CONTENT BASED IMAGE RETRIEVAL USING HYBRID FEATURE EXTRACTION TECHNIQUES

A PROJECT REPORT

Submitted by

CB.EN.U4EIE14005	**AKSHAYA. B**
CB.EN.U4EIE14034	**NIRANJANA SATHISH. A**
CB.EN.U4EIE14049	**SHOBIKA. K**

Under the guidance of
Mrs. KARTHIKA. R

in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND INSTRUMENTATION ENGINEERING

AMRITA SCHOOL OF ENGINEERING
AMRITA VISHWA VIDYAPEETHAM
COIMBATORE 641112

May 2018

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF ENGINEERING, COIMBATORE, 641112

BONAFIDE CERTIFICATE

This is to certify that the project report entitled "CONTENT BASED IMAGE RETRIEVAL USING HYBRID FEATURE EXTRACTION TECHNIQUES"

Submitted by

CB.EN.U4EIE14005	**AKSHAYA. B**
CB.EN.U4EIE14034	**NIRANJANA SATHISH. A**
CB.EN.U4EIE14049	**SHOBIKA. K**

in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in ELECTRONICS AND INSTRUMENTATION ENGINEERING is a Bonafide record of the work carried out under my guidance and supervision at Amrita School of Engineering, Coimbatore.

Project Advisor
Mrs. KARTHIKA. R
Asst. Professor

Project Coordinator
Mr. P. SUDHEESH
Asst. Professor

Chairman ECE
Dr. M. JAYAKUMAR

The project was evaluated by us on:

Internal Examiner

External Examiner

DEDICATED TO

PEOPLE IN PURSUIT OF KNOWLEDGE

ACKNOWLEDGEMENT

First and foremost, we would like to thank our guide, **Mrs. KARTHIKA. R**, Assistant Professor, Department of ECE, for her constant encouragement and guidance throughout the project.

We express our sincere gratitude to **Mr. PARGUNARAJAN. K**, Class Advisor, and **Mrs. L. VIDYA,** Class Counsellor, Department of ECE for their valuable support and advice during our tenure here.

We express our heartfelt gratitude to **Dr. M. JAYAKUMAR**, Chairperson, Department of ECE.

We also express our sincere thanks to **Dr. KARTHI BALASUBRAMANIAN**, Assistant Professor and **Mrs. KARTHIGA BALAMURUGAN**, Assistant Professor, Department of ECE for their help in correcting our mistakes and guiding us in various aspects for the project.

We would also like to thank all teaching and non-teaching staffs for providing facilities promptly and also our friends who helped us a lot in the successful completion of our project.

ABSTRACT

Images consist of visual components such as colour, shape and texture. These components stand as the primary basis with which images are distinguished. A content based image retrieval system uses these primary features of an image and checks similarity of the extracted features with those of the image given by the user. A group of images similar to the query image fed are obtained as a result. This paper proposes a new methodology for image retrieval using the local descriptors of an image in combination with one another. HSV histogram, Colour moments, Colour auto-correlogram, Histogram of Oriented Gradients and Wavelet transform are used to form the feature descriptor. In this work, a comparative study of various retrieval techniques by combination of features is performed. The result is tested with various distance metrics and a supervised learning algorithm, SVM is used for classification of the images because it has high retrieval efficacy.

TABLE OF CONTENTS

LIST OF ABBREVATION

ABBREVATION	EXPANSION	PAGE NO
SVM	Support Vector Machine	i, 4, 6, 9, 28, 29, 30, 41
CBIR	Content-Based Image Retrieval	2, 3, 4, 6, 7, 10, 41, 43
GVF	Gradient Vector Flow	6
SIMPLIcity	Sensitive Integrated Matching for Picture Libraries	43
RGB	Red, Green, Blue	10
HSV	Hue, Saturation, Value	i, 3, 10, 13
HOG	Histogram of Oriented Gradients	4, 7, 17, 18, 38
SIFT	Scale-Invariant Feature Transform	7
SURF	Speeded-Up Robust Features	7
DWT	Discrete Wavelet Transform	15, 16

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

An image can be defined as a two dimensional function f(x, y). Here, x and y are the spatial coordinates. A pixel or intensity of an image can be defined as the amplitude of f at any pair of coordinates (x, y).An image is called a digital image if x and y of f(x,y) are finite and discrete quantities. Digital image processing includes processes where inputs and outputs are images and the processes vary from extracting the attributes from images to object recognition.

1.1 CONTENT BASED IMAGE RETRIEVAL

Humans, from the earliest age, learn to communicate. Over the years, the means have refined, but the underlying necessity to communicate remains. In a world teeming with technology where cross country communication lines are established in seconds, it becomes a fundamental necessity to enhance the quality of communication. As a result of the technological outburst within the past couple of decades, humungous amounts of data, metadata and visual data are generated with every click of mouse and swipe of screen.

With a splurge of visual data, it is a vastly cumbersome task to scan tens of thousands of images manually. A content based image retrieval system extracts the basic features of every image in a dataset and compares the same with those of the image provided by the user. General flow states that post the query matching, the system ranks the images in a descending order of similarity with the given input query and the output is all the images that are ranked highest. Every image will have three basic components: Shape, Colour, and Texture. A CBIR relies on these extracted features, individually or combinations of them, to extract and run similarity algorithms on them.

While CBIR completely relies on the contents of the image itself, image retrieval using the metadata of images can also be done. There is textual data associated with each image and traditional methods of retrieval, such as retrieval using keywords, can be done. While the annotation process is time consuming and laborious, there is also an additional limitation with this system which pertains to lack of standardization. Meaning, no two users perceive an image in the same way. Since there is no standard way to perceive an image, multiple users will give multiple annotations which are not likely to match. This is will also result in large amounts of junk data which is undesirable.

Thus, CBIR is more suitable for large amounts of visual data. The most basic means of comparison between two images is a colour histogram. But, a colour histogram is sensitive to change in brightness and doesn't take spatial information into account. As a result, the accuracy of the outcome is not satisfactory. In this paper, we propose to extract colour, texture and shape features in order to maximize accuracy of extraction and reduce junk.

There are two kinds of CBIR:

General CBIR: This involves image retrieval from a vast database of mixed images.

Application specific CBIR: This involves matching of a query image from a specific database like X-ray images, brain scans, etc.

In this paper, we propose to extract colour, texture and shape features in order to maximize accuracy of extraction and efficacy of output. We propose to use the best of multiple techniques to improve quality of output and reduce error margins.

1.2 IMAGE DESCRIPTORS

The most basic means of comparison between two images is colour. Techniques used to retrieve colour feature in this paper, are HSV histogram (Colour Histogram), Colour Auto-correlogram and colour moments. Colour Histogram is a representation of colours in an image. Although simple and straightforward, colour histogram is sensitive to change in brightness and doesn't take spatial information into account. A colour auto-correlogram, which is an indexed table of colour pairs and their probabilities, includes spatial correlation of colours and is easy to compute. Colour moments are simplistic distribution of colour in the image. In order to compensate for the deficits of each technique, we propose to use a combination of three techniques and leverage their best outcomes for colour feature extraction.

Wavelets are a more general way to represent and analyze multi-resolution images. Thus, shape feature extraction is accomplished using wavelet transform and Histogram of Oriented Gradients. The HOG technique counts the number of occurrences of a specific orientation in each part of the image. And Wavelet transform is a multi-resolution filtering technique that eliminates noise efficiently. Furthermore, Gabor Wavelet is used for the extraction of texture feature. The Gabor representation minimises uncertainty in space and frequency dimensions and the micro features extracted characterize texture information.

Additionally, to enhance the results of a robust dataset, we also propose to use SVM classification. SVM is the most efficient supervised learning algorithm for image recognition, face recognition, speech recognition and face detection. It is reliable, accurate and is most efficient for binary classification. Since, more weight age is given to accuracy than response time, we propose to use SVM classification.

