I J C T A, 8(2) December 2015, pp. 811-816

Training and Testing of Dynamic Gesture Patterns

Salil Batra*, Aman Singh* and Divya Anand*

Abstract: In this paper a gesture based system is presented which works on the Dynamic gesture testing and training technique. A real time gesture input will be provided and system will detect and recognize the input. Searching will be performed in the current database, and if the gesture pattern is found, then corresponding action will be performed. If gesture pattern is not available then system will respond with informative message such as "New gesture pattern tracked, continue with the training". User can continue with the training and new course of action will be given by the user. Also, the entry for this new gesture pattern will be done in the database. Next time, if user met the same gesture pattern, the system will respond by performing the action trained by the user earlier. Results reveal that the system works efficiently in the robust conditions and it is applicable for both static and dynamic gestures.

Keywords: Marker based algorithms, ROI, Segments, Convex Hull, Convexity Defects, Watershed algorithm

I. INTRODUCTION

Human Computer Interaction provides the comprehensive study of various techniques available for establishing communication with computers. With the passage of time various new and unique ways are discovered for interaction. The journey of man-machine interaction is very vast. It has gone through many phases. Starting from normal interaction methods to advanced intelligent techniques, the current methods are equipped with new things.

There are different kinds of devices which are available for establishing communication between human and computer. These devices can vary depending upon their usage and application area. Devices such as mouse, keyboard, joystick, light pen etc are there for establishing the communication. But these are the conventional sources for giving input to the computer. Some more sophisticated and robust methods are also available to provide input to the system. For fulfilling the need of current time, HCI systems are divided into several categories, such as: 1) Visual-Based 2) Audio-Based 3) Sensor-Based [2].

Visual-Based HCI systems works on the visual inputs given by the humans and the research on these systems is broadly classified into several categories, such as: a) Emotion recognition (Facial Expression analysis) b) Gesture Recognition c) Eyes Movement Tracking. Such systems take image/video as a source of input and further detection and recognition is performed. The input for emotions generally depends upon the human behaviour and accordingly facial inputs are tracked [3]. Gesture recognition works in two steps, i.e. Gesture detection and Gesture recognition. The basic gesture that is used is hand gesture which is tracked through various mechanisms and functionality is achieved accordingly. Eyes movement tracking is based on the relative eye motion and difference in the initial and final eye position will be taken as the notable gesture.

Audio-Based HCI system works on the inputs taken through voice and speech [5]. Voice recognition works on the intensity and acoustic levels of a voice. Speech recognition works on the different kinds of dialects, words, phrases and language which are used as parameters for differentiating the inputs. These

* Department of Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab, India, 144411, *E-mail:* [1]*salilbatra40@gmail.com;* [2]*amansingh.x@gmail.com;* [3]*divyaanand.y@gmail.com*

types of systems are also fulfilling the demands as unique and novel inputs are required for Human Computer Interaction.

II. LITERATURE REVIEW

Constante [1] proposed a system which targets to reduce the problems generated using robust gesture recognition. Paper demonstrates the usage of SSM descriptors in order to deal with the problems raised due to view-invariance.

Lian and Lin [2] presented a gesture recognition system with very fast and accurate gesture recognition model. The recognition model is built using Hierarchical Hidden Markov Model (HHMM).The system recognizes Arabic numerals using the IMU sensor. Instead of recognizing the Arabic numerals directly the gesture units are divided into arcs which can be drawn clockwise and anticlockwise directions. These arcs are supported with the horizontal and vertical lines. The purpose of partitioning these numerals is to get the more optimized results.

Xu and Zhou [3] presented a gesture recognition model which MEMS accelerometer based and it recognizes seven hand gestures, i.e. left, right, up, down, circle, tick and cross. The accelerometer used here works in 3-axis which is used to recognize perpendicular directions created by hand motion. The system has advantage over other systems as its recognition model uses the sign sequence and template matching techniques. These techniques are used for the recognition of gestures for any user, instead of any specific user.

Akl, Feng and Valaee [4] presented a gesture recognition system which works on 3-axis accelerometer sensor. The sensor works in two stages: Training and Testing stage. The system uses the combination of Dynamic Time Warping (DTW) and Affinity Propagation (AP). The system uses this combination for the training purpose and for testing several gesture traces are projected into same lower dimension subspace, considering the collection as l_1-minimization problem. The system has shown better results for user-independent and mixed user recognition samples.

Elmezain, Al-Hamadi, Sadek and Michaelis [5] presented a system, formulating a non-gesture model which uses the comparative analysis of Hidden Markov Model and Conditional random fields. For reducing the number of states for HMM a new technique has been applied which combines the similar probability distribution states. The similar states are grouped according to the relative entropy measure. Results shows the individual accuracy of 93.31% and 90.49% for HMM and CRF respectively.

