

High-Performance Long Range Obstacle Detection Using Stereo Vision

Peter Pinggera^{1,2}, Uwe Franke¹, Rudolf Mester^{2,3}

Abstract—Reliable detection of obstacles at long range is crucial for the timely response to hazards by fast-moving safety-critical platforms like autonomous cars. We present a novel method for the joint detection and localization of distant obstacles using a stereo vision system on a moving platform. The approach is applicable to both static and moving obstacles and pushes the limits of detection performance as well as localization accuracy.

The proposed detection algorithm is based on sound statistical tests using local geometric criteria which implicitly consider non-flat ground surfaces. To achieve maximum performance, it operates directly on image data instead of precomputed stereo disparity maps.

A careful experimental evaluation on several datasets shows excellent detection performance and localization accuracy up to very large distances, even for small obstacles.

We demonstrate a parallel implementation of the proposed system on a GPU that executes at real-time speeds.

I. INTRODUCTION

Obstacle detection represents a fundamental problem in the areas of mobile robotics and autonomous vehicles and has been in the focus of active research for decades. It forms the basis of many kinds of high-level tasks such as situation analysis, collision avoidance and path planning. Various sensor types can be applied to the problem, from passive ones like cameras to active ones such as radar or lidar sensors. While active range sensors offer supreme accuracy in terms of point-wise distance and velocity measurement, they usually suffer from low resolution and high cost. Cameras, on the other hand, provide very high spatial resolution at relatively low cost. Stereo or multi-camera setups allow for the computation of dense range maps of the observed environment, making them increasingly popular for application in mobile robots and autonomous vehicles. Unfortunately, an inherent drawback of stereo vision systems is the comparatively low accuracy of distance measurements, especially at long ranges. However, specifically this accuracy at maximum range is crucial for timely obstacle detection and response on fast-moving and safety-critical platforms, e.g. autonomous cars.

In this work, we present an approach that pushes the limits of both detection performance and longitudinal localization accuracy of distant generic obstacles using stereo vision. Our system performs a patch-wise binary classification into either *free-space* or *obstacle* image points and simultaneously yields a corresponding estimate of each obstacle position.

¹Environment Perception, Daimler R&D, Sindelfingen, Germany {peter.pinggera,uwe.franke}@daimler.com

²VSI Lab, Computer Science Dept., Goethe Univ., Frankfurt, Germany mester@vsi.cs.uni-frankfurt.de

³Computer Vision Laboratory, E.E. Dept., Linköping University, Sweden



Fig. 1. Example of obstacles detected by the proposed system at distances between 50 and 200 m (top, red), close-up of the resulting distance estimates (red: close, green: far) and corresponding bird's eye view (right)

Only reliable obstacle detections representing a certain confidence remain in the system output. Fig. 1 shows an example of a highway driving scene featuring relevant objects up to a distance of 200 m, all of which are successfully detected and located by our system.

The proposed method employs statistical hypothesis tests based on local plane models which implicitly allow for non-flat ground surfaces. Due to the generic geometric formulation it does not suffer from restrictions such as a limited set of previously learned obstacle classes. Furthermore, detection performance and localization accuracy are optimized by operating directly on the input image data instead of precomputed stereo disparity maps. Independent processing of individual image patches enables massive parallelization and we demonstrate real-time execution using a GPU implementation. Finally, the approach is not restricted to stereo but is readily extendable to calibrated multi-camera setups.

The remainder of this paper is structured as follows: We discuss related work in Sect. II, followed by a description of the proposed approach in Sect. III. A detailed evaluation of the system is presented in Sect. IV, including a quantitative analysis of detection performance, accuracy of longitudinal localization and runtime as well as various qualitative results from challenging autonomous driving scenarios.

II. RELATED WORK

The amount of literature on obstacle detection is vast and spans multiple application areas. We focus on camera-based methods for the detection and localization of generic obstacles in 3D space, in particular by using stereo setups on autonomous ground vehicles.

Many obstacle detection approaches are based on a so-called flat-world-assumption, modeling free-space or ground as a single planar surface and characterizing obstacles by

their height-over-ground [1], [2], [3]. Geometric deviations from the reference plane can be estimated either from a precomputed point cloud or directly from image data [4]. However, the resulting detection performance strongly depends on the accuracy of the ground plane parameters as well as the validity of such a simple model. Consequently, more sophisticated ground profile models have been introduced, from piece-wise planar longitudinal profiles [5] to clothoids [6] and splines [7].

