Shape Retrieval with Flat Contour Segments

Dalong Li[1], Steven Simske
Intelligent Enterprise Technologies Laboratory
HP Laboratories Palo Alto
HPL-2002-250
September 9th , 2002*

image database, image retrieval, run length coding, Freeman coding, down-sampling

Content-based image retrieval (CBIR) is an important issue in the computer vision community. Both visual and textual content descriptions are employed when the user formulates queries. Shape feature is an important visual feature, as it corresponds to the region of interest in images. For retrieval, shape comparisons must be compact and accurate, and must be invariant to several geometric transformations such as translation, rotation and scaling, even if the particular representation may be rotated. In this paper, we propose a shape comparison technique based on the flat segments of the contour. The segmentation utilizes the Freeman coding technique and run length coding. The lengths of the flat segments make up a length vector, which are used to compare the similarity of the shapes. Experimental results from the test on the standard SQUID database are reported.

* Internal Accession Date Only
Approved for External Publication
[1] Georgia Institute of Technology

Shape retrieval with flat contour segments

Dalong Li+, Steven Simske
Intelligent Enterprise Technologies Lab
HP Laboratories Palo Alto
(+Georgia Institute of Technology / HP Labs)
July 30th, 2002

ABSTRACT

Content-based image retrieval (CBIR) is an important issue in the computer vision community. Both visual and textual descriptions are employed when the user formulates his queries. Shape feature is one of the most important visual features. The shape feature is essential as it corresponds to the region of interest in images. Consequently, the shape representation is fundamental. The shape comparisons must be compact and accurate, and must own properties of invariance to several geometric transformations such as translation, rotation and scaling, though the representation itself may be variant to rotation.

In this paper, we propose a shape comparison technique based on the flat segments of the contour. The segmentation utilizes the Freeman coding and run length coding. The lengths of the flat segments make up a length vector, which are used to compare the similarity of the shapes. Experimental results from the test on the standard SQUID are reported.

Keywords: Image Database, retrieval, run length coding, Freeman Code, down-sampling.

1 Introduction

One of the issues raised by multimedia data (such as images, video, audio, graphics, text) integration in databases is the efficient retrieval of images [3]. Several prototypes [4-7] and commercial systems [1-2] have been implemented in order to address this problem. The retrieval process is based on the content, and more particularly on the visual object features.

This process requires principally two modules: the extraction and the query. The first module extracts visual features from data by using analysis techniques for each media. Then, each extracted feature is stored in the database. For example, the color feature may be represented by a color histogram [8], statistical moments [9], etc. In the second module, users formulate their query from features previously extracted. The retrieval process computes the similarity between source and target features, and sorts the most similar objects according to their similarity value. The system must be flexible since images may belong to different domains.

For images retrieval, low-level visual features are color, texture, and shape. Among these features, shape is the most important because it represents significant regions or relevant objects in images. Extensive work has been done in shape retrieval. They include tree-pruning [14], the generalized Hough transformation [15], or pose clustering [16], geometric hashing [17], the alignment method [18], deformable template [19], relaxation labeling [20], Fourier descriptors [21], wavelet transformation [22], curvature scale space [23], neural networks [24], dynamic programming [25] and shape context [26].

In the paper, we focus on shape representation and comparisons. In section 2, we review the freeman coding and run length coding which are used in our method. Our algorithm and experiments are introduced in section 3 and section 4 respectively, followed by a conclusion.

2 Shape representation with Freeman coding

In a content-based image retrieval system, the shape matching process efficiency is essential. Consequently, a compact and reliable shape representation and a well-suited similarity distance are necessary. An interesting shape description should be invariant to translation, rotation, scaling and starting point transformations [10-11].

