Reprint of: Schomaker, L.R.B., & Teulings, H.-L. (1990). A Handwriting Recognition System based on the Properties and Architectures of the Human Motor System. *Proceedings of the International Workshop on Frontiers in Handwriting Recognition (IWFHR).* (pp. 195-211). Montreal: CENPARMI Concordia.

A Handwriting Recognition System Based on Properties of the Human Motor System

Lambert R.B. Schomaker & Hans-Leo Teulings
Nijmegen Institute for Cognition Research
and Information Technology (NICI)
P.O. Box 9104
NL 6500 HE, Nijmegen
The Netherlands
Tel: +31-80-516029 & 515476; FAX: +31-80-515938
E-mail: Schomaker@Hnykun53.Bitnet & HLT@Hnykun53.Bitnet

Abstract

The human reader of handwriting is unaware of the amount of back-ground knowledge that is constantly being used by a massive parallel computer, his brain, to decipher cursive script. Artificial cursive script recognizers do not have access to a comparable source of knowledge or of comparable computational power to perform top-down processing. Therefore, in an artificial script recognizer, there is a strong demand for reliable bottom-up processing. For the recognition of unrestricted script consisting of arbitrary character sequences, on-line recorded handwriting signals offer a more solid basis than the optically obtained grey-scale image of a written pen trace, because of the temporal information and the inherent vectorial description of shape. The enhanced bottom-up processing is based on implementing knowledge of the motor system in the handwriting recognition system. The bottom-up information will already be sufficient to recognize clearly written and unambiguous input. However, ambiguous shape sequences, such as *m* vs *n*.. or *d* vs *cl*, and sloppy stroke patterns still require top-down processing. The present paper discusses the handwriting recognition system as being developed at the NICI. The system contains six major modules: (1) On-line digitizing, pre-processing of the movements and segmentation into strokes. (2) Normalization of global handwriting parameters. (3) Extraction of motorically invariant, real-valued, feature values per stroke to form a multi-dimensional feature vector and subsequent feature vector quantization by a self-organizing two-dimensional Kohonen network. (4) Allograph construction, using a second network of transition probabilities between cell activation patterns of the Kohonen network. (5) Optional word hypothesization. (6) The system has to be trained by supervised learning, the user indicating prototypical stroke sequences and their symbolic interpretation (letter or N-gram naming).

Introduction

There are many advantages if data can be entered into a computer via handwriting rather than via typing (Teulings, Schomaker & Maarse, 1988). These advantages are acknowledged by hardware manufacturers who are testing the market with 'electronic paper' with built-in computer systems for recognizing elementary pen movements (e.g., Hayes, 1989). Electronic paper consists of an integrated LC display plus digitizer. Although the user acceptance of this kind of hardware will depend on the solution of some technical and ergonomical problems that are currently present (visual parallax, surface texture, stylus wire), it seems relevant to develop on-line handwriting recognition systems for unconstrained handwriting. Several commercial systems exist that recognize on-line handprint, but cursive script recognition has still not been solved satisfactorily (Tappert et al., 1988). Ideally, a recognition system should be able to recognize both handprint, for accuracy, and cursive script, for optimal writing speed. However, the major problem in cursive-script recognition is the segmentation of a word into its constituting allographs prior to recognizing them, while the allographs have different numbers of strokes (Maier, 1986). Indeed, even for human readers cursive script is sometimes ambiguous. One advantage of on-line recognition is that in case the system is not able to disambiguate, the correct output can be provided by the user interactively. However, the most important advantage of including on-line movement information is, that it contains more information than the unthinned, quantized images of the optically digitized pen traces. Consider for instance the final allograph m which may appear in the spatial domain as a single horizontal curl, but in the time domain still displays the three pen-speed minima. This kind of extra information is needed to compensate for the large amount of top-down processing done by the 'understanding' human reader of handwriting. The enhanced bottom-up processing is based on implementing knowledge of the motor system in the handwriting recognition system. Our efforts to introduce handwriting as an acceptable skill in the office environment has resulted in a multinational consortium (PAPYRUS) aimed at building software and hardware for a simple electronic note book, allowing the user to enter data into a computer without using a keyboard.

In Teulings et al. (1987) we introduced a modular architecture for the low-level bottom-up analysis of handwriting in our so-called Virtual Handwriting System (VHS). The present paper discusses the handwriting recognition system as being developed at the NICI. The system contains six major modules which are also found in several other recognition systems (e.g., Srihari and Bozinovic, 1987, for off-line handwriting).

1. On-line recording of handwriting, pre-processing consisting of lowpass filtering and differentiation and finally, segmentation into intended movement units ('strokes').
2. Normalization of various motorical degrees of freedom.
3. Computation of feature values ('feature vector') per stroke which are motor invariants or salient to the human perceptual system, followed by the quantization of stroke shapes using a self-organizing Kohonen network.
4. Construction of letter (allographic) hypotheses from sequences of quantized strokes.
5. Construction of word hypotheses from sequences of allographic hypotheses.
6. Supervised learning of the relation between stroke-vector sequences and allographs.

Below, these modules will be discussed in terms of their purpose, the knowledge of the motor system or the perceptual system used, its realization and its performance.

