



Literature Survey on Hand Gesture Recognition for Video Processing

C. Swapna

Computer Dept, Pune University,
AISSMS COE, Pune, Maharashtra, India

Shabnam S. Shaikh

Assistant Professor, Computer Dept. Pune University,
AISSMS COE, Pune, Maharashtra, India

Abstract— *In this busy world, gestures play an vital role in humans daily life in order to convey data and motions of human being. So gesture Recognition is a part of the Human computer Interaction (HCI). In recent years Human Computer Interaction has become an attractive field. Hardware devices like mouse, keyboard, joystick can be replaced by compatible touch less technologies. Different proposed models, algorithms and methodologies have been proposed to achieve touch less environment with human interaction. Foreground and background segmentation is an important issue in video processing. The task of video processing becomes difficult when there are moving objects and shadows in the video. In this paper we analyze the methods for recognition of hand gesture and different approaches for the segmentation of video processing. The dynamic hand gesture recognition for video processing is reviewed in this paper.*

Keywords— *Video processing, Foreground and Background Segmentation, Hand Gesture Recognition*

I. INTRODUCTION

With the large use of computers Human computer Interaction has become an important part of our daily life. The hardware devices like keyboard, mouse are used for the Human Computer Interaction. In order to use these devices a touch less technology i.e hand gesture recognition is used. Gestures are nothing but the communication of the hand with or without the object. Hand gestures have the applications in various areas like Augmented Reality (AR), Robotics, virtual environment and in electronic devices. So far there are two types of gestures, static gestures and dynamic gestures. The static gestures do not have any kind of motion. The dynamic gestures have the motion of the body parts. The use of external hardware has affected the natural use of motion of hand known as sensor based technology so this technology has restricted to use. To overcome this technology the vision based technology was introduced which take the images of the gestures from the camera to interact with the computer.

Sensor Based Approach: This approach collects the data of gesture performed by using different sensors [2][4][5]. The data is then analyzed and conclusions are drawn in accordance with the recognition model. In case of hand gesture recognition different types of sensors were used and placed on hand, when the hand performs any gesture, the data is recorded and is then further analyzed. The first sensor used was Data gloves then LED's came into existence. The invention of the first data glove was done in 1977[2]. Sensor based approach damages the natural motion of hand because of use of external hardware. The major disadvantage is complex gestures cannot be performed using this method.

Vision based approach: This approach takes image from the camera as data of gesture. The vision based method mainly concentrates on captured image of gesture and extract the main feature and recognizes it. The colour bands were used at the start of vision based approach. The main disadvantage of this method was the standard colour should be used on the finger tips. Then use of bare hands preferred rather than the colour bands. In bare hand recognition various types of skin detection algorithms [7][8][9] are used for recognition of the hand. The dynamic background condition is the main problem faced with this approach. This can be done with the subtraction of background with the bare hands.

Video processing is an integral part of multimedia technology which we are experiencing in today's life. Up gradation of the image in the video, removal of images, blur images, transmitting and compacting the image in video frame can be done in video processing. Extraction and replacement of video frames plays an important role in the video processing. These video frames can consist of constant objects and the moving objects in the foreground and the background of video. The important task in video processing is the detection of the object and segmentation from a stream of video in foreground. Generally foreground objects are extracted by considering the compression of background or by background subtraction. Most of the approaches are proposed for the extraction of foreground object and to fill the gap by the background segmentation when the object is constant in video frames. If the foreground contains a moving object or the shadows then the task of filling the gap becomes difficult in real time applications. The filling of the gap in the foreground can be done by the approach of patch based, region based and pixel based. A region based method is used for the filling the gap for the entire region which is subtracted. Patch region is used to fill the gap and compared patch by patch of the region. Every patch is In pixel based approach, the gap is filled and compared pixel by pixel. Many approaches and methodologies are proposed for this in real time applications. A review of foreground and background segmentation is done in this paper along with the hand gesture recognition.

