International Journal of Machine Learning and Computing, Vol. 8, No. 3, June 2018

Image Processing Based Ambient Context-Aware People Detection and Counting

Zeyad Al-Zaydi, Branislav Vuksanovic, and Imad Habeeb

***Abstract*—Different technologies are employed to detect and count people in various situations but crowd counting system based on computer vision is one of the best choices due to a number of advantages. These include accuracy, flexibility, cost and acquiring people distribution information. Crowd counting system based on computer vision can use closed circuit television cameras (CCTV) that have already become ubiquitous and their uses are increasing. This paper aims to develop crowd counting system that can be incorporated with existing CCTV cameras. In this paper, the extracted low-level features in a frame-to-frame analysis are processed using regression technique to estimate the number of people. Two complex scenes and environments are used to evaluate the performances of the proposed system. The results have shown that the proposed system can achieve good performance in terms of the mean absolute error (MAE) and mean squared error (MSE).**

***Index Terms*—People counting, regression technique, CCTV cameras, computer vision.**

I. Introduction

The information about the distribution and number of people is important for security, safety and operational purposes. Therefore, systems with this kind of functionality are used to establish ambient awareness [1]-[5]. Business intelligence can also be developed by this kind of information, such as counting the visitors of stores, and other applications in behavioural economics [2], [6], [7]. Furthermore, there are other applications may use this kind of information such as staff planning, transport [8]-[10], and crowd management [2], [11]. Crowds can be monitored to provide an indicator of overcrowding situations to avoid potentially disastrous incidents such as the 2015 Hajj pilgrimage disaster in Saudi Arabia, where at least 2,411 people died [12]. Finally, the number and distribution of people can also be used to develop emergency evacuation procedures or improve energy efficiency by optimising heating, lighting and air conditioning [3].

Different methods are often used for people counting, such as tally counters, cameras, differential weight, sensitive carpet, infrared beams, Bluetooth, audio tones, radio-frequency identification (RFID), wireless fidelity network (Wi-Fi) and wireless sensor network (WSN) based counters [13], [14], [23]−[26], [15]-[22]. Each method has some advantages and disadvantages but people counting systems based on conventional camera are one of the best choices because CCTV cameras are widely used. For example, there are 5.9 million CCTV cameras in the UK [27]. People counting system is one of the most challenging systems in computer vision to implement [4], [21], [28]-[30]. In comparison with the conventional camera-based method, the problem with other technologies is that they need to be carefully planned and deployed for specific purposes. In addition, their cost is prohibitive for many organisations and the accuracy is often less than the conventional camera-based method. Most of these technologies are also ineffective for acquiring people distribution such as tally, differential weight, sensitive carpet, infrared beams and WSN counters which make them an inappropriate option for various types of applications. As a consequence, conventional camera-based crowd counting systems are selected in this paper due to following reasons:

A. Cost-Efficiency

They may use the CCTV cameras that have been already installed for monitoring and there is no need for new or extra hardware.

B. High Flexibility

They work efficiently with different kinds of CCTV cameras and with different kinds of camera setting e.g. low-resolution cameras, grey or colour cameras. and vertical or tilted camera angles; they also work well in both indoor and outdoor environments, including those with a complicated background. They work without impeding traffic and without modifying the environment.

C. Ability to Provide Accurate Distribution Information

They can discover both the number of people and their distribution. Many of the other crowd counting techniques are unable to acquire information on the distribution of people.

D. Wide Coverage

In monitoring, they can cover either a small or large area. Moreover, they can work with a distant camera, or one placed very high, such that the people are so small and difficult to recognise.

E. High Accuracy

They produce a very accurate estimation of the number of people. Most of them achieve more than 90% accuracy for very crowded environments.

This paper is organised in the following way. Section II

Manuscript received April 3, 2018; revised June 12, 2018.

Z. Q. H. Al-Zaydi is with the School of Engineering, University of Portsmouth, Portsmouth, UK, and is also with the Biomedical Engineering Department, University of Technology, Baghdad, Iraq (e-mail: up714763@myport.ac.uk).

B. Vuksanovic is with the School of Engineering, University of Portsmouth, Portsmouth, UK (e-mail: branislav.vuksanovic@port.ac.uk).

