

Estimating the Unknown Poses of a Reference Plane for Specular Shape Recovery

Miaomiao Liu, Kwan-Yee K. Wong
The University of Hong Kong,
{mmliu, kykwong}@cs.hku.hk

Abstract

This paper addresses the problem of recovering the unknown poses of a moving reference plane for specular shape recovery. Given the initial pose of the reference plane with respect to the camera, a closed form solution is derived to recover its subsequent poses directly from its reflections on the specular surface observed in the image sequence. With the estimated poses of the reference plane, the specular surface can then be easily recovered using any existing ray triangulation method. The proposed method greatly simplifies the calibration problem in specular shape recovery. Experimental results on both synthetic and real data are presented, which demonstrate the effectiveness of the proposed method.

1. Introduction

The problem of shape recovery for diffuse surfaces has been well studied, and many sophisticated methods [21, 23, 20, 22] have been developed over the past few decades. Unlike a diffuse surface whose appearance is viewpoint independent, a specular surface does not have a unique appearance of its own but instead reflects its surrounding environment. Based on this special property, many researchers tried to recover the shape of a specular object by exploring the relationship between its surface geometry and its surrounding environment [2, 24, 12, 8]. Methods for specular shape recovery usually introduce motion to the surrounding environment and observe the changes in the reflections produced on the surface. Based on the assumptions made on the environment, existing state-of-the-art methods can be broadly classified into two approaches, namely shape from specular flow (SFSF) methods [13, 1, 7, 14] and shape from specular correspondences (SFSC) methods [18, 4, 5, 15, 16, 11]. SFSF methods often assume an unknown distant environment under a known continuous motion. These methods often suffer from problems in tracking dense specular flows and solving partial differential equations (PDE). When the

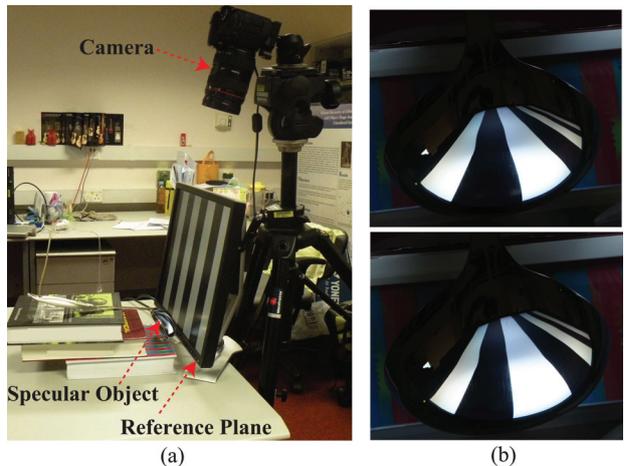


Figure 1. General setup for specular shape recovery. (a) The camera and the reference plane are put side-by-side to produce a good view of the reflection of the plane on a specular spoon. (b) Reflections produced on the spoon surface under two different poses of the reference plane.

surrounding environment is known and close to the object, SFSC methods can be derived. Like SFSF methods, SFSC methods also assume the motion (which may be discrete though) of the environment being known a priori in order to infer surface shape from the observed reflections. Commonly, a reference plane with a known pattern is used as the known environment in SFSC methods, and it is placed in at least two different poses for recovering the specular surface. In order to produce a good view of its reflection on the specular surface, the reference plane is often placed side-by-side with the camera (see Fig. 1). This results in the camera not being able to see the reference plane directly, making the calibration of the setup non-trivial. Traditional methods calibrate the poses of the reference plane by introducing an auxiliary plane into the field of view of the camera, and calibrating its pose with respect to the camera. An auxiliary camera is then used to take an image of both the reference plane and the auxiliary plane to recover their relative poses.

The pose of the unseen reference plane with respect to the camera then follows. In [19], Sturm and Bonfort used a planar mirror to allow the camera seeing the reference plane through reflection, and each pose of the reference plane can be obtained by placing the auxiliary mirror in at least three different positions. Note that in practice, it requires placing the reference plane in a large number of distinct poses to recover a complete specular shape using ray triangulation, and calibrating all the poses of the reference plane requires a tedious work. How to recover the poses of the reference plane easily and automatically hence becomes an appealing problem in specular shape recovery.

