Enriching an Intelligent Resource Management System with Automatic Event Recognition

Daniel Stein[1], Barbara Krausz[1], Jobst Löffler[1],
Robin Marterer[2], Rolf Bardeli[1], Jochen Schwenninger[1], Bela Usabaev[1]
Fraunhofer IAIS, Schloss Birlinghoven, St. Augustin, Germany[1], University of Paderborn, Germany[2]
surname.name @iais.fraunhofer.de[1], marterer @cik.uni-paderborn.de[2]

ABSTRACT

Event recognition systems have high potential to support crisis management and emergency response. Given the vast amount of possible input channels, automatic processing of raw data is crucial. In this paper, we describe several components integrated in an overall intelligent resource management system, namely abnormal event detection in audio and video material, as well as automatic speech recognition within a public safety network. We elaborate on the challenges expected from real life data and the solutions that we applied. The overall system, based on Event-Driven Service-Oriented Architecture, has been implemented and partly integrated into the end users' infrastructures. The system is continuously running since almost two years, collecting data for research purposes.

KEYWORDS Event recognition system, abnormal event detection, automatic speech recognition, TETRA channel, Event-Driven Service-Oriented Architecture, IRM.

INTRODUCTION

Event recognition systems have high potential to support crisis management and emergency response. In critical situations like congestions and extremely high pedestrian densities, which may lead to deadly stampedes and terrible crowd disasters, the quality of decision making is highly dependent on situational awareness of the emergency management team and the availability of a common operational picture. The officer-in-charge is reliant on various information sources and on smoothly running communication chains. The main challenge to the management team is to bring all information together, to filter relevant information and to come up with and communicate a coherent operational picture that is based on the information available. Technical systems for intelligent information management, especially event-driven systems including solutions for complex event processing, are very much suited to support decision making in complex operations.

In this paper, we present results from the PRONTO project, which deals with innovative methodologies and system approaches in the area of event recognition for intelligent resource management (IRM). PRONTO applies research results to the demonstration cases Emergency Rescue Operations and Public Transport Management. One main outcome of the project is an event-driven system for intelligent resource management including a digital interactive map application and a management application for human resources, vehicle fleets and specialized equipment. While the complex events derived from fine granular triggers have been described in detail in (Pottebaum et al., 2011), the focus of this paper is the general architecture, the automatic event detection and event recognition for audio and video data which usually are produced during emergency or public transport operations. Especially, we will discuss our results for robust automatic speech recognition in public safety networks like TETRA (Terrestrial Trunked Radio) and for abnormal event detection in audio and video streams.

INTELLIGENT RESOURCE MANAGEMENT SYSTEM ARCHITECTURE

Our IRM system is based upon the principles of event-driven service-oriented architectures (ED-SOA) (Taylor et al., 2009) with Complex Event Processing (CEP) (Etzion, 2010). Note that in this paper, we only briefly depict the general system architecture and backend aspects and have a stronger focus on the automatic recognition. The frontend, realized as an application framework, which is capable of managing independent web-based apps (widgets) is not described in detail. The terminology used in this paper is aligned with the Event Processing Glossary of the Event Processing Technical Society (EPTS) (Luckham and Schulte, 2011).

The system is divided into subsystems, which contain several components. Due to its modular design, arbitrary components can be added, replaced or removed without much effort. All components are connected through the Message-Oriented Middleware (MOM) component. The MOM component is based on HornetQ, which is the JBoss application server implementation of the JMS standard. JMS messages are used for transporting event objects, which are the implementation of real-world events (Friberg et al., 2010). It allows the implementation of system components in arbitrary programming languages. Most parts of the system are implemented in Java.

Following the publish-subscribe pattern, components can act as event producers or event consumers. Event producer components create simple raw events and publish them for further processing via the MOM component to the system. Event producers components can be hardware sensor components (e.g. for GPS, acceleration or CAN-Bus data) or pure software components (e.g. those located in the application subsystem) which implement the event producer interface of the MOM component.

The application subsystem, based on Google Web Toolkit, contains those components, which are directly available through a user-friendly web-based interface. The most important applications are an IRM app, which shows a sophisticated logical view onto the system's status and the MAP app, which shows a geo-based view onto the system's status. All apps act as event consumers and event producers at the same time. The system is fully functional and partly integrated into the end-users' infrastructures at two different big European cities. The installation comprises very different kinds of event producers, which send various types of simple raw events to the system, e.g., GPS, vehicle status data, speed, acceleration, weather and user interactions. Different versions of the system have been running for almost two years processing and archiving event objects. Simple derived events and complex events are computed by the Simple Event Recognition (e.g. Audio and Video, see below) respectively Complex Event Recognition subsystem (Skarlatidis et al., 2011) and then visualized in near real-time by several web-based applications. For example, the location and details of a congestion or extremely high pedestrian density can be shown on the MAP app, which is available to all involved members of the emergency management team.

