

Simulation of Gesture Recognition for Physical Impairments Peoples

Jayashree.R¹, Dinesh.L²

Student, Computer Science and Engineering, St.Joseph's College of Engineering and Technology, Thanjavur, India¹

Asst.Prof., Computer Science and Engineering, St.Joseph's College of Engineering and Technology, Thanjavur, India²

Abstract: Gesture recognition is the mathematical interpretation of a human motion by a computing device. It can originate from any bodily motion or state but commonly originate from the face or hand. It focuses in the field include emotion recognition from the face and hand gesture recognition. It presents a vision-based user interface designed to achieve computer accessibility for disabled users with motor impairments. Recognizing gestures as input allows computers to be more accessible for the physically-impaired and makes interaction more natural in a gaming or 3-D virtual world environment. Therefore, it is necessary to develop easily accessible systems for computers to achieve their inclusion within the new technologies. These applications involving hidden Markov models, particle filtering and condensation, are discussed in detail. Hidden Markov models (HMMs) and related models have become standard in statistics, with applications in areas like speech and other signal processing, bioinformatics etc. Markov chain Monte Carlo (MCMC) is great stuff. MCMC revitalized Bayesian inference and frequents inference about complex dependence.

Keyword: gesture recognition, particle filtering, HCI

I. INTRODUCTION

Pervasive and ubiquitous computing integrates computation into everyday environments. The technological progress of the last decade has enabled computerized spaces equipped with multiple sensor arrays, like microphones or cameras, and multiple human-computer interaction devices. The development of technologies relying on high usability principles, exploited new communicational channels, such as eye blinking, voice, hand gestures, sip and puff, and electromyogram as effective control modalities. The use of hand gestures provides an attractive alternative to these cumbersome interface devices for human-computer interaction (HCI). User's generally use hand gestures for expression of their feelings and notifications of their thoughts. In particular, visual interpretation of hand gestures can help in achieving the ease and naturalness desired for HCI. Sign language recognition (SLR) and gesture-based control are two major applications for hand gesture recognition technologies. SLR aims to interpret sign languages automatically by a computer in order to help the deaf communicate with hearing society conveniently. Since sign language is a kind of highly structured and largely symbolic human gesture set, SLR also serves as a good basic for the development of general gesture-based HCI. Smart environments have enabled the computer observation of human (inter)action within the environment. The analysis of (inter)actions of two and more individuals is here of particular interest as it provides information about social context and relations and it further enables computer systems to follow and anticipate human (inter)action SLR are based on hidden Markov models (HMMs) which are employed as effective tools for the recognition of signals changing over time. On the other hand, gesture-based control translates gestures performed by human subjects into controlling commands as the input of terminal devices, which complete the interaction approaches by providing acoustic, visual, or other feedback to human subjects.

II. RELATED WORK

Hand gesture recognition involves segmentation of the hands, tracking them through occlusion, and the classification of hand's dynamic trajectories and static pose. For real-time gesture-based interfaces for assistive technologies, robustness is a critical requirement for its adoption. For hand segmentation, a commonly used method is to the prebuilt skin color histogram model into new video frames. These methods are likely to fail in true world conditions, where illumination is uncontrolled and the background is cluttered. Face and hands tracking is a special case of MOT problem. If gestures in the lexicon only carry trajectory information, the hand shape does not convey extra information, classical tracking approaches can be adopted. For example, CAMSHIFT and conditional density propagation (CONDENSATION) have been shown to successfully track the hands. Another technique widely used for object tracking is particle filters. To integrated color-based appearance models to a particle filter framework to enhance tracking under complex background and occlusion problem. Then applied the particle filter framework to multiple objects tracking. All the discussed algorithms so far, attempted to solve the MOT problem. To described a method to estimating human pose from static images using body part models by using the depth information, to proposed a method to predict 3-D positions of body joints from a single depth image. They solved the pose estimation problem through a simple per-pixel \classification problem. This method is effective for static gesture recognition; One of the most widely used techniques for gesture recognition is Hidden Markov Models (HMM). Common problems with HMM approach consist of finding the optimal parameters set and trajectory spotting for gesture temporal segmentation. CONDENSATION-based trajectory gesture recognition algorithm. Interaction between hands was not specifically tackled. Recently, a new type of challenge was attracted the attention of the gesture recognition community the

One Shot Learning Challenge . The one shot learning consists of learning a gesture category by only observing one instance of that gesture, similar to how humans learn.

