Differential Geometric Consistency Extends Stereo to Curved Surfaces

Gang Li and Steven W. Zucker

Department of Computer Science,
Yale University,
New Haven, CT 06520, USA
{gang.li, steven.zucker}@yale.edu

Abstract. Traditional stereo algorithms implicitly use the frontal parallel plane assumption when exploiting contextual information, since the smoothness prior biases towards constant disparity (depth) over a neighborhood. For curved surfaces these algorithms introduce systematic errors to the matching process. These errors are non-negligible for detailed geometric modeling of natural objects (e.g. a human face). We propose to use contextual information geometrically. In particular, we perform a differential geometric study of smooth surfaces and argue that geometric contextual information should be encoded in Cartan's moving frame model over local quadratic approximations of the smooth surfaces. The result enforces geometric consistency for both depth and surface normal. We develop a simple stereo algorithm to illustrate the importance of using such geometric contextual information and demonstrate its power on images of the human face.

1 Introduction

Viewing someone's face at about 1 meter provides a rich description of its surface characteristics. While two-view dense stereo vision has achieved remarkable success [22], this success has been limited to objects with restricted geometry. Assuming a rectified stereo pair [7, 8], many stereo algorithms either explicitly or implicitly exploit the *frontal parallel plane assumption*, which assumes position disparity (or depth) is constant (with respect to the rectified stereo pair) over a region under consideration. We seek to move beyond this assumption and to develop richer descriptions of smooth surfaces curving in space (Fig.1).

Traditional area-based stereo algorithms (e.g. SSD) explicitly use the frontal parallel plane assumption by comparing a window of the same size and shape in the left and right images for the similarity measure. Results often exhibit a "staircase" effect for slanted or curved surfaces. To address this problem, [9] uses a parameterized planar or quadratic patch fit; [10] uses variable window size (but fixed shape); [5] uses disparity derivatives to deform the matching window; [2, 16] model each segmented region as a slanted or curved surface while segmentation and correspondence are iteratively performed; [25] seeks correspondence for image regions instead of individual pixels; [23] uses a PDE-based approach for wide baseline dense stereo.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part III, LNCS 3953, pp. 44–57, 2006.

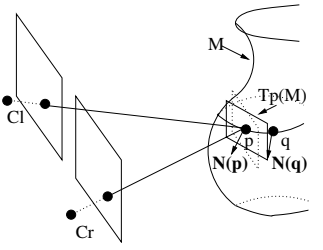


Fig. 1. Given a regular surface $M \subset \mathbb{R}^3$, the tangent plane $T_p(M)$ (in solid lines) and surface normal $\mathbf{N}(\mathbf{p})$ at a point $\mathbf{p}$ are well defined. Traditional stereo algorithms using contextual information (between $\mathbf{p}$ and a neighboring point $\mathbf{q}$) use the frontal parallel plane (in dotted lines) as the local surface model at $\mathbf{p}$. This implicit use of the frontal parallel plane assumption will result in a bias towards frontal parallel plane reconstruction, which is fundamentally flawed for curved surfaces. The correct use of contextual information should encode the change of both position and surface normal (at $\mathbf{p}$ and $\mathbf{q}$) on the surface. A differential geometric account of such contextual information is our contribution in this paper.

Since point-wise geometric constraints (e.g. epipolar constraint) and similarity measure (e.g. SSD) cannot always resolve matching ambiguities, it is natural to explore contextual information, i.e. requiring neighboring matching pairs to be "consistent". However such consistency often implicitly uses the frontal parallel plane assumption: [17] uses a local excitatory neighborhood of the same disparity level to support the current matching pair, while [27] refines local support as the sum of all match values within a 3D local support volume. [1] represents surface depth, orientation, boundaries, and creases as random variables. In the nonlinear diffusion algorithm [21], local support at different disparity hypotheses is diffused iteratively, and the amount of diffusion is controlled by the quality of the disparity estimate. In [3] a smoothness term over neighboring pixels is introduced in an energy functional minimized by graph cuts. In [24, 26] messages (similarity measures weighted by gaussian smoothed disparity differences) are passed between nearby matching pairs in a Markov network by belief propagation. These algorithms implicitly use the frontal parallel plane assumption because the neighboring matching pairs interact in a way such that the frontal parallel plane solution is preferred (Fig.1).

1.1 Our Approach

Systematic errors will be introduced by both the explicit and the implicit use of the frontal parallel plane assumption (see experiment section for details). Although the explicit use of this assumption has been addressed (e.g.[5]), the implicit use in the contextual inference stage has received little attention. To move beyond this assumption and overcome such errors, locally it implies that the tangent plane $T_p(M)$ deviates from the frontal parallel plane. Our geometric observation then arises in several forms: (i) varying the shape of matching patches in the left/right images; (ii) interpolating integer coordinates; (iii) relating disparity

derivatives to surface differential geometric properties; and (iv) (at least) surface normal consistency must be enforced over overlapping neighborhoods. Devernay and Faugeras [5] provide a solution to (i) and (ii). To take full advantage of (iii) and (iv), which follow directly from differential geometry, we exploit (Cartan) transport to combine geometric information from different surface normals in a neighborhood around a putative matching point. We describe geometric consistency between nearby matching pairs using both depth (position disparity) and surface normal, thus showing that contextual information behaves like an *extra geometric constraint* for stereo correspondence. To our knowledge this is the first time such geometric contextual constraints among nearby matching pairs have been used explicitly in stereo vision.

