As-Rigid-As-Possible Stereo under Second Order Smoothness Priors

Chi Zhang[1], Zhiwei Li[2], Rui Cai[2], Hongyang Chao[1], and Yong Rui[2]

[1] Sun Yat-Sen University, China
[2] Microsoft Research Asia, China

Abstract. Imposing smoothness priors is a key idea of the top-ranked global stereo models. Recent progresses demonstrated the power of second order priors which are usually defined by either explicitly considering three-pixel neighborhoods, or implicitly using a so-called 3D-label for each pixel. In contrast to the traditional first-order priors which only prefer fronto-parallel surfaces, second-order priors encourage arbitrary collinear structures. However, we still can find defective regions in matching results even under such powerful priors, e.g., large textureless regions. One reason is that most of the stereo models are non-convex, where pixel-wise smoothness priors, i.e., *local constraints*, are too flexible to prevent the solution from trapping in bad local minimums. On the other hand, *long-range spatial constraints*, especially the segment-based priors, have advantages on this problem. However, segment-based priors are too rigid to handle curved surfaces. We present a mixture model to combine the benefits of these two kinds of priors, whose energy function consists of two terms 1) a Laplacian operator on the disparity map which imposes pixel-wise second-order smoothness; 2) a segment-wise matching cost as a function of quadratic surface, which encourages "as-rigid-as-possible" smoothness. To effectively solve the problem, we introduce an intermediate term to decouple the two subenergies, which enables an alternated optimization algorithm that is about an order of magnitude faster than PatchMatch [1]. Our approach is one of the top ranked models on the Middlebury benchmark at sub-pixel accuracy.

1 Introduction

Stereo correspondence is a core problem in computer vision. Following the terminology of [2], existing approaches are classified as local or global methods. For most of the top-ranked global methods, a clear clue is that they usually impose smoothness priors on disparity, e.g., first order priors [2][3][4], second-order priors [1][5] and segment-based priors [6][7][8][9]. Each kind of priors has special advantages as well as limitations. The goal of this paper is to develop a new global stereo model to combine advantages of second-order and segment-based priors.

1.1 Background

For a comprehensive discussion on dense two-frame stereo matching, we refer readers to the survey conducted by Scharstein and Szelisky [2]. In this paper, we

D. Fleet et al. (Eds.): ECCV 2014, Part II, LNCS 8690, pp. 112–126, 2014.

only discuss two categories of works related to smoothness priors, i.e., second-order and segment-based priors.

In contrast to the widely used first order prior defined on two-pixel neighborhoods p and q, $S(q, \{p\}) = \|D(p) - D(q)\|_1$, which increases monotonically as the neighborhood diverges from fronto-parallel, second-order priors defined on three-pixel neighborhoods, $S(q, \{p, r\}) = \|D(p) - 2D(q) + D(r)\|_1$, increases monotonically as the neighborhood diverges from collinearity. Using second-order priors in stereo matching has a long history since the early 1980s by Grimoson [10] and Terzopoulos [11]. It has been extended to piecewise second-order by Blake and Zisserman [12]. [13] argued that it is closer to the human visual system than first-order priors.

However, effectively applying second-order priors in stereo matching is not easy. Woodford et al. [14] imposed it by considering triple-cliques, which causes the energy function to be non-submodular and the corresponding pairwise graph representation is much more complex than that of the first-order priors. The QPBO-based fusion move algorithm was adopted to minimize the energy by considering many pre-computed disparity proposals [15], which makes the approach complicated and slow.

Assigning each pixel to a tangent plane specified by three parameters becomes a recent trend of sub-pixel accurate stereo models, which is sometimes termed as 3D-label stereo [16]. Olsson et al. have proved 3D-label is an implicit way to impose second-order priors on scalar disparities and leads to submodular pairwise potentials for planar proposals [16]. The PatchMatch work [1] demonstrated that it is an effective way to handle slanted planes thanks to its pixel-wise plane induced matching cost. Although PatchMatch is a "local" method, it has shown its ability to handle very challenge cases. [5][17][16] further developed PatchMatch into global models by adding explicit smoothness terms to regularize the local neighborhoods of 3D labels, i.e., equations of disparity planes. Despite of good results achieved, the necessity of parameterizing the scalar disparity by 3D label is sometimes arguable. Regularization on 3D labels significantly increases the computational cost.

Under an MRF-like energy minimization framework, smoothness priors are regularizers to prevent the solution from over-fitting to local minimums of matching cost volume [2]. However, both first and second-order priors are defined pixel-wise, that is, they are *local constraints*. Considering the fact that most of stereo models are non-convex, local constraints are usually too "weak" to prevent the solution from trapping in bad local minimums. On the contrary, *long-range spatial constraints*, especially the segment-based priors, have advantages on this problem by imposing the smoothness constraints in a rigid form [6][7][8][9]. A basic assumption is that disparity values vary smoothly in homogeneous regions, i.e., segments. And depth discontinuities are only expected to occur on region boundaries. A disparity plane is supposed to be sufficient to model disparity changes in a segment.

In segment-based methods, a disparity plane specified by three parameters $[c_1, c_2, c_3]^\top$, is usually individually fitted to each region. Given a disparity plane,

the disparity of a pixel $[x, y]^\top$ is compute by $d = c_1 x + c_2 y + c_3$. These methods usually estimate many planes from a set of over-segmented regions according to some robust initial matches, and then determine the best matched plane for each region according to an aggregated matching cost defined inside the segment [6]. Robust initial matches are usually obtained by left-right consistency check.

