

Real Time Stereo Image Registration for Planar Structure and 3D Sensor Pose Estimation

Fadi Dornaika¹ and Angel D. Sappa²

¹*Institut Géographique National, 2 avenue Pasteur, 94165 Saint-Mandé,*

²*Computer Vision Center, UAB, 08193 Bellaterra, Barcelona,*

¹*France,*

²*Spain*

1. Introduction

In recent years, several techniques to on-board vision pose estimation have been proposed (Zhu et al., 1998; Labayrade & Aubert, 2003; Liu & Fujimura, 2004; Stein et al., 2000; Suzuki & Kanade, 1999). Vision system pose estimation is required for any advanced driver assistance application. The real-time estimation of on-board vision system pose-position and orientation is a challenging task since i) the sensor undergoes motions due to the vehicle dynamics and the road imperfections, and ii) the viewed scene is unknown and continuously changing.

Of particular interest is the estimation of on-board camera's position and orientation related to the 3D road plane. Note that since the 3D plane parameters are expressed in the camera coordinate system, the camera's position and orientation are equivalent to the 3D plane parameters. Algorithms for fast road plane estimation are very useful for driver assistance applications as well as for augmented reality applications. The ability to use continuously updated plane parameters (camera pose) will considerably make the tasks of obstacle detection more efficient (Viola et al., 2005; Sun et al., 2006; Toulminet et al., 2006). However, dealing with an urban scenario is more difficult than dealing with highways scenario since the prior knowledge as well as visual features are not always available in these scenes (Franke et al., 1999).

In general, monocular vision systems avoid problems related to 3D Euclidean geometry by using the prior knowledge of the environment as an extra source of information. Although prior knowledge has been extensively used to tackle the driver assistance problem, 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, speed bumpers), car's accelerations, uphill/downhill driving, among others.

The use of prior knowledge has also been considered by some stereo vision based techniques to simplify the problem and to speed up the whole processing by reducing the amount of information to be handled (Bertozzi & Broggi, 1998; Bertozzi et al. 2003; Nedeveschi et al., 2006). In the literature, many application-oriented stereo systems have been proposed. For instance, the edge based v -disparity approach proposed in (Labayrade et al., 2002), for an automatic estimation of horizon lines and later on used for applications such as obstacle or pedestrian detection (e.g., (Bertozzi et al., 2005; Labayrade & Aubert,

2003)), only computes 3D information over local maxima of the image gradient. A sparse disparity map is computed in order to obtain a real time performance. Recently, this v -disparity approach has been extended to a u - v -disparity concept in (Hu & Uchimura, 2005). In this work, dense disparity maps are used instead of only relying on edge based disparity maps. Working in the disparity space is an interesting idea that is gaining popularity in on-board stereo vision applications, since planes in the original Euclidean space become straight lines in the disparity space.

In (Sappa et al., 2006), we have proposed an approach for on-line stereo camera pose estimation. Although the proposed technique 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 3D plane that is assumed to be the road plane. This technique has been tested on different urban environments. The proposed algorithm took, on average, 350 ms per frame.

As can be seen, existing works adopt the following main stream. First, features are extracted either in the image space (optical flow, edges, ridges, interest points) or in the 3D Euclidean space (assuming the 3D data are built online). Second, an estimation technique is then invoked in order to recover the unknown parameters.

In this chapter, we propose a novel paradigm for on-board camera pose tracking through the use of image registration (Romero & Calderon, 2007). Since we do not rely on features, the image registration should be featureless. We solve the featureless registration by using two optimization techniques: the Differential Evolution algorithm (a stochastic search) and the Levenberg-Marquardt algorithm (a directed search). Moreover, 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. We stress the fact that our proposed methods are not restricted to the estimation of on-board camera pose/roads. Indeed, the proposed methods can be used for extracting any planar structures using stereo pairs.

The rest of the chapter is organized as follows. Section 2 describes the problem we are focusing on as well as some backgrounds. Section 3 presents the proposed approach in details. Section 4 gives some experimental results and method comparisons. Section 5 concludes the chapter.

2 Problem formulation and background

2.1 Experimental setup

A commercial stereo vision system (Bumblebee from Point Grey¹) was used. It consists of two Sony ICX084 color CCDs with 6mm focal length lenses. Bumblebee is a pre-calibrated system that does not require in-field calibration. The baseline of the stereo head is 12cm and it is connected to the computer by an IEEE-1394 connector. Right and left color images can be captured at a resolution of 640×480 pixels and a frame rate near to 30 fps. This vision system includes a software able to provide the 3D data. Figure 1(a) shows an illustration of the on-board stereo vision system as well as its mounting device.

The problem we are focusing on can be stated as follows. Given a stream of stereo pairs provided by the on-board 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 on-board vision sensor. This pose is represented by the 3D plane parameters, that is, the height d and

¹ [www.ptgrey.com]

the plane normal $\mathbf{u} = (u_x, u_y, u_z)^T$ from which two independent angles can be inferred (see Figure 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 Figure 1(b)). Due to the reasons mentioned above, 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 3D plane parameters.

