Combination of Fixed and Mobile Cameras for Automatic Pedestrian Detection

A. Alahi *   M. Bierlaire *   M. Kunt†

October 5, 2008

Report TRANSP-OR 081005
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*TRANSP-OR, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, {alexandre.alahi, michel.bierlaire}@epfl.ch

†LTS, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, murat.kunt@epfl.ch

Abstract

Pedestrian detection in the surroundings of a vehicle is highly desirable to avoid dangerous traffic situations. Typical vision-based pedestrian detection algorithms on mobile cameras suffer from the lack of a-priori knowledge on the object to be detected. The variability in the shape, pose, color distribution, and behavior affect the robustness of the detection process.

A novel vision-based system is proposed to detect pedestrians with a single mobile camera collaborating with a fixed camera observing the same scene. Nowadays, a large number of fixed cameras are installed in major cities. This work presents how features extracted from those fixed cameras can be used to detect pedestrians with mobile cameras present in the same scene. The proposed system outperforms state-of-the-art single frame pedestrian detectors using a feature-based classification framework. In addition, the system can be generalized to any object of interest. Any object detected by a fixed camera, can be detected with a mobile camera.

1 Introduction

Governmental agencies, car manufacturers, and many institutes are interested in detecting pedestrians in the surroundings of a vehicle to avoid dangerous traffic situations. In the EU, there are more than 150 000 pedestrians injured yearly and more than 8000 are killed (numbers for 2004, *Statistics of Road Traffic Accidents in Europe and North America*, 2007). Many accidents originate from the momentary distraction of the drivers. A driver assistance system that detects potential collisions with pedestrians will reduce the number of pedestrians killed on roads.

Over the past decades, progress in image and video processing algorithms has encouraged researchers to apply those techniques to transportation problems. Ali and Dagless (1990) presented a fully automatic system to detect and track pedestrians and vehicles. More recently, Velastin et al. (2006) used optical flow to avoid potentially dangerous situations involving pedestrians in public transport. However, their system was based on fixed cameras monitoring a scene. In this paper, pedestrians present in front of a moving vehicle are of interest. Detection should occur with the cameras mounted in the vehicles.

Expensive vision-based systems based on stereo cameras exist such as those proposed by Broggi et al. (2000), Zhao and Thorpe (2000) or more recently by Bota and Nedesvchi (2008). Suard et al. (2006) combined stereo cameras with other sensors. Although most of the accidents happen in daytime conditions, Far Infra-Red (FIR) thermal images have been used to detect pedestrians at night time by Bertozzi et al. (2006), and Suard et al. (2006). FIR systems exploits the radiation emitted by any warm object. Warm objects are bright whereas the rest is dark. Nonetheless, since FIR images depend on the temperature of the objects, an outdoor scene has a number of factors that affect the images. Strong sun heating can introduce texture due to different thermal behavior of different materials. Moreover, the temperature variation (due to cloud, humidity,etc.) make the road scene difficult to interpret. Near Infra-Red (NIR) sensors detect the radiation reflected by objects in the infra-red range which is close to visible light. NIR images have a higher spatial resolution than images formed by FIR. Broggi et al. (2006)

evaluate the pedestrian body and legs using geometrical moments.

Low-cost systems, *i.e.* a single low resolution camera (320x240), is not performing well enough in such applications. The variability in the appearance of pedestrians (*e.g.* clothing), their articulated structure, and the non-rigid kinematics affect the performance of the existing systems. This effect can be reduced if additional priors are integrated. Such priors can be stable features extracted from other existing cameras observing the same scene. Few years ago, such priors were not available whereas nowadays, they are. Indeed, very large number of fixed cameras have been installed in major cities (*e.g.* in 2002, approximately four millions just for the UK according to McCahill and Norris (2002)). Therefore, features extracted from those fixed cameras can be used to detect similar objects in the mobile cameras. Moving objects are more easily detected with fixed cameras. For instance, background subtraction is a natural approach to detect a moving object. Porikli (2006) presented methods to detect and track objects with a single fixed camera.

This work focuses on the detection of pedestrians within a mobile camera given their observations from a fixed camera. No training stage and data are used. However, we suppose that pedestrians are correctly detected in the fixed cameras.

The paper is structured as follows: first, a brief overview of current feature-based pedestrian detection methods is given. Then, the proposed system is presented. In section 4, the performances of both systems are evaluated on challenging data sets. Quantitative and qualitative results are given.

2 Feature-based Pedestrian Detection

Most of the low-cost vision-based systems, *i.e.* based on a single moving camera, address the pedestrian detection problem as a pattern classification one. A set of features is extracted from a large number of training samples to train a classifier (see Figure 1). Thousands of images of pedestrians and non-pedestrians are required.

Oren et al. (1997) and Papageorgiou and Poggio (1999) use Haar wavelet coefficients of a set of normalized pedestrian images. They classify the images with a support vector machine (SVM) and a "bootstrapping" method.

Gavrila (2000) uses a template matching technique based on hierarchical representation of the templates. Shape matching is based on distance transforms (chamfer distance). A reasonable shape extraction is needed.

