Open access • Journal Article • DOI:10.1109/TITS.2011.2117420

A New Framework for Stereo Sensor Pose Through Road Segmentation and Registration — Source link

Fadi Dornaika, Jose M. Alvarez, Angel D. Sappa, Antonio M. López

Institutions: University of the Basque Country

Published on: 01 Dec 2011 - IEEE Transactions on Intelligent Transportation Systems (IEEE)

Topics: Stereo camera, Image segmentation, Image registration, Feature extraction and Orientation (computer vision)

A New Framework for Stereo Sensor Pose Through Road Segmentation and Registration

Fadi Dornaika, José M. Álvarez, *Member, IEEE*, Angel D. Sappa, *Member, IEEE*, and Antonio M. López

***Abstract*—This paper proposes a new framework for real-time estimation of the onboard stereo head's position and orientation relative to the road surface, which is required for any advanced driver-assistance application. This framework can be used with all road types: highways, urban, etc. Unlike existing works that rely on feature extraction in either the image domain or 3-D space, we propose a framework that directly estimates the unknown parameters from the stream of stereo pairs' brightness. The proposed approach consists of two stages that are invoked for every stereo frame. The first stage segments the road region in one monocular view. The second stage estimates the camera pose using a featureless registration between the segmented monocular road region and the other view in the stereo pair. This paper has two main contributions. The first contribution combines a road segmentation algorithm with a registration technique to estimate the online stereo camera pose. The second contribution solves the registration using a featureless method, which is carried out using two different optimization techniques: 1) the differential evolution algorithm and 2) the Levenberg–Marquardt (LM) algorithm. We provide experiments and evaluations of performance. The results presented show the validity of our proposed framework.**

***Index Terms*—Differential evolution algorithm, featureless image registration, illuminant-invariant image, non-linear optimization, on-board stereo camera pose, road detection, road segmentation.**

I. Introduction

SINCE THE increase in automobile park during the last decades, traffic accidents have become an important cause of fatality in modern countries. According to the World Health Organization, every year, almost 1.2 million people are killed, and 50 million are injured in traffic accidents worldwide [1]. In the last decade, research by automotive manufacturers, suppliers, and universities is addressing the development of intelligent onboard systems, which are referred to as advanced driver-assistance systems (ADASs), which aim to prevent accidents or minimize their effects when they are unavoidable.

Manuscript received September 21, 2010; revised January 11, 2011; accepted February 7, 2011. Date of publication March 14, 2011; date of current version December 5, 2011. This work was supported in part by the Spanish Government under Project TRA2007-62526/AUT, Project TRA2010-21371-C03-01, and Project TIN2010-18856, and in part by the Research Program Consolider Ingenio 2010: MIPRCV (CSD2007-00018). The Associate Editor for this paper was R. I. Hammoud.

F. Dornaika is with the University of the Basque Country and the IKERBASQUE Foundation, 20018 San Sebastian, Spain.

J. M. Álvarez, A. D. Sappa, and A. M. López are with the Computer Vision Center, Universitat Autonoma de Barcelona, 08193 Bellaterra, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2011.2117420

Vision-based systems are gaining popularity in the context of the aforementioned ADAS applications. The majority of these applications require the estimation (from images) of the onboard camera's position and orientation related to the 3-D road plane. Many researchers have addressed this problem (e.g., [2]–[4]). The proposed approaches can be broadly classified into two different categories depending on the target scenario: 1) highways and 2) urban. For each category, the vision sensor can be either a monocular camera or a stereo head. Although the objectives are the same for both scenarios, it is very challenging to develop a generic algorithm that can be used for both highways and urban scenarios. Real-time estimation of the onboard vision system pose—position and orientation—is a difficult task since the sensor undergoes motions due to the vehicle dynamics and the road imperfections, and the captured scene is unknown and continuously changing.

Since the sought 3-D plane parameters are expressed in the camera coordinate system, the camera's position and orientation are equivalent to the 3-D plane parameters. Algorithms for fast road plane estimation are very useful for driver-assistance applications and augmented reality applications. For ADAS applications, the ability to use continuously updated plane parameters (camera pose) will considerably make the tasks of obstacle and object detection more efficient [5]–[7]. For example, the number of candidate regions for pedestrian detection can be considerably reduced once the road plane is known in the camera coordinate system. For augmented reality applications, synthetic objects should be inserted in the video sequences captured by the onboard camera (e.g., virtual advertisement overlaid on the road). The continuously updated pose parameters will make the inserted objects seem as a physical part of the scene. If the used road plane parameters are constant, then the inserted objects may appear as floating objects whenever the actual plane parameters change due to the car's dynamics and road's imperfections.

Obviously, dealing with an urban scenario is more difficult than dealing with a highways scenario since prior knowledge and visual features are not always available in urban scenes [8]. In general, monocular vision systems tackle the intrinsic problems related to the 3-D aspect by using prior knowledge of the environment as an extra source of information. For instance, in [9] and [10], it is assumed that the road has a constant width; in [11], it is assumed that the car is driven along two parallel lane markings, which are projected to the left and to the right of the image; and in [12], it is assumed that the camera's position and pitch angle remain constant throughout. In [13], the authors proposed a robust computation of the homography of

the ground plane between two consecutive images from reliable ground plane point correspondences. Their proposed method takes advantage of the temporal coherence of the interframe plane-to-plane homography to construct a probabilistic prediction framework based on Kalman filtering for the computation of the homography.

Although prior knowledge has been extensively used to tackle the associated problems, it may lead to wrong results. Hence, considering a constant camera's position and orientation is not a valid assumption to be used in urban scenarios since both of them are easily affected by road imperfections or artifacts (e.g., rough road and speed bumpers), the car's acceleration, uphill/downhill driving, etc. Reference [9] estimated the vehicle's yaw, pitch, and roll by using a single-mounted camera. The method was based on the assumption that some parts of the road have a constant width (e.g., lane markings).

Some stereo vision systems have also used prior knowledge to simplify the problem and to speed up the whole processing by reducing the amount of information to be handled [14]–[16]. In the literature, many application-oriented stereo systems have been proposed. For instance, the edge-based v-disparity approach proposed in [17] for automatic estimation of horizon lines and later on used for applications such as obstacle or pedestrian detection (e.g., [2] and [18]) only computes 3-D information over local maxima of the image gradient. A sparse disparity map is computed to obtain real-time performance. In [19], the authors proposed a method for robustly detecting lanes under difficult conditions. In addition, they estimate the pitch angle using stereo-based 3-D data. The proposed method is based on the principle of maximum a posteriori, where the likelihood measurement is set to a polar histogram of 3-D points in the lateral projection plane.

It should be noticed that existing works for onboard camera pose estimation adopt a two-stage approach. In the first stage, features are extracted in either the 2-D image space (optical flow, edges, ridges, and interest points) or the 3-D Euclidean space (assuming that the 3-D data are built online). In the second stage, the unknown pose parameters are estimated using an algorithm that depends on the nature of features and on the prior knowledge used.

A. Paper Contribution

This paper has two main contributions. The first contribution is combining a nonparametric model-based road segmentation algorithm with a registration technique for estimating the online stereo camera pose. The second contribution is solving the registration using a featureless method, which is carried out using two different optimization techniques: 1) the differential evolution (DE) algorithm and 2) the Levenberg–Marquardt (LM) algorithm. Although the featureless image registration is not a novelty of this paper, we believe that the proposed framework using it for real-time tracking of onboard camera pose parameters is new. This paper is an extended version of our work in [20]. In this paper, we additionally provide an elegant method for the automatic image road segmentation and some evaluations for the 3-D camera pose parameters. Our proposed frameworks can be easily used by hybrid systems for autonomous vehicle navigation (e.g., [21]).

The proposed framework consists of two main consecutive stages that are invoked for every stereo frame: 1) road segmentation and 2) 3-D stereo camera pose through road image registration.

1) Road segmentation is an essential functionality for supporting ADASs. One of the major challenges of these techniques is dealing with lighting variations, particularly shadows. In this paper, rather than using usual segmentation in a color space, we use a physics-based illumination invariant space [22] and a statistical road pixel classification for reliable road segmentation, despite illumination variations. Using this feature space, we attenuate the shadow influence from the very beginning, even using a simple road model. The invariant space consists of a grayscale image that results from projecting the $\{\log(R/G), \log(B/G)\}$ pixel values onto the direction orthogonal to lighting change. This projection greatly attenuates the shadows, and it is computable in real time using a single-sensor color camera.
2) Once the road region is segmented in one monocular image, the current camera 3-D pose is computed using a featureless registration between this segmented region and the other view. (The left image if the road is segmented in the right image.) We solve the featureless registration by using two optimization techniques: 1) the DE algorithm (a stochastic search) and 2) the LM algorithm (a directed search). Moreover, since the camera pose should be computed for every captured frame, we propose two tracking schemes based on these optimizations. The advantage of our proposed paradigm is twofold. First, it can run in real time. Second, it provides good results, even when the road surface does not have reliable features. In addition, we stress the fact that the proposed road segmentation does not belong to the feature-based approaches since all we need is a set of pixels belonging to road image in one single monocular image.

