

Yale University
Department of Computer Science

**Stereo for Slanted Surfaces:
First Order Disparities and Normal Consistency**

Gang Li Steven W. Zucker
Yale University Yale University

YALEU/DCS/TR-1330
June 2005

Stereo for Slanted Surfaces: First Order Disparities and Normal Consistency

Gang Li
Yale University

Steven W. Zucker
Yale University

Abstract

Traditional stereo algorithms either explicitly use the frontal parallel plane assumption by only considering position (zero-order) disparity when computing similarity measures of two image windows, or implicitly use it by imposing a smoothness prior bias towards frontal parallel plane solution. However this introduces two types of systematic error for slanted or curved surfaces. The first type is structural, and relates to discrete pixel coordinates and neighborhood structure. The second is geometric, and relates to differential properties of surfaces. To eliminate these systematic errors we extend stereo matching to include first-order disparities. Contextual information is then expressed geometrically by transporting surface normals over overlapping neighborhoods, which takes a particularly simple (and efficient) form in the tangent plane approximation. In particular, we develop a novel stereo algorithm that combines first-order disparity information with position (zero-order) disparity for slanted surfaces, and illustrate its use.

1 Introduction

Two frame dense stereo correspondence algorithms have made significant progress [1, 7, 24, 5]. Working with a rectified image pair [9, 11], most of these algorithms exploit the *frontal parallel plane assumption* either explicitly or implicitly. In particular, it assumes position disparity (or depth) is constant (with respect to the rectified stereo pair) over a region under consideration. However, real world objects possess surfaces rich in shape, which generically violate the frontal parallel plane assumption (Fig. 1).

Traditional area based methods compare a window of the same size and shape in the left and right images and compute the similarity measure, where the frontal parallel plane assumption is explicitly used. Several algorithms address this problem. [12] uses a parameterized planar or quadratic patch

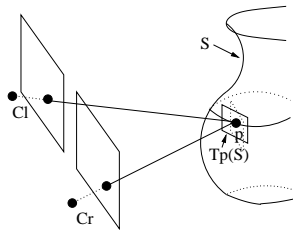


Figure 1: For a regular surface $S \subset \mathbb{R}^3$, the tangent plane $T_p(S)$ (in solid lines) at a point $p \in S$ is well defined. However traditional stereo algorithms use the frontal parallel plane (in dotted lines) to represent the (local) surface geometry at p , which is incorrect.

fit to the images as a local model for the disparity surface. In [14] variable window size (but fixed shape) is used. In [6] disparity derivatives are used to deform the matching window in a refined correlation algorithm. In [3, 17] slanted and curved surfaces are explicitly modeled for each segmented region, where segmentation and correspondence are iteratively obtained from the multiway-cut algorithm [4]. [20] develops a slanted scanline algorithm.

Taken individually, point-wise geometric constraints (e.g. epipolar constraint) and similarity measure (e.g. SSD) cannot always resolve matching ambiguities. Thus requiring neighboring matching pairs to be “consistent” is natural. An early cooperative algorithm[18] uses a local excitatory neighborhood of the same disparity level to support the current matching pair. In [28] local support is defined as the sum of all match values within a 3D local support volume. In [23] local support at different disparity hypotheses is diffused iteratively. In [4] a smoothness term over neighboring pixels is introduced in an energy function minimized by graph cuts. In [25, 26] messages (similarity measure weighted by gaussian smoothed disparity difference) are passed between nearby matching pairs in a Markov network by belief propagation. However these algorithms implicitly use the frontal parallel plane assumption since the neighboring matching pairs interact in a way such that frontal parallel plane solution is preferred.

Both the explicit and the implicit use of the frontal parallel plane assumption introduces systematic errors to stereo correspondence (see experiment section for details). To move beyond this assumption, locally it implies that the tangent plane $T_p(S)$ deviates from the frontal parallel plane. Our geometric observation then arises in several forms: (i) the shape of matching patches in the left/right image must vary; (ii) integer coordinates must be interpolated; (iii) disparity derivatives are related to surface differential geometric property; and (iv) continuity over overlapping neighborhoods must

include (at least) surface normal consistency. To take full advantage of (iii) and (iv), which follow directly from differential geometry, we further observe that some form of (Cartan) transport is required to combine information from different surface normals in a neighborhood around a putative matching point.

In this paper our task is to develop a new constraint for planar surfaces not in the frontal parallel plane using basic surface differential geometry. Specifically, we develop a novel stereo algorithm that explicitly takes into account first-order disparities for non-frontal-parallel surfaces. This amounts to: (1) the local deformation of the SSD window, which gives us continuous (interpolated) disparity as well as first order disparities (surface orientation); and (2) describing geometric consistency between nearby matching pairs by using depth (position disparity) and surface normal (first-order disparities), which *does* help the matching process. The second part behaves like an *extra geometric constraint* for stereo correspondence. While the epipolar constraint is well studied and utilized in stereo correspondence, geometric contextual constraints among nearby matching pairs are largely unexplored. In this paper we explore the simplest such geometric constraint for slanted or curved surfaces by taking into account their orientation. One implementation illustrates how well these constraints perform in practice, extending performance beyond traditional algorithms on several scenes. We hope this novelty will inspire researchers since our constraint could be used in more elaborate stereo algorithms.

