A Composite Descriptor for Shape Retrieval

Atul Sajjanhar[1†], Guojun Lu[2], Dengsheng Zhang[2], Wanlei Zhou[1]

[1]*School of Engineering and Information Technology*
Deakin University
221 Burwood Highway
Burwood, VIC 3125
Australia
{atuls, wanlei}@deakin.edu.au

[2]*Gippsland School of Information Technology*
Monash University
Northways Road
Churchill, VIC 3842
Australia
{guojun.lu, dengsheng.zhang}@infotech.monash.edu.au

Abstract

In this paper, a composite descriptor for shape retrieval has been proposed. The proposed descriptor is obtained from Generic Fourier Descriptors (GFD) for the shape region and the shape contour. A composite descriptor derived from GFD of the shape region and the shape contour is used for indexing and retrieval of shapes. Difference between two images is computed as the Euclidean distance between their composite descriptors. Experiments are performed to test the effectiveness of the proposed descriptor for retrieval of 2d images. Sets of composite descriptors, obtained by assigning different weights to the region component and the contour component, are also evaluated. Item S8 within the MPEG-7 Still Images Content Set is used for performing experiments; this dataset consists of 3621 still images. Experimental results show that the proposed descriptor is effective.

1. Introduction

Approaches for shape representation and retrieval can be broadly classified into contour based and region based. Some of the region based methods are geometric moments, moments constructed from orthogonal functions and grid based method [1]. Recently, Generic Fourier Descriptors (GFD) was proposed by Zhang and Lu [2] for region based matching of shapes. Some of the contour based methods are polygonal approximation, autoregressive model, Fourier Descriptors and distance histograms [1].

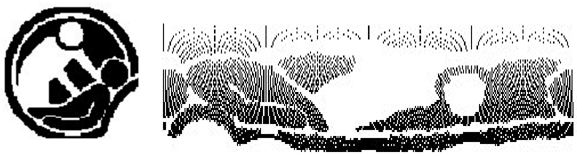
In this paper, a composite descriptor based on a region based method and a contour based method is proposed for 2d shape retrieval. The process of obtaining the contour descriptor is twofold. First, the concept of *connectivity* is introduced and it is shown how *connectivity* can be used to obtain the contour of a 2d image. Second, contour-based GFD is used for representation of the contour obtained apriori. Region-based GFD is also computed for the 2d shape [2]. A composite descriptor is constructed from the contour descriptor and the region descriptor.

It has been shown that the performance of GFD is comparable with other contemporary techniques [2]. Hence, we compare the proposed descriptor with GFD. GFD is described in Section 2. The proposed method is described in Section 3. Experimental Setup and Results are presented in Section 4. Finally, Discussion and Conclusion are presented in Sections 5 and 6 respectively.

† PhD candidate, GSIT, Monash University, Australia

2. Generic Fourier Descriptors

Generic Fourier Descriptors (GFD) is a region-based method for image retrieval [1][2]. In GFD, feature vectors are created by extracting spectral information in the frequency domain. Fourier transform is applied to the polar raster sampled shape image. Consider the image shown in Figure 1. To obtain the GFD for the image, the image is first plotted in polar space. The polar image of Figure 1(a), is shown in Figure 1(b).



(a) (b)

Figure 1. (a) An Image in Cartesian Coordinates (b) Polar Image

Before obtaining the polar image, the image is normalized for scale. 2d DFT is applied to the rectangular region in polar coordinates to obtain Fourier coefficients which are used to construct feature vectors for shape representation and similarity measure [1][2].

Polar coordinates (r, θ) are obtained from the 3d Cartesian co-ordinates (x, y) as shown below.

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \qquad (1)$$

$$\theta = \arctan\left(\frac{y - y_c}{x - x_c}\right) \qquad (2)$$

where, (x_c, y_c) is the centroid of the 2d Cartesian image.