The recent survey in [8] presents an overview of several stereo-based obstacle detection approaches that have proven to perform very well in practice. The methods are grouped into different obstacle representation categories and include Stixels [9], [10], Digital Elevation Maps (DEM) [11] and geometric point clusters [12], [13]. Notably, all of the methods rely on precomputed stereo disparity maps.

The Stixel algorithm distinguishes between a global ground surface model and a set of vertical obstacle segments of varying height. In this way a dense, compact and robust representation of the 3D scene is provided. The approach of [11] also produces a dense scene representation, but uses a DEM in combination with a quadratic ground model to detect and represent different types of obstacles. In [12] the geometric relation between pairs of 3D points is used to detect and cluster obstacle points. This point-wise representation is most similar to the one presented in this paper.

The above methods are designed for robust generic obstacle detection based on geometric criteria and work best in close- to medium range applications. Detection performance and localization accuracy drop quickly with increasing distance. In contrast, dedicated appearance-based detectors for specific object classes, e.g. vehicles, perform well even at large distances [14], [15]. Object stereo measurements of the detected objects can then be optimized by appropriate algorithms [16].

The recent work of [17] combines the generic Stixel representation with dedicated object detectors, thereby boosting the detection range for known object classes compared to the traditional Stixel algorithm. We select this method to serve as a baseline during our experimental evaluation.

Finally, impressive performance can be achieved using custom sensor configurations, such as the trinocular large-baseline tele-stereo setup shown in [18]. In contrast, the system presented here uses general-purpose stereo cameras only. A corresponding increase in performance can be expected by the straightforward extension to a trinocular setup.

III. APPROACH

We formulate the obstacle detection task as a statistical hypothesis testing problem. Independent tests are performed on small local patches distributed across the input images. Free-space is represented by the null hypothesis \mathcal{H}_f , while obstacles correspond to the alternative hypothesis \mathcal{H}_o . The hypotheses are characterized by constraints on the *orientations* of *local* 3D plane models. Each plane is defined by a parameter vector $\vec{\theta}$ holding the normal vector \vec{n} and the normal distance d from the origin: $\vec{\theta} = (n_X, n_Y, n_Z, d)^T$.

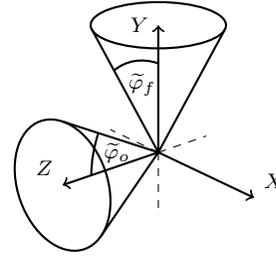


Fig. 2. The cones defined by $\tilde{\varphi}_f$ and $\tilde{\varphi}_o$ constrain the permitted plane normal orientations of the free-space and obstacle hypothesis models. The Z axis represents the optical axis of the left camera

In contrast to common global ground or obstacle models, the parameters of each individual local plane are allowed to vary within certain ranges around a hypothesis reference model. Accordingly, the parameter spaces of \mathcal{H}_f and \mathcal{H}_o are constrained by the angles $\tilde{\varphi}_f$ and $\tilde{\varphi}_o$, which define the maximum allowed deviation of the respective plane normals from their reference orientations. We set the reference values to be the Y and Z axes of the camera coordinate system, corresponding to a simple flat-ground / fronto-parallel obstacle reference model (see Fig. 2). The angles $\tilde{\varphi}_f$ and $\tilde{\varphi}_o$ can be set according to the expected shape of traversable surfaces and obstacles, respectively.

A. Generalized Likelihood Ratio Test

In order to avoid a loss in performance by intermediate processing steps, we define a test statistic based *directly* on a statistical model of the input image data. Unfortunately, the design of a likelihood ratio test yielding a provably optimal detector would require knowledge of the full PDFs of the data model under \mathcal{H}_f and \mathcal{H}_o [19].

Instead we make use of the Generalized Likelihood Ratio Test (GLRT), which offers no general optimality guarantees but has been shown to perform well in practice [20].

