THE JOURNAL OF CHINA UNIVERSITIES OF POSTS AND TELECOMMUNICATIONS
Volume 14, Supplement, October 2007

SHI Dong-cheng, XU Lan, HAN Ling-yan

Image retrieval using both color and texture features

CLC number TP391.41 **Document** A **Article ID** 1005-8885 (2007) S1-0094-06

Abstract In order to improve the retrieval performance of images, this paper proposes an efficient approach for extracting and retrieving color images. The block diagram of our proposed approach to content-based image retrieval (CBIR) is given firstly, and then we introduce three image feature extracting arithmetic including color histogram, edge histogram and edge direction histogram, the histogram Euclidean distance, cosine distance and histogram intersection are used to measure the image level similarity. On the basis of using color and texture features separately, a new method for image retrieval using combined features is proposed. With the test for an image database including 766 general-purpose images and comparison and analysis of performance evaluation for features and similarity measures, our proposed retrieval approach demonstrates a promising performance. Experiment shows that combined features are superior to every single one of the three features in retrieval.

Keywords CBIR, color histogram, edge histogram, edge direction histogram, performance evaluation.

1 Introduction

Content-based image retrieval has become a prominent research topic because of the proliferation of video and image data in digital form. Increased bandwidth availability to access the internet in the near future will allow the users to search for and browse through video and image databases located at remote sites. Therefore fast retrieval of images from large databases is an important problem that needs to be addressed.

Image retrieval systems attempt to search through a database to find images that are perceptually similar to a query image. CBIR is an important alternative and complement to traditional text-based image searching and can greatly enhance the accuracy of the information being returned. It aims to develop an efficient visual-content-based technique to search, browse and retrieve relevant images from large-scale digital image collections. Most proposed CBIR techniques automatically extract low-level features (e.g. color, texture, shapes and layout of objects) to measure the similarities among images by comparing the feature differences.

In this paper, we present a technique for image retrieval based on color and texture. Low-level visual features of the images such as color and texture are especially useful to represent and to compare images automatically. In the concrete selection of color and texture description, we use color histogram, edge histogram and edge direction histogram. We made reference to MPEG-7 standard, and use the color histogram descriptor and edge histogram descriptor.

2 Content-based image retrieval frameworks

The block diagram of our proposed approach to CBIR is shown in Fig. 1. In the CBIR system, the relevance between a query and any target image is ranked according to a similarity measure computed from the visual features. The similarity comparison is performed based on visual content descriptors including color histogram and edge histogram. The key steps we can see from the figure are features extraction and similarity measure based on content.

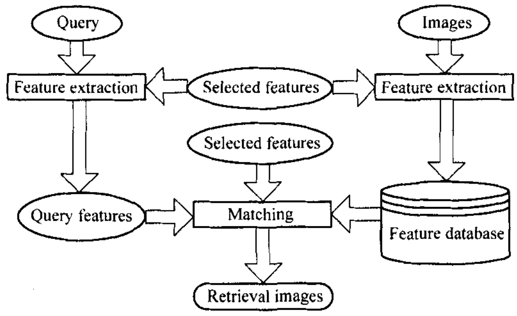


Fig. 1 Diagram for content-based image retrieved system

3 Feature extraction

3.1 Color histogram

Color is an important visual attribute for both human perception and computer vision and one of the most widely used visual features in image/video retrieval [1]. But an appropriate color space and color quantization must be specified along with a histogram representation of an image for retrieval purpose. Histogram describes the global distribution of pixels of an image [2]. The main advantage of a color histogram is its small sensitivity to variations in scale, rotation and translation of an image. We utilize different kinds of quantization schemes for the implementation of the color histograms in HSV color space. We

Received date: 2007-07-01
SHI Dong-cheng, XU Lan (✉), HAN Ling-yan
Computer Science and Engineering Institute, Changchun University of Technology, Changchun 130012, China
E-mail: xl111628@126.com

observed that the HSV color model is better than the RGB color model for our approach using the following quantization scheme where each color component is uniformly quantized: H: 16 bins; S: 4bins; V: 4bins. Finally, we concatenate this 16×4×4 histogram and get a 256-dimensional vector [3]. We do not use the Haar transformation as the MPEG-7 standard. The part 4 discusses the use of a cosine similarity coefficient and other similarity measures.

