

Handling spatial information in on-line handwriting recognition*

Sanparith Marukatat and Thierry Artières

3rd May 2004

LIP6, Université Paris 6
8, rue du Capitaine Scott, 75015
Paris, France
{Sanparith.Marukatat,Thierry.Artieres}@lip6.fr

Abstract

This paper focuses on handling the two-dimensionnal feature of on-line handwriting signals in recognition engines. This spatial information is taken into account in various ways depending on the nature of characters to be recognized. We review some technics used in the literature and investigate new ones to represent and model the spatial information in handwriting recognition engines. We compare formally and experimentally a number of solutions on various character recognition tasks.

Keywords: Online Handwriting Recognition, Spatial Relation, Complex Characters

1 Introduction

On-line handwriting recognition shares many features with other temporal sequences classification such as speech recognition. Hence many ideas and works in on-line handwriting has been inspired from previous works in speech recognition. As an illustration, today many handwriting recognition engines are based on Hidden Markov Models (HMM) a technic popularized in the speech recognition field. However, the two-dimensionnal feature of handwriting is a specificity of on-line handwriting that makes it different from other temporal sequences. An on-line handwriting signal is a series of a few strokes (each one is written without pen-up movement) arranged spatially in a particular way on a 2-dimensionnal paper. For example, the Korean character of figure 1 (a) consists in a circle above an horizontal line which itself is above an almost vertical line. There are many ways to represent, encode, and model such a spatial information, this paper aims at investigating pros and cons of technics used in the literature and some new ideas.

*Part of this work has been done in collaboration with France Télécom (grant number 021BA40).

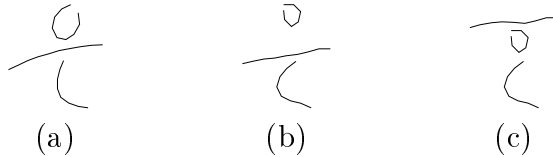


Figure 1: Two samples of the same Korean character (a) and (b). The figure (c) shows an hypothetical signal that could be confused with the character in (a) or (b) using an absolute description of the spatial information.

The 2-dimensionnal information is a fundamental feature for representing and recognizing on-line handwriting characters and a number of solutions have already been proposed which are very different according to the nature of the characters being modeled. At one extreme, considering latin handwritten characters, the spatial information is rather simple: most characters or digits are written in a single stroke, without a pen-up move so that spatial information mainly resumes to diacritical marks. Ad-hoc procedures have been used here that makes use of prior knowledge on characters shape. For example in [10], very simple heuristics are used to detect diacritics marks (e.g. short strokes written afterwards -after a pen-up- in the upper part of lowercase characters) and spatial information is represented through bitmaps that are integrated into the feature vectors [9, 10] that are fed into the classifier (e.g. HMM). This strategy gives interesting results for simple characters but is not accurate enough for more complex characters, written with many pen-up moves. Richer representation and modeling of spatial information have been proposed for Asian character recognition (Chinese, Japanese or Korean) [3, 5, 6, 7] since spatial information is a main topic for such tasks. Note that interesting work has been done to deal with “characters” or “drawings” such as mathematical formulas recognition [2].

To characterize spatial information of a handwriting signal, i.e. the positions of the *components* of a character (e.g. circle and horizontal line for the character in figure 1), one can distinguish two modeling schemes which we call in the following *absolute* and *relative* description. An absolute spatial description consists in describing the position of each component with respect to a fixed reference (generally the character’s bounding box), independently of the other components positions [6, 13]. This approach is well adapted to characters whose components positions are stable but may be inaccurate for other characters. As an illustration, figures 1 (a) and (b) show two samples of the same Korean character, one can see that absolute positions of the three main parts of the character (circle on the top, horizontal line in the middle and almost vertical line at the bottom) exhibit some variability so that an absolute modeling of these positions (e.g. with gaussian laws) could lead to an unexpected behaviour, e.g. giving a high likelihood to the positions of the three parts of the character in figure 1 (c). However, even without any knowledge of Korean characters, it should be clear that the signal in figure 1 (c) cannot be the same character as the one drawn in figure 1 (a) and (b).

