The NicIcon Database of Handwritten Icons for Crisis Management

Ralph Niels, Don Willems and Louis Vuurpijl
Nijmegen Institute for Cognition and Information
Radboud University Nijmegen
Nijmegen, The Netherlands
{r.niels;dd.willems;vuurpijl}@nici.ru.nl

Abstract

The goal of this paper is to announce the NicIcon collection of handwritten sketches containing iconic gestures. These data have recently been collected in our group with the goal to develop and assess pen input recognition technologies for the domain of crisis management. In our envisaged scenarios, users can use the pen to enter, e.g., information about the location of certain objects or the occurrence of certain events, on a depicted photograph or digital map. Typically, these sketches contain representations of drawings that are not contained in publicly available databases, which mainly contain handwritten texts. We report on our classification performances achieved for these data and make the data freely available to the handwriting recognition community.

Keywords: database,icons,sketch recognition

1. Introduction

Databases for training and testing automated reading systems have become readily available for the handwriting recognition community. These databases mostly contain either online or offline data, acquired on a multitude of acquisition devices. The first large publicly available datasets were UNIPEN [7], consisting of online isolated characters and handwritten words from Western scripts, and the CEDAR database [8], containing offline scans of address information like city names, state names, and zip codes. Other well-known databases containing Western handwriting are IRONOFF [17] and the IAM database [10]. The past decade, many other scripts like Japanese [9, 11], Tamil [2, 3], and Arabic [14] have become available. Because of these databases, many researchers in handwriting recognition have been able to develop and assess their recognition technologies on generally accepted benchmarks. Besides handwriting, other categories of pen input are musical scores, mathematical expressions, command gestures, and sketches. Unlike the availability of handwriting databases, collections of these other categories are scarcely available. The goal of this paper is to announce the NicIcon collection of handwritten sketches containing iconic gestures. Similar to the IRONOFF collection [17], the NicIcon collection comprises simultaneously acquired online pen trajectories and offline scans. In this paper, our first classification performances on these data are reported and a number of challenging future research issues are described. The NicIcon collection will be made freely available (through the `unipen.org` website) to the research community, with the aim to stimulate other researchers to develop and assess their technologies on the relatively unknown domain of iconic sketches.

Within our research on multimodal interaction, the recognition of sketches has played a prominent role. For example, in the European COMIC project, we developed a conversational bathroom salesman that provided the user with options to sketch a blueprint of their bathroom [4]. In the ICIS project [15, 22], pen-based sketches are used to indicate events on interactive maps in the domain of crisis management. In the scenarios pursued in ICIS, users can draw sketches of situations (on a pen input device like a tabletPC or PDA). We previously reported about a large study on the typical pen interactions that emerge in these scenarios in [20]. The categorization of the obtained pen gestures showed that next to route descriptions and markings of locations, the iconic sketchings of, e.g., houses, cars, fires, and persons, occurred quite frequently. We also concluded that the recognition of these unconstrained sketches (with a large variability within and between users), resulted in a recognition performance that was unacceptable for the envisaged target domain.

Therefore, we performed another significant data collection effort with the goal to reduce the variability in the data. As discussed in [6], users can convey rather complex messages using a limited set of icons. Inspired by [6], we designed a set of 14 icon shapes representing important information relevant to the domain of crisis management. In total 32 participants were asked to sketch these iconic representations of events on a digitizing tablet. The resulting NicIcon database contains 24,441 icons. In the subsequent sections, we will (i) describe the NicIcon data collection process and contents of the data, (ii) demonstrate the performance of our feature extraction and recognition techniques on the online part of the data, as a baseline for future research, (iii) provide details on how to obtain the database and (iv) conclude with a number of challenging research issues.

2. Database

A set of 14 icons that are important in the domain of crisis management and incident response systems was selected. The icons were designed such that (i) they have a visual resemblance to the objects they represent or correspond to well known corresponding symbols (so that they are easy to learn by the users), and (ii) are distinguishable by the computer. Figure 1 shows the 14 selected icons.

