



Universidad
Carlos III de Madrid



This document is published in:

Expert Systems with Applications (2014). 41(15), 6646-6661.
DOI: <http://dx.doi.org/10.1016/j.eswa.2014.04.034>

© 2014 Elsevier Ltd.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Context aided pedestrian detection for danger estimation based on laser scanner and computer vision

Fernando García ^{a,*}, Jesús García ^b, Aurelio Ponz ^a, Arturo de la Escalera ^a,
José María Armingol ^a

^a Intelligent Systems Lab, University Carlos III de Madrid, Leganés, Spain

^b GIAA Research Group, University Carlos III de Madrid, Colmenarejo, Spain

* Corresponding author. Tel.: +34 6248325; fax: +34 916249430.

E-mail addresses: fegarcia@ing.uc3m.es (F. García), jgherrer@inf.uc3m.es (J. García),
apv@ing.uc3m.es (A. Ponz), escalera@ing.uc3m.es (A. de la Escalera), armingol@ing.uc3m.es (J.M. Armingol).

Abstract: Road safety applications demand the most reliable sensor systems. In recent years, the advances in information technologies have led to more complex road safety applications able to cope with a high variety of situations. These applications have strong sensing requirements that a single sensor, with the available technology, cannot attain. Recent researches in Intelligent Transport Systems (ITS) try to overcome the limitations of the sensors by combining them. But not only sensor information is crucial to give a good and robust representation of the road environment; context information has a key role for reliable safety applications to provide reliable detection and complete situation assessment. This paper presents a novel approach for pedestrian detection using sensor fusion of laser scanner and computer vision. The application also takes advantage of context information, providing danger estimation for the pedestrians detected. Closing the loop, the danger estimation is later used, together with context information, as feed-back to enhance the pedestrian detection process.

Keywords: Data fusion, Context, Laser scanner, Computer vision, Pedestrian detection

1. Introduction

Most of the accidents in roads are connected with human errors. Wrong decision making or driver inattentions are the two main errors that cause traffic accidents. These kind of errors are related to human nature and cannot be eliminated, although efforts can be made to decrease them. Recent researches in Intelligent Vehicles have focused on using the advances in information technologies to prevent these errors. One example of these kind of applications are the Advanced Driver Assistance Systems (ADAS), which try to warn driver in case of hazardous situations.

Sensor trustability is one of the main issues when dealing with road safety applications. The ADAS systems need the most reliable set of sensors to fulfill the requirements of these demanding applications. Thus, to accomplish such a difficult task, it is mandatory to combine different information sources so we can overcome the limitations of each independent sensor. Here is where context information can be helpful for both increasing the accuracy of each sensor independently and providing a new information source to improve the performance of the Fusion process.

In this paper, a novel approach, based on data fusion for pedestrian detection is presented. It makes use of state of the art pedestrian detection algorithms for both laser scanner and computer

vision and Joint Probabilistic Data Association (JPDA) for data association, which was specially adapted to be used in a real time automotive environment. The application also takes advantage of some available contextual information, (including static knowledge as well as some online information). Thus, by combining a strong association technique and context information, classical ADAS detection is enhanced. Three main sensors were used for this application:

Laser scanner. Recent researches have focused on the use of this well-known sensor in automotive applications. Its robustness and reliability has been proved in different test and contests (e.g. DARPA Grand and Urban Challenge) (Defense & Agency, 2006; Iagnemma & Buehler, 2006; Buehler et al., 2005; Iagnemma and Buehler, 2007).

Computer vision. It is a common topic in Intelligent Vehicles research, and nowadays it can be found in commercial systems.

Inertial sensor. It is an improved GPS with inertial correction that allows accurate estimation, not only of position and velocity, but also of Euler angles, acceleration, etc. This information about vehicle state is added as dynamic contextual knowledge to enhance the fusion process.

2. Sensor fusion in vehicles. State of the art

Fusion approaches in vehicle safety can be divided in relation to the processing architecture employed. In Intelligent Vehicle applications, this is usually the way of dividing them. Most of these approaches take advantage of complementary properties of the available sensors mainly at fusion level 1 (Object Assessment), focusing on the detection and tracking of the different actors involved in road environments:

In **feature vector fusion** approaches, some preprocessing is performed for each sensor to create a set of features for each one. These individual sets are combined to create a compound set that is used to perform the obstacle detection and classification. In Premebida, Ludwig, Silva, and Nunes (2010) and Premebida, Ludwig, and Nunes (2009) features are extracted for each sensor independently and a new data set is created; authors present different approaches whether combining or not the different features of the different sensors and comparing results. The final classification after fusion is compared with alternative methods such as Naïve Bayes, Gaussian Mixture Models, Neural Networks. In Zhao, Chen, Zhuang, and Xu (2014) multiple feature fusion is performed, based on different feature extraction approach for vision approaches i.e. Support Vector Machine (SVM), Naïve Bayesian and Minimum Distance Classifier.

Decentralized fusion architecture based approaches perform detections and classifications for each sensor independently and a final stage combines the detections according to the certainty of the detections and sensors trustability. Spinello and Siegwart (2008) uses Adaboost vision based pedestrian detection and Gaussian Mixture Model classifier (GMM) for laser scanner based pedestrian detection, finally a Bayesian decisor is used to combine detections of both subsystems. In Premebida et al. (2009) pedestrians are detected by a laser scanner using multidimensional features that describe the geometrical properties of the detections, and features of Histograms of Oriented Gradients (HOG) with SVM classification for computer vision based pedestrian detection; Bayesian modeling approach is used for the final fusion. Premebida and Nunes (2013) takes advantage of lidar ROI detection, computer vision and contextual information from digital map to provide high level fusion based on a Bayesian approach.

Generally, data fusion approaches among Intelligent Vehicles researches use data from laser scanner to detect regions of interest (ROI), and computer vision to classify among different obstacles that can be found. Such as Ludwig, Premebida, Nunes, and Ara (2011) where HOG features combined with SVM approach provide pedestrian detection and Pérez Grassi, Frolov, and Puente León (2010) based on Invariant Features, again with SVM, to perform the vision based pedestrian detection. These approaches take advantage of the trustability of the laser scanner for obstacle detection but fusion is limited to speed up the process by detecting robust ROIs. Thus, the information added by the fusion process is limited and it could barely be considered real data fusion.

Some other fusion approaches in the scope of the Intelligent Transport Systems researching field take advantage of different sensors' properties in systematic approaches, although no explicit data fusion processes or algorithms are included in the works: Broggi, Cerri, Ghidoni, Grisleri, and Jung (2008) uses information from laser scanner to search those zones of the environment where pedestrians could be located and visibility is reduced (e.g. space between two vehicles) and performs detections using a vision approach. In Bohmlander, Doric, Appel, and Brandmeier (2013) mono camera and a capacitive sensor are used for pedestrian detection, and in García, Cerri, Broggi, de la Escalera, and Armingol (2012) radar and computer vision by means of featured based optical flow, are used for vehicle over-taking detection.

Fusion is also widely used for vehicle positioning, based on the fusion of different positioning and tracking techniques, such as GPS, inertial measurements and odometry. By fusing GPS signal with inertial measurements, problems regarding to GPS signal loss can be overcome, improving the precision of the positioning systems, as presented on Bhatt, Aggarwal, Devabhaktuni, and Bhattacharya (2014) and Martí et al. (2012).

Within the scope of data fusion, target tracking is one of the main aspects, several works have been presented that enhance the detections by the use of advance tracking procedures in Intelligent Transport Systems (ITS) and expert systems researching fields. By combining strong classification algorithms and trustable tracking procedures, reliable pedestrian detection can be performed: In Li, Xu, Goodman, Xu, and Wu (2013) background-foreground identifications enhances the detection, the later combination with Camshift tracker and a Kalman Filter (KF) allows trustable pedestrian detection and tracking. In Fan, Mittal, Prasad, Saurabh, and Shin (2013) deformable part models and KF are used for visual based pedestrian detection and tracking with JPDA association technique. In Schneider and Gavrila (2013) comparative between Extended KF and Interacting Multiple Models (IMM) tracking methods is provided for stereovision based pedestrian detection. Works presented on Sánchez, Patricio, García, and Molina (2009) and Gómez-Romero, Patricio, García, and Molina (2011) provide tracking procedures for surveillance applications, taking advantage of the context information in complex scenarios.

Context information can aid safety applications by both adding inference development (i.e. checking the consistence of the detections with the previously defined model) and adding explanatory aspects when the inference is consistent with the context. Some works commented before already take advantage of this useful information for pedestrian detection based on digital maps (Premebida & Nunes, 2013), video surveillance applications (Sánchez et al., 2009; Gómez-Romero et al., 2011) or for vehicle positioning (Martí et al., 2012). This work uses contextual information to complete the available sensor data, representing accurately the current situation of the vehicle and the objects surrounding it, affecting to the predictable behavior of the driver according to the safety regulations.

Finally it is important to remark that traffic and road safety applications are common topic within expert systems, the availability of modern IT technologies allows the development of modern and advance algorithms that make use of the these advances to prevent road accidents or mitigate its consequences: In Castro, Delgado, Medina, and Ruiz-Lozano (2011) an expert system based on fuzzy logic is presented, designed specifically to avoid pedestrian accidents. Guo, Ge, Zhang, Li, and Zhao (2012) presents an Adaboost and SVM based system for pedestrian detection. In Conesa, Cavas-Martínez, and Fernández-Pacheco (2013) vehicles driving in opposite direction are detected by means of agent based architecture. In Jo, Lee, Park, Kim, and Kim (2014) driver drowsiness is analyzed by means of the used of driver specific biological measurements and computer vision based algorithms for eye state and blinking detection. Finally authors in Abellán, López, and De Oña (2013) provide an algorithm to analyze and identify the severity of the accidents using decision trees.

In this work, a fusion-based expert system for pedestrian detection and danger avoidance is presented. Previous works, commented above, are particular solutions developed for specific sensor inputs not taking advantage of all the contextual knowledge. Furthermore, present work tries to focus on a system-level approach, making use of information and processes performed at different levels, and taking into account the final application of the fusion process. All these aspects are not common among ITS researches. Finally, the model makes use of a JPDA approach for multiple sensors, able to overcome difficult situations in the data

association technique. Although JPDA (Bar-Shalom & Li, 1995; Blackman & Popoli, 1999) algorithm is not a recent algorithm in the fusion field, successfully used on previously presented works, such as Fan et al. (2013), it represents a step forward in data fusion on ITS, overcoming some of the problems of the classical association techniques (Musleh, García, Otamendi, Armingol, & De La Escalera, 2010; Garcia, de la Escalera, Armingol, Herrero, & Llinas, 2011).

