Hand Gesture Recognition System for Human-Computer Interaction with Web-Cam

Narendra V. Jagtap[1] Prof. R. K. Somani[2] Prof. Pankaj Singh Parihar[3]
[1, 2, 3]Department of Computer Science & Engineering
[1, 2, 3]Institute of Technology & Management, Bhilawara (RTU), Rajasthan, India

Abstract—This paper represents a comparative study of exiting hand gesture recognition systems and gives the new approach for the gesture recognition which is easy cheaper and alternative of input devices like mouse with static and dynamic hand gestures, for interactive computer applications. Despite the increase in the attention of such systems there are still certain limitations in literature. Most applications require different constraints like having distinct lightning conditions, usage of a specific camera, making the user wear a multi-coloured glove or need lots of training data. The use of hand gestures provides an attractive alternative to cumbersome interface devices for human-computer interaction (HCI). This interface is simple enough to be run using an ordinary webcam and requires little training.

I. INTRODUCTION

Body language is an important way of communication among humans, adding emphasis to voice messages or even being a complete message by itself. Thus, gesture recognition systems could be used for improving human-machine interaction. This kind of human-machine interfaces would allow a human user to control remotely through hand postures a wide variety of devices. Different applications have been suggested, such as the contact-less control or home appliances for welfare improvement.

In order to be able to represent a serious alternative to conventional input devices like keyboards and mice, applications based on computer vision like those mentioned above should be able to work successfully under uncontrolled light conditions, no matter what kind of background the user stands in front of. In addition, distorted and joints objects like hands mean an increased difficulty not only in the segmentation process but also in the shape recognition stage.

Most work in this research field tries to solve the problem by using markers, using marked gloves, or requiring a simple background. Other approaches are based on complex representations of hand shapes, what makes them unavailable for their implementation in real – time applications. A new vision-based framework is presented in this paper, which allows the users to interact with computers through hand postures, being the system adaptable to different light conditions and backgrounds. Its efficiency makes it suitable for real – time applications. The present paper focuses on the various stages involved in hand posture recognition, from the original captured image to its final classification. Frames from video sequences are processed and analysed in order to remove noise, find skin tones and label every object pixel. Once the hand has been segmented it is identified as a certain posture or discarded, if it does not belong to the visual memory. This paper proposes a better way in which the background picture is taken in the beginning afterwards the background picture is subtracted from the picture to detect the area of interest, which makes it easier to recognize gestures.

In today's world the best means of Human Computer Interaction (HCI) is keyboard and mouse, this conventional input devices are very easy to use and are easily available and easy to learn.

But today there is very less way disabled people can communicate with machine, this lead to development of a new kind of system which makes disabled people to easily communicate with the system, Hand Gesture Recognition System would be the best means for disabled people to communicate with system.

II. GESTURE RECOGNITION

If we remove ourselves from the world of computers and consider human – human interaction for a moment we quickly realize that we utilize a broad range of gestures in communication. With respect to objects, we have a broad range of gesture that are almost universal, including pointing at objects, touching or moving objects, changing object shape, activating objects such as controls, or handling objects to others. This suggests that gestures can be classified according to their functions. This suggests that gestures can be classified according to their functions:

1) Semiotic: Those used to communicate meaningful information.[2] [3] [4]
2) Ergotic: Those used to manipulate the physical world and create artifact.[2] [3] [4]
3) Epistemic: Those used to learn from the environment through tactile or haptic exploration. [2] [3] [4]
4) Within these categories there may be further classification applied to gestures. We are primarily interested in how gestures can be used to communicate with a computer so we will be mostly concerned with empty-handed semiotic gestures. These can be further categorized according to their functionality as:
5) Symbolic Gestures: These are gestures that within each culture have a single meaning. An emblem such as the "OK" gesture is one such example.[2] [3] [4]
6) Deictic Gestures: These are the types of gestures most generally seen in HCI and are the gestures of pointing, or otherwise directing the listener's attention to specific events or objects in environment. They are gestures made when someone says "Put That There".[2] [3] [4]
7) Iconic Gestures: As the name suggests, these gestures are used to convey information about the size, shape or orientation of the object of discourse. They are gestures made when someone says "The Plane Flew Like This", while moving their hand through the air like the flight path of the aircraft.[2] [3] [4]

