

Curb Reconstruction using Conditional Random Fields

Jan Siegemund¹, David Pfeiffer², Uwe Franke², and Wolfgang Förstner¹

¹University of Bonn, Department of Photogrammetry,

Institute of Geodesy and Geoinformation, Bonn, Germany

²Daimler AG, Group Research and Advanced Engineering,
Sindelfingen, Germany

Abstract—This paper presents a generic framework for curb detection and reconstruction in the context of driver assistance systems. Based on a 3D point cloud, we estimate the parameters of a 3D curb model, incorporating also the curb adjacent surfaces, e.g. street and sidewalk. We apply an iterative two step approach. First, the measured 3D points, e.g., obtained from dense stereo vision, are assigned to the curb adjacent surfaces using loopy belief propagation on a Conditional Random Field. Based on this result, we reconstruct the surfaces and in particular the curb. Our system is not limited to straight-line curbs, i.e. it is able to deal with curbs of different curvature and varying height.

The proposed algorithm runs in real-time on our demonstrator vehicle and is evaluated in urban real-world scenarios. It yields highly accurate results even for low curbs up to 20 m distance.

I. INTRODUCTION

Robust detection of obstacles endangering the driver's safety is an essential task for driving assistance. Although curbs in general are of low height a collision is a potential risk of severe tire damage. Even minor damages bear the risk of delayed blowouts and may result in critical situations. Furthermore, curbs usually represent the boundary between driving lane and sidewalk. For this reason they are of special interest for traffic scene interpretation tasks.

Approaches modeling the car's free driving space, as for example the Stixel World representation by Badino et al. [1], are designed to model objects with a certain minimum height to be robust against artifacts caused by measurement noise. This makes them unsuitable for the curb detection task due to the curb's low height occurrence.

There are several approaches addressing this task explicitly. Se and Brady [2] detect curb candidates from clusters of parallel straight lines in the image. The lines are extracted via a Hough transformation [3] from edge points detected by means of a Canny edge detector [4]. Using brightness information exclusively is risky since straight lines in the image may also result from lane markings and pavement transitions. In a similar approach, Turchetto and Manduchi [5] additionally include 3D information. The vote for each single edge point within the Hough accumulator is weighted by a function of the brightness gradient and the 3D elevation gradient. This idea is extended by introducing the estimated surface curvature to the weighting function [6].

However, not all methods are based on the image domain. In [7] the 3D points are arranged in an horizontal height



Fig. 1. Example of a reconstructed curb. The estimated curb stone edge is displayed by red vertical lines. The attachment pieces of street and sidewalk surface are denoted by the orange horizontal lines.

grid, denoted as Digital Elevation Map (DEM). For each grid cell a height value is determined. Curb candidates are then detected from discontinuities between heights of neighboring cells. This is done by using the combination of a Canny edge detector and a Hough transformation on the DEM.

All these approaches have in common that they are restricted to the detection of straight-line curbs. Furthermore the detection of curbs via the elevation gradient is very sensitive to artifacts caused by measurement noise. The observed height discontinuity gets blurred with increasing distance to the camera. Thus, for the detection of low and distant curbs the threshold of the edge detector needs to be scaled down. However, this results in many false curb candidates. Oniga et al. show how to remove these false candidates by temporal integration if reliably egomotion information exists [8]. They present a robust and real-time capable approach to detect curbs of a constant height of at least 5 cm up to a distance of 10 meters. Even curved curbs are modeled using chains of straight-line curb segments.

In this contribution we present a new approach to detect and reconstruct curbs even of low height up to a distance of 20 meters. This includes curved curbs as well as those with non constant height as shown in Figure 1. Our method is based on 3D point observations, e.g., received from dense stereo vision, which we arrange in a DEM similar to [8]. The underlying idea is that even at great distances the average measured height levels of the curb adjacent surfaces, e.g., street and sidewalk, still differ. We exploit this fact by using a parameterized environment model, where the curb is defined as horizontal separation of its adjacent surfaces. In order to reconstruct the curb we need to determine the parameters of this model from the measured height data. Therefore we

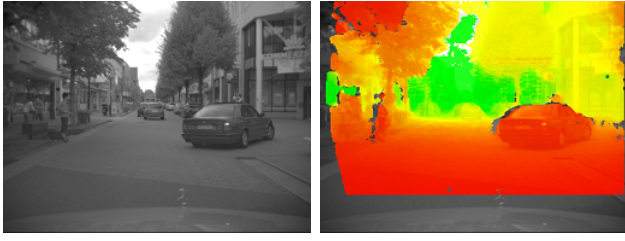


Fig. 2. Original input image and result of the SGM stereo algorithm. The color encoding denotes red for close and green for far disparity measurements.

assign each cell of the DEM to one part of the environment model using a Conditional Random Field (CRF) [9]. CRFs have successfully been used for road scene interpretation, e.g., in [10] or [11]. Contrary to these methods, we align the CRF's graph structure to the grid of the DEM, which has a considerably smaller resolution than the camera image. Thereby we reach real-time performance.

The remainder of this paper is structured as follows. Sections II and III describe the data acquisition and construction of the DEM. Section IV comprises the definition of the environment model. Section V presents an iterative two-step approach for simultaneous classification and reconstruction. Results are presented and discussed in Section VI. Section VII concludes the paper.

