A Discriminative Approach for People Detection Using Color Camera for Mobile Robot Platforms

Ali H. A. Aldabbagh and Laith A. H. Al-Shimaysawee
Faculty of Engineering, University of Kufa, Najaf, Iraq
Email: {ali.h.aldabbagh, laith.anwer}@gmail.com

Nasser Asgari
School of Computer Science, Engineering and Mathematics, Flinders University, Adelaide, Australia
Email: nasser.asgari@flinders.edu.au

***Abstract*—In this paper, we explore a new algorithm to detect people with color cameras based on our modified Implicit Shape Model (ISM) implemented for grey scale thermal images. The idea of this approach is to convert a color image to a grey scale image, invert it to appear like a thermal image, and then apply the same algorithm that we used for people detection in thermal images. As the first step, we use the ISM to define the proposed centers of people locations. Then we utilize a novel method to detect people based on the density of the concentrated proposed centers by using an auto generated threshold mechanism. Our method is easy to implement and does not require complicated computations; thus resulting in a considerable increase in the speed performance and decrease in the cost of the required hardware on mobile platforms. We evaluated our system by testing it on three image sets for indoor and three for outdoor scenarios. Our system showed promising results in detecting people on images taken by different types of color cameras under difficult scenarios. This technique is used as the vision system for a rescue assist mobile robot built at Flinders University.**

***Index Terms*—color image, people detection, ISM, rescue robot**

I. Introduction

With the rapid growth of artificial intelligence and computer vision technologies in recent years, the need for robust people and object detection techniques are increasing because of their importance in many applications such as drive assist systems in modern cars, security surveillance systems, rescue assist systems, military applications, computer and robot interactive applications, and several others. Many researches have been done on people detection using various types of cameras such as normal color, night vision and thermal cameras and so on. Each camera has its own advantage and disadvantage; however, using normal color cameras are very popular in the research because of the high resolution and low cost of these cameras compared with other types of cameras. The disadvantage is that detecting people in color cameras requires more sophisticated approaches because of the clutter in the background.

Some of the proposed techniques with thermal cameras depend on detecting the hot spots [1]-[4]. Such techniques can work very well in indoor applications as well as some outdoor applications, but they have a limitation that prevents them from being robust in many situations. These algorithms assume that the temperature of the people is higher than the surrounding background, which is not always the case in most outdoor and some indoor scenarios.

Other techniques essentially depend on the background subtraction [5]-[7]. These methods can be robust in stationary surveillance camera situations, but not in the moving camera applications because of the change in the background that makes the subtraction with the previously saved background ineffective.

One of the outstanding researches is based on computing the gradients of IR test images and applying model patches to a search window. Then, the Model Histogram Ratios (MHR) are computed between the patches and classified by a trained Support Vector Machines (SVM) classifier for each window [8]. This is similar to the original Histograms of Oriented Gradients (HOG) approach first proposed in [9]. There are other methods that depend on training the system to detect specific objects based on the similarity of the local appearance between the training samples and the test images such as the Implicit Shape Model (ISM) algorithm [10]-[12]. These algorithms have the advantage that they do not depend on the background or the hot spots. The technique discussed in this paper is based on our Modified ISM Implementation for Grey Scale Thermal Images [13] which is inspired by the ISM in [10]. It takes advantage of the ISM in defining the proposed centers of the people locations, but detects people depending on the density of the concentrated proposed centers by an auto generated threshold mechanism that we have developed. The outstanding point in this research is that the same training samples used to detect people in thermal images in [13] are also used to detect people in color images. This approach significantly decreases the amount of training images, especially for a system which

Manuscript received July 1, 2015; revised December 9, 2015.

doi: 10.18178/ijmerr.5.2.103-108

uses both thermal and normal cameras at the same time. This technique has been implemented on a prototype rescue assist mobile robot at Flinders University. The task of the robot is to go to dangerous places such as destroyed buildings or unstable structures, search for injured or trapped victims, and send their locations to a rescue team who are following the robot from a safe distance.

