Structural information implant in a context based segmentation-free HMM handwritten word recognition system for Latin and Bangla script

Abstract

In this paper, an improvement of a 2D stochastic model based handwritten entity recognition system is described. To model the handwriting considered as being a two dimensional signal, a context based, segmentation-free Hidden Markov Model (HMM) recognition system was used. The baseline approach combines a Markov Random Field (MRF) and a HMM so-called Non-Symmetric Half Plane Hidden Markov Model (NSHP-HMM). To improve the results performed by this baseline system operating just on low-level pixel information an extension of the NSHP-HMM is proposed. The mechanism allows to extend the observations of the NSHP-HMM by implanting structural information in the system. At present, the accuracy of the system on the SRTP[1] *French postal check database is 87.52% while for the handwritten Bangla city names is 86.80%. The gain using this structural information for the SRTP dataset is 1.57%.*

1. Introduction

After a remarkable success of the HMMs [7] in speech recognition, the model was borrowed and used with the same success in handwriting recognition domain too. The power of such a model resides in its capability to track the temporal aspect of the modeled signal, which is impossible in a connectionist approach. While the speech can be considered as a 1D signal, the handwriting is a much more complex. As the writing has its temporal aspect, the one dimensional models are not able to take into account this information. The HMM based models are very interesting in handwriting as are able to stock such information. In the last decade a growing interest was observed to develop new formalisms to bypass the 2D constraint. The literature proposes different methods like PHMMs [3, 5] or totally 2D models using MRF as described in [1, 4, 8].

In this paper, we describe a general method that allows to insert extra information in the system to improve its recognition capacity. The proposed model is an extension of a context based, segmentation-free HMM approach operating on pixel level described in [1, 8]. While several systems use low level pixel information, other systems use high level features like structural features, our idea is to combine these different information in the framework of the NSHP-HMM.

In order to exploit totally the pixel information coming from the analyzed shape we extended the observation of the HMM by joining the color of the pixel with its structural nature. This coupling of low-level information with a high-level one, coming from the same pixel gives a new dimension for the observations performed by the NSHP-HMM. Since our model should be able to recognize different scripts (Latin, Bangla, etc.) the method is designed to be a general one, able to exploit different type of information.

Rest of the paper is organized as follows. In Section 2 the baseline NSHP-HMM system is presented while Section 3 describes the extension of the system by the implant of the structural information in the HMM observations. Section 4 describes the used databases and the results obtained by the baseline and the extended system. Finally, Section 5 allows some discussions and conclusions on the proposed method.

2. The baseline NSHP-HMM system

The baseline system so called NSHP-HMM for handwritten word recognition has been described in [8]. The technique operates in a holistic manner, on pixels coming from height normalized images which are perceived as a random field realizations. The context based segmentation-free stochastic method proposed by the authors avoids the errors coming from the different segmentation techniques based mainly on heuristics and the errors coming from the pseudo 2D or planar HMM approaches which are sensitive to the major distortions.

2.1. Formal description

A detailed formal description of the NSHP-HMM can be found in [1, 8]. For this work, it is not necessary to describe

[1] Service de Recherche Technique de la Poste

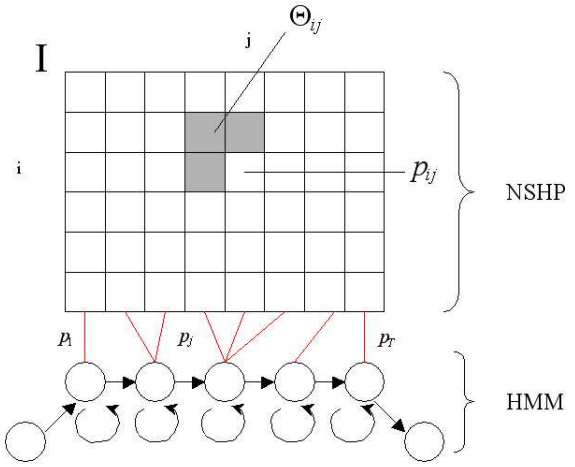

Figure 1. The scheme of the NSHP-HMM

the whole NSHP-HMM model. Just some relevant parts will be described.

