Feature Sets Evaluation for Handwritten Word Recognition

José J. de Oliveira Jr, João M. de Carvalho
UFPB - Federal University of Paraiba, Department of Electrical Engineering,
Postal Box 10105, 58109-970, Campina Grande, PB - Brazil
josemar,carvalho@dee.ufpb.br

Cinthia O. de A. Freitas
PUCPR - Pontíficia Universidade Católica do Paraná, Rua Imaculada Conceição 1155, Prado Velho
80215-901, Curitiba, PR - Brazil
cinthia@ppgia.pucpr.br

Robert Sabourin
ÉTS - École de Technologie Supérieure, 1100 Rue Notre Dame Ouest
H3C 1K3, Montreal - Canada
sabourin@gpa.etsmtl.ca

Abstract

This work presents a baseline system used to evaluate feature sets for word recognition. The main goal is to determine an optimum feature set to represent the handwritten names for the months of the year in Brazilian Portuguese language. Three kinds of features are evaluated: perceptual, directional and topological. The evaluation shows that taken in isolation, the perceptual feature set produces the best results for the lexicon used. These results can be further improved combining the feature sets. The baseline system developed obtains an average recognition rate of 87%. This can be considered a good result considering that no explicit segmentation is performed.

1. Introduction

The main objective of this work is to evaluate different feature sets for a particular handwritten recognition problem. Several authors when presenting their systems describe different kinds of features to represent the data [6]. However, comparasion of sets using a baseline system is necessary in order to answer the fundamental question: **What is the best kind of feature to represent handwritten words for a given application?**

Some authors have tried to incorporate knowledge about human reading mechanisms in their systems [4, 2, 3, 6], justifying that this duality (human-computer) has been applied in other areas with succes (for example, speech recognition). But another question appears: **Is this duality really efficient and necessary for handwritten recognition?**

To answer these questions a baseline system has been defined and used as comparative mechanism. The application chosen was recognition of the Portuguese handwritten names of the months. This is an important problem since it constitutes a sub-problem of bankcheck date recognition. Although this study deals with a limited lexicon of 12 classes, there are classes that share a common sub-string, which can affect the overall system performance: Jan***eiro***, Fever***eiro***, ***Ma***rç***o***, Abril, ***Ma***i***o***, ***Ju***n***ho***, ***Ju***l***ho***, Agosto, Set***embro***, Outu***bro***, Nov***embro*** and Dez***embro***.

Another point investigated is the applicability of global or holistic approaches to word recognition. Global or holistic techniques make no attempt to segment the word into sub-units, relying instead on word level feature extraction and matching to determine the identity of the word, therefore avoiding problems of segmentation, ambiguity and variability of segment shape. There is a consensus among researchers in this field that holistic approaches are of utility either in cases of small lexicons or in the filtering of large lexicons.

In this paper, we present and analyze a method for handwritten word recognition used to evaluate the selected feature sets. Initially, a system overview is presented in Section 2, describing the distinct stages of operation as well as the feature sets extracted. In Section 3 the experimental results are presented and analyzed, in order to determine the discriminating potential for each feature set. This is done

based on the recognition rate obtained with the baseline system, considering each set individually as well as combinations of them. Finally, this paper is concluded trying to answer the questions formulated in this Introduction.

2 System Overview

2.1 Image Acquisition

To develop the system it was necessary to construct a database that can represent as well as possible the different styles of handwriting present in the Brazilian Portuguese language. This was done by collecting 500 samples of each month name, from writers of different levels of education. Each writer was asked to fill a specific form where the word corresponding to each month name would be written once. No restrictions were imposed regarding writing style and no handwritten models were provided, which resulted in a very heterogeneous database. The words were digitilized using a scanner set to 200 DPI. Figure 1 illustrates some samples from this database.

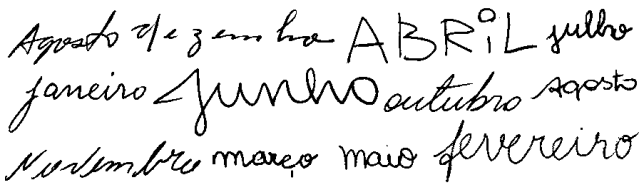

Figure 1. Sample images from the database.