3.2 Edge histogram

The edge histogram descriptor captures the spatial distribution of edges. The distribution of edges is a good texture signature that is useful for image to image matching even when the underlying texture is not homogeneous. The computation of this descriptor is fairly straightforward [4]. A given image is first sub-divided into 4x4 sub-images, and local edge histograms for each of these sub-images are computed. Edges are broadly grouped into five categories: vertical, horizontal, 45° diagonal, 135° diagonal, and isotropic (non orientation specific) (see Fig. 2). Thus, each local histogram has five bins corresponding to the above five categories. The image partitioned into 16 sub-images results in 80 bins.

To compute the edge histograms, each of the 16 sub-images is further subdivided into image blocks. The size of these image blocks scale with the image size and is assumed to be a power of 2. The number of image blocks per sub-image is kept constant, independent of the original image dimensions, by scaling their size appropriately. A simple edge detector is then applied to each of the macro-block, treating the macro-block as a 2x2 pixel image. The edge-detector operators include four directional selective detectors and one isotropic operator. Those image blocks whose edge strengths exceed a certain minimum threshold are used in computing the histogram.

Thus, for an image block, we can compute five edge strengths, one for each of the five filters from Fig. 2. If the maximum of these edge strengths exceed a certain preset threshold, then the corresponding image block is considered to be an edge block. An edge block contributes to the edge histogram bins.

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & \sqrt{2} \\ -\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

(a) (b) (c) (d) (e)

(a) Vertical edge (b) Horizontal edge (c) 45 degree edge
(d) 135 degree edge (e) Non-directional edge

Fig. 2 Filters for edge detection

Here, we utilize YCbCr color space, which is a scale and offset version of the YUV color space. The YUV space is widely used in image compression and processing applications. Y represents the luminance of a color, while U and V represent the chromaticity of a color. The luminance Y component is separated from the chromatic components in this space [5].

3.3 Edge direction histogram

Another way of edge detection is to examine each pixel and its direct adjacent area's state to decide whether the pixel is placed in a boundary of an object, if the pixel has the characteristic, it's called edge point. When the gray degree of each pixel reflects that each pixel meets the degree that the edge pixel request, this kind of picture is called edge picture or edge diagram (edge map). The coding image of edge direction not extent is called edge direction diagram.

In our experiment, we convert the RGB pictures to the YCbCr color space first, and then take out the irrelevant boundary to get a central portion luminance image. We select Sobel edge detector operator to calculate the edge image of luminance. We set a region value TH=[tan(*p*/8), tan(*p**3/8), tan(−*p*/8), Inf; tan(−*p*/8), tan(*p*/8), tan(−*p**3/8), tan(*p**3/8)]. While $-\pi/8 \leqslant q < \pi/8$, it belongs to 0 degree direction; $\pi/8 \leqslant q < 3\pi/8$, it belongs to 45 degree direction; $-3\pi/8 \leqslant q < -\pi/8$, it belongs to −45 degree direction; $3\pi/8 \leqslant \theta < \infty$, it belongs to 90 degree direction. And operation is made between the edge diagram of the whole image and each angle in order to wipe off the voice to make the edge of every direction clearer. Finally, we work out the elementary number of each direction and compute the histogram.

4 Histogram distance measures

Instead of exact matching, content-based image retrieval calculates visual similarities between a query image and images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent. Different similarity/distance measures will affect retrieval performances of an image retrieval system significantly. In this section, three well-known histogram distance measures are described.

4.1 Histogram euclidean (HE) distance

Let H_1 and H_2 represent two histograms. The Euclidean distance between the histograms H_1 and H_2 can be computed as

$$d_{\mathrm{HE}}(H_1,H_2)=\sqrt{\sum_{x\in X,\, y\in Y,\, z\in Z}\left(H_1(x,y,z)-H_2(x,y,z)\right)^2} \qquad (1)$$

This distance is the L^2 -norm of the different vector.