The architecture of the proposed CBIR system is shown in Fig.1

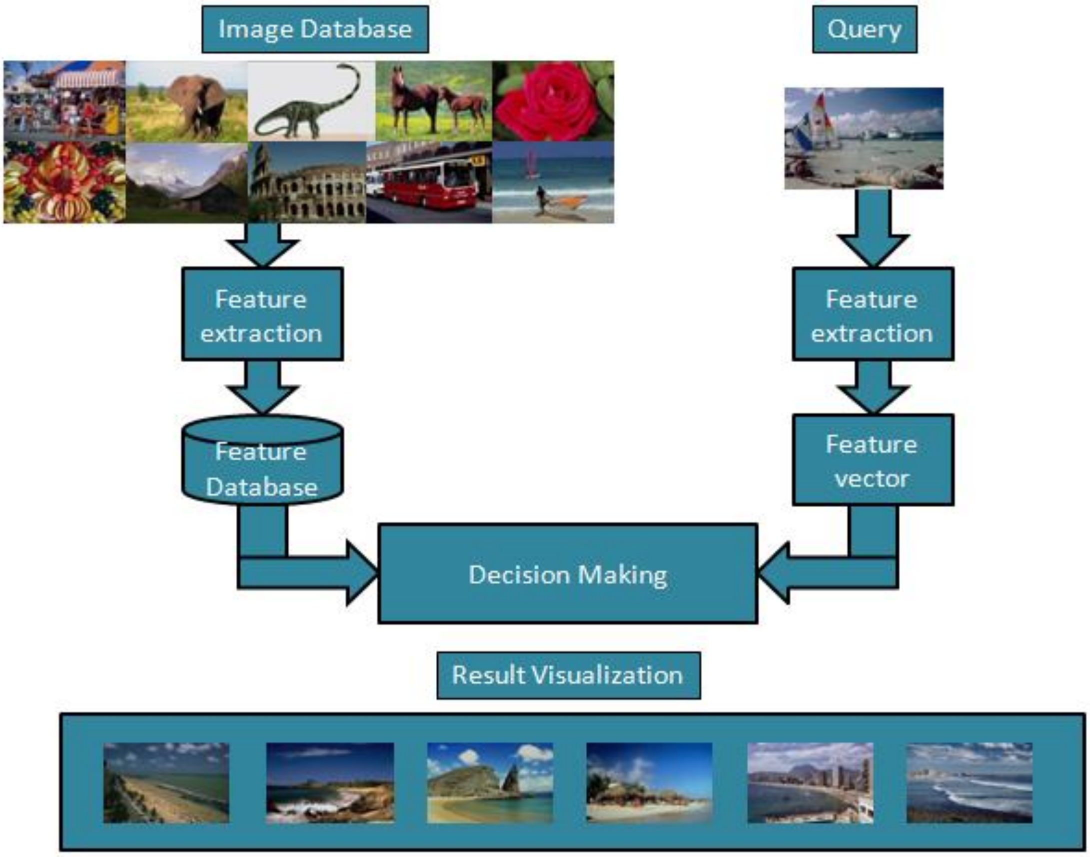

Fig. 1.2 Architecture of the proposed CBIR system

CHAPTER 2
LITERATURE SURVEY

2. LITERATURE SURVEY

The properties of Hue, Saturation and Intensity values colour space is analyzed in [1]. The values are varied and the visual perception is studied. The saturation value is used to decide if the Hue or the Intensity of the pixel is closer to human perception.

Various CBIR tools are compared in [2]. From this comparison, it is observed that most of the systems use colour and texture features. Shape feature is not as common. Layout feature is very rarely used. Retrieval techniques based on a single feature worked well only for a specific set of images.

One of the most commonly used colour feature in CBIR system is colour histogram, [3]-[5]. Colour Histogram concentrates only on the proportion of the number of various types of colours in an image, but does not focus on the spatial location of the colours. Noise is not handled efficiently by histograms because they are very sparse. To overcome the drawback of this feature, features such as colour-correlogram and colour moments are applied. Pre-processing the images will increase the accuracy.

In [6], it is observed that colour and texture features are used. Support Vector Machine (SVM) and Euclidean distance are applied to retrieve similar images.

Different approaches of different combinations of colour, shape and texture retrieval is compared in [7]. When colour (Colour histogram) and texture features (standard wavelet) are combined, accuracy was enhanced but feature set was inadequate. On merging colour, texture and shape feature (Colour moment, Gabor filter, Gradient Vector Flow (GVF), strong feature set was created.

Ecosembles is formed by concatenating different combinations of weak feature-views (colour, shape, texture, etc.), which must be extracted from images. Histograms are extracted from a group of images with varying numbers of bins for each histogram. By examining the Intelligence, Surveillance and Reconnaissance data, it is understood that Ecosembles performed slightly better than GIST descriptors coupled with SVM. This is observed in [8].

In [9], images are retrieved separately using the features like colour Histogram, Gabor and wavelet transform for texture, and Shape information from Phase congruency (edge detection for any change in illumination and contrast in the image). A combination of these produced an accuracy of 96.4%.

A system for biometric security for CBIR is developed based on the extraction of Shape (moment invariant), Colour (Histogram) and Texture (Gabor wavelet) in [10].

The colour and texture features are concatenated where WBCH (Wavelet based Colour Histogram) method is used. The precision of this proposed method is found to be better. The computational steps are reduced with the help of wavelet transform thereby increasing the retrieval speed [11].

In [12], a relative study on several features like merged colour histogram and Gabor transform is performed.

With only statistical entities of the first order such as mean and standard deviation, Gabor wavelet showed better classification results. This method is proven to be slightly superior to the co-occurrence matrix which is usually used for texture classification [13].

A CBIR system in which the features like HOG, SIFT, SURF and colour histogram are used to extract the features of the image and formed a collection of local feature vectors is observed in [14].

CHAPTER 3

PROPOSED WORK

3. PROPOSED WORK

The main objective of the proposed work is to develop an efficient image retrieval system by showing a comparative study on retrieval using different features. Initially, images are retrieved with colour, shape and texture separately. Further, all three features are combined and results are compared. They are then retrieved using various distance metrics. To increase the retrieval efficiency, Support Vector Machine (SVM) algorithm is used for classification.

The proposed system consists of the following components.

3.1 QUERY IMAGE

The query image is the input given by the user for testing. This image is found in the dataset.

3.2 IMAGE DATASET

WANGS DATASET

Wang dataset consists of 1000 images of 10 different classes. Each class has 100 images. The classes include Africa, Rose, Beach, Monument, Bus, Horse, Food, Dinosaur, Scenery and Elephant. The sizes of the images are either 384 x 256 or 256 x 384.

Fig. 3.2 Sample images from Wang Dataset

3.3 FEATURE EXTRACTION

CBIR relies on retrieving stored images from a dataset by comparing the features that are automatically extracted by the system. The low level features that are extracted include colour, shape, texture etc.

The features of an image are extracted and stored in a vector. The features used are listed below.

3.3.1 COLOUR FEATURE

Colour is one of the most common features used for CBIR. One of the most important features that make possible the recognition of images by humans is colour. The property of colour depends on the reflection of light and the processing of the information in the brain.

It is due to the following reasons:

It simplifies feature extraction and identification.

Colours can be defined in three dimensional spaces. They can be RGB (Red, Green, and Blue), HSV (Hue, Saturation, and Value) or HSB (Hue, Saturation, and Brightness). Hence, it is superior to the 1 dimensional gray scale spaces. RGB colour space corresponds closely to the way humans perceive colour.

The colour descriptors used to form the feature vector are as follows:

3.3.1.1 HSV HISTOGRAM

HSV colour model which means hue saturation and value, it is a cylindrical coordinate representation of points in an RGB colour model. This histogram is a three dimensional representation of HSV colour space is a hexacone. Consider a cylinder where the edge which is an angle, S which is a radius and V which is a height. This is the conversion from RGB colour model into a cylinder. In this H is a Hue which is the dominant colour as perceived by an observer, S is for saturation which means the amount of white light mixed with Hue and V is Value which denotes the chromatic notion of intensity of the colour, For example if we have a blue colour then you will point to this blue colour the saturation which is the amount of white light mix it with all and value is the chromatic notion of the intensity they lower the value the similar to black the higher value the more similar to the colour itself and we change these three parameters in order to create different combinations of colours and also create all different possible colour. Changing the saturation from 0 to 1 change the colour from shades of grey to pure form of any colour and similarly lowering the saturation value will change the colours back to shades of grey. Therefore, we use the saturation to determine which is more pertinent to human visual perception, hue or intensity

$$H = cos^{-1}\left\{\frac{\frac{(R-G)+(R-B)}{2}}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\} \tag{1}$$

$$S = 1 - \frac{3}{R-G-B}[\min(R,G,B)] \tag{2}$$

$$V = \frac{1}{3}(R,G,B) \tag{3}$$

The range of Hue defined is $[0,2\pi]$ with respect to Red axis where the angle of Red is at 0 and again at 2π, Green at $2\pi/3$and Blue at $4\pi/3$. When saturation value is 0 all the pixels in the image look the same, the true colours are separated and visually perceived while increasing the Hue values. The transition which is dependent on the intensity of the image is determined by Saturation threshold, for low intensity the colour of the pixel is close to grey value even if the saturation is high and vice versa. Saturation also gives details about depth of the colour which concludes that human eye is less sensitive to its variation than intensity or Hue variation. Hence saturation is used to determine which is more pertinent to human eye, Hue or intensity, and ignores the saturation if the pertinent is more. The threshold of saturation is calculated using the following formula:

$$\mathrm{th}_{\mathrm{sat}}(V) = 1.0 - \frac{0.8V}{255} \tag{4}$$

In the above formula, if V = 0 then it means that $\mathrm{th}_{\mathrm{sat}}(V)$ =1 which signifies that all colours are same as black irrespective of Hue and saturation value

Extracting colour histogram has two parts, first, representation of Hue in the range of $[\mathbf{0}, \mathbf{2\pi}]$second, quantization set of gray values. The number of components in the feature vector is determined by the following equations:

$$N_h = \lceil 2\pi MULT\_\_FCTR \rceil + 1 \quad (5)$$

$$N_g = \left\lceil \frac{Imax}{DIV\_FCTR} \right\rceil \quad (6)$$

Where,

MULT__FCTR = quantization level for Hues

DIV_FCTR = number of quantized gray level

$Imax$ = maximum value of the intensity

Each H, S, V component is quantized into (8*2*2) bin which gives a vector of 32 dimensions.