2 Background

2.1 Initial Local Information from Deformed Matching Window

Assuming a rectified stereo pair [7, 8], traditional area based methods compare a small window (e.g. 11x11) centered at (u, v) in the left image with a window of the same size and shape at $(u-d, v)$ in the right image using a similarity measure such as SSD, and select a disparity estimate d based on such a measure. When the scene within the window satisfies the frontal parallel plane assumption the above method is valid. But for slanted or curved 3D surfaces such a formulation is incorrect. Consider a small image window of a curved surface: 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 $(u+\delta u, v+\delta v)$ in the left image is $(u + \delta u - d - \frac{\partial d}{\partial u}\delta u - \frac{\partial d}{\partial v}\delta v, v + \delta v)$ 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 are a small step size in each direction.

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

$$\arg \min_{\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}} \sum_{(u+\delta u, v+\delta v) \in \mathcal{N}_{uv}} (I_l(u + \delta u, v + \delta v) - \hat{I}_r(u + \delta u - d - \frac{\partial d}{\partial u}\delta u - \frac{\partial d}{\partial v}\delta v, v + \delta v))^2 \quad (1)$$

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. We use the direction set method [19], 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). They could also be obtained by enumerating different combinations of these parameters if they are properly quantized, and selecting the set that minimizes deformed window SSD. In [5] such a deformed window was

also used. Our contribution is to relate the deformation to surface orientation and to impose geometric consistency over overlapping neighborhoods by using surface orientation, which provides extra geometric constraints for stereo correspondence. We now start to develop our contribution.

2.2 Problem Formulation in Euclidean Space

Using the left (reference) camera coordinate system as the world coordinate system, the depth z at pixel (u, v) is $z(u, v) = \frac{\alpha b}{d(u,v)}$, where $d(u, v)$ is the (positional) disparity at (u, v), α is the focal lengh (in pixels), and b is the stereo baseline. We assume such a model with known α and b; i.e. the pin-hole cameras are calibrated and the stereo pair is rectified. To work in $\mathbb{R}^3$ (not in disparity space) we need the partial derivatives of depth z with respective to x and y, respectively:

$$z_x = \frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} = -\frac{\alpha b}{d^2}\frac{\partial d}{\partial u}\frac{\alpha}{f}, \quad z_y = \frac{\partial z}{\partial y} = \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} = -\frac{\alpha b}{d^2}\frac{\partial d}{\partial v}\frac{\alpha}{f} \tag{2}$$

where $\frac{\partial u}{\partial x}$ and $\frac{\partial v}{\partial y}$ are constants determined by quantization of the image sensor, i.e. the focal length in pixels (α) and in physical unit (f) (we assume the same values in both x and y directions). A typical value is $1200 pixels/12mm = 100$.

Remark 1. Further taking derivatives shows that disparity derivatives (e.g. $\frac{\partial^n d}{\partial u^n}$) are (roughly) related to scaled physical derivatives (e.g. $\frac{\partial^n z}{\partial x^n}$) by $(\frac{f}{\alpha})^n$ (e.g.$(1/100)^n$). □

For physical objects with meaningful higher-order derivative information (e.g. $\frac{\partial^2 z}{\partial x^2}$, normal curvature in x direction), numerically it is difficult to manipulate the related higher-order disparity derivates (e.g.$\frac{\partial^2 d}{\partial u^2}$) in disparity space with image coordinates. This was a problem in [5]. We avoid working in such disparity space and chose to work in Euclidean space $\mathbb{R}^3$. First order derivatives $\{z_x, z_y\}$ are computed using the above equations after getting $\frac{\partial d}{\partial u}$ and $\frac{\partial d}{\partial v}$ from the inital deformed window SSD . A fitting process over a 3D neighborhood yields $\{z_{xx}, z_{xy}, z_{yy}\}$.

Now, for every candidate match we have estimated its depth z (disparity d), first order derivatives z_x and z_y, second order derivatives z_{xx}, z_{xy}, and z_{yy}, based on a local deformed SSD window (followed by fitting). Next we will show what it means for a candidate match with these properties to be *geometrically* consistent with its neighbors. This will enable us to eliminate inappropriate candidate matches and to refine the geometric estimates.

3 Differential Geometry of Smooth Surfaces

Assume the object under view is bounded by a smooth surface that can be described (locally) as a Monge patch. We briefly review the relevant differential geometry following [6, 4, 18] for notation. In particular, M is a regular surface in $\mathbb{R}^3$, $\mathbf{p}$ and $\mathbf{q}$ denote surface points in $\mathbb{R}^3$, $\mathbf{v} \in \mathbb{R}^3$ denotes a tangent vector in the tangent plane $T_p(M)$, $\mathbf{X}$ the position vector field (i.e. $\mathbf{X}(\mathbf{p}) = \mathbf{p}$), and $\mathbf{N}$ the unit surface normal vector field.

3.1 Surface Differential Properties

For a regular surface $M \subset \mathbb{R}^3$ the surface normal (or equivalently the tangent plane) changes as we move over it. This geometric property has been studied as the second fundamental form and the shape operator. They both encode such geometric information. In particular: The *shape operator* $S_p(\mathbf{v})$ encodes the shape of a surface M by measuring how the surface normal $\mathbf{N}$ changes as one moves in various directions from point $\mathbf{p}$ in the tangent plane $T_p(M)$. It is defined as [18]:

$$S_p(\mathbf{v}) = -\nabla_v \mathbf{N} \tag{3}$$

where $\nabla_v \mathbf{N}$ denotes the covariant derivative of the unit normal vector field $\mathbf{N}$ with respect to the tangent vector $\mathbf{v} \in T_p(M)$, i.e. the initial rate of change of $\mathbf{N}(\mathbf{p})$ as $\mathbf{p}$ moves in the $\mathbf{v}$ direction. In other words, it gives an infinitesimal description of the way surface M is curving in $\mathbb{R}^3$.

