International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-2, Issue-1, January 2016
ISSN: 2395-3470
www.ijseas.com

A Survey on Hand Gesture Recognition and Hand Tracking

Arjunlal [1], Minu Lalitha Madhavu [2]
[1] PG scholar, Department of Computer Science And Engineering,
SBCE, Alappuzha, India
[2] Assistant Professor, Department of Computer Science And Engineering,
SBCE, Alappuzha, India

Abstract
Hand gesture recognition is the natural way of human-computer interaction (HCI) and has become one of the key techniques of HCI. Many researchers in academia and industry are devoted in this area. It allows human being to interact with machine conveniently and very easily without wearing any extra hardware device. It has very huge applications in the field of sign language recognition, robot control, intelligent home systems, etc. In this paper we are discussing work done in the area of hand gesture recognition and tracking. Firstly mentioned about the history of hand gestures, then analyses the basic principles regarding it. Then moved to the approaches used for gesture recognition. Vision-based, glove-based and depth-based approaches are mentioned. Different region of interest tracking mechanisms, which are suitable for hand tracking, also discussed in this paper. Then we described about fingertip detection using convex hull.

Keywords: *Human-Computer Interaction (HCI), Hand Gesture, CAMSHIFT, Kalman Filter, Convex-hull, Kinect.*

1. Introduction

Human-computer interaction is the study of how user interacts with computers and computer's developments on successful interaction with human beings. Different techniques are used for the smooth outcome from human computer interaction. Gestures are meaningful human body motions including the movements of head, hands, fingers etc. Interact with the system using hand gesture is more reliable and natural and also has least limitation for users in HCI. There is no need of any special purpose intermediate input devices for the communication between human and computer, when our hand is used as a direct input device of a computing system. By the hand gesture, user can control the computers in natural way. The recognition of hand gestures are classified into two categories[1]. First one is static hand gesture recognition and next is dynamic hand gesture recognition. Static gestures remains unchanged with time and it is called as posture[1], but dynamic gestures changes over a period of time. Static hand gesture recognition method just recognizes pre-defined postures. We have to recognize and understanding a sequence of gestures from the input, so the real time gesture recognition points to dynamic gestures.

Nowadays hand gesture recognition has great significance for human computer interaction in the field of image processing, computer vision and artificial intelligence. Widely used hand gesture recognition techniques are based on glove, vision, and depth. In the case of glove based method, it is quite easier and fast to communicate with the computer due to the presence of special purpose sensor device called data glove. But it is quite expensive and difficult to wear an external glove all the time. In the case of vision based technique, no need of any special purpose devices, except a camera. Hence vision based method leads to natural interaction between the user and computer without any help of additional devices. Depth based recognition technique is much more robust. But the wide usage depth based recognition is limited because of the presence of expensive technologies such as structure light and 3d laser scanning.

2. About Hand Gesture

Hand gesture is a body language with specific meaning, its expressed using palm, finger position and shape. Usually gestures are of two types [1], static and dynamic hand gestures. Static hand gestures denote the individual shape of hand, and the dynamic hand gestures are refer to a series of hand movements. Because the gesture has ambiguity and diversity, and it is affected by the time and the cultural backgrounds, different people have different meanings of gestures. Throughout this paper, hand gesture is defined as the combination of all sorts of

movements and gestures which produced by hand or hand and arm. It includes static hand gestures in which the shape of hand gesture is used to express the meaning, and dynamic hand gestures, whose meanings are based on the track of the motion of hands. The main difference between posture and gesture is that posture has more importance on the shape and state of the hand and the body, but gesture has more emphasis on the hand motions [2].

3. Hand Gesture Recognition Approach

There are many techniques to detect hand from input image. Hand gesture recognition approaches were updated with technology changes. Based on these updates hand gesture recognition methods can be classified into different categories.

3.1 Data Glove Based Approaches

Data glove[3] is a dedicated hardware device which is used to detect hand motion, position and finger bending. The user have to wear this glove like device and it has the ability to collect data from different joints of hand. The collected data are analyzed usually using neural network.

Fig. 1 Data Glove.

