Shape-from-Shading under Perspective Projection

Ariel Tankus
School of Computer Science
Tel-Aviv University
Tel-Aviv, 69978
arielt@post.tau.ac.il

Nir Sochen
School of Mathematics
Tel-Aviv University
Tel-Aviv, 69978
sochen@post.tau.ac.il

Yehezkel Yeshurun
School of Computer Science
Tel-Aviv University
Tel-Aviv, 69978
hezy@post.tau.ac.il

Abstract

Shape-from-Shading (SfS) is a fundamental problem in Computer Vision. A very common assumption in this field is that image projection is orthographic. This paper re-examines the basis of SfS, the image irradiance equation, under a perspective projection assumption. The resultant equation does not depend on the depth function directly, but rather, on its natural logarithm. As such, it is invariant to scale changes of the depth function. A reconstruction method based on the perspective formula is then suggested; it is a modification of the Fast Marching method of Kimmel and Sethian. Following that, a comparison of the orthographic Fast Marching, perspective Fast Marching and the perspective algorithm of Prados & Faugeras on synthetic images is presented. The two perspective methods show better reconstruction results than the orthographic. The algorithm of Prados & Faugeras equates with the perspective Fast Marching. Following that, a comparison of the orthographic and perspective versions of the Fast Marching method on endoscopic images is introduced. The perspective algorithm outperformed the orthographic one. These findings suggest that the more realistic set of assumptions of perspective SfS improves reconstruction significantly with respect to orthographic SfS. The findings also provide evidence that perspective SfS can be used for real-life applications in fields such as endoscopy.

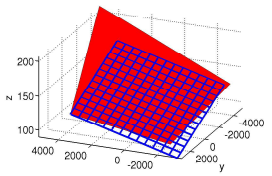
Keywords: perspective shape-from-shading, fast marching methods.

1. Introduction and Background

Recovery of Shape-from-Shading (SfS) is a fundamental problem in Computer Vision. The goal of SfS is to solve the image irradiance equation, which relates the reflectance map to image intensity, robustly. The task, however, appears to be nontrivial. This has caused most of the works in the field to add simplifying assumptions to the equation. Of particular importance is the common assumption that scene points are projected orthographically during the photographic process.

Many works in the field of Shape-from-Shading have followed the seminal works of Horn [3], [4] , [5], who initiated the subject in the 1970s, and assumed orthographic projection. Horn's book [6] reviews the early work on Shape-from-Shading (until 1989). Zhang et al. [32] surveys and classifies some of the works from the '90s and compares the performance of six of them (namely, minimization approaches: [34], [11]; propagation approach: [1]; local approach: [10]; linear approaches: [17], [28]). Kimmel & Bruckstein [8] classify image extrema and two kinds of saddle points and use these topological properties of the surface in a global Shape-from-Shading algorithm. In the current millennium Zhao & Chellappa [33] use symmetric Shape-from-Shading to develop a face recognition system which is illumination insensitive; they show the symmetric Shape-from-Shading algorithm has a unique solution. Kimmel & Sethian [9] proposed the Fast Marching method as an optimal algorithm for surface reconstruction. Their reconstructed surface is a viscosity solution of an Eikonal equation for the vertical light source case. Sethian [25] provides deep insight into Level Set and Fast Marching methods. Robles-Kelly & Hancock [20] use the Mumford-Shah functional to derive diffusion kernels that can be employed for Shape-from-Shading. Prados et al. [19] base their approach on the viscosity solution of a Hamilton-Jacobi equation. They extend existing proofs of existence and uniqueness to the general light source case and prove the convergence of their numerical scheme. Many more orthographic algorithms were suggested in the literature, but only a few can be described herein.

*This research has been supported in part by Tel-Aviv University fund, the Adams Super-Center for Brain Studies, the Israeli Ministry of Science, the ISF Center for Excellence in Applied Geometry, and the A.M.N. fund.



a. b.

Figure 1: Difference in reconstruction between perspective and orthographic SfS. **a.** Perspective projection of the filled quadrilateral is identical to orthographic projection of the meshy parallelogram. **b.** The image produced by both surfaces (light source direction: $\vec{L} = (0, 0.5, -1)$). Orthographic reconstruction of this image must produce a 3D parallelogram.

Despite all the work in this field, the comparative study [32], which dealt only with orthographic SfS, reaches the following conclusions: "1. All the SFS algorithms produce generally poor results when given synthetic data. 2. Results were even worse on real images, and 3. Results on synthetic data are not generally predictive of results on real data."

The few works that did employ the perspective projection have been too restrictive and have not addressed the general problem. Yamany et al. [30] and Seong et al. [23] assumed that distance variations between camera and surface could be ignored. Samaras & Metaxas [22] employed a deformable model for the SfS problem, so reconstruction took place in 3D space. Thus, during the deformation process, the image point onto which a 3D point was projected changed, and its new location should have been interpolated, resulting in a nonuniform sampling of the image.

Another approach to perspective SfS is piecewise planar modelling of the depth function (Lee & Kuo [12], Penna [16]). However, orthographic and perspective reflectance maps of a plane are identical, as Sect. 3.1 would show. Therefore, the two types of projection of a piecewise planar surface differ only at the edges, while fully agree at the interior of the faces.

Recently, Yuen et al. [31] proposed the use of perspective SfS with the Fast Marching method of Kimmel & Sethian [9]. This work approximated surface normals in 3D space using the neighboring pixels of the point under examination. Into these approximations the equations of perspective projection were substituted. This approach suffers two drawbacks. First, is describes a specific numerical approximation without reference to the theoretic problem (i.e., the image irradiance equation itself). Second and most importantly, neighboring pixels lie on a uniform grid (image space), while their 3D correspondents need not be so (in 3D space). The result was that depth derivatives were approximated in 3D space on a nonuniform grid, while the underlying assumption was a uniform one (image space uniformity).

Weiss [29] suggested a physical formalism which enables incorporation of invariants of the imaging processes and geometric knowledge about the surface. This work describes a theoretical method, but presents no numerical results.

Although the great majority of researches in the field of SfS rely on the orthographic projection, and the minority which applies to perspective SfS is limited in scope, no information is available on the image irradiance equation under the perspective projection model. The goal of this paper is to formulate the image irradiance equation under the perspective assumption and then to solve the resultant Shape-from-Shading problem. The proposed solution is a perspective version of the Fast Marching method of Kimmel & Sethian [9] based on the new formulation of the image irradiance equation.

To motivate why a change in the underlying assumption from orthographic to perspective projection has a strong impact on the results, let us introduce an analytic example of two Lambertian quadrilaterals (Fig. 1(a)). It can be shown analytically, that perspective projection of the filled quadrilateral onto the image plane is identical to orthographic projection of the meshy parallelogram (the mesh is for visualization purposes only). Their images (under identical directional lighting) would also be the same, as they reside on the same plane, and hence have identical normals. This stems from the image irradiance equation (see [5]) for a Lambertian surface illuminated by a point light source at infinity (Sect. 2.2 will describe the equation). Consequently, the perspective image of the quadrilateral is identical to the orthographic image of the parallelogram under the same light source (Fig. 1(b)). This implies that if the quadrilateral was photographed by a perspective camera, but reconstructed by an ideal, orthographic algorithm, the reconstruction would be that parallelogram. Thus, the shape difference between the two quadrilaterals is a reconstruction error inherent in the orthographic model, which cannot be overcome by any specific orthographic algorithm. Furthermore, it can be proved that orthographic reconstruction of a rectangular image of a 3D plane must yield a 3D parallelogram; this need not be the case if the projection is perspective, as Fig. 1 demonstrates. (The proof is omitted for brevity.)

This example (Fig. 1) suggests that the improvement in reconstruction due to the perspective projection assumption may be considerable, as it diminishes a major source of error in current SfS techniques.

Preliminary results of the work described by the current paper appeared in [26]. (There, the algorithm was a very basic one with gradient descent minimization of an energy functional). In parallel to [26], another research group, Prados & Faugeras [18], developed the perspective image irradiance equation but with a different algorithm for its solution. The current paper will compare these two perspective methods as well as the orthographic Fast Marching on synthetic data.

The practical contribution of this paper will be further evaluated by a reconstruction comparison of the proposed algorithm and the original Fast Marching method on medical images taken by endoscopy from different parts of the gastrointestinal tract. The comparison will show that perspective SfS, in contrast with orthographic SfS (see the above quote of [32]), should be adequate for real-life applications such as endoscopy.

The paper is organized as follows. We first develop the image irradiance equation under the perspective projection model (Sect. 2), and explain its dependence on the natural logarithm of the depth function (Sect. 2.3). Section 3 provides intuition for surfaces in image coordinates and their reflectance maps under the perspective model. Examples of simple surfaces (planes and paraboloids) are described. Section 4 suggests a perspective SfS algorithm based on the Fast Marching method of [9]. Section 5 describes the comparison of orthographic Fast Marching, perspective Fast Marching and the algorithm of Prados & Faugeras on synthetic images. In addition, it compares the orthographic and perspective Fast Marching algorithms on medical images taken by endoscopy. Finally, Sect. 6 draws the conclusions. Appendix A derives the perspective image irradiance equation in detail. Appendix B develops the equations for the perspective Fast Marching method and proves the relevant theorems.

2. The Perspective Image Irradiance Equation

2.1. Notation and Assumptions

Let us first describe the notation and assumptions that hold throughout this paper. Photographed surfaces are assumed representable by functions of real-world coordinates as well as of image coordinates. $\hat{z}(x,y)$ denotes the depth function in a real-world Cartesian coordinate system whose origin is at camera plane. If the real-world coordinate $(x,y,\widehat{z}(x,y))$ is projected onto image point (u,v), then its depth is denoted $z(u,v)$. By definition, $z(u,v) = \hat{z}(x,y)$. $I(u,v)$ denotes the intensity at image point (u,v). f denotes the focal length, and is assumed known. The scene object is Lambertian, and is illuminated from direction $\vec{L} = (p_s, q_s, -1)$ by a point light source at infinity. $\vec{N}(x,y)$ is the surface normal.

2.2. Equation in Image Coordinates

As a first step in solving the image irradiance equation under the perspective projection model, we convert the equation into more convenient forms. The equation is given by:

$$I(u,v) = \vec{L} \cdot \vec{N}(x,y) \tag{1}$$

where:

$$x = -\frac{u \cdot \widehat{z}(x,y)}{f} \tag{2}$$

$$y = -\frac{v \cdot \widehat{z}(x,y)}{f} \tag{3}$$

Substituting Eqs. 2, 3 and $\vec{L} \stackrel{\text{by}}{=}_{\text{def}} (p_s, q_s, -1)$ (see Sect. 2.1) into Eq. 1 yields:

$$I(u,v) = \frac{1 + p_s\widehat{z}_x + q_s\widehat{z}_y}{\sqrt{1+p_s^2+q_s^2}\sqrt{1+\widehat{z}_x^2+\widehat{z}_y^2}} \tag{4}$$

We then express $\widehat{z}_x$ and $\widehat{z}_y$ in terms of u, v, z, z_u, and z_v, and substitute the resultant expressions along with Eqs. 2, 3 into Eq. 4. Appendix A derives these expressions from the projection equations, and obtains:

$$I(u,v) = \frac{(u-fp_s)z_u + (v-fq_s)z_v + z}{\sqrt{1+p_s^2+q_s^2}\sqrt{(uz_u+vz_v+z)^2 + f^2(z_u^2+z_v^2)}} \tag{5}$$

where $z(u,v) \stackrel{\text{def}}{=} \widehat{z}(x,y)$ for (u,v) which is the perspective projection of $(x,y,\widehat{z}(x,y))$. Equation 5 is the *perspective image irradiance equation*.

2.3. Dependence on $\ln(\mathbf{z}(\mathbf{u},\mathbf{v}))$

Equation 5 shows direct dependence on both $z(u,v)$ and its first order derivatives. If one employs $\ln(z(u,v))$ instead of $z(u,v)$ itself (by definition $z(u,v) > 0$), one obtains the following equation:

$$I(u,v) = \frac{(u - fp_s)p + (v - fq_s)q + 1}{\sqrt{1 + p_s^2 + q_s^2}\sqrt{(up + vq + 1)^2 + f^2(p^2 + q^2)}} \tag{6}$$

where $p \stackrel{\text{def}}{=} \frac{z_u}{z} = \frac{\partial \ln z}{\partial u}$ and $q \stackrel{\text{def}}{=} \frac{z_v}{z} = \frac{\partial \ln z}{\partial v}$. Eq. 6 depends on the derivatives of $\ln(z(u,v))$, but not on $\ln(z(u,v))$ itself. Consequently, the problem of recovering $z(u,v)$ from the image irradiance equation reduces to the problem of recovering the surface $\ln(z(u,v))$ from Eq. 6. Because the natural logarithm is a bijective mapping and $z(u,v) > 0$, recovering $\ln(z(u,v))$ is equivalent to recovering $z(u,v) = e^{\ln(z(u,v))}$.

