Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Survey on Hand Gesture Recognition Methods

Jignasha B. Panchal*, Karshan P. Kandoriya
CSE Department, Parul Institute of Engineering & Technology,
Gujarat, India

Abstract—Hand Gesture Recognition system received great attention in recent years because it provides human computer interaction and sign language. In this paper survey of hand gesture recognition is provided. Hand Gesture Recognition is contained three stages: Pre-processing, Feature Extraction and Classification. Each stage contains different methods which are reviewed. Summary of different methods used for hand gesture recognition in existing system with comparison of all method with its benefits and drawbacks are provided.

Keywords— Hand Gesture Recognition (HGR), Human Computer Interaction (HCI), Sign Language, Pre-Processing, Feature Extraction, Classification.

I. INTRODUCTION

Image Processing is a type of signal allocation in which input is image, like photograph or video and output may be image or characteristics or attributes associated with that image. Image is function f(x, y) and gives intensity at position (x, y) [1]. Hand gesture is a body language provides specific meaning is established through our language centre by palm and finger position and shape. Hand gesture provides a natural and expressive communication mode for human dialog or speech. Gesture consists of static hand gesture and dynamic hand gesture. In Static hand gesture shape of hand gesture is used to express meaning or feeling. Dynamic Hand Gestures are composed of series of hand movements for track motion of hand [3]. Hand Gesture Recognition is topic in computer vision because of more application, such as HCI, sign language interpretation, and visual surveillance. Primary goal of gesture recognition is to create a recognition system that can identify specific human gesture and use them to convey meaningful information or to control devices which is used in HCI [5].

HGR is performed using Glove based HGR or Non Vision based HGR, Vision based HGR, Depth based HGR. (1) In Glove-based method data glove is applied to collect the data of joints in order to extract the gesture, and use the neural network to analyse these data. Then, the meaning of the gesture will show up. This method owns the advantages of less input data, high speed, and it can get 3D information about hands or fingers movement directly. The disadvantage of of method is impossible for users to wear gloves all the time in nature condition. Glove is expensive and heavy [2]-[3]. (2) Vision-based method is work well in bad condition and user take hand as direct input to system for human computer interaction. Using camera image sequence collects and identifies, analyze gesture [2] [3]. (3) Depth-based method has a high robust, real-time identification and high precision. Depth camera based on the technologies such as time of flight (TOF), structure light, 3d laser scanning is so expensive that its utility has been limited. So researchers' interest was greatly inspired when Microsoft launched the Kinect device [3].

II. HAND GESTURE RECOGNITION

Hand gesture recognition process is classified into three step: 1. Pre-processing 2. Feature Extraction and Estimation 3. Classification or Recognition as illustrated in fig.1

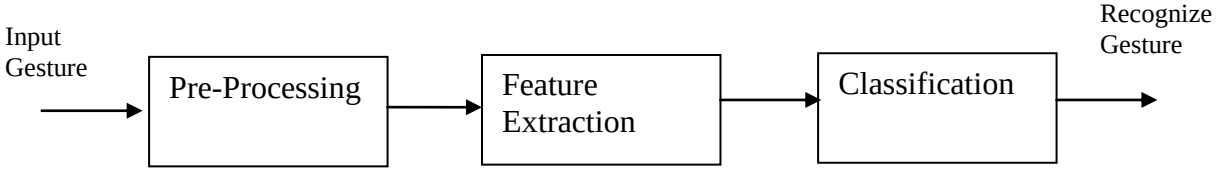

Fig.1 Gesture Recognition Process

A. Pre-Processing

In preprocessing stage operation are applied to extract the hand gesture from its background and prepare the hand gesture image for feature extraction. In this stage segmentation, edge detection as well as smoothing and other filtering process can occur [6].

Step 1: Segmentation

It is the process of dividing an image into multiple parts. This is used to identify objects or other relevant information or data in digital images. There are many different ways to perform image segmentation, including: HSV, YCbCr, Skin color detection, Depth data.

