Hindawi

Research Article

Automated Decision Technique for the Crowd Estimation Method Using Thermal Videos

Nermin K. Negied ⓘ,[1] **Ayman El-Sayed** ⓘ,[2] and **Asmaa S. Hassaan**[2]

[1]*Communication and Information Engineering Department, Faculty of Engineering,*
University of Science and Technology, Zewail, Egypt
[2]*Computer Science and Engineering Department, Faculty of Electronic Engineering, Menoua University, Menouf, Egypt*

Correspondence should be addressed to Nermin K. Negied; nnegied@zewailcity.edu.eg

Counting and detecting the pedestrians is an important and critical aspect for several applications such as estimation of crowd density, organization of events, individual's flow control, and surveillance systems to prevent the difficulties and overcrowding in a huge gathering of pedestrians such as the Hajj occasion, which is the annual event for Muslims with the growing number of pilgrims every year. This paper is based on applying some enhancements to two different techniques for automatically estimating the crowd density. These two approaches are based on individual motion and the body's thermal features. The essential characteristic of crowd counting techniques is that they do not require a previously stored and trained data; instead they use a live video stream as input. Also, it does not require any intervention from individuals. So, this feature makes it easy to automatically estimate the crowd density. What makes this work special than other approaches in literature is the use of thermal videos, and not just relying on a way or combining several ways to get the crowd size but also analyzing the results to decide which approach is better considering different cases of scenes. This work aims at estimating the crowd density using two methods and decide which method is better and more accurate depending on the case of the scene; i.e., this work measures the crowd size from videos using the heat signature and motion analysis of the human body, plus using the results analysis of both approaches to decide which approach is better. The better approach can vary from video-to-video according to many factors such as the motion state of humans in this video, the occlusion amount, etc. Both approaches are discussed in this paper. The first one is based on capturing the thermal features of an individual and the second one is based on detecting the features of an individual motion. The result of these approaches has been discussed, and different experiments were conducted to prove and identify the most accurate approach. The experimental results prove the advancement of the approach proposed in this paper over the literature as indicated in the result section.

1. Introduction

The steady population growth, along with the worldwide urbanization, has made the crowd phenomenon something that happens in a frequent manner; consequently, study and analysis are needed in both technical and social research disciplines. The best known definition of "crowd" is the presence of a large or huge number of persons gathered or considered together and linked by a common interest or activity.

This crowd takes on various aspects: cultural, economic, artistic, religious, and popular. Crowd size estimation is a method used to measure the density of individuals in a crowd. The problem of estimating an individual's number occurs in cases such as events in open areas, streets, and protests. This problem causes many different fatal accidents that in some cases may lead to real disasters [1].

Crowd counting or crowd size estimation and control methods are important to give complete safety to the individuals composing the group also increase the ability to

avoid accidents caused by overcrowding appears in such human gathering events by automatically measuring crowd density, which can allow the prediction of the disaster before it happens and consequently prevent it. This type of system and research area may also be used as the first aid for avoiding or mitigating the accidents that happen annually in the event of HAJJ. On 24 September, 2015, a crowd collapse resulted in the death of at least 2,236 people [2] in Mina, Makah, making it the deadliest Hajj in history [3–5]. Different judgments of the number of victims were obtained. The Associated Press reported that 2,411 people were killed [6, 7], whereas Agency France Press reported that 2,236 were injured . So, this paper helps in enhancing the crowd measurement calculations to overcome the crowding problems and risks by presented a robust technique for estimating the density of individuals in the overcrowded areas. This technique based on a combination between two most widely spreading approaches in the crowd estimation field, namely, individual motion characteristics and the heat signature of each individuals.

There are three different night-vision techniques. These techniques are called near infrared illumination, low light imaging, and thermal imaging [8]. Thermal imaging technique differs from the two other techniques because it can work in dark places without any light. The thermal images can also penetrate obscurants such as vapor, smoke, and fog using near-infrared lighting techniques. The technique that thermal imaging is based on is summarized in the following steps: The first one is the infrared energy emitted from the object which is widely known as the object's heat signature. The intensity of the emitted radiation depends on the temperature of the object. Thermal imaging is based on the concept of heat sensor which can discover the small bit differences in temperature degrees. This device based on collecting the object infrared radiation and producing an image which electronically depends on the differentiation of temperature degrees. Depending on the fact which reported that the object temperature degree is rarely equaled, thermal image can estimate it and distinguish them. Thermal imaging nowadays contribute to many research areas like medical, agriculture, and navigation domains [9–11]. Thermal images have also been recently preferred in crowd counting research because they ignore the effects of shadow and illumination which affect the results a lot when normal color images used.

This paper is organized as follow: Section 2 presents literature review, Section 3 presents proposed approach. Section 4 shows the results. Finally, Section 5 concludes this paper.

2. Literature Review

2.1. Crowd Counting in Visible Bands.
Zhang et al. [12] presented the newest method for solving the crowd counting problem in open scenes or outdoor scenes based on a deep involutional neural network. This deep learning method has a robust ability to characterize individuals in the crowd rather than texture feature extracted by using the traditional method. Their approach used two associated learning global crowd count and density evaluation map during the training phase. These two associated schemes that are used for training support each other and decrease the loss amount. The approach also used the data-driven methodology to select patterns from the training set as a sample to adapt the convolutional neural network (CNN) pretrained approach to cases in which the target is not visible. The authors indicated that they applied this approach to 108 different crowds with approximately 200,000 individuals per scene, and the experimental result indicated that the proposed approach achieved more precious accuracy rate. Figure 1 presents the layers of their approach.

In 2019, Basmallah et al. [13] used scale driven CNNs to detect human heads from crowded scenes to measure crowd size. The authors mentioned that they achieved good results over state-of-the-art approaches, and that the SD CNN is a very good method to measure crowd size provided that the data used contains very high-resolution images. Khan and Basmallah [14] In 2020, researchers tried to detect human head using two types of convolutional neural networks (CNNs), and they figured out that the head is very small compared to a scene taken by a very faraway camera to capture very crowded scene images. A year later, Khan et al. [15] tried the deep CNNs to detect human heads again and measure the crowd size accordingly from Hajj occasion videos, but again the authors found that the accuracy obtained by their system was lower than the ground truth because of the smallness of human heads in crowded scenes, taken by a faraway camera in the same year.

2.2. Crowd Counting in Thermal Bands.
Human biometrics recently moved from visible bands to thermal bands, especially pedestrians detection. Literature research proved that thermal bands can outperform many of the visible bands problems, a fact that has attracted researchers to use thermal bands in the field of crowd counting and control [16–18].

In 2012, Abuarafah et al. [19] proposed another approach for detecting and tracking pedestrians and evaluating crowd density using infrared thermal (IR) imaging. The main objective of their approach was to estimate pilgrims in Hajj in which about 3.5 million pedestrians gather in one place (Makkah). In order to prevent the critical disasters, the crowd monitoring systems must estimate the crowd size in a real-time manner to take an immediate and accurate decision. Thermal camera was used for the monitoring process. The special software program has been implemented to analyze thermal images in real time using Matlab. The authors confirmed that their approach achieved a satisfactory accuracy rate for crowd size estimation. Figure 2 shows the schema of this approach [20].