This approach can deliver accurate coordinates of hand and finger's position[4][5]. Due to its accuracy, in real time it can analyze and identify a variety of gestures. The main advantage of this method are high accuracy, high speed, less input data and can deliver the information about the fingers movement directly. Even if data glove has its own advantage, it is expensive and not apt for natural HCI.

3.2 Vision Based Approaches

In vision-based gesture recognition (VGR) approach user need not to wear any extra hardware devices. Just a camera is required as capturing device and takes hand as direct input equipment, so no other intermediate media is needed. In this approach camera is used to collect real time gesture image sequence. In the case of HCI using hand gesture, it has different stages. First stage is identifying and distinguishing the hand region from captured frame. Next stage is tracking this distinguished hand region throughout the time.

Hand Segmentation: Hand portion from captured image is identified using skin color. Image from the camera is in RGB format. We need to separate hand region from the complex background. Due to the variety of illumination and skin color, It is little tough to detect skin color in natural environment. So we need to sensibly choose a range of skin color. While comparing YCbCr with RGB, YCbCr[6] is insensitive to color variation. So convert the color space from RGB to YCbCr to get better result.

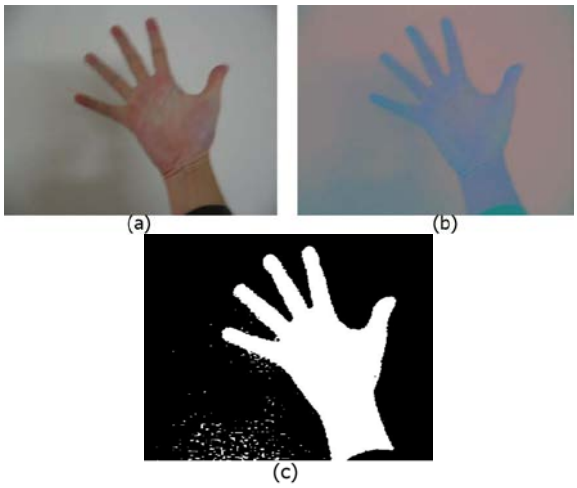

Fig. 2: Convert from original image to binary image. (a) RGB image. (b) YCbCr image. (c) Binary image.

The binary image produced by applying threshold to YCbCr will not provide a good estimate of the hand image due to presence of background noise. We can use image morphology algorithm to eliminate noise by performing image erosion and image dilation[7][8]. Background subtraction using by average filter[9] can also be used to eliminate unwanted noise.

Fig. 3: Result of noise elimination. (a) Before elimination. (b) After elimination

Accurate hand segmentation from the background and the other skin colored body parts of the video are the primary requirement for hand gesture based application. It offers a robust and efficient hand tracking as well as segmentation algorithm which efficiently handles the problems of varying lighting conditions [10].

While hand gesture recognize using skin color, it is not easy to detect only hand region. Face or other skin colored regions also detected as required region. We can eliminate face by masking that region. For this masking process at first need to detect face region from the binary image. Different face detection methods are available. Viloa and Jones [11] proposed a face detection method which is called Viloa-Jones face detector. This method contain three main ideas to get better result, they are: the integral image, classifier learning with AdaBoost, and the attentional cascade structure. Another method to detect the face is feature extraction[12]. The general practice of this method is to collect a large set of face and non face samples, and implement certain machine learning algorithms to learn a face model to perform classification. After getting face region, mask the face region from actual binary image. It will reduce the presence of extra skin portion from processed image. After the masking portion the largest connected segment corresponds to the palm region of the hand. Then the centroid[13] is determined for the palm region. Centroids of segmented region is calculated using moments calculation[13].Han-Sung Park and Kang-Hyun[9] introduce blob labelling method. In this method, for the hand region segmentation the minimum window size was used with a blob labeling. Pixels in all moving area are given a same label by labeling. Information about the minimum size window surrounding separate blobs is obtained. The biggest size of blob is extracted and it holds the hand region.

Contour finding: The contour is useful to identify edge regions, it can be defined as a sequence of points which are the boundary pixels of a region. Skin color areas are represented by white pixels of the binary image. The contour of this skin color area is identified by comparing the length of all the contours, so that we can eliminate those small areas and more focus on fore-arm area. We can use Theo Pavlidis Algorithm to find contours and bring not only lower computation cost but also comes with the contour information we need.