In general, shape representations are classified into two categories: boundary-based and region-based. The first one describes the considered region by using its external characteristics (i.e. the pixels along the object boundary) while the second one represents the considered region by using its internal characteristics (i.e. the pixels contained in the region). Several shape description approaches have been developed in the two categories. For example, area, compactness, bounding box for the region-based category, and perimeter, curvature for the boundary-based category may be cited [10-11]. Other methods of shape description are the Fourier theory-based method and the Moment theory-based method. As far as the former approach is concerned, the Discrete Fourier Transform (DFT) or the Fourier Series (FS) are generally used to describe the shape feature from its boundary. They give a sequence of complex coefficients called Fourier Descriptors [12-13]. These coefficients represent the shape of an object in the frequency domain where the lower frequencies symbolize its general contour, and where the higher frequencies represent the details of its contour. Only a few coefficients are required to describe even quite complex shapes. A modified Fourier descriptor method [13] is proposed in order to take into account the discretization noise. The moment theory-based method uses region-based moments to characterize the contour of an object. A set of 7 moments was identified by Hu, and is invariant to geometric changes. These moments are called invariant moments [10-11].

A simple method to represent a contour is Freeman code (chain code), a coding method of closed shape by approximation of the continuous contour with a sequence of numbers, each number corresponding to a segment direction. Freeman's code is usually employed in a 4-neighborhood (where 4 possible directions may be used) or in an 8-connectivity (that gives an 8-directional chain code). This representation is compact and is invariant to the geometric transformation translation. However, it depends on rotation and scaling transformations. An illustration of Freeman code is given in Figure 1.

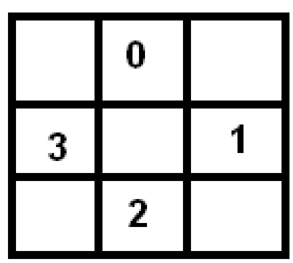
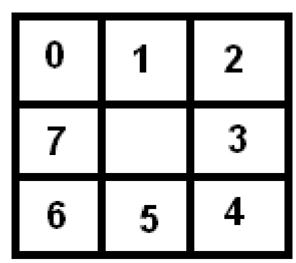
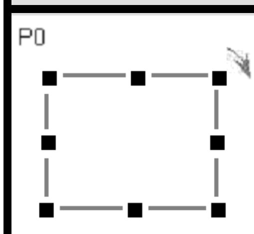
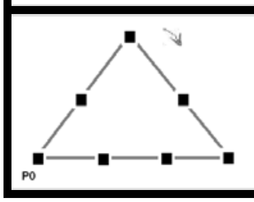
The considered object	4-neighborhood / 4-connectivity	8-neighborhood / 8-connectivity
		
	11223300	33557711
	00113333	22447777

Figure 1. Illustration of Freeman Coding

It is worth mentioning that Freeman coding is applied to a simply connected closed planar curve. Basically there are three parameters when a Freeman is created. They are start points (P_0), the step (L) between two consecutive sample points along the contour, and the tracking direction (T_d) on which the curve is traced. The tracking direction is either clockwise or anticlockwise.

Let us use $\{P_0, L, T_d\}$ in denoting the parameters of the Freeman coding. Now let us analysis how each parameter influences the coding result. Choosing a different starting point usually leads to a different (shifted) coding result. Rotation has same effect in coding as choosing a different start point.

L is the step of two consecutive sample points along the contour. In our work, we let L = 1 to have an accurate representation of the shapes.

The impact from tracking direction to coding is straightforward, {0,1,2,3,4,5,6,7} is substituted by {4,5,6,7,0,1,2,3} accordingly. Flipping can result in the same effect as choosing a different tracking direction.

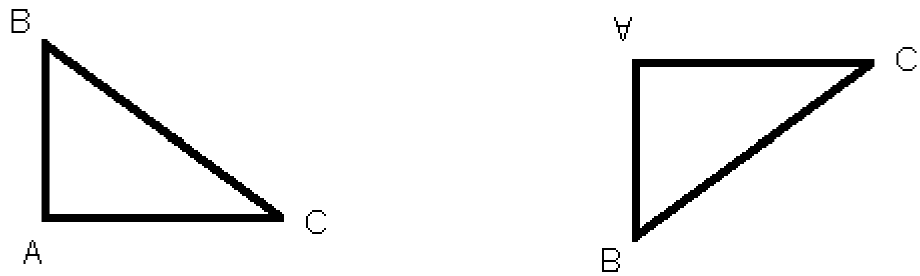


Figure 2 Flip reverses the tracking directions

What remain unchanged are the angles no matter how you geometrically transform a shape. The number in the Freeman code represents the direction. Thus the difference in the Freeman code reflects the turning angles. They are invariant to the transformations. As we know the angle is periodic modulo 360 degree. For example, -180 degree represents the same angle as –180+360(180). When an 8-connectivity is used, 360 degree is divided evenly into eight parts, each part is 45 degrees. To reflect the periodicity of the angle in the code, it is necessary to compute the complement of the angles (reflected by direction difference) in the following way:

