© 2017 IJSRSET | Volume 3 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Multi-Modal Representation of Public Safety Events Using Social Media and Surveillance

Dhanya Vijay, S. Antony Mutharasan
Department of Computer Science and Engineering, Sardar Raja College of Engineering, Alangulam, Thirunelveli, Tamilnadu, India

ABSTRACT

Public safety event is a danger and urgent event that need early detection, fast response and exact recover. The efficient method for responding to a happening public safety event is to collect and describe the related data. In addition to the surveillance cameras from the physical space, the social media data can also be used to collect and describe the related data. In addition to the surveillance cameras from the physical space, the social media data can also be used to collect and describe the related data of a public safety event. In this paper the proposed method concentrates on the steps for describing public safety events. Given a public safety event, videos from the surveillance cameras and social messages from social sensors are collected. The different mode of information including texts, videos and spatial-temporal data is mined to give a description exactly and concisely. At first, the social sensors are associated to surveillance camera by the spatial and temporal information. Second, the social messages are associated to surveillance cameras by semantic information. In third step the social messages are associated to surveillance cameras by the visual features. Apart from the text, social sensors may upload images or videos. Finally the different mode of description is driven based on the three different associations.

Keywords : Mobile crowd sensing, public safety events, social media, spatial-temporal data, surveillance cameras.

I. INTRODUCTION

Now, the public safety events have an influence on the daily life of a person in the society. We can say that no individuals are immune from the public safety events that need early detection, fast response and exact recovers [15]. The public safety event can be divided into three types by the cause reasons. The first type is caused by nature, which can be named as natural safety event. For example, On April 2015 a highly destructive earth quake struck Nepal, nearly 9,000 people are killed and nearly 22,000 people are injured. The second type is caused by disease, which can be termed as health safety event. For example, Zika virus is an infectious disease spread by mosquito bites. The Zika outbreak has raised concerns among travelers with plans to visit affected regions. The third type is caused by person themselves. The terror attacks in Pathankot and Uri are the most recent attack of external forces. This paper focus on the third category of public safety event, such as terror attacks, traffic event and non-rural emergency event.

The efficient method for response to a happening public safety event is to collect and describe the related data. Recently video surveillance system is used to collect the data of public safety event. About 95% of public safety event uses video surveillance system to get data and try to recover the situation. This is an important safety system because of its accurate, timely and rich information content [16]. But the current video surveillance system has some problem to process public safety events 1) it is highly expensive 2) less coverage and 3) it is difficult to extract the real-time related information from the surveillance videos. The important factor of a public safety event such as face of a suspect and the registration number of a vehicle are difficult to extract [50].Moreover these factors the property of a public safety event is also difficult to find. For example the surveillance camera cannot detect a fight event automatically with higher precision.

Along with the surveillance cameras the social media data can also be used to describe the related data of a public safety event [4], [1] with the development of mobile crowd sensing method the social media user can transfer the information into knowledge. The mobile crowd sensing participant can be as “social sensors” [12]. A social sensor is defined as an agent that provides information about environment on a social network after interaction with other agents [9]. The physical sensors present in mobile devices such as GPS are used to infer social relationships and human activities. Social media are the appropriate platforms. But these social sensor based descriptions have some challenges. 1) The quality of the social media data is low 2) most of the social media messages are with low value 3) the social messages about public safety events are with week association. The spatial and temporal relations between social messages are difficult to mined and related.

The data from the surveillance camera and social media are integrated in this paper, the disadvantages of surveillance camera and social media can be made up each other.

1) The mobile crowd sensing on social media can provide the data of public safety event which the surveillance camera does not cover the important features of mobile crowd sensing is pervasive. The witness can give description by social media through the place of the event may not cover by a surveillance camera.
2) The mobile crowd sensing on social media can provide different data other than only video data by surveillance cameras. The social media user can give the temporal, spatial and semantic information of a public safety event. The semantic information can improve the event detection. The social media can improve the precision of event description.
3) The surveillance camera can be used to reduce the noise and redundancy of social media data. If a public safety event can be detected by a surveillance camera, the spatial and temporal information can provide the location and time of the camera.
4) The surveillance camera can be used to link the social media data of the public safety event. The spatial and temporal information can be used to associate the related information of social media. The event information of the surveillance camera can be used to link the social media about the same event.
The proposed method focuses on the step for describing public safety. Given a public safety event, the videos from surveillance cameras and messages from social media are collected. The different information including texts, images, videos and spatial-temporal data is mined to give the description accurate and concise. Basically, the proposed method consists of four stages. First, the social sensors are linked to the surveillance camera by the spatial and temporal information. The social messages and surveillance video can be related if they are in the same time interval or the near place. Second, the social messages are associated to surveillance cameras by the semantic information. The semantic information can be extracted from video using the event detection algorithm. The social messages and surveillance videos can be related if they mention the same event. In the third stage, the social messages are linked to surveillance cameras by the visual feature. In addition to the texts social sensors may upload images or videos. The visual features such as face, vehicle register number can be extracted and related to the videos. Finally the multi modal representation step is given based on the three different associations. To remove the data noise and redundancy, data fusion method is used.