Despite of the excellent accuracy and the relatively low computational cost, segment-based approaches are often challenged by curved surfaces. The rigidness is a double-edged sword. On one side it is very helpful to prevent over-fitting, while on the other side it leads to inaccurate results. Besides, a significant advantage of segment-based approaches is that they usually have much smaller number of parameters than pixel-wise 3D-label approaches.

In summary, both pixel-wise second-order priors and segment-based priors are very effective for stereo matching. Although both of them have limitations, their advantages are complementary for each other. If they can be considered in a single model, better results are very likely to be achieved. Our approach, to the best of our knowledge, is the first one to combine second-order and segment-based priors in a unified framework.

1.2 Contribution

In this paper we propose a novel segment-based global model, in which scalar disparities and parameterized disparity surfaces are jointly modeled. Our major contributions are three-folds:

First, we demonstrate that the second-order smoothness priors can be imposed on disparities by a well-defined Laplacian operator constructed on 4- or 8-pixel neighborhoods guided by color cues. Thus, it is able to preserve sharp discontinuities on edges as well as to encourage collinearity in smooth regions.

Second, unlike previous works where each segment is modeled by a disparity plane, we model a segment by a quadratic surface. Only with two additional parameters, we demonstrate that our approach is able to handle curved surfaces, as well as keeping the advantages of rigidness brought by segments.

Third, we propose an alternated optimization approach for the problem, which is about an order of magnitude faster than PatchMatch [1], making the proposed stereo model very practical.

The paper is organized as following. We present the general framework in Section 2, and a detailed implementation in Section 3. Experiments on benchmarks are given in Section 4. Finally, we conclude the paper in Section 5.

2 As-Rigid-As-Possible Stereo

Given a rectified stereo image pair $\{I_L, I_R\}$, the goal of stereo matching is to estimate the disparity map $\mathbf{u}$ for the reference view I_L. Most global stereo methods can be formulated in an energy minimization framework

$$E(\mathbf{u}) = E_S(\mathbf{u}) + E_D(\mathbf{u}) \tag{1}$$

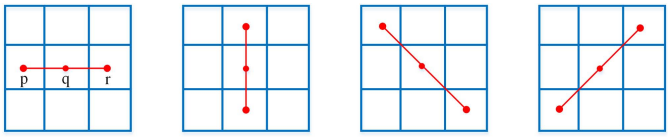


Fig. 1. Four types of Laplacian operators. From left to right are L_1, L_2, L_3 and L_4.

where E_D denotes the matching cost, and E_S denotes the smoothness constraints imposed on a local neighborhood of each pixel or segment. Different stereo models may use different terms or optimization algorithms. We present our proposal in the following sections.

2.1 Second-Order Smoothness Priors

Second-order smoothness priors are defined on three-pixel neighborhoods

$$S(q, \{p, r\}) = \|D(p) - 2D(q) + D(r)\|_1 \tag{2}$$

where p, q, r are the three pixels of a 3×1 patch, and $D(\cdot)$ is a disparity functional. We usually can define four types of 3×1 patches as shown in Fig. 1.

To avoid penalizing disparity discontinuities aligned with intensive edges, we add a weight for each 3×1 patch

$$w(q, \{p, r\}) = \exp\left(-\frac{\|c(p) - 2c(q) + c(r)\|_1}{\gamma}\right) \tag{3}$$

$$S(q, \{p, r\}) = \|w(q, \{p, r\})D(p) - 2w(q, \{p, r\})D(q) + w(q, \{p, r\})D(r)\|_1 \tag{4}$$

where $\|.\|_1$ is $L1$ normal of a vector, γ is a parameter controlling the significance of an edge, and $c(p)$ is the color vector of pixel p. To simplify notations, we will denote them by $w(q)$ and $S(q)$ at the rest of the paper. If the 3×1 patch crosses an edge between two regions, the corresponding weight $w(q)$ is likely to be zero, as well as $S(q)$. Therefore, this weighting scheme does not punish discontinuity of disparities on edges. Indeed $c(p) - 2c(q) + c(r)$ is the second derivative of the image. It is easy to verify that if the disparity function is linear, i.e., $D([x, y]^\top) = ax + by + c$, the smoothness term will be zero. Fox example, if $q = [x, y]^\top, p = [x - 1, y]^\top, r = [x + 1, y]^\top$, then

$$S(q) = \|w(q)(a(x-1)+by+c)-2w(q)(ax+by+c)+w(q)(a(x+1)+by+c)\|_1 = 0 \tag{5}$$

Rewriting Eq. (4) in matrix form, the smoothness energy can be expressed as

$$E_S(\mathbf{u}) = \mathbf{u}^\top L^\top L\mathbf{u} \tag{6}$$

where $L = D - W$ is the Laplacian matrix [18], W is the weight matrix with $w_{qp} = w_{qr} = w(q, \{p, r\})$, and D is a diagonal matrix with $d_{qq} = \sum_i w_{qi}$.

By considering smoothness along multiple directions , we get several Laplacian matrices L_i. The overall problem is a sum of individual quadratic problem. Each directional Laplacian matrix imposes smoothness along one corresponding direction, and does not punish disparity discontinuous crossing intensity edges in any directions.