Convex-hull: When we consider a nonempty set of points in a region, the convex hull of that set can be expressed as the smallest convex polygon which includes all the points in the set. We calculate the convexity defect after getting the convex hull of the fore-arm contour. The next element in this convex hull technique is depth point, the point which has the longest distance to the convex hull edge. Rather than arm part, the hand portion has more convex and concave contours and it holds the information we need. Then we find out the convexity defects around the palm area after distinguishing the fore-arm contour and its convex hull. Since the convexity defects are around the palm area, the depth points are found to be on the edge of palm portion. By these depth points, we can determine the position of palm even when the fore-arm is included in the captured frame.

Hand Tracking: After getting palm region, we need to track that hand throughout the time. Different hand tracking algorithms can use for this tracking process. Computer vision algorithms that are proposed to forma part of user interface must be fast and efficient. They must be talented to trace in real time. To trace colored objects in captured video frame sequences the color image data has to be denoted as a probability distribution[13]. The mean shift algorithm[14]is based on a robust nonparametric technique for mounting density gradients to find the mode of probability distributions[15].This algorithm is an effective pattern matching algorithm with no parameter estimation, and can be joined with other algorithms. Color distributions derived from real time

video image sequences change over time, so the mean shift algorithm has to be improved to adapt dynamically to the probability distribution that it is tracking. CAMSHIFT (continuously adaptive mean shift) algorithm is efficient colored object tracker[16].

The idea of camshift is to accomplish all the video frames MeanShift operations[17]. And the outcomes of the prior frame is taken as the initial value of the search window of the next frame's MeanShift algorithm. If this process continues, required tracking can be achieved. When the moving object is blocked by a large area, then we can overcome it by the velocity of moving object is applied linear prediction to kalman filter tracking[17]. Kalman filtering algorithm predict the most likely object location in the current frame according to the results of targets tracking in the previous frame. If there is a target existing in the search area, it will continue to process the next frame. The key procedure of kalman filter is prediction and update.

Mathias Kolsch and Matthew Turk prefer KLT feature for tracking in [22]. It can deliver accurate result on fast moving objects and can compute very effectively[18][19]. Kanade, Lucas, and Tomasi proposed this efficient tracking method. In this method if the feature match relation between two successive frames is below a particular threshold, the feature is considered as "lost". Our ultimate aim is to tack the hand, so the KLT features were initialized on the distinguished palm region and no need of restrictions during subsequent frames in the input video. During the tracking, if frames match quality from one to the next frame was below a particular threshold, they were reinitialized randomly. We can improve the efficiency of this KLT feature tracking by adding flocking behaviour and color cue[22].

Fingertip detection: We have to distinguish the fingertip points from a set of contour points by considering the end points of a sharp shape to be finger tips of our required hand. Fingertips are striking on the contour by calculating the angle of each determined point[7]. Calculate the angle of each point in the contour and a threshold is fixed for the cosine value which gives fingertip position.

3.3 Depth-Based Hand Gesture Recognition

The depth camera based technology reveals a new stage for 3D geometric information acquisition [20]. There are mainly two approaches were used currently in depth camera technology before 2010.The first one is based on the light coding and the second one is based on the time of flight principle. The main advantage of depth image over color image is the ability to reflect 3d feature straight. It would not be influenced by the factors such as shadow, color and illumination. By using different distance information from the depth image, we can separate different parts of the object, even if a covered part is present between two objects. It might not be done with the visible light image [21]. Even if this technology have lot of advantages, at this time, depth camera is too expensive to apply. Hand gesture recognition using Kinect depth data[23] is one of the most popular and accurate method in this area. When compared to the traditional sensor device, it have advantages in the case of resolution and cost.

4. Conclusions

The applications of hand gesture recognition have been spread over from intelligent home systems to medical applications. So there cannot be any compromise in the accuracy of hand gesture recognition and tracking. In this paper we did survey on different approaches in hand gesture recognition and tracking. Among them vision based approach is the natural way of human-computer interaction. But depth-based approach is more accurate than vision based approach. Depth-based approach using Kinect is an important breakthrough in this area. But it is not much comfortable in the natural conditions. Kinect's complete implementation in the natural condition is not far away with our technology growth.