II. LITERATURE SURVEY

Real time hand tracking and 3D gesture recognition for interactive interfaces using HMM [3]: The authors have introduced a system which identifies the hand gestures and captures the gestures of user having glove with color. For noise elimination this system uses 3D kalman filters and two color cameras for 3D reconstruction. Hidden Markov Model is used for 3D dynamic gesture recognition. This HMM is used for reducing the spatio-temporal variability. The spatio-temporal variability mean the duration and shape of every gesture differs when capturing the image. This system has two parts First is the segmentation .Using HMM it is hard to discover the start and end points so they have reconstructed the model with Baum-Welch algorithm. The second part is the gesture spotting and is implemented by Adaptive Threshold model. The accuracy of 98.75% was achieved with 160 trails of 8 defined gestures.

Model based Segmentation and recognition of dynamic gestures in continuous video stream[1]: In this paper the authors mainly concentrated on the segmentation and recognition of continuous gestures which effect from the spatio-temporal variations. The authors used two types of gestures one is two arm movements for contour extraction, and another one includes a single hand movement hands contour used as the feature vector. They proposed a Multi scale Gesture Model. This model proposes 3 approaches which mainly differs in end point Localization. First proposed approach is used to find the end points with Multi scale search and Motion detection strategy. Second proposed approach is used to locate the end points of the fingers roughly with Dynamic Time wrapping. Third approach is based on Dynamic programming. Using these three approaches the recognition of the hand ranges from the 88% to 96%.The author mentioned that the third approach is the best for the recognition of hand gesture in continuous video streams.

ANN for Gesture Recognition using Accelerometer Data [6]: The authors introduced an Artificial Neural network application used for the classification and gesture recognition. The gesture recognition is done through the wii remote, this remote will rotate in X,Y,Z directions. To reduce the computational cost and memory consumption the gesture recognition is processed in two levels. In first level User Authentication is done for gesture recognition. Accelerometer-Based gesture recognition method is used .In second level without any kind of signal processing for gesture recognition Fuzzy automata algorithm has been proposed. After recognizing the data of the gestures , the data was normalized and filtered by k-means and Fast Fourier transform algorithm. Using this Dynamic Bayesian Network The recognition accuracy has increased up to 95%.

Combining multiple depth-based descriptors for hand gesture recognition[10]: Based on the depth information of the image taken by the depth cameras the authors have introduced a scheme known as novel hand gesture recognition scheme. To properly recognize complex gestures by using 3-D information they used a set of 3-Dimensional features. The proposed hand gesture recognition system consist of three main steps .The first step based on colour and depth information the hand samples are segmented from the background. Wrist samples, palm and the fingers are subparts of the segmented hand samples. The proposed hand gesture recognition consists of four types of features. The second step is to extract these features for the segmentation. The first two set of features are based on the distance from elevation of finger tips to palm center. The third feature set concentrate on computed curvature features of hand contour. The fourth set is on constructed geometry of the palm region. The SVM classifier is used with constructed feature vectors to identify the hand gesture which is performed in front of the camera .The accuracy of 95% is achieved with the combination of feature set and the SVM classifier.

A Real-Time Hand Gesture Recognition System for Daily Information Retrieval from Internet [11]: In this paper the a system is proposed in such a way that with the hand movements the daily information is retrieved from the internet. Principal component analysis is used for the identifying the hand. Using YcbCr color spaces skin color detection and CAMSHIFT algorithm is used to detect and track the hand gestures. The position and the region of the hand is detected from the skin detection. It keeps on detecting the skin region until the condition of tracking trigger is enough. The CAMSHIFT algorithm is used when the tracking trigger condition is enough .Segmentation and normalization is done through the PCA. The experimental proves that the 93.1% of accuracy rate is achieved for hand gesture recognition. For processing a single frame the total time taken was in between 0.1sec to 0.3 sec.