$$f(x) = x, x>=0; \qquad \text{Eq 1}$$

$$f(x) = 8+x, x<0; \qquad \text{Eq 2}$$

Run length coding is a simple yet useful technique in lossless coding. Run Length Coding (RLC) is a conceptually simple form of compression. RLC consists of the process of searching for repeated runs of a single symbol in an input stream, and replacing them by a single instance of the symbol and a run count.

3 Shape comparisons with flat contour segments

Freeman coding itself is not robust to rotation or scaling. It can not be directly used in shape comparison. However based on freeman coding, we can create something which is robust for comparisons: the length of the flat segments of the shape contours. The flat segment is defined as " *a part of the contour which travels in a unchanged direction for at least a certain length*". In Freeman code, we use number (0-8 for 8-connectivity) in denoting the directions. Thus in the flat segments, the number (direction) is unchanged, the difference between the consecutive number must be zero. In our method, we use run length coding to detect the flat segment and compute the length of the flat segment. To get the flat segments from a contour, the first operation we apply to freeman code is a differentiation. Then the complement is computed according to Eq 1-2. Run-length coding is performed in order to extract the continuous flat parts from the contour.

With run length code and thresholding, we can get the lengths of the flat segments of the contour, which are used for shape comparisons. From the flat segment lengths, we can construct a vector. The length of the vectors may not be equal. To compare vectors with different lengths, we use down sample to make them have the same length. To compare the down-sampled vectors, circular shifts are performed. Finally the minimal standard deviation of the ratio of the shifted vectors is used for the measurement of the shapes. In general, the steps are following:

1. Track the contour and get the Freeman Code (FC)
2. Differentiate FC and obtain the difference (DFC)
3. Compute the complement of DFC (CDFC)
4. Extract the flat segment of the contour via run length coding on CDFC
5. Obtain the flat segment length vector (FSLV)
6. Down-sample the FSLV
7. Circular shift the FSLV before comparison
8. Use the minimal standard deviations of the ratios of the vectors as the measurement of the similarity

Given two vectors: $v_1 = \{a_1, a_2, a_3, a_4, a_5\}$, $v_2 = \{b_1, b_2, b_3, b_4, b_5\}$, the ratio of the two vectors $\{v_1, v_2\}$ is $\{a_1/b_1, a_2/b_2, a_3/b_3, a_4/b_4,\}$. This explains the robustness to scaling. When a shape is scaled, the elements of the obtained segment length vectors are proportionally increased or decreased. Thus the standard deviation of the ratios shall be zero. Circular shift is designed to deal with rotation.

To compare two signals with different lengths, down-sampling can be used. If the two signals are similar to each other, then the down sampled signals are still similar to each other. Down sampling provides a chance to change the length of the signals. Suppose the lengths of the original signals are L_1, L_2. What shall be the L_3, the down-sampled signal? Usually it is determined by experiment. The criteria is: L_3 shall be able to distinguish similar signals and non-similar signals. L_3 is signal depended. That is to say, for another database, it may be different from this database. There is no uniform value that is proper for any database.

4 Experiments

The parameters we need to specify for the retrieval system are minimal length ML, which is used to threshold the run length code to extract the flat segments on the contour; T, the threshold, which distinguish the shapes by the minimal standard deviation of the length ratio and SN, the Sample number used in down-sampling.

The parameters are obtained from the training set, which is made of the pairs of shapes. Samples of the training set are shown in Figure 3. We already know whether those pairs of shapes are similar. Let ML = 5. We tried several SN. The results are shown below. In all the figures, red (*) stands for the minimal standard deviation of the length ratios of the similar shapes while the blue (#) denotes that of the dissimilar shapes. From the figures, N = 10 and t = 0.0 are selected for the retrieval system.