I. Q. Habeeb is with the School of Engineering, University of Information Technology & Communications, Baghdad, Iraq (e-mail: emadkassam@yahoo.com).

doi: 10.18178/ijmlc.2018.8.3.698

describes the related work which includes visual and non-visual based crowd counting systems; Sections III describes the system design which includes regression model, feature representation and selection, and the method of prospective normalization subsections; Section IV presents the experimental results; Finally, Section V summarizes the main conclusions.

II. Related Work

The people counting task has been studied extensively using non-visual methods. In recent years, many researchers have turned to visual technologies to count people automatically using cameras. For further development of people counting systems and to provide more accurate performance estimation, there is an increasing need for a good understanding of their key characteristics, problems and limitations. Although current research into people counting in sparse environments is well established, there are still many challenges and limitations in crowded environments.

A. Non-visual Based Crowd Counting Systems

Non-visual based crowd counting systems use different methods which may include tally counters, differential weight, sensitive carpet, infrared beams, Bluetooth, audio tones, RFID, Wi-Fi and WSN based counters [13], [14], [23]-[26], [15-[22].

Tally counters provide easy and accurate counts [31]. They are mechanical, or electronic devices that incrementally count people. They can be used to manually count the number of people walking in and out of a venue. The main disadvantages of tally counters are inflexibility and in some instances inability to detect people distribution. Furthermore, they are not suitable for detailed analysis and can be a bottleneck in crowded situations.

Differential weight counters estimate the number of people by evaluating the weight variations using load cells [32]. They may be useful for carriage environments such as trains. buses or lifts. These counters are only suitable for a few types of environments. They also assume a fixed weight for each person. That is not always reliable due to the significant difference in weight between children and adults or between fat and thin people.

Sensitive carpet counters are an accurate option but involve severe modifications of the environment and they are prone to wear [32]. They use sensitive electronic sensors to count the steps of people. They are particularly useful for indoor environments. Furthermore, people stand with two feet but walking with one or both feet which lead to error in counting.

Infrared beams counters are also used to count the number of people [13], [14]. One or more horizontal infrared beams are usually used across an entrance. If the beam is broken, the counter counts a 'tick'. Multiple beams are used by many researchers to find the direction of people or to improve the accuracy. Infrared counters are still widely used due to their low cost and simplicity of installation. On the other hand, simple infrared beams counters are non-directional and their main disadvantages are that they cannot discern people walking side-by-side and they can be blocked by people standing in front of the beam. In addition, they are not suitable to work in open areas where no particular entrances and exits exist.

Bluetooth, audio tones, RFID and Wi-Fi are used as device-based methods to count or localise people [26], [33]-[35]. Device-based methods require people to carry mobile devices. They also require people to enable the Bluetooth units, use speakers or to use extra hardware such as RFID tags. The main disadvantage of this technique is that some people carry more than one mobile device and not everyone carries a device which affects the accuracy significantly.

WSN and Wi-Fi are also used as device-free techniques to count the number of people [15], [17], [26], [36], [37]. Device-free methods do not require people to carry certain devices to be counted. They usually depend on the variation of the wireless signal to find the relationship between it and the number of people. These techniques are easily affected by environmental dynamics, noise, fading and other factors that may influence signal. In addition, their application is mainly limited to indoor environments.

B. Visual Based Crowd Counting Systems

Different kinds of cameras can be used for people counting. Visual based crowd counting systems can be classified into four categories; 3D, smart, thermal and conventional camera counters.

The 3D camera counter is a technology used in people counting which can help to identify the depth information of the people [22], [38]. The release of Microsoft Kinect [39] in 2010 increased the interest in the field of 3D camera counter because Microsoft Kinect provides good quality depth images at a lower price compared to previous technologies. However, the information from Microsoft Kinect can still contain a lot of noise [40]. In addition, the practical sensors range of Microsoft Kinect is 3.5 meter which makes it useless to count people in the large areas [41]. It also cannot sense objects that are illuminated by direct sunlight so it does not work in outdoor environments [22]. Image depth can also be obtained using time-of-flight (TOF), light detection and ranging (LIDARs) and stereo cameras methods [20]. However, stereo cameras are affected by changing illumination and cannot operate in the dark. Another problem emerges when monitoring a large area with a similar colour and little edges, because it may be difficult to find features [42]. In addition, developing a stereo based depth sensing system is more complex and would, therefore, require a significant amount of knowledge and computational power [22]. On the other hand, the luminance sensitivity of TOF cameras is poor and their depth range is limited [20] and the size of LIDARs cameras are large [22]. In addition, they are expensive and their accuracies are lower than Microsoft Kinect [22], [41].

