

Counting of People in the Extremely Dense Crowd

Muhammad Arif^[1, 2], Sultan Daud^[1, 2], SalehBasalamah^[1, 2]

^[1]Center of Research Excellence in Hajj and Omrah (HajjCoRE), Umm Al-Qura University, Makkah, Saudi Arabia

^[2]College of Computer and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia

Article Info

Article history:

Received Nov 13, 2012

Revised Jan 05, 2013

Accepted Jan 12, 2013

Keyword:

People Counting
Image Processing
Blob Analysis
Median filtering
Genetic algorithm

ABSTRACT

In this paper, we have proposed a framework to count the moving person in the video automatically in a very dense crowd situation. Median filter is used to segment the foreground from the background and blob analysis is done to count the people in the current frame. Optimization of different parameters is done by using genetic algorithm. This framework is used to count the people in the video recorded in the mattaf area where different crowd densities can be observed. An overall people counting accuracy of more than 96% is obtained.

Copyright © 2013 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Muhammad Arif,
College of Computer and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia.
Email: mahamid@uqu.edu.sa

1. INTRODUCTION

Crowd monitoring is very important in many aspects especially in the areas of Airports, railway stations, sports, and rallies. In Saudi Arabia, Hajj and Ramadan are the mega events in Makkah Mukarama where lot of people get together to do umra and hajj. Crowd management during Ramadan and Hajj at the Holy Mosque in Makkah is a daunting task. Tremendous effort from the security staff is required to manage the huge crowd peacefully and smoothly. In the last decade, due to low cost of cameras, lots of cameras were used for the surveillance of public places. Lot of cameras are installed in the Grand Mosque that help the security staff to take appropriate crowd management decisions. Normally uneven crowding occurs in the Mosque due to tendency of people entering in the nearest possible entry points when they arrive in the Mosque. This situation creates suffocation in these areas and becomes prone to certain mishap or loss of precious lives. Looking at these installed cameras and estimating the crowd density and directing the people to empty places requires huge amount of manpower. Manual analysis of high quantity of visual data is not practical and an automatic decision support system is required that can guide the security staff and public to minimize the crowding of people in certain places and optimize the usage of the grand mosque. Lot of research is being done in automating the process of estimation and management of crowd using visual cameras, thermal imaging or other sensors placed at the entry points. In the last decade, due to low cost of cameras, lots of cameras were used for the surveillance of public places.

Manual monitoring of crowd is done by putting many surveillance cameras and some observers monitor the crowd density and their movement. This scenario is very costly as lot of manpower is required who can watch the monitors continuously for many hours. Hence, alertness of the observers is an important factor in good surveillance. As the working hours increases as the case of Masjid-e-Haram, fatigue and stress of the observer increases degrading their performance. The importance and demand for automated tools to manage and analyze crowd behavior and dynamics grows day by day as the population increases. People counting in the crowded areas are being done either by segmenting the people or head, or based on texture analysis or wavelet descriptors.

Journal homepage: <http://iaesjournal.com/online/index.php/IJAI>

Most of the research focused on dividing the crowd density into bins and then extracting some useful features to classify the bins correctly. Xiaohua et al [1] showed classification accuracy of 95% when crowd density is classified into four classes by using wavelet descriptors. Classification is done by support vector machine. Their method is good for estimation of crowd density for moderate crowd density. Ma et al [2] used texture descriptors called advanced local binary pattern descriptors to estimate crowd density estimation. The ground truth is manually labeled into five categories starting from low to high density. Some researchers [2, 3, 4] have used texture analysis to extract certain features from the images and have used neural networks to estimate the crowd. Cho et al 1998, 1999 [5, 6] and Huang et al 2002 [7] blended the concept of image processing and neural networks to estimate and count the crowd of people. Roqueiro et al [8] uses the foreground pixels and finds them using a Median Background computing technique. They apply classification algorithms like SVM, k-nearest, PNN, BPNN to classify the images in 2 categories first, zero persons and one and more persons. On more than zero people's categories it again applies the classification techniques to find the number of people in the scene. Zhao et al [9] proposed Bayesian model based segmentation to segment and count people but this method is not appropriate for high density crowds. Yoshinaga et al [10] proposed blob features of moving objects to eliminate background and shadow from the image. For each blob of moving people, numbers of pedestrians are estimated by using neural networks. They have shown that accuracy of 80% can be achieved by this method in the real life scenarios where maximum numbers of pedestrians are 30 in a single frame. Roqueiro et al [8] used simple background subtraction from the static images to estimate the crowd density.

