Chapter Number

Calibration and Sensitivity Analysis of a Stereo Vision-Based Driver Assistance System

András Bódis-Szomorú, Tamás Dabóczi and Zoltán Fazekas
Computer and Automation Research Institute &
Budapest University of Technology and Economics
Hungary

1. Introduction

As safety and efficiency issues of transportation - hand-in-hand with the intelligent vehicle concept – have gathered ground in the last few years in the automotive research community and have penetrated into the automotive industry, vision-based applications have become increasingly important in driver assistance systems (Kastrinaki et al., 2003; Bertozzi et al., 2002). The primary targets of the safety and efficiency improvements are intelligent cruise control (e.g. vehicle following), lane keeping and lane departure warning systems, assistance in lane changing, avoidance of collision against vehicles, obstacles and pedestrians, vision enhancement and traffic sign recognition and signalling. Our focus of interest here is stereo machine vision used in the context of lane departure warning and lane keeping assistance. The primary purposes of our vision system are to determine the vehicle's position and orientation within the current lane, and the shape of the visible portion of the actual lane on structured roads and highways. That is, we face a somewhat simplified simultaneous localization and mapping (SLAM) problem here, with the assumption of a structured man-built environment and a limited mapping requirement. The localization is achieved through the 3D reconstruction of the lane's geometry from the acquired images.

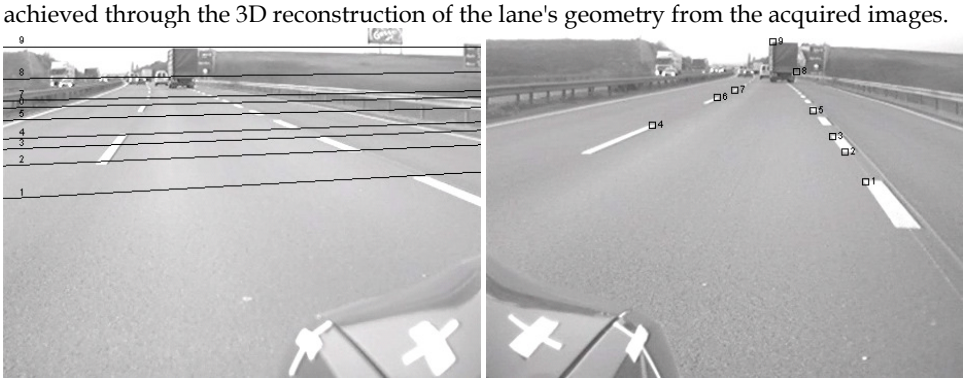

Fig. 1. Images acquired with our wide-baseline stereo vision system in a typical highway scene. The overlaid epipolar lines in the left image correspond to the marked points in the right image.

In monocular lane detection systems, reconstruction usually relies on a number of assumptions concerning the scene geometry and the vehicle motion, such as flat road and constant pitch and roll angles, which are not always valid (Kastrinaki et al., 2003; Marita et al., 2006). Stereo vision provides an effective and reliable means to extract correct range information without the need of relying on doubtful assumptions (Hartley & Zisserman, 2006). Nevertheless, there are some stereo systems that still make use of such assumptions while providing a higher level of robustness compared to monocular systems. For example, certain systems use inverse perspective mapping (IPM) to remap both the left and the right images to the road's plane, i.e. to generate a "bird's view", where lane detection is performed while stereo disparity can be exploited in vehicle detection by analysing the difference between the remapped images (Bertozzi & Broggi, 1998). Another solution is to use a stereo pair in standard configuration (parallel optical axes and image coordinate axes) and the Helmholtz shear equation to relate the precomputed sparse stereo disparity to the distances measured in the road's plane and also to classify the detected 3D points to "near-road" and "off-road" points (Weber et al., 1995). Some algorithms perform a sparse stereo reconstruction of the 3D scene through the classical stereo processing steps of feature detection, correlation-based matching to solve the correspondence problem and triangulation (Nedevschi et al., 2005). Stereo matching is performed by making use of the constraints arising from the epipolar geometry, that is, the search regions are simplifed to epipolar lines (Hartley & Zisserman, 2006), as shown in Figure 1. In the standard arrangement, the correspondence problem simplifies to a search along corresponding scanlines, thus, such a setup is more suitable for real-time applications (Weber et al., 1995). Even if the cameras are in general configuration it is possible to virtually rotate the cameras and rectify the images with a suitable homography (Hartley & Zisserman, 2006). However, this transformation is time-consuming and when sub-pixel accuracy can be achieved in feature extraction, image warping can jeopardize the overall accuracy (Nedevschi et al., 2006). This consideration is increasingly valid for lane detection systems where the scene's depth can easily reach 60 to 100 meters.

Even if up-to-date tracking methods (involving dynamical scene and ego-motion models) are used to make the reconstruction more robust, the precision and even the feasability of the reconstruction using any of the techniques mentioned above is seriously affected by the precision of the camera parameters that are typically determined prelimarily with camera calibration. IPM techiques (and also the one based on the Helmholtz shear equation) may fail when the relative orientations of cameras with respect to the road's plane are not precisely determined (Weber et al., 1995; Broggi et al., 2001). Since epipolar lines and the rays at the triangulation step are highly dependent on the camera parameters, both stereo matching and 3D reconstruction may prove unsatisfactory (Marita et al., 2006). These facts highlight the importance of an accurate calibration and motivate a thorough sensitivity analysis in the design of such safety-critical systems.

In this chapter, we investigate the effects of parameter uncertainties on the quality of 3D geometrical reconstruction. We propose an off-line and far-range camera calibration method for stereo vision systems in a general confguration. Due to the high depth range, we restrict our analysis to perspective cameras. The cameras are calibrated individually by using fairly common planar patterns, and camera poses with respect to the road are computed from a specifc quasi-planar marker arrangement. Since a precise far-range arrangement might be difficult and costly to set up, we put up with an inexpensive and less precise arrangement,

and formulate the maximum likelihood estimate of the camera parameters for it. We demonstrate how our method overperforms the widely used reprojection error minimization method when significant errors are present in the marker arrangement. By applying the presented methods to real images, we give an estimate on the precision of 3D lane boundary reconstruction.
The chapter is organized as follows. We outline our preliminary lane detection algorithm for stereovision in Section 2. Next, a brief overview is given of existing calibration methods and their applicability in the field in Section 3. Then, in Section 4, we present an intrinsic camera calibration and a single-view pose estimation method. The proposed stereo calibration method and the overall sensitivity analysis, including epipolar line uncertainty are derived in Section 5. In Section 6, evaluation of the methods based on real data is presented. Finally, Section 7 concludes the chapter.

2. The lane detection algorithm

In the current section, we suppose that our wide-baseline system with two forward-looking grayscale cameras fixed to the side mirrors of a host vehicle (we refer to Figure 1) is already calibrated and the camera poses are known. Then we return to the problem of calibration in details in Section 3. The outline of the algorithm is depicted in Figure 2.

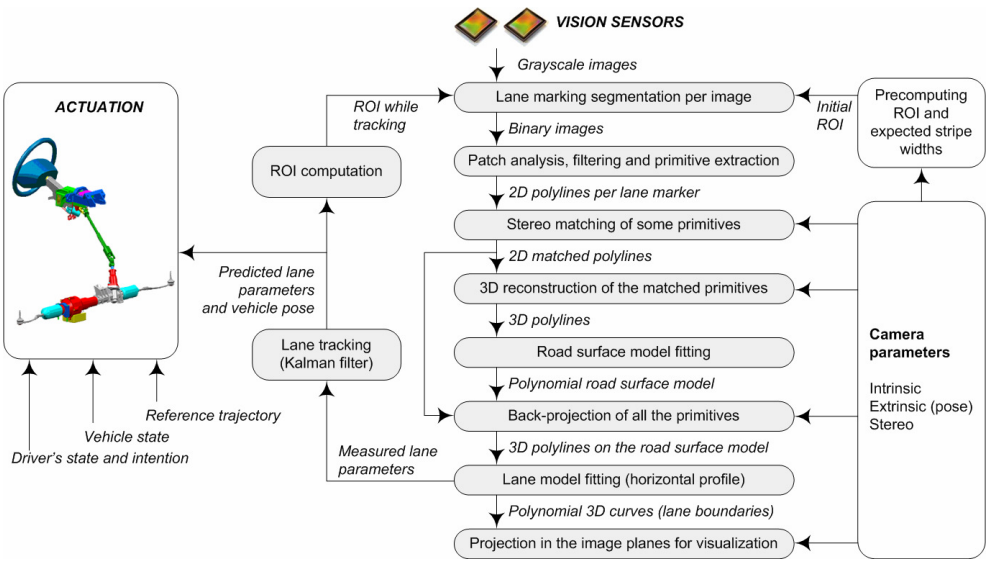

Fig. 2. Outline of a lane keeping assistant system based on the lane boundary detection and a stereo lane reconstruction algorithm. The algorithm highly depends on the camera parameters. ROI stands for Regions-of-Interest.

In order to minimize the resource requirements of the algorithm, we avoid using dense stereo reconstruction and image rectification. In sparse methods, interesting features are defined and extracted from the images by a feature detector.

2.1 Lane marking extraction

The presented algorithm currently relies on the presence of lane markings. We have developed a lane marking detector that is capable of extracting perspectively distorted bright stripes of approximately known metrical width over a darker background. The lane marking extraction is performed along scanlines with several variable-width 1-dimensional templates within precomputed regions-of-interest (ROI). The ROI computation requires the a priori knowledge of the camera parameters as the ROI boxes in the images are determined from rectangular regions specified metrically in the ground plane (which does not exactly match the road's surface). In the current study, we focus on the pure lane extraction algorithm working on independent successive frames without considering temporal relations and only using the known camera parameters as a priori information. Tracking of the lane boundary curves over time (e.g. with a Kalman filter) can be added easily to this framework. It gives more robustness and stability to the estimated parameters. Solutions for an elaborate dynamical modeling of the problem exist in the literature (Eidehall & Gustafsson, 2004; Bertozzi et al., 2002). Thus, the presented algorithm can be considered as an initialization stage for lane (and object) tracking (see Figure 2).

