

International Journal of Advance Research in Computer Science and Management Studies

Research Paper

Available online at: www.ijarcsms.com

A Review of Literature on Hand Gesture Recognition for Indian Sign Language

Neha V. Tavari¹

M.Tech Student

Department of Computer Science and Engineering
Government College of Engineering
Amravati, Maharashtra – India

A. V. Deorankar²

Associate Professor

Department of Computer Science and Engineering
Government College of Engineering
Amravati, Maharashtra – India

Dr. P. N. Chatur³

Head of Department

Department of Computer Science and Engineering
Government College of Engineering
Amravati, Maharashtra – India

Abstract: Sign language is the language of communication for deaf and dumb people. Most of these physically impaired communities are dependent on sign language translators to express their thoughts to rest of the world. This causes isolation of these people in society. Hence, Sign Language Recognition is one of the most growing fields of research today, which in fact is composed of various gestures formed by physical movement of body parts i.e. hand, arms or facial expressions. Gestures are considered as the most natural expressive way for communications between human and computers in virtual system. Hand gesture is a method of non-verbal communication for human beings for its freer expressions much more other than body parts. Hand gesture recognition has greater importance in designing an efficient human computer interaction system. In this paper a survey on various hand gesture recognition approaches is provided.

Keywords: Sign language; Hand gesture; Computer vision; Human computer interaction; deaf and dumb.

I. INTRODUCTION

Sign language is widely used by people who cannot speak and hear or people who can hear but cannot speak and is the only communication mean for those deaf-dumb people. It is composed of various gestures formed by different hand shapes, movements and orientations of hands or body, or facial expressions. These gestures are used by the deaf people to express their thoughts [1]. As sign language is well structured code gesture, each gesture has a meaning assigned to it [2]. There are different sign languages across the world. The sign language used at a particular place depends on the culture and spoken language at that place[1] such as American Sign Language (ASL), British Sign Language (BSL), Japanese Sign Language family (Japanese, Taiwanese and Korean Sign Languages), French Sign Language family (French, Italian, Irish, Russian and Dutch Sign Languages), Australian Sign Language, etc.. Indian sign language (ISL) is used by the deaf and dumb community in India. As ISL got standardized only recently, a very little research work has happened in ISL recognition [3]. In fact, researches on other sign language such as American Sign Language, Greek Sign Language etc. are reviewed in this paper.

The purpose of this paper is present a review of Hand Gesture Recognition techniques for human computer interaction, consolidating the various available approaches. Although recent reviews in computer vision based have explained the importance of gesture recognition system for human computer interaction (HCI), this work concentrates on vision based techniques method [4].

II. HAND GESTURE TECHNOLOGY

Gestures are considered as the most natural expressive way for communications between human and computers in virtual system [4] and a powerful means of communication among Humans. The key problem in gesture interaction is how to make hand gestures understood by computers [5]. For any system the first step is to collect the data necessary to accomplish a specific task. For hand posture and gesture recognition system different technologies are used for acquiring input data. Present technologies for recognizing gestures can be divided into vision based, instrumented (data) glove, and colored marker approaches. Fig 1 shows an example of these technologies.



Figure 1. (a) Data-Glove based (b) Vision based (c) Colored marker

A. Vision based approaches

In vision based methods the system requires only camera(s) to capture the image required for the natural interaction between human and computers and no extra devices are needed [4]. It is more natural and useful for real time applications[3]. Although these approaches are simple but a lot of gesture challenges are raised such as the complex background, lighting variation, and other skin color objects with the hand object, besides system requirements such as velocity, recognition time, robustness, and computational efficiency. Figure 2 shows block diagram for vision based hand gesture recognition system.

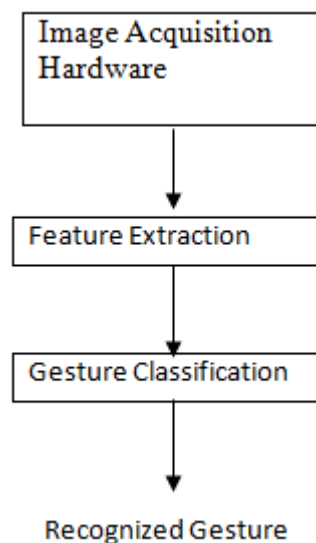


Figure 2. Vision based hand gesture recognition system

Vision based technologies use a bare hand to extract data needed for recognition, these methods are natural, easy, and the user directly interact with the system. Vision based technology deals with some image characteristics such as texture and color for acquiring data needed for gesture analyze. There are many techniques applied for detecting hand object after some image preprocessing operations, these methods can be divided into two parts[4].

