# Analytical Study of Different Techniques in Crowd People Counting Framework

Htet Htet Lin and Kay Thi Win

**Abstract**— The more increasing the world population, the higher the using of the surveillance cameras due to safe the behavior of people. The estimating numbers of world population of 2050 will grow to 9.4 billion. So, crowd analysis is more critical for peoples' safety. This paper aims to gauge all main recent terminology of the people counting systems. This describes the problems faced by data sets and contributors' contributions. This paper has also been addressed the several techniques of counting people, compared them with the help of evaluation performance measures which are widely used for counting. The paper also highlights the analysis to find out the best strategies with some prominent existing methods for the researchers.

**Index Terms**— Crowd Counting, Crowd Dataset, Feature-Based, Hardware-Based, Map-Based, Trajectory-Based

## 1 Introduction

THE world current population is about 7.7 billion due to the hot statistical reports. All areas in the world are interconnected with some form of transport systems. Where ever we go all over the world we are facing with the one or many problems of the heavy crowd situations due to increasing population and more modern development in technology. So, crowd analysis is great interest and still an active area in computer vision technology for noticing all types of environment and applications such as urban space planning, people monitoring and security, disaster prevention, audience counting, meeting room management, resource management and control, student attendance counting system, public safety, crowd behavior modeling, etc.. It is also a developing area of study which was inspired by the twenty-five safety concerns nearby the heavily or lightly crowded environments.

Existing earlier research works are clustered into four groups: (1) Feature-based counting approach, (2) Trajectory clustering-based counting approach, (3) Map-based counting approach and (4) Hardware-based counting approach. This is shown in Fig. 1. In Feature-based counting approaches, this method involve preparing the detector a visual object to search for and count all in the scene. Trajectory clustering-based counting techniques attempt to study the centroid distances among objects. On the other hand, Map-based counting techniques attempt to study the automatically related mapping of low-level features with the whole number of people in the frame or inside a frame region. Hardware-based counting approaches are depended on the hardware devices, it leads to cost expensive.

The key contribution of this paper is to investigate various methods applied in different surveillance cameras or sensors with specific attention towards people tracking or detection and people counting or density estimation in crowded scenes.

Earlier, various review papers have been issued recently about the crowd analysis. However, this paper aims to offer a comparative analysis of the crowd analysis that considers on clustering four groups of various techniques.

In this paper all main terminology of people counting system has been addressed. This paper also shows the amazing progress in people detection, tracking and counting has been achieved by new classification approaches, features, deformation models, fast algorithms and datasets. This paper organizes as follows. Section 2 presents the earlier methods and technologies and their applicable work on the crowd counting system has been discussed. Section 3 shows the diverse datasets offered to contribute in the crowd sensing field. Section 4 investigates and discusses the performance analysis of crowd analysis. The rest Section 5 has been attempted the concerns and conclusion.

## 2 Review of Counting Techniques

### 2.1 Feature-based Counting Approach

Most of state-of-the-art emphasized on detection framework that performed features (such as Haar wavelets, histogram oriented gradients, edgelet and shapelet) extracted from a fully whole body or part of the body [25], [18], [22] and [21]. And then these features are trained into various classifiers (such as adaboost boosting, random forest and Support Vector Machines) [26]. According the literature survey, detection is generally performed either in the monolithic detection approaches and part-based detection approaches. Numerous detection approaches have been suggested to evaluate the counting system.

Histogram of Oriented Gradients (HOG) descriptor is the significant development of feature descriptors for people [18]. Extension of this work has also got a significant improvement [24]. The occurrences and co-occurrences of gradient direction are computed in some parts of the image or in the entire image. Due to a dense grid of uniformly spaced cells, they have got the distinctiveness computation. The normalization of overlapping local contrasts could also improve accuracy. This descriptor has

*Htet Lin is currently pursuing Ph.D. degree program in University of Computer Studies, Mandalay, Myanmar. E-mail: htethtet.linnnnn@gmail.com*
*Kay Thi Win is currently working as Professor in University of Computer Studies, Mandalay, Myanmar. E-mail: kthiwin11@gmail.com*

also been used by the AdaBoost classifier [17] and [27], and a particle filter [24]. Their spatial space is also used for face detection in combination with the KLT tracker [17]. Jin and Bhanu [27] have tried a crowd simulation that combines the tracking of several people. Ge et al. [24] have addressed a detection and tracking framework for the fewest number of pedestrians (small crowds). They have also used HOG detector by combining a correlation tracker to detect fully body and localize people in the crowd. But, the problem is high miss detection rate due to occlusion issue. They can work well only in the small crowd (consisted the small number of people).