The image irradiance equation under orthographic projection is invariant to translation of $\widehat{z}(x,y)$, which means $\widehat{z}(x,y) + c$ (for constant c) produces the same intensity function as $\widehat{z}(x,y)$. In contrast, the perspective image irradiance equation (Eq. 5) is invariant to scale changes of $z(u,v)$. That is, the intensity functions of $c \cdot z(u,v)$ and $z(u,v)$ are identical. This follows from the properties of the natural logarithm, and can also be verified by Eqs. 5, 6. Invariance to scaling seems to be a more plausible assumption than invariance to translation when employing real cameras.

3. The Perspective Irradiance Equation of Simple Surfaces

We next provide some analytic examples of surfaces and their representation in the image coordinate system $(u,v,z(u,v))$, and their reflectance map $(R(u,v))$ under the perspective model. These formulae would sharpen the difference between the orthographic and perspective models and would give the reader some intuition for the difference between the real-world representation of a surface $(x,y,\hat{z}(x,y))$ and its representation in image coordinates $(u,v,z(u,v))$ under the perspective model (under the orthographic model, these representations are identical).

We examine two types of real-world surfaces: planes and paraboloids.

3.1. Planes

Let us consider a general plane:

$$\widehat{z}(x,y) = z_0 + a(x - x_0) + b(y - y_0)$$

where a, b, x_0, y_0, z_0 are constants. Substituting image coordinates (u,v) according to the perspective projection equations and solving for $z(u,v)$ yields:

$$z(u,v) = z_0 \frac{f + au_0 + bv_0}{f + au + bv} \tag{7}$$

where $u_0 \stackrel{\text{def}}{=} -\frac{f \cdot x_0}{z_0}$, $v_0 \stackrel{\text{def}}{=} -\frac{f \cdot y_0}{z_0}$. The last equation states that the depth of the planar surface at point (u,v) is proportional to the *reciprocal* of $au + bv$. The opposite takes place in orthographic projection: $x \propto u$, $y \propto v$, and hence depth is proportional to $au + bv = ax + by$, by definition of $\widehat{z}(x,y)$.

Under both perspective and orthographic projections, the image irradiance equation becomes:

$$R(u,v) = \frac{p_s a + q_s b + 1}{\|\vec{L}\|\sqrt{a^2 + b^2 + 1}} \tag{8}$$

This fact is trivial for the orthographic projection. In [27] we derive this equation for the perspective case as well. The equation shows that for a planar object the image irradiance is constant (i.e., independent of u and v) under both projection models.

3.2. Paraboloids

3.2.1. Canonical Paraboloids

We first consider a canonical paraboloid of the form:

$$\widehat{z}(x,y) = ax^2 + by^2$$

Its representation in image coordinates under perspective projection is:

$$z(u,v) = \begin{cases} \frac{f}{au^2+bv^2}, & \text{if } au^2+bv^2 \neq 0 \\ 0, & \text{if } au^2+bv^2 = 0 \end{cases}$$

Again, the perspective and orthographic equations are reciprocal (up to a scale factor).

The reflectance map in this case is:

$$R(u,v) = \frac{2f(p_s au + q_s bv) - (au^2+bv^2)}{\|\vec{L}\|\sqrt{au^2+bv^2}\sqrt{au^2+bv^2+4f^2}}$$

3.2.2. General Paraboloids

For a general paraboloid of the form:

$$\widehat{z}(x,y) = z_0 + a(x-x_0) + b(y-y_0) + c(x-x_0)^2 + d(y-y_0)^2 + e(x-x_0)(y-y_0)$$

the image coordinate representation is:

$$z(u,v) = \frac{S(u,v) - \sqrt{S^2(u,v) - 4T(u,v)P}}{2T(u,v)} \tag{9}$$

where:

$$\begin{aligned} T(u,v) &\stackrel{\text{def}}{=} cu^2 + dv^2 + euv \\ S(u,v) &\stackrel{\text{def}}{=} f^2 + u(fa + 2cu_0z_0 + ev_0z_0) + v(fb + 2dv_0z_0 + eu_0z_0) \\ P &\stackrel{\text{def}}{=} z_0 f(f + au_0 + bv_0) + z_0^2(cu_0^2 + dv_0^2 + eu_0v_0) \end{aligned}$$

(assuming $T(u,v) \neq 0$). The reflectance formula in this case is omitted due to its complexity. Even though there exists another solution to the quadratic equation, in the general case that solution is not physical. This is because substitution of z_0 into the other solution results in $z(u_0,v_0) \neq z_0$ (unless $f + au_0 + bv_0 = 0$), which contradicts the definition of z_0.

4 Perspective Fast Marching

This section suggests a perspective SfS algorithm. The algorithm is a modification of the Fast Marching method of Kimmel and Sethian [9] from the orthographic set of assumptions to the perspective one.

4.1 Solving The Approximate Problem

The algorithm of Kimmel and Sethian [9] stems from the orthographic image irradiance equation: $I(x,y) = \vec{L} \cdot \vec{N}(x,y)$. This equation is known as the Eikonal equation and can be written as:

$$p^2 + q^2 = \tilde{F}^2$$

where $p \stackrel{\text{def}}{=} z_u = z_x$, $q \stackrel{\text{def}}{=} z_v = z_y$ and $\tilde{F} = \sqrt{(I(x,y))^{-2} - 1}$. Similarly, the perspective image irradiance equation (Eq. 5), can be transformed into the form:

$$p^2 A_1 + q^2 B_1 = \hat{F} \tag{10}$$

where A_1 and B_1 are positive and independent of p as well as of q. $\hat{F}$, on the other hand, depends on both p and q. The complete expressions for A_1, B_1 and $\hat{F}$ appear in Appendix B.

Following [9], we use the numerical approximation (originally introduced in [21] as a modification of the scheme of [15]):

$$\begin{aligned} p_{ij} &\approx \max\{D_{ij}^{-u}z, -D_{ij}^{+u}z, 0\} \\ q_{ij} &\approx \max\{D_{ij}^{-v}z, -D_{ij}^{+v}z, 0\} \end{aligned}$$

where $D_{ij}^{-u} z \stackrel{\text{def}}{=} \frac{z_{ij}-z_{i-1,j}}{\Delta u}$ is the standard backward derivative and $D_{ij}^{+u} z \stackrel{\text{def}}{=} \frac{z_{i+1,j}-z_{ij}}{\Delta u}$, the standard forward derivative in the u-direction ($z_{ij} \stackrel{\text{def}}{=} z(i \cdot \Delta u, j \cdot \Delta v)$). $D_{ij}^{-v} z$ and $D_{ij}^{+v} z$ are defined in a similar manner for the v-direction.

The motivation for employing this numerical scheme is due to its consistency and monotonicity. For the Eikonal equation, Rouy & Tourin [21] have shown that an iterative algorithm based on this scheme with Dirichlet boundary conditions on image boundaries and at all critical points converges towards the viscosity solution with the same boundary conditions. Existence of the viscosity solution has been proven in [14] and uniqueness, in [21] and [7]. Sethian [24] have proven that the Fast Marching algorithm produces a solution that everywhere satisfies the discrete version of the Eikonal equation.

Substituting the numerical approximation into Eq. 10, we get the discrete equation:

$$\left(\max\{D_{ij}^{-u} z, -D_{ij}^{+u} z, 0\}\right)^2 A_1 + \left(\max\{D_{ij}^{-v} z, -D_{ij}^{+v} z, 0\}\right)^2 B_1 = \hat{F}_{ij} \tag{11}$$

where $\hat{F}_{ij} \stackrel{\text{def}}{=} \hat{F}(i \cdot \Delta u, j \cdot \Delta v)$. As Appendix B details, the solution of this equation at every point (i, j) is:

$$z = \begin{cases} z_1 + \sqrt{\frac{\hat{F}}{A_1}}, & \text{if } \; z_2 - z_1 > \sqrt{\frac{\hat{F}}{A_1}} \\ z_2 + \sqrt{\frac{\hat{F}}{B_1}}, & \text{if } \; z_1 - z_2 > \sqrt{\frac{\hat{F}}{B_1}} \\ \frac{A_1 z_1 + B_1 z_2 \pm \sqrt{(A_1+B_1)\hat{F} - A_1 B_1 (z_1 - z_2)^2}}{A_1 + B_1}, & \text{otherwise} \end{cases} \tag{12}$$

where $z_1 \stackrel{\text{def}}{=} \min\{z_{i-1,j}, z_{i+1,j}\}$ and $z_2 \stackrel{\text{def}}{=} \min\{z_{i,j-1}, z_{i,j+1}\}$.

4.2 The Iterative Solution

An important observation described in [9] is that information always flows from small to large values at local minimum points. Based on this, the orthographic Fast Marching method reconstructs depth by first setting all z values to infinity, and the correct height value at the local minima. Then, every step extends the reconstruction to higher depths. Reconstruction is thus achieved by a single pass.

Nevertheless, a single pass cannot solve the aforementioned formulation of the perspective problem (Eq. 11), because the approximate solution (the right-hand side of Eq. 12) depends on $\hat{F}$, which depends on both p and q. Hence, we suggest an iterative method. In every iteration, $\hat{F}$ is calculated according to the depth recovered by the previous iteration. Based on this approximation of $\hat{F}$ and on Eq. 12, a solution is calculated for the new iteration. We initialize this process by the orthographic Fast Marching method of [9].

Following each iteration, the resulting depth map was normalized (i.e., divided by the norm of all depth values). This preserves a correct reconstruction, because the perspective SfS is invariant to multiplication by constant (see Sect. 2.3).

5. Experimental Results

5.1. The Experiments

To evaluate the contribution of perspective SfS, we compared it with the Fast Marching method of Kimmel & Sethian [9]. The reason for selecting this orthographic algorithm for the comparison is triple. First, we consider the Fast Marching method a state-of-the-art technique. Second, in [26] we compared three orthographic methods (Lee & Kuo [11], Zheng & Chellappa [34] and Kimmel & Sethian [9]) with a basic perspective method that was suggested there (based on gradient descent). Among these three orthographic methods, the Fast Marching method performed best. Third, the fact that the suggested perspective method is based on this orthographic method, neutralizes the effect of the numerical scheme on the results. Therefore, any improvement would be a consequence of the transition to the perspective equation, and *not* of the different ways of solving the equations.

Recently, another perspective algorithm has been suggested by Prados & Faugeras [18] in parallel to ours. We compare our algorithm with this algorithm as well.

An important advancement over [26], which compares merely synthetic images, is the experimentation with real images. In addition to a demonstration with synthetic data, we compared the orthographic and perspective Fast Marching algorithms[1] on medical images taken by endoscopy.

[1] The algorithm of Prados & Faugeras could not be compared on real data, as it requires the exact depth function on the boundaries (Dirichlet condition).

5.1.1. Experiments with Synthetic Images

All synthetic input images were produced from an original surface $\widehat{z}(x, y)$ in the real world. The surface was projected onto plane $[uv]$ according to the perspective projection equations (Eqs. 2, 3). A rectangular area bounded by the projection and symmetric about the optical axis was uniformly sampled. The original surface $\widehat{z}(x, y)$ was then interpolated to the sampling points. The orthographic image irradiance equation then served to create the intensity at each point. This procedure was applied to avoid direct usage of the perspective formula, which the proposed algorithm attempts to recover.

A large amount of synthetic inputs was examined, but only few can fit into this paper. Section 5.2 provides representative examples.

To evaluate the contribution of the perspective Fast Marching, we compared it with two other algorithms: an orthographic algorithm (Fast Marching by Kimmel & Sethian [9]) and a different perspective algorithm (by Prados & Faugeras [18]).

We evaluated the performance of the algorithms on synthetic images according to three criteria adopted from Zhang et al. [32]: mean depth error, standard deviation of depth error, and mean gradient error. For completeness, we also supply the standard deviation of gradient error, although it is considered not physical.

Notwithstanding, the adoption of orthographic criteria (such as the above) to the perspective case is nontrivial. In contrast with a pure orthographic comparison (as in [32]), where reconstructed $[xy]$ domains are guaranteed to be rectangular, in a perspective comparison each algorithm may recover a different $[xy]$ domain. Thus, the resultant surface points need not have the same (x, y) rates as points on the original surface. Consequently, scaling the recovered surface to fit the original (due to invariance to depth scaling; see Sect. 2.3) is also more complicated. The scaling now need be calculated by surface samples at different (x, y) locations.

To best fit the reconstructed $[xy]$ domains to the true ones (in the least-squares sense), we scaled them linearly. In order to determine a scale factor for the depth functions ($\hat{z}(x, y) = z(u, v)$) we projected the reconstructed surface onto the true one, and calculated the scale factor between reconstructed points and their projection. The distance from reconstructed points to the projections was taken as the distance for mean depth error.

We considered three methods of projection:

1. The trivial one, to compare depths at points corresponding to the same image pixel. This method ignores the discrepancy in $[xy]$ domain.

