A Comparative Study on Content Based Image Retrieval Methods

Kratika Arora, Ashwani Kumar

Sant Longowal Institute of Engineering and Technology, Longowal, Punjab, India.

***Abstract*— Content-based image retrieval (CBIR) is a method of finding images from a huge image database according to persons' interests. Content-based here means that the search involves analysis the actual content present in the image. As database of images is growing daybyday, researchers/scholars are searching for better techniques for retrieval of images maintaining good efficiency. This paper presents the visual features and various ways for image retrieval from the huge image database**.

***Keywords*—CBIR, correlogram, feature extraction, descriptors, relevance feedback.**

I. INTRODUCTION

Rapidly growing in technology day by day, has led to the increase in multimedia files, digital images and visual objects. Advancement in medical field and in many other technologies has given rise to extensive images, their memory requirement and transmission capability. Increase in the usage of images in various areas, has led researchers to focus on new ways [1] by which images can be quickly and efficiently retrieved and from the large databases. So the mechanism of image processing and retrieval of the desired image from the large image database has become an serious task. For many years, researchers have been working on retrieval processes. Content-based image retrieval [2] is the modern image retrieval system. In CBIR retrieval system, different techniques are brought together for the same purpose as image retrieval, information processing, and database communities. [3]In a general content-based image retrieval system(Fig. 1), the visual content present in the image, in the database is extracted and is described by a feature vector. The content can be s, texture, shape or any other useful information that can be derived from the image.For the image retrieval, user provides the retrieval system with sample images or figures. System then internally changes these samples into its feature vectors. The similarity between the images is determined by calculating the distance between the feature vectors of the query sample or the sketch and those of the images present in the database. Image retrieval is then done with the aid of an indexing scheme. Relevance feedback is also used in CBIR in order to generate semantically more and better meaningful retrieval results.

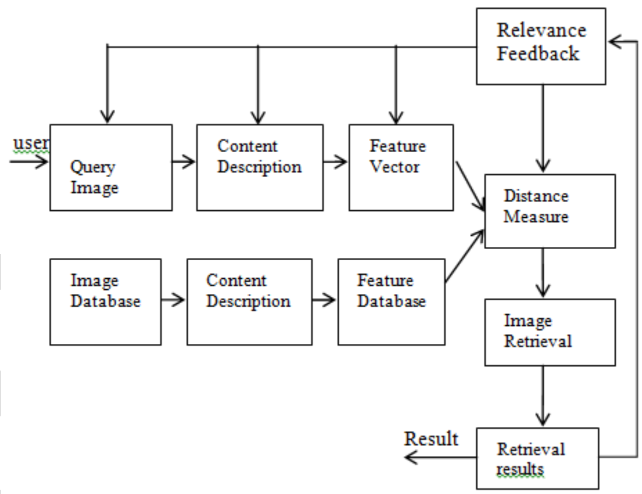

Fig. 1. Block Diagram of CBIR

II. IMAGE CONTENT DESCRIPTORS

Image content can be classified as visual and semantic. Visual content is very generic or area specific. Generally the visual content of the image means the colour, texture, and shape. Area specific visual content is application dependent like human eyes, faces and it may involve some field knowledge. This paper focuses on general visual content descriptors. A proper descriptor must be invariable to the variation which is introduced accidentally by the process (e.g., illuminance variation). Also, a visual descriptor can be local or global. A local descriptor uses the visual features of part/area or objects in the image for the description of the image content whereas global descriptor use visual features of the complete image. Below section presents some techniques which are used widely for the extraction of colour, texture and shape from the image.

A. COLOR

The color features are extensively used features in image processing, as their extraction is easy compared to extraction of image tex. and shape. Colour is comparatively robust and is also independent of image size and orientation. The following sections presents commonly used techniques for extraction of color : color moments, color histogram and color correlogram.

