This is a postprint version of the following published document:

Martín, D., et al. (2014). IVVI 2.0: An intelligent vehicle based on computational perception.
*Expert Systems with Applications*, 41(17), pp. 7927-7944.

# IVVI 2.0: An intelligent vehicle based on computational perception

D. Martín *, F. García, B. Musleh, D. Olmeda, G. Peláez, P. Marín, A. Ponz, C. Rodríguez, A. Al-Kaff,

A. de la Escalera, J.M. Armingol

*Intelligent System Lab, University Carlos III de Madrid, Leganés, Spain*

## A B S T R A C T

This paper presents the IVVI 2.0 a smart research platform to foster intelligent systems in vehicles. Com putational perception in intelligent transportation systems applications has advantages, such as huge data from vehicle environment, among others, so computer vision systems and laser scanners are the main devices that accomplish this task. Both have been integrated in our intelligent vehicle to develop cutting edge applications to cope with perception difficulties, data processing algorithms, expert knowl edge, and decision making. The long term in vehicle applications, that are presented in this paper, out perform the most significant and fundamental technical limitations, such as, robustness in the face of changing environmental conditions. Our intelligent vehicle operates outdoors with pedestrians and oth ers vehicles, and outperforms illumination variation, i.e.: shadows, low lighting conditions, night vision, among others. So, our applications ensure the suitable robustness and safety in case of a large variety of lighting conditions and complex perception tasks. Some of these complex tasks are overcome by the improvement of other devices, such as, inertial measurement units or differential global positioning sys tems, or perception architectures that accomplish sensor fusion processes in an efficient and safe manner. Both extra devices and architectures enhance the accuracy of computational perception and outreach the properties of each device separately.

*Keywords:*
Intelligent transportation systems
Computer vision,
IR detection,
Data fusion,
Vehicle positioning.

## 1. Introduction

Traffic accidents are one of the main health risk problems. Glob ally, as the number of vehicles on the road increases so does the number of fatalities and injuries. As a consequence of road acci dents, every year approximately 1.2 million people are killed and 50 million disabled or injured (WHO, 2009). Not only are road traf fic accidents the eleventh cause of death in the world, but it is the only cause of death among the worst twelve which is not related to illnesses or diseases.

Human errors are the cause of most traffic accidents. Drivers' inattention and wrong driving decisions are the two main errors. Governments are trying to reduce the said accidents with infra structure improvement and educational campaigns, but they can not be completely eliminated due to the human factor. The speed of vehicles on the road is directly related to the risk factor associ ated with accidents and is also responsible for the consequences (WHO, 2008). Excess velocity on the road within areas of a deter mined speed limit and inappropriate driving speed are the main cause of traffic accidents, i.e. driving at an excess velocity when considering parameters such as: the driver, the volume of the traf fic, and the condition of the vehicle and the road. That is why Advanced Driver Assistance Systems (ADAS) can reduce the num ber, danger and severity of traffic accidents. Several ADAS, which nowadays are being researched for intelligent vehicles, are based on Artificial Intelligence, Laser and Computer Vision technologies (Guan, Bayless, & Neelakantan, 2012; Milanés et al., 2012).

ADAS are designed to help human drivers. Thus, there are examples of road lane detection (Collado, Hilario, de la Escalera, & Armingol, 2008; Zhou, Xu, Hu, & Ye, 2006), and obstacles recog nition and avoidance in the vehicle's path such as either vehicles (Musleh, de la Escalera, & Armingol, 2012b) or pedestrians (Musleh, de la Escalera, & Armingol, 2011; Soquet, Perrollaz, Labayrade, & Auber, 2007) or other elements, like traffic lights and marks on roads (Franke et al., 2001). ADAS are on board vehi cle systems which focuses on the driving process. One of the main objectives of this technology has been to increases driver aware ness by providing useful information.

There are a large variety of systems on the market that employ cameras based on Computer Vision, Radar, Light Detection Ranging (LIDAR) and Ultrasonic Sensors to support ADAS. These sensors are critical to support the aim of intelligent vehicles, such as the IVVI vehicle that has been conceived at Intelligent Systems Lab to

_____
∗ Corresponding author.
  *E-mail address:* dmgomez@ing.uc3m.es (D. Martín).

incorporate our developed technologies to foster the implantation of next generation of vehicles and its growing potential for the Transportation Sector (Fig. 1).

The advancement of sensors and computation devices have allowed in vehicle technology rapidly advancing and a complete standardization in present and forthcoming vehicles. The applications that require object detection, such as vehicle collision avoidance (Chang, Tsai, & Young, 2010), have reached a growing interest in vehicle manufacturers due to increasing safety conditions for both passengers and pedestrians (Guo, Ge, Zhang, Li, & Zhao, 2012). These advances have influenced a number of in vehicle applications already available in the automotive sector, such as a system developed by Mercedes that monitors the space available on both sides and behind the vehicle and that is capable of detecting other vehicles. It notifies the presence of other close by vehicles when changing lane (Daimler, 2013a). The parking assistant developed by Bosch (2013) employs a total of six sensors on the front and rear of the vehicle and measures the parking space which is indicated to the driver using an acoustic emission whose characteristics are in function of the size of the space available. Driver behaviour monitoring system is another safety application to be implemented in vehicles to a widespread transition to intelligent vehicles. Another driver monitoring system has been developed for Nissan which monitors the attention of the driver and detects possible symptoms of drowsiness (Nissan., 2013). The lane departure system of Iteris (2013) warns drivers of unintentional lane changes in areas where the lanes are marked. It detects both continuous and discontinuous lines even when the road markings are not clearly visible. Other innovations have demonstrated also the possibility to improve safety and mobility, this is the case of Mercedes Corporation, where a multipurpose device has been constructed, named DistronicPlus, which among other features has an adaptive cruise control function. This is used specifically for traffic jam situations where the system can take control of the car and maintains automatic user preselected security distances between vehicles by braking and accelerating (Daimler, 2013b).

Intelligent transport system and road safety applications are common topic in expert systems, thanks to the recent advances in information technologies, modern applications are used to enhance the vehicle positioning, prevent road accidents or in the event of an accident, mitigate the harm of the agents involved: in Bhatt, Aggarwal, Devabhaktuni, and Bhattacharya (2014) authors proposed a hybrid fusion scheme based on low cost INS systems to overcome the eventual signal loss in GPS systems. In Castro, Delgado, Medina, and Ruiz Lozano (2011) a fuzzy logic base system is presented for pedestrian accidents avoidance. Adaboost and SVM system for pedestrian detection is presented in Guo et al. (2012), and also SVM algorithm fused with laser scanner information is presented by García, García, Ponz, de la Escalera, and Armingol (2014) for pedestrian detection and tracking. In Conesa, Cavas Martínez, and Fernández Pacheco (2013) vehicles driving in opposite direction are identified by means on an agent based architecture. A different approach, but also related with traffic security is presented on Abellán, López, and De Oña (2013) where an algorithm to identify the severity of the accidents based on decision trees is presented. Finally driver drowsiness is analyzed based on computer vision algorithms and biological measurements in Jo, Lee, Park, Kim, and Kim (2014). All these applications represent important advances in the latest years in the expert system field related to the road safety and intelligent transport systems topics. In further sections, ADAS systems applications developed in the research platform IVVI 2.0 are presented, each of them provides new and novel solution on their respective fields.

Following with recent innovations over the past decade, scientists and engineers at Intelligent Systems Lab have developed intelligent systems in industry and academia to solve a wide variety of safety problems in vehicles, adapting to market needs of the burgeoning vehicle industry. These in vehicle systems have favoured vehicle manufacturers due to difficulties of integrating devices together with safety applications, which can cope with the widespread changes on the vehicle environment. So, this work presents our contribution to intelligent transportation systems (ITS) by means of devices that deal with machine vision, laser scanning, inertial measurement, GPS positioning and computer based processing technology. The on board applications are obstacle and



**Fig. 1.** IVVI 2.0 intelligent vehicle from Intelligent Systems Lab.

free space detection, vehicle and pedestrian detection and classification, and drive monitoring, among others. Then, these long term research and improvements to advance in intelligent systems are condensed and presented in following sections.

This article is organised in nine parts. Section 2 describes our experimental vehicle IVVI 2.0, emphasising the availability of in vehicle low level devices and the forthcoming ROS architecture. Section 3 presents an analysis of our motivations and ITS solutions to justify approaches that are detailed in following sections by means of new in vehicle sensor data and results. Section 4 explains the stereo vision for obstacle and free space detection that is the foundation of vehicle positioning by visual odometry of following Section 8. Section 5 explains the far infrared pedestrian detection, Section 6 explains the driver safety through facial recognition, Section 7 explains the obstacle detection and classification based on data fusion with laser scanner and computer vision, and Section 8 explains the vehicle positioning in urban environments by sensor fusion and visual odometry methodologies. The conclusions and future work of this review are presented in Section 9.

## 2. IVVI 2.0: an experimental vehicle

IVVI 2.0 vehicle is the second research platform of the Intelligent System Lab for the implementation of smart systems based on computer vision and laser techniques, with the goal of developing and testing ADAS technologies. The purpose of the IVVI is to test new algorithms under real conditions. At this moment, different sensing capabilities are being researched for road lanes, pedestrians, vehicles, traffic signs, driver monitoring and positioning. They can be taken as inputs for some ADAS like Lane Keeping System, Adaptive Cruise Control, Pedestrian Protector, Traffic Sign Recognition, Driver Monitoring Systems and Autopilot Systems. All of the designed ADAS work in urban and road environments, providing real time information to the driver.

The different sensing devices include a colour CCD camera (Fig. 2) mounted on the windscreen, for the detection of traffic signs and other vertical signs; a stereo vision system for road lane detection and classification and object detection (vehicles, pedestrians, etc.) in day driving conditions has been included; an infra red camera, mounted on the rear view mirror, for pedestrian detection in night driving conditions; a Kinect system, placed on the dashboard, for face detection and driver monitoring has been chosen. A multi layer laser is mounted on the front bumper for object detection, and finally, the overall devices are completed with a GNSS receiver and an Inertial Measurement Unit (IMU), which are integrated in a platform on the roof of the IVVI vehicle and aligned with the rear axis.

The processing system is based on a computer placed in the vehicle's boot, which is used for processing all the sensorial information in real time conditions. In order to provide information to the driver, it is connected to the loudspeakers of the vehicle and different audio warning messages and alerts are broadcasted. Moreover, a display in the driver's dashboard shows visual warnings and danger alerts.

In vehicle device data are available directly from sensors to our applications, or low level data can be also acquired through Robot Operating System (ROS) (Quigley et al., 2009). The advantage of synchronizing low level data by means of ROS is the timestamp of data acquisition. So, we have upgraded our IVVI platform from stand alone data acquisition to synchronized data acquisition using ROS architecture. That is, ROS allows us a collection of drivers and middleware that aim to simplify the complex task of global data acquisition and sensor synchronization. Thus, applications presented on this paper have been created accordingly to ROS architecture (Fig. 3(a)). Moreover, sensor fusion processes allow

to improve the performance of each application in high level stages.

Thus in the designed architecture, all processes communicate with each other in order to refine information and knowledge, providing higher level information to improve the decision making process(i.e.: to warn safely the driver of a possible collision). ROS architecture allows inter process communication in an independent and modular form, allowing on board computer to run multiple and parallel algorithms in order to achieve both low level tasks (e.g.: sensor acquisition, and data preprocessing, as shown in Fig. 3(b)), and high level objectives such as pedestrian detection. Moreover, ROS scalability allows to adapt the whole system to changes of the in vehicle technology due to novel sensors or higher capabilities of the applications, i.e. the use of standard messages from sensors of the same type makes the whole system immune to hardware changes.

In the following sections, our driver assistance systems and their perception modules will be described in detail.

## 3. Our motivations and ITS solutions

Our motivation is to warn vehicle drivers and pedestrians, providing assistance to avoid potential hazards by means of intelligent systems. In this section, we introduce our last intelligent systems in the automobile industry that provide safety applications.

Our computer vision application for obstacle and free space detection in front of the vehicle is based on stereo vision, which is integrated in the IVVI vehicle and allows to monitor and analyze the environment in real time (by using Graphics Processing Unit (GPU)). Stereo rig device has two image sensors to capture images, which joints with a computer, it is possible to analyze these images and extract information of interest to create stereo vision intelligent system. So, this machine vision system based on computer aids drivers indicating the presence of obstacles and free spaces within front view of the vehicle by identifying visual attributes such as points of the road or the objects (Musleh, Martin, de la Escalera, & Armingol, 2012a). Moreover, this is the foundation for following Section 8, where visual odometry will be explained to accomplish vehicle positioning task in urban environments.