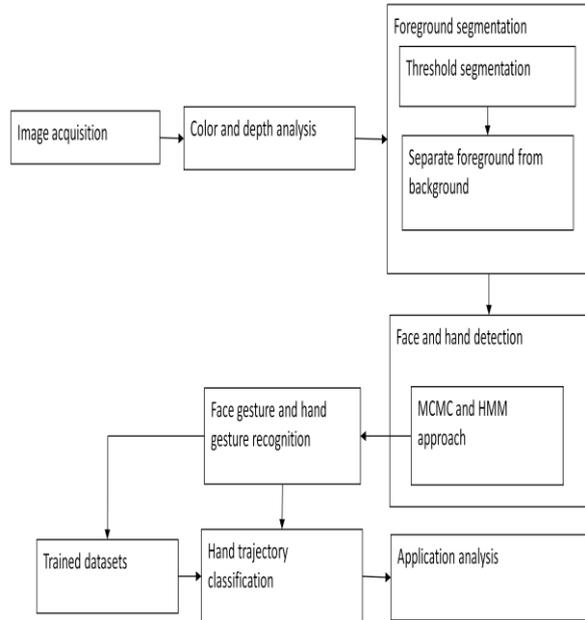


Fig.1. The flow chart of our designed gesture recognition method

III. LAYOUT-BASED GESTURE RECONGNITION

In gesture recognition an interaction model was incorporated to the color histogram-based particle filter framework to track hands through interaction and occlusion.

The machine vision-based gestural system included four parts: foreground segmentation, face and hand detection and tracking, hand trajectory classification, and robotic control policies. These parts are described in the following sections.

A. Layout Analysis of Foreground Segmentation

In the background was ruled out from the captured frames and the whole human body was kept as the foreground. Initially, the user’s body was treated as a foreground object in order to detect the user’s movements. Two steps were used to segment the foreground (refer to algorithm 1 in Table I).

In the first step, the sensed image acquired by a Kinect sensor was threshold using depth information. An example of a depth image is shown by Fig. 2(a), where the distance between objects and the depth sensor was mapped to intensity levels. The depth value of each pixel was defined as $D(i, j)$ with i and j indicating the horizontal and vertical coordinates of the pixel [Fig. 2(b)] in each frame of the video sequence.

TABLE I
Algorithm -Foreground Segmentation

Algorithm 1: Foreground Segmentation

Input: Low depth threshold T_{DL} ; High depth threshold T_{DH} ; pixel value of depth Image $D(i, j)$;

Output: pixel value of mask image $D_1(i, j)$; pixel value of foreground mask image $D_2(i, j)$.

$$D_1(i, j) = \begin{cases} 1; & T_{DL} \leq D(i, j) \leq T_{DH} \\ 0; & \text{otherwise} \end{cases}$$

$$T_{SH} = \max(\text{Area}(B_i)) \quad // B_i \text{ is the } i\text{th blob in the mask image } D_1$$

$$D_2(i, j) = \begin{cases} 1; & D_1(i, j) \in B_i \ \& \ \text{Area}(B_i) == T_{SH} \\ 0; & \text{otherwise} \end{cases}$$

Two absolute depth thresholds (a low threshold TDL and a high threshold TDH) were custom set by the user according to their relative distance to the depth sensor [Fig. 2(c)]. TDL was set to no less than a constant which was the minimum distance that can be registered by the depth sensor (due to its physical limitations). TDH was set to be the maximum distance that can be reached by the user while seated in a wheelchair.

B. Layout Analysis Face and Hand Detection

Face and hand detection was used to initialize the position of the face and hands for the tracking phase. Two 3-D histograms—a skin and a non-skin color histogram were created using the Compaq database and the HSV color space to achieve higher robustness for skin color detection. The mask image obtained from histogram back-projection is shown as in Fig. 3(a). To obtain the hand regions without the face, a face detector

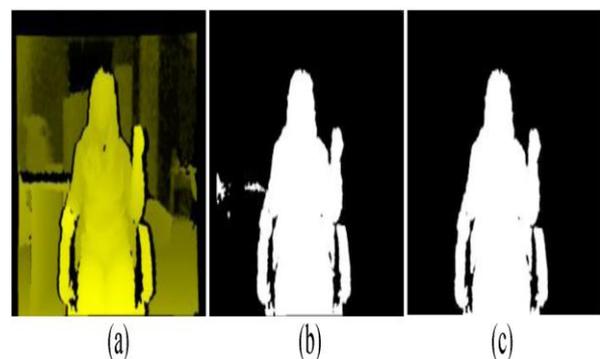


Fig. 2. Face and hand detection. (a) Skin color detection. (b) Hand extraction.(c) Face and hand localization

[Fig.3(c)] to remove the face region from the target image. Two largest blobs in the target image were then selected as hand regions [Fig. 3(b)]. This hand detection procedure was only used to provide automatic initialization to the particle filter tracking procedure. Afterwards the hands positions were continuously tracked by the particle filter.

