Classifiers for hand gesture recognition

Merav Shmueli, Helman Stern, Sigal Berman

Department of Industrial Engineering and Management, Deutsche Telekom Labs at BGU
Ben Gurion University of the Negev, Beer-Sheva, Israel
Merav207@bgu.ac.il

***Abstract*— The use of hand gestures provides an attractive alternative to cumbersome interface devices for human – machine interaction. The current research addresses real time gesture classification and aims to develop an algorithm capable of accurate classification of gestural control commands. Two different classifiers were developed for classifying a gesture vocabulary of eight dynamic hand gestures. The classifiers developed were: K-means + rule-based classifier and longest common subsequence (LCS) classifier. An experiment was performed to determine the recognition accuracy of the classifiers in which a test set of 180 trajectories were classified. The obtained accuracies are 90 and 94 percent for the K-means and LCS classifiers, respectively.**

Keywords-component: Classification, Human machine interface, LCS, K-means, Gesture recognotion

I. INTRODUCTION

The act of gesturing is an integral part of human communication used to express a variety of directives, feelings and thoughts. Hand gesture recognition research has attracted attention because it provides an attractive alternative to cumbersome interface devices for human – machine interaction. Possible application areas for control based on gesture recognition include smart home interfaces [1] [2], computer interfaces [3] [4], medicine [5] and robotics [6]. In dynamic hand gesture recognition systems (GRS) the hand must first be tracked by a tracking algorithm in order to generate a trajectory that represents the movement of the hand. This trajectory then serves as input to the gesture classifier. Classification in a GRS is stage of identifying (recognizing) the tracked trajectory as one gesture of a pre-defined gesture vocabulary or as a non-gesture (random hand movement). Many approaches were used for classifying gestures, including: Neural Networks, Fuzzy C means inference systems [7], Hidden Markov Models [8], Support Vector Machines [9], etc. While most of the listed methods show high accuracy, results their accuracy depends on the existence of a large amount of training data. This may restrict their implementation and rule out options such as self-defined gestures. In addition, for some algorithms computation times are long, precluding real - time operation. Some classifiers require a preliminary stage of gesture spotting that detect track segments which potentially contain gestures, these segments are then sent to the classifier.

A generic GRS architecture is depicted in Figure 1 [10]. The state machine is an application which is controlled by the recognized gestures.

In this paper we compare two gesture recognition classifiers developed for recognizing a gesture vocabulary of six dynamic gestures, a longest common substring (LCS) classifier and a k-means rule based classifier [11]. Section II describes the gesture vocabulary recognized by the classifiers and a brief description on the classifiers. In section III the experimental procedure and data analysis is described, and in section IV the results of the classifier accuracy experiment are given.

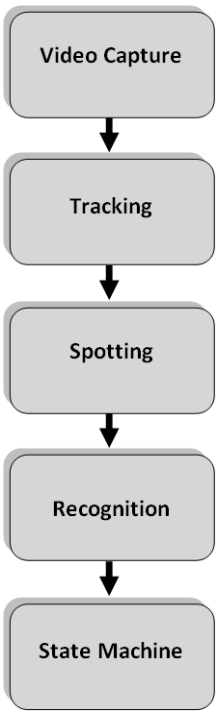

Figure 1 - Basic architecture for complete gesture recognition

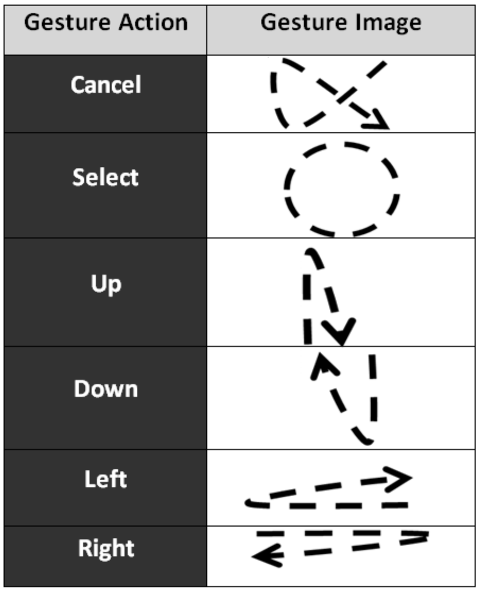

Figure 2 – Gesture vocabulary for GRS remote control

II. Methods

A. Gesture vocabulery

The gesture vocabulary is comprised of six dynamic gestures (Figure 2). The vocabulary was designed after a preliminary user experiment and the selected gestures were chosen using the criteria of: intuitiveness, memorability and distinct enough to avoid confusing between them. Two groups are defined in the gesture vocabulary: navigation (up, down, left and right) and control (cancel and select). The time length for navigation gestures is about 1 second and for control gestures about 2 seconds.

