Visual Surveillance for Hajj and Umrah: A Review

Yaser Othman

Umm Al-Qura University

Abstract

This paper presents advances on crowd management research with specific interest on high density crowds such as Hajj and Umrah crowds. In the past few years, there has been increasing interest in pursuing video analytics and visual surveillance to improve the security and safety of pilgrimages during their stay in Makkah. Most works published in these aspects addressed topics ranging from people counting, density estimation, people tracking and modeling of motion and behaviors. Despite the fact that visual surveillance research has matured significantly in the rest of the world and had been implemented in many scenarios, research on visual surveillance for Hajj and Umrah application still remains at its early stages and there are many issues that need to be addressed in future research. This is mainly because Hajj is a very unique event that shows the clustering of millions of people in small area where most advanced image processing and computer vision algorithms fail to generate accurate analysis of the image content. There is a strong need to develop new algorithms specifically tailored for Hajj and Umrah applications. This review aims to give attentions to these interesting future research areas based on analysis of current visual surveillance research. The review also pinpoint to pioneer techniques on visual surveillance in general that can be customized to Hajj and Umrah applications.

INTRODUCTION

Visual Surveillance

Visual surveillance is one of the important tools for improving public safety and security in urban areas. All major cities in the world have begun installing CCTV cameras in public areas and sensitive areas for preventing and predicting possible crimes and accidents. Moreover and due to the availability of cheap and ubiquitous surveillance camera, these cameras have been installed in shops, hotels and even small outlets [1]. Effective visual surveillance system is one of the key components for cities to be ready for major world events such a religious gathering (Hajj), sport events such as World Cup and Olympic Games as well as political and business gatherings (demonstrations, conferences etc). All cities hosting major world events proudly declare the sophistication of visual surveillance systems they implemented such as London which is the world most surveillance city and Vancouver winter Olympics [2].

Hajj Security

The city of Mecca, home for Al-masjid Al-Haram is prayer face of Muslims and to which millions of Muslims assemble at the end of every Muslim's lunar year for the Hajj. In short period, the Holy city of Mecca faces more than three times its usual capacity which poses serious security, safety and health challenges to the authorities of the Kingdom of Saudi Arabia. Hajj contains several rituals that are performed in Al-masjid Al-Haram and the holy sites (Menna, Muzadlifa and Arafat) [3]. Figure ١ shows sample images for Hajj captured at different locations, the first row shows images captured for Tawaf which is circulating around the Kabba. The second row shows images captured from al Jamarat which is the place for stoning the Devil. The third row shows images captured in Saffa and Marwa. The forth row is mount Arafat and the last row shows the tents of Menna where pilgrims stay there for three days for stoning the Devil in the Jamarat place [4].

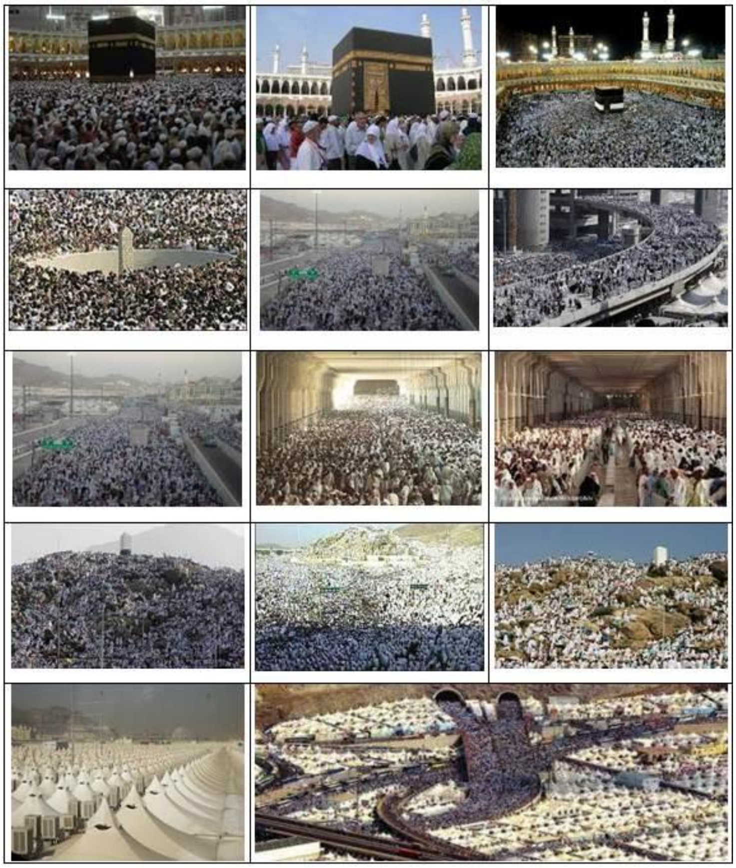

Figure 1: Crowded areas at different parts of Hajj rituals

The remaining of this paper is organized as follows; section 2 present research progresses on people counting for Hajj and its applications. Finally section 3 presents a comprehensive conclusion to this survey and it shows future research direction of visual surveillance for Hajj applications.