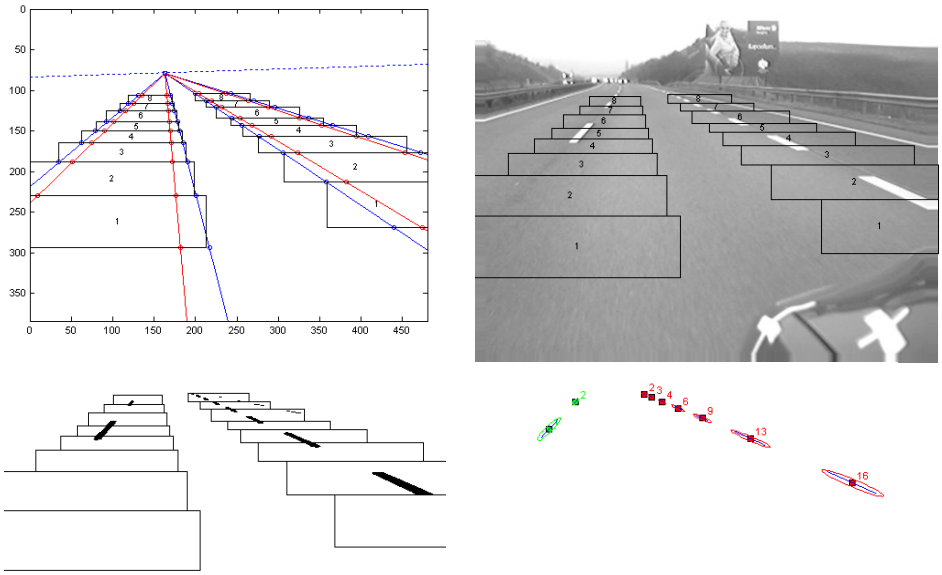

Fig. 3. Intermediate results of automatic lane marking extraction for the image of the left camera. The ROI and expected stripe width computation (top-left), ROI boxes over the input image (top-right), segmentation results (bottom-left), patch analysis results (bottom-right).

The first part of the algorithm is performed independently and simultaneously in the two images. Intermediate results are shown in Figure 3. Lane marking segmentation consists of a lane marking enhancement step and a binarization step. Binarization uses automatic threshold computation. This is followed by a patch analysis stage where several properties of the identified patches are determined, for example, patch area and eccentricity of the ellipse with equivalent second central moments (the ellipse is only a means to measure

elongatedness of the patches). These properties are then used to filter the detected patches: small and non-elongated patches are removed. Next, the ridge of each patch is determined along scanlines. After this step, the identified lane markings are represented as chains of points (primitives). If two primitives overlap in the horizontal direction, the external one is removed in order to avoid ambiguities at feature matching and to avoid the detection of the boundary of a potential exit lane (slip road) next to the current lane. The points grouped into these chains are radially undistorted before proceeding.

2.2 Lane reconstruction

The second part of the algorithm uses stereo information. The primitives are matched by cycling through the points of each primitive belonging to either boundary of the lane in the left image and then by computing the intersection of the corresponding epipolar lines with the primitives belonging to the same boundary in the right image. Presently, we describe the primitives as polylines. An alternative solution is to fit low-order models (e.g. line or polynomial curve segments) to the ridge points for each lane marking, and compute the intersection of the epipolar lines with these models. The matched points identifying lane boundary sections detected in both images are then triangulated into 3-space. The reconstructed 3D points belonging to the lane's boundary are used to find the *road*'s surface. The road surface model used is a second-order polynomial in z, which is the longitudinal coordinate of the vehicle reference frame and linear in the lateral coordinate x. The x-axis of the vehicle reference frame is defined to point from the left to the right. As we use right-handed reference frames, the remaining axis y points upwards. The road surface model explicitely contains the vehicle's roll angle φ and pitch angle ϑ measured with respect to the road, the vertical road curvature c_V and the vertical distance h between the road's surface and the origin of the vehicle's reference frame (the height coordinate):

$$y(x,z) = h + \varphi x + \vartheta z + \frac{c_v}{2} z^2 . \tag{1}$$

This model is currently fitted to the triangulated 3D boundary-points in the least-squares (LS) sense. An alternative method would be to use a robust fitting method, e.g. RANSAC (Fischler & Bolles, 1981).

Mono systems usually rely on the assumption $y(x,z) = 0$, i.e. the vehicle motion is simplified to a planar motion, the road's surface is modeled as a constant plane in the vehicle's reference frame or equivalently, vehicle pitch and roll angle, as well as, the vertical curvature of the road is neglected. This may cause instabilities in the next step, when a lane model is fitted to the back-projected feature points as depicted in Figure 4.

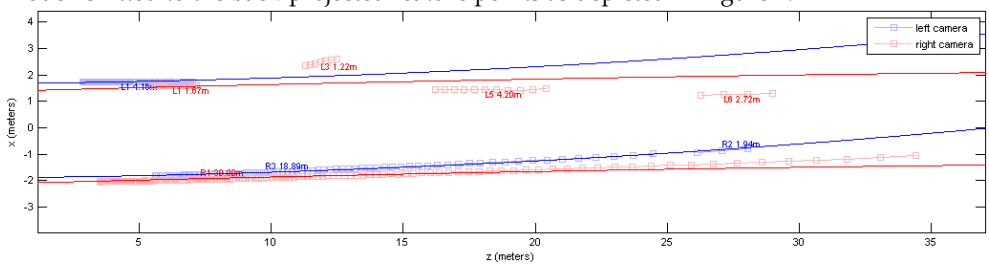

Fig. 4. An example of divergence present at lane model fitting when the left and right cameras are not interpreted as a stereo pair and instead the $y(x,z) = 0$ assumption is used.

In Figure 4., the point chains – shown as connected squares - represent the identified and reprojected lane markings while the continous double-curves represent the fitted polynomial lane model. The left and right-side chains converge due to an unmodeled pitching. It should be noted that some mono systems are able to estimate vehicle pitching by optimizing the pitch angle until the reprojected primitives become parallel. In Figure 4 at the left-side boundary, an outlier segment is present that is resulted from an imperfect segmentation.

In our stereo approach, as soon as the road's surface is found, all the detected points are projected onto it, including those that were ruled out at stereo matching. Then a double-polynomial lane boundary model is fitted to the 3D data as shown in Figure 5. Some of the roads are designed using constantly varying curvature (e.g. in Europe) while others include straight and circular segments (e.g. in the United States). Corresponding to a constantly varying curvature, clothoid lane models may be used (Nedevschi et al., 2005, Kastrinaki et al., 2003) (European case), but we experienced that the LS-fitting is relatively unstable when the polynomial order is higher than 2. Again, a robust model fitting method such as the RANSAC could be used to account for some outliers and make the detection more stable. Figure 6 shows some outputs of the discussed lane geometry reconstruction algorithm.

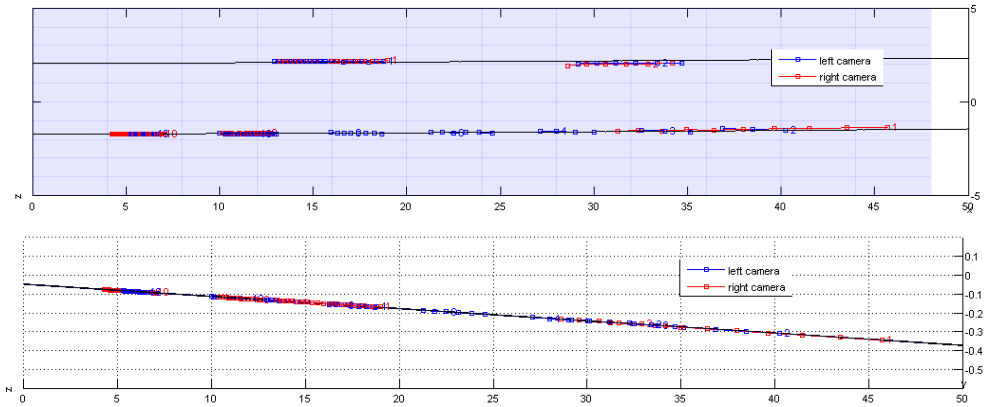

Fig. 5. Example of polynomial lane model fitting that followed the road surface detection based on stereo data. The image shown in Figure 3 and its right pair were used as input. There is a slight vehicle pitching (note the different scales on the axes) as shown in the side view of the road surface (bottom) but the lane profile (top) shows that the reprojected primitives are parallel.

2.3 On reconstruction errors

Reconstruction errors can have multiple sources. An imperfect segmentation causing outliers can disturb stereo feature matching, road surface model fitting and horizontal lane profile model fitting, the effect on the latter two being more severe. In most of the cases, the effects caused by outliers can be moderated or even eliminated by using robust algorithms at the model fitting stage. Ambiguities at feature matching are rare in the discussed lane boundary detection algorithm since horizontally overlapping boundary segments are removed as a prevention (the epipolar lines are almost horizontal). However, feature matching becomes a difficult problem when the lane detection algorithm is extended by vehicle, obstacle or guard rail detection. In such cases, area-based matching techinques are

popular and ambiguities may occur at repeating patterns. Also, imprecisions in feature matching cause errors in 3D reconstruction, especially if the 3D point is far from the vehicle. As mentioned earlier, the ROI computation, feature matching and 3D reconstruction require the knowledge of the camera parameters (see Figure 2) that are determined by calibration. Any imprecision in the camera parameters may jeopardize the whole procedure. In such a case, the computed epipolar lines do not exactly pass through the corresponding points at point feature matching. Therefore, in case of area-based matching, the correlation threshold may not be reached along an epipolar line and the point pair to match may be rejected. Globally, this results a decreased number of reconstructed feature points per snapshot. This may cause, e.g. missed obstacles if obstacles are searched based on a vicinity criterion of the reconstructed points. In the meantime, the mentioned threshold should be kept as high as possible to avoid false matches. Even if matches are accepted, their localization may be imprecise which, together with the imprecisely known camera parameters can cause signficant errors in the reconstruction by triangulation. Considering further processing stages, high reconstruction errors can affect the model fitting stage seriously (similarily to the case shown in Figure 4). Therefore, extra care is required at camera calibration.

