Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, October 3-7, 2009

Robust Ground Plane Estimation Based on Particle Filters

Andre G. Linarth
Driver Assistance and Sensor Information
Elektrobit Automotive GmbH
Erlangen, Germany

Manuel Brucker and Elli Angelopoulou
Chair of Pattern Recognition
Friedrich-Alexander University Erlangen-Nuremberg
Erlangen, Germany

Abstract—*This paper describes a Particle Filter based approach for estimating the ground plane from an image sequence. Based on a Bayesian framework, the Particle Filter provides a robust estimation of the plane parameters, since it can handle non-linearities, while allowing a high flexibility for integrating new cues into the system. Furthermore, the different modes of the resulting probability density function are segmented by means of a mean-shift algorithm, resulting in better localization of the estimate with the highest posterior probability. Our method has been tested on both synthetic and real world scenarios and has shown to be robust to missing and unstable measurements. On synthetic data of representative runs the angular error is well within 0.5° with a standard deviation of less than 0.3°.*

Ground plane estimation; Particle Filter; Sequential Monte Carlo, Mean Shift

I. Introduction

The knowledge of the ground plane parameters is widely used in automotive vision applications in order to, for example, restrict search areas, have a fast depth calculation of points lying on the plane, or segment the road area. When a camera is stably fixed in the vehicle, the ground plane parameters can be calculated a priori by means of calibration patterns. However, such a calibration pre-processing step is not always feasible. A typical case is a camera mounted in a portable device, e.g. a mobile phone or navigation system, which is set up in a vehicle. The user may change the position and orientation of the device, while the provision of calibration patterns in such case would also be unacceptable. In our work, we present a probabilistic method to automatically estimate the ground plane parameters out of forward looking camera images under vehicle motion. Such a methodology can also be used in fixed-camera systems, providing important information on the reliability of (as well as adapting) the factory calibrated parameters over time.

Ground plane estimation is often related to ego-motion estimation in the literature. For solving these problems in a real world scenario, it is not easy to apply traditional Structure from Motion algorithms, since the cluttered background present in the scenes makes the task of depth estimation more difficult [4]. Several researches proposed approaches based on optimization techniques [3, 4, 6, 8, 10]. Such algorithms need to adjust up to eight parameters, and since this is a non-convex optimization problem it may get stuck in local minimas. These approaches depend heavily on the quality of the initial guess. Another important aspect of iterative approaches is their inherent sequential architecture. Modern hardware systems maximally exploit parallelism, and such kind of sequential processes limits the scope of possible performance optimizations.

A smaller group of approaches, in which our method can be included, makes use of a probabilistic framework to robustly estimate plane and motion parameters. Torr et al. [1], for example, showed that out of multiple cues it is possible to segment different planes in image sequences by means of a Bayesian framework. Stein et al. [3] used a probabilistic approach for estimating the ego-motion. The registration error is applied as the likelihood function. The algorithm first segments the road plane by assigning weights to image patches according to their likelihood to the motion model. Secondly the motion model is refined by means of an optimization technique. The authors leave the combination of estimations of individual patches as future work.

Our technique is based on the Sequential Monte Carlo methodology, and more specifically on Particle Filters. The ground plane estimation problem is expressed in a Bayesian framework and the Particle Filter is used in obtaining an estimate of the probability density functions (pdf). Compared to simple geometric solutions, such probabilistic frameworks are usually more robust to noise in the measurements, since this is explicitly included in the model. The resulting pdf also provides important statistics on the reliability of the estimation. Among the advantages of a Particle Filter-based methodology are the flexibility to include multiple cues, the inherently parallel architecture (particles do not depend on each other) and the ability to handle non-linear systems. In our methodology, we exploit the fact that the ground plane parameters will change slowly over time. The parameters are then tracked and the resulting pdf is further processed through a fast adaptive mean shift algorithm to segment the different modes and select the best candidate for the parameter set.

