Handwritten Word Recognition Using Multi-view Analysis

J.J. de Oliveira Jr.[1], C.O. de A. Freitas[2], J.M. de Carvalho[3], and R. Sabourin[4]

[1] UFRN - Universidade Federal do Rio Grande do Norte
josejosemarjr@gmail.com
[2] PUC-PR - Pontíficia Universidade Católica do Paraná
cinthia.freitas@pucpr.br
[3] UFCG - Universidade Federal de Campina Grande
carvalho@dee.ufcg.edu.br
[4] ÉTS - École de Technologie Supérieure
sabourin@etsmtl.ca

Abstract. This paper brings a contribution to the problem of efficiently recognizing handwritten words from a limited size lexicon. For that, a multiple classifier system has been developed that analyzes the words from three different approximation levels, in order to get a computational approach inspired on the human reading process. For each approximation level a three-module architecture composed of a zoning mechanism (pseudo-segmenter), a feature extractor and a classifier is defined. The proposed application is the recognition of the Portuguese handwritten names of the months, for which a best recognition rate of 97.7% was obtained, using classifier combination.

1 Introduction

In a general way handwritten recognition systems are defined by two operations: features extraction and classification. Feature extraction is related to information extraction, creating the word representation used as input to the classifier. Thus, the goal of feature extraction is to capture the most relevant and discriminatory information of the object to be recognized, eliminating redundancies and reducing the data amount to be processed. The classifier based on this representation associates conditional probabilities to the classes by means of an estimation process.

This study deals with recognition of the Portuguese month names represented by a limited lexicon of 12 classes: Janeiro, Fevereiro, Março, Abril, Maio, Junho, Julho, Agosto, Setembro, Outubro, Novembro and Dezembro. Some of these classes share a common sub-string, which adds complexity to the problem. As can be observed in Figure 1, there is similarity between the suffix of some classes in the lexicon, which creates confusion and affects the performance of the recognizer. Another source of confusion is a common first letter (e.g. junho and

E. Bayro-Corrochano and J.-O. Eklundh (Eds.): CIARP 2009, LNCS 5856, pp. 371–378, 2009.

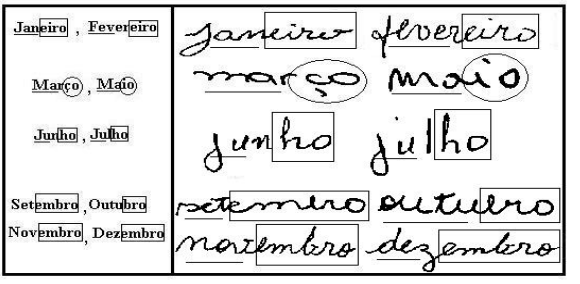

Fig. 1. Complexity of the recognition problem: prefix and suffix

julho), which plays a significant role in the word recognition process, as observed by Schomaker[1]. Further difficulty is added by the fact that the vowels (a, e, i, o) exhibit low discriminatory power in the human reading process[1].

Performance of the recognition system for the considered lexicon is limited by the type of confusion illustrated in Figure 1. To overcome these constraints, we utilize an approach based on multi-view representation and perceptual concepts, designed to avoid the intrinsic difficulties of the lexicon. This approach is one evolution of other previous systems published by the same research group[2,3], however the multi-view analysis proposed here are original and it is based on perceptual concepts. This particular choice of lexicon, does not take from the generality of the solution, since that the same problem are founded in other lexicons like in French language[4]. The proposed system can be applied equally well to any similar problem, thus bringing a true contribution to the state of the art in the area.

This paper is divided into 4 sections. Section 2 describes the overall system developed, considering the multi-view representation and system architecture. In Section 3, the experimental results are presented and analyzed. Finally in Section 4 the conclusions and suggestions for future work are presented.

2 Methodology

This section presents an overview of the proposed system based on multi-view representation. The words database used and the preprocessing operations applied are described. Next, each pseudo-segmentation scheme is defined, combined with feature vectors extraction and the classification method utilized. The classifiers outputs are combined in order to produce a final decision for the sample in analysis.

2.1 Multi-view Analysis

Usually, two main approaches are considered for Handwritten Word Recognition - HWR problems: local or analytical approaches held at character level and global approach held at word level[5].

