International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Video Inpainting Through Dynamic Hand Gesture Recognition

C.Swapna[1], S. Shabnam Shaikh[2]

[1]Computer Department, AISSMS, Collage of Engineering, Kennedy Road, Pune, India

[2] Professor, Computer Department, AISSMS, Collage of Engineering, Kennedy Road, Pune, India

Abstract: ***This paper aims at applying the hand gestures which are recognized by the camera for the removing of the object and filling that object with other object for the frames of video . The video inpainting is an important technique in video processing , applying hand gesture to this technique is important. As the world is moving towards the touchless technology, Human computer Interaction with gesture recognition for mouse control is applied to the video inpainting. A Finite state machine is used for the hand gesture recognition. For the extraction and replacement of the object the PIX-MIX algorithm is used. To improve the outline image quality of the frame this algorithm compares pixel to pixel rather than the patch to patch or region to region. The extraction and replacement of the object is carried out through dynamic hand gesture.***

Keywords: Hand gesture recognition, Video inpainting, Human computer Interaction, segmentation.

1. Introduction

The hardware parts of the computer such as the mouse, keyboard are restraining the creativity and capabilities of the end user for the human computer Interaction(HCI).For rapid development of software and hardware some new types of HCI techniques are required. This paper focuses on proposing the real time system able to understand the hand gestures which are captured by the webcam and used to communicate with the computers. In this paper a machine user interface is developed which implements gesture recognition for the Human computer Interaction. These type of systems have the applications in medical, robotics, healthcare, industries, multimedia applications, graphics.

Hand gestures are communication or physical behavior of the hand with or without object. These hand gestures can be static gestures or dynamic gestures. The static gestures depend on the constant object i.e., not on the moving object. The dynamic hand gestures depend on the moving object and they need the motion of the object. In dynamic hand gestures the different types of sensors[2] [3] such as the data gloves, LED's and markers were used to find the gestures These limit the natural motion of the hand and the complex gestures were hard to recognize. So vision based approach was used. In this method the bare hands were used without using any external resource. In this paper the bare hands are used for the recognition of the hand gestures

Dynamic hand gestures are defined as a series of hand gestures and hand trajectory. The hand trajectory is calculated by tracking the movements of the hand and this includes various skin detection and background subtraction methods. Recognition of hand gesture is the fundamental challenge that has to be solved while constructing the virtual mouse system. Different color models like RGB, HSV, YCbCr[2], are studied. The proposed system is a real time application which effect from the luminance and chrominance of light, and is normalized RGB color space the YCbCr color model is used to detect the skin region of the hand. For the gesture recognition the finite state machine is used as it is easy to handle the ordered sequence of states. These dynamic gestures are applied for the video inpainting technique of video processing.

The gestures are used for the operations like the selected object extraction, filling the gap with the background and replacing with the other object in a frame of video. After extracting the foreground region of the object the gap filling is important. For filling the gap the approaches like the pixel based method, patch based method and region based method have been studied. This paper uses the pixel based approach as it compares pixel to pixel and give the high accuracy with PIX-MIX algorithm.

The paper is organized as follows section 2 gives the review of related work on hand gesture and video inpainting section 3 gives the detailed description of our approach section 4 gives the results and finally the conclusion and future scope in section 5.

2. Related Work

For the hand detection the sensor based approach[2] [3] with different LED's, Colored gloves , and markers were introduced later the vision based approach with bare hands[11] were used. Dynamic Bayesian Network of Artificial neural network application for the recognition of hand gestures were used in the paper [9]. The Wii remote camera is used this camera rotate in all directions for gesture recognition. After gestures are recognized the K-means and Fast Fourier transform algorithm was used to normalize and filter the data and to lower the cost and increase the performance. Color gloves with Hidden markov model is used for the recognition of the gestures of the hand in the paper [8]. The authors proposed Baum-Welch algorithm with HMM as it was difficult to identify the start and end points only by using the HMM. A Kinect Sensor was used to detect even a small color and depth data information in the paper [10]. The authors used the black colored gloves to detect the

