

# Event Discovery in Multimedia Reconnaissance Data Using Spatio-Temporal Clustering

Bo Gong, Utz Westermann, Srikanth Agaram, Ramesh Jain

Department of Computer Science  
University of California, Irvine  
Irvine, CA 92697-3425, USA  
{bgong, utz.westermann, sagaram, jain}@ics.uci.edu

## Abstract

In reconnaissance application scenarios, support for the analysis of important events that happened during a mission is highly desirable. This demands techniques to discover those events from media and sensor data that have been captured during missions. Because reconnaissance missions constitute uncontrolled environments, high-level event detection based purely on media and sensor data analysis methods is difficult. In this paper, we propose the use of spatio-temporal clustering for the discovery of important mission events. We cluster basic events that occurred during a mission – such as the creation of content or basic events detected via media or sensor data analysis - according to time and location of their occurrence. Experiments performed on real-world patrol data show the efficacy of our approach. They indicate the general usefulness of spatio-temporal clustering for event detection in scenarios where media and sensor data analysis methods are not reliable, or practically absent.

## Introduction

Generating reports of military reconnaissance missions is important. In the DARPA project “Electronic Chronicling and Group Wear for Advanced Soldier Information Systems and Technology” (EC-ASSIST) – on which we work with partners from IBM T. J. Watson Research Center, Georgia Tech, MIT, and AWARE Tech – we are developing a multimedia eChronicle that helps soldiers to analyze and understand what happened during a mission. It is thus necessary to discover important events from information captured during the mission. Thanks to the advancement of technologies, wearable devices have become available that can be used for recording documentary data of various modalities during reconnaissance. For example, cameras, microphones, and camcorders can be used to capture images, audios, and videos. Clocks and GPS receivers can be used to record time and location of a soldier. Moreover, wearable accelerometer sensors can be used to detect body movement. With these devices and sensors, rich data sources are provided to analyze and understand the incidents that occurred in a mission. However, reconnaissance missions are conducted while soldiers are freely moving through incompletely mapped or

even unknown and noisy outdoors areas; the incidents that may be of highly varying nature (for example, truck rides, dialogues with inhabitants, fights, arrests, car examinations), unexpected and uncontrolled (for example, a sudden explosion forcing soldiers to duck). This not only makes it difficult to come up with a complete set of event categories that should be detected from the recorded sensor and media data. Such a dynamic environment will also cause data to be recorded with varied quality. For example, photos will be blurry, audio recordings highly noisy, and voice distorted. Unlike other application scenarios such as building surveillance that offer more controlled environments, founding the detection of high-level events in reconnaissance missions on media and sensor data analysis is thus very difficult.

In this work, we propose to utilize time and location information for high-level event discovery. Time and GPS positions can be obtained in a reliable fashion even in uncontrolled outdoors environments like reconnaissance missions. Specifically, we discover events by applying spatio-temporal clustering techniques on basic events in a mission. The basic events during a mission include the media creation events and lower-level media or sensor data analysis events. For instance, taking a photo is a media creation event; a “gunshot” event detected from audio data and a “ducking” event derived from accelerometer data are media and sensor data analysis events. We cluster these basic events according to their occurrence time and location that are obtained by cross-referencing media production and sensor data with clock and GPS data.

The assumption is that the resulting clusters consist of semantically related events and denote events of interest for mission analysis. If something of interest is going on like the arrest of a person, there will be a spatio-temporal concentration of events – either because soldiers create more media to document the events going on or simply because they stay in the area of the event for a longer time. The main contribution in this paper is a method for event discovery in scenarios where media or sensor data analysis is difficult or unavailable. We have applied this approach to patrol data obtained from a real evaluation with soldiers in a training area. The experimental results show the effectiveness of our approach.

This paper is organized as follows. In the next section, we discuss related work. Then, we discuss our representation of basic events in a reconnaissance mission. We present our event discovery method. We describe our experimental methodology and present experimental results using that methodology that show the efficacy of our approach.

## Related Work

The terms event discovery and event detection have been used in many different scenarios where different kinds of media and different kinds of events are of interest.

The topic detection community has mainly been focusing on the detection of previously unknown events and topics in news texts (Yang *et al.* 1998, Yang *et al.* 1999, Brants *et al.* 2003). In this case, the events of interest have very coarse granularity. “Asian tsunami in 2004” is a good example of this type of event.

