

# Real-Time Crowd Density Mapping using a Novel Sensory Fusion Model of Infrared and Visual Systems

S. Yaseen<sup>1,a</sup>, A. Al-Habaibeh<sup>1,b\*</sup>, D. Su<sup>1,c</sup> and F. Otham<sup>2,d</sup>

<sup>1</sup>Advanced Design and Manufacturing Engineering Centre, School of the Built Environment,  
The Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, UK

<sup>2</sup>CTHM Institute for Hajj Research, Um Al-Qura University, Makkah, Saudi Arabia

<sup>a</sup> saiedyaseen@ntu.ac.uk, <sup>b\*</sup> Amin.AL@ntu.ac.uk, <sup>c</sup> d.su@ntu.ac.uk, <sup>d</sup> fadelothman@gmail.com

**Keywords:** Crowd safety, Density mapping, Sensor fusion, Infrared thermography, Real-time monitoring, Intelligent buildings.

**Abstract.** Crowd dynamic management research has seen significant attention in recent years in research and industry in an attempt to improve safety level and management of large scale events and in large public places such as stadiums, theatres, railway stations, subways and other places where high flow of people at high densities is expected. Failure to detect the crowd behaviour at the right time could lead to unnecessary injuries and fatalities. Over the past decades there have been many incidents of crowd which caused major injuries and fatalities and lead to physical damages. Examples of crowd disasters occurred in past decades include the tragedy of Hillsborough football stadium at Sheffield where at least 93 football supporters have been killed and 400 injured in 1989 in Britain's worst-ever sporting disaster (BBC, 1989). Recently in Cambodia a pedestrians stampede during the Water Festival celebration resulted in 345 deaths and 400 injuries (BBC, 2010) and in 2011 at least 16 people were killed and 50 others were injured in a stampede in the northern Indian town of Haridwar (BBC, 2011). Such disasters could be avoided or losses reduced by using different technologies. Crowd simulation models have been found effective in the prediction of potential crowd hazards in critical situations and thus help in reducing fatalities. However, there is a need to combine the advancement in simulation with real time crowd characterisation such as the estimation of real time density in order to provide accurate prognosis in crowd behaviour and enhance crowd management and safety, particularly in mega event such as the Hajj. This paper addresses the use of novel sensory technology in order to estimate people's dynamic density during one of the Hajj activities. The ultimate goal is that real time accurate estimation of density in different areas within the crowd could help to improve the decision making process and provide more accurate prediction of the crowd dynamics. This paper investigates the use of infrared and visual cameras supported by auxiliary sensors and artificial intelligence to evaluate the accuracy in estimating crowd density in an open space during Muslims Pilgrimage to Makkah (Mecca).

## 1. Introduction

The Hajj, the Muslims pilgrimage to Makkah (Mecca) represents one of the largest annual human gatherings. Millions of people from all over the world with diversity of cultures and languages congregate to perform religious rite within a specific area and during several days. This activity is made more complex by the arranged movement of high density crowd between locations and the performance of rituals at pre-determined times which occur as part of the Islamic lunar calendar. As a result, Hajj activities include a combination of crowd dynamics including low-density unmotivated movement and high-density, driven movement (Siddiqui and Gwynne, 2012). There has been a number of safety related incidents between 1994 and 2006, see (Ministry of Hajj, 2011; AlGadhi and Still, 2003). Previous research in this area suggests pilgrim safety can be better protected and lessons that

---

\* Corresponding Author, Amin Al-Habaibeh, email: [amin.al@ntu.ac.uk](mailto:amin.al@ntu.ac.uk)

can be learned and applied more generally in other events and activities; for further details regarding Hajj activities please refer to (Siddiqui and Gwynne, 2012).

A number of studies of the Hajj crowd density and movement have been conducted, see for example (AlGadhi and Still, 2003), primarily characterising the flow-density relationships generated: these relationships were rightly seen as major contributors to the incidents that occurred. In addition, there have been several attempts at simulating the pedestrian conditions present during the Hajj in order to investigate mitigation opportunities or applying tools to better describe the make-up of the crowd, see for example (AlGadhi and Still, 2003; Koshak and Fouda, 2008; Shendarkar et al., 2008; Zainuddin et al., 2009). However, little analysis has been performed relating to real time sensory estimation of crowd density for improved confidence in the overall prognosis produced in real time to prevent accidents. The limitation of simulation algorithms is that it deals with static assumptions, estimations and off-line scenarios, which could help in designing the physical constraints of the space and understanding the generic behaviour in some situations. However, crowd management in real time would require accurate and real time estimation in different locations to allow intervention before the occurrence of any critical situation. The ultimate objective of this work is to enhance the real time decision making process from crowd density by accurate crowd density mapping. An unexpected incident or behaviour within the crowd, see for example the observations made by Siddiqui and Gwynne (2012), could result in significant change in the crowd dynamics leading to a higher and unexpected risk. Therefore, real-time density estimation and mapping could help in addressing such drawbacks, avoid such situations and provide more accurate results.

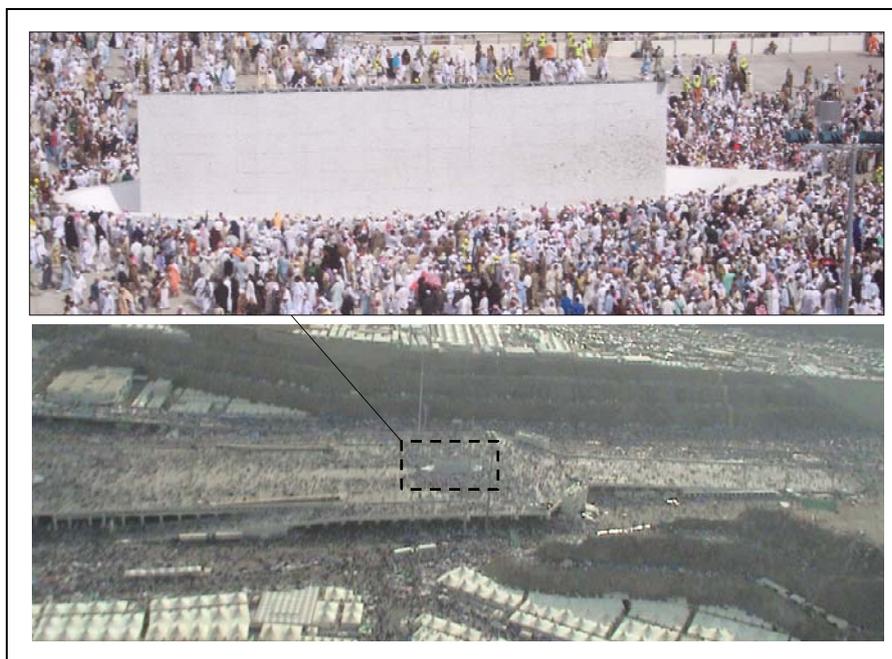


Figure 1: The Old Jamarat Bridge in Makkah during Hajj with variety of density levels.

The authors have presented several technologies to help in predicting the density of people, see for example (Al-Habaibeh, et al., 2010; Al-Habaibeh, et al., 2003; Amin, et al., 2006). To explain the context of this research, Figure 1 presents an example of crowd density during Hajj on the old Jamarat bridge. The ultimate objectives of this research is to develop a density map of the crowd in different

locations in real time in order to improve crowd management and predict any potential incidents or unusual density levels. This should help, in addition to the simulation technology, to provide more accurate prognosis.

Research work has been done in recent years towards real-time density mapping and crowd behaviour. Lee and Hughes (2006) present prediction system of human crowd pressures. Utilising the standard forward-backward autoregressive modelling techniques for spectral analysis of a measured signal, predictions of pressures generated by very high densities of pedestrians have been formulated. The study suggests that several minutes are available for corrective action to be taken to avoid an accident involving crushing of pedestrians. Crowd monitoring using image processing has been presented by Davis et al. (1995) where image processing techniques of visual images for data collection and on-line monitoring of crowds have been presented and discussed with some success.

An example of video-based people counting systems is described by Sidla et al. (2006). They have shown a system uses motion to compute regions of interest and prediction of movement extracting shape information from the video frames to detect individuals, and applies texture features to recognise people. Terada et al. (1996) have proposed an automatic method of counting the passing people through the gate by using a normal CCD camera. First, images of the passing people are obtained in series. As a next step, pixels on two measurement lines are picked up and transformed into two space-time images. Finally, the incoming and outgoing persons could be measured by counting the people data on the space time images of each moving direction. Velastin et al. (2006) have developed a people detection system to detect important situations of interest to public transport operators. This has been done by using a monochrome video camera and motion detection board developed by the authors. The analogue images then converted into digital images. The system then feeds these images onto motion detector to carry out further analyses, namely noise reduction, foreground extraction and event detection. Septian et al. (2006) have shown an investigation to counting number of people that pass the view of an overhead mounted camera. Moving people are first detected as blobs and represented by binary masks, based on which possible multi-person blobs are further segmented into isolated persons according to their areas and locations. Each single person is tracked through consecutive frames using a correlation-based algorithm. In recent years, Kilambi et al. (2007) have shown a system which can aid in monitoring people crowd, which often contains tight groups of people using prior knowledge obtained from the scene and accurate camera calibration, the system learns the parameters required for estimation. There are also systems using multiple camera to count people, see for example, Kettner and Zabih (1999), and Challa et al. (2007). Another reference, Sacchi et al. (2001), has used a change-detection algorithm for people counting. The proposed algorithm has been found capable of exploiting the low computational complexity of the pixel-based methods and, at the same time, the adaptively involved by the image partition into different luminosity regions, which is the key concept of the local methods. Masoud et al. (2008) have presented a method for the estimation of crowd densities. They use novel projection methods to compute the area that represents each blob in world coordinates. This method reduces problems due to moving objects which are not similar to human height. It also takes care of the differences between the projected areas of people at different distances from the camera. Roqueiro and Valery (2007) have presented a method of estimating the number of people visible in a camera. The method is based on the ratio of foreground pixels in compare to the background pixels in each cell of a rectangular grid. Using the foreground pixels and data mining techniques allowed reaching accuracy up to 85% for an indoor surveillance environment and can count up to 23 persons. Applying median filters to the sequence of estimation results increased the final accuracy up to 91% for exact match. Celik et al. (2007) have proposed a research where they relate proportionally the number of moving pixels to the number of persons. By extract the foreground object. According to the adopted Stauffer- Grimson adaptive background method, every pixel probability density function is modelled by means of a Gaussian Mixture Model.

