Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 806–824
Pattern Recognition Letters
www.elsevier.com/locate/patrec

# A fast marching formulation of perspective shape from shading under frontal illumination

Shiu Yin Yuen *, Yuen Yan Tsui, Chi Kin Chow

*Department of Electronic Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong, China*

Received 20 May 2004; received in revised form 3 March 2006
Available online 26 January 2007

Communicated by S. Dickinson

**Abstract**

An adaptation of the fast marching method to the perspective Shape from Shading under frontal illumination is proposed. A heuristic is proposed to handle occlusion. The method outperforms Tankus et al. [Tankus, A., Sochen, N., Yeshurun, Y., 2005. Shape-from-shading under perspective projection. Internat. J. Comput. Vision 63(1), 21–43] in both time and accuracy. Accurate methods for generating testing images are also reported.


*Keywords:* Shape from shading; Perspective projection; Fast marching; Depth discontinuity; Intensity discontinuity; Occlusion

## 1. Introduction

This paper deals with the reconstruction of a single 2D image as a 3D shape: the *shape from shading* problem. In computer vision, shape recovery is a classic problem. It aims to reconstruct a 3D shape from one or more intensity images. Such techniques are collectively known as Shape-from-X, where X includes shading, stereo, motion, texture, etc. Shape from shading (SfS) uses the pattern of lights and shades in a single image to infer the shape of the surface in view. In many studies, the Lambertian model, a known lighting direction and orthographic projection are assumed. However, it is clear that the human perception system is less restrictive.

Horn (1975) pioneered the use of shape from shading in computer vision. Reported techniques can be classified into four categories: *local*, *linear*, *minimization* and *propagation* (Zhang et al., 1999).

Local approaches recover the shape based on certain surface type assumptions. Postulating a locally spherical surface at each point, Pentland (1984) reconstructs the surface from the intensity and its first and second derivatives. However, how to make a correct guess for the surface type at each point is an open problem.

An example of the linear approach, first proposed by Pentland (1990), is Zhang et al. (1999). First order discrete approximation is applied to the shape from shading equation. Making the simplifying assumption that during the iteration, the neighboring pixels have the correct depths, a non-linear equation of a single variable is written for each pixel, which is then solved using Newton's iteration. Recently, Ikeda (2002) reports a more correct version of the Newton method that allows the depths of the neighboring pixels to be variables as well. A problem of both Zhang et al. (1999) and Ikeda (2002) is that the performance is heavily dependent on a good initialization.

Minimization approaches, pioneered by the early work of Ikeuchi and Horn (1981), attempt to optimize an energy function to obtain shape information. Brooks and Horn (1985) minimize the energy function composed of the brightness and the smoothness constraints. The two constraints are combined using variational calculus. This approach suffers from three problems. Firstly, the regular-

* Corresponding author. Tel.: +852 27887717; fax: +852 27887791.
*E-mail address:* kelviny.ee@cityu.edu.hk (S.Y. Yuen).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.11.008

ization constant is given by a heuristic rule of thumb. Secondly, it may provide over-smooth solutions. Thirdly, like the Newton's method above, there is no guarantee of convergence (an exception is Crouzil et al., 2003). Worthington and Hancock (1999) proposed a robust regularizer that partially solves the over-smooth problem, but then the regularizer has its own set of parameters to be determined.

In our opinion, the propagation approach offers the best method to date to solve the shape from shading problem. An early attempt using the characteristic strip approach (Horn, 1975) is susceptible to numerical errors. Better methods have since been reported: Bichsel and Pentland (1992) use the minimum downhill principle to propagate on a discrete grid. Rouy and Tourin (1992) obtain a solution for the shape from shading problem based on the viscosity theory and the dynamic programming principle. Dupuis and Oliensis (1992) formulate the shape from shading problem as an optimal control problem. Both approaches are iterative. Though both are guaranteed to converge correctly, the convergence speed is unknown. This problem is solved by Verbeek and Verwer (1990) and Kimmel and Sethian (2001). They reported a deterministic, non-iterative, algorithm of complexity $O(N \log N)$ for a unique surface reconstruction, where $N$ is the number of pixels in the image. The solution is based on the fast marching method (Sethian, 1996, 1999; Osher and Paragios, 2003). It is stable, monotonic, and offers a well posed solution up to the unknown depths at each local minimum (or concave) singular point, which are the brightest points in the image.

Strictly speaking, the fast marching method is only applicable for light illumination in the frontal direction, in which case the shape from shading equation is an Eikonal equation. If the lighting direction is oblique, a light source coordinate transformation is needed to transform the shape from shading equation to an approximate Eikonal equation. Recently, Prados et al. (2002) reported a method based on the viscosity theory that can handle oblique lighting directions as well. However, the approach is iterative. In this paper, we shall restrict our analysis to the case when the illumination direction is frontal. Recently, Chow and Yuen (submitted for publication) show that under perspective projection, any oblique illumination scenario is equivalent to a frontal illumination scenario.

In all of the above works, orthographic projection is assumed. In reality, human beings are often interpreting photographs taken under perspective projection. Thus it is instructive to extend the shape from shading formulation to the perspective case. It also provides a way to link up the work with shape from stereo in the future. Courteille et al. (2004) showed that perspective shape from shading can be used to unwrap a non-flat page of an open book. The application in this case clearly cannot be accomplished by shape from shading with orthographic projection. Tankus et al. (2005) also compared shape from shading using perspective and orthographic fast marching on endoscopic images taken from different parts of the gastrointestinal channel. They conclude that both the qualitative and quantitative shapes obtained with perspective shape from shading are better than that obtained with orthographic shape from shading.

Due to its solid mathematical foundation and practical merits, it is desirable to extend the fast marching method described above to perspective projection. In this paper, such an extension will be made.

The following assumptions are made in this paper:

1. The surface is opaque and Lambertian with constant albedo.
2. The image is perspectively projected.
3. The light source is far away so that it can be modeled as a lighting direction. Moreover, a frontal light source is assumed.
4. The camera is assumed to have been calibrated.
5. The heights of the local minimum (or concave) singular points are assumed to have been determined (e.g. by other methods such as stereo).

A preliminary version of this paper appeared in (Yuen et al., 2002). In that paper, we gave a novel formulation of the perspective shape from shading through a change of variables. Then we showed that it can be directly applied to the fast marching method developed for shape from shading under orthographic projection (Kimmel and Sethian, 2001). In this paper, we give a theoretical justification of our approach (see the Lemma in Section 2.3) and corrected some of the detailed equations (with no change to the basic idea). One of the well known problems with perspective projection is that it would generate occlusion, which manifests as intensity discontinuity in the image. Surprisingly, it is ignored by all the works on perspective shape from shading published so far (see below for a survey). In this paper, we propose a heuristic rule within the framework of fast marching to handle such occlusion. In (Yuen et al., 2002), only experimental results on synthetic images has been reported. In this paper, experimental result on a real image is also reported.

The pervious works on shape from shading using perspective projection (Lee and Kuo, 1994, 1997; Hasegawa and Tozzi, 1996; Yamany et al., 2000) are based on Newton's iteration. A well known problem of these methods is that they will not converge to the correct solution unless the initialization is very close to the correct solution. In contrast, the perspective fast marching method converges to the correct solution, is fast, and is deterministic. Further experimental support against Newton's type iteration schemes can be found in (Tsui, 2003). We experimented with Nelder and Mead type iteration but the experimental results show that the algorithm fails to converge properly.

Recently, Tankus et al. (2005) proposed an alternative method of perspective recovery. Their method is also based on the fast marching. This method supersedes an earlier algorithm by the same authors using gradient descent (Tankus et al., 2003). In their opinions, our approach suffers two drawbacks. "First, it 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 non-uniform grid, while the underlying assumption was a uniform one (image space uniformity)." (Tankus et al., 2005). With due respect to Tankus et al., it seems to us that the above are incorrect. Firstly, our approach is simply an application of the Eikonal equation with a change of variables, then use a first order discrete approximation of the normal in terms of the changed variables. In fact, both formulations would involve a first order numerical approximation at some stage. Secondly, fundamentally, the fast marching method does not require a uniform grid (Sethian, 1999, p. 98). This fact is also verified by experiments: Our experimental results on synthetic images show that our perspective fast marching method also gives the correct results. Further theoretical justification for our fast marching method under perspective projection can be found in the Lemma in Section 2.3.

Tankus et al. (2005) assume that the photographed surfaces $Z$ are representable by both its 3D coordinates $Z_S(X,Y)$ (S stands for scene) and its image grid coordinates $Z_I(u,v)$ (I stands for image). This is not necessarily the case for perspective projection. Occlusion may be generated due to perspective projection and the relation is not 1–1. An intensity discontinuity will be generated as a result. In this paper, a heuristic is proposed in Section 3 to handle this situation. Moreover, the formulation of Tankus et al. requires an iterative solution within the fast marching scheme. It is more computationally intensive and they provide no proof on the correct convergence and rate of convergence, though in their numerical experiments, they report that no more than five iterations are necessary. Our formulation requires no such iterative solution. Thus it does not suffer from this complication. An extensive comparison with the method of Tankus et al. is conducted in Section 4.

Another method for perspective shape from shading has also recently been proposed by Prados and Faugeras (2003). Tankus et al. has compared their method with Prados' method in details. In general, both methods give similar results, but Prados and Faugeras' method require more iterations. One difficulty of Prados and Faugeras (2003) is that the heights at the image boundary have to be known. Recently, Prados and Faugeras (2004) relax this requirement by applying state constraints and reported that the number of iterations can be reduced by three orders of magnitude by starting from a supersolution. The fast marching, used in this paper, is a single-pass non-iterative approach, which is fundamentally a different approach compared with the iterative approach of Prados and Faugeras. A comparison with this new method is out of scope of the current paper.

Another approach is reported by Courteille et al. (2004), who considered the Eikonal equation obtained under frontal illumination and perspective projection. They used a priori shape information to solve the equation.

The paper is organized as follows. In Section 2, the fast marching method is extended to perspective projection by a novel formulation that accounts for the perspective foreshortening effect. A proof that the fast marching is applicable to perspective projection with no occlusion is also constructed. The original fast marching method implicitly assumes that the surface is continuous (but not necessarily smooth). However, in real surfaces, jumps occur at depth discontinuities. In perspective projection, new intensity discontinuities may be generated due to occlusion. A simple empirical modification to the fast marching method that apparently overcomes this problem is reported in Section 3. Experimental results are reported in Section 4. To test the algorithm properly, two methods to generate perspective images accurately are reported. Finally, a conclusion is given in Section 5.

## 2. Shape from shading

### 2.1. Fast marching method

Sethian (1996) introduces the Fast Marching Method. It applies to phenomena that can be described as a wave front propagating (normal to itself) with a speed $F = F(i,j)$. The main idea is to methodically construct the solution using only upwind values. Let the solution surface $T(i,j)$ be the time at which the curve crosses the point $(i,j)$, then it satisfies the *Eikonal equation* $|\nabla T|F = 1$. The physical meaning is that the gradient of arrival time of the front is inversely proportional to the speed of the front. For an upwind scheme, the approximation to the gradient is written as

$$[\max(D_{ij}^{-x}T, -D_{ij}^{+x}T, 0)^2 + \max(D_{ij}^{-y}T, -D_{ij}^{+y}T, 0)^2]^{\frac{1}{2}} = \frac{1}{F_{ij}} \tag{1}$$

where the difference operator notation, for example $D_{ij}^{+x}T = (T_{i+1,j} - T_{i,j})/(\Delta x)$, is employed. A key insight in this scheme is that to solve this non-linear system, an optimal ordering of the grid points can be found. Starting at the singular points (also called boundary points), which are sources of the wave, the method marches the front outwards one grid point at a time. In detail, the method is as follows:

1. Assign the $T$ values of all local minimum singular points. Label the singular points as *Known*.
2. Label as *Trial* all points that are one grid point away. Calculate the $T$ value of the *Trial* points.
3. Label as *Far* all other grid points.
4. Determine the trial point $A$ that has the smallest $T$ value.

