

Crowd Monitoring System Using Unmanned Aerial Vehicle (UAV)

Ali Al-Sheary¹ and Ali Almagbile²

1. Department of Civil Engineering, Umm Al-Qura University, Makkah 21955, Saudi Arabia;

2. Department of Geography, Yarmouk University, Irbid 21163, Jordan

Abstract: Understanding and dealing with safety aspects of crowd dynamics in mass gatherings of people related to sports, religious and cultural activities is very important, specifically with respect to crowd risk analysis and crowd safety. Historical trends from the Kingdom of Saudi Arabia hosting millions of pilgrims each year during the Hajj and Omrah seasons suggest that stampedes in mass gatherings occur frequently and highlight the importance of studying and dealing with the crowd dynamics more scientifically. In this regard, efficient monitoring and other safe crowd management techniques have been used to minimize the risks associated with such mass gathering. An example of these techniques is real-time monitoring of crowd using a UAV (Unmanned Aerial Vehicle); this technique is becoming increasingly popular with the objective to save human lives, preserve environment, protect property, keep the peace, and uphold governmental authority. In this paper, a crowd monitoring system for pedestrians has been proposed and tested. The system has deployed crowd monitoring technique using real-time images taken by UAVs; the collected data was investigated, and crowd density was estimated using image segmentation procedures. A color-based segmentation method has been employed to detect, identify and map crowd density under different camera positions and orientations. Furthermore, the associated anomalies/outliers which may lead to non-classification of features have been eliminated using image enhancement tools. The paper presents a crowd monitoring system for pedestrians that can contribute to an area of research still in its infancy. The proposed system is a valuable tool in terms of facilitating timely decisions, based on highly accurate information. The results show that the used image segmentation technique has the capability of mapping the crowd density with an accuracy level up to 80%.

Key words: UAV, crowd monitoring, crowd density, geo-referencing, mapping.

1. Introduction

The research field of crowd dynamics, specifically the need for minimizing the risks associated with pedestrian crowd, has taken considerable attentions in the last few decades [1-5]. Thus, considerable effort has already been exerted in crowd monitoring and analysis in either large or small-scale areas [6-8].

The focus of previous research was to detect and track moving objects using images and videos generated by vision sensors (cameras) on board moving platforms like UAVs (Unmanned Aerial Vehicles). However, crowd management also requires

identifying the parameters for estimating safety risks such as the density of individuals per square meter [9], the direction of movement of groups of people [10] and their movement patterns [6]. To effectively respond to potential risk event situation, it is necessary to build a system that provides surveillance and crowd control. To achieve this, designing a crowd monitoring system to extract real-time information that is capable of communicating instructions to the crowd is crucial for safe crowd management process. Such a system would be a valuable tool for developing an efficient transport system in a sustainable smart city [11].

To build a crowd monitoring system, a platform which provides a comprehensive database of the area under investigation is required. One of the platforms that is capable of rapid collection of dynamic data is the UAV. The UAV flies autonomously over a defined

Corresponding author: Ali Al-Sheary, assistant professor, Ph.D.; research fields: unmanned aerial vehicle applications, GNSS real-time kinematic positioning, location-based applications, sustainability, and crowd management. E-mail: amshaery@uqu.edu.sa.

area and provides high-resolution and real-time images. The UAV needs to be equipped with different sensors, e.g., GPS (Global Positioning System), INS (Inertial Navigation System), and cameras (color or infrared) for generating pictures and detecting people in the desired location. In addition, computing and communication infrastructure is also required to transmit sensory data to the control station (first responder). The control station then processes the sensory data and builds information to alert the field staff (second responder) of any urgent situations. Ultimately, the second responder accesses the information and takes action within a short time [8].

Once the system has been designed, the UAV sensory data needs to be checked by the first responder before building the information and alerting the second responder. Otherwise, the information may develop inaccurately and improperly influence the required action. Therefore, quality control procedures are essential towards clarifying the data from the potential faults. An instance of sensor faults is anomalies/outliers in GPS data which influence the precision of positioning. In some cases, the GPS signals might be influenced by multipath error or blocked in urban environments. To overcome this problem, GPS can be integrated with other sensors such as INS so as to resolve issues relating to the absence of GPS signals. Other problems may relate to image processing, such as image mismatching, which would require further quality control procedures.