Not many papers are published related to crowd estimation or people counting in Masjid-e-Haram. Hussain et al [11] have proposed pixel based crowd density estimation system. They have used crowd foreground blobs to classify the crowd into five ranges from very low to very high using neural networks. Sarmady et al [12] has proposed an interesting model for circular tawaf around Kaaba.

In this paper, we have considered videos of the Mattaf area in the haram while hundreds of people doing tawaf around Kaaba.

Density of the crowd is very high and changing with the distance from the Kaaba. In the next section, we have presented our proposed framework. In section 3, we have discussed data and results. In the final section, paper is concluded.

2. RESEARCH METHOD

Proposed framework for people counting is shown in the Figure 1. Video from the video camera is recorded frame by frame. In the first step background image is calculated and subtracted from all the images to get the foreground images. In these images only moving people are left in the frames. Optimization of the threshold value to specify which pixel belongs to background and which pixel belongs to foreground plays an important role in segmenting the moving people in the mattaf area and to filter out the jitter or small movement of the video camera. In the next stage, blobs of particular size are identified to estimate the number of people in the crowd. In this step, proper selection of blob size is very important as larger blob size may combine two or more people together and show them as single person, whereas smaller blob size may consider a single person as two or more persons. After identifying the blobs of a particular size, these blobs are counted and we output the number of people present in the frame as the count of the blobs present in this frame.

One recorded video in the mattaf area of Al-Haram Mosque is used to assess the performance of the proposed framework. Sample frame of the video is shown in Figure 2. In the sample frame, there is a rich background, and many people are entering the tawaf area and leaving this area from different sides. Some people are standing and praying which are in relatively very slow motion while remaining on their place of prayer. In the tawaf area (middle area) many people are moving at different speed. At the outer circle they are moving fast whereas in the inner circle the motion is slow. Frame rate of the video is 50 frames per second and frame resolution is 1920×1080 . The video contains more than 10,000 frames. Since there are more than 2500 people in motion in every frame so it is extremely difficult to count in all the 10,000 frames. Moreover, in one second there are 50 frames and we expect that not much change occurs from one frame to the other frame. Hence we have decided the count the people in motion in 100 frames at almost equal interval of about two seconds. We assume that the error in the people counting occurring in these 100 frames will be the same in all 10,000 frames.