2 Background

Assume the image pair is rectified [11]. For a 3D point p with disparity d , 3D planes (centered at p) in \mathbb{R}^3 can be classified as frontal parallel, horizontally slanted, vertically slanted, and in general configuration, respectively (Fig. 2(a-d)). First order disparities ($\{\frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$) do not vanish simultaneously for non-frontal parallel planes. Thus they cannot be ignored, otherwise systematic error will arise. In this section we describe how to take this effect into account for any local (window-based) algorithms with an aggregation step. In the next section we show how these first order disparities can be used to impose geometric contextual information by transporting surface normals over overlapping neighborhoods.

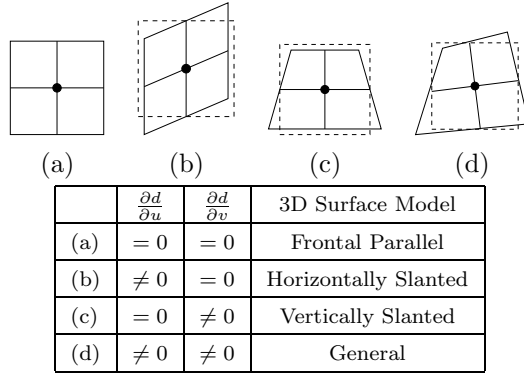


Figure 2: (LEFT: a-d) 3D plane types in \mathbb{R}^3 : frontal parallel, horizontally slanted, vertically slanted, and in general configuration, respectively (frontal parallel plane also drawn in dashed lines for comparison); (RIGHT) First order disparities for different 3D plane types.

2.1 Deforming Matching Window by First-order Disparities

To find correspondence, traditional area based methods use a small window (e.g. 9x9) centered at (u, v) in the left image, and compare it with a window of the same size and shape at $(u - d, v)$ in the right image using similarity measure such as normalized cross correlation (NCC), sum of squared difference (SSD), or sum of absolute difference (SAD), etc. Disparity estimate d is obtained by selecting the one that gives the best similarity measure. When the scene within each window satisfies the frontal parallel plane assumption the above method is valid. However when the 3D surfaces are slanted in depth or curved the problem formulation has to be modified, otherwise systematic error will arise. In Fig. 9(corridor), traditional SSD yields a stepwise scalloped pattern (frontal parallel planes at different depth) due to such systematic error.

To illustrate, we consider a small image window of a slanted planar surface (Fig. 3) for similarity measure (SSD). If the correspondence of (u, v) in the left image is $(u - d, v)$ in the right image, then to a first order approximation the correspondence of $(u + \Delta u, v + \Delta v)$ (black dots) in the neighborhood (in the left image) is $(u + \Delta u - d - \frac{\partial d}{\partial u} \Delta u - \frac{\partial d}{\partial v} \Delta v, v + \Delta v)$ (also black dots) in the right image, with $\frac{\partial d}{\partial u}$ and $\frac{\partial d}{\partial v}$ the partial derivatives of disparity d with respect to u and v , respectively; Δu and Δv the small step size in each direction. Typically $\Delta u, \Delta v \in \{0, \pm 1, \pm 2, \dots\}$. When the 3D surface is not frontal parallel, these black dots are different from the measured gray dots. Matching window in the right image must be deformed accordingly.

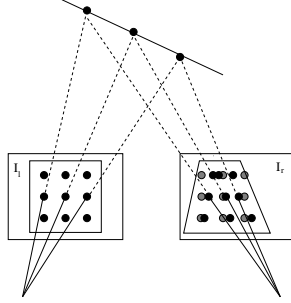


Figure 3: Computing a similarity measure using an image window of size 3x3. To overcome the limitation of the frontal parallel plane assumption, in the other image window must be deformed according to the shape of the (disparity) surface. This involves both window shape and corresponding pixels. Shown are true corresponding points (as black dots) in the right image of a slanted surface, which are different from measured gray dots.

The deformed window SSD is:

$$\begin{aligned}
 dw_SSD(u, v, d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}) = & \\
 \sum_{(u+\Delta u, v+\Delta v) \in \mathcal{N}_{uv}} & (I_l(u + \Delta u, v + \Delta v) - \\
 \hat{I}_r(u + \Delta u - d - \frac{\partial d}{\partial u} \Delta u - \frac{\partial d}{\partial v} \Delta v, v + \Delta v))^2 & \quad (1)
 \end{aligned}$$

where \mathcal{N}_{uv} denotes the window centered at (u, v) , and \hat{I}_r is the linearly interpolated intensity of two nearest integer index positions in the right image.

