Real Time Hand Gesture Recognition Based Interface for Microsoft Word Document Handling

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Software Engineering

Submitted By
Kapil Yadav
Roll No. 801331009

Under the supervision of:
Dr. Jhilik Bhattacharya
Assistant Professor, CSE Department

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

July 2015

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*Real Time Hand Gesture Recognition Based Interface For Microsoft Word Document Handling*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Jhilik Bhattacharya* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

(Kapil Yadav)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Jhilik bhattacharya)

Assistant Professor,

CSE, Department

Countersigned by

(Dr. Deepak Garg)

Head

Computer Science and Engineering Department

Thapar University

Patiala

(Dr. S. S. Bhatia)

Dean (Academic Affairs)

Thapar University

Patiala

ACKNOWLEDGEMENT

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life.

This work would not have been possible without the encouragement and able guidance of my supervisor **Dr. Jhilik Bhattacharya**. I thank my supervisor for their time, patience, discussions and valuable comments. Their enthusiasm and optimism made this experience both rewarding and enjoyable.

I am equally grateful to **Dr. Deepak Garg**, Associate Professor and Head, Computer Science & Engineering Department, a nice person, an excellent teacher and a well–credited researcher, who always encouraged me to keep going with work and always advised me with his invaluable suggestions.

I will be failing in my duty if I don't express my gratitude to **Dr. S.S. Bhatia,** Senior Professor and Dean of Academic Affairs, Thapar University, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection, which made my stay at Thapar University memorable.

Last but not least, I would like to thank my family whom I dearly miss and without whose blessings none of this would have been possible. To my parents, I own thanks for their wonderful love and encouragement. I would also like to thank my brother, since he insisted that I should do so. I would also like to thank my close friends for their constant support.

Date: July, 2015
Place: Thapar University, Patiala (Kapil Yadav)

Abstract

Development of new technologies for Man Machine Interface hardware is in tandem with the corresponding advancement of software algorithms for data interpretation and processing. Technology enhancement thus saw a leap from Swept Frequency Capacitive Sensing techniques used in conventional touch screens, to interactive hardware displays like large displays, flexible displays and wearable displays for mixed reality. Current applications of MMI include smart homes, collaborative working environments, advanced information visualization and many more. The interfacing parameters range from multimodal interaction like touch, speech and gesture, to physiological factors such as ECG, EOG, Heart Rate, Eye blinking, facial expression for example. Of these gesture based HCI is popular due to two main reasons. Primarily, sensor for gesture (camera) is easily available and attachable on laptops or desktops when compared to sensors for ECG, heart rate acquisitions. Also, they are easy to use and process. The thesis discusses a gesture driven Microsoft Word Document handling. This can be applied on other applications like powerpoint or PDF reader. The main idea is to control the document from a distance without using keyboard or mouse. This can be beneficial when (i) the user is at some distance from the system (ii) hands are dirty and the user does not want to touch the system (iii) mouse is temporarily out of order (iv) simply increasing functionality of the system and providing different interfaces like voice and gesture, besides touch. The system was developed using a two state discrete temporal model for gestures which works with distinct poses. The model fuses the state information along with individual pose recognition to activate the interfacing mechanism. It can be inferred from the experimental results that the model facilitates both accurate gesture recognition as well as promptness in response. Different feature extraction techniques like gabor, wavelet,SURF are tested on the system and a decision fusion approach for these features is proposed. Generally, gesture recognition involves a huge offline training dataset to make the system robust against skin color, illumination and hand pose changes. This thesis introduces a color and shape model such that no explicit training set to train the gesture database for realtime gesture recognition is required. The response time which varies between 2 to 2.5 second can be further improved by implementing the feature detection steps in VC++ environment instead of Matlab.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since their evolvement, computing systems have proved to be vital systems of our lives. Letter typing, web surfing, gaming or data storage and retrieval, all are some very common daily use of computers. The high demand of efficient computing systems lead to the way of economical computers for general purpose. This has also led to the requirement of easy interaction facilities with the computers. HCI has hence become a progressive field for research and innovation in recent years. To utilize this new phenomenon efficiently, many studies have examined computer applications and their requirement of increased interaction. Methods of human computer interaction have evolved greatly since last decades. Starting from regular interfaces, researchers have proposed many branches where focus is on multimodality concepts rather than uni-modality, active rather than passive interfaces and adaptive intelligent interfaces rather than command/action based ones. The methods by which human has been interacting with computers has traveled a long way[2],[3]. The journey still continues and new designs of technologies and systems appear more and more every day [4].

1.1 Human-Computer Interaction

For more interactive applications, evolution of HCI have been from graphical interface paradigm to touch screen, voice based[7], gesture based interaction systems. According to key ideas of computer, machine and systems, HCI should be a design that fits between user, the machine and the required services for achieving a certain performance both in optimality and quality of the services. Most computer applications require more and more interaction[6]. Determination of quality objectives of HCI design is mostly context dependant and subjective mostly. For an instance, an aircraft part designing tool should be supporting highly precise views while design of such parts may not need such precision. How different types of HCI are being designed for the same purposes is also a major technical aspect. For example, to access computer functionalities, menus, GUI, commands and virtual reality all are being used. More detailed review on present methods and devices being used for interacting with computers is presented later. Current technologies for HCI can be categorized by the relative human sense like vision, audition, and touch [13] using which the device is designed. Common examples of sound based HCI and touch based HCI include turn-by-turn navigation commands of a GPS device and haptic devices[17] respectively.

Figure 1.1: Two surgeons check brain by hand gesture

A solution for keyboarding has been proposed by Compaq's iPAQ called, Canesta Keyboard as a vision based HCI. It is a virtual keyboard designed by projection of a QWERTY like pattern on a solid surface with usage of red light. Then, it tracks user's finger movement while typing on the surface with a motion sensor and sends the keystrokes back to the device [19]. Vision based HCI makes use of gestures as further discussed in the next section.

1.2 Gesture based HCI

Sign language development for deaf and dumb people has been a major drive for the advent of gesture based interfaces [22]. Initially, work was done to automate communication between visually impaired and deaf people[44]. This later motivated exploration of gesture based interaction, whose primary application target was gaming and home entertainment. Work done by various researchers in this field also includes gesture based robot locomotion[45],[46], intelligent wheel chair [25], surgical systems, weather channels [9],computer games [23]. G-Speak spatial computing operating system, offers movement of data between different computing systems and displays through a gesture interface[47]. A Virtual keyboard was proposed [48] which provides a detectable surface on which user can move fingers that replicates the act of key pressing. Google and Microsoft have recently patented their concepts of 3D desktop interface for their operating system where hand movements are converted into 3D graphics; while the gaming industry appears to have commercialized 3d images for interaction. Figure 1.5 gives the core application areas of gesture based interfacing.

