

Content-Based Retrieval Using Color, Texture, and Shape Information

Ryszard S. Choraś

Faculty of Telecommunications & Electrical Engineering
University of Technology & Agriculture
S. Kaliskiego 7, 85-791 Bydgoszcz
choras@mail.atr.bydgoszcz.pl

Abstract. Current technology allows the acquisition, transmission, storing, and manipulation of large collections of images. A way to achieve this goal is the automatic computation of features such as color, texture, shape, and position of objects within images, and the use of the features as query terms.

In this paper we describe some results of a study on similarity evaluation in image retrieval using shape, texture, color and object orientation and relative position as content features. A simple system is also introduced that computes the feature descriptors and performs queries.

1 Introduction

Current technology allows the acquisition, transmission, storing, and manipulation of large collections of images. Content based information retrieval is now a widely investigated issue that aims at allowing users of multimedia information systems to retrieve images coherent with a sample image [1]. A way to achieve this goal is the automatic computation of features such as color, texture, shape, and position of objects within images, and the use of the features as query terms.

Content-based retrieval can be divided in the following steps:

Preprocessing: The image is first processed in order to extract the features, which describe its contents. The processing involves filtering, normalization, segmentation, and object identification. The output of this stage is a set of significant regions and objects.

Feature extraction: Features such as shape, texture, color, etc. are used to describe the content of the image. Image features can be classified into primitive. We can extract features at various levels.

The basic image retrieval system based on this concept is shown in Figure 1.

The main difference between our system and other is the manner in which similarity between a query image and images in a database is computed. For query images, we first compute ROI (Region of Interest) and extract a set of

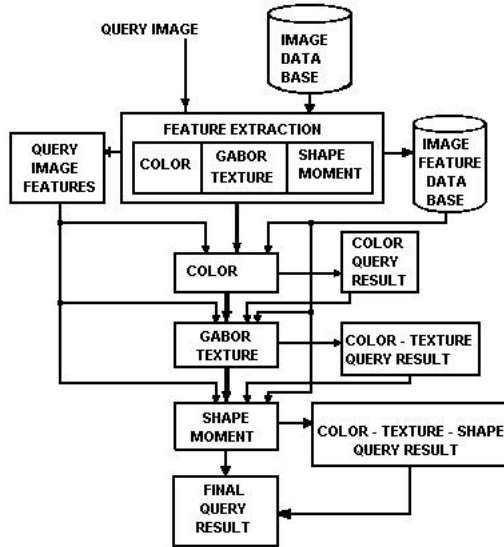


Fig. 1. Image retrieval system

color, texture and shape features by applying color histogram computation, Gabor texture extraction and shape parameters computation. The query is processed using color features computation unit, next the Gabor texture unit uses as input the query results of the color features computation unit. Gabor texture unit compares the texture information of the images and discards the images whose color information is similar to that of the query image but the texture information is much different from that of the query image. Next, the shape parameters computation unit is applied to the query results in this stage. The last step is the final query results.

2 Feature Extraction

2.1 Color Features

We propose a new color feature called color correlogram which describes the global distribution of local spatial correlations of colors and the size of this feature is fairly small [2,3,4].

For a pixel $p = (x, y) \in F$, let $F(p)$ denote its color. The histogram h of F is defined for $i \in [c]$ where c is number of colors e.g. c_1, \dots, c_c as

$$h_{c_i}(F) = N \cdot M \Pr_{p \in F}[p \in F_{c_i}] \quad (1)$$

$\frac{h_{c_i}(F)}{N \cdot M}$ gives the probability that the color of the pixel is c_i .

The correlogram of F is defined for $i, j \in [c], k \in [d]$ as

$$\gamma_{c_i c_j}^{(k)}(F) = \Pr_{\substack{p_1 \in F_{c_i} \\ p_2 \in F}} [p_2 \in F_{c_j} \mid |p_1 - p_2| = k] \tag{2}$$

Given any pixel of color c_i in the image, $\gamma_{c_i c_j}^{(k)}$ gives the probability that a pixel at distance k away from the given pixel is of color c_j . To compute the distance between images F and F' , we compare histograms and correlograms

$$|F - F'|_h = \sum_{i \in [c]} \frac{|h_{c_i}(F) - h_{c_i}(F')|}{1 + h_{c_i}(F) + h_{c_i}(F')} \tag{3}$$

$$|F - F'|_\gamma = \sum_{\substack{i, j \in [c] \\ k \in [d]}} \frac{|\gamma_{c_i c_j}^{(k)}(F) - \gamma_{c_i c_j}^{(k)}(F')|}{1 + \gamma_{c_i c_j}^{(k)}(F) + \gamma_{c_i c_j}^{(k)}(F')} \tag{4}$$