4.2 Cosine distance

The cosine similarity coefficient is used for comparing feature vectors of a query image and feature vectors of images in a database. It utilizes the angle between vectors to measure similarity. The following equation is referred to as a cosine similarity coefficient:

$$\cos\theta = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}} \tag{2}$$

$$d_{\cos} = \cos^{-1} \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}} \tag{3}$$

4.3 Histogram intersection (HI) distance [6]

The histogram intersection was proposed for image retrieval by Swain and Ballard. The intersection of histograms H_1 and H_2 was originally defined as

$$I_{H(H_1,H_2)} = \frac{\sum_{x\in X, y\in Y, z\in Z} \min(H_1(x,y,z), H_2(x,y,z))}{\sum_{x\in X, y\in Y, z\in Z} H_2(x,y,z)} \tag{4}$$

where H_1 is the histogram of a query image and H_2 is the histogram of an image in the database. This is not a distance in the strict sense, since it dose not satisfy the condition of association. However, it does satisfy the associative condition if the cardinalities of the histograms of the query image and the image in database are the same. In that case the histogram intersection is equivalent to the L^1 -norm of the difference vector.

Smith and Chang extended to the case when the cardinalities of the two histograms are different by modifying the denominator of the original definition slightly:

$$I_{H(H_1,H_2)} = \frac{\sum_{x\in X, y\in Y, z\in Z} \min(H_1(x,y,z), H_2(x,y,z))}{\min\left[\sum_{x\in X, y\in Y, z\in Z} H_1(x,y,z) \sum_{x\in X, y\in Y, z\in Z} H_2(x,y,z)\right]} \tag{5}$$

Finally, to satisfy the condition $d(H_1,H_2) = 0$, for $H_1 = H_2$, the histogram intersection distance can be defined as follows:

$$d_{HI}(H_1,H_2) = 1 - I_H(H_1,H_2) \tag{6}$$

4.4 The distance measure of combined features

When we use combined features for image retrieval, the corresponding distance measure is different as a result of different feature extraction. In order to compare different sub-features, we need make normalize for similarity distance of sub-features. The overall similarity is the sum of weighted similarity for normalized distance. The overall similarity is calculated by

$$S' = w_c S_c + w_t S_t \tag{7}$$

where w_c and w_t determine the contribution of color features and texture features respectively in measuring the similarities. S_c and S_t determine the normalize similarity of color features and texture features between two images.

For the selection of weight w_c and w_t, there is no concrete specification, it's indicated by the user. Simply, if the query image is a texture image, $w_t > w_c$; while the query image is a non-texture image, $w_t < w_c$ and satisfy the condition $w_c + w_t = 1$.

5 Performance evaluations in content-based image retrieval

The performance of a retrieval system can be measured in terms of its recall (or sensitivity) and precision (or specificity) [7]. Recall measures the ability of the system to retrieve all models that are relevant, while precision measures the ability of the system to retrieve only models that are relevant. They are defined as

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \tag{8}$$

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{9}$$

While in the system of CBIR, a sort list based on similarity is always returned to user. Commonly, we use P_N, which is the precision of the top N returned results. It defined as follows:

For the retrieval process using a query image q_i Î R, where R is a set of images including some specifically semantic meanings. The purpose for users to submit q_i is to search R, system returns the top N results p_j, j=1, 2,..., N, precision is calculated as

$$P_N(q_i) = \sum_{k=1}^{N} \frac{\psi(p_k,R)}{N}, \quad \psi(x,Y) = \begin{cases} 1; & if\ x\in Y \\ 0; & if\ x\notin Y \end{cases} \tag{10}$$

So the average precision for all test-example retrieval image sets is

$$P_N = \sum_{i=1}^{\text{Total\_Query\_Count}} \frac{P_N(q_i)}{\text{Total\_Query\_Count}} \tag{11}$$