Event extraction from text is a closely related area. Its goal is to detect certain types of events mentioned from the individual sentences in natural language text and recognize and merge them into a unified representation for each detected event (ACE 2005). The types of events to extract are known in advance. For example, “Movement” or “Death” could be possible event types to be extracted. Compared to event detection, events resulting from event extraction have a finer granularity.

There are also studies in clustering personal photos according to events (Cooper *et al.* 2005, Loui & Savakis 2003). The events in this application are personal, previously unknown types of events, which vary from a dinner party, wedding ceremony, to a trip, etc. In these approaches, temporal proximity is typically considered more important than spatial locality. The reason for this is that – in contrast to capture time – location information is rarely available with personal photographs.

Event detection in media streams, especially in video, has been studied extensively (e.g., Gong *et al.* 1995, Haering *et al.* 2000, Petkovic & Jonker 2001, Zhang *et al.* 2005, Zotkin *et al.* 2001). In these works, content features are used for detecting interesting state changes and to perform object detection, recognition, and tracking in - usually highly structured, and regular - video content of high quality, like news, sports, or surveillance video. The kinds of events to be detected are predefined and highly depend on the application domain.

The research papers mentioned above are related to our work. However, they all differ in at least one of the following important aspects: (a) Media are assumed to be captured in controlled environments and of high quality making high-level event detection possible via media analysis. In military reconnaissance missions, however, recorded media are not well structured and noisy (b) The types of events to be detected are known in advance. In our case, we have to cope with unexpected situations; possible events are unknown in advance. (c) The applied event detection methods are based on application-specific media analysis methods. Our approach of using time and location

of event occurrence for detection is largely application-independent as we do not make assumptions about the form, quality, and structure of content. It can thus also be used when media and sensor data analysis are unavailable.

## Event Description

It is hard to give an event definition according to which any application can identify its events strictly. The reasons are twofold. First, different applications consider different types of events. An object’s state change, a goal in a soccer game, and a soccer game itself can all be considered as events. Second, different persons interpret event differently. How people understand an event depends on their knowledge and perspectives. When their view angles are different, they will have different interpretations. Nevertheless, there are common characteristics among events in different applications. For example, the “when” (time) and “where” (location) are two fundamental aspects of events. The “what” and “who” are also important to identify an event. Note in this paper, we are studying only the effectiveness of spatio-temporal clustering in event detection and will consider factors like “what” and “who” in future work. Two kinds of events are defined in this paper: basic events and mission events.

**Basic Events** A basic event is a media creation event or a detected event using media or sensor data analysis. A media creation event is an event that a piece of media data is created at a specific time and location. For example, when a soldier takes a photo at a specific time and location during a reconnaissance mission, it is considered as a media creation event. A detected event is the event that is detected using media or sensor data analysis. A good example of such an event is “gunshot” detected by voice analysis. We denote a basic event by  $e$ . A basic event contains the following information:

**Time** ( $t_e = (sta_e, dur_e)$ ): The time of a basic event consists of two parts: the start time of the event ( $sta_e$ ) and the duration of the event ( $dur_e$ ). for a discrete image creation event, the event’s duration would be 0; for a continuous audio creation event, the event’s duration is the time length of the audio.

**Location** ( $l_e = (lat_e, lon_e)$ ): The location of a basic event is the place where the event occurs. We represent location using latitude ( $lat_e$ ) and longitude ( $lon_e$ ).

**Mission Events** A mission event is a higher-level incident occurring during a reconnaissance mission and is composed of semantically related basic events. For example, arresting a person is a mission event. Mission events are the events of interest that need to be detected in our work. We denote a mission event by  $E$ .

## Event Discovery

There are two kinds of events defined in this paper. Our goal is to detect mission events from the basic events in reconnaissance data. As discussed above, it is assumed that if something of interest is going on in a mission, there will

be a spatio-temporal concentration of events – either because soldiers create more media to document the events going on or simply because they stay in the area of the event for a longer time. Soldiers will take many images when an incident occurs, and even unreliable audio detection is likely to come across distinctive audio events that are comparably easy to detect such as shooting or screaming. In other words, spatio-temporal concentration of basic events suggests the happening of mission events. Moreover, time and location information can be reliably recorded with clocks and GPS receivers. Hence, we propose to discover mission events by clustering the basic events. Each resulting cluster is considered as occurrence of a mission event.

