

Combining Local and Global Features for Offline Handwriting Recognition

Rhandley D. Cajote* and Rowena Cristina L. Guevara

*Digital Signal Processing Lab
Department of Electrical and Electronics Engineering
University of the Philippines Diliman
Quezon City 1101 PHILIPPINES*

ABSTRACT

The result of investigating the performance of handwriting recognition systems using local features, global features and a combination of local and global features is presented. The global features are derived from the shape of the word contour and the local features are derived from the geometric characteristics of the word segments. The global features utilize the centroidal distance of the word contour also known as the polar-radii graph or PRG. The system was trained and tested using the demo version of the publicly available IAM database. Using the local features alone and an HMM recognizer a recognition rate of 58% was achieved using a 20-word vocabulary. Using the PRG of the word contour as a global feature and an MLP classifier a recognition rate of 78% is achieved. The PRG is then combined with the local features using a combined probabilistic framework and the hybrid handwritten word recognition system achieved a recognition rate of 72%.

Keywords: HMM, MLP, PRG, Offline Handwriting Recognition

1. INTRODUCTION

The field of handwriting recognition can be broadly classified into two: on-line and offline recognition. Online handwriting recognition deals with the problem of processing the handwriting while it is being written on a digitizing tablet. There is a common consensus among researchers that online handwriting recognition is a much easier task than the offline case [1] because more information is available to the recognizer. The position, velocity or acceleration of the pen tip as a function of time and the pen trajectory provides more information about the nature of the handwriting.

In offline handwriting recognition, only the image of the handwritten word to be recognized is available. A combination of image processing and pattern recognition algorithms must be employed to solve the problem. In general, offline handwriting recognition is a more difficult problem than the online case [2].

Traditional methods of solving handwritten word recognition (HWR) problems require two different approaches. The first approach, often called the analytic approach, segments the

*Correspondence to: Digital Signal Processing Lab, Department of Electrical and Electronics Engineering, University of the Philippines Diliman, Quezon City 1101 PHILIPPINES. email:rhandley@eee.upd.edu.ph

words into smaller subunits. The subunits can be characters or image segments smaller than the characters. The characters that comprise the word are recognized to come-up with the word-level interpretation of the handwritten word. This is also called the bottom-up approach because recognition starts from the character level then goes up to the word level. The second approach, called the global approach, treats the words as a single entity. The word is recognized based on the features extracted from the word as a whole ([2], [3]).

The analytic approach works well for applications using a large and dynamic dictionary or lexicon, such as bank checks and postal mail applications. The disadvantage is that the approach relies heavily on the ability of the recognizer to effectively segment the words into characters. The problem of segmenting a handwritten word into its character components is still a subject of research. The global approach eliminates the need to segment the words into individual characters but the application is limited to small and static lexicons, such as numeral recognition for example. Recent studies indicate that higher levels of performance can be achieved by combining the analytic approach and the global approach [1].

In this research, the performance of a handwriting recognition system that uses both global and local features is investigated. A method to combine two recognition systems, one using HMM and the other using MLP, is also implemented.

2. OFFLINE HANDWRITING RECOGNITION SYSTEM

The block diagram of the offline handwritten word recognition system is shown in Figure 1 below.

