ISBN 978-93-84422-71-4
Proceedings of 2016 International Conference on Image Processing, Production and Computer Science (ICIPCS'2016)
London (UK), March 26-27, 2016 pp.30-36

Cascade Method for Image Processing Based People Detection and Counting

Zeyad Al-Zaydi, David Ndzi and David Sanders

School of Engineering, University of Portsmouth, Anglesea Road, Portsmouth PO1 3DJ, UK.

Abstract: *People detection is of great importance in video surveillance. Different approaches have been proposed to achieve accurate detection system. The main problem in people detection systems is that it must maintain a balance between the number of false detections and the number of missing people which limits the global detection results. In order to solve this problem and add robustness to detection, we propose a multiplexor and collector model composed of multiple independent detectors. This model is used to keep the true positive detections provided by a number of detectors and reduce the miss rate. In addition, a fusion model is proposed to check the robustness of the cascaded detection system. A pipeline techniques will also be used to avoid the increasing of detection time.*

Keywords: *people detection, counting, surveillance systems, image processing, computer vision.*

1. Introduction

People detection is one of the most challenging task in computer vision [1]. Although significant research has been carried out to find an accurate solution for this task, there are still many challenges that need to be resolved. These include variability in appearances, crowded scenarios, handling complex backgrounds and occlusion which lead to high false detection and miss rate. The trade-off between false detection and miss rate renders most methods ineffective.

People detection is fundamental in intelligent video surveillance systems as it provides important information for establishing awareness. People detection can be used for people counting and tracking. An efficient and accurate people counting, tracking and distribution system would be beneficial and fundamental to a lot of applications such as in

- safety applications; e.g. an indicator of over-crowded situations, for possible emergency evacuation processes and crowd management [2], [3];
- security applications; e.g. an indicator of fighting, rioting, violent protest, mass panic and excitement [4], [5];
- business intelligence and behavioural economics applications; e.g. the distribution of costumers may be used for product placement, floor planning and staff management [6]. In addition, the overall crowd in a retail store may be monitored to assess store performance over time [7];
- Transport applications; e.g. to improve the distribution of buses over different routes which, is fundamental to the optimisation of transport network and for scheduling public transport [8], [9]; and
- energy management applications such as optimising air conditioning, lighting and heating in buildings according to occupancy density and distributions [3], [10].

http://dx.doi.org/10.17758/UR.U0316014

This paper proposes firstly, a multi stage independent detection system employed to minimise miss rate; secondly a fusion technique that can be used to compensate the limitations of each independent detector and finally, a pipeline techniques to that minimises processing, and hence detection, time.

The remainder of this paper is structured as follows: Section 2 describes the related work; Sections 3 describes the system design; Finally, Section 4 summarizes the main conclusions and future work.

2. Related Work

There is an extensive literature on people detection approaches. They can be generally grouped into six paradigms. In this section, we provide an overview on each of the paradigms.

2.1. Full Body Detection Based Algorithms

They are direct approaches to count the number of people in a scene through detection. The algorithms are trained using the full body appearance of a set of people [11]–[13]. They suffer from large pose variations and partial occlusion as the number of people increases [14]. Different features are used to represent the full body appearance such as Haar like features [15] and histogram of oriented gradient (HOG) features [13]. Different linear and nonlinear classifiers are also used to find the relationship between the features and the number of people such as linear support vector machine (SVM), neural network (NN) and Gaussian process regression (GPR) [16], [17]. The accuracy of full body detectors is acceptable in sparse environments but the accuracy decreases significantly in crowded environments.

2.2. Part Body Detection Based Algorithms

A significant amount of research has been carried out to mitigate partial occlusion by detecting only part of the body such as heads, faces, eyes and head-shoulders [18]–[20]. The shape of people's heads changes or differ with hair styles and head coverings. Hence head based detection is not robust enough for counting people [14]. On the other hand, a head-shoulder region occupies a larger proportion of a human body image than a head alone and they are more likely to be detected [14]. Faces and eyes are rarely used to count the number of people because a lot of people do not look at the cameras when passing and faces and eyes are easily occluded.