Figure 1.2: Weather information seen through hand gesture

Figure 1.3: Body motion video games

Figure 1.4: Intelligent wheelchair

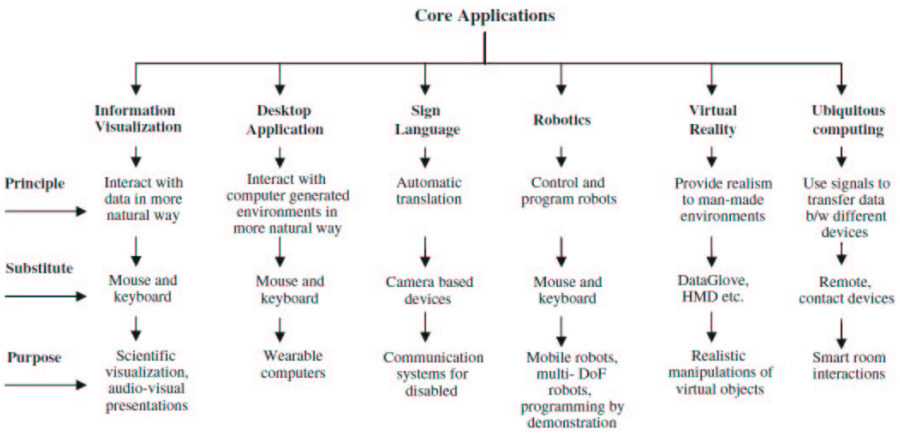


Figure 1.5: Core applications

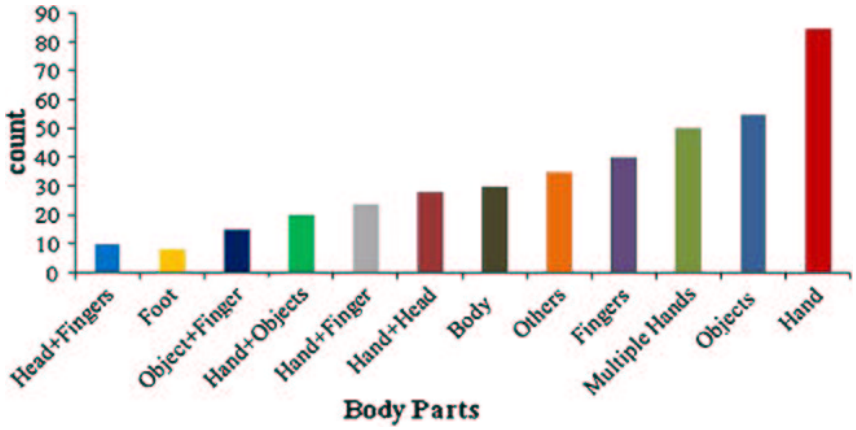


Figure 1.6: The graph shows the different body parts or objects identified in the literature employed for gesturing

1.3 Hand Gesture

The hand is extensively used for gesturing compared with other body parts because it is a natural medium for communication between humans and thus is the most suitable tool for HCI. Recently, there has been a surge in interest in recognizing human hand gestures. Hand gesture recognition has various applications like computer games, machinery control (e.g.crane), and thorough mouse replacement. One of the most structured sets of gestures belongs to sign language. In sign language, each gesture has an assigned meaning (or meanings).Computer recognition of hand gestures may provide a more natural-computer interface,allowing people to point, or rotate a CAD model by rotating their hands. Hand gestures can be classified in two categories: static and dynamic. A static gesture is a particular hand configuration and pose, represented by a single image. A dynamic gesture is a moving gesture, represented by a sequence of images.

Hand gesture analysis can be divided into glove-based analysis and vision-based analysis. The first approach employs sensors (mechanical or optical) attached to a glove that acts as transducer of finger flexion into electrical signals to determine hand posture. The relative position of the hand is determined by an additional sensor; this sensor is normally a magnetic or an acoustic sensor attached to the glove. For some dataglove application, look-up table software toolkits are provided with the glove to be used for hand posture recognition [26]. The second approach, vision-based analysis, is based on how humans perceive information about their surroundings [12] .This deals with matching a hand image in real time with a stored set of hand gestures,each of which has a specific meaning.

The thesis proposes a system which utilizes vision based hand gestures to interface Microsoft Word document in real time. The work uses gestures for opening, closing, changing font size and color, scrolling up and down in the word document. Gesture based research previously concentrated on offline gesture recognition mainly. Realtime gesture recognition

requires accurate gesture modeling techniques for prompt response to the HCI besides static gesture classification. Chapter 4 discusses the different techniques for gesture modeling and recognition which includes gesture image synchronization, hand image segmentation, feature detection, gesture classification. Related work on each of these steps are discussed in chapter 2. Chapter 5 gives the performance analysis of gesture recognition using different features and classification techniques as well as the performance of the gesture based Microsoft document handling system. Conclusion and scope of future work in this domain is presented in chapter 6.

Chapter 2

Literature Survey

This chapter discusses the different steps involved in vision based gesture recognition and reviews some of the techniques reported in literature for each of the phases. A vision based gesture recognition system consists of segmentation, feature extraction and gesture classification phases. These are further elaborated in the subsequent sections. Figure 2.1 displays some common algorithms used for segmentation, feature extraction and gesture classification. A review of some hand gesture based HCI systems are also provided towards the end of the chapter.

2.1 Hand Segmentation and Detection

Hand segmentation algorithms deal with extracting the hand region from the image acquired by the camera. Some common algorithms for segmentation are discussed in the following subsections.

2.1.1 Skin color detection

Skin color can be detected with the help of RGB color model. Skin color is mapped using an intensity range and any image is matched with this range thus extracting regions of the same color marked by the range. This generally has three problems. First illumination changes influence the intensity range. Secondly different people have different skin color thus leading to a different intensity ranges. Thirdly, if the image has other regions which are skin colored too,they will also be extracted. Red, Green, Blue (RGB) color space is the most common color space used to represent images. RGB has high correlation, non-uniformity and mixing of chrominance and luminance data. Therefore RGB is not suitable for color analysis and color based recognition[25]. To overcome this problem, normalized RGB has been introduced to obtain the chromaticity information for more robust tracking [26][27][28][29]. However, normalized RGB still suffer illuminations problems.HSV and YCbCr color spaces have the luminance and chromaticity information and is hence attractive for skin color modeling[40][41].

Figure 2.1: Common techniques involved in segmentation, feature extraction and gesture classification

2.1.2 Region growing

Region growing algorithms take one or more pixels, called seeds, and grow the regions around them based upon a certain homogeneity criteria. If the adjoining pixels are similar to the seed, they are merged with them within a single region. The process continues until all the pixels in the image are assigned to one or more regions. The focus of different region growing algorithms is on investigating how different merge criteria affect the quality of segmentation and the processing time.

2.1.3 k-Means Clustering

k-means clustering is a partitioning method. The function k-means partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. Unlike hierarchical clustering, k-means clustering operates on actual observations (rather than the larger set of dissimilarity measures), and creates a single level of clusters. The distinctions mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. Each cluster in the partition is defined by its member objects and by its centric, or center. k-means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centric, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible.

2.1.4 Thresholding

Thresholding techniques fail to segment hand regions from complex backgrounds. Other than basic thresholding algorithms, work in this area include the use of two dimensional histograms of an image. In 2D histograms, the information on point pixels as well as the local grey level average of their neighbourhood is used. The application of Fisher linear discriminant to the histogram results in an optimal projection where the data clusters are better defined and hence easier to separate by choosing appropriate thresholds.

2.1.5 Motion detection

Hand regions are often tracked across a sequence of frames in case of dynamic gestures. Assuming a static background, the hand can be extracted from each frame through frame differencing. Other tracking algorithms like kalman filter are also used in this case.

2.2 Features extraction

Feature extraction describes the relevant shape information contained in a pattern so that the task of classifying the pattern is made easy by a formal procedure. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. The main goal of feature extraction is to obtain the most relevant information from the original

data and represent that information in a lower dimensionality space. When the input data to an algorithm is too large to be processed and it is suspected to be redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Some feature extraction techniques are discussed below.

2.2.1 The scale - invariant feature transform

This algorithm smooths the sample with different scaled gaussian filters and obtains the difference of two corresponding scale. It then uses Difference of Gaussians(DoG) to find minima and maxima points. For this one pixel in the sample DoG is compared with its 8 neighbors as well as 9 pixels in the next scale and previous scales.Taylor series expansion of scale space are used to get more accurate location of extrema. Orientation Histograms from the neighborhood of the extrema points are recorded as features.

2.2.2 SURF (Speeded Up Robust Features)

SURF algorithm uses an integer approximation as the determinant of Hessian blob detector, that can be computed much faster with an integral image. Hessian matrix is computed as shown in equation 2.1 .

$$H(p,\sigma) = \begin{bmatrix} L_{xx}(p,\sigma) & L_(xy)(p,\sigma) \\ L_{xy}(p,\sigma) & L_(yy)(p,\sigma) \end{bmatrix} \tag{2.1}$$

Where $L_{xx}(p,\sigma)$ is image convolution of second derivative $\frac{dx}{dx^2}g(\sigma)$

2.2.3 Wavelet transforms

Wavelet transforms have become one of the most important and powerful tool of signal representation. It is used in image processing for data compression, fine feature extraction. Wavelet transforms are based on small waves, called wavelets. Input images are decomposed through the Wavelet Packet Decomposition using the wavelet basis function to get detailed and approximate coefficients. The Approximation (A) gives the lowpass image whereas three detail coefficients Horizintal (H), Vertical (V) and the Diagonal (D) as shown in Figure 2.2 are obtained.