With this new formulation of the similarity measure, the correspondence problem is then: for every (u, v) in the left image to select $\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ that gives the best similarity measure:

$$\arg \min_{\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}} dw_SSD(u, v, d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}) \quad (2)$$

We use direction set method [21], a multidimensional minimization method, initialized with the integer disparity d_I (obtained from traditional SSD) and zeros for the first order disparities. The results are the (interpolated) floating point disparity d and first order disparities $\{\frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ that achieve the best similarity measure at (u, v) . In [6] such a deformed window was also used. Our contribution is to relate the deformation to surface orientation

and impose geometric consistency over overlapping neighborhoods by using surface orientation, which provides extra geometric constraints for stereo correspondence. In the next section details will be given.

Note that several recent algorithms [4, 16, 25] use a pixel dissimilarity measure [2], which is insensitive to sampling by using linearly interpolated intensity functions. This measure is pixel-wise. To compute the similarity between two image windows using correlation or SSD, this measure has to be adjusted according to the deformation described above, otherwise systematic error may still arise.

3 Geometric Consistency from First-order Disparities

This is the central part of our paper. Since ambiguities (along the epipolar line) cannot always be resolved using the local measurements described previously, we require neighboring matching pairs to be consistent. Unlike others [28], we derive this consistency geometrically, with surface orientation playing an important role.

Assume cameras are calibrated. The stereo pair has a baseline B and focal length α (pixels), and the depth of a point with disparity d is [9]:

$$Z(u, v) = \frac{B\alpha}{d(u, v)}$$

Differentiating Z we have:

$$Z_u = \frac{\partial Z}{\partial u} = -\frac{B\alpha}{d^2} \frac{\partial d}{\partial u} \quad (3)$$

$$Z_v = \frac{\partial Z}{\partial v} = -\frac{B\alpha}{d^2} \frac{\partial d}{\partial v} \quad (4)$$

The surface normal at this point is then $(1, 0, Z_u)^T \times (0, 1, Z_v)^T$, which after normalization is:

$$\mathbf{N} = \frac{(-Z_u, -Z_v, 1)^T}{\sqrt{Z_u^2 + Z_v^2 + 1}} \quad (5)$$

With the (putative) surface depth and normal at hand, we can then utilize contextual information geometrically. The basic idea is in the spirit of the Cartan moving frame model [15, 8], which specifies how adapted frame fields change when they are transported along an object. And this model can be used to integrate local (geometric) information with (geometric) information in the neighborhood. We use i to denote a candidate match, i.e.

$\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ at pixel (u, v) , or equivalently $(u, v) < - > (u - d, v)$ with first order disparities $\{\frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$.

Geometric consistency means that, for a matching pair i , when it is transported along the object to a neighboring position, it is consistent with a neighboring matching pair j (Fig. 4). Note that the local surface approximation around i (computed from the information *at* i) serves as the object along which geometric information is transported, and is formally known as the *osculating object*. In this paper, the osculating object for a surface S takes its simplest form, which is the tangent plane $T_p(S)$ at $p \in S$. Since i and j refer to matching pairs, they each encode position disparity (depth) separately. Note that first order disparities at i and j further encode surface normals at these two points. Fig. 4 shows the geometric consistency between a matching pair i and its neighboring matching pair j (in solid lines). The tangent planes provide a natural description of the *geometric compatibility* r_{ij} between i and j :

$$r_{ij} = 1 - \frac{1}{m} (|\mathbf{v}_{ij} \cdot \mathbf{N}_i| + |\mathbf{v}_{ji} \cdot \mathbf{N}_j|) \quad (6)$$

where \mathbf{v}_{ij} denotes the vector from i to j , \mathbf{N}_i is the surface normal at i , and m is a normalization constant. Also shown are different neighbors j (dashed lines) of different depth and orientation, to illustrate the importance of both zero-order disparity (depth) and first-order disparities (surface orientation) in determining such consistency measure.

The geometric constraint (eqn. (6)) can be used as follows. For a matching pair i (hypothesis), we initialize its support s_i^0 according to its deformed window SSD (denoted by c_i) and iteratively update s_i by the geometric support it receives from its neighboring matching pair j :

$$s_i^0 = 1 - \frac{c_i}{c} \quad (7)$$

$$s_i^{t+1} = \frac{\sum_{j \in \mathcal{N}_i} r_{ij} s_j^t}{\sum_{j \in \mathcal{N}_i} s_j^t} \quad (8)$$

with c a normalization factor, \mathcal{N}_i denotes the neighbors of i (in our experiments we use a $5 \times 5 \times 3$ (u, v, d) region). The true correspondence will be supported by its neighbors since their local surface geometry estimates are geometrically consistent. False matches are unlikely to get support from neighbors.