5. Remove $A$ from *Trial* and add it to *Known*.
6. Label as *Trial* all neighbors of $A$ that are not already *Known*, i.e., if the neighbor is in *Far*, removes, and add to the set *Trial*.
7. If all points are labeled as *Known*, then exit, else goto step 4.

Note that the $T$ values of those local minimum singular points that are sources need to be given (step 1), and the point with the smallest $T$ is updated (step 4). The later is known as the *entropy condition*. This condition is a consequence of the Fermat's principle of least time, taking the physical analogy that the propagating particles are light particles (Verbeek and Verwer, 1990). Computationally, the fast marching is the same as the well known A* algorithm.

### 2.2. Shape from shading using fast marching

Kimmel and Sethian (2001) applied the fast marching method to shape from shading under orthographic projection. (Verbeek and Verwer (1990) use essentially the same idea but with a cruder update scheme.) The analogy is that grid points are image pixels and time $T$ is the depth. As the lighting direction is $L = (0, 0, 1)$, for image pixel $(X, Y)$ with depth $Z_S(X, Y)$, the normalized intensity $I_S$ of the classical shape from shading equation is

$$I_S(X,Y) = \frac{N \cdot L}{|N|} = \frac{N_2(X,Y)}{\sqrt{N_0(X,Y)^2 + N_1(X,Y)^2 + N_2(X,Y)^2}}$$

or

$$|\nabla Z_S(X,Y)| = \sqrt{\left(\frac{N_0}{N_2}\right)^2 + \left(\frac{N_1}{N_2}\right)^2} = \sqrt{\frac{1}{I_S(X,Y)^2} - 1} \stackrel{\text{def}}{=} \frac{1}{F_S(X,Y)} \tag{2}$$

where $N = (N_0, N_1, N_2)$ is the normal vector to the surface at point $P(X, Y, Z)$. Eq. (2) is an Eikonal equation.

The singular points are subsets of points with $I_S = 1$, the brightest points in the image. They need to satisfy the additional condition that they are also the local minimum (concave points) in depth as well.

### 2.3. Perspective projection

A perspective image of a 3D object suffers from *foreshortening*. Let a Cartesian coordinate system with the origin at the center of projection be the reference coordinate system. The 3D point $P = (X, Y, Z)$ is related to the image point $p = (u, v, f)$ by the perspective projection equations: $X = \frac{u}{f} Z_S$ and $Y = \frac{v}{f} Z_S$, where $f$ is the focal length. (Since the camera has been calibrated, we assume that $f$ is known.) Since the surface is Lambertian, $I_S(X, Y) = I_I(u, v)$ and $\frac{1}{F_S(X,Y)} = \frac{1}{F_I(u,v)}$. After a change of variables,

$$\sqrt{\left(\frac{N_0(u,v)}{N_2(u,v)}\right)^2 + \left(\frac{N_1(u,v)}{N_2(u,v)}\right)^2} \stackrel{\text{def}}{=} \frac{1}{F_I(u,v)} \tag{3}$$

Note that the physical meaning of the Eikonal equation remains unchanged. The following lemma guarantees that the fast marching method can be applied in most situations under perspective projection.

**Lemma 1.** *If $Z_S(X, Y)$ is continuous and the perspective projection $\Pi$ is bijective, the fast marching method can be applied on the $(u, v) = f\left(\frac{X}{Z_S}, \frac{Y}{Z_S}\right)$ plane.*

**Proof.** As $f$ is a positive constant, $Z_S(X, Y) > 0$ is continuous and $\Pi$ is bijective, $(X, Y)$ to $(u, v)$ is a continuous function. Suppose two points $P_1(X_1, Y_1, Z_1)$ and $P_2(X_2, Y_2, Z_2)$ satisfies $Z_2 > Z_1$, then $P_1$ and $P_2$ lie on level sets (or wavefronts) $L_1$ and $L_2$ with heights (or time) $Z_1$ and $Z_2$. Moreover, since the Eikonal equation always have positive costs, i.e. the height (time) variable is monotonically increasing, $L_1$ and $L_2$ are simple closed contour with no self intersections; has no intersections with each other; and $L_2$ encloses $L_1$. Denote projected entities by small letters. Since the perspective projection is bijective, the enclosure property is satisfied. Thus projected contour $l_2$ encloses $l_1$. Now let $Z_2 - Z_1 \to 0$. Since the transformation is continuous, $l_2 \to l_1$. This shows that (1) if a point is a neighbor of another point, it is also a neighbor of the point after the projection; (2) the order of propagation from the projected points is the same as the order of projection from the original point, i.e. $p_1$ propagates to $p_2$. Thus the fast marching can be applied on the $(u, v)$ plane after incorporating the change of variables. □

**Remarks**

1. In any discrete computer simulation of the continuous problem, one has to make numerical approximation. In this paper, a first order approximation of the normal and the four-neighborhood is used. Using the four-neighborhood, we necessarily define them to be neighbors in the discrete approximation. From the above, we know that neighborhood property is preserved. This shows that the Eikonal equation can be applied after a change of variables. However, we are non-uniformly sampling the $X$–$Y$ plane, as pointed out by Tankus et al. (2005). This, however, is the correct way to do given that we only have the image and we wish to make a numerical approximation (cf. numerical approximation is also needed in (Tankus et al., 2005) in their approximation of the partial derivatives $\frac{\partial Z}{\partial u}$ and $\frac{\partial Z}{\partial v}$).
Due to the non-uniform sampling nature, it is theoretically possible that even if $Z_1 > Z_3$, $Z_1$ may be the actual propagating source because $P_1$ is closer to $P_0$. To allow for this scenario, we can use a fast marching scheme such that all the four combinations: $(Z_1, Z_2)$, $(Z_1, Z_4)$, $(Z_3, Z_2)$, $(Z_3, Z_4)$ are used to calculate $Z_0$, and taking the smallest $Z_0$. Note: This conforms with the fast

marching principle – the Fermat's principle of least time. In our experiments, we have tried out the above scheme. However, it always gives the same result as the scheme which takes the smallest $Z$ as the propagating source. It suggests that the possibility of this particular change of order is rare in reality. Therefore, we use the simple way to determine $a$ and $b$ in step (4a) of Algorithm 1, though the above scheme is the more correct method.

2. If $Z_S$ is discontinuous at $(X, Y)$, it will be discontinuous at $(u, v)$. The fast marching will not give a meaningful numerical result at a depth discontinuity as the intensity is zero.
3. If there is occlusion, $\Pi$ will not be 1–1, and there is an intensity discontinuity newly generated due to the perspective projection. Likewise, the first marching is expected to encounter difficulty. This is illustrated in Fig. 1.

Below we give a first order discrete approximation implementation of the fast marching method in perspective projection.

### 2.4. Discrete formulation

Depending on which two of the four-neighboring points of image point $p_0(u_0, v_0, f)$ are selected, the corresponding outward normal $N$ of the 3D surface at point $P_0(X_0, Y_0, Z_0)$ shown in Fig. 2 is computed as the following:

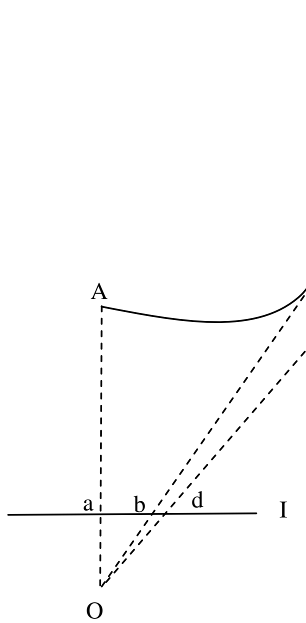


Fig. 1. In the figure, $I$ and $O$ is the image plane and the center of projection; the perspective projection is not 1–1, as $B$ and $C$ both project to the same point. Segment $AC$ and $CD$, belonging to two different parts of the surface, will be adjacent after projection. There will be a new discontinuity at $b$ due to occlusion and the fast marching cannot be applied there.

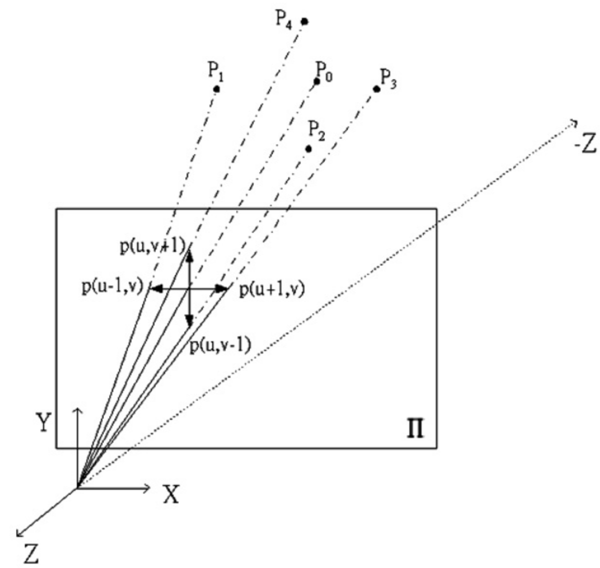


Fig. 2. The image coordinate system with its origin in the center of perspective projection.

We define the horizontal and vertical neighboring points as

$$\vec{P}_a \stackrel{\text{def}}{=} \left(\frac{u_a z_a}{f},\ \frac{v_a z_a}{f},\ z_a\right) \quad \text{and} \quad \vec{P}_b \stackrel{\text{def}}{=} \left(\frac{u_b z_b}{f},\ \frac{v_b z_b}{f},\ z_b\right)$$

respectively where $a \in [1, 3]$ and $b \in [2, 4]$. The normal $N$ is obtained as

$$\begin{aligned}
\vec{N} &= (\vec{P}_a - \vec{P}_0) \times (\vec{P}_b - \vec{P}_0) \\
&= \left(\frac{u_a z_a - u_0 z_0}{f}, \frac{v_a z_a - v_0 z_0}{f}, z_a - z_0\right) \\
&\quad \times \left(\frac{u_b z_b - u_0 z_0}{f}, \frac{v_b z_b - v_0 z_0}{f}, z_b - z_0\right) \\
&= \frac{1}{f}\begin{bmatrix} z_0(z_a(v_0 - v_a) - z_b(v_0 - v_b)) + z_a z_b (v_a - v_b) \\ z_0(z_b(u_0 - u_b) - z_a(u_0 - u_a)) + z_a z_b (u_b - u_a) \\ \frac{z_0}{f}(z_b(u_b v_0 - u_0 v_b) + z_a(u_0 v_a - u_a v_0)) + \frac{z_a z_b}{f}(v_b u_a - u_b v_a) \end{bmatrix}^{\mathrm{T}} \\
&\stackrel{\text{def}}{=} \frac{1}{f}\begin{bmatrix} z_0 A_1 + B_1 \\ z_0 A_2 + B_2 \\ z_0 A_3 + B_3 \end{bmatrix}^{\mathrm{T}}
\end{aligned}$$

And the shape from shading equation becomes

$$I = \frac{|z_0 A_3 + B_3|}{\sqrt{(z_0 A_1 + B_1)^2 + (z_0 A_2 + B_2)^2 + (z_0 A_3 + B_3)^2}} \tag{4}$$