One of the main problems in gesture recognition is to determine where the gesture begins and ends. One way resolve this problem is to impose a pause before and after a gesture in order to detect the lack of hand movement, and mark the beginning and end of a gesture. In this paper the experiment was made using pauses between gestures.

B. Classifiers tested

Two classifiers have been implemented; a k-means rule base classifier, and a LCS classifier [11]. Both do not require large training datasets and their computational load is low. The computational load of the LCS classifier is a bit higher yet it does not require a spotter while the K-means rules based classifier requires one.

1. K-means Rule based classifier – The classifier has two stages. The first is based on clustering a training set of gestures, using two features: aspect ratio and standard deviation of the distances from the trajectory boundary to its center of gravity. The gestures are divided into two groups: control and navigation. The clustering stage is followed by a rule-based stage, where specific gestures are recognized using a sequence of rule filters [11].
2. Longest Common Subsequence (LCS) classifier – The LCS method is based on dynamic programming principles. The feature used by the LCS classifier is the absolute angels of the sequence of motion vectors in a trajectory. The algorithm compares the angle sequence to pre-defined template patterns representing the six gestures seeking the highest correlation. Some gestures are represented by multiple templates due to common user variations and alternative starting points [11]. This classifier was developed due to low accuracy in recognizing the 'cancel' gesture by the k-means classifier.

III. Experiment

A. Subjects and experimental procedure

The experiment was performed in order to evaluate the classifiers accuracy rate. Four well-behaved users performed all six gestures in the gesture vocabulary. Thirty trajectory samples per gesture were recorded. In addition, thirty non gesture trajectory samples were tested to evaluate the classifiers ability to reject random hand movements, for a total of 180 gesture trajectories. The users hand positions were tracked for four seconds using a Fastrak (Polhemus) magnetic position sensor. A position sensor was used rather than a vision-based tracker in order to obtain very accurate ground truth trajectories. Sample trajectories used as input to both classifiers are shown in figure 3.

B. Data analysis

Classifier output was summarized in a confusion matrix. Four measurements are indicative of the classifier performance:

- Accuracy rate (HIT) – percentage of trajectory samples classified correctly from all trajectory samples tested.
- Error rate (MISS) - percentage of trajectory samples classified incorrectly from all trajectory samples tested.
- False alarm – percentage of non-gesture trajectories classified as control gesture from all non-gesture trajectory samples tested.
- Correct rejection – percentage of non-gesture classified as non-gestures from all non-gesture trajectory samples tested.

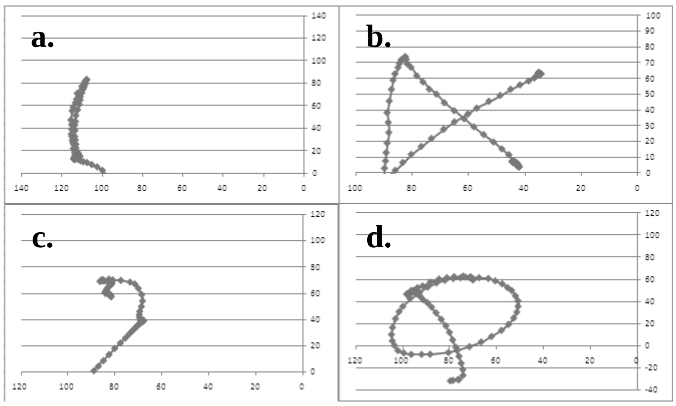

Figure 3 – Sample trajectories of: a. up, b. cancel,

c. non-gesture, d. select

Clearly a better classifier has a high accuracy rate, correct rejection, low error rate and false alarm.

IV. RESULTS AND DISCUSSION

The LCS classifier (Table 1, Figure 4) achieved high accuracy (mean number) and low error (mean number) rate. However some of the recognition rates required further investigation in order to achieve future improvements:

- Up gesture - Results show that the up gesture has the lowest "HIT" rate (84%) and the highest "MISS" rate (16%). Al the miss-classified up gestures are classified as down gestures. Further examination showed that most miss-classified samples resulted from users making a small down movement before the main up movement. Since the users were, experienced users to improve the classifier the algorithm must be improved by further examination of the users' behavior. Adding an additional pattern that represents the movement described above should improve the results.

- Down gesture – although the accuracy rate is high, 7% of misclassified gestures where found to be a result of recognizing the movement as an up gesture. This is because users moved their hand up before starting the down movement of the gesture. A possible solution to the problem is to impose a pause in the beginning of every gesture in order to detect the true gesture's beginning. A different solution is possible which consists of adding the movement up to the beginning of the gesture as part of the down gesture and thereby adding an additional pattern.

The K-means + Rule-Based classifier achieved a good accuracy (number) however its results were lower than that of the LCS classifier. The factors effecting misclassified gestures are discussed below.