Feature vectors are constructed from the polar coordinates by computing the 2d DFT. 2d DFT of the polar coordinates is defined as below.

$$PF(\rho, \tau) = \sum_r \sum_\theta f(r, \theta) e^{-j2\pi\left(\frac{r}{R}\rho + \frac{\theta}{T}\tau\right)} \qquad (3)$$

where, R and T is the radial and angular resolution. r, θ is obtained from Eqn. 1 and Eqn. 2. Feature vectors are represented as shown below.

$$F : \left(\frac{F(0,1)}{F(0,0)}, \ldots \frac{F(0,T-1)}{F(0,0)}, \ldots \frac{F(1,0)}{F(0,0)}, \ldots \frac{F(R-1,T-1)}{F(0,0)} \right)$$

where, R and T is the radial and angular resolution as used in Eqn. 3.

The difference between two images is computed as the Euclidean distance between their feature vectors as shown in Eqn. 4.

$$Dist\,(F_1, F_2) = \sqrt{\sum_{i=0}^{RT-1} \left(f_{1,i} - f_{2,i}\right)^2} \qquad (4)$$

where, $f_{x,i}$ is a descriptor within the feature vector of image *x*. $0 < i < RT$, where R, T is the radial and angular resolution.

3. Proposed Method

First, we introduce connectivity which was proposed by Sajjanhar et al [4]. An analogy is drawn from Color Coherence Vectors (CCV) proposed by Pass and Zabih [3]. CCV is used for image retrieval based on color. Pass et al [3] defined color coherence of pixels as the degree to which pixels of that color are members of a large similarly colored region. Pixels are classified as coherent or incoherent. Coherent pixels are part of a sizable contiguous region of similar color while incoherent pixels are not. In the case of shape representation, *connectivity* of pixels is defined. The state of the nearest 8-neighbours is computed for each *OFF* pixel. An *OFF* pixel is a dark pixel i.e. has intensity below a predefined threshold. Connectivity of an *OFF* pixel is obtained as the number of *OFF* pixels amongst the nearest 8-neighbours. The connectivity is obtained by convolving the image with a 3x3 mask. Figure 2 provides additional information for the image in Figure 1(a). Connectivity information is added in Cartesian coordinates. Hence, z-axis is obtained which provides information regarding connectivity of pixels. For each *OFF* pixel within the image, the connectivity can take values 0 through 8. A connectivity of 0 indicates that none of the nearest 8-neighbours are *OFF*. A connectivity of 8 indicates that all of the nearest 8-neighbours are *OFF*.

Connectivity for the image in Figure 1(a) is shown by the point cloud in Figure 2. The point cloud in 3d Cartesian coordinates contains connectivity information along the z-axis.

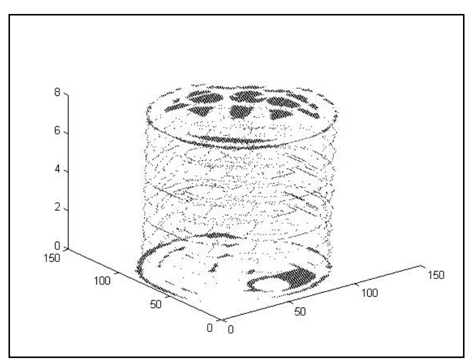


Figure 2. Connectivity Information for Image in Figure 1

The shape contour is obtained from the connectivity information as shown in Figure 2. Pixels with connectivity 0 through 7 are used to construct the shape contour. The shape contour for image in Figure 1(a) is shown in Figure 3(a) below. Pixels with connectivity of 8 are ignored when capturing contour information; these pixels are shown in Figure 3(b).

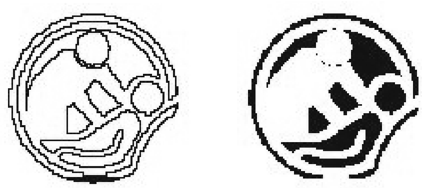


Figure 3. (a) Contour Information (b) Contour Complement

The advantage of capturing the contour as shown above is that the contour continuity is non-essential. Hence, contour propagation techniques are not required.

In the proposed method, each image is indexed using the GFD of the shape region and the GFD of the shape contour. During retrieval, images are ranked based on the two features. Ranking of images based on multiple features is a two step process. First, distances are computed for each feature thus giving a ranked list for each feature. Second, distances are combined to obtain an overall measure of distance, namely, global distance. When combining distances for different features, some features may contribute disproportionately to the distance measure because the scale of the distance will be different for different features. Celentano et al. [6] have proposed two methods to mitigate the disproportionate contribution by different features. The methods are *norm1* and *norm2*.

norm1: For each element in a ranked list of k elements having distance d_i from the query image, the normalized distance is:

$$d_i^{'} = \frac{d_i}{d_k} \tag{5}$$

where d_k is the distance for the lowest ranked element.

norm2: The normalised distance is:

$$d_i^{'} = \frac{d_i - d_1}{d_1 - d_k} \tag{6}$$

where d_1 and d_k are the distances for the highest and the lowest ranked elements.

