

PASSENGER COUNTING IN PUBLIC RAIL TRANSPORT

Using Head-Shoulder Contour Tracking

Pieterjan De Potter¹, Philippe Belet¹, Chris Poppe¹,
Steven Verstockt^{1,2}, Peter Lambert¹ and Rik Van de Walle¹

¹*Department of Electronics and Information Systems - Multimedia Lab, Ghent University - IBBT,
Gaston Crommenlaan 8 bus 201, B-9050, Ledeberg-Ghent, Belgium*

²*ELIT Lab, University College West Flanders, Ghent University Association,
Graaf Karel de Goedelaan 5, B-8500, Kortrijk, Belgium*

Keywords: Video Analytics, Public Transport, People Counting, Histogram of Oriented Gradients, Kalman Filter.

Abstract: Automated people counting has multiple applications: referring passengers to vehicles with empty seats, gathering statistical information for railway companies to improve their distribution of vehicles, etc. In this paper, a people counting algorithm for public transport vehicles is presented. First, head-shoulder contours are detected by adaboost classification of a combination of a histogram of oriented gradients features and a color histogram. An integral histogram and integral image are used to speed up the extraction of these features. The results of the classification process are clustered and these clusters are tracked by a Kalman filter using a custom error covariance matrix. Finally, the path followed by an observed person is evaluated in order to count passengers entering and exiting the vehicle. Evaluation shows that this approach performs better than previous approaches, especially in scenarios with occlusions.

1 INTRODUCTION

Over the past decade, the number of installed video surveillance cameras has grown exponentially because of the reduced cost and the fact that for some scenarios security has gained importance over privacy. This has led to the development of various video analytic systems to detect different events, mostly outdoors or in large open spaces.

In public transport as well, video surveillance cameras are being installed, and video analytic systems can be used. The primary goal of the installed cameras is to provide security, but as the cameras are already installed, they can also be used for other purposes, such as passenger counting. The conditions in vehicles are however different than in other scenarios (e.g., fast illumination changes, a lot of occlusion, a moving background through the windows), which makes modified or new algorithms necessary.

The remainder of this paper is organized as follows. In Section 2, related work is discussed. Our system is described in Section 3. In section 4, an evaluation of our approach is given. Finally, conclusions and future work are discussed in Section 5.

2 RELATED WORK

In (Vu et al., 2006), an event recognition system based on face detection and tracking combined with audio analysis is presented. Zones of interest and static objects are used as context information. The focus is on audio-video based event detection.

High accuracies have been reported counting passengers using a dedicated setup with vertically directed cameras (Yahiaoui et al., 2008). Since the cameras used for this setup can not be used for other purposes, this solution is more expensive than reusing already installed video surveillance cameras.

Regarding human and object detection in general, different feature descriptors have already been proposed. In (Li et al., 2009), people are detected based on the omega-shape features of their head-shoulder parts. Viola-Jones classification and AdaBoost classification using histogram of oriented gradients features are combined to obtain fast and reliable results. In our system, parts of this approach are adapted and used.

In previous work, we counted passengers by combining Laplacian edge detection with a median-based background subtraction technique to detect objects (De Potter et al., 2010a). Rectangle-shaped regions

