Counting People in a Crowd Using Viola-Jones Algorithm

A. Alharbi, A. Aloufi, E. Hamawi, F. Alqazlan, S. Babaeer and F. Haron

***Abstract*—** Counting the number of people in crowded areas such as the Masa'a of Al-Masjid Al-Haram has become a necessity. This is due to the increasing number of people performing Hajj or Umrah. In this paper, we propose a crowd counting system for the Masa'a using image processing techniques. The system detects the number of faces using the Viola-Jones detector then count the number of detected objects. The results show that the accuracy of the algorithm depends on the crowd density level, which are 77%, 82% and 79%, for low, medium and high density, respectively. The relatively low accuracy was due to the women wearing *niqab*, the face cover (usually black) since it is difficult to distinguish between a covered face from the back of the head. The system keeps the count results (i.e. total people heading to Safa or Marwah) in a database and presents the stored information in a histogram so that the authority is able to study the pattern of crowd flow and in case of overflow the system shows an alert window to warn the authorities of this dangerous situation in order to effectively manage the crowd at the Masa'a.

Index Terms: Viola-Jones; Crowd counting; Crowd management.

I. Introduction

Millions of Muslims visit Al-Masjid Al-Haram each year to perform different kind of religious rituals like Hajj and Umrah, and this can sometimes lead to overcrowding. It is a life-threatening situation and if it is not suitably handled, it may lead to injuries or sometimes death. We can take what happened in Mina as an example; people were killed in a stampede while undertaking the Hajj pilgrimage. This incident has sadly affected more than 1000 people. In view of the fact that the number of visitors to Al-Masjid Al-Haram is increasing each year, a crowd control and monitoring technique is required.

Several approaches for counting the number of people have been proposed, for example skeleton-based head detection as Merad et al used in [1]. In addition to that, Viewpoint Invariant Training (VIT) feature normalization method is proposed in [2] to account for distortion introduced by different camera orientation.

We based our work on a recent study by Soomro et al. [3], which uses Viola-Jones object detection framework. They detected the heads of the people in a crowd at Mochit light rail station in Bangkok, Thailand and then counted the people that are present in the image. The difference between Soomro et al's [3] and ours is that they trained the Viola-Jones detector to detect heads, but we used the Viola-Jones detector to detect faces. Arif et al [4] worked on people counting but the images are from the Mataf at Al-Masjid Al-Haram. We have used a data set that has not been used before in all the previous work. That is the videos of people doing Sa'i at the Masa'a in Al-Masjid Al-Haram.

Manuscript received November 12, 2016.

F. Haron was with School of Computer Sciences, University Sains Malaysia. She is now with the Department of Computer Science, College of Engineering and Computer.

F. Alqazlan graduated from Department of Computer Science, College of Engineering and Computer Science, Taibah University

II. Methodology

The system will help the authorities in managing the crowd in the Masa'a of Al-Masjid Al-Haram to avoid major injuries and deaths. It will count the number of people in Masa'a using face detection. The system will keep a simple query that contains data such as date, time and total count at some time interval of the video, by doing so we are converting the video of human flow to some statistical representation that can be used by the authorities.

Based on our system requirement we identify the modules of our system architecture (Fig. 1) as follow:

Detection module: Using Viola-Jones algorithm to detect the faces in each frame, the algorithm consist of the following stages:

* Haar like features: Determines adjacent rectangular regions in specific place in an image and then subtracts the sum of black pixels from the white pixels then uses the result to specify some subsections of an image.

* Integral images: Calculates the area by summing up every pixel.

* Adaboost: Filters the features and chooses only the relevant ones via forming a strong classifier from multiple weak classifiers.

* Cascading classifier: Cascades the features and decides whether the image is a head/face image or non-head/non-face image.

Counter module: Counts the bounding boxes extracted from the detection module and produces the total count of people passing through Safa and Marwah.

Store and display module: A database that contains some data such as date, hour, total people passing through Safa and Marwah per hour, day, month and year.