Real-time Dynamic Hand Gesture Recognition using Hidden Markov Models [12]: In this paper a system is proposed to recognize the English numbers from 0-9 with the use of dynamic hand gesture. The system has two steps .First step is pre-processing and the next is classification step. As gestures are of two types Key gestures and Link gestures. The key gestures are used for spotting the link gestures in continuous gestures. The path between the two points of continuous gesture is given for the classification. Discrete Hidden Markov Model is used for the classification. This DHMM is trained by the Baum-Welch algorithm. Average recognition rates using HMM ranges from 93.84% to 97.34%.

Robust Part-Based Hand Gesture Recognition Using Kinect Sensor[2]: Inexpensive depth camera -A Kinect sensor is used to build a robust part based hand gesture recognition, in this paper. As kinect sensors are of low resolution it is hard to identify the hand, but they can capture large objects easily. To deal with the noisy hand gestures which are captured by kinect sensors, the authors are proposed a novel distance metric known as Finger Earth Movers distance. Only the fingers are matched with FEMD but not the whole hand .The noisy hand shapes are managed in a better way, as

FEMD can differentiate the hand gestures with small differences. This system works perfectly and efficiently in uncontrolled environments. The accuracy of 93.2% is achieved with the experimental result.

Video Inpainting on Digitized Vintage Films via Maintaining Spatiotemporal Continuity[13]: The authors have put forward a model in which an algorithm named video inpainting is used for the restoration of damaged old films concentrating on spatio temporal continuity. To find the additional reference data from the whole existing video sequence, the gap are filled in the temporal information. There are two major steps in this video inpainting algorithm. The foremost step is the motion completion, is used to help the inpainting process to get the required reference data by substituting the missing motion information, the next step is the frame completion. It preserves the spatial continuity of the testimonial content before filling the gap area. Before these steps the preprocessing procedure is done by the motion map construct to obtain the motion information of repair part from the source area. This inpainting process recovers the gap of videos by using the real undistractedly data from all video frames whenever possible. With this approach restored videos are nearer to the real state and it reduces the block reuse state. The proposed method has increased the accuracy with the previous existed methods.

A Robust Exemplar-Based Inpainting Algorithm Using Region Segmentation [14]: In this paper the region segmentation is used for the robust exemplar image inpainting proposed algorithm. This algorithm increases the performance of the robust inpainting by using segmentation map. It continuously searches the region of the source to fill the lost or damaged area i.e. target region. Robust priority function is used to decide the size of the patch, and diminish the search region. Segmentation map is used to select repeatedly parameter values for this priority function. Graph based region segmentation algorithm is used to fill the gap by large number of parameter values. Depending on the texture similarity a number of regions of target regions are filled. The borders of adjacent segments are identified as dominant structures to fill the region. Using this system, number of iterations, incorrect matching of source region caused by error propagation is decreased.

Object Removal by Region Based Filling Inpainting [15]: In this paper a new algorithm is introduced for removing big objects from digital images .Best first algorithm is introduced to fill the gap that is left. It is a block based sampling process, rather than a region based process. Block based sampling process is used to fill the large object. The selected large object is removed and filled with background of the original image. This method follows an approach of exemplar based texture synthesis for filling of the target region with a modulated unified scheme. This procedure is able of propagating two- dimensional texture and linear structure into the target region. The results say that the selection of the fill order should be done carefully and is necessary to deal with the task.

High-Quality Real-Time Video Inpainting with PixMix[16]: High excellence real time capable of image capture and video inpainting is presented in this paper. They propose an algorithm known as PixMix which works for the changing of the live video streams, and even when applying conditions in real time. This works for the real diminished reality applications. PixMix algorithms is a pixel based approach rather than a patch based or region based. New tracking approach and frame –frame coherence is combined for the video manipulation in real time. To achieve a high coherence for rotating and translational cameras homographic based approach is used. To give best inpainting results this approach allows stability between the cost and spatial function. This inpainting method uses only the previous and the current frame for real time manipulations.

