International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1464-1469 www.ijaem.net ISSN: 2395-5252

Survey on Various Gesture Recognition Technologies and Techniques

Anju Roslin Francis[1], Ashna Ajay M[2], Denil Scaria[3]
[1,2,3]*Student, Amal Jyothi College of Engineering, Kottayam, Kerala*
Corresponding Author: Ashna Ajay M

Submitted: 01-06-2021 Revised: 14-06-2021 Accepted: 16-06-2021

ABSTRACT: Motions are one of the most regular expressive route for correspondences between humans and computers in a virtual framework. Hand gestures are a form of non-verbal correspondence as it provides a liberated articulation that is significantly more other than of any other body parts. In various applications, gestures are utilized as a distinctive interface dissecting, demonstrating, reproduction, and acknowledgment of motions. This paper provides a survey on recent techniques in gesture recognition with emphasizes on hand gestures. To review the static hand posture, various tools and algorithms are applied in the method of gesture recognition systems, including the Hidden Markov model, connectionist models, and fuzzy clustering. The challenges in each area, as well as prospective research directions, are discussed.

KEYWORDS:Gesture Recognition System, Hand Gesture Methods, Artificial Neural Network, Fuzzy Clustering, Hidden Markov Model

I. INTRODUCTION

Gestures and facial expressions can be used for daily human interactions while human-computer interactions still require understanding and examine signals to interpret the required command that made the interaction understandable and natural. Recently, the design of special input devices witnessed great thrust that further enabled simple and complex interaction between humans and computers.

With the perspective of the model framework for gesture recognition, gesture segmentation, gesture modeling and analysis, and gesture recognition, the subsequent discussion systematically summarizes the present research status of dynamic vision recognition technology in computer vision and analyses its shortcomings. To illustrate static hand posture approaches, various techniques, and algorithms are used in gesture recognition systems are elaborately discussed. The results indicate that for models performing gesture recognition supported by simple wearable devices, gesture recognition supported deep vision sensors, and multi-method cross-fusion gesture recognition are going to be trending technologies in this field in the future

II. GESTURE RECOGNITION TECHNOLOGY

The important steps of gesture recognition techniques are analysis and identification of posture, human response, and proxemics. Gesture recognition allows computers to more correctly read human body language than earlier text user interfaces or even GUIs (graphical user interfaces), which still rely on guessworkon mouse and keyboard input. Thiscould make usual input on such devices and even make them redundant.

A. Instrument Glove Approach

Hand gesture recognition can be done by utilizing wearable sensors attached directly to the hand with gloves, according to Munir Oudah et al [1].These sensors, monitor the physical response to hand movements or finger flexing. The glove-based sensor system are made wireless and mobile by attaching a sensor to microcontrollers using standard wireless protocol. It offered simple commands for a computer interface. By identifying the exact coordinates of the location of the palm and fingers, gloves use different sensor types to capture hand motion and position.

Figure 1:**Sensor-based data glove**

Curvature sensor, angular displacement sensor, optical fiber transducer, flexes sensors, and accelerometer sensor are among the sensors that operate on the angle of bending in the same way. These sensors utilize different physical aspects according to their designed type.

B. Camera Vision-Based Sensor

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1464-1469 www.ijaem.net ISSN: 2395-5252

In vision-based methods, the system may only use a single camera or several cameras to capture the image necessary for natural human-computer interaction, and no additional devices are employed. It is now easier than ever to detect hand movements that can be later on used, in a variety of applications. This is supported by a variety of open-source software libraries. These approaches use a camera to replace the instrumented glove. Cameras such as RGB camera, time of flight (TOF) camera, thermal cameras, or night vision cameras are used. To identify hand, computer vision-based algorithms are developed employing concepts like skin color, appearance, motion, skeleton, depth, to segment the hand from the captured image

Figure 2:Computer vision–based camera using a marked glove or just a naked

C. Colored Marker Approaches

According to "Survey on Various Gesture Recognition Technologies and Techniques," marked gloves or colored markers are gloves worn by the human hand that have specific colors to implement the process of tracking the hand and identifying the palm and fingers, allowing the geometric features needed to form a hand shape to be brought out

The color glove shape could be made up of small regions of contrasting colors, or it could be as simple as using three different colors to reflect the fingers and palms on a wool glove. When opposed to instrumented data gloves, this device has the advantage of being easier to use and less expensive. [2]

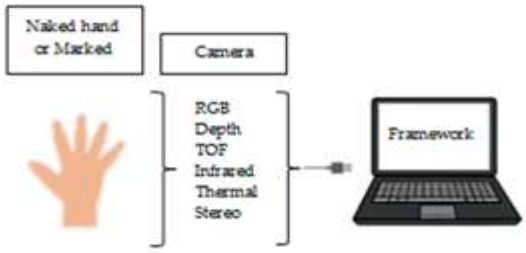

Figure 3:Color-based recognition using glove marker

III. VISSION BASED HAND GESTURE RECOGNITION

The data required for identification is obtained through vision-based technologies using only the human hand. These methods are natural, easy, and the user directly communicates with the system. Vision-based technology handle with some image feature such as texture and color for obtaining data needed for gesture analyze. Many techniques are employed for detecting hands they are discussed here.