Fig. 3.3 (a) Before applying HSV

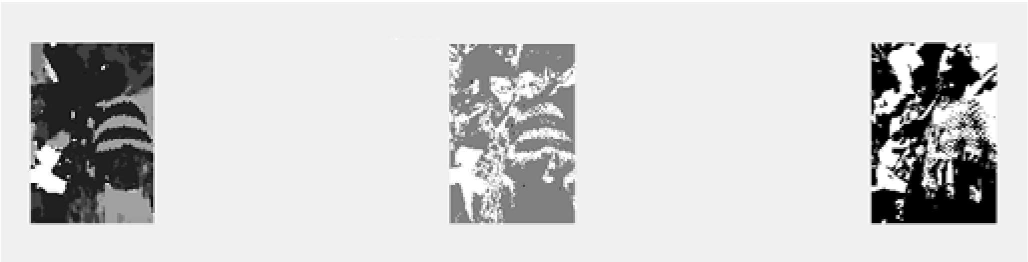

Fig. 3.3 (b) Quantized by HSV

3.3.1.2 COLOUR AUTOCORRRELOGRAM

The change in spatial correlation of colours with respect to distance is expressed through colour correlogram. The colour distribution in an image is captured using colour histogram which does not include spatial correlation. The data of correlogram are collected and stored in the form of a table which is indexed by pairs of colours in an image (*i*, *j*), the entries in the table shows the probability of finding pixels between *i* and *j* with distance *d*. Therefore, auto correlogram proved to show spatial correlation for identical colours only.

The highlights of the feature are:

1. Spatial correlations of colours are included.

2. The feature size is small.

3. The global distribution of spatial correlation of colours can be described.

4. Easy to compute.

3.3.1.3 COLOUR MOMENTS

The main purpose of colour moments is colour indexing, as a feature in image retrieval process which is done to compare how two images are similar based on colour. This process takes several steps like one image is given as query which is compared to a database of digital images with certain pre-computed features to find the similar image and retrieve it as well. Every comparison between query image and images in database is collected and stored as a similarity score. The retrieval is processed based on the similarity score, lower similarity score determines that those two images have more identical features and higher score denotes that the images have less similar features. The advantage of using colour moments are they are scaling and rotational invariant. These can be used in lighting conditions because they contain both shape and colour information of an image. Computing moments in probability distribution is similar to computing colour moments. Three colour moments are computed per channel as following methods.

MEAN

The average of the colour in an image is interpreted as first colour moment, and it is calculated using the following formula

$$E_i = \sum_{j=1}^{N} \frac{1}{N} P_{ij} \tag{7}$$

Where,

N = Number of pixels in the image

P_{ij} = Value of the j^{th} pixel in the image at i^{th} colour channel.

STANDARD DEVIATION

Standard deviation is the square root of the variance of the colour distribution in the image; standard deviation is the second colour moment. It is calculated with following formula

$$\sigma = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(P_{ij} - E_i\right)^2\right)} \tag{8}$$

where,

E_i = Mean value for i^{th} colour channel of the image

3.3.2 SHAPE FEATURE

3.3.2.1 WAVELET TRANSFORM

During the shape extraction process, edges and contours are used to identify the images. Features of the object shape are computed for each image and stored in the feature vector. These features are independent of the size and orientation of an object.

Shape features are then calculated for query image. Images retrieved from the dataset are images that are most similar to the query image.

EDGE

Points in an image where intensity of the image changes suddenly mainly due to shadows, texture or geometry are called edges. To prevent noisy pixels, images are pre-processed before applying the edge detector algorithms.

A wavelet is an irregular and asymmetric waveform of finite duration which has zero mean value. They have varying frequency. They are more flexible because functions can be altered and enlarged for signal analysis. Wavelet analysis is suitable for time-frequency analysis can be applied to 1D data and 2D data.

The key difference between Short time fourier transform and wavelet transform is that in STFT, the window is fixed whereas in wavelet transform it is not fixed.

In Discrete Wavelet Transform (DWT), the filter looks for local maxima in a wavelet domain. This technique is used for wavelet edge detection.

In DWT, discrete time signals are transformed into discrete wavelet representations. Filters are used to calculate DWT instead of matrices. Two filters: High frequency and low frequency filters are convolved with the function. In the function obtained, one sample out of two is suppressed.

In wavelet transform, it is possible to select the size of the details that will be detected. Wavelet analysis is done separately for the horizontal and the vertical directions in a two dimensional image processing. Hence, these edges are detected separately. The 2D discrete wavelet transform (DWT) breaks the images into sub-images which contains three details (horizontal detail, vertical detail and diagonal detail) and one approximation. The approximation looks similar to the input image but only 1/4 of original size. The 2-D DWT is an addition to the 1-D DWT in both the horizontal and the vertical direction. The approximation (smoothing image) contains the most information of the original image. Horizontal edge details, vertical edge details, diagonal details and the approximation image are used to form the sub-image.

These sub-images are the passed through another set of low pass and high pass filters. These iterative processes are used for multi-resolution analysis.

COIFLET

Coiflet wavelet is another form of Daubechies wavelets. Coiflet are orthogonal wavelets for which ψ and φ has several vanishing moments. Hence orthogonality condition and both the moment conditions have to be satisfied. Coiflet wavelet family is shown in Fig. 4.

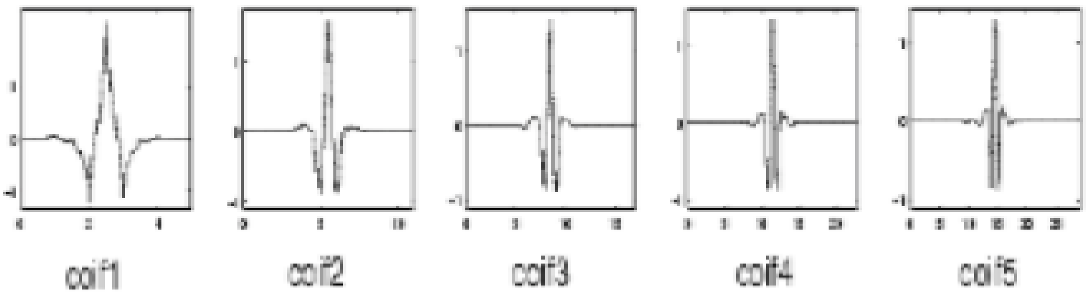

Fig. 3.3 (c) Coiflet Wavelet Family

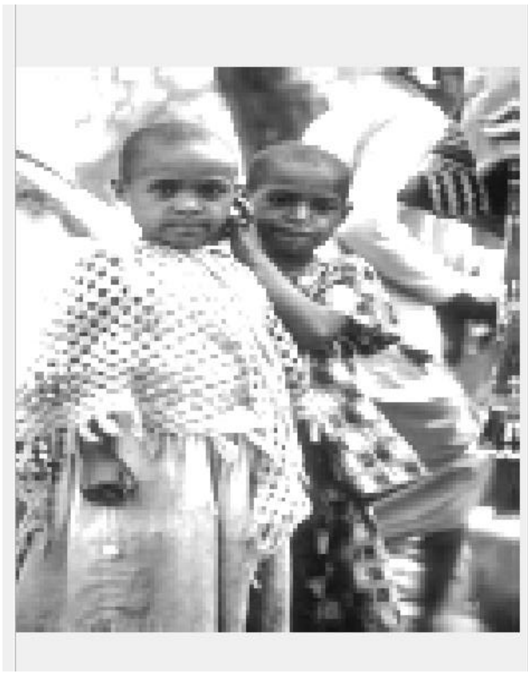

Fig. 3.3 (d) Output image after applying Wavelet Transform

IMPORTANCE OF COIFLET

1. It has higher computational overhead

2. It utilizes wavelet function coefficients and six scaling, so increment in pixel averaging and differencing prompts a smoother wavelet and expanded abilities in several image-processing techniques.

3. This filter follows same procedure as Daubaechies and Haar wavelet. It computes differences and average using only with six adjacent pixels.

4. It follows mirror technique

3.3.2.2 HISTOGRAM OF GRADIENTS (HOG)

The HOG feature is generally used for object detection. It forms a feature vector of 1-N length, where N is the length of the HOG feature.

The following steps are followed to extract HOG information of an image.

After preprocessing, the image is filtered with the following kernals to obtain the horizontal and vertical gradients.

-1	0	1

-1
0
1

A gradient of a single patch will contain two values namely, magnitude and direction. For example, a 7*7 patch will contain 7*7*3 = 147 pixel values. The gradient of this patch will contain 7*7*2 (magnitude and direction) = 98 numbers. Magnitude and direction can be found by using the formula given below:

$$g = \sqrt{g_x^2 + g_y^2} \tag{9}$$

$$\theta = \arctan \frac{g_y}{g_x} \tag{10}$$

A histogram is created by calculating the contribution of all the pixels in the 7*7 cells. The histogram is the normalized. Normalization is performed so that the histogram is invariant to light.

For example: consider a colour vector [32, 64, 128]. The length can be given by $\sqrt{}32^2+64^2+128^2$ = 146.4. A normalized vector is obtained by dividing each element by 146.4 which gives [0.22, 0.43, 0.87]. Now a normalized vector of any multiple of the original vector will give us the same value of [0.22, 0.43, 0.87].

i.e., [32, 64, 128] *2 = [64, 128, 256]

Normalizing this will give [0.22, 0.43, 0.87]

Finally the HOG vector is formed by concatenating all the vectors. The mean of this vector is taken to form a feature vector here.

Fig. 3.3 (e) Output image after applying HOG

3.3.3 TEXTURE FEATURE

3.3.3.1 GABOR WAVELET

Corners, blobs and edges in an image can be found with the help of Gabor wavelets. Gabor functions are mainly used for texture-based image retrieval. Texture features are obtained with the help of these functions. Elements of a family where Gabor functions are alike and are known as wavelets. They are formed by dilation and shift from mother wavelet. To form a feature vector, Gabor wavelet filter is applied with four scaled and six orientations.