### 2.5. Algorithm

For the perspective case, $(u, v)$ are the grid points, and their depths $Z_I(u, v)$ are the time. Let $\Delta u = 1$ and $\Delta v = 1$. In the initialization, all $Z_I(u, v)$ are set to infinity except at the local minimum singular points, which are assigned their true depth values. (Infinity does not literally mean that the depth is infinity, but that the depth is unknown.) If the perspective projection is a continuous function (see

Lemma in Section 2.3), the simple four-neighborhood is used. (The fast marching method can be designed to work on other neighborhood, as well as higher order schemes, but we shall not pursue this here.) The detailed algorithm for fast marching under perspective projection is shown below:

**Algorithm 1**
***Input***: (a) Intensity Image. (b) Initial depth values of a set of local minimum singular points.

1. Label the local minimum singular points as *Known*. Assign their true depth values.
2. Label as *Trial* all four-neighboring grid points of *Known*. Assign the depth of *Trial* points as infinity.
3. Label the rest of the grid points as *Far*. Assign the depth of *Far* points as infinity.
4. (a) Update the values of $Z_I(u,v)$ at all *Trial* points by solving Eq. (4), as follows:
   - Let $a = \arg\min_{k\in[1,3]} Z_k$ and $b = \arg\min_{k\in[2,4]} Z_k$
   - ***Solution of the main case***
     If $Z_a \neq \infty$ and $Z_b \neq \infty$, calculate $Z_I(u,v)$. As shown in Section 2.4, Eq. (4) can be rewritten as a quadric equation:
     $z_0^2 C_1 + z_0 C_2 + C_3 = 0$
     where
     $$C_1 = I^2(A_1^2 + A_2^2 + A_3^2) - A_3^2$$
     $$C_2 = 2I^2(A_1B_1 + A_2B_2 + A_3B_3) - 2A_3B_3$$
     $$C_3 = I^2(B_1^2 + B_2^2 + B_3^2) - B_3^2$$
     $$C_4 = C_2^2 - 4C_1C_3$$
     $$\text{Suppose } Z_- = \min\frac{-C_2 \pm \sqrt{C_4}}{2C_1} \quad \text{and} \quad Z_+ = \max\frac{-C_2 \pm \sqrt{C_4}}{2C_1} \tag{5}$$
     If $C_4 < 0$ or $Z_+ < Z_m$
     $$\text{where } Z_m = \begin{cases} \min(Z_a, Z_b) & \text{The heuristic is turned on} \\ \max(Z_a, Z_b) & \text{Otherwise} \end{cases}$$
     then

     $Z_i$ is reassigned as $\infty$ where $i = \arg\max_{j\in[a,b]} Z_j$ and performs the degenerate case.
     else
     $$z_0 = \begin{cases} Z_- & \text{if } Z_- \geqslant \max(Z_a, Z_b) \\ Z_+ & \text{if } Z_+ \geqslant \max(Z_a, Z_b) > Z_- \\ Z_+ & \text{if } Z_- < Z_m \\ Z_- & \text{otherwise} \end{cases}$$
   - ***Solution of the degenerate case***
     If $Z_a = \infty$, we redefine $(\vec{P}_a - \vec{P}_0)$ as $(v_b z_b - v_0 z_0, -(u_b z_b - u_0 z_0), 0)$ such that the wavefront is restricted to propagate from the direction of $\vec{P}_b$ only i.e. $(\vec{P}_a - \vec{P}_0)\cdot(\vec{P}_b - \vec{P}_0) = 0$. The normal $\vec{N}$ is rewritten as
     $$\begin{aligned} &(v_b z_b - v_0 z_0, -(u_b z_b - u_0 z_0), 0) \\ &\quad\times\left(\frac{u_b z_b - u_0 z_0}{f}, \frac{v_b z_b - v_0 z_0}{f}, z_b - z_0\right) \\ &\stackrel{\text{def}}{=} (y_b, -x_b, 0)\times\left(\frac{x_b}{f}, \frac{y_b}{f}, z_b - z_0\right) \\ &= \begin{bmatrix} -x_b(z_b - z_0) \\ -y_b(z_b - z_0) \\ \dfrac{x_b^2 + y_b^2}{f} \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{6}$$
     The shape from shading equation becomes
     $$I = \frac{x_b^2 + y_b^2}{\sqrt{f^2 x_b^2 (z_b - z_0)^2 + f^2 y_b^2 (z_b - z_0)^2 + (x_b^2 + y_b^2)^2}}$$
     $$I^2 = \frac{x_b^2 + y_b^2}{f^2(z_b - z_0)^2 + (x_b^2 + y_b^2)}$$
     Since $(u_b, v_b)$ is the neighbor of $(u_0, v_0)$, $x_b$ and $y_b$ can be expressed as:
     $$x_b = u_0(z_b - z_0) + \Delta_1 z_b$$
     $$y_b = v_0(z_b - z_0) + \Delta_2 z_b$$
     for $(\Delta_1, \Delta_2) \in \{(0, \pm 1), (\pm 1, 0)\}$. The corresponding shading equation is rewritten as:
     $$D_1(z_0 - z_b)^2 + D_2(z_0 - z_b) + D_3 = 0$$
     where
     $$D_1 = (f^2 + u_0^2 + v_0^2)I^2 - (u_0^2 + v_0^2)$$
     $$D_2 = 2z_b(1 - I^2)(\Delta_1 u_0 + \Delta_2 v_0)$$
     $$D_3 = 2z_b^2(I^2 - 1) \leqslant 0$$
     $$D_4 = D_2^2 - 4D_1D_3$$
     Thus, the solution of $z_0$ can be obtained as one of the following four cases:
     *Case* 1: $D_1 \geqslant 0$
     The non-negative value of $D_1$ leads to a non-negative determinant $D_4$ and $z_0$ is computed as
     $$z_0 = z_b + \frac{-D_2 + \sqrt{D_2^2 - 4D_1D_3}}{2D_1}$$
     *Case* 2: $D_1 < 0$ and $D_2, D_4 \geqslant 0$
     The value $z_0$ is computed as $z_b + \frac{-D_2 - \sqrt{D_2^2 - 4D_1D_3}}{2D_1}$
     *Case* 3: $D_1 < 0$ and $D_2 < 0$ and $D_4 \geqslant 0$
     In this case, $\sqrt{D_4} < |D_2|$. Thus $z_0 - z_b < 0$. This infers that there is no propagation from $(u_b, v_b)$ to $(u_0, v_0)$.
     Thus, $z_0$ should be assigned as infinity in this propagation.
     *Case* 4: $D_4 < 0$
     This refers to an invalid propagation from $(u_b, v_b)$ to $(u_0, v_0)$. Thus $z_0$ is assigned as infinity.
   - If $Z_b = \infty$, the normal $N$ at Eq. (6) is re-formulated in which the subscript $a$ is replaced as $b$. Similar to the

pervious case, $z_0$ is obtained by solving a quadric equation.
4. (b) Find the *Trial* point $A$ with the smallest $Z_0$ values.
5. Remove $A$ from *Trial* and add it to *Known*.
6. Label as *Trial* all neighbors of $A$ that are not already *Known*.
7. If all points are labeled as *Known*, then exit, else goto step 4.

***Output***: Reconstructed depth map $Z_\mathrm{I} = Z_\mathrm{I}(u, v)$.

The basic structure of the algorithm follows the structure of the original fast marching algorithm (Kimmel and Sethian, 2001) except the choice of $Z_m$ in step (4a), which will be explained in the next section.

## 3. Handling intensity discontinuities (occlusion)

As discussed above, the fast marching method will give incorrect results when there are (1) inherent depth discontinuities in the $Z_\mathrm{S}(X, Y)$ function or (2) intensity discontinuities at occlusion boundaries generated during perspective projection. The former is usually harmless as at a depth discontinuity, $I_\mathrm{S}(X, Y)$ is zero, which gives an infinite change in depth – this can also be observed in the experimental results later on. Such a change will almost certainly be rejected by the fast marching. The following heuristic is introduced to deal with the latter situation.

If there is no discontinuity, both entropy conditions

1. $Z_0 \geqslant Z_a$ (Horizontal)
2. $Z_0 \geqslant Z_b$ (Vertical)

should be true. We relax one of the entropy conditions, i.e. we only enforce

$$(Z_0 \geqslant Z_a) \text{ OR } (Z_0 \geqslant Z_b) \quad \text{(Relaxed entropy condition)}$$

for a point to be accepted in step (4a) of Algorithm 1. In other words, it accepts those points that will satisfy at least one of the entropy conditions. The reason is explained in Fig. 3.

In Fig. 3a, $Z_3 > Z_2 > Z_1$. So the natural order of propagation is from $(1,2) \to 3$. In (b), there is a discontinuity and $Z'_2$ is incorrect. However, since the discontinuity is locally a line, $Z_4$ may still be correct despite $Z_4 > Z_3$. As both $(1,2)$ and $(1,4)$ could furnish the correct depth to calculate $Z_3$, relaxing one entropy condition helps to fill in the depth at discontinuity. However, if both entropy conditions are not satisfied, the depth estimate is incorrect and the depth should not be accepted.

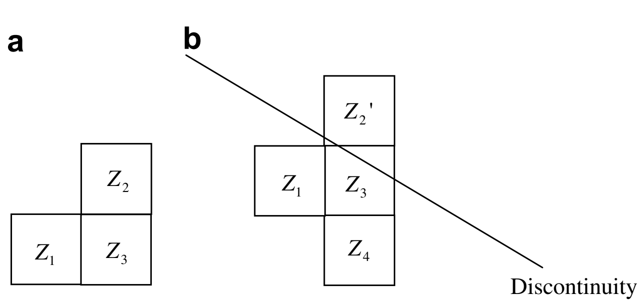


Fig. 3. (a) Normal propagation. (b) Propagation with discontinuity. One of the entropy conditions may not be satisfied.

Another rationale for using the condition is pragmatic and is concerned with numerical errors. Eq. (5) involves a fair amount of algebraic computations, and numerical errors may be significant when the actual $|Z_0 - Z_a|$ and/or $|Z_0 - Z_b|$ are small. For real images, in general, the image is not purely Lambertian. In view of this, it is safer to relax the entropy condition.

From the experimental results, it is found that without this heuristic, the fast marching method will give wrong predictions at occlusion boundaries and high intensity gradient regions (near occlusion) after the perspective projection; and it is empirically verified that the above heuristic is effective in solving this problem for both synthetic and real images. However, though the heuristic is observed to give improvement, the improvement is locally essential at those regions above. Overall, it does little harm if the heuristic is turned-off. Other suggested techniques to handle discontinuities under orthographic projection may be found in (Prados et al., 2002; Ostrov, 2000; Kain and Ostrov, 2001).

## 4. Experiments

### 4.1. Synthetic images reconstruction

#### 4.1.1. Testing images

The proposed model is tested on five synthetic images: the plane, the sphere, the vase, the four mountains and Mozart. The plane is used to validate the proposed formulae. The sphere, vase and Mozart are bench mark test images in the literature (Zhang et al., 1999). The sphere and vase are Morse surfaces. The four mountains, used by Kimmel and Bruckstein (1995) to test their singular point reasoning method, present a complicated Morse surface with convex, concave extrema and saddle points. Mozart present a even more complicated, non-Morse surface since Monkey saddles can be observed in the collar region. (The singular point reasoning method (Kimmel and Bruckstein, 1995) will fail in Mozart.) For simplicity, the self and cast shadows and inter-reflections of these surfaces are not modeled. In reality, the retinal light intensity is a function of the incident angle of the light (Forsyth and Ponce, 2003), so a mapping of light intensity is needed. This effect has not been modeled, though it is simple to account for it if necessary. The equations for the first four objects are Eqs. (7)–(10). The depth map of Mozart was obtained by a laser range finder.