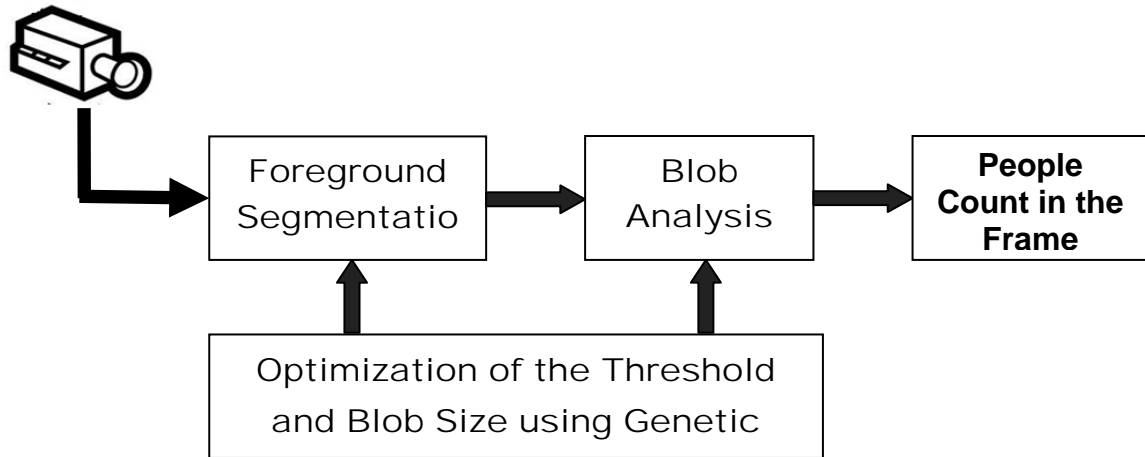


Figure 1. Proposed Framework for people counting



Figure 2. Sample frame from the video

2.1. Foreground Segmentation

Foreground segmentation is a pre-processing step to identify the moving objects from the background. Pixels in the current frame that deviate significantly from the background are considered to be moving objects. These foreground pixels are further processed for object localization and tracking. Piccardi [13] has provided a good survey by comparing different background subtraction methods. Median filter is designed by buffering N number of frames and median of these frames are calculated and a threshold is applied to detect the background of the video. This method is very effective but many frames have to be stored to calculate the median frame. In median filtering, the previous N frames of video are stored in the buffer and the background frame is calculated as the median of buffered frames. Then the background frame is subtracted from the current frame to find out the foreground pixels. The approximate median method [14] gives a good alternate solution to the buffering the N frames in the memory.

Median frame is calculated from these buffered N frames. The first frame is taken as the background frame and for the next coming frames; if the pixel value of the current frame is greater than the background pixel then the pixel value of the background image is incremented by 1. If the pixel value of the current

frame is less than the background pixel value then the pixel value of the background pixel is decremented by 1. Hence it is assumed that after sufficient number of frames, the background image will converge to the true median image of the video. Figure 3 shows background and foreground images when median filter is applied on the frame shown in Figure 2. Accuracy of foreground segmentation depends on the threshold value.