Scale: 0.4, 0.2, 0.1, 0.05

Orientations: $\theta = 0$, $\theta_{n+1} = \theta_n + \frac{6}{\pi}$

The feature vector is formed by obtaining the standard deviation and mean of the Gabor wavelet coefficients. The vector is of 48 dimensions.

3.4 SIMILARITY METRIC

The dataset images are trained and the features are stored in the database. A query image is then given by the user and the feature vector is computed and similar images are retrieved. Here, the query image and the dataset images are compared for similarity and this is computed by finding the difference between the dataset feature vector and query feature vector. The images are more similar if the difference between the images is smaller. The above process is done for various distance metrics such as Euclidean distance, City block distance (Manhattan distance) and Cosine distance. Usually, the feature vector's structure decides which distance metric should be used to check the similarity. The metric mathematically shows the resemblance between the input image and each object or region or image in a dataset. For better performance and accurate retrieval, the system should use effective similarity metric. Regardless of its success in the literature, finding a distance metric which is robust is still a challenging issue.

3.4.1 EUCLIDEAN DISTANCE

Euclidean distance is the most commonly used distance metric because of its effectiveness and efficiency. This distance is measured by taking the distance between two vectors by calculating the square root of the sum of the squared absolute differences. It is computed as:

$$D_E = \sqrt{\sum_{i=1}^{n} (|I_i - D_i|)^2} \tag{11}$$

where, I_i and D_i are vectors

3.4.2 CITY BLOCK DISTANCE

This metric is also known as Manhattan distance.

$$D_C = \sum_{i=1}^{n} |I_i - D_i| \tag{12}$$

3.4.3 COSINE DISTANCE

The distance is calculated by finding the angle between the two vectors by using dot product and magnitude.

$$d(X,Y) = 1 - cos\theta = 1 - \frac{X.Y}{\left||X|\right|.\,||Y||} \tag{13}$$

CHAPTER 4

EXPERIMENTS AND RESULT ANALYSIS

4. EXPERIMENTS AND RESULT ANALYSIS

Various experiments were performed in order to show efficiency and accuracy of the proposed method. The performance evaluation of image retrieval is calculated using precision and recall.

PRECISION

Precision is the ratio of retrieved relevant images to the total number of images retrieved.

$$Precision = \frac{no\ of\ relevant\ images\ retrieved}{total\ no\ of\ images\ retrieved} \quad (14)$$

RECALL

Recall is the measure of how many number of truly relevant results are retrieved. A high recall implies that the algorithm has returned most of the relevant images.

$$Recall = \frac{no\ of\ relevant\ images\ retrieved}{no\ of\ relevant\ images\ in\ the\ database} \quad (15)$$

4.1 RETRIEVAL RESULT USING HYBRID FEATURES

The dataset used for the comparative study of features is Wang dataset. The proposed system receives a single query image and returns 20 similar images from the database. The result is tested using test images from each class. There exist 100 images for each class where 95 images are trained. For each query image, there exists 95 relevant images. Initially, images are retrieved using colour, shape and texture features separately. Then, the features are used in tandem with one another and a comparative study is done.

4.2 RETRIEVAL RESULT USING VARIOUS DISTANCE METRICS (WANG DATASET)

It can be concluded that the distance is superior above the other metrics.

Table 1 Values of the performance measure based only on colour – Wang Dataset

Similarity Metric	Manhattan		Euclidean		Cosine	
Performance metric	Precision	Recall	Precision	Recall	Precision	Recall
Africa	0.55	0.12	0.60	0.13	0.50	0.11
Beach	0.70	0.15	0.65	0.14	0.85	0.18
Monument	0.80	0.17	0.15	0.03	0.20	0.04
Bus	0.60	0.13	0.25	0.05	0.30	0.06
Dinosaur	1.00	0.21	1.00	0.21	1.00	0.21
Elephant	0.75	0.16	0.75	0.16	0.75	0.16
Rose	0.90	0.19	0.80	0.17	0.80	0.17
Horse	0.90	0.21	0.90	0.19	0.75	0.16
Mountain	0.55	0.12	0.40	0.08	0.45	0.09
Food	0.95	0.20	0.80	0.17	0.75	0.16
Mean	0.77	0.16	0.63	0.13	0.64	0.13

Table 2 Values of the performance measure based only on shape – Wang Dataset

Similarity Metric	Manhattan		Euclidean		Cosine	
Performance metric	Precision	Recall	Precision	Recall	Precision	Recall
Africa	0.35	0.07	0.35	0.07	0.05	0.01
Beach	0.15	0.03	0.30	0.06	0.10	0.02
Monument	0.35	0.07	0.40	0.02	0.05	0.01
Bus	0.05	0.01	0.05	0.01	0.10	0.02
Dinosaur	1.00	0.21	1.00	0.21	1.00	0.21
Elephant	0.90	0.19	0.80	0.17	0.65	0.14
Rose	0.90	0.19	0.90	0.19	0.90	0.19
Horse	0.80	0.17	0.65	0.14	0.40	0.08
Mountain	0.25	0.05	0.30	0.06	0.30	0.06
Food	0.45	0.09	0.35	0.07	0.35	0.07
Mean	0.52	0.11	0.51	0.10	0.39	0.08

Table 3 Values of the performance measure based only on texture – Wang Dataset

Similarity Metric	Manhattan		Euclidean		Cosine	
Performance metric	Precision	Recall	Precision	Recall	Precision	Recall
Africa	0.25	0.05	0.20	0.04	0.30	0.06
Beach	0.30	0.06	0.15	0.03	0.10	0.02
Monument	0.40	0.08	0.35	0.07	0.35	0.07
Bus	1.00	0.21	0.90	0.19	0.80	0.17
Dinosaur	0.85	0.18	0.70	0.15	0.35	0.07
Elephant	0.65	0.14	0.40	0.08	0.50	0.11
Rose	1.00	0.21	0.95	0.20	0.95	0.20
Horse	0.25	0.05	0.40	0.08	0.40	0.08
Mountain	0.25	0.05	0.30	0.06	0.50	0.11
Food	0.50	0.11	0.30	0.06	0.25	0.05
Mean	0.55	0.11	0.47	0.10	0.45	0.09

Table 4 Values of the performance measure by combining colour, shape and texture – Wang Dataset

Similarity Metric	Manhattan		Euclidean		Cosine	
Performance metric	Precision	Recall	Precision	Recall	Precision	Recall
Africa	0.55	0.12	0.60	0.13	0.50	0.11
Beach	0.70	0.15	0.65	0.14	0.85	0.18
Monument	0.80	0.17	0.15	0.03	0.20	0.04
Bus	0.60	0.13	0.25	0.05	0.30	0.06
Dinosaur	1.00	0.21	1.00	0.21	1.00	0.21
Elephant	0.75	0.16	0.75	0.16	0.75	0.16
Rose	0.90	0.19	0.80	0.17	0.80	0.17
Horse	0.90	0.21	0.90	0.19	0.75	0.16
Mountain	0.55	0.12	0.40	0.08	0.45	0.09
Food	0.95	0.20	0.80	0.17	0.75	0.16
Mean	0.77	0.16	0.63	0.13	0.64	0.13