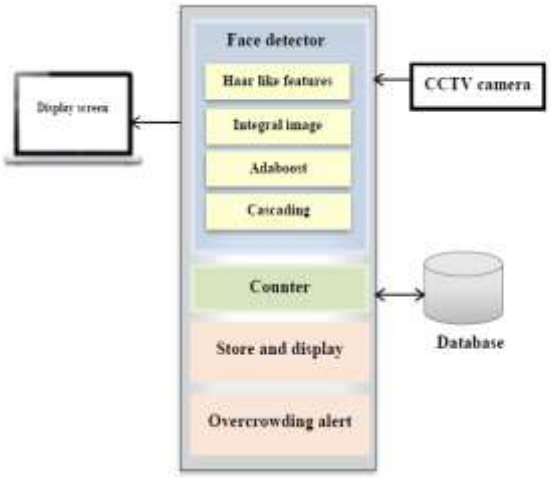

Fig. 1. System architecture and the main system components.

III. Results/Discussion

This section describes the result and the accuracy of our system. We trained a total of 4000 faces. To train the detector the user can load multiple images at one time and draw a rectangle on the region of interest and then export the region of interest information in the appropriate format for the TrainCascadeObjectDetector. We tested three videos that was provided for us by TCMcore. The videos cover both Safa and Marwah and the camera was facing the people. The first video featured a low dense crowd, the second one was with medium density crowd, the last one was with a high dense crowd, and we calculated the accuracy based on the following formula:

$$\frac{true\ positive + true\ negative}{true\ positive + true\ negative + false\ positive + false\ negative} \times 100 \quad (1)$$

The following table presents the number of people that our system detected and counted with the actual number of people in the video, in addition to the true positive (i.e. the correct detection):

TABLE 1
ACCURACY OF OUR SYSTEM

Density level	Frame number	Number of people detected	Actual	Difference	True positive
Video A (Low density crowd)	1	16	14	4	13
	2	18	20	4	16
	3	17	14	1	13
	4	3	3	0	3
	5	10	16	7	9
Total		64	67	16	54
Accuracy				77%	
Video B (Medium density crowd)	1	74	68	7	61
	2	73	76	10	66
	3	85	86	12	74
	4	81	83	5	78
	5	113	106	7	99
	6	69	70	8	62
Total		495	489	49	440
Accuracy				82%	
Video C (High density crowd)	1	162	165	24	1
	2	187	177	12	165
	3	130	144	32	112
	4	77	86	18	68
	5	96	105	21	84
	6	85	91	15	76
Total		773	768	122	646
Accuracy				79%	

As shown in the table the accuracy of our system differs depending on the density level and if the number of images used for training is increased, this approach will show more efficient result. There is exceptional cases where women who cover their faces with *niqab* will not be detected and the reason is Viola-Jones detector depends on feature extraction and if the face is covered the features will not be detected.

Soomro et al [3] used Viola-Jones detector and has an accuracy of 93.877%, and our system produced an accuracy of 79%. Soomro et al.'s result is higher than ours because the samples they used is a data set from the Mochit light rail station in Bangkok, Thailand, whereas our dataset was on Masa'a and it included women covering their faces, and since our detector depends on feature extraction it cannot detect covered faces so the accuracy decreases accordingly.