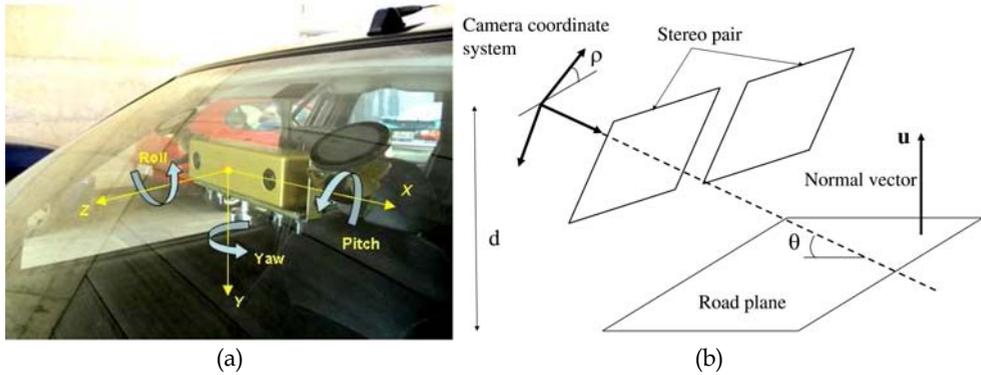


Fig. 1. (a) On-board stereo vision sensor. (b) The time-varying road plane parameters d and \mathbf{u} . θ denotes the pitch angle and ρ the roll angle.

2.2 Image transfer function

Before going into the details of the proposed approach, this section will describe the geometric relation between road pixels belonging to the same stereo pair—the left and right images. It is well-known (Faugeras & Luong, 2001) that the image coordinates of the projections of 3D points belonging to the same plane onto two different images are related by a 2D projective transform having 8 independent parameters—*homography*. In our setup, the right and left images are horizontally rectified². Let $p_r(x_r, y_r)$ and $p_l(x_l, y_l)$ be the right and left projection of an arbitrary 3D point P belonging to the plane (d, u_x, u_y, u_z) . In the case of a rectified stereo pair where the left and right cameras have the same intrinsic parameters, 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):

$$x_l = h_1 x_r + h_2 y_r + h_3 \quad (1)$$

$$y_l = y_r \quad (2)$$

where h_1 , h_2 , and h_3 are function 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:

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

$$h_1 = 1 + b \frac{u_x}{d} \quad (3)$$

$$h_2 = b \frac{u_y}{d} \quad (4)$$

$$h_3 = -b u_0 \frac{u_x}{d} - b v_0 \frac{u_y}{d} + \alpha b \frac{u_z}{d} \quad (5)$$

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 3-vector encapsulating the 3D plane parameters, that is, $\mathbf{w} = \frac{\mathbf{u}}{d}$.

$$\mathbf{w} = (w_x, w_y, w_z)^T = \left(\frac{u_x}{d}, \frac{u_y}{d}, \frac{u_z}{d} \right)^T \quad (6)$$

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 3D road plane parameters d and \mathbf{u} or equivalently the vector \mathbf{w} .

3. Approach

Since the goal is to estimate the road plane parameters \mathbf{w} for every stereo pair (equivalently the 3D pose of the stereo head), the whole process is invoked for every stereo pair. Figure 2 illustrates the tracking of the stereo head pose over time. The inputs to the algorithm are the current stereo pair as well as the estimated road plane parameters associated with the previous frame. The algorithm is split into two consecutive stages. First, a rough road region segmentation is preformed for the right image. Let \mathcal{R} denotes this region—a set of pixels. Second, 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 plane parameters—the distance d and the normal \mathbf{u} —then the registration error between corresponding road pixels in the right and left images over the 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:

$$e(\mathbf{w}) = \sum_{(x_r, y_r) \in \mathcal{R}} (I_r(x_r, y_r) - I_l(h_1 x_r + h_2 y_r + h_3, y_r))^2 \quad (7)$$

The corresponding left pixels are computed according to the linear transform given by (1) and (2). The computed $x_l = h_1 x_r + h_2 y_r + h_3$ is a non-integer value. Therefore, the grey-level, $I_l(x_l, y_l)$, is set to a linear interpolation of the grey-level of two neighboring pixels—the ones whose horizontal coordinates bracket the value x_l .

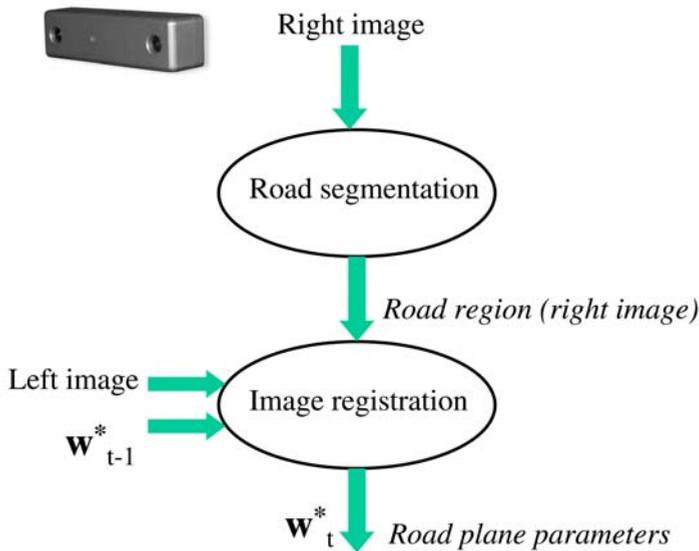


Fig. 2. The proposed approach consists of two stages: A rough road segmentation followed by image registration.

