2008 IEEE Intelligent Vehicles Symposium
Eindhoven University of Technology
Eindhoven, The Netherlands, June 4-6, 2008

Obstacle Detection and Classification fusing Radar and Vision

M. Bertozzi, L. Bombini, P. Cerri and P. Medici
VisLab – Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Parma, ITALY
http://vislab.it
{bertozzi,bombini,cerri,medici}@vislab.it

P. C. Antonello and M. Miglietta
CRF - Centro Ricerche Fiat
I-10043 Orbassano(TO), ITALY
maurizio.miglietta@crf.it
pierclaudio.antonello@crf.it

Abstract— **This paper presents a system whose aim is to detect and classify road obstacles, like pedestrians and vehicles, by fusing data coming from different sensors: a camera, a radar, and an inertial sensor. The camera is mainly used to refine the vehicles' boundaries detected by the radar and to discard those who might be false positives; at the same time, a symmetry based pedestrian detection algorithm is executed, and its results are merged with a set of regions of interest, provided by a Motion Stereo technique.**

The tests have been performed in several environments and traffic situations, their results showed how the vision based filtering provides an effective reduction of radar's false positives; furthermore, the regions of interest detected by the Motion Stereo algorithm, truly improves the pedestrian detector's performance again by keeping low the number of detection errors.

The system has been shown during the APALACI-PReVENT European IP final demonstration[1] in September 2007 in Versailles (France).

I. Introduction

This paper describes an obstacle detection and classification system using different methods to detect regions of interest. It exploits a vehicle detection algorithm, based on fusion of camera images and radar data [1], [2], to detect vehicles, while a pedestrian detection algorithm [3], [4] is exploited to detect the presence of potential pedestrians and, finally, a motion stereo technique is also used to find obstacles and to refine pedestrian detection results.

Radar is robust against bad weather, rain and fog; it can measure speed and distance of an object, but it does not provide enough data points to detect obstacle boundaries, and experimental results show that radar is not reliable to detect small obstacles like pedestrians. Vision-based system can cope with this lack in localization and, moreover, other tasks can be performed using the same sensor.

Some vision-based systems [5] for obstacle avoidance exploit stereo sensors. They performs a 3D world reconstruction of the scene through the triangulation of homologous points. In special setup calibration or using image rectification is possible to look for the same feature on the same row between the images couple with good performance and low computational resource.

[1]The work described in this paper has been developed in the framework of the Integrated Project APALACI - PReVENT, a research activity funded by the European Commission to contribute to road safety by developing and demonstrating preventive safety technologies and applications.

However, the engineering of stereo-based systems on vehicles is complex, due to the excessive cost, the connection between cameras, computation engine, and miscalibration issues.

A potential solution of this problem can be found in the use of motion stereo: a technique that allows the recovery of three dimensional informations from motion as binocular stereo vision.

Two different approaches can be used to perform motion stereo: 3-D reconstruction and warped image comparison.

In the first approach, points of interest are tracked and matched over the frames: they can be chosen and tracked using Kanade-Lucas-Tomasi technique [6] or, simply, using optical flow on strong edges, for example corners [7], [8]. Under the assumption of static world, it is possible to extract the vehicle ego-motion and to obtain a 3D scene reconstruction.

The difficulty of tracking reliable features and subsequent error propagation decrease the performance of this method. This approach provides good performance to recover Structure from Motion (SFM) like in park assistant systems, but is not generally used in more dynamic scenes like motorways. An improvement of this approach is described in [6] where a vision-radar fusion is developed and radar is used to classify features associate to static or dynamic obstacles.

In the second approach, the ego-motion is instead computed according to rotation and translation parameters provided by the inertial sensors, thus no more features tracking is needed; hence this approach is quite fragile, since it heavily relies on fine camera calibration and good odometric data, in order to provide reliable results.

Aubert et al proposed an approach of motion stereo for obstacle detection using warped images [9], that are at each cycle compared against the previous one. Since the warped images are computed under the flat-world assumption and the ego-motion compensation is applied using the data provided by odometry and a gyroscope, the differences between that two images are then attributable to vertical objects not satisfying the initial assumption; in this way it is possible to compute a V-disparity like image in order to detect obstacles. In Batavia et al [10] only edges are warped and the predicted position is compared with the current one. Pitch fluctuation produced by vehicle vibration are rejected using edge tracking, called 1-dimensional optic flow.