This paper is structured as follows: In Section II, we describe the hardware of the system. In Section III, we explain our approach in details. In Section IV, we demonstrate how the system is used and tested for detecting people. Section V presents the results and evaluation of the system, and finally, Section VI includes the conclusions and the future work.

II. System Hardware Description

The prototype that we have built for a rescue assist mobile robot at Flinders University is shown in (Fig. 1). The robot can explore an area autonomously or can be controlled remotely by an operator. It scans the surrounding by rotating the camera and provides a wireless live video feed to a computer carried by the rescue team. The computer applies the vision process on the received real time video from the robot and displays the live video feed on its screen in which the detected subjects are highlighted. The details of the vision process are demonstrated in the next section.

Figure 1. A rescue assist mobile robot built at Flinders University.

III. Implementing the Vision System

The vision process in the system can be summarized into three distinct parts: Codebook Generation, Learning an Implicit Shape Model, and the Detection Process. These parts are similar to the ones used in our modified Implicit Shape Model (ISM) implementation for grey scale thermal images [13], except that the method is extended for processing the color images as shown in Fig. 2. The details are as follows:

In the recognition process, the first step as shown in Fig. 2b is to convert the color image to a grey scale image with 0 for black to 255 for white.

In the next step, the grey scale image is inverted by subtracting its pixel values from 255 to get 0 for white to 255 for black as shown in Fig. 2c. This makes a normal image look like a thermal image. Hence, we can apply the same algorithm that we have developed to detect people in grey scale thermal images [13] as shown in Fig. 2d and is summarized below:

Firstly, an interest point detector is applied on the test image and patches around these points are extracted. The interest point detector used in our system is the Harris detector [14]. There are more robust feature extractors instead of Harris detector such as Difference of Gaussian (DoG), Harris-Laplace, and Hessian-Laplace detectors in [11], [15], however, these detectors make the system slower because of the large number of points that they generate, requiring more computations. There is also a combined detector/descriptor termed as Speeded Up Robust Features (SURF) in [12], [16], [17]. Although Harris Detector produces less interest points, the use of it in our system was sufficient for detecting people in difficult scenarios as shown in the results in section V.

Secondly, each extracted patch is matched with the entries in the codebook using the Normalized Greyscale Correlation (NGC) function [10]. If the similarity is greater than a specified threshold, this codebook entry is activated. All the possible locations of the people centers for each activated codebook entry are included in the detection process [13]. This approach is different from the one used in the standard ISM version [10] and later versions [11], [12], [16], [17].

Next, the concentration of the proposed centers are used to form 3D density areas by putting a block on each center [13]. This block has a square base and the height represents the probability. As the blocks overlap, the probability of the positions that have concentrated points are increased more than the other positions, creating piles of density areas [13].

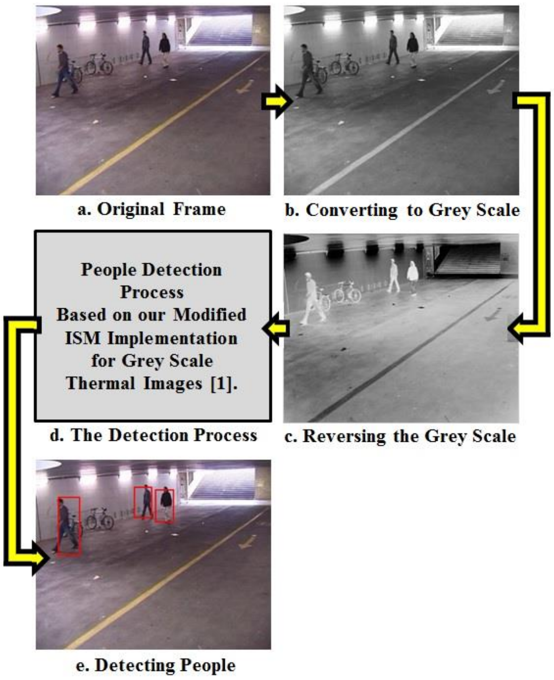


Figure 2. The recognition process.

