Combining texture, shape and spatial information for image retrieval

A. Georgakis, M. E. Osadebey and H. Li
Digital Media Laboratory (DML)
Dept. of Applied Physics and Electronics
Umeå University, Umeå SE-90187, Sweden.
apostolos.georgakis@tfe.umu.se

Abstract

Most Content-Based Image Retrieval (CBIR) systems employ color as primary feature with texture and shape as secondary features. Very few systems exploit spatial features. None of the available systems combines all three visual features, texture, shape and location, for organization and retrieval. Moreover relatively few systems use Gabor filters in texture extraction, despite the widely acclaimed efficiency; Gabor filters are confined only in pure texture images. In this paper a simple, robust, flexible and effective image retrieval system is presented. The proposed system uses weighted combination of integrated Gabor texture features, shape features of texture regions and spatial information features of the texture regions.

1 Introduction

In the computer age virtually all domains of human life including commerce, government, academics, crime prevention, engineering, etc are in need of, and use of images for efficient services. An active research area that is in the heart of all the above domains is the *extraction* or *retrieval* of images from a database given a query (image retrieval system). An image retrieval system is a system for searching and retrieving images from a large database of digital images. The most common method of image retrieval utilize some method of annotation such as keywords, or descriptions to the images so that retrieval can be performed over the labels. Unfortunately manual annotation is time-consuming, laborious and expensive. The answer to the previous difficulty is termed CBIR.

CBIR describes the process of retrieving desired images from the image database on the basis of syntactical image features. Early research comprise of systems such as TRADEMARK [Kat92], QBIC [NBE+93] and Photobook [PPS94]. More recent research can be found in [SWS+00]. Most current CBIR techniques are geared towards retrieval by some aspect of image appearance, depending on the automatic extraction and comparison of image features judged most likely to convey that appearance. The features most often used include color, texture, shape, spatial information and multi-resolution pixel intensity transformations such as wavelets or multi-scale Gaussian filtering.

2 Proposed system

2.1 Texture

Texture is a repeated pattern of pixels over the spatial domain; the pattern can possibly be contaminated by noise. The repetition frequencies might result in textures that appear to be random and unstructured. Texture are the visual patterns in an image that have properties of homogeneity that do not result from the presence of only a single color or intensity. Today, the most commonly used methods for texture feature description are statistical and transform based methods. In the present work a transformed based method is used.

The state-of-the-art in transformed based texture feature extraction uses Gabor wavelets. This is due to physiological research evidence that Gabor filters model the neurons in the visual cortex of the human visual system. Furthermore, Manjunath *et al.* in [MM96] showed that Gabor features performs better than using pyramid-structured, tree-structured wavelet transform features and multi-resolution simultaneous autoregressive model.

A total of twenty four wavelets were generated from the "mother" Gabor function using four scales of frequency and six orientations. Redundancy, which is the consequence of the nonorthogonality of Gabor wavelets, was addressed by choosing the parameters of the filter bank to be set of frequencies and orientations that cover the entire spatial frequency space so as to capture texture information as much as possible. The lower and upper frequencies of the filters were set at 0.04 octaves and 0.5 octaves respectively, the orientations were at intervals of 30 degrees, and the half-peak magnitudes of the filter responses in the frequency spectrum are constrained to touch each other.

Each image $I(x,y)$ in the database is convolved with each wavelet in the filter bank according to the convolution equation $G_{mn}(x,y) = \int I(x-s, y-t)\,\bar{\phi}(s,t)$ where s and t are the dimensions of the filter and $\bar{\phi}(s,t)$ is the complex conjugate of the Gabor wavelet. Furthermore, $m \in \{1,2,3,4\}$, $n \in \{1,2,\ldots,6\}$ correspond to the scales of frequency and orientations respectively.

By assuming spatial homogeneity of texture regions the mean and the std. deviation of the magnitude of the transformed coefficients was computed according to:

$$\mu_{mn} = \int |G_{mn}(x,y)|\,dxdy,$$

$$\sigma_{mn} = \sqrt{\int\int \left(G_{mn}(x,y) - \mu_{mn}\right)^2 dxdy}.$$

Finally, the texture feature vector for each image is constructed using the computed values for the mean μ_{mn} and std. deviation σ_{mn} according to:

$$I_{fr}(x,y) = [\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \ldots, \mu_{35}, \sigma_{35}].$$

2.2 Shape

Shape of an object is the characteristic surface configuration as represented by the outline or contour. Shape recognition is one of the modes through which human perception of the environment is executed. Shape is important in CBIR systems because it corresponds to region of interests in images. In CBIR system designed for specific domain such as trademarks and silhouettes of tools, shape segmentation can be automatic and effective. However this is not the case for a system having heterogeneous database. In this case shape segmentation may be difficult or sometimes impossible.