Related to safe daylight and nightlight pedestrian detection, the IVVI vehicle integrates another computer vision system, that is, a single far infrared camera based on a microbolometer, which can establish more than the presence of an object. This system can classify the object based upon the requirements of an intelligent application (e.g.: pedestrian detection and tracking for night driving (Olmeda, de la Escalera, & Armingol, 2011)). So, an infrared camera is attached to left rear view mirror for pedestrian detection, which can operate in low light conditions. The presented method can determine if a pedestrian is in vehicle trajectory by means of the temperature of the human body. Moreover, the pedestrian trajectory is estimated in subsequent steps to warn the driver of the vehicle and to reduce break reaction time if it is necessary to stop the vehicle.

Moreover, in vehicle computer vision's advantages can be applied to driver safety using facial recognition. Thus, driver's face recognition is determined by means of an integrated XBOX 360 Kinect (Microsoft., 2014) over the dashboard of IVVI vehicle, where the camera allows us to obtain the depth perception feature besides the colour image. This information leads to a 3D model of the face (Peláez, García, de la Escalera, & Armingol, 2013).

Sensor fusion composes following IVVI sophisticated intelligent system, which is based on laser scanner and computer vision, that is, environment perception based on Laser Imaging Detection and Ranging (LIDAR) sensor and digital camera that allows safe obstacle detection and classification (Garcia, de la Escalera, Armingol,
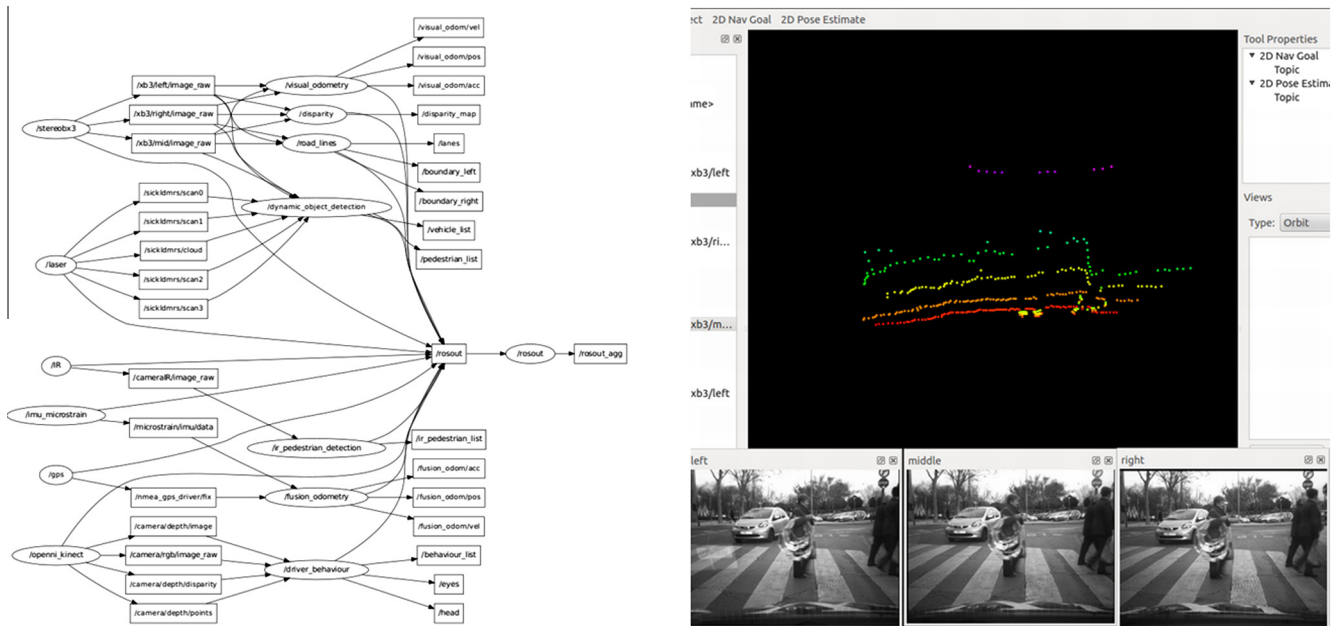
**Fig. 2.** IVVI on-board devices.



**Fig. 3.** (a) Left image, ROS diagram of global data acquisition and synchronization of low-level processes, (b) Right image, example of real-time data acquisition from in-vehicle laser scanner and stereo camera using ROS.

Herrero, and Llinas, 2011; García et al., 2013). So, in intelligent transportation systems, our sensor fusion technology can be applied to detect objects which may represent safety risks to drivers and pedestrians. This fused information provides full understanding of the environment.

The improvement of positioning in urban environments by means of visual odometry and sensor fusion is our solution for enhanced urban navigation, that is, our aim is to improve the efficiency of IVVI vehicle positioning in complex urban canyons where GNSS signals are high degraded or even loss for seconds (Martí et al., 2012). We deploy a visual odometry application that uses the movement of the vehicle to provide 2D visual ego motion estimation. This method is based on a stereo vision system where the road feature points are tracked frame to frame in order to estimate

4

the movement of the vehicle, avoiding outliers from dynamic obstacles. The second method is an advanced GNSS/IMU fusion system based on a context aided Unscented Kalman filter, where new scenarios have been analyzed to be presented in this work. The architecture of the system is designed to use contextual knowledge that reasons about sensor quality and driving context adapting to complex urban environment, besides to provide a continuous estimation and correction of INS drift errors.

Summarizing, our in vehicle applications present advantages over many other devices or technologies (e.g. advanced automotive radar). Both, vision and low cost laser scanner systems, are composed by devices which are easily installed on a vehicle. Vision systems detect and identify pedestrians, vehicles, road signs and traffic signals, enriching the available data of the vehicle environment. That is, vision systems capture a tremendous wealth of visual information beyond the range of automotive radar, allowing our computer vision algorithms cope with this wealth of visual information to identify and classify more subtle changes and distinctions between pedestrians and vehicles, enabling safe applications to vehicle manufacturers. Moreover, together with computer vision, laser scanner data, provide IVVI useful information to detect pedestrians and vehicles under a wide variety of changing environmental conditions (e.g. illumination or weather), besides GNSS positioning, which improves location in urban environments by fusing GNSS and IMU data.

Following Sections provide a detailed description of each technology.

## 4. Stereo vision for obstacle and free space detection

Research in obstacle and free space in front of the vehicle have advanced significantly in recent years. Most research, however, has focused on the development of new applications rather than optimising the algorithms of existing methodology. Thus, our research addresses these algorithms, exploring the challenges of the real world through the configuration of specific settings and smart techniques to accomplish the safe obstacle and free space detection. These robust results are later applied to develop our visual odometry application, which is explained in Section 8 as a vehicle positioning method in urban environments.

The dense disparity map and the $u$ $v$ disparity (Hu, Lamosa, & Uchimura, 2005) are developed in order to detect obstacles in front of the vehicle. The system is composed of two cameras, which are necessary to calculate the depth ($Z$) for a point $P$ $(X, Y, Z)$ in world coordinates by means of the stereo Eq. (1). The projection of the point P over the image plane is $(u_L, v_L)$ for the left camera and $(u_R, v_R)$ for the right one, $d$ is the disparity, $f$ is the focal length and $B$ is the baseline between both cameras.

$$Z \quad \frac{f \cdot B}{d} \quad \frac{f \cdot B}{u_L \quad u_R} \tag{1}$$

The disparity map is obtained by means of the rectified images that are supplied by the stereo system, where the disparity ($d$) is represented in the disparity map for every pixel of the image. So, the pixels corresponding to shortest distances between a world point and the camera have a bigger grey level and the largest ones, an inferior grey level (Fig. 4). Following, once the disparity map has been obtained, the $u$ $v$ disparity can be created: the $v$ disparity expresses the histogram over the disparity values for every image row ($v$ coordinate), whereas the $u$ disparity is calculated for every column ($u$ coordinate). The systematic use of the disparity map and its corresponding $u$ $v$ disparity allow us to extract useful information of the world, such as the obstacles that are located perpendicularly in front of the vehicle. These obstacles appear as horizontal lines in the $u$ disparity and as vertical lines in the $v$ disparity

(Broggi, Caraffi, Fedriga, & Grisleri, 2005), in their corresponding values of disparity. So, this method for detecting obstacles has, as a result, two different dense disparity maps: the obstacle map (Fig. 4(c)) and the free map (Fig. 4(d)). The obstacle map is a disparity map where only the obstacles appear, whereas the free map is simply the opposite of the obstacle map, where only the empty space ahead of the vehicle appears. This useful information can be utilised by an intelligent system for detection and localization of obstacles, as presented in Musleh et al. (2011).

Therefore, the obstacle map is one of the results of the obstacle detection step. The obstacle map is composed of two stages: firstly, the $u$ disparity is thresholded, using as a threshold the minimum height of an obstacle measured in pixels. The requirement for the selection of the threshold value is that it must be small enough to detect every obstacle, which blocks the movement of the vehicle, and large enough so as not to detect possible obstacles which do not avoid the movement of the vehicle, for example a speed bump. Thus, secondly, once the u disparity has been thresholded, every pixel of the disparity map that does not belong to an obstacle is removed, getting the obstacle map. The second result of the obstacle detection is the free map. The process to obtain the free map follows the same method that the first stage of the obstacle map, however, in the second stage, every pixel belonging to an obstacle is removed. The free map corresponds usually to the road, although it is necessary to highlight that the free map is the whole empty space ahead of the vehicle. So, in addition to the information of the obstacles, it is possible to obtain information about the road from the $v$ disparity, that is, the road appears in the $v$ disparity as an oblique line called the road profile (Labayrade, Aubert, & Tarel, 2002). Then, if a flat ground assumption is performed, this road profile can be expressed as a straight line (Eq. (2)), where $v$ is the vertical coordinate of the image, $m$ is the slope of the road profile and $b$ is the theoretical value of the horizon of the stereo system.

$$v \quad m \cdot d + b \tag{2}$$

A difficult scenario to apply this method is in urban environments, where the estimation the road profile $v$ disparity by means of the disparity map is a difficult task, that is, the detection of the oblique line corresponding to the road profile is difficult. The analysis of this complex task is presented in Musleh et al. (2012b), where the conclusion is that it is better to use the free map in order to generate the $v$ disparity, because the obstacles are removed from the $v$ disparity, and then it is easier to estimate the road profile.

Finally, the information provided by the road profile can be used for many different issues. For example, Eqs. (1) and (2) allows to obtain the position of the obstacles regarding to the vehicle, and it is also possible to classify the obstacles into two different types: obstacles that are on the ground and obstacles that are elevated. However, the main issue is visual odometry for vehicle positioning that is explained in Section 8.

## 5. Far infrared pedestrian detection

Nowadays IR vision technology is ripe for ITS applications due to properties such as, flexibility in daytime or nighttime obstacle detection, or the ability to support a variety of algorithms. So, IR vision provides drivers with additional information which is relevant to the driving task, where the IR information available, especially at night, is richer than what may be obtained by the driver himself.

