Hybrid Fusion Scheme for Pedestrian Detection based on Laser Scanner and Far Infrared Camera

F. Garcia, D. Olmeda, J. M. Armingol and A. de la Escalera.
Intelligent System Lab, Universidad Carlos III de Madrid.

Abstract—**The lack of trustworthy sensors makes ADAS applications a tough task. In this paper a fusion approach that uses two sensors: far infrared camera vision and a laser scanner to detect pedestrian for ADAS applications is presented. Both sensors have different field of view thus different detection zones are created according to the number of sensors available. The proposed algorithm combines low and high level information, providing a hybrid fusion scheme. Experimental results show that pedestrian detections are improved thanks to the use of several sensors.**

I. INTRODUCTION

TRUSTWORTHY sensors are key elements regarding current road safety applications. In recent years, advances in information technologies have lead to more intelligent and complex applications which are able to deal with a large variety of situations. These new applications are known as ADAS (Advance Driver Assistant Systems). In order to provide reliable ADAS application one of the principal task involved are obstacles detection, especially those obstacles that represent the most vulnerable road users, pedestrians. However, the lack of trustworthy sensors capable of detecting those users and anticipate to potentially dangerous situations underlines the need to use different sensors combined to provide a reliable and accurate application.

In this paper a novel fusion method is proposed. The method combines the information provided by a 2D laser range finder and far infrared camera vision to detect pedestrians in a road environment. Taking advantage of the different capabilities of each sensor and detection ranges, a pedestrian detection is performed independently. In regions where ranges overlap, a fusion algorithm deals with the redundant information, increasing the positive detection results and avoiding false positives.

II. STATE OF THE ART

Fusion methods are typically divided in two sets according to the level in which fusion is performed. Low level Fusion schemes, so called centralized fusion schemes [1]. They perform fusion using a set of features extracted from both sensors. High level fusion, so called decentralized schemes perform different classifications with data provided by each sensor separately. A final stage combines information from all classifications. Each configuration has its own advantages and disadvantages [2]. Low level fusion combines information from both sensors creating a new set of data with more information. But problems related to data association arise. High level fusion schemes allow fusion in an easier and more scalable way, new sensors can be added more easily but with less information to do the classification.

High level fusion schemes can be differentiated in track based fusion and cell based fusion schemes. First tries to associate the different objects found in each sensor [3]. Second one [4] uses occupation grids, adding confidence according to the kind of sensor that detects the obstacle, but losing the geometrical structure.

Low level approaches that take advantage of statistical knowledge are presented in [1], and [5] , these approaches obtain information from all sensors and combine the information using whether Bayes formula, SVM, Neural Networks, etc …

Other works related to fusion schemes take advantage of laser scanner trustworthiness to select regions of interest where vision based systems try to detect pedestrians [6] and [7]. In [8] Detection of especially dangerous zones is done using laser scanner information integrated along time.

[9] Performs what some authors call medium level, it processes information from different sensors creating a feature vector which is used to perform an unique classification.
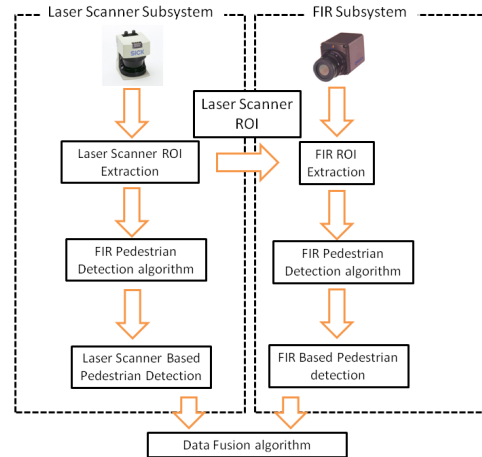

Fig. 1. Flow diagram for the whole system.

All authors are members of Intelligent Systems Laboratory, Universidad Carlos III de Madrid, Leganes, 28911 Madrid, Spain. E-mail: {fegarcia,dolmeda,armingol,escalera}@ing.uc3m.es.

III. General System Description

A hybrid fusion system is presented in this paper. As it is shown in figure 1, two detection subsystem where created each of them detects pedestrian separately using different sensors.

Detection algorithms usually are not able to detect pedestrians in the whole range, although sensors can operate in greater regions. Limitations inherent to the sensors and to the algorithms create different zones where subsystems can operate.

A. Fusion Zone

In this zone both sensors are able to give detection information. Its range is from 5m to 30m.

