

3D Vehicle Extraction and Tracking from Multiple Viewpoints for Traffic Monitoring by using Probability Fusion Map

Zhencheng Hu* and Chenhao Wang* and Keiichi Uchimura*

* *The Graduate School of Science and Technology, Kumamoto University, Kurokami 2-39-1, Kumamoto, 860-8555 Japan*

Received 15th May 2008; accepted 12th June 2008

Abstract

This paper presents a novel solution of vehicle occlusion and 3D measurement for traffic monitoring by data fusion from multiple stationary cameras. Comparing with single camera based conventional methods in traffic monitoring, our approach fuses video data from different viewpoints into a common probability fusion map (PFM) and extracts targets. The proposed PFM concept is efficient to handle and fuse data in order to estimate the probability of vehicle appearance, which is verified to be more reliable than single camera solution by real outdoor experiments. An AMF based shadowing modeling algorithm is also proposed in this paper in order to remove shadows on the road area and extract the proper vehicle regions.

Key Words: Probability Fusion Map, 3D Modeling, Multiple View, Traffic Monitoring

1 Introduction

Intelligent traffic monitoring is an active area of research given the increasing volumes of vehicular traffic around the world leading to the problems as environmental degradation and economic inefficiency. For this purpose, the acquisition of accurate traffic data is essential for optimizing traffic management systems.

Approaches to traffic monitoring system can be classified into one of two major categories: spot monitoring and area monitoring systems. Microwave transducer, underground magnetic loop sensors are examples of spot monitoring system. Area monitoring systems generally rely on video cameras. Ground (magnetic) loop detectors are accurate but have some major drawbacks. Installation of these detectors requires the excavation of road surfaces and are, therefore, constructively expensive and complex. In addition, loop detectors can only acquire the vehicle count and speed. Video (camera) monitoring systems, on the other hand, have several advantages. They are easy to be installed and maintained.

In addition, visual information has potential ability to provide more information including lane changing frequency, vehicle trajectories, and driving behavior analysis through continuous tracking of the target vehicle. However, previous approaches are generally based on single camera's output, which is sensitive to the problem of occlusion, shadows and various illumination conditions. And traditional video monitoring systems cannot measure target vehicle's 3D size for vehicle classification and recognition.

Correspondence to: <hu@cs.kumamoto-u.ac.jp>

Recommended for acceptance by <João Manuel Tavares>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

In the work presented here, we describe a 3D vehicle extraction and tracking approach based on fusion of video data acquired from multiple un-calibrated cameras. We are using the fact that the images of the same scene acquired by the different cameras must have some degree of self-consistency when projected onto a common framework. Because points on the road surface can satisfy the constraint that the inverse mapped points lie on a horizontal plane, a set of common road features can be defined for the purpose of registering the inverse mapped image from the different cameras. Thus, the individual background-subtracted and inverse mapped images can be merged to create a Probability Fusion Map (PFM) where the intensity represents the probability of a vehicle being present. Length and width data of the target vehicle can be directly derived from the PFM, and height data is calculated through the original background-subtracted binary result. In this paper, we also proposed an Applied Median Filter (AMF) based shadowing modeling algorithm to remove shadows in order to extract the proper vehicle regions.

Result of real road tests shows more than 95 percent of correct recognition and 92 percent of vehicle speed estimation correctness. Vehicle 3D measurements will be fed into the further recognition process for vehicle classification and recognition.

2 Review of Previous Works

Tracking of moving vehicles from video stream has been an active area of research both in traffic surveillance and computer vision. Recently, many single camera based works focus on the improvement of robustness and precision in different weather conditions and traffic. The most popular self-adaptive background subtraction algorithms are concluded in [1]. Non-recursive techniques, such as Frame Differencing (FD), Median Filter (MF), Linear Predictive Filter (LPF), and recursive techniques, such as Approximated Median Filter (AMF), Kalman filter (KF), Mixture of Gaussians (MoG), are evaluated in retrieval measurement (recall and precision), to quantify how well each algorithm matches with the ground truth in different weather conditions.