There are also systems using multiple camera to count people such as Kettner and Zabih (1999), Challa et al. (2007) and Yang et al. (2003). Yang et al. (2003) have shown an alternative approach of counting people that directly estimates the number of people using groups of image sensors. It has developed a people counting system in crowds based on the computation of bounds on the number of objects in a region and not on localizing individual objects. The system consists of 8 calibrated cameras arranged around a rectangular room. The image processing technique has been based on background subtraction. Hashimoto et al. (1998) have presented people counting system with a multi-element infrared sensor made of  $\text{PbTiO}_3$  ceramics in order to establish non-contact and non-blocking automatic real-time counters with high accuracy. They have described a people-counting system using a human information sensor consisting of one piezoelectric array detector, which can be used for detecting the movement directions and the number of people passing a 100 cm then a 200 cm wide door.

This paper address some of the drawbacks in the current knowledge and technologies by investigating the use of a sensor fusion system for monitoring people's density in one of the Hajj activities. The demand for intelligent and integrated monitoring systems of people has significant importance during *Hajj*. The experimental work for this paper has been done on the old Jamarat bridge in Saudi Arabia in 2006 which has been demolished to build the new multi-story high capacity Jamarat bridge. The old Jamarat bridge has seen high levels of density during its operation and significant and unique data has been captured to investigate different techniques for density estimation for future use on other Hajj sites .

## **2. Infrared and Visual Imaging Systems**

The majority of people counting systems in the market use visual cameras, see for example, Rooqueiro and Petrushin (2006), Cho and Chow (2002). It has been found that the available systems are mainly expensive vision systems that need to be calibrated in advance. Nevertheless, visual automated counting systems can only work in the presence of a good intensity of light and the change in ambient light would influence the results of the system. A major problem with vision-based monitoring systems is the inaccuracy associated with visual detection of people. If a person is wearing the same shades of grey as the background it is difficult to discriminate between the background and the person. These objects in the background are of the main concerns, commonly affected the accuracy of the automated people counting systems. By all means it can be said that background separation is not an easy task. Also, counting people based on image processing and number of pixels is not applicable in the case of high density of people as occlusion would occur, which makes the relation between the number of pixels and the number of people in the selected region of interest (ROI) nonlinear.

There are many techniques for people detection such as infrared sensors and the ultrasonic sensors, however, each system has its own problems. For the thermal imager, it is easy to identify the existence of a person from a thermal image compared with a visual image. In spite of infrared detectors being very suitable as human information sensors, conventional point-type infrared sensors have low accuracy for counting the number of people passing a door. For advanced control of the thermal environment using infrared sensors, the overall conditions of the occupants in a room, such as temperature and light, should be taken into account.

To date, the use of thermography has been limited due to the high cost of instrumentation. The advent of low cost sensing systems is opening up many new opportunities for condition monitoring and the development of intelligent machines and processes (Al-Habaibeh et al. 2004). For the thermal imager, it is easy to identify the existence of a person from a thermal image compared with a visual image

(Hashimoto et al., 1998) . Espina and Velastin (2005) stated that "there is not much work reported on the integration of different types of sensors". The infrared technology has also been found to be lagging behind in current industrial and academic research and it could provide essential solutions to some current recognised problems of crowd monitoring. The introduction of low-cost infrared systems could play an important role in reducing the cost of the integration of infrared systems to crowd monitoring.

### **3. The use of neural networks for crowd monitoring**

The main advantage of using neural networks is the full automation of the learning and classification processes, therefore, they can be implemented in fully automated condition monitoring system to recognize and classify patterns without human involvement.

Neural Networks (NN), also known as Artificial Neural Networks (ANN), consist of “n” numbers of computing cells namely artificial neurons, which connected together through weighted links ( Ghosh and Chakravarthy, 1992). Neurons are connected together to form an input layer, hidden layers and an output layer. Neural Networks are classified by their learning method to: supervised learning and unsupervised learning. In supervised learning the error is fed back to the network and used to adjust the weights, such error decreases with each iteration and the neural model gets closer and closer to producing the desired output. This process is known as “training”. In contrast to supervised learning, unsupervised learning does not require a specific target for training; it requires guidelines to determine how it will classify groups. Grouping may be based on colour, shape, material or on some other property of the object. The most common supervised neural networks are (BP) network , and the radial basis function (RBF) neural network. (BP) and (RBF) can be use for many applications such as: temperature control, fault detection , monitoring of flow rate (Meireles et al., 2003), and hydrocarbon density estimation ( Al Quraishi and Shokir, 2011). Unsupervised neural networks include examples such as LVQ Learning Vector Quantization and self organizing neural networks.

Researchers have used neural network for people density and counting activities. For example, Yoshinaga and Shimada (2010) have analysed a combination of blob area, total edge length and perimeter length using Resilient back propagation neural network (RPROP) to estimate the number of pedestrians. Chow and Cho on (2002) and (2004) have analysed a combination of number of pixels of edge and number of pixels from crowd objects using a feed forward neural network. The systems perform good results in estimation the crowd density. However, these types of systems cannot be use in large public places like Jamarat site, where there are thousands of people at extreme high densities. Hussain et al., (2011) have shown an alternative approach of estimating the density of people using a combination of image processing and artificial intelligence technologies. A back propagation (BP) neural network used to estimate the number of people. The number of people is then used to calculate the people density and classify the people density based on the area region of interest. The system has performed good results in estimating the people density which range between 1 and up to 60 persons, achieving 81.67% accuracy of true detection with 78 images used in the training phase. In this research work, back propagation (BP) neural network is used to estimate the density of people. Using the area of consideration, the crowd density is calculated and classified into four densities from very low to extremely high.

### **4. The Experimental Work**

The development of an autonomous monitoring system technology is essential to provide the means for capturing crowd data and integrating information for the decision making. Figure 2 presents the

autonomous monitoring system that is developed for crowd monitoring for this research work. The system includes a monochrome visual camera attached to a small computer system via a video capture card type National Instruments NI PCI-1407. A low cost infrared camera type IRISYS IRI1002 is connected to the same computer system using an RS232 connection. Temperature and light sensors, designed by IENSYS Ltd, are interfaced to the computer system via IENSYS board to the RS232 port of the system. The computer system is Ethernet enabled so that data can be transferred to remote central analysis point as described in Figure 2.

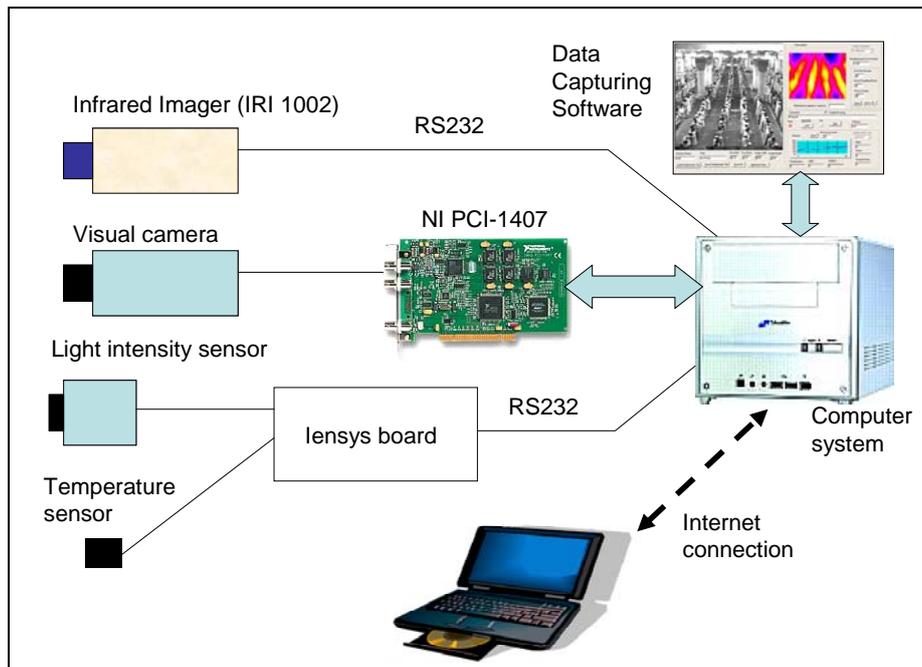


Figure 2: The autonomous data capturing system

The Experimental work of data capturing described in this research has been performed at the old Jamarat bridge in Saudi Arabia which has been removed following this experimental work to be replaced by much more sophisticated multi-story bridge. However, the old bridge provided a unique opportunity to capture data representing extreme levels of densities that is now being analysed to design future monitoring systems for other sites for Hajj. The system is installed to view people from a high corner as shown in Figure 3. Data is captured for different densities of people over several days during Jamarat Hajj activities. Every 30 seconds a frame of data is captured and saved to the computer for analysis. The selected view is realistic and includes many sources of noise such as the sunlight, temperature and shadow. Figure 3 presents the infrared image and the associated visual image at the old Jamarat bridge. The idea of using infrared images to capture density information is related to the concept that every person is considered a heat source and the more people you have in a specific area the more heat (i.e. infrared radiation) is produced. Notice that the infrared image is “warm” in a specific linear pattern based on the density of people. Point (1) presents the cold Jamarat wall, point (2) presents high density of people and points (3 and 4) present slower level of density. It is essential to recognise that density colour-map is only valid for regions associated with crowd, other areas such as point (1) is related directly to the temperature of the object. One of the common problems is that people closer to the camera will look warmer and larger in size (i.e. will occupy more pixels). People in the far-side of the site will look smaller and at less temperature level. Also the sunlight will generate a noise on the infrared images. A calibration and analysis process therefore will be needed to normalise