For each hypothesis $\mathcal{H}_i = \mathcal{H}_{\{f,o\}}$ the GLRT replaces the unknown parameters $\vec{\theta}_i$ by their Maximum Likelihood Estimates (MLEs) $\hat{\vec{\theta}}_i$, assuming the respective hypothesis to be true. We then decide for \mathcal{H}_o if

$$L(\vec{\mathcal{I}}) = \frac{p(\vec{\mathcal{I}}; \hat{\vec{\theta}}_o, \mathcal{H}_o)}{p(\vec{\mathcal{I}}; \hat{\vec{\theta}}_f, \mathcal{H}_f)} > \gamma, \quad (1)$$

or equivalently if

$$\ln \left(p(\vec{\mathcal{I}}; \hat{\vec{\theta}}_o, \mathcal{H}_o) \right) - \ln \left(p(\vec{\mathcal{I}}; \hat{\vec{\theta}}_f, \mathcal{H}_f) \right) > \ln(\gamma), \quad (2)$$

where the likelihood ratio $L(\vec{\mathcal{I}})$ represents the test statistic and $\vec{\mathcal{I}}$ is the grey-scale data vector of the stereo image pair.

Note that for patches classified as obstacle, the MLE $\hat{\vec{\theta}}_o$ implicitly provides an optimized estimate of the obstacle position in 3D space.

For nested parameter models it is possible to determine the decision threshold γ and the corresponding detection performance immediately, as $L(\vec{\mathcal{I}})$ can be shown to be asymptotically χ^2 distributed [20]. However, the parameter

spaces of our models are disjoint by design and we determine the optimal value of γ from an empirical analysis of the detection performance on relevant data (see Sect. IV-B).

B. Data Model

We formulate a statistical image formation model to define the likelihood terms in (1). The discrete left and right image patch values $I_l(\vec{x})$ and $I_r(\vec{x})$ are considered as noisy samples of the observed continuous image signal f at position \vec{x} . The terms $\alpha_l(\vec{x})$ and $\alpha_r(\vec{x})$ model a potential local intensity bias and $\eta(\vec{x})$ represents samples from a noise distribution with zero mean and an assumed variance σ^2 :

$$I_l(\vec{x}) = f(\vec{x}) + \alpha_l(\vec{x}) + \eta(\vec{x}) \quad (3)$$

$$I_r\left(W(\vec{x}, \vec{\theta})\right) = f(\vec{x}) + \alpha_r(\vec{x}) + \eta(\vec{x}). \quad (4)$$

The warp W transforms the image coordinates \vec{x} from the left to the right image, according to the plane model of the true hypothesis and the camera parameters $\mathbf{P} = \mathbf{K}[\mathbf{R} | \vec{t}]$. For the used models the warp represents a multiplication by the plane-induced homography

$$\mathbf{H} = \mathbf{K} \left(\mathbf{R} - \frac{1}{d} \vec{t} \vec{n}^T \right) \mathbf{K}^{-1}. \quad (5)$$

First, to compensate for a potential local bias, the mean intensity for each considered patch is removed. Treating all pixels in the patch area Ω as i.i.d. samples, we get from (4):

$$\begin{aligned} & \ln \left(p(\vec{\mathcal{L}}; \vec{\theta}_i, \mathcal{H}_i) \right) \\ &= \sum_{\vec{x} \in \Omega} C_1 - C_2 \cdot \rho \left(I_r \left(W(\vec{x}, \vec{\theta}_i) \right) - f(\vec{x}) \right), \end{aligned} \quad (6)$$

where C_1 and C_2 are constants and ρ represents the characteristic loss function of the assumed noise model, e.g. quadratic for a Gaussian or the L1 norm for a Laplacian distribution.

Finding the MLE of the parameter values then corresponds to the bound-constrained non-linear optimization problem

$$\hat{\vec{\theta}}_i \leftarrow \arg \min_{\vec{\theta}_i} \left(-\ln \left(p(\vec{\mathcal{L}}; \vec{\theta}_i, \mathcal{H}_i) \right) \right) \quad s.t. \quad |\varphi_i| \leq \tilde{\varphi}_i. \quad (7)$$

C. Parameter Optimization

We approach the bound-constrained non-linear parameter optimization problem of (7) using a modified trust region algorithm [21]. To enforce the constraints during the iterative optimization procedure, a basic trial step method is employed. Parameter steps resulting in infeasible configurations are rejected and the step size adapted appropriately for the subsequent iteration. In our experiments this simple method showed similar performance but faster execution than a variant of the more involved optimization approach of [22].