Broggi et al. (2000) and Bertozzi et al. (2003) detect pedestrians without any training. Their detection is based on morphological characteristics of pedestrians (size and aspect ratio), vertical linear filter, and the strong vertical symmetry of the human shape. Moreover, an assumption about the region where a pedestrian can be found is done. Thus, their system only operates on flat roads with smoothly varying slope. In addition, multiple detections of the same pedestrian occurs, and pedestrian with monochrome clothing are hardly detected.

Shashua et al. (2004), Dalal and Triggs (2005), Suard et al. (2006) have shown that histogram of oriented gradient (HOG) is an efficient and robust shape-based cue. Recently, Tuzel et al. (2008) obtain the best performance with a novel object descriptor

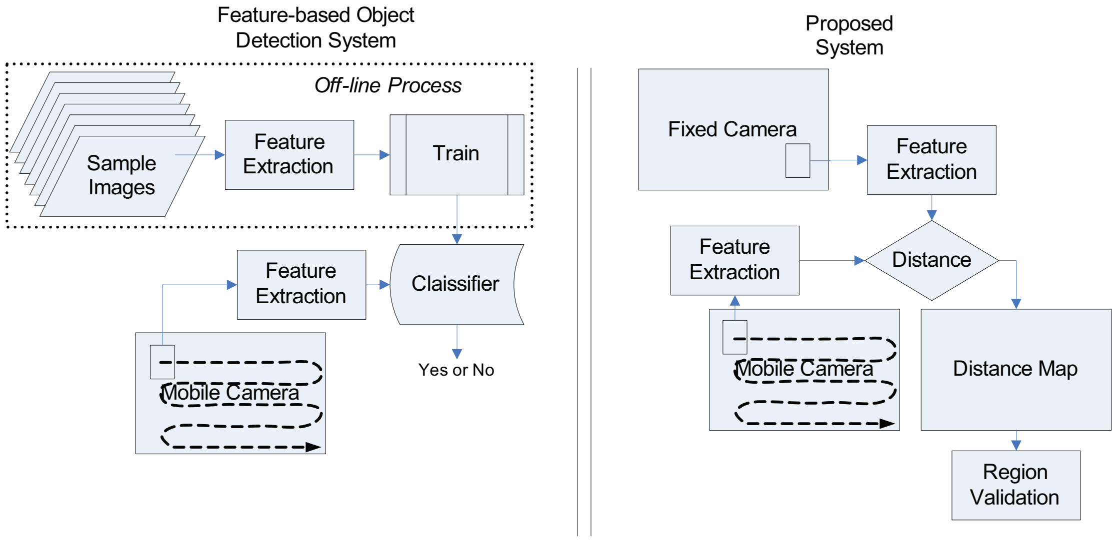


Figure 1: Feature-based object detection framework and the proposed system to detect a pedestrian in a mobile camera

based on covariance matrices. They outperform previous approaches. In section 4, the performance of their work is compared with our proposed system.

3 Collaborative low-cost Pedestrian Detector

3.1 Problem Formulation

Given an observation x of an object O in a fixed camera, we wish to detect its presence in the view of a mobile camera, and if present, locate it in its image plane.

Let y_i be a potential region in the mobile camera. x and y_i are subsets of an image bounded by a rectangular bounding box.

We define the "Region Matching" operator, Φ, which maps a region x to the N_y most similar regions in a given image I_m:

$$\Phi(x, I_m, N_y) = \{y_1, y_2, ..., y_{N_y}\} = Y_x \tag{1}$$

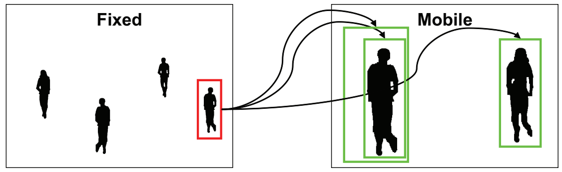
with I_m the image plane of the mobile camera. The precise notion of similarity will be described in section 3.2.

The same operator Φ can be used to map any y_i to a set of $\widehat{x}_i$ referred in this paper as the dual problem:

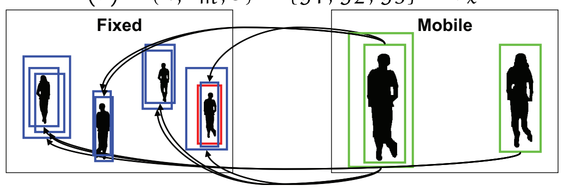
$$\Phi(y_i, I_f, N_x) = \{\widehat{x}_1, ..., \widehat{x}_{N_x}\} = \widehat{X}_i, \tag{2}$$

where $\widehat{X}_i$ are the regions in the fixed camera similar to y_i.

If a region $\widehat{x}_i$ matches x, then the corresponding y_i should be the region bounding object O in the mobile camera (see Figure 2). If none of the $\widehat{x}_i$ coincides with x, object O is probably not present in the view of the mobile camera.



(a) $\Phi(x, I_m, 3) = \{y_1, y_2, y_3\} = Y_x$



(b)For $i = 1:3$: $\Phi(y_i, I_f, 3) = \{\widehat{x}_1, \widehat{x}_2, \widehat{x}_3\} = \widehat{X_i}$

Figure 2: Illustration of the Φ operator. (a) An object x, highlighted in the fixed camera, is mapped to the best 3 regions in the mobile camera. (b) Then, each region y_i is mapped back to 3 regions in the fixed camera. If those regions coincide with x, there is a match.