After normalising the distances the global distance for an image which appears in n ranked lists may be computed from *mean1* or *mean2* which are computed as shown in Eqn. 7 and Eqn. 8.

$$d = \frac{d_1 + d_2 ... d_n}{n} \tag{7}$$

$$d = \frac{d_1 + d_2 ... d_n}{n^2} \tag{8}$$

Experimental results provided by Celentano et al. suggest that *norm1* and *mean1* perform better than *norm2* and *mean2*. A method similar to *norm1* has also been used by Jain et al [8].

Another method for integration of disparate feature vectors is used in the QBIC system [9][10][11]. The distance between two images is computed using a weighted Euclidean distance with the inverse of feature variances used for normalization. The distance between two images is computed as,

$$d_{ij} = \sum_k \frac{(k_i - k_j)^2}{\sigma^2} \tag{9}$$

where k is a feature and i and j are the two images.

4. Experimental Results

Three sets of experiments are conducted. In the first set of experiments, effectiveness of two methods for the integration of disparate feature vectors (namely, *Contour GFD* and *Region GFD*) are compared. In the second set of experiments the effectiveness of the composite descriptor for GFD is compared with traditional GFD. In the third set of experiments, the proposed descriptor is evaluated when different weights are assigned to *Contour GFD* within the composite descriptor.

Experiments are performed on Item number S8 within the MPEG-7 Still Images Content Set; this is a collection of trademark images which was originally provided by the Korean Industrial Property Office. S8 consists of 3621 still images. It is divided into sets A1, A2, A3, A4 to test the robustness of methods to geometric and perspective transformations.

Set A1 consists of 2881 shapes from the whole database, it is for test of scale invariance. 100 shapes in Set A1 are organized into 20 groups (5 similar shapes in each group) which can be used as queries for test of retrieval. In our experiment, all the 100 shapes from the 20 groups are used as queries to test the retrieval.

Set A2 consists of 2921 shapes from the whole database, it is for test of rotation invariance. 140 shapes in Set A2 are organized into 20 groups (7 similar shapes in each group) which can be used as queries for test of retrieval. In our experiment, all the 140 shapes from the 20 groups are used as queries to test the retrieval.

Set A3 consists of 3101 shapes from the whole database, it is for test of rotation/scale invariance. 330 shapes in Set A3 are organized into 30 groups (11 similar shapes in each group) which can be used as queries for test of retrieval. In our experiment, all the 330 shapes from the 30 groups are used as queries to test the retrieval.

Set A4 consists of 3101 from the whole database, it is for test of robustness to perspective transform. 330 shapes in Set A4 are organized into 30 groups (11 similar shapes in each group) which can be used as queries for test of retrieval. In our experiment, all the 330 shapes from the 30 groups are used as queries to test the retrieval.

The effectiveness of two methods to integrate Contour GFD and Region GFD are compared. The result for recall and precision for *norm1* and *QBIC method (variance based)*, as described in Section 3, are shown below. The results are averaged over 220 queries from the dataset described above. The queries are divided into 20 classes of shapes with each class containing 11 member shapes which are generated through geometric and perspective transformations. The results are shown in Table 1 where Method A refers to *norm1* method and Method B refers to *variance based method.* Weight refers to the weight assigned to the Contour GFD when computing the distance between two images. *norm1* method clearly outperforms the other method.