1) HSV: was widely used for image segmentation. HSV means Hue refers to different colors(such as red, green), S refers to saturation (for instance, the difference between deep blue and light blue) and V refers to value (also considered as brightness).It is represent image accurately with institutive values. Consider (r, g, b) are the coordinates of red, green and blue respectively. The values are from 0 to 1. Max represents the maximum value among r, g and b. Min represents the minimum value among the r, g and b. The *h* means the hue angle (between 0-360), *S* means the saturation, calculated as follow [2].

$$H=\begin{cases} 60*\frac{g-b}{max-min}-0 & if\ max=r\ and\ g\geq b \\ 60*\frac{g-b}{max-min}-360 & if\ max=r\ and\ g<b \\ 60*\frac{b-r}{max-min}-120 & if\ max=g \\ 60*\frac{r-g}{max-min}-240 & if\ max=b \end{cases} \quad (1)$$

$$S=\begin{cases} \frac{max-min}{max}=1-\frac{min}{max} & otherwise \\ 0 & if\ max=0 \end{cases} \quad (2)$$

$$V=max \quad (3)$$

2) YCbCr (luminance, blue difference, red difference): Widely used color spaced method used for image segmentation. To obtain YCbCr, RGB colour space transformed into as follow[2]:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix}=\begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}*\begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix}*\begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

3) Skin color detection: It is based on threshold technique. The segmentation method can be divided into two groups, one is the threshold value only used in one region and the other is the threshold value used in the whole image [1]. The threshold value is used to classify the points in this way similar features on the image into the same class. Feature of the image is usually considering colour, gray scale or other X factor is one particular characteristic of image. Consider input image is *f*(*x, y*), the output is *f'*(*x, y*), the threshold is *T*, the segmentation process using threshold can be described as following equation (5) [2].

$$F'(x,y)=\begin{cases} a1\ f(x,y)<T \\ a2\ f(x,y)\geq T \end{cases} \quad (5)$$

Assume the threshold is described as : *T*(*C*(*x, y*),*G*(*x, y*),*X*(*x, y*))

If T=T (G(x, y)), it means the threshold is only related to the gray scale of the isolated image. If T=T (G(x, y),X(x, y)), which means the threshold will be decided by the gray scale and other characteristics If T=T(C(x, y),G(x, y),X(x, y)), implies that the threshold will be decided by the position, gray scale and other characteristics.

4) Depth Data: This is one type of colour limitation method usually limits the environment by wearing coloured markers or using a fixed colour background. Availability of depth information of image objects can overcome hand movement difficulty and noise these difficulties easily. The grey scale of the pixel in depth image is related to the distance between the surface plane and camera, so it is not be affected by space colour, illumination and other colour factors. It can be used for gesture recognition in 3D space [2].

Step 2: Filter

Filter is most used for remove noise in digital image. Various type of filtering technique is available like: Spatial filter, Mean Filter, Median Filter, Morphological filter, Gaussian filter.

1) Spatial Filter:

It can be effectively used to remove various types of noise in digital images and it is typically operate on small neighbourhoods ranging from (3x3) to (11x11).Spatial filters are implemented with convolution masks for provides a result that is a weighted sum of the values of a pixel and its neighbours. It is also called linear filter [7].

2) Mean Filter:

The mean filter is averaging filters and it has operated on local groups of pixel called neighbourhood and replace the centre pixel with the average of the pixel in this neighbourhood. Calculate average of intensity values in a m x n region of each pixel (usually m = n)[7] take the average as the new pixel value using following equation (6)

$$h(i,j)=\frac{1}{mn}\sum_{k\epsilon m}\sum_{l\epsilon n}f(k,l) \quad (6)$$

The normalization factor m, n preserves the range of values of the original image.

3) Median Filter:

It is nonlinear filter. In median filter size of local neighbourhood is defined and then operates on it. The centre pixel is replaced by the median or the centre value present among its neighbour, unlike by average. Median filter is excellent in removing isolated extreme noise pixels while substantially retaining spatial details [7].

