

Real-time Crowd Monitoring using Infrared Thermal Video Sequences

Adnan Ghazi Abuarafah¹; Mohamed Osama Khozium² and Essam AbdRabou²

¹Vice Dean for Academic Affairs, Faculty of Computer of and IS Makkah, Umm Al- Qura University. Makkah, SA

²Center of Research Excellence in Hajj and Omrah (HajjCore) Makkah, Umm Al- Qura University, Makkah, SA
agabuarafah@uqu.edu.sa, osama@khozium.com

Abstract:Monitoring people in a crowded environment is a critical task in civilian surveillance. Most vision-based counting techniques depend on detecting individuals in order to count their number. Counting becomes inefficient when it is required in real-time and when the crowd is dense. This paper proposes a novel technique for monitoring and estimating the density of crowd in real-time using infrared thermal video sequences. The research targets monitoring the crowd in Muslims' pilgrimage event (Hajj) while almost 3.0 million Muslims gather in Makkah to perform Hajj. During different Hajj phases the movement of the gathered Muslims is required at the same time from a place to another. Thus monitoring their crowd in real-time is crucial in order to take immediate decisions to prevent crowd disasters. A state of the art thermal camera has been acquired for the surveillance process. In addition, special software modules have been developed to analyze the captured thermographic video sequences in real-time. The results show high accuracy of the estimation of the crowd density in real-time.

[Adnan Ghazi Abuarafah; Mohamed Osama Khozium and Essam AbdRabou **Real-time Crowd Monitoring using Infrared Thermal Video Sequences**] Journal of American Science 2012; 8(3): 133-140]. (ISSN: 1545-1003).
<http://www.americanscience.org>.

Keywords: Hajj, Islamic informatics, crowd density estimation, crowd monitoring, computer image understanding, Thermography, computer vision.

1. Introduction

Around the 9th day of Hajj month, more than 3.0 million Muslims gather in Makkah to perform Hajj. This number will continue to increase in the next few years to reach almost 3.75 million Muslims by the year 2019[1]. Moving this giant number of people with uncontrolled manner resulted in many accidents in the past twenty years [2]. For example, in 2004, 251 died in a stampede at Mena Valley in Makkah[2]. Avoiding critical crowd densities and triggering rapid group movement in a relatively short period of time have become critical tasks[3] required to avoid future accidents and keeping the sanctity of emotions at its best. So providing real-time information about the density and behavior of the crowd in a certain place or route is the aim of this research.

This paper proposes a computer vision methodology for estimating of crowd density. Crowd density is defined as the number of people in the audience relative to the facility's capacity[3]. The target is to integrate the proposed technique with an intelligent decision support system for crowd management during Hajj rituals. The proposed technique uses a far infrared camera (sometimes also called forward looking infrared or FLIR camera) for monitoring and estimating the density of crowd in real-time. Although far infrared cameras were originally developed for military use, decreasing their prices make them now available for civilian use. There are some applications of it in industry especially in quality control. However, its technology is not yet widely

available, and many researchers lack the experience in acquiring far infrared (thermographic) images[4].

The proposed technique analyzes video sequences captured by a FLIR camera and calculates the occupied area of land in the captured scenes. Then, the crowd density of that area can be calculated. The analyzing algorithm works in real-time by continuously analyzing the camera output. A graph can be plotted to show the crowd density of a certain area against time. From this graph, the behavior of the crowd whether it is accelerating or decelerating can be determined. The results of the proposed technique are validated in two experiments which show high accuracy of providing crowd density and behavior estimation in real-time.

This paper is arranged in six sections. Section one is this introduction, section two discusses infrared Thermography from a theoretical point of view for Thermography. Section 3 lists some related work. Section four describes the methodology. Section five discusses the experimental work. Finally, section 6 concludes the work and discusses the future of the research.