The *second fundamental form* II_p is defined in $T_p(M)$ as the quadratic form $II_p = -\langle d\mathbf{N}_p(\mathbf{v}), \mathbf{v}\rangle$, where $d\mathbf{N}_p$ is the differential of the Gauss map [6].

For tangent vectors $\mathbf{v}$ and $\mathbf{w}$ (both in $T_p(M)$) these two concepts are related by $II_p(\mathbf{v}, \mathbf{w}) = S_p(\mathbf{v}) \cdot \mathbf{w}$.

3.2 Second Fundamental Form for Monge Patch

We now switch to a convenient form for computation. In the (reference) camera coordinate system we can represent the surface as a Monge patch, $\mathbf{r}(x, y) = (x, y, z(x, y))$. Taking partial derivatives:

$$\mathbf{r}_x = (1, 0, z_x), \quad \mathbf{r}_y = (0, 1, z_y)$$
$$\mathbf{r}_{xx} = (0, 0, z_{xx}), \quad \mathbf{r}_{xy} = \mathbf{r}_{yx} = (0, 0, z_{xy}), \quad \mathbf{r}_{yy} = (0, 0, z_{yy})$$

Unit surface normal $\mathbf{N} = \frac{\mathbf{r}_x \wedge \mathbf{r}_y}{\|\mathbf{r}_x \wedge \mathbf{r}_y\|} = \frac{(-z_x, -z_y, 1)}{\sqrt{1+z_x^2+z_y^2}}$, where $\wedge$ is vector cross product.

The matrices of the first fundamental form I and the second fundamental form II are:

$$I: \quad \begin{bmatrix} \mathbf{r}_x \cdot \mathbf{r}_x & \mathbf{r}_x \cdot \mathbf{r}_y \\ \mathbf{r}_y \cdot \mathbf{r}_x & \mathbf{r}_y \cdot \mathbf{r}_y \end{bmatrix} = \begin{bmatrix} 1 + z_x^2 & z_x z_y \\ z_x z_y & 1 + z_y^2 \end{bmatrix}$$
$$II: \quad \begin{bmatrix} \mathbf{r}_{xx} \cdot \mathbf{N} & \mathbf{r}_{xy} \cdot \mathbf{N} \\ \mathbf{r}_{yx} \cdot \mathbf{N} & \mathbf{r}_{yy} \cdot \mathbf{N} \end{bmatrix} = \frac{1}{\sqrt{1 + z_x^2 + z_y^2}} \begin{bmatrix} z_{xx} & z_{xy} \\ z_{xy} & z_{yy} \end{bmatrix}$$

respectively. In the basis $\{\mathbf{r}_x, \mathbf{r}_y\}$, $d\mathbf{N}$ is given by the matrix $-I^{-1}II$, and the matrix of the shape operator S is $I^{-1}II$. This matrix is relative to the tangent vectors $\mathbf{r}_x$, $\mathbf{r}_y$ (as basis vectors) in the tangent plane $T_p(M)$ of M at $\mathbf{p}$. The matrix of the shape operator S is:

$$I^{-1}II: \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \frac{1}{(1 + z_x^2 + z_y^2)^{3/2}} \begin{bmatrix} 1 + z_y^2 & -z_x z_y \\ -z_x z_y & 1 + z_x^2 \end{bmatrix} \begin{bmatrix} z_{xx} & z_{xy} \\ z_{xy} & z_{yy} \end{bmatrix} \tag{4}$$
$$= \frac{1}{(1 + z_x^2 + z_y^2)^{3/2}} \begin{bmatrix} (1 + z_y^2) z_{xx} - z_x z_y z_{xy} & (1 + z_y^2) z_{xy} - z_x z_y z_{yy} \\ (1 + z_x^2) z_{xy} - z_x z_y z_{xx} & (1 + z_x^2) z_{yy} - z_x z_y z_{xy} \end{bmatrix}$$

Note that this matrix is not necessarily symmetric, unless $\{\mathbf{r}_x, \mathbf{r}_y\}$ is an orthonormal basis. Typical values are given in Section 5.

Remark 2. When the object is a planar surface ($z_{xx} = z_{xy} = z_{yy} = 0$), elements of matrices II and S are all zeros. Observe that S still encodes the geometric property for such planar surfaces. The special case of a frontal parallel plane ($z_x = z_y = 0$) is also encoded in S. □

Remark 3. At the occluding boundaries of the surface (when $\mathbf{N}$ is orthogonal to the line of sight), we can not represent the surface as a Monge patch. As a result matrices of I, II, and S are not well defined in above formulas. This is an implementation limitation but not a theoretical one. □

4 Differential Geometric Consistency for Curved Surfaces

The intuition behind geometric consistency is that the measurement (position, normal) information at each point, when transported along the surface to neighboring points (as described previously), should agree with the measurements at those points. We now develop this intuition. At a given point $\mathbf{p}$ we would like to study how the position and surface normal change as we move in various directions in the tangent plane (Fig. 2(a)). $\mathbf{X}$ denotes the position vector field (thus $\mathbf{X}(\mathbf{p}) = \mathbf{p}$) and $\mathbf{N}$ the unit surface normal vector field. We require explicit formulas for their change as we move along $\mathbf{v}$ in the tangent plane $T_p(M)$ (i.e.$\nabla_v \mathbf{X}$ and $\nabla_v \mathbf{N}$).

