This is a postprint version of the following published document:

Garcia, F.; Ponz, A.; Martin, D.; Escalera, A.; Armingol, J. M. (2014). Computer vision and laser scanner road environment perception. *IWSSIP 2014 Proceedings*, Dubrovnik, 2014, pp. 63-66.

Computer Vision and Laser Scanner Road Environment Perception

Fernando Garcia, Aurelio Ponz, David Martin, Arturo de la Escalera & José María Armingol

Universidad Carlos III de Madrid,

Avenida de la Universidad, 30 (28911), Leganés. Madrid

fegarcia@ing.uc3m.es

Abstract - Data fusion procedure is presented to enhance classical Advanced Driver Assistance Systems (ADAS). The novel vehicle safety approach, combines two classical sensors: computer vision and laser scanner. Laser scanner algorithm performs detection of vehicles and pedestrians based on pattern matching algorithms. Computer vision approach is based on Haar-Like features for vehicles and Histogram of Oriented Gradients (HOG) features for pedestrians. The high level fusion procedure uses Kalman Filter and Joint Probabilistic Data Association (JPDA) algorithm to provide high level detection. Results proved that by means of data fusion, the performance of the system is enhanced.

***Keywords* - Pedestrian Detection; Computer Vision; Data Fusion; Laser Scanner; ADAS.**

I. INTRODUCTION

Traffic accidents are one of the main problems related with road transport. In the latest years, the efforts to improve the security in both vehicles and roads have lead to a reduction of the number of fatalities, but there is still a lot of work to be done. Recent advances on information technologies have lead to new applications that apply to the road transports these advances, helping to increase the security and assisting in the driving process. These applications are called Advanced Driving Assistance Systems (ADAS).

The strong requirements of road safety applications and the extreme difficulty to provide reliable detections, leads to the necessity of combining different set of sensors. By means of the fusion of the information from the different sensors sets, the drawbacks of the each sensor are overcome, allowing to fulfill these strong requirements, and thus providing reliable detection and tracking of the different users in the road.

The present paper explains a novel approach for road safety, based on data fusion using laser scanner and computer vision for vehicle and pedestrian detection. On section II, state of the art is presented, section III provides general description of the system and the research platform IVVI 2.0 is presented. Section IV provides description of the low level detection and section V details the tracking algorithm. Finally section VI provides results of the different test performed and some conclusions.

II. STATE OF THE ART

Data Fusion approaches can be divided on low, medium and high level data Fusion. First is related to the fusion of the raw data received from the sensors, thus a new set of more complex data is created. In [2] and [3] stereovision is used to provide pedestrian detection. Stereovision is a typical example of low level data fusion, and has been widely used for road security applications.

A preprocessing stage is necessary for medium level data fusion. The data is preprocessed and features based on each sensor are obtained, a fusion feature vector based of features from both sensors is created. In [3] and [4] medium level fusion works are presented, they combine features and perform classification based on different machine learning algorithms.

In high Level fusion approaches, classification is performed based only on the information provided by each sensor independently, a final fusion stage performs the fusion based on the reliability of the detection and the sensor itself. In [5] pedestrian detection is performed based on visual Adaboost detection and Gaussian Mixture Model (GMM) for laser scanner, finally a Bayesian decisor is used to combine detections at high level. In [6] multidimensional features are used for laser scanner detection; Histograms of Oriented Gradients (HOG) features and Support Vector Machine (SVM) for computer vision detection, finally high level fusion is performed based on a Bayesian Model.

Figure 1. Research platform IVVI 2.0.

III. GENERAL DESCRIPTION

The presented application is a high level data fusion approach, for vehicle and pedestrian detection. The system provides trustable detection by means and accurate allocation of the obstacles in the surroundings. The application is included in the scope of the IVVI 2.0 project (Figure 1). The IVVI 2.0 project is a platform of a commercial vehicle, in which several sensors were included providing a test bench for the development and test of ADAS applications.

Within the scope of the project, several applications were developed, including driver monitoring systems for both vision [7] and time of flight camera [8], night-vision pedestrian detection [9], stereovision based visual odometry [10], pedestrian detection [2] and obstacle detection [11], advance positioning systems [12], lane detection [13] and traffic sign detection [14] and [15], among other works.

IV. LOW LEVEL DETECTION

As mentioned before, the low level detection was based on laser scanner and computer vision, performing independent detection.

In order to provide trustable detection, Region Of Interest (ROI) detection is based on laser scanner due to the high reliability of the laser scanner detection and the accuracy of distances provided to the obstacles.