II. DATA ACQUISITION

In our contribution the 3D point cloud data is computed from dense stereo vision. Other sensors that capture punctual data in real-time, e.g., laser scanners, would also be suitable. We use the implementation of [12] which is based on the Semi-Global Matching algorithm (SGM) of [13]. This implementation runs on a Xilinx FPGA platform at 25 Hz with a power consumption of less than 3 W. The image data (VGA images) is captured by a stereo camera system, with 0.3 m baseline. The cameras are mounted behind the windshield of a test vehicle, pointing in the direction of the vehicle's front. In Figure 2 an exemplary scene is illustrated with the corresponding SGM result given besides. We define the coordinate system for the 3D reconstructed points relative to the viewing direction of the camera as follows. The x -axis points right, the y -axis points upwards, and the z -axis completes the right-handed system pointing into the camera and thus in negative viewing direction. The origin is positioned on ground level straight under the projection center of the left camera.

III. DIGITAL ELEVATION MAPS

Grid based representations are an established method to model the environment and to concentrate and prepare information for data fusion. In [14], Badino et al. present a technique for modeling occupancy evidence for the vehicle environment. Oniga et al. employ Digital Elevation Maps not to model the degree of occupancy but to determine the height of the area observed [15], [7].

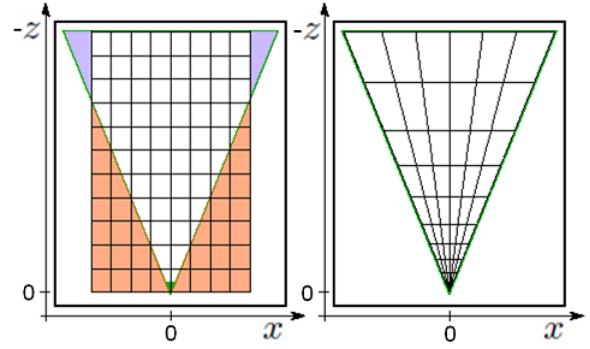


Fig. 3. A Cartesian grid (left) and a column disparity grid (right) are illustrated in a top down view. The green cone represents the field of view (FOV) of the camera. The colors denote areas that lie either outside of the FOV (red), or are not covered by the grid (blue).

These grids can be defined over different coordinate systems. A common approach is to align the DEM grid paraxial to the Cartesian xz -plane and to assign a constant width and length to every cell of the DEM. These grids are comfortable to handle but entail the issue that distant cells tend to get too small (within the image), i.e., only a few stereo disparities can be associated with them, while close cells - the ones that are of particular interest in driver safety tasks - are still too clumsy. A possible way out is to define different cell sizes and/or to use hierarchical grid structures like quadrees or kd-trees. However, this would derogate the convenience in the usage of the DEM due to the resulting irregular grid structure and would cause a higher computational effort.

Another option is to define the grid to be regular and paraxial to the column disparity space (u, d) . Using this domain for the grid structure, one intrinsically deals with those mentioned perspective effects. All cells remain the same size within the image, resulting in small close and bigger distant cells with respect to the Cartesian xz -plane. These two different methods are illustrated within Figure 3. Another benefit of the column disparity grid representation is, by definition, that it does not hold any cells that lie outside the field of view of the camera.

In this paper the column disparity based grid representation is used. We assign all stereo disparities d_k belonging to image pixels $(u_k, v_k)^T$ within the rectified image Ω to their corresponding grid cells. A grid cell is represented by its center $(u_i, d_i)^T$ and its index $i \in I$. Then we triangulate $(u_k, v_k, d_k)^T$ to their 3d coordinates $(x_k, y_k, z_k)^T$ and register the obtained height value y_k to the corresponding cell i . The coordinates of the cell centers with respect to our 3d world coordinate system are denoted by $\mathbf{x}_i = (x_i, h_i, z_i)^T$. Here we use the alias $h_i := y_i$ for the cell's height value and denote the vector of all height values of the grid by \mathbf{h} . Since x_i and z_i are independent of the image row they are directly derived from $(u_i, d_i)^T$.

Numerous strategies exist to determine the final height measurement h_i from all height values y_k registered to a cell i . A mean based approach is a straightforward method, but suffers from outliers and is likely to result in values that have

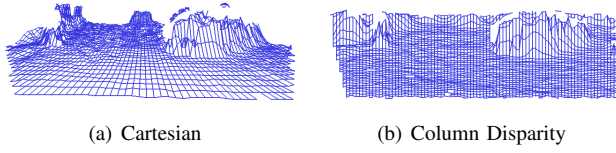


Fig. 4. (a) Cartesian and (b) column disparity based DEM of the scene illustrated in Figure 2. Both grids consist of the same amount of cells. One can observe, that the Cartesian cells are relatively big in the front and get tiny in the distance, while the column disparity based cells roughly remain the same size.

not been measured at all. The use of median or histogram based approaches is more reliable. For our purpose we use histograms registering all heights y_k in the range of $y_k \in [-0.5, 1.5]$ m with a discretization of 1 cm. We obtain the final height measurement from the histogram bin containing the maximum value, after smoothing the histogram by means of a Gaussian kernel.

