A Recent Trend in Individual Counting Approach Using Deep Network

Anahita Ghazvini*, Siti Norul Huda Sheikh Abdullah, Masri Ayob

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM) 43600, Bangi Selangor (Malaysia)

Received 18 February 2019 | Accepted 12 April 2019 | Published 17 April 2019

Abstract

In video surveillance scheme, counting individuals is regarded as a crucial task. Of all the individual counting techniques in existence, the regression technique can offer enhanced performance under overcrowded area. However, this technique is unable to specify the details of counting individual such that it fails in locating the individual. On contrary, the density map approach is very effective to overcome the counting problems in various situations such as heavy overlapping and low resolution. Nevertheless, this approach may break down in cases when only the heads of individuals appear in video scenes, and it is also restricted to the feature's types. The popular technique to obtain the pertinent information automatically is Convolutional Neural Network (CNN). However, the CNN based counting scheme is unable to sufficiently tackle three difficulties, namely, distributions of non-uniform density, changes of scale and variation of drastic scale. In this study, we cater a review on current counting techniques which are in correlation with deep net in different applications of crowded scene. The goal of this work is to specify the effectiveness of CNN applied on popular individuals counting approaches for attaining higher precision results.

Keywords

Analysis of Individuals, Automatic Video Surveillance, Counting Individuals, CNN, Deep Learning.

DOI: 10.9781/ijimai.2019.04.003

I. Introduction

Individual counting scheme has an extensive application in various domains such as public stations, shopping mall, universities, etc. [1]. The analysis of crowd has been presented surprising assessment to get knowledge about the crowded scene, though comprising the behavior analysis of crowd [2][3], tracking individuals [4][5], and segmenting the crowd [6][7].

The region of interest (ROI) and line of interest (LOI) are considered as two broad categories which are employed for individuals counting in the video. The ROI approximates the total value of individuals in some areas at a specific time whereas LOI obtains the total value of each individual in a certain period of time when they cross a detecting line [8]. To deal with these two categories, researchers employed either human detection, feature regression, or clustering techniques respectively [8]. Numerous schemes are proposed to count the individual on the basis of ROI and LOI categories. The general techniques are clustering [9], detection [10], and regression [11][12].

The crowd analysis has contributed profusely towards understanding the scene with overpopulation via linking the behavior analysis [2] [3], tracking [4][5], and segmenting the crowd [13][6]. Based on the literature [14][15], the different crowd schemes had similar principles which enable them to precisely describe by the attributes. Currently [16], multiple studies struggle on profiling attributes of the crowd, due to the limitation to the value of the attribute and the indifference to the dissimilarity in the scene.

* Corresponding author.

E-mail address: p86698@siswa.ukm.edu.my

The research of [17], suggested a deep multitask method to establish a good comprehension towards crowd by attaining information, mix procedure and features movement. The consequence of this letter demonstrates significant advances on the evaluation of recognizing the attribute in cross-scene by the suggested deep scheme.

The letter of [18] compared the most well-known counting approaches such as detection, clustering, and the latest technique, regression, which is gaining traction in handling the overpopulated area. The results of [18] showed that the regression scheme presented a higher performance in contrast to other approaches. Among the well-known counting schemes, the performance of detection and clustering based approaches are significantly marred on overpopulated scenes. In contrast to these approaches of the regression techniques, [19] is applicable to the overpopulated area and overlapping among the objects in crowded scenes. On the other hand, of all the regression techniques, the partial least square regression (PLSR) can resolve the collinearity difficulty with fewer factors, also requires fewer computations and converges speedily compared to other methods [20]. The major restriction of PLSR is at higher risk of overlooking 'real' correlations and sensitivity to the relative scaling of the descriptor variables [21]. Still, these approaches fall short of localization task due to their ineffectiveness to specify the location data of each individual in the scene.

The density maps technique exhibits great success in tackling the issue of individual counting in comparison to regression-based methods, specifically where the scene consists of a few complications such as heavy inter-occlusion between individuals, unclear resolution in videos [22]-[25][7]. Additionally, these methods can specify spatial

data about the crowd.

Density estimation-based approach is normally utilized in public places like malls, the squares, and stations. This approach is highly effective in establishing a better understanding of individuals' behavior, which can elevate the individuals' safety in a public area. As the crowds may vary in distributions and shape patterns, the problem of recognizing pattern arises [26]. The crowd density is attested to be at greater advantage in comparison to crowd counting methods as it provides location data of the crowd. Counting individuals can be predicted with greater ease by utilizing a density map for some specific region. The total number of individuals in the video frame/image was achieved simply by determining the integral of density function over that frame/image [27].

Deep learning has shown a great performance in various area like speech and pattern recognition while among these techniques of deep learning, the CNN fares the best in the task of the image classification. The study of [28] proposed a novel approach based on deep techniques to classify a face. This work boosts the network performance through three various techniques of Deep belief Net (DBN), Stacked Auto-Encoder (SAE) besides of Back Propagation Neural Networks (BPNN) while utilizing sigmoid objective function. The high-level features are extracted via a deep algorithm which were utilized as inputs of the classifier to distinguish among the faces and non-faces. The outcome of this letter [28] attested to the robustness and proficiency of proposed technique for facial classification task on two datasets of BOSS and MIT.