Smart cameras (intelligent cameras) can also be used for people counting. They refer to cameras that have processing capabilities so there is no need to an external processing unit such as computers [43]. The main disadvantage of these cameras is the cost because they are expensive. In addition, this is not a very convenient option because most of the current surveillance systems use conventional cameras. To use this option, the current CCTV cameras would have to be replaced which make this option not very practical due to the scalability.

Thermal cameras are also used to count the number of people [19], [22]. They are usually positioned at an entrance or a gate and they detect people's body heat. Accuracy can be affected if the ambient temperature within the counting area is above a certain threshold. Heat sources and external weather conditions may affect the accuracy of detecting the emitted heat from people. In addition, they have narrow fields and may not cover wide spaces. Thermal counters have the advantage that they are not affected by changing illumination and do not need background subtraction algorithm, therefore have a shorter processing time [22].

For counters based on conventional camera, different algorithms have been introduced to increase the accuracy of counting [16], [18], [21], [23]-[25]. Most of them are proposed to work in both indoor and outdoor environments whereas some algorithms are proposed to only work in indoor environments [44], [45]. Conventional camera based people counters can be classified depending on the area of view into the region of interest (ROI) and line of interest (LOI) [23]. In the ROI methods, people in a specific region are counted [24] whereas, in the LOI methods, people who cross a real or virtual line are counted [16]. These counters can also be classified into four categories: features regression-based, features trajectories based, people detection based and pixel-wise based algorithms [27]. Features regression-based algorithms start by extracting useful features and then use them to count people [21]. Features trajectories based algorithms involve tracking people and counting them [46]. People detection based algorithms detect people and then counting them [47]. In pixel-wise optimisation based algorithms, the density of each pixel in a frame is determined and then integrated [48].

III. System Design

This section presented the proposed system starting with the regression model. Secondly, the feature representation and selection approach adopted for the proposed system. Thirdly, the method of prospective normalization is described. The flow diagram of the proposed system is given in Fig. 1.

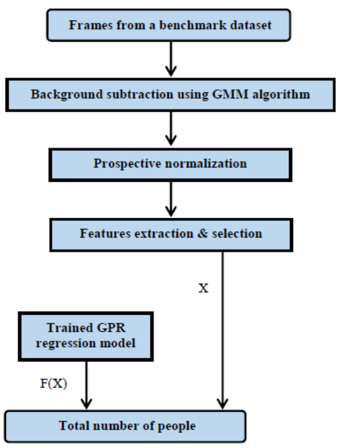


Fig. 1. Flow diagram of the proposed system.

A. Regression Model

In order to train the crowd counting system, a regression function has to be learned using a set of training samples to find the relationship between the extracted features and the number of people. Gaussian process regression (GPR) has been selected in this system. GPR does not use any prior assumptions about the relationship between the features and the crowd size and can achieve high accuracy so it has been chosen in the proposed system [4], [49]–[51]. Mathematically, the estimated number of people in GPR is given by [52];

$$y_*|y \sim N\left(K_* K^{-1} y, K_{**} - K_* K^{-1} K_*^{T}\right) \quad (1)$$

The best estimate is represented by y_* [49]:

$$y_* = K_* K^{-1} y \quad (2)$$

The uncertainty of the estimated number of people is given by [49]:

$$var(y_*) = K_{**} - K_* K^{-1} K_*^{T} \quad (3)$$

where $x1, x2, x3, \dots \dots \dots xn$ are the training sets and x_* is the test set. $k(x, x')$ is the kernel. A combination of linear and RBF kernels is used with the GPR model. Mathematically, this combination of kernels is given by [50], [52]:

$$K = \begin{bmatrix} k(x1,x1) & \cdots & k(x1,xn) \\ \vdots & \ddots & \vdots \\ k(xn,x1) & \cdots & k(xn,xn) \end{bmatrix} \quad (4)$$

$$K_* = k(x_*, x1) \quad \dots \quad k(x_*, xn) \quad (5)$$

$$K_{**} = (x_*, x_*) \quad (6)$$

where $x1, x2, x3, \dots \dots \dots xn$ are the training sets and x_* is the test set. $k(x, x')$ is the kernel. A combination of linear and RBF kernels is used with the GPR model. Mathematically, this combination of kernels is given by [50], [52]:

$$k(x, x') = \alpha_1 \left(x^T x' + 1\right) + \alpha_2{}^2 \exp\left[\frac{-1}{2\alpha_3{}^2} (x - x')^2\right] \quad (7)$$

where α_1, α_2 and α_3 are the kernals parametes.

