Crowd Analysis of Almasjid Alnabawi using convolutional neural networks of CCTV footage

Mohammed J. Abdulaal
Department of Electrical and Computer Engineering
King Abdulaziz University
Jeddah, Saudi Arabia
mjabdulaal@kau.edu.sa

2020 8th International Conference on Intelligent and Advanced Systems (ICIAS) | 978-1-7281-7666-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIAS49414.2021.9642651

***Abstract*—In recent months, crowd management has become more important than ever, given the spread of contagious diseases such as COVID-19. The Hajj, in Saudi Arabia, is one of the largest gatherings in the world; it happens annually and is getting bigger every year. The development of radio-frequency identification (RFID) and mobile apps has been investigated to help estimate crowd movements in and among the holy sites. However, network-based technologies require large infrastructures and are therefore very costly. In this paper, a system is proposed to use existing closed-circuit television (CCTV) to accurately visualize the movements of crowds in the Almasjid Alnabawi, also known as The Prophet's Mosque. The proposed neural network is trained with large datasets of crowd images to produce estimates of the number of pilgrims in an image. Images are then integrated to produce crowd level models throughout the building. The system has been tested on two instances and showed high performance.**

***Index Terms*—Almasjid Alnabawi, Hajj, crowd management, CNN, CCTV, AI**

I. Introduction

We face crowding in many aspects of our lives. Events across the globe require enormous management efforts to avoid accidents associated with over-crowding, such as stampedes [1]. One of the largest gatherings in the world is the annual Hajj, in Saudi Arabia. In 2015, more than 700 pilgrims died in the Mina Valley stampede during the Hajj [2]. Another accident happened in 2006, when 363 people perished at the eastern entrance to the Jamarat Bridge, due to overcrowding [3]. As a result, The Ministry of Municipal and Rural Affairs of the Kingdom of Saudi Arabia initiated programs to monitor pilgrim flows for improved routes, schedules, and accident predictions at the holy sites [1]. Another concern associated with crowds is the spread of viruses and contagious diseases [4]. The most recent COVID-19 pandemic required dramatic measures to mitigate the spread among pilgrims during the 2020 Hajj. The government of Saudi Arabia significantly reduced the number of pilgrims allowed to participate, as a measure for reducing the spread of COVID-19 [5]. Other measures included enforcing social distancing during the performance of Tawaf around the Kaaba, where pilgrims circle around the black building at the center, resulting in the most iconic image of enforced precautionary measures, as shown in Fig 1.

King Abdulaziz University funding under grant no. (J: 11-135-1441).

978-1-7281-7666-6/18/$31.00 ©2021 IEEE

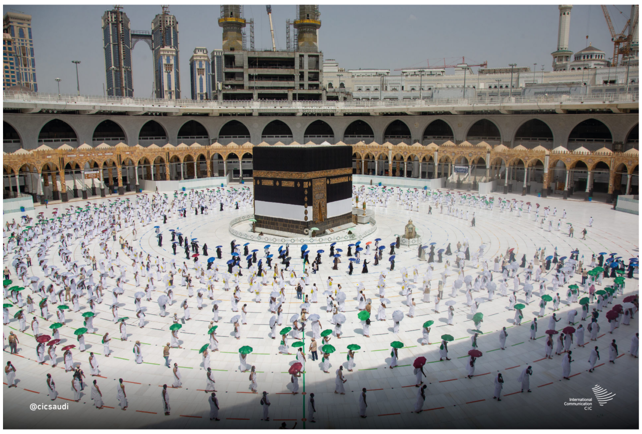

Fig. 1. Social distancing in Tawaf during the COVID-19 pandemic [5].

Given that millions of pilgrims visit Mecca and Medina every year, it is important to explore more smart technologies, such as artificial intelligence (AI), to maintain safer and healthier rituals. Various studies have investigated the use internet of things (IoT) systems to enable crowd management during the Hajj, such as see [6]–[13].

