Monocular Vision-Based Target Detection on Dynamic Transport Infrastructures

S. Álvarez, M.A. Sotelo, D.F. Llorca, R. Quintero, and O. Marcos

Department of Automation, University of Alcalá, Alcalá de Henares, Madrid, Spain
{sergio.alvarez,sotelo,llorca}@aut.uah.es

Abstract. This paper describes a target detection system on transport infrastructures, based on monocular vision. The goal is to detect and track vehicles and pedestrians, dealing with objects variability, different illumination conditions, shadows, occlusions and rotations. A background subtraction method, based on GMM and shadow detection algorithms are proposed to do the segmentation of the image. Finally a feature extraction, optical flow analysis and clustering methods are used for the tracking step. The algorithm requires no object model and prior knowledge and it is robust to illumination changes and shadows.

Keywords: GMM, shadow removal, feature detection and tracking.

1 Introduction

In recent years, the use of cameras for traffic scene analysis has greatly promoted the development of intelligent transportation systems. In result, video sequences are used to detect vehicles and pedestrians for traffic flow estimation, signal timing, safety applications or video surveillance, among others. The challenge and the main task to solve are the object segmentation and tracking.

As the traffic monitoring systems normally use fixed cameras, most of the named applications above are based on the *background subtraction* algorithm. The idea is to subtract the current image from a reference image, which is a representation of the scene background, to find the foreground objects. The technique has been used for years in many vision systems as a preprocessing step, and the results obtained are fairly good. However the algorithm is susceptible to several problems such as sudden illumination changes, cast shadows, camera jitter or image noise; which often cause serious errors due to misclassification of moving objects.

In this paper, an approach for detecting moving objects from a static background scene is presented. The algorithm differs from the previous ones in some aspects. It gives more robust foreground extraction, refining shadow/highlight detection with a novel method, which does not need any prior knowledge or threshold. This skill make the system as "plug&play" due to the possibility to use it in a wide range of environments and illumination conditions, without setting any parameter.

R. Moreno-Díaz et al. (Eds.): EUROCAST 2011, Part I, LNCS 6927, pp. 576–583, 2012.

The remainder of the document is organized as follows. Section 2 provides a brief summary of the related work. Section 3 describes the proposed method. Implementation and experimental results are described in Section 4, and finally, Section 5 summarizes the conclusions and future work.

2 Related Work

The main related work in traffic monitoring, using vision-based systems with fixed cameras, is based on the background subtraction method. Stauffer et al. [1] present a method that model each pixel intensity by a mixture of K Gaussian distributions, and Zivkovic [2] improve the method incorporating a model selection criterion to choose the proper number of components for each pixel on-line, obtaining an automatic full adaptation to the scene. Although these methods show interesting results in good illumination conditions they are vulnerable to sudden changes, and shadows cast by moving objects can easily be misinterpreted as foreground.

Many efforts have been made to solve the problem of the illumination changes. Algorithms can be classified as model-based or property-based. On the one hand, model-based methods use prior knowledge of scene geometry, target objects or light sources to predict and remove shadows. For example Joshi et al. [3] propose an algorithm that could detect shadows by using Support Vector Machine (SVM) and a shadow model, learned and trained from a database.

On the other hand, property-based approaches use features like geometry, brightness or color to detect illumination changes. Horprasert et al. [4] propose a computational color model to classify each pixel as foreground, background, shadowed background, or highlighted background. Salvador et al. [5] use the idea that a shadow darkens the surfaces on which it is cast, to identify an initial set of shadowed pixels, that is then pruned by using color invariance and geometric properties of shadows. In [6], Cucchiara et al. use the hypothesis that shadows reduce surface brightness and saturation while maintaining hue properties in the HSV color space. These methods can deal with illumination noises and soft shadows but they fail representing heavily shadows where color and chromaticity information are totally lost.

