**PAPER • OPEN ACCESS**

# Congestion analysis of pilgrims in Hajj and Umrah congregation using block matching and optical flow

View the article online for updates and enhancements.

# Congestion analysis of pilgrims in Hajj and Umrah congregation using block matching and optical flow

**Sumaiyya Farooq, Shoab A Khan and M Usman Akram**

Department of Computer Engineering, College of E&ME, National University of Sciences and Technology, Islamabad, Pakistan

E-mail: sumaiyya20@hotmail.com

**Abstract.** A novel method has been proposed to classify the motion of pilgrims with respect to congestion level in the holy mosque of Makkah. Millions of Muslims visit this mosque during Hajj and Umrah every year. A large number of security personnel is required to maintain the smooth flow of pilgrims in order to avert any catastrophe. Therefore, it is inevitable to design a computer aided system to reduce human effort. The proposed system pre-processes input images to segregate the moving shadows and pilgrims in order to nullify the false motion due to moving shadows. A hybrid method consisting of block matching and optical flow techniques has been used for the computation of motion vectors. Decision tree classifier is used on the number of motion vectors having non-zero magnitude. Experiments show that the proposed system has promising results yielding an accuracy of 90.58% for the congestion classification of pilgrims.

## 1. Introduction

A huge number of Muslims from all over the world gather at the holy mosque of Makkah every year to perform Hajj and Umrah. The number of pilgrims remains very high at both these events. A large number of security personnel is needed to be deployed to manage smooth flow of pilgrims to avoid undesirous situations. A large number of surveillance cameras are installed in the mosque. But manual monitoring of enormous data acquired from these cameras is a highly tedious job. Improvements in the fields of image processing, computer vision and machine learning has sparked a lot of scope to explore this area. More novel, robust and efficient techniques can be developed that can help security staff to monitor the congestion level and suggest some suitable measures in case some emergency situation takes place.

Chen and Huang proposed a motion-based system for detection of abnormal activity in human crowd [1]. Optical flow was computed and the characterization of crowd was performed using unsupervised learning techniques on the features obtained by optical flow. Prediction of crowd behaviour was done using the force field model. H. Idrees et al. [2] incorporated multiple bases of information to make a reliable crowd estimation system in which texture analysis, head count using HOG features and frequency domain analysis was merged. Direct counting of people was not possible because pixel count per person was low and images were highly occluded.

Counting of people was done using head detection method by V.B. Subburaman [3]. Interest points were detected by calculating gradient of the image and head contours were used to train SVM classifier to detect and count human heads. J. Yang et al. estimated the crowd density using the texture features based on gray level co-occurrence matrix (GLCM) in [4]. Texture features such as energy,

entropy, contrast, homogeneity, were passed to SVM classifier for density detection and linear regression model was used to count number of people.

To deal with Hajj and Umrah data is much different and challenging as compared to other crowd scenarios. Two reasons for this strenuous situation include: 1) low contrast between the white marble floor and the clothing of pilgrims [5], 2) the motion of pilgrims is highly congested and 3) heavy congestion leads to highly occluded scenes. Figure 1(a) shows the occlusion problem in which moving pilgrims overlap each other and different congestion levels such as high and low are depicted in (b). Many researchers have touched this area in recent years and a brief review of their work is presented below.
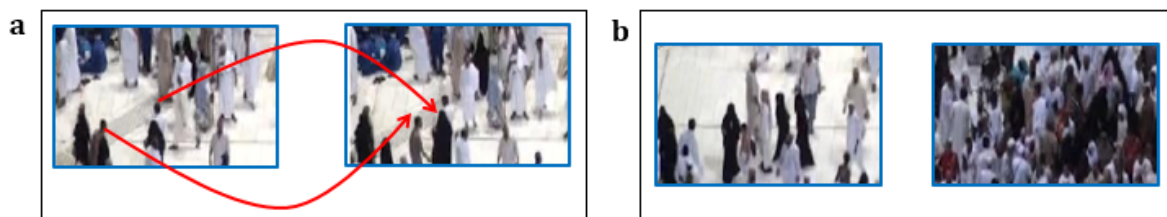


**Figure 1.** Problems of (a) occlusion (b) low and high congestion scenes.

Hussain et al., in 2011, proposed a crowd density estimation system (CDES) [6] for medium density level specifically for Al-Haram mosque, Makkah. Pixel based features of foreground blobs were extracted and density was classified into five levels by building backpropagation neural network on foreground pixel features. In 2013, M. Arif et al. presented a framework to count the number of pilgrims moving in the Mattaf area with more than 2000 persons per frame. The estimation was done using blob analysis techniques on the segmented image data because the size of the blob was found to be correlated with the actual number of people [7].