Viola and Jones [25] have been addressed a face detection framework built upon Haar wavelets features to reveal the integral images. They aimed to compute feature and a cascade arrangement for fast effective detection. AdaBoost have been used with this feature for automatic feature utilizing and selection. These concerns have also come as the continuous foundation of modern detectors. Nevertheless, they have the high runtime to give the efficiency of current detectors.

Nevatia [22] have presented a human detection framework in crowded scenes from static images. They introduced the "edgelet" features, to locally curve and segment shape representation. They focused to tackle the inter occlusion problem. But the view is limited not to exceed about 45 degrees of front, rear, upright standing or walking pose. D. Gavrila and V. Philomin [8] and [9] have been employed the Hausdorff distance transform to introduce contour features also a frequent cue for detection and rapidly matching image edges with this set of shape templates. They have the miss detection issue when the situation that the person wearing clothes with the same color to the background or the floor or etc.

Ryan, David, et al. [20] has presented a multiple local feature based method to count each foreground blob segment of the scenes. They could use all various areas of the scene or the multiple camera location. This method reduces the required training data. Due to imperfect foreground segmentation, some blobs are disposed to errors such as splitting, fading and noise which reduces overall precision in counting.

Researchers are trying to solve this problem by applying detection methods based on methods [12], [18] and [23], where one constructs to estimate the number of people in the designated area, head and shoulder [14]. In another method using shape learning, Zhao et al. [28] simulated humans using a three-dimensional shape of assembled ellipsoids, and participated in a stochastic process to estimate the true amount of land using the foreground mask shape in a video sequence. Ge and Collins [13] further extended the idea by using a flexible and practical shape model.

F. Duc, et al. [10] have suggested a counting people model in which the first step is foreground segmentation and the diverse blobs gets the estimated head and ground planes. Later predictions are used to calculate the person's number. This estimation count is jointed with tracking algorithm to get a smooth estimate count. They couldn't work well in public places such as railway stations or airports, etc.

These methods only got high performance in low density scenes, i.e. four or five people in the scenes. Although they used part based or shape based detector, they were not mitigated the problem of occlusion, illuminations and not suitable for crowded scene.

### 2.2 Trajectory-based Counting Approach

For the trajectory-based counting method, it is also known as tracking approach, is the estimating trajectory problem of people motion in the video surveillance camera [3] and [11]. In these approaches, there have many key challenges such as environmental challenges (i.e. snow, rain or shadows, etc.), various viewpoint variations, lighting issues, noisy condition, occlusion and crowd density. Several motions based models (optical flow or background modeling and subtracting) have been developed for the tracking approach.

C. Yang et al. [7] have presented the tracking based counting approach. Trajectories feature is tracked and clustered into object tracks or based on extracting the temporal slice and counting blobs from the video. R. Vincent, et al. [20] has addressed a paralleled form of the KLT method to assess the video into a trajectory feature set. They presented a simple means of spatially and temporally condition of the motion trajectories for subsequence processing. Then they have combined it with a learned people descriptor to get the constituent motion segmentation. But they have the identifying problems of a complex appearance and motion.

Gianluca et al. [2] have attempted automatic clustering methods for pedestrians counting in video sequences. These techniques are used to reduce the resulting trajectory bias between the actual number of target people and the number of tracking people. They have the problems that the trajectories of the same person's body are similar with the trajectories of other different peoples. This leads to the decreasing tracking rate. B. Gabriel et al [5] have introduced a probabilistic clustering framework that uses low level image features that is well for finding a first estimate of the number and location of individual entities in crowded surveillance scenes. Occlusion that contains the footage of complex pedestrian traffic condition, insects, and animals is preserved as a one-shot fashion, without the benefit of training data or any notion of an appearance model.

The problem of tracking approach is limited by the scene complexity, and they have less number of individuals. And intra occlusion and inter occlusion is also facing as the main challenges for this approach.

### 2.3 Map-based (Regression-based) Counting Approach

To overcome the occlusion issue, most of the researcher emphasized on regression or mapping. They have proposed to count by regression based method of learning a mapping between the actual counts with features extracted from local image patches [13] and [1]. They can independent on learning detectors that is a comparatively difficult task. These frameworks used various classifier (such as linear regression, piecewise linear regression, Gaussian process regression, ridge regression and neural network) [4] to study low-level feature with crowd count. Their framework has focused on the concept of discriminatory attributes used to solve sparse learning data. Their methods can essentially handle

ISSN 2229-5518

unbalanced data. They have a less estimate of people region in real time video due to inter-class and intra-class variations.