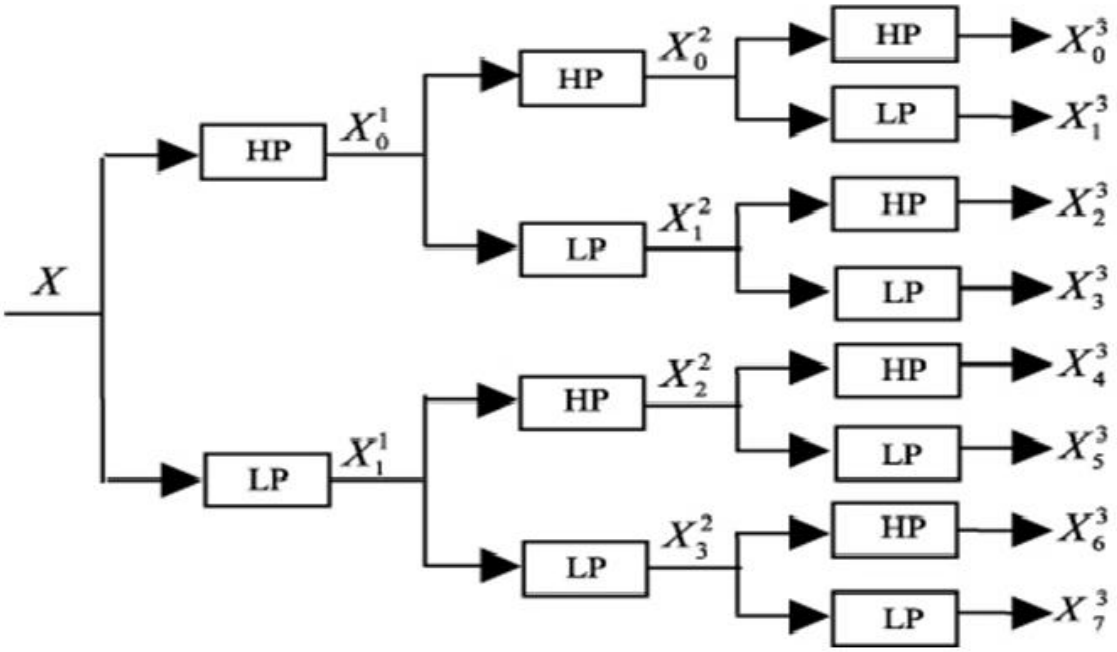


Figure 2.2: Schematic diagram of 2D wavelet transform

$$wavelet f(\tau) = f_{approx}(\tau) = region(\rho) g(2\tau - \rho) \tag{2.2}$$

$$wavelet f(\tau) = f_{detail}(\tau) = region(\rho) h(2\tau - \rho) \tag{2.3}$$

$$f(\tau) = [f_{approx}(\tau) f_{detail}(\tau)] \tag{2.4}$$

where$\rho \rightarrow polar(x, y), g \rightarrow highpassfilter, h \rightarrow lowpassfilter$

2.2.4 Gabor filter

A Gabor filter, named after Dennis Gabor, is a linear filter. Gabor filter gives pertinent feature information and is ideal for texture representation. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The 2D Gabor filter is applied on a sample image with different scales and orientations. The real and the imaginary parts are convoluted to form the final filtered image. The Gabor filtered image is a result of the original image convoluted using the gabor kernel. Usually the gabor kernel $\psi_{\gamma,v}$ is constructed using five scales $v = 0, 1, 2, 3, 4$ and eight orientations $\gamma = 0, 1, ..7$ hence resulting in a very high dimensional representation.

$$\psi_{\gamma,v}(z) = \frac{\| \kappa_{\gamma,v} \|^2}{\delta^2} e^{-\frac{\|\kappa_{\gamma,v}\|^2 \|z\|^2}{2\delta^2}} [e^{\iota \kappa_{\gamma,v} z} - e^{-\frac{\delta^2}{2}}] \tag{2.5}$$

where $\kappa_{\gamma,\upsilon} = \kappa_\upsilon e^{\iota\phi_\gamma}$, $\delta = 2\pi$, $\kappa_\upsilon = \frac{\kappa_{max}}{f^\upsilon}$, $f = \sqrt{2}$, $\kappa_{max} = \frac{\pi}{2}$,$\phi_\gamma = \frac{\pi\gamma}{8}$, $z = (r, c)$ and $\| \|$ denotes the norm operation.

2.3 Hand gesture Classification

Classification methods referred here are either distance classifiers like euclidean and mahalanobis or learning algorithms like neural network and support vector machines. These are further elaborated in the following subsections.

2.3.1 Distance Metrics

Distance metrics is used as a classifier to choose the best class given n classes and a test sample. Two common distance metrics are discussed below.

Euclidean Distance

The Euclidean distance gives the straight line distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm. If the two pixels that we are considering have coordinates (x_1, y_1) and (x_2, y_2) then the Euclidean distance is given by:

$$D_{Euclid} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2.6}$$

Mahalanobis distance

The Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. This distance is zero if P is at the mean of D, and grows as P moves away from the mean. Along each principal component axis, it measures the number of standard deviations from P to the mean of D. If each of these axes is rescaled to have unit variance, then Mahalanobis distance corresponds to standard Euclidean distance in the transformed space. Mahalanobis distance is thus unitless and scale-invariant, and takes into account the correlations of the data set. The Mahalanobis distance of an observation $x = (x_1, x_2, x_3, x_4, ...x_N)^T$ and covariance matrix S is defined as

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \tag{2.7}$$

Mahalanobis distance (or "generalized squared interpoint distance" for its squared value can also be defined as a dissimilarity measure between two random vectors $\overrightarrow{x}$ and $\overrightarrow{y}$ of the same distribution with the covariance matrix S. If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance. If the covariance matrix is diagonal, then the resulting distance measure is called a normalized Euclidean distance.

$$d(\overrightarrow{x}, \overrightarrow{y}) = \sqrt{(\overrightarrow{x} - \overrightarrow{y})^T S^{-1} (\overrightarrow{x} - \overrightarrow{y})} \tag{2.8}$$

where s_i is the standard deviation of the x_i andy_i over the sample set.

2.3.2 Support vector machine

SVM is a non-linear analysis first proposed by Burges in 1998. Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane separates between object sets having different class memberships. One of the bottleneck of the SVM is the many number of support vectors used for the training set to perform classification (regression). Given a set of instruction examples, each marked as to belong to one of the two category, an SVM training algorithm check a model that assigns new examples into one category or the other. An SVM model is a demonstration of the examples as points in space, a map so that the examples of the separate categories are divided by a clear gap that is as wide as probable.

2.3.3 Adaboost Algorithm

Adaboost, that stands for adaptive boosting is a machine learning algorithm that is used for finding strong classifier from weighted sum of weak classifiers. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favour of those instances misclassified by previous classifiers, and there weight will be increased so that next classifier concentrates more on those instances. Weak classifiers are rectangle features. Adaboost assign each of weak classifier weight to make a strong classifier.

2.3.4 Neural Network

Neural networks are composed of simple elements operating in parallel. Neural networks are models that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected neurons that can compute values from inputs by feeding information through the network. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are used, in this supervised learning, to train a network. Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech and vision and control systems.

There are two modes of learning: Supervised and unsupervised. In Supervised learning, the weights of the network are learned given a set of input and output through a number of iterations. The data flow starts from the input to the next hidden layer neurons and then the output layer. The error is obtained from the predicted and observed output difference, and weights between the neurons in each layer is corrected till a fixed number of iterations or error threshold. Neural networks which use unsupervised learning are most effective for describing data rather than predicting it. Here no output is given to the system during training. It basically works as a classifier where the input data is grouped into distinct separable classes.