finger tips after detecting the hand. Fingers Earth Movers algorithm was introduced by the authors which is used which discards the region of hand by selecting only the finger tips. In the paper[11] hand gestures were used for information retrieval of the electronic devices. YCbCr color model is used for the skin detection, CAMSHIFT algorithm was used for the recognition and tracking of the hand. Principal Component Analysis was used for gesture recognition. Kinect sensor were used in the paper[12] for the Zhou Ren et al [10] used a Kinect Sensor to capture even a small color and depth data. Hand is detected by the camera then the black color gloves are used to find the finger tips. The authors introduced Fingers Earth Movers algorithm which is used to track only the finger points discarding the whole hand. In the paper [11] the authors proposed a model such a way that hand gestures are used for information retrieval from electronic device. For skin detection the authors used the YCbCr color model, and for detecting and tracking the hand CAMSHIFT algorithm is used. Principal Component Analysis is used for gesture recognition. Kinect sensor[12] is used for the depth information by applying the Principal Component analysis(PCA) hand samples of palm, arm and wrist are segmented from the background. Then to recognize the hand gestures Multi class Static Vector Machine (SVM) is used. In the present paper for skin detection the YCbCr color model with local adaptive technique is used to get the range of skin color of the hand and Finite state machine is used for the recognition of real time hand gestures.

The video in-painting algorithm can be of pixel approach, region approach or patch approach. The study of the papers say many of them depend on the region based approach. In the paper [13] the region segmentation is used for proposed Robust exemplar based image in-painting algorithm. To increase the performance robust inpainting is used with the segmentation map. In [6] Block based sampling process is used . Best first algorithm was used to fill the gap after extracting the object from the digital image frames. In paper [14] the authors used Kinect sensors to take the information of the depth data. To fill the gap from the extracted region the combination of classifiers were proposed with average weight of depth data and color data. In [15] region segmentation is used background subtraction method was introduced. Region based foreground prediction method with a combination of background segmentation was used for the filling the gap. The foreground object is detected by the background method depending on the scene at pixel at region level. In the paper Nick C. Tang et al [16] introduced a model by taking into consideration a spatio-temporal continuity for the restoration of old images and of old videos. From the sequence of neighboring video frames the temporal information of the gaps are filled for the extra reference. In region based approach the entire block of image frame is compared with the other block of image so accuracy was less. In patch based approach the image is divided into patches and then compared with the other patches of the image so it is hard to paint the edges of the object so the gap cannot be painted properly so in the present paper a pixel based approach is used with PIX-MIX algorithm is used

3. Proposed Architecture

The Architecture of the proposed system mainly consist of three modules Hand gesture recognition, video as the input, video in painting

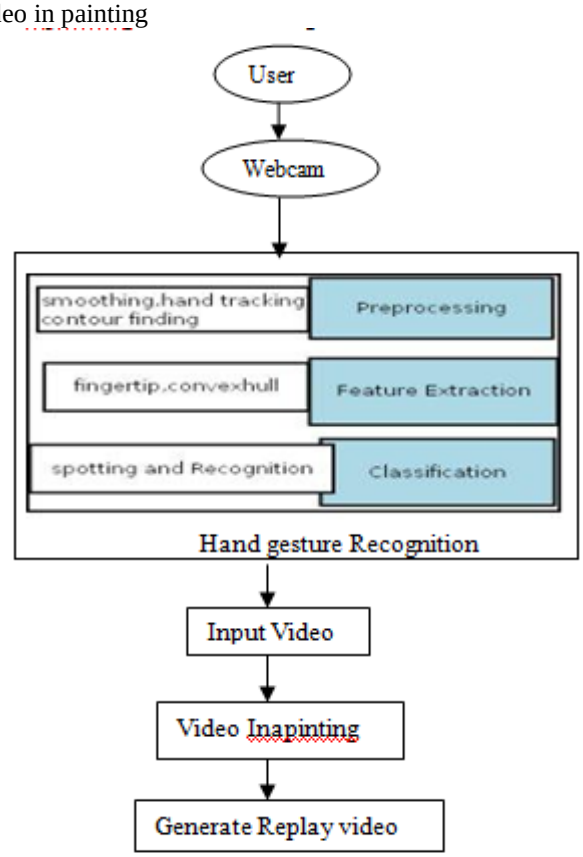

Figure 1: Architecture of Proposed System

3.1 Preprocessing

The hand of the end user should be captured as the frames of video by the webcam. For the detection of hand the skin detection method is used with YCbCr color model. As YCbCr color model is a normalized form of RGB and it works prominent for the luminance and chrominance of light as our system works with real time hand capturing. For the skin detection we consider seven patches of the hand to get the range of the skin color i.e., local adaptive technique is used for the detection of skin color. Then this frames of video of hand is used further for preprocessing.

