M. Tech (SPDD)

HAND POSTURE RECOGNITION USING SKIN, TEXTURE AND SALIENCY MAP

DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
SIGNAL PROCESSING AND DIGITAL DESIGN

Submitted by:

HARIKESH

2K12/SPD/07

Under the supervision of

Sh. D. K. VISHWAKARMA
(Assistant Professor, ECE)

HARIKESH

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

2014

2014

CERTIFICATE

DELHI TECHNOLOGICAL UNIVERSITY,
(Formerly Delhi College of Engineering)
BAWANA ROAD, DELHI – 110042
Department of Electronics & Communication Engineering

This is to certify that the report entitled **"Hand Posture Recognition using Skin, Texture and Saliency map"** submitted by Harikesh, Roll. No. *2K12/SPD/07*, in partial fulfillment for the award of degree of Master of Technology in Signal Processing & Digital Design at **Delhi Technological University, Delhi**, is a bonafide record of student's own work carried out by him under my supervision and guidance in the academic session 2013-14. The matter embodied in dissertation has not been submitted for the award of any other degree or certificate in this or any other university or institute.

Mr. D.K.Vishwakarma (Project Supervisor)
Assistant Professor
Department of Electronics and communication Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
New Delhi -11004

ACKNOWLEDGEMENT

I express my sincere gratitude to my thesis supervisor **Sh. D. K. Vishwakarma** for giving his valuable guidances, suggesations, encouragement, consistance moral support during the course of the investigation.His enormous knowledge and investigation always helped me unconditionally to solve the various problems. I would like to thank him for introducing me with the problem and providing valuable advice throughout the course of work. I truly admire his depth of knowledge and strong dedication to students and research that has made him one of the most successful professors ever. His mastery of any topic is amazing, but yet he is such a humble and down to earth person. I am glad that I was given the opportunity to work with him. He surely brings out the best in his students.

I am greatly thankful to **Prof. Rajiv Kapoor, Professor and Head,** Department of Electronics & Communication Engineering, the entire faculty and staff of Electronics & Communication Engineering Department for their continuous support, encouragement and inspiration in the execution of this "**thesis**" work.

I would like to thank my parents who bestowed upon me their grace and were the source of my inspiration and encouragement. It would be my fault if I didn't mention my friends here. I would like to thank them for their support and ideas whenever I stuck. Its because of them my work was completed without any hurdles.

Harikesh

ABSTRACT

Hand Gesture Recognition (HGR) systems are paying more attention in the present scenario due to a wide range of application in real life. These systems are based on the processing of an incoming digital image. In this thesis, the first task is to separate the hand from rest of the image. This can be achieved in several ways and depends on whether the image includes only a hand against a background or the entire person. All the techniques proposed have some limitations. Their performance is affected by the illumination or lighting conditions and complex backgrounds of the hand images.

To overcome the above mentioned limitations, a new technique is proposed for Hand Posture Recognition using Skin, Texture and Saliency map which involve classification of hand gesture. It efficiently makes the use of different color spaces, skin detection, saliency map and texture of an image. To recognize the hand gesture classes the features are extracted using the proposed methodology. These features are used for the classification purpose. Finally the application of multi-class SVM classifier is employed to recognize the hand gesture class. The experiments show that the proposed model has stable performance for a wide range of images.

Keywords: Hand Posture Recognition, Texture, Saliency Map, SVM and Gabor Filter

CONTENTS

List of Figures

List of Abbreviations

HGR: Hand Gesture Recognition

HCI : Human Computer Interface

FEMD: Finger-Earth Mover's Distance

R: Red

G: Green

B: Blue

HSI: Hue-saturation Intensity

STFT : Short-time Fourier transform

SVM: Support Vector Machine

HMM: Hidden Markov Model

NUS: National University of Singapore

List of Tables

CHAPTER 1

INTRODUCTION

1.1 HUMAN COMPUTER INTERACTION SYSTEM

HCI has the very vast range of application of the computer systems in everyday of life. So in the present scenario, human computer interaction is very important Zhang et al [17]. In the early days the interaction between them has happened through some devices, they are called as interface devices, for instance mouse, keyboard etc. Mitra and Tinku et al. [18]. A new developed device will try to make the computer system more reliable and efficient and now one can interact with the computer in a simple way. Complex communication with the computer systems is possible to a great extent because of emerging new techniques, which has made the computer systems even more intelligent. It is possible due to the constant efforts made by researchers and professionals to create efficient human computer interfaces. Now a days human needs are growing wildly and continues to grow. The computer programming will ease the information processing between computer system and human being. At present, the products with latest technologies have been emerging which are easing the human computer interaction. For example, it has helped in facilitating virtual reality/computational video, video confessing, surveillance and robots for physically disabled persons Sigal and Helman et al [15]. In the early days, the computer programmers focus mainly on speed and avoid complex algorithms. To create a user friendly environment motivated them to focus on new technologies with larger speed and more effective control. So the ultimate goal is to train the system like human to make computers understand the human language, human gestures, understand speech, facial expressions so that human computer interface (HCI) becomes efficient. Gestures are some expression to exchange meaningful information. Human can make numerous gestures at a particular instant. As the gestures are interpreted through vision, so it becomes a new area of interest for computer vision research. Our project will focus on processing the different human gestures to create an effective and reliable HCI. The encoding of the different gestures into an objective language well understood by the computers has to be done and it demands a complex programming algorithm.

1.2 GESTURES

Gestures are some expressions to convey some information. A human can do gesture can be large in number and vary in terms of their meaning and applications, and so every gesture contains some information Pramod and Shen et al [2] [9]. For the communication takes place, it is necessary that at both end, i.e. Transmitter and receiver should have the same meaning of the gesture. So finally we may define a gesture as some movement of the body or parts such as head or hand conveying some meaningful information.

The Gestures are mainly of three types as explained below Mitra and Tinku et al [18]:

1) **Hand and arm gestures:** Hand postures, sign languages, and entertainment applications

2) **Head and face gestures:** the examples of expressing this type of gesture are:

   a) Moving of the head
   b) Eye gazes
   c) The eyebrow expressions
   d) Opening the mouth to speak
   e) Flaring the nostrils and
   f) Expressions of happiness, sadness, fear, etc.

2) **Body gestures:** gesture in which the whole body movement is observed. For example movements of athletes in sports activities;

The hand gestures may be Static and Dynamic type Pramod et al [2].
These can be described as follows:

(a) Static Gesture: static gestures are observed at only that particular instant of time i.e. static in their meaning.

Static Gestures are characterized by Pramod et al [2]:

- Shape
- Texture
- Skin colour
- Finger flex‘s angle
- Orientation

(b) Dynamic Gesture: A dynamic gestures are the gesture defined for a small interval of time X. Shen et al [9]. For instance, moving head to accept or reject is a fine example of dynamic gesture

Dynamic Gestures are characterized by Pramod et al [2]:

- Shape
- Texture
- Skin color
- Finger‘s flex angle
- Hand trajectory
- Scale

Static and dynamic gestures can be used in a large number of applications such as in activity recognition and sign languages, etc. A gesture can be defined both in time and in space. A particular gesture may affect the gesture which follows or precedes the current gesture, i.e. gestures may be dependent. There are different type of human races exists, so the meanings of gestures for different community may be different from the other. Thus, it is necessary to interpret all types of gestures for a short time so as to decode meaning. This process is known as gesture recognition. Hence we can put it in this way; the gesture recognition is the way of detecting, recognizing and then interpreting a continuous sequential gesture.

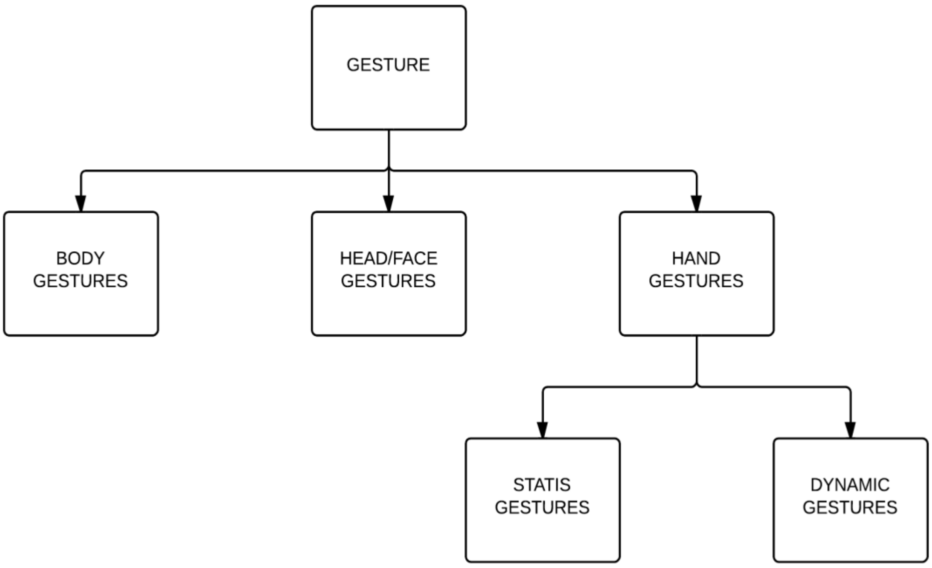


Figure 1.1 Classifications of Gestures

1.3 THESIS OVERVIEW

Chapter 2: Presents the details of literature survey on hand gesture recognition systems.