2.4 Related work on Hand Gesture based HCI

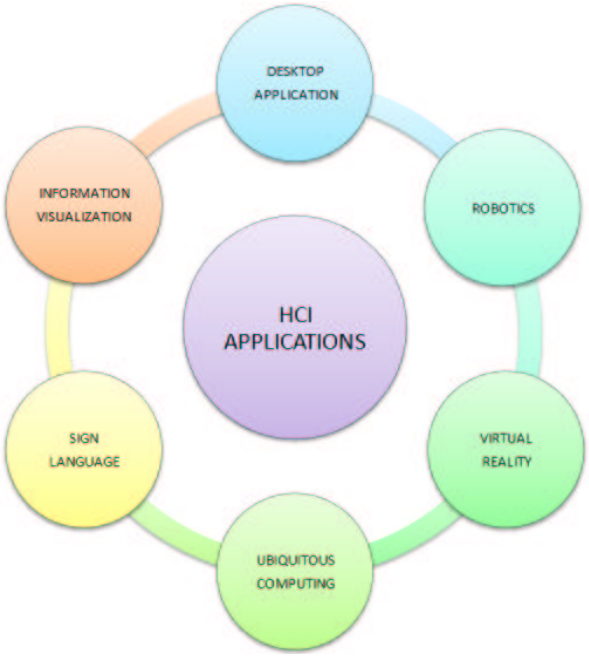


Figure 2.3: Hand Gesture based HCI

1. Hand gestures have been a substitute of mouse and keyboard for many desktop applications . Some of these include stroke gestures for marking menu selection, circular motion gestures acting like radial scrolling effect for navigating through documents [32] .

2. Virtual reality interactions use hand gestures to enable realistic manipulations of virtual objects using ones hands, for 3D display interactions . Non-immersed interactions, semi-immersive interactions as well as fully-immersive interactions with the virtual world was presented here. Non-immersed interaction involved handling virtual objects without being a part of the scene, where as in semi -immersed interaction the user is represented in the scene as an avatar. Fully immersed interaction enables navigating around a 3D information space using a stereoscopic display [33].

3. A fully-immersed virtual reality based robot control was presented where the robot environment is provided to the user through video feed. The user movement in replicated to the robot which moves and actually grabs the objects the user is virtually touching. [34].

4. Users can control smart home appliances such as TV, fan, lighting, doors and change channels, temperature, and volume by just hand gestures using a hand gesture interface

system via a depth imaging sensor. Depth images are recognized using random forests (RFs) classifiers and mapped to generate control commands for the appliance control interface[36].

5. A vision based hand gesture interface for robot control using a fixed set of manual commands and a reasonably structured environment was presented [37].

6. A human-robot communication by bare hand dynamic gestures was also proposed using a Bag-of-Features and three dimensional histograms of a gradient orientation as features and SVM as classifier [38].

7. Gesture based interaction with a mobile device was proposed by Prasuhn, Lukas, et al [39].

8. Siddharth S. et al. presented a hand gesture HCI system for controlling virtual games, browsing images [30].Hand gestures are tracked using camshift and Haar features.

9. A multi-gesture based interaction for 3-D File System Navigation and "Midgard" Geographic Information Visualization was presented by Michael J. Reale et al [31]. In this research the current directory is represented as a large 3-D sphere, subdirectories are represented as smaller orbs and files as cubes. Opening a subdirectory is analogous to zooming into this smaller orb. For both the applications, head pose controls the rotation of the folders in the sphere or map for GIV, head position allows the user to zoom out or out , mouth opening switches modes, the hand controls the primary cursor, and the eyes control the secondary cursor.

10. Hand gestures are also used to play computer games. Here the players hand and body position is tracked to control objects such as car[35].

Chapter 3

Problem Statement

With the development of present computing, current user interaction approaches with keyboard, mouse and pen are not sufficient. Due to the limitation of these devices the usable command set is also limited. Direct use of hands can be used as an input device for providing natural interaction. A thorough literature review suggests the use of touch, speech, gesture based applications like smart homes, collaborative working environments, advanced information visualization, gaming systems, virtual keyboard and many more. A lot of gesture based applications for desktop computing systems, mobile devices are reported. As an extension to these, this thesis aims to present a gesture based HCI system for interfacing Microsoft Word document. This involves two main operations:

- synchronization between, gesture recognition and action invocation module
- accuracy of gesture recognition.

The former deals with interpretation of gesture forms and flows. A form gesture may have a distinct path without and distinct pose or a distinct pose without any distinct path. Gesture flows are discrete or continuous depending on whether the action is invoked at the end of the gesture or during the gesture. The second operation stems down to any gesture recognition problem where the main criteria is using robust features, efficient learning algorithms, and huge training set so that a high accuracy performance is achieved subject to varying illumination, hand orientations and skin color. The main contributions of this work are:

- a two state discrete distinct pose based model for gestures
- a robust feature set for unique gesture representation.
- a decision fusion approach for gesture classification which facilitates decision making using a multimodal feature set.
- a color and shape model for handling real time skin color, illumination and hand pose orientation variations, hence cutting down the requirement for any explicit training set to train the gesture database for realtime gesture recognition .

Chapter 4

Methodology

This chapter discusses the gesture based HCI system in detail. The system consists of a camera which will acquire images in real time. The software consists of three modules (as seen in figure 4.1). (i) A feature database which was trained offline (ii) A module called synchronize which detects gestures from captured images and matches it with the database for a valid gesture.(iii) The third module actor performs the desired event based on the matched gesture. The feature database generation and synchronization is done using Matlab. The actor needs to interface Microsoft OLE for the corresponding events to be handled in the word document. Document opening, closing, scroll up, scroll down, font color and size are the different events handled by the interface currently. The entire work can be divided in two phases (as seen in Figure4.2). In the first phase, a set of gesture images are trained so that each gesture corresponds to a particular pattern. This includes hand region segmentation from an image, extracting features from the segmented region, using the feature vector for training. In the second phase a hand gesture has to be mapped to an event in realtime and the corresponding event needs to be performed. Three kinds of feature extraction and classification is discussed here. These are marked as FECI,FECII and FECIII. These are elaborated in the sections below. Later in the chapter, the shape and color model used to escape large training dataset is elaborated.

- SURF, Wavelet, features and their Canonically correlated feature with neural network
- Decision fusion of SURF & Wavelet features using euclidean distance metrics
- Wavelet-Gabor, Gabor-PCA and Wavelet-Gabor-PCA features with euclidean classifiers

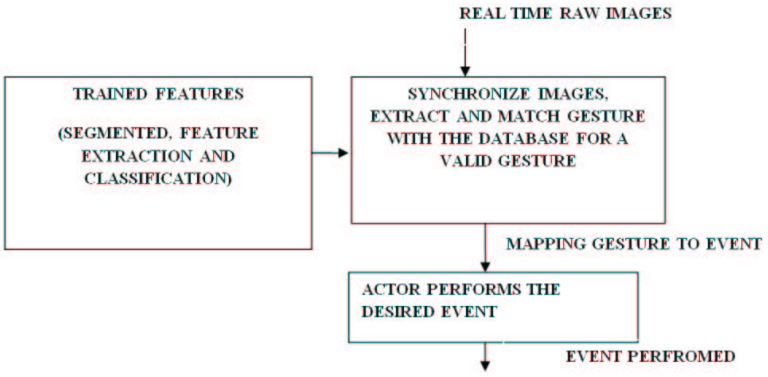


Figure 4.1: Gesture based HCI system for Word Document Handling

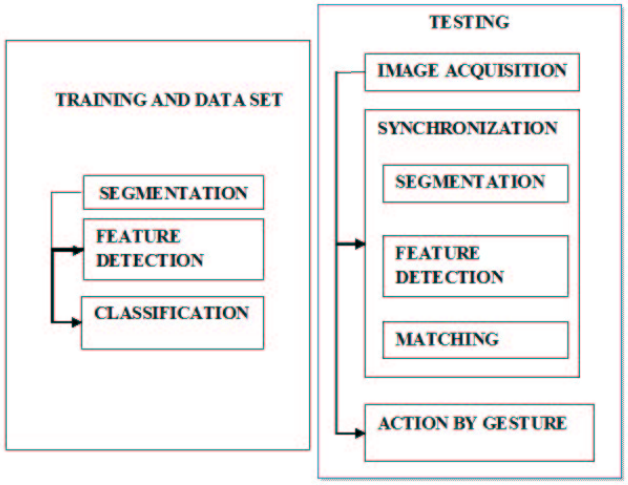


Figure 4.2: Flow Chart Of The Proposed Gesture Recognition System

4.1 Image sequence by camera and Acquizition

Images in realtime are acquired using the webcam provided in the laptop. The image acquisition program is looped to continuously capture snapshots every second. Any variations in gesture representation due to snapshot intervals are handled by the synchronization module.