II. Ground Plane Estimation

Given a single camera mounted in an unknown position in a moving vehicle, we want to estimate the parameters describing the road plane relative to the camera coordinate system. We assume perspective projection and a pinhole camera model. A 3D world point $\boldsymbol{a}_w$ is projected to a point $\boldsymbol{a}_i$ in the image plane as follows:

$$\boldsymbol{a}_i = \boldsymbol{K}[\boldsymbol{R}_w | \boldsymbol{\tau}_w]\boldsymbol{a}_w \tag{1}$$

978-1-4244-5521-8/09/$26.00 ©2009 IEEE

where $\boldsymbol{K}$ is the 3×3 matrix representation of the camera intrinsic parameters, $[\boldsymbol{R}_w | \boldsymbol{\tau}_w]$ is a 3×4 matrix transformation that maps the 3D world point $\boldsymbol{a}_w$ to the camera coordinate system, formed by the rotation matrix $\boldsymbol{R}_w$ and the translation vector $\boldsymbol{\tau}_w$. Both $\boldsymbol{a}_i$ and $\boldsymbol{a}_w$ are represented in homogeneous coordinates. Given this model and considering a moving camera, one can compute a geometric relationship between two consecutive image planes [8]. In this paper we are interested on a specific transformation, the homography $\boldsymbol{H}$, which maps a 2D homogeneous point $\boldsymbol{a}_{\pi_1}$ on the plane π_1 to the homogeneous point $\boldsymbol{a}_{\pi_2}$ on the plane π_2.

$$\boldsymbol{a}_{\pi_2} = \boldsymbol{H}\,\boldsymbol{a}_{\pi_1} \tag{2}$$

When π_1 and π_2 are the image planes at two different viewing positions of a static scene (see Fig. 1), Faugeras and Lustman first showed in [12], that the homography is described by:

$$\boldsymbol{H} = \boldsymbol{K}\left(\boldsymbol{R}_c - \frac{\boldsymbol{\tau}_c}{d}\boldsymbol{\eta}^T\right)\boldsymbol{K}^{-1} \tag{3}$$

where $\boldsymbol{\eta}$ is the normal of an inducing plane in the scene, located at a distance d from the origin of the first camera. $\boldsymbol{R}_c$ and $\boldsymbol{\tau}_c$ give respectively the rotation and translation of the camera centers between the two views. From at least four non-collinear point correspondences on the scene, the homography can be calculated by means of a least square approach [8]. The parameters can be further optimized by minimizing the error between the reference frame and the second frame warped in terms of the estimated homography. The forming components of the inducing plane be decomposed out of the homography by applying the methods described in [11, 12].

Our approach is also based on evaluating this error, but it evolves in the opposite direction. We randomly generate candidates for the parameters and evaluate how each of these candidates fits the geometric/photometric model. The system then concentrates the computational effort on the most probable regions, and after a number of frames, a refined estimation of the parameters can be derived.

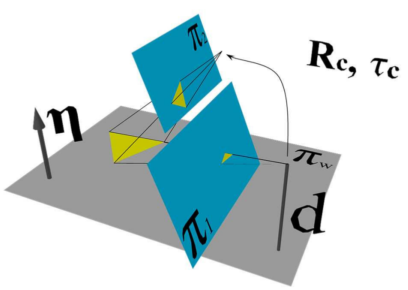

Fig. 1 – Homography induced by plane

III. Particle Filter Based Ground Plane Estimation

A. Bayesian Tracking

We will begin this section with a brief review of Bayesian Tracking, which can be seen as the analytical basis of Particle Filtering. The overall goal of Bayesian Tracking is to obtain the pdf $p(\boldsymbol{x}_t | \boldsymbol{z}_{1:t})$ describing the probability of a dynamic system to end up in a certain state $\boldsymbol{x}_t$ given all the observations $z_{1:t} = \{z_i, i=1,\ldots,t\}$ up to time t. This pdf can be obtained by alternating between two steps: prediction and update. Under the Markovian assumption, $p(\boldsymbol{x}_t | \boldsymbol{x}_{1:t-1}) = p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$, the prediction step can be expressed by:

$$\begin{gathered} p(\boldsymbol{x}_t | z_{1:t-1}) = \\ \int p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})\, p(\boldsymbol{x}_{t-1} | z_{1:t-1})\, d\,\boldsymbol{x}_{t-1} \end{gathered} \tag{4}$$

and the update step by:

$$p(\boldsymbol{x}_t | z_{1:t}) = \frac{p(z_t | \boldsymbol{x}_t)\, p(\boldsymbol{x}_t | z_{1:t-1})}{p(z_t | z_{1:t-1})} \tag{5}$$

Since

$$p(z_t | z_{1:t-1}) = \int p(z_t | \boldsymbol{x}_t)\, p(\boldsymbol{x}_t | z_{1:t-1})\, d\,\boldsymbol{x}_t \tag{6}$$

is a normalizing constant, one only needs to model the state transition probability $p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$ and the observation probability $p(z_t | \boldsymbol{x}_t)$ in order to recursively calculate $p(\boldsymbol{x}_t | \boldsymbol{z}_{1:t})$.