The target function derivatives necessary for optimization can be determined analytically, but results are omitted for lack of space. Similar formulations can be found in [23].

Since finding the global optimum of the parameter values cannot be guaranteed, suitable initialization is necessary. We initialize the free-space models from a coarse global ground

plane estimate and obstacles from fronto-parallel plane models at a certain distance. The required initial distance values can be extracted e.g. from a coarse disparity map.

D. Joint Signal Estimation

Due to the symmetric formulation of the data model, optimizing (7) involves the unknown signal f . We apply an approach as in [16] to optimize the parameter vector and estimate f at the same time. The unknown signal is simply computed as the mean of the correspondingly aligned input images after each update of $\hat{\vec{\theta}}_i$.

E. Verification

After the hypothesis test has been completed successfully, the pixel-wise residuals for the winning hypothesis are checked for consistency with the assumed image formation model. To this end, the sample mean and variance as well as the number of outlier pixels are analyzed.

Additionally, patches where the parameter constraints result in invalid solutions are discarded. This can occur due to our initialization method, e.g. for points high above the horizon that cannot be reached by the free-space model.

F. Patch Prefiltering

To ensure reliable results and well-conditioned systems as well as to avoid unnecessary computations, the initial set of patches is filtered prior to testing. Homogeneous image patches, i.e. patches with insufficient image texture are removed using a basic interest operator based on the horizontal image gradient.

G. Parallel Implementation

Since all tested image patches are processed independently, the described approach lends itself to massive parallelization. Thus, we implement our bound-constrained non-linear optimization algorithm on a GPU using the NVIDIA CUDA framework, enabling real-time execution.

IV. RESULTS AND ANALYSIS

A. Experimental Setup and Data

First, in Sect. IV-B we perform a quantitative evaluation of detection performance on the dataset of [17]. The dataset consists of 2,000 frames of manually labeled stereo images, taken from a test car in a highway scenario. Relevant obstacles are represented by other traffic participants. Non-occluded vehicles up to distances of more than 300 m are labeled with pixel-accuracy in every frame, every 10th frame also includes a pixel-wise free-space labeling. For evaluation, we directly analyze the patch-wise coverage of ground truth objects, i.e. the number of correct patch responses per object.

Furthermore, we use the dataset and evaluation framework of [16] to analyze the accuracy of estimated longitudinal obstacle positions in Sect. IV-C. The data consist of approximately 70,000 frames of stereo images, also taken in a highway setting. In the existing framework, individual vehicles at distances between 50 and 160 m are detected

and tracked based on appearance cues and stereo measurements. For each vehicle, we compute an associated distance measurement using a robust mean of the center positions of all corresponding obstacle patch detections. A long range radar sensor serves as ground truth for the resulting object distances.

Both datasets use a stereo setup with a baseline of 38 cm and a focal length of 1240 pixels, with spatial and radiometric resolutions of 1024×440 pixels and 12 bits. Distance estimates for initialization are computed using the approach of [24].

Finally, in Sect. IV-D we qualitatively evaluate the detection performance of our approach on a set of particularly challenging obstacles. Detecting even smallest objects turns out to be critical to fully autonomous driving functionality. Here we use a camera system with a resolution of 2048×1024 pixels, a focal length of 2270 pixels and a baseline of 21 cm.

Note that for many practical obstacle detection scenarios it might be unnecessary to optimize the full parameter vector $\vec{\theta}$. For our experiments, we reduce the number of free parameters by setting $n_X = 0$, which preserves sufficient flexibility for the considered scenarios. Only for extreme vehicle roll angles or off-road terrain it might be necessary to optimize the full parameter vector.

The exact values of the maximum angles defining the hypothesis parameter spaces turn out not to be too critical in practice. We set $\tilde{\varphi}_f = \tilde{\varphi}_o = 30^\circ$ (cf. Fig. 2).

To formulate the data model (Sect. III-B), we use a robustified Gaussian noise model by defining ρ as the Huber norm with an outlier threshold of 3σ . The variance σ^2 is estimated on test data from samples of correct hypotheses.

In our experiments we also consider different subsampling factors k . At $k = 1$, the overlapping patches are placed at every image pixel, while $k = 2$ uses only every second image line and column, etc. Naturally, this reduces the absolute number of possible patch detections but improves computational efficiency.

