Media Player Control Through Hand Recognition And Hand Detection

Thesis submitted in partial fulfillment

of the requirements of the degree of

Master of Technology

in

Very-Large-Scale Integration

by

Afsha Khan

(Roll Number: 2015PEV5130)

based on research carried out

under the supervision of

Prof. Vineet Sahula

June, 2017

Department of Electronics and Communication Engineering

Malaviya National Institute Of Technology,

JAIPUR 302017

Department of Electronics and Communication Engineering
Malaviya National Institute Of Technology, JAIPUR 302017

June 30, 2017

Certificate

Roll Number: *2015PEV5130*
Name: *Afsha Khan*
Title of Dissertation: *Media Player Control Through Hand Recognition And Hand Detection*

We the below signed, after checking the thesis mentioned above and the official record book (s) of the student, hereby state our approval of the thesis submitted in partial fulfillment of the requirements of the degree of *Master of Technology* in *Very-Large-Scale Integration* at *Malaviya National Institute Of Technology, JAIPUR 302017*. We are satisfied with the volume, quality, correctness, and originality of the work.

Prof.Vineet Sahula

Declaration of Originality

I, *Afsha Khan* , Roll Number *2015PEV5130* hereby declare that this thesis entitled *Media Player Control Through Hand Recognition And Hand Detection* presents my original work carried out as a Postgraduate student of Malaviya National Institute Of Technology and, to the best of my knowledge, contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of Malaviya National Institute Of Technology or any other institution. Any contribution made to this research by others, with whom I have worked at Malaviya National Institute Of Technology or elsewhere, is explicitly acknowledged in the dissertation. Works of other authors cited in this dissertation have been duly acknowledged under the sections "Reference" or "Bibliography". I have also submitted my original research records to the scrutiny committee for evaluation of my dissertation.

I am fully aware that in case of any non-compliance detected in future, the Senate of Malaviya National Institute Of Technology may withdraw the degree awarded to me on the basis of the present dissertation.

June 30, 2017
MNIT Jaipur

Afsha Khan

Acknowledgment

I would like to convey my sincere gratitude to Mr. Vineet Sahula, Professor, Department of Electronics and Communication Engineering for his keen interest in guiding me on such a progressive topic with his great dedication, expertise and knowledge throughout the process of this research. Without his support and timely guidance, the completion of this project would have seemed a farfetched dream. He has guided me not only with the subject matter, but also taught me the proper style and techniques of working.

I am also thankful to Prof. D. Boolchandani, Professor of Electronics and Communication Engineering Department and Dr. C. Periasamy for their Co-operation and help rendered in numerous ways for the successful completion of this research and allowing me to work on this topic and providing me all possible support.

I thank all Technical, Non-Technical Staff of the Laboratory as well as office of Electronics and Communication Engineering Department MNIT, Jaipur for supporting and encouraging me throughout the research work.

Abstract

This thesis focuses on recognition of all the alphabets of ASL correctly. The proposed work is to take the image from the input folder that is to be recognized and then apply different algorithms to the testing image to find its match among the database image. Database images include all the possible type of images that is all the gestures of ASL languages. This thesis also concentrates on controlling the media player through the hand recognition. Also this thesis contains the detection process of the hand posture taken from the webcam. That is detection of hand in real time from the captured webcam image or the real time video capturing through webcam.

Our hand gesture recognition system is done by point pattern matching algorithm along with the SIFT algorithm. The SIFT algorithm is used to find the key point of the image to be tested hence convert the image into key point readable format. Then the key points falling in the low contrast regions are eliminated and also the key points falling at the edge are discarded. Then the features are extracted from the remaining key points and are saved. These features represent the information about the image like its color, shape and location.

Then the SIFT matching algorithm is applied to match the features of the input image with the database images. The database images are trained images. Key points of database images are also extracted. The features are saved into vector or array format. Then these arrays are matched for recognizing the similar image from the database.

The database image with the maximum key points matched is selected. Then this image and the input image are checked through Mk-Rod algorithm. It calculates the validity ratio for the two images that represents the relative distance between the key points of both the image. This algorithm check if the matched key points are really from the similar image or similar key points are false recognized as similar image. With these algorithms the input image is recognized correctly. Then the media player is controlled by the recognition process. The basic features of media player that is play and stop are controlled through this process. Two ASL language gestures are assigned for play and stop. When the input image is matched to any one of these then command is given to the media player to perform the required action.

Hand detection is also done in this thesis where the real time image captured from the webcam is converted to grey format and hand is detected with the skin color

properties. It is also able to convert the real time video from webcam simultaneously into the grey format and detects the skin color.

—

***Keywords*: *HGR (hand gesture recognition)*; *HPR (hand posture recognition)*; *feature extraction*; *SIFT (Scale Invariant Feature Transform)*; *Mk-Rod*; *Template matching*; *PCA*.**

Contents

List of Figures

Chapter 1

Introduction

The most common input device has not been changed very much in recent years. This might be because the present input devices are user friendly and don't require much effort while using. The most common input device now days are keyboard, mouse, light pen, keypads etc. the hardware introduced with such devices may change with time but the basic process of giving input is same. Nowadays computers are basic requirement of everyday life and the way of communicating with them is restricted to these input devices till now. These devices are familiar to user but also they restrict the natural way of communicating of humans with the computer. Since the technical industry follows the Moore's law then more and more device are now integrating on same chip and also many other peripherals are introduced. Also many vision based systems were developed where computer is able to see. Hence new types of interactions are continuously developing. These new peripherals and technology requires new and advanced commands. These commands might not be possible with the same input existing devices. Hence advancement in input devices is also required. With these advanced input devices also plenty of time will be saved.

The new advancement in this field is towards hand gesture. Recognition of human hand gestures can be used as the input device for the interaction with the computer. Hand gesture recognition can be used in many fields like gaming, controlling objects at field, and replacement of other input devices like mouse and keyboard. The most common set of hand gestures include sign languages. In sign languages each posture describes a particular meaning or letter so it will be easy for computer to understand through sign language. Hand recognition approach may provide more natural interaction with the computer. By using hand gestures we can command the devices of instruct the computer to do what is required. Hand recognition is divided in two categories, static and dynamic. Static hand gesture recognition involves a constant posture of hand that is shown by a single image whereas dynamic hand gesture recognition involves moving hand postures that is shown by a sequence of images. In this thesis we will concentrate on recognition of static hand gestures. These types of interacting technologies come with many challenges like delay. There should ideally be no delay between when the user gives command and when the computer responds.

It should be as fast as possible then only these interactions will be helpful. Also the algorithms used should be accurate and work for different people. Hence speed and accuracy are the main factors while creating such applications.

There are also economic factors. Obviously new interfaces and new applications require additional hardware which will increase the cost. Like vision based systems are replacing remote used in television or video games that are very cheap. So why user will spend more money for same controlling by using vision based systems while they are getting these features with lower cost. Similar constraints will be there for regular input devices and advanced hand gesture based input commands.

The problem is large as present algorithms are complex or slow or unreliable. And the algorithms that are fast and accurate are very expensive as compared to present hand operated input devices.

1.1 BACKGROUND

Regular human computer interactions that are traditionally used presently became ineffective when we talk about virtual environment where we have to run 3D applications. These applications need a new interacting device with different features. These applications require more natural devices which will provide effective human computer interaction. Among all human computer interactions hand gesture has gained maximum popularity. The aim for using hand gesture is to use the posture or movement of human hand or arm to interact with computer and to control the application or to give the command. Hand gestures are a collection of various types of postures of hand that will be interpreted according to our need. Static postures are generally used to give a particular command while the dynamic postures are used for communication. Recognizing these hand gestures require to convert them into spatial and temporal domain. Researches are still in progress for recognizing hand gestures using different algorithms. The hand gesture control information in both spatial and temporal mode. Recognition methods of hand gesture are divided into two categories. First is data-Glove based approach. This uses mechanical or optical sensors connected to glove that converts finger movement into electrical signals and then recognizes the hand posture. This method is more complex in terms of hardware and ease of using and the naturalness of human computer interaction. Also the hardware carry cable wires attached to the computer and also the setup is relatively expensive. Second category is vision based method. This method uses only a camera which is more near to the natural human computer interaction. Also no more extra hardware is required as camera already present in computer. This method requires implementation of artificial vision based systems. There is another problem with this method that the system should be independent of the user and the camera, the variations in the size and

color of the hand, variations in the background, lightning condition to make system real time. The problem arises while comparing the posture since human hand differs in shape and formation and color. The vision based approach is the most natural way of creating human computer interaction. However it is also the most difficult way since the vision based machines are less accurate. So these systems are less satisfactory. Also to increase the reliability and accuracy more than one camera should be employed so that the motion of hand can easily be detected. The vision based approaches are of various types.

They differ among themselves in many ways like –

- The position and number of cameras used.
- Type and number of features that the algorithm is extracting from the image.
- Constraints related to background, lightning, motion speed etc.
- Image type (only hand part, with arms, with sleeves, with watch rings etc.)
- 2D or 3D image difference.
- Time and space representation

3D images require more feature extraction and space representation. Also the tracker should be independent of shape and size of the hand. In this thesis we are concentrating on 2D images. The method and algorithm used here is independent of the color and lightning condition. But depend on camera quality that is the resolution. Hence this system is not accurate for real time images taken from webcam. This thesis concentrates on recognizing the posture from the database and then giving command to computer to control the media player.

The technique involved here used the point pattern matching algorithm where the features of the image are extracted and then the patterns of feature are matched. The image having maximum matched features is selected.

The group of postures used is American Sign Language. There are total 26 hand positions in ASL used to represent 26 alphabets of American English language. The technique used here is vision based. Since the vision base technique is more user friendly and more near to natural human computer interaction. But it includes some complexity relates to machine vision because human hand differs in color, shape, size, rigidness etc. therefore this technique is less reliable and less accurate compared to the glove based approach. Here we are using vision based approach because it is based on how human receives information from surrounding. However satisfactory implementation of this approach is not simple.

1.2 MOTIVATION

1.2.1 HUMAN COMPUTER INTERACTION

The method of receiving information by computer from human categorizes the system of human computer interaction. Different types of devices used have effect on the output or the response of the system. Input device type effect the speed of response and also the accuracy. Traditionally used peripherals for interacting with computer are keyboard and mouse. But for working with virtual environment or 3d images these devices would not work. Hence more interacting methods are required. Gesture recognition based methods are one of those. The initial beginning of human computer interaction starts with the Shakel's article on 'The ergonomics of a computer '. Since then new methods have been developed. The first integrated keyboard came in 1981 used for personal computer and the first mouse that was widely used was on Apple Lisa PC in 1983. As the machine learning capability rises then more methods were developed to interact with computer and to command it with various input methods. Computer vision based methods are most promising new generation of human computer interaction. In this human hand gesture recognition is the most widely used interacting method.

1.2.2 APPLICATION

There are many applications based on the human gesture recognition. The human recognition is itself categorized into hand recognition and face recognition. The hand recognition techniques are widely used in gaming, controlling machines, interacting with sign language. Sign languages are related to ancient time and also with the persons not able to speak. The main area of focus of this thesis is toward American Sign Language. It is the language for most of the deaf people living in United States. There are many sign languages used worldwide. ASL is one of them. Like people speaking one language can't understand the other language, similarly people using one sign language are not able to understand other sign language. In this thesis we are including postures of ASL that represents alphabets of English language. ASL also has its own grammar like any other language. ASL consists of around 6000 gestures representing common words or phrases. Some of the signs are shown below

In ASL finger are used to spell a proper word or noun. ASL can also describe a person , a thing or place. ASL also uses facial expressions to differentiate between questions, proverbs, directives or statement. For example for asking a question, eyebrows are raised. There are various methods and algorithms developed separately for recognizing facial expressions and features. All these hand gestures and facial expression are defined in full ASL dictionary.

Figure 1.1: ASL Gestures

1.3 PROBLEM STATEMENT

• Vision based recognition of hand gesture presently became a highly researched area with many interacting applications like ASL , gaming, robotics, machine controlling, virtual environment controllers, remote controllers, HCI etc. Hand gestures are a powerful mean of interacting with human. Hand gestures can generally be categorized into four parts. These are controlling gestures, manipulative gestures, conversational gestures and communicative gestures. Hand gestures have begun to evolve but till now also these systems are not robust or much effective and are weak in analyzing gestures in different environment. Hence real time applications are not easily accessible. More techniques are needed to be evolved to make hand gesture a universally used language that is more frequent, robust, accurate, fast, simple and interactive. Many advanced techniques have been developed that attains the required accuracy and robustness but they are not easily achievable and hence not frequently used techniques.

