Int J Comput Vis
DOI 10.1007/s11263-009-0207-3

A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes

Chenyu Wu · Srinivasa G. Narasimhan · Branislav Jaramaz

Received: 10 February 2008 / Accepted: 6 January 2009

Abstract This article formulates a near-lighting shape-from-shading problem with a pinhole camera (perspective projection) and presents a solution to reconstruct the Lambertian surface of bones using a sequence of overlapped endoscopic images, with partial boundaries in each image. First we extend the shape-from-shading problem to deal with perspective projection and near point light sources that are not co-located with the camera center. Secondly we propose a multi-image framework which can align partial shapes obtained from different images in the world coordinates by tracking the endoscope. An iterative closest point (ICP) algorithm is used to improve the matching and recover complete occluding boundaries of the bone. Finally, a complete and consistent shape is obtained by simultaneously regrowing the surface normals and depths in all views. In order to fulfill our shape-from-shading algorithm, we also calibrate both geometry and photometry for an oblique-viewing endoscope that are not well addressed before in the previous literatures. We demonstrate the accuracy of our technique using simulations and experiments with artificial bones.

Keywords Multi-image · Shape-from-shading · Near-lighting · Perspective projection · Calibration · Endoscope · Bone

C. Wu (✉) · S.G. Narasimhan · B. Jaramaz
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: cywu@cs.cmu.edu

S.G. Narasimhan
e-mail: srinivas@cs.cmu.edu

B. Jaramaz
Institute of Computer Assisted Surgery, 3380 Boulevard of the Allies, Suite 270, Pittsburgh, PA 15213, USA
e-mail: branko@icaos.org

Published online: 12 February 2009

1 Introduction

One of the main goals of orthopedic surgery is to enable minimally invasive procedures. As a key tool, endoscope is attracting increasing attention for its potential role in computer aided surgeries. Besides visualizing the interior of the anatomy, its capability can be significantly enhanced by tracking the endoscope in conjunction with surgical navigation systems.

Figure 1 shows an endoscope illuminating and observing an artificial spine. The endoscope uses one or more point light sources placed at its tip for illumination. Due to the small distance between the sources, camera and the scene, the endoscopic images are very different from the images we experience of natural scenes under distant lighting such as from the sun or the sky. Since the field of view of the endoscope is small and the bone is usually a few millimeters away, we only observe a small part of the bone. As a result, it can be difficult even for a skilled surgeon to infer bone shape from a single endoscopic image. Better visualizations are achieved by overlaying endoscopic images with 3D surfaces obtained from CT (Computed Tomography) scanned images (Clarkson et al. 1999; Dey et al. 2002). However, this technique requires us to solve a complex registration problem between the CT and endoscopic images during the surgery. Thus, there is an immediate need for explicit computer reconstruction of bone shapes from endoscopic images.

Since bone surfaces have few identifiable features, surface shading is the primary cue for shape. Shape-from-shading has a long and rich history in computer vision and biological perception (Zhang et al. 1999; Durou et al. 2008), with most of the work focusing on distant lighting, orthographic projection and Lambertian surfaces (Horn and Brooks 1989; Kozera 1998; Zhang et al. 1999). Few papers

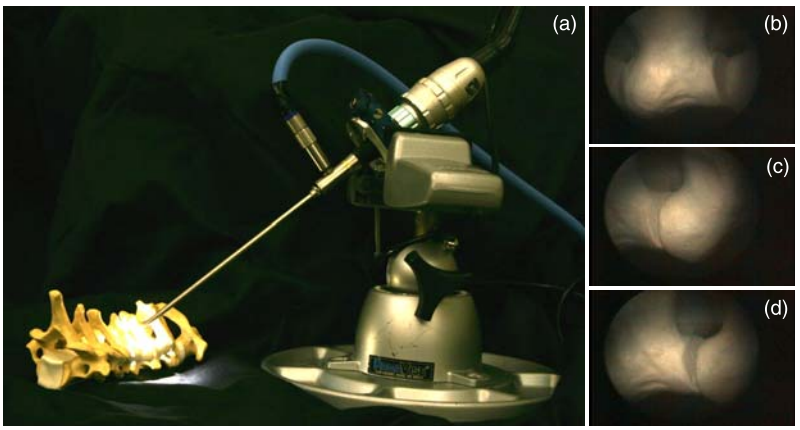

Fig. 1 Illustration of an endoscope (**a**) and endoscopic images of an artificial spine (**b**, **c**, **d**) (It is difficult to perceive the shape of the object in the images because of small field of view, distortion, lack of texture, partial boundaries, and shading caused by near-field lighting.)

have considered the more realistic scenarios. Real materials can be non-Lambertian model (Lee and Kuo 1997). With a pin-hole camera, the projection is perspective instead of orthographic (Penna 1989; Lee and Kuo 1994; Hasegawa and Tozzi 1996; Okatani and Deguchi 1997; Samaras and Metaxas 1999; Tankus et al. 2004). Considering the special imaging devices such as endoscopes, the light sources are not located at infinity (Okatani and Deguchi 1997; Prados and Faugeras 2005), thus $1/r^2$ attenuation term of the illumination need to be taken into account (Prados and Faugeras 2005). Three papers have established the equations for the perspective projection (Prados and Faugeras 2003; Tankus et al. 2003; Courteille et al. 2004). A variety of mathematical tools have been utilized to solve the perspective shape from shading (PSFS) problem (Kimmel and Sethian 2001; Courteille et al. 2004; Tankus et al. 2004, 2005; Prados and Faugeras 2005). Most relevant to this work, shape-from-shading under a near point light source and perspective projection for an endoscope has been addressed in Okatani and Deguchi (1997), Forster and Tozzi (2000), Mourgues et al. (2001) and Tankus et al. (2004). But all these works assume the light source and camera center are co-located. Note that this assumption is inaccurate in our setting since the scene is 5–10 mm away from the endoscope and the source and camera separation distance (3.5 mm) is of the same magnitude as the distance to the scene. In this paper we consider the perspective projection and near-point lighting without assuming that the light source is located at the center of the projection. Other works like photometric stereo can be used to estimate surface orientation (Woodham 1980; Kozera 1991, 1992) but it requires the switch of two light sources. However it is hard to achieve with a conventional endoscope since the light sources are designed to be switched on or off at the same time. Structured light has also been used to recover the bone shape (Fuchs et al. 1998; Iyengar et al. 2001) but the system requires specialized setups that are expensive and cumbersome to use in the operating room.

Due to the small field of view of the endoscope, only a partial shape from a single image can be obtained. By capturing image sequences as the endoscope is moved, it is possible to cover a larger part of the shape that can be more easily perceived (Seshamani et al. 2006). Once again, there is a long and rich history of shape-from-motion in vision (both calibrated and uncalibrated) (Poelman and Kanade 1997; Pollefeys et al. 1999; Hartley and Zisserman 2004). In the context of anatomical reconstruction, the heart coronary (Mourgues et al. 2001) and tissue phantom (Stoyanov et al. 2005) have been reconstructed using motion cues (image features and correspondences). However, it is difficult to establish correspondences for textureless (or featureless) bones.

Neither shape-from-shading nor shape-from-motion can individually solve the problem of bone reconstruction from endoscopic images. The key idea in this work is to utilize both strengths to develop a global shape-from-shading approach using multiple partial views. There are two contributions in this paper.

1. We establish the reflectance function R for a perspective projection and near-point light sources, without assuming that the light sources are located at the center of the projection. We solve an image irradiance equation based on this reflectance function. In our near-lighting perspective SFS problem, we explicitly include depth z in the reflectance function R, image irradiance equation and later as an unknown in the minimization function. In order to compute the reflectance map, we also calibrate the photometry of endoscope including camera response function, light source intensity and spatial distribution in advance.
2. We present a multi-image shape-from-shading framework. Given the global boundary constraints obtained from aligned individual contours in the world coordinates, by tracking the endoscope when capturing a sequence of images. As a clarification, note that our global shape-from-shading framework is not really merging

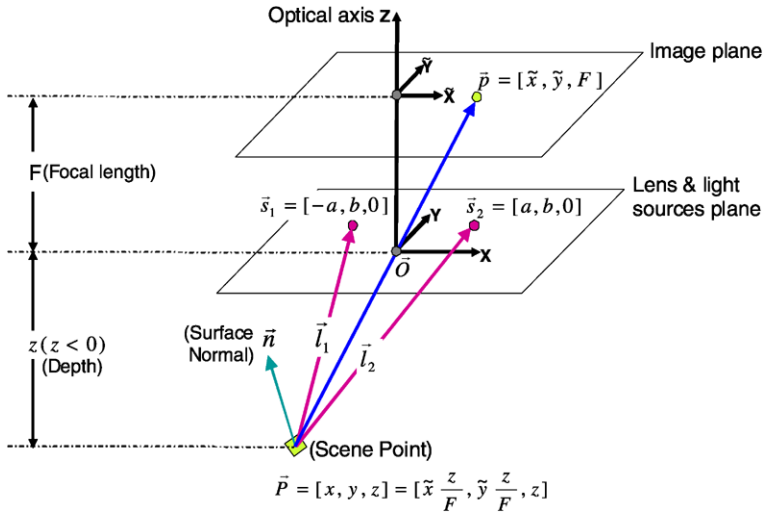

Fig. 2 Perspective projection model for endoscope imaging system with two near point light sources: $\vec{O}$ is the camera projection center. $\vec{s_1}$ and $\vec{s_2}$ are two light sources. We assume the plane consisting of $\vec{O}$, $\vec{s_1}$ and $\vec{s_2}$ is parallel to the image plane. The camera coordinate system (**X**–**Y**–**Z**) is centered at $\vec{O}$ and **Z**-axis is parallel to the optical axis and pointing toward the image plane. **X**-axis and **Y**-axis are parallel to the image plane. F is the focal length. a and b are two parameters related to the position of the light sources. Given a scene point $\vec{P} = (x, y, z)$, the projected image pixel is $\vec{p} = (\tilde{x}, \tilde{y}, F)$, where $(\tilde{x}, \tilde{y})$ are image coordinates. Assuming a Lambertian surface, the surface illumination therefore depends on the surface albedo, light source intensity and fall-off, and the angle between the normal and light rays

shape-from-shading and shape-from-motion in the classical sense of the terms. We perform the shape-from-shading for all images simultaneously and use global boundary constraints in each iteration. Meanwhile, in order to align the individual shapes, we calibrate the geometric parameters of endoscopes as well.