the data based on the geometry of the view and to compensate the noise of the sunlight on the infrared data. Typical crowd densities are presented in Figure 4. In general, it is estimated that an extremely high density is 5 or more person/m<sup>2</sup>, high density is 4 person/m<sup>2</sup>, medium density is 3 person/m<sup>2</sup>, low density is 2 person/m<sup>2</sup>, and very low density is expected to be 1 or less person/m<sup>2</sup>. Above 6 person/m<sup>2</sup> is extremely dangerous and may lead to an accident if maintained for a long time. However, due to the size of the covered area, its open nature and the number of people involved, a qualitative judgment has been used to categorise the different densities.

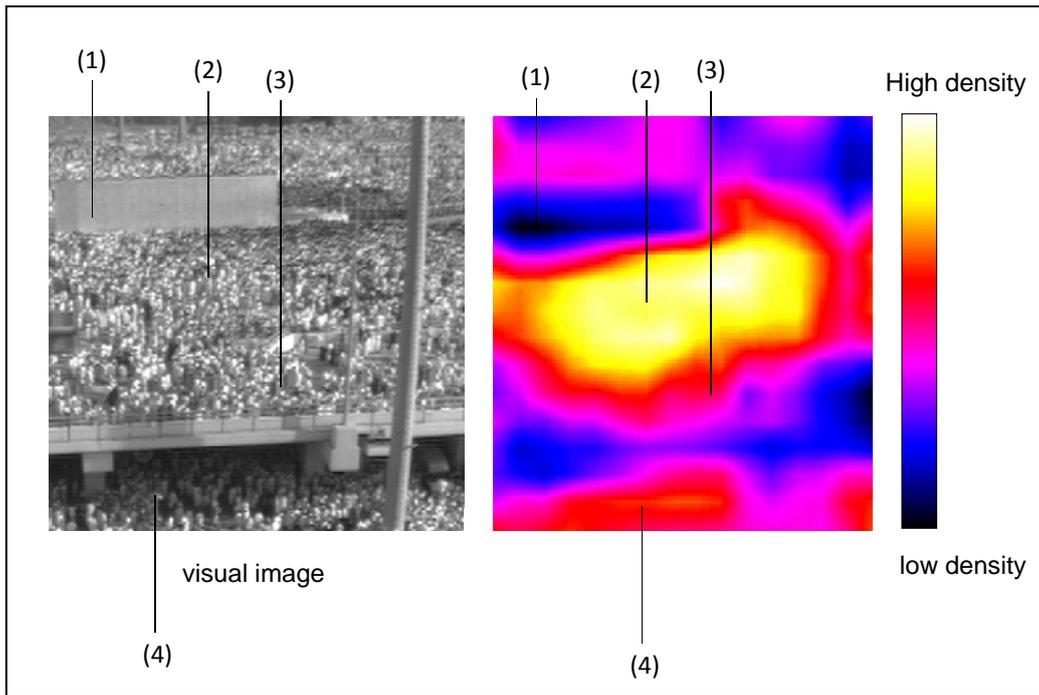


Figure 3: Examples of the images captured.

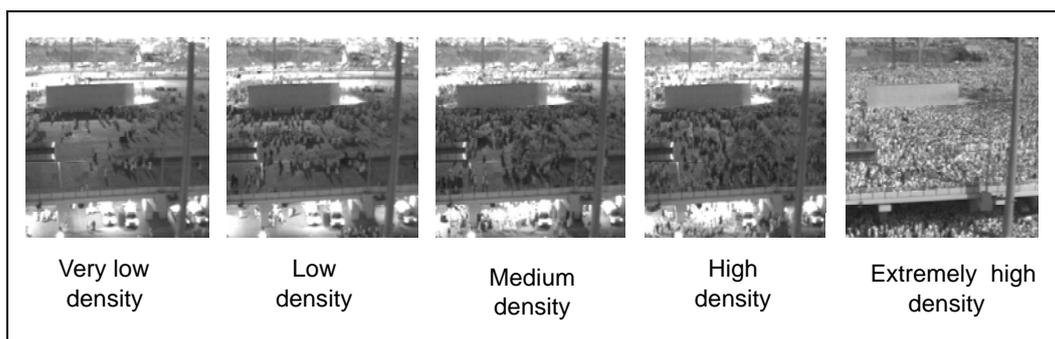


Figure 4: The crowd densities estimated by human scene are represents five different densities.

This paper investigates five main set of experiments:

- The use of visual images for estimation of density.
- The use of visual images with neural networks to estimate the density.
- The use of infrared images to estimate the density.
- The use of infrared images with neural networks to estimate the density.

- A sensor fusion system using a combination of visual and infrared images combined with neural networks to estimate the density.

## 5. Visual Image Processing and Analysis

The first set of experiments focuses on the analysis of visual images by applying image processing methods and analysis of the raw data in an attempt to predict the density of people. The visual images have many sources of noise such as: sunlight intensity variation, temperature and shadow. For that a calibration process will be needed to compensate for this noise. In this proposed system, the results from the feature extraction process is a number that represents the crowd density of the corresponding image. Figure 4 presents the original images of the Jamarat site at different densities and different times of the day. The block diagram shown in Figure 5 presents the general feature extraction system implemented in this paper for density estimation for visual images. As shown in the block diagram, the regions of interest are selected from the original image with reduction in number of pixels to speed up the computational time depending on the required accuracy and application. The edge detection process is implemented in order to enhance and detect sharp changes in image brightness that is related to the people in the image. Edges are those places in an image that correspond to object boundaries. Edge detection also reduces the size of the memory needed for the following image processing steps. All pixels values smaller than this threshold value will be eliminated and viewed as a background image. This process is very important because it will eliminate too much pixels and data which lead to decrease the data processing time and increase the system efficiency. Most of the eliminated pixels did not relate to the density of people, for example, some of these pixels represent shadows of the objects. The number of pixels in the outcome image represents the density of people in the image. Light compensation is used, as discussed in the next section, to reduce the effect of light noise on the predication process during day time. The next step is to subtract the background image from each individual image for the same time of the day. The subtraction will result in a binary image, by counting the number of pixels in each image this will indicate the change between the current and the background image which is as a result represent the density of people in an image in most cases. Pixel counting is used which is then transferred to a density estimation of five different levels of density as presented in Figure 4. Figure 6 presents an example of the edge detection and background subtraction of the images.

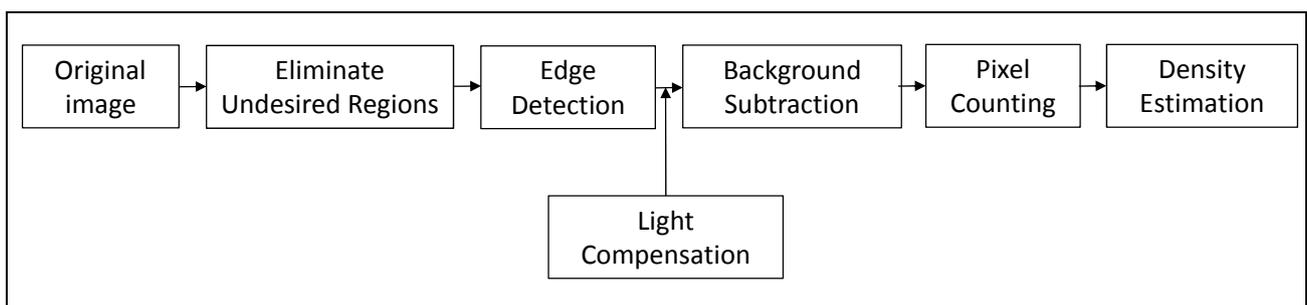


Figure 5: The block diagram of visual image processing.

Pixel counting  $N_{pixels}$  of ‘white’ pixels, see Figure 6, has been found, as will be discussed later in this paper, a useful technique for predicting people’s density. The calibration of number of pixels  $N_{pixels}$  in relation to the density of people is expressed by equation (1) below:

$$Density = \frac{(N_{pixels} - \alpha)}{\beta} \quad (1)$$

Where  $\alpha$  is a constant which is a function of the geometry of the public place including distance and the angle of the camera; and  $\beta$  is a constant factor related to the assumed density of people under consideration. Equation (1) ignores the effect of noise from the change in light intensity.

The original captured visual images are 344 x 344 pixels. However, a reduction in the image size to 256 x 256 pixels is found to decrease significantly the required processing time by approximately 55% without significantly reducing the results from the image processing in this case.

