Supplemental Material for Submission #383: A Stochastic Cost Function for Stereo Vision

BMVC 2012 Submission # 383

In the following we present more experiments and other results which did not fit into the paper. The results of section 3 were generated with the identical implementation. However, please note that the results of section 1 and 2 were obtained previously with slightly different parameters, but the claims made are still valid and representative.

1 Parameter Analysis

The walk length N and the parameter σ_C of the random walk transition probability are the most relevant ones for the performance of our method. In Fig. 1, we present exemplary results for varying values of N and σ_C using the datasets *Tsukuba* and *Cones*. We chose these two datasets because we measured the biggest differences for them.
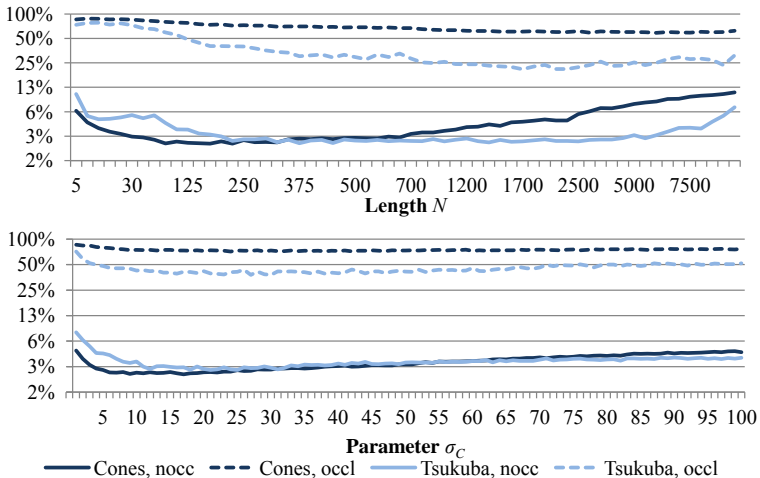

Figure 1: The charts display the influence of the parameters N and σ_C on the errors in non-occluded (nocc) and occluded (occl) areas. Errors are defined as percentages of disparities that differ by more than 1 from the ground truth and we used a logarithmic scale of the vertical axis for a better visualization. To generate these charts, we did not use the random walk based propagation. For the first chart we fixed $\sigma_C = 15$ and for the second chart $N = 200$.
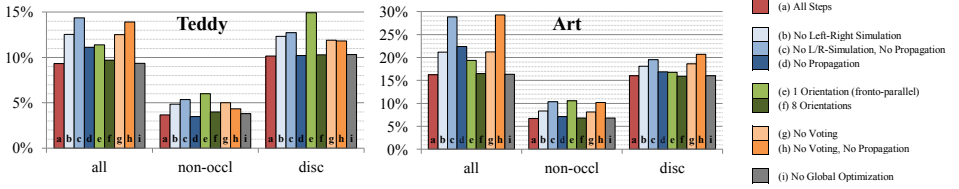
Figure 2: The charts display the effect of disabling processing steps of our cost function. Error bars show percentages of disparities that differ by more than 1 from the ground truth in the whole image (all), non-occluded pixels (non-occl) and regions near discontinuities (disc).

Small values of N introduce errors because aggregated matching costs are less discriminative in this case and thus, many wrong minimum values receive votes. Larger values of N have a positive effect on the performance in occluded regions, which can be explained by the voting strategy. If a walk covers occluded pixels and if the correct disparity is selected for voting then these occluded pixels will receive support for the correct disparity. At the same time, the errors increase in non-occluded regions because it becomes more likely that a random walk steps over discontinuities. The steeper increase at *Tsukuba* might be explained by the simple geometric structure of *Tsukuba* (there are less discontinuities than in *Teddy*).