AUTOMATIC SPEECH RECOGNITION IN PUBLIC SAFETY NETWORKS

In scenarios of large-scale emergencies, the radio operator is in charge of transcribing information streams submitted over the public safety network and passing the written information to the operation control. Even nowadays, this is mostly done manually. While for safety reasons a human should always be in charge of this transcription procedure, reliable keyword extraction taken from a strong automatic speech recognition system can greatly speed-up and enhance this step. However, real-life data poses several challenges like a distorted voice signal, background noise and several different speakers. Moreover, the domain is out-of-scope for common language models, and the available data is scarce. We proceed to describe our efforts for an ASR system on a real-life fire-fighter exercise scenario.

Most European and Asian government networks and public safety networks employ the Terrestrial Trunked Radio (TETRA) standard published in the mid 90s by the European Telecommunications Standards Institute (ETSI). See (ETSI, 1998) for an overview of the codec. While TETRA has been optimized for speech intelligibility while maintaining a relatively low bit rate of 4.567 kBit/sec and a sample rate of 8 kHz, it introduces several distortions that pose challenges for common ASR systems. Moreover, scientific papers on the impact of natural language processing by automatic means are scarce. (Euler et al., 1994) is one of the few papers employing actual TETRA data in their recognition setup. On a small corpus of spoken German digits, they show that the TETRA codec performs poorly in comparison to the plain signal, to a 16 kBit/s Code-Excited Linear Predictive (CELP) and to a GSM codec.

We recorded the TETRA radio communication for ten fire fighter exercises. The material consists of status reports on the place of accident, conveyance of contaminant analysis, requests for backup, et cetera. Of the initial set of 1,769 transcribed sentences, 1,272 (71.9%) sentences are unique. The data has been randomly split into 769 sentences for the development set and 1,000 sentences for the test set. The audio samples have been recorded under real-life conditions. Slip of the tongue happens occasionally, hesitations occur frequently. Some parts are recorded indoors, others on the street. Background noise occurs frequently; occasionally there are co-interference phenomena from different channels or mobile phones. Sirens from emergency vehicles are audible in several instances. Since a radio button on the handheld device has to be pressed before speech is recorded, the beginning and the end of an utterance is often truncated. Parts of the material, especially places and numbers, are spoken with a local dialect. The material consists of many fire-fighter terms. Longer words that are common are often abbreviated. Due to the two-way radio systems, voice procedure (e.g. "affirmative", "over and out") is used very frequently. The grammar is often quite basic, verbs are often in infinite form.

We use an ASR system as described in (Schneider et al., 2008). In order to connect the ASR system to the remaining parts of the event-recognition and IRM solution, the transcriptions are sent as messages to the MOM component. For the acoustic model, we use a state-of-the-art German broadcast system, and a large language model derived from online newspapers and RSS feeds. We also extended the in-domain written text collection by crawling firefighter websites for operational reports. In total, this resulted in 30,791 running sentences, containing 318,954 words.

There is an obvious mismatch between the clean acoustic material of broadcast news and the one taken from the fire fighter scenario, but with standard adaption and interpolation techniques, a Word Error Rate (WER) of already 51.8% can be achieved. Several domain-specific problems like dialectal variance and voice procedure grammar arise where we can employ knowledge of the material to extend the dictionary with multiple pronunciations and enhance the language model by using a rule-based approach. This approach gives a further increase to 47.9% WER. Manually checking the recognizer output, especially the standardized voice procedures that request attention of one party for the other, and the standardized calls for backup, are among the sentences that perform best in accuracy. More problematic sentences include those that are specific to the situation, e.g., the exact nature of the current emergency situation, and unforeseen issues like locked doors or leaking chemical barrels. From the vocabulary, we identified five classes of words which convey a certain kind of information: equipment, emergency, location, entity, and urgency.

ABNORMAL EVENT DETECTION IN AUDIO STREAMS

Processing of audio streams in public safety networks for logging and transcribing purposes is one important aspect of an IRM system, but continuous monitoring for abnormal events of pre-installed audio and video surveillance can also submit crucial information for the operation control. For example, in many train disasters like the terrible accident of Eschede, Germany 1998, or the train derailment in Cologne, Germany 2008, abnormal sounds had been noted by passengers and personnel, but no appropriate action was taken.

