ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

Real Time Hand Tracking System

Aisha Meethian[1], Dr. B.M.Imran[2]

M Tech Student, Department of Applied Electronics, Illahia college of Engineering and Technology, Mulavoor, Kerala, India[1]

Professor, Department of P.G, Applied Electronics, Illahia college of Engineering and Technology, Mulavoor, Kerala, India[2]

Abstract: This paper investigates a real time gesture recognition system. Hand tracking is first step for gesture recognition and hand tacking system is based on video acquisition, skin colour segmentation, foreground detection and background elimination. The proposed Gesture Recognition Hand Tracking (GR-HT) system take less training time compared to existing system. Moreover, other feature of GR-HT system is less memory requirement. Disadvantage of GR-HT system is it is affected by environmental condition, to reduce the effects, we use Gaussian Mixture Model (GMM) for background elimination and foreground detection. The experimental result shows that the proposed GR-HT system achieves the results up to 98% in recognition and our work is a novel method for real time gesture recognition applications.

Keywords: Neural network, GMM, Centroid, Euclidean distance, Gesture.

I. Introduction

The tracking systems benefits lie in its speed, its robustness against background noise, and its ability to track objects that undergo arbitrary rotations, motions and fast environmental changes. Hand gesture recognition as human-computer interfaces have the potential to open new realms of applications and functionalities, especially for mobile and worn computing devices. This system open a new Application area for Controlling mouse and/or keyboard functionality, mechanical system, 3D World, Manipulate virtual objects, Human/Robot manipulation and Instruction Communicate at a distance. Recognizing hand motions and configurations by means of computer vision is a particularly promising approach as it allows a maximum of versatility without encumbering the user. In recent years, researches are going on to open more natural, human-centred means of interaction with computers. A particular important direction is that of gesture recognition using hand tracking, where the computer is endowed with perceptive capabilities that allow it to acquire tracking information.

Several hand tracking methods has been proposed recent years. Kolsch and Turk [1] ,proposed new method flocks of features, a fast tracking method for non-rigid and highly articulated objects to detect a hand using image cues obtained from the optical flow and a color probability distribution. The Kanade–Lucas–Tomasi (KLT) feature tracking was employed as it shows excellent performance on quickly moving rigid objects, and can be processed very efficiently. The method benefits lie in its speed, its robustness against background noise, light changing conditions, and its ability to track objects that undergo arbitrary rotations and vast and rapid deformations. The effect of this method is that individual features can latch on to arbitrary artifacts of the object being tracked, such as the fingers of a hand. They can then move independently along with the artifact, without Disturbing most other features and without requiring the explicit updates. Donoser and Bischof [2], proposed a real-time method for tracking hands through image sequences. They applied appearance-based approach to combine a state-of- the-art interest point tracker which efficiently calculated color likelihood. Tracking is based on Maximally Stable Extremal Region (MSER). The MSER algorithm allows to robustly tracking hands through image sequences and additionally provides accurate hand segmentations per frame. The color likelihood map is used to compute ordered set for each pixel. Based on the pixel ordering, MSERs connect regions which can be detected in any image and the hand region can be segmented with high accuracy. After detection the system tracks the hand based on the detected targets in frame. However, the system cannot deal with the nonhand targets. Chen and Georganas [3], proposed gesture recognition using Harr-like features for hand tracking.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

They collected Positive samples (including four different postures) with different scales, and negative samples (only backgrounds) for experiment. Compared with other methods which operate on multiple scales, the approach has high accuracy, which reduces the processing time by using the integral image. However, as described in their paper [3], the positive samples used for training do not contain any complex background, and the results show that the detection is simply on uniform back-ground. Thus, the background subtraction is required before detecting, and the Gaussian filter is employed to reduce noise. Thus, the overall performance of the system highly relies on the subtraction.

In this paper, we focus our attention to sign language recognition of hand gestures. The first part of the paper provides hand tracking recorded by 1 video cameras. The second part of the paper presents gesture Recognition for human interaction Most of the complete hand tracking systems comprises of three layers: detection, tracking and recognition. The detection layer is responsible for defining and extracting visual features attributed to the presence of hands in the field of view of the camera(s). The tracking layer is responsible for performing temporal data association between successive image frames, so that, at each moment in time, the system may be aware of ‘‘ what is where’’ Last, the recognition layer is responsible for grouping the spatiotemporal data extracted in the previous layers and assigning the resulting groups with labels associated to particular classes of gestures.