Wen and Niu [6] presented a system which recognizes the hand gesture from the complex background. A combination of morphological operations is applied such as erosion and dilation for detecting the hand counters. The angles of finger-tip are calculated in order to mark out the fingertips from the captured gesture image. The system presented here is not workable in a real time video processing environment.

Yun and Peng [7] suggested a system with two sub systems for detecting and recognizing the hand gestures respectively. The detection of hand gesture is done using the Viola-Jones method which is one of the fastest and most accurate techniques for object detection. For the purpose of final recognition SVM (Support Vector Machines) classifier is used. It gives high generalization performance without the additional requirement of knowledge base. The system also uses the Hu invariant moments for the purpose of feature extraction. The system is expected to include more gesture patterns as it was trained for only three types of gesture images such as Palm, Index, LPalm.

Joslin, El-Sawah, Chen and Georganas [8] presented a system which recognizes the hand gestures using real time camera. The input is tracked and later this input is classified into meaningful gestures by separating the fingers and hands, then these gestures are further projected into 3D space and Hidden Markov Model is used for further processing.

III. METHODS AND TECHNIQUES

The proposed work focuses on capturing both types of gestures: Static as well as dynamic. The static gestures can be given through array of images. This predefined array acts as a training data for the proposed system. Otherwise, user can also issue some online gestures and their capturing is performed [6]. The retrieval process for the gestures can be performed through web camera and later the recognition algorithm is applied. The recognition algorithm will identify the gesture issued by the user and after this required functioning can be performed. The main task here is to refine the gesture input, such as reducing the noise level [8]. In the proposed system Non local means de-noising algorithm is used for reducing the noise levels and further refinement of the image frames captured.

The captured frames from the video sequence consist of various unwanted regions. For obtaining the region of interest various techniques can be applied. Marker based image segmentation technique has been applied in the proposed system. In order to achieve best results watershed algorithm has been applied to a greater effect [4]. As, this algorithm is one of the marker based technique, so different labels are decided for marking the unwanted and wanted regions. The unwanted regions are marked with 0 label and wanted regions are given a different colour mark.

Watershed algorithm consists of three main steps:

1. Calculating gradient image values
2. Using watershed algorithm step
3. Merging step

Consider the system is tracking for the hand gesture then the unwanted background objects will be removed and only the required hand region will be tracked. The same can be applied for other gestures also, like face or even the eye movements.

The next step is to apply the recognition algorithm. The recognition algorithm works in two steps:

1. Contour Extraction
2. Detection of Convex Hull and Defects points

For hand gestures the contours should be extracted and then the points for convex hull and defects are also identified. The proposed system uses the combination of Laplace and Canny algorithms available under OpenCV for achieving the refined results. The Laplace algorithm is used for line and edge detection whereas the canny algorithm is used for finding the intensity gradient of the image using the following equations:

Edge_Gradient $(G) = \sqrt{(G_x^2 + G_y^2)}$, where both terms in the equation are horizontal and vertical derivates respectively [4].

After performing the edge detection, Contour Extraction can be done using the findContours() and drawContours() methods of OpenCV.

Once the contours are extracted the next step is to detect the convex hull and defect points. It is defined as a bounded subset of the plane X. It can be visualized as the shape formed by a rubber band stretched around X. The convex hull can be drawn around the hand using the convexHull() method and later convexity defects points are identified using the convexityDefects() method in OpenCV.

Figure 1: Convex Hull

Figure 2: Convexity Defects of a Hand Contour

A. Proposed Algorithm

Step 1: Capture gesture patterns through web cam.

Step 2: Turn on the training module of the system and train the system with the live gestures

Step 3: Set the time period for each gesture pattern and repeat step 2 until unless all the gesture patterns are not recorded

Step 4: Select course of action against each gesture pattern, from the available options like: Playing music, recording some text in the notepad, clicking of the picture, etc.

Step 5: After completing the training with all course of actions, convert the gesture patterns into binary format and save them into the database

Step 6: Start the testing module and turn on the recognition module

Step 7: Give gesture patterns as input and recognition algorithm will also start working

Step 8: After recognition the corresponding gesture pattern will be tracked and corresponding action will be performed

Step 9: Perform Step 7 and 8 iteratively and check the accuracy counter.

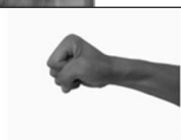
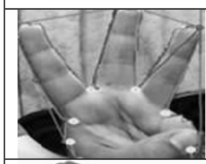
Step 10: If the accuracy counter is above 90%, then freeze the gesture pattern and move the system into real mode.

Step 11: After Step 10, the corresponding gesture pattern can be utilized to perform real time actions.