Infrared Thermography

Thermographic images represent the electromagnetic radiation of an object in the far infrared range, which is 6 – 15 μ m. The principle of Thermography is based on the physical phenomenon that any body of a temperature above absolute zero (-273.15 °C) emits electromagnetic radiation. The electromagnetic spectrum is the range frequencies of electromagnetic radiation and is divided, lowest

wavelength to highest wavelength, in Gamma rays, X-rays, Ultraviolet, Visible, Infrared, Microwaves and Radio waves. All of these radiations emit energy that is described by wavelength according to the following equation:

$$E = h \times \frac{C}{\lambda}$$

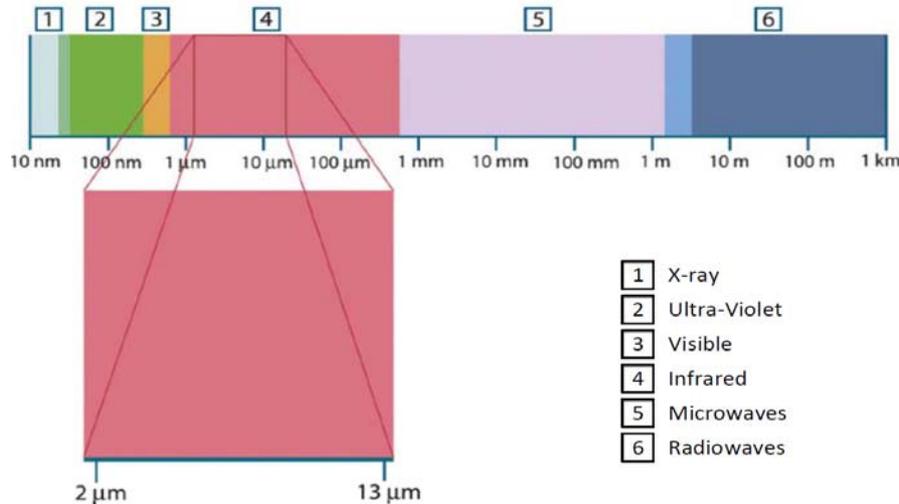


Fig.1: The Electromagnetic Spectrum[5]

There is clear correlation between the surface of a body and the intensity and spectral composition of its emitted radiation[6]. At the room temperature (300 K), objects emit the wavelengths with the peak at 9.7 μm, which is in the middle of far infrared range so they appear to shine without any external source of light[4]. The thermographic image depicts emitted energy from an object, and the relation between the temperature of an object and the emitted energy is given by the Stefan-Boltzmann formula:

$$W = \epsilon \sigma T^4 (\text{Watt}/m^2)$$

Where T represents temperature, W represents energy, σ is the Stefan-Boltzmann constant, and ϵ is the object's spectral emissivity.

Hence, thermography can be simply interpreted as a measurement technique, which, in most cases, is able to quantitatively measure surface temperatures of objects with different pixel intensities representing different temperatures. Thermography makes it possible to see one's environment with or without visible illumination. The amount of radiation emitted by an object increases with temperature; therefore, thermography allows one to see variations in temperature. When viewed through a thermal imaging camera, warm objects stand out well against cooler backgrounds; humans and other warm-blooded animals become easily visible against the environment, day or night.

As a result, thermography is particularly useful to users of surveillance cameras. Surveillance cameras are

available with both 320×240 or 640×480 pixels and a variety of lenses and fields of view[6]. Fig.2 shows an example of thermal image of some people moving in Haram area.



Fig.2: A Thermal Image of some People moving in Haram Area

Related Work

Certain objects in dynamic scenes have been making an active research topic in computer vision in the last few years. The target is not only to replace the ineffective passive video surveillance but also to develop intelligent visual surveillance in order to achieve the full automation for the entire surveillance task[7]. Four basic visualization resources have been used by researchers; they are visual images, video sequences, satellite images, and thermography.

The use of still images has been extended to analyzing the work from multiple cameras using information fusion. A tool for estimating behaviors of a crowd derived from distributed and heterogeneous sensors has been presented By Andersson, and Rydell[8]. The tool doesn't need to identify specific persons or decide their exact positions in the scene but it aims to become aware of that something abnormal has occurred. The concept is used for automatically alerting operators when abnormal behaviors occur, or are about to occur. Authors Yang *et al.*[9]estimated the number of people directly from groups of image sensors. For each sensor, foreground objects are segmented from the background, and the resulting silhouettes are combined over the sensor network. A geometric algorithm is then introduced to limit the number and possible locations of people using silhouettes extracted by each sensor. Mahalingam *et al.*,[10]have presented a simplified method for tracking people in crowded scene from a video sequence. The method is based on computing the Minimum Mean Square Error (MMSE) between frames of the video sequence to identify people in subsequent frames. They have succeeded to handle small occlusions, varied lighting conditions and camera motions. A system for crowd detection from a moving platform has been presented By Piniet *et al.*[11]. The system uses slices in the spatiotemporal domain to detect inward motion as well as intersections between multiple moving objects. The system calculates probability distribution functions for left and right inward motion and uses these probability distribution functions to deduce a decision about inward motion or crowd detection. The system has succeeded to automatically detect scenes that contain crowd consisting of multiple Attempts to detect, recognize and track pedestrians moving in opposite directions, even at a large distance

