CAMAS: A Citizen Awareness System for Crisis Mitigation

S. Mehrotra, C. Butts, D. Kalashnikov, N. Venkatasubramanian, K. Altintas, Haimin Lee,
J. Wickramasuriya, R. Hariharan, Y.Ma
University of California, Irvine[1]

R. Eguchi, C. Huyck
ImageCat, Inc.[2]

ABSTRACT
This demo paper provides a brief description of the intuition and design philosophy of CAMAS, one of the main testbeds being developed in the context of the RESCUE (Responding to the Unexpected) project [9]. The goal of our work is to enhance the mitigation capabilities of first responders in the event of a crisis by dramatically transforming their ability to collect, store, analyze, interpret, share and disseminate data. CAMAS specifically is a system designed to allow a variety of users, including the average citizen to report incidents and potentially hazardous situations, analyze and evaluate these reports and notify appropriate personnel for further action. The multidisciplinary approach incorporates a variety of information technologies: networks; distributed systems; databases; image and video processing; GIS; and machine learning, together with subjective information obtained through social science. Besides providing an overview of the CAMAS architecture, we describe the demonstration platform and specific experiments that will illustrate issues of event extraction, ranking, access control and visualization.

1. INTRODUCTION
Responding to natural or man-made disasters in a timely and effective manner can reduce deaths and injuries, contain or prevent secondary disasters, and reduce the resulting economic losses and social disruption. During a crisis, responding organizations confront grave uncertainties in making critical decisions. They need to gather situational information (e.g., state of the civil, transportation and information infrastructures), together with information about available resources (e.g., medical facilities, rescue and law enforcement units). Clearly, there is a strong correlation between the accuracy, timeliness, and reliability of the information available to the decision-makers, and the quality of their decisions. The ‘Responding to Crises and Unexpected Events’ (RESCUE) Project [7] was recently conceived with the objective of radically transforming the ability of organizations to gather, manage, analyze and disseminate information when responding to man-made and natural catastrophes. Dramatic improvements in the speed and accuracy at which information about the crisis flows through the disaster response networks has the potential to revolutionize crisis response, saving human lives and property.

In the RESCUE Project, our focus is to radically transform the speed and accuracy with which information flows through disaster response networks, networks that connect multitudes of response organizations as well as the general public. We are working to develop information technology solutions that dynamically capture and store crisis-relevant data as it is generated, analyze this data in real-time, interpret it, and disseminate the resulting information to decision makers in the forms most appropriate for their various tasks. Challenges in realizing such IT solutions arise due to the scale and complexity of the problem domain, the diversity of data and data sources, the state of the communication and information infrastructures through which the information flows, and the diversity and dynamic nature of the responding organizations.

2. THE CAMAS TESTBED
The Crisis Assessment, Mitigation and Analysis System (CAMAS) testbed is one of several testbeds that are being developed in the context of the RESCUE project. In CAMAS, our focus is on leveraging the existing fixed infrastructure for mitigating incident-level crisis in instrumented spaces such as airports, malls, nuclear facilities, research facilities, etc. Such instrumented spaces may have an existing surveillance and network infrastructure that may be partially/completely operational and can be used by first-responders for crisis mitigation. To simulate such an environment, we will instrument a significant portion of the UCI campus to create smart corridors and open spaces. This provides us with an ideal experimental testbed for many of the information technology solutions (e.g., collecting multimodal data in dynamic settings, capturing live events and testing multimodal event extraction algorithms, studying collaboration among user groups in dynamic settings in the context of specific events, privacy issues in information gathering, real-time voice analysis, models of reliability of human input, etc.) we are developing as part of the RESCUE project. CAMAS also incorporates passive information collection from human sensors, in the form of an integrated problem reporting system which translates raw text messages (and, ultimately, speech) into information regarding problem events on campus. Research currently underway includes an automated data analysis system to convert such event reports into probabilistic information regarding unknown problem states. This system factors informant and problem attributes into the analysis, thereby allowing for automatic down weighing of suspect information and robustness to denial of service attacks. The information resulting from the

[1] ITR-Rescue, 5251 California Avenue, Suite 210, Irvine CA 92612-2815, Phone: 949-824-1147 http://www.itr-rescue.org
[2] Union Bank Building, 400 Oceangate, Suite 1050, Long Beach CA 90802 Phone: 562-628-1675, http://imagecatinc.com

analysis can then be utilized by the automatic alert dissemination system to notify on-campus responders of potential problems, while minimizing false positives. We now discuss in more detail key components of the CAMAS infrastructure that deal with multimodal data collection, event extraction and analysis.