4) Morphological Filter:

A morphological filtering [9] was developed to obtain a smooth, closed, and complete contour of a gesture by using a sequence of dilation, opening, closing and erosion operations. In general, the dilation and erosion operations on a binary image A and with a structuring element B are defined as follows.

Dilation:

If A and B are sets in the 2-D integer space Z^2; x=(x1,x2) and Φ is the empty set, then, the dilation of A by B is

$A \oplus B=\{x \mid (B')x \cap A \neq \Phi\}$

Where B' is the reflection of B. Dilation consists of obtaining the reflection of B about its origin and then shifting this reflection by x. The dilation of A by B is the set of all x displacements such that B and A overlap by at least one nonzero element. Set B is commonly referred to as the structuring element.

Erosion:

The erosion of A by B is

$A \otimes B = \{x \mid (B)x \subseteq A\}$

it indicates that the erosion of A by B is the set of all points x such that B, translated by x, is contained in A. Note that dilation expands an image and erosion shrinks it.

Opening:

The opening of set A by structuring element B is

$A \circ B = (A \otimes B) \oplus B$

thus, the opening of A by B is simply the erosion of A by B followed by a dilation of the result by B Opening generally smoothes the contour of an image, breaks narrow isthmuses, and eliminates thin protrusions.

Closing:

The closing of set A by structuring element B is

$A \bullet B = (A \oplus B) \otimes B$

this says that the closing of A by B is simply the dilation of A by B followed by the erosion of the result by B. Closing tends to smooth sections of contour but, and opening, it generally fuses narrow breaks and long thin gulfs, eliminates small holes, and fills gaps in the contour.

5) Gaussian Filter:

Gaussian smoothing is very effective for removing noise. It can provide weights give higher significance to pixels near the edge .It is linear low pass filters. It is computationally efficient and rotationally symmetric, degree of smoothing is controlled by filtering with m x m mask, the weights are computed according to a Gaussian function:

$g(i,j)=c.e^{\frac{i^2+j^2}{2\sigma^2}}$ σ is user defined (7)

Step 3: Edge detection

Edges are basic feature and carry useful information about object boundaries. It can be used for image analysis object identification and filtering. In edge detection first edge point corresponding image find and then merged to form line and object outlines. There are various type of edge detection method like: Sobel, Prewitt, Laplacian, Canny edge detection [6].

1) Sobel Operator:

It is recognized as one of the best operators. It is utilize two (3x3) masks. The sobel edge detection masks look for the horizontal and vertical direction and then combine this information into a single metric [6].

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

(a) Row mask (b) Column mask

At each pixel location there are two numbers S1 corresponding to the result from the row mask and S2 corresponding to the column mask these numbers are used to compute two metrics, the edge magnitude and the direction which are defined as follow

Edge Magnitude=$\sqrt{S1^2 + S2^2}$ (8)

Edge Direction: $\tan^{-1}\frac{S1}{S2}$ (9)

2) Prewitt Operator:

It is similar to sobel operator but with difficult mask coefficients. The masks are defined as follow the edge magnitude and direction are defined as follows.[6]

$$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

(a) Row mask (b) Column mask

Edge Magnitude=$\sqrt{P1^2 + P2^2}$ (10)

Edge Direction: $\tan^{-1}\frac{P1}{P2}$ (11)

3) Laplacian Operator:

Laplacian masks are rotationally symmetric that means edges at all orientation contribute to result. They are applied by selecting one mask and convolving provides directional information and provide which edges is brighter [6].