2.2 As-Rigid-As-Possible Smoothness and Data Cost

Segment-based smoothness prior is an implicit rigid regularization on disparities, which has been demonstrated to be very robust even in hard cases, e.g., autonomous driving stereo estimation [19]. Compared with pixel-wise smoothness priors, it has two major benefits 1) it is more robust to bad local minimum, and 2) the number of parameters is significantly reduced, which usually leads to faster models.

In order to exploit such advantages while avoiding over-rigidness, we reparameterize the disparity map $\mathbf{u}$ by a set of quadratic disparity surfaces that are fitted to each segment. In other word, we try to be "as-rigid-as-possible" (thus the name of the paper) by keeping using the segment-based scheme, while circumventing over-rigidness by fitting quadratic surfaces instead of planes. Specifically, given a segment list S of the input view, we define the segment-based data cost term, which has "built-in" as-rigid-as-possible smoothness, as

$$E_D(\mathbf{\Lambda}) = \sum_{\mathbf{s}} \sum_{p \in \mathbf{s}} \rho \left(\begin{bmatrix} p_x \\ p_y \end{bmatrix}, \begin{bmatrix} p_x - D_{\mathbf{s}}(p - \bar{p}_{\mathbf{s}}) \\ p_y \end{bmatrix} \right) \tag{7}$$

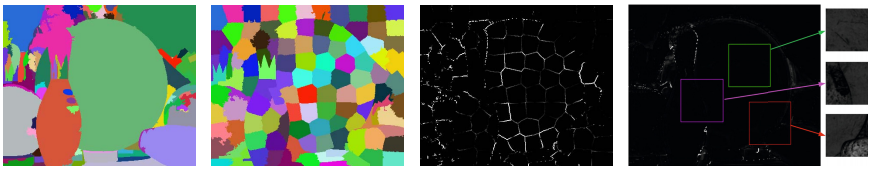
where $\mathbf{s}$ is a segment, p is a pixel in $\mathbf{s}$, $\bar{p}_{\mathbf{s}}$ is the barycenter of $\mathbf{s}$, and $D_{\mathbf{s}}([x, y]^\top) = dx^2 + ey^2 + ax + by + c$ is the disparity functional, $\mathbf{\Lambda}$ is a $|S| \times 5$ matrix representing the quadratic coefficients for all the $|S|$ segments, and $\rho(p, q)$ is the commonly used matching cost [1]

$$\rho(p, q) = (1 - \alpha) \min(\|I_1(p) - I_2(q)\|_1, \tau_{col}) + \alpha \min(\|\nabla I_1(p) - \nabla I_2(q)\|_1, \tau_{grad}) \tag{8}$$

where parameters τ_{col} and τ_{grad} truncate costs for robustness in occlusion regions. For clarity, we use $\rho(p, p - D_{\mathbf{s}}(p))$ to denote the ρ in the right hand side of (7) at the rest of the paper.

Actually, many different types of functionals can be adopted here to model a curved surface, e.g., the multiple-RBF proposed in [20]. Considering its simplicity, we choose the quadratic functional. If the two second-order parameters d and e are zero, the functional degenerates to a plane. Thus, it is a natural extension of the classical plane-fitting approaches [6]. In section 3.1 we will further show a straightforward way to initialize the functional.

However, the surfaces types that a quadratic functional can model are limited. To circumvent this problem, we propose to segment images to almost regular blocks by the super-pixel approach SLIC [21]. As shown in Fig. 2(a-b), the segmentation obtained by the SLIC is better than mean-shift [22] for our application. In each small regular grid, a quadratic functional could be a good approximation to the disparity surface.



(a) mean-shift [22] (b) SLIC [21] (c) plane-fit [6] (d) after optimization

Fig. 2. Some results of the Middlebury dataset Bowling2 [23]. (a-b) are segmentations. (c-d) shows the absolute residuals $\sum_i |L_i \mathbf{u}|$ with respect to disparity map $\mathbf{u}$ obtained by plane-fit[6] and our approach respectively. See text for more details.

2.3 Overall Energy

Combining the pixel-wise second-order smoothness (6) and the matching cost (7) with built-in "as-rigid-as-possible" smoothness , we obtain the energy

$$E_1(\mathbf{u}, \mathbf{\Lambda}) = \mathbf{u}^\top (\sum_i L_i^\top L_i)\mathbf{u} + \lambda \sum_{\mathbf{s}} \sum_{p \in \mathbf{s}} \rho(p, p - D_{\mathbf{s}}(p)) \tag{9}$$

where $\mathbf{u} = \mathbf{D}(\mathbf{\Lambda})$. However, this results in a tight dependence between the two energy terms, which is difficult to optimize. Inspired by the work [24][5], we introduce an intermediate term to decouple the tight dependence of the two subenergies. In summary, the overall energy is defined as

$$E_2(\mathbf{u}, \mathbf{\Lambda}) = \mathbf{u}^\top (\sum_i L_i^\top L_i)\mathbf{u} + \theta(\mathbf{u} - \mathbf{v})^\top G(\mathbf{u} - \mathbf{v}) + \lambda \sum_{\mathbf{s}} \sum_{p \in \mathbf{s}} \rho(p, p - D_{\mathbf{s}}(p)) \tag{10}$$

where $\mathbf{v} = \mathbf{D}(\mathbf{\Lambda})$ is the disparity map determined by the surface equations $\mathbf{\Lambda}$, G is an identity matrix[1], θ is a scalar determining the tightness of the couple between $\mathbf{u}$ and $\mathbf{v}$. When θ increases from 0 to a large value, the two disparity maps gradually converge to an agreement.