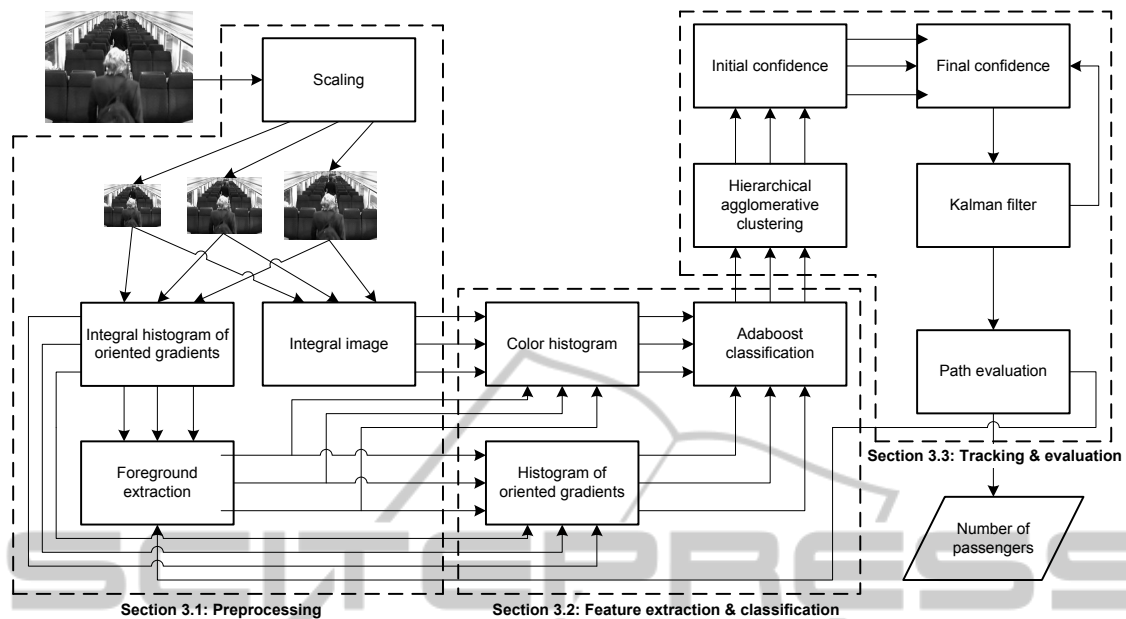


Figure 1: The general architecture of our system. The input images are first preprocessed, after which features can be extracted to classify persons. These persons are tracked and their paths are evaluated to obtain a right passenger count.

were defined in the seat regions to detect seating and leaving actions. Drawbacks were the need for a manual calibration of the seat regions and the need for a training phase for the used background subtraction technique. We also compared a Laplacian of Gaussian edge detector, a non-linear difference of Gaussians edge detector and a mixture of models background subtraction technique (De Potter et al., 2010b). The results of these techniques were used by a bounding box based tracker to count passengers in a vehicle. The main disadvantage is that the bounding box tracker is based on the detection of blobs: occlusions in the camera view hinder correct detections.

In this paper, we counter the occlusion problem by counting people when they are best visible: at the entrances of each vehicle. Once the classification unit is well trained, the manual calibration is limited to the definition of the entry zone.

3 SYSTEM OVERVIEW

Figure 1 gives an overview of our system. Different preprocessing steps are described in Section 3.1. The extraction of the histogram of oriented gradients (HOG) features and the color histogram, and the classification of persons is described in Section 3.2. The persons are tracked and the paths are evaluated to obtain the passenger count, as discussed in Section 3.3.

3.1 Preprocessing

Since HOG features are not scale-invariant, the input images are rescaled to three smaller dimensions to enable the detection of head-shoulder contours closer and further away from the camera.

A lot of HOGs will need to be calculated for the head-shoulder detection. For this reason, an integral histogram (Porikli, 2005) is used on each scale to store the magnitude and orientation of the gradients. At each pixel, the magnitude of the gradient is calculated for each color channel. Only the color channel leading to the largest magnitude is used in the further calculations (Dalal and Triggs, 2005). For this color channel, the orientation of the gradient is calculated.

Color histograms are used in addition to the HOG features for classification and also need to be calculated many times. In order to cope with illumination variances, a normalized color space is used to construct these histograms. Since the transformation to this color space uses mean values and standard deviations, an integral image (Viola and Jones, 2004) is constructed for each $K \in [R^2 G^2 B^2]$. Using this integral image, the normalized color histograms can be calculated much faster.

Classification is the slowest step in our system. In order to limit the number of partial images for which the features need to be extracted and classified, the positions of foreground objects in the input images are calculated. In order to cope with illumination changes in the input images, the foreground detection is edge-

based. In addition to these detected foreground regions, regions that are predicted as foreground regions by the person tracker (see Section 3.3) are indicated as foreground regions as well.

3.2 Feature Extraction and Classification

There is a lot of occlusion in vehicles for public transport. In our approach, the parts of the body that are visible most of the time are used to detect persons: the head and shoulders. HOG features are used to detect these head-shoulder contours in 32x32 pixels partial images (Li et al., 2009). In addition, a normalized color histogram is calculated per partial image.

The adaboost boosting algorithm (Freund and Schapire, 1997) is used to create a robust classifier out of multiple weak classification units. Some results of the classification are shown in Figure 3.2.



Figure 2: The input image at the largest scale and the results of the Adaboost classification at all three scales.

3.3 Tracking and Counting

The results of the classification are clustered using single linked agglomerative clustering (Manning et al., 2008). This way, the number of clusters does not need to be known in advance.

For each cluster, the area of the minimal enclosing ellipse and the ratio of the number of actual points versus holes are taken into account to obtain the initial confidence in that cluster. The initial confidence and the centers of the minimal enclosing ellipse of the different clusters on the different scales are scaled to the smallest scale. This enables an equal comparison of clusters over the different scales. For each cluster, corresponding clusters on all scales are searched and the maximal initial confidence of these corresponding clusters is used for further calculations.

The final confidence is represented by an ellipse, with a smaller ellipse representing a greater confidence. The minimal enclosing ellipse corresponding

to the cluster with the maximal initial confidence is first rescaled so its area matches a predefined area. Then it is scaled to represent the number of points in that cluster, relative to a predefined number of points. The generalized conjunction / disjunction function (Dujmović, 1996) is used to take other factors (i.e. the proximity of the centers of the enclosing ellipses on the different scales, the initial confidence, and the proximity of the center the predicted position by the Kalman filter) into account as well, resulting in a second scaling factor. The final confidence is obtained after scaling with both scaling factors. From this ellipse, the corresponding covariance matrix is retrieved. This covariance matrix is used as the measurement error covariance matrix in the tracking stage.

