Analysis of Hand Gesture Recognition

Sakshi Gupta
Electronics & Communication Department, Lingaya’s University, Faridabad, India.

Sunil Kumar
Electronics & Communication Department, Krishna Institute of Engineering & Technology, Ghaziabad, India.

Abstract—Human gesture recognition has been an exciting research area. Hand gesture and head gesture recognition can have tremendous applications in Human Computer interface and other robotic machineries. The keyboard and mouse are currently the main interfaces between man and computer. There has been a need of robotic hand that can perform actions alike human hand in real time applications, as it is not possible for human to reach up to every object due to difficult environment. In other areas where 3D information is required, such as computer games, robotics and design, other mechanical devices such as roller-balls, joysticks and data-gloves are used. User would perform gesture according to the action as he wants to be done by robotic hand. The ability to recognize human gestures open up a wide range of possible applications such as automatic recognition of sign language to facilitate communication with the hearing impaired, using gestures as input to explain the emotion of a gesturing person. Various researchers have proposed different approaches for real time gesture recognition. This paper compares the approaches developed as yet for hand gesture recognition.

Keywords—Neural Network, Fuzzy Logic, Gesture Recognition, Harris Corner Detector, Eigen Vector.

I. INTRODUCTION

Gesture recognition in real time from video sequences is one of the most important challenges in computer vision and behaviour understanding since it offers to the machine the ability to identify, recognize and interpret the human gestures in order to control some devices, to interact with some human machines or to monitor some human activities. Generally defined as any meaningful body motion, gestures play a central role in everyday communication and often convey emotional information about the gesticulating person. During the last decades, researchers have been interested to recognize automatically human gestures for several applications: sign language recognition, socially assistive robotics, directional indication through pointing, control through facial gestures, alternative computer interface, immersive game technology, virtual controllers, affective computing and remote control. Recently human gesture recognition catches the peak attention of the research in both software and hardware environment. Gesture recognition is a form of biometric identification that relies on data acquired from the gesture of an individual. This data, which can be either two- dimensional or three-dimensional in nature, is compared against a database of individuals or compared with respective thresholds based on the way of solving the riddle. In recent years, face and gesture recognition has gained popularity among researchers all over the world. With applications ranging from security to entertainment, face recognition is an important subset of biometrics.

Gesture made by human being can be any but few have a special meaning. Human hand can have movement in any direction and can bend to any angle in all available coordinates. Chinese sign language as shown in figure 1, used hand gestures to represents digits as well as alphabets. Many researchers [3][6][15][39][40] have tried with different instruments and equipment to measure hand movements like gloves, sensors or wires, but in these techniques user have to wear the device which doesn’t make sense in practical use. So people thought about a way of contact less gesture recognition that could be considered as a research area in Machine Vision or Computer Vision and which would be as natural as human to human interaction. According to Mitra [5] gesture recognition is a process where user made gesture and receiver recognize it. Using this technique, we can easily interact with machines and can give them particular message according to the environment and application syntax. Even people who can’t communicate orally (sick, old or young child), they would also get benefit from this technology. It is possible to make a gesture recognition system for these people. Mobile companies are trying to make handsets which can recognize gesture and could operate from little distance also [1][47]. Here we are focusing on human to machine interaction (HMI), in which machine would be able to recognize the gesture made by human.

There are approaches of two types.

a) Appearance based approaches where hand image is reconstructed using the image properties and extraction.

b) Model based approaches where different models are used to model image using different models to represent in Computers.

Here we are dividing approaches based on the method used in it not on how it is treating the image. Many approaches have been developed to interact with machines from glove based [3] to neural networks [2]. Users always like the easy and naturalness of technology in HMI and it was more convenient to interpret visual inputs [6]. As Pickering stated [20] that initially touch based gesture interfaces would be popular but, non-contact gesture recognition...
technologies would be more attractive finally. Input to a machine using gesture is simple and convenient, but the communication includes many difficulties. “The human hand is a complex deformable object and gesture itself has many characteristics, such as diversities, ambiguities, temporal and spatial differences and human vision itself is an ill-posed problem” [9]. Pickering [20] described a real time gesture based driving system simulator developed at Carnegie Mellon University with the help of General Motors. Many researchers [a.25][12][19][26][34][44][57][38] have used a color strip or a shirt to detect hand image in the captured image. For a detailed survey of gesture recognition you can see [5][6][13]. Gesture segmentation a part of the gesture recognition process, have been reviewed in [9] and [10] based on color spaces.