In some places, crowd management is necessary to minimize the risks associated with the mass gathering and movement of public to provide a safe and peaceful movement. Thus, an accurate decision must be made based on accurate information provided by the control station. In this study, we carry out an investigation of crowd monitoring with different camera positions and orientations using UAV systems. These include the altitude of UAV, the use of horizontal and vertical cameras, and the use of different camera angles. The purpose is to construct a system that monitors and

responds to risks associated with crowds gathered at sports, cultural, and religious events. Particularly for Saudi Arabia, which hosts millions of pilgrims each year during the Hajj season, an efficient crowd monitoring and management system is inevitably required. Also, an adapted system will be introduced to meet Hajj-specific challenges and its harsh environment. Once UAV images are received, they will be processed to estimate the crowd density. Ultimately, each image is georeferenced using ground control points and then mapped to estimate the crowd density levels. This paper presents crowd monitoring from a researcher's point of view followed by methodology of crowd density estimation.

2. Crowd Monitoring

Various investigations have been directed towards developing an approach that provides a comprehensive solution for crowd management and analysis. These studies have focused on three main principles of crowd analysis including estimation of the density of a group of people per square meter, finding the direction of motion of the crowds and geo-referencing the crowd images in the real world coordinate system. It can be noticed that most of the studies have been conducted in either computer vision or navigation disciplines. Thus, crowd analysis approaches have been based on images processing and positioning solutions. In this regard, several sensors such as camera (color, infrared), GPS and IMU (Inertial Management Unit) have usually been mounted on a moving vehicle to detect and analyze the density along with the motion of the crowd in a specific location.

In recent years, UAVs have been extensively used for crowd analysis due to its low cost and fast, autonomous and flexible flying ability. Thus, crowd motion have been tracked via UAVs sensors. Talukder et al. [12] used optical flow and dense stereo as a combination technique for estimating object background motion at each pixel. This, in turn, increased the probability of detection of small or low

texture objects where feature selection schemes might fail. While Rodriguez-Canosa et al. [7] proposed a real-time approach to detect and track moving objects from UAV. This approach was entirely dependent upon creating artificial optical flow by estimating the camera motion between two subsequent video frames and then comparing the artificial flow with the real optical flow directly obtained from the video feed.

The process of estimation of density and the people count per square meter have been investigated in detail in various studies (e.g., Hinz [13]; Sirmacek and Reinartz [14]; Perko et al. [6]). Hinz [13] estimated the background by applying a gray-level bounded region-growing approach based on sequential overlapped images. In order to filter out less crowded objects, a blob detector was used on the foreground pixels. The crowd density was then estimated by using Gaussian smoothing kernel with a fixed standard deviation/bandwidth process. Sirmacek and Reinartz [14] proposed an automated detection technique based on FAST (Features from Accelerated Segment Test) features detector to detect the images of blob-shaped and cornered structures. Using airborne video recordings, Perko et al. [6] presented a novel framework for estimating human density and motion, based on custom tailored object detection techniques, a regression based density estimation and a total variation based optical flow extraction.

Other studies focused on sensor fusion techniques for direct geo-referencing. They presented an overview of understanding algorithms developed at CMU (Carnegie Mellon University) to perform cooperative and multi-sensor surveillance. A network of smart sensors was deployed that are independently capable of performing real-time, autonomous object detection, tracking, classification, and gait analysis. Attya et al. [15] proposed a new framework based on constructing a 3D model of the desired location, and then used this model for geo-referencing the collected images and finally used the same for estimation of the crowd volume.

In the same line of this research, Helbing et al. [9] used video recordings of the crowd disaster incident in Mina/Makkah during the Hajj in 2006. This was to analyze the time frame of the crowd as well as the reason behind it.

3. Methodology of Crowd Estimation

In this paper, we focused on estimating and geo-referencing the crowd density levels through using image processing tools. This includes image segmentation for people detection, crowd density estimation and finally geo-referencing for mapping the crowd levels.

3.1 Image Segmentation

This process has been initiated in computer vision which basically changes the image into such feature that can be easily identified. Various methods have been proposed for image segmentations. In this paper, $L^*a^*b^*$ color space method have been used to classify the content of the image into readable features. This method consists of three-colored channel namely " L^* " indicating the luminosity, and the other two channels " a^* " and " b^* " indicating the chromaticity. Thus, the color is either full along " a^* " (red-green axis) or " b^* " (the blue-yellow axis) space.

In order to classify the content of an image into " a^* " and " b^* ", a space reference must be selected first. This reference includes a color region as a sample and hence the average of each sample's region is calculated in terms of " a^* " and " b^* " space. Then each image pixel is classified using nearest neighbor rule. This will allow to identify each color marker with respect to " a^* " or " b^* " value. The smallest distance will tell that the pixel most closely matches that color marker. Detailed information about LAB method can found in Bora et al. [16].