PEOPLE COUNTING

This section covers some of the published works on people counting from images captured using surveillance cameras. The section starts by scanning the published research articles in this matter. Then it provides a detailed discussion and analysis to these works and finally it points out what methods among these are suitable for Hajj and Umrah applications and how they can be approached.

2.1 Related Works

Table 8. Comparison of research articles published on people counting

Paper	Algorithm used	Advantages	Disadvantages	Accuracy
[5]	Used Bayesian model for image segmentation	-Bayesian model can learn complex image features	Not good for high density crowd	N/A
[6]	- Used blob extraction from sequence of images - Counting using trained model with neural networks	-Removing shadow and background	-Requires training phase and training data -Assumed maximum of 30 pedestrians per image	80%
[7]	-Passing people counting using overhead stereo camera	- No issues of occlusion - Gives count as well as the direction of movement	- Limited testing was performed - Narrow field of view for cameras	N/A
[8]	-Used specialized IR sensor for detecting and counting humans	- Fast processing - Accurate for low density crowd	- Narrow field of view - Not suitable for large crowd	N/A
[9]	- Used blobs extraction from sequence of image with a known background image	- Joint estimation of density and count	- Not suitable for high density crowd - Errors due to occlusions	85%
[10]	- Used a camera mounting on a moving car to detect and count crowd	- Simple method to detect a moving person -Can distinguish between vehicle and human	- It assumes movement of the camera - It fails to detect crowd moving in undetermined direction	N/A
[12]	-Used Haar wavelets to detect head-like features and filter it using SVM	-People counting form single image	-Requires training phase -Verification was done with human likes puppet	Above 90%

	classification - Apply perspective correction		not real crowd scenarios	
[13]	-Use background differencing to detect people -Use foreground ratio in small blocks are recorded for small moving window	-Radial Basses Functions (RPF) features learn good model for filtering out false blobs	-Uses a sequence of only 7 frames to do neural network classification of detected blobs -Requires training phase	89%
[14]	-Use a group of sensors to extract the foreground image -Used neural network with the extracted silhouette to project the visual hull of the scene	-Real time counting performance	-Using multiple sensors induces high cost -Cameras calibration overhead -Testing was done with limited data	N/A
[15]	-Fusion of IR with visual camera to detect and count people	-IR can work in total darkness	-IR images does not provide sharp edges for body silhouette	N/A
[16]	-Employed histogram filter to extract human sized blobs from foreground image	-Ultra low computations been implemented on Imote2 sensor node -histogram is robust to intensity fluctuations	-suitable for counting few peoples only	N/A
[17]	-Employed median filtering for selecting the background -Genetic algorithm was used for selecting foreground threshold and blob size	-The algorithm has been developed for real crowd scenarios with thousands of peoples	-Limited training data was used for genetic algorithm training -Not robust to illumination variations	N/A
[18]	Used local features of the object blob such as (area and perimeter) with camera calibration as prior step	-Invariant to the scene by taking knowledge of the camera position with respect to the scene (scene invariant) -Applied perspective correction to the image	-Requires camera calibration -Requires a training step using annotated set of data -It relies on accurately detecting the human in the image	N/A

Discussion and Summary

This previous literature reviews showed rich and diverse attempts to people counting from images that employs different computer vision and image processing algorithms. Table 1 summarizes the previous listed works. The accuracy report is based on what was reported on the paper with their dataset. This mean the accuracies are not comparable with each other across different works as some used simple data while other used high density crowd images. Simple people counting approach where performed by subtracted a known or trained background of the scene from each new frame and then counting the number of valid blobs in the foreground image [9]. This is only viable in low density crowd where all people are clearly visible to the camera and they can easily be distinguished from the background of the scene. Some worked tried tuning the background removal and blobs filtering stages in order to get accurate count by using genetic algorithm optimization [16] and histogram filters [15]. Another works performed people counting at gates using the concept of virtual gates with overhead cameras [7] or with specialized IR cameras [8]. Some researcher had proposed preprocessing steps to improve the counting such as [6] which removed shadow and [5] which presented Bayesian estimators for image segmentation.

Another class of method learned the counting of crowd from low level image features. The motivation of these works was the difficulty in detecting the presence of people in high density crowds due to severe occlusion [11]. Image features could be in form of texture or color histogram and they have learned it using regression method such as support vectors regression or linear regression. Such algorithms are mostly common for computing the crowd density but they can also be employed for crowd counting application. New research trends on people counting for Al-masjid Al-Haram should use image features instead of detecting people because of the large number of people in one image. These image features can be frequency properties of textures or color distribution or interest point detectors such as HOG or SIFT or other low level or high level image features that can be combined with machine learning to produce accurate count. In addition to that, using this kind of algorithms should keep in mind that the density is not uniformly distributed all over the image as some parts of the image tends to be with no people due to barriers. Interest point detector can false detect people in these areas which produce wrong count. To overcome those local features can be processed in small and overlapping image blocks with associate confidence level of each block that can be later aggregated to produce the final count [35].