An example of this is the use of radio frequency identification (RFID) systems. RFIDs have been used to collect data involving crowd movement and individuals' biometric information [9]. Each pilgrim is provided with a unique RFID tag that is activated near readers and transmits its unique ID, enabling the counting and estimating of crowds. The system can also retrieve information about the pilgrim and help with safety, medical, transportation, visa, and security management [9]. However, this system requires a costly network of readers distributed in all areas of the holy sites. For this system to work efficiently, a large portion of pilgrims need to receive personalized tags. This would not only be costly, it also includes the risk of misplacing the tags or their accidental destruction. An alternative to RFID technology is the use of existing mobile apps for crowd management [8]. This approach could use existing mobile phone networks to read the locations of phones and build a model of the crowds, without the need to develop new dedicated networks. A study [8] proposed using existing social media platforms and free wi-fi to enable crowd management by reading live GPS data and using push notifications to swiftly manage pilgrim movements to keep them organized. One concern with this technology, however, is that it requires educating end-users; this might be difficult, as a significant portion of pilgrims are elderly [14]. Another study [12] has proposed the use of unmanned aerial vehicles for image analysis, to estimate the levels of crowd density in Hajj sites based on views from above. This method could provide real-time crowd models with reduced cost, if used in more depth and accuracy.

In this study, crowd estimation was achieved using convolutional neural networks from existing CCTV cameras to accurately model crowd movements in the Almasjid Alnabawi without incurring any additional costs. The proposed system was tested on videos freely available online, allowing us to plot crowd levels throughout the mosque.

II. Methodology

Counting the number of people in the pictures was accomplished by using the VGG19 image classification model [15] as the base for the network. This model was trained on density maps obtained from crowd images. Let $\mathbf{D}(\mathbf{x}_m) >= 0 : m = 1, 2, ..., M$ be density map function that represents the crowds in a picture $\mathbf{x}_m$. The density map, as described by [16], represented the ground-truth of the number of people in a picture. The sum of all point annotations in the map represented the number of people. A Gaussian filter was applied to the density map, to remove sparks caused by point annotations that are difficult to train on [17].

$$\mathbf{D}^{gt} = \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{||\mathbf{x}_m - \mathbf{z}_n||_2^2}{2\sigma^2}), \tag{1}$$

where $\mathbf{z}_n$ is the point annotation map representing the location of each head in the crowd, and σ is the covariance matrix of $\mathbf{x}_m$. However, the above provided a fixed value for the Gaussian filter and, because head sizes vary among people, using an adaptive filter [18], [19] for each $n : \sigma_n \propto d_n$, where d_n is the distance to the nearest neighbor was recommended. As a result, to train the model we used the loss function:

$$\mathbf{L} = \sum_{m=1}^{M} F(\mathbf{D}^{gt}(\mathbf{x}_m) - \mathbf{D}^{est}(\mathbf{x}_m)), \tag{2}$$

where $F(.)$ is the distance function and $\mathbf{D}^{est}$ is the estimated crowd density map.

The convolutional neural network was trained using two datasets. First, we used UCF-QNRF [20], which included 1535 images of various crowds, including crowds in Al-Haram in Makkah, Saudi Arabia. We also used the ShanghaiTech A and B datasets [18], [21], which contained 1198 images of crowds from various sources, such as sporting events and scenes from crowded cities. Both datasets are widely used in crowd estimation and have been reported to have good performance when tested with similar image classification models [16], [18]. The final step was to use the trained model on new images from the Almasjid Alnabawi, to produce an estimated density map. The number of people in an image was then calculated by summing the number of predicted annotations from the density map.

The previous step was repeated to estimate crowds from all CCTV cameras. See Fig 2 for a sample of predictions. The actual numbers were obtained by manually annotating each person and visually counting the people, to validate the accuracy of the system.