This paper proposes a novel method for describing public safety events using multi-modal data. These data is from different space (cyber space and physical space) is used.

The four different stages of the proposed method are given. The four stages aim at linking related information from social media and surveillance camera.

II. METHODS AND MATERIAL

In this section, the problem formulation is given first, and then the basic definitions of the proposed method are listed. Finally the basic elements of the social media and surveillance videos of public safety events are also introduced. The proposed description of public safety events based on the integration of social media and surveillance videos are given.

Input: The public safety event, the related social messages of the event and the related videos from surveillance camera.

Output: The different description of the given public safety events including spatial-temporal description, visual description and semantic description.

A) Semantic Description

Semantic description is output to detect which kind of event happens. The semantic description can be extracted from the keywords of social media and the event detected from the surveillance cameras.

B) Spatial-Temporal Description

The spatial description including the happened and spread places of the public safety event. The spatial description can be extracted from the information of social media users or the information from the surveillance camera.

The temporal data include the time stamp information of the public safety event. This data can be extracted from the post time of social message or the system time of the surveillance camera.

C) Visual Description

This data include the image and video of the public safety event. The visual description can be extracted from the uploaded visuals of the social messages or surveillance camera.

PROBLEM DEFINITION

A public safety event is something that happens at some specific time and often at some specific places [17] is defined as follows.

Definition 1) Public Safety Event (p) a public safety event is something that happens at some specific time and at some specific places.

Definition 2) Social Message Set of a Public Safety Event (MS (p)) the social message set contains the related messages of a public safety event. Social message mention the place, timestamp, text or visual information of the public safety event. The message set is denoted as

$$MS\ (p) = \{ms_1, ms_2 \ldots ms_{|MS\ (p)|}\} \tag{1}$$

Definition 3) Surveillance Camera Set of public safety event CS(p)

The nearby camera set contains the cameras near the happening place public safety event which is defined as

$$CS\ (p) = \{cs_1, cs_2 \ldots cs_{|CS\ (p)|}\} \tag{2}$$

Definition 4) Semantic Element of the Public Safety Event

The semantic elements include Keywords (KW) Patterns (P) and Sentences (S) from MS (p) these elements are denoted as

$$KW\ (p) = \{kw_1, kw_2 \ldots kw_{|KW\ (p)|}\} \tag{3}$$

$$P\ (p) = \{p_1,\ p_2 \ldots p_{|P(p)|}\} \tag{4}$$

$$S\ (p) = \{s_1, s_2 \ldots s_{|S(p)|}\} \tag{5}$$

The semantic elements also includes event type (ET) from CS (p) which is denoted as

$$ET\ (p) = \{et_1, et_2 \ldots et_{|ET(p)|}\} \tag{6}$$

Definition 5) Spatial Elements of Public Safety Event

The spatial elements include Location Stamps (LS) Check-in (CI) from MS (p). These elements are denoted as

$$LS\ (p) = \{ls_1, ls_2 \ldots ls_{|LS(p)|}\} \tag{7}$$

$$CI\ (p) = \{ci_1, ci_2 \ldots ci_{|CI(p)|}\} \tag{8}$$

The spatial elements also include Geo information (GI) of cameras from CS (p) which is denoted as

$$GI\ (p) = \{g_1, g_2 \ldots gi_{|GI(p)|}\} \tag{9}$$

Definition 6) Temporal Elements of Public Safety Events

The temporal elements provide the information of the public safety event. The temporal elements include Posted Time (PT) from MS (p) which is denoted as

$$PT\ (p) = \{pt_1, pt_2 \ldots pt_{|PT(p)|}\} \tag{10}$$

The temporal elements also include System Time (ST) of camera from CS (p) which is denoted as

