Socializing Multimodal Sensors for Information Fusion

Yuhui Wang
NUS Graduate School for Integrative Sciences and Engineering
National University of Singapore, Singapore
wangyuhui@u.nus.edu

ABSTRACT

In the modern big data world, development of a huge number of physical or social sensors provides us a great opportunity to explore both cyber and physical situation awareness. However, social sensor fusion for situation awareness is still in its infancy and lacks a unified framework to aggregate and composite real-time media streams from diverse sensors and social network platforms. We propose a new paradigm where sensor and social information are fused together facilitating event detection or customized services. Our proposal consists of **1)** a tweeting camera framework where cameras can tweet event related information; **2)** a hybrid social sensor fusion algorithm utilizing spatio-temporal-semantic information from multimodal sensors and **3)** a new social-cyber-physical paradigm where human and sensors are collaborating for event fusion. Our research progress and preliminary results are presented and future directions are discussed.

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous

Keywords

social sensor network; multimodal fusion; information fusion;

1. INTRODUCTION

We are witnessing a world of big social and sensor data explosion. All kinds of sensors that are constantly monitoring our surrounding situations. ***Physical sensors*** like visual sensors, humidity sensors, GPS sensor, mobile phones are widely used in capturing ambient signals. ***Social sensors*** such as Twitter, Facebook, Sina Weibo, Flickr are fed with exploding voluntarily user-generated data. These sensors, though working in different way, often provide complementary event related information. Their ambient sensing capabilities provide the opportunity that allows humans and machines working together to make sense of real-time situation. This has provided us an opportunity to socialize sensors with humans by which human can establish connections with sensors in social network and "follow", interact or get updates from them. Socializing physical sensors with human and fusing social sensor would not only lead to a better understanding of dynamically occurring situations but also a new social-cyber-physical ecosystem that connects sensors and people for human's needs as their situations evolve.

However, due to the diversity of these sources, currently physical sensors and social sensors are capturing information separately in their individual silos. The information captured by sensors of different modalities are residing as heterogeneous data. They are not combined or fused which impedes event detection and understanding in a comprehensive manner. Without aggregating relevant these multimodal sensors signals, it is hard to obtain a holistic and comprehensive event or situation overview. Moreover, though these sensors are observing and monitoring our physical world, there is a lack of effective interaction between them and humans for situation information retrieval and propagation.

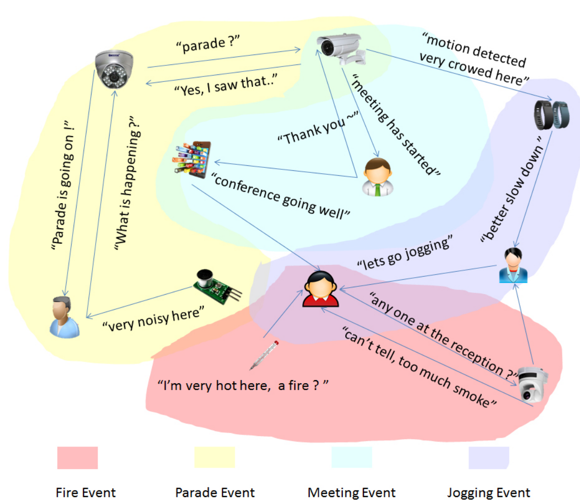


Figure 1: Social Sensors Fusion for Event Detection

This thesis proposes to extend the traditional sensing paradigm by combining sensors of different modality and socializing them with human so as to facilitate real-world event information fusion and analysis. Our conceptual idea is shown in 1. Specifically, we propose to socialize multimodal sensors for information fusion such that **1)** camera network-

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807995.

s "tweet" information when things of interest happens; **2)** spatio-temporal semantic information is fused in a pseudo-image representation when particular physical event happens, giving more precise and complementary information; **3)** humans interact with sensors through social networks by following sensors, receiving event related alert or notification from sensors, and providing semantic knowledge to sensors, making them smart so that useful information could be best gathered, utilized and disseminated through the whole human-sensor social network.

2. RELATED WORK

With an increasing number of sensors having capabilities of sensing, processing, communicating, usage of sensors in event detection and situation awareness is spreading. Massively distributed visual sensors (webcams) are utilized for environment understanding [8]. A multi-tier network SensEye of heterogeneous cameras has been proposed to overcome the disadvantage of single-tier networks in surveillance application, performing object detection, recognition and tracking tasks [11]. Distributed smart cameras [4] are also being widely used in camera sensor networks to produce alerts if certain types of unusual behaviour or abnormal events occur. To bridge the gap between machine-oriented low level features and human readable high level semantics, a number of works concern image captioning [6] and video concept detection [9, 3], where the task is to assign concept labels to an input image/video along with their associated probabilities. Moreover, multi-modal sensor fusion has been studied for combining multiple level of information for various multimedia tasks [1]. Fusion of mobile computing and social networking services for traffic anomaly detection has been proposed in [12]. Crowd-sourced sensing and collaboration systems over Twitter are designed and implemented which opens publish-subscribe infrastructure for sensors to be combined with social platform [5].

