Crowd Reckoning towards Preventing the Repeat of ‘2015 Hajj Pilgrims’ Stampede’

Blinded for Review

*Abstract -* The issue of crowd reckoning has as of late gotten noteworthy consideration in the current era. Given an image of a crowded environment, the objective is to evaluate the density of the crowd, by counting the quantity of individuals that it contains. There are some approaches to solve crowd reckoning problem. They intend to reckon crowd for a medium sized place. But crowd detection at Hajj Pilgrimage site has never been tried before as far we have studied. Hajj Pilgrimage sites are different than most of the places. The aim of this work is to prevent any future stampede at Hajj Pilgrimage sites. This journal proposes two methods where head detection technique is used for the crowd reckoning. The proposed methods use Erosion Technique and Convolutional Neural Networks for head detection. Both methods are experimented using still images. The crowd reckoning will determine the possibility of a stampede and can alarm the people to take necessary steps. Assessment of the proposed models demonstrates worthy performance in realistic operating conditions.

***Keywords—Erosion, CNN.***

# INTRODUCTION

The objective of crowd reckoning, is to counting the number of people that it contains by a given video or an image of a crowded environment. Nowadays, this is certainly a great concern about human safety where overcrowded is a common scenario [1], [2]. On 24 September 2015, a crowd catastrophe caused the losses of life measurably 2,236 hajjis who were stampeded during the annual Hajj pilgrimage in Mina of Mecca, making it the deadliest Hajj tragedy ever. Estimation of the number of death may vary; the Associated Press reported 2,411 deaths. The definite reason for the congestion that led to an enormous crush on Mina is yet to be determined. To overcome this issue, we need to develop a method which can warn us of overcrowded scenario. The key idea is to count people and warn if the selected place is overcrowded. Because of the extensive application of people reckoning, most computer-based systems undergo from the following problems First, mutual barrier among people causes momentous distinctions in their presences and the loss of wrest features. It often results in a miscalculation of the quantity of individuals. Second, the difficulties initiated by images which have low resolution or blurry condition, especially for people who are far distant from the camera, generally reduce the strength of a counting system. Thirdly, women wearing hijab or people having umbrella may cause wrong counting. Finally, large variations in the appearance of individuals and lighting conditions, and in addition chaotic backgrounds, make crowd reckoning more challenging [1]. Crowd counting problem has been tried to evaluate many times. There are some approaches to solve crowd counting problem. Crowd detection at Hajj Pilgrimage site has never been tried before. In recent times, a large number of approaches have been projected for crowd reckoning. The methods proposed by Ma at el. [3] and Atnic at el. [4] intend to count crowd for a medium sized place. But Hajj Pilgrimage sites are different than most of the places. The exiting methods may provide wrong information due to large variation in the images at these sites.

This journal proposes two new methods which aim to reckon crowd at Hajj Pilgrimage sites. The goal is to prevent the repeat of 2015 Hajj Pilgrims Stampede.

Erosion Technique uses the segmentation of head regions. CNN Technique trains the CNN using several image sets of head and non-head regions. The result of crowd reckoning will determine the possibility of a stampede and can alarm the people to take necessary steps.

# Related Work

In this section, we survey the research related to the advancement of the proposed methodologies.

Ankan at el. [5] introduced a crowed counting method that can detect high density crowds. But it is not so efficient for Hajj pilgrimage sites because mutual occlusion among peoples causes loss of information.

Fu at el. [6] proposed a crowd counting approach which can only estimate crowd in low density. It captures images in low area segments but it can’t work in high density area. Huiyuan Fu [7] proposed another promising crowd counting method which uses depth camera to find the possible head regions. But it cannot detect crowd in large region. The depth camera is costly which incurs the cost overhead of the total system.