$$ST\ (p) = \{st_1, st_2 \ldots st_{|ST(p)|}\} \tag{11}$$

Definition 7) Visual Elements of the Public Safety Event

The visual elements include Uploaded Images and video (UI) from MS (p) which is denoted as

$$UI(p) = \{ui_1, ui_2 \ldots ui_{|UI(p)|}\} \quad (12)$$

Visual elements also include camera videos (CV) from CS (p) which is denoted as

$$CV(p) = \{cv_1, cv_2 \ldots cv_{|CV(p)|}\} \quad (13)$$

3. The Proposed Method

The proposed method consists of four layers including spatial-temporal integration layer, semantic integration layer, visual layer and multi- modal representation layer. The proposed framework is set as a hierarchical model including three layers.

Multi-modal Representation Layer

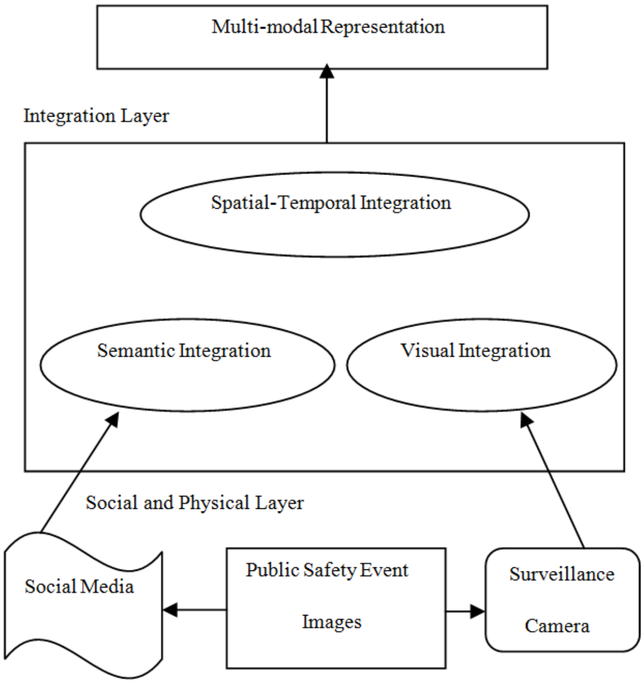

Multi-modal Representation
Integration Layer
Spatial-Temporal Integration
Semantic Integration
Visual Integration
Social and Physical Layer
Social Media
Public Safety Event Images
Surveillance Camera

Figure 1: Structure of the Proposed Method

A) Social and Physical Layer

In this layer the related data to a public safety event is collected. The data space includes physical and social space. The social space collects the data from social media. The users may post related messages when they see a real- time happening public safety event. The physical space collects data from the surveillance cameras. The cameras may record the video about the public safety event.

B) Integration layer

In this layer the basic elements are extracted. The integration method based on semantic, spatial, temporal and visual elements from social and physical layer is given.

C) Multi-Modal Representation Layer

In this layer multi-modal representation of public safety event is launched. This includes semantic, spatial, temporal and visual aspects.

3.1 Integration of Semantic Elements

The semantic elements include patterns (P), keywords (KW) and Sentences (S) from social media. Keywords can describe the type of a public safety event. For example the keyword “FIRE” and “Fire Man” can reflect a real time fire event. The semantic elements include event type (ET) from surveillance cameras. The integration of social media and surveillance camera only based on semantic elements is not enough for the description. The same type of event may happen in different places. So the spatial and temporal information should be considered.

3.2 Integration of Spatial-Temporal Elements

Spatial elements include check-in (CI) and location stamp (LS) information from social media. The location stamp is the world; it can be a building, a road, or a famous place etc. The location stamp is usually a spatial field.

Besides the location stamp, the social media users also provide check-in information. This is usually a tuple with the longitude and latitude data. The check-in information is a unique spatial data. The spatial elements also include Geo information (GI) from surveillance cameras. The Geo information is similar to the check-in information of a social user.

The integration of the spatial information from social media and surveillance camera is based on map distance. The spatial information can be seen as a node and the distance between nodes can be seen as a link. In the proposed system the spanning tree algorithm is used to simplify the network. The weight of the minimum spanning tree is set as the direct distance in the map.

