while local methods extract features from a region (portion) of an image and compute the similarity between regions. For the CBIR task that the user is only interested in a portion of an image, it is defined as localized content-based image retrieval [11]. In localized CBIR, an image is segmented into regions. However, it is hard to locate the ROI in the image when the interested object/region occupies only a small part of the image or the image background has dominant impact on the feature extraction. The most direct way to solve the problems is to let the user select a ROI while conducting a query, which is used in our approach.

B. Segmentation

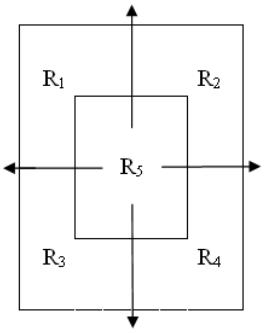


Fig 1.The five rectangular regions used in our approach.

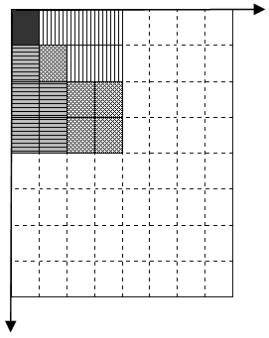


Fig. 2. The upper left DCT coefficients used in our approach: (a) DC, (b) vertical texture feature, (c) horizontal texture feature, and (d) diagonal texture feature.

Though many automatic segmentation algorithms were proposed in localized CBIR, they are often too complex and computation intensive and the retrieval results are often not correct. To solve this, some approaches break images into a fixed number of regular rectangular regions. It is expectable that using more regions better results may be produced but the execution speed becomes unsatisfactory slow for a large database. In our approach, segmentation into homogeneous regions is obtained by dividing the image into four non-overlapping regions and one central region with the same size as others, which is similar to the layout of the IBM TRECVID video retrieval system [1]. Figure 1 illustrates the five rectangular regions used in our approach.

III. FEATURE EXTRACTION

Image contents can be defined at different levels of abstraction. At the first lowest level, an image is a collection of pixels. Pixel level content is rarely used in retrieval tasks. The raw data can be processed to produce numeric descriptors capturing specific visual characteristics called features. The most important features for image databases are color, texture and shape. In general, a feature-level representation of an image requires significantly less space than the image itself. Some transform type feature extraction methods can be applied to reduce the number of dimensions, such as Karhunen-Loeve (KLT), discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet transform (DWT), etc. Among these methods, DCT has been known for its excellent energy compacting property. It has received a great deal of attention and is widely used in image compression. For most images, most significant DCT coefficients are concentrated around the upper left corner; the significance of the coefficients decays with increased distance. The DCT techniques can be applied to extract dominant directional texture features from images, where the DC coefficient (V_1) represents the average energy of the image and all the remaining AC coefficients contain three directional feature vectors: vertical (V_2), horizontal (V_3), and diagonal (V_4). To ease the computation load, only a block size of 4x4 is considered in our approach, as shown in Fig. 2.

Before the feature extraction process, the images have to be converted to the suitable color space. There are some existing color models to describe images, known as color spaces, such as RGB, HSV, HIS, YUV, etc. RGB is perhaps the simplest color space for people to understand because it corresponds to the three colors that the human eyes can detect. However, the RGB color model is unsuitable for similarity comparison. The luminance and saturation information are implicitly contained in the R, G, and B values. Therefore, two similar colors with different luminance may have a large Euclidean distance in the RGB color space and are regarded as different.

In our approach, the YUV color space is used for two reasons: 1) efficiency and 2) ease of extracting the features based on the color tones. Psycho-perceptual studies have shown that the human

brain perceives images largely based on their luminance value (the Y component), and only secondarily based on their color information (the U and V components); therefore, only the Y component is used in our approach. After an image is converted to the YUV color space, it is equally divided into four rectangular regions and one additional central region. Then the DCT is performed over the Y component for a whole image (global features) and five regions (regional features). Therefore, an image is represented by one global feature and five regional features, each of which is constituted by a block of 4x4 DCT coefficients. As a result, only 80 DCT coefficients are needed for each image. In addition, the importance of each directional feature in a region is also taken into account. The user can give different weight for each texture feature based on their perceptions for a region.