8) Pantomimic Gestures: These are the gestures typically used in showing the use of movements of some invisible tool or object in speaker's hand. Like when a speaker says "I turned the steering wheel hard to the left" while making the action of turning a wheel with both hands.[2] [3] [4]

III. EXITING METHODS FOR GESTURE RECOGNITION

With the help of serious improvement in image capturing and processing technology, hand gestures become an easy and popular tool in human machine interaction systems. Recently, human machine interfaces are based on and limited to use keyboards and mice with some additional tools such as special pens and touch screens. Although these devices are well designed to interact with machines and are used in daily life, they are not perfect for natural quality of human communication. Main application areas of hand gesture recognition in human machine interface systems are keyboard – mouse simulations, special game play without joysticks, sign language recognition, 3D animation, special HCI for disabled users, etc. Although in daily life people usually do not want to touch buttons or touch screens in public areas like screens in planes or buttons in automatic teller machine (ATM) because of hygienic consideration, hand gestures would be an ideal replacement in that manner. Many researchers have proposed numerous methods for hand gesture recognition systems. Generally such systems are divided into two basic approaches namely –

1) Glove – based approach
2) Vision – based approach

In glove – based analysis, detection of the hand is eliminated by the sensors on the hand. Such systems are optimal for body motions capture purposes and widely used in industry. On the other hand vision – based analysis is more natural and useful for real – time applications.

A healthy human can easily identify a hand gesture, however for a computer to recognize hand gesture, first the hand should be detected in the acquired image and recognition of that hand should be done in a similar way as humans do. Yet this is a more challenging approach to implement because of the limitations of such a natural system.

Fig. 1: Glove with Sensor

Detection and gesture analysis of the hands is a growing topic and has many user environment limitations for most of the studies. Segmentation of the hand is the first step of such systems. Exceptionally, such systems in [5], [6] and [7] gloves or finger marks have been used to extract the hand gesture information in the frame and ease the hand detection process by eliminating the different skin colour issue problem. This technique allows the system to detect hands in a straightforward manner and it is more robust to change in lightning conditions and it is also independent of the user's skin colour.

Easiness is carried out running the hand gesture recognition system in front of a simple and uniform background like a black curtain [8]. Such systems need to distinguish skin colour but since background could be estimated easily, it would be very easy to segment the hand region from the background. On the other hand, such systems involving gloves or uniform background ruin the natural behavior of gesture applications by limiting the user environment.

In recent years, with the introduction of a new approach [9], which has a high detection rate, new studies are mostly concentrated on Boosting and HMM. The most positive side of using these methods is that they usually work with gray scale images instead of coloured images and thus it eliminates the drawbacks of such colour based noise issues. This innovative approach is using a well-known technique namely Adaboost classifier which was mentioned in [10].This feature extraction technique does not need skin colour information and have less computation time with the use of integral image concept. But the drawback of this method is that it requires a training process. This process often needs huge-sized sample images to have a high detection rate.

Fig. 2: Hand Gesture with Black Background

According to hand detection method, coloured images are investigated in 2 steps namely coarse skin colour segmentation and fine skin colour segmentation. For coarse skin colour segmentation fixed skin colour thresholds in nRGB colour space are used (skin locus).Many studies in literature use different skin locus thresholds for different colour spaces to locate faces or hands in images. [11] Starts with an RGB image and by apply a dimension reduction algorithm to propose its own skin locus in two dimensions to compare its performance with HSV skin locus. [12] And [13] compare HSV/HSI, RGB, TSL and YCbCr colour space skin locus performances. [14]starts the segmentation of skin colour in YUV colour space to have a quick result and then tune the current skin colour with a quasi-automatic method which needs some user input. In [15], chrominance

along with luminance information in YCbCr colour space was used to segment the skin colour for current conditions and histogram clustering is used for fine skin segmentation.

