Designing interactive maps for crisis management

Don Willems
Nijmegen Institute for Cognition and Information
Radboud University Nijmegen
dd.willems@nici.ru.nl

Louis Vuurpijl
Nijmegen Institute for Cognition and Information
Radboud University Nijmegen
vuurpijl@nici.ru.nl

ABSTRACT

This paper describes the design, implementation, and evaluation of pen input recognition systems that are suited for so-called interactive maps. Such systems provide the possibility to enter handwriting, drawings, sketches and other modes of pen input. Typically, interactive maps are used to annotate objects or mark situations that are depicted on the display of video walls, handhelds, PDAs, or tablet PCs. Our research explores the possibility of employing interactive maps for crisis management systems, which require robust and effective communication of, e.g., the location of objects, the kind of incidents, or the indication of route alternatives. The design process described here is a mix of "best practices" for building perceptive systems, combining research in pattern recognition, human factors, and human-computer interaction. Using this approach, comprising data collection and annotation, feature extraction, and the design of domain-specific recognition technology, a decrease in error rates is achieved from 9.3% to 4.0%.

Keywords

Interactive maps, pen input recognition systems, Bayesian networks, mode detection, crisis management.

1. INTRODUCTION

Because of the current need for improved incident response systems and ongoing advances in the technological capabilities of such systems, the use of information systems supporting various services in crisis management is rapidly increasing. In the ICIS project, novel interactive collaborative information systems are being designed for such applications. The ICIS/CHIM research cluster combines a number of partners from research and industry and pursues the development of multimodal interfaces for crisis management tasks and incident response systems. The technologies developed in CHIM are mainly targeted at professional users operating in crisis control rooms and mobile communication centers at the scene of an incident. The motivation for this research is based on the fact that for conveying information in such applications, robust and efficient human to computer and computer to human interaction is vital. It is known that humans interact best with computer systems when they are allowed to do so in a natural way, just as when they interact with other humans (Cohen, Johnston, McGee, Oviatt, Pittman, Smith, Chen, and Clow, 1997a, 1997b; Oviatt, 2003; Willems and Vuurpijl, 2006). Since for this purpose, a varied set of modalities (like speech, gesture, drawing, and writing) is available for communication, research on multimodal systems has become increasingly important in the last decade. Nevertheless, still no guidelines for the design of multimodal systems exist. Experiences from other projects, like SMARTKOM (Wahlster, 2006), COMIC (den Os and Boves, 2004), and the works from Oviatt (Oviatt, 2003), show that designing multimodal systems requires co-operation between various research groups, requiring significant collaborative software engineering efforts to ensure the correct mutual operation of distinct processing modules involved. Furthermore, choosing the optimal modality for conveying information depends on a myriad of factors, like the type of information to be relayed, the human preference and capabilities, local environmental conditions, and the availability of the proper interaction devices.

This paper reports on the design of a pen interface for crisis management. Advances in pen-aware systems like interactive video walls, PDA's, handhelds, and tablet PCs have lead to the possibility of capturing pen-based information on the display of computer systems. Such systems, by which users can annotate objects on rendered maps or visualized photographic content (Schomaker, Vuurpijl, and Leau, 1999), or draw maps or blueprints (den Os and Boves, 2004, Rossignol, Willems, Neumann, and Vuurpijl, 2004), are called *interactive maps*. Pen input is particularly appropriate for the communication of, e.g., the location of objects or events, or the specification of routes (Willems and Vuurpijl, 2006). Interactive maps are important tools to enhance interaction between different actors, in particular where spatial information is concerned (Montello, 2001, Cohen et al, 1997a). In interactive maps, different broad types of pen gestures can be distinguished. The classification of these types of pen gestures is called *mode detection*. By reliably detecting the mode of a pen gesture and subsequently engaging the appropriate mode-specific recognizer, the efficiency and robustness of interactive map applications can be enhanced (Willems, Rossignol, and Vuurpijl, 2005a). Typical modes of pen gestures that can be distinguished are: deictic gestures, handwritten texts, and iconic objects. Deictic gestures are pen gestures that are used to identify spatial information. Handwritten text and iconic objects are mostly used for tagging objects or adding new descriptive information. The latter may include sketches of people, cars, fires, etc. (Willems and Vuurpijl 2006). Research on mode detection is relatively new. Previous work includes research by Jain (Jain et al., 2001), Rossignol (Rossignol, et al., 2004) and Bishop (Bishop, Svensèn, and Hinton, 2004), and focused on the distinction between handwriting and drawings.