4) Canny edge detection [6]:

Step 1: Filter out any noise. The Gaussian filter is used for this purpose. An example of a Gaussian kernel of size=5 that might be used is shown below:

$$K=\frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \quad (12)$$

Step 2: Find the intensity gradient of the image. For this, we follow a procedure analogous to Sobel:

1. Apply a pair of convolution masks (in x and y directions):

$$Gx=\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix} \quad Gy=\begin{bmatrix}-1 & -2 & -1\\ 0 & 0 & 0\\ 1 & 2 & 1\end{bmatrix} \qquad (13)$$

2. Find the gradient strength and direction with:

$$\text{Edge Magnitude}=\sqrt{Gx^2+Gy^2} \qquad (14)$$

$$\text{Edge Direction: } \tan^{-1}\frac{Gx}{Gy} \qquad (15)$$

The direction is rounded to one of four possible angles (namely 0, 45, 90 or 135)

Step 3: Non-maximum suppression is applied. This removes pixels that are not considered to be part of an edge. Hence, only thin lines (candidate edges) will remain.

Step 4: Hysteresis: The final step. Canny does use two thresholds (upper and lower):

1. If a pixel gradient is higher than the upper threshold, the pixel is accepted as an edge
2. If a pixel gradient value is below the lower threshold, then it is rejected.
3. If the pixel gradient is between the two thresholds, then it will be accepted only if it is connected to a pixel that is above the upper threshold.

B. Feature Extraction and Estimation

Feature extraction is a complex problem and in which whole image or transformed image is taken as input. The goal of feature extraction is to find the most differentiate information in the recorded images. Feature extraction operates on two-dimensional image array but produces a feature vector [7]. Feature Extraction has various type of method like: Contour detection, movement invariant, Histogram orientation gradient.

1) Contour detection:

Contour detection in images is a fundamental problem in computer vision tasks. Contours are distinguished from edges as follows.

Contour tracking algorithm [9]:

Step 1: Scan the image from top to bottom and left to right to find first contour pixel marked as X1(i, j). Then record position of X1 into x and y arrays, respectively, i.e., x[1] = I and y[1] = j and set i1 = 0, i2 = 0, j1 = 0, j2 = 0 and search next contour pixel.

Step 2: To scan clockwise direction searching PJ sequences are XJ (i, j − 1), XJ (i − 1, j − 1), XJ (i − 1, j), XJ (i − 1, j +1), XJ (i, j+1), XJ (i+1, j+1), XJ (i+1, j) and XJ (i+1, j−1) respectively. where j = 2, 3, 4, . . . ,N. and N is total number of detected contour pixel.

Step 3: Scan clockwise until the next pixel value Xj = 1.

Step 4: If XJ = 1, position of XJ ≠ (i1, j1) and position of XJ≠ (i2, j2), Store position of XJ into x and y arrays and set (i, j) = position of XJ . When J > 3 set (i1, j1) = position of XJ−1 and (i2, j2) = position of XJ−2.

Step 5: If step 4 become false then set x[J] = x[J−1], y[J] = y[J − 1], (i, j) = position of XJ−1, (i1, j1) = position of XJ and (i2, j2) = position of XJ−2.

Step 6: Repeat steps 2 to 5 until position of XJ = position of X1.

By computing proposed contour tracking algorithm, the position of all contour pixel is stored into x and y arrays.

2) Moments Invariant:

Features associated with images are called invariant if they are not affected by certain changes regarding the object view such as translation, scaling, rotation and light conditions [7].

Hu proposed 7 famous invariant moments. It is used in many image classification approaches and not affected by translation, scale or rotation [2].

Regular moment is defined as follow:

$$m_{pq=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^p y^q f(x,y)dxdy,p,q=0,1,2,..} \qquad (16)$$

Central moment is defined as follow:

$$\mu_{pq=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{x})^p(y-\bar{y})^q f(x,y)dxdy,p,q=0,1,2,..} \qquad (17)$$

$$\bar{x}=\frac{m_{10}}{m_{00}} \quad \bar{y}=\frac{m_{01}}{m_{00}}$$

Normalized Central moment is defined as follow:

$$\eta_{pq=\frac{\mu_{pq}}{\mu_{00}^{\gamma}}},\ \gamma=\frac{p+q+2}{2}, p+q=2,3,\ldots \qquad (18)$$

Normalized Central moment, which is invariant to image scale change.