Similarly, the recall of the top N returned results is defined as follows:

$$R_N(q_i) = \sum_{k=1}^{N} \frac{\psi(p_k,R)}{\|R\|} \tag{12}$$

while $\|R\|$ determines the count of image sets, so the average recall for all test-example retrieval image sets is

$$R_N = \sum_{i=1}^{\text{Total\_Query\_Count}} \frac{R_N(q_i)}{\text{Total\_Query\_Count}} \tag{13}$$

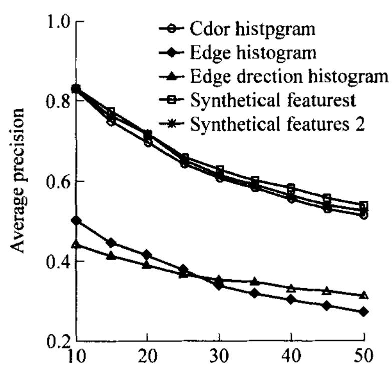
6 Experimental results

To date, we have tested our retrieval algorithm on a general-purpose image database with 766 images from Calphotos and Internet. These images are stored in BMP format with size 198x198 or 256x384. The entire database has 15 categories with 40–170 images in each category. Most categories contain distinct semantics including horse, flowers, lion, elephant, birds, fish, penguin and sunset and so on.

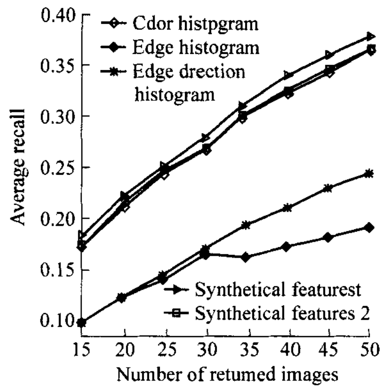
To qualitatively evaluate the retrieval effectiveness of our algorithm over the 766-imge database, we randomly select 10 query images with different semantics on 5 distinct image categories respectively. For each query image, we examine the precision and recall of the query results based on the relevance of the image semantics and then calculate the average precision and the average recall of each category. Here, a retrieved image is considered as a correct match if and only if it is in the same category as the query image.

6.1 Comparison of overall retrieval performance based on image features

Performance evaluation experiment based on image features is used to evaluate the retrieval performance of different features on the same image database, which uses the same similarity measure. The experiment selects five methods for image retrieval which are color histogram, edge histogram, edge direction histogram, combined features1 (color histogram + edge histogram) and combined features 2 (color histogram + edge direction histogram). Figure 3 compares the overall average retrieval precision and recall of the 5 distinct image categories from top 10, 15,...,50 returned images. Since color is more important than texture in natural images, is empirically set to be 0.85 in our system. Here, we use histogram euclidean distance for similarity measure.



(a) Comparison of the retrieval precision of different features



(b) Comparison of the retrieval recall of different features

Fig. 3 Performance evaluation based on image features

The experiment results indicate that the precision of color feature is superior to texture feature with the same number of returned images, the retrieval result obtained from combined features fits more closely with human perception than the retrieval result obtained from single-feature. When N increases, the precision of a same feature presents downtrend, the reason is that ground truth set keeps invariable while N increases, after sorting, the most correct images line in the front of the queue. The closer to the end of the queue, the accurate probability is lower. So N is bigger, precision is lower. Recall tends to increase as the number of retrieved images increase. Further, when the number of relevant images is greater than the number of the retrieved images, recall is meaningless.

6.2 Performance evaluation of similarity measure

Performance evaluation experiment based on similarity measure is used to evaluate the retrieval performance of different similarity measures for the same image feature. We mainly compare the performance of Euclidean, cosine and histogram intersection. Figure 4 shows the comparative results which utilize single feature (color histogram), combined features1 and combined features 2 (w_c =0.85). Where C means color histogram, C-EH means color histogram + edge histogram, C-ED means color histogram + edge direction histogram.

From the retrieval result variation curves of Fig. 5, we can see:

1) Combined features are superior to every single one of the three features in retrieval.