IV. SIMILARITY MEASUREMENTS

In CBIR systems, image features are in general organized into *n*-dimensional feature vectors. Thus the query image and the database images can be compared by evaluating the distance between their corresponding feature vectors. It is hard to define a distance between two sets of feature points such that the distance could be sufficiently consistent with a person's concept of semantic closeness of two images. Therefore, there are very few theoretical arguments supporting the selection of one distance over the others; computational cost is probably a more important consideration in the selection. To exploit the energy preservation property of DCT, we use the sum of squared differences (SSD) as the distance function. Using a simpler distance on lower dimensional features means that computation can be saved both in the evaluation of distance and in the number of comparisons to be performed. For the regional search, similarity measurement is performed based on region similarity. For the global search, the whole image is regarded as a "large" region. The distance function is defined as follows.

Let Q and X denote the query image and a database image, respectively. V_k is the k-th feature vector of an image or a region (In our approach, k = 1 to 4). C_n is a vector component in V_k. Assume the distance d_k is the distance between the k-th feature vector V_k^q of the ROI (e.g., the i-th region R_i^q) in Q and the k-th feature vector V_k^x of the j-th region in X. Then,

$$d_k = d(R_i^q, R_j^x) = \sum_{C_n^q \in V_k^q, C_n^x \in V_k^x} (C_n^q - C_n^x)^2. \qquad (1)$$

Similarity is evaluated as a weighted aggregation of image features. W_k is the weight assigned to the k-th feature vector in a query to express its importance. Thus the overall distance between two regions is:

$$D(R_i^q, R_j^x) = \sum_{k=1}^{4} w_k d_k. \qquad (2)$$

When the user selects an interest region R_i^q in the query image Q and issues a query, candidate images are hence analyzed, by applying equations (1) and (2) for each region in turn, to find the best matching region in an image X, which having the smallest distance with the query region. The distance between Q and X can thus be defined as:

$$D(Q, X) = \underset{j=1 \text{ to } 5}{Min}(D(R_i^q, R_j^x)). \qquad (3)$$

V. EXPERIMENTAL RESULTS

The conventional computer vision recognition-based task looks for the object to be searched with as small and as accurate a retrieved list as possible. But in CBIR, the goal is to extract as many "similar" objects as possible, the notion of similarity being very loose as compared to the notion of exact match. To evaluate our work, an experimental CBIR system has been implemented with a general-purpose image database including 1,000 color images, which was downloaded from the WBIIS database [16]. Unlike recognition-based systems, CBIR systems require versatility and adaptation to the user, rather than the embedded intelligence desirable in recognition tasks. Therefore,

design efforts in our CBIR system are devoted to combine light computation, great flexibility and friendly user interface.

The ROI capabilities in the system, allowing the expression of an interested region in the query image, are highly appealing to capture a certain level of semantics and can be used much in the same way as words. For example, the system will accept queries like “Find me all the images containing the contents as in the upper left region of the query image.” To accomplish this, the user can select any one of the regions (upper left, upper right, lower left, lower right, and center) for regional search. For those query images without clear objects, the user can select the option “whole” for global search. For a single region query (i.e., regional search), there are two options” “same” and “any”, which means the ROI of the query image is compared with the “same” or “any” region of the candidate images. Figure 3 is the main screen of our system, where the user can specify the ROI, adjust the weight of each feature, and inspect the retrieved results.