Table 1. Retrieval Performance of Two Methods to Integrate Feature Vectors

Method, Weight	Recall (%)				
	10	**20**	**30**	**40**	**50**
A, 0.1	100	100	100	100	100
B, 0.1	100	100	100	100	99.2
A, 0.4	100	100	100	100	100
B, 0.4	100	100	98.9	99.1	97.4
A, 0.8	100	100	97.7	96.2	96.1
B, 0.8	100	98.6	94.7	85.6	76.2

Method, Weight	Recall (%)				
	60	**70**	**80**	**90**	**100**
A, 0.1	100	100	99.4	81.3	58.0
B, 0.1	99.3	94.7	82.4	64.5	43.6
A, 0.4	100	99.4	96.5	73.3	49.3
B, 0.4	86.8	84.0	67.9	47.5	19.4
A, 0.8	94.9	92.3	88.5	69.3	42.2
B, 0.8	69.1	59.5	47.6	33.9	16.5

In the proposed method, normalised distance between two images is computed for *Region GFD* and *Contour GFD* using the *norm1* method described in Section 7.3. The total distance between two images is computed as the by MPEG-7. 60 features (reflecting 5 radial frequencies and 12 angular frequencies) each are selected for the Contour GFD and the Region GFD. Queries are performed using GFD; these are represented by *'traditional'* within the legends. Another set of queries are performed using the proposed method; these are represented by '*composite'* within the legends. Average recall-precision plots for queries in Sets A1, A2, A3 and A4 are shown in Figure 4. Precision along the vertical axis is plotted for recall along the horizontal axis.

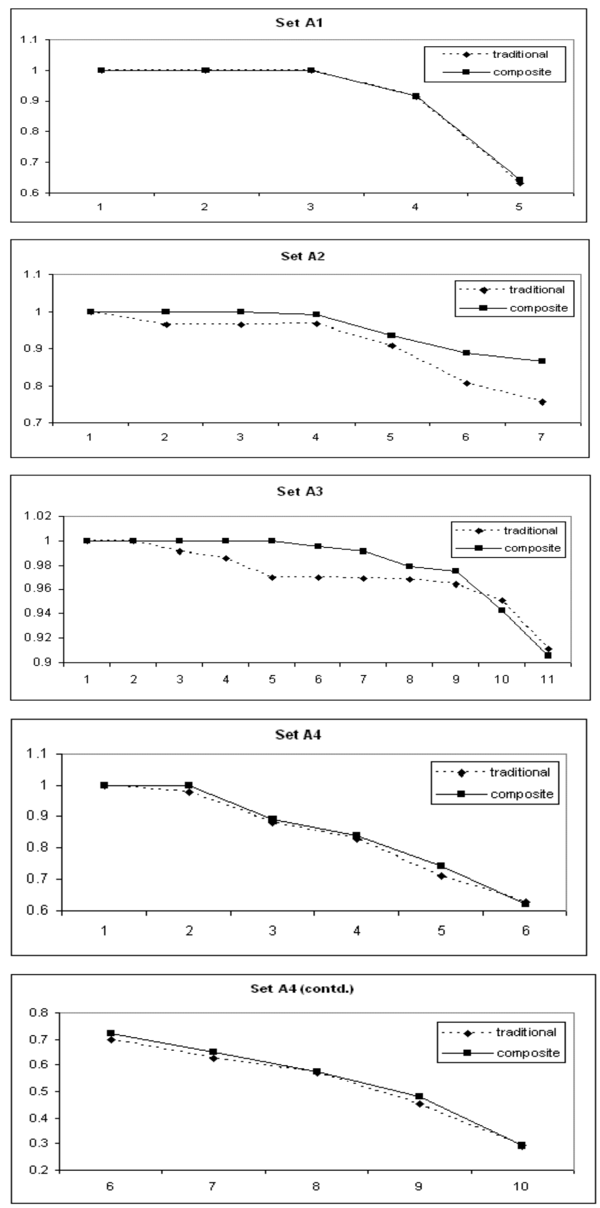


Figure 4. Recall-Precision Plots for Queries in Set A1, A2, A3, A4

From the results, we observe that there is no significant difference in performance for Set A1. The proposed method shows significant improvement for Sets A2 and A3. The proposed method shows a small improvement for Set A4. In Figure 4, the weight assigned to *Contour GFD* is 0.1. Experimental results for weights of 0.4 and 0.8 are shown in Figure 5.