1. Color Moments

Color moments are being used successfully in many CBIR systems [4] mainly when image has only the object. These have been efficient in representation of colour distributions of images. In this,1st order (mean), 2nd order (variance) and 3rd order (skewness), are calculated.
Mathematically, these 3 moments are defined as:

$$E_{r,i} = \frac{1}{N}\sum_{j=1}^{N} I_{ij} \tag{1}$$

$$\sigma_{r,i} = \left(\frac{1}{N}\sum_{j=1}^{N}(I_{i,j} - E_{r,i})^2\right)^{1/2} \tag{2}$$

$$S_{r,i} = \left(\frac{1}{N}\sum_{j=1}^{N}(I_{i,j} - E_{r,i})^3\right)^{1/3} \tag{3}$$

where N represents no. of pixels present in the image and I_{ij} represents the value of i-th colour component for the image pixel j. Using the 3rd -order moment in addition i.e skewness, compared to using only the 1st and 2nd order moments improves the performance. But sometimes it makes the representation more sensitive and illumination changes and decreasing the overall performance.
Since, only 9 numbers (3 moments for each of the 3 color components) has been used to represent the colour content of every image, colour moments becomes a packed representation compared to all other colour methods. Due to the packedness, it may decrease the discrimination power. Hence, color moments is often used as the 1st pass to narrow down the search before using other colour extraction methods for retrieval.

2. Color Histogram

Histogram gives the distribution of number of pixels present in an image. These are easier in computation and are effective in describing both the local and global colour distribution in an image. For huge image database, histogram comparison saturates the discrimination. For solving such problem, the joint histogram technique can be used. [5] As the colour histogram doesn't takes spatial information into account, thus different images can have similar color histogram. This problem becomes especially acute for large databases. To increase the discrimination power, many improvements are designed to consider spatial information. A simple approach includes dividing an image into sub-parts and computing the histogram for every sub-part separately. This division can be simple as vertical or horizontal partition, or as complex as an area or even object segmentation. Also, increase in the no. of sub-parts increases the information content and also the memory and computational time at the same time.

3. Color Correlogram

The color correlogram [6] characterizes the distributions of pixels as well as the spatial correlation of color pairs. The 1st and the 2nd dimension of the 3-D histogram represents the colours of any pixel pair and the 3 dimension represents the spatial distance. A color correlogram is in a table format, in which the r-th entry for (i, j) is the probability of finding a pixel of colour j at a distance r from a pixel of colour i . Let I be the entire set of image pixels and Ib(i) the set of pixels whose colours are b(i). For this the colour correlogram is as shown below:

$$\gamma_{i,j}^{(s)} = \Pr\ [|p_2 \in I_{b(j)}|\,|\, p_2 - p_1|] \tag{4}$$

where i, j ∈ {1, 2, 3,… M}, s∈ {1, 2, 3, …, d}, and | p2 – p1 | mentions the distance between pixels p2 and p1. Considering all the combinations of colour pairs, makes the size of the colour correlogram large (O(M2d)), therefore a simplified version of the correlogram called the color autocorrelogram is often used. The color autocorrelogram captures only the spatial correlation between same colors and thus it reduces the dimension to O(Md). Compared to the other color extraction methods, the autocorrelogram gives the better retrieval results but at the same time it is also computational very expensive due to its large dimensionality.

B. Texture

The texture is the visual and especially tactile quality of a surface. It contains information about the structural arrangement and their relation to the surrounding pixel. It is property of virtually all surfaces including flowers, bricks, skin, fabric, etc. There are 3 levels at which texture analysis [7] can be discussed : On the structural level, texture is the primitives of the image and their placement rules. On a statistical level, a set of statistics value extracted from the image. On the spectral level, set of coefficients in the transform domain is texture. With the aid of above levels, the textures can be identified but it may not agree with the human perception of evaluating the textures. Reasons for this include semantic gap and human perception Texture feature [8] describes spectral features which are taken using wavelet transform, statistical features. In the following section, we discuss Haar wavelet Transforms and Gabor Filter Features for extraction of texture.

1. Haar wavelet Transforms

Wavelets give a multi-resolution approach for texture classification and analysis. For a 2-dimensional image, wavelet transform computation is also a multi resolution approach, that applies recursive filtering and sampling. At every sub level, the image is decomposed into 4 frequency sub-bands, LL, HH, HL, and LH where H is the high frequency and L is the low frequency. Let a data set X1, X2,…Xn.. contains M elements, so there will be M/2 averages and M/2 wavelet coefficient values. The average is stored in the 1st half of the M element array, and the coefficient is stored in the 2nd half of the M element array. The averages are the input in the next step in the calculation. For a 1D Haar transform of an array of M elements, determine the average of every pair, and find the

difference between every pair of, and divide it by 2. Now fill the 1st half with the averages, the 2nd half of the array with coefficients. Repeat the process till a single average and a single coefficient are determined. For a 2D Haar transform, Compute 1D Haar wavelet decomposition for every row of the original pixel values, then compute1D Haar wavelet decomposition for every column of the row-transformed pixels. RGB components are extracted from the images. Now apply the 2D Haar transform to every colour matrix.
A feature vector is determined from every image present in the database. Set of all the features is organized in a database index. When similar images has to be searched with a query image, a feature vector is again extracted from the query image and is then matched against the feature vectors in the index. Measure between feature representation of the query image and feature representation of the database image is find, if it is small, then it is considered similar. In this way, Haar wavelet can be used for matching images from the database.

