Hand Gesture Recognition Systems: A Survey

Surve Pranjali, Ubale V.S

Abstract:*The first mode of communication in the early ages was gesture. Later on the verbal communication was developed very well. But still non-verbal communication has its importance. Non – verbal communication is used for the physically challenged people, for different applications. For e.g. aviation, surveying, music direction etc. It is the method to interact with other peripheral devices, such as keyboard, mouse. The important steps involved in HGR systems are data acquisition, gesture modelling, feature extraction and hand gesture recognition. Different algorithm are followed by different researchers or sometimes used their own algorithm. This paper studies different methods of HGR and finally the desired characteristics of a robust and efficient hand gesture recognition system have been described.*

Index terms: *Hand gesture recognition, Principal component analysis (PCA), Rotation invariant*

I. INTRODUCTION

Non-verbal communication includes various gestures of body parts, mostly face and hand. In human being the oldest method of communication is gesture. In early ages the way of communication for men to share the information of food, source of water, information about their enemy, request for help etc. Though gestures are used widely for different applications. This includes human-robot interaction, sign language recognition, interactive games, vision-based augmented reality etc. Another major application of gestures is found for making the passengers aware about the safety features by the airhostess. For communication by the people at a visible, but not audible distance (surveyors) and for the physically challenged people (mainly the deaf and dumb) gesture is the only method. Another area where hand gesture recognition has been successfully used is human robot interaction. Keyboard and mouse can be used in 2D world, but the control of a robot needs a 3D space. Hand gesture is most suitable for such purposes. Different algorithms and features are used by different researchers for recognition purpose. The present work reviews a numbers of researches on hand gesture recognition systems along with the different steps of the recognition systems. We will study different methods that will give best match results of recognition. In this for the recognition purpose real time camera is used, so the process becomes fast. This system can be used anywhere such as home, office because in this USB camera is used. MATLAB 7.1 software is used in this system.

Manuscript Received on February 2015.

Surve Pranjali, B.E. (Electronics), Department of Electronics (Digital Systems), A V C O E, Sangamner, Affiliated to University of Pune, India.

Ubale V.S, Asst. Prof., Department of Electronics (Digital Systems), A V C O E, Sangamner, Affiliated to University of Pune, India.

Database is created for all the four hand gestures using ten different images. The images are labelled using integer numbers starting from 1. Main aim of hand gesture recognition system is to identify hand gestures and to classify them accurately. Performance of most algorithms is reported on identification task since characterization of task holds on identification. To successfully meet verification, for hand gesture recognition systems, it is needed to create testing and training databases that specifically address these applications. The process will start with image processing techniques such as, segmentation, noise removal (low-level) feature extraction.

II. SYSTEM OVERVIEW

Hand gesture recognition system has four different phases to find out the gesture. They are data acquisition, hand segmentation and pre-processing, feature extraction and finally the recognition. The hand image is captured by suitable input device. The image is segmented to locate the hand from the (cluttered) background and other parts of the body and thereafter the image is processed to remove noises, to detect edges/ contours, to normalize for generating the simplest and desired model. The features are extracted from the segmented and pre-processed image for recognition. Any gesture recognition system works according to following steps first acquiring the input image from camera, filtering, segmentation, feature extraction and classification. To control robot or to convey meaningful information we have to create human computer interaction to recognize gesture for physically disable people. Our aim is to make computer understand any human gesture to control various applications

A) Extraction method and image preprocessing(gesture modelling)

It is the next step after data acquisition and the success of the gesture recognition mostly depends on this stage. Different data received through the input devices are to be modelled properly depending up on the type of applications. Gesture modelling has four different steps, viz. hand segmentation, filter/ noise removal, edge/ contour detection and lastly normalization.

B) Hand Segmentation

Hand segmentation (or hand localization) refers to locate the hand or hand sequences in the images. There are different methods for hand segmentation and all of them classify the image in two homogeneous parts, foreground containing the hand and background containing the rest. Backgrounds are of two types, uniform and cluttered. A very good segmentation is required to choose threshold of gray level to extract hand from background. i.e. no part of object should be present in the background and vice a versa. With help of segmentation we can separate regions from boundaries. Various segmentation algorithms are there for various applications. Depending upon type of hand gesture, for

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

dynamic hand gesture, we have to locate and track hand gesture. For static gesture (posture) first input image is segmented. A bounding box is created to locate a hand firstly depending on the skin colour and secondly, we need to track the hand. If we want to track hand from a video, the video is divided into frames and all the frames are processed individually considering it as a static posture.