A. Appearance Base Recognition

To represent visual appearance such as the hand, this method relies on bringing out picture features and distinguishing these parameters with attributes, collected from the input image frames. The properties are derived from the pixel intensities without any prior segmentation. Due to the ease with which 2D image properties can be retrieved, the approach runs in real-time and is regarded easier to implement than the 3D model. This model is also capable of distinguishing between different skin tones.

B. 3D Model-Based Recognition

As discussed by Munir Oudah, et al [1], the 3D model essentially depends on the 3D Kinematic hand mode which has a huge degree of freedom Here hand parameter estimation is obtained by comparing the input image with the two-dimensional aspects projected by the three-dimensional hand model. The 3D model often integrates human hand characteristics like pose estimation by shaping a volumetric, skeletal, or 3D model that is similar to the user's hand. The depth parameter is added to the model to improve accuracy, with the 3D model parameter renovating through the matching process.

IV. GESTURE RECOGNITION TECHNIQUES

For gesture recognition, several concepts are needed, including pattern recognition, motion detection and analysis, and machine learning based recognition. Different tools and techniques are implemented in the gesture recognition process such as computer vision, image processing, pattern recognition, and modeling based on statistics.

A. Hidden Markov Models

The Hidden Markov Model (HMM) is based on the Markov chain [3] elaborated Daniel Jurafsky & James H. Martin. A Markov chain is a model that estimates the probabilities of sequences of random variables, or states, each of which may take on values from a collection of possibilities.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1464-1469 www.ijaem.net ISSN: 2395-5252

These sets can be words, or tags, or symbols representing the given environment or object. A Markov chain makes the very strong assumption that the current state is the only thing that matters if we want to predict the future in the series. The states that came before the current state have no impact on the future unless they are influenced by the current state.

When computing the probability of a succession of recognised occurrences, a Markov chain is required.In many other circumstances, however, the relevant events are concealed or not visible at all. A Hidden Markov Model (HMM) analyses both seen and hidden events in our given probabilistic model. An HMM is described by the following.

$Q = q_1, q_2 \ldots q_N$	a set of N states
$A = a_{11} \ldots a_{ij} \ldots a_{NN}$	a transition probability matrix A, each ai j representing the probability of moving from state i to state j, s.t. $\sum_{i=1}^{N} a_{i,j} = 1 \ \forall i$
$O = o_1 o_2 \ldots o_t$	a sequence of T observations, each one drawn from a vocabulary $V = v_1, v_2, \ldots, v_V$
$B = b_i(o_t)$	A collection of observation likelihoods, also known as emission probabilities, each of which expresses the likelihood of an observation o_t being made from a state i.
$\pi = \pi_1, \pi_2, \ldots, \pi_N$	An initial probability distribution over states. π_i is the probability that the Markov chain will start in state i. Some states j may have $\pi_j = 0$, meaning that they cannot be initial states. Also, $\sum_{i=1}^{n} \pi_i = 1$

Two simplifying assumptions a first-order hidden Markov model is formulated. For starters, just like in a first-order Markov chain, the likelihood of a particular state is exclusively controlled by the state preceding it:

Markov Assumption:

$P(q_i | q_1 \ldots q_{i-1}) = P(q_i | q_{i-1})$

Second, the probability of an output observation is solely determined, by the state that generated the observation qi, not by any other states or observations:

Output Independence:

$P(o_i | q_1 \ldots q_i, \ldots, q_T, o1, \ldots, o_i, \ldots, o_T) = P(o_T | qi)$ [3]

B. Artificial Neural Networks

Artificial neural networks, proposed by Rob J Hyndman and George A [4], are forecasting approaches based on simple mathematical models of the brain.They allow for complex nonlinear interactions between the response variable and its predictors. A neural network is a layered collection of "artificial neurons."The predictors (or inputs) are at the bottom of the pyramid, while the forecasts (or outputs) are at the top.There may also be “hidden neurons” in intermediate levels.