• Also it is difficult to make vision based approach real time since then we have to take image from camera and simultaneously process it. This type of system should be camera quality such as pixels and resolutions independent. There are many phases that an image has to undergo like processing, detection, motion capture, tracking, recognition of shape and size, pattern matching. Hence the problem is quite challenging because of the complex static and dynamic captures of gestures, cluttered backgrounds, lighting, transformations and occlusions. To solve such problems we need more elaborate technique for hand detection and recognition.

• Capturing hand gestures from a video frame is one of the most challenging parts of whole recognition system. It simultaneously require the computer to detect, track and recognize the hand gesture and then at the same time interpreting the gesture and to control various devices or to interact with other interfaces of human machine.

• The purpose of this thesis is to recognize the hand gesture more accurately from the given set of data and then to control a media player according to the pattern recognized.

• Further real time recognition can be done. This type of recognition is complex and is measured by frames per second. If the speed of the generation of hand gestures is too fast than it may not even detect and recognize. Hence stronger algorithm is needed for this.
• Next requirement is the system should be flexible so that it can be combined with any application that we want to command through hand recognition.
• The algorithm should be accurate and precise enough so that we can use it practically. It should recognize correctly more than 90 • Next the system should be reliable and robust so that it can recognize accurately without any effect of lightning and background noise. So that it can be used with different persons in different environment. It should be independent of environment conditions.
• The algorithm should be scalable. It should operate on different size images and must scale them on a single platform.
• It should be user independent and must work on various person hands of different size, shape and color.

1.4 OBJECTIVE

• This thesis completes the recognition of hand gestures of ASL. The algorithm used is resolution dependent therefore it is not used as real time algorithm. Hence images are prerecorded with particular resolution. Then these input images are matched with all database images that are alphabets of ASL. It shows how many key features of a particular input image matched to each database image. And shows the exact match of the input image sign.
• With these matches the media player controller is interfaced. The media player gets ON and OFF when a particular sign matches.
• Hand detection is also done separately for the real time images taken from webcam but it is not recognized due to different resolution. For recognition input saved images will be used.
• For recognition the key features are extracted from the input image and the database images by using Scale Invariant feature Transform (SIFT). By using this algorithm the image get converted into key features. Then these features are saved in array form and matched with features of database images.
• The key features extracted are independent of color of image, scale invariant, independent of orientation of hand and illumination
• Then the input and matched image are shown together and the command passes to media player accordingly.
• Hand gesture set is restricted to static hand postures which describes the alphabet

of a language.

1.5 THESIS FLOW

This thesis develops a system that is capable of recognition of ASL gestures and controls the media player through this. The flow of the thesis is as follows-

- Chapter 1 gives the introduction of the method and the done work. It also defines the need of the work and the research done in this field. It describes the problem statement and the objective and purpose of the thesis. In addition to this it gives the motivation behind this.

- Chapter 2 gives the theoretical concepts based on the project and the related work done till now that is required to complete this thesis. Theory explains the need and application of the proposed work in day to day life. It in short also explains the steps involved during the recognition phase.

- Chapter 3 represents the proposed work on the basis of which this thesis is formed. It also describes the different technique that can be used and the motive behind the algorithm used here. It describes step by step the work that is done and the phases of the proposed algorithm.

- Chapter 4 describe the implementation of the project. It gives the exact method and steps followed for completion of the proposed work. It explains each algorithm, factors involved in them and the working of the algorithm.

- Chapter 5 gives the results of the project and also discusses the future work that can be done to enhance the idea of human computer interaction and to make it more effective and vast. This part concludes the thesis and the work done in the project and the results.

Chapter 2
THEORY AND RELATED WORK

2.1 THEORY

2.1.1 HAND GESTURE APPLICATIONS

Here we discuss the applications where hand gesture is used or can be used further for interaction. It can be used in gaming, medical field, mechanical controlling field, computer interaction, robotics etc. In robotics, in future human controlled robots can be formed which act according to the human gestures. General and widely used application areas are discussed below

MEDICAL SYSTEM AND RELATED FIELDS

Hand gestures can be utilized to operate medical equipment, in screen visualization, and to help the handicapped persons in their work. Some ideas related to these have been used nowadays to enhance the medical facilities and systems. A system of physician and computer interaction has been developed where hand gestures are interfaced to operate the mouse functionality like curser movement and pressing a button. A hand gesture machine for tracking is designed named Gestix where doctor can study MRI images in the operating room through a natural interface using hand postures as shown in the fig 2.1.

Figure 2.1: HGR application in medical field

ENTERTAINMENT PURPOSE

Hand gestures can be used in gaming field or media player type applications which comes under entertainment category.

Gaming field has a vast interfacing possibility where various types of interacting devices can be developed for different type of games. Hand gestures can be used to control the player of the game. This technology can be further enhanced in virtual reality gaming where a person can operate his virtual character through his actions. These type of vision based hand gesture controlled games must have a fast, reliable and robust algorithm. Speed in such application is the major factor to be controlled. The controlling action should be performed immediately after the gesture input by the player. Also these applications require real time compatible software.

Intuitiveness is another major factor involved in computer gaming systems. The intuitive feature is acquired using gesture vocabulary. For this a training session must be carried out so that the machine can easily understand the action of the user. This type of intuitive feature of hand gesture recognition is employed in a children's game named QuiQui's Giant Bounce. In this user interacts through a hidden interface with the gaming application. Training session for user is also carried out to understand and to become use to with the speed, posture and configuration used by the machine.

Hand gesture can also be used to control the media player that we are doing in this thesis. We can operate the media player functions like play, stop, fast forward, backward etc.

HUMAN - ROBOT INTERACTION

Another major application of human interacting with computer is robot control. Many researches have been done regarding this and also a lot papers have published related to this. Gesture recognition is main part of controlling the robot action that might be fixed or portable.

A similar application has been developed where hand gestures that are visual commands and voice commands are combined to guide the robot. A human robot system was designed to track the postures of the moving hand and arm. These type of applications have learning robot that learns from the human arm movement and recognize the trajectories formed by the human hand.

2.1.2 METHODOLOGIES OF APPEARANCE BASED HAND RECOGNITION APPROACH

There are many approaches for hand recognition. One of them is appearance based approach. In this section appearance based HGR method is discussed. There are three

steps involved in this approach. These are discussed one by one below. Analysis of pros and cons of this approach is also done here.

SEGMENTATION AND TRACKING OF HAND

For tracking the hand from the image it is first segmented from the background that is background is subtracted and then hand image is detected and tracked in the video and its trajectory is tracked. Then the features of the tracked image from the video is detected and then compared with the feature of the database image. The features of the tracking image are called special tracking feature and the features of gesture classified image are called temporal trajectory features. The target hand is distinguished from the background through the skin color. Hence skin color detection is the most common and widely used feature in HGR techniques. Researches also going on for skin detection methodologies. Human skin color has different distributions. The hue of different skin tones is same but the saturations are different. The most widely used color space is HSV for skin detection. The tracking and recognition of image should be very fast for time sensitive applications. Detection time should be minimum for such applications. Generally used method for skin detection is setting threshold. The color threshold is set very carefully then the color hue and composition values that are falling under that threshold are detected as skin. In this thesis also for hand detection through webcam we are using threshold method.

For hand posture recognition, if the images are still and not moving then there is no need to perform tracking. For still images direct HGR methods can be applied after the feature extraction. For moving images, first the hand part is detected and then it is tracked. Hand tracking is similar to other tracking such as object tracking etc. there are various methods used for hand tracking. One of them is Mean Shift method for analysis of feature space. Basic idea is to detect a particular feature of the target object such as contour or color and then search for the optimized matched area in the complete frame.

FEATURE EXTRACTION

While recognizing the hand posture, its features are the data representation of the posture or shape of the hand. Therefore at first the features of the hand image is extracted and saved which gives the information about the hand posture in data format. There are generally two types of features of any image, contour features and texture features. Contour feature is related to the boundary of the target image of hand that is it is the output representation of the target hand. Contour features are more often described by Fourier descriptors. The basic method is to represent the contour by a one dimensional vector. Then the Fourier transform of the vector is taken. The Fourier coefficients of this vector are treated as the features of the contour

of the hand image. For describing properties of contour, Moments are used. They give properties like sum of vertical and horizontal directed variance. There are more contour features such as shape signatures like the average of the distances between the pixels on the contour, contour's centroid, wavelet descriptors etc. Texture features represent the pattern of the image. They are the representation of various gradient patterns. One popular and widely used texture feature is Histogram of gradient. It is the representation of local gradient in form of histogram.

Another popular texture feature is SIFT that is Scale Invariant Feature Transform. In this thesis we are using this texture feature. In SIFT various key points of image represent the coordinates and gradient properties of the image. The key points are generally the high contrast region of the image that gives the feature of the image. SIFT has two advantages that make it very useful. First is it is independent of the scale rotation of the image. Hence the SIFT key points are the representation of the edge or high contrast region and are invariant towards the scale and orientation of image. Second is it is also tolerant to certain level when the view point changes as it is independent of rotation of the screen and scale invariant. These properties make it desirable for using it as a feature extraction method for HGR. Here 128 bits descriptors are used to represent each key point feature. Hence for K number of key points of an image, all key points are represented by a K by 128 size matrix.

Another method for descriptor is Binary Robust Independent Elementary Feature (BRIEF). This is the method of building a short binary descriptor with independent bits. This method does not use the Euclidean distance between the descriptors. Instead of this it uses the hamming distance. It speeds up the matching process of the texture feature. But the problem with this is that the descriptors are not scale and rotation invariant. Later an improved binary descriptor method is proposed that is independent of the rotation called as Oriented Fast and Rotated BRIEF (ORB). This is also robust towards the noise which makes it better than previous one. Further a method called Binary robust Invariant Scalable Key points (BRISK) that is both scale invariant and rotation invariant. Hand tracking process locates the position of the hand in the image. After the hand position is located in the frame the temporal features of the hand are extracted that represents the trajectories. There are only a few trajectory features that are commonly used unlike the contour and texture features. These trajectory features are divided into two categories, local and global feature. Local trajectory features include hand movement direction, speed and location. The coordinate displacement of hand, orientation of movement of hand and hand velocity between the adjacent frames are used to represent the elementary trajectory segments. Whereas global features describes the shape of the hand and are extracted from the complete trajectory of hand. Contour and texture features can also be used as the global trajectory features because trajectories can also be taken as still image just like

the hand postures are stationary images.

GESTURE CLASSIFICATION

In all type of classifiers the feature pattern of the training image are summarized and then the feature pattern of the testing image is matched with patterns of the trained images one by one. For HGR, for specifying both spatial and temporal features, hand trajectories are used. For classifying the hand trajectories, the classifier must process the sequential features for HGR. But for classifying the features for the HPR, classifier is not requiring to process the sequential features because the postures are the still images and they don't contain any temporal information. For directly measuring the distance between the testing image and the trained image, template matching is a method widely used. It has two advantages. First is that minimum training is required because this method directly measures the distance between the vector feature of the two images instead of measuring elementary local patterns by extracting latent patterns. The templates of gesture are the feature vectors that are pre-processed. Hence there is no need for the statistical data for the feature vectors. Second advantage is that the inference time cost is generally low. Also the calculation of the distance between feature vectors does not require complex computation. The statistical model approach is considered to be more computational intensive then the template matching methods. One of the most popular template matching methods is Continuous Dynamic Programming (CDP). In this to represent trajectories, a set of sequence patterns are used in spatial and temporal space. And to match the sequence pattern, A dynamic programming based method is used which sums the distance between corresponding elements in sequence pattern.

2.1.3 RELATED WORK

1. Deval G Patel [1] introduced the system able to recognize the ASL language gestures using the point pattern matching algorithm. The algorithm used effectively recognizes the gestures by taking them through input image folder and also it works for real time recognition. It also compares the work with other algorithms and defines the accuracy percentage of all methods.

2. Talla Bhuvana vaibhav Reddy et al. [2] proposed a system that is able to detect and recognize the gestures of human and then convert it to speech. This system is useful for deaf and mute people. In this project the system at transmitter side recognizes the human gesture taken through webcam. The computer installed with MATLAB selects the matches sign from the database and then the transceiver employed would send the information to the other hand

through Zigbee transceiver. On the other hand the signal gets converted into corresponding speech signal that can be heard. Reverse process is also possible.

3. Pallavi Gurjal and Kiran Kunnur[3] introduces improvised SIFT algorithm base recognition system that is able to detect and recognize the ASL gestures. This system works in real time environment and able to take image from the webcam. The system propose is also have fast processing with capturing 40 frames from a video in 40 seconds, the system is also invariant to the scale, orientation.