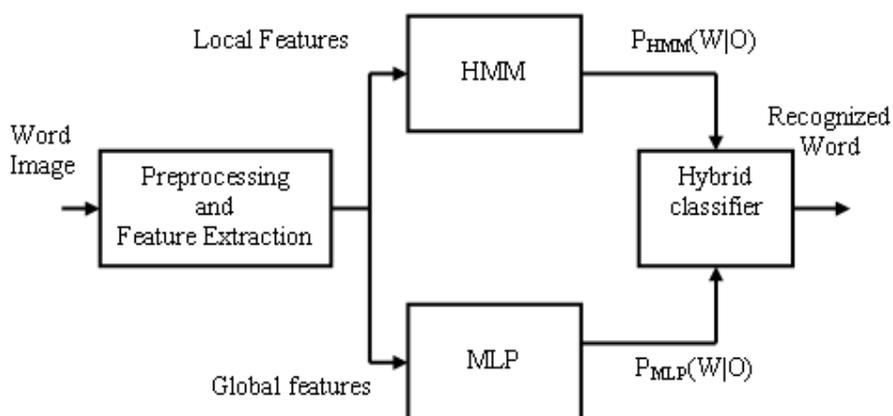


Figure 1. Block Diagram of the Offline Handwritten Word Recognition System

The recognition system takes as its input the isolated handwritten word image, the word images are taken from a public database of handwritten words. The demo version of the IAM database was used to test and train the system [4]. The handwritten word image undergoes image preprocessing and feature extraction. Two sets of features were extracted, the local features from the sub-character units of the word and the global features from the word as a whole. The local features are used as input to the discrete HMM recognizer, the recognizer

outputs posterior probabilities of the n-best recognition results using the maximum likelihood criterion. The global features are used as input to a multi-layer perceptron (MLP), which also outputs posterior probabilities. The results of the HMM and MLP recognizers are combined, using the framework introduced in [13], to recognize the handwritten word.

3. IMPLEMENTATION DETAILS

3.1. The IAM Database

The database of the Institute of Computer Science and Applied Mathematics (IAM) was used to test and train the system. The IAM database was derived from the (Lancaster-Oslo/Bergen) LOB corpus developed at the Computer Science and Applied Mathematics in University of Bern. The database contains large amounts of general unconstrained handwritten modern English texts. The demo version of the IAM database, around 30% of the total database, is publicly available and was used to test and train cursive handwritten words. The database provides scanned form documents with an XML file containing the transcription of the handwritten text on the form documents. The total number of words extracted from the form documents available is 21,119 for a 500-word vocabulary. Since a large number of the words occurred only once, only the words that occurred more than ten times were considered. This considerably reduced the number of word used for training and testing. Shown in Table I is the summary of the datasets used to train and test the system. The first column is the name of the dataset, the second column is the number of words in the dictionary, the third column is the total number of words used for training and the last column is the number of words used for testing. The number of occurrences of each word in the datasets varies proportionately with respect to their usage in the English language.

Train Data	Words	# of Training set	# of Testing set
Trainset1	10	2860	100
Trainset2	20	5780	639
Trainset3	25	6278	694
Trainset4	35	9364	1172
Trainset5	50	5830	815
Trainset6	100	11830	1533

Table I. Datasets Used for Training and Testing

3.2. Preprocessing

The word images from the IAM database underwent preprocessing to reduce the variability in the handwriting process and eliminate unnecessary artifacts in the word image.

The word images extracted from the IAM database are represented as 8-bit gray scale images. For most handwriting recognition systems, the information of interest is primarily the location of the black ink pixels on the white background. An optimal thresholding operation based on Otsu's algorithm [5] was used to convert the grayscale image to binary images. An example illustrating the result of the binarization process is illustrated in Figure 2.



Figure 2. Example Results of the Binarization Process

Handwritten words are usually slanted or italicized due to the mechanism of the handwriting process. Slant correction is performed after the binarization process in order to reduce some of the handwriting variability. The average slant of the word is estimated using the 8-directional chain code method used by [6]. In order to correct the slant of the image, a shear transformation is applied. The result of the shearing operation is shown in Figure 3. The slant correction algorithm helps normalize the word image and it is necessary because local feature extraction involves a segmentation process.



Figure 3. Example of the Slant Angle Estimation and Correction Algorithm

The image undergoes further processing depending on whether local or global feature will be extracted.