2. To interpolate and extrapolate the original surface by a Thin-Plate spline, and approximate the z value of the original surface at the (x, y) rates where the reconstructed surface is provided. Thus, projection is vertical (i.e., parallel to the z-axis).

3. To project reconstructed points onto the true surface using an approximation of the Moving Least Squares (MLS) method [13]. The main idea is to project a point onto a surface by finding the nearest neighbor of the point among surface points, approximating a plane in its vicinity (from surface points), and projecting the point onto this plane, perpendicularly. Then, the surface is approximated by Weighted Least Squares in a local coordinate system (defined by this plane) at the point of projection. This type of projection is locally perpendicular to the target surface.

When comparing orthographic and perspective algorithms, measures based on the first two method led to inconclusive results. The comparison we describe hereafter would therefore be based upon the third projection, Moving Least Squares.

5.1.2. Experiments with Real Images — Endoscopy

We studied endoscopic images taken from different parts of the gastrointestinal channel.

Endoscopy is a practical field of life on the one hand, while it has the advantage of a controlled light source environment, on the other hand. The light source can be considered a point light source, but not an infinitely distant one. To overcome this limitation, we worked on a small portion of the original endoscopic image at a time. This had a double effect. First, light in this case came from a narrow range of directions, which could be approximated by a constant lighting direction. Second, because the light source and camera were adjacent, a narrower range of distances from object to camera meant a narrower range of distances to the light source as well. This diminished the decay of illumination strength with distance.

5.1.3. Algorithm Implementation

We tested the algorithm of Kimmel & Sethian using two implementations. First, we extended the implementation of the Fast Marching method by the Technical University of Munich[2] to accommodate the oblique light source case as well. Then, to ensure the correctness, we re-implemented the algorithm from scratch. Both implementations gave similar results. In the comparison, we quote our own implementation, but results are practically the same with both.

The code of the orthographic Fast Marching served as the basis for the implementation of the perspective Fast Marching too. Thus again, two implementations were produced and verified.

The implementation of the algorithm of Prados & Faugeras [18] is courtesy of the original authors. To be consistent with the original paper, the implementation starts from a subsolution and uses Dirichlet boundary data on image boundaries and at all the critical points. The stopping criterion was a threshold of 10^{-10} on the difference between the surfaces reconstructed by two successive iterations.

5.1.4. Parameters and Visualization Issues

The three algorithms under study assume that light source direction is known. For the synthetic images the true direction was provided, but for the endoscopic images these data were unavailable. We therefore utilized very rough estimations of light source directions. A human viewer estimated the azimuth and elevation of the light source direction from the endoscopic image itself in multiples of $\frac{\pi}{8}$ or $\frac{\pi}{6}$ radians. The same estimated direction was supplied to all methods.

In addition, perspective SfS requires the knowledge of the focal length f. Our implementation arbitrarily set an identical value for all examples.

Another kind of data required by all three algorithms is the points of local minimal depth. Again, for the synthetic examples the true data was supplied, while for the real ones a human viewer visually located the points in the photographs, and set their depth to an arbitrary constant (identical for all real images).

The algorithm of Prados & Faugeras was supplied with the Dirichlet boundary data extracted from the synthetic images. It was also supplied with the true $[uv]$ grid size used to construct the images.

As a post-processing step, all real-image reconstructions underwent a translation and a rotation to convert camera coordinates to object coordinates, for better visualization.

The suggested algorithm converges very fast. No more than 2 iterations, in addition to the orthographic stage, were necessary for the perspective Fast Marching method to converge on real-life images. We demonstrate this in our comparison by inclusion of images of 5 iterations per example, all of which appear to be visually the same. We exploit the excessive images to provide more viewing angles of the perspective reconstruction. Viewing angles were selected so as to let the reader appreciate the three-dimensionality of reconstructed surfaces.

In all real examples, the orthographic reconstruction and the perspective reconstruction after 1 iteration were plotted from an identical viewpoint to allow their visual comparison. Also, the same illumination and albedo were used to reproduce the orthographic and perspective surfaces.

5.2. Comparative Evaluation of Synthetic Examples

The synthetic surfaces we study are described in Table 1. Figure 2 shows the original image of each example (size: 50×50 pixels), the real surface and reconstruction by three algorithms: orthographic Fast Marching, perspective Fast Marching and the perspective algorithm of Prados and Faugeras [18]. The reconstructed surfaces and the real ones are juxtaposed in Fig. 3. Tables 4–6 summarize the error rates according to the aforementioned criteria.

Example #1: Perspective Fast Marching and Prados & Faugeras gained error rates lower than those of the orthographic Fast Marching according to all measures. Perspective Fast Marching performed better than Prados & Faugeras according to mean and standard deviation of depth error. Prados & Faugeras performed better than perspective Fast Marching according to mean gradient error. In general, both perspective algorithms are better than orthographic Fast Marching, while they equate with each other, both being based on the same equation.

Example #2: Perspective Fast Marching has lower mean depth and gradient errors than the orthographic version. The orthographic Fast Marching has lower standard deviation of depth error than the perspective Fast Marching. Prados & Faugeras obtains lowest mean and standard deviation of the depth error, but highest mean gradient error. In a visual inspection

[2] Folkmar Bornemann, Technical University of Munich, WiSe 00/01, 11.12.2000, http://www-m8.mathematik.tu-muenchen.de/m3/teaching/PDE/begleit.html

	Formula:	$\vec{L} =$	$x \in$	$y \in$
1.	$\hat{z}(x,y) = 300 + 30(\sin(2x) + \sin(2y))$	$(0,0,-1)$	$[-3.0788, 3.054]$	$[-3.0788, 3.054]$
2.	Vase image. The image was slanted by 20^{o} about the x-axis, because otherwise the background would have a constant depth and need be supplied to Fast Marching as minima points.	$(0,0,-1)$	$[-12.3654, 12.3654]$	$[-12.3654, 12.3654]$
3.	$\hat{z}(x,y) = 5(\cos(\sqrt{x^2+(y-2)^2}) + \cos(\sqrt{x^2+(y-1)^2}) + \cos(\sqrt{x^2+(y+2)^2})) + 100$	$(0,0,-1)$	$[-2.9016, 2.9016]$	$[-2.9478, 2.9486]$
4.	$\hat{z}(x,y) = \ln(\sqrt{x^2+y^2})$	$(0,0,-1)$	$[-15.3283, 15.3283]$	$[-15.3283, 15.3283]$
5.	$\hat{z}(x,y) = \sin(2x)$, $\hat{z}(x,y)$ is then rotated by 20^{o} about the x-axis and the result is scaled by factor of 2 and translated by 20	$(0,1,-1)$	$[-2.5820, 2.5276]$	$[-1.9440, 2.3582]$

Table 1: *The formulae and parameters of four typical synthetic examples. These examples were part of a much larger comparison, and would be described in detail herein.*

Algorithm:	No. of Iterations:	Mean Depth Error:	Std. Dev. of Depth Error:	Mean Gradient Error:	Std. Dev. of Gradient Error:
Kimmel & Sethian:	1	0.46340	0.31221	51.79201	1862.09956
Perspective FM:	$1^{\underline{\text{st}}}$	0.44414	0.31049	21.42945	202.78671
Perspective FM:	$2^{\underline{\text{nd}}}$	0.31569	0.23315	58.57205	1112.16660
Perspective FM:	$3^{\underline{\text{rd}}}$	0.31123	0.22648	22.57623	376.79622
Perspective FM:	$4^{\underline{\text{th}}}$	0.30853	0.22492	19.27577	254.64257
Perspective FM:	$5^{\underline{\text{th}}}$	0.30787	0.22477	29.18984	623.03837
Prados & Faugeras:	169	0.32077	0.25668	26.82533	274.27769

Table 2: *Comparison of algorithms on example #1.*

Algorithm:	No. of Iterations:	Mean Depth Error:	Std. Dev. of Depth Error:	Mean Gradient Error:	Std. Dev. of Gradient Error:
Kimmel & Sethian:	1	4.66332	2.47092	0.08462	0.59064
Perspective FM:	$1^{\underline{\text{st}}}$	3.02464	3.29558	0.06061	0.69701
Perspective FM:	$2^{\underline{\text{nd}}}$	3.11067	3.28568	0.06012	0.69318
Perspective FM:	$3^{\underline{\text{rd}}}$	3.11062	3.28573	0.06010	0.69300
Perspective FM:	$4^{\underline{\text{th}}}$	3.11062	3.28573	0.06010	0.69301
Perspective FM:	$5^{\underline{\text{th}}}$	3.11062	3.28573	0.06010	0.69301
Prados & Faugeras:	89	1.80394	1.15217	0.30808	2.06713

Table 3: *Comparison of algorithms on example #2.*

Algorithm:	No. of Iterations:	Mean Depth Error:	Std. Dev. of Depth Error:	Mean Gradient Error:	Std. Dev. of Gradient Error:
Kimmel & Sethian:	1	0.58363	0.44509	12.16410	120.05034
Perspective FM:	$1^{\underline{\text{st}}}$	0.41374	0.27841	8.13954	284.45168
Perspective FM:	$2^{\underline{\text{nd}}}$	0.09686	0.06990	0.40490	2.60023
Perspective FM:	$3^{\underline{\text{rd}}}$	0.09474	0.06947	0.32935	1.00981
Perspective FM:	$4^{\underline{\text{th}}}$	0.09455	0.06935	0.31795	0.80138
Perspective FM:	$5^{\underline{\text{th}}}$	0.09455	0.06938	0.31755	0.80125
Prados & Faugeras:	356	0.03068	0.03564	0.15031	0.25134

Table 4: *Comparison of algorithms on example #3.*

Algorithm:	No. of Iterations:	Mean Depth Error:	Std. Dev. of Depth Error:	Mean Gradient Error:	Std. Dev. of Gradient Error:
Kimmel & Sethian:	1	0.16896	0.10483	0.08623	0.16119
Perspective FM:	$1^{\underline{\text{st}}}$	0.08131	0.06237	0.04360	0.09746
Perspective FM:	$2^{\underline{\text{nd}}}$	0.07401	0.05411	0.02924	0.06335
Perspective FM:	$3^{\underline{\text{rd}}}$	0.07418	0.05436	0.03056	0.06359
Perspective FM:	$4^{\underline{\text{th}}}$	0.07419	0.05437	0.03058	0.06361
Perspective FM:	$5^{\underline{\text{th}}}$	0.07419	0.05437	0.03058	0.06361
Prados & Faugeras:	35	0.07950	0.06459	0.03628	6.94938

Table 5: *Comparison of algorithms on example #4.*

Algorithm:	No. of Iterations:	Mean Depth Error:	Std. Dev. of Depth Error:	Mean Gradient Error:	Std. Dev. of Gradient Error:
Kimmel & Sethian:	1	0.35744	0.21890	9.10098	284.53313
Perspective FM:	$1^{\underline{\text{st}}}$	0.34954	0.23540	6.74028	88.13782
Perspective FM:	$2^{\underline{\text{nd}}}$	0.40338	0.25898	2.28858	4.98138
Perspective FM:	$3^{\underline{\text{rd}}}$	0.40178	0.26164	16.31858	164.50187
Perspective FM:	$4^{\underline{\text{th}}}$	0.40222	0.26133	8.77626	38.31631
Perspective FM:	$5^{\underline{\text{th}}}$	0.40249	0.26089	14.39487	216.13707
Prados & Faugeras:	73	0.38651	0.25717	5.27366	32.45018