This paper presents a different approach for the motion

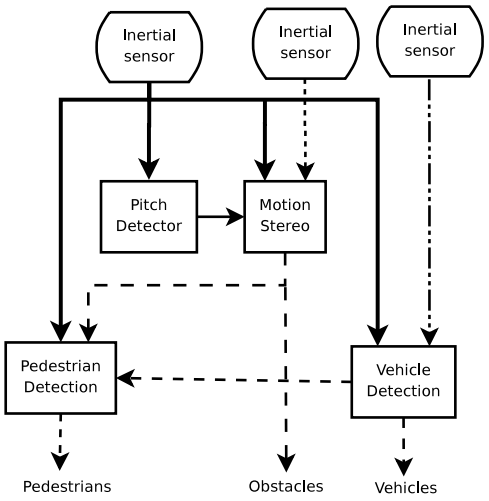

Fig. 1. Obstacle detection and classification algorithm

stereo: images are not directly warped, but the operation is instead executed on its inverse perspective mapping. So, once image's bird's eye view is computed, it is rotated and translated according to odometric data provided by the inertial sensors.

Hence, this bird's eye approach doesn't allow to consider the whole image; however the loss of information is negligible, since, because of the flat-world assumption, the only pixels of interest are those which are under the horizon line.

Generally if there is not enough camera motion between two frames, images differences are very low and displacement of vertical objects are comparable with image noise. In the opposite case, large motion makes feature detection difficult. Big interest on motion stereo are given on key-frame selection [11]; but while key-frame selection is a good approach for static scenarios, that's not true when considering moving obstacles. Proposed approach do not use key-frames and all the frames are used to generate the bird's eye view of the scene using a reinjection technique.

This paper is organized as follows: in section II the system organization is presented (a pitch detector for structured environment, the Motion Stereo Inverse Perspective Mapping technique, a vehicle detection system based on radar, and a pedestrian detection based on image edges symmetry), in section III some results in different situation and environments are discussed, while section IV ends the paper with some final remarks.

II. Algorithms

The sensor system is composed by two medium range (40 meters) 24 GHz radar, a gyroscope, and a commercial grey-level firewire camera.

The whole algorithm flowchart is presented in image 1 and in the following sections different subsystems implemented in the system are described.

A. Pitch Detector

One of the problems that must be faced during the processing of images for intelligent vehicles is the miscalibrations

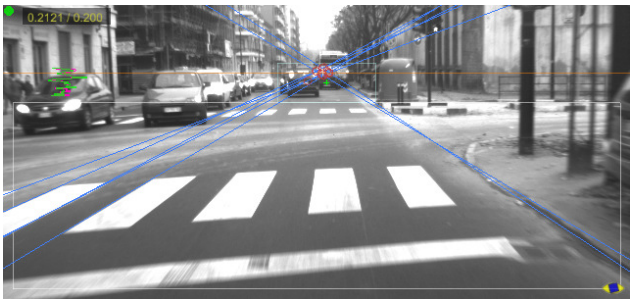

Fig. 2. Pitch detection algorithm: lines hints (blue), intersecting points (red circles), vanishing point suggested by calibration (green cross) and vanishing point provided by algorithm (red cross and orange line).

introduced by the vehicle's movements. A possible solution is the real time computations of the calibration parameters for each image. In particular, it can be noticed that a large amount of edges are oriented toward the vanishing point of images and this feature can be used for computing the horizon position and thus computing the correct vision system pitch that is the most important calibration parameter used in the following processings.

Several approaches have been already discussed for retrieving the vanishing point using a lane detection algorithm [12]. In such a case, the result depends on the presence of lane marking on the road and on their correct detection.

Other research groups use Hough transform to directly detect vanishing point [13]. In this case all edges on images are involved in vanishing point generation and therefore this approach is too computational heavy for our hardware constraints.

In the system presented in this paper, vertical edges computed by means of a Sobel operator are used. The sign of edges is taken into account but only edges with an absolute value higher than a given threshold are considered. Beginning from the bottom part of the image, contiguous edge pixels having the same sign are joint to form a large cluster of pixels

For each clusters bigger than a given threshold (40 pixels in the current setup) a linear regression that minimizes the square deviation on columns is computed in the form $ax + by + c = 0$. The size of the cluster is used as confidence value.