However, the vital problem of these single camera's solutions is still existed: the target areas will be occluded with each other because of the limitation of camera's viewpoint and angle (Figure 1). Occlusion will cause under segmentation of extracted regions and lead to miss counting and miss tracking. Most of the single camera solutions try to use target tracking to overcome the occlusion problem within certain frames. Jung [2] provided a real time system for measuring traffic parameter through adaptive background extraction, and used the concept of explicit occlusion and implicit occlusion to deal with occlusion problem. However, this work was effective only to the cases of straight line movement in highway traffic and there was no discussion on the dense traffic. Malik [3] in University of California, Berkeley, used a feature-based method along with occlusion reasoning for tracking vehicles in congested traffic. And in [5], vehicle parameters are extracted and followed by Kalman Filtering. The University of Reading has done extensive work in three-dimensional tracking of vehicles and classification of the tracked vehicles using 3-D model matching methods.

However, tracking is only effective to less dense traffic scenes and has to make sure the targets are not occluded in the first couples of frames. In addition, because of the drawback of less depth information from single camera, 3D measurement of target vehicle is extremely difficult. Stereo cameras are rarely used for traffic monitoring, since the inherent complexity of stereo algorithm and the need to solve the correspondence problem makes them difficult for real-time applications.

Another critical issue is the shadow problem. Shadows will significantly effect the extraction of target region, especially in urban areas where the targets are frequently moving in and out from shadows of roadside buildings. Geometric Project [7] uses binocular vision by two cameras registration to eliminate shadows. Color Constancy [8] defines the pixel and chromaticity and their distortion in HSV. And in [9], shadow model is concerned with the derivation of a progression of shadow-free image representation. However, none of the shadow-removing algorithms so far can provide an effective solution for traffic surveillance applications since the environmental light is changing continuously and the algorithm has to deal with the uneven shadowed regions for up to 100 meters.

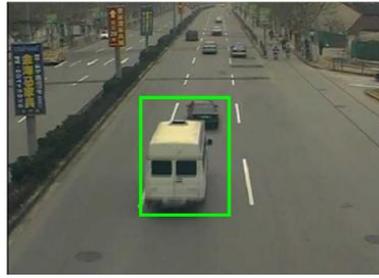


Figure 1: Example of occlusion in vision-based traffic monitoring system

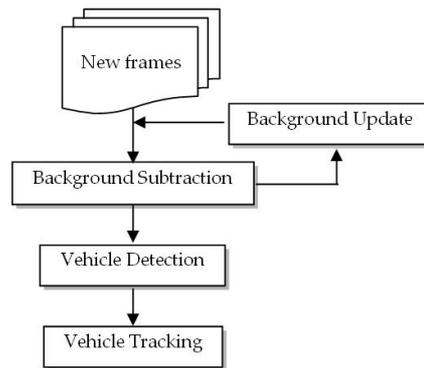


Figure 2: Block diagram of conventional traffic analysis algorithm

3 Probability Fusion Map (PFM)

To overcome the problems of single camera solution, fusion of multiple viewpoints' visual data is employed based on the concept of probability fusion map (PFM). PFM represents the probability of a vehicle being present in the scene, and is calculated by merging individual inversely mapped images within a common framework. In addition, since PFM is based on the inversely mapped image, the target 3D information including length, width and height could be directly derived through the process of blob analysis on PFM.

In the following sections, Section 3.1 describes the common processing steps of a conventional traffic analysis system. Section 3.2 describes the PFM-based solution and an effective shadowing removing algorithm is provided in Section 3.3.

3.1 Conventional Traffic Analysis

The common processing steps for single camera based vehicle detection and tracking is listed here and a block diagram of this procedure is shown in Figure 2.

Background Generation and Update: AMF or MoG based automatic background generation algorithms are generally employed to adaptively track the illumination change of background.

Background subtraction: Each new frame will be fed into Background Subtraction Module and the result will be binarized adaptively to extract the foreground regions. Morphological processing like opening/closing is necessary to segment the proper target regions.

Vehicle Detection: To improve the accuracy of identification and classification, a vehicle is recognized and analyzed by blob analysis from image coordinates and projection coordinates.

Tracking: Vehicle's position, velocity and driving state are derived in real time through vehicle tracking.