Table 6: *Comparison of algorithms on example #5.*

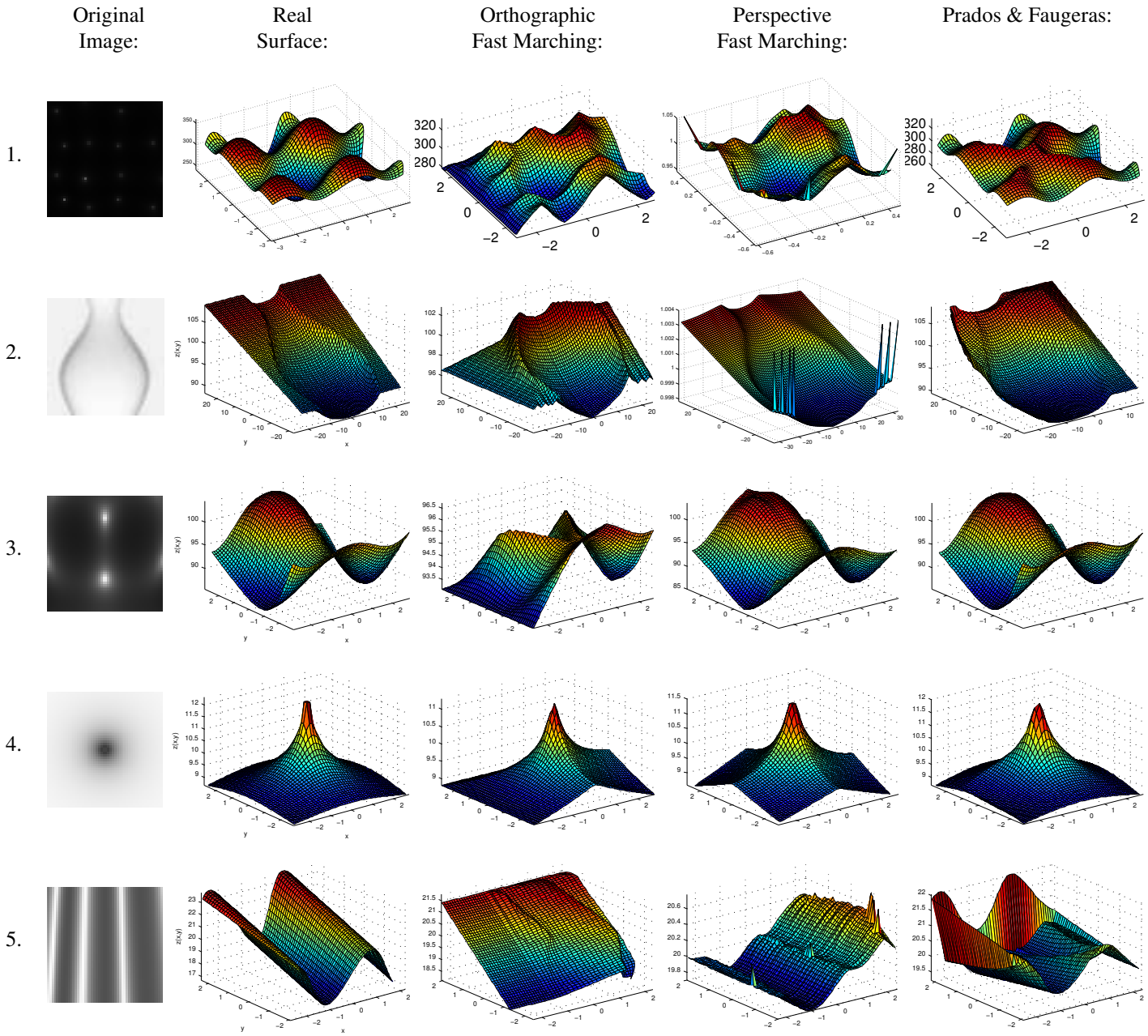


Figure 2: Comparison of surfaces reconstructed by the orthographic Fast Marching, perspective Fast Marching, and the perspective algorithm of Prados and Faugeras [18]. The leftmost column indicates the serial number of the example (see Table 1). In Examples #1,#2, some spikes in the reconstruction by perspective Fast Marching were cropped for better visualization only.

it is clear that except for some peaks, the perspective Fast Marching generated a reconstruction most similar to the original. (Fig. 2; Indeed, removing the peaks would result in lowest error rates according to all criteria.)

Example #3: Perspective Fast Marching and Prados & Faugeras performed significantly better than orthographic Fast Marching according to all error criteria. Prados & Faugeras performed better than perspective Fast Marching according to all error criteria.

Example #4: The two perspective methods perform better than the orthographic. Perspective Fast Marching gained lower error rates than Prados & Faugeras according to all criteria. This can be seen in Fig. 2 especially by the more accurate shape

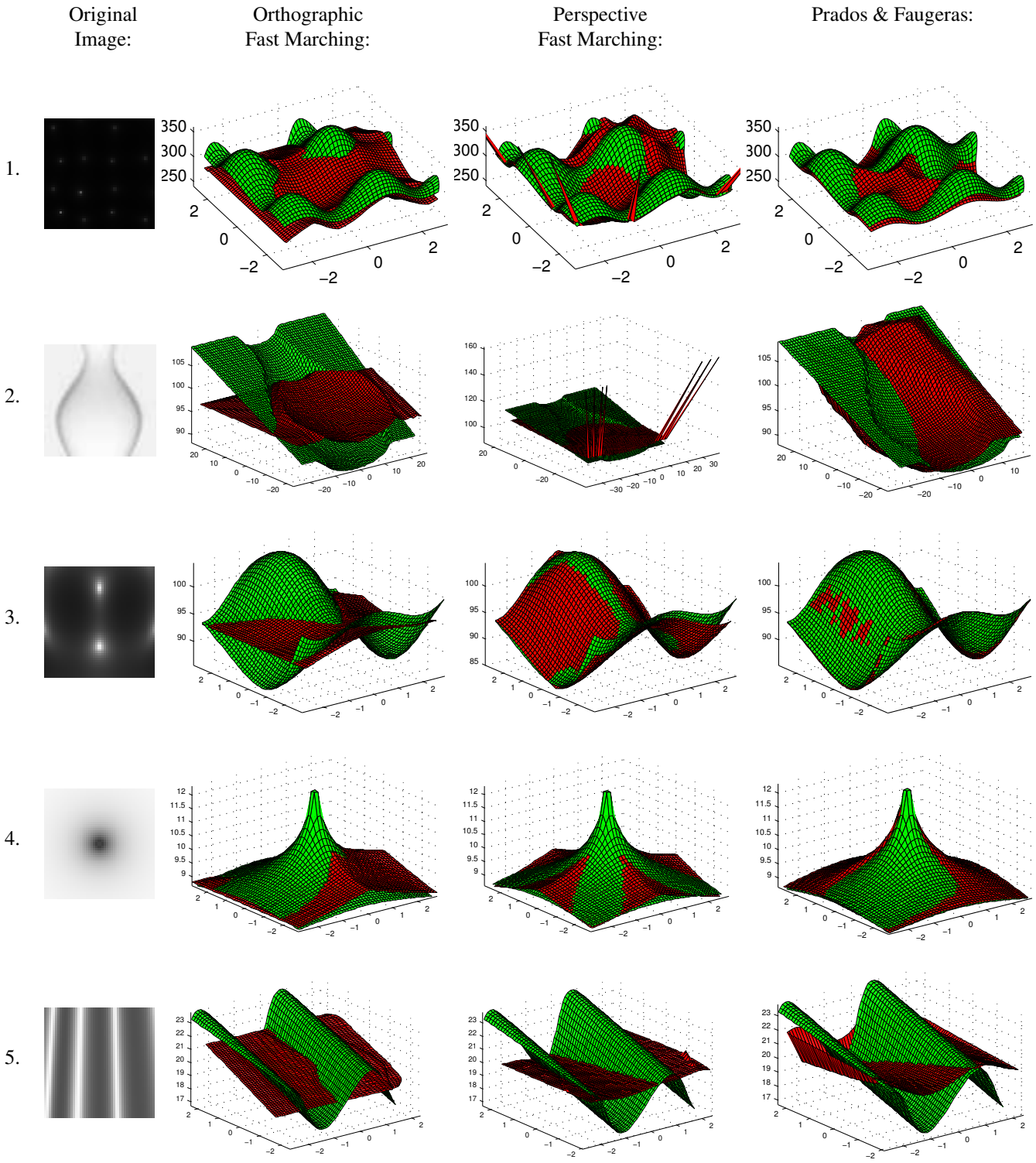


Figure 3: Juxtaposition of the surfaces reconstructed by the three algorithms and the true surfaces. The leftmost column indicates the serial number of the example (see Table 1). In Example #1, some spikes in the reconstruction by perspective Fast Marching were cropped for better visualization only.

of perspective Fast Marching at the vicinity of the peak. Bear in mind that the smoother surface at image boundaries obtained by Prados & Faugeras is due to their requirement of Dirichlet boundary condition at both critical points and image boundaries (i.e., true depth is supplied to Prados & Faugeras on the boundaries as well).

Example #5: In this example orthographic Fast Marching obtained lowest error rates according to mean and standard deviation of depth error. It gained lower mean gradient error than perspective Fast Marching, but higher than Prados & Faugeras. Nevertheless, visual inspection of the perspective Fast Marching and the algorithm of Prados & Faugeras (Fig. 2) reveals, that both algorithms managed to reconstruct the sinusoidal structure of the true surface (with some errors), while orthographic Fast Marching did not recover this structure at all. The original surface consists of parallel sinusoidal waves, but due to the perspective projection, the images of the crests become unparallel. The perspective methods succeeded to recover this structure, but the orthographic Fast Marching reconstructed waves which are neither parallel nor uniform along the y-axis. It reconstructed one of the sinusoidal waves approximately half-sized and with incorrect orientation (unparallel to the y-axis). The other wave is again unparallel to the y-axis, with one orientation for positive y-rates and another for negative ones.

We see that in most examples, the perspective methods obtained lower error rates than the orthographic one. Nevertheless, there were cases when the orthographic method gained lower error rates according to all or some criteria. In these cases (as exemplified by Examples #2, #5), the orthographic reconstruction appears to be inferior to the perspective ones in visual inspection. This demonstrates why error measures common in the literature, such as mean and standard deviation of depth or gradient errors, disagree with human vision. While the errors ranked the orthographic Fast Marching as best in Example #5, visual inspection revealed its failure to recover the underlying sinusoidal structure (in contrast with the perspective methods). In Example #2, the measures failed to show that except for a very small portion of the image (the peaks) the reconstruction by perspective Fast Marching was very accurate. This mainly stem from incorrect scaling in the first stage of the computation of the measures (cf. the perspective Fast Marching in Fig. 2 with its appearance in Fig. 3, where the peaks were not cropped).

Another important factor in a reconstruction comparison is the projection of the reconstructed surface onto the original one. This is especially important for perspective algorithms, where scaling of the surface and its comparison in real-world coordinates are sensitive to this projection. An inaccurate scaling process can change comparison results drastically. Improving the error measures or the projection model is beyond the scope of this paper, and is a subject for future research.

While the perspective Fast Marching and the algorithm of Prados & Faugeras equate when considering the quality of the reconstructions they produce, perspective Fast Marching has three important advantages over the algorithm of Prados & Faugeras:

1. It requires a significantly lower number of iterations to converge, at least 1–2 orders of magnitude on the simple synthetic images examined.

2. The algorithm of Prados & Faugeras uses the full knowledge of the focal length and the $[uv]$ grid on which the image has been produced. The perspective Fast Marching, on the other hand, lacks the grid size information, and thus spaced the grid with 1 unit intervals. This is equivalent to lack of knowledge of the focal length f.

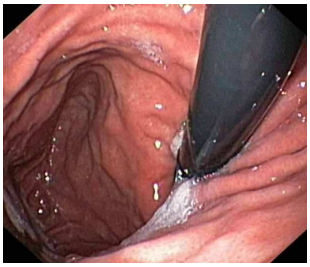
3. Perspective Fast Marching does not require a Dirichlet boundary condition on image boundaries. Knowledge of the true boundary depth is not trivial to obtain (unlike depth at minima points, where a global topology solver can be applied [8], [2], [9]). [If one had an algorithm to obtain the boundary depth, one could run this algorithm, then crop the boundary of the image, re-run the algorithm, etc. and thus build the depth map from the boundary inward.]

5.3. Comparison on Real Medical Images

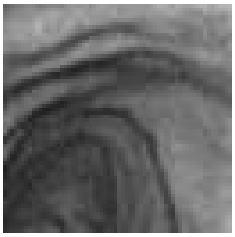
Figure 4A shows the gastric fundus[3]. The cropped version of this image focuses on a cavity with folded walls (Fig. 4B). The orthographic Fast Marching method failed to recover the cavity. Straight horizontal folds were recovered instead of curved gastric folds along cavity walls (Fig. 4C). These showed little match to the true ones. The perspective method presented high correspondence of both the cavity and its folds to the contents of the original image (Figs. 4D–H). Figures 4D–H show perspective reconstruction after 1–5 perspective iterations. As no significant improvement occurred at the second or higher iterations, we use different viewing angles to emphasize the 3D structure. Figures 4C,D have an identical viewpoint, which enables their visual comparison.

Figure 5A introduces the gastric angulus[3]. The cropped version of this image contains three folds (Fig. 5B). The ortho-

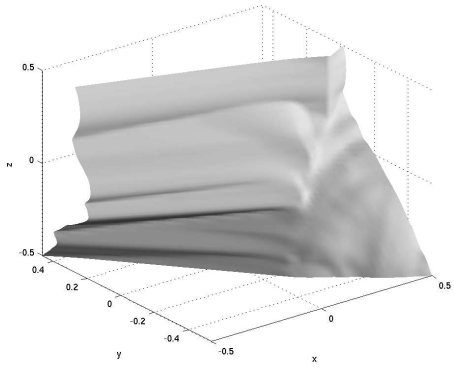
[3]Image from www.gastrolab.net, courtesy of The Wasa Workgroup on Intestinal Disorders, GASTROLAB, Vasa, Finland.