3. MULTIMODAL EVENT EXTRACTION AND ANALYSIS

One of the integral components of our research is an end-to-end data analysis system that captures and analyzes multimodal data (e.g., voice and video input from in-field officers and cameras, GPS, sensor data), extracts meaningful events/information from transcriptions, populates key databases, and uses this information in real-time as input into a damage and impact assessment system. Such an analysis system is at the very core of the human-as-sensor concept, whereby input from the response and rescue teams based on their personal observations, local "eye-witnesses," and personal communications from others on the ground is used for situation assessment. Such input from first responders or informants with the additional benefit of human interpretation, has the potential to provide more robust information gathering capability under adverse circumstances, resulting in more effective and timely response. Figure 1 shows various components of the data analysis system we are building. Below we focus on only two of the components - event extraction and analysis. Besides the extraction and analysis framework, the system consists of various other components whose functionality is described below.

- *Data Mediation Infrastructure* – a system that enables dynamic access to heterogeneous databases in the context of diverse tasks and varying access control policies to data. Such a mediation infrastructure can provide access to the knowledge/database useful in information extraction as discussed below.
- *Event Data Management System* -- an XML based model for representing, querying, and analyzing spatio-temporal events.
- *Event Visualization System* -- a GIS based framework that supports spatio-temporal event monitoring in the context of specific tasks such as activity planning, activity monitoring, etc using multi projector displays.
- *Event Mining system* -- that looks for unusual patterns and correlations among events.

Event Extraction from Multimodal Data Streams: One of the first challenges in designing the data analysis system we envision is to accurately extract meaningful events from the multimodal input. Since transcribed voice input is among the most valuable sources of information, we focus on event extraction from text. Information extraction from unstructured text - even in restricted domains - is notoriously difficult [10][5]. In crisis-specific event extraction, these difficulties are compounded due to the inherent noise in voice transcriptions. The approach we are exploring is to first develop a taxonomy of event types with each of which an event model is associated. The disaster response experts may initially specify such event models. Classification approaches based on the event models are used to determine the type of events. Over time, the system will come across event types that cannot be classified automatically. In such cases, the event taxonomy as well as the corresponding models will evolve over time possibly with the help of analysts/users.

Once an event is appropriately classified, the set of properties associated with the event are extracted. (Almost) all event types share some properties (such as spatial and temporal location of the event). Other properties might be event specific. For example, with a report about a fire, one may associate properties such as the extent of the structural damage, interrupted service (e.g. network, electrical) etc. Data extraction accuracy can be significantly improved by exploiting multiple information sources. For example, textual data can be supplemented with location information (from GPS), video data, redundant observations across other text streams, as well as knowledge existing in current knowledge/data bases (available via the mediation component); the context provided by this additional information can be used to restrict the space of possible events and their properties improving the accuracy of the extractor. While context and multimodal data, has the potential of dramatically improving text extraction accuracy, presence of textual data can also benefit multimedia analysis. In particular, video scene analysis can benefit from other information sources such as maps, aerial images, localization information, video from fixed and calibrated cameras and analyzed text. Such input can be systematically incorporated for landmark-based registration of the video acquired by low cost small field of view cameras carried by first responders. Such an approach will improve the interpretability of the video as well as help in improving its perceptual quality [11][6].

Event Analysis: Once basic events have been extracted, they are subjected to variety of analysis tasks (see figure above). For example, a same event may be reported multiple times (e.g., due to multiple observers making and reporting the same event). Alternatively, two events might be linked or correlated in some ways. Similarity matching and data mining techniques properly enhanced to handle spatio-temporal events can be used for this purpose. Furthermore, multiple tightly coupled events might correspond to a larger fused event [8]. These analysis tasks are vital in proper interpretation, ranking, prioritization, and triaging of information to decision makers. We emphasize below two such that highlights the synergistic fusion of our research in social science and information technology – the essence of our multidisciplinary approach.