Although the conventional solutions have been enhanced in precision and efficiency, the occlusion problem still affect the results of detection.

3.2 Image data fusion solution

Our solution (PFM) focuses on data fusion provided by different viewpoints, and the probability of vehicle presence is calculated to judge vehicle and related parameters depending on fusion results. Therefore, even if one of the viewpoints were to fail because of occlusion or backlighting, the object could still be extracted correctly by some viewpoints, as the final detection result is synthesized referring to weight and probability from each viewpoint. Furthermore, it could simplify the traditional algorithms in terms of object extraction and tracking. Noise caused by weather, sudden light changing, and weak shadows, could be eliminated effectively.

The elementary of data fusion by PFM, is to set up a common inverse image for a same road scene from different viewpoints with omnidirectional monitoring. Common feature points are set on the road ahead or be referred to landmarks on the road. They are used for registering the cameras through inverse projection. And the probability of vehicle appearing on the road is calculated by fusion map of multiple views. Finally, all information about vehicle including: location, 3D size, velocity and so on, could be acquired as a direct result of PFM.

In the following sections, theoretical explanations of inverse projection will be described in subsection 3.2.1, and the calculation of PFM and vehicle extraction algorithm will be described in subsection 3.2.2 and 3.2.3 respectively.

3.2.1 Inverse Projection

The relationship among cameras is realized by transform matrix calculated by way of inverse projection referring to the feature points on the road. And the model function of inverse projection is expressed as follows:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = AP \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} P \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (1)$$

where $i = 0, 1, 2, \dots$ is the number of cameras. $(X_i, Y_i, Z_i, 1)^T$ is the position of feature in the 3D world coordinate and the corresponding position of image coordinate is $(x_i, y_i, 1)$. The matrix A is camera's intrinsic parameter matrix. Focal of lens is f and the optical center coordinate is (u_0, v_0) .

The extrinsic matrix P is acquired by following steps:

1. The 3D world coordinates $(X_i, Y_i, Z_i, 1)^T$ of feature point s are acquired on the road pavement surface. Assuming the road surface is flat, the parameter Z_i could be set as 0 and the plane position among these points are acquired by pre-knowledge or real measurements.
2. At the mean time, feature points could be marked on the scene and the positions of image coordinate are acquired.
3. Finally, the relationship between image and real world is calculated easily through Mean-least Algorithm in [6].

Figure 3 illustrates the inverse projection from different viewpoints.

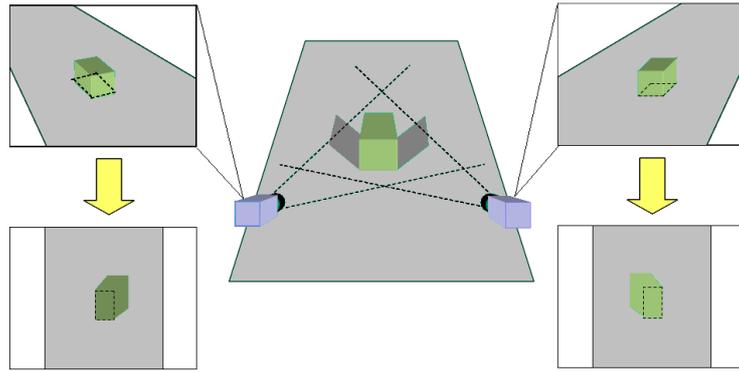


Figure 3: Inverse projection on the common framework derived from different viewpoints

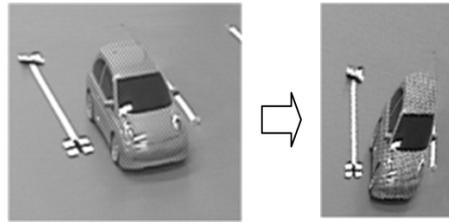


Figure 4: Inverse projection from real scene

3.2.2 Fusion of multiple views

With the solution of inverse projection and the transform matrix, every pixel on the scene from different viewpoints could be transformed by individual inverse matrix onto a common framework that is similar to a virtual platform scene from a virtual bird viewpoint. Equation (2) can be derived from Eq.(1), where Z_i is set to zero.

$$\begin{bmatrix} X_i \\ Y_i \\ 0 \\ 1 \end{bmatrix} = A^{-1} P^{-1} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{2}$$

As a result, each road scene from different viewpoints could be transformed on the virtual platform through respective inverse projection. An example of inverse projection result is shown in figure 4.