Not much work has been done to design a viewpoint dependent framework which is able to characterize the motion of pilgrims in extremely dense and occluded scenarios. This paper addresses the above mentioned scenario by analysing the level of congestion of pilgrims. To achieve the above said task, computation of motion vectors (MVs) is performed and number of motion vectors having non-zero magnitude is used to classify the congestion level.

The paper is organized in the following manner: Section 2 describes the proposed methodology thoroughly. In section 3, results of the proposed framework are presented. Section 4 discusses the results and Section 5 concludes the paper.

## 2. Proposed methodology
The proposed system aims to estimate and analyse the congestion level of pilgrims in Hajj and Umrah congregation. Top level diagram of the proposed framework is shown in figure 2. The input images are taken in temporal sequence and display the outside view of holy mosque of Kabah near the entrance gate and have a size of 1440x1080 pixels. It is assumed that the camera is held stationary and mounted at an elevated place while capturing the images.

### 2.1. Preprocessing
Preprocessing includes following steps:
   (i) Images are first converted to gray scale and cropped to limit the ROI to 432x1440 pixels.
   (ii) An adaptive binary mask is generated to separate the pilgrims from their moving shadows to discard false motion.
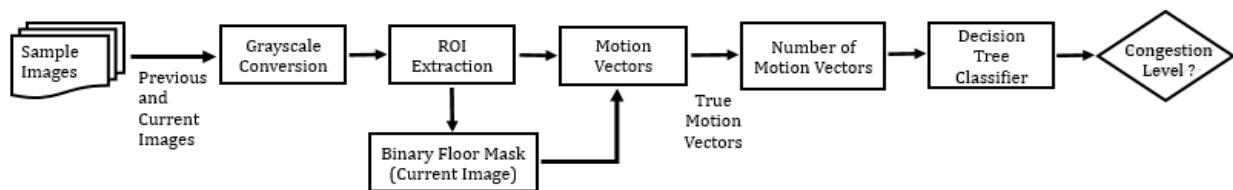    (a)  Input image is divided into 50x50 sized patches.

**Figure 2.** Flow of proposed framework for congestion analysis of pilgrims.

(b)  An adaptive threshold 'T' is set equal to 80% of maximum intensity level as shown in equation (1).

$$T = 0.8* \text{ (Max Intensity Level)} \qquad (1)$$

(c)  Binary mask is generated using equation (2). Take the complement of the mask.

$$\text{mask} = \begin{cases} 0; & pixel\ Intensity\ < T \\ 1; & pixel\ Intensity\ \geq T \end{cases} \qquad (2)$$

### 2.2. Estimation of motion

MVs have been computed using the hybrid technique proposed by S.H. Chan et al. in [8]. At first step, block matching technique based on Sum of Absolute Differences (SAD), block size of 8x8 and search window of 16x16 pixels computes the integer part of displacement. After that, Taylor refinement has been applied on the translated reference image block to calculate the sub-pixel motion.

A significant modification has been proposed to reduce the computation time of calculating motion vectors. Motion vectors have been calculated only for those blocks that hold `'1' in the binary mask. The modified algorithm to calculate motion vectors is given below:

**Algorithm 2.1:** modified algorithm to compute $(V_x, V_y)$

**for** $i \leftarrow 1$ **to** rows of target image
    **for** $j \leftarrow 1$ **to** columns of target image
        **if** mask $(i, j) = 1$
            $(V_x, V_y) \leftarrow$ x and y components of motion vector for block $(i, j)$ respectively

