García, F.; Escalera, A.; Armingol, J. M. (2013). "Joint Probabilistic Data Association fusion approach for pedestrian detection," *Intelligent Vehicles Symposium (IV), 2013 IEEE*, Gold Coast, QLD, 23-26 June 2013, pp. 1344-1349.
DOI: 10.1109/IVS.2013.6629653

Joint Probabilistic Data Association Fusion Approach for Pedestrian Detection

Fernando García, Arturo de la Escalera and José María Armingol

[1]Intelligent Systems Lab. Universidad Carlos III de Madrid, Spain.

Abstract— **Fusion is becoming a classic topic in Intelligent Transport System (ITS) society. The lack of trustworthy sensors requires the combination of several devices to provide reliable detections. In this paper a novel approach, that takes advantage of the Joint Probabilistic Data Association technique (JPDA) for data association, is presented. The approach uses one of the most powerful techniques of Multiple Target Tracking theory and adapts it to fulfill the strong requirements of road safety applications. The different test performed proved that a powerful association technique can enhance the capacity of Advance Driver Assistance Systems.**

Two main sensors are used for pedestrian detection: laser scanner and computer vision. Furthermore, the approach takes advantage of the availability of other information sources i.e. context information and online information (GPS). The detections are fused using JPDA, enhancing the capacities of classical pedestrian detection systems, mainly based in visual information.

The test performed also showed that JPDA improved the results offered by other data association techniques, e.g. Global Nearest Neighbors.

I. INTRODUCTION

FUSION is a classical topic in Intelligent Transport System (ITS) society. The lack of trustworthy sensors requires the combination of several devices to provide reliable detections, able to fulfill the strong requirements of road safety applications.

In the present paper, two classic sensors are used i.e. Computer Vision and 2D Laser Scanner. First provides a considerable high amount of unstructured information, thus any classification requires a high computational cost and lacks of reliability. On the other hand, information provided by the 2D laser scanner is more reliable, thanks to the trustworthiness of the technology used, but limited to a single layer. Modern laser scanners overcome this last problem by providing 3D detection, but this technology is still economically unaffordable for road applications. Combining the information provided by the two sensors, the classic Advance Driver Assistant Systems (ADAS) can be enhanced and thus the limitations of each sensor can be overcome.

Typically fusion applications are solution oriented and does not pay attention to the classical Data Fusion (DF) methodology. DF tries to provide a general framework to the fusion problem. In this context the Joint Probabilistic Data Association is presented as one of the most powerful tools for data association. It represents a highly adaptable solution that provides very good results even in the most demanding situations. The present work provide solution, by using JPDA, to the fusion problem of laser scanner and computer vision in road scenarios.

II. GENERAL DESCRIPTION

Two sensors were available for environment detection and classification, laser scanner and computer camera. Laser scanner is mounted in the bumper of the test platform IVVI 2.0 (Figure 1) and camera is installed in the front windshield.

Figure 1. Test platform IVVI 2.0. (Intelligent Vehicle based on Visual Information). Left the vehicle. Right the laser scanner mounted in the bumper of the vehicle. Center, closer look of the laser scanner sensor.

Each sensor provides single sensor detection (low level detections). A subsequent stage combines the information from low level, providing fused detections (tracks).

After testing different configurations, the laser scanner provided a higher reliability in the detection of objects, thus it was used to provide Region Of Interest (ROI) to the images. This configuration can be easily changed by using pin-hole model for distance estimation, in the case that the laser scanner is not available.

Fusion stage provides estimation of the movement of the pedestrians by a Kalman Filter. The association of the new detections with the previous detections (tracks) is performed using the JPDA approach.

III. STATE OF THE ART

Fusion approaches can be divided in decentralized and centralized schemes:

In **centralized** scheme fusion is performed by an unique classification system, able to retrieve information from several sensors, providing a single classification based on

the features obtained from the combined set of information. It generally requires a preprocessing stage that creates the features vector based in information from the sensors. In [1] and [2] authors present and compare decentralized schemes, that performs pedestrian classification in different ways i.e. Naïve Bayes, Gaussian Mixture Model Clasifiers, Neural Networks, Fuzzy Logic Decision Algorithm and Support Vector Machines.