Maximum distance is given by far infrared detection algorithm performance, over 30 meters it has been estimated that information is not enough to give a good detection. Below 5 meters, the camera is not able to give the complete shape of a typical pedestrian (figure 2). It has been estimated average height of a pedestrian 1,70 meters, according to this distance and given a vision angle of the camera (19°) minimum distance is 5m.

$$\text{tg}\left(\frac{\propto}{2}\right) = \frac{\frac{h}{2}}{d} \;,\; d \approx 5 \; meters \tag{1}$$

where h= 1.70 meters and α=19°.

Although eq. 1 supposes that far infrared (FIR) camera is located at the medium height of a human being. In real world it is not always in this way, but it is a good estimation of the distance in which the system has information enough to perform detection. Test performed to the FIR algorithm corroborates these assumptions and sets minimum detection
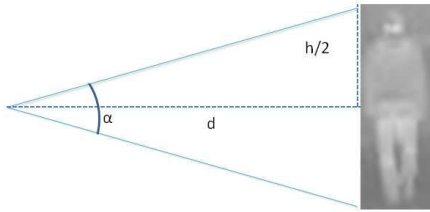

Fig. 2. Scheme of estimation of minimum distance camera detection.

distance at 5m.

B. Laser Scanner Zone

Laser Scanners have a wider field of view where detections can be performed. For the present work, laser used was a LMS 291 S-05 from SICK. The laser scanner configuration was 100° of field of view, a maximum distance of 80 meters and an angular resolution of 0.25°. The algorithm created to detect pedestrians, presented in this work, is able to detect pedestrians at maximum distance if data provided by the laser is enough.

IV. Laser Subsystem

The algorithm consists on low level detection and identification, tracking information is also computed. According to the movement of the obstacles their position in the following detection can be predicted. This tracking stage not only records and predicts the movement of vehicles and pedestrians, it is also useful to provide more accurate detections.

The algorithm is composed by two stages (figure 3). In the first stage, the data is received and a low level identification is performed. In the second stage, the data is integrated along time thus a higher classification level is obtained. This stage uses information provided by the previous detections to predict the movement of the pedestrian giving more robust classification.

A. Stage 1.Low Level Detection

Low level detection consists of four steps.

Egomotion Correction

To avoid misdetection due to the time difference between the spots, data received by the laser is corrected according to the movement of the vehicle. Egomotion information is recorded by a GPS sensor included in the system. This GPS Sensor from Xsens (MTI- G) is equipped with gyroscopic and magnetic sensors that correct and improve the data measured from the GPS system also included. It provides positioning data and egomotion information such as velocity, acceleration, Euler angles (absolute and angular velocity) etc... Velocity and yaw angle measurements are used to compensate the movement of the car for each pulse separately (eq. 2 to 5):
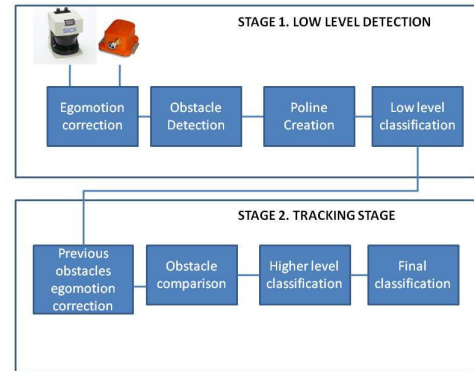

Fig. 3. Processes involved in detection and the classification algorithm using laser scanner.

Translation compensation:

$$x = x_0 - v \cdot T_i \cdot cos(\Delta\varphi) \tag{2}$$
$$y = y_0 - v \cdot T_i \cdot sen(\Delta\varphi) \tag{3}$$

Rotation compensation:

$$x = \cos(\Delta\varphi) \cdot x_0 - \sin(\Delta\varphi) \cdot y_0 \qquad (4)$$
$$y = \sin(\Delta\varphi) \cdot x_0 + \cos(\Delta\varphi) \cdot y_0 \qquad (5)$$

Where v is the velocity of the car, T_i the time between the given point and the first point and $\Delta\varphi$ is the increment in the yaw angle during a period of time Ti.

Obstacle Detection

The resulting points are joined according to the distance among them after egomotion correction. Clustering algorithm based on the Euclidean distance. A threshold which is distance dependant has been used (eq. 6).

$$th = th_0 + K \cdot dist \qquad (6)$$

here th_0 is the base threshold and K is a proportional constant which is multiplied by the distance.

