

IMAGE RETRIEVAL SYSTEM USING HYBRID FEATURE EXTRACTION TECHNIQUE

Vadhri Suryanarayana¹, Dr. M.V.L.N. Raja Rao², Dr. P. Bhaskara Reddy³ and
Dr. G. Ravindra Babu⁴

¹Dept. of CSE & MCA, NRI Institute of Technology, Agiripalli(V),
Vijayawada, A.P, India.

s_vadhri@yahoo.co.in

²Dept. of IT, Gudlavalleru Engineering College, Gudlavalleru, A.P, India

rajarao.mamidanna@gmail.com

³Director, MLR Institutions, Hyderabad, A.P, India

pbhaskarareddy@rediffmail.com

⁴Dept. of CS&E, S.C.C.E.T, Hyderabad, A.P, India.

grbabu22@gmail.com

ABSTRACT

Image Retrieval system is an effective and efficient tool for managing large image databases. A content based image retrieval system allows the user to present a query image in order to retrieve images stored in the database according to their similarity to the query image. In this paper content based image retrieval method is used on digital image data set. The main objective of this paper is to evaluate the retrieval system based on Hybrid features. The texture features are extracted by using pyramidal wavelet transform and the shape features are extracted by using Fourier descriptor. And the hybrid technique is the combination of both texture and shape. The major advantage of such an approach is that little human intervention is required. It is ascertained that the performance is superior when the image retrieval based on the Hybrid features, and better results than primitive set.

KEYWORD

Feature Extraction, Hybrid features, Image Retrieval

1. INTRODUCTION

Image Retrieval System provides an effective and sophisticated tool for maintaining and managing image databases. There is a significant amount of increase in the use of medical images in clinical medicine and disease research. Image retrieval (IR) is one of the most exciting and fastest growing research area in the field of medical imaging [2]. The goal of CBIR is to retrieve images from a database that are similar to an image placed as a query. In CBIR, for each image in the database, features are extracted and compared to the features of the query image. A CBIR method typically converts an image into a feature vector representation and matches with the images in the database to find out the most similar images. A comparative study has given on multiple databases [3,6,10] it is concluded that performance of DRD image is less compared to other database images.

The goal of CBIR is to retrieve images from a database that are similar to an image placed as a query. In CBIR, for each image in the database, features are extracted and compared to the features of the query image. A CBIR method typically converts an image into a feature vector representation and matches with the images in the database to find out the most similar images. In various studies different databases have been used to compare the study. The similarity

between features are to be calculated by using algorithms i.e., used by well known CBIR systems such as IBM's QBIC[28]. For each specific feature there is a specific algorithm for extraction and another for matching.

In this paper evaluation of retrieval system based on hybrid features is carried out. The major advantage of this approach is that little human intervention is required. The databases used here is digital data of 200 images with multiple contexts.

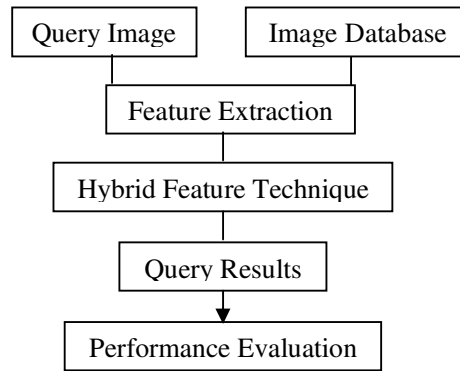


Figure 1. Block diagram of Image Retrieval System

Figure 1 show the basic block diagram used in this work. Shape and texture features are extracted for both query image and images in the database. The distance (ie. similarities) between the features vectors of the query image and database are then computed and ranked. The database images that have highest similarity to the query image are retrieved. Then the performance analysis is carried out using precision and recall.

2. FEATURE EXTRACTION

Feature Extraction is the process of creating a representation, or a transformation from the original data. The images have the primitive features like color, texture, shape, edge, shadows, temporal details etc. The features that were most promising were color, texture and shape/edge. The reasons are color can occur in limited range of set. Hence the picture elements can be compared to these spectra. Texture is defined as a neighborhood feature as a region or a block. The variation of each pixel with respect to its neighboring pixels defines texture. Hence the textural details of similar regions can be compared with a texture template. shape/edge is simply a large change in frequency. The three feature descriptors mainly used most frequently during feature extraction are color, texture and shape.