Reference	Primary Method of Recognition	Number of Gestures Recognized	Background to gesture Images	Additional Markers Required	Number of Training Images
18	Hidden Markov Models	97	General	Multi – colored gloves	400
17	Entropy Analysis	6	No	No	400
19	Linear approximation to non – linear point distribution models	26	Blue Screen	No	7441
20	Finite State Machine Modeling	7	Static	Marker on glove	10 sequences of 200 frames each
21	Fast Template Matching	46	Static	Wrist Band	100 examples per gesture

Table. 1: Comparison of existing methods of recognition

For hand gesture recognition part of this study, HMM or Adaboost classifier type training based methods have very limited usage because of the non-strict structure of the hand. To recognize a gesture once need to train positive images (include the defined gesture) and negative images (do not include the defined gesture). But negative images have a serious role at this point. Since many hand poses might yield similar training data, reliance to the training data would be limited. So an adaptation of a well-known hand gesture recognition method was implemented in this study. According to the proposed methods in prior studies [16] and [17] centroid profile extraction of the hand is extracted around the centre of the palm and histogram clustering is applied to the resultant data to recognize the gesture. Such an algorithm would typically count the number of fingers being shown to the camera. This can capture 6 gestures for each hand namely 1, 2, 3, 4 or 5 fingers or no fingers (punch) conditions. In this paper, an adaptation of that method is proposed and a higher correct detection rate is provided.

There is a limited amount of studies in literature for the hand gesture recognition. Recognition methods, like in the detection procedure, are mainly relying on algorithms, which need training or different environmental constraints. A clear summary of such algorithms is shown in table.

A. HAND SEGMENTATION:

Detection is one of the most challenging steps in many image-processing systems. To analyse and extract valuable information from the acquired image, one needs to find the desired data in the entire set of pixels. Since the hand has not a strict shape like face, hand gesture recognition has less spot then face detection and recognition in literature. Most systems avoid handling hand systems because of this fact. To detect the hand(s) in the image a two steps system was designed. First, skin colour locus for the current is extracted for the user's skin color, lightning condition and camera parameters. Then as the second step, eliminating false positive skin pixels and identifying hand(s) and other real skin colour regions detect hand. Different proposed techniques to detect hand in images are as follows:

B. HAND GESTURE RECOGNITION SYSTEM:

Hand gesture recognition is an improving topic, which has many applications areas in human computer interaction systems. It is believed that future systems will have so many frameworks however applications will be less useful. In modern systems design and interaction are crucial keywords to attract consumer attention and gesture recognition is bringing an advanced interaction environment in many systems. People would want to control sound or stop/play features with hand gestures while they are watching movies on their home theatre systems or many people do not want to touch screens in ATMs while they are drawing money or some disable people cannot use mouse or keyboards but they can still have some defined gestures for themselves to control computers. As you see in these examples, there are millions of application areas of hand gesture systems. It is very critical that those systems would work in environment free conditions and they would work in an efficient way to recognize gestures. Today's systems have many limitations especially for environmental conditions. We have proposed new methods to decrease the dependency to lightning conditions, camera parameters or user specific environment to detect hands.

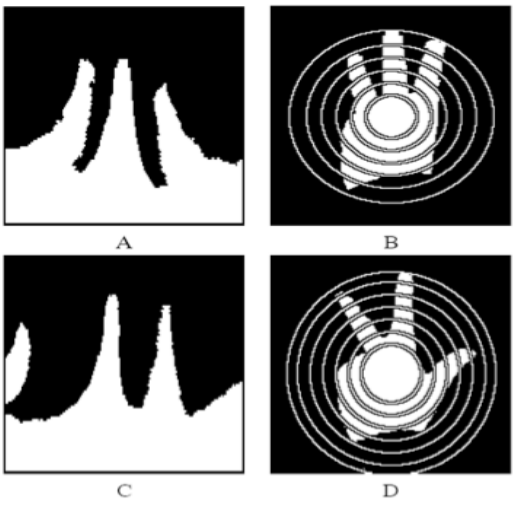

Fig. 3: Polar Transformation of Hand [16]

Once the hand is located in the acquired image, rest is regarding the gesture recognition procedure. Hands are searched by a special profile extraction technique and gesture is estimated. The proposed method for gesture recognition in this study is based on a procedure called centroidal profile extraction which was mentioned in [17] and [16]. According to the centroidal profile extraction

method, growing circles are drawn around the palm of the hand. Each circle is considered as a contour to move on and a polar transformation is used to count the number of fingers being shown. If a point on a circle is a skin pixel then the corresponding angle on the Number of Skin Pixels vs. Angle graphic will be increased by one.