The lines are then compared each others and their intersection angle is computed. If this angle is below a 0.02 radians threshold the two lines are merged and a new line with a new confidence equal to the sum of two originals is generated. The new line parameters are computed as follows:

$$\begin{aligned} a &= a_1 \frac{1-\alpha}{\sqrt{a_2^2+b_2^2}} + a_2 \frac{\alpha}{\sqrt{a_1^2+b_1^2}} \\ b &= b_1 \frac{1-\alpha}{\sqrt{a_2^2+b_2^2}} + b_2 \frac{\alpha}{\sqrt{a_1^2+b_1^2}} \\ c &= c_1 \frac{1-\alpha}{\sqrt{a_2^2+b_2^2}} + c_2 \frac{\alpha}{\sqrt{a_1^2+b_1^2}} \end{aligned} \quad (1)$$

where $\alpha \in [0, 1]$ is based on the relative confidence of the two original lines.

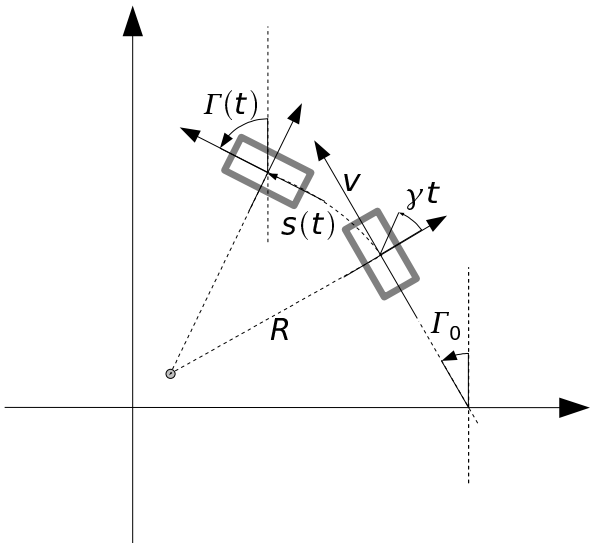

Fig. 3. Approximated model.

After this step only the 7 lines with the highest confidence are evaluated for determining the vanishing point. The intersection points generated by those lines falling in a region around the vanishing point measured by calibration are considered. They are averaged and weighted according to the minimum confidence of the intersecting lines (results in figure 2) in order to compute the current vanishing point.

The camera pitch ϑ is calculated using vanish point row v_p :

$$\vartheta = \arctan\left[\left(1 - \frac{v_p}{v_0}\right)\tan\frac{\beta}{2}\right] \tag{2}$$

where β is the vertical field of view and v_0 row coordinate of principal point.

Confidence of intersection is low when no perspective features are present, or when the vehicle is approaching to curves. When the confidence values are too low, vanishing point is not considered for pitch detection and an horizon tracking is performed instead.

The edges histogram computed on rows is compared against the same obtained from previous images . A Vertical offset is computed minimizing the differences and used to compute the horizon position stabilizing the image. This second approach is generally affected by a drift error and therefore the first approach is preferred when possible.

For both approaches the edges computed where a vehicle is detected by fusion vision and radar, described in section II-C, are not considered.

B. Motion Stereo IPM

The gyroscope attached to the CAN bus provide yaw rate γ and vehicle speed v. A simple vehicle model has been chosen; a more sophisticated model has been considered too, anyway due to odometry imprecision the model easier to manage has been preferred.

Therefore, a vehicle is approximated as a point (see figure 3) on the rotation center where gyroscope is installed,