2.3. Shape Matching Detection Based Algorithms

Ellipses are used by some researchers to count the number of people [21]. In this approach, the background subtraction method is applied to segment the foreground blobs [22], [23] and ellipse detection is applied to identify the number of people in each blob. Other shapes such as Bernoulli shapes have been used by some researchers to count the number of people [24]. The accuracy of shape matching detectors is acceptable in sparsely occupied environments but the accuracy decreases significantly in crowded environments.

2.4. Multi Camera Detection Based Algorithms

Many research studies have focused on counting the number of people using a single camera which can fail in crowded environments (i.e., heavy occlusion occurs). Some researchers have used multiple cameras to count people to avoid occlusion [25]. The cost of hardware and the multi-camera set-up is the main disadvantages of this approach.

2.5. 3D camera detection based algorithms

Depth information has great potential to improve people counting for many reasons e. g. [26], it can be used to;

- Improve foreground segmentation; and
- Handle occlusions more efficiently.

This third dimension allows time-of-flight (TOF) and stereo vision features to be used to obtain image depth [27]. For instance, Microsoft Kinect devices could be used to obtain image depth, which provides high quality images at a lower price compared to previous technologies [28]. The performance of 3D camera based detection techniques is affected by changing illumination and when monitoring a large area of similar colors and little edges, because it may be difficult to find features [29]. In addition, developing a people counting algorithm based on a depth sensing system is more complex and would therefore require a significant amount of computational time [30].

2.6. Density-Aware Detection Based Algorithms

This approach combines full or part body detection based algorithms and crowd density estimation [31]. Full body, head and head-shoulder detection based algorithm can be improved and the accuracy can be increased by using density-aware information [31]. The aim of this approach is to reduce the false positive per image (FPPI) in low crowd density locations in the frame. This occurs when it incorrectly detects the presence of a person, when there is actually nobody. In addition, this approach reduces the miss rate in high crowd density locations in the frame.

3. System Design

The main aim of classical cascade classifiers is to reduce the FPPI. In this approach FPPI will be improved, obviously, but the miss rate of detection (FNR) will increase rapidly. The FNR is measured using [13]:

$$FNR = \frac{FN}{FN+TP} \quad (1)$$

Where FN is the number of times it incorrectly indicates that there is nobody, TP is the number of correct detections. In addition, a classical cascade classifier usually consists of multiple stages of weak classifiers of the same kind but do not use multiple independent detectors. As shown in Figure 1, all detected windows of Non-Person (NP) are rejected directly at each classifier while the windows of Candidate Person (CP) are passed to the next stage for more checking by the cascaded classifiers.

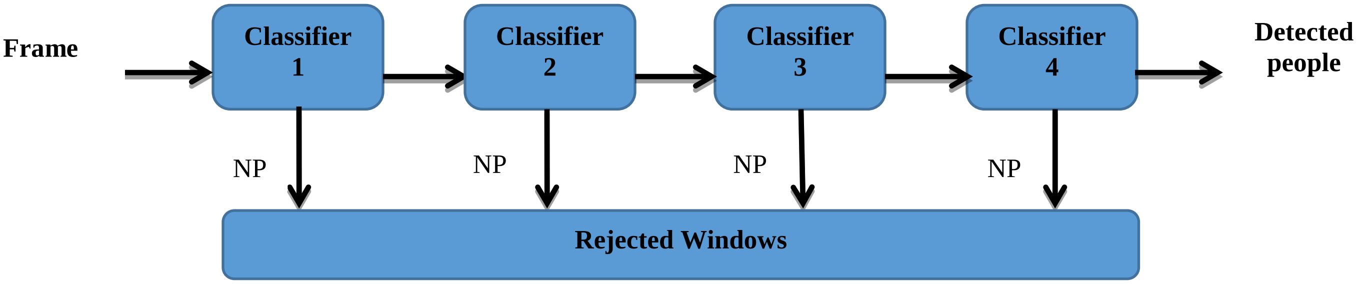


Fig. 1: Flow diagram of a cascade classifier.