3. General description

System architecture is based on a decentralized scheme, where each sensor (scanner laser and computer vision) performs independent detections. Later fusion method performs plot-to-track fusion combining the information of both subsystems to give a more robust and reliable detections. Inertial system is used in level 0 to compensate the movement of the vehicle when using laser scanner and in level 2 and 3 is combined with context information to provide danger estimation of the different pedestrian detected. In Fig. 1, system architecture is presented. The nature and the requirements of the application demand a multilevel approach that can give a complete solution to the safety application. Data fusion levels definition can be found in Hall and Llinas (2001):

In **level 0** inertial measurements are used to compensate the movement of the vehicle during the laser scan and also to avoid unreliable reading (for instance while significant movement in the pitch angle due to obstacles). Besides, data alignment with video input is performed using pin-hole model to transform laser scanner coordinates to vision coordinates. This allows avoiding false positives in the camera approach and increases the efficiency of the vision system by reducing the search space in the image for obstacles detected by the laser.

Level 1 First, each sensor performs classification. A later fusion approach combines the systems' detection in a plot-to-track

approach and performs tracking algorithm. Here, some context information is used to help in the classification, e.g. pedestrian size.

Level 2 and 3 Interactions and situation assessment are made using detection tracks; context information is augmented with the information of the inertial system that provides information about the movement of the vehicle.

Vision approaches have proved to be very robust for pedestrian detection in the latest years. But computer vision has a limited field of view and it is strongly affected by weather conditions (e.g. snow, dust, solar reflections, etc.). The present approach tries to take advantage of different information sources available (computer vision, inertial system, laser scanner and context) to add robustness to the classic computer vision based approaches. Pedestrian detection is performed by the two main sensors (laser scanner and computer vision) independently; thus, as both sensors have different field of view, different detection zones can be found. A fusion zone where both sensors are available is created, and a single sensor zone where only the laser scanner can be used is also created (Fig. 2). In this way, thanks to the laser scanner pedestrian detection subsystem, the pedestrians can be found even when they are out of the camera field of view or when weather conditions or system failure makes impossible the computer vision detection, adding robustness and reliability to the system. In the fusion zone, the redundant pedestrian detection helps to provide a more robust detection preventing false positives and decreasing the amount of misdetections.

In order to give situation awareness for a complete system-level approach, danger estimation was performed accordingly to the different danger zones created. Here is where context takes a relevant role in the application, since vehicle movement information, pedestrian detection and other relevant information (e.g. vehicle mass, braking distance, response time, etc.) are integrated in a common reasoning framework to obtain pedestrian danger

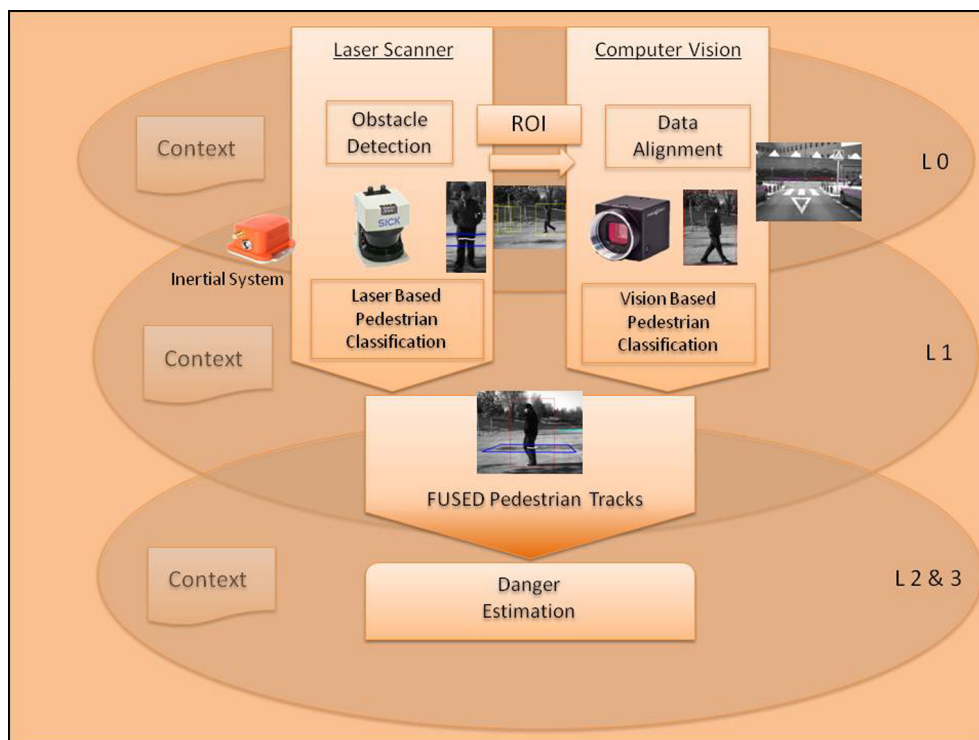


Fig. 1. Overall System Diagram with JDL levels correspondence. The process diagram shows the different tasks involved in any of the levels, with the corresponding information sources used.

estimation and relevant distances. These relevant distances are braking distance and response distance. The former is the time elapsed until vehicle completely stops, and the latter is the mean response time of the driver to an external stimulus. These relevant distances are used to create three danger regions, where pedestrian detection is situated: safe region, danger region and imminent collision region. According to the zone where a given pedestrian is located, different actions should be performed.

Tests were performed in the test platform IVVI 2.0 (Fig. 3), property of the Intelligent Systems Lab of University Carlos III of Madrid.

In Section 4, pedestrian detection subsystems and data alignment are explained. In Section 5, the tracking process is detailed together with the fusion algorithm. Section 6 depicts the fusion levels 2 and 3, giving description of the relevant distances taken into account for danger estimation. Finally, test results and conclusion are given in Sections 7 and 8 respectively.

4. Pedestrian detection and data alignment

As it was previously mentioned, Intelligent Vehicles fusion approaches are mainly focused on levels 0 and 1, but they rarely present a systematic description of fusion processes accordingly to JDL Fusion model. This works tries to give a complete application that takes into account all the levels of the Fusion process. In this section, Level 0 of data fusion is going to be detailed.

First, pedestrian detection system is described for each subsystem. Later, procedures to extrapolate laser scanner detections to computer vision coordinate system are detailed. Finally, all these detections should be extrapolated to vehicle coordinates system that is the front part of the vehicle.

4.1. Laser scanner pedestrian detection

The laser scanner was mounted in the bumper of the test platform IVVI 2.0 (Fig. 4). The model selected for this application was a single layer laser scanner from SICK, model: LMS 291-S05. This laser scanner provides angular resolution of 0.25° , a field of view of 100° and a detection distance up of 82 m.

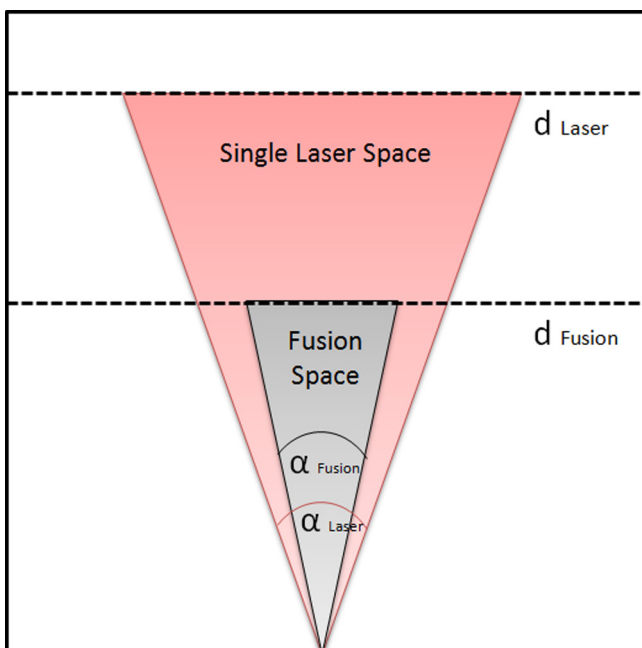


Fig. 2. Detection zones according to the different fields of view of the different subsystems.



Fig. 3. Test platform IVVI 2.0 (Intelligent Vehicle based on Visual Information 2.0).

Pedestrian detection algorithm is composed by two stages. First, obstacle segmentation and shape estimation is performed. Obstacle classification is performed in the next stage, taking into account the shape estimated in the previous stage.

4.1.1. Obstacle segmentation and shape estimation

Laser scanner provides 401 detection points per scan, each one of them with a given time delay respect to the others. Thus, after data are received from the laser scanner, the movement of the vehicle should be compensated using the information given by the inertial system MTI-G. Euler angles, displacement and velocity should be taken into account in order to avoid deformation of the reconstructed shapes.

Laser scanner is also very sensitive to pitching movements, so it is mandatory to check if there is a pitching movement that would make impossible the detections. In these situations, the laser scanner pedestrian detection is disconnected so the detection is performed based on computer vision system.

4.1.2. Movement compensation

Errors due to strong pitching movements are avoided using the inertial system. Pitch movement is checked and when a strong pitching movement is detected, laser scanner detection is disabled to avoid misdetections and errors.



Fig. 4. Platform IVVI 2.0 with the laser scanner mounted in the front bumper and the computer camera system in the windshield.

Euler angles computed by the inertial system are used to correct the displacement of the measures due to the movement of the vehicle. Eq. (1) depicts the compensation with the rotation and translation matrixes needed to correct this movement. This way, the points are referenced to the position of the last point received (Fig. 5(a)).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T + T_0 \right)$$

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

$$\cdot \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ \sin(\Delta\theta) & \cos(\Delta\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$T_0 = \begin{bmatrix} vT_i \cdot \cos(\Delta\theta) \\ vT_i \cdot \sin(\Delta\theta) \\ 0 \end{bmatrix}, \quad T_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix},$$

in Eq. (1) $\Delta\delta$, $\Delta\varphi$ 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 the vehicle movement compensation. R is the rotation matrix, T_v is the translation matrix according to the movement of the test vehicle, and T_0 is the translation matrix according to the position of the laser and the inertial sensor.

The point clouds are clustered using Euclidean distance and a threshold that is dependent on the distance to the laser scanner (Eq. (2)).

$$th = th_0 + K \cdot dist \quad (2)$$

Thus, for a given point p_i , defined by its coordinates (x_i, y_i) , belongs to a segment S_j if there is a point p_j included on this segment that satisfies:

$$p_i(x_i, y_i) \in S_j \rightarrow \{\exists [p_j(x_j, y_j) \in S_j] : d(p_j, p_i) < th\} \quad (3)$$

A new segment is created if a given point is not included in any segment. Finally, a filtering mechanism eliminates all the single point segments, considered false detections from the laser scanner.

