https://doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-014

Perspectively Correct Bird's Views Using Stereo Vision

Christian Fuchs and Dietrich Paulus; Active Vision Group, University of Koblenz-Landau; Koblenz, Germany

Abstract

In modern vehicles bird's view systems are widely used to show the direct car surroundings to the driver. However, state-of-the-art methods for bird's view computations suffer from heavy distortions and unnatural warping. We propose an approach towards perspectively correct bird's view images for vehicular applications. Our method uses stereo images as input and is tested using stereo datasets.

Introduction

Nowadays cameras are of common use in vehicular applications. Many approaches concerning the environmental perception using optical sensors have been published and are being discussed. In summary, means of image processing and computer vision have become common use. In order to achieve a maximum coverage of the vehicle surroundings, a mostly exposed position for the camera installation is required. Yet, the geometric limitations caused by vehicle shapes and safety considerations constitute the need for optimization of the geometric setup. The problem raises when large combination vehicles, for example trucks with attached trailers, are taken into account. In order to generate appropriate views, the computation of virtual camera views, which consist of one or multiple transformed camera views, becomes necessary.

A common scenario is the computation of virtual bird's view images (also known as top view or 360° images) which visualize the direct car surroundings. Of course optical sensors cannot be installed so that a direct bird's view is captured. This matter is commonly addressed with a multiple camera setup with sensors distributed around the car. The images are transformed accordingly to the camera's properties and installation position and are finally combined into one virtual camera image. However, virtual bird's view images computed using state-of-the-art technologies suffer from several drawbacks which include imperfect transformations and unnatural distortion of *3-D* objects due to violated assumptions. An example of unnatural distortions for a single image is shown in Figure 1. The ground plane (for which an underlying assumption holds) is correctly transformed. All other objects, such as persons/legs, are wrongly transformed.

In this publication we propose an approach towards perspectively correct top view images. It is meant to improve the correctness of virtual camera views for the vehicle surroundings while raising safety especially in complex driving scenarios. We use axial stereo data to generate the bird's eye views, exploit the depth information from stereo correspondences and test our method on established datasets. In this publication, we do not focus on the issue of stitching a surround view image from several cameras. Instead, we focus on the correct transformation of single (stereo) images to a virtual bird's view.

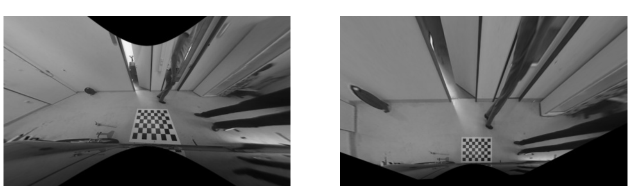


Figure 1: Left: Original image (undistorted); Right: corresponding homography top view

At first, related work and state-of-the-art methods are introduced. The drawbacks of homography based approaches are explained afterwards. The approach towards perspectively correct bird's views is introduced and our experimental results using the dataset are presented and finally concluded.

Related Work

Different research topics have to be taken into account in order to compute perspectively correct bird's views.

Perspective 2-D Warping

Perspective geometry and perspective transformations are a common approach for virtual camera views. Hartley and Zisserman [6] and Vincent and Laganiere [28] describe the principle behind it: Given two cameras C_1 and C_2. The image captured by camera C_1 shall be converted to the view described by camera C_2. Assuming all objects visible from C_1 are located on a plane in *3-D* space. This plane can be described using at least four image points $Q_{C_1} = \left\{q_i \in \mathbb{R}^2 \,\middle|\, i \in \mathbb{N}^+\right\}, |Q_{C_1}| \geq 4$ on the image plane. The image plane of C_1 (respectively C_2) is interpreted as a projective plane. Given corresponding image points $Q_{C_2} = \left\{q_i \in \mathbb{R}^2 \,\middle|\, i \in \mathbb{N}^+\right\}$ with $|Q_{C_2}| = |Q_{C_1}|$ in image coordinates of C_2, a transformation from camera C_1 to camera C_2 can be formulated using a homography matrix $H_{C_1 \to C_2} \in \mathbb{P}^{2\times 2}$. Assuming the cameras to have fixed lenses and a rigid affine transformation between their poses, matrix $H_{C_1 \to C_2}$ can be considered constant.