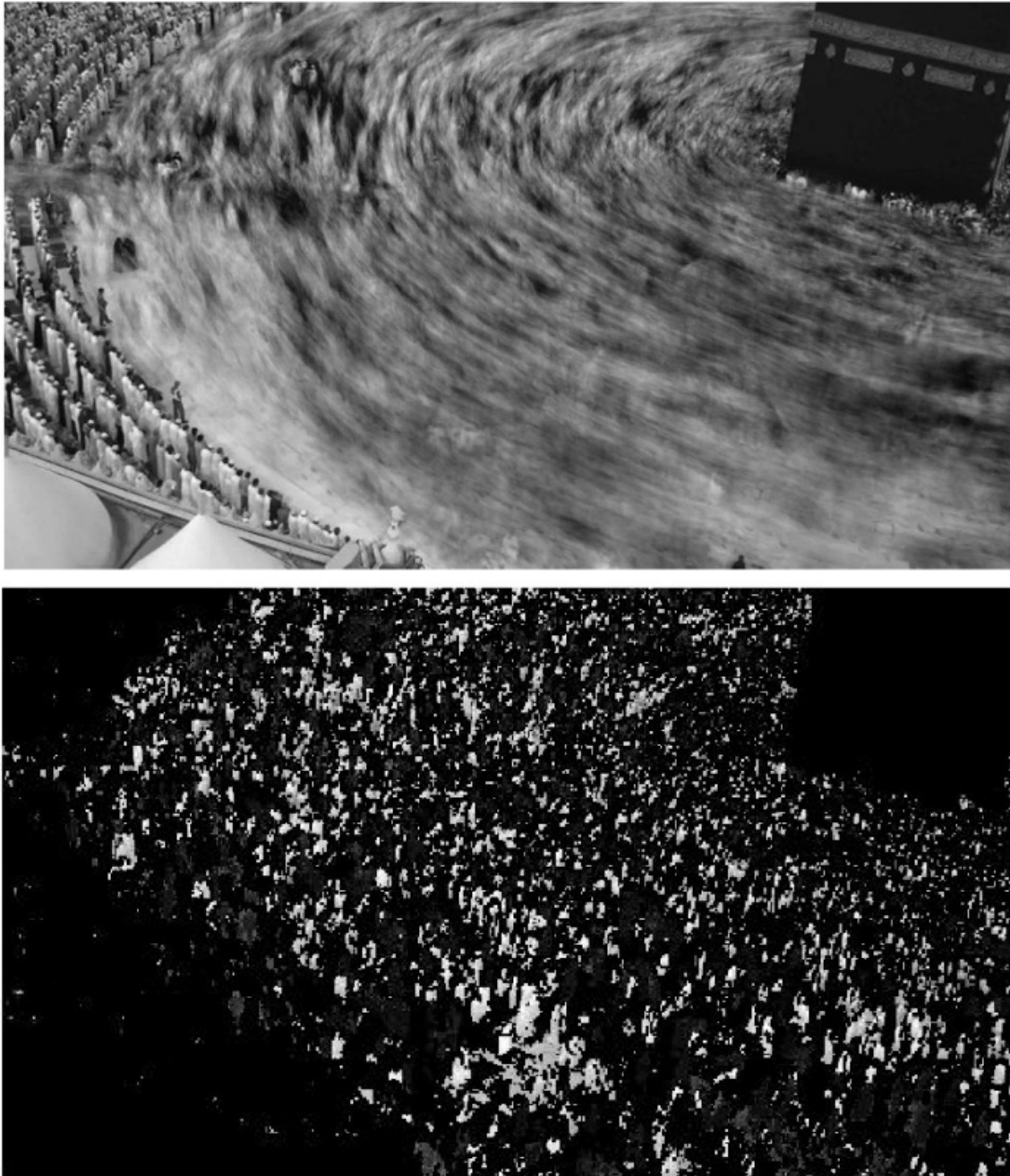


Figure 3. Background (top) and foreground frames (bottom) using median filter

2.2. Finding Blobs in the Foreground Frame

Blobs are the connected regions in a binary image. Blob analysis process is aimed at detecting point and/or regions in the image that differ in properties like brightness or area etc. For blob detection, image is first converted to binary image. Then next step is finding the connected components in the binary image. To find the connected components in the binary image, we start with the unlabeled pixel and find the

neighboring connecting pixels and label them connected. We keep doing it until there is no neighbor. Then we search from the next unlabeled pixel and repeat the same process [15]. This process can be done by the matlabcommand [16] “bwconncomp”. Then properties of the connected regions can be obtained by the matlabcommand “regionprops”. Area is the number of pixels in the region. Each binary image has a lot of connected components of variable size. In our particular application, we are interested in the area of the connected regions which we call blobs. We discard all the blobs whose area is below a particular user defined blob area cutoff value α and count all the blobs having area above α . This corresponds to the number of moving people in a particular frame.

2.3. Optimization of the Threshold Value γ and Blob Area Cutoff Value α

It can be seen from previous discussion that threshold value γ for foreground segmentation and blob area cutoff value α are very critical for correctly counting the people in the frame. Hence, we have optimized these values using genetic algorithm. Genetic algorithm is widely used in the optimizations problems and gives very promising results [18, 19]. Basic steps in the genetic algorithm consist of initialization of a population of solutions, assessing the fitness of this population, and based on the fitness values, make crossovers and mutations to create a new population. In our optimization algorithm, intermediate crossover [17] with roulette wheel selection method for parents is used to perform the crossover as given below,

$$\begin{aligned} O_1 &= \beta P_1 + (1 - \beta)P_2 \\ O_2 &= \beta P_2 + (1 - \beta)P_1 \end{aligned}$$

Where P_1 and P_2 are the parent and O_1 and O_2 are the offspring. Gaussian noise based mutation operator is used to introduce the mutation in the population by adding random Gaussian noise to the individual. Parameters of the genetic algorithm to optimize γ and α are shown in Table 1.

Individuals of the population comprise of threshold value for median filter and blob area cutoff value in the range of [10, 80] for the threshold and [10, 50] for the blob area cutoff. Both values are optimized for the training frames and then used as fixed values for the testing frames of the video.

Table 1. Parameter values for genetic algorithm

Parameter	Value
Maximum Generations	10
Population size	14
Crossover Probability	0.9
Mutation Probability	0.2
Mutation Scale	0.3