2.4 Interactions between the Two Priors

Without the segment-wise "as-rigid-as-possible" (ARAP) prior, the solution may easily get trapped in bad local minimums, such as those shown in Fig. 4, because the pixel-wise second-order prior itself is too flexible to handle this issue. Without the second-order prior term, the energy is almost the same as traditional segment-based approaches [6], which usually suffer from misalignments along segment boundaries. Fig. 2(c) shows the absolute residual $\sum_i |L_i \mathbf{u}|$ of a disparity map obtained by plane-fitting. We can easily find strong seams between regions while inside each region the residual is zero due to the collinearity. By combining the two priors, the over-flexible issue of the second-order prior is restricted by the ARAP prior. At the same time, the seams caused by the ARAP

[1] Each g_{ii} could be a confidence assigned to pixel i if we have that kind of information.

prior are smoothed out by the second-order prior, as shown in Fig. 2(d). It is interesting to note that in some regions on the bowling ball non-zero residuals appear, see the three patches of Fig. 2(d). Since the overall intensities are very small, we clip three patches and visualize their intensities with a large scaling factor. The reason is that after the optimization, the disparity functionals become slightly non-linear. Although the second-order prior punishes such changes, the overall energy decreases due to the smaller data cost. In all, the functionalities of the two priors are complementary to each other. As demonstrated in section 4, the two priors work together to generate accurate and seamless dense correspondences.

2.5 Optimization

The problem is solved by an alternative optimization approach. First, by fixing $\mathbf{u}$ we minimize the energy in (10) with respect to $\mathbf{\Lambda}$

$$E_D(\mathbf{\Lambda}) = \theta(\mathbf{u} - \mathbf{D}(\mathbf{\Lambda}))^\top G(\mathbf{u} - \mathbf{D}(\mathbf{\Lambda})) + \lambda \sum_{\mathbf{s}} \sum_{p \in \mathbf{s}} \rho(p, p - D_{\mathbf{s}}(p)) \tag{11}$$

where $\mathbf{D}(\mathbf{\Lambda})$ denotes the functional vector. It is easy to verify that, given $\mathbf{u}$, the optimizations for different segments are independent. Therefore, minimization can be conducted in parallel. Any derivative-free algorithms can be adopted to minimize it. We will present a specially designed approach in section 3.2. When θ is zero, optimizing (11) is similar as the plane-fitting approaches [6].

Second, by fixing $\mathbf{D}(\mathbf{\Lambda})$, that is $\mathbf{v}$, we solve a quadratic programming problem with respect to $\mathbf{u}$

$$E_S(\mathbf{u}) = \mathbf{u}^\top (\sum_i L_i^\top L_i)\mathbf{u} + \theta(\mathbf{u} - \mathbf{v})^\top G(\mathbf{u} - \mathbf{v}) \tag{12}$$

which has a closed-form solution by solving the linear equation

$$(\sum_i L_i^\top L_i + \theta G)\mathbf{u} = \theta G \mathbf{v} \tag{13}$$

Since $\sum_i L_i^\top L_i + \theta G$ is a sparse positive-semidefinite matrix, the equation can be efficiently solved by the Cholesky decomposition [25].

The alternated optimization is enclosed by an outer loop, which increases θ from zero to a large value. Thus, at the beginning, due to the loose couple, i.e., $\theta(\mathbf{u} - \mathbf{D}(\mathbf{\Lambda}))^\top G(\mathbf{u} - \mathbf{D}(\mathbf{\Lambda}))$ is small, both the two functionals $\mathbf{u}$ and $\mathbf{\Lambda}$ have freedom to change their values to decrease the total energy. While with the increase of θ, the couple of them becomes tighter. Finally, $\mathbf{u}$ and $\mathbf{D}(\mathbf{\Lambda})$ converge to an agreement. The whole algorithm is summarized in Algorithm 1.

3 Implementation

In this section, we present some further implementation notes to allow the readers to more easily replicate our method.

Algorithm 1. As-Rigid-As-Possible Stereo

```
compute Laplacian matrix L_i
set θ to zero
repeat
  repeat
    minimize Eq. (11) with respect to Λ by a derivative-free algorithm
    solve a quadratic programming Eq. (12) by sparse Cholesky decomposition
  until converged
  increase θ
until converged
```

3.1 Initialize

A good initialization will make the model converge fast. We adopt a similar plane-fitting approach as proposed in [6] to estimate a plane equation for each segment. Our approach consists of three steps:

1. A matching cost volume is computed based on Eq. (8).
2. Some robust matches are obtained by the WTA strategy followed by the left-right consistency check. The matches are not necessary to be accurate because the plane equations will further be optimized in the model.
3. A plane equation is estimated by RANSAC for each segment. Six matches are randomly sampled to estimate the three parameters of a plane equation, whose cost is then computed by querying pre-computed cost volume [6]. This process is repeated about 20 times[2], and the plane with lowest cost is kept as the result.

It is noted that, unlike in a traditional RANSAC algorithm where the number of inliers is used to evaluate the quality of an estimation, we use the matching cost as the quality measure. The objective of step 3 is coherent with the Eq. (9) where we try to minimize the matching cost too. For small segments or segments without initial matches, we just copy equations of the most similar (measured by mean colors) segment from the local neighborhood.

