A Stochastic Cost Function for Stereo Vision

Christian Unger
http://campar.in.tum.de/Main/ChristianUnger
BMW Group
Munich, Germany

Slobodan Ilic
http://campar.in.tum.de/Main/SlobodanIlic
Siemens AG
Research & Technology Center
Munich, Germany

Abstract

In this paper, we use random walks to infer information about the disparities in the image. While the random walks help to maintain sharp object boundaries, we introduce a correlation step that handles occlusions and slanted surfaces. The random walks are then used to build a stochastic cost function which serves to identify the most probable disparities. This also delivers valuable information about the reliability of the depths by exploiting their consistency statistically.

Our method delivers very good results by means of local matching, but we also demonstrate that the obtained cost function is well suited for global optimization techniques. Moreover, our consistency maps deliver reliable statistical information about the confidence of disparity maps. In our paper we provide extensive evaluations with challenging images and show that our cost function based on random walks is very useful.

1 Introduction

The goal of this paper is to present a novel stochastic cost function for binocular stereo vision that delivers statistics about the most probable disparities on the pixel level. We drive these statistics by many independent stochastic processes so that robustness to outliers can be achieved. Each of these stochastic processes may be understood as an individual who is requested to deliver his opinion about the depth. Finally, the idea is to fuse all these individual measurements into one global disparity map.

In this paper, we use random walks for this. A random walk randomly traverses the image where at each step of the walk, an adjacent pixel location is chosen based on color similarity. In this sense, a random walk can be viewed as a local segmentation which is assumed to be robust along discontinuities. The set of pixels which is covered by the random walk is used to infer information about the disparity. In order to handle perspective distortions from slanted surfaces we incorporate orientations and occluded regions into the correlation step. The global cost function is finally obtained by a novel voting technique based on these random walks. This serves to identify the most probable disparities by collecting the statistics of all walks. After this voting step, a global optimization may be run or the disparity map may be extracted using a winner-takes-all (WTA) strategy. Finally, confident disparities may be propagated into inconsistent regions. In this paper, we demonstrate that random walks are useful for gathering important statistics about disparities. One strong property of our method is that our cost function is statistically motivated and we show that our proposed statistical

consistency is a powerful and very useful confidence measure with which occlusions may be filtered out effectively.

We provide an extensive evaluation with elaborate details. Our method compares very well with the current state-of-the-art and is capable of achieving top rankings at the Middlebury benchmark. Our results demonstrate that our cost function can deal with notoriously difficult situations like occlusions, discontinuities and slanted surfaces.

2 Related Work

Since stereo vision is a very broad area we focus here only on the most relevant works (an overview on early methods is given in [16]).

The main driving aspect in our paper is the use of random walks to retrieve matching costs. This is in some sense related to the works which use support weights, like [10, 22]. However, in their approaches no slanted surfaces are addressed and therefore known limitations exist in such situations. Moreover, no statistical reasoning is possible in these approaches. The problem of slanted surfaces was later addressed in [4], where [22] is combined with the iterative PatchMatch algorithm. After a random initialization of depths and orientations, the disparity map is iteratively updated by several propagation and refinement steps using relatively large matching windows. The results of this approach are quite good but also no statistical reasoning is provided for this method. For the sake of completeness recent extensions should also be mentioned, where the original idea of [4] is combined with different optimization approaches [1, 8] or with edge-aware filtering based on super pixels [14].

Shen *et al*. [17] use the improved random walks algorithm of Grady [7] to solve a global energy minimization problem in order to compute reliable matches. In a second step, they interpolate matches in ambiguous regions. Their approach is very different from our work because we explicitly simulate random walks and use them as matching primitives. Further, we incorporate surface orientations and we also introduce a novel voting technique based on simulated random walks.

Oh *et al*. [15] used random walks for cost aggregation, but with a completely different principle. In their work, the aggregated matching costs are computed using a global iteration. In every iteration, costs are summed by using previously computed aggregated costs. While this is a relatively efficient structure, our method is very different because for every pixel we perform, in some sense, a local segmentation of the image and use it for cost aggregation and voting. Furthermore, we explicitly address slanted surfaces and perform a statistical reasoning among all walks using our voting technique. In the results we will show that this method has weaknesses at slanted surfaces.