3.1 Road segmentation

In this section, we briefly describe how the road region \mathcal{R} is detected in the right images. Road segmentation is the focus of many research works (Lombardi et al., 2005; Jansen et al., 2005; Guzman & Parra, 2007; Alvarez et al., 2008). In our study, the sought segmentation should meet two requirements: (i) it should be as fast as possible, and (ii) it should be as generic as possible (both urban roads and highways). Thus our segmentation scheme will be a color-based approach which works on the hue and saturation components. The segmentation stage is split 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 on the hue and saturation 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 3 and 5. The cluster having the largest number of pixels will be assumed to belong to the road. Once the cluster is identified, the mean and the 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 quantify the likelihood of an arbitrary pixel to be a road pixel. Thus, the pixels within the ROI are labelled as road pixels if their *Mahalanobis* distance to the mean is below a certain threshold. Figure 3 shows the segmentation results obtained with the proposed scheme. Detected road pixels are shown in white within the ROI of two different frames. As can be seen, all detected pixels belong to the road plane.

3.2 Image registration

The optimal current road parameters are given by:

$$\begin{aligned}
\mathbf{w}^* &= \arg \min_{\mathbf{w}} e(\mathbf{w}) \\
&= \arg \min_{\mathbf{w}} \sum_{(x_r, y_r) \in \mathcal{R}} (I_r(x_r, y_r) - I_l(h_1 x_r + h_2 y_r + h_3, y_r))^2
\end{aligned} \tag{8}$$

where $e(\mathbf{w})$ is a non-linear function of the parameters $\mathbf{w} = (w_x, w_y, w_z)^T$. In the sequel, we describe two minimization techniques: i) the Differential Evolution minimization, and ii) the Levenberg-Marquardt minimization. The first one is a stochastic search method and the second one is a directed search method. Moreover, we present two tracking schemes.



Fig. 3. Rapid road segmentation associated with two frames.

3.2.1 Differential evolution minimization

The Differential Evolution algorithm (DE) is a practical approach to global numerical optimization that is easy to implement, reliable and fast (Price et al., 2005). We use the DE algorithm (Das et al. 2005; Storn & Price, 1997) in order to minimize the error (8). This is carried out using generations of solutions-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 as well as 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 $1m$ and $(0, 1, 0)^T$, respectively.

The optimization adopted by the DE algorithm is based on a population of N solution candidates $\mathbf{w}_{n,i}$ ($n = 1, \dots, N$) 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 generating new solutions iteratively for each candidate.

Calibration. Since the stereo camera is rigidly attached to the car, the differential evolution algorithm can also be used as a calibration tool by which the camera pose can be estimated off-line. To this end, the car should be at rest and should face a flat road. Whenever the car moves, the off-line 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.

3.2.2 Levenberg-Marquardt minimization

Minimizing the cost function (8) can also be carried out using the Levenberg-Marquardt technique (Fletcher, 1990; Press et al., 1992) —a well-known non-linear minimization

technique. One can notice that the Jacobian matrix only depends on the horizontal image gradient since the right and left images are rectified.

3.3 Tracking schemes

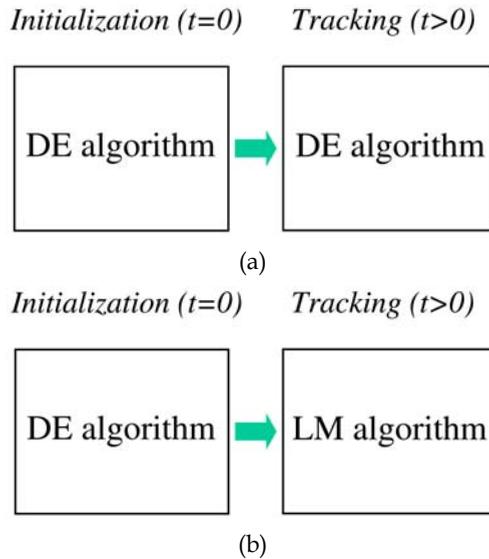


Fig. 4. Parameter tracking using two strategies. (a) The tracking is only based on the Differential Evolution search. (b) The tracking is based on the Differential Evolution search and on the Levenberg-Marquardt search.

The unknown parameters (road parameters/camera pose) should be estimated for every stereo pair. Thus, we will adopt a tracking strategy in which the estimated parameters \mathbf{W}_{t-1}^* associated with the previous frame will be handed over to the current frame.