Finally, an auto threshold technique is applied for the density area to detect people [13]. At this stage, if there are groups with less than 100 pixels, they are discarded to improve the detection accuracy [13].

IV. Experiment

A. Training Images

For the training images used in the codebook, we selected images of various people captured by different thermal cameras from the internet. We extracted people samples from these images and converted the background to black. In total, we only used eight samples, as shown in Fig. 3. One aspect of this codebook is the use of the black background to remove the interest points on them. This eliminates the need to separate the training images into positives (for people) and negatives (for background) which was the case in [10]-[12], [17]. As a result, in the codebook images, the Harris points are distributed over each person rather than the black background. The other aspect is that the number of images used in this codebook is extremely small in comparison with other ISM systems in [10]-[12], [17], which will be explained in the result section.

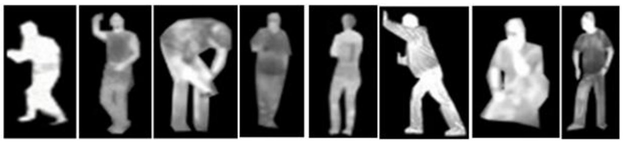


Figure 3. The training images used in this system

B. Testing the System

After preparing the codebook, we tested our system with colored images from various indoor and outdoor scenarios. For indoors, three image sets were used. The first set had 150 indoor images captured using the IP camera mounted on the robot at the Tonsley campus of the Flinders University. The other two sets had 250 images in total taken evenly from the sources in [18], [19]. There were 224 people in the first image set, 444 people in the second set, and 239 people in the third set.

Similarly, three image sets were used for outdoor testing. The first set had 150 images which were taken from the camera on the robot at Flinders University. The other two sets had 250 images in total taken evenly from the databases [18], [20]. There were 441 people in the first image set, 146 people in the second set, and 191 people in the third set.

As mentioned above, in addition to the use of images captured by the camera on the robot, we used the images from other databases to test our system more rigorously with a variety of images resolution, different capturing angles, various people scales, and dissimilar scenarios. For evaluating this system, the same evaluation rules presented in [13] were used here. Those rules are:

- A person is considered detected if half or more of the body is inside the rectangle, otherwise it is considered as not detected.
- At high threshold values, a person might be detected by two or more small boundaries on the body. This is considered neither a correct detection nor a false one since those boundaries would combine later to form a rectangle around the person when the threshold is lowered.
- If there is a rectangle surrounding two or more people, they are all considered as detected.

