



# Early Recognition of Handwritten Gestures based on Multi-classifier Reject Option

Zhaoxin Chen, Harold Mouchère, Eric Anquetil, Christian Viard-Gaudin

## ► To cite this version:

Zhaoxin Chen, Harold Mouchère, Eric Anquetil, Christian Viard-Gaudin. Early Recognition of Handwritten Gestures based on Multi-classifier Reject Option. 14th IAPR International Conference on Document Analysis and Recognition (ICDAR2017), Nov 2017, Kyoto, Japan. 10.1109/ICDAR.2017.43 . hal-01653154

**HAL Id: hal-01653154**

**<https://hal.science/hal-01653154>**

Submitted on 5 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Early Recognition of Handwritten Gestures based on Multi-classifier Reject Option

Zhaoxin Chen  
and Eric Anquetil  
INSA de Rennes  
Rennes, France 35043  
Email: firstname.lastname@irisa.fr

Harold Mouchère  
and Christian Viard-Gaudin  
University of Nantes, LS2N  
Nantes, France 44300  
Email: firstname.lastname@univ-nantes.fr

**Abstract**—In this paper a multi-classifier method for early recognition of handwritten gesture is presented. Unlike the other works which study the early recognition problem related to the time, we propose to make the recognition according to the quantity of incremental drawing of handwritten gestures. We train a segment length based multi-classifier for the task of recognizing the handwritten touch gesture as early as possible. To deal with potential similar parts at the beginning of different gestures, we introduce a reject option to postpone the decision until ambiguity persists. We report results on two freely available datasets: MGSet and ILG. These results demonstrate the improvement we obtained by using the proposed reject option for the early recognition of handwritten gestures.

## I. INTRODUCTION

Generally, an online gesture recognition system produces its result just after the user has completed his/her gesture. However, some applications need to predict the intention of users without waiting for the completion of their gestures. For example, in [1], [2] handwriting gestures are used for both abstract actions (e.g. shortcut for command or character input) and direct manipulation (drag, rotation, etc.). The co-existence of these two usages requires a feedback as soon as possible to be consistent with a direct manipulation. Hence, an Early Recognition (ER) strategy is desirable to cope with these two kinds of commands.

ER works have been developed for motion prediction problems [3], [4]. A basic idea is to employ a partial matching method, where the recognition result of an input pattern is determined by the matching distance of its beginning part from reference patterns. The Dynamic Time Warping (DTW) algorithm is a widely used method to search for an optimal partial alignment. Another approach is combination of classifiers  $\{h_1, \dots, h_t, \dots, h_T\}$  [5], [6], where  $h_t$  is a weak frame classifier at  $t^{th}$  frame (i.e., time  $t$ ). The recognition result at the  $t^{th}$  frame will be determined by combining  $t$  recognition results provided by  $\{h_1, \dots, h_t\}$ . This method assumes the input pattern having the same performing speed as the reference pattern.

Unlike the previous works that investigate the early recognition based on time frame, we believe that the time frame does not represent the motion information. We study in this work early recognition for handwritten touch gestures and do not make any assumption regarding the size and the speed of these gestures. In other words, a handwritten gesture may vary from

writing speed and gesture size. Gesture which moves very slowly contains less information frame by frame. We propose that the early recognition should be investigated based on difference of the motion rather than difference of the time. A difficulty is the gesture size normalization. In training process, usually a classification system normalizes each gesture to a fixed size bounding box before feature extraction so that the value of features can be unified to a fixed scale. However, in early recognition it is difficult to normalize the early part of a gesture without knowing the size of the full gesture. Fig. 1 shows an example of this problem. Fig. 1(a) is a normalized gesture which is assumed to be a template. Fig. 1(b) (c) are two unknown gestures with different sizes. As displayed on this figure, without normalization, considering a partial trajectory with a length  $l$ , it could be a small part of a large gesture or a large part of a small gesture. Therefore, the problem to solve corresponds to achieve a partial matching in a size free context.

