REDUCTION OF THE FRONTO-PARALLEL BIAS FOR WIDE-BASELINE SEMI-GLOBAL MATCHING

Lukas Roth, Helmut Mayer

Institute for Applied Computer Science, Bundeswehr University Munich, Neubiberg, Germany
{lukas.roth, helmut.mayer}@unibw.de

Commission II, WG II/4

KEY WORDS: Wide-Baseline Image Matching, Dense Image Matching, Semi-Global Matching, Fronto-Parallel Bias Reduction

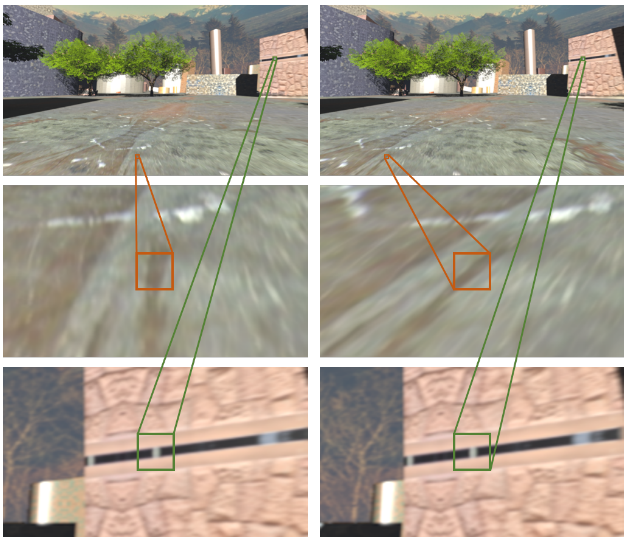
ABSTRACT:

Semi-Global Matching (SGM) is a widely-used technique for dense image matching that is popular because of its accuracy and speed. While it works well for textured scenes, it can fail on slanted surfaces particularly in wide-baseline configurations due to the so-called fronto-parallel bias. In this paper, we propose an extension of SGM that utilizes image warping to reduce the fronto-parallel bias in the data term, based on estimating dominant slanted planes. The latter are also used as surface priors improving the smoothness term. Our proposed method calculates disparity maps for each dominant slanted plane and fuses them to obtain the final disparity map. We have quantitatively evaluated our approach outperforming SGM and SGM-P on synthetic data and demonstrate its potential on real data by qualitative results. In this way, we underscore the need to tackle the fronto-parallel bias in particular for wide-baseline configurations in both the data term and the smoothness term of SGM.

1. INTRODUCTION

Structure from Motion (SfM) and Multi-View Stereo (MVS) are fundamental tasks in Computer Vision and Photogrammetry. In order to obtain dense 3D information about a scene from a set of 2D images, SfM first simultaneously estimates sparse 3D geometry (structure) and camera poses (motion). From this, MVS then reconstructs a dense 3D point cloud. Both steps are based on the establishment of correspondences between images (image matching). While there are approaches that can cope with wide baselines for sparse image matching (Mishkin et al., 2015; Roth et al., 2017), it still remains a challenging problem for dense image matching.

Dense image matching aims at computing the apparent motion between as many individual pixels of two images as possible. In the case of rectified images of a rigid scene, this motion is called disparity. One of the most widely used techniques for dense image matching is Semi-Global Matching (SGM) proposed by Hirschmüller (2005). It is popular because of its accuracy and speed and is, therefore, employed in a broad spectrum of applications ranging from 3D mapping (Hirschmüller, 2008; Rothermel et al., 2012; Kuhn et al., 2017), the navigation of robots and UAVs (Unmanned Aerial Vehicle) (Schmid et al., 2012) to autonomous driving (Franke et al., 2013). SGM has been implemented on different hardware architectures like GPU (Graphics Processing Unit) (Banz et al., 2011) and FPGA (Field-Programmable Gate Array) (Gehrig et al., 2009). While it works well for aerial images and terrestrial images with small baselines and sufficiently textured scenes mainly consisting of fronto-parallel surfaces, its performance drops significantly for wide-baseline images, in particular at higher resolutions, and with slanted, weakly-textured surfaces.



(a) Left image with details (b) Right image with details

Figure 1. Illustration of the fronto-parallel bias: In the bottom row, the image patches marked by the green squares are almost identical, because the disparity inside the window is constant. The disparity can be reliably determined. The image patches marked by the orange squares (middle row) are hardly similar, since image plane and object plane are not parallel. The disparity cannot be reliably determined.

The reason for this is that SGM, just like all other local or window-based methods for dense image matching, has the underlying implicit assumption that the disparity within the window being considered for the calculation of the matching cost is constant (fronto-parallel bias). However, this assumption is only fulfilled if image plane and object plane are parallel (fronto-parallel). This is for instance approximately true for the wall shown in the bottom row of Figure 1. Comparing the two windows marked by the green squares, it is clear that the disparity at this position can be reliably determined: The image patches are almost identical, i.e., all pixels have the same disparity as the center pixel. On the other hand, for the ground shown in the middle row of Figure 1, image

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-W5-69-2019 | © Authors 2019. CC BY 4.0 License.

plane and object plane are not parallel. If one compares the two windows marked by the orange squares, it is obvious that the disparity at this position cannot be reliably determined: The image patches are hardly similar, as most of the pixels have a disparity different from the center pixel.

The aspect above refers to the data term of SGM describing the cost of matching a pixel at a certain disparity. However, SGM also incorporates a smoothness term that penalizes changes in neighboring pixels' disparity similar to global methods. Since the fronto-parallel bias also occurs in the smoothness term, our main motivation is to reduce the fronto-parallel bias in both the data term and the smoothness term of SGM.