The study of [29] suggested the estimation of density methods to assess the density maps on dissimilar crowd analysis tasks. Since the proficiency of density methods is firmly based on the features forms [22], this paper estimated a classical CNN to produce the original resolution of density map in contrast to current CNN approaches, which caused the reduction of density maps resolution through down-sampling strides in convolution processes. The obtained result from this work proved that the performance of counting remained unaffected on the unclear light of density maps, it, however, did not work for the localization tasks, like detection and tracking.

By producing the information of high-density, the utmost of research concentrated on an approximation of a density map to obtain the total number of the individuals [30]. These approaches were totally depending on the forms of features while CNN was utilized to obtain the valuable info from input automatically and also very efficiently to deal with overlapping among objects, non-uniform lighting, and altering scales [30]. However, the current CNN approaches may err in resolving the two issues of non-stable density distributions and dissimilarity in scale. The letter of [30] suggested the multi-column multi-task convolutional neural network (MMCNN) to solve these problems. This paper suggested three innovative methods, namely to offer a new density map which could focus on location and full data, propose a multi-column CNN to get beneficial data from different scales and lastly, to estimate the density map, level of crowded environment, as well as background or foreground mask. The accuracy of the density map improved by additional tied objectives. The suggested scheme of this work displayed to be more proficient in contrast to conventional Gaussian density map and the recently utilized shaped of individual density map. The method of this work proved that current CNN-based individual counting methods were capable in dealing with the issue of non-uniform distributions and scale variations. The study of [13], proposed a framework to disentangle the individual counting difficulties in cross-scene by using CNN with no additional annotations for a new target scene. The CNN model of this letter was trained based on two objectives of the density map and individual counts where these objectives assisted each other to gain higher local optima. Also, this model could learn certain crowd features more efficiently compared to the handcraft features. To tackle the gap among various scenes, the pre-trained CNN was adjusted to all scenes. The texture of the crowd would be taken through the CNN technique which yielded an accurate counting outcome without considering the foreground segmentation results, as the method of this study was solely based on appearance information. The experiment result of this study showed the efficiency and consistency of the proposed method. As a consequence, in comparison to different counting approaches, density map methods were more proper for an overpopulated area, overlap scenes. Moreover, it seemed very potential in specifying spatial data to locate each individual in the overcrowded scene. This technique is accurate in defining the crowds through density distribution, as it provides the data on the location.

Nevertheless, it may collapse where only heads of individual are obvious. As the performance of this method is strictly dependent on the features types, the CNN can provide useful information automatically in contrast to shallow techniques. Moreover, the CNN techniques are highly beneficial to deal with difficulties that involve overlapping among objects, non-uniform lighting, and altering scales. Although the MMCNN is able to resolve these difficulties, they are restricted to the scales which are applied in overtraining and their ability to understand well-generalized schemes. In this letter, we aim to provide a review on recent individuals counting techniques which use deep networks in various applications of the crowded scene, and also to specify the effectiveness of deep CNN for attaining higher precision while utilizing popular individuals counting approaches. This paper is segregated into five parts, specifically, introduction, analysis of individuals counting, discussion, outcomes and conclusion which are specified in parts I, II, III, IV, and V correspondingly.

II. Individual Counting Analysis

Individuals counting approaches are classified into three techniques, namely, "Detection techniques", "Clustering techniques", "Regression techniques", and "Density Map techniques".

A. Detection Technique

The detection-based technique assigns the detector to recognize the individuals in the video scene for obtaining the total value of individuals with their location [31].

The well-known detection methods commonly include detection based on the body [32], shoulders [33] and heads [34]. In heavy occlusion condition, only the heads of individuals might be appearing so, head detection methods mostly show well performance in contrast to the other schemes of shoulder and body detection. The problem in detection technique is to identify an individual by itself, mostly in the existence of crowds and overlapping among objects [35]. Although this kind of method shows promising outcomes, the carefully hand-engineered designed indicator is not precise in the blurred scene where the quality of utmost videos is relatively low [1][35][36]. Furthermore, this scheme is unable to deal with the overpopulated region or unstable climate as well as to count the total value of individuals with the different flow direction, and it limits its operation in real scenes [15]. These schemes are efficient to produce appropriate detections in sparse scenes, providing the total value of individuals and location, in addition, to posture of all individuals in a scene. However, these approaches are travail from sight clutter. A multi-camera adjustment with overlying views is not available in many circumstances, and also scene not being able to promise satisfactory cross-dataset generalization [10], while training of a specific scene indicator for counting is difficult [18]. The tracking and detecting of individuals are getting more complex while the crowded environment is getting denser and larger [37]. Nevertheless, these approaches require huge computing resource and are often limited by an overlapping and complex background in

realistic situations, subsequent a low accuracy [38]. This owes to the fact that these methods would have to scan each frame via a trained detector, which normally takes more time to obtain the total value of individuals [39]. The individuals counting through detection-based approaches need more time whilst applied to the scene with occlusion and lighting changes [40]. These approaches are able to provide higher accuracy in the crowd scene with low density in comparison to the scene with high crowd density [41]. Additionally, it is essential to have scene with high resolution to get high precisions [42]. The pros and cons of prevalent detection techniques are shown in Table I.