The experimental result of this approach denotes that the crowd estimation accuracy rate is not accurate in the case that uses thermal camera without removing the background. Also, the accuracy rate decreases by using normal images and without removing the background because the individual's shadow which is eliminated when using thermal cameras. Figure 3 shows a sample result from their approach.
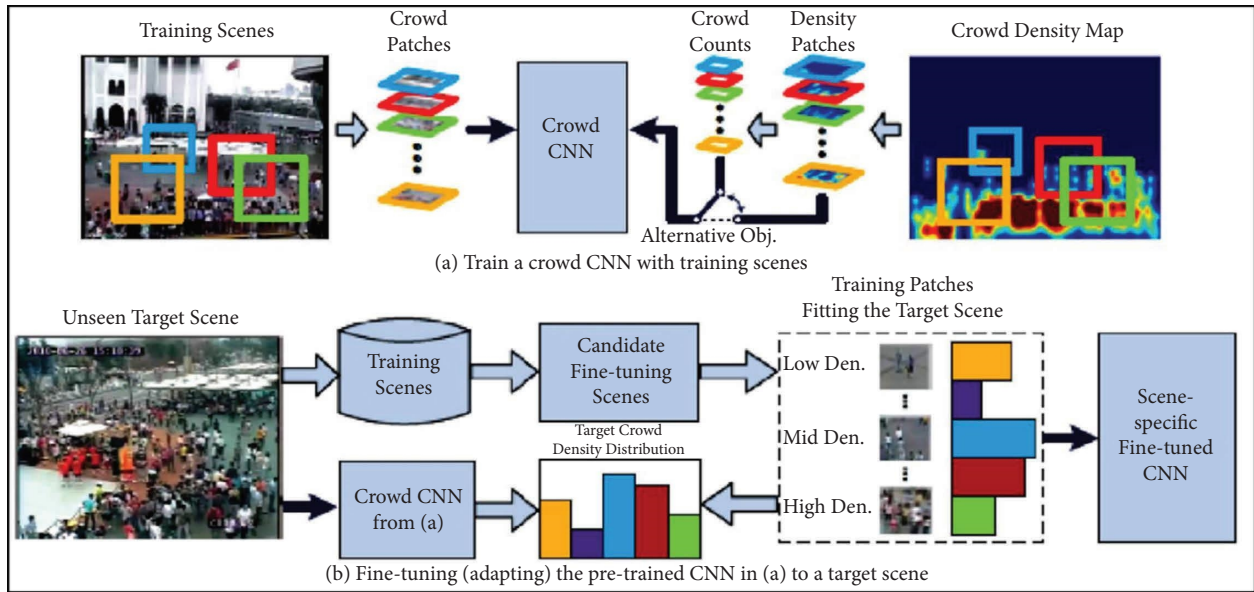
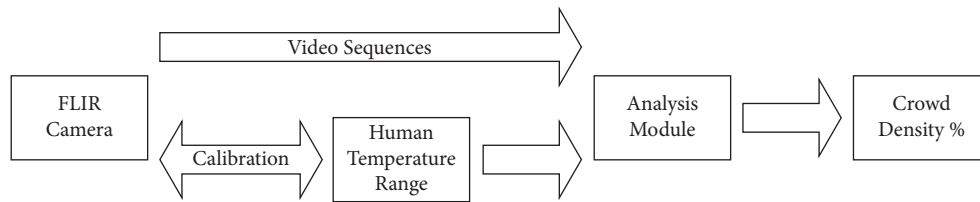FIGURE 1: The scheme that presents the Zhang et al. proposed approach [12].

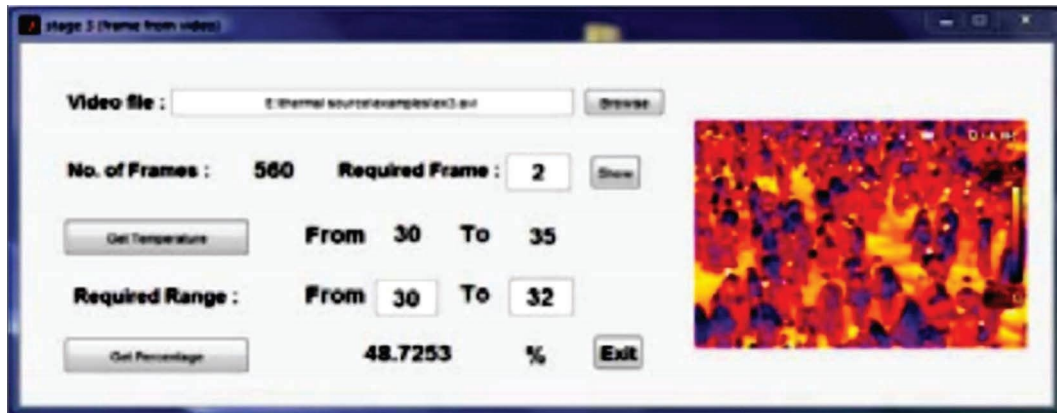FIGURE 2: The framework of the proposed algorithm is given in [20].

FIGURE 3: CMINS sample result [19].

In 2016, Negied et al. [20] have proposed a new hybrid methodology to combine the human's heat signature with background subtraction (HSBS) to give more accurate results than crowd monitoring using infrared thermal video sequences (CMINS).

In cases such as fully mobile crowds, the avoidance of false positive and false negative cases was done by merging and combining those two techniques. The motion detection techniques cannot solve the cases in which the individuals are static, but HSBS solved this problem. Motion detection techniques enable in solving the problem of objects that have temperature degree which is equal to the moving individual degree. A fully automatic system which consists of combining these two techniques and taking the difference between the outputs result enables the detection of the crowd size, nonhomogenous, homogeneous, and classify the crowd as fully static, fully dynamic, or mixed. To discover the abnormal cases and changes in the crowd, graphical user interface (GUI) is used monitor the crowd, as shown in Figure 4.

FIGURE 4: GUI representing the crowd density used in HSBS show classification from the people motion point of view [20].

3. Proposed Approach

This section presents proposed approaches and presents the merging mechanism between them.

3.1. The First Proposed Approach. The first approach in this paper is human detection using the heat signature. This work essentially based on enhancing HSBS research which was launched in [20] as it obtained the best results in this field of study till the publish time of it. Their approach did not consider some important facts, such as the following:

(i) First, HSBS considered the human body temperature range on covered and uncovered as one entity, while the big variance between the two temperatures may require dividing the range into two different ranges.

(ii) Second, HSBS ignored the fact that the temperature of the individual's body which is located nearer to the thermal camera is different than that of the one who has located far away from it; this fact was proven by the experiments conducted in this study.

This paper aims to introduce an enhanced approach to consider these facts. The experiments conducted to check the system's efficiency to show the higher accuracy rates resulted in the proposed approach.

The two mentioned facts are discussed in different studies are as follows:

(1) The first fact tells there is a difference between the temperatures of covered skin and uncovered skin. It is mentioned that the normal case covered skin temperature range less than uncovered body areas such as face and arms, as shown in Figure 5 [17, 18, 21 22].

(2) The second fact says there is a difference in temperature between faraway objects and the objects that are closer to the IR camera [23, 24]. This
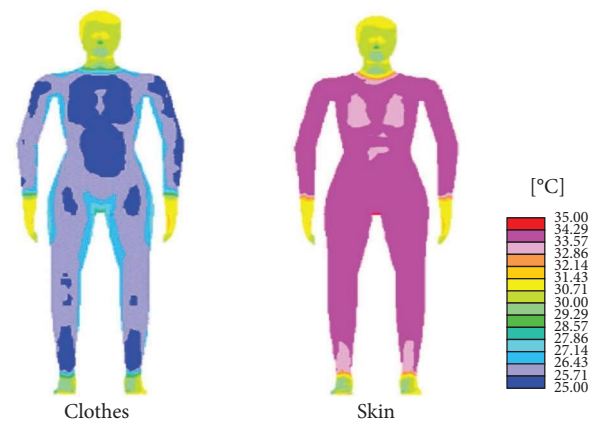

FIGURE 5: Covered vs. uncovered human skin temperatures [20].

technique has a mathematical equation which is used to calculate the difference in temperature according to the distance from a heat source. To address such an issue, the video was divided into two different parts, the first one at the bottom of the frame (closer to the camera) and the other at the top (far away from the camera). In this way, fewer temperature ranges were assigned to objects far away from the camera, and higher temperature ranges were assigned to closer objects [20].

3.2. The Implementation Methodology Used in This Approach. Crowd measure in thermal video frames use the heat signature by applying the following steps of the proposed approach that presents the basic implementation and the proposed enhancements mentioned above.

3.2.1. Frame Acquisition. Starting by the frame features' analysis, it is indicated that each frame consists of $240 * 320 * 3$ RGB pixels. Each frame contains two different boxes

to indicate the range of the temperatures in this frame. These boxes represent upper and lower degrees of temperature, as shown in Figure 6. Those boxes were then localized and cropped to get the temperature range of this frame.

3.2.2. Image Processing. Image processing is an essential part that can be divided into several steps:

(i) Converting the RGB image of the frame to gray scale image.

(ii) Then converts the image from gray to binary to become purer and clearer.

(iii) And then applying some morphological operations to the binary image resulting from the previous step to make enhancements on the image such as fill gaps by removing isolated pixels.

3.2.3. Numbers Recognition. In this stage, some processing is done to divide the resulting binary images of the lower or upper temperature value into two parts each of which contains one digit represents only one number its value between 0 and 9. The two digits are compared with the prestored digit images with the same size and shape for the ten values from 0 to 9 to recognize the accurate value for the numbers in the images, as shown in Figure 7.

The previous steps are applied to both the upper and lower boxes that represent minimum and maximum temperature values are shown in Figure 8.

3.3. Temperature Range Colors Acquired. After cropping those boxes and applying the previous steps, the temperature range of the video frame was automatically recognized.

For each frame, the maximum temperature (MAXIMUM (T)) and the minimum temperature (MINIMUM (T)) as well as the higher and lower degrees for the required temperature range (MAXIMUM (R)) and (MINIMUM (R)) values are identified. The following equations are used to calculate the difference of the temperature degrees in the frame and the difference of the required temperature range to be used in calculating the density of individuals:

$$\Delta T = MAXIMUM\,(T) - MINIMUM\,(T), \qquad (1)$$

$$\Delta R = MAXIMUM\,(R) - MINIMUM\,(R), \qquad (2)$$

where ΔT Denotes difference of temperature degrees, and ΔR Denotes difference of temperature range required.