Camera Geometry

The correct geometric interpretation of image data requires precise camera models and calibration techniques. Tsai and Lenz [12] and Tsai [27] proposed calibration models especially for the pinhole camera model. Their work can be considered fundamental in this area. Using their methods, camera parameters can be estimated. Their approach has been extended by Zhang [30], which is widely used in computer vision.

Different camera models need adapted calibration methods. The omnidirectional camera model by Scaramuzza [21] is a suitable fit for both catadioptric and fisheye cameras. It is often applied in vehicular applications as wide angle lenses are commonly used there.

The camera parameters estimated using the methods mentioned are prerequisite for the connection of camera images to the

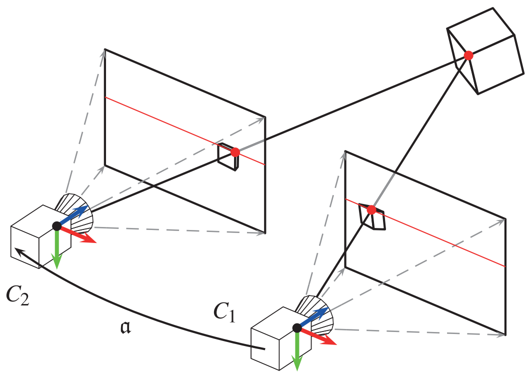


Figure 2: Stereo camera setup for disparity estimation. The point in the images and on the cube marks the correspondence between the images.

3-D world.

Virtual Camera and Bird's View Computation

Various publications address the issue of computing the view of a virtual camera using images taken by cameras at different locations. The utilization of homographies is a widely used approach for the warping process. Homographies include a flat world assumption, which seems reasonable on a first look for vehicles, as the street can be assumed to be a plane. However, artifacts for objects violating these assumptions are included.

An approach using cameras installed around the car is used by Liu, Kin and Chen [14]. They use homography matrices to transform the images and finally stitch the images to a top view image. To create a more natural look of the resulting image, they propose a virtual fisheye view.

The integration of a homography based approach with a hardware setup is proposed by Luo *et al.* [16]. Thomas *et al.* [26] focus on stitching top view images on cost-efficient computation systems. Sato *et al.* [20] utilize fish-eye cameras together with homographies on spatio-temporal data, whereas Li and Hai [13] focus on the calibration of a multi-view bird's eye view.

However, these approaches show heavy distortions for objects which violate the flat world assumption (of the homography). Yet, they are used in bird's view systems nowadays. Up to the best of the author's knowledge, no approach towards perspectively correct virtual bird's views was published, yet.

Stereo Vision

Multiple view geometry incorporates the use of multiple camera setups in order to compute depth information from multiple images. Hartley and Zisserman [6] have published fundamental work on the topic and summarize the principle behind this approach. In general, the idea is to make use of known geometric relations between calibrated cameras and to match the image frames taken at a time. Keypoints in corresponding image frames are matched by solving the point correspondence problem. The information gained is used together with the camera calibration data to estimate the *3-D* position of the point relative to the cameras. The principle of a stereo setup and the correspondence problem is depicted in Figure 2.

Two major types of algorithms for correspondence matching exist: On the one hand, keypoint-based approaches using algorithms like *Speeded Up Robust Features* (SURF) [1] or *Scale-Invariant Feature Transform* (SIFT) [15] exist. In general, image features of any kind can be used, assuming a proper assignment between the images can be achieved. Grimson [5] uses image features in order to find stereo correspondences. Horaud and Skordas [9] group features first in order to extract correspondences. Approaches using keypoints and/or image features usually show sparse *3-D* data with high accuracy.

On the other hand, block matching can be utilized for correspondence matching. Various algorithms and improvements have been developed and published so far. The results tend to show dense *3-D* data.

Hirschmüller *et al.* [8] use mutual information and pixel-wise matching in their *semiglobal matching (SGM)*. Their method shows promising precision properties. The algorithm has become popular and has already been adapted to particular scenarios, for example for in-vehicle applications [24, 7]. Pantilie and Nedevschi [17] propose an optimized version of the SGM. Einecke and Eggert [3] utilize a local correspondence approach and significantly reduce the execution while maintaining correspondence quality.

