Integration of Contextual Information in Handwriting Recognition Systems

F. Grandidier [1,2] R. Sabourin [1,2] C.Y. Suen [1]

[1] CENPARMI, Concordia University
1455 de Maisonneuve Blvd West
Montréal, H3G 1M8, Canada
{grandidier, suen}@cenparmi.concordia.ca

[2] LIVIA, ETS
1100, rue Notre Dame Ouest
Montréal, H3C 1K3, Canada
Robert.Sabourin@etsmtl.ca

Abstract

This paper investigates different strategies allowing integration of contextual information during the feature extraction stage of a cursive handwriting HMM-based recognition system. First we propose to use linear discriminant analysis (LDA) in order to integrate the class information during feature set building. Secondly several zoning strategies are used to integrate local contextual information. Finally, a weighting technique is proposed in association with zoning with the aim of integrating handwriting style. Some experiments were carried out and the results show the interest of the proposed strategies.

1. Introduction

The problem of handwriting recognition has been studied for several decades but is still open [6]. The main difficulty lies in the large variability of the handwriting signal, especially for applications without writing constraints like bank cheque processing and mail sorting.

Different sources of variability can be pointed out. The writing tool and the support used can modify the shape of characters. The writer mood can also influence his writing, but the main variability source is the writer himself. As proof, handwriting is used for authentication purpose. Each writer has his own style. Tappert *et al.* [9] identify five different styles: boxed discrete, spaced discrete, run-on discretely written, pure cursive script and mixed cursive, discrete run-on discrete. However, their study was done for on-line handwriting systems. In such industrial application as mail sorting or bank cheque processing, the three last styles are the most used and correspond to natural writing when the two first are usually associated with constrained writing, like on some forms.

Another problem comes from the segmentation step. In large vocabulary application, like mail sorting, the model must be done at the character level. Thus a segmentation process must be carried out. Segmentation is a complicated task still not solved. Generally, over-segmentation of word in graphemes is preferred ; then dynamic programming is used in order to evaluate character hypothesis.

At this level, the knowledge of the grapheme shape may not help in character identification. Then the main difficulty lies in extracting discriminant information for the classification step. Contextual information becomes really important in this scheme. It will help to recombine the graphemes in characters and improve the discrimination power of features.

In this paper we will introduce several strategies allowing improvements of the discriminant power of feature sets by integrating contextual information. In the next section we will expose our motivation. Then, in section 3, the proposed strategies will be developed. Several experiments were carried out ; they will be described in section 4. Finally some conclusions will be presented.

2. Problem statement

Our main objective is to improve the performance of the handwriting recognition system described in [2]. This is a discrete HMM-based off-line handwriting recognition system, using an analytic approach with explicit segmentation. After some pre-processing and segmentation steps, the basic units processed by this system are graphemes, *i.e.* part of a character, a whole character or several characters fully connected. Each grapheme is represented by two symbols from different sets of feature. The first one E_1 (27 symbols) is based on global features: ascenders, descenders and loops, and is more dedicated to cursive handwriting. The second feature set E_2 (14 symbols) is based on the analysis of horizontal and vertical contour transition histograms of each segment. It better characterizes hand-printing. Those two sets are combined by Cartesian multiplication, thus allowing their integration into the recognition system. The resulting set contains 378 features. In order to train and test this system, three data corpuses are used: 12023 city names for learning, 3475 for validation, and 4674 for testing. Performances of this system, using a combination of the feature sets $E_1 \times E_2$, are shown in Table 1, according to the different writing styles and lexicon sizes. Samples treated as *hand-printed* are composed of upper-case letters only when those designated as *cursive* are mainly composed of lower-case letters, only the first letter of each word can be upper-case. *Mixed* clusters all the other samples. There are approximatively 38% of hand-printed samples in our corpus and respectively 51.5% and 10.5% of cursive and mixed.

Lexicon size	10	100	1 000	10 000
All Samples	**98.9**%	**95.7**%	**89.5**%	**77.3**%
Hand-printed	99.5%	97.9%	93.4%	82.0%
Cursive	98.3%	93.6%	85.4%	72.3%
Mixed	99.7%	98.0%	95.6%	85.1%

Table 1. Performances of the original system.