The temporal elements include posted time (PT) from social messages. Each social message has a post time stamp. The temporal data also include system time (ST) of surveillance cameras. The integration of temporal information is based on time range. The time range is from the starting time to the ending time of the public safety event. The important time stamp must contain new location check-in information or new keywords.

3.3 Integration of Visual Elements

The integration of visual elements between social media and surveillance camera is based on person, vehicles and scene. These are extracted from uploaded image by saliency analysis method [13]. Content Based Image Retrieval (CBIR) system is based to integrate the element from social media and surveillance camera. The person and vehicle extracted from the uploaded image by social media users are set as the search query to the CBIR system. The videos from the cameras are set as data base to the system. Visual statements are built using the extracted interest points. Later the visual statements are used to estimate the similarity distance which are used to create the image data set clusters.

3.4 Multi-Modal Representation

The multi-modal representation of public safety event is based on the spatial-temporal information, semantic information and visual information. The data is from social media and surveillance camera. The representation is different where the proposed method chooses different space as the event detector.

Steps of multi-modal representation

1) Some keywords are used to search the related social message of a public safety event. For example “Flood” is used as a keyword to detect a flood related event.
2) Not all social messages provide useful information. It is needed to remove noisy messages, for this the filtering method is used.
3) The spatial-Temporal, semantic and visual elements are extracted from the selected social messages. The surveillance cameras near the spatial information from social messages are selected. After collecting the information integration is conducted to give the multi-modal representation. The integration step improves the accuracy of event detectors.

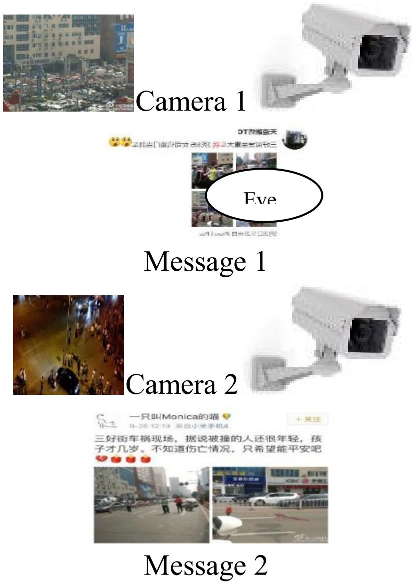
Another method uses surveillance camera as the detector. This method is different from social media based scenario. The initial event detector in this method is manually. Since the precision is very low. It is unpractical to use automatic surveillance cameras. The camera can only provide videos compared to social media.

III. RESULTS AND DISCUSSION

4.1 Data Sets

Four types of public safety events are selected as the experimental data including fire, car crash, crowd and protest. These four types of events can be detected by detectors. Four real events are taken as events to collect related data.

Data from social media and surveillance cameras are collected as the dataset. The social media data is collected by the keywords of these four kinds of events. The video from the surveillance cameras are collected from the time range of the happening time of the public safety event.



一只叫Monica的猫
+ 关注
9-26 12:19 来自小米手机4
三好街车祸现场，据说被撞的人还很年轻，孩子才几岁。不知道伤亡情况，只希望能平安吧

Figure 2. The illustration of The “procession event”

4.2 **Evaluation of Semantic Integration**

Several social messages using the keywords of the public safety events are collected. Some social messages are selected using filtering method [14]. The idea of evaluating the semantic integration method is based on whether the accuracy of detecting public safety events improves or not. When the surveillance cameras are used as the event detection initiator, the evaluation steps are as follows.

1) In a given time range of the event, the detectors are used to detect the event in the surveillance cameras. The average event detector result of the surveillance camera is captured and the event detector result of each camera is ranked.
2) Doing the semantic integration of social messages and surveillance cameras, the integration result of each camera is ranked.
3) Comparing the new ranking based on the ranking in first and second step.

4.3 Evaluation of Spatial-Temporal Integration

The spatial temporal integration is similar to semantic integration. Social messages are selected from all social messages which provide spatial or visual information from these messages. Check-in positions are extracted. These positions are the real positions of the social media users. When the social messages are used as the event detection initiator, the evaluation steps are as follows.

1) The check-in information is extracted from social messages. The check- in position is with the longitude and latitude information.
2) The related surveillance camera are selected for each check-in position. The selection method is set as distance and time range. The surveillance videos from the cameras are selected before and after the time stamps of the social message.
3) Searching the detection result of the nearby surveillance cameras to check whether the social messages mention a real happening event.
The integration of spatial-temporal information can filter the noisy information from social media.