Lin at el. [8] used counting by detection method that has low-level features to detect human heads or moving substances Since the training samples are ordinarily of a high determination without impediments, the finders' execution falls apart essentially when the focused on individuals are incompletely blocked or in blur images Besides, the computational overhead of the detection stage is too high to sustain real-time responses. Erosion technique is faster because it’s time complexity is minimal. But it may over count head regions if the test image has lots of variations.

Russell Stewart [9] proposed a model that is based on decoding an image into a set of people detections. They have used a recurrent LSTM layer for sequence generation and trained their model end-to-end with a new loss function that operates on sets of detections.

However, CNN technique is more trustworthy. It gives more accurate result though it takes more computational time than Erosion technique.

# Methodology

This section presents an overview of the proposed methodology of this journal.

Crowd Reckoning is counting the number of people that it contains by a given video or an image. Two methods are proposed in this journal. The first method uses Erosion techniques. The second method trains CNN using head and non-head images. Erosion techniques are computationally lightweight, so it is faster. But the second method guarantees more accurate result. The result of our approach is then compared with the estimated number of people safe for a specific region. The estimated safe number of people for a place is predetermined by the hajj authority. Our method is then applied to that scene and thus we take decision based on the supplied information.

## *Crowd Reckoning Using Erosion Technique*

This method involves some basic image processing operations. To count the number of people in an image, we have to ensure that each head is identified individually. Fig. 1 shows the flow diagram of crowd reckoning using Erosion technique.

To accomplish the research, a sample image is taken and it is then converted to gray scale image which is shown in Fig. 2(a). The several steps of the proposed methodology are then carried out and these are outlined shortly as follows.

*Thresholding*

Thresholding is the easiest technique for image segmentation and building binary images from grayscale images*.* We apply thresholding on our image to segment head region from other body parts.

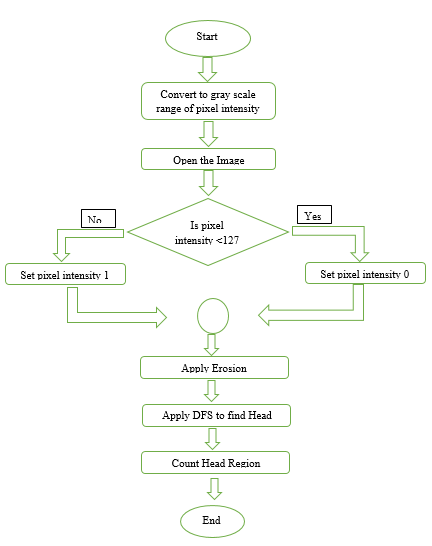
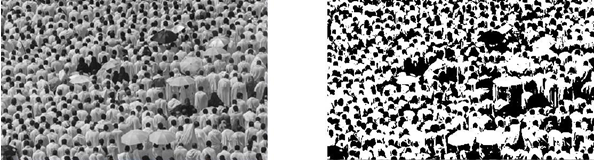


Fig 1. Flow Diagram of Crowd Reckoning Using Erosion Technique

We have chosen thresholding value 127. If a pixel intensity is less than 127, then it is considered to be a head region, otherwise it is considered background or other body parts. The resulting binary thresholding image is displayed in Fig. 3.1(b).



1. (b)

Fig 2. Thresholding a Sample Image

*Segmentation*

For dense crowds, people are standing closely to each other. That’s why it is toughto detect individual’s head. So, we applied erosion to segment head regions of individuals. Erosion is a morphological image processing technique. The key idea in binary morphology is to probe an image with a template image, conclude on how this template fits or misses the shapes in the image. Template itself is a binary image (i.e., a subset of the space or grid). It is a filtering process where image details that are smaller than the template are filtered.

If two heads are connected as shown in Fig. 3(a), then by applying erosion, we are removing those connecting lines and separate the two heads as shown in Fig. 3(b). We are using 3x3 kernel for this purpose. Assuming that the origin of this kernel is at its center, for each pixel in the image, we superimpose the origin of kernel. If kernel is completely contained by the image then the pixel is retained, else it is deleted. By applying this procedure two or three times, we get our resulted image where we can apply our head detection technique. We take a 3x3 kernel to perform erosion which will move over the image. If all the pixels inside the kernel equals to high density (binary value = 1), then the center pixel is selected, otherwise the center pixel is ignored. The ignored pixel will have value zero at the end of the process.

