Multimedia Surveillance in Event Detection: Crowd Analytics in Hajj

Layla Al-Salhie, Mona Al-Zuhair, and Areej Al-Wabil

Software Engineering Department, College of Computer and Information Sciences,
King Saud University, Riyadh, Saudi Arabia
aalwabil@ksu.edu.sa

Abstract. Multimedia surveillance systems have evolved in recent years to capture, process and analyze multimedia data coming from heterogeneous sensors in the context of the annual pilgrimage to Mecca in Hajj. Systems in these contexts are often designed to support decision making, such as responding to alerts triggered by sensors and incidents detected by surveillance systems as well as to provide useful information for monitoring and emergency-response teams concerned with health and public safety. Various tools and techniques from different fields such as operations research, computer vision, image and video processing, pattern recognition, and multimedia fusion have contributed to the proliferation of such systems in the context of Hajj. In this paper, a systematic review and synthesis of the representative works that have been done in the field of multimedia surveillance for event detection in Hajj is presented. Observations and reflections on these works are discussed in the context of Hajj rituals' distinctive characteristics, crowd-management challenges, and multimedia issues related to event detection and surveillance systems in Hajj.

Keywords: Multimedia, multimodal fusion, Hajj.

1 Introduction

Hajj is an annual event that takes place in Saudi Arabia, where millions of Muslims congregate in the sacred sites of Mecca. Risks of incidents related to crowd-management often rise during this annual event such as stampedes in over-crowded routes of synchronized flow during rituals and health-related incidents (e.g. sunstrokes). Saudi Arabia receives every year around three million pilgrims from all over the world to participate in the Hajj event [15]; and this number is expected to reach about 3.75 million pilgrims by 2019 [19]. Crowd management in Hajj is particularly important because of the unidirectional flow of crowd in most rituals which often breaks down into dynamic stop-and-go waves [13]. Crowd turbulence occurs in some synchronized rituals which have been linked to incidents involving injuries and fatalities in stampedes. In 2004, 251 pilgrims died under foot from a stampede that lasted about 27 minutes; and 363 pilgrims lost their lives in

A. Marcus (Ed.): DUXU 2014, Part II, LNCS 8518, pp. 383–392, 2014.

crowd-related incidents in 2006. Technologies have been increasingly utilized in the past decade for crowd management and event detection systems during Hajj. However, public health and safety issues remain a challenge and advances in multimedia can be leveraged for surveillance and event detection in Hajj.

In this paper we provide a systematic review of the representative works that have been conducted in the field of multimedia event detection in Hajj. Based on that, we provide number of observations on these works according to Hajj distinctive characteristics and according to the multimedia issues related to surveillance systems for event detection. We conclude with a discussion on future directions for multimedia surveillance and event detection in Hajj.

2 Crowd Analytics

Several challenging factors contribute to the complexity in managing crowds in Hajj events; namely, constraints in location and timing of rituals. In addition to the heterogeneous demographic profile of pilgrims who often vary across the spectrum in literacy and socio-economic backgrounds and in many cases, Hajj is often their first trip abroad [6]. One of the earliest efforts in Hajj safety control was in 2007 where the first real-time analytics system was installed in the Jamarat area to analyze the live video feeds coming from almost 100 closed-circuit television (CCTV) video cameras around the Jamarat bridge [13].

Information and communication technologies (ICT) have been considered for addressing Hajj-related challenges for crowd safety and surveillance as described in [11]. Crowd analytics have ranged from solutions that require deal with knowledge management, representation and manipulation of data in different multimedia forms (e.g. audio and video), to solutions that require high capacity for communications and data transmission such as augmented reality. The advances in multimedia technologies made it feasible to utilize different modalities and surveillance systems have evolved in recent years in efficiently acquiring, processing and analyze multimedia streams coming from multiple and heterogonous sensors in Hajj events. Multiple works have been proposed in this field such as the works in [1], [8], [15], [16], [2], [14] and [3]. These kinds of systems are often designed to detect abnormal patterns in crowds, incidents and public health issues that are triggered by sensors or field-initiated reporting such as systems in [1] and [8].

Various tools and techniques from different fields such as operations research, computer vision, image and video processing, image background removal, pattern recognition, and multimedia fusion have contributed to the proliferation of such systems in the context of Hajj. Large-scale multimedia surveillance systems are mainly systems consisting of number of spatially distributed and pre-installed video camera sensors that are connected to a central control station, where human operators (security personnel usually) remotely monitor the areas captured by the cameras [4]. Multimedia surveillance systems ' design considerations often involve issues related to cost, environmental conditions, optimal placement and combination of different

sensors. Another issue is the selection of the best units of the multimedia data and the best fusion method for improved accuracy in surveillance.