In some works image segmentation has been incorporated into global approaches with excellent results [3, 5, 12, 18, 19, 20]. However, in practice, it is difficult to identify the optimal segmentation parameter set for a broad spectrum of image data.

3 Method

Fig. 1 shows an overview of our method. (a) We first compute pixel-wise matching costs. (b) In our cost aggregation stage, we simulate a random walks for every pixel of the image. Since random walks rarely cross large image gradients, this can be understood as a pre-segmentation step to increase robustness at discontinuities. We also explicitly consider

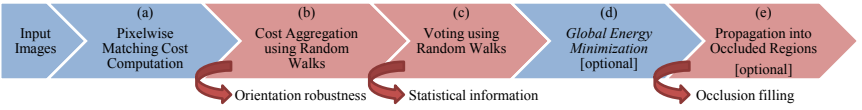


Figure 1: The processing steps of our method.

slanted surfaces by evaluating different surface orientations. (c) Once we computed a cost function for every random walk, we introduce a novel voting technique that fuses information of all random walks into one global voting space. The collected votes contain statistical information about the likelihood of every disparity at every pixel location and also reveal inconsistencies in places where matching is ambiguous. (d) The votes may be used in a global optimization or for direct disparity selection. (e) After this step, random walks may be used to propagate reliable matches into inconsistent regions.

3.1 Basic Definitions

Matching Costs We compute pixel-wise matching costs $\mathcal{C}_M(x,y,d)$ where (x,y) is defined for all possible image locations of the left image $\mathcal{I}_L$ and d iterates over the set of possible disparity values: $d_{\min} \leq d \leq d_{\max}$. $\mathcal{C}_M(x,y,d)$ expresses the dissimilarity of a correspondence between the left image pixel $\mathcal{I}_L(x,y)$ and right image pixel $\mathcal{I}_R(x-d,y)$ using the rectified image pair. In practice, we use the sampling invariant differences of [2].

Random Walks We define a random walk as an ordered sequence of pixel locations $\mathcal{R}(\mathbf{x}_S) = \langle \mathbf{r}_i \rangle_{0 \leq i \leq N}$ starting at $\mathbf{r}_0 := \mathbf{x}_S$ and with a length of N steps. At each step of the walk, *i.e.* $\mathbf{r}_i \mapsto \mathbf{r}_{i+1}$, a new pixel is randomly selected for $\mathbf{r}_{i+1}$ from the four-connected neighbourhood of $\mathbf{r}_i$ based on *transition probabilities* $p_T(\mathbf{r}_{i+1} \,|\, \mathbf{r}_i)$. It is common practice to assume that depth discontinuities coincide with large color gradients and therefore we define the transition probability p_T as a function of the color similarity:

$$p_T(\mathbf{r}_{i+1} \,|\, \mathbf{r}_i) = \frac{1}{C(\mathbf{r}_i)} \cdot \exp\left(-\frac{d_C(\mathcal{I}(\mathbf{r}_i) - \mathcal{I}(\mathbf{r}_{i+1} + (\mathbf{r}_{i+1} - \mathbf{r}_i)))}{\sigma_C} \right) \tag{1}$$

where $\mathbf{r}_{i+1}$ is a pixel from the four-connected neighborhood $\mathcal{N}(\mathbf{r}_i)$ of $\mathbf{r}_i$. The value $C(\mathbf{r}_i)$ is a normalization factor, such that the transition probabilities sum up to one for every pixel. The function d_C computes the color norm and in practice we use $d_C(\mathbf{c}) = \sqrt{\mathbf{c}^T \mathbf{c}}$, where the vector $\mathbf{c}$ contains the differences of the individual color channels. The parameter σ_C controls how likely it is that a random walk steps towards a pixel with a higher color difference. Further, in (1) we used a step size of 2 pixels for the probability computation because along depth discontinuities, there is a 1 pixel wide region of interpolated, unreliable color values, which is a result of pixel sampling. By using a step size of 2 for probability computation, the risk is reduced that a random walk crosses an object boundary.