Thus, for a given point $p(x_i, y_i)$ it may be treated as belonging to a segment S_j if it satisfies eq. 7:

$$p_i(x_i, y_i) \epsilon S_j \rightarrow \{\exists [p_j(x_j, y_j) \epsilon S_j]: d(p_j, p_i) < th\} \qquad (7)$$

The algorithm checks for all of the segments already created, if the case arises where a point is not included within any segment, a new segment is then created. Finally segments with only one point are eliminated while they are considered as false detection points.

Polyline creation

Once the segments are created, the points contained within each segment are merged using lines known as polylines. These lines are merged together according to the distance between the points included in it. This process is a variation of the Ramer algorithm [10]. The first and last points are merged using a line, for each point contained within this segment, the distance to the line is computed and if it is higher than a given threshold two new lines are created merging these three points. This process is repeated for every point within these new lines

Low level classification

Low level classification is performed with the information provided by the polylines according to the shapes of the different obstacles. Here different obstacles can be differentiated: **Small obstacles**, **road borders, possible vehicles**, **L shaped obstacles**, **pedestrians** and **other obstacles**.
For the present work, pedestrians and little obstacles are the obstacles that should be taken into account, other obstacles description can be found in [11] and [12].

1) small obstacles: This obstacle represents the regions of interest (ROI) that the subsystem considers could represent a pedestrian. This classification is based on the size of the detected obstacle.

2) Pedestrian: According to the shape of the detected polyline, a pedestrian can be differentiated. The algorithm used for this approach is based on the detection of two legs. The movement of a pedestrian is modeled using five possible situations.
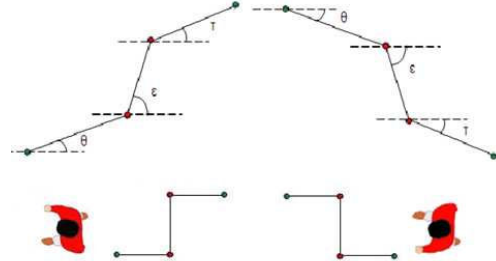

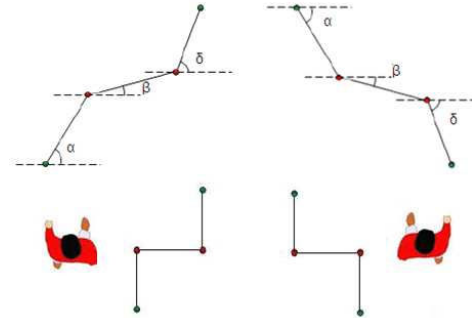Fig. 4. Models for movements perpendicular to the laser.

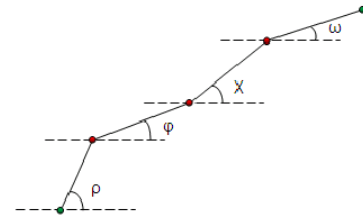Fig. 5. Models for movements parallel to the laser.

Fig. 6. General model for segments with more than three polylines.

Figure 4 and 5 shows movements perpendicular and parallel to the laser represented by three polylines. Model used in case of four polylines is presented in figure 6. The angles allowed for those models are shown below:

For parallel movements:

$$45^o \leq \alpha \leq 90^o$$
$$0^o \leq \beta \leq 55^o$$
$$45^o \leq \delta \leq 90^o$$

For perpendicular movements:

$$0^o \leq \theta \leq 55^o$$
$$35^o \leq \varepsilon \leq 90^o$$
$$0^o \leq \tau \leq 55^o$$

For four polylines segments two cases taken into account:

$$35^o \leq \rho \leq 90^o$$
$$0^o \leq \varphi \leq 45^o$$
$$35^o \leq X \leq 90^o$$
$$0^o \leq \omega \leq 45^o$$

$$0^o \leq \rho \leq 45^o$$
$$35^o \leq \varphi \leq 90^o$$
$$0^o \leq X \leq 45^o$$
$$35^o \leq \omega \leq 90^o$$

B. Stage 2. Higher Level Classification

A higher level stage is required to observe the behavior of different obstacles during a specific time period. Previously scanned obstacles are stored and verified using the new low level detection.

Previous obstacle egomotion correction

The egomotion correction is performed thus all the previously detected obstacles can be referenced to actual vehicle position.

The movement of pedestrians detected in previous scans is computed thus the following position is predicted for every obstacle. Obstacles are searched within a window according to the size of the obstacles in previous scans. If an obstacle is found within this window of the current scan, a comparison algorithm is used to verify if the obstacle is the same. If several obstacles have been found, most similar according to several parameters is considered as being the same obstacle.