In this paper, multiple independent detectors and a novel multiplexer cascade is proposed. Different detectors are used to detect people. Each one has some advantages and disadvantages mainly because each technique is based on different features extraction, learning method and person models. Frame-by-frame, different independent detectors may produce different results so a novel cascade of different independent detectors can be implemented to improve the true positive detections provided by a number of detectors. That will reduce the FNR especially when the advantages of each detector is exploited in this cascaded model. In addition, the rejected windows from all detectors can be fused and compared with a predefined threshold for detection purposes. Figure 2 shows the block diagram of the proposed method which consists of three

http://dx.doi.org/10.17758/UR.U0316014

independent detectors and one fusion model. Where CP is the windows of candidate person, LCP is the windows of low-level candidate person, DP is the windows of detected people and RP is the windows of rejected people. Two fusion models are developed in this paper.

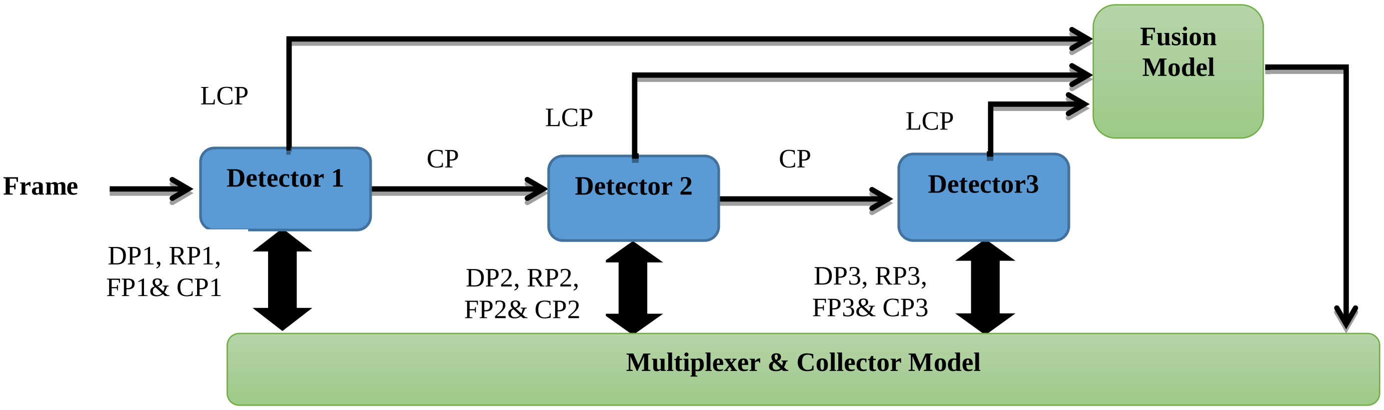


Fig. 2: Flow diagram of a novel multiplexer cascade detector.

The confidence level of the detectors will be used to classify windows into detected person windows, rejected person windows and candidate person windows. In addition, three predefined threshold will be used in this classification; high-quality, medium-quality and low-quality thresholds. The multiplexer and collector model will use the following rules to classify the windows;

$\boldsymbol{if}\ (CFW > High\ quality\ threshold)\ \boldsymbol{Then}\ DP$

$\boldsymbol{if}\ (CFW < High\ quality\ threshold)\ \boldsymbol{Then}\ RP$

$\boldsymbol{if}(CFW < High\ quality\ threshold\ \boldsymbol{and}\ > Low\ quality\ threshold)\ \boldsymbol{Then}\ CP$