Since the unknown parameters (road parameters/camera pose) are estimated by two optimization techniques, we propose two tracking schemes which are illustrated in Figure 4. The first scheme (Figure 4(a)) is only based on the Differential Evolution 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.

The second scheme (Figure 4(b)) uses the Differential Evolution minimization for the first stereo frame only. It utilizes the Levenberg-Marquardt for the rest of the frames where the initial solution for a given frame is provided by the solution \mathbf{W}_{t-1}^* associated with the previous frame.

Although the first scheme might have better convergence properties than the second scheme, the latter one is better suited for real-time performance since the Levenberg-Marquardt algorithm is faster than the Differential Evolution search (the corresponding CPU times are illustrated in Section 4.2). In both tracking schemes, the pose parameters associated with the first stereo pair are estimated by the DE search. The Differential

Evolution algorithm performs a global search whereas the Levenberg-Marquardt performs a directed and local search.

4. 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.

4.1 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 . Figure 5(a) depicts the estimated camera's height as a function of the sequence frames. Figures 5(b) and 5(c) depict the estimated pitch and roll angles as a function of the sequence frames, respectively. The dotted curves correspond to the first scheme that is based on the Differential Evolution minimization. The solid curves correspond to the second scheme which is based on both the Differential Evolution algorithm and the Levenberg-Marquardt algorithm. As can be seen, the estimated parameters are almost the same for the two proposed schemes. However, as we will show, the second scheme is much faster than the first scheme (the stochastic search).

Differential Evolution convergence. Figure 6 illustrates the behavior of the Differential Evolution algorithm associated with the first stereo pair of the above 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 $d = 1m$ and $\mathbf{u} = (u_x, u_y, u_z)^T = (0, 1, 0)^T$. The algorithm converged in five iterations (generations) when the population size was 30 and in two iterations when the population size was 120. At convergence the solution was $d = 1.25m$ and $\mathbf{u} = (u_x, u_y, u_z)^T = (-0.03, 0.99, -0.02)^T$. Note that even the manually provided initial camera's height has 25cm discrepancy from the current solution, the DE algorithm has rapidly converged to the actual solution. Also, we have run the Levenberg-Marquardt algorithm with the same starting solution but we get at convergence $d = 1.09m$ and $\mathbf{u} = (u_x, u_y, u_z)^T = (0.01, 0.99, -0.02)^T$.

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 3D 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:

$$v_h = v_0 + \frac{\alpha d}{u_y Z_\infty} - \frac{\alpha u_z}{u_y} \approx v_0 - \frac{\alpha u_z}{u_y} \quad (9)$$

The above formula is derived by projecting a 3D point $(0, Y_p, Z_\infty)$ belonging to the road plane and then taking the vertical coordinate $v = \alpha \frac{Y_p}{Z_\infty} + v_0$. 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. Figure 7 illustrates 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.

Approach behavior in the presence of road segmentation error. In order to study the algorithm behavior in the presence of significant segmentation errors or non-road objects, we conducted the following experiment. We used a video sequence corresponding to a flat road (see Figure 3). We run the proposed technique described in Section 3 twice. We used the second tracking scheme (DE-LM). The first run was a straightforward run. In the second run, the right images were corrupted to simulate a significant registration error (road segmentation 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.

Figure 8 compares the pose parameters obtained in the two runs. The solid curves were obtained with the non corrupted images. The dotted curves were obtained when the right images of the same sequence are artificially corrupted. The simulated corruption 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 perturbing factor has disappeared. Figure 9 shows the registration error 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.

Figure 10(a) illustrates the registration error (8) as a function of the camera's height while the orientation is kept fixed. Figure 10(b) illustrates 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 slope of the error function we can see that the camera height will not be recovered with the same accuracy as the plane orientation. This will be confirmed in the accuracy evaluation section (see Section 4.3).

4.2 Method comparison

The second experiment has been conducted on a short sequence of stereo pairs corresponding to a typical urban environment (see Figure 3). The stereo pairs are of resolution 320×240 . Here the road is almost flat and the changes in the pose parameters are mainly due to the car's accelerations and decelerations. Figures 11(a) and 11(b) depict 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 3D data based approach (Sappa et al., 2006). This approach uses a dense 3D reconstruction followed by a RANSAC-based estimation of the dominant 3D plane—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 PC, the proposed approach processes one stereo pair in about 20 ms assuming that the ROI size is 190×90 pixels and the number of the detected road pixels is 11000 pixels (3 ms for the fast color-based segmentation and about 17 ms for the Levenberg-Marquardt minimization). One can notice that this is much faster than the 3D data based approach, which needs 350 ms. Moreover, the Levenberg-Marquardt 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 11000). 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.