B. Clustering Technique

The total number of individuals in the clustering-based techniques are obtained through tracking techniques algorithms [54]. In these approaches the useful information/features are followed out frame by frame, then by using spatial and temporal constancy heuristics, they are able to cluster the direction and also use extra elements to accomplish the spatial path for each person [9][54][55]. The total number of individuals is indicated based on the number of clusters [56]. As these methods highly depend on tracking approaches, they are time-consuming and need high computation resource. The performance of these techniques degrades excessively while they are applied under certain conditions of variations in illumination, low resolution, and motion imaging platforms, which reduces the stability of the tracking algorithm [30][54][57]. These techniques are based on extracting and counting the individuals' blobs of a temporal slice of the video, the certain blobs which consist of many individuals are not able to provide precise individual counting due to severe occlusion [58][59]. Furthermore, the accuracy of these techniques is influenced by the inaccuracy of coherently unbalanced info which is not matched with the exact object [60]. Mainly in a real scene, the handcraft information is hardly capable to adapt with different climate conditions, and lighting [58][59]. However, a model propounds movement coherency, thus the probability of improper approximation increases when objects staying constant in a scene, signifying objects which allocating similar info at a certain period [61]. These methods are capable to handle the consecutive image frames while other counting approaches are not facing this restriction [18]. Also, these approaches are able to provide an accurate result when dependable trajectories can be extracted [39].

C. Regression Technique

The regression-based techniques, by gaining the low-level features map, are able to count the individuals. Typically, the regression algorithms encompass the appropriate features such as texture features [11][62][63] and key points [11][62]-[64] which are extracted from the foreground through background subtraction approaches. These techniques obtain the correlation amongst features and count individuals through the training of extracted features without considering the individuals' identification [54]. These approaches provide superior performance under over-crowded scene circumstances [65] by escaping the issue of hard detection. However, these approaches lack of the ability to provide the information about the individual count, and also, disability to specify the position of each individual, which restricts these approaches for localization tasks [7][66]. It is crucial for video surveillance systems to have information about the position of each person in the scene in order to acquire the spatial distribution of individuals [11][24][67]. The computational cost of these approaches is very low as these methods do not need to detect and trace the individual [37]. Table II summarizes the advantages and disadvantages of six popular regression-based techniques.

D. Density Based Technique

In the Density-based techniques, the total value of the individuals is equivalent to the integral of the density map around sub-region. These approaches are employed for both counting and localization by retaining the spatial information which renders these methods very effective for describing the density distribution of crowds. Notwithstanding, these approaches might collapse where only heads of individuals can be observed. These methods are able to tackle the individual's counting issue, under high occlusion, and low resolution in the video scene [22]-[25][7]. The performance of this technique is extremely based on the types of features [22]-[25][7], where the convolutional neural network

TABLE I. The Pros and Cons of Individuals' Detection-Based Techniques

No	Detection Approach	Function	Pros	Cons	Paper
1	Monolithic	• Use the appearance of full body to train the classifier. The quality and speed of detection is on the bases of classifier's choice.	• Provides reasonable detection in sparse scenes.	• Not applicable under overlapping situation among the individuals in scenes. • Not being able to deal with clutter areas.	[43][44][45][46]
2	Part Based	----------	• More robust in comparison to monolithic method as whole body is observable.	• Unable to provide precise detection solely based on head region.	[47]
3	Shape Matching	----------	• It provides the info about pose as well as count and location of each individuals.	----------	[48][49]
4	Multi-Sensor	• Accessible to multiple camera where the uncertainties caused by overlap among individuals can be dissolved via one camera.	----------	• Not applicable if multiple camera is not available.	[50][51]
5	Transfer Learning	• Detecting the individuals in new scenes without controlling by human.	----------	• Not applicable if scene contains low resolution, varying in viewpoints and illumination.	[45][52][53]

(CNN) is known as the best technique for extracting the features [30]. However, the CNN method cannot efficiently tackle three difficulties of distributions of non-uniform density, scale changes, and drastic scale changes [30] and also causes the resolution of the density maps to decrease which has no discernible effect on the development of the precise individual counting, even though it prevents from localizing the individual satisfactorily [36]. By considering the mentioned difficulties as a preventive factor from attaining higher precisions, the CNN schemes particularly handle these difficulties through multi-column or multi-resolution network [36]. Though these approaches exhibit their strength to non-uniform distribution density, scale changes, and drastic scale variation, they still pose some limitation to the size that is applied during training and thus their proficiency delimited for learning better-common approach [39]. In comparison to regression techniques, the density maps have been shown to be very efficient at solving the issue of individual counting, especially where the scene contains high inter-occlusion, with low-resolution surveillance videos [22]-[25][7] and also, these methods are providing spatial information about the crowd.