The first such task is ascertaining the *reliability of human reports*. While human input can play a vital role in crisis response, it must be used with care since informant accounts are known to be highly distorted in a number of ways [2] due to a multitude of factors. Exacerbating circumstances include intentional abuse (whereby a certain category of people may try to negatively affect the system on purpose), confusion and errors in recall, linguistic ambiguity, biases due to prior expectancy, etc. Due to the flow of information through social networks, individual errors in perception may be compounded by factors such as rumors, exaggerations, and differential information availability [3]. Informant reliability can be gauged based on knowledge about the informant, such as a group he/she belongs to, the history and prior reliability reports of similar/related events, the informant's reporting on events known

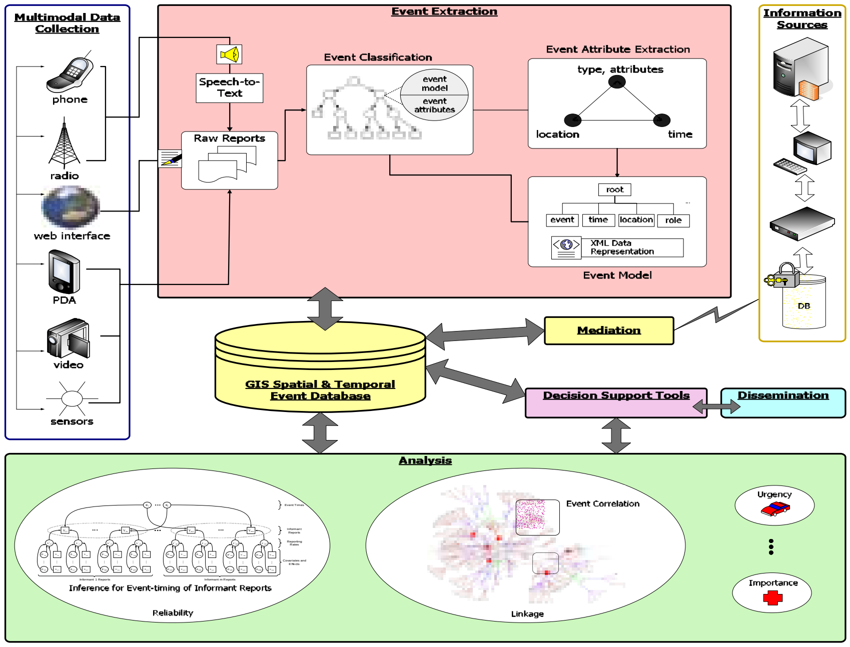

Figure 1: CAMAS Testbed Framework

to be exaggerated or caused by rumor, etc. Although sophisticated methods exist for integrating multiple, erroneous informant reports [1], much of the work has not incorporated models of social context in which information is generated. To cope with these conditions, information collection and analysis procedures must be developed which can not only model cognitive and perceptual biases, but also social structural effects.

Another related task is determining i*mportance and urgency* of an event. Importance and urgency are two tightly coupled factors that play a crucial role in information triaging/filtering. Events could be important but not urgent, or vice versa. Urgency and importance depend upon factors such as the type of event, its attributes, potential impact, the vulnerability of the affected population, etc. These factors ultimately combine to determine utility/loss functions for intertemporal choices regarding crisis interventions.

Within the RESCUE project, we are currently developing an inferential modeling framework for analysis of personal accounts in order to simultaneously infer the true state of a focal event; the social network of information flow among informants; and determinants of informant accuracy, urgency and importance. Some preliminary work in this context has recently been done in the context of network inference [4] and is being built upon here. In conjunction with these inferential models, we are in the process of devising active sampling methods to be deployed by crisis response organizations (e.g., via field agents) for the rapid collection of data for subsequent analysis. By deploying the inferential models and sampling strategies in a variety of event contexts, we can obtain background information regarding informant error processes and initial predictors of informant network structure. Such analyses will help in designing information technology solutions for information triage to responding organizations, and will ultimately be incorporated into decision support systems for organizations involved in crisis response.