Most state of the art has observed that most of the regression based methods ignored the global spatial information. Lempitsky et al [5] have presented a linear framework by learning a mapping between object density maps with corresponding local patch features. Their issue is to optimize a convex cutting plane. Pham et al. [38] have presented a non-linear framework by using random forest to map local patch features with density maps.

The creation of high-quality density and error maps along with a calculation error will be another important problem that must be solved in the future. This model is more sophisticated and time consuming with the two previous approaches.

### 2.4 Hardware-based Counting Approach

In current years, most of the researchers have been proposed the counting system with the help of hardware devices. Xaung et al. [29] have proposed an apparatus hardware-assisted framework in the graph for local triangle counting. They have used the rule patterns by translating the graph vertex relationships. These patterns are prearranged into a finite state machine dependent of the hardware device (a hardware pattern matching accelerator).

SensingKit [39] is a well-organized and does not depend on platform client-server system supported for iOS and Android devices to access motion (Magnetometer Accelerometer and Gyroscope), location (Global Positioning Satellite System) and nearness to other smart-phones (Bluetooth Smart). It has planned for a crowd, sensing application and has used for a system mobile device. The ability of continuous sensing also consisted in this approach. These are application definite tools as for shopping malls, railway stations, music concerts, etc. Another application is an adaptable android based mobile sensing platform, MobiSens [40]. This has been intended for real world applications. The next one is EmotionSense [45] was presented for the social psychology studies based sensing program. It already perceived personal emotions, as well as actions between members of social groups, verbal and close interactions.

Although this approach gave the fasted result depend on the hardware devices, they still face with the cost expensive and less accurate problem due to hardware error.

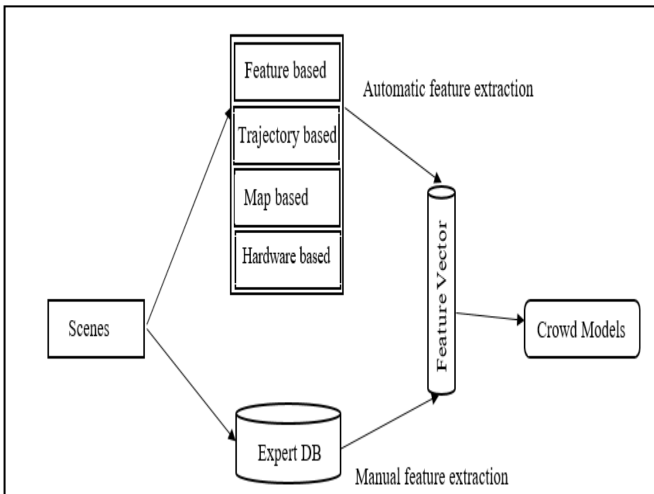


Fig. 1. Architecture of the crowd counting system

## 3 DATASETS

There have various public datasets for people detection and counting framework to evaluate performance test and evaluations. Fig. 2 shows some of the example images of the following benchmark datasets.

The Mall dataset has been proposed by Chen et al. [30]. This has been by captured a publicly accessible surveillance camera in a shopping mall. Grand Central Dataset: Zhou et al. [31] have introduced Grand-Central dataset to consider and learn the heavy crowd cluster's behaviors. It was taken from a scale video (thirty-three minutes) from New York's Grand Cen translation. Ryan et al. [32] have attempted QUT (Queensland University of Technology) dataset that got from their university's campus. It comprises three camera views A, B and C. This video sequence contains some harsh scenes like complex illumination fluctuations, shadows and reflections to create crowd counting more challenging. Moreover, this dataset has focused more occlusion than other datasets.

Tan et al. [33] have addressed Fudan Pedestrian dataset to capture the Guanghua Tower' entry view, which is located at Fudan University, Shanghai, China. They also offer the ground truth file of all the images, the foreground masks and various featured such as Edge, Area, Minkowski, Perimeter, Statistical Landscape Features (SLF) and Ratio. Pets 2009 dataset (Ferryman et al. [34]) has published from the 11st IEEE International Work shop on performance Evaluation of Tracking and Surveillance (PETS) and captured at nights in the University of Reading Campus, UK. This consists of multiple sensor series (Three dissimilar crowd sequences: S1 is the estimation and counting of the people density, size, S2 is the tracking process of the people, and S3 is the event recognition and the flow analysis). The UCSD dataset was initiated by Chan et al. [35] and includes pedestrian's walkways video recorded two videos by a stationary camera at the opposite directions of the UCSD Street.