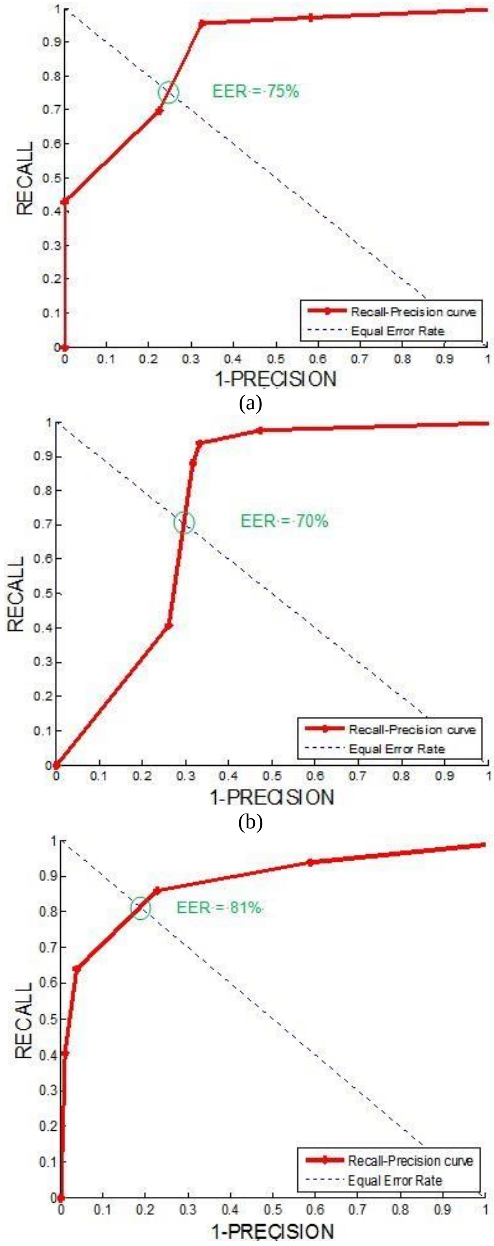


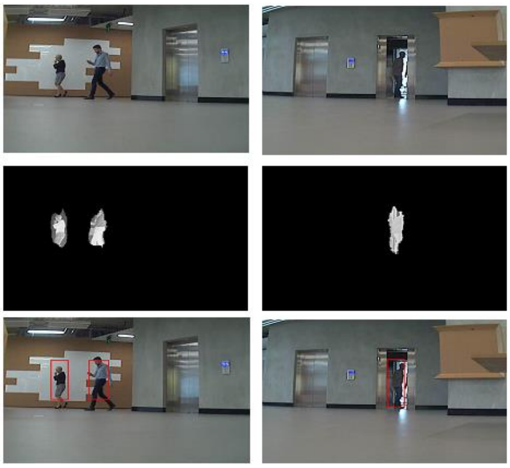
Figure 4. The recall-precision curves for the indoor image sets: (a) first set, (b) second set, and (c) third set.

V. Results and Discussion

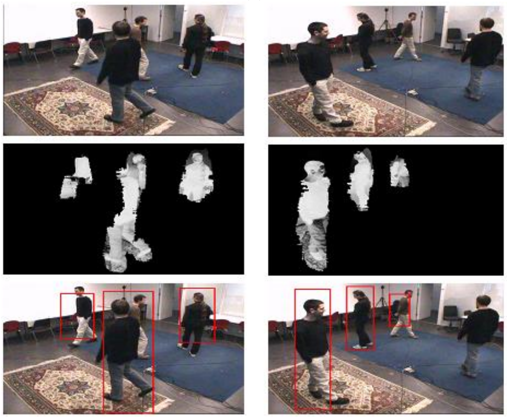
The recall-precision curves for the indoor detection are shown in Fig. 4 and for the outdoor detection in Fig. 6. The results are discussed separately in the following sections:

A. Indoor Detection Performance

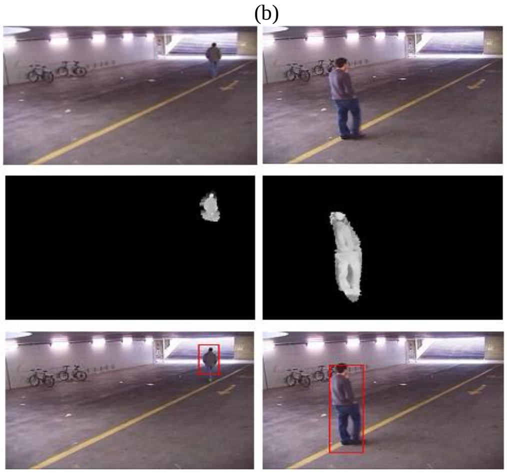
Our system showed a good performance for the first and second indoor image sets. As shown in Fig. 4a and 4b, the system achieved an Equal Error Rate (EER) of 75% for the first set and 70% for the second set. For the third image set, the system performance was higher with EER of 81% as shown in Fig. 4c despite its rather difficult conditions due to the poor lighting as explained in [19]. Fig. 5 shows samples of the system detection for the indoor image sets.