4. CAMAS DEMONSTRATION PLATFORM AND EXPERIMENTS

The prototype CAMAS application that will be demonstrated will be a problem reporting facility for the UCI campus. The choice of this specific initial domain was decided upon by scale its similarity to the expected CAMAS problem domain and availability of geospatial data for accurate mapping of reported problems and easy access to buildings and facilities. We were also able to obtain significant domain expertise from maintenance, computer support and campus networking personnel. Users, acting in the context of a role are allowed to report using multiple modalities (e.g. telephone call, email or via a web interface) a variety of problems ranging from functioning of public facilities (e.g., damaged phones, broken windows, electrical problems, parking problems) to computer support issues (e.g. inability to access wireless networks) as well as potentially serious issues that could lead to fires, gas leaks etc. The reports are made in natural language text which is then parsed, analyzed, classified and ranked. We briefly describe the specific user group and databases in our example demonstration first and describe some specifics of components mentioned in the overall CAMAS architecture. The querying infrastructure is integrated with an interactive GIS tool that allows for easy visualization of specific events on a campus map.

We broadly consider two user groups in the initial prototype, a reporter group and a manager group. In the reporter group, a user may report the observed problem (e.g. leaking roof in CS building 2nd floor) as an anonymous observer or within a specified role such as graduate student, department staff or faculty. Members of the manager group are allowed access to reported events (raw or processed) based on their role, that correspond their particular responsibilities, e.g. a network manager will see only network related events. The events will be filtered and ranked using appropriate techniques to allow for prioritized processing of urgent events.

CAMAS consists of a event database where events are defined as spatiotemporal occurrences represented using 3 components –

location, time and event type/attributes. Efficient event extraction implies determining the above attributes using NLP or IR methods. The CAMAS event extraction component extracts three factors: exact location from the location tree, event occurrence time, and the problem type from the problem type tree. A reporter's input in natural language, either as a text sentence or speech data is often ambiguous in terms of event location or the specifics of the problem. In our prototype, we apply a simple but powerful technique often used in information retrieval literature called TF/IDF to locate appropriate nodes in the problem-type and location trees.

Event Filtering & Ranking: One common filtering criteria uses problem type and role information to retrieve appropriate query results. Analysts frequently deal with large volumes of streaming data from a variety of information sources. The data corresponding to a given time window is normally batched and periodically processed by the analyst. Given the size of data and rate at which it arrives, it is usually impossible to manually process every record or case. Instead, automated filtering (classification) mechanisms are employed to identify information relevant to the analyst's task. We have developed a flexible framework as well as algorithms that facilitate this filtering process by: (1) Allowing analysts to drive the process by expressing their domain knowledge (ie filtering constraints) to the system in the form of similarity queries that enable data to be ranked based on relevance to the target class,(2) Grouping the ranked data in pages (data groups) of user specified size to enable analysts to determine relevance at data group level,(3) Providing a refinement mechanism (consisting of algorithms to iteratively modify the query) that allows the evolution of the analysts' filtering constraints over time as well as feedback-based modification of the constraints when they fail to achieve desired precision.

In the crisis environment, ranking for determining importance of problems is a very complex and important issues that involves integrating user reliability and event urgency parameters. Once reports (of emergency events) are made, the responder faces the task of assessing what the real situation is and send out the help deemed necessary. Often a crucial interpretation is to decide whether a change in the status-quo (from "normal" to "emergency") has indeed occurred, which would justify sending agents to the field. This problem relates to the one referred to in the literature as the change-point problem. Our approach to this problem provides an assessment of how much we believe a change of that nature has really happened, given the information - even if scarce - on the history of the event, its (reported) location, the reporter's location, and the population in the area affected by it. Other factors that determine ranking include problem history, projected damages, resource constraints etc.

The CAMAS prototype currently supports triaging of urgent events via a notification-based event dissemination module. This module is capable of alerting appropriate personnel using a variety of modalities such as email, telephone calls using standard text-to-speech translation systems. The CAMAS prototype also includes GIS-based visualization tools (See Figure 2) that assist in decision support.