4.3 Accuracy evaluation

The evaluation of the proposed approach has been carried out on real video sequences, including a comparison with a 3D data based approach (Section 4.2). However, it is very challenging to get ground-truth data for the on-board camera pose. In this section, we propose a simple scheme giving the ground-truth data for the road parameters through the use of synthetic stereo sequences. To this end, we use a 1000-frame real video captured by the on-board stereo camera. For each stereo pair, we set the distance (camera height) and the plane normal—the ground-truth 3D plane road parameters. Those ones can be constant 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 the road parameters. The obtained stereo pairs are then perturbed by adding Gaussian noise to their grey levels.

Figure 12 depicts a perturbed stereo pair. The Gaussian noise standard deviation is set to 20. Here the grey-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.

Figure 13 summarizes the obtained errors associated with the synthetic stereo pairs. Figure 13(a) depicts the distance error and Figure 13(b) the orientation error. Here one percent error corresponds to 1.2cm. Each point of the curves—each noise level—corresponds to 10000 stereo pairs corresponding to 10 realizations, each of which is a sequence of 1000 perturbed stereo pairs. The solid curves correspond to the global average of errors over the 10000 stereo pairs and the dashed curves correspond to the maximum error. As can be seen, the performance of the method downgrades gracefully with the image noise. Moreover, one can appreciate the orientation accuracy.

4.4 Convergence study

In order to study the convergence behavior of the two optimization techniques 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 degrees for the plane normal. This shifted solution is used as the starting solution for the Levenberg-Marquardt technique and as the center of the first generation for the Differential Evolution technique. Table 1 depicts the average height and orientation errors obtained with the LM and DE minimizations. As can be seen, the DE minimization has better convergence properties than the LM minimization which essentially looks for a local minimum.

5. Conclusion

A featureless technique for real time estimation of on-board stereo head pose has been presented. The method adopts a registration scheme that uses images' brightness. The advantages of the proposed technique are as follows. First, the technique does not need any specific visual feature extraction neither in the image domain nor in 3D space. Second, the technique is very fast compared to almost all proposed stereo-based techniques. The proposed featureless registration is carried out using two optimization techniques: the Differential Evolution algorithm (a stochastic search) and the Levenberg-Marquardt algorithm (a directed search).

A good performance has been shown in several scenarios—uphill, downhill and flat roads. Although it has been tested on urban environments, it could be also 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 3D pose algorithms. The provided experiments tend to confirm that (i) the Differential Evolution search was crucial for obtaining an accurate parameter estimation, and (ii) the Levenberg-Marquardt technique was crucial for obtaining a 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 as well as the recovery solution from a tracking discontinuity adopting the Levenberg-Marquardt algorithm.

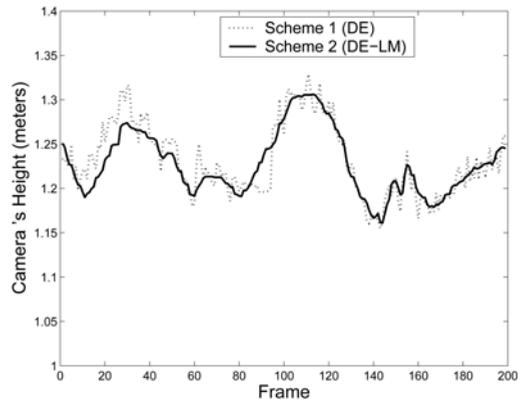
We stress the fact that our proposed framework is not restricted to the estimation of on-board camera pose/roads. Indeed, the proposed methods can be used for extracting any planar structures using stereo pairs.

6. Acknowledgment

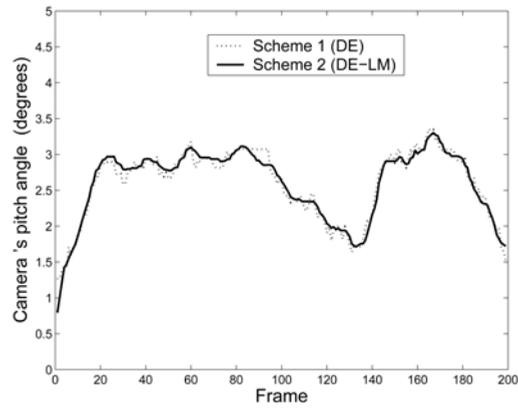
This work was partially supported by the Government of Spain under MEC project TRA2007-62526/AUT and research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

1000 stereo frames	<i>LM minimization</i>	<i>DE minimization</i>
Ave. height error (%)	26.6	3.5
Ave. orientation error (degrees)	10.9	0.41

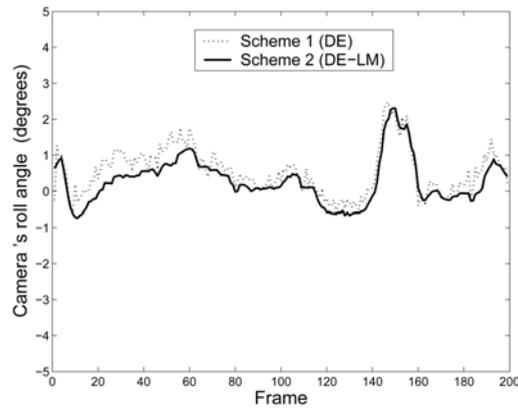
Table 1. Average camera pose errors. The first column corresponds to the Levenberg-Marquardt minimization and the second column to the Differential Evolution minimization.