After the user loads a query image and selects the ROI, the system first initializes a set of uniformly distributed weights for features. Then the user’s specific information needs can be described by adjusting the weight of each feature. The search consists of the comparison of the ROI against all of the predefined regions of candidate images. The similarity is computed on the basis of weights, and retrieval results are displayed to the user. The retrieved images are the top ten in similarity, ranked in the ascending order of the distance to the query image from the left to the right and then from the top to the bottom. An image is deemed as a “correct” retrieval if it contains objects similar to the query, as judged subjectively.

Several queries are conducted to examine the retrieval quality. Because the retrieval performance is subjective to users, the query targets are manually picked from some commonly accepted categories. For the first query, a scene of the declining sun is used as the query image. The user selects “upper left” as the ROI. Fig. 4(b) shows the results for the option “same” while Fig. 4(c) shows the results for the option “any”. It can be seen that the number of “correct” images is improved from 6 to 8.

For the second query, a round ball is used as the query image. The user selects “upper left” as the ROI, and “any” as the regional option. From Fig. 5, it is observed that a number of balls at other regions are also collected in the output list. An image of a mountain scene is given for the third query. Since no obvious object/region appears in the image, the user cannot pick the region which is perceptually meaningful as the ROI. Thus, the user selects option “whole” to conduct a global search. Figure 6 gives a very promising result.

Since one cannot expect results obtained in response to a query to be fully satisfactory, the system allows a form of interaction by adjusting the weight of each feature or choosing a different region to improve the quality of retrieval. For expressing users' perceptions on each individual feature, a weighting vector W in form of (w_1, w_2, w_3, w_4) is used to indicate significant levels for grayness, vertical texture, horizontal texture, and diagonal texture, respectively. In our system, visual interface are employed in order to ease the task of resubmitting queries again and again. The fourth query is used to examine the power of the weighting vector. Figure 7(b) is the results for the initial vector W=(1,1,1,1). The user might try to eliminate the horizontal texture feature by giving a vector of (1,1,0,1). Figure 7(c) shows a better result with more interesting images in the list.

VI. DISCUSSION AND CONCLUSION

It is generally agreed that one of key challenges in CBIR is how to reduce the semantic gap between user expectation and system support, especially in nonprofessional applications. To find a model to enhance the intelligence of CBIR systems, some researchers study in sophisticate image analysis and retrieval techniques to identify the images that contain the query object, but the performance is limited and only appropriate for narrow domains such as trademarks, textiles, etc.

In practice, the ROI is easy to observe but hard to isolate through automatic region analysis. To solve this problem, we propose an efficient region-based approach that provides the ROI capabilities, allowing the expression an of interested

region in the image. In this approach, queries are formulated at a simple semantic level. The system will accept queries like “Find me all the images containing the contents as in the upper left region of the query image.” To accomplish this, a friendly user interface is provided in the system. After all, it is the user, being in the retrieval loop, analyzes system responses, refines the query, and determines relevance. This implies the need for intelligence and reasoning capabilities inside the system can be reduced. We built an experimental system to demonstrate the effectives of our approach. It is observed that the use of ROI rather than the entire image increases the retrieval performance.

According to these results, one important fact worth mentioning is that the ultimate end user of the CBIR system is human, and the image is inherently a subjective medium, that is, the perception of image content is very subjective, and the same content can be interpreted differently. Therefore, users may use different search criteria for the same query image. This human perception subjectivity has different levels to it: one user might be more interested in a different dominant feature of the image from the other, or two users might be interested in the same feature (e.g., texture), but the perception of a specific texture might be different for the two users. To tackle this problem, our system provides a set of weights to characterize the relative importance of the features in a query image. From another point of view, each weight can be regarded as the fuzziness of the cognition to the associated feature. The user can emphasize the features that are relatively important based on his/her interests. Though it is still difficult to translate “soft” feelings into “hard” values based on human perceptions, it does play an important role in the multiple passes of refining the retrieval. The experimental results indicate that by adaptively adjusting the weighting vector, the retrieval performance can be further improved.