2 Shape-from-Shading under Near Point Sources and Perspective Projection

Given the special design of endoscopes, we formulate the shape-from-shading under near point lighting and a perspective projection, where the light sources are not located at the projection center. As Fig. 2 shows, given two point light sources $\vec{s_1}$ and $\vec{s_2}$, we can compute the scene radiance emitted by the surface point $\vec{P}$ according to the Lambertian cosine law and inverse square distance fall-off law of isotropic point sources (Horn and Sjoberg 1979):

$$R = I_0 \rho \left(\frac{\cos\theta_1}{r_1^2} + \frac{\cos\theta_2}{r_2^2} \right) \tag{1}$$

where I_0 is the light intensity of $\vec{s_1}$ and $\vec{s_2}$, ρ is the surface albedo. We use the unit vector $\vec{n}$ to represent the surface normal at $\vec{P}$, $\vec{l_1}$ and $\vec{l_2}$ are two light rays incident at $\vec{P}$, and r_1 and r_2 are the distances from each light source to the surface, and we have

$$\begin{aligned} \cos\theta_1 &= \frac{\vec{n} \cdot \vec{l_1}}{\|\vec{l_1}\|}, \quad \vec{l_1} = \vec{s_1} - \vec{P}, r_1 = \|\vec{l_1}\| \\ \cos\theta_2 &= \frac{\vec{n} \cdot \vec{l_2}}{\|\vec{l_2}\|}, \quad \vec{l_2} = \vec{s_2} - \vec{P}, \ r_2 = \|\vec{l_2}\| \end{aligned} \tag{2}$$

According to Horn and Brooks (1989), surface normal $\vec{n}$ can be represented in terms of partial derivatives of depth z w.r.t. x and y ((x, y, z) are camera coordinates):

$$\vec{n} = \left[-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1 \right] \Big/ \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 + 1} \tag{3}$$

Note that for the orthographic projection $\partial z/\partial \tilde{x}$ and $\partial z/\partial \tilde{y}$ are used to represent surface normals, where $(\tilde{x}, \tilde{y})$ are image coordinates. Under the perspective projection we have

$$\begin{aligned} x &= \tilde{x}\frac{z}{F} \\ y &= \tilde{y}\frac{z}{F} \end{aligned} \tag{4}$$

where F is the focal length and $z < 0$. We take the derivative of both sides of (4) w.r.t. x and y and obtain

$$\begin{aligned} \frac{\partial z}{\partial x} &= \frac{1}{\tilde{x}} \left(F - \frac{z}{\partial x / \partial \tilde{x}} \right) \\ \frac{\partial z}{\partial y} &= \frac{1}{\tilde{y}} \left(F - \frac{z}{\partial y / \partial \tilde{y}} \right) \end{aligned} \tag{5}$$

We take the derivative of both sides of (4) w.r.t. $\tilde{x}$ and $\tilde{y}$ and obtain

$$\begin{aligned}\frac{\partial x}{\partial \tilde{x}} &= \frac{1}{F}\left(z+\tilde{x}\frac{\partial z}{\partial \tilde{x}}\right)\\ \frac{\partial y}{\partial \tilde{y}} &= \frac{1}{F}\left(z+\tilde{y}\frac{\partial z}{\partial \tilde{y}}\right)\end{aligned} \tag{6}$$

Let $p = \partial z/\partial \tilde{x}$ and $q = \partial z/\partial \tilde{y}$, from (5) and (6) we have

$$\begin{aligned}\frac{\partial z}{\partial x} &= \frac{Fp}{z+\tilde{x}p}\\ \frac{\partial z}{\partial y} &= \frac{Fq}{z+\tilde{y}q}\end{aligned} \tag{7}$$

Given two light sources $s_1 = [-a, b, 0]$ and $s_2 = [a, b, 0]$ (calibration of a, b, F will be discussed later in Sect. 3.1), we can explicitly write light source vectors as follows:

$$\begin{aligned}\vec{l_1} &= \left[-a-\tilde{x}\frac{z}{F}, b-\tilde{y}\frac{z}{F}, -z\right]\\ \vec{l_2} &= \left[a-\tilde{x}\frac{z}{F}, b-\tilde{y}\frac{z}{F}, -z\right]\end{aligned} \tag{8}$$

Combine (1, 2, 3, 7) and (8), we obtain the reflectance map R as a function of $\tilde{x}, \tilde{y}, z, p, q$:

$$\begin{aligned}R(\tilde{x},\tilde{y},z,p,q) &= I_0\rho\left(\frac{\vec{n}(\tilde{x},\tilde{y},z,p,q)\cdot\vec{l_1}(\tilde{x},\tilde{y},z)}{r_1(\tilde{x},\tilde{y},z)^3}+\frac{\vec{n}(\tilde{x},\tilde{y},z,p,q)\cdot\vec{l_2}(\tilde{x},\tilde{y},z)}{r_2(\tilde{x},\tilde{y},z)^3}\right)\\ &= I_0\rho\left(\frac{p(z+\tilde{y}q)(a+\tilde{x}\frac{z}{F})-q(z+\tilde{x}p)(b-\tilde{y}\frac{z}{F})-\frac{z}{F}(z+\tilde{x}p)(z+\tilde{y}q)}{\sqrt{p^2(z+\tilde{y}q)^2+q^2(z+\tilde{x}p)^2+\frac{(z+\tilde{x}p)^2(z+\tilde{y}q)^2}{F^2}}\cdot\left(\sqrt{(-a-\tilde{x}\frac{z}{F})^2+(b-\tilde{y}\frac{z}{F})^2+(z)^2}\right)^3}\right.\\ &\quad\left.+\frac{-p(z+\tilde{y}q)(a-\tilde{x}\frac{z}{F})-q(z+\tilde{x}p)(b-\tilde{y}\frac{z}{F})-\frac{z}{F}(z+\tilde{x}p)(z+\tilde{y}q)}{\sqrt{p^2(z+\tilde{y}q)^2+q^2(z+\tilde{x}p)^2+\frac{(z+\tilde{x}p)^2(z+\tilde{y}q)^2}{F^2}}\cdot\left(\sqrt{(a-\tilde{x}\frac{z}{F})^2+(b-\tilde{y}\frac{z}{F})^2+(z)^2}\right)^3}\right)\end{aligned} \tag{9}$$

Equation (9) is similar to (2) in Tankus et al. (2004) and (2) in Prados and Faugeras (2003), but we consider the $1/r^2$ fall-off since the light sources are very close to the bone surface. Equation (9) can be easily extended to multiple point light sources as follows (M is the number of sources):

$$R(\tilde{x},\tilde{y},z,p,q) = I_0\rho\sum_i^M\left(\frac{\vec{n}(\tilde{x},\tilde{y},z,p,q)\cdot\vec{l_i}(\tilde{x},\tilde{y},z)}{r_i(\tilde{x},\tilde{y},z)^3}\right) \tag{10}$$

2.1 Solve Image Irradiance Equation

Given the image irradiance function $I(\tilde{x},\tilde{y})$, the image irradiance equation (Horn and Brooks 1989) is

$$R(\tilde{x},\tilde{y},z,p,q) = I(\tilde{x},\tilde{y}) \tag{11}$$

According to Durou et al. (2008), different mathematical methods have been proposed to solve the equation for PSFS, including methods of resolution of PDEs (Tankus et al. 2004, 2005; Kimmel and Sethian 2001; Prados and Faugeras 2005) and methods using optimization (Courteille et al. 2004). The optimization methods based on the variational approach can work in most general case of PSFS and there is no need to set values to the singular and local minimum points (Durou et al. 2008), so we adopt the optimization method and plug it in the multi-image framework described later in Sect. 3. Specifically, we solve (11) by minimizing the error between the image irradiance $I(\tilde{x},\tilde{y})$ and the reflectance map $R(\tilde{x},\tilde{y},z,p,q)$ given in (9). Image irradiance is obtained from the image intensity using the camera response function (see details in Sect. 2.2). Different from the previous optimization methods (Ikeuchi and Horn 1981), depth z is explicitly included in R. The error can be computed as:

$$\mathbf{e}(z,p,q) = \lambda\mathbf{e_i}(z,p,q) + (1-\lambda)\mathbf{e_s}(z,p,q) \tag{12}$$

and

$$\begin{aligned}&\mathbf{e_i}(z,p,q)\\ &= \int\int_{image}[I(\tilde{x},\tilde{y}) - R(\tilde{x},\tilde{y},z,p,q)]^2 d\tilde{x}d\tilde{y}\\ &\mathbf{e_s}(z,p,q)\\ &= \int\int_{image}[(z_{\tilde{x}}^2+z_{\tilde{y}}^2)+(p_{\tilde{x}}^2+p_{\tilde{y}}^2)+(q_{\tilde{x}}^2+q_{\tilde{y}}^2)]d\tilde{x}d\tilde{y}\end{aligned} \tag{13}$$

where $\mathbf{e_i}(z,p,q)$ is the irradiance error and $\mathbf{e_s}(z,p,q)$ is the smoothness constraint for z, p and q. λ is a Lagrange multiplier. We find a solution to (z,p,q) by minimizing the error $\mathbf{e}(z,p,q)$ (Leclerc and Bobick 1991):

$$[z^*, p^*, q^*] = \arg\min_{z,p,q}(\lambda \mathbf{e_i} + (1-\lambda)\mathbf{e_s}) \tag{14}$$

Similar to Horn and Brooks (1986), we discretize (12) and obtain:

$$\mathbf{e}(z_{k,l}, p_{k,l}, q_{k,l}) = \sum_k \sum_l (\lambda \mathbf{e_i}_{k,l} + (1-\lambda)\mathbf{e_s}_{k,l}) \tag{15}$$

The solution to (15) is to find $[p^*_{k,l}, q^*_{k,l}, z^*_{k,l}]$ that minimize $\mathbf{e}$:

$$\frac{\partial \mathbf{e}}{\partial p_{k,l}}\Big|_{p^*_{k,l}} = 0, \qquad \frac{\partial \mathbf{e}}{\partial q_{k,l}}\Big|_{q^*_{k,l}} = 0, \qquad \frac{\partial \mathbf{e}}{\partial z_{k,l}}\Big|_{z^*_{k,l}} = 0 \tag{16}$$

Combing (15) to (16), yields the update functions for $p_{k,l}$, $q_{k,l}$ and $z_{k,l}$ in each iteration (Horn and Brooks 1986):

$$\begin{aligned}
p^{n+1}_{k,l} &= \bar{p}^n_{k,l} \\
&\quad + \frac{\lambda}{4(1-\lambda)}[I_{k,l} - R(k,l,\bar{z}^n_{k,l}, \bar{p}^n_{k,l}, \bar{q}^n_{k,l})]\frac{\partial R}{\partial p_{k,l}}\Big|_{\bar{p}^n_{k,l}} \\
q^{n+1}_{k,l} &= \bar{q}^n_{k,l} \\
&\quad + \frac{\lambda}{4(1-\lambda)}[I_{k,l} - R(k,l,\bar{z}^n_{k,l}, \bar{p}^n_{k,l}, \bar{q}^n_{k,l})]\frac{\partial R}{\partial q_{k,l}}\Big|_{\bar{q}^n_{k,l}} \\
z^{n+1}_{k,l} &= \bar{z}^n_{k,l} \\
&\quad + \frac{\lambda}{4(1-\lambda)}[I_{k,l} - R(k,l,\bar{z}^n_{k,l}, \bar{p}^n_{k,l}, \bar{q}^n_{k,l})]\frac{\partial R}{\partial z_{k,l}}\Big|_{\bar{z}^n_{k,l}}
\end{aligned} \tag{17}$$

$\bar{p}^n_{k,l}$, $\bar{q}^n_{k,l}$ and $\bar{z}^n_{k,l}$ are local 8-neighborhood mean of values around pixel position (k,l). Detailed derivation can be found in Appendix A. The value for the $(n+1)$th iteration can be estimated from the nth iteration. During the iteration, the Lagrange multiplier λ is gradually increased such that the smoothness constraint is reduced as well. In our experiment the setting of Lagrance multiplier is empirical. We set it to 0.005 at the beginning and increase it by 0.02 whenever the error in (15) is reduced by 1%.

Given initial values of p, q and z on the boundary, we can compute a numerical solution to the shape. Variational methods usually rely on good initial guesses, so we manually label the boundaries of each images. After that, z is set to 0 for each pixel in the image. p and q are set to 0 for pixels that are not on the occluding boundaries. p and q at each pixel on the occluding boundaries are computed by performing the cross product of the viewing vector (starting from the optical center) and edge vector. Since it is difficult to estimate the initial values of z, we recompute z values by integrating the normals using the method in Frankot and Chellappa (1988), after several iterations when the error in (15) is reduced by 10%. We keep performing this adjustment until the algorithm converges. Results for real endoscopic images are showed in Fig. 3. It is still challenge to recover the entire shape of the spine from the individual reconstructions due to the small field of view.

2.2 Photometric Calibration of Endoscope for Image Irradiance

In order to compute the image irradiance $I(\tilde{x}, \tilde{y})$, we need to calibrate the photometry of the endoscope, which is seldom addressed in the literature. We propose a method to compute the radiometric response function of the endoscope, and also simultaneously calibrate the directional/spatial intensity distribution of the near light sources (the camera and the sources are "on/off" at the same time) and light intensities, inspired by Litvinov et al.'s (2005) work.