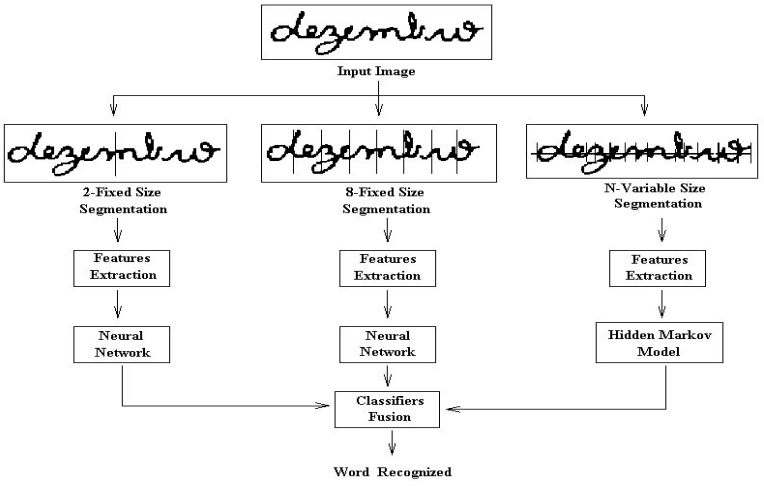

Fig. 2. System block diagram

The global approach extracts features from the words as a whole, therefore making it unnecessary to explicitly segment words into characters or pseudo-characters. This approach seeks to explore information from the word context, allowing aspects based on psychological models to be considered. The global or word level allows to incorporate principles of the human reading process into computational methods[5].

The local approach utilizes the word basic elements in the recognition strategy. These elements can be characters or segments of characters (pseudo-characters). This approach is characterized by the difficulty to define the segmentation or separation points between characters. Therefore, success of the recognition method will depend on success of the segmentation process[5].

Our focus is on the global level, as it seeks to understand the word as a whole, similarly to the human reading process where the reader uses previous knowledge about the word shape to perform recognition.

Although the global analysis does not use segmentation, pseudo-segmentation (or zoning) mechanisms can be added to produce a robust recognition system[3]. Zoning basically consists of partitioning the word image into segments (or sub-images) of equal or variable size. Three different zoning schemes have been employed, as described next and illustrated in Figure 2.

- 2-FS (2 fixed size sub-regions): Each image is split in two, to the right and to the left of the word center of gravity[2];
- 8-FS (8 fixed size sub-regions): Each image is divided in 8 sub-regions of equal size. This number corresponds to the average number of letters in the lexicon;
- N-VS (N variable size sub-regions): The words horizontal projection histogram of black-white transitions is determined. The line with maximum

histogram value is called Central Line (CL). A segment is defined by two consecutive transitions over the CL.

Multi-view analysis therefore, seeks to provide different simultaneous approximations for the same image. For each zoning procedure, one specific feature vector and classifier are defined, all based on global word interpretation. At the end, the classifiers outputs are combined to produce the final decision, therefore taking advantage of the zoning mechanisms complementarity.

2.2 Word Database and Preprocessing

To develop the system it was initially necessary to construct a database that can represent the different handwriting styles present in the Brazilian Portuguese language for the chosen lexicon. The words were digitized at 200 dpi. Figure 3 illustrates some samples from this database. To reduce the variability, slant and baseline skew normalization algorithms were applied, using inclinated projection profiles and shear transformation.

2.3 2-FS Feature Set

In this word representation, perceptual features and characteristics based on contour as concavities/convexities are represented by the number of their occurrences. The features extracted from each word form a vector of dimension 24. Perceptual features are considered high-level features due to the important role they play in the human reading process, which uses features like ascenders, descenders and estimation of word length to read handwritten words[5].

The components of the feature set can be described as following[2]:

- Number of concave and convex semicircles, number of horizontal and vertical lines, number of ascenders and descenders with loop in the left/right areas, respectively;
- Number of crossing-points, branch-points, end-points, loops, ascenders and descenders on the left/right areas, respectively;
- Number of horizontal axis crossings by stroke;
- Proportion of white/black pixels inside the word bounding box.

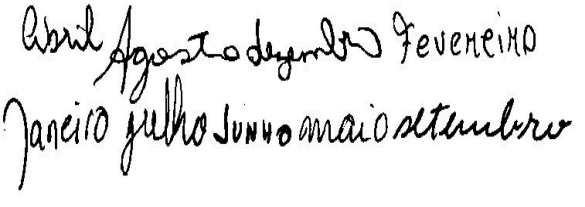

Fig. 3. Sample images from the database

2.4 8-FS Perceptual Feature Set (P)

In this zoning mechanism, ten patterns are defined for each sub-region, thus forming for each image a feature vector containing 80 patterns.