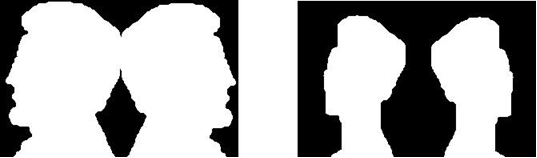
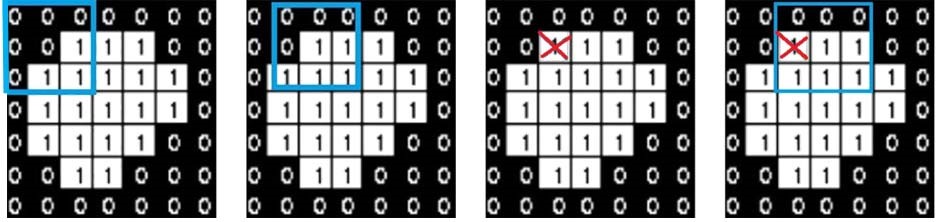
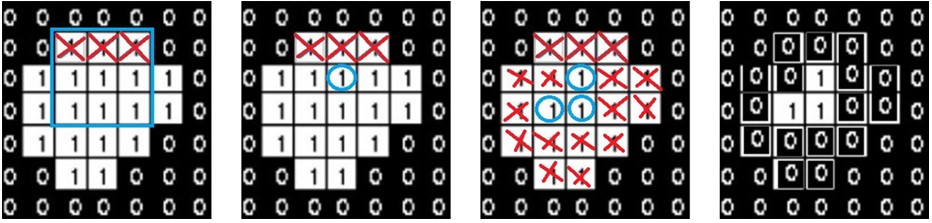


Fig 3.Segmentation of Two Heads Using Erosion.

At first, we set our kernel over the image. Now before we examine all contained pixels, we see that the center pixel is already zero, so we ignore this pixel, as shown in Fig. 4(a). Now we move the kernel to one column right. Again, we check where this kernel contained pixels, as shown in Fig. 4(b). Here not all of the pixels are of high density, so we ignore the center pixel too. We set this pixel intensity to zero. Fig. 4(c) and Fig. 4(d) follow the same scenario. After some iterations, we have a situation where all the kernel contained pixels are equal to one, as shown in Fig. 4(e). So, we select the center pixel and move the kernel to the right as shown in Fig. 4(f). Fig. 4(g) shows the image with the selected three pixels after all the iterations are completed. Fig. 4(h) is the final result where ignored pixels are set to zero.



(a) (b) (c) (d)

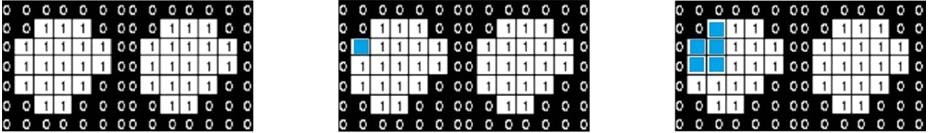


(e) (f) (g) (h)

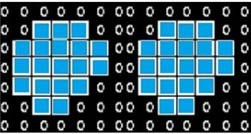
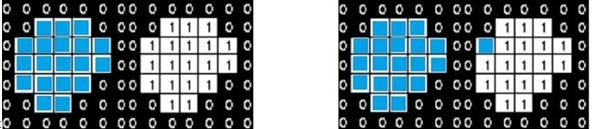
Fig 4.Working Procedure of Erosion.