Finally, after clustering, the shapes of the different obstacles are estimated using polylines (Fig. 5(b)). The idea of polylines consists of a recursive method that estimates the shape of a given obstacle

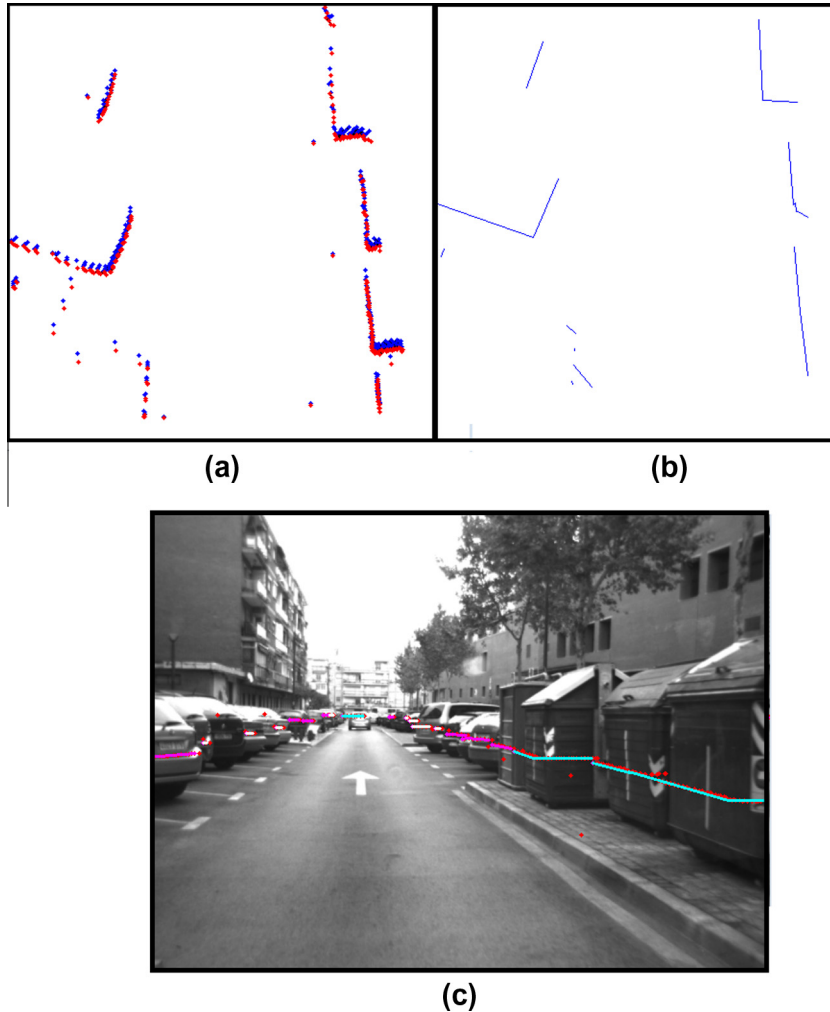


Fig. 5. Vehicle movement compensation and data alignment. (a) Shows the detection points, in blue before the vehicle movement compensation and red for compensated. (b) Shows the shape reconstructed after the movement compensation. (c) Shows the final alignment of the laser scanner data and the image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

by merging the points included in a given segment with straight lines.

In a given segment, the first and last point are merged by a straight line (Fig. 6(a)).

Next step consists of selecting the next point in the segment. If the distance from this point to the line is lower than a threshold, two new lines are created from the first point to the selected one and from it to the last one Fig. 6(b) and (c). Otherwise the point is not connected by any line.

A new point is selected to check the distance to the closest polyline and the previous step is repeated until all the points are checked (Fig. 6).

This process is presented for its use in automotive applications in Broggi et al. (2008). Its main drawback is that it is dependent on the threshold for an accurate shape reconstruction, although the threshold strategy can be used to configure the precision of the reconstruction. Thus, for an accurate detection where the original shape is important, it can be chosen a low threshold. On the other hand, when dealing with bigger distances, or low detail of the reconstruction is less important, the threshold can be increased allowing the algorithm to focus on the general shape or skeleton of the shape. For the present application, the higher threshold allows to create a general pattern for the leg movement. Other approaches (Shao et al., 2007; Xavier, Pacheco, Castro, Ruano, & Nunes, 2005) are based on closer distances where a higher amount of points allows statistical differentiation. As it is shown in Fig. 7, the polyline pattern obtained for the case of pedestrians is less variable in relation to the distance when the threshold selected is high. Thus, for the present approach, high threshold was used to avoid variations of the polylines for pedestrians due to the distance. Besides, as this approach does not pay attention to the details, it is not dependant on the shoes or even the clothes that the pedestrian is using (e.g. skirts, trousers, etc.). On the other hand, the simplicity of the reconstructions leads to misdetections. To solve this, a tracking stage was added in at low level to analyze the movement of the pedestrian along time e.g. velocity, changes in the shape, impossible movements, etc.

4.1.3. Obstacle classification

Pedestrian classification is performed in two steps, taking into account a priori context information. First, obstacles with a size proportional to a pedestrian are selected among the different

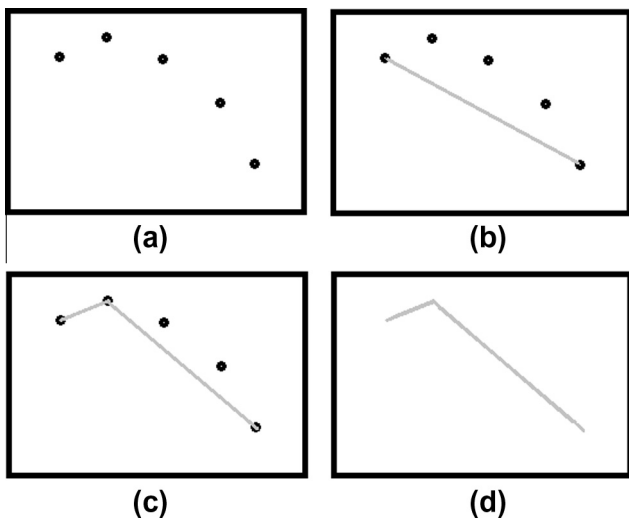


Fig. 6. Polyline creation process. (a) Segment without polyline. (b) Segment with first polyline. (c) A new point is added to the polyline creation process and (d) the final shape estimation.

obstacles and then the shape of the polyline is checked against a typical pedestrian pattern. It has to be remarked that the purpose of this approach is to detect single pedestrians. Further approaches, already under development, deals with groups of pedestrians.

4.1.3.1. Contextual knowledge for pedestrian detection. The pedestrian size that was used in the present application to select possible pedestrians is based on the model of human body (Skehill, Barry, & Mcgrath, 2005; Highway Capacity Manual, 2000) that models human body as an ellipse. In Skehill et al. (2005) an study of the physical dimension of human being is detailed. It is usually accepted that physical dimension for pedestrian was given in the early 70s in Highway Capacity Manual (2000) and are corresponding to an ellipse whose two main axes are (57.9 cm \times 33 cm); that ellipse includes the body of a dressed human being. Other researches Still (2000) uses worldwide anthropological studies to conclude that this ellipse is (45.58 cm \times 28.20 cm). Finally both Skehill et al. (2005) and Highway Capacity Manual (2000) conclude that this dimension is (0.6 \times 0.5). In this work, the latter assumption will be the model used to perform the pedestrian detection.

Finally, a study of the different patterns given by pedestrians was performed, giving the conclusions found in Fig. 4.

4.1.3.2. Pattern matching. In this pattern, three polylines are presented, and the angles that connects the polylines are included within the limits of $[0, \frac{\pi}{2}]$.

A pattern matching process computes the two angles and gives a similarity score where 1 means 100% match (see Fig. 8)

$$\text{Similarity} = \frac{2\theta_1}{\pi} \cdot \frac{2\theta_2}{\pi}, \quad (4)$$

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

This similarity is computed among any two consecutive polylines that represent the shape of the pedestrian. If the result is bigger than a given threshold the obstacle is considered to be a pedestrian.

It is assumed that the previous pattern is very common when dealing with laser scanner, thus false positives are expected. To overcome this problem, a low level tracking stage was created. This stage allows tracking the movement of pedestrians along time to remove false detections, and checking whether the pedestrian is performing unexpected movement or if the size of the pedestrians changes and it does not fulfill the human being constraints. It is based on KF with constant velocity model. The new detections are searched within a window whose size depends on the size of the pedestrians.

Finally, the correlation with previous detections from the laser scanner tries to check if we are dealing with the same segment, or if the segment changed its size. This correlation is also very helpful to overcome the problem of two detections within the same window.

The algorithm for obstacle correlation is based on the size of the object found (Eq. (5)). It is remarkable that most of the features used for this correlation were based on the internal points and the distances in x coordinate. Y coordinates are subject to a higher variability due to frequent occlusions:

$$\text{Corr} = \gamma_1 \bar{N} + \gamma_2 \text{width} + \gamma_3 \delta_{+x} + \gamma_3 \delta_{-x} + \gamma_4 \sigma + \gamma_5 d + \gamma_6 \rho, \quad (5)$$

where γ_i represents the weight for a given parameter, N is the medium number of points, width is the size of the obstacle, σ is the standard deviation of the points to the center of the obstacle ρ . Is the radius of the circle surrounding the obstacle and d is the distance to the estimation of the KF. Finally δ_{+x} , δ_{-x} are the number of points to the left or to the right of the center.

The correlation value used in Eq. (5) is also used to eliminate tracked pedestrians with high variability.