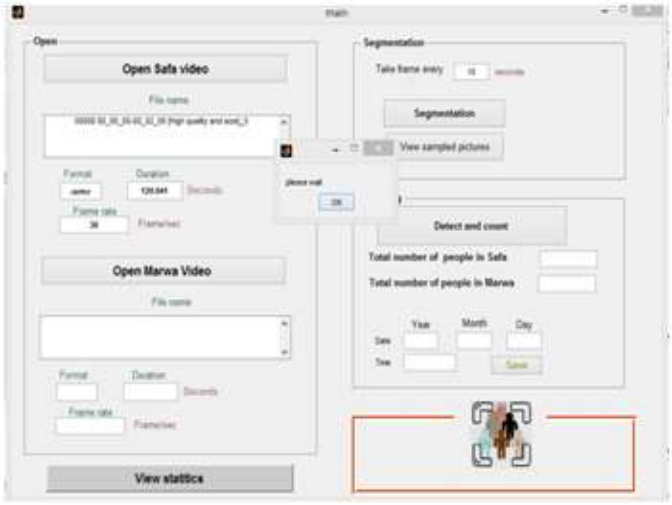

Fig. 2. The main page of the system

Fig. 3. Sample of a video being

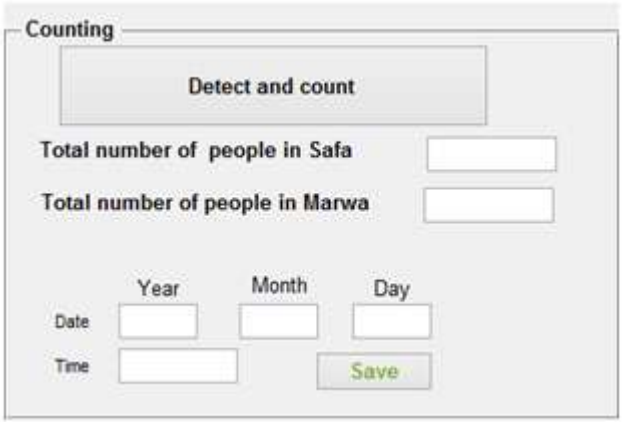

Fig. 4. Starting the count/saving the data.

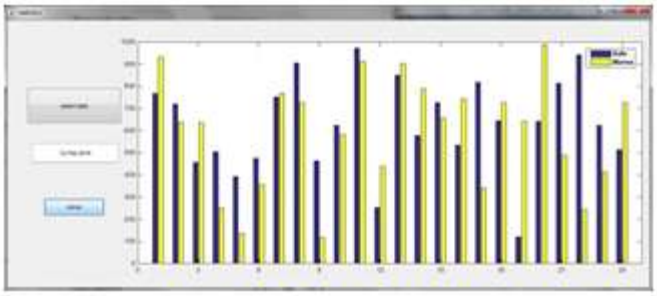

Fig. 5. Displaying the histogram of the count result.

Fig. 6. Image before face detection.

Fig. 7. Image after face detection

IV. Conclusion

Our work is about introducing a system that uses image processing techniques to solve the problem of crowd management in the Haram by counting the number of people passing through Safa and Marwah. We have proposed a crowd counting system at the Masa'a, which counts the number of people passing through Safa and Marwah. The system detect the faces at every interval using Viola-Jones detector. To the best of our knowledge, our work is the first published work on counting at the Masa'a. The accuracy of the detection is 77% for low dense, 82% for medium dense and 79% for high density crowd of the Masa'a. We highlighted that the accuracy is very much affected by the number of women wearing the niqab, hence reducing the number of detected faces. Our future work includes addressing this issue.

Acknowledgment

We would like to thank Dr. Lili Nurliyana for her help in this project and the TCMcore, Umm-Alqura University, for providing us with the video of the Masa'a.

References

[1] D. Merad, K. E. Aziz and N. Thome, "Fast People Counting Using Head Detection from Skeleton Graph," Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, Boston, MA, 2010, pp. 233-240. https://doi.org/10.1109/avss.2010.77

[2] Kong, D., Gray, D. and Tao, H. (2005), Counting Pedestrians in Crowds Using Viewpoint Invariant Training. In: Proc. British Machine Vision Conference (BMVC).Oxford: Citeseer, pp. 8. https://doi.org/10.5244/c.19.63

[3] Soomro, P., Memon, U. and Memon, S. (2014). Human Detection and Counting in Crowded Scene. Asian Journal of Engineering, Sciences & Technology, 4 (1), pp. 18-22.

[4] Arif, M., Daud, S. and Basalamah, S. (2013). Counting of People in the Extremely Dense Crowd using Genetic Algorithm and Blobs Counting. IJ-AI, 2(2). pp. 4-6 https://doi.org/10.11591/ij-ai.v2i2.1793

Afrah Alharbi, Alaa Alaufi, Esraa Hamawi, Fatma Alqazlan and Samah Babaeer received their B.Sc. (in computer Science) on 2016 from Taibah University, Madinah, Kingdom of Saudi Arabia. Their Graduation project is mainly about image processing using viola-Jones at the Masa'a in Al-Masjid Al-Haram.

Fazilah Haron received her B.Sc. (in Computer Science) from the University of Wisconsin-Madison, U.S.A. and her Ph.D. (in Parallel Computing) from the University of Leeds, U.K. She is currently an Associate Professor at the Department of Computer Sciences, Taibah University, Madinah, Saudi Arabia. She previously served at the School of Computer Science, University Sains Malaysia, Penang. Her research interests include modeling and simulation of crowd, parallel processing, image processing, bioinformatics and cloud computing. She has published more than 15 papers specifically, on the research related to crowd at the Masjid Al-Haram and Masjid An-Nabawi.