A Bayesian network approach to mode detection for interactive maps

Don J.M. Willems and Louis. G. Vuurpijl
Nijmegen Institute for Cognition and Information
Radboud University Nijmegen
P.O.Box 9104, 6500 HE Nijmegen, The Netherlands

Abstract

This paper describes a mode detection system for online pen input that employs a Bayesian network to combine classification results and context information. Previous monolithic classifiers were not able to provide sufficient performance to be used in the domain of crisis management, where robust interaction is extremely important. To enhance mode detection for the intended target domain of crisis management, domain specific pen gesture data was used to train the four different classifiers and to calculate the conditional probabilities used in the Bayesian network. Mode detection, which is used to distinguish between different types of pen input such as deictic gestures, handwritten text, and iconic objects, clearly profited from this new approach. The error rate dropped from 9.3% for a monolithic system to 4.0% for the new mode detection system.

1. Introduction

Interactive maps are especially suited for conveying spatial information between human and computer. Using a digital pen on an electronic tablet, one can mark or add objects on a map or on visualized photographic content. Computer systems that provide for this type of interaction with the user, need to be able to recognize the pen gestures that are produced. Unfortunately, when users are unconstrained in the types of gestures that they can use, recognition becomes problematic. In a recognition system that needs to recognize not only different gestures, but also the type (or mode) of the pen gestures, mode detection [3, 7] is employed before specific classifiers are used for actual recognition [12]. Mode detection should for instance be able to determine whether a user is producing deictic gestures (e.g. to mark an object on a map or to specify a route), handwritten text, or iconic object drawings (people, cars, etc.).

The domain in which our pen gesture recognition system will be employed is crisis management. In earlier work [13], we concluded that two thirds of the pen gestures used in crisis management situations are deictic gestures. When a deictic gesture is detected by the mode detection system, it is not required to recognise the exact pen gesture (for instance an encirclement) but the context related to the pen gesture (the object that is encircled). Based on these observations, it is apparent that mode detection is of prime importance for successful pen interaction in the crisis management domain.

Previous work on mode detection includes research by Jain [7] and Bishop [3] who distinguished between handwritten text and lines, and handwritten text and drawings, respectively. Using geometrical features with kNN and MLP classifiers we were able to obtain a performance of 98.7% for mode detection between handwritten text, arrows, lines, and geometric objects [11]. Using a hierarchical mode detection system and expanding the recognized classes to include four geometric shapes, an overall performance of 95.6% was reached [12]. The problem with these mode detection systems was that they were trained and tested using data that was not obtained from the crisis management domain. The human factors experiment [13] provided us with data specific to our target domain.

The mode detection systems we presented in [12] was tested with the new domain specific data set that resulted from the human factors experiment. As expected the performance was not good. The mode detection performance dropped from 95.6% (using the original data set) to 84.8%. The reason for this decline in performance was twofold: (i) some of the modes (for instance geometric objects) that could be recognized by the previous systems were not relevant for the newly acquired data and (ii) many pen gestures are ambiguous in that they can be assigned to different modes (see Figure 1). Using a monolithic classifier using the same geometric features, we were able to reach a performance of 90.7% for the recognition of deictic gestures, handwritten text, and objects [13].

In the domain of crisis management, where lives may depend on correct and efficient communications, an error rate of 9.3% is not acceptable. To increase the performance we decided to combine the results of the different classifiers and information from map, photographic, and task context. The fact that a combination of classifiers often achieves better classification results than any of the individual classifiers by themselves has been well established [9]. One approach that can be used to combine different information sources is the use of Bayesian belief networks (BBNs) [2, 8]. BBNs have been successfully implemented for pattern recognition tasks [5, 6], and because of their probabilistic nature, BBNs lend themselves very well for Decision Support Systems [10]. We will use Bayesian networks to combine different classifiers and context information.