Very small values of σ_C lead to high errors in non-occluded regions because the walks are then very sensitive to image noise. Especially at *Tsukuba*, there are some vertical artifacts present in the images, presumably a result from a Bayer-pattern, which seem to lead to higher errors for small values of N and σ_C. Larger values of σ_C gradually increase the error because it is more likely that walks cross object boundaries. In practice, good values for σ_C and N, which minimize errors in non-occluded regions, can be obtained relatively efficiently in an iterative manner. With a dense discrete parameter exploration we found that the trends described above are still valid for other parameter combinations.

2   Analysis of the Different Method Steps

Please note that in this experiment a different global energy minimization technique based on belief propagation was used. Since a very weak smoothness constraint was used, practically no influence on the results can be observed and thus, the global optimization can be ignored here.

In Fig. 2, we analyze the influence of the different processing steps of our cost function on the quality in different image regions. The red bars (a) show the performance of the full method with all processing steps.

The blue-colored bars (b-d) show the impact of the simulation in left and right images and of the propagation. The left-right simulation helps in most parts of the image because some false matches in regions near discontinuities are avoided. Due to the voting, both occluded and non-occluded regions benefit from that. The propagation clearly improves the occluded regions by comparing (a) and (d), but may also slightly degrade non-occluded areas because occasionally false matches are diffused into the neighborhood.

The green-colored bars (e-f) show the influence of the a priori surface orientations. For (e) we used only a fronto-parallel prior $\Delta_1 = \{(0,0)^T\}$, for (a) 4 orientations $\Delta_4 = \Delta_1 \cup$

$\{(0,1)^T, (\pm\frac{1}{2},0)^T\}$, and for (f) 8 orientations $\Delta_8 = \Delta_4 \cup \{(\pm\frac{1}{3},0)^T, (\pm\frac{3}{4},0)^T\}$. It is clearly visible that the addition of only a few orientations using Δ_4 results in a huge improvement in performance on these datasets. The negative side-effects of adding more orientations using Δ_8 is surprisingly low, since we would have expected a larger degradation due to a higher matching ambiguity. However, there is nearly no improvement of using 8 orientations at these datasets, mainly because the disparity gradients on surfaces are relative small at these datasets and thus, less orientations suffice. At the dataset *Flowerpots* of Fig. 3 larger gradients occur and in non-occluded regions we measured an error of 7.8%, 6.5% and 4.8% for 1, 4 and 8 orientations respectively. These experiments also provide some evidence for our assumption that random walks usually cover only a small region and thus, perspective distortions have to be considered only for larger disparity gradients.

The orange-colored bars (g-h) display the effect of the voting technique. In this case, the data term was initialized using $E_D(\mathcal{D}) = \sum_{\mathbf{x}} \min_\delta C_A(\mathbf{x}, \mathcal{D}(\mathbf{x}), \delta)$. For (g) we only disabled the voting but left the propagation enabled and for (h) we disabled both voting and propagation. When comparing (g) to (h) it is noticeable that for *Teddy* in non-occluded regions the propagation reduced the quality, which was due to a false match which was propagated into the neighborhood. But also here the picture is clear that the propagation mainly improves in occlusions. The impact of voting is best measured by comparing (h) to (d) because in both cases no propagation is performed. From that, a considerable influence on the quality can be observed in all image regions. This can be explained by the random walks: if a random walk covers occluded and non-occluded pixels, all of them will receive support for the correct depth if the true disparity and orientation is contained in the set S.
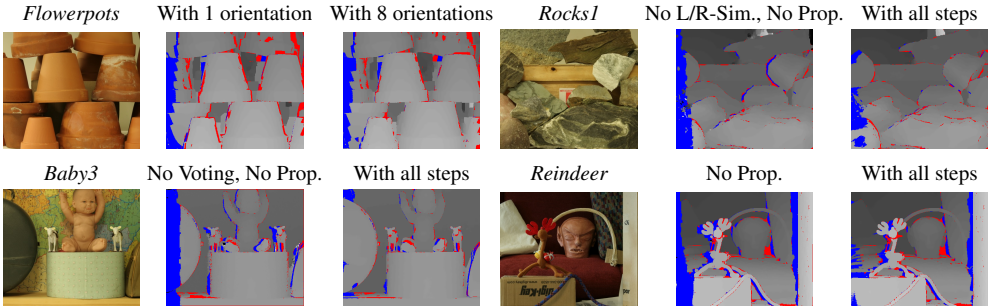

