

Novel Method for Vehicle and Pedestrian Detection Based on Information Fusion

Fernando García, Arturo de la Escalera and José María Armingol

Abstract A novel approach for vehicle and pedestrian detection based on data fusion techniques is presented. The work fuses information from a 2D laser scanner and a computer camera, to provide detection and classification of vehicles and pedestrians in road environments. Thanks to the data fusion approach, the limitations of each sensor are overcome. Thus a reliable system is provided, fulfilling the demands of road safety applications. Classification is performed using each sensor independently. Laser scanner approach is based in pattern matching and vision approach is based in the classical Histogram of Oriented Gradients features approach. A higher stage performs data fusion using Kalman Filter and Global Nearest Neighbors.

1 Introduction

Roads are the transport with more fatalities. It is estimated that more than 40,000 people die every year in Europe in traffic accidents. During the latest years, the advances in both vehicles and roads helped to reduce the numbers of deaths in roads, but there is still a lot of work to be done.

F. García (✉) · A. de la Escalera · J. M. Armingol
University Carlos III of Madrid, Madrid, Spain
e-mail: fegarcia@ing.uc3m.es

A. de la Escalera
e-mail: escalera@ing.uc3m.es

J. M. Armingol
e-mail: armingol@ing.uc3m.es

Fig. 1 Intelligent vehicle based on visual information (IVVI) 2.0



In recent years, efforts have focused in creating applications that use the advances in information technologies to help to increase the security in roads. One example of these applications are the Advance Driver Assistant Systems (ADAS), which purpose is to help and warn the driver in case of a hazardous situations.

Among all the available sensors for road safety applications, it is difficult to find a system able to fulfill the strong requirements of these applications. In the present article an approach based in data fusion is presented. This system tries to overcome the limitations of each sensor by fusing the information provided. A classic vision based ADAS is enhanced by adding a 2D laser scanner, providing pedestrians and vehicles detection with a high positive rate. The resulting application is already available in the Platform IVVI 2.0 (Fig. 1).

2 State of the Art

Fusion approaches can be divided according to the level where fusion is performed:

In **Low Level** approaches raw data is fused, creating a new data set, which combines information from different sources. Usually these methods depend on the technology or sensor used. In computer vision, stereovision is an example of low level fusion. Images from two cameras are used to create a more complete set of information able to provide 3D information; in [1, 2] this information is used to provide pedestrian detection.

Medium Level fusion requires preprocessing stages for each sensor separately, creating a feature set, based in features form the different sensors, this set is used to perform the final classification. In [3, 4]authors present works combining the features and performing classification by different ways: Naïve Bayes, GMMC, NN, FLDA, and SVM.

High Level fusion approaches perform detection and classifications for each sensor independently and a final stage combines them [5]. Performs pedestrian detection, using visual Adaboost detection and Gaussian Mixture Model (GMM) for laser scanner, a Bayesian decisor is used to combine detections at high level. In [6] pedestrians are detected using laser scanner by multidimensional features; Histograms of Oriented Gradients (HOG) features and Support Vector Matching (SVM) for computer vision detection; finally Bayesian model provides high level fusion.

The work presented here is an example of high level detection, with independent classifiers for pedestrian and vehicle detection, providing a robust system able to fulfill requirements of safety applications. Besides, the independence of the low level classifiers allow to use them separately even in extreme situations were one of them is not available.

3 Low Level Detection

As it was remarked before, the first stage of the approach consisted in a low level detection, based in the information given by the laser scanner and the camera independently. Later a higher level stage fuses this information.

Several configuration were tested for obstacle detection: pattern based monocular camera detection, stereo-based obstacle detection and laser scanner obstacle detection. The final configuration used the laser scanner to provide obstacle detection to both systems, laser scanner and camera. The higher trustability of the laser scanner helps to reduce the amount of false positives in the vision, since only the regions in the image where there is certainty given by the laser are check. Besides, laser scanner provides obstacle detection faster and more efficient in comparison to the stereo based system.

3.1 *Laser Scanner Detection*

The laser scanner is mounted in the bumper of the vehicle, thus the delayed detections given should be corrected according to the movement of the vehicle. Translation and rotation should be performed according to the information provided by a GPS with inertial measurement available in the platform (1) and Fig. 2a. Later, shapes of the detected obstacles should be estimated, this shape reconstruction is based in polylines [7], as it is shown in Fig. 2b.

