

# A Temporal Filter Approach for Detection and Reconstruction of Curbs and Road Surfaces based on Conditional Random Fields

Jan Siegemund, Uwe Franke, and Wolfgang Förstner

**Abstract**—A temporal filter approach for real-time detection and reconstruction of curbs and road surfaces from 3D point clouds is presented. Instead of local thresholding, as used in many other approaches, a 3D curb model is extracted from the point cloud. The 3D points are classified to different parts of the model (i.e. road and sidewalk) using a temporally integrated Conditional Random Field (CRF). The parameters of curb and road surface are then estimated from the respectively assigned points, providing a temporal connection via a Kalman filter.

In this contribution, we employ dense stereo vision for data acquisition. Other sensors capturing point cloud data, e.g. lidar, would also be suitable.

The system was tested on real-world scenarios, showing the advantages over a temporally unfiltered version, due to robustness, accuracy and computation time. Further, the lateral accuracy of the system is evaluated. The experiments show the system to yield highly accurate results, for curved and straight-line curbs, up to distances of 20 meters from the camera.

## I. INTRODUCTION

Robust registration and modeling of the ego vehicle’s free driving space provides the basis for many high-level driving assistance applications, such as path planing and collision avoidance.

Naturally, curbs play an important role in this context and should essentially be considered when searching for street delimiting objects. However, many existing systems for obstacle detection classify curbs as road inliers, due to their low height occurrence. Therefore, there exist several dedicated approaches, addressing the problem of curb detection directly.

Beside cameras, miscellaneous sensor types are used for this task, e.g. lidar [1] [2], time-of-flight cameras [3], or sensor fusion [4]. However, stereo camera systems are getting affordable and provide several advantages, such as a high data rate and a low requirement of space inside the vehicle.

Most of the recent stereo vision based approaches for curb detection utilize so called Digital Elevation Maps (DEM). These maps are horizontal grids, holding a local height value for each grid cell, computed from the triangulated 3D points.

In [5], edge detection is applied on the DEM to detect curb candidates as chains of straight-line segments of local height discontinuities. Temporal filtering is utilized, to deal with the blurring of the observed discontinuities, growing with increasing distance to the camera.

J. Siegemund and W. Förstner are with the University of Bonn, Institute of Geodesy and Geoinformation, Department of Photogrammetry, Germany. {jansiegemund, wf}@ipb.uni-bonn.de

U. Franke is with Daimler AG, Group Research and Advanced Engineering, Sindelfingen, Germany. uwe.franke@daimler.com



Fig. 1. A typical reconstruction result. The reconstructed curbstone is marked by red vertical lines, while the orange horizontal lines mark the attachment pieces of the neighboring surfaces.



Fig. 2. Two examples for partial occluded curbs. The stabilizing cubic curb model provides a prediction of the curb’s characteristics in occluded regions.

Michalke et al. [6] show how to extend the detection range by fusing 3D geometry with further vision based information. First, curb candidates near the vehicle are extracted from the DEM. The appearance of these candidates is then tracked within the image to predict their position in greater distances. Again, temporal integration is used for robustification.

An alternative solution to increase the detection range is presented in [7], demonstrating the advantages of a model-based approach. A parameterized curb model is fitted into the DEM providing a robust reconstruction of the curb’s horizontal and vertical geometry up to 20 meters from the camera (Figure 1). The measured height data is assigned to the different parts of the model using a Conditional Random Field (CRF) [8]. The curb’s horizontal shape is represented by a third order polynomial, stabilizing the reconstruction result even in the case of partial occlusions (Figure 2). In [9], Oniga et al. confirm the choice of a polynomial curb representation.

So far, the reconstruction in [7] is performed for each frame independently. In this contribution, we propose a temporal filtering process as extension for this algorithm to:

- Improve the robustness and accuracy, especially at the case of curbs of low height and missing or erroneous measurements.
- Reduce the computation time by distributing the computational effort over successive time steps.

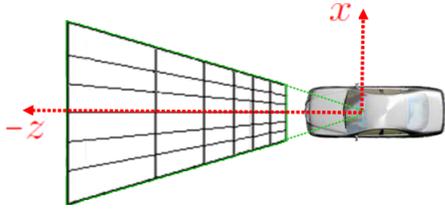


Fig. 3. Birdseye view of the DEM's horizontal grid structure, being regular and paraxial in the column-disparity domain.

The paper is structured as follows. First, the data acquisition and model assumptions are described in sections II and III. Then, a brief overview of the curb reconstruction approach presented in [7] is given in Section IV. In Section V we present the temporal filtering extension of this approach. The benefit of this extension is evaluated in Section VI, with respect to the lateral reconstruction accuracy in real world scenarios. Section VII concludes the paper and gives an outlook to future work.