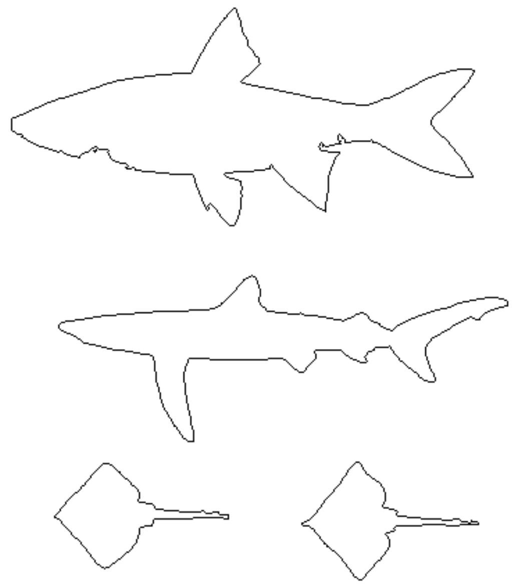


Figure 3 The samples of the training set

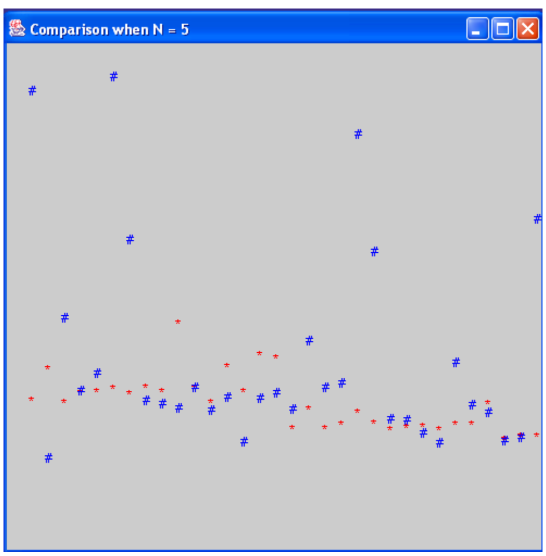


Figure 4 – 1 The comparisons when N = 5

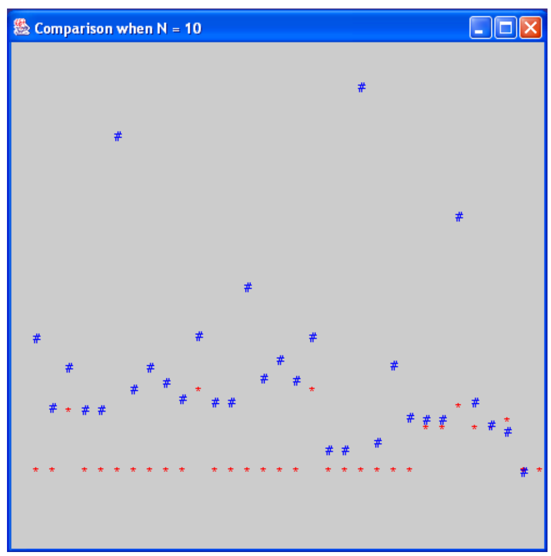


Figure 4 – 2 The comparisons when N = 10

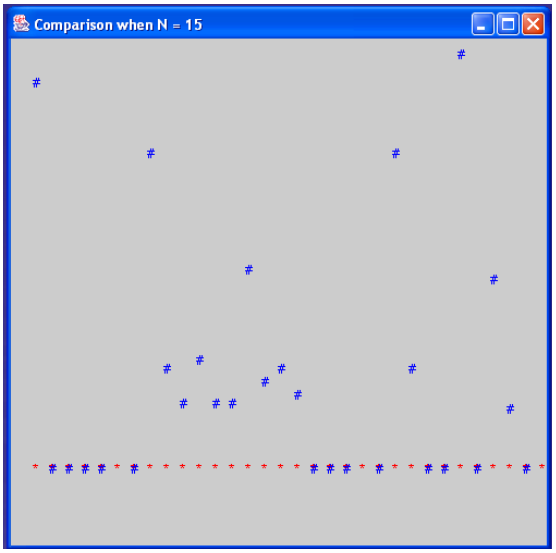


Figure 4 – 1 The comparisons when N = 15

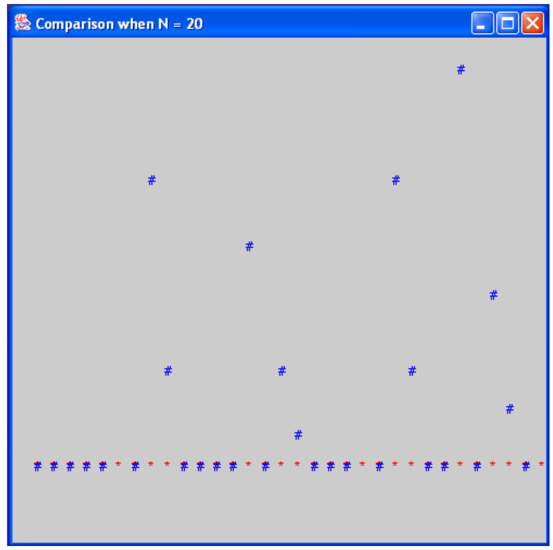


Figure 4 – 4 The comparisons when N = 20

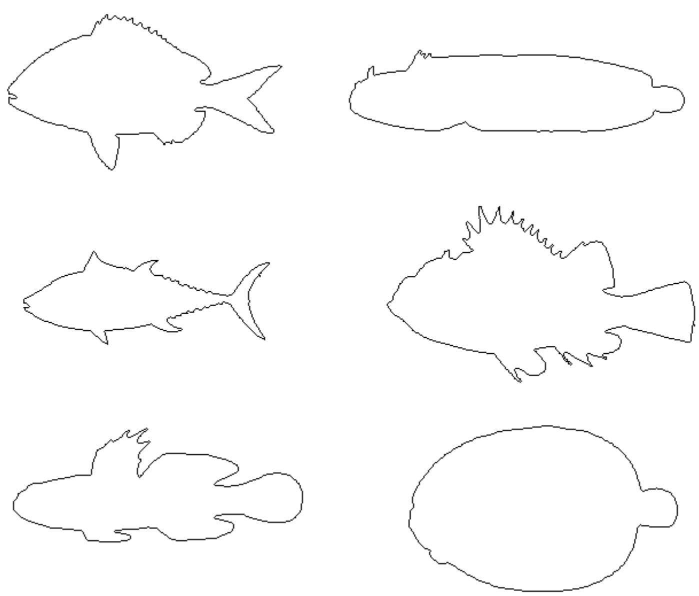


Figure 5 The index fish

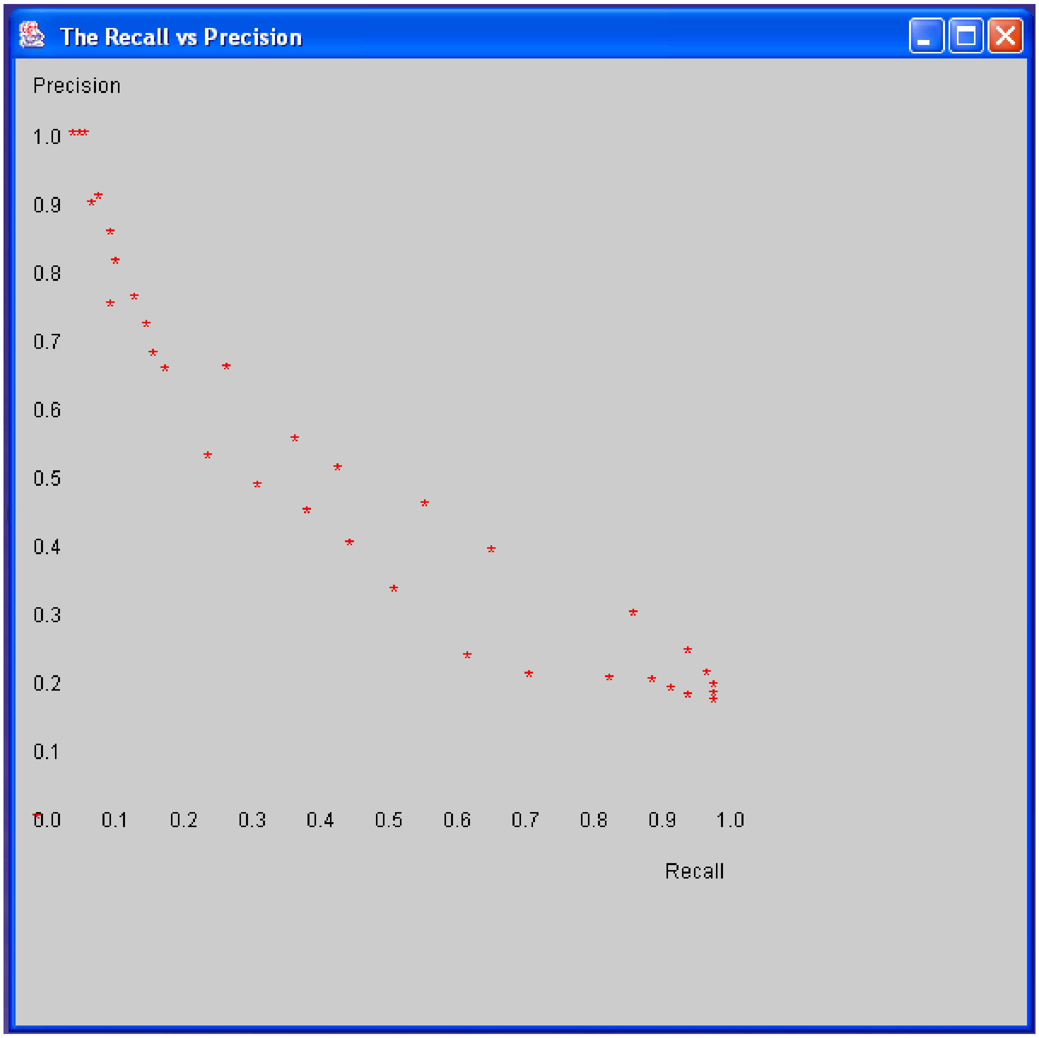


Figure 6 The recall precision plot

We performed the retrieval test on a set made of 110 fish from the SQUID. The index fish are shown in figure 5. Like for all retrieval systems, we compute the recall-precision. Given a query Q: R is the answer set the system has retrieved; C is the set of correct answers in R; A is the set of the right answers for the query. Precision and recall measures can be defined as follows:

$$\text{Precision} = C/R \qquad \text{Eq 3}$$

$$\text{Recall} = C/A \qquad \text{Eq 4}$$