B. Sequential Importance Sampling

Given a set of samples (particles) $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N\}$ drawn according to a pdf $p(\boldsymbol{x})$, it is possible to discretely approximate the probability distribution function by:

$$p(\boldsymbol{x}) \approx \frac{1}{N}\sum_{i=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}^i) \tag{7}$$

where N denotes the number of samples and δ is the *dirac* delta function. If drawing samples from $p(\boldsymbol{x})$ is not possible or difficult, but the density function can be evaluated at a given point, one can instead draw the samples from an arbitrary density $q(\boldsymbol{x})$, the so called importance density and approximate $p(\boldsymbol{x})$ by:

$$p(\boldsymbol{x}) \approx \sum_{i=1}^{N} \omega^i \delta(\boldsymbol{x} - \boldsymbol{x}^i), \omega^i \propto \frac{p(\boldsymbol{x}^i)}{q(\boldsymbol{x}^i)}, \sum_{i=1}^{N} \omega^i = 1 \tag{8}$$

where ω^i is the importance weight of the *i*-th sample. Note, however, that the support of $q(\boldsymbol{x})$ has to be equal to or bigger than the support of $p(\boldsymbol{x})$ to be able to represent the target function. This is known as the principle of importance sampling [13]. The posterior distribution can then be approximated by:

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \approx \sum_{i=1}^{N} \omega_t^i \delta(\boldsymbol{x}_t - \boldsymbol{x}_t^i) \tag{9}$$

where ω_t^i is the importance weight of the *i*-th particle in the *t*-th timestep. If the importance density is constructed such that it can be factorized in two parts:

$$q(\boldsymbol{x}_t^i|\boldsymbol{z}_{1:t}) = q(\boldsymbol{x}_{t-1}^i|\boldsymbol{z}_{1:t-1})\, q(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1}^i, \boldsymbol{z}_t) \tag{10}$$

it can be shown [17], that the corresponding weights of each sample can be calculated recursively according to

$$\omega_t^i \propto \omega_{t-1}^i \frac{p(\boldsymbol{z}_t|\boldsymbol{x}_t^i)\, p(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1}^i)}{q(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1}^i, \boldsymbol{z}_t)}, \sum_{i=1}^{N} \omega_t^i = 1 \tag{11}$$

Note that as $N \to \infty$, the approximation of eq. (9) approaches the true posterior density $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$. By choosing

$$q(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1}^i, \boldsymbol{z}_t) = p(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1}^i) \tag{12}$$

equation (11) can be rewritten as:

$$\omega_t^i \propto \omega_{t-1}^i\, p(\boldsymbol{z}_t|\boldsymbol{x}_t^i) \tag{13}$$

The process for calculating the posterior distribution is then achieved by recursively predicting the new state through $p(\boldsymbol{x}_t^i|\boldsymbol{x}_{t-1}^i)$, updating the measurements with $p(\boldsymbol{z}_t|\boldsymbol{x}_t^i)$ and calculating the weights using eq. (13).

C. *Sequential Importance Resampling*

One of the main problems of the sequential importance sampling scheme is that the weights tend to become highly degenerate after some iterations. This means that a small portion of the particles will contain nearly all the probability mass, and therefore the majority of the particles will almost not contribute to the estimates [2]. An idea for avoiding this problem is to move the particles towards regions with a higher probability mass. One way to achieve this is to draw a new set of particles proportionally to their previous weights. In other words, the new set is formed by sampling with replacement *N* times from the discrete representation of $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$. Once the results are i.i.d. samples of the discrete distribution, we can then set the corresponding weights to *1/N*.

D. *Effective Sample Size*

A good measurement of the degeneracy problem described above is the effective sample size [13]:

$$ESS_t = \frac{N}{1 + \frac{1}{N}\sum_{i=1}^{N} (N\omega_t^i - 1)^2} \tag{14}$$

The effective sample size can be seen as a measure of how many of the samples significantly contribute to the estimated target pdf $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$. Large values indicate more uniformly distributed weights, while small values imply a bigger variance on their distribution.