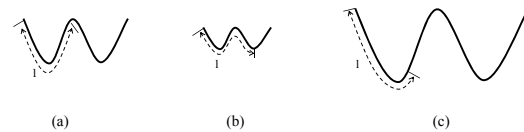


Fig. 1. (a) A normalized gesture as a template. (b) (c) In a size free context, due to the input gestures having a variety of sizes, a trajectory with a length of  $l$  may cover different parts of a same type gesture.

On the other hand, as shown in Fig.2, it is a very general case that two gestures have similar shape in their beginning part. An ER system should be able to reject for ambiguous recognition and wait for enough information to make a decision. To the best of our knowledge, this problem has few discussion in the related works.

In brief, we propose to build an early recognition system being able to deal with ambiguous common parts under free drawing context for handwriting gesture recognition. We control the progress of the gesture using its length instead of time duration. We propose a multi-classifier system to deal with the early part normalization problem. A reject option is involved to postpone the decision until enough confidence is achieved. The rest of the paper is organized as follows.

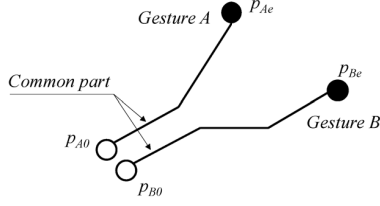


Fig. 2. Ambiguity of two gestures for early recognition problem.  $p_{A0}$  and  $p_{B0}$  are the finger down points.  $p_{Ae}$  and  $p_{Be}$  are the finger up points. The two gestures have a common beginning part.

Section II introduces the related work of handwritten gesture recognition and early recognition. Then section III presents the structure of the multi-classifier method and the reject option strategy. Next in section IV, we report the experimental result to show the earliness and accuracy of the system. Experiments are conducted on two freely available dataset **ILG** [7] and **MGSet** [8]<sup>1</sup>. The ILG dataset contains common mono-touch gestures which are assumed for abstract command while the MGSet contains special multi-touch gestures which can be both used for abstract command and direct manipulation. Finally, we conclude this work and discuss the perspectives.

## II. RELATED WORK

A typical partial matching method can be found in the work of Mori *et al.* [3], they use DTW to recognize a set of 18 categories of hand postures in a video stream. The partial matching is achieved by traversing all time frames of the reference pattern to find the minimum distance. In [4], they propose to compare the posture at each individual time  $t$  rather than comparing the posture sequence. They try to solve the ambiguous beginning part by calculating the distance gap between the most and second most similar gestures, i.e. the result is determined when the difference of the first two best classes is larger than a threshold.

In [6] a multi-classifier method is proposed to take early decision. Each frame classifier  $h_t$  is trained using the feature vectors of the  $t$ th frame. The early recognition result at an arbitrary frame  $t$  is obtained from a weighted combination of the frame classifier  $h_0, \dots, h_t$ . However, this method assumes that all the training patterns have the same time length  $T$ . Su *et al.* [9] also use boost algorithms to make early recognition for facial expression. A traditional boosting technique is applied to select a set of weak classifiers. Then a weight propagation is adopted to temporally combine the classifiers into a strong classifier. More works on boost classifier can be found in [5].

In another way, Weber *et al.* [10] apply Long Short Term Memory (LSTM) for the task of recognizing motion templates from body as early as possible. They achieve 92% accuracy rate with average earliness less than one second (24.7 frames).

Note that most of the works mentioned above deal with the ER problem for video stream. There are few ER works on on-line 2D handwritten gesture whose signal is captured

on touch screen. Since the length of on-line 2D trajectory is measurable, we propose that the ER for handwritten gesture should be investigated based on difference of motion rather than difference of time. We aim to give a generic framework of multi-classifier early recognition which can be implemented with different kind of classifiers.