## II. DATA ACQUISITION

The presented approach is based on 3D point clouds, which, in our experiments, are received from dense stereo vision. The image data is captured by a stereo camera system that is mounted behind the wind shield of a test vehicle, yielding 25 images per second. The baseline is approximately 0.3 m. For stereo computation we employ the implementation of Gehrig [10].

The origin of the local coordinate system  $K$ , containing the 3D points, is placed at ground level, straight under the origin of the left camera. The  $x$ -axis points right, the  $y$ -axis upwards, while the  $z$ -axis completes the right handed system pointing into negative driving direction as illustrated in Figure 3.

From the triangulated point cloud a DEM is generated, which we denote by  $\mathcal{M}$ . This grants two advantages. First, the amount of data is significantly reduced, providing real-time performance. Second, spatial neighborhood relations are modeled explicitly. The utilized DEM is defined as a horizontal grid, that is regular and paraxial in the column-disparity space  $(u, d)$  as demonstrated in Figure 3. This ensures an approximate constant number of observations assigned to each grid cell, opposed to a grid being regular and paraxial to the horizontal world axes.

We assign all image pixels  $(u_k, d_k)^T, k \in \Omega$  of a region of interest  $\Omega$  that have valid disparity values  $d_k$  to their nearest grid cells. Each grid cell  $i \in I$  is represented by its center  $(u_i, d_i)^T$ , or  $(x_i, z_i)^T$  with respect to the Cartesian space. From all triangulated height values  $y_k$  of the image points assigned to  $i$  a common height value  $y_i$  is computed using a histogram based approach. For a shorter notation we denote the vector of all height values  $y_i$  by  $\mathbf{h}$ .

Further a theoretical height accuracy  $\sigma_{y_i}$  is computed for each cell. This is done by error propagation via the triangulation concept, assuming a measurement accuracy  $\sigma_u = \sigma_v = 1/4$  pel within the image.

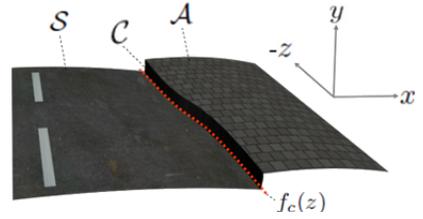


Fig. 4. Perspective view of the utilized environment model.

## III. ENVIRONMENT MODEL

As mentioned in Section I, model assumptions are introduced to stabilize the reconstruction result. We assume a curb  $C$  to be a vertical structure, that separates the street surface  $S$  from the adjacent horizontal surface  $\mathcal{A}$  (e.g. sidewalk or traffic isle) with respect to the  $x$ -axis, as illustrated in Figure 4.

More precisely, we define the horizontal shape  $f_c$  of the curb by the coefficients  $\mathbf{c} = [c_0, \dots, c_3]^T$  of a third order polynomial

$$x = f_c(z) = [z^3, z^2, z, 1]\mathbf{c}, \quad (1)$$

specifying a left-right separation of  $S$  and  $\mathcal{A}$ . In the remainder of this contribution, the curb is assumed to be located on the right hand side of the road. A curb on the left hand side can be dealt with in an analogous way.

Using the vector  $\mathbf{q} = [x^2, z^2, xz, x, z, 1]^T$ , we represent  $S$  and  $\mathcal{A}$  by bounded second order surfaces  $g_s$  and  $g_a$

$$\mathcal{A} = \{[x, y, z]^T \mid y = g_a(x, z) := \mathbf{a}^T \mathbf{q}, x \geq f_c(z)\} \quad (2)$$

$$\mathcal{S} = \{[x, y, z]^T \mid y = g_s(x, z) := \mathbf{s}^T \mathbf{q}, x \leq f_c(z)\}, \quad (3)$$

with  $\mathbf{a} = [a_0, \dots, a_5]^T, \mathbf{s} = [s_0, \dots, s_5]^T$  being the unknown surface parameters.

We use the notation  $\Theta = (\mathbf{c}, \mathbf{a}, \mathbf{s})$  to combine the set of unknown model parameters.

## IV. CURB RECONSTRUCTION APPROACH

In this section we briefly recapitulate the algorithm presented in [7]. The total workflow, including the temporal filtering presented in Section V, is illustrated in Figure 6.