(a)



(b)



(c)

Fig. 5. Camera's height and orientation computed by the proposed tracking schemes.

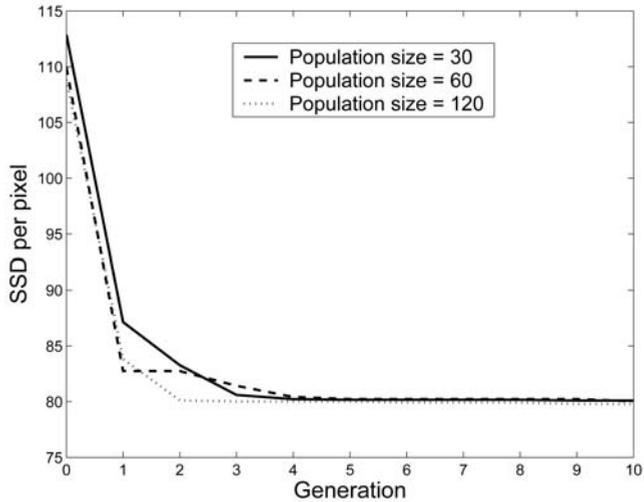


Fig. 6. The evolution of the best registration error obtained by the Differential Evolution algorithm associated with the first stereo pair. The algorithm has converged in 5 iterations (generations) when the population size was 30 and in two iterations when the population size was 120.



Fig. 7. The estimated horizon line associated with frames 10 and 199. The sequence corresponds to an uphill driving.

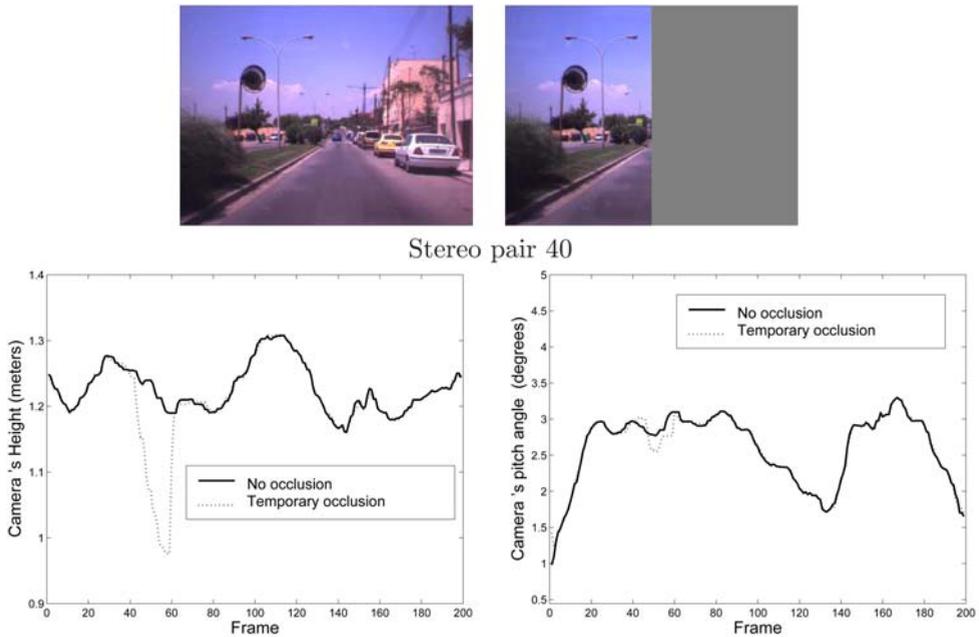


Fig. 8. The camera pose parameters in the presence of a significant corruption (road segmentation errors). The solid curves are obtained with the non corrupted images. The dotted curves are obtained when 20 frames of right images of the same sequence are artificially corrupted. The corruption is simulated by setting the vertical half of the right images to a fixed color. This corruption starts at frame 40 and ends at frame 60.

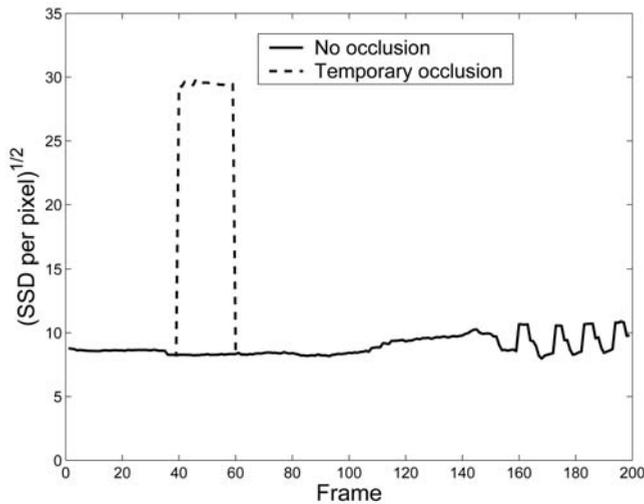


Fig. 9. The registration error obtained at convergence as a function of the sequence frame. The second tracking scheme is used.