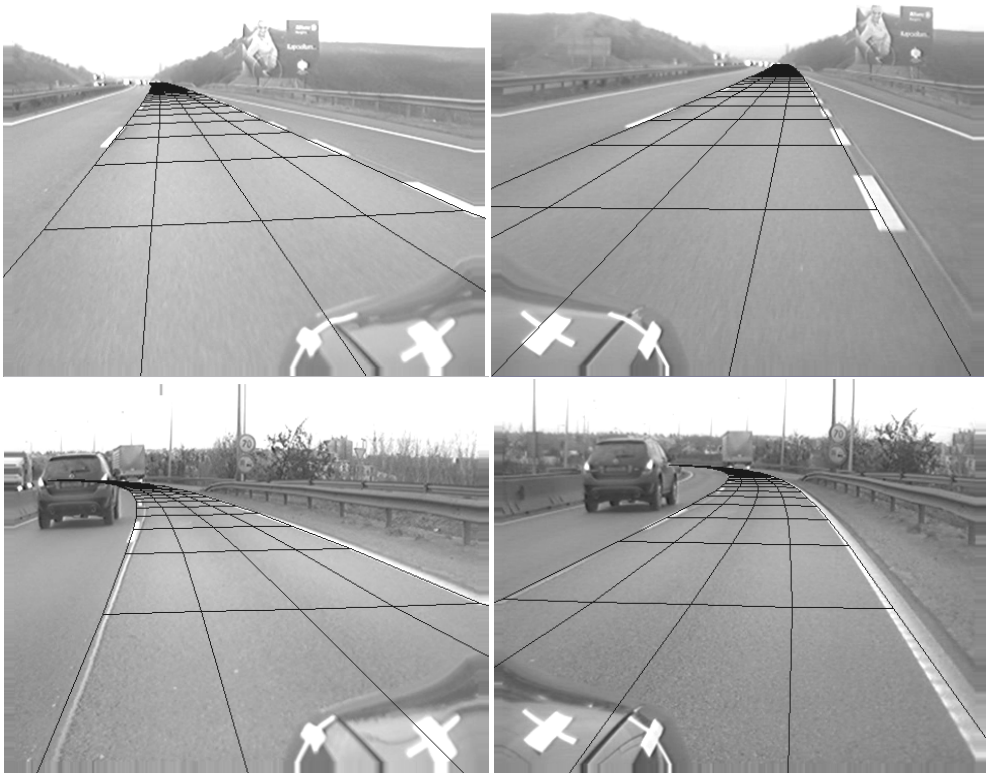

Fig. 6. The result of the discussed stereo reconstruction algorithm. The reconstructed 3D lane geometry is reprojected to the source images.

3. Camera calibration preliminaries

There are several common ways to calibrate a stereo rig. For example, it is possible to compute the fundamental matrix $\mathbf{F}$ from point correspondences (e.g. by using the well known normalized 8-point algorithm or the 7-point algorithm) without any knowledge of the scene or motion. It has been shown that the reconstruction based on this information only is possible up to a projective transformation (Hartley & Zisserman, 2006). Moreover, the camera matrices determine $\mathbf{F}$ up to scale, but not vice-versa. If additional knowledge either of the motion or of the scene's geometry (e.g. parallel scene lines or planes) an affine or a metric reconstruction may be reached. This is still not enough to achieve a "ground truth" (true Euclidean) lane reconstruction. Moreover, the reconstruction algorithm should not rely on such scene constraints as these might not always be available in a road scene. We can state that exact information on the vehicle's reference frame is required in such applications. $\mathbf{F}$ does not provide information on the 3D Euclidean reference frame but camera models do (Hartley & Zisserman, 2006). Consequently, we need to determine the two camera models first, only then can we proceed with computing $\mathbf{F}$ and the epipolar lines for stereo matching. In general form, the model of a single camera relating a 3D world point $\mathbf{W} \in R^3$ (given in metrical coordinates) to its 2D image denoted by $\mathbf{I} \in R^2$ is as follows:

$$\mathbf{I} = \Phi\,(\mathbf{p}_{in}, \mathbf{p}_{ex}, \mathbf{W}), \tag{2}$$

where $\mathbf{p}_{in}$ is a vector formed of the intrinsic camera parameters and $\mathbf{p}_{ex}$ is the 6-vector of the extrinsic camera parameters, the latter representing the 6-Degrees-of-Freedom (DoF) Euclidean transformation (i.e., a 3D rotation and a 3D translation) between camera and world reference frame. The mapping Φ may model certain non-linear effects (e.g., radial and tangential lens distortions), as well. These distortions can be - in certain cases - neglected, or can be removed from Φ by an adequate non-linear image warping step. The remaining Φ_{lin} is linear in a homogeneous representation:

$$\tilde{\mathbf{I}} = \mathbf{KR}[\,\mathbf{E}_{3x3} \,|\, -\mathbf{t}]\tilde{\mathbf{W}} \qquad \mathbf{K} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where $\mathbf{E}_{3x3}$ is the identity matrix, $\tilde{\mathbf{I}} \in P^2$ and $\tilde{\mathbf{W}} \in P^3$ are homogeneous representations of the points and $\mathbf{K}$ is the camera calibration matrix incorporating the relative focal lengths α and β, the skew γ and the principal point $(u_0, v_0)^T$. These are all intrinsic parameters contained in $\mathbf{p}_{in}$. $\mathbf{R}$ is the 3-DoF rotation matrix, and $\mathbf{t}$ is the camera position. $\mathbf{R}$ and $\mathbf{t}$ correspond to $\mathbf{p}_{ex}$.

Calibration of a monocular camera consists of determining the camera parameters $\mathbf{p}_{in}$ and $\mathbf{p}_{ex}$ from properly measured $\mathbf{W}_i \mapsto \mathbf{I}_i$ point correspondences. We note here that for the intrinsic parameters, novel methods tend to use planar calibration patterns instead of 3D calibration objects (Malm & Heyden, 2003; Zhang, 2000; Sturm & Maybank 1999). A single planar arrangement does not provide enough information for estimating all the intrinsic paremeters, however, a solution for $\gamma = 0$ and known aspect ratio β / α exists (Tsai, 1987). Algorithms for calibrating from planar patterns are developed for the stereo case, as well

(Malm & Heyden, 2001), but those determine the relative pose between the cameras while the absolute poses are required in our application.

As to lane-related applications, Bellino et al. have used a single planar pattern and the aforementioned method for calibrating a monocular system used in heavy vehicles (Bellino et al, 2005). The authors used a fixed principal point and did not involve γ. They have investigated the reconstruction errors of two target points on the ground plane vs. the tilt angle of a calibration plane given with respect to the vertical position. The target points were placed up to 11.5 m from the camera and their positions were measured with a laser-based meter for validation. They found that the inclination of the plane had considerable effect on the quality of the result.

A more general intrinsic parameter estimation is given in (Hartley & Zisserman, 2006) for multiple planes with unknown orientations. Homographies (i.e., perspective 2D-to-2D mappings) $\mathbf{H}_j$ between the planes and the image are estimated first. Each $\mathbf{H}_j$ gives rise to two constraints on the image of the absolute conic (IAC) $\boldsymbol{\omega}$. The IAC is formulated as $\boldsymbol{\omega} = (\mathbf{K}\mathbf{K}^T)^{-1}$. The constraints are linear in the elements of $\boldsymbol{\omega}$, namely $\mathbf{h}_1^T\boldsymbol{\omega}\mathbf{h}_2 = 0$ and $\mathbf{h}_1^T\boldsymbol{\omega}\mathbf{h}_1 = \mathbf{h}_2^T\boldsymbol{\omega}\mathbf{h}_2$, where $\mathbf{h}_i$ is the i-th column of $\mathbf{H}_j$. Since $\boldsymbol{\omega}$ is the homogeneous representation of a conic, it has 5 DoF, so 3 planes suffice to estimate $\boldsymbol{\omega}$. $\mathbf{K}$ can be computed from $\boldsymbol{\omega}^{-1}$ by Cholesky-factorization (Hartley & Zisserman, 2006), or by using direct non-linear formulas (Zhang, 2000). In the literature, some comparisons can be found between the method of Tsai, Zhang (Sun & Cooperstock, 2005; Zollner & Sablatnig, 2004). It turns out, that Zhang's model and his method overperforms the others with respect to residual errors and convergence. The price payed is the relatively high number of iterations, which is not a serious problem in case of off-line applications.

Having determined the intrinsic camera parameters, the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ need to be recovered. This is called pose estimation and a single general planar arrangement with at least 4 points suffices for a unique solution (Lepetit & Fua, 2005). A rough pose estimation can be performed by first estimating the homography between the world plane and the image and then by re-using the orthogonality constraints with known camera calibration matrix $\mathbf{K}$ (Malm & Heyden, 2003; Lepetit & Fua, 2005).

The planar pattern used in intrinsic calibration is unsuitable for far-range systems, because it minimizes errors for close-range (Marita et al., 2006; Bellino et al., 2005). For this purpose, Broggi et al. used a medium-range grid painted on the ground (Broggi et al., 2005) while Marita et al. used vertical X-shaped markers placed on the ground in front of the vehicle in a distance up to 45 m (Marita et al., 2006). In the latter work, the intrinsic and the extrinsic parameters of each camera are computed separately. Extrinsic parameters are computed by minimizing the reprojection error in the image for the available control points. A constrainted Gauss-Newton minimization is used with the constraints $\mathbf{R}^T\mathbf{R} = \mathbf{I}$. The calibration is validated by comparing the 3D reconstruction errors of the control points to the actual 3D measurements. However, there is no information available about the accuracy of the control point setup. Alternatively, line features can also be used for pose estimation (Kumar & Hanson, 1994).

With the road/lane following application in mind, and making use of the calibration methods used in computer vision, we present here a calibration scheme and method that is optimal under some reasonable assumptions. It should be emphasized that the proposed method takes into consideration the errors present in the 3D setup.

4. Calibration of a single camera

4.1 Camera model

We use a pinhole camera - extended with a fifth-order radial distortion model and a tunable distortion centre (Hartley & Zisserman, 2006) - as our camera model:

$$\mathbf{D} = \begin{pmatrix} x_D \\ y_D \end{pmatrix} = (1 + d_1 r^2 + d_2 r^4) \begin{pmatrix} x_P - c_x \\ y_P - c_y \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}, \tag{4}$$

where $(c_x, c_y)^T$ is the distortion center, d_1 and d_2 are the distortion coefficients, while $(x_P, y_P)^T$ is the point perspectively projected to the normalized image plane and $\mathbf{D}$ is the corresponding distorted point on this plane. Thus, the distortion model is applied between the projection step and the rasterization step, the latter being modeled with the homogenous transformation represented by the camera calibration matrix $\mathbf{K}$, like in equation (3). Lens distortion models involving tangential distortion are also available in the literature (Heikkila & Silvén, 1997). An explicit tangential distortion model is not required in our case, as the distortion center, with its two additional parameters, models imperfect alignment of the lenses and of the sensor. Therefore, we have, in total, the nine intrinsic parameters $\mathbf{p}_{in} = (\alpha, \beta, \gamma, u_0, v_0, d_1, d_2, c_x, c_y)^T$ for each camera.

4.2 Intrinsic calibration

Intrinsic calibration is performed separately for the two cameras. We used a hand-held checkerboard pattern shown in m different orientations to the camera. Alternatively, a pattern with circular patches could have been used (Heikkila & Silvén, 1997, Zollner & Sablatnig, 2004). Corners in the checkerboard pattern can be localized with sub-pixel accuracy with our corner detector based on a robust saddle-point search in small regions around the corners.