1. Recording, Pre-processing, and Segmentation

Purpose

The pre-processing stage consists of all operations needed to provide a solid base for further processing. At this stage the data consist of a continuous signal without any structure. The first operation is to split the continuous signal into batches that can be processed separately. We suggest that a word is the easiest batch to be processed. Then for each word, the signal, containing noise from different sources (e.g., the digitizing device), is low-pass filtered. Finally the continuous movement is segmented into basic movement units. Knowledge of the human motor system provides an empirically and theoretically basis for the segmentation heuristics.

Motor system

The control of the muscles involved in producing the writing movements is of a ballistic nature: each stroke has only a single velocity maximum (Maarse et al., 1987) and a typical duration between 90 and 150 ms. Shorter-lasting motorical actions are very unlikely to be the result of intentional muscle contractions. For an appropriate pre-processing it is relevant to understand the frequency spectrum of handwriting movements. The displacement spectrum contains a large portion of very low-frequency activity, mainly due to the ramp-like shape of the horizontal displacement. This is not true for pen movement direction and velocity. The latter signal is estimated by calculating the first time derivative. The differentiation suppresses the low-frequency components that are present in the displacement spectrum, and a more informative spectral shape emerges. In Teulings & Maarse (1984) it has been shown that the velocity amplitude spectrum is virtually flat from 1 to 5 Hz where it has a small peak and then declines to approach the noise level at about 10 Hz. Therefore, a low-pass filter with a flat pass band from zero to 10 Hz will remove the high-frequency noise portion of the signal while leaving the relevant spectral components of the handwriting movement unaltered. In order to prevent oscillations (Gibbs phenomenon) it has been shown that the transition band should not be too narrow (e.g., at least 8/3 of the width of the passband). From the bandwidth of at least 5 Hz follows that the movement can be most parsimoniously represented by about 10 samples per second. Since the endpoints of the strokes appear to be about 100 ms apart, the time and position of the stroke endpoints as determined by two consecutive minima in the absolute velocity are a good basis for reconstruction (Plamondon & Maarse, 1989). Points of minimum velocity correspond with peaks in the curvature (Thomassen and Teulings, 1985).

Realizing that the vertical movements appear to be less irregular than the horizontal progression, Teulings et al., (1987) suggested to weigh the vertical component higher than the horizontal component in the calculation of a biased absolute velocity signal (up to factor of 10).

Realization

Handwriting movements are recorded on a CalComp2500 digitizer with a resolution of about 0.1 mm and a sampling frequency of 125 Hz using a pen which contains a solid state transducer to measure the axial pen pressure synchronously with pen tip position. A pressure threshold serves as a sensitive pen on/off paper detector. The data were not corrected for non-simultaneous sampling of x and y (Teulings and Maarse, 1984) nor for variations of pen tilt (Maarse et al., 1988).

Filtering, and time derivation are done using frequency domain fast Fourier transforms. In stroke segmentation, time points are chosen which are about 100 ms or more apart. This is done by selecting the lowest absolute velocity minimum within a time window of 50 ms around a given minimum (Teulings & Maarse, 1984).

Word segmentation is not based on particular information of the motor system but rather on perceptual cues. It is done by detecting a fixed horizontal displacement while the pen is travelling above the paper beyond the right or the left boundary of the last pen-down trajectory.

Performance

The performance of this straight-forward pre-processing does not appear to be the main source of recognition error in the present system.

2. Normalization

Purpose

A particular problem in handwriting recognition is its extensive variability. A given letter can be produced in several ways, each having its own typical shape, e.g., lower case a vs upper case A or the well-known different variants of the t. The shape variants for a given letter are called allographs. Thus, first there is the between-allograph variability (I): a writer might select different letter shapes in different conditions or at free will. Second, there is the within-allograph shape variation in which the topology of the pattern is not distorted (II), the error source being (psycho)motor variability. Topology can be defined as the number of strokes and their coarsely quantized relative endpoint positions. Third, there is the within-allograph shape variation which actually does distort the topology of the pattern, by the fusion of two consecutive strokes into a single ballistic movement (III) in fast and/or sloppy writing. These three types of variabilities will all be prevalent to some degree under different conditions. Table 1. gives an impression of the estimated order of these variabilities depending on context and writer. The context of a given allograph is defined as the identity of the allographic neighbours and the serial position of the target allograph.

Table 1. The estimated order of the degree of handwriting variability that a script recognition system has to handle, under different conditions, for all three types (I-III) of variability (1= minimum variability, 4=maximum variability).

	Context	
Writer	Identical	Different
Identical	1	2
Different	3	4

The problem of allographic variation (I) can only be solved by presenting to the recognition system at least one prototype for each allograph, as they act as different symbols and we do not suppose that an artificial system will be able to generalize totally different allographs of the same letter. Also, within-allographic variation leading to different topologies (III) is handled by presenting each variant to the system separately in the training stage. Theoretically, however, it should be possible to recover a 'clean' topological representation of a fused allograph by deconvolution, or by auto-regressive techniques (Kondo, 1989). However, a large part of the within-allographic variation (II), can be solved in the bottom-up analysis by normalizing the writing pattern, prior to extracting the features and by choosing relatively invariant features. The importance of normalization will be discussed here and the choice of invariant features in the next section.