Choi [14] brings attention of researchers pointing out an old problem of the incrementing processing time of algorithm’s complexity and say “the most important issue in field of the gesture recognition is the simplification of algorithm and the reduction of processing time”. He used morphological operation to implement his system using the center points extracted from primitive elements by morphological shape decomposition. Lu [18], Gastaldi [26], Ozer [27] used parallel approach in the design and implementation of their system. Different threads are implemented in such way that they can run in parallel and can compute faster. Shin [37] presented a 3D system HGR system with the application in fruit fly chromosomes based on 2D slices of CT scan images. Lee [16] describes his system which he developed for remote control systems which worked for motion recognition also. He uses 3D systems with two or more cameras to detect command issued by hand. Villani [8] has tried to develop a system for teaching mathematics to the deaf with an easy user interface. Morimoto [43] made interesting virtual system, in which he pushed virtual buttons using fingers in the air and recognized it using 3D sensors.

II. HAND DETECTION APPROACHES

There are many techniques to detect hand in the acquired image after pre-processing. As shown above, we divide these approaches into two parts:

Appearance Based Approaches: Many researchers have used fingertip detection for the hand image construction [2][7][11][12][16][19][26][31][34][44][57][38]. As we are also using fingertip detection technique for our research work, this paper devotes great attention to work done by other researchers using this technique. Nolker [2] focuses on large number of 3D hand postures in her system called GREFIT. She used finger tips in hands as natural determinant of hand posture to reconstruct the image. In her system she suggests few approaches to locate fingertip in hand, such as marked fingertips colored and making histogram and using different templates or images of a prototype. It takes 192x144 size gray scale image to process. Verma [7] extract features from image as fingertip, edges and vectors for 2D modelling. He used harris corner detector to extract fingertips corresponding to corners. Nguyen [11] used gray scale morphology and geometric calculations to relocate fingertip locations using learning based model on 640x480 pixel size frame. Here Author use similar approach to hand detector given by shin [12] to detect both hands based on skin color. To recognize hands Nguyen [11] used skin segmentation technique using Gaussian model. Density function of skin color distribution is as defined.

$$P(c_{lsk}) = \sum_{i=1}^{K} \pi_i P_i(c_{lsk})$$

Where $K$ is the number of components and $\pi_i$ are the weight factors of each component. He used CIELUV color space to represent skin. Interestingly he used palm to finger length ratio to construct the hand image. Zhou [19] worked with 320x240 size 24 bit image frames. Zhou used Markov Random Field to remove noise component in processed image. Gastaldi [26] find perimeter using Gaussian filters and freeman’s algorithm [28] to localize fingertips in that image for 3D detection. Kim [34] tried to recognize gesture in a dark room on black projection for his system. Although the system was vision based but he used florescent white paper to mark finger tips in the captured image, which is not practical for generic purpose as user have to wear white florescent strips. Kim used kalman filter for finding fingertips and their

![Fig 1.](image_url)
correct positions in a recursive manner. Stefan [4] implemented a system which can detect motion of fingers in the air visually. Ng [54] developed a system to recognize the 14 predefined gestures in 320x240 pixel sizes in 24 bit color, where hands were moving and the system was able to work with one or both hands. Ng performed a wrist cutting operation on hand images to make both image invariable.

Model Based Approaches: Sawah [31] used histogram for calculating probability for skin color observation. Hu [35] take Gaussian distribution for background pixels marking then he subtracted the pixels from the new image to acquired gesture image. Lee [17] used the same technique to get gesture image

\[ \Delta = | I_n - B | \]

In the modeling of his application of human activity monitoring, Hu [35] applied Genetic Algorithm (GA) to Chromosome pool with Pco and Pmo as crossover and mutation rate respectively which he founded using different statistic attributes. Crossover creates new chromosomes while mutation in this case introduces new genes into chromosome. Lee [a.44] use YCbCr skin color model to detect hand region and then he applied distance transform. Tarrataca [47] used RGB and HSI color space model based algorithm for skin detection. Malassiotis [55] developed a system to recognize real time hand gestures in German sign language in 3D using a sensor enabled camera which can find the pattern based on illumination and computes the 3D coordinates of each point on surface. The details about the pattern finding 3D coordinates are given in his other paper [56]. Lien [38] presented a model base system for HGR where the joints in fingers had one DOF (degree of freedom), effective joins had 2 DOF and spherical joints had 3 DOF. So fingers had 4 DOF while thumb had 5. Then he defined local coordinate systems with the origins on the joints (figure 2).