The plane is generated using the formula:

$$Z_\mathrm{S}(X, Y) = B + aX + bY \tag{7}$$

where $0 \leqslant X,\ Y \leqslant 127$ and $r$ is the radius.

The sphere is generated using the formula:

$$Z_S(X,Y) = \sqrt{r^2 - (Y-64)^2 - (X-64)^2} \tag{8}$$

where $0 \leqslant X, Y \leqslant 127$ and $r$ is the radius.

The vase is generated using the formula provided by Ascher and Carter (1993):

$$Z_S(X,Y) = \sqrt{G(Y)^2 - X^2} \tag{9}$$

where $G(Y) = 0.15 - 0.1 * Y * (6Y+1)^2 * (Y-1)^2 * (3Y-2)$, $-0.5 \leqslant X \leqslant 0.5$, and $0.0 \leqslant Y \leqslant 1.0$.

The four mountains are generated using the formula provided by Kimmel and Bruckstein (1995)

$$\begin{aligned} Z(X,Y) = {} & 1.4\exp(-(2(X+0.4))^2 - (2(Y+0.5))^2) \\ & - \exp(-(3(X+0.2))^2 - (2(Y+0.2))^2) \\ & + 1.4\exp(-(3(X-0.6))^2 - (2(Y+0.7))^2) \\ & + 2\exp(-(2(X+0.4))^2 - (2(Y-0.4))^2) \\ & - 1.4\exp(-(5(X+0.52))^2 - (6(Y-0.5))^2) \\ & + 1.7\exp(-(3(X-0.5))^2 - (2(Y-0.6))^2) \end{aligned} \tag{10}$$

where $-1 \leqslant X,\ Y \leqslant 1$.

### 4.1.2. A method for generating perspective images

The simplest way to generate a synthetic image is by first order discrete approximations of the surface gradients ($\frac{\partial Z_S}{\partial X}$ and $\frac{\partial Z_S}{\partial Y}$), followed by applying the shape from shading equation. This method is used in some of the testing images in the literature for methods that work on orthographic projection. Most of the suggested approaches to solve the problem of shape reconstruction assume the orthographic projection instead of the perspective projection. To tackle the shape recovery under perspective projection, two accurate image generation methods for perspective images are desirable.

For synthetic images generated by analytical equations (namely, the plane, the sphere, the vase and the four mountains), in principle, a ray casting algorithm (Hearn and Baker, 2004) can be used with the intersection points calculated analytically. This solves the depth equation directly. (To generate a more realistic lighting model that includes higher order reflection, the radiosity approach can be used.) For testing purposes, an algorithm for generating the noise-free perspective depth map and image of a planar object is shown below:

**Algorithm 2**
***Input***: A plane equation $H:\ ax + by + z + d = 0$
For each $(u,v)$

Step (1) Let $Z$ be the depth at $(u,v)$
Step (2) Obtain the 3D point $(X,Y,Z)$ from $X = \frac{uZ}{f}$ and $Y = \frac{vZ}{f}$
Step (3) Based on the given plane equation, $X$ and $Y$ are expressed as

$$X = -\frac{u(d+aX+bY)}{f} \quad \text{and} \quad Y = -\frac{v(d+aX+bY)}{f} \tag{11}$$

Step (4) Eq. (11) can be rewritten as

$$\begin{bmatrix} f+au & bu \\ av & f+bv \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} -ud \\ -vd \end{bmatrix}$$

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} f+au & bu \\ av & f+bv \end{bmatrix}^{-1} \begin{bmatrix} -ud \\ -vd \end{bmatrix}$$

Step (5) Obtain

$$Z = -d - \begin{bmatrix} a \\ b \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} f+au & bu \\ av & f+bv \end{bmatrix}^{-1} \begin{bmatrix} -ud \\ -vd \end{bmatrix}$$

***Output***: The depth map $Z_I(u,v)$.

On the other hand, if the equations are complicated, the root finding will become non-trivial or even impossible. An additional problem is that numerical solution may give a root that is not the visible surface point. Consequently, a simpler algorithm that is accurate to a grid resolution $\delta_2$ is reported below:

Given an analytical equation, the $Z$ buffer method is used for visible surface determination. Let $Z_I(u,v)$ be the depth buffer value at grid point $(u,v)$. The backprojected point $(X_{bp}, Y_{bp}) = \left(u\frac{Z_I}{f}, v\frac{Z_I}{f}\right)$ is used as the center of a fine grid search (Rao, 1984) to find the exact intersection point in an $2\delta_1 \times 2\delta_1$ area with a fixed step size $\delta_2$. For each point $(X,Y) = (a,b)$ in the fine grid, $Z_S(a,b)$ is calculated using the analytical equation. The perpendicular distance of the fine grid point $(X,Y,Z)$ and the ray passing through the center of projection and the center of the image pixel is determined, and the intersection point is the one with the minimum perpendicular distance. At this point, the analytical equation is applied to determine the normal, and the corresponding intensity is generated.

The algorithm is a bit more complicated than the above. Since a perspective projection is involved, the ray may cut through more than one point of the surface. Thus sometimes the fine grid search may return an occluded point, which leads to the wrong intensity being generated. To preclude this, the following measure is taken: If the absolute difference between the depth of the fine grid point $Z_S$ and the $Z$ buffer value is larger than a limit $\varepsilon$, the point is excluded from further consideration.

The detailed algorithm is shown below:

**Algorithm 3**
***Input***: An analytical $Z_S = Z_S(X,Y)$ depth function
For each $(X,Y)$

Step (1) Calculate $Z_S(X,Y)$.
Step (2) Obtain image points $(u,v)$ from $u = \left\lfloor \frac{X}{Z_S} f \right\rfloor$ and $v = \left\lfloor \frac{Y}{Z_S} f \right\rfloor$, where $\lfloor\,\rfloor$ denotes the floor function
Step (3) ($Z$-buffer method) Record the $Z_I$ value of the visible point at $(u,v) = \left(\left\lfloor \frac{X}{Z_S} f \right\rfloor, \left\lfloor \frac{Y}{Z_S} f \right\rfloor\right)$, i.e., $Z_I(u,v) = \min(Z_S(X,Y))$
For each $(u,v)$

Step (4) (Back projection) Obtain $X_{bp} = \frac{u}{f} Z_{\mathrm{I}}(u, v)$ and $Y_{bp} = \frac{v}{f} Z_{\mathrm{I}}(u, v)$

Step (5) (Fine grid point search)
For $a = X_{bp} - \delta_1$ to $X_{bp} + \delta_1$, step $\delta_2$
  For $b = Y_{bp} - \delta_1$ to $Y_{bp} + \delta_1$, step $\delta_2$
  {
   compute $Z_{\mathrm{S}}(a, b)$
   Let $L$ be the line passing through $(0, 0, 1)$ and $(u, v, f)$, compute the perpendicular distance $D(a, b)$ between $L$ and $(a, b, Z_{\mathrm{S}}(a, b))$, i.e.,

$$D(a,b) = \frac{|(u,v,f) \times (a,b,Z_{\mathrm{S}}(a,b))|}{|(u,v,f)|}$$

  }
Select the minimum distance $D_{\min}(a, b)$ satisfying $|Z_{\mathrm{I}}(u, v) - Z_{\mathrm{S}}(a, b)| < \varepsilon$ and record its $(a, b)$

Step (6) Compute $p = \frac{\partial Z_{\mathrm{S}}}{\partial X}\Big|_{(a,b)}$ and $q = \frac{\partial Z_{\mathrm{S}}}{\partial Y}\Big|_{(a,b)}$ using the given analytical equation, and calculate the intensity value

$$I(a,b) = \frac{1}{\sqrt{p^2 + q^2 + 1}}$$

Record the depth and intensity values in the $Z$ buffer and the intensity map: $Z_{\mathrm{S}}(a, b) \rightarrow Z_{\mathrm{I}}(u, v)$ and $I_{\mathrm{S}}(a, b) \rightarrow I_{\mathrm{I}}(u, v)$.

***Output***: $Z$ buffer depth map $Z_{\mathrm{I}}(u, v)$ and intensity map $I_{\mathrm{I}}(u, v)$.

The fine grid search cannot be applied to the test image Mozart since the data given is a set of sampled depth values obtained by a laser range finder. Thus for Mozart, step 5 is omitted and in step 6, $p$ and $q$ are estimated using first order discrete approximation.

#### 4.1.3. Results

$128 \times 128$ images are used for the plane, the sphere, the vase, the four mountains, and a $256 \times 256$ image is used for Mozart. The following settings are used: $\delta_1 = 2.0$, $\delta_2 = 0.005$. The value of $\varepsilon$ is image dependent. The lighting direction is $(0, 0, 1)$.

Fig. 4a shows a plane with the normal $[-0.1, -0.1, 1]$ and bias $-100$. The focal length and offset for $Z_s$ are 50 and 200 respectively. The corresponding depth map and

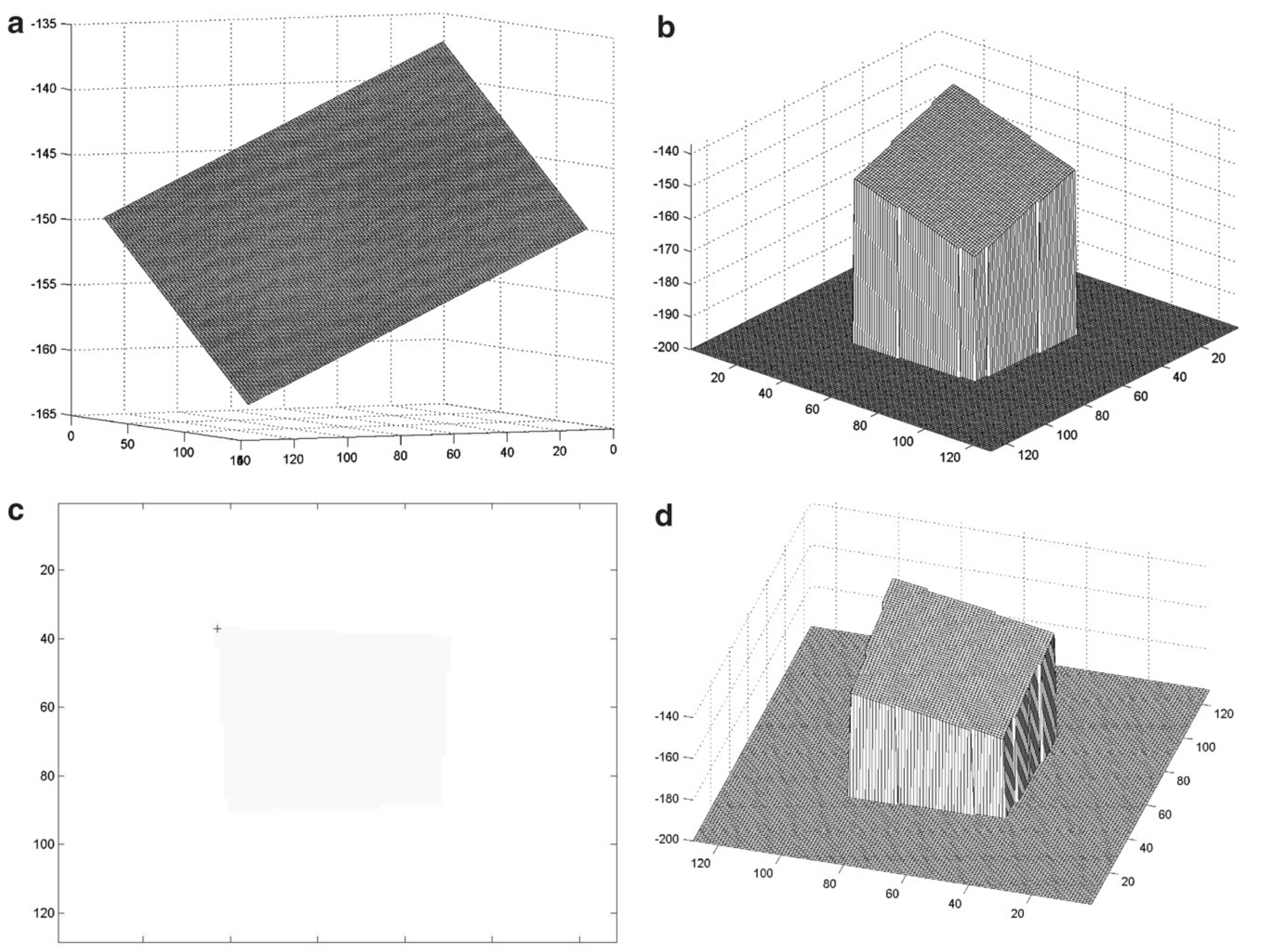


Fig. 4. Plane: (a) the original surface generated from its standard equation, (b) the transformed surface, (c) the shading image for the perspective model and (d) the result of the algorithm.