Chapter 3: Describes the various colour spaces used for HGR, skin detection methods, saliency map and texture analysis.

Chapter 4: Explain the classifiers: Analysis of Linear SVM and Multi-class SVM.

Chapter 5: Proposed methodology is described that includes the method for feature extraction using skin, saliency and texture. The classification of hand gesture is done using multi-class SVM classifier.

Chapter 6: Result and discussion along with the conclusion of the work is presented.

CHAPTER 2

LITERATURE SURVEY

Hand gesture recognition systems are paying more attention in the present scenario due to a wide range of application in real life. Vision-based methods are popularly used in the hand recognition system. In vision-based methods, the analysis is natural and useful for real time applications. A human can easily identify a hand gesture, however, for a computer to recognize hand gesture first the hand should be detected in the acquired image and recognition of that hand in the image be done in a similar way as humans do. This is a more challenging approach to implement because of the limitations of such a natural system.

Various methods for hand gesture recognition already exist. Pramod and Prahlad et al [2] proposed a learning based method for hand detection against complex backgrounds. Dynamic hand gesture recognition by X. Shen et al [9] introduces a technique for hand motions represented by the use of the motion divergence fields. These are to be normalized to grayscale images. In this method salient regions are detected based on the motion divergence maps. Shikha Gupta et al. [7] describes the use of Gabor filter for hand gesture recognition. The Gabor filter finds the useful information and thus helps in reducing the redundant data. A novel method by S. Padam et al [6] uses the geometry based normalizations and Krawtchouk moments of the hand for extracting the regions of the hand and the forearm. The Krawtchouk features are invariant to viewpoint changes of hand and achieve a good recognition when the training samples are small. Wein and Lee et al [14] constructed a self- eigen hand recognizer with genetic algorithms (GA) for selecting discriminant eigenvector subsets for classification. This approach maximizes the differences in hand images for various hand shapes, it also minimizes variations in lighting and pose for the same hand shape. Ren et al. [4] proposed the use of Kinect sensor in the hand recognition system. A hand gesture recognition system using Kinect sensor is capable of handling the hand shapes obtained from the Kinect sensor. Here Finger-Earth Mover's Distance (FEMD) is calculated which gives the measure the dissimilarity between hand shapes. It measures the similarity in the finger parts, not the whole hand so it can easily distinguish the slight differences in the hand gestures. Chuang and Y.Chuang et al. [1] constructed the integration of image saliency and skin information. Skin model is a simple and efficient strategy by which to locate skin regions within images, it is easily affected by complex backgrounds, e.g. skin-like background regions and various lighting conditions. Since the method considers only the first-order average colour, it could be insufficient to analyse the complicated situations that frequently occur in images with natural surroundings. Valenti et al [30] adopt an isocentric feature approach to representing image saliency in a global manner. Since the isocentric feature is seriously influenced by the type of background in the image and the noise level, the results of might contain a substantial amount of irrelevant information. Zhang et al [31] propose a novel method for detecting image saliency by exploring spectral components in an image. This method is very fast, but because it is based on global considerations, detailed information regarding salient objects could be overlooked. Schauerte and Stiefelhagen et al. [32] [33] propose a spectral-based image saliency detection for eye fixation prediction. However the method should define an appropriate colour/feature space. These methods have limitations against the variation against the environment and illumination conditions. Here we try to answer these problems.

We propose a method to detect the salient pixels in images using skin saliency and texture features. In our method, an isophote-based operator is employed to capture the potential structure and global saliency information related to each pixel. The potential structure is used to determine the centre–surround contrast that is then combined with global saliency to determine the final saliency information. We calculate the texture information of the image by means of Gabor filter. This improves the detection of hand in an image against natural backgrounds.

CHAPTER 3

SKIN COLOUR DETECTION, SALIENCY MAP AND TEXTURE ANALYSIS

3.1 SKIN COLOUR DETECTION

3.1.1 INTRODUCTION

Skin detection is used to find skin pixels in an image Y. Ban et al.[3]. This method is used as a pre-processing step to find the regions where human faces and limbs are present in images. So skin detection means to find out the pixels and regions in an image or video that contain skin Ting-hui et al [12]. There exists many approaches for efficiently skin detection. First we transform the image into an appropriate color space in which effect of illumination is minimum and then we apply a skin classifier to decide the input pixel is a skin or non-skin pixels Y. Ban et al.[3]. As human skin occupy a limited intensity range in the color space so by the skin classifier one can define a decision boundary of the skin color, based on a training database of skin-colored pixels.

Skin color and texture can be used as an important feature that can be used for various human computer application system. Skin color and texture can vary due to race, age, health, etc. Pramod et al [2]. In images and videos, skin colour can be used as an indicator of the existence of humans in video or image. Therefore, this can be a very important tool for an application so extensive research has been focused to efficiently detect the skin in images.

Detecting skin pixels in an image is a simple task and can be done very efficiently. So this can be used as a feature that in many image and video applications. In the applications, where the background of image is uniform or does not contain skin-colored regions, skin-color detection becomes very easy and so we can easily find hands or other skin regions in images Y. Ban et al. [3]. So skin-colored region detection seems to be an easy task, but there may be many constraints such as background, illumination, etc. which makes this quite challenging H. Han et al [11]. The skin region detection is dependent on the illumination conditions, i.e. in which condition they are taken. Humans can easily identify the color of objects in different illumination conditions. This is known as color constancy. Color constancy is a very important for the system to efficiently find the skin regions Jose et al [19]. So here the challenges are how to represent the color information in such a way which is least sensitive to variation in the illumination conditions. The color space selected plays an important role for the performance of any skin detector and its sensitivity to variations in illumination conditions H. Han et al [11]. There may be objects which have similar color of that human skin e.g. tree, wood, clothes, hair, sand, etc. This may lead to false skin region detection under the uncontrolled environment or complex environment.

3.1.2 A SKIN DETECTION METHODS

Given a skin detector, to identify skin region in an image involves H. Han et al [11]:

1. First transform the image into the suitable color space.
2. Decide that whether the input pixel is a skin or non-skin using a suitable skin classifier.
3. Post processing may be needed using morphological operation on the detected skin regions.

In any color space, the skin color occupies a small portion of such a space, which is compact region in the color space. This region is known as the skin color cluster. A pixel in an image is classified as a skin pixel or a non-skin in a given color space Ting-hui Huang et al. [12]. In the skin detection, skin the pixels which are correctly classified as skin by the classifier are true positives. True negatives are non-skin pixels which are wrongly classified by the classifier. No classifier can classify all pixels correctly so there are some errors in the classification. It may classify a non-skin pixel as skin or a skin pixel as a non-skin. The former type of error is known as false positives or false detections while the latter is false negatives. A classifier must have small error either false positive or false negative. So there is a trade-off between false positives and false negatives errors in a classifier. The decision boundary for a skin classification decides the error rate Jose et al [19]. So larger the class boundary, there is less false negatives rates and more the false positives rates. The smaller the class boundary, there is more the false negatives rates and less the false positives rates. So the color space choice plays important role in skin detection. Such color space is chosen in which the skin occupy the most compact region. So the color space chosen is extremely important and this directly affects the performance of the classifier.

3.1.3 COLOUR SPACES USED FOR SKIN DETECTION

The human skin color occupies a small portion of a color space. Therefore, the human skin color is not randomly distubuted in a given color space and occupies a small region in the color space R. Khan et al [10]. Different color space may have different range of skin cluster. Variety of color space exists which can find application in skin detection. The aim is to find such a color space in which variation of the skin color to illumination conditions is minimal H. Han et al [11].

The different color spaces and their properties are explained as follows:

3.1.3.1 RGB COLOR SPACE

RGB color space is commonly used color space in digital images. In this color space all information is a combination of three primary colors given by Ting-hui Huang et al. [12]:

(i) Red

(ii) Green and

(iii) Blue

In RGB Color space, there are 3D cube representing the R, G and B on three perpendicular axes. The RGB space is very popular for its simplicity. But this is not a uniform color space. The RGB color space all information such as color, luminance and chrominance, are represented in the combination of three color components. As all the information is in term of R, G, and B components so all information is highly correlated. The luminance component of a RGB pixel comprises a linear combination of the R, G, and B values. Therefore, if we change the luminance of a image then it will affects all component of pixel i.e. the value of R, G, and B components changes due to change in illumination. This leads to a much stretched skin colour cluster in the RGB color cube. So, in the RGB color space the skin clusters varies for different races as this is located at different locations.

NORMALIZED RGB

Normalized RGB can be derived from the RGB values as follows:

$$r = \frac{R}{R+G+B}, \qquad g = \frac{G}{R+G+B}, \qquad b = \frac{B}{R+G+B}$$

Also

$$r + g + b = 1$$

The advantage of this color space is that if the sum of r, g and b components is known then third component may not be required, so space dimensionality is reduced. The remaining components are known as “pure colors”, for the dependence of r and g on the brightness of the source RGB color is diminished by the normalization. The normalized RGB is invariant to change in orientation of space relative to the light source.

