A SURVEY ON GESTURE RECOGNITION TECHNIQUES IN GESTURE-BASED HUMAN COMPUTER INTERACTION INTERFACES

Sunanda Biradar[1], Ashwini M Tuppad[2]

[1,2] *Dept. of Computer Science & Engg., BLDEA's VP Dr. P G Halakatti Engg. College, (India)*

ABSTRACT

With the advent of newer techniques for the design of human computer interaction interfaces, there has been a surge in modern technologies that enable more intuitive form of communication forms. This is in contrast to the traditional console and graphical user interfaces, which restricts the process to physical device input. With availability of powerful platforms coupled with end user's demands, user interfaces are coming up with convenience incorporated within for its users. One such interface becoming more widely used nowadays is gesture based interface, where user's gestures mostly using the hands, are recognized and mapped into appropriate commands for computer operation. This paper introduces the concept of gesture recognition and reviews the literature on gesture recognition approaches, technologies utilized to implement the same to meet today's demands of human computer interaction.

Keywords: Gesture, Interface, Intuitive, Recognition, Taxonomy

I INTRODUCTION

Computers have evolved from a long history of technological advances and attempts by the designers and developers of new ways of its usage, and in proving their usefulness and ease of operation to the end users. In any field, computers need to have some sort of communication by its users and vice-versa.

The user interfaces are the means that make possible human computer interaction(HCI). Of these the traditional ones were the cumbersome console interfaces, that proved to be inconvenient for the general end users. They usually tend to have lack of computer command-specific knowledge and weren't used to the idiosyncracies inherent in command line consoles. But the developers were very well versed in it. Next, came the graphical user interfaces (GUIs) bringing revolution in computer industry by making possible Human Computer Interaction smooth and appealing to non-developer end user community due to features like icons, windows, menus and a mouse as a pointing device that could select a graphical object and navigate as per the user's choice. Again GUIs limited the scope of interaction to keyboard and mouse or such sort of physical devices. Others HCI interfaces include switch interfaces, which used buttons, audition interfaces in the form of beeps, alarms, turn by turn navigation commands when using GPS and haptic devices that generate sensations from body parts like skin [14]. The gesture-based HCI interfaces have begun to flourish and are the main research sources in this area.Interaction between any two agents, either both humans or between a human and a machine can be accomplished in two modes. First mode is verbal communication that is in terms of written or vocal

natural language sentences. The second being non-verbal communication [1], which can take the forms of sign language, gestures made by hands or other body parts, body language etc.

Hand gestures are widely used in gesture recognition systems. Again the hand gestures are categorized as static ones that involve motionless hand gestures and the dynamic category involves gestures made while hand is in motion [1]. To understand gesture based interface, the prime requirement obviously has to be understanding gestures, their interpretation, gesture enabling technologies, system response of input gestures and the application domain for which this interface is being designed [14].

II BACKGROUND

This section will throw light on the essential concepts involved in understanding different classes of gestures. In order to interpret the hand gestures by a hardware or software system, it must primarily detect the gesture based on certain features of the hand. Thus the anatomy of human hand is necessary, which will also be dealt with here.

1. Gesture Style Classification

Classification of anything becomes very easy and better if a taxonomy is presented. [14] presents a well-organized gesture style taxonomy based on the means used to make the gestures.