J. Shao et al. [36] have been proposed CHUK crowd dataset. This is captured at the outdoor of 215 independent scenes. They have also described the ground truth file and aimed to work the academic research. WorldExpo'10 dataset has been proposed by C. Zang et al. [37] and aimed for focusing on cross-scene counting system.

M. Hashemzadeh et al. [41] have been introduced Gate and Bridge Datasets by setting the camera on the top of building gate. This had been captured the video of people entering and exiting a building by the different way of directions. The ground truth is manually count 1000 frames per second. The crowded size of Gate is 1 to 22 people and Bridge is 6 to 30 people.

In spite of the ease of use of an extensive variety of crowd datasets, mainly at hypothetically various crowd density ranks, the accessible public PETS and UCSD datasets have a dominant dataset for most of the previous crowd system. The highlight is that there is very small indication of the earlier research work that has studied the density dependence of

ISSN 2229-5518

crowd analysis applications and consequently the restricted usage of such datasets. The selection of the dataset has been influenced by the selection of crowd framework. Fig. 3 shows to explore that fact.

Fig. 2. Example images of the crowd counting system

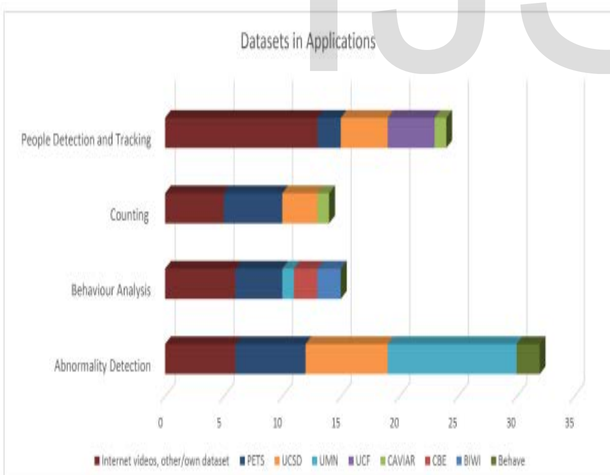


Fig. 3. Application statistics of the crowd datasets

## 4 Performance Analysis and Discussions

In order to choose a suitable crowd dataset, it is vital to learn the diverse assessment criteria applied for determining performance and standard methods adjacent to their particular competing baseline. The two evaluation methods are popular in computer vision. They are qualitative and quantitative evaluation. Some metrics applied to evaluate for crowd analysis are ROC curves, Recall and precision, Detection rate, and Error rate. It is essential to highlight that quantifiable evaluation on crowd detection, tracking and modeling are done another way of each other. It is achievable that the metrics evaluation could be executed on separable pixels in the some case of the background modeling approaches, or using bounding box for the appearance model methods. This is the focus of the accessibility and layout of the ground truth, the execution mechanism of the algorithm and the protocol.

### 4.1 Receiver Operating Characteristic

ROC (Receiver Operating Characteristic) curves: It is a graphical performance of a binary classifier system that is changing the discrimination threshold. The true positive rate (TPR) is designed against the false positive rate (FPR) at several thresholds for creating the curve. They are calculated as follows:

$$TPR = TP/(TP + FN) \tag{1}$$

$$FPR = FP / (FP + TN) \tag{2}$$

where TPR is the true positive rate, TP is the true positive, FN is the false negative, FRP is the false positive rate, FP is the false positive rate and TN is the true positive rate.

Recall (sensitivity) and Precision (positive predictive value): Recall is the proportion of the corresponding instances that are extracted. Precision is the proportion of relevant instances that are extracted. Moreover, F-measure combines recall and precision and provides their harmonic average.

Recall, precision and F-measure are calculated as follows:

$$R = TPR \tag{3}$$

$$P = TP/ (TP + FP) \tag{4}$$

$$F = 2TP /(2TP + FP + FN) \tag{5}$$

$$A = TP + TN /(P + N) \tag{6}$$

where R is the recall, P is the precision, F is the F1 measure, A is the accuracy, TPR is the true positive rate, TP is the true positive, FN is the false negative, FP is the false positive, and TN is the true positive rate.

### 4.2 Error

There have a variety of error calculation types used for the quantifiable evaluation of previous crowd counting approaches. These errors are the mean squared error (MSE), mean absolute error (MAE) and MRE (mean relative error).

MAE (Mean Absolute Error): It is used to evaluate the system performance. MAE can give clear insight that the predicted value is the same with the ground truth data. MRE (Mean Relative Error) is the rate of the comparative the predictor value with the actual value. MAR and MRE are

ISSN 2229-5518

calculated as follows:

$$MAE = 1/TN \left(\sum_{t=1}^{TN} |h_a - h_b|\right) \quad (7)$$

$$MRE = 100\ /TN\ \left(\left(\sum_{t=1}^{TN} |h_a - h_b|\right)/h_a\right) \quad (8)$$

where TN is the total number of testing frames, $h_b$ is the ground truth and $h_a$ is the estimated count of person in frame t.