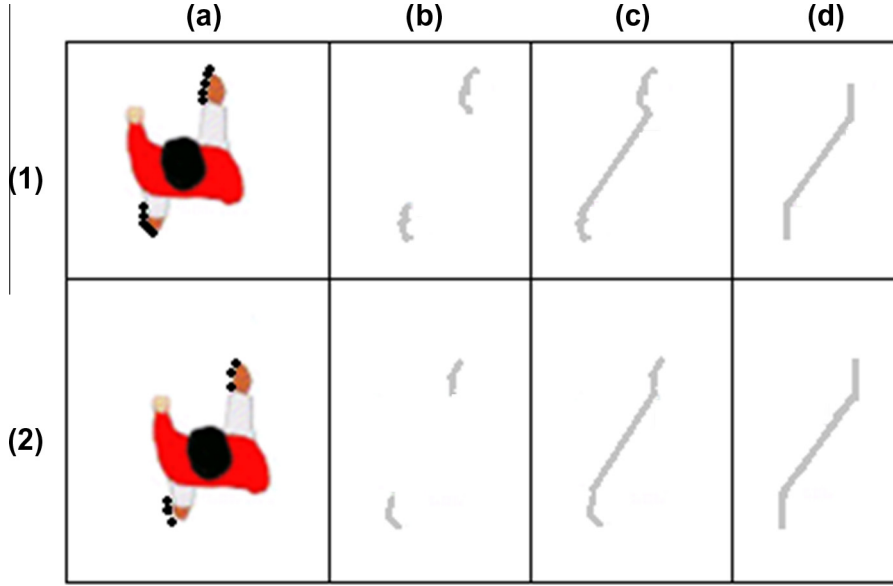


Fig. 7. Representation of the polyline creation according to the thresholds. (a) Representation of the detection points according to the pedestrian legs, black dots represent laser scanner detections. (b) Polyline reconstruction with low thresholds for both segment creation and polyline creation. (c) Polyline reconstruction with low thresholds for polyline creation and high threshold for segment creation. (d) Polyline reconstruction with high thresholds for both segment creation and polyline creation. (1) Example at closest distance (more detection points). (2) Example for long distance (less detection points).



Fig. 8. (a) Pattern for pedestrian detection. (b) Different examples of different patterns given by pedestrians with different leg positions.

The final classification is provided taking into account the last ten detections by a voting scheme. The voting scheme tries to overcome the limitations in the information provided by a single scan by taking into account the last 10 detections:

$$V_i = \delta_i N_i, \quad (6)$$

where V_i is the weight of each kind of obstacle and δ_i is the weight considered for the given kind of obstacle, and finally N_i is the number of votes for each type of obstacle. The kind of obstacle with higher V_i is the final selection for each obstacle.

4.2. Vision based pedestrian detection

Vision based pedestrian detection is based on the HOG descriptor and SVM classification (Dalal & Triggs, 2005). This classification requires a high computational cost, thus ROI detection is performed before. This approach uses the laser scanner detections and the field of view of the camera to reduce the amount of data to process in the image; thus, only obstacles given by the laser scanner and extrapolated to the image are processed to check whether they are pedestrians or not.

Before ROI detection, some data alignment should be performed since sensors do not share the same coordinate system.

It is interesting to notice that this system can work in two configurations: based only in computer vision, or together with the

information provided by the laser scanner. The latter allows the system to reduce false positives, since only obstacles detected by the laser scanner are used for computer vision detection. In this way, errors or multiple detections are reduced since only regions with obstacles in are taken into account. In extreme situations where the laser scanner is not available, i.e. strong pitching movements, the camera is still able to work using the pin-hole model to estimate the distance. Furthermore, in normal conditions, the laser scanner represents a very accurate sensor to estimate the distance to the obstacle. It is important to remark that here, the region of interest provided by the laser scanner corresponds to those obstacles with sizes similar to pedestrians, and the pedestrians classification presented is an independent process. Fusion of the classifications of each sensor independently will be explained in further sections.

Space alignment was performed taking into account the pin-hole model and using the rotation matrix. Both coordinate frames of the two different subsystems were translated to the reference point of the vehicle which is the central point of the front bumper (Fig. 9).

To perform this coordinate change, transformation and rotation matrix should be used,

$$\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix} = R_i \left(\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + T_i \right), \quad (7)$$

where T_i is the translation matrix and R_i the rotation matrix that corresponds whether to the laser or the camera (R_l or R_v). These rotation matrixes are equivalent to the rotation matrix used in Eq. (1) but in this cases Euler angles corresponds to the angular deviation among coordinates system of the camera and of the laser scanner with respect to the vehicle coordinate system. T_i is equivalent to the displacement in cartesian coordinate system for each sensor (T_l for the laser and T_c) for the camera:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ z_c \\ y_c \end{bmatrix} \quad (8)$$

where u, v are the image coordinates in pixels, f the focal distance and (x_c, y_c, z_c) the cartesian coordinates of the image detections in the image coordinate system. u_0 and v_0 are the coordinates of the center of the image.

Eqs. (7) and (8) were also used to transform ROIs detections from laser scanner space to camera space to perform vision based classification. This way, data association from both sensors is implicit since they perform classification for each sensor for the given obstacles detected by the laser scanner (Fig. 10).

Finally, once the bounding boxes provide the region of interest, the HOG features algorithm perform the pedestrian detection based on the HOG features algorithm presented in Dalal and Triggs (2005).

The theory after the HOG feature description is based on local appearance and shape of all objects in an image, which can be described by the distribution of intensity gradients or edge directions. The implementation divides the image into small-connected regions (cells) that can have different shapes (circles or squares). For each cell a histogram of gradient directions (or edge orientations) for the pixels within the cell is compiled (Fig. 11(c)). These cells are later divided into blocks. These blocks represent the descriptor of the image, with different regions. Regions can overlap, thus some cells can belong to more than one block. The final descriptor is the normalized histogram of the cells belonging to all the blocks that represent the image.

These descriptors are trained according to a database of images, to select those regions of the image having more weights to differentiate a pedestrian, and this process is performed by a Support Vector Machine (SVM) classifier. SVM is a binary classifier that looks for the optimal hyperplane for a decision function. It is a machine-learning algorithm widespread in computer vision approaches. Fig. 11(c) depicts the HOG features for a given bounding box.



Fig. 9. Different coordinate system (x_c, y_c, z_c) are camera coordinate system, (x_l, y_l, z_l) is the laser coordinate system, and (x_v, y_v, z_v) is the coordinate system. Coordinates (u, v) are the image coordinates.

5. Fusion system

The sensor fusion algorithms (fusion at JDL level 1) are detailed here, including an extensive explanation of the estimation filter used to track the movement of the pedestrians found. Plot to track configuration is detailed paying attention to track creation and deletion policy as well as to data association and gating procedures.

5.1. Estimation filter

Thanks to the high update rate of the laser scanner, KF was considered to be a reliable choice for pedestrian tracking in the road environment. We used a model similar to Kohler (1997) to track pedestrians assuming constant velocity and modeling accelerations as system process noise.

5.2. JPDA for data association

Data association is based on JPDA, adapted to the problem of pedestrian detection with laser and camera.

First, a squared-gates approach is used to perform gating (9):

$$K_{Gl} \sigma_r \quad (9)$$

where σ_r is the residual standard deviation and K_{Gl} is a empirically chosen constant. Next step is the association process, it is performed based on JPDA approach, adapted to this specific application.

JPDA is a well-known association technique (Blackman, 1986; Blackman & Popoli, 1999). Its probabilistic approach helps to obtain better results when dealing with specifically difficult situations than the classic approach based on Global Nearest Neighbors.

The joint association probability, defined in (10), represents the probability of a joint association event $\theta_{k_j}^m$ that associated measurement m to track j at a time k :

$$P(\theta|Z_k) = P_D^{M-n} (1 - P_D)^n P_{FA}^{m_k - (1-M)} \prod_{j=1}^{m_k} g_{ij}, \quad (10)$$

where g_{ij} is defined assuming a 2 dimensional Gaussian association likelihood, for all the measurements to the target:

$$g_{ij} = \frac{1}{(2\pi)^{N/2} \sqrt{|S_{ij}|}} e^{-\frac{d_{ij}^2}{2}}, \quad (11)$$

where d_{ij} is the Euclidean distance between the prediction and the observation. S_{ij} is the residual covariance matrix. Assuming independence of errors in Cartesian coordinates, we would have $\sqrt{|S_{ij}|} = \sigma_x \sigma_y$ and $N = 2$.

Finally, all the association hypotheses are weighted in the updating stage of the KF. The innovation is calculated using all possible combinations, weighted accordingly to the association likelihood values:

$$I_k = \sum_{i=1}^m \left[P(\theta|Z_k) \left(Z_{i,k} - H_k \hat{X}_{k|k-1} \right) \right], \quad (12)$$

where I_k is the innovation for the KF of a given track.

In certain situations, the previous algorithm can lead to unstable behaviors. When several tracks compete for a single observation, the clutter (no assignment) is the most powerful option due to the weight of the joined probabilities of the other options, different from the joint to be calculated. To overcome this problem, any associated track-new detection pair is eliminated from the assignment process after it is assigned. Thus, in further assignments all the joined probabilities are recalculated with the remaining tracks and detections. This way, the problem is avoided by eliminating the weight of the already assigned solutions in subsequent

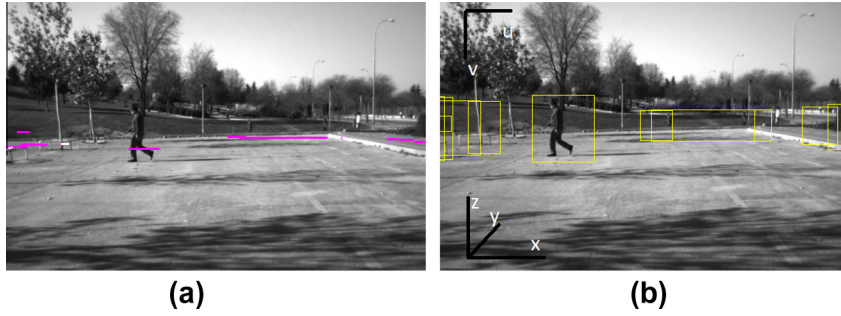


Fig. 10. Laser scanner ROI detection with the different coordinate system of the different sensors highlighted in the image (u, v) for image space and (x, y, z) for laser scanner. (a) Shows the laser scanner detections in the image coordinate system. (b) Shows the boxes that represent the ROIs.