(a)



(b)

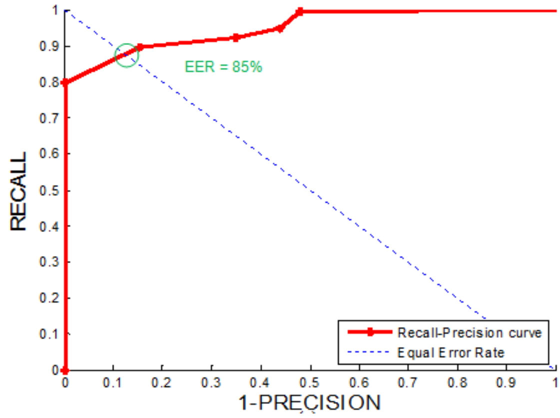


(c)

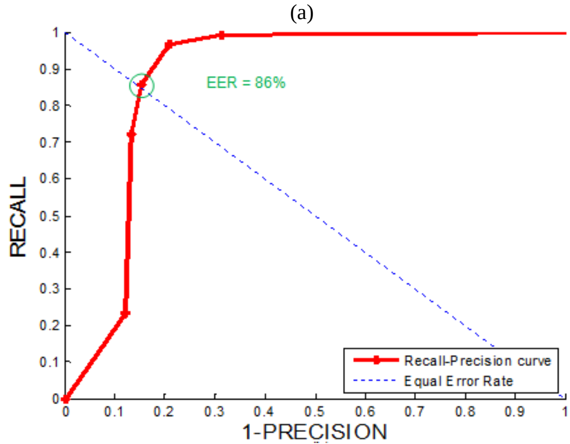
Figure 5. Samples of the indoor images used in testing the system: the images in the first set (a) are captured by the IP camera mounted on the robot, in the middle set (b) people have different scales with complicated background, and in the last set (c) images have poor lighting and low contrast. The first row in each set shows the tested image, the second row, the density locations, and the third row, the system detections.

A. Outdoor Detection Performance

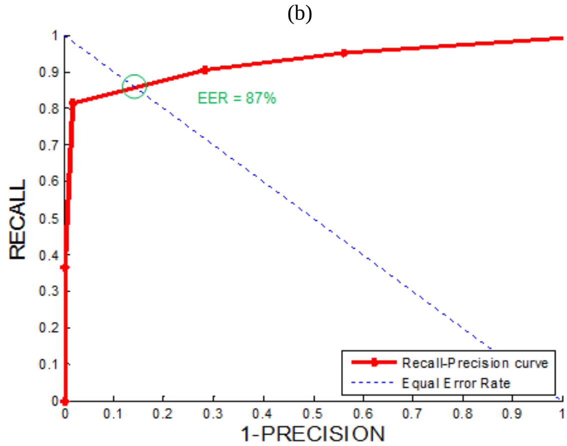
In the first image set, the system reached a very good performance with an EER of 85% as shown in Fig. 6a. For the second image set, the system performance was slightly better with an EER of 86% as shown in Fig. 6b. For the third set, the system performed marginally better with an EER of 87% as shown in Fig. 6c. Samples of the system detection for the outdoor image sets are shown in Fig. 7.



(a)



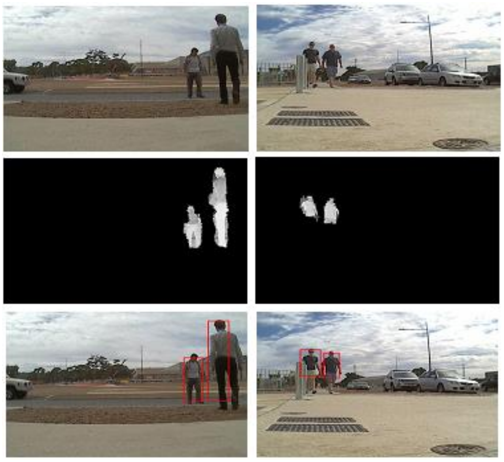
(b)



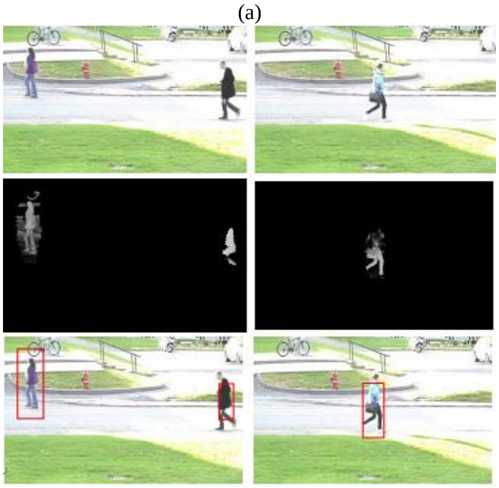
(c)

Figure 6. The recall-precision curves for outdoor image sets: (a) first set, (b) second set, and (c) third set.