Sirmacek and Reinartz[12]have introduced a novel approach to detect crowded areas automatically from very high resolution satellite images. Although resolutions of those images are not enough to see each person with sharp details, they can still notice a change

of color components in the place where a person exists. Therefore, they developed an algorithm which is based on local feature extraction from input images. They have tested our algorithm on panchromatic Worldview-2 satellite image dataset, and also compared with an algorithm result obtained from an airborne image of the same test area. The presented results indicate possible usage of the algorithm in real-life events.

Thermographic technology is not yet widely available, and researchers still don't have enough experience in acquiring far infrared images[4].So, there is no many available literature that uses real-time thermography in crowd monitoring. Al-Habaibe *et al.*[13] have described a novel application of low-cost infrared system for estimating people's density using infrared thermography. They have conducted their experiments inside Madina Mosque in Saudi Arabia. They designed a fusion of three sensors low-cost infrared sensor, light intensity sensor and temperature sensor. By taking one shot every one minute they have been able to predict the people density by plotting the number of warm infrared pixels which are found to be very representative to the density of people in the mosque. Because of the air condition inside the mosque, the temperature sensor has no use in their experiment which leaves a question about its role in the fusion system. In addition, the change in light intensity has caused significant source of error in their experiments. The obtained results did not give an indication about the critical crowd density level; however they only showed that in different prayer times, people are increasing in the mosque.

Methodology

The MATLAB® package including toolboxes for the acquisition of video data and pictures is selected as the development environment. Several stages have been accomplished in order to reach the real-time analysis of thermal video sequences starting from single image and ending with real-time video sequence. Fig.3 illustrates a schematic diagram of the software module.

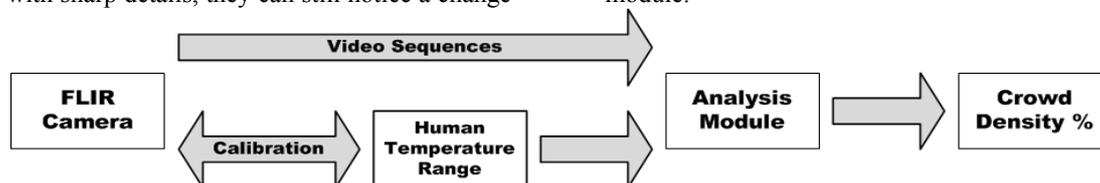


Fig.3: Analysis Module Schematic Diagram

The software module requires two different inputs. The first input is the video sequence captured by the FLIR camera whether it is in Audio Video Interleave (AVI)[14]or in MPEG-4 Part 14 or MP4[15]formats. The second input is the range of human temperature. To determine this range a calibration step is required.

Human temperature differs on FLIR cameras according to the time of capturing the image day or night, the ground temperature, and the crowd density. In addition, the human temperature is affected by the crowd density ratio at the target scene. High crowd density will raise the human temperature and vice versa. So, the calibration step is required. The timing

for the calibration step is chosen to be every 30 minutes. Several images are taken at the target scene and the required temperature range of humans can be calculated. Fig.4 shows some thermal images during the calibration step.