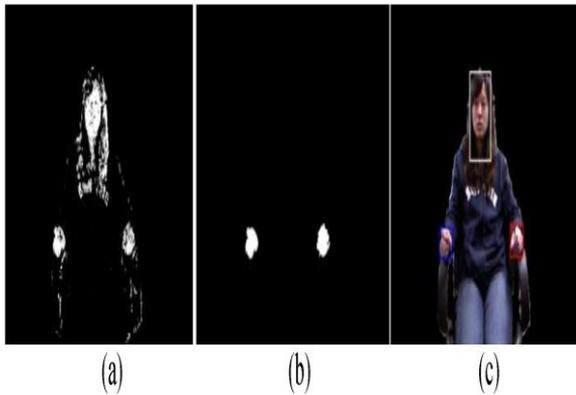


Fig.3. Face and hand tracking. (a) Skin color detection. (b) Hand extraction. (c) Face and hand localization

C. Layout Analysis of Face and Hand Tracking

A 3-D particle filter framework based on color, depth, and spatial information was used to track the face and hands through video sequences the particle filter tracking process consists of three main phases: predicting, measuring and resampling. In the proposed system, for the predicting phase, a second order auto regressive (AR) model. Many appearance-based models, such as contour, edge, and piece-wise, were used in object tracking. Color-based preprocessing using HSV space can facilitate the extraction of the aforementioned features for face and hands tracking. The extracted face and hands regions were used to compute the reference HSV histogram models (H^*_{f} , H^*_{h1} , and H^*_{h2}) for tracking initialization. During the resampling phase, each particle, assigned in the predicting phase, was reweighted by the observation likelihood function.

The contribution of this paper is three-fold: 1) prove the effectiveness of hand gestures as an 2) solve the frequently hand gesture interaction and occlusion problem through integration of color and 3-D spatial information 3) new gestures can be created and learned through the one shot learning paradigm, leading to an almost effortless training process .

IV. STRUCTURE-BASED GESTURE RECONGNITION

The architecture of the system is illustrated in Fig. 1. Eight gestures were selected to constitute the gesture lexicon which in turn was used to control the robots

A. Hand Tracking Through Interaction and Occlusion

Color-based particle filter tracking was effective for multiple independent objects tracking when the objects did not interact or occlude each other. The false merging problem denotes the situation that the tracker shift from the object being tracked to a different object that has higher observation likelihood. Conversely, the false labeling problem denotes the situation that the objects being tracked exchange their labels after interaction or occlusion occurred. In the proposed system, the face and both hands were tracked.

1) *Hidden Markov Models (HMM)*: HMM is a doubly stochastic model and is appropriate for coping with the stochastic properties in gesture recognition. Instead of using geometric features, gestures are converted into sequential symbols.. The concept of HMM can be used in solving three basic problems: the evaluation problem, the decoding problem, and the learning problem. In the learning problem, we provide model parameters in such a way that the model possesses a high probability of generating the observation for a given model and a set of observations. Therefore, the learning process is to establish gesture models according to the training data. In the evaluation problem we can score the match between a model and an observation sequence, which could be used for isolated gesture recognition. In the decoding problem we can find the best state sequence given an observation sequence, which could be used for continuous gesture recognition. The HMM approach to gesture recognition is motivated by the successful application of hidden Markov modeling techniques to speech recognition problems. The similarities between speech and gesture suggest that techniques effective for one problem may be effective for the other as well. First, gestures, like spoken languages, vary according to location, time, and social factors. Second, body movements, like speech sounds, carry certain meanings. Third, regularities in gesture performances while speaking are similar to syntactic rules. Therefore, linguistic methods may be used in gesture recognition. To develop a gesture interface, some criteria are needed to evaluate its performance such as meaningful gestures, suitable sensors, efficient training algorithms, and accurate, efficient, on-line/real-time recognition. Meaningful gestures may be very complex, containing simultaneous motions of a number of points. However, these complex gestures should be easily specifiable. The trained models are the representations of all gestures that the system must recognize. In the latter method of specification, a description of each gesture is written in a gesture description language, which is a formal language in which the syntax of each gesture is specified.. This problem would be avoided if the model parameters were determined by the most likely performance criterion. Because gesture is an expressive motion, it is natural to describe such a motion through a sequential model. Based on these considerations, HMM is appropriate for gesture recognition. A multi-dimensional HMM is able to deal with multi-path gestures which are general cases of gesture recognition.