Motor system

In order to extract the sequence of feature vectors of the handwriting input, several normalization steps can be performed (See Thomassen et al., 1988, for an overview). The reason is that a sample of a person's handwriting contains various global subject-specific parameters, like slant or width of the allographs (e.g., Maarse et al., 1988). Also, the motor system is able to transform handwriting deliberately, e.g., changing orientation, size or slant (e.g., Pick & Teulings, 1983). However, these global parameters do not contain any information about the identity of the characters. Therefore, the handwriting patterns have to be normalized in terms of orientation, vertical size, and slant (Thomassen et al., 1988).

It may be anticipated that several alternative normalization procedures can be proposed. We require the system to try them all and to learn to use the most appropriate ones. As such it resembles Crossman's (1959) statistical motor-learning model: a person has a repertoire of several methods for every action and learns with time which of those is most appropriate.

Realization

Orientation is defined as the direction of the imaginary base line. Vertical size consists of three components: body height, ascender height and descender height relative to the base line. Slant is defined as the general direction of the vertical down strokes in handwriting (e.g., Maarse and Thomassen, 1983). The normalization consists of estimating these parameters and then performing a normalization by a linear planar transformation towards horizontal orientation and upright.

Various algorithms to estimate the parameters for each normalization step are available and not every algorithm may be appropriate in all conditions. Averaging these estimates is probably not the best choice because one estimator ('demon') may be totally wrong. A sub-optimal choice of the orientation, for instance, has dramatic effects in the subsequent normalization of size or slant. The solution we propose is to have the system select the best available, unused estimator using the estimators' current confidence and the proven correctness in the past using a Bayesian approach (Teulings et al., in prep.). This prevents an exponential increase in computational demands with an increasing number of estimator algorithms (demons).

Performance

The normalization estimators have not yet been evaluated statistically. However, both in artificial data (using bimodally distributed estimates of different variance) and in handwriting data (using a prototype system with parallel processes), the system produces stable and optimized estimates within 30 trials. We observe that the system backtracks immediately to the normalization level where apparently an inappropriate estimator was chosen first, after which the second best alternative is evaluated. Even though calculation is reduced by taking the 'best first' approach, a multiple estimator scheme requires a lot of computation. However, due to the modularity of the approach, a solution by means of a network of transputers is very well possible. As we are still in a stage of testing with only two writers this system was not used currently. Only vertical-size normalization was performed using one estimator. The effects of vertical size normalization are relatively small as it is only one of several features. Orientation was standardized by lined paper on the digitizer and slant can be assumed approximately constant within a writer in a standard condition (Maarse et al., 1988). However, slant does seem to be influenced by the orientation of the digitizer if it is located more distally than normal, e.g., to the right of the keyboard in a typical workstation setting instead of directly in front of the writer. It was observed that the feature quantization network partially counteracted these slant variations as evidenced by reconstruction of the handwriting trace.

3. Feature extraction

Purpose

Each stroke of the normalized handwriting pattern must be quantified in terms of a set of features, a feature vector, that describes the raw coordinates in a more parsimonious way.

It is important to use features that show a relative invariance across replications and across different contexts. As a check for the completeness of the feature set the original pattern must be reconstructable from these features. Finally, in order to facilitate the subsequent classification and recognition stages the feature vector itself should be quantized into a lower-dimensional representation space.

Motor system

We employ a set of features which is related to the underlying hypothetical motor commands and which is complemented by a few visual features. The feature vector comprises 14 features. Only nine of them are related to the stroke itself whereas five refer to the the previous or the following stroke and are included to capture between-stroke context effects. The procedure to select appropriate features is to write a number of identical patterns (e.g., 16) at two speed conditions (normal and at higher speed, respectively). The invariance of a feature of a particular stroke in those patterns can be tested by estimating its Signal-to-Noise Ratio (SNR, See Footnote 1) (Teulings et al., 1986). The advantage is that SNRs of totally different features can be compared and the ones with the highest SNR can be selected. The preliminary data presented here are based on the central 28 strokes of the word 'elementary' produced by one subject. It appears that the SNRs are remarkably constant between the two speed conditions such that only the averages are presented. In order to assess the invariance across conditions, the between-condition correlation of the average stroke patterns of a feature is employed.

The features currently employed are:

(a) The vertical positions of the beginning (Y_b) and end of a stroke (Y_e) relative to the base line and the path length of the stroke (S) all scaled to the average body height, also called x-height, referring to the lower case x. In Teulings et al. (1986) it has been indicated that especially the relative (vertical) stroke sizes are invariant. The SNRs of Y_e or Y_b are 4.9, and the SNR of S is 4.7, which are typical values for spatial characteristics. The between-condition correlations are as high as 0.99.

(b) The directions ϕ_n of the five, straight stroke segments between two subsequent points corresponding with the time moments

$$t = t1 + (n/5) * (t2 - t1),$$

where $t1$ and $t2$ are the time moments of beginning and end of the stroke and $n = [0, 1, ..., 5]$, i.e., $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$. Here we explicitly use dynamic movement information. The rationale is that in equal time intervals the movement direction is changing a relatively constant amount (e.g., Thomassen and Teulings, 1984) such that each new stroke segment adds an equal amount of new information. The two previous and the two following stroke segments (respectively, ϕ_{b4}, ϕ_{b5}, and ϕ_{e1}, ϕ_{e2}) are included as well in order to capture the stroke's context. The SNRs of $\phi_1, ...\phi_5$, are, 7.2, 8.7, 6.3, 2.1, and 1.2, respectively, and the between-condition correlations are higher than 0.92. It can be seen that the directions of the first three stroke segments are highly invariant both within and between conditions. However, the latter two stroke segments show a relatively low SNR but they are kept in the feature vector as they are important to reconstruct the stroke shapes.