The temperature ruler that is placed in the video frames is located at the right part of the image frame (exactly located starts in 307 in X-axis and 70 in Y-axis). The width and height of the temperature ruler are 8 and 101 pixels, respectively. Figure 9 shows a sample of the used temperature ruler.

In the following equation , a ratio value is calculated as ratio between range variation and total variation of temperature degree:
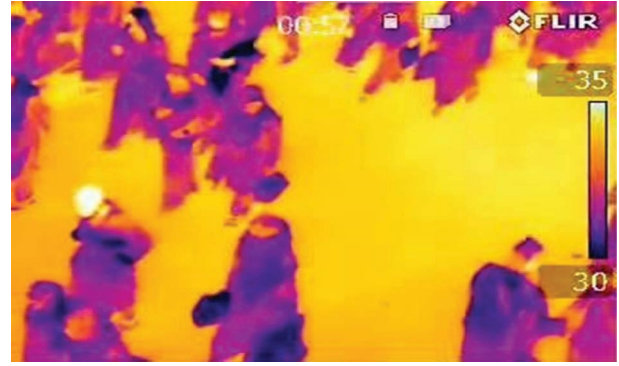

FIGURE 6: The heat temperature representation of one frame scene extracted from thermal video.

FIGURE 7: The prestored ten-digit images.

FIGURE 8: Sample of cropped boxes represent upper and lower temperature degrees, converted from RGB to gray and then to binary levels.

$$R = \frac{\Delta R}{\Delta T}, \qquad (3)$$

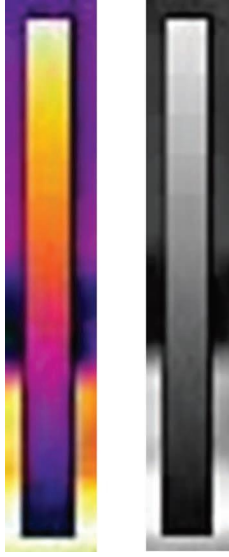where R is the ratio between range variation and total variation of temperature degree.

FIGURE 9: The heat and temperature ruler representation of a one-frame scene.

The height of the required range temperature is calculated by multiplying the previous ratio value with the whole temperature ruler height (which is 101). This height is used

to find out the size of the required range from the size of the whole ruler.

The following equation shows how to calculate the value of the range height:

$$r_h = R * 101, \tag{4}$$

Where r_h denotes range height value.

In the following equations the start position of the ruler is determined to know the part of the temperature colors representing the required temperature range as in the following steps:

(i) First, the number of rows for each temperature degree is calculated by dividing the ruler height (101) by the difference in the temperature degree in the frame.

$$rows\,for\,each\,degree = \frac{101}{\Delta T}. \tag{5}$$

(ii) Then the upper part from the ruler which is above the required range is calculated from the difference between higher temperature value in the frame (MAXIMUM T), and the higher temperature value in the range (MAXIMUM R).

$$rows\,from\,the\,top\,of\,rulers = (MAXIMUM\,T - MAXIMUM\,R) * (ows\,for\,each\,degree). \tag{6}$$

(iii) Finally, the start position of the required temperature range from the ruler is calculated by adding the rows value calculated in the previous step on the row value of the top of the ruler (Y-axis value which is 70).

$$start = rows\,from\,the\,top\,of\,rulers + 70. \tag{7}$$

So, as indicated from the above equations, the needed portion of the ruler representing the required temperature range colors are cropped from the whole temperature ruler according to the calculated start point and the range height, as shown in Figure 10.

3.4. Crowd Density Calculation. The crowd density calculation is based on getting the number of pixels with color values in the whole frame which is similar to the color values in the temperature ruler portion determined in the previous part.

This number of pixels is divided by the total number of pixels in the frame which is 320 for columns and 240 for rows. The density percentage value is estimated by multiplying the total value by 100 as indicated in the following equation:

$$Dp = \left(\frac{pixel}{r * c}\right) * 100, \tag{8}$$

where Dp refers to the density percentage value of people in the frame.

3.5. Applying Proposed Enhancements. The first enhancement is using two different ranges (covered and uncovered) of temperature instead of the one applied in the previous studies. The steps mentioned in the previous section are applied 2 times in this approach due to calculating two different temperature ranges for covered and uncovered skin, as mentioned above.

In the following figures, there are different samples for applying this enhancement by using Range 1 for covered temperature range and Range 2 for skin temperature range, as shown in Figure 11.

The second enhancement depends on dividing the frame into two parts (top and bottom). As shown in the previous figures, the percentage of the density is still high when compared with the calculation of the manual estimation. To solve this problem, another enhancement is considered to reach to more accurate results. This enhancement is based on dividing the frames of the proposed video into two parts. One part indicates the frame's top half, and the second is for the bottom. This idea is used because the acquired degree of temperature from the thermal camera for the farthest individual is lower than the nearest one. The thermal camera position and angle show that the individuals who are in the top part of the video frame appear farther away than the individuals in the bottom. The proposed approach in this paper is based on using a small

FIGURE 10: Sample for the cropping range height from ruler height.

difference in the range of temperature degrees which can be one degree as a difference from top and bottom halves of the video frame, as indicated in the following samples figures, with differentiation between top half and bottom half temperature ranges as shown in Figure 12.

So, the steps mentioned in the previous section are applied 4 times in this approach due to dividing the frame into two parts and calculating two different temperature ranges for each part to cover skin and clothes ranges, as mentioned above.

The final percentage is the summation of the average skin percentage calculations and the average covered skin calculations to get the final density percentage of the heat signature calculation.

3.6. The Second Proposed Approach. The second proposed approach in this paper uses motion analysis by merging background subtraction and frame differencing techniques.

To calculate the crowd density of individuals, we used two techniques of individual motion analysis. These two techniques are known as frame difference and background subtraction techniques. These two techniques are based on the individual motion speed value. This speed value is used to estimate the value of a threshold used to indicate which pixels represent the moving human body. The previous studies used the mechanism of tracking the individuals to estimate the individuals' speed value. This mechanism has two different disadvantages which are as follows:

(i) First, it needs high speed processors and the huge number of calculations in every frame to track the individuals' movement which make this mechanism take more time in processing than the time represented in the difference between the compared frames. This makes this mechanism hard when used as a real-time application.

(ii) Second, the user has to enter the width and height of the covered area in the video frame to calculate the speed value from the individual tracking algorithm. That is because the speed value is calculated from

knowing the distance that the individual moves in the specific period of time. This makes this mechanism cannot be run automatically in real-time applications without user input for different videos.

This paper proposed a new method for estimating the speed value from its relationship with the crowd density value to decrease processing time which fixes the first disadvantage [18].

As the speed of walking individuals and the crowd density are proportional, they have an inverse relationship. The more crowded the scene, the slower the individual, and vice versa. The maximum speed of walking individuals reached about 4.8 km/hr in the case of a very light crowd, while the minimum speed reached about 1.2 km/hr in the case of very crowded scenes. Accordingly, the threshold is chosen to be determined according to the speed of walking pilgrims. So, the required speed value is calculated from the relation between the speed of the moving people and the crowd density which is calculated using the heat technique. This speed value is used in the threshold calculation needed in frame difference and background subtraction techniques.

For the fastest movement of about 4.8 km/hr, this happens with a density percentage less than 25%, and for the slowest movement about 1.2 km/hr, this happens with a density percentage greater than 75%.

So, two more speed ranges are proposed between minimum and maximum speed values to cover different crowd density values between very light and very crowded cases, which are as follows:

(i) Less than 25% crowd percentage > Speed is maximum which is about 4.8 km/hr.

(ii) Between 25% and 50% crowd percentage > Speed about 3.6 km/hr.

(iii) Between 50% and 75% crowd percentage > Speed about 2.4 km/hr.

(iv) More than 75% crowd percentage > Speed is minimum which is about 1.2 km/hr.

3.7. The Implementation Methodology Used in This Approach. Crowd measure in the thermal video frame using motion analysis by applying the following steps of the proposed approach that presents the basic implementation and the proposed enhancements as mentioned above.

3.7.1. Frame Acquisition. This step is similar to its equivalent in the previous approach (crowd measure in thermal video frames using the heat signature technique), but without cropping the boxes and ruler representing the temperature degree.