In this paper, we propose an extension of SGM that utilizes image warping to reduce the fronto-parallel bias in the data term, such that the calculation of the matching cost is no longer affected by it. For this purpose, we generate hypotheses for dominant slanted planes, using them for warping the images and as surface priors improving the smoothness term of SGM. We calculate disparity maps for each dominant slanted plane and fuse them to obtain the final disparity map.

2. RELATED WORK

In this section, we review related work, focusing on dense image matching methods that address the problem of the fronto-parallel bias in general and on SGM-based methods in particular.

Dense image matching methods are usually classified (Scharstein and Szeliski, 2002) into local and global methods. The former, also termed window-based methods, make implicit smoothness assumptions by aggregating the matching cost over a local window. Global methods, on the other hand, make explicit smoothness assumptions and then solve an optimization problem over all pixels.

Among the local methods, several approaches have been proposed that reduce the effect of the fronto-parallel bias through targeted warping of the input images. Burt et al. (1995) recommend to warp the right image of a stereo image pair to align it with a reference plane, such as the ground, before performing dense image matching. They report improved performance at lower computational cost due to the reduced disparity range. Einecke and Eggert (2013) as well as Ranft and Strauss (2014) adopt the idea of warping one of the input images, parameterized by horizontal shear and shift. While the former set the parameters manually, the latter propose a procedure that dynamically generates hypotheses based on the scene structure. Disparity maps from both, the differently warped image pairs and the original image pair, are fused to avoid that the final disparity map deteriorates in image regions not belonging to one of the planes.

Other local methods that aim at reducing the effect of the fronto-parallel bias use oriented matching windows that adapt to the scene structure. PatchMatch stereo (Bleyer et al., 2011) initializes each pixel with a random disparity as well as a randomly slanted plane and iteratively propagates these parameters to neighboring pixels. Sinha et al. (2014) perform local slanted plane sweeps around disparity planes that are estimated from sparse feature correspondences. For the final disparity map, each pixel is assigned to one of the local plane hypotheses by an efficient optimization technique based on SGM. Among all the plane-sweeping approaches that succeeded Collins (1996), the one of Gallup et al. (2007) was the first to explicitly handle slanted planes. In (Bulatov et al., 2011), triangular meshes from a sparse point cloud are used to compensate for the fronto-parallel bias.

SGM-based methods address the problem of the fronto-parallel bias either by replacing the unweighted sum over the aggregated cost of each direction in SGM with a weighted sum or by manipulating the penalties of SGM's smoothness term. Michael et al. (2013) introduce both path-dependent weights and penalties resulting in 20 parameters that are optimized by an evolutionary algorithm. Spangenberg et al. (2013) propose to weight the aggregated cost of each direction according to its compliance with the scene structure. While the above two approaches use global weights for each path, Poggi and Mattoccia (2016) predict per-pixel weights for each path, using random forests based on several disparity-based features. Random forests are also employed in (Schönberger et al., 2018), where disparity proposals estimated using features based on the aggregated cost of each direction are fused directly. SGM-Net (Seki and Pollefeys, 2017) is a CNN-based (Convolutional Neural Network) method that predicts the penalties of SGM's smoothness term. SGM-P (Scharstein et al., 2017) instead utilizes surface orientation priors to modify the penalties to favor surfaces coinciding with the expected scene structure.

Just like Burt et al. (1995), Einecke and Eggert (2013) and Ranft and Strauss (2014), our approach uses image warping to reduce the fronto-parallel bias. Nevertheless, this is novel in the context of SGM. Arguing that it is necessary to tackle the effect of the fronto-parallel bias in both the data term and the smoothness term, we adopt the approach of Scharstein et al. (2017) and incorporate it into our proposed extension. As we correct the effect of the fronto-parallel bias beforehand, there is no need to introduce a weighted sum in SGM's sum-based aggregation over the paths, such as, e.g., in (Spangenberg et al., 2013). Finally, we note that we fuse disparity maps similarly to Sinha et al. (2014).

3. ALGORITHM

Before describing our proposed extension, we first give a review of SGM and SGM-P.

3.1 SGM and SGM-P

SGM is an efficient algorithm for approximate energy minimization of a 2D Markov Random Field (MRF). It defines the energy function

$$E\left(D\right) = \sum_{\mathbf{p}} C_{\mathbf{p}}\left(d_{\mathbf{p}}\right) + \sum_{\mathbf{q} \in N_{\mathbf{p}}} V\left(d_{\mathbf{p}}, d_{\mathbf{q}}\right), \tag{1}$$

where $C_{\mathbf{p}}\left(d\right)$ is a unary data term representing the cost of matching pixel $\mathbf{p}$ at disparity $d \in \mathcal{D} = \{d_{\min}, \dots, d_{\max}\}$ and $V\left(d, d'\right)$ is a pairwise smoothness term penalizing changes in neighboring pixels' disparity:

$$V\left(d, d'\right) = \begin{cases} 0 & \text{if } d = d' \\ P_1 & \text{if } |d - d'| = 1 \\ P_2 & \text{if } |d - d'| > 1. \end{cases} \tag{2}$$

It adds a constant penalty P_1 for small changes in disparity and a larger constant penalty P_2 for all larger disparity changes. This allows an adaption to slanted surfaces, while preserving discontinuities at the same time. Unfortunately, this also introduces a fronto-parallel bias in the smoothess term.