B. Detection Performance

1) *Patch-Based Performance:* In the first step, we determine a suitable operating point regarding the GLRT threshold γ as well as the used patch size. To this end, we analyze the Receiver Operating Characteristics (ROC) of three different patch sizes over multiple values of γ (see Fig. 3). We set $k = 2$ and consider all patches beyond a distance of 30 m.

As can be seen in Fig. 3, using too small patches causes a significant increase in the number of false positives. Too large patches, on the other hand, limit the achievable true positive rate irrespective of γ , since they tend to include an excess of background pixels around small objects and thus fail the verification step. Overall, the patch size should be chosen based on expected obstacle sizes and could further be scaled adaptively according to an estimate of the scene layout. As

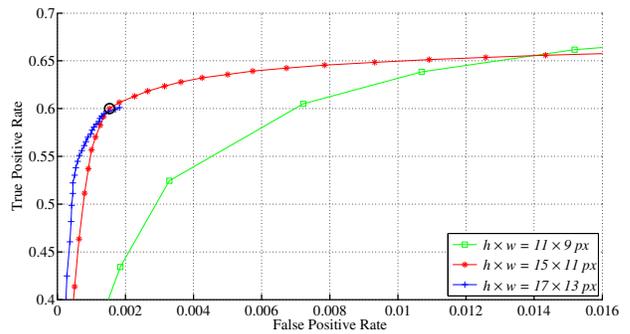


Fig. 3. ROC curve of patch detections for different patch sizes. The selected operating point is marked by a black circle

a suitable trade-off in the following experiments, we select a constant size of 15×11 pixels at a false positive rate of $1.5 \cdot 10^{-3}$ /patch. While the true positive rate of 0.6 at this point might appear relatively low, note that this represents only the detection rate for *single* patches beyond 30 m and also considers patches which were rejected during the verification step, representing neither \mathcal{H}_f nor \mathcal{H}_o .

Above the selected γ , most remaining false positives represent border artifacts around true obstacles, due to the necessary patch size, or objects like roadside tufts of grass.

Fig. 1 and Fig. 6 show sample images of the test dataset, evaluated at $k = 2$ with patch detections overlaid in red and distance estimates color coded from red (close) to dark green (far). The corresponding bird's eye views illustrate the obtained localization accuracy up to large distances.

2) *Object-Based Performance:* Using the selected operating point, detection performance on an object level can be analyzed. Fig. 4 displays the overall object detection rate depending on a given threshold, i.e. on the number of correct patch detections per ground truth object. As a state-of-the-art baseline we use the upper limits of the results from [17], including the traditional Stixel algorithm [10] and Stixels enhanced by appearance-based vehicle detector responses [17]. Notably, our generic approach outperforms even the combined Stixel and detector approach up to a threshold of 20 detections per object for low subsampling factors, and up to a threshold of 5 detections for higher k .

Additionally, we analyze the average number of patch detections depending on the absolute object distance (see Fig. 5). Since the ground truth does not include distance measurements, we use distances estimated directly from the corresponding patch detections. It can be seen that even for $k = 3$ and $k = 4$ we achieve an average of more than 10 patch detections per object up to a distance of 250 m. However, for such large distances the analysis actually benefits from several trucks occurring in the dataset, which are detected at much larger distances than cars.

C. Accuracy of Obstacle Position Estimates

To assess the accuracy of the estimated longitudinal obstacle positions, we follow [16] and consider the absolute

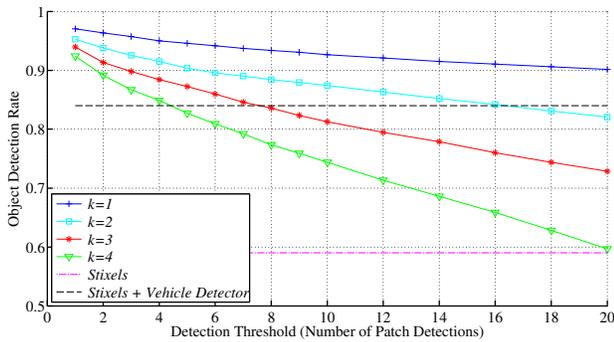


Fig. 4. Object detection rate using the number of patch detections per object as threshold. Results for different subsampling factors are compared to Stixels [10] and the extension using dedicated vehicle detectors [17]