Given the histograms for a template T and an image F , the intersection of these two histograms is defined as

$$H_{c_i}(T \cap F) = \min\{H_{c_i}(T), H_{c_i}(F)\} \tag{5}$$

and

$$h_{c_i}(T \cap F) = \frac{H_{c_i}(T \cap F)}{|T|} \tag{6}$$

The intersection correlogram is defined as the correlogram of the intersection $T \cap F$. The intersection correlogram is defined as

$$\gamma_{c_i c_j}^{(k)}(T \cap F) = \frac{\Gamma_{c_i c_j}^{(k)}(T \cap F)}{H_{c_i}(T \cap F) \cdot 8k} \tag{7}$$

where

$$\Gamma_{c_i c_j}^{(k)}(F) = |\{p_1 \in F_{c_i}, p_2 \in F_{c_j} \mid |p_1 - p_2| = k\}| \tag{8}$$

2.2 Gabor Features

The differential structure of an image is completely extracted by the convolution with the Gaussian filter family. We use Gabor filters in our image retrieval system. This results in the family of Gabor filters covering the total spatial frequency plane nearly uniformly. Filtering an image with Gabor kernel can be interpreted as local Fourier analysis. The known good characteristics of Gabor filters for image analysis can be justified in scale space framework [5,6]. Gabor filters are used in analysing the property of an object in the selected image because they have optimal joint localization (resolution) in both the spatial and the spatial frequency domains. The frequency tuning of filters allows an axiomatic

characterization of Gabor filters being the linear, shift invariant family of transformations which is (i) parameterized by a scale parameter with a semi-group structure, (ii) is scale invariant i.e. the function that relates the observable is independent of the choice of dimensional units.

Gabor functions are Gaussians modulated by complex sinusoids. In its general form, the two-dimensional Gabor function and its Fourier transform can be written as [7,8]

$$h(x, y) = g(x, y) \exp(j2\pi Fx') \tag{9}$$

F the radial center frequency and $g(x, y)$ is the 2D Gaussian

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x}{\sigma_x} \right)^2 + \left(\frac{y}{\sigma_y} \right)^2 \right] \right\} \tag{10}$$

where (σ_x, σ_y) characterize the spatial extent and bandwidth of Gabor filter $h(x, y)$.

The aspect ratio of $g(x, y)$ is given by $\lambda = \sigma_y/\sigma_x$, which gives a measure of filter's symmetry. In the frequency domain,

$$H(u, v) = \exp \left\{ -2\pi^2\sigma^2 \left[(u - F)^2 \lambda^2 + (v)^2 \right] \right\} \tag{11}$$

The set of self-similar Gabor filters is obtained by appropriate rotations and scalings of through the generating function:

$$g_{mn}(x, y) = a^{-m}g(x', y') \quad a > 1 \quad m, n = integer \tag{12}$$

where

$$(x', y') = (a^{-m}[x \cos \theta + y \sin \theta], a^{-m}[-x \sin \theta + y \cos \theta]) \tag{13}$$

where a is the scale factor, $n = 0, 1, \dots, K - 1$ is the current orientation index, K is the total number of orientations, $m = 0, 1, \dots, S - 1$ is the current scale index, S is the total number of scales, and $\theta = \frac{n\pi}{K}$.

The scale factor a^{-m} in equation (12) ensures that the filter energy is independent of m .

$$E_{mn} = \int \int |g_{mn}(x, y)|^2 dx dy \tag{14}$$

$a = \left(\frac{f_h}{f_l} \right)^{\frac{-1}{S-1}}$, $F = f_h$, f_l and f_h are the lower and upper center frequencies of interest. In our implementation f_l and f_h are equal respectively $f_l = 0.05$ and $f_h = 0.4$ and $a = 2$.

Gabor filtered output of the image is obtained by the convolution of the image with Gabor function for each of the orientation/spatial frequency (scale) orientation. Given an image $I(x, y)$, we filter this image with $g_{mn}(x, y)$

$$G_{mn} = \sum_k \sum_l I(x - k, y - l)g_{mn}^*(x, y) \tag{15}$$

where * indicates the complex conjugate.