As minimizing $E(D)$ from Eq. (1) is NP-hard, SGM divides the grid-shaped problem into multiple one-dimensional problems that can be efficiently solved via dynamic programming by defining an aggregated cost $L_{\mathbf{r}}(\mathbf{p}, d)$ along a path in the direction $\mathbf{r}$:

$$L_{\mathbf{r}}(\mathbf{p}, d) = C_{\mathbf{p}}(d) + \min_{d' \in \mathcal{D}} \left(L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d') + V(d, d') \right). \tag{3}$$

The aggregated cost $L_{\mathbf{r}}(\mathbf{p}, d)$ is recursively computed from the image boundaries for eight cardinal directions $\mathbf{r}$ and summed up at each pixel, resulting in the aggregated cost volume

$$S(\mathbf{p}, d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d). \tag{4}$$

The final disparity at each pixel is chosen by a winner-takes-all strategy:

$$d_{\mathbf{p}} = \arg\min_{d} S(\mathbf{p}, d). \tag{5}$$

The sum of the minima of the aggregated cost $L_{\mathbf{r}}(\mathbf{p}, d)$ of these eight paths represents a lower bound for the minimum of the aggregated cost volume $S(\mathbf{p}, d)$ for each pixel $\mathbf{p}$. The difference between these two quantities defines an uncertainty measure $U_{\mathbf{p}}$ (Drory et al., 2014):

$$U_{\mathbf{p}} = \min_{d} \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d) - \sum_{\mathbf{r}} \min_{d} L_{\mathbf{r}}(\mathbf{p}, d). \tag{6}$$

If the minima of the aggregated cost of all eight directions agree (they all occur at the same disparity), then $U_{\mathbf{p}}$ equals zero. This is often the case in image regions with textured, fronto-parallel surfaces, where wrong disparities would lead to high matching costs. In image regions with weakly-textured, slanted surfaces, instead, different disparities can cause similarly high matching costs. Therefore, the minima of the aggregated costs probably occur at different disparities, causing $U_{\mathbf{p}}$ to be different from (greater than) zero. We use this uncertainty measure to fuse different disparity maps.

In SGM-P, surface priors are utilized to modify the penalties in SGM's smoothness term to favor these surfaces. This is done by first rasterizing a real-valued disparity surface prior S, as SGM uses discrete (integer) disparities:

$$\hat{S}(\mathbf{p}) = \operatorname{round} S(\mathbf{p}) \tag{7}$$

with steps (or jumps)

$$j_{\mathbf{p}} = \hat{S}(\mathbf{p}) - \hat{S}(\mathbf{p} - \mathbf{r}) \tag{8}$$

for the discretized disparities $\hat{S}$. The original smoothness term V is then replaced with

$$V_S\left(d_{\mathbf{p}}, d'_{\mathbf{p}}\right) = V\left(d_{\mathbf{p}} + j_{\mathbf{p}}, d'_{\mathbf{p}}\right). \tag{9}$$

By this means, the zero-cost transitions coincide with the disparity jumps. As we want to tackle the fronto-parallel bias in both the data term and the smoothness term, we incorporate SGM-P into our proposed extension and feed it with the same hypotheses for dominant slanted planes that we use for warping the images.

3.2 Generation of Hypotheses for Dominant Slanted Planes

Since our approach (see Algorithm 1) is to be used in the classic SfM/MVS pipeline, we assume that a sparse point cloud is available from SfM. From this sparse point cloud, we generate hypotheses for dominant slanted planes Π and use them for warping the images. It is not our goal to improve the disparity map over the entire image by finding as many planes as possible. We aim at improving the disparity map in image regions that could only be poorly reconstructed or are partially or even completely missing due to the fronto-parallel bias by only considering dominant slanted planes. In urban environments, these often are the ground, facade or roof planes. We use RANSAC (Random Sample Consensus) (Fischler and Bolles, 1981; Schnabel et al., 2007) to find these dominant slanted planes in the sparse SfM point cloud.

```
Input: rectified stereo image pair, sparse SfM point cloud
  (optional)
Output: disparity map $D$
Variables: SGM parameters, RANSAC parameters, $\alpha_{\text{fp}}$

Calculate disparity map $D_0$ and uncertainty map $U_0$
  with original SGM
Generate hypotheses for dominant slanted planes $\Pi$
  with RANSAC from sparse SfM point cloud (or disparity
  map) discarding almost fonto-parallel planes
for each dominant slanted plane $\pi_i \in \Pi = \{\pi_1, \ldots, \pi_n\}$
  do
    Estimate approximate image extent of $\pi_i$ with
      GrabCut
    Calculate disparity map $D_{\pi_i}$ and uncertainty map
      $U_{\pi_i}$ with our proposed extension of SGM improving
      the data term by image warping with $H_{\pi_i}$ and
      improving the smoothness term by manipulating
      penalties according to $S_{\pi_i}$
end
Fuse disparity maps $D_0, D_{\pi_1}, \ldots, D_{\pi_n}$ to final disparity
  map $D$ based on uncertainty maps $U_0, U_{\pi_1}, \ldots, U_{\pi_n}$
  with SGM
```

Algorithm 1. Our proposed method.

If no sparse point cloud is available, the disparity map D_0 calculated with original SGM in the first step of our algorithm (cf. Algorithm 1) is used to search for dominant slanted planes. In this case, the search is performed in disparity space rather than in 3D space. Since we only consider dominant slanted planes, calculating the disparity map D_0 is always necessary to obtain a complete disparity map D at the end. Almost fronto-parallel planes, for which the angle between the normal and the cameras' orientation is smaller than α_{fp} (we used an empirically determined angle $\alpha_{\text{fp}} = 60°$ in our experiments), are discarded and not further considered. For these planes, reliable estimates should already be obtained by the disparity map D_0.

3.3 Improving SGM's Data Term and Smoothness Term

Based on the generated hypotheses for dominant slanted planes Π, we utilize image warping to improve the data term of SGM. In our proposed extension, we particularly warp the right image so that the window which is considered for the calculation of the matching cost coincides with the left image with respect to the considered plane. We use a plane-induced homography (Hartley and Zisserman, 2004). With the left camera placed at the origin and the camera projection matrices $P_1 = K_1 [I \mid \mathbf{0}]$ and $P_2 = K_2 [R \mid \mathbf{t}]$ for the left camera and the right camera, respectively, the plane-induced homography from the left image to the right image is given by

$$H_{\pi} = K_2 \left(R - \frac{\mathbf{t}\mathbf{n}^{\top}}{a} \right) K_1^{-1} \tag{10}$$

for a plane $\pi = \left(\mathbf{n}^\top, a\right)^\top$ with normal $\mathbf{n}$ and distance a to the origin. As we map from the right image to the left image, we use the inverse of matrix H_π from Eq. (10).