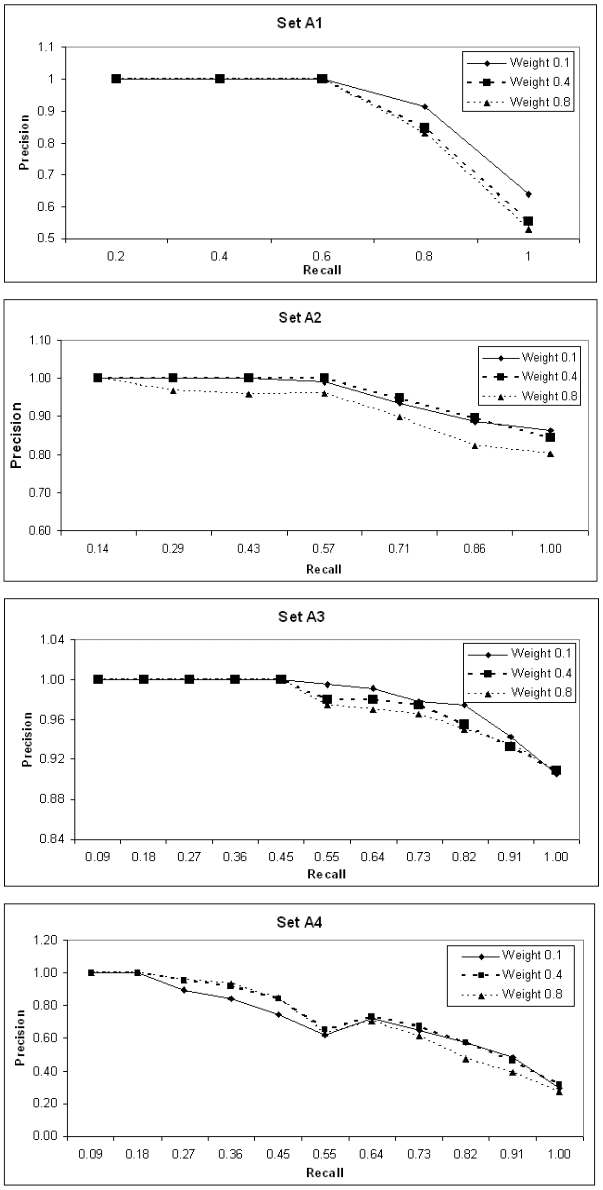


Figure 5. Recall-Precision for varying weights of Contour GFD

From Figure 5, it is observed that different datasets respond differently to the change in the weight assigned to *Contour GFD*. As expected, the effectiveness drops when the weight of *Contour GFD* is increased beyond a threshold. The optimal value of the weight assigned to *Contour GFD* needs to determined empirically and it will depend on the dataset.

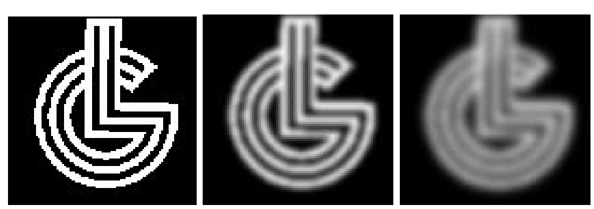
5. Discussion

There are two factors which contribute to the relative improvement of the proposed method when compared with traditional GFD. First, additional information is captured by connectivity; descriptors which encode connectivity are able to discriminate better between shapes [4].

Improvement of the proposed method is also attributed to the use of spectral analysis. Spectral analysis of images has been widely used for image retrieval. There are two advantages of spectral features. First, they are robust compared with spatial features. Second, spectral features are inherently multi-resolutional and this property can be leveraged to determine the degree of detail encoded during indexing.

Computational expense of the proposed method also needs to be addressed. The proposed method requires substantial processing compared with other techniques for 2d image retrieval. Processing overheads of the proposed method include decomposition of images by connectivity. Computation of connectivity has complexity O(n) where n is the number of foreground pixels in the image. The intense processing requirements may be prohibitive for some applications. However, for applications where accuracy of retrieval is important, the improvement in effectiveness may outweigh the processing complexity.

We note that the dataset does not contain fine contours. In Figure 2, we see that the pixel density is high for connectivity=0 and connectivity=8. Datasets where intermediate values of connectivity are high will warrant an investigation to obtain GFD for the intermediate values of connectivity; composite descriptor based on GFD for intermediate connectivities can be constructed. However, it is likely that these descriptors are highly sensitive to noise.