The literature becomes comparatively sparse when it comes to automatic pose estimation of the reference plane in specular shape recovery. Liu et al. [10] proposed an automatic motion estimation method by constraining the motion of the reference plane to a pure translation, with the assumption that the initial pose of the plane is known a priori. Although they can achieve a simple closed form solution for the motion estimation problem, their method cannot handle the general motion case. In [3], Bonfort proposed a method to calibrate the poses of the two reference planes by observing their reflections on the specular surface. However, their method is quite sensitive to noise, making it not quite applicable in practice. In this paper, we propose a *novel* and *practical* solution for estimating the poses of a reference plane based on observing its reflections on a specular surface. Given the initial pose of the reference plane with respect to the camera, a closed form solution is derived to recover its subsequent poses directly from its reflections on the specular surface observed in the image sequence. Unlike [10], however, the reference plane is not constrained to undergo a pure translation. With the estimated poses of the reference plane, the specular surface can then be easily recovered using any existing ray triangulation method. The proposed method greatly simplifies the calibration problem in specular shape recovery.

The major contributions of this paper are

- A novel and practical method for automatically recovering the poses of a reference plane in specular shape recovery. In particular, the proposed method does not require the use of any auxiliary reference plane, camera nor mirror and can directly recover the poses from the observed reflections of the reference plane on the specular surface.
- A closed form solution for recovering the poses of a reference plane undergoing a general motion.

The rest of the paper is organized as follows. Section 2 describes the physical configuration of our proposed approach and introduces a motion constraint based on the observed reflections of a moving reference plane. Based on such a constraint, a closed form solution for recovering the

poses of the reference plane is derived in Section 3, together with an analysis on the degenerate cases. Section 4 describes implementation issues related to the extraction of the sparse reflection correspondences and recovery of the specular surface given the estimated poses of the reference plane. Experimental results on both synthetic and real data are presented in Section 5, followed by discussions and conclusions in Section 6.

2. Geometric Constraint

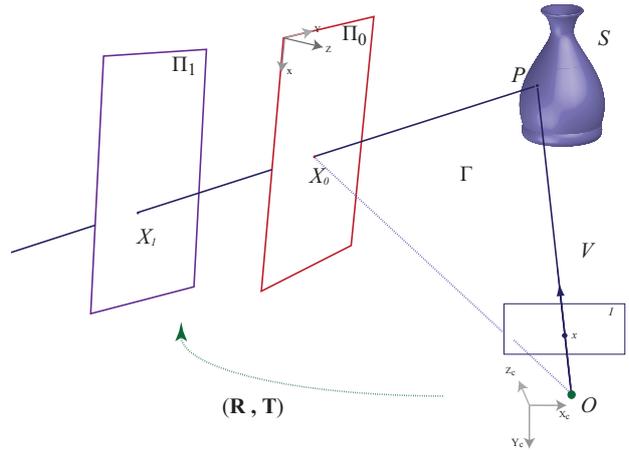


Figure 2. A pinhole camera centered at O is viewing a specular object S which reflects a nearby reference plane Π_0 to the image I . Π_1 denotes the plane at a new pose after undergoing an unknown rigid body motion. X_0 and X_1 are points on Π_0 and Π_1 respectively which are reflected at a point P on S to the same image point x on I . They are referred to as *reflection correspondences*. X_0 and X_1 define the incident ray. V denotes the visual ray. Both the incident ray and visual ray lie in the plane Γ called the *reflection plane*.

Consider a pinhole camera centered at O observing the reflection of a reference plane Π_0 produced on a specular object S (see Fig. 2). Let X_0 be a point on Π_0 which is reflected at a point P on S to a point x on the image I . Suppose now Π_0 undergoes an *unknown* rigid body motion, and let Π_1 denote the reference plane at its new pose. Let X_1 be a point on Π_1 such that it is collinear with X_0 and P . According to the law of reflection, X_1 will also be reflected at P to the same image point x on I . As shown in [18, 5], the 3D position of P can be obtained by intersecting the incident ray (defined by X_0 and X_1) with the visual ray (defined by O and x). In order to construct the incident ray, the 3D positions of both X_0 and X_1 are needed, and their accuracies would directly affect the accuracy of the estimated position of P . The local positions of X_0 and X_1 on the reference plane can be resolved relatively easily by encoding a grid pattern on the reference plane using Gray code and observing the intensity patterns at x . In or-

der to recover the 3D positions of X_0 and X_1 , the poses of Π_0 and Π_1 with respect to the camera must also be known. When the reference plane is directly visible in the image, the calibration of its pose is trivial. As mentioned previously, however, the camera usually cannot see the reference plane directly in the general setup for specular shape recovery. In [19], Sturm and Bonfort showed how the pose of a reference plane without a direct view in the image can be recovered with the help of an auxiliary planar mirror.