The demonstration will walk a viewer through the processing of information in the CAMAS system starting from a raw user report (e.g. speech, text) followed by the event extraction and analysis phases and ending with contextualized dissemination of alerts (if necessary). Through examples and experiments, the walkthrough will illustrate the strength of our extraction module, the capabilities of the event refinement interface and the adaptivity of our system to unexpected events.

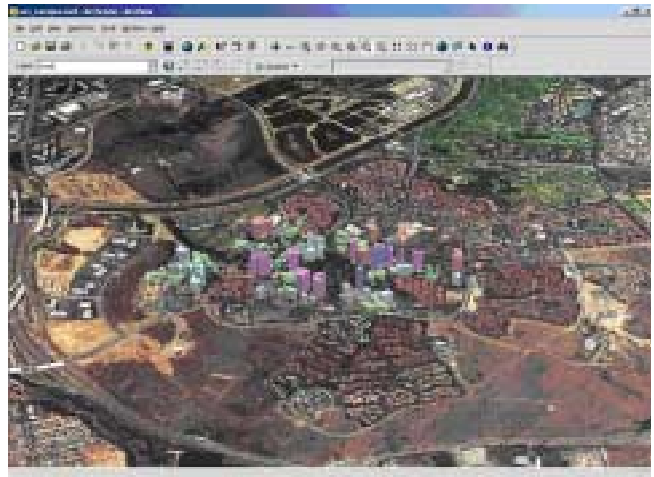

Figure 2: CAMAS Visualization Subsystem

ACKNOWLEDGEMENTS

We would like to acknowledge all the members of the RESCUE project research team (http://www.itr-rescue.org) and our government partners from the cities of Irvine, Los Angeles, and San Diego, the County of Los Angeles, and the State of California. Many of the ideas presented in this paper have evolved from discussions with them. We would also like to acknowledge the help of Lynn Harris in the preparation of this document. Finally, the research described in this paper is funded through NSF Grant UCI-0331707.

REFERENCES

[1] W. Batchelder, and Romney, A. K., "Test theory without an answer key", *Psychometrika*, 53(1), 1988.

[2] H. R. Bernard, Killworth, P., Kronenfeld, D., and Sailer, L. "The problem of informant accuracy: the validity of retrospective data". *Annual Review of Anthropology*, 13, 1984.

[3] C. T. Butts, "A Bayesian model of panic in belief", *Computational and Mathematical Organization Theory*, 4(4), 1998.

[4] C. T. Butts, "Network inference, error, and informant (in)accuracy: a Bayesian approach", *Social Networks, forthcoming*, 2003.

[5] Mary Elain Califf and Raymond Mooney, "Relational Learning of Pattern-Match Rules for Information Extraction", *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pp. 328-334, 1999

[6] K. Huang, and M. Trivedi, "Video Arrays for Real-Time Tracking of Persons, Head and Face in an Intelligent Room," Machine Vision and Applications, Omni-directional Vision, 2002.

[7] S. Mehrotra, C. Butts, D. Kalashnikov, N. Venkatasubramanian, R. Rao, G. Chockalingam, R. Eguchi, B.

Adams, Ch. Huyck, “Project RESCUE: Challenges in Responding to the Unexpected.”, Submitted for publication

[8] Andrew Moore, “Rule-Based Anomaly Pattern Detection for Detecting Disease Outbreaks”, *Proceedings of the 35th Symposium on the Interface of Computing and Statistics*, 2003.

[9] RESCUE (Information Technology Research in Responding to Crisis & Unexpected Events) *website*: http://www.itr-rescue.org

[10] M. Shinozuka, “On-line Damage Identification of Water Delivery Systems”, *13th ASCE Engineering Mechanics Conference*, 1999.

[11] M. M. Trivedi, I. Mikic, S. Bhonsle: "Active Camera Networks and Semantic Event Databases for Intelligent Environments", *Proc. IEEE Workshop on Human Modeling, Analysis and Synthesis*, 2000.