Figure 3: A qualitative comparison to visualize the effect of disabling processing steps of our cost function. For each dataset we show the left image, a disparity map where specific steps of the algorithm were disabled and a disparity map of the full method. Blue and red pixels are wrong disparities in occluded and non-occluded regions respectively (*i. e.* the disparity error is greater than one). We processed *Flowerpots* with one and eight a priori surface orientations. At *Rocks1* we disabled the simulation in left and right images and the propagation. At *Reindeer* we disabled the voting and at *Baby3* the propagation additionally.

In Fig. 3, we give a qualitative impression using difficult Middlebury datasets which underline the previous observations.

3   More Results

In the following we present more disparity maps. For the method of Oh *et al.* [2] we include the results with activated occlusion filling. It is clearly visible that their technique drastically reduces the quality. We also include the results for the method of Woodford *et al.* [3] which fails at some curved objects, especially for *Baby3*. We also show results for our method including our proposed global optimization and including our hole filling technique which we described in the paper (but it should not be compared directly to the other methods).
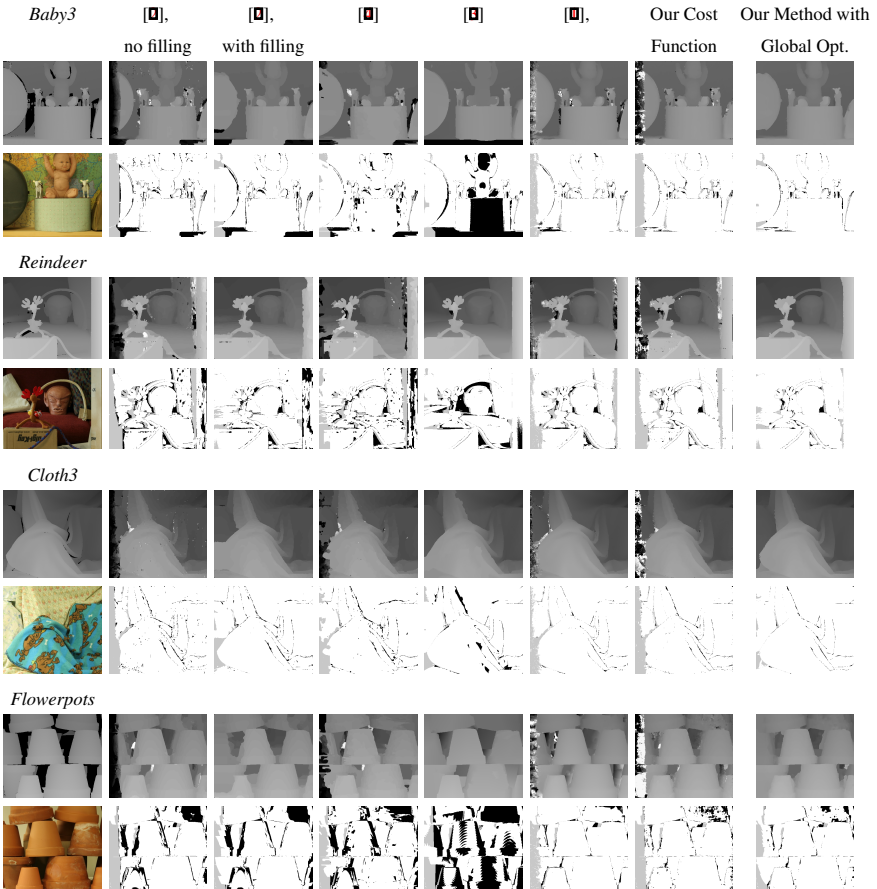