Based on normalized central moments, Hu introduced seven nonlinear functions:

$$\emptyset_1=\eta_{20}+\eta_{02} \qquad (19)$$

$$\emptyset_2=(\eta_{20}-\eta_{02})^2+4\eta_{11}^2 \qquad (20)$$

$$\emptyset_3=(\eta_{30}-3\eta_{12})^2+(3\eta_{21}-\eta_{03})^2 \qquad (21)$$

$$\emptyset_4=(\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2 \qquad (22)$$

$$\emptyset_5=(\eta_{30}-3\eta_{12})(3\eta_{30}+\eta_{12})+[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2]+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2] \qquad 23)$$

$$\emptyset_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (24)$$

$$\emptyset_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) + [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (25)$$

Hu's seven moment invariants are invariant to image scaling, translation, and rotation. It can be used for disjoint shapes. Disadvantages of hue is high computation cost; the recovery of image is difficult and computation expensive.

3) Histogram Orientation Gradient:

Feature extraction using HOG is first find hand area and generate orientation map of that and compute histogram for input as feature vector and classify using correlation function. Orientation template correlation (OTC) is a technique which operates on the orientation map of an image not on the image itself. It is indicate 2D vector contain the strengths and angles that present locally throughout the image [8].

The mean and standard deviation of orientation histogram about the x and y directions, correlation function, Cor (x, y) according as j shifts.

$$\text{Cor(j)} = \frac{1}{A * \sqrt{(E(q1x) - E(q2x))^2 + (\sigma(q1x) - \sigma(q2x))^2}} + \frac{1}{B * \sqrt{(E(q1y) - E(q2y))^2 + (\sigma(q1y) - \sigma(q2y))^2}} \quad (26)$$

E() is the expectation operation and $\sigma()$ is the standard deviation operation. q1 and q2 are the patches of input images about each x-axis and y-axis. q1 and q2, are the patches of model images and **A,** B are constants. After correlation function calculate orientation histogram for each block calculated and classified.

C. Classification or Recognition

Recognition process is used by feature extraction method and classification algorithm. In recognition stage, new feature vector can be classified after system be trained with enough. So, classification accuracy is good. Classification is done using popular methods like, Artificial neural network, Hidden Markov Model, Template matching, Dynamic time warping.

1) Artificial Neural Network (ANN):

An Artificial neural network is an information processing system and has performance characteristics in common with biological neural networks. It is developed as generalizations of mathematical models of human cognition or neural biology based on the assumptions that 1. At many simple elements Information processing occurs called neurons. 2. Over connection links signals are passed between neurons. 3. Weight is associated with each connection link, which in a typical neural net, multiplies the signal transmitted. 4. Applies an activation function (usually nonlinear) on each neuron to its net input (sum of weighted input signals)[7].

Nodes are the basic units in the neural network and in between node weight is assigned. Consider the input are *X1*, *X2*, … , *Xk*, the value vector between input layer and hidden layer is *Wih* and its formula is as shown below:

$$\text{input} = \sum_1^h W_{ih} X_i \quad (27)$$

The output from input layer to hidden layer is calculated as follow:

$$W_i^m = f\left(\sum_{m=1}^m W_{ih} X_k - \theta_i^1\right), i = 1,2,.. \quad (28)$$

Then output of hidden layer is calculated as follow:

$$\text{output} = \sum_1^k W_{im} X_k \quad (29)$$

If the output is not exactly the same with respect to results then turn to operate reversely, it means the error signals will be returned along the original link and repeatedly change the coefficients of all layers until that the actual output is equal to the input.

2) Hidden Markov Model (HMM) [2]:

HMM is also one of the most popular approaches and also used in dynamic hand gesture. A HMM is a Markov Model (MM) with hidden parameters treated as the state. The state transitions produce variant results are visible but state cannot be visible. The sequence of state produce sequence of output includes hidden information. A HMM can be shown as Figure. Set of hand positions in each state *A1* to *Ai* represent the input states. The probability of transferring from one state to another state *Pab*, P*ba* represent the state transitions. The one specific posture or one gesture *B1* to *Bi* represents the corresponding outputs. So the database of HMM has many samples per single gesture, the relation between the number of samples and the accuracy is directly proportional, and the speed is inversely proportional.