3.2 Optimize E_D

By the initialization step, we have obtained an approximately good equation for the three 1st-order parameters of each segment, then we randomly assign small values for the rest two 2nd-order parameters, d and e. As we have mentioned the analytical derivatives of Eq. (11) is difficult to compute, we adopt a derivative-free approach to optimize it. The Nelder-Mead simplex algorithm [26] is chosen because a good property of it is that we are able to customize the initialization

[2] Different from previous segment-based methods where the number is usually large to guarantee a good estimation, we only need a rough estimation because the surface will be optimized by consecutive steps.

of the $N+1$ vertices of the simplex, where $N = 5$ is the dimension of the variable being optimized. Starting from the space spanned by the initial simplex, better solutions are searched by effective rules of Nelder-Mead algorithm.

To minimize E_D with respect to a segment, we randomly pick six surface equations from its neighboring segments to initialize the Nelder-Mead simplex. This idea is motivated by the PatchMatch approach [1] where good estimations are propagated among adjacent pixels. The basic assumption is that adjacent segments are likely to share similar surface equations.

3.3 Optimize E_S

For large images, although a directional Lalpacian matrix L_i is very sparse, the $\sum_i L_i^\top L_i$ becomes a little dense. This fact causes the Cholesky decomposition slow. To address the problem, we suggest not to solve the linear equation exactly as in Eq. (13). Since Eq. (12) is a convex problem, we can use the gradient descent approach to find a better $\mathbf{u}^{(t+1)}$ from a start point instead of getting the best $\mathbf{u}$

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} - \tau \nabla E_S \tag{14}$$

where $\nabla E_S = 2(\sum_i L_i^\top L_i)\mathbf{u}^{(t)} + 2\theta G(\mathbf{u}^{(t)} - \mathbf{v})$ is the gradient of E_S, and τ is a step size.

3.4 Post-process

After optimization, we perform a left-right consistency check to label inconsistent pixels as unknown. The proposed model has a nature way to re-fill unknown pixels. In the G matrix of Eq. (12), we set the corresponding g_{ii} of unknown pixels to zero, then solve the quadratic programming again. Unknown pixels will then be filled by disparities of neighboring pixels with large weights in W. Therefore, disparities of occluded pixels are likely be filled by background instead of foreground, since the intensity edge between them leads to a small weights in W. No other post-processing is conducted.

4 Experiments

We tested our approach on the Middlebury stereo benchmark [2]. The experiments are conducted on a PC equipped with an i7-2600 3.40GHz CPU and 8GB memory. For the parameter setting, we choose $\{\alpha, \tau_{col}, \tau_{grad}\} = \{0.85, 20, 4\}$. The gradient feature is a 2-dimensional vector computed by the 3×3 Sobel operators in horizontal and vertical directions, preceded by a gaussian blur. The value of α is set relatively large to favor the gradient feature in order to deal with radiometric difference between left and right frames. On the other hand, the truncations τ_{col} and τ_{grad} are set to relatively small values to increase the robustness in occlusion. The weight λ which balances the smoothness and matching cost is set to 2. For the construction of the Laplacian matrix, we choose the

Table 1. Top ranked entries of the Middlebury stereo benchmark evaluated at 0.5 error threshold. Our method currently is ranked at the third place out of 143 competing algorithms. In each cell the number denotes bad pixel rate, and subscript denotes ranking.

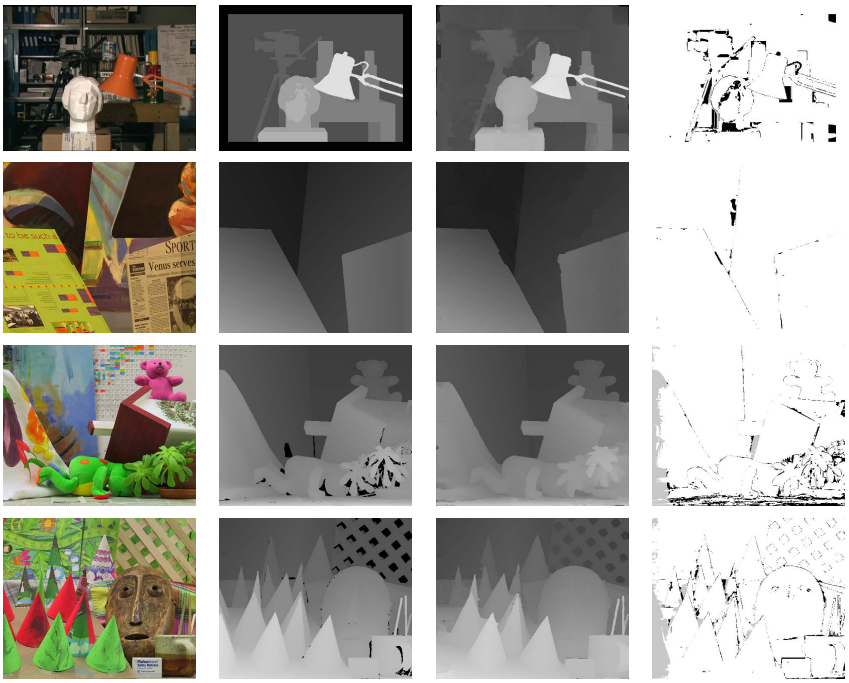
	Avg.	Tsukuba			Venus			Teddy			Cones		
	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
1.GC+LSL	3.8	5.04_2	5.56_2	14.0_{10}	0.66_3	0.88_3	5.82_4	4.20_1	7.12_1	12.9_1	3.77_5	9.16_5	10.4_8
2.PM-Huber	6.0	7.12_9	7.80_{10}	13.7_8	1.00_9	1.40_{10}	7.80_{12}	5.53_4	9.36_2	15.9_5	2.70_1	7.90_1	7.77_1
3.Ours	6.1	7.17_{11}	7.67_8	16.0_{25}	0.64_2	0.87_2	6.17_5	5.52_3	10.7_4	15.6_4	3.00_3	8.55_3	8.35_3
⋮			⋮			⋮			⋮			⋮	
7.PMBP	14.8	11.9_{44}	12.3_{40}	17.8_{48}	0.85_7	1.10_5	6.45_7	5.60_5	12.0_7	15.5_3	3.48_4	8.88_4	9.41_4
⋮			⋮			⋮			⋮			⋮	
9.PatchMatch	22.2	15.0_{63}	15.4_{62}	20.3_{74}	1.00_{10}	1.34_9	7.75_{11}	5.66_6	11.8_6	16.5_6	3.80_6	10.2_7	10.2_6