4. Hamid A Jalab and Herman K Omer[4] introduces a HGR system for interaction between human and computer through neural networks. The proposed system also controls media player through neural network. The system works in four phases namely image acquisition, hand segmentation, feature extraction and classification. The neural network used provides good accuracy. The proposed work includes detection of the skin part.

5. Reza Azad et al.[5] proposed a real time system to recognize the human hand and face gesture. In this the face recognition is used to detect and identify the person and then hand recognition is used to control the media player application like play stop. In the proposed system first the image is captured and hand and face are detected and recognized separately.

6. Manuj Paliwal et al.[6] proposed a hand recognition dynamic system. This system is low cost with effective computation. The proposed system detects the skin part of the gesture. It uses the central computation module and the approximate median technique. This system uses a decision tree for recognition that uses features extracted from segmentation.

7. Siddharth S Rautaray and Anupam Agrawal[7] proposed a system that works with multiple gestures. The system uses the basic phases of recognition after feature extraction and then also integrated with several applications like virtual game, image browser etc.

8. MS Kalyani et al.[8] introduces a system that recognizes the ASL gestures using the neural network. Here aal gestures representing 26 characters of language are successfully recognized by means of static posture. This method is also capable of classifying the gestures.

9. Prof. Suganthi S and Prof redappa BC[9] proposed a dynamic hand gesture recognition system for controlling the media player application. This method does not utilizes the database for comparing the images. It uses color detection method and threshold concept. For interacting with media player also they have

used ActiveX control method. Hence making it independent of the cluttered or changing background and noise. This system also accepts the dynamic gestures for controlling media player.

10. Siddharth S Rautaray and Anupam Agarwal [10] proposed a noval human computer interface system using hand vision techniques for recognizing hand gestures. The system uses a central computational module which uses the PCA. It finds the feature vector of the gesture and XML file is used to save the vector. Recognition is done by K nearest neighbor algorithm and hand motion is detected using Lucas Kanade pyramidical optical flow algorithm.

11. Mrs AR Patil and Dr SS Subbaraman[11] proposed a vision based HG system that uses support vector machines. This system used in various applications like sign language and virtual reality. The method is based on SVM that is machine learning algorithms that analyze data and recognize patterns.

12. Shahzad malik et al.[12] proposed a system that uses hand tracking for interactive pattern based augmented reality. This pattern based augmented reality method most promising for recognizing the virtual objects with real time video feeds. The method describes a fast and accurate method for tracking the 3D augmented virtual reality objects.

13. X Zabulis et al.[13] proposed a vision based HGR for human computer interaction. This represents the natural interaction with the autonomous robots to guide the visitors in the museums. It is based on the probabilistic framework in which it deals with multiple information. Tracking is achieved by a system that is capable of detecting multiple hands.

Chapter 3

PROPOSED WORK

3.1 HAND DETECTION APPROACHES

There are many approaches for hand detection. Each approach is based on any one or more feature. Through that feature the hand is detected from each image. The same feature is matched in all images. Each method depends on different visual effects. These features can be color, motion, skin or shape. Some of the approaches used for detection are discussed below. In this thesis the method used for hand detection is skin detection.

3.1.1 COLOR

In many hand detection applications skin color detection is used frequently. The major part in this is to decide the color space that must be utilized in order to detect the skin color. For this a threshold is set with upper and lower value. The color coming under this space will be considered as the skin color. There are many color spaces introduced like HSV, RGB, normalized RGB, YUV etc. Some color spaces split the chromaticity and luminance very effectively. These type of methods are preferable. In these approaches only chromaticity dependent components are utilized hence it will become illumination independent to some degree.

The problem with this approach is that many times the object in the background having color similar to skin color can be confused with the hand. There are some solutions for this problem. Some techniques are using background subtraction approach. But some background approaches depend on the static background and the camera system. For this also more advanced background subtraction techniques are utilized. Another solution involves dynamically correcting the background models or compensating the background.

3.1.2 SHAPE

Another technique used to detect hand is detecting the shape. For this the boundary features are utilized. Hand shape features are employed per frames. For this the

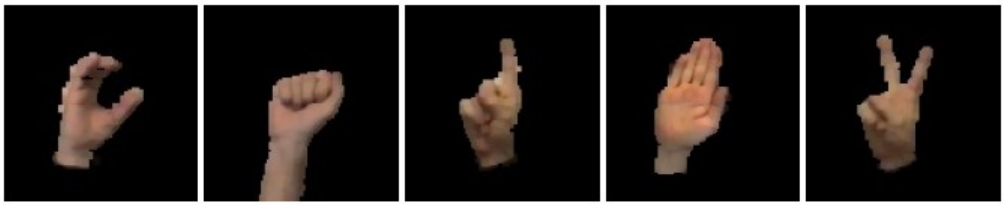

Figure 3.1: Detection of skin using color detection

contour features are used to accurately detect the boundary and shape of the target object. Contours are extracted from each frame. Then the contours are detected. If they are detected accurately then they provide good features related to shape. This type of detection does not depend on color, rotation or illumination. But the 2D shape image extraction can be limited by occlusions or viewpoints. The shape extraction method involves usually the edge detection where the edges of the hands are detected and there features are extracted. These edges should also be independent of background objects. For more accuracy, more approaches can be combined or post processing is required. For more accuracy sometimes skin color detection can be combined with the contour detection.

Some methods also use some part of hand to detect the hand like fingertips. There are some methods that utilize the feature related to curve of the finger so that the hand shape can be detected. Another method that uses fingertip detection uses template matching. Templates may include fingertip images. These methods can be improved by adding contour matching methods.

3.2 HAND GESTURE RECOGNITION

3.2.1 BACKGROUND KNOWLEDGE

Hand posture recognition is the method to compare the two image by matching there features that are extracted from any HGR technique. HPR is the technique used for the still images. It does not contain any temporal information because the target image here is stationary. In this thesis we are dealing with the hand poses that are still and hence do not require to be tracked. HPR techniques are simpler than HGR because of this reason. Still the background subtraction is an issue here. Hence before recognition of the image the background need to be subtracted from each image during detection process.

Appearance based methods have the problem of cluttered backgrounds. The HPR methods some time use the boosting based method. Through this the optimized feature set can be selected which in invariant to rotation and scale. In boosting based methods, all the points of interest are used to strengthen the classifier of each posture.

Hence the most effective and strong features are combined. This method reduces the cost of computational inference.

3.2.2 POINT PATTERN MATCHING ALGORITHM

The HGR system in this thesis recognizes the 36 ASL hand gestures. It takes the hand as the input and performs several actions that depend on the category of the pattern of the image. The key points (the points that are present at the high contrast region of the image) of the input image are calculated and then matching is done using the point pattern matching algorithm.

Here the algorithm used is point pattern matching algorithm and SIFT algorithm is used to transform the image into key points and then to recognize the gesture. Point pattern matching approach is simpler and fast approach. In this the features of the key points are matched with the pre-existing features.

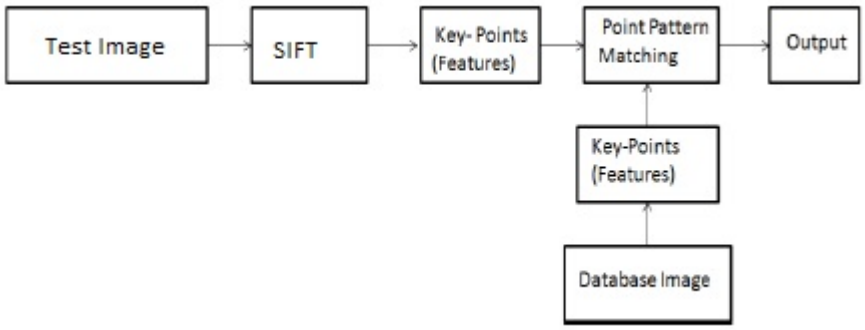

Figure 3.2: Block diagram of HGR technique using SIFT

As shown in the figure 3.2 above, the input image is taken from the webcam or from the saved input images. These image will then transformed using SIFT algorithm. This transform will give the key points of the image in the form of a feature vector. Then this vector is compared with the other feature vectors of the database image of ASL gestures. For comparison, point pattern matching algorithm is used.

HGR techniques require matching of two sets of point in space since the images are represented by their key point features. Point pattern matching approach is very efficient and robust in nature. So we are using this algorithm with the SIFT transform approach to recognize the ASL hand gestures. The flow chart of the system is shown below The working of this algorithm is described below in steps

- Take a input image that is to be tested
- Transform the test image called pre-processing
- Initialize the distance ration with value 0.65
- Initialize the threshold value of the algorithm with 0.035
- Run the SIFT algorithm for matching

Figure 3.3: Flow chart of system

- We will get the key points matched between the test image and all the database trained images.
- Now we calculate the validity ratio
- If we got more than one result then re perform the process with new set values
- Increment the distance ratio of algorithm by 0.05 and decrement the threshold value by 0.005 and repeat the above process from point 5.
- Repeat the process until we got only one result
- Then display the result

SIFT algorithm is used for converting the image into key points. Then the features of these key points are extracted and then saved into array format. SIFT based algorithm are rotation variant and scale variant. Hence image having different hand size can also be recognized accurately. This algorithm works in four steps. At first the points of local interest are identified and located by calculating the maxima and minima of a set of difference of Gaussian filters. These filters are applied all over the image at different scales. Among all the points detected and located the points falling in low contrast regions are eliminated. The points having high contrast are only considered as the key points of a particular image. For more accuracy and reliability more number of key points should be detected. These key points are stored and then used to compare the image with the database images.

Then the orientation is assigned to each key point that depends on the image feature. Then a image feature descriptor is formed for each key point. This descriptor contains the information of the feature of the key point. The content of the feature descriptor is based on the local gradient transform that depends on the orientation assigned to the key point. The size of the descriptor depends on the number of feature extracted for each key point.

Here we are using the SIFT algorithm to first select the key points for each database image. Then a key point vector is created containing features of all key point of a particular image.

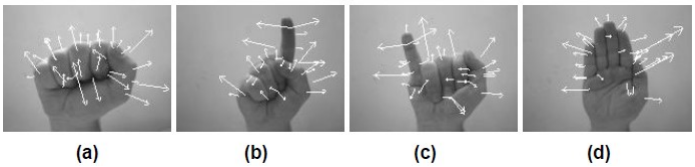

Figure 3.4: key points of some gestures

The number of key points extracted depends on the quality and condition of the image and also on the position of image from the camera. The number of key points decreases when the hand moves away from the camera and the number of key point increases as the hand moves towards the camera. This is because of the change in the area of the hand that the camera captures. Among all the parts of the hand the maximum numbers of key points are present in the palm.

For increasing the reliability of the system and for making the system to detect hand of different persons and of different scale, we should train our system with more training images. Hence the database of such applications should be as large as possible and should cover all the possibilities of the posture that a hand can have. We can train the system to work for all sizes, orientation and illumination. When the number of images with different condition is more in the database the more accurate will be the k-matrix that is the descriptor matrix.

First the database images are stored and then in the training stage all database images are trained according to the feature extracted from their key points. Then the k-matrix is created for each image that contains the features of each key point present in the image.

Then the input folder is formed that contains the image to be tested. These entire images during recognition phase get converted into key point vector that contains their feature information. Hence now all database and input image get stored in key point readable format. Now matching of the input image starts with each database image. This matching is done between the key points of the images. Hence the output will be the number of key point of input image matched with each database image.

The number of key points matched gives the number of features matched between two images. This approach will be used during HGR phase.

3.2.3 FORMING K-VECTOR

Cluster means a subgroup where all the elements are similar to each other. We do clustering of the features to divide the groups into subgroups that contain similar

features. This is done to simplify the matching process. There are many ways for doing clustering. K-means clustering is one of them.

Clustering is used in many fields such as pattern recognition, machine learning, bioinformatics, data mining, image analysis etc. Clustering makes the process of identification and classification very simple.

The cluster size depends on the amount of data to be represented. If the cluster size is too small then it cannot represent all the features or key points that are extracted. And if the cluster size is too large then it will create over burden because of less number of key points extracted from the training image.

The cluster size required will be less if the background color is white or black. Then it would require less number of key points to represent the features of the whole image. This is because in this case only features of the hand or palm need to be detected. So the key point extracted will only represents the features of the palm. So in this case number of key points will not exceed 75 for the palm posture. The palm has maximum number of key points. Hence for HGR the palm features are enough for matching with the trained database images. For this we can assume that the cluster size should be larger than 75.