3 Multimedia Surveillance in Hajj

The purpose of Hajj multimedia surveillance varied in the representative works reviewed in this paper. Crowd density and behavior estimation was the main purpose of systems described in [1-2] and [15-16] to address the overcrowdings in Hajj rituals. Decision support systems are often embedded in the functionality of these surveillance systems such as in [15-16] and [2] which provided extra features of supporting decisions regarding the regulation of site access at crowded routes of "Nafra". The purpose of the work [8] was different since it was focusing on individual's monitoring which included counting and tracking individuals in Hajj sites. The work in [14] has another surveillance purpose which is the extraction of number of parameters such as speed, direction, distance between pedestrians, and the elapsed time taken to reach certain areas. The system was designed for calibrating crowd simulation models that will help in understanding patterns in Hajj crowds. Extracting speed parameters of individual pilgrims has also been shown to provide insights for detecting incidents because pedestrians tend to increase their speed when something abnormal occurs. Although pilgrims in crowded areas often tend to be in the "stop and go waves" flow, individuals have little control over changing their speed in synchronized rituals. Considering these flow patterns, the work in [3] examined walking themes in Hajj by providing a framework for group motion detection by exploiting related techniques in fluid flow mechanics.

3.1 Modalities

Video sequences were the main multimedia modality processed in Hajj surveillance literature. However, the types of the produced videos differ from one work to another. As we can see from works [1], [16], [15] and [2], infrared thermal video sequences were the most commonly used multimedia modality in Hajj multimedia surveillance systems; this is because thermal imaging depends on the amount of radiation (temperature) emitted by an object in that only warm objects (humans) stand out well against cooler backgrounds in thermal images. Therefore the produced thermal images allow sufficient information for allocating the target without the need of more complex recognition techniques. Furthermore, infrared thermography detection systems are not affected by the illumination conditions in that they produce similar accuracies day or night. Moreover, thermal imaging doesn't capture the shadow of a human and therefore it is better used with background removal techniques. However, the use of thermography has some issues related to climatic conditions that will be discussed later. The work in [8] utilized MPEG4 and H264 video sequences. MPEG4 video format supports intellectual property management and protection (IPMP) which is not a key concern in surveillance systems but on the other side it supports low-bandwidth (less than 1.5MBit/sec bitrate) which could work with bad network

coverage [29].The work in [14] used CCTV camera video sequences. Multimedia processing of videos acquired from CCTV Video cameras placed at the gates of the main mosque in Mecca was described in [14]. The work in [3] utilized CCD cameras' video sequences which provide videos of relatively high quality for the context of Hajj facilities. Audio-only media can be utilized for analyzing the crowd behavior; and this was applied in [8] by examining audio signals for detecting abnormal incidents and consequently determining the exact position of the location's coordinates.

3.2 Design Considerations in Hajj Multimedia Systems

Design considerations for Hajj multimedia surveillance systems are described in the following sections.

Time, date and place limitations. Hajj activities involve performing a sequence of tasks at specific times, within a period of five days and space constraints of a 4 square kilometer area [6]. This creates overcrowding challenges in synchronized activites that have been known to lead to incidents such as stampedes, bridge collapses and health-related incidents such as fainting. Multimedia studies in Hajj surveillance have addressed this challenge by investigating how technologies can assist in the estimation of crowd density and behavior (accelerating or decelerating) and also by using camera sensors that do not necessarily have to be close to the targets. Knowledge management of in real-time monitoring was also examined in these studies.

Rituals are performed in a specific unchangeable sequence. While Hajj rituals hinder response teams from changing the sequences within activities or re-routing and closing certain areas, systems have developed decision support systems (DSS) to aid in the decision making process regarding controlling the crowd in these contexts such as in the works reported in [16] and [15]. These systems provided a DSS embedded engine that aids in regulating site access by controlling the movement of crowds.

Climatic conditions in Mecca. Temperatures in Mecca often reaches 44 degrees Celsius in the summer and can rises close to fifty degrees. This climatic nature poses risks of health-related incidents such as sunstrokes; heat stress and also fires in cases of overheating. It also imposes some design considerations on the design of surveillance systems. The cameras used as sensors need to be heat-resistant in order to adapt to the extremely high day temperature and the acute sun rays. For the works that utilized infrared thermography, the use of thermography in the extremely hot climates may affect its accuracy. This is because in such conditions the temperature of the ground and some other solid objects would be relatively high and these systems would need to be re-calibrated to adapt to the climatic conditions. Using heat sensors in [8] also point to the need for configuring sensors to distinguish the normal and abnormal temperatures in this specific context of surveillance.