Proposition 1. *Let* $\mathbf{X}$ *be the special vector field* $\Sigma_{i=1}^{3} x_i \mathbf{E}_i$*, where* x_1*,* x_2*, and* x_3 *are the Euclidian coordinate functions of* $\mathbb{R}^3$*. Then* $\nabla_v \mathbf{X} = \mathbf{v}$ *for every tangent vector* $\mathbf{v}$*.*

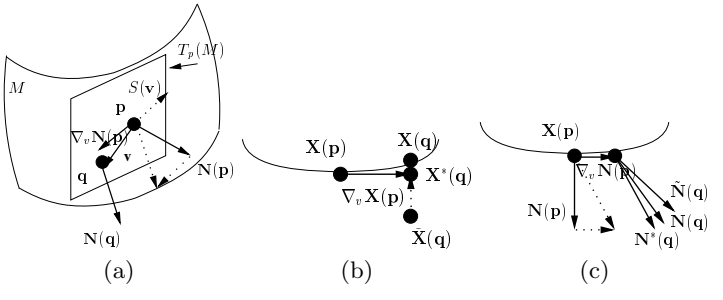


Fig. 2. (a) Shows the change of position vector field $\mathbf{X}$ and the unit surface normal vector field $\mathbf{N}$ by moving $\mathbf{v}$ in the tangent plane. (b) The predicted position $\mathbf{X}^*(\mathbf{q})$ in the neighborhood can be obtained from $\mathbf{X}(\mathbf{p})$ and $\nabla_v \mathbf{X}(\mathbf{p})$. Geometric consistency in position is determined by comparing $\mathbf{X}^*(\mathbf{q})$ with the true measurements $\mathbf{X}(\mathbf{q})$ in the neighborhood. Also shown is a less consistent one $\tilde{\mathbf{X}}(\mathbf{q})$. (c) The predicted surface normal $\mathbf{N}^*(\mathbf{q})$ in the neighborhood can be obtained from $\mathbf{N}(\mathbf{p})$ and $\nabla_v \mathbf{N}(\mathbf{p})$. Geometric consistency in orientation is determined by comparing $\mathbf{N}^*(\mathbf{q})$ with the true measurements $\mathbf{N}(\mathbf{q})$ in the neighborhood. Also shown is a less consistent one $\tilde{\mathbf{N}}(\mathbf{q})$.

Proof: Rewrite $\nabla_v \mathbf{X}$ according to its definition and express it as the sum of directional derivatives. We have $\nabla_v \mathbf{X} = \Sigma_{i=1}^3 \mathbf{v}[x_i]\mathbf{E}_i(\mathbf{p})$. Further expand the directional derivative part by $\mathbf{v}_p[f] = \Sigma_{j=1}^3 v_j \frac{\partial f}{\partial x_j}(\mathbf{p})$; we have $\Sigma_{i=1}^3 \mathbf{v}[x_i]\mathbf{E}_i(\mathbf{p}) = \Sigma_{i=1}^3 \Sigma_{j=1}^3 v_j \frac{\partial x_i}{\partial x_j}(\mathbf{p})\mathbf{E}_i = \Sigma_{i=1}^3 v_i \mathbf{E}_i = \mathbf{v}$. □

From point $\mathbf{p}$ on M, if we move along $\mathbf{v}$ in the tangent plane, then to first order approximation the new position is:

$$\mathbf{X}^*(\mathbf{q}) = \mathbf{X}(\mathbf{p}) + \nabla_v \mathbf{X}(\mathbf{p}) = \mathbf{X}(\mathbf{p}) + \mathbf{v} \tag{5}$$

Interpret this computed position $\mathbf{X}^*(\mathbf{q})$ as the "transported" geometric information to a neighboring position $\mathbf{q}$ along the surface (from the measurements) at $\mathbf{p}$. Since direct measurements are also available at $\mathbf{q}$ (denoted $\mathbf{X}(\mathbf{q})$), the discrepancy between $\mathbf{X}^*(\mathbf{q})$ and $\mathbf{X}(\mathbf{q})$ can be used to measure the geometric consistency between nearby candidate matching points $\mathbf{p}$ and $\mathbf{q}$. Fig. 2(b) illustrates this point using two (possible) measured points $\mathbf{X}(\mathbf{q})$. Clearly the one on the same surface as $\mathbf{p}$ should be very close to the transported position, i.e. $\mathbf{X}^*(\mathbf{q})$.

And similarly for surface normal. Given the shape operator $S_p(\mathbf{v}) = -\nabla_v \mathbf{N}$ of M at $\mathbf{p}$, the change of surface normal $\mathbf{N}$ is characterized by the covariant derivative $\nabla_v \mathbf{N}$ for any $\mathbf{v}$ in the tangent plane $T_p(M)$. To emphasize its importance we show it as a proposition.

Proposition 2. *Let* $\mathbf{N}$ *be the unit normal vector field. Then for every tangent vector* $\mathbf{v} = \delta t_1 \mathbf{r}_x + \delta t_2 \mathbf{r}_y$ *in the tangent plane,* $\nabla_v \mathbf{N}$ *is given by:*

$$\begin{aligned} \nabla_v \mathbf{N} = \nabla_{\delta t_1 \mathbf{r}_x + \delta t_2 \mathbf{r}_y} \mathbf{N} &= (\delta t_1 \nabla_{r_x} \mathbf{N} + \delta t_2 \nabla_{r_y} \mathbf{N}) \\ &= -(\delta t_1 a_{11} + \delta t_2 a_{12})\mathbf{r}_x - (\delta t_1 a_{21} + \delta t_2 a_{22})\mathbf{r}_y \end{aligned} \tag{6}$$

where a_{ij}*'s are given in equation* (4).