Many publications concerning stereo reconstruction and disparity estimation have already been presented. Therefore, this work does not focus on this issue. Of course, a lot of approaches towards stereo processing in vehicular environments (e. g. Broggi *et al.* [2], Pfeiffer and Franke [18] or Keller *et al.* [10]) have been published. Yet, no publication addresses perspectively correct bird's views so far.

In addition, several datasets for the evaluation of stereo reconstruction datasets have been published with an appropriate ground truth. The *KITTI Stereo Benchmark* as published by Geiger, Lenz and Uratsum [4] provides stereo datasets from road scenes for example. It is commonly used for benchmarking stereo correspondence algorithms. Pfeiffer, Gehrig and Schneider [19] published the *Ground Truth Stixel Dataset*, which contains annotated stereo sequence datasets of road scenes. The *Daimler Urban Segmentation Dataset* by Scharwächter *et al.* [22, 23] provides stereo video sequences which were recorded in urban traffic scenes. It contains stereo images with precomputed disparity maps as well as camera calibration information. We use the 2014 version in this publication for evaluation purposes.

Image-based Rendering

The computation of virtual camera views is a topic discussed in the field of image-based rendering. Shum and Kang [25] give a survey of different approaches towards the interpolation of views. They categorize the existing techniques using the grade of geometric modeling underneath. As in vehicular applications no pre-computed or pre-modeled geometric knowledge is possible due to the dynamic scenery and the target is to get to real-time processing time, only implicit geometry methods are reasonable. Laveau and Faugeras [11] propose view prediction based on the fundamental matrix using two captured images. Zinger, Do and De With [31] discuss a free-viewpoint depth based rendering for *3-D*-TV applications. Their method relies on disparity maps. Vogt *et al.* [29] use image-based rendering with light-fields to improve image quality in image sequences.

However, these publications discuss the rendering of virtual

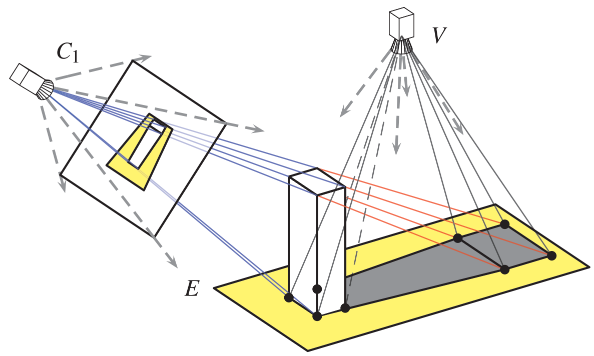


Figure 3: The Homography Shadowing Effect

camera views from camera poses close to the original camera's views, e. g. light positions shift and/or light rotation. In case of a virtual bird's eye computation in vehicular applications, extensive shifts and rotations become necessary. Yet, the publications contain fundamental work on the issue.

Homography Shadowing

As introduced, homography transformations are the state-of-the-art approaches for virtual top view computations in vehicles. As already implied, their use has several drawbacks due to the underlying assumption for the visible objects to be located on a single plane. This assumption is mostly fulfilled for roads and other flat ground surfaces. Yet, it is mostly violated by any obstacle the vehicle might come across. The violation of the plane assumption leads to an unnatural warping of all respective objects, thus producing unnatural looking images (Figure 1). We denote the underlying effect as the **Homography Shadowing Effect**: The shape of the unnaturally looking object matches the shape of the resulting shadow that would be caused by the light source at the virtual cameras position. Figure 3 illustrates the effect. The image taken by camera C_1 is transformed to the view of the virtual camera V using a homography with the underlying plane assumption for plane E. The viewing rays of C_1 (blue) of course end at the object. As no depth information is available, the object will be mapped onto plane E (flat world assumption) when applying the transformation. The view of virtual camera V is computed using a homography transformation. Therefore, V will see the object in shape of its *homography shadow* (gray).

In case of vehicular use, the shadowing effect may lead to a deceiving feeling of safety: A moving passenger or vehicle might be hidden behind and object. Nevertheless will the top view itself show a filled image while omitting the potential dangerous situation. It is therefore desirable to eliminate the effects of the obviously violated homography top view assumptions.

Virtual Top View Computation

A prerequisite for the computation of a virtual top view is the definition of the virtual camera itself. The two main projections to consider are perspective and orthographic projections. In general, a perspective projection is derived from the pinhole camera model.

