Collaborative Emergency Management with Multimodal GIS

Sven Fuhrmann[1&2], Isaac Brewer[1&2], Ingmar Rauschert[1&3], Alan M. MacEachren[1&2], Guoray Cai[1&4], Rajeev Sharma[1&3], Hongmei Wang[1&4], Levent Bolelli[1&4] & Benyah Shaparenko[1&3]

Abstract

GIS provide critical support for decision makers during emergencies. Emergency management is a collaborative effort requiring coordination among federal, state, and local specialists in planning, logistics, operations, etc. Thus, GIS should provide access to geospatial information quickly and support collaborative work between domain experts. However, conventional GIS are not suited for multi-user access. In order to overcome these limitations, a new interface paradigm for GIS is required. The research reported in this paper attempts to overcome unimodal mouse and keyboard controlled GIS by designing a multimodal, multi-user GIS interface. Our implementation of a multimodal GIS uses ArcIMS as the mapping engine serving an interactive, large-screen display.

Introduction

Rapid access to geospatial information is crucial in emergency situations when decision makers need to work collaboratively, using GIS for emergency management. Current GIS provide mostly unimodal interaction tools, i.e. classical input devices like keyboard and mouse that support only a single user. The result of this situation is that GIS functionalities are not accessible to all participants at the same time and do not allow users to express meaning, extract important information or develop planning scenarios effectively (Brewer, et al. 2000).

Collaborative emergency task situations require a new paradigm for human-computer interaction. This paradigm replaces unimodal (and WIMP-based) and unidirectional interaction with multimodal dialog-based interaction and supports collaborative (rather than individual) work. Multimodal interactions consist usually of two or more combined user input modes, such as speech, gesture, gaze or body movements (Oviatt 2002). In order to meet these goals, this research represents part of an ongoing effort to develop principles for implementation and to guide in assessing natural, multimodal, multi-user dialog-enabled interfaces to GIS that make use of large screen displays. The Dialog-Assisted Visual

Ingmar: "There is hurricane approaching, and it is likely to affect this area (*gesture circling region of interest*) in the southeast."
DAVE_G: "Is this the region you mean, Ingmar?" (*outline appears and blinks twice*)
Ingmar: "No, I'm not interested in that area to the west of the Interstate." (*gesture indicating a general direction to the north and west*)
DAVE_G: (the *outline adjusts, shifting the center of focus*)
Ingmar: "That's better."
DAVE_G: (*the map zooms in, and an inset appears in the corner*)
Ingmar: "Let's look at the population distribution here (*gesture circling region of interest*) in the southeast."
(…)

Figure 1: A scenario about DAVE_G usage during an emergency event

Environment for Geoinformation (DAVE_G) has been developed as a prototype testbed environment that uses integrated speech-gesture interaction modalities, domain knowledge and task context for a dialog management, and that supports collaborative group work with GIS in emergency management situations. Figure 1 includes a scenario in which DAVE_G is used during emergency management.

Related Work

User studies and evaluations (Lamel, et al. 1998, Schapira and Sharma 2001) of recently developed multimodal human-computer interfaces emphasize the use of natural and conversational human-computer interfaces, such as speech and free-hand gestures. However, in spite of calls for research to make GIS easier to use (Krahnstoever, et al. 2002), limited work has focused on multimodal GIS interaction (Schlaisich and Egenhofer 2001).

Lokuge and Ishizaki (1995) developed a speech-based interactive visualization system (GeoSpace) for accessing information spaces. The initial system featured spoken queries on geo-spatial databases including an approach to dialog management. Since GeoSpace relied only on speech as an interaction modality, it remained unimodal. Oviatt (1999), Oviatt and Cohen (2000) concluded that speech recognition, as a single interface modality is error-prone and often not efficient, especially if used in context with interactive maps. Speech provides an effective way of expressing actions, pronouns and abstract relations, but it often fails when spatial relations or locations have to be specified (Oviatt 1999). In order to explain spatial relations or locations to someone, gestures are very often used. In combination, speech and gesture recognition might provide a multimodal interface that is more suitable for recognizing spatial expressions than unimodal speech-based system approaches (Oviatt 1996). In order to overcome the speech-based unimodality, Cohen, et al. (1997) developed a pen and speech-based multimodal interface to generate interactive maps. Since speech often provides a very vague concept of spatial relations, Cohen, et al. (1997) allowed users to query spatial data directly on the interactive map, applying an electronic pen. The usability assessment of this multimodal interaction research showed that participants were able to express spatial relationships more efficiently.