### 4.3 Analysis

There have a variety of proposed approaches. Most of the work has been tested on one of the public challenging crowd counting dataset such as only on PET 2009 or only on UCSD or etc. Others have been tested on two or three datasets. There have been many datasets of the crowd counting fields. Among them, PETS and UCSD datasets are popular and widely used in the most of the previous crowd system. TABLE. 1 shows the detail nature of crowded counting datasets.

To analyze which crowded datasets get the best performance, this paper shows the result that is used the methods of our previously published papers [42] and [43]. The papers [42] and [43] evaluated the accuracy performance of our work on the challenging PET 2009 dataset, UCSD dataset, Mall dataset, Gate dataset and Bridge dataset. As an experiment, MAE (Mean Absolute Error) and MRE (Mean Relative Error) is used to evaluate the performance of the various challenging crowds counting dataset nature. MAE can give clear insight that the predicted value is the same with the ground truth data. Otherwise, there exists higher miss rate. MRE is like as the relative square error. This is the rate of the comparative the predictor value with the actual value. The MAE performance analysis of the challenging datasets by using the method of [42] is shown in Fig. 4. And the MRE performance is also shown in Fig. 5.

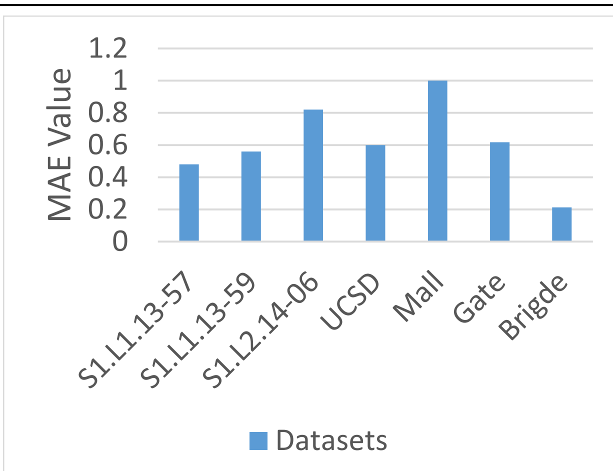


Fig. 4. Various crowded datasets analysis of MAE results [42]

The performance analysis of various challenging crowded counting datasets of MAE and MRE value by using the methods of [43] is shown in Fig. 6 and Fig. 7. Our previous

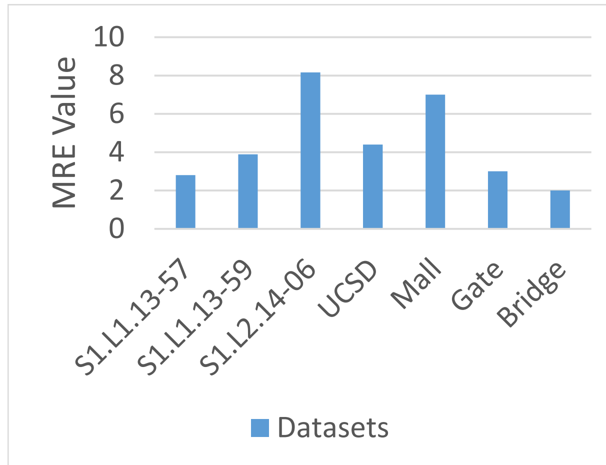


Fig. 5. Various crowded datasets analysis of MRE results [42]

published paper [43] proposed a new intuition Color Deep

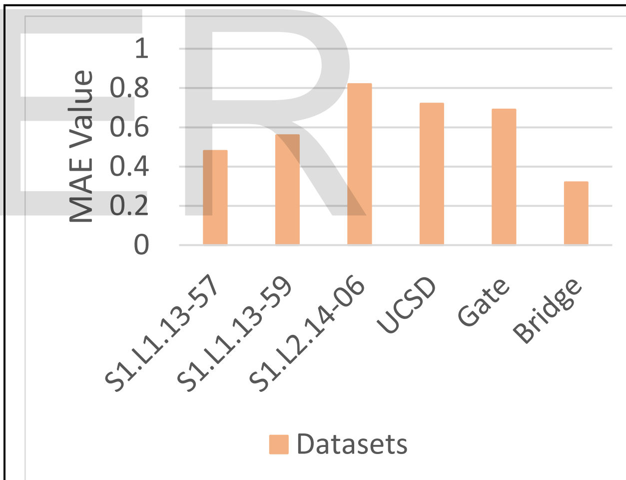


Fig. 6. Various crowded datasets analysis of MAE results [43]

system which utilizes based on the color-deep based feature for detecting and estimating the people numbers.