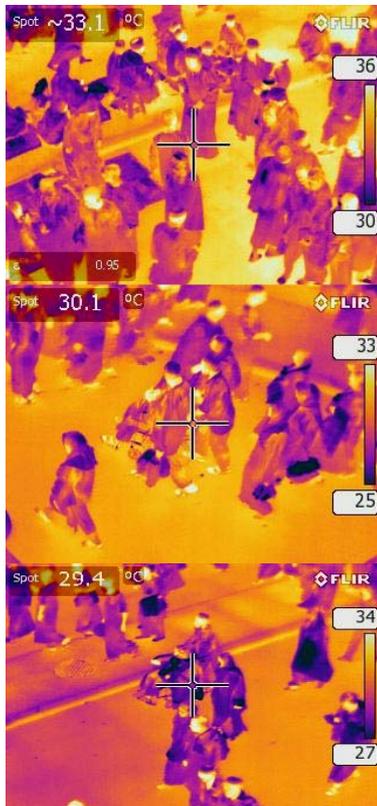


Fig.4: Calibration of the FLIR Camera

Frame Temperature Extraction:

At any frame (320 x 240 pixels) of the entered video sequence, the module can determine the minimum and the maximum temperature. Given a temperature range of humans, the module is able to determine the percentage of this temperature range. To do so, the module is performing the following steps:

Given

The minimum T_l and maximum T_h temperature difference in the working frame

$$\Delta T = T_h - T_l$$

The higher R_h and lower R_l temperature range required to calculate its occupancies

$$\Delta R = R_h - R_l$$

Converting from Color to Gray:

Convert the working frame into gray scale values by forming a weighted sum of the Red (R), Green (G), and Blue (B) components of each pixel as follows;

$$C_r \times R + C_g \times G + C_b \times B$$

Where $C_r = 0.2989$, $C_g = 0.587$ and $C_b = 0.114$

Locate the temperature ruler at the right most of the working frame which has a fixed position of (x = 307, y = 70) in pixels and a fixed size of (width = 8, height = 101) in pixels.

Calculate pixels ratio according to temperature ruler color ranges:

Since the height of the required range is a part of the ruler height according to the ratio between the specified range difference and the working frame temperature difference in the frame, calculate the ratio,

$$ratio = \frac{\Delta R}{\Delta T}$$

Calculate the specified range height by multiplying the calculated ratio by the height of the ruler H,

$$H_r = \frac{\Delta R}{\Delta T} \times H$$

Determine the y coordinate of the specified temperature range on the working frame by determining its start point,

$$S = \frac{(T_h - R_h)}{\Delta T} \times H$$

Crop the specified temperature range from the ruler starting from its y coordinate up to its specified height;

Start searching the working frame image by the cropped range by counting the number of pixels P which have color values equal to all the color values in the cropped part;

$$P = \sum_{k=m}^n X(k)$$

Where X(k) is the number of pixels for the color k in the frame, m & n are the minimum and maximum color values according to the required temperature range colors respectively.

Calculate the occupancy density percentage (D) by dividing P by the total number of pixels in the working frame which is equal to the multiplication of 320x240,

$$D = \frac{P}{FrameSize} \times 100$$

Where Frame Size=320 x 240

The calculated ratio is simply the crowd density if the provided temperature range is the approximate range of the human temperature. So the crowd density ratio is given by

$$Crowddenesityratio = OccupancyRatio$$

The module calculates the crowd density in real-time in pre-defined steps; every step includes a specific number of frames pre-defined in system configuration. The less number of frames in the step will result in increasing the accuracy of calculating the average crowd ratio. The module will show alert lamps which indicate the crowd density percentage in different colors as shown in Fig.5. Showing a specific color means that there is at least one frame within the

selected step that hits the crowd density range specified by that color. For example, showing an orange bulb means that within the specified step there is one scene (frame) that its calculated crowd density is between 75% and 85%.

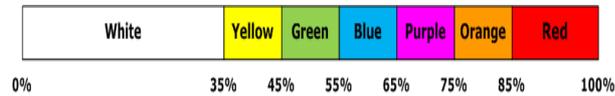


Fig.5: Color bulbs according to crowd density

In addition to the generated warnings, the module generates a plot of number of frames against the crowd density ratio as shown in Fig.6.