Crowd Reckoning

After exceeding the erosion method, we apply head detection technique. Fig. 5 describes the process of head region counting. At first, we invert the image to make our work easier, as shown in Fig. 5(a). Then we select the first pixel as shown in Fig. 5(b) whose intensity is high, because this can be a possible head region. After that we traverse the eight connected neighbor in recursive manner. If the neighboring pixels’ intensity is also high, then we mark these pixels as visited as shown in Fig. 5(c). We terminating the recursive process for this pixel when no more neighbor pixels can be selected. So after this iteration with a pixel, we get the result as shown in Fig. 5(d). Then we calculate the area of the marked region. If this area value is not sufficient to be a predefined head region, then we discard this considering as a noise. Otherwise we increment our counter of head region.



(a) (b) (c)



(d) (e) (f)

Fig 5.Detection of Head Region Using DFS.

Now the process is again on loop with the next pixel whose intensity is high and the pixel is not visited as shown in Fig. 5(e). Finally, after traversing all the pixels in the image, the result is shown in Fig. 5(f). We have found two possible head regions with this sample image.

Then we calculate the area of the marked region. If this area value is not sufficient to be a predefined head region, then we discard this considering as a noise. Otherwise we increment our counter of head region. This process is continued for all high intensity pixels in the image.

## *Crowd Reckoning using Convolutional Neural Network*

In our second approach we trained a convolutional neural network. Our pre-trained CNN will take an input image and estimate the number of people in this image. This procedure is time expensive but it gives more accuracy.

Dataset Preparation

Our train data set consist of 1000 images of head and non-head regions of which 500 are head and 500 are non-head images. We have labeled the dataset using a 21000 matrix. The head images are labeled with 1 and non-head images are labeled with 0. Each image in the dataset is 2525 pixels.



Fig 6.Sample Images Containing Head Regions.



Fig 7.Sample Images That Do Not Contain Head Regions.

Our test data set consist of 16 images. Two sample image is given below.

Fig 8.Sample Test Images

### *Convolutional Neural Network*

Convolutional Neural Networks (ConvNets or CNNs) [10] are a category of Neural Networks that have been proven very effective in areas such as image recognition and classification [11]. Because of its classifying visual patterns such as pixel images with a very few pre-processing, it has added a new dimension in image classification tasks [12].

CNN applied here in our experiment is demonstrated in Fig 8. This network consists of two convolutional layers and two subsampling layers each following only one convolutional layer. For both convolutional layers, the kernel size remains fixed and is where in both subsampling layers, the size of the pooling area is . Following this is a dense layer, containing a linear representation of the terminal subsampled feature map’s units; those in the end are connected to the neurons in the output layer for classifying images into classes. Head regions are determined from a particular neuron in the output layer and non-head regions are determined from another neuron. When the head region neuron value outputs 1, the other neuron value goes 0 and vice versa. The kernels as well as the hidden-output weights are updated while training process continues for a defined number of epochs until gaining the desired accuracy.

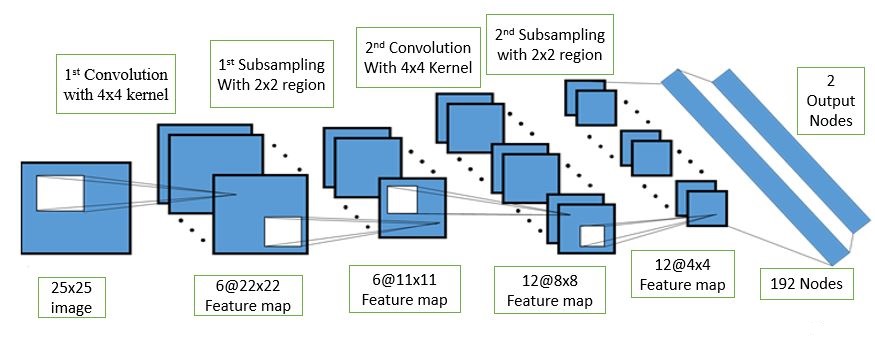


Fig 9.CNN Architecture for Classifying Dataset.