III. Experimental Results

This research aims to cater a review based on recent counting approaches and work is to specify the effectiveness of CNN used on popular individuals counting approaches for attaining higher precision results. In the field of computer vision, individual counting is considered as a fundamental task. There are several techniques such as counting by clustering, detection, and regression that exist to resolve the issue of counting individuals. However, these techniques fall short of overcoming the difficulties such as overlapping objects, the difference in illumination, unstable weather, overpopulated scene. These techniques, albeit, have shown astonishing performance while incorporated with CNN to attain useful data from the image /video frames which play an important role in both ROI and LOI [8]. Through utilizing the advantage from CNN for representation of the image, the individuals counting has shown great achievements in contrast to shallow approaches to tackle the issue of the populated scene. Table III shows some certain individual counting techniques mentioning the quantitative outcomes from the corresponding reference.

IV. Discussion

There are several techniques for individual counting namely, detection, clustering, and regression. These methods were employed to overcome the difficulties of counting, which have dissimilar reasons under different conditions in an overpopulated area. The detection-based approaches specify the total number of individuals in the crowded scene by assigning a detector for recognizing each individual [73]. These approaches are only practical to a specific scene and will fail for some certain real-life applications [73][74]. These approaches have difficulty to obtain precise outcome under various conditions such as unstable weather or individuals with opposite flow path [15]. Besides, they are time-consuming as the detector scans every frame of the scene [39]. The clustering methods obtain the total number of persons in a populated scene via employing tracking systems [73]. These approaches also take a long time to process and suffer from computation difficulty as they are closely dependent on the tracking algorithms. Moreover, their performance degrades significantly when employed to scene with variation in illumination, unclear resolution, and motion imaging platforms, which causes the tracking algorithm to be unsteady. The regression-based methods obtain the total value of

TABLE II. The Advantages and Disadvantages of Individuals Regression-Based Methods

No	Detection Approach	Function	Pros	Cons	Paper
1	Linear Regression	----------	• Yields high performance in sparse scene where the crowds are small and there is less occlusion among the objects.	• Some objects are not beneficial for counting estimation.	[68]
2	Partial Least Square Error (PLSR)	----------	• Produces good performance under various crowdedness levels and unseen density. • Being capable to overcome the collinearity issue with fewer factors. • Needs less computations. • Converges very fast.	• Sensitive to the proportion of positive and negative class in training data.	[69]
3	Kernel Ridge Regression (KRR)	• Its extended nonlinear form of ridge regression which obtained from kernel tricks.	• Diminish the issue of multiple colinear.	----------	[47]
4	Support Vector Regression (SVR)	----------	• Requires less time testing to approximate the solution.	----------	[70]
5	Gaussian Process regression (GPR)	• Most prevalent regression-based counting technique.	----------	• Not reliable to deal with big datasets.	[69]
6	Random Forest regression (RFR)	• -Attain to nonlinear scalable regression.	• Less susceptible to the parameters. • Being to be scaled to big dataset.	• Disable to work with the points that are out of the range of target value.	[71]

TABLE III. Some Certain Individual Counting Techniques with Quantitative Outcomes

Year	Paper	Description	Method	Challenges	Datasets	Results
2016	[1]	Count the people on basis of head detection techniques by composition of Adaboost algorithm and the CNN.	Detection based methods.	Low resolution data, body occlusion and not limited imaging viewpoints.	Real classroom surveillance.	**Stage**=15 **Recall**=0.86 **Precision**=0.33
2015	[8]	Proposed a new technique to approximate the overall value of arriving and leaving crowd flow with three CNN methods.	Line of feature (LOI).	High flow density crowd, various illuminations and different mal-weather.	Large dataset that consists of various real videos of public gates.	**Precision**= 95.06%
2017	[59]	Introduces passenger counting system using CNN and Spatio-temporal Context.	Region of interest (ROI).	Complex low-resolution scene for public transportation.	Public bus transportation in China.	**Recall**= 94.215 **Precision**= 92.486
2017	[60]	Evaluation of produced density maps thru density estimation techniques on different crowd analysis tasks, such as counting, detecting, and tracking.	Density map based.	Reduction of density map resolution while using CNN based method which can degrade the performance of localization tasks (detection and counting).	UCSD [62] UCF CC 50 [18] WorldExpo'10 [7] TRANCOS [40]	**(CNN- Pixel) Error Distance (ED)**= 3.61±0.72 **(CNN- Pixel) Error Difference Distance**= 2.90±0.83 **(FCNN-Skip) Error Distance (ED)**= 3.61±0.72 **(FCNN-Skip) Error Difference Distance**= 3.38±1.01
2018	[30]	Proposed multi-column multi-task convolutional neural network (MMCNN) for counting the crowd.	Density estimation.	Drastic size change and unsteady density distribution.	UCSD [63] UCF CC 50 [18] WorldExpo'10 [7] Shanghai Tech [36] MALL [73]	**UCSD MAE**= 1.02 **UCSD MSE**= 1.18 **MALL MAE**= 1.98 **MALL MSE**= 5.68 **UCF CC 50 MAE**= 320.6 **UCF CC 50 MSE**= 323.8 **WorldExpo'10 MAE**= 9.1 **WorldExpo'10 MSE**= 18.7 **Shanghai Tech Part A MAE**= 91.2 **Shanghai Tech Part A MSE**= 128.6 **ShanghaiTech Part B MAE**= 18.5 **ShanghaiTech Part B MSE**= 29.3