3.7.2. Image Processing. This step converts color (RGB) images into gray-level images.

3.7.3. Background Construction. The main step of this approach is calculating the average frame image using a simple average model where the average intensity values for each
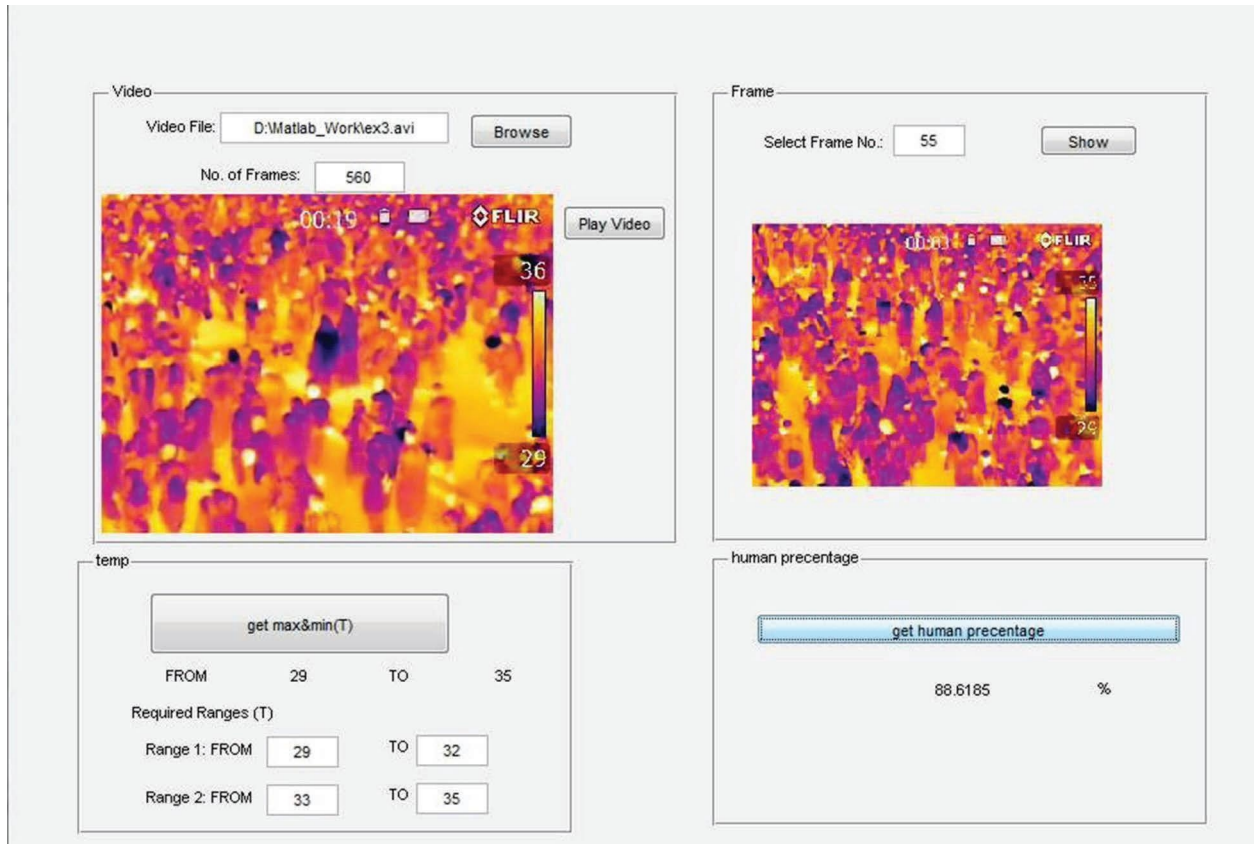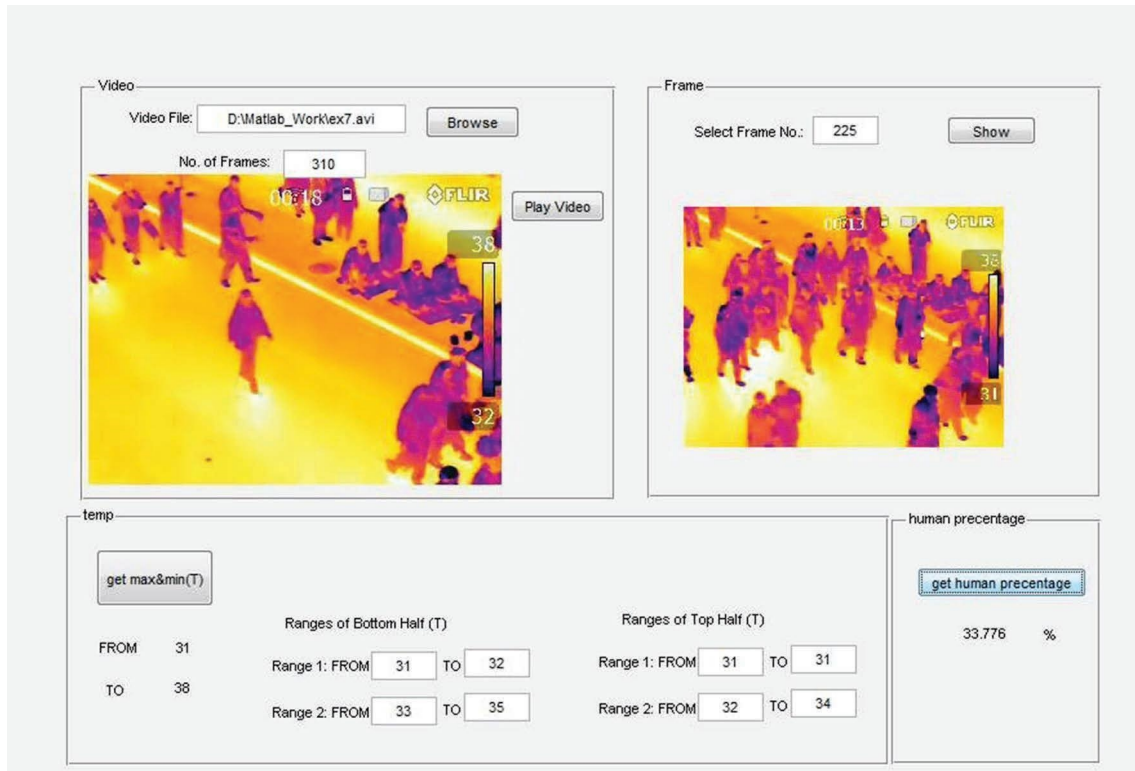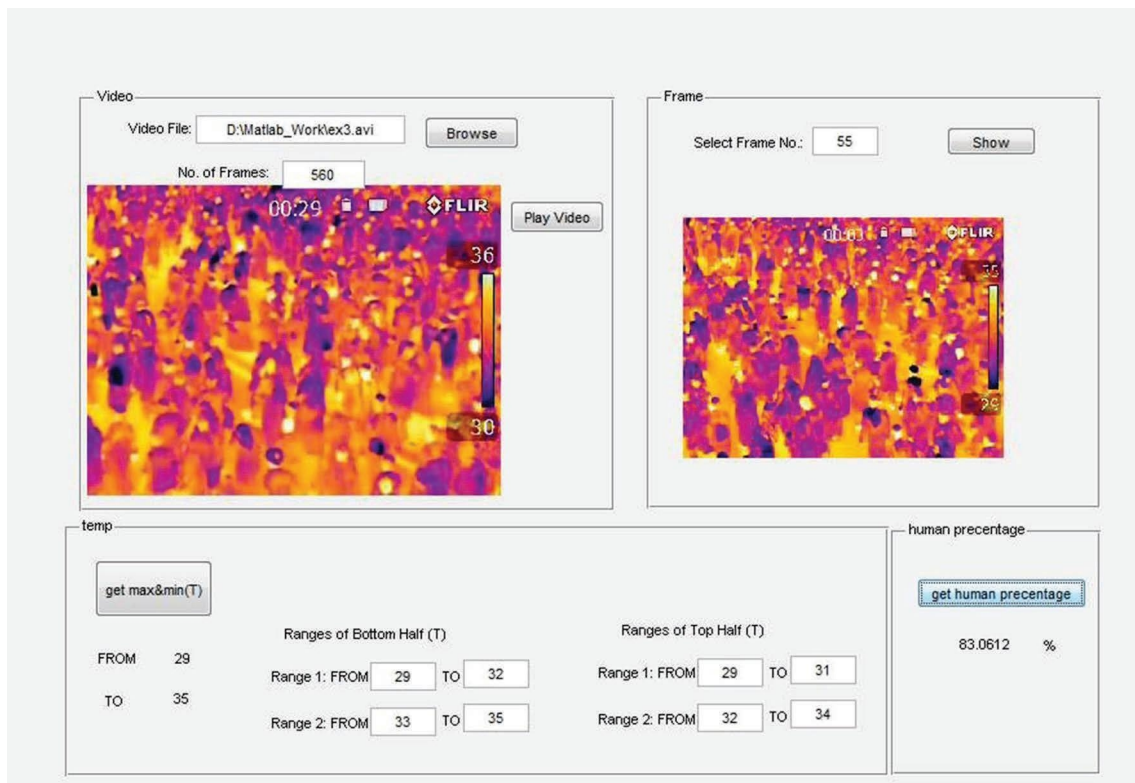
(a)
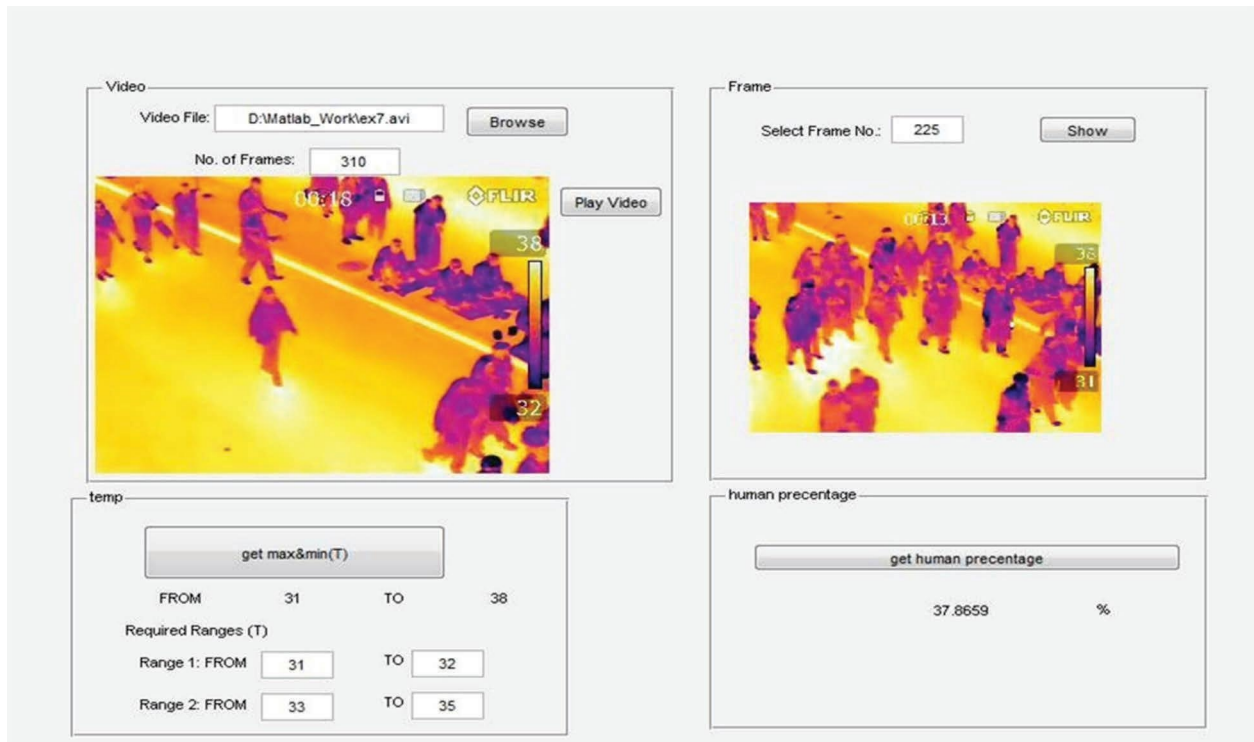
(b)