The paper is organized as follows. Section II, describes about proposed hand tracking system. Section III describes Real-Time hand gesture recognition, section IV describes Results and Analysis, and section V draws the Conclusion.

II. PROPOSED HAND TRACKING SYSTEM

The differences between hand detection and face detection most of the positive samples used in hand detection [14] are with complex background. Most of the positive samples of face detection [15] contain only face area. The positive samples for training contain various backgrounds it considered as a part of object. This result the false detection and thus reduces the tracking rate. Also the appearance of background in the positive training samples can benefit the proposed method to gain the advantage of detecting hand against complex background. This is the main difference to other methods when most of the approaches tried to separate the pure hand region from background, which made their system complex. Thus, the proposed GR-HS system employs the video acquisition, skin colour segmentation, foreground detection and background elimination as organized in Fig. 1 to yield effective tracking performance.

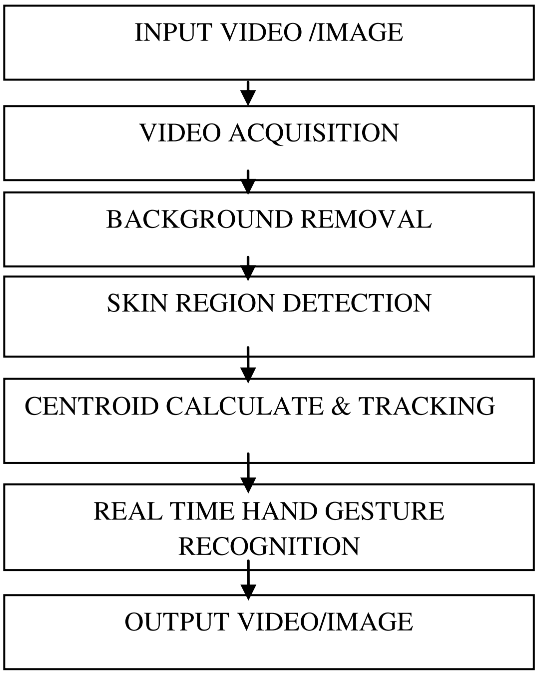

Fig.1. Flow chart of the proposed GR-HS SYSTEM

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

A. Video acquisition

Video acquisition module contains three sub processes. They are Read video, Frame extraction, Pre-processing. The initiation of the acquisition is being done manually. A camera sensor is needed in order to capture the features. Local changes due to noise and digitization errors should not alter the image scene and information. In order to satisfy the memory requirements and the environmental scene conditions, pre-processing of the raw video content is highly important [16]. Various factors like illumination, background, camera parameters, and viewpoint of camera add complexity. These conditions affect images of the object dramatically. The first most step of pre-processing block is filtering a moving average or median filter is used to remove the unwanted noise from the image scenes. Background subtraction is the next step after pre-processing block. Running Gaussian mixture method [17] is used in order to obtain the background subtraction as it is very fast and consumes low memory when compared to other methods.

B. Background Removal

The first phase of the tracking system involves separating hand pixels from nonhand pixels. Before segmentation occurs, we first convolve all captured images with a 5x5 Gaussian filter and then scale this filtered image by one half in each dimension in order to reduce noisy pixel data. All subsequent image processing operations are then performed on this scaled and filtered image. GMM is used for background removal. GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system.

A Gaussian mixture model (GMM) is a weighted sum of M component Gaussian densities given by equation,

$$p(x|\lambda) = \sum_{i=1}^{M} wig(x\,|\,\mu i, \Sigma i)$$

Where x is a D-dimensional continuous-valued data vector, $wi, i =$ 1...M, are the mixture weights, and $g(x\,|\,\mu i, \Sigma i), i = 1$ ....M, are the component Gaussian densities.