## III. MULTI-CLASSIFIER EARLY RECOGNITION

To deal with the size normalization problem for ER, we propose a multi-classifier recognition system as shown in Fig. 3. The global gestures are divided into several parts, and each classifier is trained to recognize a different sub-part, representing incomplete gestures. However a trajectory with a length  $l$  could be a small part of a large gesture or a large part of small gesture. In other words, a coming incomplete gesture (i.e. an incomplete trajectory) can not be explicitly sent to a certain classifier according to its length because we can not estimate the size of its potential full gesture. Therefore, in the recognition stage, and unlike what has been done in the training stage, the successive incomplete trajectories will be processed by all classifiers. The recognition result is determined by a fusion of the results from all classifiers. Meanwhile, each classifier incorporates two reject options: *ambiguity rejection* and *outlier rejection*. The former one aims at dealing with the common beginning parts. It avoids to make the decision if a gesture is similar to another one. The latter one rejects the gestures which are out of its scope so that it would let other classifiers make the decision. The classifier training will be detailed in section II.A. The rejection algorithm and fusion of classifiers will be proposed in sections II.B and C.

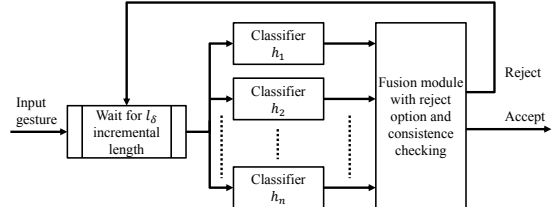


Fig. 3. The structure of our proposed multi-classifier early recognition system.

### A. Segment Classifier

Consider a set of  $N$  training gestures  $x_i | i = 1, \dots, N$ , each gesture  $x_i$  is a sequence of points  $x = p_0, \dots, p_e$  normalized and centered in the unit square bounding box. As we discussed in section I, users may perform a same gesture at different speeds. In other words, during a fixed time interval  $\Delta t$  the length of the gestures performed by different users may be different. We consider a curvilinear distance segmentation rather than a time segmentation, i.e. a gesture will be early recognized every  $\Delta l$  length of its trajectory.

As displayed in Fig.3, early recognition is realized by combining classifiers  $h_1, \dots, h_n$ . Since we control early recognition with a length information, each classifier is trained

<sup>1</sup><https://www-intuidoc.irisa.fr/en/category/bases-de-donnees/>

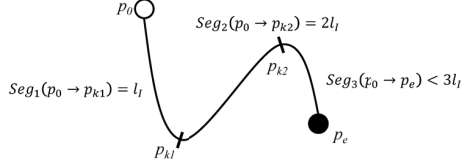


Fig. 4. Trajectory of an example gesture.  $p_0$  and  $p_e$  are the starting and ending point, respectively.  $p_{k1}$  is the keypoint where the length of  $seg_1$  (from  $p_0$  to  $p_{k1}$ ) is  $l_I$ .  $p_{k2}$  represents the point at  $2l_I$ . Since the total length is less than  $3l_I$ , this trajectory will offer three segments for training.

with different lengths of the gestures. Thus, we first segment the training gestures under a fixed length interval  $l_I$ . Fig. 4 illustrates the segmentation of a gesture.

From all the training gestures, we build  $n$  segment sets  $S_i$ , with  $i = 1, \dots, n$ , where each  $S_i$  set represents all the segments of  $Seg_i$  whose length is equal or less than  $il_I$ . Since the training gestures are not assumed to have the same length, the number of training gestures in different  $S_i$  may be different.  $S_1$  will always cover the beginning part of all the training gestures, while  $S_i$  only contains the gestures which are longer than  $(i-1)l_I$ . We denote the segment classifiers as  $\{h_1, \dots, h_i, \dots, h_n\}$ , each classifier  $h_i$  is trained by the feature vectors of the  $i$ th segment set  $S_i$ .

In the recognition step, as we explained in Fig. 1 an arbitrary input gesture  $x$  at length  $l$  ( $(i-1)l_I < l < il_I$ ) can not be specifically recognized by classifier  $h_i$  because of the size free context. Therefore, an arbitrary gesture  $x$  should be processed by all the classifiers and determined by the one giving the highest probability value. Let  $h_j(x, c_i)$  be the probability of the best class  $c_i$  obtained by the classifier  $h_j$ , the result of multi-classifier is

$$H(x, c_i) = \max_{j=1, \dots, n} h_j(x, c_i). \quad (1)$$

### B. Rejection Algorithm

Referring to the work in [11] [12], our reject option is designed from two aspects: ambiguity and outlier. Fig.5 illustrates reject option boundary based on a classification space. The ambiguity refers to the patterns which are near the pair-wise classification hyperplane. These patterns reflects the common part ambiguity as depicted in Fig. 2. The outlier refers to the patterns which are far away from the training samples. Because of the size free context that an input pattern can not be specified to a certain classifier, the outlier rejection is used by each classifier to explicitly reject the pattern which does not belong to the scope of this classifier.