Several query examples have shown that our approach is particularly useful in qualitative query. It is especially suited for images with regions having features which significantly differ from the global image features. However, only a positive impression of the abilities of our approach is given in this paper. In the future, a quantitative performance evaluation will be given to examine retrieval quality more intensively. In addition, we can also find that the color tones of the retrieved images are not always similar to that of the query image even though they are similar from the viewpoint of texture or shape. This is because only Y-component is used in the feature vector. Our future work includes exploring the U and V components to further improve the retrieval performance.

* M.-J. Hsiao is currently an instructor at Kang-Ning Junior College of Medical Care and Management.

REFERENCE

[1] A. Amir, M. Berg, S.-F. Chang, W. Hsu, G. Iyengar, C.-Y. Lin, M. Naphade, A. Natsev, C. Neti, H. Nock, J. Smith, B. Tseng, Y. Wu, D. Zhang, “IBM research Trecvid-2003 video retrieval system,” Proc. of NIST TrecVid 2003, Nov. 2003.

[2] R. Datta, J. Li and J. Z. Wang, “Content-based image retrieval - approaches and trends of the new age,” Proc. of the ACM Int. Workshop on Multimedia Information Retrieval, pp.253-262, 2005.

[3] T. Deselaers, D. Keysers, and H. Ney, “Features for image retrieval: an experimental comparison,” Information Retrieval, Vol. 11, No. 2, pp.77-107, Apr. 2008.

[4] J. Guan and G. Qiu, “Learning user intention in relevance feedback using optimization,” Proc. of ACM Int. workshop on multimedia information retrieval, pp.41-50, 2007.

[5] R.-B. Huang, S.-L. Dong, and M.-H. Du, “A semantic retrieval approach by color and spatial location of image regions,” Proc. of the IEEE Congress on Image and Signal Processing, Vol. 2, pp.466-470, May 2008.

[6] X.-Y. Huang, Y.-J. Zhong, and D. Hu, “Image retrieval based on weighted texture features using DCT coefficients of JPEG images,” Proc. of the Joint Conf. of the 4th Int. Conf. on Information, Communications and Signal Processing, and the 4th Pacific Rim Conf. on Multimedia, Vol. 3, pp.1571-1575, Dec. 2003.

[7] M.-J. Hsiao, Y.-P. Huang, and T.-W. Chiang, “A region-based image retrieval approach using block DCT,” Proc. of IEEE the 2nd Int. Conf. on Innovative Computing, Information

and Control, Sep. 2007.

[8] M. M. Islam, D. Zhang, and G. Lu, “Comparison of retrieval effectiveness of different region based image representations,” Proc. of the 6th Int. Conf. on Information, Communications, and Signal Processing, pp.1-4, Dec. 2007.

[9] W.-J. Li and D.-Y. Yeung, “Localized content-based image retrieval through evidence region identification,” Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, June 2009.

[10] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, “A survey of content based image retrieval with high-level semantics,” Pattern Recognition, Vol. 40, No. 1, pp.262-282, 2007.

[11] R. Rahmani, S. A. Goldman, H. Zhang, S. R. Cholleti, and J. E. Fritts, “Localized content-based image retrieval,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 30, No. 11, pp.1902-1912, 2008.

[12] S. Rudinac, M. Uscumlic, M. Rudinac, G. Zajic, and B. Reljin, “Global image search vs. regional search in CBIR systems,” Proc. of the 8th Int. Workshop on Image Analysis for Multimedia Interactive Services, June 2007.

[13] G. Sheikholeslami, W. Chang, and A. Zhang, “SemQuery: Semantic clustering and querying on heterogeneous features for visual data,” IEEE Trans. on Knowledge and Data Engineering, Vol. 14, No. 5, pp.988-1002, 2002.

[14] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, “Content-based image retrieval at the end of the early years,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, pp.1349-1380, 2000.