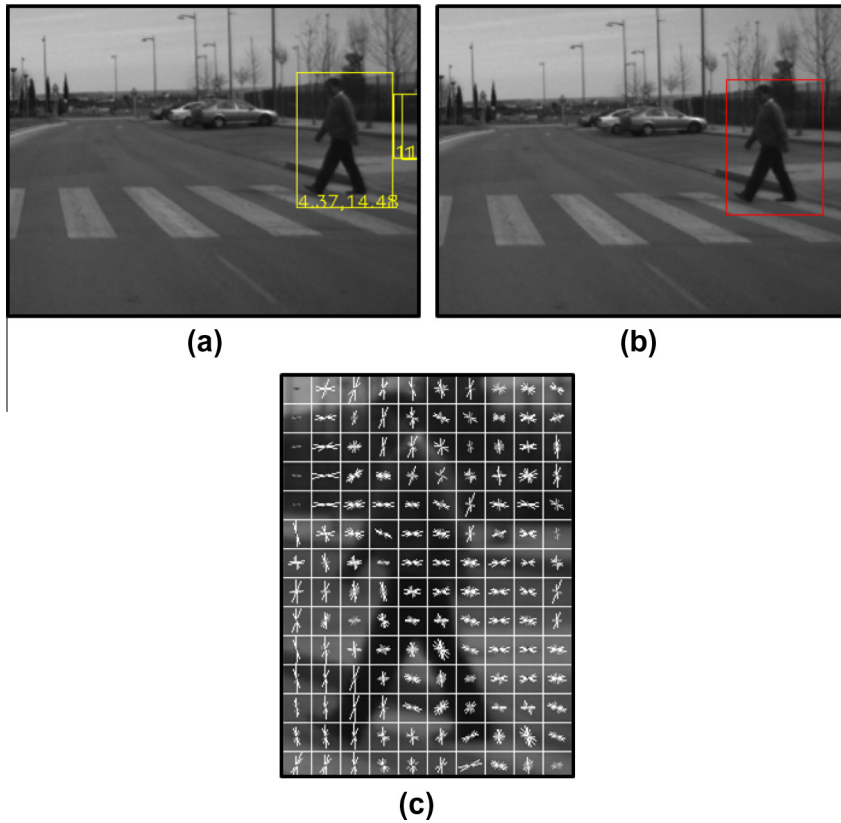


Fig. 11. Visual pedestrian detection. (a) Region of interest calculation with the distance to the obstacle in meters indicated in cartesian coordinates. (b) The detected pedestrian highlighted. (c) The HOG features description of the bounding box, with the histogram of gradients represented by lines included in the cells, whose length and intensity are equivalent to the strength of the given gradients.

assignments. In the case of several tracks pointing to a single observation, this solution would first assign clutter to the less probable detection and eliminate the weight of this detection in subsequent assignments, until one of them is selected as more likely than the clutter. Different tests proved both the stability of the system, and that the computational cost added due to the necessity to recalculate the joining probabilities is negligible.

The solution based on a JPDA approach, adapted to this specific road safety application, helped to overcome difficult situations originated due to the specific nature of the sensors used, some of them are:

5.2.1. Clustering errors

These errors are produced due to the difficulty to separate close pedestrians by the laser scanner algorithm (Fig. 12(a)). JPDA deals with this problem better than other approaches. In these situations the two pedestrians tracked use the single detection, when they

are not differentiated, to update the KF. Thus the estimation process suffers no degradation.

5.2.2. Double detections

This error can be produced by two factors: The distance from the laser scanner to the camera can produce that several detections from the laser scanner fall in the same location within the camera plane, or due to clustering errors of the laser scanner that creates several ROIs including the same pedestrian, e.g. dust or rain. In these situations, no new track would be created, since in most of the situations both detections falls into the gate of the pedestrian track. Furthermore, the effect of the misdetection in the updating process is negligible, because its probability would be lower, so its weigh in Eq. (12) will be minor.

5.2.3. Occlusions or crossings

On this typical problem, the single detection given when one pedestrian is occluded by another one (Fig. 12(b)) can be used to

update the estimation filter of both pedestrians. Thus the tracking would not suffer any variation by this misdetection.

5.3. Track management

Track creation and deletion policy follows a logic that was set empirically. According to the detection zone where a pedestrian is found, different policies are applied. Also zone switching should be taken into account. The following table depicts the logic followed to track creation and deletion.

Table 1 depicts the track management logic according to the region where the laser scanner is located. The values of N_{DS} , N_{CS} , N_{CF} and N_{DF} vary according to the danger region where the pedestrian is located according to the movement of the vehicle. Thus, when the vehicle is closer, the risk involving any of the consolidated tracks is higher. Thus, the higher the danger, the higher the values of these parameters.

Another aspect that has to be managed is when a track changes the sensor zones: If a track switches from laser scanner to fusion zone, any track is considered not consolidated, thus vision should corroborate the detections from laser scanner in the single sensor zone. On the other hand, when the tracks change from fusion zone to single laser scanner zone, no changes are performed; consolidated or not, the track maintains its status, and only the updating policy changes.

The use of consolidated and non consolidated tracks adds robustness to the system. Only consolidated tracks are considered trustable detections, thus they are the only ones reported. This way, false positives, mainly from laser scanner, are discarded because detections that are not corroborated by the vision system are not reported. Furthermore, the use of both sensors allows that, once a track has been consolidated and a pedestrian has been detected, he can be tracked even if he becomes undetectable by one of the subsystems (e.g. he goes out of the camera field of view). These implications of the data fusion algorithm are proved in Section 7 where the different tests performed are detailed. The results of the fusion algorithm obtained enhance the basic performances of the different sensors independently.

6. Danger estimation

In Section 2, we presented the state of the art in pedestrian detection systems for road safety applications. They are mainly focused on low-level fusion, while other processes in higher fusion levels are scarcely detailed. In this section, situation and threat assessment are intended, using context information to augment the capacity of the system, creating a reliable and robust application. Context information used is related to vehicle safety information that allows not only to detect the pedestrians in the

surroundings, but also to estimate the danger of any of the given detections.

Before giving an estimation of the danger that involves any detection, two distances should be taken into account: braking distance and response distance. These distances represent context information that should be taken into account to calculate the danger that any detection involves. The former represents the space elapsed before the vehicle is completely stopped, thus collision with any pedestrian farther than this distance can be totally avoided by stopping the vehicle. The latter represents the distance covered in the time elapsed until the driver responds to a given stimulus.

These distances are not the only context information used in this danger estimation. Other information relative to vehicle safety is necessary to calculate this danger. Response time for drivers and some traffic accident reconstruction mathematics are two examples of the context information necessary for danger estimation that will be explained in this section.

6.1. Response distance

Researches generally accept response time up to 0.66 s, as it was showed by Johansson and Rumar (1971). In this article, authors calculated this mean response time for human beings when driving by means of an auditory stimulus. Other authors in recent works have made similar tests with similar results (Makishita & Matsunaga, 2008). Thus, response distance is the distance that a vehicle with a given velocity would cover during the response time.

6.2. Braking distance

It is the distance, in meters, that the vehicle would cover until it completely stops. Many different variables would affect this calculation. This approach uses basic traffic accident reconstruction mathematics (Collins, 1979) based on the worst case scenario, i.e. when the vehicle is fully loaded. Weather may also vary the conditions (e.g. road friction coefficient). Here some on-line contextual information provided by the inertial device could be useful, such as the temperature measurement.

In traffic accident reconstruction, worst case scenario means that only front wheels are blocked when braking, this fact displaces the weight of the car to the front of the vehicle. This weight displacement is represented as a change in the friction coefficient; this change is depicted in Eq. (13):

$$\eta = \frac{b_2}{L - h\mu}, \quad (13)$$

where η is the corrected and μ the real friction coefficient, b_2 is the distance to the rear axis from the mass center, L is the length of the vehicle and h is the height of the mass center. Mass center has to be

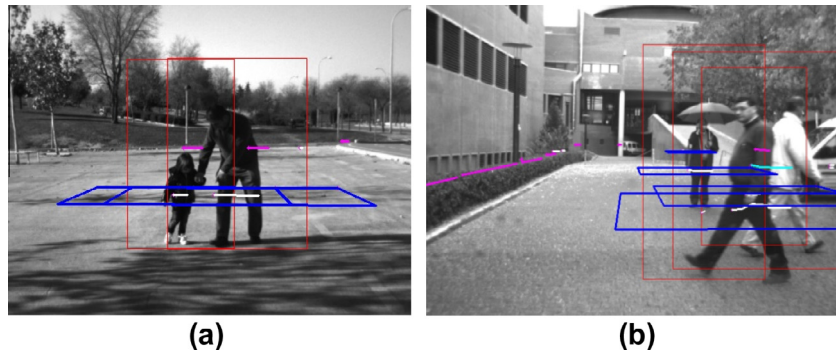


Fig. 12. Clustering errors (a) and occlusions (b).

Table 1
Track management logic.

	Single sensor zone	Fusion zone
Track creation	Every new track is considered consolidated after N_{CS} consecutive positive detections.	Every new track is considered consolidated when both subsystems give positive detection.
Track deletion	Not consolidated. If there are not N_{CS} consecutive scans the track is deleted. Consolidated. After N_{DS} no updates.	Not consolidated. If there is not match after N_{CF} consecutive scans the track is deleted. Consolidated. After N_{DF} no updates.
Track maintenance	Laser scanner track matching.	Track matching with any of the sensors.

calculated, but several authors (Collins, 1979) give an approximation of 0.4 the height of the vehicle that was used for this application. Using this approximation, the distance of the vehicle to completely stop is shown in Eq. (14):

$$d_{stopping} = \frac{v^2}{\eta \mu g}, \quad (14)$$

where v represents the velocity of the test vehicle and g the gravity acceleration.

But Eq. (14) is not the braking distance, since the response time presented before should be taken into account because it is the time before the driver starts pressing the braking pedal (15).

$$d_{braking} = vt_{response} + d_{stopping}. \quad (15)$$

6.3. Danger regions

Danger regions are created according to the previous relevant distances. These regions help the system to evaluate the danger degree associated to a given detection. Each one of the regions is created according to the different actions to perform in case that a pedestrian is found in the region. Table 2 depicts the relation between the zones and the distances.

Safe region detections are those pedestrians that are in a distance far enough to warn the driver and completely stop the vehicle before hitting the pedestrian. Danger region represents the region where it is possible to warn the driver before hitting the pedestrian, but the vehicle is not going to be able to stop on time before hitting him. Finally, imminent collision region is the region in the environment where it is not possible to warn the driver with time in advance that allows him to react, nor stop the car before colliding with the pedestrian. Furthermore, the previous section explained that these regions are very important in the tracking process since they are used to determine the thresholds for track creation and deletion process. The risk associated to a detection in the imminent collision zone is higher than the risk involving the other two regions. This way, context plays an important role in the tracking stage, enhancing the detection and maintenance of the relevant tracks. This solution can be defined as part of level 4 solution, since the detection and tracking of the pedestrians are enhanced thanks to the information obtained in the lower levels.

Finally, it is important to remark that the scope of this application is to detect and warn drivers, giving estimation of the danger involved in the detections. Thus, it is out of the scope the actions to perform in any of the detected cases. Further works should deal with this issue but a first approximation should involve:

Safe region: These detection involve no imminent danger, thus some visual or acoustic warning may be enough, paying attention to not to saturate the driver with irrelevant information.

Danger region: Here it is important to warn both, driver and pedestrian to try to avoid the possible collision. Recent works try to determine the safest trajectory to perform automatic avoiding maneuvers, allowing the vehicle to take control over the driver and prevent to harm pedestrians.