3.1.3.2 HUE SATURATION INTENSITY MODEL

Hue-saturation color spaces describe color with intuitive values, based on the idea of tint, saturation and tone. Hue component specify the color information (red, green, blue and yellow etc.) in an image. Saturation gives the colorfulness of an image in proportion to brightness. The “intensity” or “lightness” is resembled to the luminance H. Han et al [11]. The color space components properties and distinction between luminance and chrominance component made HSI color spaces more popular in skin detection. Basic properties of Hue are given as: hue component is invariant to light intensities and orientation of surface relative to the light source. The disadvantages of this color space are hue discontinuities and the brightness component (lightness, value), which are contradictory with the properties of color vision. The polar coordinate system are popularly used in Hue-Saturation spaces. This results in cyclic nature of the color space which makes it unsuitable for skin color models as this type of color space there is a need of compact cluster of skin colors for efficient performance Ting-hui Huang et al. [12].

3.1.3.3 YCbCr MODEL

YCbCr is a form of nonlinear RGB color space Aibinu et al [8]. This is commonly used for image processing application. Y represents the luminance component. It can be calculated from RGB components as a weighted sum of the RGB values. Cr component gives red difference and Cb gives blue difference. The Cr can be found by subtracting Y from RGB red and Cb can be found by subtracting Y from RGB blue components. This can be represented as X. Shen et al [9]

(i) $Y = 0.299R + 0.587G + 0.114B$

(ii) $Cr = R - Y$

(iii) $Cb = B - Y$

This color space find the application in skin based application because of its simplicity and this also separates the luminance and chrominance components of the image so that effect of illumination on color is reduced Ting-hui Huang et al.[12].

3.1.3.4 PERCEPTUALLY UNIFORM COLOUR SYSTEM

The skin color is a perceptual phenomenon what human eye sees. So, color space representation corresponds to the sensitivity of human vision for efficient performance of the system. CIELAB and CIELUV are known as uniform colour spaces because a change of the same amount in a color value produces change of about the same visual importance Ting-hui Huang et al. [12]. The commonly used RGB color space is not uniform so a non-linear transformation is needed convert to CIELAB and CIELUV. The components in Lab are given as:

(i) The 'L' component represent the lightness or seprate out luminance,
(ii) 'a' gives the levels from green to red and
(iii) 'b' gives the levels from blue to yellow.

3.1.4 EXPLICITLY DEFINED SKIN REGION

The skin is defined by the boundaries of skin cluster in any color space e.g. RGB color space skin is classified as if:

(a) R > 95 ; G > 40 and B > 20 and
(b) $Max\{R, G, B\} - Min\{R, G, B\} > 15$
(c) $|R - G| > 15$; $R > G \text{ and } R > B$

It is very simple method for skin detection. For reliable recognition rates with this method, a suitable color space is needed.

3.1.5 SKIN DETECTION APPLICATIONS

Skin color finds several applications in human computer interface. In the situation where background is uniform, skin detection is an efficient method to locate hand and face in the images. The skin color processing is employed because it makes the processing much faster. So this is used as a preliminary process for hand detection and face detection techniques. Skin detection also find the use in locating body limbs, such as hands, as a part of hand segmentation and tracking systems Jose et al [19]. Another application can be Hand Tracking for

a) Gesture recognition

b) Robotic controls

c) Other human computer interaction.

3.2 IMAGE SALIENCY

The sensors generate many different signals and it is not possible to process all this incoming information all the same time. So it is very important for the nervous system to select on which part of the available information for further processing, and which are to be ignored. Furthe, the selected information signal must be given priority so that the most relevant signal to be processed first and the less important are given less priority. So, in this way, a sequential treatment of different parts of the visual scene is given. This process of prioritization is known as selective attention.So according to findings from neurobiology and psychophysics, image saliency can be defined as regions that can be easily differentiated from their surroundings, these differentiators being color, orientation, and intensity Ren et al. [4].

The key issues in detecting image saliency include the following:

(1) How to determine the position and size for both the center regions and neighboring regions,

(2) How to compute differences between center regions and adjacent regions for the three feature channels (color, intensity, and orientation).

3.2.1 ISOPHOTE-BASED CENTER AND SURROUND REGION DETECTION

Isophotes are contour lines that connect points of equal intensity. An image may be described by isophotes because:

(1) Isophotes do not intersect each other, and

(2) The shape of isophote in an image is independent of variations in contrast and brightness.

Schauete et al. [27] make use of the features of isophotes for the task of object-based scene analysis. Moreover, it has been observed that, for highly curved isophotes, their osculating circle stend to concentrate on small regions around an object's corners, while for minimally curved isophotes, their osculating circles tend to concentrate on large regions around an object's centre. These pattern simply that the curvature of the isophotes could indicate an object's scale. Because of these properties, we use an isophote-based operator to detect a pixel's centre and surrounding regions.

Given an input $L(x, y)$, where x, y are the Cartesian coordinates in the image plane, an isophote is defined as $L(x, y)=L_0$ (where L_0 stands for a specific value of the luminance, and the range is (0–255). The curvature c of L_0 is obtained by Y.Chuang et al. [1]:

$$c = y''/(1 + y'^2)^{3/2} \qquad (3.1)$$

Where $y' = \frac{dy}{dx}$ and $y'' = \frac{d^2y}{dx^2}$

The first derivative of y with respect to x is arrived at by implicit differentiation of isophote.

$$\frac{dy}{dx} = -\frac{L_x}{L_y} = -L_x L_y^{-1} \qquad (3.2)$$

$$\frac{d^2y}{dx^2} = \frac{-L_y^{\,2}L_{xx} + 2\,L_x L_y L_{xy} - L_x^{\,2}L_{yy}}{(L_x^{\,2} + L_y^{\,2})^{3/2}} \qquad (3.3)$$

Where L_x and L_y are the first derivative of L_0 with respect to x and y.

The second derivative is the following:

$$\frac{d^2y}{dx^2} = -L_y^{-1}L_{xx} - L_y^{-1}L_{xy}\frac{dy}{dx} + L_x L_y^{-2}L_{yx} + L_x L_y^{-2}L_{yy}\frac{dy}{dx} \qquad (3.4)$$

Where L_{xx}, L_{xy} and L_{yy} are the second partials derivative in x and y.

So $\frac{d^2y}{dx^2}$ becomes

$$\frac{d^2y}{dx^2} = \frac{-L_y{}^2 L_{xx} + 2\,L_x L_y L_{xy} - L_x{}^2 L_{yy}}{L_y{}^3} \tag{3.5}$$

The curvature c is given by Y.Chuang et al. [1]

$$c = \frac{-L_y{}^2 L_{xx} + 2\,L_x L_y L_{xy} - L_x{}^2 L_{yy}}{(L_x{}^2 + L_y{}^2)^{3/2}} \tag{3.6}$$

We are interested in the osculating circle and use it to represent the objects scale information. Because the curvature c is the inversely proportional to the radius of the osculating circle and the isophote curvature depends on the intensity of the outer side of the curve X. Shen et al [9].

The final formulation of the radius is obtained by multiplying the gradient with the inverse of the isophote curvature c is given by:

$$d(\mathrm{x}, \mathrm{y}) = \frac{\{L_x, L_y\}}{\sqrt{L_x{}^2 + L_y{}^2}} \times \frac{(L_x{}^2 + L_y{}^2)^{3/2}}{-L_y{}^2 L_{xx} + 2\,L_x L_y L_{xy} - L_x{}^2 L_{yy}} \tag{3.7}$$

$$d(\mathrm{x}, \mathrm{y}) = \frac{\{L_x, L_y\}(L_x{}^2 + L_y{}^2)}{-L_y{}^2 L_{xx} + 2\,L_x L_y L_{xy} - L_x{}^2 L_{yy}} \tag{3.8}$$

where ‘d’ is a displacement vector to estimate the position of a potential structure. Once d is obtained, a potential structure is available. Potential structures are distributed in different parts of an image. We focus on the potential structures that are different from their surroundings. Thus, the potential structures are used as centre regions for computing centre–surround contrast. With respect to the surrounding region, d can be used as a cue to set its size: the radius of the surrounding region can be set by a constant multiplicative factor k $(k > 1)$:

$\mathrm{d}_{outer} = k \times d$ and the value of k is set based on experiments.

After determining the centre and surrounding regions, the centre– surround contrast can be computed, which is used to represent the corresponding centre pixel’s saliency information. Because the centre–surround contrast computation is a local operation, the obtained saliency information is referred to as the local saliency. Another point that should be mentioned is the number of pixels that belong to a potential structure (having the same displacement to a centre pixel). The larger this number is, the more significant the corresponding centre pixel is. Compared to the center surround contrast computation, the number of pixels that belong to it is computed by the isophote-based operator in a global manner. We therefore call it global saliency S_g. directly used this global saliency as the image saliency. However, this strategy is liable to be affected by noise, especially for images of nature. In contrast to, a pixel's saliency is determined by the combination of its global and local saliency information, which could contribute to noise reduction Y. Chuang et al. [1].