For an initial guess of $\mathbf{p}_{in}$, we used the method of Zhang, however we extended it with some optional constraints: $\omega_{12} = \omega_{21}$ (for $\gamma = 0$), $\omega_{11} = \omega_{22}\beta^2\alpha^{-2}$ (for a fixed aspect ratio, e.g. square pixels), and finally $\omega_{13} = -u_0\omega_{11}$ and $\omega_{23} = -v_0\omega_{22}$ (for a fixed principal point). The problem is usually overdetermined because of the great number of the control points available. Having carried out a homography estimation for each image, the solution for $\boldsymbol{\omega}$ is obtained with a simple SVD-based method that gives a least-squares solution while exactly fulfilling the constraints. $\mathbf{K}$ is determined using the formulae arising from $\boldsymbol{\omega}^{-1} = \mathbf{K}\mathbf{K}^T$.

Next, for each of the m views, the 6 extrinsic parameters $\mathbf{p}^*_{ex,j}$ ($j = 1, 2 ... m$) are determined, that is, the planar pose estimation problem is solved for each view. For this, we used the orthogonality constraints satisfied by $\mathbf{K}^{-1}\mathbf{H}_j$, where $\mathbf{H}_j$ denote the homography matrix estimated for the j-th view. Since measurements are noisy and orthogonality is not exactly satisfied in practice, some tolerances were used in the orthogonality test. In the representation of the rotation, we used the Rodrigues-vectors $\mathbf{r}_j = \varphi_j \mathbf{a}_j$, where $\mathbf{a}_j$ represents the rotation axis and $\varphi_j = \| \mathbf{r}_j \|$ is the rotation angle. This has several advantages over the rotation matrix-based representation (Lepetit & Fua, 2005).

As the set of parameters $\hat{\mathbf{p}}_{in} \in R^9$ and $\hat{\mathbf{p}}^*_{ex} = (\hat{\mathbf{p}}^{*T}_{ex,1}, ..., \hat{\mathbf{p}}^{*T}_{ex,m})^T \in R^{6m}$ determined earlier minimize an algebraic error, a refinement of the parameters is preferable by minimizing a

geometrically meaningful error. A reasonable choice is to minimize the sum of the reprojection errors in all the m images for all the n feature points:

$$f(\mathbf{p}_{in}, \mathbf{p}^*_{ex}) = \sum_{i,j} d^2(\mathbf{I}^*_{ij}, \hat{\mathbf{I}}^*_{ij}) = \sum_{i,j} \| \mathbf{I}^*_{ij} - \hat{\mathbf{I}}^*_{ij} \|_2^2 = \| \mathbf{I}^* - \hat{\mathbf{I}}^* \|_2^2 , \tag{5}$$

where $\mathbf{I}^*_{ij}$ is the measured location of the i-th corner in the j-th image and $\hat{\mathbf{I}}^*_{ij}$ is the reprojection of the world point $\mathbf{W}^*_{ij}$ using the parametrized camera model (2). $\mathbf{I}^*$ is the measurement vector containing all the $\mathbf{I}^*_{ij}$'s and $\hat{\mathbf{I}}^*$ contains all the $\hat{\mathbf{I}}^*_{ij}$'s. d denotes Euclidean distance in the pixel reference frame. The cost function (5) can be minimized using a gradient-based iterative method. It can be shown, that (5) is a Maximum Likelihood (ML) cost function, provided the checkerboard pattern is precise, significant errors are due to the corner localization, and the errors have uniform Gaussian distribution all over the images. These assumptions make it possible to estimate the deviation σ of the detection noise in the image simply as the standard deviation of the residual errors $(\mathbf{I}^*_{ij} - \hat{\mathbf{I}}^*_{ij})$ evaluated at the optimum. In order to determine the quality of the calibration, the estimated noise deviation $\hat{\sigma}$ is then back-propagated to the camera parameters through a linearized variant of the camera model (2).

4.3 Optimal pose estimation per view

The relative orientations of the views with respect to the imaged checkerboards (incorporated in $\hat{\mathbf{p}}^*_{ex}$) are of no interest for us from the point of view of our application. Instead, for both cameras, we need an estimate of the pose $\hat{\mathbf{p}}_{ex}$ with respect to the vehicle's reference frame. To determine the poses, the vehicle with the mounted cameras is stopped over an open flat area and marker plates - with an X-shape on each - are placed in front of the vehicle (see Figure 7). Similar arrangements has already been used (Marita et al., 2006).

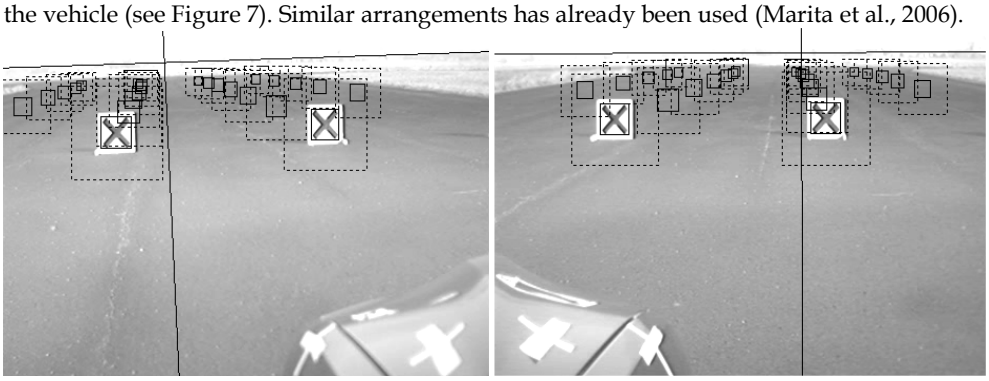

Fig. 7. An image pair of the calibration scene with overlayed marker detection and calibration results. The dashed boxes are the search regions for template matching, while the solid boxes correspond to the detected markers. The solid lines represent the vanishing lines of the world reference planes.

The feature points are the centres-of-gravity (CoG's) of the individual X-shapes. The 3D marker locations $\mathbf{W}_i$ were measured by a laser-based distance meter from two reference

points. The locations of the reference points are measured with respect to the vehicle. $\mathbf{I}_i$'s are extracted from the images using normalized cross-correlation with an ideal X-shaped template. We derive the pose estimation formula for one camera first and based on that we derive it for the stereo case. The initialization of the algorithm is done with the pose estimation method based on the orthogonality criteria discussed in Section 4.2. Therefore, an initial guess is available for $\mathbf{p}_{ex}$, and this is to be fine-tuned by using a geometrically meaningful expression. The measurement of the 3D point locations and the detection of their image are two independent measurements modeled with two independent Gaussian distributions. If we have N markers and we introduce the vector of all the measured 3D coordinates $\mathbf{W} := (\mathbf{W}_1^T,...,\mathbf{W}_N^T)^T$, and the vector of all the measured pixel coordinates $\mathbf{I} := (\mathbf{I}_1^T,...,\mathbf{I}_N^T)^T$, then the likelihood functions for $\mathbf{W}$ and $\mathbf{I}$ are:

$$L(\mathbf{W} \mid \bar{\mathbf{W}}) = \frac{1}{\left(\sqrt{2\pi}\right)^{3N}\sqrt{\det(\mathbf{C}_W)}} \exp\left\{-\frac{1}{2} \| \mathbf{W} - \bar{\mathbf{W}} \|^2_{C_W}\right\}, \tag{6}$$

$$L(\mathbf{I} \mid \bar{\mathbf{I}}) = \frac{1}{\left(\sqrt{2\pi}\right)^{2N}\sqrt{\det(\mathbf{C}_I)}} \exp\left\{-\frac{1}{2} \| \mathbf{I} - \bar{\mathbf{I}} \|^2_{C_I}\right\}. \tag{7}$$

Here, $\| \mathbf{a} - \mathbf{b} \|^2_C = (\mathbf{a} - \mathbf{b})^T \mathbf{C}^{-1} (\mathbf{a} - \mathbf{b})$ denotes the squared Mahalanobis distance between the vectors $\mathbf{a}$ and $\mathbf{b}$ with respect to the covariance matrix $\mathbf{C}$. In equations ($\mathbf{C}_I$ and $\mathbf{C}_W$ denote the covariance matrices of the measurement vectors $\mathbf{I}$ and $\mathbf{W}$, respectively. The likelihood function of all the measurements is the product of $L(\mathbf{I} \mid \bar{\mathbf{I}})$ and $L(\mathbf{W} \mid \bar{\mathbf{W}})$ because of the independence. Therefore, the Maximum Likelihood Estimate (MLE) for the expected 2D and 3D locations $\bar{\mathbf{I}}$ and $\bar{\mathbf{W}}$ of the feature points can be found by minimizing the function

$$g(\bar{\mathbf{I}}, \bar{\mathbf{W}}) = \| \mathbf{I} - \bar{\mathbf{I}} \|^2_{C_I} + \| \mathbf{W} - \bar{\mathbf{W}} \|^2_{C_W}. \tag{8}$$

$\bar{\mathbf{I}}$ and $\bar{\mathbf{W}}$ are related by the camera model (2), so $\bar{\mathbf{I}} = \Phi(\mathbf{p}_{in}, \mathbf{p}_{ex}, \bar{\mathbf{W}})$. Both $\mathbf{C}_I$ and $\mathbf{C}_W$ are block-diagonal provided the measurement of the control points are independent from each other. The blocks in the diagonals are $\mathbf{C}_{Ii}$ (the 2x2 covariance matrices of each markers' localization errors in the image) and $\mathbf{C}_{Wi}$ (the 3x3 covariance matrices of the markers' localization errors in 3-space), respectively. Thus, (8) can be rewritten as

$$g(\mathbf{p}_{ex}, \bar{\mathbf{W}}) = \sum_{i=1}^{N} \left\{ \| \mathbf{I}_i - \Phi(\mathbf{p}_{in}, \mathbf{p}_{ex}, \bar{\mathbf{W}}_i) \|^2_{C_{Ii}} + \| \mathbf{W}_i - \bar{\mathbf{W}}_i \|^2_{C_{Wi}} \right\}. \tag{9}$$

Several important conclusions can be drawn from equation (9). The first is that the covariance matrices of the measurements are involved, so this ML cost function can not be used when no uncertainty information is available of the measurements. In such cases the cost function (5) could be used for pose estimation, as well. The second one is that the expected 3D locations represented by $\bar{\mathbf{W}}_i$ are involved, therefore not only the searched 6 extrinsic parameters in $\mathbf{p}_{ex}$ are to be optimized, but the marker locations of the 3D calibration arrangement, as well. This adds extra 3N dimensions to the parameter space. Optionally, the 3N extra dimensions can be eliminated by approximating $\bar{\mathbf{W}}_i$ with the

closest point to the measured 3D point $\mathbf{W}_i$ on the ray through the radially corrected image point $\mathbf{I}_i$.