Although guidelines are lacking for the design of such systems, a suitable approach may be found in (den Os and Boves, 2004; Vuurpijl, ten Bosch, Rossignol, Neumann, Pfleger, and Engel, 2004; Willems and Vuurpijl, 2006). This approach consists of the following steps: (i) use a set of recognizers to determine a baseline performance, (ii) collect and analyze experimental data from human subjects within the given application domain and possibly in interaction with the recognition system, (iii) further improve and train the recognition technologies on the basis of these data, and (iv) asses the performance of the improved recognizers. The research presented in this paper follows this approach, which is typical for the design of any perceptive system. Below, in Section 2, an introduction of our pen-input recognition technology for interactive maps is provided. These mode detection systems have been trained on data acquired from other projects, concerned with applications differing from crisis management. Subsequently, in Section 3, we report on the results of a human-factors experiment which has yielded a taxonomy of interaction data acquired in the context of interactive maps. Section 4 describes the results of our current experiments which focus on the design of improved pen-input mode-detection technology, based on the newly acquired data. This paper is concluded in Section 5, which contains a discussion and pointers to future research.

2. BASELINE PEN-INPUT RECOGNITION TECHNOLOGY

The baseline mode detection system distinguished between four classes: (i) handwritten text, (ii) arrows, (iii) lines, and (iv) geometric objects like rectangles, circles, and triangles (Willems et al., 2005a). The data used for training and testing originated from the COMIC project (Boves, Neumann, Vuurpijl, ten Bosch, Rossignol, Engel, and Pfleger, 2003), a dataset provided by Fonseca and Jorge. (Fonseca and Jorge, 2001) (containing arrows, lines, and geometric objects), and handwritten text data from the UNIPEN database (Guyon, Schomaker,Plamondon, Liberman, and Janet, 1994). A kNN and a multi-layered perceptron classifier were trained with eight simple geometric features, like the length, the area, and the curvature of the pen trajectory (see Figure 1).

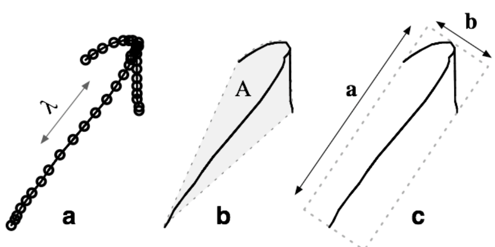

Figure 1. Examples of geometric features: (a) The length (λ) of the pen trajectory. (b) The area (A) of the convex hull of a pen trajectory. (c) The ratio of the principle axis of the bounding box (a/b).

The obtained maximum mode detection performance was 98.7% for the kNN classifier. Among the 14 errors that were produced by this system, ten were misclassifications of arrows. To improve on mode detection and to distinguish more objects, a hierarchical mode detection system with three feature classifiers and one shape matching classifier (see Figure 2) was designed (Willems, Rossignol, and Vuurpijl, 2005b). The first classifier (HWR-DRAW) distinguished between handwritten text (HWR) and drawings. If the pen gesture was classified as a drawing (DRAW), the next classifier (LINEAR-GEOM) was used to distinguish between linear drawings (lines and arrows) and geometric objects. Linear drawings were classified as lines or arrows by the third feature classifier (LINE-ARROW). Geometric objects (rectangles, triangles, ellipses, or diamonds) were fed into a shape matching classifier. The recognition performance of this hierarchical system for the same data that was used to test our first system was 99.2%. The overall mode detection performance decreased to 95.6%, mainly because the classifier had to distinguish between four more classes (the geometric objects).