3.2.2 CENTRE–SURROUND CONTRAST COMPUTATION AND FINAL SALIENCY CONSTRUCTION

We introduce an isophote-based operator to determine both the center and surround regions. Here, an integral image-based operator is employed to compute center–surround contrast for all three of the feature channels, which are then fused into a final saliency map. The overall proposed process of detecting image saliency is shown in Fig.: an image is first converted into three sub-images based on the YCbCr colour space; the isophote-based operator is then employed to determine potential structures and global saliency information for each sub-image. These two steps are described in detail as follows:

The potential structures are used to compute center–surround contrast, which is then combined with global saliency to build a final saliency map Y.Chuang et al [1].

(1) Given an image I, we first convert it into sub-images based on the YCbCr color space: The reason that we chose the YCbCr colour space is its perceptually uniformity, indicating that a change in a colour value would be perceived as approximately a change of the same amount in the human visual perception system .

(2) For each sub-image I sub image A {Y, Cb, Cr}, the isophote-based operator is adopted to extract the potential structure and global saliency information S_g related to each pixel. The potential structure is then used to determine the centre and surrounding regions for each pixel, as mentioned above. For the intensity and colour feature channels, an integral image mechanism is directly employed to compute the centre–surround contrast. To extract the most salient pixels, the global saliency information S_g is used as a weighting for the centre–surround contrast computation:

$$R_{\text{int}}(\text{x},\text{y}) = S_g^L(\text{x},\text{y}).\left|r_{center}^L(\text{x},\text{y}) - r_{surround}^L(\text{x},\text{y})\right| \quad (3.9)$$

$$R_{Col}(\text{x},\text{y}) = \frac{1}{2}\sum_{\eta=\{\text{Cb},\text{Cr}\}} S_g^\eta(\text{x},\text{y}).\left|r_{center}^\eta(\text{x},\text{y}) - r_{surround}^\eta(\text{x},\text{y})\right| \quad (3.10)$$

Where R_{int} and R_{Col} represent the intensity and color feature channel map respectively; S_g is the global saliency information; r_{center} and $r_{surround}$ represent the average of the center and surround regions around the position (x, y) respectively.

The orientation computation method proposed by is based on an integral image strategy . It is naturally integrated into the method to compute orientation centre–surround contrast, because all of the three feature channel maps can be computed by the integral image-based operator, which could be conducive to the fusion of them. Instead of computing the angle and magnitude, we compute the orientation contrast between the centre and surrounding regions by the Euclidean distance between their orientation vectors. The orientation feature map is obtained by Y.Chuang et al [1]:

$$R_{Ori}(\text{x},\text{y}) = \frac{1}{3}\sum_{\eta=\{\text{Y},\text{Cb},\text{Cr}\}} S_g^\eta(\text{x},\text{y}).\text{r}^\eta(\text{x},\text{y}) \quad (3.11)$$

Where R_{Ori} is the orientation feature map, r stands for the orientation centre–surround contrast at position(x, y). Because all of the three types of feature channels are computed by an integral image strategy, it is convenient to construct a final saliency map across multiple channels. The final saliency map is a linear combination of all of the three feature maps Y.Chuang et al [1]:

$$M_{saliency} = \frac{1}{3}(\text{R}_{\text{int}} + \text{R}_{Col} + \text{R}_{Ori}) \quad (3.12)$$

The combination of global and local saliency information can effectively detect the most salient pixels.

3.3 GABOR FILTER FOR TEXTURE ANALYSIS

Gabor filter is widely used for image description Timo and Matti et al [27]. Gabor filter banks are very important tool for extraction of information from a visual scene. In Gabor filters the mean and standard deviation are used for texture description Angelet al [26]. In texture analysis, a very common method is to convolve an image with N filters whose responses at a certain position (x, y) form an N-dimensional vector. Gabor transform plays a very important role in image processing and for texture analysis because of their excellent properties. Initially the 2-D Gabor filters are used for understanding the orientation-selective and spatial–frequency selective receptive field properties of neurons in the Brains' visual cortex Pradeep and Jayant et al [28]. To extract the texture features, one has to select bandwidth, the frequency (f_0) and orientation (θ) of Gabor filter. By this one can extract the local and global texture features.

Wavelet transforms are used to extract the spatial and frequency information from a signal. In comparison with the other wavelet transforms, the Gabor wavelet transform has important properties in terms of mathematics and biology which are very useful for texture description. Furthermore, the Gabor wavelet gives the optimal resolution in time and frequency domain. The Gabor filters also calculate multi-orientation information from a hand posture images at different scales and orientations. The extracted information may be local in nature. The approach for designing the Gabor filter for hand posture or other applications e.g. face recognition is to construct a filter bank with different-2 scales and orientations. Hence by the Gabor filters, we can extract features at different-2 angles and scales Angel et al. [26]. The main objective of Gabor filters is to capture the visual properties e.g. spatial locality, orientation selectivity, and spatial frequency characteristics. Because of these important characteristics various applications employ Gabor filter for feature representation Xuewen et al [29]. The Gabor filter is expressed as the multiplication of a 2D Gaussian function and a complex sinusoidal function also called as a complex exponential function. Thus the Gabor filter is represented as Pradeep and Jayant et al [28]:

$$G(x, y) = \exp(-(x_0^2 + \kappa^2 y^2) / 2\sigma^2) \times \cos(2\pi\, x_0 / \lambda) \tag{3.13}$$

$$x_0 = x\cos\theta + y\sin\theta$$

$$y_0 = -x\sin\theta + y\cos\theta$$

Where λ is the wavelength and σ is the standard deviation of the Gaussian factor. σ gives the information about the size of the receptive field. The bandwidth is represented as σ/λ, which gives the visible parallel stripes. In the proposed method this ratio has been fixed to $\frac{\sigma}{\lambda} = 0.50$. κ represents aspect ratio and gives the ellipticity of the receptive field. Gabor filter bank with different scale and orientations is being used. We use four different scales and five orientations.The Gabor function is shown on the output window.

These steps can be easily understood as:

(1) The input image is first filtered by the Gabor filter which creates different intensity profile. Then we smoothen it by a Gaussian filter which brings the intensities in the discrete levels.

(2) Now as the intensity are differentiated so we calculate the features vectors which are used for texture representation and hence a texture based segmented image is generated.

Some properties of Gabor filters:

- A tunable band pass filter
- Similar to a STFT or windowed Fourier transforms
- Satisfies the lower-most bound of the time-spectrum resolution (uncertainty principle)
- It‘s a multi-scale, multi-resolution filter
- Has selectivity for orientation, spectral bandwidth and spatial extent.
- Has response similar to that of the Human visual cortex (first few layers of brain cells)
- Used in many applications – texture segmentation; iris, face and fingerprint recognition.
- Computational cost often high, due to the necessity of using a large bank of filters in most applications.

As in case of hand posture recognition, Gabor filter is used for the feature extraction and the output vector of Gabor filter has high dimensionality. If the dimension of input image is given 160x120 pixels then after the convolution of the image with Gabor filter bank of 3 scales and 5 orientations the dimensionality will become 16x 160x 120=307200. The 2D Gabor filtered images are presented as a pattern vector and this process is repeated for all other 16 responses. At last all the pattern vectors for 16 responses are arranged either in rows or columns as features.

CHAPTER 4

CLASSIFIERS

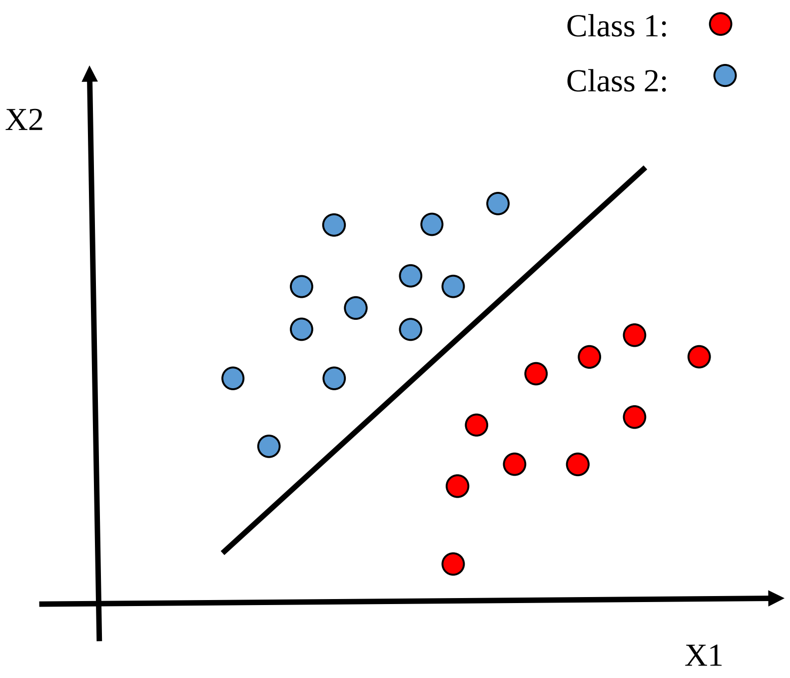
4.1 SUPPORT VECTOR MACHINE

For the classification of different classes an intelligent classifier is used. Various kinds of classifier are used for hand gesture classication like HMM, Neural Network, Bayesian Network etc. The classifier is used because of their simplicity and computational attractiveness. The classifiers like Neural Network and HMM are slow to converge at a solution. They take more time to classify the data. So we employ such a classifier in which computational time required is less and also give the correct classification. SVM satisfies all the requirement that is why we employ this classifier.