For each pixel in the input image, we encoded the pixel's intensity as the value for a corresponding neuron in the input layer. For the 25×25 pixel images we've been using, this means our network has 625 (=25×25) input neurons. We then trained the network's weights and biases so that the network's output would correctly identify the input image as 0 (non-head) , 1 (head).

### *Head Counting*

 At first, we take a sample image and convert it to gray scale.





Fig 10.Gray Scale Conversion

Now, we slide over the image and obtain a 25x25 window which will be tested using the model we just trained.



(a)



(b)

Fig 11.Extracting 25x25 Window

The output of our neural network will be either 1 or 0, which will determine if the window contains a head or not. If the window has a head in it, we increment our head\_counter. Then, again we move our window to the right. Again send it to the model to test the window image.

## *Prevention of Pilgrims’ Stampede*

The proposed methods process still images collected from pilgrimage sites and then estimate the number of people in a scene. But the number of people safe for a place cannot be fixed. It varies from place to place.

The Hajj Institution will provide information about the threshold number which means the maximum number of people safe for the particular region. To detect whether the situation is out of control or not, the number of people in a scene is needed to be checked with the threshold value. If there are more people than this safety number, then the place is overcrowded. In that case, the concerned persons are notified and they will take necessary steps. Some of the steps are as follows:

* Closing entry points
* Opening evacuation points

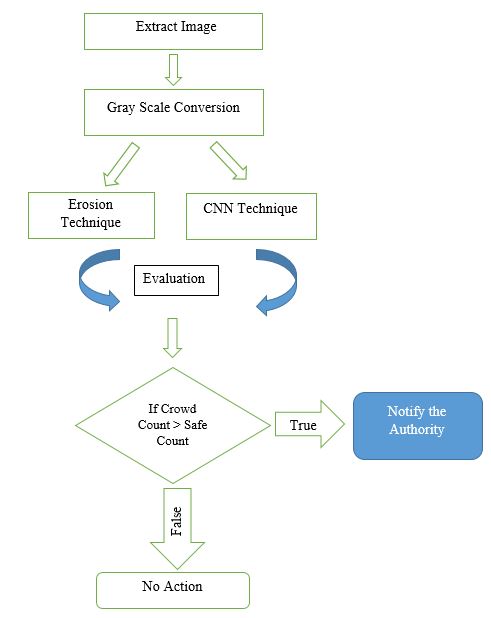


Fig 12. Flow Diagram of Stampede Prevention

# Experimental Analysis

In this section, we discuss about experimental setup and overall analysis of our proposed methods. The outcome of the implementations of both methods are discussed here briefly.

1. *Erosion Technique*

Erosion technique was implemented in C++ with open cv library on a standard PC with Intel core i3 CPU with 4 GB memory.

From (Fig. 13[)](#page8) we can see that at our first iteration, the image has a lot of noise elements indicated by big orange circles. Then the next iteration (Fig. 13), we search for head regions, and count them. In Fig. [13](#page9) and Fig. 14, small red circle indicate individuals head regions and big orange circles indicate noise elements.



Fig 13. 1st Iteration Image Contains too Much Noise.

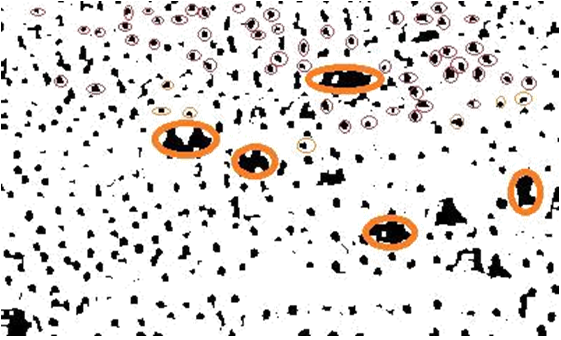


Fig 14. Noise Element Marked on 2nd Iteration

In Fig. 15 it is clear that Head detection result on 1st iteration is long distant from original count due to heavy noise element. But 2nd and 3rd iteration is showing promising results. So we take the average of them as our estimated original count.