In our proposal the shape features are extracted using local mean and std. deviation in a search 5×5 neighborhood employing the following formulas:

$$\mu_I(x,y) = \sum_{\substack{g=-2 \\ h=-2}}^{2} I(x-g, y-h) \times Z(5,5),$$

$$\sigma_I(x,y) = \sum_{\substack{g=-2 \\ h=-2}}^{2} I(x-g, y-h) \times K(5,5)$$

where Z and K are the impulse responses of the mean and std. deviation filters respectively. For that, each image in the database passes through a 5×5 grid size Gabor filter bank. Twenty four output images are then obtained. Afterwards, similarly to the pre-filtered image the local mean and std. deviation of each output of the filter bank is also calculated using a 5×5 neighborhood according to:

$$\mu_{pmn}(x,y) = \sum_{\substack{g=-2 \\ h=-2}}^{2} G_{pmn}(x-g, y-h) \times Z(5,5),$$

$$\sigma_{pmn}(x,y) = \sum_{\substack{g=-2 \\ h=-2}}^{2} G_{pmn}(x-g, y-h) \times K(5,5)$$

Thus for each pixel in the image there are twenty four reference pixels. Consider the pixel in the image and the twenty four corresponding pixels in the output of the filter bank, the distance between the image feature vector and any of the corresponding pixels feature vector is computed as:

$$D_{pmn}(x,y) = |\mu_{pmn} - \mu_I| + |\sigma_{pmn} - \sigma_I|$$

The computed distance is a measure of similarity of the texture between a pixel in the original image and the corresponding pixel at the output of the Gabor filter bank. For every pixel for each distance $D_{pmn}(x,y)$ with corresponding pixels in the output of the Gabor filter, the pixel whose texture is most likely similar to that of the database image is

$$G_{qmn}(x,y) \approx I(x,y),\ q = arg \min_{i=1}^{24} \{D_{imn}(x,y)\}$$

This results in a filtered image called "texture classified image."

2.3 Spatial information

Spatial information is the spatial relationship existing among properties characterizing image regions within the image and is commonly used to

Figure 1: The GUI of the proposed CBIR system.

addresses the problem of discriminating similar images in homogeneous or non databases. Feature like centroid, area and other geometric properties of local image regions are prime location candidates; they are also the basis for deriving spatial layout or information.

The two region properties adopted to describe the spatial features are elementary spatial feature descriptors, centroid and spatial extent. This is because the image database is complex and heterogeneous. Though elementary features, the spatial descriptors are invariant to rotation and translation. The steps taken in spatial information extraction are as follows: (i) Compute the centroid distances of regions in the binary image obtained from the texture segmented image, (ii) Thereafter the spatial extents of the texture regions are also separately computed. Finally, the number of elements for each region property was normalized to 100.

2.4 The CBIR system

Have the above feature descriptors as the basis for CBIR we proceed to develop the system. Figure 1 depicts the GUI of the system which is used by the user when querying the database. In doing so the user can select and image and then select whether to use any of the individual features alone or combined together using the weights in the upper right corner.

3 Experimental result

The proposed CBIR system was tested using the test database of University of Washington [Was] that consists of 977 images. The images are manually annotated in 110 categories. The performance of the proposed system is assessed using recall-precision curves with the help of randomly selected query-images [Kor97].

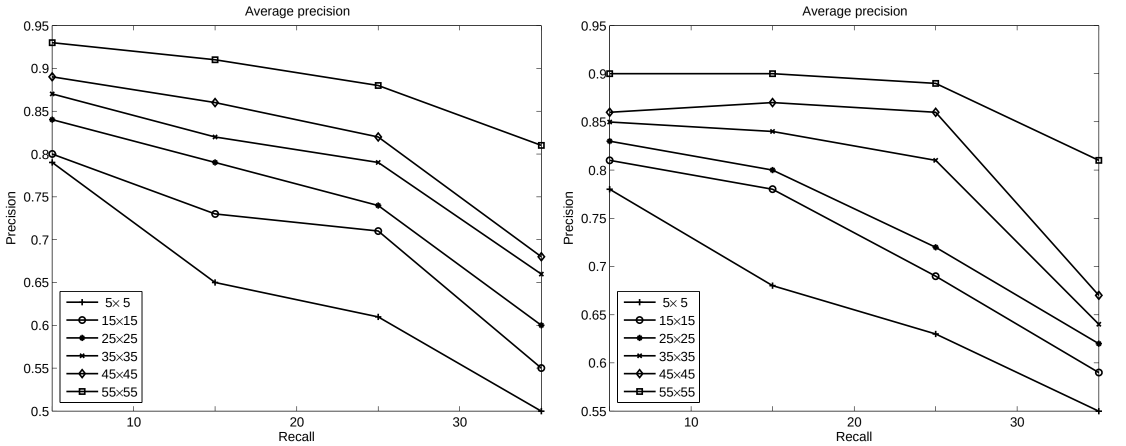
Recall is defined as the fraction of relevant objects that are retrieved whereas *precision* is the fraction of retrieved objects that are relevant to the query. In the case under consideration the relevance or not of an image to a query-image is assessed using the annotation assigned to each individual image.