FIGURE 11: (a), (b) Samples for using two (covered and uncovered) ranges, respectively.

(a)

(b)

FIGURE 12: (a), (b) Samples of the accuracy rate for the enhancement which is based on differentiating between the individual distances from the thermal camera which is required to estimate the intensity of individuals on it.

pixel over a window of N frames are regarded as the background model [19]. The entire background modeling and segmentation process were carried out on grayscale images. Fast moving objects do not contribute much to the average intensity value, and also, very crowded videos have a background image which is not clear enough and may have many noises that affect the subtraction results. If $I^j(x, y)$ is the intensity level at coordinates $X = x$ and $Y = y$ of frame j in the sequence and bg (x, y) is the average background model value at (x, y), then use the following equation (9) to extract it.

$$bg(x, y) = \frac{1}{N} \sum_{j=1}^{N} I^j(X, Y). \tag{9}$$

After applying the above equation, the background image was extracted, as shown in Figure 13.

3.8. Foreground Extracting. Then, the gray level of the selected frame is subtracted from the average frame to calculate the difference between them which will be the moving object pixels in this frame. For each pixel in $I(t)$, take the pixel value denoted by $P[I(t)]$ and subtract it from the corresponding pixel at the same position on the background image denoted as $P[B(t)]$, as shown in equation (10).

$$[F(t)] = P[BG(t)] - P[I(t)], \tag{10}$$

where $P[F(t)]$ denotes Frame containing moving pixels. $P[I(t)]$ denotes selected frame. $P[BG(t)]$ denotes Frame background.

3.9. Calculating Motion Threshold. A threshold is put on this difference image to determine whether each difference is a real moving object or a just small difference between images due to any type of noise using the estimated speed as mentioned above and from the following equation:

$$threshold = \frac{speed}{FPS}, \tag{11}$$

where FPS denotes frames per second.

3.10. Frame Difference. Here the gray level of the selected frame subtracted from the average frame (background) and a previous frame is subtracted also from the background. The two differences subtracted from each other and compared the final difference with the calculated threshold to know the moving objects pixels between the two frames, as shown in the following equations:

$$P[F(t1)] = P[BG(t)] - P[FN1(t)], \tag{12}$$

$$P[F(t2)] = P[BG(t)] - P[FN2(t)], \tag{13}$$

$$P[FD(t)] = [P[F(t1)] - P[F(t2)]] > threshold, \tag{14}$$

where $P[F(t1)]$ denotes frame containing moving pixels only at time $N1$ $P[F(t2)]$ denotes previous frame containing

FIGURE 13: Sample of the extracted background image.

moving pixels only at time $N2$. $P[FN1(t)]$ denotes selected frame at time $N1$. $P[FN2(t)]$ denotes selected frame at time $N2$ $P[BG(t)]$ denotes background frame. $P[FD(t)]$ denotes the frame containing the difference in movement between time $N1$ and time $N2$.

3.11. Crowd Density Calculation. The crowd density calculation in this approach depends on finding the number of pixels that have the difference more than the mentioned threshold. Then, dividing this value over the total number of pixels in the frame which is $240 * 320$ (rows and columns of the image) and multiplying the result by 100 to get it as a percentage using the following equation:

$$Dp = \left(\frac{no\,of\,pixel\,of\,movement\,object}{size\,of\,image} \right) * 100, \tag{15}$$

where Dp denotes density percentage of moving pixels.

The following figure shows proposed motion technique using the estimated speed value from heat density calculations to treat the disadvantage of the mechanism mentioned above. Figure 14

4. Result

This section shows the results of the two proposed approaches, but first the hardware and software tools, dataset, and the evaluation metrics that are used to present the results are introduced.

4.1. Software and Hardware Tools. The performance of the previously discussed proposed approach tested using the MATLAB (R2015a) version 8.5.0.197613 (video and image processing toolbox,.m files), installed on PC processor Intel (R) Core (TM) i7 (M620) 2.67 GHZ/4 MB Cache, RAM 8 GB.

4.2. Data Base. Forward looking infrared camera (FLIR) is the tool that is used to collect the database which used in this paper to test and make the experiment results of this paper. Generally, the FLIR camera is usually dependent on utilizing the far infrared region. Universities in Kingdom of Saudi
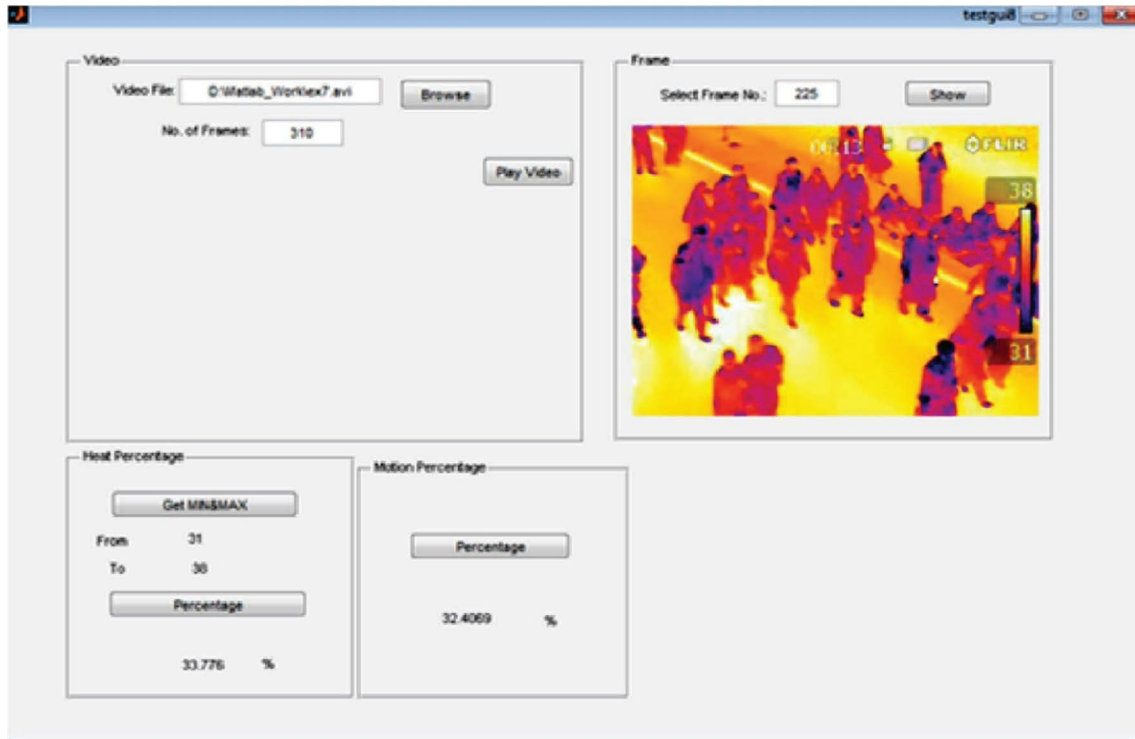
FIGURE 14: The crowd percentage for the enhanced motion subtraction approach using the frame difference.