TABLE I: List of the Extracted Features

Features	Description
Foreground segment	segment area segment perimeter perimeter orientation histogram (90 degrees) perimeter orientation histogram (120 degrees) perimeter orientation histogram (150 degrees) perimeter orientation histogram (0 degrees) perimeter orientation histogram (30 degrees) perimeter orientation histogram (60 degrees) perimeter-area ratio Blob count
Edge	internal edge length internal edge orientation histogram (90 degrees) internal edge orientation histogram (120 degrees) internal edge orientation histogram (150 degrees) internal edge orientation histogram (0 degrees) internal edge orientation histogram (30 degrees) internal edge orientation histogram (60 degrees)
Texture	GLCM energy (0 degrees) GLCM homogeneity (0 degrees) GLCM entropy (0 degrees) GLCM energy (45 degrees) GLCM homogeneity (45 degrees) GLCM entropy (45 degrees) GLCM energy (90 degrees) GLCM homogeneity (90 degrees) GLCM entropy (90 degrees) GLCM energy (135 degrees) GLCM homogeneity (135 degrees) GLCM entropy (135 degrees)
Keypoints	SIFT

B. Feature Representation and Selection

The low-level features are used with the proposed system to describe the visual properties in the frames as texture, edge, shape, size and colour [5], [53]. Blob size histogram and edge orientation histogram are also used with the proposed system as intermediate features between low and high-level features.

A combination of several features is used with this system to achieve high accuracy. The performances of crowd counting systems significantly depend on selected combinations of features [53]. Table I shows the list of features that used with the proposed system. The features used in this paper can be categorised into four categories:

1) Foreground segment features

These features are usually extracted after using a background subtraction algorithm. The segment properties frames are captured by these features. There are two types of foreground segment features; size and shape [53]. Size features refer area, perimeter, perimeter-area ratio and blob count of the foreground pixels. Shape features include the perimeter orientation histogram [5].

2) Texture features

They are the general description of a frame and have a strong relationship with the crowd size [5]. local binary pattern (LBP) and Gray-level co-occurrence matrix (GLCM) are usually used to find texture features [5], [54], [55]. The texture features that used in this work include texture randomness (entropy), total sum-squared energy (energy) and texture smoothness (homogeneity), [5], [50], [53].

The GLCM starts by creating a co-occurrence matrix $(P(i,j|\theta))$ through quantizing each frame into eight grey levels. θ represents the orientations of the co-occurrence matrix. The symmetric co-occurrence matrix $(P_s(i,j|\theta))$ is given by [53];

$$P_s(i,j|\theta) = P(i,j|\theta) + P(i,j|\theta)^T \tag{8}$$

The normalized co-occurrence matrix is then calculated by;

$$P_n(i,j|\theta) = \frac{P_s(i,j|\theta)}{\sum_{i,j} P_s(i,j|\theta)} \tag{9}$$

where $P_n(i,j|\theta)$ is the normalised co-occurrence matrix. The energy, homogeneity and entropy are given by;

$$\text{Energy}_\theta = \sum_{i,j} P_n(i,j|\theta)^2 \tag{10}$$

$$\text{Homogeneity}_\theta = \frac{P_n(i,j|\theta)}{1+|i-j|} \tag{11}$$

$$\text{Entropy}_\theta = \sum_{i,j} -P_n(i,j|\theta)\log P_n(i,j|\theta) \tag{12}$$

3) Edge features

The relative change in pixel intensities across a frame represents the edge pixels [1]. There is a strong dependency between the complexity of crowds and the number of people. Coarse edges are extracted from Low crowd size while complex edges are extracted from high crowd size [5]. The edge features that used in this paper include edge orientation histogram and total edge pixels [56]. Edge orientation histograms are used to distinguish people with other structures in the frame [4].

4) Keypoints

They are specific pixels of interest in a frame [53]. They have been used in many crowd counting studies [57]-[60]. They usually include BRISK, FAST, SIFT, SURF and Harris corner points [28], [53], [61]. SIFT points have been used in this work because they are invariant to image scale, translations and rotations [62]. In addition, they are robust with perspective transformations and moderate illumination variations [63].