However, because of the vehicle height and camera location, the inverse projected images from different viewpoints will not corresponding to each other except for the region lying on the road. Here, we proposed a judgmental factor called PFM factor, which synthesizes the inverse projected images with considerations of Inverse Projector Factors (IPF) and Perspective Accuracy Factors (PAF).

IPF (expressed as $\alpha_i(x, y)$) is decided by the property of pixel. For example, if the pixel belongs to vehicle, its IPF value is set to 0.95, and 0.05 for other pixels. PAF is calculated based on the Projection Principle and expressed as $\beta_i(x, y)$ here.

$$\beta_i(x, y) = 1 - \kappa_i \log_{10} \frac{P_i(x, y)}{P_{i0}} \tag{3}$$

Where P_{i0} is the camera distance to the nearest feature points. κ_i is an adjustment weight parameter. $P_i(x, y)$ is the distance from target to camera calculated on the inverse projected map.

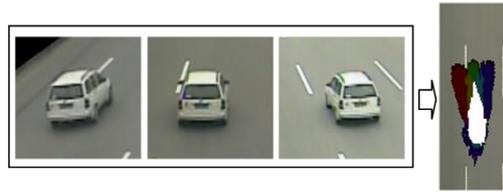


Figure 5: PFM result from 3 different viewpoints

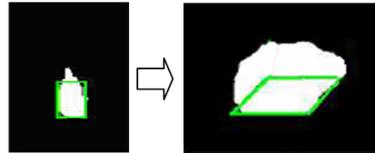


Figure 6: Bounding box in PFM is re-projected to the original images

As a result, the probability of each camera is calculated as follows:

$$P_i|_{(x,y)} = \alpha_i(x, y) \cdot \beta_i(x, y) \quad (4)$$

Finally, a target pixel's PFM value is acquired by (5).

$$P|_{(x,y)} = \sum_{i=0}^n P_i|_{(x,y)} = \sum_{i=0}^n \alpha_i(x, y) \cdot \beta_i(x, y) \quad (5)$$

Figure 5 shows the result of PFM. The red, green and blue region on the right map indicate different inverse projected result from left, middle and right camera respectively. The vehicle extraction result is shown as a white region with higher probability value calculated by PFM. Since three cameras are located behind the vehicle, there is a salient region in the rear part. The salient region could be eliminated if another camera is installed on the opposite direction.

3.2.3 Vehicle 3D Measurement from PFM

Vehicle region, as the high probability blob region in PFM can be extracted with the general morphological process and blob analysis. In order to obtain 3D information of the target, 4 corners of the target bounding box in PFM are re-projected to their original images (shown in Figure 6). And the height of vehicle is then calculated as the lower coordinate by blob analysis of rest part of object, while the length and width of vehicle is acquired by PFM directly. The 3D information of vehicle is shown on the Figure 7.

3.3 Shadow Removing

Shadow will influence the accuracy of detection and PFM solution may not work well if the vehicles are connected by shadow regions since shadow will appear from any viewpoints. Normally, shadow in traffic surveillance is projected by two kinds of object: structures on the roadside and moving objects such as vehicles or clouds. Because of outdoor monitoring, size, intensity and direction of shadow would be different when illumination changes. Fortunately, the shadow by structure could be eliminated as a road background during the background update. However, shadow which may accompany the moving objects has to be removed.

In this paper, a global filter for shadow removing is proposed which is not necessary to have previous knowledge like road surface reflection factor and shadow orientation. The main idea of this algorithm is to create

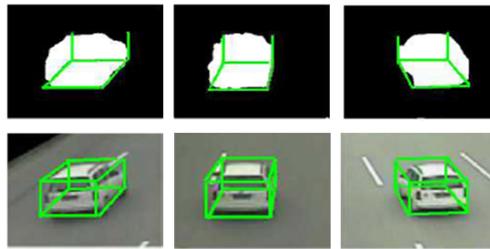


Figure 7: Vehicle 3D information extracted from PFM

a shadowing model with the intensity of typical shadow area, which could be extracted from shadow in foreground if the shadow region is confirmed. Based on the AMF algorithm, parameters of this filter will be updated automatically with illumination changing and distance from cameras.