Implementing a CBIR system is a painstaking process. The reason is that the Gabor filter dictionary adopted for the system design indicates the frequency of operation and the number of filters for optimal performance but it does not have a readymade answer for the filter grid size that gives optimal performance. On that, it is generally acceptable that larger Gabor grids are capable of capturing slowly varying levels than a lower grid size filter. Therefore, this aspect has been taken care by computing texture feature of the images in the database using Gabor filters of grid sizes 5×5, 15×15, 25×25, 35×35, 45×45 and 55×55.

In assessing the performance of the proposed system a series of query-images are given to the system and then one precision curve is computed per query. Afterwards all the curves are averaged and this yields the so-called *average recall-precision* curve. Figure 2 depicts the average curve for the CBIR system when only texture features were used. On the other hand Fig. 3(a) and 3(b) depict the average curves when all three features are utilized. In the case of Fig. 3(a) the weights assigned to texture, shape and spatial features are 0.8, 0.1 and 0.1 respectively whereas in Fig. 3(b) the weights are 0.7, 0.15 and 0.15 respectively.

4 Conclusion

In this paper a novel CBIR system has been presented. The system puts emphasis on texture, which is the primary feature, and less emphasis on shape and spatial as secondary features. Texture features derived from six grid sizes of independent and different Gabor filter banks were incorporated into the CBIR system. The results show that the combination of all three features provides higher performance than anyone individually. In the near future we are planning to add an extra module in order to allow for the user of the system to provide feed-back to the query along. Furthermore, soon we will have a web based demo of our system.



(a) (b)

Figure 3: The average precision when all three features are utilized with weights: (a) 0.8, 0.1 and 0.1, (b) 0.7, 0.15 and 0.15.

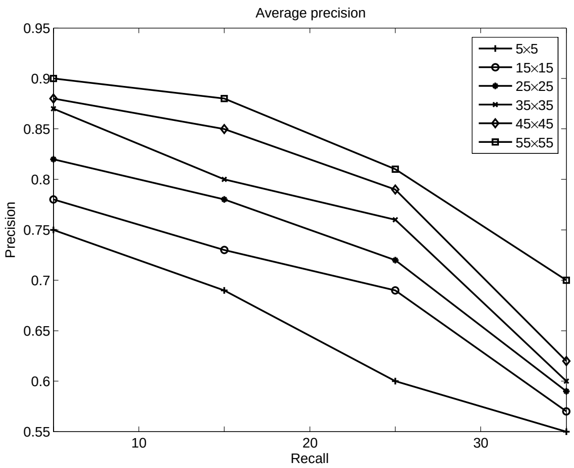


Figure 2: The average precision of the CBIR system when using only texture.

5 Acknowledgment

This work was financed by a research grant from the Faculty of Science, Umeå University.

References

[Kat92] T. Kato. Database architecture for content-based image retrieval. In A.A. Jambardino and W.R. Niblack, editors, *Proc. SPIE 1662*, volume Image Storage and Retrieval Systems, pages 112–123. Image Storage and Retrieval System, 1992.

[Kor97] R. R. Korfhage. *Information Storage and Retrieval.* NY: J. Wiley, 1997.

[MM96] B. S. Manjunath and W. Y. Ma. Texture features for Browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[NBE+93] C. W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: querying images by color, texture and shape. Technical report, IBM Research Report RJ-9203, 1993.

[PPS94] A. Pentland, W. Picard, and S. Sclaroff. Photobook - tools for content-based manipulation of image databases. In *Storage and Retrieval for Image and Video Databases II*, pages 34–47. Proc SPIE 2185, 1994.

[SWS+00] 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 Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[Was] University Washington. Groundtruth database. `http://www.cs.washington.edu/research/imagedatabase/groundtruth/`.