When the number of heights registered to a cell is beneath a certain value (e.g. due to occlusions) the cell is flagged as invalid and no height information is determined for that cell. These invalid cells are ignored at all further computation steps.

The Cartesian and column disparity DEMs, computed from the disparity input image depicted in Figure 2, are illustrated in Figure 4.

The projection of each cell center x_i onto the left and right image plane is given from the known camera geometry. By means of the triangulation concept and an assumed measurement accuracy σ_u and σ_v within the image, we derive the theoretical height accuracy σ_{h_i} for each cell by error propagation. In our experiments, we assume $\sigma_u = \sigma_v = 1/4$ pixel. The covariance matrix containing all variances $\sigma_{h_i}^2, i \in I$, is denoted by $\Sigma_{hh} = \text{diag}(\sigma_{h_1}^2, \dots, \sigma_{h_I}^2)$.

IV. STATIC ENVIRONMENT MODEL

For simplicity we assume for now that there is just a single curb to find on the right hand side of the car. The detection of an additional curb on the left hand side will be addressed in Section V-D. In general a curb \mathcal{C} separates the street surface \mathcal{S} from a street adjacent surface \mathcal{A} , such as a sidewalk or a traffic isle, as shown in Figure 5. We model \mathcal{C} as a vertical structure, that follows a third order polynomial in the xz -plane with respect to the z -axis. The vertical extent of the structure is limited by upper and lower bounds $U(z)$ and $L(z)$. Therefore the curb model is given by

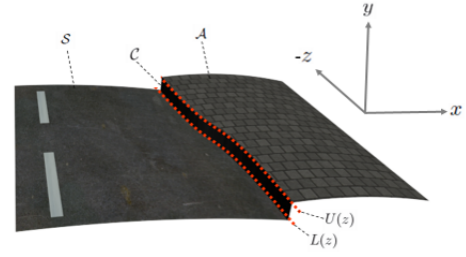
$$\mathcal{C} = \{[x, y, z]^T | x = f_c(z), L(z) < y < U(z)\}, \quad (1)$$

where $f_c(z) := c_0 z^3 + c_1 z^2 + c_2 z + c_3$, with the parameters $\mathbf{c} = [c_0, c_1, c_2, c_3]^T$, defines the third order polynomial. Further we model \mathcal{S} and \mathcal{A} by bounded second order surfaces. Using the substitution $\mathbf{q} = [x^2, z^2, 2xz, x, z, 1]^T$ we have

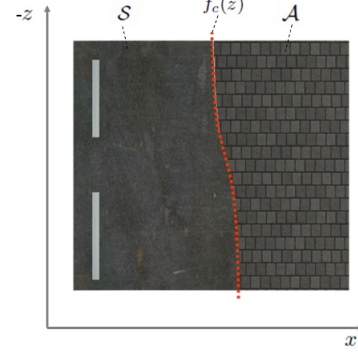
$$\mathcal{A} = \{[x, y, z]^T | y = g_a(x, z) := \mathbf{a}^T \mathbf{q}, x \geq f_c(z)\} \quad (2)$$

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

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



(a) Environment Model: Perspective View



(b) Environment Model: Birds Eye View

Fig. 5. Illustration of the environment model comprising the surfaces curb \mathcal{C} , street \mathcal{S} , street adjacent \mathcal{A} , the upper and lower bounds $U(z), L(z)$ of the curb and its projection $f_c(z)$ onto the xz -plane.

Since the upper and lower bounds of the curb are given by the heights g_s and g_a of the surfaces \mathcal{S} and \mathcal{A} we may express L and U in terms of the surface and curb parameters by combining equation (1) with (2) and (3) to

$$L(z) = \min(g_a(f_c(z), z), g_s(f_c(z), z)) \quad (4)$$

$$U(z) = \max(g_a(f_c(z), z), g_s(f_c(z), z)) \quad (5)$$

Note that the entire model is sufficiently described by the parameters \mathbf{a}, \mathbf{s} and \mathbf{c} .

When dealing with slanted curbs, our model limits the reconstructed curb to be vertical. Nevertheless, this constraint will not cause the algorithm to fail in such a situation, since the average position and the height of the curb can still be described by the separation of \mathcal{S} and \mathcal{A} .

V. SIMULTANEOUS CLASSIFICATION AND RECONSTRUCTION

In order to reconstruct the curb using the specified model from IV, we need to estimate the parameters \mathbf{a}, \mathbf{s} and \mathbf{c} from the measured height data \mathbf{h} provided by the DEM. This is realized in several steps:

- Classification $\mathbf{l} = [l_1, \dots, l_n]$ of all valid DEM cells to the labels $l_i \in \{\text{'street'}, \text{'street adjacent'}, \text{'unassigned'}\}$.
- Estimation of the parameters \mathbf{a} and \mathbf{s} of the street and street adjacent surface \mathcal{S} and \mathcal{A} .
- Estimation of the curb parameters \mathbf{c} .