Each component density is a D-variate Gaussian function

$$g(x|\mu i, \Sigma i) = \frac{1}{(2\pi)^{d/2}\,|\,\Sigma i\,|\,\frac{1}{2}} exp\{-\frac{1}{2}(-i)'^{\Sigma_i - 1(x-\mu i)}\}$$

With mean vector μ , Covariance matrix ∑ i.

Modelling the values of all the pixels as one particular type of distribution, for this we simply model the values of a particular pixel as a mixture of Gaussians. Based on the persistence and the variance of each of the Gaussians of the mixture, we determine which Gaussians may correspond to background colors Pixel values that do not fit the background distributions are considered foreground until there is a Gaussian that includes them with sufficient, consistent evidence supporting it. Fig 2 shows GMM algorithm used in GR-HT system.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

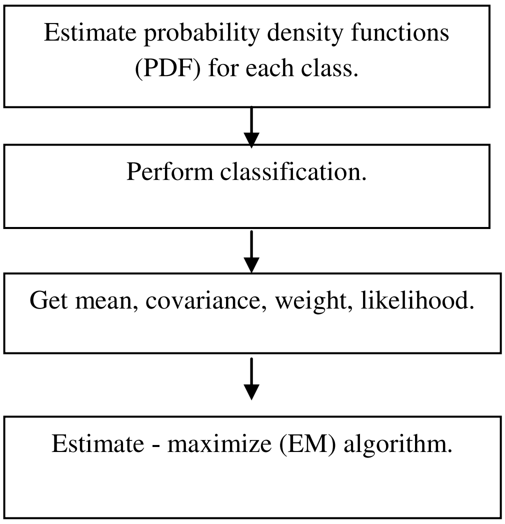

Fig. 2. GMM algorithm

C. Skin region detection

HSV is used to detect skin region. In this stage, a detected hand by the GMM Algorithm is further filtered by the skin color to reduce the false positives. The range of the skin colors heavily rely on the lighting conditions, thus the HSV color model which is robust to lighting Change is adopted for skin color localization. One of advantages of this color model in skin color segmentation is that it allows users to intuitively specify the boundary of the skin color class in terms of the hue and saturation. In this paper, the ranges of the hue and saturation are set in between 0° and 5° and 0.23 to 0.68, respectively, as specified in [18]. Fig. 3 shows one of the segmentation results.

(a) (b)

Fig.3 Skin color segmentation with background. (a) Original image. (b) Segmented result.

D. Centroid calculate & tracking

Euclidean distance is used to calculate the Centroid and to track the hand movement. Hand tracking is an application in field of object tracking. In the proposed system, the hand tracking phase is the second step after hand detection.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

In this paper, the Euclidean distance is chosen to track the hand because of its simplicity and effective result in practice. The Euclidean distance between two points A(x1, y1), B(x2, y2) as shown in Fig. 4(a), is Calculated as follows:

$$Euciliddist(A,B) = \sqrt{(x1-x2)^2 + (y1-y2)^2}$$

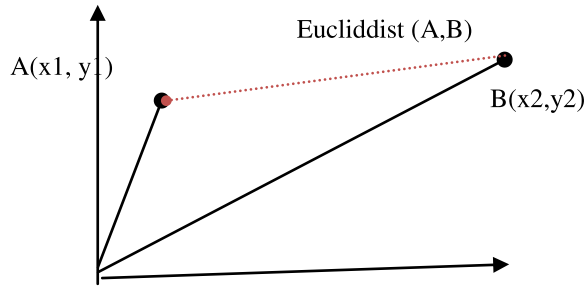

Fig.4. Hand tracking example. (a) Euclidean distance.

III. REAL TIME HAND GESTURE RECOGNITION

The primary step in gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. A large number of methods have been proposed in the literature that utilize a several types of visual features and, in many cases, their combination. In this paper Neural Network is used to recognize hand gesture. Neural Network creates new networks. Basic of Neural Network is Euclidean distance. There are three layers in neural network. They are –Input layer, Inner layer, Output layer. The number of inner layers gives the perfection of the system. Properties of neural network are Training and Testing. In the gesture classification stage, a simple neural network model is developed for the recognition of gestures signs using the features computed from the video captured.