In vehicle IR sensors can cover a wider area than that available from the driver's perception, and these sensors are not limited to the visible light spectrum, unlike in the case of a human driver. This section explains our research in classification of pedestrians
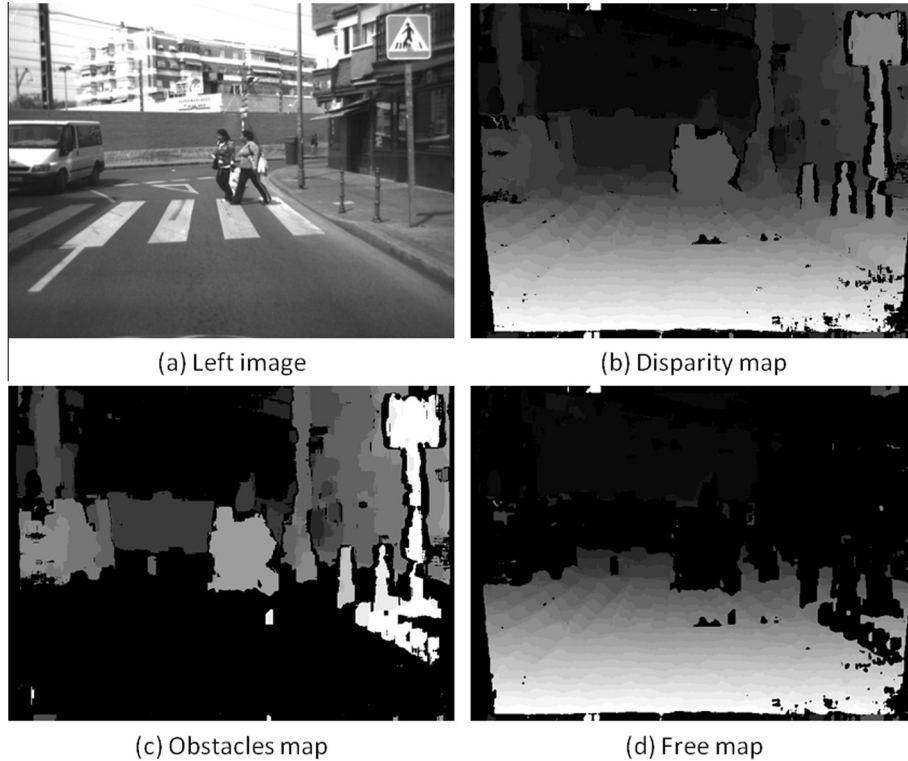
(a) Left image      (b) Disparity map

(c) Obstacles map      (d) Free map

**Fig. 4.** Example of the obstacles and road detection.

that uses information acquired by a camera sensitive to the thermal infrared or far infrared range (FIR). The recognition algorithm has been integrated as an ADAS module for intelligent transportation system applications. The aim is to detect pedestrians in low light conditions, but the method presented here is valid for any luminosity and temperature.

Illumination conditions have a major influence on the number of traffic accidents, this is particularly significant in the case of accidents involving pedestrians. With less light it takes longer for a driver to perceive a pedestrian on the road. Another important factor is exhaustion on the part of the driver. In this case the reaction time to a stimulus is much higher and chances of causing an accident grow. This type of accident is more common in conditions of limited visibility, even though there are fewer pedestrians and vehicles than during the day. Low cost cameras sensitive to far infrared spectrum are often based on microbolometer sensors. The images obtained by this kind of device represent the heat emission of the objects; thus, it makes them useful in low light or even in total absence of illumination. The relation between the measured and the actual temperature of an object is nonlinear and depends on the specific sensor being used. As such, temperature based detection would require periodic radiometric calibration of the sensor. Moreover, temperature based object detection in uncontrolled environments, such as in ITS applications, is not robust because the temperature of the object depends heavily on the ambient temperature. For example, in the case of pedestrians, the temperature of the skin and the kind of clothing is significantly different depending on the season.

In Olmeda, Premebida, Nunes, Armingol, and de la Escalera (2013) present an in depth analysis of several well known VL pedestrian classifiers applied to FIR images:

(a) LBP: Local Binary Patterns (LBP), as introduced in Ojala, Pietikainen, and Maenpaa (2002), represent the image as a similarity vector of each pixel with their surroundings. This descriptor encodes information as a binary number. For each pixel, the neighbours with a gray value higher or equal contribute with one in their position in the binary number, otherwise with zero. Each sample is divided in $3 \times 3$ pixel non overlapping cells.

(b) HOG: In this work we have used $5 \times 5$ pixel non overlapping cells. In our $64 \times 32$ dataset this means 104 cells per image, once removed the border blocks. Within each cell a 9 bins histogram of orientation between 0 and $2\pi$ radians is calculated.

(c) HOPE: Pedestrian detection algorithms usually encode the appearance of a person based on the normalised gradient (Dalal & Triggs, 2005). In the case of FIR images, this approach has some drawbacks. First, the gray level scale of a scene depends on the temperature of the sensor. Furthermore, the emission spectrum to which this type of sensor is sensitive is much wider than in the case of cameras sensitive to visible light. To overcome these problems, the authors propose a contrast invariant descriptor for object detection in far infrared images based on its phase congruency. The phase congruency features of an image are proportional to its local symmetry (Kovesi, 1999). As such, the resulting edges are not biased by the difference in temperature between the objects and the background. Because of it the resultant magnitude is also invariant to the sensor temperature. To overcome these challenges, we proposed in Olmeda, de la Escalera, and Armingol (2012) a contrast invariant descriptor for pedestrian classification in FIR images called HOPE. Basically, the HOPE descriptor encodes a grid of local oriented histograms extracted from the phase congruency of the images, which is computed from a joint of Gabor filters.

The histograms are calculates in $5 \times 5$ pixel non overlapping cells with 9 bins of orientation between 0 and $2\pi$ radians, for a total of $10 \times 4$ cells per image. No normalisation step is

applied. Phase congruency was calculated out of a set of 30 complex Gabor filters, divided in 5 scales ranging between a minimum wavelength of 2 pixels and a maximum of 10, and 6 orientations, ranging from 0 to $2\pi$ radians. Two different kernels were used for benchmarking: a linear classifier, hereafter called SVM Lin, and a radial basis function (RBF) kernel (Adeli & Karim, 2000; Karim & Adeli, 2002) designated by SVM Rbf. In this implementation the radial Gaussian function kernel $K(x,y) \quad e^{-\gamma\|x-y\|^2}$ has a scale parameter $\gamma = 1$. Both linear and RBF kernel have a regularisation trade off parameter $C = 0.05$.

(d) PCA: We treat PCA (Ghosh Dastidar, Adeli, & Dadmehr, 2008; Martinez & Kak, 2001) eigenvectors as a grey level feature vector. The initial motivation for applying this approach is that PCA tends to disregard small details at high frequency, while FIR images usually have poor levels of detail, as they present softness due to motion blur, especially at low resolutions. We retain the 30 most significant eigen vectors, that is, those with the largest eigenvalues.

(e) Feature Concatenation: Descriptor fusion is explored as feature vector concatenation, resulting in a new higher dimension feature vector with different kinds of complementary information, which can improve the overall performance.

The authors conclude that histogram based features perform best than LBP or PCA features. Among the features, HOPE performs better both for classification and detection problems. LBP and PCA features get worse overall performance, though some gains can be achieved by merging them with HOG or HOPE.

## 6. Driver safety through facial recognition

A safety focused ITS application, which depends upon computer vision technology, is facial recognition for driver protection. Our warning system is composed by the XBOX 360 Kinect from Microsoft, which takes the depth information into account, and so it is possible to fuse colour images and 3D information at the same time. Thus, the IVVI vehicle includes a drowsiness or distraction warning system based on the study of the facial features of the driver to avoid a possible vehicle accident.

Recently, Kinect applications are increasing, as is demonstrated in Frati and Prattichizzo (2011), where the tracking of a hand is used for applications oriented to virtual reality. In Santos, Lamounier, and Cardoso (2011), authors provide a solution which uses gestures as a way to interact with virtual objects in an augmented reality application. The detection of a human presence is achieved in Xia, Chen, and Aggarwal (2011), where the depth information and 2D information is associated to the contour of the head. Moreover, the Kinect device can also be used as a complementary sensor in a complex system (Ganganath & Leung, 2012), where a mobile robot uses a Kinect device for the location of land marks to correct the robot's position. Despite Kinect is not time of flight sensor, it is used for 3D environment mapping, that is, there are others and different applications that have been developed, such as Soutschek, Penne, Hornegger, and Kornhuber (2008), so a touch less user interface is developed for certain medical applications where sterility requirements in the operation's room prevent direct contact. Another application is found in Keller, Orthmann, Kolb, and Peters (2007) that explains the development of a simulation of camera like time of flight sensors. In Garcia et al. (2011), the time of fly and data fusion features are used in order to improve the road safety by detecting pedestrians (see Fig. 5).

Our application obtains a single structure through the Point Cloud Library (PCL), which contains information about the video and depth perceived. The RGB information is extracted from this structure to build a $640 \times 480$ picture. The cloud obtained from the sensor (Fig. 6), is a point structure representing Euclidean XYZ coordinates, according to a reference system where the centre of the cloud corresponds to "0" in each of the 3 axis and the RGB colour. A cloud with the Euclidean coordinates and colour information for each point is used in our application (XYZRGB cloud).

After the cloud is obtained, the colour image is built from the cloud by unpacking the RGB values. That is, from the cloud to a matrix with the same dimensions as the cloud and 3 colour channels. This matrix is converted to an image to apply computer vision algorithms; moreover, a direct association between a detected object in the colour image and its corresponding set of points in the cloud is done without additional conversions.

Then, the image is obtained and the following step is to search the facial features into the image. The process to face this step is a series of searches with different Region of Interest (ROI). That is, the image is subjected to a search of a larger object, and once detected, a more specific search is done in the area corresponding to the previous large object detected. So, the algorithm doesn't go straight for the facial features, but instead goes for a constant reduction of the searching space followed by more specific searches. The first object to be searched into the image is the upper body, following the face, and finally the eyes and mouth.

The detection is performed by a haar like features method, which allows the search of a part of the human body in the whole area of the image. This detection algorithm starts with a window of configured size, which is moved over the input image, and for each subsection of the image, the object specified in the configuration file is searched. In case of no object is found into the image, the area of the window is increased automatically by a percentage specified in the configured parameters, and following the search is repeated. So, to find an object of an unknown size in the image, the scan procedure is done several times at different scales. Our algorithm focus in the upper body, the face, the eyes and the mouth, where the ROI of the resulting object has different properties, such as: width ($w'$), height ($h'$), and the coordinates of the upper left corner. This ROI is a sub image and therefore has the 3 colour channels ($c'$).

Once the facial features are obtained, a new cloud is constructed with those points that belong to the face including depth information. Our algorithm allows to obtain more than one face and its respective point cloud, however only one is obtained to supervise supervising the driver's gestures and gaze direction.

The system has been tested with different lighting conditions and positions of the face, where if the face is detected in the RGB image then a point cloud from the face is obtained. Moreover, we



**Fig. 5.** Pedestrian detection in an infrared image in sunny conditions.
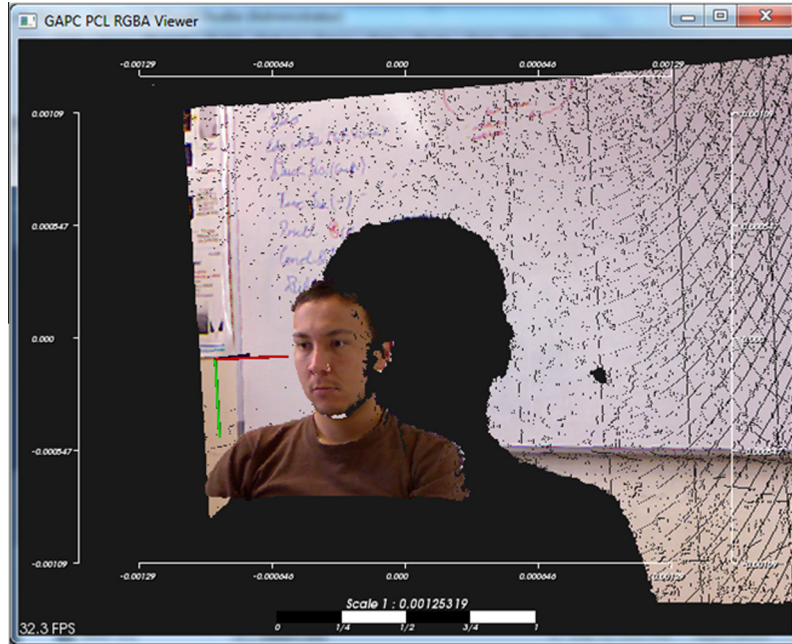
**Fig. 6.** *XYZRGB* cloud with the coordinates system.

wish to remark that the computation time of the algorithm depends on the configured parameters of the face detection library. The learned lessons are the following: Fig. 7

(i) Kinect device is designed originally for analyzing objects like human body size, however we have demonstrated that Kinect presents suitable results as the facial 3D model provides some distinguishable features such as the nose and the ocular cavities as can be observed in Fig. 7.

(ii) The second lesson is the possibility of the false positive reduction related to the detected faces inside the RGB image. That is, a 2D image mask with the depth image delivers a reduced search space for the face, and therefore, the false candidates that have geometrical and chromatic similitude with a face, are removed.