Hence the training stages inform us about the minimum number of clusters that we have to take. When the image is taken from the webcam then it also contains objects in the background. So the number of key points increases since they also represent the features of the background objects. In such case there will be around 400 key points in total for one image. Hence the number of clusters required increases.

Here in this thesis we are representing the features of an image into k vectors. First of all the key points are selected that are the points present in high contrast region. Then the features of those key points are extracted and are represented in an array format. An array is formed containing features of all the key points that are detected from an image.

The array formed for representing feature information is of the size K x 128 , where k is the number of key points present in an image and 128 is the number of features that represent a particular key point properties.

Another array is also formed that gives the information about the location of the key point in the frame containing the hand posture. This vector size is K x 4. Here each key point is represented by four points. The four values of each key point represent the row, column, scale and orientation of each key point in the frame. The K x 128 matrix is known as the descriptor array whereas the K x 4 matrix is termed as the location array.

These arrays are used for the recognition process. The trained database images and the input images are first converted into key point readable form and then the features of each key point are represented with the help of these arrays. When the

matching process begins, it will check how many key point vectors are matching and hence how many features are matching with each database image. In this way the image having maximum matched key points is further checked for the exact posture recognition.

3.2.4 MK-ROD ALGORITHM

After the k vector is formed and the matching is done, the output will be the number of key points matched with each database image for the input image. The image that is having the maximum number of matched key points is then further suspected for the exact recognition. For this validity ratio is calculated. For getting the validity ration Mk-Rod algorithm is used.

This ratio is found out between two images that have maximum key points matched. The two images having similar key point features is then checked for the relative distance and the position of the key point relative to each other and the distance and position of each key point from the center of the image. Hence in this way we can found out if the key points that are matching between both images are actually at same position or not. If there relative position is matched then only we can say that the two images are same.

Distance mask is used to mask the distances that are below the algorithm threshold after taking the absolute of the difference between the ratios of two images.

This process is done to determine if the pattern of the matched key points is similar or not. This is determined by finding the relative distance of the matched key points from the center. Hence it will determine if the pattern of key points is also matching or not.

The absolute of the distance of the matched key points from the center is taken and the matched key point that have absolute distance value below the algorithm threshold value is treated as the valid matched point that is the matched key point that is also at the similar position as that of the database image.

3.3 MEDIA PLAYER CONTROLING

In this thesis we proposed a system that uses point pattern matching algorithm to recognize the hand gesture and the matched images are used to control the media player. We have to control the basic functions of the media player like play and stop. This we have to do with the help of hand recognition. We have to take two gestures that are used to play and stop the media player when appears at the input of the HGR system. Hence our system is able to control the application without manually pushing the play and stop buttons. ASL postures are used here for hand recognition and hence through this we can control other applications also like gaming, robots etc. hence it

will provide more interactive system between human and the computer or any other controller or machine.

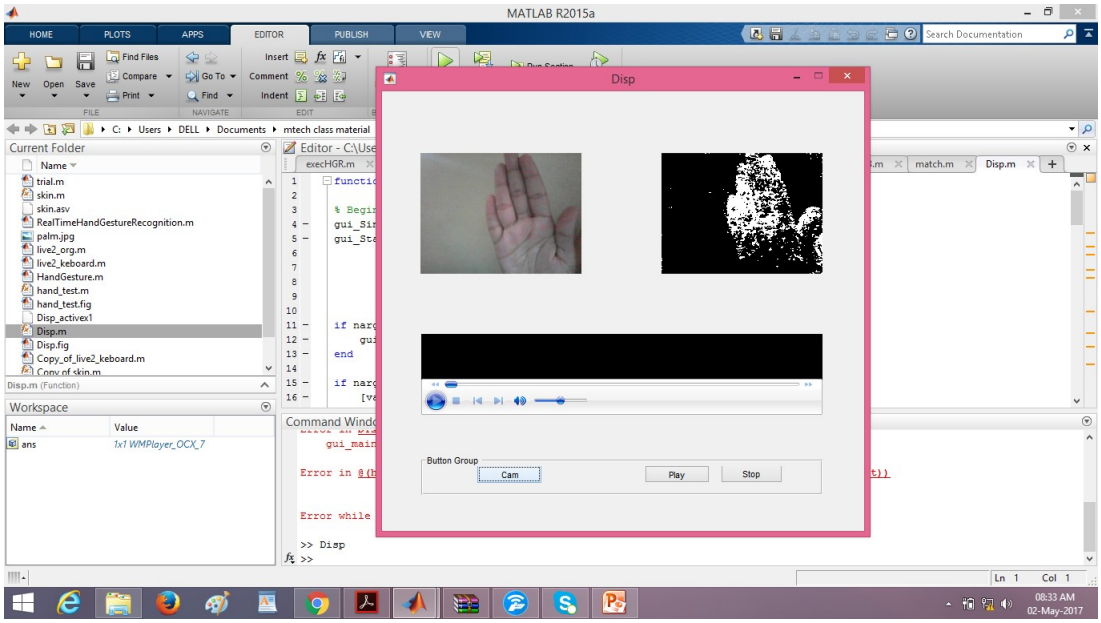

Figure 3.5: Media Player Interfaced

Chapter 4

IMPLEMENTATION

4.1 IMPLEMENTING POINT PATTERN MATCHING ALGORITHM

The algorithm is used for hand recognition. During testing phase, the input image from folder is selected and its key points are extracted. Then during the matching phase, its key points are matched with key points of all other database images by using point pattern matching algorithm. In this way the matches ASL gesture is recognized from the database images. After this the output is given in the form of matched image and also the corresponding ASCII character of the gesture. The algorithm is having two parts

1. SIFT implementation
2. Mk-Rod algorithm implementation

4.1.1 IMPLEMENTING SIFT ALGORITHM

For any image, feature can be described by extracting the point of interest from that image. Features of that point represent the information of the image. This feature description extracted from the input image can then be used to match it from all other database images. If we have to develop more accurate and reliable and robust system then it must be able to detect those features in any light condition, noise, or scale compositions. Also the relative position of the key points from each other should remain same. SIFT is used to detect the key points and their features from the image. It detects a large number of key points hence the points having low contrast can also be eliminated. The SIFT algorithm constitutes of four parts-

- A scale space is constructed
- Approximating the Log scale
- Key points detection
- Eliminating low contrast and edges key points
- Orientation assignment to key points
- SIFT feature generation

During the implementation, point pattern matching algorithm uses the SIFT algorithm to locate the key points and then to extract the features from the key points. These key points are having scale variant feature and are located at the high contrast region of the image.

In this thesis, two images are taken to match; first one is from the input image folder and another from the database folder. Distance ration is a parameter of the SIFT algorithm. Generally this is kept constant but in this thesis we are making it variable for recursivity.

Threshold is the parameter of the second algorithm used that is Mk-Rod algorithm. It is also a variable parameter. Here for finding the key points SIFT function is called. It finds the key points in addition with the features. These features are represented by image descriptors and the location array. These terms are described below:

1. IMAGE (im): this gives the image to be tested in double array format.

2. DESCRIPTORS (des): it is a matrix array of size K x 128, where K represents the number of key points extracted and for each key point there 128 descriptors or features. Hence each row represents a descriptor vector having 128 values normalized to unit length.

3. LOCATION(loc): it is given by a matrix array of size K x 4. Where each row represents four values of the key point related to its location.

These values are row, column, scale and orientation. The orientation is in the range of [-PI, PI] radians.

The SIFT is implemented in following steps:

SCALE INVARIANT FEATURE DETECTION

In this process the image is converted into collection of feature vectors, these should be scale invariant, rotation invariant, independent of illumination. Key locations are the maxima and minima of the result of the Gaussian function used in scale space to the images.

DIFFERENCE OF GAUSSIAN (DOG)

From the input image, the scale space Gaussian function is constructed. For this the image has to go from the convolution or filtering stage with Gaussian function of variable widths. Here $D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$ $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$. In the equations, D is the difference between two filtered images, one is k multiplied by the scale of other, calculated as Gaussian difference. I is the input image and G is the Gaussian function given by $G(x, y, \sigma) = -exp(-(x^2 + y^2)/2)$

EXTREMA DETECTION

In this stage the extreme points of DOG pyramid is found out. Each point of image is compared with its neighbor point to compute the local maxima and minima of the D function. If the value is minimum or maximum then it is called as extrema. Then the improvement of the location of the key point is done by using a second order Taylor series expansion. From this we will get the exact extrema locations.

This is given as $Y = -(d^2z/dx^2)$ $Z = 1/y(dD/dx)$ Here D and its derivatives are evaluated at the sample points and x is the offset from the sample point.

KEY POINT LOCALIZATION AND ELLIMINATION

In this stage the scale space extrema points are identifies and selected as the key points for the feature detection. Then this stage eliminates some points from all the

Figure 4.1: Feature of image extracted using SIFT

key points extracted for a particular image. These points are the points present at the low contrast region of the image or are poorly localized like present at the edge of the figure. Because the points present at the low contrast region would be sensitive towards noise. The value of the key points that are present at the extreme positions of the DOG pyramid is given as $D(z) = D + (1/2)(dD^{-1}/dx)z$ Now for the elimination of poorly localized extrema points we are using the fact that in the direction of difference of Gaussian function, the points across the edge are having large principal curvature as compared to the points in the perpendicular direction of the difference of the Gaussian function.

ORIENTATION ASSIGNMENT

Based on the local image properties the key points are now assigned a particular orientation. The magnitude of the gradient m, and the orientation, are pre-computed using pixel differences $m(x, y) = sqroot((L(x+1, y)–L(x−1, y))^2+(L(x, y+1)–L(x, y−1))^2)$ $\theta(x, y) = arctan((L(x, y + 1) − L(x, y − 1))/(L(x + 1, y) − L(x − 1, y)))$

We will select the peak point of the histogram and assign an orientation to it. The points coming under 80% of the height of the peak are also assigned the same orientation. Some points are also assigned multiple orientations if for the same magnitude there are multiple peaks.

These all processes compute the orientation, location, and scale of the key points in the image that are selected after elimination.

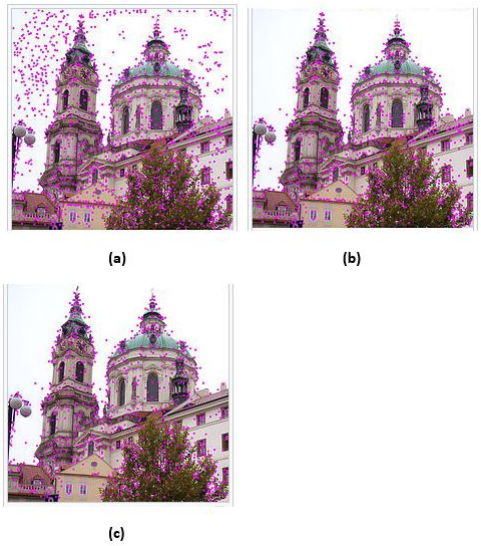

Figure 4.2: (a) All key points extracted (b) Only high contrast key point after eliminating low contrast key points (c) After removing low contrast edge located key points

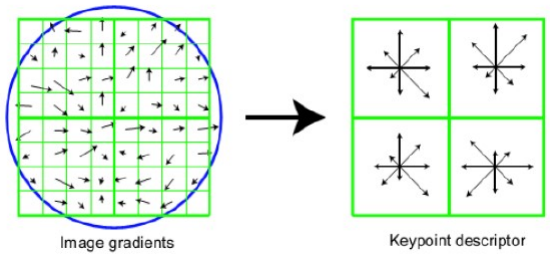

Figure 4.3: Assignment of orientation to key points

4.1.2 IMPLEMENTING Mk-Rod ALGORITHM

After the k vector is formed and the matching is done, the output will be the number of key points matched with each database image for the input image. The image that is having the maximum number of matched key points is then further suspected for the exact recognition. For this validity ratio is calculated. For getting the validity ration Mk-Rod algorithm is used.

This ratio is found out between two images that have maximum key points matched. The two images having similar key point features is then checked for the relative distance and the position of the key point relative to each other and the distance and position of each key point from the center of the image. Hence in this way we can found out if the key points that are matching between both images are actually at same position or not. If there relative position is matched then only we can say that the two images are same.

The below figure represent the two images having matched key points and that are now tested with the validity ratio through Mk-rod algorithm.