Many online social network services are prevalent nowadays where users share personal opinions, breaking news. Twitter, as one of the most important social sensors, has been leveraged to generate a large number of event detection or situation understanding systems such as Twitris [14] and Evenshop [15]. It is regarded as a social sensor [13] in detecting and tracking earthquakes, typhoons or traffic jams. Also a situation awareness algorithm is proposed to detect geo-spatial events in a given monitored geographic area, which offers good summary of events [16] .

With the Internet of things (IoT), everyday objects now have the ability to interconnect not only among themselves but also humans. Social networking concepts have been integrated into IoT [2], which establishes Social Internet of Things (SIoT). Things not only sense but start to update their status in social networks. Kranz et al. [10] make both humans and technical systems together to form a socio-technical network by describing cognitive office, where the states of the plant, windows and doors are posted via Twitter accounts. Many other such accounts have been created. For example, @VedamsIoTEdison tells if room lights are off or on; @VedamsIoTRPi tweets when there was a power cut in office hours; @MoneyPlantTrack posts not only plants temperature, humidity, status but also uploads images captured by camera when there are abnormal situations.

Summing up, aforementioned works consider information from either physical sensors or social sensors and do not deal with fusing these two types of sources in a unified framework and there is lack of interaction between humans and sensors which impede information fusion and processing for event recognition. Also, sensors or smart things tweet information related to predefined tasks and have limitations in application wise extension. We contribute to this part by proposing a new sensing paradigm where both social and physical sensor information are represented through unified data structure which enable multimodal information fusion and user customized situation understanding.

3. APPROACH

In order to socialize multimodal sensors and establish connections among them for event detection and fusion, we propose to first combine and process information from different sensor modalities, particular from visual sensor and social sensors to enhance event detection and situation understanding by building tweeting cameras, which tweet useful information among sensor networks or to the social networks. Second, we fuse spatio-temporal-semantic information from different modalities via a unified image-like representation named Cmage (Concept Image), so as to filter out sensor noise and give precise event situation. Third, we propose a new paradigm of sensing mechanism where people and physical sensors are collaborating for information mining, propagation and knowledge learning.

3.1 Tweeting Cameras for Event Detection

Combining these two complementary sensor streams can significantly improve the task of event detection and aid in comprehending evolving situations. However, the different characteristics of these social and sensor data make such information fusion for event detection a challenging problem. To tackle this problem, we have established connection between information from visual sensors (camera) and social sensors (Twitter), and have proposed an innovative multi-layer tweeting cameras framework integrating both physical sensors and social sensors to detect various concepts of real-world events [17]. In this framework, shown in Figure 2, visual concept detectors are applied on camera video frames and these concepts can be construed as "camera tweets" posted regularly. These tweets are represented by a unified probabilistic spatio-temporal (PST) data structure which is then aggregated to a concept-based image (Cmage) as the common representation for visualization. To facilitate event analysis, we have defined a set of operators and analytic functions that can be applied on the PST data by the user to discover occurrences of events and to analyse evolving situations. We further leveraged on geo-located social media data by mining current topics discussed on Twitter to obtain the high-level semantic meaning of detected events. The information captured by sensors of different modality is then combined to facilitate situation awareness in a comprehensive manner.

3.2 Hybrid Social Sensor Fusion

As geo-tagging techniques and mobile sensors become prevalent, spatial temporal and semantic patterns of various multimedia sensor data are now available for event detection and situation understanding. Therefore, we propose to extract spatial temporal and semantic patterns from information captured by sensors of different modalities, and represent those information in a pseudo-image based structure, where

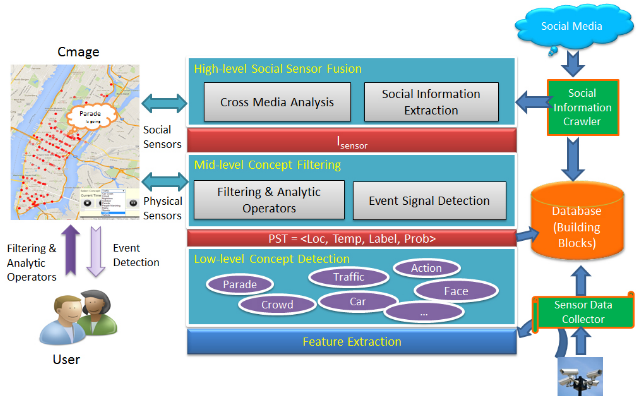


Figure 2: Multi-layer Tweeting Cameras Framework

data structure from multi-sensor streams is expressed in a unified format and easy to be fused in a way similar to image fusion. We conceive each "pixel" in an image presents an event or concept "signal" with semantic meaning, and consider semantic relatedness among signals when fusing situation images provided by particular sensor streams.

