Proceedings of
National Conference on Innovative Trends in Computer Science Engineering (ITCSE-2015)
held at BRCMCET, Bahal on 4th April 2015

HAND GESTURE RECOGNITION TECHNIQUES: A REVIEW

[1]Yogita Bhardwaj, [2]Sukhdip Singh
[1]M.Tech. Scholar, [2]Associate Professor
[1,2]DCRUST Murthal
[1]yogitabhardwaj123@gmail.com , [2]sukhdeepsingh.cse@dcrustm.org

Abstract:The development of computing technique needs the interaction between the user with the machine. Hand gesture is used in regular life. This is used for interaction between humans and computing devices. The hand gesture recognition is the most reliable technique for interaction. It can be used for the interaction using the sign language. This paper describes about the hand gesture and their different techniques used for the hand gesture recognition.

The scale invariant feature extraction algorithm is also explained in this paper.

Keyboard: Gesture, Hand Gesture, Hand Gesture Recognition Techniques.

1. INTRODUCTION

Human gestures constitute a space of motion expressed by the body, face, and/or hands. Among a variety of gestures, hand gesture is the most expressive and the most frequently used. The gestures have been used as an alternative form to communicate with computers in an easy way. This kind of human-machine interfaces would allow a user to control a wide variety of devices through hand gestures. Most work in this research field tries to elude the problem by using markers, marked gloves or requiring a simple background. Glove-based gesture interfaces require the user to wear a cumbersome device and generally carry a load of cables that connect the device to a computer.A real-time gesture recognition system which can recognize 46 ASL letter spelling alphabet and digits was proposed. The gestures that are recognized are static gestures without any motion [1]. The remarkable ability of the human vision is the gesture recognition. It is mainly noticed by deaf people when they communicating with each other via sign language and with hearing people as well [1]. Gesture is defined as an expressive movement of body parts which has a particular message that can be communicated precisely between a sender and a receiver.

2. HAND GESTURE RECOGNITION

Hand Gesture Recognition plays a vital role in human computer interaction (HCI). Hand Gesture Recognitions have variety of applications in virtual reality, computer games and sign language recognition. Despite lots of previous work, conventional vision-based hand gesture recognition methods [3] are still far from satisfactory for real-life applications. The quality of the captured images is sensitive to lighting conditions because of the limitations of the optical sensors, thus it is not able to detect and track the hands robustly that largely affects the performance of hand gesture recognition. As compared to optical sensors, these sensors are more reliable and they are not affected by lighting conditions or cluttered backgrounds [2]. They require the user to wear a data glove and sometimes calibration.

3. HAND GESTURE RECOGNITION TECHNIQUES

Hand gesture recognition techniques can be further classified under static and dynamic gestures. To detect static gestures (i.e. postures), a general classifier or a template-matcher can be used. However, dynamic hand gestures have a temporal aspect and require techniques that handle this dimension like Hidden Markov Models (HMM) unless the temporal dimension is modeled through the hand gesture representation.

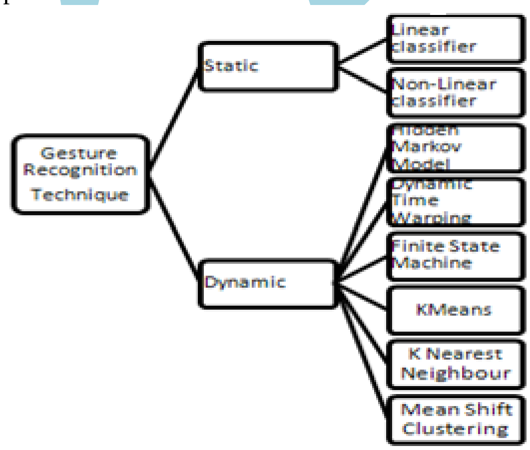

Figure 1 : Techniques of hand gesture recognition [4]

The static hand gestures are further classified into linear learner and non-linear learner. This is suited for linearly separable data and the latter for the other cases. The other way to categorize learning algorithms is to consider their outcome. They distinguishes supervised learning (i.e. matching samples to labels), unsupervised learning (i.e. only sample clusters without labels), semi-supervised learning (i.e. mix of labelled and un-labelled data), reinforcement learning (i.e. learns policies given observations, transduction (i.e. supervised learning with prediction, and learning to learn (i.e. learns his own inductive bias based on previous experience. The choice of the learning algorithm depends mainly on the chosen hand gesture representation. For instance, FSMs, HMMs, PNF (i.e. Past-Now-Future network) are sort of automaton with a set of states and a set of transitions. The states represent static hand gestures (i.e. postures) and transitions represent allowed changes with temporal and/or probabilistic constraints.