(c) The size of the enclosed area between the end of the stroke and the subsequent stroke (λ_e) is rather a visually salient feature. The SNR of λ_e is 5.6 and the between-condition correlation is as high as 0.999.

(d) A pen up indicator (P), which shows whether the pen is predominantly up or down during a stroke. It may be noted that strokes above the paper also count as strokes. As this is a rather coarse binary signal we refrained from presenting any statistics.

In summary, the selected features show absolutely high SNRs and high between-condition correlations which indicates that these features contain the basic information, which constrains the actual movement. As such, these features are attractive to use in a recognition system. Whether this set of features is also a complete one, can only be demonstrated empirically.

Realization

It is trivial to estimate the feature values per stroke. It is, however, less trivial to quantify the distance between feature vectors. An elegant method to solve the problem of irregularly shaped probability distribution functions of the feature vector of classes is vector quantization by an artificial self-organizing neural network (Kohonen, 1984; Morasso, 1989; Morasso et al., 1990). This type of network performs, in a non-supervised way, a tesselation of cell units into regions, each corresponding to a particular prototypical feature vector. The statistical properties of the training set of feature vectors will determine the emergence of the prototypical feature vector set. We have used a 20x20 network. Bubble radius and learning constant α decrease linearly with the number of iterations, from 20 to 1 and from 0.8 to 0.2, respectively. The shape of the connectivity within a bubble was a monopolar and positive rectangular boxcar. The total set of strokes was presented 100 times to the network. Cells representing a quantized vector were arranged in a hexagonal grid.

The completeness of the reduced data is tested by two reconstruction methods. In the first method, the writing trace is reconstructed from the sequence of feature vectors. An average Euclidean distance measure between reconstructed and original pattern is used to express the accuracy of reconstruction, and thus, the quality of the segmentation procedure as well as the information value of the selected features. In the second method, each feature vector is presented to the Kohonen network, and will be substituted by the nearest prototypical feature vector. The sequence of strokes thus yields a sequence of prototypical feature vectors that can be used to reconstruct the original trace in a similar way as described above. The accuracy of this reconstruction yields a second distance measure. It indicates the quality of the feature vector quantization imposed by the Kohonen network.

Performance

The patterns produced by both reconstruction methods are legible, which is in fact the crucial criterion rather than a spatial goodness of fit. Furthermore, the reconstructed patterns lack individual and context-dependent characteristics which stresses that the selected features reduce the writer dependence as well. For example, slant variations due to imperfect normalization will be counteracted by the Kohonen network as single strokes are attracted to their closest, general prototypes.

4. Allograph hypothesization

Purpose

At this stage the writing pattern is represented as a sequence of prototypical strokes. In earlier experiments, we have used a Viterbi algorithm using a lexicon of allographs. Each prototypical allograph was represented by its average feature vector (no feature vector quantization was performed). A Euclidian distance measure was used that was adapted to angular measures (Teulings et al,, in prep.). The problem with this approach was, that for a given stroke position, there is a distance measure with each of the M=26 prototypes. Solution space is a matrix of MxN, where N is the number of stroke positions. Since the allographs mostly have an unequal number of strokes, the plain Viterbi algorithm could not be used. Instead an iterative version was developed, trying to recognize 1-stroke solutions, 2-stroke-solutions, and so on, until the N-stroke solution. The path cost factor was the modified Euclidean distance, optionally combined with a digram transition probability, each term having its own weight. The results of this technique were rather poor so we decided to find a method that yields a smaller solution space, on the basis of quantized feature vectors. Another approach used was to use 6 feedforward perceptrons, (Nx400)x160x26, trained by back propagation, one perceptron for each class of N-stroked allographs, N=1,...,6. This approach, too, yielded too many hypotheses in the MxN matrix. This problem can possibly be alleviated to some extent by introducing competition among the output layers of different perceptrons. Another solution is proposed by Skrzypek & Hoffman (1989), who introduce a final judgment perceptron to combine the output of the N lower layers. The problem is, however, that for the recognition of varying-length temporal patterns, an optimal neural architecture does not exist, yet. Of the known architectures, recurrent nets (Jordan, 1985) are hampered by their limited ability to handle long sequences. Temporal flow nets (Stornetta et al., 1987; Watrous & Shastri, 1987) are currently being tested in speech recognition.

Motor system

In Teulings et al. (1983) we indicated that complete allographs are probably stored at the level of long-term motor memory. An interesting question is to what extent the strokes belonging to one allograph have to be kept together and whether the strokes of different realizations of the same allograph may be assembled to yield a new allograph. The directions of the stroke segments introduced before (i.e., $\phi_{b4}, \phi_{b5}, \phi_1, ...$) show that the correlations between subsequent stroke segments within one stroke range between 0.69 and 0.90 (mean 0.80) whereas the correlations between subsequent stroke segments across the separation of two strokes range between 0.47 and 0.53 (mean 0.50). This implies that even in identical contexts, subsequent strokes are relatively independent. This suggests indeed that allographs are probably built up of different strokes that may be assembled from other similar allographs.