An analysis of those results shows that recognition rates associated with cursive samples are lower than other styles. This performance difference is inherent in handwriting styles (samples from the same city name are shown in Figure 1). Shapes of hand-printed samples are naturally more discriminant than cursive ones. Usually cursive handwriting is realized in a more rapid movement than hand-printed, resulting in more ambiguous shapes. Moreover, after the segmentation process it is difficult, indeed impossible, to distinguish graphemes coming from different letter, *e.g.* "`i`, `u`" or "`n`, `m`". Those observations lead us to base our reflection on the analysis of the system behavior over cursive samples.

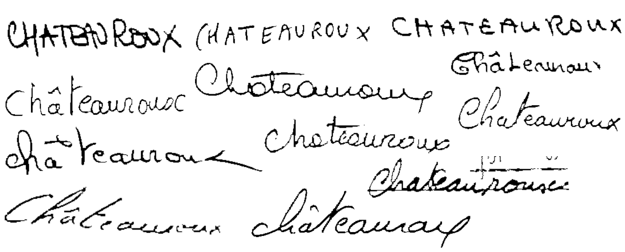
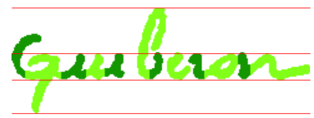


Figure 1. Several examples of the same city name.

In order to illustrate difficulties associated with cursive sample recognition, an example is shown in Figure 2. We can observe that 75% of graphemes are characterized by the same feature "-", corresponding to the absence of ascenders, descenders and loops. After system training, we noticed that more than 50% of all graphemes are characterized by this feature. Although the definition of E_1 is based on human perception of handwriting, this feature set is not discriminative enough to perform good classification. The combination with the feature set E_2 allows a better information representation but still not discriminative enough.



h B - - - H- - - o - -

Figure 2. A segmented city name sample and the corresponding E_1 feature sequence.

To improve recognition of cursive samples, one way is to define features with a better discrimination power. Another possibility is to take in account some contextual information.

3. Integration of contextual information

What is contextual information? In the case of our handwriting recognition system using an explicit segmentation, the basic entity is grapheme. Thus we are talking about the context of graphemes. It can come from the neighboring graphemes and also from the whole word image. We are dealing with low level contextual information, not high level contextual information as grammatical or semantic information.

In order to integrate this kind of contextual information during the feature extraction step, several strategies have been developed.

3.1. Introducing the class information during feature extraction

The task of pattern recognition lies in assigning a sense or a class to a given shape, according to some observations or measures evaluated on this shape. Thus the knowledge of this class, for a set of training samples, can help to build some discriminant feature sets. One way is to perform a linear discriminant analysis (LDA) [1]. This statistical data analysis seeks components (or directions) that are useful for discriminating data, according to modalities of a qualitative variable. By choosing the different classes of the recognition problem, the feature space can be transformed into a new space where samples are clustered according to classes. As principal component analysis (PCA), LDA can be used to reduce the dimensionality, but we did not investigate this possibility here.

For real problems, the class information is usually not available during the feature extraction step. In our application, the database is labelled but at the city name level *i.e.* only the character string present in each image and not the label of each grapheme.

In order to obtain grapheme labels, we propose to use a potential of Markovian modeling. After system training we can obtain the alignment between graphemes and model classes conducting to the better observation probability *via* the Viterbi procedure. Then this step can be used to label each grapheme with the corresponding modeling class.

An effective recognition system is needed in order to carry out this procedure. Thus different steps are required in order to integrate the class information in a HMM-based recognition system :

1. ***System Building***
2. ***Grapheme Labelling*** using the backtracking step of the Viterbi algorithm
3. ***LDA Calculation*** resulting in a transformation matrix
4. ***Sample Projection*** using the transformation matrix
5. ***Improved System Building***

In our discrete framework, first and fifth steps are achieved by constructing a feature set from a given feature space by vector quantization. It can then be used to carry out the system training.

A recognition system is used to obtain grapheme labels. Thus the labelling is not guaranteed to be correct. However it is an acceptable alternative to manual labelling. The verification of several samples showed that alignments given by the first system is generally accurate. Thus we assume that globally the obtained labelling of graphemes is correct.

3.2. Using zoning strategy

One way to integrate contextual information is to divide patterns into several parts and to extract features from each of them individually [8]. The resulting information representation allows to obtain the relative positions of specific characteristics and to integrate global information. For the particular case of cursive handwriting, the presence of ascenders and descenders can be held by this strategy. The resulting representation of information is more rich and precise.