1) *Ambiguity rejection*: We deduce from Fig.5 that a good ambiguity rejection solution is to define a class-pair dependent threshold which rejects the ambiguous patterns close to the pair-wise hyperplane. However, it is inefficient to maintain the entire class pair space when the class number is large. A trade-off is to use class dependent threshold that defines an ambiguity boundary against all the other classes. We define as in [12] the reliability function  $\psi_i^{Amb}$  to well interpret the

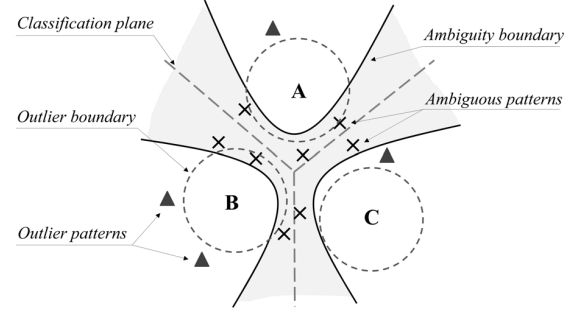


Fig. 5. Ambiguous patterns and outlier patterns in multi-class recognition rejection problem. The dotted straight lines represent the pair-wise hyperplanes to separate two classes. The curves are class-wise ambiguity rejection boundary. The dot circles are class-wise outlier boundary.

ambiguity condition. The ambiguity determines if a shape is near the decision boundaries. So let  $S = (s_1, s_2, \dots, s_n)$  be the confidence for each class given by the classifier. We can use the difference between the best class  $C_1$  and the second one  $C_2$  to form the reliability function as:

$$\psi_i^{Amb} = \frac{s_i - s_j}{s_i}, \quad (2)$$

where  $i = C_1$ ,  $j = C_2$ . Thus, the rejection decision is:

$$r_i^{Amb} = \psi_i^{Amb} < T_i^{Amb}, \quad (3)$$

where  $T_i^{Amb}$  is the class dependent threshold.

2) *Outlier rejection*: Since each classifier is trained by a set of segments of a certain length, a classifier should be able to reject a gesture which is not similar to any of the training data. It ensures that in the multi-classifier structure, only the relative classifier will give response to an input gesture while others would make rejection.

Shown in Fig.5, the outlier samples locate far from the center of each class. Therefore, the most important information for this rejection option is the intrinsic description of the learned data. Depending of the used classifier, this information is not always directly available [12]. If the classifier outputs approximate the density probability of the learned data as in classifiers like RBFNN or Mixture of Gaussian, then the reliability function and reject decision can be defined as:

$$\psi_i^{Out} = s_i, \quad r_i^{Out} = \psi_i^{Out} < T_i^{Out}, \quad (4)$$

where  $s_i$  is the output score of the best class.

3) *Threshold optimization*: We firstly define some notations to better explain the result of the reject option. Considering a set of  $N$  training samples, Table I shows the notations to represent the number of samples in different conditions after recognition and reject options.