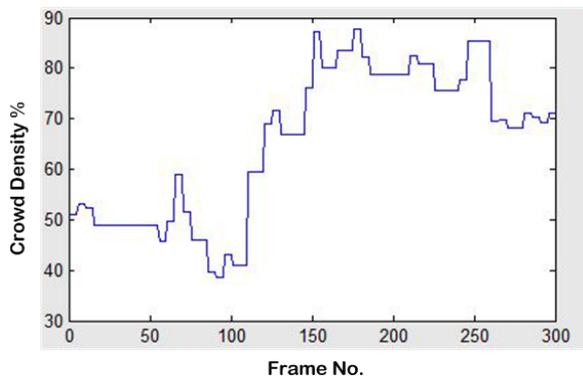


Fig.6: Final plot of 300 frames against crowd density percentage

The X-axis represents the frame number and Y-axis represents the percentage of the crowd density. This means in frame number 0 the area is 51% crowded. In frame number 150 the crowd density ratio is almost 89%. In the last frame no. 300 the crowd density ratio is 71%. So, the movement behavior can be inferred from the plot that between frame 0 and 150 there is accelerating crowd while between frame 150 and 300 there is decelerating crowd.

Experimental work

The major objective of controlling crowd during Hajj events is to let people move freely, avoiding over crowd and stampede. This can be achieved by proving the standard space required by humans to move. This raises a question about the maximum number of people to occupy a certain area. The universal space requirements for human beings in different situations have been discussed in [16]. As shown in sketch 16 in Fig.7, the maximum density is six persons per square meters. But these persons are not moving they are standing still in, for example, a cable railway.

To deduce the number of persons to occupy one square meter during walking in steps, sketches (6) and (13) in Fig.7 are used. Adding 10% to the dimension shown in sketch (6) for the moving activity will bring the dimension to 1.87m so combining the two sketches as shown in Fig.8 will bring a total area of 1.87m x (3 x 0.75m) which equals to 4.2m². This holds nine moving humans; thus for comfortably moving humans, it is required to have 2.14 persons in one square meter. For normal spacing between humans, sketches (6) and (9) are used. As shown in Fig.8 a total area of 1.7m x 1.875m equals to almost 3.2m² occupied by 12 persons which gives 3.75 persons per square meter.

Since the target is controlling a moving event during Hajj, the focus will be providing enough space for moving humans to move in a normal way so we will use the number of two persons per square meter to represent 100% crowd density.

Most of the Hajj phases take place in the Mashaer, three open sites about 5 km south east of Makkah. These sites are Mina, Muzdalefa and Arafa. In the 9th day of the Hajj month all the gathered people are required to move together after the sunset from Arafa to Muzdalefa using several roads. This movement which is called "Nafra" is the target of this study.

A FLIR E60bx thermal camera is fixed over an elevation about 10m above one specific pedestrian road. The camera is connected to a computer that runs the software modules. Two experiments A and B have been conducted by capturing two videos for known areas with known number of people. Assuming that occupancy of two persons in square meter constitutes 100% crowd density, Fig.9 and Fig.10 show experiment A and experiment B respectively.

Both experiments show good results. Experiment A examines an area of a square with dimensions of 7m by 7m that gives an area of almost 50 square meters. There are 23 persons occupy the area. So, assuming that two persons in a square meter makes 100% crowd, the actual crowd density is 23%. The calculated crowd density is shown as 22% which gives an error percentage $\pm 1\%$. Experiment B examines an area of a square with dimensions of 9m by 9m that gives an area of almost 80 square meters. There are 78 persons occupy the area. So, the actual crowd density is 48.76% with the same assumption. The calculated crowd density is shown as 48.73% which gives an error percentage almost 0%. The experiments show acceptable results in the case of dense crowd areas. The result is sufficient enough to take movement decisions. As there is no available related work, the results cannot be validated against previously obtained results.