Table 2
Relation between danger regions and relevant distances.

	From	To
Safe region	Infinite	Braking distance
Danger region	Braking distance	Response distance
Imminent Collision region	Response distance	0 m

Imminent collision region: Here the only action to take is to trigger any automatic measure available in order to mitigate the harm to be produced to the pedestrian.

6.4. Danger estimation

Taking previous regions as reference and with the purpose of giving danger estimation to upper layer applications that have to deal with the previously presented situation, a danger estimation function was created. The idea is to give an estimation of the danger involving any detection. The estimation should grow exponentially the closer the pedestrian is to the vehicle. Furthermore, two values should be taken as reference: In the border between safe region and danger region (braking distance) this estimation should be bigger than 0.5 (it was chosen 0.6); second a priori assumption was that at 0 distances this function should be 1. Taking all these considerations into account following function was created:

$$f(r) = \begin{cases} e^{-\lambda(r-d_r)}, & \text{for } 80 \geq r \geq d_r, \\ 1, & \text{for } d_r > r \geq 0, \end{cases} \quad (16)$$

where r is the distance of the pedestrian to the vehicle, d_r is the distance that the vehicle covers during the reaction time, and λ was a value to calculate that would assure the previously presented assumptions (17):

$$e^{-\lambda(d_b-d_r)} = 0.6, \text{ thus} \\ \lambda = \frac{-\ln 0.6}{(d_b - d_r)}, \quad (17)$$

where d_b is the braking distance.

Table 3
Results for the computer vision-based pedestrian detection, with the laser based obstacle segmentation.

	% Of positive detection	% Of false positives (per frame)
Test	78.01	5.19
Interurban	73.19	3.91
Urban	70.72	6.72
Total	73.97	5.27

Table 4
Laser Scanner pedestrian detection performance.

	% Of positive detection	% Of false positives (per frame)
Test	79.71	6.23
Interurban	70.35	16.96
Urban	73.61	16.72
Total	74.56	13.3

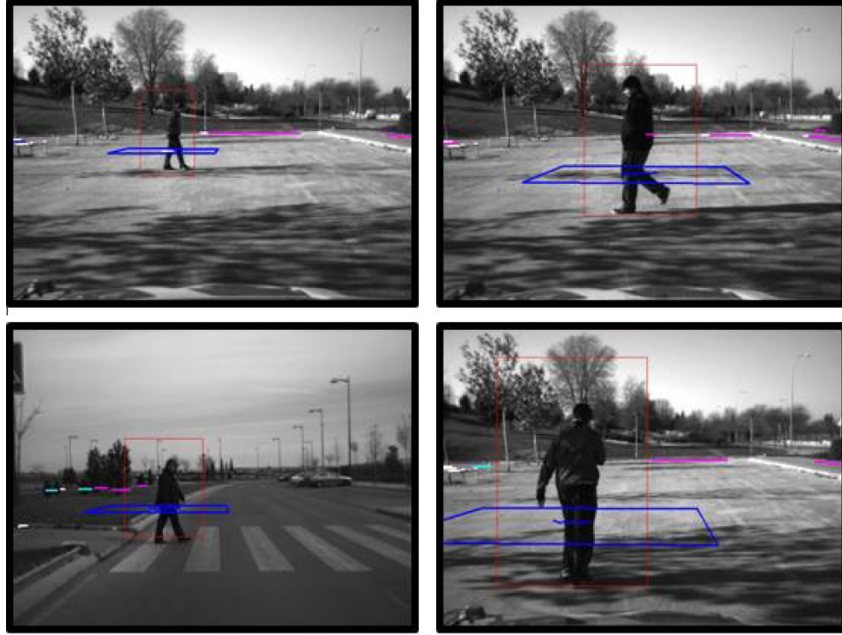


Fig. 13. Positive detections examples. Blue boxes represent laser scanner positive detections. Red boxes the image positives. Also laser scanner polyline reconstructions are showed in the image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

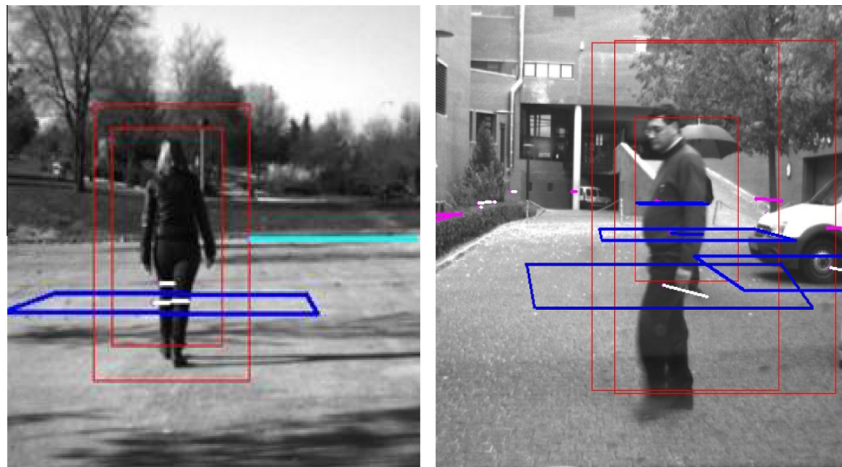


Fig. 14. Example of misdetections of the camera due to multiple ROIs pointing to a single pedestrian. Two laser scanner points to the pedestrian, thus both of them return positive detections. In these situations it could be considered a positive detection by the basic HOG feature approach, although they were considered misdetections of the vision system in the present approach that uses laser scanner ROI detection.

7. Results

Experiments were performed in more than 50 sequences in different scenarios, including more than 10,000 frames with pedestrians involved. They were divided into three kinds of scenarios: controlled environment, urban scenarios and interurban scenarios. The former represents the less challenging scenarios, typically parking lots, with the vehicle stopped, and single pedestrians performing lateral and vertical movements. Interurban scenarios are more challenging scenarios with several pedestrians in real traffic situations. Finally, urban scenarios represent the most challenging situations with a higher number of pedestrians and other obstacles that makes detection and classification tasks extremely difficult.

Tests focused on two main points. The first one was checking the viability of the low level approaches; the different low level

algorithms were checked in the different scenarios separately; later in subsequent tests these low level algorithms were compared with the fusion system to check the improvement introduced by the fusion process. The second set of tests was performed to check the behavior of the fusion algorithm. The per-

Table 5

Overall results, for all the sequences, including low level (laser scanner and camera algorithms) and the fusion procedure (JPDA) compared with other approaches (GNN).

	% Of positive detections	% False positives (per frame)
Camera	73.97	5.27
Laser scanner	74.56	13.3
Fusion (GNN)	79.62	2.21
Fusion (JPDA)	82.29	1.11

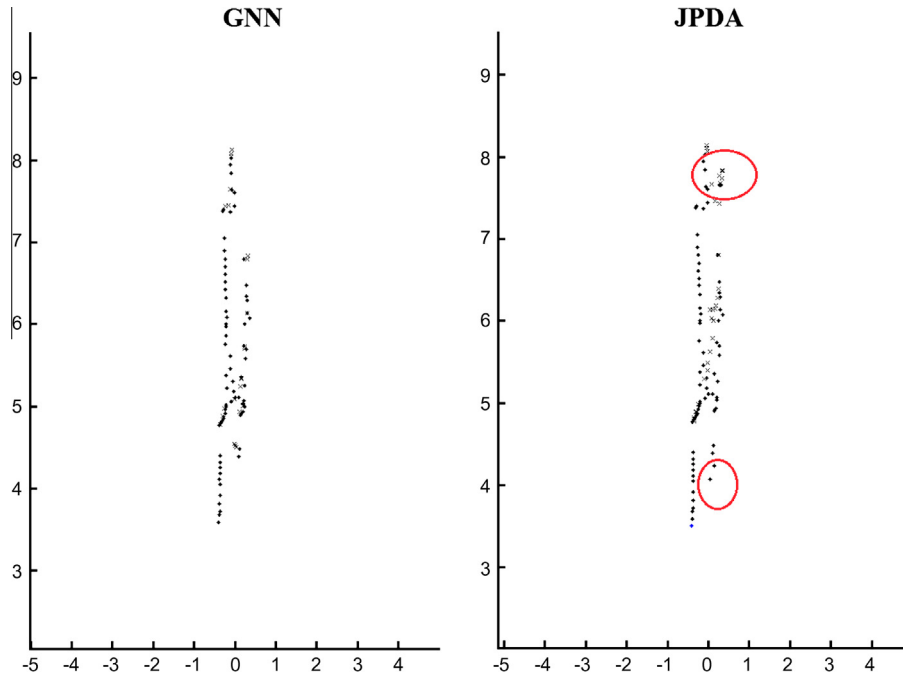


Fig. 15. Example of the performance of the algorithms GNN and JPDA, with the same sequence, with two pedestrians walking very close to each other. (·) Represent actualized tracks, and (×) represents non actualized tracks.

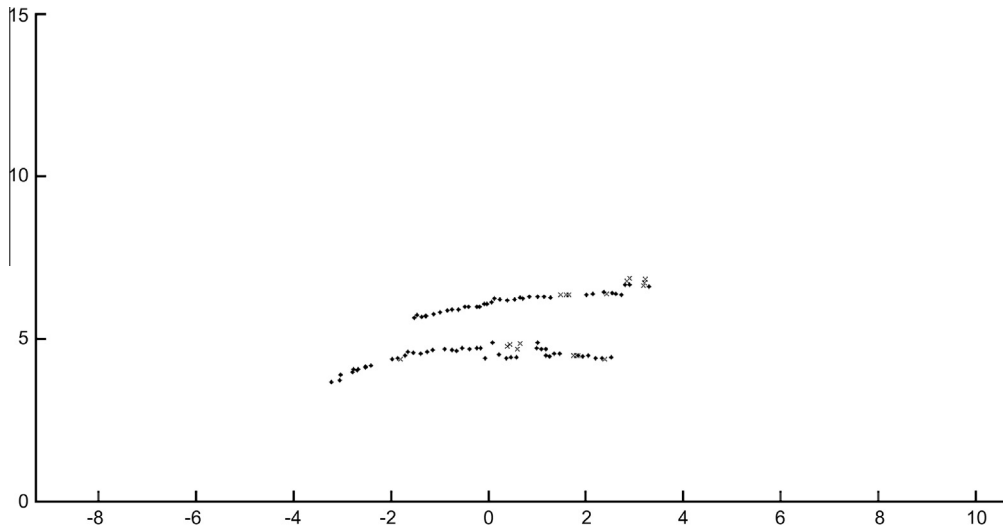


Fig. 16. Results for the tracking of the pedestrians, where two pedestrians crossing are tracked. Dots represent normal detections. Crosses represent detections where the track is not actualized. The vehicle was moving at 20 km/h and stops before the pedestrians.

formance of the fusion process was studied, including an extended comparison with other fusion processes based on Global Nearest Neighbors (GNN). In this way, it is possible not only corroborating the improvement that fusion can provide to the classic ADAS systems based on single sensor approaches, but also helping to understand the performance of the novel fusion procedure presented in this paper.