Assuming the noise in the surface normals is roughly zero mean Gaussian i.i.d. (independent and identically distributed), the “best fit” (in a least-

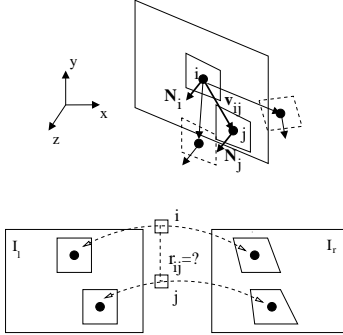


Figure 4: Illustration of the geometry for the basic matching constraint. Surface patch i on a slanted surface in \mathbb{R}^3 projects to patches in the left and right images. For a nearby patch j in \mathbb{R}^3 , the surface normal \mathbf{N}_j must be parallel to \mathbf{N}_i , and the two patches must both lie on the same surface. Possible patches (dashed lines) which lie off the (tangent) plane at i , or at the wrong orientation, are not consistent. Both zero-order (position) disparity and first-order disparities are essential in determining such geometric consistency.

squares sense) unit normal at i is updated as [22]:

$$\mathbf{N}_i^{t+1} = \frac{\sum_{j \in \mathcal{N}_i} \mathbf{N}_j^t}{\sigma} \quad (9)$$

with $\sigma > 0$ and

$$\sigma^2 = \left(\sum_{j \in \mathcal{N}_i} N_{jx} \right)^2 + \left(\sum_{j \in \mathcal{N}_i} N_{jy} \right)^2 + \left(\sum_{j \in \mathcal{N}_i} N_{jz} \right)^2$$

Note that even without calibration information, such geometric consistency support can still perform in the disparity space. Surface normal in the disparity space is then $\mathbf{N} = \frac{(-\frac{\partial d}{\partial u}, -\frac{\partial d}{\partial v}, 1)^T}{\sqrt{(\frac{\partial d}{\partial u})^2 + (\frac{\partial d}{\partial v})^2 + 1}}$.

For a line segment in the 3D plane, the orientation difference between its projections in the left and right images is called the *orientation disparity* [13, 27], which is widely studied in psychophysics. Note that first order disparities $\{\frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ encode such orientation disparity.

3.1 Relation to Energy Minimization Formulation

Our problem formulation consists of two parts: (1) It requires zero and first order disparities ($\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$) to be consistent with image measurement, or “data” consistency; and (2) It requires neighboring matching pairs i and

j to be geometrically consistent, or “smoothness” consistency. According to the taxonomy [24], our algorithm is neither a *local method* (e.g. SSD) nor a *global method* (e.g. graph cuts). It is in the spirit of a cooperative algorithm [18, 19, 28], which iteratively performs local computations and uses nonlinear operations resulting in a final effect similar to global optimization. Nevertheless, to facilitate intuition here we show its relation to global optimization method (graph cuts) that uses an energy minimization formulation. Graph cut algorithms [4, 16] for stereo find the disparity labeling d for every pixel p in the reference image such that the following energy is minimized:

$$\begin{aligned} E(d) &= E_{smooth}(d) + E_{data}(d) \\ &= \sum_{\{p,q\} \in \mathcal{N}_{\mathcal{I}}} V_{p,q}(d_p, d_q) + \sum_{p \in \mathcal{I}} D_p(d_p) \end{aligned} \quad (10)$$

where in the first term $\mathcal{N}_{\mathcal{I}}$ is the set of all neighboring pairs of image pixels $\{p, q\}$ (determined by 4-neighborhood, for instance), and E_{smooth} measures the extent to which d is not smooth; E_{data} measures the disagreement between d and the observed data, and the sum is over all image pixels p in the reference image \mathcal{I} . A popular smoothness prior is the Potts model [4] which is a piecewise constant model. To overcome the difficulty of such a model for handling non-frontal parallel planes, a piecewise smooth prior has been introduced [4] by using either a truncated quadratic or a truncated linear function ($V_{p,q}(d_p, d_q) = C \cdot \min(K, |d_p - d_q|)$) centered at zero disparity difference. But still this model prefers a frontal parallel plane solution. To illustrate, Fig. 5 shows a truncated linear piecewise smooth prior $V_{p,q}(d_p, d_q)$ centered at disparity d_p of pixel p . Since this smooth model is not oriented according to surface orientation at (p, d_p) , a point that comes from the same surface at a neighboring pixel q with disparity d_q still has a large penalty in the smooth term. This model still prefers an erroneous space point with disparity d_p at q , which lies in the frontal parallel plane at (p, d_p) . It is just the consideration of the surface orientation that allows us to encode the true (up to first order approximation) geometric contextual information in the algorithm.