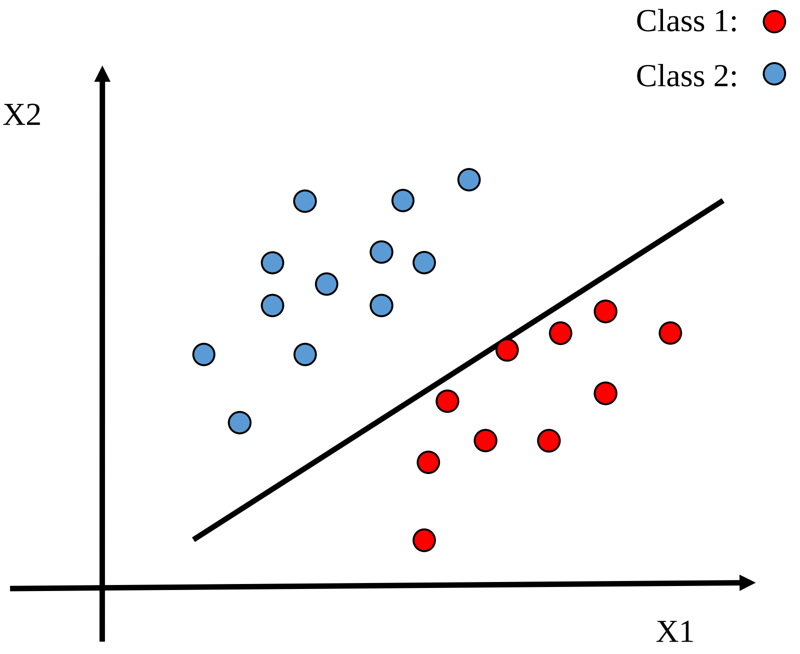
Support Vector Machine (SVM) is a very important tool used for classification and regression in which machine learning theory is employed to maximize accuracy. Machine learning is basically a computer program by which the system learns from dataset having some class of the tasks and performance measure. Support Vector machines viewed as the system in which some hypothesis space of linear discriminant functions in a high dimensional feature space is used and the system is trained with a learning algorithm. Each machine learning algorithm focuses how to represent the simple functions in high dimensionality space. So here, our aim is to generate an output hypothesis which can classify the training data correctly. The ability to classify the data that is not used for training set of the hypothesis is known as its generalization. A learning bias derived from statistical learning theory is being used in SVM Yu-Chieh et al. [22].

In support vector machines we choose a decision boundary for which the margin is maximized. Maximum margin is taken to minimize the classification error of different classes. Consider a simple two class problem. A machine has to learn the boundary between the two classes so that given a new sample as input, it should indicate in which class it actually belongs to, i.e. either class 1 or class 2 for two classes. There may be infinite many hyper planes that exist for separating the classes exactly. Figure (4.1) shows the some possible hyper planes. Figure 4.2 shows, there may be an infinite number of hyper-planes that can separate the two classes. If there exist multiple solutions which can classify the training dataset correctly, then our aim is to find such a hyper plane in which the generalization error is minimized. The maximum margin concept is employed in the support vector machine Madzarov et al [23]. This may be defined as the smallest distance between the decision hyper-lane boundary and any of the data points. However, only one of hyper-plane achieves the maximum separation between the samples of the two distinct classes. This hyper plane is known as maximum margin classifier, shown in figure (4.3).

Support Vector Machines use geometric properties to exactly calculate the optimal separating hyper-plane from the training data. There may also be non-linearly separable cases, i.e. there exists no straight line which can correctly classify the data point. So there can be several such hyper-planes that may separate the two classes prominently. But not all of them are good hyper-planes because the points which are near the hyper-plane and the points nearer to them may be classified differently where as they belong to the same class.



(a)



(b)

Figure 4.1: (a), (b) Hyper planes separating the classes

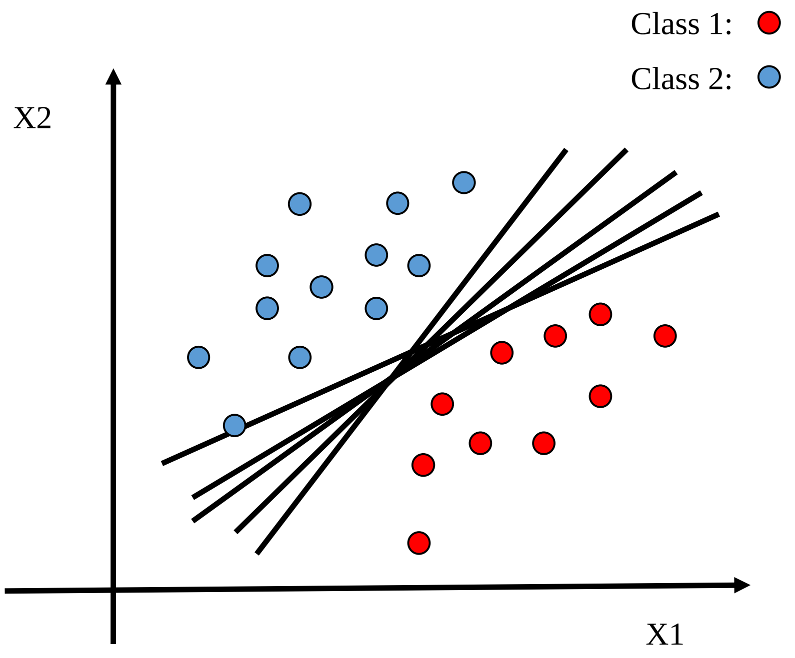


Figure 4.2 Several possible hyper plane for class separation

Therefore, we choose such a hyper- plane in which the margin is maximum i.e. the hyper-plane with largest possible margin. If we randomly choose a hyper plane to classify the dataset, it may be close to one set of dataset points as compared to others and this is not desirable. So the maximum margin classifier is a possible solution to correctly classify ted at points. This maximum margin hyper plane is known as the optimal separating hyper plane as shown in figure 4.3. The margin may be defined as the width that the boundary could be increased by before hitting a data point. The linear discriminant function (classifier) with the zone of maximum margin is the best. Suppose x is a vector point and w is the weight. Now using the equation of a straight line, we have,

$w^T x^+ + b = 1$ line which touches class1 (4.1)

$w^T x^- + b = -1$ line which touches class 2 (4.2)

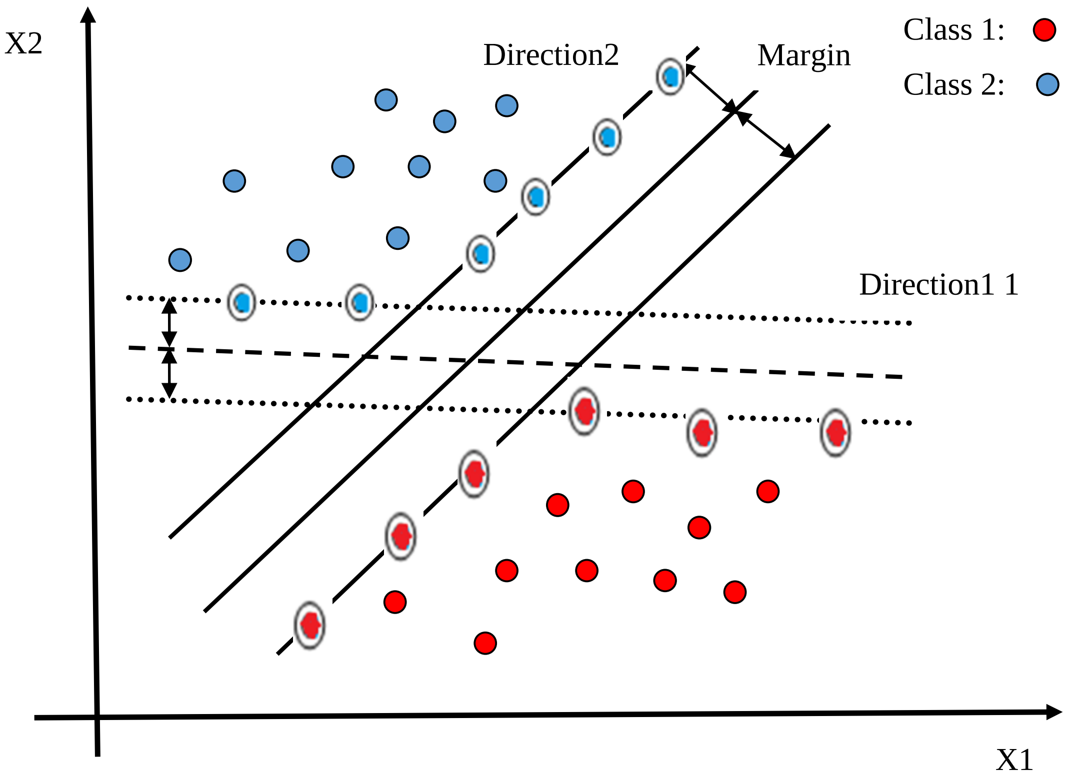


Figure 4.3. The bounding planes of a linear SVM with a soft margin (i.e. with some errors)

4.2 MULTI CLASS SVM

Support vector machines (SVM) [22] initially designed for binary classification problems, i.e. a two-class classifier. In practice, it is possible to have tackled problems involving $N > 2$ classes. Many methods already has been proposed to combine multiple two-class SVMs to build multiclass classifier.