4.3 COMPARISON OF IMAGE RETRIEVAL USING DISTANCE METRICS (WANG DATSET)

1. MANHATTAN DISTANCE

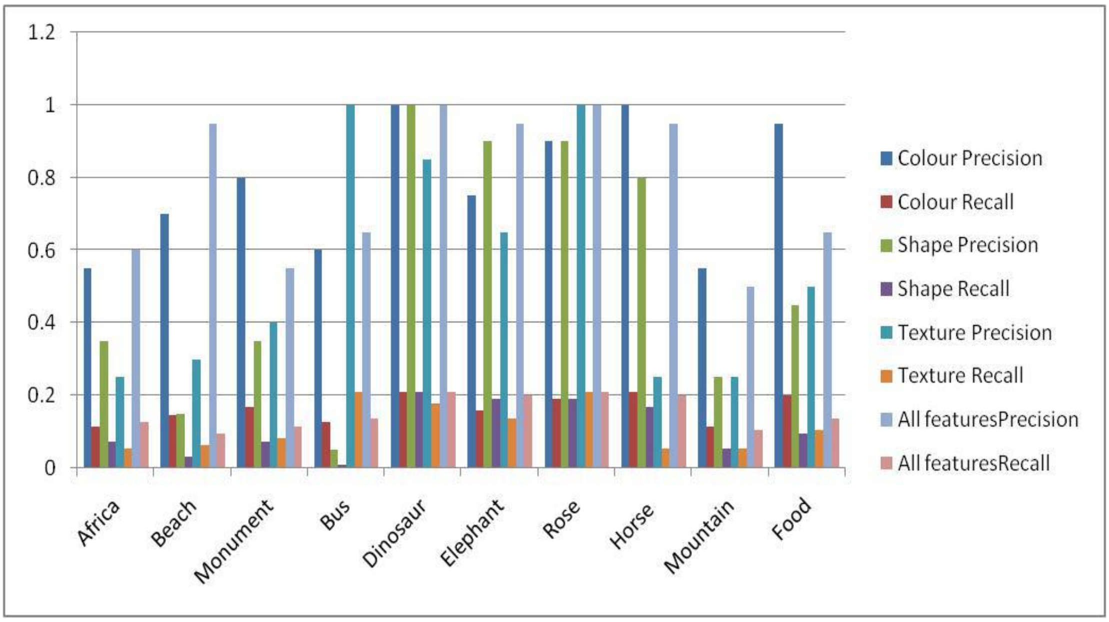

Fig. 4.3 (a) Retrieval using Manhattan Distance

2. EUCLIDEAN DISTANCE

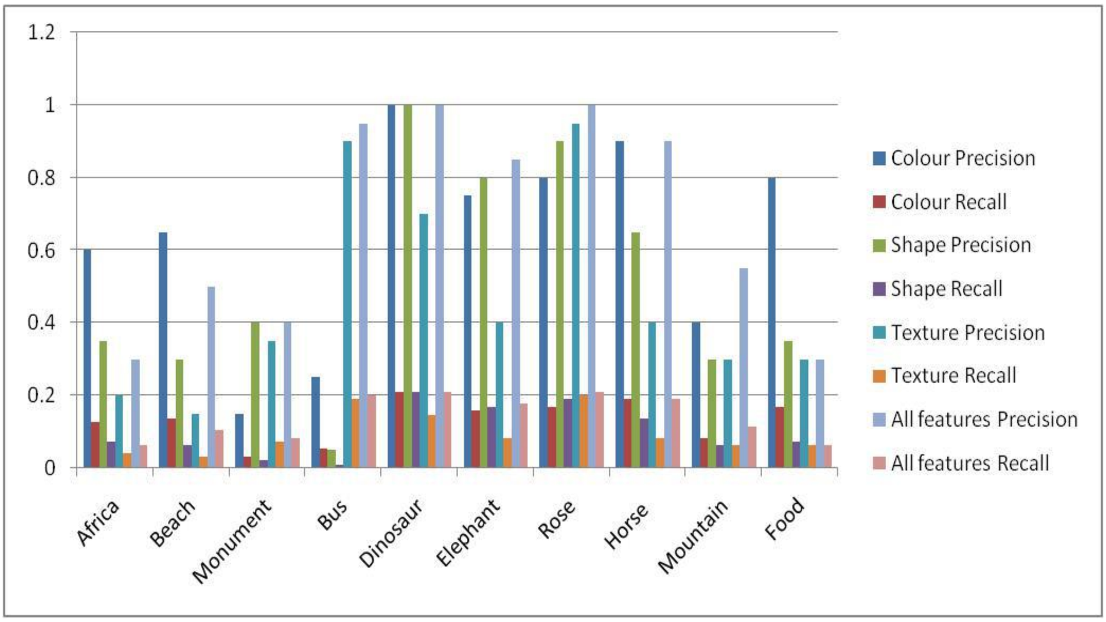

Fig. 4.3 (b) Retrieval using Euclidean Distance

3. COSINE DISTANCE

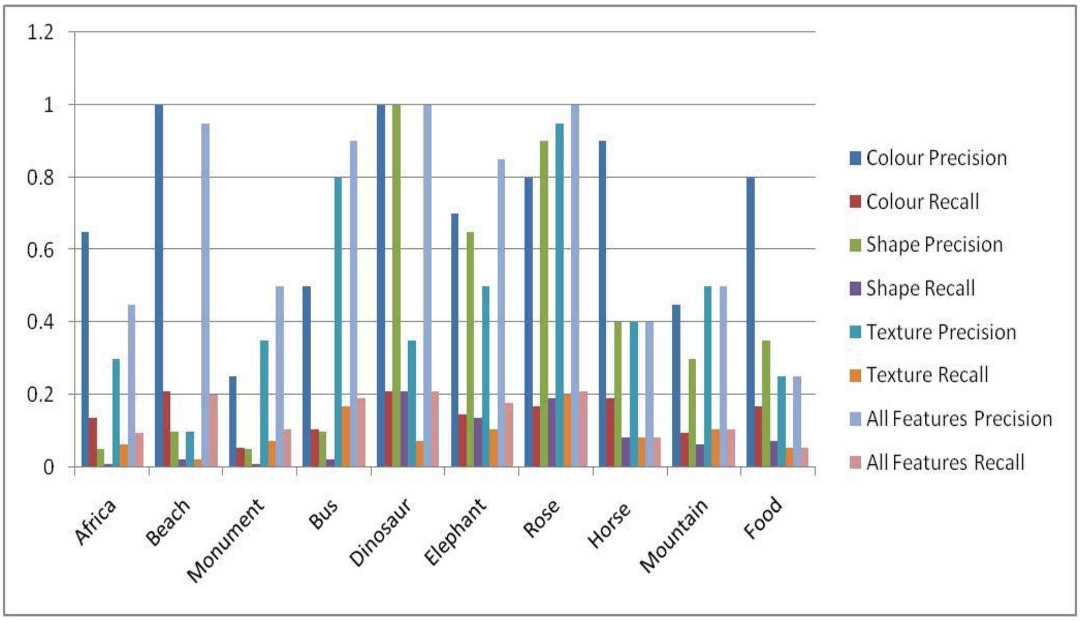

Fig. 4.3 (c) Retrieval using Cosine Distance

It can thus be seen that retrieval using Manhattan distance is the best giving a mean precision of 0.62 and mean recall of 0.13 for colour, shape and texture feature used separately. Combining the features, we get a mean precision of 0.78 and recall of 0.15. The above figures show the comparison of Manhattan, Euclidean and Cosine distance metrics.

From the Table 1, Table 2, Table 3 and Figure 8, it can be noted that the retrieval is based on colour feature using Manhattan distance alone provides a precision of 0.77 and recall of 0.16 whereas the precision and recall values for shape and texture using Manhattan distance are 0.52,0.11, 0.25 and 0.05 respectively. It is clearly shown that colour feature produces better retrieval results when compared to texture and shape when used on Wang dataset. It is to be noted that different features can be used for different datasets.

CHAPTER 5

IMAGE CLASSIFICATION

5. IMAGE CLASSIFICATION

5.1 SUPPORT VECTOR MACHINES

SVM expanded as support vector machine learning is a supervised learning algorithm that is used for classification and regression analysis. SVM works on the idea of separating data with a gap, otherwise known as 'margins'.

In this case, we have to classify images from different classes according to their classes.

Example: In the figure given below, we have two classes of images i.e., positive and negative. Let us consider that circles belong to the positive class and triangles belong to the negative class. We have a hyper plane and three labelled data points separating them.

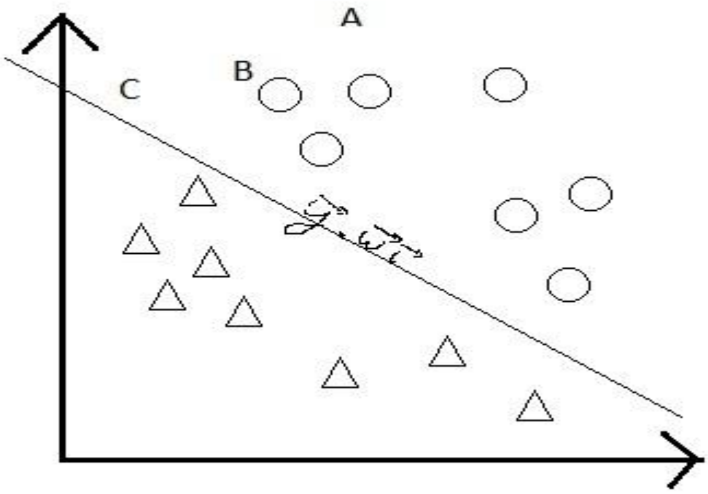

Fig. 5.1 Support vectors and separating hyperplane.

We can see that point A is farthest from the decision boundary. Therefore, one can predict the value of the label y_i =1 at point A. On the other hand, point ‘C’ lies closer to the decision boundary. The value of the label y_i will still remain to be 1. But if there is a small change in the decision boundary the value of the label y_i would be y_i = -1. This shows that we are less confident about the prediction at ‘C’ when compared to ‘A’. Point ‘B’ is located between these two points. Hence, we can conclude that one can be more confident about the predictions when the point lies furthest from the hyperplane. In this case, we have to find a decision boundary for our given training set which will help us make correct predictions.

5.2 IMPLEMENTING SVM WITH RBF KERNAL FUNCTION

$$K(x_i, x_j) = e^{-(\frac{\|x_i - x_j\|^2}{2\sigma^2})}$$

After solving the optimization we divide the hyperplane as

$$\omega^* = \sum_{i \epsilon SV} h_i y_i x_i$$

Offset of the hyperplane as

$$b^* = \frac{1}{|SV|} \sum_{i \epsilon SV} (y_i - \sum_{j=1}^{N} (h_i y_i x_j^T x_i))$$

where,

x = training vector

y = respective labels

h = Lagrangian coefficients

SV = set of support vectors

The space where the data resides is represented by $\mathcal{X}$.