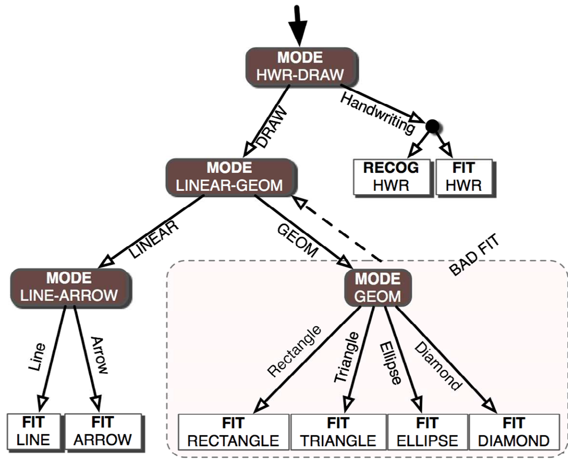

Figure 2. A hierarchical classification system for pen gestures. This system consists of three feature classifiers and one shape matching classifier. After classification takes place, a fit of the gestures to template objects may be performed.

3. ACQUISITION, ANNOTATION, AND RECOGNITION OF DOMAIN-SPECIFIC INTERACTION DATA

A significant problem in pen input recognition concerns the multitude of different pen gestures users may generate if they are totally unconstrained in the gesture vocabulary that they may use. This problem holds for any perceptive system which has to understand the meaning of user input: The larger and more complex the possible input categories, the more recognition errors are bound to result, due to the variability that has to be accounted for by the classification system. Three solutions to this problem are: (i) do not constrain the user and handle errors when they occur, (ii) provide a limited gesture vocabulary, to be learned by the user, or (iii) constrain the number of possible gestures depending on the domain and overall user preferences. The first option would imply complex error detection and correction techniques. Such techniques require extensive knowledge and representations of the domain and potential dialogues with the user, and imply more complex and integrated system components in which dialogue management plays a more central role (McTear, 2002). Unfortunately, still little is known about the gesture repertoires and interaction dialogues people employ when using interactive maps. Furthermore, since the user would have to correct the recognition system relatively often, effectiveness and confidence in the system will be low. The second option is employed by a number of devices such as Palm PDA's, which use a limited constrained vocabulary (Goldberg and Richardson, 1993) to improve recognition accuracy. While such a vocabulary is easily recognized, it requires a definition of a proper gesture lexicon and furthermore, users have to be engaged in extensive training sessions before the system can be used at a reliable level of interaction. The latter is an important factor for the acceptance of novel interaction technology like pen-aware systems: Participants engaged in crisis management have frequently reported that modifications in their modus operandi are not appreciated.

In our approach, the design of pen gesture types that are the result of user preferences and adapted to the domain in which the recognition system will be employed. These gestures are collected through experiments in which users interact with maps and photographic image content from the domain of crisis management. The acquired gestures are subsequently optimized on minimal complexity, thereby adhering to effectiveness, easy-to-learn, easy-to-remember, and easy-to-use principles (Dix, Finlay, Abowd, and Beale, 2004). The system will, therefore, eventually become better adapted to the user, enabling pen-based interactions that are natural to use. To facilitate improved accuracy, the design of the gesture set should maximize the distinction between gestures, thereby creating robust and reliable pen gesture recognition.

The participants were seated in front of an LCD-tablet (basically an LCD-screen on which one can write with a digital pen), and presented with a map or photograph. They had to perform tasks (using a pen) such as "Indicate the location of the fire" or "Indicate the route from A to B". Twelve people participated in this experiment resulting in a set of 14,210 annotated pen gestures at all levels in the taxonomy (Figure 3) that was extracted from analysis of the data. Both authors of this paper annotated all acquired data. Annotation was guided by the expected gestures corresponding to the tasks at hand (for example, when indicating the location of a fire, deictic gestures were expected as well as iconic drawings of bonfires). Furthermore, in cases where annotations between the authors differed, resolution of such a conflict was performed by a second visual inspection. This process guaranteed a high quality of the segmented and labeled data.