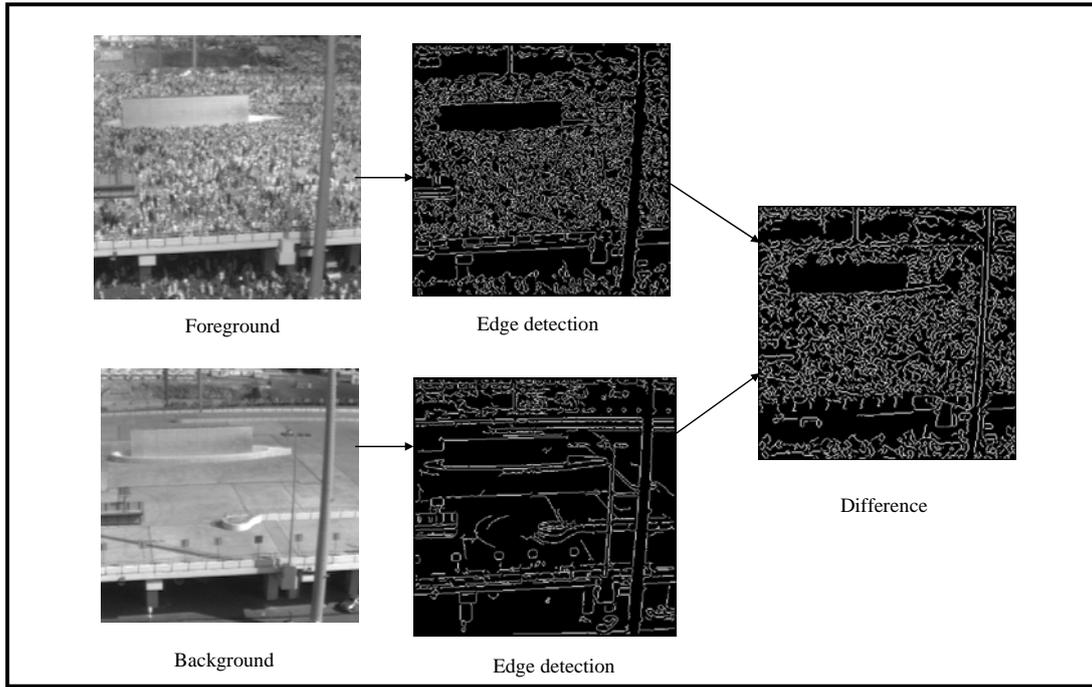


Figure 6: Example of edge detection and background subtraction of visual images.

### 5.1 Light compensation for visual image processing

When monitoring the edge detection algorithm for the empty location, it has been found that light intensity and shadows has significant influence on the accuracy of results. Therefore, sensor fusion of visual images combined with the data from the light intensity sensor are used to compensate for this error. Equation (2) presents the methodology used to compensate for light variation.

$$N_{\phi} = N_{pixels} - C x \quad (2)$$

Where  $N_{\phi}$  is the compensated number of pixels above a specific threshold value,  $N_{pixels}$  is the initial number of pixels,  $x$  is the light intensity, and  $C$  is a calibration constant. Therefore,  $N_{\phi}$  will replace  $N_{pixels}$  in equation (1) to have the general equation for light compensation as described in equation (3).

$$Density = \frac{(N_{\phi} - \alpha)}{\beta} \quad (3)$$

In order to study the effect of the sunlight on the images and the associated image processing, a total of 2880 images that present 24 hour period is monitored during one day before the use of the Jamart bridge. In these images the background scene contains no people, while these images under a variety of sunlight and shadow conditions. Figure 7 presents the results of the light compensation system. As shown in Figure 7, the number of pixels successfully present that very low density from midnight until morning. From morning until sunset noise has been detected as a result of light and shadow change, this is clear in the second curve in Figure 7. After sunset, the system accurately detects no objects. In order to eliminate noise from the light, equation (3) has been implemented by utilisation the light intensity curve in the third graph of Figure 7 to produce a much improved results in the fourth curve of Figure 7. Quantised Density  $QD$  estimation is presented in Equation (4).

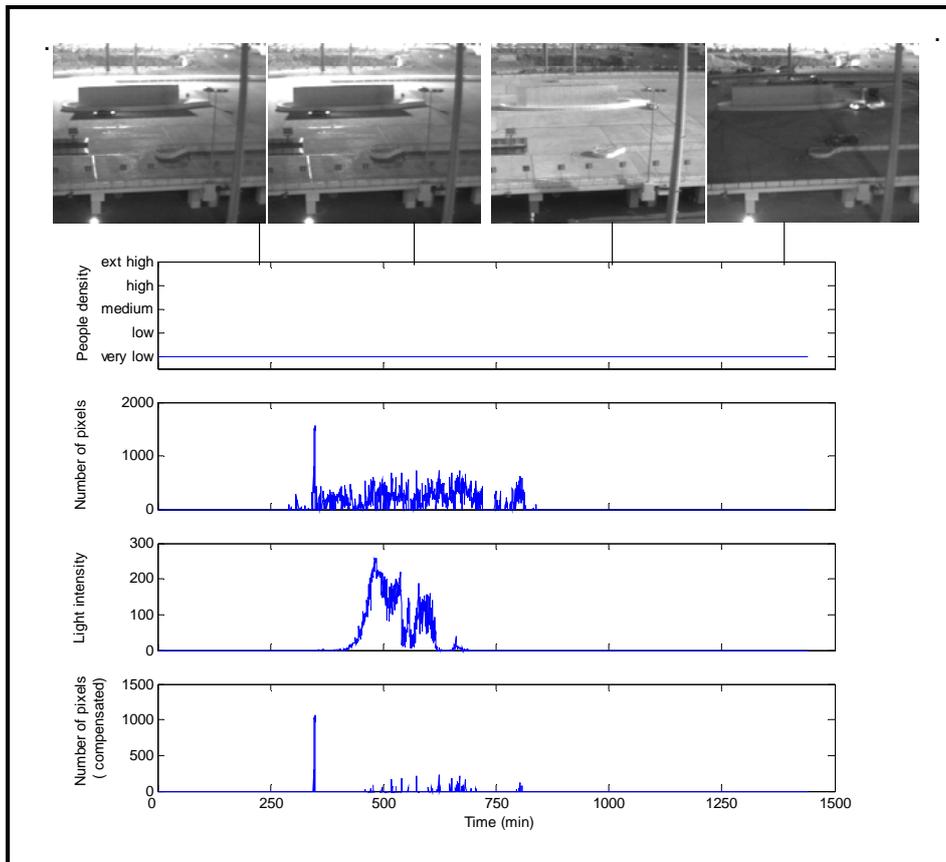


Figure 7: The effect of the sunlight on the counting system.

$$Quantised\ Density\ QD = 1 + 4 \left[ \frac{Density - \lambda}{\psi - \lambda} \right] \quad (4)$$

Equation (4) will result in a Quantised Density  $QD$  with a number between 1 and 5, where one is very low density and 5 is extremely high density. Where  $\lambda$  is the minimum Density level observed and  $\psi$  is the maximum Density level observed.

## 6. The use of visual images for estimation of density.

Figure 8 presents an example of the captured data for a full day. The first graph, graph (1), represents the density of people in each image that is captured every 30 seconds period, the people density is estimated manually by an experienced person in order to be used as the benchmark for the automated results. The visual images present different levels of density during different times of the same day.

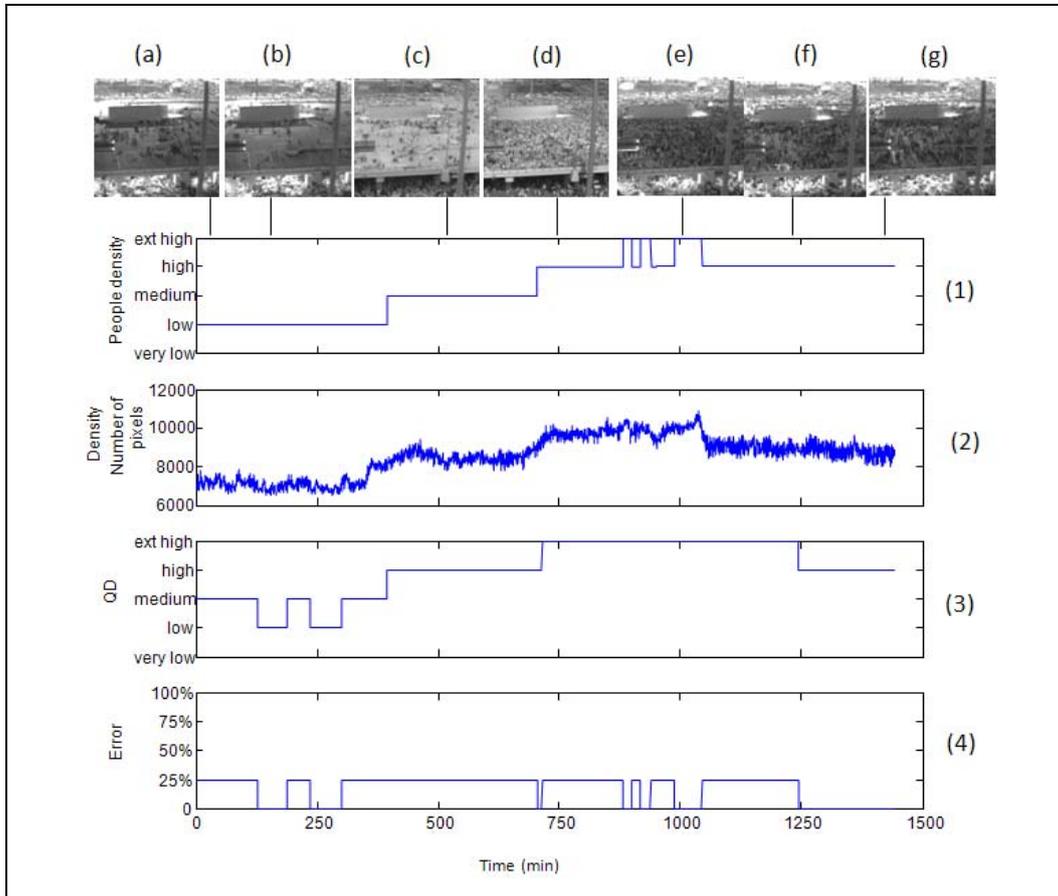


Figure 8: Density estimation results using visual images and combined with light intensity algorithm.