According to the information used, there are three possible ways to do clustering: spatial clustering, temporal clustering, and spatio-temporal clustering. Clearly spatial clustering is not suitable for this application. In reconnaissance, soldiers move back and forth. At the same location, there might be different mission events happening at different time points. So spatial clustering can not separate such events. Temporal clustering methods cluster the basic events with temporal closeness. Because they do clustering according to time without consideration of location, they can not differentiate events happening at different locations while at about the same time. For example, if two soldiers patrol at different locations but at the same time, it is hard to distinguish events recorded by these two soldiers using temporal clustering only. So temporal clustering methods might work well when mission events do not overlap in time, but would fail if there is overlap in time between mission events. Spatio-temporal clustering methods cluster the basic events according to both time and location. They cluster events together only when these events are close in time and location. Thus, in this work, we will use spatio-temporal clustering approach to cluster basic events.

### Event Similarity

We cluster the basic events according to the similarity. The similarity between two events is based on how close they are in time and location. We define the similarity  $S(e_i, e_j)$  between event  $e_i$  and  $e_j$  as follows.

$$S(e_i, e_j) = \begin{cases} 1 & t(e_i, e_j) < T, \quad d(e_i, e_j) < D; \\ 0 & \text{otherwise} \end{cases}$$

where  $t(e_i, e_j)$  is the time difference between  $e_i$  and  $e_j$ ,  $d(e_i, e_j)$  is the distance between locations of  $e_i$  and  $e_j$ ,  $T$  and  $D$  are predefined thresholds. The time difference  $t(e_i, e_j)$  is the start time of the later event minus the end time (i.e., the sum of the start time and duration) of the earlier event. If  $e_i$  occurred before  $e_j$ ,  $t(e_i, e_j)$  is calculated as

$$t(e_i, e_j) = sta_j - (sta_i + dur_i)$$

where  $sta_i$  and  $sta_j$  are the start time of  $e_i$  and  $e_j$ ,  $dur_i$  is the duration of event  $e_i$ . The distance between

locations of  $e_i$  and  $e_j$  is their physical distance. We assume that the earth is a perfect sphere when calculating the distance. Suppose the earth radius is  $R$ , then the distance  $d(e_i, e_j)$  can be computed as the following:

$$d(e_i, e_j) = \arccos[\cos(lat_i) * \cos(lon_i) * \cos(lat_j) * \cos(lon_j) + \cos(lat_i) * \sin(lon_i) * \cos(lat_j) * \sin(lon_j) + \sin(lat_i) * \sin(lat_j)] / 360 * 2 * \pi * R$$

where  $lat_i$ ,  $lon_i$  and  $lat_j$ ,  $lon_j$  are latitude and longitude in degrees of  $e_i$  and  $e_j$ .

### Threshold Selection

According to the formula, the similarity between two events is determined by the thresholds  $T$  and  $D$ . So selection of  $T$  and  $D$  will affect the event clustering. As  $T$  and  $D$  increase, the number of clusters decreases. However, when  $T$  and  $D$  are too large, it tends to group events that should be separated. On the other hand, if  $T$  and  $D$  are too small, it tends to separate events that should be clustered together.

We choose  $T$  and  $D$  based on the data of each mission: between different missions, the temporal and spatial characteristics of basic events may be different. The thresholds then need to be adjusted accordingly. For example, when the patrolled area is bigger, the temporal and spatial distances between basic events might be bigger as well, demanding larger thresholds. In this work, we estimate  $T$  and  $D$  as follows:

$$T = \theta_T * \max(t(e_i, e_j));$$

$$D = \theta_D * \max(d(e_i, e_j))$$

where  $\max(t(e_i, e_j))$  is the maximal temporal distance between any two basic events;  $\max(d(e_i, e_j))$  is the maximum of spatial distance between any two basic events.  $\theta_T$  and  $\theta_D$  are the coefficients with a value between 0 and 1. We have estimated them from our experience from previous training data.