The dynamic hand gesture is considered as a path between an initial state and a final state. Authors [5] proposed an approach for gesture recognition using HMM based threshold. Authors [6] present a method for recognizing human gestures using PCA-HOG global descriptor. The limitation of the approaches (based on automata) is that the gesture model must be modified when a new gesture needs to be recognized. The computational complexity of such approaches is generally huge since it is proportional to the number of gestures to be recognized which is not the case for methods based on other techniques. The most common techniques used for static and dynamic hand gesture recognition are as follows:

3.1 HIDDEN MARKOV MODEL

HMM were introduced in the mid 1990s and speedily became the recognition method of choice, due to its implicit solution to the segmentation problem. A hidden Markov model (HMM) can be considered a generalization of a Markov chain without this Markov chain restriction. Since HMMs can have more than one arc with the same output symbol which are nondeterministic and it is impossible to directly determine the state sequence for a set of inputs simply by looking at the output. HMM is defined as the set of states of that contains an initial state, set of output symbols and a set of state transitions. Each state transition is described by the state from which the transitions start and transition resultant states, the output symbol generated and the probability that the transition is taken. In the context of hand gesture recognition, each state shows a set of possible hand positions.

3.2 K-MEANS

The k-means problem is to determine k points called centers so as to minimize the clustering error, defined as the sum of the distances of all data points to their respective cluster centers. This classification finds statistically similar groups in multi-spectral space. The algorithm starts by randomly locating k clusters in spectral space. In the input image group each pixel is then assigned to the nearest cluster centre and the cluster centre locations are moved to the average of their class values. This process is repeated until a stopping condition is met. The stopping condition may either be a maximum number of iterations (specified by the user) or a tolerance threshold which designates the smallest possible distance to move cluster centres before stopping the iterative process. The most commonly used algorithm for solving this problem is the Lloyd's k-means algorithm which iteratively assigns the patterns to clusters and computes the cluster centers.

MacQueens k-means algorithm is a two-pass variant of the Lloyd's k-means algorithm: choose the first k patterns as the initial k centers and the assign each of the remaining N–k patterns to the cluster whose center is closest. Calculate the new centers of the clusters obtained. Each of the N patterns are assigned to one of the k clusters obtained in step 1 based on its distance from the cluster centers and recompute the centers.

3.3 K-NEAREST NEIGHBOR

It is a method for classifying objects based on closest training examples in the feature space. It is based on instance-based learning or lazy learning, in which the function is only approximated locally and all computations are deferred until classification. Object can be classified by the majority vote of its neighbors along with the object being assigned to the class most common amongst its k nearest neighbors; k is a positive integer, typically small. If k = 1, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful in choosing k to be an odd number as this avoids tied votes. Similar method can be used for regression, by assigning the property value for the object to be the average of the values of its k nearest neighbors. It is useful to weight the contributions of the neighbors, therefore the nearer neighbors contribute more than the distant ones. The neighbors were choosen from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. It is thought of as the training set for the algorithm, where no explicit training step is required. To identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is necessary to use the Euclidean distance, though other distance measures, such that the Manhattan distance could in principle be used. The KNN algorithm is very sensitive to the local structure of the data.

3.4 MEAN SHIFT CLUSTERING

It is a nonparametric clustering technique [7] which does not require prior knowledge of the number of clusters, and it is not constrain the shape of the clusters. The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. They perform a gradient ascent procedure on the local estimated density until convergence for each data point in the feature space. This procedure represents the modes of the distribution. The data points associated (at least approximately) with the same stationary point are considered members of the same cluster.