4.2 Segmentation

In this phase, hand regions need to be extracted from the background for feature extraction and recognition purposes. The performance of the feature extractor largely varies on the segmentation algorithm selected. The task becomes challenging with cluttered background. Also, the algorithm should be robust against scene illumination and skin variations. Background modelling with K Gaussian distributions [51], connected-component labelling algorithm [52], Automatic Seeded Region Growing Algorithm (ASRGC) [54] with YCbCr model, meanshift filters are some of the mostly used segmentation algorithms. K-Means algorithm is used for segmentation in the present work. Given a set of observations $(y_1, y_2, , y_n)$, where each observation is a 3-dimensional real vector, k-means clustering aims to partition the n observations into $k(\leq n)$sets $v = v_1, v_2, , v_k$ so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find σ.

$$\sigma = \sum_{i=1}^{k} \sum_{y \varepsilon v_i} \|y - \mu_i\|^2 \tag{4.1}$$

where μ_i is the mean of points in V_i.

(a) Hand image (b) Clustered image (c) Returned skin color

Figure 4.3: k-means on RGB color image for skin color segmentation

4.3 Features extraction and Classification

Previous work on Gesture recognition has used shape based , keypoint based as well as region based algorithms. Shape based algorithms used different shape signatures (fourier descriptors for example) to detect hand shape while SIFT [53] is an example of keypoint based techniques. Region based algorithms like wavelet descriptors, PCBR are also used by many researchers.
Two types of classification are generally used by recognition applications. One category uses various distance metrics like Euclidean and Mahalanobis to compute the difference between the test vector with the different classes of vectors and assigns the test vector to the class having the least distance. Another category uses machine learning algorithms like SVM,

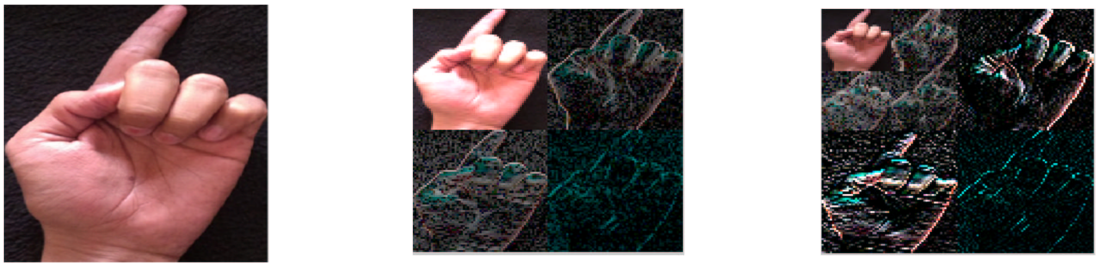
NN, AdaBoost, HMM to classify the data. The feature extraction and classification used in this work is further elaborated.

4.3.1 FECI

Here Surf (figure 4.4) and Wavelet features (figure 4.5) are separately tested with neural network. Also, the SURF and Wavelet features are canonically correlated (CC) and tested with neural network(equation 4.3).

(a) Segmented image (b) rgb to gray image (c) SURF FEATURE

Figure 4.4: SURF feature on segmented image



(a)) Segmented image (b)) Wavelet transforms for decomposition of image (c) 2nd level decomposition

Figure 4.5: 2D wavelet transform

CC creates a new feature vector for each set of SURF and wavelet vectors such that the correlation between these variables is maximized and independent of affine transformation. Equation 4.2 gives the canonically correlated variable Z.

$$Z_i = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}^T \begin{bmatrix} X_i \\ Y_i \end{bmatrix}$$

$$C_x y = \frac{1}{L} \sum_{i=1}^{L} x_i y_i^t \tag{4.2}$$

Where x_i and y_i denote the SURF and wavelet feature vector of the i^{th} image respectively. A and B are the eigenvectors of $C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{xy}^T$ and $C_{xx}^{-T} C_{xy} C_{yy}^{-1} C_{xy}$ respectively. C_{xy} gives the covariance matrix of x and y and L denotes the total number of training images.

$$x = [-1, x_1, x_2, x_3, x_4, ...x_p]^T$$
$$\text{weights } w = [-1, w_1, w_2, w_3, w_4, ...w_p]^T$$
$$\text{output } odx = \sum_{i=0}^{P} x_i * W_i$$
$$\text{error } e_x = ox - odx \text{ odx is the actual output}$$
$$\text{updated weights } \Delta w_i = -\rho * \frac{\delta e}{\delta W_i} \tag{4.3}$$

Table 4.1 gives the parameters for neural network performance.

Parameters	Values
Performance	91.54
Training	0.9871
Validation	0.9218
Stop error	0.01
Learning rate	.09

Table 4.1: Neural Network Performance Parameters

4.3.2 FECII

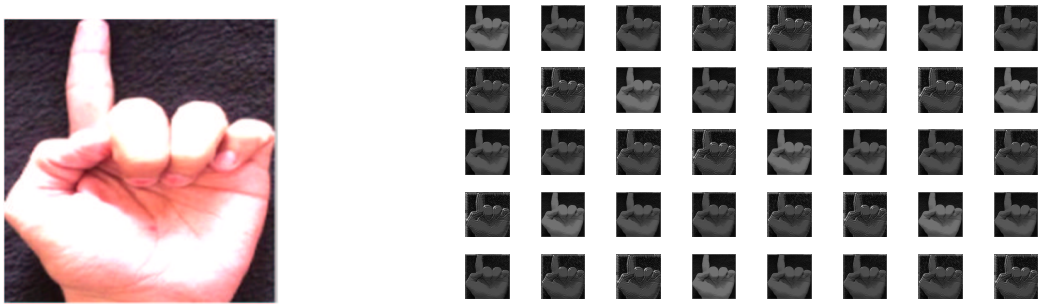
Given any test image, SURF and Wavelet features are computed and euclidean distance between each training SURF and wavelet vector is computed respectively. Decision fusion is carried out using equation 4.4.

$$d_i^{comb} = \frac{1}{d_i^y} + \frac{1}{d_i^x} \tag{4.4}$$

where d_i^x is the euclidean distance computed for all image vectors x_i and test vector x^t. The i for which d^{comb} is the maximum is considered as the correct match.

4.3.3 FECIII

Three types of feature vectors are considered here for classification using euclidean distance.
(a) PCA applied with Gabor
(b) Gabor applied on low pass coefficients of wavelet image
(c) PCA applied on (b)



(a) Hand image

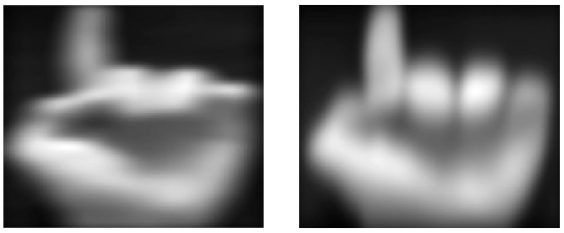
(b) Gabor images

Figure 4.6: Gabor filter is applied on the image with 5 scales and 8 orientations.