Fig. 6 shows the problem when first order disparities (surface orientation) are not considered. It consists of a horizontally slanted plane with $\frac{\partial d}{\partial u} = 0.25$ and $\frac{\partial d}{\partial v} = 0$. Fig. 6(d) shows the result by graph cuts [4]. Their result has staircases because their contextual structure is biased towards frontal parallel planes. Without modeling the orientation of the 3D surface, such results will generically arise. On the other hand, by considering

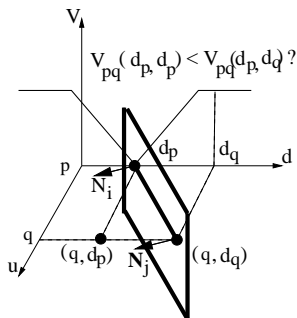


Figure 5: Shown is the smoothness prior (V) as a function of disparity difference ($d_p - d_q$) for neighboring pixels p and q along the epipolar line. The classical prior V has its valley oriented parallel to the frontal parallel plane. Our constraint amounts to a rotation of the valley according to surface orientation (surface shown in bold).

the orientation, a correct disparity map can be obtained (Fig. 6(c), please see electronic version for better comparison), also note the smooth disparity change. This agrees with the observation by Tappen and Freeman [26]: “the greatest increase in performance will be found by improving the formulation of the MRF, rather than improving the solutions found for the MRF’s currently being used.”

In summary, without considering surface orientation, the piecewise smooth model still implies frontal parallel plane assumption. Zero-order disparity (position disparity) alone does not suffice, first-order disparities (surface orientation) have to be considered as well.

In the context of the energy minimization problem, our new formulation of stereo correspondence, without a bias towards frontal parallel plane solution, is to find the disparity labeling d and its differential d' to minimize the energy functional:

$$\begin{aligned}
 E(d, d') &= E_{smooth}(d, d') + E_{data}(d, d') \\
 &= \sum_{\{p,q\} \in \mathcal{N}_{\mathcal{I}}} V_{p,q}(d_p, d_q, d'_p, d'_q) + \sum_{p \in \mathcal{I}} D_p(d_p, d'_p)
 \end{aligned} \tag{11}$$

An example smooth prior is $V_{p,q}(d_p, d_q, d'_p, d'_q) = 1 - r_{ij}$, with r_{ij} the compatibility (eqn. (6)) between i (pixel p with disparity d_p and differential d'_p), and j (pixel q with disparity d_q and differential d'_q). To make the data term meaningful here D_p could be the aggregated data cost. Note that when surface patch at i and j are frontal parallel ($\mathbf{N}_i = \{0, 0, 1\}^T$, $\mathbf{N}_j = \{0, 0, 1\}^T$), in the disparity space the new smooth prior becomes

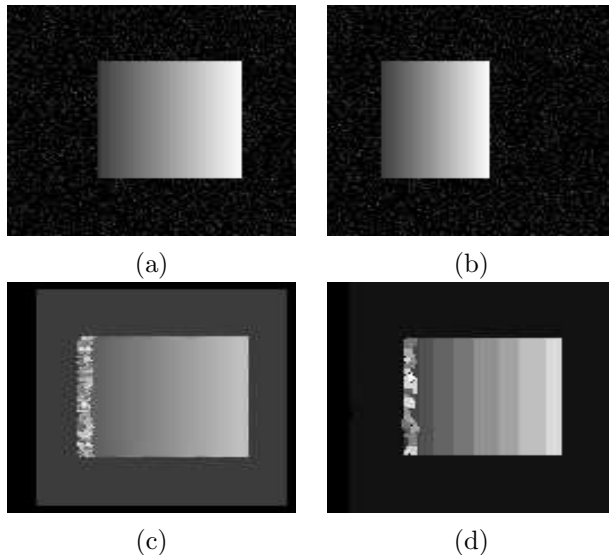


Figure 6: Synthetic pair to illustrate the importance of surface orientation on smoothness term: (a)(b) Left and right images. (c) Our result. (d) Disparity map by graph cuts.

$\frac{1}{m}(|\mathbf{v}_{ij} \cdot \mathbf{N}_i| + |\mathbf{v}_{ji} \cdot \mathbf{N}_j|) = \frac{1}{2m}|d_p - d_q|$, which after truncation is the piecewise smooth prior used in [4].

This new geometric compatibility requirement between nearby matching pairs, or contextual information expressed geometrically, can be used in graph cuts algorithms [4, 16] or belief propagation algorithms [25, 26]. It amounts to the formulation of different Markov Random Fields. However, it would be expensive to compute directly. In the next subsection we describe an approximation that efficiently accomplishes “data” consistency and “smoothness” consistency separately. Although our solution cannot guarantee the global optimization, experimental results show it works well in practice.