These regions of interest are sorted according to the size, eliminating the non-likely obstacles in the environment. This way two sets of possible obstacles are created, one for pedestrians and another for vehicles. These obstacles are later checked using information from each sensor independently.

A. laser scanner

The nature of the application and the laser scanner used implies the necessity of movement compensation, due to the delayed acquisition of the information of the distances to the obstacles. The movement of the vehicle is estimated using a GPS combined with inertial measurements, thus the movement in the time elapsed between detections is compensated using this information. This compensation is based on rotation and translation matrixes, as shown in (1).

$$
\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T_v + T_0) \text{ , with } T_v = \begin{bmatrix} vT_i \cdot cos(\Delta\theta) \\ vT_i \cdot sin(\Delta\theta) \\ 0 \end{bmatrix},
$$

$$
T_0 \quad \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} \text{, and} \tag{1}
$$

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

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

where $\Delta\delta$, $\Delta\varphi$ and $\Delta\theta$ are the changes of the Euler angles roll, pitch and yaw respectively for the time elapsed T_i in relation to the last distance received by the laser scanner, thus all the detections are extrapolated to the time when the data is received. (x_0,y_0,z_0) are the Cartesian coordinates of a given point before movement compensation and coordinates (x,y,z) are the coordinates after the movement compensation. The parameter R is the matrix that defines the rotation, T_v is the translation vector due to the velocity of the vehicle, T_0 (x_t,y_t,z_t) is the translation vector between the position of the laser scanner and the inertial sensor. v is the velocity of the car.

After movement compensation, shapes of the obstacles are estimated. This shape estimation is based on polylines [16]. On

this approach, two main obstacles are searched, vehicles [16] and pedestrians [17]:

- For vehicles, the pattern matching algorithm is presented in [16]. The pattern is based on the specific pattern given by the delayed distance points due to the movement of the vehicle. The delayed spots allow not only to detect the vehicles, but also to identify the velocity and direction of them.
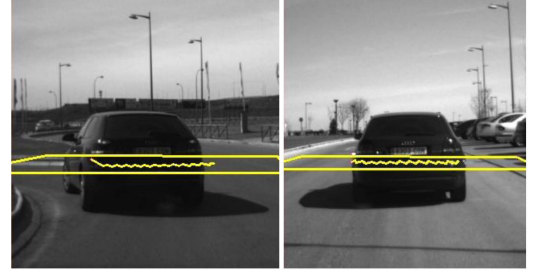

Figure 2. Laser scanner vehicle detection with the given pattern.

- Pedestrian are detected according to the pattern of a walking pedestrian. The pattern is based on three polylines connected, representing the leg of a human being, the angles that connect the consecutive polylines are checked to be included within the angles of $[0,\pi/2]$ and a Similarity score is created (2).

$$
\text{Similarity} \quad \frac{2\theta_1}{\pi} \cdot \frac{2\theta_2}{\pi} \tag{2}
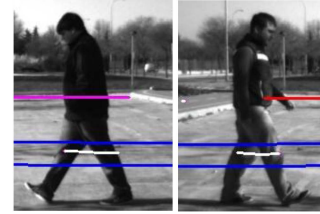$$

Figure 3. Laser scanner pedestrian detection.

A tracking stage is added based on the laser scanner information, this way the limited information provided by the laser scanner can be integrated along time, providing more robust classification. A voting scheme is used to classify pedestrians, based on the obstacle decision in the last 10 detections, a distributed multi-features approach is used for obstacle correlation along time [16]. Finally, some filters are added in order to reduce the false positives detections, these filters check impossible movements, such as lateral movements in vehicles, or big changes of shapes to reduce the miss-detection rates.

B. Computer vision

Region of interests received from the laser scanner, are used to search the obstacles within the image. The use of these ROIs helps to reduce the computational cost of the approach and adds reliability, since thanks to the trustability of the laser scanner to detect obstacles, the probability of false detections is also reduced.

Once the ROIs are received, coordinate changes should be performed using pin-hole model and accurate extrinsic calibration. The extrinsic calibration process is based on the rotational and translation equations presented on (1) to provide information from the laser scanner to the camera coordinate system. An online calibration process was used (Figure 4):
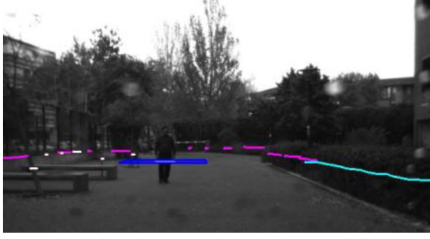

Figure 4. Laser scanner polilynes extrapolated to the camera field.