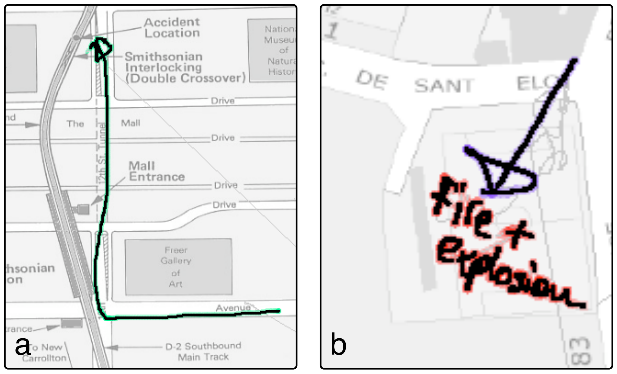

Figure 1. Two examples where context information can be of use in gesture recognition. (a) An arrow used to specify a route. (b) An arrow used to mark a location. The arrow in (a) follows the street pattern, the arrow in (b) does not.

Context information has been used in sketch recognition before [1, 14]. Our goal is to use context information to enable correct recognition of ambiguous gestures. If the gesture is an encirclement, the pen gesture will probably encircle an object on a map or photograph. If that object is known from context information, mode detection may be enhanced by using that information.

In this paper we describe our new mode detection system that employs a Bayesian network to combine information from different classifiers and from context. Furthermore, we will present the evaluation of this mode detection system, including the results.

2. Context information

2.1. Spatial context

Geographical information systems (GIS) can provide spatial information on a multitude of object types such as houses, public buildings, industrial objects, and infrastructure. This information can be used to identify relations between generated pen gestures and objects on a map. In real-time photograph or video annotation, spatial context is much more difficult to extract, since that information can only be obtained from visual recognition systems and is not as readily available as geographical information.

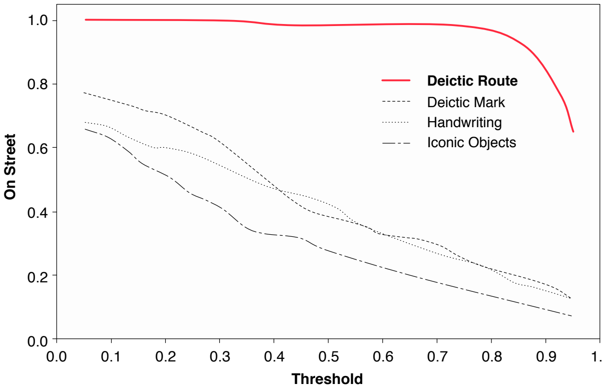

Figure 2. The proportion of route and other gestures evaluated as ’on street’ for different values of the threshold θ_{street} (see text below). Ideally, a large proportion of deictic routes and a small proportion of other gestures should be evaluated as ’on street’.

Object context is determined in two ways: (i) comparing the centroid of the pen gesture with map or photograph objects, and (ii) by calculating the proportion of an object that is encircled by the convex hull of the pen gesture. If the centroid is located within an object or if the proportion of encirclement is higher than the threshold θ_{enc}, then the gesture is evaluated as ’on object’.

Street context is determined by calculating the proportion of samples (the points on screen that make up the pen trajectory) that are situated within a street object. If this proportion is higher than a predetermined threshold (θ_{street}), the pen gesture is evaluated as ’on street’. In Figure 2, the relation between the threshold and street context is depicted. We expect that most routing gestures will be evaluated as ’on street’. Unfortunately other gestures (cars or victims) may also be drawn on street objects, and routing gestures may be (partly) drawn outside street objects [13].

As you can see in Figure 2, there is a big difference between deictic routing gestures and other types of gestures. Routes can be easily distinguished with street context information, but many objects and non-routing deictic gestures will be confused with routing gestures.

2.2. Task context

Other important context information is task related. In [13], we have found that for map and photograph annotation tasks, different distributions between pen gesture modes were found. This difference is even more pronounced when considering the type of task a participant had to perform. Examples of task types are marking tasks, where a user has to mark an existing object on a map or photograph, and routing tasks where the user has to specify a route.

The data we will use to evaluate our new mode detection system was taken from the experiment we conducted in 2005 [13]. In this data set, the task-type is predetermined for each task the participants had to perform. In real-life situations, the task-type is not readily available, and should be determined from the context of the dialog the user has with the dialog action manager, which is the part of the computer-human interaction system that is responsible for steering the interaction with the human user [4].