Advantages

- Effective
- Can switch variations in proof structure
- Optional fields
- Varying field ordering

2) *Markov chain Monte Carlo (MCMC)*: The readership of the Proceedings with a class of simulation techniques known as Markov chain Monte Carlo (MCMC) methods. These methods permit a practitioner to simulate a dependent sequence of random draws from very complicated stochastic models. The main emphasis will be placed on one MCMC method known as the Gibbs

sampler. The first model was called the competition potential (CP) model. The idea of this model comes from the joint Markov random fields (MRF) theory. Our goal in this paper is to explore the prospects for rational process models of perceptual inference based on MCMC. MCMC refers to a family of algorithms that sample from the joint posterior distribution in a high dimensional model by gradually drifting through the hypothesis space of complete interpretations, following a Markov chain that asymptotically spends time at each point in the hypothesis space proportional to its posterior probability. MCMC algorithms are quite flexible, suitable for a wide range of approximate inference problems that arise in cognition, but with a particularly long history of application in visual inference problems. The chains of hypotheses generated by MCMC shows characteristic dynamics distinct from other sampling algorithms: the hypotheses will be temporally correlated and as the chain drifts through hypothesis space, it will tend to move from regions of low posterior probability to regions of high probability; hence hypotheses will tend to cluster around the modes. Here we show that the characteristic dynamics of MCMC inference in high-dimensional, sparsely coupled spatial models correspond to several well-known phenomena in visual perception, specifically the dynamics of multi stable percepts. Our goal here is a simpler analysis that comes closer to the standard MCMC approaches used for approximate inference in Bayesian AI and machine vision, and establishing a clearer link between the mechanisms of perception in the brain and rational approximate inference algorithms on the engineering side.. A variety of psychological phenomena have natural interpretations in terms of Monte Carlo methods, such as resource limitations, stochastic responding and order effects. The Monte Carlo methods that have received most attention to date as rational process models are importance sampling and particle filtering, which are traditionally seen as best suited to certain classes of inference problems: static low dimensional models and models with explicit sequential structure, respectively.

Advantages:

- Distinct molecular dynamics imitations, Monte Carlo simulations are modern from the limitations of solving Newton’s equations of motion.
- This liberty allows for skill in the offer of moves that produce trial configurations within the numerical mechanics assembly of choice.

B, Hand Trajectory Classification

Hand tracking results were segmented as trajectories, compared with motion models, and decoded as commands for robotic control.in each frame in the video sequence, the centroids of the face and hands were obtained from the tracking stage. The motion model for each gesture trajectory was created based on the data collected from gestures performed by ten subjects. Even though the trajectories for each gestures performed by different subjects or the same subject in different instances

may look similar, the precise duration of each sub trajectory within t trajectory were different.

The CONDENSATION algorithm was employed to classify hand gesture trajectories in the lexicon it employs a set of weighted samples to fit the observed data. The original algorithm in was extended to work for two hands. The original expression $St = (\mu, \phi, \alpha, \rho)$ (the state at time t) was extended to

$$St = (\mu, \phi^i, \alpha^i, \rho^i) \\ (\mu, \phi^{right}, \phi^{left}, \alpha^{right}, \alpha^{left}, \rho^{right}, \rho^{left})$$

where, μ is the index of the motion models, ϕ is the current phase in the model, α is an amplitude scaling factor, ρ is a time dimension scaling factor, and $i \in \{\text{right hand, left hand}\}$. The gestures in the lexicon were spotted using a rest position gesture as when the subjects put their hands on the arm rest (neutral position) with no hand movement. A dynamic motion model was created for the rest position gesture. The segment between two recognized discontinuous rest position gestures is treated as a spotted gesture.