3.2 Estimation of Crowd Density

The density of crowd has been estimated through

converting the segmented image into a binary image. The crowd in the image appears as white region while other image features appear in black color. Since the black and white image (BW) allows to calculate image properties (e.g., area), the density of crowd is measured through calculating the ratio between the white and black regions. Thus, crowd level density per meter square can be measured. Regardless the number of people in the scene, the density is simply classified into low, moderate, and high.

3.3 Geo-referencing and Mapping Levels of Crowd Density

For mapping the levels of crowd density, the need for mapping is to find out the object coordinates. Note that each region is considered as an object and then it is georeferenced through using GCPs (ground control points). With known GCPs, a triangulation can be performed through mapping the XYZ of image objects with XYZ coordinates of the real world object. The point of geo-referencing is to show each object coordinate in order to identify the particular area with respect to the crowd level density. When UAV equipped with GPS and INS, direct geo-referencing can be achieved without need of GCPs. In this case, an automatic system for real-time crowd monitoring is realistic. In this paper, however; due to lack of UAV sensory data, we investigated indirect geo-referencing for mapping the crowd density.

3.4 Anomalies Detection

In any image processing procedures, anomalies/outliers may occur in the new segmented images which may lead to ambiguous information. In color-based segmentation method, classification process becomes difficult due to the similarity of colored-classes between features. For example, the color which is basically used to identify people might classify other features as people too. This problem leads to the introduction of a technique that distinguishes between feature classes. In this paper,

parameters like size, shape and position of the object have been used to achieve such tasks. A constrain based on these parameters is conditioned in the classifications. For example, the position of the crowd in the image is normally shown in flat areas (e.g., streets), thus surrounding environments such as buildings, trees and vehicles can be eliminated from image segmentation process.

4. Experiments

In this paper, different datasets have been used in order to show the effectiveness of the method for crowd density estimation. These datasets were:

- Real UAV dataset: the images were acquired via Pix4d UAV. The experiment was conducted by UNSW group. Table 1 illustrates the details of these experiments (<https://www.sensefly.com/drones/ebec.html>);
- UAV dataset: the crowd images were extracted from the internet using UAVs which were processed to show the effectiveness of the method. A video was downloaded and paused at specific time to get the picture of the crowd. The images are without GCP;
- Dataset from Hajj: images were downloaded from the internet. The images were processed to show the effectiveness of the method for crowd density estimation.

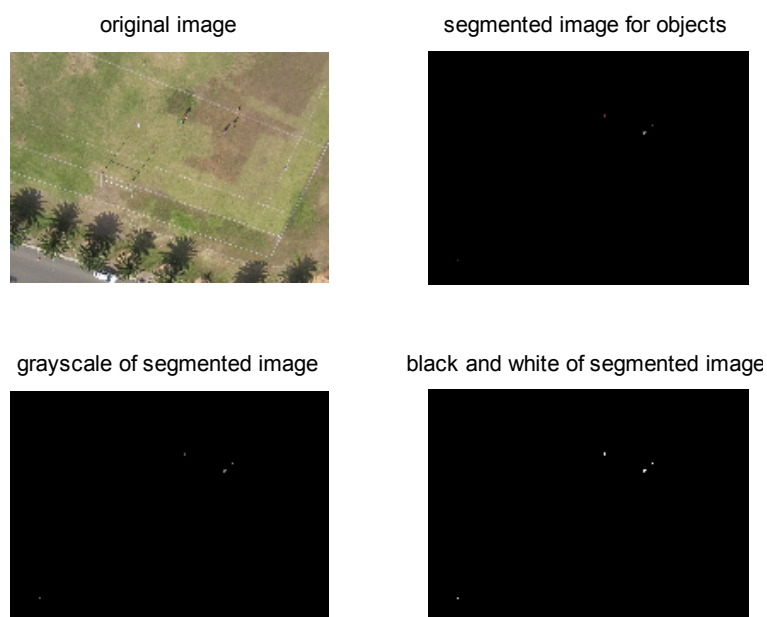
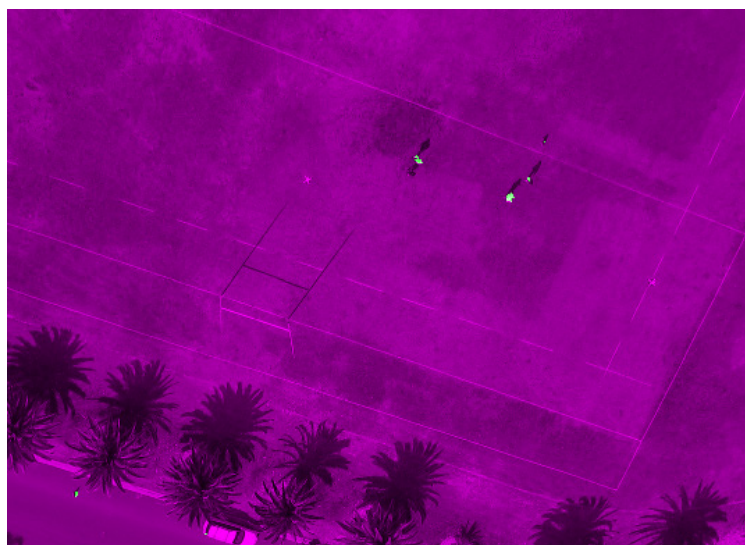
5. Results and Analysis