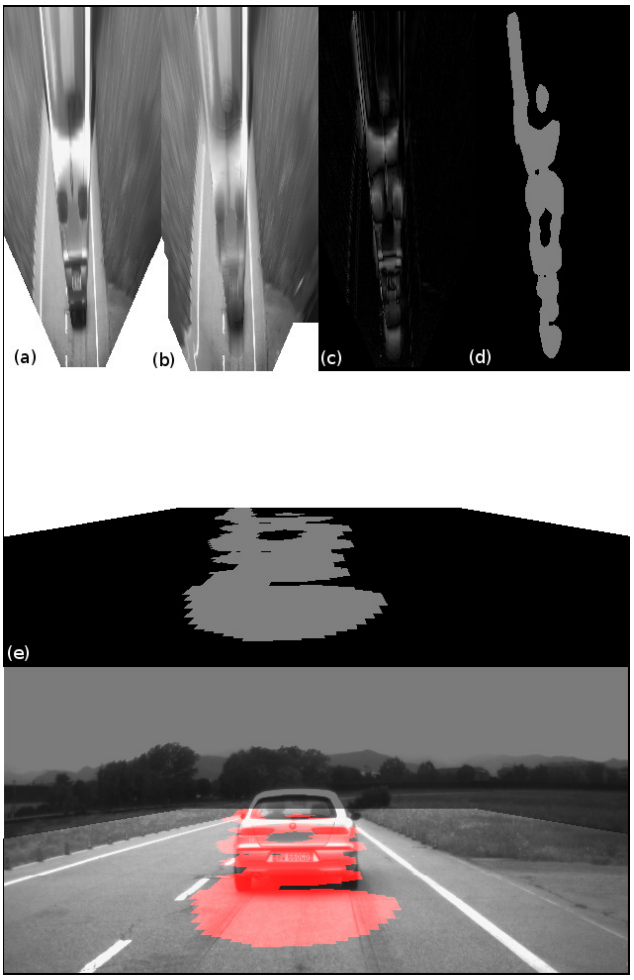

Fig. 4. Motion stereo step: *(a)* bird's eye view, *(b)* warped background, *(c)* differences, *(d)* binarized version, and *(e)* reprojected

with speed v in direction Γ. For this point the following equation is used:

$$\begin{aligned}\Gamma(t) &= \Gamma_0 + \gamma t \\ \vec{s}(t) &= \vec{v}t\end{aligned} \tag{3}$$

In the time interval $[0, T]$ between two image frames, γ and v are considered as constant, and therefore it is possible to integrate those equations between 0 and T. Under those assumptions, integrals of translations and rotation can be written as

$$\begin{aligned}\Delta x &= S\cos\left(\Gamma_0 + \tfrac{\gamma}{2}T\right) \\ \Delta y &= S\sin\left(\Gamma_0 + \tfrac{\gamma}{2}T\right) \\ \Gamma &= \Gamma_0 + \gamma T\end{aligned} \tag{4}$$

where Δx is the translation along X axis, Δy the translation along Y axis, Γ_0 the initial orientation of the vehicle, Γ the final orientation, and S is defined as:

$$S = 2\frac{v}{\gamma}\sin(\frac{\gamma}{2}T) \overset{\gamma\to 0}{\simeq} vT \tag{5}$$

Using off-line calibration (for both intrinsic and extrinsic parameters) and pitch calculated in real-time, from source

Fig. 5. Vehicle detection using radar and vision fusion: the orange box contains the final result of vehicle detection, the yellow cross is the refined radar beam associated to vehicle and the blue crosses correspond to discarded beams.

Fig. 6. The pedestrian detection algorithm and filtering provided by motion stereo. Boxes are proved by pedestrian detection algorithm, light green boxes are definitely detected, dark green boxes are the rejected by motion stereo

image an inverse perspective mapping view [14] A is computed (fig. 4.a). The size of this bird's eye view is chosen considering odometry accuracy.

A bird's eye view of the background B, generated during the processing (fig. 4.b), is rotated and translated according to movement of vehicle following equation 4 where a 2×2 rotation matrix and translation vector is used. In the first frame B is placed equal to A. In fact the following processing steps can be performed only after the first reference background has been computed.

A difference image C (fig. 4.c) is then produced computing absolute differences between A and B pixels. The presence of differences can be used as an indicator of potential obstacles: for each pixel of C a square area (3×3) centered on it is considered; the average value m_i of all the pixels in that area is computed and a threshold ξ is applied on m_i in order to produce a binarized image D (fig. 4.d) that encodes the result of motion stereo.

Moreover, the binarized image D is used to mask the image A that is then used to update the background view B as follows:

$$B'_i = w_i A_i + (1 - w_i) B_i \tag{6}$$

where weights w_i are calculated as:

$$w_i = \begin{cases} 0.25 & \text{if } m_i > \xi \\ 1 & \text{if } m_i < \xi \end{cases} \tag{7}$$

In such a way a new background image B' is computed considering the portion of image A where obstacles are not detected. This image will be used as background B in the following iteration.