2. Gabor Filter Features

The Gabor filter [10] is used world-wide to extract features of image, especially the texture features. It is optimal in minimization of the joint uncertainty in frequency and in space and is often used as an scale and orientation tunable edge and bar/line detector. There have been many approaches which are designed to characterize tex. of images based on Gabor filters.
A 2-dimensional Gabor function is :

$$g(x,y) = 1/2\pi\sigma_x\sigma_y \exp[-\tfrac{1}{2}((x/\sigma_x)^2 + (y/\sigma_y)^2) + 2\pi jWx] \quad (5)$$

where, σ_x and σ_y represents the standard deviations of Gaussian envelopes along the x, y direction.
Then a list of Gabor filters is obtained by appropriate rotations and dilations of g(x, y):

$$g_{mn}(x,y) = a^{-m}\, g(x1, y1) \quad (6)$$

$$x1 = a^{-m}(x\cos\theta + y\sin\theta) \quad (7)$$
$$y1 = a^{-m}(-x\sin\theta + y\cos\theta) \quad (8)$$

where $a > 1$, $\theta = n\pi/K$, $n = 0, 1, 2, 3 \ldots, K-1$, and $m = 0, 1, 2, 3 \ldots, S-1$. S and M are the no. of scales and orientations. The scale factor a^{-m} is ensures that energy is independent of m.
Given an image I(x', y'), its Gabor transform is defined as:

$$W_{mn}(x', y') = \int I(x', y') g_{mn}^{*}(x' - x_1, y' - y_1)\, dx_1 dy_1 \quad (9)$$

where * is complex conjugate. Mean μ_{mn} and standard deviation σ_{mn} of the $|W_{mn}|$ is $f = [\mu_{00}, \sigma_{00}, \ldots \mu_{mn}, \sigma_{mn}, \Lambda, \mu_{s-1k-1}, \sigma_{S-1K-1}]$ represents the texture feature vector of the texture area.

C. Shape

The third feature shape can also be used for the image retrieval process. Shape of an image may not refer to the shape of an image but to the shape of some particular area or object in the image. Depending on the job or application, sometimes user requires the shape representation to be invariable to rotation, translation and scaling. Shape features [9] of objects or areas have been used in many CBIR systems. Unlike to colour and text. features, shape features are generally described after image has been segmented into parts or objects. Shape features are categoriezed as area based and boundary based. Boundary based uses only the outer boundary of the shape whereas area-based shape features use entire shape area. The shape is represented through clubbed geometric cues such as corners, joints and polygonal areas extracted from image. Such a clubbing may serve as a spatial layout or as a rough sketch by additional processing. Commonly used shape feature moment invariants and Fourier Descriptors are discussed below.

1. Moment *Invariants*

Moment invariants are used for the shape representation of an image [11]. If the object S is a binary image, the central moments of order m+n for the shape of object S are defined as:

$$\mu_{m,n} = \sum_{x,y\in R}(x-x_c)^m(y-y_c)^n \quad (10)$$

Based on the above moments, set of invariants for rotation, translation, and scale can be determined.

$$Ø_1 = \mu_{2,0} + \mu_{0,2} \quad (11)$$

$$Ø_2 = (\mu_{2,0} - \mu_{0,2})^2 + 4\mu^2_{1,1} \quad (12)$$

$$Ø_3 = (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2 \quad (13)$$

$$Ø_4 = (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2 \quad (14)$$

2. *Fourier Descriptors*

Fourier descriptors describes the shape of a area or object with Fourier transform at its edge or boundary. Considering the contour of a 2D object as a sequence of successive pixels (x_i, y_i), where $0 \le i \le M-1$ and M is the no. of pixels in the boundary. The 3 types of contour representations can be defined : centroid distance, curvature and complex coordinate function.
The curvature C(s) at a point 'i' along the contour is change in tangent direction of the contour, i.e.

$$C(s) = \frac{d}{ds}\theta(s) \quad (15)$$

where θ(s) : turning function.