$\boldsymbol{if}(CFW < High\ quality\ threshold\ \boldsymbol{and}\ > miduim\ quality\ threshold)\ \boldsymbol{Then}\ LCP$

where CFW is the confidence level of a window. The multiplexer and detection windows model will use the following equation to collect the results of all detectors and the fusion model;

$$Total\ detected\ people = DP1 + DP2 + DP3 + DP4 \qquad (2)$$

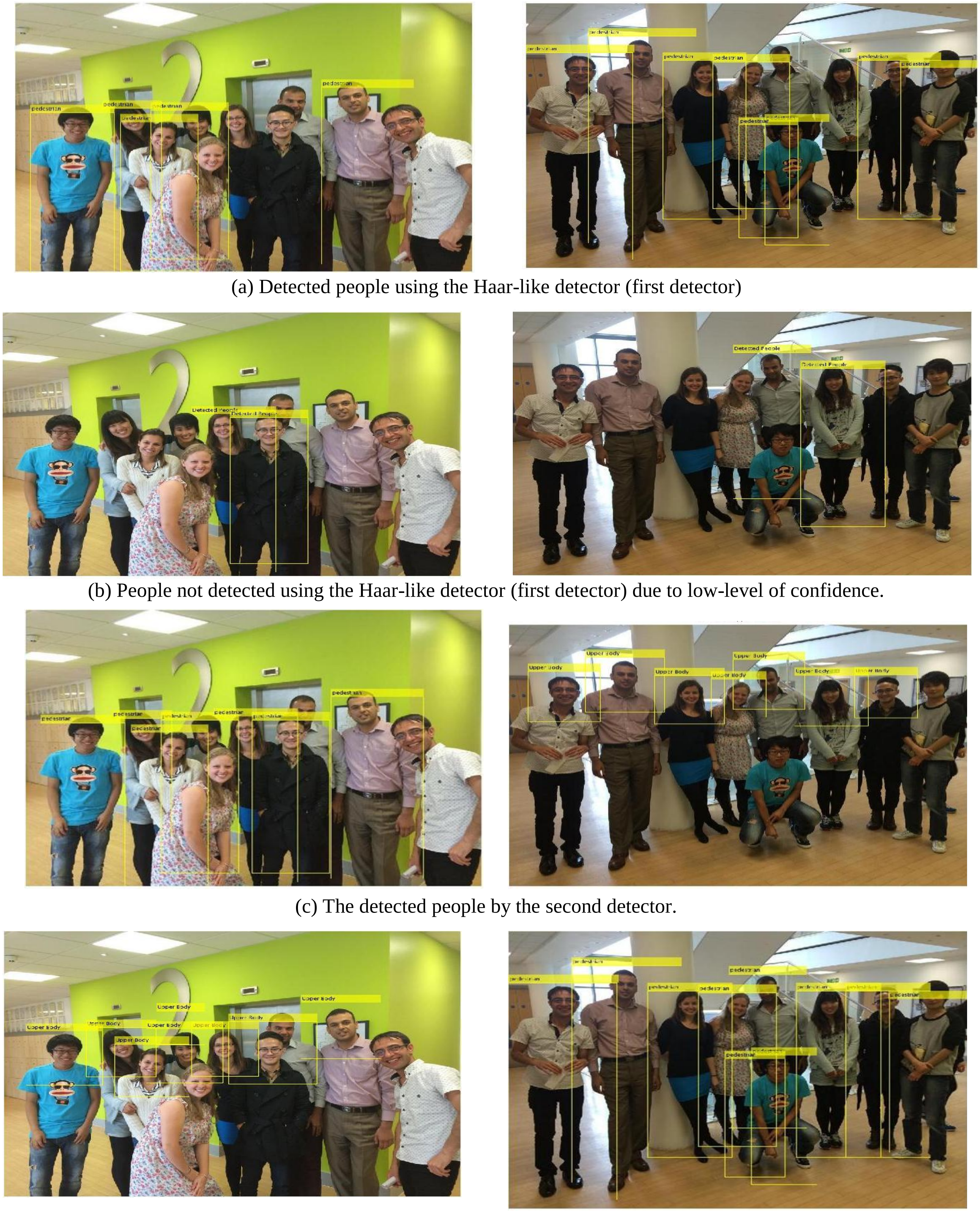
The fusion model measures the robustness of the LCP windows by comparing them with the medium-quality threshold. All LCP windows that get more than this threshold at all detectors will be considered as DP windows. The fusion model will use the following rule;