The formal NSHP-HMM can be described as follows:

$V = \{0, 1\}$ or $\{black, white\}$ the set of observable symbols

$S = \{s_1, \cdots, s_N, \Gamma, \Lambda\}$ the set of normal states and two specific states

$A = \{a_{ij} \cup \{a_{\Gamma i}, a_{i\Lambda}\}\}_{1 \leq i,j \leq N}$ where

$a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i); 1 \leq i, j \leq N$

$a_{\Gamma i} = P(q_1 = s_1 \mid \Gamma), a_{i\Lambda} = P(\Lambda \mid q_T = s_i)$

$B = \{b_i(y, \Theta, c)\}$ is the probability to observe in a state i (s_i) a pixel of color c at height y knowing the neighborhood Θ_{ij} where $s_i \in S_i, s_i \notin \{\Gamma, \Lambda\}$

To simplify the notation we denote by $b_i(O_t)$ the the column observation probability observed by the HMM.

$$b_i(O_t) = \prod_{y=1}^{M} b_i(y, \Theta_{ij}, c) = \prod_{y=1}^{M} P(X_{iy} \mid X_{\Theta_{iy}}, q_i) \quad (1)$$

N denotes the number of states, while M is the height of the NSHP-HMM model. A basic scheme of the NSHP-HMM operating on pixels is presented in Fig.1. Let I be the image having m rows and n columns observed by the NSHP. The joint field mass probability $P(I)$ of the image I can be computed following the chain decomposition rule of conditional probabilities:

$$P(I) = \prod_{j=1}^{n} \prod_{i=1}^{m} P(X_{ij} \mid X_{\Theta_{ij}}) \quad (2)$$

Let the conditional pixel probability of a pixel (i, j) be denoted by p_{ij}:

$$p_{ij} = P(X_{ij} \mid X_{\Theta_{ij}}) \quad (3)$$

and the column probability:

$$P_j = \prod_{i=1}^{m} p_{ij} \quad (4)$$

Considering the equations (3) and (4) the equation (2) can be computed as follows:

$$P(I) = \prod_{j=1}^{n} P_j \quad (5)$$

The notation used in (2-5) is similar as depicted in the Fig. 1. In this case P_j denotes the column observation given by the equation (1). In order to simplify the notation in the further discussions just the notation (4) will be used.

The results obtained by the baseline system (85.95% for the Latin and 96.40% for Bangla) hawe shown the model limits. Using just low level pixel information seems to be not sufficient to reach higher scores. This insufficiency is coming from the MRF and its re-estimation.

3. The NSHP-HMM extension with structural information

To improve the system we propose to extend the observations of the model described in (4) by inserting high-level information coming from the structural nature of the pixels. This extra information allows to precise the quantity and quality of the information perceived by the HMM. This implant of high-level information can be done inside or outside the model. The challenge of this approach is how to introduce such information in the model or how to tranform the model itself to accept such extra information.

3.1. The different weight mechanisms

The possible structural information carried out by each pixel can be transformed in some kind of weight. This weight derived from the structural information could be descriptive for each pixel of a column (e.g. each conditional pixel probability can be weighted individually by calculating a weight for each pixel) or factorized along the column (e.g. the whole column observation probability is weighted by a weight calculated in function of the different pixels' structural capacity belonging to the column).

This weight factor can be interpreted:

- If the weight is at pixel level, we can accentuate a pixel giving it an extra power, which can be translated in physical terms like the HMM have seen the same pixel several times or with a weight, where the weight is the importance of the pixel among the others.

- If the weight is global for the column, we can accentuate a column giving it an extra power which can be translated in physical term that the HMM has seen the same column several times or with a weight, where the weight is the importance of the column among the others.

This weighting mechanism should not disturb neither the Baum-Welch training nor the Viterbi decomposition. This means that if such weighting is applied the basic Markov constraints should be satisfied [7].
Let denote p^{inf} the weight derived from the extracted structural features. Different weighting mechanism are proposed in function of the global or local nature of the weight.

1. If the structural weight is global for the column j we propose to transform the equation (4) into:

$$\overline{P_j} = \left(\prod_{i=1}^{m} p_{ij} \right) \times p_j^{inf} \tag{6}$$

   where p_j^{inf} is considered as being the weight calculated for the column j considering all the pixels (i, j) and their structural properties.