Fig 15. Noise Element Marked on 3rd Iteration

From the output image we calculate Precision and Recall where precision is the fraction of retrieved instances that are relevant and recall is the fraction of relevant instances that are retrieved.

1. (b)

Fig 16.(a) Input Image (b) Result of Erosion Technique

…………………………… (1)

…………………………………(2)

If a true head is perfectly identified then it is counted as True Positive (*tp*). But if a non-head region is identified as a head region, then it is called False Positive (*fp*). Also, there is False Negative (*fn*) associated with the output which occurs if our algorithm misses a head region.

In Fig. 16(b), two or more head regions are connected that’s why the output image failed to count individual heads. These connected heads are counted as one region.

Table 4.1: The Precision-Recall Observed for Different Sample Images Using Erosion Technique

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Image no** | **Ground Truth** | **Erosion Technique** | **Precision** | **Recall** |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 8 | 14 | 0.57 | 1 |
| 3 | 18 | 20 | 0.65 | 0.72 |
| 4 | 10 | 10 | 0.6 | 0.6 |
| 5 | 27 | 25 | 0.8 | 0.74 |
| 6 | 47 | 47 | 0.7 | 0.7 |
| 7 | 19 | 25 | 0.56 | 0.74 |
| 8 | 19 | 22 | 0.6 | 0.68 |

1. *CNN Technique*

We have applied CNN on the resized and normalized grayscale image files. The experiment has been conducted on HP pro laptop machine (CPU: Intel Core i3 @ 3.50 GHz and RAM: 4.00 GB) in Window 10(64bit) environment using Matlab R2013a. Experimental results using the proposed recognition scheme have been collected based on the samples of the prepared dataset discussed earlier. Our test image set contains 16 images. The result using our CNN is more accurate than Erosion technique.

1. (b)

Fig 17.(a) Input image, (b) Result of CNN Technique

In Fig. 17(b), the head regions are identified and marked with a rectangle. But in some cases some head regions are missed and some are multiple counted. But this method is able to count head regions more precisely.

Table 4.2: The Precision-Recall Observed for Different Sample Images Using CNN Technique

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Image no** | **Ground Truth** | **CNN Technique** | **Precision** | **Recall** |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 8 | 14 | 0.57 | 1 |
| 3 | 18 | 22 | 0.68 | 0.83 |
| 4 | 10 | 9 | 1 | 0.9 |
| 5 | 27 | 29 | 0.83 | 0.89 |
| 6 | 47 | 51 | 0.8 | 0.87 |
| 7 | 19 | 20 | 0.95 | 1 |
| 8 | 19 | 21 | 0.86 | 0.95 |

Fig. 18 showing the result of applying three different procedures along with the actual count.

Along the y axis, the number of individuals showing the differences between the results obtained from different methods. Without applying Erosion it is difficult to detect individual head region accurately. After applying Erosion the system produces more accurate results.

Fig 18.Crowd Counting Results for 16 Images

Fig. 18 indicates that CNN technique is closer to ground truth. Both precision and recall value is important in people counting results. Fig. 19 shows the Precision-Recall chart for our two proposed methods. Erosion technique is not able to accurately identify all the true head regions. But CNN technique is more successful in detecting true positives and also the precision is higher than Erosion technique. It is also evident that CNN technique has higher recall value than Erosion technique.