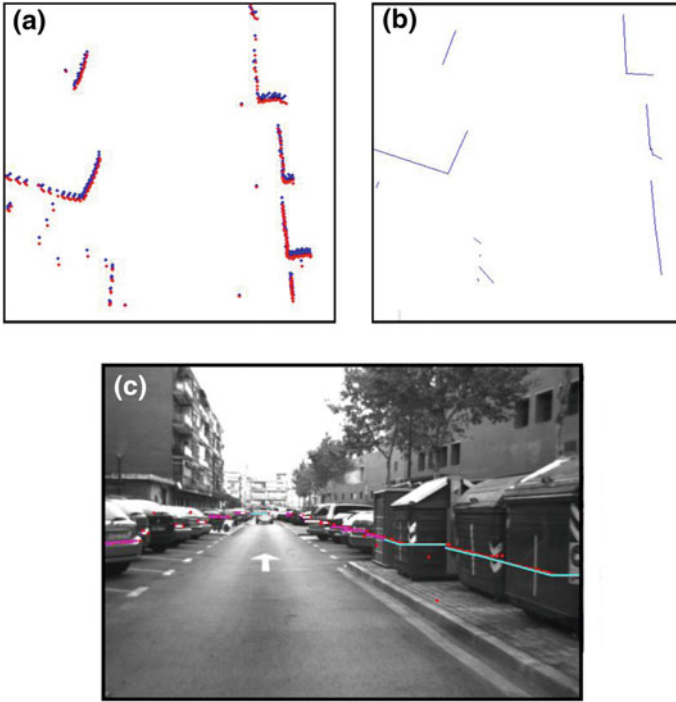


Fig. 2 Vehicle movement compensation of laser scanner information. **a** Shows the detection points, in *blue* the raw data and in *red* the compensated. **b** Shows the shape reconstructed after the movement compensation. **c** Shows the alignment of the laser scanner data and the image

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T_v + T_0 \right), \text{ with } T_v = \begin{bmatrix} vT_i \cdot \cos(\Delta\theta) \\ vT_i \cdot \sin(\Delta\theta) \\ 0 \end{bmatrix}, T_0 = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}, \text{ and} \quad (1)$$

$$R = \begin{bmatrix} \cos(\Delta\delta) & 0 & \sin(\Delta\delta) \\ 0 & 1 & 0 \\ -\sin(\Delta\delta) & 0 & \cos(\Delta\delta) \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & \cos(\Delta\varphi) & -\sin(\Delta\varphi) \\ 0 & \sin(\Delta\varphi) & \cos(\Delta\varphi) \end{bmatrix} \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ \sin(\Delta\theta) & \cos(\Delta\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where $\Delta\delta$, $\Delta\varphi$ and $\Delta\theta$ corresponds to the increment of the Euler angles roll, pitch and yaw respectively for a given period of time T_i . Coordinates (x, y, z) and (x_0, y_0, z_0) are the Cartesian coordinates of a given point after and before respectively to the vehicle movement compensation. R is the rotation matrix, T_v the translation matrix according to the velocity of the vehicle, T_0 the translation matrix according to the position of the laser and the inertial sensor, v is the velocity of the car, T_i the time between the given point and the first one in a given scan. Finally, (x_t, y_t, z_t) is the distance from the laser scanner coordinate system to the inertial measurement system.

After the shape reconstruction, the classification is performed using a pattern matching approach. The different obstacles that are possible to be differentiated are: big obstacles, small obstacles, road borders, L shaped, pedestrians or vehicles. The most important ones for this application are pedestrians and vehicles. Both detections were based in different patterns, obtained thanks to detailed studies in the movement of the different obstacles.

3.1.1 Vehicles

The pattern is based in the delay of the spots given by the laser scanner detection, this way, this spots provide a given pattern that depends on the movement of the vehicle. Thanks to this special pattern, classification can be performed as well as information about the movement of the vehicle estimated (Fig. 3). Deeper information of this algorithm is provided in [7].

3.1.2 Pedestrians

A pattern for pedestrians was defined based in the position of legs of the pedestrians (Fig. 4).

In this pattern, three polylines are presented, and the angles that connect the polylines are included within the limits of $[0, \pi/2]$ (2).