### Event Clustering Algorithm

In this work, mission event discovery is done via clustering of basic events. We use a simple single-pass clustering method (van Rijsbergen 1979). The process of the clustering is as follows. First, all basic events are sorted according to their start times in chronological order as a sequence  $\{e_0, e_1, \dots, e_n\}$ . Second, the thresholds  $T$  and  $D$  are computed. Then, for each incoming event  $e_i$ , we determine if it should be in a new cluster or not. This process continues until the last event is finished. We assume mission events do not overlap in time for data recorded by individual soldier. So if any event  $e_i$  is not in the cluster that  $e_{i-1}$  belongs to,  $e_{i+1}$  will not be in the same cluster as  $e_{i-1}$ . At the same time, if  $e_{i+1}$  and  $e_i$  belong to different clusters,  $e_{i+1}$  and  $e_{i-1}$  will not be in the same cluster. Therefore, in order to judge whether or not  $e_i$  should be in a new cluster, we only need to determine if it is in the same cluster as  $e_{i-1}$ . If the similarity  $S(e_{i-1}, e_i)$  between  $e_i$  and  $e_{i-1}$  is equal to 1, they are in the same cluster and  $e_i$  is assigned to the cluster of  $e_{i-1}$ .

Otherwise, a new cluster is formed with  $e_i$ . Algorithm 1 details the steps of computation.

**Algorithm 1. Spatio-Temporal Clustering**

1. Sort basic events in chronological order according to their start times,  $\{e_0, e_1, \dots, e_n\}$ .
2. Compute  $T$  and  $D$ .
3. Form a cluster with  $e_0$  which occurred the earliest.
4. For  $i = 1, \dots, n$ 
  - a. Compute the similarity  $S(e_{i-1}, e_i)$ .
  - b. If  $S(e_{i-1}, e_i) = 1$ , assign  $e_i$  to the cluster of  $e_{i-1}$ . Otherwise, form a new cluster with  $e_i$ .

## Experimental Methodology

### Experiment Datasets

The datasets for our experiments are patrol data obtained from a real DARPA evaluation with soldiers in a training area. During the reconnaissance mission, two soldiers recorded what happened with cameras and microphones. There are also accelerometers attached to the bodies to monitor the state of soldiers. Image data is produced mainly by an automatic camera at the helmet taking images every 5 seconds. Audio is continuously recorded and segmented into 30 second intervals. This gives the temporal view of the data a very regular appearance. This also means that temporal clustering can not be expected to provide very good results. The time and location of a soldier were continuously captured using clock and GPS receiver and cross-referenced with the basic events that occurred during the mission.

The mission was a short patrol of a village involving the capture of two insurgents. Both soldiers started the mission with a ride in a humvee that ended close to the village. The soldiers then walked to the edge of the village and split up into two teams. The first team provided cover for the second, while second explored the village grounds and buildings. The second team then provided the first team cover from the second floor of a building while the first team entered another and captured the first insurgent. After that the second team entered a third building and captured the second insurgent. The whole mission lasted 30 minutes and there are 1112 basic events. We prepared three datasets. The first is the data recorded by the first soldier, which consists of 643 basic events. The second is the data recorded by the other soldier. It contains 469 basic events. The third dataset is the combination of the data recorded by both soldiers. Table 1 shows the spatio-temporal statistics of the three datasets.

Table 1. Spatio-temporal statistics of datasets

	Average $t(e_i, e_j)$ (second)	Maximal $t(e_i, e_j)$ (second)	Average $d(e_i, e_j)$ (m)	Maximal $d(e_i, e_j)$ (m)
Dataset1	3.735	11	27.73	151.22
Dataset2	4.164	30	15.57	126.88
Dataset3	2.259	10	27.95	151.22

## Experiments Design

In this section, we now explain how we carried out the experiments. In the first experiment, we use the spatio-temporal clustering method to cluster the basic events on Dataset 1 and Dataset 2. We compare the generated results with the ground truth to see how the approach performs. Since we assume that mission events do not overlap in time for individually recorded data, temporal clustering method can be used as well. It is of interest to compare temporal clustering method with the spatio-temporal clustering approach. Thus, in the second experiment we take temporal clustering method as the baseline for comparison and apply it on Datasets 1 and 2. The temporal clustering method is similar with the spatio-temporal clustering approach described before. The only difference is that temporal clustering method does not consider the location. So  $S'(e_i, e_j)$  denoting the similarity in temporal clustering is defined as the following:

$$S'(e_i, e_j) = \begin{cases} 1 & t(e_i, e_j) < T; \\ 0 & otherwise. \end{cases}$$

In the third experiment, we applied both temporal clustering method and spatio-temporal clustering approach on Dataset 3.