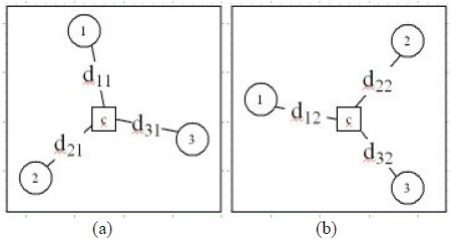

Figure 4.4: Mk-Rod algorithm

In the above figure-

- C represents the center point of the image
- D represents the distance mask
- T represents the number of the images that has to be matched
- M represents the number of matched points between the two images. These matches points are the subset of the key point of both the images that are exactly matched between the two images

Among the above two images one is the input image which is to be matched or to be recognized and the other one is the image from the database that has maximum number of key points matched with the input image.

Image (a) represents the trained database image and image (b) represents the input image to test with the key points.

The procedure of finding the validity ration between the two given images is described below.

- d(T1) = summation (from i=1 to i=M) d (i1)
- d(T2) = summation (from i=1 to i=M) d(i2)
- $Ratio1 = [d11/dT1 \;\; d21/dT1 \;\; d31/dT1]$
- $Ratio2 = [d12/dT2 \;\; d22/dT2 \;\; d32/dT2]$
- Distance mask,$D = absolute[Ratio1–ratio2] \leq (Thresholdvalue)$ • Valid points = sum (Distance mask)
- Validity ratio = number of valid points / number of matched points

4.2 IMPLEMENTATION OF OTHER METHODS

4.2.1 TEMPLATE MATCHING

In this method there is a stored set of images. Template matching determines whether the input image's gesture can be classified under the stored set of gestures of images or not. HGR using template matching can be done in two steps. First one is creating the templates by collecting the data values of the gestures from the stored image set. Each image goes through this process a few number of times after applying the slight

natural variations in the image. Second part is comparing the input image with the stored template of the gesture set and obtains the template closely matching with the given input image.

Figure 4.5: Hand posture templates

4.2.2 PRINCIPAL COMPONENT ANALYSIS (PCA)

In PCA the principal component of any image is its Eigen vectors. So in PCA the Eigen vector of every image is computed and hence image is represented in the form of its Eigen vectors. In this method the dataset is formed with the help of images having very high resolution so that minimum number of database is needed for recognition. Then the mean is calculated and subtracted from the data dimension. Then in the next step covariance matrix is calculated for the principal Eigen vector of the database. Covariance matrix represented by C for a matrix A is calculated as

$$C = A * A'$$

Then the principal components of the data set are the Eigen values and the Eigen vectors of the matrix C. Among these components the best components are chosen for the feature vector. Then the Eigen vector with which the data is presented with minimum information loss is selected. Then finally the new data set is formed that is now used to compare with the data set of other images containing best Eigen vectors. The comparison process here involves Euclidian distance calculation between the coefficients of Eigen vectors.

Method	Time	Scale	Rotation	Illumination
SIFT	Good	Best	Best	Common
PCA-SIFT	Good	Common	Good	Good
SURF	Best	Good	Common	Best

Figure 4.6: Comparison of SIFT, PCA and SURF

4.3 IMPLEMENTING MEDIA PLAYER

in this thesis media player is implemented to be controlled by the HGR. The HGR algorithm is able to recognize all 36 gestures of ASL. Among these some signs are used to control the media player. In this thesis the play and stop features are controlled

through the hand recognition. First of all a GIF file is created in the MATLAB software. In this a media player is interfaced.

With this two push buttons are introduced in the GIF. These are named as PLAY and STOP. The functions of these push buttons are defined. The play and stop signs of media player are not used because we have to control them with the recognition program.

The whole system works in the following manner

- The input image to be testes is first converted into key points.
- The features of key points are extracted.
- The input image is matched with the database images and the image having maximum matched points is selected.
- Then this database image and the input image are checked for the validity ratio which compares the relative distance between the key points of both images.
- Hence this will complete the recognition phase and provides the exact matched image.
- Among all the database images, two are selected to control the media player.
- When the recognition phase detects any one of these image then the command goes to the media player controller.
- If the command is for play then it will call the PLAY pushbutton callback function and then the media player starts playing.
- If the command is for stop then it will call the STOP pushbutton callback function ad hence the media player will stop.

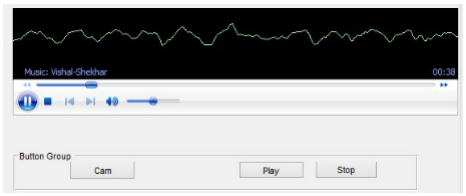

Figure 4.7: Media Player Implemented

4.4 HAND DETECTION THROUGH WEBCAM

Though in this thesis we are not recognizing the hand posture taken from the webcam, we have done with the detecting of hand posture from the webcam.

Here another GIF file is formed in which webcam is interfaced. When the GIF file is called then the webcam opens and captures the image of the hand. Then this image is converted to the grey format simultaneously when the image is taken. Also we have done the same thing for a real time video. The camera inputs are

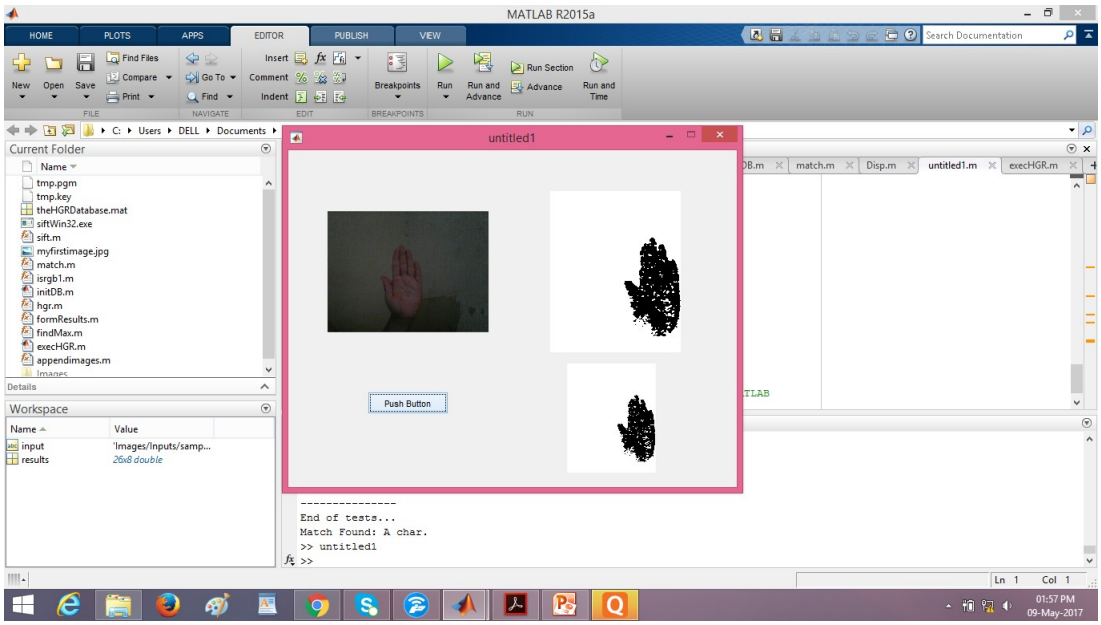

Figure 4.8: Hand detection GIF: captured, converted recalled image after saving

simultaneously converted into grey format continuously. Then this captured and grey converted image automatically get saved into one of the folder from where it can be given to the recognition algorithm.

The problem we are facing for such type of real time hand detection and recognition is the resolution. The resolution of pre-processed database image is different from the image we are taking from the webcam. And further converting it to the grey format disturbs its resolution. For exact recognition, the input images should have the same resolution as the database images in this algorithm. Hence this system is only suitable for recognition of existing images. And we can only detect and save the image from the webcam through this work but cannot accurately recognize them.

Chapter 5

RESULTS, CONCLUSION AND FUTURE WORKS

5.1 RESULTS

5.1.1 RESULTS OF HAND RECOGNITION IMPLEMENTATION

The database image having maximum validity ratio is selected and said to be matched with the test input image. Then the output is displayed by appending both images side by side that is the matched database image and the input test image. Here we have taken input images as the ASL gestures. Hence the matched image will represent the character matched. Below figures from 5.1 to 5.14 represents the output for the characters matched.

