

# PatchMatch Stereo - Stereo Matching with Slanted Support Windows

Michael Bleyer<sup>1</sup>  
bleyer@ims.tuwien.ac.at  
Christoph Rhemann<sup>1</sup>  
rhemann@ims.tuwien.ac.at  
Carsten Rother<sup>2</sup>  
carrot@microsoft.com

<sup>1</sup>Institute of Software Technology  
Vienna University of Technology  
Vienna, Austria  
<sup>2</sup>Microsoft Research Cambridge  
Cambridge, UK



Figure 1: Advantage of our method. (a) The disparity map computed with fronto-parallel windows approximates slanted surfaces via many fronto-parallel ones. (b) Our slanted support windows correctly reconstruct the scene as a collection of slanted planar surfaces.

**Introduction** In local stereo matching, a support window is centered on a pixel of the reference frame. This support window is then displaced in the second image to find the point of lowest color dissimilarity, which represents the matching point. There is an implicit assumption in this procedure, i.e., all pixels within the support window have constant disparity. This assumption is broken for slanted surfaces where the window captures many different disparities. The effect is that slanted surfaces are approximated by several fronto-parallel planes as seen in Figure 1a. This work overcomes this problem by estimating an individual slanted 3D plane at each pixel onto which the support region is projected. The advantage is shown in Figure 1b where our slanted support windows correctly reconstruct the highly slanted surfaces of the Corridor scene with sub-pixel precision (see floor, ceiling and walls).

Our method is not the first one that applies slanted support windows. Previous methods (e.g., [2, 4]) typically select a set of planes in the first step. In the second step, each pixel is assigned to one of the selected planes. These algorithms fail if a plane required to correctly reconstruct the scene has been missed in step 1. In this work we propose an algorithm based on PatchMatch [1] to effectively solve the problem of finding a slanted support plane at each pixel. This enables a one-shot optimization where planes and assignments of pixels to planes are estimated jointly. We extend PatchMatch’s ideas of random search and propagation to find the nearest neighbor on the epipolar line according to a plane.

**Method** The basic idea is that in natural stereo pairs relatively large regions of pixels can be modeled by approximately the same plane. We find the plane for a region by initializing each pixel with a random plane. The hope is that after this random initialization at least one pixel of the region carries a plane that is close to the correct one. Note that this is very likely because we have many guesses, i.e., each pixel of the region represents one. Having a single “good” guess is already enough for the algorithm to work, since there is a propagation step that passes this plane on to the other pixels of the region. Apart from propagating the plane among spatial neighboring pixels, this paper introduces two new propagation steps. These are: (1) view propagation where planes are propagated among left and right views of the stereo pair and (2) temporal propagation where planes are propagated from preceding and consecutive frames of a video when doing temporal stereo. Finally, there is also a plane refinement step where we alter plane parameters to get closer to the optimal plane.

Our algorithm uses adaptive support weights [3] in the matching cost computation to considerably improve matching results in the proximity of disparity discontinuities. Apart from applying our slanted adaptively weighted support windows for local stereo matching, we show that they can be used to compute a cost volume for global stereo methods. The advantage of our cost volume is that it incorporates color segmentation into a global algorithm in a simple way and without introducing a bias towards fronto-parallel surfaces. Moreover, our cost volume allows global algorithms to use robust radiometric insensitive matching costs (e.g., NCC or Census) that only work when operating on matching windows.

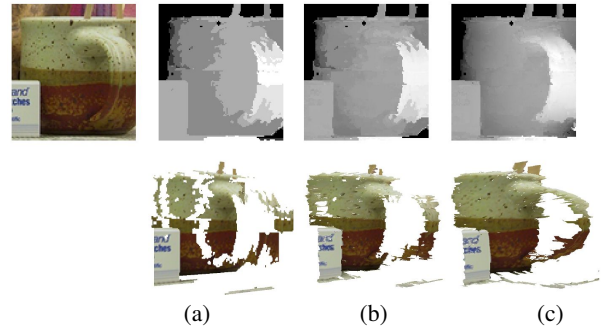


Figure 2: Results of different methods on a crop of the Middlebury Cones image. (a) Fronto-parallel windows matched at integer-valued disparities. (b) Fronto-parallel windows matched at continuous sub-pixel disparities. (c) Our slanted support windows.

**Results** To evaluate our algorithm, we implement two competitors in our PatchMatch framework, i.e. (1) fronto-parallel windows matched at integer disparities and (2) fronto-parallel windows matched at continuous sub-pixel disparities. Figure 2 compares our results against the two competitors and quantitative Middlebury results are found in the paper. In the Middlebury table, our method is currently the top-performer among local methods and takes rank 11 out of 110 submissions in the overall ranking. Our algorithm performs specifically well on the complex Teddy set where it is the Middlebury top-performer (see Figure 3). PatchMatch Stereo shows its full potential when looking at sub-pixel precision (also see reconstruction results in Figure 2c and in the paper). For the Middlebury error threshold 0.5 that is used for measuring sub-pixel performance, our method takes rank 2 and is the clear top-performer for the more complex Middlebury scenes Teddy and Cones. It takes approximately a minute to compute a disparity map for the Middlebury set using our CPU-based implementation. In the paper and supplementary material, we also show results for high-resolution videos.

- [1] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. Patch-match: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (SIGGRAPH)*, 2009.
- [2] D. Gallup, J. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys. Real-time plane-sweeping stereo with multiple sweeping directions. In *CVPR*, 2007.
- [3] K.J. Yoon and I.S. Kweon. Locally adaptive support-weight approach for visual correspondence search. In *CVPR*, 2005.
- [4] Y. Zhang, M. Gong, and Y. Yang. Local stereo matching with 3D adaptive cost aggregation for slanted surface modeling and sub-pixel accuracy. In *ICPR*, 2008.

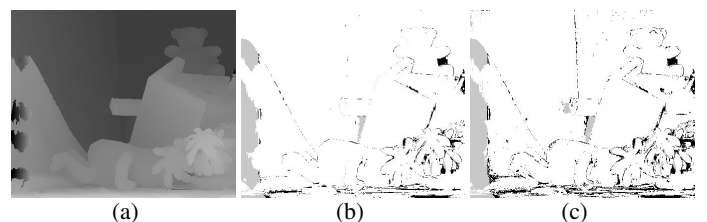


Figure 3: Results on the Middlebury Teddy set. Our method is the Middlebury top-performer on Teddy among 110 submissions (according to the disparity error in unoccluded regions). (a) Disparity Map. Note that our algorithm can correctly reconstruct the highly-slanted ground plane. (b) Error map. We plot pixels having a disparity error > 1.0 pixel (Middlebury default threshold) in black. (c) We plot pixels having a disparity error > 0.5 pixels for measuring sub-pixel performance.