In this paper, Π_0 is assumed to have been calibrated against the camera initially. We would like to estimate the pose of Π_1 by deriving constraints using the observed reflections of both Π_0 and Π_1 on the unknown specular surface. Consider the reflection plane Γ defined by the incident ray and the visual ray at P (see Fig. 2). Let \mathbf{N} be a unit normal vector to Γ . Note that since both OX_0 and Ox lie on Γ , \mathbf{N} can be expressed as

$$\mathbf{N} = \frac{Ox \times OX_0}{\|Ox \times OX_0\|}. \quad (1)$$

Note that Ox is simply the visual ray constructed through the image point x . Given the pose of Π_0 , the 3D position of X_0 can be readily obtained from the Gray code pattern observed at x , and OX_0 can then be constructed. Hence, \mathbf{N} can be recovered even when point P and the surface normal at point P are not known a priori. By definition, X_0X_1 lies on Γ . It follows that

$$X_0X_1 \cdot \mathbf{N} = 0. \quad (2)$$

The above constraint will be used in the next section to derive a closed form solution for recovering the pose of Π_1 .

3. Recovering the Unknown Pose

Consider again the configuration shown in Fig. 2, and assume the world coordinate system coincides with the local coordinate system defined on Π_1 . Let \mathbf{X}_0 and \mathbf{X}_1 denote, respectively, the position vectors of X_0 and X_1 in the camera coordinate system, and \mathbf{N} denote the unit normal vector to the reflection plane in the camera coordinate system. The constraint in Eq. (2) can be rewritten as

$$(\mathbf{X}_1 - \mathbf{X}_0) \cdot \mathbf{N} = 0. \quad (3)$$

It follows directly from Eq. (1) that

$$\mathbf{X}_0 \cdot \mathbf{N} = 0. \quad (4)$$

Substituting Eq. (4) into Eq. (3) simplifies the constraint to

$$\mathbf{X}_1 \cdot \mathbf{N} = 0. \quad (5)$$

Let $\hat{\mathbf{X}} = (\hat{x}, \hat{y}, 0)^T$ denote the coordinates of X_1 in the world coordinate system, and \mathbf{R} and \mathbf{T} denote the rotation matrix and translation vector, respectively, relating the

world coordinate system to the camera coordinate system, such that

$$\mathbf{X}_1 = \mathbf{R}\hat{\mathbf{X}} + \mathbf{T}. \quad (6)$$

Substituting Eq. (6) into Eq. (5) gives

$$(\mathbf{r}_1\hat{x} + \mathbf{r}_2\hat{y} + \mathbf{T}) \cdot \mathbf{N} = 0, \quad (7)$$

where $\mathbf{r}_i = (r_{1i}, r_{2i}, r_{3i})^T$, $i \in \{1, 2\}$, denotes the i th column of \mathbf{R} . Given an image point x , the vector \mathbf{N} can be obtained using Eq. (1). The unknowns in Eq. (7) are therefore \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{T} , which have nine degrees of freedom in total if the constraints on the elements of \mathbf{R} are ignored. Since one image point will define one linear constraint in the form of Eq. (7), the pose problem can be solved when at least nine image points showing the reflections of the moving reference plane are available. Given the reflections observed at $m \geq 9$ image points x_i , the pose of Π_1 relative to the camera can be estimated by solving the following over-constrained system:

$$\begin{pmatrix} \hat{x}_1\mathbf{N}_1^T & \hat{y}_1\mathbf{N}_1^T & \mathbf{N}_1^T \\ \hat{x}_2\mathbf{N}_2^T & \hat{y}_2\mathbf{N}_2^T & \mathbf{N}_2^T \\ \vdots & \vdots & \vdots \\ \hat{x}_m\mathbf{N}_m^T & \hat{y}_m\mathbf{N}_m^T & \mathbf{N}_m^T \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{T} \end{pmatrix} = \mathbf{0}, \quad (8)$$

where $(\hat{x}_i, \hat{y}_i, 0)^T$ denotes the world coordinates of the reflection correspondence of x_i on Π_1 , and \mathbf{N}_i denotes the unit normal vector to the reflection plane of x_i . Eq. (8) can be succinctly expressed as

$$\begin{pmatrix} \bar{\mathbf{X}}_1^T \otimes \mathbf{N}_1^T \\ \bar{\mathbf{X}}_2^T \otimes \mathbf{N}_2^T \\ \vdots \\ \bar{\mathbf{X}}_m^T \otimes \mathbf{N}_m^T \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{T} \end{pmatrix} = \mathbf{0}, \quad (9)$$

where $\bar{\mathbf{X}}_i = (\hat{x}_i, \hat{y}_i, 1)^T$ and \otimes denotes the Kronecker product. A linear least squares solution can be obtained by performing a singular value decomposition (SVD) on the $m \times 9$ coefficient matrix in Eq. (8). Constraints on the rotation matrix \mathbf{R} can then be applied to ensure it is an orthogonal matrix. The pose of Π_1 can be further improved by a nonlinear optimization, in which the optimization parameters are the three parameters defining the rotation matrix and three parameters defining the translation vector. Note that there exists some degenerate cases which will be described below in Proposition 1.

