

# A hybrid method for hand gesture recognition

Yu Huang, Dorothy Monekosso, Hui Wang, Juan Carlos Augusto  
*School of Computing and Mathematics, University of Ulster, Jordanstown,  
Co. Antrim, BT37 0QB, UK*  
{huang-y4, DN.Monekosso, h.wang, jc.augusto}@ulster.ac.uk

**Abstract**—Hand gesture recognition aims to recognize the meaningful expressions of hand motion. It is widely used in information visualization, robotics, sign language understanding, medicine and healthcare. Some methods have been proposed for hand gesture recognition. But no single algorithm can handle all kinds of situations, because of the complex environment.

In this study, we propose a hybrid method for hand gesture recognition, which extends our previous work on a gesture recognition method based on concept learning by the addition of an association learning process. We use association learning to reveal the frequent patterns in gesture sequences, and then use such patterns to help recognize incomplete gesture sequences. Experiments show the use of association learning does indeed improve recognition accuracy. Experiments also show the hybrid method is comparable to two state of the art methods (HMMs and DTW) for hand gesture recognition, but outperforms them in the larger datasets.

**Keywords**- *self-training, association rules, clustering ensembles, gesture recognition*

## I. INTRODUCTION

Gesture recognition is one of the most important tasks in artificial intelligence and data mining area. Many applications of gesture recognition play an interesting role in intelligence environments, such as hand gesture recognition, audio expression recognition and facial expression recognition. In our daily lives, hand gestures are frequently used to interact with the environment around us, so it's an important function in this field. In the past few years, some gesture recognition methods have been proposed to recognize meaningful expressions of hand motion, and they are relevant to information visualization, robotics, sign language understanding, medicine and healthcare area. In general, these approaches can be classified into four categories: FSMs (Finite State Machines)[1], Particle Filtering and Condensation Algorithm[2], Connectionist Approach[3] and HMMs (Hidden Markov Models)[4]. Recently, CRF model[5] is proposed in glove-based gesture recognition area, and a study in the use of reusable gestures for interaction has been reported in [6].

Because of the complexity of the problem, no single algorithm can handle all kinds of situations. Each method has its own particular view to understand and

describe the dataset. In the real world, the sensory data is usually very noisy and the hand gesture instances vary from person to person. Thus, we need to combine some different methods to take their advantages to get a better result.

Our gesture recognition research is focused on developing technologies for studying the motion and interaction with a data glove which can augment the capabilities of some users to perform some tasks. In the previous work, we use clustering ensembles to get the high confidence clusters which can be recognized as gestures. Because of the difference between different users, we need a method to reuse and update a set of defined gestures for interacting with the dynamic environment. Thus, we employ a self-training approach to renew the training set, and then improve the accuracy of recognition. Hand gesture spotting implies determining the start and end points of a meaningful gesture pattern from a continuous stream of input signals. We adopt the idea of association rule to discover some relationship between different postures, and help us to identify some unclear gestures. In this paper, we propose a hybrid method for gesture recognition. The experiment shows that the integration can be done efficiently and the performance is comparable to other classical methods.

## II. RELATED RESEARCH

### A. Self-training Methods

Self-training is an important semi-supervised learning method[8,9]. It starts with a small set of labeled data and trains the initial classifier with them. Then it uses the classifier to classify the unlabeled data. Only the instances which have high confidence are added to the labeled dataset with their predicted labels. The self-training method will retrain the classifier with the new training set, and repeat the process for several rounds. In the training procedure, the classifier teach itself by its own predictions and that improve the classification performance.

Ng and Cardie[8] use bagging and majority voting methods to implement self-training algorithm and define it as a “single-view weakly supervised algorithm”. A different definition is given by Clark et al.[9]- self-

training is a procedure in which "a tagger that is retrained on its own labeled cache on each around". Self-training plays an important role in the field of pattern recognition. In language processing area, Riloff et al. [10] uses self-training to identify subjective nouns. Maeireizo et al. [11] uses self-training to classify 'emotional dialogues' and 'non-emotional dialogues'. In this paper, we use the self-training method to rebuild the training set in order to improve the gesture recognition performance.