The generated hypotheses for slanted planes Π are also used to calculate the surface priors to manipulate the smoothness term in SGM to favor these surfaces. The surface prior S_π for a plane π with the plane equation $n_x x + n_y y + n_z z + a = 0$ can be calculated in the following way. For a perspective camera with focal length f, we have $x = (u - u_0)\, z/f$ and $y = (v - v_0)\, z/f$ for image coordinates (u, v) with (u_0, v_0) being the camera's principal point. Substituting these quantities into the plane equation results in

$$z = -af / \left(n_x (u - u_0) + n_y (v - v_0) + f n_z\right). \qquad (11)$$

For a rectified stereo pair, we also have $z = bf/d$, where b and d are the baseline between the cameras and the disparity, respectively. The disparity d to be expected for an image point (u, v) lying on plane π is then given by

$$d\,(u, v) = -\frac{b}{a}\left(n_x (u - u_0) + n_y (v - v_0) + f n_z\right). \qquad (12)$$

Besides calculating the surface prior S_π for a plane π to modify the penalties in the smoothness term, our proposed method uses Eq. (12) to limit the disparity search space.

For each dominant slanted plane $\pi_i \in \Pi = \{\pi_1, \ldots, \pi_n\}$, our proposed extension of SGM calculates a disparity map D_{π_i} as well as an uncertainty map U_{π_i} using Eq. (6). Thus, the fronto-parallel bias is reduced in both the data term and the smoothness term.

In order to limit the calculation of the disparity map D_{π_i} to the corresponding image regions, we estimate the approximate extent of the considered dominant slanted plane π_i in the images. For this, we use image segmentation, particularly GrabCut (Rother et al., 2004) applied to the down-scaled images with the foreground pixels being initialized with regions around the 3D points belonging to the considered plane projected into the images.

3.4 Fusion of Disparity Maps

In the last step of our algorithm (see Algorithm 1), the disparity maps $D_0, D_{\pi_1}, \ldots, D_{\pi_n}$ are fused to form the final disparity map D based on the uncertainty maps $U_0, U_{\pi_1}, \ldots, U_{\pi_n}$. We follow the idea of Sinha et al. (2014) and formulate this fusion as a pixel labeling problem, where each pixel $\mathbf{p}$ has to be assigned to a label l. In our case, these labels l are equivalent to the disparity maps $D_0, D_{\pi_1}, \ldots, D_{\pi_n}$. The optimal assignment L is computed by minimizing the energy function

$$E\,(L) = \sum_{\mathbf{p}} U_{\mathbf{p}}\left(l_{\mathbf{p}}\right) + \sum_{\mathbf{q} \in N_{\mathbf{p}}} V\left(l_{\mathbf{p}}, l_{\mathbf{q}}\right), \qquad (13)$$

where the uncertainty maps $U_0, U_{\pi_1}, \ldots, U_{\pi_n}$ are used as unary data term. As there is no order among the labels l, the pairwise smoothness term differs from Eq. (2) in equally penalizing varying labels between neighboring pixels with a constant penalty P. We also use SGM to efficiently obtain an approximate solution for this optimization problem and the final disparity map D.

Figure 2 exemplarily shows the individual steps of our approach for the image pair from Figure 1.

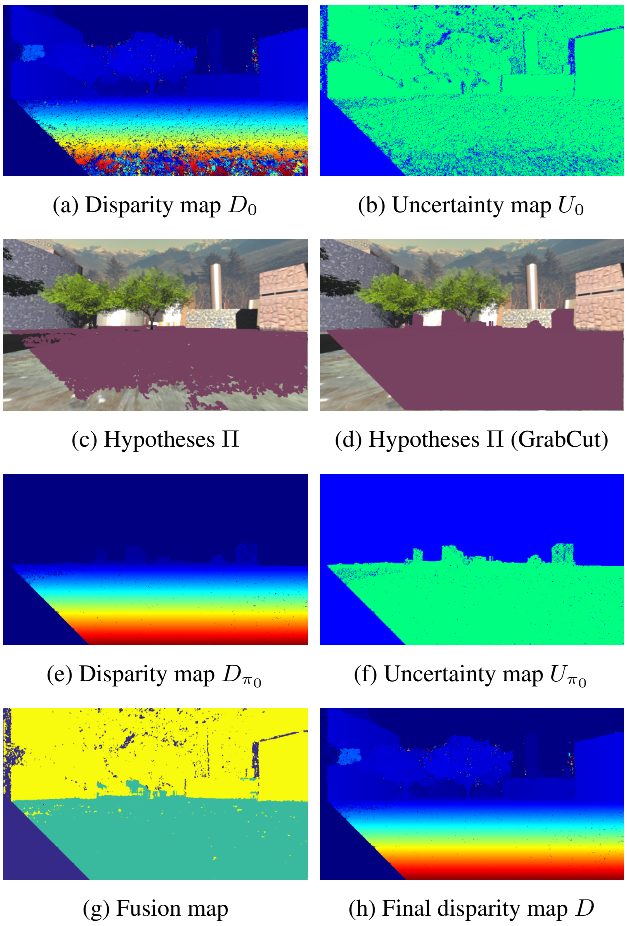

(a) Disparity map D_0
(b) Uncertainty map U_0
(c) Hypotheses Π
(d) Hypotheses Π (GrabCut)
(e) Disparity map D_{π_0}
(f) Uncertainty map U_{π_0}
(g) Fusion map
(h) Final disparity map D

Figure 2. Individual steps of our approach exemplarily shown for the image pair from Figure 1.