C. Prospective Normalization

The extracted features of the same person who exist at different distances to the camera have significantly different values due to the change of his/her size. Different weights have been used for the pixels to solve this problem. Fig. 2 shows the change of size of people at different positions. The weight of the pixels at any line is given by [5]:

$$\text{weight}_{\text{line}} = \frac{h_{ab}\, w_{ab}}{h_{line}\, w_{line}} \tag{10}$$

where the pixels at (ab) line are assigned weight equal to one. w_{line} and w_{ab} are the width of the rectangle at the line of interest and line (ab), respectively. h_{line} and h_{ab} are the heights of the same person on the line of interest and line (ab), respectively.

Fig. 2. The change of size of people at different distances to the.

IV. Experimental Results

The University of California (UCSD) and the Peds2 datasets have been used to evaluate the proposed system [51], [64]. UCSD dataset has been widely used for testing and validating people counting methods [65]. It represents an oblique view of a walkway. The peds2 dataset is newer and containing a fewer number of people than the UCSD dataset. It represents a side view of a walkway. Each dataset is split into a training set and testing set. 800 training frames and 1200 testing frames are used in the UCSD dataset. For the Peds2 dataset, 1200 frames are used for training and 2800 frames for testing. Table II shows the features of each dataset.

Two metrics are used to test and evaluate the proposed system; mean absolute error (MAE) and mean squared error (MSE) [5]. The mean absolute error is given as:

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|y_n - \hat{y}_n| \tag{11}$$

The mean squared error is given as:

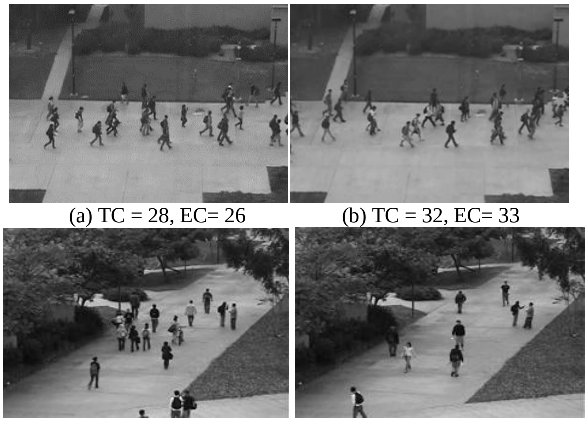
$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 \tag{12}$$

where N is the total number of the test frames, y_n is the actual count, and $\hat{y}_n$ is the estimated count.

The MAE and MSE for the proposed system are 1.945 and 6.056, respectively, for the UCSD dataset and 0.746 and 1.161, respectively, for the Peds2 dataset. The results have shown that the proposed system achieves good results in heavily occluded environments with perspective distortions. Fig. 3 shows two frames from the UCSD dataset and two frames from the Peds2 dataset. The true number of people is represented by TC and the estimated counting is represented by EC. The error of the proposed system meets the requirements of system operators and is significantly less than the acceptable error (0.2) [53].

TABLE II: THE FEATURES OF THE BENCHMARK DATASETS [49], [64]

	Peds2 dataset	UCSD dataset
Year	2010	2008
Length (frames)	4000	2000
Frame rate (fps)	10	10
Resolution	238×158	238×158
Colour	Grey	Grey
Location	Outdoor	Outdoor
Shadows	No	No
Reflections	No	No
Loitering	No	No
Frame type	.tiff	.png



(a) TC = 28, EC= 26 (b) TC = 32, EC= 33

(c) TC = 17, EC= 20 (d) TC = 9, EC= 10

Fig. 3. Examples of the Peds2 and UCSD datasets frames and their results.

V. CONCLUSIONS

The objective of described work was to develop a people counting system that can be incorporated with existing CCTV cameras, which are already widely used, to provide the information about the number of people in a given space. This paper describes the system and some results achieved using the UCSD and Peds2 datasets. Experimental results on those datasets demonstrated that the proposed system can achieve high accuracy even under heavy occlusions and perspective distortions.

The MAE and MSE of the proposed system are a very good and meet the requirements of system operators. Furthermore, the number of people at specific locations in a scene can also be estimated using the proposed system. This shows significant promise because this can be used to detect localized abnormalities in different applications such as evacuation planning, product displays and crowd control.

REFERENCES

[1] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Scene invariant multi camera crowd counting," *Pattern Recognit. Lett.*, vol. 44, pp. 98–112, 2014.