As related works, we could mention [23] and [24]. In [23], we proposed an approach for online stereo camera pose estimation. Although [23] does not require the extraction of visual features in the images, it is based on dense depth maps and on the extraction of a dominant 3-D plane that is assumed to be the road plane. This technique has been tested on different urban environments. The proposed algorithm took, on average, 78 ms/frame. Compared with [23], our current proposed framework has the following advantages: 1) There is no need to compute a dense 3-D reconstruction of the scene viewed by the onboard camera. 2) The assumption of a very small roll angle is released. In [23], this assumption is needed to obtain efficient road plane extraction. 3) Our approach directly infers the geometric parameters from image rawbrightness without using an intermediate stage, i.e., the estimation of the 3-D coordinates of every pixel. Thus, there is no error propagation in the estimation of the camera pose parameters. 4) The use of road segmentation in one monocular image guarantees that most of available data are contributing to the solution. In [24], we proposed a framework that tracks the pose parameters using a sequential Monte Carlo filter based on a featureless criterion.

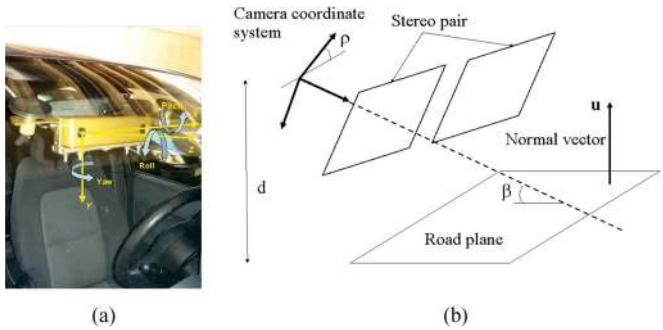

Fig. 1. (a) Onboard stereo vision sensor. (b) Time-varying road plane parameters d and $\mathbf{u}$. β denotes the pitch angle, and ρ denotes the roll angle.

Although [24] and our current proposed framework are both using featureless registration, our current proposed framework has the following advantages: 1) The accurate road region in every video frame is obtained by using a statistical road pixel classification, whereas in [24], this road region was a constant user-defined region. 2) The current proposed framework uses a deterministic tracking algorithm (directed search minimization) that is faster than the stochastic tracking algorithm presented in [24], where a probability distribution is propagated over time.

The rest of this paper is organized as follows: Section II describes the problem we are focusing on. Section III presents the used road segmentation approach and some experimental results. Section IV presents the online camera pose estimation through road image registration. Section V gives some experimental results and method comparisons. Section VI concludes this paper.

II. Problem Formulation

A. Experimental Setup

A commercial stereo vision system (Bumblebee from Point Grey[1]) was used. It consists of two Sony ICX084 color charge-coupled devices with 6-mm-focal-length lenses. Bumblebee is a precalibrated system that does not require in-field calibration. Fig. 1(a) shows an illustration of the onboard stereo vision system and its mounting device.

The problem we are focusing on can be stated as follows: Given a stream of stereo pairs provided by the onboard stereo head, we like to recover the parameters of the road plane for every captured stereo pair. Since we do not use any feature that is associated with road structure, the computed plane parameters will completely define the pose of the onboard vision sensor. This pose is represented by the 3-D plane parameters, i.e., the height d and the plane normal $\mathbf{u} = (u_x, u_y, u_z)^T$ from which two independent angles can be inferred [see Fig. 1(b)]. In the sequel, the pitch angle will refer to the angle between the camera's optical axis and the road plane, and the roll angle will refer to the angle between the camera horizontal axis and the road plane [see Fig. 1(b)]. Due to the reasons previously mentioned, these parameters are not constant and should be estimated online for every time instant. Note that the three angles (pitch, yaw, and roll) associated with the stereo head orientation can vary. However, only the pitch and roll angles can be estimated from the 3-D plane parameters.

[1][www.ptgrey.com]

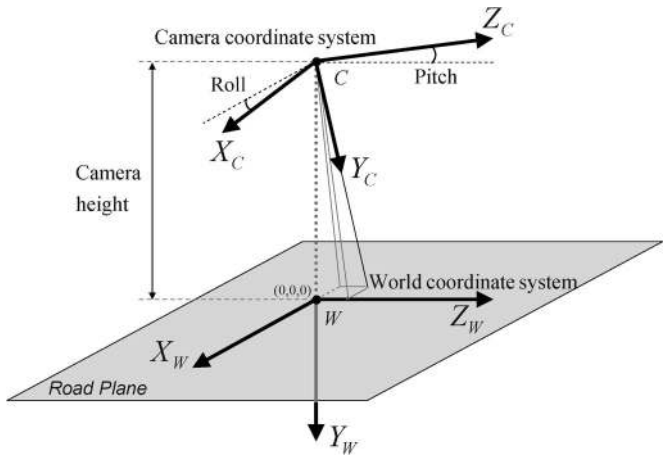

Fig. 2. Moving world coordinate system.

Since there is no fixed world coordinate system, our problem is not equivalent to the classical extrinsic camera calibration in which the six degrees of freedom of the camera pose should be estimated. However, this link is possible by adopting a moving world coordinate system $(X_W; Y_W; Z_W)$ (see Fig. 2) that was defined for every acquired stereo image in such a way that the $X_W Z_W$ plane is contained in the current road fitted plane, just under the camera coordinate system. The origin of this coordinate system is the orthogonal projection of the camera center. Thus, by adopting this world coordinate system, the extrinsic parameters reduce to just three parameters, i.e., a translational distance and two independent angles. Notice that the yaw angle is zero for all stereo frames.

B. Proposed Framework

Since the goal is to estimate the road plane parameters for every stereo pair (equivalently, the 3-D pose of the stereo head), the whole process is invoked for every stereo pair. The inputs to the process are the current stereo pair and the estimated road plane parameters associated with the previous frame. The proposed framework is shown in Fig. 3. The process is split into two consecutive stages. First, a road region segmentation is preformed for the right image (see Section III). Second, the current camera 3-D pose is computed using a featureless registration between this segmented region and the other view (see Section IV).

III. Road Segmentation

Our proposed framework for onboard camera pose estimation requires a road segmentation in one monocular view. In this section, we describe the method selected to detect the road region in the right images. Vision-based road segmentation is a very challenging problem since the road is in an outdoor scene imaged from a mobile platform [25], [26]. Thus, the segmentation algorithm should be able to deal with a continuously changing background, the presence of different objects (e.g., vehicles and pedestrians) with unknown movement, different road types (e.g., urban, highways, and off-roads), different road attributes

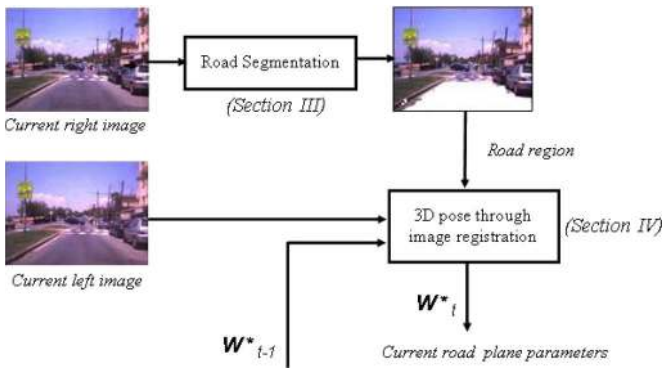

Fig. 3. Proposed framework consists of two stages. First, a road region segmentation is preformed for the right image (see Section III). Second, the current camera 3-D pose is computed using a featureless registration between this segmented region and the other view (see Section IV).

(e.g., shape and color), and different imaging conditions (varying illumination, different viewpoints, and weather conditions). Common vision-based approaches integrate different cues such as shape [27], color [27], [28], texture [29], or time coherence [28], which lead to constrained systems (i.e., shape cue limits the type of the road and scenarios, texture limits the distance ahead the vehicle because of the perspective effect, and time coherence limits the speed of the vehicle).

In this section, the color-based approach for the road detection proposed in [30] is considered. Color provides powerful information about the road to be detected, even in the absence of shape information. In addition, color imposes less physical restrictions, leading to more versatile systems. The underlying idea of the algorithm is to map the original colors into another illuminant-invariant feature space in which the road ahead has some homogeneous color, which is used to classify pixels as belonging or not to the road class. Nevertheless, the perception of the road surface depends not only on its own features, which, in fact, are not constant, but on unknown imaging conditions (shadows and highlights among others) as well. This variability is reduced by selecting the most appropriate color space to characterize the input data.

