

Spatiotemporal Stereo Matching with 3D Disparity Profiles

Yongho Shin
yongho@gist.ac.kr
Kuk-Jin Yoon
kjyoon@gist.ac.kr

Computer Vision Laboratory
Gwangju Institute of Science and
Technology (GIST), South Korea

Adaptive support weights and over-parameterized disparity estimation truly improve the accuracy of stereo matching by enabling window-based similarity measures to handle depth discontinuities and non-fronto-parallel surfaces more effectively. Nevertheless, a disparity map sequence obtained in a frame-by-frame manner still tends to be inconsistent even with the use of state-of-the-art stereo matching methods. To solve this inconsistency problem, we propose a window-based spatiotemporal stereo matching method exploiting 3D disparity profiles.

Energy modeling and minimization For a rectified stereo image sequence, we employ an energy minimization scheme for each frame. Our energy function is defined by an over-parameterized disparity $\mathbf{h} = [z \ n_{xz} \ n_{yz}]$ and a guidance disparity $\tilde{\mathbf{h}} = [\tilde{z} \ \tilde{n}_{xz} \ \tilde{n}_{yz}]^T$ as

$$E(\mathbf{H}, \tilde{\mathbf{H}}) = \sum_{p \in P} D_p(\mathbf{h}_p) + \sum_{p \in P} \sum_{q \in N_p} V_{p,q}(\mathbf{h}_p, \mathbf{h}_q) + \sum_{p \in P} T_p(\mathbf{h}_p, \tilde{\mathbf{h}}_p). \quad (1)$$

D_p is the data cost, and it is computed by the combination of census transform and mutual information. $V_{p,q}$ is the smoothness cost that encourages a smooth disparity map. T_p is the temporal cost that encourages consistent disparity values between frames, and it is defined with guidance disparity and normal values, \tilde{z}_p , $\tilde{n}_{xz,p}$ and $\tilde{n}_{yz,p}$, as

$$T_p(\mathbf{h}_p, \tilde{\mathbf{h}}_p) = \left[\min(|z_p - \tilde{z}_p|, \tau_z) + \min(\rho \cdot (|n_{xz,p} - \tilde{n}_{xz,p}| + |n_{yz,p} - \tilde{n}_{yz,p}|), \tau_n) \right]. \quad (2)$$

Ideally, \tilde{z}_p , $\tilde{n}_{xz,p}$ and $\tilde{n}_{yz,p}$ should be ground truth values. However, since ground truth values are not given, we determine guidance values by exploiting a 3D disparity profile which is a structure representing disparities and normal vectors of a window over multiple frames. In general, objects and cameras in a scene can be assumed to move with smooth variations in direction and velocity. In this case, the window shape and location must also be changed smoothly. It means that smoothly varying disparity and normal values between frames are more reliable than largely varying disparity and normal values. Hence, we set values that can make disparity and normal values between consecutive frames smoothly varied as guidance values. To obtain such guidance values, we compose a disparity profile, and fit the profile to a polynomial function.

We find the optimal solution to Eq.(1) in the PatchMatch Belief Propagation (PMBP) framework. PMBP starts with the initialization of particles with random sampling. Afterwards, in each iteration, particles are propagated from neighborhoods q to pixel p . Our propagation step consists of spatial, view, and temporal propagation. If the disbelief of a propagated state is lower than that of current states of p , the propagated state is selected as a new member of the particle set of p . After the propagation step, a random search around the state of p is performed for state refinement as in [2]. Before the next iteration, we update data with an intermediate disparity map sequence having the minimum energy in the current iteration. First, we update optical flows. Next, we update joint histograms for mutual information. Lastly, we update guidance values by fitting the disparity profile from the intermediate disparity map sequence. After a few iterations, the particle with the minimum disbelief is selected as the optimal solution of p .

Temporal correspondence establishment To obtain the 3D disparity profile, we must establish the temporal correspondences across frames. To this end, we employ the LDOF method [3] to handle large displacements by rapidly moving object or cameras. And, a forward-backward optical flow consistency check is used for filtering out incorrect flows. However, the consistency check makes the profile fragmentary. To resolve this problem, we propose an optical flow transfer method for establishing more reliable disparity profile by using optical flows of both views.