ISSN 2229-5518

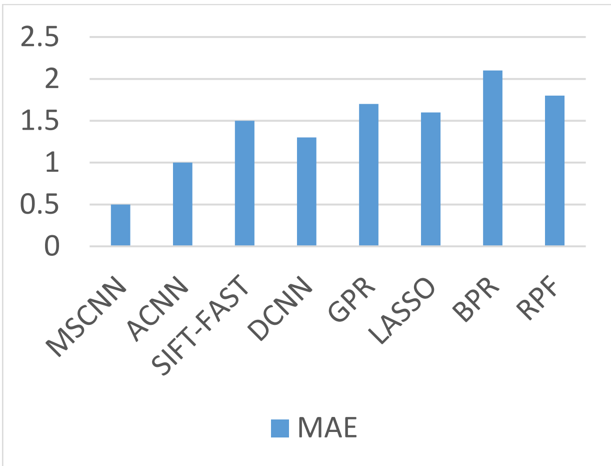


Fig. 8. Various evaluation performance tested only PET dataset

According to the experimental result of the two approaches [42] and [43] based on the challenging crowd counting datasets, Bridge dataset gets the lowest error rate on both MAE and MRE value. Because this dataset includes the lowest issues among the other rest six datasets. The highest error rate can clearly see in the Mall dataset.

Although the PET and UCSD have many challenges and widely used to develop the counting system, the Mall is greater issues than these two datasets. From the results in Fig. 4, 5, 6 and 7, it can be clearly seen that in all test videos, there is a tradeoff between the performance and the nature of the datasets (light crowd, medium crowd, heavy crowd, light challenges, medium challenges, heavy challenges, etc.). This paper aims to highlight this fact for many researchers who interest to propose the new system of estimating or tracking of population density and identifying specific traffic flows and specific situations of crowd events in the real world.

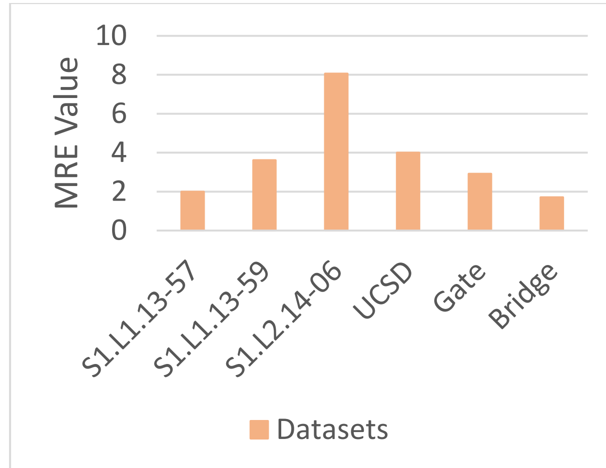


Fig. 7. Various crowded datasets analysis of MRE results [43]

To prove these highlighting points, focused only on PET 2009 dataset, there have many video sequences (S1.L1.13-57 set, S1.L1.13-59 set and S1.L2.14-06 sets). Among them, S1.L1.13-57 set has little issue and light crowd. This gets the lowest error rate in the three PET sequence sets. S1.L2 14-06 set has heavy issue and heavy crowd. Fig. 8 shows the evaluation performance of many previous works only on the PET dataset. In this figure, MSCNN means the mean subtraction CNN approach [42], ACNN means Adapting CNN approach, SIFT-FAST means the combining of SIFT and FAST detector, DCNN mean dynamic CNN approach, GPR mean Gaussian probability regression, LASSO means A least absolute shrinkage and selection operator, and BPR means Bayesian probability regression.

As a discussion, from the results in Fig. 4, 5, 6, 7 and 8, it can be clearly seen that the light weight crowd size and light challenge can achieve the highest performance (i.e. the lowest error rate).

## 5 Conclusion

The population of the world is emerging day per day. So, video surveillance works are becoming very essential in monitoring security. This paper discusses a lot of previous methodologies, algorithms, approaches or framework of people crowd counting systems. These methodologies and approaches are still necessary to develop a system to handle the issues of the massive crowd, illuminations, various variations and heavy occlusions on both static and dynamic crowd counting for all kinds of environments. This paper aims to highlight to develop a new system for less computational cost and efficient time with good performance for the researcher. The future work will involve the following steps for creating a novel people counting system. The first step is to use the best approaches for extracting background information from moving images. The second step is to describe the local or global features such as foreground pixels, relative height or width, crowd distribution, horizontal or vertical mean kinetic energy and crowd density are extracted for the person count. The feature descriptor like BRISK can be used for calculating the histogram and represented in a vector. After sampling, the features, the learning approaches like classifiers can be used for the extraction of the people count.