Decentralized schemes are based in independent classifiers, that perform the classification according to the information of one or several sensors independently. A final stage performs the final classification, according to the low level classifications and their certainties. [3] 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 [4] 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. In [5] low level detection is provided based in pattern matching for laser scanner, and stereovision with vertical projection of human silhouette for computer vision detection, the fusion stage is based in a Global Nearest Neighbor (GNN) approach.

Other approaches takes advantage of the special behavior of the different sensors to solve different situations, enhancing the capacities of the single sensor based approaches: [6] uses information from laser scanner to search particular zones of the environment where pedestrians could be located and visibility is reduced, such as the space between two vehicles, and performs detections in these regions using a vision approach.

The work presented is an example of decentralized scheme based in two independent low level classifiers (one for laser scanner and another one based in computer vision) and a final fusion stage, based in a powerful Multiple Target Tracking (MTT) algorithm, JPDA. The decentralized approach represents a more robust application, able to provide detection even in extreme situations, when any of the sensors is not available. Furthermore, the JPDA approach represents a highly adaptable algorithm, able to overcome difficult situations in the tracking stage.

IV. LOW LEVEL DETECTION

As it was depicted before, low level detection is performed independently by each sensor, allowing to have a more robust system, able to provide detection even in situations where one of the sensors is not available.

A. Laser scanner System

Before reconstruction, the information retrieved by the laser scanner is provided with a given delay among the distances provided by the laser scanner. This delay has to be corrected according to the movement of the vehicle. This movement was corrected with a GPS device with inertial enhancement from Xsens. It was mounted attached to the laser scanner, providing accurate on-line velocity and Euler angles estimation. Equations (1-3) depicts the different corrections performed to the distances provided by the laser scanner, according to the euler angles of the movement of the vehicle.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R\left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T_v + T_0\right) \tag{1}$$

$$R = \begin{bmatrix} \cos(\Delta\delta) & 0 & sin(\Delta\delta) \\ 0 & 1 & 0 \\ -\,sin(\Delta\delta) & 0 & cos(\Delta\delta) \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 \\ 0 & cos(\Delta\varphi) & -sin(\Delta\varphi) \\ 0 & sin(\Delta\varphi) & cos(\Delta\varphi) \end{bmatrix} \cdot \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ sin(\Delta\theta) & cos(\Delta\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$T_v = \begin{bmatrix} vT_i \cdot cos(\Delta\theta) \\ vT_i \cdot sin(\Delta\theta) \\ 0 \end{bmatrix}, T_0 = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} \tag{3}$$

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 movement compensation, laser scanner detection is performed, based in the movement of the pedestrian. A deep study of the movement of the pedestrian was performed that allowed to create a pattern for pedestrian classification. This pattern is based in the movement of the two legs while walking (Figures 2 and 3).

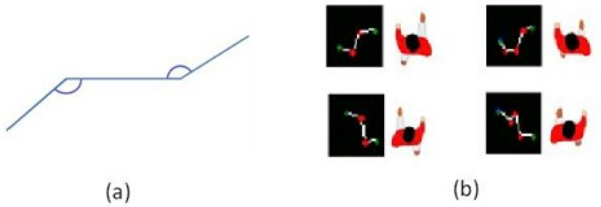

Figure 2: Laser scanner pattern (a), examples of leg movement (b).

The pattern consists on consecutive polylines fulfilling several constraints regarding to angles and sizes [5]. Rotation of the pattern allows to extend the detection to lateral and diagonal movements.

A higher stage computes the movement of the pedestrian along time. This way the final classification takes into account the last 10 detections by a voting scheme. Besides

several filters eliminates false positive detections by detecting impossible movements, accelerations, velocities, etcetera. All this information was based in context information, regarding to physical constraints and road information [5].