3.3. Local Feature Extraction

Prior to the local feature extraction algorithm, the word image undergoes a segmentation process. The ideal segmentation process would be to segment the word into isolated characters then perform character recognition on the individual characters. But due to the large variability and complex nature of the handwriting process, automatic character segmentation is still a very difficult task. A simpler approach to segmentation was adapted based on the methods of [7]; a sliding window segments the word image into objects that are smaller than characters. After the binarization and slant correction, the image is vertically segmented using a fast left-to-right slicing method as described in [7] and is illustrated in Figure 4. The width ω of the sliding windows is dependent on the height of the core zone h_{cz} of the image; typical values are 0.2 to 0.4 of the core zone height with allowable overlap from 20%-40% of ω . This method effectively segments the word into sub-character components.

For each windowed segment 139 local geometric features are extracted. The features can be categorized and summarized as shown in Table II and the details of each are described in [7]. The local geometric features derived from the sub-character components of the word is used as input to the discrete HMM recognizer.

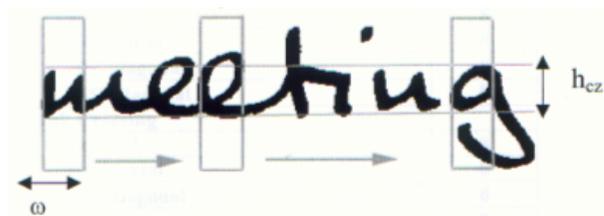


Figure 4. A Sliding Window Segments the Image from Left to Right

Features	Number
Dimension and Aspect Ratio	3
Center of Gravity	2
Distance to Core Zone	2
Ink Pixel Density	
Ascender, Core, Descender Zone, All	4
Vertical	8
Horizontal	8
-45° diagonal	8
+45° diagonal	8
Ink Crossing	
Vertical	8
Horizontal	8
-45° diagonal	8
+45° diagonal	8
Profiles	
Left	8
Right	8
Top	8
Bottom	8
-45° diagonal	8
-135° diagonal	8
+45° diagonal	8
+135° diagonal	8
Total	139

Table II. Summary of the 139 Local Feature Categories [7]

3.4. Global Feature Extraction

The global features derived from the contour of the word image are used to describe the over-all shape characteristics of the word. In [8], it has been shown that the centroidal distance or the polar-radii graph (PRG) can be used to successfully recognize handwritten characters. After the binarization and slant correction stage, the image undergoes morphological processing to estimate the contour of the word. An example of an estimate of a word contour for the word

”‘meeting’” is shown in Figure 5.



Figure 5. Estimate of the Word Contour for the Word ”‘meeting’”

The PRG [8] of the word contour is calculated to derive the global feature vectors for each word. The PRG as the global feature vector is used as the input for the MLP classifier.

3.5. The Discrete HMM Classifier

The local geometric features are used as input to a discrete HMM [9] recognizer. The basic idea is that handwriting can be interpreted as a left-right sequence of ink signals analogous to a speech signal. Hidden Markov Models are known to perform well in modeling signals that vary with time and has become a popular tool for speech as well as in handwriting recognition.

The discrete HMM recognizer was developed using the publicly available HMM toolkit (HTK) for speech recognition developed at Cambridge [10]. Each of the 52 lower and upper case characters were modeled using a 5-state left-to-right, with state skips, HMM model similar to the ones used in [7] as illustrated in Figure 6. Instead of having different number of states as used in [7] for the different characters of the English alphabet, the HMM model is allowed to skip states to represent characters that are normally shorter than the others.

The five possible states of the character HMM model are q_1 , q_2 , q_3 , q_4 and q_5 . The arrows indicate the probability of moving from one state to another called the state transition probabilities. For example P_{13} is the probability of moving from state q_1 to state q_3 . The arrows leaving and entering the same state are called self probabilities; it is the probability of remaining in that particular state.