- i. **Appearance based approaches:** In appearance based approaches, the visual appearance of the input hand image is modeled using the feature extracted from the image, which will be compared with the features extracted from stored image. Appearance based approaches have the advantage of real time performance [5] and easier than 3D model based approaches

due to the easier extraction of features in 2D image. The common method used this approach is to detect skin colored regions in the image; however this method is affected with changing illumination conditions and other background objects with skin like color. At present a lot of study efforts have been grown on approaches that apply invariant features, such as AdaBoost learning algorithm. The use of invariant features enables the identification of regions or points on a particular object, rather than modeling the entire objects. With this method the problem of partial occlusion has been overcome.

- ii. **3D model based approaches:** Model based approaches used 3D model description for modeling and analysis the hand shape. In these approaches search for the kinematic parameters are required by making 2D projection from 3D model of the hand to correspond edges images of the hand, but a lot of hand features might be lost in 2D projection. 3D Model can be classified into volumetric and skeletal models. Volumetric models deal with 3D visual appearance of human hand and usually used in real time applications. The main problem with this modeling technique is that it deals with all the parameters of the hand which are huge dimensionality. Skeletal models overcome volumetric hand parameters problem by limiting the set of parameters to model the hand shape from 3D structure. Fig 3 shows 3D model approaches.

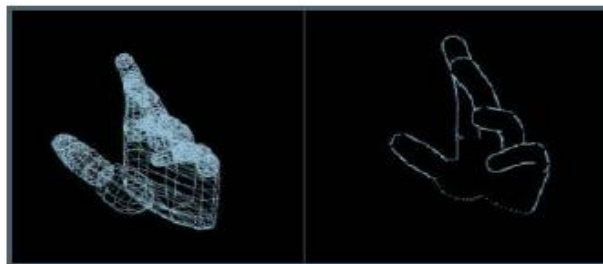


Figure 3. 3D model (left) and the generated contour (right)

In vision based hand gesture recognition system, a video camera used to record hand movements, and the input video is partitioned into frames, for each frame, a set of features are extracted. After some preprocessing operations, the hand object is localized and segmented and the necessary features are extracted and stored in the computer as a trained set. Then each input image pass through the previous steps to extract its features, and classification algorithms are applied by comparing the extracted features from input image with the training set, to interpret the gesture meaning according to a specific application.

B. Instrumented glove approaches

Data glove approaches use sensor devices for capturing hand position, and motion[4]. In this approach, detection of hand is eliminated by the sensors on the hand[3] and it can easily provide exact coordinates of palm and finger's location and orientation, and hand configurations, however these approaches require the user to be connected with the computer physically which obstacle the ease of interaction between users and computers, besides the price of these devices are quite expensive, it is inefficient for working in virtual reality.

C. Colored markers approaches

Marked gloves or colored markers are gloves that worn by the human hand with some colors to direct the process of tracking the hand and locating the palm and fingers, which provide the ability to extract geometric features necessary to form hand shape. The color glove shape might consist of small regions with different colors or as applied in where three different colors are used to represent the fingers and palms, where a wool glove was used. The amenity of this technology is its simplicity in use, and cost low price comparing with instrumented data glove. However this technology still limits the naturalness level for human computer interaction to interact with the computer.

III. RELATED WORK

A. Recognition based on 'UP' and 'DOWN' positions of Fingers

In the proposed method, a set of 32 signs, each representing the binary 'UP' & 'DOWN' positions of the five fingers is defined. In the proposed method, 32 combinations of binary number sign are developed by using right hand palm image, which

are loaded at runtime. An image captured at run time is scanned to identify finger tip positions of the five fingers. Measuring the heights according to a reference point at the bottom of the palm close to the wrist, the tip of fingers is identified. The heights are determined by Euclidean distance measurements.

The edge images are scanned to locate the tip positions of the 'UP' finger and to locate the reference point (x_0, y_0) at the bottom of the palm. The edge images are further taken through scan process and detection phase. The procedure of scan process includes i) marking of feature point ii) determination of heights of fingers in 'UP' position iii) determination of angle between the line joining the feature point of 'UP' fingers with the reference point and the horizontal line passing through the reference point and iv) determination number of instances 'd' which is used to limit number of searches among the 32 signs is considered. The scanning procedure requires either one of the two or both the scan modes namely the Left – Right and Right – Left scan to determine the feature points as required by the sign under process out of the 32 signs defined. The feature points located by Left-Right scan are marked 'GREEN' color, those located by Right-Left scan are 'BLUE' color and the reference point is marked 'RED' color[2].