The image D can be reprojected in the perspective view (fig. 4.e) and used as region of interest for the pedestrian detection algorithm.

The motion stereo processing can detect both moving and still obstacles when the vehicle on which the vision system is installed is moving. When the vehicle is still, the proposed system falls in the background subtraction algorithm and only moving obstacles can be detected.

C. Vehicle Detection using Radar

Radar data are used to locate areas of interest on images that can contain vehicles [15]. Two scanned radars with a 24 GHz frequency mounted above the front bumper are used.

The first step of the algorithm maps radar objects into the image reference system, using the same perspective mapping transformation already computed for motion stereo. Then, an area of interest is built around each radar point and furtherly investigated. In order to simplify and speed up the following steps of the algorithm and to delete details of too close vehicles, all these areas are resampled to a fixed size.

Vehicle search in these areas is performed computing vertical symmetry; only vertical binarized edges are used in order to further reduce execution time. Symmetry is computed for every column, on different sized bounding boxes whose height matches the image height and with a variable width ranging from one pixel to a given maximum number of pixels. Axis corresponding to a high symmetry content are considered. Vehicle width is computed as the first width at which the symmetry value is greater than a threshold. In order to determine the position of the base of the vehicle, the algorithm looks for the vehicle shadow.

When all radar data have been examined, all the boxes framing the detected vehicles are resampled to their original size and mixed together, and a series of filters is applied in order to delete false detections. Reversing the inverse perspective mapping transformation, real width and position of vehicles can be computed. In the final output, the radar is used to provide distance while vision outputs position and width, so that the radar precision on distance measurement and the vision refinement ability are capitalized together.

Figure 5 shows a simple example of detection.

D. Pedestrian Detection

An algorithm for human shape detection based on symmetry is used to detect potential pedestrians [3], [4].

Specific characteristics of pedestrians, such as vertical symmetry and strong presence of edges, are used to select interesting regions likely to contain pedestrians. More precisely, the acquired image is scanned and symmetries and edges are extracted; since a human shape is characterized by a strong vertical symmetry, symmetrical areas with a specific aspect ratio identify possible candidates.

The knowledge about system calibration and, specifically, current pitch is used to further refine and filter the detected areas of attention. In fact, human body may present a

sufficiently high symmetry to be detected, but the detected area may not be precise. This generally happens to the legs, which can be in different positions. In these cases, the detected area of attention does not comprise the bottom portion of the human shape and therefore the knowledge of the position of the road surface can be used to extend the bottom portion of the area of attention. Size and perspective constraints are also adopted to ease and speed up the search.

This approach is performed using a single camera and is highly affected by the presence of a cluttered or noisy background. Therefore, the final list of areas of attention is filtered according to the motion stereo and vehicle detection results to remove false positives. Only areas of attention that are also detected as obstacle by the motion stereo approach (see fig. 6) are considered while the others are discarded. The surviving areas of attention are further filtered using the result of vehicle detection; in fact, areas detected where also a vehicle has been detected, are removed as false positives.

III. Experiments

The system has been tested on several sequences in different environmental conditions (urban and rural scenarios), vehicle speeds, and traffic situations.

The processing timings of the different algorithm steps for 640×300 images using standard Pentium 4 at 3.0 GHz are:

algorithm step	timings
Pitch detector	4 ms
Motion Stereo	43 ms
Vehicle detection	1 ms
Pedestrian Detection	83 ms

Therefore, the whole system is able to reach a 7 Hz rate.

The motion stereo algorithm is able to correctly detect moving and stationary vertical objects. The results show only few false positives thanks to the fact that this approach is not aimed at providing a complete 3D reconstruction of the scene, but only a list of regions where obstacles are present. Unfortunately, miscalibrations due to odometry error or pitch estimation failures do not guarantee a perfect matching of road plane over some frames. In this case false positive on motion stereo image can appear (typical examples are due to lane marking mismatches). The use of more aggressive filter to remove these artefacts has been tested but leads to miss small objects.

Concerning the pedestrian detection subsystem, the use of the results provided by the Motion Stereo processing allows to filter out a large amount of false positives that are due to a cluttered background. The most critical situations are encountered with the presence of obstacles that feature a size and a symmetry content similar to a human shape (i.e. poles or tree trunks). It has to be considered that the radar data can not be used for pedestrian detection, since the human shape generally does not reflect enough energy.