While public transport is full of noise which does not carry crucial information, other noise that deviates from the usual background may indicate damage on roads, tracks, or engines. Statistics on such abnormal audio events can point public transport managers to issues that need to be repaired or can be improved, ideally at such an early stage that accidents can be prevented long before they occur. For the PRONTO system, we collected data from various means of public transport used by one of the authors on his daily commute to and from work. On a route of about 75 km we collected synchronized audio recordings and GPS data. Audio recordings were conducted using a mobile recorder standing on the floor, close to the engine level of the train. By wrapping the recorder inside a backpack, it was ensured that utterances from fellow travelers are too muffled to be intelligible, in order to avoid privacy concerns. Each recording contains approx. 2 hours and includes bus, tram and train commutes. On the content level, the data contains a wide variety of acoustic events including engine and vibrational noise, (muffled) utterances from fellow travelers and public announcements of stops and connection options, as well as background noises like ambulance alarms and more.

Abnormal events are by their nature hard to define. In our approach, we use a history of audio data of predefined length to learn what the normal acoustic situation at the moment looks like and update it by a sliding window approach. Clustering the spectral vectors collected from the history, we build a spectral dictionary that is used to describe the current acoustic situation as well as possible. It is described by a linear combination of the dictionary vectors plus an error term such that the norm of the error term is minimized (Zdunek and Cichocki, 2008). The error terms are then used as an indicator of the presence of abnormal events. The general idea is that whenever the description of the current acoustic situation by representative vectors of the history leaves a large error term, we face an abnormal event. In practice, this is implemented by change point detection (De Oca et al., 2010) on the spectral flatness feature of the spectral vectors representing the error term.

To give a comprehensive visualization of our method, we combine the collected GPS data and the detected abnormal events on a map. We provide an interface, see Figure 2, which allows showing where abnormal events occur and to select and inspect them by playing back the respective part of the audio and showing a spectrogram of the event. The nature of the abnormal events can be visualized in the spectrogram by exploiting the fact that the error terms indicating the event are themselves spectral vectors. Preliminary results show that many of the events that would be expected to be reported by the system can be indeed indentified. This includes ambulance alarms and engine noises that occur only during acceleration. The next step will be to annotate several days of data by ground truth events and quantify these findings. It will be an interesting aspect to filter the system output by location, time, and audio similarity and thus find out which events are indeed seldom and which occur repeatedly and in specific places.

ABNORMAL EVENT DETECTION IN VIDEO SURVEILLANCE

Although video surveillance is an integral part of mass events, train stations, airports and other pedestrian facilities, security personnel are faced with the short attention span of human operators as well as large gathering areas with many thousands of visitors. A real-time surveillance system can greatly support the security personnel by analyzing multiple video streams automatically and raising an alert if necessary.

In this section, we describe the abnormal events automatically detected by our system when working on real-life video surveillance footage of the Loveparade 2010: In the crowd disaster in Duisburg, Germany, a 240-meter tunnel was the only entrance and exit point of the festival area. When the pedestrian density reached a critical level, a stampede occurred resulting in 21 casualties and over 500 injured visitors. Having access to 20 hours of video surveillance material, taken from seven different camera positions (with three static cameras), we evaluated if and where an automated system would have detected abnormal events, thus possibly preventing a stampede. The videos are recorded with a frame rate of 25 fps from different viewpoints and have a resolution of 1280 x 720 pixels.