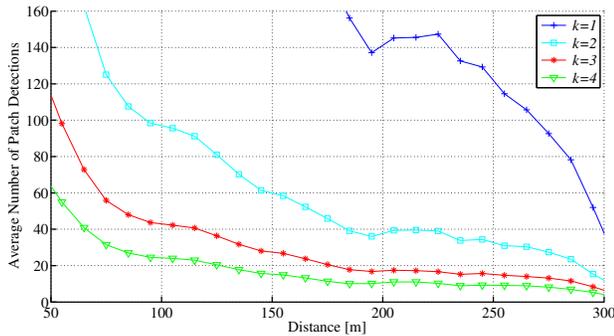


Fig. 5. Average number of patch detections per detected object, depending on its estimated absolute distance

stereo disparity error ϵ_d and its variation over time $\nabla\epsilon_d$. The estimator S_n [25] provides robust estimates of the corresponding standard deviations. By transforming the 3D patch center positions into respective disparity values and computing the interquartile mean over all detections per object, we achieve disparity error and temporal variation scales of 1/10 and 1/20 pixel, respectively (see Table I). Notably, this is on par with the limits presented in [16].

D. Detection of Small Generic Obstacles

For the last set of experiments on small generic obstacles, the patch size and subsampling rate are adapted to the modified camera setup, using a size of 21×17 pixels and $k = 4$. Fig. 7 shows example results for the considered obstacles, including a child-sized dummy, a bicycle lying on its side, and a dirt bucket. Our system is able to consistently detect and locate the bicycle and the bucket at a distance of 100 m, followed by the dummy at approximately 70 m.

TABLE I

SCALE ESTIMATES OF THE ACHIEVED OBJECT STEREO DISPARITY ERROR ϵ_d AND THE TEMPORAL DISPARITY ERROR VARIATION $\nabla\epsilon_d$

$S_n(\epsilon_d)$ [px]	$S_n(\nabla\epsilon_d)$ [px]
0.096	0.056

TABLE II

AVERAGE NUMBER OF OPTIMIZED MODELS AND CORRESPONDING RUNTIMES PER FRAME (PATCH SIZE 15×11)

Subsampling k	1	2	3	4
Number of Models	100,000	25,000	12,000	6,000
Runtime t_{avg} [ms]	400	100	60	40

E. Runtimes

The optimization of the hypothesis model parameters for each patch (see Sect. III-C) is the computationally most expensive step and is performed on a NVIDIA GeForce GTX TITAN GPU. Consequently, depending on the chosen subsampling factor k and patch size, rates of up to 20 frames per second are obtained. Approximate overall runtimes per full image of the used highway data are shown in Table II.

V. CONCLUSIONS

In this work, we present a novel method for the joint detection and localization of distant obstacles using a stereo vision system on a moving platform. The proposed algorithm is based on statistical hypothesis tests using local geometric criteria and can implicitly handle non-flat ground surfaces. It operates directly on image data, leading to excellent detection performance and localization accuracy up to large distances. Careful experimental evaluation shows consistent detections, even of smallest obstacles. Our generic approach even manages to outperform systems that leverage dedicated appearance-based detectors of known object classes. Real-time execution is obtained by massive parallelization on a GPU. Future work includes integration into a higher-level object reasoning framework.