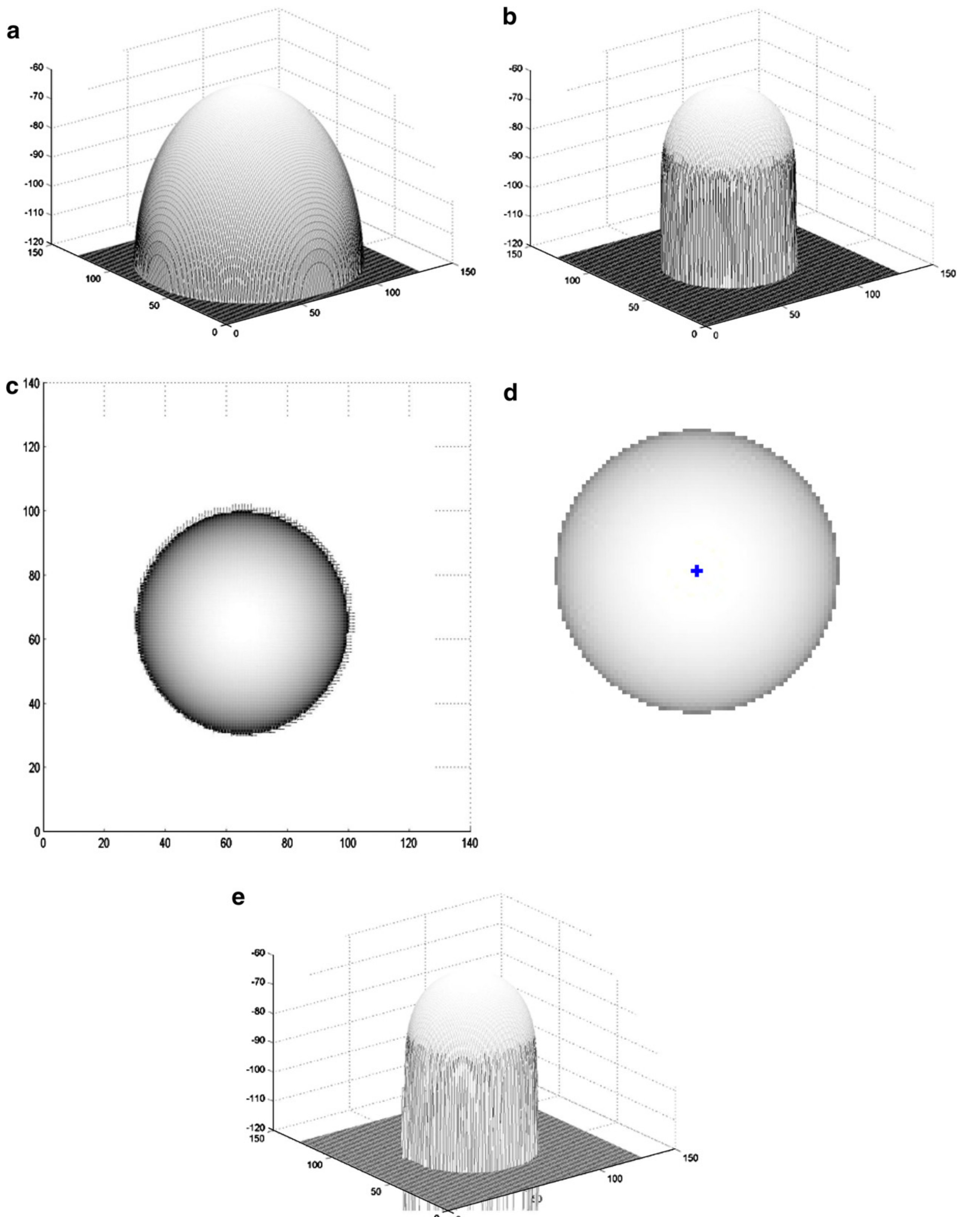


Fig. 5. Sphere: (a) the original surface generated from its standard equation, (b) the transformed surface, (c) the shading image for the perspective model, (d) singular point located at the center and (e) the result of the algorithm.

image under perspective projection are shown in Fig. 4b and c. Fig. 5a shows a sphere with radius $r$ of 60 positioned at the center of the image. A $Z_S$ offset of 120 and a focal length of 60 give the transformed depth map (Fig. 5b)

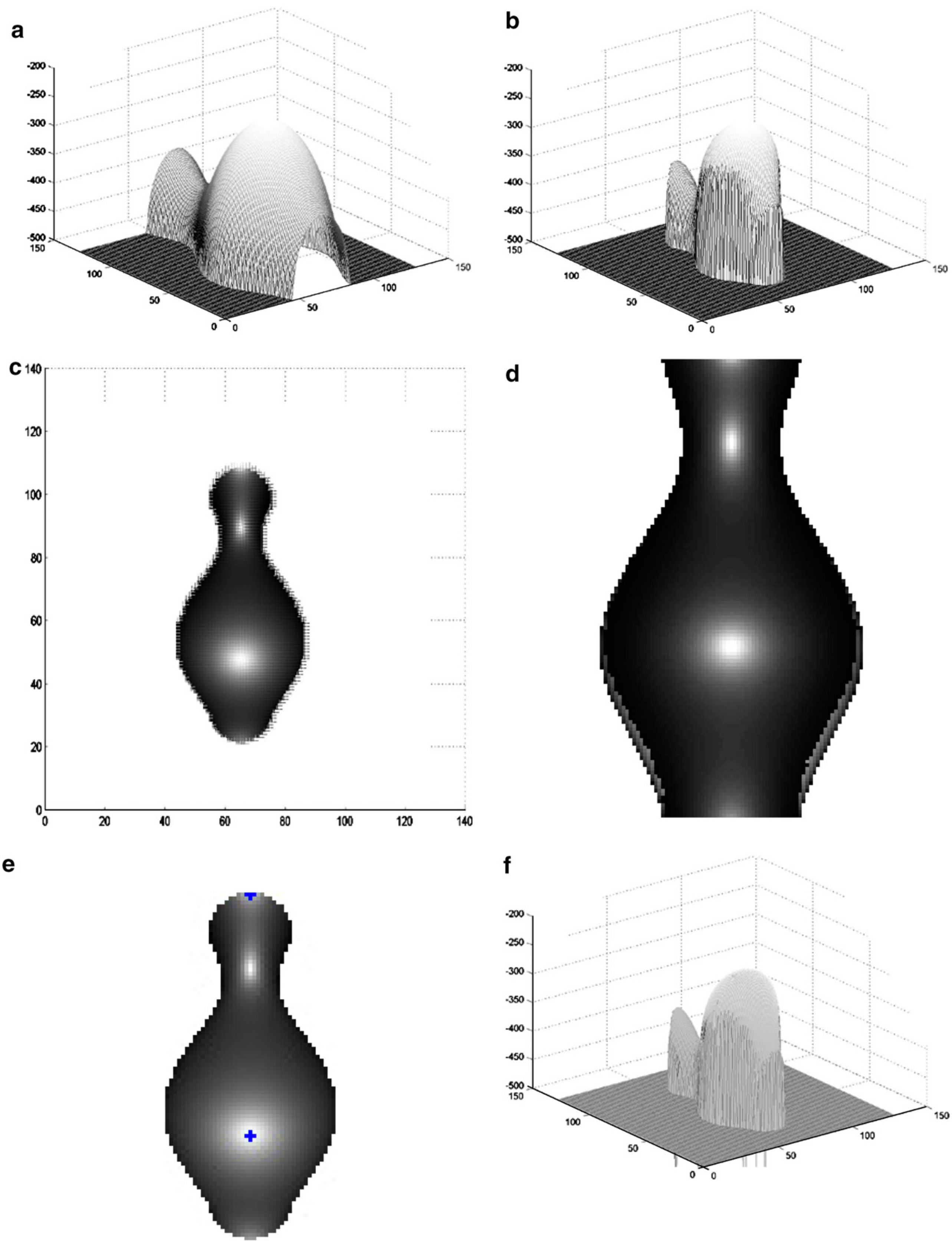


Fig. 6. Vase: (a) the original surface generated from its standard equation, (b) the transformed surface, (c) the shading image for the perspective model, (d) the occluded region (depth tolerance = 20), (e) two singular points and (f) the result of the algorithm.