C) Thresholding

In this method the image is divided in two regions of interest, background and foreground, based up on a particular value (termed as thresholding value). Thresholding is based on different real-time parameters, some of them use range/ depth thresholding, some of them use colour(RGB or HSV) thresholding,some use speed thresholding.

D) Skin-based

The human skin colour can be used to separate the hand, head or body from the back ground using RGB or Grey or HSV (hue, saturation and value) (also known as HSB i.e. hue, saturation and brightness) colour space representation. It can be used to separate the foreground from background with reference to predefined range of the colour.

E) Colour normalization

The illumination is a factor that affects the distribution of colour values in an image. These values change depending on different lighting conditions or different cameras. Colour normalization allows for object recognition techniques based on RGB colour, to compensate for these variations.

F) Noise removal

Noise is an important factor which affects the quality of an image. Noise is introduced either during the capturing or processing or transmission of the image. Noise removal or reduction is necessary for successful hand gesture recognition. Judicial selection of noise filter will lead to effective noise removal. There are many methods for noise removal.

G) Salt and pepper

It is the most common noise often found in an image. This appears as randomly occurring white and black pixels. This can be reduced by median filter, morphological filter, and contra harmonic mean filter.

H) Morphology erosion

Morphological operations are mathematical operations designed in the context of set theory. Presently, this is used for noise reduction in image processing through two basic morphological operators, erosion and dilation. An erosion filter tends to reduce the sizes of bright image features by correlation with adjacent dark areas, where as the dilation filter does the opposite that means it constrains dark features with pixels from surrounding brighter areas. The opening of an image is defined as the erosion of the image followed by subsequent dilation using the same structural element.

III. FEATURE EXTRACTION

Features are the crucial elements for hand gesture recognition. Large number of features, such as, shape, orientation, textures, contour, motion, distance, centre of gravity etc. can be used for hand gesture recognition. Hand gesture can be recognized using geometric features, like, hand contour, fingertips, finger detections. But these features may neither be always available nor reliable due to occlusions and illuminations. Some non-geometric features (such as colour, silhouette, and texture) are also available for recognition. But they are inadequate for the purpose. Therefore, the image or the processed image can be fed to the recognizer to select the features automatically and implicitly, rather than using single type of feature alone. Following three approaches are useful for extraction of features. A very important step in gesture recognition is feature extraction. After extraction of features it is given as an input to the classifier. If the segmentation is done perfectly it produces features that play important role in recognition process. In feature extraction first we have to find edge of the segmented and filtered image. We come to know about boundaries of different objects due to edge. Edge can be said as sudden change in the intensity from one pixel to other pixel. Edge detection leads to reduction in some amount of data but same shape is maintained. Depending upon the application there are many different ways of feature extraction of the segmented image.

IV. GESTURE CLASSIFICATION

Once the appropriate features have been extracted from the images and a suitable data set have been selected, the gestures can be recognized using standard machine learning techniques or a special-purpose classifiers. Several methods have been used for gesture recognition: template matching, dictionary look-up, statistical matching, linguistic matching, neural network and ad hoc method. The task of assigning a feature vector or a set features to predefined classes in order to recognize any gesture is called as classification. After analysis and modelling of the input hand image is done, to recognize the gesture, a gesture classification method. Many classification methods have been proposed and tested successfully in various recognition systems. Process of recognition is affected with the proper selection of suitable classification algorithm and features parameters. Gestures can be meaningful and expressive body motions involving physical movements of the body, face, arms, head, fingers, hands with the intent of:

Interacting with the environment.

1) Interaction with physically disable people.
2) Conveying meaningful information

A class is a set of reference features that were obtained during the training phase using a set of training images. For the features extracted in the previous phase finding the best matching reference features is classification. For the information to subsequently reconstructed nd to be transmitted elsewhere by the receiver a gesture can be perceived by the environment as a technique of compression. Gesture recognition has wide-ranging applications as following:

1) Visualization
2) Computer games
3) Man-machine interface
4) 3D animation
5) Control of mechanical systems

Gestures can be static or dynamic, the different types of gestures are:

A. Head and face gestures

1) Nodding or shaking the head

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

2) Raising the eyebrows

B. Hand and arm gestures

1) Pointing gestures
2) Isolated signs

C. Body gestures in which there is involvement of full body motion as in

1) For medical rehabilitation and athletic training, recognizing human gaits.
2) Tracking movements of two people interacting outdoors;
3) Analysis of movements of a dancer for generating graphics and matching music.