The 10 patterns used in the perceptual feature set are:

- x_1, x_2, x_3, x_4 **- Ascender (and Descender) position and size:** Position and height of the ascender (and descender) central pixel;
- x_5, x_6, x_7 **- Closed loop size and location:** Number of pixels inside a closed loop and coordinates of the closed loop center of mass;
- x_8, x_9 **- Concavity measure:** Initially the convex hull is constructed starting at the bottom-most point of the boundary. The leftmost and rightmost points in the hull are detected and the angles (relative to the horizontal) defined by the line segments joining them to the starting point are measured;
- x_{10} **- Estimated segment length:** Number of transitions (black-white) in the central line of the sub-region outside of the closed loops.

2.5 8-FS Directional Feature Set (D)

The directional features can be considered intermediate-level features, conveying relevant information about the image background[5]. In this paper, the directional features defined are based on concavity testing, where for each white image pixel (or background pixel) it is tested which of the four main directions (NSEW) leads to a black (contour) pixel.

Representation is made labeling the background pixels, that depends on the combination of the open directions. The components of the feature vector for each sub-region are obtained by counting the number of pixels for each label.

2.6 2-FS and 8-FS Classifier

To each 2-FS and 8-FS features set one classifier based on Class-Modular MLP was defined. It follows the principle that a single task is decomposed into multiple subtasks and each subtask is allocated to an expert network. In this paper, as well as in Oh et al.[6], the K-classification problem is decomposed into K 2-classification subproblems. For each one of the K classes, one 2-classifier is specifically designed.

Therefore, the 2-classifier discriminates that class from the other $K-1$ classes. In the class-modular framework, K 2-classifiers solve the original K-classification problem cooperatively and the class decision module integrates the outputs from the K 2-classifiers.

2.7 N-VS Feature Set

The features utilized by the N-VS zoning mechanism are the same as those presented in sections 2.4 and 2.5, though different extraction and representation methods were used and adapted to this approach. A symbol is designated to represent the extracted set of features for each segment, building up a grapheme.

In the case where no feature is extracted from the analyzed segment, an empty symbol denoted by **X** is emitted. This feature set is capable of representing the link between letters and separating graphemes[4,7].

2.8 N-VS Classifier

The N-VS zoning mechanism defines a variable number of sub-regions which makes neural network application difficult. Therefore, Hidden Markov Model (HMM) classifiers are more recommended in this case[4,7]. The entire and definitive alphabet is composed of 29 different symbols selected from all possible symbol combinations, using the mutual information criterion[4,7].

Our HMM word models are based on a left-right discrete topology where each transition can skip at the most two states. Model training is based on the Baum-Welch Algorithm and the Cross-Validation process is performed on two data sets: training and validation. After the Baum-Welch Algorithm iteration on the training data, the likelihood of the validation data is computed using the Forward Algorithm[4,7]. During the experiments, the matching scores between each model λ_i and an unknown observation sequence O are carried out using the Forward Algorithm.

3 Experimental Results

For the experiments, the database was randomly split into three data sets: Set 1 - Training Base with 6,120 words; Set 2 - Validation Base and Set 3 - Testing Base, both with 2,040 words. For each set, the words are evenly distributed among the classes.

For each feature set considered in the system (2-FS and 8-FS), 12 (twelve) Class-Modular MLP classifiers were trained and tested. In the Class-Modular approach, the classifier that presents the maximum output value indicates the class recognized[6]. The amount of neurons in the hidden layer was empirically determined, different configurations being tested. Each K 2-classifier is independently trained using the training and validation sets. The Back-propagation Algorithm was used in all cases. To train a 2-classifier for each word class, we reorganized the original training and validation sets into 2 sub-sets: Z_0 that contains the samples from current class and Z_1 that contains the samples from all other K-1 classes. To recognize the input patterns, the class decision module considers only the O_0 outputs from each sub-network and uses a simple winner-takes-all scheme to determine the recognized class[6].

The N-VS classifier was evaluated with the N-VS feature set and for each class one HMM was trained and validated. The model that assigns maximum probability to one test image represents the class recognized.

Table 1 shows the results obtained for each classifier individually. It can be seen that the best result was obtained using 8-FS classifier with directional features.