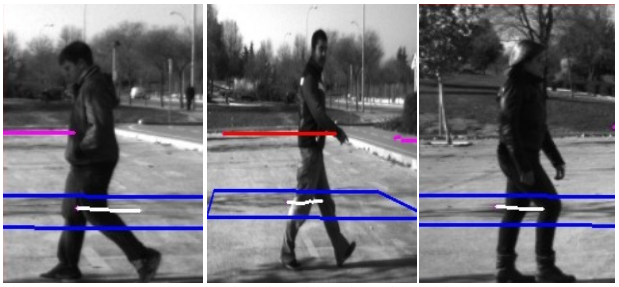

Figure 3: Laser scanner pedestrian detections based in legs pattern.

B. Camera

The camera system is based in HOG features approach [7]. Using the information retrieved by the laser scanner to limit the region of the image to search. This way the false positives are reduced, since only obstacles with size similar to a pedestrian are checked (Figure 4).

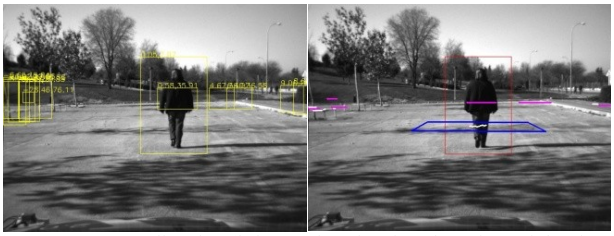

Figure 4: A. Bounding box(left) and pedestrian detection (right) in red box.

V. FUSION PROCEDURE

Fusion procedure is based in a Multiple Target Tracking (MTT) approach. Typically, MTT approaches have two stages, estimation and data association. In the present work, first is based in a Kalman Filter approach with constant velocity model. It resulted accurate enough thanks to the fast acquisition frequency of the laser scanner (~ 20 Hz). Second is based in the JPDA approach.

A. Estimation

As it was remarked before, estimation is based in constant velocity model system and Kalman Filter. For completeness the model is presented in (3-7):

$$\hat{} = \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} \tag{3}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{4}$$

$$F = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$Q = \begin{bmatrix} \frac{{a_x}^2 t^3}{3} & \frac{{a_x}^2 t^2}{2} & 0 & 0 \\ \frac{{a_x}^2 t^2}{2} & {a_x}^2 & 0 & 0 \\ 0 & 0 & \frac{{a_y}^2 t^3}{3} & \frac{{a_y}^2 t^2}{2} \\ 0 & 0 & \frac{{a_y}^2 t^2}{2} & {a_y}^2 \end{bmatrix} \tag{6}$$

$$R = \begin{pmatrix} {\sigma^2}_{\epsilon,x} & 0 \\ 0 & {\sigma^2}_{\epsilon,y} \end{pmatrix} \tag{7}$$

where ${\sigma^2}_{\epsilon,x}$ y ${\sigma^2}_{\epsilon,y}$ is the standard deviation for the measurements in x, y coordinates. $\hat{X}$ is the state vector of the Kalman Filter, H the observation model and F the state transition model. The errors are modeled by Q and R which are the covariance of the process noise and the covariance of the measurement noise respectively.

B. JPDA for pedestrian detection

Association techniques in ITS are typically based in best distance matching, this approach is known as Global Nearest Neighbors (GNN). In the present approach GNN is used for results comparison, based in the work presented in [5].

JPDA is an extension of PDA Filter([8] and [9]) which was developed for single target tracking. JPDA extends PDA to a number of targets M. The measurements at time k are denoted as $z_k = \{z_k^j\}$, where j goes from 0 to m_k. A clutter (z_0) is introduced (artificial measurement to provide mathematical completeness).

By assuming a Markovian process and using Bayes theorem, the joint association probability of an association can be described as follows.