individuals in the scene through mapping amongst low-level features [73]. Such techniques generally achieve higher performance for the overcrowded scene. However, they lack in providing the information of individuals location, hence not being used for object localization [66] [7]. As the goal of this study is to introduce the most suitable counting methods, the density map estimation has shown encouraging results in comparison to the other techniques, which can provide the location of the individuals in a very crowded scene. These approaches are very potential in tackling the individuals counting issue where heavy inter occlusion and unclear scene exist among the objects. However, they are unable to provide the precise individuals counting where only heads of individuals are obvious. The effectiveness of individual counting methods is indicated in Table IV.

V. Conclusion

This study endeavors to provide a revision on the existing individuals counting techniques namely, clustering, detection, regression, and density map-based methods. Among these techniques, the regression-based ones show great performance under overcrowded area. The regression-based approaches are capable to count the individuals in a crowded environment, although they are not applicable for localization task. The density map estimation approaches are very promising among other counting methods since they preserve the beneficial spatial data for two tasks of counting and localization. The density map technique is very efficient to resolve the issue of individuals counting under different circumstances such as the massive overlapping among the objects and unclear scene in consequence frames. This scheme is proper for defining individual's density distribution since it focuses on both details on location and spatial data but it might fail once the heads of individuals appear. While the functioning of this method relies on the type of features, the convolutional neural networks (CNN) are capable to extract the useful information. The CNN based counting approaches are not being able to tackle three problems, i.e., distributions of non-uniform density, a variation of scale. Furthermore, applying CNN-based approaches to density estimation techniques can reduce the resolution of these techniques rendering it ineffective for localization tasks. By taking these problems into account as a restrictive

TABLE IV. The Effectiveness of Individual Counting Methods

Methods	Task	Processing Period	Computational Complication	Performance					Scene Type	
				Illumination Alteration	Low Resolution	Occlusion	Overcrowded area	Small Object Size	Consecutive	Fixed
Clustering	Counting by tracking techniques	High	High	Low	Low	Low	Low	Low	Applicable	Inapplicable
Corresponding Authors	[54]	[54]	[54]	[58][59]	[54][57]	[58][59]	[73]	[29]	[18]	[18]
Detection	Count by scheming detector to identify individuals	High	--------	--------	Low	--------	Low	Low	Applicable	Applicable
Corresponding Authors	[31]	[31]	--------	--------	[19]	--------	[73]	[29]	[18]	[18]
Regression	Approximate density of crowd on the basis of holistic and collective description of pattern	--------	--------	--------	--------	High	High	High	Applicable	Applicable
Corresponding Authors	[55]	--------	--------	--------	--------	[39]	[65]	[29]	[18]	[18]
Density Map	Preserves spatial info which makes it useful for both counting and localization tasks	--------	--------	--------	High	High	High	High	Applicable	Applicable
Corresponding Authors	[30]	--------	--------	--------	[22]-[25], [7]	[22]-[25], [7]	[29]	[29]	[18]	[18]

clause to accomplish superior precisions, particular CNN techniques accurately resolve the three mentioned difficulties via multi-column or multi-resolution net. Though these approaches showed robustness to distributions of non-uniform congestion and size variations, they are yet restricted to the size of the dataset used for training which makes their ability to be limited to achieved better methods. Based on the experimental results of previous work we found that the crowd density approach is more beneficial in comparison to other counting methods as it provides information about the location of the crowd. However, the performance of this technique is highly depending on the types of features in which the usage of best deep learning technique (CNN) can be very valuable for this method to extract the important feature from each video frame. The usage of (CNN) with a density map approach will help the future research in scheming more precise counting techniques to approximate density maps with high quality for both tasks of counting and localization.

Acknowledgment

This research is based on two research grants with code of DIP-2014-039 and AP2017005/2.

References

[1] Ch. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, People counting based on head detection combining Adaboost and CNN in crowded surveillance environment,Journal of Neurocomputing, Vol. 208, no.C, pp.108–116, 2016.

[2] B. Zhou, X. Tang, H. Zhang, and X. Wang, Measuring crowd collectiveness, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, pp. 1586–1599, 2014.

[3] B. Zhou, X. Tang, and X. Wang, Coherentfiltering: detecting coherent motions from crowd clutters, In: European Conference on computer Vision, LNCS, vol. 7573, pp. 857-87, 2012.