E. *Observation and Motion Models*

We define a five dimensional state space, which consists of: two dimensions for the unit normal vector to the ground plane η_θ and η_ϕ, given in spherical coordinates with the length being equal to one; two dimensions for the translation vector τ_θ and τ_ϕ, also in spherical coordinates, whose magnitude is extracted from the velocity and time data; and one dimension for the distance d of the camera from the ground plane. In order to reduce our search space, we assume, for the time being, no camera rotation between two consecutive frames. Such an assumption is valid for trajectories with no sharp curves. Our *i*-th particle at time *t* is then expressed by:

$$\boldsymbol{x}_t^i = \{\eta_{\theta,t}^i \quad \eta_{\phi,t}^i \quad d_t^i \quad \tau_{\theta,t}^i \quad \tau_{\phi,t}^i\} \tag{15}$$

Given the observation z_{t-1}, which in this case is the Image I_{t-1} acquired by the camera at timestep $t-1$, one can estimate $\hat{I}_t^i$ by warping the previous image I_{t-1} according to the homography $\boldsymbol{H}_{x_t^i}$ given by the *i-th* particle $\boldsymbol{x}_t^i$, and calculated through eq. (3):

$$\hat{I}_t^i = W(I_{t-1}, \boldsymbol{H}_{x_t^i}) \tag{16}$$

Upon the arrival of the new observation I_t it is possible to calculate the registration error ζ_R^i between estimated and current frames:

$$\zeta_R^i = \frac{\sum_M (I_t - \hat{I}_t^i)^2}{M} \tag{17}$$

where M is the number of pixels considered in the registration step. Equation (17) measures the distance between predicted and current measurements, and it represents our first cue for the estimation process. This measurement, by itself, could suffice for calculating the probability of the observation at time t given a specific particle. However, in the Particle Filter framework, at this stage, it is possible to combine multiple pieces of information one can derive out of the measurements, to increase the robustness of the estimation. In our methodology we define a second measurement for the pitch angle, based on the road's vanishing point. The pitch is defined as the rotation around the x-axis, and our coordinate system is defined at the camera center at time $t-1$ (see Fig. 2). The intersection of the projection of parallel road markers in the image plane gives the road vanishing point. Some advanced approaches for calculating vanishing points are presented in [14, 15, 16]. With known intrinsic camera parameters, the pitch angle α can be derived by the trigonometric function:

$$\alpha = \arctan\left(\frac{c_y - v_y}{f_y}\right) \tag{18}$$

where c_y and v_y are the *y* coordinates of, respectively, the camera principal and the road vanishing points. f_y Is the focal length in pixels. The pitch error ζ_α^i can be defined as:

$$\zeta_\alpha^i = |\alpha - \hat{\alpha}^i| \tag{19}$$

where $\hat{\alpha}$ is the pitch angle given by the *i*-th particle and calculated out of the normal vector by:

$$\hat{\alpha}^i = \arctan\left(\frac{\cos \eta_\theta^i}{\sin \eta_\phi^i \sin \eta_\theta^i}\right) \tag{20}$$

Priors can also be incorporated at this stage. One may restrict the search space by penalizing the observation probability, when the particles move out of an expected target area. We define the following penalty term based on a Gaussian function:

$$\rho_\eta(\boldsymbol{x}_t^i) = e^{-\left[\frac{(\eta_{\theta,t}^i - \eta_{\theta,0})^2}{2\sigma_{\eta_\theta}^2} + \frac{(\eta_{\phi,t}^i - \eta_{\phi,0})^2}{2\sigma_{\eta_\phi}^2}\right]} \tag{21}$$

where index *0* indicates the corresponding parameter values at $t=0$. σ_{η_θ} and σ_{η_ϕ} are the standard deviation of, respectively, η_θ and η_ϕ . These terms define the size of the expected search-window for the corresponding parameter. In our setup we set it to a quarter of the expected range, such that at the boundaries of the expected values, the penalties reach values close to zero.