3.2 Stereo Algorithm

In this section we describe our stereo algorithm:

- (1) Use traditional similarity measure (e.g. SSD) for integer disparity values at each (u, v) , keep only the top $\delta\%$ (we use 3 non-immediate neighboring ones in our experiments) as the initial disparity hypotheses.
- (2) For each disparity hypothesis at every (u, v) , use optimization method (we use direction set method [21]) to obtain $\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ that minimizes

dw_SSD in equation (1) based on the deformed window. The input to the optimization method are the initial integer disparity hypotheses, together with the first order derivatives initialized to zeros. Note that d obtained this way is interpolated in the continuous domain. At each pixel (u, v) we could have several local minima based on dw_SSD. Geometric contextual information will be explored in the next few steps. $\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ could also be obtained by enumerating different combinations of these parameters if they are properly quantized, and selecting the set that minimizes dw_SSD.

(3) Compute the initial support s_i^0 at each matching pair i by equation (7), which encodes the similarity measure based on the deformed window SSD.

(4) Iteratively update the geometric support s_i at every i by equation (8) until it converges (in practice we run a preset number (e.g. 8) of iterations), using the compatibilities between nearby matching pairs r_{ij} (eqn. (6)) which denotes how geometrically consistent they are. Also update surface normal \mathbf{N}_i at i (eqn. (9)) based on the normals of neighbors, to reduce the effect of local noisy measurements.

(5) For each (u, v) select the the disparity with the highest support s , output disparity and surface normal.

Observe that steps (3)-(5) are the unique geometric content of our algorithm.

4 Experimental Results

Fig. 9 shows three examples. The first example (first row) is the synthetic ‘‘Corridor’’ pair [10] from University of Bonn, The image size is 256x256 pixels with a disparity range 11. Next to the original image pair we show the disparity map and surface normal by our algorithm. For comparison, in the second row we show disparity maps of traditional SSD (same window size, 9x9), graph cuts [4], belief propagation [26], respectively. They are obtained (and similarly for all other examples) from the stereo package provided by Scharstein and Szeliski [24]. In particular, the α - β -swap algorithm [4] for graph cuts, and max-product (contributed by [26]) for belief propagation. Also shown is the cooperative algorithm [28] (software kindly provided by the authors) result. Note that other algorithms obtain stepwise scalloped pattern because of the frontal parallel plane assumption being used, either explicitly or implicitly. Also note the gradual (continuous) disparity change in our result, we achieve such better result because we explicitly model 3D surface geometry.

Our algorithm takes 538 seconds; the synchronous belief propagation algorithm [26] takes 552 seconds, while the accelerated version [26] takes 75 seconds; the graph cuts algorithm [4] takes 95 seconds. The results are obtained with Pentium III M 1.0GHz CPU. The disparity map scale factor is 21.25. We use the taxonomy package [24] to compute the disparity error statistics (Fig. 7). Two measures are used here: (i) RMS (root-mean-squared) error (measured in disparity units); and (ii) Percentage of bad matching pixels with absolute disparity error larger than a threshold δ_d . Note that in [24] δ_d is set to 1.0, here we also report the results when δ_d is 0.5. Our algorithm outperforms other leading algorithms and has only about half of their errors. (cooperative result is included for visual inspection, its error statistics is not reported, partially due to the unspecified built in disparity scale factor in their software [28].) Also note that our goal is not to provide a complete stereo algorithm but to emphasize the importance of considering surface differential geometry (e.g. surface normal consistency) in dealing with slanted surfaces. The performance comparison should be viewed with this in mind. As we stated earlier, our new geometric constraint could be used in more elaborate algorithms (e.g. belief propagation [25, 26], graph cuts [4, 16]), but it is beyond the discussion of the current paper.

Disparity Error	SSD	GC	BPA	BPS	Our
RMS error (pixel)	1.31	0.65	0.75	0.62	0.35
% of errors $> \pm 1$	14.0	7.4	10.1	5.4	3.4
% of errors $> \pm 0.5$	32.5	26.4	30.3	24.3	11.6

Figure 7: Disparity error statistics on Corridor pair. RMS error and Percentage of bad pixels ($\delta_d=1$ and $\delta_d=0.5$), for SSD, Graph cuts, Belief propagation (accelerated), Belief propagation (synchronous), and Our algorithm, respectively. Our algorithm outperforms other algorithms for such scenes with slanted surfaces because we explicitly model 3D surface geometry.

The second example is the “Parking meter” pair from the well-known JISCT database. As in the previous example, we show our results (third row), together with results from other algorithms (fourth row). Once again other algorithms obtain stepwise scalloped pattern because of the frontal parallel plane assumption being used, either explicitly or implicitly. But we achieve better result because we explicitly model 3D surface geometry.