5.1 Identifying People in Image

In this section, two examples of real UAV images were presented to show the capability of the method for identifying people in the images. In the first example, Fig. 1 shows the original image, segmented image, grey-scale image and binary image. As shown in this example, the segmented image has identified 5 persons who already appear in the original image. In addition, the binary image indicates those persons as dots due to the low resolution. Fig. 2 presents the overlay of the segmented image on the original image to show the real position of the persons alongwith the anomalies. As shown, the green stars detect the persons in the image

Table 1 UAV experimental details.

Experiment	Using Pix4d formapping/surveying
Place	Leftous, NSW, 2013
UAV used	Sensfly eBee
Weight	700g
Sensors	GPS (5-10 metre level accuracy)
Altitude	Flying at lower altitudes, it can acquire images with a GSD of down to 1.5 cm/pixel
GCP	Available
Camera	CanonDIGITALIXUS120IS_5.0_3000x4000
Image coordinate system	WGS84
Initial values of Focal	3338.692 pixel
Optimized values	5.173 [mm]

**Fig. 1** Original image, segmented image, grey-scale image and BW image for identifying people.**Fig. 2** Image overlay that show people in original image with green stars.

with some anomalies. These anomalies caused by the shadow of each person and hence the number of objects increased to 8 instead of 5. The area of each object has been depicted in Table 2. As shown, Object No. 5 has the biggest area while the smallest area is of Objects 3 and 7. The variation in the object areas is related to the position of the object in the image as the area of person who is close to the camera is larger than those who are far away. The other reason is related to the connected objects which in this case merged in one object and thus lead into the increased area.

In the second example, the number of people and the associated anomalies have been presented. This example also demonstrates the capability of the method of identifying people in the image. The original image as well as the overlay image has been shown in Fig. 3. Almost all the people who appeared in the original image have also been identified in the overlay image along with some anomalies. These anomalies are

caused due to the similarity of color between other features and people.

Note that the purpose of presenting these two examples is to show that capability of the method of detecting a small number of people from UAV images. This method was also tested under different conditions such as high altitude, low resolution and vegetative environment.

5.2 Crowd Monitoring under Different Cases of Camera Position and Orientation

In this section, three examples for crowd monitoring were presented. Each example is representing a camera results in terms of vertical images (camera facing the ground), horizontal images from long distance and horizontal images from short distance. The details of each image including the crowd and other features such as buildings, streets, vegetation have been demonstrated in each example.