[2] A. Technology. (2013). *Our Customers.* [Online]. Available: http://www.peoplecounting.co.uk/our-customers

[3] M. Wang, "Data Assimilation for Agent-Based Simulation of Smart Environment," Dissertation, Georgia State University, 2014.

[4] D. A. Ryan, "Crowd monitoring using computer vision," Dissertation, Queensland University of Technology, 2013.

[5] C. Loy, K. Chen, S. Gong, and T. Xiang, "Crowd counting and profiling: Methodology and evaluation," *Modeling, Simulation and Visual Analysis of Crowds*, pp. 347–382, 2013.

[6] ShopperTrak. (2013). *ShopperTrak Solutions.* [Online]. Available: http://www.shoppertrak.com/

[7] Biodata Ltd. (2013). *Use CCTV to Count People.* [Online]. Available: http://www.videoturnstile.com/

[8] DILAX Intelcom. (2015). *Public Transport.* [Online]. Available: https://www.dilax.com/

[9] J. S. Lumentut, F. E. Gunawan, and Diana, "Evaluation of recursive background subtraction algorithms for real-time passenger counting at bus rapid transit system," *Procedia Comput. Sci.*, vol. 59, pp. 445-453, 2015.

[10] G. Garcia-Bunster, M. Torres-Torriti, and C. Oberli, "Crowded pedestrian counting at bus stops from perspective transformations of foreground areas," *IET Comput. Vis.*, vol. 6, no. 4, p. 296, 2012.

[11] S. Longo and B. Cheng, "Privacy preserving crowd estimation for safer cities," in *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 1543–1550.

[12] J. Gambrell and B. Ahmed, "Hajj Stampede in September killed over 2,400, new count finds," *The New York Times*, 2015.

[13] F. Z. F. Zhu, X. Y. X. Yang, J. G. J. Gu, and R. Y. R. Yang, "A new method for people-counting based on support vector machine," *Int. Conf. Intell. Networks Intell. Syst.*, vol. 1, no. 5, pp. 5–8, 2009.

[14] N.-N. Li, J. Song, R.-Y. Zhou, and J.-H. Gu, "A people-counting system based on BP neural network," in *Proc. International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2007, vol. 3.

[15] M. Nakatsuka, H. Iwatani, and J. Katto, "A Study on passive crowd density estimation using wireless sensors," in *Proc. International Conference on Mobile Computing and Ubiquitous Networking*, 2008, no. 2, pp. 1–6.

[16] Z. Ma and A. B. Chan, "Crossing the line: Crowd counting by integer programming with local features," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2539–2546.

[17] Y. Yuan, C. Qiu, W. Xi, and J. Zhao, "Crowd density estimation using wireless sensor networks," in *Proc. Seventh International Conference on Mobile Ad-hoc and Sensor Networks*, 2011, pp. 138–145.

[18] R. Shbib, S. Zhou, D. Ndzi, and K. Al-kadhimi, "Distributed monitoring system based on weighted data fusing model," *Am. J. Soc. Issues Humanit.*, vol. 3, pp. 53–62, 2013.

[19] W. C. Lin, W. Seah, and W. Li, "Exploiting radio irregularity in the Internet of Things for automated people counting," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2011, pp. 1015–1019.

[20] V. Gandhi, J. Čech, and R. Horaud, "High-resolution depth maps based on TOF-stereo fusion," in *Proc. IEEE International Conference on Robotics and Automation*, 2012, pp. 4742–4749.

[21] Y. L. Hou and G. K. H. Pang, "People counting and human detection in a challenging situation," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 41, no. 1, pp. 24–33, 2011.

[22] T. Tikkanen, "People detection and tracking using a network of low-cost depth cameras," Aalto University, 2014.

[23] J. Li, L. Huang, and C. Liu, "Robust people counting in video surveillance: Dataset and system," in *Proc. International Conference on Advanced Video and Signal Based Surveillance*, 2011, pp. 54–59.

[24] J. Tu, C. Zhang, and P. Hao, "Robust real-time attention-based head-shoulder detection for video surveillance," in *Proc. IEEE International Conference on Image Processing*, 2013, pp. 3340–3344.

[25] A. Adegboye, G. Hancke, and G. H. Jr, "Single-pixel approach for fast people counting and direction estimation," in *Proc. Southern Africa Telecommunication Networks and Applications*, 2012.