IV. RESULTS AND DISCUSSION

After implementing the above methods and techniques, quite meaningful results were obtained. The system was tested repeatedly for five different types of gesture patterns. Following table illustrates the results:

Table I

Gesture pattern	**Gesture type**	**No. of times pattern tested**	**No. Of times correctly recognized**	**Accuracy**
	Dynamic	100	92	0.92
	Dynamic	100	87	0.87
	Dynamic	100	90	0.90
	Dynamic	100	95	0.95
	Dynamic	100	98	0.98
			Net Accuracy	0.928

The system is producing accuracy of 92.8% which is quite good, subjected to the background noise. The interpretation suggests us that more filters and noise reduction algorithms can be applied to provide the better results. Despite of the fact that dynamic gestures were fed as input the accuracy seems to be quite brilliant and can be utilized to greater effect.

Table II

Gesture pattern	*Gesture type*	*No. of times pattern tested*	*No. of times correctly recognized*	*Accuracy*
Open Palm	Static	100	98	0.98
Closed Palm	Static	100	96	0.96
Three fingers raised	Static	100	95	0.95
Two fingers raised	Static	100	92	0.92
One finger raised	Static	100	95	0.95
			Net Accuracy	0.952

The results for the static gesture inputs were good as compared to the dynamic one. The static gestures were fed as static noise free images, so the recognition algorithm performed well for the static gestures. The net accuracy of 95.2% is far much better than the systems described in the review. The marker based technique has given very good results, along with that the recognition model with de-noising techniques has given push to the accuracy.

V. CONCLUSIONS AND FUTURE WORK

The proposed system can be applicable for desktop as well as smart phone applications. Although smart phones are implementing this technology to a great effect, but still the customization of gestures is not present. Live gestures can be tracked and user can customize the course of action according to its need. Smart phones are using pre defined set of gesture patterns. Training and testing modules can be used to create personalized gestures. Still the dynamic gestures are lacking at the accuracy levels, so this can be improved. The recognition model can be improved using erosion and dilation algorithms.

ACKNOWLEDGMENT

This research was supported by my fellow research mates I would like to thank my professors who helped me in every possible way to achieve the desired results.

References

[1] G. Constante, V. Galieni, Y. Yan, M. L. Fravolini, E. Ricci, P. Valigi, “Exploiting transfer learning for personalized view invariant gesture recognition”,” IEEE International Conference on Acoustic, Speech and Signal Processing”, 2014.

[2] K. Y. Lian, B. H. Lian, “Gesture recognition using improved hierarchical hidden markov algorithm”, “IEEE International Conference on Systems,Man, and Cybernetics”, 2013.

[3] R. Xu, S. Zhou, W. J. Li, “MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition”, “*IEEE Sensors Journal*”, Vol. 12, No. 5, 2012.

[4] A. Akl, C. Feng, S. Valaee, “A Novel Accelerometer-Based Gesture Recognition System”, *IEEE Transactions on Signal Processing*, Vol.: 59, No. 12, 2011.

[5] M. Elmezain, A. Al-Hamadi, S. Sadek, B. Michaelis, “Robust Methods for Hand gesture Spotting and Recognition Using Hidden Markov Models and Conditional Random Fields”, 978-1-4244-9991-5, “IEEE”, 2011.

[6] X.Wen, Y. Niu, “A method for Hand Gesture Recognition Based on Morphology and Fingertip-Angle”,”IEEE”,2012.

[7] L.Yun, Z.Peng, “An Automatic Hand Gesture Recognition System Based on Viola-Jones Method and SVMs”, “International Workshop on Computer Science and Engineering”, 2009.

[8] C. Joslin, A. El-Sawah, Q. Chen, N. Georganas, “Dynamic Gesture Recognition”, “Instrumentation and Measurement Technology Conference Otawa,Canada”, 2005.

[9] X. Deyou, A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG, International Conference on Pattern Recognitioin, pp. 519-522, 2006.

[10] X. Zhang, X. Chen, W. W. hui, J. Y. hai, V. Lantz, and K. W. qiao, “Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors,” in *Proc. 13th Int. Conf. Intell. User Interfaces (IUI ’09)*, New York, 2009, pp. 401–406.

[11] T. Schlömer, B. Poppinga, N. Henze, and S. Boll, “Gesture recognition with a Wii controller,” in *Proc. 2nd Int. Conf. Tangible and Embedded Interaction (TEI’08)*, 2008, pp. 11–14.

[12] M.H. Yang and N. Ahuja. Detecting human faces in color images. In Proc. Of ICIP’98, volume 1, pages 127–130, 1998.

[13] L. Zelnik-Manor and M. Irani, “Statistical analysis of dynamic actions,”IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, no. 9, pp. 1530–1535, 2006.

[14] J.C. Terrillon, M.N. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In Proc. of FG’00, pages 54–61, 2000.

[15] Cen Rao Alper Yilmaz and Mubarak Shah, “Viewinvariant representation and recognition of actions.,” *International Journal of Computer Vision*, Vol. 50, No. 2, pp. 203–226, 2002.