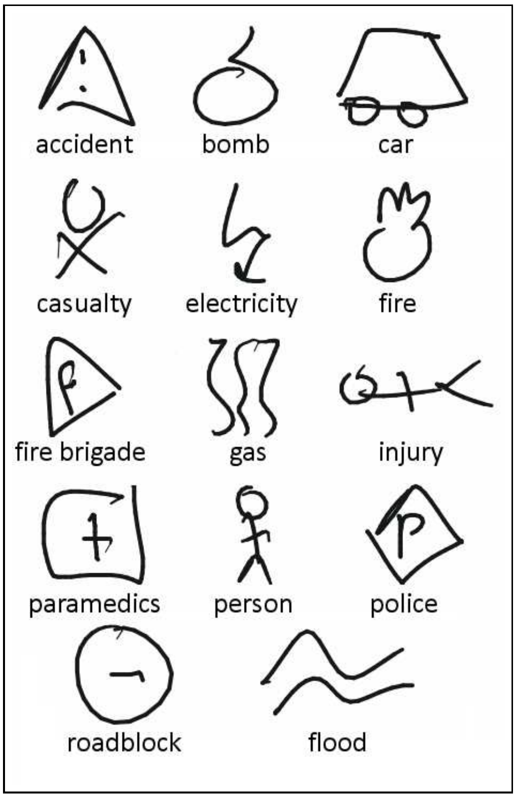

Figure 1. Examples of the 14 different icons, as they were produced by one of the participants. The numbers correspond to the numbers in Table 1.

In total, 32 participants, all volunteers, participated in the experiment. They were all Dutch students in the age range of 19 to 30 ($\mu = 21.63, \sigma = 2.35$), of which 29 were male, and 3 were female.

2.1 Forms

Each participant was asked to fill out 22 paper forms, using an inking stylus. Each form contained 35 boxes arranged in 7 rows and 5 columns, two calibration crosses and an identification area. Each box measured 22x22 mm in size. Left to each row, the stimulus (the to-be-drawn icon and a size instruction ('small', 'medium' or 'large')) was specified. The participants were asked to fill each row of boxes by drawing one instance of the specified icon in each box. The first two forms were intended as trial forms, which would prepare the participants to the real task. The forms contained one row for each of the 14 different icons, and no size specification was given. Since the quality of the icons produced on the trial forms was comparable to the quality of the rest of the icons, we decided to add the trial icons to the database. They were marked as being trial icons, however (see Section 2.4).

2.2 Procedure

For each form (see Figure 2), the participants were asked to first touch the centers of the calibration crosses with the pen tip. Knowing the location of these crosses on the tablet allowed us to automatically segment the online data, using knowledge of the form layout and box locations. Furthermore, these locations are required for aligning the online trajectories to the corresponding offline scanned images. Subsequently, the participants were asked to mark the empty spaces in the binary code box with small crosses. By doing this, they attached the unique code of the form to the online data. After the calibration process of the form, the participants were asked to fill each row of query boxes by drawing five instances of the icon specified for that row. No strict instructions were given about these size specifications which were varied to get some size variations. Each participant was asked to draw the same icons/size combinations, but the order in which they were drawn was randomized.

2.3 Equipment

A *Wacom Intuos2 A4 oversize* tablet was as digital input device. This tablet has a resolution of 100 lp/mm, an accuracy of +/- 0.25, a reading height of 10 mm, and a maximum data rate of 200 pps. The device can distinguish 1024 pressure levels. The tablet was connected to a computer running Microsoft Windows XP. Specially developed software was used to record spatial coordinates, time coordinates and pressure coordinates. An A4-sized sheet of 160 grams/m^2 paper was clamped to the tablet. A Wacom writing pen, with a green *Lamy M21* pen tip[1], was used by the participants to draw the data.

Each sheet of paper was scanned using a HP Scanjet

[1] www.lamy.com/eng/b2c/Refills and inks/M 21

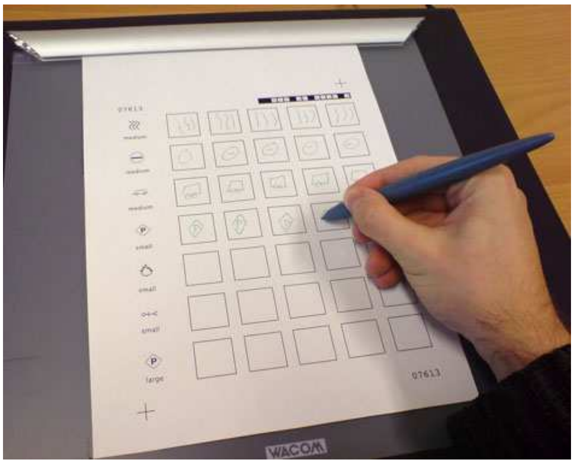

Figure 2. A piece of paper was clamped to a writing tablet. The subjects used an inking stylus to draw the icons. The resulting data contains online trajectories and offline scans of the paper.