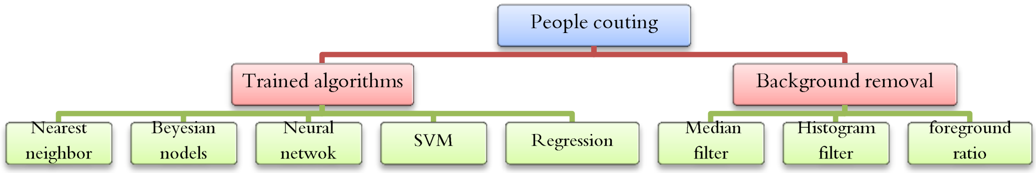

Figure 5. Hierarchical classification of techniques and methods used for people counting from images.

CONCLUSION

This paper surveyed research work on visual surveillance with focus on surveillance of Hajj and Umrah. The paper also addressed papers about dense crowd surveillance to expand the content as the number of research articles published on Hajj and Umrah surveillance is not sufficient enough. Most of the published work on people counting relied on detecting the object blob then counting it which is not appropriate for large crowds. Recent trends extracted local image features and directly related them to the crowd counting using machine learning tools. However these techniques have not been implemented on people counting in Hajj and Umrah and this direction needs to be pursued further to adopt these methods to Hajj and Umrah crowd counting.

REFERENCES

[1] "UK has 1% of world's population but 20% of its CCTV cameras," The Daily Mail News paper, 2007.

[2] C. J. Bennett and K. D. Haggerty, "Security Games: Surveilance and control at mega-events," University of Alberta, pp. 1–36, 2010.

[3] S. Sarmady, F. Haron, A. Zawawi, and A. Z. Talib, "A cellular automata model for circular movements of pedestrians during Tawaf," Simulation Modelling Practice and Theory, vol. 19, no. 3, pp. 969–985, Mar. 2011.

[4] T. Tawaf, "Agent-based Simulation of Crowd at the Tawaf Area," pp. 129–136.

[5] M. E. Leventon and W. T. Freeman, "Bayesian Estimation of 3-D Human Motion," 1998.

[6] S. Yoshinaga, A. Shimada, and R. Taniguchi, "Real-time people counting using blob descriptor," Procedia Social and Behavioral Sciences, vol. 2, pp. 143–152, 2010.

[7] K. Terada, D. Yoshida, S. Oe, and J. Yamaguchi, "A method of counting the passing people by using the stereo images," in International conference on image processing, 1999, pp. 0–7803–5476–2.

[8] S. Hamshimoto, K. Morinaka, K. Yoshiike, N. Kawaguchi, and C. Matsueda, "People count system using multi-sensing application."

[9]D. Roqueiro and V. A. Petrushin, "Counting People using Video Cameras," 2006.

[10]P. Reisman, O. Mano, S. Avidan, and A. Shashua, "Crowd detection in video sequences," in International Symposium on Intelligent Vehicles, 2004, pp. 66–71.

[11]A. Marana, L. Costa, R. Lotufo, and S. Velasin, "On the efficacy of texture analysis for crowd monitoring," in International Conference on Proc. Computer Graphics, Image Processing, and Vision, 1998, pp. 354–361.

[12]D. Yan, H. Gonzales, and L. Guibas, "Counting people in crowds with a real-time network of simple image sensors," in 9th IEEE International Conference on Computer Vision, 2003, pp. 122–129.

[13]L. Xiaohua, S. Lansun, and L. Huanqin, "Estimation of Crowd Density Based on Wavelet and Support Vector Machine," Transactions of the Institute of Measurement and Control, vol. August, pp. 299–308, 2006.
[14]J. Andersson, M. Rydell, and J. Ahlerg, "Estimation of crowd behavior using sensor networks and sensor fusion," in 12th International Conference on Information Fusion, 2009, pp. 396–403.
[15]T. Teixeira and A. Savvides, "Lightweight People Counting and Localizing for Easily Deployable Indoors WSNs," vol. 2, no. 4, pp. 493–502, 2008.
[16]M. Arif, S. Daud, and S. Basalamah, "Counting of People in the Extremely Dense Crowd using Genetic Algorithm and Blobs Counting," IAES International Journal of Artificial Intelligence, vol. 1, no. 1, pp. 1–8, 2012.
[17]A. G. Abuarafah and M. O. Khozium, "Integration of background removal and thermography techniques for crowd density scrutinizing," International Journal of Computing Academic Research, vol. 2, no. 1, pp. 14–25, 2013.
[18]H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-Source Multi-Scale Counting in Extremely Dense Crowd Images," in IEEE International Conference on Computer Vision and Pattern Recognition, 2013.