TABLE II
Hand Tracking through Interaction and Occlusion using 3-D particle filter tracking algorithm

Algorithm 2: 3D Particle Filter tracking
<p>Input: Reference HSV histogram models H_f^r, H_{h1}^r, and H_{h2}^r; Optimal parameter $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$.</p> <p>Output: Centroids and the associated bounding box of the face and hands.</p> <p>1. Initialize: //Initialize particle states and weight for face and both hands as: $x_0^i = x_0^*, \omega_0^i = \frac{1}{n}$, where $i = 1, \dots, n$</p> <p>2. Predict, Measure and Resample: //Select k; for $i=1,2,3$ //(1-face, 2-right hand, 3-left hand) for $r=1$ to N $x_{i,t}^r = A_1(x_{i,t-1}^r - x_0^i) + A_2(x_{i,t-2}^r - x_{i,t-1}^r) + x_{i,0}^r + BV_t$ //Compute candidate histograms H^r $D_i(H^*, H^r) = [1 - \sum \sqrt{H^* H^r}]^{\frac{1}{2}}$ //Calculate the weight: $\omega_{i,t}^r = k \exp(-\lambda D_{i,t}^2)$ end for Normalize the weights and resample the particles Estimate $\hat{x}_{i,t} = \sum_{r=1}^N \omega_{i,t}^r x_{i,t}^r$ //Check interaction if interactions happens for object i and j for $q=1, \dots, N$ //computer interaction likelihood ψ_1 and ψ_2: Compute $\psi_{1,t}^q(X_{i,t}^q, X_{j,t}^q)$ and $\psi_{2,t}^q(X_{i,t}^q, X_{j,t}^q)$ using (8) and (9) //Calculate the weight: $\omega_{i,t}^q = \omega_{i,t}^q \cdot \psi_{1,t}^q \cdot \psi_{2,t}^q$ end for Normalize the weights and resample the particles. Estimate $\hat{x}_{i,t} = \sum_{r=1}^N \omega_{i,t}^r x_{i,t}^r$ end if end for</p>

C, Robotic Control Policies

The commands decoded by gesture recognition results were sent to control the mobile robot and the robotic arm. A gesture lexicon was designed such that users with physical impairments can perform the gestures with minimal effort. These gestures were found through a series of interviews conducted with subjects with upper mobility impairments. Users to perform lab experiments without the need to physically attend them. In the laboratory case study experiment, a mobile robot was controlled by the gesture algorithm to transport a beaker to a position near a robotic arm. The robotic arm was activated by the operator to add a reagent to the beaker and then, the mobile robot was brought back to its original position. The gestures (a)–(h) (from the lexicon in Fig. 4) were used and mapped to the commands: change mode, robotic arm action, go forward, go backward, turn left, turn right, stop, and enable robotic arm.

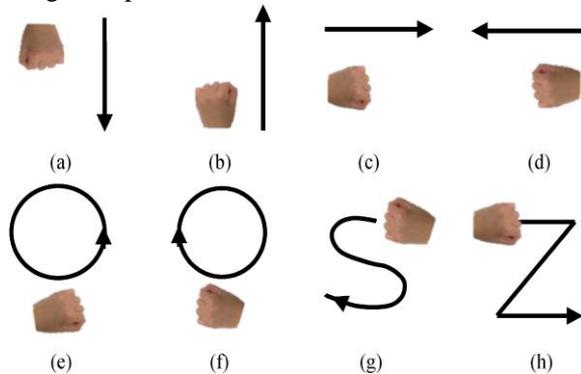


Fig. 4. Gesture lexicon. (a) Downward. (b) Upward. (c) Rightward. (d) Leftward. (e) Counter-clockwise circle. (f) Clockwise circle. (g) S. (h) Z.

The two robots were controlled by three modes—discrete, continuous, and hybrid mode (discrete plus continuous mode). In discrete mode, for each issued command, the mobile robot moved a fixed increment of distance. While in continuous mode, the mobile robot responded to a given command, until the stop command was issued.

V. QUANTITATIVE EXPERIMENTAL ANALYSIS

Gesture recognition consists of detection & tracking for hand and face and recognition. However, the main technical contributions of this paper are the two main methods for recognition schemes compatible with Real time applications. We perform experiments to evaluate the two schemes over benchmark datasets.