7400C flatbed scanner at 300dpi resolution in 24bits color. Note that, although we simultaneously captured both online and offline data, in this paper we focus only on the online data. Both online and offline data are available for download as described in Section 5.

2.4 Data

Each paper sheet contained seven rows of five columns, resulting in 35 drawing areas. Each of the 32 participants had to fill in 22 paper sheets, resulting in 770 icon gestures per participant. Some participants appeared to have skipped certain gestures (199 gestures in total), so the total number of usable iconic ink gestures is 24,441. The distribution of the data per icon can be seen in Table 1.

Table 1. Distribution of the 24,441 gestures over the 14 icon classes.

Description	Icon	n	Description	Icon	n
Accident		1736	Gas		1745
Bomb		1750	Injury		1775
Car		1720	Paramedics		1780
Casualty		1750	Person		1755
Electricity		1735	Police		1740
Fire		1740	Roadblock		1750
Fire brigade		1725	Flood		1740

The data was saved in an ASCII-based HWR-format and was later converted to Unipen format [7] format. Using the calibration crosses and unique code, we were able to segment and label each data sample automatically. Each icon was labeled with the participant ID, the page number, the row and column number, the type of the icon, and the size (S/M/L/Trial).

3. Recognition experiment

In this section, we report on the typical recognition performances that can be obtained on these data. Three different classifiers were trained to distinguish between the different icons. We used a multilayered perceptron (MLP) and a linear multi-class SVM classifier using three types of features: the 28 features described by Willems [19, 21], and two instances of feature sets computed as described in [18]. Furthermore, we used the Dynamic Time Warping (DTW) implementation described by Niels et al. [13].

The classifiers were organized in a multiple-classifier system (MCS), which employed majority voting [16] for combining classifier outputs. In case of a tie, the output of the DTW classifier was used to rule the outcome.

For a writer dependent (WD) test, the full set of icons was randomly divided into a train set of 60%, and a test set of the other 40% of the icons. For a writer independent (WI) test, the train set contained all icons written by 60% of the writers, and the test set contained all icons written by the other 40% of the writers.

3.1 Features

Three feature sets were computed from the icon data, respectively referred to as g-28, t-30, and t-60. Note that the g-28 features are global features computed from each total trajectory, whereas the t-30 and t-60 features comprise local features running along each trajectory. The g-28 features are described in [19, 21] and employed in a pen gesture recognition system used for interactive map applications. Most pen gesture data, on which these features were tested, consisted of deictic gestures [20] that users produced when marking objects on the interactive maps, like arrows, encirclements, and crosses. Handwriting and some iconic gestures were also produced while using the interactive map application. These 28 features included spatial features: (1) length of pen stream, (2) area of convex hull, (3) compactness, (4) eccentricity, (5) ratio of the coordinate axes, (6) closure, (7) circular variance, (8) curvature, (9) average curvature, (10) perpendicularity, (11) average perpendicularity, (12) centroid offset along major axes, (13) length along first principle axis, (14) rectangularity, (15) initial horizontal offset, (16) final horizontal offset, (17) initial vertical offset, (18) final vertical offset, (19) straight line count, (20) largest straight line ratio, (21) straight line ratio, (22) largest straight line, (23) macro perpendicularity, (24) average macro perpendicularity, and (25) ratio of the principal axes. Also included were three features based on the pressure data also available in the online data: (1) average pressure, (2) pen down count, (3) pen up/down ratio.

The t-30 and t-60 features are described in [18]. These trajectory-based features are computed from spatially resampled pen strokes. The features that are calculated for each resampled data sample containing n points are the the x, y and z-coordinates, the running angles, and angular velocities, resulting in $3*n+2*(n-1)+2*(n-2)$ features. As explained in [18], a typical resampling of Western characters requires $n = 30$ (204 features). Given that many of the collected iconic gestures have a more complex shape than the average Western character, we also explored resampling to $n = 60$ (414 features), resulting in a better coverage of the original trajectory with resampled points.

