Dynamic Hand Gesture Recognition Using Kinect Sensor

Devendrakumar H. Pal
Electrical Engineering Department,
College of Engineering Pune,
Pune, India
Email: devendrapal116@gmail.com

S. M. Kakade
Electrical Engineering Department,
College of engineering Pune,
Pune, India
Email: smk.elec@coep.ac.in

***Abstract*—Continuous trajectory path made by hand over a period of time is considered for gesture recognition purpose. Dynamic gesture recognition has always been critical, since conventional method utilize separation of hand from surrounding environment and then finding of palm points make it hard. In this paper real time dynamic hand gesture recognition system using kinect sensor has been proposed. In this the movement of human hand will be captured by kinect sensor and it will be processed in OpenCV. Signal generated by OpenCV are given to Arduino and it will be utilized for controlling DC servo motor action. The computer will generate appropriate signals for controlling the action of robotic arm. The movements of robotic arm links are controlled by DC servo motors. Real-time hand tracking technique is used for object detection in the range of vision.**

***Keywords*—Hidden Markov Model; Kinect Sensor; Hu moments.**

I. Introduction

Human gesture interpretation by a computer is utilized human-computer interaction in the area of computer vision[1]. The motive of gesture recognition system is to recognizing a specific human gesture and conveys information to the user pertaining to individual gesture. A particular gesture of interest can be recognized from collection of gesture, and depending on that, robotic arm is given a particular command for execution [2]. Human body language recognition using computer is over all aim of project, thus making a link between using human-machine interaction. Human machine interaction can be enhanced by hand gesture recognition using without relying on conventional input devices such as keyboards and mouse. Telerobotic control applications utilize extensive use of hand gesture [1][4]. Telerobotic communication can control naturally and intuitively a robotic system. Geometrical information to the robot can be sent in natural way which is dominant benefit of such system. Hand gesture can be utilized to control robotic arm. Extensive research work has been carried out in this field. Hand gesture recognition using computer vision is quite challenging.

For sensing hand movements several methods have been proposed [5][6]. Hand gestures can be very well recognized by glove based technique. Hand and/or arm angles and spatial position can be measured in this method using mechanical glove devices [7]. Although more precession can be attained by this method but it limit the freedom it as user need to wear it while performing gesture. Robot is controlled by data obtained by glove which contains hand position and arm gesture. Proper lightening condition and camera angles are required for detecting hand gesture correctly.

In proposed system, movement of hand is captured using kinect sensor. Various preprocessing algorithm such as skin and motion detection techniques are applied on images to detect hand region. After hand detection, Hu moments are applied to extract feature from images. After feature extraction Hidden markov model is applied due to its ability to overcome the problems of spatio-temporal variabilities. For efficient gesture recognition multiple training sequence are required to train HMM param-

978-1-5090-0467-6/16/$31.00 ©2016 IEEE

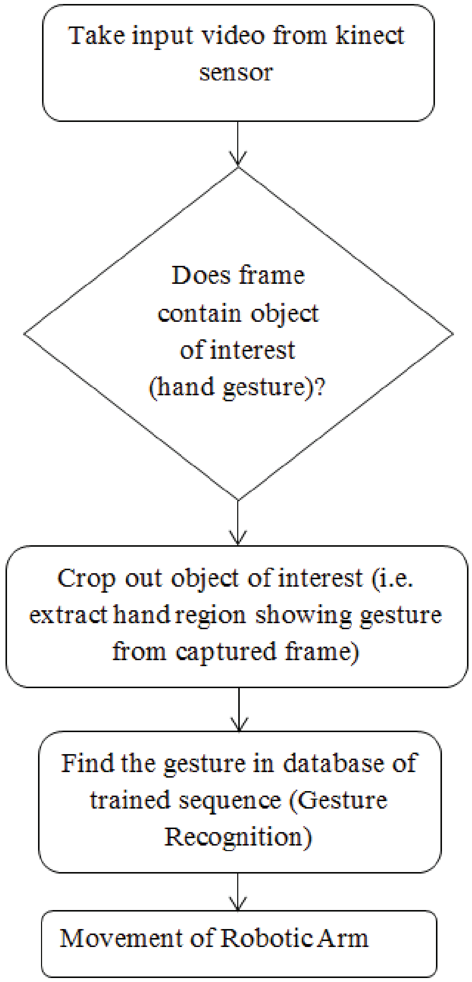


Fig. 1. Flowchart of Robot control Using hand Gesture

eters. These HMM parameters are utilized detecting particular gesture.