In case of a virtual top view an orthographic projection is of advantage: The goal is to create a flat view inspired by a bird's view from high above (assumption: infinite distance). An orthographic projection does not show perspective properties (or at least at infinite distance) and is therefore to prefer.

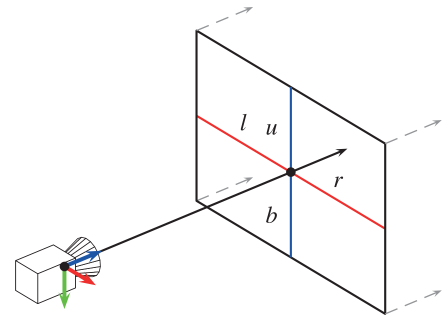


Figure 4: Parameters for an orthographic camera

The orthographic projection poses a rather simple projection: Objects are transformed onto the image plane using a direct orthogonal projection. The principle of an orthographic camera is illustrated in Figure 4. The camera is generally described using four parameters: $u, b, l, r \in \mathbb{R}$.

These parameters define the view frustum of the virtual camera. It is reasonable to assume the frustum to be symmetric ($u = b$ and $l = r$). Assuming the target pixel grid to be orthonormal and equidistant in both directions and the target camera resolution to be $(w_V, h_V)^{\mathrm{T}} \in \mathbb{N}^2$, the camera configuration space can be narrowed down to the width and height of the image in pixels and a scaling factor $s \in \mathbb{R}$ for the imaging measurements. The configuration space of an orthographic virtual camera is defined as:

$$\mathrm{conf}_o(V) = (w_V, h_V, s)^{\mathrm{T}} \tag{1}$$

The corresponding matrix for orthographic projection $O_V \in \mathbb{P}^{2\times 2}$ can be defined as:

$$O_V = O(\mathrm{conf}_o(V)) = \begin{pmatrix} 2(sw_V)^{-1} & 0 & 0 \\ 0 & 2(sh_V)^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

The position and orientation of an object in a *3-D* space defined by a Cartesian coordinate system can be described using a translation and a rotation relative to the orthonormal bases of the coordinate system. The components can be combined to a so called pose $\mathfrak{a} \in \mathbb{R} \times \mathbb{H}, \mathfrak{a} = (t, \phi)$, with $t \in \mathbb{R}^3, t = (t_x, t_y, t_x)^{\mathrm{T}}$ the translation and $\phi \in \mathbb{H}, \phi = \phi_w + \mathrm{i} \cdot \phi_x + \mathrm{j} \cdot \phi_y + \mathrm{k} \cdot \phi_z$ the rotation as a unit quaternion. The pose definition is used to describe the affine transformations between two cameras C_1 and C_2. Given a vector $b \in \mathbb{R}^n$, the vector transformed to homogenous coordinates is represented by $\tilde{b} \in \mathbb{P}^n$. For readability reasons, we use this notation in the following paragraphs and perform some implicit conversions between vectors and their homogenous representation.

Now let C_1 be a calibrated camera with disparities available. These disparities are used for the re-projection to *3-D* space. The goal is to create a perspectively correct view of virtual camera V:

Given a set of points $P_{C_1} = \left\{ p_{C_1} \middle| p_{C_1} \in \mathbb{R}^3 \right\}$ in the coordinate system of C_1. The transition to the coordinate system of a virtual camera V, whereby $\mathfrak{a}$ is the transition from C_1 to V, is defined as a function $\Upsilon_{\mathfrak{a}} : \left\{\mathbb{R}^3\right\} \rightarrow \left\{\mathbb{R}^3\right\}$:

$$P_V = \Upsilon_{\mathfrak{a}}(P_{C_1}) \tag{3}$$

$$\Upsilon_{\mathfrak{a}}(P) = \{\upsilon_{\mathfrak{a}}(p) \,|\, \forall p \in P\} \tag{4}$$

with $\upsilon_{\mathfrak{a}}(p) : \mathbb{R}^3 \to \mathbb{R}^3$ the transformation of a single vector from C_1 to V. Let $\theta : \mathbb{R}^3 \times \mathbb{H} \to \mathbb{P}^{3\times 3}$ compute the transformation matrix according to pose $\mathfrak{a}$:

$$\upsilon_{\mathfrak{a}}(p) = \theta(\mathfrak{a}) \cdot \tilde{p} \tag{5}$$

$$\theta(\mathfrak{a}) = \begin{pmatrix} 1-2\phi_y^2-2\phi_z^2 & 2\phi_x\phi_y-2\phi_z\phi_w & 2\phi_x\phi_z+2\phi_y\phi_w & t_x \\ 2\phi_x\phi_y+2\phi_z\phi_w & 1-2\phi_x^2-2\phi_z^2 & 2\phi_y\phi_z-2\phi_x\phi_w & t_y \\ 2\phi_x\phi_z-2\phi_y\phi_w & 2\phi_y\phi_z+2\phi_x\phi_w & 1-2\phi_x^2-2\phi_y^2 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6}$$

The projection onto the image plane of camera V is described by function $\Gamma_V : \{\mathbb{R}^3\} \to \{\mathbb{R}^2\}$:

$$Q_V = \Gamma_V(P_V) \tag{7}$$

$$\Gamma_V(P_V) = \{\gamma_V(p) \,|\, \forall p \in P_V\} \tag{8}$$

with $\gamma_V : \mathbb{R}^3 \to \mathbb{R}^2$ the projection function for camera V for a single point p. Let λ_V describe the projection matrix for C:

$$\gamma_V(p) = \lambda_V \cdot p \tag{9}$$

Matrix λ_V for the view of V is dependent on the desired projection:

$$\lambda_V = Z_V \cdot O_V \tag{10}$$

with $Z_V \in \mathbb{P}^{2\times 2}$ the transformation matrix to pixel coordinates with respect to the image resolution of V:

$$Z_V = \begin{pmatrix} 0.5w_V & 0 & 0.5w_V \\ 0 & 0.5h_V & 0.5h_V \\ 0 & 0 & 1 \end{pmatrix} \tag{11}$$

The resulting transform from the *3-D* points P_{C_1} to the view of V is:

$$Q_V = \Gamma_V \circ \Upsilon_{\mathfrak{a}}(P_{C_1}) \tag{12}$$

The algorithm defined above addresses the transformation from one camera C_1 into the view of virtual camera V. As one camera might not be able to cover the area which should be visible in V's view, our approach can be extended to work with multiple cameras which are joined to one virtual camera.

Given a set of cameras $\{C_1, ..., C_n\}$ with $n \in \mathbb{N}^+$ and the corresponding set of poses $\{\mathfrak{a}_1, ..., \mathfrak{a}_n\}$, the view of virtual camera V is defined as:

$$Q_V = \Gamma_V\left(\bigcup_{i=1}^{n} \Upsilon_{\mathfrak{a}_i}(P_{C_i})\right) \tag{13}$$

Using the transformations described, the view of a virtual camera can be computed with one or more cameras, when depth data is available.

A large rotational change caused by pose $\mathfrak{a}$ will result in holes in the resulting image. A post-processing routine for hole filling (e. g. single pixel holes) is needed in order to compute dense results. In a first approach, we only fill pixels which are fully enclosed by neighbors using a mean filter. Figure 5 illustrates the overall steps to take:

After an initial acquisition the raw images from the stereo pair(s) have to be undistorted and rectified. Afterwards, a stereo correspondence algorithm computes dense disparity data. The image data is then re-projected to *3-D* using depth estimates. The camera transformation is applied as stated above and the virtual camera is used to render the perspectively correct bird's view. Of course, the computed image is subject to post-processing steps thus improving the image quality. The post-processing of bird's view images (such as stitching, hole filling, etc.) is not the subject of this publication and therefore not discussed. The final image can then be used as the input for e. g. advanced driver assistance systems or visualizations.

Experimental Results

The evaluation of the system proposed is difficult when no ground truth data is available. For the results presented in this paper, we focus on commonly used datasets. In particular, we use the *Daimler Urban Segmentation Dataset* in its 2014 version [23]. The results presented rely on the calibration, disparity and image data provided by the datasets. Of course the datasets only contain data recorded in driving direction of the vehicle. Yet, no dataset providing stereo data recorded sideways or in backwards direction is publically available up to the best of the authors knowledge. However, we decided to rely on published and established datasets for comparability and reproducibility reasons. As no ground truth to compare against is available, a quantitative evaluation is not possible yet, but planned for future work.