D. Hand and Arm gestures:

1) Recognition of sign languages, hand poses, and entertainment applications
2) Head and Face gestures: someexamplesare:

a) Raising the eyebrows;
b) Winking;
c) Nodding or shaking head
d) Opening the mouth to speak;
e) Direction of eye gaze,
f) Looking surprised; and
g) Flaring the nostrils, happiness, disgust, fear, anger, sadness, contempt, etc.

Camera is used to pick the gravity center of the points and hand with maximal distances, in a multisystem, from the center gives the locations of tips of the finger, which are then utilized to obtain a skeleton image. In a special camera that supplies dept information was used to identify hand gestures. Other computer vision methods used for hand gesture recognition include particle filters, orientation histogram, neural networks, Fourier descriptors, principal component analysis, specialized mappings architecture. The classification is performed in the curvature space in visual hand recognition system. It involves to find the boundary contours of the hand and it is robust in rotation, scale, translation, it is extremely demanding. We focus on hand gesture classification using inputs from low-resolution digital cameras and classify the acquired images using features that are extracted by non-computationally intensive image processing techniques.

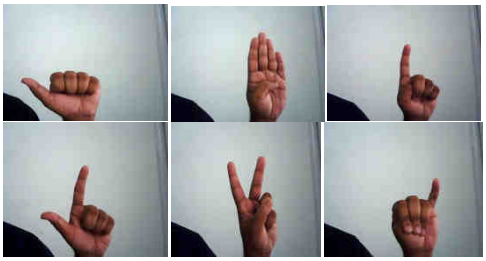

Fig. 1. General hand images used for various method

V. DIFFERENT METHODS OF CLASSIFICATION

A) Fuzzy C-means clustering algorithm

The input image is converted into HSV colourspace and segmented using Thresholding technique. After removing the noise from the image, 13elements were extracted as the feature vector, first element represents the aspect ratio of thebounding box of the hand, and the rest 12 parameters represent grid cells of the image. Eachcell represents the mean gray level in the 3 by 4 block partition of the image, where the meanvalue of each cell is the average brightness of these pixels, FCM clustering algorithm used forclassification gestures. The system implemented under intricate background and invariant lightconditions. The system trained with 120 samples, 20 samples for 6 gestures. The systemimplemented with recognition accuracy of 85.83% and recognition time 2-4 seconds. The system consists of training phase, and running phase. In training phase,several feature vectors of input gestures has been stored in a database. In testing phase the inputimage features vector compared with those stored in the database for classification using Euclidean distance.

B) Hidden Markov Model (HMM)

A time-domain process demonstrates a Markov property if the conditional probability density of the current event, given all present and past events, depends only on the *j*th most recent event. If the current event depends solely on the most recent past event, then the process is termed a first order Markov process. This is a useful assumption to make, when considering the positions and orientations of the hands of a gesturer through time.

The HMM [3], [7] is a double stochastic process governed by:

1) An underlying Markov chain with a finite number ofstates

2) A set of random functions, each associated withone state. In discrete time instants, the process is in one ofthe states and generates an observation symbol according tothe random function corresponding to the current state.

Eachtransition between the states has a pair of probabilities, definedas follows:

1) Transition probability, which provides the probability forundergoing the transition;

2) Output probability, which defines the conditional probabilityof emitting an output symbol from a finite alphabetwhen given a state.

The HMM is rich in mathematical structures and has beenfound to efficiently model spatio–temporal information in a naturalway.The model is termed “hidden” because all that can beseen is only a sequence of observations.

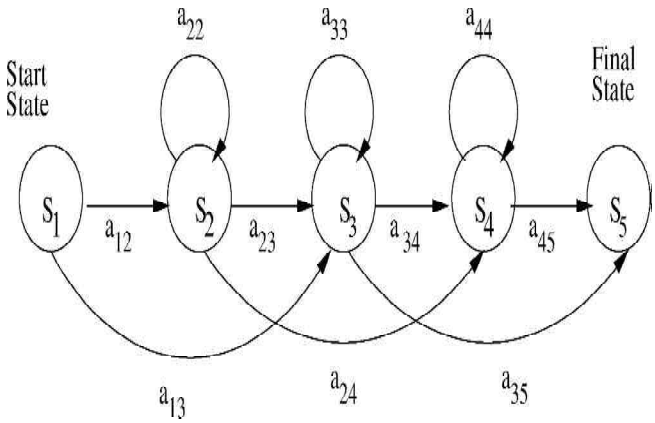

Fig. 2. Five-state left-to-right HMM for gesture recognition

The generalized topology of an HMM is a fully connectedstructure, known as an ergodicmodel, where any

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

state can bereached from any other state. When employed in dynamic gesturerecognition, the state index transits only from left to rightwith time, as depicted in Fig. 1. The start state $s1$ and final state sN, for $N = 5$, are indicated on the figure1. The global structure of the HMM is constructed by parallelconnections of eachHMM($\lambda 1, \lambda 2, \ldots, \lambda M$), whereby insertion (Or deletion) of a new (or existing)HMMis easily accomplished.