The second graph, graph (2), represents the density as number of pixels as obtained from equation (3) following the edge detection system. The third graph represents the Quantised Density (QD). The error, the difference between graph (1) and graph (3), is presented in the fourth graph. Notice that the error is within 25% which is one density category within the estimation of density. Notice that the QD of equation (4) method have some error in discriminating between high and extremely high density.

## 7. The use of visual images with neural networks to estimate the density.

In order to investigate the effect of using artificial intelligence on the effect of density prediction of visual images for the same scenario as Figure 8, a feed-forward back-propagation network is considered for the modelling of the people's density monitoring. The selected neural network has one input in this case, one hidden layer of ten neurons with *Tansig* functions and an output neurons with *Purelin* function. The selected learning rate of 0.05 and 5 samples of different densities are used. A total of 100 epochs are implemented with a training performance goal of 0.01. Matlab software and neural networks toolbox has been implemented for this analysis. The first curve in Figure 9 presents

the density of people as estimated by the experienced person for the 2880 images. The second curve presents the total number of pixels ( $N_\phi$ ). A total of 5 densities from graph (2), the number of pixels ( $N_\phi$ ), is used for the training of the neural network. The target has been selected to be the associated densities from the manually predicted densities, graph (2). Following the training, the complete data (2880 values) are fed to the neural network. The results are shown in the third graph of Figure 9. The error between the manually classified data and the neural networks is shown in the fourth graph of Figure 9. The overall accuracy of the system is found to be 97.3%. The only sources of error were the estimation of high density or extremely high density. Part of this could be a direct result from the fact that images are taken during night time where higher light intensity is needed. To solve this problem, infrared technology will be investigated.

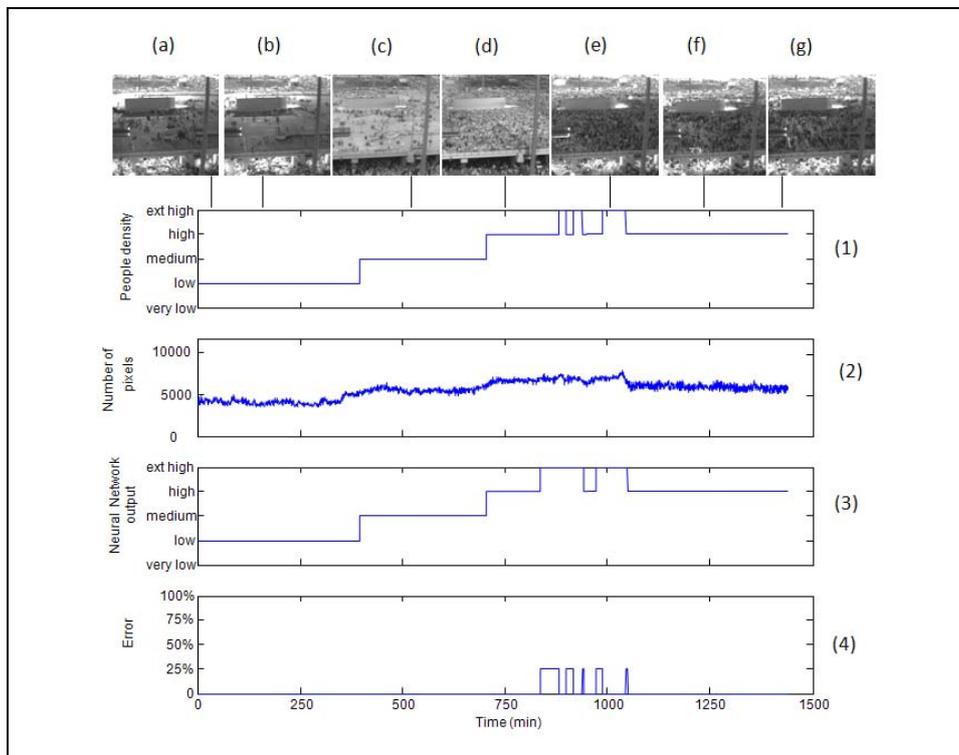


Figure 9: Visual neural network simulation result and error.

## 8. Infrared analysis

In order to enhance the system accuracy and reliability, particularly during night time, an infrared system is introduced and investigated. Figure 10 presents an example of the captured data for a full day. In Figure 10, the first graph presents the density as manually estimated by an expert. The second graph presents the average infrared data of the infrared image that is captured for a full day every 30 second period. The third graph presents the number of pixels above a specific threshold which represents human temperature. The average value in this experiment is 30°C, however this value is dependent on the distance, ambient temperature and the geometry of the view. In this case, the number of pixels is counted for an extrapolated infrared images between 16x16 pixels to 128x128 pixels. The

fourth graph presents the intensity of light measured by the light sensor. The values are normalised to the same level as the number of pixels.

In order to compensate for background heat, the average heat for a complete empty day before the Jamarat activity was captured, and the average infrared radiation (heat) for the 16x16 pixels infrared is calculated by equation (5) below:

$$\text{Average heat} = \frac{\sum \text{Pixels}}{256} \tag{5}$$

The average infrared radiation calculated by finding the average infrared for the foreground image and the background image at the same time of the day and then subtract the two average heat.

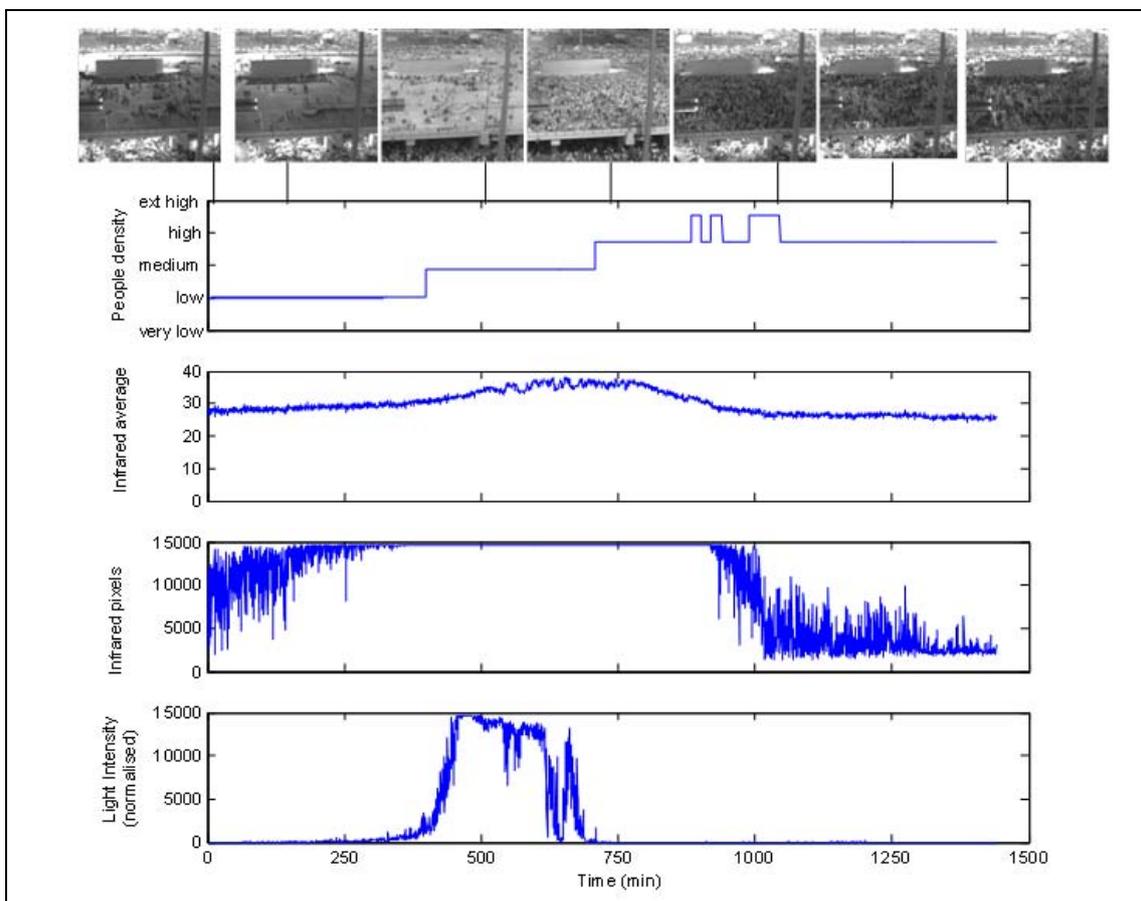


Figure 10: Infrared data after compensate the light intensity.

Notice that at time of about 10 minutes, the density of people on site is very low before sunrise and this is indicated by low level of infrared and light intensity. At time of 250 minutes, the sun starts to rise, notice that the light sensor is detecting an increase in the light values. This is also reflected in gradual increase in the infrared radiation. Notice the large peak in the average infrared radiation. The infrared graphs indicate the maximum densities. However, it is evident that the crowd in the afternoon has a medium density.