same $\gamma = 20$ in Eq. (3) for all directions. To segment the input by SLIC [21], we use $\{k, m\} = \{500, 10\}$, where k is the number of desired superpixels, and m is a compactness parameter to regularize the shape of each superpixel. All these parameters are kept constant for the online evaluation.

4.1 Comparisons of Results

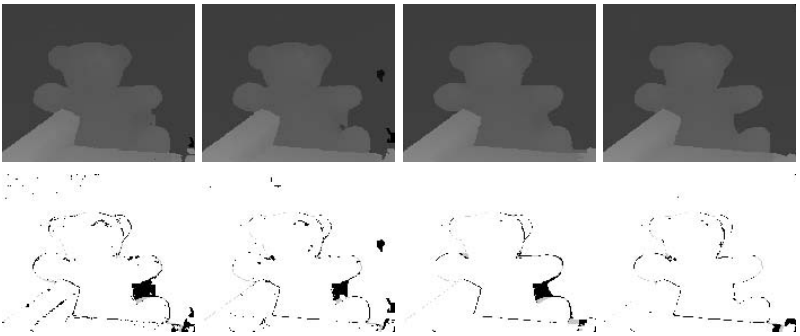
Our approach currently is ranked at the third place evaluated at the 0.5 error threshold. Table 1 details the error rates of top ranked methods. Disparity maps and bad pixel maps of our method are shown in Fig. 3. Fig. 4 compares results of our model and three PatchMatch-based models in a textureless region, the small region under the left arm of the teddy bear. It is clearly seen that PatchMatch model mislabeled this region. Although the matching cost with slanted support window defined in PatchMatch is generally very powerful, it still suffers from the lack of regularization limitation in textureless regions, which is a common problem for local stereo methods. More interestingly, the two global extensions of PatchMatch (i.e., PM-Huber [5] and PMBP [17]) which have an explicit smoothness regularization of the 3D label still failed in this region. This may support our claim that pixel-wise smoothness constraints are somehow too flexible to prevent the solution from getting trapped in bad local minimums. In contrast, our method can correctly recover this area due to the "as-rigid-as-possible" constraints introduced by segment-based priors.

To demonstrate that our model can correctly handle curved surfaces, we show point clouds generated from our disparity maps, as well as their Phong shaded versions of the Bowling, Baby, and Cloth [23]. As observed in Fig. 5, our method successfully recovered the surface of the bowling ball, while PatchMatch introduced two pits on the left side of the ball. An interesting observation is that if we take a careful look at the left image, the two fault regions of PatchMatch



(a) Left View (b) Ground Truth (c) Ours (d) Bad Pixels

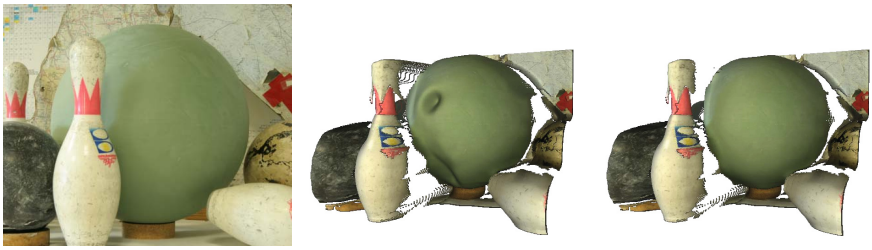
Fig. 3. Our results on the four online testing images of Middlebury. The error maps shown in the last column are evaluated at 0.5 error threshold.