Therefore, the color-based road detection algorithm, which combines a physics-based illuminant-invariant color space with a model-based binary classifier in a frame-by-frame framework (see Fig. 4), is used [30]. More precisely, our approach exploits the lighting-invariant benefits of the color space introduced by Finlayson *et al.* [22] and a binary classifier based on a nonparametric model [30]. The choice of the color space is motivated by having sunlight as the main light source. Sunlight is considered to be Planckian, which is one of the requirements of the selected color space. Furthermore, the classifier uses the invariant color representation, together with an online built likelihood to decide whether a pixel belongs or not to the road class. This likelihood measure is given by a probability distribution that is built online for every frame using scattered region seeds in the bottom part of the image. Thus, the algorithm uses the only assumption that the bottom region of the image is road to build the road model. In fact, the lowest row of our images corresponds to a distance of about 4 m away from the camera placement, which is a reasonable assumption most of the time.

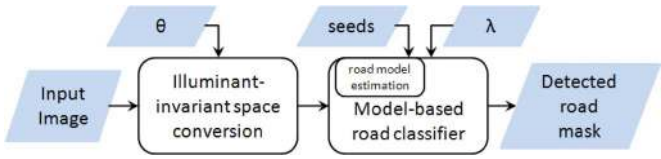

Fig. 4. Road segmentation algorithm. First, each RGB image is converted onto an illuminant-invariant image using the invariant direction θ in the 2-D $\{\log(R/G), \log(B/G)\}$ space, which is an intrinsic parameter of the camera. Each pixel is then classified as road or nonroad according to a nonparametric road model (a likelihood measure) and a fixed threshold λ.

In the remainder of this section, we first describe the illuminant-invariant feature space, then the proposed nonparametric classifier, and finally, some segmentation results are shown.

A. Illumination-Invariant Space

Finlayson *et al.* have shown that, if *Lambertian* surfaces are imaged by a three delta-function sensor under approximately Planckian light sources, it is possible to generate an illuminant-invariant image ($\mathcal{J}$) [22]. Under these assumptions, a log–log plot of 2-D $\{\log(R/G), \log(B/G)\}$ values for any surface forms a straight line, provided camera sensors are fairly narrowband. Thus, lighting change is reduced to a linear transformation along an almost straight line (see Fig. 5). In practice, empirical results prove that this theory holds, even for real-world scenes (roughly Lambertian surfaces) imaged using a regular camera (i.e., having only approximately narrowband sensors) under approximately Planckian illumination [30].

In short, $\mathcal{J}$ is a grayscale image that is obtained from projecting the $\{\log(R/G), \log(B/G)\}$ pixel values of the incoming data onto the direction orthogonal to the lighting change lines, i.e., *invariant direction* θ from now on. This direction is device dependent and is estimated offline using the calibration procedure in [22].

B. Model-Based Road Classifier

The aim of the model-based road classifier is assigning to each image pixel one of the two possible classes, i.e., road and background. Two entities are used to decide whether a pixel belongs or not to the road class. The first entity, i.e., the road model (likelihood measure), is a probabilistic description of the road that provides insight for predicting the label (road or background) of each image pixel. The second is a fixed threshold λ on this model. Hence, a road label is assigned to a pixel exhibiting a support provided by the model higher than λ. Otherwise, a background label is assigned to that pixel.

The road model is built online for each image based on a training set. The training set consists of surrounding areas of several pixels (seeds) placed at the bottom part of the image to be segmented. Thus, the algorithm considers only road pixels (positive examples) and assumes that the bottom area of the image belongs to the road surface. In this paper, nine seeds are placed using an equidistant distribution along two rows in the bottom part of each frame (see Fig. 6). The size of the surrounding region of each seed is fixed to 11×11 pixels for an image of 640×480 pixels. Then, the road model is

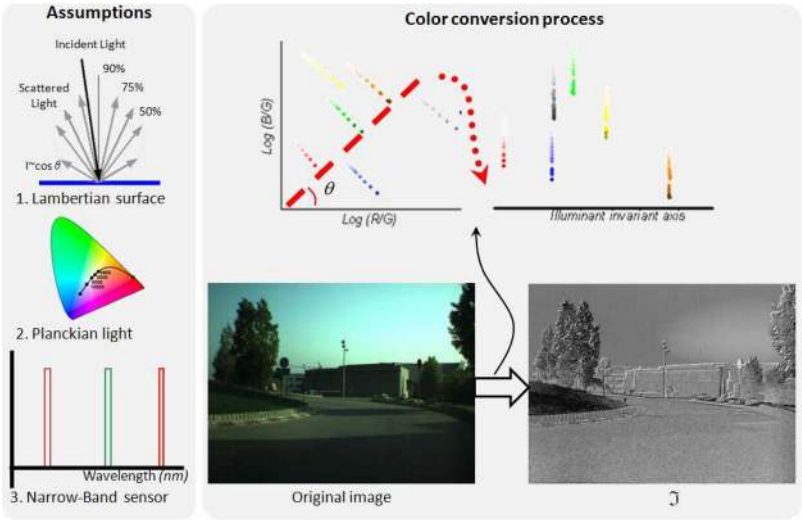

Fig. 5. Illuminant-invariant image can be obtained under the assumptions of Planckian light, Lambertian surface, and narrowband sensors. This image is almost shadow free.

Fig. 6. Road model example. The model is built using the histogram formed with the surrounding region (white blocks) of several seeds. We have used nine seeds placed at the bottom part of $\mathcal{J}$, which is the illuminant-invariant image.

computed as the normalized histogram of these pixel values. There are two important advantages in using histograms (nonparametric methods) over parametric approaches (e.g., mixture of Gaussians): nonparametric methods are faster in training and usage, and they are independent of the shape of the data distribution. Hence, the normalized histogram is used as a likelihood function, indicating the support of each possible illuminant-invariant value depicting road surface: the probability of observing $\mathcal{J}(p)$ given that the pixel p belongs to the road. As a result, the road model (the normalized histogram) is highly adaptive and can cope with sudden changes.

Finally, the algorithm for road detection is shown in Fig. 7. First, the incoming image is converted to the illuminant-invariant space. Then, the road model is applied to obtain a road confidence map [see Fig. 7(a)] depicting the probability of a pixel being road. This map is binarized using a fixed threshold λ [see Fig. 7(b)]. Then, connected components (region growing) is applied to the binary image using the same set of seeds used to build the road model [see Fig. 7(c)]. Finally, a hole-filling process using simple mathematical morphology operations is applied to obtain the final result [see Fig. 7(d)].

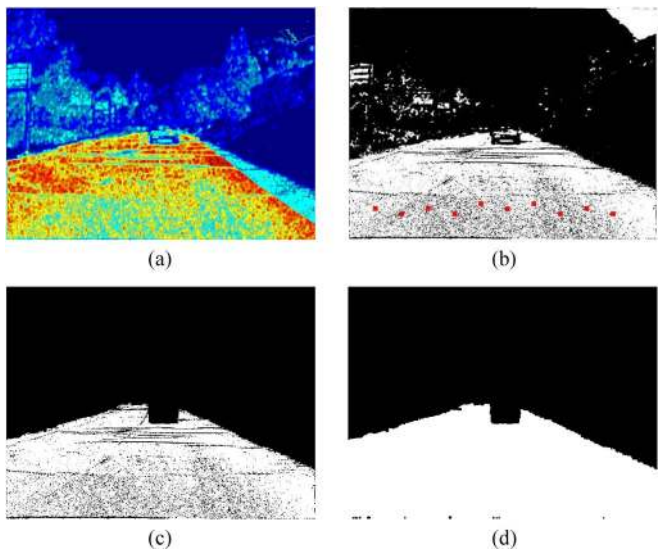

Fig. 7. (a) Road probability map. (b) Binarized image using a fixed threshold and set of seeds overlapped (red blocks). (c) Mask image after applying connected components. (d) Detected road mask obtained applying standard mathematical morphology.

C. Road Segmentation Results

In this section, we present qualitative results to validate our proposal. The algorithm has been tested using different sequences acquired using the system described in Section II. These images cover approximately the nearest 80 m ahead of the vehicle. The threshold λ of the algorithm is fixed using an offline learning approach. This approach consists of processing and evaluating a set of images using all possible values within the range of the parameter $[0, 0.05, \ldots, 1]$. The optimal threshold is that which maximizes the average performance (see Fig. 8). That is, the arithmetical mean of performance values obtained for each image in the set. The optimal threshold is then chosen for subsequent segmentation at runtime.