The parameter estimation depends on the labeling results and vice versa. Therefore we use an iterative two step approach in the manner of the well known Expectation Maximization

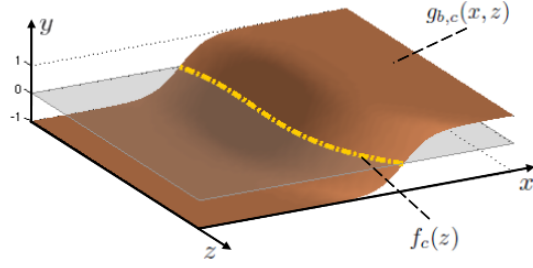


Fig. 6. Example of the sigmoidal function $g_{b,c}$. Its zero level function f_c is marked in yellow.

algorithm [16]. Based on an initial labeling $\mathbf{l}^{(0)}$ we perform the following two successive steps for every iteration $\nu \in \{1, \dots, \nu_{max}\}$.

First we estimate the parameters of the surfaces $\mathbf{a}^{(\nu)}, \mathbf{s}^{(\nu)}$ and the parameters of the third order polynomial $\mathbf{c}^{(\nu)}$, depending on the labeling of the last iteration $\mathbf{l}^{(\nu-1)}$. In the second step we search for a new labeling of the valid DEM cells with maximized probability given the estimated surface parameters and height observations.

The labels 'street' and 'street adjacent' represent the assignment of the cells to the surfaces \mathcal{S} and \mathcal{A} . The label 'unassigned' represents those cells which are not assignable to one of the surfaces caused by measurement errors, or since they contain vertical structures. In the following, we denote the set of indices of all cells assigned with label 'street' in iteration ν by $I_s^{(\nu)} \subseteq I$. Analog, the aliases $I_a^{(\nu)}$ and $I_u^{(\nu)}$ represent the cells assigned with the labels 'street adjacent' and 'unassigned' respectively.

The iterative process stops after a maximum number of iterations ν_{max} is reached, or if a termination criterion is fulfilled.

The remainder of this Section explains the single steps in more detail.

A. Parameter Estimation

1) *Estimation of the surface parameters:* We estimate the parameters $\mathbf{s}^{(\nu)}$ in a weighted least squares sense given the height values h_i of the cells assigned with label 'street':

$$\mathbf{s}^{(\nu)} = \underset{\mathbf{s}}{\operatorname{argmin}} \left(\sum_{i \in I_s^{(\nu)}} \frac{1}{\sigma_{h_i}^2} (h_i - \mathbf{s}^\top \mathbf{q}_i)^2 \right), \quad (6)$$

with $\mathbf{q}_i = [x_i^2, z_i^2, 2x_i z_i, x_i, z_i, 1]^\top$. Further the variance $\sigma_s^{2(\nu)}$ of the height data with respect to the estimated surface is given by the normalized minimum in (6). From the cells assigned with label 'street adjacent' we derive $\mathbf{a}^{(\nu)}$ and $\sigma_a^{2(\nu)}$ analog.

In general the curvature of the street and sidewalk surfaces is small. Thus, we extend (6) to a Bayesian estimation by adding a priori constraints on the second order surface parameters, i.e., we a priori assume the parameters $s_0^{(\nu)}, \dots, s_2^{(\nu)}$, and $a_0^{(\nu)}, \dots, a_2^{(\nu)}$ respectively, to be zero with a small variance.

2) *Estimation of the curb parameters:* As mentioned above the parameters $\mathbf{c}^{(\nu)}$ define the separation function f_c which divides the xz -plane projection of \mathcal{S} from the projection of \mathcal{A} . Given the classification result we search for the function f_c which is the most suitable separation of the DEM cells assigned with label 'street' and those assigned with label 'street adjacent'. Therefore, we define f_c to be the zero level of the sigmoidal function

$$g_{b,c}(x, z) = \frac{2}{1 + \exp(b(f_c(z) - x))} - 1 \quad (7)$$

that separates all points $[x, z]$ with $g_{b,c}(x, z) < 0$ from the points with $g_{b,c}(x, z) > 0$ as illustrated in Figure 6. The usage of the parameter b , which controls the steepness of the sigmoid, will be discussed later in Section V-B.3. From those cells that are assigned with the label 'street' or with the label 'street adjacent', we obtain the curb parameters by

$$\mathbf{c}^{(\nu)} = \underset{\mathbf{c}}{\operatorname{argmin}} \left(\sum_{i \in I_a^{(\nu)} \cup I_s^{(\nu)}} w_i^2 (\phi_i - g_{b,c}(x_i, z_i))^2 \right), \quad (8)$$

where ϕ_i is the selective function

$$\phi_i = \begin{cases} -1 & , \text{if } l_i^{(\nu)} = \text{'street'} \\ 1 & , \text{if } l_i^{(\nu)} = \text{'street adjacent'} \end{cases} \quad (9)$$

We choose the weights w_i to be the probabilities of the particular assigned labels $w_i = p(l_i^{(\nu)})$ as described in Section V-B.5.

Further, we assume the cubic characteristic of the curb to be low and thus constrain the third order parameter $c_0^{(\nu)}$ to be zero with a small variance, analog to the second order parameters of the curb adjacent surfaces.

B. Classification

1) *Initial labeling:* If no labeling from the previous time step is available, or if there has not been any curb detected in the last frame, we start the iterative process with an initial labeling. This is simply 'street' for all cells left of the ego vehicle's center and 'street adjacent' for all cells right of the vehicle's center, with respect to the lateral axis. Otherwise we use the final labeling of the last frame as initial input.