TABLE 1
Lists of Crowded Datasets

| Dataset | Yr. | Tot Fr. | Res: | C | P | Den: | CV |
|---|---|---|---|---|---|---|---|
| Mall | 2012 | 2000 | 640*480 | 3 | In | 15-53 | 1 |
| Grand | 2012 | 50010 | 720*480 | G | In | 250 | 1 |
| QUT:A | 2011 | 31000 | 704*576 | 3 | In | 3-23 | 3 |
| QUT:B | 2011 | 10000 | 352*288 | 3 | In | 3-23 | 3 |
| QUT:C | 2011 | 6100 | 352*288 | 3 | In | 3-23 | 3 |
| Fudan | 2011 | 1500 | 320*240 | G | Ot | 2-18 | 4-8 |
| PETs | 2009 | 4896 | 768*576 | 3 | Ot | 0-42 | 1 |
| UCSD | 2008 | 2000 | 238*158 | G | Ot | 11-46 | 1 |
| LIBRA | 2006 | 1000 | 640*480 | G | Ot | 20-50 | 1 |
| W.Xpo | 2010 | 1127 | 640*480 | 3 | Ot | 0-600 | 2 |
| CHUK | 2014 | - | 640*480 | 3 | Ot | 1-500 | 474 |
| Gate | 2014 | 1000 | - | G | Ot | 1-22 | 1 |
| Bridge | 2014 | 1000 | - | G | Ot | 6-30 | 1 |

*Statements that are brief in the above table are:*
*Yr. = year, Tot Fr. = Total Frames, Ree = Resolution, C = Color, P = Place that captured the scenes, Den: = Density, CV = Camera captured view, In = Indoor, and Ot = Outdoor.*

## References

[1] Arteta, C., Lempitsky, V., Zisserman, A., Counting in the wild, in: European Conference on Computer Vision, Springer, pp. 483–498, 2016.

[2] Antonini, Gianluca, and Jean Philippe Thiran, Counting pedestrians in video sequences using trajectory clustering. IEEE Transactions on 16 Circuits and Systems for Video Technology, no. 8, 1008-1020, 2006.

[3] A. Yilmaz, O. Javed, and M. Shah, Object tracking: A survey, New York, NY, USA, vol. 38, ACM, and Dec 2006.

[4] Benabbas, Y., Ihaddadene, N., Djeraba, C., Motion pattern extraction and event detection for automatic visual surveillance, EURASIP Journal on Image and Video Processing, 2011.

[5] Bernal, E.A., Li, Q., Loce, R.P., System and method for video based detection of drive-offs and walk-offs in vehicular and pedestrian queues, US Patent App. 14/279,652, 2014.

[6] Brostow, G.J., Cipolla, R., Unsupervised bayesian detection of independent motion in crowds, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), IEEE. pp. 594–60, 2006.

[7] Cong, Yang, Haifeng Gong, Song-Chun Zhu, and Yandong Tang, Flow mosaicking: Real-time pedestrian counting without scene-specific learning, Conference on Computer Vision and Pattern Recognition, (CVPR), IEEE, PP.1093-1100, 2009.

[8] D.M. Gavrila and V. Philomin, Real-Time Object Detection for Smart Vehicles, Proc. IEEE Int'l Conf. Computer Vision, pp. 87-93, 1999.

[9] D.M. Gavrila, A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pp. 1408-1421, Aug. 2007.

[10] Fehr, Duc, RavishankarSivalingam, Vassilios Morellas, Nikolaos Papanikolopoulos, Osama Lotfallah, and Youngchoon Park, Counting people in groups, Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS'09, IEEE PP.152-157, 2009.

[11] F. Castaldo, F. A. N. Palmieri, V. Bastani, L. Marcenaro, and C. S. Regazzoni, Abnormal vessel behavior detection in port areas based on dynamic bayesian

networks, in 17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, 2014, pp. 1–7, July 7-10, 2014.

[12] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., Object detection with discriminatively trained part-based models, IEEE transactions on pattern analysis and machine intelligence 32, 1627–1645, 2010.

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

[14] Li, M., Zhang, Z., Huang, K., Tan, T., Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection, in: Pattern Recognition, ICPR, 19th International Conference on, IEEE. pp. 1–4, 2008.