[26] W. Xi *et al.*, "Electronic frog eye: Counting crowd using WiFi," in *Proc. - IEEE INFOCOM*, pp. 361–369, 2014.

[27] Z. Q. H. Al-Zaydi, D. L. Ndzi, M. L. Kamarudin, A. Zakaria, and A. Y. M. Shakaff, "A robust multimedia surveillance system for people counting," *Multimed. Tools Appl.*, 2016.

[28] S. A. M. Saleh, S. A. Suandi, and H. Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 103–114, 2015.

[29] J. Wang, W. Fu, J. Liu, H. Lu, and S. Member, "Spatiotemporal Group Context for Pedestrian Counting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1620–1630, 2014.

[30] X. Hu, H. Zheng, Y. Chen, and L. Chen, "Dense crowd counting based on perspective weight model using a fisheye camera," *Int. J. Light Electron Opt.*, vol. 126, no. 1, pp. 123–130, 2015.

[31] L. Lev, L. J. Brewer, and G. O. Stephenson, "Tools for Rapid Market Assessments," 2008.

[32] J. Vasco Dantas dos Reis, "Image descriptors for counting people with uncalibrated cameras," University of Porto, 2014.
[33] P. G. Kannan, S. P. Venkatagiri, M. C. Chan, A. L. Ananda, and L.-S. Peh, "Low cost crowd counting using audio tones," in *Proc. the 10th ACM Conference on Embedded Network Sensor Systems*, 2012, p. 155.
[34] N. Lionel, L. Yunhao, C. L. Yiu, and P. Abhishek, "LANDMARC: Indoor location sensing using active RFID," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, pp. 407–415, 2003.
[35] J. Weppner and P. Lukowicz, "Collaborative crowd density estimation with mobile phones," in *Proc. ACM Phone Sense*, 2011.
[36] S. Di Domenico, G. Pecoraro, E. Cianca, and M. De Sanctis, "Trained-once device-free crowd counting and occupancy estimation using WiFi : A doppler spectrum based approach," in *Proc. International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016.
[37] Y. Yuan, J. Zhao, C. Qiu, and W. Xi, "Estimating crowd density in an RF-based dynamic environment," *IEEE Sens. J.*, vol. 13, no. 10, pp. 3837–3845, 2013.
[38] L. Del Pizzo, P. Foggia, A. Greco, G. Percannella, and M. Vento, "Counting people by RGB or depth overhead cameras," *Pattern Recognit. Lett.*, vol. 81, pp. 41–50, 2015.
[39] Microsoft, "Kinect," 2011. [Online]. Available: www.xbox.com/en-US/kinect/. [Accessed: 01-Aug-2016].
[40] X. Zhang, J. Yan, S. Feng, Z. Lei, D. Yi, and S. Z. Li, "Water filling: Unsupervised people counting via vertical kinect sensor," in *Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2012, pp. 215–220.
[41] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, 2013.
[42] C. R. Sensors, R. B. Fisher, and K. Konolige, *Handbook of Robotics Chapter 22 - Range Sensors*, 2008.
[43] M. Valera and S. A. Velastion, "Intelligent distributed surveillance systems: a review," in *Proc. Vision, Image and Signal Processing*, 2005, vol. 152, pp. 192–204.
[44] J. Luo, J. Wang, H. Xu, and H. Lu, "Real-time people counting for indoor scenes," *Signal Processing*, vol. 124, pp. 27–35, 2016.
[45] H. H. Cetinkaya and M. Akcay, "People counting at campuses," *Procedia - Soc. Behav. Sci.*, vol. 182, pp. 732–736, 2015.
[46] S. Yoshinaga, A. Shimada, and R. I. Taniguchi, "Real-time people counting using blob descriptor," *Procedia - Soc. Behav. Sci.*, vol. 2, no. 1, pp. 143–152, 2010.
[47] Z. Al-zaydi, D. Ndzi, and D. Sanders, "Cascade Method for Image Processing Based People Detection and Counting," in *Proc. International Conference on Image Processing, Production and Computer Science*, 2016, pp. 30–36.
[48] V. Lempitsky and A. Zisserman, "Learning to count objects in images," *Advances in Neural Information Processing Systems*, 2010, pp. 1324–1332.
[49] Z. Q. H. Al-Zaydi, D. L. Ndzi, Y. Yang, and M. L. Kamarudin, "An adaptive people counting system with dynamic features selection and occlusion handling," *J. Vis. Commun. Image Represent.*, vol. 39, pp. 218–225, 2016.
[50] A. B. Chan and N. Vasconcelos, "Counting people with low-level features and bayesian regression," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2160–2177, 2012.
[51] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
[52] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regressionn: A Quick Introduction," 2008.
[53] D. Ryan, S. Denman, S. Sridharan, and C. Fookes, "An evaluation of crowd counting methods, features and regression models," *Comput. Vis. Image Underst.*, vol. 130, pp. 1–17, 2015.
[54] S. Ershad, "Texture classification approach based on combination of edge & co-occurrence and local binary pattern," *arXiv Prepr.*, pp. 626–629, 2012.
[55] B. Wang, H. Bao, S. Yang, and H. Lou, "Crowd density estimation based on texture feature extraction," *J. Multimed.*, vol. 8, no. 4, pp. 331–337, 2013.
[56] A. N. Marana, L. D. F. Costa, R. A. Lotufo, and S. A. Velastin, "Estimating crowd density with Minkowski fractal dimension," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, vol. 6, pp. 3521–3524.
[57] A. Albiol, M. J. Silla, J. M. Mossi, and A. Albiol, "Video analysis using corner motion statistics," in *Proc. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2009, pp. 31–38.
[58] A. Albiol, A. Albiol, and J. Silla, "Statistical video analysis for crowds counting," in *Proc. International Conference on Image Processing (ICIP)*, 2009, pp. 2569–2572.
[59] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, "Counting moving people in videos by salient points detection," in *Proc. International Conference on Pattern Recognition*, 2010, pp. 1743–1746.
[60] S. W. Jeong, C. Y., Choi, S., & Han, "A method for counting moving and stationary people by interest point classification," in *Proc. IEEE International Conference on Image Processing*, 2013, pp. 4545–4548.
[61] C. Y. Jeong and S. Choi, "A Comparison of Keypoint Detectors in the Context of Pedestrian Counting," in *Proc. IEEE Information and Communication Technology Convergence (ICTC)*, 2016, pp. 1179–1181.
[62] D. Giveki, M. A. Soltanshahi, and G. A. Montazer, "A new image feature descriptor for content based image retrieval using scale invariant feature transform and local derivative pattern," *Opt. - Int. J. Light Electron Opt.*, vol. 131, pp. 242–254, 2017.
[63] S. H. Zhong, Y. Liu, and Q. C. Chen, "Visual orientation inhomogeneity based scale-invariant feature transform," *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5658–5667, 2015.
[64] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1975–1981, 2010.
[65] C. Zhang, H. Li, and X. Wang, "Cross-scene Crowd Counting via Deep Convolutional Neural Networks," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Zeyad Qasim Habeeb Al-Zaydi received his B.S. in computer engineering, from University of Technology, Iraq in 2003. He received his M.S. in computer engineering from the University of Baghdad, Iraq in 2012. He received his PhD from the School of Engineering, University of Portsmouth, UK in 2018. His research interests are in the areas of image processing, computer vision, image representation, pattern recognition, artificial intelligence, people detection and crowd counting.