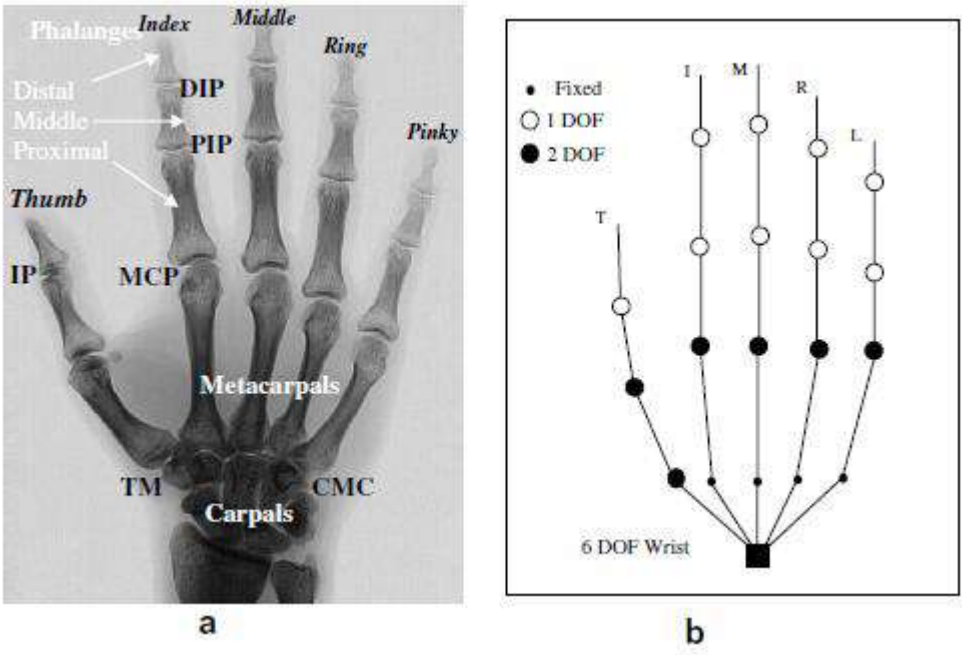
- Deictic gestures: Characterized by pointing using fingers, which would be a direction towards the position of an entity on a display. Their common use is to move the entity over the screen or display virtually within the problem domain.
- Gesticulation: Constitutes hand movements along with user's speech in order to interpret one's gestures. It would be more advantageous as speech and hand movements will result in more clearer understanding, but both inputs need to be synchronized.
- Manipulative gestures: Involve mapping of movements of the arm onto a location on coordinate system determined by the movements. The coordinate system refers to one that is internally used by the computer system display.
- Semaphoric gestures: Those that use hand or arm symbols that are universal across communities of people in their interpretation.
- Language gestures: Involve a set of hand gestures exclusively for one particular language, whose sequences represent distinct grammatical structures. There can be many different linguistic gesture systems for many different languages.

The interface must be designed based on application domain within which it has to work and based on suitable technologies for recognition and mapping gestures into outputs.

2. Anatomy of the human hand

Anatomy is a word from biology, which is a branch that relates to the study of the structure of a body part. As the focus of this survey is on hand gestures, structure of hand and its features must be understood thoroughly. The reason behind this is, while detection of gestures the features or signals generated by the hand guide the

process. They act as inputs to the hand gesture recognition techniques, which will ultimately be interpreted correctly and mapped onto commands for computer operation



1. a: Hand gesture anatomy 1. b: Hand kinematic model.

Fig.1. structure of human hand from [2]

The skeleton of the human hand consists of 27 bones: the eight short bones of the wrist or carpus organized into a proximal, which articulates with the skeleton of the forearm, and a distal row, which articulates with the bases of themetacarpal bones (i.e. the bones of the palm or "hand proper"). The heads of these Metacarpal bones will each in turn articulate with the bases of the proximal phalanx of the phalanges. These articulations result in the formation of themetacarpophalangeal joints, which are colloquially referred to as the knuckles of a clenched fist. The fixed and mobile parts of the hand adapt to various everyday tasks by forming bony arches: longitudinal arches (the rays formed by the finger bones and their associated metacarpal bones), transverse arches (formed by the carpal bones and distal ends of the metacarpal bones), and oblique arches (between the thumb and four fingers). Of the longitudinal arches or rays of the hand, that of the thumb is the most mobile (and the least longitudinal). While the ray formed by the little finger and its associated metacarpal bone still offers some mobility, the remaining rays are firmly rigid. The phalangeal joints of the index finger, however, offer some independence to its finger, due to the arrangement of its flexor and extension tendons [14].

The hand kinematic model [2] depicts the degrees of freedom, the number of directions of movements at that point. The study of the kinematic model is important as it helps to know the kinds of motion of hand and the parameters that characterize any movement, in turn a dynamic gesture features.

II GESTURE RECOGNITION : APPROACHES AND TECHNIQUES

Gesture recognition process itself includes a number of stages, starting from hand gesture image acquisition with final stage being the mapping of recognized gesture into appropriate computer commands, which will

achieve the goal of interacting with computers using gestures. The fig.2 shows the stages involved in this process.