The image irradiance $I(\tilde{x}, \tilde{y})$ is related to the image intensity or gray level v via the camera response function $H(\cdot)$:

$$I(\tilde{x}, \tilde{y}) = \frac{H^{-1}[v(\tilde{x}, \tilde{y})]}{M(\tilde{x}, \tilde{y})} \tag{18}$$

where, $M(\tilde{x}, \tilde{y})$ represents the anisotropy of the source intensity. The two sources are identical and are oriented in the same way so we assume their intensity distributions $M(\tilde{x}, \tilde{y})$ are the same. From (1, 2, 11) and (18) we have:

$$\begin{aligned}
&H^{-1}[v(\tilde{x}, \tilde{y})] = \rho \cdot I_0 \cdot \tilde{M}(\tilde{x}, \tilde{y}) \quad \text{where} \\
&\tilde{M}(\tilde{x}, \tilde{y}) = M(\tilde{x}, \tilde{y}) \cdot \left(\frac{\vec{n} \cdot \vec{l_1}}{r_1^3} + \frac{\vec{n} \cdot \vec{l_2}}{r_2^3} \right)
\end{aligned} \tag{19}$$

For calibration, we use a Macbeth color chart with the known albedo for each patch. We capture a set of images by varying the source intensity for each patch (the exposure time of the camera cannot be controlled in our endoscope). We apply log to both sides of (19) to obtain a linear system:

$$h[v^j_i(\tilde{x}, \tilde{y})] = \eta_i + \gamma_j + \tilde{m}(\tilde{x}, \tilde{y}) \tag{20}$$

where, i indicates the surface albedo and j indexes the light intensity. $h[v^j_i(\tilde{x}, \tilde{y})] = \log\{H^{-1}[v^j_i(\tilde{x}, \tilde{y})]\}$, $\eta_i = \log(\rho_i)$, $\gamma_j = \log(I_{0j})$ and $\tilde{m}(\tilde{x}, \tilde{y}) = \log[\tilde{M}(\tilde{x}, \tilde{y})]$. The unknowns $(h(\cdot), \gamma_j, \tilde{m}(\tilde{x}, \tilde{y}))$ can be estimated by solving this linear system of equations. The cosine term is then estimated by physically measuring the distance to the chart from the scope tip and finally $M(\tilde{x}, \tilde{y})$ is recovered. Details of solutions to $h(\cdot)$, γ_j and $\tilde{m}(\tilde{x}, \tilde{y})$ are given in Appendix B.

A series of images of color chart are used for photometric calibration. We use 6 different levels of light intensity (see Appendix B). Figure 11(a), (b) and (c) show the camera response functions with Red, Green, Blue channels. Figure 11(d) shows the recovered light intensities in different levels and a comparison to the ground truth. Smaller numbers in x-axis correspond to higher intensities. We observe a small amount of variation when the light intensity is high, which can be caused by the saturation. Figure 11(e) shows the original image and Fig. 11(f) shows $\tilde{M}(\tilde{x}, \tilde{y})$. Figure 11(g) shows the cosine term $\frac{\vec{n} \cdot \vec{l_1}}{r_1^3} + \frac{\vec{n} \cdot \vec{l_2}}{r_2^3}$ and

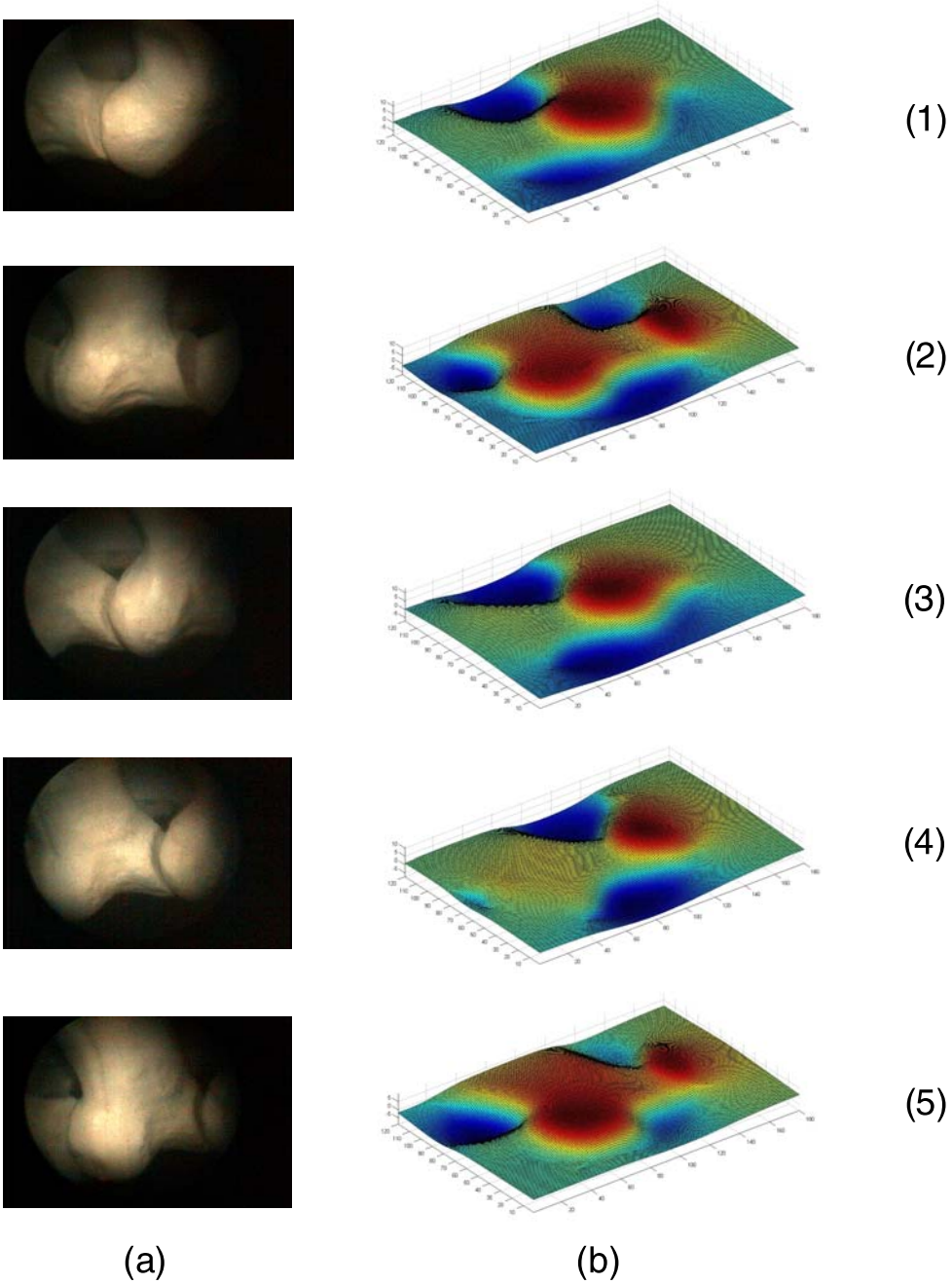

Fig. 3 Results of shape from shading from single image. (**a**) Input image. (**b**) Shape from shading. (1)–(5) are captured from different positions

Fig. 11(h) shows the spacial distribution function $M(\tilde{x}, \tilde{y}) = \tilde{M}(\tilde{x}, \tilde{y})/(\frac{\vec{n}\cdot\vec{l}_1}{r_1^3} + \frac{\vec{n}\cdot\vec{l}_2}{r_2^3})$.

3 Global Shape-from-Shading Using Multiple Partial Views

As shown in Fig. 3, most images include a small fraction of visible contours. How do we estimate the complete shape reliably? An intuitive solution is to merge individual shapes recovered from different views as shown in Fig. 4.

However, each image is defined in a view dependent local coordinates. We must first transform them to the world coordinates before we combine individual shapes. We solve this problem by using a navigation tracking system. We attach optical markers to the endoscope (as showed in Fig. 10) and track them to obtain the transformation between the local and world coordinates for each image.

Consider a 3D-point set $[S^i]_{3\times N_i}$ reconstructed from the image i, where N_i is the number of points in shape S^i. If all the shapes are reconstructed correctly, and without any errors from the tracking system and the calibration process,

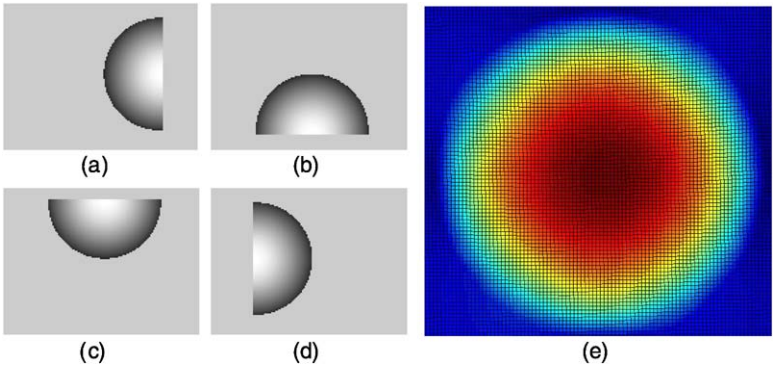

Fig. 4 Simulation results of shape from shading from multiple views. (**a**)–(**d**) Synthesized images of different parts of a sphere. (**e**) Reconstructed sphere

they should be perfectly aligned in the world coordinates. Let ${}^{w}T_{m_i}$ denote the transformation from the marker at the position taking ith image to the world coordinates, and M denote the transformation from the marker coordinates to the camera coordinates (image distortion has been corrected in advance). Thus, any 3D-point set S^i in the local camera coordinates can be transformed to the world coordinates $\hat{S}^i_w$ by:

$$\hat{S}^i_w = {}^{w}T_{m_i} \cdot [M^{-1} \cdot S^i; \mathbf{1}_{1\times N_i}] \tag{21}$$

where, $\hat{S}^i_w$ is represented using homogenous coordinates in the world reference frame. As shown in Fig. 5, each shape from single image is transformed to the world coordinates. However, due to the tracking error (less than 0.3 mm) and calibration error (less than 6 pixels), $\hat{S}^i_w$ are not initially well matched (e.g. Fig. 5(c) and (d)). Since the bone is rigid, we use ICP (Besl and McKay 1992) to improve the matching between different views by ensuring overlapping fields of view in adjacent images. Then, for $\hat{S}^i_w$, ICP yields an aligned shape ${}^{i-1}\hat{S}^i_w$ with respect to the adjacent shape $\hat{S}^{i-1}_w$:

$${}^{i-1}\hat{S}^i_w = {}^{i-1}T_i * \hat{S}^i_w \tag{22}$$

where, ${}^{i-1}T_i$ is the rigid transformation computed from ICP.

Using ICP, we can align all the shapes from one view with respect to the shape in the reference view, i.e. the shape $\hat{S}^1_w$ in the first image. Then, by combining (21) and (22), we have

$${}^{1}\hat{S}^i_w = \prod_{j=1}^{i-1} {}^{j}T_{j+1} \cdot {}^{w}T_{m_i} \cdot [M^{-1} \cdot S^i; \mathbf{1}_{1\times N_i}] \tag{23}$$

where ${}^{1}\hat{S}^i_w$ is the aligned shape of $\hat{S}^i_w$ with respect to the reference shape $\hat{S}^1_w$, as shown in Fig. 6(a) and (b).

Once the contours are aligned, we "re-grow" the shape in the local views but under the global constraints. We compute the average value in (17) using the neighbors in the world coordinates. The key idea is to keep the global constraints updated according to the shape developed in the local view, and therefore SFS in the local views are guided towards convergence under global constraints.

Table 1 Global shape-from-shading: key steps

*** Loop $k = 1$: number of images**
(1) Compute SFS from single image using (17)
(2) Track endoscope motions
(3) Transform local views to the world coordinates using (21)
(4) Register each shape to the previous shape using ICP using (22)
End loop
*** Compute global boundaries using (23)**
*** Compute global constraints (z, p, q) in the world coordinates**
*** Loop until converge**
(1) **Loop $k = 1$: number of images**
(a) Update (z, p, q) in the local views according to the global constraints
(b) Compute new (z, p, q) by using (17)
End loop
(2) Update (z, p, q) in the world coordinates
End loop

We restart the SFS in the local coordinates simultaneously for all images. After initializing (z, p, q) in each image, we initialize the values on the corresponding global boundaries. If there are two or more points from different views that are mapped to the same point in the world coordinates, (z, p, q) at that point are computed by the average. In each iteration, we grow each image and update the global constraints. Compared with the constraints for single shape-from-shading (see Fig. 5(a)), we can obtain more information from global alignments (see Fig. 6(c)). At the end of each iteration, we update the local (z, p, q) according to the global constraints. We proceed this until the convergence. With global constraints, our algorithm converges quickly though not in real-time (around 5 minutes for 18 images in Matlab with P4 2.4 Ghz CPU). Table 1 lists the key steps for global shape from shading.