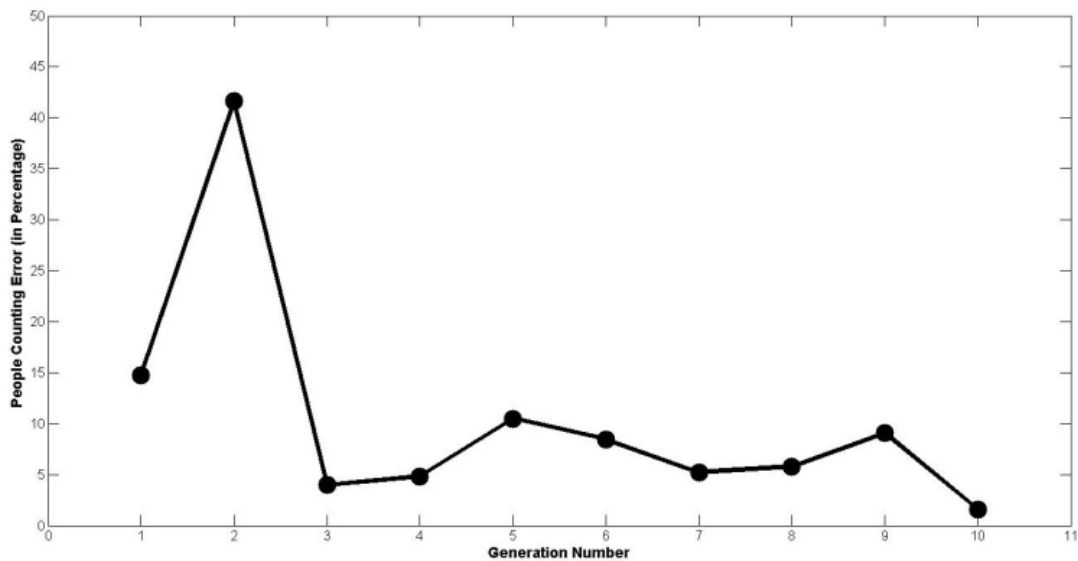


Figure 4. Average people counting error on training frames during genetic algorithm run (Generation wise)