There are no hidden layers in the simplest networks, which are similar to linear regressions. Figure 4 depicts the neural network linear regression model with four predictors. The “weights” are the coefficients associated with these predictors. A linear combination of the inputs is used to construct the forecasts. In the neural network system, the weights are chosen using a "learning algorithm" that minimizes a "cost function" such as the MSE. A much more efficient method of training the model we can use linear regression. [4]

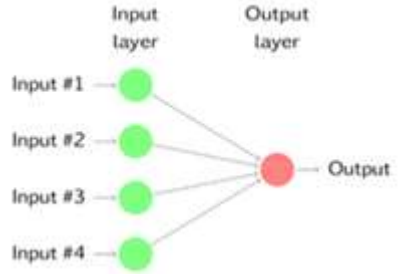

Figure 4:A simple neural network equivalent to a linear regression.

When we add an intermediate layer containing hidden neurons to the neural network, it becomes non-linear. Figure 5 depicts a basic example.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1464-1469 www.ijaem.net ISSN: 2395-5252

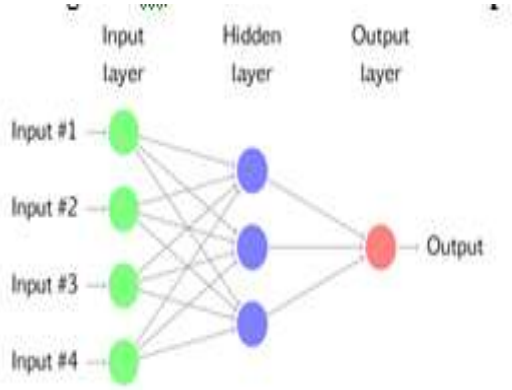

Figure 5:A neural network with four inputs and one hidden layer with three hidden neurons

Each layer of nodes in a multilayer feed-forward network receives input from the layers before it. The outputs of one layer's nodes are used as inputs in the next layer. The inputs to each node are combined using a weighted linear combination.Before being output, the result is changed by a nonlinear function. For example, Rob J et al suggests, the inputs into hidden neuron jj in Figure 5 are combined linearly.

$$Z_L = b_j + \sum_{i=1}^{4} w_{i,j}, X_i$$

This is then updated in the hidden layer using a nonlinear function like a sigmoid to provide the input for the next layer.This lessens the impact of severe input values, makingthe network more resistant to outliers

$$s(Z) = \frac{1}{1 + e^{-Z}}$$

The parameters b_1,b_2,b_3,b_1,b_2,b_3 and $w_{1,1},\ldots,w_{4,3}$,$w_{1,1},\ldots,w_{4,3}$ are "learned" from the data. The values of the weights are often restricted to prevent them from becoming too large. The "decay parameter" is a parameter that limits the weights and is commonly set to 0.1.

To begin, the weights are given random values, which are then modified based on the observed data. As a result, the predictions made by a neural network include a degree of randomness. As a result, the network is trained several times with different random start points, with the results summed. [4]

C. Fuzzy Clustering Algorithm

The work by A.K. Jain et al [5], reviews the data-clustering algorithm. In this, based on the data points and the distance between the cluster center, this algorithm assigns membership to each data point. The closer the data is near the cluster center, the more likely it is to be associated with it.Each data point's membership should be equal to one when added together. After each iteration membership and cluster centers are reconditioned according to the formula

$$\mu_{ij} = 1/\sum_{k=1}^{c} (d_{ij}/d_{ik})^{\frac{2}{m-1}}$$

$$v_j = \frac{\sum_{i=1}^{n} (\mu_{ij})^m x_i}{\sum_{i=1}^{n} (\mu_{ij})^m}, \forall_j = 1,2,3,\ldots c$$

Where,

'n' is the number of data points.

'm' is the fuzziness index m ϵ [1, ∞].

μ_{ij}' represents the membership of i^{th} data to j^{th} cluster center.

'v_j' represents the j^{th} cluster center.

'c' represents the number of cluster center.

'd_{ij}' represents the Euclidean distance between i^{th} data and j^{th} cluster center.

D. Histogram Based Feature

Histogram-based thresholding is mainly based on characteristics of the image involves shape, intensity form of pixels. Rosenfeld [7] and Lee [6] derived the optimal threshold using histogram concavity analysis for the given image. In entropy-based thresholding, the distribution of grey levels in images has a different aspect. They considered the image pixels as background and foreground classes. In each class, entropy has been determined and summed to obtain total entropy. Finally, an optimal threshold has been evaluated by maximizing the total entropy.