(iii) Our algorithm has been accomplish in nearly real time, with a frequency of 10 Hz, to obtain the coloured cloud of the face and the detection of the facial features such as eyes and mouth.

(iv) The obtained cloud has the same information as a 2D image, but it also includes the depth perceived by the sensor, and therefore, it offers a new method for detecting and monitoring facial gestures. That is, driver's assistance applications, where the face is analyzed constantly, are benefited from the additional information of the perceived depth.

The proposed algorithm was able to provide the analysis in 2D from the image and to associate this information with the corresponding 3D point from the depth image. Furthermore, the system is able to identify relevant parts of the human face, such as nose or eyes. It provides accurate results with relatively low processing time and based on a extremely low cost acquisition system: Kinect. This technology was designed and tested based on the low cost sensor, although it could be adapted to any 3D sensor device, such as stereo camera, or time of flight cameras presented before. The high performance and the 3D information based with low cost nature of the sensor represented an important advance in relation to other similar works, such as those presented by Jo et al. (2014) or Flores, Armingol, and Escalera (2009).

## 7. Obstacle detection and classification based on data fusion with laser scanner and computer vision

Laser devices are common in ITS applications due to their reliability. Laser scanner provide accurate and trustable range infor
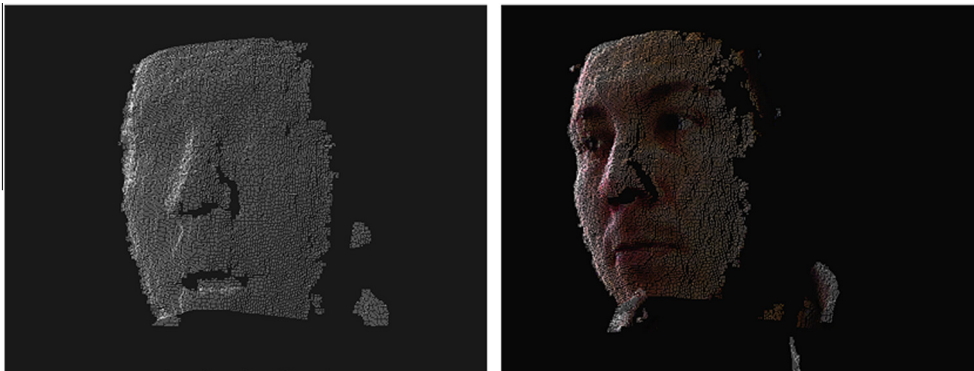


**Fig. 7.** Point Clouds of the face, *XYZ* and *XYZRGB*.

mation of the surroundings. This information can be used alone or can be fused with other sensors and vehicle data to enhance the understanding of the environment. However, the main disadvantage of the laser scanner is the small amount of information provided. By fusion of the information with other sensors e.g.: computer vision algorithms, it is possible to overcome the limitations of each one. Fig. 8 depicts the data fusion architecture used.

Following, the data fusion architecture is described accordingly. The first stage is related to the laser scanner detection and later the computer vision and tracking algorithms are detailed:

## 7.1. Laser scanner detection and obstacle classification

The laser scanner is mounted on the bumper of the vehicle. Thus, the displacement of the vehicle causes variations and displacement in the estimation of the shape of the obstacles found. In order to avoid the misinterpretation of the data, the displacement of the vehicle has to be compensated. Euler angles, obtained by the inertial system mounted on the vehicle, are used to correct the displacement of the measurements, compensating the movement of the vehicle. Eq. (3) depicts the rotation and translation to correct this displacement, this way the points are referenced to the position of the last point received (Fig. 9(a)).

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

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

(3)

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

where $\Delta\delta$, $\Delta\varphi$ and $\Delta\theta$ are the increment of the Euler angles, roll, pitch and yaw respectively, for a given period of time $T_i$, and 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 vehicle; $T_i$ the time between the given point and the first one in a given scan, and finally $(x_t, y_t, z_t)$ is the distance from the laser scanner coordinate system to the inertial measurement system.

The clouds of points are clustered using Euclidean distance and a threshold that is distance dependant (Eq. (4)), where $th_0$ is the threshold base and $K$ is a proportional constant which is multiplied by the distance.

$$th \quad th_0 + K \cdot dist \tag{4}$$

So, each point $p(x_i, y_i)$ must be treated as belonging to a segment $S_j$ if it satisfies:

$$p_i(x_i, y_i)\epsilon S_j \rightarrow \{\exists [p_j(x_j, y_j)\epsilon S_j] : d(p_j, p_i) < th\} \tag{5}$$

The algorithm verifies the points of the segments and the points among them, so if a point is not included within any segment then a new segment is created. Finally the algorithm searches for segments containing only one point, which are removed as they are considered as false detection points.

## 7.2. Polyline creation process

Polyline creation process is explained on García et al. (2012) and based on the work presented on Broggi, Cerri, Ghidoni, Grisleri, and Jung (2008). The process consists on connecting the points within a segment using straight lines, stating by merging first and last point of a segment. If a given point is too close to an existing line, no line is created, merging this point with the existing line (Fig. 9(b) shows the polylines created from the points given by the laser scanner detections depicted in Fig. 9(a)). Some
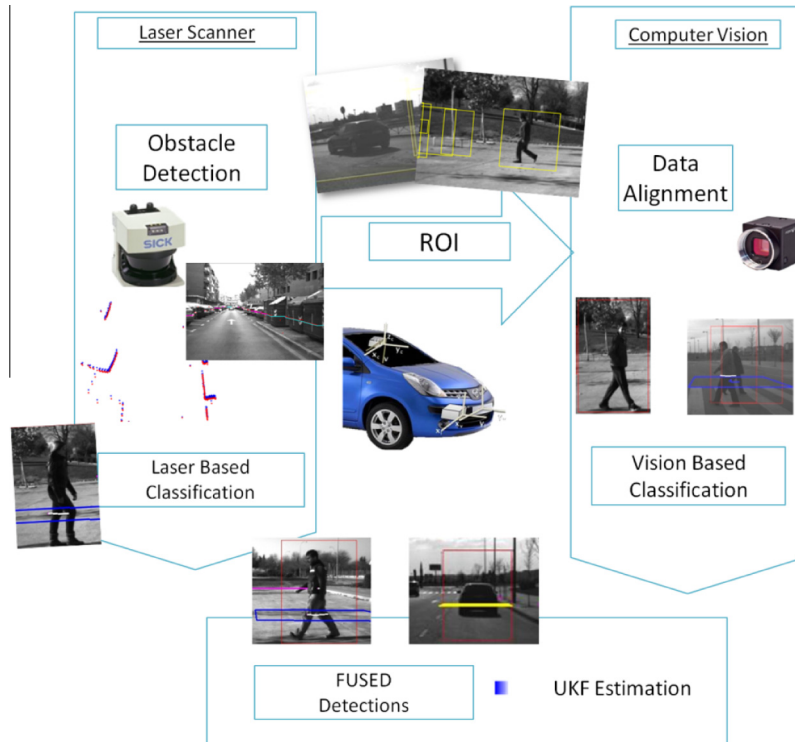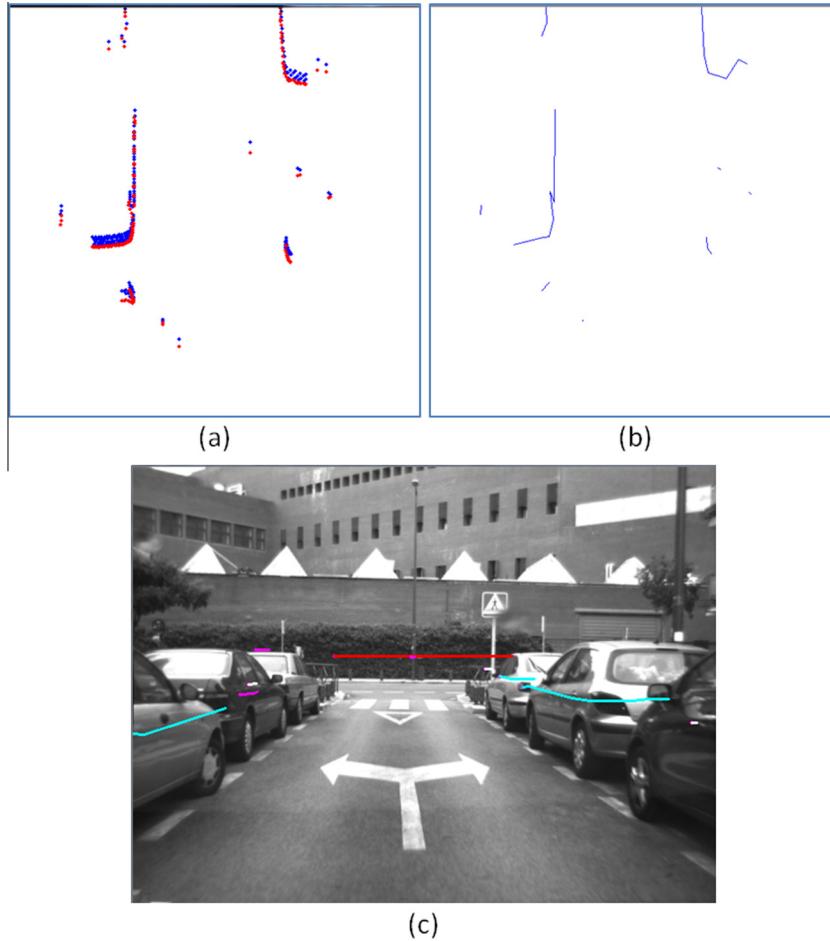


**Fig. 8.** Data fusion architecture for road understanding.

9

**Fig. 9.** Vehicle displacement compensation and data alignment: (a) the detection points, in blue colour before the vehicle displacement compensation, and in red, the compensated points, (b) the reconstructed shape is shown after the compensation, and (c) the alignment of the laser scanner data and the visual image is displayed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hints of the laser scanner obstacle estimation algorithm is provided below, although further explanation is found on García et al. (2012) for vehicles and on García et al. (2013) for pedestrians.

### 7.3. Laser scanner classification

Low level classification uses the information of the previous stage, where the obstacles can be differentiated as:

(1) Small obstacles: this type of obstacles are considered whenever the size is compatible with that of a small obstacle. So, these type of obstacles can belong to pedestrians, lampposts, milestones, trees, traffic sings, and other small obstacles that can be found in road environments. Moreover, in the case of pedestrians, a specific algorithm is applied later to determine if the obstacle is truthfully a pedestrian. It is done by comparing the resulting polyline with a pattern. We address this specific case subsequently.

(2) Road limits: this case is composed of two different possibilities according to the following procedures: first big obstacles located parallel to the road (parallel to the movement of the vehicle) and second small obstacles parallel to the road, this detection is based on histogram detection. So, if the frequency of small obstacles is sufficient it may be considered that the obstacles found on the road borders can be considered as road limits.

(3) Vehicles: the pattern is based on the fast movement of the vehicles and the delayed detection points provided by the laser scanner. The obstacles are detected using the rotation technique of the SICK LMS291 laser scanner, which performs 4 independent scans that provides 4 sets of delayed spots. After a scan, the laser scanner returns a complete set of spots separated by 0.25°. In this configuration, if a dynamic obstacle is found, the four scans sets for a single detection appear with a variation that is proportional to the velocity and direction of the detected object or vehicle. After the pattern is found, the velocity of the vehicle can be estimated by measuring the distance between two consecutive points (Fig. 10).