[15] R. C. Veltkamp and M. Tanase, “Content-based image retrieval systems: a survey,” Technical Report UU-CS-2000-34, Utrecht University, Available at http://give-lab.cs.uu.nl/cbirsurvey/cbir-survey.pdf.

[16] J. Z. Wang, Content Based Image Search Demo Page, Available at http://bergman. stanford.edu/~zwang/project/imsearch/WBIIS.html, 1996.

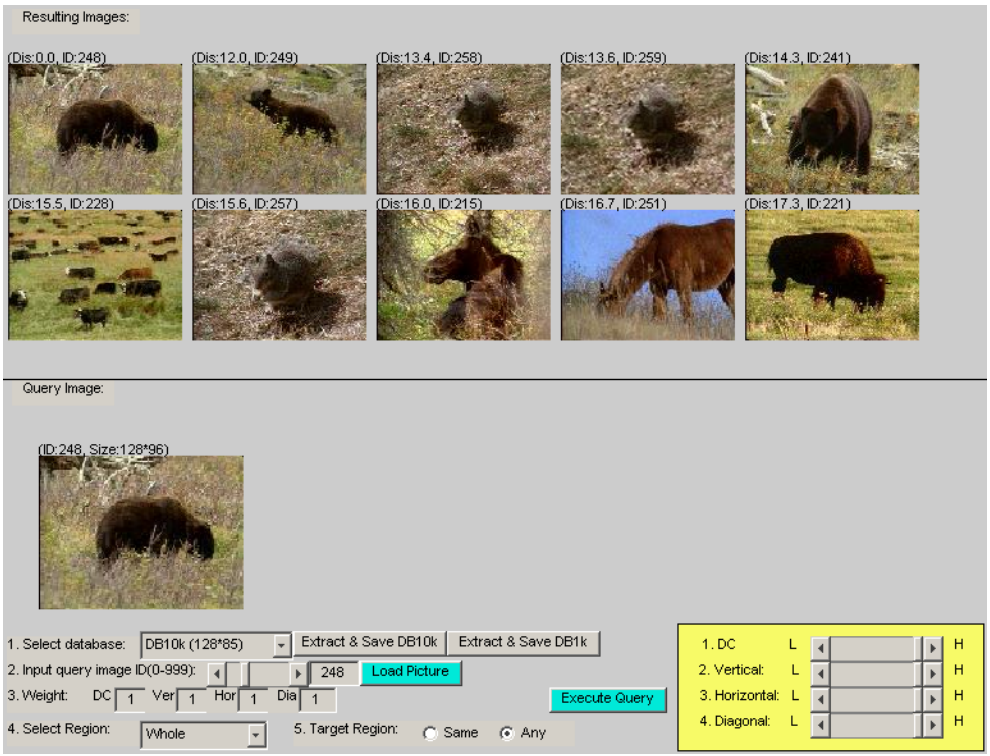
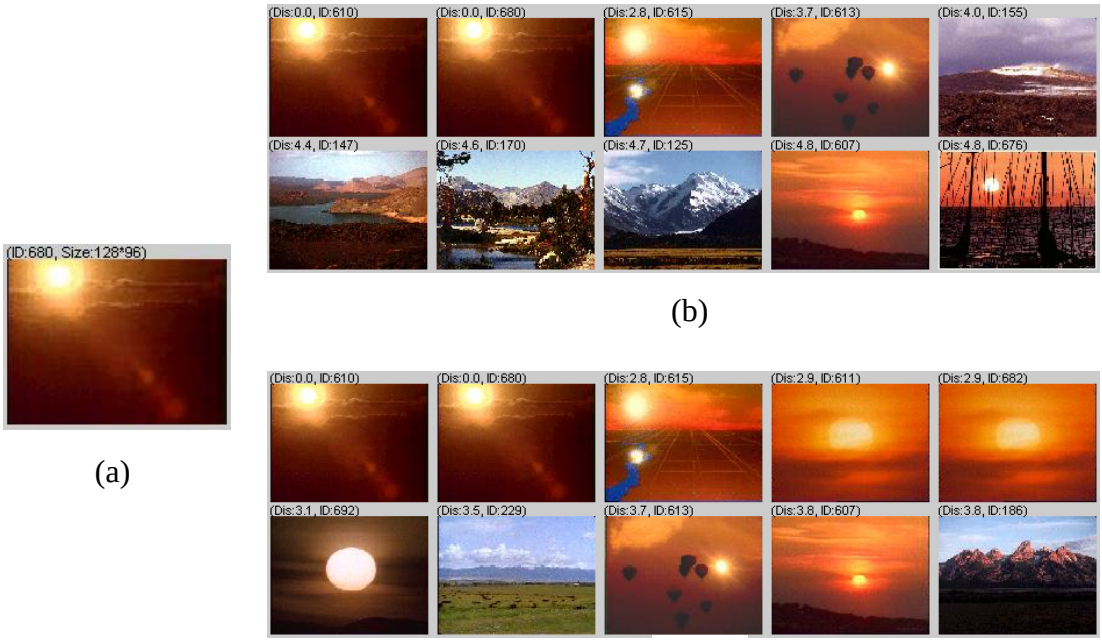
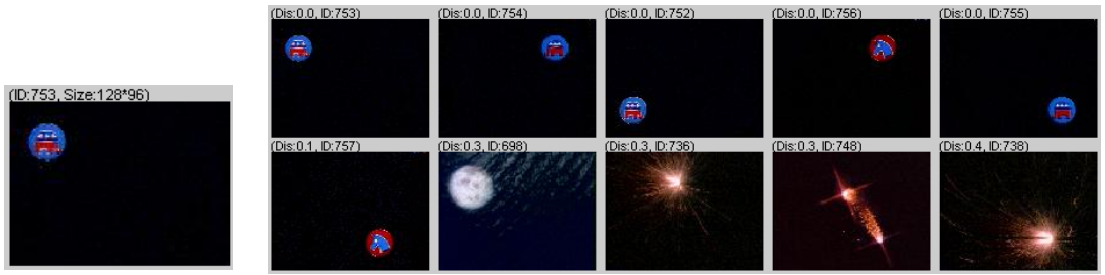