(a) Original image

(b) Wavelet low image with low pass component highlighted

(c) Gabor filtered images with different theta orientation



(d) Gabor filtered images with different theta orientation

(e) Gabor filtered images with different theta orientation

Figure 4.7: Lowpass DWT coefficients for applying gabor filter

Principal Components Analysis (PCA) is utilized here for dimension reduction as well as improved feature set in a eigen space. PCA is applied over the training data set of each feature vector where rows correspond to number of images and columns correspond to size of each vector. Each feature $\mathbf{x}_i$ in the feature vector is further transformed in the eigen space using equations 4.5 to 4.9. The mean vector of each feature is calculated using equation 4.5.

$$\mu_i = \sum_{j=1}^{m} x_i^j \tag{4.5}$$

The covariance $\mathbf{C}_i$ of the feature i in the training set is calculated using the mean adjusted data φ_i

$$\varphi_i = \mathbf{x}_i - \mu_i \tag{4.6}$$

$$\mathbf{C}_i = \frac{1}{m} \varphi_i^T . \varphi_i \tag{4.7}$$

The feature is then transformed to the eigen space using the eigen vectors ω_i of the covariance matrix $\mathbf{C}_i$ as shown in equation 4.8.

$$\Omega_i = \omega_i^T . \varphi_i^T \tag{4.8}$$

For any test image the transformed vector in the eigen space is calculated from the feature $\hat{\mathbf{x}}_i$ using equation 4.9.

$$\widehat{\Omega}_i = \omega_i^T . \widehat{\varphi}_i^T \tag{4.9}$$

where

$$\widehat{\varphi}_i = \hat{\mathbf{x}}_i - \mu_i \tag{4.10}$$

The distance matrix is calculated by finding the distance between each training image principal component vector Ω_i^j and test image principal component vector $\widehat{\Omega}_i$.

4.4 Synchronization

A software-based system for the real-time synchronization of images captured by a lowcost camera framework is presented (as seen in figure 4.8). It is highly recommended for cases where special hardware cannot be used. Every gesture is identified in two steps. A start symbol denotes the start of the gesture, which is followed by appropriate document open, close or other gestures. As shown in the figure 4.9, interim invalid gestures may be captured in snapshots while the user is forming the particular gesture. All these gestures will not find any match in the database, however the recognition module will run unnecessarily wasting processor time. Hence a frame is passed for recognition only after it becomes static. Thus only when last three captured frames have no change, it is forwarded for segmentation, feature detection and gesture recognition. The start state is maintained as long as a valid gesture is not recognized. Once a gesture is recognized and the corresponding event is invoked, the cycle is complete. The next event will be marked by another start state.

4.5 Document Handling

The work uses existing MS Office Automation (OLE) modules and interfaces it with corresponding gesture recognition events. This is shown in figure 4.10 and table4.2. For example when a gesture denoting document open OLE is invoked, it is recognized as 001 which in turn calls the corresponding module to open the document, and the document is opened. Same is the case for the font color, size or scroll gestures.

GESTURE	RECOGNITION	FUNCTION INVOKE	Action perform
1	001	$CMSWord :: OpenDocument$	Microsoft word open
2	010	$SpinButton1_SspinDown()$	Microsoft word scroll up
3	011	$SpinButton1_SspinUp()$	Microsoft word scroll down
4	100	$CMSWord :: SetFont$	Microsoft word font change
5	101	$CMSWord :: CloseDocuments()$	Microsoft word close

Table 4.2: The table shows the output for each recognized gesture and the corresponding function it invokes

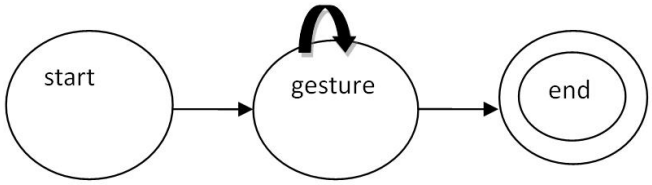


Figure 4.8: A start gesture and another corresponding gesture is required for an event invocation.

Figure 4.9: Transition between 2 gestures captured by the camera.

4.6 Color Model

This model handles skin color variations and illumination changes between the stored image and any test image in the following manner. The test input is tuned according to the stored lighting by synchronizing the color component of the training image with the testing image. Color synchronization involves modifying the color component obtained using the fundamental frequency component of an image.

The hand region in the single training image is manually extracted and represented as

$$M = (\mu_c, \sigma_c) \tag{4.11}$$

Where μ_c and σ_c denote the mean and variance of the intensity segment of each color channel $L_c - H_c$, $C\varepsilon$ red ,green and blue. $L_c - H_c$ encompasses the entire range of the model excluding outliers. In the present case outliers are selected as intensities with probabilities less than a threshold t, selected experimentally. Regions in the test image are extracted based on the model M. Pixels which have intensities in the range $\mu_c \pm \pm d\sigma_c$, where d is taken as 2 in the current results.

(a) Start gesture 000

(b) Document open 001

(c) Document scroll up 011

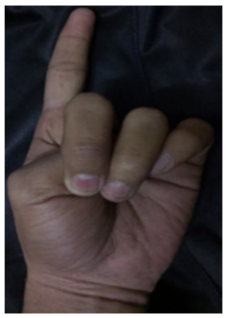
(d) Document scroll down 010

(e) Document font change 100

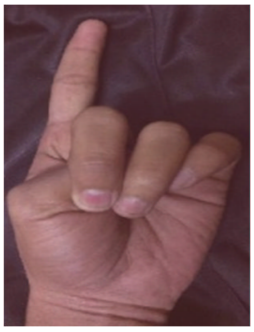
(f) Document close 101

Figure 4.10: The different gestures used for particular events

(a) Sample training image



(b) Sample test image



(c) Modified test image

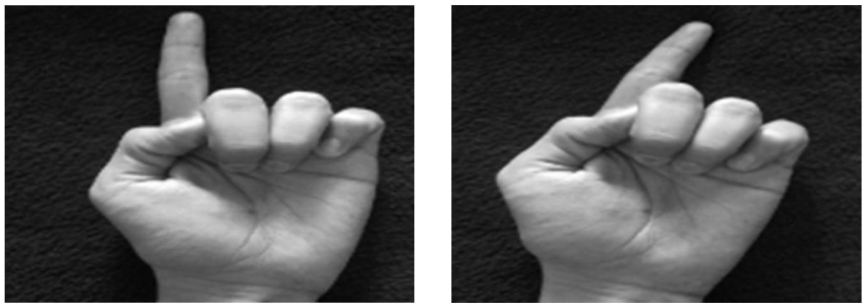
Figure 4.11: Skin color variations and illumination

(a) Testing rgb image (b) Segmented image (c) Returned skin color

Figure 4.12: RGB color model for segmentation.

4.7 Shape Model

Each testing image is reconstructed based on the single training image before feature vector generation. SURF features are used for feature matching and registration of test image. This is further shown in figure4.13



(a) Stored image

(b) Testing image 1

(c) Testing image 2

(d) Reconstruction image of (b)

(e) Reconstruction image of (c)

Figure 4.13: SURF features are used for feature matching and registration of test image

The shape model is implemented using a look up module which has three entries per gesture, with a total of 6 gestures G1,G2,...G6. A surf vector, a wavelet-gabor vector, and the pattern denoted by the gesture. Any test gesture is first segmented using the color model. It is then registered individually with all the six gestures using surf features generating 6 registered images R1,R2,...R6. Wavelet-gabor vectors are created for all the six registered images and a one to one matching is done, i.e. wavelet-gabor vector of L1 is matched with G1,L2 with G2 and so on. The best match gives the matching gesture.

Chapter 5

Experimental Results

The system is tested in realtime using FECI,FECII,FECIII and Shape model and the results obtained are illustrated.

5.1 Results with FECI and FECII

A dataset of 300 images are considered for training. Three types of sample images as shown in figure5.1, are considered for experimental purpose for all the methods and procedures explained in this paper namely clear background (C1) , slightly cluttered background (C2), and slightly cluttered background with changing lighting conditions (C3). Some of the images are captured in home environment without any special lighting using a consumer quality web camera. The resolution of the images considered for processing after segmentation is 128 X 128.