3.2 Classifiers

Three algorithms were used to experiment with the three feature sets discussed above. The first two classifiers are relatively standard algorithms which require little explanation. For the first classifier (SVM), we used public domain software implementing the linear multi-class LIBSVM-classifier [5]. The data in the train sets (writer dependent or writer independent) was used to develop the SVM models, and the data in the test sets (writer dependent or writer independent) was used to evaluate the performance of the classifier. For the second classifier (MLP), several architectures and learning parameters were varied to train multilayered perceptrons on the three feature sets. The third classifier uses the dynamic time warping (DTW) algorithm described in [13], which calculates the DTW-distance between two data samples by summing the normalized Euclidean distances between the matching coordinates of two data samples. Whether or not two coordinates i of sample A, and j, of sample B match, is decided using three conditions: (i) the continuity condition, which is satisfied when coordinate i is on the same relative position on A as coordinate j is on B, (ii) the boundary condition, which is satisfied if i and j are both at the first, or both at the last position of their sample, (iii) the penup/pendown condition, which is satisfied when both i and j are produced with the pen on the tablet, of when they are both produced with the pen above the tablet. i and j match if either the boundary condition, or both other conditions are satisfied. Classification of a test sample was performed through nearest neighbor matching with the DTW distance function.

3.3 Multi-classifier system

As described above, 7 classifiers were used to classify the data: the SVM-classifier and the MLP-classifier (each using the 3 described feature sets), and the DTW-classifier. The resulting classifications were combined in a simple majority voting MCS. Table 2 shows the correct classification performance for each classifier (where a classifier is the combination between the used classification method and the used features) and the overall performance of the MCS, for both the writer dependent and the writer independent setting.

Table 2. Correct classification performances of the individual classifiers using the different feature sets and the multi-classifier system on writer dependent (WD) and writer independent (WI).

Classifier	Feature set	WD perf.	WI perf.
SVM	g-28	84.39%	79.99%
	t-30	98.57%	92.63%
	t-60	98.51%	92.16%
MLP	g-28	94.00%	90.72%
	t-30	96.63 %	92.40%
	t-60	97.79%	92.92%
DTW	-	98.06%	94.70 %
MCS	-	99.32%	96.49%

When observing Table 2, it shows that the trajectory-based feature sets t-30 and t-60 result in a better performance than the global features g-28. Furthermore, the results of the MCS significantly improve each single classifier. Although this can be expected, leaving any of the classifiers out of the MCS results in a lower performance. Apparently, although the g-28 features are less discriminative, they provide necessary extra information that improves recognition.

3.4 Conflicting classes

To explore conflicting classes between the different icon shapes, below the confusion matrix for the writer dependent setting is presented (see Table 3). From this confusion matrix, it can be concluded that especially the 'boxed' icons (icons which consist of a base in a frame: , , and ) are harder for the system to distinguish. As illustrated in Figure 3, the main reason for this confusion is that users in some occasions mix up the shape and orientation of the boxes. Two other categories of typical misclassifications are cases where users retrace the pen trajectories and cases where the pen input is sloppy.

4. Discussion and future challenges

In this paper, we have introduced a new dataset containing 24,441 hand drawn icons from the domain of crisis management. The 14 icon classes were designed and constrained such that they (i) were easy to learn by human subjects and (ii) they can be distinguished well by computer software. We have demonstrated high recogni-

Table 3. Confusion matrix of the writer dependent classification. Hyphens ('-') denote confusions of 0.0%.

	Predicted						
Real							
	99.4	-	-	-	-	0.1	0.3
	-	100.0	-	-	-	-	-
	0.1	-	99.9	-	-	-	-
	-	0.1	-	99.6	-	-	-
	-	-	-	0.1	99.6	-	0.1
	-	-	-	0.1	-	99.9	-
	0.3	-	-	0.2	-	0.2	98.9
	-	-	-	-	-	-	0.1
	-	-	0.1	-	-	-	0.1
	0.4	-	0.1	-	-	0.3	0.4
	-	-	-	0.4	0.1	0.1	-
	-	-	0.1	-	-	0.4	0.7
	-	0.1	-	-	-	-	0.7
	-	-	0.4	-	-	0.1	0.1
Real							
	-	-	-	-	-	0.1	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	0.3	-	-	-
	-	-	-	-	0.1	-	-
	-	-	-	-	-	-	-
	-	-	0.2	-	0.3	-	-
	99.7	-	0.1	-	-	-	-
	-	99.7	-	-	-	-	-
	-	-	98.4	0.1	-	0.1	-
	-	-	-	99.3	-	-	-
	-	-	0.1	0.3	98.2	-	-
	-	-	0.6	-	-	98.6	-
	-	-	-	-	-	-	99.3