The heterogonous demographics of pilgrims. The heterogonous profiles of pilgrims affect Hajj scenarios on many aspects including the organization and orientation efforts. The works [16] and [15] focused on DSS for organizations managing the pilgrims from different regions of the world, by providing decision-support systems that suggest alternate decisions to regulate the movement of pilgrims in Nafra route. This flow is often organized by groups and this systems aims to respond with recommendations for these entities int his synchronized activity. These decisions were resulted by the integration of a fuzzy logic, operations research and decision support modules. Further, language variations among pilgrims also present challenges for audio surveillance in addition to patterns of movement or behavior since it may vary across cultures; for examples, video surveillance trained to automatically detect abnormal gestures or behaviors of individuals/crowds.

3.3 Technologies in Hajj Surveillance

Underlying infrastructure. The work in [8] relied on 4G connections as a connection infrastructure between some parts of the system. However, considering the relatively large number of pilgrims using this network in Hajj, the load on this 4G networks provided by different telecommunication companies would need to be considered to ensure adequate service for these systems to work in the context of Hajj. Further, the mountainous nature of Mecca would need to be considered as it relates to the infrastructure of telecommunications and quality of the signals for individual pilgrims' mobile devices if sensors are embedded as well as the transmitted streams of surveillance data via these channels of information.

Camera parameters and positioning. The system described in [14] installed the camera sensors at Gate no.1 of al-Masjid al-Haram. It was suggested to position scene cameras for at the gates by directing the lenses to the people entering the gate so that the dimensions are reduced to 2D while still being able to detect the pedestrian parameters such as speed and direction. The cameras they used were assumed to be on relatively high position above ground so as to capture the dimensions of people of varying heights. For all works reviewed, the cameras were assumed to be fixed - not movable – and over a fixed elevation above the pedestrians (10 meters in case of works in [1], [16], [15] and [2]. However, the camera position is often related to the number of cameras covering the area under surveillance. The work in [3] justified having a stationary camera in the case of detecting GOP in a crowded scene in that the flow of crowd must be visualized from certain height for the visibility of the spatial organization of the flow. Further, the variability of the images captured will also produce significant noise sources to the overall analysis. Therefore a stationary camera was used to minimize the noise factors related to scene-capture.

Optimal media selection and multimodal fusion techniques. Multimedia modality and fusion methods are key components in the design of surveillance systems; however it has been inadequately examined in the context of Hajj multimedia surveillance systems. In [5], the authors reported fusion methods - at the feature

level- for surveillance tracking and detection purposes: linear weighted fusion, Baysian inference, dynamic Baysian networks, neural networks and Kalman filter. These method in particular can be applied in the context of video and audio surveillance systems for Hajj in scenarios of usage described in the systems of [1], [8], and [14-16].

Sensors in surveillance. The type and allocation of sensors is critical for accuracy in surveillance systems. Although this was not directly addressed in many fo the Hajj surveillance research, some studies alluded to opportunities of examining these issues in more depth. The analysis results of archived videos in [8] can be utilized in selecting the critical places that require mores sensors. The collection of information from multiple synchronous cameras located far away from each other is essential in surveillance systems, and these features were proposed by in [21] but the context of this study varied from Hajj conditions and a closer examination of adaptation for contexts-of-use would be needed for Hajj surveillance.

As for the type of sensors, all the works reviewed utilized camera sensors with variability in the type of cameras (forward looking infrared FLIR) thermal cameras in works [1], [16], [15] and [2], IP cameras and range-finder cameras in [8], CCTV cameras in [14] and CCD cameras in work [3]. Further sensors were used in [8] such as heat sensors. Rangefinder camera sensors along with white laser illuminator were used in [8] to measure the pilgrim's distance and detect him/her. Internet protocol camera IP camera is a type of digital video camera commonly employed for surveillance. Unlike the analog CCTV cameras, this camera can send and receive data via a computer network and the Internet. Flexibility and distributed intelligence are advantages of using IP cameras. However, the high initial cost and high network bandwidth requirements would need to be considered. Cameras were used as sources of live stream data in [3] where CCD cameras provided sufficient accuracy, light sensitivity and low-noise images. CCDs are known to consume relatively more powerthan alternative sensors that can be considered in Hajj scenarios such as satellite image sensors, normal (camcorder) video sensors, lighting intensity sensors [10] and microphone sensors.