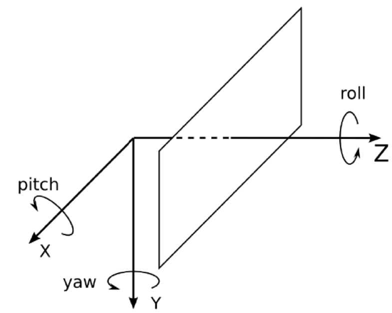

Fig. 2 – Pitch, roll and yaw angles

For more details the reader is referred to section IV. Finally, we combine all the different cues in a single observation probability as follows:

$$p(\boldsymbol{z}_t | \boldsymbol{x}_t^i) \propto e^{\frac{-\zeta_R^i}{\gamma_R}} e^{\frac{-\zeta_\alpha^i}{\gamma_\alpha}} \rho_\eta \tag{22}$$

where γ_R and γ_α are empirically determined values related to the confidence of, respectively, the registration and pitch measurements.

Regarding the motion model, in our approach, without further knowledge of the movement of the camera relative to the ground plane, the state transition is unpredictable. However the plane parameters should not change significantly, as well as no big changes on the translation vector are expected between two consecutive frames. Therefore, we consider that changes of the parameters are exclusively caused by the noise component of our model, which, without loss of generality is assumed to be Gaussian:

$$p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_{t-1}, \Sigma) \tag{23}$$

where Σ is the covariance matrix containing empirically determined variances (see section IV).

F. Mean Shift

Once our approximation of the pdf $p(\boldsymbol{x}_t | \boldsymbol{z}_{1:t})$ is calculated, it is important to decide which mode of our estimation represents our final solution. According to the Bayesian framework, we should select the estimate with the highest posterior probability value. We propose in this paper the application of a mean-shift [7] segmentation algorithm in order to robustly identify the different modes of the pdf. Particles will move and concentrate proportionally around highly probable regions. The mode with the biggest concentration of particles will contain our final solution. The convergence point with the highest number of particles resulting from the segmentation is then chosen as the set of estimated parameters. In our implementation we have applied the fast adaptive mean-shift algorithm described in [19].

Algorithm 1 shows one iteration of the complete algorithm.

Algorithm 1 : Iteration of the method

for i = 1:N

Propagate $\tilde{\boldsymbol{x}}_t^i \sim p(\boldsymbol{x}_t^i | \boldsymbol{x}_{t-1}^i)$

Compute weight $\tilde{\omega}_t^i = \omega_{t-1}^i \, p(z_t | \tilde{\boldsymbol{x}}_t^i)$

end for

Normalize weights for each particle i $\omega_t^i = \dfrac{\tilde{\omega}_t^i}{\sum_{j=1}^{N} \tilde{\omega}_t^j}$

Resample according to normalized weights

Segment $\boldsymbol{x}_t^i, i=1,\ldots,N$ with FAMS

Select the convergence point of the mode with most particles as the current estimation $\boldsymbol{x}_t$

IV. Results

We have tested our system on both synthetic and real data. In the first subsection we concentrate on the comparison of the estimated values with the available ground truth, while in the second subsection, we evaluate the system in a real sequence. A reference implementation of an optimization based approach is also presented.

The registration cue based on the homography mapping needs to be applied exclusively in points belonging to the road plane. For this reason, for all the experiments, a rough segmentation of the road plane is applied. The left picture in Fig. 5 shows the segmented area overlayed on the driving road.

As discussed in section III-E, we selected σ_{η_θ} and σ_{η_ϕ} to be a quarter of our expected range in order to have the penalty term close to zero in the boundaries of the range. For our experiments we defined the range to be approximately ±15° around the y-axis. Therefore, $\sigma_{\eta_\theta} = \sigma_{\eta_\phi} = 3.75°$ and $\boldsymbol{x}_{\eta_\theta,0} = \boldsymbol{x}_{\eta_\phi,0} = 0°$. The standard deviation terms should be large enough to avoid (or to reduce significantly) the bias added by the prior cue on the final estimation. Regarding the covariance matrix Σ in the propagation step, we set its entries such that the standard deviation of the angular error and distance to the ground are respectively 0.1° and 5mm. These values should be sufficiently high to cover parameter changes between two consecutive timesteps. Finally the gains γ_R and γ_α were empirically selected based on statistics of previous runs. For the experiments they were set respectively to 2.5×10^{-4} and 10. Such factors include information of the confidence of the measurement and determine its influence on the final estimation.

A. *Synthetic Data*

Synthetic road scenes were generated with known motion and plane parameters. The images simulate a vehicle driving 20s on a typical road, at a 22 fps camera frame rate. In all the experiments on simulated data 1000 particles were employed.