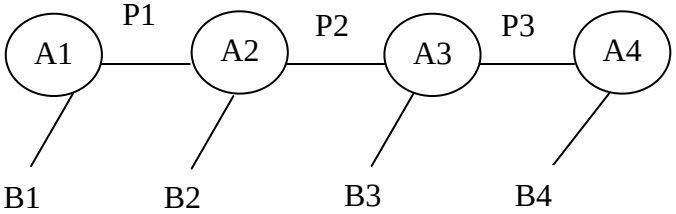

Fig. 2 HMM

3) Template Matching:

Pattern recognition is based on template matching. Matching is used to determine the similarities between two entities (points, cures or shapes) of the same type. In template matching, a template (typically a 2D shape) or a prototype of the pattern to be recognized is available and pattern to be recognized is matched against the stored template and moves

the template image to all possible positions in a larger source image and computes a correlation value that indicates how well the template matches the image in that position like translation and rotation and scale changes. Template matching is a computationally demanding process, but using faster processors has made more feasible. It has a number of disadvantages. For instance, when changes in the imaging process viewpoint or large interclass variations among the patterns, it would fail if the patterns were distorted because of [2].

4) Dynamic time Warping (DTW):

DTW has been used to find the optimal alignment of two signals. The DTW algorithm calculates the distance between each possible pair of points out of two signals and that signal is as defined as their associated feature values. Using these distances to calculate a cumulative distance matrix and find the least expensive path through this matrix and represent ideal warp of synchronization of the two signals that causes the feature distance between their synchronized points to be minimized. This process is done after the signals are normalized and smoothed. DTW is useful in various applications, such as speech recognition, data mining and movement recognition [5].

III. LITERATURE REVIEW OF HAND GESTURE RECOGNITION

Mohammad et. al.[10] proposed the HGR using in 3D environment using depth camera. For segmentation it is uses YCrCb color space for accurate skin detection. It is use encoded nonlinear RGB. After detecting hand remove noise from image. For contour detection find boundary region of detected hand then draw rectangular box around the contour and find center of hand. Three features are effectively used to recognize the gesture: orientation, Area of hand, Angle of Box. Calculate orientation of hand then check current hand area with previous are if previous area is greater than current are then moving away from camera else moving towards camera. After find area of hand find angle of box by using draw box around hand contour area. Classification is done using Hidden Markov Model (HMM). In this markov chain emits a sequence of observable outputs. It is model parameter $\lambda = (A, B, \pi)$ Where A is transition Matrix, B is Observation Matrix and π is initial probability distribution. Probability of each class sequence can be computed and highest probability is considered as output. After that polling method is used for classify gesture. This proposed method is provide 80.67% accuracy. It is not more dynamic.

Rajat et. al.[11] proposed system provide dynamic HGR by using intel image processing OpenCV. It has three stage: 1. detection & tracking: $L\alpha\beta$ color space is used for hand detection where α component represents pixel position between red and green and β is component represents between yellow and blue. Thresolding is combine with color space to extract information. CAMSHIFT (Contionously Adapted Mean Shift) Algorithm to track hand. 2. feature extraction: Hu moment invariant has 7 moments which are invariant to rotation and translation. 3. training & recognize. HMM with left-right banded model (LRB) technology. In this model state is not directly visible but output is visible and it is dependent on state. HMM has three topologies and it is fully connected and each state can be go back to itself. After completion of process maximal likelihood from gesture set selected. It has 90% accuracy. It is focus on speed not recognition rate.

Saad [12] proposed the process for detecting, understanding and translation sign language gesture to vocal language. There are two mode recording and translation mode. In recording mode, user adds gesture to dictionary. In translation mode, gesture is compared with gesture stored in dictionary. In this method call this gesture "Recording Translation Gesture" or "RTG". This method contains four step: 1.Getting joint of interest 2. Normalize the skeleton frame data 3. Build link list of temporary storage data 4. Detecting gesture. Dynamic Time Wrapping algorithm is used to compare gestures. It is provide 91% accuracy. It is not suitable for finger movement.