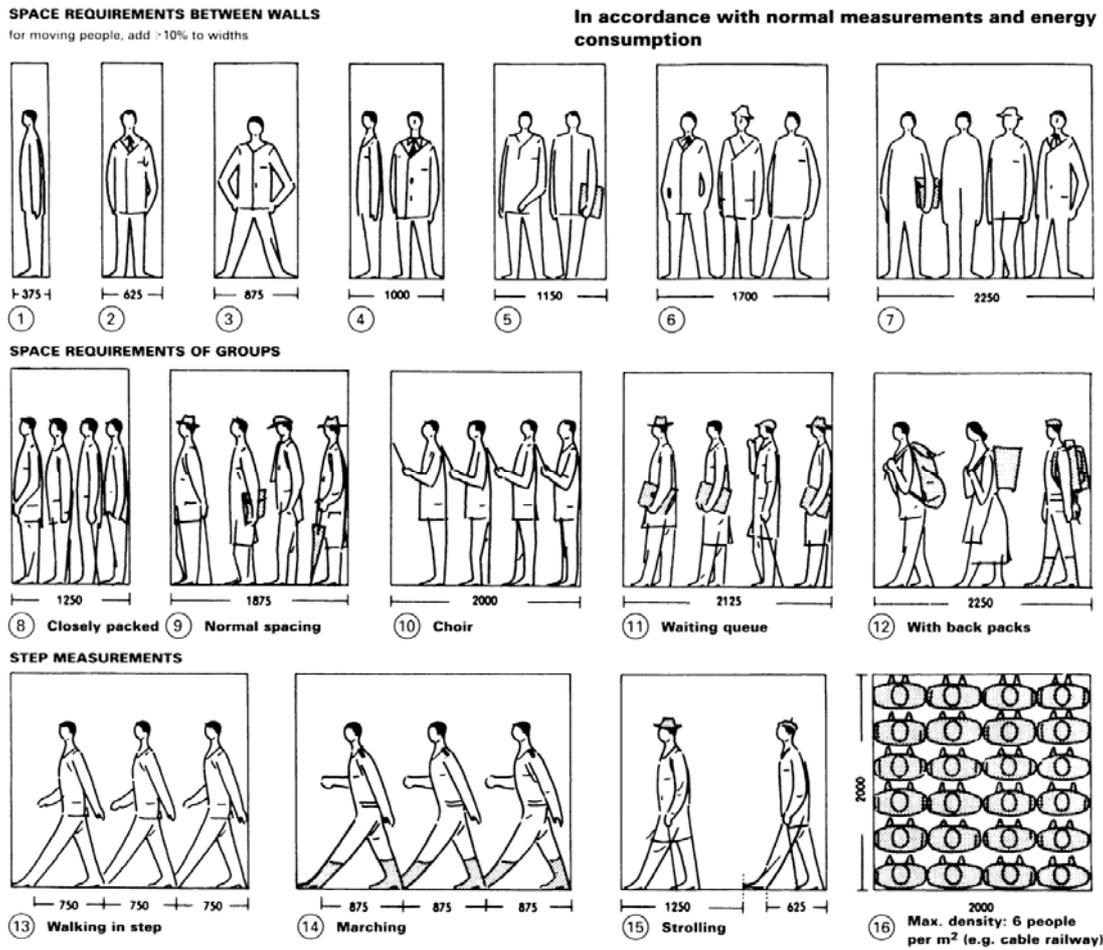


Fig.7: Man: Dimensions and Space Requirements [16]