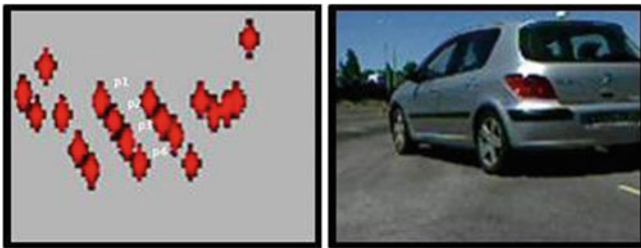


Fig. 3 Typical pattern for vehicle detection

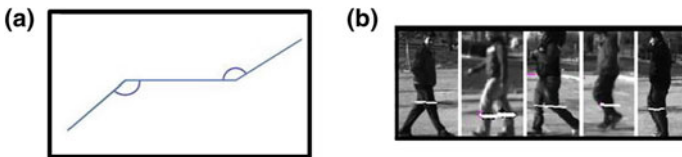


Fig. 4 Pattern for pedestrian detection. a The pattern used. b Examples of the pattern with real pedestrians

$$\text{Similarity} = \frac{2\theta_1}{\pi} \cdot \frac{2\theta_2}{\pi} \quad (2)$$

where θ_1 and θ_2 are the angles that connect two consecutive lines.

This similarity is computer between two consecutive angles and if the case arises that they match the pattern it is labeled as a possible pedestrian.

Finally a simple tracking stage is added at this level, to track the movement of the obstacle providing reliable detection. A voting scheme that takes into account the last 10 frames was created. Furthermore several filters are added to avoid false positives, these filters check wrong behavior along time, such as impossible movements, velocities or accelerations.

3.2 Computer Vision

As it was remarked, the vision approach makes use of the reliability of the laser scanner to provide region of interest. Thus only the obstacles detected by the laser scanner, with size similar to the object to be found, are provided to the vision system, allowing reducing the parts of the image where the algorithms perform the search. Thus computation cost and false positives are reduced (Fig. 5).

Coordinate change should be performed using pin-hole model and accurate extrinsic calibration (3) and (4) to provide information from the laser scanner to the camera coordinate system:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \left(\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} + T \right) \quad (3)$$

where R is the rotation matrix shown in (1) corresponding to the Euler angles that represent rotation between the different coordinate systems. T is the translation

vector $T = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}$ that corresponds to the distance between the coordinate systems,

(x_c, y_c, z_c) are the camera coordinates and (x_l, y_l, z_l) the laser scanner coordinates.



Fig. 5 Examples of obstacle sets for pedestrians (*left*) and vehicles (*right*)

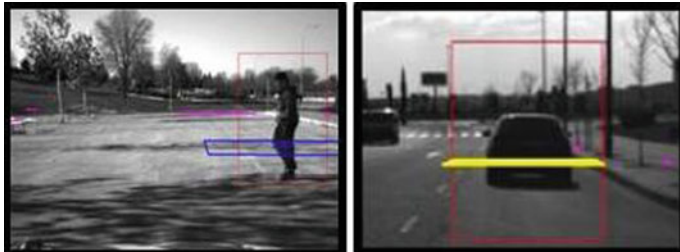


Fig. 6 Visual detection examples for pedestrians (*left*) and vehicles (*right*). *Red boxes* represents for visual detections, *blue boxes* represents laser scanner based pedestrian detections and *yellow boxes* vehicle detections

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ z_c \\ y_c \end{bmatrix} \tag{4}$$

where u_0 and v_0 are the center coordinates of the camera coordinate system in pixels. (u, v) are the coordinates in the camera coordinate system in pixels. x_c, y_c and z_c are the Cartesian coordinates from camera. And f is the focal length.

The computer vision algorithms used were different according to the obstacle to be found:

- **Vehicles.** Haar-Like features based approach with cascade classifiers was used (Fig. 6, right). The common features that can be found in the back of the vehicle allowed to used this fast algorithm originally used for face detection [8].
- **Pedestrian.** Based in Histogram of Oriented Gradients (HOG) features (Fig. 6 left). This approach is classical in Intelligent Vehicles, and was proposed in [9].

4 Fusion

Fusion stage retrieves the detections from both subsystems, providing fused detections (tracks). Kalman filter was used to estimate the movement of the different obstacles.

Two kind of tracks were defined, consolidated and non-consolidated. First corresponds to those tracks detected by both sensors, second correspond to the detected by a single sensor.

The association technique used to match the new detection with the old ones (tracks) was Global Nearest Neighbors (GNN), based in the distance between the estimation of the track and the position of the detections. A distance based in the stability of the measurements was defined (5) and the gate used to eliminate the non likely pairs was based in a square approach (6).

$$d^2 = \frac{(x_i - \bar{x})^2}{\sigma_x^2} + \frac{(y_i - \bar{y})^2}{\sigma_y^2} + \ln(\sigma_x \sigma_y) \quad (5)$$

$$K_{Gl} \sigma_r \quad (6)$$

where σ_r is the residual standard deviation and K_{Gl} is a constant that was empirically chosen, d is the computed distance between previous and presented tracks to be associated. And (σ_x, σ_y) the appropriate values of covariance matrix of Kalman Filter.