Fig. 9 shows the segmentation results associated with a sequence of images with extreme light variations. These images include nonhomogeneous roads due to extreme shadows and the presence of other vehicles. Fig. 10 shows the segmentation results associated with another challenging road sequence

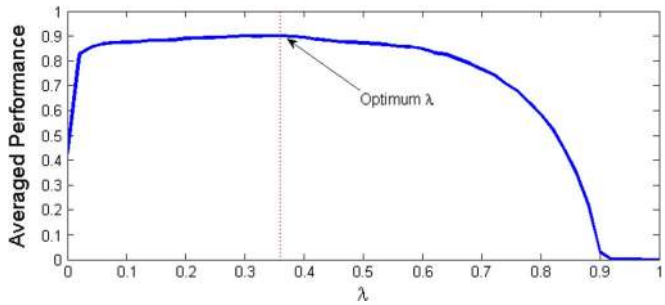

Fig. 8. Variation of the averaged performance in the process of learning the threshold λ. This process consists of processing and evaluating a set of images. The optimum λ is the one maximizing the average performance (mean of performance obtained for each image in the data set).

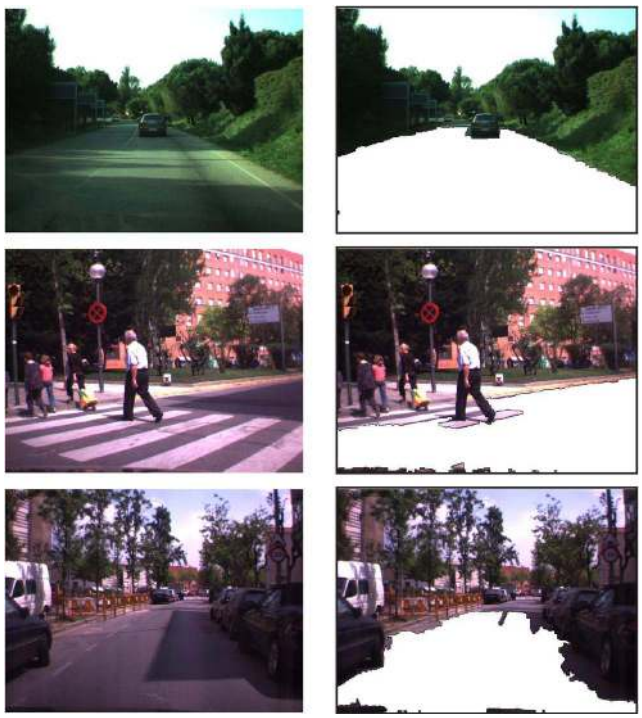

Fig. 9. Segmentation results associated with extreme light variations. The left column is the original image that covers the nearest 80 m ahead of the vehicle. The right column shows in white the corresponding segmented road.

acquired when the car is performing a turning maneuver. The challenge relies on the fact that, in some frames, the road region is not well defined and is very small compared with the background. Despite this challenge, the proposed segmentation algorithm has provided the correct segmentation. The nonroad small regions can be discarded by a simple analysis.

For evaluation purposes, the proposed method is compared with a simple approach that does not need offline learning and calibration. It works on the hue and saturation components (referred as HS-based algorithm hereinafter). HSV color space has been used for scene segmentation under varying illumination conditions [31], [32]. The used HS-based algorithm splits the segmentation stage into two phases. The first phase is only invoked every T frames for updating the color model and for obtaining a real-time performance. The second phase exploits the road color consistency over short time. The first phase consists of a classical K-means algorithm that is applied

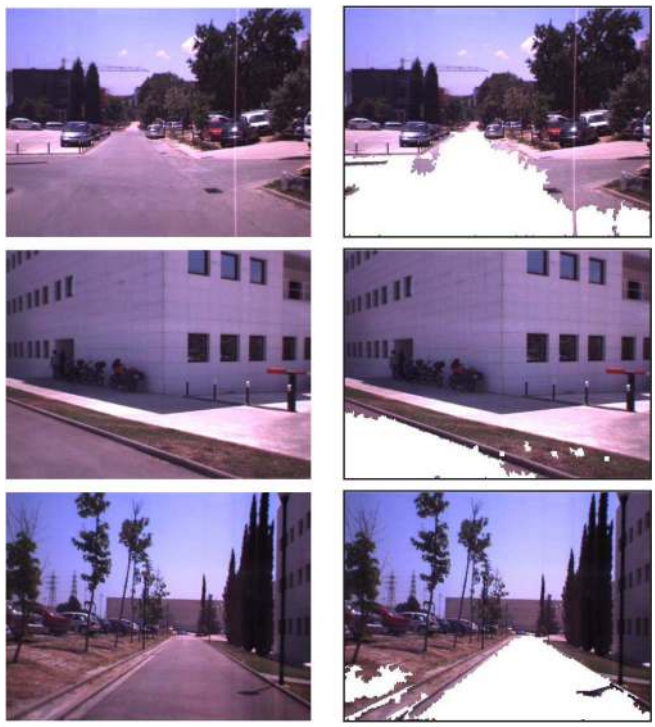

Fig. 10. Challenging road sequence. The left column shows the original image. The right column shows in white the corresponding segmented road.

on the hue H and saturation V values of the pixels belonging to a predefined region of interest (ROI) that is centered at the bottom of the image. The number of classes can be between three and five. The cluster having the largest number of pixels will be assumed to belong to the road. Once the cluster is identified, the mean and covariance of its color (hue and saturation components) can be easily computed. In the second phase (invoked for every frame), by assuming that the color distribution of the detected cluster is Gaussian, we can classify any pixel. Thus, pixels within the ROI are labeled as road pixels if their *Mahalanobis* distance to the mean is below a certain threshold.

Fig. 11 shows comparison results associated with two frames: Fig. 11(a) depicts the original frames, (b) shows the segmentation results obtained with the HS-based algorithm, and (c) shows the segmentation results obtained with the proposed method. As can be seen, both methods have succeeded to segment a large part of the road. The side part of the road has not been detected by the HS-based algorithm since the used ROI is a rectangular window and the method assumes that the road color has one mode.

All these results suggest that a reliable road segmentation algorithm is obtained by combining the illuminant-invariant image space and the online model-based classifier. Road surface is well recovered most of the time, with the segmentation stopping at road limits and vehicles, and can deal with complex road shapes. Nevertheless, the algorithm fails for areas where there is a lack of color information leading to nonvalid invariant image areas. However, this issue can be addressed by improving the acquisition system (i.e., cameras with higher dynamic range). For a detailed quantitative and failure case study, see [30].

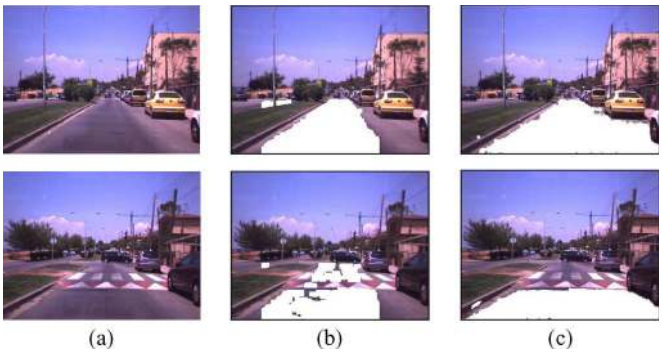

Fig. 11. Road segmentation associated with two frames. (a) Original images. (b) Segmentation results obtained with the color consistency method (HS method). (c) Segmentation results obtained with the histogram road model.

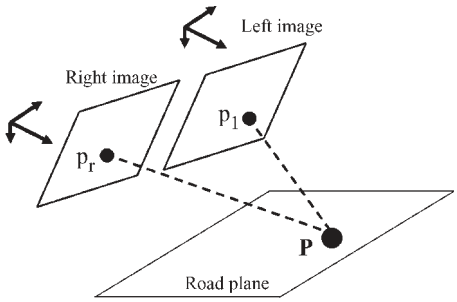

Fig. 12. Mapping between the corresponding left and right road pixels is given by a linear transform.