Consider a function

$$\phi: \mathcal{X} \rightarrow \mathcal{F}$$

Taking the points from the input space $\mathcal{X}$ maps to $\mathcal{F}$, let's say that all points from $\mathcal{X}$are mapped to a new space $\mathcal{F}$, now solving the SVM in this plane instead of the previous plane, the working of the SVM will look same as before except that all points of x_i are represented as $\phi(x_i)$ and replacing $\langle \phi(x) | \phi(y) \rangle$, which is inner product in new space with $x^T y$, which is the natural inner product for Euclidean space and the resulting ω^* looks like the following

$$\omega^* = \sum_{i \epsilon SV} h_i y_i \phi(x_i)$$

$$\langle \omega^* | \phi(x) \rangle = \sum_{i \epsilon SV} h_i y_i \langle \phi(x_i) | \phi(x) \rangle$$

Similarly,

$$b^* = \frac{1}{|SV|} \sum_{i \epsilon SV} (y_i - \sum_{j=1}^{N} (h_i y_i \langle \phi(x_j) | \phi(x_i) \rangle))$$

Now, replacing all the dot product $K(x_i, x_j)$ with kernel function, changes ω^* and b^* as following equations

$$\langle \omega^* | \phi(x) \rangle = \sum_{i \epsilon SV} h_i y_i k(x_j, x_i)$$

$$b^* = \frac{1}{|SV|} \sum_{i \epsilon SV} (y_i - \sum_{j=1}^{N} (h_i y_i k(x_j, x_i)))$$

The above equations compute the inner product of *w* with *x* instead of computing *w* explicitly. SVM can be analysed theoretically from the concept of statistical learning theory, it is advantageous to problems with limited training samples in the higher dimensional space which results in good performance for retrieval of similar images.

The approach used here is “one-vs.-one” where $\frac{n!}{(n-k)!k!}\frac{n!}{(n-k)!k!}$ binary classifiers have to be trained for a k - way problem. It differentiates the samples of a pair of classes at a time. When a query image is given as input, a voting scheme is applied to all $\frac{n!}{(n-k)!k!}$ $\frac{n!}{(n-k)!k!}$ classifiers. The predicted output by the classifier is the class that gets the highest number of ‘+1’ predictions.

The classifier results are analysed below. The classifier results are also tested using the precision and recall metric.

The images are divided into as training and testing. 95 images from each class are taken for training and 5 are taken for testing. Hold out validation technique is implemented to test the results. It prevents the over fitting problem.

This is called hold out validation. It uses only a single validation training pair. The confusion matrix obtained from testing with colour, shape and texture separately are shown in fig 7, fig 8, fig 9 respectively. Fig 10 shows the confusion matrix obtained by combining the colour, shape and texture features.

The precision, recall and accuracy are given by

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \tag{16}$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \tag{17}$$

$$Accuracy = \frac{true\ positive + true\ negative}{total\ no\ of\ images} \tag{18}$$

The sample output using colour, shape and texture features separately is shown in Fig 12, Fig 13, Fig 14 respectively. Fig 15 shows the output combining all colour, shape and texture feature.

Fig. 5.2 (a) Retrieved result using colour

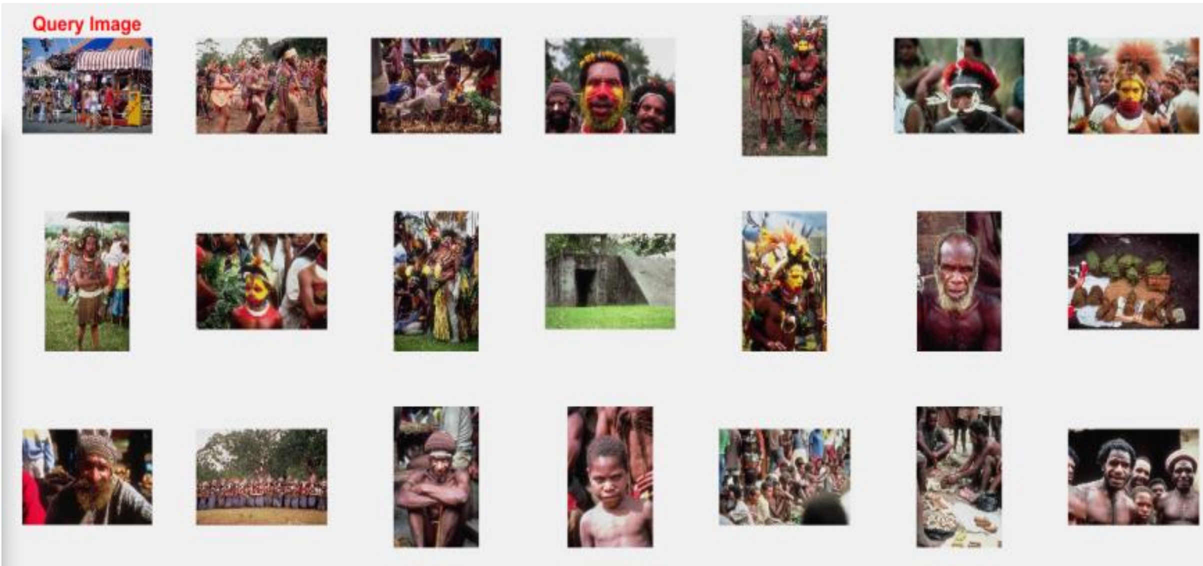

Fig. 5.2 (b) Retrieved result using shape

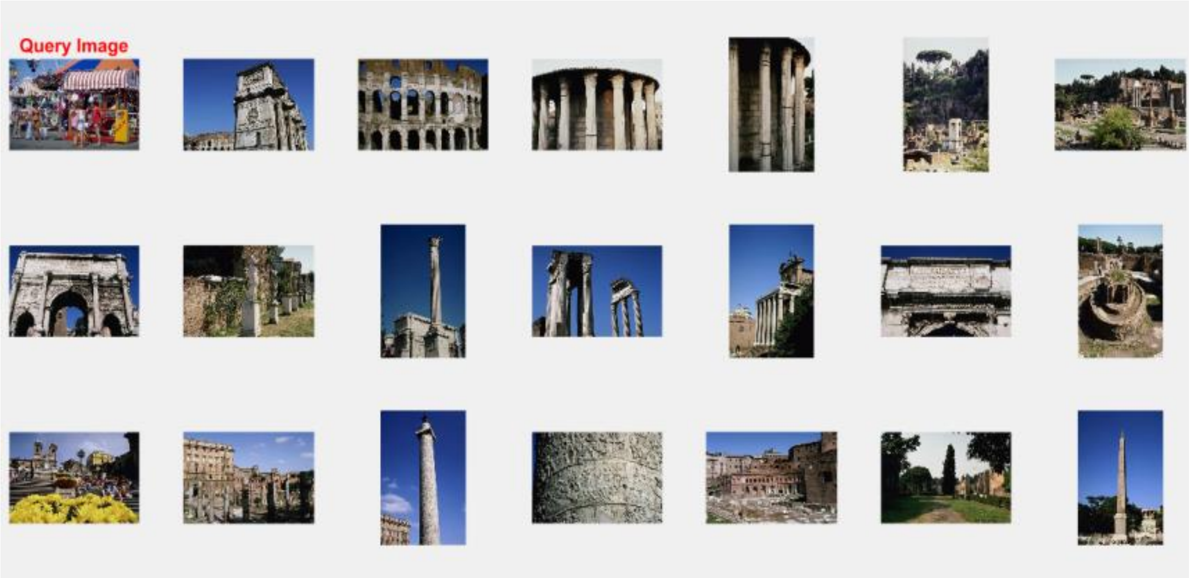

Fig. 5.2 (c) Retrieved result using texture

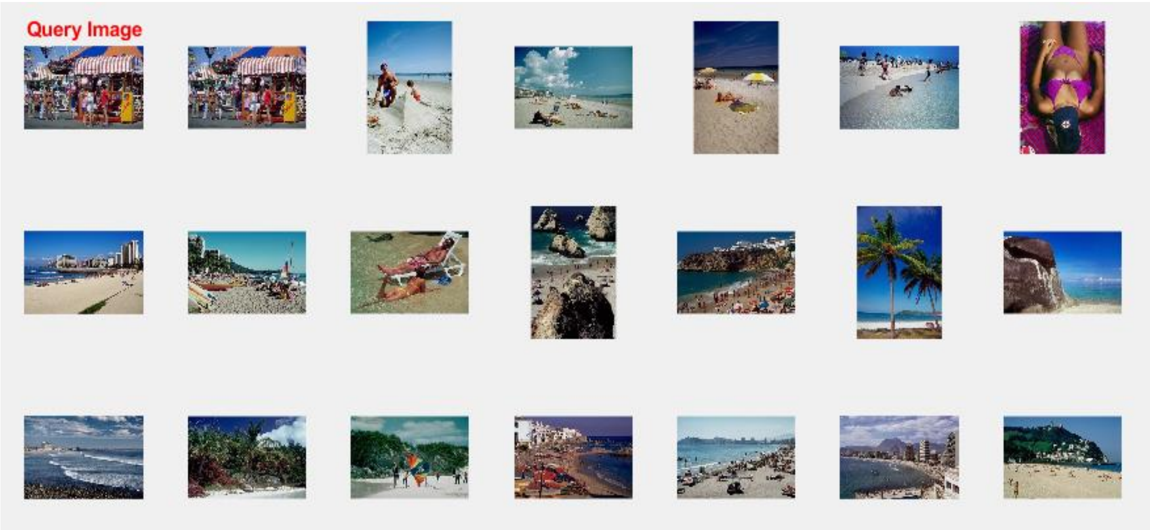

Fig. 5.2 (d) Retrieved result using combination of colour, shape and texture

Table 12 represents the values of the performance measure based only on colour feature. Table 2 represents the values based only on texture. Table 3 represents the values based only on shape. Table 4 shows values based on combination of colour, texture and shape feature.