Table 2 The number of people (objects) and the area/pixel for each object

Obj/people	1	2	3	4	5	6	7	8
Area/pixel	57	28	1	151	225	53	1	3

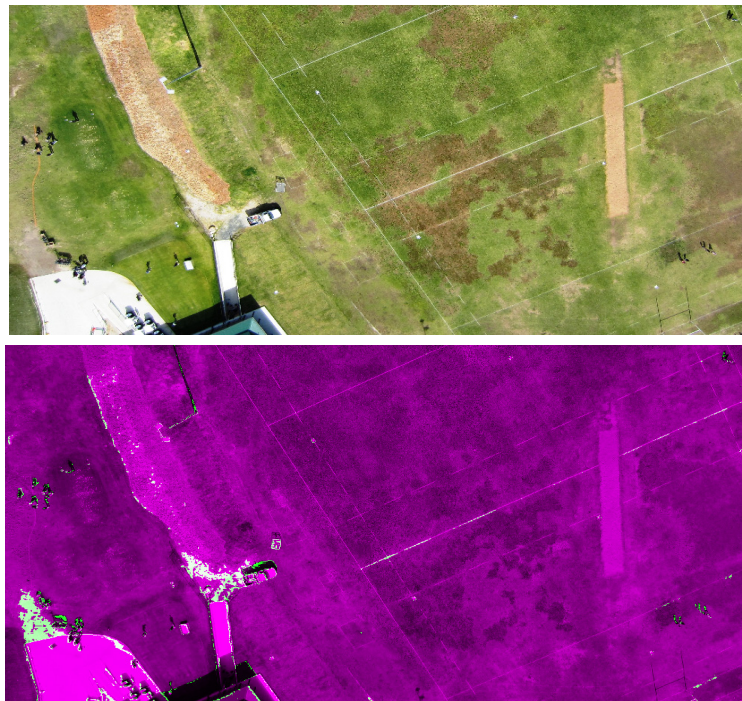


Fig. 3 Original image (top) and image overlay (bottom) that show people with green stars.

5.2.1 Example 1: Vertical Image/Camera Face Down

In this example, the image was taken from vertical camera facing the ground. As shown in Fig. 4, the content of the image includes buildings, green area, streets and the crowd. It can be noticed that the crowd is appearing as a strip following the direction of the street. This can also be seen in the segmented image in Fig. 4. In comparison between the areas of the crowd in pixels with other feature areas, it is observed that approximately 13% crowd is involved whereas the building showed the highest percentage (38%). This has been clearly shown in Fig. 5 and Table 3.

5.2.2 Example 2: Horizontal Image from Long

Distance/Camera Angle 45 Degrees

As shown in Fig. 6, the image is appearing in horizontal direction with relatively longer distance than that in previous section. The segmented image filtered out the features related to crowd only while all other features were eliminated. It can be noticed that, method for segmentation has also classified the features of the building as crowd. This is due to the color similarity. In Fig. 7, the green color which represents the crowd has been observed in other features also. Thus misleading information may give erroneous results demanding for anomaly detection and filtering procedure. This paper utilizes an auxiliary parameter for determining the crowded area based on the position of the object/crowd.

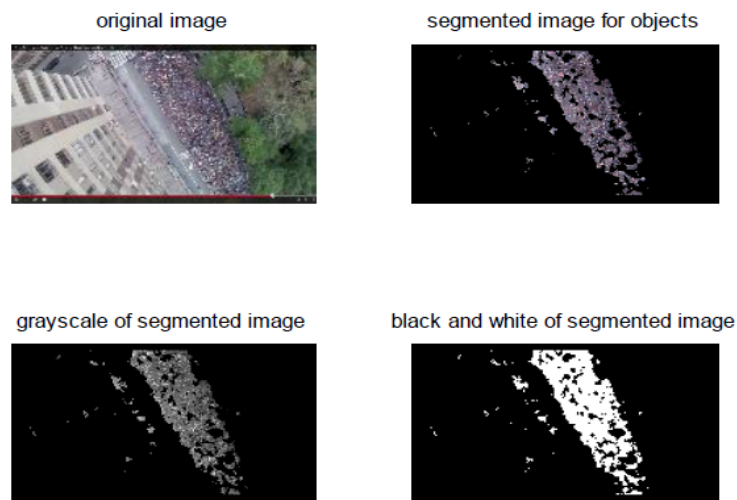


Fig. 4 Original image, segmented image, grey-scale image and BW image for crowd monitoring using UAV.

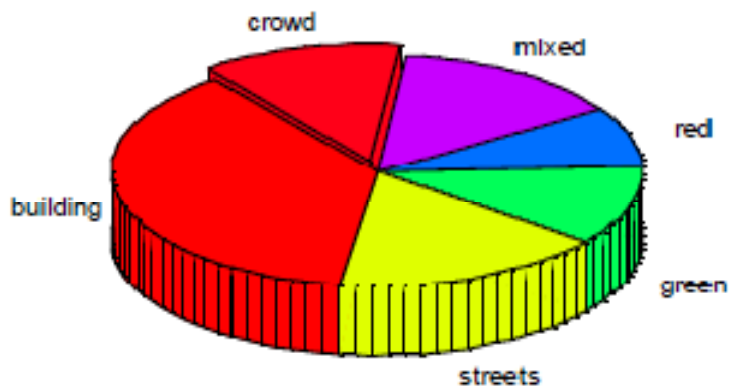


Fig. 5 Pie chart for all image classes comparing with crowd.

Table 3 The area/pixel and the percentage of crowd comparing with other classes in the image.