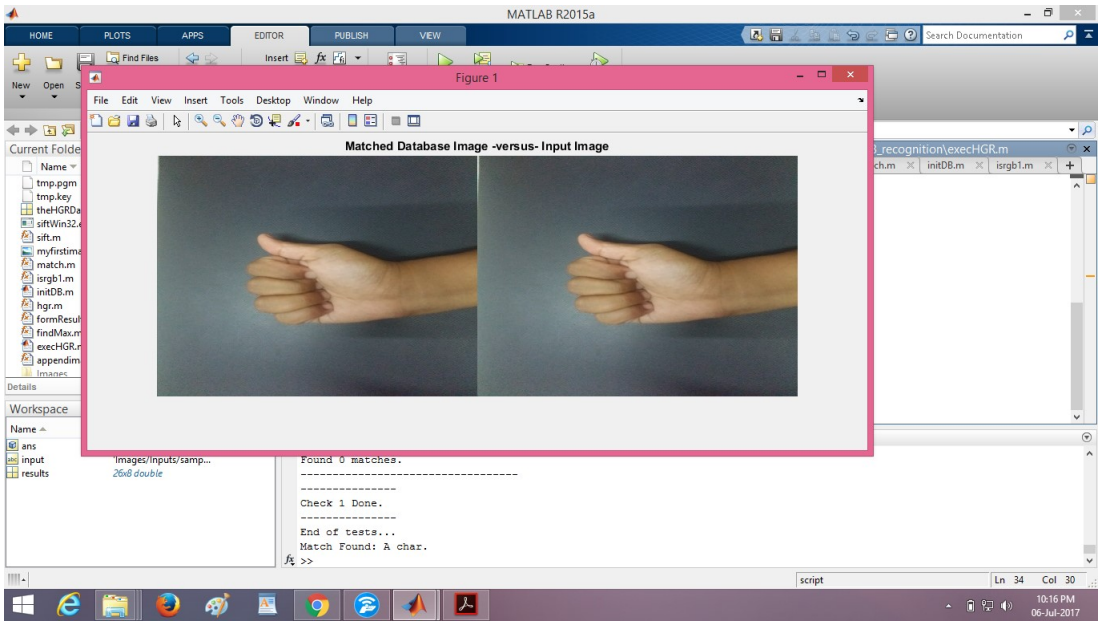

Figure 5.1: 'A' character matched

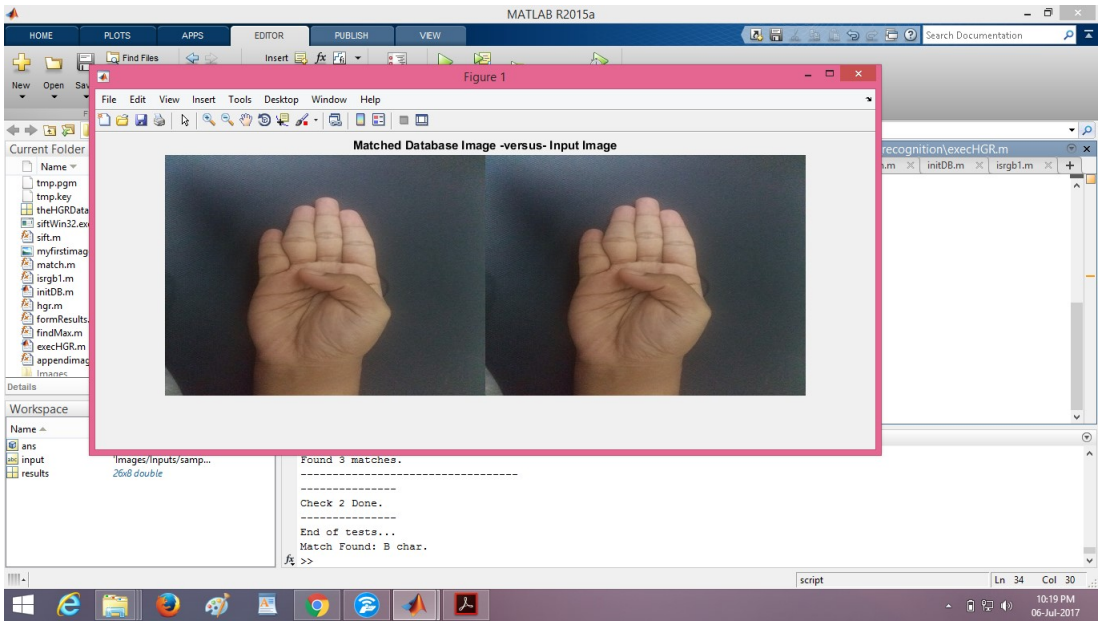

Figure 5.2: 'B' character matched

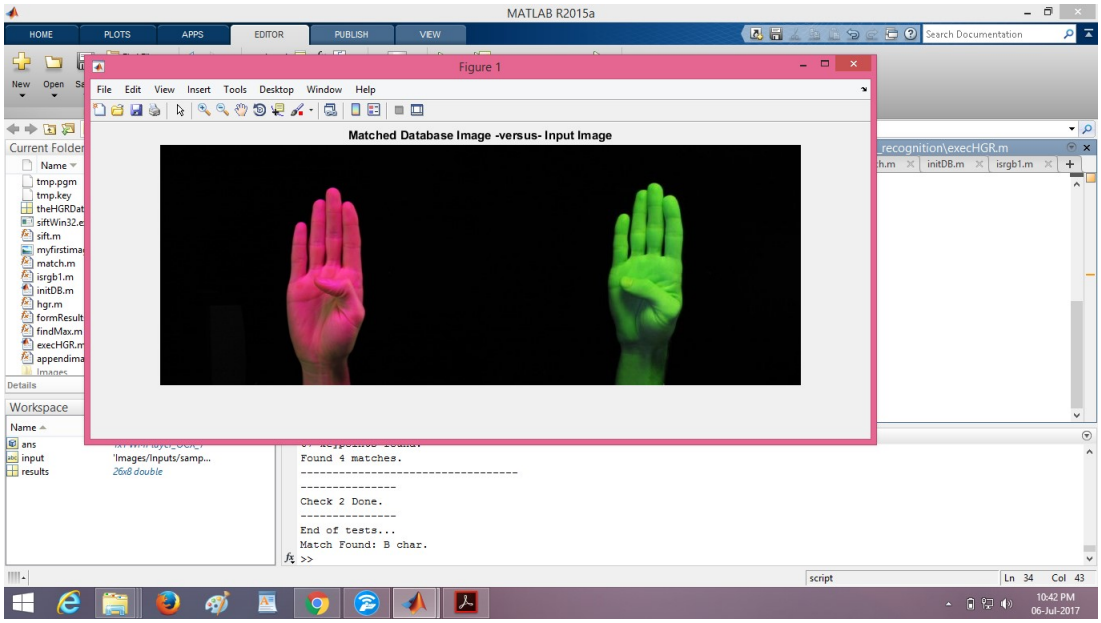

Figure 5.3: 'B'character matched with different color

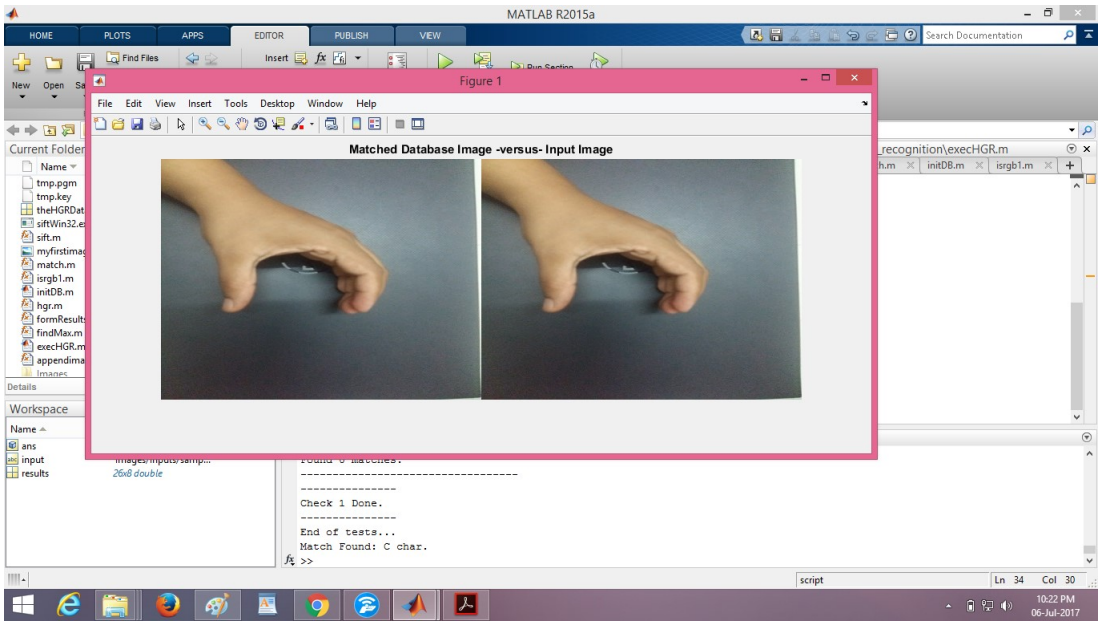

Figure 5.4: 'c' character matched

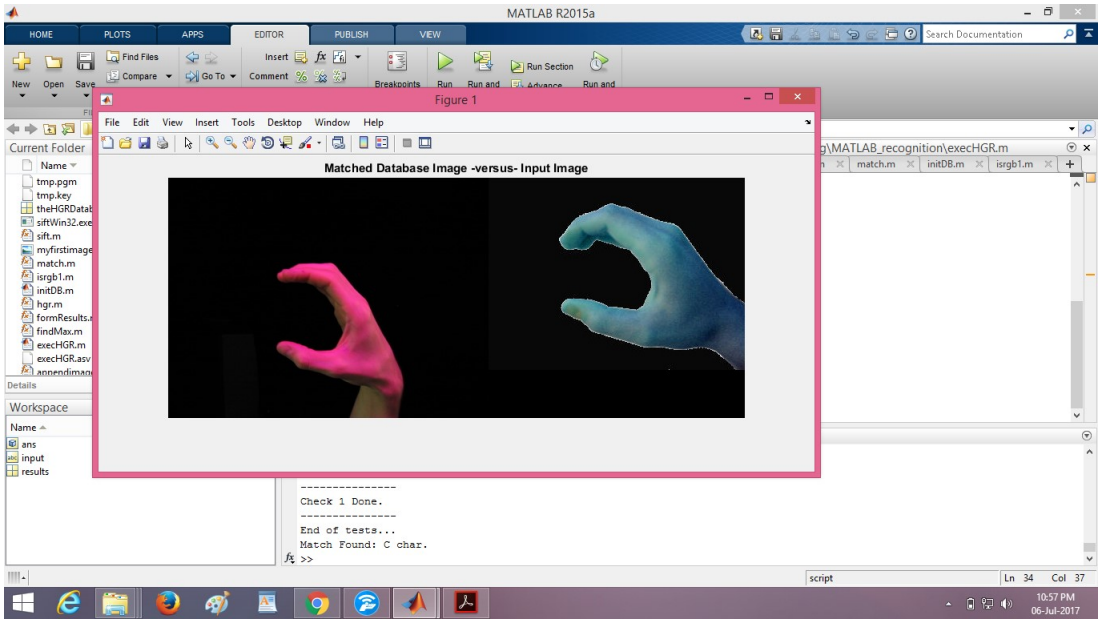

Figure 5.5: 'c' character matched with different color and shape

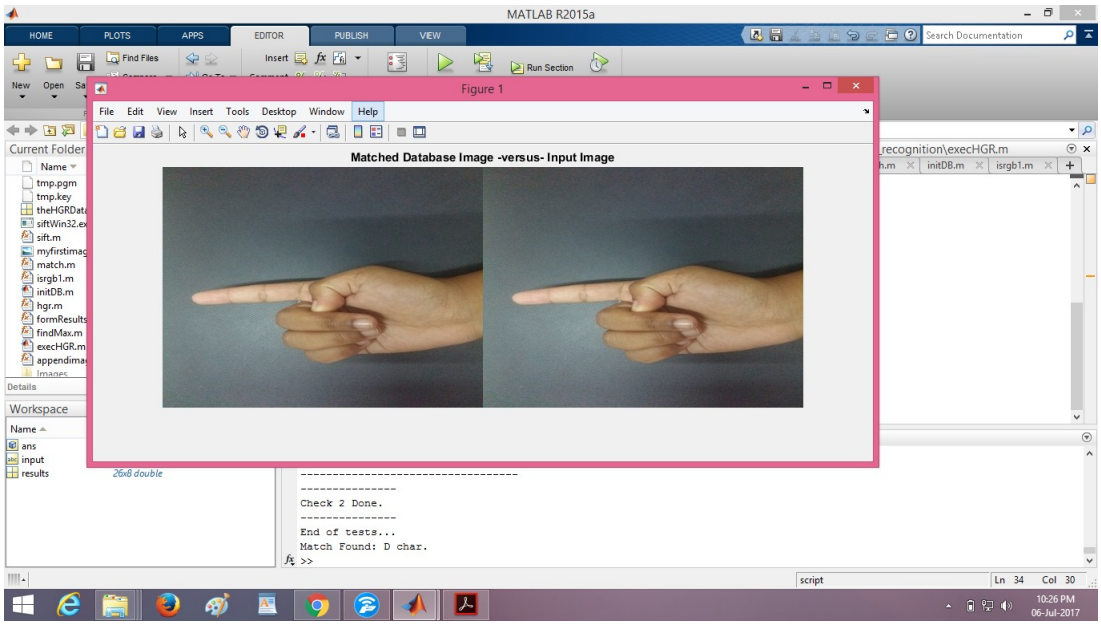

Figure 5.6: 'D' character matched

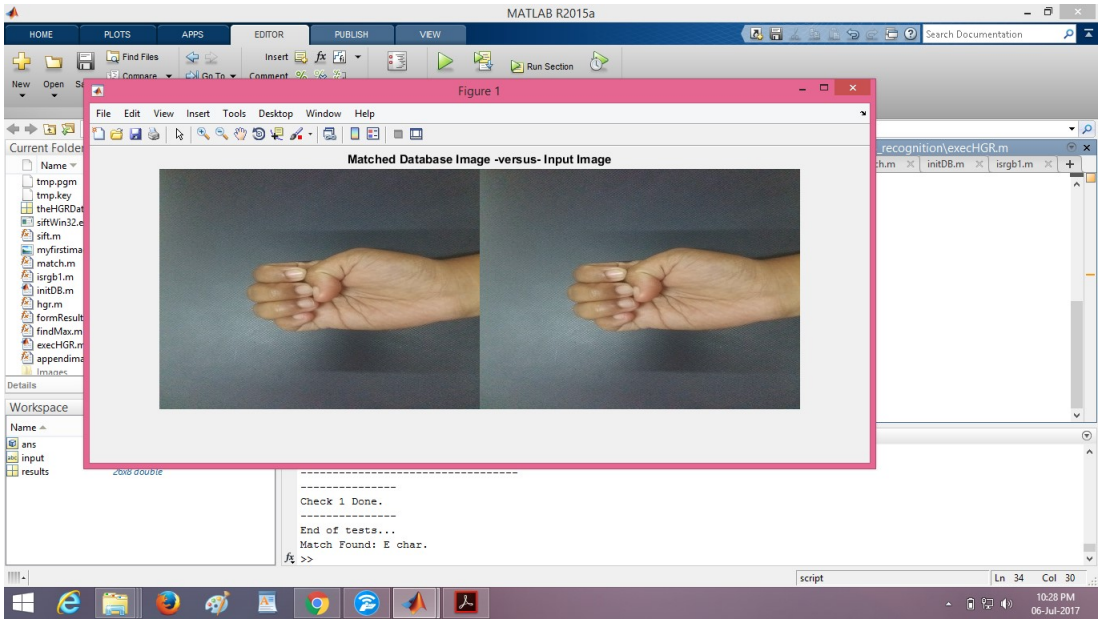

Figure 5.7: 'E' character matched

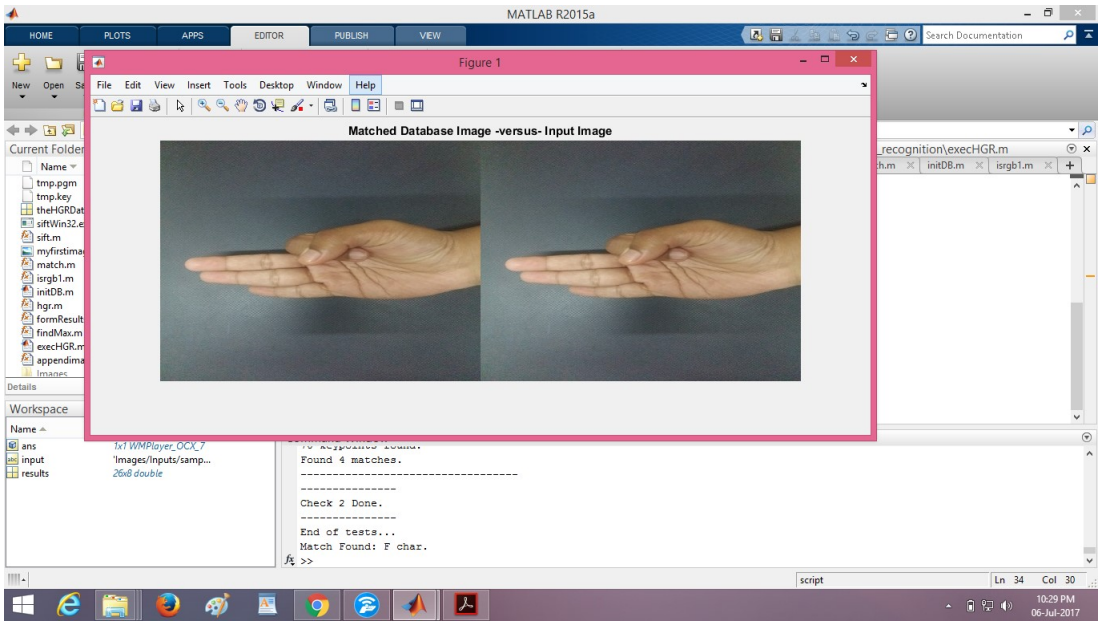

Figure 5.8: 'F' character matched

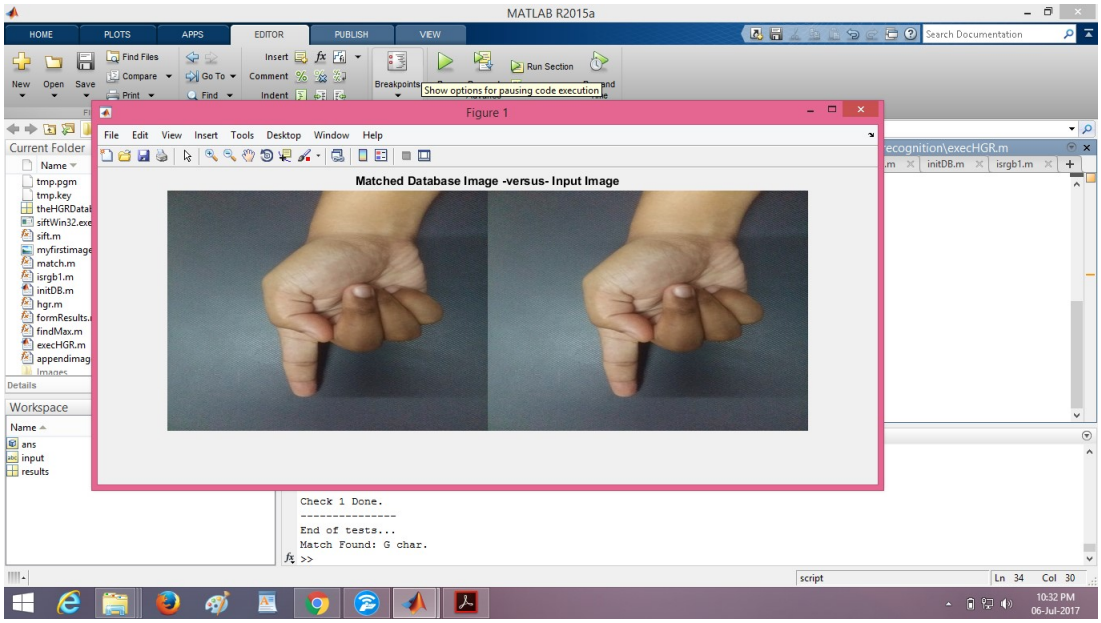

Figure 5.9: 'G' character matched

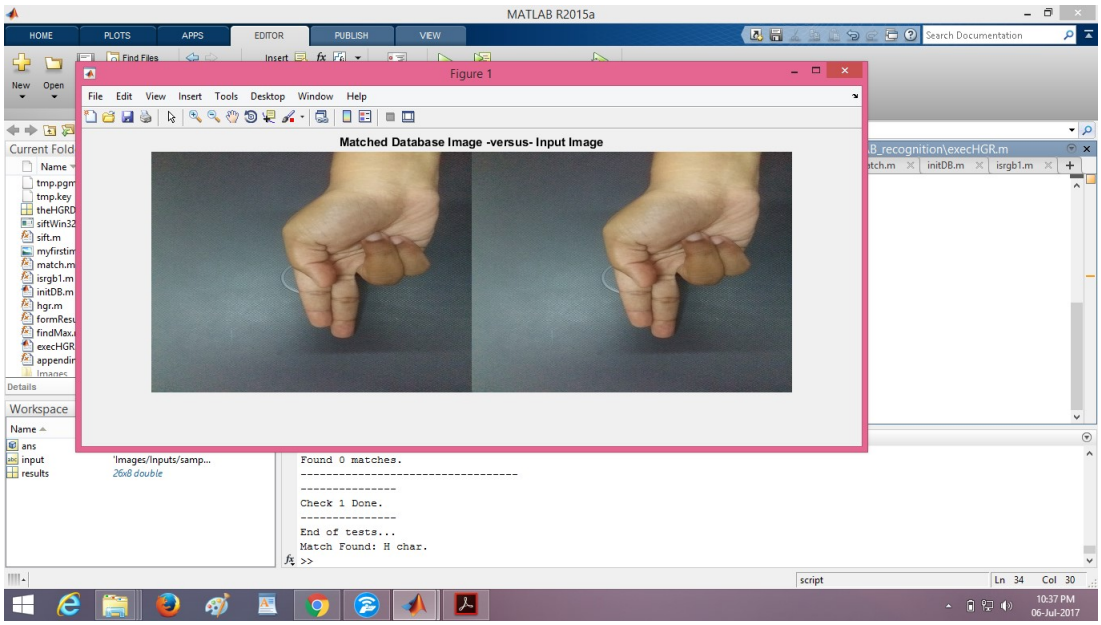

Figure 5.10: 'H' character matched

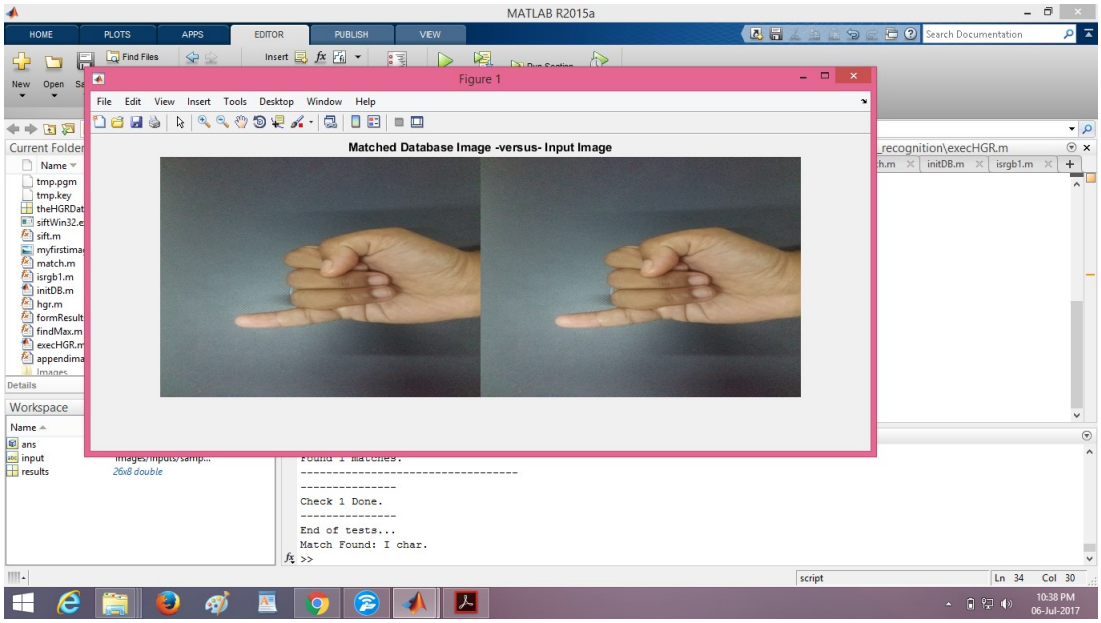

Figure 5.11: 'I' character matched

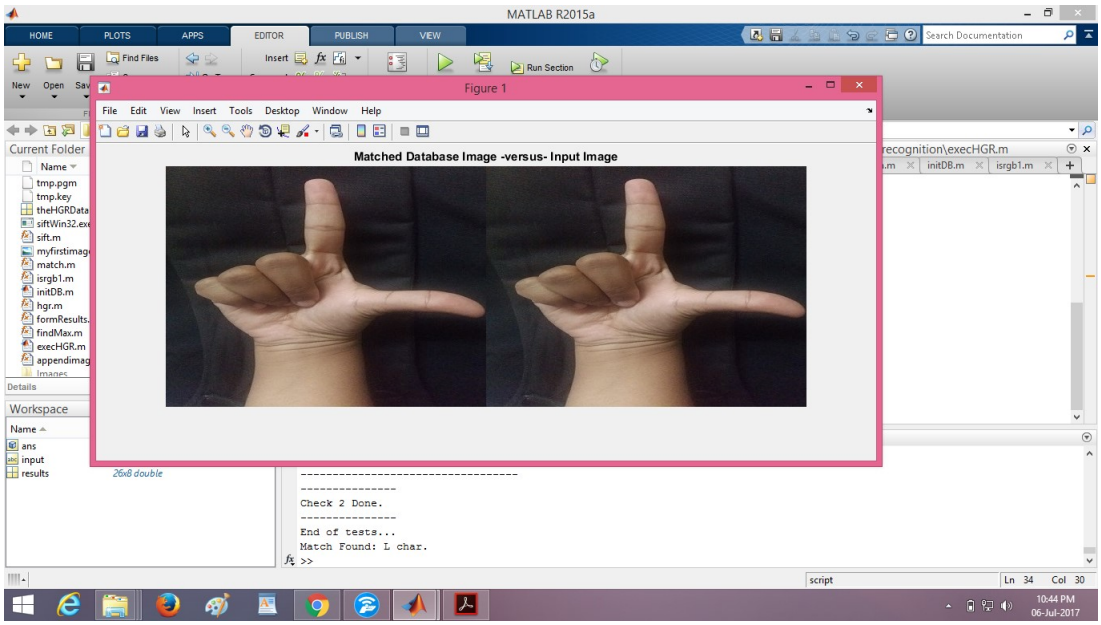

Figure 5.12: 'L' character matched

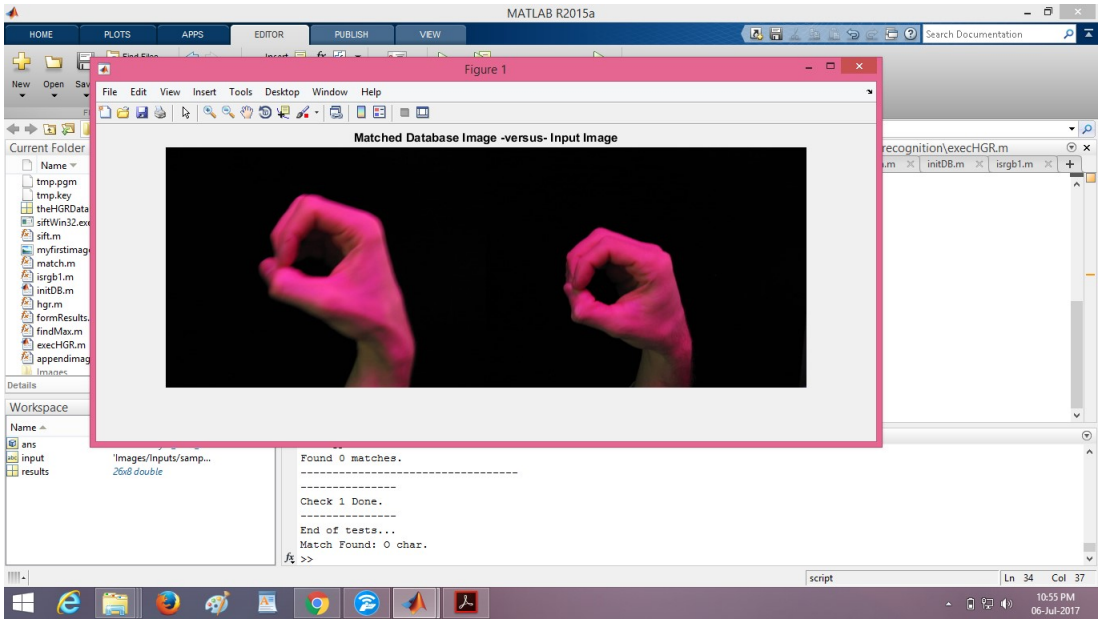

Figure 5.13: 'O' character matched

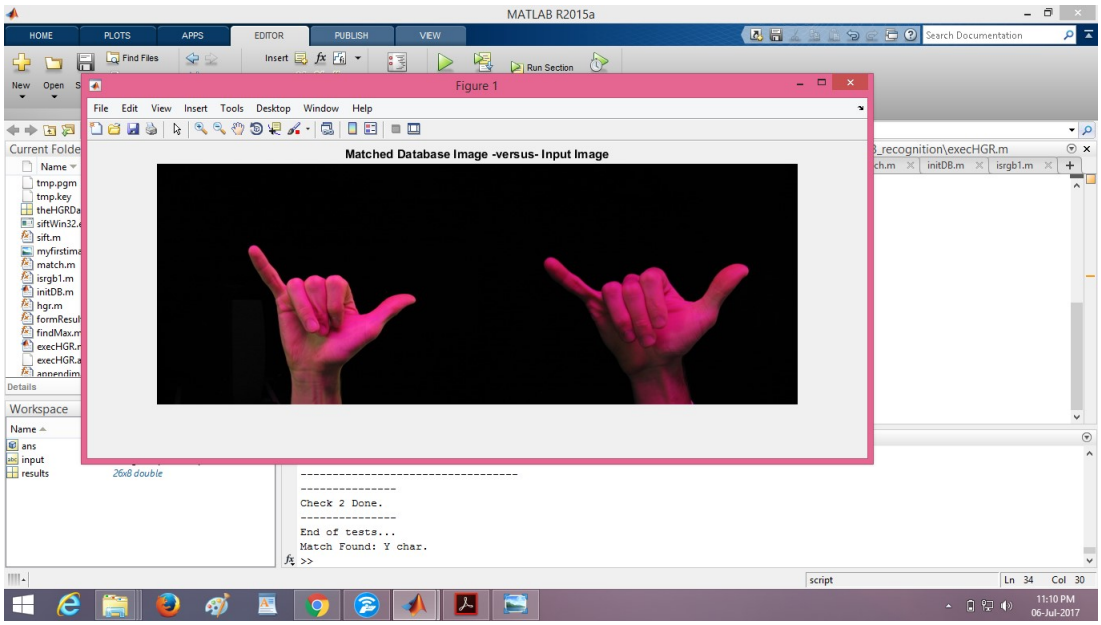

Figure 5.14: 'Y' character matched

5.1.2 RESULTS WITH THE MEDIA PLAYER CONTROLLING

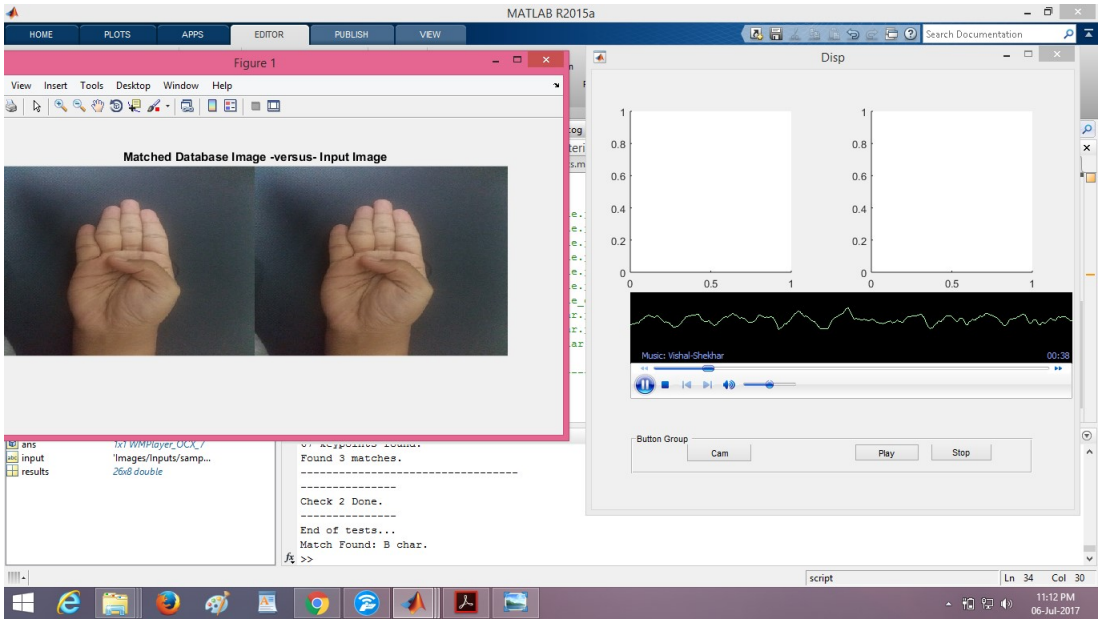

Figure 5.15: Media player playing when 'B' character matched

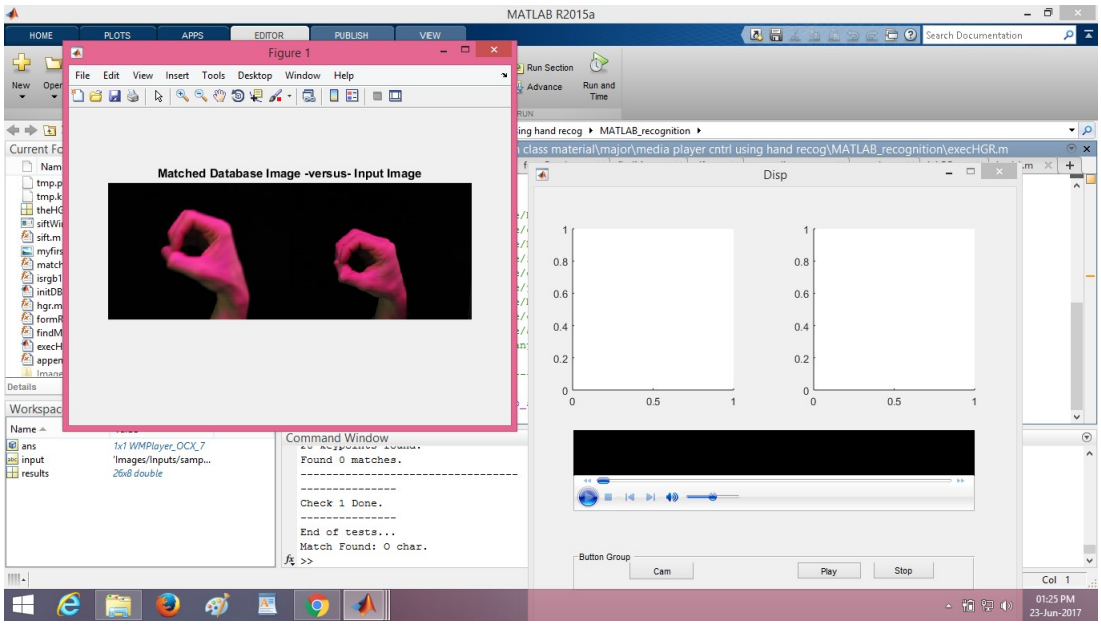

Figure 5.16: Media player playing when 'O' character matched

5.2 CONCLUSIONS AND FUTURE WORKS

The main objective to create such system is to provide a more interactive environment for human and computer interface. The HGR or HPR methods work by taking the input from human gestures. In these human interact with computer in more natural way as they communicate with other human beings. So it will require less effort in communication. Therefore these type of communications need to be progresses and researched.

Nowadays many researches are going on in this area to make these types of systems more effective and popular so that it can be used in day to day life. Also these systems help the handicapped persons in operating any device that is based on input taken from gestures.