Live and passive monitoring. Most of the works reviewed provided real-time monitoring solutions for Hajj events. In [15], a cost-effective method for decision making purposes was utilized (i.e. matrix-based CWP). Passive video recording might be also beneficial for performing further analysis on the archived videos to gain insights into critical behavioral patterns for rituals in previous years and for building estimation models from known behaviors. The work in [8] provided a storage feature of video sequences for later analysis to improve services. Further, due to the annual occurrence of Hajj, large number of examples of incidents will be useful for archiving and analysis. In such passive recording contexts, the video quality can be processed in higher resolutions for facilitating more accurate analysis.

Group tracking and mobility directions. People in Hajj have different movement patterns according to their customs and social cultures; however, it is notable that

most pilgrims walk in organized groups. This movement pattern suggests opportunities for providing surveillances systems designed for group motion detection such as the work in [3]. The study in [18], although not in the context of Hajj, has focused on group tracking in crowded environments where groups split and merge during the tracking process. In [3], authors exploited the shared properties between high density crowd flow with the fluid flow in that they applied some techniques available in fluid mechanics to study the flow properties of group of people (GOP) moving in relatively large crowds as is the case in Hajj. One of these techniques applied was the Lagrangian Coherent Structure LCS. The basic inspiration behind using LCS on large crowd flow scenario was to treat a flow as a single entity globally and to track the unstable groups inside the flow. Movement directions is also a concern that should be considered when developing people detection and tracking systems and also when placing different sensors. However, movement direction is different according to the area in Hajj rituals in that it is multi-directional inside Al-Haram (e.g. rotations around the Kaaba, bi-directional in Safa and Marwa), two directions at the Masjid Alharam gates, and often uni-directional at Nafra and al-Jamarat routes. This movement parameter should be considered when developing Hajj event and person detection systems. As for crowd estimation, the multiple directions in movement are often not a critical issue in estimation models. The speed of the crowd has been examined in one Hajj surveillance study in [14].

Shadows and high occlusions in surveillance. Shadows of people in video surveillance has been addressed in [2] where a background removal technique was applied with thermography instead of normal camcorder cameras. They justified their choice in that when using background removal techniques with normal video camera which captures shadows, the shadow effect could not be removed efficiently from the image. High occlusion among pilgrims is expected in Hajj contexts due to the overcrowding in the synchronized events. This issue impacts the detection and tracking efficiency and was not addressed in the works [1], [16], [15] and [2] because these studies examined crowd estimation rather than detection. In contrast, the work in [14] addressed the high occlusion as a major issue from computer vision lenses of investigation and successfully used the active contours technique [7] to handle occluded persons. This technique achieves successful segmentation by evolving a contour near the boundary of the object. In the related work [20], the system showed good results for handling the occlusion and illumination changes by using single view camera. However, this approach presents its own challenges for the Hajj context where there are large crowded areas that need multiple cameras' coverage for efficient multimedia surveillance.

Predefined locations in surveillance. Considering that Hajj is performed at predefined fixed places, this can help in fixing the background elements of the surveillance media produced. Therefore, the secondary background removal method that was mentioned in work [2] can be considered for Hajj scenarios because in that model, the value of the background in the previous frame can be considered as the background in the current frame. Another method in the related work [12] may also

be applied since it assumed a permanent static background. However, when we consider the reengineering works applied usually in Hajj areas, the applicability of those methods will be debatable in that there could be changes in the background images. Furthermore, when we have a moving object that becomes part of the seen (e.g. health-care vehicles and mobile clinics), the two methods above will present computational challenges in the background removal aspect of their multimedia processing. This is because both methods assume a static background to be removed during scenes' processing; therefore their efficiency will decrease when the background elements are changing.

Multiple postures and visual appearance of pilgrims. The study in [14] applied the silhouette tracking technique for handling large variety of human poses. In Hajj, pilgrims adhere to some constraints in clothing in some rituals (e.g. omra) and similarity in color and garment flow can be considered when designing a computer vision application that analyzes images in monitoring and detection systems, counting and the crowd estimation algorithms. The issue of low contrast of clothing colors of different pedestrians, and low contrast of clothes with the whitish color of the ground marble was highlighted in [14]. However, the issue of contrast and color-based detection was not addressed. Further, the same work tested a tracking method CAMSHIFT to track specific objects in Hajj event. The method failed because of two reasons; there was high occlusion of the object being tracked and there were many objects with similar color features appearing in the same frame.