2) *Classification using a Conditional Random Field:* For the classification task we determine the labeling $\mathbf{l}^{(\nu)}$ which is most probable given the measured heights \mathbf{h} and the set of model parameters $\Theta^{(\nu)} = \{\mathbf{a}^{(\nu)}, \mathbf{s}^{(\nu)}, \mathbf{c}^{(\nu)}, \sigma_s^{2(\nu)}, \sigma_a^{2(\nu)}, \Sigma_{hh}\}$:

$$\mathbf{l}^{(\nu)} = \underset{\mathbf{l}}{\operatorname{argmax}} \left(p(\mathbf{l} | \mathbf{h}, \Theta^{(\nu)}) \right) \quad (10)$$

We align the graph structure of the random field to the grid structure of the DEM. This means that each cell represents a node and each pair of neighbors out of the 4-neighborhood $N_4 \subset I \times I$ represents an edge. With this, the probability

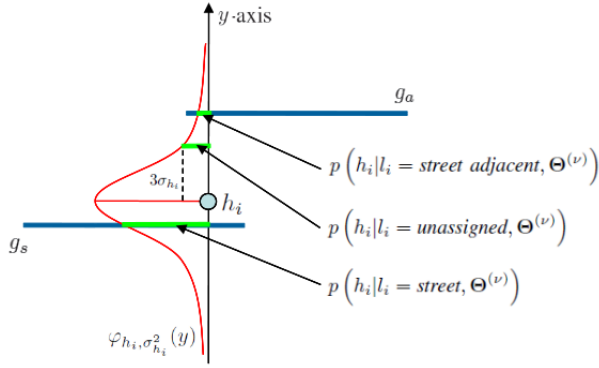


Fig. 7. Illustration of the height data likelihood for cell i sampled from the Gaussian function $\varphi_{h_i, \sigma_{h_i}^2}$. The likelihood of a cell assigned with 'street' or 'street adjacent' is sampled at the estimated height of the corresponding surface. The likelihood of a cell labeled to 'unassigned' is sampled at $3\sigma_{h_i}$ distance.

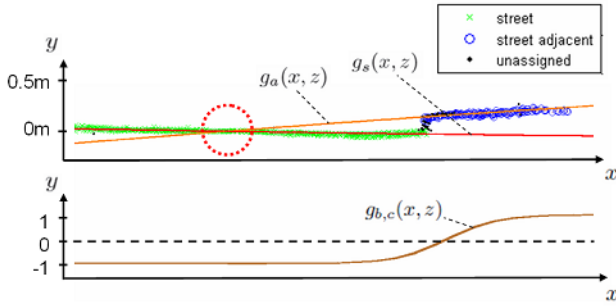


Fig. 8. Top: Front view of a labeled DEM in combination with the reconstructed surfaces. The height data likelihood is insufficient to properly classify the cells near the intersection of the estimated surfaces g_s and g_a encircled red. For this reason we introduce the dependency of the a priori term to the sigmoidal function $g_{b,c}$ illustrated below.

in (10) factorizes to a product of unary and binary terms

$$p(l|h, \Theta^{(\nu)}) \propto \prod_{i \in I} p(l_i | h_i, \Theta^{(\nu)}) \prod_{(i,j) \in N_4} p(l_i, l_j | h_i, h_j, \Theta^{(\nu)}) \quad (11)$$

The unary terms model the label probabilities with respect to the local height value measurement while the binary terms model the dependencies between neighboring labels. We do not model higher order dependencies to keep the computational effort of the inference manageable.

3) *Unary terms*: Using the Bayesian rule, the unary terms factorize into

$$p(l_i | h_i, \Theta^{(\nu)}) \propto p(h_i | l_i, \Theta^{(\nu)}) p(l_i | \Theta^{(\nu)}). \quad (12)$$

The first term describes how likely the measured height value h_i of a specified cell i is under the class assignment l_i and given the parameters $\Theta^{(\nu)}$. We assume the h_i to be normal distributed around the reconstructed height value of the assigned surface $g_s^{(\nu)}(x_i, z_i)$, or $g_a^{(\nu)}(x_i, z_i)$ respectively. Thus, we obtain this likelihood by sampling the Gaussian function $\varphi_{h_i, \sigma_{h_i}^2}(y)$ centered at h_i , with standard deviation taken from the cell's height accuracy σ_{h_i} . The sampling procedure is illustrated in Figure 7. Simply put, neglecting

ι	$p(h_i l_i = \iota, \Theta^{(\nu)})$	$p(l_i = \iota \Theta^{(\nu)})$
'street'	$\varphi_i(g_s^{(\nu)}(x_i, z_i))$	$\frac{1}{\xi} \sigma_s^{2(\nu)} \ 1 - g_{b,c}^{(\nu)}(z_i)\ $
'street adjacent'	$\varphi_i(g_a^{(\nu)}(x_i, z_i))$	$\frac{1}{\xi} \sigma_a^{2(\nu)} \ -1 - g_{b,c}^{(\nu)}(z_i)\ $
'unassigned'	$\varphi_i(h_i + 3\sigma_{h_i})$	$\frac{1}{\xi}$

TABLE I

UNARY TERMS DIVIDED IN HEIGHT DATA LIKELIHOOD $p(h_i | l_i, \Theta^{(\nu)})$ AND A PRIORI INFORMATION $p(l_i | \Theta^{(\nu)})$.