Proposition 1 *If the matrix composed of the normals of the reflection planes is not of full rank, the rotation matrix \mathbf{R} and translation vector \mathbf{T} cannot be determined uniquely using the proposed theory.*

Proof. Consider the matrix

$$\tilde{\mathbf{N}}_{m \times 3} = (\mathbf{N}_1 \quad \mathbf{N}_2 \quad \cdots \quad \mathbf{N}_m)^T \quad (10)$$

composed of the normals of the reflection planes, and suppose $\text{rank}(\tilde{\mathbf{N}}) \leq 2$. Let

$$\tilde{\mathbf{X}}_{m \times 3} = (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_m)^T, \quad (11)$$

$$\mathbf{A}_{m \times 9} = (\tilde{\mathbf{X}}_1 \otimes \mathbf{N}_1, \tilde{\mathbf{X}}_2 \otimes \mathbf{N}_2, \dots, \tilde{\mathbf{X}}_m \otimes \mathbf{N}_m)^T. \quad (12)$$

Since \mathbf{A} is a sub-matrix of $\tilde{\mathbf{X}} \otimes \tilde{\mathbf{N}}$, it follows that $\text{rank}(\mathbf{A}) \leq \text{rank}(\tilde{\mathbf{X}} \otimes \tilde{\mathbf{N}})$. It has been proved in [9] that $\text{rank}(\tilde{\mathbf{X}} \otimes \tilde{\mathbf{N}}) = \text{rank}(\tilde{\mathbf{X}})\text{rank}(\tilde{\mathbf{N}})$. Since $\text{rank}(\tilde{\mathbf{X}}) \leq 3$ and $\text{rank}(\tilde{\mathbf{N}}) \leq 2$, $\text{rank}(\mathbf{A}) \leq \text{rank}(\tilde{\mathbf{X}} \otimes \tilde{\mathbf{N}}) = \text{rank}(\tilde{\mathbf{X}})\text{rank}(\tilde{\mathbf{N}}) \leq 6$.

Let $\mathbf{n}_i = (n_{1i} \dots n_{mi})^T$, $i \in \{1, 2, 3\}$, denote the i th column of $\tilde{\mathbf{N}}$. When $\text{rank}(\tilde{\mathbf{N}}) = 2$, \mathbf{n}_3 can be expressed as a linear combination of \mathbf{n}_1 and \mathbf{n}_2 , i.e.,

$$\mathbf{n}_3 = a\mathbf{n}_1 + b\mathbf{n}_2. \quad (13)$$

Eq. (8) can be re-written as $\mathbf{B}\mathbf{Q} = 0$, where

$$\mathbf{B} = \begin{pmatrix} \hat{x}_1 n_{11} & \hat{x}_1 n_{12} & \hat{y}_1 n_{11} & \hat{y}_1 n_{12} & n_{11} & n_{12} \\ \hat{x}_2 n_{21} & \hat{x}_2 n_{22} & \hat{y}_2 n_{21} & \hat{y}_2 n_{22} & n_{21} & n_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{x}_m n_{m1} & \hat{x}_m n_{m2} & \hat{y}_m n_{m1} & \hat{y}_m n_{m2} & n_{m1} & n_{m2} \end{pmatrix},$$

$$\mathbf{Q} = (Q_1, Q_2, Q_3, Q_4, Q_5, Q_6)^T,$$

$$Q_1 = r_{11} + ar_{31},$$

$$Q_2 = r_{21} + br_{31},$$

$$Q_3 = r_{12} + ar_{32},$$

$$Q_4 = r_{22} + br_{32},$$

$$Q_5 = t_x + at_z,$$

$$Q_6 = t_y + bt_z,$$

and $\mathbf{T} = (t_x, t_y, t_z)^T$. It is noted that $\text{rank}(\mathbf{B}) \leq 6$. When $\text{rank}(\mathbf{B}) = 6$, $\mathbf{Q} = \mathbf{0}$ and this provides six constraints on the unknowns (i.e., four on \mathbf{r}_1 and \mathbf{r}_2 , and two on \mathbf{T}). Three more constraints on \mathbf{r}_1 and \mathbf{r}_2 can be derived from the fact that \mathbf{R} is an orthogonal matrix. This allows the recovery of \mathbf{r}_1 and \mathbf{r}_2 , and \mathbf{r}_3 can be obtained readily from the cross product $\mathbf{r}_1 \times \mathbf{r}_2$. It is, however, not possible to recover the translation vector \mathbf{T} from the remaining two constraints. Similar proof goes to the case where $\text{rank}(\mathbf{B}) < 6$. When $\text{rank}(\tilde{\mathbf{N}}) = 1$, $\text{rank}(\mathbf{B}) \leq 3$ and there are not enough constraints for solving \mathbf{R} and \mathbf{T} .