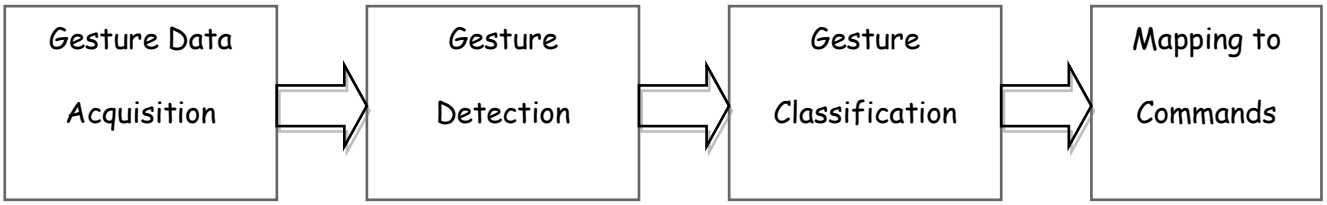


Fig.2. Gesture Recognition Process Stages

2.1 Gesture Data Acquisition

The first stage is the acquisition of data representing the gestures made by the human's hand. Gesture data acquisition accomplishment is made possible by hardware driven acquisition approach, like the sensors in the form of electrical signals or by computer vision approach where a video camera would capture an image(s) of gesture(s).

- *Hardware driven acquisition approach*

The gesture data is in the form of signals acquired by sensors held or worn on hand. Electromyogram (EMG) sensors, accelerometers, data gloves are some common means of getting signals generated by the hand while gesturing in the form of equivalent electrical signals.

Accelerometers are the electromechanical sensors that percept large, distinct gestures formed by the trajectory of forearm while it is moved [3]. Accelerometers come as analog and digital devices. The acceleration of the hand while movements is converted into a continuous voltage signal. In case of digital one, the result is a square wave, the height of whose pulse indicates the acceleration of hand [3]. But accelerometers might not sense some subtlelities, gestures with minute details involved in finger or hand movements. This can be overcome if EMG sensors are used that capture various size and scale gestures [3]. EMG sensors work by producing electrical impulses proportional to the muscular movements of the hand. Apart from these, data gloves have flex sensors embedded into it, whose resistance varies by twisting [11]. The limitation of data glove lies in the fact that they are wired and inconvenient to bear them for the duration of interaction.

- *Computer Vision Approach*

In contrast to former approach, computer vision acquisition of gesture data relies on imaging of gestures and using image analysis techniques to classify them correctly. They utilize mostly video cameras as a separate unit or web cameras mounted onto the computer as tools to acquire images. Computer vision approach can be further divided into two techniques: appearance-based and 3D model-based techniques [12]. In appearance-based technique, the gesture input image is compared with a reference hand gesture image previously stored in the computer for match of features like contours, color distribution for recognition of specific gestures [12]. The images for specific gestures are stored a priori in the database and multiple sample images for each gesture can be taken for empirical analysis. The 3D model based technique focuses on working on the degrees of freedom of

hand and its various joints while a gesture is made [12]. The analysis of image over 2- or 3- dimensions is done to infer the results by comparing with actual values of parameters and those seen in image.

2.2 Gesture Detection, Recognition

This stage involves detection of presence of gesture, not the specific gesture identification. It is done by detecting the electrical signals or image parameters that would track that the signal or image represents some or the other gesture. It is decided by looking at the pattern of magnitude, shape and other characters of signal or of the features found in the input image.

A gesture detection technique may use more of hardware components or software modules or both coupled together. It depends on the application domain, precision of recognition required, budget and system resource availability constraints, and lastly parameters like convenience, robustness etc. Generally a system that is hardware-based is costly compared to purely software like based on image processing, neural network or other concepts used as the foundation. The paper by T. Ulu et. al.[11] uses mostly hardware implementation using data gloves for gesture recognition to achieve HCI. The glove has five flex sensors to detect gesturing, embedded in it: two for index and middle fingers, with one for the thumb. They are sensitive to movement of joints in these fingers and record the bending in terms of its resistance that is proportional to bending in the fingers. In order to send the detected data to computer via microprocessor, a bridge circuit and a power amplifier circuit is used. A fast analog multiplexer to reduce 5 repetitions of the circuits for five sensors is prevented by using multiplexer that connects amplifier & bridge. An artificial neural network(ANN) is used to create a neural circuit representing the network of connections between the various bones and joints of the hand. The hidden layer of ANN has been built using 50 neurons apart from 5 neurons for input and 5 for output. The angular positions of 8 junction points detected by sensor is sent to ANN, which simulates the output hand image with sensed angular positions displayed on output part of ANN. But these circuits are prone to noisy input signals due to addition of sensor noise, impedance noise, interference noise etc. The ANN was found less sensitive to noise errors generated by sensors and tolerated use of different users[11] as experimented by the authors.