Experimental results We evaluated our method using a synthetic dataset presented in [4] and a real dataset provided by KITTI. For comparison,

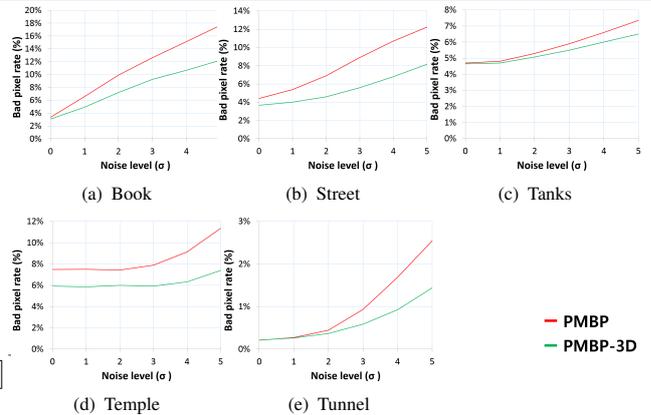


Figure 1: Average bad-pixel rates for synthetic data sets according to increasing noise level

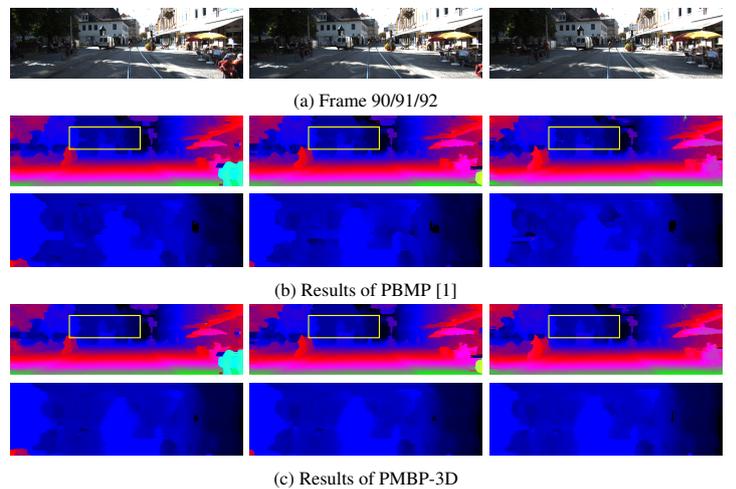


Figure 2: Results for real dataset.

we selected the original PMBP-based method [1] as a baseline. Here, the baseline and proposed methods are referred to as PMBP and PMBP-3D, respectively. To see clearly the consistency improvement, we analyzed the accuracy of disparity maps for synthetic image sequences with noise. Figure 1 shows average bad-pixel rates according to different noise levels. As the noise level increases, average bad-pixel rates of PMBP increase faster than those of PMBP-3D. Next, we analyzed the results for a real dataset. Figure 2 illustrates disparity maps for three successive frames. To clearly compare the performance of PMBP and PMBP-3D, we provide magnified results with contrast adjustment for a specific area of the disparity map. The specific area is marked with a yellow square in the disparity map. Because the image captured for a real scene contains ambiguous regions and inevitable noise, PMBP is likely to provide temporally inconsistent disparity maps. Whereas, PMBP-3D generates a consistent disparity map sequence.

- [1] Frederic Besse, Carsten Rother, Andrew Fitzgibbon, and Jan Kautz. Pmbp: Patchmatch belief propagation for correspondence field estimation. *IJCV*, 110(1):2–13, 2014.
- [2] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo - stereo matching with slanted support windows. In *BMVC*, 2011.
- [3] Thomas Brox and Jitendra Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE TPAMI*, 33(3):500–513, 2011.
- [4] Christian Richardt, Douglas Orr, Ian Davies, Antonio Criminisi, and Neil A. Dodgson. Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid. In *ECCV*, 2010.