First we evaluated the system with different configurations, in terms of the relative orientation of the camera to the ground plane. For each scenario, we have performed ten experiments, with different seeds for the random number generators, in order to analyze accuracy of the estimated values and the convergence of the particles. We show in Table 1 the results of four indicative runs out of 40, one for each test configuration. The first column shows the ground truth values for the setup. The following columns present the mean value and standard deviation of the pitch and roll angles, as well as of the distance to the ground. The selected output is the convergence point of the mode with the most particles, as segmented by the mean shift algorithm. Each mean value shown in the table is the average of single-frame estimations over time. In our experiments we drop the first ten seconds, when the particles are still moving to their stationary state. An example of the evolution of the estimation is presented in Fig. 3, which shows the distance to the ground parameter for the first configuration in Table 1.

All the experiments exhibited very similar results, with a maximum standard deviation of 0.215° for the pitch angle, 0.211° for the roll angle and 0.019m for the height. The maximum mean estimation error was 0.170° for the pitch angle, 0.441° for the roll angle and 0.028m for the height. To further validate our stochastic process, the results of all ten different runs for each configuration, each initialized with a different seed, were combined. Table 2 shows the average and standard deviation of the root mean square errors (RMSE) of the experiments for the given configurations. The resulting very small standard deviations show that the method is robust to the initialization of the particles.

Table 1 – Estimated parameters on synthetic data

Configuration (Pitch, Roll, Height)	Pitch	Roll	Height
10°, 0°, 1.25m	9.963° ±0.098°	0.032° ±0.071°	1.244m ±0.014m
0°, -10°, 1.25m	0.006° ±0.068°	-9.584° ±0.193°	1.245m ±0.010m
0°, 0°, 1.40m	0.008° ±0.195°	0.012° ±0.085°	1.386m ±0.016m
10°, -10°, 1.40m	10.028° ±0.067°	-9.641° ±0.181°	1.411m ±0.008m

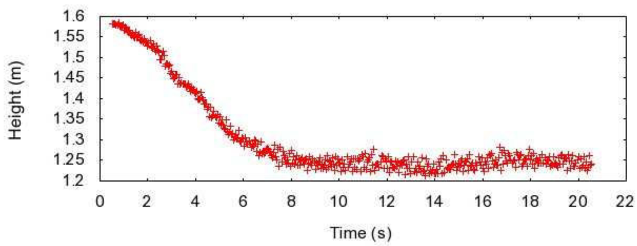

Fig. 3 - Evolution of distance to ground estimation over time

Table 2 – RMSE average and standard deviation

Configuration (Pitch, Roll, Height)	Pitch	Roll	Height
10°, 0°, 1.25m	0.083° ±0.002°	0.058° ±0.003°	0.013m ±0.0007m
0°, -10°, 1.25m	0.054° ±0.001°	0.421° ±0.008°	0.009m ±0.0004m
0°, 0°, 1.40m	0.162° ±0.003°	0.075° ±0.007°	0.020m ±0.0034m
10°, -10°, 1.40m	0.061° ±0.002°	0.367° ±0.007°	0.012m ±0.0004m

B. Real Data

In this subsection we present how the proposed system behaves in a real driving scenario. As a reference for the experiments, the camera intrinsic and extrinsic (related to road plane) parameters were previously calibrated. The intrinsic parameters K are applied to calculate the homography in equation (3). The extrinsic parameters are only shown as reference in Table 3. The test sequence consists of 65.5s of a typical driving scenario. Fig. 4 shows the parameter development over time for the whole video. The top graph shows the quality of the particle estimations, as given by the effective sample size measurement. The second subplot depicts the evolution of the distance between camera and ground plane. The next plot shows the evolution of the pitch angle estimate, while information on the roll angle is included at the bottom graph.

By reaching its maximum value, the ESS measurement clearly depicts the situations where the employed pre-processing method fails to estimate a road region (see Fig. 5 right). This is interpreted as a possible failure of a given sensor. In this case it is not possible to calculate the observation probability. All the particles receive the same weight and hence re-sampling does not occur. The particles are just propagated due to the noise model describing the uncertainty of the measurement. Therefore estimations during this time are discarded. It is important to note that at these failure points, the particles need time to converge back to their final stationary stage. We, however, consider for the final estimation all the values in valid regions, segmented by means of the ESS measurement, and discarded only the first ten seconds of the sequence to be consistent with the results of the synthetic analysis.