$\boldsymbol{if}\ (LCP1\ \boldsymbol{and}\ LCP2\ \boldsymbol{and}\ LCP3 > miduim\ quality\ threshold\ )\ \boldsymbol{Then}\ DP$

The proposed method will use pipeline techniques to avoid increasing the detection or processing time. It is an implementation technique in which multiple frames are overlapped during execution. Three frames are processed simultaneously using different detectors. In this case, the processing time is not increased.

4. Experimental Results

Matlab software is used to implement the proposed system. Preliminary results of the proposed method have been obtained. The proposed approach is implemented and tested using pictures. Two detectors are used in to produce the results presented in this section; the Haar-like detector, which is very famous and widely used in detection systems and the second one is a full body detector. As shown in Figure 3, the miss rate of the proposed approach is 0.3 and 0 for the first and second pictures. From the results, it can be seen that the miss rate is lower than each detector individually.



(a) Detected people using the Haar-like detector (first detector)

(b) People not detected using the Haar-like detector (first detector) due to low-level of confidence.

(c) The detected people by the second detector.

(d) The detected people by the proposed approach.

Fig 3: The performance of the proposed approach.

5. Conclusion

This paper has outlined the general principles of a new approach for people detection and counting using video surveillance. This approach combines multiple independent detectors using a multiplexer & collector model, fusion model and a pipeline technique in order to keep the correct positive detections by a number of detectors and, at the same time, reduce the miss rate. This integration technique performs better, provides more accurate results and enhances the detection rate. The proposed approach has been implemented and tested using frames as well as pictures. Initial results are very promising and shows that the proposed approach reduces the miss detection rate and hence, improves accuracy. Further development and evaluation of the proposed technique using different videos in different environments with different crowd densities in real-time environments is currently being carried out.

6. References

[1] S. Mukherjee and K. Das, "Omega Model for Human Detection and Counting for application in Smart Surveillance System," arXiv Prepr. arXiv1303.0633, vol. 4, no. 2, pp. 167–172, 2013.

[2] A. Technology, "Our customers," 2013. [Online]. Available: www.peoplecounting.co.uk/our-customers. [Accessed: 23-Mar-2015].

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

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

http://dx.doi.org/10.1016/j.patrec.2013.10.002

[5] Z. Zhang, M. Wang, and X. Geng, "Crowd counting in public video surveillance by label distribution learning," Neurocomputing, vol. 166, pp. 151–163, 2015.

[6] C. Loy, K. Chen, S. Gong, and T. Xiang, Crowd counting and profiling: Methodology and evaluation. 2013.

[7] D. A. Ryan, "Crowd Monitoring Using Computer Vision," Queensland University of Technology, 2013.

[8] J. Zhang, B. Tan, F. Sha, and L. He, "Predicting pedestrian counts in crowded scenes with rich and high-dimensional features," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1037–1046, 2011

http://dx.doi.org/10.1109/TITS.2011.2132759

[9] A. Bartolini, F., Cappellini, V., & Mecocci, "Counting people getting in and out of a bus by real-time image-sequence processing," Image Vis. Comput., pp. 36–41, 1994.

http://dx.doi.org/10.1016/0262-8856(94)90053-1

[10] K. Hashimoto, K. Morinaka, N. Yoshiike, C. Kawaguchi, and S. Matsueda, "People count system using multi-sensing application," Proc. Int. Solid State Sensors Actuators Conf. (Transducers '97), vol. 2, pp. 1291–1294, 1997.

http://dx.doi.org/10.1109/SENSOR.1997.635472

[11] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian Detection via Classification on Riemannian Manifolds," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 10, 2008.

http://dx.doi.org/10.1109/TPAMI.2008.75

[12] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," Comput. Vis. Pattern Recognition, 2005. CVPR 2005. IEEE Comput. Soc. Conf., vol. 1, pp. 878–885, 2005.