Let θ denote the joint association event, and ${\theta_k}_j^m$ the particular event that associates measurement m to a track j. The joint association probabilities are defined as:

$$P(\theta|Z_k) = \frac{1}{K} p(z_k|\theta, X_k) P(\theta|X_k) \tag{8}$$

where K is a normalization constant, X_k is the target state vector. $P(\theta|X_k)$ is the probability of the assignment θ conditioned to the sequence of the target sequence states vector which is defined as:

$$P(\theta|X_k) = {P_D}^{M-n} (1 - P_D)^n {P_{FA}}^{m_k-(M-n)} \tag{9}$$

where P_D is the probability of detection, which can be empirically calculated. n is the number of assignments to the clutter (z_0) and P_{FA} is the false alarm probability that also can be obtained empirically.

Finally the association likelihood ($p(z_k|\theta, X_k)$) is defined assuming a 2 dimensional Gaussian association likelihood, for all the measurements to the target. The joint probability of a single measurement j to the target i would be:

$$g_{i,j} = \frac{1}{(2\pi)^{N/2}\sqrt{|S_{ij}|}} e^{-\frac{d_{i,j}^2}{2}} \quad (10)$$

where $d_{i,j}$ is the Euclidean distance between the prediction and the observation. $S_{i,j}$ is the residual covariance matrix. Since a Cartesian approach was used $\sqrt{|S_{ij}|} = \sigma_x\sigma_y$ and N=2.

Thus finally the resulting $P(\theta|Z_k)$ is:

$$P(\theta|Z_k) = P_D{}^{M-n}(1-P_D)^n P_{FA}{}^{m_k-(1-M)} \prod_{j=1}^{m_k} g_{i,j} \quad (11)$$

Finally all the association hypotheses are weighted in the updating stage of the Kalman filter. The innovation is calculated using all possible combination weighted for the likehood for this association.

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

where R_k is the innovation covariance for the Kalman Filter of a given track.

C. Tack management

Track management is based in the definition of consolidated and non consolidated tracks. First refers to those tracks with positive detection provided by both sensors. The second are those tracks detected by a single sensors, thus with not enough certainty to be reported.

Track creation and deletion policy has a key role in the algorithm:

- A new track is created when a given detection falls out of the gates of all the available tracks i.e. There is no match for the given detection.
- A track is eliminated if no detection is included within the gate after a given number of frames. This process is defined as track maintenance. It refers to the process of maintaining a given track along time, if a new observation falls within the gate. The track logic defined limited the new detections to be used only for the maintenance of a single track. Thus when a given detection is included in the gate of more than one track, it is used for maintenance only of the highest match. Although in the updating process of the filter this observation is used in all the.

Test demonstrated that the presented algorithm could, in certain situations, reach to unstable behavior. This is when several tracks compete for a single observation. In these situations, the cluster 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, a special behavior was created. It consisted on that once a given association is assigned, the associated pair track-new detection is eliminated from the assignation process. So for the next assignment all the joined probabilities are recalculated with the remaining tracks. This way the problem is avoided by eliminating the weight of the already assigned solutions in subsequent assignations. In the case of several tracks pointing to a single observation, this solution would first assign cluster to the less probable detection and eliminate the weight of this detection in subsequent assignations, until one of them is selected as more likely than the cluster. 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.

VI. Results

Several test were performed including comparison with other data association approach i.e. Global Nearest Neighbors (GNN) [5]. Both systems provided similar results, but there are specific situation where JPDA provides special behavior that helps the system to overcome specific problems. These situations are clustering errors (Figure 5), double detections (Figure 6) and crossings or occlusions (Figure 7).

A. Clustering errors

These errors are produced due to the difficulty of separating different obstacles by the laser scanner, when they are very close to each other and at a certain distance. This problem is very difficult to solve using the fusion approach presented, based in laser scanner clustering. But the association algorithm helps to overcome the problems generated due to this inconvenient. In these situations, where two pedestrians merge and separate into a single obstacle several times in the same sequence, the updating stage of the Kalman Filter uses the single observation given by the laser scanner to update both tracks, thus the errors produced by this inconvenient are negligible.