3.2 Left-Right Simulation

In this section, we take a closer look at the behavior of random walks near occlusions. Including occluded pixels during cost aggregation is highly problematic, because pixels with

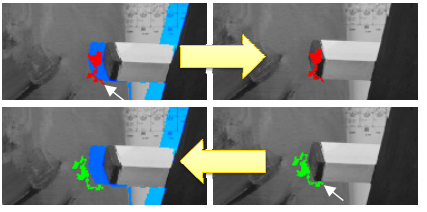


Figure 2: Left-right simulation of random walks for the image pair *Teddy*. White arrows indicate the pixels at which random walks were started. Occluded pixels are painted in blue. In the top left image the random walk was started at a non-occluded pixel, but crosses the area of occluded blue pixels. In the bottom right a random walk does not step into an occluded area because no occlusions are present there in the right image.

no physical relation to the other image would be correlated with completely wrong surfaces and the cost values would not be reliable. In Fig. 2, we show a magnified region of the stereo image pair *Teddy* from Middlebury [16], where random walks are simulated in both left and right images starting at the pixel labeled with a white arrow. In the upper left image a *left random walk* painted in red was simulated. In the lower right image a *right random walk* painted in green was simulated. The left random walk crosses occluded pixels which are painted in blue. This does not happen for the right random walk because no occluded pixels are present in the vicinity in the right image. In general, both the left and right images contain occluded pixels, however at different locations in the images. Therefore, we simulate random walks for the left and right images independently and denote them as $\mathcal{R}^L$ and $\mathcal{R}^R$ respectively.

3.3 Cost Aggregation using Multiple Surface Orientations

One of the big challenges for stereo methods is the treatment of slanted surfaces and in the following we describe how we tackle this problem. First, we assume that the surface shape can be linearly approximated for the region covered by a random walk. While this assumption might be violated for some walks that traverse a large image region, we argue that the failure of some walks is negligible for the final result, due to our voting technique presented in the next section. Additionally, random walks are known to stay asymptotically at the same place, so the majority of pixels is usually found near the start pixel. Hence, the idea is that only a few surface orientations should be sufficient, because for small image patches only a discrete, finite number of possible perspective distortions can happen between a given, small number of pixels. In the worst case, the aggregated cost will have a reduced performance similar to methods that perform disparity estimation using a fronto-parallel prior. Our experiments clearly show that our approach can handle very difficult geometries.

To compute the aggregated cost $\mathcal{C}_A(\mathbf{x}, d, \delta_k)$ for a given pixel $\mathbf{x}$ and disparity d, we use the random walks defined in the left and right image and take different a priori surface orientations $\delta_k \in \Delta$ into account. To perform left to right correlation, we use the random walk $\mathcal{R}^L(\mathbf{x}) = \langle \mathbf{r}_i \rangle_{0 \leq i \leq N}$ and sum the pixelwise costs $\mathcal{C}_M$ along the walk:

$$\mathcal{C}_A^L(\mathbf{x}, d, \delta_k) = \sum_{i=0}^{N} \mathcal{C}_M(\mathbf{r}_i, d + \delta_k^T(\mathbf{r}_i - \mathbf{r}_0)) \tag{2}$$

The expression $d + \delta_k^T(\mathbf{r}_i - \mathbf{r}_0)$ in (2) adapts the disparity value to the given surface orientation δ_k, which is defined as the disparity gradient in horizontal and vertical direction. Note that in this formulation some pixels may occur multiple times in the correlation sum.

For right to left correlation we simulate a random walk at each pixel in the right image. In this case, the random walk is defined as: $\mathcal{R}^R(\mathbf{x} - (d,0)^T) = \langle \mathbf{r}_i \rangle_{0 \le i \le N}$. Using this random walk, we compute the cost as

$$\mathcal{C}_A^R(\mathbf{x}, d, \delta_k) = \sum_{i=0}^{N} \mathcal{C}_M(\mathbf{r}_i + (d,0)^T, d + \delta_k^T(\mathbf{r}_i - \mathbf{r}_0)) \tag{3}$$

Similar to $\mathcal{C}_M$, $\mathcal{C}_A^R(\mathbf{x}, d, \delta_k)$ is defined for *left image* pixels to make the reasoning on cost values easier. Therefore, we have to transform right image positions $\mathbf{r}_i$ of the random walk $\mathcal{R}^R$ into left image positions using $\mathbf{r}_i + (d,0)^T$.