(a) PatchMat.[1] (b) PMBP[17] (c) PM-Huber[5] (d) Ours

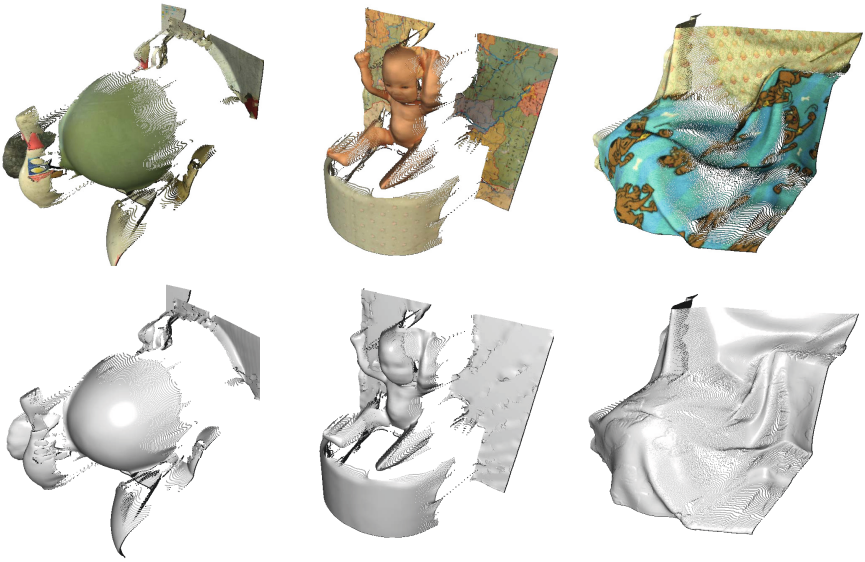
Fig. 4. Comparison of results in a textureless region. First row is the disparity maps. Second row is the bad pixel maps.

happen to be shadowed by the standing bottle, which probably results in ambiguous matching cost in these area. Local methods such as PatchMatch may probably fail in this situation. Fig. 6 shows more results. For the Baby case, the



(a) Left view (b) PatchMatch[1] (c) Ours

Fig. 5. Comparison of the reconstructed point cloud between PatchMatch and our method

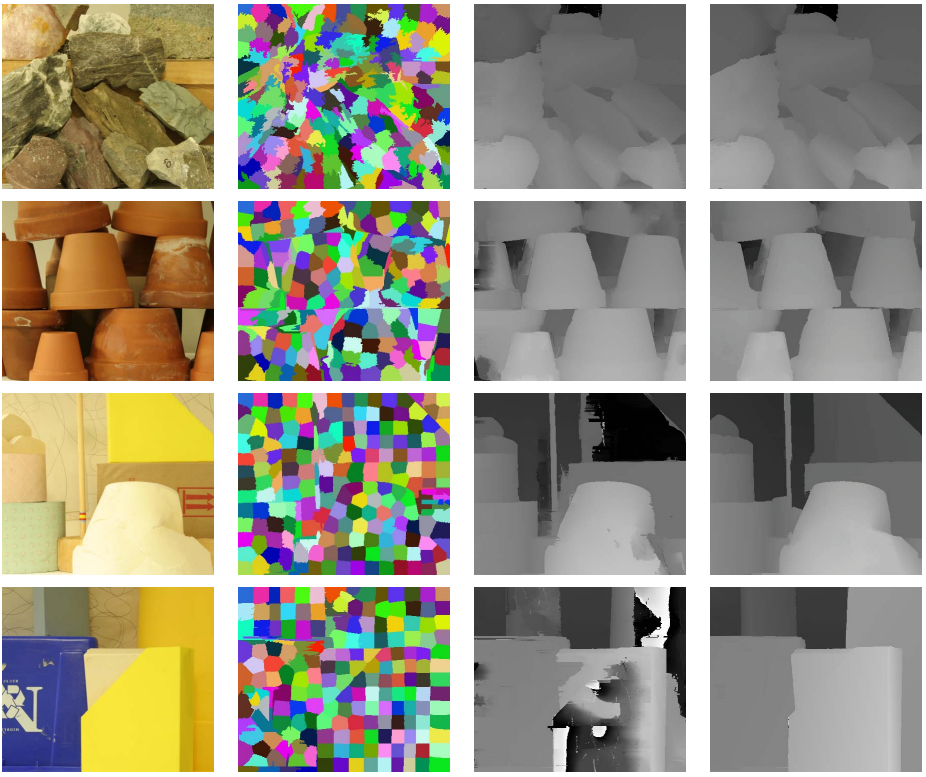


(a) Bowling (b) Baby (c) Cloth

Fig. 6. Point clouds and their Phong shaded versions obtained by our model

surface of the round table is correctly recovered, and the shape of the baby's body is well preserved too. The Cloth case is more challenging due to the surfaces with large curvature. However, our approach still works quite well under this hard condition.

In Fig. 7, we show more results of our method, compared with PatchMatch. We implemented the PatchMatch stereo algorithm following the parameters in the original paper [1]. Weighted median filtering is performed to pixels that fail to pass the left-right consistency check. For well textured scene such as the Rock case in the first row, both methods work well and the results are comparable. For weakly textured scene such as the Flowerpots case in the second



(a) Left View (b) SLIC[21] (c) PatchMatch[1] (d) Ours

Fig. 7. More results. From the first to the last row are the Rocks, Flowerpots, Lampshade and Plastic datasets, respectively.

row, PatchMatch can still work considerably well due to its powerful matching cost. But the result is not perfect. It can be clearly seen that the upper-right most pot is not correctly recovered, while this is not a problem for our method. For scenes containing almost textureless regions, such as the Lampshade and Plastic case in the third and last row, disparities computed by PatchMatch are almost completely erroneous in those regions. In contrast, our method can naturally handle this problem. These results again support our basic idea that "as-rigid-as-possible" is effective for stereo matching.