In preprocessing these skin color are converted to black and white image i.e., binary image the skin pixels are set to 1 as white image and other pixels are set to 0 as black image. These frames of video are set to morphological operations of XOR operation and converted to single image. While detecting the hand and in converting to the threshold image the noise may occur in binary image. Median filter is used to remove the noise from the binary image.

After getting the binary image the contour is formed for the hand by detecting the strong points. Suzuki's algorithm[17] is used for detecting the strongest points of the hand.

3.2 Feature Extraction

After contour formation the fingertips are identified by considering the X , Y positions of the hand. The highest X, Y position is the middle finger tip. Depending upon the angle and the distance between the fingers the other fingertips are calculated. To extract the good features or patterns these contour image and the fingertips which are recognized are calculated by the convex hull [17] identify the concave points after obtaining the hull to detect the exact fingers.

3.3 Gesture Recognition

After getting the convex hull and detecting the concave points the features extracted should be handled by the events. For the recognition of the gestures and for the handling of the events Finite state machine[1] was used.

To have the region of the hand the bounding box and total line length is calculated. The four mouse events are handled through the five recognized gestures. By determining the counter as numbers to the events these four mouse events are handled. For the mouse move we set the counter as 4. If any of the four fingers are up then the mouse move event is activated. If the counter is 2, two fingers are up then the right click is activated. In the same way for the left click counter value one is activated and if counter value is three then double click is activated. Again in the mouse move the ordered sequence of states should be handled to move mouse move right, mouse move left, mouse move up, mouse move down. For the mouse move left a counter of 1 is set in the mouse move operation.

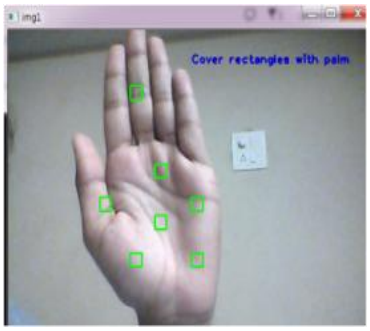

Figure 2: Image of Hand Detection

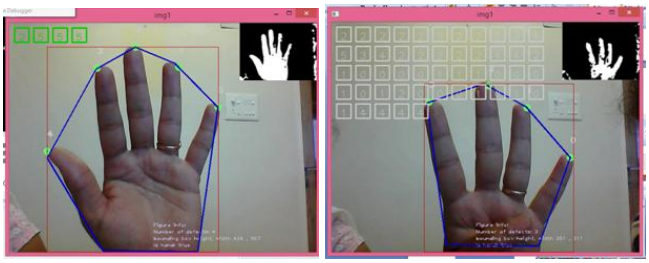

Figure 3: Results For the Hand Gesture for Reset and Mouse Move

Figure 2 and 3 shows the hand gesture results obtained using the proposed system. In Fig 2 the seven patches of the hand takes the range of skin color . In Fig 3 the binary image is shown and the events are handled for the "Reset " and "Mouse move". The contour formed with the blue lines is shown and the green dots shows the finger tips formed by the convex hull. The bounded box area of hand region is shown with red box.

3.4 Input Video

After handling the events through the hand gestures these events are applied for the extraction and replacement of the object in a frame of video so the input video is given for the present system. The system can take any type such as .JPEG, .MPG or .avi of the video file. Then the input video is extracted into frames to do the operations of the video inpainting.

3.5 Video Inpainting

The video inpainting is the most enhancement technique in the video processing. The proposed system is able to select the constant object and extract it from the current frame and track in the remaining frames and inpaint that object with the background. Using the key frame as a reference model, the system is able to handle the static elements. The main steps in the Video inpainting technique in this system are 1. Object selection 2. Object tracking 3. Video pipelining.