2.1 Texture Feature Extraction

The feature of Texture is describe the visual patterns with property of homogeneity. It contains important information about the structural arrangements of the surface like bricks, etc., It is also possible to depicts the relationship of the surface to the in and around environment. In short, it is a feature that describes the distinctive physical composition of a surface. The most popular statistical representations of texture are: Co-occurrence Matrix, Tamura Texture and Wavelet Transform.

2.2 Feature Extraction Methods for Texture Analysis

In Texture Analysis the main feature extraction methods are spectral approach, structural (or syntactic) approach and statistical approach. The spectral approach to texture analysis deals with images in the frequency domain. Therefore, this approach requires Fourier transform to be

carried out on the original images to acquire their corresponding representations in the frequency space. Fourier transform based methods usually perform well on textures showing strong periodicity, however their performance deteriorates as the periodicity of textures weakens. The classification of all methods among the two other approaches: structural and statistical.

The statistical point of view, an image is a complicated pattern on which statistics can be obtained to characterize these patterns. The techniques used within the family of statistical approaches make use of the intensive values of each pixel in an image and apply various statistical formulae to the pixels in order to calculate feature descriptors. Texture feature descriptors, extracted through the use of statistical methods, can be classified into two categories according to the order of the statistical function that is utilized: First-Order Texture Features and Second Order Texture Features. First Order Texture Features are extracted exclusively from the information provided by the intensity histograms, thus yield no information about the locations of the pixels. Another term used for First-Order Texture Features is Grey Level Distribution Moments.

In contrast, Second-Order Texture Features take the specific position of a pixel relative to another into account. The most popularly used of second-order methods is the Spatial Grey Level Dependency Matrix (SGLDM) method. It consists of constructing number of matrices by counting the occurrences of pixel pairs of given intensities at a given displacement.

2.3 Shape Feature Extraction

The Feature of Shape may be defined as the characteristic surface configuration of an object. It permits an object to be distinguished from its surroundings by its outline. Shape representations are two categories. They are 1. boundary-based and 2. region-based. Boundary-based mathematical representations are Polygonal Models, boundary partitioning, Fourier Descriptors, Splines, higher order constructs, Curvature Models. Region-based mathematical representations are Super Quadrics, Fourier Descriptors, Implicit Polynomials, Blum's Skeletons. The successful representations are Fourier Descriptor and Moment Invariants. The Fourier Descriptor is used to boundary as the shape feature and the Moment invariants is used to region-based shape feature.

Edge detection is used to detect sharp changes in an image brightness is to capture important events and changes in properties of the world. It can be exhibited that the general assumptions for any image formation model, discontinuities in image brightness are likely to correspond to discontinuities in depth. Canny considered the mathematical problem of deriving an optimal smoothing filter given the criteria of detection, localization and minimizing multiple responses to a single edge. This method showed that the optimal filter given these assumptions is a sum of four exponential terms. It also showed that this filter can be well approximated by first-order derivatives of Gaussians. This paper emphasizes on the image retrieval based on the extraction of features of texture and shape of a specific image database.

2.4 Texture Feature Extraction

The feature of Texture is describe the visual patterns with property of homogeneity. It contains important information about the structural arrangements of the surface like bricks, etc., It is also possible to depicts the relationship of the surface to the in and around environment. In short, it is a feature that describes the distinctive physical composition of a surface. The most popular statistical representations of texture are: Co-occurrence Matrix, Tamura Texture and Wavelet Transform.

2.5 Feature Extraction Methods for Texture Analysis

In Texture Analysis the main feature extraction methods are spectral approach, structural (or syntactic) approach and statistical approach. The spectral approach to texture analysis deals with images in the frequency domain. Therefore, this approach requires Fourier transform to be carried out on the original images to acquire their corresponding representations in the frequency space. Fourier transform based methods usually perform well on textures showing strong periodicity, however their performance deteriorates as the periodicity of textures weakens. The classification of all methods among the two other approaches: structural and statistical.