Fig 19.Crowd Counting PR Chart for Erosion Technique

Fig 20.Crowd Counting PR Chart for CNN Technique

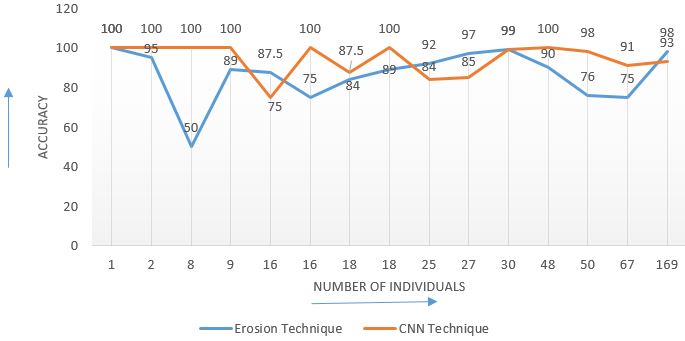


Fig 21.Accuracy Curve for Proposed Method

The accuracy is measured by taking the deviation of the proposed methodology from ground truth.

Fig. 21 shows the accuracy curve for test images using the proposed methods. Both methods show promising results but CNN technique is more accurate. The average accuracy of Erosion technique is 76%. On the other hand, the average accuracy of CNN technique is 94%.

CNN technique has a major drawback. To train a network and test images on this network is time consuming. Generally to train a CNN using 1000 test images, it takes a few hours. The procedure for counting head regions also involve checking all possible rectangular regions. Thus the time complexity increases. But Erosion technique is faster because the method Erosion has low time complexity.

# Conclusions

We consider the application of proposed method based on some preconditions. Since the Erosion technique is very simple, it can be applied for real time head counting in pilgrimage sites to prevent any possibility of a stampede. To get more accurate result, CNN technique is preferable. We evaluate our methods with respect to ground truth and the experimental results are convincing.

The cameras on the Hajj Pilgrimage sites are at a fixed distance above the ground. The head shapes are almost of same sizes so the detection step becomes easy. But if the camera position were not fixed, then the head shapes would have been of different sizes and thus it would have been difficult to identify them. So in order to detect head regions irrespective of image source the proposed methodology needs to be adjusted.

# References

1. Nick C. Tang, Yen-Yu Lin, Ming-Fang Weng, and Hong-Yuan Mark Liao, “Cross-Camera Knowledge Transfer for Multiview People Counting,” *IEEE Transactions On Image Processing*, Vol. 24, No. 1, January 2015.
2. B. Liu and N. Vasconcelos, “Bayesian Model Adaptation for Crowd Counts,” 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 4175-4183.
3. H.D. Ma, C.B. Zeng, and C.X. Ling, “A reliable people counting system via multiple cameras,” *ACM Transactions on Intelligent Systems and Technology*, vol. 3,pp. 1-22, 2012
4. B. Antic, D. Letic, D. Culibrk, and V. Crnojevic, “K- means based segmentation for real time zenithal people counting,” *IEEE International Conference on Image Processing,* pp 2565-2568, 2009.
5. Ankan Bansal and K S Venkatesh, “People counting in High Density crowds from still images*,” IEEE International Conference on Computer Vision and Pattern Recognition*, pp 1093-1100, 2009.
6. Huiyuan Fu, Huadong Ma, Hongtian Xiao, “Crowd Counting via head detection and motion ow estimation,” *IEEE International Conference on Image Processing*, 2014.
7. Huiyuan Fu, Huadong Ma, Hongtian Xiao, “Real-time accurate crowd counting based on rgb-d information,” IEEE International Conference on Image Processing, pp 2585-2568, 2012.
8. S.-F. Lin, J.-Y. Chen, and H.-X. Chao, “Estimation of number of people in crowded scenes using perspective transformation*,” IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 31, no. 6, pp. 645–654, Nov. 2001.
9. Stewart, Russell, Mykhaylo Andriluka, and Andrew Y. Ng. “End-to-end people detection in crowded scenes.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
10. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. “Imagenet classification with deep convolutional neural networks.” In *Advances in neural information processing systems*, pp. 1097-1105. 2012.
11. Simonyan, Karen, and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition.” *arXiv preprint arXiv:1409.1556* (2014).
12. Lawrence, Steve, et al. “Face recognition: A convolutional neural-network approach.” *IEEE transactions on neural networks* 8.1 (1997): 98-113