A Kalman filter (Welch and Bishop, 1995) is used to track clusters over time. The center of the final confidence ellipse is used as measured position of the person; the corresponding covariance matrix is used as measurement for the prediction error.

During the tracking phase, path information is collected. This information then evaluated to count passengers. Following variables are taken into account during this evaluation: the start- and endpoint of the path, the total distance over which a person was tracked and the average confidence during the tracking. By taking the total distance and the average confidence into consideration, some paths that result from false positives in the person classification process can be eliminated. The evaluation of the start- and endpoint of the path gives an indication of the action (entering/exiting) that has taken place.

4 EVALUATION

We evaluated our system on six acted sequences with increasing difficulty in terms of occlusion.

In Table 1, the results from our approach described in this paper and the results of previous approaches are listed. In (De Potter et al., 2010b), the entire carriage is processed by each camera. The results for this approach are the averages of the results of both cameras.

As can be seen in Table 1, our new approach deals better with scenarios where occlusions happen, especially when persons are exiting the vehicle. Improvements are still possible, mostly in the area of tracking and track evaluation.

Table 1: Performance evaluation: the number of persons entering and exiting the vehicle that are counted correctly (CC), too much (CP) and too less (CM) are listed. Average values are used for (De Potter et al., 2010b).

	Sequence						Total
	1	2	3	4	5	6	
<i>Ground truth</i>	10	10	10	12	14	14	70
<i>(De Potter et al., 2010a)</i>							
CC	5	9	7	6	7	7	40
CP	0	1	4	2	1	0	8
CM	5	1	3	6	7	7	30
<i>(De Potter et al., 2010b)</i>							
CC	7	4	2.5	1	0	0	14.5
CP	0	0	0	0	0	0	0
CM	3	6	7.5	11	14	14	55.5
<i>Current system</i>							
CC	10	7	9	11	13	11	61
CP	0	0	1	0	1	0	2
CM	0	3	1	1	1	3	9

5 CONCLUSIONS

In this paper, a people counting algorithm for vehicles in public transport is described. It uses head-shoulder detection by adaboost classification of histograms of oriented gradients and color histograms to detect people. A Kalman filter is used to track these people, after which the paths of the observed people are evaluated in order to count them. The evaluation shows that this approach works better than previous approaches, especially in scenarios with occlusions.

Future work consists of evaluating the use of a particle filter (Isard and Blake, 1998) instead of the Kalman filter, the use of an attentional cascade (Viola and Jones, 2004) instead of the adaboost classifier to decrease the computation time, and the use of face detection to improve the results for exiting passenger counting. More test videos need to be recorded to obtain better training sets and to make a more realistic evaluation possible.

ACKNOWLEDGEMENTS

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

REFERENCES

- Dalal, N. and Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 886–893.
- De Potter, P., Billiet, C., Poppe, C., Stubbe, B., Verstockt, S., Lambert, P., and Van de Walle, R. (2010a). Available Seat Counting in Public Rail Transport. In *Progress in Electromagnetics Research Symposium*, pages 1294–1298.
- De Potter, P., Kypraios, I., Verstockt, S., Poppe, C., and Van de Walle, R. (2010b). Automatic Available Seat Counting In Public Rail Transport Using Wavelets. In *Electronics in Marine*, pages 79–83.
- Dujmović, J. J. (1996). A Method For Evaluation And Selection Of Complex Hardware And Software Systems. In *Computer Measurement Group Conference*, pages 368–378.
- Freund, Y. and Schapire, R. (1997). A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119–139.
- Isard, M. and Blake, A. (1998). CONDENSATION - Conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28.
- Li, M., Zhang, Z., Huang, K., and Tan, T. (2009). Rapid and Robust Human Detection and Tracking Based on Omega-Shape Features. In *IEEE International Conference on Image Processing*, pages 2545–2548.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*, pages 378–382. Cambridge University Press.
- Porikli, F. (2005). Integral Histogram: a Fast Way to Extract Histograms in Cartesian Spaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol. 1*, pages 829–836.
- Viola, P. and Jones, M. (2004). Robust Real-time Face Detection. *International Journal of Computer Vision*, 57(2):137–154.
- Vu, V.-T., Bremond, F., Davini, G., Thonnat, M., Pham, Q.-C., Allezard, N., Sayd, P., Rouas, J.-L., Ambellouis, S., and Flancquart, A. (2006). Audio-video Event Recognition System for Public Transport Security. In *IET Conference on Crime and Security*, page 6 pp.
- Welch, G. and Bishop, G. (1995). An Introduction to the Kalman Filter. Technical report, University of North Carolina at Chapel Hill.
- Yahiaoui, T., Meurie, C., Khoudour, L., and Cabestaing, F. (2008). A people Counting System Based on Dense and Close Stereovision. In *International Conference on Image and Signal Processing*, pages 59–66.