The statistical point of view, an image is a complicated pattern on which statistics can be obtained to characterize these patterns. The techniques used within the family of statistical approaches make use of the intensive values of each pixel in an image and apply various statistical formulae to the pixels in order to calculate feature descriptors. Texture feature descriptors, extracted through the use of statistical methods, can be classified into two categories according to the order of the statistical function that is utilized: First-Order Texture Features and Second Order Texture Features. First Order Texture Features are extracted exclusively from the information provided by the intensity histograms, thus yield no information about the locations of the pixels. Another term used for First-Order Texture Features is Grey Level Distribution Moments.

In contrast, Second-Order Texture Features take the specific position of a pixel relative to another into account. The most popularly used of second-order methods is the Spatial Grey Level Dependency Matrix (SGLDM) method. It consists of constructing number of matrices by counting the occurrences of pixel pairs of given intensities at a given displacement.

The Feature of Shape may be defined as the characteristic surface configuration of an object. It permits an object to be distinguished from its surroundings by its outline. Shape representations are two categories. They are 1. Boundary-based and 2. Region-based.

The Boundary-based shape representation uses the outer boundary of the shape. This can be done by describing the concerned region using its external characteristics i.e., the pixels along the object boundary. Region-based shape representation uses the entire shape region by describing the concerned region using its internal characteristics i.e., the pixels contained in that region. The Boundary-based mathematical representations are Polygonal Models, Fourier Descriptors, etc., Region-based mathematical representations are Super Quadrics, Fourier Descriptors, Implicit Polynomials and Blum's Skeletons. The major important representations for shape categories are Fourier Descriptor and Moment Invariants. The Fourier Descriptor is used to boundary as the shape feature and the Moment invariants is used to region-based shape feature.

Edge detection is a terminology in image processing and computer vision, particularly in the areas of feature detection and feature extraction. The purpose of detecting sharp changes in an image brightness is to capture important events and changes in properties of the world. It can be exhibited that the general assumptions for any image formation model, discontinuities in image brightness are likely to correspond to discontinuities in depth. The major important representations for shape categories are Fourier Descriptor and Moment Invariants. The main idea of Fourier Descriptor is to use the Fourier transformed boundary as the shape feature. The important idea of Moment invariants is to use region-based moments.

Canny Edge detection is an optimal smoothing filter given the criteria of detection, localization and minimizing multiple responses to a single edge. This method showed that the optimal filter given these assumptions is a sum of four exponential terms. It also showed that this filter can be well approximated by first-order derivatives of Gaussians. The Canny Edge detection also

introduced the notion of non-maximum suppression, which means that given the pre-smoothing filters are defined as points.

The Sobel Edge detection operations are required to perform on the data. The processed data is sent back to the computer. All the digital logics implemented and verified on the field programmable gate array kit was described using the Verilog® Hardware Description Language.

To estimate image gradients from the available input image, different gradient operators can be applied. The simplest approach is to use central differences:

$$L_x(x, y) = -1/2 \cdot L(x - 1, y) + 0 \cdot L(x, y) + 1/2 \cdot L(x + 1, y).$$

$$L_y(x, y) = -1/2 \cdot L(x, y - 1) + 0 \cdot L(x, y) + 1/2 \cdot L(x, y + 1).$$

corresponding to the application of the following filter masks to the image data:

$$L_x = [-1/2 \quad 0 \quad 1/2] * L \quad \text{and} \quad L_y = \begin{bmatrix} +1/2 \\ 0 \\ -1/2 \end{bmatrix} * L$$

The well-known and earlier Sobel operator is based on the following filters:

$$L_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * L \quad \text{and} \quad L_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * L$$

Given such estimates of first- order derivatives, the gradient magnitude is then computed as:

$$|\nabla L| = \sqrt{L_x^2 + L_y^2}$$

while the gradient orientation can be estimated as

$$\theta = \text{atan2}(L_y, L_x)$$

2.6 Thresholding and linking

After the project computed as a measure of edge strength, the next stage is to apply a threshold. The lower boundary the threshold. Conversely a high threshold may miss subtle edges, or result in fragmented edges.