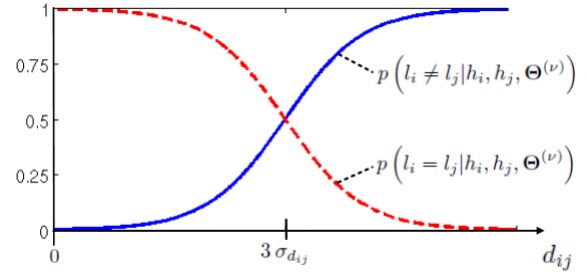


Fig. 9. Likelihood functions for assigning neighboring cells with equal or unequal labels dependent to the difference of the cells height values.

the second term, this yields the following properties to the a posteriori probability $p(l_i | h_i, \Theta^{(\nu)})$:

- A cell is the more likely assigned to a surface the smaller the distance between the cell's measured height value and the estimated height of the surface is.
- The probability of the label 'unassigned' is dominant if the distance to both surfaces is larger than $3\sigma_{h_i}$.

The second term of Equation (12) is the a priori information obtained by the previous iteration. Therefore, we prefer the surface with lower variance of the assigned cell heights represented by $\sigma_s^{2(\nu)}$ and $\sigma_a^{2(\nu)}$. In addition the a priori term represses multi-regional labeling. An exemplary situation for that to happen is illustrated in Figure 8, where the two surfaces g_a and g_s intersect. For this purpose we reduce the labeling to a region competition around the zero level of $g_{b,c}^{(\nu)}$ by weighting the probabilities with the distance of $g_{b,c}^{(\nu)}(x_i, z_i)$ to 1 for $l_i = \text{'street'}$ and respectively to -1 for $l_i = \text{'street adjacent'}$. The range of this region competition is implicitly controlled by the steepness parameter b defined in Equation (7). A small value of b allows large changes to the last labeling while a large value leads to a sharp separation of the labels along $f_c^{(\nu)}$. In our implementation we start the iterative process with a small value and increment b with each iteration.

In summary we achieve the probabilities charted in Table I where we use the short term $\varphi_i(y)$ for $\varphi_{h_i, \sigma_{h_i}^2}(y)$. The factor ξ ensures that the a priori probabilities sum to one.

4) *Binary terms*: The binary terms consider the height difference information of neighboring cells. We assume neighboring cells i and j to be more likely assigned with the same labels if the height difference $d_{ij} = \|h_i - h_j\|$ is small. Vice versa, we assume them to be labeled different if the height difference is large.

To distinguish real height differences from measurement

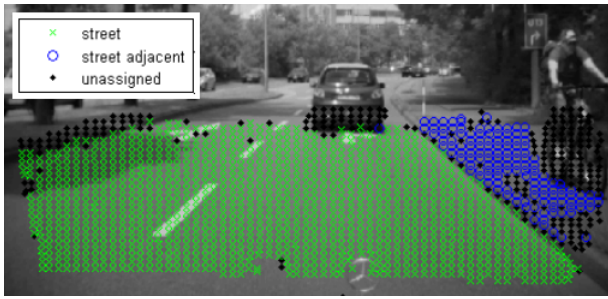


Fig. 10. Exemplary classification result of a DEM, providing 3D height information up to a distance of 20 m, projected into the image plane. Cells assigned with the label 'street' are plotted as green crosses and those assigned with 'street adjacent' appear as blue circles. The cells labeled to 'unassigned' are represented by black dots. Note, that some cells do not contain enough height values in the DEM and therefore are not used in the classification.

noise, the value of d_{ij} where both options are equiprobable is set to $3\sigma_{d_{ij}} = 3\sqrt{\sigma_{h_i}^2 + \sigma_{h_j}^2}$. The resulting likelihood functions regarding equal and unequal labeling are plotted in Figure 9.

5) *Inference*: Modeling the problem by means of a Conditional Random Field allows us to make use of a broad variety of procedures to estimate the best labeling. In our implementation we use the Sum-Product algorithm, also known as (Loopy) Belief Propagation. The algorithm is described for example in [17] (pp. 334-340). The decision is based on the fact, that this algorithm can be implemented efficiently exploiting the potential for parallel processing of the multi-core processor architecture of today's computers, which leads to the desired real time performance. As result we obtain an estimation of the most probable labeling $l_i^{(\nu)}$ for each cell along with the assigned probability $p(l_i^{(\nu)})$. Figure 10 illustrates the projection of a labeled DEM to the image plane.

C. Termination Criteria

The iterative process stops if a maximum number of iterations is reached or the following termination criteria is fulfilled. This criterion is designed to check for the absence of a curb. For this purpose we verify if one of both surfaces g_s and g_a sufficiently describes a certain percentage (say 99 percent) of all DEM cells either labeled to 'street' or to 'street adjacent'. In such a case all sufficiently described cells are reassigned with the label 'street' while the rest of the cells is labeled to 'unassigned'. We define a cell i to be sufficiently described by a surface if the difference between h_i and the surfaces height value $g_a^{(\nu)}(x_i, z_i)$ and $g_s^{(\nu)}(x_i, z_i)$ respectively is smaller than σ_{h_i} .