References

[1]Weng C, Li Y, Zhang M, et al. Robust hand posture recognition integrating multi-cue hand tracking[M] Entertainment for Education. Digital Techniques and Systems. Springer Berlin Heidelberg, 2010: 497-505

[2]S. Mitra, T. Acharya. "Gesture Recognition: A Survey", IEEE Transactions on Systems, Man,

and Cybernetics, Part C: Applications and Reviews, pp. 311-324, 2007.

[3] Sturman, David J., and David Zeltzer. "A survey of glove-based input." Computer Graphics and Applications, IEEE 14.1 (1994): 30-39.

[4] Harshith.C, Karthik.R.Shastry, Manoj Ravindran, M.V.V.N.S Srikanth, Navee Lakshmikhanth "Survey on various Gesture Recognition Techniques for Interfacing Machines based on Ambient Intelligence" (IJCSES) Vol.1, No.2, November 2010.

[5] PragatiGarg, Naveen Aggarwal and SanjeevSofat, 2009. Vision Based Hand Gesture Recognition, World Academy of Science, Engineering and Technology 49, pp. 972-977.

[6] Hojoon Park, ''A Method for Controlling Mouse Movement using a Real-Time Camera''

[7] Imran Hussain, Anjan Kumar Talukdar and Kandarpa Kumar Sarma, "Hand Gesture Recognition System with Real-Time Palm Tracking" Annual IEEE India Conference, 2014.

[8] Computer vision based mouse, A. Erdem, E. Yardimci, Y. Atalay, V. Cetin, A. E. Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASS). IEEE International Conference.

[9] Han-Sung Park and Kang-Hyun Jo, "Real-Time Hand Gesture Recognition for Augmented Screen using Average Background and CAMshift", The 19th Korea -Japan Joint Workshop on Frontiers of Computer Vision.

[10] Dharani Mazumdar, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, "Gloved and Free Hand Tracking based Hand Gesture Recognition" , ICETACS 2013.

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. of CVPR, 2001. 1, 3, 4, 5, 7, 8, 11, 13.

[12] Cha Zhang and Zhengyou Zhang,"A Survey of Recent Advances in Face Detection", June 2010, Technical Report, MSR-TR-2010-66.

[13] "A Comparative Analysis of Hand Tracking Algorithms for Gesture Recognition", Ambika. L. A and H. C hidananda, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 5, May-2014.

[14] Sulochana M.Nadgeri, Dr.S.D.Sawarkar and Mr.A.D.Gawande, "Hand Gesture Recognition Using CAMSHIFT Algorithm". Third Intenational Conference onEmerging Trends in Engineering and Technology.

[15]Bradski, Gary R. "Computer vision face tracking for use in a perceptual user interface." (1998).

[16] G. R. Bradski, S. Clara, ―Computer Vison Face Tracking For Use in a Perceptual User Interface‖, Intel Technology JournalQ2„98, 1998.

[17] Huang, Shengluan, and Jingxin Hong. "Moving object tracking system based on camshift and Kalman filter." Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on. IEEE, 2011.

[18] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proc. Imaging Understanding Workshop, pages 121–130, 1991.

[19] F. K. H. Quek. Unencumbered Gestural Interaction. IEEE Multimedia, 4(3):36–47, 1996.

[20] XIANG Xue-qin, PAN Zhi-geng, TONG Jing. " Depth Camera in Computer Vision and Computer Graphics：An Overview", [J].Journal of Frontiers of Computer Science & Technology, pp.5-16,2011.

[21] ZHU Yan-ming,YANG Zhi-bo,YUAN Bo. "Vision Based Hand Gesture Recognition", Published by International Conference on Service Science,2013.

[22] Mathias Kolsch and Matthew Turk, "Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration" [19] Sturman, David J.,

[23] Ni Chen,Qiu Peng et al ."The Research And Application of Kinect Technology in the Field of Human-computer Interaction",Information Technology and Informatization,2013.