Stereo Image Processing for ADAS and Pre-Crash Systems

Sergiu Nedevschi[1], Radu Danescu[1], Ciprian Pocol[1], Marc Michael Meinecke[2]
[1]Technical University of Cluj Napoca, Computer Science Department
e-mail: Sergiu.Nedevschi@cs.utcluj.ro
[2]Volkswagen AG, Group Research Driving Assistance Electronics
e-mail: Marc-Michael.Meinecke@volkswagen.de

***Abstract*-- Stereovision is a technique for range detection able to provide accurate 3D description of the environment in real time. In the last years due to the progresses in computer and camera technology the stereovision have became a robust alternative for environment sensing in driving assistance systems. The dense stereo sensor presented in this paper answers to the divers requirements of the urban scenario through a multitude of detection modules, built on top of a hybrid (hardware plus software) dense stereo reconstruction engine. The sensor is able to detect, track and classify clothoid and non-clothoid lanes, drivable areas in the absence of lane markings, cars, pedestrians and road maintenance signs. The hybrid stereovision engine and the proposed detection algorithms allow accurate sensing of the demanding urban scenario at a high frame rate being suitable for use in ADAS. The presented sensor is tested for precrash applications.**

***Index terms*—stereovision, ADAS, precrash**

I. INTRODUCTION

Advanced driving assistance systems (ADAS) are a complex multidisciplinary research field, aimed at improving traffic efficiency and safety. A realistic analysis of the requirements and of the possibilities of the traffic environment leads to the establishment of several goals for traffic assistance, to be implemented in the near future (ADASE, INVENT, PREVENT, INTERSAFE) including: high-way, rural and urban assistance, intersection management, pre-crash.

While there are approaches to driving safety and efficiency that focus on the conditions exterior to the vehicle (intelligent infrastructure), it is reasonable to assume that we should expect the best results from the in-vehicle systems. Traditionally, vehicle safety is mainly defined by passive safety measures. Passive safety is achieved by a highly sophisticated design and construction of the vehicle body. The occupant cell has become a more rigid structure in order to mitigate deformations. The frontal part of vehicles has been improved as well, e.g. it became specially designed “soft” areas to reduce the impact in case of a collision with a pedestrian. In the recent decades a lot of improvements have been done in this field.

Similarly to the passive safety systems, primitive active safety systems, such as airbags, are only useful when the crash is actually happening, without much assessment of the situation, and sometimes against the well-being of the vehicle occupants. It has become clear that the future of the safety systems is in the realm of the artificial intelligence, systems that sense, decide and act.

Sensing implies a continuous, fast and reliable estimation of the surroundings. The decision component takes into account the sensorial information and assesses the situation. For instance, a pre-crash application must decide whether the situation is of no danger, whether the crash is possible or when the crash is imminent, because depending on the situation different actions are required: warning, emergency braking or deployment of irreversible measures (internal airbags for passenger protection, or inflatable hood for pedestrian protection). While warning may be annoying, and applying the brakes potentially dangerous, deploying non-reversible safety causes permanent damage to the vehicle, and therefore the decision is not to be taken lightly. However, in a pre-crash scenario it is even more damaging if the protection systems fail to act. Therefore, it is paramount that the protection systems act when needed, and only when needed, a decision that cannot be taken in the absence of reliable sensor data.

The sensorial systems for driving assistance (highway and urban) are today the focus of large, joint research projects, which combine active and passive sensors, GPS navigation, and telematics. Projects such as CARSENSE (www.carsense.org) INVENT (www.invent-online.de), PREVENT (www.prevent-ip.org), bring together car manufacturers and research partners for the common goal of solving the driving assistance problem.

In order to provide support for these applications, a sensorial system must provide an accurate and continuously updated model of the environment fitted for high level reasoning. The environment description should include:

- Lane detection/ Lane parameters estimation
- Navigable channel detection and channel parameters estimation in crowded environments
- Vehicle detection and tracking
- Detection of fixed (non-moving) obstacles
- Pedestrian detection and tracking.

There are many types of sensors that can be used for advanced driving assistance systems. The most known are:

- Long range radar: with a range of 1 to 200 m, and a response time of around 40 ms, it is a highly accurate ranging sensor, with a narrow field of view, suitable for detection of radar-reflecting targets such as vehicles in highway environments.