REFERENCES

- [1] Z. Zhang, R. Weiss, and A. Hanson, "Obstacle Detection Based on Qualitative and Quantitative 3D Reconstruction," *TPAMI*, vol. 19, no. 1, pp. 15–26, 1997.
- [2] M. Lourakis and S. Orphanoudakis, "Visual Detection of Obstacles Assuming a Locally Planar Ground," in *ACCV*, 1998.
- [3] S. Nedeveschi, R. Schmidt, R. Danescu, D. Frentiu, T. Marita, T. Graf, F. Oniga, and C. Pocol, "High Accuracy Stereo Vision System for Far Distance Obstacle Detection," in *IV*, 2004, pp. 292–297.
- [4] H. S. Sawhney, "3D Geometry from Planar Parallax," in *CVPR*, 1994.
- [5] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through "V-disparity", in *IV*, 2002.
- [6] S. Nedeveschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, T. Graf, and R. Schmidt, "High Accuracy Stereovision Approach for Obstacle Detection on Non-Planar Roads," *INES*, 2004.
- [7] A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke, and D. Cremers, "B-Spline Modeling of Road Surfaces with an Application to Free Space Estimation," *TITS*, 2009.
- [8] N. Bernini, M. Bertozzi, L. Castangia, M. Patander, and M. Sabbatelli, "Real-Time Obstacle Detection using Stereo Vision for Autonomous Ground Vehicles: A Survey," in *ITSC*, 2014.
- [9] H. Badino, U. Franke, and D. Pfeiffer, "The Stixel World - A Compact Medium Level Representation of the 3D-World," in *DAGM*, 2009.
- [10] D. Pfeiffer and U. Franke, "Towards a Global Optimal Multi-Layer Stixel Representation of Dense 3D Data," in *BMVC*, 2011.
- [11] F. Oniga and S. Nedeveschi, "Processing Dense Stereo Data Using Elevation Maps: Road Surface, Traffic Isle and Obstacle Detection," *Veh. Technol.*, 2010.

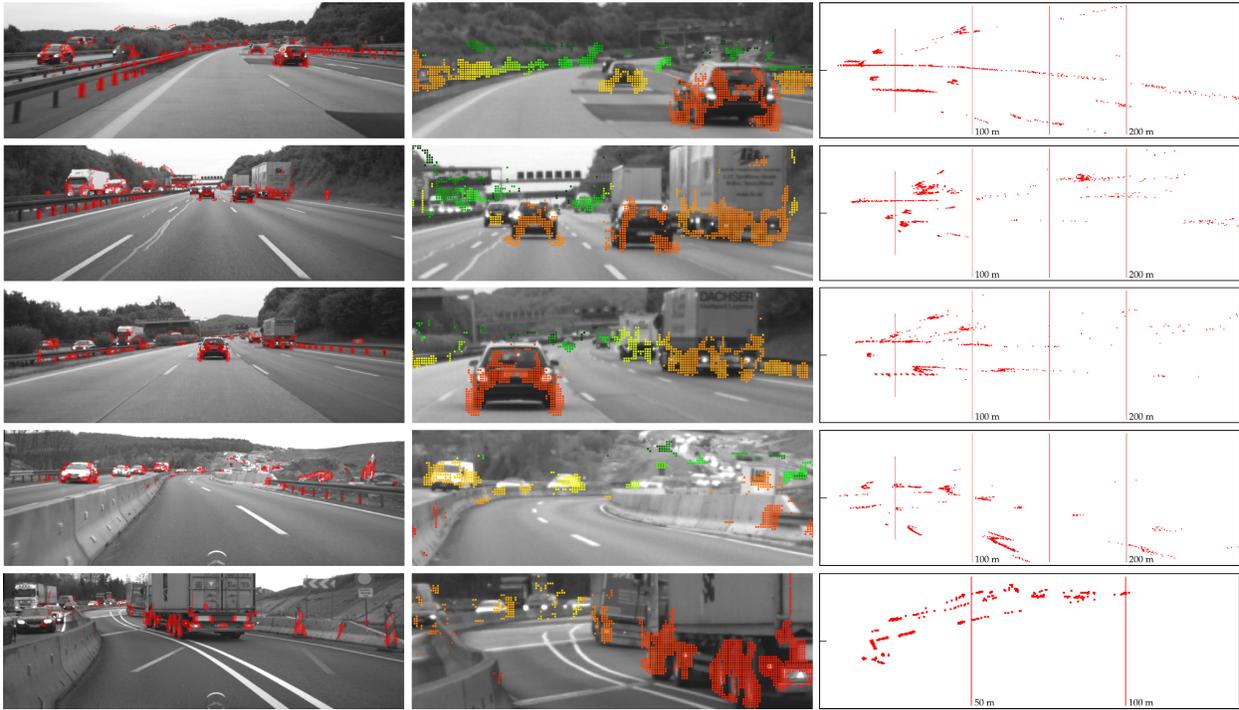


Fig. 6. Long range detection of vehicles and infrastructure: Patch detections (left, red), close-up of color-coded distance estimates (middle), and corresponding bird's eye view (right). Rows 1-4 demonstrate detected obstacles even beyond 200 m, rows 4 and 5 show scenes which strongly violate a common flat-world-assumption