### B. Associative classification

In order to build an accurate classifier, the aim of classification rule mining is to discover a small subset of rules in the dataset, and usually miss some rules which are useful in some cases. In classification rules, there is only one pre-specified target, i.e. the class. But association rule mining can find all the rules in the dataset that satisfy some conditions, and its target is not pre-specified. Thus, Liu[11] propose an integrated method, called associative classification, which combine the association rule mining and classification.

Recently, some effective associative classification methods have been presented. CPAR[13] uses a greedy algorithm to generate and tests rules to avoid missing important rules. CMAR[14] extends the FP-growth method and adopts the CR-tree to store and retrieve mined association rules. MCAR[15] employs a rule ranking method to build the classifier with the high confidence rules. Many experiments showed that the associative classification method is efficient and in some cases better than some traditional classification approaches. In our proposed method, we use the relationships in the gestures to get the useful rules then use the rules to classify the uncertain instances.

## III. THE PROPOSED METHOD

### A. The framework

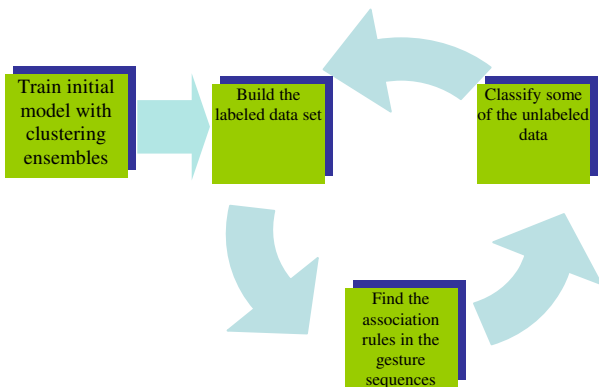


Figure 1. The framework of this hybrid method.

In order to understand the method, we describe the process with a workflow. First we use clustering ensembles to build an initial model and obtain labeled

gestures. In a second step, we make use of these to find some useful association rules in the gestures sequences. The third step is to classify some of the unlabeled data based on the rules, and then add the labeled data into the training dataset to rebuild the model.

### B. Definitions and notations

First, we give some definitions and notations here:

#### [Definition 1] Gesture

Let  $D$  be a gesture dataset,  $D=\{d_1, d_2, \dots, d_n\}$  with  $n$  instances. Each instance can be considered as a simple gesture.

#### [Definition 2] Sequence

Each sequence is composed by some gestures, it can be represented as a label list, such as:  $l_1 l_2, l_1 l_3 l_6$ .

#### [Definition 3] Cover operator ' $\prec$ '

Suppose  $Y$  is a sequences with  $m$  labels, and  $Z$  is a sequences with  $n$  labels.

$$g(Y, Z, i, j) = \begin{cases} 1 & Y_j = Z_{i+j-1} \text{ and } Y_j \neq -1 \\ 0 & Y_j \neq Z_{i+j-1} \end{cases}$$

$$\text{If } m < n \text{ then } \Delta = \max_{i=1}^n \left( \sum_{j=1}^m g(Y, Z, i, j) \right)$$

$Y \prec Z$  is called  $Y$  is covered by  $Z$  if and only if  $\Delta=m$ .

For example:  $l_1 l_2 \prec l_1 l_2 l_3 l_4, l_2 l_3 l_4 \prec l_1 l_2 l_3 l_4$

#### [Definition 4] Rule

$R=\{r_1, r_2, \dots, r_n\}$  is a rules set,  $r_i$  is a rule which is composed by two sequence patterns  $A$  and  $B$ , denoted as  $r_i=\langle A \rightarrow B \rangle$ . For rule  $r_i$ , the support of  $r_i$  denoted as  $sup(r_i)$ , and the confidence of  $r_i$  denoted as  $conf(r_i)$ .