If the edge thresholding is applied to just the gradient magnitude image, the resulting edges will be in general be thick and some type of edge thinning post-processing is necessary. For edges detected with non-maximum suppression however, the edge curves are thin by definition and the edge pixels can be linked into edge polygon by an edge linking (edge tracking) procedure. On a discrete grid, the non-maximum suppression stage can be implemented by estimating the gradient direction using first-order derivatives, then rounding off the gradient direction to multiples of 45 degrees, and finally comparing the values of the gradient magnitude in the estimated gradient direction.

The most frequently used approach to manage the problem of appropriate thresholds is by using thresholding with hysteresis. This method uses multiple thresholds to find edges and begin by using the upper threshold to find the start of an edge. If it has the starting point then trace the path of the edge pixel by pixel and mark an edge whenever it has above the lower threshold. It stops marking our edge only when the value falls below our lower threshold. This approach makes the assumption that edges are likely to be in continuous curves, and allows us to follow a faint section of an edge it has previously seen, without meaning that every noisy pixel in the image is marked down as an edge. This project may have the problem of choosing appropriate thresholding parameters and suitable thresholding values may vary from image to image.

2.7 Second-Order Approaches to Edge Detection

Edge-detection operators are based upon second-order derivatives of the intensity. It captures the rate of change in the intensity gradient. Thus, in the ideal continuous case, detection of zero-crossings in the second derivative captures local maxima in the gradient. Marr-Hildreth operator is based on the detection of zero-crossing of the Laplacian operator. However, this operator will also return false edges corresponding to local minima of the gradient magnitude. Moreover, this operator will give poor localization at curved edges.

2.8 Differential Edge Detection

A more refined second-order edge detection approach, which has automatically gives edges with sub-pixel accuracy, is by using the following *differential approach* of detecting zero-crossings of the second-order directional derivative in the gradient direction: Following the differential geometric way of expressing the requirement of non-maximum suppression proposed by Lindeberg. Let us introduce at every image point a local coordinate system (u,v) , with the v -direction parallel to the gradient direction. Assuming that the image has been presmoothed by Gaussian smoothing and a scale-space representation $L(x,y;t)$ at scale t has been computed, this project can require that the gradient magnitude of the scale-space representation, which is equal to the first-order directional derivative in the v -direction L_v , should have its first order directional derivative in the v -direction equal to zero

$$\partial_v(L_v) = 0$$

while the second-order directional derivative in the v -direction of L_v should be negative, i.e.,

$$\partial_{vv}(L_v) \leq 0.$$

written out as an explicit expression in terms of local partial derivatives $L_x, L_y \dots L_{yyy}$, this edge definition can be expressed as the zero-crossing curves of the differential invariant

$$L_v^2 L_{vv} = L_x^2 L_{xx} + 2 L_x L_y L_{xy} + L_y^2 L_{yy} = 0,$$

that satisfy a sign-condition on the following differential invariant

$$L_v^3 L_{vvv} = L_x^3 L_{xxx} + 3 L_x^2 L_y L_{xxy} + 3 L_x L_y^2 L_{xyy} + L_y^3 L_{yyy} \leq 0$$

where L_x, L_y, \dots, L_{yyy} denote partial derivatives computed from a scale-space representation. In this way, the edges will be automatically obtained as continuous curves with subpixel accuracy. Hysteresis thresholding can also be applied to these differential and subpixel edge segments.

First-order derivative approximations can be computed by central differences. Second-order derivatives can be computed from the scale-space representation.

The development in edge detection techniques takes a frequency domain approach to find edge locations. Phase Coherence method is to attempt to find locations in an image. The advantage

of this technique is that it responds strongly to Mach bands, and avoids false positives. A roof edge is a discontinuity in the first order derivative of a grey-level profile.

3. IMAGE RETRIEVAL

Image retrieval algorithms based on Image Primitive features like color, texture and shape. Combinational features are specified to give good performance. On each feature more efficient algorithms are used to retrieve the information from data set.

3.1 Image Retrieval Based on Texture

3.1.1. Pyramid-Structured Wavelet Transform

The Wavelet Transform is one of the suitable approach for signals consisting of components with information concentrated in lower frequency channels. Because of an innate - image properties it allows for most information to exist in lower sub-bands then the pyramid-structure wavelet transform is highly sufficient.