2.2 Preprocessing

The form does not provide reference lines to the writer, which causes the words to present different baseline skew and slant. To reduce pattern variability a slant and baseline skew normalization algorithm was applied [2], which uses inclinated projection profiles and shear transformation.

Judging by subjective visual inspection, the preprocessing applied produces good results in 99% of the images forming the database. Figure 2 examplifies the results obtained at this stage.

2.3 Implicit Segmentation and Feature Extraction

A limitation with neural classifiers is the need for a fixed size input vector. To solve this problem an implicit segmentation is performed by which each sample image is divided in 8 sub-regions, as shown in Figure 3. This number corresponds to the average number of letters in the lexicon words. For each sub-region, ten patterns are defined denominated $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$, thus forming

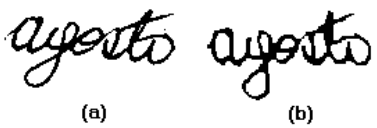

Figure 2. Example of preprocessing: (a) original image and (b) preprocessed image.

a feature vector containing 80 patterns for each image. Another requirement of the neural classifier are normalized input patterns, implying that all components of the feature vector need to be normalized accordingly with the features definition, as described in the next section.

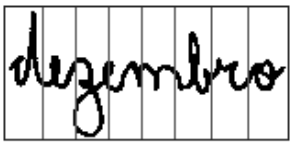

Figure 3. Example of implicit segmentation.

Three different feature sets were determined based in the holistic features classification proposed by Madhvanath [6], which consists of three classes: High-level, Intermediate-level and Low-level features. The three sets defined are named as perceptual, directional and topological, and are described next.

2.4 Perceptual Features

The perceptual features are considered high-level features and their utilization is justified by the human reading process, which uses features like ascenders, descenders and estimation of word lenght to read handwritten words [6].

To extract ascenders and descenders it is necessary to determine the image reference lines. To do this, the words horizontal projection histogram of black-white transitions is initially determined. The line with maximum histogram value is called Central Line (CL). After this, a smoothing procedure is applied to eliminate histogram discontinuities. The Upper (UL) and Lower (LL) Lines are the ones above and below CL, respectively, with 70% of the maximum histogram value. This percentage was obtained heuristically by Freitas [3]. An example of this procedure is presented in Figure 4.

The 10 patterns used in the perceptual feature set are:

- x_1 **- Ascender position:** Position of the ascender central pixel, normalized by the sub-region width;

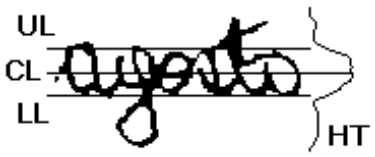

Figure 4. Example of reference lines detection.

- x_2 - **Ascender size:** Height of ascender normalized by the height of central sub-region;
- x_3, x_4 - **Descender position and size:** Same as defined for ascenders, considering the descender sub-region;
- x_5 - **Closed loop size:** Number of pixels inside the closed loop normalized by the sub-region area. A closed loop is defined as the region where from an internal pixel a black pixel is always reached for any moving direction;
- x_6, x_7 - **Closed loop location:** Coordinates of the closed loop center of mass. The x and y coordinates are normalized by the sub-region width and height, respectively;
- x_8, x_9 - **Concavity angles:** Initially the convex hull is constructed starting at the bottom-most point of the boundary as shown in [9]. The leftmost and rightmost point 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. The angles are normalized by 90°;
- x_{10} - **Estimated word lenght:** Number of transitions (black-white) in the central line of the sub-region, normalized by the total number of transitions in the central line of the word. One transition is defined as the *background-foreground* or *foreground-background* transition outside of the closed loops.