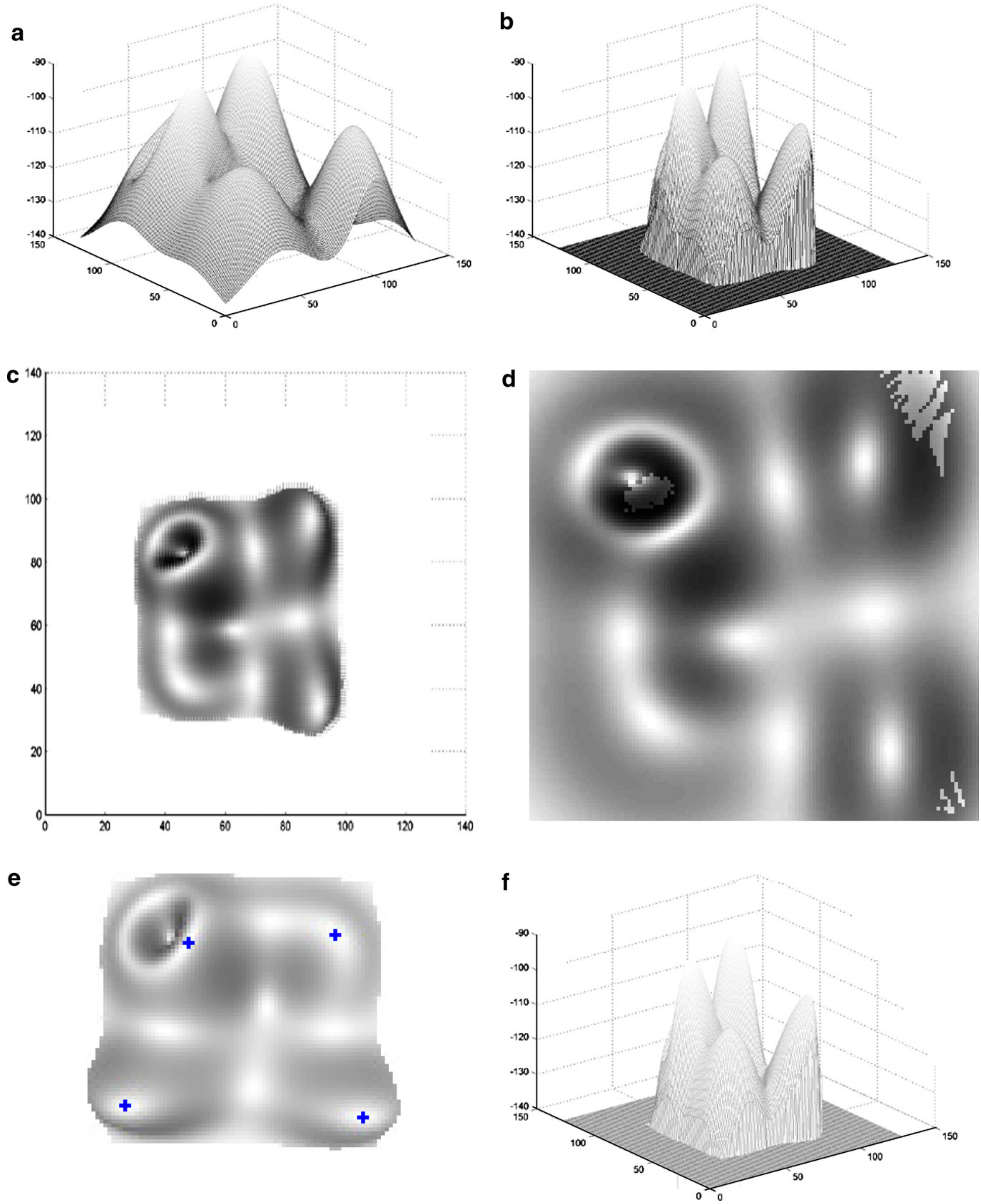


Fig. 7. Four mountains: (a) the original surface generated from its standard equation, (b) the transformed surface, (c) the shading image for the perspective model, (d) the occluded region (depth tolerance = 10), (e) four singular points, (f) the result of the algorithm, (g) the error surface at the region of occlusion boundary determined by the proposed method and (h) the error surface at the region of occlusion boundary determined by the algorithm of Tankus et al.

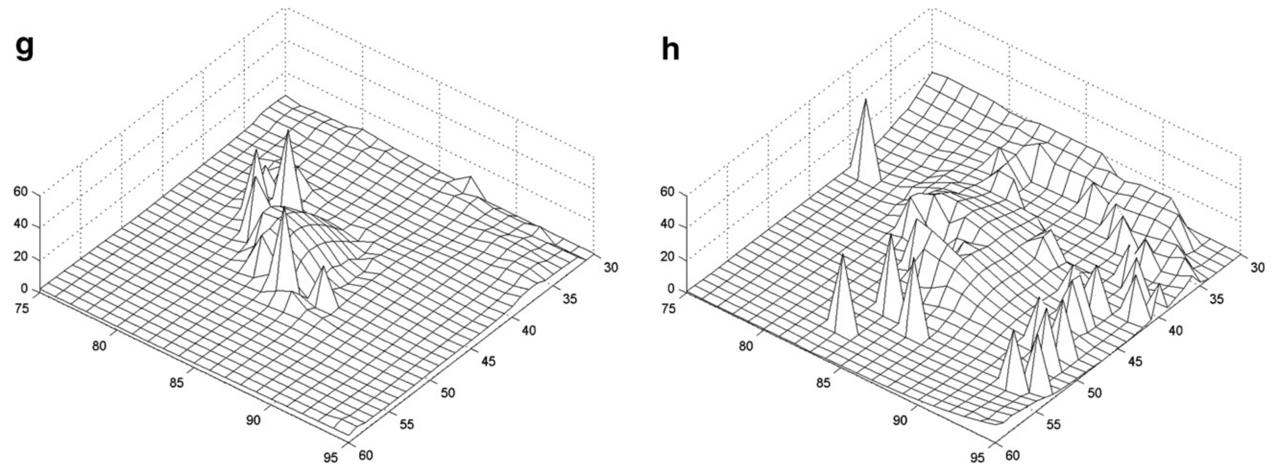


Fig. 7 (*continued*)

and the perspective image (Fig. 5c). Fig. 6a shows the vase. The $X-Y$ range is mapped to $[0,127]\times[0,127]$ and $Z_S$ is scaled by 894. Applying an $Z_S$ offset of 500 and a focal length of 250 give the transformed depth map (Fig. 6b) and the perspective image (Fig. 6c). Fig. 7a shows the four mountains with $X-Y$ mapped to $[0,127]\times[0,127]$ and $Z_S$ scaled by 28. The offset for $Z_S$ is 140 and the focal length is 70 (Fig. 7b and c). The original depth map of Mozart is shown in Fig. 8a. The offset is 200 and the focal length is 65. The intensity image generated is shown in Fig. 8b. The transformed depth map, the side view and the perspective image are shown in Fig. 8c, d and e respectively. In Fig. 8d, the face of Mozart is deliberately stretched to give a better visualization of the salient features.

Due to the influence of perspective projection, some points on the original models are occluded after the transformation. A point $(X,Y)$ is said as occluded if there exists another pixel $(X_0,Y_0)$ such that:

$$\left|\frac{X}{Z_S(X,Y)}f-\frac{X_0}{Z_S(X_0,Y_0)}f\right|<1$$
$$\left|\frac{Y}{Z_S(X,Y)}f-\frac{Y_0}{Z_S(X_0,Y_0)}f\right|<1$$

and

$$Z_S(X,Y)-Z_S(X_0,Y_0)>\varepsilon_D$$

where $\varepsilon_D$ is the depth tolerance. The highlighted region of Fig. 6d shows the occluded points of the vase with $\varepsilon_D=20$. Similarly, Figs. 7d and 8f show the occluded point sets of the four mountains and the Mozart with $\varepsilon_D=10$ and 30 respectively. The values of $\varepsilon_D$ are arbitrarily chosen for better visualization and they do not affect the reconstruction results.

To reconstruct the surface, the selected local minimum (or concave) singular points are assigned the true depth values. The singular points whose depths are given are shown as crosses in Figs. 4c, 5d, 6e, 7e and 8g. One singular point is chosen for the plane image (Fig. 4c) and the sphere image (Fig. 5d). The vase image has two singular points (Fig. 6e). The four mountains have four singular points (Fig. 7e). In the Mozart image (Fig. 8g), five out of thirty eight singular points are chosen.

The reconstructed depth maps of the plane, the sphere, the vase, the four mountains and Mozart are shown in Figs. 4d, 5e, 6f, 7f and 8h respectively. Fig. 8i shows the side view of the reconstruction of Mozart (compare with Fig. 8d). In the original depth map, there is a normal discontinuity at the silhouette and hence intensity is equal to zero; the Eikonal equation degenerates and no meaningful solution should be expected. This is compounded by the perspective projection, which further introduces a depth discontinuity as well as an intensity discontinuity. Therefore, the depth near the silhouette is set to the background in these figures. Qualitatively, it can be seen that excellent results are obtained for all the figures.

The performance of the proposed method is further studied by a quantitative comparison with the perspective Fast Marching method of Tankus et al. (2005). Instead of representing the performance as the error mean, error standard deviation and the complicated moving least square measurement suggested in (Tankus et al., 2005), root mean square depth error (RMSE) is used, which gives a more straight forward and fair comparison.

The reconstructions of the four models are repeated by the algorithm of Tankus et al. using the same desired depth maps and the images as the pervious experiment. We initialize the processes by the orthographic Fast Marching method of Kimmel and Sethian. The resulting depth maps are obtained after five iterations since they report that their algorithm usually converges after five iterations. In each model reconstruction, the singular points are defined as the points of local minimal depth as suggested in (Tankus et al., 2005). Totally 1, 1, 2, 4 and 38 singular points are given for the plane, the sphere, the vase, the four mountains and the Mozart respectively. In addition, since the proposed method uses only five singular points to reconstruct the Mozart, which is a more stringent setting, we repeat the Mozart reconstruction by the algorithm of Tankus using the same set of singular points. Both algorithms are implemented using the C language. The code is avail-

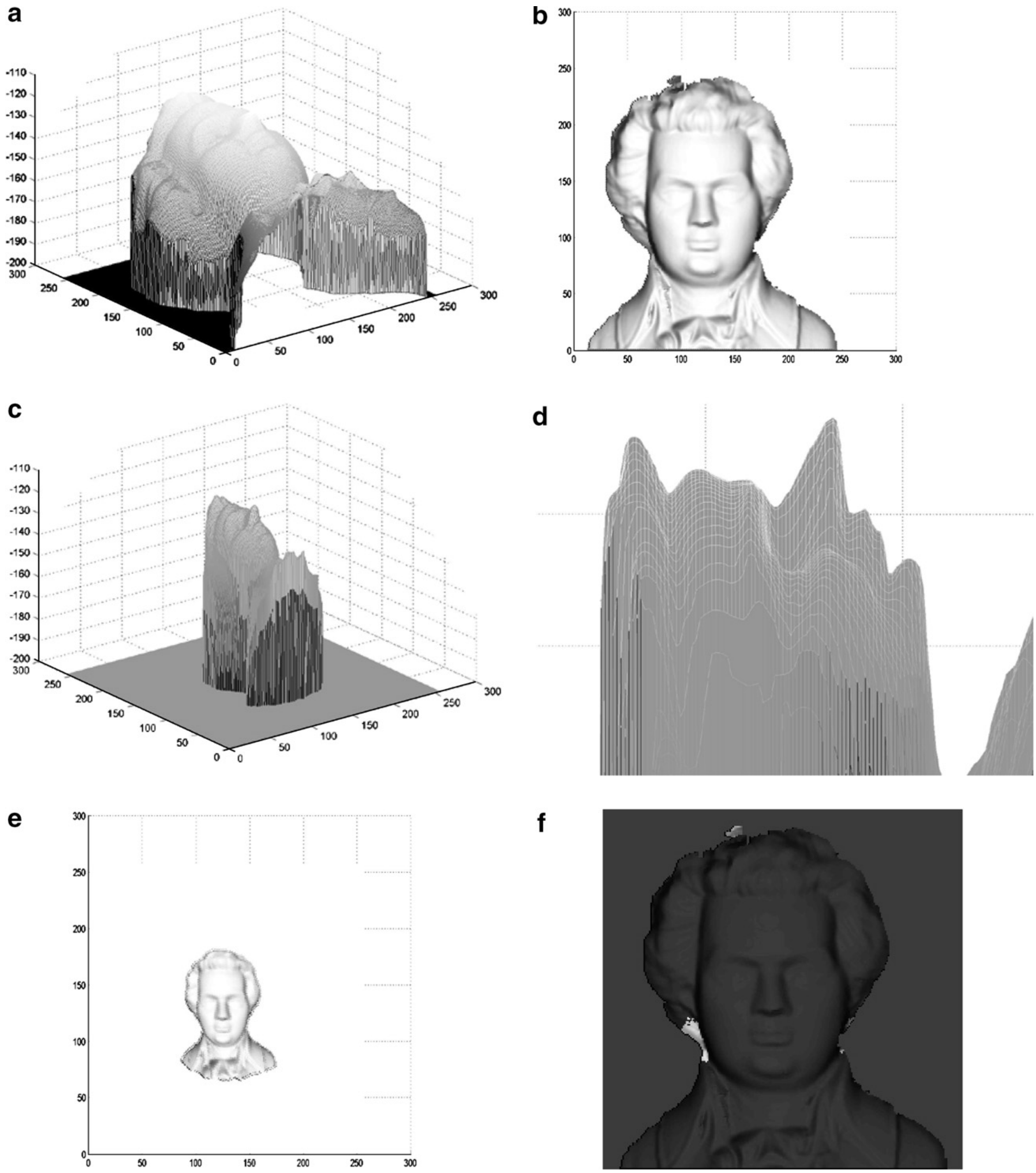


Fig. 8. Mozart: (a) the original depth map, (b) the intensity image generated by the true depth map, (c) the transformed surface, (d) the transformed surface in the side view, (e) the shading image for the perspective model, (f) the occluded region (depth tolerance = 30), (g) five singular points, (h) the result of the algorithm and (i) the result of the algorithm in the side view.

able in http://www.ee.cityu.edu.hk/~syyuen/Public/SfS/SfSSim.html. All reconstruction processes are performed on a PC with Pentium IV 1.7 GHz CPU and 512MB memory.