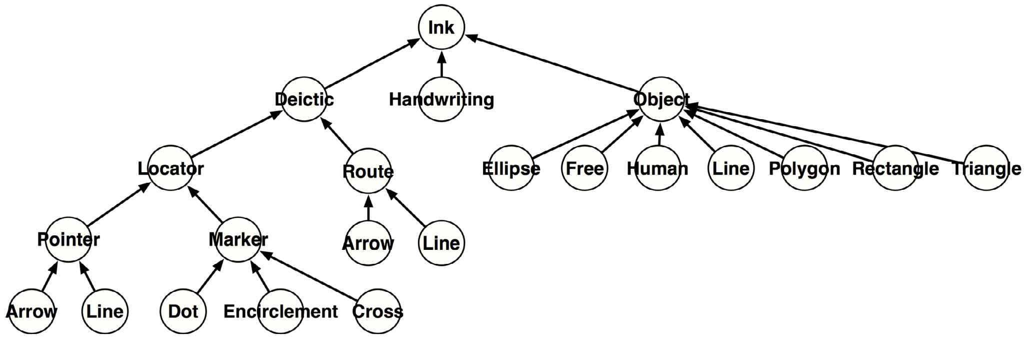

Figure 3. The hierarchical taxonomy of the acquired pen gestures, based on data analysis of the collected data. Pen input was distinguished in several deictic gestures, handwritten text, and various object classes.

In Figure 3, the resulting taxonomy of pen input categories is depicted. Classification results of these data using the systems described in Section 2 confirmed our assumption that the relevant modes to be distinguished should be adapted to data acquired in crisis management scenarios (Willems and Vuurpijl 2006). A hierarchical classification was made corresponding to the taxonomy yielded by the data analysis (Figure 3). The main categories of modes that should be distinguished were deictic gestures, handwritten-text, and (iconic) objects. About two thirds of the pen gestures belonged to the deictic gesture class. Another problem that became apparent is that a lot of pen gestures are ambiguous. It is difficult to distinguish between lines that are used as deictic gestures and lines that are used as objects. It should be noted that the newly acquired data contained many additional gesture shapes for which the original classification system (presented in (Willems et al., 2005b)) was not (and could not have been) adequately trained. To be able to make a proper performance comparison with that system, all such gestures were excluded from the new performance test data. The overall performance, using the original classification system, was therefore, only 84.8%. Compared to the performance on the original data set 95.6%, this was rather meager. Distinction between deictic gestures, handwritten text, and objects, achieved a performance of 90.7%.

The mode detection classifier for distinguishing deictic gestures, handwritten text, and iconic objects was only able to reach a performance of 57.6% for iconic objects. The number and variability of the different (iconic) object gestures cause this low accuracy. It appeared that people use many different iconic representations of for instance a fires (flames or even bonfires), cars, or victims. This is probably the area where it is best to restrict the gesture set people may use, e.g. using a limited number of iconic representations for indicating specific events or objects. Research by Fitriani (Fitriani, Datcu, and Rothkrantz, 2006), provides a set of iconic objects used in a visual language for crisis management. This set may provide a indication on which iconic object shapes are important to incorporate in the pen gesture vocabulary used for interactive maps in crisis management.

4. IMPROVING RECOGNITION: EXPERIMENT AND RESULTS

A performance of 90.7% is not acceptable for a recognition system that should robustly function during crisis management tasks. To improve the system, we decided to explore more and better distinguishing features. Furthermore, a Bayesian belief network (BBN) was implemented to combine the results of four different classifiers. Below, in Section 4.1, the data used for this experiment is described. Section 4.2 contains the architecture of the employed BBN. In Section 4.3 our new results are discussed.