- Vehicles visual identification was performed based on Haar-Like features [18]. The system was trained for both frontal and back panel detection. These parts of the vehicles have very similar characteristics, such as lights or bumpers that make them easy to detect by this fast algorithm .

- Pedestrian detection is based on Histogram of Oriented Gradients (HOG) features. This approach, proposed on [19], is a classical approach for human detection.
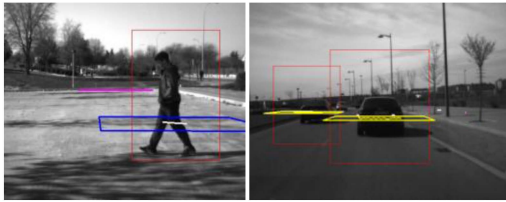

Figure 5. Example of visual pedestrian detection (left), and vehicle (right).

V. TRACKING ALGORITHM

Tracking algorithm is based on a unimodal MTT algorithm Joint Probabilistic Data Association Algorithm (JPDA) and the estimation Kalman Filter (KF). The movement model was defined using the constant velocity target model shown in equations 3 to 7.

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

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

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

$$R = \begin{pmatrix} \sigma^2_{\epsilon,x} & 0 \\ 0 & \sigma^2_{\epsilon,y} \end{pmatrix} \tag{6}$$

$$Q = \begin{bmatrix} \frac{a_x^2 t^3}{3} & \frac{a_x^2 t^2}{2} & 0 & 0 \\ \frac{a_x^2 t^2}{2} & a_x^2 & 0 & 0 \\ 0 & 0 & \frac{a_y^2 t^3}{3} & \frac{a_y^2 t^2}{2} \\ 0 & 0 & \frac{a_y^2 t^2}{2} & a_y^2 \end{bmatrix} \tag{7}$$

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

The JPDA algorithm adapted for ADAS application was based on the definition of consolidated and non consolidated tracks. First refers to those tracks where the positive detections are provided by both sensors. Non consolidated are those tracks that were positive by a single sensor, but they are tracked for some time steps, in case the other sensor also provides positive detections. In that case the non consolidated track becomes consolidated track.

Data Association algorithm, based on the JPDA technique ([20] and [21]), adapted here for ADAS application, allows tracking even in case of difficult situations such as occlusions and double detections. JPDA is defined as follows:

Lets denote z_k $\{z_k^j\}$, to the detections given by the sensors at a time k. With j going from 0 to m_k. A clutter (artificial measurement that represents no association, j=0) is introduced. Furthermore, assuming a Markovian process and using Bayes theorem, the joint association probability for a given association can be described as :

Defining θ as the joint association, and $\theta_k{}^m_j$ the particular association of measurement m to a track j. The joint association probabilities can be written as:

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

where, X_k is the target state vector, C the normalization constant. $P(\theta|X_k)$ is the probability of the assignment θ given the states of the targets at a given time k, defined as:

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

where P_D is the detection probability for a given target and P_{FA} is the false alarm probability, both given by the definition of the sensors. n is the number of assignments to the clutter, m_k is the number of detections given by the sensors and M is the number of targets being tracked.

Finally the association likelhood ($p(z_k|\theta, X_k)$) is defined here, based on a bidimensional gaussian association likelihood. Thus the joint probability of a measurement j to a target i is:

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

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

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

Finally all the association hypotheses that falls within the gate of a single track are weighted in the updating stage of the KF, thus all possible hypothesis are taken into account in the updating process of the KF:

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

where I_k is the innovation covariance for the KF of every track.

VI. Results and conclusions

Test performed in both urban and interurban scenarios in more than 1,000 frames showed that the improvement of the fusion procedure led to positive detections of 77.69% for pedestrians and 88.25 % for vehicles. The misdetection frames was of 3.11 % of misdetection per frame for vehicles and 2.59% for pedestrians.

The results proved that the system was able to enhance the low level approaches, increasing the positive rate, thus providing better results. Thus as a conclusion, a novel fusion procedure was presented which combines information from the camera and the laser scanner to overcome the limitations of each sensor, providing enhanced detection and improving classical ADAS application .

Acknowledgements

References

[1] M. Bertozzi, A. Broggi, M. Felisa, S. Ghidoni, P. Grisleri, G. Vezzoni, C. Hilario Gómez, and M. Del Rose, "Multi Stereo based Pedestrian Detection by means of Daylight and Far Infrared Cameras," in *Object Tracking and Classification Beyond the Visible Spectrum*, R. I. Hammoud, Ed. Springer Verlag, 2009, pp. 371 401.