Video abstraction based on the visual attention model and online clustering [17]: Video abstraction is used to fast browse the video from the big video database and to get the fast access and real content. Based on the Visual attention and clustering a algorithm is proposed. Many ideas are put in this paper. The first idea is to select the key frame; an dissimilar content distance selection principle is used. High level same content is retrieved to represent the main content into video frame, this is done by regions of extract extraction method, non similar content is eliminated for the from important changes. Online clustering is used decrease the same redundancies, time cost which is used for real time video processing. First shot boundaries are detected, then from this boundaries the key frames are taken. To achieve this shot boundaries key frame selection algorithm based on isocontent distance principle is used to get similar key frames from the video. For both effectiveness and performance the summarization algorithms meet the requirements.

Multi-sensor background subtraction by fusing multiple region-based probabilistic classifiers[18]: Based on the fusion of different region based classifiers which process the colour and depth data which are captured by the RGB cameras, a capable background subtraction method is proposed. With a combination of region based foreground prediction which is based on the depth data, and with different depth and colour models i.e. background models the foreground objects are detected based on the scene at pixel and region level. The data collected from these models are used by the mixture of experts to increase the foreground detection accuracy. This proposed model mainly focuses on the foreground and background region based models.

Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers [19]: Foreground/ background segmentation methods are proposed on the basis depth information captured by the kinect sensors without considering the colour data. Based on the combination of classifiers a unique approach is proposed to

increase the accuracy of background subtraction with taking colour and depth information. The false detections are reduced by combining the two classifiers one is of depth data and the other is the colour data. The combination of the classifiers is based on the average weight which allows to modify the classifier importance by considering the edges of colour and depth data. Due to critical Issues, It is able to reduce the false detection. To test the proposed strategy the benchmark database is used.

Table I Comparison of different techniques for video processing

METHOD	ACCURACY	PURPOSE
Video inpainting	Medium	Filling the gap at the foreground
Graph based region segmentation, with priority function	High	Number of iterations of filling the gaps are reduced.
Best first algorithm, Block based algorithm	High	Filling the gap for large objects
Pix Mix algorithm Pixel based approach	High	To change the live video streams, to increase the accuracy in real time applications.
Background subtraction method, region based approach	Medium	Filling the gap by region based
Foreground and background method, classifiers	Medium	To reduce the false detection of the object.

Table II Comparison of different techniques for handgestures

METHOD	ACCURACY	PURPOSE
Hidden Markov Model for data glove.	98.7%	Spatio temporal variability is reduced.
Multiscale Gesture Model.	88%-96%	Segmentation and recognition of the hand.
Accelerometer-Based gesture recognition, k-means and Fast fourier transform algorithm.	Upto 95%	Recognition and normalization and filtering the gestures.
Novel hand gesture recognition scheme, SVM classifier.	95%	3D recognition of hand gestures.
CAMSHIFT algorithm, PCA algorithm.	93.1%	Recognition , Segmentation and normalization of hand gestures.
Discrete Hidden Markov Model.	ranges from 93.84% to 97.34%	Recognition for dynamic hand gesture
Finger Earth Movers distance metric method.	Upto 93.2%	Only fingers of the hand are recognized.

III. CONCLUSIONS

The hand gesture recognition methods discussed in this paper are better in their own way with some advantages and disadvantages. Almost all recognition of hand gestures are similar in outline but their algorithms used are different. Dynamic hand gestures have many constraints such as background change, light conditions, processing speed, and resolution of camera. Due to this constraint any algorithm may not get 100% accuracy. With the discussion in the paper discrete hidden Markov model is better than the other methods for recognition of hand gestures. For the video processing the algorithm which is based on the pixel based method is better than the patch based and region based method. The region based method is used generally for filling the large objects. By pixel based approach even a small portion of the object can be detected or filled by the other object in the target region So pixel based approach is better than the other methods.

ACKNOWLEDGMENT

I am Thankful to my guide Prof. Shabnam S Shaikh , Asst Professor, Computer Department, for helping me in preparation of this paper and her valuable suggestions and encouragement.