Realization

Rather than performing a template matching between prototypical allographs and an input sequence of strokes, the method we developed at this stage is based on the idea of an active construction of allograph hypotheses. This is done by a neurally inspired algorithm. Once the writer has labeled allographs interactively, and thus created a data base covering a wide range of allographs in different contexts, the system collects, for each prototypical stroke, its possible interpretations. The representation is based on the reasonable assumption that the fundamental (root) feature of an allograph is its number of strokes. Thus, two allographs are definitely different if their number of strokes differs. Each stroke interpretation has the general form $Name(I_{stroke}/N_{stroke})$. Thus, a given stroke may be interpreted as representing one element of the set $\{a(1/3), d(1/3), o(1/2), c(1/1)\}$. The construction of an allograph is a left-to-right process, where the activation level of an allograph hypothesis increases stepwise with each interpretation that is a continuation of a previously started trace. The advantage over storing prototypical allographs is evident: after labeling three, 3-stroked sequences, each representing the allograph a, the network will recognize an a that corresponds to any one of the 27 combinations. The method does not exclude the use of digrams or trigrams as graphical entities. However, the computational load on a sequential computer will increase quadratically with an increasing number of interpretations per prototypical stroke, so the use of trigrams is impractical.

Performance

Table 2 presents the recognition results of two types of handwriting. In Section 6. the training procedures have been reported for each of the two writers. It may be stated that these results have been achieved on unrestricted cursive script of lower case letters without the use of linguistic post processing by means of a lexicon. On the other hand, the data are optimistic as in case of alternative allograph hypotheses (on average about 2 alternatives) the appropriate one was accepted. This was done under the assumption that only linguistic post-processing will be able to solve these true ambiguities. For instance, *u* abd *n* are sometimes written identically.

Table 2. The recognition rates of allographs and of allograph strokes of five different text samples from two writers.

Writer	Text	Words	Time	Allo-graphs	Recog-nized		Strokes	Strokes in allographs	Recog-nized	
		#	(s)	#	#	%	#	#	#	%
A	AP87	40	281	236	153	65	668	525	368	70
A	FE90	51	224	275	243	88	862	702	657	94
A	TEKST	51	89	262	198	76	755	633	542	86
B	TEST1	54	221	299	209	70	987	786	544	69
B	TEST2	60	234	300	229	76	965	813	608	75

Note the difference between the number of strokes that is actually part of an allograph and the total number of strokes. Apparently 18.4% of all strokes cannot be attributed to letters because they are connecting strokes, hesitation fragments, or editing movements. Note that there has been no post processing in any sense. Figure 1 gives an impression of the processing stages and the solution space for the word *aquarel*. Going from bottom to top the solution space (d) is liberally filled with hypotheses of decreasing length as expressed in number of strokes. Shorter hypotheses may 'fall down' in holes that are not filled by hypotheses of greater length.

5. Optional word hypothesization

Purpose

Apart from yielding a list of hypothesized allographs the bottom-up information contains also information to narrow down the number of possibly written words in a word lexicon. The word in the list with the minimum distance from a word in the lexicon can be selected. However, if the bottom-up process is rather certain of a given hypothesized word, then it seems superfluous to use additional lexical top-down processing.

Motor system and Perceptual system

It is known that when writing redundant character sequences (i.e., words or parts of words that could be recovered with a lexicon of words) the writer uses less efforts to produce the allographs neatly.

From human reading research we know that ascenders and descenders (i.e., the contour) are strong cues to recognize the presented word, similar to the function of consonants in speech recognition.