4.1 The data set

For training and testing the BBN, we used the data described in the previous section. The feature set was increased from eight (Willems et al., 2005a) to 26 by adding new distinguishing features. These new features are the (1) ratio between the major axes, (2) the orientation of the major axis, (3) the length of the major axis, (4) the rectangularity (correspondence of the convex hull with the bounding box), (5) maximum curvature between trajectory segments, (6) sharp turn (<60°) count, (7) last sharp turn offset, (8) pen down count, (9) initial horizontal co-ordinate offset from centroid, (10) final horizontal co-ordinate offset from centroid, (11) average pen pressure, (12) average straight line length, (13) straight line ratio, (14) largest straight line ratio, (15) trajectory crossing count, (16) loop count, (17) perpendicularity, and (18) average curvature.

The data set was randomly divided in three subsets: (i) a development set containing 265 samples, (ii) a test set (1050 samples), and (iii) an evaluation set (1325 samples). The development set was used for testing the suitability of the different features. Furthermore, these data were used to train the individual classifiers. The test set was used to assess the recognition accuracy of the individual classifiers. The results of these classification tests were used to create the probability tables of the classifier nodes used in the BBN. The tables of the other nodes were determined from statistical analysis of the data in both the development and the test set.

Before the final evaluation phase, the classifiers were trained with both the second set and the development set, which probably increased the performance since more samples were available for training. The gestures were randomly assigned to each data set, except that the same distribution over the different gesture modes was enforced. Final evaluation was performed on the evaluation set of which 871 (65.7%) were deictic gestures, 290 (21.9%) handwritten text gestures, and 164 (12.4%) objects.

4.2 Classifier combination through Bayesian belief networks

Bayesian belief networks (BBNs) use prior and conditional probabilities to calculate the probability for a state in a variable depending on the available evidence (Jensen, 2001). This may be implemented as a hierarchically organized pattern recognition system (Heskes, 1998), where the results of different classifiers are used as evidence. A BBN is a directed acyclic graph with nodes that represent variables and arcs that represent statistical dependence relations between those nodes (Jensen, 2001). Bayesian statistics are used for calculating the probability for each state of each node. In the construction of the Bayesian network used for mode detection, causal relations and the resulting probability tables were determined between the nodes of the network (see Figure 4).

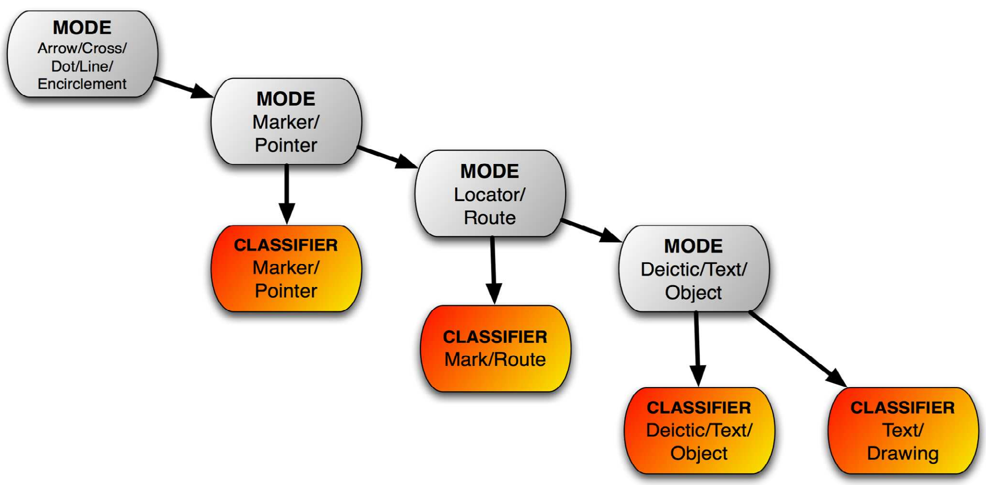

Figure 4. The Bayesian belief network used to combine the results of four different classifiers.

The arcs between the nodes define these causal relations so that for instance the variable CLASSIFIER Mark/Route depends on the MODE Locator/Route. The state of the classifier node is *caused,* therefore, by the state of the mode node. If evidence is found, probabilities for each state in the network need to be updated. For instance when a classifier returns a result, the state corresponding with that result is entered as evidence for that state in the corresponding CLASSIFIER node. In the CLASSIFIER node, the probability for that state will be set to 1.0, and for all other states in that node to 0.0. The changes in probabilities in one node will be propagated back to other nodes.