Sample frames of the bird's view computation are shown in Figure 7. The presentation of the results is intended to show single computed images together with colored variants which highlight some geometry details. Sample 1 (Figure 7) shows the perspective correctness of the transformation, especially in a range of up to 15 meters from the camera's position. The arrows on the ground are projected correctly to a bird's view. Samples 2 and 3 illustrate the effect of the stereo data integration. The passengers/cyclists on the image are correctly projected and areas which are not visible are not filled with wrong information. Of course, hidden geometry results in no displayed data which we consider to be true compared to distorted areas in images transformed using homographies. For further development, these areas could be integrated with image data from frames without occlusion.

An example with a classical homography top view and a top view computed by the method proposed is shown in Figure 6.

Conclusion

We presented an approach towards perspectively correct bird's view images using stereo data as input. The experimental results show promising results and will be further subject of our research. We have shown a proof of concept concerning the perspectively correct bird's view using the *Daimler Urban Segmentation Dataset* [23]. The dataset was recorded for segmentation purposes. Unfortunately, it does not cover the direct vehicle surroundings due to the cameras' installation positions. It is planned to extend the method using multiple stereo camera pairs around the vehicle. But as already stated, no public dataset is available yet. Our plan is to

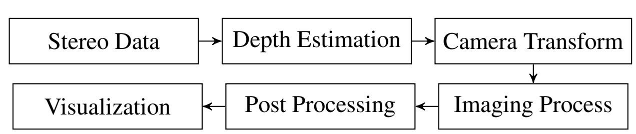


Figure 5: Scheme for the computation of perspectively correct bird's view images

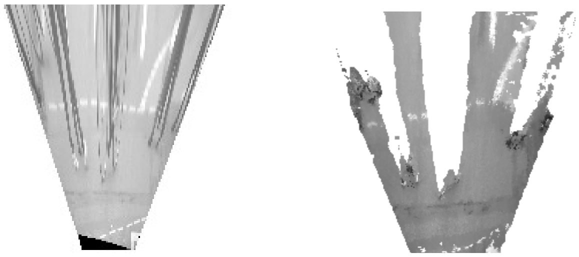


Figure 6: Classical homography top view (left) and perspectively correct view (raw output without post-processing; right). Frame 656 from dataset “test_2” [23]. The bicycle and the passengers in the image are warped rapidly in the left image due to the Homography Shadowing Effect.

record appropriate test data and to publish it.

The range of the bird’s view is far beyond the range of traditional bird’s view systems. The faraway areas (approx. 15 m and above) could be used for obstacle detection (e. g. Sample 1, Figure 7) so that further applications can be derived from the system. The registration of the multiple stereo camera setup is another challenging task, which will be of further investigation.

References

[1] H. Bay, T. Tuytelaars, and L. V. Gool. SURF : Speeded Up Robust Features. In *European Conference on Computer Vision*, pages 404–417, 2006.

[2] A. Broggi, C. Caraffi, R. I. Fedriga, P. Grisleri, and I. Parma. Obstacle Detection with Stereo Vision for Off-Road Vehicle Navigation. In *IEEE Conference on Computer Vision and Pattern Recognition*, page 65, San Diego, 2005.

[3] N. Einecke and J. Eggert. A Multi-Block-Matching Approach for Stereo. In *IEEE Intelligent Vehicles Symposium, Proceedings*, pages 585–592, 2015.

[4] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[5] W. E. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE transactions on pattern analysis and machine intelligence*, 7(1):17–34, 1985.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[7] S. Hermann and R. Klette. Iterative semi-global matching for robust driver assistance systems. In *Asian Conference on Computer Vision*, pages 465–478, 2013.

[8] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, 2008.

[9] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1168–1180, 1989.

[10] C. G. Keller, C. Hermes, and D. M. Gavrila. Pattern Recognition: 33rd DAGM Symposium, Frankfurt/Main, Germany, August 31 – September 2, 2011. Proceedings. In R. Mester and M. Felsberg, editors, *DAGM*, chapter Will the Pedestrian Cross? Probabilistic Path Prediction Based on Learned Motion Features, pages 386–395. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[11] S. Laveau and O. Faugeras. 3-D scene representation as a collection of images. *Proceedings of 12th International Conference on Pattern Recognition*, 1, 1994.