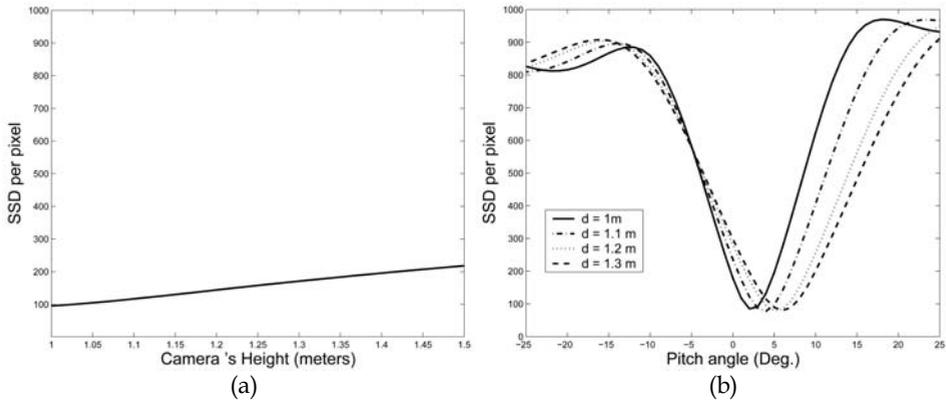


Fig. 10. The registration error as a function of the camera pose parameters. (a) Depicts the error as a function of the camera height with a fixed orientation. (b) Depicts the error as a function of the camera's pitch angle associated with four different camera heights.

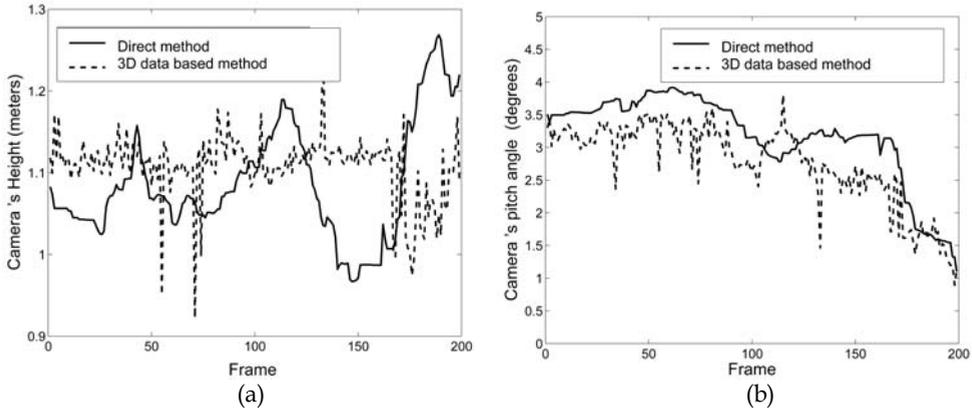


Fig. 11. Camera's height and orientation using two different methods.

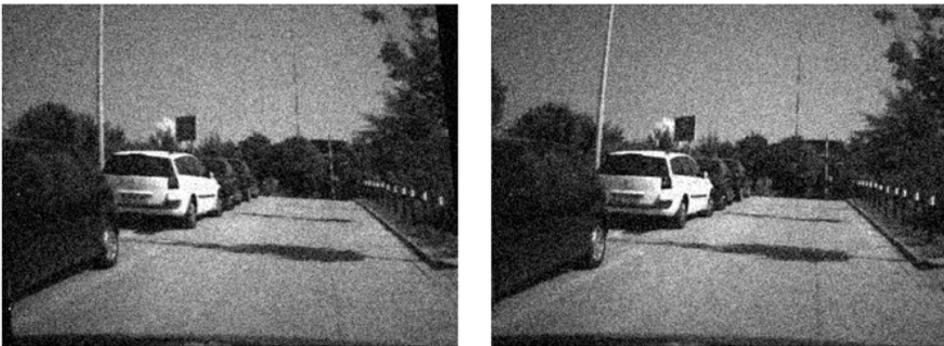


Fig. 12. A stereo pair from a perturbed 1000-frame video. The standard deviation of the added Gaussian noise is 20. The left images are synthesized using the ground-truth road parameters.