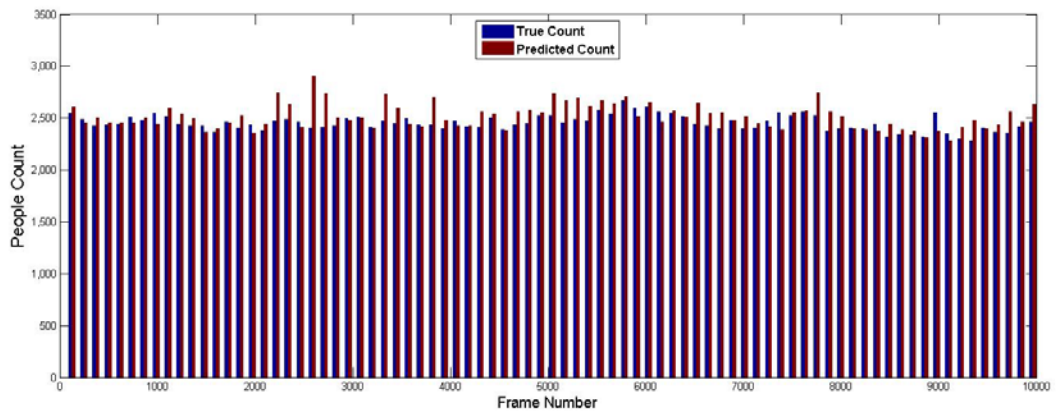


Figure 5. True and predicted people count in the video

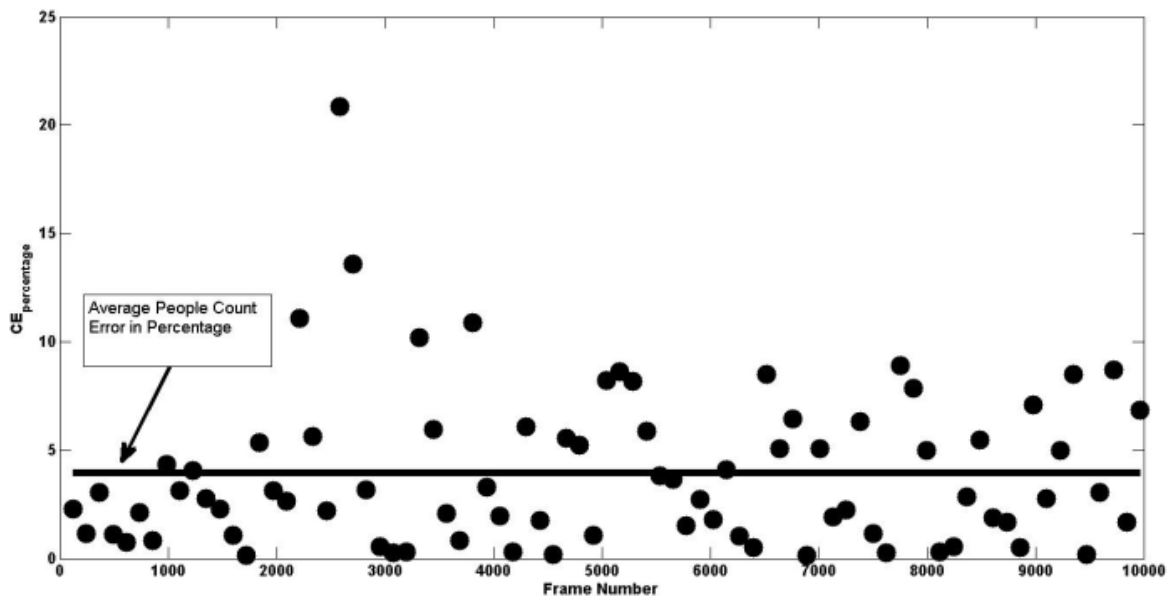


Figure 6. People count error in percentage

3. RESULTS AND ANALYSIS

To assess the performance of our proposed framework, selected video contains 10000 frames at the sampling rate of 50 Hz. Every frame contains more than 2000 persons in the mattaf area. So it was not possible to manually annotate all the frames of the video. Moreover, the sampling rate of 50 Hz means there are 50 frames per second and movement of the people are very fast, so we expect that number of people will not change so abruptly. So we have annotated the frames at constant intervals. One person has counted the persons in a frame and second person check the annotation to remove any human error in counting. Total 82 frames are annotated. Initial four annotated frames, frame number 124, 247, 370 and 493 of the video are used to optimize the threshold value from the median filter and blob area cutoff value for the blob analysis.

Genetic algorithm is run with the settings shown in Table to find out the threshold and blob area cutoff values using four frames mentioned above. Accuracy of people counting is used as fitness function to evolve the population. Figure 4 shows average people count error of the whole population versus generation. It can be seen that the error is dropped considerably to 2%.

The threshold value of the best individual in the 10th generation is found to be 45.6 and blob cutoff area is 21.3. These settings are used to test the people counting accuracy of all 81 frames equally spaced in 10000 frames of the video. True and predicted people count is shown in Figure 5 for all 81 frames. It can be seen from the figure that predicted people count is very near to the actual people count in the frame.

People counting error is defined as follows,

$$CE_{\%age} = \frac{|PC_{True} - PC_{Pred}|}{PC_{True}} \times 100$$

Where $CE_{\%age}$ is the people counting error (in percentage), PC_{True} is the true people count from the manual annotation and PC_{Pred} is the predicted people count from our proposed framework. $CE_{\%age}$ is plotted in Figure 6 for all 81 frames.

Average people counting error is found to be only 3.9% which is quite low keeping in mind the scenario of the video and the extreme crowd present in the video.

4. CONCLUSION

A good accuracy of more than 96% is observed in the video which show the effectiveness of the proposed framework. It can be applied to the different focus settings of the camera just by optimizing the threshold and blob area values. In this video we have assumed that camera is watching the crowd from approximately top position and hence the effect of angled view is not significant.

In this paper, we have emphasized that by using video at a particular location moving people can be counted easily by our framework with slight modification and optimization. Hence in the future work, the video streaming from the whole mattaf area is used with pre-defined locations and angles to count the number of people doing tawaf and number of people present in a particular location. Empty spaces in the tawaf area, local density of the people doing tawaf and number of people entering and exiting from the tawaf area will also be counted.

ACKNOWLEDGEMENTS




This research has been supported by the Center of Research Excellence in Hajj and Omrah (HajjCoRE), Umm Al-Qura University. Under Project number P1119, titled "Automatic Decision Support System for Crowd Estimation and Management in Masjid-e-Haram".