Proof: This follows from the linearity of covariant derivative and the calculations in the previous section. □

Again, if we move along $\mathbf{v}$ in the tangent plane from $\mathbf{p}$, then the surface normal at the new position $\mathbf{N}^*(\mathbf{q})$ can be computed from $\mathbf{N}(\mathbf{p})$ and $\nabla_v \mathbf{N}(\mathbf{p})$. To first order approximation the new normal is:

$$\mathbf{N}^*(\mathbf{q}) = \mathbf{N}(\mathbf{p}) + \nabla_v \mathbf{N}(\mathbf{p}) = \mathbf{N}(\mathbf{p}) - (\delta t_1 a_{11} + \delta t_2 a_{12})\mathbf{r}_x - (\delta t_1 a_{21} + \delta t_2 a_{22})\mathbf{r}_y \tag{7}$$

After nomalization this computed unit surface normal $\mathbf{N}^*(\mathbf{q})$ is the "transported" geometric information along the surface (from the measurements) at $\mathbf{p}$. Since direct measurements are also available at $\mathbf{q}$ to obtain $\mathbf{N}(\mathbf{q})$, the discrepancy between $\mathbf{N}^*(\mathbf{q})$ and $\mathbf{N}(\mathbf{q})$ can be used to measure the geometric consistency between nearby candidate matching points p and q. Fig. 2(c) illustrates this point by showing the transported normal $\mathbf{N}^*(\mathbf{q})$, which should agree with the geometrically consistent normal at $\mathbf{q}$. Observe that for planar surfaces (all zeros for matrix II) this implies constant surface normal (e.g. see Remark 2), which was discussed in [14].

The principle of *geometric consistency* between two neighboring points $\mathbf{p}$ and $\mathbf{q}$ holds that, based on the geometric information at $\mathbf{p}$ (i.e. $\mathbf{X}(\mathbf{p})$, $\mathbf{N}(\mathbf{p})$, $\nabla_v \mathbf{X}(\mathbf{p})$, and $\nabla_v \mathbf{N}(\mathbf{p})$), the transported (computed) geometric information at $\mathbf{q}$ (i.e. $\mathbf{X}^*(\mathbf{q})$ and $\mathbf{N}^*(\mathbf{q})$) should agree with the measurements at $\mathbf{q}$ (i.e. $\mathbf{X}(\mathbf{q})$ and $\mathbf{N}(\mathbf{q})$) if it is on the same surface as $\mathbf{p}$.

4.1 Geometric Contextual Information for Stereo

Our geometric way of using contextual information is in the Cartan moving frame model [6, 11]. It specifies how adpated frame fields change when they are transported along an object, and is concisely encoded in the connection equations. This model can be used to integrate local geometric information with geometric information in the neighborhood. Given candidate matches (obtained from initial local measurements), now we can impose the smoothness constraint in the neighborhood based on the geometric study just performed. Note that this is our unique construction in using contextual information geometrically. Both the position and the normal should be used in defining such geometric consistency.

Definition. The *geometric compatibility* between candidate match points $\mathbf{p}$ and $\mathbf{q}$ is:

$$r_{pq} = \frac{1}{2}((1 - \frac{1}{m}\|\mathbf{X}^*(\mathbf{q}) - \mathbf{X}(\mathbf{q})\|) + |\mathbf{N}^*(\mathbf{q}) \cdot \mathbf{N}(\mathbf{q})|) \tag{8}$$

where m is a normalization constant related to the neighborhood size.

Remark 4. $0 \leq r_{pq} \leq 1$, with $r_{pq} = 1$ for consistent $\mathbf{p}$ and $\mathbf{q}$, while $r_{pq} = 0$ for inconsistent $\mathbf{p}$ and $\mathbf{q}$. We use a mixed norm in defining such geometric consistency, but other formulas are also possible. □

The geometric constraint (eqn. (8)) can be used in the cooperative framework. For a candidate match point $\mathbf{p}$ (hypothesis), we initialize its support s_p^0 according to its deformed window SSD (denoted by c_p) and iteratively update s_p by the geometric support it receives from its neighboring candidate matching point $\mathbf{q}$:

$$s_p^0 = 1 - \frac{c_p}{c} \tag{9}$$

$$s_p^{t+1} = \frac{\sum_{q \in \mathcal{N}_p} r_{pq} s_q^t}{\sum_{q \in \mathcal{N}_p} s_q^t} \tag{10}$$

with c a normalization factor, $\mathcal{N}_p$ denotes the neighbors of $\mathbf{p}$ (in our experiments we use a 21x21x7 (u, v, d) region). Note that here we use subscript to denote the measure with respect to candidate match point $\mathbf{p}$ (not the partial derivatives!). 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. We also experimented with a two label relaxation labeling algorithm [15, 13], and observed similar results. According to the taxonomy [22], such an iterative 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 [17, 27], which iteratively performs local computations and uses nonlinear operations resulting in a final effect similar to global optimization.

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-squares sense) unit normal at $\mathbf{p}$ is updated as [20]: $\mathbf{N}_p^{t+1} = (\sum \mathbf{N}_q^t)/\|\sum \mathbf{N}_q^t\|$, with $\mathbf{q}$ points in the neighborhood of $\mathbf{p}$ and within a normal threshold (e.g. $\pi/4$).

4.2 Stereo Algorithm

A simple algorithm illustrates how such geometric contextual information could be used.

(1) Use deformed window SSD to get the intial candidate matches. We first use a traditional SSD (15x15 window) to get integer disparity values at each (u, v) and only keep the top $\delta\%$ (we use 3 non-immediate neighboring ones) as the initial guesses. Then as explained in Section 2, for each disparity guess at every (u, v), we obtain $\{d, \frac{\partial d}{\partial u}, \frac{\partial d}{\partial v}\}$ (interpolated in the continuous domain) that minimizes deformed window SSD in equation (1). Several local minima could exist at each pixel (u, v). Geometric contextual information will be explored in the next few steps.