![Fig 2 (a) Hand model (left) and (b) Local coordinate frames for the joint position for middle finger (right) [38]](image)

III. SOFT COMPUTING APPROACHES

Under the umbrella of soft computing principal constituents are Neural Networks, Fuzzy Systems, Machine Learning, Evolutionary Computation, Probabilistic Reasoning, etc. and their hybrid approaches. Here we are focusing on mainly three components: - Artificial Neural Networks, Fuzzy Logic, Genetic Algorithm

Artificial Neural Network: An Artificial Neural Network (ANN) is made of many highly interconnected processing elements, which are working in together to solve specific problems [32]. ANN can be configured for problems like pattern recognitions or data mining through learning based models. Also ANN has capabilities like adaptive learning, self-organizing and real time operations using special hardware. Nolker [2] used ANN based layer approach to detect fingertips. After obtaining fingertips vectors, it is transformed into finger joint angles to an articulated hand model. For each finger separate network were trained on same feature vectors, having input space 35 dimensional while output dimensional as only 2. Lee [16] used Hidden Markov Model (HMM) for gesture recognition using shape feature. Gesture state is determined after stabilizing the image component as open fingers in consecutive frames. He also used maxima and minima approach like Raheja for construction the hand image and FSM like Verma [7] for gesture finalization. Wang [29] proposed an optical flow based powerful approach for human action recognition using learning models. It labels hidden parts in image also. This mas-margin based algorithm can be applied to gesture recognition. Kim [34] in his system used learning model for dynamic gestures recognition. Ng [52] used HMM and RNNs for gesture classification from the collected vectors of hand pose frames. Outputs of both classifiers were combined to get better result and it was input to the developed GUI. They used Fourier descriptors to represent the boundary of extracted binary hand and trained Radial Basis Function consisted of 56 input nodes, 38 hidden layer and five output nodes. The activation function of the jth hidden node was given by

\[ \Phi_j(x) = \exp \left( \frac{|x-c_j|^2}{2\sigma_j^2} \right) \]
Where \( x \) in input vector, \( c_j \) is the center and \( \sigma_j \) is the spread of \( \phi_j(x) \). Just [36] has presented a comparative study of HMM and IOHMM HGR techniques on the two openly accessible databases, and came with the conclusion that HMM is a better choice for the HGR modeling. Stergiopoulou [59] used unsupervised Self-Growing and Self-Organized Neural Gas network for 31 pre specified gesture. Although he made several assumptions like arm should be vertical, and user is using only his right hand, while this system have a problem when left handed users is showing the gesture. The raised fingers detection in the hand is done by finding the fingertip neuron, which would be followed by the other neurons chain as shown in figure 3. The center of palm could be calculated by gravity method from neuron only in palm area and distance from the fingertips to palm center is calculated, but the main problem in gesture recognition is that only raised fingers would be counted in presented algorithm (figure 3) and gesture would be recognized accordingly. Then he applied a likely hood classification to get the gestures which are predefined based on raised fingers.

Fig. 3 SGONG network working (a) start with two points (b) growing stage with 45 neurons (c) output with 83 neurons (d) hand gesture (e) only raised fingers would be counted [59].

**Fuzzy Logic Based Approaches:** A Professor from UCB USA, Lotfi Zadeh presented fuzzy logic in an innovative way. His view was that for processing precise and accurate information is not necessary, we can perform it with imprecise data also. It is near to natural thinking. As described in [32] “Fuzzy logic is a multivalued logic that allows intermediate values to be defined between conventional evaluations”. Verma [7] used c-mean fuzzy clustering based finite state machines (FSM) to recognize hand gestures. Formula for centroid calculation of fuzzy c-means clusters is that centroid would be mean of all points weighted by their degree of belongings to the cluster center. For each point \( x \), a coefficient giving the degree in the \( k^{th} \) cluster \( U_k(x) \) [21]. Here \( x_k = k^{th} \) trajectory point, so

\[
\text{Center}_k = \frac{\sum U_k(x)^m x}{\sum U_k(x)^m}
\]

In second phase these cluster maps onto FSM states and final state show gesture recognition, although Verma [7] didn’t implement it. Schlomer [45] used k-mean algorithm on clusters, then he applied HMM and Bayes-classifier on vector data. Trivino [33] tried to make a more descriptive system which can convert human gesture positions into a linguistic description using fuzzy logic. He related it to Natural Language Processing (NLP). He used sensors and took only few positions in sitting and standing, into consideration.

**Genetic algorithm based approaches:** Genetic Algorithm comes from biology but it is very influential on computational sciences in optimization. This method is very effective to get optimal or sub optimal solutions of problems as it have only few constraints [32]. It uses generate and test mechanism over a set of probable solutions (called as population in GA) and bring optimal acceptable solution. It executes its three basic operations (Reproduction, Crossover and Mutation) iteratively on population. Sawah [31] has focused on a very generic scenario where he used generic non-restricted environment, generic not-specific application for gesture recognition using genetic programming. He used crossover for noise removal in gesture recognition, while Dynamic Bayesian Network (DBN) for gesture segmentation and gesture recognition with the fuzzification. Hu [35] applied Genetic Algorithm on his system which make 2D parametric model with human silhouette in his application of Human Activity Monitoring. The best point about GA is that it work parallel on different points for faster computation.
IV. OTHER APPROACHES