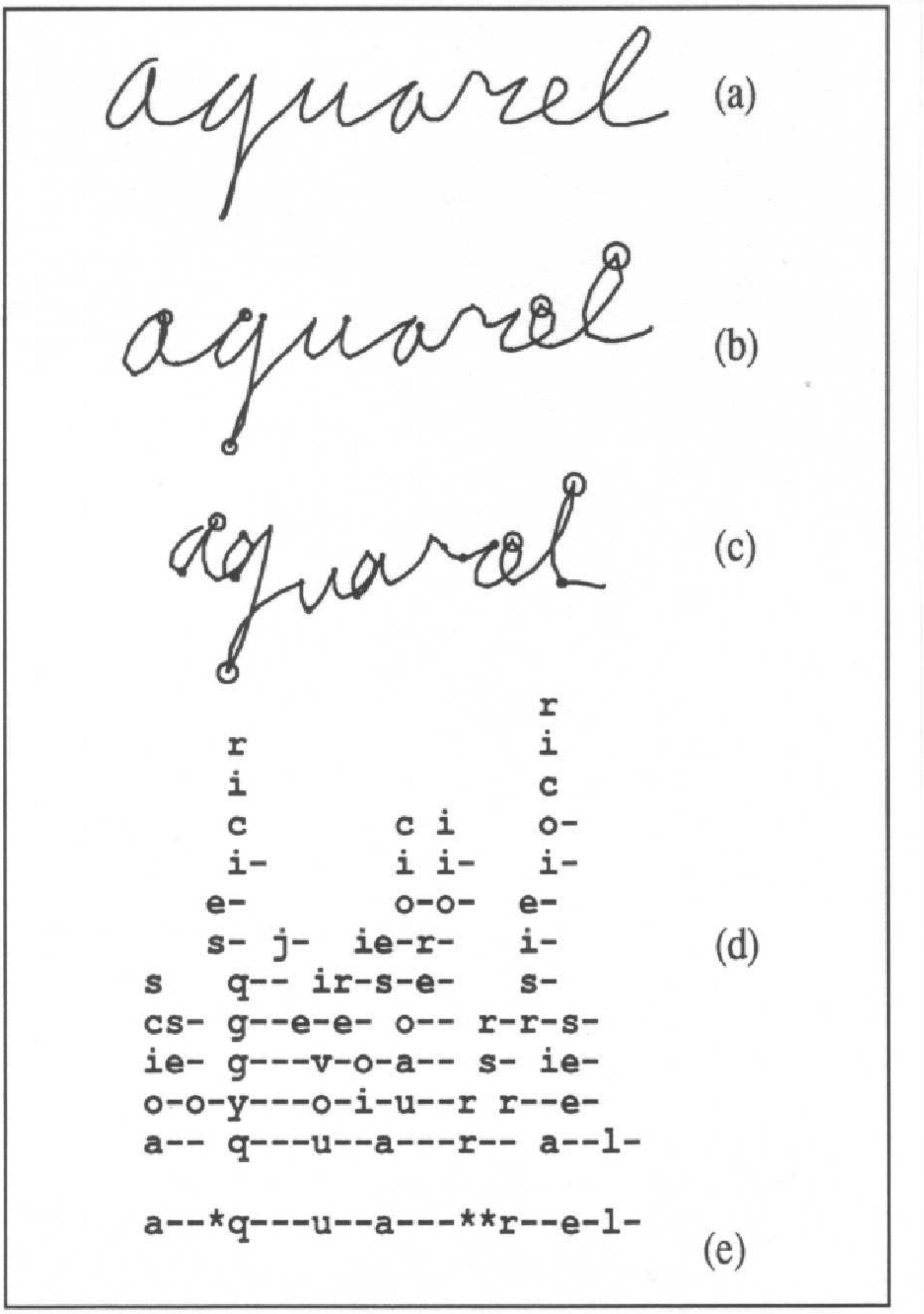

Figure 1: A sample of an on-line recorded word (*aquarel*) (a), its reconstruction from the sequence of feature vectors (b), and its reconstruction from the sequence of prototypical feature vectors as quantized by the Kohonen network (c), the solution space, each '-' indicating an allograph stroke (d), the target classification as provided by the user, each '*' indicating strokes that are not part of an allograph (e). Circles indicate the existence of loops as coded by the loop area feature λ_e.

Realization

For the coding of the ascender and descender contour of a word the following coding scheme is proposed. Contours are assumed to be equal if their pattern of ascenders, descenders, and body-sized objects correspond. The body-size characters $(a, c, e, i, m, n, o, r, s, u, v, w, x, z)$ are recoded as `o`, the descender stroke of (g, j, p, q, y) is recoded as `j`, the ascender stroke of (b, d, h, k, l, t) is recoded as `l`, whereas the (f) is a unique class `f` because it spans both the ascender and descender area in cursive script. In this coding, a b is a combination of an ascender object and a body-sized object, i.e., `lo`. This coding assumes that the letters as such have been identified. However, if a repetition of N body-size characters N`o` is coded by `x`, a compressed coding is formed which is not based on the number of letters in a word. For instance, the word '*they*' can be coded by `llooj` in letter-dependent code, and by `llxj` in compressed code. For the time being no special attention is paid to the allographs with dots (i, j).

Performance

Although the word-hypothesization stage has not yet been integrated it is of interest to mention its potential performance. Letter-dependent contour coding of a Dutch lexicon of 48000 common words yielded a collision of 4 word hypotheses on average for a given code pattern, with a worst case of 398 collisions for the code `oooooo`. A number of 84.5% of the codes has a number of collisions less or equal to the average of 4. Modal code pattern length was 9 codes.

Compressed contour coding yielded an average collision of 24 word hypotheses, with a worst case of 1953 collisions for the code `xlx`. In this case, a number of 89.3% of the codes has a number of collisions less or equal to the average of 24. Modal code pattern length was 5 codes.

The consequences of these figures for recognition are the following. First, letter-dependent coding is practically of no use since it is the letter identification itself which is the objective in cursive script recognition. Thus, only compressed contour coding is useful. The actual gain depends on the linguistic frequencies of the words in the different code groups. These frequencies are currently being analyzed.

6. Supervised learning

Purpose

Before a cursive-script recognition system is ready to work, it has to learn how to segment a writing pattern in the to-be-recognized allographs. The segmentation into allographs of handprint, with sufficient distance between individual allographs (e.g., spaced discrete characters, Tappert, 1986), would be relatively straightforward. If the written text is available, the learning module could just assign each allograph within the context of a word to a character. Although it is a rather cumbersome task to teach a system each allograph that may occur in a person's handwriting, it is currently still the most reliable procedure. The reason is that the allograph boundaries in cursive script have to be specified somehow (Footnote 2).

Several methods for performing this task in a non-supervised fashion are being developed (Morasso, 1990, Teulings et al., in prep.). Maier (1986) tried to segment an unknown writing trace into allographs using a-priori assumptions about the shape of the connection strokes between allographs. However, such a method produces persistent errors (e.g., segmenting allographs like cursive b, v, w, u, or y into two parts). Therefore, teaching is presently done interactively by the user.