Related to tracking, Bayesian filtering, and in particular Kalman filter, is extensively used to predict the position of the targets. The state vector can be modeled with data directly available from blobs such as kinematic parameters [7]. However, the most interesting works combine background subtraction and feature tracking to take advantage when partial occlusion occurs; since some of the features of the object remain visible. Kanhere et al. [8] use the background subtraction result to estimate the 3D height of corner features by assuming that the bottom of the foreground region is the bottom of the object. ZuWhan [9] present a dynamic multi-level feature grouping approach that provides high-quality trajectories. The problem is the system needs 3D information and a semi-supervised learning procedure at the beginning.

3 System Description

In this section, the implemented method is described in separated subsections.

3.1 Background Subtraction

The basic idea of background subtraction is to subtract the current image from a reference image that models the background scene. Obviously the capturing system has to be fixed and the background static. Although pedestrians and vehicles are the only objects which are moving in the field of view, the algorithm is susceptible to both global and local illumination changes such as shadows, so a detection and treatment of these problems is needed to achieve satisfying results.

Rather than explicitly modeling the values of the pixels as one particular kind of distribution, like average, mean, etc., each pixel is modeled by a mixture of K Gaussian distributions [1], whose mean and variance is adapted over time. The probability that a certain pixel has a value X_t at time t can be written as:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \tag{1}$$

where the mean $\mu_{i,t}$, the covariance $\Sigma_{i,t}$ and the weight $\omega_{i,t}$ (with $0 < \omega_{i,t} \leq 1$), are the parameters of the k^{th} gaussian component, and η is the gaussian probability density function:

$$\eta(X_t, \mu, \sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)} \tag{2}$$

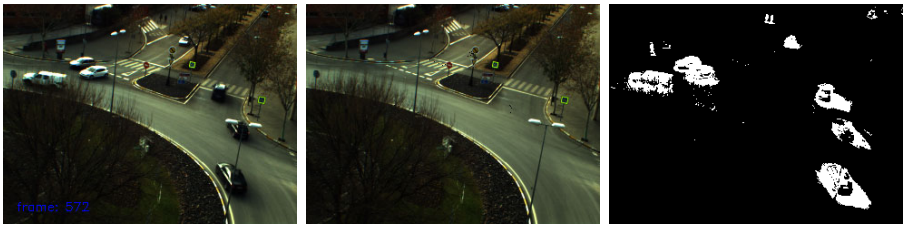
Given a new data sample X_t at time t, the recursive equations to update the model are [2]:

$$\omega_i = \omega_i + \alpha(\theta_i - \omega_i) \tag{3}$$

$$\mu_i = \mu_i + \theta_i (\frac{\alpha}{\mu_i}) \delta_i \tag{4}$$

$$\sigma_i^2 = \sigma_i^2 + \theta_i (\frac{\alpha}{\mu_i})(\delta_i^T \delta_i - \sigma_i^2) \tag{5}$$

where α is the learning rate and $\delta_i = X_t - \mu_i$. For a new sample the ownership θ_i is set to 1 if the sample matches with a component of the mixture (sorted by the value of $\frac{\omega}{\sigma}$) and 0 for the remaining models. The matching is defined by the Mahalanobis distance between the sample and the gaussian component of the mixture and a threshold. If there is no matching, a new component is generated with $\omega_{i+1} = \alpha$, $\mu_{i+1} = X_t$ and $\sigma_{i+1} = \sigma_0$, where σ_0 is a predefined initial variance. If the maximum number of components has been reached, the component with the smallest weight is discarded. Fig. 1 shows the result of this step.



(a) (b) (c)

Fig. 1. Background subtraction result on a tested sequence. (a) Original image; (b) modeled background; (c) extracted foreground.

3.2 Shadow and Highlight Detection

Background subtraction step detects all the moving objects that do not belong to any component of the mixture. Despite the robust detection in good illumination conditions, the algorithm suffers with the presence of shadows and sudden illumination changes (see Figure 1(c)). For this reason, a shadow and highlight detection algorithm is implemented, based on texture matching.