3. A Bayesian network for combining multiple information resources

Bayesian belief networks (BBNs) are directed acyclic graphs containing nodes and directed arcs between those nodes [2, 8]. Each node represents a variable (for instance the mode of a pen gesture) that can have different states (for instance: deictic, handwriting, or object). The BBN uses prior and conditional probabilities to calculate the probability of a state given the available evidence, using Bayesian statistics. The different prior and conditional probabilities are gathered in a probability table for each node.

Four different types of node are used in the BBN of our mode detection system (see Figure 3):

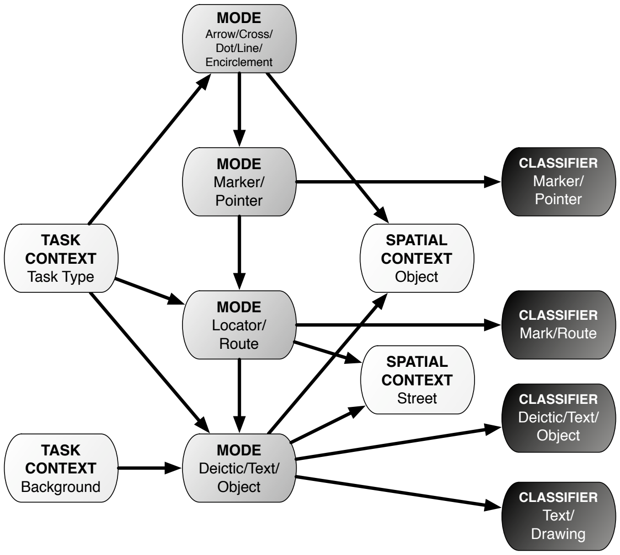

Figure 3. The Bayesian belief network (BBN) used to combine context information with classification results from different classifiers. The four MODE nodes provide us with the desired output.

1. Task context nodes (task type and background) are used to provide task context information to the BBN. Context may be predefined (as in an experimental setup) or available from dialog context. Background context specifies whether the user is drawing on a map or photograph. The probability tables are calculated from statistical analysis of domain specific data [13].

2. The four mode nodes provide the mode detection system with its results. The result for each mode is specified by the state of the node with the highest probability. The mode nodes represent the pen trajectory. Because the pen trajectory depends on the intent of the user, and therefore, on the task, mode nodes depend on the task context nodes. As with the task context nodes, the probability tables are calculated from statistical analysis of the data.

3. Spatial context nodes are used to add evidence from spatial context. If, for instance, the proportion of samples in a pen trajectory drawn on a street is higher than the threshold, the state 'on street' is entered as evidence. Because the evaluation of spatial context uses the pen trajectory as input and can therefore be said to be caused by the pen trajectory, spatial context nodes depend on mode nodes. The probability tables for the two association nodes are determined by evaluating spatial context for each gesture in the data set used during development (see Section 4.1).

4. Classifier nodes are used to enter evidence depending on the results of the classifiers. If the Text/Drawing classifier returns 'Text' as result, the state representing text in the classifier node will be entered as evidence. Like spatial context nodes, classifier nodes depend on mode nodes, because classification results can be said to be caused by the pen trajectory. The probability table for each classifier node is determined by testing the classifier on the data set and is equal to the confusion matrix of the classifier. The four classifiers use kNN (with k=3, determined by trial an error), using all geometric features presented in [11, 12].

For evaluation of the new mode detection system two BBNs were created. This enabled us to distinguish the contribution of (i) the BBN and (ii) the context information. The first BBN was created without context nodes and the second with context nodes. Apart from the existence of these context nodes, both networks were the same.

4. Experiments and results

4.1. Data

Before evaluating the mode detection system, the probability values in the probability tables for each of the nodes in the BBN needed to be determined, and the classifiers needed to be trained. For these tasks and final evaluation, three different data sets were randomly taken from the full data set which resulted from the 2005 experimen [13], a development set, a testing set, and an evaluation set.