B. Kohonen self-organizing map algorithm

In this hand gesture recognition system, we have used an intensity (grayscale) representation of the segmented image for further processing. This grayscale version, also called a "skin map," contains intensity values for skin pixels and the background is represented as black. Then, the Two-Dimensional Discrete Cosine Transform (2D-DCT) for each region is computed, and feature vectors are formed from the DCT coefficients. The DCT can be extended to the transformation of 2D signals or images. This can be achieved in two steps:

- i) By computing the 1D-DCT of each of the individual rows of the two dimensional image,
- ii) After the above step, by computing the 1D-DCT of each column of the image.

Self-organizing map (SOM)[3] is used and an unsupervised learning technique in Artificial Neural Network (ANN) is used to classify DCT-based feature vectors into groups to classify whether the sign mentioned in the input image is "present" or "not present" in the ISL database. As SOM is based on unsupervised learning, no manual intervention is needed during the learning and little need to be known about the characteristics of the input data. Hence SOM is used for clustering data without knowing the class memberships of the input data. The SOM is also known as SOFM, the Self-Organizing Feature Map (SOFM) as it can be used to detect features belonging to the problem. The particular kind of SOM known as a Kohonen Network is used here having feed-forward structure with a single computational layer arranged in rows and columns.

C. Novel segmentation algorithm

In order to simplify the process of gesture identification, special modified white color woolen hand gloves [6] were used. One of the most important features of sign language is that each finger in a gesture conveys a particular message and hence each and every finger has to be individually identified as well. In order to assist this requirement, the woolen hand gloves were modified. This was done by replacing and sewing each finger of the glove with a colored cloth for each digit of the hand. Here, we have utilized a unique color coding for each finger of our hand in order to assist in identifying the fingers. Therefore, segmentation based on various color spaces would be a viable option. In proposed work, two of the most popular color spaces used other than RGB and their conversion from the RGB color space.

i) YCbCr Color Space

A prominent color space used in video and digital photography is the YCbCr color space which consists of Y, Cb and Cr components. 'Y' or luma component is described as the representation of brightness of an image and is different from Y or luminance which is described as the light intensity encoded nonlinearly using gamma correction, Cb is described as blue difference chroma component and Cr is described as red difference chroma component.

ii) Hue, Saturation and Intensity (HSI) Color Space

HSI color space is a color space commonly used in computer graphics applications which represents the points of RGB color space in cylindrical co-ordinates. The components of this color space consists of H or Hue which is described as an optical perception wherein a region seems to be comparable to that of one of the primary colors red, green and blue, S or Saturation described as total amount of purest color distributed throughout the image and I or Intensity is described as the total amount of light leaving through the specific region. With the conversion to any of these above mentioned color spaces it is possible to obtain additional valuable information in order to perform perfect segmentation.

D. Transition movement model

Transition movement models (TMMs)[7] are proposed to handle transition parts between two adjacent signs in large-vocabulary continuous SLR. For large-vocabulary continuous SLR, TMMs were proposed for continuous Chinese SLR. Sign samples taken from input devices are fed into the feature extraction unit and then input into two related parts i.e.TMM training and recognition based on TMMs. In the TMM training part, sign/sentence samples are trained into sign models and TMMs by the model training module (no TMMs in the first run). Then, these models are used to segment continuous sentence samples into sign parts and transition parts. Transition parts are clustered using the temporal clustering algorithm. We iterate this process until the convergence criterion is met. In the recognition part based on TMMs, the estimated TMMs and sign models obtained from the training part are viewed as candidates of the Viterbi search algorithm, together with language models (Bigram) for recognizing large-vocabulary continuous sign language.