- Cancel gesture - Results show that the cancel gesture has the lowest "HIT" rate (70%). Low identification of the cancel gesture was one of the reasons for developing a second classifier. Further examination of the miss-classified cancel gestures indicates that the main reason comes from the "harsh" conditions of the rules distinguishing the cancel gesture from random hand movements such as: requiring a single intersection within the gesture trajectory, and the rule requiring the end and beginning of the gesture to be on the same side of the trajectory's bounding box, etc. A possible way of fixing this problem is by optimizing the rule threshold parameter values.

- Left gesture – examination of the testing set reviles that the some of the users, execute a curved left gesture. This movement should be considered in future research and since the rules in the k-means classifier for a horizontal navigation gesture require a large standard deviation of the distances from the gesture center of gravity to the trajectory's boundary, and a small aspect ratio, two conditions that do not exist.

Measurements	LCS Classifier	K-means + Rule-Based Classifier
Accuracy Rate	94%	90%
Error Rate	6%	10%
False Alarm	16%	16%
Correct Rejection	84%	84%

Table 1 – Testing classifiers experiment measurements

LCS Classifier

	Cancel	Select	Up	Down	Left	Right	Non-Gesture
Cancel	93%	0%	7%	0%	0%	0%	0%
Select	0%	97%	0%	0%	0%	3%	0%
Up	0%	0%	84%	16%	0%	0%	0%
Down	0%	0%	7%	93%	0%	0%	0%
Left	0%	0%	0%	0%	100%	0%	0%
Right	0%	0%	0%	0%	3%	97%	0%
Non-Gesture	0%	0%	3%	10%	3%	0%	84%

K-means + Rule-Based Classifier

	Cancel	Select	Up	Down	Left	Right	Non-Gesture
Cancel	70%	7%	0%	0%	0%	0%	23%
Select	0%	100%	0%	0%	0%	0%	0%
Up	0%	0%	97%	0%	0%	0%	3%
Down	0%	0%	0%	90%	0%	0%	10%
Left	3%	0%	0%	0%	83%	0%	13%
Right	0%	0%	0%	0%	0%	100%	0%
Non-Gesture	0%	13%	3%	0%	0%	0%	84%

Figure 4 – Testing classifier experiment results

The LCS classifier proved to be superior to the K-means rules based classifier. However, examining the confusion matrices we can see that for different gestures different classifiers gave better results. Therefore augmenting both classifiers may improve recognition. .

V. CONCLUSIONS

This paper discusses the classification stage of a GRS developed for remote control of a TV channel selection interface [11]. The paper presents the comparative experiment of the two types of candidate classifiers: a k-mean rule based classifier and a LCS classifier. The experimental results prove the superiority of the LCS classifier. The experiment identified deficiencies in both classifiers, and provides directions for further improvements.

ACKNOWLEDGMENT

This work was supported by Deutsche Telekom AG. We would like to acknowledge Darya Frolova, Tom Godo, Eti Almog and Omri Mendels for their help in this study.

REFERENCE

[1] T. Baudel and M. Baudouin-Lafon, Charade: remote control of objects using free-hand gestures. Comm. ACM **36** 7 (1993), pp. 28–35.

[2] Yoon, H. S., Soh, J., Bae, Y. J., & Seung Yang, H. (2001). Hand gesture recognition using combined features of location, angle and velocity. *Pattern Recognition, 34*(7), 1491-1501.

[3] Kang, H., Woo Lee, C., & Jung, K. (2004). Recognition-based gesture spotting in video games. *Pattern Recognition Letters, 25*(15), 1701-1714.

[4] Geer, D. (2004). Will gesture-recognition technology point the way? *Computer, 37*(10), 20-23.

[5] Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., et al. (2007). Gestix: A doctor-computer sterile gesture interface for dynamic environments. *Soft Computing in Industrial Applications: Recent and Emerging Methods and Techniques,* 30–39.

[6] Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: Remote control of objects using free-hand gestures. *Communications of the ACM, 36*(7), 28-35.

[7] Freeman, W. T., & Weissman, C. (2000). Television control by hand gestures. Paper presented at the *Proc. of Intl. Workshop on Automatic Face and Gesture Recognition,* 179-183.

[8] H. Stern, D. Frolova, S. Berman, 2010. Hand Gesture Recognition for TV Remote Control using Tree-Based Ensemble and LCS Classifiers, WORLDCOMP'10 The 2010 World Congress in Computer Science, Computer Engineering, and Applied Computing, July 12-15, Las Vegas, USA.

[9] Starner, T. E. (1995). *Visual Recognition of American Sign Language using Hidden Markov Models,*

[10] Vapnik V. Statistical Learning Theory. John Wiley and Sons, New York, 1998.

[11] Yu, J. (2005). General c-means clustering model. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*

[12] Manigandan M., I. Manju Jackin, "*Wireless Vision Based Mobile Robot Control Using Hand Gesture Recognition Through Perceptual Color Space"* ace, pp.95-99, 2010 International Conference on Advances in Computer Engineering, 2010

This research was supported by Deutsche Telekom AG.