We determined the thresholds θ_{street} and θ_{enc}, used in context evaluation, by analysing the correspondence of pen gestures and map context in both the development and test sets. The probability tables for the non-classifier nodes used in the BNN were calculated using the data in the development and testing sets. The classifiers were trained with the development set and tested with the testing set. The resulting confusion matrices are equal to the probability tables for the classifier nodes. The evaluation set was only used for the final evaluation phase. The evaluation set was the largest and contained 1325 gestures with 871 (65.7%) deictic gestures, 290 (21.9%) handwritten text gestures, and 164 (12.4%) object gestures. The test set contained 1050 gestures, and the development set 265 gestures.

4.2. Results

Two mode detection systems with BBNs were tested, the first without, and the second with context nodes. First, we discuss the results of mode detection without context nodes.

Previously, mode detection for distinguishing between deictic gestures, handwritten text, and objects with the original mode detection system reached a performance of only 90.7% [13]. Using the new mode detection system without context nodes, the performance was enhanced to 96.0%.

When considering the confusion table (Table 1), one can see that deictic gestures and handwritten text are recognized quite well (99.0% and 97.2% respectively), but that mode detection on object gestures are problematic (only 78.0% recognition). Nevertheless, this is a big improvement over the 57.6% recognition rate that was reached with the original mode detection systems [13].

The accuracy of mode detection between marking gestures and routing gestures increased to 96.8%. The recognition of marking gestures decreased somewhat, but the recognition of routes increased to 64.8% (see Table 1).

The second BBN that was tested was the network with context nodes. Surprisingly, the mode detection system performance with context information was lower than without context information. 95.5% was recognized correctly.

As we can see in the confusion table between deictic gestures, text, and objects (Table 2), the confusion of deictic gestures to the other two modes stayed the same. The recognition rate of handwritten text increased because less text was confused with deictic gestures. The recognition performance of objects on the other hand decreased by 5.5% to 72.5%. Analysis of the object gestures that were not recognized correctly, shows that object gestures like rectangles and free-form objects were more often recognized as deictic gestures in the mode detection system with context nodes. Some of these gestures that were recognized correctly when not using context, were now misclassified because they were drawn on the street pattern (cars or victims), and other gestures were found to encircle a map object (house icons on the location of an address on a map).

Table 1. The confusion matrices for the mode detection system without context nodes. The test class is presented horizontally and the recognized class vertically.

Type	Deictic	Text	Object	N
Deictic	99.0%	0.3%	0.7%	871
Text	2.4%	97.2%	0.3%	290
Object	15.9%	6.1 %	78.0%	164
Correct	96.0%			

Type	Marking	Routing	Other	N
Marking	98.9%	0.2%	0.9%	817
Routing	31.5%	64.8%	3.7%	54
Correct	96.8%			

Table 2. The confusion matrices for the mode detection system with context nodes. The test class is presented horizontally and the recognized class vertically.

Type	Deictic	Text	Object	N
Deictic	99.0%	0.3%	0.7%	871
Text	1.7%	97.9%	0.3%	290
Object	21.3%	6.3 %	72.5%	164
Correct	95.5%			

Type	Marking	Routing	Other	N
Marking	97.9%	1.5%	0.6%	817
Routing	22.2%	70.4%	7.4%	54
Correct	96.2%			

Apparently, spatial context information enhances recognition when the pen gesture is related to spatial context, but the recognition rate decreases for gestures that are not related to context. This can also be seen in the recognition of marking gestures and routing gestures (Table 2). Street context enhances the recognition of routing gestures, while lowering the recognition rate of marking gestures.

5. Discussion

A new mode detection system was presented in this paper that employs a Bayesian network combining the results of multiple classifiers and different types of context. Bayesian networks have clearly proven their worth by more than halving the error rate on the data set acquired from the target domain of crisis management. With this mode detection system we were able to obtain a performance of 96.0%.