The technique used is the normalized cross correlation, and particularly *color normalized cross correlation* (CNCC). The idea is based on the fact that a shadow or a highlight changes color properties of the objects, but not their surface properties such as texture. The algorithm uses this method to compare the texture of every foreground pixel, by a neighbourhood window, with the correspondent one in the background model.

Two different space colors are used to compute the correlation. On the one hand, RGB is chosen for soft shadows and sudden illumination changes; and on the other hand, for strong shadows the international standard CIE 1931 XYZ color space has been tested empirically with better results; so two different matching analysis are done. Figure 2 shows the result of removing strong shadows.

After the matching, shadows and illumination changes are detected and removed, except for the external edge of the strong shadows (Figure 2(c)). A last simple step then is needed: the external contours of the initial foreground are extracted and dilated. After that, these contours are subtracted from the matching foreground image, and finally the resultant image is dilated again to recover the original size of the detected objects (Figure 2(d)).

Avoiding the problems of this technique due to the absence of imaging scale, rotation, and perspective distortions, the method works fairly good in every tested situations, under different illumination conditions.

3.3 Feature Extraction and Tracking

After extracting correctly the image foreground, a new step to distinguish between different objects is done. Moreover, due to partial and global occlusions,

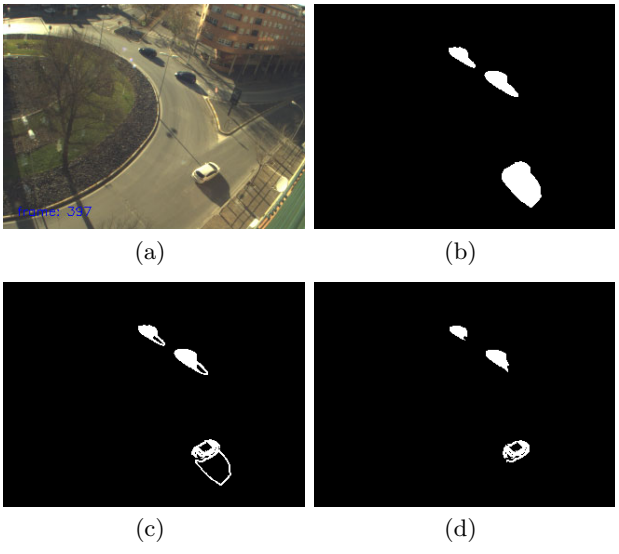


Fig. 2. Strong shadow removal process. (a) Original image with strong shadows; (b) GMM initial foreground; (c) foreground after NCC matching; (d) final foreground.

detected objects could be fragmented, joined with a close one or even lost; so a tracking algorithm is needed. Feature-based tracking gives up the idea of tracking objects as a whole, after obtain the different regions through background subtraction. The idea of this algorithms is to extract and track foreground features and group them into objects using proximity, motion history, velocity and orientation.

The proposed method is called *flock of features* and it is based on the work of Kölsch et al. [10]. The concept comes from natural observation of flocks of birds or fishes. It consists of a group of members, similar in appearance or behaviour to each other, which move congruously with a simple constraint: members keep a minimum safe distance to the others; but not too separated from the flock. This concept helps to enforce spatial coherence of features across an object, while having enough flexibility to adapt quickly to large shape changes and occlusions.

Pyramid-based KLT feature tracking (Kanade, Lucas and Tomasi [11]), based on "good features to track" [12]; is chosen as the main tracker where the flock constrains are applied. Features are extracted from the foreground regions and tracked individually frame to frame. Moreover, for each feature, a level of life is computed. This level increases if the feature has a match in the next frame with a new extracted feature, and decreases otherwise. If this level reaches 0, feature is removed or reallocated inside the object depending on the constrains of the flock. Figure 3 depicts an example of the tracking step with a truck in a highway.