Finally, we obtain a global cost volume which assigns a dissimilarity score to every disparity and orientation given a pixel location $\mathbf{x}$ of the left image using $\mathcal{C}_A(\mathbf{x}, d, \delta_k) = \min(\mathcal{C}_A^R(\mathbf{x}, d, \delta_k), \mathcal{C}_A^L(\mathbf{x}, d, \delta_k))$. Here, the basic idea is that the aggregated cost of a random walk is usually higher if the walk covers occluded pixels. In the next section, we analyze the cost volume $\mathcal{C}_A$ to determine depths and orientations for every random walk of the left image.

3.4 Voting using Random Walks

In this step, we transform the volume of aggregated costs $\mathcal{C}_A$ into a *voting space* $\mathcal{V}$ using a novel technique based again on random walks. The main intuition consists of two simple observations. Firstly, the minima of the aggregated costs $\mathcal{C}_A$ provide depth and orientation hypotheses for *all* pixel locations along a random walk. Secondly, every pixel is covered by many different random walks. Therefore, for every pixel $\mathbf{x}$, there are multiple depth and orientation hypotheses that are contributed by different random walks which start in the vicinity of $\mathbf{x}$ and which cover $\mathbf{x}$. By collecting this information in the voting space at every pixel we obtain statistics about the disparities and ideally, the true disparity will receive many votes from the random walks crossing the pixel.

The voting space $\mathcal{V}(\mathbf{x}, d)$ can be best understood as a data structure that holds a histogram for disparity values at every pixel location and is initialized to zero. First, for a given pixel $\mathbf{x}$ of the left image, we collect a set S of relevant depth and orientation hypotheses:

$$S = \{(d, \delta) \,|\, \mathcal{C}_A(\mathbf{x}, d, \delta) \le \hat{s} + N\Theta, d \in D, \delta \in \Delta\} \tag{4}$$

with $\hat{s} = \min_{d,\delta} \mathcal{C}_A(\mathbf{x}, d, \delta)$ and $D = [d_{\min}, d_{\max}]$. The parameter Θ specifies a small corridor within which the hypotheses may be located relative to the minimum $\hat{s}$, and N is the length of the random walks.

The second step is to update the voting space with a random walk starting at pixel $\mathbf{x}$ based on simulations using the hypotheses in S. For every hypothesis $(d, \delta) \in S$ for the start pixel $\mathbf{x}$, we update every pixel of the random walk $\mathcal{R}(\mathbf{x}) = \langle \mathbf{r}_i \rangle_i$ by simulating the random walk again with the given depth and orientation prior using: $\mathcal{V}(\mathbf{r}_i, d_i) \mapsto \mathcal{V}(\mathbf{r}_i, d_i) + 1$ with $d_i = d + \delta^T(\mathbf{r}_i - \mathbf{x})$. Note that for a given random walk, $\mathcal{V}(\mathbf{r}_i, d_i)$ is updated only once per pixel (different to $\mathcal{C}_A$), even if a pixel occurs multiple times in the walk.

3.5 Statistical Consistency

By normalizing the values in $\mathcal{V}$ for every pixel we obtain a probability for the disparities at every pixel and allows a statistically motivated consistency measure. For that, we define the *consistency* at pixel $\mathbf{x}$ as: $\frac{\mathcal{V}(\mathbf{x},\hat{d})}{1+\sum_i \mathcal{V}(\mathbf{x},i)}$, where $\hat{d}$ is the disparity with most votes at pixel $\mathbf{x}$. Please note that by defining the consistency in this way, we assume that there is always virtually one random walk that is wrong and the value will never be 1.0. By defining it this way the consistency depends not only on a few (or even only one) random walks. So to achieve a very high consistency value it is rather required that many random walks confirm to the same specific disparity. In this sense, the consistency will asymptotically reach 1 if more and more random walks confirm the same disparity for a specific pixel. Later, we will analyse this statistical measure in more detail.

3.6 Global Optimization and Propagation into Occluded Regions

Until now we presented a novel cost function which delivers statistics about the likelihood of every disparity. From these the most probable disparities may be extracted directly using WTA but in many cases mismatches can be reduced by enforcing a smoothness constraint. We decided for a recent variational technique [13] based on the second-order total generalized variation [6] because of its strengths at slanted surfaces and its practicability. For the optimization, we use our voting space in their energy functional:

$$\min_{\mathbf{u},\mathbf{v}} \int_\Omega \lambda_s |G(\nabla \mathbf{u} - \mathbf{v})| + \lambda_a |\nabla \mathbf{v}| + \lambda_d \tilde{\mathcal{V}}(\mathbf{x}, \mathbf{u}(\mathbf{x})) d\mathbf{x} \tag{5}$$