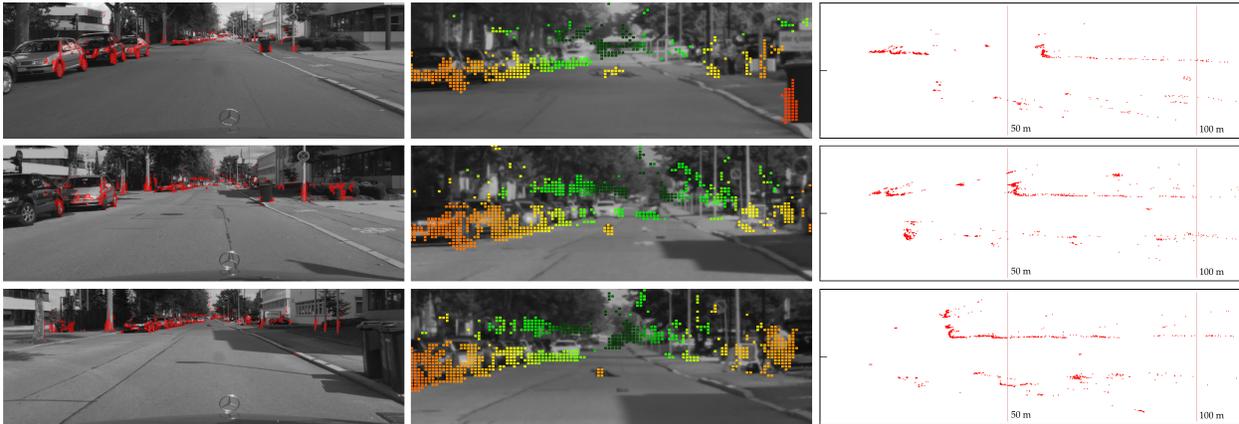


Fig. 7. Early detection of small but potentially critical objects. Top row: Bicycle lying on its side at 100 m. Middle row: Dirt bucket at 90 m. Bottom row: Child-sized dummy at 70 m. Also note the correct detections on vehicles in the far background

[12] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle Detection and Terrain Classification for Autonomous Off-Road Navigation," *Auton. Robots*, vol. 18, pp. 81–102, 2005.

[13] A. Broggi, M. Buzzoni, M. Felisa, and P. Zani, "Stereo Obstacle Detection in Challenging Environments: The VIAC Experience," in *IROS*, 2011, pp. 1599–1604.

[14] Z. Sun, G. Bebis, and R. Miller, "On-Road Vehicle Detection: A Review," *TPAMI*, vol. 28, no. 5, pp. 694–711, 2006.

[15] S. Sivaraman and M. M. Trivedi, "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis," *TITS*, vol. 14, no. 4, pp. 1773–1795, 2013.

[16] P. Pinggera, D. Pfeiffer, U. Franke, and R. Mester, "Know Your Limits: Accuracy of Long Range Stereoscopic Object Measurements in Practice," in *ECCV*, 2014, pp. 96–111.

[17] M. Cordts, L. Schneider, U. Franke, and S. Roth, "Object-level Priors for Stixel Generation," in *GCPR*, 2014.

[18] T. Williamson and C. Thorpe, "A Trinocular Stereo System for Highway Obstacle Detection," in *ICRA*, 1999.

[19] J. Neyman and E. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Math. Phys. Eng. Sci.*, vol. 231, 1933.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1998, vol. II.

[21] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.

[22] C. Kanzow, N. Yamashita, and M. Fukushima, "Levenberg-Marquardt Methods with Strong Local Convergence Properties for Solving Non-linear Equations with Convex Constraints," *J. Comput. Appl. Math.*, vol. 173, pp. 321–343, 2005.

[23] O. Kähler and J. Denzler, "Tracking and Reconstruction in a Combined Optimization Approach," *TPAMI*, vol. 34, no. 2, pp. 387–401, 2012.

[24] S. K. Gehrig, F. Eberli, and T. Meyer, "A Real-Time Low-Power Stereo Vision Engine Using Semi-Global Matching," in *ICVS*, 2009.

[25] P. J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," *J. Am. Stat. Assoc.*, vol. 88, no. 424, 1993.