To control the recall rate, the thresholds are modified each time. Generally the higher the threshold, the higher the recall rate. We compute the average recall and precision of the 6 retrievals each time with a different T. T is evenly increasing so that the recall ratio is changed in the experiment. The precision and recall plot is shown in figure 7. SQUID is a challenging database for retrieval since many fish are similar. As a result, the precision recall plot usually resembles. Different recall-precision plots are obtained when tested on different databases.

5 Conclusion

The shape feature is essential in the content-based image retrieval. It must be accurate and compact, and it should be invariant to certain geometric transformations. In our proposal, we use freeman code and run length code in segmenting the contours. From the length of the extracted flat segments, we get the measurements of the content of the shape. Robustness to translation is obvious since the length of the segments has nothing to do with the position of the shape. Circular shifting makes it robust to rotation, and the length ratio is designed to handle scaling.

Acknowledgement

The authors would like to thank Prof. Mokhtarian, Centre for Vision, Speech, and Signal Processing Laboratory, University of Surrey, U.K., for providing the SQUID database for our experiments.

References

[1]Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., and Shu, C.F. "The Virage Image Search Engine: An Open Framework for Image Management", In *Proc. Storage and Retrieval for Still Image and Video Databases* IV, SPIE, San Diego, CA, 1996, Vol. 2670, 76-87.

[2]Flickner, M., et al. "Query by Image and Video Content: The QBIC System." *Computer*, September 1995.

[3]Khoshafian, S., and Baker, A.B. "Multimedia and imaging databases", *Morgan Kaufmann*, San Francisco, California, 1996.

[4] Smith, J.R., and Chang, S.F. "VisualSEEK: a fully automated content-based image query system", *ACM Multimedia'96*, November 1996.

[5] Pentland, A., Picard, R.W., and Sclaroff, S. "Photobook: Tools for Content-Based Manipulation of Image Databases", *Proc. Storage and Retrieval for Image and Video Databases, SPIE*, Bellingham, Washington, 1994, Vol. 2, 34-47.