Jaya [13] proposed method for HGR using Microsoft Kinect Sensor. In preprocessing stage, Kinect depth feature for background segmentation of hand gesture. Feature Extraction is done using find contour of hand then calculate convex hull, convexity defects. After calculating defects extract image features. Classification is done using naïve Bayesian classifier. As a feature set Convex hull and convexity defects are are considered as two attributes to classify data. Number of images for each attribute set is considered as weighted sum. It is provide good classification rate. It is not recognize hand orientation and not recognize dynamic gesture.

Archana [14] introduces a HGR system to recognize the alphabets of Indian Sign Language. There are five modules: 1. Hand tracking: Camshift algorithmis used to hand tracking.. 2. Hand Segmentation: HSV color space is applied on track hand for segmentation. 3. Feature Extraction: Shape representation techniques are used for feature extraction and define relationship between features. 4. Recognition: Genetic algorithm is better choice for managing randomness for natural samples and hand gesture analysis. In this algorithm first initialize population and evaluate fitness of each individual. Choice of parent based on fitness. Create new individual using 2 point cross over. Chose elements randomly and choose worst one and replace with new individual and check stop if best solution. It is easy to use and inexpensive approach but not more dynamic and applies on all.

Ayan [15] proposed hand gesture recognition under varying illumination. Overall process has two step: 1. Segmentation: To achieve segmentation of hand from complex background using Information Measurement Ratio based threshold technique. It is work more efficient in manner. 2. Training & recognition : Principle Component Analysis (PCA), Euclidian distance is used for recognize hand accurately and efficiently. In training phase, first convert image matrix to column vector and calculate average matrix and difference matrix. Find covariance matrix using Difference matrix and calculate Eigen value and Eigen vector. In recognition phase, transform image into its components and compute vector and its Euclidian distance. PCA is fast, simple and accurate. It has also reduced dimensionality of picture and cheaper in terms of time. It is give reduce time complexity but not apply on orientation of hand gesture recognition.

Parul [16] This paper provide method for recognize sign language. It has three steps: 1. Pre-processing: First input sign image in RGB to convert into Lab colour space where L is lightness a and b are two colour channel. Lab color space

for convert image into binary image and hand region is cropped and perform filter for remove noise. 2. Feature Extraction: It is done using Area, height, centered-origin Euclidance distance, Average height. 3. Classification: Feed forward back propagation algorithm is used for training and classification. Feed forward back propagation training algorithm is supervised learning algorithm. In this input and output vector are provided for training network. Target vector is provided for each input gesture. It is provide 85% accuracy. It is not recognize dynamic gesture.

IV. COMPARISON OF EXISTING METHODS

TABLE I. COMPARISON OF EXISTING METHODS

Method	Segmentation	Feature Extraction	Classifier	Accuracy	Advantages	Disadvantages
[10]	YCrCb	Orientation hand, area of hand, angle of palm	HMM	80.67%	Effective, costless	Cannot more dynamic, depth camera is needed
[11]	Lαβ	Hu invariant moment, hand orientation	HMM with LRB	90%	Good result for real time gesture, focus on speed not on recognition rate	Suffer from noise
[12]	Depth data	Kinect, Gesture dictionary	DTW	91%	Detect all gesture, good result	not detect finger moment
[13]	Depth data, median filter	Contour, convex hull, convexity defect	Naive Bayesian classifier	-	Good classification rate	Two hand cannot recognize, cannot recognize dynamic gesture, not recognize orientation
[14]	HSV, Camshift	Shape representation techniqe	Genetic	-	Inexpensive approach, recognize single and double hand	Not use for dynamic gesture, not apply on all
[15]	IMR based Thresholding	PCA	Euclidance distance	98%	Simple, accurate, reduce dimension of picture	Time complexity high, not recognize dynamic, cannot recognize orientation of hand
[16]	Lab	Orientation, Euclidance distance	Neural Network	85%	Simple & efficient method	Not recognize dynamic gesture