If the Mark/Route classifier classifies a pen gesture as “Route”, then this evidence will increase the probability for the state “route” in the mark/route MODE node. In this manner, evidence (results) from the different classifiers can be used to influence the probabilities in the MODE nodes. These probabilities of the MODE nodes specify the results required of the mode detection system. The probability tables that determine the causal relations between the nodes consist of a set of conditional probabilities. For instance, the Mark/Route CLASSIFIER node has the conditional probabilities: P(CLASS=Mark|MODE=mark), P(CLASS=Route|MODE=mark), P(CLASS=Mark|MODE=route), and P(CLASS=Route|MODE=route), where P(CLASS=Mark|MODE=mark) reads as the probability that the classifier result is ‘Mark’ *given* that the mode is ‘mark’. For each node, a probability table needs to be specified. In our network, probabilities were calculated, in the case of MODE nodes, from the statistical distribution of the modes in the development and training sets, or, in the case of the CLASSIFIER nodes, from the classifier results on the training set (being trained using the development set).

4.3 Results

To determine the result of mode detection, the probability for each state in the MODE nodes was calculated after evidence from the classifiers had been entered as evidence. If the deictic gesture state in the MODE Deictic/Text/Object node had the highest probability in the three possible states, the pen gesture was classified as a deictic gesture. By classifying all pen gestures in the test data set in this manner, a performance was obtained of 96.0% for classification between deictic gestures, handwritten text, and objects. Compared to the original performance of 90.7% on this data, this may be considered as a significant improvement.

	Deictic	Text	Objects	Evaluation Set Size
Deictic	99.0 [95.3]	0.3 [0.9]	0.7 [3.8]	871 (65.7%)
Text	2.4 [2.4]	97.2 [96.5]	0.3 [1.0]	290 (21.9%)
Objects	15.9 [29.7]	6.1 [12.8]	78.0 [57.6]	164 (12.4%)

Table 1. The confusion matrix for the deictic, handwritten text, and object modes. The type of the gesture is depicted horizontally, and the recognized class vertically. Numbers represent percentage correct classification. Numbers between square brackets are the results from the former mode detection system.

When considering the confusion table (Table 1), it can be observed that deictic gestures and handwritten text were recognized quite well (99.0% and 97.2% respectively), but that the recognition of objects (78.0%) poses mode detection problems. Objects were often confused for deictic gestures (15.9%). This strengthens the idea that it may be advantageous to restrict the number of accepted object gestures. However, even though the classification of object gestures reached only 78.0%, it is a big improvement over the 57.6% recognition rate of the original classification tests.

The classification between locating and routing gestures increased slightly from 96.5% to 96.8%. The recognition rate of locating gestures decreased from 99.6% to 98.9%, but the recognition rate of routing gestures increased from 58.2% to 64.8%. One possibility of increasing the performance is to use context information such as may be available from geographic information systems (GIS). Using such information, it can be assessed, for example, whether a pen gesture follows a road-pattern. It is likely that the trajectory of routing gestures will follow the road pattern much more closely compared to locator gestures. Preliminary results employing a road context detector as an extra node in the Bayesian network are promising: The recognition of routing gestures increased to 68.5%.

The other two MODE nodes in the Bayesian network distinguish between marker (encirclements, crosses, etc.) and pointer gestures (arrows and lines that point to an object on the map or photograph to mark it), and between different types of deictic marker gestures. The recognition of these modes is less important except that they influence the probabilities of the states in the other modes. In the final implementation, it will not be important whether an object is marked by a gesture that points to the object or by a gesture that encircles that object.

Overall the error rate has been more than halved, which seems to prove the worth of a Bayesian belief network in the implementation of a pattern recognition system for pen gesture interaction.