	Building	Street	Green	Red	Mixed	Crowd
Area/pixel	381,380.4	170,831.4	115,660.25	86,443.63	143,375.25	12,175.5
Area/percentage	37.31	16.71	11.31	8.46	14.11	12.31

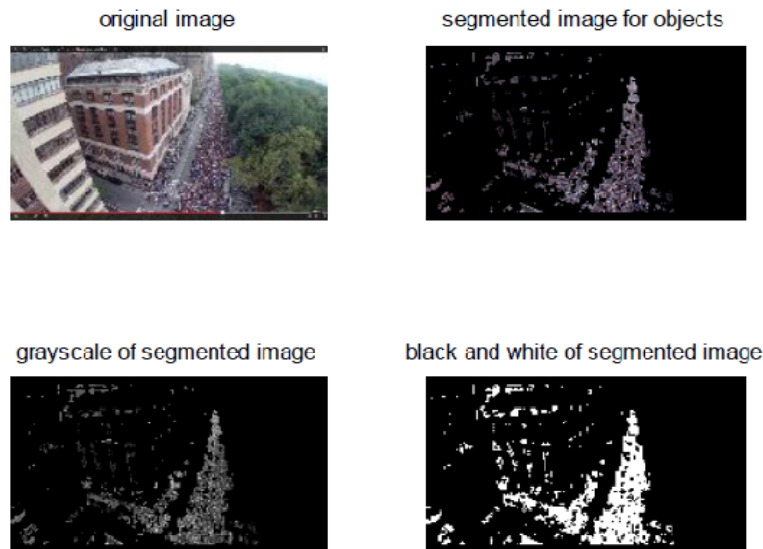


Fig. 6 Original image, segmented image, grey-scale image and BW image for crowd monitoring using UAV

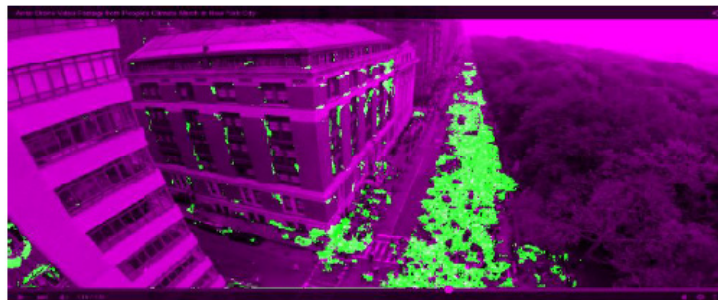


Fig. 7 Images overlay that show crowd in original image with green color.

Table 4 The area/pixel and the percentage of crowd comparing with other classes in the image.

	Building	Street	Green	Red	Mixed	Crowd
Area/pixel	328,039.9	235,014.1	20,0761.5	49,844.63	88,764.75	124,819.125
Area/percentage	31.94	22.88	19.56	4.86	8.65	12..16

As a result, the surrounding environment is eliminated and only the crowded areas remain. The image enhancement tools are then used to enhance image segmentation.

In this example, according to the positioning constrains, the building features can be removed from the image because it does not belong to the crowd area. Thus, a huge number of anomalies can be eliminated. As can be seen in Table 4, the crowd takes 12% of the total area in comparison with the other

feature classes. If the surrounding environment is removed, the accuracy of the segmentation will be improved.

5.2.3 Example 3: Horizontal Image from Short Distance/Camera Angle 45 Degrees

This example is similar to Example 2 except the only difference here is the distance between the camera and the crowd. As can be seen in Fig. 8, the crowd was distinguished from other feature classes with some anomalies that give erroneous results and misleading

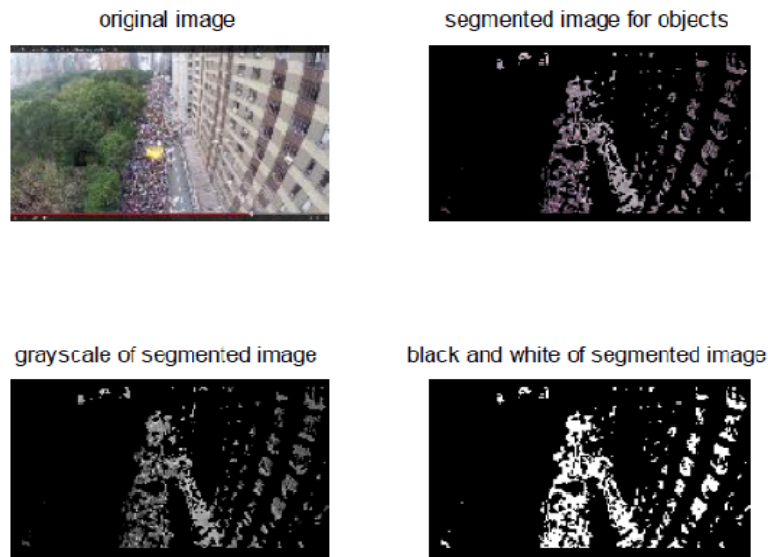


Fig. 8 Original image, segmented image, grey-scale image and BW image for crowd monitoring using UAV.