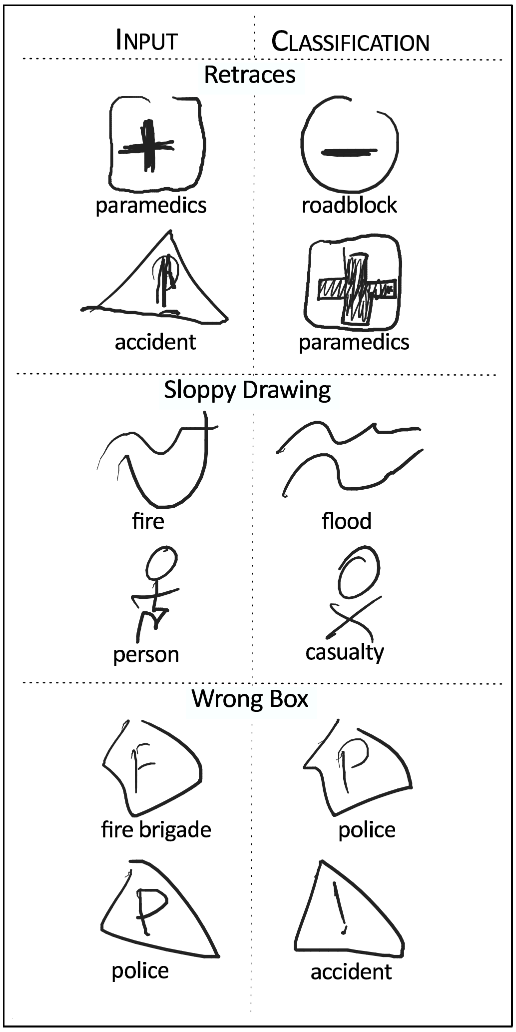

Figure 3. Examples of the three main conflicting cases: (i) retraces, (ii) sloppy drawings and (iii) wrong box shapes.

tion performances by using three feature sets and combining three different classifiers. Using a multiple-classifier system ruled by majority voting, 99.3% of the writer-dependent and 96.5% of the writer independent test samples were classified correctly. Trajectory-based features result in a significantly better performance than when using global features. An analysis of the conflicting error classes revealed that — apart from apparent slips of the user —, in particular the boxed classes and cases where a user retraces the trajectories yield conflicts. One approach to solve the confusion of boxed classes, might be to remove the box part and apply character recognition to the remaining part.

We are currently further developing and assessing our pen input recognition technologies in more elaborate experiments, involving pen input data acquired in less constrained situations. Two settings will be explored: (i) an evaluation of our new recognition systems on the less constrained gesture repertoire described in [20], and (ii) a new interactive experiment in which new data will be collected from novel writers in a less-constrained setting as well. Both datasets will be made available as part of the NicIcon collection.

Our second research interest in these data is concerned with our work on UPX, an upcoming standard for the specification of online data[1]. A relatively underspecified issue in UPX is the specification of links between multimodal data like the NicIcon collection introduced here. Our goal is to use the NicIcon data as a testbed for developing a specification format for linking regions of interest (in the scanned image) to online trajectories specified in inkML.

The third direction concerns the combination of the online data recorded on the digitizing tablet and the offline scanned images. Here, not only offline recognition of iconic gestures will be pursued, but the availability of pen input data acquired in both modalities also opens up the possibilities of (i) exploring the relation between ink distribution as present in the offline scan and the online trajectories containing pressure information, and (ii) performing research on the extraction of dynamic information from the scanned images. The problem of extracting the intended trajectory from offline data is still unsolved and having the new NicIcon collection available may contribute to new steps ahead in this area.

5. Obtaining the database

The NicIcon dataset introduced in this paper is freely available for download. Online data is stored in the Unipen [7] format and the PNG image format is used for the offline images. Please visit `http://unipen.nici.ru.nl/NicIcon` for download instructions and further details about the datasets, like writer population, class distributions and the coupling between online and offline samples.

Acknowledgments

This research is sponsored by the projects ICIS-CHIM [15] and NWO Trigraph [12]. The authors would like to thank Gerard van Oijen and Chris Bouwhuisen for their help in setting up the experiment.

References

[1] M. Agrawal, K. Bali, S. Madhvanath, and L. Vuurpijl. UPX: A new XML representation for annotated datasets of online handwriting data. In *Proc. Intl. Conf. on Document Analysis and Recognition (ICDAR'05)*, pages 1161–1165, Seoul, Korea, August 2005. IEEE.