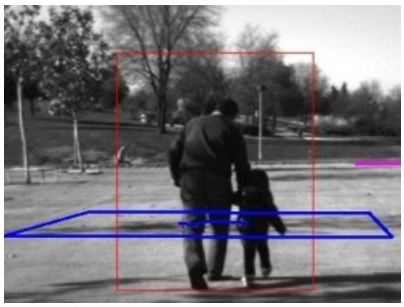

Figure 5: Example of challenging situation for data association, laser scanner clustering errors.

B. Double detections

In this case, the situations are related to the superposition of several bounding boxes, as it is shown in (Figure 6). This error can be caused due to two factors: First the laser scanner is not located in the same plane than the camera, thus the projection of several obstacles can be superposed in the camera plane. Second cause is clustering errors or misdetections of the laser scanner due to dust or other particles e.g. rain and fog. In all these situations several bounding boxes include the same pedestrian. The JPDA approach deals with this situation in an efficient way. Since generally both detection falls in the same gate, they are computed for the same obstacle, but as the wrong detection

have a lower probability due to the higher diversity with the expected value, the effect of the misdetection is low. Other approaches, such as GNN, would generate a new track, maintaining the false positive for several frames.

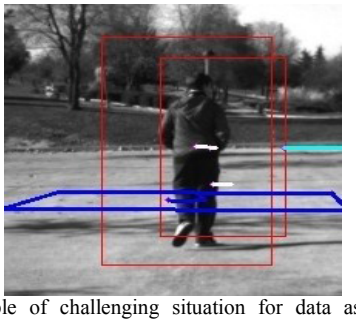

Figure 6: Example of challenging situation for data association, laser scanner errors.

C. Occlusions or crossings

In these situations, the pedestrian at the back is not visible due to the occlusion by another pedestrian. Both pedestrians are generally close, so the single detection falls into the gate of the missing pedestrian. This way, the single detection obtained is used for both sensors to update the Kalman Filter of the misdetected pedestrian, allowing more accurate movement estimation.

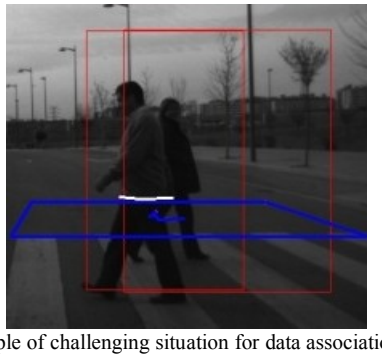

Figure 7: Example of challenging situation for data association, crossings or occlusions.

Figure 8 and Figure 9 depicts an example of a sequence where the laser scanner algorithm has difficulties to differentiate among two very close pedestrians. In the results provided in Figure 9, it is highlighted the main differences, showing the better performances of the JPDA approach to deal with specifically difficult situations. This improvement by the JPDA approach was visible in numerous sequences, as it is depicted in the results provided by Table 1.

Several test were preformed, including urban and interurban scenarios with more than 10,000 frames, the results for both low and high level are depicted in table 1.

	% of positive detections	% misdetections (per frame)
Camera	72.97	5.27
Laser Scanner	74.56	13.3
Fusion (GNN)	79.62	2.21
Fusion (JPDA)	**82.29**	**1.11**

Table 1. Test Results.

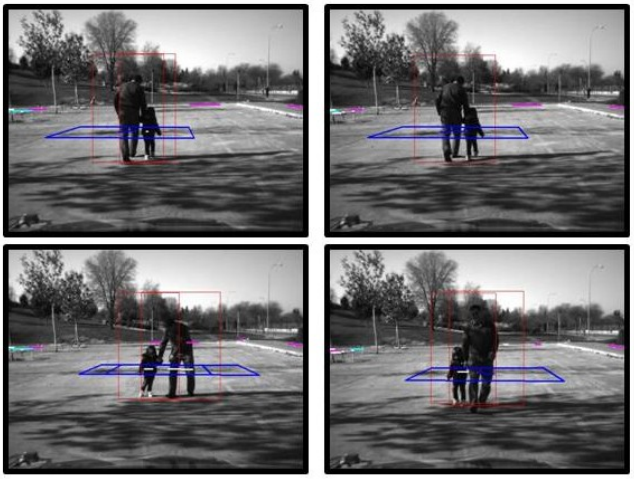

Figure 8: Images of a sequence with two pedestrians. The two pedestrians walk very close, in several situations the laser scanner is unable to separate among them. Blue boxes represents laser scanner detection, red boxes represents the vision detections.