For comparison, we choose only the points that are on the surface of the spine. We use a laser range scanner to obtain

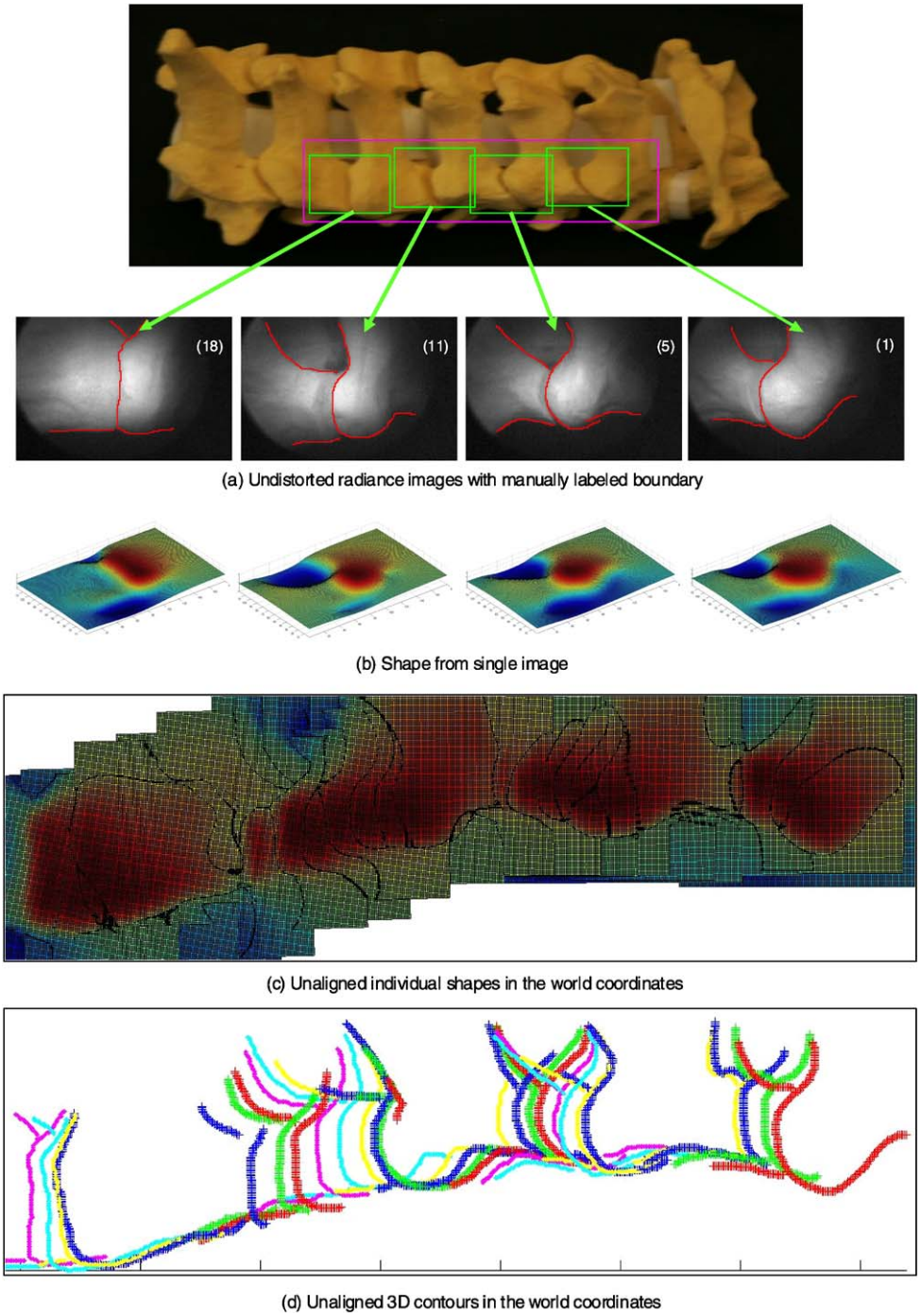

Fig. 5 Illustration of the problem about directly merging the individual shapes in the world coordinates. 18 images are captured by moving the endoscope horizontally (only translation). Four of them are shown as an illustration. (**a**) After removing the distortion and illumination effects, the boundaries in each image are labelled by hand, and the initial (p, q) are computed automatically on the boundaries. (**b**) Shape from each single image are reconstructed using the method described in Sect. 2. (**c**) Unaligned shapes in the world coordinates. (**d**) Unaligned 3D contours in the world coordinates

a ground truth shape of the spine. Then, the maximum, minimum, mean and RMS errors are 2.8 mm, 0.0 mm, 1.24 mm and 1.45 mm respectively. This level of accuracy is practical for surgery. More views of the reconstructed surface against the ground truth are shown in Fig. 7. We also compare our result with the result under the orthographic projection assumption in Fig. 8. Compared with the real shape captured by a regular camera, we can tell that the shape under the orthographic projection is over grown near the boundary.

3.1 Geometric Calibration of Endoscope for Aligning Multi-Views

In order to align the individual shapes, we need to know the camera parameters. Oblique-viewing endoscopes (oblique scope) are widely used in computer assisted surgeries. Viewing direction of an oblique scope can be changed by rotating the scope cylinder, which extends the field of view, but makes the scope geometric calibration process more difficult. An oblique endoscope is illustrated in Fig. 9.

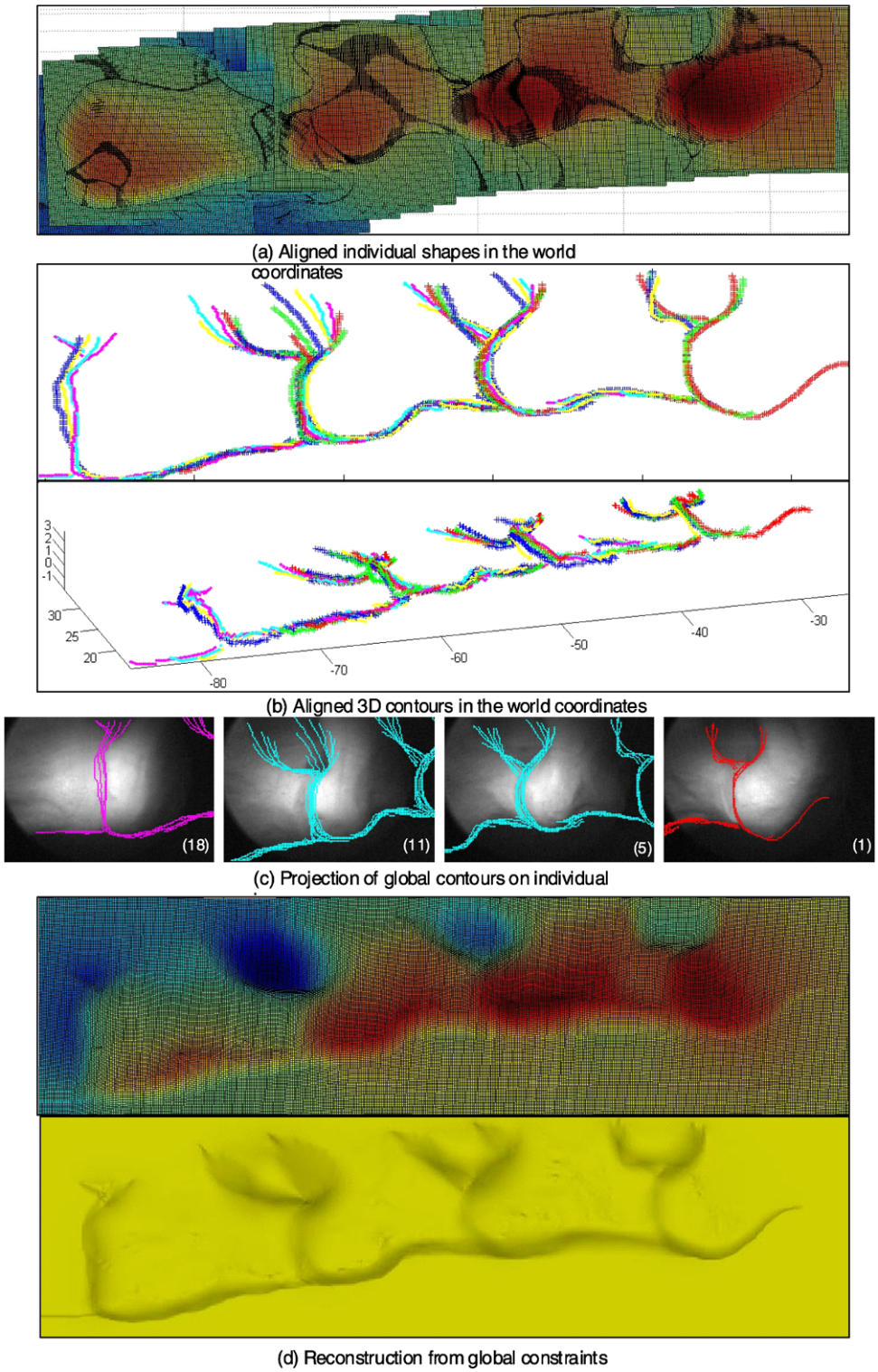

Fig. 6 Illustration of the multi-image shape-from-shading. (**a**) Aligned shape in the world coordinates. (**b**) Aligned 3D contours in the world coordinates. (**c**) Projection of the global constraints onto each image. (**d**) Final shape are reconstructed using the method described in Sect. 3

Unlike regular cameras with the forward viewing direction, most current endoscopes are designed for oblique viewing with a tilt from the scope axis at the tip (see Fig. 9(a)). The viewing direction can be changed by simply rotating the scope cylinder without having to move it. To increase flexibility of imaging in terms of focus and resolution, the lens system (in the cylinder) and the CCD plane (in the camera head) can be rotated with respect to one another (see Figs. 9 and 10(a)). While they are desirable, these properties make the geometric calibration harder than in regular cameras.

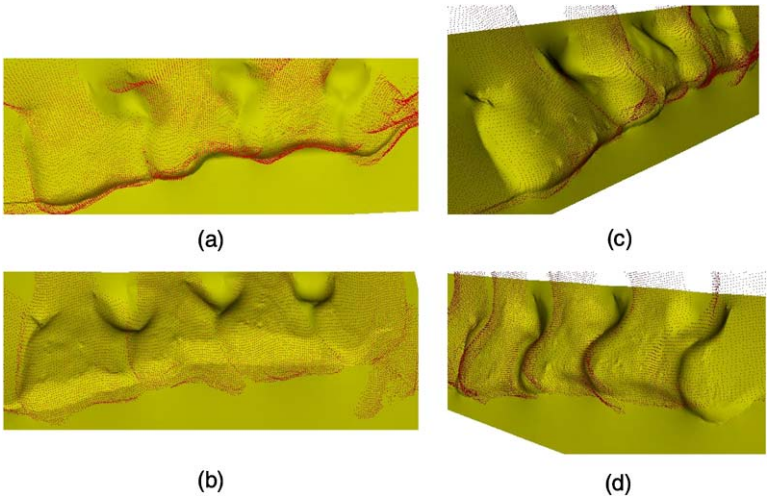

Fig. 7 (Color online) Different views of the reconstructed surface (*yellow surface*) against the ground truth (*red points*). (**a**) View from the top, (**b**) view from the bottom, (**c**) view from the left side, (**d**) view from the right side

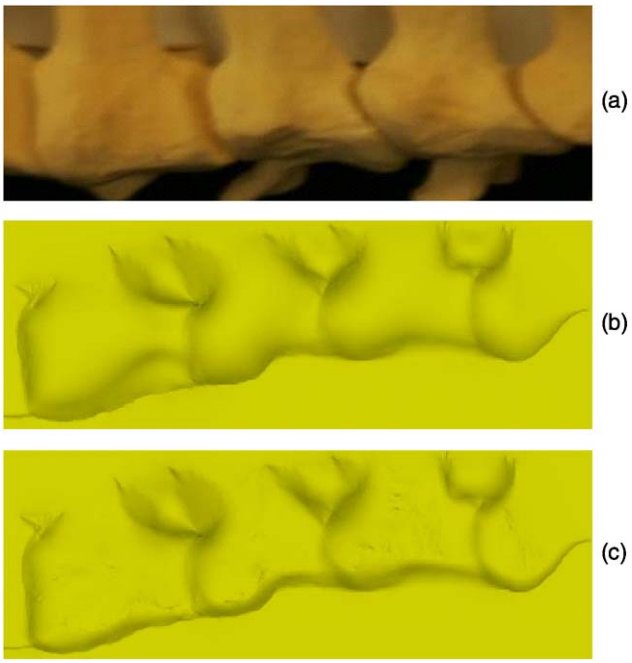

Fig. 8 (**a**) Real shape captured by a regular camera. (**b**) Reconstructed shape under orthographic projection. (**c**) Reconstructed shape under perspective projection

Yamaguchi et al. first modelled and calibrated the oblique scope (Yamaguchi et al. 2004). They formulated the rotation parameter of the scope cylinder as another external parameters in Tsai's camera model (Tsai 1987; Zhang 1998). They used two extra transformations to compensate the rotation θ of the lens system and still of the camera head. Yamaguchi et al.'s camera model successfully compensates the rotation effect but their method requires five additional parameters and the model is complicated.