Hence, absolute positions of components is not always an accurate representation and relative positions (the circle is above the horizontal line etc) seems to be much more related to our perception of spatial arrangement similarity. In the relative description approach, one is interested in the spatial relation between components rather than their absolute position; The fact that the circle is above the horizontal stroke is more important than the real absolute position of these components. A number

of ideas of this kind have been studied ranging from characterizing the pen-up movement between two written strokes to a more complex description of spatial relationships.

This paper aims at comparing the effectiveness and the accuracy of some modeling approaches for the spatial information in handwriting. However, such an evaluation is not easy. To do this, we integrated these spatial information modeling schemes in a Markovian-based character recognition system and compare recognition results on various isolated character recognition tasks. In the following, we first present how spatial information is integrated in a Markovian recognition engine (§2). Then, we describe various modeling of the spatial information using an absolute approach (§3) and a relative approach (4) and explain how these modeling may be embedded into the recognition engine. Finally, we present experimental results (§5) gained with various handwriting signals in order to put in evidence the strength and weakness of the various solutions we propose.

2 Integrating spatial information in an on-line handwriting recognition engine

For clarity of presentation, we describe here how spatial information modeling is integrated into our HMM based recognition engines but the principle holds for many other HMM based recognition systems. In our systems, a character model is a mixture of left-right HMMs. The likelihood of a T -lengthed handwriting signal X (e.g. a sequence of T points $X = (p_1, p_2, \dots, p_T)$), computed by one of these left right HMM λ is given by :

$$P(X/\lambda) = \sum_{q_1^T} P(X/q_1^T, \lambda).P(q_1^T/\lambda) \quad (1)$$

where $q_1^T = (q_1, q_2, \dots, q_T)$ is a segmentation of X in the HMM λ (i.e. a sequence of states of λ). Let K be the number of states of λ and, using the left-right topology of λ , let b_i and l_i be the beginning and leaving times in state s_i along the segmentation q_1^T , i.e. $q_t = s_i, \forall t \in [b_i, l_i]$ and $\forall i = 1..K-1, b_{i+1} = l_i + 1$. In a classical HMM system, the handwriting signal X is preprocessed and transformed in a sequence of feature vectors (f_1, f_2, \dots, f_T) and, using an independence assumption of these observations (feature vectors) conditioned on the segmentation q_1^T , $P(X/q_1^T, \lambda)$ is computed through :

$$P(X/q_1^T, \lambda) = \prod_{i=1}^{i=K} p(f_{b_i}^{l_i}/s_i) = \prod_{i=1}^{i=K} p(f_{b_i}, f_{b_i+1}, \dots, f_{l_i}/s_i) = \prod_{i=1}^{i=K} \prod_{t=b_i}^{l_i} p(f_t/s_i) \quad (2)$$

To integrate an explicit spatial information modeling in this formalism, we chose to separate the shape information modeling and the spatial information modeling using the following formalism:

$$P(X/q_1^T, \lambda) = P_{shape}(X/q_1^T, \lambda).P_{spatial}(X/q_1^T, \lambda)$$

where $P_{shape}(X/q_1^T, \lambda)$ is the probability of the shape of this handwriting signal, without considering the spatial information (pen-up moves etc), and $P_{spatial}(X/q_1^T, \lambda)$ is related to the spatial information only. The first part the likelihood is computed as usual through equation 2, where feature vectors f_t may include direction features but not spatial information features (e.g. coordinates).

$$P_{shape}(X/q_1^T, \lambda) = \prod_{i=1}^{i=K} p_{shape}(f_{b_i}, f_{b_i+1}, \dots, f_{l_i}/s_i)$$

The question is how to compute the spatial part of the likelihood. Before going further, recall that in left-right HMM models, each state models a particular part (e.g. beginning, middle, end...) of the drawing of a character. It is then natural to define a model of spatial information as a function of the parts of an on-line signal that are assigned, through a segmentation, to the states of the HMM. Let seg_1^Q, \dots, seg_K^Q be the K segments of points of X associated to states s_1, s_2, \dots, s_K according to the segmentation $Q = q_1^T$: the first segment seg_1^Q is (x_1, \dots, x_{l_1}) , the second segment seg_2^Q is $(x_{b_2}, \dots, x_{l_2})$... and :

$$P_{spatial}(X/q_1^T, \lambda) = P_{spatial}(seg_1^Q, \dots, seg_K^Q/\lambda) \quad (3)$$

In the next sections, we will define different ways to model such a spatial information and according methods to score an observed spatial information, given the model.