With the help of this transform, the texture image is decomposed into four sub images i.e., low-low, low-high, high-low and high-high sub-bands. At this point, the energy level of each sub-band is calculated. This is the first level decomposition. With the help of the low-low sub-band for further decomposition then reached fifth level decomposition. The main reason for this is the basic assumption that the energy of an image is concentrated in the low-low band.

3.1.2. Energy Level

Energy Level Algorithm:

- Decompose the image into *four* sub-images
- Calculate the energy of all decomposed images at the same scale, using [2]:

$$E = \frac{1}{MN} \sum_{i=1}^m \sum_{j=1}^n |X(i, j)|$$

where M and N are the dimensions of the image, and X is the intensity of the pixel.

- Repeat from step 1 for the low-low sub-band image until index $ind = 5$. Increment ind .

By using the above algorithm, the energy levels of the sub-bands were calculated. This will be repeated for five times, to reach fifth level decomposition.

3.1.3. Euclidean Distance

Euclidean Distance Algorithm:

- Decompose query image.
- Get the energy of the first k channels.
- For image i in the database obtain the k energy.
- Calculate the Euclidean distance between the two sets of energies, using [2]:

$$D_i = \sum_{k=1}^k (x_k - y_{i,k})^2$$

- Increment i . Repeat from step 3.

With the help of this algorithm, the query image is searched for in the image database. The Euclidean distance will be calculated between the query image and every image in the database. This process must be repeated until all the images in the database have been compared with the query image. Based on the completion of implementation of the Euclidean distance algorithm, we have an array of Euclidean distances, which is then sorted.

4. RESULTS AND ANALYSIS

The image database has used to retrieve the relevant images based on query image. The test image database contains 200 images on various categories. Image retrieval has performed on hybrid feature texture and shape. The results of Image Retrieval are shown below:

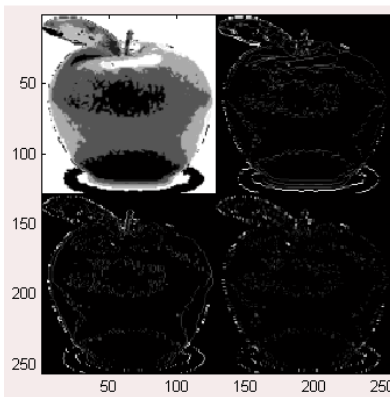


Figure 2: Query Image of Apple

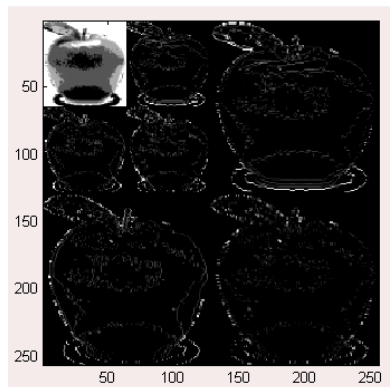


Figure 3: First Decomposition Figure

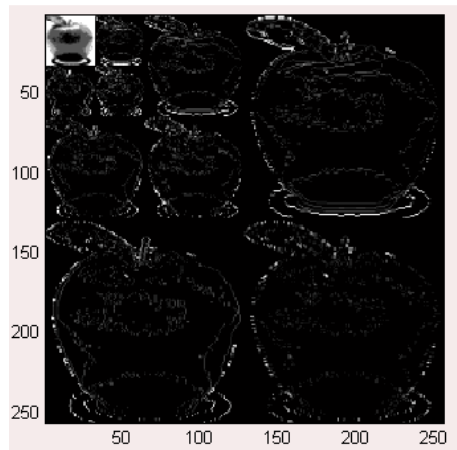


Figure 4: Second Decomposition

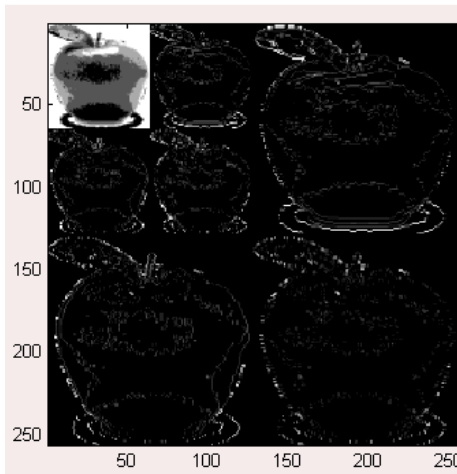


Figure 5: Third Decomposition

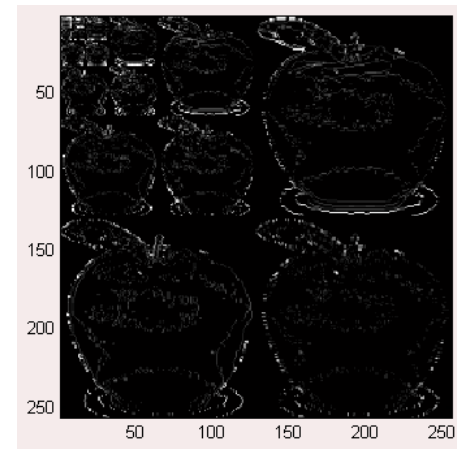


Figure 6: Fourth Decomposition

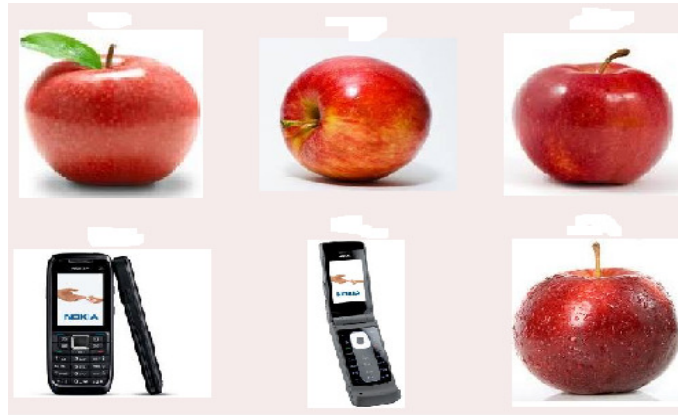


Figure 7: Fifth Decomposition of Apple Image

From the figure, the application called decomposition uses the wavelet decomposition technique to decompose the query image that passes through the five levels of decomposition, this wavelet decomposition technique shows that the texture features of a selected image. After decomposition of an image the user will get the desired images, which has same texture and shape features, for this purpose the user using wavelet decomposition technique.

5. PERFORMANCE ANALYSIS

Based on commonly used performance measures in information retrieval, two statistical measures were computed to assess system performance namely Recall and Precision. For good retrieval system ideal values for recall is 1 and precision is of low value.

5.1 Precision

Precision gives the accuracy of the retrieval system. Precision is the basic measures used in evaluating the effectiveness of an information retrieval system. Precision consists of the proportion of relevant images that are retrieved.

$$\text{Precision} = \frac{(\text{Number of Relevant Images Retrieved})}{(\text{Total Retrieved Images})}$$

Table 1: Precision on Hybrid Features(Texture and Shape)

Data base category	P(1)	P(2)	P(3)	P(4)	P(5)
Image1	0.75	0.857	0.833	0.833	0.833
Image2	0.625	0.714	0.833	0.833	0.666
Image3	0.625	0.714	0.666	0.833	0.833
Image4	0.75	0.857	0.833	0.833	0.833
Image5	0.625	0.714	0.833	0.833	0.833

Table 1 shows the precision of the retrieved images. Here P(1) is for texture decomposition at first level and P(2),P(3),P(4),P(5) are for further decompositions. It can be observed that the

average precision value P (1) is 1 for all the classes which means that all the retrieved images are relevant. On the whole, average precision at first three images, about half of the images retrieved are relevant to the query images. The average precision rate is nearly 66% when first three images are retrieved.

5.2 Recall

Recall consists of the proportion of target images that have been retrieved among all the relevant images in the database. Recall gives the measurement in which how fast the retrieval system works. It also measures how well the CBIR system finds all the relevant images in a search for a query image.

$$\text{Recall} = \frac{(\text{Number of Relevant Images Retrieved})}{(\text{Total Number of Relevant Images})}$$

Table 2 shows the recall of the retrieved images. It gives the average recall for each class of image. Recall 1 means all the relevant image in the database are retrieved. From the table it can be observed that almost all the relevant images are retrieved from the database when the first seven images are retrieved.