[15] Lin, S.F., Chen, J.Y., Chao, H.X., Estimation of number of people in crowded scenes using perspective transformation, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 31, 645–654, 2001.

[16] M. Enzweiler, P. Kanter, and D.M. Gavrila, Monocular Pedes-trian Recognition Using Motion Parallax, Proc. IEEE Intelligent Vehicles Symp., pp. 792-797, 2008.

[17] H. M. Dee, A. Caplier, Crowd behaviour analysis using histograms of motion direction, IEEE International Conference on Image Processing, 1545–1548, 2010.

[18] N. Dalal, Triggs, B., Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, (CVPR), IEEE Computer Society Conference on, IEEE. pp. 886–893, 2005.

[19] Rabaud, Vincent, and Serge Belongie, Counting crowded moving objects, in Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, vol. 1, PP. 705-711, 2006.

[20] Ryan, D., Denman, S., Fookes, C., Sridharan, S., Crowd counting using multiple local features, in: Digital Image Computing: Techniques and Applications, DICTA'09. IEEE. pp. 81–88, 2009.

[21] Sime, J.D., Crowd psychology and engineering, Safety science 21, 1-14, 1995.

[22] Wu. B., Nevatia, R., Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors, in: Computer Vision, ICCV 2005. Tenth IEEE International Conference on, IEEE. pp. 90–97, 2005.

[23] Wu, B., Nevatia, R., Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors, International Journal of Computer Vision 75, 247–266, 2007.

[24] W. Ge, R. T. Collins, R. B. Ruback, Vision-based analysis of small groups in pedestrian crowds, IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 1003–1016, 2012.

[25] Viola, P., Jones, M.J., Robust real-time face detection, International journal of computer vision 57, 137–154, 2004.

[26] Viola, P., Jones, M.J., Snow, D., Detecting pedestrians using pat-terns of motion and appearance, International Journal of Computer Vi-sion 63, 153–161, 2005.

[27] Z. Jin, B. Bhanu, Single camera multi-person tracking based on crowd simulation, International Conference on Pattern Recognition, 3660–3663, 2012.

[28] Zhao, T., Nevatia, R., Wu, B., Segmentation and tracking of multiple humans in crowded environments, IEEE transactions on pattern analysis and machine intelligence 30, 1198–1211, 2008.

[29] https://www.google.com/patents/US20130013534

[30] http://www.eecs.qmul.ac.uk/ccloy/dowloads_mall_ dataset.html

[31] http://www.ee.cuhk.edu.hk/xgwang/grandcentral.html

[32] https://www.wiki.qut.edu.au/display/saivt/SAIVT-QUT/CrowdþCountingþDatabase

[33] http://www.iipl.fudan.edu.cn/zhangjp/Dataset/fd_pede_dataset_intro.htm

[34] http://www.ftp.pets.rdg.ac.uk/pub/PETS2009/Crowd_PETS09_dataset/a_data/a.html

[35] http://www.svcl.ucsd.edu/projects/ peoplecnt/

[36] http://www.ee.cuhk.edu.hk/~jshao/CUHKcrowd_files/cuhk_crowd_dataset.htm

[37] http://www.ee.cuhk.edu.hk/~xgwang/expo.html

[38] Pham, V.Q., Kozakaya, T., Yamaguchi, O., Okada, R., 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, pp. 3253–3261, 2015.

[39] K. Katevas, H. Haddadi, and l. Tokarchuk, Poster: Sensingkit: A multi-platform mobile sensing framework for large-scale experiments, in Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, New York, NY, USA, 2014, MobiCom '14, pp. 375–378, ACM.

[40] K. R. Kiran, M. Musolesi, C. Mascolo, P.J. Rentfrow, C. Longworth, and A. Aucinas, Emotionsense: A mobile phones based adaptive platform for experimental social psychology research, in Proceedings of the 12th ACM International Conference on Ubiquitous Computing, New York, NY, USA, 2010, UbiComp '10, pp. 281–290, ACM.

[41] M. Hashemzadeh, G. Pan, M. Yao, "Counting moving people in crowds using motion statistics of feature-points", in: Multimed Tools Application, 2014, pp. 453-475.

[42] H.H. Lin and K.T. Win, "People counting with extended convolutional neural network", in 27th International Conference on Computer Theory and Applications (ICCTA), Alexandria, Egypt, 2017, pp. 99-104, ISBN 978-1-5386-3794-4.

[43] H.H. Lin and K.T. Win, "People counting system with C-Deep feature in dense crowd views", in 17th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2018), Singapore, 2018, pp. 99-104, ISBN 978-1-5386-5892-5.