7.1. Low level algorithms

Results for the tests for the low level algorithms are depicted in Tables 3 and 4, and some examples are shown in Fig. 13.

Given the results, some conclusions can be obtained from the low level approaches:

- Laser scanner is a novel approach giving very good performance, although the results obtained lack reliability since a percentage of 13.3% of false positives means that less than every ten frames a false positive is given. Consequently, false positives are frequent.
- Performance of the computer vision system is high and reliable, mainly in more structured scenarios, but in more complex scenarios such as urban environment, this reliability decreases mainly due to the abundant information present in this environment that can lead to misdetection.

It has to be remarked that the system setup was performed taking into account these situations, i.e. the threshold for positive detection of the computer vision was increased to add reliability to the detections of the laser scanner during the fusion process.

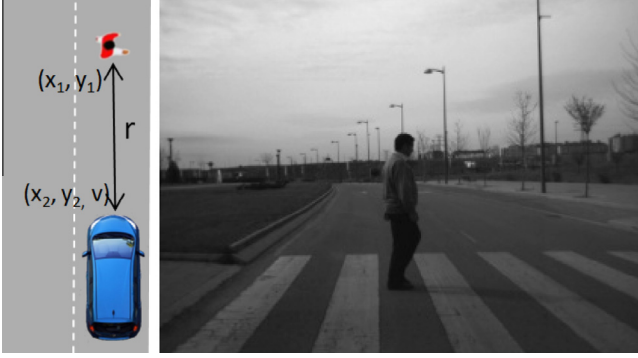


Fig. 17. Test scenario, with a pedestrian crossing the road.

- False positives provided by the camera were mainly unrelated to the HOG features algorithm presented in chapter 4, but related to the nature of the entire algorithm that was based on the ROIs selected by the laser scanner. As explained before, The ROIs are chosen according to the lectures from the laser; thus, if a spurious observation is returned by the laser, and a pedestrian is included in more than one ROI, all of them are going to be considered pedestrian, and as a result a false positive is returned. Although they could be considered positive detection for the computer vision algorithm itself, the approach provided here is based on both sensors, and therefore, they were included in false positive detection (Fig. 14).

7.2. Tracking and fusion

After single-sensor tests, the fusion system was studied using the same sequences indicated above. Besides individual sensor performance, a GNN implementation based on the one presented in Garcia et al. (2011) was used for comparison. This way the performance of the system was tested, not only in relation to the isolated sensors but also to the effect of other data association approaches. Results are depicted in Table 5.

The improvement of the algorithms with respect to the performance of the low level approaches is shown in Table 5. This shows

that the JPDA approach used in this paper represents an improvement to the basic association based on GNN. The improvement that JPDA represents with respect to GNN is related to the better behavior of the algorithm in specifically difficult situations (i.e. clustering errors, double detections and occlusions).

Figs. 15 and 16 depicts two examples of sequences.

Fig. 15 provides a comparison of two algorithms that ran in the same sequence, one used GNN approach and the other one used JPDA. In this sequence, two pedestrians are walking toward the vehicle (Fig. 12(a)). Differences are highlighted in red. The overall performance is better in the case of JPDA, providing tracking for longer time, even in this extremely difficult situation.

Fig. 16 depicts another complex situation with crossing pedestrians involved. In this case, even with occlusions all the pedestrians are able to be tracked during the entire sequence.

With the provided results, the enhanced performance of the system, compared to the low level approaches, and to other data fusion algorithms, is proved. By combining laser scanner with the vision based approach, the reliability of the overall system is increased. It can be done thanks to the laser scanner based obstacle detection, and by requiring both sensors to confirm a given detection. Finally it was also proved that JPDA can improve performance in highly complex situations.

7.3. Danger estimation tests

Danger estimation algorithms were tested, based on a pedestrian crossing scenario (Fig. 17). The context information used was based on the on-line information of the vehicle, using the GPS-INS system and the information of the test platform (e.g. mass center and dimensions).

Fig. 18 depicts the different results of the test with the estimation values according to the velocity of the test vehicle and the distance to the detected pedestrian. As the scope of the application is urban scenarios, where pedestrians are commonly found, test performed were based on specific scenarios where velocity is limited to 40 km/h, and distances are smaller (less than 40 m).

Finally, Fig. 19 shows the evolution of the estimation according to these values, proving the utility of the danger estimation. It provides a fast way to represent the danger that involves any detection according to the context information.

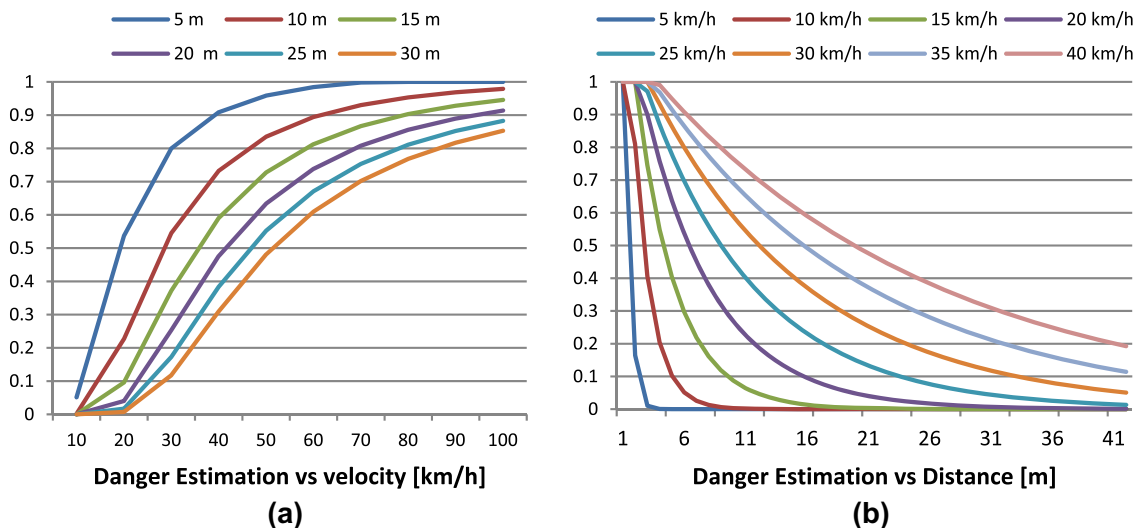


Fig. 18. (a) Danger estimation versus velocity, based on different cases, where pedestrian is located at distances from 5 to 30 m from the vehicle. (b) Danger estimation versus distance to the detected pedestrian, based on different situation, with vehicle velocity from 5 km/h to 40 km/h.

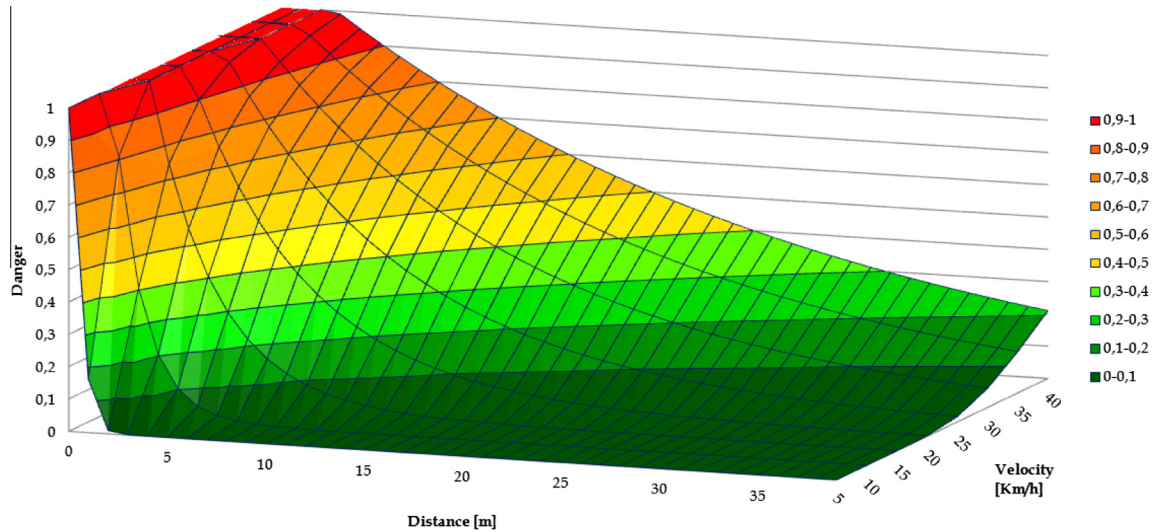


Fig. 19. Danger estimation evolution according to the two on-line variables (i.e. distance to the pedestrian and vehicle velocity).

8. Conclusions

A novel fusion approach for pedestrian detection in road safety application has been presented. Results have proved that, by adding laser scanner pedestrian detection and context information, both performance and trustability are increased, fulfilling the demanding requirements of these applications. The presented work gives a solution based on processes performed at different JDL levels.

The sensor fusion tests provided compare two data association techniques, the method proposed in the contribution (consisting in a JPDA approach) gave more reliable results than classic GNN.

Context aids at level 2 and level 3 have been also presented, giving a novel solution that takes advantage of the available information and knowledge, to provide a new context based danger estimation. Thus, by fusing three sensors and context information, trustable pedestrian detection is provided and danger associated to these detections is estimated.

Enhancement of the classic ADAS system was achieved thanks to the use of laser scanner, a proved trustable tool in novel road applications, for the ROI detection and independent pedestrian identification. Thus the presented algorithm provided more reliable and robust detection thanks to this high level fusion process. Besides, the independent detection, provides back up detection in case one of the sensors is temporary unavailable. The JPDA algorithm tested, proved to be more reliable in specific stress conditions, such as crossing or occlusions, improving the classical performance of pedestrians tracking algorithms. Finally the danger identification, allows to focus the attention in those pedestrians that may interfere in the driving process.

The main drawback of the presented work relies in the costs of the technologies and its high computational requirements. Laser scanner technologies are still too expensive to be included in commercial vehicles, although in latest years the increasing interest of the scientific community have lead to more robust and interesting applications that may provide added value worth to be included in commercial vehicles. Regarding to the computational costs, modern parallel processing techniques and powerful computers can allow to provide a real time application.