2. If the structural weight is local for the pixel (i, j) we propose to transform the equation (4) into:

$$\overline{P_j} = \prod_{i=1}^{m} (p_{ij} \times p_{ij}^{inf}) \tag{7}$$

   where p_{ij}^{inf} is considered as being the weight calculated for the pixel (i, j) belonging to the column.

In the same manner, we can establish two other equations:

3. If the structural weight is global for the column j we propose to transform the equation (4) into:

$$\overline{P_j} = \left(\prod_{i=1}^{m} p_{ij} \right)^{p_j^{inf}} \tag{8}$$

   where p_j^{inf} is considered as being the weight calculated for the column j considering all the pixels (i, j) and their structural properties.

4. If the structural weight is local for the pixel (i, j) we propose to transform the equation (4) into:

$$\overline{P_j} = \prod_{i=1}^{m} (p_{ij})^{p_{ij}^{inf}} \tag{9}$$

   where p_{ij}^{inf} is considered as being the weight calculated for the pixel (i, j) belonging to the column.

Generally the weight p^{inf} can be calculated considering the quantity and the quality of the information. As just two high-level features were extracted, we considered just the quantity of the information without making any difference between pixels having different characteristics.

As the approaches proposed in (7) and (9) have some technical limitations, we limit the further discussions to the equations (6) and (8).
In order to obey the Markov constraints a normalization process is necessary. As the structural information is extracted from the height normalized images the normalization is ensured. In order to distinguish between a pixel column where no structural point exists and a column where pixels carrying structural information, the weight p_j^{inf} for (6) is calculated as follows:

$$p_j^{inf} = \frac{1}{nbFeature + 1} \tag{10}$$

where $nbFeature$ denotes the number of pixels having a structural feature in the column j. Some other normalization schemes were also tested. Finally, our column based observation for the structural NSHP-HMM is:

$$\overline{P_j} = (\prod_{i=1}^{m} p_{ij}) \times \frac{1}{nbFeature + 1} \tag{11}$$

In the equation (8) the p_j^{inf} is calculated as follows:

$$p_j^{inf} = \begin{cases} \eta & nbFeatures > \kappa \\ 1 & \text{otherwise} \end{cases} \tag{12}$$

where $nbFeature$ denotes the number of pixels having a structural feature in the column j, while η and κ are some parameters set to suitable values based on trial runs. In that case the column observation can be described as follows:

$$\overline{P_j} = \begin{cases} \left(\prod_{i=1}^{m} p_{ij} \right)^{\eta} & nbFeatures > \kappa \\ \prod_{i=1}^{m} p_{ij} & \text{otherwise} \end{cases} \tag{13}$$

Once we have defined these observations defined by (11) and (13) the same train/test mechanism developed and described in [1, 8] can be used. We used as extra information the structural features extracted from the different word shapes, as we consider than these high-level features are sufficiently descriptive for handwriting. Moreover, many HWR systems use such features to discriminate the different forms. As the method is general any other kind of information can be used instead of the information selected by us.

Concerning the model complexity, the memory complexity of the new model will be similar to the case of the

former system $\mathcal{O}[N(N + 2^V Y]$ while the computational complexity will grow in function of the features which will be extracted. For the calculus we have considered a model having N normal states, analyzing Y pixels in each column using a neighborhood of order V.

4. Results

4.1. Databases description

The tests were performed on two different handwritten word datasets. The Latin one is the SRTP dataset containing handwritten French bank cheque amounts. The 7031 images are distributed not uniformly in 26 classes. The 26 classes correspond to the different French words describing the different legal amounts. The second dataset is a Bangla city name database containing Indian city names written in Bangla script, collected in Kolkata, West Bengal, India. The dataset contains 7500 postal documents and we have used just the different Bangla city names extracted manually. We have identified 76 different city names. In order to have a uniform distribution of city names (100 images/class) some extra images were necessary. In both cases the image acquisition was off-line at 300 dpi. We have used 66% of the images to train the systems and the 34% remaining images were used to test the system.

4.2. Image normalization

As the NSHP-HMM operates on pixel columns is necessary to perform operations like angle correction, slant correction, as the model is sensitive to such kind of distorsions. In order to reduce the computational complexity of the model a differential height normalization has been used based on the middle zone of the writing. The normalization gives as result images with the same height but with proportionally different widths.