4.2 Running Time

Speaking of running time, our current implementation takes on average 10 seconds for a Middlebury pair. It runs considerably faster than previous methods, e.g., PatchMatch [1] about 1 minute and PM-Huber [5] about 2 minutes. A key reason is that we handle segments, whose number (e.g., hundreds or thousands) is orderly smaller than image pixels. Let us give more details regarding to the

implementation. First, the iteration number of the outer loop in Algorithm 1 is not necessarily to be large, since the disparity map usually converges in a few iterations. We increased θ from 0 to 1 by 10 uniform levels and performs two iterations of the inner loop for each level. Second, when solving for smoothness term in Eq. (12), it is not necessary to solve the large sparse linear system in Eq. (13) exactly. Since the energy in (12) is convex, we only need to perform a few steps of gradient descent instead. This results a huge acceleration. For an image of size 375×450, solving the sparse linear system requires approximately 4 seconds on our PC [25]. In contrast, we perform 10 steps of gradient descent, which only cost 0.1 seconds. Third, when solving for the segment-based matching cost term, we use the current best solutions of neighboring segments to initialize the simplex of current segment, which results in a faster convergence during Nelder-Mead search. We therefore keep the number of max iteration of Nelder-Mead low, i.e., 50 in out experiments. Finally since each segment can be optimized individually, we adopted OpenMP [27] for the parallelism.

5 Conclusion

We have presented a novel stereo model which simultaneously incorporates pixelwise second-order smoothness priors and the "as-rigid-as-possible" segment-based priors. The two priors work together to generate seamless, accurate and robust stereo correspondences. The model can be efficiently optimized by alternating between a quadratic programming and many parallel Nelder-Mead simplex search of surface equations. Experiments showed that our method can correctly handle curved surfaces, as well as large textureless regions. Compared to the powerful PatchMatch stereo algorithm, our method is about an order of magnitude faster and generally outperforms it.

References

1. Bleyer, M., Rhemann, C., Rother, C.: Patchmatch stereo - stereo matching with slanted support windows. In: BMVC, pp. 1–11 (2011)
2. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47(1-3), 7–42 (2002)
3. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l^1 optical flow. In: DAGM-Symposium, pp. 214–223 (2007)
4. Ranftl, R., Gehrig, S., Pock, T., Bischof, H.: Pushing the limits of stereo using variational stereo estimation. In: Intelligent Vehicles Symposium, pp. 401–407 (2012)
5. Heise, P., Klose, S., Jensen, B., Knoll, A.: Pm-huber: Patchmatch with huber regularization for stereo matching. In: ICCV, pp. 2360–2367 (2013)
6. Klaus, A., Sormann, M., Karner, K.F.: Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: ICPR (3), pp. 15–18 (2006)
7. Hong, L., Chen, G.: Segment-based stereo matching using graph cuts. In: CVPR (1), pp. 74–81 (2004)

8. Tao, H., Sawhney, H.S., Kumar, R.: A global matching framework for stereo computation. In: ICCV, pp. 532–539 (2001)
9. Wang, Z.F., Zheng, Z.G.: A region based stereo matching algorithm using cooperative optimization. In: CVPR (2008)
10. Grimson, E.: From images to surfaces: a computational study of the human early visual system. MIT Press (1981)
11. Terzopoulos, D.: Multilevel computational processes for visual surface reconstruction. Computer Vision, Graphics, and Image Processing 24(1), 52–96 (1983)
12. Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press (1987)
13. Ishikawa, H., Geiger, D.: Rethinking the prior model for stereo. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 526–537. Springer, Heidelberg (2006)
14. Woodford, O.J., Torr, P.H.S., Reid, I.D., Fitzgibbon, A.W.: Global stereo reconstruction under second order smoothness priors. In: CVPR (2008)
15. Rother, C., Kolmogorov, V., Lempitsky, V.S., Szummer, M.: Optimizing binary mrfs via extended roof duality. In: CVPR (2007)
16. Olsson, C., Ulén, J., Boykov, Y.: In defense of 3d-label stereo. In: CVPR, pp. 1730–1737 (2013)
17. Besse, F., Rother, C., Fitzgibbon, A., Kautz, J.: Pmbp: Patchmatch belief propagation for correspondence field estimation. International Journal of Computer Vision, 1–12 (2012)
18. Forsyth, D.A., Ponce, J.: Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference (2002)
19. Yamaguchi, K., McAllester, D.A., Urtasun, R.: Robust monocular epipolar flow estimation. In: CVPR, pp. 1862–1869 (2013)
20. Zhou, X., Boulanger, P.: New eye contact correction using radial basis function for wide baseline videoconference system. In: Lin, W., Xu, D., Ho, A., Wu, J., He, Y., Cai, J., Kankanhalli, M., Sun, M.-T. (eds.) PCM 2012. LNCS, vol. 7674, pp. 68–79. Springer, Heidelberg (2012)
21. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34(11), 2274–2282 (2012)
22. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603–619 (2002)
23. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: CVPR (1), pp. 195–202 (2003)
24. Steinbrücker, F., Pock, T., Cremers, D.: Large displacement optical flow computation withoutwarping. In: ICCV, pp. 1609–1614 (2009)
25. Chen, Y., Davis, T.A., Hager, W.W., Rajamanickam, S.: Algorithm 887: Cholmod, supernodal sparse cholesky factorization and update/downdate. ACM Trans. Math. Softw. 35(3) (2008)
26. Nelder, J., Mead, R.: A simplex method for function minimization (1965)
27. OpenMP Architecture Review Board: OpenMP application program interface version 3.0 (May 2008)