D. Extensions

In previous parts of this Section we demonstrated how to reconstruct the environment model specified in Section IV from an observed DEM. This approach is very general and may easily be extended regarding the environment model as well as additional input data. To consider further observation data, which may be extracted from the images or given by



Fig. 11. Reconstruction result using the extended model described in Section V-D for a scene containing curbs on both sides of the road.

additional sensors, we need to model its influence on the unary and binary terms defined in the Sections V-B.3 and V-B.4.

In the following, we exemplarily extend the environment model, by adding another curb on the left hand side.

We define this second curb C_{left} analog to equation (1) using the street surface \mathcal{S} and an additional street adjacent surface $\mathcal{A}_{\text{left}}$ on the left side of \mathcal{S} . The reconstruction procedure then is analog to the case of a single curb.

The parameter estimation step requires to supplementary estimate a_{left} and c_{left} , which are the parameters of the additional surface and separating function. For the classification step we introduce a class-label 'street adjacent left' and model the unary and binary factors as in the case of a single curb. Further we define the initial labeling by a tripartition along the x -axis, assigning the new label to the left third, the label 'street' to the middle third and 'street adjacent right' to the right third. Figure 11 illustrates the result of this extension for a scene with curbs on both sides of the street.

VI. RESULTS

The proposed method was implemented in C++ and integrated into our Mercedes Benz demonstrator vehicle. Tested on extensive runs in suburban environments, the system has proven to yield reliable results evaluated by live inspection of a human expert.

For all tests we use a DEM having 64×32 cells with respect to the column-disparity domain. It provides 3D height information up to a distance of 20 meters from the camera. Since the ground surface is occluded by the engine hood we choose a minimum distance of 6 meters.

We have benchmarked the computation time on recent PC hardware (4×3 GHz Intel Core2 Quad). It takes 2-3 ms to calculate the DEM and 6-7 ms for each iteration of the simultaneous classification and reconstruction process. The process usually converges after 3-4 iterations. Otherwise, we stop the iterative process after a maximum number of 7 iterations. Thus, the system fulfills our real-time requirements.

In order to evaluate the accuracy of the reconstruction with respect to the height of the curb, we have applied the algorithm to a set of eight different scenes, containing curbs of various heights. The curb heights are constant within each scene and vary from 4 cm to 16 cm. Each scene comprises 20 frames. To compare the obtained results to the manually measured reference height we calculate the joint Root Mean Squared Error (RMSE) concerning all frames of all scenes

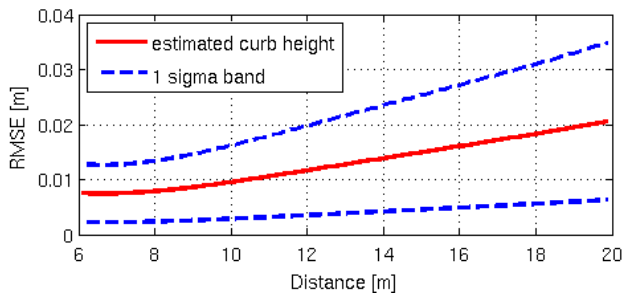


Fig. 12. Joint RMSE of the measured heights regarding all frames in the test set plotted against the distance to the camera. The 1-sigma interval is marked by the blue dashed lines.

of the test set. In Figure 12 the joint RMSE is plotted as function of the distance to the camera. The plot shows the RMSE to be smaller than 1 cm up to a distance of 10 meters and around 2 cm at 20 meter distance.

For comparison, we implemented a less complex approach based on the detection of local height discontinuities in the DEM, similar to [7]. The algorithm performs edge detection on a Cartesian DEM by means of a Canny edge detector. In a second step it extracts curb candidates from the set of edges using a Hough transformation. We applied this algorithm to the test set mentioned above and, for fairness, manually adapted the edge detector's parameters to the particular curb height in each scene. While this approach performs well at close-up range, it often fails to detect low curbs at distances of more than 10 meters.

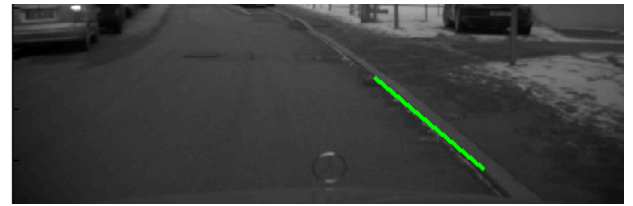
This problem is exemplary illustrated in Figure 13. The algorithm successfully detects the curb of 5 cm height up to a distance of approximately 11 meters 13(a). Lowering the edge detector's thresholds results in many false curb candidates without significant improvement regarding the real curb 13(b). Applying the approach to a DEM that is aligned to the column disparity domain yields similar results 13(c). This problem is caused by measurement noise which blurs the height gradients especially those of small curbs at great distances.

Since our method additionally considers the average measured height levels of curb adjacent surfaces we gain robustness with respect to measurement noise. Figure 13(d) illustrates the result of our method with regard to the challenging scene. Figure 14 demonstrates another benefit of this property. Although the curb is covered by snow and no sharp height discontinuity can be detected around the curbs position, our approach yields reliable reconstructions of the actual street boundary.