Figure 4: Results of different methods for the datasets *Baby3*, *Reindeer*, *Cloth3* and *Flowerpots*. The first column shows the ground truth disparities and the left images. Then disparity maps and bad pixel images of the methods are presented. The bad pixel images show disparity errors > 1 pixel.
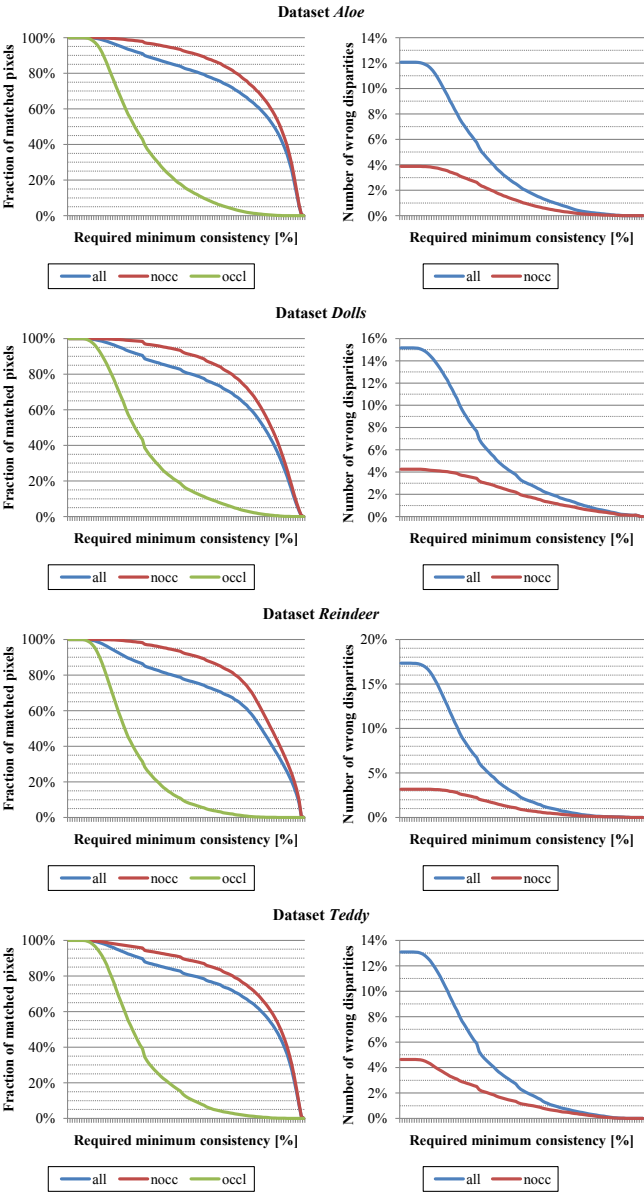
Figure 5: An analysis of the statistical consistency measure at the datasets *Aloe*, *Dolls*, *Reindeer* and *Teddy*. It can be seen that there are no significant differences between different datasets. The left charts show the number of matched pixels over consistency. The fraction of matched pixels is defined as the number of valid disparities (*i. e.* which are considered *consistent*) divided by the total number of pixels of the corresponding region. It can be seen that occluded regions (occl) can be filtered out effectively and that non-occluded regions (nocc) are much less affected by filtering. The right charts show the number of wrong disparities over consistency. The overall error (all) drops quickly, because the number of occluded pixels decreases dramatically.

References

[1] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo - stereo matching with slanted support windows. In *BMVC*, pages 14.1–14.11, 2011.

[2] Changjae Oh, Bumsub Ham, and Kwanghoon Sohn. Probabilistic correspondence matching using random walk with restart. In *BMVC*, pages 1–10, 2012.

[3] Oliver J. Woodford, Philip H. S. Torr, Ian D. Reid, and Andrew W. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. In *CVPR*, 2008.

[4] Qingxiong Yang. A non-local cost aggregation method for stereo matching. In *CVPR*, pages 1402–1409, 2012.