[2] C. Hilario, J. Collado, J. Armingol, and A. La Escalera, "Pedestrian Detection for Intelligent Vehicles Based on Active Contour Models and Stereo Vision," *Computer Aided Systems Theory – EUROCAST 2005*. pp. 537 542, 2005.

[3] C. Premebida, O. Ludwig, M. Silva, and U. Nunes, "A Cascade Classifier applied in Pedestrian Detection using Laser and Image-based Features," *IEEE Intell. Transp. Syst. Conf. ITSC*, pp. 1153 1159, 2010.

[4] C. Premebida, O. Ludwig, and U. Nunes, "LIDAR and Vision Based Pedestrian Detection System," *J. F. Robot.*, vol. 26, no. Iv, pp. 696 711, 2009.

[5] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, "A Lidar and Vision based Approach for Pedestrian and Vehicle Detection and Tracking," *IEEE Intell. Transp. Syst. Conf. ITSC*, pp. 1044 1049, 2007.

[6] L. Spinello and R. Siegwart, "Human detection using multimodal and multidimensional features," *2008 IEEE Int. Conf. Robot. Autom.*, pp. 3264 3269, 2008.

[7] M. J. Flores, J. M. Armingol, and A. de la Escalera, "Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions," *EURASIP J. Adv. Signal Process.*, vol. 2010, p. 3, 2010.

[8] G. Peláez, F. García, A. Escalera, and J. M. Armingol, "Obtaining a 3D Model from a Facial Recognition in 2D," in in *Computer Aided Systems Theory EUROCAST 2013 SE 5*, vol. 8112, R. Moreno Díaz, F. Pichler, and A. Quesada Arencibia, Eds. Springer Berlin Heidelberg, 2013, pp. 33 38.

[9] D. Olmeda, A. de la Escalera, and J. M. Armingol, "Far infrared pedestrian detection and tracking for night driving," *Robotica*, vol. 29, no. 04, pp. 495 505, 2011.

[10] B. Musleh, D. Martin, A. Escalera, D. Guinea, and M. Garcia Alegre, "Estimation and Prediction of the Vehicle's Motion Based on Visual Odometry and Kalman Filter," in in *Advanced Concepts for Intelligent Vision Systems SE 43*, vol. 7517, J. Blanc Talon, W. Philips, D. Popescu, P. Scheunders, and P. Zemčík, Eds. Springer Berlin Heidelberg, 2012, pp. 491 502.

[11] B. Musleh, A. D. La Escalera, and J. M. Armingol, "Obstacle Detection and Localization Using U V disparity with Accelerated Processing Time by CUDA," in *IEEE International Conference on Robotics and Automation 2011 ICRA*, 2011.

[12] E. D. Martí, D. Martín, J. García, A. de la Escalera, J. M. Molina, and J. M. Armingol, "Context aided sensor fusion for enhanced urban navigation.," *Sensors (Basel).*, vol. 12, no. 12, pp. 16802 37, Jan. 2012.

[13] J. M. Collado, C. Hilario, A. De La Escalera, and J. M. Armingol, "Adaptative road lanes detection and classification," in *Transform*, 2006, vol. 4179 LNCS, pp. 1151 1162.

[14] G. A. Pelaez, M. R. Munoz, J. M. Armingol, A. De la Escalera, J. J. Munoz, W. van Bijsterveld, and J. A. Bolano, "Road sign recognition for automatic inventory systems," *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*. pp. 1 4, 2011.

[15] J. P. Carrasco Pascual, "Advanced Driver Assistance System based on Computer Vision using Detection, Recognition and Tracking of Road Signs," Universidad Carlos III de Madrid, 2009.

[16] F. García, F. Jiménez, J. E. Naranjo, J. G. Zato, F. Aparicio, J. M. Armingol, and A. de la Escalera, "Environment perception based on LIDAR sensors for real road applications," *Robotica*, vol. 30, no. 02, pp. 185 193, 2012.

[17] F. García, F. Jiménez, J. J. Anaya, J. M. Armingol, J. E. Naranjo, and A. de la Escalera, "Distributed Pedestrian Detection Alerts Based on Data Fusion with Accurate Localization," *Sensors*, vol. 13, no. 9, pp. 11687 11708, 2013.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR 2001*, vol. 1, no. C, pp. I 511 I 518, 2001.

[19] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Comput. Vis. Pattern Recognition, 2005. CVPR 2005. IEEE Comput. Soc. Conf.*, vol. 1, pp. 886 893, 2005.

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

[21] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Artech House, Norwood, MA, 1999.