with $\mathbf{u} \in D^{|\mathcal{I}|}$ the smoothed disparity map to solve for, an auxiliary vector field $\mathbf{v}$, weighting factors λ_s, λ_a, λ_d and G the anisotropic diffusion tensor, a linear operator that adapts and directs the amount of regularization locally. The data term $\tilde{\mathcal{V}}(\mathbf{x}, \mathbf{u}(\mathbf{x}))$ is based on the voting space $\mathcal{V}$ and is defined as $\tilde{\mathcal{V}}(\mathbf{x}, \mathbf{u}(\mathbf{x})) = \frac{\mathcal{V}_{max} - \mathcal{V}(\mathbf{x}, \mathbf{u}(\mathbf{x}))}{\mathcal{V}_{max}}$.

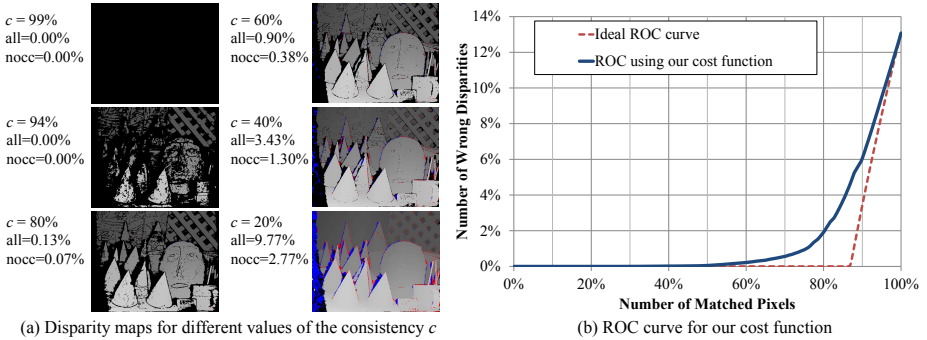
In occluded areas it is impossible to obtain depth information using binocular correlation. The common strategy is to identify occluded pixels using the well known left-right consistency check [9]. In our work, if no global optimization is used, implausible disparity matches may be identified using our statistical consistency measure. In practice we determine for all those pixels a disparity for which the consistency is greater than a specific threshold and leave a hole for all other pixels. Then we propagate depth information into the holes by again using random walks using the framework proposed by Grady [7].

4 Experiments

We use the Middlebury benchmark [16] to compare our method to related works. We used constant parameters for all image pairs[1]. It must be pointed out that the memory consumption is relatively large when naively implemented. In practice, if an approximation is tolerated, the memory consumption of $\mathcal{C}_A$ can be dramatically reduced by only storing the minimum among the orientations (*i.e.* storing $\mathcal{C}_A(\mathbf{x}, d)$ instead of $\mathcal{C}_A(\mathbf{x}, d, \delta_k)$). It could even be further reduced by just maintaining the set S on a per pixel basis. Using parallelization (using OpenMP on the scanline level) we achieve on the CPU running times that are similar to many other state-of-the-art methods (20s for a Middlebury pair on a dual Intel X5690).

[1] With $\Delta = \{(0,0), (\pm\frac{1}{3}, 0), (\pm\frac{1}{2}, 0), (0, \pm\frac{1}{3}), (0, \frac{1}{2}), (0,1)\}$, $N = 200$, $\sigma_C = 17.7$

4.1 Analysis of the Statistical Consistency Measure



(a) Disparity maps for different values of the consistency c

(b) ROC curve for our cost function

Figure 3: Disparity maps of different values of the consistency c and the ROC curve for our method computed using the definitions in [].

First, we analyse our novel statistical consistency measure that serves as a confidence for every disparity value. In order to evaluate the influence of the consistency filtering on disparity density and matching quality separately, we compute the disparity errors only for the portion of estimated pixels. We define the disparity density (or fraction of matched pixels) as the number of estimated disparities (whose consistency is greater than the consistency threshold) divided by the total number of pixels of the corresponding region.