(2) Compute differential properties (e.g. surface normal $\mathbf{N}$, shape operator S) for every candidate match point $\mathbf{p}$ (Section 3).

(3) Compute the initial support s_p^0 for each candidate match point $\mathbf{p}$ by equation (9), which encodes the similarity measure based on deformed window SSD.

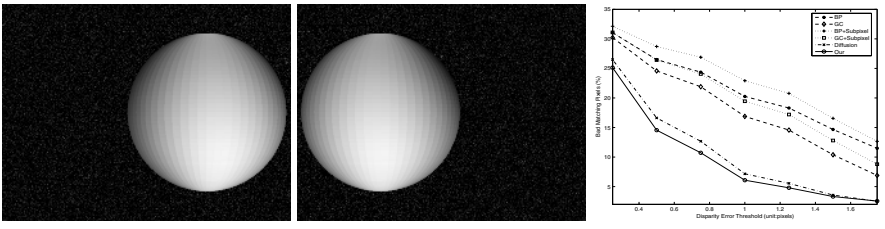
(4) Iteratively update the geometric support s_p at every $\mathbf{p}$ by equation (10) until it converges (in practice we run a preset number (e.g. 8) of iterations). To get the geometric compatibilities between nearby putative matches r_{pq} (eqn. (8)): first project $(\mathbf{q} - \mathbf{p})$ onto the tangent plane of $\mathbf{p}$, resulting in the displacement vector $\mathbf{v} \in T_p(M)$; then compute the predicted position and normal according to eqn. (5)(7); and finally use eqn. (8). Also update surface normal $\mathbf{N}_p$ at $\mathbf{p}$ based on the normals of neighbors, to reduce the effect of local noisy measurements.

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

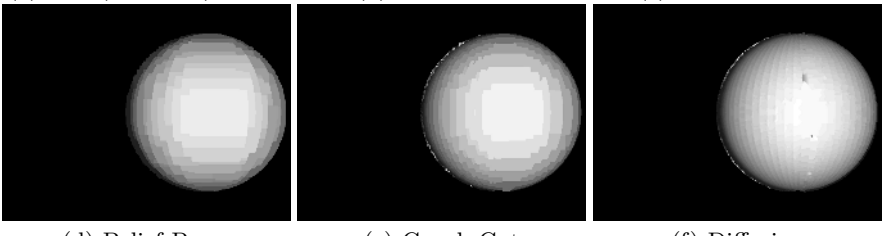
Observe that steps (2)-(5) are the unique geometric content of our algorithm.

5 Experimental Results

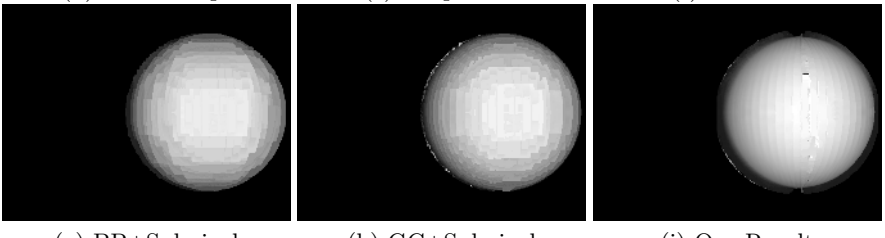
Fig. 3 provides a comparison between algorithms guided by the frontal parallel plane assumption for contextual interaction (e.g., graph cuts [3, 12], belief propagation [24, 26]) and those designed for more general surfaces (e.g., diffusion [21] and our algorithm). As expected, the first group yields "scalloped" surfaces broken into piecewise frontal parallel planar patches, even with subpixel interpolation (by parabola fitting of the costs of a 15x15 SSD window), while the second group follows the surface more robustly.



(a) Left (reference) image (b) Right image (c) Error statistics



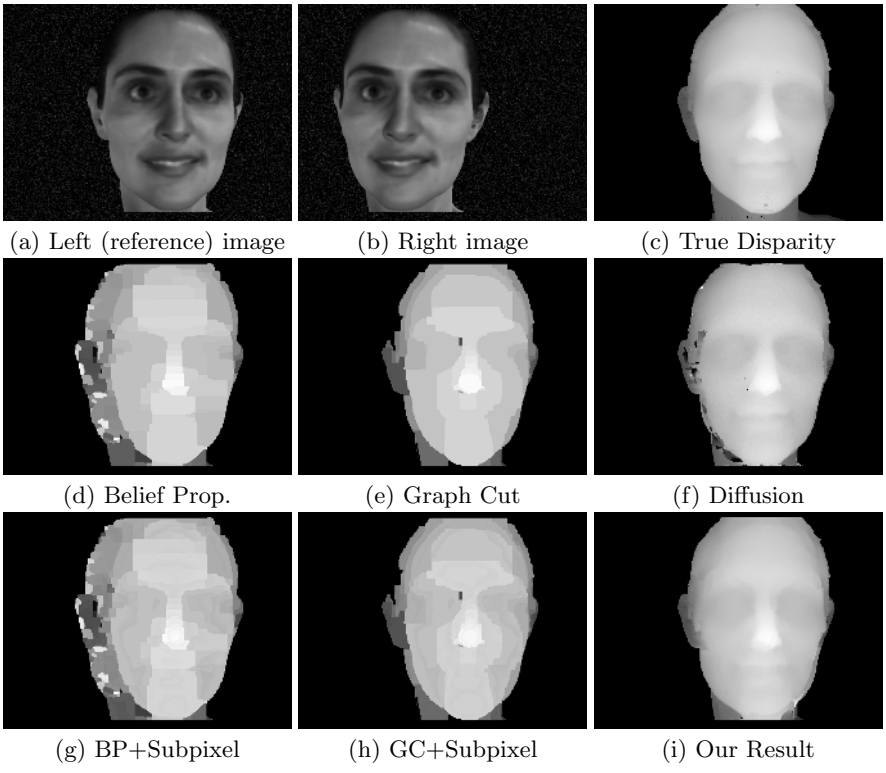
(d) Belief Prop. (e) Graph Cut (f) Diffusion