Perceptual system

Although this stage is rather artificial it is still important to make the job as ergonomic as possible. During supervised learning the experimenter has to tell the system which parts of the handwriting trace belong to which allograph. It is relatively easy for the perceptual system if the user has to point only to complete strokes belong to a certain allograph. The initial connection stroke of the cursive allographs $a, c, d, g, i, j, m, n, o, p, q, u, x$ and y is not included and the initial connection stroke of the cursive allographs $b, e, f, h, k, l, r, s, t, v, w$ and z is included because it forms a strong perceptual cue for these allographs.

Realization

The software tool to teach the system the allographs displays a writing pattern with small circle markers on each stroke. The markers indicating the initial and final strokes of an allograph and the name of the allograph are successively clicked by using the mouse. Occasionally, N-gram names have to be entered by means of the keyboard. The naming of N-grams is needed when two allographs regularly 'melt' together because of increased writing speed. Typical fused digrams in Dutch handwriting are *or*, *er*, and *en* in many writers.

Performance

Once the procedure is running smoothly it takes on average 5 s per allograph to teach the system. After the teaching phase all allographs and their names can be made visible in order to assure that no mistakes have been made. Two handwritings were trained. The first handwriting (Writer A) was a neat constant-size handwriting and was trained incrementally up to 1671 prototypes by exposing the system to characters it could not discriminate or recognize well. The average number of strokes per allograph was 4.7. The second handwriting (Writer B) was a normal handwriting with considerable variation of allograph sizes with words. The allographs were trained from an a priori determined story of 240 words with low word frequencies. The script contained 1366 allographs (a posteriori), the average number of strokes per allograph being 2.9. The total script was written in 16 minutes.

Conclusion

It seems that the complex software system requires a powerful machine. A system inspired by the human motor system and the human perceptual system may seem to confine itself artificially. However, we see that the architecture is a very modular one (vertical modularity) and allows parallel modules (horizontal modularity). Problems can be very well located in one or two levels of the system. As such it seems that can be extended and tested relatively easily. The word hypothesization based on varying-length input sequences containing meaningless objects (e.g., connecting strokes) is currently a problem that has been solved only partially. It is to be hoped that robust artificial neural network models, handling noisy sequential data of unbounded lenght, will evolve in the future. This capability will be of special importance in languages like, e.g., German and Dutch, where nouns and prepositions plus nouns may be concatenated to form strings that are unlikely to be an entry in a standard lexicon.

References

Crossman, E.R.F.W. (1959). *A theory of the acquisition of a speed-skill.* Ergonomics, 1959, 2, 153-166.

Hayes, F. (1989). *True notebook computing arrives.* Byte,14 (Dec.), 94-95.

Jordan, M.I. (1985). *The learning of representations for sequential performance.* Phd Thesis. University of California, San Diego, pp. 1-160.

Kohonen, T. (1984). *Self-organisation and associative memory.* Berlin: Springer.

Maarse, F.J., Janssen, H.J.J., & Dexel, F. (1988). A special pen for an XY tablet. In F.J. Maarse, L.J.M. Mulder, W.P.B. Sjouw & A.E. Akkerman (Eds.), *Computers in psychology: Methods, instrumentation, and psychodiagnostics* (pp. 133-139). Amsterdam: Swets.

Maarse, F.J., Meulenbroek, R.G.J., Teulings, H.-L., & Thomassen, A.J.W.M. (1987). Computational measures for ballisticity in handwriting. In R. Plamondon, C.Y. Suen, J.-G. Deschenes, & G. Poulin (Eds.), *Proceedings of the Third International Symposium on Handwriting and Computer Applications* (pp. 16-18). Montreal: Ecole Polytechnique.

Maarse, F.J., Schomaker, L.R.B., & Teulings, H.L., (1988). Automatic identification of writers. In G.C. van der Veer & G. Mulder (Eds.), *Human-Computer Interaction: Psychonomic Aspects* (pp. 353-360). New York: Springer.

Maarse, F.J., & Thomassen, A.J.W.M. (1983). Produced and perceived writing slant: Difference between up and down strokes. *Acta Psychologica,* 54, 131-147.

Maier, M. (1986). Separating characters in scripted documents. *8th International Conference on Pattern recognition* (ISBN: 0-8186-0742-4), 1056-1058.

Morasso, P., Kennedy, J., Antonj, E., Di Marco, S., & Dordoni, M. (1990). Self-organisation of an allograph lexicon. Submitted to International Joint Conference on Neural Networks, Lisbon, March.

Morasso, P., Neural models of cursive script handwriting (1989). *International Joint Conference on Neural Networks,* Washington, DC, June.

Pick, H.L., Jr., & Teulings, H.L. (1983). Geometric transformations of handwriting as a function of instruction and feedback. *Acta Psychologica,* 54, 327-340.

Skrzypek, J., & Hoffman, J. (1989). *Visual Recognition of Script Characters; neural network architectures.* Technical report UCLA MPL TR 89-10, Computer Science Department University of California, Los Angeles

Srihari, S.N., & Bozinovic, R.M. (1987). A multi-level perception approach to reading cursive script. *Artificial Intelligence,* 33, 217-255.