The unknown model parameters  $\Theta$  are estimated in an iterative two step approach, in the manner of an Expectation Maximization algorithm [11]. Starting with an initial labeling  $\mathbf{l}^{(0)}$  the two successive steps

- Estimation of the unknown model parameters  $\Theta^{(\nu)}$  based on the labeling  $\mathbf{l}^{(\nu-1)}$ .
- Classification  $\mathbf{l}^{(\nu)} = [l_1^{(\nu)}, \dots, l_I^{(\nu)}]$  of the DEM cells to labels  $l_i^{(\nu)} \in \Lambda = \{\text{'street'}, \text{'street adjacent'}, \text{'unassigned'}\}$  based on information extracted from  $\Theta^{(\nu)}$ .

are iteratively performed, until a termination criteria is fulfilled or a maximum number of iterations is reached. The variable  $\nu \in \{1, \dots, \nu_{max}\}$  denotes the iteration counter.

The labels 'street' and 'street adjacent' represent the affiliation of cells to the surfaces  $S$  and  $\mathcal{A}$ , while 'unassigned' tags cells containing vertical structures or measurement errors. The initial labeling  $\mathbf{l}^{(0)}$  is given by the final labeling result

of the last frame, if available. Otherwise it is simply 'street' for all cells left of the cars lateral center and 'street adjacent' for those on the right hand side.

In the remainder of this section we refer to both successive steps in more detail.

### A. PARAMETER ESTIMATION STEP

1) *Estimation of the surface parameters:* Given the class assignment  $\mathbf{l}^{(\nu-1)}$ , the model parameters  $\Theta^{(\nu)}$  can be estimated directly from the DEM.

The surface parameters  $\mathbf{s}^{(\nu)}$  are estimated from the Cartesian coordinates  $[x_i, y_i, z_i]$  of all cells assigned to 'street' in a weighted least squares sense

$$\mathbf{s}^{(\nu)} = \underset{\mathbf{s}}{\operatorname{argmin}} \left( \sum_{i \in I_s^{(\nu)}} \frac{1}{\sigma_{y_i}^2} (y_i - \mathbf{s}^\top \mathbf{q}_i)^2 \right), \quad (4)$$

with  $\mathbf{q}_i = [x_i^2, z_i^2, x_i z_i, x_i, z_i, 1]^\top$ . In a similar manner, we derive  $\mathbf{a}^{(\nu)}$  from all cells assigned to 'street adjacent'.

Further, the variances  $\sigma_s^{2(\nu)}$  and  $\sigma_a^{2(\nu)}$  of the measured height values with respect to the estimated surfaces are computed.

2) *Estimation of the curb parameters:* The parameters  $\mathbf{c}^{(\nu)}$  of the horizontal shape of the curb can be estimated as the horizontal left-right separation of the classes 'street' and 'street adjacent'. This is done by logistic regression, formulating  $f_c$  as zero level of the sigmoidal function

$$g_{b,c}(x, z) = \frac{2}{1 + \exp(b(f_c(z) - x))} - 1, \quad (5)$$

The constant  $b$  controls the steepness of the sigmoid. Using artificial height values  $y_i = -1$  for all cells assigned to 'street' and  $y_i = +1$  for all those assigned to 'street adjacent', we compute a least squares estimate  $\mathbf{c}^{(\nu)}$  similar to Equation (4).

Finally, lower and upper bounds of the curb can be directly computed from the vertical intersection of  $f_c$  with  $\mathcal{S}$  and  $\mathcal{A}$ , as shown in Figure 5.

### B. CLASSIFICATION STEP

In this step, the objective is to find a labeling  $\mathbf{l}^{(\nu)}$  from the set of all possible labelings  $\mathcal{L}$ , that maximizes the conditional probability  $p(\mathbf{l}|\mathbf{h}, \Theta^{(\nu)})$ , i.e.  $\mathbf{l}^{(\nu)} = \operatorname{argmax}_{\mathbf{l} \in \mathcal{L}} p(\mathbf{l}|\mathbf{h}, \Theta^{(\nu)})$ . We model this probability by means of a CRF that is aligned to the DEM's grid. Each DEM cell corresponds to one graph node, while each pair of neighboring cells, according to the DEM's 4-neighborhood  $\mathcal{N}_4$ , corresponds to one edge. With this,  $p(\mathbf{l}|\mathbf{h}, \Theta^{(\nu)})$  can be written as a product of unary potential functions  $\Phi$  and binary potential functions  $\Psi$

$$p(\mathbf{l}|\mathbf{h}, \Theta^{(\nu)}) \propto \prod_{i \in I} \Phi(l_i|y_i, \Theta^{(\nu)}) \prod_{(i,j) \in \mathcal{N}_4} \Psi(l_i, l_j|y_i, y_j, \Theta^{(\nu)}). \quad (6)$$

In the following we will briefly sketch the influence of the unary and binary functions and refer to [7] for a detailed formulation.