Table 2: Recall on Hybrid Features(Texture and Shape)

Database category	R(1)	R(2)	R(3)	R(4)	R(5)
Image1	1	1	1	1	1
Image2	0.833	0.833	0.833	0.833	0.8
Image3	1	1	1	1	0.833
Image4	1	0.857	0.833	0.666	0.714
Image5	1	0.833	0.833	0.833	0.833

6. CONCLUSION AND FUTURE ENHANCEMENTS

This paper elucidates the potentials of extraction of features of the image using texture and shape for retrieving the images from the specific image databases. The images are retrieved from the given database of images by giving the query image using texture and shape features. These results are based on various digital images of dataset. The performance of the image retrieval was assessed using the parameters recall rate and precision. It was ascertained that the recall rate and precision are high when the image retrieval was based on the Hybrid Features texture and shape than primitive features alone. This work can be extended further on huge data bases for retrieving relevant image. This can be extended for pixel clustering to obtain objects using different combination of weight for color and texture and shape features. It can maximize the performance by choosing the best combination between these weights.

REFERENCES

- [1]. "Adaptation in the Quincunx wavelet filter bank with in image denoising,"in Proc. of the 2004 Int. TICSP Workshop on Spectral Methods and Multirate Signal Processing, SMMSP 2004, pp. 245- 252., 2004.
- [2]. Aigrain, P et al (1996) "Content-based representation and retrieval of visual media – a state-of-the-art review" *Multimedia Tools and Applications*

- [3]. Alsuth, P et al (1998) "On video retrieval: content analysis by Image Miner" in *Storage and Retrieval for Image and Video Databases*
- [4]. Androutsas, D et al (1998) "Image retrieval using directional detail histograms" in *Storage and Retrieval for Image and Video Databases*
- [5]. Bjarnestam, A (1998) "Description of an image retrieval system", presented at *The Challenge of Image Retrieval*
- [6]. Craig Nevill-Manning (Google) and Tim Mayer (Fast). *Anatomy of a Search Engine, Search Engine Strategies Conference, 2002*
- [7]. Daubechies, I., (1992), "Ten Lectures on Wavelets," *Capital City Press, Montpelier, Vermont*
- [8]. Drimbarean A. and Whelan P.F. (2001) Experiments in color texture analysis, *Pattern Recognition Letters*, 22: 1161-1167.
- [9]. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. and Yanker, P., (1995), "Query by image and video content: The QBIC system," *IEEE Computer*.
- [10]. Hermes, T et al (1995) "Image retrieval for information systems" in *Storage and Retrieval for Image and Video Database*
- [11]. J. Smith and S. Chang. *VisualSEEK: A fully automated content-based image query system*. ACM Multimedia: 87-98, November 1996.
- [12]. Jain, R (1993) "Workshop report: NSF Workshop on Visual Information Management Systems" in *Storage and Retrieval for Image and Video Databases*
- [13]. John R. Smith and Shih-Fu Chang. VisualSEEk: a fully automated content-based image query system, *ACM Multimedia 96, Boston, 1996*
- [14]. Petkovic, D (1996) "Query by Image Content", presented at *Storage and Retrieval for Image and Video Databases*
- [15]. Sergey Brin and Lawrence Page (Google Founders). *Brisbane, Australia, 1998*.
- [16]. Shou-Bin Dong and Yi-Ming Yang. Hierarchical Web Image Classification by Multi-level Features, *Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, 2002*
- [17]. Unser, M., (1995), "Texture classification and Segmentation Using Wavelet Frames," *IEEE Transaction Image Processing*
- [18]. W. Ma, Y. Deng and B. S. Manjunath. *Tools for texture/color based search of images*. SPIE International conference - Human Vision and Electronic Imaging: 496-507, February 1997.
- [19]. W. Niblack, R. Barber, and et al. *The QBIC project: Querying images by content using color, texture and shape*, Feb 1994.