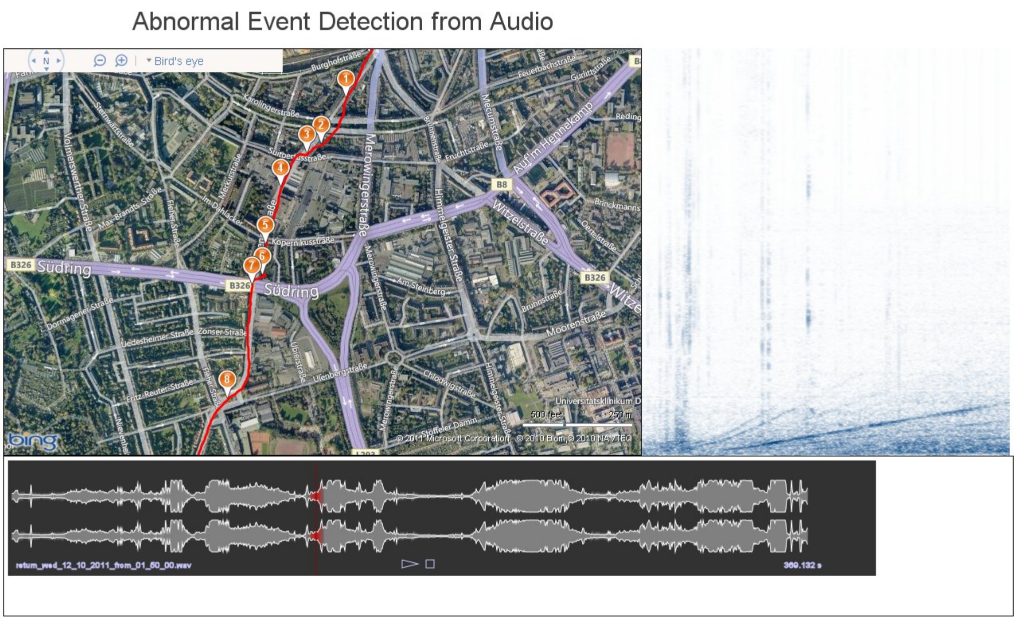

Abnormal Event Detection from Audio
Bird's eye
1
2
3
4
5
6
7
8
Südring
B326
B8
Merowingerstraße
bing
return_wed_12_10_2011_from_01_50_00.wav
369.132 s

Figure 1. Visualization of abnormal events detected. Numbers on the map represent events, and the current selected event is shown as a red line in the wave form below and in the spectrogram to the right. Map material (c) Microsoft Corp.

The video analysis method makes use of typical human motion behavior similar to those described by (Krausz and Bauckhage, 2011): Essentially, people do not move along a straight line, instead, it is a characteristic of human gait that they tend to swing laterally where the amplitude of lateral swaying depends linearly on their velocity. In congested areas, we observe very characteristic motion patterns: People go very slowly while the amplitude of lateral swaying increases resulting in oscillating motions. That is, in congested areas, people are stepping from one foot to the other in order to keep their balance. The proposed system computes dense optical flow fields where two subsequent video frames are compared in order to compute motion vectors for each pixel. Next, we design a feature for detecting changes in pedestrian flow. This feature is based on the observation of lateral oscillations given above. Essentially, the feature measures the amount of left-to-right motion vectors which is directly related to lateral oscillations of pedestrians and their velocity. Using a sequential change-point detection method with an adaptive threshold (De Oca et al. 2010), we learn typical feature values and detect deviations from that value indicating changes in pedestrian flow and congestions in particular. The system assesses the severity of the change in pedestrian flow by measuring the deviation from the learned values. However, the security personnel finally decides if the situations is critical and takes necessary actions.

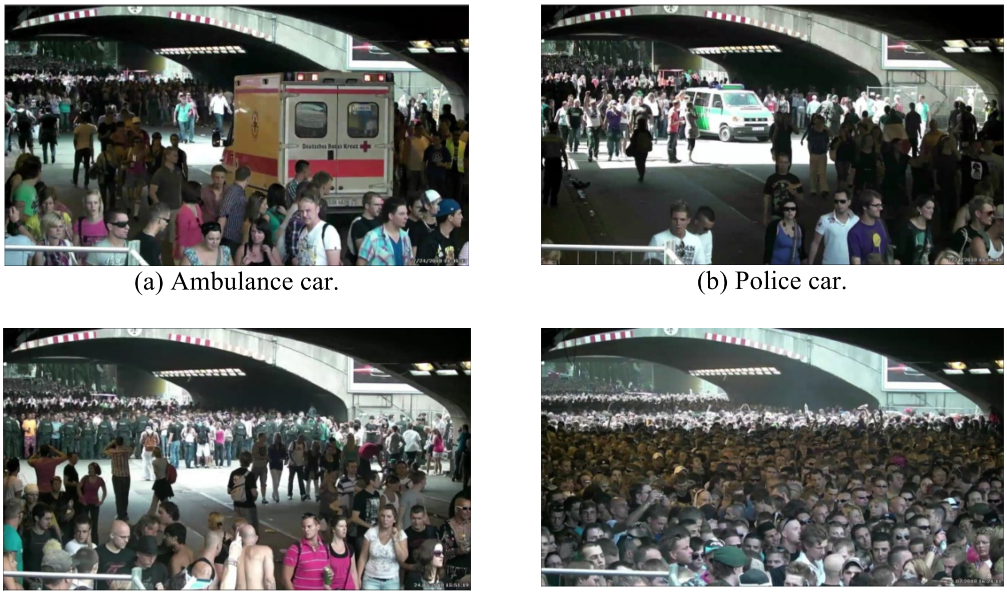
We tested our approach on 6 hours of video taken from the Loveparade dataset. The system detects abnormal events such as the crossing of police cars, a police cordon and, most importantly, a congestion. In this situation, our system would have detected a very critical situation and alarmed the security personnel to take necessary actions in order to prevent a deadly stampede. Figure 3 shows exemplary screenshots corresponding to raised alarms.