[4] M. Rodriguez, J. Sivic, I. Laptev, and J. Y. Audibert, Data driven crowd analysis in videos, In: International Conference on Computer Vision, Barcelona, Spain, 2011.

[5] F. Zhu, X. Wang, and N. Yu, Crowd tracking with dynamic evolution of group structures, In: European Conference on computer Vision, LNCS, vol. 8694, pp. 139-154, 2014.

[6] K. Kang, and X. Wang, Fully convolutional neural networks for crowd segmentation, Computer Vision and Pattern Recognition, 2014.

[7] C. Zhang, H. Li, X. Wang, and X. Yang, Cross-scene crowd counting via deep convolutional neural networks, In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07, pp. 833–841, 2015.

[8] L. Cao, X. Zhang, W. Ren, and K. Huang, Large scale crowd analysis based on convolutional neural network, Pattern Recognition, vol. 48,

no.10, pp. 3016–3024, 2015.

[9] V. Rabaud, and S. Belongie, Counting crowded moving objects, In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 705–711, 2006.

[10] P. Dollár, C. Wojek, B. Schiele, and P. Perona, Pedestrian detection: An evaluation of the state of the art, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 743–761, 2012.

[11] A. B. Chan, and N. Vasconcelos, Counting people with low-level features and bayesian regression, In: IEEE Transactions on Image Processing, vol. 21, pp. 2160–2177, 2012.

[12] A, C. Davies, J, H. Yin, and S. A. Velastin, Crowd monitoring using image processing, Electronics & Communications Engineering Journal, vol. 7, no.1, pp. 37-47, 1995.

[13] C. Zhang, H. Li, X. Wang, and X. Yang, Cross-scene crowd counting via deep convolutional neural networks, In: IEEE Conference on Computer Vision and Pattern Recognition (*CVPR)*, pp. 833–841, 2015.

[14] C. Castellano, S. Fortunato, and A. Loreto, Statistical physics of social dynamics, Journal of Reviews of modern physics, vol. 81, no. 2, pp. 81–591, 2009.

[15] H. Chat´e, F. Ginelli, G. G´egoire, and F. Raynaud, Collective motion of self-propelled particles interacting without cohesion, Journal of Reviews of modern physics, vol. 77, no.4, 2008.

[16] J. Shao, C. C. Loy, and X. Wang, Scene-independent group profiling in crowd, In: IEEE Conference on Computer Vision and Pattern Recognition *(CVPR*), 2014.

[17] J. Shao, K. Kang, Ch.Ch. Loy, and X. Wang, Deeply Learned Attributes for Crowded Scene Understanding, In: IEEE Conference on computer vision and pattern recognition (CVPR), Boston, MA, USA, 2015.

[18] Ch, Ch. Loy, K. Chen, Sh. Gong, and T. Xiang, Crowd Counting and Profiling: Methodology and Evaluation, In: IEEE Transactions on Information Technology in Biomedicine, vol. 2, 2013.

[19] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, Deep People Counting in Extremely Dense Crowds, In: Proceedings of the 23rd ACM international conference on Multimedia, pp. 1299–1302, 2015.

[20] Ö. Yeniay, and A. Götas, A Comparison of Partial Least Squares Regression with Other, In: Journal of Mathematics and Statistics, vol. 31, pp. 99–111, 2002.

[21] R.D. Cramer III, Partial Least Squares (PLS): Its strengths and limitations, In: Perspectives in drug discovery and Design, vol.1, pp. 269-278, 1993.

[22] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, Interactive object counting. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, vol. 8691, pp. 504–518, 2014.

[23] L., Nair, R., Koethe, U., F. A. Hamprecht, Learning to Count with Regression Forest and Structured Labels, In: International conference on pattern recognition (ICPR), pp. 2685–2688, 2012.

[24] V. Lempitsky, and A. Zisserman, Learning To Count Objects in Images, In: Advances in Neural Information Processing System, pp. 1324–1332, 2010.

[25] V. Q. Pham, T. Kozakaya, O. Yamaguchi, and R. Okada, COUNT forest: Co-voting uncertain number of targets using random forest for crowd density estimation, In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2015, pp. 3253–3261. 2015.

[26] Sh. Pu, T. Song, Y. Zhang, and D. Xie, Estimation of crowd density in surveillance scenes based on deep convolutional neural network convolutional neural network, In: 8th International Conference on Advances in Information Technology, IAIT2016, pp. 154–159, Macau, China, 2016.

[27] F. Xiong, X. Shi, and D. Yeung, Spatiotemporal Modeling for Crowd Counting in Videos, In: International Computer Vision (ICCV), 2017.

[28] N. Bouchra, A. Aouatif, N. Mohammed, and H. Nabil, Deep Belief Network and Auto-Encoder for Face Classification, International Journal of Interactive Multimedia and Artificial Intelligence, http://dx.doi.org/10.9781/ijimai.2018.06.004, 2018.