Yamaguchi et al.'s camera model is based on Tsai's model:

$$\lambda\vec{\mu}_i = A\,{}^cT_m(\theta)\,{}^mT_w\,\vec{P}_w$$
$${}^cT_m(\theta) = T_R(-\theta;\vec{l}_h(\theta))T_R(\theta;\vec{l}_s)\,{}^cT_m(0) \tag{24}$$

where λ is an arbitrary scale factor, $\vec{P}_w$ is a 3D point in the world coordinates, $\vec{\mu}_i$ is the corresponding 2D image pixel. A is the camera intrinsic matrix. ${}^mT_w$ is the rigid transformation from the world coordinates to the optical marker coordinates, ${}^cT_m(\theta)$ is a rigid transformation from the marker (camera head) to the camera coordinates. ${}^cT_m(\theta)$ is dependent on the rotation angle θ. By considering the marker coordinates (camera head) as a reference, only the lens system rotates while the camera head, i.e., the image plane, remains fixed irrespective of the rotation. They describe such a transformation due to the rotation by decomposing the one physical rotation into two mathematical rotations. $T_R(\theta;\vec{l}_s)$ is a rotation of both scope cylinder and the camera head (image plane) around the axis of cylinder $\vec{l}_s$. $T_R(-\theta;\vec{l}_h(\theta))$ is an inverse rotation of the image plane around the z-axis of lens system $\vec{l}_h$. Both $\vec{l}_s$ and $\vec{l}_h$ have two unknown parameters. Although this model works well, it is very complicated.

As Fig. 10(b) shows, in our work, we attach an optical marker to the scope cylinder instead. Our model is still an extension of Tsai's model but much simpler. The geometric model illustrated in Fig. 10(a) can be written as:

$$\lambda\vec{\mu}'_i = A\cdot{}^cT_m\cdot{}^mT_w\cdot\vec{P}_w$$
$$\vec{\mu}_i = R(\theta)\cdot(\vec{\mu}'_i - cc) + cc \tag{25}$$

where $\vec{P}_w$ is a 3D point in the world coordinates, $\vec{\mu}'_i$ is the corresponding 2D image pixel without any rotation, $\vec{\mu}_i$ is the image pixel with rotation θ. ${}^mT_w$ is the rigid transformation from world coordinates to the optical marker coordinates (scope cylinder). ${}^cT_m$ is the rigid transformation

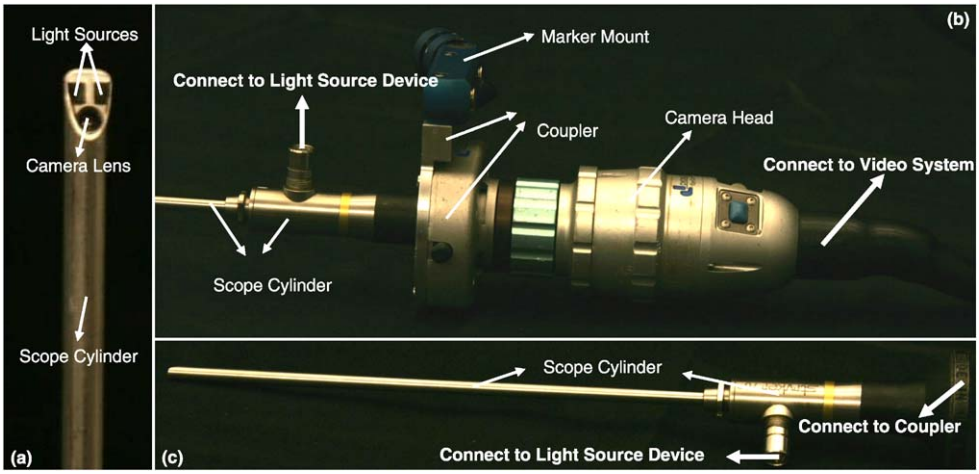

Fig. 9 An oblique endoscope consists of a scope cylinder with a lens and two point light sources at the tip (the tip has a tilt from the scope cylinder) and a camera head that captures video. The scope cylinder is connected flexibly to the camera head via a coupler. The scope cylinder or the camera head can be rotated together or separately

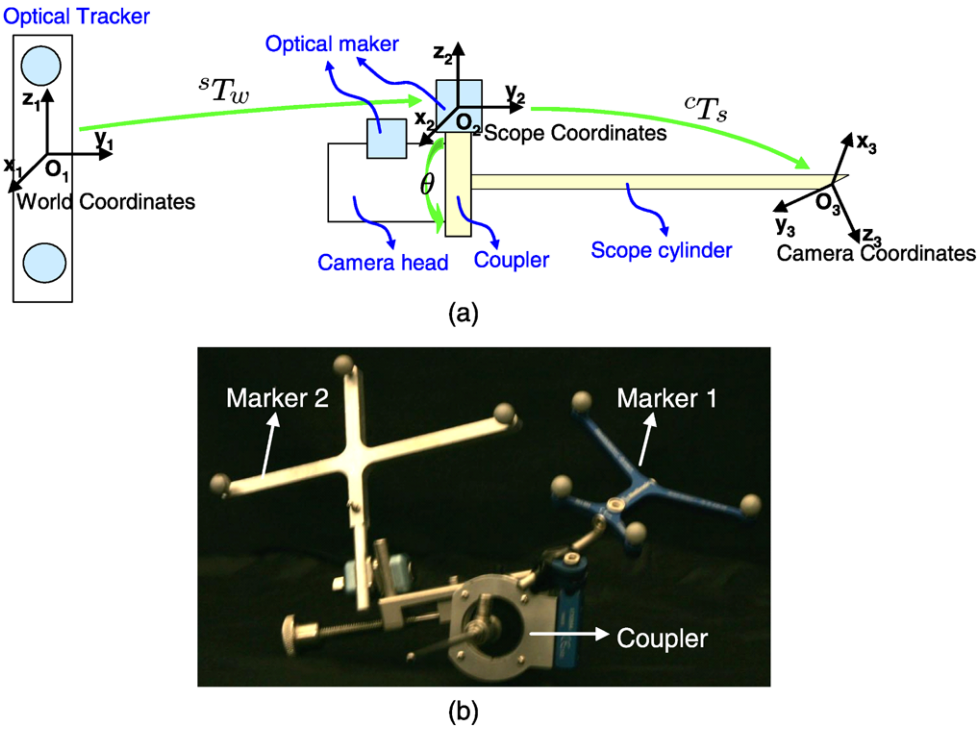

Fig. 10 The geometric model of endoscope based on a tracking system. A new coupler (see Fig. 9(**b**)) is designed to mount an optical marker to the scope cylinder which ensures that the transformation from scope cylinder(marker) coordinates $\mathbf{O}_2$ to the lens system (camera) coordinates $\mathbf{O}_3$ is fixed. World coordinates $\mathbf{O}_1$ is defined by the optical tracker. Two optical markers are attached to the coupler and camera head separately in order to compute the rotation θ in between

from the marker to the camera coordinates and independent on θ. cc is the principal point which is an intrinsic parameter. $R(\theta)$ represents a rotation of the image plane around cc by θ. Camera intrinsic matrix A and external matrix ^cT_m can be calibrated by using Zhang's (1998) method and ^mT_w can be obtained directly from the tracking system. In our model we only need to estimate the rotation angle.

Rotation angle can be estimated by using a rotary encoder, as Yamaguchi et al. (2004) suggested. When it is not available, the rotation angle can be estimated by using two optical markers: one attached to the scope cylinder and the other one on the camera head, see details in Appendicies C and D.

We tested our algorithm using two different setups. In the first experiment, we tested the system in the controlled environment. We used Stryker 344-71 arthroscope Vista (70 degree, 4 mm) oblique-viewing endoscope, DYONICS DyoCam^{TM} 750 video camera, DYONICS DYOBRITE 3000 light source, Polaris (Northern Digital Inc., Ontario, Canada) optical tracker. In the second experiment we tested it in the operating room. We used Smith & Nephew video arthroscope—autoclavable SN-OH 272589 (30 degree, 4 mm), DYONICS video camera and light source, OPTOTRAK (Northern Digital Inc., Ontario, Canada) optical tracker.

For each rotation angle of the endoscope, we computed the average back projection error for this angle

$$\epsilon(\theta) = \frac{1}{M}\sum_{i=1}^{M} |\vec{\mu}_i - \vec{\mu}(\vec{P}_i, \theta)| \tag{26}$$

where $\vec{P}_i$ is a 3D point in the world coordinates, $\vec{\mu}_i$ is the corresponding 2D image pixel, $\vec{\mu}(\vec{P}_i, \theta)$ is the back projected 2D image pixel of $\vec{P}_i$ computed by using (25). M is the number of corners on the calibration pattern. We have used different grid patterns (3×4, 4×5, 5×6, 6×7. The size of each checker is 2 mm $\times$ 2 mm). In order to obtain enough light on the grid pattern, the endoscope needs to be placed very close to the target (usually 5–15 mm). Smaller grid cannot capture the radial distortion but the bigger grid will exceed the field of view. The 5×6 grid gave the best results.

We did many trials by moving and rotating the endoscope randomly and estimate θ simultaneously. The average back projection error with respect to different rotation angles are shown in Fig. 14 (see Appendix D). Figure 14(a) shows the result using Stryker 344-71 arthroscope Vista (70 degree, 4 mm) and Polaris optical tracker. Figure 14(b) shows the result using Smith & Nephew video arthroscope—autoclavable SN-OH 272589 (30 degree, 4 mm) and OPTOTRAK optical tracker. The red curve (square mark) represents the back projection error without taking into account of the rotation angle, and the blue curve (circular mark) shows the error with considering the rotation angle. The results show that including the rotation angle into the camera model significantly improve the accuracy of the calibration.

Figure 14 shows that two experiments have different accuracy. The reason is that endoscopes have different magnification and optical trackers have different accuracy (according to the manufacturer, RMS error is 0.1 mm for OPTOTRAK and 0.3 mm for Polaris). Yamaguchi et al. (2004) used an OTV-S5C laparoscope (Olympus Optical Co. Ltd., Tokyo, Japan) and Polaris optical tracker. They have achieved a high accuracy of less than 5 mm back projection error when the rotation angle is within 140 degrees. Our results show that we can achieve the same level of accuracy when the rotation angle is within 75 degrees. Beyond this range, due to the bigger magnification, larger radial distortion and poorer lighting, the back projection error is increased to 13 mm when the rotation angle is 100 degrees. When given the same quality endoscopes, we should be able to achieve the same level of accuracy.

4 Conclusion and Discussions

Shape-from-shading and shape-from-motion have been successfully used in many vision applications, but both have difficulties for 3D reconstruction from orthopedic endoscopy due to the featureless bone surface and partial occluding boundaries in a small field of view. In this work, we propose a method to combine the strengths of both approaches to solve our problem: we formulate the shape-from-shading for endoscopes under near point light sources and perspective projection, and develop a multi-image shape-from-shading framework. To deal with the tracking and calibration errors, we use ICP to improve the alignment of partial shapes and contours in the world coordinates. As a result, we can reconstruct the shape of a larger bone area and provide useful visualizations for surgical navigation in the minimally invasive procedures.