We hence define an operator ϑ to validate if a region y_i matches x:

$$\vartheta(y_i|x, \widehat{X_i}) = \vartheta(y_i|x, \widehat{x}_1, ..., \widehat{x}_j) \in [0, 1] \tag{3}$$

As a result, the problem can be formulated as follows: for a given x, find the region y_x in the mobile camera that maximizes $\vartheta(y_i|x, \widehat{X_i})$ for all $y_i \in Y_i$:

$$y_x = \arg\max_{y_i \in Y_i} \vartheta(y_i|x, \widehat{X_i}) \tag{4}$$

If such a y_x does not exist, it means that the object is not present in the image plane of the mobile camera.

3.2 Region Matching

An object descriptor (OD) is created from the region bounding the object of interest in the fixed camera. Then, all possible regions in the image plane of the mobile camera are compared with the OD. A window of size proportional to the object bounding box scans the image plane of the mobile camera at different scales. For each region, its similarity with the OD is computed to find the region with highest similarity. Therefore, a discriminative region descriptor is needed.

3.2.1 Covariance Matrices as Region Descriptors

Covariance matrices are a very attractive descriptor first used by Tuzel et al. (2006). For each pixel, a set of features are extracted. Alahi, Marimon, Bierlaire and Kunt (2008) use the grayscale intensity, I, and the norm of the first order derivatives with respect to x and y, I_x and I_y:

$$f_n = (x, y, I, I_x, I_y). \tag{5}$$

Other features such as the R,G,B values or the second order derivatives, the gradient magnitude, mg, and its angle, θ, can also be used according to Alahi, Bierlaire and Kunt (2008). The pixel coordinates, x and y, are integrated in the feature vector to consider the spatial information of the features. Finally, the covariance of a region is computed as:

$$C_i = \frac{1}{N-1}\sum_{n=1}^{N}(f_n - m)(f_n - m)^T, \tag{6}$$

where N is the number of points in the region, and m the mean vector of all the feature vectors.

With covariance matrices, several features can be fused in a lower dimensionality without any weighting or normalization. They describe how features vary together.

Similarity between two regions B_1 and B_2 is given by the following distance proposed by Forstner and Moonen (1999):

$$\sigma_1(B_1, B_2) = \sqrt{\sum_i \ln^2 \lambda_i(C_1, C_2)} \tag{7}$$

where $\lambda_i(C_1, C_2)$ are the generalized eigenvalues of the covariance matrices C_i

3.2.2 A Collection of Grids of Descriptors

An object descriptor (OD) is used taking into account local and global information. It is a collection of grids of region descriptors (see figure 3). Each grid segments the object into a different number of sub-rectangles of equal sizes (called *blobs*). Grids of finer blob size describe local information whereas grids of coarse blob size describe a more global behavior.

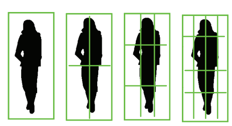


Figure 3: A collection of grids of descriptors

Similarity between two objects, $\phi(x, y_i)$, is computed by summing distance between corresponding blobs segmenting the grids. Since, many objects do not have a rectangular shape and some can be partially occluded, only the most similar blobs are kept. Thereupon, blobs belonging to the background can also be discarded.

3.3 Matching Process

3.3.1 Preprocessing step: Edge Filtering

Some regions in the mobile camera do not need to be compared with the ODs. They can be discarded with a simple preprocessing. The difference between the proportion of edges in two regions can give a quick indication about their similarity. If the proportion

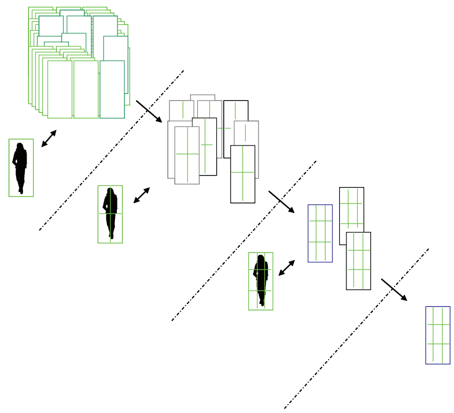


Figure 4: A three stages cascade of coarse to fine descriptors

of edges is not similar, the region is discarded. As a result, fewer regions remain to be analyzed and it increases the likelihood to detect the right object by reducing the search space.

3.3.2 Cascade of Coarse to Fine Descriptors

Many regions can be easily discarded without knowing the local information. Therefore, an approach similar to a cascade of classifier is proposed. "Easy regions" are discarded with coarse grids (*i.e.* grids with small number of blobs). More challenging regions require the use of finer grids (*i.e.* more number of blobs).

The detection process is divided into several stages. At each stage, a finer grid is used. After each stage, only the best candidates, *i.e.* regions with highest similarity (top ρ% of the evaluated regions), remain.

ρ is chosen such that after each stage the same percentage is kept:

$$N_r \times \rho^{N_s} = 1 \tag{8}$$

where N_r is the total number of regions in the mobile camera to compare with the object descriptor, and N_s is the total number of stages to use.

$$\rho = N_r^{-1/N_s} \tag{9}$$

3.4 Region Validation

The validation operator, ϑ, evaluates the likelihood that object x matches region y_i in the mobile camera. It considers the dual problem by analyzing the set obtained by $\Phi(y_i, I_f, N_x) = \{\widehat{x}_1, \widehat{x}_2, ..., \widehat{x}_{N_x}\}$. Alahi, Vandergheynst, Bierlaire and Kunt (2008) studied the impact of the choice of N_x on the performance.