4 Conclusion

In this paper we presented a review of studies that examined multimedia surveillance and event detection in Hajj. Observations and reflections on multimedia design and multimodal processing, and how they relate to Hajj contexts, were described in the design considerations' sections.

From the experimental sections of the reviewed systems for multimedia surveillance, it seems that the issue of having relatively overcrowded contexts as in Hajj (e.g. having more than three people in one square meter) was inadequately examined. According to [22], 2.14 persons is the maximum number of comfortably moving persons per square meter. This observation suggests that more work needs to examine the applicability of the proposed methods for crowd situations in these contexts. The work in [14] suggested that there is a lack of ground truth data for performing experiments on the surveillance systems proposed for event detection in Hajj. Examples on ground truth data for such cases as in [14] is to have a short sequence of the video samples containing the pedestrians at specific areas with bounding boxes drawn around each pedestrian. Moreover, cultural models of behavior, such as Hofstede's cultural dimensions model [9], can be utilized to understand the movement nature of pilgrims from different cultures and then map these insights to parameters in the design of Hajj event detection systems.

Acknowledgement. The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for partially funding the work through the research group project number RGP-VPP-157.

References

1. Abuarafah, A.G., et al.: Real-time Crowd Monitoring using Infrared Thermal Video Sequences. Journal of American Science 8(3), 133–140 (2012)
2. Abuarafah, A.G., Khozium, M.O.: Integration of Background Removal and Thermography Techniques for Crowd Density Scrutinizing
3. Arshad, M.R., Soori, U.: Group Tracking Algorithm for Crowded Scene
4. Atrey, P.K., et al.: Effective multimedia surveillance using a human-centric approach. Multimedia Tools and Applications 51(2), 697–721 (2011)
5. Atrey, P.K., et al.: Multimodal fusion for multimedia analysis: a survey. Multimedia Systems 16(6), 345–379 (2010)
6. Bahurmoz, A.M.A.: A strategic model for safety during the Hajj pilgrimage: An ANP application. J. Syst. Sci. Syst. Eng. 15(2), 201–216 (2006)
7. Caselles, V., et al.: A geometric model for active contours in image processing. Numerische Mathematik 66(1), 1–31 (1993)
8. Fardoun, H.M., et al.: Mecca Access and Security Control System. In: Proceedings of the 13th International Conference on Interacci & Oacute; N Persona-Ordenador, pp. 30:1–30:7. ACM, New York (2012)
9. Geert, H., Jan, H.G.: Cultures and organizations: Software of the mind. McGraw-Hill, New York (1991)
10. Al-Habaibeh, A., et al.: The design of a novel sensor fusion model for monitoring people's density in public places using infrared thermography. Key Engineering Materials 419, 377–380 (2010)
11. Hameed, S.A.: ICT to serve Hajj: Analytical study. In: 2010 International Conference on Computer and Communication Engineering (ICCCE), pp. 1–7. IEEE (2010)
12. Horprasert, T., et al.: A robust background subtraction and shadow detection. In: Proc. ACCV, pp. 983–988 (2000)
13. Johansson, A.: Crowd control in mass gatherings. In: The Lancet Conference of Mass Gathering Medicine, Implications and Opportunities for Global Health Security, Jeddah, Saudi Arabia (2010)
14. Al-Khaffaf, H.S.M., Haron, F., Sarmady, S., Talib, A.Z., Abu-Sulyman, I.M.: Crowd Parameter Extraction from Video at the Main Gates of Masjid al-Haram. In: Sambath, S., Zhu, E., et al. (eds.) Frontiers in Computer Education. AISC, vol. 133, pp. 727–736. Springer, Heidelberg (2012)
15. Khozium, M.: A Hybrid Intelligent Information System for the Administration of Massive Mass of Hajjis. Life Science Journal 9(4), 171–180 (2012)
16. Khozium, M.O., et al.: A Proposed Computer-Based System Architecture for Crowd Management of Pilgrims using Thermography. Life Science Journal 9(2), 377–383 (2012)
17. Lau, B., et al.: Tracking groups of people with a multi-model hypothesis tracker. In: IEEE International Conference on Robotics and Automation, ICRA 2009, pp. 3180–3185. IEEE (2009)
18. Ben-Mahmoud, C., et al.: Holy Cities: Saudi's Unique Real Estate Markets. In: on. point Jones Lang LaSalle IP, Inc. (2010)

19. Mirabi, M., Javadi, S.: People tracking in outdoor environment using Kalman filter. In: 2012 Third International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 303–307. IEEE (2012)
20. Mittal, A., Davis, L.S.: M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 18–33. Springer, Heidelberg (2002)
21. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 747–757 (2000)