When a pattern does not occurs in a sub-region it is necessary to assign a value to represent this absence. The zero value is not a good choice, because the occurrence of many null patterns would degrade the NN performance. Therefore, it was decided to assign unity value to indicate absence of a pattern, the same value used when a given pattern assumes value greater than 1.0.

2.5 Directional Features

The directional features can be considered intermediate-level features, conveying relevant information about the image background. In this paper, the directional features defined are inspired by the idea of concavity measurements [9], where for each white image pixel it is verified in each of the main four directions (NSEW) if a black pixel can be reached, as shown in Figure 5.

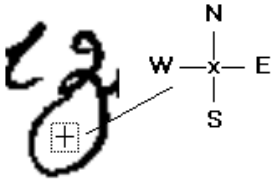

Figure 5. Example of directional features extraction.

Depending on the number and combination of the open directions the background pixels are labeled by the convention depicted in Table 1.

Table 1. Convention used for directional feature set.

Label	Type
0	Closed in all directions
1	Open down
2	Open up
3	Open right
4	Open left
5	Open right and up
6	Open left and up
7	Open left and down
8	Open right and down
9	Open down and up

Label 9 is used to represent letters without ligature strokes. The components of the feature vector $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$ for each sub-region are obtained by counting the number of pixels attributed to each label, normalized by the sub-region area. When there are no pixels of some label, the corresponding value mapped to the vector is 1.0.

2.6 Topological Features

Topological features reflect pixel density over the image regions, being classified as low-level features. To determine these features a zoning was performed, dividing each sub-region in two parts, above and below the word central line.

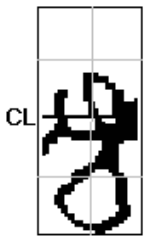

Figure 6. Example of zoning used.

After this, the upper and lower parts were each divided in 4 zones, as shown in Figure 6.

The feature vector components $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ are obtained counting the number of black pixels in each of the eight zones, normalized by the respective zone area. Components (x_9, x_{10}) correspond to the zone center of mass coordinates normalized by the sub-region weight and height, respectively. When the number of black pixels is zero, the value mapped to the vector is **1.0**.

2.7 Neural Classifier

The neural classifier was chosen to this application because the lexicon presents a small and limited number of classes. Also, this classifier has not been largely used for handwritten word recognition. Thus, the main motivation was to evaluate its adequability to this kind of problem.

The neural network (NN) used was of the MLP-type implemented via the SNNS simulator program [1]. Each NN is composed by 80 patterns in the input layer, one hidden layer of variable size and 12 patterns in the output layer. Input data is shuffled before presentation and the back-propagation with momentum algorithm plus one update function for optimizing the adaptation weights were used for training. Training with validation was employed in order to avoid over-learning. The error obtained in the validation set for each training epoch was used as stop criterion.

3 Experimental Results

Initially, the database was randomly split in three sets, namely: **Training**, **Validation** and **Test** sets, that contains 300, 100 and 100 samples for each class, respectively.

For each feature set presented in Section 2.4, the NN is trained and tested. The class that presents the maximum output value is the class recognized. The quantity of neurons in the hidden layer is empirical, different configurations being tested. The best results were obtained using 75, 80 and 85 neurons for perceptual features (NN-P), directional features (NN-D) and topological features (NN-T). Table 2 shows the results obtained for each feature set individually. It can be seen that the best results were obtained using the perceptual feature set. Tables 3, 4 and 5 show the confusion matrix obtained for each feature set.

Table 2. Average recognition rate obtained for each scheme individually.