He is a lecturer at the Biomedical Engineering Department, University of Technology, Iraq.

Branislav Vuksanovic graduated from the University of Belgrade, Serbia with degree in electrical and power engineering. He holds MSc degree in measurement and instrumentation from South Bank University, London and a PhD in active noise control from the University of Huddersfield, UK.

Previously, he worked as a project engineer for Croatian Electricity Board in Osijek, Croatia. During his academic career he worked as a research fellow at Sheffield and Birmingham Universities on Optical Brain Imaging and Medical Video Compression projects. He also worked as a lecturer at the University of Derby where he was a member of Sensors and Controls Research Group. Currently he works as a senior lecturer at the University of Portsmouth, School of Engineering. He has published papers in the field of active noise control, biomedical signal processing and pattern recognition for intrusion detection and knowledge based authentication. He published one book in Digital Electronics and Microcontrollers field.

His current research interests are in the application of pattern recognition techniques for power systems and analysis of ground penetrating radar and ECG data.

Imad Qasim Habeeb is a lecturer in the Engineering College, University of Information Technology and Communications, Iraq. He received his B.S. in computer and control engineering, from University of Technology, Iraq. He recieved his PhD and MSc degree in information technology from the Universiti Utara Malaysia, Malaysia. His research interests are in optical character recognition, automatic speech recognition, image processing, computer vision, natural language processing, and strings alignment.