The distance between boundary pixels and the centroid (x_o, y_o) of the object is centroid distance given below:

$$R(s) = ((x_s - x_o)^2 + (y_s - y_o)^2)^{1/2} \tag{16}$$

The Fourier transforms of these 3 types of contour representations generate 3 sets of complex coefficients, representing the shape in the frequency domain. Lower frequency coefficients signify the general shape while higher frequency coefficients signify shape details. The Fourier descriptor are shown below:

For Curvature:

$$F_K = |F_1| , |F_2| , | F_3| \;..., |F_{M/2}| \tag{17}$$

For centroid distance :

$$f_R = \frac{|F1|}{|F0|}, \frac{|F2|}{|F3|}, \dots\dots, \frac{|FM/2|}{F0} \tag{18}$$

where F_i : ith component of Fourier transform coefficients. Here Fourier transforms exhibit symmetry, i.e., $|F\text{-}j| = |Fj|$.

III. RELEVANCE FEEDBACK

The perception of human on image similarity is subjective, semantic and job-dependent. Although methods of (CBIR) provide challenging directions for retrieval of image, generally, the results based on the similarities of pure features are not inevitably semantically meaningful. Also, each type of visual feature captures only certain aspect of image property and it usually becomes difficult to specify how different aspects are combined. With relevance feedback, [12] the link between low-level and high-level features. Is established. It is a supervised learning mechanism which is used to improve the accuracy of CBIR systems. To improve system performance, the main idea lies in using positive and negative examples from the user. For a given query image, firstly the system retrieves a set of ranked images according to a predefined similarity measure. In the second step, the user labels the retrieved images as relevant (positive examples) or not relevant (negative examples). Then the system refines the retrieval results based on the relevance feedback and presents a new list of images to the user. Hence, the main point in relevance feedback is incorporating both the examples to refine the query and/or adjusting the similarity metrics.

IV. CONCLUSION

There are several methods available in CBIR for extraction of feature In this paper, these are identified and explored to understand the extraction and retrieval process in the CBIR systems. From the several experiment performed it is clear that the method based on the mixed combination of color, texture and shape features have higher retrieval accuracy compared to the other methods based on single feature. It is difficult to say that one method is superior to others. To combine color descriptors produces better retrieval rate as compared to one color descriptors. The combination of Colour histogram and color moments features can be combined to produce better results. Colour histograms and coherence vector can be combined retaining advantages of histograms with the spatial layout. Similarly, texture feature can be combined with colour histogram to get efficient results for image retrieval There is a considerable increase in retrieval rate when mixed combination of colour and texture features are used.

REFERENCES

[1] R.Datta, D. Joshi, J.Li, J.Z. Wang, " Image retrieval: ideas, influences, and trends of the new age ", ACM Computing Surveys 40(2), 2008, pp. 1-60.
[2] V.N. Gudivada, and V.V. Raghavan, “Content based image retrieval systems", IEEE Computer, Vol. 28, No. 9, 1995, pp.18-22.
[3] Guoyong Duan, Jing Yang, Yilong Yang: Content-Based Image Retrieval Research, 2011 International Conference on Physics Science and Technology (ICPST 2011)
[4] J.L. Shih, and L.H. Chen, “Color image retrieval based on primitives of color moments”, IEEE Proceedings online no.20020614., 2002.
[5] G.Pass, and R. Zabith, "Comparing images using joint histograms," Multimedia Systems, Vol.7, pp.234-240, 1999.
[6] H.A. Moghadam, T. Taghizadeh, A.H. Rouhi, M.T.Saadatmand, “Wavelet correlogram: a new approach for image indexing and retrieval”, J. Elsevier Pattern Rec.38(2005), pp. 2506 – 2518.
[7] B.S. Manjunath, and W.Y. Ma, " Texture Features for browsing and retrieval of image data", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, 1996, pp. 837-842.
[8] P. Howarth, and S. Ruger, “Evaluation of texture features for content based image retrieval”, Image and video retrieval. Springer LNCS 3115: pp. 326-334.
[9] J. E. Gary, and R. Mehrotra, "Shape similarity-based retrieval in image database systems," Proc. of SPIE, Image Storage and Retrieval Systems, Vol. 1662, pp. 2-8, 1992.
[10] D. Tegolo, "Shape analysis for image retrieval," Proc. of SPIE, Storage and Retrieval for Image and Video Databases -II, no. 2185, San Jose, CA, pp. 59-69, February 1994.
[11] Source: Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng,Fundamentals Of Content-Based Image Retrieval
[12] Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns " IEEE, and Vincent S. Tseng, Member, Vol. 23, No. 3, March 2011