To evaluate the threshold, we compute the False Accept Rate (FAR), and False Reject Rate (FRR) as:

$$FAR = \frac{N_A^F}{N} \quad FRR = \frac{N_R^F}{N} \quad (5)$$

TABLE I  
NOTATIONS TO REPRESENT THE NUMBER OF SAMPLES IN DIFFERENT  
CONDITIONS. WITH THESE NOTATIONS:  
 $N = N_A + N_R = N_{cor} + N_{err} = N_A^T + N_A^F + N_R^F + N_R^T$

Sample set ( $N$ )	Reject option	
	Accept ( $N_A$ )	Reject ( $N_R$ )
Correctly classified ( $N_{cor}$ )	True Accept ( $N_A^T$ )	False Reject ( $N_R^F$ )
Mis-classified ( $N_{err}$ )	False Accept ( $N_A^F$ )	True Reject ( $N_R^T$ )

For ambiguity rejection,  $N_R^F$  are the training samples which are correctly classified but wrongly rejected by reject option while the  $N_A^F$  are wrongly classified but accepted. Note that it is better to prepare a validation dataset since the high precision in training data makes  $N_A^F$  close to 0, which leads to unavailable optimization. For outlier rejection, since each classifier is trained with a set of segments in certain length, the positive samples are the classifier's training set while the negative samples are the training sets for other classifiers. The acceptance of negative samples will be count for  $N_A^F$  and the rejection of positive samples will be  $N_R^F$ . In rejection, the aim is to obtain the lowest error rate while rejecting least correct results. Intuitively, the optimization of the threshold is to find a trade-off between the  $FAR$  and  $FRR$ . Therefore, the optimal threshold for class  $i$  is defined as:

$$T_i^{opt} = \arg \min_{T_i} \sqrt{\alpha_e FAR_i^2(T_i) + \alpha_r FRR_i^2(T_i)} \quad (6)$$

where the weights  $\alpha_e$  and  $\alpha_r$  are used to balance the impact of each rate. In general case, these parameters are set to 1. Since we use the class-wise threshold, we measure the  $FAR_i$  and  $FRR_i$  based on each class  $i$  to learn the threshold. The two thresholds are learned independently.

Finally, the rejection of an input gesture is made if it is rejected by either reject option.

$$r = \max(r^{Amb}, r^{Out}) \quad (7)$$

If the input gesture is accepted, the probability  $h_j(x, c_i)$  of the class  $i$ , as shown in (1), is the conjunction of both reliability function:

$$h_j(x, c_i) = \psi_i^{Amb} * \psi_i^{Out} \quad (8)$$

With the reject option, the equation (1) will be changed to

$$H(x, c_i) = \begin{cases} \text{Reject,} & \text{if } \prod_j r_{j,i}(x) = 1 \\ \max_j (\bar{r}_{j,i}(x) * h_j(x, c_i)), & \text{otherwise} \end{cases} \quad (9)$$

#### IV. EXPERIMENTAL RESULT

The evaluation experiment has been conducted on the **MGSet** and **ILG** datasets. The **MGSet** is a dataset of multi-touch gestures. This dataset contains 45 different multi-touch

gestures collected from 33 users. 3589 gestures are for training and 2259 for testing. This dataset assumes that multi-touch gestures can both be used for shortcut command and object manipulation. The early recognition problem is significant on this dataset since system should give the feedback for object manipulation gesture as soon as possible. A basic recognition result has been reported in [8].

The **ILG** dataset [7] is a collection of mono-stroke pen-based gestures. 38 users were asked to perform gestures for triggering 21 different commands in a simulated image edition software. The dataset is partitioned to 3 groups. Note that the first two groups contain user-defined gestures (user is allowed to design own gestures to trigger commands, see the details in [7]). Obviously, this two groups can only be used for writer-dependent experiments since it only offers few training samples for each class. To obtain a more general result, we select the third group (1926 samples, 693 for training and 1233 for testing) which has more classical properties: users all performed the same 21 gestures. Note that for both datasets, we partition 20% of the data from the training set as a validation set to learn the ambiguity threshold. Fig. 6 shows some examples from the two datasets. For both datasets, we partition 20% of the data from the training set as a validation set to learn the ambiguity threshold.