3 Absolute spatial information

Using an absolute spatial description relies on an absolute reference [6, 13], we chose to use a normalized bounding box : we first determine the bounding box of the handwriting signal, then we rescale the signal by setting the bounding box to an height and a width equal to 1. Hence, there is no explicit dependency among parts of signals assigned to different states of an HMM so that :

$$P_{spatial}(X/Q = q_1^T, \lambda) = \prod_{i=1}^{i=K} P_{Absolute}(seg_i^Q/s_i)$$

A few choices are possible to define the probability distribution $P_{Absolute}(seg_i^Q/s_i)$, we used to define it as a function of the position of the center of the segment which we model with a gaussian law, leading to:

$$p_{Absolute}(seg_i^Q/s_i) = \mathcal{N}(\mu(s_i), \sigma(s_i); center(seg_i^Q))$$

where $\mu(s_i)$ represents an “ideal” position for the center of the i th segment of the signal and $\sigma(s_i)$ is the standard deviation around this ideal position, $center(seg_i^Q)$ is the center of mass of the sequence of points $(x_{b_i}, \dots, x_{l_i})$. All HMM parameters, gaussian laws parameters as well as topology are learned from the data [11].

4 Relative spatial information

The aim of the relative approach is to describe the spatial information as relative positions of strokes with respect to each others. Before describing various ways to characterize such a relative position -which we call in the following an *elementary spatial relation (ESR)*- we first discuss how to characterize the whole spatial information in an handwriting signal as a set of ESR.

4.1 Global Description

An ESR allows characterizing the position of a stroke (or segment) with respect to another stroke (or another segment): is it above, on the right ...? Considering the discussion at the end of §2, assume ESR between all segments of the handwriting signal (i.e. sequence of points associated, along a particular segmentation Q , to the K states of the HMM) are available. Using such ESR, the question is how to describe the whole spatial information in an on-line handwriting signal.

It is an intuitive idea that the whole spatial information may be represented as the set of all ESR between all segments of the handwriting signal [8], this would lead to $\frac{K \times (K-1)}{2}$ ESR to describe the spatial information of a handwriting signal with K strokes. This means that (3) may be rewritten as

$$P_{spatial}(seg_1^Q, \dots, seg_K^Q/\lambda) = P_{spatial}(esr(seg_1^Q, seg_2^Q), \dots, esr(seg_1^Q, seg_K^Q), esr(seg_2^Q, seg_3^Q), \dots, esr(seg_2^Q, seg_K^Q), \dots, esr(seg_{K-1}^Q, seg_K^Q)/\lambda)$$

Using an independence assumption between all ESR, this may be rewritten as:

$$P_{spatial}(X/q_1^T, \lambda) = \prod_{i=1}^{i=K} \prod_{j=1, j \neq i}^{j=K} p_{spatial}(esr(seg_i, seg_j)/\lambda, s_i)$$

Furthermore, one can argue that ESR are more or less symmetrical such that only half of these ESR are necessary. Then a good choice for the global description is the set of all ESR between a segment associated to a state and all segments associated to previous states:

$$P_{spatial}(X/q_1^T, \lambda) = \prod_{i=2}^{i=K} \prod_{j=1}^{i-1} p_{spatial}(esr(seg_i, seg_j)/\lambda, s_i) \quad (4)$$

This allows using extension of classical dynamic programming routines such as Viterbi algorithms, but with a increased cost however. A simplified choice consists in considering only *ESR* between any segment and the preceding segment.

$$P_{spatial}(X/q_1^T, \lambda) = \prod_{i=2}^{i=K} p_{spatial}(esr(seg_i, seg_{i-1})/\lambda, s_i)$$

This choice leads to an easier integration in dynamic programming procedures but gives poorer results form some complex characters, so that in the remaining of this work, we provide experimental results for the more complete modeling of (4) only. Note that the above modeling is close to the one proposed in [3] where the position of a stroke with respect to the previous and the next stroke is modeled through Bayesian networks.