(a) 20 gesture clear background (C1)

(b) 20 gesture clear background (C2)

(c) 20 gesture clear background (C3)

Figure 5.1: Training images with different skin color, illumination, background and hand angle

Figure 5.2: Sample test Images used for testing the system in realtime.

The accuracy and performance of the proposed system was further verified using realtime test cases. Hundred gestures (some of them are shown in figure5.2) , randomly selected and performed, by different people at different illumination and backgrounds were used for generating the test results shown in table5.1 and 5.2. Table 5.2 gives a comparison of different feature vectors used. The technique which gives the best performance (decision fusion approach in the present case) is further used for word document interfacing. The time and performance accuracy for each event is shown in table5.2.

No	Method	Percentage
1	wavelet transformation	93
2	Speeded Up Ro- bust Features	90.5
3	feature fusion	87
4	decision fusion	96

Table 5.1: Performance comparison of individual methods for correct gesture recognition.

It was seen that the decision fusion approach with wavelet and SURF features gives a satisfactory performance on the current dataset

Gesture Action	perform	Percentage	Time in second
a	Microsoft word open	95	4.5
b	Microsoft word scroll up	92	4
c	Microsoft word scroll down	91	3.5
d	Microsoft word font change	93	3
e	Microsoft word close	90	3.25
f	Start gesture	98	2

Table 5.2: Performance of event invocation on gesture recognition, using decision fusion of SURF and Wavelet feature

5.2 Results with FECIII

The same dataset used for FECI and FECII was considered. The vector sizes used were 160 X 160. The accuracy and performance of the proposed system using different feature

vectors was verified using 100 realtime test cases as shown in figure 5.2). Table 5.3 gives a comparison of different feature vectors used. The technique which gives the best performance (wavelet-gabor- pca) is further used for application invocation as shown in table 5.4.

Method	Recognition Rate
Wavelet Gabor	97.2
Gabor Pca	94.5
Wavelet Gabor Pca	98.3

Table 5.3: Comparison of gesture recognition using different feature detection methods

Gesture Action	perform	Percentage	Time in second
a	Microsoft word open	95	4
b	Microsoft word scroll up	92	4
c	Microsoft word scroll down	91	4.5
d	Microsoft word font change	93	5
e	Microsoft word close	90	3
f	Start gesture	98	2

Table 5.4: Performance of event invocation on gesture recognition, using wavelet-gabor-pca feature and distance classifier

5.3 Results for Shape-Color Model

The system is tested with 100 test samples. Hundred gestures randomly selected and performed, by different people at different illumination, poses and backgrounds were used for generating the test results shown in table5.5. The time and performance accuracy for each event is shown here. It is observed that the system is robust even without a training set .

Gesture Action	perform	Percentage	Time in second
a	Microsoft word open	95	2
b	Microsoft word scroll up	92	2
c	Microsoft word scroll down	91	2.5
d	Microsoft word font change	93	2.5
e	Microsoft word close	90	2
f	Start gesture	98	1

Table 5.5: Performance of event invocation on gesture recognition using color-shape model,wavelet-gabor feature,distance classifier

Chapter 6

Conclusion And Future Work

The thesis proposes a gesture based Microsoft Document handling system which operates using a two state gesture model. The performance of the system depends on the gesture synchronization and the gesture recognition algorithms. The former have been handled by a temporal modelling of gestures. This can further be improved by combining temporal model along with form and path. The gesture recognition shows a performance collation between SURF and wavelet transformation along with neural network, feature fusion and decision fusion approaches. The system has also been tested with a combination of different scale space features like gabor and wavelet. Combinations of PCA over wavelet images, gabor feature on wavelet images, PCA on wavelet gabor have been evaluated with distance classifiers. It has been observed that the decision fusion of Wavelet and SURF features provide the best results with the test set used. The number of features chosen for decision or feature fusion can be increased based on the available hardware for implementation. It was limited to two in the present work, implemented, on a laptop with Intel core i3 processor (CPU 2.27GHz) on a 64 bit windows platform. The use of color and shape model instead of a large training set also shows a stable performance. This facilitates two factors:- Primarily the space requirements of the system to save huge offline databases is excluded. Secondly, any new gesture can be added to the system without re-training of the former model. The color and shape model hence satisfactorily handles illumination and pose orientation changes. Wavelet-Gabor features, chosen for feature extraction in this case helps in extracting pertinent feature information.
A number of extensions of the current work can be pursued:

- The work can be implemented on other softwares like Acrobat Reader, PowerPoint,Excel for example
- Two dimensional image information processed by the recognizer can be extended to include the depth dimension. This can help estimate user distance from the screen. Thus there will be an automatic zoom in or out depending on this distance
- Dynamic gestures can be considered instead of static ones. As a result the same gesture will work for scroll up or down depending whether the hand is moving upwards or downwards

Bibliography

[1] Just. A (2006) Two-handed gestures for human-computer interaction. Research report IDIAP 06-73, EPFL

[2] Hasan H, Abdul Kareem S (2012) Static hand gesture recognition using neural networks. Artif Intell Rev. doi:10.1007/s10462-011- 9303-1

[3] Vsrkonyi-Kczy AR, Tusor B (2011) Human-computer interaction for smart environment applications using fuzzy hand posture and gesture models. IEEE Trans Instrum Meas 60(5):1505-1514

[4] Karam M (2006) A framework for research and design of gesturebased human-computer interactions. PhD Thesis, University of Southampton

[5] Symeonidis K (1996) Hand gesture recognition using neural networks. Neural Netw 13:1-5

[6] Derpanis KG (2004) A review of vision- based hand gestures. http://cvr.yorku.ca/members/gradstudents/kosta/publications/file Gesturereview.pdf

[7] Rautaray, Siddharth S., and Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey." Artificial Intelligence Review 43.1 (2015): 1-54.

[8] K.l.boz, Nurettin a.r, and U?ur Gdkbay. "A hand gesture recognition technique for human-computer interaction." Journal of Visual Communication and Image Representation 28 (2015): 97-104.

[9] Callies, Rodolphe, Burkhard Claus Wnsche, S. Marks, Christof Lutteroth, and Lindsay Alexander Shaw. "Challenges in virtual reality exergame design." (2015).

[10] Ogunyemi, Abiodun, and David Lamas. "Interplay between human-computer interaction and software engineering." Information Systems and Technologies (CISTI), 2014 9th Iberian Conference on. IEEE, 2014.

[11] Sinha, Subarna, et al. "Analysis of a new paradigm for depth vision application in augmented HCI." Communications and Signal Processing (ICCSP), 2014 International Conference on. IEEE, 2014.

[12] Bhaltilak, Kavita V, Harleen Kaur, and Cherry Khosla. "Human Motion Analysis with the Help of Video Surveillance A Review" International Journal of Computer Science and Information Technologies 5.5 (2014).

[13] Navarra, Jordi, Salvador Soto-Faraco, and Charles Spence. "Discriminating speech rhythms in audition, vision, and touch." Acta psychologica 151 (2014): 197-205.

[14] Khosravy, Moe, Lev Novik, and Darryl E. Rubin. "Mobile computing devices, architecture and user interfaces based on dynamic direction information." U.S. Patent No. 8,700,301. 15 Apr. 2014.

[15] Harish, R., et al. "Human computer interaction-A brief study." International Journal of Managment, IT and Engineering 3.7 (2013): 390-401.

[16] Ramakrishnan, S., and Ibrahiem MM El Emary. "Speech emotion recognition approaches in human computer interaction." Telecommunication Systems 52.3 (2013): 1467-1478.

[17] Gedawy, Hend K. Designing an Interface and Path Translator for a Smart Phone-Based Indoor Navigation System for Visually Impaired Users. Diss. Qatar Foundation, 2011.