Tables 1–5 list the comparisons of three algorithms: orthographic Fast Marching, the perspective Fast Marching and the proposed method. The computational time of our proposed single-pass method is shorter than that of the iterative algorithm of Tankus et al. On the other hand, since the discontinuities occur in the region near the silhouette, huge errors are introduced in the algorithms, such that the resultant RMSEs are non-conclusive.

In order to illustrate the performance of the algorithms in a conclusive manner (i.e. ignore the meaningless solution at the silhouette), the error measurement considers only the valid pixels. $P(i,j)$ is valid if none of the pixel within the

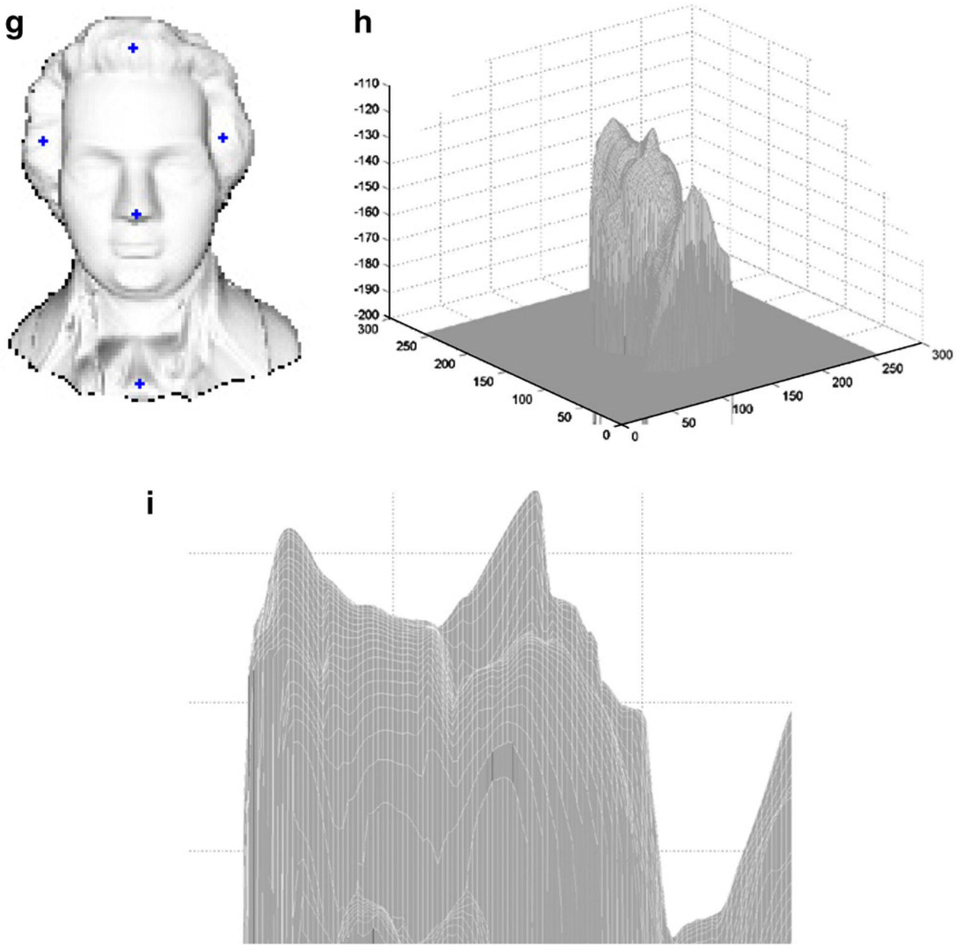


Fig. 8 (*continued*)

Table 1
Plane: comparison of algorithms

| | Kimmel and Sethian | Tankus et al. | The proposed method |
|---|---|---|---|
| RMSE | 311 | 324 | 297 |
| Computation time (s) | 0.89 | 3.98 | 0.97 |

Table 2
Sphere: comparison of algorithms

| | Kimmel and Sethian | Tankus et al. | The proposed method |
|---|---|---|---|
| RMSE | 374 | 316 | 341 |
| Computation time (s) | 0.874 | 4.355 | 0.952 |

Table 3
Vase: comparison of algorithms

| | Kimmel and Sethian | Tankus et al. | The proposed method |
|---|---|---|---|
| RMSE | 2623 | 2641 | 2598 |
| Computation time (s) | 1.164 | 5.913 | 1.432 |

Table 4
Four mountains: comparison of algorithms

| | Kimmel and Sethian | Tankus et al. | The proposed method |
|---|---|---|---|
| RMSE | 1185 | 1180 | 1097 |
| Computation time (s) | 1.243 | 6.038 | 1.507 |

template centered at $P$ is a silhouette pixel. Fig. 9 shows an example of pixel validation with template width $w = 3$. Given that white-filled boxes represent the pixels with value depth, the solid-line boxes are regarded as the valid pixels for measuring the RMSE while the dotted-line boxes are the invalid pixels that are ignored in the calculation of RMSE. The comparisons are repeated in which only the RMSEs of valid pixels are studied. Seen from Table 6, the nearly error-free result verifies the correctness of the proposed formulae. Clearly, the proposed method performs better than Kimmel and Sethian, as well as Tankus et al. in terms of RMSE for all tested images, especially in the four mountains and the Mozart. The main reason is that Tankus' algorithm is an iterative method with an incorrect initialization and no proof of correct conver-

Table 5
Mozart: comparison of algorithms

| | Kimmel and Sethian | Tankus et al. | The proposed method |
|---|---|---|---|
| RMSE | 2598 (38 singular points)<br>2614 (5 singular points) | 2588 (38 singular points)<br>2603 (5 singular points) | 2523 |
| Computation time (s) | 2.327 (38 singular points)<br>2.298 (5 singular points) | 12.056 (38 singular points)<br>11.673 (5 singular points) | 4.873 |

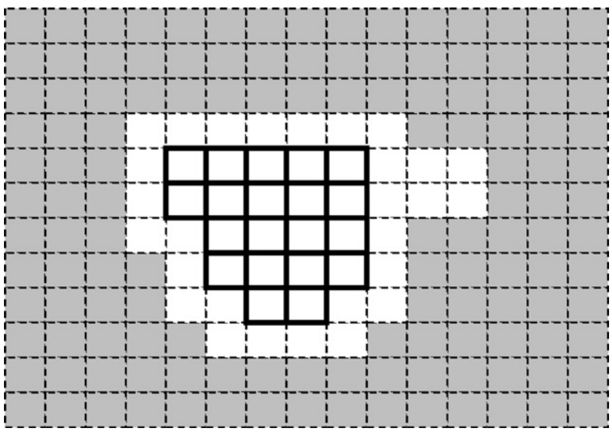

Fig. 9. Example of pixel validation.

gence. Another reason, but minor, is that the algorithm of Tankus et al. cannot handle the occlusion problem while the proposed method can. The effectiveness of the occlusion heuristic can be seen by studying the volcano region of the four mountains, where there is an area with occlusion (upper left of Fig. 7d). Fig. 7g and h show the respective error surfaces of our method and Tankus's method at the volcano. It can be seen clearly that our method handles the intensity and depth discontinuities well, resulting in a surface with smaller errors. Furthermore, the comparisons on the Mozart show that the proposed method has an advantage over the algorithm of Tankus et al. as it is more robust to missing singular points. By comparing the results obtained from the proposed algorithm with and without the heuristic for handling occlusion (the last two columns of Table 6), it indicates that the heuristic does not improve much the overall error, but further enhances the performance of the proposed algorithm in local regions where occlusion occurs.

Though that the results shown in (Tankus et al., 2005) converge to correct solutions, its correct convergence is doubtful as large error is obtained in the reconstruction of the planar object.

Finally, the proposed method is 2–3 times faster than Tankus's algorithm in all tested images, but is somewhat slower than Kimmel's method. These results are reasonable as both Kimmel and our methods are single pass, whilst Tankus' are iterative. Our method takes somewhat longer than Kimmel's method because of the more complicated equation solving.

### 4.2. Real image reconstruction

A real $128 \times 128$ image of a white mask (Fig. 10a) is used. Since ground truth is unavailable, we guess the depths to be 150 at two singular points, as shown in Fig. 10b. The focal length and the offset are assumed to be 150 and 300 respectively. These parameters will affect the scale of the object, so one should only look at the shape of the reconstruction.

The recovered depth map of the mask is shown in Fig. 10c. As shown in the close up view (Fig. 10d), the recovered shape is faithful to the actual shape. To verify the results, the intensity image reconstructed from the result is shown in Fig. 10e. Not with standing some slight creases, this image depicts faithfully the intensity undulations. It is somewhat darker than the original image. This may be due to the deviation from the Lambertian assumption of the real image.

## 5. Conclusions

A novel single-pass method for shape from shading under perspective projection is reported. It is an extension of the method of Kimmel and Sethian (2001) to the

Table 6
RMSE: comparisons of algorithms with pixel validation

| Model | Kimmel and Sethian | Tankus et al. | The proposed method | |
|---|---|---|---|---|
| | | | Without heuristic | With heuristic |
| Plane | 1.348 | 0.827 | 0.06 | 0.05 |
| Sphere | 3.774 | 3.684 | 0.7138 | 0.7138 |
| Vase | 22.228 | 22.687 | 4.0984 | 4.0514 |
| Four mountains | 10.039 | 9.823 | 2.2560 | 1.9970 |
| Mozart | 7.063 (38 singular points)<br>8.786 (5 singular points) | 9.271 (38 singular points)<br>11.384 (5 singular points) | 6.7094 | 6.4010 |