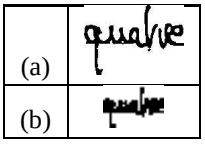

Figure 2. (a) original image and (b) normalized image of the word "four" in French

The original threshold mechanism to find the middle zone was adapted to handle Bangla script also. While for Latin, besides the middle zone, the upper part and lower part contains the ascenders respectively descenders, in Bangla, the major part of the information is located in the middle zone and the lower part of the writing. A detailed description of the normalization can be found in [1].

4.3. Test results

We have tested the different methods on SRTP and the Bangla dataset.

Method	SRTP	Bangla
Classic	85.95%	86.40%

Table 1. Recognition scores of the NSHP-HMM based on pixel information

The overall recognition accuracy using the classical recognition scheme for the different datasets is given in Table 1. We can observe that the system was not sensitive to the vocabulary opening. The model gives more or less the same accuracy for the SRTP (26 class) and Bangla word dataset (76 class) which is a considerable for such a holistic approach.

To test the implant mechanism proposed in this paper, some feature extraction was necessary. As our main goal was to propose a new and general mechanism to implant extra information in the NSHP-HMM model, we limited our feature extraction to the descenders and ascenders (two basic feature often used in the literature). The extraction of the ascenders and descenders is based on the middle zone of the writing already used for the normalization. A pixel was considered as being a structural pixel if it belongs to an ascender or a descender.

The performance of the NSHP-HMM using the structural information is as follows. The achieved accuracy using the observation defined by the equation (11) is 87.52% for the SRTP dataset and 86.80% for the Bangla city name dataset. Using the definition given by equation (13) the achieved recognition is 86.39% for the SRTP dataset and 86.52% for the Bangla dataset. The results given by the different improved (extended) observations are summarized in Table 2.

Method	SRTP	Bangla
Improvement1	87.52%	86.80%
Improvement2	86.39%	86.52%

Table 2. Recognition scores of the NSHP-HMM based on pixel and structural information

The improvement reached by the implant of the ascender and descender in the column observation in the NSHP-

HMM is much more considerable (1.57%) in the case of the SRTP database. For the Bangla city name dataset the improvement is just 0.4%. The difference is due to the nature of the scripts and the used structural features. While in case of the SRTP bank cheque dataset, the words are Latin words so the notion of ascender/descender is clearly distinguishable; the same notion has not the same signification in the case of the Bangla script. In order to reach higher results for Bangla script, some other kind of structural features should be extracted as water reservoir features [6] or matra feature which can better describe the Bangla script.

5. Conclusions

In this paper, we described a general technique to implant high-level information is the baseline NSHP-HMM. The described technique improves the discriminating capability of the system by combining low-level features with high-level features extracted from the analyzed shape.

While the encouraging results achieved for the Bangla dataset can not be compared with other methods as no previous work exists in this field, the result for the SRTP dataset outperforms all the results reported in [1, 2, 8].

Generally, to get more important improvements, more adequate structural information should be extracted, like convex and concave sectors, cross points, cutting points, etc which better describes the given script. Extracting a huge variety of features the normalization process can be also refined as different weights can be assigned to the different features in function of their discriminating power.

References

[1] C. Choisy and A. Belaïd. Cross-learning in analytic word recognition without segmentation. *IJDAR*, 4(4):281–289, 2002.

[2] M. Gilloux, B. Lemarié, and M. Leroux. A hybrid radial basis function network/hidden markov model handwritten word recognition system. In *ICDAR*, pages 394–397, 1995.

[3] S. S. Kuo and O. E. Agazzi. Keyword spotting in poorly printed documents using pseudo 2-d hidden markov models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):842–848, 1994.

[4] J. Li, A. Najmi, and R. M. Gray. Image classification based on a multiresolution two dimensional hidden markov model. *IEEE Transactions on Signal Processing*, 48(2):517–533, 2000.

[5] H. Miled and N. E. B. Amara. Planar markov modeling for arabic writing recognition: Advancement state. In *ICDAR*, pages 69–73, 2001.

[6] U. Pal, A. Belaïd, and C. Choisy. Touching numeral segmentation using water reservoir concept. *Pattern Recognition Letters*, 24(1-3):261–272, 2003.

[7] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[8] G. Saon and A. Belaïd. High performance unconstrained word recognition system combining hmms and markov random fields. *IJPRAI*, 11(5):771–788, 1997.