(4) Pedestrians: The pedestrian classification is composed of two steps. First, based on context information, obstacles with a size proportional to a pedestrian are selected among the different obstacles found, and secondly, the shape of the polyline is compared with a pedestrian pattern. Human size was estimation was based on anthropometric works (Highway Capacity Manual., 2000; Skehill, Barry, & Mcgrath, 2005) defining the human being dimension as an ellipsis which main axes are 0.6 and 0.5 m. Pattern matching algorithm was based on the study of the patterns given by pedestrians with different leg positions where the information from laser scanner is translated to the image (García et al., 2013) (Fig. 11).
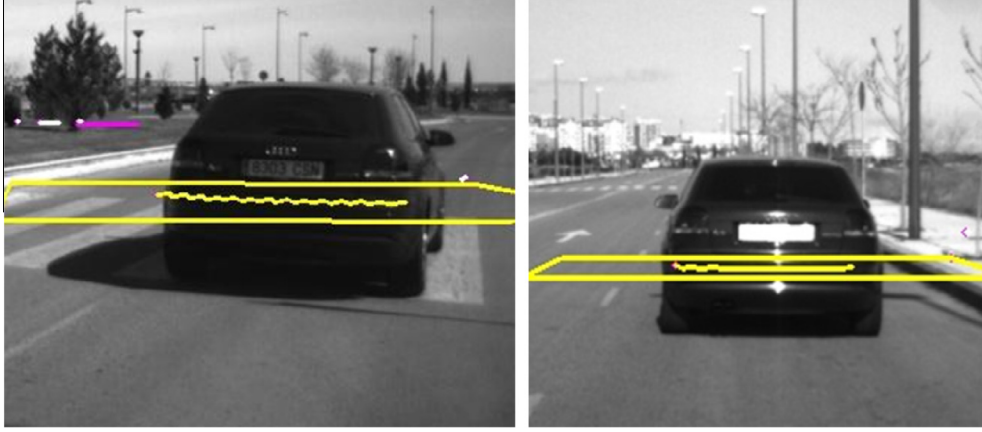
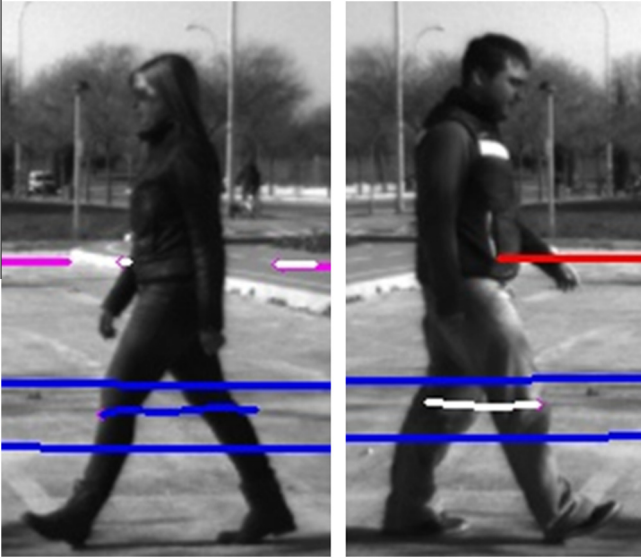Fig. 10. Laser scanner vehicle detection with the given pattern.



Fig. 11. Laser scanner pedestrian detection.

(5) Big obstacles: These obstacles are the obstacles that, due to their dimensions, could not be classified as any of the afore mentioned kind of obstacles.

Given the limitation of the laser scanner information, false pos itives were expected, based on the aforementioned pattern match ing algorithms. To reduce the rate of false positives, a high level stage was added. In this stage data is integrated into a specific per iod, by correlating obstacles in subsequent scans. So, this higher level stage cope with the behaviour of different obstacles for a spe cific time period. At this stage, a voting scheme is used to classify the obstacles, based on the low level decision in the last 10 detec tions, a multi feature approach is used for obstacle correlation along time (García et al., 2012). Finally, the obstacles being tracked (vehicles and pedestrians), are studied according to their behav iour, eliminating those performing impossible manoeuvres (eg: impossible lateral movements for vehicles or velocities)

## 7.4. Fusion with computer vision algorithms

In order to increase the accuracy of the algorithms, and taking advantage of the experience of the vision algorithms previously motioned. Fusion stage was added to the laser scanner that pro vides redundancy to the vehicle and pedestrians detections, increasing the performance of the whole system. By means of the fusion of the information from the camera and the laser scanner, the limitations inherent to each sensors are overcome.

### 7.4.1. Computer vision algorithms

The obstacles detected by the laser scanner are extrapolated to the camera field of view based on the pin hole model creating regions of interest (ROI). Later, vision algorithms are used to iden tify these obstacles in the computer vision space. The use of these ROIs helps to reduce the computational cost of the vision approach and adds reliability, thanks to the trustability of the laser scanner.

To provide accurate localization provided by the laser scanner, extrinsic calibration has to be performed. The extrinsic calibration process is based on rotational and translation equations presented on (Eq. (3)), an online calibration process was used. Later two dif ferent detection algorithms were preformed on the ROIs, for each kind of obstacle to be detected:

- Vehicle classification was performed based on Haar Like fea tures (Viola & Jones, 2001), trained for both frontal and back panel detection. These parts of the vehicles have specific char acteristics, such as lights or bumpers that make them easy to be detected with this algorithm (Fig. 12 (b)).
- Pedestrian detection is based on Histogram of Oriented Gradi ents (HOG) features. This approach, proposed on Dalal and Triggs (2005), is a classical approach for human detection (Fig. 12(a)).

### 7.5. Tracking algorithm

After the pedestrians and vehicles are detected, they are tracked, based on an unimodal (Multiple Target Tracking) MTT algorithm: Global Nearest Neighbours (GNN) and Unscented Kal man Filter (UKF). The movement model was defined using the con stant velocity target model shown in Eqs. (6) (10).

$$\widehat{X} \quad \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} \tag{6}$$

$$H \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{7}$$
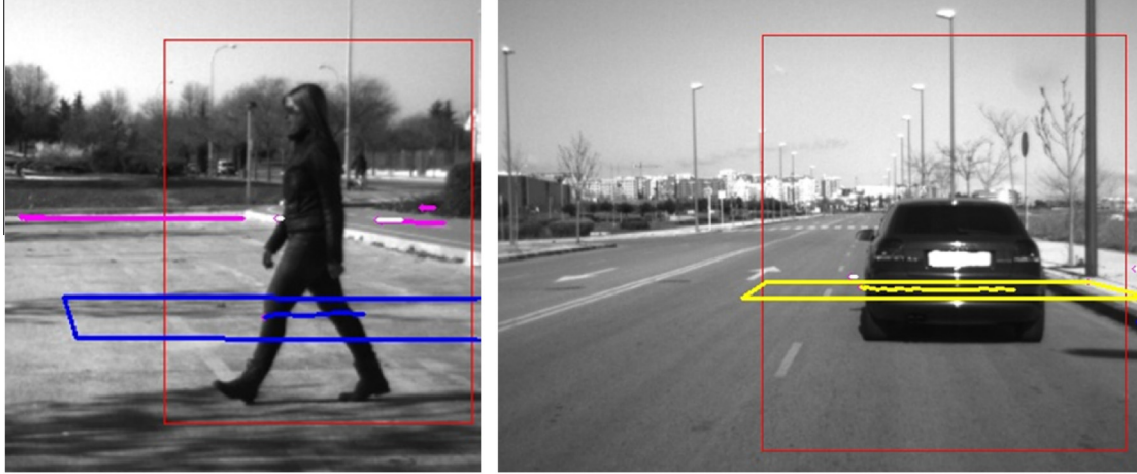
11

**Fig. 12.** Example of visual pedestrian detection (left), and vehicle (right).

$$F \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$$R \begin{pmatrix} \sigma_{\epsilon,x}^2 & 0 \\ 0 & \sigma_{\epsilon,y}^2 \end{pmatrix} \tag{9}$$

$$Q \begin{bmatrix} \frac{a_x^2 t^3}{3} & \frac{a_x^2 t^2}{2} & 0 & 0 \\ \frac{a_x^2 t^2}{2} & a_x^2 & 0 & 0 \\ 0 & 0 & \frac{a_y^2 t^3}{3} & \frac{a_y^2 t^2}{2} \\ 0 & 0 & \frac{a_y^2 t^2}{2} & a_y^2 \end{bmatrix} \tag{10}$$

where $\sigma_{\epsilon,x}^2$ y $\sigma_{\epsilon,y}^2$ are the standard deviation for the measurements in $x$, $y$ coordinates. The matrixes $\hat{X}$, $H$ and $F$ are the matrixes that define the UKF movement: State vector, observation model matrix and state transition model matrix respectively. $Q$ and $R$ are the covariance matrixes of the process noise and measurement noise of the system. Following the constant velocity model, $Q$ was selected to model the changes in the velocity of the target, the parameters defining this model was different for vehicles and pedestrians.

The GNN algorithm used on this application was based on the definition of consolidated and non consolidated tracks. First refers to those tracks with positive detections provided by both sensors. Non consolidated are tracks that were positively detected by a single sensor among the two available. The last case, is tracked, although they are considered not reliable to be reported. This way, in the case that the other sensor provides positive detection, they become consolidated tracks.

The GNN consist on the assignation of the detected track to the closest obstacle, based on the distance definition given by (11). This normalised distance uses a stability factor that gives less priority to less stable tracks:

$$d^2 \quad \frac{(x_i \quad x)^2}{\sigma_x^2} + \frac{(y_i \quad y)^2}{\sigma_y^2} + \ln (\sigma_x \sigma_y) \tag{11}$$

### 7.6. Conclusion

The presented computer vision approach was able to enhance classical computer vision algorithms, providing reliable and trustable pedestrian and vehicle detection. Although typical vision works for vehicle and pedestrian detection have successfully proved their performance in literature, the novel fusion algorithm presented on this section provides an advanced in the following lines: First by the use of laser scanner to reduce the region to search and thus speeding up the system and reducing the false positives; and second by providing original laser scanner detection algorithms which can act as backup system in case of absence of visual information. All these advances, together with the powerful tracking algorithm provides an advance respect to other vision based systems, such as those presented on Guo et al. (2012)) or García et al. (2012).

## 8. IVVI vehicle positioning in Urban Environments by visual odometry and sensor fusion

The improvement of positioning in urban environments is accomplish by visual odometry and sensor fusion. Our continuous goal is to enhance urban navigation and the efficiency of IVVI vehicle positioning in complex urban canyons where GNSS signals are high degraded or even loss for seconds, as has been studied initially in previous works (Martí et al., 2012; Musleh et al., 2012a). So, we present our long term research in advanced visual odometry and GNSS/IMU fusion system based on a context aided Unscented Kalman filter and embedded into IVVI vehicle. The visual odometry system uses the road profile (explained in Section 4) to obtain the world coordinates of the road feature points as a function of its left image coordinates, where it is only necessary to search feature points in the lower third of the left images. Otherwise, the sensor fusion architecture of the system has been designed to use contextual knowledge, which reasons about sensor quality and driving context adapting to complex urban environment.

Nowadays, vehicle positioning by Global Navigation Satellite System (GNSS) has become a wide scale device in intelligent vehicles. However, GNSS receiver presents significant variations of accuracy depending on the environment conditions and available enhancements. In urban conditions the accuracy is typically around 20 m, or even more, depending on following characteristics: (i) the number of available satellites and geometrical configuration (dilution of precision, DOP); (ii) signal propagation; and (iii) multipath conditions, a very common situation in urban conditions, whose the worst case is referred as the "urban canyon" problem (Morrison, Renaudin, Bancroft, & Lachapelle, 2012). So, our methods based on visual odometry and sensor fusion overcome specially "urban canyon" problems, when the direct path is totally

occluded and receivers only make use of signals bounced off walls of close buildings, with the corresponding degradation or even loss of any positioning solution.

Then, the use of only GNSS has been demonstrated as an insufficient positioning solution at complex urban environments. Our GNSS rover receiver calculates solutions based on two modes depending on environment characteristics: (i) single point position mode (SINGLE mode), which utilises all available GPS satellites to calculate the position solution without differential corrections, and (ii) differential mode (DGPS mode), where the base station is positioned at accurately known location and transmits differential corrections in real time to a rover receiver that is positioned in a complex urban environment. The rover receiver automatically switches between both modes, but DGPS mode has priority if appropriate corrections are received. In following tests, the rover receiver has been configured to use L1 C/A code data for differential solution due to advantages in urban environment instead of using carrier phase DGPS (Martí et al., 2012). Then, the experiments of this work have been performed with a DGPS system using L1 C/A code data for differential solution where accuracy is less than 1 m.

Therefore, following scenario delves into technical details of the GNSS receiver that causes inaccuracy of position in urban environments and other special cases selected to demonstrate the improvement of visual odometry and sensor fusion. Some cases of low accuracy caused by switching from DGPS to SINGLE mode are also presented. DGPS mode loss can be caused shortly by GPS signal unavailability or 3G network outages, and thus the DGPS recovery time is variable and depends on the technical details of the receiver. Another characteristic is the age of the corrections in DGPS mode, and its effect in the accuracy of calculated solution has been also analyzed. This characteristic predicts anomalous behaviour, since high values of differential ages cause a decrement of accuracy in DGPS mode. The accuracy in DGPS mode can be also affected by baseline length effect, atmospheric errors, satellite clock and ephemeris errors.