4. EXPERIMENTS

We report a quantitative evaluation on synthetic data as well as qualitative results on real data demonstrating the potential of our proposed method. In this way, we underscore the need to tackle the fronto-parallel bias in both the data term and the smoothness term of SGM, in particular for wide-baseline configurations.

4.1 Implementation Details

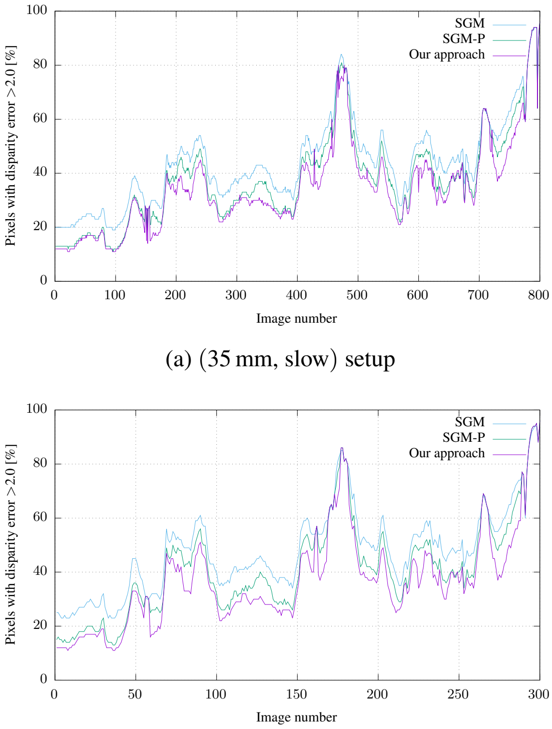
We compare our approach against SGM and SGM-P, emphasizing the individual contributions of reducing the fronto-parallel bias in the data term and the smoothness term. In order to ensure an unbiased evaluation, we build on the same implementation of SGM for all experiments. We use the OpenCV implementation (Bradski, 2000) extended to allow the Census transform (Zabih and Woodfill, 1994) to be used as a cost function and to calculate uncertainty maps. For all experiments, the employed parameters of SGM's smoothness term are $P_1 = 8$ and $P_2 = 32$. For the data term, we rely on the Census transform calculated on patches of a size of 7×7 pixels. Since we use SGM also for the fusion of disparity maps, we note that, in this case, the constant penalty in the smoothness term is $P = 32$.

When comparing our proposed method against SGM-P, we use identical hypotheses for dominant slanted planes for both. In contrast to (Scharstein et al., 2017), where a single surface prior is derived, we create a surface prior for each dominant slanted plane and fuse the corresponding disparity maps, just like for our approach, to allow a fair comparison. By this means, differences in the final disparity maps from SGM-P and our proposed method express the effect of additionally reducing the fronto-parallel bias in the data term. On the other hand, the effect of reducing the fronto-parallel bias in both the data term and the smoothness term can be seen when comparing the final disparity maps from original SGM and our approach.

If the dominant slanted planes are estimated in the (isotropic) disparity space, the RANSAC parameters are fixed to 0.99 for the confidence threshold and to 2.0 for the distance threshold. For sparse (anisotropic) SfM point clouds, the RANSAC parameters have to be adapted. We discard almost fronto-parallel planes, for which the angle between the normal and the cameras' orientation is smaller than $\alpha_{\text{fp}} = 60^\circ$ (empirically determined). These parameters are the same for SGM-P as for our proposed method.

4.2 Quantitative Evaluation

We start by evaluating our approach on the Driving dataset of Mayer et al. (2016). This synthetic dataset inspired by the KITTI dataset (Geiger et al., 2012) provides between 300 and 800 stereo image pairs with a resolution of 960 × 540 pixels for each setup. Besides the virtual focal length (15 mm or 35 mm), the setups differ in the "speed" they were recorded (fast or slow), causing more or less motion blur and defocus blur. We consider the following four setups, using only the backwards scenes: (35 mm, slow), (35 mm, fast), (15 mm, slow) and (15 mm, fast).

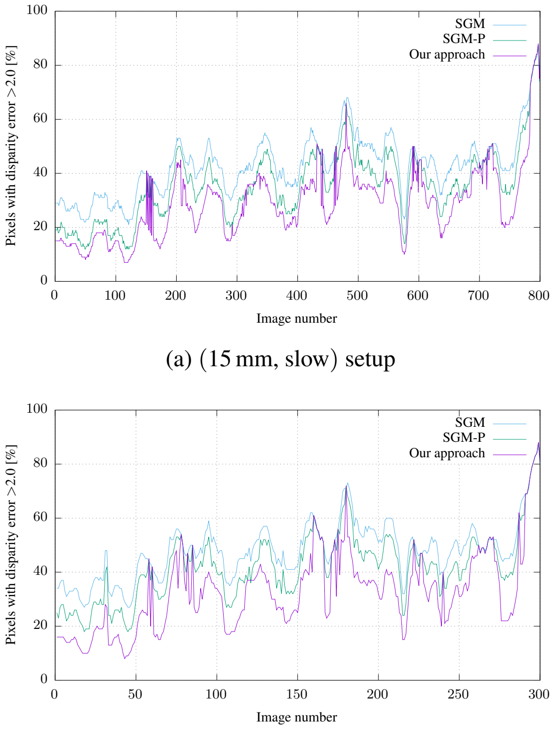


(a) (35 mm, slow) setup

(b) (35 mm, fast) setup

Figure 3. Performance on the Driving dataset (35 mm setups).