5. Optimal stereo calibration and sensitivity analysis

5.1 Optimal two-view calibration and pose estimation

Up to this point we considered the two cameras independently. Clearly, it is possible to determine the extrinsic parameters of the left ($\mathbf{p}_{exL}$) and right cameras ($\mathbf{p}_{exR}$) independently (Marita et al., 2006). However, this model does not take into consideration that the 3D control point setup is common for the two views. Also, equation (9) requires the knowledge of the exact intrinsic parameters $\mathbf{p}_{in}$, however, only its estimation $\hat{\mathbf{p}}_{in}$ is available from intrinsic calibration. These problems can be solved by formulating the MLE for the overall problem that involves all the measurements including those available from checkerboard-based calibration and for both cameras. As a result, the cost function to minimize becomes slightly more complex:

$$\begin{aligned} h(\mathbf{p}_{inL},\mathbf{p}_{exL},\mathbf{p}_{inR},\mathbf{p}_{exR},\mathbf{p}^*_{exL},\mathbf{p}^*_{exR},\bar{\mathbf{W}}) &= \frac{1}{\sigma_L^2}\|\mathbf{I}^*_L-\bar{\mathbf{I}}^*_L(\mathbf{p}_{inL},\mathbf{p}^*_{exL})\|_2^2+\frac{1}{\sigma_R^2}\|\mathbf{I}^*_R-\bar{\mathbf{I}}^*_R(\mathbf{p}_{inR},\mathbf{p}^*_{exR})\|_2^2+ \\ &+\|\mathbf{W}-\bar{\mathbf{W}}\|^2_{C_W}+\|\mathbf{I}_L-\bar{\mathbf{I}}_L(\mathbf{p}_{inL},\mathbf{p}_{exL},\bar{\mathbf{W}})\|^2_{C_{IL}}+\|\mathbf{I}_R-\bar{\mathbf{I}}_R(\mathbf{p}_{inR},\mathbf{p}_{exR},\bar{\mathbf{W}})\|^2_{C_{IR}} \end{aligned}, \quad (10)$$

where the first two terms come from the intrinsic calibration performed independently for the left and right camera (see the right side of equation (5)), the last two terms represent the localization errors in the images in the X-marker-based calibration and the third term represents the errors in the world point locations in the X-marker-based pose estimation. Each measurement influences the solution weighted with the inverse of its uncertainty. The outline of the proposed algorithm is as follows.

1. Perform an intrinsic calibration independently for the two views by using planar patterns and by minimizing the cost function (5). Compute $\hat{\sigma}_L$ and $\hat{\sigma}_R$.
2. Set up a far-range planar arrangement of visible control points for stereo pose estimation (as suggested by Figure 8). Locate the markers and estimate the measurement covariances both for the images and for the 3D arrangement.
3. Initialize the yet unknown extrinsic parameters $\mathbf{p}_{exL}$ and $\mathbf{p}_{exR}$ with respect to the road by solving the pose estimation problem based on the orthogonality criteria independently for the two views. Initialize $\bar{\mathbf{W}}$ with the measured 3D locations and minimize cost function (10).

5.2 Sensitivity of the camera parameters

Supposing that the optimal solution of (10) has been found (this can be checked with a residual analysis together with a linearity test of the cost function h), the overall quality of the calibration can be characterized by performing a sensitivity analysis,. First of all, the the cost function (10) can be approximated as

$$h(\mathbf{q}) = \|(\mathbf{m}-\bar{\mathbf{m}})-(\hat{\mathbf{m}}-\bar{\mathbf{m}})\|^2_{C_m} \approx \|(\mathbf{m}-\bar{\mathbf{m}})-\mathbf{J}_m(\mathbf{q}-\bar{\mathbf{q}})\|^2_{C_m}, \quad (11)$$

where $\mathbf{q}$ is the vector of all the parameters to optimize, $\mathbf{m}$ is the vector of all the measurements $\mathbf{I}_L, \mathbf{I}_R, \mathbf{I}^*_L, \mathbf{I}^*_R$ and $\mathbf{W}$. $\mathbf{C}_m$ is the block-diagonal covariance matrix of all these

measurements. $\hat{\mathbf{m}}$ contains, on the one hand, the image points $\hat{\mathbf{I}}_L, \hat{\mathbf{I}}_R, \hat{\mathbf{I}}_L^*, \hat{\mathbf{I}}_R^*$ reprojected using the camera model, on the other hand, the optimized world point coordinates $\hat{\mathbf{W}}$. $\bar{\mathbf{m}}$ represents the "ground thruth" values of the parameters that are always unknown and $\mathbf{J}_m$ is the analytically computed Jacobian matrix of the mapping $\mathbf{q} \mapsto \hat{\mathbf{m}}$ evaluated at the optimal parameter vector $\hat{\mathbf{q}} \approx \bar{\mathbf{q}}$.

The measurement uncertainty incorporated in $\mathbf{C}_m$ can be back-propagated to the parameters, as $\mathbf{C}_q = (\mathbf{J}_m^T \mathbf{C}_m^{-1} \mathbf{J}_m)^{-1}$. Since we are primarily interested in the uncertainty of the intrinsic parameters $\mathbf{p}_{inL}$ and $\mathbf{p}_{inR}$ and of the poses $\mathbf{p}_{exL}$ and $\mathbf{p}_{exR}$, the corresponding two 30x30 sub-matrix should be extracted from $\mathbf{C}_q$. It is important that nothing prevents the parameters of the two cameras to cross-correlate. Therefore, at the uncertainty estimation of any computation involving the parameters of both cameras, the 30x30 covariance matrix has to be considered instead of the two 15x15 blocks corresponding to the two cameras, independently.

5.3 Uncertainty of the epipolar lines

In the feature matching stage of the lane detection algorithm, we compute the intersection of epipolar lines and 2D primitives (polylines or curves). It is well known that the fundamental matrix required for epipolar line computation can be derived from the camera parameters (Hartley & Zisserman, 2006). We use the formulation

$$\mathbf{F} = \mathbf{K}_R^{-T} \mathbf{R}_R \mathbf{R}_L^T [\mathbf{R}_L (\mathbf{t}_R - \mathbf{t}_L)]_\times \mathbf{K}_L^{-1}, \tag{12}$$

where the indices L and R refer to the left camera and right camera, respectively. $[\cdot]_\times$ denotes the 3x3 matrix of rank 2 corresponding to the cross product operator, so that $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b}$. Then the epipolar line $\mathbf{l}_R$ in the right image corresponding to a point $\tilde{\mathbf{I}}_L$ in the left image can be computed as $\mathbf{l}_R = \mathbf{F}\tilde{\mathbf{I}}_L$, where both $\mathbf{l}_R$ and $\tilde{\mathbf{I}}_L$ are homogeneous 3-vectors. The epipolar constraint $\mathbf{l}_R^T \tilde{\mathbf{I}}_R = 0$ assures that the line $\mathbf{l}_R$ passes through the point $\tilde{\mathbf{I}}_R$, which is the right image of the same world point, as shown in the left side of Figure 8.

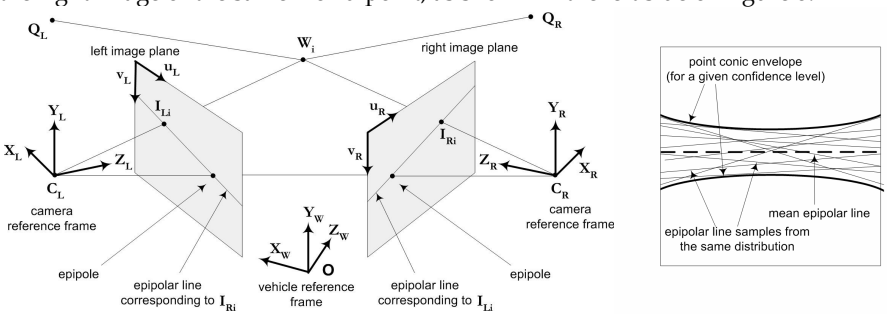

Fig. 8. Epipolar geometry of cameras in general configuration (left). The 3D point $\mathbf{W}_i$ given in the vehicle reference frame is "seen" as $\mathbf{I}_{Li}$ and $\mathbf{I}_{Ri}$ in the pixel reference frames. The nature of epipolar line uncertainty (right).

Uncertainty in the camera parameters propagates to the derived fundamental matrix and to the epipolar lines. If the vector of the involved camera parameters is denoted by $\mathbf{p}$, then the formula (12) can be interpreted as a mapping $\mathbf{p} \mapsto \mathbf{f}$ where $\mathbf{f}$ is a vector formed from the elements of $\mathbf{F}$. If this mapping can be approximated by a linear mapping in the range of the noise, and Gaussian distributions are supposed just like in earlier derivations, then the forward propagation of covariance can be given as

$$\mathbf{C}_f = \mathbf{J}_f \mathbf{C}_p \mathbf{J}_f^T , \tag{13}$$

where $\mathbf{C}_p$ is the covariance matrix of $\mathbf{p}$, $\mathbf{C}_f$ is a covariance matrix associated to $\mathbf{F}$ and $\mathbf{J}_f$ is the Jacobian matrix of the mapping $\mathbf{p} \mapsto \mathbf{f}$ evaluated at the mean values $\hat{\mathbf{p}} \approx \bar{\mathbf{p}}$ and $\hat{\mathbf{f}} \approx \bar{\mathbf{f}}$ (hat over the letter denotes estimated values and bar denotes the unknown ground thruth). $\mathbf{J}_f$ is either available analytically as the partial derivatives of (12) with respect to the camera parameters or it can be approximated numerically. To avoid ambiguities, $\mathbf{F}$ is always normalized so that $\|\mathbf{F}\| = 1$.

Next, the uncertainty present in the fundamental matrix can be forward-propagated to the epipolar lines very similarily, by linearizing the mapping $\mathbf{f} \mapsto \mathbf{l}_R / \|\mathbf{l}_R\|$. It is known that the set of epipolar line samples corresponding to a given confidence level form a line conic that is bounded by a point conic envelope of 5 DoF (Hartley & Zisserman, 2006). This is illustrated in the right side of Figure 8. The point conic can be analytically derived if the covariance matrix $\mathbf{C}_{lR}$ of the epipolar line $\mathbf{l}_R$ is known. The envelope conic for a given confidence level λ (99% for example) characterises very well the sensitivity of an epipolar line. However, it is more practical to use the distribution of the angle between the epipolar line and the u-axis of the image reference frame to characterize uncertainty. If $\mathbf{l}_R = (a, b, c)^T$, this angle can be expressed as

$$\Theta = \arctan(-a / b) . \tag{13}$$

Moreover, the mapping $\mathbf{l}_R \mapsto \Theta$ should also be linearized in order to propagate the covariance of the epipolar lines to Θ. As a result, the complete chain of the uncertainty propagation from the camera parameters to the angle Θ is $\mathbf{p} \mapsto \mathbf{F} \mapsto \mathbf{l}_R \mapsto \Theta$.