[29] D. Kang, Z. Ma, and A. B. Chan, Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks - Counting, Detection, and Tracking, In: IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), pp. 1–14, 2017.

[30] B. Yang, J. Cao, N. Wang, Y. Zhang, and L. Zou, Counting challenging crowds robustly using a multi-column multi-task convolutional neural network, Journal of signal processing: image communication, vol. 64,pp. 118-129, 2018.

[31] O. Sidla, Y. Lypetskyy, N. Brändle, and S. Seer, Pedestrian detection and tracking for counting applications in crowded situations, In: Proceedings IEEE International Conference on Video and Signal Based Surveillance 2006, AVSS' 06, 2006.

[32] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, A method for counting people in crowded scenes, In: Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010, pp. 225–232, 2010.

[33] W. Li, X. Wu, H, A. Zhao, New techniques of foreground detection, segmentation and density estimation for crowded objects motion analysis, Journal of Information and Media Technologies, vol. 6, no. 2, pp. 528–538, 2011.

[34] T. Van Oosterhout, S. Bakkes, and B. Krse, Head Detection in Stereo Data for People Counting and Segmentation, In: Proceedings of the International Conference on Computer Vision Theory and Applications, pp.620–625, 2011.

[35] H. Fradi, and J. L. Dugelay, Low level crowd analysis using frame-wise normalized feature for people counting, In: Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS 2012), pp.246–251, 2012.

[36] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, Single-Image Crowd Counting via Multi-Column Convolutional Neural Network, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 589–597, 2016.

[37] H. Foroughi, N. Ray, and H. Zhang, Robust people counting using sparse representation and random projection, Journal of Pattern Recognition, vol. 48, no.10, pp.3038–3052, 2015.

[38] L. Zeng, X. Xu, B. Cai, S. Qiu, and T. Zhang, Multi-scale Convolutional Neural Networks for Crowd Counting, Retrieved from http://arxiv.org/abs/1702.02359, 2017.

[39] B. Xu, and G. Qiu, Crowd density estimation based on rich features and random projection forest, In: IEEE Winter Conference on Applications of Computer Vision(WACV 2016), 2016.

[40] K. Chen, C. C. Loy, S. Gong, and T. Xiang, Feature mining for localised crowd counting, In: Procedings of the British Machine Vision Conference 2012, vol.1, pp. 1–11, 2012.

[41] Y. L. Hou, and G. K. H. Pang, People counting and human detection in a challenging situation, In: IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, vol. 41, pp. 24–33, 2011.

[42] 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, Journal of Visual Communication and Image Representation,vol. 39, pp. 218–225, 2016.

[43] O. Tuzel, F. Porikli, and P. Meer, Pedestrian detection via classification on Riemannian Mani folds, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no.10, pp. 1713–1727, 2008.

[44] P. Sabzmeydani, and G. Mori, Detecting pedestrians by learning shapelet features, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2007.

[45] P. Dollar, C. Wojek, B. Schiele, and P. Perona, Pedestrian detection: An evaluation of the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, pp. 1–1, 2011.

[46] M. Pa¨tzold, R. Evangelio, and T. Sikora, Counting people in crowded environments by fusion of shape and motion information, In: IEEE International Conference on Advanced Video and Signal based Surveillance, pp. 157–164, 2010.

[47] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.

[48] C. Lampert, Kernel methods in computer vision, vol. 4. Now Publishers Inc, 2009.

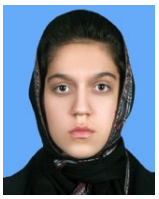
[49] W. Ge, R. Collins, Marked point processes for crowd counting, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2913–2920, 2009.

[50] W. Ge, and R. Collins, Crowd detection with a multi view sampler, European Conference on Computer Vision, pp. 324–337, 2010.

[51] R. Benenson, M. Mathias, R. Timofte, and L.V. Gool, Pedestrian detection at 100 frames per second, In: IEEE Conference Computer Vision and Pattern Recognition, 2012.