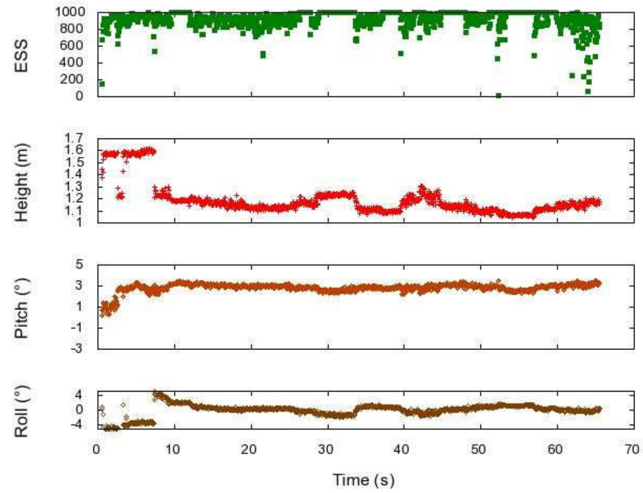

Fig. 4 – Parameter evolution over time in a real scenario

Table 3 shows the mean and standard deviation of the derived parameters for the whole sequence, together with values obtained from an off-line calibration and estimates derived by an optimization method (see sub-section IV-C). The sequence was tested with 1000 and 10000 particles. As the number of particles increases, a better representation of the real pdf is achieved. The changes on roll and height between both experiments were mainly caused due to the higher sensitivity of the methodology while running with a larger amount of samples. In a few frames, particles converged faster to a wrong minima following the measurement error.

The method has shown however, to be able to robustly estimate the parameters in the big majority of segmented frames. At the same time, it showed a good resistance to problems related to bad distribution of features on the plane (see center picture in Fig. 5). In such a case where features concentrate in specific areas, e.g. near to the roll axis, it may happen that the measurement function becomes ill-conditioned. Big changes of the roll estimation only lead to a small change of the error measurement. In such a situation, noise may have a big influence on the parameter estimation compared to geometric constraints. In our method, the smoothing provided by the Particle Filter compensates for measurements with high frequency noise.

Table 3 – Estimated parameters on real data

	Pitch	Roll	Height
Offline Calibration	3.827°	0.676°	1.257m
1000 particles	2.958° ±0.187°	0.250° ±0.600°	1.146m ±0.041m
10000 particles	2.972° ±0.190°	0.047° ±0.750°	1.158m ±0.083m
Optimization Method	2.992° ±0.337°	1.166° ±4.884°	1.087m 0.070m

Fig. 5 – Frames at times 3.95s, 15.70s and 9.59s

C. Optimization approach

A direct comparison between an optimization based approach and our methodology is not easily feasible. In order to fairly evaluate the strengths and weaknesses of each approach, both methodologies need to be tested on a large number of sequences with diverse driving scenarios. Furthermore, there are many different optimization based techniques, each with its own advantages and disadvantages. To our knowledge, there exists no publicly available dataset of diverse road sequences. Nonetheless, a comparison with an optimization approach applied in our sequence can still provide an insight on how a Particle Filter method performs relative to the more widely used optimization-based techniques.

In our optimization-based implementation we chose to use well-established sub-processes which, at the same time, fit our data set (e.g. small set of features, etc.). We applied exactly the same pre-processing for establishing a search region as we did in our approach. Within that region a Harris corner-detector [9] is applied. The detected corners are tracked between two consecutive frames using a sparse iterative implementation of the Lukas-Kanade algorithm [5]. The resulting correspondences are used in calculating the initial homography as described in [8]. The homography estimate is then optimized with the Levenberg-Marquardt algorithm as described in [18]. The derived homography is then decomposed to its rotation and translation components by applying the analytical solution from Malis [11]. The estimates obtained from our optimization-based implementation can be seen in Table 3. The presented values are obtained out of only 13% of all the frames in the sequence. The majority of the frames were dropped by plausibility and quality checks. Although a direct comparison would be unfair, these results indicate the weakness of the forward-looking registration measurement function, a problem already indicated in [4].