A. Dataset

A dataset of 16 videos (4 subjects x 4 activities) was used to evaluate the proposed tracking algorithm. The local likelihood $p(z_t|x_t)$ was calculated using the 3-Dcolor histograms and two interaction models as the algorithm mentioned in Table II. The performance of the proposed method—competition potential and motion consistency

(CPMC)—was compared to other existing methods, such as Markov Chain Monte Carlo (MCMC)-based particle filter tracking. The tracking performance of these algorithms was evaluated by employing three metrics: false merging, false labeling and tracking accuracy. The false merging is defined as the situation where the tracker of one hand occupies 80% of the area of the other hand. The false labeling is defined as the situation where the trackers of both hands change positions during/after interaction or occlusion. The tracking accuracy is defined by

$$\text{Tracking Accuracy} = \frac{\text{Total number of, (true positives + true negative)}}{\text{Total number of tracked frames}}$$

where a true positive is defined as the situation where as a target object is present and the tracker was able to find it. True negatives are instances where the target object is not present.

B. Gesture Recognition Performance

The recognition performance for the CONDENSATION algorithm with our training procedures (CONDENSE) was compared to four other existing state-of-the-art recognition algorithms: 1) Basic motion 2) Motion-based PCA 3) DTW and 4) HMM. After applying each gesture recognition method to our data set, the results shown in Table III were obtained.

TABLE III
Gesture Recognition Performance

Method	Basic	PCA	DTW	HMM	CONDENSE
Accuracy (%)	54.6	66.0	67.4	94.2	95.9

Continuous and hybrid modes require commands to be issued only when the robot needs to change directions or stop, therefore fewer operations were required for continuous and hybrid modes than for discrete mode for the task observed.

VI. CONCLUSION

The importance of gesture recognition lies in building efficient human-machine interaction. Its applications range from sign language recognition through medical rehabilitation to virtual reality. Since skin and non-skin shade histogram models were utilized to instate the face and hands' centroid, the execution of the framework may be influenced when the clients wear short sleeves. Moreover, it was normal that clients will be situated inside the working separation to go determined by the Kinect sensor. A connection model was joined into the shade based molecule channel schema for hand following. At the point when there was no communication between the face and hands, various autonomous molecule channels followed the clients' developments. At the point when cooperation was available, the various free molecule channel trackers were joined with a collaboration model to comprehend false fusing and false naming issues. Results demonstrated that HMM-based distinguish systems may convey comparable results to our strategy. Consequently, higher distinguish could be attained by utilizing trajectories order based strategy.

REFERENCES

- [1] M. R. Ahsan, "EMG signal classification for human computer interaction: A review," *Eur. J. Sci. Res.*, vol. 33, no. 3, pp. 480–501, 2009.
- [2] J. A. Jacko, "Human–computer interaction design and development approaches," in *Proc. 14th HCI Int. Conf.*, 2011, pp. 169–180.
- [3] I. H. Moon, M. Lee, J. C. Ryu, and M. Mun, "Intelligent robotic wheelchair with EMG-, gesture-, and voice-based interface," *Intell. Robots Syst.*, vol. 4, pp. 3453–3458, 2003.
- [4] M. Walters, S. Marcos, D. S. Syrdal, and K. Dautenhahn, "An interactive game with a robot: People's perceptions of robot faces and a gesturebased user interface," in *Proc. 6th Int. Conf. Adv. Computer–Human Interactions*, 2013, pp. 123–128.
- [5] O. Brdiczka, M. Langet, J. Maisonnasse, and J. L. Crowley, "Detection human behavior models from multimodal observation in a smart home," *IEEE Trans. Autom. Sci. Eng.*, vol. 6, no. 4, pp. 588–597, Oct. 2009.
- [6] M. A. Cook and J. M. Polgar, *Cook & Hussey's Assistive Technologies: Principles and Practice*, 3rd ed. Maryland Heights, MO, USA: Mosby Elsevier, 2008, pp. 3–33.
- [7] G. R. S. Murthy, and R. S. Jadon, "A review of vision based hand gesture recognition," *Int. J. Inform. Technol. Knowl. Manage.*, vol. 2, no. 2, pp. 405–410, 2009.

BIOGRAPHY



Jayashree.R received her B.E degree in computer science and engineering in Anjalai Ammal mahalingam engineering college, Kovilvenni, Thiruvarur, in 2013, Tamilnadu, India. Now she is doing her master in engineering in St.

Joseph's college of engineering and technology, Thanjavur, Tamilnadu, India.