A similarity measure σ between the original x and each $\widehat{x}_i$ is estimated based on the spatial arrangement of their bounding boxes:

$$\sigma(x, \widehat{x}_i) = \frac{1}{1 + c_1 e^{-c_2 \cdot O}} w_o + \frac{1}{1 + c_1 e^{-c_2 \cdot C}} w_c + \frac{1}{1 + c_1 e^{-c_2 \cdot D_c}} w_d \tag{10}$$

where

- C is a percentage which represents how much of the original bounding box of x is covered by the bounding box of $\widehat{x}_i$. Likewise, O is the percentage which represents how much $\widehat{x}_i$ is covered by x. (see figure 5)
- D_c measures the similarity of the center of two bounding boxes. The smallest is the euclidian distance between the center, the highest is D_c.

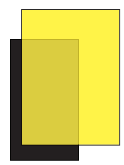


Figure 5: An example of the bounding box of x (in black) and $\widehat{x}_i$ where $C \approx 0.75$, $O \approx 0.4$

c_1 and c_2 are the parameters of the logistic function leading to the curve illustrated in figure 6.

$\sigma(x, \widehat{x}_i) > 0$ if and only if C and $O > 30\%$ and $D_c < 0.75 * max(width_x, height_x)$.

A weight $w_.$ is given to each factor to emphasize priority. In this work, focus is first on a high cover of x, then a similar center of mass, finally $\widehat{x}_i$ should not be too big with respect to x (decent O).

The logistic operator is used to reduce sensitivity to two regions overlapping with a slight difference. A linear σ_l such as:

$$\sigma_l(x, \widehat{x}_i) = 1 - (\frac{1 - O}{1 - c_1} w_o + \frac{1 - C}{1 - c_2} w_c + \frac{D_c}{c_3} w_d) \tag{11}$$

was too sensitive to differences. Figure 6 plots the two operators, and figure 7 presents an example of the value obtained with σ and σ_l.

Finally, $\vartheta(y_i | x, \Phi(y_i))$ is computed as follows:

$$\vartheta(y_i | x, \Phi(y_i)) = \max_{\widehat{x}_i \in \Phi(y_i)} \sigma(x, \widehat{x}_i) w(y_i) \tag{12}$$

where $w(y_i)$ weights region y_i with respect to other y_j based on the similarity measurement computed by $\Phi(x)$ (in section 3.2.2):

$$w(y_i) = \frac{\phi(x, y_i)}{\max_{y_j \in \Phi(x)} \phi(x, y_j)} \tag{13}$$

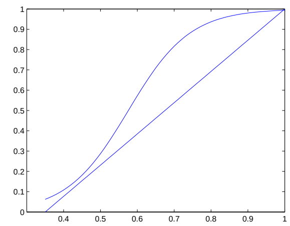


Figure 6: x-axis represents C or O; y-axis represents its contribution to σ and σ_l. It can be seen that for values of C or O close to 1, the contribution remains 1 (full) for the logistic operator.

Figure 7: The linear σ_l gave 0.63% and the proposed σ gives 0.86%

4 Performance Evaluation

4.1 Data Sets

Several data sets composed of video sequences captured by a fixed and a mobile camera are considered[1]. Fixed cameras are located at a height equivalent to the first floor of a building. Mobile cameras are held by pedestrians or mounted on cars. The images are recorded at 25fps with a resolution of 320×240. Figure 9 to 13 present examples of images captured by the cameras.

The data sets used have meaningful changes in viewpoint, illumination, and color distribution between fixed and mobile cameras. Sensing devices are also different. Indeed, mobile cameras have a cheap capturing device and hence provide noisy images.

4.2 Experiments

First, the state-of-the-art single frame pedestrian detector (referred to as SPD in this section) proposed by Tuzel et al. (2008) is evaluated. The true positive rate (%TP) and false positive rate are given in table 1. For comparison purposes, the performance of a pedestrian detector based on Haar features and Adaboost classifier is also presented (Viola and Jones, 2002). The SPD has clearly a much better detection rate compared to the detector based on Haar features. However, the false positive rate is still very high (more than 4 #FP/frame).

[1] Data sets are available at http://lts2www.epfl.ch/˜ alahi/data.htm

Method based on	%TP	#FP/frame
Feature-based Pedestrian detector based on:		
Haar features (Viola and Jones (2002))	11.0	2.1
Covariance features (Tuzel et al. (2008))	57.4	4.32
Covariance features (Tuzel et al. (2008)) filtered out	44.1	0.75
Proposed Collaborative Pedestrian Detector based on:		
Covariance matrices of I, $\lvert I_x\rvert$, $\lvert I_y\rvert$	65.8	0.63
Covariance matrices of I, $\lvert I_x\rvert$, $\lvert I_y\rvert$, mg, θ	69.7	1.4

Table 1: Performance measurement. %TP and #FP/frame on the complete data sets.