There are following types of multiclass SVM Yiguang and Chen et al [24] [25]:

(i) One-against-all

For the given N-class problems where $N > 2$, N two-class SVM classifiers may be constructed to finally build a multiclass classifier Chen et al [25]. When we train the i^{th} SVM classifier the i^{th} class samples are considered as positive examples while all the rest samples of other classes are considered as negative examples. For the recognition part, one can give a test sample as input to N SVMs classifiers and output class is assigned according to the maximum of N classifier output. The drawback of this method is that when the training samples of classes is large then training becomes very difficult.

(ii) One-against-one

Here one can build $N(N-1)/2$ two-class classifiers. All classifiers used in SVM are the binary pairwise combinations of the N classes. While training the each classifier, the data points of the first class are considered as positive examples and the data points of the second class are considered as negative examples. When we combine these $N\ (N-1)\ /2$ classifiers, we adopt Maximum Wins algorithm to find the output class by voting the classes according to the results of each of the classifier and then finding the most voted class. The disadvantage of this method is that every test sample has to be presented to the large number of classifiers $N\ (N-1)\ /2$. This results in faster training but slower testing, especially when the number of the classes in the problem is big Chen et al [25].

So to construct a multi-class classifier, one way is to construct N separate SVMs. Here the N^{th} model is trained from the N^{th} class as the positive examples and all the remaining $N-1$ classes are treated as the negative examples.

The other way to build a multi-class classifier is to decompose an N-class problem into a number of two-class problems, in which one-against-all the most commonly used implementations Pramod et al [2]. The results from the multiple classifiers are used in the final decision as $D_i(\mathrm{x}) \geq -1$.

Let the maximum margin output that separates i^{th} class from the remaining classes, is given by :

$$D_i(x) = w_i^T \varphi(x) + b_i \tag{4.3}$$

Where w_i is the l-D vector, $\varphi(x)$ is known as the mapping function that maps the data points x into the l -D feature space, and b_i is known bias. The optimal hyper plane is selected for classification, and if it is possible to separate classes then the training data belonging to class i satisfy $D_i(\mathrm{x}) = 0$ and for the remaining classes satisfy $D_i(\mathrm{x}) \leq 1$. If it is not possible to the dataset into classes then unbounded support vectors satisfy $|D_i(\mathrm{x})| = 1$ and bounded support vectors belonging to class i satisfy $D_i(\mathrm{x}) \leq 1$and those belonging to a class other than class i satisfy $D_i(\mathrm{x}) \geq -1$. Data points from the different classes, x is classified into the class is given by Gjorgji et al [23]:

$$\arg {}_{i=1,..n}^{max} D_i(x) \tag{4.4}$$

How to effectively cast the multiclass problems is still an on-going issue.

CHAPTER 5

THE PROPOSED FRAME WORK

5.1 THE PROPOSED FRAME WORK

In this project, our analysis is based on the hand gestures that a person can do to interact with computer system. Our project main goal is on the recognition of hand postures in any complex background. The detection of hand under complex background is very challenging task, and it is notably the most difficult to implement in a satisfactory way. Several different approaches have been proposed so far. We use NUS data-set II [21], and on the hand images we perform some preprocessing operation such as Skin detection, Saliency and texture on the classes available in the NUS data-set II [21] and extract some feature vectors. Now these feature vectors are used as input for classifier.

The block diagram of the proposed method is shown below:

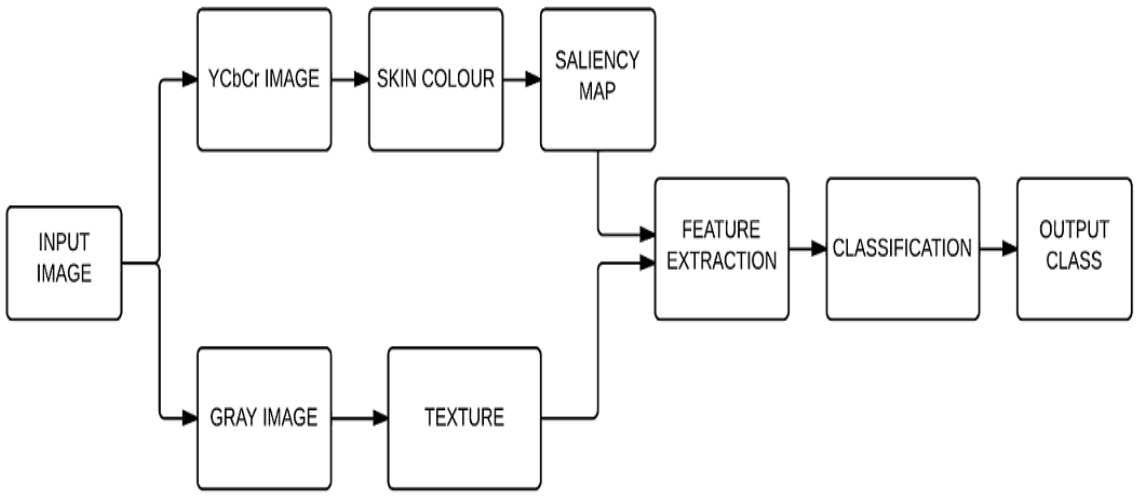


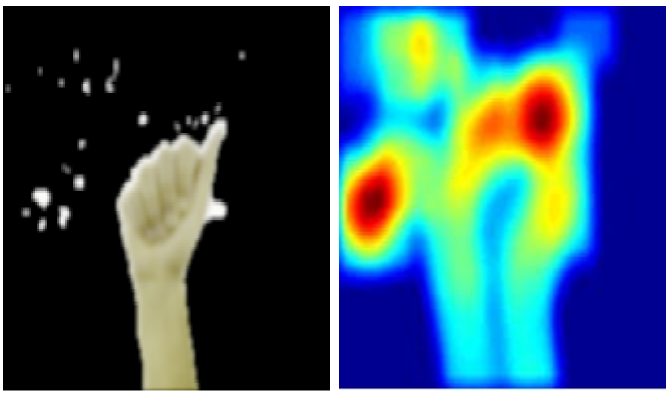
Figure 5.1 Block diagram of proposed method

To start with, the skin based features are extracted from a map which represents the similarity of pixels to human skin color, using a computational model [19]. Some preprocessing is also done on the image to improve the skin detection. We improve the contrast of image so that skin information can be easily extracted. Here we convert the input image which is RGB colour space from the dataset in YCbCr colour space. In YCbCr the effect of illuminance is minimum.. Then, the color features are extracted by the discretization of chrominance color components YCbCr color spaces, and the similarity to skin map. Skin region are detected using thresholding method as skin has a limited range of intensity in the space. This skin map is very useful to detect and identify the hand region in complex background images. This information is extracted using feature based visual attention. Skin region detected in the images are used for further processing. Figure 5.2 shows the extracted skin region.

(a) (b) (c)

Fig 5.2 (a) original image (b) preprocessed image (c) skin image

Saliency map is detected from the skin detected regions. This gives the information about the pixels which are useful in hand detection. So here we can extract some feature which are useful in the representation of hand in the image. Once the saliency map is computed, we extract some features and named as $M_{saliency}$. Output saliency image of the skin detected image is shown below in figure 5.3.



(a) (b)

Fig 5.3 (a) Skin preprocessed Image, (b) Saliency map

Now texture feature are extracted from the input image. To extract the texture feature, first we convert the input image into a gray level image, then texture feature are extracted using Gabor filter. In Gabor filter four different scales and five orientations are used for texture analysis. Finally we calculate the texture feature named as ‘Texture feature’.

The features are extracted which are based on skin color, saliency map and texture, which are extracted from the specific hand gestures and then they form the basic components which are used to recognize the other similar gestures. Then we calculate the saliency map of the skin region. Feature are extracted from the saliency

Next, the texture feature is extracted using Gabor filter. The Gabor filters are convolved with hand posture images to extract texture features of the different class. Then we take the average of Saliency map feature and Texture feature to calculate the finally feature vectors of the hand postures.

Final feature are calculated as follows:

$$\text{Final feature}=\frac{1}{2}(M_{saliency}+\text{Texture feature}) \qquad (5.1)$$

Now, as we have gathered all the concerned features. Now we employ multi-class SVM classifier which is trained with the extracted features and used to perform the hand gesture classification. Consequently, the hand postures are recognized using skin color, saliency map and texture based features (of the hand region), with a Multi-class Support Vector Machines (SVM) classifier.

In hand gesture recognition system we used the database from the standard hand gesture database, NUS-HANDSET database-II [21]. The proposed algorithm is reliable against complex and skin color backgrounds as the segmentation of the hand region employ the efficient use of the combination of color, saliency as well as texture features. The experimental results show that the algorithm has a person independent performance. The proposed algorithm has robustness against variation in hand size and its position in the image and it is naturally the case, as we have worked upon the general dataset.