Real time vision based systems are most challenging part of research in this area. This is because real time environment suffers with change in many factors such as lighting, illumination, background, rotation, scale etc. For making such real time systems the algorithm used should be more effective and trained images must complete all the aspects. This thesis completes the recognition of all the gestures of ASL and our system is also capable of controlling media player application without any human interference. It also covers the detection of hand from the webcam and converting and storing the captured image or video in grey format. But this does not include recognition of the webcam input image.

The HGR algorithm used in this thesis uses SIFT algorithm for features extraction and matching and also uses Mk-Rod algorithm for detecting the relative distances and find the validity ratio for exact recognition. Hence this HGR system is capable

of recognizing all static postures of ASL correctly using the point pattern matching algorithm with SIFT. This algorithm provides a simple and robust way for recognizing the patterns. Comparing this matching algorithm with other methods conclude that this method has higher rate of recognition than the template matching algorithm. It is also capable of dealing with the environment changing factors like illumination, scaling, orientation, viewpoint etc. this feature is not present in algorithms like template matching and PCA. In this method the gestures that are similar to each other can also be recognized separately if good quality and high resolution images are used.

Extension fir this work can be making this same system real time where media player can be controlled instantaneously by recognizing image taken through webcam. This system needs to be fast and should take very less response time. Another extension can be detecting and recognizing two hand postures at the same time. This system will be helpful in applications like gaming and robotics. The more advanced reseach can be 3D image processing in the real time environment.

Chapter 6

References

[1] Deval G Patel, “Point Pattern Matching algorithm for recognition of 36 ASL gestures”, International Journal of Science Modern Engineering (IJISME), Volume 1, Issue 7, June 2013.

[2] Talla Bhuvana Vaibhav Reddy, Vijaykumar Gooli, Stephen R, Yogeeswar Reddy P, Shivprakash.C, “Senso less virtual talking system for differently ABLED ”, International Journal Of Engineering Research in Electronics Communication Engineering (IJERECE), Vol 3, Issue 6, June 2016.

[3] Pallavi Gurjal, Kiran Kunnur, “Real time Hand Gesture Recognition using SIFT”, International Journal Of Electronics Electrical Engineering, Vol 2, Issue 3, March 2012.