Our proposed system can be used to reduce the false positive rate of the SPD. All detected regions by the SPD can be filtered out by the validation operator ϑ described in section 3.4. Applying the operator to the detected regions reduce the false positive rate by 83%. Figure 8 illustrates the regions filtered out with our validation scheme. However, with such approach, the detection rate is likely to decrease. Setting $N_x = 5$, the %TP is reduced by 23%. Running the SPD and filtering out its result is not the best approach since the detection rate is bounded by the performance of the classification approach.

The performance of our proposed system (without any SPD running) gives the best performance compared to other approaches (see table 1). Only observations from the fixed camera are used. The true positive rate is increased while the false positive rate is decreased.

Figure 9 to 12 illustrate the output of our proposed system compared with the SPD. Our proposed system is able to detect pedestrians who are not detected by the SPD due to poor contrast and image quality without increasing the number of false positives. Extracting features from the fixed camera enable the detection of pedestrians in challenging situations.

On the one hand, Figure 9 presents the success of our approach when pedestrians viewed by the mobile camera are of very low resolution and image quality. The mobile camera is as far as the fixed camera to the pedestrians of interest. On the other hand, pedestrians are still correctly detected when they are much closer to the mobile camera (Figure 10). Figure 11 shows that even when the viewpoint is drastically different, the algorithm succeeds in detecting the pedestrians. Finally, figure 12 evaluates the strength of the validation operator. None of the objects present in the fixed camera are wrongly found in the mobile camera.

The proposed system can be generalized to any objects of interest. The focus of this work is on pedestrian. Nevertheless, any other object observed by the fixed camera is correctly detected in the mobile camera. Figure 13 presents some random examples. The object descriptor is generic to any object of interest.

Figure 8: First column: objects detected by the SPD. Second column: Remaining objects kept after proposed validation scheme Third column: Objects detected by the fixed camera

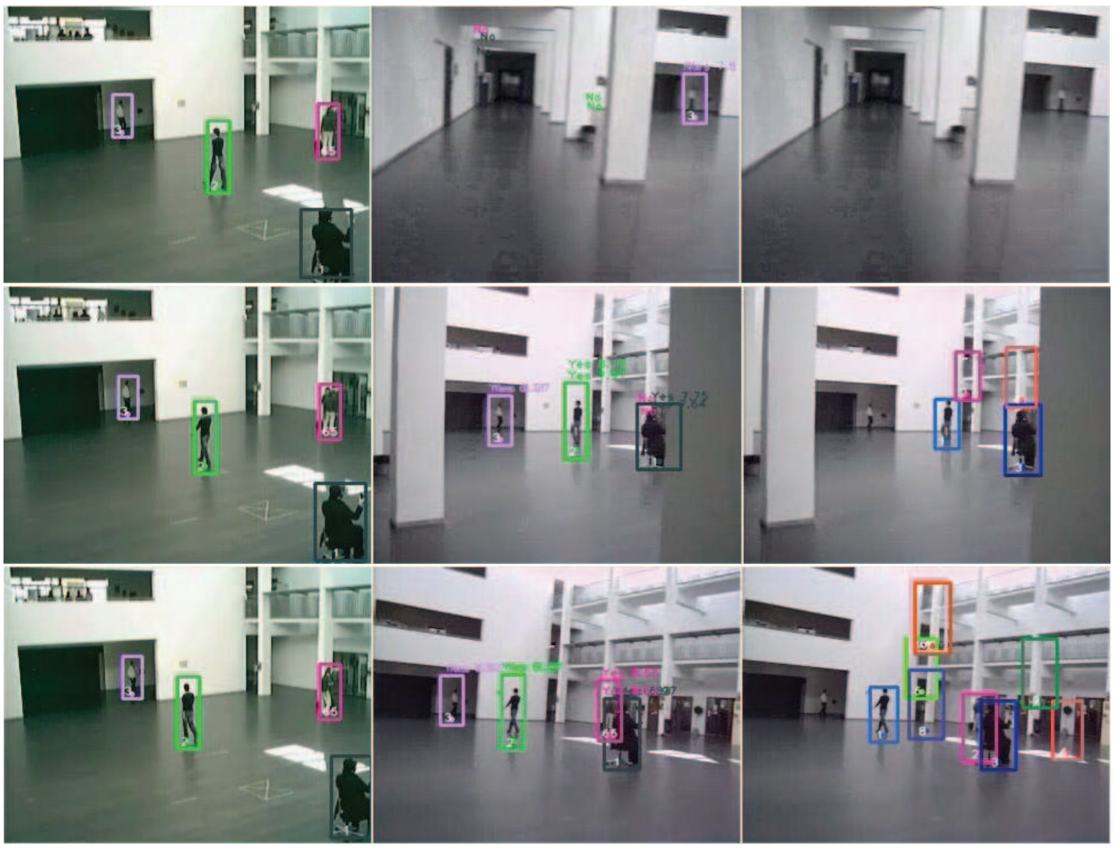


Figure 9: First column: objects detected by fixed camera. Second column: corresponding objects detected with the mobile camera. Third column: output of the SPD proposed by Tuzel et. al. (2008)

Figure 10: First column: objects detected by fixed camera. Second column: corresponding objects detected with the mobile camera. Third column: output of the SPD proposed by Tuzel et. al. (2008)