Sign language recognition using neural networks is based on the learning of the gestures using a database set of signs. In the case where the database is small, a sequential exhaustive search can solve the problem [21]. There is necessity of universal database as the applications grow and in that case such sequential search algorithms fail to meet the timing and memory constraints. In order to reduce the size of search space, a new search algorithm called combinational neural networks is being used. This architecture of CNN is based on the cache search memory concept of a CPU.

In the proposed approach we use a new scheme called combinational neural networks [22]. A parallelism exists between the feature extraction layer and the neural network layer as shown in Figure 5. A dual bus is provided in order to allow the flow of data in bidirectional. The feature vector computation involves time and memory.

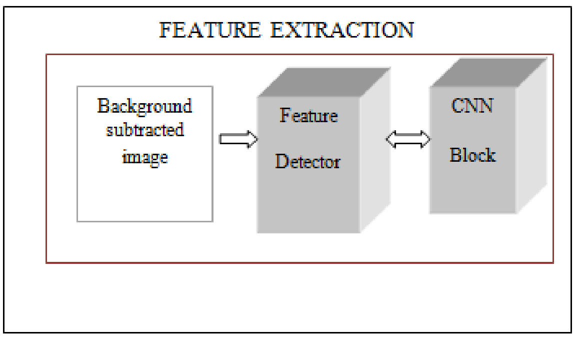

Fig.5. Feature extraction and Combinational Neural Network blocks of GR-HT system.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

A three layer network called back propagation is used to build the CNN. The transfer function used for the back propagation layers is typically a step function, such as the hard-limit transfer function, or a sigmoid function such as log-sigmoid.

$$f(x) = \frac{1}{1 + e^{-x}}$$

The network layer consists of 3 stages: stage 1, stage 2, and stage 3 as in [22]. Each stage acts as a back propagation neural network layer that takes the elements of feature vector as input and outputs the class of object. Each stage receives their input elements from the feature extraction layer.

IV. Experimental Results and Analysis

This paper, is designed to recognise hand detection and hand tracking and also for the hand gesture recognition. In this the American sign language of English alphabets is used in our experiment, and some of the samples are show in figure 6.

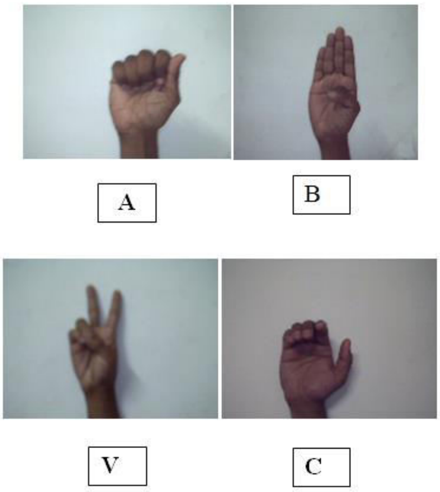

Fig.6 Four hand postures examples.

In this implementation, the Logitech webcam is used for testing in real time, and the captured images are of size 640 × 480 under 30 f/s with natural lighting conditions. The Toshiba laptop with CORE i3 2.4GB and 4GB RAM is employed as the platform.

In our paper, the processing time highly depends on the complexity of the background. The Figure 7 represents the output of the Gesture Recognition Hand Tracking System Performed using the MATLAB.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

Fig .7 .The Gesture Recognition Hand Tracking System.(a) hand tracking system,(b) gesture recognition.

V. CONCLUSIONS

In this paper, a hand tracking system was proposed by using the video acquisition, skin colour segmentation, foreground detection and background. The goal of this paper was to reduce the required tracking time and further reducing the complexity in computation at tracking phase. According to the experimental results, the above tasks were achieved; meanwhile the tracking accuracy was still maintained in high level as that speed which can produce results in real-time manner. In GS -HS system signs for all the alphabets from A to Z are being recognized using the combinational neural networks architecture. The advantage of using the neural networks is high processing speed so that the results in real-time manner. For future extensions processing of words/ sentence and in voice format gestures can be included. In those cases the grammar plays an important role in deciding the efficiency and speed of the system.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