Krahnstoever, et al. (2002) developed a multimodal human-computer interaction framework for large screen displays. The framework used natural speech and hand gestures to enable interactions with data displayed on a large screen. Krahnstoever, et al. (2002) designed two prototypes. One prototype was an “intelligent” map of the Penn State University campus, in which visitors could post queries and ask for assistance on locations and directions. The second application was a crisis management simulation that allowed users to direct police cars, fire trucks and other emergency vehicles using gesture and speech (Kettebekov, et al. 2000). In spite of these first attempts, the integration of multimodal interfaces with GIS and collaborative use of GIS in emergency situations remain elusive.

DAVE_G – the multimodal, multiuser GIS

Designing a multimodal collaborative GIS application for emergency management is a “chicken and egg” problem, where neither such a system nor relevant and exhaustive user studies are available. Sharma, et al. (1999) addressed this problem by analyzing weather broadcasts in which individuals talk and gesture at large screen maps as a starting point to build a computational model for speech-gesture integration. This model was implemented as iMap and that prototype (which used a simple, non-interactive map display) served as a starting point for our work to extend the ideas to GIS. A current challenge is in transitioning from the presentational style of weather broadcast to interactive situations for single and multi-user tasks.

In order to design a user and task centered application, cognitive systems engineering (CSE) methods were applied to the process of task analysis and scenario development (Brewer

2002). The goal of CSE was to collect, represent and analyse information about domain experts (emergency management planners in Florida, South Carolina, Washington D.C. and Pennsylvania) and their task and problem solving knowledge (Rasmussen, et al. 1994). The generated information on users and tasks was than used for prototype design.

Functionalities in DAVE_G

At the current stage, DAVE_G (Dialog-Assisted Visual Environment for Geoinformation) provides a variety of data querying, navigating and drawing functions targeted to emergency management activities. Gestures and natural speech allow users to express GIS-based requests that are closely tied to emergency management tasks. Two types of requests can be distinguished; one that relies on spatial references that need to be specified by gesturing (e.g. pointing and outlining), and a second that allows requests to be expressed solely by speech. The interaction with DAVE_G is not limited to making requests; users can provide input: free hand drawing is supported and an annotation functionality is under development. The current prototype of DAVE_G features the following requests

- Data query (show/hide features, select/highlight features)
- Buffering (create/show buffers),
- Map navigation (pan left/right/up/down, center at, zoom in/out/area/full extend) and
- Drawing (circle, line, free hand).

Some of the above functions can be initiated by speech only, e.g. visualizing themes (highways, counties, etc.), creating buffers around selected features and navigating the map (zoom, pan, etc.). Gestures and speech are integrated when a spatial reference is needed for accomplishing a task, e.g. selecting a number of distributed objects within a region. For fulfilling a successful task, e.g. visualizing a theme, the user does not need to provide all required information at once. DAVE_G incrementally collects information about the intended task by supporting a human-computer-based dialog. As an example, the user might want to display population data, DAVE_G might have different kinds of population information available (e.g., census tracts, counties), thus DAVE_G would ask which of the available population data the user wants to have displayed. The dialog manager of DAVE_G supports indirect and implicit references to themes and features on the map that were mentioned or used in previous requests. Parameters that are not obtained through the dialog can be provided by expert-defined default values or by intelligent guessing within the context domain, avoiding time-consuming dialogs.

The architecture of DAVE_G

The architecture of DAVE_G is based upon three modules: Human Interaction Handling, Human Collaboration and Dialog Management, and Information Handling (see Figure 2). Communication among these modules is established as request-response message pairs using a predefined XML encoded protocol. The Human Interaction Handling Module has two components: a human reception control and a display control. Since DAVE_G is a multi-user environment, each instance of the human reception control component captures a single user's speech and gesture input and generates descriptions of recognized words, phrases, sentences, and gestures to be used for processing. Gesture is captured using a non-calibrated active camera (for each user), which tracks the user's head and hand skin color and motion. Speech is captured by microphones and processed by speech recognition software. A context-free grammar defines syntactically correct phrases and sentences that are to be recognized. Speech and gesture are combined in the reception control unit, which forwards recognized gesture descriptions and verbal expressions to the display control and the Human Collaboration and Dialog Management Module.

Screen rendering is handled in the display control. It receives the responses of the system, e.g. GIS-based maps and textual messages and coordinates the system part of the computer-human interaction. The display control also controls the direct feedback in response to users actions, e.g. cursor movements.