Detail processing steps are shown as follows:

1. Background subtraction and foreground extraction as described in Section 3.1.
2. Blob analysis on the foreground extraction result and removed the blobs which have low average luminance values (most of these regions are the dark vehicles). Result is shown in Figure 8.
3. Extracting the pixels from the rest of foreground regions with the constraint that the pixel value is lower than its corresponding background pixel's luminance. And these pixels are belonging to the typical shadow regions.
4. Put the average luminance value of the shadow into non-shadow pixels as the initial shadow value to build a shadow background. Figure 9 shows one result of the shadow background.
5. Shadow background value are kept updating with Adaptive Median Filter algorithm as Eq.(6).

$$S_i(x, y) = \begin{cases} S_{i-1}(x, y) + k & \text{if } F_i(x, y) > S_{i-1}(x, y), \\ S_{i-1}(x, y) - k & \text{if } F_i(x, y) < S_{i-1}(x, y), \\ S_{i-1}(x, y) & \text{if } F_i(x, y) = S_{i-1}(x, y) \end{cases} \quad (6)$$

Where $S_i(x, y)$ is the updated shadow background value on pixel (x, y) at time i , and $F_i(x, y)$ is the newly detected shadow pixel's luminance. k is the updating weight.

6. Extracted foreground regions on each frame are subtracted again with the shadow background to remove shadows.

The proposed shadow removing filter updates itself with the real shadow pixels' luminance value, which does not rely on the previous knowledge like road pavement refectance and shadow orientation. Therefore, it is robust to the variation on size, intensive and orientation of shadow area.

4 Experiment and Result

To evaluate the performance of our approach, outdoor experiments were carried out. We tested different width and shape of roads. Figure 10 shows the extraction and tracking result of a three-lane road in Shanghai, China. The left part shows the extracted vehicles and the right part is the result of PFM. It is obvious to see that even the target are occluded visually from one viewpoint, our approach can still obtain the correct result since the target regions are well isolated on PFM.



Figure 8: (a) Background subtraction result. (b) Extracted foreground. (c) Extracted shadow regions

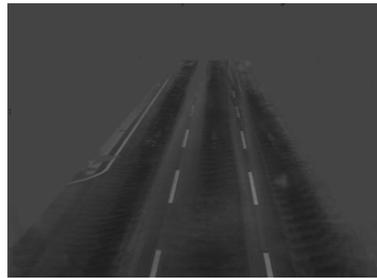


Figure 9: Shadow background generated for shadow removing

Table 1 compares the result of vehicle account between conventional method using single camera and our method using three cameras. PFM shows very high accuracy in almost all time segments and is much stable than single camera based solution since our approach eliminates the effect of occlusion and shadows. And Table 2 shows the accuracy of vehicle speed by PFM estimation.

The performance of PFM is related to two factors: one is the density of vehicle on the road and the other is arrangement of cameras. The occlusion vehicle may be distinguished easily just using two cameras if the traffic flow is in low density. On the other hand, PFM will be failed even with 2 or 3 viewpoints in the cases of heavy traffic congestion. Further discussion on optimizing the number of cameras and the distribution of camera location will be done in our future work.

Table 1: Comparison result between PFM and conventional solution

Time segments (minutes)	Ground truth	PFM solution		Conventional solution	
		Detected vehicles	Recall rate	Detected vehicles	Recall rate
0-10	215	213	99.5 %	231	92.7 %
10-20	182	179	98.3 %	191	89.5 %
20-30	151	151	100 %	159	94.4 %
30-40	158	155	98.1 %	174	90.0 %
40-50	214	213	99.5 %	239	88.6 %
50-60	154	155	99.2 %	168	90.2 %
Total	1074	1066	99.3%	1162	91.8%