5.3 RETRIEVAL RESULT USING CLASSIFICATION (WANG DATASET)

Table 5 shows the average values of the precision and recall on Wang dataset. It shows the retrieval performance when colour, texture and shape features are used in isolation. The average value is taken for each class. Table 6 shows values based on combination of colour, texture and shape feature.

Table 5 Classification using colour

Class	Precision	Recall
Africa	0.76	0.79
Beach	0.57	0.71
Monument	0.52	0.63
Bus	0.93	0.91
Dinosaur	1.00	0.96
Elephant	0.92	0.92
Rose	0.84	0.98
Horse	1.00	0.98
Mountain	1.00	1.00
Food	1.00	1.00
Mean	0.85	0.89
Accuracy = 74.68 %		

Table 6 Classification using shape

Class	Precision	Recall
Africa	0.27	0.38
Beach	0.47	0.68
Monument	0.31	0.29
Bus	0.82	0.82
Dinosaur	0.98	1.00
Elephant	0.91	0.88
Rose	0.93	0.96
Horse	0.98	0.97
Mountain	0.92	0.85
Food	1.00	1.00
Mean	0.76	0.78
Accuracy = 57.45 %		

Table 7 Classification using texture

Class	Precision	Recall
Africa	0.50	0.62
Beach	0.43	0.59
Monument	0.71	0.67
Bus	0.96	1.00
Dinosaur	1.00	1.00
Elephant	0.70	0.74
Rose	1.00	1.00
Horse	0.97	0.94
Mountain	0.68	0.72
Food	1.00	1.00
Mean	0.80	0.83
Accuracy = 66.17 %		

Table 8 Classification using combination of colour, shape and texture

Class	Precision	Recall
Africa	0.74	0.83
Beach	0.67	0.78
Monument	0.77	0.87
Bus	1.00	1.00
Dinosaur	1.00	0.98
Elephant	0.93	1.00
Rose	1.00	1.00
Horse	0.98	1.00
Mountain	1.00	1.00
Food	1.00	1.00
Mean	0.91	0.95
Accuracy = 85.32 %		

Important results can be drawn from these tables. It is clearly seen that the performance is the best when shape, texture and colour features are used in tandem with one another. It is seen that the mean precision of 0.90 and recall of 0.94 is obtained when all features are combined which is much higher than the mean precision and recall values when the features are used separately. A confusion matrix for classification on the Wang dataset is shown below.

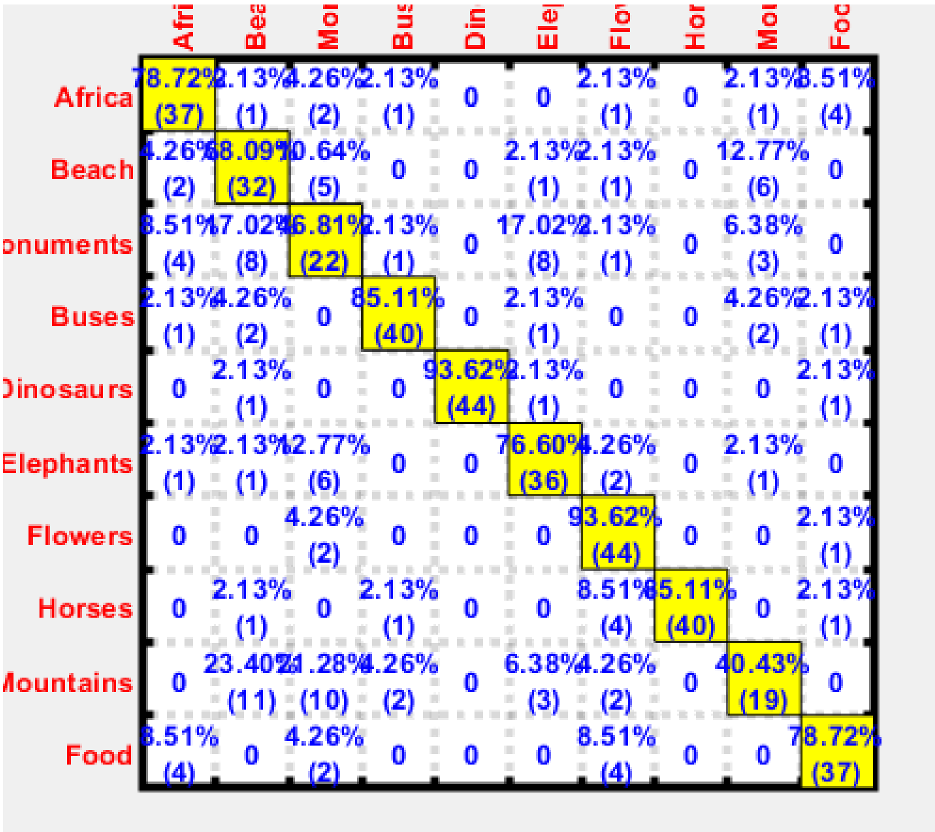

Fig. 5.3 (a) Confusion Matrix using only colour feature

The above confusion matrix shows classification using colour feature separately. It can be noted that in the Africa class, only 37 images are classified correctly. The class where the least number of images classified correctly is mountain. This could be due to the various colours in the class. It can easily get mixed with the beach class.

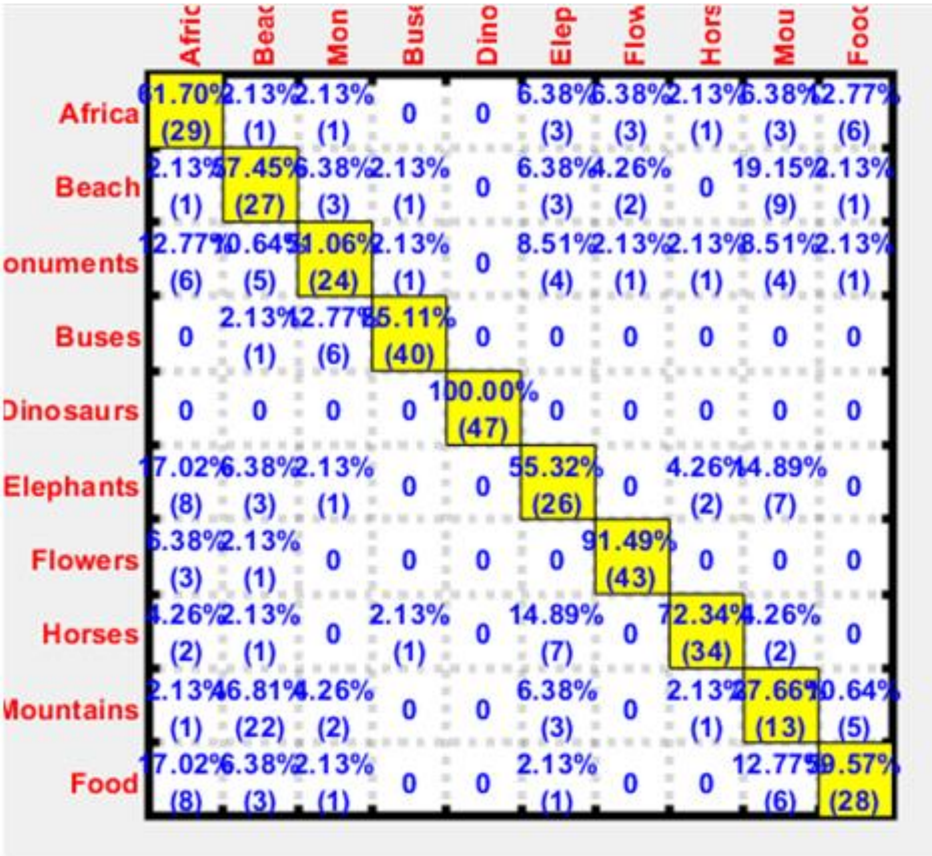

Fig. 5.3 (b) Confusion Matrix using only texture feature

The above confusion matrix shows classification using the texture feature separately. It can be noted that all images in the dinosaur class is classified correctly. The true positive rate is the least in mountain class.