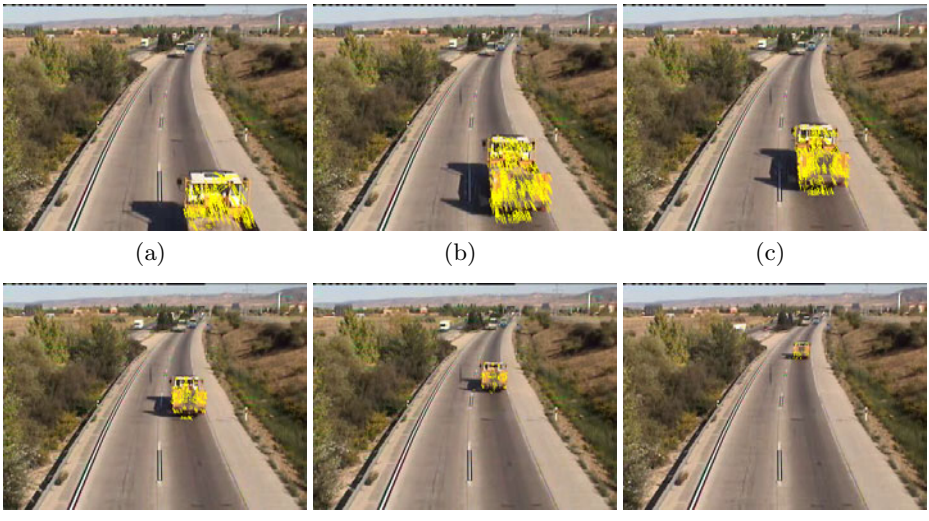


Fig. 3. Feature tracking sequence

4 Implementation and Results

This section demonstrates the performance of the proposed algorithm, on sequences acquired from the same place, but in different day times and illumination conditions. Although the section shows only two different videos, other results of the algorithm have been depicted in previous pages of the paper, in different environments and conditions, trying to cover as much situations as possible.

The system has been implemented on a Pentium IV PC at 2.4 GHz, running Kubuntu/Linux Operating System and OpenCV libraries, with a 320x240 CMOS camera. The sequences (Figures 4 and 5), have been chosen to represent heavy and light shadows in two opposite illumination conditions: with sun and at dusk.

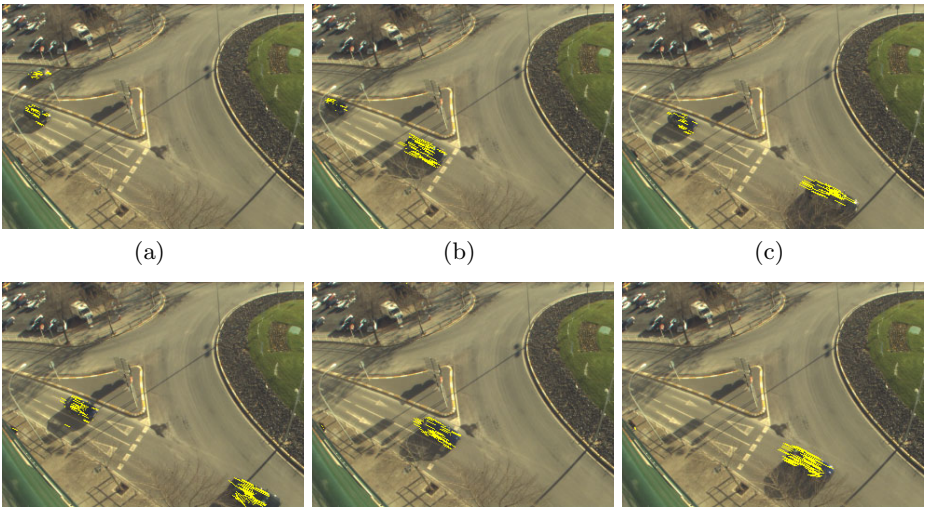
The results are very interesting (see Figure 6) and give the chance to extend and improve the system for night conditions.

5 Summary and Conclusions

In this work, a real-time monocular method has been developed to detect and track vehicles and other moving objects as pedestrians, for applications in the framework of Intelligent Transportation Systems (ITS).

The algorithm requires no object model and prior knowledge and it is robust to illumination changes and shadows. Therefore it can work indoor and outdoor, in different conditions and scenarios.