Table 2: Accuracy of speed estimated by PFM

No.	Truth (km/h)	PFM (km/h)	Accuracy
1	41	45.9	88.0 %
2	64	65.5	97.7 %
3	54	59.7	89.9 %
4	64	62.1	97.2 %
5	36	38.1	94.2 %
6	64	59.0	92.2 %
7	54	55.5	97.2 %

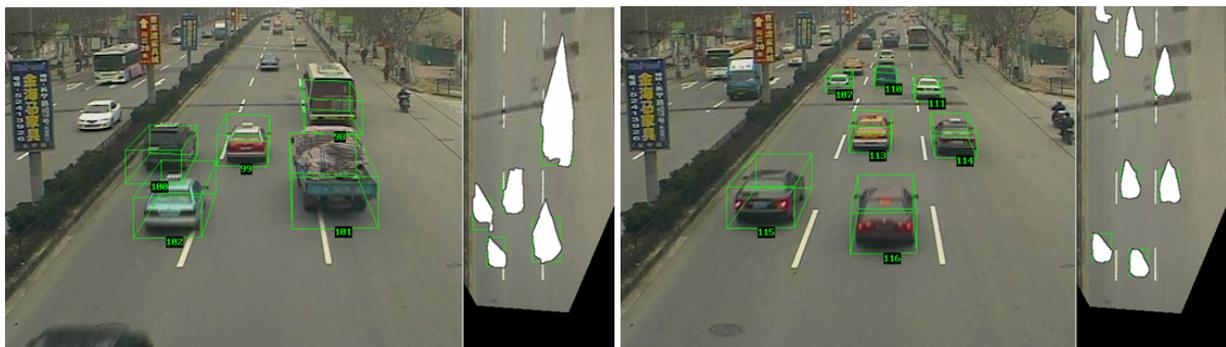


Figure 10: Fusion result of vehicle identification and tracking

5 Conclusion

A novel solution of vehicle occlusion and 3D measurement for traffic monitoring is proposed in this paper. Comparing with single camera based conventional methods in traffic monitoring, our approach fuses video data from different viewpoints into a common probability fusion map (PFM) and extracts targets. The proposed PFM concept is efficient to handle and fuse data in order to estimate the probability of vehicle appearance, which is verified to be more reliable than single camera solution by real outdoor experiments. An AMF based shadowing modeling algorithm is also proposed in this paper in order to remove shadows on the road area and extract the proper vehicle regions. Real outdoor experiment result shows the excellent performance of our approach.

6 Acknowledgment

This work was partially supported by the MEXT Grant under the grant number of 18700184.

References

- [1] S. Sen-Ching, C.K. Cheung, "Robust Techniques for Background Subtraction in Urban Traffic Video", *Proceedings of SPIE*, 2004.
- [2] Y.K. Jung, Y.S. Ho, "Traffic Parameter Extraction using Video-based Vehicle Tracking", *Proc. IEEE Intelligent Transportation System Conf.'99*, 1:746-769, 1999.

- [3] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, "A real-time computer vision system for measuring traffic parameters", *IEEE Proc. of Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, 1:496-501, 1997.
- [4] G. Surendra, M. Osama, F.K.M. Robert, and P.P. Nikolaos, "Detection and Classification of Vehicles", *IEEE Transactions on Intelligent Transportation System* 3(1), 2002
- [5] Z. Qiu, D. Yao, "Kalman Filtering Used in Video-Based Traffic Monitoring System", *Journal of Intelligent Transportation Systems* 10(1), 2006
- [6] Z. Zhang, "A flexible new technique for camera calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11):1330-1334, 2000
- [7] K. Onoguchi, "Shadow Elimination Method for Moving Object Detection", *Proc. Int'l Conf. Pattern Recognition*, 1(2):583-587, 1998
- [8] R. Cucchiara, C. Grana, M. Piccardi, A. Parati and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information", *Proc. IEEE Intelligent Transportation System Conf.*, 1:334-339, 2001
- [9] G.D. Finlayson, S.D. Hordley, C. Lu and M.S. Drew, "On the Removal of Shadows from Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(1), 2006
- [10] J.M. Ferryman, S.J. Maybank and A.D. Worrall, "Visual Surveillance for Moving Vehicles", *Intl. Journal of Computer Vision*, 37(2):187-197, 2000