### 2.3. Classification of Congestion

Number of MVs having non-zero magnitude "N" is a potential indicator of congestion level in the scene. The MV grid is divided into 3x6 blocks and MVs are calculated for each block separately. In this way, congestion level is analysed in local blocks. Decision tree classifier is used to divide the congestion level of each block into five classes (figure 3). The upper branch of the tree is followed if the condition of relevant node is true and vice-versa.
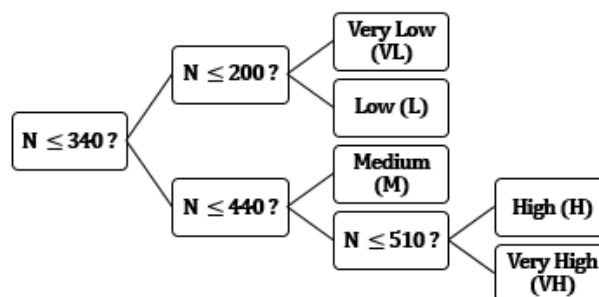


**Figure 3.** Decision tree for congestion classification.

## 3. Experimental Results

The formulation of congestion analysis system is tested on images captured at the entrance gate of Al-Haram mosque. Since this data is not publicly available, it is manually annotated to set the ground truth for congestion analysis.

Figure 4(a) shows the original image with red box indicating the ROI. Floor mask is generated for the extracted ROI as shown in figure (b) and (c) to segment the pilgrims and their moving shadows on the floor.
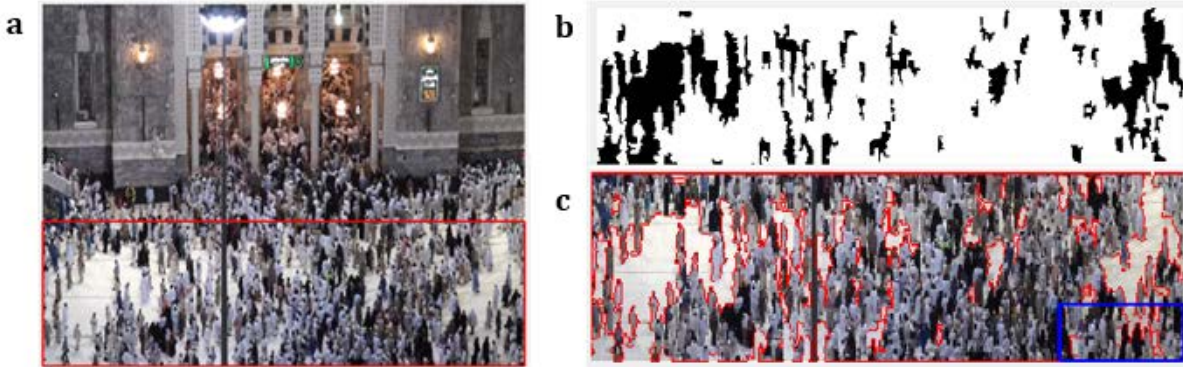


**Figure 4.** Preprocessing results (a)original frame (b)floor mask (c)result of floor segmentation.

MVs are calculated using binary mask as a prior check to eliminate the false motion. Results of it are presented in figure 5. It is quite clear that redundant MVs are cleaned up to a greater extent.
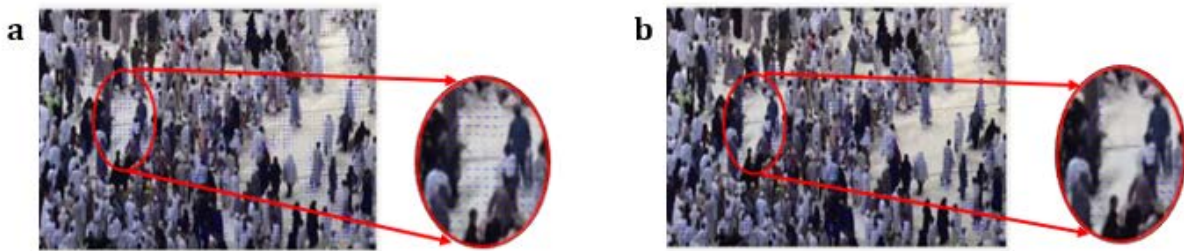