An appropriate signal is sent to a robot after hand gesture is identified. Once signal reaches to the robot, it performs a predefined task. A flowchart for overall controlling of a robot is given in Fig 1.

A. Kinect Sensor

Kinect is launched by Microsoft in June 2010 as XBOX360 somatosensory peripheral. The Kinect is a depth camera. Kinect has two cameras and one IR projector. One camera gives RGB image and other camera calculates depth of image. IR projector projects IR laser pattern towards objects, which are bounced back towards IR camera. Kinect takes 30 frame per second for 640x480 resolutions [8].

Using IR camera, depth of each joint is calculated. Algorithm to track the human arm by forward and reverse kinematics is implemented in OpenCV. C++ is utilized as coding language. The coordinates of joints of hand are detected from depth and color images. Depending on a particular hand gesture signal is provided to servo motor.

Fig. 2. Kinect Sensor

B. Robotic Arm

This robotic Arm is a 5 Axis robotic arm with servo gripper. Robot arm has 5 degrees of freedom which includes base rotation, shoulder rotation, elbow rotation, wrist pitch and roll. It uses 9 metal gear servo motors. There are two 15Kg/cm torque servo motors in parallel in shoulder rotation, elbow rotation, wrist pitch for enhancing the performance of the robotic arm. The range of motion per axis is 180 degrees. Robotic arm can lift maximum payload of 50 grams.

II. Hand Tracking and Handshape Extraction

In this paper, single handed postures are considered. A gesture is a specific combination of hand position, orientation, and flexion observation at some time instance. Hand gestures are recognized based on time based sequence of hand. Users hand

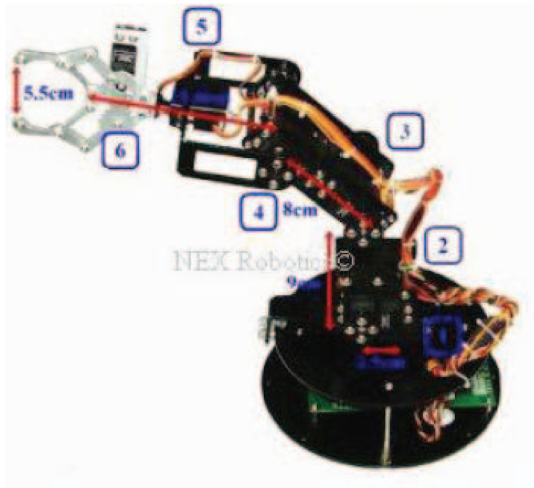


Fig. 3. Robotic Arm

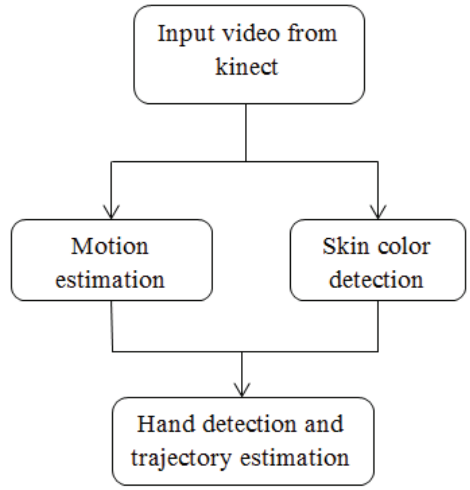


Fig. 4. Hand Detection and Tracing Algorithm

tracking is done in gesture recognition process. Skin color and motion of hand are used for tracking the hand region. Shape and motion information are utilized for recognition of the gesture using Discrete Hidden Markov Model(DHMMs).

A. *Thresholding*

In this system, object localization and extraction information is provided by motion of the object. Moving objects are tracked by observing motion detector when hand is moving. Let $I_i(x, y)$ the ith frame of the scene and $D_i(x, y)$ be the difference image between the ith and the $(i+1)th$ frame. This gives motion information in the image. It is defined as

$$D_i(x, y) = T_i\{|I_i(x, y) - I_{i+1}(x, y)|\} \quad (1)$$

where T_i is a thresholding function, $F_i(x, y)$ and $D_i(x, y)$ are all 640×480 images, and $D_i(x, y)$ is binary image defined as follows[9].

$$D_i(x, y) = \begin{cases} 1, & |F_i(x, y) - F_{i+1}(x, y)| \geq threshold \\ 0 & otherwise \end{cases} \quad (2)$$

B. *Skin color detection*

Skin can be easily detected by using its color information. Images obtained from video are in RGB form which is highly affected by lightening condition. Thus images obtained are converted from RGB color space to $L\alpha\beta$ color space because skin color comprises of red and yellow color and α component comprises of red and green while β component comprises of yellow and blue thus making it most useful color space for skin color $S_i(x, y)$ detection.