The above description schemes are what we call *flat* descriptions to emphasize the difference with *hierarchical* description schemes we discuss now. For example, a flat description of each signal in figure 3 (b) consists in ESR between all 5 strokes whereas a hierarchical description would include two levels, the first one related to the spatial relation between the two main components of the drawing (the box and the horizontal line), the second being related to spatial relation between strokes within

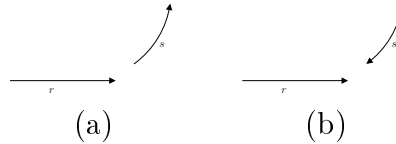


Figure 2: Example of two stroke sequences.

a same component (the four strokes of the boxes). The idea behind hierarchical description is the definition of a more generic representation able to cope with more complex characters or drawings. Indeed, one might expect that, for complex drawings involving many strokes and pen-up moves, all spatial relation between strokes may not be relevant and could be “noisy”. The main problem with such a hierarchical modeling lies in that one must be able to identify components in an handwriting signal. We used in this preliminary study a simple heuristic that consists in considering as components the parts of the handwriting signal that are written within a pen-down and a pen-up moves.

4.2 Elementary Spatial Relation (ESR)

The simplest way to characterize the relative position of strokes is to compute the translation vector corresponding to pen-up moves. The translation may be computed between center of strokes or between the last point of a stroke and the first one of the next stroke and such translation may be modeled with for example a gaussian law [1]. Another possibility is proposed in [7, 12], the handwriting signal is represented as a sequence of direction codes and particular direction codes are used for pen-up moves. However, such representations are adapted to pen-up moves only while we are interested here in more general representation and modeling of spatial relation between strokes.

One popular and simple way to describe the position of a stroke with respect to another stroke (or between two groups of strokes) is to use a combination of a few discrete attributes such as vertical position VP (with values above/aligned/below), horizontal position HP (left/aligned/right) and connexity $Connex$ (touching/not touching). A relative spatial position is then a triplet $(HP, VP, Connex)$ that is easily computed from bounding boxes. For example, in figure 2 (a) and (b) stroke r is below, on the left and not touching stroke s . Such a simple spatial relation does not allow to distinguish between the two configurations in figure 2 (a) and (b). A more accurate spatial relation consists in taking into account the direction of the strokes [8]. Let consider now three attributes which are the longitudinal position (in front/aligned/behind), lateral position (left/aligned/right) and connexity (touching/not touching). These attributes are computed to characterize the position of a stroke (e.g. r) with respect to another stroke (e.g. s) and its orientation. For example, in the figure 2 (a) the stroke r is behind, on the left and not touching stroke s while in the figure 2 (b) it’s in front, on the right and not touching stroke s . To distinguish between these two ESR described above, we will call direction independent elementary spatial relation (DIESR) for the first one and direction-dependent elementary spatial relation (DDESR) for the second one. Such ESR may be used in global spatial information modeling schemes (4), where $p_{spatial}(esr(seg_i, seg_j)/s_i)$ are discrete probabilities learned from data as in [11].

These two ESR are based on discrete attributes that are computed from the handwriting signal.



Figure 3: Samples of three similar but different Korean characters (a) and of six symbols (b).

While these ESR give an understandable description of the spatial information these are not robust enough to noise and variability in handwriting signals. We investigated the use of continuous attributes instead. For a DIESR, we use three numerical values to encode the vertical position instead of one discrete attribute. These 3 values are the ratios of the bounding box of r (noted b_r) that is above the bounding box of s (noted b_s), the ratio of b_r that is aligned vertically with b_s and the ratio of b_r that is below b_s . We use three numerical values, computed similarly, to represent the horizontal position of r relative to s . These six values together constitute an ESR between two strokes or two groups of strokes, this computation can also be applied to the direction dependent ESR.

At the end, this lead to four ESR that we will call discrete DIESR (DDIESR), continuous DIESR (CDIESR), discrete DDESER (DDDESER) and continuous DDESER (CDDESER). Continuous ESR may be used in global spatial information modeling schemes (4), where $p_{spatial}(esr(seg_i, seg_j)/s_i)$ are modeled with gaussian laws whose parameters are learned from data.