Table 5 The area/pixel and the percentage of crowd comparing with other classes in the image.

	Building	Street	Green	Red	Mixed	Crowd
Area/pixel	300,936.4	205,097.5	149,221.5	51,137.38	211,812.25	128,050
Area/percentage	28.77	19.61	14.27	4.89	20.25	12.24

information. As can be seen in Table5, the crowd again takes 12% of the total area in comparison with the other feature classes.

5.3 Crowd Monitoring in Hajj Area

Fig. 9 shows the segmented image of crowd during Hajj using close range images. It can be noted that the segmentation method distinguishes the crowd accurately. This is due to limited number of feature classes in this image. Mostly in Hajj areas, the color of feature classes is either white or black or one or two classes with mostly black and white. Thus, distinguishing people in this area is relatively easier by color based segmentation method. In Fig. 10, an overlay image is presented to show the anomalies. It can be noted that surrounding environment are shown as crowd due to similarity in color. Once again, when processing for crowd detection in an image, the position of crowd in specific area must be specified so as to eliminate other surrounding environment. If in this case, the building is eliminated, the results will be more accurate.

In Hajj season, crowd management is a severe problem which needs an effective action. While preparing this paper, hundreds of pilgrims lost their lives in Makkah/Mena due to a stampede. Thus, finding an effective method for estimating the crowd density in the area within a specific time is extremely important. In order to manage the crowd, the following shall be given careful consideration:

- Estimating the density of the crowd per square meter;
- Estimating the flow of the crowd in terms of time and direction;
- Georeferencing the crowd in order to find the position of crowd;
- Mapping the area in order to find alternative paths if the density of crowd is high.

In any case, the system must be comprehensive enough to tackle all crowd management problems. The system must include three principle capabilities including sensing, alerting, and action. By sensing, the purpose is to gather dataset using sensors such as

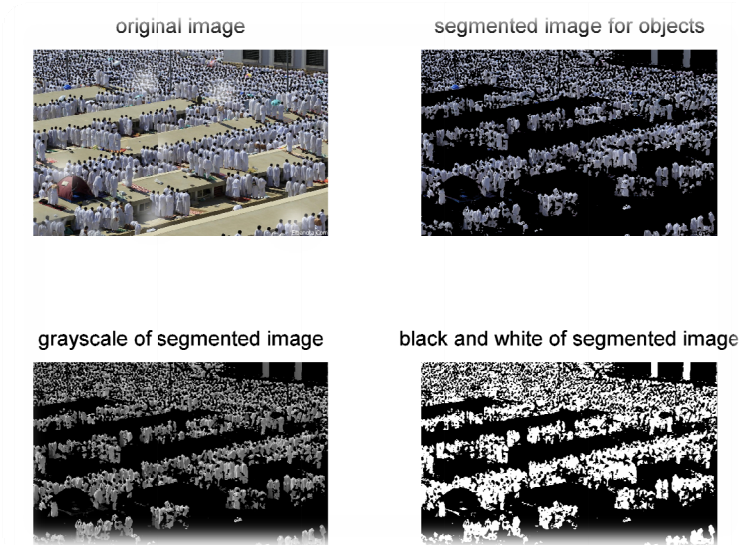


Fig. 9 Original image, segmented image, grey-scale image and BW image for crowd monitoring using close range camera.



Fig. 10 Image overlay (green color) on the original image.

digital cameras, positioning sensors (GPS/INS, communication devices) mounted on a vehicle (e.g., UAV) and transmitting the data to the first responder. The first responder (ground control station) then processes the data and convert it to information necessary to find the severity of the problem and then alerts the second responder. The second responder is then required to take an action based on the received information.