A couple of issues worth further discussion in our framework. We build a near-lighting perspective shape-from-shading (NLPSFS) model without assuming that the light sources are located in the optical center. Our contribution is to present a new model for medical endoscopes, instead of a new solution to any existing models. We adopt the variational optimization method since it is effective for most general cases and no pre-defined values are involved for the singular and local minimum points. Since our multi-frame approach is very general, other PSFS solutions such as Prados et al.'s (2005) and Tankus et al.'s (2004) approach can be employed to solve the NLPSFS problem as well. We plan to compare different SFS methods in the future work. In this paper we focus on the formulation of the NLPSFS model and multi-image framework.

As mentioned in the previous section, in the real operating environment, the bone surfaces are usually 5–10 mm away from the endoscope. The light source and camera separation distance is 3.5 mm in our case. These two numbers are in the same magnitude as the distance to the scene. The approximation becomes inaccurate if we ignore the parameters a and b in (9). However, it is a reasonable approximation when the endoscope is far away from the scene. In this paper we attempt to build a general and accurate model to handle more realistic cases, since we can calibrate all parameters beforehand.

One advantage of our method is that the computation of single image SFS in the first step of Table 1 can be parallelized since the error function in (15) uses only one single image. Note in the last step of Table 1, the computation of single image SFS cannot be parallelized because of the global constraints need updates for each iteration.

By using ICP in our multi-image framework, we are able to deal with the tracking and calibration errors. Moreover, we can tolerant some shape errors caused by the single image SFS in the first step of Table 1. Therefore we use an early stop strategy during the single image SFS optimizations by terminating the iterations before the error in (15) reaches zero (we set the threshold to 15% of the initial error). By this means we significantly speed up the algorithm.

This is one of the reasons why the result from a single image tends to be flat. Other reasons include partial occluding boundaries and the smoothness constraints.

The results from multi-image SFS still lack high frequency shape details due to the over smoothing. We are working on applying shape prior to our multi-image framework and will describe the details in the future work. Dealing with image noise due to blood and tissues (Kozera and Noakes 2004) will be addressed in the future work as well.

Acknowledgements This work is supported in parts by an NSF Grant #0325920 ITR: Data-driven Human Knee Modeling for Expert Surgical Planning Systems, an NSF CAREER Award #IIS-0643628, an NSF Award #CCF-0541307 and an ONR Award #N00014-05-1-0188.

Appendix A: Iteration of Solutions to Shape Parameters (z, p, q)

In this appendix we will show the derivation for the update functions in (17). To minimize the error $\mathbf{e}$, we take the derivative of $\mathbf{e}$ w.r.t. $p_{k,l}$, where (k,l) is the position of an image pixel:

$$\begin{aligned}\frac{\partial \mathbf{e}}{\partial p_{k,l}} &= \frac{\partial}{\partial p_{k,l}}[(1-\lambda)\mathbf{e_s} + \lambda \mathbf{e_i}] \\ &= (1-\lambda)\frac{\partial \mathbf{e_s}}{\partial p_{k,l}} + \lambda \frac{\partial \mathbf{e_i}}{\partial p_{k,l}}\end{aligned} \tag{A.1}$$

where for each pixel (k,l) we have

$$\begin{aligned}\mathbf{e_s}_{k,l} &= (p_{k+1,l} - p_{k,l})^2 + (p_{k,l+1} - p_{k,l})^2 \\ &\quad + (q_{k+1,l} - q_{k,l})^2 + (q_{k,l+1} - q_{k,l})^2 \\ &\quad + (z_{k+1,l} - z_{k,l})^2 + (z_{k,l+1} - z_{k,l})^2 \\ \mathbf{e_i}_{k,l} &= [I_{k,l} - R(k,l,z_{k,l},p_{k,l},q_{k,l})]^2\end{aligned} \tag{A.2}$$

For smoothness constraints:

$$\begin{aligned}\frac{\partial \mathbf{e_s}}{\partial p_{k,l}} &= \frac{\partial \mathbf{e_s}_{k,l}}{\partial p_{k,l}} + \frac{\partial \mathbf{e_s}_{k-1,l}}{\partial p_{k,l}} + \frac{\partial \mathbf{e_s}_{k,l-1}}{\partial p_{k,l}} \\ &= 2((p_{k,l} - p_{k+1,l}) + (p_{k,l} - p_{k,l+1})) \\ &\quad + 2(p_{k,l} - p_{k-1,l}) + 2(p_{k,l} - p_{k,l-1}) \\ &= 8\left(p_{k,l} - \frac{1}{4}(p_{k+1,l} + p_{k,l+1} + p_{k-1,l} + p_{k,l-1})\right) \\ &= 8(p_{k,l} - \bar{p}_{k,l})\end{aligned} \tag{A.3}$$

where $\bar{p}_{k,l}$ is 4-neighbor average and we extend it to 8 neighbors:

$$\begin{aligned}\bar{p}^n_{k,l} &= \frac{1}{5}(p^n_{k,l-1} + p^n_{k,l+1} + p^n_{k+1,l} + p^n_{k-1,l}) \\ &\quad + \frac{1}{20}(p^n_{k-1,l-1} + p^n_{k-1,l+1} + p^n_{k+1,l-1} + p^n_{k+1,l+1})\end{aligned} \tag{A.4}$$

For brightness error:

$$\frac{\partial \mathbf{e_i}}{\partial p_{k,l}} = \frac{\partial \mathbf{e_i}_{k,l}}{\partial p_{k,l}} = -2(I_{k,l} - R(k,l,z_{k,l},p_{k,l},q_{k,l}))\frac{\partial R}{\partial p_{k,l}} \tag{A.5}$$

From (A.3) and (A.5) we have

$$\frac{\partial \mathbf{e}}{\partial p_{k,l}} = 8(1-\lambda)(p_{k,l} - \bar{p}_{k,l}) - 2\lambda(I_{k,l} - R_{k,l})\frac{\partial R}{\partial p_{k,l}} \tag{A.6}$$

and setting it to zero yields:

$$p_{k,l} = \bar{p}_{k,l} + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,z_{k,l},p_{k,l},q_{k,l}))\frac{\partial R}{\partial p_{k,l}} \tag{A.7}$$

In order to make the algorithm robust we also use the average $\bar{p}$, $\bar{q}$ and $\bar{z}$ when computing the reflectance map $R(\tilde{x},\tilde{y},z,p,q)$ in (A.7). We use the same method to derive the update for q and z and finally we have:

$$\begin{aligned}p_{k,l} &= \bar{p}_{k,l} \\ &\quad + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,\bar{z}_{k,l},\bar{p}_{k,l},\bar{q}_{k,l}))\frac{\partial R}{\partial p_{k,l}}\Big|_{\bar{p}_{k,l}} \\ q_{k,l} &= \bar{q}_{k,l} \\ &\quad + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,\bar{z}_{k,l},\bar{p}_{k,l},\bar{q}_{k,l}))\frac{\partial R}{\partial q_{k,l}}\Big|_{\bar{q}_{k,l}} \\ z_{k,l} &= \bar{z}_{k,l} \\ &\quad + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,\bar{z}_{k,l},\bar{p}_{k,l},\bar{q}_{k,l}))\frac{\partial R}{\partial z_{k,l}}\Big|_{\bar{z}_{k,l}}\end{aligned} \tag{A.8}$$

Appendix B: Solutions to Photometric Calibration

In this appendix, we will show how we separate and solve three unknown photometric parameters (response function, light intensity and light distribution).

B.1 Solution to $h(\cdot)$

Given the same light intensity γ_j and pixel value $v(\tilde{x},\tilde{y})$ but two different albedos η_{i_1} and η_{i_2}, we have

$$\begin{cases} h[v^j_{i_1}(\tilde{x},\tilde{y})] - \gamma_j - \tilde{m}(\tilde{x},\tilde{y}) - \eta_{i_1} = 0 \\ h[v^j_{i_2}(\tilde{x},\tilde{y})] - \gamma_j - \tilde{m}(\tilde{x},\tilde{y}) - \eta_{i_2} = 0 \end{cases} \tag{B.1}$$

Subtract the first line from the second line of (B.1) we obtain

$$h[v^j_{i_2}(\tilde{x},\tilde{y})] - h[v^j_{i_1}\tilde{x},\tilde{y})] = \eta_{i_2} - \eta_{i_1} \tag{B.2}$$

We use different pixels in the same image (albedo) or different images (albedos) to make many equations as (B.2),

as long as we fix the light intensity for each pair of albedos. Since $v_{i_1}^j(\tilde{x},\tilde{y})$ changes from 0 to 255 (image intensity), we only need 256 such equations and stack them as:

$$\begin{bmatrix} \cdots & & 1^* & & -1^* & \cdots \\ \cdots & -1^{\#} & & 1^{\#} & & \cdots \\ & \cdots & & & \cdots & \end{bmatrix} \cdot \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(255) \end{bmatrix} = \begin{bmatrix} \eta_{i_2} - \eta_{i_1} \\ \eta_{i_4} - \eta i_3 \\ \vdots \end{bmatrix} \tag{B.3}$$

where 1^* and -1^* correspond to the column $v_{i_2}^j(\tilde{x},\tilde{y})+1$ and $v_{i_1}^j(\tilde{x},\tilde{y})+1$, respectively. $1^{\#}$ and $-1^{\#}$ correspond to the column $v_{i_4}^j(\tilde{x},\tilde{y})+1$ and $v_{i_3}^j(\tilde{x},\tilde{y})+1$, respectively. Therefore $h(v)$ is solved from (B.3) and $H(\cdot)=\{\exp[h(v)]\}^{-1}$.

B.2 Solution to γ_j

Given the same albedo η_i and pixel value $v(\tilde{x},\tilde{y})$ but two different light intensities γ_{j_1} and γ_{j_2} we have

$$\begin{cases} h[v_i^{j_1}(\tilde{x},\tilde{y})] - \gamma_{j_1} - \tilde{m}(\tilde{x},\tilde{y}) - \eta_i = 0 \\ h[v_i^{j_2}(\tilde{x},\tilde{y})] - \gamma_{j_2} - \tilde{m}(\tilde{x},\tilde{y}) - \eta_i = 0 \end{cases} \tag{B.4}$$

Subtract the first line from the second line of (B.4):

$$h[v_i^{j_2}(\tilde{x},\tilde{y})] - h[v_i^{j_1}(\tilde{x},\tilde{y})] = \gamma_{j_2} - \gamma_{j_1} \tag{B.5}$$

We use the minimum light intensity γ_1 as a reference, for other light intensities γ_j, $j=2,\ldots,N_{light}$, we have

$$\gamma_j = \gamma_1 + \{h[v_i^j(\tilde{x},\tilde{y})] - h[v_i^1(\tilde{x},\tilde{y})]\} \tag{B.6}$$

Given estimated $h[v(\tilde{x},\tilde{y})]$ and by changing the albedos and pixels, we compute the average value for each γ_j as follows and $\bar{I_{0j}} = \exp(\gamma_1)\cdot\exp(\bar{\gamma_j})$.

$$\bar{\gamma_j} = \gamma_1 + \frac{1}{N_{albedo}} \cdot \frac{1}{N_{pixels}} \times \sum_{i}^{N_{albedo}} \sum_{\tilde{x},\tilde{y}}^{N_{pixels}} \{h[v_i^j(\tilde{x},\tilde{y})] - h[v_i^1(\tilde{x},\tilde{y})]\} \tag{B.7}$$

B.3 Solution to $\tilde{m}(\tilde{x},\tilde{y})$

Again, Given the same albedo η_i and light intensity γ_j but two different pixels $(\tilde{x}_p,\tilde{y}_p)$ and $(\tilde{x}_q,\tilde{y}_q)$ we have

$$\begin{cases} h[v_i^j(\tilde{x}_p,\tilde{y}_p)] - \gamma_j - \tilde{m}(\tilde{x}_p,\tilde{y}_p) - \eta_i = 0 \\ h[v_i^j(\tilde{x}_q,\tilde{y}_q)] - \gamma_j - \tilde{m}(\tilde{x}_q,\tilde{y}_q) - \eta_i = 0 \end{cases} \tag{B.8}$$