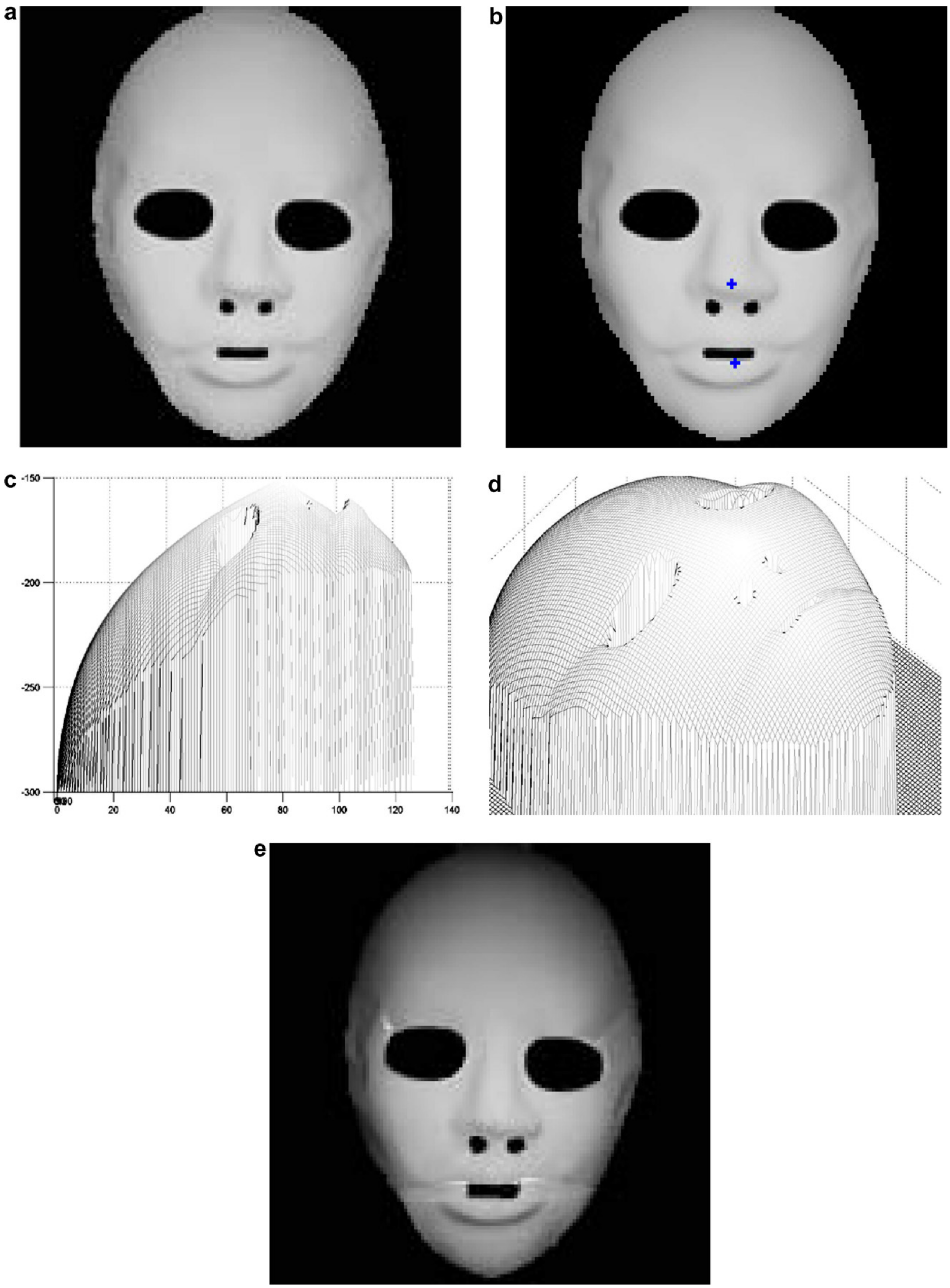


Fig. 10. Mask: (a) real intensity image, (b) two singular points, (c) reconstructed depth map in the side view, (d) reconstructed depth map in close up and (e) intensity image from reconstructed depth map.

perspective case through a simple change of variables. The method is based on the fast marching method. Compared with other existing algorithms, the method enjoys the following intrinsic advantages: (1) It is a deterministic method with $O(N \log N)$ complexity, where $N$ is the total number of pixels. (2) It is guaranteed to obtain the correct solution, provided that the heights of local minimum singular points are given. These unknown heights are the free boundary conditions of the Eikonal PDE and are explicitly isolated out. A proof that the fast marching method with change of variables is applicable to perspective projection is given. We also introduce a novel technique for handling intensity discontinuities due to occlusion. Implicit in the formulation of fast marching is the assumption of continuous function. This assumption will become false when occlusion occurs, at which time the perspective projection is not bijective. A heuristic method is proposed to handle this problem by enforcing either the horizontal or the vertical entropy condition.

Experiments are conducted using synthetic data, laser scanner data and real data. It is found that the qualitative result is excellent. The shape features are nearly perfectly recovered for each of the five figures tested, including some fairly complicated synthetic data. Also, the quantitative error in the interior of the surface is reasonably small. However, at the silhouette, the method suffers from large errors. It should not be over emphasized that it is not the fault of the algorithm, as there is a large discontinuity in the normal, depth, and intensity at the silhouette in a perspective projection.

The performance of the proposed method is further studied by a quantitative comparison with (1) the orthographic Fast Marching method of Kimmel and Sethian (2001) and (2) perspective Fast Marching method of Tankus et al. (2005). In all the images that we tested, our method has the smallest RMSE. It is also about two to three times faster than Tankus et al. but is somewhat slower than Kimmel and Sethian method.

It is natural that our method has smaller RMSE than Kimmel's method as it uses a more correct model. One reason why our method is better than Tankus's method is that it computes the direct solution, rather than uses an iterative method to approximate the solution. A minor reason is that our method explicitly handles occlusions, whilst there is no such mechanism in Tankus's method.

Our method is definitely faster than Tankus's method for the simple reason that our method computes the solution in one step directly, whilst Tankus's method relies on iterative solution (though experimentally, we observe that convergence to a stationary point is fast) with yet no theoretical proof of correct convergence.

In addition, it is also observed that our method works more robustly than Tankus's method. For example, our results in Mozart, which uses just five singular height points, is better than Tankus's method using 38 singular height points.

Two simple methods for generating synthetic perspective images are also reported. It is useful for a careful testing of shading algorithms in future research.

Recently, we have obtained some result for oblique lighting. Based on the light source coordinate transformation, we show that the oblique and frontal problems are equivalent under perspective projection (Chow and Yuen, submitted for publication). Our experimental results indicate that the fast marching method is also valid when the lighting direction is highly oblique.

## Acknowledgements

The work described in this paper was supported by a grant from the CityU Strategic Research Grant CityU 7001424. Tsui Yuen Yan is supported by a Research Studentship and Research Scholarship by the City University of Hong Kong. We thank the editor and reviewers for their valuable comments on this work. Prof. Richard M.M. Chen pointed out a mistake in our early work. Leung Yiu Wah checked some of our programs. We are grateful to Dr. Ron Kimmel for providing the four mountains model. Mozart can be obtained by anonymous ftp from custis@cs.ucf.edu under the directory /pub/shading.

## References

Ascher, U.M., Carter, P.M., 1993. A multigrid method for shape from shading. SIAM J. Numer. Anal. 30 (1), 102–115.

Bichsel, M., Pentland, A.P., 1992. A simple algorithm for shape from shading. In: Proc. CVPR, pp. 459–465.

Brooks, M.J., Horn, B., 1985. Shape and source from shading. In: Proc., IJCAI, pp. 932–936.

Chow, Chi Kin, Yuen, Shiu Yin, submitted for publication. Equivalence of oblique and frontal illumination in perspective shape from shading. IEEE T. PAMI.

Courteille, F., Crouzil, A., Durou, J.-D., Gurdjos, P., 2004. Towards shape from shading under realistic photographic conditions. In: Proc. 17th ICPR, vol. 2, pp. 277–280.

Crouzil, A., Descombes, X., Durou, J.-D., 2003. A multiresolution approach for shape from shading coupling deterministic and stochastic optimization. IEEE Trans. Pattern Anal. Machine Intell. 25 (11), 1416–1421.

Dupuis, P., Oliensis, J., 1992. Direct method for reconstructing shape from shading. In: Proc. CVPR, pp. 453–458.

Forsyth, D.A., Ponce, J., 2003. Computer Vision, a Modern Approach. Prentice Hall.

Hasegawa, J.K., Tozzi, C.L., 1996. Shape from shading with perspective projection and camera calibration. Comput. Graphics 20 (3), 351–364.

Hearn, D., Baker, M.P., 2004. Computer Graphics with OpenGL. third ed. Prentice Hall.

Horn, B., 1975. Obtaining shape from shading information. In: Winston, P.H. (Ed.), The Psychology of Computer Vision. McGraw-Hill, New York, pp. 115–155.

Ikeda, O., 2002. A novel shape-from-shading algorithm using Jacobi iterative method and bi-directional estimation. In: Proc. 5th IASTED Internat. Conf. Computer Graphics and Imaging, pp. 56–61.

Ikeuchi, K., Horn, B., 1981. Numerical shape from shading and occluding boundaries. Artif. Intell. 17 (1–3), 141–184.

Kain, J., Ostrov, D.N., 2001. Numerical shape-from-shading for discontinuous photographic images. Internat. J. Comput. Vision 44 (3), 163–173.

Kimmel, R., Bruckstein, A.M., 1995. Global shape from shading. Comput. Vision Image Understanding 62 (3), 360–369.

Kimmel, R., Sethian, J.A., 2001. Optimal algorithm for shape from shading and path planning. J. Math. Imaging Vision 14, 237–244.

Lee, K.M., Kuo, C.C.J., 1994. Shape from shading with perspective projection. Comput. Vision Graphics Image Process.: Image Understanding 59 (2), 202–212.

Lee, K.M., Kuo, C.C.J., 1997. Shape from shading with a generalized reflectance map model. Computer Vision Image Understanding 67 (2), 143–160.

Ostrov, D.N., 2000. Extending viscosity solutions to Eikonal equations with discontinuous spatial dependence. Nonlinear Anal. 42, 709–736.

Osher, S., Paragios, N., 2003. Geometric Level Set Methods in Imaging, Vision, and Graphics. Springer.

Pentland, A.P., 1984. Local shading analysis. IEEE Trans. Pattern Anal. Machine Intell. 6, 170–187.

Pentland, A.P., 1990. Linear shape from shading. Internat. J. Comput. Vision 4 (2), 153–163.

Prados, E., Faugeras, O., 2003. Perspective shape from shading and viscosity solutions. In: Proc. 9th ICCV, vol. 2, pp. 826–831.

Prados, E., Faugeras, O., 2004. Unifying approaches and removing unrealistic assumptions in shape from shading: Mathematics can help. In: Proc. ECCV (LNCS 3024), pp. 141–154.

Prados, E., Faugeras, O., Rouy, E., 2002. Shape from shading and viscosity solutions. Lect. Notes Comput. Sci. 2351, 790–804.

Rao, S.S., 1984. Optimization: Theory and Applications. second ed. Wiley, New York, pp. 214–255.

Rouy, E., Tourin, A., 1992. A viscosity solutions approach to shape-from-shading. SIAM J. Numer. Anal. 29 (3), 867–884.

Sethian, J.A., 1996. A fast marching level set method for monotonically advancing fronts. Proc. Natl. Acad. Sci. 93, 1591–1595.

Sethian, J.A., 1999. Level Set Methods and Fast Marching Methods. Cambridge University Press.

Tankus, A., Sochen, N., Yeshurun, Y., 2003. A new perspective on shape-from-shading. In: Proc. 9th ICCV, vol. 2, pp. 862–869.

Tankus, A., Sochen, N., Yeshurun, Y., 2005. Shape-from-shading under perspective projection. Internat. J. Comput. Vision 63 (1), 21–43.

Tsui, Y.Y., 2003. MPhil Dissertation, City University of Hong Kong, PRC.

Verbeek, P.W., Verwer, B.J.H., 1990. Shape from shape, the Eikonal equation solved by gray-weighted distance transform. Pattern Recognition Lett. 11, 681–690.

Worthington, P.L., Hancock, E.R., 1999. Needle map recovery using robust regularizers. Image Vision Comput. 17, 545–557.

Yamany, S.M., Farag, A.A., Tasman, D., Farman, A.G., 2000. A 3-D reconstruction system for the human jaw using a sequence of optical images. IEEE Trans. Med. Imaging 19 (5), 538–547.

Yuen, S.Y., Tsui, Y.Y., Leung, Y.W., Chen, R.M.M., 2002. Fast marching method for shape from shading under perspective projection. In: Proc. 2nd Internat. Conf. Visualization, Imaging, and Image Processing, pp. 584–589.

Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M., 1999. Shape-from-shading: a survey. IEEE T. PAMI 21 (8), 690–706.