V. Conclusion

In this paper we have presented a method based on Particle Filters for automatically estimating the ground plane parameters from a video sequences and a velocity sensor. We showed how the probabilistic nature of the Particle Filters is well suited, in terms of robustness, to this application. Estimating the ground layer and ego-motion based on the information of the road area is challenging due to the lack of texture and non-linear perspective distortion. We believe that the results can be further improved through the use of a better measurement function. An extension we are currently investigating is the proposal of Ke and Kanade [4]. They suggest a better estimation of the motion components through the simulation of an orthogonal projection of the ground plane area. Furthermore, the entire estimation is currently based only on information provided by the road itself. It is clear that other components in the scene, e.g. buildings, vehicles or traffic signs, shall provide important information that can be combined for the estimation.

References

[1] P. H. S. Torr, R. Szeliski, P. Anandan, “An Integrated Bayesian Approach to Layer Extraction from Image Sequences”, IEEE Trans. Pattern Analysis And Machine Intelligence, vol. 23, no. 3, pp. 297-303, March 2001.

[2] O. Cappé, S. J. Godsill, E. Moulines, “An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo”, Proceedings of the IEEE, vol. 95, no. 5, pp. 899-924, May 2007.

[3] G. P. Stein, O. Mano, A. Shashua, “A Robust Method for Computing Vehicle Ego-Motion”, in proc. IEEE Intelligent Vehicles Symposium, pp. 362-368, October 2000.

[4] Q. Ke, T. Kanade, “Transforming Camera Geometry to a Virtual Downward-Looking Camera: Robust Ego-Motion Estimation and Ground-Layer Detection”, in proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 390-397, June, 2003.

[5] J. Bouguet, “Pyramidal Implementation of the Lucas-Kanade Feature Tracker: Description of the Algorithm.” Technical report, OpenCV Document, Intel Microprocessor Research Labs, 2000.

[6] J. Zhou, B. Li, “Robust Ground Plane Detection with Normalized Homography in Monocular Sequences from a Robot Platform”, in proc. IEEE Conf. Image Processing, pp. 3017-3020, October 2006.

[7] D. Comaniciu, P. Meer, “Mean Shift: A Robust Approach Toward Feature Space Analysis”, IEEE Trans. Pattern Analysis And Machine Intelligence, vol. 24, no. 5, pp. 603-619, May 2002.

[8] R. Hartley, A. Zisserman, “Multiple View Geometry in Computer Vision", Cambridge University Press, 2nd edition, 2003.

[9] C. Harris, M.J. Stephens. “A Combined Corner and Edge Detector”, in proc. Alvey Vision Conference, pp. 147–151, 1988.

[10] C. Mei, S. Benhimane, E. Malis, P. Rives, “Efficient Homography-based Tracking and 3-D Reconstruction for Single Viewpoint Sensors”, IEEE Trans. Robotics, vol. 24, no. 6, pp. 1352-1364, December 2008.

[11] E. Malis, M. Vargas, “Deeper Understanding of the Homography Decomposition for Vision-based Control”, Technical report, no. 6303 INRIA Sophia Antipolis, September 2007.

[12] O. Faugeras and F. Lustman, “Motion and Structure from Motion in a Piecewise Planar Environment”, Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 2, no. 3, pp. 485-508, 1988.

[13] J. S. Liu, R. Chen, T. Logvinenko, “A Theoretical Framework for Sequential Importance Sampling and Resampling”, Technical report, Stanford University, Department of Statistics, January 2000.

[14] J. A. Shufelt, "Performance Evaluation and Analysis of Vanishing Point Detection Techniques," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 3, pp. 282-288, March, 1999.

[15] J. Kogecka, W. Zhang, “Efficient Computation of Vanishing Points”, in proc. IEEE Conf. Robotics and Automation”, pp. 223-228, 2002.

[16] A. Almansa, A. Desolneux, S. Vamech, “Vanishing point detection without any a priori information”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 502-507, April, 2003.

[17] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, “A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking”, IEEE Trans. Signal Processing, vol. 50, no. 2, pp. 174-188, February 2002.

[18] S. Baker, I. Matthews, “Lucas-Kanade 20 Years On: a Unifying Framework”, International Journal of Computer Vision, vol. 56, no. 3, pp. 221-255, 2004.

[19] B. Georgescu, I. Shimshoni, P. Meer, “Mean Shift Based Clustering in High Dimensions: A Texture Classification Example”, in proc. IEEE Conf. Computer Vision, pp. 456-463, October 2003.