2) While measuring by Euclidean distance, the retrieval performance of combined features 2 is superior to single feature slightly, the precision is improved by 0.37%, recall is improved by 0.21%; the retrieval performance of combined features 1 is superior to single feature obviously, the precision is improved by 1.16%, recall is improved by 0.70%.

3) While measuring by cosine distance, the retrieval performance of combined features1 is superior to single feature, the

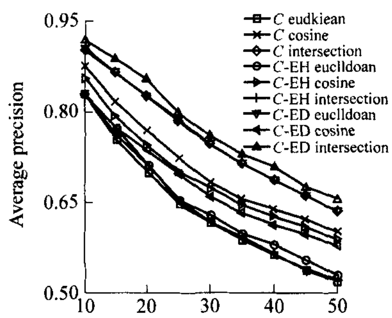
precision is improved by 1.37%, recall is improved by 0.69%; the retrieval performance of combined features 2 is superior to single feature obviously, the precision is improved by 3.01%, recall is improved by 1.69%.

4) While measuring by histogram intersection distance, the retrieval performance of combined features1 is close to single feature, the precision is improved by only 0.08%, recall is improved by 0.02%; the retrieval performance of combined features 2 is superior to single feature obviously, the precision is improved by 1.88%, recall is improved by 1.12%.

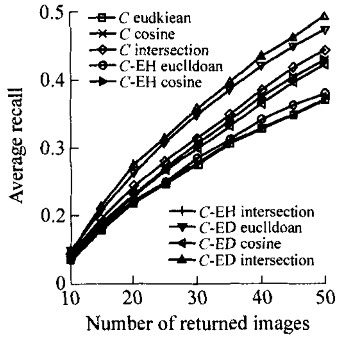
5) In the three similarity measure methods, the retrieval performance of cosine distance is superior to Euclidean distance, histogram intersection distance is superior to the former two, but its speed is slower correspondingly. Where, single feature using cosine distance measure compared with Euclidean distance, the precision is improved by 3.97%, recall is improved by 2.45%, and using histogram intersection measure compared with cosine distance, the precision is improved by 11.79%, recall is improved by 6.65%; combined features1 using cosine distance measure compared with Euclidean distance, the precision is improved by 4.18%, recall is improved by 2.44%, and using histogram intersection measure compared with cosine distance, the precision is improved by 10.71%, recall is improved by 5.97%; combined features 2 using cosine distance measure compared with Euclidean distance, the precision is improved by 6.61%, recall is improved by 3.92%, and using histogram intersection measure compared with cosine distance, the precision is improved by 13.30%, recall is improved by 7.55%.

6) Image retrieval using our proposed combined features method, the average precision is above 64%, and the average recall is above 27%, which demonstrates a promising performance.

We can see the retrieval results via an example in Fig. 5 visually. The results are close to human visual perception.