E. CAMSHIFT algorithm

In the proposed system, an attempt is made to recognize alphabet characters (A-Z) in real time from color image sequences using “Continuous Adaptive Mean Shift Algorithm (CAMSHIFT)” tracking algorithm [8]. The mean shift algorithm operates on probability distribution. To track colored objects in video frame sequences, the color image data has to be represented as probability distribution by using color histogram. Color distributions acquired from video image sequence change over time, so the mean shift algorithm has to be adapted dynamically to the probability distribution which it is tracking. The new algorithm that meets all these requirements is called CAMSHIFT. Here the input video is converted into 50 frames and we apply CAMSHIFT tracking algorithm to all frames and find out one frame that covers the maximum hand area under search window. Tracing and extracting hand from an image sequence achieved by implementation of CAMSHIFT algorithm for hand tracking which works as follows: i) choose the initial location of search window. ii) Perform mean shift and store zeroth moment. iii) Set search window size equal to a function of the zeroth moment found in step 2. iv) Repeat step 2 and step 3 until convergence.

F. Intrinsic-Mode sample entropy on sEMG and accelerometer data

In this work, data from five-channel surface electromyogram and 3-D accelerometer [9] from the signer’s dominant hand were analyzed using intrinsic mode entropy (IMEn) for the automated recognition of Greek sign language (GSL) isolated signs. As the gesture is directly connected to hand movement, measurement of the latter could contribute to the gesture representation in the recognition space. This was the motivation to explore the capabilities of sEMG and 3-D-Acc data in SL recognition. To this end, the movements of the arm can be captured with the help of a 3-D Accelerometer, whereas the motions of the wrist and the fingers can be obtained by the corresponding muscles on the arm. After experimentation on the exact placement and type of the sEMG electrodes, a position that provides with high signal quality and discrimination among the performed motions per channel was identified.

G. Eigen value weighted euclidean distance based classification technique

Eigen values and Eigen vectors are a part of linear transformations. Eigen vectors are the directions along which the linear transformation acts by stretching, compressing or flipping and Eigen values gives the factor by which the compression or

stretching occurs. For recognition of hand gestures, only hand portion till wrist is required, thus the unnecessary part is clipped off using this hand cropping technique. After the desired portion of the image is being cropped, feature extraction phase is carried out. Here, Eigen values and Eigen vectors are found out from the cropped image. Here is designed a new classification technique[10] that is Eigen value weighted Euclidean distance between Eigen vectors which involved two levels of classification.

i) Classification based on Euclidean Distance:

Euclidean distance was found out between the Eigen vectors of the test image and the corresponding Eigen vectors of the database image. As five Eigen vectors were considered, we get five Euclidean distances for each database image and then the minimum of each was found out. Mathematically,

$$E.D. = \sqrt{\sum_{n=1}^m (EV1(n) - EV2(n))^2} \quad (1)$$

where EV1 represents the Eigen vectors of the test image and EV2 represents the Eigen vectors of the database image in Eq. (1).

ii) Classification based on Eigen value weighted Euclidean distance:

The difference of Eigen values of the test image and the Eigen values of the database image was found out. Then, it was multiplied with the Euclidean Distance obtained in the first level of classification given as C2 in equation below. Then sum of results obtained for each image were added and minimum of them was considered to be the recognized symbol. Mathematically,

$$C2 = (E.D.) * |E1 - E2| \quad (2)$$

where E1 and E2 in Eq. (2) are the Eigen values of the test images and database images respectively.

H. K-Mean based radial basis function neural network

To obtain a rotation invariant gesture image, a novel technique is proposed in this paper by coinciding the 1st principal component of the segmented hand gestures with vertical axes. A localized contour sequence (LCS) based feature is used here to classify the hand gestures. A k-mean based radial basis function neural network (RBFNN)[11] is also proposed here for classification of hand gestures from LCS based feature set.

The shape of the contour is an important property that can be used to distinguish of the static hand gestures from one class to another. The localized contour sequence (LCS), which has been confirmed to be a very efficient representation of contours, is selected as a feature set of the hand gesture. A well established canny edge detector used to detect the edge of preprocessed hand gesture. A contour tracking algorithm is proposed to track the contour of a gesture in the clockwise direction and the contour pixels are numbered sequentially starting from the topmost left contour pixel. After successfully extracting a normalized LCS feature vector of the static hand gesture, The classification job is done via k-mean based radial basis function neural network (RBFNN). Radial basis function neural network (RBFNN) is widely used in pattern recognition tasks for its fast learning algorithms.

I. Finger detection

It states an efficient algorithm for identification of the number of fingers opened in a gesture representing an alphabet of the American Sign Language (ASL) and introduces a robust and efficient technique for finger detection [12]. The method has three main phases of processing viz., Edge Detection, Clipping and Boundary Tracing. The first phase having canny edge operator and produces an edge detected image which reduces the number of pixels to be processed at runtime. The second phase clips the

undesirable portion of the edge detected image for further processing and the final phase traces the boundary of the image and detects finger tips which aid in finger detection.