In recent years the Middlebury stereo database [24] has been very influential in providing a fair testbed and a taxonomy [24] for quantitatively evaluating algorithms’ performance. However we observe that this dataset is very limited in terms of 3D geometry of objects such as slanted or curved

surfaces, as objects are largely frontal parallel in this dataset. (This is why we include images from other databases as well.) Without the “slant” effect, our algorithm degrades roughly to the cooperative algorithm [28], and has similar error statistics using the taxonomy [24]. Due to space limit here we only report the results on one example from this dataset — the “sawtooth” pair. We show our results (fifth row), together with results from other algorithms (sixth row). Since the object is roughly frontal parallel, the difference between our result and other results is not obvious, as expected. Fig. 8 reports the error statistics (cooperative algorithm result provided by the authors [28]). For these last three examples, the geometric consistency support is performed in the disparity space.

Disparity Error	SSD	GC	BPA	BPS	CO	Our
RMS error (pixel)	1.65	1.42	1.67	1.45	1.46	1.30
% of errors $> \pm 1$	8.7	3.9	4.5	4.7	4.2	4.5

Figure 8: Disparity error statistics on Sawtooth pair. RMS error and Percentage of bad pixels ($\delta_d=1$), for SSD, Graph cuts, Belief propagation (accelerated), Belief propagation (synchronous), Cooperative algorithm, and Our algorithm, respectively. Our algorithm is comparable to other algorithms for such scenes with frontal parallel surfaces.

Note that the disparity sub-pixel refinement post-processing stage cannot be performed for belief propagation and graph cuts here. At pixel (u, v) , sub-pixel refinement stage takes the “cost” at the winning disparity $C(d)$, and the “costs” at the immediate two disparities $C(d-1)$, $C(d+1)$ (also at pixel (u, v)), and then fits a parabola for these three costs. Finally it outputs the floating number disparity which gives the minimum of this parabola. Essentially this step uses the frontal parallel plane solution at the floating number disparity to replace the frontal parallel plane solution at integer disparity. Although this step *does* alleviate some problems caused by the frontal parallel plane assumption, it *does not* solve the problem. The remaining “soft staircasing” is still noticeable (see [24] for examples). Furthermore, this post-processing step *does not* help the matching process. We use the Scharstein and Szeliski [24] software package (for SSD and Graph Cuts) with extension (for Belief Propagation) by Tappen and Freeman [26]. In this package, both Belief Propagation and Graph Cuts find the Maximum A Posteriori (MAP) estimates, which means the “costs” at other disparities are not necessarily meaningful, thus sub-pixel refinement cannot be performed. As a result, we did not perform such a sub-pixel refinement step in our comparison (except for cooperative algorithm, which has this step built in).

5 Conclusion

In this paper we show a novel stereo algorithm that explicitly takes into account first-order disparities of non-frontal parallel surfaces; it then relates these to surface normal and further impose geometric consistency between neighboring matching pairs. Our algorithm outperforms state-of-the-art algorithms on general surfaces (e.g. slanted surfaces), which suggests the power of surface differential geometry in obtaining a smooth solution for stereo correspondence. Our constraints could be used in more elaborate graph cuts or belief propagation algorithms.

However, there are several limitations of the current algorithm: First, as most local area based methods, it needs texture or shading variations to get a reliable local estimation. Possible solution to this limitation is to combine surface occluding contours where reliable information can be propagated to unreliable estimates. Second, occlusion is not considered in the current model. And third, object boundaries are not used, which provide information on depth discontinuities. We consider all these as our future work.

References

- [1] Stephen T. Barnard and Martin A. Fischler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, 1982.
- [2] Stan Birchfield and Carlo Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):401–406, 1998.
- [3] Stan Birchfield and Carlo Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *Proc. of IEEE International Conference on Computer Vision*, 1999.
- [4] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [5] Myron Z. Brown, Dariusz Bruschka, and Gregory D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.
- [6] Frederic Devernay and Olivier D. Faugeras. Computing differential properties of 3-d shapes from stereoscopic images without 3-d models.

- In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1994.
- [7] Umesh R. Dhond and J. K. Aggarwal. Structure from stereo – a review. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.
 - [8] Manfredo P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Inc., 1976.
 - [9] Olivier Faugeras. *Three-Dimensional Computer Vision*. The MIT Press, 1993.
 - [10] T. Frohlinghaus and J. M. Buhmann. Regularizing phase-based stereo. In *Proc. of ICPR*, 1996.
 - [11] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2000.
 - [12] William Hoff and Narendra Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):121–136, 1989.
 - [13] David G. Jones and Jitendra Malik. Determining three-dimensional shape from orientation and spatial frequency disparities. In *Proc. of European Conference on Computer Vision*, 1992.
 - [14] Takeo Kanade and Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994.
 - [15] Jan J. Koenderink. *Solid Shape*. The MIT Press, 1990.
 - [16] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *Proc. of IEEE International Conference on Computer Vision*, 2001.
 - [17] Michael H. Lin and Carlo Tomasi. Surfaces with occlusions from layered stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1073–1078, 2004.
 - [18] David Marr and Tomaso Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