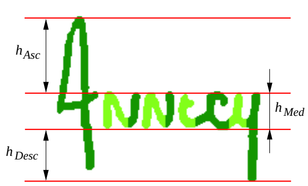
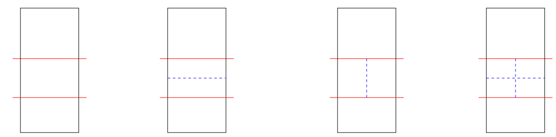


Figure 3. Definition of writing zones.

In order to divide graphemes, we propose to use the writing zones (ascender, median and descender zones, see Figure 3) defined during pre-processing steps. This solution guarantees extraction of similar information (or features) for samples of different sizes. From the three writing zones, the median one is certainly the more informative. Simon [7] states that word body is the regular part of the writing and ascenders and descenders are singularities. For a cursive sample, the median zone contains all the information about the connection or ligature between graphemes.

As the median zone is more informative, different zoning strategies (see Figure 4) are proposed in order to obtain a more accurate information representation from the word body.



(a) 3 zones (b) 4 zones Hor. (c) 4 zones Vert. (d) 6 zones

Figure 4. Different zoning strategies.

3.3. Integrating handwriting style during extraction

The zoning strategy proposed allows to extract more precise information for cursive samples. However, for hand-printed one it is not so obvious. Indeed, for this style only the median zone must be significant. As writing style is unknown during pre-processing steps, the determination of the three zones is done according to some heuristics.

An analysis shows that the size of the ascender zone is bigger than half of the median zone for 50% of the hand-printed samples. This observation means that the ascender zone holds a significant proportion of information concerning hand-printed samples. Similar conclusions were drawn about the descender zone. Thus the use of a zoning strategy can bring confusions during recognition of hand-printed samples.

During the feature extraction step, the knowledge of sample style can help to obtain a more discriminative information representation, leading to better recognition results. However this information is really difficult, indeed impossible, to obtain with high reliability. Thus we proposed an alternative : define a weighting strategy in order to reduce the influence of zoning when hand-printed samples are processed.

Our reflection is based on some observations : for hand-printed samples the height of ascender and descender zones is smaller than the median one. We evaluate two height ratios ρ_A and ρ_D defined as :

$$\rho_A = \frac{h_{Asc}}{h_{Med}} \qquad \rho_D = \frac{h_{Desc}}{h_{Med}} \tag{1}$$

where h_{Asc}, h_{Med} and h_{Desc} are defined in Figure 3. These quantities were evaluated over our entire training corpus. Their analysis showed that 15% of the hand-printed samples have an ascender zone bigger than median one versus 80% for cursive samples. Only 5% of the hand-printed samples has a descender zone bigger than median one versus 35% for cursive samples.

This information was taken into account to define a weighting strategy of feature vector components. Let $\mathbf{x}_i = [x_1^i, x_2^i, \ldots, x_N^i]^t$ be the vector associated with one sample in a N-dimension space where all x_j^i are in the range $[0, 1]$. Let $\mathbf{z}_k = [z_1^k, z_2^k, \ldots, z_N^k]^t$ be a centroid resulting of the vector quantization process. Centroid components are obtained from this equation :

$$z_j^k = \frac{1}{N_{Vk}} \sum_{i=1}^{N_{Vk}} x_j^i \qquad (2)$$

where N_{Vk} is the number of samples related to centroid $\mathbf{z}_k$. Let n_A, n_M and n_D be the number of components associated with ascender, median and descender zones respectively ($N = n_A + n_M + n_D$). Let $\mathbf{p}_i = [p_1^i, p_2^i, \ldots, p_N^i]^t$ be the weighting vector associated with sample $\mathbf{x}_i$. Its components are obtained in the following way :

$$\begin{cases} p_k^i = \min(1, \rho_A^i) & \text{if} \quad 0 < k \leq n_A \\ p_k^i = 1 & \text{if} \quad n_A < k \leq n_A + n_M \\ p_k^i = \min(1, \rho_D^i) & \text{if} \quad n_A + n_M < k \leq N \end{cases} \qquad (3)$$