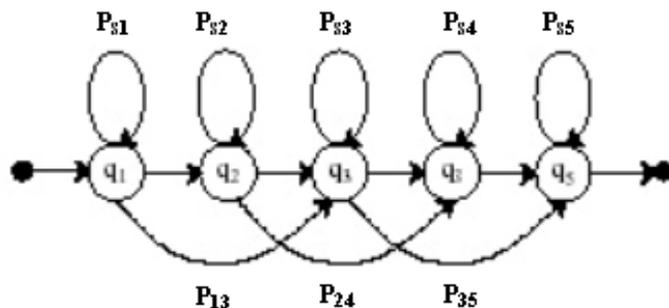


Figure 6. Left-to-Right HMM Model for the Upper and Lower Case Characters

Word HMM models are derived as concatenation of the character HMM models that

comprise the word. The Baum-Welch algorithm was used to re-estimate the parameters of the HMM by adjusting the model parameters to maximize the probability of the observation sequence given the model [9] starting from an initial estimate. This is how the character HMM models are trained to recognize the handwritten word.

In the decoding process, the objective is to find the best state sequence that will maximize the observation probability given the HMM model parameters trained using the Baum-Welch algorithm. Finding the best state sequence given the model parameters is similar to finding the best character sequence that matches the word to be recognized. The Viterbi algorithm, using the maximum likelihood criterion, was designed for this purpose and is the one used to recognize the handwritten word.

The discrete HMM recognizer outputs the posterior probabilities $P(O_j | W_i)$, that is the probability that the local features as observation sequence W_i belong to the class O_j of word. In addition the Viterbi decoder was configured to produce the N-best results of the recognition. Producing the N-best recognition results means that given the local features as input to the HMM, the recognizer can produce the posterior probabilities of the best recognized word, the second best, and the third best results.

3.6. The MLP Classifier

The PRG derived from the word contour is used as the input to the MLP recognizer. The Multi-Layer Perceptron or MLP is a type of neural network with a single layer of hidden units. The publicly available SPRACHcore tools [11] developed in the International Computer Science Institute in Berkeley were used to develop the MLP recognizer. The MLP used has 360 input units that correspond to the PRG of the word contour measured at discrete angles from 0° to 359° at 1° interval. The number of hidden units varies from 150-200 units as determined empirically from experiments that yield the best results. The number of output units is equal to the number of words in the dictionary. The MLP also produces posterior probabilities that may be used to generate the n-best recognition results based on the global features as inputs.

3.7. The Combined MLP-HMM Classifier

The output of the two recognizers, the HMM and the MLP, are the posterior probabilities of the N-best recognition results using the local and global features as inputs. A method to combine these results is needed for the hybrid classifier. Several methods have been used in literature to combine the outputs of two classifiers, using *a posteriori* probability, to generate a new metric for the hybrid classifier ([12], [13], [14]). In [12], three combination schemes have been developed, the maximum voting, Linear Confidence Accumulation (or LCA) [14] and weighted multiplication method. The idea is to derive a new probability measure from the output posterior probabilities of an HMM and MLP classifier and to determine if the combination will yield better recognition rates.

Another method that combines an HMM and a neural network classifier was used in [13] for large vocabulary word recognition. In [13], the HMM classifier produces at the output a ranked list on the N-best results with estimated *a posteriori* probabilities. A neural network that also outputs N-best results is then combined with the HMM output to generate a composite score

by a weighted combination of the outputs of both classifiers using Equation 1.

$$R_{score} = \alpha \log(HMM_{score}) + \beta \log(MLP_{score}) \quad (1)$$

where $\alpha + \beta = 1$

The new score R_{score} , is the composite score of the word. The values of α and β are the weights associated with the HMM and MLP recognizers. It is based on the idea that if both classifiers estimate Bayesian *a posteriori* probabilities at the output, they can be combined in a probabilistic framework [15]. The result of using Equation 1 is that we now have a rescored N-best list that can be re-ordered based on the composite score. The effect is that the second or third best choice of the recognizer has a chance to become the first choice.

4. PERFORMANCE EVALUATION

The performance of the discrete HMM recognizer, the MLP recognizer and the combined results are discussed. The system was trained and tested using the datasets in Table I with 10, 20, 25, 35, 50 and 100-word dictionary. Because of the limitations of the demo version of the IAM database the number of words used for training can no longer be increased.