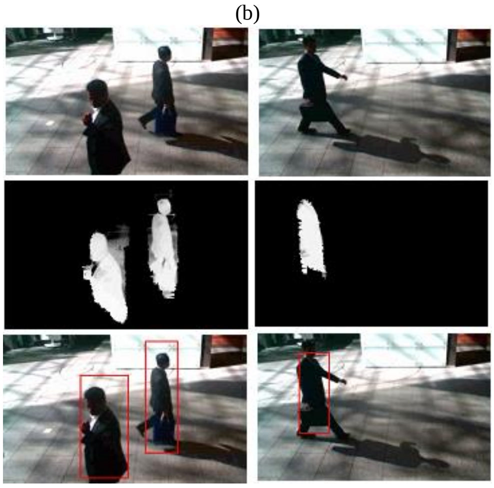
It is apparent from the results that the system performed better in outdoor images compared with the indoor ones. This is because in the outdoor environment, especially in the sun light, the environment is brighter than people and this creates a high contrast between them in the image. By inverting the image as part of the detection process, people become brighter than the environment (which is similar to the greyscale white hot thermal images), and this makes their detection easier. However, for the indoor environment, people and the environment usually have similar brightness, and this makes the people detection more challenging for the system.



(a)



(b)



(c)

Figure 7. Samples of the outdoor images used in testing the system: the images in the first set (a) are taken by the camera on the robot, in the middle set (b) people are walking in a park, and in the last set (c), people are shown with their shodows. The first row in each set shows the tested image, the second row, the density locations, and the third row, the system detections on the test image.

The noticeable feature of the codebook used in our system is that the training images were not taken from the same image source used to test the system, but were selected from unrelated sources. Another feature is the low number of training images used to create the codebook which was eight in total. It is worth mentioning that 300 training images were used in [17]. Moreover, these algorithms also required negative training images for the background which was 250 in [17], increasing the total training images to 550. Our system does not require the negative background training images and using the eight training images is adequate. As the number of training images increases, the system becomes slower. Naturally, a system with a small number of training images does not require a large processing power and will run faster. This makes it ideal for implementing the system on small mobile robot platforms with limited processing power. Finally, it is worth mentioning that this was the first time that people detection was performed on color images based on a codebook generated from thermal image samples.

VI. Conclusion and Future Work

There are many color patterns that complicate a color image and make people recognition difficult. Converting a color image to greyscale will remove most of these patterns. Moreover, inverting a greyscale image in most cases results in people showing in high contrast with respect to the environment and this simplifies their detection. We used these techniques for the pre-processing of a color image and then applied our approach used for detecting people in thermal images. This was the first time that a system used a codebook generated with samples from thermal images to work on both thermal and color images for detecting people, and it resulted in reducing the system complexity and increasing the speed and performance. We are working on the simultaneous use of color and thermal cameras on a single system to increase the detection accuracy even further.

Acknowledgment

This work was supported by the Higher Committee for Education Development in Iraq (hcediraq.org) and the Flinders University of South Australia.

References

[1] G. Cielniak and T. Duckett, "People recognition by mobile robots," *Journal of Intelligent and Fuzzy Systems*, vol. 15, no. 1, pp. 21-27, 2004.

[2] J. C. Castillo, J. Serrano-Cuerda, A. Fernández-Caballero, and M. T. López, "Segmenting humans from mobile thermal infrared imagery," in *Bioinspired Applications in Artificial and Natural Computation*, Springer Berlin Heidelberg, 2009, pp. 334-343.

[3] J. Serrano-Cuerda, M. T. Lopez, and A. Fernández-Caballero, "Robust human detection and tracking in intelligent environments by information fusion of color and infrared video," in *Proc. 7th International Conference on IEEE*, July 2011, pp. 354-357.