Corollary 1 *If the specular object is composed of a single plane or a surface of revolution with its revolution axis passing through the camera center (e.g., a sphere), the matrix formed from the normals of the reflection planes will have a maximum rank of two and therefore the pose of the reference plane cannot be uniquely determined using the proposed method.*

4. Reflection Correspondences

As mentioned before, a simple way to establish reflection correspondence between an image point and a point on the reference plane is by encoding a grid pattern on the reference plane using Gray code and observing the intensity pattern at the image point. In practice, a computer monitor is commonly used as a reference plane by displaying a set of binary images constructed from Gray code patterns, once original and once inverted [17]. The total number of images taken under each pose of the reference plane is therefore twice the binary resolution in each direction. As stated in [19], the resolution of the codes and the width of the lowest order stripes must be chosen according to the shape of the specular object and the resolution of the camera. Too high resolution codes tend to be blurred out and become unusable, whereas too coarse ones lack in precision. Binary images taken under the same pose of the reference plane are combined into a single image from which the Gray code for each pixel can be retrieved. Generally, a patch of pixels in the image corresponds to a patch of points on the reference plane with the same Gray code. In order to obtain more accurate reflection correspondences, we first extract grid corners up to sub-pixel accuracies in each of the combined images. The 2D positions of their reflection correspondences can be obtained with a high accuracy. Note that, however, an image point which corresponds to a grid corner in one combined image will generally not correspond to a grid corner in another combined image. In this paper, the 2D position of the reflection correspondence of such a point is obtained by interpolation using the detected grid corners in the same combined image. The reflection correspondences of all the detected grid corners can then be used to estimate the pose of the reference plane. For specular shape recovery, dense reflection correspondences can be obtained in a similar manner by considering all the image points.

5. Experimental Results

5.1. Synthetic Experiment

The specular surface used in the synthetic experiment was composed of two spheres, with parametric equations

$$\begin{cases} x_1^2 + y_1^2 + (z_1 - 10)^2 = 9 \\ (x_2 + 1)^2 + (y_2 + 0.5)^2 + (z_2 - 11)^2 = 9 \end{cases} \quad (14)$$

The perspective camera used had a focal length of 24mm and an image resolution of 1600 × 1200, and was centered at the world origin. Fig. 3(a) shows the reference plane Π_0 at its initial pose as well as the reference plane Π_1 resulting from a rigid body motion of Π_0 . The average distance between Π_0 and Π_1 was 4.3943cm. The software *Pov-ray* was used to render the images capturing the reflections of the reference plane on the specular sur-

face. Four different grid resolutions, with square patches having a side of 0.4cm , 0.6cm , 0.8cm and 1.0cm respectively, were used in encoding 2D positions on the reference plane using Gray code. To simulate the effects of noise, a normally distributed noise with a standard deviation $\sigma \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ (in pixels) was added independently to each of the extracted grid corners in the images. 150 independent trials were carried out for each of the eleven noise levels and each of the four grid resolutions respectively. Fig. 4 shows three plots of pose estimation error (averaged over the 150 trials) against noise level for each of the grid resolutions. Fig. 4(a) shows the rotation angle error $r_ang_err = \text{rot_ang}(\mathbf{R}_e^T \mathbf{R}_g)$ where \mathbf{R}_e and \mathbf{R}_g are the estimated rotation matrix and the ground truth respectively, and $\text{rot_ang}(\mathbf{R})$ returns the rotation angle of the rotation matrix \mathbf{R} . Fig. 4(b) and (c) show the translation angle error $t_ang_err = \arccos\left(\frac{\mathbf{T}_e \cdot \mathbf{T}_g}{\|\mathbf{T}_e\| \|\mathbf{T}_g\|}\right)$ and the translation scale error $t_sca_err = \frac{\|\mathbf{T}_e - \mathbf{T}_g\|^2}{\|\mathbf{T}_g\|^2}$ respectively, where \mathbf{T}_e and \mathbf{T}_g are the estimated translation vector and the ground truth respectively. Note that small errors still existed in the estimated poses even when the noise level was zero. This can be explained by the inaccuracies in the 2D positions of the reflection correspondences on the reference plane caused by the limited resolution of the Gray code. As expected, the pose estimation errors decreased with the size of the square patches of the encoding grid pattern (i.e., a higher grid resolution and smaller errors in the 2D positions of the reflection correspondences). Using a grid with square patches having a side of 0.4cm and under a noise level of $\sigma = 1.0$ pixel, the average pose estimation errors are $r_ang_err = 0.4574^\circ$, $t_ang_err = 0.3081^\circ$, and $t_sca_err = 0.55\%$. Fig. 3(b) shows a histogram of distance errors between reflection correspondences on Π_1 and Π_e (i.e., the estimation of Π_1). It can be seen that over 82.0594% of the correspondences have a distance error less than 0.08cm , which demonstrates the accuracy of the estimated pose.