As ground-truth disparity maps are available for all setups, we use the Driving dataset to quantitatively evaluate our proposed method. For this purpose, we consider the disparity error, i.e., the percentage of pixels in the image whose disparity differs by more than 2.0 pixels from the ground truth. In Figures 3 and 4, this disparity error is plotted against the image number for the 35 mm and the 15 mm setups, respectively, comparing our approach with SGM and SGM-P. For all four setups, our approach's curve is below that of SGM and SGM-P, indicating that it clearly outperforms the other two. Our proposed method performs virtually never worse than SGM or SGM-P. It performs worst if no dominant slanted planes are found. In this case, our approach just returns the SGM disparity map. This happens several times for image pairs around 150 in the slow setups. Since the fast and slow setups come from the same trajectories, but with different distances, the shape of the curves is similar (cf., e.g., Figures 3a and 3b).



(a) (15 mm, slow) setup

(b) (15 mm, fast) setup

Figure 4. Performance on the Driving dataset (15 mm setups).

Table 1 shows the mean disparity error reduction ranging from 23 to 35% over SGM and from 8 to 24% over SGM-P. It is evident that the improvement decreases from the 15 mm (wide-baseline) to the 35 mm (small-baseline) setups. This is in particular true for SGM-P. The results prove what intuition tells us: Our approach is particularly suitable for wide-baseline image pairs, whereas for small-baseline image pairs, the improvement over SGM-P is significantly smaller. This is the reason for us refraining from a quantitative evaluation on well-known stereo benchmarks such as KITTI or ETH3D (Schöps et al., 2017) with small-baseline image pairs. Instead, we focus on qualitative results demonstrating the potential of our proposed method on meaningful examples.

	Error reduction over SGM [%]	Error reduction over SGM-P [%]
(35 mm, slow)	23.03	8.31
(35 mm, fast)	25.40	11.45
(15 mm, slow)	35.04	20.25
(15 mm, fast)	34.35	24.90

Table 1. Mean disparity error reduction over SGM and SGM-P on the Driving dataset.

https://doi.org/10.5194/isprs-annals-IV-2-W5-69-2019

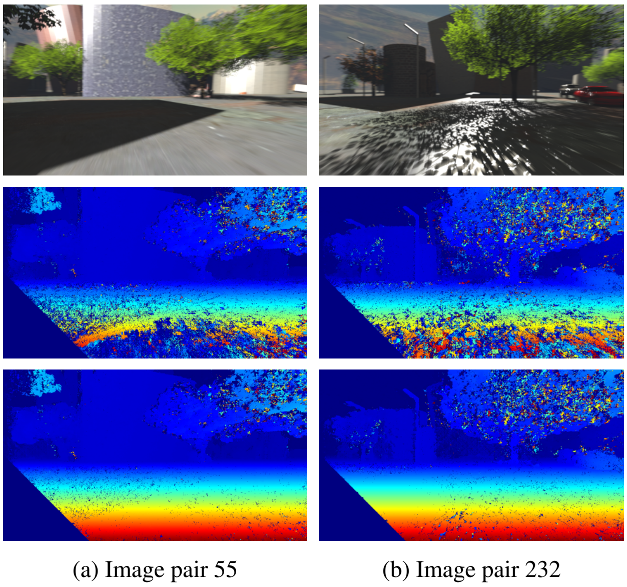

(a) Image pair 55 (b) Image pair 232

Figure 5. Qualitative results for two image pairs from the Driving dataset, (15 mm, fast) setup. *Top:* Left image. *Middle:* Disparity map (SGM). *Bottom:* Disparity map (our approach).

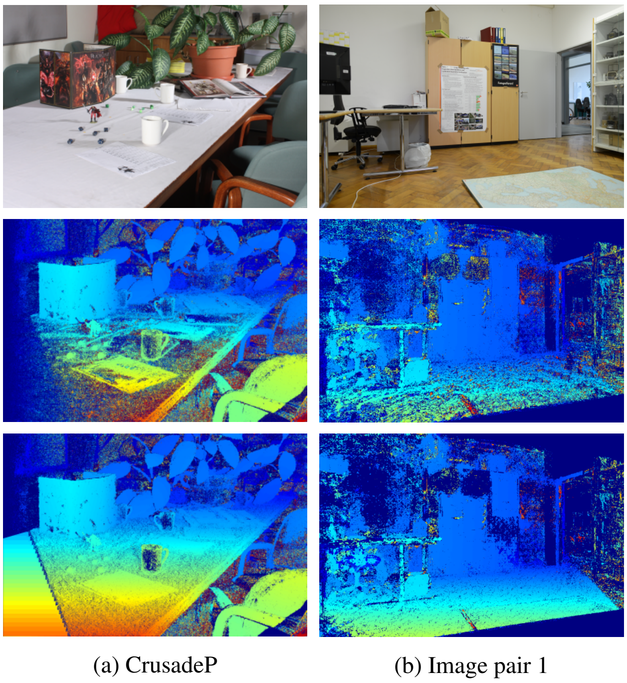

(a) CrusadeP (b) Image pair 1

Figure 6. Qualitative results for two image pairs. CrusadeP is from the Middlebury dataset, image pair 1 acquired by us. *Top:* Left image. *Middle:* Disparity map (SGM). *Bottom:* Disparity map (our approach).

4.3 Qualitative Results

Besides examples from the synthetic Driving dataset, we demonstrate the potential of our approach on image pairs from the Middlebury dataset (Scharstein et al., 2014), from the multi-view dataset of Strecha et al. (2008), and from own images. While for the Driving and the Middlebury dataset, dominant slanted planes are estimated in the disparity space, for the others these are estimated in the sparse point cloud obtained using the wide-baseline SfM technique of Mayer et al. (2012) as well as Michelini and Mayer (2016). The resolution is about six megapixels across all image pairs.