### C. Initial model

In previous work[7], we employ the clustering ensembles to obtain the high confidence instances which can be recognized as labeled gesture. And these labeled gestures can be used to find the association rules in the next stage. Assume  $H$  is the number of partitions (i.e., ensemble size),  $k$  is the number of gesture clusters.  $L$  is a set of cluster labels denoted as  $L=(l_1, l_2, \dots, l_k)$  and  $f$  is a mapping function, it satisfies  $f(d_i) \in L$ . Assume  $\Pi=\{\pi_1, \pi_2, \dots, \pi_H\}$  is the set of  $H$  partitions,  $\pi_i = \{C_1^i, C_2^i, \dots, C_n^i\}$  ( $1 \leq i \leq H$ ), where  $C_j^i$  is the  $j^{\text{th}}$  gesture cluster in the partition  $\pi_i$ . Let  $\pi_i(d)$  be the cluster label of a gesture  $d$  in partition  $\pi_i$ .  $\theta$  is a predefined threshold of the clustering consistency index. The processing procedure is given as follow:

**Input:**  $D, H, k, \Pi, \theta, CL$

**Step 1:** Generate  $H$  different partitions on the dataset  $D$ .

**Step 2:** Match clusters in different partitions with Jaccard index method[16]

**Step 3:** For every gesture  $d$  in  $D$ , calculate  $CI(d)$  -the value of the clustering consistency index. After the clusters match, we can find that some

instances are steadily assigned to the same cluster and some instances are not. So we use the **CI** function to select the instances with the stable cluster assignment to create the high confidence clusters.

$$CI(d) = \frac{1}{H} \max \left\{ \sum_{i=1}^H \delta(\pi_i(d), l) \right\}_{l \in L},$$

$$\delta(a, b) = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{if } a \neq b \end{cases}$$

$f(d)$  shows the cluster label for  $d$ :

$$f(d) = \begin{cases} \arg \max_L \left\{ \sum_{i=1}^H \delta(\pi_i(d), l) \right\}_{l \in L}, & \text{if } CI(d) \geq \theta \\ -1, & \text{if } CI(d) < \theta \end{cases}$$

**Step 4:** For each gesture  $d$  with  $f(d) \neq -1$  in  $\mathbf{D}$ ,  
Output the value of  $f(d)$  as cluster labels of gesture  $d$ .

**Step 5:** For each gesture  $d$  with  $f(d) = -1$  in  $\mathbf{D}$ ,  
Output  $d$  as an unlabeled gesture.

#### D. Self-training with the association rule

The hand gestures can be recognized as gesture sequences. Usually, we use our hands to do a task in specified gesture order. Thus, we can import the association rule method to help us discover some interesting rules in the the gesture sequences, and then make use of the rules to classify some uncertain gestures.

Assume dataset  $\mathbf{S} = \{s_1, s_2, \dots, s_q\}$  with  $q$  sequences. The minimum support threshold is denoted **MinSupp**. The minimum confidence threshold is denoted **MinConf**.  $\varphi$  is the threshold of the difference between two sequence and  $t$  is the number of times of the self-training process. The algorithm is given as follow:

**Input:**  $\mathbf{S}, q, L, \text{MinSupp}, \text{MinConf}, \varphi, t$

**Step 1:** Generate the rule set  $\mathbf{R}$  in the dataset  $\mathbf{S}$  with the apriori method.

**Step 2:** Clean the rule set  $\mathbf{R}$ .

For each rule  $r = \langle A \rightarrow B \rangle$  in  $\mathbf{R}$ ,

1. Delete the rule if  $\text{conf}(r) < \text{MinConf}$  or  $\text{sup}(r) < \text{MinSupp}$ ,
2. Delete the rule if it contains unlabeled instances in  $A$  or  $B$ .
3. Because the hand gestures are generated in specified order, we only need the ordered association rules. Thus we should delete the rule if  $\text{order}(A) = 0$  or  $\text{order}(B) = 0$ . Assume  $x$  is a sequence,

$$\text{order}(x) = \sum_{i=1}^q b(x, s_i, i), \quad b(x, s_i, i) = \begin{cases} 1 & x \prec s_i \\ 0 & \text{otherwise} \end{cases}$$

**Step 3:** Rank the rules.

Given two rules,  $r_a$  and  $r_b$ ,  $r_a$  precedes  $r_b$  if :

1.  $\text{conf}(r_a) > \text{conf}(r_b)$

2.  $\text{conf}(r_a) = \text{conf}(r_b)$ , but  $\text{sup}(r_a) > \text{sup}(r_b)$

3.  $\text{conf}(r_a) = \text{conf}(r_b)$  and  $\text{sup}(r_a) = \text{sup}(r_b)$ , but  $r_a$  was generated before  $r_b$ .