Stornetta, W.S., Hogg, T., & Huberman, B.A. (1987). A dynamical approach to temporal pattern processing *Proceedings of the IEEE conference on Neural Information Processing Systems, Denver.*

Tappert, C. (1986). An adaptive system for handwriting recognition. In H.S.R. Kao, G.P. van Galen, & R. Hoosain (Eds.), *Graphonomics: Contemporary research in handwriting* (pp. 185-198). Amsterdam: North-Holland.

Tappert, C.C., Suen, C.Y., & Wakahara, T. (1988). On-line handwriting recognition: A survey. *IEEE*, 1123-1132.

Teulings, H.L., Schomaker, L.R.B. Gerritsen, J., Drexler, H., & Albers, M. (in prep.). An On-line handwriting-recognition system based on unreliable modules. In R. Plamondon and G. Leedham (Eds.), *Computer processing of handwriting.* Singapore: World Scientific. Presented at the 4th International Graphonomics Society Conference, Trondheim, July, 1989.

Teulings, H.L., Schomaker, L.R.B., Morasso, P., & Thomassen, A.J.W.M. (1987). Handwriting-analysis system. In R. Plamondon, C.Y. Suen, J.-G. Deschenes, & G. Poulin (Eds.), *Proceedings of the Third International Symposium on Handwriting and Computer Applications* (pp. 181-183). Montreal: Ecole Polytechnique.

Teulings, H.L., Schomaker, L.R.B., & Maarse, F.J. (1988). Automatic handwriting recognition and the keyboardless personal computer. In F.J. Maarse, L.J.M. Mulder, W.P.B. Sjouw, & A.E. Akkerman (Eds.), *Computers in psychology: Methods, instrumentation, and psychodiagnostics* (pp. 62-66). Amsterdam: Swets & Zeitlinger.

Teulings, H.L., Thomassen, A.J.W.M., & Van Galen, G.P. (1983). Preparation of partly precued handwriting movements: The size of movement units in writing. *Acta Psychologica,* 54, 165-177.

Teulings, H.L., Thomassen, A.J.W.M., & Van Galen, G.P. (1986). Invariants in handwriting: The information contained in a motor program. In H.S.R. Kao, G.P. Van Galen, & R. Hoosain (Eds.), *Graphonomics: Contemporary research in handwriting* (pp. 305-315). Amsterdam: North-Holland.

Teulings, H.L., & Maarse, F.J. (1984). Digital recording and processing of handwriting movements. *Human Movement Science,* 3, 193-217.

Thomassen, A.J.W.M., Teulings, H.-L., Schomaker, L.R.B., Morasso, P., & Kennedy, J. (1988). Towards the implementation of cursive-script understanding in an online handwriting-recognition system. In Commission of the European Communities: D.G. XIII (Ed.), *ESPRIT '88: Putting the technology to use.* Part 1 (pp. 628-639). Amsterdam: North-Holland.

Thomassen, A.J.W.M., Teulings, H.L., Schomaker, L.R.B., Morasso, P., & Kennedy, J. (1988). Towards the implementation of cursive-script understanding in an online handwriting-recognition system. In Commission of the European Communities: D.G. XIII (Ed.), *ESPRIT '88: Putting the technology to use.* Part 1 (pp. 628-639). Amsterdam: North-Holland.

Thomassen, A.J.W.M., & Teulings, H.L. (1985). Time, size, and shape in handwriting: Exploring spatio-temporal relationships at different levels. In J.A. Michon & J.B. Jackson (Eds.), *Time, mind, and behavior* (pp. 253-263). Heidelberg: Springer.

Watrous, R. & Shastri, L. (1987). Learning phonetic features using connectionist networks. *Proceedings of the 1987 IJCAI, Milano*, pp. 851-854.

Footnote 1.

In order to derive an estimation of the SNR, each replication's pattern of stroke data is considered as the sum of an invariant component (signal) and an uncorrelated, movement-impaired component (noise). Of course, the signal representing metric data per stroke (e.g., S) may have a small over-all multiplied parameter and the signal representing directions per stroke (e.g., ϕ_1) may have a small over-all added parameter. Therefore, it is wise to correct for any overall parameter. The variance of the pattern equals the variance of the signal plus the variance of the noise. Furthermore, the variance of the pattern of stroke data averaged over n replications equals the variance of the signal plus only 1/n times the variance of the noise. From these two equations, the standard deviations of the signal and the noise can be calculated as well as their ratio, the SNR.

Footnote 2.

The on-line digitized handwriting training and test sets are available on paper and on-line digitized with x, y and pressure data in a selfexplanatory, documented ASCII format on an MSDOS 3.5" DS diskette.

About the authors:

Lambert Schomaker has studied muscle control from the psychophysiological point of view and is currently studying handwriting movement control and artificial neural modeling aspects from the psychomotor perspective. His main interests are neural network models for movement control, robotics, and signal processing in movement analysis.

Hans-Leo Teulings received his PhD in 1988 on Handwriting-Movement Control at the University of Nijmegen, The Netherlands. Current interests range from the human representation of handwriting movements, the recognition aspects, to handwriting impairment in Parkinson patients, and handwriting-movement training using computers.

Acknowledgment

This research was supported by ESPRIT Project 419, Image and Movement Understanding.