CHAPTER 6

RESULTS & DISCUSSION

6.1 Results

In our project, we considered total 2000 images of 10 classes from NUS data-set II [21] to classify a hand posture. There are 2000 images of hands with different background. We applied the algorithm on the hand images of ten classes in the dataset. The output images of skin, saliency and texture output are shown below:

Skin and Saliency Map output:

(a) Saliency of skin image output of class A

(a) Saliency output of class A

(b) Saliency of skin image output of class B

(c) Saliency output of class B

(d) Saliency of skin image output of class C

(f) Saliency output of class C

(e) Saliency of skin image output of class D

(f) Saliency output of class D

(g) Saliency of skin image output of class E

(h) Saliency output of class E

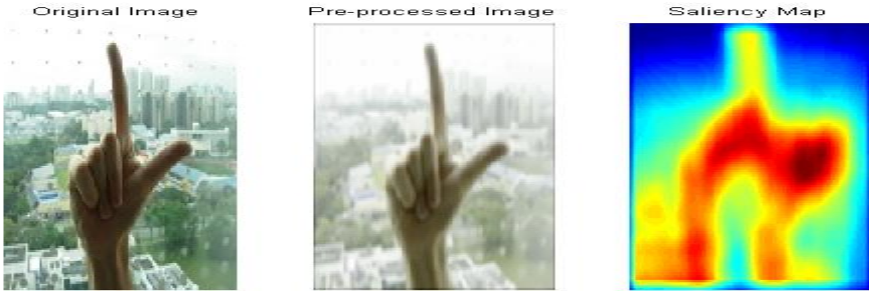
(i) Saliency of skin image output of class F

(j) Saliency of skin image output of class F

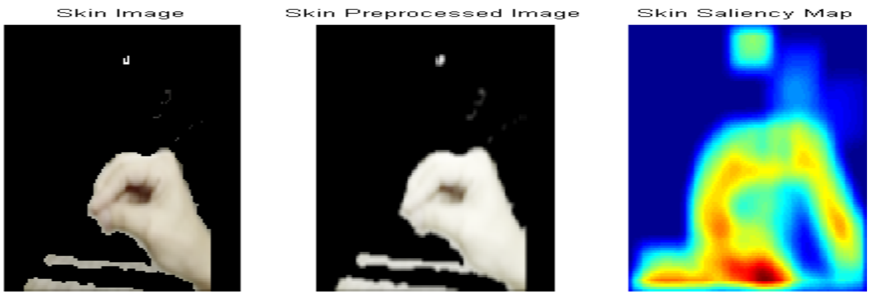
(k) Saliency of skin image output of class G

(l) Saliency output of class G

(o) Saliency of skin image output of class H



(p) Saliency output of class H



(q) Saliency of skin image output of class I

(r) Saliency output of class I

(s) Saliency of skin image output of class J

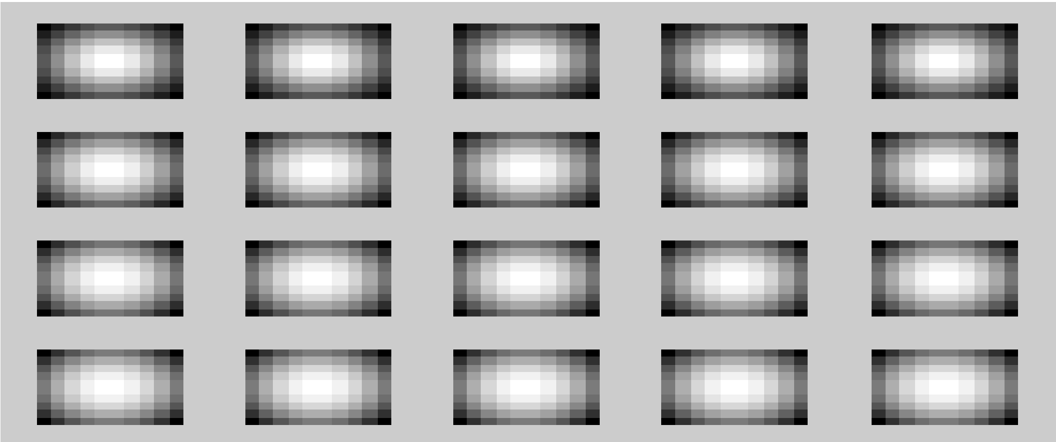
(t) Saliency output of class J

Figure 6.1 (a) to (t) show Skin Saliency and Saliency output images from different classes in the dataset.

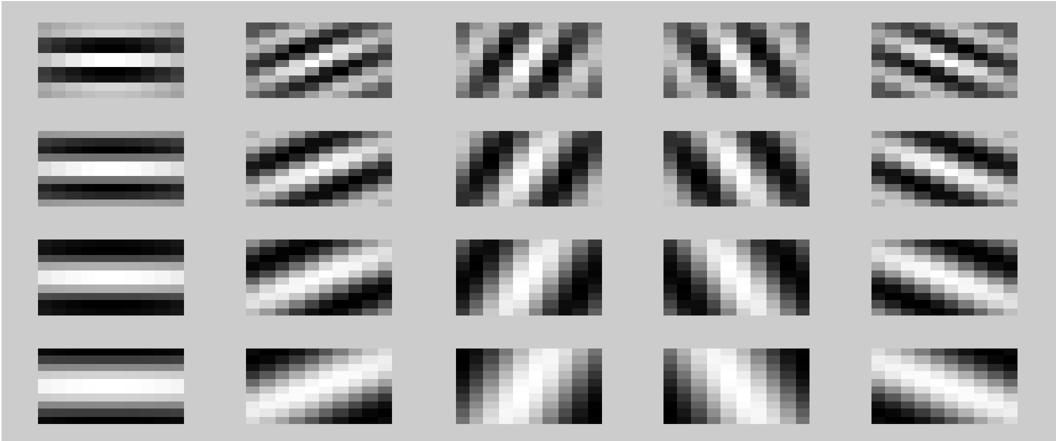
Figure 6.1 shows the original images, pre-processod image, skin region extracted from the pre-processed image , saliency map of skin extracted image. It also shows the the saliency of the original pre-processed image.

Gabor filter output:

The response of Gabor filter is shown in figure 6.2. In the figure 6.2 the magnitude of Gabor filter for four scales and five different oreintations are shown.



(a)



(b)

Figure 6.2: Response of Gabor filter (a) showsthe magnitude of Gabor filter; (b) shows 4 different scales and 5 different orientations

(a)

(b)

Figure 6.3: output of Gabor filter on hand image with complex background (a) shows the real part of Gabor filter output; (b) shows the magnitude of Gabor filter output

(a)

(b)

Figure 6.4: output of Gabor filter on hand image with human noise (a) shows the real part of Gabor filter output; (b) shows the magnitude of Gabor filter output

We apply the Gabor filter on the hand gesture images of different classes. The output of Gabor filter on hand image is shown in fig. 6.3. in the figure four scale and five different orientations are shown.

Figure 6.4 shows the output of Gabor filter when applied on a hand image with human noise. Finally feature are extracted as average of the saliency and texture features.

6.2 DATABASE DESCRIPTION

In this project we basis the work on the standard database of NUS hand posture dataset-II [21]. The dataset contains 10 classes hand posture. The hand postures of various classes are captured in the National University of Singapore (NUS), with different-different backgrounds, with varying hand size and shapes. The Classes in the dataset are represented as 'a' to 'j'. There are total 2000 hand posture images with 200 images per posture in the dataset. The classes are shown in figure 5.1. The other images in the dataset are the images with human noise. Total 750 images are of this type 75 per class. These hand posture images contain human noises such as face of the posturer, human in the background etc. The Classes are represented as 'a_HN' to 'j_HN' respectively. The hand postures images are taken from 40 subjects, all of these with different ethnicities and against different complex backgrounds. There are both males and females included in the dataset in the age of 22 to 56 years. These are asked to present the 10 hand postures, 5 times each. The hand muscles are also loosen after each shot so as compare to the natural variations in the postures. Also, the hand images have indoor as well as outdoor complex backgrounds. The hand postures have wide intra class variations in hand sizes and appearances. Furthermore, the database also contains a set of background images which is used to test the hand detection capability of the proposed algorithm. The recognition algorithm is tested with the new dataset using a 10 fold cross validation strategy. The hand postures are chosen such that the inter class variation in the appearance of the postures is less, which makes the recognition task challenging. Our project work is to recognize the classes. One can give any test image as input to the system. Then the system gives the information about the gesture image that was given as input. The system performance is dependent on the type of data. The images are represented in an RGB colour image and then converte it into YCbCr image. The images of different gestures/signs are captured in a background closely resembling to natural environment.

Class A **Class B** **Class C** **Class D** **Class E**

Class F **Class G** **Class H** **Class I** **Class J**

Figure 6.5 Samples of Images of different classes from database

Output of Classifier

We apply the algorithm on the NUS dateset-II. We extracted the feature of 2000 images of different hand classes and finally classify the hand gesture by using multi-class SVM. While classification we made a confusion matrix for the classification of gestures. The confusion matrix is a table with rows and columns that gives the idea of classification of data. What percentage of data is correctly classified can be estimated from this matrix. This gives the reports about the number of false positives, false negatives, true positives, and true negatives. This allows more detailed analysis than the mere proportion of correct guesses (accuracy).