**Step 4:** Identify some unlabeled gestures.

If we can find  $A$  pattern in a gesture sequence, we may also find  $B$  in the same sequence based on the high confidence rules. But sometimes the  $B$  sequence is mixed with some unlabeled instances in the gesture sequence, so that we can classify these instances with the  $B$  sequence.

For each rule  $r = \langle A \rightarrow B \rangle$  in  $\mathbf{R}$ ,

For  $i=1$  to  $q$  do

If  $A \prec r_i$  and not( $B \prec r_i$ ) then

get the sub-sequence  $Sb$  behind  $A$  with the same length (if it has enough length) in  $s_i$ , and set  $\text{diff}=0$ ,  $\text{err}=\text{false}$ .

For  $j$  for 1 to  $|B|$  do

If  $Sb_j \neq B_j$  then  $\text{diff}++$

If  $Sb_j = -1$  then  $\text{err}=\text{true}$ ;

End;

If  $\text{diff}/|B| < \varphi$  and  $\text{err}=\text{false}$  then

For  $j$  for 1 to  $|B|$  do

If  $Sb_j = -1$  then  $Sb_j = B_j$ . (Give the class labels to relevant instances)

End;

End;

**Step 5:** 1. Generate the new dataset  $\mathbf{S}$ .

2. Repeat the process for  $t$  times.

**Step 6:** Classify  $S_i$  with nearest neighbour method

## IV. EXPERIMENTS AND EVALUATION

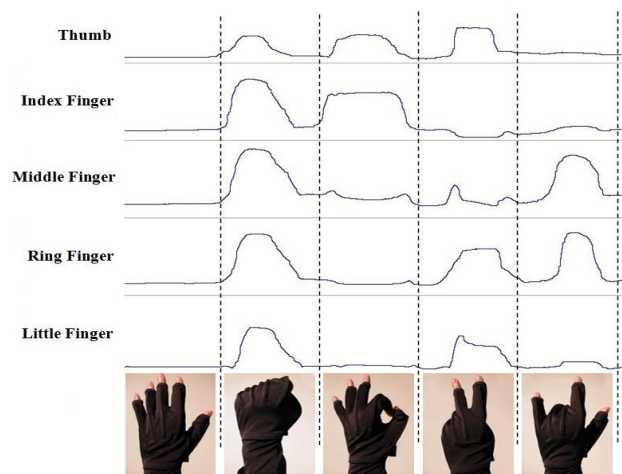


Figure 2. The 5 gestures and the data.

In the real world, it is difficult to obtain perfect data from the data glove. Gesture data generated by a 5DT data glove in Figure 2 shows five gestures which are represented as waves. The data glove receives the flexure signals from the 5 fingers sensors when we bend the fingers. We can see that sometimes the signals are

not clear, because of the disturbance from the movements of other fingers. In some cases, it could be considered as noisy data or uncertain instances in the previous algorithm.

The Australian Sign Language dataset[17] is used to test the method. The gesture data was collected from a native Auslan signer using the data gloves. In the experiment, several subsets are extracted from the database. Each subset contains some word classes, each word class has 27 instances. We used the 10-fold cross validation method to compare the classification accuracy of this method with HMMs[4] and DTW[18]. We used k-means to generate multiple partitions for the initial model with random initialization of cluster centers. Parameter  $t$  is the number of self-training times. The numbers with bold-font show the best classification accuracy among the three methods.

TABLE I. A COMPARISON OF THREE METHODS

Size of subset	Classification Accuracy(%)					
	HMMs	DTW	Hybrid method			
			t=0	t=1	t=2	t=3
5	95.6	<b>96.3</b>	79.3	80.7	80.7	80.7
10	90.4	<b>91.9</b>	75.2	76.3	75.2	75.2
15	<b>85.9</b>	84.9	74.1	79.0	76.8	76.8
20	81.9	83.7	75.9	78.9	<b>84.8</b>	80.0
25	74.5	77.3	70.1	<b>79.9</b>	72.9	70.1
30	71.0	74.1	66.9	72.0	<b>76.3</b>	74.3

The experimental results show that the proposed method can improve the performance after self-training. It doesn't perform well when the size of subset is small, because there is not enough rules for it to retrain itself. But in the larger subsets, the hybrid method is better than the other two methods, and we notice that, the accuracy may be reduced when the number of self-training times is increased.