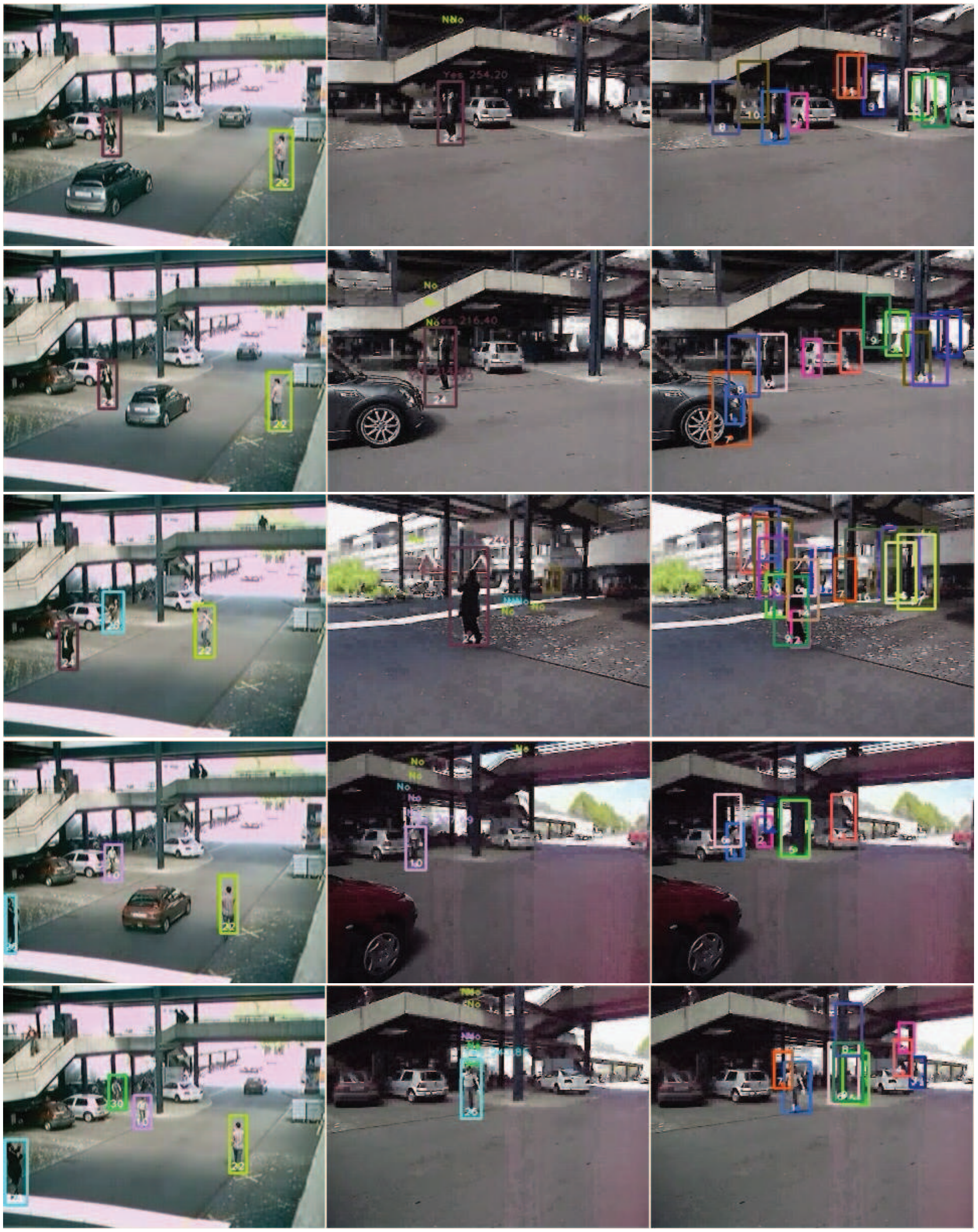


Figure 11: First column: objects detected by fixed camera. Second column: corresponding objects detected with the mobile camera. Third column: output of the SPD proposed by Tuzel et. al. (2008)