3.5.1 Object selection

For a selection of the object the frame should be selected from a video through the hand gesture events which are detected. Many of the methods for the selection of the object include painting based technique[18] gives the results slower for video streams for segmentation. The combination of mean-shift and graph is the other method for the segmentation. In the current system for selecting the object need to form the contour i.e., shape of the object , a contour algorithm [19] is used. The fingerprint application of [20] is applied to find the binary mask. This binary mask is formed by considering all the pixels of the object. After selecting the object by forming the contour, the clustering technique with Nearest Neighbor Field(NNF) is used to find the exact boundaries of the object, and to match the nearest pixel with highest probability. Many methods of clustering techniques have been studied like k-means and hierarchical in this the main drawback is the number of clusters should be defined at the initial stage of the clustering algorithm. The segmentation method along with the multi resolution approach[20] is used to fine the object borders. This approach begin from the core layer and forward to the next layer. At this stage of finer layer the object border pixels are investigated.

3.4.2. Tracking the object

The median filter is used to remove the noise and small or tiny gaps . The calculated object contour should be tracked in neighboring frames of the video after selecting the object. A contour tracking approach[20] of two phase is used in this system. In the first phase homography based contour tracking approach is used while in second phase a counter is refined and adjusted with undesired object area. For the homography determination[19] the strongest contour points are tracked in two frames by pyramid base motion detection. The homography is calculated on the strongest points between the contour point whereas the mask is based on the object contour.

3.4.3 Video Pipelining

A contour is formed through the selection of object and the mask synthesize is invoked. The object of the frame I is a combination of the source region (S) and the target region (T) and is given by the equation

$$I=TUS \quad (1)$$

All pixels defined in source region are replaced from the target region. F denote the mapping between the target and the source region and is given by the equation

$$F=T\rightarrow S \quad (2)$$

F has been determined by the equation 2. To create the final image the target pixels are replaced by the source pixels. The key frame K_f is stored as the synthesize frame. A reference model R_f is defined by calculated homogrophy with binary synthesis Mask M_n for each new frame F_n. The desired pixels $(M_n(p)=1)$ the current frame pixels are copied for undesired pixels $(M_n(p)=0)$ and are replaced by the key frame F K_{kf} interpolation information. The method of mapping forwarding by homography H_k [20] for the reference model is given by the equation 3

$$R_f(P) = \begin{cases} F_n(P), & M_n(P)=1 \\ K_f(H_k.P), & M_n(P)=0 \end{cases} \quad (3)$$

To find a transformation for the new frame the H_k is used to find the number of iterations. After filling the gap with the source region the new object is replaced. The selection of object and replacement of object is done through the real time hand gestures with which events are handled.

4. Results

The results of hand detection compared with the other papers are shown in table 1.

Table 1: Comparision of Hand Detection Accuracy

Sr no	Algorithm used in Papers	Static/dynamic	Accuracy
1	PCA[11]	dynamic	93.1%
2	Finger Earth Movers classification[10]	Dynamic/ black belt	93.2%
3	HMM[8]	dynamic	93.84%
4	SVM [12]	dynamic	92.5%
5	FSM	dynamic	94-96%

For the hand gesture detection , each event is tested by the five persons for the 10 times and the accuracy for the each event is shown in table 2. and the Overall percentage for the events handled ranges from 94% to 96%.

Table 2: Results for the Hand gesture Recognition

Events Detected	No .of Persons	Input Images	Output Images	% of accuracy
Left click	3	50	47	94%
Right click	3	50	46	92%
Mouse move	3	50	47	94%
Reset	3	50	48	96%