IV. Three-Dimensional Pose Parameters Through Image Registration

A. Right-Image-to-Left-Image Mapping

This section describes the 2-D mapping between road pixels belonging to the same stereo pair: the left and right images. This mapping implicitly depends on the camera pose parameters. It is well known [33] that the 2-D projections of 3-D points belonging to the same plane onto two different images are related by a 2-D projective transform having eight independent parameters, i.e., *homography*. In our setup, the right and left images are horizontally rectified.[2] Let $p_r(x_r, y_r)$ and $p_l(x_l, y_l)$ be the right and left projections of an arbitrary 3-D point P belonging to the road plane (d, u_x, u_y, u_z) (see Fig. 12). In the case of a rectified stereo pair where the left and right cameras have the same intrinsic parameters, the 3×3 homography matrix will have the following form:

$$\mathbf{H} = \begin{pmatrix} h_1 & h_2 & h_3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1}$$

In other words, the right and left coordinates of corresponding pixels belonging to the road plane are related by the following linear transform (the homography reduces to a linear mapping):

$$x_l = h_1 x_r + h_2 y_r + h_3 \tag{2}$$
$$y_l = y_r \tag{3}$$

[2]The use of nonrectified images will not have any theoretical impact on our developed method. However, the image transfer function will be given by a general homography.

where h_1, h_2, and h_3 are functions of the intrinsic and extrinsic parameters of the stereo head and of the plane parameters. For our setup (rectified images with the same intrinsic parameters), those coefficients are given by

$$h_1 = 1 + b\frac{u_x}{d} \tag{4}$$
$$h_2 = b\frac{u_y}{d} \tag{5}$$
$$h_3 = -bu_0\frac{u_x}{d} - bv_0\frac{u_y}{d} + \alpha b\frac{u_z}{d} \tag{6}$$

where b is the baseline of the stereo head, α is the focal length in pixels, and (u_0, v_0) is the image center (principal point). Let $\mathbf{w}$ be the three-vector encapsulating the 3-D plane parameters, i.e., $\mathbf{w} = \mathbf{u}/d$. Thus, $\mathbf{w}$ is given by

$$\mathbf{w} = (w_x, w_y, w_z)^T = \left(\frac{u_x}{d}, \frac{u_y}{d}, \frac{u_z}{d}\right)^T. \tag{7}$$

Note that the vector $\mathbf{w}$ fully describes the current road plane parameters. The problem can be stated as follows: Given the current stereo pair, estimate the corresponding 3-D road plane parameters d and $\mathbf{u}$ or, equivalently, the vector $\mathbf{w}$.

B. Approach

Let $\mathcal{R}$ denote the segmented road region in the right image (e.g., the white regions in the right column of Fig. 9). This segmentation is obtained using the approach described in the previous section. Recovering the plane parameters from the rawbrightness of a given stereo pair will rely on the following fact: *if the parameter vector* $\mathbf{w}$ *corresponds to the actual road plane parameters—the distance* d *and the normal* $\mathbf{u}$*—then the registration error between corresponding pixels in the right* and *left images over the road region* $\mathcal{R}$ *should correspond to a minimum*. In our work, the registration error is set to the sum of squared differences (SSD) between the right image and the corresponding left image computed over the road region $\mathcal{R}$. The registration error is given by

$$e(\mathbf{w}) = \sum_{(x_r, y_r) \in \mathcal{R}} \left(I_{r(x_r, y_r)} - I_{l(h_1 x_r + h_2 y_r + h_3, y_r)}\right)^2. \tag{8}$$

The corresponding left pixels are computed according to the linear transform given by (2) and (3). The computed $x_l = h_1 x_r + h_2 y_r + h_3$ is a noninteger value. Therefore, the gray level $I_l(x_l, y_l)$ is set to a linear interpolation of the gray level of two neighboring pixels, i.e., the ones whose horizontal coordinates bracket the value x_l.

The optimal current road parameters are given by

$$\begin{aligned} \mathbf{w}^\star &= \arg\min_{\mathbf{w}} e(\mathbf{w}) \\ &= \arg\min_{\mathbf{w}} \sum_{(x_r, y_r) \in \mathcal{R}} \left(I_{r(x_r, y_r)} - I_{l(h_1 x_r + h_2 y_r + h_3, y_r)}\right)^2 \end{aligned} \tag{9}$$

where $e(\mathbf{w})$ is a nonlinear function of the parameters $\mathbf{w} = (w_x, w_y, w_z)^T$. In the sequel, we describe two minimization

techniques: 1) the DE minimization and 2) the LM minimization. The first one is a stochastic search method, and the second one is a directed search method. Moreover, we present two tracking schemes. Recall that the goal is to compute the 3-D pose or the plane parameters for all frames. Finally, we stress the fact that our 3-D pose estimation relies only on the raw-brightness of the region $\mathcal{R}$, i.e., a subset of the right image and the rawbrightness of the left image. We should notice that the choice of SSD criterion is justified by the fact that the scene is imaged by two identical cameras having the same orientation. In other words, the difference in gray levels of a given pair of corresponding points is modeled by a Gaussian noise. It is worth noticing that the preceding SSD function can be replaced by any M-estimator [34].

1) DE Minimization: The DE algorithm is a practical approach to global numerical optimization that is easy to implement, reliable, and fast [35]. We use the DE algorithm [36], [37] to minimize the error (9). This is carried out using generations of solutions, i.e., population. The population of the first generation is randomly chosen around a rough solution. We point out that even the exact solution for the first frame is not known, the search range for the camera height and for the plane normal can be easily known. For example, in our experiments, the camera height and the normal vector are assumed to be around 1 m and $(0, 1, 0)^T$, respectively.

The optimization adopted by the DE algorithm is based on a population of N solution candidates $\mathbf{w}_{n,i}$ $(n = 1, \ldots, N)$ at iteration (generation) i, where each candidate has three components. Initially, the solution candidates are randomly generated within the provided intervals of the search space. The population improves by iteratively generating new solutions for each candidate.

Calibration. Since the stereo camera is rigidly attached to the car, the DE algorithm can also be used as a calibration tool by which the camera pose can be estimated offline. To this end, the car should be at rest and should face a flat road. Whenever the car moves, the offline calibration results can be used as a starting solution for the whole tracking process. Note that the calibration process does not need to run in real time.

2) LM Minimization: Minimizing the cost function (9) can also be carried out using the LM technique [38], which is a well-known nonlinear minimization technique. One can notice that the Jacobian matrix only depends on the horizontal image gradient since the right and left images are rectified.

C. Tracking Schemes

The DE algorithm performs a global search, whereas the LM algorithm performs a directed and local search. Since the unknown parameters (road parameters/camera pose) should be estimated for every stereo pair, we propose two tracking schemes, which are shown in Fig. 13. The first scheme [see Fig. 13(a)] is only based on the DE minimization. In other words, the solution for every stereo frame is computed by invoking the whole algorithm where the first generation is generated by diffusing the previous solution using a normal distribution. A uniform distribution is used for the first stereo frame.

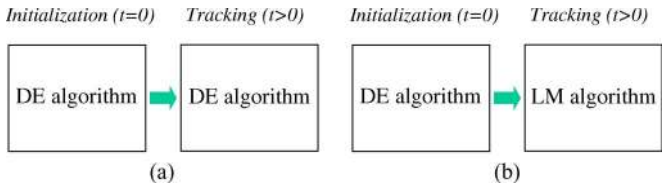

Fig. 13. Parameter tracking using two strategies. (a) Tracking is only based on the DE search. (b) Tracking is based on the LM algorithm, although it is initialized with the DE search.

The second tracking scheme [see Fig. 13(b)] uses the DE minimization for the first stereo frame only. It utilizes the LM algorithm for the rest of the frames where the initial solution for a given frame is provided by the solution $\mathbf{w}^{\star}_{t-1}$ associated with the previous frame.

Although the first scheme might have better convergence properties than the second scheme, the latter scheme is better suited for real-time performance since the LM algorithm is faster than the DE search. (The corresponding central processing unit (CPU) times are illustrated in Section V.) In both tracking schemes, the pose parameters associated with the first stereo pair are estimated by the DE search.

V. Experimental Results

The proposed technique has been tested on different urban environments since they correspond to the most challenging scenarios. In this section, we provide results obtained with two different videos associated with different urban road structures. Moreover, we provide a performance study using synthetic videos with ground-truth data.

A. Tracked Road Parameters

The first experiment has been conducted on a sequence corresponding to an uphill driving. The stereo pairs are of resolution 320 × 240. Fig. 14(a) shows the estimated camera's height as a function of the sequence frames. Fig. 14(b) and (c) shows the estimated pitch and roll angles as a function of the sequence frames, respectively. The dotted curves correspond to the first tracking scheme that is based on the DE minimization. The solid curves correspond to the second tracking scheme, which is essentially based on the LM algorithm. As can be seen, the estimated parameters are almost the same for the two proposed tracking schemes. However, as we will show, the second scheme is much faster than the first scheme (the stochastic search).

1) DE convergence: Fig. 15 shows the behavior of the DE algorithm associated with the first stereo pair of the preceding stereo sequence. This plot depicts the best registration error (SSD per pixel) obtained by every generation. The three curves correspond to three different population sizes. The first generation (iteration 0) has been built using a uniform sampling around the solution $d = 1$ m and $\mathbf{u} = (u_x, u_y, u_z)^T = (0, 1, 0)^T$. The algorithm has converged in five iterations (generations) when the population size was 30 and in two iterations

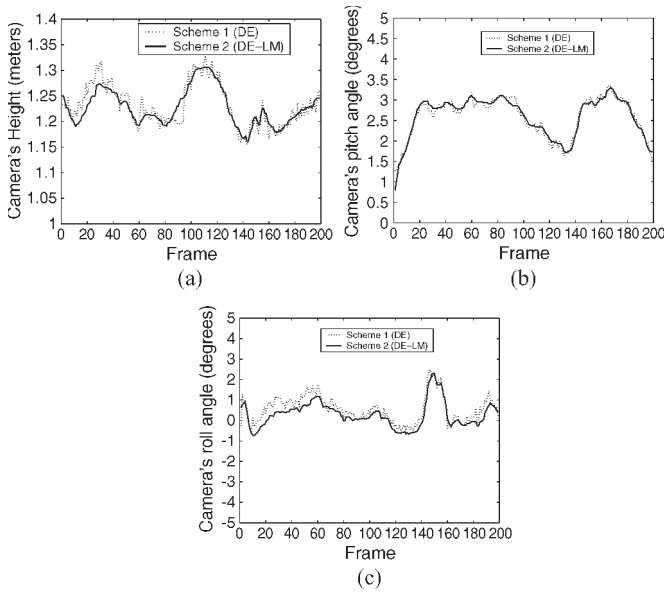

Fig. 14. Camera's height and orientation from the proposed tracking schemes.