The vehicle detection subsystem has been demonstrated to be the most reliable functionality. Thanks to the fusion between radar and vision, the final result shows correct distance and size estimations.

Figure 7 shows some examples of the vehicles and pedestrians detections superimposed onto original images; the detected vehicles are marked using an orange marker, while detected pedestrians are marked using a bright green box. A dark green box is also used for marking non valid pedestrians detections that are filtered out by the fusion steps. The pedestrians search area corresponds to the grey strip in the images. The pedestrian detection results contain some false positives (figures 7.d, 7.f, 7.g, and 7.h) that are anyway validated since they are contained in areas where obstacles are detected by the motion detection subsystem. Figure 7.a shows a false positive for pedestrian detection that is correctly filtered out by the fusion with vehicle detection results.

IV. Conclusion

Fig. 8. The CRF prototype and the camera position.

This paper presented an obstacle detection and classification system for road environment. This system acquires images from a monocular camera installed on a vehicle, radar data from a 24 GHz radar installed in the front part of the vehicle, and cinematic informations from an inertial sensor.

Artificial vision is used to refine the pitch of the vision system in order to compensate the vehicle movements. Thanks to the availability of inertial data, vision is also exploited for the detection of obstacles by means of a motion stereo technique.

Radar data and images, acquired by the camera, are fused together to detect the presence of vehicles and estimate their position and size.

Vision is also used to preliminary detect the presence of pedestrians in a specific region of interest in the image; the result of this subsystem is filtered out using data coming from the motion stereo and vehicle detection subsystems in order to remove false positives due to complex backgrounds.

The whole system has been installed on a CRF prototype (see figure 8) and demonstrated during the final exhibition of the European funded project PReVENT-APALACI (18–22 September 2007, Versailles, France).