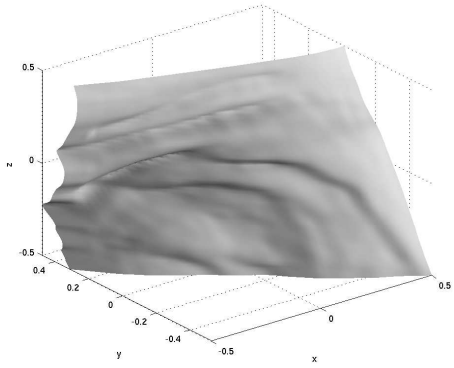
A. Original.



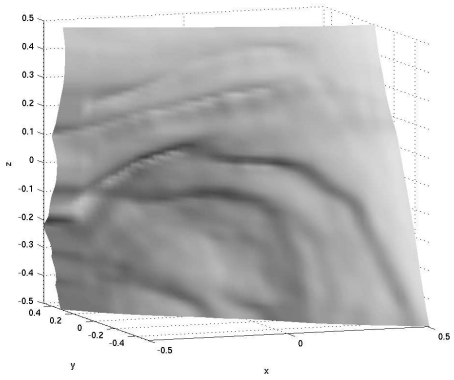
B. Cropped image.



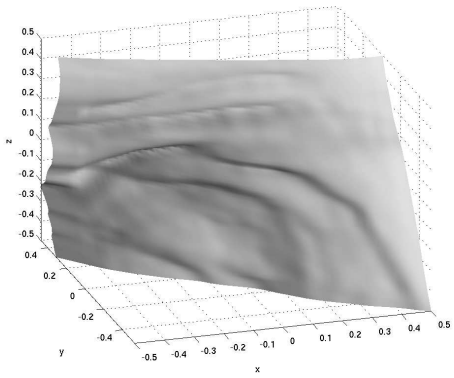
C. Orthographic reconstruction.



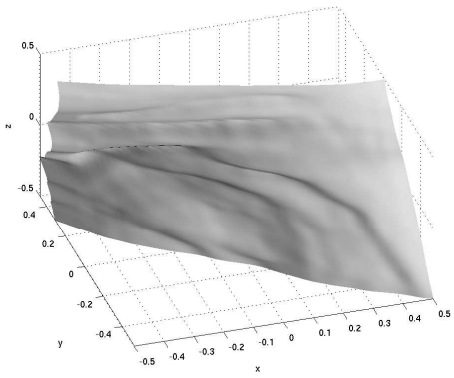
D. Perspective reconstruction (1 iter.).



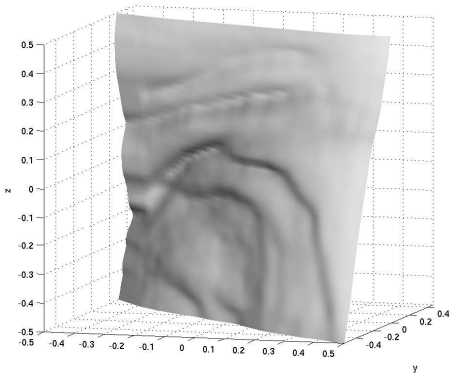
E. Perspective reconstruction (2 iters.).



F. Perspective reconstruction (3 iters.).



G. Perspective reconstruction (4 iters.).



H. Perspective reconstruction (5 iters.).

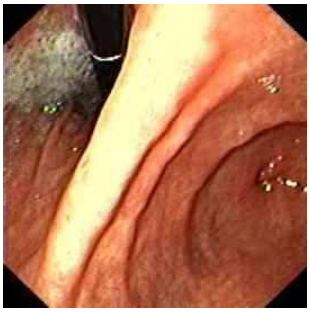
Figure 4: The gastric fundus (cropped image size: 64×64 pixels). Perspective reconstruction is visually the same at all iterations; we exploit this to display more viewing directions of the reconstructed surface. The viewpoint in (C) and (D) is identical.

graphic method reconstructed one fold, but instead of the second one, a bend of the surface was recovered. Between the first and second folds there was a very prominent pyramid-like cavity. The third fold is missing (Fig. 5C). Perspective reconstruction clearly recovered all three folds (Figs. 5D–H). In Figs. 5D–H, pay special attention to the change in width of the shadow casted by the folds. The visible part of the shadow alters between different viewing directions. This implies that the reconstructed folds are indeed three dimensional. Fig. 5H emphasizes the height of each fold above the gastric wall.

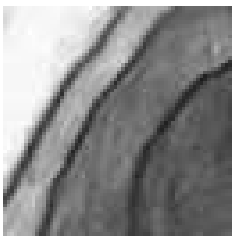
Figure 6A exhibits the descending duodenum[3]. The cropped version contains three plicae circulares (folds typical of the small intestine; Fig. 6B). The orthographic method yielded two surfaces with a sharp edge between them, which did not appear in the original data (Fig. 6C). In contrast, the perspective version recovered all three folds correctly (Figs. 6D–H). Fig. 6H is a side view of the folds which lets their different heights be appreciated.

Figure 7A presents an inverted appendix[4]. Figure 7B focuses on the appendiceal orifice; its image is of low quality.

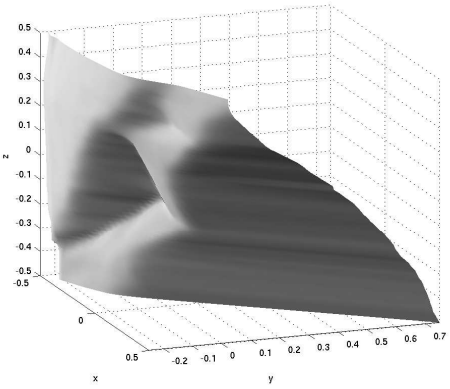
[4]Image from www.gastrointestinalatlas.com, courtesy of the Department of Gastroenterology, Hospital Centro de Emergencias, Jerusalem Medical



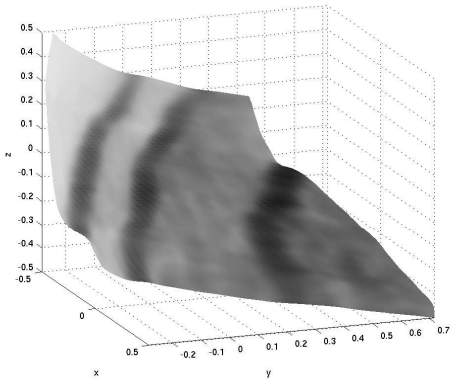
A. Original.



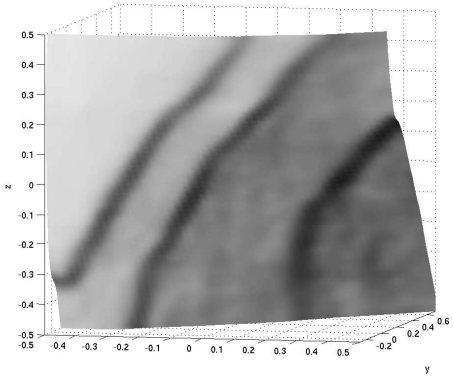
B. Cropped image.



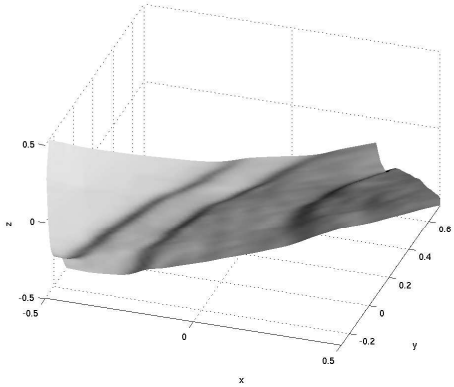
C. Orthographic reconstruction.



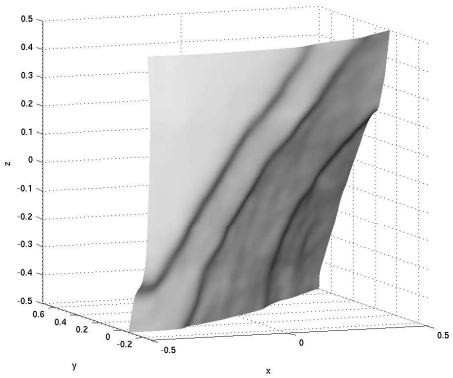
D. Perspective reconstruction (1 iter.).



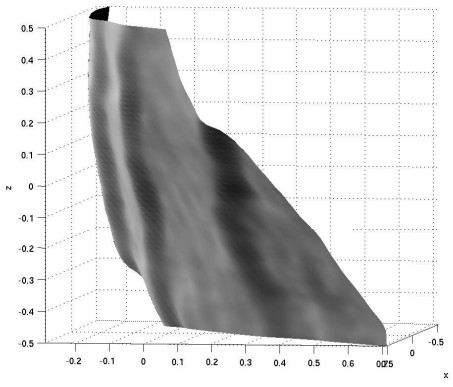
E. Perspective reconstruction (2 iters.).



F. Perspective reconstruction (3 iters.).



G. Perspective reconstruction (4 iters.).



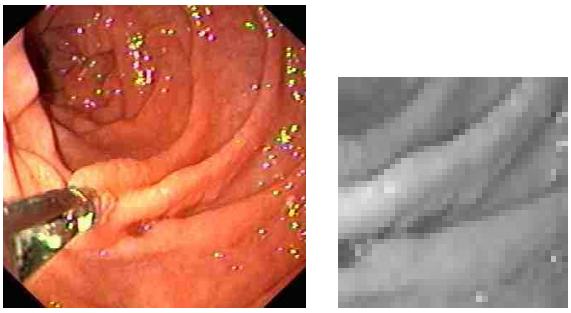
H. Perspective reconstruction (5 iters.).

Figure 5: The gastric angulus (cropped image size: 64×64 pixels). The perspectively reconstructed surface is visually the same at all iterations; we use this fact to show more viewing angles of the surface. The viewpoint in (C) and (D) is identical.

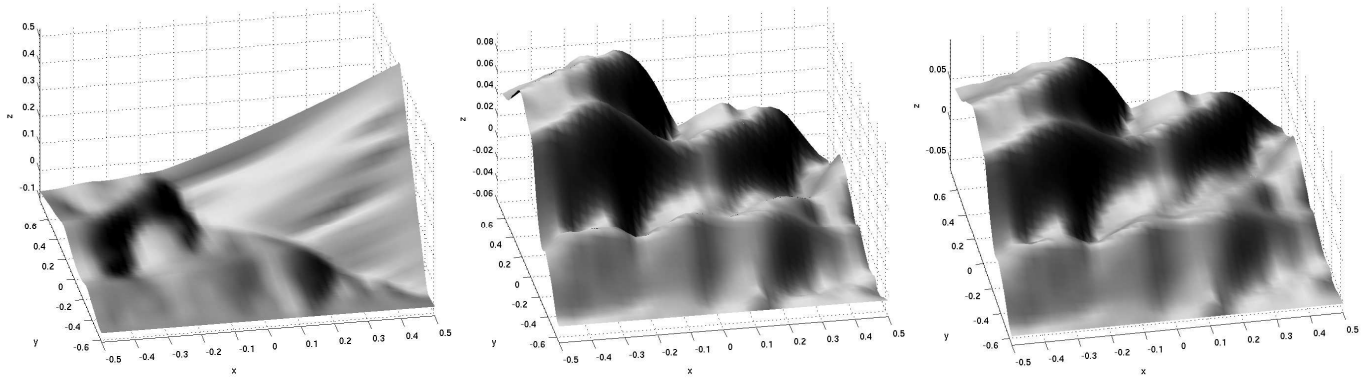
Figure 7C presents the reconstruction by the orthographic Fast Marching method, which recovered merely horizontal folds. Figs. 7D–H introduce perspective reconstruction. The appendiceal orifice was faithfully reconstructed.

Figure 8A shows the colon ascendens[3]. Figure 8B shows three plicae semicircularis (typical folds of the large intestine) from Fig. 8A. Even though this image visually resembles Fig. 5B, bear in mind that it is of a different part of the gastrointestinal tract. Figure 8C shows the orthographic reconstruction, while Figs. 8D–H, the perspective. The orthographic reconstruction produced horizontal and vertical folds which did not exist in the original image. Two of the main folds could difficultly be noticed in the reconstruction (center and bottom–left of Fig. 8C). Both of these folds suffered strong noise in the form of short vertical folds. In the perspective reconstruction, all three folds were prominent.