In Fig. 3a we show a comparison of disparity maps of the dataset *Cones* that were generated using different values for the consistency. One key result is that consistency values towards 100% lead to highly reliable disparity maps. While for a consistency of 94% the disparity density was relatively small with 29%, it is on the other hand impressive that not a single pixel was wrongly estimated. In general, it can be seen clearly that occluded pixels can be filtered out very well. The overall picture for the occlusion filtering also holds similarly for many datasets and more details can be found in the supplemental material.

In Fig. 3b we show for the dataset *Teddy* the ROC curve of our cost function using our confidence measure and we use the same definitions as in []. It can be seen that the envelope of the graph is relatively close to the ideal ROC curve, which indicates that our cost function in combination with the statistical confidence is quite reliable. In [] more metrics are described such as the *Improvement* over random selection. While the best value reported in the paper lies at 0.630 (with a theoretical optimum of 1.0), our method achieves a value of 0.886. Moreover, the best reported *Performance* relative to the optimal confidence was reported as 0.331 (with an optimum of 1.0) while our method achieves a value of 0.601. Certainly, the values are not directly comparable, because the underlying cost function plays an important role. However, being clearly better than the best reported numbers in [] is a strong argument for the reliability of both our stochastic cost function in combination with our confidence.

4.2 Analysis of the Stochastic Cost Function

Now we analyze the strengths and weaknesses our cost function (sections 3.1-3.4) and compare to other related methods. So in particular we compare our cost function with a winner-

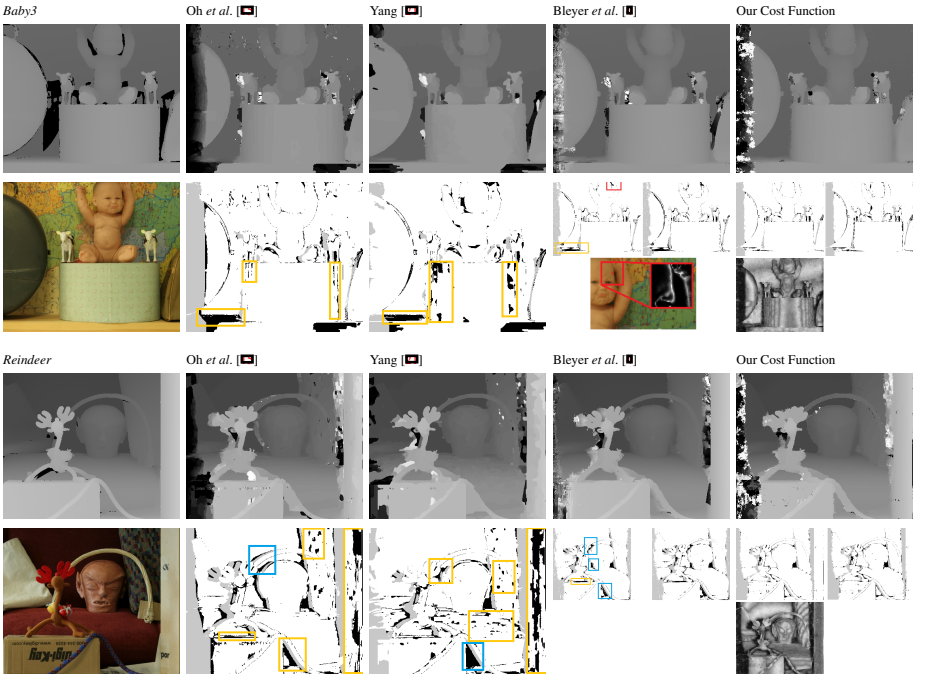


Figure 4: Results for the datasets *Baby3* and *Reindeer*. *Our cost function* refers to the steps described in sections 3.1-3.4. The first column shows ground truth disparities and left images. The first and third row show the disparity maps. The second and fourth rows show the disparity errors: at every method for error threshold 1.0; for [4] and our method we also include error threshold 0.5 (left is 1.0; right is 0.5). For our method we also include the confidence maps. The rectangles indicate errors: red: similarly colored objects; yellow: slanted surfaces; blue: inter-object occlusions.

takes-all strategy to [4, 15, 21]. For [15] we obtained disparity maps directly from the authors and for [21] we used the publicly available implementation. For the comparison between the cost functions we used disparity maps without hole-filling to make the comparison independent of different hole-filling approaches. In Fig. 4 we show disparity maps for *Baby3* and *Reindeer* and we highlight different challenging regions with differently colored rectangles. Blue rectangles for inter-object occlusions, yellow for slanted surfaces and red for similarly colored objects.