- Short/mid range radar: having a working range of 0-80 m, a fast response time, high accuracy and a medium width field of view, it is suitable for near range detection of vehicles in crowded urban scenarios. Both near range and far range radars have an increased reliability when detecting moving objects.

- Laser scanner: a high precision ranging sensor, working in near or far distance ranges, it is not limited to the metallic surfaces like the radar, but has considerable difficulty with low albedo objects.

- Monocular video sensors: employed in the visual or in the infrared light spectrum, the visual sensors can have a high field of view and can extract almost any kind of information relevant for driving assistance. The main problem of these sensors is that it cannot rely on accurate 3D information, having to infer it indirectly, usually poorly.

A stereovision sensor adds the 3D information to the visual, thus becoming the most complex and complete sensor for driving assistance. It is capable of detecting any type of obstacle that falls inside its adjustable field of view, the road and lane geometry, the free space ahead, and it is also capable of visual classification, for pedestrian recognition.

II. RELATED STEREOVISION WORK

The stereovision-based approaches have the advantage of directly estimating the 3D coordinates of an image feature, this feature being anything from a point to a complex structure. Stereovision involves finding correspondents from the left to the right image, and the search for correspondence is a difficult, time demanding task, which is not free from the possibility of errors. Obstacle detection techniques involving stereovision use different approaches in order to make some simplifications of the classic problem and achieve real-time capabilities. For instance, [9] uses stereovision only to measure the distance of an object after it has been detected from monocular images, [10] detects the obstacle points from their stereo disparity compared to the expected disparity of a road point, [11] detects obstacle features by performing two correlation processes, one under the assumption that the feature is part of a vertical surface and another under the assumption that it is part of a horizontal surface, and comparing the quality of the matching in each of the cases. A stereovision system that uses no correspondence search at all, but warps images instead and then performs subtraction, is presented in [12].

Processing 3D data from stereo (dense or sparse) is a challenging task. A robust approach can prove of great value for a variety of applications in urban driving assistance. There are two main algorithm classes, depending on the space where processing is performed: disparity space-based and 3D space-based. Most of the existing algorithms try to compute the road/lane surface, and then use it to discriminate between road and obstacle points.

Disparity space-based algorithms are more popular because they work directly with the result of stereo reconstruction: the disparity map. The "v-disparity" [11] approach is well known and used to detect the road surface and the obstacles in a variety of applications [12]. It has some drawbacks: is not a natural way to represent 3D (Euclidian) data, it assumes the road is dominant along the image rows, and it can be sensitive to roll angle changes.

The 3D space algorithms have also become popular among the researchers in recent years. Obstacle detection and 3D lane estimation algorithms using stereo information in 3D space are presented in [5], [6], ego pose estimation algorithms are presented in [4] [3], and unstructured environment estimation algorithms are presented in [7] and [8].

III. Dense Stereovision-Based Stereo System for Urban Driving Assistance

The Technical University of Cluj Napoca has been actively involved in the field of stereovision for driving assistance systems. A general geometry stereovision algorithm, based on software correlation around edge points, is described in [5]. This system was designed for high accuracy obstacle detection at large distances in highway scenarios, and it was later adapted for 3D lane detection, as described in [6].

The urban traffic scenario proved to be way more complex that the highway, and the edge-based, software correlation-driven stereo system was not up to the task, due to the large processing time required by correlation, which left little time for more complex algorithms, and to the sparseness of the stereo data which proved to be insufficient for accurate discrimination between obstacles in the urban traffic environment.

The solution to the urban traffic was the development of the dense stereovision-based algorithms [13].