Fig. 3. The main screen of our system.



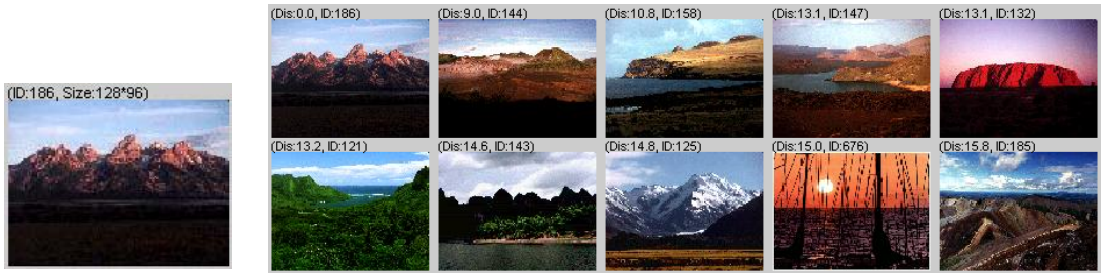
(a) (b) (c)

Fig. 4. (a) The query image; (b) retrieval results for option “same”; (c) retrieval results for option “any”. (ROI is “upper left”)



(a) (b)

Fig. 5. (a) The query image; (b) retrieval results for option “any”. (ROI is “upper left”)



(a) (b)

Fig. 6. (a) The query image; (b) retrieval results for *W*=(1,1,1,1). (ROI is “whole”)

(a)

(b)

(c)

Fig. 7. (a) The query image; (b) retrieval results for W=(1,1,1,1) (c) retrieval results for W=(1,1,0,1). (ROI is “whole”)