The method of Oh *et al.* [15] captures surface curvature well for small disparity gradients but errors appear at steeply slanted surfaces. A clear limitation appears also near discontinuities which can be directly seen at *Reindeer*. Tab. 1 confirms that in difficult images more errors appear on object boundaries.

We measured the non-local filter of Yang [21] worst in non-occluded regions among the methods in Tab. 1, which is obviously to due many errors at slanted surfaces. Also visually the artifacts can be clearly seen in Fig. 4 when looking at the yellow rectangles.

In Fig. 4 we also show results for the method of Bleyer *et al.* [4] for which we performed

Table 1: The performance of different methods at the Middlebury benchmark [16]. We tested our method with and without global optimization (GO).

Algorithm	Error Thresh.	Avg. Error	Tsukuba nocc	Tsukuba disc	Venus nocc	Venus disc	Teddy nocc	Teddy disc	Cones nocc	Cones disc	Baby3 nocc	Baby3 disc	Reindeer nocc	Reindeer disc
Our Method, with GO	0.5	8.77	8.68	**17.4**	**0.96**	8.98	5.98	15.0	3.81	**9.94**	3.61	16.1	**2.92**	**11.8**
Our Method, no GO	0.5	8.57	**8.64**	17.6	2.10	9.91	5.96	**14.9**	4.32	11.6	**3.58**	**16.1**	4.12	14.2
Bleyer *et al.* [4]	0.5	9.91	15.0	20.3	1.00	**7.75**	**5.66**	16.5	**3.80**	10.2	4.18	16.6	4.73	14.8
Bleyer *et al.* [4]	1.0	4.59	2.09	9.31	0.21	2.62	2.99	9.62	2.47	7.11	2.90	**10.8**	3.95	12.3
Our Method, with GO	1.0	4.61	2.17	10.0	0.27	3.49	**2.75**	**8.31**	**2.18**	**6.50**	**2.34**	10.9	**2.33**	**10.1**
Our Method, no GO	1.0	4.90	2.14	10.1	0.45	3.75	3.16	8.68	2.54	7.49	2.42	11.0	**3.19**	11.2
Yang [22]	1.0	5.48	**1.47**	7.88	0.25	2.60	6.01	14.3	2.87	8.10	7.67	15.1	12.3	19.4
Oh *et al.* [17]	1.0	5.38	1.60	**6.44**	**0.20**	**2.51**	6.15	15.8	2.60	7.48	7.23	14.0	11.2	17.0

10 iterations to achieve a better quality[2]. Qualitatively and also quantiatively in Tab. 1, our method compares very well with [4] and this further underlines the potential of random walks for stereo matching. We also want to point out one detail in our comparison indicated by a red rectangle together with an illustration of the support weights. At this place, different similarly colored objects are present within the matching window and even though an intensity gradient exists, high support weights are assigned to different objects. Random walks are more robust than the support weights of [22] in this case, because a random walk is required to consist of *spatially connected*, similarly colored pixels.

We also would like to point out that the occlusions are clearly visible as dark "shadows" in our confidence maps which qualitatively underlines the reliability of our confidence measure. In practice our results are relatively independent of the randomization. To assess this, we ran our method 100 times on the same stereo pair with constant parameters and different random seeds. The standard deviation of the overall disparity error ranged from 0.005% up to 0.05%. Finally, we present results on standard image pairs and results with global optimization in Fig. 5 and Tab.1. The optimization removes small matching errors and improves smoothness. On the Middlebury benchmark we currently achieve the 8$^{\text{th}}$ place for the error threshold of 0.5 which is another evidence that our method compares well and that our cost function can be useful in other stereo methods.

So as a bottomline, our proposed stochastic cost function performs quite well at the recovery of complex scene geometry which is especially visible at difficult inter-object occlusions. Some weaknesses appear in homogeneous regions (which can be addressed with global optimization) and at field of view occlusions (which can be addressed with hole-filling). For more experiments, also on method parameters and method steps, we would like to refer the reader to the supplemental material.