4.1. Recognition Results of the Discrete HMM Recognizer

Shown in Table III is the summary of the performance of the discrete HMM recognizer developed using HTK with the local geometric features as input. The highest recognition rate is 66% for a 10-word vocabulary using Trainset1. By using a larger vocabulary size, Trainset 2 for example, the highest recognition rate obtained is 58% for a 20-word dictionary with 5,780 training samples. An increase in the size of the dictionary lowers the recognition performance because the number of choices of possible words also increases.

Train data	words	Training	Test set	HMM
Trainset1	10	2860	100	66%
Trainset2	20	5780	639	58%
Trainset3	25	6278	694	55%
Trainset4	35	9364	1172	48%
Trainset5	50	5830	815	25%
Trainset6	100	11830	1533	34%

Table III. Summary of the Discrete HMM Results

4.2. Recognition Results of the MLP Recognizer

A summary of the characteristics and performance of the MLP recognizer is shown in Table IV. The cross-validation (CV) and the Train columns indicate how well the MLP is able to capture the global features of the word during training. The test results show the performance of the global features in recognizing the word using the test data. For the 10-word vocabulary, a recognition rate of 95% was obtained and for the 20-word vocabulary the recognition rate is 78%.

Train Data	words	Total	Test	CV	Train	Test
Trainset1	10	2860	100	96%	95%	95%
Trainset2	20	5780	639	84%	92%	78%
Trainset3	25	6278	694	81%	93%	77%
Trainset4	35	9364	1172	76%	87%	68%
Trainset5	50	5830	815	68%	76%	42%
Trainset6	100	11830	1533	62%	78%	56%

Table IV. Summary of the MLP Recognition Results

4.3. Recognition Results of the Combined HMM-MLP Recognizer

The output *a posteriori* probabilities of the MLP and the discrete HMM recognizer were combined to create a composite score using Equation 1. The composite score is used to re-order the N-best list of words to obtain the recognized word. A summary of the performance of the combined HMM-MLP recognizer is shown in Table V. The recognition rate of the combined classifier using the local and global features and using Equation 1 is 72% for a 20-word vocabulary with $\alpha = 0.1$ to 0.4. This result is a significant increase in the 58% performance of the discrete HMM recognizer alone. Varying the values of from 0.4 to 0.9 has little effect on the recognition rate that varies from 72% to 71.5%. This result indicates that there is not much correlation between the HMM and MLP output posterior probabilities for the features used in the experiments.

Train data	words	Training	HMM	MLP	Hybrid
Trainset1	10	2860	66%	95%	68%
Trainset2	20	5780	58%	78%	72%
Trainset3	25	6278	55%	77%	70%
Trainset4	35	9364	48%	68%	50%
Trainset5	50	5830	25%	42%	28%
Trainset6	100	11830	34%	56%	35%

Table V. Recognition Rate of the Combined Hybrid HMM-MLP Classifier

5. CONCLUSIONS AND RECOMMENDATIONS

Based on the results previously discussed we can conclude that global features based on the PRG of the word contour is not complementary to the local features chosen in this research since the recognition rate of the combined HMM-MLP is lower than the MLP recognizer using PRG. Future work can explore other local features and classifiers to create a better combination. The experiments also showed that increasing the number of words in the vocabulary inherently decreases the performance of the recognizers because more choices are made available to the recognizer. Thus there is an inherent advantage in combining the local and global features for handwriting recognition compared to using the global features only. When local features are used for recognition, and using the analytic approach the number of

entities that needs to be trained is fixed; only the 52-character HMM models need to be trained. The system trained using the local features can easily incorporate new words in the dictionary by training the character HMM models using the new words only and not the entire database. Thus the increase in the vocabulary because of the new words will not have a significant effect on the recognition performance.