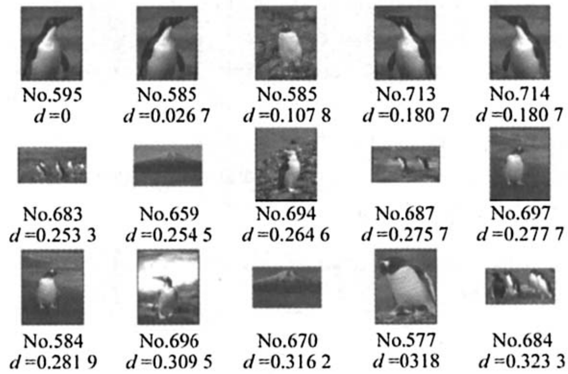


(a) Comparison of the retrieval precision of similarity measure

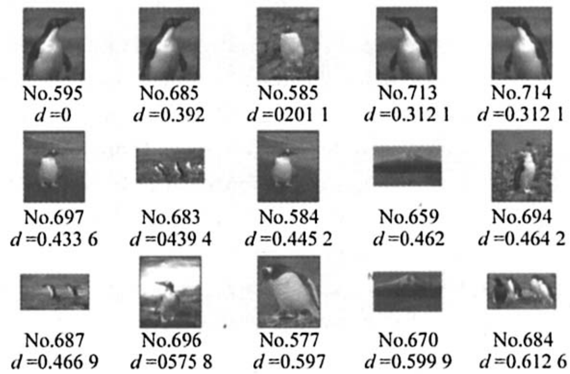


(b) Comparison of the retrieval recall of similarity measure

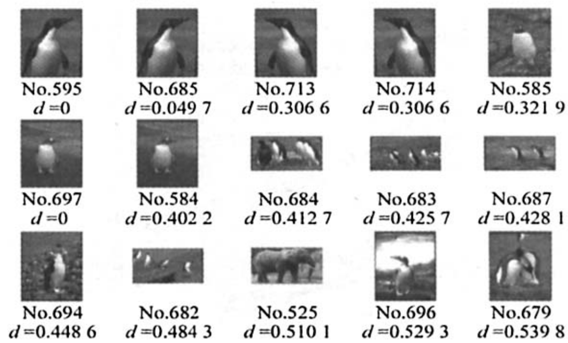
Fig. 4 Performance evaluation of similarity measure



(a) Retrieval results by euclidean distance



(b) Retrieval results by cosine distance



(c) Retrieval results by histogram intersection

Fig. 5 Retrieval results by using our proposed method (combined features1)

7 Conclusions

The traditional image retrieval mainly depends on color, texture and shape. For these basic visual features are just parts of image information, the retrieval results are not so perfect. This paper proposes a novel approach for image retrieval, which takes combined features as retrieval foundation. It extracts combined features through getting the single features such as color and texture. We tested 766 pictures, the performance and accuracy is reasonable as expected. But our experimental system is still far from practice application. The future work includes the investigation of other properties of the image pixels that can be exploited in order to have a more precise description of the image's visual content. In addition, the similarity measures between visual features do not necessarily match human perception. Users are interested in are semantically and perceptually similar images, the retrieval results of low-level feature based retrieval approaches are generally unsatisfactory and often unpredictable. We also plan to investigate the use of relevant feedback to reduce the semantic gap between the low-level visual features automatically extracted from images and the human interpretation of the image's visual content.

References

1. Rui Y, Huang T S, Chang S F. Image retrieval: current techniques, promising directions and open issues, Journal of Visual Communication and Image Representation, 1999, 10(1): 39–62
2. Smith J R, Chang S F. Automated image retrieval using color and texture, Tech. Rep., Columbia University, 1995
3. Smith J R, Chang S F. Tools and techniques for color image retrieval, in: IST/SPIE-Storage and Retrieval for Image and Video Databases IV, San Jose, CA, 2670, 1996, 426–437
4. Manjunath B S, Jens-Rainer Ohm, Vasudevan V, et al. Color and texture descriptors, IEEE Transactions on circuits and systems for video technology, 2001, 11(6), 703–715
5. Timothy K SHIH, Huang Jiung-yao, Wang Ching-sheng, et al. An intelligent content-based image retrieval system based on color, shape and spatial relations, in: Proc. Natl. Sci. Counc. ROC (A), 25(4), 2001, 232–243
6. Sangoh Jeong, Chee Sun Won, Robert M Gray. Image retrieval using color histograms generated by Gauss mixture vector quantization, Computer vision and image understanding, 94, 44–66
7. Kulwinder Singh, Ming Ma, Dong-Won Park, Syungog An. Image indexing based on Mpeg-7 scalable color descriptor, Key Engineering Materials, 2005, 375–382

Biographies: SHI Dong-cheng, assistant dean of Computer Science and Engineering Institute, Changchun University of Technology, professor. Main research areas: image processing and pattern recognition, multimedia communication technology, speech signal digital processing. He has performed research project 7 items, published about 40 journal papers.

XU Lan, received the B. S. degree in the North University for Nationalities in 2005, and now she is studying her M. S. degree in Changchun University of Technology. Her current research interests are content-based image retrieval, image processing and computer vision.

HAN Lingyan, female, received the B. S. degree in the Anhui University in 2003, and now she is studying her M.S. degree in Changchun University of Technology. Her current research interests are pattern recognition, image processing and computer vision.