REFERENCES

- [1] Hong Li, Michael Greenspan. "Model based Segmentation and recognition of dynamic gestures in continuous video streams". Springer publications in Pattern recognition,44(2011)1614-1628.
- [2] Zhou Ren, Junsong Yuan, JingjingMeng, Zhengyou Zhang. "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor". IEEE Transactions on Multimedia, Vol.15, No.5, August 2013.
- [3] C. Keskin, A. Erkan, L. Akarun ."Real time hand tracking and 3D gesture recognition for interactive interfaces using HMM". IEEE Transaction.
- [4] Popa, M. "Hand Gesture recognition based on Accelerometer Sensors". IEEE Networked Computing and Advanced Information Management(NCM),2011.
- [5] Thomas Sholmer, Benjamin Poppinga, NielsHenze, Susanne Bol. "Gesture Recognition with a Wii controller". In TEI'08 Proceedings of the 2nd international conference on Tangible and embedded interaction Pages 11-14.
- [6] Blanca Miriam Lee-Cosioa, Carlos Delgado-Mataa, Jesus Ibanezb. "ANN for Gesture Recognition using Accelerometer Data". Elsevier Publications, Procedia Technology 3 (2012) 109 – 120.
- [7] Fabio Dominio, Mauro Donadeo, Pietro Zanuttigh. "Combining multiple depth-based descriptors for hand gesture recognition". Elsevier Publications, Pattern recognitions 2013.
- [8] Guoliang Yang, Huan Li, Li Zhang, Yue Cao. " Research on a Skin Colour Detection Algorithm Based on Self-Adaptive Skin Colour Model". IEEE Communications and Intelligence Information Security(ICCIIS),2010.
- [9] Lei Yang, Hui Li, Xiaoyu Wu, Dewei Zhao, Jun Zhai. " An algorithm of skin detection based on texture". IEEE Image and Signal Processing(CSIP),2011.
- [10] Fabio Dominio, Mauro Donadeo, Pietro Zanuttigh," Combining multiple depth-based descriptors for hand gesture recognition", Elsevier, Pattern recognition Letters 2013.
- [11] Sheng-Yu Peng, Kanoksak Wattanachote, Hwei-Jen Lin, Kuan-Ching Li, "A Real-Time Hand Gesture Recognition System for Daily Information Retrieval from Internet", IEEE Fourth International Conference on Ubi-Media Computing, 978-0-7695-4493-9/11 © 2011.
- [12] M.M.Gharasue, H.Seyedarabi, "Real-time Dynamic Hand Gesture Recognition using Hidden Markov Models" , 8th Iranian Conference on Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP), 978-1-4673-6184-2/13/\$31.00 ©2013 IEEE.
- [13] Nick C. Tang, Chiou-Ting Hsu, Chih-Wen Su, Timothy K. Shih, and Hong-Yuan Mark Liao. "Video Inpainting on Digitized Vintage Films via Maintaining Spatiotemporal Continuity" published at IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 4, AUGUST 2011.
- [14] Jino Lee, Dong-Kyu Lee, and Rae-Hong Park, "A Robust Exemplar-Based Inpainting Algorithm Using Region Segmentation". IEEE Transactions on Consumer Electronics, Vol. 58, No. 2, May 2011.
- [15] N.Neelima , M.Arulvan, B. S Abdur, "Object Removal by Region Based Filling Inpainting" . IEEE transaction 978-1-4673-5301/2013 IEEE.
- [16] Jan Herling, and Wolfgang Broll," High-Quality Real-Time Video Inpainting with PixMix" , IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 6, JUNE 2014
- [17] Qing-Ge Ji , Zhi-DangFang , Zhen-HuaXie , Zhe-MingLu ," Video abstraction based on the visual attention model and online clustering" .Published at Elsevier, Signal Processing: Image Communication 28 (2013)241–253 in 2013.
- [18] Massimo Camplani , Carlos R. del Blanco , Luis Salgado , Fernando Jaureguizar , Narciso García, "Multi-sensor background subtraction by fusing multiple region-based probabilistic classifiers", published at Elsevier Pattern Recognition Letters 2013.
- [19] Massimo Camplani , Luis Salgado, "Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers". published at Elsevier J. Vis. Commun. Image R. 25 (2014) 122–136.