Gesture classification is extremely difficult as ambiguity between similar gestures that have to be distinctly recognized is essential. Adding to the complexity is the scientific reason that between two gesture actions, the muscular movements vary instantaneously. The magnitude of muscular motion of hand motion instantly drops to near zero when the arm relaxes after a gesture, then it increases to its actual value for second gesture [3]. Clarity in the magnitudes of hand movements gets seriously reduced. Zhang Xu et. al. [3] have presented a hand gesture recognition and virtual game control system using Electomyogram (EMG) signals and 3D Accelerometer(ACC) sensors as fusion sensor system to widen applications of it. The advantage of the process is it uses multi-channel EMG sensors thus amplifying the signals. EMG sensors are in the form of a band put on forearm of the hand, which have multiple (four) channels for signal passage to amplify the signals and provide clarity. They have used the term active segments to refer to the multi-channel signals of gestures, which represent some semantics with each such gesture. The 3D accelerometer used is a two mutually perpendicular 2D accelerometer combination, placed on the back of forearm near the wrist. Segmentation of EMG signals follows the procedure of dividing active segments into frames and representing each frame as 4n- dimensional feature vector. The average signal of the multiple EMG channel is calculated and thresholding is described for segmentation of ACC signal stream in sync with EMG signal stream. Feature extraction of the 3D ACC stream

in each active segment consists of two steps: scaling and extrapolation. Scaling of amplitude by linear min-max scaling method followed by linear extrapolation to get temporal lengths of all 3D ACC data sequences same is employed. Recognition is using Hidden Markov Model(HMM) [12], a stochastic process that takes time series of observation data as input. The output of the HMM is the probability that the input data is generated by that model. The testing is done on controlling virtual Rubik's cube game. The results of the proposed method in paper [3] for EMG+ACC were recorded the highest accuracy, nearly 100%. Results for EMG-only condition were between 65.9-80.3% and for ACC-only condition between 85.5-90.7%. The large standard deviations of the accuracies for EMG-only and ACC-only indicated that several gestures were unclassifiable. The recognition results achieved by the author's proposed system were considered satisfactory as the overall accuracy was 91.7%.

Computer vision based gesture recognition systems use image processing principles and methods. One important challenge here is the effect of background illumination over correct detection and gesture recognition. It has been addressed by devising a technique by Yoo-Joo Choi et. al. [8], the object(hand) detection phase has been done using image processing of input hand gesture image. The background image region is first separated by extracting the region based on the difference of mean and standard deviation of hue, hue-gradient of background image pixels in input image and background image. A background model of image is built for this purpose. To tightly extract the foreground image, an object bounding box is created using eigen value and vector of initially extracted image. Hand region is obtained by segmenting foreground object region into 16 sub-regions, whose histogram is produced based on number of edges in each sub-region Recognition of gestures is based on support vector machine(SVM) that is trained with sample hand shape features, multiple class for distinct hand shapes are trained followed by testing and classification. 1620 images – 180 images per hand sign for 9 different classes were captured. The mean success rate of recognition on the 9 hand signs was as seen as 92.6%.