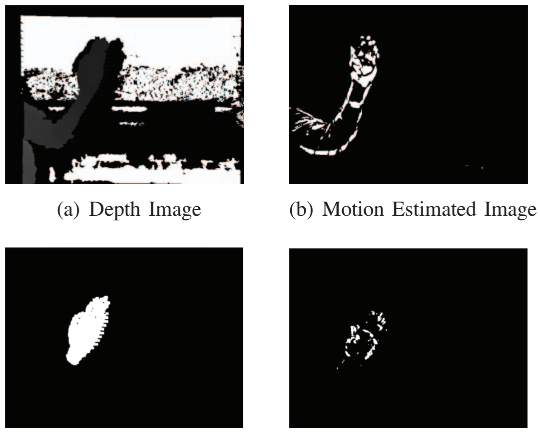
Movement and skin color feature of hand are utilized to detect hand. Logic AND is used to combine these three types of information, that is

$$C_i(x, y) = D_i(x, y) \wedge S_i(x, y) \quad (3)$$

where $D_i(x, y)$ gives information about movement of hand and $S_i(x, y)$ gives information about skin color detection.The combined image $C_i(x, y)$ represents feature extracted for particular gesture. Large region of palm and small part of arm region is represented by the combined image $C_i(x, y)$. Using this motion of hand can be traced out.

III. Feature Extraction

Features are extracted from the image are obtained using Hu moments and orientation an-



(a) Depth Image (b) Motion Estimated Image

(c) Skin Detected Image (d) Skin Detected Image

Fig. 5. Hand Detection

gle. Extracted features are fed to Hidden Markov Model(HMM) for the recognition purpose. For effective recognition, the feature selection process should have follow the given criteria.

- Features should be preferable sovereign on translation, rotation.
- Features should be simply determined.
- Features should be selected such that they do not reproduce each other.

Hu moments are 7 moments derived which rotational and translation invariants. Moments $n_{i,j}$ where $i + j \geq 2$ are given as [3]

$$n_{i,j} = \frac{\mu_{i,j}}{\mu_{00}^{(i+j)/2+1}} \tag{4}$$

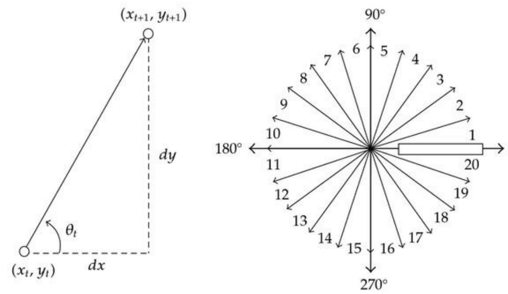


Fig. 6. The orientation and its codewords

Hu moments are

$$\begin{aligned}
\Phi_1 &= n_{0,2} + n_{2,0} \\
\Phi_2 &= (n_{2,0} - 3n_{2,0})^2 + 4n_{1,1} \\
\Phi_3 &= (n_{3,0} - 3n_{1,2})^2 + (3n_{2,1} - n_{0,3})^2 \\
\Phi_4 &= (n_{3,0} + n_{1,2})^2 + (n_{2,1} + n_{0,3})^2 \\
\Phi_5 &= (n_{3,0} - 3n_{1,2})(n_{3,0} + n_{1,2}) \\
&\quad [(n_{3,0} + n_{1,2})^2 - 3(n_{2,1} + n_{0,3})^2] \\
&\quad +(3n_{2,1} - n_{0,3})(n_{2,1} + n_{0,3}) \\
&\quad [(n_{3,0} + n_{1,2})^2 - (n_{2,1} + n_{0,3})^2] \\
\Phi_6 &= (n_{2,0} - n_{0,2})[(n_{3,0} + n_{2,1})^2 - (n_{2,1} + n_{0,3})^2] + \\
&\quad 4n_{1,1}(n_{3,0} - 3n_{1,2})(n_{3,0} + n_{1,2}) \\
\Phi_7 &= (3n_{1,2} - n_{0,3})(n_{3,0} + n_{1,2})[(n_{3,0} \\
&\quad +n_{1,2})^2 - 3(n_{2,1} + n_{0,3})^2] + \\
&\quad (3n_{1,2} - n_{3,0})(n_{2,1} + n_{0,3})[(n_{3,0} \\
&\quad +n_{1,2})^2 - (n_{2,1} + n_{0,3})^2]
\end{aligned}$$