3.3 New Paradigm of Social Sensor Network

Establishing connections between people and the resources they need is a fundamental task in society, which leads to the problem of "Social Life Networks" (SLN) [7]. We imagine and propose a new paradigm where people are not only connected to people through social network, but also have access to customized services provided by all kinds of widely used physical sensors. Node pairs in these network are temporally, spatially, semantically or information-wise close to them. E.g., as in Figure 1, a sensor could directly tell a person a parade in going on at some specific place; a meeting has started or abnormal things are detected. An end user could specify their information needs from sensors by assigning customized tasks or share personal information to sensors making them learn to be smarter for future event detection or fusion tasks. Such a new sensor-involved social network would greatly improve our situation understanding and effectively satisfy people's information and resources needs through human sensor interaction and collaboration.

4. EXPERIMENT RESULTS

We have crawled live feeds from 150 public CCTV traffic cameras distributed on the roads all over the Manhattan district of New York City as well as Manhattan geotagged tweets from October 08, 2014 onwards, and also collected NUS campus foodcourts video dataset consisting of video feeds from 73 cameras located at 9 different foodcourts. We quantitatively evaluated proposed tweeting cameras framework in Figure 2 with this large-scale dataset.

4.1 Events Ground Truth

We use the notices posted on the "Weekend Traffic Advisory" website of the New York City Department of Transportation for obtaining the ground truth.[1] The ground truths for the events that we try to detect are shown in Table 1.

4.2 Applying Analytic Functions

We apply concept detectors on raw camera feeds and we defined a set of analytic functions and operators such as

[1]http://www.nyc.gov/html/dot/html/motorist/wkndtraf.shtml

Table 1: Real-world Events Ground Truth

Event	***Date***	***Time***	***Location***
CBGB Music Festival	12 Oct	10am-7pm	Broadway 51 Street
Hispanic Parade	12 Oct	12pm-5pm	5th Avenue
Columbus Day Parade	13 Oct	11am-5pm	5th Avenue
Saint Patrick's Day Parade	17 Mar	12pm-5pm	5th Avenue
Million March NYC Protest	13 Dec	2pm-5pm	Washington Square Park, 5th Avenue, Foley Square

"smooth" , "extreme", "trend" that can be applied to obtain meaningful information such as event pattern, concept trending etc. Results of "smooth" and "trend" function is shown in Figure 3 and Figure 4 which reduce sensor noise in "St Patrick's Data Parade" event data and describes university Foodcourt crowdedness trend respectively.

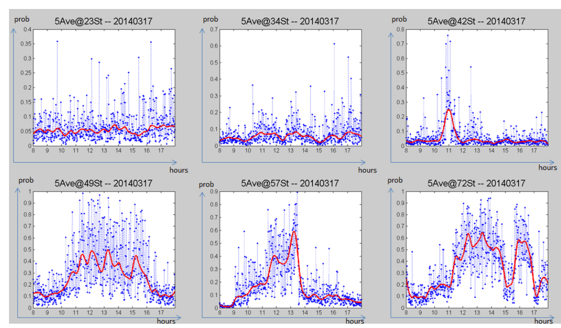


Figure 3: *People Marching* Concept Results from 8:00 to 18:00 in March 17th, during Saint Patrick's Day Parade Event

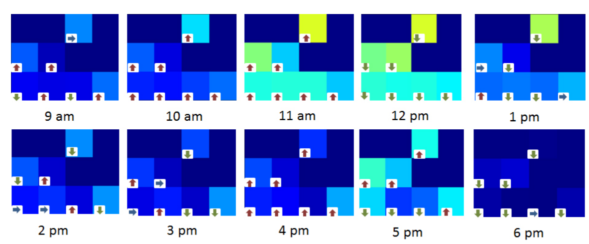
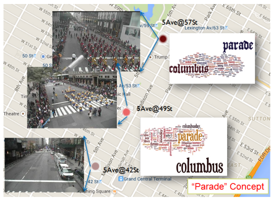


Figure 4: A Campus Foodcourt Cmage *Crowdness* with Arrows indicating Ascending, Descending and Plateau Trend.

4.3 Social Sensor Fusion

To fuse sensor and social information, we detect events spatial and temporal information from sensors and then generate related representative terms from social media by using tweets posted nearby the places where an event occurs. Results are shown in Figure 5. As can be seen, social tweets posted near an event location are able to give high-level semantic meaning of the event, which enhance the understanding of situations.