[4] Hamid A Jalab, Herman K Omer, “Human Computer Interface Using Hand Gesture Recognition Based on Neural Network”.

[5] Reza Azad, Babak Azad, Nabil Belhaj Khalifa, Shahram Jamali, “Real Time Human Computer Interaction based on face and Hand Geasture Recognition”, International Journal in Foundation of Computer Science and Technology(IJFCST), Vol 4, No.4, July 2014.

[6] Manuj Paliwal, Gaurav Sharma, Dina Nath, Astitva Rathore, Himanshu Mishra, Soumik Mondal, “A Dynamic Hand Gesture Recognition System For controllimg VLC Media Player”.

[7] Siddharth S Rautaray, Anupam Agrawal, “Real Time Multiple Hand Gesture Recognition system for Human Computer Interaction”, Published Online in May 2012 in MECS.

[8] MS Kalyani P Sable, DR. H R Deshmukh, Prof N S Band, Prof R N Gadbail, “Pattern Matching Algorithm for ASL Sign Gesture Recognition Using Neural Network”, International Journal Of Pure And Applied Research In Engineering And Technology (IJPRET), May 2014.

[9] Prof Sugandhi S, Prof Redappa BC, “Hand Gesture Recognition System For Creating and Controlling Media Player Using MATLAB Tools”, IOSR Journal Of Computer Engineering (IOSR-JCE), Vol 18, Issue 4, Jul-Aug 2016.

[10] Siddharth Swarup Rautaray, Anupam Agrawal, “A Noval Human Computer Interface Based on Hand Gesture Recognition using Computer Vision Techniques ”.

[11] Mrs AR Patil, Dr SS Subbaraman, “A Review on Vision Based Hand Gesture Recognition Approach using Support Vector Machines”, IOSR Journal of Electronics And Communication Engineering (IOSR-JECE).

[12] Shahzad Malik, Chris McDonald, Gerhard Roth, “hand Tracking for Interactive Pattern based Augmented Reality”, National Research council Canada, Published in IEEE and ACM International Symposium on Mixed and Augmented Reality, Oct 2002.

[13] X Zabulis, H baltzakis, A Argyros, “Vision based Hand Gesture Recognition for Human Computer Interaction”, Computer Science Dept, University of Crete Heraklion, Crete, Greece.

[14] Jun Yang, Yu Gang Jiang, Alexander G Hauptmann, Chong Wah Ngo, “Evaluating Bag-Of-Visual-Words Representations in Scene Classification ”, Published in Proceedings of International workshop on Multimedia Information Retrieval.

[15] M Hasanuzzaman, Tao Zhang, V Ampornaramveth, M A Bhuiyan, Yoshiaki Shirai, H Ueno, “Gesture Recognition for Human-Robot Interaction through a Knowledge based Software Platform”.