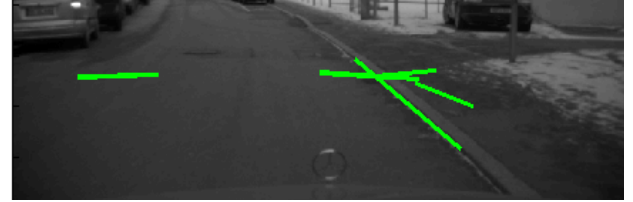
Further results of our method with respect to straight-line and curved curbs of various heights are shown in Figure 16.

VII. CONCLUSION AND FUTURE WORK

We proposed a novel approach for real-time reconstruction of curbs based on 3D point clouds. So far the method is not restricted to a special sensor. Although we use 3D data obtained from dense stereo vision in this contribution, other



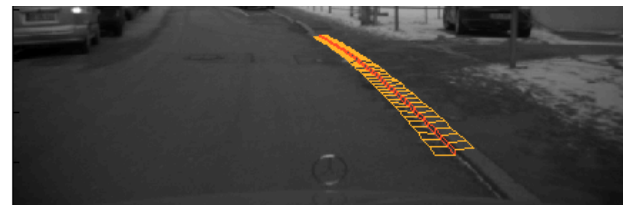
(a) Canny-Hough on Cartesian DEM



(b) Canny-Hough on Cartesian DEM (low edge detector threshold)



(c) Canny-Hough on column disparity DEM (low edge detector threshold)



(d) Our approach

Fig. 13. Example scene containing a low curb of 5 cm height. The curb candidates detected by the Canny-Hough approach are marked by green lines. In (a) and (b) a Cartesian DEM is used, while in (c) the DEM is aligned to the column disparity domain. The reconstruction result of our algorithm is illustrated in (d).

sensors that capture punctual data in real-time, e.g., laser scanners, would also be suitable.

Tests in real-world scenarios have shown the system to yield reliable results for curved and straight-line curbs up to a distance of 20 meters. We do not use any explicit threshold for the curbs minimum height and found the system to work fine even for low curbs of just 4 cm height.

In comparison to a less complex approach which extracts curb candidates from local height discontinuities, our method appears to be more robust with respect to measurement noise, especially at great distances. This is founded by the fact that we model the static environment of the curb as a whole and discover the global context of the curb's adjacent surfaces by means of an interpretation step.

However, the used environment model also limits the flexibility of the reconstruction. The assumption that the curb's characteristic can be modeled using a third order polynomial may not hold in all situations, as illustrated in Figure 15. A more flexible formulation, e.g., based on splines, is part of future work.



Fig. 14. Reconstruction results of scenes with snow covered curbs. Although the height discontinuity around the curbs position is not sharp, the actual street boundary is well estimated.



Fig. 15. Example for a curb which cannot be modeled using a polynomial of third degree with respect to the z -axis. Thus our algorithm fails to yield a precise reconstruction.

Further steps will exploit the generic concept of our approach to detect and reconstruct other static parts of the scene, e.g., parking cars or chuckholes. Another interesting task is to determine how additional information extracted from the images, e.g., image gradients or texture boundaries, support the reconstruction task.

REFERENCES

- [1] H. Badino, U. Franke, and D. Pfeiffer, "The stixel world - a compact medium level representation of the 3d-world," in *31st DAGM Symposium*, September 2009.
- [2] S. Se and M. Brady, "Vision-based detection of kerbs and steps," in *8th British Machine Vision Conference (BMVC)*, 1997, pp. 410–419.
- [3] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, pp. 11–15, 1972.
- [4] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.
- [5] R. Turchetto and R. Manduchi, "Visual curb localization for autonomous navigation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, Oct. 2003, pp. 1336–1342.
- [6] X. Lu and R. Manduchi, "Detection and localization of curbs and stairways using stereo vision," in *IEEE International Conference on Robotics and Automation (ICRA)*, April 2005, pp. 4648–4654.
- [7] F. Oniga, S. Nedeveschi, and M. Meinecke, "Curb detection based on elevation maps from dense stereo," in *IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, Sept. 2007, pp. 119–125.
- [8] —, "Curb detection based on a multi-frame persistence map for urban driving scenarios," in *11th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct. 2008, pp. 67–72.
- [9] 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.
- [10] C. Wojek and B. Schiele, "A dynamic conditional random field model for joint labeling of object and scene classes," in *10th European Conference on Computer Vision (ECCV)*, 2008, pp. 733–747.
- [11] P. Sturges, K. Alahari, L. Ladicky, and P. Torr, "Combining appearance and structure from motion features for road scene understanding," in *20th British Machine Vision Conference (BMVC)*, 2009.
- [12] 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*, 2009.
- [13] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [14] H. Badino, U. Franke, and R. Mester, "Free space computation using stochastic occupancy grids and dynamic programming," in *Workshop on Dynamical Vision (ICCV)*, October 2007.
- [15] F. Oniga, S. Nedeveschi, M. M. Meinecke, and T. B. To, "Road surface and obstacle detection based on elevation maps from dense stereo," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, Sept. 30 - Oct. 3 2007, pp. 859–865.
- [16] 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.
- [17] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.



Fig. 16. Reconstruction results of scenes comprising straight-line and curved curbs of various heights.