After applying Gabor filters on the image we obtain an array of magnitudes

$$E_{mn} = \sum_x \sum_y |G_{mn}(x, y)| \tag{16}$$

The magnitudes of the Gabor filter responses are represented by three moments:

- the mean μ_{mn}

$$\mu_{mn} = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N G_{mn}(x, y) \tag{17}$$

- the standard deviation σ_{mn}

$$\sigma_{mn} = \sqrt{\sum_{x=1}^M \sum_{y=1}^N [|G_{mn}(x, y)| - \mu_{mn}]^2} \tag{18}$$

- the skewness κ_{mn}

$$\kappa_{mn} = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N \left(\frac{G_{mn}(x, y) - \mu_{mn}}{\sigma_{mn}} \right)^3 \tag{19}$$

The feature vector (FV) is represented as follows

$$FV = [\mu_{11}, \sigma_{11}, \kappa_{11} \dots, \mu_{SK}, \sigma_{SK}, \kappa_{SK}] \tag{20}$$

The similarity of a query image Q and a image T in the database is defined as T where

$$d^{(Q)(T)}(\mu, \sigma, \kappa) = \sum_m \sum_n d_{mn}^{(Q)(T)} \tag{21}$$

where

$$d_{mn}^{QT} = \left| \frac{\mu_{mn}^{(Q)} - \mu_{mn}^{(T)}}{\Xi(\mu_{mn})} \right| + \left| \frac{\sigma_{mn}^{(Q)} - \sigma_{mn}^{(T)}}{\Xi(\sigma_{mn})} \right| + \left| \frac{\kappa_{mn}^{(Q)} - \kappa_{mn}^{(T)}}{\Xi(\kappa_{mn})} \right| \tag{22}$$

where $\Xi(\mu_{mn})$, $\Xi(\sigma_{mn})$ and $\Xi(\kappa_{mn})$ are respectively mean ,the atandard deviation and the skewness of the transform coefficients over the database.

The number of scales chosen is 4 and orientations are 6. Thus 24 Gabor filters are used in the experiments, which give 48 dimensional feature vector texture classification. The proposed features are found to give 94.35% correct classification rates.

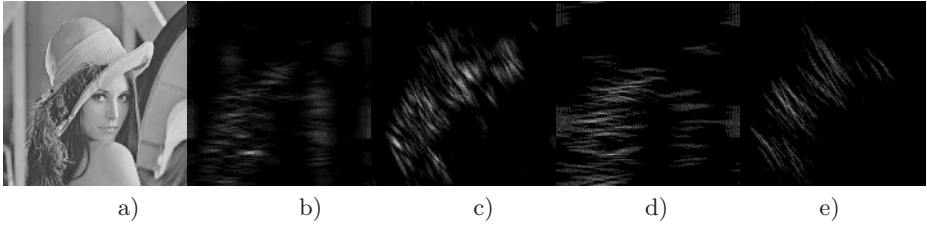


Fig. 2. Texture image a). The power spectrum of the Gabor transform with 0° (respectively b) and d)) and 60° (respectively c) and e)) orientation for various scale.

2.3 Shape Features

Basically, shape based image retrieval is the measuring of similarity between shapes represented by their features. Shape is an important visual feature and it is one of the primitive features for image content description. However, shape content description is a difficult task because it is difficult to define perceptual shape features and measure the similarity between shapes. To make the problem more complex, shape is often corrupted with noise, deflection, arbitrary distortion and occlusion. Therefore, two steps are essential in shape based image retrieval, they are, feature extraction and similarity measurement between the extracted features.

To characterize the shape we used following descriptors: principal axis ratio, compactness, circular variance which are translation, rotation and scale invariant shape descriptors, and seven Hu moments [9,10].

The principal axes ratio (*par*)

$$par = \frac{c_{yy} + c_{xx} - \sqrt{(c_{yy} - c_{xx})^2 - 4(c_{xx}c_{yy} - c_{xy}^2)}}{c_{yy} + c_{xx} + \sqrt{(c_{yy} - c_{xx})^2 - 4(c_{xx}c_{yy} - c_{xy}^2)}} \tag{23}$$

where covariance matrix *C* of a contour is defined

$$C = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{xy} & c_{yy} \end{bmatrix} \tag{24}$$

Compactness (*comp*) is the ratio of the perimeter of a circle with equal area as the original object and the perimeter of original contour

$$comp = \frac{P_{circle}}{P} = \frac{2\sqrt{A_{circle}\pi}}{P} \tag{25}$$

Circular variance (*cv*) is the proportional mean-squared error with respect to a solid circle

$$cv = \frac{1}{N\mu_r^2} \sum_i (||p_i - \underline{\mu}|| - \mu_r)^2 \tag{26}$$

where N is the number of contour points, $p_i = (x_i, y_i)^T$ is the contour point, μ is the centroid and μ_r is the mean radius of the contour.