Static gestures are focused by many researchers to establish HCI. But a dynamic user interface involves real-time gesture tracking with all the complexities previously cited. S.M. Hassan Ahmed et. al. [6] have discussed real-time, static and dynamic hand gesture recognition for Human-Computer Interaction with the aim of developing a prototype system for controlling Microsoft PowerPoint^{TM} presentations. They have used motion detection by Fast Accelerated Segment Test(FAST) to detect the finger tips. The zest of FAST is to operate on a binary input image, traversing over the region, identifying white pixels(hand region) within a circle of radius r_{finger} drawn using Bresenham's algorithm[15] and noting the maximum number of black pixels on the circle as N_b. N_b should be greater than the maximum threshold N_{min} based on geometry of fingers. If this condition is true, then the outline separating black and white pixels would be a potential fingertip corner. Bresenham's circle algorithm exploits the circle property of being highly symmetrical, and uses it to drawing them on a display screen. It calculates the locations of the pixels in the first 45 degrees[15]. The circle is translated to a location {(x+cx), (y+cy)}, where (x,y) is current position and (cx,cy) its actual center. It then calculates pixels similarly in each of the remaining octants of the circle. Adaptive resonance theory (ART) is used for gesture classification a theory based on aspects of how the brain processes information. The ART model is that object identification and recognition generally occur as a result of the interaction of 'top-down' observer expectations with 'bottom-up' sensory information. The model postulates that 'top-down' expectations take the form of a memory template or prototype that is then compared with the actual features of an object as detected by the senses and are

recognized based on match between template and extracted object of interest. Dynamic gestures were detected using the trajectory formed by the center of the hand over a finite amount of time. Mutiple classes of gestures were given for test of which error occurred when there was little space between the fingers and ambiguity in gesture interpretation. A classification rate of 75 % was achieved for identifying the gestures used as described by the authors [6].

Routaray, S., Agarwal, A. [1] have proposed a dynamic user interface design for HCI using gestures. The image of hand acquired by a web camera is subjected to background elimination by converting each frame into two level gray scale image by removing static backgrounds. Region-based segmentation is used to extract region of interest, with object tracking using average shift calculation. Features like contours of hands are extracted, whose convex hull is obtained and recognition is comparing with predefined classes of gesture images. Recognition extracted features were classified into 11different classes, interpretation was done by mapping to particular class. Generating actions related to gesture as commands was taken as execution of the process. Recognition rate for gestures used in system were 92% for move backward gesture, followed by move forward gesture with 83.3% recognition, zoom out gesture having a rate of less than 66.8% , zoom in, rotate clockwise and rotate anticlockwise were depicted having recognition rates of 73.3%, 70% and 80% respectively on 30 users [1].

Gesture recognition involves basically complex pattern matching in terms of hand posture's static and dynamic features. Pengyu Hong et. al. [4] propose one such method based a finite state machine modelled as a sequence of states in spatial-temporal space. Each state can jump to either itself or its next state. The spatio-temporal information of a state and its neighbour states specifies the motion and the speed of the trajectory within a certain range of variance. Each state *Si* has 5 parameters : a 2D spatial centroid of a state, spatial covariance matrix, spatial threshold and a pair of minimum and maximum temporal units. The system is trained for each possible gesture that it is expected to recognize. Gesture recognition is thought as string matching between a data sequence and the state sequence of an FSM. The Knuth-Morris-Pratt (KMP) algorithm a fast string-matching algorithm, to speed up the recognition procedure. The algorithm uses a prefix function, that encapsulates the information about how a pattern matches against shifts of itself.

The experimental results for hand gestures were recognition rates with 90% or better. For mouse gestures, rates were as low as 70% for complex gestures and 90-100% for simpler gestures [4].

Virtual reality applications are potential applications of HCI through gestures. They involve moving, pointing and manipulating virtual 3D objects by the user giving a real world feel. They are also called augmented reality. Augmented reality (AR) is this technology to create a "next generation, reality-based interface"[9] and is moving from laboratories around the world into various industries and consumer markets. AR supplements the real world with virtual (computer-generated) objects that appear to coexist in the same space as the real world [10].

R.G. O'Hagan et. al [5] have presented a paper on visual gesture interface for virtual environments. The authors use a Barco Baron projection table, that provides a virtual working environment as horizontal table/vertical wall/as an inclined wall. Stereo shutter glasses are the products that are capable of creating a virtual 3D view of objects, is used to display the images of hand on the projector. A twin-camera system, mounted on projection table and right-hand coordinate system with cameras at x-axis, y-axis and the z-axis moving through the scene on display. The images are sent to the computer via the processor for image processing. A model is developed to

create templates for each of the gestures to use it as a reference to narrow down the search during feature extraction. Color-based segmentation is employed. Normalized skin color detection to track the hand is used. The pixels in skin color region are extracted. The largest connected region is detected as hand and small holes are filled. Feature extraction includes moment, high curvature detection, area, principal axes are used to obtain regions of interest like wrist, finger bases, wrist. Template matching with a confidence value indicating the amount of precision of matched feature is used for tracking. Classification of gestures is based on a statistical classifier : a logistic regression which uses a probability-based equation that takes value 0 for image features predictor parameter when it doesn't match for a certain class of gesture; it takes 1 if the image features match the class of gesture for which probability is being calculated. Positional accuracy as found out by experiment by author by measuring the location of 120 points in a plane at 100 mm intervals from 1300 to 1700 mm away from cameras [5].