Each camera covered a designated area inside the building. A survey of the building was done to find the occupancy for each of these designated areas during prayers. This survey was completed by counting the people inside each area during prayers. The crowd model was obtained by calculating the occupancy rate (i.e., dividing the predicted number of people in the area by the maximum capacity) and allocating a color code to indicate the rate. This was done for each area inside the building. For areas covered by more than one camera, the average ratio was taken. Once all areas were covered, a Gaussian filter was applied to the mesh to smooth its appearance and fill any blind spots.

For the purposes of this study, two YouTube videos from the official channel of Almasjid Alnabawi were used to estimate the crowds inside the Almasjid Alnabawi. The first video was recorded live during a busy time before the pandemic in 2019, while the second video was recorded while the COVID-19 precautionary arrangements were in place to enforce social distancing. The analysis shows the differences between the crowds; see Fig 2. There were 19 CCTV scenes, some obtained from the same camera for different angles. However, not all areas of the Almasjid Alnabawi were available for analysis, as some were closed on the day when the video was taken. The present analysis only includes the areas shown in the video. The allocation of each camera angle and its area was done manually. The next section demonstrates crowd estimates for normal days and during COVID-19 precautions, in addition to showing how these estimates were used to model crowds inside the Almasjid Alnabawi.

III. Results

The system was tested to produce estimates of the number of people in images and to model crowds throughout the building. Fig 2 shows the output for three images taken from YouTube videos on normal days during busy hours and three other pictures taken during COVID-19 social distancing precautions. The first three pictures were fed to the algorithm to predict the density maps, and the output clearly showed areas where the heads of the pilgrims were located. The orange/red areas indicated that there were more people further from the camera. The integration of these head locations resulted in numbers between 442 and 739. These values are close to the actual numbers. However, the larger the value of the crowd, the greater the margin of error. This is due to the difficulty of counting people far from the cameras and toward the back of the building. The second of the three images was also taken during a busy time, but with social distancing enforced. The head annotations showed pilgrims distancing further from one another, compared to their actions on normal days; this indicated the system's capability of detecting people's locations accurately. The crowd numbers varied from 190 to 230. These numbers account for around 30% of the size of a crowd on a normal day; this was done by leaving every other row empty and imposing a distance of around 1m surrounding each person. The predicted numbers were very close to the actual numbers.

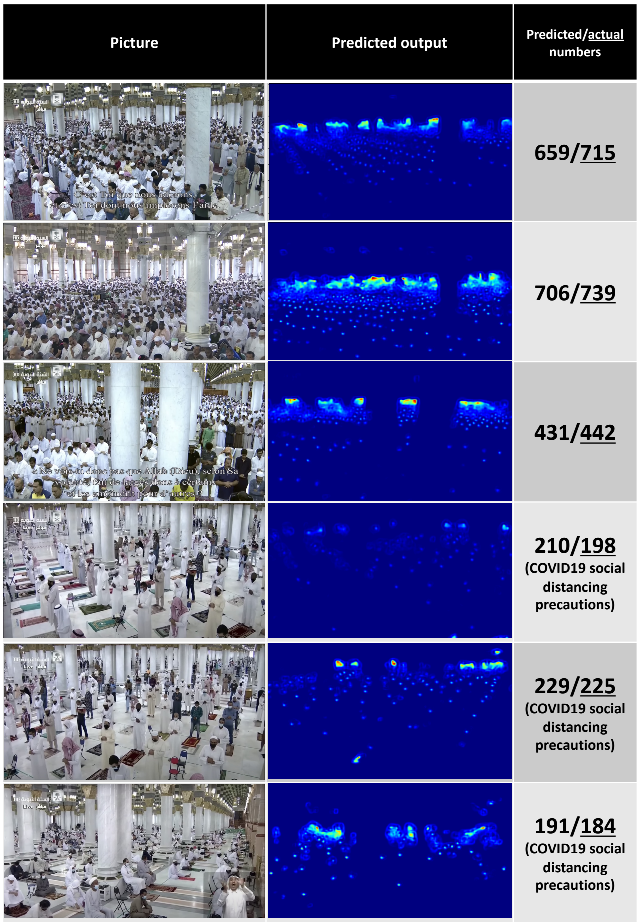

Fig. 2. Input images to the algorithm and the predicted density map. The predicted number was achieved by summing all annotations in the predicted density map.