A practical verification of the uncertainty computations is done by Monte-Carlo simulations. A high number of samples $\{\mathbf{p}_k\}$ were generated of $\mathbf{p}$ with mean the estimated camera parameters $\hat{\mathbf{p}}$ and covariance matrix $\mathbf{C}_p$ that is extracted from the computed $\mathbf{C}_q$ matrix as described in Section 5.2. Then by selecting a point $\mathbf{I}_L$ in the left image, the corresponding epipolar line and Θ_k is computed for each sample $\mathbf{p}_k$. Finally, the standard deviation of the set $\{\Theta_k\}$ is calculated. This way, to any point $\mathbf{I}_L$ in the left image, a single value is associated that characterizes the uncertainty of the corresponding epipolar line. If the output of the computationally expensive Monte-Carlo approach coincides with the results received from the discussed cheaper linear approximation method, then the non-linear $\mathbf{p} \mapsto \mathbf{F} \mapsto \mathbf{l}_R \mapsto \Theta$ mapping is nearly linear in the range of the uncertainities around $\hat{\mathbf{p}}$.

5.4 Uncertainty in the reconstruction

The main reason why camera parameter uncertainties are studied is to predict reconstruction errors due to an imprecise knowledge of the camera parameters. The errors

in stereo point reconstruction can be simulated with a Monte-Carlo method that is very similar to the one discussed in Section 5.3. The estimated camera parameters are perturbed corresponding to the estimated parameter covariance matrix and for each parameter set, several 3D points are reconstructed. As a result, point clouds are formed in 3-space that correspond to reconstruction errors. Similarily, one can go further, and apply further steps of those detailed in Section 2, e.g. road surface model fitting and lane model fitting for each generated set of parameters. Some results based on real data are presented in Section 6.

6. Evaluation on real images

6.1 Numerical results of the two-step camera calibration method

A setup with two analog 1/3" b&w CCD cameras with a resolution of 720x576 pixels and 8 mm lenses with 34° horizontal field-of-view were mounted on the side mirrors of a test vehicle. For various reasons, the acquired images were resized to a size of 480x384 pixels when the recorded videos were post-processed in order to remove interlacing effects. For intrinsic calibration, we used a checkerboard pattern with 11x7 control points and a square size of 3 cm. Images were taken in 16 different views (which is much more than required) and a constrainted minimization with $\gamma = 0$ has been carried out by using the Levenberg-Marquardt method. As a result, the estimated focal lengths are (777.6, 849.8) ± (3.3, 4.1) pixels for the left camera, and (776.1, 847.2) ± (3.0, 3.5) pixels for the right, the principal point is located at (215.7, 201.9) ± (10.4, 7.6) pixels in the image of the left camera and at (236.0, 168.7) ± (8.8, 6.9) pixels in that of the right one. The uncertainties given here correspond to the 99% confidence interval of the uncertainty in the parameter space that is calculated by back-propagating the standard deviation of the residual reprojection errors given in the image to the parameters, as mentoned in Section 4.2. The above uncertainties are only given for reference; clearly they do not completely characterise the uncertainty (cross-covariance information is also required). The standard deviations of the residual errors were 0.26 and 0.23 pixels for the left and right cameras, respectively. The distortion centers are located approximately at the principal points while the estimated radial distortion coefficients are (-0.505, 0.878) ± (0.044, 0.540) for the left camera and (-0.516, 0.957) ± (0.041, 0.540) for the right one. Note that the second coefficient is estimated with a relatively high uncertainty (alternatively, it could be forced to zero before the estimation).

In the second step, 24 X-markers of size 50x50 cm were placed in front of the vehicle along four lines at a distance of 3 meters laterally and in a depth range from 10 to 40 meters. Single rows of markers were placed at a time to have a clear view on each, that is, to avoid masking due to the perspective effect (see Figure 7). Marker distances from two reference points were measured with a laser-based distance meter and the 3D locations were computed by triangulation. Also, errors in the reference point locations, distance measurements, non-ideal marker-placing and deviations with respect to the planar ground assumption were estimated and forward-propagated to the computed 3D locations resulting in an estimate of each $\mathbf{C}_{Wi}$ $(i = 1, 2 ... N)$ (see Bodis et al., 2007, for more details). The resulted 99% covariance ellipsoids of the 3D measurements are plotted in Figure 9C for some markers. Using a quasi-Newton optimization with a termination criterion of 10^{-5} on the relative change in the parameter values, the minimization of (10) converged in 116 iterations. For comparison, we also minimized (5), like if there would be no covariance information. A comparison of the results provided by the two approaches is shown in Fig. 9A-B.

It is clear from the results (Figure 9A-B) that inappropriate modeling of the problem or the lack of covariance information misleads the optimization when the 3D marker locations are not known precisely. Naturally, the high reprojection errors shown in Figure 9A is undesirable because it is directly related to the corresponding 3D reconstruction errors.
Because camera skews were forced to be zero, we had 2x8 intrinsic parameters, 2x16x6 extrinsic parameters with respect to the checkerboards and 2x6 extrinsic parameters with respect to the road (or stopped vehicle). As to the measurements, we had 2x2x11x7 coordinates from the checkerboard corners, while 2x2x24 image coordinates and 3x24 world coordinates were available from the X-marker based measurement. In total, there were 220 parameters and 476 measurements. In Figure 9C, we can see that the covariance ellipsoids of the estimated 3D locations are much smaller than those of the measured locations. This is what we expected from a regression-like problem.

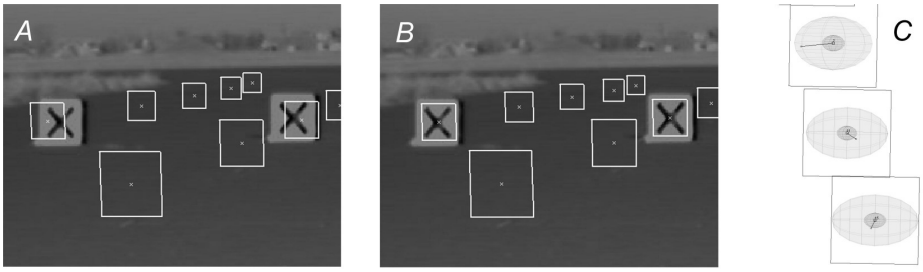

Fig. 9. A-B) Markers reprojected to one of the source images of the right camera using the camera parameters that resulted from the minimization of A) the cost function given in (5), B) the proposed cost function (10). C) 99% condence levels of the 3D location measurements for three 50x50 cm marker plates placed at 20, 25 and 35 m (outer ellipsoids) and 99% confidence level of the position estimate after optimization, from $\mathbf{C}_q$ (inner ellipsoids).

6.2 The sensitivity of point reconstruction

The evaluation of the calibration is performed by simulating the effects of parameter uncertainties represented by $\mathbf{C}_p$ on 3D point recunstruction by triangulation.

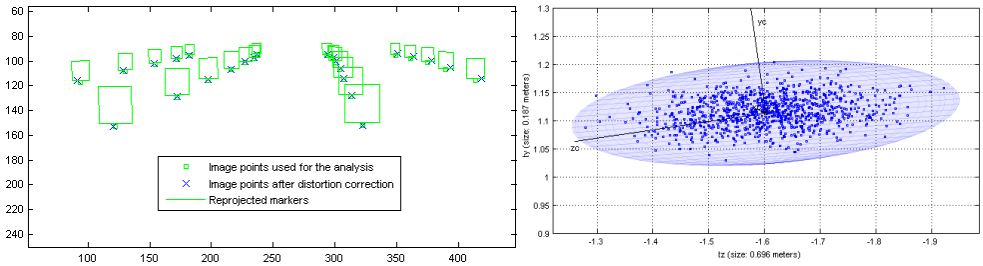

Fig. 10. The points to be reconstructed (triangulate) by Monte-Carlo simulations are the marker centers shifted vertically to the ground surface (left side). 99% covariance ellipsoid of the left camera's position and verification of the generated noise in parameter space, 1000 experiments (right side). 99.4% of the points fell inside the ellipsoid in this realization.

The reference points to reconstruct by Monte-Carlo simulations were chosen from the calibration scene: the estimated marker centers were shifted by 25 cm vertically to lie on the

ground surface. The left and right images of these 3D points were computed by using the set of the optimal camera parameter estimates (see the left side of Figure 10). Then, 1000 perturbed parameter sets were generated around the estimated parameters corresponding to $\mathbf{C}_p$. There are 2x14 estimated camera parameters that are to be considered here, zero skew being fixed, so $\mathbf{C}_p$ is a 28x28 matrix. A 3D subspace of the generated distribution is shown in the right side of Figure 10. This corresponds to the 99% confidence level of the 3D location of the left camera given in the vehicle's reference frame. The sizes of this ellipsoid are ±14 laterally (X), ±35 cm in depth (Z) and ±10 cm in the direction perpendicular to the ground (Y). The ellipsoid is elongated in the longitudinal (Z) direction because the estimation of the camera positions in this direction is more sensitive to the uncertainty in 3D marker locations of the planar and far-range calibration scene. Interestingly, this does not mean that any 3D point can be reconstructed with a maximum precision of ±35 cm, because reconstruction quality is affected by the covariances between all the 28 parameters, as well.
After that the points were radially corrected in both images by using the perturbed radial distortion parameters in each experiment, they were reconstructed by using the simple mid-point triangulation method. The resulted 3D point clouds are shown in Figure 11 and their measured extent is plotted in Figure 12.