**Figure 5.** Motion vectors (a)without floor mask condition (b)with floor mask condition.

To make the system more accurate, number of non-zero MVs are calculated for all 3x6 sub-images separately. The classification results for one sub-image from each class of congestion are shown in figure 6. The title of each sub-image includes number of MVs having non-zero magnitude 'N', ground truth (GT) class of congestion and predicted class respectively.
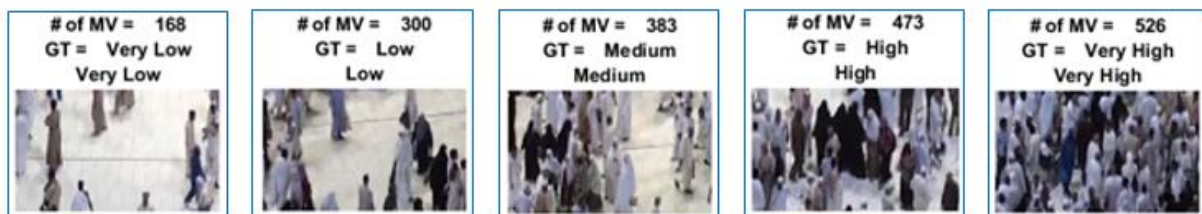


**Figure 6.** Results of congestion classification for each class.

A total of 1050 images are tested and number of non-zero MVs are averaged out for a group of 10 consecutive images to discard any irregularities. Each image further consists of 18 sub-images. Hence we have a total of (1050/10) x 18 = 1890 test subjects. The classification accuracy is found to be 90.58%.

## 4. Discussion

Illumination variation and occlusion are the two main reasons for wrong classification. Illumination does not remain constant across the image. This poses a challenge on the floor extraction method. The reason is primarily the inclusion of static floor along with moving pilgrims because shadow makes the floor appear a bit darker than the rest of the floor. The blue box in the figure 4(c) region explains the scenario. Images are more prone to wrong classification if actual number of motion vectors lie at the periphery of two classes. In such a case, a slight increment in their number throw them in next class of congestion.

Second reason of anomalous results in some cases is the occlusion of pilgrims in highly dense areas of images. Despite the fact that camera is mounted at an elevated place, it still occludes the persons and remains problematic for the correct classification of congestion level of pilgrims. Because of large density of persons, no algorithm of image processing can overcome the problem of occlusion.

To my best knowledge, a significant work has not been done yet to build a crowd analysis system for Hajj and Umrah events where number of pilgrims are in millions. Therefore, a true comparison cannot be performed between our proposed methodology and the existing state-of-the-art techniques.

## 5. Conclusion

We have proposed a novel framework to classify the congestion level of moving pilgrims at the entrance gate of Al-Haram mosque. Floor mask is generated to segment out the static floor and shadow of moving pilgrims. A hybrid technique comprising of block matching and optical flow methods is used to calculate the motion of pilgrims. Block matching computes the integer part of displacement while sub-pixel motion has been computed using optical flow. Further, a modification is proposed to calculate the motion vectors. Floor mask is incorporated as a prior check to see whether to compute motion vectors for a given block of input image or not. To classify the congestion level of pilgrims, number of motion vectors having non-zero magnitude is considered. In future, there is a great scope to work in this area. For example, decision support systems can be built on the basis of such statistics of motion of pilgrims to avoid any unfortunate situation.

## References

[1]    Chen D Y and Huang P C 2011 *J. Visual Communication and Image Representation* **22** pp 178-86
[2]    Idrees H, Saleemi I, Seibert C and Shah M 2013 *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 2547-54
[3]    Subburaman V B, Descamps A and Carincotte C 2012 *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE 9th Int. Conf. on (IEEE)* pp 470-75
[4]    Yang J, Li J and He Y 2014 *J. of Multimedia* **9** 1152-59
[5]    Al-Khaffaf H S, Haron F, Sarmady S, Talib A Z and Abu-Sulyman I M 2012 *Frontiers in Computer Education (Springer)* pp 727-36
[6]    Hussain N, Yatim H S M, Hussain N L, Yan J L S and Haron F 2011 *Safety Science* **49** 824-33
[7]    Arif M, Daud S and Basalamah S 2013 *IAES Int. J. of Artificial Intelligence* **2** 51
[8]    Chan S H, Nguyen T Q et al. 2010 *2010 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (IEEE)* pp 722-25