5. DISCUSSION

In this paper we have demonstrated a best-practices approach to the design of interactive maps that can be used in crisis management applications. We have argued that for this domain, interactive maps provide a promising tool for improving the efficiency and robustness of communicating spatial information. The data collection process reported in this paper resulted in natural pen input because users were not constrained in the gesture repertoires they were allowed to use. Please note that these data were acquired in laboratory conditions and it should be researched whether pen input in more stressful situations has similar characteristics.

In order to deal with the huge variability in pen input, comprising handwriting, drawing, and gestures, we have presented three different mode detection systems with increasing complexity, both in the system architectures and in the pen gestures that are successfully recognized. A decrease in average mode detection error rates was reported from 9.3% to 4%. While this still requires improvement before interactive maps can be employed in a crisis management system, the increased performance can be considered as a promising baseline for future research.

The results presented in this paper clearly indicate how the gesture repertoire can best be constrained. Since the mode detection system performs well on distinguishing deictic gestures and handwritten text, for these classes, users can use their preferred gesture repertoire and handwriting. Iconic objects are not very well distinguished. Restricting the user to a small set of iconic objects that can easily be distinguished and are still intuitive representations to the user, will undoubtedly increase performance. The selection of the appropriate classes and shapes of iconic object gestures should depend on the gestures preferred by the users in the field of crisis management and iconic representations as used in schematic renderings of situational maps. New experimental studies following the approach presented in this paper are currently being undertaken to obtain such a set of constrained iconic gesture shapes and corresponding recognition technology.

Furthermore, as part of the ICIS/CHIM project, we will evaluate the usability of interactive maps in an interactive pen recognition system. The participants will be able to enter annotations that are interpreted by the system. The results of the interpretations will be rendered on the map using symbolic gestures and icons, e.g., using the IconMap system developed by a partner in our project (Fitrianie et al., 2006). The eventual pen input recognition system will implement (i) mode detection, (ii) recognition of map objects that are marked or pointed at, (iii) the recognition of the routes specified by the user, (iv) the recognition of the content of handwritten text, and (v) the recognition of iconic objects. The architecture and improved performance of the system presented here provide promising guidelines for the successful accomplishment of this project, which may eventually lead to the successful employment of interactive maps in the world of crisis management.

ACKNOWLEDGEMENTS

This work is supported by the Dutch ICIS project (grant BSIK03024). Data was kindly provided by the COMIC project, the UNIPEN database, and Fonseca and Jorge.

REFERENCES

1. Bishop, C.M., Svensèn, M., and Hinton, G.E. (2004) Distinguishing Text from Graphics in On-line Handwritten Ink. In: *Proceedings of the ninth International Workshop on Frontiers in Handwriting Recognition, IWFHR 2004*, Tokyo, Japan, pp. 142-147.
2. Boves, L., Neumann, A., Vuurpijl, L., ten Bosch, L., Rossignol, S., Engel, R., and Pfleger, N. (2003) Multimodal Interaction in Architectural Design Applications. In: *8th ERCIM Workshop on ``User Interfaces for All''*, Vienna, Austria.
3. Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. (1997) Quickset: Multimodal interaction for Distributed Applications. *Proc. Of the Fifth ACM International Multimedia Conference*, ACM Press New York, 31-40.
4. Cohen, P. R., Johnston, M., McGee, D., Smith, I., Oviatt, S., Pittman, J., Chen, L., and Clow, J. (1997) QuickSet: Multimodal interaction for simulation set-up and control., in *Proceedings of the Fifth Applied Natural Language Processing meeting*, Association for Computational Linguistics, Washington.
5. Dix, A., Finlay, J., Abowd, G., and Beale, R. (2004) *Human-Computer Interaction*, 3rd edition, Prentice Hall.
6. Fitrianie S., Datcu D. and Rothkrantz L.J.M. (2006) Constructing Knowledge of the World in Crisis Situations using Visual Language, *IEEE SMC'06*, 121-126.
7. Fonseca, M.J. and Jorge, J.A. (2001) *Experimental evaluation of an on-line scribble recognizer*, Pattern Recognition Letters, **22** (12), 1311-1319.
8. Goldberg, D. and Richardson C. (1993) Touch-Typing With a Stylus. In: *Proceedings of the CHI'93 Conference on Human Factors in Computing Systems*, pp 80-87, Addison Wesley.
9. Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., and Janet, S. (1994) UNIPEN project of on-line data exchange and benchmarks. In: *Proc. of the 12th Int. Conf. on Pattern Recognition*, pp. 29—33.