Arabia especially (Umm El Qura University) depend on using an enormous amount of FLIR cameras in different roads which are connected between Arafat to Muzdalefa [19]. These cameras are used to capture and count the walking pilgrims' status in this specific area which is the total thermal videos scenes from the Mecca road. All the videos are with dimensions of 320 pixels width and 240 pixels height. All the captured videos were in AVI format with a rate of 18 frames per second. The lengths of the video are from 11 seconds up to 59 seconds.

4.3. Evaluation Metrics. The proposed approach was examined on the dataset of Umm El Qura University mentioned before, and the following are samples of the GUIs of the system representing how the crowd size percentage estimated by the system compared with the crowd size estimated by human eyes. Statistics of human eye estimation were collected from six different persons.

4.4. Experimental Results. It presents the real-time estimation for the video frame. This is considered as an enhancement on the previous research without depending on the width and height of the covered area. The following Figures 15–17 show samples of the real-time GUI for applying the contribution on heavy, medium, and few crowded videos. The combination of these figures depends on taking the values of heat percentage and motion percentage to overcome the problem if one technique gets a nonaccurate result. The final GUI just takes the number of seconds required to calculate the

density and lets it work in real-time, showing the video frames being analyzed with minimal calculations and processing.

The comparison of the calculated number of persons counted by eye estimation with the actual number gives the required accuracy rate to test this approach and previous studies. The difference of the accuracy rates between the proposed approach and previous approaches are afore-monitored . As shown in Figures 18–20, there are 3 results for the same video sample in the previous approaches [19–21, 24–26]. Also, the new proposed approach in this paper works on the same video sample, as shown in Figure 21.

As shown in the previous GUI (see Figures 18–21) mentioned above, the percentage of people in this video frame is exceeding the 3/4 of the overall frame. CMINS gives a crowd size of only about 48.7% by heat approach and about 65% by motion approach. While in HSBS, they got about 58.8% using the heat signature approach and about 69.8% using the motion analysis one. But in the proposed approach, the results are about 80.29% using the heat signature approach and about 72% using the motion analysis approach. Also, the proposed approach presents that it is a heavy dynamic crowd. The system obtains a new result every 30 seconds according to the calculated values in the frames passed in those 30 seconds to achieve real-time crowd surveillance and control.

As mentioned in the accuracy summary Table 1, a clear enhancement appears in heat calculations in the proposed approach about 99.7% and a higher accuracy rate value rate of motion calculations of about 92%.
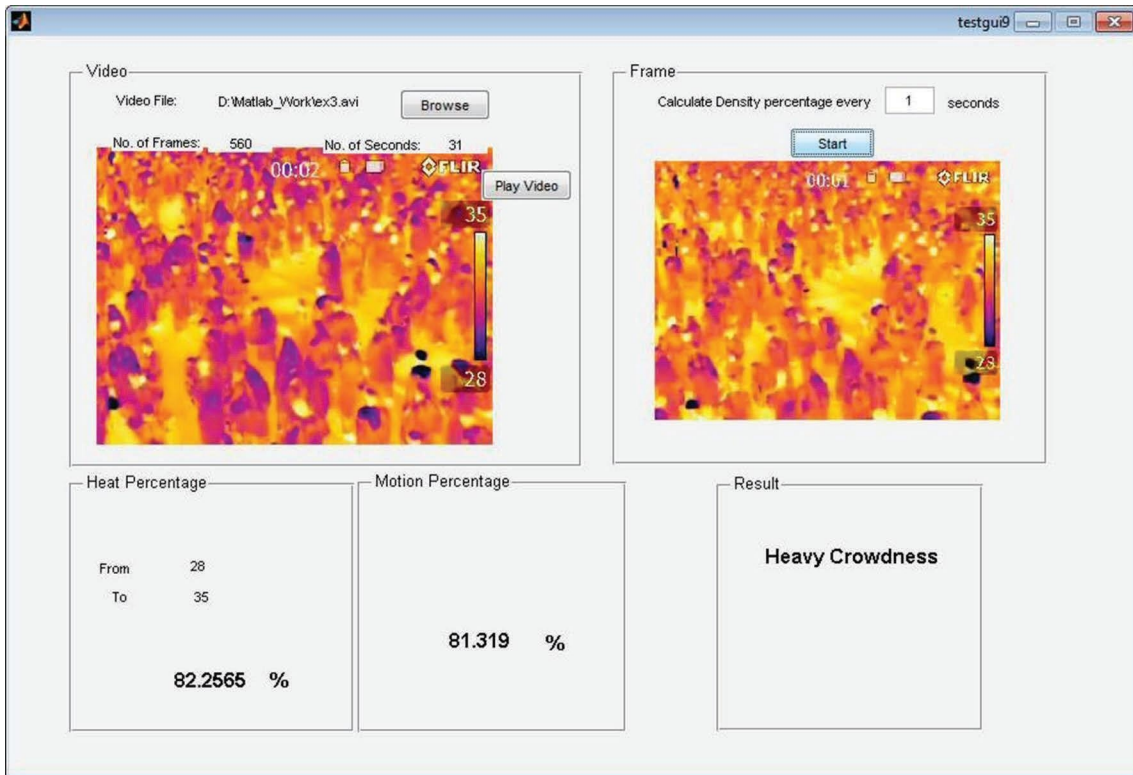
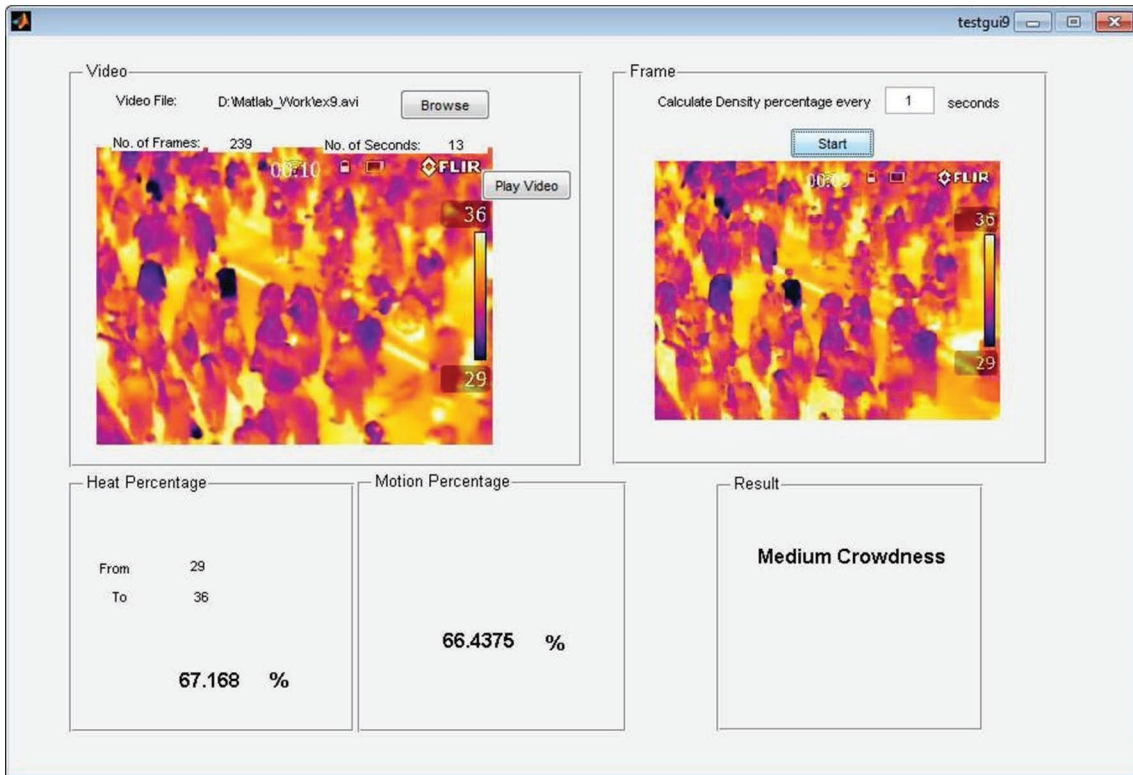FIGURE 15: Sample result of real-time GUI for heavy crowd.