From Figure 10, it can be observed that the number of infrared pixels has significant noise and distortion. This is caused by a combination of the change in the background temperature and the intensity of light. In the following sections the compensation for light intensity and background temperature is discussed. Significant part of the noise is related to the infrared radiation of the empty background. Therefore, in order to enhance the accuracy and eliminate the noise, a localised pixels selected in the image of the background in order to compare it with foreground localised pixels as shown in the Figure 11.

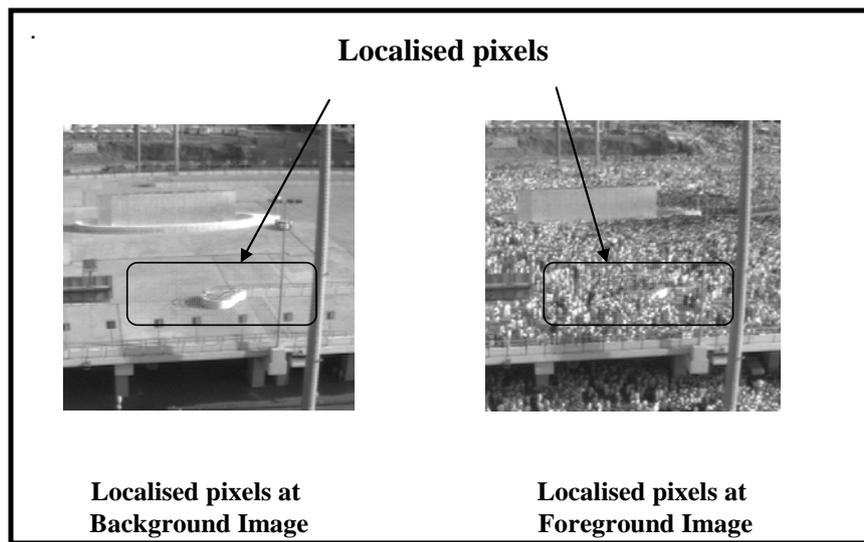


Figure 11: Localised pixels for the background and foreground images.

Figure 12 presents the results obtained using localised area of interest that represents people for the same scenario as Figure 10. The first graph of Figure 12 shows the expected density and the second graph represents the infrared average for the localised pixels. The third graph represents the infrared average for the background and the fourth curve represents the sensor fusion model output for the localised pixels. The fourth graph represents the difference between the background of the empty location and the location during the normal use. The difference is used as an indication of density estimation and has been quantised using equation (4) where Density in this case is the difference between the average infrared radiation. The system accuracy is 53% as shown in the fifth graph. The classification error was about 25% in most cases (i.e. one category) but it has also reached 50% (two categories of error).

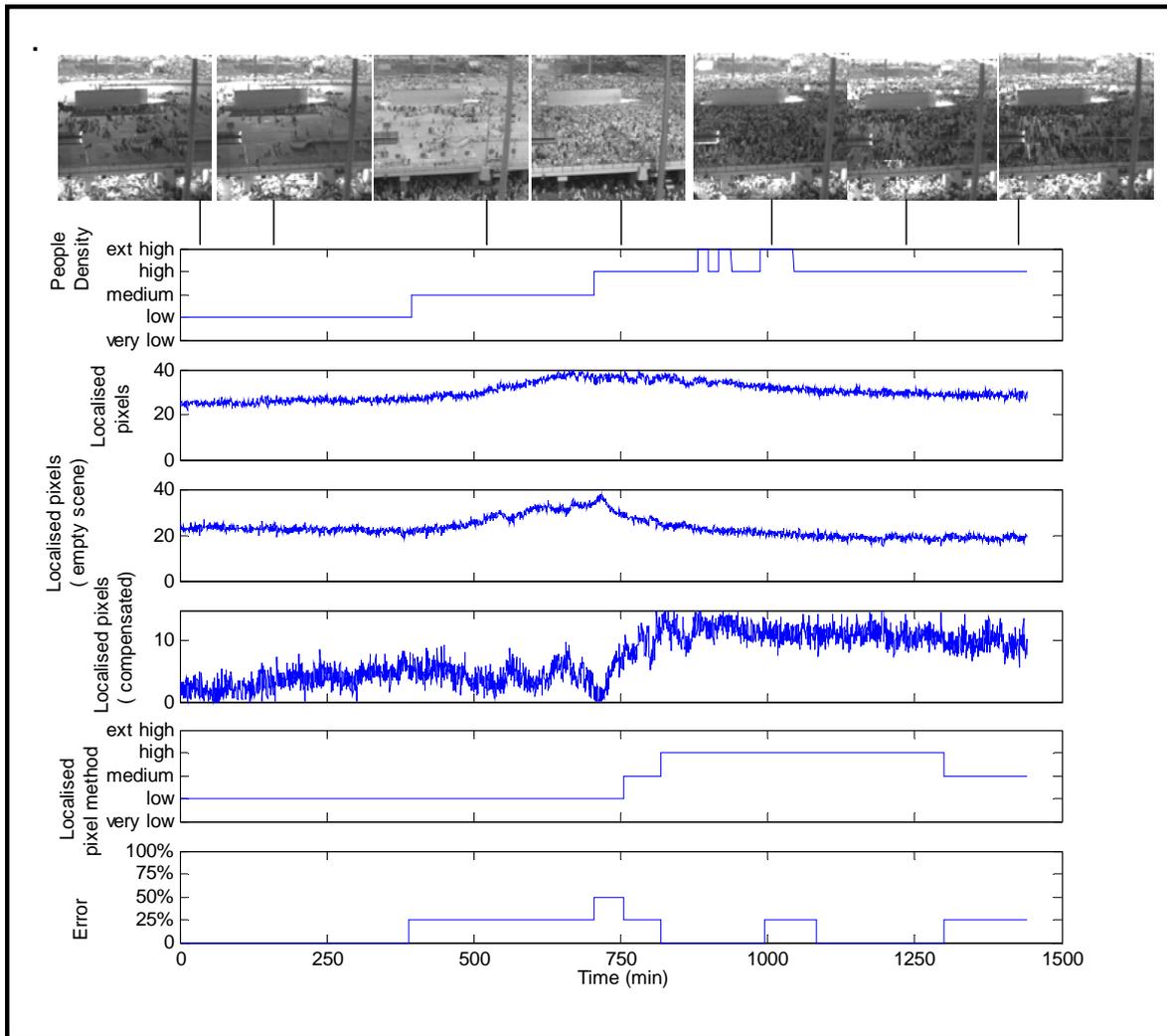


Figure 12: Localised infrared pixels after compensation.

## 9. Infrared neural network results

To investigate the use of artificial intelligence, a feed-forward back-propagation network is considered for the modelling of the people's density monitoring using infrared data. Similar to the visual analysis, the selected neural network has one input, one hidden layer of ten neurons with *Tansig* functions and an output neurons with *Purelin* function. The selected learning rate of 0.05 and 5 samples of different densities are used. Matlab software and neural networks toolbox has been implemented for this analysis. Five samples only with densities at different five levels are used as targets based on the first graph of Figure 12. Ten associated samples from the fourth graph of Figure 12 are used for training. Following the training, the complete 2880 samples were used for testing.

Figure 13 presents the results acquired from the neural network. The results are plotted based on the simulation of 2880 samples. It is evident that the maximum error in the classification is within one category (25%). However, it is evident that the system failed to recognise the extremely high density and in some cases low and medium densities.