(a) (b) (c)

Figure 6. Gaussian Smoothing Image #308

The proposed method was tested for sensitivity of the contour. In order to alleviate any possible effect of noise, image smoothing was performed at the preprocessing stage. Image smoothing was performed by convolving the original image with a Gaussian filter [7]. The result of image smoothing on Figure 6(a) (Image #308) is shown in Figure 6(b) and Figure 6(c).

Contour extraction was done for the smoothed image; the contours thus generated are shown in Figure 7. With the exception of Set A1, there was no noticeable difference in performance when contour smoothing was used. The robustness of the method is attributed to the spectral nature of the descriptors. Ignoring high resolution features in the descriptors makes the method robust. Set A1 experienced a drop in performance when image smoothing was used.

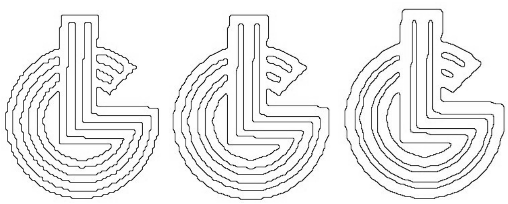


Figure 7. Contour Extraction after Smoothing

In order to explain the drop in performance for Set A1, it should be noted that Set A1 consists of scaled transformations of images. Consider scaling down the images shown in Figure 6; when contour extraction is done for the scaled down images, we obtain the results shown in Figure 8.

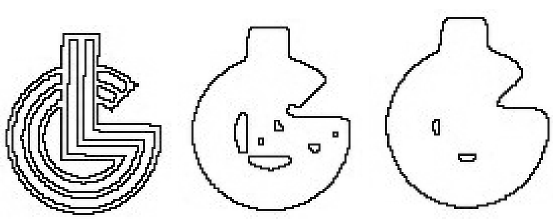


Figure 8. Contour Extraction after Smoothing the Scaled-down Images

From Figure 8, we observe that there is loss of information in the contour as the scaled down image is smoothed; degraded performance for scaled down images (Set A1) is attributed to this information loss.

6. CONCLUSION

In this paper, we have proposed a new index which complements GFD by capturing contour information. Contour information is captured based on connectivity. Experiments have been performed on the MPEG-7 Still Images Content Set. Experimental results prove that the proposed method is promising. The proposed method may be modified to incorporate other shape retrieval techniques such as geometric moments, however, this will need further investigation.

6. References

[1] D. Zhang, *Image Retrieval Based on Shape*, PhD Thesis, Monash University, Australia, 2002.

[2] D. Zhang and G. Lu, *Generic Fourier Descriptors for Shape-based Image Retrieval*, IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, August 26-29, 2002.

[3] G. Pass and R. Zabih, *Histogram refinement for content-based image retrieval*, IEEE Workshop on Applications of Computer Vision, pp. 96-102, December 1996.

[4] A. Sajjanhar, G. Lu and D. Zhang, *Discriminating Shape Descriptors Based on Connectivity*, IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, 2004.

[5] A. Sajjanhar, G. Lu and D. Zhang, *Spherical Harmonics Descriptors for 2D Image Retrieval*, IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 2005.

[6] A. Celentano and S. Sabbadin, *Multiple Feature Indexing in Image Retrieval Systems*, First European Workshop on Content-Based Multimedia Indexing, Toulouse, France, October 25-27, 1999.

[7] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[8] A. K. Jain; A. Vailaya, *Shape-based retrieval: a case study with trademark image databases*, Pattern Recognition, Volume 31, Number 9, September 1998, pp. 1369-1390

[9] C. Faloutsos, R. Barber, M. Flickner, J. Hafner,W. Niblack, D. Petkovic, and W. Equitz, *Efficient and effective querying by image content*, Journal of Intelligient Information Systems, 3(3/4):231-262, July 1994.

[10] J. Ashley, R. Barber, M. Flickner, J. Hafner, D. Lee,W. Niblack, and D. Petkovic, *Automatic and semi-automatic methods for image annotation and retrieval in QBIC*, Proceedings of Storage and Retrieval for Image and Video Databases III, Volume 2420, pages 24-35, February 1995.

[11] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, *The QBIC project: querying images by content using color, texture and shape,* Storage and Retrieval for Image and Video Databases, Volume 1908, pages 173-187, 1993.