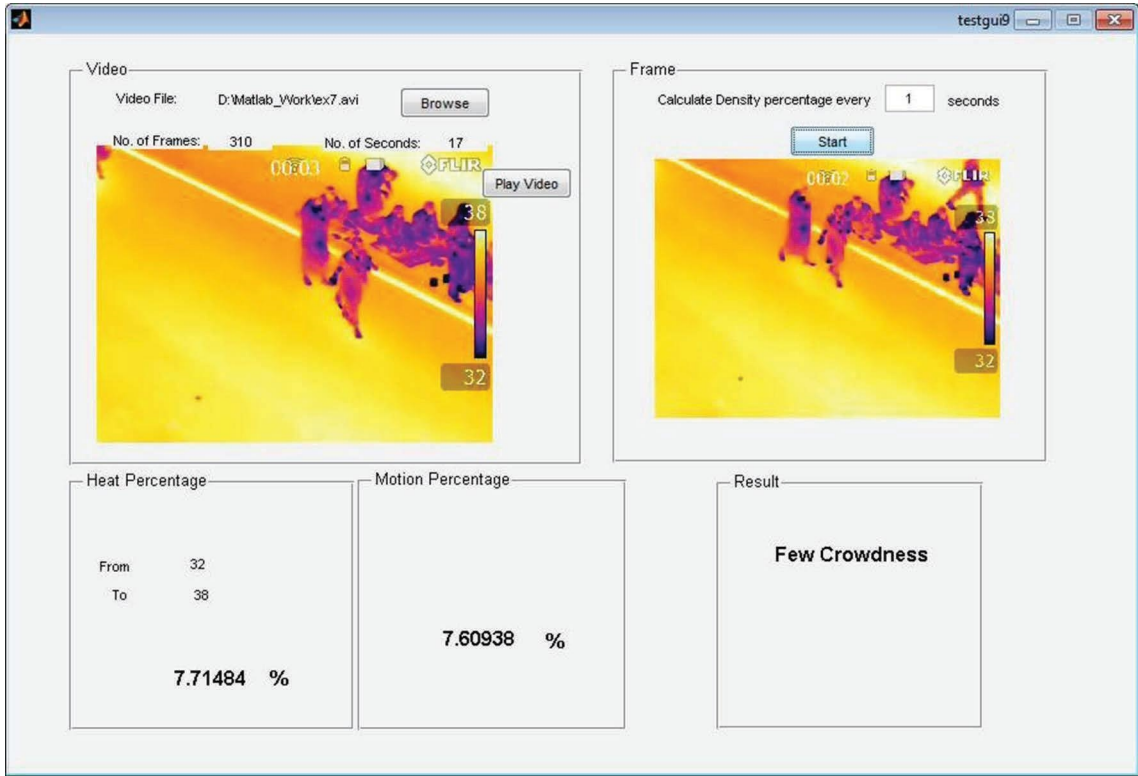FIGURE 16: Sample result of real-time GUI for medium crowd.

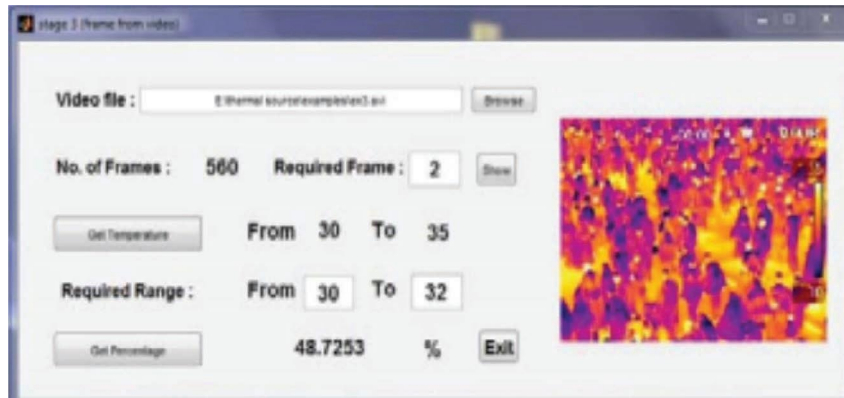Figure 17: Sample result of real-time GUI for low crowd.

Figure 18: Density calculation using a thermal image utilizing heat signature by CMINS [13].

Figure 19: Density calculation using a thermal image utilizing background subtraction by Abu Arafah [20].

Figure 20: Density calculation using a thermal image utilizing heat signature and background subtraction by HSBS [14].
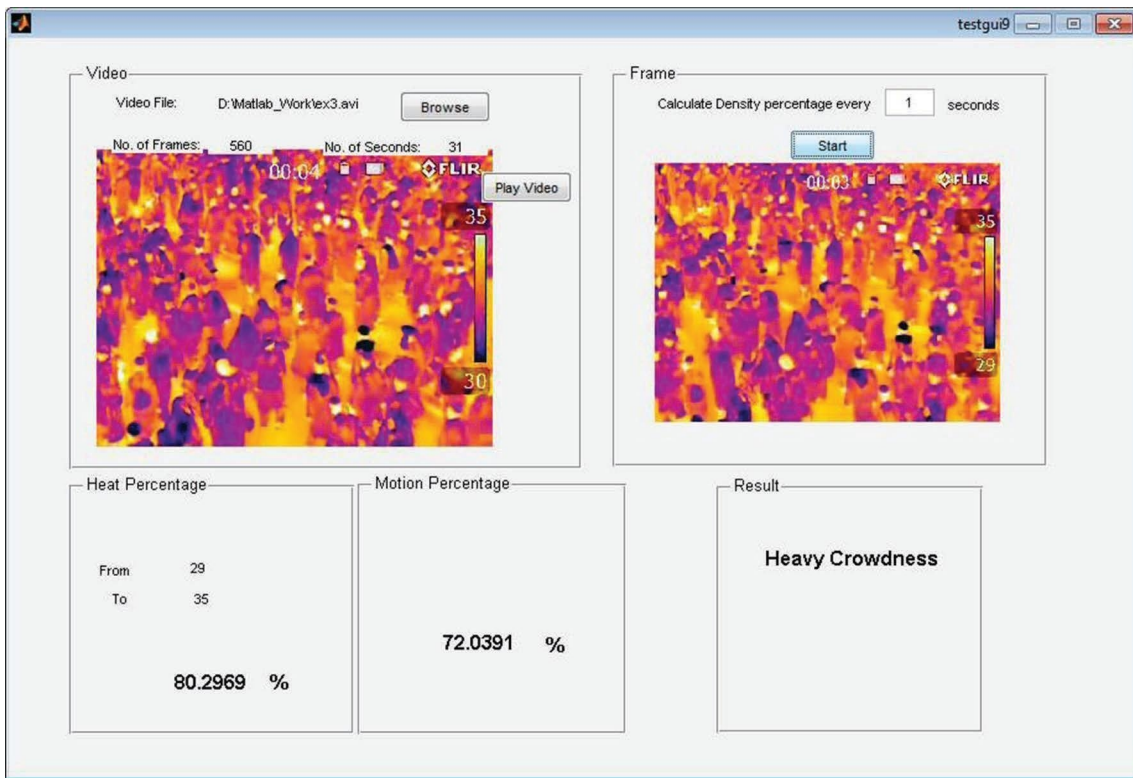

Figure 21: Density calculation using a thermal image utilizing heat signature and background subtraction by proposed approaches.

Table 1: Accuracy comparisons of different approaches.