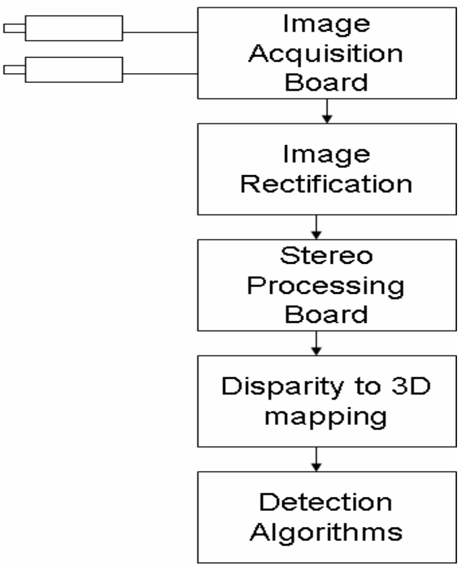

Fig. 1. The dense stereovision system architecture

The hardware acquisition system (fig. 1) includes two cameras with 2/3" (1380x1030) CCD sensors and 6.5 mm fixed focal length lenses, allowing a horizontal field of view (HFOV) of 72 [deg]. The cameras are mounted on a rigid rig with a baseline of 320 [mm]. The images are acquired at full resolution with digital acquisition board with a maximum frame rate of 24 fps.

The images are further enhanced by lens distortion correction and rectified in order to fulfill the dense stereo reconstruction requirements (canonical images). A down-sampling step is used to adapt the image size to the stereo processing board parameters (512 pixels width) and to minimize the noise introduced by the digital rectification and image correction. The whole process is reduced to an image warping approach performed in a single step (fig. 1) using reverse mapping and bilinear interpolation.

The 3D reconstruction of the scene is performed using a dedicated hardware board. The input of the board consists in two rectified images and the output can be either a disparity or a Z map (expressed in the left camera coordinate system). Our system uses 3D points set for scene representation; therefore the preferred output is the Z map. Using the Z coordinate value, the X and Y coordinate can be computed and then transformed into the car coordinate system using the extrinsic camera parameters.

With the current system setup a detection range optimally suited for the urban environments is obtained:

-minimum distance: 0.5 m in front of the ego car (approximately 2.5 m in front of the cameras) – the near range distance is limited by the baseline, focal length and maximum disparity allowed by the hardware board.

-delimiters of the current lane (considered approximately of 3,5 m wide) are visible at 1.0 m;

-reliable detection range: 0.5 … 35 m, with a maximum detection range (up to witch 3D points can be reconstructed) of 50 m

-large horizontal field of view, 72 degrees

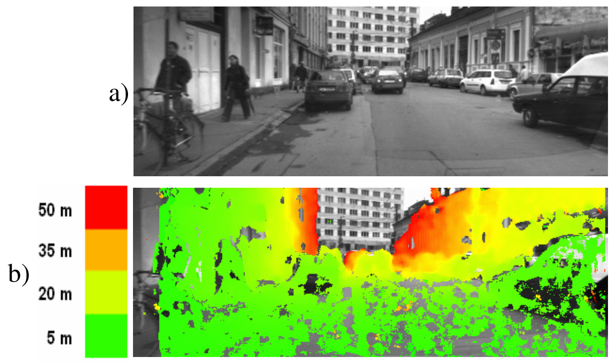

Fig. 2. Depth map generated from stereovision.

The stereovision system is used to provide two types of environment description: a structured description consisting of parametrical lanes, tracked cuboids, and pedestrians, and an unstructured description based on a dense elevation map with drivable and non-drivable areas, suitable for very difficult scenarios.

IV. Lane detection

The lane detection system is organized as an integrator of multiple sensors. Instead of having multiple physical sensors, we have multiple detection stages, which all deliver results that will be used to update the lane model state parameters. The cycle begins with the prediction, and continues with all the detection algorithms, until the final update. When one algorithm updates the lane state, the resulted estimation becomes the prediction for the next stage. In this way, we can insert any number of algorithms into the processing chain, or we can temporary disable some of them, for testing or speedup purposes.

The detection of the pitch angle and of the vertical curvature is done using the same stereovision-based algorithm that we have used for the highway scenario [6], adjusted for the distance range of the urban traffic. This step will mark each edge point that has 3D information associated to it as either “road point” or “above road point”. The road points are of interest in lane detection, the others are used for the obstacle detection routines.

Lane Marking Point Extraction: together with edge detection, stereo reconstruction and road/above road labeling done by the vertical profile detection, the lane marking point extraction (classification) algorithm is part of the feature extraction methodology for urban lane estimation. This algorithm detects lane markings as pairs of 3D road points of similar in value but opposing in sign gradients, placed at the proper distance. This step is independent of the prediction, as it has to have universal, model-free application.