C) K-Nearest Neighbor (KNN)

K-Nearest Neighbours. This classification method uses thefeature-vectors gathered in the training to find the k nearestneighbours in a n-dimensional space. The training mainlyconsists of the extraction of (possible good discriminable)features from training images, which are then stored for later classification. Due to the use of distance measuringsuch as the Euclidian or Manhattan distance, the algorithmperforms relatively slowly in higher dimensional spaces orif there are many reference features. In an approximate nearest neighbours classification was proposed, whichprovides a better performance.

D) Multilayer Perceptron.

A Multilayer Perceptron (MLP)classifier is based on a neural network. Therefore, MLPsrepresent a trainable classifier (similar to Hidden MarkovModels). They use three or more layers of neurons that areall connected. During the training phase, theweights of theconnections between the neurons are adapted, based on the feedback that describes the difference between the output and the expected result. In a MLP classifier was used torecognize 26 different ASL gestures with a recognition rateofup to 98.7%, depending on the number of features usedto describe the gesture.

E) Self-organizing Map

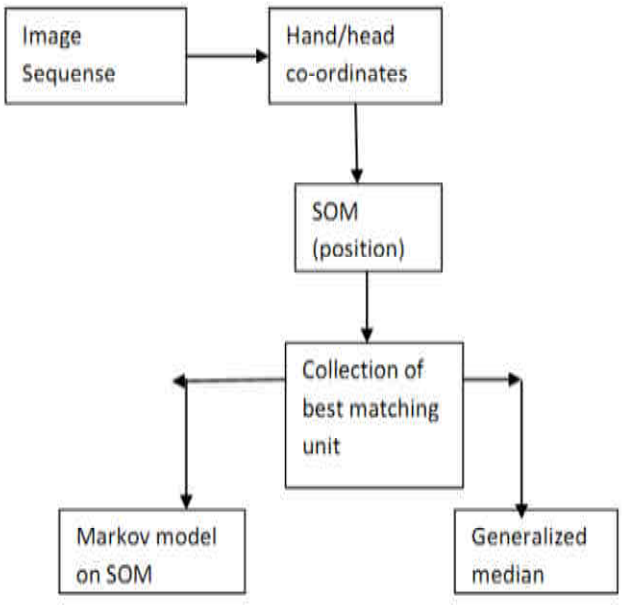

Fig. 3. Flow of self-organising map

In gesture recognition module systems find out the particular gesture moment. System first take image as input do extraction on it with colour detection technique. After doing extraction segmentation is done. In this it divide image in various object and recognize objects of image in various direction. System also gives the particular axis of image movement. It involves two gesture module for finger recognition and head recognition.

F) Support Vector Machine

Machine learning is known as subfield of artificial intelligence. Through machine learning we can develop methods for enabling a computer to learn. Over the period there are so many techniques developed for machine learning. Support vector machine (SVM) has been firstly introduced by Vapnik and gained popularity because of its exiting feature such as better empirical performance. Support vector machine (SVM) is a classification and regression technique that uses machine learning theory to maximize the accuracy of prediction. Support vector machines is discussed for two-class problems. First, support vector machines, in which training data are linearly separable in the input space and the second one support vector machines for the case where training data are not linearly separable and map the input space into the high-dimensional feature space to enhance linear separability in the feature space. For a two-class problem, a support vector machine is trained so that the direct decision function maximizes the generalization ability.

G) Principal Component Analysis (PCA)

In this section, we will study the hand gesture recognition through Principal Components Analysis, but we will need some mathematical background to understand the method. This method is called: PCA or Eigenfaces. It is a useful statistical technique that has found application in different fields (such as face recognition and image compression). This is also a common technique for finding patterns in data of high dimension too. Before realizing a description of this method, first mathematical concepts that will be used in PCA are described.

1. Standard Deviation:

In statistics, we generally use samples of population torealize the measurements. For the notation, we will use thesymbol X to refer to the entire sample and we will use thesymbol Xi to indicate a specific data of the sample.

2. Variance

Variance is another measure of the spread out of data in aset. In fact it is quite the same as the standard deviation.

3. Eigenvectors

The eigenvector of a linear operator are non-vectorswhich, when operated on by the operator, result in a scalarmultiple ofthemselves. The scalar is then called theEigenvalue associated with the eigenvectors.