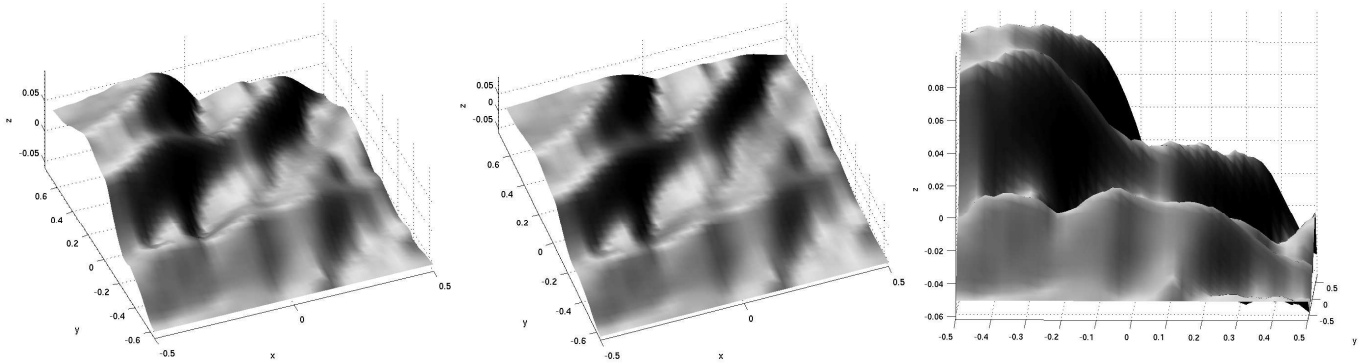
Center, San Salvador, El-Salvador.



A. Original. B. Cropped image.



C. Orthographic reconstruction. D. Perspective reconstruction (1 iter.). E. Perspective reconstruction (2 iters.).

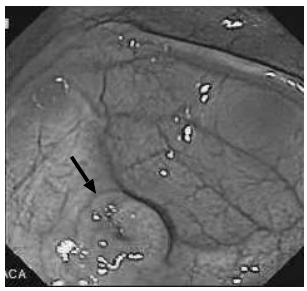


F. Perspective reconstruction (3 iters.). G. Perspective reconstruction (4 iters.). H. Perspective reconstruction (5 iters.).

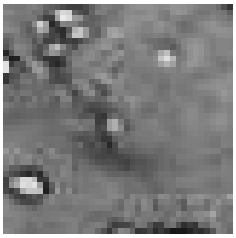
Figure 6: The descending duodenum (cropped image size: 40×40 pixels). The appearance of the perspective reconstruction is similar at all iterations; we use this fact to present more viewing directions of the surface. The viewpoint in (C) and (D) is identical.

Even though both algorithms use the same numerical methodology, the perspective Fast Marching appears to outrank the orthographic one. This suggests that the assumption of a perspective rather than an orthographic image irradiance equation yields an important improvement in reconstruction.

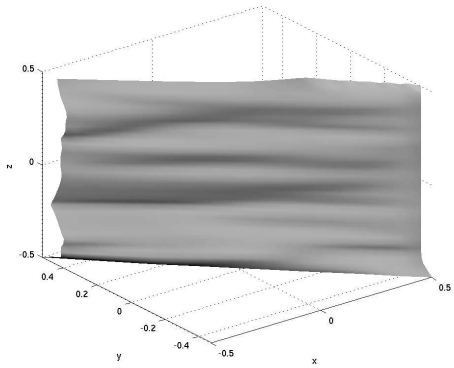
While many orthographic algorithms rival the best numerical way to solve the classic equation, the suggested one adopts its numerical scheme from Kimmel and Sethian [9] and thus avoids competition in the numerical arena. Instead, it demonstrates that the perspective equation may be better suited for the task of SfS.



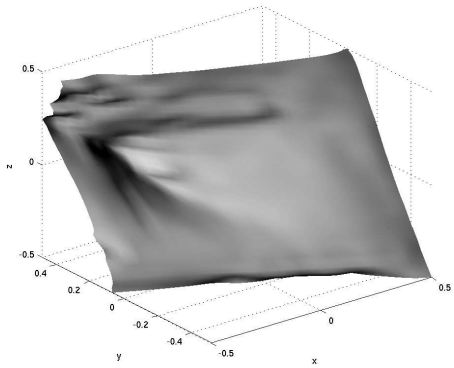
A. Original.



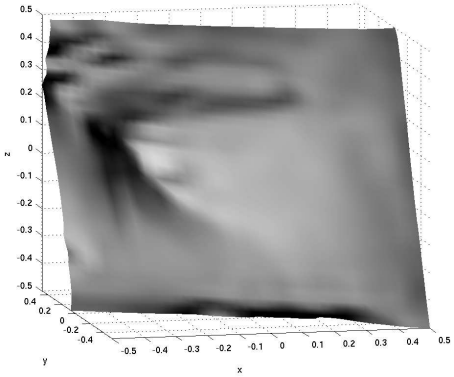
B. Cropped image.



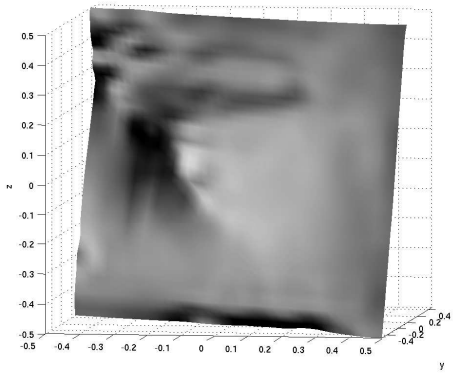
C. Orthographic reconstruction.



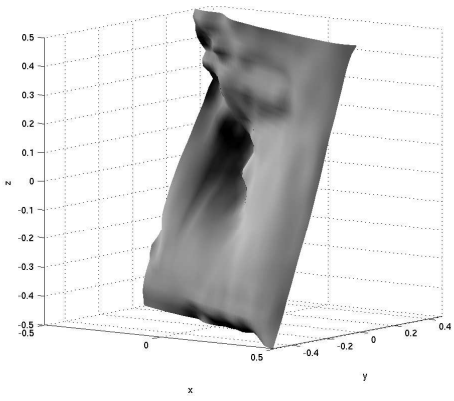
D. Perspective reconstruction (1 iter.).



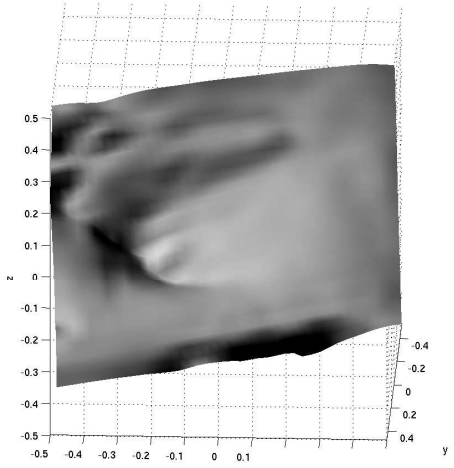
E. Perspective reconstruction (2 iters.).



F. Perspective reconstruction (3 iters.).



G. Perspective reconstruction (4 iters.).



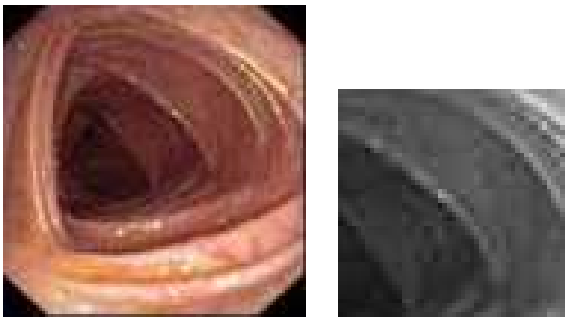
H. Perspective reconstruction (5 iters.).

Figure 7: An inverted appendix (cropped image size: 40×40 pixels). Perspective reconstruction is visually the same at all iterations; we exploit this to show more viewing angles of the reconstructed surface. The viewpoint in (C) and (D) is identical.

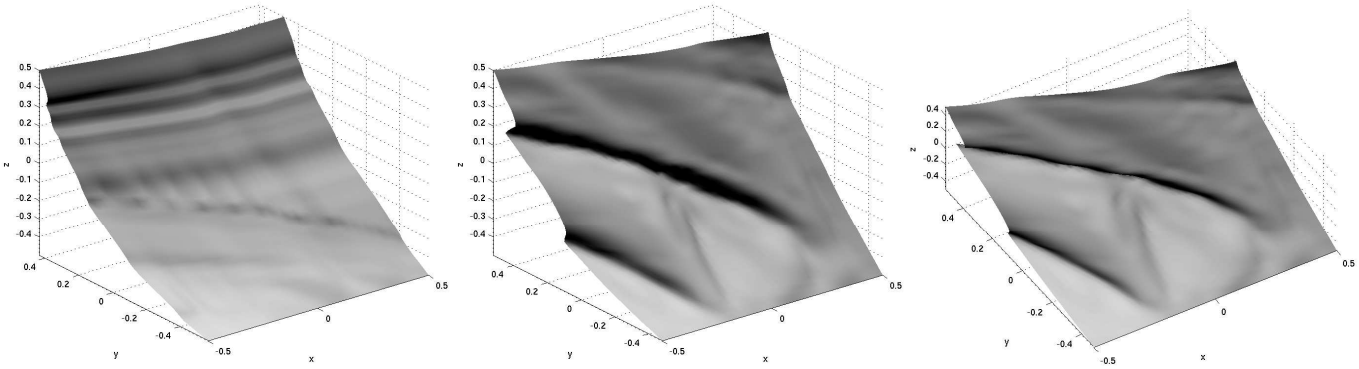
6. Conclusions

This research re-examined the roots of the field of Shape-from-Shading, the image irradiance equation. We reformulated the equation under the assumption of perspective projection and showed its dependence on the natural logarithm of the depth function. Based on this equation, a perspective variant of the Fast Marching method of Kimmel & Sethian [9] was developed.

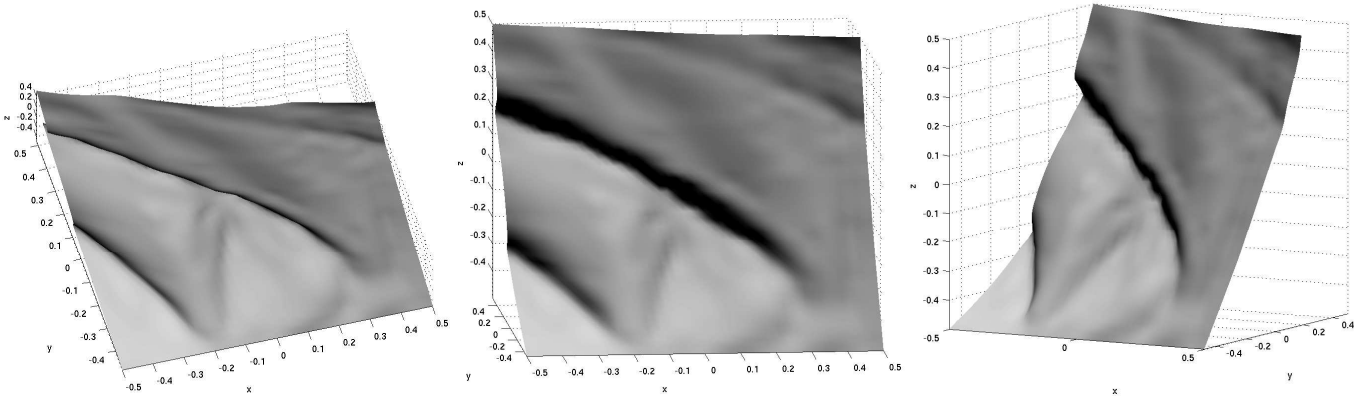
We compared three algorithms: the orthographic Fast Marching, the perspective Fast Marching and the perspective algorithm of Prados & Faugeras [18]. In general, the two perspective methods showed lower error rates than the orthographic, while they equated in accuracy with each another.



A. Original. B. Cropped image.



C. Orthographic reconstruction. D. Perspective reconstruction (1 iter.). E. Perspective reconstruction (2 iters.).



F. Perspective reconstruction (3 iters.). G. Perspective reconstruction (4 iters.). H. Perspective reconstruction (5 iters.).

Figure 8: The colon ascendens (cropped image size: 50×50 pixels). Perspective reconstruction is visually the same at all iterations; again, more viewing angles of the reconstructed surface are introduced this way. The viewpoint in (C) and (D) is identical.

Even though the perspective Fast Marching and Prados & Faugeras exhibited similar accuracy performance, the perspective Fast Marching has three important advantages. First and foremost, perspective Fast Marching needs orders of magnitude less iterations to converge than the algorithm of Prados & Faugeras (each iteration has similar complexity for both algorithms). Indeed, proving the convergence of the perspective Fast Marching is still an open issue, but in practice no more than 5 iterations were ever necessary (on both synthetic and real images). Second, the algorithm of Prados & Faugeras requires prior knowledge of the true depth on image boundary (Dirichlet boundary condition). This requirement is nontrivial and in practice cannot be obtained for real-life images. This requirement does not exist for the perspective Fast Marching method. Third, the algorithm of Prados & Faugeras uses the grid size on which the image was constructed. This data cannot be obtained unless exact camera parameters are available. Lack of these data (as is the case for the perspective Fast Marching

method) is equivalent to lack of knowledge of the true focal length of the camera.

Despite the comparison results described above, it is important to pay attention to the fact, that a comparison of perspective algorithms is more complex than in the orthographic case. Perspective algorithms produce their own $[xy]$ domain, which has to be fit to the original one. This scaling of the domain, along with scaling of the surface to fit the original (due to invariance to depth scaling) are sensitive to the method used for projecting one surface onto the other.