The orientation θ_t is given as

$$\theta_t = tan^{-1}\left(\frac{y_{t+1} - y_t}{x_{t+1} - x_t}\right)$$

IV. Hidden Markov Model

Hidden Markov Model(HMM) is a doubly stochastic model and is appropriate for coping with the stochastic properties in gesture recognition. Hand gestures are transformed into sequential symbols instead of using geometric features. The gestures are used to represent HMM, and the training data is used for learning parameters of HMMs. The gestures can be identified by evaluating the trained HMMs based on most likely criteria. First-order HMM are utilized to encode the dynamic gestures of the system [10][11].

Discrete Hidden Markov Model(DHMM) is defined as follows: λ=(N, M, A, B, n) where

- N represents the number of states in the model. The state set is defined as $S = \{S_1, S_2, ..., S_N\}$.
- M represents the number of distinctive observations in the model. The observation set is defined as $V v_1, v_2, ...v_M$.
- A represents the transition probability matrix, A = a_{ij}, where $a_{ij}$is the transition probability of taking the transition from state i to state j.

$$a_{ij} = P[s_{t+1} = j | s_t = i] \qquad 1 \leq i, j \leq N \tag{5}$$

- B represents the observation symbol probability distribution. B = $b_j(k)$ where $b_j(k)$ is the observation symbol probability and

$$b_j(k) = P[o_t = v_k | s_t = j] 1 \leq k \leq M, 1 \leq j \leq N \tag{6}$$

- Π represents the initial state distribution. $\Pi = \Pi_i$ in which

$$\Pi_i = P[s_1 = i] \qquad 1 \leq i \leq N \tag{7}$$

In practice,to specify complete DHMM, all 5 items are required but two parameters N and M are ignored since they are constants. For a discrete HMM, a_{ij} and $b_j(O_k)$ have the following properties[11].

$$a_{ij} \geq 0, b_j(O_k) \geq 0 \forall i, j, k \tag{8}$$

$$\Sigma_j a_{ij} = 1 \forall i \tag{9}$$

$$\Sigma_j b_j(O_k) = 1 \forall j \tag{10}$$

If the initial state is of distribution $\Pi = \Pi_i$, an HMM can be written in a compact notation

$$\lambda = (A, B, \Pi) \tag{11}$$

HMM is usually applied to solve three kinds of problems:

- Evaluation problem: Given the observation sequence O=$o_1 o_2 o_3 ....o_N$ and the HMM $lambda$, calculates $P(O|\lambda)$. This problem in gesture recognition is mainly used identify given gesture. Forward-Backward algorithm can be solve this problem.
- Decoding problem: Given the observation sequence O=$o_1 o_2 o_3 ....o_N$ and the HMM λ, find out the most likely hidden state sequence. This problem in recognition is mainly used to find the ultimate recognition in implicit sequence. Viterbi algorithm can be utilized to solve this problem.
- Learning problem: Given some training observation sequence O=$o_1 o_2 o_3 ....o_N$ and general structure of HMM, determine HMM parameters λ to make $P(O|\lambda)$ is maximum. This problem is mainly used for gesture recognition in the beginning of training. With help of the BaumWelch algorithm this problem can be solved[13].

V. Results

For training of the HMM parameter (A,B,Π), first, initial values of the matrices are selected according to the conditions given equation (7),(8) and (9). Here 4 and 5 States Left Right Bakis (LRB) and Fully Connected models are trained with Δ = 1. For the above trained model, a test observation sequence is obtain in real time. This test sequence in then used to calculate observation probability for each trained HMM for given numbers. For calculating probability, again forward or backward computation is used. The HMM having highest probability is chosen as a recognized gesture.