http://dx.doi.org/10.1109/cvpr.2005.272

[13] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," CVPR '05 Proc. 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 1, pp. 886–893, 2005.

http://dx.doi.org/10.1109/cvpr.2005.177

[14] J. S. C. Yuk, K. Y. K. Wong, R. H. Y. Chung, F. Y. L. Chin, and K. P. Chow, "Real-time multiple head shape detection and tracking system with decentralized trackers," Proc. - ISDA 2006 Sixth Int. Conf. Intell. Syst. Des. Appl., vol. 2, pp. 384–389, 2006.

[15] P. Viola and M. Jones, “Robust real-time face detection,” Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, 2004. http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb

[16] P. Viola, M. J. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” Int. J. Comput. Vis., vol. 63, no. 2, pp. 153–161, 2005.

http://dx.doi.org/10.1007/s11263-005-6644-8

[17] 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.

http://dx.doi.org/10.1016/j.cviu.2014.07.008

[18] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object Detection with Discriminative Trained Part Based Models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, 2010.

http://dx.doi.org/10.1109/TPAMI.2009.167

[19] S. Lin, J. Chen, and H. Chao, “Estimation of Number of People in Crowded Scenes Using Perspective Transformation,” IEEE Trans. Syst. Man. Cybern., vol. 31, no. 6, pp. 645–654, 2001.

http://dx.doi.org/10.1109/3468.983420

[20] B. Wu and R. Nevatia, “Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors,” Int. J. Comput. Vis., vol. 75, no. 2, pp. 247–266, 2007.

http://dx.doi.org/10.1007/s11263-006-0027-7

[21] J. Li, L. Huang, and C. Liu, “Robust people counting in video surveillance: Dataset and system,” 2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011, pp. 54–59, 2011.

http://dx.doi.org/10.1109/avss.2011.6027294

[22] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, “Real-time foreground-background segmentation using codebook model,” Real-Time Imaging, vol. 11, no. 3, pp. 172–185, 2005.

http://dx.doi.org/10.1016/j.rti.2004.12.004

[23] A. Ilyas, M. Scuturici, and S. Miguet, “Real time foreground-background segmentation using a modified codebook model,” 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2009, pp. 454–459, 2009.

http://dx.doi.org/10.1109/avss.2009.85

[24] W. Ge and R. T. Collins, “Marked point processes for crowd counting,” 2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009, pp. 2913–2920, 2009.

[25] H. Ma, C. Zeng, and C. X. Ling, “A Reliable People Counting System via Multiple Cameras,” ACM Trans. Intell. Syst. Technol., vol. 3, no. 2, pp. 1–22, 2012.

http://dx.doi.org/10.1145/2089094.2089107

[26] M. Harville, “Stereo person tracking with adaptive plan-view statistical templates,” Proc. ECCV Work. Stat. Methods Video Process., pp. 67–72, 2002.

[27] V. Gandhi, J. Čech, and R. Horaud, “High-resolution depth maps based on TOF-stereo fusion,” Proc. - IEEE Int. Conf. Robot. Autom., pp. 4742–4749, 2012.

http://dx.doi.org/10.1109/icra.2012.6224771

[28] Microsoft, “Kinect,” 2011. [Online]. Available: www.xbox.com/en-US/kinect/. [Accessed: 23-Mar-2015].

[29] C. R. Sensors, R. B. Fisher, and K. Konolige, “Handbook of Robotics Chapter 22 - Range Sensors,” 2008.

[30] T. Tikkanen, “People detection and tracking using a network of low-cost depth cameras,” 2014.

[31] M. Rodriguez, E. N. Superieure, I. Laptev, J. Sivic, and J.-Y. Audibert, “Density-aware person detection and tracking in crowds,” Comput. Vis. (ICCV),IEEE, pp. 2423–2430, 2011

http://dx.doi.org/10.1109/iccv.2011.6126526