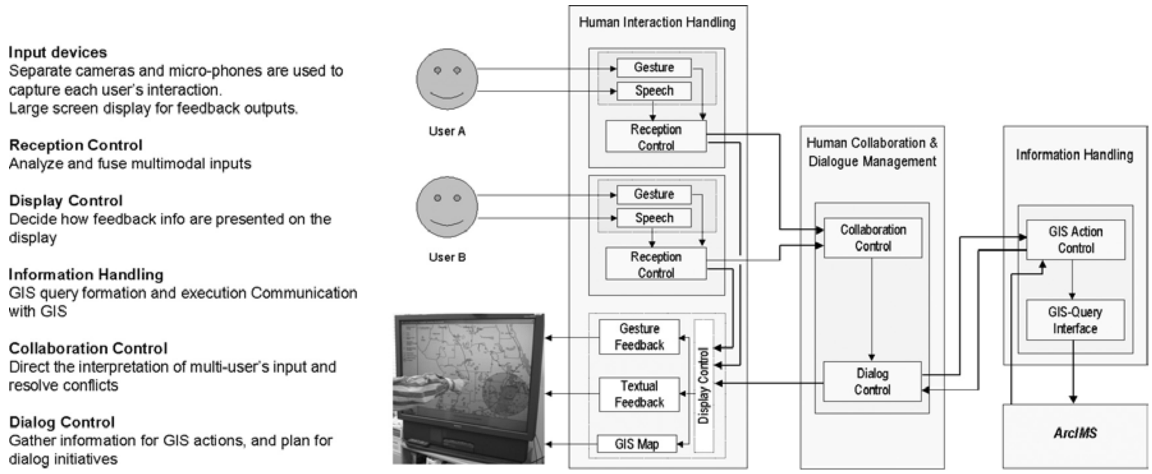

Figure 2: The architecture of DAVE_G

The Human Collaboration and Dialog Management Module receives the recognized gesture descriptions and verbal utterances from the reception control components. It coordinates the execution of these commands through two components: the collaboration manager and the dialog manager. The collaboration manager is involved in the conflict management if two or more persons are interacting with DAVE_G. In this prototype it is assumed that the persons involved in the task work collaboratively and their interactions are interpreted through one semantic frame representing the collective intentions in the formation and execution of one GIS command. The dialog manager checks the received information on consistency and establishes a dialog with the users if the information is not consistent. When sufficient information is collected on a task, the information is passed to the Information Handling Module for the formation and execution of a corresponding GIS query. Information returned from the Information Handling Module can be maps and textual messages (successful query) or error messages. This information is passed to the display control.

The Information Handling Module processes the received commands by forming GIS queries and sending the query to ArcIMS through the GIS query interface. The GIS action control takes care of all GIS related queries and maintains information regarding the current status of the GIS. The GIS query interface is a library that uses ArcXML to expose the GIS functionality through the ArcIMS Server. Successful GIS queries return a map generated by ArcIMS. The map display of the current prototype includes standard information, such as legends and theme information.

Conclusions and continuing research

Current GIS provide a range of geospatial tools and analysis methods but the rather complex functionalities are usually available to a single, simultaneous user. In collaborative emergency management, current GIS restrict the range of users and narrow chances for efficient task solving.

Therefore, a multimodal, multi-user GIS interface was developed, which supports collaborative emergency management on large screen displays. This research focuses on research problems in three different fields: 1) The development of collaborative and natural gesture speech recognition, 2) The design of a human centered interface by applying cognitive system engineering methods to domain analysis in the emergency management domain, and 3) The development of an intelligent dialog system that can respond and interact with a group of users.

The DAVE_G prototype will undergo a range of usability assessments, since our goal is to create a user-centered, usable, and robust system. Methods to be applied include

- expert guideline-based evaluations (user interface experts identify potential usability problems using heuristics),
- formative usability assessments (domain users apply task-based scenarios, discuss and test the prototype) and
- summative usability assessments (the prototype is tested against other comparable tools) (Nielsen 1993, Rubin 1994, Torres 2002).

The qualitative and quantitative results of the usability assessments will support the refinement of DAVE_G and provide a good basis for future developments of user-centered collaborative, multi-user GIS for emergency management. Our working group will continue developing robust stationary and mobile tools in order support group work with geospatial data during highly volatile emergency management situations.