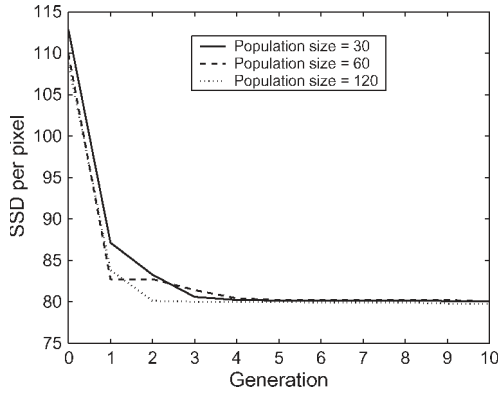

Fig. 15. Evolution of the best registration error obtained by the DE algorithm associated with the first stereo pair. The algorithm has converged in five iterations (generations) when the population size was 30 and in two iterations when the population size was 120.

when the population size was 120. At convergence, the solution was $d = 1.25$ m, and $\mathbf{u} = (u_x, u_y, u_z)^T = (-0.03, 0.99, -0.02)^T$. Note that, even if the manually provided initial camera's height has 25-cm discrepancy from the current solution, the DE algorithm has rapidly converged to the actual solution. In addition, we have run the LM algorithm with the same starting solution, but we get, at convergence, $d = 1.09$ m, and $\mathbf{u} = (u_x, u_y, u_z)^T = (0.01, 0.99, -0.02)^T$. This is not surprising since the LM algorithm can get stuck at nondesired local minima.

2) Horizon line: In the literature, the pose parameters—plane parameters—can be used to compute the horizon line. In our case, since the roll angle is very small, the horizon line can be represented by an horizontal line in the image. Once the 3-D plane parameters d and $\mathbf{u} = (u_x, u_y, u_z)^T$ are computed, the vertical position of the horizon line will be given by

$$v_h = v_0 + \frac{\alpha d}{u_y Z_\infty} - \frac{\alpha u_z}{u_y} \approx v_0 - \frac{\alpha u_z}{u_y}. \qquad (10)$$

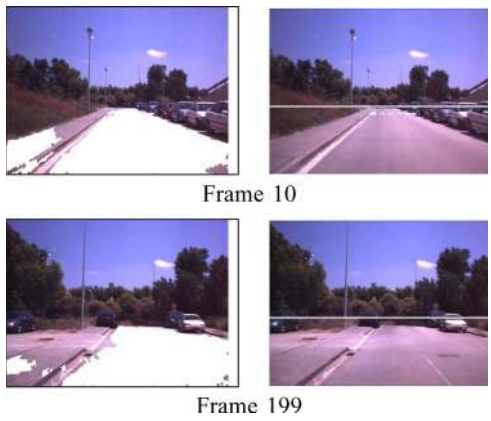

Fig. 16. Estimated horizon line associated with frames 10 and 199. The sequence corresponds to an uphill driving.

The preceding formula is derived by projecting a 3-D point $(0, Y_p, Z_\infty)$ belonging to the road plane and then taking the vertical coordinate $v = \alpha(Y_p/Z_\infty) + v_0$. Z_∞ is a large depth value. The right-hand expression is obtained by using the fact that u_y is close to one and Z_∞ is very large. Fig. 16 shows the computed horizon line for frames 10 and 199. The whole video illustrating the computed horizon line can be found at www.cvc.uab.es/~asappa/HorizonLine.avi.

3) Occlusions: To study the algorithm behavior in the presence of significant occlusions or significant segmentation errors, we conducted the following experiment: We used a video sequence corresponding to a flat road (see Fig. 11). We run the proposed featureless registration technique twice. (The ROI was defined manually.) We used the second tracking scheme differential evolution/Levenberg–Marquardt (DE-LM). In the first run, the stereo images were used as they are. In the second run, the right images were modified to simulate a significant registration error. To this end, we set the vertical half of a set of 20 right images to a fixed color. The left images were not modified. The road region is kept fixed to a rectangular window centered at the bottom of the image.

Fig. 17 compares the pose parameters obtained in the two runs. The solid curves were obtained with the noncorrupted images. The dotted curves were obtained when the right images of the same sequence are artificially corrupted. The simulated occlusion starts at frame 40 and ends at frame 60. The upper part of the figure illustrates the stereo pair 40. As can be seen, the only significant discrepancy has affected the camera height. Moreover, one can see that the correct parameters have been recovered once the occlusion has disappeared. Fig. 18 shows the optimized registration error, which was obtained at convergence, as a function of the sequence frames. As can be seen, the obtained registration error has suddenly increased, which can be used for validating the estimated parameters.

We have plotted the registration error in the neighborhood of the optimal solution $d \approx 1.05$ m and $\beta \approx 3.5^\circ$ (pitch angle). Fig. 19(a) shows the registration error (9) as a function of the camera's height while the orientation is kept fixed. Fig. 19(b) shows the registration error as a function of the camera's pitch angle for four different camera's height. In both figures, the depicted error is the SSD per pixel. From the slop of the error function, we can see that the camera height will not be

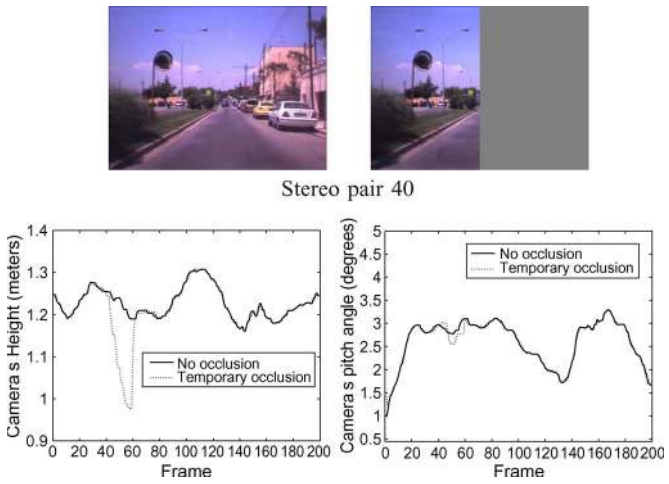

Fig. 17. Comparing the pose parameters when a significant occlusion occurs. The solid curves are obtained with the noncorrupted images. The dotted curves are obtained when 20 frames of right images of the same sequence are artificially corrupted. The occlusion is simulated by setting the vertical half of the right images to a fixed color. This occlusion starts at frame 40 and ends at frame 60.

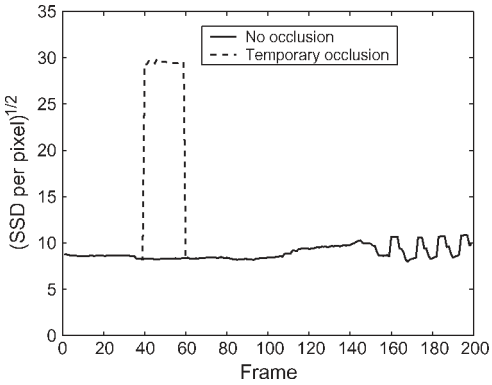

Fig. 18. Registration error obtained at convergence as a function of the sequence frame. The second tracking scheme is used.

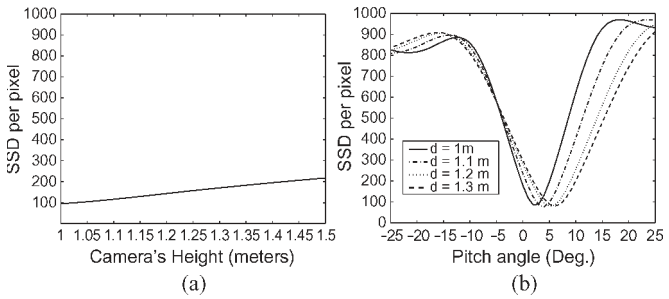

Fig. 19. Registration error as function of the camera pose parameters. (a) Error as a function of the camera height with a fixed orientation. (b) Error as a function of the camera's pitch angle associated with four different camera heights.

recovered with the same accuracy as the plane orientation. This will be confirmed in the accuracy evaluation section.