(g) BP+Subpixel (h) GC+Subpixel (i) Our Result

Fig. 3. Synthetic sphere example separates the performance of algorithms with the frontal parallel plane assumption from those designed for smooth surfaces. (c) shows the percentage of bad matching pixels (occluded region not counted) using the taxonomy package [22] at 7 different thresholds ranging from 0.25–1.75 pixels. Performance for the diffusion algorithm and our approach were similar for this spherical surface. Other algorithms were obtained from the stereo package provided by Scharstein and Szeliski [22] for nonlinear diffustion (the membrane model) [21], and max-product ([26]) for belief propagation; The α-expansion algorithm [12] for graph cuts. The stereo pair was rendered with 152mm baseline and focal length 1303 pixels (obtained from real calibration data). Image size is 640x480 pixels, disparity range 41 pixels; Sphere has radius 100mm and center at 750mm distance.

Remark 5. To illustrate the numerical stability of our computation we pick a typical point $(u, v) = (526, 240)$, i.e. $(x, y) = (1.90, 0.0)$. From the ground truth we obtain $z = 687.29mm$, and $(z_x, z_y) = (1.242, 0.0)$. In our result we get $(z_x, z_y) = (1.239, 0.0)$ and $(z_{xx}, z_{xy}, z_{yy}) = (0.0406, 0, 0.0159)$, further computation shows matrix I is $\begin{bmatrix} 2.5432 & 0.0000 \\ 0.0000 & 1.0000 \end{bmatrix}$, matrix II is $\begin{bmatrix} 0.0254 & 0.0000 \\ 0.0000 & 0.0100 \end{bmatrix}$, and the matrix of the shape operator is: $\begin{bmatrix} 0.0100 & 0.0000 \\ 0.0000 & 0.0100 \end{bmatrix}$. These numbers are clearly meaningful numerically; however, by Remark 1, note that previous attempts



(a) Left (reference) image (b) Right image (c) True Disparity

(d) Belief Prop. (e) Graph Cut (f) Diffusion

(g) BP+Subpixel (h) GC+Subpixel (i) Our Result

Fig. 4. Human face example. Shown are results from other algorithms (as in previous caption), and our result. While the scalloping remains present in belief propagation and graph cuts, again diffusion and our algorithm appear smoothest. A detailed analysis of the statistical data reveals the difference (in the next Fig.). Ground truth data from $Cyberware^{TM}$ laser scanner dataset. Timing: Our algorithm takes 982.69 seconds on a Intel Xeon 2.4GHz CPU; accelerated belief propogation takes 1977.57 sec.; graph cuts takes 221.28 sec. and diffusion 59.82 sec.

(e.g.,[5]) at surface computations in (u, v)-space would have to multiply the above entries by (about) 10^{-4} for second order properties, thus placing them right at the limit of measurable quantities even for this idealized example. □

The second set of examples illustrates the difference between our approach and diffusion. Faces have rich surface geometry needed to support graphical rendering and different types of recognition, and the 3D details matter. Ground truth data (3D geometry and texture map) were obtained from the $Cyberware^{TM}$ laser scanner dataset. The true disparity map is then computed. The stereo pair has a baseline of 6cm and focal length 1143 pixels. The human head ranges from 26.5cm to 53.5cm in front of the camera. The original image size is 1024x768 pixels but is then subsampled to 512x384 pixels with a disparity range 66 pixels. Results are in Fig. 4; once again, the diffusion algorithm is closest to ours but differences are emerging.

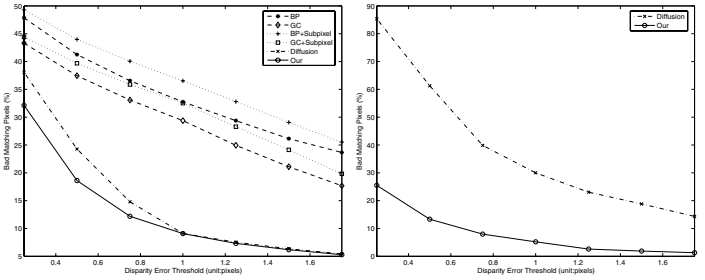


Fig. 5. Error statistics of bad matching pixels: (LEFT) Whole image. (RIGHT) A 30x30 pixel region around the side of the nose. Notice in particular how the statistics diverge for the nose region, where surface normal is changing rapidly. It is in places such as this that our algorithm noticeably outperforms diffusion.