In summary, an offline handwriting recognition system was successfully developed using both the local geometric features and the global features based on the PRG. The local features are used as input to the discrete HMM recognizer with a recognition rate of 58% for a 20-word vocabulary. The global features was used as input to the MLP classifier and obtained a recognition rate of 78% for the 20-word vocabulary. By combining the results of the classifiers an over-all recognition rate of 72% is achieved. Among the three methods the MLP classifier using PRG had the best performance. The recognition rate of the MLP classifier may be improved by using a larger training dataset. This method is ideal for recognition tasks with small vocabularies such as recognition of courtesy amount on bank checks.

This research utilized some of the publicly available tools for speech recognition to develop a handwriting recognition system. There software tools are the HTK and the SPRACHcore tools. Other parameter refinement techniques, which are normally used in speech recognition, available in the HTK and SPRACHcore tools can be used to improve the accuracy of the recognition system.

This research used the demo version of the publicly available IAM database for training and testing. It is recommended that the system utilize the full version of the IAM database, when it becomes publicly available, to significantly increase the number of words available for training and testing the system. This will also consequently increase the recognition rate of the system.

REFERENCES

1. R. Plamondon, S. Srihari, "Online and Offline Handwriting Recognition: A comprehensive survey", *IEEE Transaction on PAMI*, vol. 22, no.1, 2000
2. S. Madvanath, J. Govindaraju, "The Role of Holistic Paradigms in Handwritten Recognition", *IEEE Transactions on PAMI*, vol. 23, no. 2, 2001
3. C. Parisse, "Global word shape processing in Offline Recognition of Handwriting", *IEEE Transaction on PAMI*, vol. 18, no. 4, 1996
4. U. Marti, H. Bunke, "A full English sentence database for offline handwriting recognition", *ICDAR '99*, Bangalore, pages 705-708, 1999
5. N. A. Otsu, "Threshold selection method from Gray-level histograms", *IEEE Transaction on Systems, Man and Cybernetics*, vol. 15, no. 5, 1979
6. T. Ding, F. Kimura, Y. Miyake, M. Shridhar, "Accuracy Improvement of Slant Estimation for Handwritten Word", *ICPR*, 2000
7. Y. H. Tay, P. Lallican, M. Khalid, C. Viard-Gaudin, C. Knerr, "An Offline Cursive Handwritten Word Recognition System", *5th IEEE TENCON*, 2001
8. J. Oche, R. Cajote, R.C.L. Guevara, "Polar-Radii Graphs: A Geometric-invariant shape based approach to Handwritten Character Recognition", *TENCON 2003*, Bangalore, India, 2003
9. L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition", *Proceeding of the IEEE*, vol. 7, no. 2, 1989
10. S. Young, G. Everman, *et al.*, "The HTK Book (for HTK version 6.2)", Cambridge University Engineering Department, 2002. <http://htk.eng.cam.ac.uk/docs/docs.shtml>
11. ICSI Speech Recognition: Embedded Training, <http://www.icsi.berkeley.edu/dpwe/isrinfo/embedded.html>
12. J. H. Kim, K. K. Kim, C. P. Nadal, C. Y. Suen, "A Methodology for Combining HMM and MLP Classifiers for Cursive Word Recognition", *International Conference on Pattern Recognition*, Barcelona, Spain, 2000

13. Koerich, A.L., Leydier, Y., Sabourin, R., Suen, C.Y., "A Hybrid Large Vocabulary Handwritten Word Recognition System Using Neural Networks and Hidden Markov Models", ICDAR 2001
14. Y.S. Huang, C.Y. Suen, "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals", *IEEE Transactions on PAMI*, vol. 17, no. 1, pp.90-94, 1995
15. G. Zavaliagos, Y. Zhao, R. Schwartz and J. Makhoul, "A hybrid segmental neural net/hidden markov model system for continuous speech recognition", *IEEE Transaction on Speech and Audio Processing*, 2(1):151-160, 1994