## Evaluation Methodology

In order to evaluate the performances of different clustering methods, we need to compare the generated results with the ground truth. In our experiments, the basic events in all three datasets have been subjectively clustered into mission events by a human annotator who participated in the DARPA evaluation. The labeled data are considered as the ground truth. Figure 1 shows the mission events in Datasets 1 and 2. Because of problems with the microphone setup, the voices recorded during the mission turned out to be very noisy and unclear making it difficult to label them. So in our experiments, only images are labeled. We only use images for comparison; however, the clustering is applied to all basic events regardless of modality.

Precision ( $P$ ), recall ( $R$ ), and  $F_1$  are well-known metrics in evaluating clustering results. Thus, we adapted them as the evaluation measures in this work. They are defined as follows.

$$P = \frac{T^+}{T^+ + F^+}, \quad R = \frac{T^+}{T^+ + F^-}, \quad F_1 = \frac{2 * P * R}{P + R}.$$

$T^+$ ,  $F^+$ ,  $T^-$ , and  $F^-$  in the formula above stand for true positives, false positives true negatives and false negatives respectively. These notions are summarized in Table 2. In our experiments, all of them are calculated for each computed mission event first, then the measures are computed.

Table 2. Event contingency table

	In cluster	Not in cluster
In ground truth	$T^+$	$F^-$
Not in ground truth	$F^+$	$T^-$

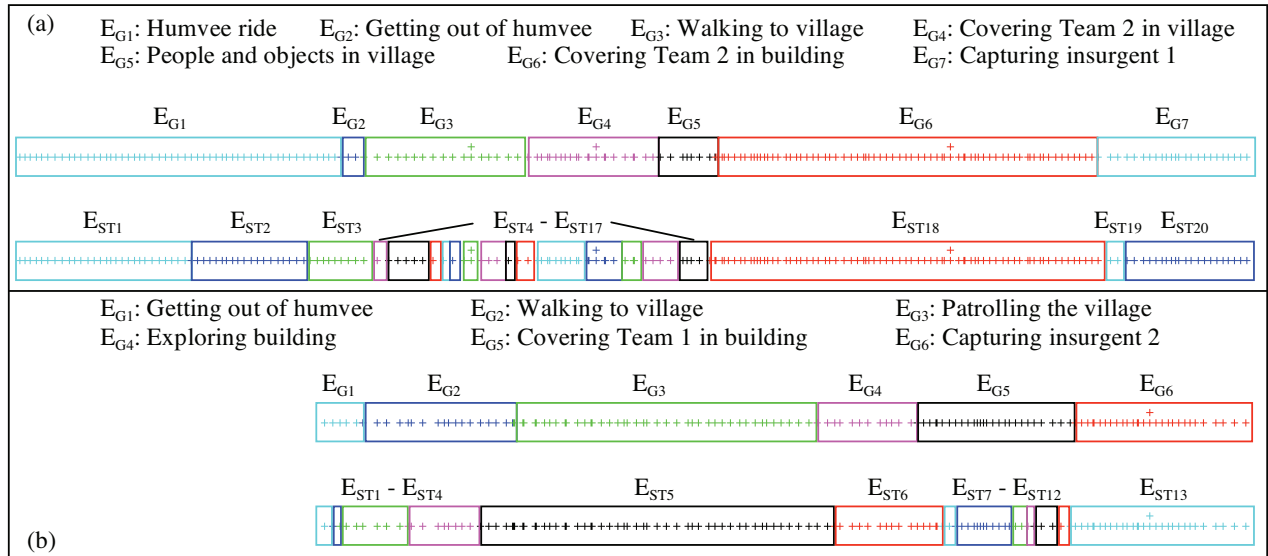


Figure 1. Drawings of generated clusters using spatio-temporal clustering approach (below) and the ground truth clusters (above) in time: (a) Dataset 1. (b) Dataset 2.