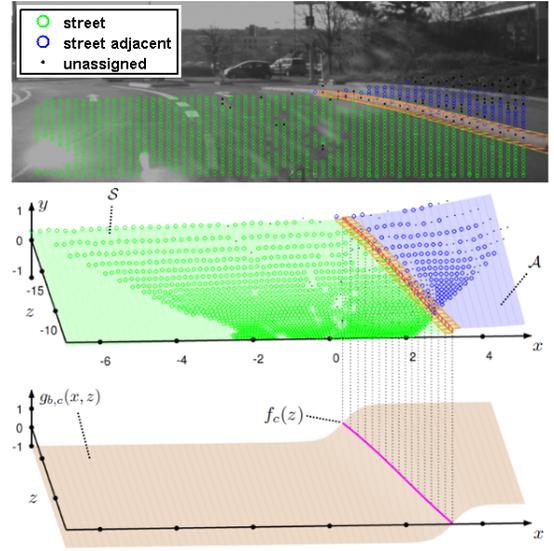


Fig. 5. Visualization of the reconstruction result. Top: Projection of the final labeling and estimated curb into the image. Center: 3D view of the labeled DEM, containing the reconstructed surfaces  $\mathcal{S}$  and  $\mathcal{A}$ . Bottom: Estimated sigmoidal function  $g_c$ , defining the curbs horizontal position by its zero level  $f_c$ . The curbstone is reconstructed from the vertical intersection of  $f_c$  with  $\mathcal{S}$  and  $\mathcal{A}$ .

The unary potential functions  $\Phi$  define the local, individual labeling decision at each cell, which is based on several criteria:

- A cell is the more likely assigned to a surface the smaller the distance between the cell's measured height value  $y_i$  and the estimated height of the surface  $g_s^{(\nu)}(x_i, z_i)$ , respectively  $g_a^{(\nu)}(x_i, z_i)$ .
- The probability of the label 'unassigned' is dominant if the distance to both surfaces is larger than  $3\sigma_{y_i}$ .
- The surface having a lower height variance  $\sigma_s^{2(\nu)}$ , respectively  $\sigma_a^{2(\nu)}$ , is preferred.
- The more the cell lies on the right side of  $f_c^{(\nu)}$ , the more unlikely is its assignment to 'street' and vice versa for 'street adjacent'.

The binary potentials  $\Psi$  are defined using an height difference sensitive Potts model. I.e., neighboring cells  $i$  and  $j$  are the more likely assigned with the same label, the smaller the weighted height difference  $\frac{|y_i - y_j|}{\sigma_{d_{ij}}}$ , with  $\sigma_{d_{ij}} = \sqrt{\sigma_{y_i}^2 + \sigma_{y_j}^2}$ .

For inference we utilize Loopy Belief Propagation (LBP) [12] (pp. 334-340). As result, we obtain an estimation of the most probable labeling  $\mathbf{l}^{(\nu)}$ , as well as estimated marginal probabilities for each cell and for all possible assignments of labels

$$\mathcal{P}^{(\nu)} = \left\{ p(l_i^{(\nu)} = \iota | \mathbf{h}, \Theta^{(\nu)}) \mid i = 1, \dots, I, \iota \in \Lambda \right\}. \quad (7)$$

### V. TEMPORAL FILTERING

The results in [7] are derived from the model presented in the last section. Observe no temporal continuity is guaranteed, as curb and surface estimation and classification does not use information from the past time step.

In this section, a temporal filtering process for both steps of the presented curb reconstruction approach is proposed. The schematic overview of the total process is given in Figure (6).

We assume the variation of the observed curb and surface characteristics to be small between successive frames. Thus, using information about the cars ego motion, the position and shape of the curb, as well as the class regions at the current time step  $t$  can be predicted from the results estimated at the previous acquisition time  $t - 1$ .

Let  $\gamma$  be the rotation angle and  $\mathbf{T} = [T_x, T_z]^T$  be the translation vector defining the planar motion of the ego car from  $t - 1$  to  $t$ . Then, the coordinates of a point in the current reference system  $K_t$  can be transformed into the previous reference system  $K_{t-1}$  using

$$\begin{bmatrix} x_{i,t-1} \\ z_{i,t-1} \end{bmatrix} = \begin{bmatrix} \cos(\gamma) & \sin(\gamma) \\ -\sin(\gamma) & \cos(\gamma) \end{bmatrix} \begin{bmatrix} x_{i,t} \\ z_{i,t} \end{bmatrix} + \begin{bmatrix} T_x \\ T_z \end{bmatrix}. \quad (8)$$

In our experiments we derive  $\gamma$  and  $\mathbf{T}$  from inertial sensors. Here, a circular path motion model, assuming constant yaw rate and velocity, is used.