Two examples from the Driving dataset are shown in Figure 5. Due to the characteristics of the dataset, usually the ground and, more rarely, facade planes are found as dominant slanted planes. In particular, the strongly distorted image regions in the foreground are completely reconstructed by our proposed method in contrast to SGM. As the CrusadeP image pair from the Middlebury dataset and image pair 1 in Figure 6 prove, our approach is not limited to synthetic KITTI-like data. Nevertheless, our proposed method strongly relies on the scene structure, presuming dominant slanted planes. We aim at reconstructing these image regions, as they are potentially missing in the disparity map due to the fronto-parallel bias. For most of the image pairs from the Middlebury dataset, we did not succeed in finding dominant slanted planes. Please note that, in this case, our approach still returns the SGM disparity map. As use case we mainly concentrate on scenes in urban environments, Figure 7 shows two examples from the multi-view dataset of Strecha et al. (2008), more precisely the fountain-P11 (7,4) image pair and the Herz-Jesu-P25 (15,16) image pair, along with two more examples acquired by us. Our approach significantly improves the disparity maps in image regions belonging to the ground for all four image pairs. In addition, the roof which is largely missing in the disparity map of SGM is reconstructed for image pair 3.

5. CONCLUSION

We have proposed an extension of SGM that tackles the fronto-parallel bias in both the data term and the smoothness term. It utilizes image warping to reduce the fronto-parallel bias in the data term. Hypotheses for dominant slanted planes are generated either from the sparse SfM point cloud or from the SGM disparity map, being used as surface priors to improve the smoothness term. Our approach calculates disparity maps for each dominant slanted plane and fuses them to obtain the final disparity map.

Our proposed method has been quantitatively evaluated on synthetic data, where it outperforms SGM and SGM-P, underscoring the need to tackle the fronto-parallel bias in both the data term and the smoothness term of SGM, in particular for wide-baseline configurations. Qualitative results on real data demonstrate its potential.

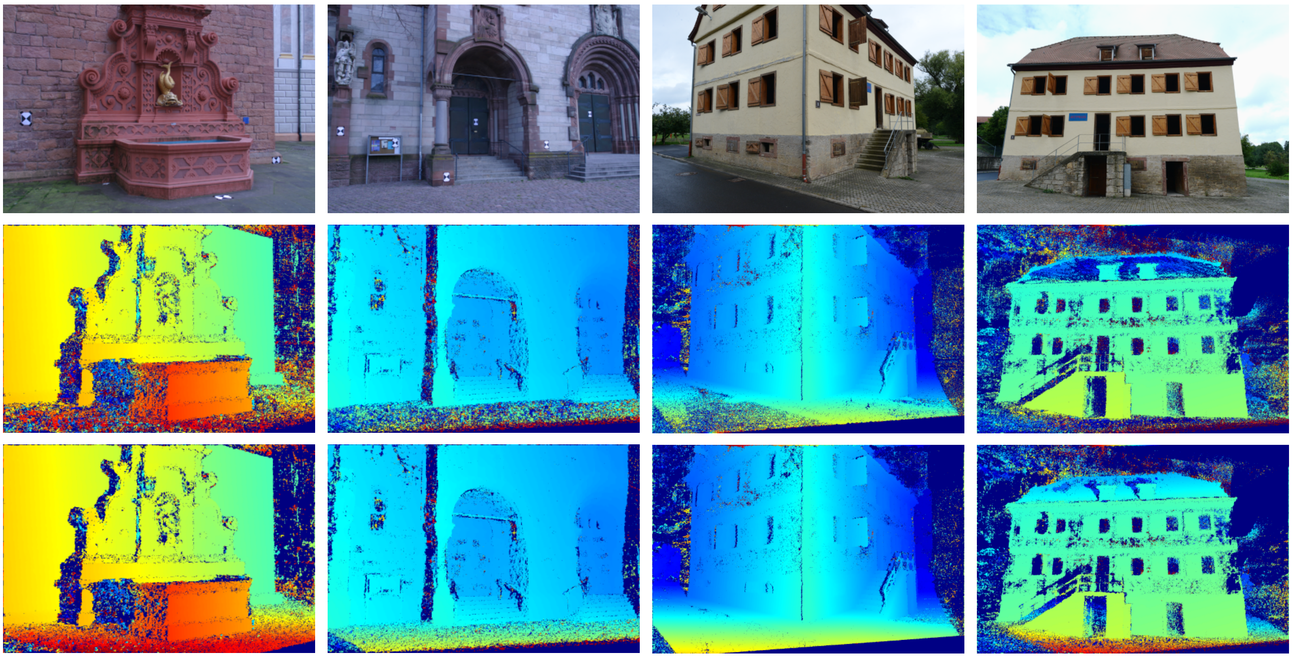
As our approach strongly relies on the robust detection of dominant slanted planes, future work includes assisting their detection by semantic image analysis.

References

Banz, C., Blume, H. and Pirsch, P., 2011. Real-Time Semi-Global Matching Disparity Estimation on the GPU. In: *2011 IEEE International Conference on Computer Vision Workshops*, IEEE, pp. 514–521.

Bleyer, M., Rhemann, C. and Rother, C., 2011. PatchMatch Stereo - Stereo Matching with Slanted Support Windows. In: *British Machine Vision Conference 2011*, British Machine Vision Association, pp. 14.1–14.11.

Bradski, G., 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.



(a) Fountain-P11 (b) Herz-Jesu-P25 (c) Image pair 2 (d) Image pair 3

Figure 7. Qualitative results for four image pairs. *Top:* Left image. *Middle:* Disparity map (SGM). *Bottom:* Disparity map (our approach).

Bulatov, D., Wernerus, P. and Heipke, C., 2011. Multi-View Dense Matching Supported by Triangular Meshes. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(6), pp. 907–918.

Burt, P., Wixson, L. and Salgian, G., 1995. Electronically Directed ”Focal” Stereo. In: *1995 IEEE International Conference on Computer Vision (ICCV)*, IEEE, pp. 94–101.

Collins, R. T., 1996. A Space-Sweep Approach to True Multi-Image Matching. In: *1996 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 358–363.