IV. PROPOSED HAND GESTURE RECOGNITION SYSTEM DESIGN

Software design is a process of problem – solving and planning for a software solution. It includes low-level component and algorithm implementation issues as well as the architectural view. Design is a meaningful engineering representation of something that is to be built. It can be traced to a customer's requirements and at the same time assessed for quality against a set of predefined criteria for 'good' design.

A. SYSTEM ARCHITECTURE DESIGN:

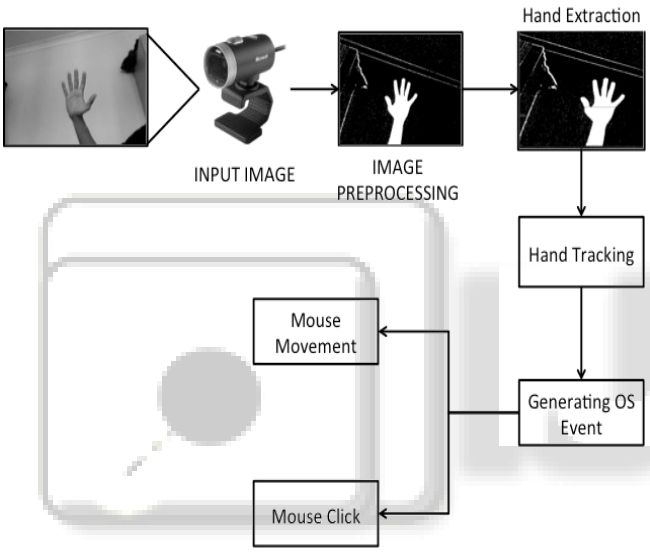

Fig. 4: System Architecture

B. COMPONENTS:

1) Input Image:

Get the image from the web cam to be processed for detection of hand

1) Role: Act as the source image to be processed for recognition of gestures
2) Function:
   a. Get the video stream from real world
   b. Store a frame of video into image
3) Output: The output would be a single image, which would be further processed for gesture recognition.

   a. *Image Processing:*

4) Role: This part deals with processing the color image into binary version for further hand detection.
5) Input: The color image is given as input for processing.
6) Function:
   a. Get the color image
   b. Convert it into black and white image
7) Output: The output would be a binary image, which is further send for hand detection.

C. Hand Extraction:

1) Role: This part is responsible for extracting the hand region from the image captured and converted to binary format.
2) Input: The input would be the pre – processed image, which would be in binary format.
3) Function:
   a. The binary image is received from previous phase
   b. This image is processed to check the hand region in image.
   c. The hand region is selected as origin.
4) Output: The output would be the hand region, which would be used as a point of origin to control the mouse events.

D. Hand Tracking:

1) Role: This part is responsible for tracking the point of origin to track the position of the hand in the image.
2) Input: The input would be the binary image with point of origin recognized.
3) Function:
   a. The binary image with point of origin is received.
   b. This point of origin is tracked in the image to calculate the position of mouse.
4) Output: The output would be the position of hand in the image to calculate the position of mouse.

   a. *Generating OS Event:*

5) Role: The role of this part would be to generate OS Events for controlling the mouse movement.
6) Input: The input would be the position of hand in the image
7) Function:
   a. The hand position is recognized
   b. This position is used to generate OS events
8) Output: The output would be the OS events either the mouse movement or mouse click generated by system.

E. Requirements Traceability:

To ensure the proper running of the system the requirements traceability is extremely important. The next table lists all the requirements mapped by each system component used in the Gesture Recognition System.