IV. CONCLUSION

In this framework a new method for representing public safety event with multi-modal information; given a public safety event the information from cyber space and physical space integrated. The proposed model including text, images, videos and spatial-temporal data is mined to give a description accurate and concise. The social messages and surveillance cameras are integrated by semantic, spatial-temporal and visual information. The multi-modal representation step has been given based on the four kinds of integration.

V. REFERENCES

[1]. Azfar A, Choo K-K R and Liu L 2016. An Android Social App Forensics Adversary Model. In Proceedings of 49th Annual Hawaii International Conference on System Sciences (HICSS 2016),pp. 5597–5606, 5–8 January 2016, IEEE Computer Society Press.

[2]. Ariffin A, Choo K-K R,and Yunos Z,2016. CCTV Forensic Readiness: A Case Study on Digital CCTV Systems. In Choo K-K R and Dehghantanha A, editors, Contemporary DigitalFo-rensic Investigations of Cloud and Mobile Applications, Syngress, an Imprint of Elsevier.

[3]. Bai L, Lao S, Jones G, and Smeaton A. Video Semantic Content Analysis Based on Ontology. InProceedings of the 11th Inter-national Machine Vision and Image Processing Conference, 2007, 117-124.

[4]. Chin P Y, Evans N and Choo K-K R 2015. Exploring Factors Influencing the Use of Enterprise Social Network in Multinational Professional Service Firms. Journal of Organizational Computing and Electronic Commerce 25(3): 289-315

[5]. Chin C P-Y, Evans N and Choo K-K R 2015. Enterprise Social Networks: A Successful Implementation within aTelecommu-nication Company. In Proceedings of 21st Americas Conference on Information Systems (AMCIS 2015), 13–15 August 2015, As-sociation for Information Systems.

[6]. David A. Forsyth. Object Detection with Discriminatively Trained Part-Based Models. IEEE Computer 47(2): 6-7 (2014).

[7]. Fan J, Aref W, Elmagarmid A, Hacid M, Marzouk M, and Zhu X. Multiview: Multilevel Video Content Representation andRe-trieval. Journal of Electronic Imaging, 2001, 10(4):895-908.

[8]. C. Hu, Z.Xu, et al. Video Structured Description Technology for the New Generation Video Surveillance System. Frontiers of Computer Science, 2015, 9(6): 980-989.

[9]. V. Krishnamurthy and H. Vincent Poor. A Tutorial on Interac-tive Sensing in Social

Networks. IEEE Transactions on Computa-tional Social Systems, 1(1):3-21, 2014.

[10]. X. Lu and C. Brelsford. Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami. Scientific Reports, Article number: 6773, 2014.

[11]. Marszalek M, Schmid C, and Inria M. Semantic hierarchies for visual object recognition. InProceedings of 2007 IEEE Comput-er Society Conference on Computer Vision and Pattern Recog-nition, 2007.

[12]. G. Valkanas and D. Gunopulos. How the Live Web Feels about Events. In Proceedings of ACM Conference of Information and Knowledge Management, pp. 639-648, 2013.

[13]. S.Wang, Y.Shang, J.Wang, L.Mei, and C.Hu. Deep Features for Person Re-identification,SKG 2015: 244-247.

[14]. Z.Xu,H.Zhang,C.Hu,L.Mei,J.Xuan,K.R.Choo,V.S ugumaran,and Y.Zhu. Building knowledge base of urban emergency events based on crowdsourcing of social media. Concurrency and Computation: Practice and Experience, 2016,28(15):4038-4052.

[15]. J. Xuan, X. Luo, G. Zhang, J. Lu, and Z.Xu. Uncertainty Analy-sis for the Keyword System of Web Events. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 46(6):829-842, 2016.

[16]. Z.Xu et al. Semantic based representing and organizing sur-veillance big data using video structural description technology.The Journal of Systems and Software, 2015, 102, 217-225.

[17]. C. Yang and X. Shi. Discovering event evolution graphs from newswires. In Proceedings of the 15th international World Wide Web conference, pp. 945-946, 2006.

[18]. Yu-Gang Jiang, Jun Yang, Chong-Wah Ngo, Alexander G. Hauptmann. Representations of Keypoint-Based Semantic Con-cept Detection: A Comprehensive Study. IEEE Transactions on Multimedia, vol. 12, issue 1, pp. 42-53, 2010.