Table 1. Recognition rate obtained for each classifier individually

Classifier	2-FS	8-FS(P)	8-FS(D)	N-VS
RR	73.9%	86.3%	**91.4%**	81.7%

Table 2. Recognition rate obtained using classifiers combination

Classifiers	RR (%)
2-FS and 8-FS(P)	90.5
2-FS and 8-FS(D)	94.4
8-FS(P) and 8-FS(D)	93.6
8-FS(P) and N-VS	93.5
8-FS(D) and N-VS	95.6
2-FS and N-VS	90.5
2-FS, 8-FS(P) and 8-FS(D)	95.4
2-FS, 8-FS(P) and N-VS	95.8
2-FS, 8-FS(D) and N-VS	97.2
8-FS(P), 8-FS(D) and N-VS	96.9
2-FS, 8-FS(P), 8-FS(D) and N-VS	97.7

3.1 Classifiers Fusion

To obtain the hybrid classifier it is necessary to define a combination rule for the classifiers's outputs. Initially, we made the assumption that an object Z must be assigned to one of the K possible classes $(w_1, \cdots, w_K)$ and assume that L classifiers are available, each one representing the given pattern by a distinct measurement vector. Denote the measurement vector used by the ith classifier as x_i and the *a posteriori* probability $P(w_j|x_1, \cdots, x_L)$[8]. Therefore, the combining rule is defined as:

– Weighted Sum (WS): Assigns Z to class w_j if

$$\sum_{i=1}^{L} \alpha_i \cdot p(w_j|x_i) = \max_{k=1}^{K} \sum_{i=1}^{L} \alpha_i \cdot p(w_k|x_i); \quad (1)$$

where α_i, $i = 1, \cdots, L$ are weights for each classifier.

To guarantee that the classifier outputs represent probabilities, an output normalization was performed: $P^*(w_j|x_i) = \frac{P(w_j|x_i)}{\sum_K P(w_j|x_i)}$. The best weights were obtained by an exhaustive search procedure, considering for each classifiers combination, 2,000 different n-upla of weight vectors with random adaptation.

The average recognition rates obtained considering different classifiers combination are presented in Table 2. It can be seen that the best result was obtained using combination for 2-FS, 8-FS(P), 8-FS(D) and N-VS classifiers.

4 Discussion and Conclusions

This paper presents a hybrid system using a methodology based on multi-view analysis, applied to the recognition of the Portuguese handwritten names of the

months. This system is based on a Global Approach, which extracts global features from the word image, avoiding explicit segmentation. This approach is more that one simple combination of classifiers since that explores word context information, while allows incorporating aspects based on perceptual concepts. Therefore, unlike other proposed systems, we have a computational approximation inspired in the human reading process.

We have evaluated the efficiency of multiple architectures using Neural Network and Hidden Markov Models classifiers for the handwritten word recognition problem. The main conclusion obtained is that the analyzed classifiers are complementary and the combining strategy proposed enhances their complementarity. Therefore, the classifiers arranged in the multi-view analysis are a better solution for our problem than any of the classifiers applied individually. This result indicates that a similar strategy can be applied to other restricted lexicons. Future work will focus on the analysis of adaptative models that will be applied to large lexicons.

References

1. Schomaker, L., Segers, E.: A Method for the Determination of Features used in Human Reading of Cursive Handwriting. In: IWFHR 1998, The Netherlands, pp. 157–168 (1998)
2. Kapp, M.N., de Almendra Freitas, C.O., Sabourin, R.: Methodology for the Design of NN-based Month-Word Recognizers Written on Brazilian Bank Checks. International Journal on Image and Vision Computing 25(1), 40–49 (2007)
3. de Almendra Freitas, C.O., Oliveira, L.S., Aires, S.K., Bortolozzi, F.: Handwritten Character Recognition Using Non-Symmetrical Perceptual Zoning. International Journal on Pattern Recognition and Artificial Intelligence 21(1), 1–21 (2007)
4. de Almendra Freitas, C.O., Bortolozzi, F., Sabourin, R.: Study of Perceptual Similarity Between Different Lexicons. International Journal on Pattern Recognition and Artificial Intelligence 18(7), 1321–1338 (2004)
5. Madhvanath, S., Govindaraju, V.: The Role of Holistic Paradigms in Handwritten Word Recognition. IEEE Trans. on PAMI 23(2), 149–164 (2001)
6. Oh, I., Suen, C.-Y.: A Class-Modular Feedforward Neural Network for Handwriting Recognition. Pattern Recognition 35(1), 229–244 (2002)
7. de Almendra Freitas, C.O., Bortolozzi, F., Sabourin, R.: Handwritten Isolated Word Recognition: An Approach Based on Mutual Information for Feature Set Validation. In: ICDAR 2001, Seattle - USA, pp. 665–669 (2001)
8. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. IEEE Trans. on PAMI 20(3), 226–239 (1998)