Architecture Components	Requirements addressed by the Components
Web Cam	A normal web cam is required to capture video stream
Computer System	A system with minimum following requirements are needed: OS Supported: 98/ME/NT4/2000/XP/2003 Hardware Requirements: Intel Pentium III 850MHz or Above, or any compatible Athlon 128 MB RAM or above Webcam VGA or higher resolution

Table. 2: system component used in the Gesture Recognition System

V. SOFTWARE ARCHITECTURE DESIGN

Based on the hierarchical composite property of hand gestures, it is natural to use a divide-and-conquer strategy to break hand gestures into their constituent hand postures and movements. To implement this strategy, a two – level architecture shown in Figure is proposed. This architecture differentiates hand gesture recognition into two levels: the low – level hand posture detection and tracking and the high-level hand gesture recognition and motion analysis.

Because there is no obvious structural information for hand postures, it is more appropriate to use a statistical approach for the low-level of the architecture. Two components are included in this approach: the posture detection and tracking component. The high-level of the architecture is responsible for gesture recognition and motion analysis. A syntactic approach is selected to fully use the composite property of hand gestures. The detected hand postures and motion tracking from the low-level are sent to the high-level of the architecture as primitives for syntactic analysis so that the whole gesture can be recognize. There are two components included in the high-level of the architecture: the local finger motion analysis component and the global hand motion analysis component.

The block diagram of the two-level architecture is shown in Figure. At the low – level of the architecture, hand gesture images are collected from the web-camera this web-camera provides video capture with maximum resolution of 640 × 480 up to 15 frames-per-second.

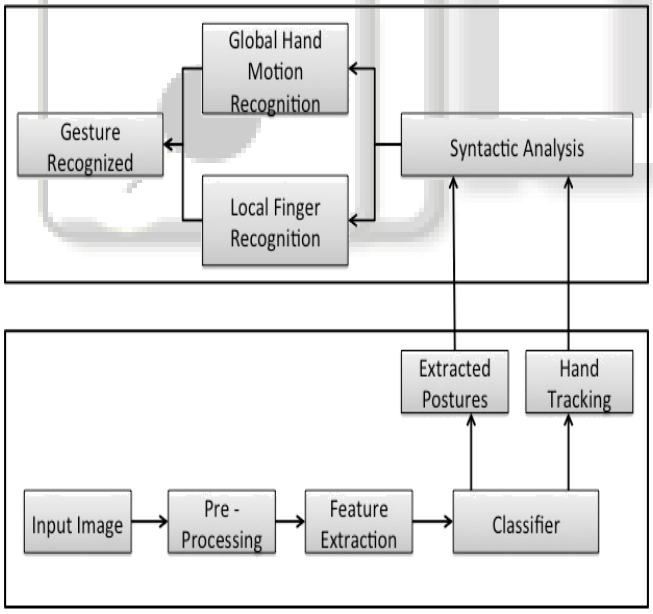

Fig. 5: Software Architecture

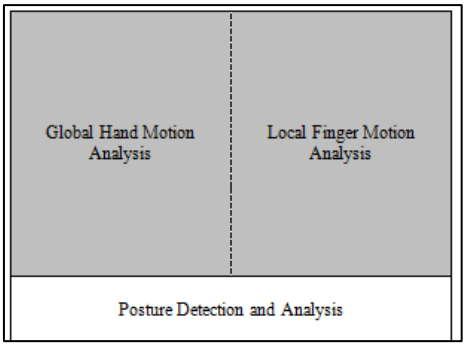

Fig. 6: Level Architecture of System

VI. CONCLUSION

A new hand gesture recognition system which works under most lightning conditions with different skin colored users and with different camera parameters is aim. It should not need any training or not make the user wear a special glove etc. Also the system will aim to work in or nearly real time to be applicable in human computer applications. Finally it should work in a typical PC with a cheap USB webcam.

REFERENCES:

[1] The Site [Online]. Available at: http://jareddonovan.com/index.html

[2] Gesture-based communication in human-computer interaction: 5th International Gesture Workshop, GW 2003: Genova, Italy, April 2003: selected revised papers

[3] Fels, S. S. and Hinton, G. E. (1993). Glove-Talk: A Neural Network Interface between a Data glove and a Speech Synthesizer. IEEE Transactions on Neural Networks, 4, 2-8.