[52] M. Wang, and W. Li, X. Wang, Transferring a generic pedestrian detector

towards specific scenes, In: IEEE Conference Computer Vision and Pattern Recognition, 2012.
[53] M. Wang, and X. Wang, Automatic adaptation of a generic pedestrian detector to a specific traffic scene, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3401–3408, 2011.
[54] I. S. Topkaya, H. Erdogan, and F. Porikli, Counting people by clustering person detector outputs, In: 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014, pp. 313–318, 2014.
[55] A. M. Cheriyadat, B. L. Bhaduri, and R. J. Radke, Detecting multiple moving objects in crowded environments with coherent motion regions, In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, 2008.
[56] Z. Ma, and A, B. Chan, Crossing the line: Crowd counting by integer programming with local features, In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2539–2546, 2013.
[57] G. Antonini, and J. P. Thiran, Counting pedestrians in video sequences using trajectory clustering, In: IEEE Transactions on Circuits and Systems for Video Technology,vol.16, pp.1008–1020, 2006.
[58] Y. Cong, H. Gong, S. C. Zhu, and Y. Tang, Flow mosaicking: Real-time pedestrian counting without Scene-specific learning, In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, pp. 1093–1100 ,2009.
[59] G. Liu, Z. Yin, Y. Jia, and Y. Xie, Passenger Flow Estimation Based on Convolutional Neural Network in Public Transportation System, Journal of Knowledge-Based Systems, vol. 123, pp. 102–115, 2017.
[60] R. Shbib, S. Zhou, D. Ndzi,and K. Al-kadhimi, Distributed Monitoring System Based On Weighted Data Fusing Model, American Journal of Social Issues and Humanities, pp. 53–62, 2013.
[61] A. E. Hoerl, and R. W. Kennard, Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, vol. 12, no. 1, pp. 69–82, 1970.
[62] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, Privacy preserving crowd monitoring: Counting people without people models or tracking, In: 26th IEEE Conference on Computer Vision and Pattern Recognition,CVPR, 2008.
[63] K. Chen, C. C. Loy, S. Gong, and T. Xiang, Feature Mining for Localised Crowd Counting, In: Procedings of the British Machine Vision Conference, 2012.
[64] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, Crowd counting using multiple local features, In: DICTA 2009 - Digital Image Computing: Techniques and Applications, pp. 81–88, 2009.
[65] A. Adegboye, G. Hancke, and G. H. Jr, Single-pixel approach for fast people counting and direction estimation, In: Southern Africa Telecommunication Networks and Applications, 2012.
[66] X. Zeng, W. Ouyang, M. Wang, and X. Wang, Deep learning of scene-specific classifier for pedestrian detection, In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, Vol. 8691, pp. 472–487, 2014.
[67] H. Idrees, K. Soomro, and M. Shah, Detecting humans in dense crowds using locally-consistent scale prior and global occlusion reasoning, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, pp. 1986–1998, 2015.
[68] K. De Brabanter, J. De Brabanter, J. Suykens, and B. De Moor, Approximate confidence and prediction intervals for least squares support vector regression, IEEE Transactions on Neural Networks, vol. 99, pp. 1–11, 2011.
[69] N. Dalal, and B. Triggs, Histogram so oriented gradients for human detection, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 886–893, 2005.
[70] R. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, vol. 3, no.6, pp.610–621, 1973.
[71] S. Cohen, Background estimation as a labeling problem, In: IEEE International Conference on Computer Vision, vol. 2, pp. 1034–1041, 2005.
[72] D. Hal, Frustratingly Easy Domain Adaptation, Journal of ACL, 2009.
[73] Z. Zhang, H. Gunes, and M. Piccardi, Head detection for video surveillance based on categorical hair and skin colour models, In: Proceedings - International Conference on Image Processing, ICIP, pp. 1137–1140, 2009.
[74] D. Merad, K. E. Aziz, and N. Thome, Fast people counting using head detection from skeleton graph, In: Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010, pp. 233–240, 2010.



Anahita Ghazvini

PhD candidate in computer science at Universiti Kebangsaan Malaysia (UKM). Received the bachelor degree with honours in information technology (computer science) at Universiti Kebangsaan Malaysia (UKM) in 2013. Received a master degree in information technology (artificial intelligence) at University Kebangsaan Malaysia (UKM) in 2016.

Siti Norul Huda Sheikh Abdullah

Received her first Degree in Computing at University of Manchester Institute of Science and Technology, United Kingdom. She furthered her master study in the area of Artificial Intelligence in Universiti Kebangsaan Malaysia. Later, she continued her Phd Study in the area of Computer Vision at Faculty of Electrical Engineering, Universiti Teknologi Malaysia. Starting the career, she involved in conducting national and international activities such as Royal Police Malaysia, Cyber Security Malaysia, Cyber Security Academia Malaysia, Federation of International RobotSoccer Association (FIRA), Asian Foundation, Global Ace Professional Certification Scheme, MIAMI, MACE and IDB Alumni. She is now holding a post as the Chairperson of Center for Cyber Security. Her research focuses are Digital Forensics, Pattern Recognition, and Computer Vision Surveillance System. She has published two books entitled “Pengecaman Pola” or “Pattern Recognition” and “Computational Intelligence for Data Science Application” and more than 50 and 100 of journal and conferences manuscripts correspondingly.

Masri Ayob

Prof Dr Masri Ayob is a lecturer in the Faculty of Information Science and Technology at the Universiti Kebangsaan Malaysia (UKM) since 1997. She has obtained her PhD in Computer Science at The University of Nottingham in 2005. Her main research areas include meta-heuristics, hyper-heuristics, scheduling and timetabling, especially educational timetabling, healthcare personnel scheduling and routing problems, and Internet of Things. She has published more than 100 papers at international journals and at peer-reviewed international conferences. She has been served as a programme committee for more than 50 international conferences and reviewers for high impact journals. She was a member of ASAP research group at the University of Nottingham. Currently, she is a principle researcher in Data Mining and Optimisation Research Group (DMO), Centre for Artificial Intelligent (CAIT), UKM.