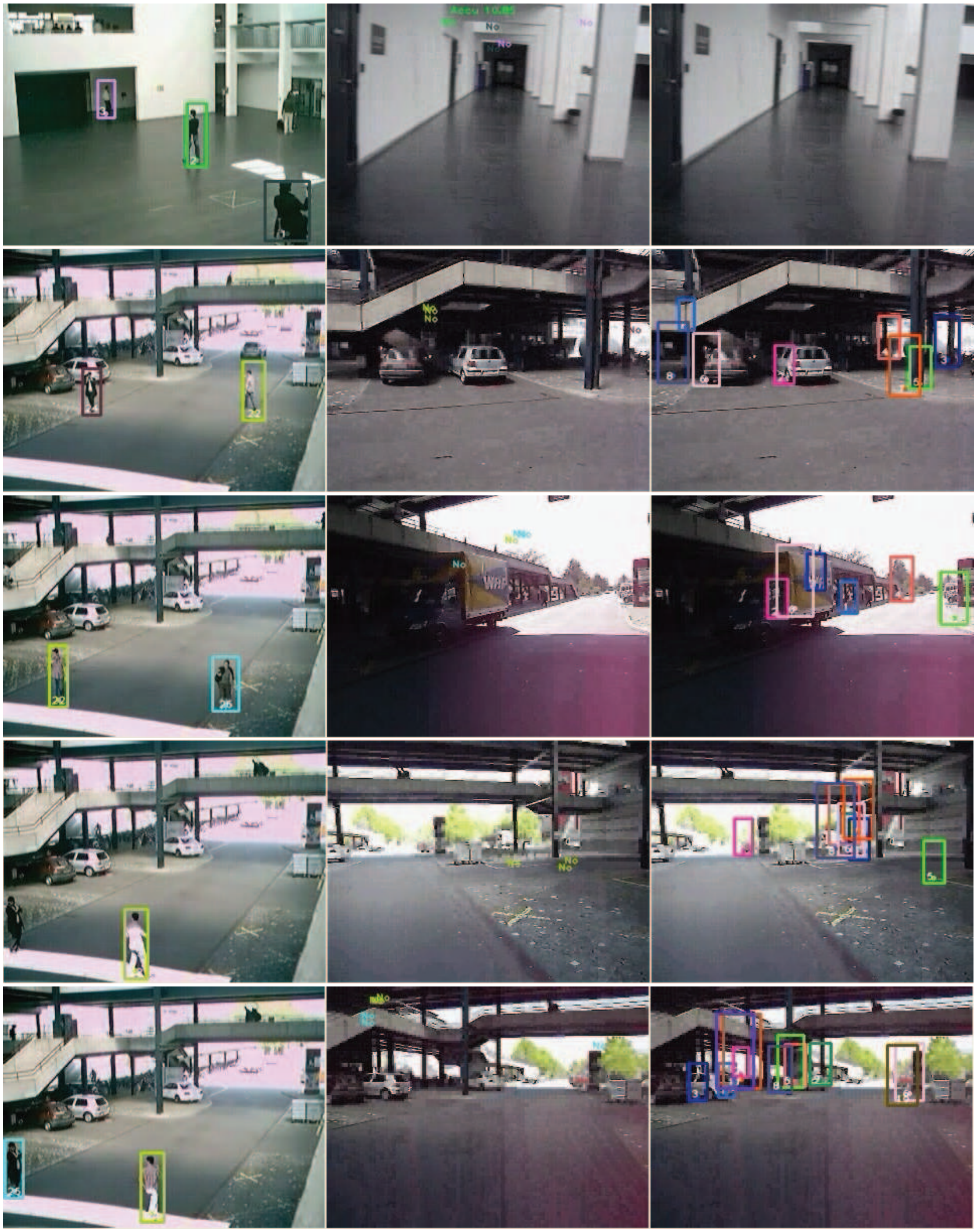


Figure 12: First column: objects detected by fixed camera. Second column: corresponding objects detected with the mobile camera. Third column: output of the SPD proposed by Tuzel et. al. (2008)

Figure 13: First column: objects detected by fixed camera. Second column: corresponding objects detected with the mobile camera

4.3 Discussions

The proposed system outperforms the feature-based pedestrian detection approach since higher detection rate and lower false positive rate is obtained. In fact, the number of potential false positives is upper bounded by the number of objects observed by the fixed camera. In this work, it is supposed that fixed cameras have correctly detected the pedestrians since they can benefit from their static behavior to model the background and segment efficiently all moving objects, hence pedestrians (Porikli, 2006).

One of the main advantage of such system is its generalization to any objects of interest. Regarding the driver assistance application, all potential collisions are of interest. A stroller, a bicycle, or an animal (e.g. a dog) passing in front of a vehicle can be detected with our system since they are detected in the fixed camera with a simple background subtraction algorithm. The limit of our system is hence the detection result from the fixed cameras. In addition, the complexity cost depends on the number of objects to look for in the mobile camera.

Also, some applications require a correspondence within the detected objects across the cameras. Typically, if an object observed by a fixed camera is correctly detected and matched in a mobile camera, higher resolution features can be extracted leading to a better analysis of the object (such as facial expression analysis for instance). Techniques to detect an object with a camera cannot be used to match the objects across cameras. By definition, they remove the discriminative parts between two objects of the same category. To find correspondence between two views, most of the systems suppose static and calibrated cameras. Our proposed system, not only detect the objects but match them across cameras.

Finally, the use of mathematical models describing the content of the scene can be combined with the proposed method to make the results even more robust. In particular recent research developments on pedestrian walking behavior (Antonini, Bierlaire and Weber, 2006, Robin et al., forthcoming) have proved to be highly relevant and efficient in the context of pedestrian tracking and detection (Antonini, Venegas, Bierlaire and Thiran, 2006, Venegas et al., 2005).

5 Conclusions

This work presents how fixed cameras installed in major cities can enhance the performance of an automatic vision-based pedestrian detector. A novel system is presented to detect in the image plane of a mobile camera pedestrians observed by a fixed camera. No calibration among the cameras is needed. No training data is used. It outperforms previous systems based on a single low resolution mobile camera. Future work can compare our system with more sophisticated systems based on stereo cameras or infrared cameras.

Moreover, the strength of our proposed system is its generalization to any object of interest. Any object observed by a fixed camera can be detected in the mobile camera. In this work, a master-slave approach is used between the fixed and mobile cameras. Detected objects with the mobile cameras do not influence the detection process within the fixed cameras. Hence, future work can also evaluate detection in the mobile cameras to enhance results in the fixed cameras.

Acknowledgments

We are very grateful to Fatih Porikli who provided us with the state of the art approach for single frame pedestrian detection. We also thank Pierre Vandergheynst for useful discussions.