#### A. FILTERING OF THE CLASSIFICATION STEP

For the temporal filtering of the classification step, we assume the labels of the considered regions to be constant with respect to a fixed world frame. Thus, given the ego motion, we are able to predict the class probabilities for the current DEM cells.

The basic idea is to connect the CRF to the results of the last time step and transfer the information contained in the node marginals  $\mathcal{P}_{t-1}$  into the current labeling task. We model this connection using an additional potential function  $\Upsilon$ , extending the probability defined in (6) to

$$p\left(l_t | \mathbf{h}_t, \Theta_t^{(\nu)}, \mathbf{h}_{t-1}, \Theta_{t-1}\right) \propto p\left(l_t | \mathbf{h}_t, \Theta_t^{(\nu)}\right) \prod_{i \in I} \Upsilon(l_{i,t} | y_{i,t-1}, \Theta_{t-1}). \quad (9)$$

We determine  $\Upsilon$  from the marginals  $\mathcal{P}_{t-1}$  performing the following steps for all cells  $i$  of the current DEM  $\mathcal{M}_t$ :

- Compute the Cartesian coordinates  $[x_{i,t-1}, z_{i,t-1}]^T$  with respect to  $K_{t-1}$ , using (8).
- Transform these coordinates into column disparity space  $[u_{i,t-1}, d_{i,t-1}]^T$ .
- If  $[u_{i,t-1}, d_{i,t-1}]^T$  lies inside the previous DEM  $\mathcal{M}_{t-1}$ :
  - Identify its four nearest neighbor cells  $n_1, \dots, n_4 \in \mathcal{M}_{t-1}$ .
  - For all  $\ell \in \Lambda$ : Interpolate  $\Upsilon(l_{i,t} | y_{i,t-1}, \Theta_{t-1})$  from the marginals  $p(l_{j,t-1} = \ell | \mathbf{y}_{t-1}, \Theta_{t-1}) \in \mathcal{P}_{t-1}$ , with  $j = n_1, \dots, n_4$ .
- Otherwise: Set  $\Upsilon(l_{i,t} | y_{i,t-1}, \Theta_{t-1}) = 1, \forall \ell \in \Lambda$ , i.e. the influence of the prior term for the current cell is switched off.

Linking the CRF to the results of the previous time step in the presented manner results in several advantages. First, the robustness of the classification increases, especially for curbs of low height. Second, the inference procedure sets up on

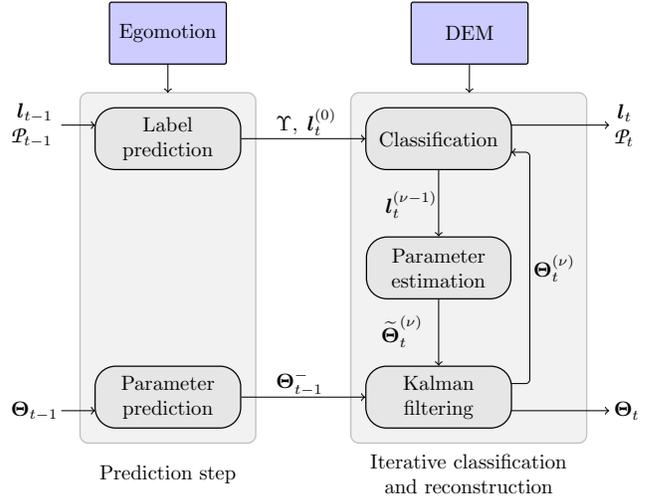


Fig. 6. Workflow of the proposed reconstruction approach.

the classification result of the last time step instead of propagating the information from scratch. With this, the amount of iterations needed by the inference method is significantly reduced, amortizing the additional computational effort of the temporal integration. We can think of this as distributing the classification task over successive time steps. Further, for each cell, we can use the label maximizing the respective prior term  $\Upsilon$  to define the initial labeling  $l_t^{(0)}$  discussed in section IV.

#### B. FILTERING OF THE PARAMETER ESTIMATION STEP

We make two assumptions for the filtering of the model parameters. First, the position and shape of the retrieved parts of the model is assumed to be constant with respect to the world frame. Second, the change of the characteristics in the new observed parts is assumed to be small.

The filtering is performed by means of a Kalman filter. In the following, we use the notation  $\Theta$  for parameters estimated by the approach presented in Section IV-A,  $\Theta^-$  for predicted parameters and  $\Theta$  for the final, filtered parameters.