5.2. Real Experiment

A specular spoon and a polished metallic sphere were used as the specular objects in the real experiment (see Fig. 5(a)). A 17-inch LCD monitor displaying a set of Gray code pattern images together with their negative images was used as the reference plane. The encoding square patches had a side of 0.59cm , which was chosen to maximize the Gray code resolution while ensuring its distinguishability in the images. A Canon 40D digital camera equipped with a 24mm lens was used to capture the reflections of the reference plane on the spoon and the sphere. The image resolution used was 3888×2592 . In order to obtain the ground truth data for evaluation, another Nikon D3100 digital camera was used to calibrate the poses of

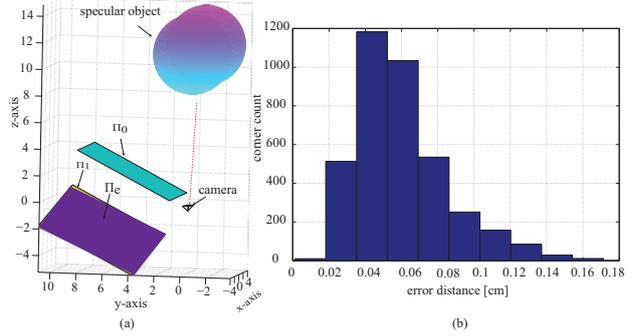


Figure 3. Synthetic experiment. (a) Configuration of the synthetic experiment. Π_0 and Π_1 denote the poses of the reference plane before and after the rigid body motion respectively. Π_e denotes the estimated pose of the reference plane after the rigid body motion. (b) Histogram of distance errors between reflection correspondences on Π_1 and Π_e .

the reference plane. The intrinsic parameters of both cameras were calibrated using [6]. Fig. 5(b) shows an image of the spoon and the sphere reflecting the finest Gray code pattern. The pose of the reference plane after its rigid body motion was estimated using the reflection correspondences of the extracted grid corners in the images. Using the same error measures defined in Section 5.1, the errors in the estimated pose were $r_ang_err = 0.8398^\circ$, $t_ang_err = 0.2863^\circ$, and $t_sca_err = 0.6625\%$ respectively. The specular spoon and sphere are reconstructed by using ray triangulation without making use of the spherical information for the sphere. In order to validate the pose estimation error on shape recovery, we fit a sphere to the recovered sphere points. Over 87.6997% of the 1252 recovered points were less than 0.05cm away from the fitted sphere and 74.7604% less than 0.02cm . The approximate radius of the reconstructed sphere was 3.15cm , resulting in a 1% error compared with the ground truth. Fig. 5(c) and (d) show the normal maps of the spoon and the sphere, respectively, recovered using the estimated pose of the reference plane. (e) and (f) show the normal maps of the spoon and the sphere using the pose of the reference plane calibrated by the auxiliary camera. Fig. 5(g) shows the reconstructed surfaces of both the spoon and the sphere, together with the reference plane at its initial pose, its estimated new pose by the proposed method, and its pose calibrated by the auxiliary camera.