IV. CONCLUSION

In this paper, we give brief summary about various methods and technique which are provided by various authors for recognition of hand gesture. The ultimate goal of hand gesture recognition system is to identify language of physically impaired people as well as to build an efficient human computer interaction system. Computer Vision methods for hand gesture interfaces must surpass current performance in terms of robustness and speed to achieve interactivity and usability. Taking into consideration the research related to vision based hand gesture recognition an observable progress has been made and it can be implemented as a real time application. Hand gesture is an active area of research in the computer vision. To continue with the efficiency further research in the areas of feature extraction, classification methods and gesture representation are required to realize the ultimate goal of humans computer interface in the field of sign language recognition for physically impaired peoples.

References

1. G Adithya V., Vinod P. R., Usha Gopalakrishnan , “Artificial Neural Network Based Method for Indian Sign Language Recognition” , IEEE Conference on Information and Communication Technologies (ICT) ,2013, pp. 1080-1085.
2. Rajam, P. Subha and Dr G Balakrishnan , "Real Time Indian Sign Language Recognition System to aid Deaf and Dumb people", 13th International Conference on Communication Technology(ICCT) ,2011, pp 737-742.
3. Deepika Tewari, Sanjay Kumar Srivastava , “A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self- Organizing Map Algorithm”, International Journal of Engineering and Advanced Technology (IJEAT), Vol.2, Dec 2012, pp. 165-170.
4. Noor Adnan Ibraheem, RafiqulZaman Khan, “Survey on Various Gesture Recognition Technologies and Techniques”,International Journal of Computer Applications,Vol. 50, No.7, July 2012, pp. 38-44.
5. Pragati Garg, Naveen Aggarwal and Sanjeev Sofat, “Vision Based Hand Gesture Recognition”,World Academy of Science, Engineering and Technology 49, 2009, pp.972-977.
6. Dhruva N. and Sudhir Rao Rupanagudi, Sachin S.K., Sthuthi B., Pavithra R. and Raghavendra , “Novel Segmentation Algorithm for Hand Gesture Recognition”, IEEE,,2013, pp. 383-388.
7. Gaolin Fang, Wen Gao, and Debin Zhao, “Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models”, IEEE transactions on systems, man, and cybernetics—part a: systems and humans, Vol. 37, No. 1, Jan 2007.
8. Sulochana M. Nadgeri,Dr. S.D.Sawarkar,Mr.A.D.Gawande, “Hand gesture recognition using CAMSHIFT Algorithm”,Third International Conference on Emerging Trends in Engineering and Technology,IEEE ,2010, pp.37-41.
9. Vasiliki E. Kosmidou, and Leontios J. Hadjileontiadis, “Sign Language Recognition Using Intrinsic-Mode Sample Entropy on sEMG and Accelerometer Data”, IEEE transactions on biomedical engineering, Vol. 56, No. 12, Dec 2009.
10. Joyeeta Singha, Karen Das, “Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique”,(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 2 , 2013, pp. 188-195.
11. Dipak Kumar Ghosh , Samit Ari , “A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network”, IEEE,2011.
12. Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R, Sudheender S, Nitin V. Pujari, “Finger Detection for Sign Language Recognition”,Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1, March 2009.

AUTHOR(S) PROFILE



Neha V. Tavari, has received her B.E. degree in Computer Engineering from B.D. college of Engineering ,Sewagram, India in 2011 and now she is pursuing M.Tech in CSE branch from Government college of Engineering, Amravati . Her area of research includes image processing, pattern recognition, and neural networks.



Prof. A. V. Deorankar, has received his M.E. degree in Electronics Engineering from Govt. College of Engineering, Amravati, India. He has published nine national level papers and seven international papers. He is also patent of one paper. His area of research includes Computer Network, Web Mining. Currently he is working as an Associate Professor at Govt. college of Engineering, Amravati, India.



Dr. P. N. Chatur, has received his M.E. degree in Electronics Engineering from Govt. College of Engineering, Amravati, India and PhD degree from Amravati University. He has published twenty national level papers and fifteen international papers. His area of research includes Neural Network, data mining. Currently he is head of Computer Science and Engineering department at Govt. College of Engineering, Amravati, India.