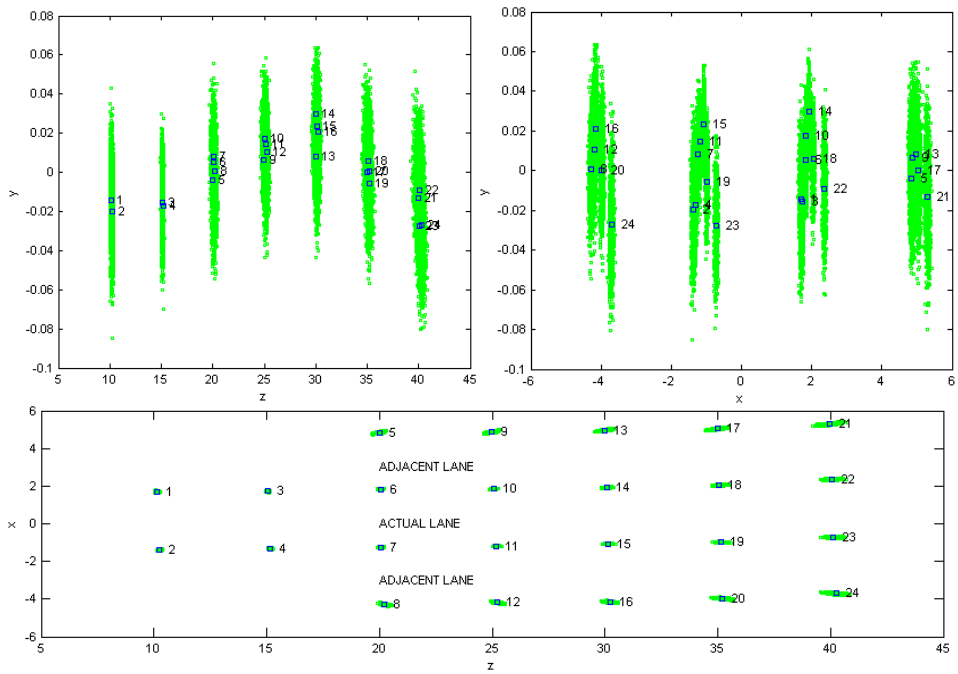

Fig. 11. Different views of the reconstructed 3D point clouds resulted from the Monte-Carlo simulations. XZ (top-left), XY (top-right) and XZ (bottom) view, where X is the lateral, Y is the height and Z is the longitudinal (depth) coordinate in meters. 1000 experiments were carried out.

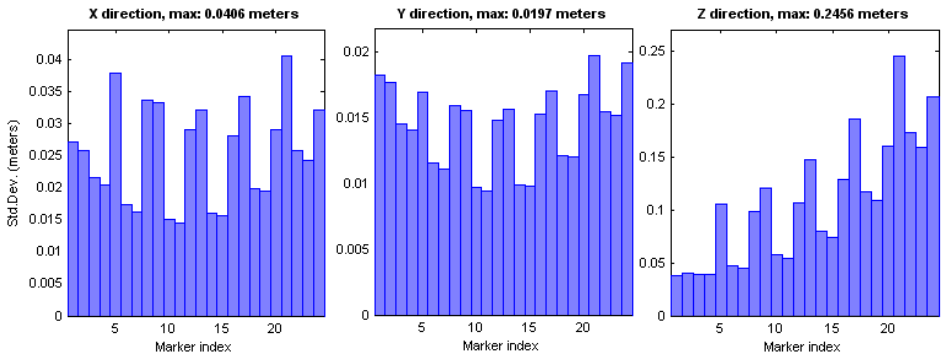

Fig. 12. Standard deviation (in meters) of the point clouds' sizes for each of the 24 reference points when only the optimal camera parameters are perturbed.

The largest deviation is 25 cm in the longitudinal (Z) direction at a distance of 40 meters from the car. This means that the "extent" (the 99% confidence interval) of the corresponding 3D point cloud is ±64 cm. In the direction perpendicular to the ground, the 99% confidence interval is ±5 cm and in the lateral direction, it is ±10 cm. We should emphasize that this only characterizes reconstruction errors due to uncertainty present in the camera parameters and not due to errors in stereo matching.

In order to simulate the effects of a random error (but not outliers) present at the stereo matching of point features, as well, random 2D Gaussian noise has been added to the image point locations in each experiment. We repeated the whole experiment with different standard deviations that ranged from 0 to 0.66 pixels in 0.11 pixels steps. In the case of a 2D Gaussian distribution, 0.33 pixels correspond to a 99% confidence interval of ±1 pixels while 0.66 corresponds to ±2 pixels. This is the simulated precision of the feature localization and stereo matching solution. The resulted point reconstruction errors are shown in Figure 13.

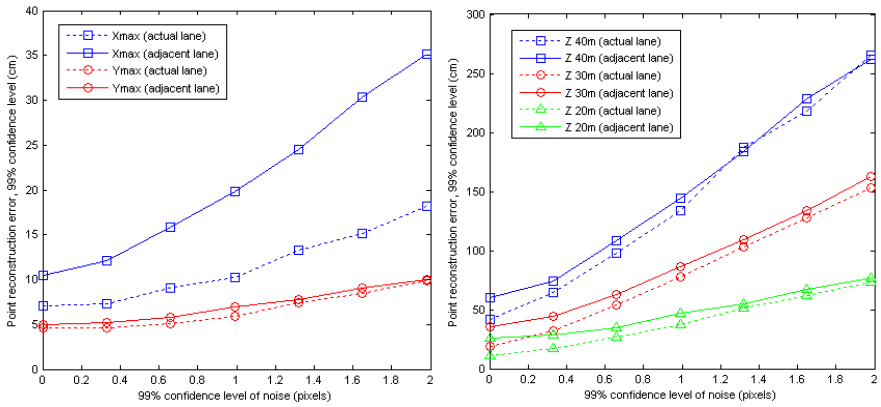

Fig. 13. Maximum reconstruction errors in the X and Y (left side) and Z (right side) directions vs. the simulated point feature localization and stereo matching errors in the image while the optimal camera parameters are perturbed corresponding to $\mathbf{C}_p$.

In Figure 13, both 3D and 2D errors are given as the 99% confidence interval of the corresponding distribution. It can be seen that the effects of the localization and feature

matching noise in the image starts to dominate the uncertainty present in the camera parameters from ±0.5 pixels (Z plot). As we expected, the depth coordinate is the most sensitive one.
It should be noted that although parameter uncertainty is simulated as a random noise in order to measure the uncertainty of the parameter estimates with respect to the true parameters, the error of a single realized calibration remains constant when the calibrated system is on-line. In contrast, feature localization is realized in every acquired frame. However, the random perturbation is still valid, since we are interested in the deviation of the reconstructed features from the true ones.

6.3 Uncertainty of the epipolar lines

Next, the sensitivity of the epipolar lines was analysed as described in Section 5.3. The epipolar line uncertainty was characterized by the deviation of the line's angle with respect to its mean value. To every pixel center in the left image, the angle deviation of the corresponding epipolar line in the right image is associated. The resulting surface is shown in Figure 14.

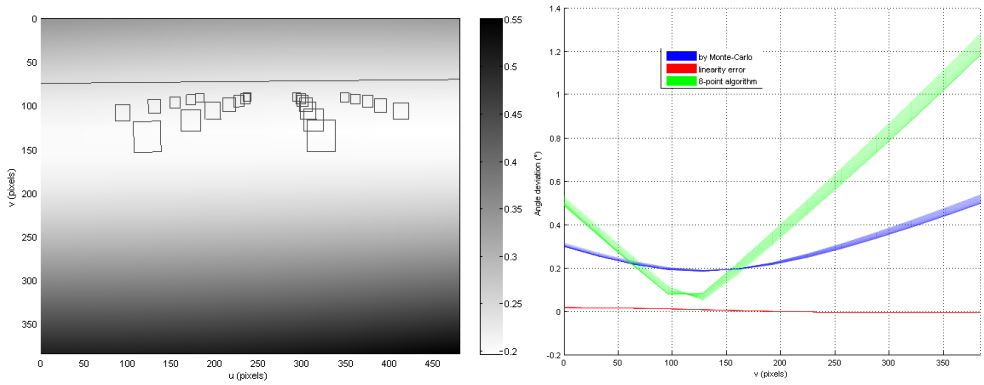

Fig. 14. Epipolar line uncertainties. Every pixel in the left image has an associated epipolar line uncertainty. The uncertainty is encoded in gray level values over the pixels of the left image (left side), while the side view of the resulting surface is compared with that of the eight-point algorithm (right side).

The angle deviation was not only computed by Monte-Carlo simulations but also with the linear covariance-propagation method detailed in Section 5.3. The difference between the two resulted surfaces is the linearity error surface that is also shown in the right side of Figure 14 (the side view of this error surface is a line near zero degrees).
As a reference, we plotted the surface received from the eight-point algorithm used to determine the fundamental matrix. Since this method breaks down in the case of flat arrangements, we used the center and all the four corners of the markers in both images (five times more reference points than those used in the second step of the calibration procedure). We should also mention that the eight-point algorithm, or more generally the fundamental matrix, in itself does not suffice for the specific purpose because – as discussed in Section 3 - it does not provide Euclidean information about the camera poses with respect to the scene.

The uncertainty in the angle of epipolar lines does not significantly depend on the horizontal coordinate of the corresponding point (the epipolar lines are almost all horizontal). Although the objective function during optimization was not the uncertainty in the epipolar lines itself, it is clear from Figure 14 that this is minimal in the interesting zone. This due to the specific arrangement (the marker locations and the horizon are overlayed for this purpose). The minimum of the angle deviation is around 0.2° and the maximum is 0.5°.

6.4 The sensitivity of road surface detection and lane model fitting

In order to see how the uncertainty in the camera parameters affect the quality of road surface reconstruction, we have used the Monte-Carlo technique, once again. In each of the 100 experiments, the feature (primitive or point chain) matching, the stereo point reconstruction and the road surface model were recomputed. The computations were performed for each frame over a 50 frames sequence, which corresponds to 2 seconds in real-time. The computed optimal camera parameters were perturbed corresponding to the estimated parameter covariance $\mathbf{C}_p$. The pitch angle, roll angle and height parameters resulted from the road surface model fitting are shown in Figure 15. Although there are some outliers (e.g. at frame indices 4, 10 and 49, that may correspond to surfaces with relatively high residual errors), the sensitivity of the estimation remains constant. In other words, the LS-fitting, in itself, is not very reliable in all circumstances, but the sensitivity estimation still remains stable over time. The standard deviations are around 0.11°, 0.36° and 5.4 cm, for the pitch, roll and height parameters respectively. The stability of fitting could be increased by using a robust fitting method or a weighted least-squares (WLS) method by giving more weight to the farther reconstructed points or primitives. This is because much more points constitue closer primitives than the farther ones so that farther points are not really involved in model shaping (we refer to Figure 5).

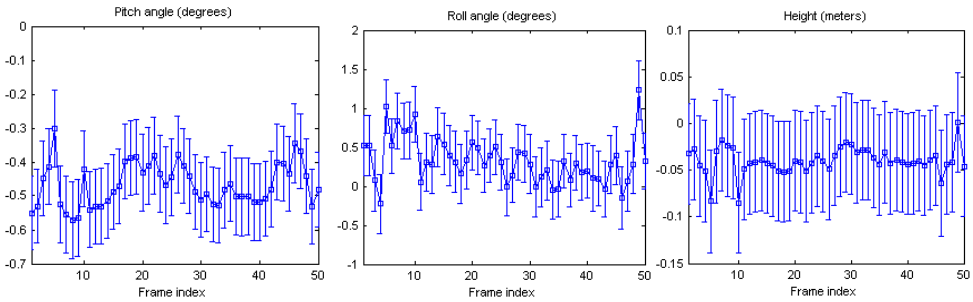

Fig. 15. Uncertainties in road surface model fitting due to uncertainties present in the camera parameters. 100 experiments in each of the 50 successive frames are evaluated. The curves represent mean values and the bars represent the standard deviations. There are outiers in model fitting but the computed sensitivities remain stable over time.