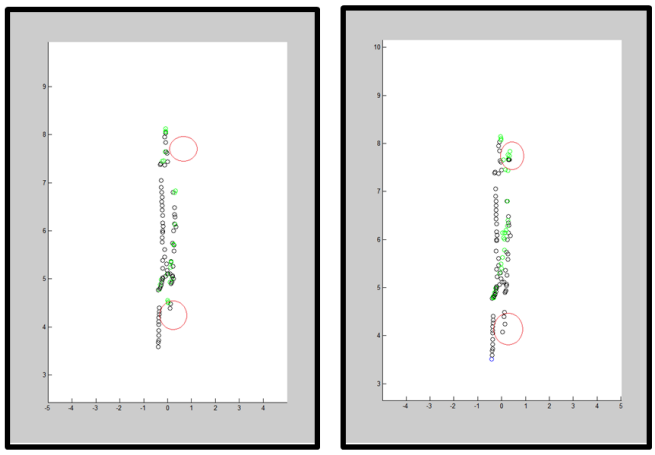

Figure 9: Results for the tracking and data association for test sequence (Figure 8), GNN (left) and JPDA (right). The main differences are highlighted. The axis represents the distance in meters to the laser scanner in y and x coordinates. Green detections are tracks with no match with the new detections, black are tracks with match.

It is interesting to highlight the high positive rate for laser scanner, although a high false positive rate was also expected due to the extremely difficulty of classifying pedestrians using the limited information provided by the laser scanner. Here is where fusion with the camera detection algorithm has an important role.

It has to be remarked that the training process for the camera was performed taking into account the laser scanner results, thus the system was trained to provide good false positive performance, penalizing the high positive rate.

The results depicted in Table 1 show that by the fusion scheme, it is possible to increase the positive detection rate of single sensor systems and to improve the misdetection rate. Besides, JPDA proved to be a more efficient tracking approach than classic approaches such as GNN.

VII. CONCLUSIONS

A fusion system based in laser scanner and camera detection is presented. The system provides decentralized

scheme able to perform independent detection for each sensors and to fuse information at high level. The system is able to enhance the subsystems detections, overcoming the limitation of each one. Besides, it was proved that JPDA approach represents a better association problem than GNN for this specific fusion process.

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] C. Premebida, O. Ludwig, M. Silva, and U. Nunes, "A Cascade Classifier applied in Pedestrian Detection using Laser and Image-based Features," *Transportation*, pp. 1153-1159, 2010.

[2] C. Premebida, O. Ludwig, and U. Nunes, "LIDAR and Vision-Based Pedestrian Detection System," *Journal of Field Robotics*, vol. 26, no. Iv, pp. 696-711, 2009.

[3] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, "A Lidar and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking," *IEEE Intelligent Transportation Systems Conference ITSC*, pp. 1044-1049, 2007.

[4] L. Spinello and R. Siegwart, "Human detection using multimodal and multidimensional features," *2008 IEEE International Conference on Robotics and Automation*, pp. 3264-3269, 2008.

[5] F. Garcia, B. Musleh, A. D. Escalera, and J. M. Armingol, *Fusion procedure for pedestrian detection based on laser scanner and computer vision*. IEEE, 2011, pp. 1325-1330.

[6] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H. G. Jung, "Localization and analysis of critical areas in urban scenarios," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 1074-1079.

[7] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05*, vol. 1, no. 3, pp. 886-893, 2004.

[8] Y. Bar-Shalom and X.-R. Li, *Estimation and tracking - Principles, techniques, and software*. Boston: Artech House.

[9] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems(Book)*. 1999.