[18] Hale, Kelly S., and Kay M. Stanney, eds. Handbook of virtual environments: Design, implementation, and applications. CRC Press, 2014.

[19] Bretz E.A.(2002), "When work is fun and games", IEEE Spectrum, 39(12), pp 50-50

[20] Te'eni Dov , Carey , Jane M. , Zhang Ping (2007), ?Human computer interaction: developing effective organizational information systems", University of British Columbia A Balanced Look at HCI in Business

[21] Pavlovic, Vladimir I., Rajeev Sharma, and Thomas S. Huang. "Visual interpretation of hand gestures for human-computer interaction: A review."Pattern Analysis and Machine Intelligence, IEEE Transactions on 19.7 (1997): 677-695.

[22] Ghotkar, Archana S., and Gajanan K. Kharate. "Study Of Vision Based Hand Gesture Recognition Using Indian Sign Language." Computer 55 (2014): 56.

[23] Grau, Oliver, and Thomas Veigl, eds. Imagery in the 21st Century. Mit Press, 2011.

[24] Soussi, Iheb, et al. "Expert system for the decision on the ability to drive power wheelchair based on fuzzy logic." Electrical Engineering and Software Applications (ICEESA), 2013 International Conference on. IEEE, 2013.

[25] Y. Zhang. Image Engineering (III): Image Understanding, 2nd edition, Beijing, Tsinghua University Press, 104-109, 2007.

[26] Beetz, M., B. Radig, and M. Wimmer. A person and context specific approach for skin color classification, 18th International Conference on Pattern Recognition, 2006 (ICPR 2006). 2006. Hong Kong.

[27] Soriano, M., et al., Skin detection in video under changing illumination conditions, 15th International Conference on Pattern Recognition 2000. 2000. Barcelona.

[28] Kawato, S. and J. Ohya., Automatic skin-color distribution extraction for face detection and tracking, 5th International Conference on Signal Processing Proceedings 2000 (WCCC-ICSP 2000). 2000. Beijing.

[29] Park, J., et al., Detection of human faces using skin color and eyes, 2000 IEEE International Conference on Multimedia and Expo 2000 (ICME 2000). 2000. New York, NY

[30] Rautaray SS, AgrawalA,Anovel humancomputer interface based on hand gesture recognition using computer vision techniques.International conference on intelligent interactive technologies and multimedia (IITM-2011), pp 292296

[31] Reale MJ, Canavan S, Yin L, Hu K, Hung T , A multigesture interaction system using a 3-D Iris disk model for gaze estimation and an active appearance model for 3-D hand pointing,IEEE Trans Multimed 13(3):474486

[32] Lenman S, Bretzner L, Thuresson B (2002) Using marking menus to develop command sets for computer vision based hand gesture interfaces. In: Proceedings of the second Nordic conference on humancomputer interaction, ACM Press, pp 239242

[33] Osawa N, Asai K, Sugimoto YY (2000) Immersive graph navigation using direct manipulation and gestures. In: ACM symposium on virtual reality software and technology. ACM Press, pp 147152

[34] Goza SM, Ambrose RO, Diftler MA, Spain IM (2004) Telepresence control of the nasa/darpa robonaut on a mobility platform. In: Conference on human factors in computing systems. ACM Press, pp 623629

[35] William T. Freeman, P. A. Beardsley, H. Kage, K. Tanaka, K. Kyuma, C. D. Weissman,Computer vision for computer interaction,SIGGRAPH Computer Graphics magazine, November 1999

[36] Dinh, Dong-Luong, Jeong Tai Kim, and Tae-Seong Kim. "Hand Gesture Recognition and Interface via a Depth Imaging Sensor for Smart Home Appliances." Energy Procedia 62 (2014): 576-582.

[37] Malima, Asanterabi, Erol Ozgur, and Mjdat etin. "A fast algorithm for vision-based hand gesture recognition for robot control." Signal Processing and Communications Applications, 2006 IEEE 14th. IEEE, 2006.

[38] Abid, Muhammad R., et al. "Dynamic hand gesture recognition for human-robot and inter-robot communication." Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2014 IEEE International Conference on. IEEE, 2014.

[39] Prasuhn, Lukas, et al. "A HOG-based hand gesture recognition system on a mobile device." Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014

[40] Phung, S. L., Bouzerdoum, A., and Chai, D., A novel skin color model in YCbCr color space and its application to human face Detection, IEEE International Conference on Image Processing (ICIP2002), vol. 1, 289292. 2002.

[41] Zarit, B. D., Super, B. J., and Quek, F. K. H., Comparison of five color models in skin pixel classification, ICCV99 Intl Workshop on recognition, analysis and tracking of faces and gestures in Real-Time systems, 5863. 1999.

[42] Cui, Bo, and Tongze Xue,Design and realization of an intelligent access control system based on voice recognition, Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on. Vol. 1. IEEE, 2009.

[43] Sumathi, S., S. K. Srivatsa, and M. Uma Maheswari. ,Vision based game development using human computer interaction, arXiv preprint arXiv:1002.2191 (2010).

[44] Ghotkar, Archana S., et al. ,Hand gesture recognition for indian sign language.,Computer Communication and Informatics (ICCCI), 2012 International Conference on. IEEE, 2012.

[45] Lee, Christopher, and Yangsheng Xu. ,Online, interactive learning of gestures for human/robot interfaces., Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on. Vol. 4. IEEE, 1996.

[46] Nosowitz, D. ,Video: MIT's Kinect Hack Tracks All Ten Fingers Simultaneously., (2010).

[47] Malik, Shahzad, and Joe Laszlo. ,Visual touchpad: a two-handed gestural input device., Proceedings of the 6th international conference on Multimodal interfaces. ACM, 2004.

[48] Samanta, Debasis, Sayan Sarcar, and Soumalya Ghosh. ,An approach to design virtual keyboards for text composition in Indian languages., International Journal of Human-Computer Interaction 29.8 (2013): 516-540.

[49] Yin, Ying, and Randall Davis. ,Real-time continuous gesture recognition for natural human-computer interaction., Visual Languages and Human-Centric Computing (VL/HCC), 2014 IEEE Symposium on. IEEE, 2014.

[50] Wobbrock, Jacob O., Meredith Ringel Morris, and Andrew D. Wilson. ,User-defined gestures for surface computing., Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2009

[51] Lucchi, Aurlien, et al. ,Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features., Medical Imaging, IEEE Transactions on 31.2 (2012): 474-486.

[52] Shapiro and G. Stockman, Computer Vision, Prentice Hall, (2002).

[53] Kook-Yeol Yoo,,Robust hand segmentation and tracking to illumination variation,, Consumer Electronics (ICCE), 2014 IEEE 1nternational Conference on, vol., no., pp.286,287, 10-13 Jan (2014)

[54] Yang, Guoliang, et al. ,Research on a skin color detection algorithm based on self-adaptive skin color model., Communications and Intelligence Information Security (ICCIIS), 2010 International Conference on. IEEE, 2010.

[55] Qi, Feng, Xu Weihong, and Li Qiang. ,Research of Image Matching Based on Improved SURF Algorithm., TELKOMNIKA Indonesian Journal of Electrical Engineering 12.2 (2014): 1395-1402.

[56] Suaib, Norhayati Mohd, et al. ,Performance evaluation of feature detection and feature matching for stereo visual odometry using SIFT and SURF., Region 10 Symposium, 2014 IEEE. IEEE, 2014

List of Publications

1. Kapil Yadav and Jhilik Bhattacharya "Real Time Hand Gesture Recognition Based Interface for Microsoft Word Document Handling," accepted at Fourth International Conference on Advances in Computing, Communications and Informatics (ICACCI-2015), Kochi, 2015 will be published by "Springer as a special volume in the prestigious Advances in Intelligent Systems and Computing Series".

2. Kapil Yadav and Jhilik Bhattacharya "HCI System for Windows Application", accepted at International Journal of Innovations & Advancement in Computer Science (IJIACS-15), Vol. 4 ISSN (2347 – 8616) 2015.

Video URL

https://youtu.be/XfurXjdZHQA