Drory, A., Haubold, C., Avidan, S. and Hamprecht, F. A., 2014. Semi-Global Matching: A Principled Derivation in Terms of Message Passing. In: *Pattern Recognition*, Lecture Notes in Computer Science, Vol. 8753, Springer International Publishing, Cham, pp. 43–53.

Einecke, N. and Eggert, J., 2013. Stereo Image Warping for Improved Depth Estimation of Road Surfaces. In: *2013 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 189–194.

Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* 24(6), pp. 381–395.

Franke, U., Pfeiffer, D., Rabe, C., Knoeppel, C., Enzweiler, M., Stein, F. and Herrtwich, R. G., 2013. Making Bertha See. In: *2013 IEEE International Conference on Computer Vision Workshops*, IEEE, pp. 214–221.

Gallup, D., Frahm, J.-M., Mordohai, P., Yang, Q. and Pollefeys, M., 2007. Real-Time Plane-Sweeping Stereo with Multiple Sweeping Directions. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1–8.

Gehrig, S. K., Eberli, F. and Meyer, T., 2009. A Real-Time Low-Power Stereo Vision Engine Using Semi-Global Matching. In: *Computer Vision Systems*, Lecture Notes in Computer Science, Vol. 5815, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 134–143.

Geiger, A., Lenz, P. and Urtasun, R., 2012. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 3354–3361.

Hartley, R. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge.

Hirschmüller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 807–814.

Hirschmüller, H., 2008. Stereo Processing by Semi-Global Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), pp. 328–341.

Kuhn, A., Hirschmüller, H., Scharstein, D. and Mayer, H., 2017. A TV Prior for High-Quality Scalable Multi-View Stereo Reconstruction. *International Journal of Computer Vision* 124(1), pp. 2–17.

Mayer, H., Bartelsen, J., Hirschmüller, H. and Kuhn, A., 2012. Dense 3D Reconstruction from Wide Baseline Image Sets. In: *Outdoor and Large-Scale Real-World Scene Analysis*, Lecture Notes in Computer Science, Vol. 7474, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 285–304.

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A. and Brox, T., 2016. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 4040–4048.

https://doi.org/10.5194/isprs-annals-IV-2-W5-69-2019

Michael, M., Salmen, J., Stallkamp, J. and Schlipsing, M., 2013. Real-Time Stereo Vision: Optimizing Semi-Global Matching. In: *2013 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 1197–1202.

Michelini, M. and Mayer, H., 2016. Efficient Wide-Baseline Structure from Motion. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III-3, pp. 99–106.

Mishkin, D., Matas, J. and Perdoch, M., 2015. MODS: Fast and Robust Method for Two-View Matching. *Computer Vision and Image Understanding* 141, pp. 81–93.

Poggi, M. and Mattoccia, S., 2016. Learning a General-Purpose Confidence Measure Based on O(1) Features and a Smarter Aggregation Strategy for Semi-Global Matching. In: *2016 International Conference on 3D Vision (3DV)*, IEEE, pp. 509–518.

Ranft, B. and Strauss, T., 2014. Modeling Arbitrarily Oriented Slanted Planes for Efficient Stereo Vision Based on Block Matching. In: *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1941–1947.

Roth, L., Kuhn, A. and Mayer, H., 2017. Wide-Baseline Image Matching with Projective View Synthesis and Calibrated Geometric Verification. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 85(2), pp. 85–95.

Rother, C., Kolmogorov, V. and Blake, A., 2004. Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Transactions on Graphics* 23(3), pp. 309.

Rothermel, M., Wenzel, K., Fritsch, D. and Haala, N., 2012. SURE: Photogrammetric Surface Reconstruction from Imagery. In: *LC3D Workshop*, pp. 1–9.

Scharstein, D. and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision* 47(1/3), pp. 7–42.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X. and Westling, P., 2014. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. In: *Pattern Recognition*, Lecture Notes in Computer Science, Vol. 8753, Springer International Publishing, Cham, pp. 31–42.

Scharstein, D., Taniai, T. and Sinha, S. N., 2017. Semi-Global Stereo Matching with Surface Orientation Priors. In: *2017 International Conference on 3D Vision (3DV)*, IEEE, pp. 215–224.

Schmid, K., Hirschmüller, H., Dömel, A., Grixa, I., Suppa, M. and Hirzinger, G., 2012. View Planning for Multi-View Stereo 3D Reconstruction Using an Autonomous Multicopter. *Journal of Intelligent & Robotic Systems* 65(1-4), pp. 309–323.

Schnabel, R., Wahl, R. and Klein, R., 2007. Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum* 26(2), pp. 214–226.

Schönberger, J. L., Sinha, S. N. and Pollefeys, M., 2018. Learning to Fuse Proposals from Multiple Scanline Optimizations in Semi-Global Matching. In: *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, Vol. 11217, Springer International Publishing, Cham, pp. 758–775.

Schöps, T., Schönberger, J. L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M. and Geiger, A., 2017. A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 2538–2547.

Seki, A. and Pollefeys, M., 2017. SGM-Nets: Semi-Global Matching with Neural Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 6640–6649.

Sinha, S. N., Scharstein, D. and Szeliski, R., 2014. Efficient High-Resolution Stereo Matching Using Local Plane Sweeps. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1582–1589.

Spangenberg, R., Langner, T. and Rojas, R., 2013. Weighted Semi-Global Matching and Center-Symmetric Census Transform for Robust Driver Assistance. In: *Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, Vol. 8048, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 34–41.

Strecha, C., Hansen, W. v., van Gool, L., Fua, P. and Thoennessen, U., 2008. On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1–8.

Zabih, R. and Woodfill, J., 1994. Non-Parametric Local Transforms for Computing Visual Correspondence. In: *Computer Vision — ECCV '94*, Lecture Notes in Computer Science, Vol. 801, Springer-Verlag, Berlin/Heidelberg, pp. 151–158.