Set	Recognition rate
Perceptual (NN-P)	81.8 %
Directional (NN-D)	76.6 %
Topological (NN-T)	73.0 %

The tables are analyzed considering the classes grouped in clusters, defined by the presence of a common sub-string:

- Jan**eiro**-Fever**eiro**: The results show that most of the confusion for these classes happenned between themselves. This was expected because they present ascenders in the same position, difficulting that feature extraction. However, for the topological set (T), the word *Janeiro* produced considerable confusion with other classes, like *Junho*, for example.
- **Ma**r**ço**-**Ma**i**o**: Despite the similarity of the words, confusion is relatively low for this pair, likely due to the descender in *Março*. However, when the D and T sets are considered the word *Maio* produces considerable confusion with several other classes, consequence of the fixed segmentation procedure that over-segments this word particularly.
- **Ju**n**ho**-**Ju**l**ho**: The main confusions for these classes occur between themselves, as expected since they are very similar. Also, depending on the segmentation, the ascenders can be located in the same sub-region thus difficulting class separation. However, for set T a high occurence of confusion between *Junho* and *Janeiro* is observed.
- **A**bril-**A**gosto: This words have not similarities, thus the occurence of confusions between them is low. The word *Abril*, shows no confusion with any determined class, specifically. For *Agosto* a high level of confusion with *Dezembro* can be observed for sets D and T, maybe caused by the close placement of ascenders/descenders, or by the segmentation procedure.
- Set**embro**-Outu**bro**-Nov**embro**-Dez**embro**: The main confusions for those words are among themselves, mainly, between *Setembro* and *Outubro* for set D.

This analysis shows that the main problems of the system occur generally between classes in the same cluster,

Table 3. Confusion matrix obtained with perceptual feature set (NN-P).

Month	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**75**	10	2	1	2	4	4				2	
Fevereiro	7	**77**	1	1	5	1	2			1	4	1
Março		1	**84**	5	4	1	1	2			1	1
Abril		1	3	**88**	2	1		2	1	1	1	
Maio	2	1		2	**82**	6	1	2	2	2		
Junho	3		1	1	1	**82**	4	2	2	2	1	1
Julho				2	3	6	**87**	1	1			
Agosto	1		2	2	1		2	**88**			1	3
Setembro	1	3	1	1	1	4			**76**	6	4	3
Outubro		3		3		3	1		5	**85**		
Novembro	1	1	2	3		2			6	1	**82**	2
Dezembro	3	2		1	1	3	1	3	1	5	4	**76**

Table 4. Confusion matrix obtained with directional feature set (NN-D).

Month	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**74**	10		1	2	5	2		2	3		1
Fevereiro	8	**77**	4	1	3	2	1	2			2	
Março		2	**86**	1	3	2		3		1		2
Abril		1	1	**89**	5			1	1	1	1	
Maio	1		2	2	**78**	3	1	2	2	5	2	2
Junho	5		1	1		**69**	14	1	2	2	4	1
Julho	1	1	1		3	17	**65**	3	1	5	2	1
Agosto	2	3		3	1		2	**79**			1	9
Setembro	1	3		1	1	3		1	**64**	19	6	1
Outubro		2	1		2		1		13	**79**	2	
Novembro	2	1		2		3		1	4		**86**	1
Dezembro	4	3			1			8	3	5	2	**74**

Table 5. Confusion matrix obtained with topological feature set (NN-T).

Month	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**53**	14	2	1	5	11	5		3		1	5
Fevereiro	7	**71**	6	1		4	1	2	1	3	2	2
Março	1	2	**82**	3	3	2	1	3				3
Abril	2	2	1	**87**	3	1			2	1	1	
Maio	4		2	8	**56**	2	3	3	7	6	8	1
Junho	8	1		1	1	**72**	4	1	5	1	5	1
Julho	1		1	2	1	10	**79**	2	2	1		1
Agosto		3	1	2	1	3		**74**		1	4	11
Setembro	1	2		1		3			**76**	7	9	1
Outubro	1		1			1	1	1	8	**81**	5	1
Novembro	4	5	1		2	4		1	9	2	**67**	5
Dezembro	1	1	1	1	2	2	2	5	2		5	**78**

i.e., classes that present a common sub-string. To improve the results a selection of representative features inside each cluster is necessary. The defined feature sets are consistent, the perceptual set producing better results than the others. This fact shows that the incorporation of human reading knowledge is significant to obtain good results in handwriting recognition.