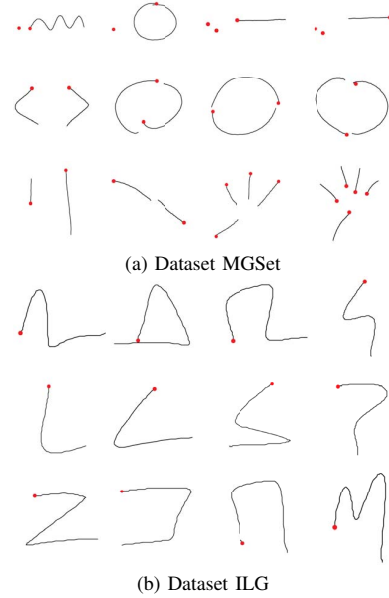


Fig. 6. Examples of the gestures in MGSet and ILG dataset.

The classifier we used for each independent classifier is a Graph + LibSVM described in [8]. The confidence scores for ambiguity threshold learning are the probabilities from LibSVM. For outlier threshold, we use clustering algorithm to find three centers for each class, and compute the distances of an input gesture to the centers. The minimum distance is used as  $s_i$  in (4) to learn the outlier threshold. These can be replaced by any classifiers which give output confidence score for each class. The number of classifiers is set to 3. We firstly evaluate

the early recognition with regarding to the different lengths of input gestures. As we illustrated above, in the training process classifiers are trained by normalized gestures. But in testing step, since there is no clue to obtain their real size before ending, early recognition has to be applied on their original size. Fig.7 shows the gesture's original length distribution for the test set. Each gesture is recognized on every 50 pixels of its incremental length. Referring to the notations in table I, we measure the False Accept Rate ( $FAR = N_A^F/N$ ) and Reject Rate ( $RR = N_R/N$ ) when using the reject option and compare them with the traditional Error Rate ( $ER = N_{err}/N$ ) without reject strategy. The recognition results according to different lengths are shown in Fig. 8.

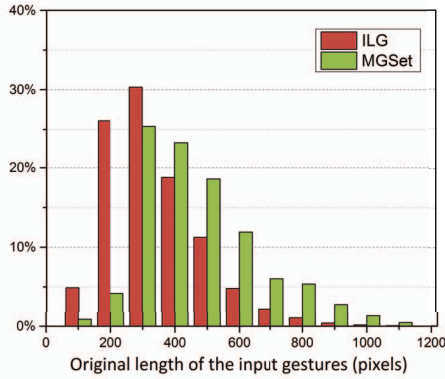
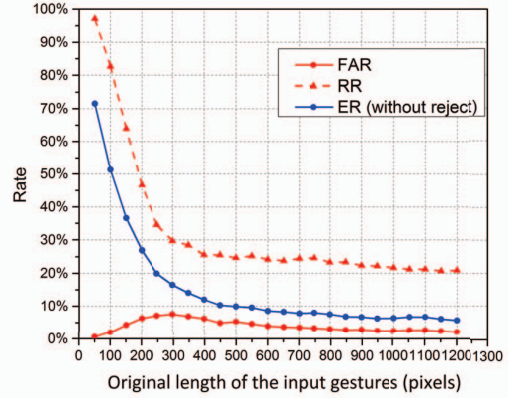


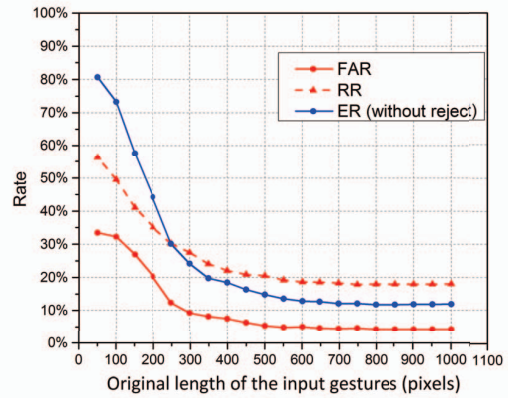
Fig. 7. Histogram for the distribution of test gestures according to their length on the two datasets.

Both results show that without the rejection algorithm, the ER is very high at the beginning since gestures are still ambiguous to take a distinction. Accordingly, the rejection algorithm is effective to reject most of the gestures at beginning. The RR decreases along with the decreasing of ER (without reject) which means that it well rejects the ambiguous gestures but accept the gesture as soon as it has enough distinctive information. This strategy leads to a good performance of FAR which is very low at the beginning and always lower than ER at any input length. Meanwhile, the RR is always higher than ER in the ending part, which means that some correctly classified gestures are rejected. This is the negative effect of the reject option; a low error rate is obtained at the cost of a high reject rate.