Finally, Figure 16 shows the effects of the computed camera parameter uncertainties on lane geometry reconstruction demonstrated on the frames already shown in Figure 6. Figure 16 demonstrates that the proposed off-line calibration method together with the discussed stereo lane reconstruction method gives acceptable lane reconstruction accuracy, but in the meantime, the derived errors are not insignificant, and thus, they can not be neglected, even if special care has been taken at calibration.

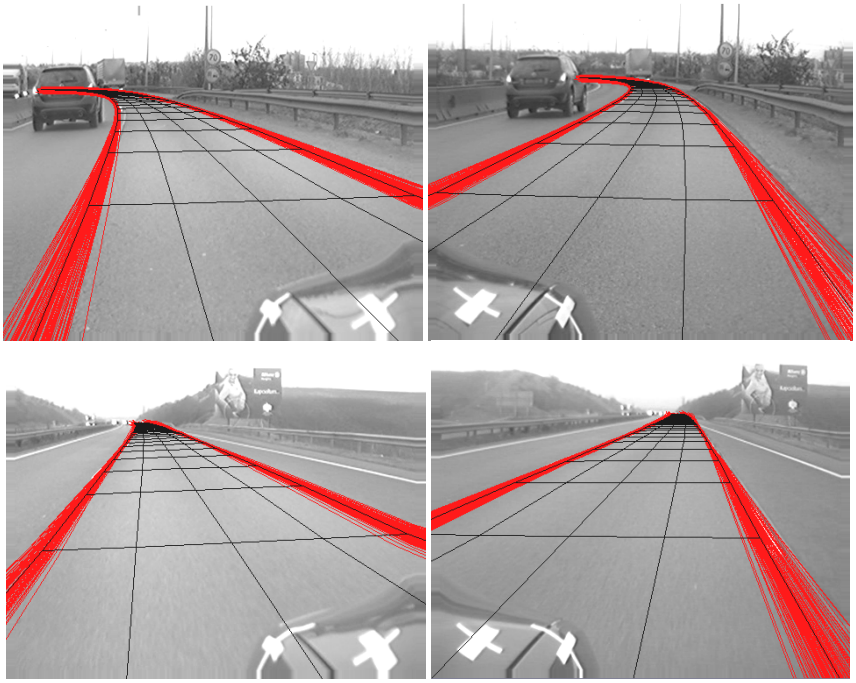

Fig. 16. Uncertainties in lane reconstruction due to uncertainties present in the camera parameters. 100 experiments are overlayed.

7. Conlusions

A novel off-line static method has been proposed for calibrating the cameras of a stereo vision-based driver assistance system. We formulated the maximum likelihood cost function for the stereo calibration problem. The resulting method involves the optimization of the 3D marker locations and covariance information of the measurements. Therefore, the method is only applicable, when an appropriate preliminary estimation of the uncertainties of the calibration measurements can be given. Moreover, a method for estimating the sensitivity of the parameters has been presented. It has been shown on real data, that when measurement uncertainties are available, our approach co-minimizing errors in the image together with errors in 3-space gives significantly better results than one can achieve by using the common reprojection error minimization. Thus, we have put extra effort in estimating measurement uncertainties at calibration. A stereo lane reconstruction algorithm has also been presented and by Monte-Carlo simulations of a triangulation method, we have demonstrated how the computed parameter uncertainties affect the precision of 3D reconstruction. The estimated reconstruction errors can be used when defining the safety margins in a decision algorithm that may trigger an actuation in a critical situation (e.g. unexpected lane departure or collision). The study should draw attention to the errors remaining in a vision system which is increasingly important in safety-critical systems.

The procedure followed in the estimation of 3D point reconstruction uncertainties can be applied to estimate the output quality of a vehicle or obstacle detection algorithm. This is

meant without a tracking algorithm that should decrease the errors by involving temporal information. It should be noted that, in general, tracking increases robustness, but the vision algorithm without tracking should still be reliable in itself, as well.
The proposed optimization method and far-range calibration arrangement of "ground thruth" control points is relatively elaborate compared to markerless on-line methods, while it is indispensable to integrate some kind of on-line parameter estimation - or at least parameter checking – in such systems. This is critical because the cameras are subject to shocks and vibrations and the parameters (mostly the extrinsic parameters) may change over time. Thus, the presented methods and results will serve as a reference to evaluate some on-line calibration methods that are presently developed.

8. Acknowledgements

This work was supported by the Hungarian National Office for Research and Technology through the project "Advanced Vehicles and Vehicle Control Knowledge Center" (OMFB-01418/2004) which is gratefully acknowledged.

9. References

Bellino, M.; Meneses, Y. L.; Kolski, S. & Jacot J. (2005). Calibration of an embedded camera for driver assistant systems, *Proceedings of the IEEE Intelligent Transportation Systems*, pp. 354-359, ISBN: 0-7803-9215-9, Vienna, Austria, 13-15 Sep 2005

Bertozzi, M. & Broggi, A. (1998). GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, Vol. 7, No. 1 (Jan. 1998), 62-81, ISSN: 1057-7149

Bertozzi, M.; Broggi, A.; Cellario, M.; Fascioli, A.; Lombardi, P. & Porta, M. (2002). Articial vision in road vehicles. *Proc. of the IEEE*, Vol. 90, No. 7, 1258-1271, ISSN: 0018-9219

Bodis-Szomoru, A.; Daboczi, T.; Fazekas, Z. (2007). A far-range off-line camera calibration method for stereo lane detection systems, *Proceedings of the IEEE Conference on Instrumentation and Measurement Technology (IMTC'07)*, pp. 1-6, ISBN: 1-4244-0588-2, Warsaw, Poland, 1-3 May 2007

Broggi, A.; Bertozzi, M. & Fascioli, A. (2001). Self-calibration of a stereo vision system for automotive applications, *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3698-3703, ISBN: 0-7803-6578-X, Seoul, Korea, 21-26 May, 2001

Eidehall, A. & Gustafsson, F. (2004). Combined road prediction and target tracking in collision avoidance, *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'04)*, pp. 619-624, ISBN: 0-7803-8310-9, Parma, Italy, 14-17 June 2004

Fischler, A. & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, vol. 24, no. 6 (June 1981), 381-395, ISSN: 0001-0782

Hartley, R. & Zisserman, A. (2006). *Multiple View Geometry in Computer Vision, Second Edition*, Cambridge University Press, ISBN: 0521-54051-8, Cambridge, United Kingdom

Heikkila, J. & Silvén O. (1997). A Four-step camera calibration procedure with implicit image correction, *Proceedings of the Conference on Computer Vision and Patter Recognition (CVPR'97)*, pp. 1106-1112, ISBN: 0-8186-7822-4, San Juan, Puerto Rico, 17-19 June 1997

Kastrinaki, V.; Zervakis, M. & Kalaitzakis, K. (2003). A survey of video processing techniques for traffic applications, *Image and Vision Computing*, Vol. 21, No. 4 (April 2003), 359-381, DOI: 10.1016/S0262-8856(03)00004-0

Kumar, R. & Hanson, A. R. (2004). Robust methods for estimating pose and a sensitivity analysis, *Computer Vision Graphics and Image Processing: Image Understanding*, Vol. 60, No. 3 (Nov. 1994), 313-342, ISSN: 1049-9660

Lepetit, V. & Fua, P. (2005). Monocular model-based 3D tracking of rigid objects: A survey, *Foundations and Trends in Computer Graphics and Vision*, Vol. 1, No. 1, 1-89, ISSN: 1572-2740

Malm, H. & Heyden, A. (2001). Stereo head calibration from a planar object, Proceedings of the Conference on IEEE Computer Society, pp. 657-662, ISBN: 0-7695-1272-0, Kauai, Hawaii, 8-14 December 2001

Malm, H. & Heyden, A. (2003). Simplified intrinsic camera calibration and hand-eye calibration for robot vision, *Proceedings of the Conference on IEEE Intelligent Robots and Systems*, pp. 1037-1043, ISBN: 0-7803-7860-0, Las Vegas, Nevada, USA, Oct. 2003

Marita, T.; Oniga, F.; Nedevschi, S.; Graf, T. & Schmidt, R. (2006). Camera calibration method for far range stereovision sensors used in vehicles, *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'06)*, pp. 356-363, ISBN: 4-901122-86-X, Tokyo, Japan, 13-15 June 2006

Nedevschi, S.; Danescu, R.; Marita, T.; Oniga, F.; Pocol, C.; Sobol, S.; Graf, T. & Schmidt, R. (2005). Driving environment perception using stereovision, *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'05)*, pp. 331-336, ISBN: 0-7803-8961-1, Las Vegas, Nevada, USA, 6-8 June 2005

Nedevschi, S.; Oniga, F.; Danescu, R.; Graf, T. & Schmidt, R. (2006). Increased accuracy stereo approach for 3D lane detection, *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'06)*, pp. 42-49, ISBN: 4-901122-86-X, Tokyo, Japan, 13-15 June 2006

Sun, W. & Cooperstock, J.R. (2005). Requirements for camera calibration: Must accuracy come with a high price?, *Proceedings of the Seventh IEEE Workshop on Application of Computer Vision*, pp. 356-361, Breckeneidge, Colorado, USA, 5-7 Jan. 2005

Sturm, P. F. & Maybank, S. J. (1999). On plane-based camera calibration: A general algorithm, singularities, applications, *Proceedings of the Conference on Computer Vision and Patter Recognition (CVPR'99)*, pp. 1432-1437, ISBN: 0-7695-0149-4, Fort Collins, Colorado, USA, 23-25 June 1999

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Journal of Robotics and Automation*, Vol. 3, No. 4 (Aug. 1987), 323-344, ISSN: 0882-4967

Weber, J.; Koller, D.; Luong, Q.-T. & Malik, J. (1995). New results in stereo-based automatic vehicle guidance, *Proceedings of the Intelligent Vehicles Symposium (IV'95)*, pp. 530-535, ISBN: 0-7803-2983-X. 386, Detroit, Michigan, USA, 25-26 Sep. 1995

Zhang, Z. (2000). A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11 (Nov. 2000), 1330–1334, ISSN: 0162-8828.

Zollner, H. & Sablatnig, R. (2004). Comparision of methods for geometric camera calibration using planar calibration targets, *Proceedings of the 28th Workshop of the Austrian Association for Pattern Recognition*, pp. 237-244, Hagenberg, Austria, 2004