This weighting vector is taken into account directly during the vector quantization process by integrating it during distance evaluation between samples and centroids. The used metric is the Euclidean distance. The weighting version can be written as :

$$d_{Ep}(\mathbf{x}_i, \mathbf{z}_k) = \sqrt{\frac{1}{P_i} \sum_{j=1}^{N} p_j^i \left(x_j^i - z_j^k\right)^2} \qquad (4)$$

$$P_i = \sum_{j=1}^{N} p_j^i = (\rho_A^i \times n_A) + n_M + (\rho_D^i \times n_D) \leq N \quad (5)$$

The advantage of this approach lies in the fact that data extracted from graphemes (the feature vector) are not modified. However this weighting strategy can not be used in association with LDA because in the resulting space, each component is a linear combination of the original ones. In order to compensate this problem, another approach is proposed : weight directly feature vector components. This approach generates a new vector $\mathbf{y}_i = [y_1^i, y_2^i, \ldots, y_N^i]^t$ where each component is obtained from the original $\mathbf{x}_i$ vector :

$$y_j^i = p_j^i \times x_j^i \qquad (6)$$

where the weighting vector is the same (see equation 3).

The two weighting strategies will conduct to build two distinct recognition systems

4. Experiments

In order to evaluate the influence of the proposed strategies, we developed two new numerical feature spaces. The first is based on concavity extraction (CCV). White pixel labelling as proposed in [3] was used. Only continuous concavity configurations with more than 2 rays are used to build the feature vector. The black pixel ratio is also taken in account.

The second feature space used was proposed by Oh [5] and is called Directional Distance Distribution (DDD) feature. By opposition to the original method, we do not perform image size normalization before evaluating distances. Instead, distances are normalized according to the size of the analysis zone : width for horizontal distances,*etc*. Tilling strategies are not used. As required, the zone of analysis is divided into several sub-zones. We choose to cut each side by 2, resulting in 4 sub-zones.

Our system uses a discrete modeling, a vector quantization is necessary to build feature set from a continuous feature space. We chose the LBG algorithm [4] for speed and simplicity of its use. One associated constraint is the cardinality of the resulting feature set : it must be a power of 2. Feature sets with different cardinality were evaluated. Finally, all results presented in this section are obtained using 256 features. This size leads usually to the higher recognition rates. Moreover we used a lexicon size of 1 000 during testing.

		Before LDA			After LDA		
Lines	Samples	All	Hand.	Curs.	All	Hand.	Curs.
1	CCV-BB	88.3	94.1	82.7	89.0	95.2	83.1
2	DDD-BB	83.6	92.1	75.7	92.7	96.5	89.1
3	CCV-3z	89.6	92.6	86.4	89.9	94.2	85.7
4	DDD-3z	86.3	88.8	83.0	93.2	95.6	90.5
5	CCV-4zH	90.9	94.2	87.3	90.8	95.2	86.6
6	DDD-4zH	87.5	90.7	83.6	93.6	95.9	91.3
7	CCV-4zV	89.9	93.1	86.2	90.4	95.2	85.9
8	DDD-4zV	88.7	92.2	85.0	93.8	96.8	90.9
9	CCV-6z	90.9	94.4	87.3	91.0	95.2	86.7
10	DDD-6z	88.5	93.5	83.5	92.9	96.1	89.8

Hand.: Hand-Printed; Curs.: Cursive; BB: Bounding Box; 3z: 3 zones

Table 2. Evaluation of class integration and zoning

Table 2 presents the evaluation of class information integration during feature extraction and also the effect of the different zoning strategies on our recognition system. In table 3 we can observe results obtained when weighting strategies are used in association with zoning. The label "Weighting 1" stands for the first weighting strategy described, without modification of feature vector components. All values presented are recognition rates in %.

The analysis of lines 1-2 of table 2 (features extracted from bounding boxes) shows that the use of LDA algorithm leads to improvements of recognition rates whatever the feature space used and whatever the writing style. From this observation we can conclude that the proposed strategy must be used in a discrete HMM-based recognition system when features are extracted from bounding boxes.

A comparison of recognition rates from table 2 (all lines) and table 1 (column 4) shows that the different feature sets based on CCV and DDD lead to better performance than the combination $E_1 \times E_2$ with less features.