The core of the model-based lane estimation process is the linear model matching. This algorithm fits two line segments (for the left and right lane border) to the perspective-projected road points, under several constraints that will ensure that these two segments are very likely to be the 2D projection of a section of the lane. First the linear matching is attempted to a range segment close to the ego vehicle, to ensure a minimum detection in restricted visibility conditions. If the near range linear model matching succeeds, the same algorithm is run for the next road section, in order to refine the estimation of curved roads.

Free Form Left Border / Right Border Detection – These routines are independent of the model-based prediction and of the linear model matching algorithms, but they rely heavily on the lane marking extraction results. Each lane border is estimated independently as a chain of 3D points. The results of these routines are used for updating the lane model parameters, but they can also be used as standalone output.

A detailed description of the lane detection algorithms can be found in [14].

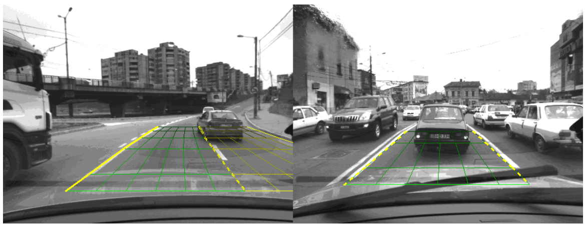

Fig. 3. Lane detection in urban environments

V. Obstacle Detection and Tracking

The obstacle detection algorithm receives the vertical road profile from road detection, and uses this profile to identify the obstacle points. The only 3D points used by the obstacle detection algorithm are those situated above the road and below the height of the ego car.

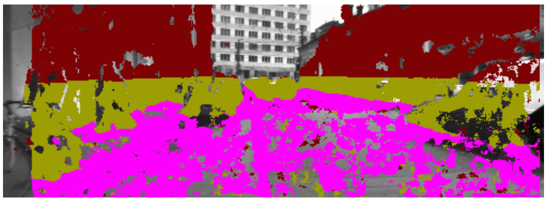

Fig. 4. Road points, relevant points for obstacle detection and irrelevant high points

The local density and vicinity of the relevant 3D points is analyzed in a special compressed top view space. The compressed space keeps a constant density of the 3D points and neutralizes the error of the reconstructed depth. On the compressed space, a labeling algorithm is applied, on the cells with high density of points, determining the occupied areas.

The occupied areas are fragmented into obstacles that are suitable for the cuboidal model. For this fragmentation, the visible shape (towards the camera) of the 3D points is analyzed and two criteria are used: the shape must have no concavities and the cuboid fitted on the shape must not contain significant free (drivable) space.

The orientation of the obstacles is determined, when possible.

Most of the processing is done on the compressed space which concentrates the useful information of the set of 3D points and leads to a fast computation.

After the cuboids are generated, they are used as measurement in a model-based tracking algorithm. Tracking is initiated for objects that fit the size requirements for vehicles or pedestrians, and that have at least three consistent measurements from consecutive frames. The measurements are associated to the tracks corner by corner, partial associations and updates being allowed. This ensures a robust behaviour in case of occlusions and overtakes.

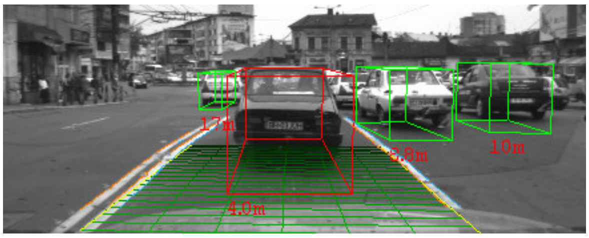

Fig. 5. Tracked obstacles

VI. Pedestrian Recognition

The pedestrian recognition algorithm is applied to the cuboids resulted from obstacle detection. The following steps seem to be of interest at the moment:

1.Simple form features extraction: The height of grouped objects and their base radius are simple, if not powerful features, that can be used for pedestrian detection, especially when rejecting too large or too small objects. These features are taken directly from the output of the grouping module.

2.Tracking: coarse objects are tracked across multiple frames using a Kalman filter–based multi–tracker. This step ensures stable detection and easier optical field computation (because of the inherent motion compensation resulting from the tracking). All tracked objects are considered as possible pedestrian candidates.

3.Object speed extraction: From the tracker's output, and knowing the ego vehicle speed, object speeds can be computed. Currently, we use 2 types of speeds for our classifier: one that is parallel to the ego vehicle's axis, the longitudinal speed, and one that is perpendicular to the first.