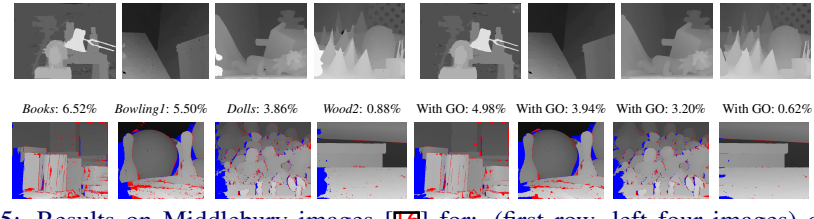


Figure 5: Results on Middlebury images [16] for: (first row, left four images) our cost function with propagation and (first row, right four images) our cost function with global optimization. In the second row we show disparity maps of our method with and without global optimization for other images with percentages of disparity error > 1 in non-occluded regions. Red pixels indicate errors in non-occluded regions; blue pixels in occluded regions.

[2] In [4] three iterations were used with 35x35 matching windows.

5 Conclusion

In this paper, we proposed a novel stochastic cost function based on random walks which enables statistical reasoning on the discovered depth measurements. In particular, we introduced (1) a cost aggregation technique based on random walks which is orientation- and occlusion-robust, (2) a novel voting technique based on random walks to obtain statistical information about the disparity likelihood and (3) a strong novel statistical consistency measure. Finally, in the experiments we show impressive results on challenging stereo images. Given the obtained results we believe that our cost function together with the confidence is useful for other stereo methods and is valuable in practical applications.

Acknowledgements

We would like to express our thanks to Georg Kuschk and Daniel Cremers for fruitful discussions and for performing many experiments with their method using our cost spaces. We also want to thank Changjae Oh and Philipp Heise for providing disparity maps of their methods. This work is supported by the BMW Group.

References

[1] Frederic Besse, Carsten Rother, Andrew W. Fitzgibbon, and Jan Kautz. Pmbp: Patchmatch belief propagation for correspondence field estimation. In *BMVC*, pages 1–11, 2012.

[2] Stan Birchfield and Carlo Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *PAMI*, 20(4):401–406, 1998.

[3] Michael Bleyer, Carsten Rother, and Pushmeet Kohli. Surface stereo with soft segmentation. In *CVPR*, 2010.

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

[5] Michael Bleyer, Carsten Rother, Pushmeet Kohli, Daniel Scharstein, and Sudipta Sinha. Object stereo: Joint stereo matching and object segmentation. In *CVPR*, 2011.

[6] K. Bredies, K. Kunisch, and T. Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[7] Leo Grady. Random walks for image segmentation. *PAMI*, 28(11):1768–1783, 2006.

[8] Philipp Heise, Sebastian Klose, Brian Jensen, and Alois Knoll. Pm-huber: Patchmatch with huber regularization for stereo matching. In *ICCV*, pages 2360–2367, 2013.

[9] Heiko Hirschmüller, Peter R. Innocent, and Jon Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *IJCV*, 47(1-3):229–246, 2002.

[10] Asmaa Hosni, Michael Bleyer, Margit Gelautz, and Christoph Rhemann. Local stereo matching using geodesic support weights. In *ICIP*, 2009.

[11] Xiaoyan Hu and Philippos Mordohai. A quantitative evaluation of confidence measures for stereo vision. *PAMI*, 34(11):2121–2133, 2012.

[12] Andreas Klaus, Mario Sormann, and Konrad Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *ICPR*, pages 15–18, 2006.

[13] G. Kuschk and D. Cremers. Fast and accurate large-scale stereo reconstruction using variational methods. In *ICCV Workshop on Big Data in 3D Computer Vision*, 2013.

[14] Jiangbo Lu, Hongsheng Yang, Dongbo Min, and Minh N. Do. Patch match filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation. In *CVPR*, pages 1854–1861, 2013.

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

[16] Daniel Scharstein, Richard Szeliski, and Ramin Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2002.

[17] Rui Shen, Irene Cheng, Xiaobo Li, and Anup Basu. Stereo matching using random walks. In *ICPR*, pages 1–4. IEEE, 2008.

[18] Y. Taguchi, B. Wilburn, and C. L. Zitnick. Stereo reconstruction with mixed pixels using adaptive over-segmentation. In *CVPR*, 2008.

[19] 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.

[20] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister. Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling. *PAMI*, 31 (3):492–504, 2009.

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

[22] Kuk-Jin Yoon and In-So Kweon. Adaptive support-weight approach for correspondence search. *PAMI*, 28(4):650–656, 2006.