CONCLUSION

In this paper, we have shown that there is great potential for enrichment of an IRM system with automatic recognition techniques. As heterogeneous as the requirements within emergency rescue operations and city transport management are, as individual are the solutions for the various applications. We focused on three aspects of Complex Event Processing, namely the logging of communication via ASR, and the abnormal event detection in audio and in video material. For the ASR component within a TETRA channel, both the material and the domain of firefighter radio transmission are challenging, but standard methodology already leads to promising results. With suitably tailored domain-specific enhancements, the recognizer can already substantially support further annotation procedure. For the abnormal event detection in audio, first preliminary experiments have been presented where the computer is able to recognize previously unseen events and correlate them with GPS data. Last, for the abnormal event detection in video material, our system makes use of characteristic human motion patterns and computes optical flow fields in real-time enabling us to continuously monitor multiple video streams. Using a change-point detection algorithm, we can detect unusual events and critical crowd behavior.

ACKNOWLEDGMENTS

This work has been partly funded by the European Community's Seventh Framework Programme (FP7-ICT-2007-3. Cognitive Systems, Interaction, Robotics – STREP) under grant agreement n◦ 231738.



(a) Ambulance car.

(b) Police car.

(c) Police cordon.

(d) Congestion.

Figure 2. Examples for detected unusual events from camera 15 of the Loveparade dataset.

REFERENCES

1. ETSI. **Terrestrial Trunked Radio (TETRA); Speech Codec for Full rate Traffic Channel; Part 2: TETRA Codec**. Technical Report ETS 300 395-2, European Telecommunication Standard, February 1998.

2. O. Etzion, and P. Niblett. **Event Processing in Action**, Manning, Stanford, 2010.

3. S. Euler and J. Zinke. **The Influence of Speech Coding Algorithms on Automatic Speech Recognition**. In Proc. ICASSP, April 1994. Volume i, pages I/621–I/624.

4. T. Friberg, B. Birkhäuser, J. Pottebaum, and R. Koch. **Using Scenarios for the Identification of Real-World Events in an Event-Based System**, Proceedings of the 7th International ISCRAM Conference, 2010.

5. B. Krausz and C. Bauckhage. **Automatic Detection of Dangerous Motion Behavior in Human Crowds**, In Proc. Advanced Video and Signal-Based Surveillance (AVSS), Klagenfurt, Germany, August 2011.

6. D. Luckham, and W.R. Schulte. **Event Processing Technical Society: Event Processing Glossary – Version 2.0**, 2011.

7. V.M. De Oca, D.R. Jeske, Q. Zhang, C. Rendon, and M. Marvasti. **A Cusum Change-point Detection Algorithm for Non-stationary Sequences with Application to Data Network Surveillance.** Journal of Systems and Software, Volume 83, Issue 7, July 2010, Pages 1288-129.

8. J. Pottebaum, A. Artikis, R. Marterer, G. Paliouras, and R. Koch. **Event Definition for the Application of Event Processing to Intelligent Resource Management.** In Proc. ISCRAM, Lisbon, Portugal, 2011.

9. A. Preti, B. Ravera, F. Capman, and J.-F. Bonastre. **An Application Constrained Front End for Speaker Verification**. In Proc. EUSIPCO, Lausanne, Switzerland, August 2008.

10. H. Taylor, A. Yochem, L. Phillips, and F. Martinez. **Event-Driven Architecture: How SOA enables the Real-Time Enterprise**, Addison-Wesley, Boston, 2009.

11. R. Zdunek, and A. Cichocki. **Nonnegative Matrix Factorization with Quadratic Programming.** Neurocomputing*,* Vol. 71, No. 10-12, pp. 2309 - 2320, 2008.

12. D. Schneider, J. Schon, and S. Eickeler. **Towards Large Scale Vocabulary Independent Spoken Term Detection: Advances in the Fraunhofer IAIS Audiomining System**, in Proc. SIGIR, Singapore, 2008.

13. A. Skarlatidis, G. Paliouras, G. Vouros, A. Artikis. **Probilistic Event Calculus based on Markov Logic Networks**, in Proc. Of 5th International Symposium on Rules (RuleML 2011) Part 2, Fort Lauderdale, USA, 03-05 November 2011.