Based on the aforementioned points, the presented work represents a step forward in pedestrian detection for ADAS systems in three fronts. First providing a multi level approach that integrates low levels (pedestrian detection and tracking) with high levels (situation analysis and danger estimation). By this multilevel

approach, a full expert system, able to detect and identify the danger situations is provided. Second adding context information, which plays an important role for correct situation awareness that could prevent road accidents, enhancing the pedestrian detection and danger estimation with expert information, able to adapt the system performance to the specific application of road safety. Finally by adapting a powerful association technique (JPDA) for its use as data fusion technique for road safety. Proving advance performance under the most demanding situations: i.e. occlusions and misdetections.

By means of the data fusion algorithm presented, computer vision detection has been enhanced, providing reliable and trustable detection with danger estimation. Although vision approaches for pedestrian detection have proved to be a very useful application, ready for commercial application, most of the previous integration approaches try to speed up this detection by using ROIs given by a laser scanner. The presented fusion algorithm goes beyond this point, providing improved detection in two ways: By reducing the false positive rate by creating regions of interest based on a reliable laser scanner system and by providing a laser scanner pedestrian detection. The latter allows using the system in situations where computer vision is not available. Finally it is important to remark that the presented work provides a unique and novel multilevel expert solution taking advantage of the most powerful data fusion techniques.

Further steps are under consideration and may extend context-based fusion to augment the performance of the fusion process by adapting the procedures to the real time information (which is labeled as "level 4" fusion accordingly to JDL model). This adaptation may take under consideration the available context information. Furthermore, these next steps will take advantage of available information thanks to the new technologies, such as digital maps and driver identification. New and advance expert system, able to adapt the danger estimation to the real situation of the vehicle and the driver needs will be created, reducing the stress to the driver by the identification of those pedestrians that may interfere with the trajectory of the vehicle, and adapting the response of the system to the driver state.

Acknowledgments

This work was supported by the Spanish Government through the Cicyt projects (GRANT TRA2010-20225-C03-01), (TEC2011-28626-C02-02) and (GRANT TRA 2011-29454-C03-02), CAM

through SEGVAUTO-II (S2009/DPI-1509) and mobility program of “Fundación Caja Madrid”.

References

- Abellán, J., López, G., & De Oña, J. (2013). Analysis of traffic accident severity using decision rules via decision trees. *Expert Systems with Applications*, 40(15), 6047–6054.
- Bar-Shalom, Y., & Li, X.-R. (1995). *Multitarget-multisensor tracking: principles and techniques*. YBS Publishing (Vol. 16, no. 1, p. 600).
- Bhatt, D., Aggarwal, P., Devabhaktuni, V., & Bhattacharya, P. (2014). A novel hybrid fusion algorithm to bridge the period of GPS outages using low-cost {INS}. *Expert Systems with Applications*, 41(5), 2166–2173.
- Blackman, S. S. (1986). *Multiple-target tracking with radar application* (Vol. 1). Norwood, MA: Artech House.
- Blackman, S., & Popoli, R. (1999). *Design and analysis of modern tracking systems*. Norwood, MA: Artech House.
- Bohmlander, D., Doric, I., Appel, E., & Brandmeier, T. (2013). Video camera and capacitive sensor data fusion for pedestrian protection systems. In *2013 Proceedings of the 11th workshop on intelligent solutions in embedded systems (WISES)* (pp. 1–7).
- Broggi, A., Cerri, P., Ghidoni, S., Grisleri, P., & Jung, H. G. (2008). Localization and analysis of critical areas in urban scenarios. In *IEEE intelligent vehicles symposium* (pp. 1074–1079).
- Buehler, M., Iagnemma, K., Singh, S., Thrun, S., Montemerlo, M., Dahlkamp, H., et al. (2005). DARPA grand challenge 2005. *Industrial Robot: An International Journal*, 31(5), 1–43.
- Castro, J. L., Delgado, M., Medina, J., & Ruiz-Lozano, M. D. (2011). An expert fuzzy system for predicting object collisions. Its application for avoiding pedestrian accidents. *Expert Systems with Applications*, 38(1), 486–494.
- Collins, J. C. (1979). *Accident reconstruction*. Charles C. Thomas Publisher Limited.
- Conesa, J., Cavas-Martínez, F., & Fernández-Pacheco, D. G. (2013). An agent-based paradigm for detecting and acting on vehicles driving in the opposite direction on highways. *Expert Systems with Applications*, 40(13), 5113–5124.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE comput. soc. conf. comput. vis. pattern recognition, 2005. CVPR 2005* (Vol. 1, pp. 886–893).
- Defense, T., & Agency, P. (2006). Foreword for journal of field robotics – special issue on the DARPA grand challenge. *Journal of Field Robotics*, 23(9), 657–658.
- Fan, X., Mittal, S., Prasad, T., Saurabh, S., & Shin, H. (2013). Pedestrian detection and tracking using deformable part models and Kalman filtering. *Journal of Computer-Mediated Communication*, 10, 960–966.
- García, F., de la Escalera, A., Armingol, J. M., Herrero, J. G., & Llinas, J. (2011). Fusion based safety application for pedestrian detection with danger estimation. In *2011 Proceedings of the 14th international conference on information fusion (FUSION)* (pp. 1–8).
- García, F., Cerri, P., Broggi, A., de la Escalera, A., & Armingol, J. M. (2012). Data fusion for overtaking vehicle detection based on radar and optical flow. In *2012 IEEE intelligent vehicles symposium (IV)* (pp. 494–499).
- Gómez-Romero, J., Patricio, M. A., García, J., & Molina, J. M. (2011). Ontology-based context representation and reasoning for object tracking and scene interpretation in video. *Expert Systems with Applications*, 38(6), 7494–7510.
- Guo, L., Ge, P. S., Zhang, M. H., Li, L. H., & Zhao, Y. B. (2012). Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine. *Expert Systems with Applications*, 39(4), 4274–4286.
- Hall, D. L., & Llinas, J. (2001). *Handbook of multisensor data fusion*. CRC Press (pp. 3–18).
- Highway capacity manual 2000*. (2000). Transportation Research Board, National Academy of Sciences, no. 193.
- Iagnemma, K. & Buehler, M. (Eds.). (2006). *Special Issue on the DARPA Grand Challenge, Part 2, Journal of Field Robotics*, (Vol. 23, no. 9, pp. 655–656).
- Iagnemma, K., & Buehler, E. M. (2008). Special issue on the 2007 DARPA urban challenge. *Journal of Field Robotics*, 25, 423–860.
- Jo, J., Lee, S. J., Park, K. R., Kim, I.-J., & Kim, J. (2014). Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications*, 41(4), 1139–1152 (Part 1).
- Johansson, G., & Rumar, K. (1971). Drivers' brake reaction times. *Human Factors*, 13(1), 23–27.
- Kohler, M. (1997). *Using the Kalman filter to track human interactive motion - modelling and initialization of the Kalman filter for translational motion*. Citeseer.
- Li, D., Xu, L., Goodman, E. D., Xu, Y., & Wu, Y. (2013). Integrating a statistical background-foreground extraction algorithm and SVM classifier for pedestrian detection and tracking. *Integrated Computer-Aided Engineering*, 20(3), 201–216.
- Ludwig, O., Premebida, C., Nunes, U., & Ara, R. (2011). Evaluation of boosting-SVM and SRM-SVM cascade classifiers in laser and vision-based pedestrian detection. In *IEEE intelligent transportation systems conference ITSC* (pp. 1574–1579).
- Makishita, H., & Matsunaga, K. (2008). Differences of drivers' reaction times according to age and mental workload. *Accident Analysis and Prevention*, 40(2), 567–575.
- Martí, E. D., Martín, D., García, J., de la Escalera, A., Molina, J. M., & Armingol, J. M. (2012). Context-aided sensor fusion for enhanced urban navigation. *Sensors (Basel)*, 12(12), 16802–16837.
- Musleh, B., García, F., Otamendi, J., Armingol, J. M., & De La Escalera, A. (2010). Identifying and tracking pedestrians based on sensor fusion and motion stability predictions. *Sensors*, 10(9), 8028–8053.
- Pérez Grassi, A., Frolov, V., & Puente León, F. (2010). Information fusion to detect and classify pedestrians using invariant features. *Information Fusion*, 12(4), 284–292.
- Premebida, C., Ludwig, O., Silva, M. & Nunes, U. (2010). A cascade classifier applied in pedestrian detection using laser and image-based features. In *IEEE intell. transp. syst. conf. ITSC* (pp. 1153–1159).
- Premebida, C., Ludwig, O., & Nunes, U. (2009). LIDAR and vision-based pedestrian detection system. *Journal of Field Robotics*, 26(4), 696–711.
- Premebida, C., & Nunes, U. J. C. (2013). Fusing LIDAR, camera and semantic information: A context-based approach for pedestrian detection. *International Journal of Robotics Research*.
- Sánchez, A. M., Patricio, M. A., García, J., & Molina, J. M. (2009). A context model and reasoning system to improve object tracking in complex scenarios. *Expert Systems with Applications*, 36(8), 10995–11005.
- Schneider, N., & Gavril, D. M. (2013). Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study. In *Pattern Recognition* (pp. 174–183). Springer.
- Shao, X. S. X., Zhao, H. Z. H., Nakamura, K., Katabira, K., Shibasaki, R., & Nakagawa, Y. (2007). Detection and tracking of multiple pedestrians by using laser range scanners. In *2007 IEEE/RSJ int. conf. intell. robot. syst.* (no. 1, pp. 2174–2179).
- Skehill, R. J., Barry, M., & Mcgrath, S. (2005). Mobility modelling with empirical pedestrian and vehicular traffic characteristics. *WSEAS Transactions on Communications*, 4(10).
- Spinello, L. & Siegart, R. (2008). Human detection using multimodal and multidimensional features. In *2008 IEEE int. conf. robot. autom.* (pp. 3264–3269).
- Still, G. K. (2000). *Crowd dynamics*. University of Warwick.
- Xavier, J., Pacheco, M., Castro, D., Ruano, A., & Nunes, U. (2005). Fast line, arc/circle and leg detection from laser scan data in a player driver. In *Proc. 2005 IEEE int. conf. robot. autom.* (Vol. 4, no. April, pp. 3930–3935).
- Zhao, J., Chen, Y., Zhuang, X., & Xu, Y. (2014). Posterior probability based multi-classifier fusion in pedestrian detection. In J.-S. Pan, P. Krömer, & V. Snášel (Eds.), *Genetic and evolutionary computing SE – 35* (Vol. 238, pp. 323–329). Springer International Publishing.