10. Heskes, T. (1998) Solving a huge number of similar tasks: a combination of multi-task learning and a hierarchical {B}ayesian approach, Proceedings of the 15th International Conference on Machine Learning, Morgan Kaufmann, San Francisco, CA, pp. 233-241.
11. Holland J. (1975) *Adaptation in Natural and Artificial Systems*, The University of Michigan..
12. Jain, A.K., Namboodiri, A.M., and Subrahmonia, J. (2001) Structure in on-line documents, In: *Proceedings of the 6th International Conference on Document Analysis and Recognition (ICDAR'01).*
13. Jensen, F.V. (2001) *Bayesian Networks and Decision Trees*, Springer Verlag, New York Inc.
14. Kettebekov, S., Krahnstoever, N., Leas, M., Polat, E., Raju, H., Shapira, E., and Sharma, R. (2000) *iMap: Crisis management using a multimodal interface.* Presented at: *ARL Federated Laboratory 4th Annual Symposium, College Park*, MD.
15. McTear, M.F. (2002), Spoken Dialog Technology: Enabling the Conversational User Interface, *ACM Computing Surveys*, Vol. 34, No. 1, pp. 91--169.
16. Montello, D. R. (2001) Spatial Cognition. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*, pp. 14771-14775. Oxford: Pergamon Press.
17. den Os, E. and Boves, L. (2004) Natural multimodal interaction for design applications. In: *Proceedings, eChallenges e2004*, Vienna, 27-29.
18. Oviatt, S.L. (2003) Multimodal interfaces. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. Jacko and A. Sears, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, chap.14, 286-304.
19. Rossignol, S., Willems, D.J.M., Neumann, A., and Vuurpijl, L. (2004) Mode detection and incremental recognition. In: *Proceedings of the ninth International Workshop on Frontiers in Handwriting Recognition, IWFHR 2004*, Tokyo, Japan, pp. 697-602.
20. Schomaker, L., Vuurpijl, L., and de Leau, E. (1999) New use for the pen: outline-based image queries, In: *Fifth International Conference on Document Analysis and Recognition (ICDAR'99)*, IEEE, pp. 293-296.
21. Sharma, R., Yeasin, M., Krahnstoever, N., Rauschert, I., Cai, G., Brewer, I., Maceachren, A.M., and Sengupta, K. (2003) Speech-Gesture Driven Multimodal Interfaces for Crisis Management. In: *Proceedings of the IEEE* 92 (9), pp 1327-1354.
22. Vuurpijl, L., ten Bosch, L., Rossignol, S., Neumann, A., Pfleger, N., and Engel, R. (2004) Evaluation of multimodal dialog systems, In: *Proceedings, LREC 2004 Workshop on Multimodal Corpora*.
23. Wahlster, W. (ed.) (2006) *SmartKom, Foundations of Multimodal Dialogue Systems*, Cognitive Technologies, Springer.
24. Willems, D.J.M., Rossignol, S., and Vuurpijl, L. (2005a) Features for mode detection in natural online pen input, In: *Advances in Graphonomics Proceedings of the International Graphonomics Society (IGS2005)*, June 26-29, Salerno, Italy, pp. 113-11.
25. Willems, D.J.M., Rossignol, S., and Vuurpijl, L. (2005b) Mode detection in online pen drawing and handwriting recognition, In: *Proceedings of the Eight international conference on document analysis and recognition (ICDAR'05)*, IEEE, Aug 29-Sep 1 2005, Seoul, Korea, pp. 31-36.
26. Willems D.J.M and Vuurpijl L.G. (2006) Pen Gestures in Online Map and Photograph Annotation Tasks, *Proc. of the Tenth International Workshop on Handwriting Recognition IWFHR 2006*, France, 397-402.