5 Experiments

In our experiments we systematically report results gained with two recognition systems, one “small” system A requiring few memory and processor, and a more expensive system B . These two systems are built using the approach described in ([11]) and spatial information is integrated in such systems as explained in previous sections. Note that a two-fold cross validation is used in all our experiments and averaged accuracies are reported.

We considered three databases of characters involving spatial information of various complexity : standard latin characters from the UNIPEN database [4], Korean character from the KAIST database¹, and a small home made database of miscellaneous symbols.

Standard latin characters (lowercase, uppercase and digits) involve simple spatial information. In our experiments, we used 16000 digits samples, 27000 uppercase samples and 60000 lowercase samples. Korean characters are written in several pen lifts, with or without ligature, and spatial information is an important feature of the writing of a character (see figure 3 (a)). We used signals corresponding to 83 Korean characters (those with at least 50 samples), with about 13000 samples. Finally, the symbols database consists in writings of six very similar symbols, except from the spatial information point of view (see figure 3 (b)). We designed these symbols to put in evidence fundamental differences between the various modeling schemes that we investigated, especially for the global description. There are about 30 samples per symbol, written by 3 writers.

In a first experiment, we investigated the effectiveness of various ESR in character recognition tasks (lowercase and Korean characters) and for the two systems A and B (Table 1). First, one may

¹<http://ai.kaist.ac.kr/>

elementary spatial relation		Database and Recognition engine			
		lowercase		Korean characters	
		A	B	A	B
DIESR	discrete	76.4	82.1	83.7	89.3
	continuous	77.2	82.7	84.9	89.6
DDESR	discrete	73.4	80.9	84.3	88.9
	continuous	70.7	79.4	84.7	89.3

Table 1: Recognition rates for lowercase and Korean characters for various ESR and with two recognition engines (A and B).

see that Direction Independent ESR almost always overcome Direction Dependent ones, whatever the characters, the recognition engine, the variant used (discrete vs. continuous). This is an interesting fact since one could imagine that the Direction Dependent ESR allow richer description. In our experiments it seems that DDESR may be more accurate but are also less robust. A second interesting point is that continuous variants most often overcome discrete ones, it is expected that continuous modeling is more robust against variability. Finally, it is worth noting that the above comments hold whatever the system A and B , meaning that these trends are independent of the recognition engine but are rather intrinsic properties of these elementary spatial relations.

In a second series of experiments we compared absolute and relative spatial description with a flat global scheme (Table 2). According to previous results, we use here the continuous variant of DIESR for the relative spatial information. There are a few points that we can notice. First, recognition rates are significantly higher when using spatial information (from 5% to 9% accuracy). Second, at first glance systems using absolute and relative modeling achieve rather similar results but looking deeper, the absolute approach is more efficient for simpler characters such as digits and lowercase letters while the relative approach leads to better results for more complex characters (with more pen-up moves) such as uppercase and Korean characters. This suggests that relative approach allows a better description for complex spatial arrangement. Furthermore, the two approaches are complementary. To make use of this complementarity, we considered a combination of the two modeling where the likelihood of the spatial information is computed as the product of absolute and relative modeling likelihoods. The results for this combination scheme are given in column “mixed” in table 2. As may be seen, this strategy is the best one in any case. Finally, it is interesting to notice, here again, that all the above comments hold whatever the recognition engine, A or B.

Finally, we conducted a third series of experiments in order to investigate the more efficient way to model the global spatial information in the writing of a character as a set of ESR. To put in evidence main trends, we used here characters for which spatial information is a main feature, Korean characters and our home-made symbols database. We provide comparative results of hierarchical spatial description and flat spatial description, with CDIESR but similar trends have been observed with other kind of ESR. Table 3 sums up results for Korean characters and symbols. One sees that hierarchical description is less accurate than a flat global spatial description for Korean characters.

characters	system	spatial information			without spatial information
		absolute	relative	mixed	
digits	A	91.4	90.0	93.5	87.2
	B	94.9	94.0	96.6	91.0
uppercase letters	A	85.1	86.0	88.3	77.4
	B	89.1	89.8	91.9	81.4
lowercase letters	A	79.8	77.2	80.9	72.9
	B	84.4	82.7	86.7	77.3
Korean characters	A	85.1	84.9	86.5	80.5
	B	89.6	89.6	91.0	84.6

Table 2: Comparison of absolute and relative spatial description. The column entitled “mixed” designs the system combining both absolute and relative spatial information.