The results for video Inpainting through this events are shown in the figure 4

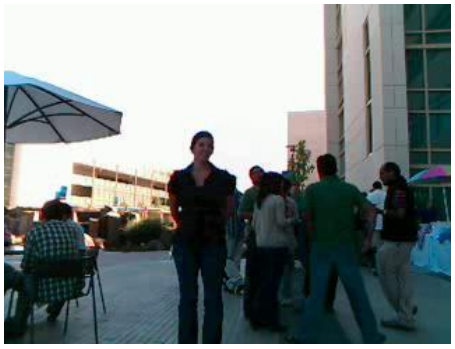

Figure 4.1: Frame before Inpainting

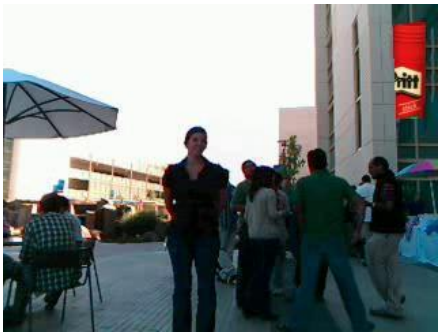

Figure 4.2: Frame after Inpainting

The fig 4.1.shows the frame of video in this the half of the mirror is replaced with the other object of red color as it is shown in the fig 4.2 i.e., after replacement of object.

The input videos are taken from different locations for testing the accuracy ranges from 92% to 95% the time required for the larger number of frames of video is greater compared to less frames of video

5. Conclusion

Dynamic hand gesture recognition system is proposed whose aim is to identify human generated gestures for device control and for video inpainting technique. Gestures are identifies and recognized by the Finite State Machine. . We use a Pix-Mix algorithm for the video inpainting and is a pixel based approach. This system can be used in multimedia applications, robot applications, and healthcare e.t.c very useful for the physically disabled users who lack the strength and precision used to operate the traditional input devices. This paper works well with minimal hardware requirements but under some constraints of light and background when detecting the hand through the camera.

References

[1] Alper Aksa Öztürk, and Tansel Özyer, "Real-time Multi-Objective Hand Posture/Gesture Recognition by Using Distance Classifiers and Finite State Machine for Virtual Mouse Operations", ieee Electrical and Electronics Engineering (ELECO) , 2011 7thInternational conference on 1-4 Dec 2011.

[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 2011.
[3] Popa, M. "Hand Gesture recognition based on Accelerometer Sensors". IEEE Networked Computing and Advanced Information Management(NCM),2011.
[4] 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.
[5] 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.
[6] N.Neelima , M.Arulvan, B. S Abdur, "Object Removal by Region Based Filling Inpainting" . IEEE transaction 978-1-4673-5301/2OI3 IEEE.
[7] Jan Herling, and Wolfgang Broll," High-Quality Real-Time Video Inpainting with PixMix" ,IEEE Transcations On Visualization and computer graphics , Vol. 20, No. 6, June 2014.
[8] M.M Gharasuie, h sayedrabi ."A Real time hand gesture recognition using HMM ".IEEE Transaction- 2013 .
[9] Blanca Miriam Lee-Cosioa, Carlos Delgado-Mataa, Jesus Ibanezb. "ANN for Gesture Recognition using Accelerometer Data". Elsevier Publications, Procedia Technology 3 (2012).
[10] 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.
[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] Fabio Dominio, Mauro Donadeo, Pietro Zanuttigh," Combining multiple depth-based descriptors for hand gesture recognition", Elsevier, Pattern recognition Letters 2013.
[13] 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.
[14] 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.
[15] 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.
[16] 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.
[17] Ogata.K Futatsugi.k, "Analysis of the Suzuki-Kasami algorithm with SAL model checkers", Software Engineering Conference 12th Asia Pacific. Published in 2005.
[18] R. Tong, Y. Zhang, and M. Ding, "Video Brush: A Novel Interface for Efficient Video Cutout," Computer Graphics Forum, vol. 30, no. 7, pp. 2049-2057, 2011.
[19] J. Herling and W. Broll, "Advanced Self-Contained Object Removal for Realizing Real-Time Diminished Reality in Unconstrained Environments," Proc. IEEE Ninth Int'l Symp. Mixed and Augmented Reality (ISMAR '10), pp. 207-212, Oct. 2010.
[20] Jan Herling, and Wolfgang Broll," High-Quality Real-Time Video Inpainting with PixMix" , IEEE Transaction On Visualisation of Computer Graphics, Vol. 20, No. 6, June 2014.
[21] R. Szeliski, "Computer Vision: Algorithms and Applications". Springer,2010.