4.Depth Masks: Object masks are computed for all tracked objects. Only points for which their 3-D coordinates lie inside the tracked cuboids are considered. This step is important as it eliminates spurious background points and deals with partial occlusions.

5.Optical Flow: A pyramidal, corner-based optical flow detection algorithm is used to compute optical flow in all corner points belonging to tracked objects. Only optical flow vectors starting and ending on non-masked points are considered.

6.3-D Velocity: The true 3-D velocity of the considered points is computed, using the 2-D optical flow, stereo depth and frame time-stamps.

7.PCA: Principal component analysis is used to find the principal direction of the 3D velocity field variation for each individual object. Variance is smoothed across frames, to increase its stability. We call the magnitude of this principal component a "motion signature". This motion signature is much smaller for non–pedestrians as compared to pedestrians, and is thus a powerful feature for pedestrian detection.

8.Motion History: Using tracking information, we record the motion signature across multiple frames, to determine its history

9.Motion Spectrum: We compute the spectrum of the motion signature variation in time. Pedestrians display a typical periodic motion signature, while other types of objects display only impulsive noise.

10. Bayesian Classification: A naive Bayesian classifier is used to combine the extracted features. The prior pedestrian probability is also an input of the Bayesian classifier.

A detailed description of the approach is found in [15]

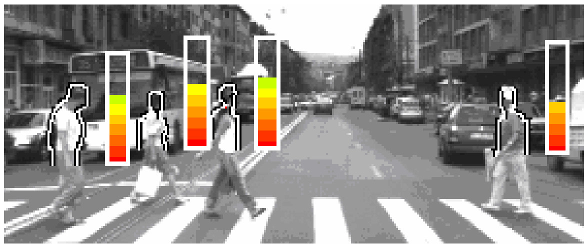

Fig. 6. Pedestrian recognition results

VII. Unstructured Environment Description

There are some urban scenarios where the 3D lane cannot be detected, especially when not enough lane delimiters exist (ex. road crossing). An alternative method must be used to detect elevated areas (obstacles), regions where the ego vehicle cannot be driven. Complementary, the obstacle-free road areas can be considered as drivable.

The dense stereo engine usually reconstructs most of the road surface points even if lane markings are not present. Thus, the surface of the road can be computed by fitting a geometric model to the 3D data. A (bird-eye rectangular, 3x35 meters) region of interest of the 3D space can be represented similar to a digital elevation map. An image of elevations is formed, with each pixel (cell) having the intensity proportional to the 3D height. If a cell has more than one 3D point, then the greatest height is used. Morphological dilation is used to fill voids and compensate for the perspective effect (the 3D space gets sparser with the depth).

The road surface is fitted to the cell heights in a restricted region in front of the vehicle, where most of the 3D points belong to the road. The RANSAC technique is used for fitting.

The final classification of the cells into drivable/non drivable areas is performed using the deviation of the height from the road model, using a height uncertainty measure computed from the stereo reconstruction uncertainties and the camera perspective model. A detailed description of the algorithms can be found in [8]. The elevation map and road are also used for curb detection, as described in [9].

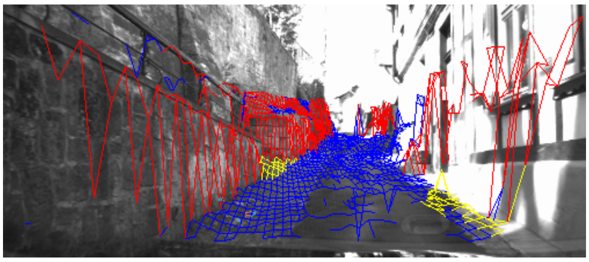

Fig. 7. Drivable areas (blue), non-drivable areas (red) and curb areas (yellow) in a complex non-structured environment

VIII. ADAS APPLICATIONS OF THE DENSE STEREO SENSOR

The dense stereovision system is able to detect all visible obstacles. The measurement error is dependent on the distance, ranging from 0.1 m for close objects to 1 m for the objects near the limit of the detection range (40 m with the 72 degrees wide field of view). The reliable detection combined with the fast response time (frame rate is 20 fps) make it a reliable sensor for pre-crash applications, and the capability of pedestrian detection ensure that the proper measures can be taken for their protection.