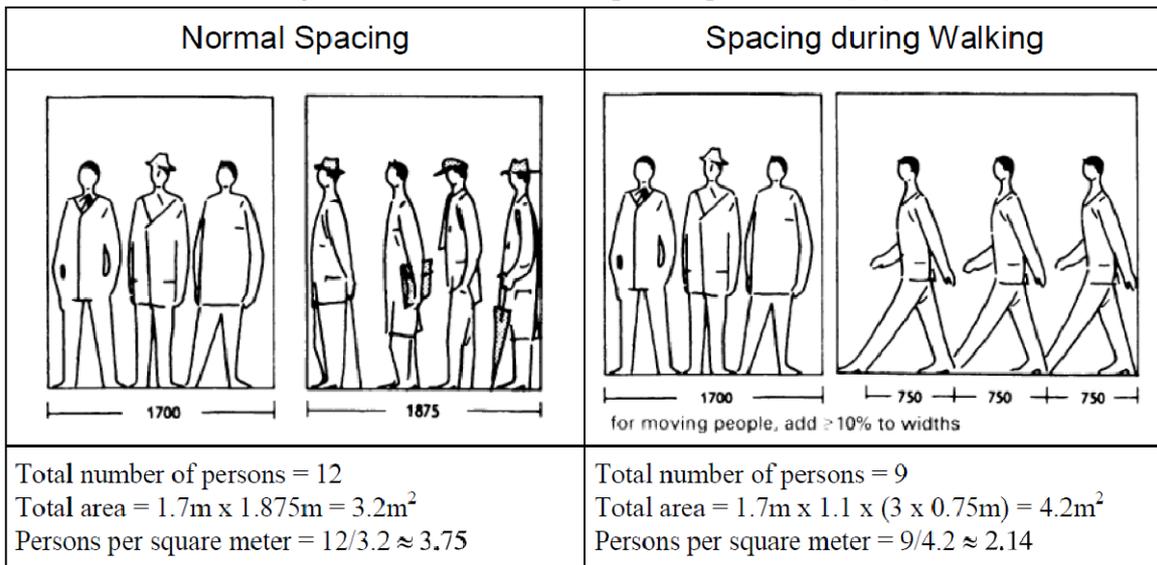


Fig.8: Space Required by Standing and Moving People

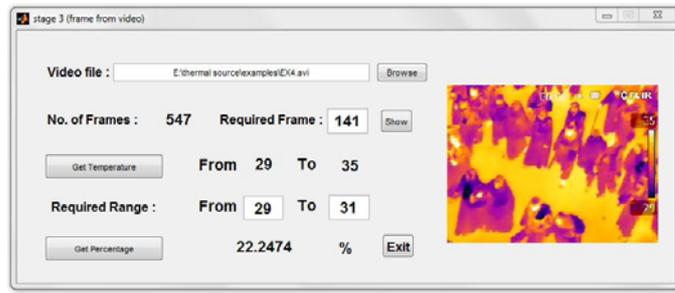


Fig.9: Experiment A

Working Frame	141
Approximate Area (m ²)	50
No. of persons for 100% crowd	100
Actual No. of Persons	23
Actual Crowd Density Ratio	23%
Min.- Max. Temp. Range °C	29 – 35
Specified Temp. Range°C (Human Temperature)	29 – 32
Calculated Occupancy Ratio of the Specified Range	22%
Error %	±1%

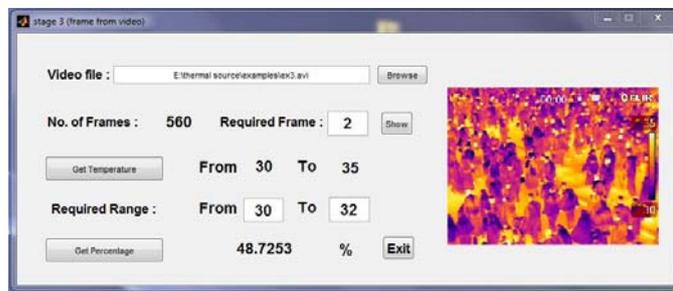


Fig.10: Experiment B

Working Frame	2
Approximate Area (m ²)	80
No. of persons for 100% crowd	160
Actual No. of Persons	78
Actual Crowd Density Ratio	48.76%
Min.- Max. Temp. Range °C	30 – 35
Specified Temp. Range°C (Human Temperature)	30 – 32
Calculated Occupancy Ratio of the Specified Range	48.73%
Error %	±0.03%

Conclusion and Future Work

In this paper a novel technique for crowd density and behavior estimation in real-time has been proposed. The proposed technique employs a far infrared FLIR camera. The technique used is able to process the thermal camera output such as video sequences in real-time and produce an estimation of the crowd density. By plotting a graph of the calculated density against frames which in fact represent time,

crowd movement behavior also can be determined whether accelerating or decelerating.

Several reasons make thermography is the most suitable technique for the Hajj event. First thermal imaging is non-contact, i.e. uses remote sensing so it keeps the user out of danger. Meanwhile it does not intrude upon or affect the target at all so it keeps people privacy intact. Also the produced images allow for excellent overview of the target without the need of intelligent recognition of faces or body parts.

For the future work the proposed technique needs to be integrated with an intelligent decision support system approach for crowd management during Hajj rituals in order to direct the crowd towards safe movement decisions.

Acknowledgement

This research has been supported by the center of research excellence in Hajj and Omrah (HajjCore), Umm Al-Qura University, Makkah, Saudi Arabia. Under project number p1122, entitled (Intelligent-based notification system for crowd management using video analytics approach).