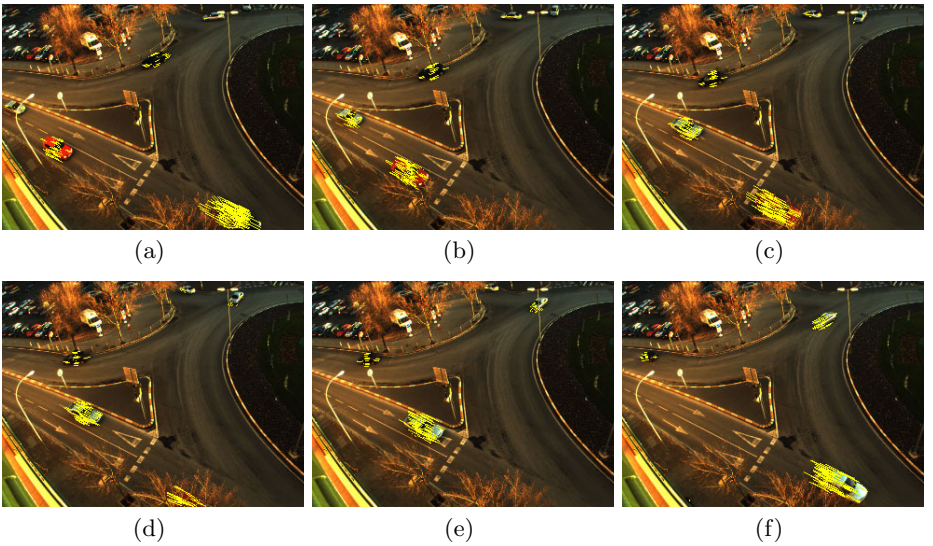
The performance of the system is demonstrated via several sequence images. Experimental results show different environments and illumination conditions and the proposed technique performs well in all of them; even with shadows.



(a) (b) (c)

(d) (e) (f)

Fig. 4. Result of the system in sunny sequence



(a) (b) (c)

(d) (e) (f)

Fig. 5. Result of the system at dust sequence

Future work will include applying the algorithm to a larger number of data and performing comparative studies on various applications, besides extending the approach for night conditions.

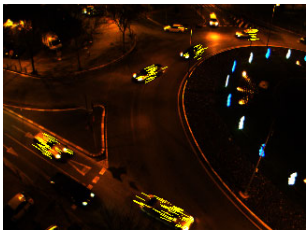


Fig. 6. Result of the system in night conditions

References

1. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (1999)
2. Zivkovic, Z., van der Heijden, F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letter (2006)
3. Joshi, A.J., Papanikolopoulos, N.: Learning to detect moving shadows in dynamic environments. IEEE Transactions on Pattern Analysis and Machine Intelligence (2008)
4. Horprasert, T., Harwood, D., Davis, L.S.: A statistical approach for real-time robust background subtraction and shadow detection. In: Proc. IEEE Int. Conf. Computer Vision FRAME-RATE Workshop (1999)
5. Salvador, E., Cavallaro, A., Ebrahimi, T.: Cast shadow segmentation using Invariant color features. Computer Vision and Image Understanding (2004)
6. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with HSV color information. In: Proceedings of Intelligent Transportation Systems Conference (2001)
7. Bas, E., Tekalp, M., Salman, F.S.: Automatic vehicle counting from video for traffic flow analysis. In: Proceedings of IEEE Intelligent Vehicles Symposium (2007)
8. Kanhere, N.K., Pundlik, S.J., Birchfield, S.T.: Vehicle segmentation and tracking from a low-angle off-axis camera. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR (2005)
9. Kim, Z.: Real time object tracking based on dynamic feature grouping with background subtraction. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2008)
10. Kölsch, M., Turk, M.: Fast 2D hand tracking with flocks of features and multi-cue integration. In: IEEE Workshop on Real-Time Vision for Human-Computer Interaction (2004)
11. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of Imaging Understanding Workshop (1981)
12. Shi, J., Tomasi, C.: Good features to track. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR (1994)