[12] R. K. Lenz and R. Y. Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):713–720, 1988.

[13] S. Li and Y. Hai. Easy calibration of a blind-spot-free fisheye camera system using a scene of a parking space. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):232–242, 2011.

[14] Y. C. Liu, K. Y. Lin, and Y. S. Chen. Bird’s-eye view vision system for vehicle surrounding monitoring. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4931 LNCS:207–218, 2008.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[16] L. Luo, I. Koh, S. Park, R. Ahn, and J. Chong. A software-hardware cooperative implementation of bird’s-eye view system for camera-on-vehicle. *2009 IEEE International Conference on Network Infrastructure and Digital Content*, pages 963–967, 2009.

[17] C. D. Pantilie and S. Nedevschi. SORT-SGM: Subpixel Optimized Real-Time Semiglobal Matching for Intelligent Vehicles. *IEEE Transactions on Vehicular Technology*, 61(3):1032–1042, 2012.

[18] D. Pfeiffer and U. Franke. Towards a Global Optimal Multi-Layer Stixel Representation of Dense 3D Data. *Procedings of the British Machine Vision Conference 2011*, pages 51.1–51.12, 2011.

[19] D. Pfeiffer, S. Gehrig, and N. Schneider. Exploiting the power of stereo confidences. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 297–304, 2013.

[20] T. Sato, A. Moro, A. Sugahara, T. Tasaki, A. Yamashita, and H. Asama. Spatio-temporal bird’s-eye view images using multiple fish-eye cameras. *Proceedings of the 2013 IEEE/SICE International Symposium on System Integration*, pages 753–758, 2013.

[21] D. Scaramuzza. *Ominidirectional vision: from calibration to robot estimation*. PhD thesis, Citeseer, 2008.

[22] T. Scharwächter, M. Enzweiler, U. Franke, and S. Roth. Efficient Multi-Cue Scene Segmentation. *German Conference on Pattern Recognition (GCPR)*, 2013.

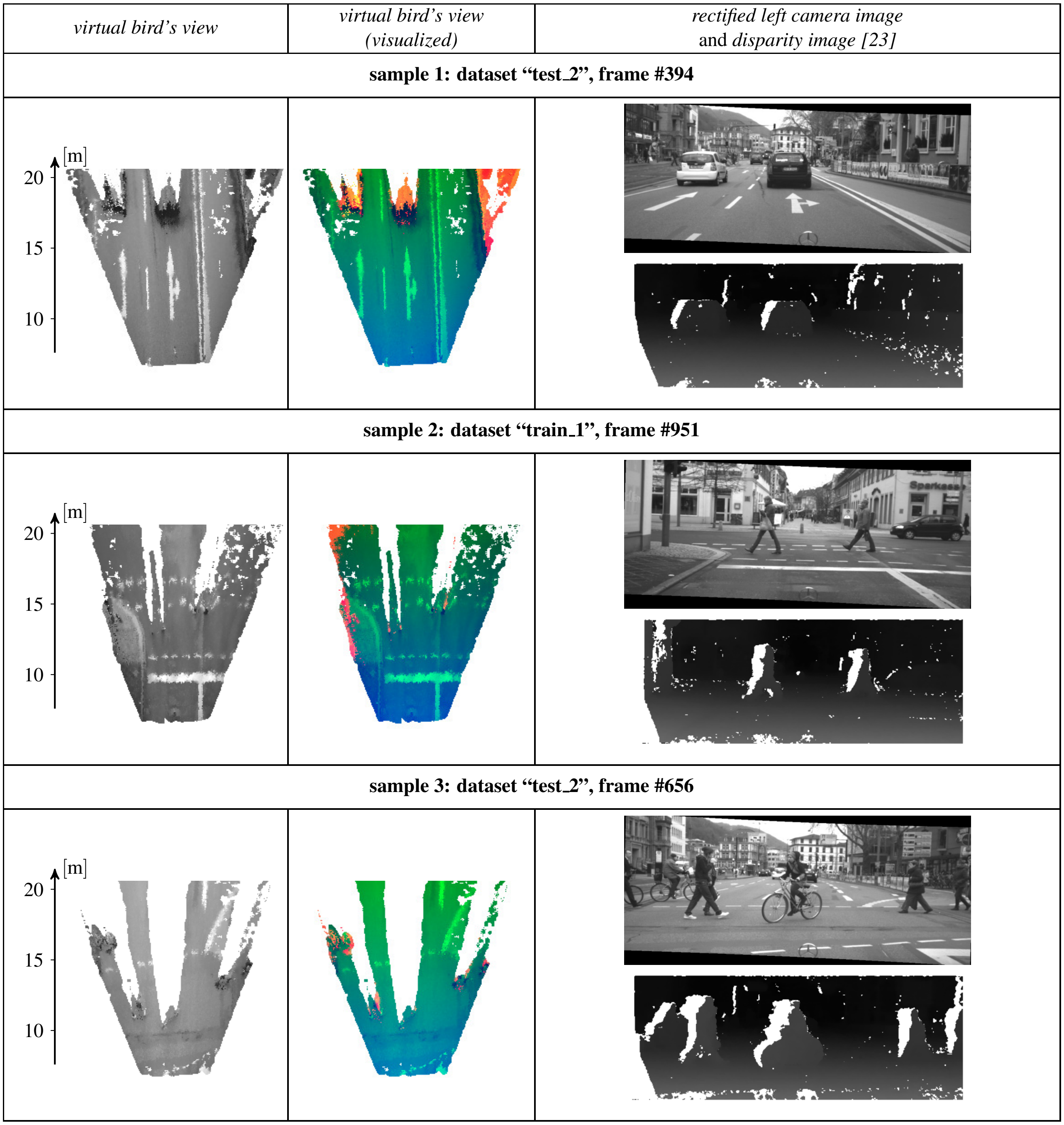
[23] T. Scharwächter, M. Enzweiler, S. Roth, and U. Franke. Stixmantics: a medium-level model for real-time semantic scene understanding. In *Computer Vision ECCV 2014*, pages 533–548. Springer International Publishing, 2014.