6. Conclusions and Discussions

This paper introduces a novel and practical method for estimating the unknown poses of a reference plane in specular shape recovery. Unlike existing methods, it does not require the use of any auxiliary reference plane, camera or mirror, and allows the reference plane to undergo unrestricted general motion. Given the initial pose of the refer-

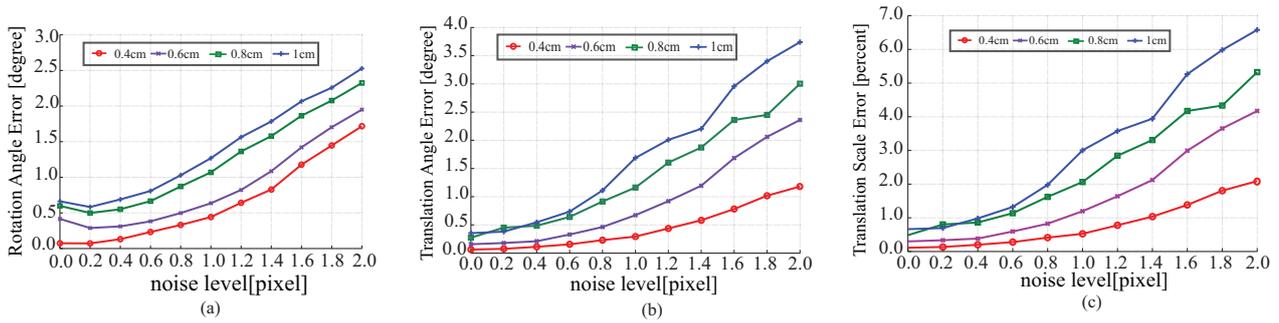


Figure 4. Three plots of pose estimation error (averaged over 150 independent trials) against noise level for four different encoding grid resolutions. (a) Relative rotation angle between the estimated rotation matrix and the ground truth. (b) Angle between the estimated translation vector and the ground truth. (c) Ratio of the squared norm of the difference between the estimated translation vector and the ground truth to the squared norm of the ground truth.

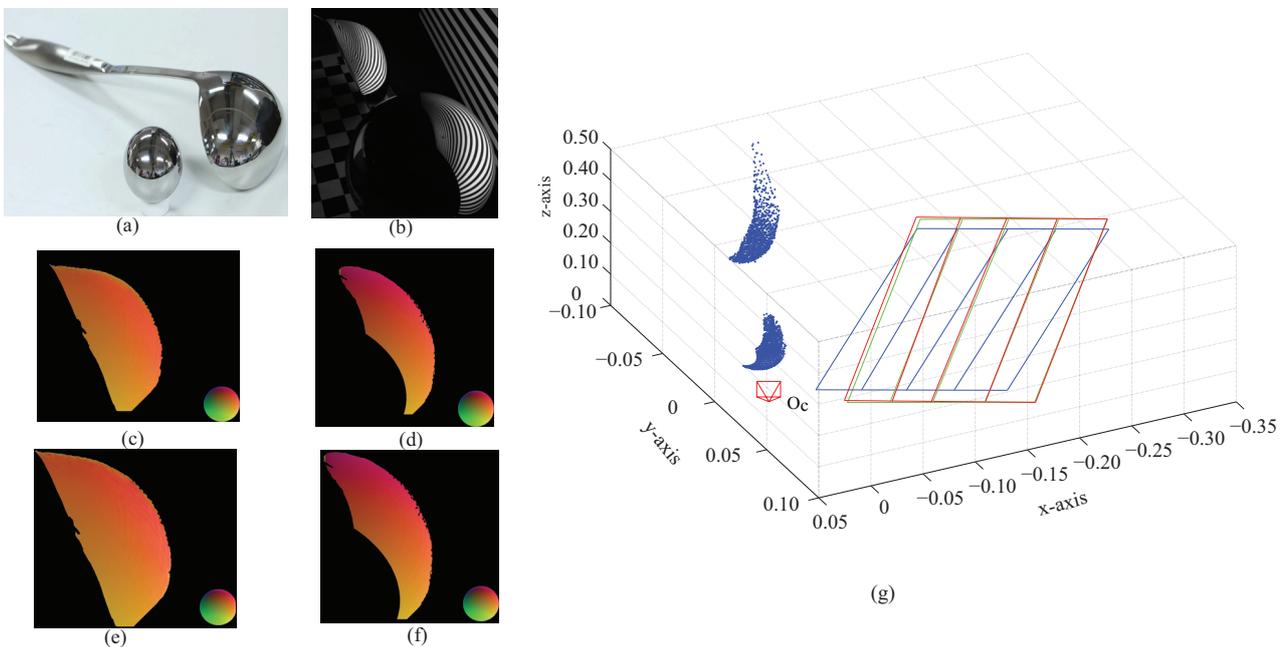


Figure 5. Real experiment. (a) A specular spoon and a polished metallic sphere. (b) An image of the spoon and the sphere reflecting the finest Gray code pattern. (c)-(d) Normal maps recovered for the spoon and sphere by using the estimated pose. (e)-(f) Normal maps constructed by using the pose calibrated by the auxiliary camera. (g) rendered setting. The blue rectangle denotes the reference plane in its initial pose. The green rectangle denotes the reference plane at its estimated new pose (our method). The red rectangle denotes the reference plane at the pose calibrated by the auxiliary camera. O_c indicates the camera position.

ence plane with respect to the camera, the proposed method gives a closed form solution for recovering its subsequent poses directly from its reflections on the unknown specular surface observed in the image sequence. It greatly simplifies the calibration problem in specular shape recovery. With the estimated poses of the reference plane, the specular surface can then be easily recovered using any existing ray triangulation method. Experimental results demonstrate that the poses of the reference plane can be estimated with high accuracy.