[4] Fels, S. & Hinton, G. (1995). Glove-Talk II: An adaptive gesture-to-formant interface. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'95), 456-463.

[5] Oya Aran, Cem Keskin, Lale Akarun. Computer Applications for Disabled People and sign Language Tutoring. Proceedings of the Fifth GAP Engineering Congress, 26 – 28 April 2006, Sanhurfa, Turkey.

[6] Oya Aran, Lale Akarun. Recognizing Two Handed Gestures with Generative, Discriminative and Ensemble Methods via Fisher Kernels. Multimedia Content Representation, Classification and Security International Workshop, MRCS 2006, Istanbul, Turkey.

[7] Aykut Tokatlı, Ugŭr Halıcı. 3D Hand Tracking in Video Sequences, Middle East Technical University.

[8] B. Ionescu, D. Coquin, P. Lambert, V. Buzuloiu. Dynamic Hand Gesture Recognition using the Skeleton of the Hand. EURASIP Journal on Applied Signal Processing 2005:13, 2101-2109, Hindawi Publishing Corporation.

[9] P. Violaand M. Jones. Rapid object detection using a boosted cascade of simple features. Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, 2001.

[10] Y. Freund and R. Schapire. A decision theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci., vol.55 no.1, pp. 119-139, August 1997.

[11] C. H. Kim, J. H. Yi. An Optimal Chrominance Plane in the RGB Color Space for Skin Color Segmentation. International Journal of Information Technology vol.12 no.7, pp.73-81, 2006.

[12] L. Sabeti, Q. M. Jonathan Wu. High-Speed Skin Color Segmentation for Real-Time Human Tracking. IEEE International Conference on Systems, Man and Cybernetics, ISIC2007, Montreal, Canada, 7-10 Oct. 2007.

[13] J. C. Terrillon, S. Akamatsu. Comparative Performance of Different Chrominance Spaces for Color Segmentation and Detection of Human Faces

in Complex Scene Images. Proc. of Vision Interface '99, pp. 180-187, Trois-Rivieres, Canada, 19-21 May 1999.

[14] S. Askar, Y. Kondratyuk, K. Elazouzi, P. Kauff, O. Scheer. Vision-Based Skin-Color Segmentation of Moving Hands for Real-Time Applications. Proc. Of. 1st European Conference on Visual Media Production, CVMP, London, United Kingdom, 2004.

[15] A. Albiol, L. Torres, E. J. Delp An Unsupervised Color Image Segmentation Algorithm For Face Detection Applications. Proc. of International Conference on Image Processing 2001, 7-10 Oct 2001.

[16] Moritz Storring, Thomas Moeslund, Yong Liu, and Erik Granum. Computer Vision- Based Gesture Recognition For an Augmented Reality Interface. In 4th IASTED Inter- national Conference on Visualization, Imaging and Image Processing, pages 766-771, Marbella, Spain, Sep 2004

[17] Jae-Ho Shin, Jong-Shill Lee, SeKee Kil, Dong-Fan Shen, Je-Goon Ryu, Eung Hyuk Lee, Hong-Ki Min, Seung Hong. Hand Region Extraction and Gesture Recognition Using Entropy Analysis, IJCSNS International Journal of Computer Science and Network Security, Vol.6 No.2A, February, 2006.

[18] Bauer, Hienz. Relevant feature for video-based continuous sign language recognition. Department of Technical Computer Science, Aachen University of Technology, Aachen, Germany, 2000.

[19] Bowden, Sarhadi. Building temporal models for gesture recognition. In proceedings British Machine Vision Conference, pages 32-41, 2000.

[20] Davis, Shah. Visual gesture recognition. In proceedings IEEE Visual Image Signal Process, vol.141, No.2, pages 101-106, 1994.

[21] R. Lockton, A. W. Fitzgibbon. Hand Gesture Recognition Using Computer Vision. B.Sc. Graduation Project, Oxford University

[22] U. Ahlvers, U. Zolzer, R. Rajagopalan. Model-free Face Detection and Head Tracking with Morphological Hole Mapping, EUSIPCO'05, Antalya, Turkey