REFERENCES

- [1] Xiaohua L, Lansun S, Huanqin L. Estimation of crowd density based on wavelet and support vector machines. *International Conference on Intelligent Computing*. 2005.
- [2] Ma W, Huang L, Liu C. Advanced local binary pattern descriptors for crowd estimation. *IEEE Pacific Asia Workshop on Computational Intelligence and Industrial application*. 2008: 958-962.
- [3] Marana AN, Costa LF, Lotufo RA, Velastin SA. *On the efficacy of texture analysis for crowd monitoring*. Proc. Computer Graphics, Image Processing, and Vision. 1998: 354-361.
- [4] Marana AN, Velastin SA, Costa LF, Lotufo RA. Automatic estimation of crowd density using texture. *Safety Science*. 1998; 28: 165-175.
- [5] Cho Y, Chow TWS. A fast neural learning vision system for crowd estimation at underground stations platform. *Neural processing letters*. 1999; 10(2): 111-120.
- [6] Cho Y, Chow TWS, Leung CT. A neural based crowd estimation system by hybrid global learning algorithm. *IEEE Trans on Systems, Man, and Cybernetics, Part B*. 1999; 29(4): 535-541.
- [7] Huang D, Chow TWS, Chau WN. Neural network based system for counting people. *IEEE Trans on Systems, Man, and Cybernetics, Part B*. 2002; 31(4): 2197-2200.
- [8] Roqueiro D, Petrushin VA. Counting people using video cameras. *International Journal of Parallel, Emergent and Distributed Systems (IJPEDES)*. 2007; 22(3): 193-209.
- [9] Zhao T, Nevatia R. Bayesian human segmentation in crowded situations. *IEEE Conference on Computer Vision and Pattern Recognition*. 2003; 2: 495-466.
- [10] Yoshinaga S, Shimada A, Taniguchi R. *Real-time people counting using blob descriptor*. Procedia Social and Behavioral Sciences. 2010; 2: 143-152.
- [11] Hussain N, Yatim HSM, Hussain NL, Yan JLS, Haron YF. CDES: A pixel-based crowd density estimation system for Masjid al-Haram, Safety Science, doi:10.1016/j.ssci.2011.01.005, 2011.
- [12] Sarmady S, Haron F, Talib AZ. A cellular automata model for circular movements of pedestrians during Tawaf. *Simulation Modeling Practice and Theory*. 2011; 19: 969-985.
- [13] Piccardi M. Background subtraction techniques: a review. *IEEE International Conference on Systems Man and Cybernetics*. 2004: 3099-3104.
- [14] McFarlane NJB, Schofield CP. Segmentation and tracking of piglets in images. *Machine Vision and Applications*. 1995; 8(3): 187-193.
- [15] Haralick RM and Shapiro LG. *Computer and Robot Vision*. Volume 1, Addison-Wesley. 1992: 28-48.
- [16] MATLAB version 2011a. Natick, Massachusetts: The MathWorks Inc., 2011.

- [17] Mühlenbein H and Schlierkamp-Voosen D. Predictive Models for the Breeder Genetic Algorithm: I. Continuous Parameter Optimization. *Evolutionary Computation*. 1993; 1(1): 25-49.
- [18] Xuesong Yan and Qinghua Wu and Can Zhang and Wei Li and Wei Chen and Wenjing Luo. An Improved Genetic Algorithm and Its Application. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1081-1086.
- [19] Xuesong Yan and Qinghua Wu and Hammin Liu. An Improved Robot Path Planning Algorithm Based on Genetic Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(8).

BIOGRAPHIES OF AUTHORS

	<p>Dr Muhammad Arif is Professor in the Department of Computer Science, Umm Alqura University, Makkah, Saudi Arabia. He has done his PhD in System Information Sciences, from Tohoku University, Japan in 1999. He has published more than 100 papers in various journal and international conference proceedings. His research interests are Intelligent Monitoring, Pattern Recognition, Evolutionary Computing and Biometrics.</p>
	<p>Sultan Daud Khan received the BS degree in Computer Engineering from University of Engineering & Technology, Peshawar, in 2005 and MS degree in Electronics & Communication Engineering from Hanyang University, South Korea, in 2010. During his MS studies, his research was mainly focused on Off-Chip memory access optimization for MPSoCs. Currently he is a Lecturer and Research Assistant in Department of Computer Engineering of Umm Al-Qura University, Saudi Arabia since Jan, 2011.</p>
	<p>Dr Saleh Basalamah received his PhD from Imperial College London, UK in 2005. Presently He is Deputy Director, Geoinformatics Center, Umm Alqura University, Makkah Saudi Arabia. Currently He is assistant professor at the Department of Computer Engineering at Umm Al-Qura University. He was the Dean of College of Computing and Information Systems from 2009 to 2012. His research interests include Medical Image Analysis, Computer Vision and Visual Sensor Networks.</p>