Acknowledgements

This work is based upon work supported by the National Science Foundation under Grants No. 0113030, IIS-97-33644, and IIS-0081935. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF). ESRI's software contribution to the project is gratefully acknowledged.

References

Brewer, I. (2002). Cognitive Systems Engineering and GIScience: Lessons learned from a work domain analysis. Proceedings of GIScience 2002, Boulder, CO.

Brewer, I., A. M. MacEachren, H. Abdo, J. Gundrum and G. Otto (2000). Collaborative geographic visualization: Enabling shared understanding of environmental processes. IEEE Information Visualization Symposium, Salt Lake City, IEEE.

Cohen, P., M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen and J. Clow (1997). QuickSet: Multimodal interaction for distributed applications. Proceedings of the Fifth International Multimedia Conference (Multimedia '97), Seattle, ACM Press.

Kettebekov, S., N. Krahnstoever, M. Leas, E. Polat, H. Raju, E. Schapira and R. Sharma (2000). i2Map: Crisis management using a multimodal interface. ARL Federate Laboratory 4th Annual Symposium, College Park, MD.

Krahnstoever, N., S. Kettebekov, M. Yeasin and R. Sharma (2002). A real-time framework for natural multimodal interaction with large screen displays. Fourth IEEE International Conference on Multimodal Interfaces (ICMI 2002), Pittsburgh.

Lamel, L., S. Bennacef, J. L. Gauvain, H. Dartiguest and J. N. Temem (1998). User Evaluation of the MASK Kiosk. ICSLP 1998, Sydney.

Lokuge, I. and S. Ishizaki (1995). Geospace: An interactive visualization system for exploring complex information spaces. Conference Proceedings CHI 1998, New York.

Nielsen, J. (1993). Usability Engineering. Boston, AP Professional.

Oviatt, S. (1996). Multimodal interfaces for dynamic interactive maps. Proceedings of the Conference on Human Factors in Computing Systems (CHI'96), New York, ACM Press.

Oviatt, S. (1999). "Ten myths of multimodal interaction." Communications of the ACM 42(11): 74-81.

Oviatt, S. (2002). Multimodal interfaces. The Human-Computer Interaction Handbook. J. A. Jacko and A. Sears. Mahwah, Lawrence Erlbaum.

Oviatt, S. and P. Cohen (2000). "Perceptual user interfaces: multimodal interfaces that process what comes naturally." Communications of the ACM 43(3): 45-53.

Rasmussen, J., A. M. Pejtersen and L. P. Goodstein (1994). Cognitive engineering: Concepts and applications. New York, J. Wiley & Sons.

Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York, J. Wiley & Sons, Inc.
Schapira, E. and R. Sharma (2001). Experimental evaluation of vision and speech-based multimodal interfaces. Workshop on Perceptive User Interfaces, November 15-16, Lake Buena Vista.
Schlaisich, I. and M. J. Egenhofer (2001). Multimodal spatial querying: What people sketch and talk about. 1st International Conference on Universal Access in Human-Computer Interaction, New Orleans. http://www.spatial.maine.edu/~max/UAHCI.pdf.
Sharma, R., I. Poddar, E. Ozyildiz, S. Kettebekov, H. Kim and T. S. Huang (1999). Toward interpretation of natural speech/gesture: Spatial planning on a virtual map. Proceedings of ARL Advanced Displays Annual Symposium, Adelphi, MD.
Torres, R. J. (2002). Practitioner's handbook for user interface design and development. Upper Saddle River, Prentice Hall PTR.

Author Information

Sven Fuhrmann[1&2], Isaac Brewer[1&2], Ingmar Rauschert[1&3], Alan M. MacEachren[1&2], Guoray Cai[1&4], Rajeev Sharma[1&3], Hongmei Wang[1&4], Levent Bolelli[1&4] & Benyah Shaparenko[1&3]

[1] GeoVISTA Center, Penn State University, University Park, PA 16802, {fuhrmann, maceachren, isaacbrewer, gxc26, rxs51, rauschei, bolelli, benyah, hongmeiw}@psu.edu

[2] Department of Geography, 302 Walker, Penn State University, University Park, PA 16802, {fuhrmann, maceachren, isaacbrewer}@psu.edu

[3] Department of Computer Science and Engineering, Penn State University, University Park, PA 16802, {rxs51, rauschei, benyah}@psu.edu

[4] School of Information Science and Technology, Penn State University, University Park, PA 16802, {gxc26, bolelli, hongmeiw}@psu.edu