1) *FILTERING OF THE CURB PARAMETERS*: The mathematical proper approach for a temporally filtered estimation of  $\mathbf{c}_t$  would be to completely reformulate the procedure described in IV-A.2 using a Kalman filter. Unfortunately, this would require a high computational effort since the innovation matrix, whose dimension is equal to the number of observed height values  $|I|$ , must be inverted in each iteration.

We avoid this by directly filtering the estimated parameters  $\tilde{\mathbf{c}}$ . This leads to a simple measurement model that is given by the identity function  $\tilde{\mathbf{c}}_t = \mathbf{c}_t + \mathbf{v}_t$ , with a Gaussian white noise term  $\mathbf{v}_t$ .

The system model is derived from substituting the Cartesian coordinates in Equation (1) using the right side of (8), yielding

$$\begin{aligned} \cos(\gamma)x_{i,t} + \sin(\gamma)z_{i,t} + T_x = \\ f_{c_{t-1}}(-\sin(\gamma)x_{i,t} + \cos(\gamma)z_{i,t} + T_z) \end{aligned} \quad (10)$$

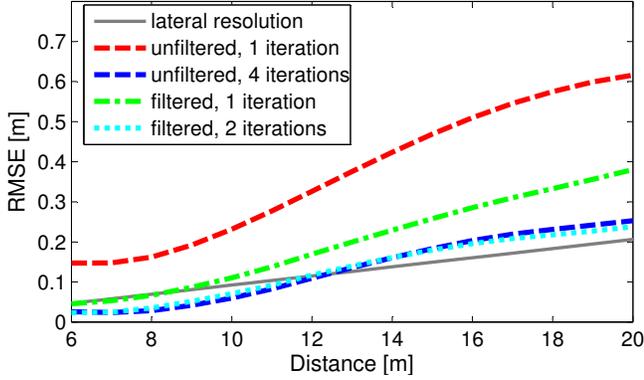


Fig. 7. Joint lateral RMSE derived from a comparison with manually generated ground truth, plotted against the distance to the camera. The RMSE is plotted for the filtered (light blue and green) and unfiltered approach (dark blue and red), using varying numbers of iterations. The utilized lateral DEM resolution is plotted by the solid gray line.

Neglecting all terms including products of  $x_{i,t}$  and  $z_{i,t}$ , we can reorganize (10) to

$$x_{i,t} = [z_{i,t}^3, z_{i,t}^2, z_{i,t}, 1] \mathbf{c}_t^- = f_{\mathbf{c}_t^-}(z_{i,t}). \quad (11)$$

where  $\mathbf{c}_t^-$  is given as

$$\mathbf{c}_t^- \approx \begin{bmatrix} c'_0 \cos^3(\gamma) \\ (3c'_0 T_z + c'_1) \cos^2(\gamma) \\ (3c'_0 T_z^2 + 2c'_1 T_z + c'_2) \cos(\gamma) - \sin(\gamma) \\ c'_0 T_z^3 + c'_1 T_z^2 + c'_2 T_z + c'_3 - T_x \end{bmatrix} / \tau \quad (12)$$

$$:= \hat{h}(\mathbf{c}_{t-1}), \quad (13)$$

using the substitution

$$\tau = (3c'_0 T_z^2 + 2c'_1 T_z + c'_2) \sin(\gamma) + \cos(\gamma) \quad (14)$$

and the short notation  $[(c_0)_{t-1}, \dots, (c_3)_{t-1}] = [c'_0, \dots, c'_3]$ .

We obtain the final system model from (13) by adding a Gaussian white noise term  $\mathbf{w}_t$  to capture the influence of the neglected terms

$$\mathbf{c}_t^- = \hat{h}(\mathbf{c}_{t-1}) + \mathbf{w}_t. \quad (15)$$

This influence is assumed to be small, since the neglected terms all contain a factor that is a power of the sine of the small angle  $\gamma$ .

## 2) FILTERING OF THE SURFACES PARAMETERS:

For the filtering of the surface parameters  $\mathbf{s}_t$  and  $\mathbf{a}_t$  a full 3D ego motion estimation is required, because the influence of pitch and roll rotations cannot be neglected as by the horizontal structures discussed before.

In case reliable ego-motion information is available one can proceed as follows for filtering of the street surface parameters  $\mathbf{s}_t$  (the parameters  $\mathbf{a}_t$  can be dealt with in an analogous way).

Assume a 3D ego motion matrix  $M$  to be given, defining the homogeneous transformation

$$\mathbf{X}_{i,t-1} = M \mathbf{X}_{i,t}, \quad (16)$$

with the homogeneous points  $\mathbf{X}_{i,t} = [x_{i,t}, y_{i,t}, z_{i,t}, 1]^T$ .