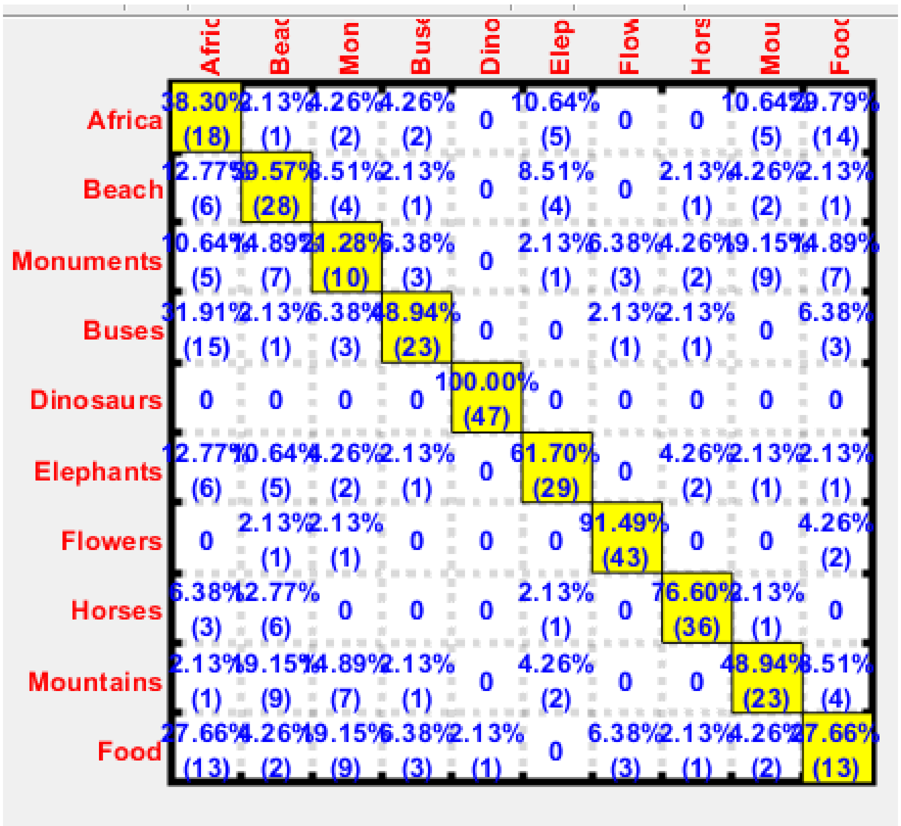

Fig. 5.3 (c) Confusion Matrix using only shape feature

The above confusion matrix shows classification using the shape feature separately. Coiflet wavelet and HOG are used to form the feature vector. It can be noted that the accuracy produced by the shape feature is the least when compared to colour and texture.

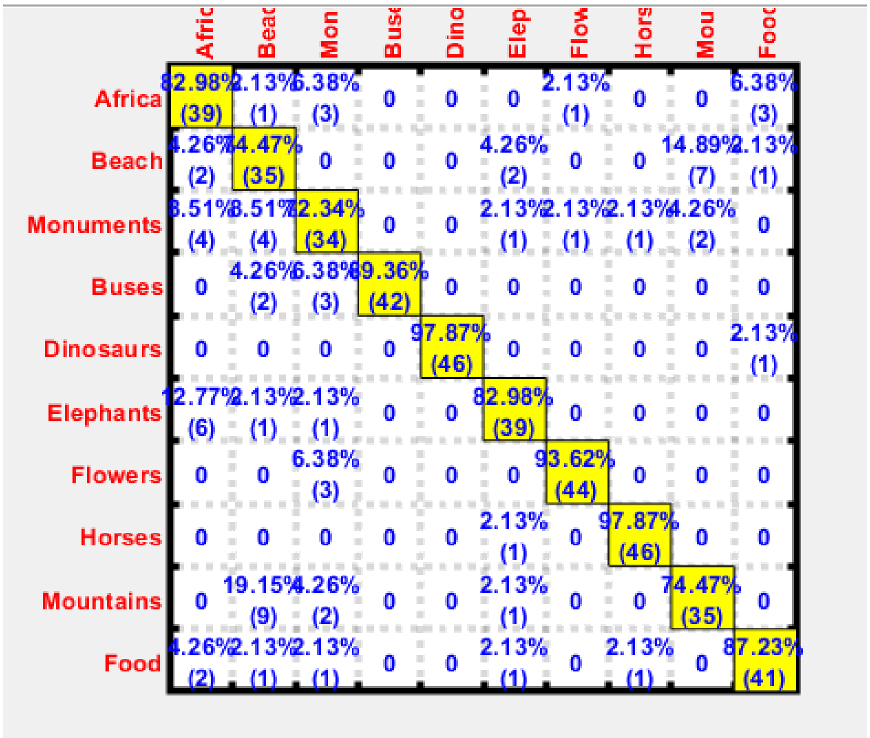

Fig. 5.3 (d) Confusion Matrix using color, shape and texture features

The above confusion matrix shows classification using the shape, texture and colour feature in tandem with one another. The classification accuracy is found to be 85.32%.

CHAPTER 6

CONCLUSION

6. CONCLUSION

The goal of this project is to retrieve images from a database with reliable accuracy by using multiple techniques in tandem with one another. The proposed work introduces an integrated approach to CBIR which helps retrieve similar images from a database.

A comparative study on retrieval results was performed and the results using colour, shape and texture features in tandem with one another was performed and the proposed system combining all the features was found to be superior with a precision of 0.91 and recall of 0.95.

A comparative study on distance metrics for retrieval was also performed and it was found that Manhattan distance metric achieved best results with a precision of 0.77 and recall of 0.16 compared to Euclidean and Cosine distance metric.

Further, SVM was used for classification as a measure to achieve higher retrieval efficacy and it was found that classification combining all the features gave the best accuracy of 85.32% which superseded previous research on CBIR.

Future refinement of the work will involve research using bag of words as a feature for larger datasets. Integration of the features can be made according to the application. For example, detection of glaucoma can be made using the shape feature alone. One can also perform context based image retrieval that concentrates on the context rather than the content of an image.

CHAPTER 7

REFERENCES

7. REFERENCES

1. Katta Sugamya, Suresh Pabboju, A. Vinaya Babu, A CBIR classification using Support vector machines, International Conference on Advances in Human Machine Interaction, March 3-5, Bangalore, India, (2016).

2. Syed Hamad Shirazi, Noor ul Amin Khan, Arif Iqbal Umar, Muhammad Imran Razzak, Saeeda Naz, Bandar AlHaqbani, Content-Based Image Retrieval Using Texture Color Shape and Region, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 7, No. 1, (2016).

3. Qasim Iqbal, J. K. Aggarwal, CIRES: A system for Content-Based Retrieval in digital image libraries, Seventh International conference on Control, Automation, Robotics, and Vision (ICARCV), December, Singapore, (2002).

4. Anil Balaji Gonde, R.P. Maheshwari, R. Balasubramanian, Modified curvelet transform with vocabulary tree for content based image retrieval, Digital Signal Processing, (2013).

5. B. S. Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan, Akio Yamada, Color and Texture Descriptors, IEEE Transactions on circuits and systems for video technology, Vol. 11, No. 6, (2001).

6. Aditi Giri, Yogesh Kumar Meena, Content Based Image Retrieval using integration of Color and Texture features, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 4, (2014).

7. Dharmendra Pandey, Shivpratap kushwah, A Review on CBIR with its Advantages and Disadvantages For Low-level Features, International Journal of Computer Sciences and Engineering, Vol. 4, Issue 7, (2016).

8. Adrian Rosebrock, Tim Oates, Jesus Caban, Ecosembles: A Rapidly Deployable Image Classification System Using Feature-Views, 12th International Conference on Machine Learning and Applications, (2013).

9. K. Satya Sai Prakash,RMD. Sundaram, Combining Novel features for Content Based Image Retrieval, EURASIP Conference focused on Speech and Image Processing, (2007).

10. James Z. Wang, Jia Li, Gio Wiederhold, SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries, IEEE Transactions on Pattern analysis and Machine Intelligence", Vol. 23, No. 9, (2001).

11. Aravind G, Andan H M, Tripty Singh, Glence Joseph, Development of Biometric Security System using CBIR and EER, IEEE International Conference on Communication and Signal Processing (ICCSP), (2015).

12. Manimala Singha, K. Hemachandran, Content Based Image Retrieval using Color and Texture, Signal & Image Processing : An International Journal (SIPIJ), Vol. 3, No. 1, (2012).

13. G. Raghupathi, R.S. Anand, and M.L Dewal, M. L., Color and Texture Features for content Based image retrieval, Second International conference on multimedia and content based image retrieval, (2010).

14. S. Arivazhagan, L. Ganesan, S. Padam Priyal, Texture classification using Gabor wavelets based rotation invariant features", Pattern Recognition Letters 27, (2006).

15. T.Bagyammal, Latha Parameswaran, Context Based Image Retrieval using Image Features, International Journal of Advanced Information in Engineering Technology (IJAIET), Vol. 9, No. 9, (2015).

16. http://wang.ist.psu.edu/

International Conference on ISMAC in Computational Vision and Bio-Engineering (ISMAC - CVB 2018)

ISMAC-CVB 2018

ACCEPTANCE LETTER

APRIL 12, 2018

To
Karthika. R , Akshaya. B , Niranjana Sathish. A , Shobika. K ,
Sruthi Sri. S , Latha Parameswaran

Acknowledgement Number/Paper ID: ISMAC/CVB:1226

Subject: Acceptance Letter – 2018 International Conference on ISMAC in Computational Vision and Bio-Engineering (ISMAC - CVB 2018) – **Reg.**

Dear Author,

This is the notification to inform you that your Oral presentation proposal entitled " CONTENT BASED IMAGE RETRIEVAL USING HYBRID FEATURE EXTRACTION TECHNIQUES " **submitted to the International Conference on ISMAC in Computational Vision and Bio-Engineering (ISMAC -CVB 2018) organized by SCAD Institute of Technology, Palladam, Tamilnadu, India. on 16-17 May, 2018 has been accepted as a result of blind reviews.**

After registration, your paper will be published in the following springer series

Springer - Lecture Notes in Computational Vision and Biomechanics.

On behalf of the organization committee I would like to congratulate you.

Yours sincerely,

M. [signature]

Dr. M. Durai Pandian,
Vice Principal,

SCAD Institute of Technology, Palladam, Tamilnadu, India. svsduraipandian@gmail.com

TECHNICAL SPONSORS

SCAD INSTITUTE OF TECHNOLOGY
Palladam, COIMBATORE
Approved by AICTE, Affiliated to Anna University

Springer