The results demonstrates the feasibility and effectiveness of our proposed framework and the idea of combining camera tweets (with spatial and temporal information) and social tweets (high level semantic meanings) are shown to be promising for detecting real-world events.



(a) Social Sensor Fusion for "Columbus Day Parade"

(b) Social Sensor Fusion for "Hispanic Parade" Event

(c) Crowd Concept and Salient Topic Words during "CBGB Musical Festival"

(d) Sensor and Social Information of "Million March NYC" Protest Event

Figure 5: Social Sensor Fusion for Real-world Events

5. CONCLUSIONS AND FUTURE WORK

We propose an new paradigm that extend traditional sensor sensing mechanism by combining sensor of different modality and socializing them with human to facilitate real-world event information fusion. The work includes a novel tweeting camera frame that combines both physical sensor and social sensor together for event detection, which is demonstrated through data analysis in real-world events, and a pseudo-image based social sensor fusion algorithm.

Future directions:

As we are solving multimodal fusion problem, our future works mainly direct in two folds: 1) In social sensor fusion, we will explore how social sensors information is diffusing spatially and temporally in social network and how this diffusion should improve physical sensors detection task. The reliability of social sensors for event detection and the system scalability issues will also be explored. 2) We will build a new social-cyber-physical paradigm where human and sensor are interacting in information sharing, fusion and recommendation. Besides spatial and temporal correlation, we will explore semantic correlation among event related signals from different sensing modalities to provide deeper understanding of going situation. Such spatio-temporal-semantic pattern mining using multi-modal and complementary sensors could offer more proper, prompt and accurate recommendation for action takings.

6. ACKNOWLEDGEMENT

This research was conducted at NUS-ZJU SeSaMe Centre, which is supported by the Singapore NRF under its IRC@SG Funding Initiative and administered by the IDMPO.

7. REFERENCES

[1] P. Atrey, M. Hossain, A. El Saddik, and M. Kankanhalli. Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems*, 16(6):345–379, 2010.

[2] L. Atzori, A. Iera, and G. Morabito. From "smart objects" to "social objects": The next evolutionary step of the internet of things. *Communications Magazine, IEEE*, 52(1):97–105, January 2014.

[3] C. A. Bhatt, P. K. Atrey, and M. S. Kankanhalli. A reward-and-punishment-based approach for concept detection using adaptive ontology rules. *ACM Transactions on Multimedia Computing, Communications and Applications*, 9(2):10:1–10:21, 2013.

[4] M. Bramberger, A. Doblander, A. Maier, B. Rinner, and H. Schwabach. Distributed embedded smart cameras for surveillance applications. *Computer*, 39(2):68–75, 2006.

[5] M. Demirbas, M. Bayir, C. Akcora, Y. Yilmaz, and H. Ferhatosmanoglu. Crowd-sourced sensing and collaboration using twitter. In *International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM)*, pages 1–9, 2010.

[6] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. *PAMI*, 35(8):1915–1929, 2013.

[7] A. Gupta and R. Jain. Social life networks: A multimedia problem? In *ACM MM*, pages 203–212, 2013.

[8] N. Jacobs, W. Burgin, N. Fridrich, A. Abrams, K. Miskell, B. H. Braswell, A. D. Richardson, and R. Pless. The global network of outdoor webcams: Properties and applications. In *ACM SIGSPATIAL*, pages 111–120, 2009.

[9] Y.-G. Jiang, J. Yang, C.-W. Ngo, and A. Hauptmann. Representations of keypoint-based semantic concept detection: A comprehensive study. *IEEE Transactions on Multimedia*, 12(1):42–53, 2010.

[10] M. Kranz, L. Roalter, and F. Michahelles. Things that twitter: Social networks and the internet of things. In *CIoT Workshop*, 2010.

[11] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu. Senseye: A multi-tier camera sensor network. In *ACM MM*, pages 229–238, 2005.

[12] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi. Crowd sensing of traffic anomalies based on human mobility and social media. In *ACM SIGSPATIAL*, pages 344–353, 2013.

[13] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *WWW*, pages 851–860, 2010.

[14] A. Sheth, A. Jadhav, P. Kapanipathi, C. Lu, H. Purohit, G. A. Smith, and W. Wang. 'twitris- a system for collective social intelligence'. Encyclopedia of Social Network Analysis and Mining, 2013.

[15] V. K. Singh, M. Gao, and R. Jain. Social pixels: Genesis and evaluation. In *ACM MM*, pages 481–490, 2010.

[16] M. Walther and M. Kaisser. Geo-spatial event detection in the twitter stream. In *European Conference on Advances in Information Retrieval*, pages 356–367, 2013.

[17] Y. Wang and M. S. Kankanhalli. Tweeting cameras for event detection. In *WWW*, pages 1231–1241, 2015.