From the column marked "All", we can notice that zoning strategies lead to an improvement of recognition rates, before and after LDA and whatever the feature set used. If we analyze performances according to writing style (columns marked "Hand." and "Curs." in table 2), conclusions are not the same. As expected the zoning strategies are favorable only to cursive samples. For hand-printed ones, a loss can be observed. The reason is that writing zones have no real significance for these samples. We can not say that one zoning strategy is better than the others : "6z" seems the better for concavity features (line 9 in table 2) but "4zV" seems better than the DDD features (line 8 in table 2). Thus we can conclude that the zoning strategy must be chosen according to the nature of used features.

		Before LDA			After LDA		
Lines	Samples	All	Hand.	Curs.	All	Hand.	Curs.
1	CCV-3z	89.6	92.6	86.4	89.9	94.2	85.7
2	Weighting 1	89.5	92.8	85.8	89.7	94.2	85.3
3	Weighting 2	89.7	93.2	86.2	90.0	94.5	85.9
4	DDD-3z	86.3	88.8	83.0	93.2	95.6	90.5
5	Weighting 1	87.1	89.8	83.9	93.3	95.8	90.7
6	Weighting 2	86.1	89.3	82.3	93.1	95.5	90.7

Table 3. Evaluation of weighting strategies

In order to avoid the effect of zoning on hand-printed samples, we proposed two weighting strategies. From columns marked "All." of table 3 we can observe that global recognition rates are not really affected by these strategies. The analysis according to writing style (columns marked "Hand." and "Curs." in table 3) shows that both weighting strategies lead to some improvements of recognition rates associated with hand-printed samples. However both have a bad effect on cursive sample recognition since we can observe a loss in performance. The reason is that for some cursive samples the weighting strategies reduce the influence of significant ascenders or descenders and thus introduce confusion to the system. Another remark about hand-printed samples is that even if the use of weighting strategies allows to reduce the influence of zoning, recognition rates are far from those obtained when features are extracted from bounding boxes (94.1%).

5. Conclusions

In this paper we proposed several strategies to integrate of contextual information in a handwriting recognition system. First a scheme to introduce the class information during feature extraction, in a discrete HMM-based recognition system, was developed. The observed performance improvement indicates clearly the interest of using this strategy during the building of feature sets.

Then we evaluated zoning strategies to introduce contextual information during the feature extraction step. This technique leads globally to performance improvements. However only recognition rates associated with cursive samples are increased. Those associated with hand-printed samples suffer from this technique.

In order to reduce the loss of performance on hand-printed samples, we proposed two weighting strategies, allowing to take into account the writing style during feature set building. This technique allows to reduce the influence of zoning on hand-printed sample recognition. However we can observe a reduction in performance associated with cursive samples. A better discrimination between cursive and hand-printed samples must be carried out in order to use these weighting strategies.

Acknowledgements

This work was supported by the Service de Recherche Technique de La Poste (SRTP) at Nantes, France, the Ecole de Technologie Supérieure and the Centre for Pattern Recognition and Machine Intelligence at Montréal, Canada.

References

[1] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, USA, 2001. 2nd edition.

[2] A. El-Yacoubi, M. Gilloux, R. Sabourin, and C. Suen. An HMM-based approach for off-line unconstrained handwritten word modeling and recognition. *IEEE Trans. on Pattern Analysis and Machine Recognition*, 21(8):752–760, 1999.

[3] L. Heutte. *Reconnaissance de caractères manuscrits : application à la lecture automatique des chèques et des enveloppes postales*. PhD thesis, Université de Rouen, 1994.

[4] Y. Linde, A. Buzo, and R. Gray. An algorithm for vector quantizer design. *IEEE Trans. on Communication*, 28(1):84–95, 1980.

[5] I. Oh and C. Suen. Distance features for neural network-based recognition of handwritten characters. *International Journal on Document Analysis and Recognition (IJDAR)*, 1(2):73–88, 1998.

[6] R. Plamondon and S. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. on Pattern Analysis and Machine Recognition*, 22(1):63–84, January 2000.

[7] J.-C. Simon. Off-line cursive word recognition. *Proceedings of the IEEE*, 80(7):1150–1161, July 1992.

[8] C. Suen, J. Guo, and Z. Li. Analysis and recognition of alphanumeric handprints by parts. *IEEE Trans. on Systems, Man, and Cybernetics*, 24(4):614–631, April 1994.

[9] C. Tappert, C. Suen, and T. Wakara. The state of the art in on-line handwriting recognition. *IEEE Trans. on Pattern Analysis and Machine Recognition*, 12(8):787–808, 1990.