The predicted output data from all cameras were correlated to their locations inside the building and then used to produce crowd models in their designated areas. Fig 3 represents the data from 19 locations inside the Almasjid Alnabawi under COVID-19 precautions. The figure shows the floorplan of the building, and the colormap indicates crowd levels, which have been color-coded, in each area. The two top corners of the building were female sections and had no broadcast CCTV, while the bottom side is the old mosque, which was not shown on YouTube that day. The top middle section had only one camera, shown on the left, which explains the low number of crowds on the right. This chart was taken from one moment in the video, but we could observe the live movement of crowds based on live CCTV.

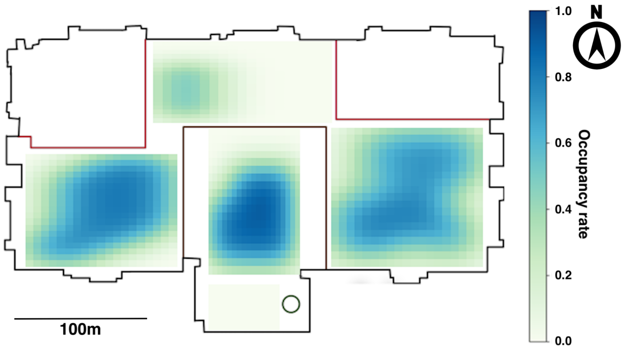

Fig. 3. Crowd modelling of the visitors to the Almasjid Alnabawi under COVID-19 precautions. The colors indicate the occupancy rate. The top two corners are female sections and were not broadcast, while the bottom part is the old mosque part and was not available on the day of the broadcast.

IV. Discussion

The system used CCTV footage to make real-time estimates of spatial crowds across the Almasjid Alnabawi in Medina, Saudi Arabia. The prediction performance was robust enough to produce accurate estimates of the crowds. This system used existing infrastructure to solve the problem of a lack of crowd information about the entire building. This solution to crowd modelling problems requires no extra infrastructure or major expenses, yet it is still capable of providing valuable data. This is unlike many other solutions, which require smartphone data plans and apps to provide these data. This system also does not require pilgrims to wear or attach an RFID tag to themselves. Moreover, it saves the need to educate hundreds of thousands of pilgrims on how to use the technology, making it the most efficient way of providing accurate crowd information indoors. It is easily expandable to areas outside the mosque.

During a pandemic such as COVID-19, it is essential to have information about crowds that are not adhering to social distancing rules, such as people gathering near one another and chatting. Such a system could easily help authorities detect these overcrowded areas and to act swiftly to reduce the spread of disease.

Despite its many benefits, there are limitations to this study. First, the current CCTV recordings were obtained from YouTube; we did not have access to all of the CCTV cameras. The use of the actual system could fill this gap and produce real-time data for crowd modelling. Second, the system requires high computational power to predict each frame in real-time for 30+ cameras. This will require more sophisticated graphical processing units for the faster processing of images.

Future implementations of the system could use accumulated data to train a neural network on the behavior of crowds and provide a model for predicting such behaviors in the future. For example, it could help authorities to avoid stampedes and overcrowding in small areas by giving prior warnings. This could help save lives and make the pilgrims' experience of the Hajj rituals even better and more comforting.

V. Conclusion

This paper introduced a method for using existing an CCTV system to provide real-time crowd modelling inside the Almasjid Alnabawi in Medina, Saudi Arabia. The large mosque serves millions of pilgrims annually, thus it requires sophisticated crowd monitoring and management systems. The proposed system uses live feeds of images to count the number of people within an area. The system was based on the VGG19 image classification neural network and was trained using existing detests. The results showed accurate estimates of the crowds and a graphical interface to represent the crowds' movements.