In addition, error measures common in the literature (see [32]), namely: mean and standard deviation of depth and gradient errors, are also sensitive to noise. This is due to their pixel-wise nature. Translation of a reconstructed feature of the surface by 1 pixel from the original may cause a drastic change in mean and standard deviation of the depth error, for example. In addition, the need for scaling described above increases the sensitivity of the measures. Their inaccuracy is further demonstrated by disagreements between different measures when comparing the perspective Fast Marching method with the algorithm of Prados & Faugeras (see Example #1, Sect. 5.2): some measures rendered the perspective Fast Marching method better, while others, that of Prados & Faugeras. Another important deficiency in these measures is their discordance with human visual inspection. As Example #5 shows, lower error rates according to these measures not necessarily reflect better correspondence to the original surface from a human point of view. Development of more adequate error measures is beyond the scope of the current paper.

As a result of the above, visual inspection remains a major evaluation technique. As such, fitness of SfS algorithms for real-life tasks should mainly be evaluated visually for the specific task under consideration. Zhang et al. [32] draws similar conclusions, saying synthetic images has low predictive power for real-life images.

To show the aptness of perspective SfS to real-life tasks, we compared reconstruction by the orthographic and perspective variants of the Fast Marching method on endoscopic images from different parts of the gastrointestinal channel. It appears that perspective SfS outperformed the orthographic Fast Marching method. As we compared two variants with a similar numerical basis, the results seem related to the underlying assumptions, rather than to the numerical methodology. Consequently, we infer that the perspective assumption yields a significant improvement in reconstruction.

From the practical point of view, the comparison demonstrated that perspective SfS could be used for real-life images. The application to endoscopy suggests that, unlike orthographic SfS, perspective SfS should be robust enough to handle real-life images.

7 Acknowledgements

We would like to thank Dr. Daniel Reisfeld from VisionSense for his inspiring discussions. We thank Mr. Emmanuel Prados and Prof. Olivier Faugeras from INRIA for providing an implementation of their algorithm [18]. We thanks Prof. David Levin from Tel-Aviv university for helpful discussions and assistance with Moving Least Squares [13], and for providing the code for it.

A Deriving The Perspective Image Irradiance Equation

We next develop the formula of the perspective image irradiance equation in image coordinates u, v.

Let us define a scene surface $S = \{(x, y, \hat{z}(x, y)) : (x, y) \in \overline{\Omega}_{scene}\}$ where Ω_{scene} is an open domain. Due to the perspective projection equations this surface can be written as: $S = \{(\frac{-uz}{f}, \frac{-vz}{f}, z(u, v)) : (u, v) \in \overline{\Omega}_{image}\}$ where Ω_{image} is an open domain. Let us assume the surface is differentiable with respect to (u, v) and also with respect to (x, y). The surface is depicted in Fig. 9.

In addition, let us assume that surface S is Lambertian and visible from all points of $\overline{\Omega}_{image}$ under the perspective projection model. A point light source at infinity illuminates the scene from direction: $\vec{L} = (-p_s, -q_s, 1)$. The image intensity in image coordinates is a function $I : \overline{\Omega}_{image} \longmapsto [0, 1]$, which maps the brightness of S as observed at point $(u, v, -f)$ to image coordinate (u, v).

Theorem 1 *Under the above definitions and assumptions, the perspective image irradiance equation is:*

$$I(u,v) = \frac{(u - fp_s)z_u + (v - fq_s)z_v + z}{\sqrt{1 + p_s^2 + q_s^2}\sqrt{(uz_u + vz_v + z)^2 + f^2(z_u^2 + z_v^2)}} \tag{13}$$

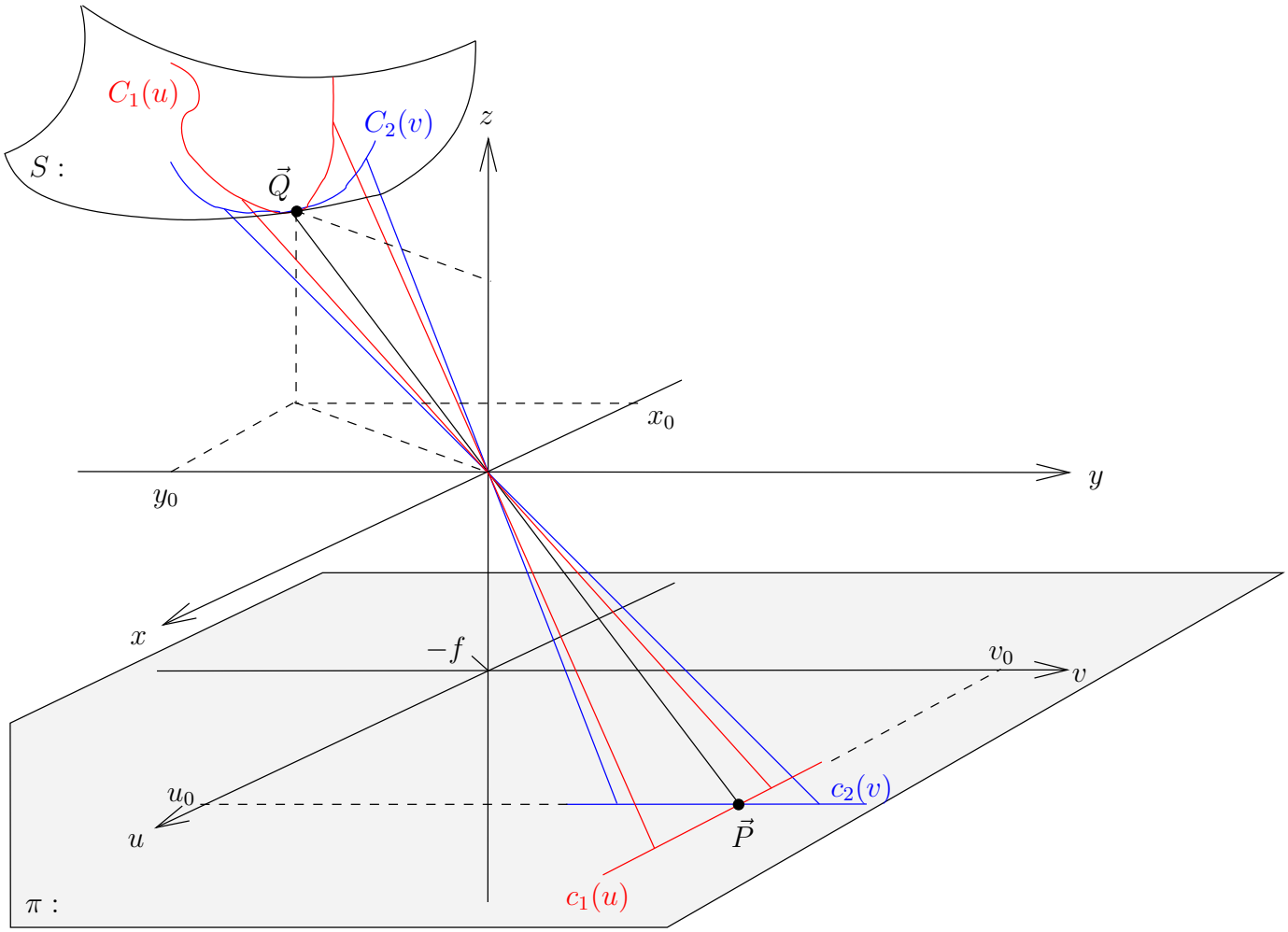


Figure 9: The image plane is $\pi = \{(u, v, -f)\}$ where f is the focal length. Point $\vec{Q} = (x_0, y_0, \hat{z}(x, y))$ on scene surface S is projected onto point $\vec{P} = (u_0, v_0, -f)$ on image plane π. The curves $c_1(u), c_2(v) \in \pi$ (red and blue) are parallel to axes x, y (or u, v). The curves $C_1(u), C_2(v) \in S$ (red and blue) are the curves on the object whose perspective projections on π are curves $c_1(u), c_2(v) \in \pi$, respectively. The tangents to $C_1(u), C_2(v)$ at point $\vec{Q}$ are computed from $\vec{P}$ and the perspective projection equations. The normal to the two tangents is the normal to S at point $\vec{Q}$.

Proof: Let us examine a curve on the projection plane:

$$c(s) \stackrel{\text{def}}{=} (u(s), v(s), -f)$$

with parameter s. This curve is the projection of a curve on the real-world surface S. Due to the perspective projection equations, the real-world curve can be written as:

$$C(s) = (-\frac{u(s)z(s)}{f}, -\frac{v(s)z(s)}{f}, z(s)) = \frac{z(s)}{f}(-u, -v, f)$$

A tangent to the real-world curve $C(s)$ is:

$$\frac{dC(s)}{ds} = \frac{1}{f}(-u_s(s)z(s) - u(s)z_s(s), -v_s(s)z(s) - v(s)z_s(s), fz_s(s)) \tag{14}$$

Now, let us consider two different curves through image point $\vec{P} = (u_0, v_0, -f)$ (where $(u_0, v_0) \in \Omega_{image}$). We define the

curves to be parallel to the x and y axes at the vicinity of $\vec{P}$. Thus,

$$\begin{aligned} c_1(u) &= (u, v_0, -f) \\ c_2(v) &= (u_0, v, -f) \end{aligned}$$

where u and v are the parameters of the curves (see Fig. 9). Substituting these curves into Eq. 14 provides two tangents to surface S at point $\vec{P}$:

$$\begin{aligned} \frac{dC_1(u)}{du} &= \frac{1}{f}(-z - uz_u, -vz_u, fz_u) \\ \frac{dC_2(v)}{dv} &= \frac{1}{f}(-uz_v, -z - vz_v, fz_v) \end{aligned}$$

A normal to the surface is therefore parallel to the cross product:

$$\vec{N} = \frac{dC_1(u)}{du} \times \frac{dC_2(v)}{dv} = \frac{z}{f^2}(fz_u, fz_v, uz_u + vz_v + z)$$

A unit normal is thus given by:

$$\hat{N} = \frac{(fz_u, fz_v, uz_u + vz_v + z)}{\sqrt{(uz_u + vz_v + z)^2 + f^2(z_u^2 + z_v^2)}}$$

The image irradiance equation thus becomes:

$$\begin{aligned} I(u,v) &= \hat{N} \cdot \hat{L} = \frac{(-p_s, -q_s, 1) \cdot (fz_u, fz_v, uz_u + vz_v + z)}{\sqrt{1 + p_s^2 + q_s^2}\sqrt{(uz_u + vz_v + z)^2 + f^2(z_u^2 + z_v^2)}} \\ &= \frac{(u - fp_s)z_u + (v - fq_s)z_v + z}{\sqrt{1 + p_s^2 + q_s^2}\sqrt{(uz_u + vz_v + z)^2 + f^2(z_u^2 + z_v^2)}} \end{aligned}$$

□

B Perspective Fast Marching

B.1 The Equation

We raise the image irradiance equation to the power of 2, and rearrange the terms:

$$p^2 A + q^2 B + 2pqC + 2pD + 2qE + F = 0$$

where:

$$\begin{aligned} A &\stackrel{\text{def}}{=} I^2\|L\|^2(u^2 + f^2) - (u - fp_s)^2 \\ B &\stackrel{\text{def}}{=} I^2\|L\|^2(v^2 + f^2) - (v - fq_s)^2 \\ C &\stackrel{\text{def}}{=} I^2\|L\|^2 uv - (u - fp_s)(v - fq_s) \\ D &\stackrel{\text{def}}{=} I^2\|L\|^2 u - (u - fp_s) \\ E &\stackrel{\text{def}}{=} I^2\|L\|^2 v - (v - fq_s) \\ F &\stackrel{\text{def}}{=} I^2\|L\|^2 - 1 \end{aligned}$$

We would like to have the left-hand side of this equation positive semidefinite. We therefore transfer non positive definite terms to the right-hand side:

$$p^2 A_1 + q^2 B_1 = p^2 A_2 + q^2 B_2 - 2pqC - 2pD - 2qE - F$$

where:

$$\begin{aligned} A_1 &\stackrel{\text{def}}{=} I^2\|L\|^2(u^2+f^2) \\ A_2 &\stackrel{\text{def}}{=} (u-fp_s)^2 \\ B_1 &\stackrel{\text{def}}{=} I^2\|L\|^2(v^2+f^2) \\ B_2 &\stackrel{\text{def}}{=} (v-fq_s)^2 \end{aligned}$$

Let us also define:

$$\hat{F} \stackrel{\text{def}}{=} p^2A_2 + q^2B_2 - 2pqC - 2pD - 2qE - F$$

so the equation becomes:

$$p^2A_1 + q^2B_1 = \hat{F} \tag{15}$$

where A_1 and B_1 are positive definite, by definition. It therefore also implies that $\hat{F} \geq 0$.