6. Conclusion

Crowd management can be performed through the estimation of crowd density level. Different image cases were presented in this paper in order to estimate

the crowd density levels. These simulated scenarios were based on the position and the orientation of the camera onboard of UAV. The image classifier method based on color segmentation was employed to detect and identify people/objects and their area in pixels. It has been concluded that image segmentation used herein has the capability to detect and distinguish people from other surrounding environment. In addition, the crowd density estimated from binary image show up to 80% accuracy level when using other parameters for image segmentation like the position of the crowd.

Furthermore, image segmentation in Hajj cases works well due to limited number of feature classes

which were mostly black and white. Future work will include mapping the levels of crowd based on the density of crowd and geo-referencing the area in order to find alternative paths for crowd.

References

- [1] Fruin, J. J. 1993. "The Causes and Prevention of Crowd Disasters." In *Engineering for Crowd Safety*, edited by Smith, R. A., and Dickie, J. F. Amsterdam: Elsevier, 99-108.
- [2] Helbing, D., and Johansson, A. 2010. "Pedestrian, Crowd and Evacuation Dynamics." *Encyclopedia of Complexity and Systems Science* 16: 6476-95.
- [3] Helbing, D., Buzna, L., Johansson, A., and Werner, T. 2005. "Self-organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions." *Transportation Science* 39 (1): 1-24.
- [4] Schadschneider, A., Klingsch, W., Klupfel, H., Kretz, T., Rogsch, C., and Seyfried, A. 2009. "Evacuation Dynamics: Empirical Results, Modeling and Applications." *Encyclopedia of Complexity and Systems Science* 3: 3142ff. Berlin: Springer, Berlin.
- [5] Schreckenberg, M., and Sharma, S. D., eds. 2002. *Pedestrian and Evacuation Dynamics*. "Simulation of Pedestrian Crowds in Normal and Evacuation Situations.", by Helbing, D., Farkas, I., Moln'ar, P., and Vicsek, T. Heidelberg: Springer-Verlag, 21-58.
- [6] Perko, R., Schnabel, T., Fritz, G., Almer, A., and Paletta, L. 2013. "Counting People from above: Airborne Video Based Crowd Analysis." Presented at OAGM/AAPR Workshop 2013.
- [7] Rodriguez-Canosa, G., Thomas, S., Del Cerro, J., Barrientos., A., and MacDonald., B. 2012. "A Real-Time Method to Detect and Track Moving Objects (DATMO) from Unmanned Aerial Vehicles (UAVs) Using a Single Camera." *Remote Sens.* 4: 1090-111.
- [8] Witayangkurn, A., Nagai, M., Honda, K., Dailey, M., and Shibasaki, R. 2011. "Real-Time Monitoring System Using Unmanned Aerial Vehicle Integrated with Sensor Observation Service." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-1/C22 UAV-g* 2011. Presented at Conference on Unmanned Aerial Vehicle in Geomatics, Zurich, Switzerland.
- [9] Helbing, D., Johansson, A., and Zein Al-Abideen, H. 2007. "Dynamics of Crowd Disasters: An Empirical Study." *Physical Review E* 75 (4): 1-7.
- [10] Helbing, D., Molna, P., Farkas, H., and Bolay, K. 2001. "Self-organizing Pedestrian Movement." *Environment and Planning B: Planning and Design* 28 (3): 361-83.
- [11] Al-Shaery, A. 2017. "Embedding the Dimensions of Sustainability into City Information Modelling." Presented at the 19th International Conference on Sustainable Architecture, Environment and Engineering, Paris, April 18-19.
- [12] Talukder, A., Goldberg, S., Matthies, L., and Ansar, A. 2003. "Real-Time Detection of Moving Objects in a Dynamic Scene from Moving Robotic Vehicles." Presented at IEEE Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, October 2003.
- [13] Hinz, S. 2009. "Density and Motion Estimation of People in Crowded Environments Based on Aerial Image Sequences." In: *ISPRS Hannover Workshop on High-Resolution Earth Imaging for Geospatial Information*, Vol. 1.
- [14] Sirmacek, B., and Reinartz, P. 2011. "Automatic Crowd Density and Motion Analysis in Airborne Image Sequences Based on a Probabilistic Framework." In: *Computer Vision Workshops (ICCV Workshops), IEEE International Conference on, IEEE*, 898-905.
- [15] Attya, H., Habib, A., Detchev, I., and Rawabdeh, A. 2012. "Crowd Volume Estimation Using Photogrammetric Techniques." Presented at ASPRS 2012 Annual Conference, Sacramento, California, March 19-23.
- [16] Bora, D., Gupta, A., and Khan, F. 2015. "Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation." *International Journal of Emerging Technology and Advanced Engineering* 5 (2): 192-203.