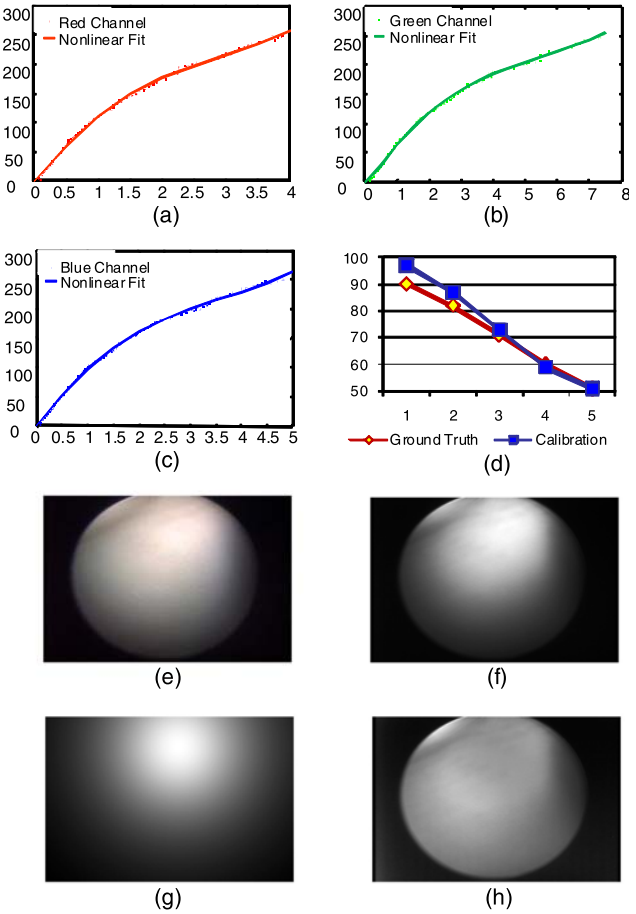

Fig. 11 (Color online) Results of photometric calibration. (**a**) Camera response function in Red channel. Red dots represents the data points and magenta line represents the nonlinear fit. x-axis represents the image irradiance, and y-axis represents the image intensity, same in (**b**) and (**c**). When gray level is equal to 0, the image irradiance is restricted to 0 no matter what we compute from the nonlinear fit. (**b**) Camera response function in Green channel. Green dots represents the data points and magenta line represents the nonlinear fit. (**c**) Camera response function in *blue* channel. *Blue dots* represents the data points and *magenta line* represents the nonlinear fit. (**d**) Calibrated light intensity in different level (*blue*) and ground truth (*green*). x-axis is the trail index and y-axis is the relative light intensity. We use level 6 as a reference and plot level 1–5 with the small levels corresponding to high light intensities. A bit variation in range of the high intensities may be caused by saturation. (**e**) Original image of color chart. (**f**) $\tilde{M}(\tilde{x},\tilde{y})$. (**g**) Cosine term $\frac{\vec{n}\cdot\vec{l_1}}{r_1^3} + \frac{\vec{n}\cdot\vec{l_2}}{r_2^3}$. (**h**) Spacial distribution function $M(\tilde{x},\tilde{y}) = \tilde{M}(\tilde{x},\tilde{y})/(\frac{\vec{n}\cdot\vec{l_1}}{r_1^3} + \frac{\vec{n}\cdot\vec{l_2}}{r_2^3})$

Subtract the first line from the second line of (B.8):

$$h[v_i^j(\tilde{x}_q,\tilde{y}_q)] - h[v_i^j(\tilde{x}_p,\tilde{y}_p)] = \tilde{m}(\tilde{x}_q,\tilde{y}_q) - \tilde{m}(\tilde{x}_p,\tilde{y}_p) \tag{B.9}$$

For each γ_j and η_i, instead of using $C^2_{N_{pixels}}$ different pairs of pixels, we choose only $N_{pixels} = 720 \times 480$ pairs

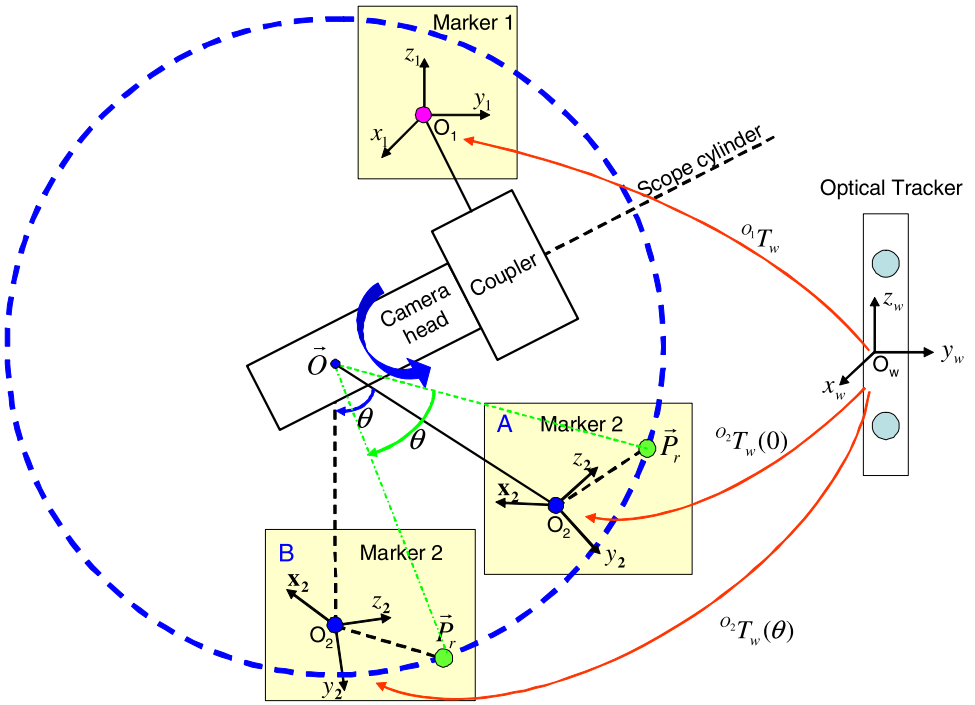

Fig. 12 Illustration of the relationship between the rotation angle θ and two markers' coordinates. Marker 1 is attached to the scope cylinder and Marker 2 is attached to the camera head. "A" indicates the position of Marker 2 when $\theta = 0$ and "B" indicates the position of Marker 2 given a rotation θ. For any point $\vec{P}_r$ in Marker 2's coordinates, its trace with the rotation of the camera head is a circle in Marker 1's coordinates (It moves from position $\vec{P}_r^A$ to $\vec{P}_r^B$ in the figure). This circle is also on the plane perpendicular to the axis of scope cylinder. $\vec{O}$ indicates the center of the circle

and stack the equations like the following:

$$\underbrace{\begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & & . & \\ & & & & . \\ -1 & & & & 1 \end{bmatrix}}_{\Phi} \cdot \begin{bmatrix} \tilde{m}(\tilde{x}_1, \tilde{y}_1) \\ \tilde{m}(\tilde{x}_2, \tilde{y}_2) \\ \vdots \\ \tilde{m}(\tilde{x}_{N_{pixels}}, \tilde{y}_{N_{pixels}}) \end{bmatrix}$$

$$= \begin{bmatrix} h[v_i^j(\tilde{x}_1, \tilde{y}_1)] - h[v_i^j(\tilde{x}_2, \tilde{y}_2)] \\ h[v_i^j(\tilde{x}_2, \tilde{y}_2)] - h[v_i^j(\tilde{x}_3, \tilde{y}_3)] \\ \vdots \\ h[v_i^j(\tilde{x}_{N_{pixels}}, \tilde{y}_{N_{pixels}})] - h[v_i^j(\tilde{x}_1, \tilde{y}_1)] \end{bmatrix} \quad \text{(B.10)}$$

It's not practical to solve (B.10) using SVD directly since matrix Φ requires a huge memory. However we notice that the matrix Φ in (B.10) is a special N_{pixels} by N_{pixels} matrix such that we could get the inverse directly by using Gauss-Jordan Elimination:

$$\Phi^{-1} = \begin{bmatrix} 0 & & & & 0 & -1 \\ -1 & 0 & & & 0 & -1 \\ -1 & -1 & 0 & & 0 & -1 \\ \vdots & \vdots & \vdots & \cdots & 0 & -1 \\ -1 & -1 & -1 & \cdots & -1 & -1 \end{bmatrix} \quad \text{(B.11)}$$

Thus we can efficiently compute each element of $\tilde{m}(\tilde{x}, \tilde{y})$ independently and again, $\tilde{M}(\tilde{x}, \tilde{y}) = \exp[\tilde{m}(\tilde{x}, \tilde{y})]$.

Appendix C: Estimation of Rotation Angle Using Two Optical Markers

Let the marker attached to the scope cylinder be Marker 1 and the marker to the camera head be Marker 2 (Fig. 10(b)). As Fig. 12 shows, when we rotate the camera head around the scope cylinder from the position "A" to "B" by θ, point $\vec{P}_r$ in Marker 2's coordinates $\mathbf{O}_2$ will move along a circle with respect to a point $\vec{O}$ on the axis of the scope cylinder, in Marker 1's coordinates $\mathbf{O}_1$. We can estimate the center $\vec{O}$ of the circle first and compute θ

$$\theta = \arccos \frac{\|\vec{OP}_r^A\|^2 + \|\vec{OP}_r^B\|^2 - \|\vec{P_r^A P_r^B}\|^2}{2\|\vec{OP}_r^A\| \cdot \|\vec{OP}_r^B\|} \quad \text{(C.1)}$$

The center of the circle can be represented in terms of the transformations from the world coordinates $\mathbf{O}_w$ to Marker 1's coordinates $\mathbf{O}_1$, and Marker 2's coordinates $\mathbf{O}_2$. At least 3 different positions of Marker 2 ($\mathbf{O}_2$) (with different θ) are necessary.

Appendix D: Estimation of the Center of Rotation Circle in 3D

We rotate the camera head around the cylinder to acquire 3 different positions of Marker 2. Let the transformation matrix from the world coordinates $\mathbf{O}_w$ to both Marker 1's coordinates $\mathbf{O}_1$ and Marker 2's coordinates $\mathbf{O}_2$ for position i be $({}^{o_1}T_{o_w}{}^i, {}^{o_2}T_{o_w}{}^i)$ $(i = 1, 2, 3)$. Given any point $\vec{P}_r$ in $\mathbf{O}_2$, we

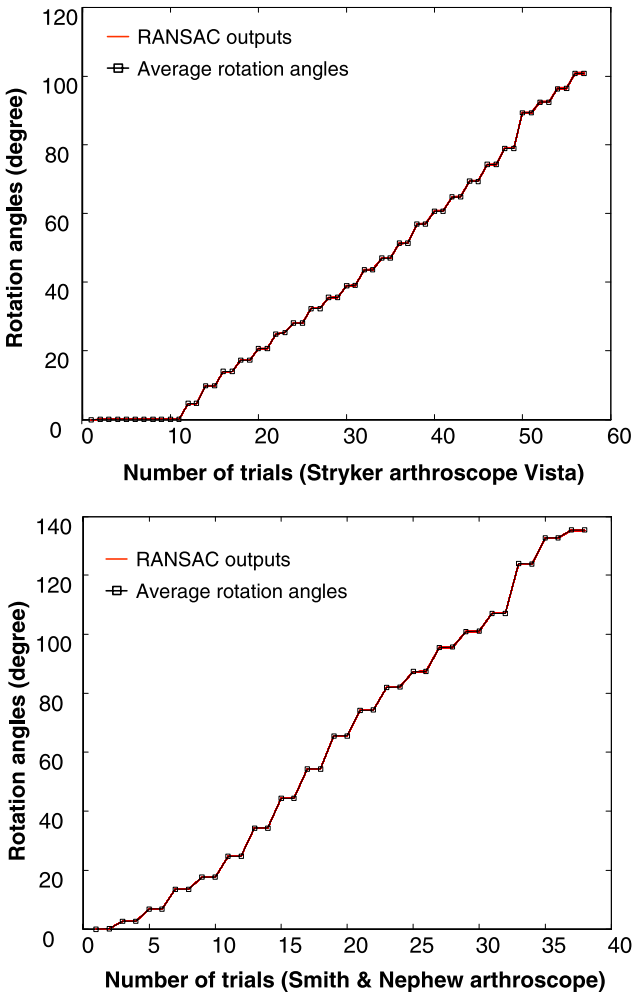

Fig. 13 (Color online) Estimated rotation angles for two endoscopes. In each trial we rotated the camera head with respect to the scope cylinder and captured an image. We captured a few images for the initial position. After that we took two images for each rotation angle. The *red curves* are estimated rotation angles from different RANSAC iterations. The *black curve* is the average rotation angle