It is noted that the accuracy of the pose estimation will

greatly depend on the accuracy in locating the 2D positions of reflection correspondences on the reference plane. However, most of the practical encoding strategies, if not all, cannot achieve a point-to-point correspondence between the image and the reference plane. In our experiments, in order to improve the accuracy, we considered using interpolation to achieve approximate point-to-point correspondences at grid corners. As for the future work, we are now trying to solve the reconstruction problem of a complete specular object.

References

- [1] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *IEEE International Conference on Computer Vision*, pages 1–8, Brazil, Oct. 2007. 1
- [2] A. Blake and G. Brelstaff. Geometry from specularities. In *IEEE International Conference on Computer Vision*, pages 394–403, 1988. 1
- [3] T. Bonfort. *Reconstruction de surfaces réfléchissantes partir d'images*. PhD thesis, Institut National Polytechnique de Grenoble, 2006. 2
- [4] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *IEEE International Conference on Computer Vision*, pages 691–696, Nice, France, Oct. 2003. 1
- [5] T. Bonfort, P. Sturm, and P. Gargallo. General specular surface triangulation. In *Asian Conference on Computer Vision*, volume II, pages 872–881, Hyderabad, India, Jan. 2006. 1, 2
- [6] J.-Y. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/. 5
- [7] G. D. Canas, Y. Vasilyev, Y. Adato, T. Zickler, S. Gortler, and O. Ben-Shahar. A linear formulation of shape from specular flow. In *IEEE International Conference on Computer Vision*, pages 191–198, Kyoto, Sep. 2009. 1
- [8] R. W. Fleming, A. Torralba, and E. H. Adelson. Specular reflections and the perception of shape. *Journal of Vision*, 4(9):798–820, 2003. 1
- [9] R. A. Horn. *Topics in Matrix Analysis*. Cambridge University Press New York, NY, USA, 1986. 4
- [10] M. Liu, K.-Y. Wong, Z. Dai, and Z. Chen. Specular surface recovery from reflections of a planar pattern undergoing an unknown pure translation. In *Asian Conference on Computer Vision*, volume 2, pages 647–658, Queenstown, New Zealand, Nov. 2010. 2
- [11] D. Nehab, T. Weyrich, and S. Rusinkiewicz. Dense 3d reconstruction from specular consistency. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, Alaska, Jun. 2008. 1
- [12] M. Oren and S. K. Nayar. A theory of specular surface geometry. *International Journal of Computer Vision*, 24(2):105–124, 1997. 1
- [13] S. Roth and M. J. Black. Specular flow and the recovery of surface structure. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1869–1876, New York, Jun. 2006. 1
- [14] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal. Specular surface reconstruction from sparse reflection correspondences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1245–1252, San Francisco, Jun. 2010. 1
- [15] S. Savarese and P. Perona. Local analysis for 3d reconstruction of specular surfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 738–745, Hawaii, Dec. 2001. 1
- [16] S. Savarese and P. Perona. Local analysis for 3d reconstruction of specular surfaces - part ii. In *European Conference on Computer Vision*, pages 759–774, London, UK, May 2002. 1
- [17] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 195–202, Madison, WI, USA, Jun. 2003. 4
- [18] E. Steger and K. N. Kutulakos. A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*, 76(1):13–29, 2008. 1, 2
- [19] P. Sturm and T. Bonfort. How to compute the pose of an object without a direct view. In *Asian Conference on Computer Vision*, pages 21–31, Jan. 2006. 2, 3, 4
- [20] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *European Conference on Computer Vision*, pages 709–720, 1996. 1
- [21] C. Tomasi. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9:137–154, 1992. 1
- [22] B. Triggs. Factorization methods for projective structure and motion. In *IEEE Conference on Computer Vision*, pages 845–851, 1996. 1
- [23] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 1
- [24] A. Zisserman, P. Giblin, and A. Blake. The information available to a moving observer from specularities. *Image Vision Computing*, 7(1):38–42, 1989. 1