Obstacle comparison

The Comparison process is carried out according to shape characteristics, i.e. the width and position. If the case arises where there are several possible candidates, the one with the closest value is considered as being the same obstacle (eq. 7).

$$Difference = \gamma_1 \cdot d_{width} + \gamma_2 \cdot d_{xmed} + \gamma_3 \cdot d_{ymin}$$
(7)

here γ_i is the weight applied to each distance of the different parameters that have been considered. d represents the distance from the previous detections to actual detection values.

Higher level classification

Once a pedestrian low level detection is performed, in the subsequent scans a similar obstacle representing whether a small obstacle or a pedestrian is considered a pedestrian if satisfies previously comparison algorithm according to some constraints relative to the movement and size. For each obstacle, a record of latest 10 low level detection is stored to give a higher level detection. This higher level detection takes into account ten movements using eq. 8.

$$V_i = \delta_i N_i$$
(8)

here V_i represents the number of votes for each type of obstacle, δ_i is a gain factor associated with each obstacle, and N_i is the number of times that an obstacle has been considered as being this type during last ten low level detections. Higher level classification chooses the higher V_i for all possible classes of obstacles.

$$p = 100 \cdot \frac{V_i}{(10 \cdot \delta_i)}$$
(9)

According to the p formula (eq. 9), for values where the occurrence is 10, the certainly (p) will be 100%.

V. FAR INFRARED SUBSYSTEM

Camera model

The proposed system makes use of raw images of the microbolometer, obtained through a 14 bit A/D. The gray level of each pixel of these images represents the amount of heat that the sensor captures. The camera used is based on the non-refrigerated microbolometer and, as such, its sensibility changes in a way that is function of its temperature. It is necessary, then, to calibrate that sensibility.

The camera is modeled as a pin-hole. The intrinsic characteristics are known and so is the position and orientation of the camera.

The homography of the ground plane onto the sensor is calculated for each frame (eq. 10) to determine the position. The projection of a 3D point in the image plane can be done if it is known its relative position to a certain plain. Camera position related to the ground is known considered constant. The rotation of the camera is known via the gyroscopes included in MTI-G previously introduced.

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = P \cdot W \begin{bmatrix} X \\ Y \\ Z \\ S \end{bmatrix}$$
(10)

where P is the intriniscs matrix. The camera movement between frames is modelled as W, which is comprised of the rotation matrix R and the translation vector T from the camera coordinate system to the ground. U and V are the image homogenous coordinates, being the true pixel coordinates $u = \frac{U}{S}$ and $v = \frac{V}{S}$.

$$P = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$
(11)

$$W = [R \quad T] =$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(\frac{\pi}{2} + \theta) & -sin(\frac{\pi}{2} + \theta) & t_y \\ 0 & sin(\frac{\pi}{2} + \theta) & cos(\frac{\pi}{2} + \theta) & 0 \end{bmatrix} \quad (12)$$

The results of the projection of a world point in the image plane are especially sensitive to variations of the skew angle θ. In the presented system this angle is known for each capture frame with the help of an accelerometer with three degrees of freedom attached to the same base as the camera, these accelerometers are also included in MTI-G system form Xsens. A point on the ground plane is projected on the image as:

$$u = c_u - \frac{X \cdot f_u}{Y \cdot sin(\frac{\pi}{2} + \theta)} \quad (13)$$

$$v = c_v - \frac{t_y \cdot f_v}{Y \cdot sin(\frac{\pi}{2} + \theta)} - \frac{f_v}{tan(\frac{\pi}{2} + \theta)} \quad (14)$$

where f_u and f_v are the focal lengths on the u and v directions of the image; c_u and c_v are the coordinates of the center of the images. These four parameters are measured in pixels.

The intrinsic are obtained in a calibration process that involves the use of a special chessboard pattern, a matrix of incandescent lamps.
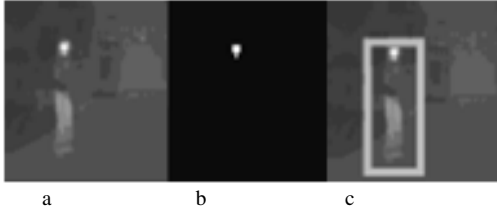

a   b   c

Fig. 7. (a) Image received from the camera. (b) Binarized image with a head. (c) ROI selected.