B.2 Solution of the Main Case

Similarly to [9], we estimate the directional derivatives by:

$$\begin{aligned} p_{ij} &\approx z_{ij} - z_1 \\ q_{ij} &\approx z_{ij} - z_2 \end{aligned}$$

where $z_{ij} \stackrel{\text{def}}{=} z(i\cdot\Delta u, j\cdot\Delta v)$ is the depth of the pixel (i,j), $z_1 \stackrel{\text{def}}{=} \min\{z_{i-1,j}, z_{i+1,j}\}$ and $z_2 \stackrel{\text{def}}{=} \min\{z_{i,j-1}, z_{i,j+1}\}$. Substituting into Eq. 15, we get:

$$A_1(z_{ij}-z_1)^2 + B_1(z_{ij}-z_2)^2 = \hat{F_{ij}}$$

where $\hat{F_{ij}} \stackrel{\text{def}}{=} \hat{F}(i\cdot\Delta u, j\cdot\Delta v)$. Solving this equation for z_{ij} yields:

$$z_{ij} = \frac{A_1z_1 + B_1z_2 \pm \sqrt{(A_1+B_1)\hat{F_{ij}} - A_1B_1(z_1-z_2)^2}}{A_1+B_1} \tag{16}$$

B.3 Solution of the Degenerate Cases

The degenerate cases of the solution (Eq. 16) result from negative discriminant $\Delta \stackrel{\text{def}}{=} (A_1+B_1)\hat{F}_{ij} - A_1B_1(z_1-z_2)^2 < 0$. This case can be written as:

$$|z_1 - z_2| > \sqrt{\frac{\hat{F}_{ij}}{A_1} + \frac{\hat{F}_{ij}}{B_1}} \tag{17}$$

We next consider three lemmas which solve the equation (Eq. 15) for the degenerate case. We adopt the following notation from [9]: $D_{ij}^{-u}z \stackrel{\text{def}}{=} \frac{z_{ij}-z_{i-1,j}}{\Delta u}$ is the standard backward derivative approximation and $D_{ij}^{+u}z \stackrel{\text{def}}{=} \frac{z_{i+1,j}-z_{ij}}{\Delta u}$ is the standard forward derivative approximation in the u-direction. $D_{ij}^{-v}z$ and $D_{ij}^{+v}z$ are defined in a similar manner for the v-direction. W.L.O.G we assume $\Delta u = \Delta v = 1$.

Lemma 1 *If $z_2 - z_1 > \sqrt{\frac{\hat{F}_{ij}}{A_1}}$, then $z_{ij} \stackrel{\text{def}}{=} z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}}$ is a solution of the equation:*

$$\left(\max\{D_{ij}^{-u}z, -D_{ij}^{+u}z, 0\}\right)^2 A_1 + \left(\max\{D_{ij}^{-v}z, -D_{ij}^{+v}z, 0\}\right)^2 B_1 = \hat{F}_{ij}$$

Proof:

The estimate of the u-derivative is:

$$\begin{aligned} \max\{D_{ij}^{-u}z, -D_{ij}^{+u}z, 0\} &= z_{ij} - \min\{z_{i-1,j}, z_{i+1,j}, z_{ij}\} \\ &= z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}} - \min\{z_1, z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}}\} \\ &= z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}} - z_1 = \sqrt{\frac{\hat{F}_{ij}}{A_1}} \end{aligned}$$

The estimate of the v-derivative is:

$$\max\{D_{ij}^{-v}z, -D_{ij}^{+v}z, 0\} = z_{ij} - \min\{z_{i,j-1}, z_{i,j+1}, z_{ij}\} = z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}} - \min\{z_2, z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}}\}$$

Now, because $z_2 - z_1 > \sqrt{\frac{\hat{F}_{ij}}{A_1}}$, it follows that:

$$\max\{D_{ij}^{-v}z, -D_{ij}^{+v}z, 0\} = z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}} - (z_1 + \sqrt{\frac{\hat{F}_{ij}}{A_1}}) = 0$$

If we substitute into Eq. 15:

$$\left(\max\{D_{ij}^{-u}z, -D_{ij}^{+u}z, 0\}\right)^2 A_1 + \left(\max\{D_{ij}^{-v}z, -D_{ij}^{+v}z, 0\}\right)^2 B_1 = \left(\sqrt{\frac{\hat{F}_{ij}}{A_1}}\right)^2 A_1 + 0^2 B_1 = \hat{F}_{ij}$$

□

Lemma 2 *If $z_1 - z_2 > \sqrt{\frac{\hat{F}_{ij}}{B_1}}$, then $z_{ij} \stackrel{\text{def}}{=} z_2 + \sqrt{\frac{\hat{F}_{ij}}{B_1}}$ is a solution of the equation:*

$$\left(\max\{D_{ij}^{-u}z, -D_{ij}^{+u}z, 0\}\right)^2 A_1 + \left(\max\{D_{ij}^{-v}z, -D_{ij}^{+v}z, 0\}\right)^2 B_1 = \hat{F}_{ij}$$

The proof is similar to that of Lemma 1.

Lemma 3 *If $\Delta < 0$ then necessarily either $z_2 - z_1 > \sqrt{\frac{\hat{F}_{ij}}{A_1}}$ or $z_1 - z_2 > \sqrt{\frac{\hat{F}_{ij}}{B_1}}$ holds. In other words, any degenerate case is contained in one of the cases of Lemmas 1 or 2.*

Proof:

By definition: $A_1 > 0$ and $B_1 > 0$. Therefore, any $\hat{F}_{ij}$ which satisfies Eq. 15 is positive: $\hat{F}_{ij} > 0$.

From Eq. 17, if $\Delta < 0$ then $|z_1 - z_2| > \sqrt{\frac{\hat{F}_{ij}}{A_1} + \frac{\hat{F}_{ij}}{B_1}}$. Hence, $|z_1 - z_2| > \sqrt{\frac{\hat{F}_{ij}}{A_1} + \frac{\hat{F}_{ij}}{B_1}} > \max\{\sqrt{\frac{\hat{F}_{ij}}{A_1}}, \sqrt{\frac{\hat{F}_{ij}}{B_1}}\}$. We distinguish three cases:

1. If $z_1 > z_2$, then $z_1 - z_2 = |z_1 - z_2| > \sqrt{\frac{\hat{F}_{ij}}{A_1} + \frac{\hat{F}_{ij}}{B_1}} > \sqrt{\frac{\hat{F}_{ij}}{B_1}}$.
2. If $z_1 < z_2$, then $z_2 - z_1 = |z_1 - z_2| > \sqrt{\frac{\hat{F}_{ij}}{A_1} + \frac{\hat{F}_{ij}}{B_1}} > \sqrt{\frac{\hat{F}_{ij}}{A_1}}$.
3. If $z_1 = z_2$, then $\Delta = (A_1 + B_1)\hat{F}_{ij} > 0$.

It follows, that in any case where $\Delta < 0$, either $z_2 - z_1 > \sqrt{\frac{\hat{F}_{ij}}{A_1}}$ or $z_1 - z_2 > \sqrt{\frac{\hat{F}_{ij}}{B_1}}$ holds.

□

Lemmas 1 and 2 found solutions for the cases $z_2 - z_1 > \sqrt{\frac{\hat{F}_{ij}}{A_1}}$ and $z_1 - z_2 > \sqrt{\frac{\hat{F}_{ij}}{B_1}}$, respectively. Lemma 3 showed that these cases contain the degenerate case ($\Delta < 0$), which means that the solutions introduced by Lemmas 1 and 2 cover the degenerate case $\Delta < 0$.

References

[1] M. Bichsel and A. P. Pentland. A simple algorithm for shape from shading. In *Computer Vision and Pattern Recognition*, pages 459–465, 1992.

[2] M. J. Brook and W. Chojnacki. Direct computation of shape from shading. In *Proceedings of the International Conference on Pattern Recognition*, pages 114–119, Israel, October 1994.

[3] B. Horn. Obtaining shape from shading information. In P. H. Winston, editor, *The Psychology of Computer Vision*, Computer Science Series, chapter 4, pages 115–155. McGraw-Hill Book Company, 1975.

[4] B. K. P. Horn. Image intensity understanding. *Artificial Intelligence*, 8(2):201–231, Apr. 1977.

[5] B. K. P. Horn. *Robot Vision*. The MIT Press/McGraw-Hill Book Company, 1986.

[6] B. K. P. Horn and M. J. Brooks, editors. *Shape from Shading*. The MIT Press, 1989.

[7] H. Ishii. A simple, direct proof of uniqueness for solutions of the hamilton-jacobi equations of eikonal type. *Proceedings of the American Mathematical Society*, 100(5):247–251, June 1987.

[8] R. Kimmel and A. M. Bruckstein. Global shape from shading. *Computer Vision and Image Understanding*, 62(3):360–369, Nov. 1995.

[9] R. Kimmel and J. A. Sethian. Optimal algorithm for shape from shading and path planning. *Journal of Mathematical Imaging and Vision*, 14(3):237–244, 2001.

[10] C.-H. Lee and A. Rosenfeld. Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence*, 26:125–143, 1985.

[11] K. M. Lee and C.-C. J. Kuo. Shape from shading with a linear triangular element surface model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(8):815–822, Aug. 1993.

[12] K. M. Lee and C.-C. J. Kuo. Shape from Shading with a Generalized Reflectance Map Model. *Computer Vision and Image Understanding*, 67(2):143–160, Aug. 1997.

[13] D. Levin. Mesh-independent surface interpolation. In G. Brunnett, B. Hamann, H. Mueller, and L. Linsen, editors, *Geometric Modeling for Scientific Visualization*, Mathematics and Visualization. Springer-Verlag, 2004.

[14] P.-L. Lions. *Generalized Solutions of Hamilton-Jacobi Equations*. Pitman, London, 1982.

[15] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[16] M. A. Penna. A Shape from Shading Analysis for a Single Perspective Image of a Polyhedron. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):545–554, June 1989.

[17] A. P. Pentland. Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):170–187, Mar. 1984.

[18] E. Prados and O. Faugeras. Perspective shape from shading and viscosity solutions. In *Proceedings of the $9^{\underline{th}}$ IEEE International Conference on Computer Vision*, volume 2, pages 826–831, October 2003.

[19] E. Prados, O. Faugeras, and E. Rouy. Shape from shading and viscosity solutions. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, $7^{\underline{th}}$ *European Conference on Computer Vision*, volume II, pages 790–804, Copenhagen, Denmark, May 2002.

[20] A. Robles-Kelly and E. R. Hancock. Model acquisition using shape-from-shading. In F. J. Perales and E. R. Hancock, editors, *The $2^{\underline{nd}}$ International Workshop on Articulated Motion and Deformable Objects*, pages 43–55, Palma de Mallorca, Nov. 2002. Springer.

[21] E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal of Numerical Analysis*, 29(3):867–884, June 1992.

[22] D. Samaras and D. Metaxas. Illumination constraints in deformable models for shape and light direction estimation. *IEEE Transactions on Pattern Analisys and Machine Intelligence*, 2003.

[23] I. Seong, S. Hideo, and O. Shinji. A Divide-and-conquer Strategy in Shape from Shading Problem. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 413–419, 1997.

[24] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Science of the USA*, 93:1591–1595, Feb. 1996.

[25] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge Monograph on Applied and Computational Mathematics. Cambridge University Press, 2 edition, 1999.

[26] A. Tankus, N. Sochen, and Y. Yeshurun. A new perspective [on] Shape-from-Shading. In *Proceedings of the $9^{\underline{th}}$ IEEE International Conference on Computer Vision*, volume II, pages 862–869, Nice, France, Oct. 2003.

[27] A. Tankus, N. Sochen, and Y. Yeshurun. Perspective shape-from-shading. Technical Report TR-001-2004, Tel-Aviv University, 2004.

[28] P.-S. Tsai and M. Shah. Shape from shading using linear approximation. *Image and Vision Computing*, 12(8):487–498, Oct. 1994.

[29] I. Weiss. A perspective 3D formalism for shape from shading. In *Proceedings of DARPA Image Understanding Workshop*, volume 2, pages 1393–1402, 1997.

[30] S. M. Yamany, A. A. Farag, E. Rickard, D. Tasman, and A. G. Farman. A Robust 3-d Reconstruction System for Human Jaw Modeling. In *Proceedings of the Second International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 778–787, Berlin, Germany, 1999. Springer-Verlag.

[31] S. Y. Yuen, Y. Y. Tsui, Y. W. Leung, and R. M. M. Chen. Fast marching method for shape from shading under perspective projection. In V. J. J., editor, *Proceedings of Second IASTED International Conference Visualization, Imaging, and Image Processing*, pages 584–589, Malaga, Spain, 2002. ACTA Press, Anaheim, CA, USA.

[32] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–705, August 1999.

[33] W. Zhao and R. Chellappa. Face recognition using symmetric shape from shading. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 4, pages 286–293, Hilton Head, SC, June 2000.

[34] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):680–702, July 1991.