[4] M. Teutsch, T. Mueller, M. Huber, and J. Beyerer, "Low resolution person detection with a moving thermal infrared camera

by hot spot classification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2014, pp. 209-216.

[5] J. Davis and V. Sharma, "Robust background-subtraction for person detection in thermal imagery," in *Proc. IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.

[6] J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery," in *Proc. 17th International Conference on Pattern Recognition*, August 2004, vol. 4, pp. 713-716.

[7] E. J. Carmona, J. Martínez-Cantos, and J. Mira, "A new video segmentation method of moving objects based on blob-level knowledge," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 272-285, 2008.

[8] R. Miezianko and D. Pokrajac, "People detection in low resolution infrared videos," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[10] B. Leibe, A. Leonardis, and B. Schiele, "Combined object categorization and segmentation with an implicit shape model," in *Proc. Workshop on Statistical Learning in Computer Vision*, May 2004, vol. 2, no. 5, p. 7.

[11] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 259-289, 2008.

[12] A. Lehmann, B. Leibe, and L. Van Gool, "Fast prism: Branch and bound hough transform for object class detection," *International Journal of Computer Vision*, vol. 94, no. 2, pp. 175-197, 2011.

[13] L. A. H. Al-Shimaysawee, A. H. A. Aldabbagh, and N. Asgari, "A novel algorithm for people detection in grey scale thermal images," in *Proc. International Conference on Control, Automation and Robotics*, Singapore, May 2015, pp. 238-243.

[14] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vision Conference*, Aug. 1988, p. 50.

[15] E. Seemann, B. Leibe, K. Mikolajczyk, and B. Schiele, "An evaluation of local shape-based features for pedestrian detection," in *Proc. British Machine Vision Conference*, Oxford, 2005.

[16] K. Jungling and M. Arens, "Feature based person detection beyond the visible spectrum," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2009, pp. 30-37.

[17] A. Konigs and D. Schulz, "Evaluation of thermal imaging for people detection in outdoor scenarios," in *Proc. IEEE International Symposium on Safety, Security, and Rescue Robotics*, November 2012, pp. 1-6.

[18] Y. Wang, P. M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014, An expanded change detection benchmark dataset," in *Proc. IEEE Workshop on Change Detection (CDW-2014) at CVPR-2014*, 2014, pp. 387-394.

[19] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera people tracking with a probabilistic occupancy map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[20] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

Ali H. A. Aldabbagh was born in Najaf-Iraq on 28/07/1988. Aldabbagh awarded bachelor degree in electrical engineering from the University of Kufa, Najaf, Iraq, 2010. Currently, Aldabbagh is studying at Flinders University in Adelaide, Australia for the award of master's degree in electronic engineering.

He worked at Najaf governorate as an Adviser Engineer for nine months on electrical foundations at buildings which were under construction. Then, he worked as a Maintenance Engineer in the labs of the Electrical Engineering Department in the Faculty of Engineering, University of Kufa. He is a member of the Golden Key International Honour Society since Feb. 2015. He published the paper that has the citation number [13].

Laith A. H. Al-Shimaysawee was born in Najaf-Iraq on 22/03/1988. Al-Shimaysawee awarded bachelor degree in electrical engineering from the University of Kufa, Najaf, Iraq, 2010. Currently, Al-Shimaysawee is studying at Flinders University in Adelaide, Australia for the award of master's degree in electronic engineering.

He worked as a Maintenance Engineer in labs of the Electrical Engineering Department in the Faculty of Engineering, University of Kufa. He is a member of the Golden Key International Honour Society since Feb. 2015. He published the paper that has the citation number [13].

Nasser Asgari was born on 20/05/1959. Asgari awarded PhD degree in the areas of Computer Engineering from the University of Adelaide, Australia in 2003. He has been lecturing at Flinders university since 1998. He published the paper that has the citation number [13]. He has other publications in Robot Localization, and Interprocessor Communications using Multiport Memories.