[6]Rui, Y., Huang, T.S., Mehrotra, S., and Ortega, M. "A Relevance Feedback Architecture in Content-Based Multimedia Information Retrieval Systems", *Proc. IEEE Workshop Content-Based Access of Image and Video Libraries, IEEE,* 1997.

[7]Nastar, C., Mitschke, M., Meilhac, C., and Boujemaa, N. "Surfimage: a Flexible Content-Based Image Retrieval System". *The 6th ACM Int'l Multimedia Conf. (MM'98)*, Bristol, England, September 1998, 339-344.

[8]Swain, M.J., and Ballard, D.H. "Color indexing". *Int'l Journal of Computer Vision*, 1991, Vol.7(1), 11-32

[9]Stricker, M., and Orengo, M. "Similarity of Color Images". *Proc. Storage and Retrieval for Still Image and Video Databases III*, SPIE, San Diego, CA, 1995, 381-392.

[10]Jähne, B. "Digital Image Processing - Concepts, Algorithms, and Scientific Applications". 4th Edition, *Springer*, 1997.

[11] Khoshafian, S., and Baker, A.B. "Multimedia and imaging databases", *Morgan Kaufmann*, San Francisco, California, 1996

[12]Zahn C.T., and Roskies, R.Z. "Fourier Descriptors for Plane Closed Curves". *IEEE Transactions on Computers*, March 1972, Vol. C-21, N°. 3, 269-281.

[13]Rui, Y., She, A.C., and Huang, T.S. "Modified Fourier Descriptors for Shape Representation - A Practical Approach". *Proc. of First Int'l Workshop on Image Databases and Multi Media Search*, Amsterdam, The Netherlands, 1996.

[14] S. Umeyama, "Parameterized point pattern matching and its application to recognition of object families," *IEEE Transaction on pattern Analysis and Machine Intelligence*, 15(1):136-144,1993.

[15] D.H.Ballard, "Generalized Hough transform to detect arbitrary patterns", *IEEE Transaction on pattern Analysis and Machine Intelligence*, 13(2): 111-122,1981.

[16] G.Stockman, "Object recognition and localization via pose clustering", *Computer vision, Graphics, and Image Processing*, 40(3):361-387,1987.

[17] H.Wolfson and I. Rigoutsos, "Geometric hashing; an overview", *IEEE Computational Sciences & Engineering* , pages 10-21, October-December 1997.

[18] D. Huttenlocher and S Ullman, "Object recognition using alignment", *Proceeding of international conference on computer vision*, London, pages 102-111,1987

[19] S. Sclaroff and A. P. Pentland, "Modal matching for correspondence and recognition", *IEEE Transaction on pattern Analysis and Machine Intelligence*, 17(6) June 1995

[20] S. Ranade and A Rosenfeld, “Point pattern matching by relaxation”, *Pattern recognition*, 12:269-275,1980.
[21] Persoon, E., and Fu, K.S. “Shape Discrimination Using Fourier Descriptors”. *IEEE Transactions on Systems, Man, and Cybernetics*, March 1977, Vol. SMC-21, N° 3, 170-179.
[22] C. Jacobs, A. Finkelstein, and D. Salesin, “Fast multi-resolution image querying”, *Computer Graphics Proceeding SIGGRAPH*, page 277-286,1995.
[23] F. Mokhtarian, S. Abbasi, and J. Kittler, “Efficient and robust retrieval by shape content through curvature scale space”, *In image databases and multi media search, proceeding of the first international workshop IDB-MMS’96*, Amsterdam, the Netherlands, pages 35-42,1996.
[24] S. Gold, “Matching and learning structural and spatial representation with neural networks”, *PHD thesis*, Yale University, 1995
[25] Evangelos Milios and Euripides G. M. Petrakis “Shape Retrieval Based on Dynamic Programming”, *IEEE Transaction on image processing*, Vol 9, No 1, pp 141-147, Jan 2001.
[26] S. Belongie, J. Malik, J. Puzicha, “Shape matching and object recognition using shape context”, *IEEE Trans. PAMI* April 2002, 24(4), pp. 509-522.