B. Method Comparison

The second experiment has been conducted on a short sequence corresponding to a typical urban environment (see Fig. 11). The stereo pairs are of resolution 320 × 240. Here,

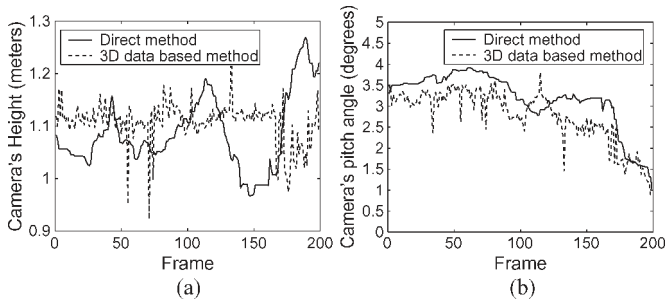

Fig. 20. Camera's height and orientation using two different methods.

the road is almost flat, and the changes in the pose parameters are mainly due to the car's accelerations and decelerations. Fig. 20(a) and (b) show the estimated camera's height and orientation as a function of the sequence frames using two different methods. The solid curves correspond to the developed direct approach (DE-LM), and the dashed curves correspond to a 3-D data based approach [23]. This approach uses a dense 3-D reconstruction, followed by a random sampling consensus (RANSAC)-based estimation of the dominant 3-D plane, i.e., the road plane. One can see that, despite some discrepancies, the proposed direct method is providing the same behavior of changes.

On a 3.2-GHz personal computer, the proposed approach processes one stereo pair in about 20 ms, assuming that the ROI size is 190 × 90 pixels and that the number of the detected road pixels is 11 000 pixels (3 ms for the fast color-based segmentation and about 17 ms for the LM minimization). The same time becomes about 50 ms if the histogram-based segmentation is used.

Moreover, the LM algorithm is faster than the DE algorithm, which needs 120 ms, assuming that the number of iterations is 5 and the population number is 30. (The number of pixels is 11 000.) Obviously, devoting a very small CPU time for estimating the road parameters/camera pose is advantageous for real-time systems since the CPU power can be used for extra tasks such as pedestrian or obstacle detection.

C. Accuracy Evaluation

The evaluation of the proposed method has been carried out on real video sequences, including a comparison with a 3-D 3-D data based approach (see Section V-B). However, it is very challenging to get ground-truth data for the onboard camera pose. In this section, we propose a purely vision-based scheme giving the ground-truth data for the road parameters using synthesized images. To this end, we use 1000 frames captured by the onboard stereo camera. For each stereo pair, we fix the distance (camera height) and the plane normal, i.e., the ground-truth 3-D plane road parameters. Those can be fixed for the whole sequence or can vary according to a predefined trajectory. In our case, we keep them constant for the whole synthesized sequence. Each left image in the original sequence is then replaced with a synthesized one by warping the corresponding right image using the image transfer function encapsulating road parameters. The obtained stereo pairs are then perturbed by adding Gaussian noise to their gray levels.

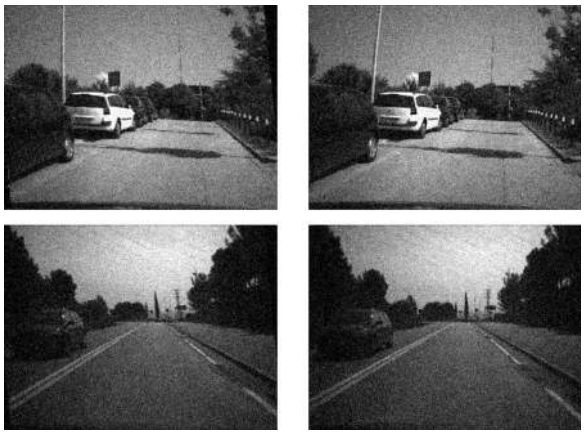

Fig. 21. Two stereo pairs from a perturbed 1000-frame video. The standard deviation of the added Gaussian noise is 20. The left images are synthesized using the ground-truth road parameters.

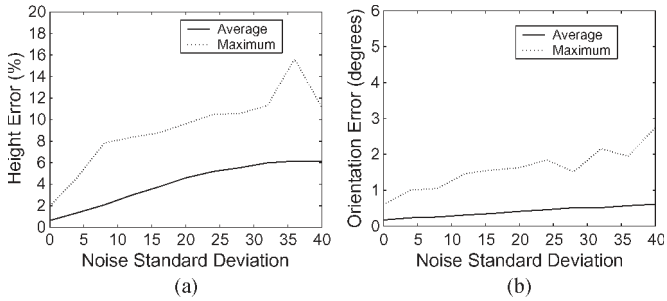

Fig. 22. Errors associated with the plane parameters as a function of the noise standard deviation using synthesized video sequences. (a) Height errors. (b) Plane orientation errors. Each point of the curves—each noise level—corresponds to 10 000 stereo pairs corresponding to ten realizations, each of which is a sequence of 1000 perturbed stereo pairs.

Fig. 21 shows two perturbed stereo pairs. The Gaussian noise standard deviation is set to 20. Here, the gray level of the images has 256 values. The noise-free left image is synthesized using the ground-truth road parameters. The proposed approach is then invoked to estimate the road parameters from the noisy stereo pair. The performance can be directly evaluated by comparing the estimated parameters with the ground-truth parameters. The camera height error is simply the absolute value of the relative error. The orientation error is defined by the angle between the direction of the ground-truth normal and the direction of the estimated one.

Fig. 22 summarizes the obtained errors associated with the synthetic stereo pairs. Fig. 22(a) shows the distance error, and Fig. 22(b) shows the orientation error. Here, one percent error corresponds to 1.2 cm. Each point of the curves—each noise level—corresponds to 10 000 stereo pairs corresponding to ten realizations, each of which is a sequence of 1000 perturbed stereo pairs. The solid curves correspond to the global average of errors over the 10 000 stereo pairs, and the dashed curves correspond to the maximum error.

D. Convergence Study

To study the convergence behavior of the two optimization techniques (the DE algorithm and the LM minimization technique), we run the following experiment. We used the same synthetic stereo sequence containing 1000 stereo frames. The standard deviation of the added image noise is kept fixed to 4. For every stereo frame in the sequence, the starting solution was shifted from the ground-truth solution by 20 cm for the camera height and by 10° for the plane normal. This shifted solution is used as the starting solution for the LM technique and as the center of the first generation for the DE technique. Table I depicts the average height and orientation errors obtained with the LM and DE minimizations for our artificial scenario. As can be seen, the DE minimization has better convergence properties than the LM minimization, which essentially looks for a local minimum. We stress the fact that global minimum of registration error is the same for both search algorithms. However, the LM algorithm has difficulty reaching it if the starting solution is so far from it. Moreover, we can observe that the average error associated with the LM minimization is roughly equal to the introduced shift.

TABLE I
AVERAGE CAMERA POSE ERRORS. THE FIRST COLUMN CORRESPONDS TO THE LM MINIMIZATION, AND THE SECOND COLUMN CORRESPONDS TO THE DIFFERENTIAL EVOLUTION MINIMIZATION. IN THIS EXPERIMENT, THE INITIAL SOLUTION FOR EVERY FRAME WAS ARTIFICIALLY CORRUPTED

1000 stereo frames	*LM mini.*	*DE mini.*
Ave. height error (%)	26.6	3.5
Ave. orientation error (degrees)	10.9	0.41

VI. CONCLUSION

This paper has proposed a new framework for the real-time estimation of the onboard stereo head's position and orientation. This framework can be used with all road types: highways, urban, etc. This paper has provided two main contributions. The first contribution is combining a nonparametric model-based road segmentation algorithm with a registration technique for estimating the online stereo camera pose. The second contribution is solving the registration using a featureless method, which is carried out using two different optimization techniques: 1) the DE algorithm and 2) the LM algorithm.

The method adopts a registration scheme that uses images' brightness. The advantages of the proposed framework are given as follows. First, the road region is segmented by using an illuminant-invariant road model classification where the model is built for every acquired frame. Second, the registration technique that determines the pose parameters does not need any specific visual feature extraction neither in the image domain nor in 3-D space. Third, the technique is fast compared with almost all proposed stereo-based techniques. A good performance has been shown in several scenarios—uphill, downhill, and flat roads. Although it has been tested on urban environments, it could also be useful on highways scenarios. Experiments on real and synthetic stereo sequences have shown that the accuracy of the orientation is better than the height accuracy, which is consistent with all 3-D pose algorithms. The provided experiments tend to confirm the following: 1) The DE search was crucial for obtaining accurate parameter estimation, and

2) the LM technique was crucial for obtaining real-time tracking. As a consequence, the DE optimization can be used as a complementary tool to the LM optimization in the sense that it provides the initialization and the recovery solution from a tracking discontinuity adopting the LM algorithm.