Local thresholding such as Niblack Thresholding (NT) determined threshold value by sliding a window area over the grayscale image. Mean and the standard deviation of all the pixels in the window area was used to compute the final threshold. Sauvola thresholding (ST) was used to improve the NT method by a dynamic range of standard deviation of grayscale images for determining the threshold. The feature extraction is intermediate between pre-processing and recognition system based on techniques of local or global features. Local features extract the specific key points based on the magnitude and direction of the images. Local feature descriptors such as the local binary pattern (LBP), histogram of gradients (HOG), scale-invariant feature transform (SIFT), and speeded-up robust features (SURF) were commonly used. M. Hruz et al. used Local binary patterns for monochrome vision-based images and produced a high discriminative feature for recognition [8]. The merits of LBP are sufficient for invariant grayscale images.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1464-1469 www.ijaem.net ISSN: 2395-5252

David G. Lowe described [9] that SIFT feature descriptors use Gaussian operators to find the intensities of gradients. Based on gradients intensity, determine the dominant key location by comparing the neighborhood pixels using suppression of local maxima. Then, descriptors were generated by the computation of magnitude and direction around the key points of the images. The demerits of SIFT are a time-consuming process. These limitations are avoided by SURF local features, in which H. Bay calculated the interest points using a Hessian function, involve the second order of Gaussian derivative [10]. Then, select the appropriate interest points using local maxima. Finally, determine the dominant orientation of relevant key points by Haar features response which represents 64- dimensional descriptor vector using the sum of Haar response. The HOG feature descriptors were evaluated by the intensity gradient distribution of an image. First, divide the images into cells, and in each cell of the pixels, the histograms of gradients orientation were computed. Secondly, feature descriptors were determined by histogram bins. Global features describe the topological and statistical measurements based on the properties such as geometric features, shape, texture, and color of the entire image. Shape based descriptors such as Fourier descriptors, curvature scale, space descriptors, angular radial transform, and image moments.

V. DISCUSSION

Various software tools are used for implementing gesture recognition systems. Implementation tools may be programming languages like C, C++, Java and Python. For simplifying the work in image processing Matlab® can also be used. While in hardware implementation tool for recognizing gesture various systems are used according to their requirements. Mainly cameras are used for recognizing posture. 3D cameras, stereo cameras are generally used. Other than camera sensors are used it can provide correct information about the position and orientation of the hand using magnetic tracking. Sensors are attached in gloves. Flex sensors, accelerometer sensors, tactile sensors are generally used.

HMM is widely used in the field of gesture recognition. Combining particle filtering along with HMM for finding dynamic gesture recognition trajectories. Neural networks have a high degree of parallelism, self-adaptability, and certain learning capabilities. Some scholars have applied this method to the field of gesture recognition.In fuzzy clustering, the partitioning of sample data into groups in a fuzzy way is the main difference between fuzzy clustering and other clustering algorithms, where the single data pattern might belong to different data groups.

VI. CONCLUSION

Hand gesture recognition tries to addresses a shortcomings in HCI systems. Controlling things by hand is more natural, easier, more flexible, and cheaper, and there is no need to fix problems caused by hardware devices since none is required. Building an efficient human-machine interaction is an important goal of a gesture recognition system. Many applications of gesture recognition systems ranging from virtual reality to sign language recognition and robot control. In this paper, a survey of gesture recognition system, tools and techniques is presented, with a focus on hand gesture gestures. The major methods discussed include HMMs, ANN, and fuzzy clustering

REFERENCES

[1]. Munir Oudah, Ali Al-Naji, and Javaan Chahl, “Hand Gesture Recognition Based on Computer Vision: A Review of Techniques”, Journal of Imaging 6(8):73, 2020

[2]. Noor Adnan Ibraheem, RafiqulZaman Khan, “Survey on Various Gesture Recognition Technologies and Techniques”,International Journal of Computer Applications, Vol 50, 2012

[3]. Daniel Jurafsky & James H. Martin, “Hidden Markov Models”, Speech and Language Processing, 2020

[4]. Rob J Hyndman and George Athanasopoulos, “Forecasting: Principles and Practice”, Third Edition, 2008

[5]. A.K. Jain, M.N. Murty, P.J. Flynn, “Data Clustering: A Review”, ACM Computing Surveys, 1999

[6]. CK Lee, FW Choy, and HC Lam. Real-time thresholding using histogram concavity [image processing]. In Industrial Electronics, Proceedings of the IEEEInternational Symposium, pp 500–503., 1992

[7]. A. Rosenfeld and de la Torre, “Histogram concavity analysis as an aid in threshold selection(in image processing)”, IEEE Transactions on Systems, Man, and Cybernetics,13:231–235, 1983.

[8]. Hrúz M., Trojanová J., and Elezn M., "Local Binary Pattern based features for Sign

Language Recognition", Pattern Recognition And Image Analysis, Vol. 22, No. 4, 2012

[9]. David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", IJCV, 2004

[10]. H. Bay, A. Ess, T. Tuytelaars, and L. van Gool.,"Speeded-up robust features (Surf)",Computer Vision and Image Understanding (CVIU), 110(3):346–359, 2008.