A M/N policy was used to create and eliminate the tracks, thus a given track is created after M detection and is eliminated if N number of frames provide no matches for the track. In the case of a non-consolidated track which is not corroborated by the other sensor it is considered a false positive.

5 Results

Different test were performed in both urban and interurban scenarios with more than 10,000 frames in real road situations. Result comparison is showed in Table 1.

The results proved that the system was able to enhance the low level approaches providing better results. In the low level detections, it was very interesting the high positive rate obtained with the limited information provided by the laser scanner, mainly in the case of vehicles. On the other hand the amount of misdetections was also very high.

It has to be remarked that the training process created for the camera approaches were performed taking into account the results of the laser scanner. Due to the high amount of false positives given by the laser scanner based system, the vision system was trained to obtain the lowest false positive rate possible. Besides, the camera systems did not include a tracking stage, thus the positive results expected were lower. This situation is visible in the case of vision based vehicle detection. Although the amount of no-detection errors for vehicles was high, it was proved that all the vehicles in the images were positively detected. Thus even in the worst case scenario any vehicle is detected after one or two frames. Finally the low

Table 1 Results

	Camera		Laser scanner		Fusion	
	% of positive detections	% of misdetections (per frame)	% of positive detections	% of misdetections (per frame)	% of positive detections	% of misdetections (per frame)
Pedestrians	72.97	5.27	74.56	13.3	77.69	3.11
Vehicles	47.72	1.13	91.03	8.19	88.25	2.59

amount of false positives in the visual approaches allowed overcoming the excessive number of these errors in the laser scanner approach.

6 Conclusions

Finally we can conclude, given the results presented in Table 1, that the fusion process allows to combine information from the camera and the laser scanner and helps to overcome the limitations of the system, enhancing the capacities of classic ADAS and providing reliable detection, ready to be included in a real road application.

Limitation inherent to each sensors and its algorithms are overcome thanks to the use of data fusion approach. First, computer vision have the trustworthiness limitations due to the unstructured information. The trustworthiness of the laser scanner detection allows to reduce the false positives. On the other hand, the limited information given by the laser scanner is completed thanks to the high amount of data given by the laser scanner.

Although nowadays a road application based in both technologies would represent a high cost, due to the high cost mainly of the laser scanner in comparison to the computer based approaches. Modern cars are already available with this technology to detect obstacles and perform avoiding maneuvers.

Acknowledgments This work was supported by the Spanish Government through the Cicyt projects (GRANT TRA2010-20225-C03-01) and (GRANT TRA 2011-29454-C03-02). CAM through SEGAUTO-II (S2009/DPI-1509).

References

1. Bertozzi M, Broggi A, Felisa M, Ghidoni S, Grisleri P, Vezzoni G, Hilario C, Del Rose M (2009) Multi stereo-based pedestrian detection by means of daylight and far infrared cameras. In: Hammoud RI (ed) Object tracking and classification beyond the visible spectrum. Springer, Berlin, pp 371–401
2. Hilario C, Collado JM, Armingol JM, de la Escalera A (2005) Pedestrian detection for intelligent vehicles based on active contour models and stereo vision. Computer aided systems theory at “EUROCAST 2005”. pp 537–542
3. Premebida C, Ludwig O, Silva M, Nunes U (2010) A cascade classifier applied in pedestrian detection using laser and image-based features. Intelligent Transportation Systems (ITSC), 13th International IEEE Conference on 19–22 Sept. 2010, pp.1153–1159
4. Premebid C, Ludwig O, Nunes U (2009) LIDAR and vision-based pedestrian detection system. J Field Robot 26(4):696–711

5. Premebida C, Monteiro G, Nunes U, Peixoto P (2007) A lidar and vision-based approach for pedestrian and vehicle detection and tracking. In: IEEE intelligent transportation systems conference ITSC, pp 1044–1049
6. Spinello L, Siegwart R (2008) “Human detection using multimodal and multidimensional features. In: Proceedings of 2008 IEEE international conference on robotics and automation. pp 3264–3269
7. García F, Jiménez F, Naranjo JE, Zato JG, Aparicio F, Armingol JM, de la Escalera A (2012) Environment perception based on LIDAR sensors for real road applications. *Robotica* 30(2):185–193
8. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition (CVPR 2001) 1(C):511–518
9. Dalal N, Triggs W (2004) Histograms of oriented gradients for human detection. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition CVPR05 1(3):886–893