The stereo sensor's capabilities make it suitable for several other ADAS applications. The ability to track the objects, especially those in front of the vehicle, recomment it for ACC applications, the wide field of view makes it suitable for go inhibit applications (where it is vital to detect last minute crossing pedestrians), and the lane estimation capabilities make it suitable for lane following applications. The stereo system is a complex integrated sensor, suitable for integrated ADAS applications.

REFERENCES

[1] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection on non flat road geometry through V-disparity representation," in IEEE Intelligent Vehicles Symposium, Versailles, June 2002, pp. 646–651.

[2] A. Broggi, C. Caraffi, P. Paolo Porta, and P. Zani, "The Single Frame Stereo Vision System for Reliable Obstacle Detection used during the 2005 DARPA Grand Challenge on TerraMax", in IEEE Intelligent Transportation Systems Conference, Toronto, Canada, September 17-20, 2006, pp. 745-752.

[3] A. Sappa, D. Gerónimo, F. Dornaika, and A. López, "Real Time Vehicle Pose Using On-Board Stereo Vision System", Int. Conf. on Image Analysis and Recognition, LNCS Vol. 4142, Portugal, September 18-20, 2006, pp. 205-216.

[4] M. Cech, W. Niem, S. Abraham, and C. Stiller, "Dynamic ego-pose estimation for driver assistance in urban environments". In IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004, pp. 43-48.

[5] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, R. Schmidt, T. Graf, "High Accuracy Stereo Vision System for Far Distance Obstacle Detection", IEEE Intelligent Vehicles Symposium (IV2004), Parma, Italy, pp. 292-297, 2004.

[6] S. Nedevschi, R..Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, "3D Lane Detection System Based on Stereovision", IEEE Intelligent Transportation Systems Conference (ITSC), Washington, USA, pp.161-166, 2004.

[7] F. Oniga, S. Nedevschi, M-M. Meinecke, T-B. To, "Road Surface and Obstacle Detection Based on Elevation Maps from Dense Stereo", The 10th International IEEE Conference on Intelligent Transportation Systems, Sept. 30 - Oct. 3, 2007, Seattle, Washington, USA.

[8] F. Oniga, S. Nedevschi, M-M. Meinecke, "Curb Detection Based on Elevation Maps from Dense Stereo", The 3rd International IEEE Conference on Intelligent Computer Communication and Processing, pp.119-125, 6-8 Sept. 2007, Cluj-Napoca, Romania.

[9] M. Bertozzi, A. Broggi, A. Fascioli, and S. Nichele, "Stereo Vision-based Vehicle Detection", in Procs. IEEE Intelligent Vehicles Symposium 2000, pages 39-44, Detroit, USA, October 2000.

[10] U. Franke, D. M. Gavrila, S. Görzig, F. Lindner, F. Paetzhold and C. Wöhler, "Autonomous Driving Approaches Downtown", IEEE Intelligent Systems, vol.13, nr.6, pp. 40-48, 1998.

[11] T. A. Williamson, "A high-performance stereo vision system for obstacle detection", September 25, 1998, CMU-RI-TR-98-24. Robotics Institute Carnegie Mellon University. Pittsburg, PA 15123.

[12] M. Bertozzi and A. Broggi, "GOLD: a Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection", IEEE Trans. on Image Processing, 7(1):62-81, January 1998

[13] S. Nedevschi, R. Danescu, T. Marita, F. Oniga, S. Sobol, C. Tomiuc, C. Vancea, M.M. Meinecke, T. Graf, T.B. To, M.A. Obojski, "A Sensor for Urban Driving Assistance Systems Based on Dense Stereovision", Proceedings of IEEE Intelligent Vehicles Symposium, (IV 2007), 2007, Istanbul, Turkey.

[14] R. Danescu, S. Nedevschi, M.M. Meinecke, T.B. To, "Lane Geometry Estimation in Urban Environments Using a Stereovision System", Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC 2007), Seattle, USA, 2007

[15] S. Bota, S. Nedevschi, "Walking Pedestrians Detection Using Motion Field and Dense Stereo", IEEE 2nd International Conference on Intelligent Computer Communication and Processing, pp 45-51, 1-2 September, Cluj-Napoca, Romania,2006