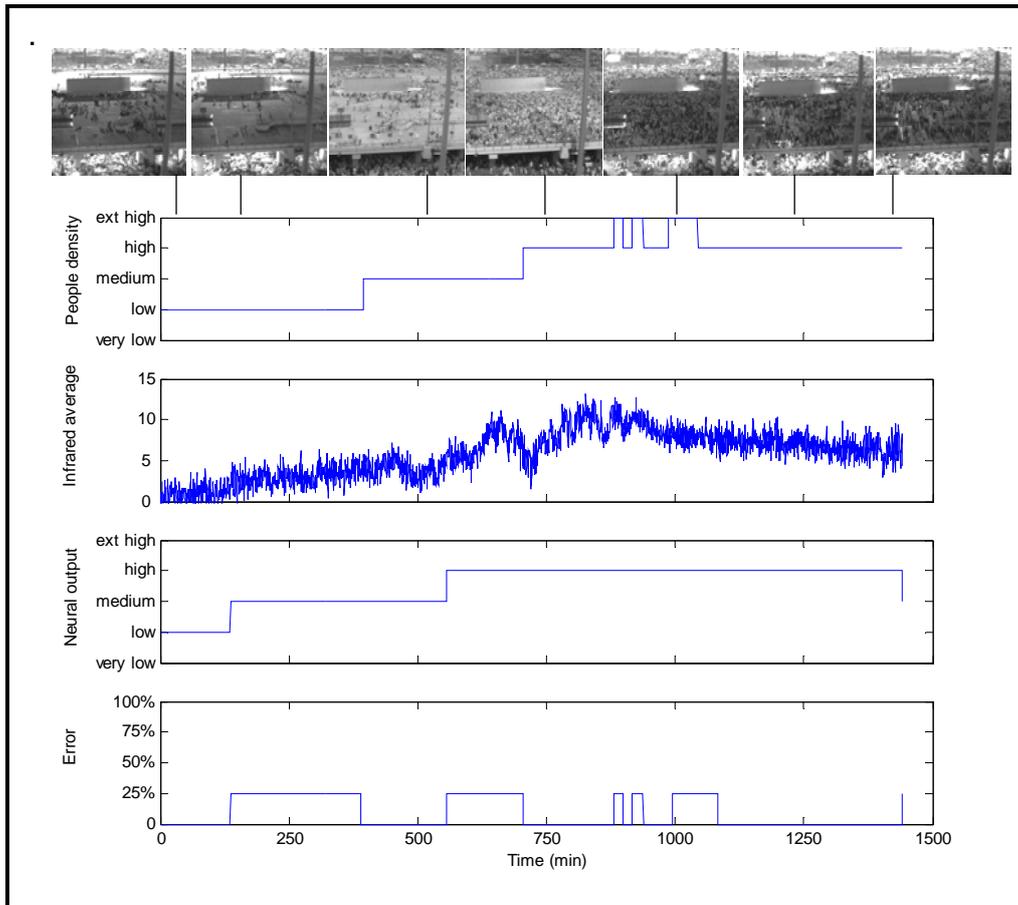


Figure 13: Infrared neural network simulation results and error

### 10. Sensor Fusion of visual and infrared systems

In order to optimise the overall accuracy of visual and infrared systems, the sensor fusion is investigated by combining both data of Figures 9 and 13. A feed-forward back-propagation network is considered with two inputs, one hidden layer of ten neurons with *Tansig* functions and an output neurons with *Purelin* function. The selected learning rate of 0.05 and 10 samples of different densities are used. Similar to the previous section, five samples only with densities at different five levels are used as targets based on the first graph of Figure 14. Ten associated samples are used for training. The two inputs are the visual number of pixels and the average infrared radiation as presented in Figure 14, the second and third graph respectively. Following the training, the complete 2880 samples were used for testing. The fourth graph in Figure 14 represents the results obtained using the combined neural network of visual and infrared systems; as shown the maximum percentage error has been reduced to zero, even for scenarios containing very low or extremely high density of people. A comparison between the five discussed systems is shown in Figure 15. This result proves that the sensor fusion can significantly enhance the density estimation of the crowd which will help in the design of safety systems for crowd monitoring and management.

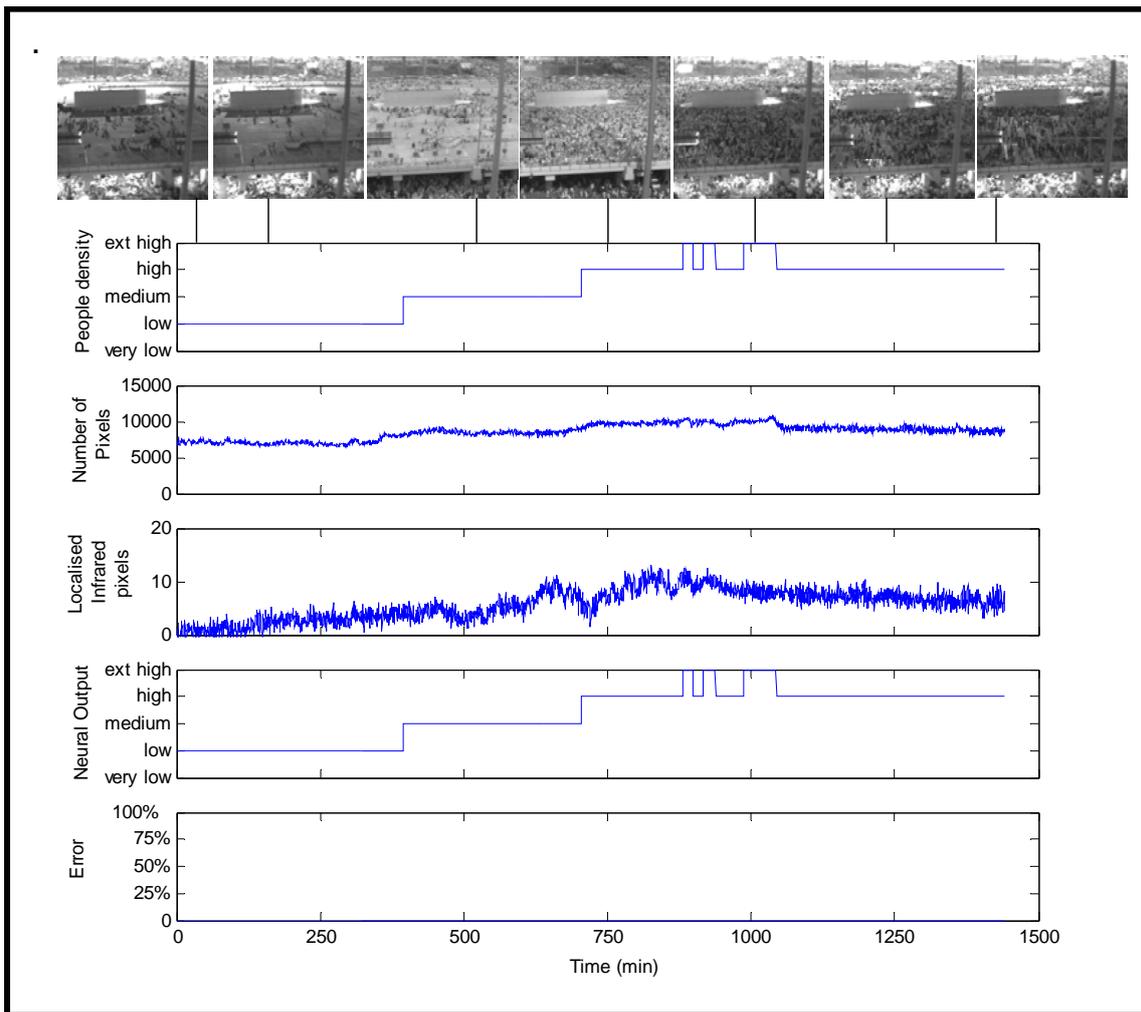


Figure 14: Combined neural network of visual and infrared systems.

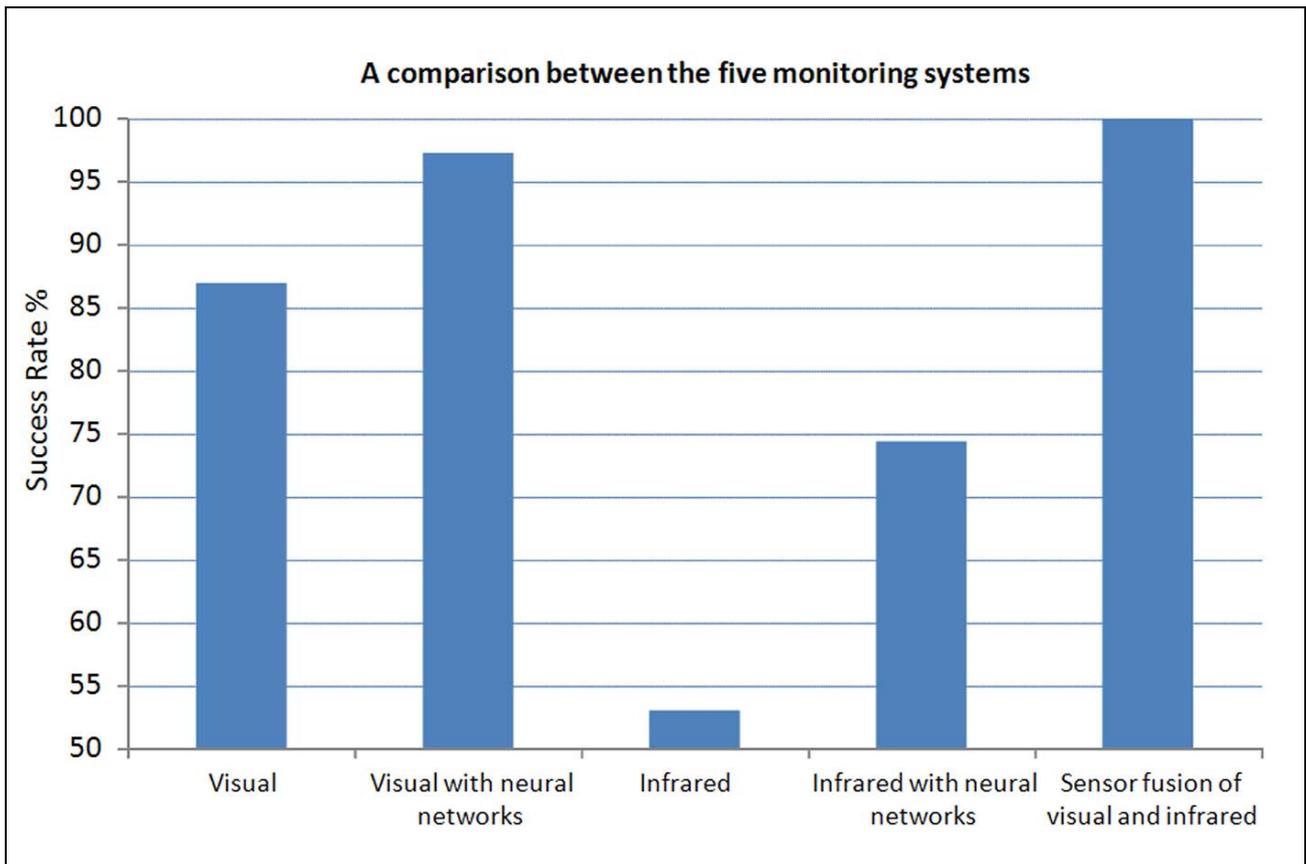


Figure 15: A Comparison between the five monitoring systems for density estimation.