Generally, some researches use only the Dilution of Precision (DOP) to characterise the position accuracy, as proposed in Fakharian, Gustafsson, and Mehrfam (2011), which is a numerical representation of satellite geometry. However, lower DOP values generally represent better position accuracy, although a lower DOP value does not automatically mean a low position error (NovAtel Inc., 2012). So, to face position error, our system uses five parameters related to accuracy of position and its reliability, which are provided by OEMV 1G NovAtel receiver: the standard deviation of latitude, longitude and height errors; the age of corrections; and the number of satellites used in solution. Among these parameters, the accuracy of GPS and DGPS positions calculated by rover receiver depends mainly on the number of satellites that are used in the solution, which can be reduced caused by satellite signal obstructions (unintentional obstacles, such as buildings or trees), and the incorrect geometry of the satellites that is quantified by lower dilution of precision (DOP) parameter. There are other factors, which affect also the accuracy of GPS, such as multipath, ionospheric or tropospheric effects. The positioning error due to multipath problem cannot be reduced by the use of DGPS in urban environments, due to this type of error depends on local reflection geometry near each receiver antenna (Grewal, Weill, & Andrews, 2007).

Following experiment shows some factors that affect accuracy in urban environments and have been observed by means of the measured magnitudes: the standard deviations (latitude, longitude and height), the age of corrections and the number of satellites used in solution (Fig. 13). This trajectory shows the behaviour of the positioning system in a complex urban trajectory with variable number of satellites and DGPS correction. The accuracy is affected by complex urban environment, so through the whole trajectory it

can be observed that DGPS accuracy is reduced when only four satellites are available. This well known effect of inaccuracy can be observed in Fig. 13; where the left graph shows DGPS trajectory in Universal Transverse Mercator (UTM) geographic coordinate system. The middle graph shows latitude and longitude accuracy in metres. The right graph presents the performance of the receiver in DGPS mode, where differential ages and number of satellites are according with accuracy, however when four satellites are only used in solution, the DGPS mode loses best conditions and accuracy is reduced. Moreover, the behaviour of the receiver changes several times from differential mode to GPS SINGLE mode, causing again more reduction of accuracy. This effect can be noticed in right Fig. 13 where the value of differential age is zero.

Following the same trajectory, the exit of the urban environment is displayed in Fig. 14, where middle graph presents an increase of accuracy at 60 s of the experiment. The effect is observed again in right graph with the increase of satellites used in differential solution that leads to recovery of accuracy in differential mode.

Our first IVVI 2.0 research to face the GNSS inaccuracy in Urban Environments and enhance IVVI vehicle positioning is visual odometry. Thus, this method allows to know the 2D displacements of the vehicle in urban environments with a great accuracy, overcoming the difficulties of vehicle positioning by GNSS systems. Our visual odometry method acquires images from a stereo vision system (Fig. 15(a)), and is performed by means of detecting and tracking feature points between consecutive frames (Parra, Sotelo, Llorca, & Ocana, 2010). The algorithm uses a dense disparity map (Scharstein & Szeliski, 2002) to detect the road in front of the vehicle in order to use only the feature points that belong to the road, avoiding feature points of obstacles that can be a source of outliers if the obstacles are moving (Wangsiripitak & Murray, 2009). Moreover, another advantage of using only feature points that belong to the road, is that it is not necessary to search feature points in the whole image, as our algorithm processes only the lower third of the image. The information of the road profile (Labayrade et al., 2002) and the position of the feature points on the left image are used to obtain the world coordinates of the feature road points. Moreover, our feature points are close to the vehicle reducing the uncertain in depth estimation. So, the disparity estimation is simplified unlike the usual visual odometry algorithms, which need to perform a matching between the images of the stereo pair in order to obtain the disparity for each feature point.

Our algorithm tracks feature points between consecutive frames of the left camera that belong to the road. These road points are located on world coordinates by using the coordinates of points on the left image (Musleh et al., 2012a). An implementation of the Scale Invariant Transform Feature (SIFT) detector and descriptor (Lowe, 2004) has been used in order to detect the feature points of the images. Moreover, we have tested our algorithm with in vehicle stereo rig images that have been extracted from the visual odometry benchmark of Karlsruhe Institute of Technology (Geiger, Lenz, & Urtasun, 2012) (Fig. 15(b)).

The kinematic motion of the vehicle is modelled by the Ackerman's steering model (Borenstein, Everett, & Feng, 1996). In order to simplify the visual odometry estimation, two assumptions are applied: (i) the movement of the vehicle between two consecutive frames can be divided into two stages, whose velocity is constant in each one: a rotation around the centre of the motion of the rear axle, and a forward translation after the rotation, and (ii) there is no wheel slip in any direction.

Following the visual odometry between consecutive frames is estimated, where the vehicle motion estimation between two consecutive frames ($t$ and $t + 1$) is performed in two steps. In the first step, the road feature points have to be detected on the two left
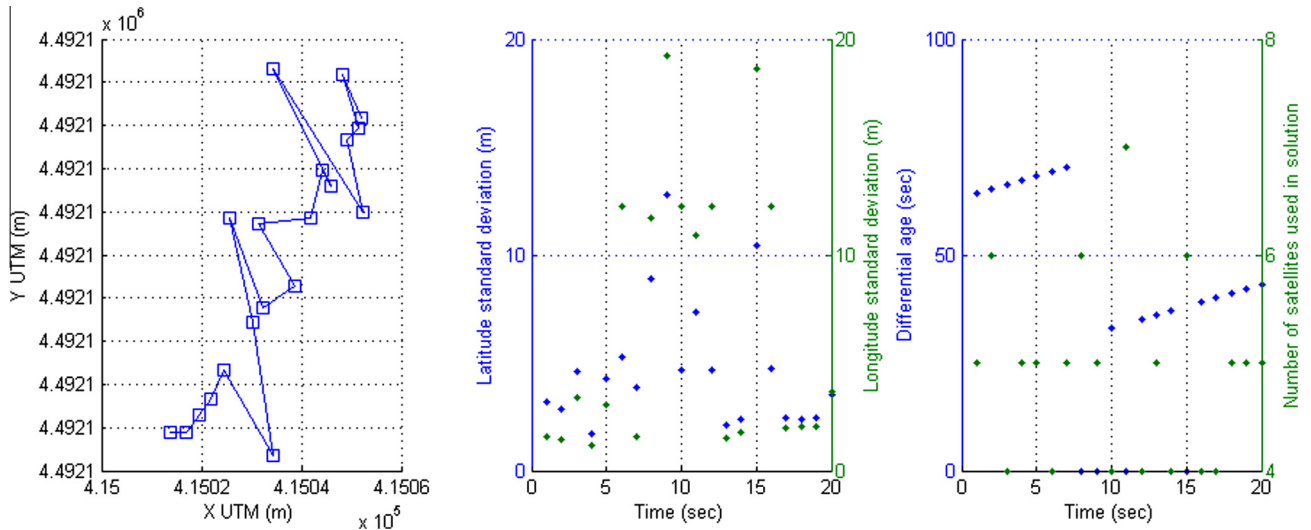
**Fig. 13.** Urban environment where differential correction is inactive in some cases and the change to GPS SINGLE mode reduces accuracy.
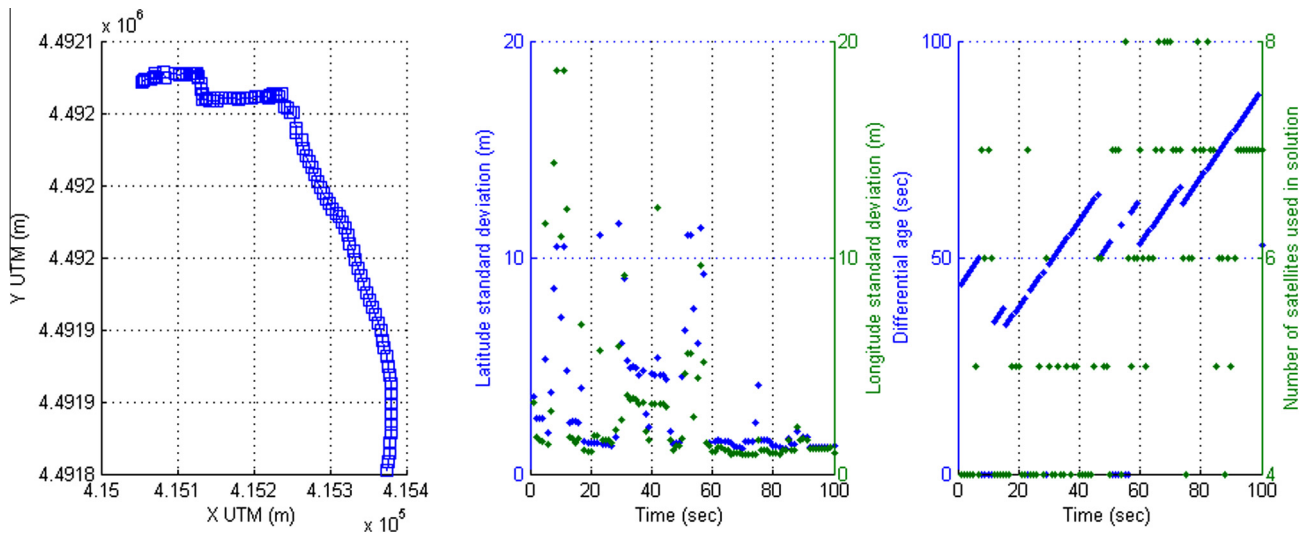


**Fig. 14.** Effect of leaving the urban environment.
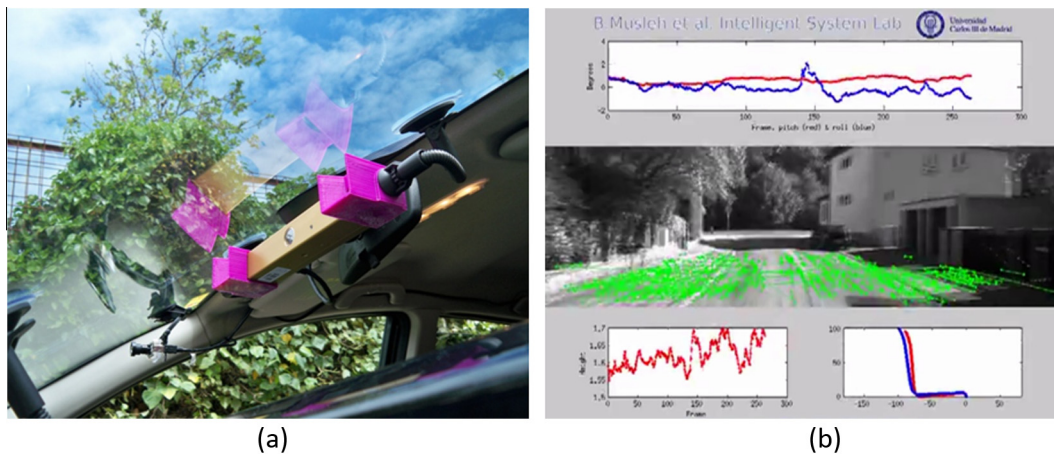


(a)

(b)

**Fig. 15.** (a) IVVI stereo rig, (b) Example of road feature points detecting and matching between consecutives frames.

images of the stereo pair at instants $t$ and $t+1$. The correspondence between the two sets of feature points ($t$ and $t+1$) is established in order to know the displacement of the road feature points within both images. Thus, our method uses only the points of the road that are located in the bottom of the image fostering the reduction of the computation time. The detection of the road feature points at the bottom of the image is accomplished by means of checking what feature points belong to the free map obtained using stereo camera. Once the image coordinates of the road feature points have been obtained, it is possible to calculate the world coordinates $\{X,Z\}$ of these road feature points in the instants $t$ and $t+1$.