III CONCLUSION

As the human society is moving towards modernization and computerized, the demand for innovation and ease of use is overly increasing. An intuitive, realistic interface for human-computer interaction is acutely needed. In this regard, the evolution of user interfaces plays an important clue in determining the way progress is seen in user interface design, driven by end user needs. The survey on all the technologies that have come so far will make us to anticipate future desires. Gesture recognition is advancing towards providing the end users a comfortable interface that would free a user from the peculiarities of the machine world. The smart technological developments in this field will bring the coming generations to contribute for making the world much easier, much appealing than ever.

REFERENCES

[1] Rautaray, S.S. and Agrawal, A., "Design of gesture recognition system for dynamic user interface", IEEE International Conference on Technology Enhanced Education, pp. 1-6, 2012.

[2] Rafiqul Z. Khan & Noor A. Ibraheem,"Gesture Algorithms based on Geometric Features Extraction and Recognition"

[3] Zhang Xu, Chen Xiang, Wang Wen-hui, Yang Ji-hai, Vuokko Lantz, Wang Kong-qiao, "Hand Gesture Recgnition and Virtual Game Control Based on 3D Accelerometer and EMG Sensors", Proceedings of the 14th international conference on Intelligent user interfaces, ACM, pp. 401-406, 2009.

[4] Hong, P., Turk, M. and Huang, T. S., " Constructing finite state machines for fast gesture recognition", Proc. ICPR '00. Los Alamitos, CA: IEEE Press, pp. 691—694, 2000.

[5] O´Hagan, R.G., Zelinsky, A. & Rougeaux, S. , " Visual Gesture Interfaces for Virtual Environments", Interacting with Computers, Vol. 14, Nr 3, pp. 231-250, 2002.

[6] S.M. Hassan Ahmed, Todd C. Alexander, and Georgios C. Anagnostopoulos, "Real-time, Static and Dynamic Hand Gesture Recognition for Human-Computer Interaction", Electrical Engineering, University of Miami, Miami, 2009.

[7] Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models", Image and Vision Computing 21, Elsevier, pp. 745–758, 2003.

[8] Choi, Yoo-Joo., Lee, Je-Sung, Cho. We-Duke, "A Robust Hand Recognition In Varying Illumination", Advances in Human Computer Interaction, Shane Pinder (Ed.), 2006.

[9] T. Jebara, C. Eyster, J. Weaver, T. Starner, and A. Pentland. Stochasticks, "Augmenting the billiards experience with probabilistic vision and wearable computers", In ISWC'97: Proc. Int'l Symp. On Wearable Computers, pp. 138–145, IEEE CS Press, pp.13-14, 1997.

[10] D.W.F. van Krevelen and R. Poelman. "A Survey of Augmented Reality Technologies, Applications and Limitations", The International Journal of Virtual Reality, 9(2), pp.1-20, 2010.

[11] T. Ulu, G. Cansever, IB Kucukdemiral, "An ANN Based Electronic Glove for Human-Machine Interaction", Proc. Int. Symp. on Innovations in Intelligent Systems and Applications (INISTA'05), Istanbul, Turkey, pp. 191-194, 2005.

[12] Noor A. Ibraheem & Rafiqul Z. Khan, "Vision Based Gesture Recognition Using Neural Networks Approaches: A Review", International Journal of human Computer Interaction (IJHCI)), Volume (3) : Issue (1) , 2012.

Books :

[13] Karam, Maria and Schraefel, M. C., A Taxonomy of Gestures in Human Computer Interaction, (ACM Transactions on Computer-Human Interactions, 2005).

Website links:

[14] Wikipedia website

[15] Blog link: http://www.asksatyam.com/2011/01/bresenhams-circle-algorithm_22.html