3.5 DYNAMIC TIME WARPING

It has long been used to find the optimal alignment of two signals. The DTW algorithm calculates the distance between each possible pair of points out of two signals in terms of their associated feature values. By using these distances they calculate a cumulative distance matrix and find the least expensive path through this matrix. This path represents the ideal warp—the synchronization of the two signals which causes the feature distance between their synchronized points to be minimized. The signals are normalized and smoothed before the distances between points are calculated. DTW has been used in various fields, such as speech recognition, data mining, and movement recognition. DTW mainly focused on speeding up the algorithm whose complexity is quadratic in the length of the series. Examples are applying constraints to DTW, approximation of the algorithm and lower bounding techniques.

3.6 FINITE STATE MACHINE

A finite state machine is one that has a limited or finite number of possible states that can be used both as a development tool for approaching and solving problems and as a formal way of describing the solution for later developers and system maintainers. There are many ways to show state machines through graphically animated illustrations using the simple tables. Usually, the training of the model is done off-line, using many possible examples of each gesture as training data, and the parameters (criteria or characteristics) of each state in the FSM are derived. The hand gestures recognition can be performed online using the trained FSM. When input data (feature vectors such as trajectories) are supplied to the gesture recognizer, the latter decides whether to stay at the current state of the FSM or jump to the next state based on the parameters of the input data.

4. RELATED WORK

Qing Chen et al. [8] (2009) discusses two body-language Human–Computer Interaction (HCI) modalities, namely facial expressions and hand gestures, for healthcare and smart environment applications.

Pragati Garg et al. [9] (2009) said that with the development of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient. Due to the limitation of these devices the useable command set is also limited. The author reviews vision based hand gesture recognition. The existing approaches are categorized into 3D model based approaches and appearance based approaches, highlighting their advantages and shortcomings and identifying the open issues.

Jagdish Lal Raheja et al. [10] (2010) recognize that controlling a robot arm, in real time, through the hand gestures is a novel approach. The technique proposed here was tested under proper lighting conditions that they created in their laboratory. Experimental results show that the system can detect hand gestures with an accuracy of 95 % when it is kept still in front of the camera for 1 second.

Luigi Lamberti et al. [11] (2011) described real-time hand gesture recognition system based a color Glove. The system is formed by three modules. The first module identifies the hand image in the scene. The second module performs the feature extraction and represents the image by a nine-dimensional feature vector. The third module, the classifier, is performed by means of Learning Vector Quantization.

Alsheakhali, M. et al. [12] (2011) proposed a new technique to increase the adaptability of a gesture recognition system. They have implemented a real-time version, using an ordinary workstation with no special hardware beyond a video camera input. The technique works well under different degrees of scene background complexity and illumination conditions with more than 94% success rate.

Meenakshi Panwar et al. [13] (2012) proposed a shape based approach for hand gesture recognition with several steps including smudges elimination orientation detection, thumb detection, finger counts etc.

Manhal S. Almohammad et al. [14] (2013) discussed multimodal biometrics to increase the security level. With the fusion of multiple biometrics they can minimized the system error rates. Abhinandan Julka et al. [15] (2013) said that today world is running behind the computer industries and pattern recognition is one of the important and vast fields of computer intelligence. The technique given in the paper provides a human hand interface with computer which can recognize static gestures from American Sign Language. Since 24 gestures from American Sign Language (ASL) are static so, they was able to recognize them. Their objective is to develop a hand gesture recognition system which can recognize most of the static characters from ASL with a good accuracy which can only work offline and is mainly dependent on database.

Yongjing Liu et al. [16] (2013) proposed a method to recognize multifarious static hand gesture. In this research, the preprocessing for the captured video image followed by feature extraction and classification.

5. DRAWBACK OF EXISTING TECHNIQUES

The existing techniques provide the maximum recognition rate of 85.9%. In the existing technique the system is not adaptive. The variation in the scale of the image doesn't recognize the gesture. The system must be robust enough to tolerate such errors.

6. SIFT

The SIFT algorithm takes an image and transforms it into a collection of local feature vectors. All feature vectors is supposed to be distinctive and invariant to any rotation or translation of the image. For implementation, they use these features to find distinctive objects in different images and the transform can be extended to match faces in images [17].