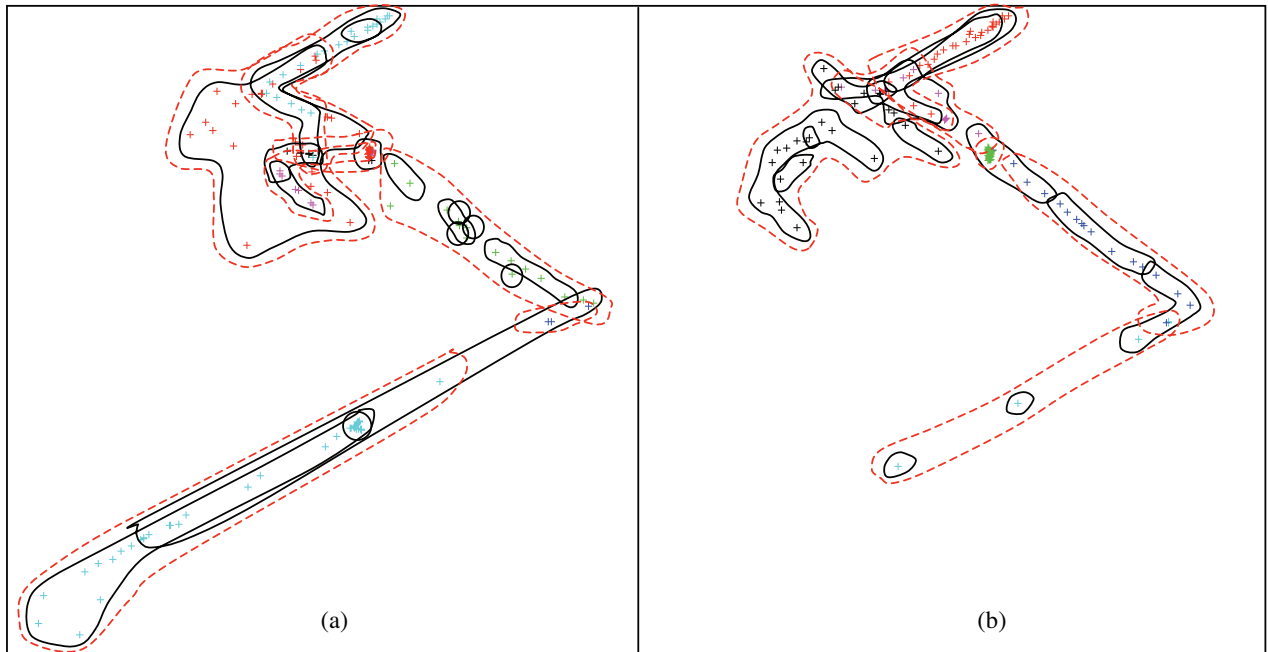


Figure 2. Drawings of generated clusters using spatio-temporal clustering approach (solid line) and the ground truth clusters (dotted line) in space: (a) Dataset 1. (b) Dataset 2.

## Experimental Results

In the first experiment, the spatio-temporal clustering approach was applied to Datasets 1 and 2. Figures 1 and 2 overlay the generated clusters with the ground mission events in both temporal view and spatial view. As shown in the figures, the generated clusters match well with the ground truth clusters.

In the second experiment, we applied temporal clustering method to Datasets 1 and 2 as well. Table 3 shows the comparisons. The temporal clustering method is far inferior to spatio-temporal clustering approach for two reasons. In these datasets, the temporal distribution of events is very uniform. Moreover, soldiers have been moving at different speeds: they rode a vehicle, ran, and walked. Given the regular temporal distribution of events

in the datasets, these changes in movement, which certainly are interesting incidents, are hard to detect without considering location.

Table 3. Performance comparison of temporal clustering and spatio-temporal clustering on Dataset 1 and Dataset 2

	Clustering method	Precision	Recall	$F_1$
Dataset1	Temporal	.74	.84	.78
	Spatio-temporal	.94	.72	.81
Dataset2	Temporal	.17	1.0	.29
	Spatio-temporal	.86	.79	.83

We applied both the temporal and spatio-temporal clustering methods to Dataset 3 in the third experiment. Table 4 shows the results. Although the spatio-temporal clustering approach is still significantly better compared to purely temporal clustering, its performance is not as good as it was for Datasets 1 and 2. The result is not completely surprising because our approach assumed that there is no overlap in time between mission events. However, Dataset 3 is the mixture of Dataset 1 and Dataset 2. As both soldiers have split up during missions, they have produced basic events at the same time at different location. Thus, this basic assumption is no longer valid. The situation is further complicated by the fact that the GPS positions of both soldiers differ even when being at the same place due to error.