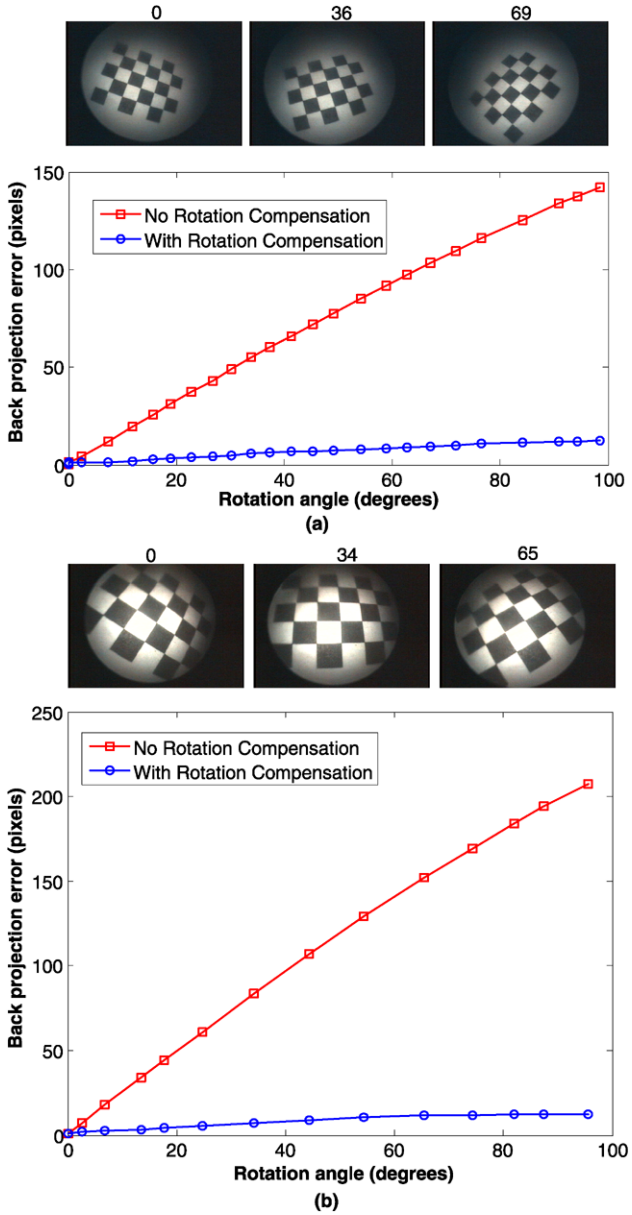

Fig. 14 (Color online) Back projection errors with respect to the rotation angles for two systems. (**a**) Stryker 344-71 arthroscope Vista and Polaris optical tracker. (**b**) Smith & Nephew video arthroscope and OPTOTRAK optical tracker in the operating room. Images in the top row of (**a**) and (**b**) correspond to different rotation angles (the angle is shown on the top of each image). The *red curves* in (**a**) and (**b**) represent the errors without a rotation compensation. The *blue curves* in (**a**) and (**b**) represent errors with a rotation compensation

first compute the position $\vec{P}_i$ in $\mathbf{O}_1$ corresponding to different rotations as:

$$\vec{P}_i = {}^{o_1}T_{o_w}{}^{i} \cdot ({}^{o_2}T_{o_w}{}^{i})^{-1} \cdot \vec{P}_r, \quad i = 1, 2, 3. \tag{D.1}$$

$\vec{O}$ is the center of the circumcircle of the triangle $(\Delta \vec{P}_1 \vec{P}_2 \vec{P}_3)$.

Let $\vec{R}_1 = \vec{P}_1 - \vec{P}_3$, $\vec{R}_2 = \vec{P}_2 - \vec{P}_3$, the normal of the triangle is $\vec{n}_\Delta = \vec{R}_1 \times \vec{R}_2$. The perpendicular bisector $\vec{L}_1$ of $\vec{R}_1$ and $\vec{L}_2$ of $\vec{R}_2$ can be computed as:

$$\begin{aligned} \vec{L}_1 &= \vec{P}_3 + \vec{R}_1/2 + \lambda_1 \cdot \vec{n}_\Delta \times \vec{R}_1 \\ \vec{L}_2 &= \vec{P}_3 + \vec{R}_2/2 + \lambda_2 \cdot \vec{n}_\Delta \times \vec{R}_2 \end{aligned} \tag{D.2}$$

where λ_1 and λ_2 are parameters of the line $\vec{L}_1$ and $\vec{L}_2$. The intersection of these two lines is the center of the circle. From (D.2) we can derive the center of the circle

$$\begin{aligned} \vec{O} = {} & \frac{(\vec{R}_2 - \vec{R}_1) \cdot \vec{R}_1/2}{|\vec{R}_1 \times \vec{R}_2|^2} \cdot (\vec{R}_1 \times \vec{R}_2) \times \vec{R}_2 \\ & + \vec{R}_2/2 + \vec{P}_3 \end{aligned} \tag{D.3}$$

It can be easily proved that $\vec{O}$ does not depend on the selection of $\vec{P}_r$. Since at least 3 different positions are nec-

essary, we rotate the camera head around the scope cylinder by N different angles. We apply a RANSAC algorithm to estimate $\vec{O}$ using random positions, and select $\vec{O}^*$ which corresponds to the smallest variance as the center of the circle. The similar RANSAC algorithm is also used to compute θ.

Figure 13 shows the estimated rotation angle using RANSAC algorithm for two different endoscopes. The red curves are output angles from different RANSAC iterations, the black curve is the average angle. We can see the variance of the estimation is very small (less than 0.2 degree).

References

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256.

Clarkson, M. J., Rueckert, D., King, A. P., Edwards, P. J., Hill, D. L. G., & Hawkes, D. J. (1999). Registration of video images to tomographic images by optimising mutual information using texture mapping. In *LNCS: Vol. 1679. Proceedings of the second international conference on medical image computing and computer-assisted intervention (MICCAI'99)* (pp. 579–588). Berlin: Springer.

Courteille, F., Crouzil, A., Durou, J.-D., & Gurdjos, P. (2004). Towards shape from shading under realistic photographic conditions. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, *2*, 277–280.

Dey, D., Gobbi, D. G., Slomka, P. J., Surry, K. J. M., & Peters, T. M. (2002). Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas. *IEEE Transactions on Medical Imaging*, *21*(1), 23–30.

Durou, J.-D., Falcone, M., & Sagona, M. (2008). Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, *109*(1), 22–43.

Forster, C. H. Q., & Tozzi, C. L. (2000). Toward 3d reconstruction of endoscope images using shape from shading. In *Proceedings of the 13th Brazilian symposium on computer graphics and image processing (SIBGRAPHI'00)* (pp. 90–96).

Frankot, R. T., & Chellappa, R. (1988). A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *10*(4), 439–451.

Fuchs, H., Livingston, M. A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J. R., Rademacher, P., Drake, S. H., & Meyer, A. A. (1998). Augmented reality visualization for laparoscopic surgery. In *LNCS: Vol. 1496. Proceedings of the first international conference on medical image computing and computer-assisted intervention (MICCAI'98)* (pp. 934–943). Berlin: Springer.

Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd edn.). Cambridge: Cambridge University Press.

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

Horn, B. K. P., & Brooks, M. J. (1986). The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing (CVGIP'86)*, *33*(2), 174–208.

Horn, B. K. P., & Brooks, M. J. (1989). *Shape from shading*. Cambridge: MIT.

Horn, B. K. P., & Sjoberg, R. W. (1979). Calculating the reflectance map. *Applied Optics*, *18*(11), 1770–1779.

Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, *17*(1-3), 141–184.

Iyengar, A. K. S., Sugimoto, H., Smith, D. B., & Sacks, M. S. (2001). Dynamic in vitro quantification of bioprosthetic heart valve leaflet motion using structured light projection. *Annals of Biomedical Engineering*, *29*(11), 963–973.

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

Kozera, R. (1991). Existence and uniqueness in photometric stereo. *Applied Maths and Computation*, *44*(1), 1–103.

Kozera, R. (1992). On shape recovery from two shading patterns. *International Journal of Pattern Recognition and Artificial Intelligence*, *6*(4), 673–698.

Kozera, R. (1998). An overview of the shape from shading problem. *Machine Graphics and Vision*, *7*(1), 291–312.

Kozera, R., & Noakes, L. (2004). Noise reduction in photometric stereo with non-distant light sources. *Proceedings of the International Conference on Computer Vision and Graphics (ICCVG'04)*, *32*, 103–110.

Leclerc, Y. G., & Bobick, A. F. (1991). The direct computation of height from shading. In *Proceedings of IEEE computer society conference on computer vision and pattern recognition (CVPR'91)* (pp. 552–558).

Lee, K. M., & Kuo, C.-C. J. (1994). Shape from shading with perspective projection. *Computer Vision, Graphics, and Image Processing: 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 and Image Understanding*, *67*(2), 143–160.

Litvinov, A., & Schechner, Y. Y. (2005). Addressing radiometric nonidealities: a unified framework. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 52–59.

Mourgues, F., Devernay, F., Malandain, G., & Coste-Manière, E. (2001). 3d reconstruction of the operating field for image overlay in 3d-endoscopic surgery. In: *Proceedings of the IEEE and ACM international symposium on augmented reality (ISAR'01)* (pp. 191–192).

Okatani, T., & Deguchi, K. (1997). Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer Vision and Image Understanding*, *66*(2), 119–131.

Penna, M. A. (1989). 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.

Poelman, C. J., & Kanade, T. (1997). A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(3), 206–218.

Pollefeys, M., Koch, R., & Gool, L. V. (1999). Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, *32*(1), 7–25.

Prados, E., & Faugeras, O. (2003). Perspective shape from shading and viscosity solution. *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, *2*, 826–831.

Prados, E., & Faugeras, O. (2005). Shape from shading: a well-posed problem? *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 870–877.

Samaras, D., & Metaxas, D. N. (1999). Coupled lighting direction and shape estimation from single images. *Proceedings of the International Conference on Computer Vision (ICCV'99)*, *2*, 868–874.

Seshamani, S., Lau, W., & Hager, G. (2006). Real-time endoscopic mosaicking. In *LNCS: Vol. 4190. Proceedings of the Ninth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)* (pp. 355–363). Berlin: Springer.

Stoyanov, D., Darzi, A., & Yang, G. Z. (2005). A practical approach towards accurate dense 3d depth recovery for robotic laparoscopic surgery. *Computer Aided Surgery*, *10*(4), 199–208.

Tankus, A., Sochen, N., & Yeshurun, Y. (2003). A new perspective [on] shape-from-shading. *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, *2*, 862–869.

Tankus, A., Sochen, N., & Yeshurun, Y. (2004). Reconstruction of medical images by perspective shape-from-shading. *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*, *3*, 778–781.

Tankus, A., Sochen, N., & Yeshurun, Y. (2005). Shape-from-shading under perspective projection. *International Journal of Computer Vision*, *63*(1), 21–43.

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, *3*, 323–344.

Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, *19*(1), 139–144.

Yamaguchi, T., Nakamoto, M., Sato, Y., Konishi, K., Hashizume, M., Sugano, N., Yoshikawa, H., & Tamura, S. (2004). Development of a camera model and calibration procedure for oblique-viewing endoscopes. *Computer Aided Surgery*, *9*(5), 203–214.

Zhang, R., Tsai, P. S., Cryer, J. E., & Shah, M. (1999). Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(8), 690–706.

Zhang, Z. (1998). *A flexible new technique for camera calibration* (Technical Report MSR-TR-98-71). Microsoft Research.