- [19] David Marr and Tomaso Poggio. A computational theory of human stereo vision. *Proc. Royal Soc. London B*, 204:301–328, 1979.
- [20] Abhijit S. Ogale and Yiannis Aloimonos. Stereo correspondence with slanted surfaces: Critical implications of horizontal slant. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
- [21] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.
- [22] Peter T. Sander and Steven W. Zucker. Inferring surface trace and differential structure from 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):833–854, 1990.
- [23] Daniel Scharstein and Richard Szeliski. Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, 28(2):155–174, 1998.
- [24] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, 2002.
- [25] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.
- [26] Marshall F. Tappen and William T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *Proc. of IEEE International Conference on Computer Vision*, 2003.
- [27] Richard P. Wildes. Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):761–774, 1991.
- [28] Charles Zitnick and Takeo Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684, 2000.

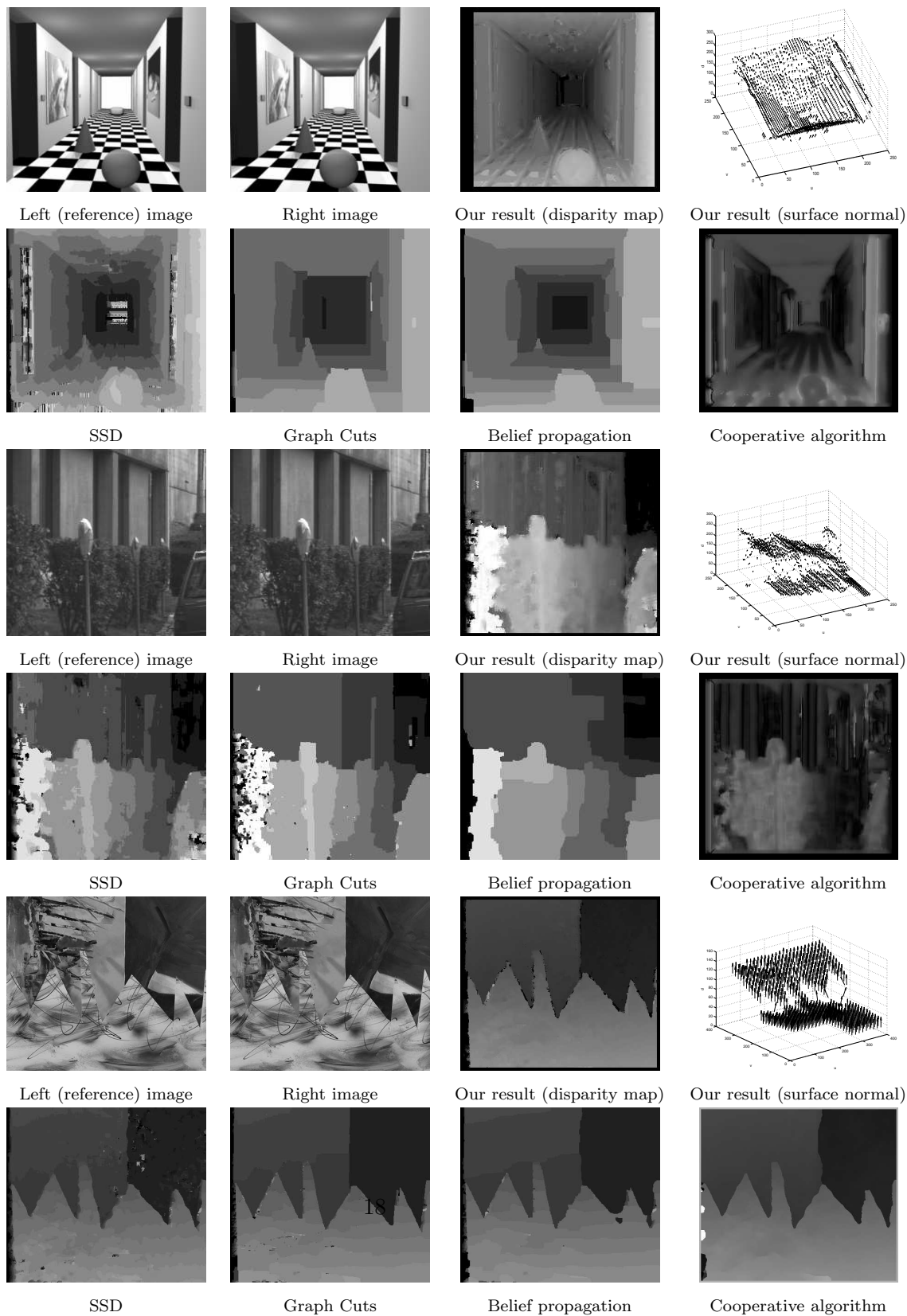


Figure 9: Input image pair, disparity map and surface orientation by our algorithm, disparity maps by traditional SSD, graph cuts, belief propagation, and cooperative algorithm, respectively. See text for details.