## V. CONCLUSIONS

This paper proposed a hybrid method for hand gesture recognition. This algorithm integrate the ideas of associative classification and self-training approaches to find the useful rules in the gesture sequences, and then improve the performance. However, the hybrid method has high computing cost because of the pre-modelling and self-training for several rounds, which need to be improved in our future work.

## References

[1] P. Hong, M. Turk and T. S. Huang, "Gesture modeling and recognition using finite state machines," 4th IEEE Int. Conf. Autom. Face Gesture Recog., Grenoble, France, pp. 410-415, Mar. 2000.

[2] C. Kwok, D. Fox and M. Meila, "Real-time particle filters," Proc. IEEE, vol. 92, no. 3, pp. 469-484, Mar. 2004.

[3] M. S. Yang and N. Ahuja, "Recognizing hand gesture using motion trajectories," IEEE CS Conf. Comput. Vis. Pattern Recog., vol. 1, pp. 466-472, Jun (1998)

[4] M. Elmezain, A. Al-Hamadi, J. Appenrodt and B. Michaelis, "A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory," International Conference on Pattern Recognition (ICPR), pp. 519-522, 2008.

[5] H. Yang, S. Sclaroff, S. Lee, "Sign Language Spotting with a Threshold Model Based on Conditional Random Fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1264-1277, July. 2009.

[6] R. Vatavu, "Reusable Gestures for Interacting with Ambient Displays in Unfamiliar Environments," Advances in Intelligent and Soft Computing, pp. 157-164, 2011.

[7] Y. Huang, D. Monekosso and H. Wang, "Clustering Ensembles Based on Multi-classifier Fusion," Intelligent Computing and Intelligent Systems (ICIS), pp.393-397, 2010.

[8] Ng and C. Cardie, "Weakly supervised natural language learning without redundant views," Human Language Technology Conference of the North American, Edmonton, Canada, May .2003.

[9] S. Clark, J. R. Curran and M. Osborne, "Bootstrapping POS taggers using unlabelled data," Walter Daelemans and Miles Osborne, editors, Proceedings of CoNLL-2003, pages 49-55. Edmonton, Canada, 2003.

[10] E. Riloff, J. Wiebe and T. Wilson, "Learning subjective nouns using extraction pattern bootstrapping," the Seventh Conference on Natural Language Learning, 2003.

[11] B. Maeireizo, D. Litman and R. Hwa, "Co-training for predicting emotions with spoken dialogue data," The Companion Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), 2004.

[12] B. Liu, W. Hsu and Y. Ma, "Integrating Classification and association rule mining," KDD'98, New York, NY, Aug. 1998.

[13] X. Yin, J. Han, "CPAR: classification based on predictive association rules," Proceeding of the 2003 SIAM international conference on data mining (SDM'03), San Francisco, CA, pp 331-335, 2003.

[14] W. Li, J. Han and J. Pei, "CMAR: Accurate and efficient classification based on multiple-class association rule," ICDM'OI, pp. 369-376, San Jose, CA, 2001.

[15] F. Thabtah, P. Cowling and Y. Peng, "MCAR: multi-class classification based on association rule," Proceedings of the ACS/IEEE 2005 International Conference on Computer Systems and Applications, p.33-I, January 03-06, 2005.

[16] A. Topchy, B. Minaei-Bidgoli, A.K. Jain and W.F. Punch, "Adaptive Clustering Ensembles," Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), vol. 1, pp. 272-275, 2004.

[17] UCI repository:  
<http://kdd.ics.uci.edu/databases/auslan2/auslan.data.html>

[18] G.A. ten Holt, M. Reinders and E. Hendriks, "Multi-dimensional dynamic time warping for gesture recognition," Thirteenth annual conference of the Advanced School for Computing and Imaging, 2007.