V. CONCLUSIONS

In this paper various method for each stage of Hand Gesture Recognition system used. In pre-processing stage, perform sequence of steps like segmentation, filter and edge detection which is used different method and algorithm to detect hand region and filter noise. In Feature extraction stage most commonly used contour tracking algorithm for identify contour of hand, moment invariant for identify feature of time series based capture gesture and histogram orientation gradient are used. For classification HMM, ANN, Template matching and DTW is used. ANN is provided good classification and learning rate. Detail discussion of recent recognition system is provided in the paper. Comparison of existing methods is listed with its advantages and disadvantages.

ACKNOWLEDGMENT

I would like to acknowledge Mr. Karshan Kandoriya, Assistant Professor at Parul Institute of Engineering and Technology, Vadodra for suggesting this idea to work upon.

REFERENCES

[1] Rafael C. Gonzalez, Richand E.Woods ,*Digital Image Processing*, 3rd edition, 2009.
[2] Shaui Yang, Parashan, Peter, *Hand Gesture Recognition: An Overview*, IEEE, pg. 63-69, 2013.
[3] Lingchen Chen, Feng Wang, Hui Deng, Kaifan Ji, *A Survey on Hand Gesture Recognition*, International conference on computer sciences and applications, IEEE, pg. 313-316, 2013.
[4] Rafiqul Z.Khan, Noor A. Ibraheem, *Hand Gesture Recognition : A Literature Review*, International general of Artificial Intelligence and Application, Vol. 3, No. 4, pg. 162-174, July 2012.

[5] Haitham S. Badi, Sabah Hussein, "Hand Posture and Gesture Recognition Technology", Springer, pg. 1-8, 2014.
[6] Raman Maini, Himansu Aggraval, *A study and comparison of various image edge detection technique*, International General of Image Processing, vol. 3, Issue 1, pg. 1-12, 2010.
[7] Haitham Hasan, S. Abdul-Kareem, *Static hand gesture recognition using neural networks*, springer, pg.147-181, 2012.
[8] Hyung-Ji Lee, Jae-Ho Chung, *Hand Gesture Recognition using Orientation Histogram*, IEEE, pg. 1355-1358,1999.
[9] Dipak K. Ghosh, Samit Ari, *A Static Hand Gesture Recognition Algorithm Using K-Means Based Radial Basis Functional Neural Network*, IEEE, Dec-Jan. 2011
[10] Mohammad I. Khan, Sheikh Md. Ashan, Razib Chandra, *Recognition of Hand Gesture Recognition Using Hidden Markov Model*", IEEE, pg. 150-154, 2012.
[11] Rajat Shrivastava, *A Hidden Markov Model Based Dynamic Hand Gesture Recognition System using OpenCV*, International Advance Computing Conference, IEEE, pg. 947-950, 2012.
[12] Saad, Majid, Bilal, Salman, Zahid, Ghulam, *Dynamic Time Wrapping based Gesture Recognition*, IEEE, pg.205-210, 2014.
[13] Jaya, Ashutosh, *A Method for Hand Gesture Recognition*, International Conference on Communication System and Network Technology, IEEE, pg. 919-923, 2014.
[14] Archana, Rucha, Sanjana, Surbhi, Mithila, *Hand Gesture Recognition for Indian Sign Language*, International Conference on Computer Communication and Informatics, IEEE, pg. 1-4, 2012.
[15] Ayan, Pragya, Rohit, *Information Measure Ratio Based Real Time Approach for Hand Segmentation with a Focus on Gesture Recognition*, International Conference on Intelligent Systems, Modeling and Simulation, IEEE, pg. 172-176, 2011.
[16] Parul, Hardeep, *Neural Network Based Static Sign Gesture Recognition System*, International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 2, Issue 2, pg. 3066-3072, 2014.