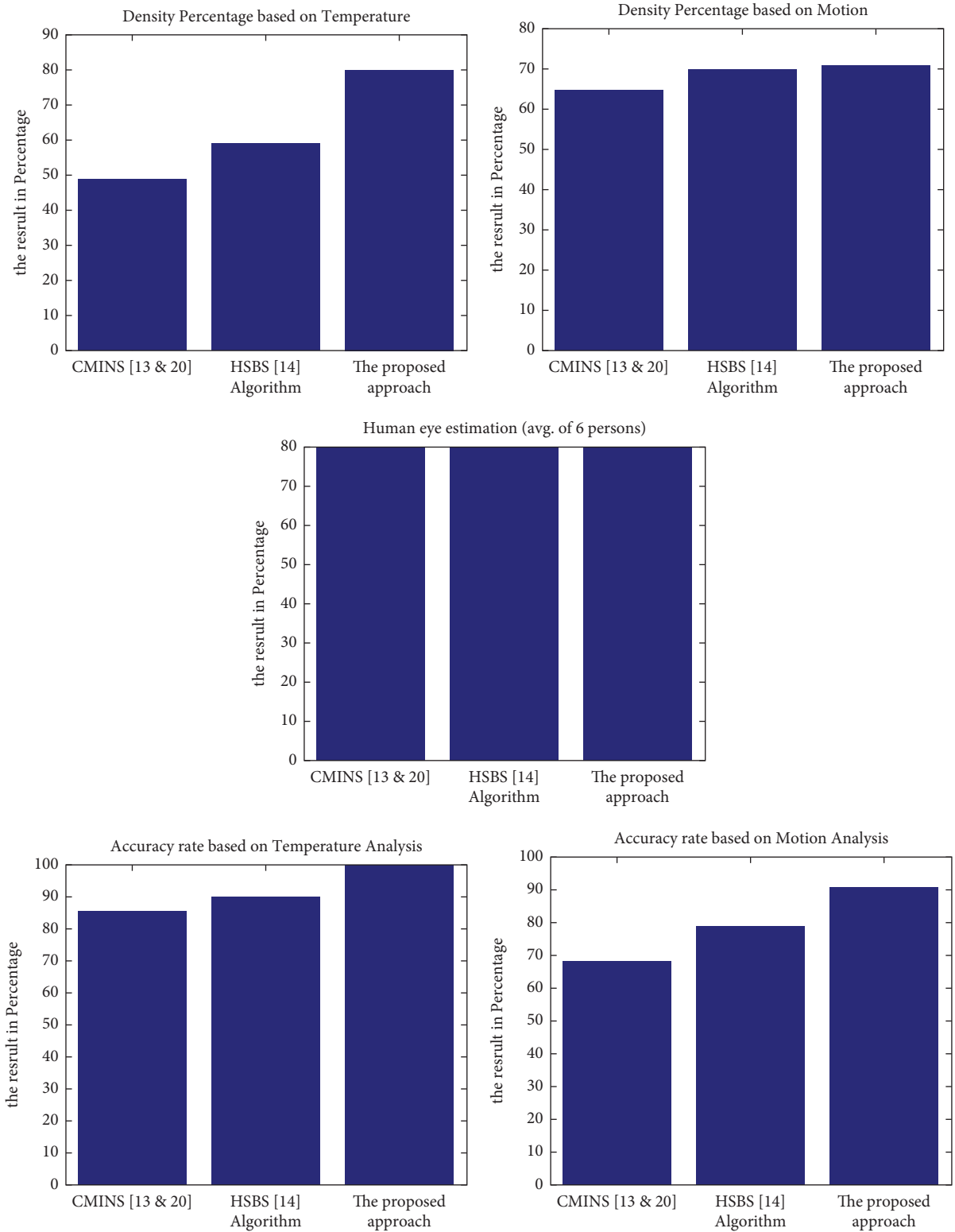| Approach	Density percentage		Human eye estimation (%)	Accuracy rate	
	Temperature result (%)	Motion result (%)		Temperature result (%)	Motion result (%)
CMINS [13, 20]	48.7	65.02		68.7	85.02
HSBS [14]	58.87	69.79	80	78.87	89.79
The proposed approach	80.29	72		99.71	92

FIGURE 22: The comparison between the two previous algorithms and the proposed approach.

Figure 22 shows the comparison among the two previous algorithms and the proposed approach result in the different metrics.

The following table compares the two previous algorithms results and the proposed approach for the same video sample at the same frame.

5. Conclusion

This paper presents enhancements to two different approaches for crowd counting estimation. This paper is also based on using video frames which are extracted using thermal video frames instead of fixed images. The enhancements on the first approach depend on measuring density by differentiating between an individual's skin temperature in both covered and uncovered situations. Also differentiate between far and near objects' temperatures. The second approach is based on measuring crowd in thermal images using background subtraction and frame difference without applying huge calculations of object tracking technique. As the experimental result indicates that the accuracy rate mentioned in this paper is more robust than the accuracy rate which was mentioned in previous different studies especially in real-time estimation and individual counting in crowded areas.

Data Availability

The data is found in literature and well described in the paper.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] B. Zhan, P. Remagnino, D. Monekosso, and S. Velastin, *The Analysis of Crowd Dynamics: From Observations to Modelling*, pp. 441–472, Springer-Verlag, Berlin, Germany, 2009.

[2] "Iran says tests will show cause of diplomat's death in Saudi," *AFP*, 2015.

[3] J. Gambrell, M. Ahmed, M. B. Osman, and A. Mazen, *Saudi Crush was Deadliest Hajj Tragedy Ever*, Yahoo! News Associated Press, New York, NY, USA, 2015.

[4] H. Gayathri, P. Aparna, and A. Verma, "A review of studies on understanding crowd dynamics in the context of crowd safety in mass religious gatherings," *International Journal of Disaster Risk Reduction*, vol. 25, pp. 82–91, 2017.

[5] J. Gambrell and A. Batrawy, *New Tally Shows at Least 1,621 Killed in Saudi Hajj Tragedy*, Business Insider Associated Press, New York, NY, USA, 2015.

[6] J. Gambrell and B. Ahmed, *Hajj Stampede in September Killed over 2,400, New Count Finds*, The New York Times, New York, NY, USA, 2015.

[7] *Iran Holds Funeral for Diplomat Killed in Saudi Hajj Crush*, Yahoo! News, New York, NY, USA, 2015.

[8] 2017 https://whatis.techtarget.com/definition/thermal-imaging.

[9] K. Nermin, "Infrared thermography based breast cancer detection comprehensive investigation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 6, 2018.

[10] R. Vadivambal and D. S. Jayas, "Applications of thermal imaging in agriculture and food industry—a review," *Food and Bioprocess Technology*, vol. 4, no. 2, pp. 186–199, 2011.

[11] S. Ward, J. Hensler, B. Alsalam, and L. F. Gonzalez, "Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision," in *Proceedings of the IEEE Aerospace Conference*, Montana, MT, USA, March 2016.

[12] H. Zhang, X. W. Li, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.

[13] S. Basmallah, S. D. Khan, and H. Ullah, "Scale driven convolutional neural network model for people counting and localization in crowd scenes," *IEEE Access*, vol. 7, 2019.

[14] S. D. Khan and S. Basalamah, "Sparse to dense scale prediction for crowd couting in high density crowds," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3051–3065, 2020.

[15] S. D. Khan, Y. Salih, B. Zafar, and A. Noorwali, "A deep fusion network for crowd counting in high density crowded scenes," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, p. 168, 2021.

[16] K. Nermin, A. Wahab, E. E. Hemayed, and M. B. Fayek, "HEARD: an automatic human EAR detection technique," in *Proceedings of the International Conference on Engineering and Technology (ICET)*, New Jersey, NJ, USA, October 2012.

[17] K. Nermin, "Human biometrics: moving towards thermal imaging," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 2, no. 6, 2014.

[18] N. K. Negied, E. E. Hemayed, and M. B. Fayek, "Pedestrians' detection in thermal bands – critical survey," *Journal of Electrical Systems and Information Technology*, vol. 2, no. 2, pp. 141–148, 2015.

[19] A. G. Abuarafah, M. O. Khozium, and E. AbdRabou, "Real time crowd monitoring using infrared thermal video sequences," *Journal of American Science*, vol. 3, 2012.

[20] N. K. A. W. Negied, E. B. Hemayed, and M. Fayek, "HSBS: a human's heat signature and background subtraction hybrid approach for crowd counting and analysis," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 8, Article ID 1655025, 2016.

[21] G. Li, J. Zhang, H. Lin, D. Tu, and M. Zhang, "A moving object detection approach using integrated background template for smart video sensor," in *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference*, vol. 1, pp. 462–466, Como, Italy, May 2004.

[22] G. Keith Still, *Crowd Safety and Risk Analysis*, https://www.gkstill.com/Support/crowd- density/, 2017.

[23] K. Stephan and A. Laesecke, "The thermal conductivity of fluid air," *Journal of Physical and Chemical Reference Data*, vol. 14, no. 1, pp. 227–234, 1985.

[24] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Molleda, and F. Bulnes, "Infrared thermography for temperature measurement and non destructive testing," *Sensors*, vol. 14, no. 7, pp. 12305–12348, 2014.

[25] A. G. Abuarafah, *Background Extraction Technique in Thermal Imagery; Case Studies during Hajj*, https://www.sciencepub.net/researcher, 2013.

[26] Abuarafah and M. Khozium, "Integration of background removal and thermography techniques for crowd density scrutinizing," *International Journal of Disaster Risk Reduction*, vol. 2, no. 1, pp. 14–25, 2013.