Secondly, the estimation of the vehicle movement between two consecutives frames ($t$ and $t+1$) is calculated by using the different locations in the instant $t$ and $t+1$ of both road points. Thus, the rotation angle $\theta$ of the vehicle is calculated by means of Eq. (12), where $Z_T$ and $X_T$ represent the translation after the rotation. Besides, it is possible to express $Z_T$ and $X_T$ as a function of $\theta$ and the locations of a road point, in the instant $t$ and $t+1$ by using Eq. (13). So, we use Eqs. (12) and (13) to obtain (14), where angle $\theta$ is the unknown variable of the expression. So, it is possible to obtain firstly $\theta$ solving the second order Eq. (15) and then $Z_T$ and $X_T$ by using (13).

$$\theta = \arctan\left(\frac{X_T}{Z_T}\right) \Rightarrow \tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{X_T}{Z_T} \tag{12}$$

$$\begin{bmatrix} X_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} X_{t+1} \\ Z_{t+1} \end{bmatrix} + \begin{bmatrix} X_T \\ Z_T \end{bmatrix} \tag{13}$$

$$\frac{X_T}{Z_T} = \frac{\sin\theta}{\cos\theta} = \frac{X_t - X_{t+1}\cos\theta - Z_{t+1}\sin\theta}{Z_t - Z_{t+1}\cos\theta + X_{t+1}\sin\theta} \tag{14}$$

$$\left(X_t^2 + Z_t^2\right)\sin^2\theta + (2 \cdot X_{t+1} \cdot Z_t)\sin\theta + \left(X_{t+1}^2 - X_t^2\right) = 0 \tag{15}$$

The result is a set of solutions $\{\theta, Z_T, X_T\}$ to the visual odometry estimation, where a solution $\{\theta, Z_T, X_T\}_k$ has been calculated for each pair of points $\{X_t, Z_t\}_k$ and $\{X_{t+1}, Z_{t+1}\}_k$. Following, different methods can be selected in order to choose a unique solution $\{\theta, Z_T, X_T\}$ from the set of solutions, and then, the final result of the visual odome try between two consecutive frames. The simplest method to choose a unique solution from the set of solutions is the mean, however our algorithm uses the median because it is more robust to possible outliers.

Our second approach to face the GNSS inaccuracy in Urban Environments is sensor fusion architecture for enhancing IVVI vehicle positioning, which has been designed to use contextual knowledge and to reason about sensor quality and driving context adapting to complex urban environments. Thus, besides the GNSS receiver, we use an Inertial Measurement Unit (IMU). This inertial device obtains the attitude of vehicle by integrating angular rate measurements in time, and the position is computed and updated with respect to an initial solution. Then, GPS and IMU devices are complementary key technologies, and a carefully designed sensor fusion architecture can be used to provide an enhanced vehicle positioning solution.

The aim of the sensor fusion architecture is to enhance GNSS with dead reckoning capability, so that accurate vehicle navigation remains available for a certain amount of time when the GNSS sig nal data becomes unavailable or seriously degraded. However, GPS/IMU fusion is vulnerable to residual errors so a continuous monitoring of the process is necessary to guarantee that the quality of the vehicle positioning is acceptable, minimizing the effect of these factors during GPS or DGPS availability drops. The integration of both devices (DGPS + IMU) for sensor fusion that solves the unavailability of GPS positioning data in urban environments and improve its accuracy is presented in Fig. 16. The IMU axes are indi cated accordingly on image with red arrows and their rotations (roll, pitch and yaw) in blue colour.

The fusion methodology uses adaptive non linear filters (Unscented Kalman Filter, UKF) (Julier & Uhlmann, 2004), which are continuously monitored by a contextual reasoning process to provide improved performance. This architecture separates atti tude and kinematic filters to create a loosely coupled closed loop scheme that continuously estimates the IMU biases to correct them and exploit whenever the GPS data is degraded or unavail able. The aim of the architecture is to inject contextual knowledge about vehicle dynamics to adapt the model to the real conditions. Therefore, conditions such as stops, straight motion, lane changes, turns, roundabouts, are considered in the model. Besides, as it has been previously explained, it is integrated a GPS reasoning module with rules depending on conditions based on extra information, such as the availability and age of differential corrections, the num ber of satellites, the standard deviation, among others. This infor mation is applied to weight the fusion parameters or switch the bias estimation processes accordingly to the conditions (Martí et al., 2012).

The new experimental validation has been performed in com plex scenarios, such as urban canyons, to show the reliability of our second approach to face the GNSS inaccuracy. Thus, the follow ing result displays the performance of the filters when GNSS sig nals are unavailable or severely degraded in complex urban environments (Fig. 17).

This scenario is related to a complex urban environment where the vehicle is passing through urban canyons with low visibility of
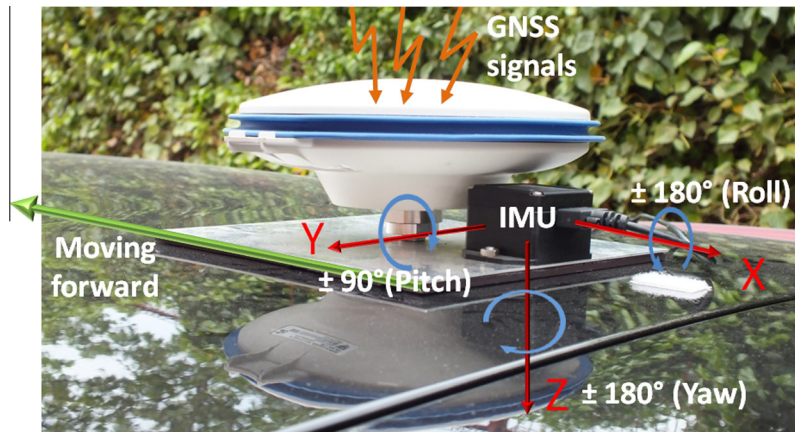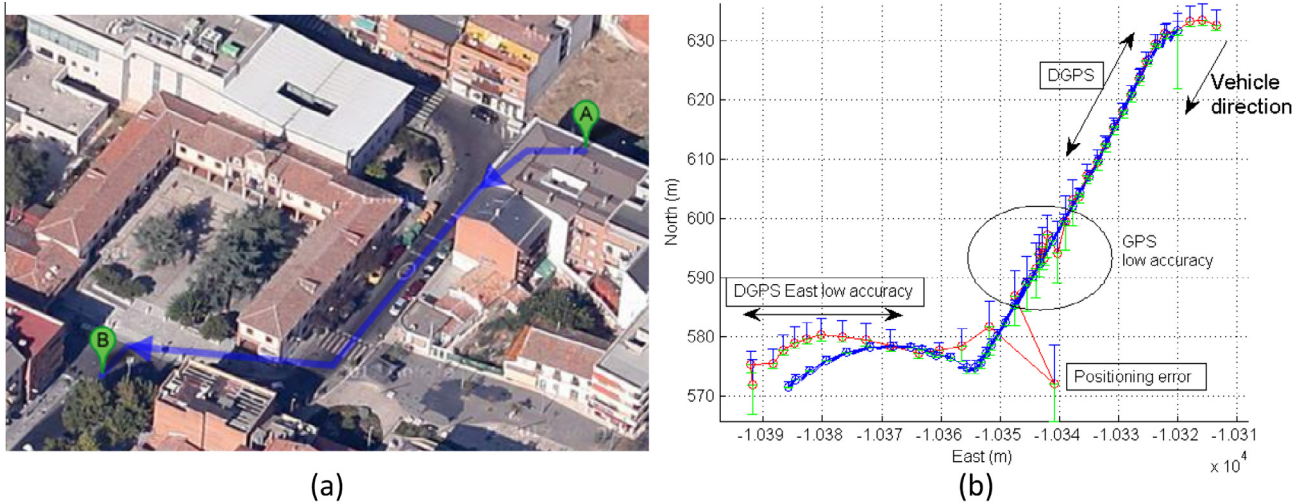


**Fig. 16.** Integration of GNSS antenna of the in-vehicle receiver and IMU in a platform over the roof of the IVVI vehicle.

**Fig. 17.** (a) Rover trajectory within complex urban canyon, (b) DGPS and GPS solutions (red circle), UKF filter solution (blue circle), and standard deviations of DGPS and GPS: East (upper blue bar) and North (lower green bar). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

satellites. The Fig. 17 shows initially vehicle positioning with active DGPS mode, cases with inactive DGPS mode and active SINGLE GPS mode solution, positioning errors due to GPS outages, and cases with DGPS East low accuracy caused by the reduction of visibility from 9 to 5 satellites for 1 s. The trajectory can be observed in the left graph (Fig. 17(a)) where the vehicle is into complex urban canyons, and the rover receiver is changing frequently their mode depending on conditions through complex urban environment. The right graph (Fig. 17(b)) displays the diversity of cases presented in this experiment, where it is difficult to obtain optimal DGPS conditions in complex urban canyons. The red trajectory (Fig. 17(b)) displays the difficult calculation of vehicle positioning by rover DGPS system caused by the inaccuracies of the complex urban area. The estimated solution using UKF filter is blue trajectory (Fig. 17(b)). Initially the DGPS trajectory is the same that the UKF filter trajectory, but positioning inaccuracy appears when the rover changes from DGPS to GPS, caused by the movement of the vehicle in a complex canyon where close buildings cause GPS and DGPS inaccuracies, and outages. The UKF solution is presented in Fig. 17(b) (blue trajectory), and shows the filter reliability with a smooth trajectory that corresponds to the real trajectory following by the vehicle, as can be observed in Fig. 17(a). Moreover, the GPS and DGPS standard deviations are presented in Fig. 17(b) to show the positioning errors that are solved by sensor fusion architecture. The last part of the trajectory displays an increase of DGPS East standard deviation caused to use five satellites in the solution for 1 s, where once more, the UKF filter solution displays again reliability to estimate the position of the vehicle.

The algorithms presented in this section represent a wide and complete set of solutions developed to overcome the classical GNSS problems algorithms in three ways: First, by the use of DGPS correction to increase the accuracy of the vehicle localization, second by combining the GNSS information with visual odometry for vehicle localization in the event of GNSS signal loss, and finally, by the fusion of the INS data with GNSS information for accuracy and localization improvement. The combination of all these technologies allows IVVI 2.0 to be able to adapt to all the different situations and configurations, representing an important advance to other works presented with a limited scope due to the use of a single solution, or with less advanced algorithms, such as those presented in Musleh et al. (2012a), Martí et al. (2012) or Bhatt et al. (2014). The novel and accurate results proved the performance of the system in the most challenging situations.

## 9. Conclusions

In this article, IVVI 2.0 Intelligent Vehicle has been presented as a whole framework to cope with cutting edge ADAS technology. Vehicle perception, data fusion and positioning has been studied as complex and essential tasks for intelligent transportation systems, where presented ADAS take advantages of heterogeneous sensors to accomplish complex tasks from safe pedestrian detection to reliable vehicle positioning.

The usefulness and advantages of the presented reliable solutions have been demonstrated through real results under demanding circumstances, such as, complex obstacles and close vehicles, cloud of pedestrians, different environment temperatures, complex driver's head poses, or GNSS outages, while IVVI vehicle maintains all applications working together reaching the aim of a safe intelligent vehicle. So, complex scenarios have been evaluated and difficulties have been be overcome successfully. The different performed tests proved both, the trustable performance of the different solutions provided and the improvements in relation to the previous works available in literature.

The different ADAS techniques mounted in the IVVI 2.0 represent cutting edge applications in intelligent transport systems and expert system research fields. The cooperation of all these applications helps to overcome classical problems of road transportation and safety: By the use of fusion techniques between laser scanner and computer vision, together with thermal camera algorithm presented, users of the road (pedestrians and vehicles) are protected, thus safety of the road, both in daylight and nightlight conditions is increased. Additionally, accurate detection and localization is achieved by means of the advanced tracking algorithms provided. Furthermore, the use of advance techniques based on data fusion for advance positioning, and the stereo camera system for visual odometry help to overcome the signal loss in extreme conditions, enhancing the data positioning systems available in the market. Finally, the advanced 3D driver monitoring system allows full understanding of the driving context, based in an extremely low cost sensor.