Character type	System	Hierarchical		Flat	
		discrete	continuous	discrete	continuous
Korean	A	81.4	83.7	83.7	84.9
	B	87.2	88.4	89.3	89.6
Symbols	A	82.2	92.5	69.3	84.7
	B	85.0	95.7	76.8	91.3

Table 3: Comparison of hierarchical and flat global spatial description for Korean character and symbols recognition.

But, in the case of the six symbols of figure 3 we found that the hierarchical description improves significantly accuracy especially with discrete ESR. We believe these contradictory results may come from the heuristic we used to identify components in character drawings and we are working at the improvement of this decoding step.

6 Conclusion

This paper discussed the representation and modeling of spatial information in on-line handwriting modeling and recognition. We reviewed a number of propositions that have been used in the handwriting literature and organized this review around the concepts of elementary spatial relation and of global description scheme. We investigated the definition and comparison of various elementary spatial relations. We also proposed a few global description schemes, i.e. minimal and efficient sets of elementary spatial relations allowing defining the whole spatial information in an handwriting signal. We performed experimental comparisons on various character recognition tasks. Handling the spatial information in a recognition engine allows systematic improvements whose significance depends on the nature of the characters being recognized. Going further absolute improvements, we discussed and put in evidence experimentally some general trends concerning the modeling accuracy of various spatial information representation and investigated how to deal with more complex

characters using a hierarchical modeling of the spatial information. Although this latter idea did not bring systematic improvements it gave promising results for symbols recognition for which accurate spatial information modeling is a cue feature.

References

- [1] T. Artières and P. Gallinari. Stroke level hmms for on-line handwriting recognition. In *International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, 2002.
- [2] K.-F. Chan and D.-Y. Yeung. Mathematical expression recognition: A survey. *International Journal on Document Analysis and Recognition (IJDAR)*, 3(1):3–15, 2000.
- [3] S. J. Cho and J. H. Kim. Bayesian network modeling of strokes and their relationships for on-line handwriting recognition. In *International Conference on Document Analysis and Recognition (ICDAR)*, 1999.
- [4] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet. Unipen project of on-line data exchange and recognizer benchmark. In *International Conference on Pattern Recognition (ICPR)*, 1994.
- [5] H.-Y. Kim and J. H. Kim. Hierarchical random graph representation of handwritten characters and its application to hangul recognition. *Pattern Recognition*, 34:187–201, 2001.
- [6] K. Kuroda, K. Harada, and M. Hagiwara. Large scale on-line handwritten chinese character recognition using successor method based on stochastic regular grammar. *Pattern Recognition*, 32:1307–1315, 1999.
- [7] J. J. Lee, J. Kim, and J. Kim. Data-driven design of hmm topology for on-line handwriting recognition. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 15(1):107–121, 2001.
- [8] C.-L. Liu, I.-J. Kim, and J. H. Kim. Model-based stroke extraction and matching for handwritten chinese character recognition. *Pattern Recognition*, 34(12):2339–2352, 2001.
- [9] S. Manke, M. Finke, and A. Waibel. Combining bitmaps with dynamic writing information for on-line handwriting recognition. In *ICPR*, 1994.
- [10] S. Marukatat, T. Artières, B. Dorizzi, and P. Gallinari. Sentence recognition through hybrid neuro-markovian modelling. In *ICDAR*, 2001.
- [11] S. Marukatat, R. Sicard, T. Artières, and P. Gallinari. A flexible recognition engine for complex on-line handwriting character recognition. In *ICDAR*, 2003.
- [12] M. Nakai, N. Akira, H. Shimodaira, and S. Sagayama. Substroke approach to hmm-based on-line kanji handwriting recognition. In *ICDAR*, 2001.

- [13] J. Zheng, X. Ding, Y. Wu, and Z. Lu. Spatio-temporel unified model for on-line handwritten chinese character recognition. In *ICDAR*.