Our proposed generic framework has required that the road geometry in the vicinity of the car is mainly a planar structure and that at least one camera is calibrated for the direction of the invariant axis in the 2-D $\{\log(R/G), \log(B/G)\}$ space. One can notice that, in general, these two requirements are not restrictive.

We believe that the size of the road in the stereo pair has no major impact on the accuracy of the camera pose parameters. Indeed, there are three degrees of freedom that are estimated through image registration of a region having thousands of pixels in both images. The main limitation of the proposed framework is mainly linked to the limitation of the road segmentation in the sense that, if the road segmentation stage fails for some reason, the proposed framework will not able to estimate the 3-D stereo camera pose. A typical segmentation failure may occur when the road is almost totally occluded. However, even in this extreme case, a simple analysis of the obtained segmented road region can be used to invoke or not invoke the registration-based pose estimation.

REFERENCES

[1] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A. Hyder, E. Jarawan, and C. Mathers, *World Report on Road Traffic Injury Prevention*. Geneva, Switzerland: World Health Org., 2004.

[2] R. Labayrade and D. Aubert, "A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision," in *Proc. IEEE Intell. Vehicles Symp.*, 2003, pp. 31–36.

[3] X. Liu and K. Fujimura, "Pedestrian detection using stereo night vision," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1657–1665, Nov. 2004.

[4] Z. Zhu, S. Yang, G. Xu, X. Lin, and D. Shi, "Fast road classification and orientation estimation using omni-view images and neural networks," *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 1182–1197, Aug. 1998.

[5] D. Gerónimo, A. Sappa, D. Ponsa, and A. López, "2D-3D based on-board pedestrian detection system," *Comput. Vis. Image Understand.*, vol. 114, no. 5, pp. 583–595, May 2010.

[6] G. B. Z. Sun and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 125–137, Jun. 2005.

[7] G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, and A. Broggi, "Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2364–2375, Aug. 2006.

[8] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler, "Autonomous driving goes downtown," *IEEE Intell. Syst.*, vol. 13, no. 6, pp. 40–48, Nov./Dec. 1998.

[9] P. Coulombeau and C. Laurgeau, "Vehicle yaw, pitch, roll and 3D lane shape recovery by vision," in *Proc. IEEE Intell. Vehicles Symp.*, Versailles, France, Jun. 2002, pp. 619–625.

[10] J. Collado, C. Hilario, A. Escalera, and J. Armingol, "Adaptive road lanes detection and classification," in *Proc. Adv. Concepts Intell. Vis. Syst.*, 2006, pp. 1151–1162.

[11] Y. Liang, H. Tyan, H. Liao, and S. Chen, "Stabilizing image sequences taken by the camcorder mounted on a moving vehicle," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Shanghai, China, Oct. 2003, pp. 90–95.

[12] D. Ponsa, A. López, F. Lumbreras, J. Serrat, and T. Graf, "3D vehicle sensor based on monocular vision," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Vienna, Austria, Sep. 2005, pp. 1096–1101.

[13] J. Arrospide, L. Salgado, M. Nieto, and R. Mohedano, "Homography-based ground plane detection using a single on-board camera," *IET Intell. Transp. Syst.*, vol. 4, no. 2, pp. 149–160, Jun. 2010.

[14] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 62–81, Jan. 1998.

[15] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, "Shape-based pedestrian detection and localization," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Shanghai, China, Oct. 2003, pp. 328–333.

[16] S. Nedevschi, F. Oniga, R. Danescu, T. Graf, and R. Schmidt, "Increased accuracy stereo approach for 3D lane detection," in *Proc. IEEE Intell. Vehicles Symp.*, 2006, pp. 42–49.

[17] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through 'V-disparity' representation," in *Proc. IEEE Intell. Vehicles Symp.*, Versailles, France, Jun. 2002, pp. 646–651.

[18] M. Bertozzi, E. Binelli, A. Broggi, and M. Del Rose, "Stereo vision-based approaches for pedestrian detection," in *Proc. Comput. Vis. Pattern Recog.*, San Diego, CA, Jun. 2005.

[19] R. Danescu and S. Nedevschi, "Probabilistic lane tracking in difficult road scenarios using stereovision," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 272–282, Jun. 2009.

[20] F. Dornaika and A. Sappa, "Real time on board stereo camera pose through image registration," in *Proc. IEEE Intell. Vehicles Symp.*, 2008, pp. 804–809.

[21] D. Schleicher, L. Bergasa, M. Ocana, R. Barea, and M. E. López, "Real-time hierarchical outdoor SLAM based on stereovision and GPS fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 440–452, Sep. 2009.

[22] G. Finlayson, S. Hordley, C. Lu, and M. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, Jan. 2006.

[23] A. Sappa, F. Dornaika, D. Ponsa, D. Gerónimo, and A. López, "An efficient approach to on-board stereo vision system pose estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 476–490, Sep. 2008.

[24] F. Dornaika and A. Sappa, "A featureless and stochastic approach to on-board stereo vision system pose," *Image Vis. Comput.*, vol. 27, no. 9, pp. 1382–1393, Aug. 2009.

[25] P. Jansen, W. van der Mark, J. van den Heuvel, and F. Groen, "Colour based off-road environment and terrain type classification," in *Proc. IEEE Intell. Transp. Syst.*, 2005, pp. 216–221.

[26] P. Lombardi, M. Zanin, and S. Messelodi, "Switching models for vision-based on-board road detection," in *Proc. IEEE Intell. Transp. Syst.*, 2005, pp. 67–72.

[27] M. Sotelo, F. Rodriguez, and L. Magdalena, "Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 2, pp. 69–83, Jun. 2004.

[28] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 309–318, Dec. 2004.

[29] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," in *Proc. CVPR*, 2004, vol. 1, pp. 470–477.

[30] J. M. Álvarez and A. M. Lopez, "Road detection based on illuminant invariance," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 184–193, Mar. 2001.

[31] Z. Huang and D. Liu, "Segmentation of color image using EM algorithm in HSV color space," in *Proc. Int. Conf. Inf. Acquisition*, 2007, pp. 316–319.

[32] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proc. Int. Conf. Image Process.*, 2002, pp. II-589–II-592.

[33] O. Faugeras and Q. Luong, *The Geometry of Multiple Images*. Cambridge, MA: MIT Press, 2001.

[34] J. Chen, C. Chen, and Y. Chen, "Fast algorithm for robust template matching with M-estimators," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 230–243, Jan. 2003.

[35] K. V. Price, J. A. Lampinen, and R. M. Storn, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer-Verlag, 2005.

[36] S. Das, A. Konar, and U. Chakraborty, "Two improved differential evolution schemes for faster global search," in *Proc. Genetic Evol. Comput.*, 2005, pp. 991–998.

[37] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, Dec. 1997.

[38] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1990.

Fadi Dornaika received the Ph.D. degree in signal, image, and speech processing from Institut National Polytechnique de Grenoble, Grenoble, France, in 1995.

He is currently an Ikerbasque Research Professor with the University of the Basque Country, San Sebastian, Spain. He has published more than 120 papers in the field of computer vision. His research interests include geometrical and statistical modeling with focus on 3-D object pose, real-time visual servoing, calibration of visual sensors, cooperative stereo motion, image registration, facial gesture tracking, and facial expression recognition.

José M. Álvarez (M'11) received the M.Sc. degree in computer science from La Salle School of Engineering, Barcelona, Spain, in 2005 and the Ph.D. degree from the University of Barcelona in 2010.

He is currently a Postdoctoral Researcher with the Advanced Driver Assistance Systems Group, Computer Vision Center, Universitat Autonoma de Barcelona, Bellaterra, Spain. His research interests include road detection, color, photometric invariance, machine learning, and fusion of classifiers.

Angel D. Sappa (S'93–M'99) received the electromechanical engineer's degree from the National University of La Pampa, General Pico, Argentina, in 1995 and the Ph.D. degree in industrial engineering from the Polytechnic University of Catalonia, Barcelona, Spain, in 1999.

In 2003, after holding research positions in France, the U.K., and Greece, he joined the Computer Vision Center, where he is currently a Senior Researcher. He is a member of the Advanced Driver Assistance Systems Group. His research interests span a broad spectrum within 2-D and 3-D image processing.

Antonio M. López received the B.Sc. degree in computer science from the Universitat Politécnica de Catalunya, Barcelona, Spain, in 1992 and the M.Sc. degree in image processing and artificial intelligence and the Ph.D. degree from the Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain, in 1994 and 2000, respectively.

Since 1992, he has been giving lectures with the Computer Science Department, UAB, where currently he is currently an Associate Professor. In 1996, he participated in the founding of the Computer Vision Center, UAB, where he held different institutional responsibilities and is currently responsible for the research group on advanced driver-assistance systems by computer vision. He has been responsible for public and private projects. He is a coauthor of more than 100 papers, all of which are in the field of computer vision.