Fig. 8. Example showing the limitations of the utilized curb model. The assumptions are violated to much by the geometry of the traffic isle to allow a proper reconstruction.

The surface equation  $y_{i,t-1} = g_{s_{t-1}}(x_{i,t-1}, z_{i,t-1})$  defined in (3) can be formulated as implicit condition

$$0 = \mathbf{X}_{i,t-1}^T \mathbf{S}_{t-1} \mathbf{X}_{i,t-1}, \quad (17)$$

using

$$\mathbf{S}_{t-1} = \begin{bmatrix} (s_0)_{t-1} & 0 & \frac{1}{2}(s_2)_{t-1} & \frac{1}{2}(s_3)_{t-1} \\ 0 & 0 & 0 & -\frac{1}{2} \\ \frac{1}{2}(s_2)_{t-1} & 0 & (s_1)_{t-1} & \frac{1}{2}(s_4)_{t-1} \\ \frac{1}{2}(s_3)_{t-1} & -\frac{1}{2} & \frac{1}{2}(s_4)_{t-1} & (s_5)_{t-1} \end{bmatrix} \quad (18)$$

Substituting (16) into (17) we obtain

$$0 = \mathbf{X}_{i,t}^T M^T \mathbf{S}_{t-1} M \mathbf{X}_{i,t} = \mathbf{X}_{i,t}^T \mathbf{S}_t \mathbf{X}_{i,t} \quad (19)$$

and can directly extract the predicted parameters  $\tilde{\mathbf{s}}_t$  from the respective elements of  $\mathbf{S}_t$ . With this, the measurement model and system model of the Kalman filter can be defined analogously to the filtering of the curb parameters in the previous section.

In our experiments, we forbear from filtering the surface parameters  $\mathbf{s}_t$  and  $\mathbf{a}_t$  due to two reasons. First, the 3d-ego-motion estimation requires an additional computational effort. Second, since the labeling decision depends on the distance of the measured height values  $y_i$  to the estimated surfaces, it reacts very sensitive to effects of slightly biased ego motion parameters.

## VI. RESULTS

The proposed method was implemented in C++ and compared to the original approach presented in [7].

We use a DEM of  $64 \times 32$  (column  $\times$  disparity) cells, providing height information up to 20 meter distance. The resulting lateral resolution is plotted in Figure 7. It takes 2-3 ms to compute the DEM. The computation time for a single iteration of the reconstruction process on recent PC hardware ( $4 \times 3\text{GHz}$  Intel Core2 Quad), is given by 6-7 ms for both approaches. This is because the additional computational effort needed for the filtering steps is amortized by the resulting speedup of the inference method, as mentioned in Section V-A. Further, the propagated information reduces the number of iterations needed for convergence. In our experiments, we can restrict the filtered approach to a maximum number of 2 iterations, while the unfiltered approach usually needs 4 iterations to converge.

In comprehensive tests on real-world scenarios, the filtered method show considerable improvements concerning the smoothness and robustness of the reconstruction result. Jerky leaps and discontinuities between the curbs geometry