Corresponding author

Mohamed Osama Khozium¹

Center of Research Excellence in Hajj and Omrah (HajjCore) Makkah, Um Al- Qura University, Makkah, SA

osama@khozium.com , agabuarafah@uqu.edu.sa

References

- [1] Ben-Mahmoud C., M. AbouChalbak, C. Plumb, and K. Moore(2010): Holy Cities: Saudi's Unique Real Estate Markets. In *on.point Jones Lang LaSalle IP, INC*. Retrieved from <http://www.joneslanglasalle-mena.com/MENA/EN-GB/Pages/Home.aspx>.
- [2] Still K.(2011): Crowd Safety in the Complex and Built Environment. *Crowd Disasters*.Retrieved Sep. 19, 2011, from <http://www.gkstill.com/CrowdDisasters.html>.
- [3] Fruin J. (1993): The causes and prevention of crowd disasters. *Engineering for crowd safety*. Elsevier, New York, , pp. 1-10.
- [4] Perš J., M. Kristan, M. Perše, and S. Kovačič, Observing Human Motion Using Far-Infrared (FLIR) Camera- Some Preliminary Studies. In *Proceedings of the Thirteenth International Electrotechnical and Computer Science Conference ERK 2004, Portorož, Slovenia*, 2004, pp. 187-190.
- [5] ThermaCam Reporter: *Users's Manual Version 7.0*. Flir Systems, 2004.
- [6] Vollmer M., and K.-P.Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications*, 1st Ed. Weinheim, Germany, 2010.
- [7] Hu W., T. Tan, L. Wang, and S. Maybank, A Survey on Visual Surveillance of Object Motion and Behaviors. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 34(3), 2004, pp. 334-352.
- [8] Andersson M., and J. Rydell, Estimation of crowd behavior using sensor networks and sensor fusion. *Fusion, 2009, FUSION*'09, 2009, pp. 396-403.
- [9] Yang D., H. González-Baños, and L. Guibas(2003): Counting people in crowds with a real-time network of simple image sensors. In *Proceedings of Ninth IEEE International Conference onComputer Vision*, , pp. 122–129.
- [10]Mahalingam G., C. Kambhamettu, and B. Aguirre(2009): Crowd Analysis from Video. *2009 NSF Engineering Research and Innovation Conference*. Hawaii,.
- [11]Pini R., M. Ofer, A. Shai, and S. Amnon, Crowd detection in video sequences.*IEEE Intelligent Vehicles Symposium*, 2004, pp. 66-71.
- [12]Sirmacek B., and P. Reinartz(2011): Automatic crowd analysis from very high resolution satellite images. In *Proceedings of Photogrammetry and Image Analysis Conference*, Munich, Germany,.
- [13]Al-Habaibeh A., S. Yaseen, R. Parkin, M. Otham, and O. Albar(2009): The Design of a Novel Sensor Fusion Model for Monitoring People's Density in Public Places Using Infrared Thermography. *Key Engineering Materials*, 419-420, , pp.: 377-380.
- [14]Microsoft Corporation (1998). *WAVE and AVI Codec Registries - RFC 2361*.IETF.Retrieved 2011-09-26.
- [15]International Organization for Standardization, ISO (2003).*MPEG-4 Part 14: MP4 file format; ISO/IEC 14496-14:2003*. Retrieved 2011-09-26.
- [16]Neufert E., P. Neufert, B. Baiche, and N. Walliman2002, *Architects' Data*. 3rd Ed., , Wiley-Blackwell.
- [17]Williams T., *An Introduction to Thermal Imaging, Thermal Imaging Cameras*. 1st Ed., 2009, CRC Press, FL.
- [18]Yin J., S. Velastin, and A. Davies(1995): Image Processing Techniques for Crowd Density Estimation Using a Reference Image. In Invited Session Papers from the Second Asian Conference on Computer Vision: *Recent Developments in Computer Vision (ACCV '95)*. Li, S., Z., Mital, D., P., Teoh, E., K., and Wang, H., (Eds.). Springer-Verlag, London, UK, , pp. 489-498.

Biography

Adnan G. Abuarafah completed his PhD from Illinois Institute of Technology, Chicago, Illinois, USA in 1995 in AI area; he is currently associate professor, Faculty of Computer Science and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia.

Mohamed O. Khozium completed his PhD from Cairo University (EGYPT) in 2005 in information system area; he is currently associate professor at the department of engineering, Faculty of Society, Umm Al-Qura University, Makkah, Saudi Arabia.

EssamAbdrabou received his PhD in computer science in 2011, Ain Shams University, Cairo, Egypt. He is currently a Research Associate in the Center of Research Excellence in Hajj and Omrah (HajjCore), Makkah,SaudiArabi.