Accuracy is not a reliable metric for the real performance of a classifier, because it will yield misleading results if the data set is unbalanced (that is, when the number of samples in different classes vary greatly). Accuracy can be of little help, if classes are severely unbalanced.

The True Positive (TP) and True Negative (TN) are the correct classifications. But, False Positive (FP) and False Negative (FN), give the incorrectly classified data. False Positive (FP) is the lower risk. It‘s similar to the false probability in the Bayesian Theory. It means, that the dataset doesn‘t belong to the actual class, but the classifier has recognized it as belonging to that class. Furthermore, accuracy assumes an equal cost for both kinds of errors (FP and FN).

Confusion Matrix for the classification

	TABLE-I Confusion Matrix of Input Posture v's Recognized Posture									
Recognized Posture class / Input Gesture class	A	B	C	D	E	F	G	H	I	J	RR (%)
A	50	0	0	0	0	0	0	0	0	0	100
B	3	47	0	0	0	0	0	0	0	0	94
C	1	1	48	0	0	0	0	0	0	0	96
D	1	0	1	48	0	0	0	0	0	0	96
E	0	0	0	1	49	0	0	0	0	0	98
F	0	0	0	1	1	48	0	0	0	0	96
G	0	0	0	2	2	0	46	0	0	0	92
H	3	0	2	2	0	0	0	43	0	0	86
I	0	0	0	0	0	0	0	0	50	0	100
J	0	0	1	0	0	1	0	0	0	48	96
Overall recognition rate (%) = 95.4%											

Table 1: Confusion Matrix for the classification

Accuracy(%)= (Accurately recognized gestures/ Total recognized gestures) * 100%

= (477/500)*100%

= 95.4%

CHAPTER 7

CONCLUSION

7.1 CONCLUSION

In this project we present a hand gesture algorithm, which involves classification of a gesture. It efficiently makes the use of different color spaces, skin detection, saliency map and texture of an image. To recognize the hand gesture classes the features are extracted using the proposed methodology. These features are used for the classification purpose. Finally the application of multi-class SVM classifier is employed to recognize the hand gesture class. This is a new approach which works well even in complex backgrounds. Our algorithm also minimizes the effect of the intensity of light in the image. It is also invariant of lateral orientation, which has a lot of significance as it makes the hand gesture recognition system realistic.

7.2 FUTURE SCOPE

From the previous discussion we can emphasize that the hand recognition against a complex background is a very challenging work. Our technique finds little difficulty in a few situations when human noise is present. So our future work mainly focuses on sorting out the limitations when human noise does not affect the recognition.

Also computation complexity is high in our method so reduction of complexity leads us to a less computation time and hence a fast hand recognition algorithm, so we may easily go for real time recognition.

So our hand recognition system can find application in many fields, e.g. robotics, human computer interaction.

REFERENCES

[1] Yue long Chuang, Ling and Gencai "Saliency-guided improvement for hand posture detection and recognition", *Neuro computing* 133(2014) pp. 404–415.

[2] Pramod Kumar Pisharady Prahlad Vadakkepat Ai Poh Loh, "Attention Based Detection and Recognition of Hand Postures Against Complex Backgrounds" *Int. J Computer Vision* (2013) 101: pp. 403–419.

[3] Yuseok, Sang-Ki, Sooyeon, Kar-Ann and SangyounLee, "Face detection based on skin color likelihood" *Pattern Recognition* 47(2014) pp. 1573–1585.

[4] Zhou Ren, Junsong Yuan,Jingjing Meng and Zhengyou Zhang, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor" *IEEE Transactions on Multimedia*, VOL. 15, NO. 5, AUGUST 2013, pp. 1110-1120.

[5] Arpita Ray, G. Sanyal and S. Majumder, "Hand Gesture Recognition Systems: A Survey" *International Journal of Computer Applications* (0975 – 8887) Volume 71– No.15, May 2013, pp. 26-37.

[6] S. Padam and Prabin Kumar, "A robust static hand gesture recognition system using geometry based normalizations and Krawtchouk moments", *Pattern Recognition* 46 (2013) pp. 2202–2219.

[7] Shikha, Jafreezal, Wan Fatimah Wan "Static Hand Gesture Recognition Using Local Gabor Filter" *Procedia Engineering* 41 (2012) pp. 827 – 832.

[8] A. M. Aibinu, A. A. Shafie and M. J. E. Salami "Performance Analysis of ANN based YCbCr Skin Detection Algorithm" *Procedia Engineering* 41 (2012) pp.1183 – 1189.

[9] Xiaohui, Gang, Lance and Ying "Dynamic hand gesture recognition: An exemplar-based approach from motion divergence fields" *Image and Vision Computing* 30 (2012) pp. 227–235.

[10] Rehanullah, Allan, Julian and Abdul "Color based skin classification" *Pattern Recognition Letters* 33 (2012) pp. 157–163.

[11] Hu Han, Shiguang, Xilin Chen and Wen "A comparative study on illumination preprocessing in face recognition" *Pattern Recognition* 46 (2013) pp. 1691–1699

[12] Ting-hui, Yi-min Yu, Xing-guo Qin "A High-Performance Skin Segmentation Method" *Procedia Engineering* 15 (2011) pp. 608 – 612

[13] Serge Belongie, Jitender Malik and Jan Puzicha, "Shape matching and Recognition using Shape Contexts" *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.24, NO. 24, April 2002, pp. 509-522.

[14] Jing-Wein, Chou-Chen,Jiann-Shu Lee, "Genetic Eigen hand selection for hand shape classification based on compact hand extraction" *Engineering Applications of Artificial Intelligence* 26(2013) pp. 2215–2226

[15] Sigal Berman and Helman Stern, "Sensors for Gesture Recognition Systems" *IEEE Transactions on Systems, man and cybernetics- part C: Application and Reviews*, Vol. 42, No. 3, May 2012, pp. 277-290.

[16] Nasser H. Dardas and Nicolas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques", *IEEE Transactions on Instrumentation and Measurement*, Vol. 60, No. 11, Nov. 2011, pp. 3592-3607.

[17] Xu Zhang, Xiang Chen, Yun Li, Vuokko Lantz, Kongqiao Wang, and Jihai Yang, "A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors" *IEEE Transactions on Systems, Man and Cybernetics- part A: Systems and Humans*, Vol. 41, No. 6, Nov 2011, pp. 1064-1076.

[18] Sushmita Mitra and Tinku Acharya, "Gesture Recognition: A Survey" *IEEE Transactions on Systems, man and cybernetics- part C: Application and Reviews*, Vol. 37, NO. 3, MAY 2007, pp.311-324.

[19] Jose M. Chaves, J. M., Vega-Rodrígueza, M. A., Gómez- Pulidoa, J. A., & Sánchez-Péreza, "Detecting skin in face recognition systems: a colour spaces study", *Digital Signal Processing* 20(2010), pp. 806–823.

[20] Andrew Wilson and Aaron Bobick, "Learning visual behavior for gesture analysis", *In Proceedings of the IEEE Symposium on Computer Vision*, November 1995, pp. 19-21,.

[21] http://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/.

[22] Yu-Chieh, Yue-Shi, Jie-Chi, "Robust and efficient multiclass SVM models for phrase pattern recognition", *Pattern Recognition* 41 (2008) pp. 2874 – 2889.

[23] Gjorgji Madzarov, Dejan Gjorgjevikj, Ivan Chorbev, "Multi-class Classification using Support Vector Machine in Binary Tree Architecture", *International Scientific Conference*, 21 – 22 November 2008, pp. 413-418.

[24] Yiguang Liua, Zhisheng and Liping, “Anovel and quick SVM-based multi-class classifier”, *Pattern Recognition* 39 (2006) pp. 2258 – 2264.

[25] Chen ping, Feiping, Changshui and Dongyun, “Multiple rank multi-linear SVM for matrix data classification”, *Pattern Recognition* 47(2014) pp. 454–469.

[26] Angel, Isaac, Cristina and Enrique, “Analysis of variance of Gabor filter banks parameters for optimal face recognition”, *Pattern Recognition Letters* 32 (2011) pp. 1998–2008.

[27] Timo and Matti, “Image description using joint distribution of filter bank responses”, *Pattern Recognition Letters* 30 (2009) pp. 368–376

[28] Pradeep and Jayant, “Rotation and intensity invariant shoeprint matching using Gabor transform with application to forensic science”, *Pattern Recognition* 42 (2009) pp.1308-1317

[29] Xuewen, Xiaoqing and Changsong, “Gabor filters-based feature extraction for character recognition”, *Pattern Recognition* 38 (2005) pp. 369 – 379

[30] R.Valenti, N.Sebe, and T.Gevers, “Image saliency by isocentric curvedness and color”, *IEEE International Conference on Computer Vision*, 2009, pp. 2185–2192.

[31] X.Hou and L.Zhang, “Saliency detection: a spectral residual approach”, *IEEE Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[32] B.Schauerte and R.Stiefelhagen, “Predicting human gaze using quaternion DCT image signature saliency and face detection”, *IEEE Workshop on the Applications of Computer Vision*, 2012, pp. 9–11.

[33] B.Schauerte and R.Stiefelhagen, “Quaternion-based spectral saliency detection for eye fixation prediction”, *European Conference on Computer Vision*, 2012, pp. 7–13