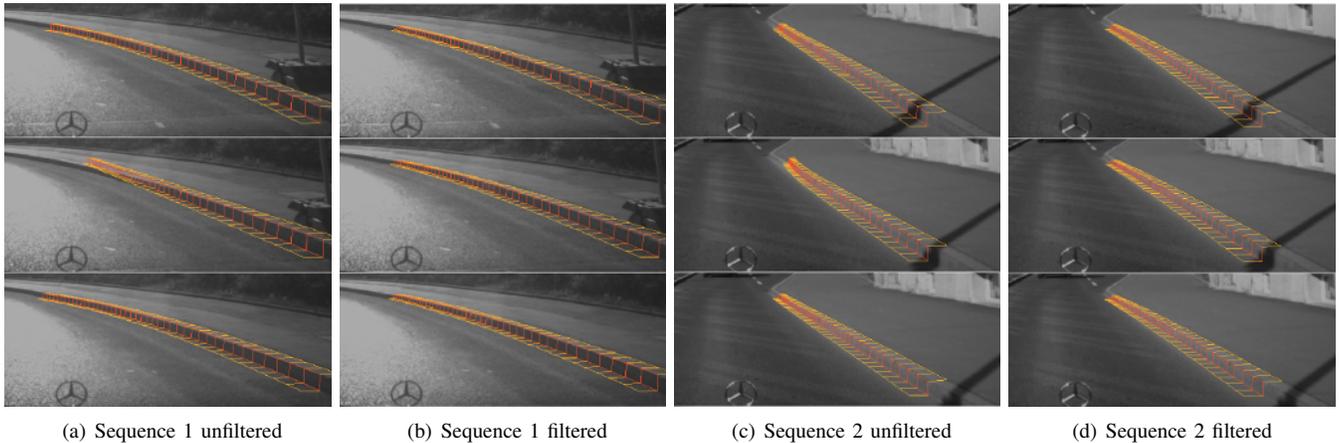


Fig. 9. Two exemplary image sequences showing jerky leaps between the reconstruction results of successive frames (column-wise), using the unfiltered approach (columns (a) and (c)). These effects are not shown by the respective results of the temporal filtered method (columns (b) and (d)).

estimated from successive frames are reduced significantly as demonstrated in Figure 9.

The lateral accuracy was evaluated on a set of nine different real-world scenarios, containing curved and straight line curbs, having heights between 4 and 15 cm. Each scenario consists of a sequence of 20 frames. For all images the position of the curbs bottom edge was annotated by an expert for each image row. These annotated bottom edge is then projected on the estimated street surface, yielding lateral ground-truth information. From this, we evaluate the lateral accuracy by computing the lateral error of the respective curb positions estimated by the proposed reconstruction approach. In Figure 7, the joint Root Mean Squared Error (RMSE) concerning all frames of all scenarios is plotted against the distance to the camera.

The result shows that the filtered approach significantly outperforms the unfiltered method, when using just one iteration (green and red dashed lines). Further, the unfiltered method requires four iterations, to reach results comparable to the filtered method after two iterations (blue dashed and dotted lines). We observe the lateral error to be approximately given by the half of the lateral resolution near the camera, exceeding this resolution in greater distances. This is an intuitive result, considering the interpolation properties of the sigmoid (5) on the one hand and the exponential growing of the measurement noise with increasing distance to the camera on the other hand.

## VII. CONCLUSION AND FUTURE WORK

We proposed a temporal filter approach for robust detection and reconstruction of curbs and street surfaces in real-time. In experiments on real-world scenarios, the advantages over an unfiltered version, due to robustness, accuracy and computational effort, were demonstrated.

The evaluation of the lateral accuracy have shown the ability of the method to yield accurate reconstruction results up to distances of 20 meters to the camera.

Future work is required for the development of a more general and flexible curb model. The utilized curb model

is designed for curbs being collateral to the vehicles driving corridor. This limits the ability of the approach to reconstruct curbs violating this assumption to much, such as small traffic isles as illustrated in Figure 8.

Furthermore, the fusion with additional vision based information, e.g. texture and gradient information, will be part of future work.

## REFERENCES

- [1] C. Yu and D. Zhang, "Road curbs detection based on laser radar," in *8th International Conference on Signal Processing (ICSP)*, 2006, pp. 16–20.
- [2] W. Zhang, "Lidar-based road and road-edge detection," in *Intelligent Vehicles Symposium (IV)*, 2010, pp. 845–848.
- [3] O. Gallo, R. M. and A. Rafii, "Robust curb and ramp detection for safe parking using the canesta tof camera3rd," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008, pp. 1–8.
- [4] A. Huang and S. Teller, "Lane boundary and curb estimation with lateral uncertainties," in *Intelligent Robots and Systems (IROS)*, 2009, pp. 1729–1734.
- [5] F. Oniga, S. Nedeveschi, and M. Meinecke, "Curb detection based on a multi-frame persistence map for urban driving scenarios," in *11th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE Computer Society, 2008, pp. 67–72.
- [6] T. Michalke, R. Kastner, J. Fritsch, and C. Goerick, "A self-adaptive approach for curbstone/roadside detection based on human-like signal processing and multi-sensor fusion," in *Intelligent Vehicles Symposium (IV)*. IEEE Computer Society, 2010.
- [7] J. Siegemund, D. Pfeiffer, U. Franke, and W. Förstner, "Curb reconstruction using conditional random fields," in *Intelligent Vehicles Symposium (IV)*. IEEE Computer Society, 2010, pp. 203–210.
- [8] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *18th International Conference on Machine Learning (ICML)*. Morgan Kaufmann Publishers Inc., 2001, pp. 282–289.
- [9] F. Oniga and S. Nedeveschi, "Polynomial curb detection based on dense stereovision for driving assistance," in *13th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE Computer Society, 2010, pp. 1110–1115.
- [10] S. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching," in *International Conference on Computer Vision Systems (ICVS)*. Springer-Verlag, 2009, pp. 134–143.
- [11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [12] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.