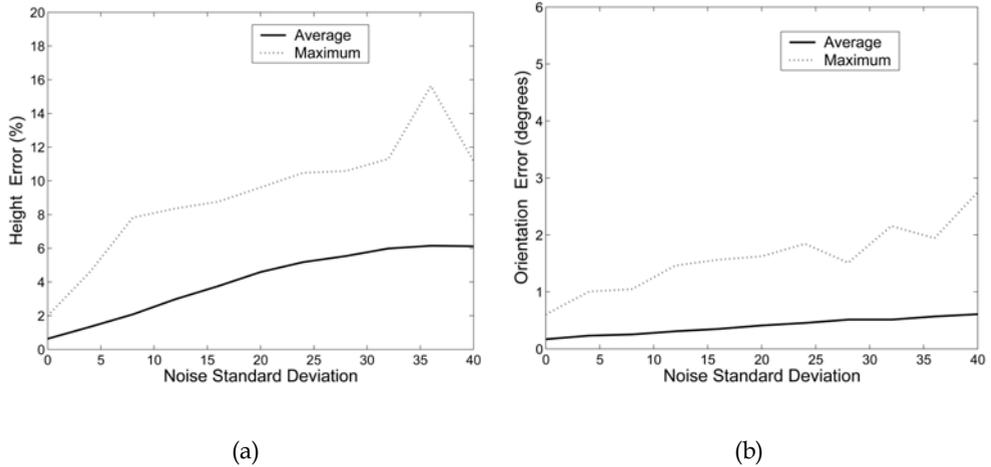


Fig. 13. The errors associated with the plane parameters as a function of the noise standard deviation using synthesized video sequences. (a) Depicts the height errors. (b) Depicts the plane orientation errors. Each point of the curves—each noise level—corresponds to 10000 stereo pairs corresponding to 10 realizations, each of which is a sequence of 1000 perturbed stereo pairs.

7. References

- J. M. Alvarez, A. López, and R. Baldrich. Illuminant-invariant model-based road segmentation. In *IEEE Intelligent Vehicules Symposium*, 2008.
- M. Bertozzi, E. Binelli, A. Broggi, and M. Del Rose. Stereo vision-based approaches for pedestrian detection. In *Procs. Computer Vision and Pattern Recognition*, San Diego, USA, June 2005.
- M. Bertozzi and A. Broggi. GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Trans. on Image Processing*, 7(1):62-81, January 1998.
- M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi. Shape-based pedestrian detection and localization. In *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems*, pages 328-333, Shangai, China, October 2003.
- S. Das, A. Konar, and U. Chakraborty. Two improved differential evolution schemes for faster global search. In *Genetic and Evolutionary Computation*, 2005.
- O. Faugeras and Q.T. Luong. *The Geometry of Multiple Images*. The MIT Press, 2001.
- R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, 1990.
- U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler. Autonomous driving approaches downtown. *IEEE Intelligent Systems*, 13(6):1-14, 1999.

- A. Guzmán and C. Parra. Extraction of roads from outdoor images. In *Vision Systems: Applications*, pages 101-112. 2007.
- Z. Hu and K. Uchimura. U-V-Disparity: An efficient algorithm for stereovision based scene analysis. In *Procs. IEEE Intelligent Vehicles Symposium*, pages 48-54, Las Vegas, USA, June 2005.
- P. Jansen, W. van der Mark, J.C. van der Heuvel, and F.C.A. Groen. Colour based off-road environment and terrain type classification. In *IEEE Intelligent Transportation Systems*, 2005.
- R. Labayrade and D. Aubert. A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision. In *Proc. IEEE Intelligent Vehicles Symposium, Columbus, OH, USA*, pages 31-36, June 2003.
- 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 Intelligent Vehicles Symposium, Versailles, France*, pages 646-651, June 2002.
- X. Liu and K. Fujimura. Pedestrian detection using stereo night vision. *IEEE Trans. on Vehicular Technology*, 53(6):1657-1665, November 2004.
- P. Lombardi, M. Zanin, and S. Messelodi. Switching models for vision-based on-board road detection. In *IEEE Intelligent Transportation Systems*, 2005.
- S. Nedeveschi, F. Oniga, R. Danescu, T. Graf, and R. Schmidt. Increased accuracy stereo approach for 3D lane detection. In *IEEE Intelligent Vehicles Symposium*, 2006.
- W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery. *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, New York, 1992.
- K. V. Price, J. A. Lampinen, and R. M. Storn. *Differential Evolution: A Practical Approach To Global Optimization*. Springer, 2005.
- L. Romero and F. Calderón. A tutorial on parametric image registration. In *Scene Reconstruction, Pose Estimation and Tracking*, pages 167-184. 2007.
- A. Sappa, D. Gerónimo, F. Dornaika, and A. López. On-board camera extrinsic parameter estimation. *Electronics Letters*, 42(13):745-747, June 2006.
- G. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. In *IEEE Intelligent Vehicles Symposium*, 2000.
- R. Storn and K. Price. Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341-359, 1997.
- Z. Sun, G. Bebis, and R. Miller. Monocular precrash vehicle detection: Features and classifiers. *IEEE Trans. on Image Processing*, 15(7):2019-2034, 2006.
- T. Suzuki and T. Kanade. Measurement of vehicle motion and orientation using optical flow. In *IEEE Intelligent Vehicles Symposium*, 1999.
- G. Toulminet, M. Bertozzi, S. Mousset, A. Benschrair, and A. Broggi. Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis. *IEEE Trans. on Image Processing*, 15(8):2364-2375, 2006.
- P. Viola, M.J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153-161, 2005.

-
- 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. on Image Processing*, 7(8):1182-1197, 1998.