Acknowledgment

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (J: 11-135-1441). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

References

[1] K. Haase, M. Kasper, M. Koch, and S. Müller, "A pilgrim scheduling approach to increase safety during the Hajj," *Operations research*, vol. 67, no. 2, pp. 376–406, 2019.

[2] BBC, "Hajj stampede: At least 717 killed in Saudi Arabia," 2015. [Online]. Available: https://www.bbc.com/news/world-middle-east-34346449

[3] D. Helbing and A. Johansson, "Encyclopedia of Complexity and Systems Science, vol. 16, chap. Pedestrian, Crowd and Evacuation Dynamics, 6476–6495," *Springer*, vol. 26, pp. 28–34, 2009.

[4] M. Yamin, "Secure and healthy Hajj management: a technological overview," *American Academic & Scholarly Research Journal*, vol. 7, no. 3, pp. 195–202, 2015.

[5] Ministry of Media, "Muslim Pilgrims Perform Tawaf of Arrival in Makkah on First Day of Hajj 2020," 2020. [Online]. Available: https://hajj.media.gov.sa/en/news/muslim-pilgrims-perform-tawaf-of-arrival-in-makkah-on-first-day-of-hajj-2020/

[6] M. F. Mohamed, A. E.-R. Shabayek, and M. El-Gayyar, "IoT-Based Framework for Crowd Management," in *Mobile Solutions and Their Usefulness in Everyday Life*. Springer, 2019, pp. 47–61.

[7] M. Yamin, M. Mohammadian, X. Huang, and D. Sharma, "RFID technology and crowded event management," in *2008 International Conference on Computational Intelligence for Modelling Control & Automation*. IEEE, 2008, pp. 1293–1297.

[8] M. Yamin, H. M. Al-Ahmadi, and A. Al Muhammad, "Integrating social media and mobile apps into Hajj management," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2016, pp. 1368–1372.

[9] A. H. Al-Hashedi, M. R. M. Arshad, A. S. Baharudin, and H. H. Mohamed, "RFID applications in Hajj management system," in *2013 IEEE International Conference on RFID-Technologies and Applications (RFID-TA)*. IEEE, 2013, pp. 1–6.

[10] N. Nasser, A. El Ouadrhiri, M. El Kamili, A. Ali, and M. Anan, "Crowd Management Services in Hajj: A Mean-Field Game Theory Approach," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–7.

[11] N. Nasser, M. Anan, M. F. C. Awad, H. Bin-Abbas, and L. Karim, "An expert crowd monitoring and management framework for Hajj," in *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*. IEEE, 2017, pp. 1–8.

[12] A. Almagbil, "Detecting and Estimating the Levels of Crowd Density from UAV Imagery," *Dirasat, Human and Social Sciences*, vol. 46, no. 1, 2019.

[13] H. Eladly, "RIDR Model for Crowd Management using RFID Technology," *International Journal of Control Systems and Robotics*, vol. 4, 2019.

[14] A. A. Owaidah, "Hajj crowd management via a mobile augmented reality application: a case of The Hajj event, Saudi Arabia," Ph.D. dissertation, University of Glasgow, 2015.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] Z. Ma, X. Wei, X. Hong, and Y. Gong, "Bayesian loss for crowd count estimation with point supervision," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6142–6151.

[17] Y. Li, X. Zhang, and D. Chen, "Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1091–1100.

[18] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 589–597.

[19] L. Liu, Z. Qiu, G. Li, S. Liu, W. Ouyang, and L. Lin, "Crowd counting with deep structured scale integration network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1774–1783.

[20] H. Idrees, M. Tayyab, K. Athrey, D. Zhang, S. Al-Maadeed, N. Rajpoot, and M. Shah, "Composition loss for counting, density map estimation and localization in dense crowds," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 532–546.

[21] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 833–841.