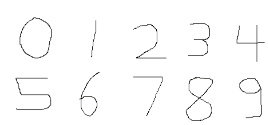


Fig. 7. Defined Gestures

The recognized gesture number is sent to microcontroller after gesture recognition through serial port communication. Depending on gesture number, microcontroller gives PWM pulses to servos of robotic arm. According to PWM pulse width, servo motor moves to particular angle. Thus making different movement for different gestures. Accuracy of system for each gesture for different models is listed below.

TABLE I
GESTURE RECOGNITION RATE WITH DHMM

Gestures Recognition Rate(%)				
Gesture	**Model**			
Gesture	**FL-4S**	**LRB-4S**	**FL-5S**	**LRB-5S**
0	80	87	87	93
1	80	80	87	87
2	87	87	87	93
3	87	87	93	93
4	80	87	87	87
5	80	80	87	93
6	87	93	93	93
7	80	80	87	87
8	93	93	100	100
9	80	87	87	93

FL-Fully Connected, LRB- Left Right bakis, 4S-States, 5S- 5 States

VI. CONCLUSION

A technique for representing, identifying, and understanding human gestures utilizing the HMM(hidden Markov model) is developed. The gestures are transformed into sequential symbols rather than utilizing geometric feature. To illustrate the gestures, and their parameters are understood from the learning data, HMMs are utilized. The gestures identified by estimating the learned HMMs depends on the most possible performance criterion.

To substantiate the usefulness of HMM-based gesture recognition technique, a prototype is developed. Kinect sensor is used as gesture input device and several digits are defined as various gestures. Then HMM is applied to understand and to identify estimated gestures. The suggested technique can be utilized for understanding and identifying gestures in continuous cases as par results shown by experiment. The suggested technique is applicable to single dimensional signal recognition and understanding problems, and will be of significance in designing gesture interface in human-computer interfacing.

In this Hu moments are used to extract feature of image not palm position and its coordinates. The features are given to Discrete hidden marokov model for recognition. From table I it can seen that Left Right Bakis (LRB) has better accuracy compared with fully connected model since there is backward movements gesture. As the no states increases accuracy of system also increases but it will increase the computation required.

REFERENCES

[1] Ng, Chan Wah, and Surendra Ranganath. *"Real-time gesture recognition system and application."* Image and Vision computing 20.13 (2002): 993-1007.

[2] Ramamoorthy, Aditya, et al. *"Recognition of dynamic hand gestures."* Pattern Recognition 36.9 (2003): 2069-2081.

[3] Shrivastava, Rajat. *"A hidden Markov model based dynamic hand gesture recognition system using OpenCV."* Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2013.

[4] Raheja, Jagdish Lal, et al. *"Real-time robotic hand control using hand gestures."* Machine Learning and Computing (ICMLC), 2010 Second International Conference on. IEEE, 2010.

[5] Ren, Zhou, Jingjing Meng, and Junsong Yuan. *"Depth camera based hand gesture recognition and its applications in human-computer-interaction."* Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on. IEEE, 2011.

[6] Wang, Xiaoyan, et al. *"Hidden-markov-models-based dynamic hand gesture recognition."* Mathematical Problems in Engineering 2012 (2012).

[7] Liu, Kui, et al. *"Fusion of inertial and depth sensor data for robust hand gesture recognition."* Sensors Journal, IEEE 14.6 (2014): 1898-1903.

[8] Borenstein, Greg. *Making things see: 3D vision with kinect, processing, Arduino, and MakerBot.* " O'Reilly Media, Inc.", 2012.

[9] Chen, Feng-Sheng, Chih-Ming Fu, and Chung-Lin Huang. *"Hand gesture recognition using a real-time tracking method and hidden Markov models."* Image and vision computing 21.8 (2003): 745-758.

[10] Yang, Zhong, et al. *"Dynamic hand gesture recognition using hidden Markov models."* Computer Science and Education (ICCSE), 2012 7th International Conference on. IEEE, 2012.

[11] Yang, Jie, and Yangsheng Xu. *Hidden markov model for gesture recognition.* No. CMU-RI-TR-94-10. CARNEGIE-MELLON UNIV PITTSBURGH PA ROBOTICS INST, 1994.

[12] Bradski, Gary, and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library.* " O'Reilly Media, Inc.", 2008.

[13] Wang, Youwen, et al. *"Kinect based dynamic hand gesture recognition algorithm research."* Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on. Vol. 1. IEEE, 2012.