To evaluate the combined potential of the features the individual networks outputs (one for each feature set) were combined by either multiplying or averaging the confidence measure of each class. The class that presents the maximum value after this operation is the class recognized. Table 6 shows the results obtained with this experiment. lt shows that the best results are produced by the perceptual and directional feature sets combination.

Table 6. Average recognition rate obtained using sets combination.

	By multiplication	By averaging
Sets	Recognition rate (%)	Recognition rate (%)
P-D	87.0	**87.2**
P-T	85.7	85.2
D-T	82.2	81.9

4 Discussion and Conclusions

In this section we try to answer the questions initially formulated. The results show that the perceptual features set is the one that produces the best recognition rate for the baseline system presented. This indicates that the classifier presents an information processing mechanism similar to the human reading system to discriminate different words, which relies heavily on the discriminating power of perceptual features. Therefore, the incorporation of human reading knowledge, represented by the perceptual features, on artificial reading systems seems to be worthwhile. However, that feature set is not always enough, specially for words that either do not have ascenders/descenders or have them in the same position. For those case, good classification rates can only be obtained with the use of other features.

To conclude, this paper presented a baseline system for recognition of the Portuguese handwritten names of the months. This system was used to evaluate three different feature sets, obtaining a best average recognition rate of 87.2%. Due to the particularity of the database it is difficult to compare this work with others reported in the literature. Morita et alli [7, 8] utilizes the same lexicon from a distinct database, having achieved recognition rates of 81.7% and 90.0%, using a global and an analytic approach, respectively. Kim et alli [5] have combined HMM and MLP classifiers for a similar lexicon, extracted from the CENPARMI database. For the MLP classifier alone with 12 classes, corresponding to the full english month names, a recognition rate of 79.4% was acchieved. For the combined HMM-MLP classifier this rate goes to 88.8%. This comparison shows that our baseline system obtain rates compatible or better than other systems recently reported. Our future work will focus on the analysis of the combination of different classifiers. The goal is to improve understanding of the confusion mechanism.

Acknowledgements

The authors would like to thank CNPq and CAPES/PROCAD for the financial support of this work.

References

[1] A. Zell et al. *SNNS - Stuttgart Neural Network Simulator, User Manual, Version 4.2*. University of Stuttgart, 1994.

[2] Côté, M. *Utilisation d'un Modèle d'Accès et de Concepts Perceptifs pour la Reconnaissance d'Images de Mots Cursifs*. PhD thesis, École Nationale Supérieure des Télécommunications, France, 1997.

[3] Freitas, C. O. de A., Bortolozzi, F. e Sabourin, R. An Approach Based on Mutual Information for Feature Set Validation. In *ICDAR'2001*, pages 665–669, Seattle - USA, Setembro 2001. International Conference on Documents Analysis Recognition.

[4] Guillevic, D. *Unconstrained Handwriting Recognition Applied to the Processing of Bank Cheques*. PhD thesis, Department of Computer Science at Concordia University, Canada, 1995.

[5] Kim, J. H., Kim, K. K., Nadal, C. P. e Suen, C. Y. A Methodology of Combining HMM and MLP Classifiers for Cursive Word Recognition. In *ICPR'2000*, Barcelona - Spain, Setembro 2000. International Conference on Pattern Recognition.

[6] Madhvanath, S. e Govindaraju, V. The Role of Holistic Paradigms in Handwritten Word Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2):149–164, 2001.

[7] Morita, M. E., El Yacoubi, A., Bortolozzi, F. e Sabourin, R. Handwritten Month Word Recognition on Brazilian Bank Cheques. In *ICDAR'2001*, pages 972–976, Seattle - USA, Setembro 2001. International Conference on Documents Analysis Recognition.

[8] Morita, M. E., Lethelier, E., El Yacoubi, A., Bortolozzi, F. e Sabourin, R. An HMM-based Approach for Date Recognition. In *DAS'2000*, pages 233–244, Rio de Janeiro - Brazil, Dezembro 2000. International Workshop on Document Analysis Systems.

[9] Parker, J. R. *Algorithms For Image Processing and Computer Vision*. Jonh Wiley & Sons, 1997.