References

Alahi, A., Bierlaire, M. and Kunt, M. (2008). Object Detection and Matching with Mobile Cameras Collaborating with Fixed Cameras.

Alahi, A., Marimon, D., Bierlaire, M. and Kunt, M. (2008). A master-slave approach for object detection and matching with fixed and mobile cameras, *Accepted IEEE Int. Conf. on Image Processing (ICIP), San Diego, CA, USA*.

Alahi, A., Vandergheynst, P., Bierlaire, M. and Kunt, M. (2008). Object Detection and Matching in a Mixed Network of Fixed and Mobile Cameras.

Ali, A. and Dagless, E. (1990). Vehicle and pedestrian detection and tracking, *Image Analysis for Transport Applications, IEE Colloquium on* p. 5.

Antonini, G., Bierlaire, M. and Weber, M. (2006). Discrete choice models of pedestrian walking behavior, *Transportation Research Part B: Methodological* **40**(8): 667–687.

Antonini, G., Venegas, S., Bierlaire, M. and Thiran, J.-P. (2006). Behavioral priors for detection and tracking of pedestrians in video sequences, *International Journal of Computer Vision* **69**(2): 159–180.

Bertozzi, M., Broggi, A., Chapuis, R., Chausse, F., Fascioli, A. and Tibaldi, A. (2003). Shape-based pedestrian detection and localization, *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, Shangai, China, pp. 328–333.

Bertozzi, M., Broggi, A., Felisa, M., Vezzoni, G. and Del Rose, M. (2006). Low-level pedestrian detection by means of visible and far infra-red tetra-vision, *Procs. IEEE Intelligent Vehicles Symposium 2006*, Tokyo, Japan, pp. 231–236.

Bota, S. and Nedesvchi, S. (2008). Multi-feature walking pedestrians detection for driving assistance systems, *Intelligent Transport Systems, IET* **2**(2): 92–104.

Broggi, A., Bertozzi, M., Fascioli, A. and Sechi, M. (2000). Shape-based pedestrian detection, *Proc. IEEE Intelligent Vehicles Symp* pp. 215–200.

Broggi, A., Fedriga, R. I., Tagliati, A., Graf, T. and Meinecke, M. (2006). Pedestrian detection on a moving vehicle: an investigation about near infra-red images, *Procs. IEEE Intelligent Vehicles Symposium 2006*, Tokyo, Japan, pp. 431–436.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection, *CVPR05*, pp. I: 886–893.

Forstner, W. and Moonen, B. (1999). A metric for covariance matrices, *Qua vadis geodesia* pp. 113–128.

Gavrila, D. (2000). Pedestrian detection from a moving vehicle, *ECCV00*, pp. II: 37–49.

McCahill, M. and Norris, C. (2002). Cctv in london.

Oren, M., Papageorgiou, C., Sinha, P., Osuna, E. and Poggio, T. (1997). Pedestrian detection using wavelet templates, *Proc. Computer Vision and Pattern Recognition* **97**: 193–199.

Papageorgiou, C. and Poggio, T. (1999). Trainable pedestrian detection, *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on* **4**.

Porikli, F. (2006). Achieving real-time object detection and tracking under extreme conditions, *Journal of Real-Time Image Processing* **1**(1): 33–40.

Robin, T., Antonini, G., Bierlaire, M. and Cruz, J. (forthcoming). Specification, estimation and validation of a pedestrian walking behavior model, *Transportation Research Part B: Methodological* . Accepted for publication.

Shashua, A., Gdalyahu, Y. and Hayun, G. (2004). Pedestrian detection for driving assistance systems: Single-frame classification and system level performance, *IVS04*, pp. 1–6.

Statistics of Road Traffic Accidents in Europe and North America (2007). Economic Commission for Europe, United Nations, Geneva.

Suard, F., Rakotomamonjy, A., Bensrhair, A. and Broggi, A. (2006). Pedestrian detection using infrared images and histograms of oriented gradients, *Procs. IEEE Intelligent Vehicles Symposium 2006*, Tokyo, Japan, pp. 206–212.

Tuzel, O., Porikli, F. and Meer, P. (2006). Region covariance: A fast descriptor for detection and classification, *Proc. 9th European Conf. on Computer Vision* .

Tuzel, O., Porikli, F. and Meer, P. (2008). Pedestrian detection via classification on riemannian manifolds, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**.

Velastin, S., Boghossian, B. and Vicencio-Silva, M. (2006). A motion-based image processing system for detecting potentially dangerous situations in underground railway stations, *Transportation Research Part C* **14**(2): 96–113.

Venegas, S., Antonini, G., Thiran, J.-P. and Bierlaire, M. (2005). Automatic pedestrian tracking using discrete choice models and image correlation techniques, *in* B. S. and B. H. (eds), *Machine Learning for Multimodal Interaction*, Vol. 3361 of *Lecture Notes in Computer Science*, Springer, pp. 341 – 348. ISBN:978-3540245094.

Viola, P. and Jones, M. (2002). Robust real-time object detection, *International Journal of Computer Vision* **57**(2): 137–154.

Zhao, L. and Thorpe, C. (2000). Stereo-and neural network-based pedestrian detection, *Intelligent Transportation Systems, IEEE Transactions on* **1**(3): 148–154.