The use of context information did not enhance the mode detection system as we had expected. While pen gestures that are related to context have indeed been recognized with higher accuracy, the confusion within the pen gestures that are not related to context has resulted in a worse performance then when using the same system without context information. Nevertheless, after analyzing the misclassifications that are due to context information, we are convinced that context information may enhance mode detection when context detection is improved. This may be achieved, for instance, by using an extra classifier that is used to distinguish long pen gestures that are parallel to the streets on the map (routing gestures) from small compact objects drawn on a street (such as cars). When that classifier is combined with context detection for streets, we expect that street context may indeed enhance mode detection.

Because of the prevalence of deictic gestures over handwritten text and object gestures, mode detection is very important for the interpretation of pen gestures in interactive maps. Future development will use the current mode detection system to create a fully featured pen interaction system that can be used in crisis management situations. This pen interaction system will employ an improved context detection algorithm to recognize the objects that are related to the produced gestures. It will also use existing handwriting recognizers to facilitate the recognition of handwriting.

Our results indicate that the recognition of iconic objects is still problematic. We are currently pursuing two directions to improve accuracy. First, we propose to enhance object recognition by improving the feature set and using other classification methods in conjunction with the present classifiers. The Bayesian networks discussed in this paper provide a suitable framework for adding such new technologies. Second, we are considering a suitable constrained vocabulary of iconic object gestures, adapted to the preferences of the users and optimized on distinctiveness between the gestures. The choice for iconic object shapes is explored in cooperation with our project partners, who are experts in the domain of crisis management.

Acknowledgments

This work is supported by the Dutch ICIS project (grant BSIK03024).

References

[1] C. Alvarado, M. Oltmans, and R. Davis. A framework for multi-domain sketch recognition. In *AAAI Spring Symposium on Sketch Understanding*. AAAI, 2002.

[2] C. M. Bishop. *Pattern recognition and Machine Learning*. Information Science and Statistics. Springer Verlag New York, LLC, 2006.

[3] C. M. Bishop, M. Svensén, and G. E. Hinton. Distinguishing text from graphics in on-line handwritten ink. In *Proceedings of the ninth International Workshop on Frontiers in Handwriting Recognition*, pages 142–147, 2004.

[4] T. H. Bui, M. Poel, A. Nijholt, and J. Zwiers. A tractable (DDN-POMDP) approach to affective dialogue modeling for general probabilistic frame-based dialogue systems. In *Proceedings of the 5th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, India, January 2007. IJCAI 2007.

[5] S.-J. Cho and J. H. Kim. Bayesian network modeling of strokes and their relationships for on-line handwriting recognition. *Pattern Recognition*, 37(2):253–264, 2004.

[6] T. Heskes. Solving a huge number of similar tasks: a combination of multi-task learning and a hierarchical Bayesian approach. In *Proceedings of the 15th International Conference on Machine Learning*, pages 233–241, San Francisco, CA, 1998. Morgan Kaufmann.

[7] A. Jain, A. Namboodiri, and J. Subrahmonia. Structure in on-line documents. In *Proceedings of the 6th International Conference on Document Analysis and Recognition (ICDAR'01)*, Seattle, Washington, 2001.

[8] F. Jensen. *Bayesian Networks and Decision Trees*. Springer Verlag New York Inc., Secaucus, NJ, USA., 2001.

[9] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 3 1998.

[10] E. J. M. Lauría and P. J. Duchessi. A bayesian belief network for IT implementation decision support. *Decision Support Systems*, 42(3):1573–1588, 2006.

[11] D. Willems, S. Rossignol, and L. Vuurpijl. Features for mode detection in natural online pen input. In *Proceedings of the 12th Biennial Conference of the International Graphonomics Society*, pages 113–117, 2005.

[12] D. Willems, S. Rossignol, and L. Vuurpijl. Mode detection in online pen drawing and handwriting recognition. In *Proceedings of the Eight international conference on document analysis and recognition*, pages 31–35, 2005.

[13] D. Willems and L. Vuurpijl. Pen gestures in online map and photograph annotation tasks. In *Proceedings of the tenth International Workshop on Frontiers in Handwriting Recognition*, pages 397–402, 2006.

[14] E. Yi-Luen Do. Design sketches and sketch design tools. *Knowledge-Based Systems*, 18(8):383–405, 2005.