[2] A. S. Bhaskarabhatla and S. Madhvanath. Experiences in collection of handwriting data for online handwriting recognition in indic scripts. In *Proc. of the 4th Int. Conf. Linguistic Resources and Evaluation (LREC)*, 2004.

[3] U. Bhattacharya. Handwritten character databases of Indic scripts, 2004. www.isical.ac.in/~ujjwal/download/database.html.

[4] L. Boves, A. Neumann, L. Vuurpijl, L. ten Bosch, S. Rossignol, R. Engel, and N. Pfleger. Multimodal interaction in architectural design applications. In *8th ERCIM Workshop on User Interfaces for AI*, Vienna, Austria, 2003.

[5] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at www.csie.ntu.edu.tw/~cjlin/libsvm.

[6] S. Fitrianie and L. Rothkrantz. A visual communication language for crisis management. *Int. Journal of Intelligent Control and Systems (Special Issue of Distributed Intelligent Systems)*, 12(2):208–216, June 2007.

[7] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet. UNIPEN project of on-line data exchange and recognizer benchmarks. In *Proc. ICPR'94*, pages 29–33, Oct. 1994.

[8] J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), May 1994.

[9] S. Jaeger and M. Nakagawa. Two on-line japanese character databases in unipen format. In *Proc. Intl. Conf. Document Analysis and Recognition, ICDAR'01*, pages 566–570. IEEE Computer Society, 2001.

[10] U. Marti and H. Bunke. A full english sentence database for off-line handwriting recognition. In *Proc. of the 5th Int. Conf. on Document Analysis and Recognition (ICDAR'99)*, pages 705–708, Bangalore, India, 1999.

[11] M. Nakagawa and K. Matsumoto. Collection of on-line handwritten Japanese character pattern databases and their analysis. *Int. Journal on Document Analysis and Recognition*, 7(1), 2004.

[12] R. Niels, L. Vuurpijl, and L. Schomaker. Introducing trigraph - trimodal writer identification. In *Proc. European Network of Forensic Handwr. Experts*, Budapest, Hungary, 2005.

[13] R. Niels, L. Vuurpijl, and L. Schomaker. Automatic allograph matching in forensic writer identification. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 21(1):61–81, 2007.

[14] M. Pechwitz, S. Maddouri, V. Mrgner, N. Ellouze, and H. Amiri. Ifn/enit-database of handwritten Arabic words. In *7th Colloque Int. Francophone sur l'Ecrit et le Document (CIFED02)*, Hammamet, Tunis, Oct. 21-23 2002.

[15] The ICIS project. Interactive collaborative information systems ICIS, 2004. www.icis.decis.nl/.

[16] M. van Erp, L. Vuurpijl, and L. Schomaker. An overview and comparison of voting methods for pattern recognition. In *Proc. of IWFHR8*, pages 195–200, Niagara-on-the-Lake, Canada, August 2002.

[17] C. Viard-Gaudin, P. M. Lallican, P. Binter, and S. Knerr. The ireste on/off (ironoff) dual handwriting database. In *Proc. Intl. Conf. Document Analysis and Recognition, ICDAR'99*, pages 455–458, Bangalore, India, 1999.

[18] L. Vuurpijl and L. Schomaker. Finding structure in diversity: A hierarchical clustering method for the categorization of allographs in handwriting. In *Proc. ICDAR4*, pages 387–393. IEEE Computer Society, Aug. 1997.

[19] D. Willems, S. Rossignol, and L. Vuurpijl. Mode detection in online pen drawing and handwriting recognition. In *Proc. of the 8th Int. Conf. on Document Analysis and Recognition (ICDAR2005)*, pages 31–35, Seoul, Korea, 2005.

[20] D. Willems and L. Vuurpijl. Pen gestures in online map and photograph annotation tasks. In *Proc.of the tenth Int. Workshop on Frontiers in Handwriting Recognition (IWFHR06)*, pages 297–402, La Baule, France, October 2006.

[21] D. Willems and L. Vuurpijl. A bayesian network approach to mode detection for interactive maps. In *Proc. of the 9th Int. Conf. on Document Analysis and Recognition (ICDAR2007)*, Curitiba, Brazil, 2007.

[22] D. Willems and L. Vuurpijl. Designing interactive maps for crisis management. In *Proc. of the 4th Int. Conf. on Information Systems for Crisis Response and Management (ISCRAM2007)*, pages 159–166, 2007.