Table 4. Performance comparison of temporal clustering and spatio-clustering on dataset 3

Clustering method	Precision	Recall	$F_1$
Temporal	.28	.99	.44
Spatio-temporal	.56	.69	.62

## Conclusions and Future Work

In this paper, we studied the problem of mission events discovery for generating reports of reconnaissance missions. Because of the uncontrolled environment in military reconnaissance mission, media and sensor data captured during the missions are of varied quality. Methods to discover events based purely on media and sensor analysis results are difficult.

We proposed to utilize time and location information to discover the mission events by clustering the basic events of the mission. We developed a spatio-temporal clustering approach that takes advantage of time and location and showed that it improved significantly over the baseline of temporal clustering method.

However, the experiments also showed that this basic approach needs further refinement. The basic assumption that basic events do not temporally overlap is not valid when considering events of different soldiers. Moreover, we used fixed temporal and spatial thresholds in clustering. But fixed spatial thresholds, for instance, cannot take full account of varying movement speeds and transportation methods of soldiers. In our data sets, soldiers changed from riding vehicles to walking, from walking to running, etc.

Adaptive thresholds could gracefully adapt to the resulting changes of spatial distances between basic events.

Finally, in this work, the spatio-temporal clustering approach was proposed with the assumption that event detection via sensor data and media analysis in reconnaissance patrols is very difficult. It is of interest to study how spatio-temporal clustering can be combined with media and sensor data analysis results. The consideration of media and sensor data analysis results will also help in characterizing the kinds of events detected by clustering. So far, we leave the interpretation of the meaning of a mission event cluster to the user.

## References

- Automatic Content Extraction (ACE). 2005. The ACE 2005 (ACE05) Evaluation Plan. <http://www.nist.gov/speech/tests/ace/>.
- Brants, T., Chen, F., and Farahat, A. 2003. A System for New Event Detection. In *Proc. of the SIGIR Conference on Research and Development in Information Retrieval*.
- Cooper, M., Foote, J., Girgensohn, A., and Wilcox, L. 2005. Temporal Event Clustering for Digital Photo Collections. In *ACM Transactions on Multimedia Computing, Communications and Applications*.
- Gong, Y., Sin, L., Chuan, C., Zhang, H., and Sakauchi, M. 1995. Automatic Parsing of TV Soccer Programs. In *Proc. International. Conference on Multimedia Computing and Systems*.
- Haering, N., Qian, R., and Sezan, M. 2000. A Semantic Event-Detection Approach and Its Application to Detecting Hunts in Wildlife Video. In *IEEE Transactions on Circuits and Systems for Video Technology*.
- Loui, A. and Savakis, A. 2003. Automatic event clustering and quality screening of consumer pictures for digital albuming. In *IEEE Trans. Multimedia*.
- Petkovic, M. and Jonker, W. 2001. Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events. In *IEEE Workshop on Detection and Recognition of Events in Video*.
- van Rijsbergen, C.J. 1979. Information retrieval.
- Yang, Y., Carbonell, J., Brown, R., Pierce, T., Archibald, B. T., and Liu, X. 1999. Learning approaches for detecting and tracking news events. In *IEEE Intelligent Systems Special Issue on Applications of Intelligent Information Retrieval*.
- Yang, Y., Pierce, T., and Carbonell, J.G. 1998. A study on retrospective and on-line event detection. In *Proc. of the SIGIR Conference on Research and Development in Information Retrieval*.
- Zhang, D., Gatica-Perez, D., Bengio, S., and McCowan, I. 2005. Semi-supervised Adapted HMMs for Unusual Event Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Zotkin, D., Duraiswami, R., Davis, L. 2001. Multimodal 3-D Tracking and Event Detection via the Particle Filter. In *IEEE Workshop on Detection and Recognition of Events in Video*.