Finally, knowing that the image coordinates are $u = \frac{U}{S}$ and $v = \frac{V}{S}$. The relation between those and the world coordinates of the ground plane can be calculated with eq. 13 and 14.

ROI check

Objects within the normal temperature of the human body are thresholded. The result is a binarized image, containing blobs that can represent parts of the human body with a high temperature difference, especially heads and hands (figure 7). Since this first phase searches for the human heads, only blobs in the upper half of the image are taken into account. Also blobs that are not within a reasonable size are also excluded.

Both ROIs given by the laser and heads detected by the far infrared camera were taken into account thus missdetection due to a bad thresholding when looking for heads where avoided, and also missdetection due to bad laser segmentation.

Once the head candidates have been selected, a first set of

regions of interest are generated and a bounding box created. Top of the box is chosen highest point of the head, while the lowest point is at the closest point of the ground at that resolution. This way, the whole body of the pedestrian is included in the box, if there is any The width of the bounding boxes is set to be 3/7 of the height, as it is a usual proportion of the human body. A first approach is to suppose that the head is 200cm from the ground plane on which the pedestrian is walking. The distance of the pedestrian to the camera is given by eq. 15, where w_z = 2m. The base of the region of interest is calculated with equation 15, for this new distance.

$$w_y = \frac{(wz - h) \cdot \alpha_u}{v - v_0} \quad (15)$$

A second set of region of interest is created using information given by the laser. If this region has no correspondence with far infrared region detected, new boxes are created. With laser information width is provided, but no height. An inverse proceeding is performed taking into account that height is 7/3 times the width

The regions of interest generated from the original image are now binarized with a threshold that is the lower temperature established for the human body.

Vertical Symmetry Pedestrians edges on far infrared images are usually very well defined against a cool background. Human beings present a very high vertical symmetry, so it can be used to separate them from the non-pedestrian class. Only the ROIs obtained in previous steps of the algorithm are considered, so vertical edges of each pedestrian have to be symmetrical around an axis located approximately at w_i /2, where w_i is the width of each ROI i.

Eq. 16 defines this symmetry, normalized by the size of the box.

$$S = \sum_{i=0}^{h} \left[\sum_{j=0}^{w} \left(I_P \left(i, \frac{w}{2} + j \right) \cdot \left(I_N(i, \frac{w}{2} - j) \right) \right) \right] \cdot \frac{1}{w \cdot h} \quad (16)$$

where IP and IN are the images of the positive and negative Sobel filter, respectively; w and h are the width and height of the ROI. If the symmetry S is not over a certain threshold, the ROI is discarded.

Correlation with non-deformable models

Final verification of the extracted regions is done by means of gray scale correlation with some precomputed models. Correlation takes place between the final ROIs extracted thresholded and the models, whose creation process is explained in eq. 17.

$$S = \sum_{i=0}^{N} \frac{R_i(x,y)}{N} \quad (17)$$

where N is the number of ROIs selected for the creation of each model. In this case, every model has been generated out

of N = 50 candidates.

ROIs that eventually contain a pedestrian are grouped in four different categories: open, almost open, almost closed and closed legs. An example of these models can be seen on figure 8 (a to d). To reduce the number of calculations a fifth model is created for a common characteristic of pedestrians: the head. An example of this model can be seen on figure 8 (e). For each candidate, the top third of the bounding box is correlated with this head model before carrying on with the other four.
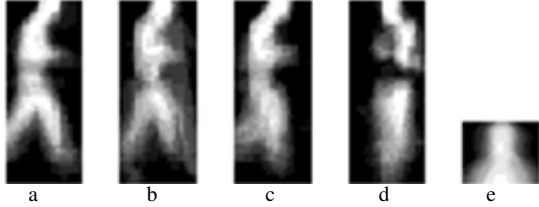

Fig. 8. Pedestrian models.

Besides computational time improvement, the use of this head model can prevent misdetections of due occlusions, i.e. standing behind a parked car

To take account of all distances, models have been created for four distance ranges: 5 to 15m, 15 to 25m, 25 to 40m.

Correlation is done by eq. (18), introduced in [12]:

$$c = \frac{\sum_{i=0}^{N}[(p(x,y)_i - 0.5)(M(x,y)_i - 0.5)]}{\sum_{i=0}^{N}[(p(x,y)_i - 0.5)]} \qquad (18)$$

where p(x, y) is each pixel of candidate ROI and M (x, y) is each pixel of the model.