4. Eigenvalue

Each eigenvector is associated to an Eigenvalue. TheEigenvalue could give us some information about theimportance of theeigenvector. The Eigenvalues are reallyimportant in the PCA method, because they will permit torealize some threshold to filterthe non-significanteigenvectors, so that we can keep just the principal ones.

VI. RESULTS

1. Fuzzy C means clustering

The successful recognition ratio does not drop below 86% while the false negative percentage remains in low level as well. In some cases the percentage of false negative decisions for the preparation and retraction phase was higher than 15%.

2. Hidden Markov Model (HMM)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

It is demonstrated that HMM is a feasible parametric model for building a gesture based system. HMM is a doubly stochastic model and is appropriate for coping with the stochastic properties in gesture recognition. HMM is more suitable for continuous gesture recognition because no hand marking is needed.

3. K- Nearest Neighbor (KNN)

KNN has better recognition rate depending on k value. In this algorithm system is tested with different k values. In KNN method for k=2 recognition rate are between 96.66% to 99.58%. In this the algorithm is robust to traditional problems of gesture recognition like illumination variations, hand sizes, hand initial positions and can be used for HCI efficiently.

4. Multilayer Perceptron (MLP)

Pattern recognition is a powerful technique for harnessing the information in the data and generalizing about it. The system is developed through learning rather than programming. Performance of MLP is at least as good as classical statistical modelling and better on most of problems.

5. Self-organizing Map (SOM)

Input hand gesture images are changed and the network is simulated repeatedly and the SOM neural network generates correct answer for most of the trials. SOM is one of the most powerful neural network models. It is competitive learning network.

6. Support Vector Machine (SVM)

The online classification showed the total accuracy 72%, this can be improved. The lower accuracy rate of the online classification compared with offline classification is due to the use of lower resolution camera in the classification process in testing of online accuracy.

7. Principal Component Analysis (PCA)

Accuracy rate of PCA is 72.22% as tested in the project. Depending on illumination variations conditions, brightness and contrast, inverting images, gray images the result is obtained.

VII. COMPARISON OF VARIOUS METHODS

Segmentation	Features Vector Representation	Classifier
HSV threshold	13 parameters as a feature vector; the first parameter represents the ratio aspect of the bounding hand box and 12 parameters are the mean values of brightness pixels in the image	Fuzzy C-Means (FCM) algorithm
thresholding	Three features vector: Boundary Chord's size FFT, Boundary Chord's size, Boundary Chord's size histogram	MLP Neural Network / Dynamic Programming (DP) matching
YCbCr color space	Three angles from hand shape: RC Angle, TC Angle, Distance from the palm center.	Gaussian distribution
Threshold	13 data item for postures/ 16 data item for gestures	Back propagation network / Elman recurrent network
GMM and YCbCr space	Orientation quantization	HMM
thresholding	Template for posture/ hand location for gesture	structural analysis for postures / HMM for gestures.
Segmented manually	Three different features group; Invariant Moments, K-curvature Geometric, Shape Descriptors.	Multilayer Perceptron Artificial Neural Network.

Table 1. Summary of segmentation, features vector representation, and recognition of hand gesture recognition system.

VIII. SUMMARY AND CONCLUSION

Constructing an efficient hand gesture recognition system is an important aspect for easily interaction between human and machine. In this work we provided a comparative study on various gesture recognition systems with emphasis on segmentation and features detection and extraction phases which are essential for gesture modelling and analysis.

REFERENCES

[1] RafiqulZaman Khan & Noor Adnan Ibraheem (2012), omparative study of hand gesture recognition system" Department of Computer Science, A.M.U., Aligarh, India.

[2] SushmitaMitra, Senior Member, IEEE, and Tinku Acharya, Senior Member, IEEE (2007), "Gesture Recognition: A Survey"

[3] PrateemChakraborty, PrashantSarawgi, AnkitMehrotra, Gaurav Agarwal, Ratika Pradhan (2008), "Hand Gesture Recognition: A Comparative study"IMECS 2008, 19-21 March, 2008, Hong Kong.

[4] Sudheesh.P, Gireesh Kumar T. July 2012, "Vision based Robot Navigation for Disaster Scenarios"AmritaVishwa Vidyapeetham Coimbatore, India.

[5] Kwang-HoSeok, Chang-Mug Lee, Oh-Young Kwon, Yoon Sang Kim (2009),"A Robot Motion Authoring using Finger robot interaction" Korea University of Technology and Education Cheonan, Korea.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.