## 11. Conclusion

This paper has investigated crowd density mapping using a novel sensory fusion model of infrared and visual data combined with artificial intelligence system to estimate of crowd's density for Hajj activities. The old Jamarat Bridge has been used to monitor different levels of density. The novel application of the sensor fusion model has been found to be successful in detecting people's density in a large open public place. The sensor fusion model has been shown to give significant improvements in the accuracy of density estimation in comparison with the separate applications of visual or infrared systems. This result proves that the sensor fusion can significantly enhance the density estimation of the crowd which will help in the design of safety systems for crowd monitoring and management.

The sensory fusion model is associated with additional sensory and computer cost. However, with the continuous reduction in cost of sensor and computer technology in recent years, this is expected to make the sensory fusion systems more common particularly in intelligent buildings application. The uniqueness of the sensing fusion model is its expected reliability to function in dark or light conditions at different levels including smoke interference, which makes it useful in emergency situations where power loss or thick smoke can affect visibility.

## 12. References

AlGadhi, S.A., Still, K., 2003. Jamarat Bridge: Mathematical Models, Computer Simulation and Hajjis Safety analysis, Ministry of Public Works and Housing, Saudi Arabia

Al-Habaibeh, A., Cai,R., Jackson,M.R., and Parkin, R.M., 2004, Proc. of 7th Int. Conference on Monitoring And Automatic Supervision In Manufacturing, Zakopane, Poland, 19 - 21 August 2004, p 49 .

Al-Habaibeh, A., Parkin, RM., Othman, F. and Albar, O., 2003. Investigation of an automated system for estimating people's density using infra-red thermal imaging . In: International Conference on Mechatronics, ICOM 2003, Loughborough, UK, 18 June 2003 . Parkin, RM., Al-Habaibeh, A. and Jackson, MR., eds., Professional Engineering Publishing, , Bury St Edmunds. pp. 371-376 .

Al-Habaibeh, A., Yaseen, S., parkin, RM., Othman, F. and Albar, O., 2010. The design of a novel sensor fusion model for monitoring people's density in public places using infrared thermography . Key Engineering Materials. vol 419-420, pp. 377-380.

Amin, I., Taylor, A., Junejo, F., Al-habaibeh, A. and Parkin, RM., 2006. Automated people counting by using low-resolution infrared and visual cameras . Measurement. vol 41 (6) , pp. 589-599.

AlQuraishi,A.A., Shokir,E.M.,2011, Artificial neural networks modeling for hydrocarbon gas viscosity and density estimation. Journal of King Saud University, Engineering Sciences (2011).

BBC,1989, [http://news.bbc.co.uk/onthisday/hi/dates/stories/april/15/newsid\\_2491000/2491195.stm](http://news.bbc.co.uk/onthisday/hi/dates/stories/april/15/newsid_2491000/2491195.stm).

BBC, 2002 , [http://news.bbc.co.uk/1/hi/world/middle\\_east/1830877.stm](http://news.bbc.co.uk/1/hi/world/middle_east/1830877.stm)

BBC, 2004,[http://news.bbc.co.uk/2/hi/middle\\_east/3448779.stm](http://news.bbc.co.uk/2/hi/middle_east/3448779.stm).

BBC, 2010,<http://www.bbc.co.uk/news/world-asia-pacific-11814894>.

BBC, 2011, <http://www.bbc.co.uk/news/world-asia-15633204>.

Celik, H.; Hanjalic, A.; Hendriks, E.A.,2006, Towards a Robust Solution to People Counting, , 2006 IEEE International Conference on Image Processing.

Challa, S.; Aboura, K.; Ravikanth, K.; Deshpande, S., 2007, Estimating the Number of People in Buildings Using Visual Information, Page(s):124 – 129.

Challa, S.; Aboura, K.; Ravikanth, K.; Deshpande, S.,2007, Estimating the Number of People in Buildings Using Visual Information, Information, Decision and Control, 2007. IDC '07,12-14 Feb. 2007 Page(s):124 – 129

Cho, S.Y., Chow, T.W.S., 2004. A Fast Neural Learning Vision System for Crowd Estimation at Underground Stations Platform Neural Processing Letters, vol. 10 (2), pp. 111–120.

Chow, T. W. S. , Cho, S.Y. ,2002, Industrial neural vision system for underground railway station platform surveillance. Advanced Engineering Informatics, 2002: 73~83.

Chow, T.W.S., Cho, S.Y., 2002. Industrial Neural Vision System For Underground Railway Station Platform Surveillance Advanced Engineering Informatics, vol. 16 (1), January 2002, pp. 73–83.

Davies, A. C. , Yin, J. H. , and Velastin, S. A.,1995, Crowd monitoring using image processing, IEE Electronic and Communications Engineering Journal, Vol. 7, No. 1 (Feb), pp.37-47, 1995.

Espina,M.V., Velastin,S.A. , 2005, Intelligent distributed surveillance systems: A Review, 'IEE Proceedings - Vision, Image and Signal Processing', 152(2) IEE, April 2005, pp. 192-204.

Ghosh,J., and Chakravarthy,V.S.,1992, Rapid Kernel Classifier: A link between the Self-Organizing Feature Map and the Radial Basis Function network. In Proc. of ANNIE 92, pp 15-20, Nov 1992.

Hashimoto,K., Kawaguchi, C. , Matsueda, S., Morinaka, K., Yoshiike, N. ,1998, People counting system using multisensing application, Sensors and Actuators , A 66,( 1998) 50-55 .

Hussain N, Yatim HSM, Hussain NL, Yan JLS, Haron F., 2011, A pixel-based crowd density estimation system for Masjid al-Haram, Safety Science 49 (2011) 824–833.

Kettnaker, V.; Zabih, R, 1999, Counting people from multiple cameras, Multimedia Computing and Systems. IEEE International Conference on Volume 2, 7-11 June 1999 Page(s):267 - 271 vol.2.

Kilambi, P., Ribnick,E., Joshi, A. J. , Masoud,O., and Papanikolopoulos,N.,2007 , Estimating pedestrian counts in groups Computer Vision and Image Understanding, In Press, Corrected Proof, Available online 12 March 2007.

Koshak, N.A., Fouda, A., 2008. Analyzing pedestrian movement in Mataf using GPS and GIS to support space redesign. In: Proceedings of the Ninth International Conference on Design and Decision Support Systems in Architecture and Urban Planning, Netherlands.

Lee, R. S.C., Hughes, R. L. ,2006, Accident Analysis & Prevention, Volume 38, Issue 4, July 2006, pp. 712-722.

Ministry of Hajj, 2011. <http://www.hajjinformation.com/main/1.htm>.

Masoud, O. , Kilambi, P. , Ribnick, E. , Joshi, A.J. , Papanikolopoulos, N.P.,2008, Estimating pedestrian counts in groups, Computer Vision and Image Understanding, 110, 2008.

Meireles, M.R.G. Almeida, P.E.M. Simoes, M.G,2003, A comprehensive review for industrial applicability of artificial neural networks ,Industrial Electronics, IEEE Transactions on Colorado Sch. of Mines, june 2003, Vol 3, page : 585 - 601 Golden, CO, USA

ROQUEIRO,D., and PETRUSHIN,V.A.,2006, Counting people using video cameras, The International Journal of Parallel, Emergent and Distributed Systems, Vol. 00, No. 0, 2006, 1–17.

Roqueiro,D., Valery A.P.,2007, Counting People using Video Cameras, International Journal of Parallel, Emergent and Distributed Systems Volume 22 Issue 3, May 2007.

Sacchi,C., Gera,G., Marcenaro,L., Regazzoni, C.S. ,2001, Advanced image-processing tools for counting people in tourist site-monitoring applications, Signal Processing 81 (2001) 1017-1040.

Septian, H.; Ji Tao; Yap-Peng Tan, 2006, People Counting by Video Segmentation and Tracking Control, Automation, Robotics and Vision, ICARCV, 9th International Conference on 5-8 Dec. 2006 Page(s):1 – 4.

Shendarkar A., Vasudevan, K., Lee,S. , Son, Y.J , 2008, Simulation Modelling Practice and Theory, Volume 16, Issue 9, October 2008, pp. 1415-1429.

Siddiqui, A.A. ; Gwynne, S.M.V.; 2012, Employing pedestrian observations in engineering analysis, Safety Science, Volume 50, Issue 3, March 2012, Pages 478-493.

Sidla, O.; Lypetsky, Y.; Brandle, N.; Seer, S,2006, Pedestrian Detection and Tracking for Counting Applications in Crowded Situations, Video and Signal Based Surveillance. AVSS, IEEE International Conference on Nov. 2006 Page(s):70 – 70.

Terada, K., Fujikawa,M. and Yamaguchi J.,1996, An Automatic Method of Counting the Passing People through the Gate by Using Two Space-time Images, MVA '96 IAPR Workshop on Machine Vision Applications. November. 12-14, 1996. Tokyo, Japan.

Velastin, S.A. , Boghossian, B. A., Vicencio-Silva, M. A.,2006, A motion-based image processing system for detecting potentially dangerous situations in underground railway stations, Transportation Research Part C 14 (2006) 96–113.

Yang, D. , Gonzalez-Banos,H., and Guibas,L., 2003, Counting People in Crowds with a Real-Time Network of Image Sensors, Int. Conference on Computer Vision (ICCV), pp. 122-129, 2003.

Yoshinaga, S., Shimada,A.,2010 , Real-time people counting using blob descriptor, Procedia Social and Behavioral Sciences 2 (2010) 143–152.

Zainuddin, Z., Thinakaran, K., Abu-Sulyman, I.M., 2009. Simulating the Circumambulation of the Ka'aba using SimWalk. European Journal of Scientific Research 38 (3), 454–464.