In operational use case, a reasonable strategy to prevent noisy decisions consists in filtering the decision by considering several consecutive outputs of the classifier. Consequently, a decision is finally accepted when the classifier gives  $t$  consecutive times the same output. Results on the two datasets are shown in Table II for  $t = 1$  to 6. Referring to the notation of Table I, TAR is True Accept Rate ( $TAR = N_A^T/N$ ), FAR is False Accept Rate, RR is Reject Rate which represent the percentage of gestures which are rejected at every length until their completion, CR is correct rate ( $CR = N_{cor}/N$ ). The



(a) Dataset MGSet



(b) Dataset ILG

Fig. 8. Recognition results with respect to the length of the input gesture on two datasets. FAR and RR are obtained using the reject option while ER is the traditional mis-classified rate.

Avg. length percent indicates the earliness of recognition that the decision is made at this percentage before the completion of the gesture. Since the ILG data does not contain the time label, the average decision time (Avg.T) is only measured on MGSet.

It shows an acceptable result on MGSet dataset that the accuracy rate of first time decision is 81.89% which is obtained with an average of 37.04% length of gestures. Comparing to the third result with no reject option, where the decision is also achieved around 33.98% length by 3 consecutive same results, the FAR is less than half of the ER. With the increasing of the time for consistence checking, the decision is postponed to obtain less errors. The FAR decreases from 14.56% to 3.41% while the RR increases from 3.54% to 19.17%. It indicates that we have to find a trade-off between the error rate and the reject rate. The result by  $t = 2$  may be considered as an acceptable one where the FAR is 10.85% and RR is 5.71%. The comparable result from no reject option is shown at  $t = 4$  where the CR is 79.59% (3.85% lower than TAR:83.44%) and ER is 20.41% (9.56% higher than FAR:10.85%). In other words, the reject options minimize the error rate by offering



TABLE II  
RECOGNITION RATE WITH CONSISTENCE CHECKING.

Dataset	t	Reject Option					No Reject Option			
		TAR	FAR	RR	Avg. length percent	Avg. T(ms)	CR	ER	Avg. length percent	Avg. T(ms)
MGSet	1	81.89%	14.56%	3.54%	37.04%	456.21	24.88%	75.15%	8.13%	297.23
	2	83.44%	10.85%	5.71%	46.82%	523.34	48.78%	51.22%	21.32%	368.07
	3	<b>82.38%</b>	<b>8.85%</b>	<b>8.77%</b>	<b>55.89%</b>	<b>591.33</b>	67.60%	32.40%	33.98%	437.85
	4	82.20%	6.06%	11.73%	66.16%	669.86	79.59%	20.36%	45.44%	518.21
	5	80.35%	4.60%	15.05%	71.03%	738.17	<b>85.83%</b>	<b>13.72%</b>	<b>54.93%</b>	<b>598.04</b>
	6	77.42%	3.41%	19.17%	77.54%	811.38	88.62%	10.00%	62.34%	660.90
ILG	1	30.65%	67.15%	2.20%	34.81%	N/A	21.22%	78.78%	18.03%	N/A
	2	64.15%	26.42%	9.43%	75.53%	N/A	42.85%	57.15%	56.17%	N/A
	3	73.98%	11.22%	14.80%	92.24%	N/A	68.29%	31.71%	82.16%	N/A
	4	77.72%	6.26%	16.02%	97.62%	N/A	79.51%	20.49%	92.67%	N/A
	5	77.80%	4.88%	17.32%	99.19%	N/A	85.45%	14.55%	97.27%	N/A
	6	77.72%	4.55%	17.72%	99.68%	N/A	87.40%	12.44%	99.08%	N/A

reject Although there are 5.71% samples are rejected during the recognition, we believe that in a real practice it is better to reject an input and provide some ambiguous options to select than giving a wrong result. By this way, user only need to make a selection instead of removing the wrong input and re-draw it again. However, the result on ILG shows not as good as MGSet. The accuracy rate is only 30.65% for the first decision. From the Fig. 8(b), the FAR is around 20% to 30% from 50 to 200 pixels. Decisions made on this stage cause much more errors than MGSet. Therefore, the first time decision may not be acceptable in this situation. The accuracy rate on ILG dataset shows a great improvement using consistence checking. With  $t = 2$ , the TAR is 33.5% higher than  $t = 1$  while the FAR decreases 52.73% comparing to  $t = 1$ . A higher time of consistence checking seems not useful since the corresponding Avg. length is over 90%.