The strengths of the presented applications were clearly stated in the manuscript: a safer and more efficient road transportation systems based on close to market sensors such as laser scanner, computer vision cameras and advance positioning systems. However, all these applications have specific drawbacks which should be taken into account, mainly related with the nature of the sens

16

ing devices used: First, the camera based systems have the main advantage of the low cost sensor devices, with high amount of information, but with the drawback of the high sensitivity to lighting conditions (e.g. direct sun light may lead to lack of information with both the Kinect device and colour vision camera). On the other hand, laser scanners provide accurate and reliable detections with the problem of the high sensitivity to strong pitching movements. Finally, both laser scanner and inertial systems provided high accurate measurements, resulting interesting solutions for future road applications, although the actual high costs of these sensors difficult the implementation of the presented applications for commercial purposes. Nonetheless, the novel applications recently developed, such as the applications presented here, provide an important added value that brings these sensing devices closer to the final market.

All the presented algorithms give IVVI 2.0. full understanding of the driving situation, the accurate location and detection systems have the practical implication of a complete modern vehicle solution, with several close to market applications that provides an added value to the driving process, by means of a safer and more trustable driving experience. IVVI 2.0 is the second platform of the Intelligent System Lab, and represents the next generation of vehicles that are forthcoming, as an intermediate vehicle prior to the future autonomous upcoming vehicles. IVVI 2.0 was designed as a copilot vehicle, able to help and understand the driver necesities during the driving experience. Besides IVVI vehicle applications for ADAS can be applied, using presented available sensors, in forthcoming vehicles that will require reliable positioning in urban environments, automatic manoeuvres for pedestrian safety, autonomous vehicles, and collision avoidance, among other ITS applications. Furthermore, the whole deployment of an architecture based on ROS software to gather and manage the enormous data and knowledge of IVVI vehicle, will foster the improvement of all applications by fusing and integrating computer vision, laser scanner, GNSS, and digital map information, which could influence a number of ADAS applications in the forthcoming automotive industry.

Future researches, regarding to IVVI projects are already in progress, and are focused in three main lines: First refers to the improvement of the available systems, adding new sensing capabilities and developing more robust and complex algorithms that helps to overcome the actual limitations, e.g. new laser scanner device, able to provide higher amount of information, or stereo vision based on thermal camera, able to provide more accurate pedestrian localization. The second research line focus on increasing the sensing capabilities of the platform, helping to increase the environment understanding of the vehicle, these new applications refers to road understanding (e.g. road lanes and traffic signs identification) and the combinations of all this information with online context information, such as digital maps, traffic sign databases, among others. All this information, combined with the ROS based architecture, able to provide inter process communication, will enhance the information provided by the platform. All these working lines certainly converge to the next generation of intelligent vehicle i.e. autonomous vehicles, Intelligent System Lab is currently working in providing a fully autonomous vehicle, which inherits all the presented applications, and evolves them to allow safe and reliable autonomous driving.

## Acknowledgements

## 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*, 6047–6054.

Adeli, H., & Karim, A. (2000). Fuzzy-wavelet RBFNN model for freeway incident detection. *Journal of Transportation Engineering, 126*, 464–471.

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*, 2166–2173.

Borenstein, J., Everett, H., & Feng, L. (1996). Where am i? sensors and methods for mobile robot positioning. *University of Michigan*, 119 (p. 120).

Bosch (2013). Parking assistant device. <http://www.boschautoparts.com/parkassist/pages/parkassist.aspx>.

Broggi, A., Caraffi, C., Fedriga, R., & Grisleri, P. (2005). Obstacle detection with stereo vision for off-road vehicle navigation. In *IEEE computer society conference on computer vision and pattern recognition-workshops* (pp. 65).

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).

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*, 486–494.

Chang, B. R., Tsai, H. F., & Young, C.-P. (2010). Intelligent data fusion system for predicting vehicle collision warning using vision/GPS sensing. *Expert Systems with Applications, 37*, 2439–2450.

Collado, J. M., Hilario, C., de la Escalera, A., & Armingol, J. M. (2008). Multi-lane visual perception for lane departure warning system. In *Third international conference on computer vision theory and applications* (pp. 360–367).

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*, 5113–5124.

Daimler (2013a). Mercedes blind spot warning system. <http://www.daimler.com/technology-and-innovation>.

Daimler (2013b). Distronic plus. <http://www.daimler.com/technology-and-innovation>.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE computer society conference on computer vision and pattern recognition* (Vol. 1, pp. 886–893).

Fakharian, A., Gustafsson, T., & Mehrfam, M. (2011). Adaptive Kalman filtering based navigation: An IMU/GPS integration approach. In *Proceedings of the 2011 international conference on networking, sensing and control* (pp. 11–13).

Flores, M. J., Armingol, J. M., & Escalera, A. (2009). Real-time warning system for driver drowsiness detection using visual information. *Journal of Intelligent & Robotic Systems, 59*, 103–125.

Franke, U., Gavrila, D., Gern, A., Gorzig, S., Janssen, R., Paetzold, F., et al. (2001). From door to door, principles and applications of computer vision for driver assistant systems. In L. Vlacic, F. Harashima, & M. Parent (Eds.), *Intelligent vehicles technologies* (pp. 131–188). Oxford, UK: Butterworth Heinemann.

Frati, V. & Prattichizzo, D. (2011). Using kinect for hand tracking and rendering in wearable haptics. In *World Haptics Conference (WHC)* (pp. 317–321).

Ganganath, N., & Leung, H. (2012). Mobile robot localization using odometry and kinect sensor. In *International Conference on emerging signal processing applications (ESPA)* (pp. 91–97).

Garcia, 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 *Proceedings of the 14th international conference on information fusion* (pp. 1–8).

García, F., Jiménez, F., Naranjo, J. E., Zato, J. G., Aparicio, F., Armingol, J. M., et al. (2012). Environment perception based on LIDAR sensors for real road applications. *Robotica, 30*, 185–193.

García, F., Jiménez, F., Anaya, J. J., Armingol, J. M., Naranjo, J. E., & de la Escalera, A. (2013). Distributed pedestrian detection alerts based on data fusion with accurate localization. *Sensors, 13*, 11687–11708.

García, F., García, J., Ponz, A., de la Escalera, A., & Armingol, J. M. (2014). Context aided pedestrian detection for danger estimation based on laser scanner and computer vision. *Expert Systems with Applications, 41*, 6646–6661.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In *IEEE conference on computer vision and pattern recognition* (pp. 3354–3361).

Ghosh-Dastidar, S., Adeli, H., & Dadmehr, N. (2008). Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering, 55*, 512–518. Guan, A., Bayless, S. H., & Neelakantan, R. (2012). Connected vehicle insights. Trends in computer vision. Technology scan series 2011–2012. The intelligent transportation society of america (ITS America).

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*, 4274–4286.

Grewal, M. S., Weill, L. R., & Andrews, A. P. (2007). *Global positioning systems, inertial navigation, and integration* (2nd ed.). New York, NY, USA: Wiley-Interscience.

Highway Capacity Manual. (2000). *Transportation research board. National academy of sciences (ISBN 0-309-06681-6)*. United States of America: Library of congress cataloging in publication data.

Hu, Z., Lamosa, F., & Uchimura, K. (2005). A complete uv-disparity study for stereovision based 3D driving environment analysis. In *Fifth international conference on 3-d digital imaging and modeling* (pp. 204–211).

Iteris (2013). Lane departure system. <http://www.iteris.com/press.aspx?q=117>.

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*, 1139–1152.

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE, 92*, 401–422.

Karim, A., & Adeli, H. (2002). Comparison of fuzzy-wavelet radial basis function neural network freeway incident detection model with California algorithm. *Journal of Transportation Engineering, 128*, 21–30.

Keller, M., Orthmann, J., Kolb, A., & Peters, V. (2007). A simulation framework for time-of-flight sensors. In *International symposium on signals, circuits and systems* (Vol. 1, pp. 1–4).

Kovesi, P. (1999). Image features from phase congruency. *Videre: Journal of Computer Vision Research, 1*, 1–26.

Labayrade, R., Aubert, D., & Tarel, J. (2002). Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation. In *IEEE intelligent vehicle symposium* (Vol. 2, pp. 646–651).

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision, 60*, 91–110.

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, 12*, 16802–16837.

Martinez, A., & Kak, A. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*, 228–233.

Microsoft. (2014). Kinect for Xbox 360. <http://www.xbox.com/en-US/kinect>.

Milanés, V., Llorca, D. F., Villagrá, J., Pérez, J., Fernández, C., Parra, I., et al. (2012). Intelligent automatic overtaking system using vision for vehicle detection. *Expert Systems with Applications, 39*, 3362–3373.

Morrison, A., Renaudin, V., Bancroft, J. B., & Lachapelle, G. (2012). Design and testing of a multi-sensor pedestrian location and navigation platform. *Sensors, 12*, 3720–3738.

Musleh, B., de la Escalera, A., & Armingol, J. M. (2011). Real-time pedestrian recognition in urban environments. In *15th International forum on advanced microsystems for automotive applications* (pp. 139–148).

Musleh, B., Martin, D., de la Escalera, A., & Armingol, J. M. (2012a). Visual ego motion estimation in urban environments based on uv disparity. In *IEEE intelligent vehicles symposium (IV)* (pp. 444–449).

Musleh, B., de la Escalera, A., & Armingol, J. M. (2012b). U-v disparity analysis in urban environments. In *Computer aided systems theory EUROCAST 2011: Vol. 6928. LNCS* (pp. 426–432).

Nissan. (2013). Driver monitoring <http://www.nissan-global.com/EN/TECHNOLOGY/OVERVIEW/dpcc.html>.

NovAtel Inc. (2012). OEMV family of receivers—firmware reference manual. Publication Number: OM-20000094. <www.novatel.com>.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*, 971–987.

Olmeda, D., de la Escalera, A., & Armingol, J. M. (2011). Far infrared pedestrian detection and tracking for night driving. *Robotica, 29*, 495–505.

Olmeda, D., de la Escalera, A., & Armingol, J. M. (2012). Contrast invariant features for human detection in far infrared images. In *IEEE intelligent vehicles symposium (IV)* (pp. 117–122).

Olmeda, D., Premebida, C., Nunes, U., Armingol, J. M., & de la Escalera, A. (2013). A pedestrian detector using histograms of oriented gradients and a support vector machine classifier. *Integrated Computer-Aided Engineering, 20*, 347–360.

Parra, I., Sotelo, M., Llorca, D., & Ocana, M. (2010). Robust visual odometry for vehicle localization in urban environments. *Robotica, 28*, 441–452.

Peláez, G., García, F., de la Escalera, A., & Armingol, J. M. (2013). 3D based driver detection. In *Fourteenth international conference on computer aided systems theory* (pp. 10–15).

Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R., & Ng, A. (2009). ROS: An open-source robot operating system. In *ICRA workshop on open source software* (Vol. 3, p. 3.2).

Santos, E. S., Lamounier, E. A., & Cardoso, A. (2011). Interaction in augmented reality environments using kinect. In *XIII Symposium on virtual reality (SVR)* (pp. 112–121).

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision, 47*, 7–42.

Skehill, R. J., Barry, M., & Mcgrath, S. (2005). Mobility modelling with empirical pedestrian and vehicular traffic characteristics. *Wseas Transactions On Communications, 4*, 971–979.

Soquet, N., Perrollaz, R., Labayrade, R., & Auber, D. (2007). Free space estimation for autonomous navigation. In *Fifth international conference on computer vision system* (pp. 1–6).

Soutschek, S., Penne, J., Hornegger, J. & Kornhuber, J. (2008). 3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In *IEEE computer society conference on computer vision and pattern recognition workshops* (pp. 16–23).

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE comput. soc. conf. comput. vis. pattern recognit. CVPR 2001* (Vol. 1, no. C, pp. 511–518).

Wangsiripitak, S. & Murray, D. W. (2009). Avoiding moving outliers in visual SLAM by tracking moving objects. In *2009 IEEE international conference on robotics and automation* (pp. 375–380).

WHO. (2009). Global status report on road safety: Time for action. WHO library cataloguing-in-publication data, World Health Organization 2009, ISBN 978-9-241563-84-0, Geneva, Switzerland.

WHO, (2008). Speed management: a road safety manual for decision-makers and practitioners. Global Road Safety Partnership, World Health Organization 2008, ISBN 978-2-940395-04-0, Geneva, Switzerland.

Xia, L., Chen, C. C., & Aggarwal, J. K. (2011). Human detection using depth information by Kinect. In *Computer vision and pattern recognition workshops (CVPRW)* (pp. 15–22).

Zhou, Y., Xu, R., Hu, X., & Ye, Q. (2006). A robust lane detection and tracking method based on computer vision. *Measurement Science and Technology, 17*, 736–745.