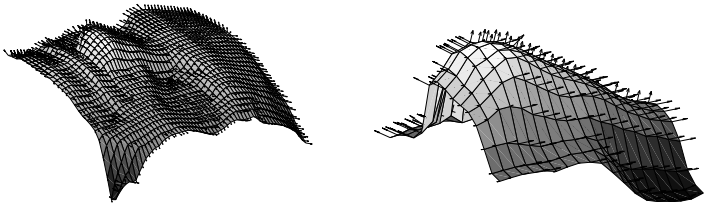
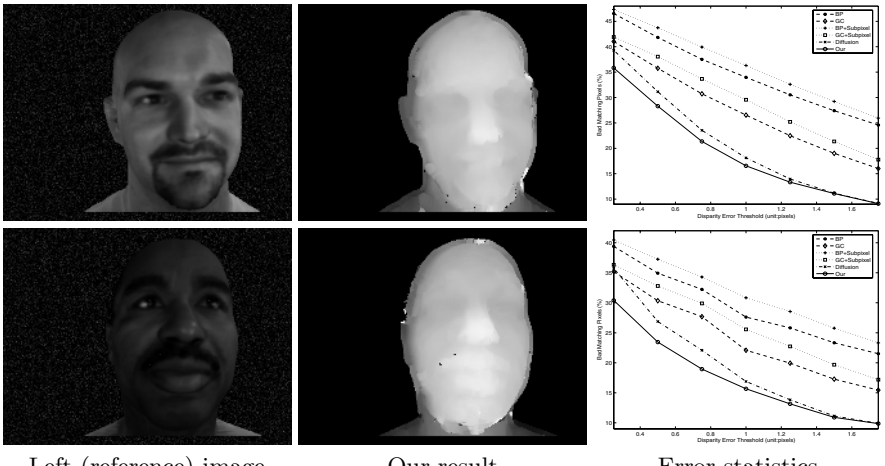


Fig. 6. Reconstruction results. (LEFT) Reconstructed surface normal. (RIGHT) Zoom in of nose region. For display purpose surface normal and depth are subsampled to one in five pixels in both x and y directions.



Left (reference) image Our result Error statistics

Fig. 7. More results on face stereo pair with ground truth

The membrane model underlying the diffusion algorithm applies uniform smoothing in proportion to iteration number. Our algorithm, by contrast, implements regularization in proportion to surface geometry, as a more detailed analysis indicates. The error statistics are shown in Fig. 5. While our algorithm differs from diffusion when averaged across the entire image, it differs *sharply* in those regions where the surface normal is rapidly changing (Fig. 5(RIGHT)). Diffusion oversmoothes these regions to get the smoothing right in larger, less varied regions. Our reconstructed surface normals are shown in Fig. 6; note how exquisitely the normal follows the nose in the blow-up.

Several other stereo face pairs are basically the same; see Fig. 7. Due to space limits we only report our result and error statistics of bad matching pixels; again, zooms on rapidly curving regions are informative.

6 Conclusion

We introduced the principle of geometric consistency to stereo, which holds that local observations of spatial disparity and surface shape should agree with neighboring observations; and that agreement between these neighboring observations can be implemented with a transport operation. In effect, nearby normals can be transported along (estimates of) the surface to be compared with directly measured normals. We provided direct calculations of these transport operations, and demonstrated their efficacy with a simple stereo algorithm. The geometric compatibility function could also be used in more powerful inference frameworks [26, 3], or developed into a richer probabilistic form.

Several limitations remain, though. Occlusion is not considered currently, nor the object boundaries, which provide information about depth discontinuities. The geometry underlying these will be studied in our next paper.

References

1. P. N. Belhumeur. A bayesian approach to binocular stereopsis. *IJCV*, 19(3):237–262, 1996.
2. S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *Proc. ICCV*, 1999.
3. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*, 23(11):1222–1239, 2001.
4. R. Cipolla and P. Giblin. *Visual Motion of Curves and Surfaces.* Cambridge Univ. Press, 2000.
5. F. Devernay and O. D. Faugeras. Computing differential properties of 3-d shapes from stereoscopic images without 3-d models. In *Proc. CVPR*, 1994.
6. M. P. do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall, Inc., 1976.
7. O. Faugeras. *Three-Dimensional Computer Vision.* The MIT Press, 1993.
8. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2000.

9. W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *IEEE Trans. on PAMI*, 11(2):121–136, 1989.
10. T. Kanade and M. Okutomi. A stereo maching algorithm with an adaptive window: Theory and experiment. *IEEE Trans. on PAMI*, 16(9):920–932, 1994.
11. J. J. Koenderink. *Solid Shape*. The MIT Press, 1990.
12. V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Proc. ICCV*, 2001.
13. G. Li and S. W. Zucker. A differential geometrical model for contour-based stereo correspondence. In *Proc. IEEE Workshop on Variational, Geometric, and Level Set Methods in Computer Vision (at ICCV'03)*, 2003.
14. G. Li and S. W. Zucker. Stereo for slanted surfaces: First order disparities and normal consistency. In *Proc. EMMCVPR, LNCS 3757*, 2005.
15. G. Li and S. W. Zucker. Contextual inference in contour-based stereo correspondence. *IJCV, in press*, 2006.
16. M. H. Lin and C. Tomasi. Surfaces with occlusions from layered stereo. *IEEE Trans. on PAMI*, 26(8):1073–1078, 2004.
17. D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.
18. B. O'Neill. *Elementary Differential Geometry*. Academic Press, 2nd edition, 1997.
19. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Reciples in C*. Cambridge University Press, second edition, 1992.
20. P. T. Sander and S. W. Zucker. Inferring surface trace and differential structure from 3-d images. *IEEE Trans. on PAMI*, 12(9):833–854, 1990.
21. D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. *IJCV*, 28(2):155–174, 1998.
22. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1/2/3):7–42, 2002.
23. C. Strecha, T. Tuytelaars, and L. V. Gool. Dense matching of multiple wide-baseline views. In *Proc. ICCV*, 2003.
24. J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo matching using belief propagation. *IEEE Trans. on PAMI*, 25(7):787–800, 2003.
25. H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *Proc. ICCV*, 2001.
26. M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *Proc. ICCV*, 2003.
27. C. Zitnick and T. Kanade. A cooperative algorithm for stereo mathching and occlusion detection. *IEEE Trans. on PAMI*, 22(7):675–684, 2000.