A object can be represented by the spatial moments of its intensity function. In the spatial case

$$m_{pq} = \sum_{x=1}^m \sum_{y=1}^n x^p y^q f(x, y) \tag{27}$$

The central moments are given by

$$m_{pq} = \sum_{x=1}^m \sum_{y=1}^n (x - X)^p (y - Y)^q f(x, y) \tag{28}$$

where (X, Y) are

$$X = \frac{m_{10}}{m_{00}} \quad \text{and} \quad Y = \frac{m_{01}}{m_{00}} \tag{29}$$

Normalized central moment μ_{pq}

$$\mu_{pq} = \frac{m_{pq}}{(m_{00})^\alpha} \quad \alpha = \frac{p + q}{2} + 1 \tag{30}$$

Using nonlinear combinations of the lower order moments, a set of moment invariants (usually called geometric moment), which has the desirable properties of being invariant under translation, scaling and rotation, are derived. Hu [11] employed seven moment invariants, that are invariant under rotation as well as translation and scale change, to recognize characters independent of their position size and orientation.

$$\begin{aligned} \phi_1 &= \mu_{20} + \mu_{02} \\ \phi_2 &= [\mu_{20} - \mu_{02}]^2 + 4\mu_{11}^2 \\ \phi_3 &= [\mu_{30} - 3\mu_{02}]^2 + [3\mu_{21} - \mu_{03}]^2 \\ \phi_4 &= [\mu_{30} + \mu_{12}]^2 + [\mu_{21} + \mu_{03}]^2 \\ \phi_5 &= [\mu_{30} - 3\mu_{12}][\mu_{30} + \mu_{12}] \times [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + \\ &+ [3\mu_{21} - \mu_{03}][\mu_{21} + \mu_{03}] \times [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ \phi_6 &= [\mu_{20} - \mu_{02}][(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}[\mu_{30} + \mu_{12}][\mu_{21} + \mu_{03}] \\ \phi_7 &= [3\mu_{21} - \mu_{03}][\mu_{30} + \mu_{12}] \times [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &- [\mu_{03} - 3\mu_{12}][\mu_{21} + \mu_{03}] \times [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \tag{31}$$

To characterize the shape we used a feature vector

$$SFV = (\phi_1, \phi_2, \dots, \phi_7, par, comp, cv) \tag{32}$$

consisting of the seven moment invariants, principal axis ratio, compactness and circular variance descriptors. This vector is used to index each shape in the database. The distance between two feature vectors is determined by city block distance measure.

3 Conclusion

A retrieval methodology which integrates color, texture and shape information is presented in this paper. Consequently, the overall image similarity is developed through the similarity based on all the feature components. Experimental evaluation based on our image database shows that our method promisingly outperforms the retrieval systems from the literature.

References

- [1] M. S. Lew(ed), *Principles of Visual Information Retrieval* (London: Springer-Varlag, 2001).
- [2] M. J. Swain and D. H. Ballard, Color Indexing, *International Journal of Computer Vision*, 7(1), 1991, 11-32.
- [3] A. Drimbarean and P.F. Whelan, Experiments in colour texture analysis, *Pattern Recognition Letters*, 22, 2001, 1161- 1167
- [4] T. Gevers and A.W.M. Smeulders, Colour based object recognition, *Pattern Recognition*, 32, 1999, 453-464.
- [5] George M. Haley and B. S. Manjunath, Rotation-Invariant texture classification using a complete space-frequency model, *IEEE Transactions on Image Processing*, 8(2), 1999, 255-269.
- [6] A. Bovik, M. Clark, W. Geisler, Multichannel texture analysis using localized spatial filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(1), 1990, 55-73.
- [7] A. K. Jain and F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, *Pattern Recognition*, 24(12), 1991, 1167-1186.
- [8] B. S. Manjunath and W. Y. Ma, Texture features for browsing and retrieval of data, *IEEE Trans. Pattern Analysis Machine Intelligence*, 18(8), 1996, 837-842.
- [9] F. Mokhtarian and A. Mackworth, Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1), 1986, 34-43.
- [10] E. Persoon and K. S. Fu, Shape Discrimination Using Fourier Descriptors, *IEEE Trans. On Systems, Man and Cybernetics*, SMC-7(3), 1977, 170-179.
- [11] M. Hu, Visual pattern recognition by moment invariants, *IEEE Trans. on Inf. Theory*, 8, 1962, 179-187.
- [12] R. T. Hu, The revised fundamental theorem of moment invariants, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 1991, 830-834.
- [13] T. N. Tan, Rotation Invariant Texture Features and Their Use in Automatic Script Identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7), 1998, 751-756.