## V. CONCLUSION

We propose a reject option based multi-classifier system for handwritten gesture early recognition. The reject option is designed to deal with the ambiguous gestures which have a similar beginning part. We believe that the intrinsic value of early recognition is to recognized a pattern with few information. Therefore, we study the early recognition based on the input length instead of the input time. The multi-classifier are trained with different lengths of segments and fused with reject option to find the optimal result. The experiment gives a promising result on MGSet. The system achieve 83.44% accuracy rate with 46.82% average length of input gestures. Comparing to the no reject option system, the error rate is very low at the beginning part which proves that our reject algorithm works well to reject the ambiguous gestures.

Our future work will first focus on investigating the automatic selection of the optimal number of classifiers and segment length for training instead of using empirical selection as in our experiment. We will also try to use boost ing as in

previous studies [6] [5] for our multi-classifier combining with the reject option.

## REFERENCES

- [1] E. Petit and C. Maldivi, "Unifying gestures and direct manipulation in touchscreen interfaces," Dec. 2013.
- [2] G. Kurtenbach and W. Buxton, "Issues in combining marking and direct manipulation techniques," in *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '91. New York, NY, USA: ACM, 1991, pp. 137–144.
- [3] A. Mori, S. Uchida, R. Kurazume, R. Taniguchi, T. Hasegawa, and H. Sakoe, "Early recognition and prediction of gestures," in *18th Intern. Conf. on Pattern Recognition (ICPR)*, vol. 3, 2006, pp. 560–563.
- [4] M. Kawashima, A. Shimada, H. Nagahara, and R. I. Taniguchi, "Adaptive template method for early recognition of gestures," in *17th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, Feb 2011, pp. 1–6.
- [5] K. Ishiguro, H. Sawada, and H. Sakano, "Multi-class boosting for early classification of sequences," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2010, pp. 24.1–24.10.
- [6] S. Uchida and K. Amamoto, "Early recognition of sequential patterns by classifier combination," in *19th International Conference on Pattern Recognition (ICPR)*, Dec 2008, pp. 1–4.
- [7] N. Renau-Ferrer, P. Li, A. Delaye, and E. Anquetil, "The ILGDB database of realistic pen-based gestural commands," in *21st Intern. Conf. on Pattern Recognition (ICPR)*, Tsukuba, Japan, 2012, pp. 3741–3744.
- [8] Z. Chen, E. Anquetil, H. Mouchère, and C. Viard-Gaudin, "Recognize multi-touch gestures by graph modeling and matching," in *17th Biennial Conference of the International Graphonomics Society*, Pointe-à-Pitre, France, Jun. 2015.
- [9] L. Su and Y. Sato, "Early facial expression recognition using early rankboost," in *10th IEEE Intern. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, April 2013, pp. 1–7.
- [10] M. Weber, M. Liwicki, D. Stricker, C. Scholzel, and S. Uchida, "Lstm-based early recognition of motion patterns," in *22nd International Conference on Pattern Recognition (ICPR)*, Aug 2014, pp. 3552–3557.
- [11] Y. Zhu, J. Sun, Y. Hotta, and S. Naoi, "Rejection optimization based on threshold mapping for offline handwritten chinese character recognition," in *ICFHR*. IEEE Computer Society, 2010, pp. 72–77.
- [12] H. Mouchère and E. Anquetil, "A unified strategy to deal with different natures of reject," in *18th International Conference on Pattern Recognition (ICPR)*, vol. 2, 2006, pp. 792–795.