[24] T. Scharwachter, M. Schuler, and U. Franke. Visual guard rail detection for advanced highway assistance systems. *IEEE Intelligent Vehicles Symposium, Proceedings*, (Iv):900–905, 2014.

[25] H.-Y. Shum and S. B. Kang. A review of image-based rendering techniques. *Proc. SPIE Visual Communications and Image Processing*, pages 2–13, 2000.

[26] B. Thomas, R. Chithambaran, Y. Picard, and C. Cougnard. Development of a cost effective bird’s eye view parking assistance system. *2011 IEEE Recent Advances in Intelligent Computational Systems*, pages 461–466, 2011.

[27] R. Y. Tsai. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344 (22 pages), 1987.



Color Legend: *Red*: Indicates objects/obstacles above ground (via thresholding)
Green: Gray level intensity from original image
Blue: Proximity to the vehicle

The images show preliminary results using the method proposed. Our current work focuses on image quality improvement methods such as hole filling.

Figure 7: Experimental bird's view results together with the input images from [23]. The distance to the camera is meters given on the side axis.

[28] E. Vincent and R. Laganière. Detecting planar homographies in an image pair. *2nd International Symposium on Image and Signal Processing and Analysis*, 0(2):182–187, 2001.

[29] F. Vogt, S. Krüger, J. Schmidt, D. Paulus, H. Niemann, W. Hohenberger, and C. H. Schick. Light fields for minimal invasive surgery using an endoscope positioning robot. *Methods of information in medicine*, 43(4):403–408, 2004.

[30] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[31] S. Zinger, L. Do, and P. H. N. De With. Free-viewpoint depth image based rendering. *Journal of Visual Communication and Image Representation*, 21(5-6):533–541, 2010.

Author Biography

Christian Fuchs received a Diploma degree in Computer Science from the University of Koblenz-Landau in 2011. He works as a research associate in the Active Vision Group. His primary research interests are 3D pose estimation, stereo vision and driver assistance systems.

Dietrich Paulus obtained a Bachelor degree in Computer Science from University of Western Ontario, London, Canada, followed by a diploma (Dipl.-Inf.) in Computer Science and a PhD (Dr.-Ing.) from Friedrich-Alexander University Erlangen-Nuremberg, Germany. He obtained his habilitation in Erlangen in 2001. Since 2001 he is at the Institute for Computational Visualistics at the University Koblenz Landau, Germany where he became a full professor in 2002. His primary interests are computer vision and robot vision.