Tracking stage

An Unscented Kalman Filter was used to perform tracking in fir stage. The Unscented Kalman Filter (UKF) [13] and [14] extends the general Kalman filter to non-linear transformations of a random variable without the need of linearization, as the Extended Kalman Filter (EKF) in eq. (13) and (14), this transformation is highly non-linear. Thus, the use of the UKF is justified over the EKF.

VI. FUSION ALGORITHM

A coordinates association should be done. Laser coordinate system was considered as reference. So obstacles are referred to the bumper of the car where the LIDAR was mounted. Fusion sensor calibration is performed using Least Mean Square (LMS)algorithm. The calibration algorithm takes a sample sequence with a single pedestrian. This pedestrian is detected by both subsystems separately and a LMS is used to obtain the coefficients to convert coordinates from fir coordinate system to laser scanner coordinate system.

At this point a feature vector can be obtained using information from both subsystems, these features are, velocity, probability from fir subsystem, probability from

laser scanner subsystem, width (from each subsystem), position (x,y), and previous detection results. They were introduced in an AdaBoost decision tree to perform the final decision based on both high level and low level information.

VII. RESULTS AND CONCLUSIONS

Results obtained for every subsystem are a detection percentage of 90% for pedestrian detection using laser scanner in distances shorter than 30 meters. Fusion regions results show that more than 96% of pedestrian has been detected.

REFERENCES

[1] C. Premebida, O. Ludwig and U. Nunes. "LIDAR and vision-based pedestrian detection system" in *Journal of Field Robotics*, vol. 26 Issue 9. Wiley, Sep. 2009. pp 696-711.

[2] N. Floudas, A. Polychronopoulos, O. Aycard, J. Burlet and M. Ahrholdt, "High Level Sensor Data Fusion Approaches For Object Recognition Road Environment," Intelligent Vehicles Symposium, 2007 IEEE , vol., no., pp.136-141, 13-15 June 2007.

[3] N. Kaempchen and K. Dietmayer, "Fusion of laserscanner and video for advanced driver assistance systems," in Proceedings of ITS 2004, 11th World Congress on Intelligent Transportation Systems, Nagoya, Japan, October 2004.

[4] O. Aycard,A. Spalanzani, J. Burlet, C. Fulgenzi, Dung Vu; D. Raulo, M. Yguel, "Grid Based Fusion & Tracking," *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE* , pp.450-455, 17-20 Sept. 2006.

[5] L. Spinello and R. Siegwart, (2008, May). "Human detection using multimodal and multidimensional features". *In IEEE International Conference on Robotics and Automation, 2008. ICRA 2008*, Pasadena, CA (pp. 3264-3269).

[6] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, D. "Cooperative Fusion for Multi-Obstacles Detection With Use of Stereovision and Laser Scanner". *Auton. Robots* 19, 2 .Sep. 2005, 117-140.

[7] M. Mahlisch, R. Hering, W. Ritter and K. Dietmayer, "Heterogeneous Fusion of Video, LIDAR and ESP Data for Automotive ACC Vehicle Tracking," *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on,* pp.139-144, Sept. 2006.

[8] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, Ho Gi Jung, "Localization and analysis of critical areas in urban scenarios*," Intelligent Vehicles Symposium, 2008 IEEE* , pp.1074-1079, 4-6 June 2008.

[9] N. Kaempchen, M. Buehler and K. Dietmayer, "Feature-level fusion for free-form object tracking using laserscanner and video," *in Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* , pp. 453-458, 6-8 June 2005.

[10] F. Nashashibi and A. Bargeton, "Laser-based vehicles tracking and classification using occlusion reasoning and confidence estimation*," Intelligent Vehicles Symposium, 2008 IEEE*, pp. 847–852, June 2008.

[11] F.Garcia, P. Cerri, A. Broggi, J. M. Armingol and A. de la Escalera, "Vehicle Detection Based on Laser Radar". Lecture Notes in Computer Science. Computer Aided System theory. Eurocast 2009. 393-397.

[12] D. Olmeda, C. Hilario, A. de la Escalera, and J. M. Armingol. "Pedestrian detection and tracking based on far infrared visual information". *Advanced Concepts for Intelligent Vision Systems, 2008*.

[13] S Julier and J Uhlmann. "A new extension of the kalman Filter to nonlinear systems. Int. Symp". *Aerospace/Defense Sensing*, Jan 1997.

[14] M. Meuter, U. Iurgel, S. Park, and A. Kummert. "The unscented kalman filter for pedestrian tracking from a moving host". *Intelligent Vehicles Symposium*, Jan 2008.