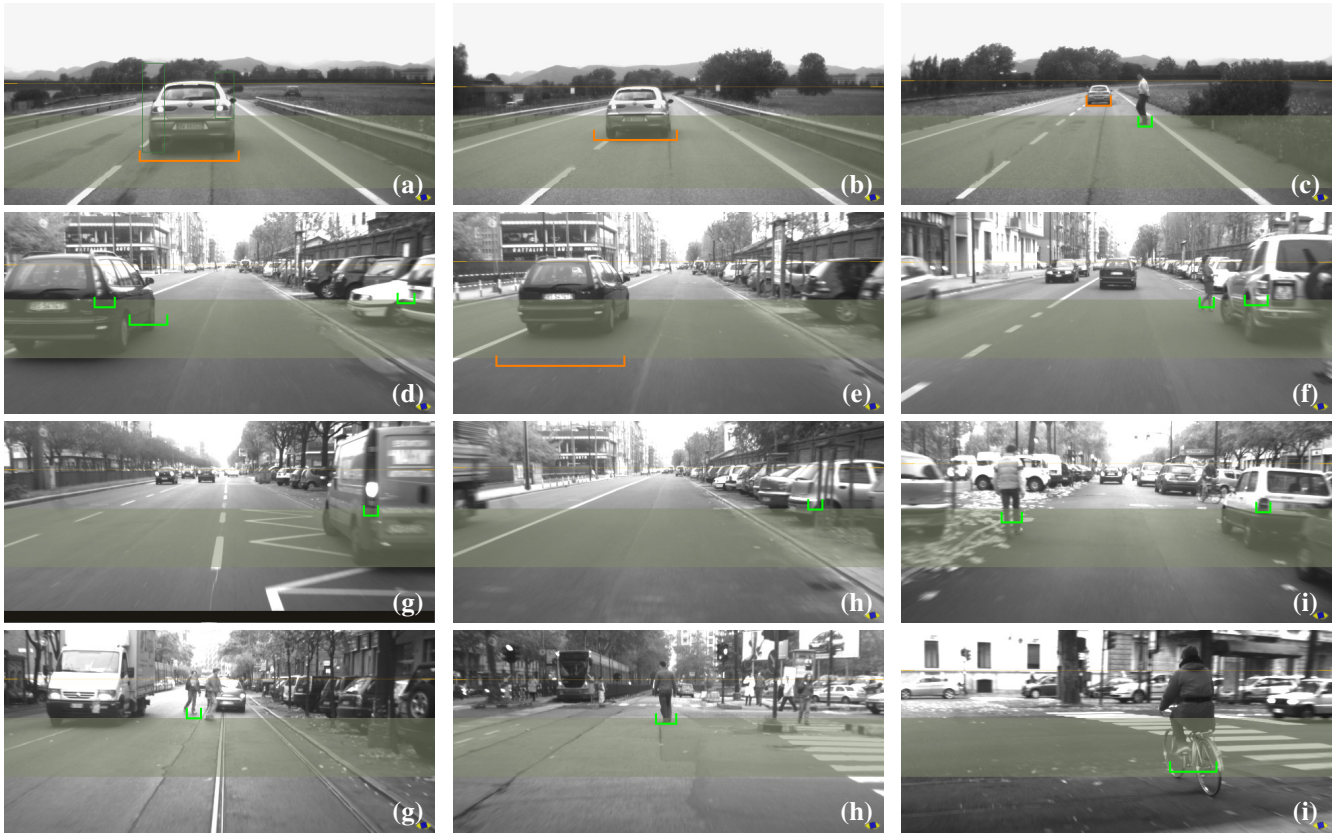

Fig. 7. Result of whole processing in different environmental conditions.

REFERENCES

[1] A. Broggi, P. Cerri, and P. C. Antonello, "Multi-Resolution Vehicle Detection using Artificial Vision," in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, pp. 310–314.

[2] L. Bombini, P. Cerri, P. Medici, and G. Alessandretti, "Radar-Vision Fusion for Vehicle Detection," in *Procs. Intl. Workshop on Intelligent Transportation*, Hamburg, Germany, Mar. 2006, pp. 65–70.

[3] M. Bertozzi, A. Broggi, A. Fascioli, and M. Sechi, "Shape-based Pedestrian Detection," in *Procs. IEEE Intelligent Vehicles Symposium 2000*, Detroit, USA, Oct. 2000, pp. 215–220.

[4] A. Broggi, M. Del Rose, A. Fascioli, I. Fedriga, and A. Tibaldi, "Stereo-based Preprocessing for Human Shape Localization in Unstructured Environments," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 410–415.

[5] A. Broggi, C. Caraffi, P. P. Porta, and P. Zani, "A Single Frame Stereo Vision System for Reliable Obstacle Detection during Darpa Grand Challenge 2005," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2006*, Toronto, Canada, Sept. 2006, pp. 745–752.

[6] T. Kato and Y. N. A. I. Masaki, "An obstacle detection method by fusion of radar and motion stereo," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, pp. 182–188, 2002.

[7] K. Fintzel, R. Bendahan, C. Vestri, S. Bougnoux, and T. Kakinami, "3d parking assistant system," *Intelligent Vehicles Symposium, 2004 IEEE*, pp. 881–886, 14-17 June 2004.

[8] E. Mouragnon, F. Dekeyser, P. Sayd, M. Lhuillier, and M. Dhome, "Real time localization and 3d reconstruction," *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, pp. 363–370, 17-22 June 2006.

[9] R. Alix, F. L. Coat, and D. Aubert, "Flat world homography for non-flat world on-road obstacle detection," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 310–315.

[10] P. H. Batavia, D. A. Pomerleau, and C. E. Thorpe, "Overtaking Vehicle Detection using Implicit Optical Flow," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems'97*, Boston, USA, Nov. 1997, pp. 729–734.

[11] D. Nistér, "Frame decimation for structure and motion," in *SMILE '00: Revised Papers from Second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*. London, UK: Springer-Verlag, 2001, pp. 17–34.

[12] S. Nedevschi, C. Vancea, T. Marita, and T. Graf, "On-line calibration method for stereovision systems used in vehicle applications," *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pp. 957–962, 2006.

[13] D. Schreiber, B. Alefs, and M. Clabian, "Single camera lane detection and tracking," *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pp. 302–307, 13-15 Sept. 2005.

[14] M. Bertozzi, A. Broggi, and A. Fascioli, "Stereo Inverse Perspective Mapping: Theory and Applications," *Image and Vision Computing Journal*, vol. 8, no. 16, pp. 585–590, 1998.

[15] G. Alessandretti, A. Broggi, and P. Cerri, "Vehicle and Guard Rail Detection Using Radar and Vision Data Fusion," *IEEE Transaction on Intelligent Transportation System*, vol. 8, no. 1, pp. 95–105, Mar. 2007.