REFERENCES

[1] M. Kolsch and M. Turk, "Fast 2D hand tracking with flocks of features and multi-cue integration," in Proc. CVPRW, vol. 10. Jun. 2004, p. 158.
[2] M. Donoser and H. Bischof, "Real time appearance based hand tracking," in Proc. Int. Conf. Patt. Recog. Dec. 2008, pp. 1–4.
[3] Q. Chen, N. D. Georganas, and E. M. Petriu, "Hand gesture recognition using Haar-like features and a stochastic context-free grammar," IEEE Trans. Instrument. Measurement, vol. 57, no. 8, pp. 1562–1571, Aug.2
[4] K. K. Byong and H. S. Yang, ''Finger mouse and gesture recognition system as a new human computer interface," in Computer & Graphics, Vol. 21, No. 5, PP. 555 - 561, 1997.
[5] Manar Maraqa and Dr. Raed Abu-Zaiter, "Recognition of Arabic sign language (ArSL) using recurrent neural networks" Applications of Digital Information and Web Technologies, pp: 478 – 481, 2008.
[6] Yang quan, "Chinese Sign Language Recognition Based on Video Sequence Appearance Modelling", Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on,pp: 1537 –15
[7] K.Kawahigasi, Y.Shirai, J.Miura and N.Shimada "Automatic Synthesis of training Data for Sign Language Recognition Using HMM", PROC.ICCHP, pp.623-626, 2006.
[8] R Liang and M. Ouhyoung (1998), "Real-time continuous gesture recognition system for sign language", Proc Third. IEEE International Conf: on Automatic Face and Gesture Recognition, pp. 558-567.
[9] Bauer and Heinz "Relevant features for video-based continuous sign language recognition", Automatic Face and Gesture Recognition, 2000. Proceedings, pp: 440 – 445, 2000.
[10] L Rabiner and B. Juang. An introduction to hidden markov models.IEEE ASSP Magazine, pages 4–16, January 1986.
[11] K. Grobel.Videobasierte Geb ardenspracherkennung mit Hidden Markov–Modellen. Ph. D. Thesis, Aachen University of Technology, VDI-Verlag, Dusseldorf, 1999.
[12] Maebatake, Suzuki, Nishida, Horiuchi and Kuroiwa S. "Sign Language Recognition Based on Position and Movement Using Multi-Stream HMM", Universal.
[13] E.-J.Holden, G.Lee and R.Owens "Australian Sign Language Recognition, Machine Vision and Applications", Vol.16, pp. 312- 320, 2005.42.008.
[14] Hand Posture Database [Online]. Available: http://www.idiap.ch/resources/gestures
[15] Face Database [Online]. Available: http://www.face-rec.org/databases.
[16] Brian L. Pulito, Raju Damarla and Sunil Nariani, " 2-D Shift Invariant image Classification Neural Network, which overcomes Stability, Plasticity Dilemma", Vol 2, International Joint Conference on Neural Network, San Deigo, June 17-21,1990.
[17] Jong Bae Kim, Hye Sun Parkman Ho Park and Massimo Piccardi "Background subtraction techniques: a review', Systems, manand Cybernetics, vol.4, 2004, IEEE International conference, pp: 3099-3104.
[18] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: Analysis and comparison," IEEE Trans. Patt.Anal. Mach. Intell., vol. 27, no. 1, pp. 148–154, Jan. 2005.Conference, pp: 3099-3104.
[19] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," ACM Comput. Survey, vol. 38, no. 4, pp. 1–45, Dec. 2006.
[20] Lubin, Jones, and Kornhauser, "Using back-propagation networks to assess several image representation schemes for object recognition", Neural Networks, 1989. IJCNN. International Joint conference, 1989
[21] Tzay Y. Young, King-Sun Fu,"Handbook of Pattern Recognition and Image Processing", Academic Press Inc, 1986.oint Conference, 1989.
[22] P. Mekala, S. Erdogan, Jeffrey Fan, "Automatic object recognition using combinational neural networks in surveillance networks", IEEE 3rd International Conference on Computer and Electrical Engineering (ICCEE'10), Chengdu, China, Vol. 8, pp. 387-391, November 16-18, 2010.
[23] http://www.colourlovers.com/blog/2008/01/22/color-by-hand-theamerican-sign-language-spectrum