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description which is extracted from a training image, used to identify the object when attempting to locate the object in a test image containing many other objects [17]. To perform reliable recognition, the features extracted from the training image are detectable even under changes in image scale, noise and illumination. These points usually lie on high-contrast regions of the image, like an object edges. Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. Consider an example, if the four corners of a door were used as features, they continuous to work without take care of the door's position; but if points in the frame were also used, the recognition fails if the door is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. SIFT detects and uses a much larger number of features from the images that reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors [17].

7. CONCLUSION

In today's digitized world, processing speeds have increased dramatically, with computers being advanced to the levels where they can assist humans in complex tasks. Still input technologies seem to cause a major bottleneck in performing some of the tasks under-utilizing the available resources and restricting the expressiveness of application use. This paper discusses about the gesture, hand gesture and their different techniques. The paper brief the finding of various existing authors. The drawback of the existing work along with the sift is also explained. In future the existing techniques can be extended by using the SIFT to make the robust.

REFERENCES

[1] Lowe, D.G., "Distinctive Image Features from Scale-Invariant Keypoints", January 5, 2004.

[2] Zhou Ren et al. ,"Robust Hand Gesture Recognition Based on Finger-Earth Mover's Distance with a Commodity Depth Camera", MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

[3] C. Chua, H. Guan et al. ,"Model-based 3d hand posture estimation from a single 2d image", Image and Vision Computing, 20:191, 2002.

[4] Rautaray, Siddharth S., and Anupam Agrawal ,"Vision based hand gesture recognition for human computer interaction: a survey", Artificial Intelligence Review (2012): 1-54.

[5] Lee H-K et al. ,"An hmm-based threshold model approach for gesture recognition", IEEE Trans Pattern Anal Mach Intell 21(10):961–973, 1999.

[6] Lu W-L et al. ,"Simultaneous tracking and action recognition using the pca-hog descriptor", In: The 3rd Canadian conference on computer and robot vision, 2006. Quebec, pp 6–13.

[7] Derpanis KG (2005) Mean shift clustering, Lecture Notes. http://www.cse.yorku.ca/~kosta/CompVis_Notes/mean_shift.pdf

[8] Chen, Q., Cordea, M. D., Petriu, E. M., & Varkonyi-Koczy, A. R. (2009). Human Computer interaction for smart environment applications using hand gestures and facial expressions. *International Journal of Advanced Media and Communication*, *3*(1), 95-109.

[9] Garg, P., Aggarwal, N., & Sofat, S. (2009). Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology*, *49*(1), 972-977.

[10] Raheja, J. L., Shyam, R., Kumar, U., & Prasad, P. B. (2010, February). Real-time robotic hand control using hand gestures. In *Machine Learning and Computing (ICMLC), 2010 Second International Conference on* (pp. 12-16). IEEE.

[11] Lamberti, L., & Camastra, F. (2011). Real-time hand gesture recognition using a color glove. In *Image Analysis and Processing–ICIAP 2011* (pp. 365-373). Springer Berlin Heidelberg.

[12] Alsheakhali, M., Skaik, A., Aldahdouh, M., & Alhelou, M. (2011). Hand Gesture Recognition System. *Information & Communication Systems*, 132.

[13] Panwar, M. (2012, February). Hand gesture recognition based on shape parameters. In *Computing, Communication and Applications (ICCCA), 2012 International Conference on* (pp. 1-6). IEEE.

[14] Almohammad, M. S., Salama, G. I., & Mahmoud, T. A. Face and Gait Fusion Methods: A Survey.

[15] Abhinandan Julka, Sandeep Bhargava, (2013). A Static Hand Gesture Recognition Based on Local Contour Sequence. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013 ISSN: 2277 128X.

[16] Liu, Y., Yang, Y., Wang, L., Xu, J., Qi, H., Zhao, X., ... & Wang, J. (2013, July). Image Processing and Recognition of Multiple Static Hand Gestures for Human-Computer Interaction. In *Image and Graphics (ICIG), 2013 Seventh International Conference on* (pp. 465-470). IEEE.

[17] Gurjal, P., & Kunnur, K. (2012). Real Time Hand Gesture Recognition using SIFT. *International Journal for Electronics and Engineering*, 19-33.