Raheja proposes a new methodology for real-time robot control using Principal Component Analysis (PCA) for gesture extraction and pattern recognition with saved images in database in 60x80 image pixels formats. He used syntax of few gestures and decides corresponding actions of robot. He claims that PCA method is very faster than neural network-based methods which require training database and more computation power. Huang [52] used PCA for dimensionality reduction and SVM for gesture classification in using skin color model switching for varying illumination environment. In Huang approach (Figure 4) image sequences were sent for skin-color detection, hand pose angle estimation and hand region segmentation. Then it divides resultant image into 40, 20x20 pixels size and run Gabor filter. Morimoto [43] also used PCA and maxima methods. Gastaldi [26] used PCA to compress five image sequences into one and get eigen vectors and eigen values for each gesture. He used statistical HMM model for gesture recognition. Zaki [58] used PCA where hand representation is transformed from the image coordinates to eigen vector space. After vector rotation the largest eigen vector was aligned with the mid of data as shown in figure 5. He used three HMMs for every sign, one for each feature PCA sequence, Kurt Pos sequence and MCC sequence found.

Shin [30] shows gesture extraction and recognition using entropy analysis and low-level image processing functions. Lee [17] also used entropy to get color information. He used PIM to quantify the entropy of image using the following equation.

\[ PIM = \sum_{i=1}^{l-1} h(i) - \max_j h(j) \]
Where $h(i)$ is the $i^{th}$ histogram value of each image or block. To acquire PIM value, subtracting all pixels in each block from maximum frequency in histogram model. Lu [18] implemented system for 3D gesture recognition where he fused different positions of gesture using coordinate transformations and then use stored presupervised gestures for gesture recognition. Stefan [4] has used Dynamic Space-Time Warping (DSTW) [42] to recognize a set of gestures. This technique doesn’t require hands to be correctly identified at each frame. Zou [46] used Deterministic Finite State Machine (DFSM) to detect hand motion and then apply rule based techniques for gesture recognition. He defines gesture into two category based on motion linear and arc shaped gestures. Tarrataca [47] used convex hull method based clustering algorithm Graham’s Scan [48] for posture recognition. Chang [49] used a feature alignment approach based on curvature scale space to recognize hand posture.

V. Conclusions

GREFIT [2] system was able to detect finger tips even when it was in front of palm, it reconstruct the 3D image of hand that was visually comparable. Nguyen [11] claimed results 90-95% accurate for open fingers that is quite acceptable while for closed finger it was 10-20% only. As shown in figure 6 closed or bended finger are coming in front of palm, so skin color detection would not make any difference in palm or finger. According to him image quality and morphology operator was the main reason for low detection. Raheja claims about 90% accuracy in the result, if the lighting conditions are good. Hu [35] used six different parameters to control the performance of system, if he found much noise there, he could control it using two parameters called as $c$ and $\beta$ respectively. Stergiopoulos [59] claims about 90.45% accuracy, through hidden finger was not detected in his approach. Morimoto [43] claimed for his system near 91% accuracy after he applied normalization. Ng and Ranganath [54] showed 91.9% correct results with the combination of HMM and RNNs in their approach. Huang [52] claimed 93.7% result using Gabor filters. Lee [17] showed results for six kinds of gesture with recognition rate of more than 95% but it recognized bended finger as bended, no matter the morphology recognize a set of gestures. This is quite acceptable in general environments, Greece, Italy.

The running gestures like Bye and showing a direction, Suk [53] claims 90.77% results. Where $h(i)$ is the $i^{th}$ histogram value of each image or block. To acquire PIM value, subtracting all pixels in each block from maximum frequency in histogram model. Lu [18] implemented system for 3D gesture recognition where he fused different positions of gesture using coordinate transformations and then use stored presupervised gestures for gesture recognition. Stefan [4] has used Dynamic Space-Time Warping (DSTW) [42] to recognize a set of gestures. This technique doesn’t require hands to be correctly identified at each frame. Zou [46] used Deterministic Finite State Machine (DFSM) to detect hand motion and then apply rule based techniques for gesture recognition. He defines gesture into two category based on motion linear and arc shaped gestures. Tarrataca [47] used convex hull method based clustering algorithm Graham’s Scan [48] for posture recognition. Chang [49] used a feature alignment approach based on curvature scale space to recognize hand posture.

REferences


[26.] Gastaldi G. and et al., A man-machine communication system based on the visual analysis of dynamic gestures, International conference on image processing, Genoa, Italy, 11-14 Sep, 2005, pp. 397-400.


Gupta et al., International Journal of Advanced Research in Computer Science and Software Engineering 3(5), May - 2013, pp. 1332-1339


