Mirror Surface Reconstruction under an Uncalibrated Camera

Kai Han[1], Kwan-Yee K. Wong[1], Dirk Schnieders[1], Miaomiao Liu[2]
[1]The University of Hong Kong,* Hong Kong,
[2]NICTA[†] and CECS, ANU, Canberra

[1]{khan, kykwong, sdirk}@cs.hku.hk, [2]miaomiao.liu@nicta.com.au

Abstract

This paper addresses the problem of mirror surface reconstruction, and a solution based on observing the reflections of a moving reference plane on the mirror surface is proposed. Unlike previous approaches which require tedious work to calibrate the camera, our method can recover both the camera intrinsics and extrinsics together with the mirror surface from reflections of the reference plane under at least three unknown distinct poses. Our previous work has demonstrated that 3D poses of the reference plane can be registered in a common coordinate system using reflection correspondences established across images. This leads to a bunch of registered 3D lines formed from the reflection correspondences. Given these lines, we first derive an analytical solution to recover the camera projection matrix through estimating the line projection matrix. *We then optimize the camera projection matrix by minimizing reprojection errors computed based on a cross-ratio formulation. The mirror surface is finally reconstructed based on the optimized cross-ratio constraint. Experimental results on both synthetic and real data are presented, which demonstrate the feasibility and accuracy of our method.*

1. Introduction

3D reconstruction of *diffuse* surfaces has enjoyed tremendous success. Diffuse surfaces reflect light from a single incident ray to many rays in all directions, resulting in a constant appearance regardless of the observer's viewpoint. Methods for diffuse surface reconstruction can therefore rely on the appearance of the object.

This paper considers mirror surfaces, which exhibit *specular* reflections and whose appearances are a reflection of the surrounding environment. Under specular reflection, an
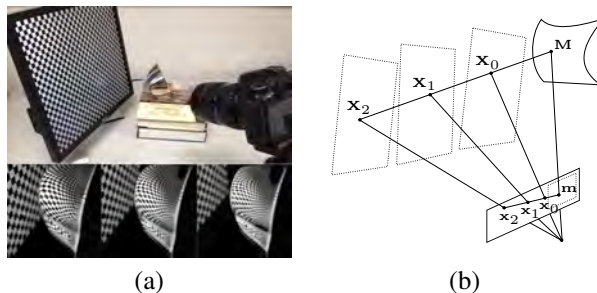
(a)   (b)

Figure 1: (a) A stationary uncalibrated camera observing the reflections of a reference plane undergoing an unknown motion. (b) Surface points can be recovered using the cross-ratio between a surface point $\mathbf{M}$ and its reflection correspondences $\{\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2\}$.

incoming ray is reflected to a single outgoing ray. This special characteristic leads to different appearances of the mirror surface under different viewpoints, and renders diffuse surface reconstruction methods useless. Meanwhile, there exist many objects with a mirror surface in the man-made environment. The study of mirror surface reconstruction is therefore an important problem in computer vision.

In this paper, we assume the mirror surface reflect a light ray only once, and tackle the mirror surface reconstruction problem by adopting a common approach of introducing motion to the environment. Unlike previous methods which require a fully calibrated camera and known motion, we propose a novel solution based on observing the reflections of a reference plane undergoing an unknown motion with a stationary uncalibrated camera (see Fig. 1(a)).

2D correspondences between the image and the reference plane are established by displaying a sweeping line on the plane (we use a computer screen as the reference plane in practice). The relative poses of the reference plane are then estimated [16], and rays piercing the plane under different poses are determined for each image point on the mirror surface.

Given the rays and their corresponding image points, we

first derive an analytical solution to estimate the camera projection matrix through estimating the *line projection matrix*. Such a line projection matrix can then be transformed to a corresponding camera (point) projection matrix [11]. To make our solution more robust to noise, we use this closed-form solution as an initialization and optimize the camera projection matrix by minimizing reprojection errors computed based on a cross-ratio formulation for the mirror surface (see Fig. 1(b)). The mirror surface is finally reconstructed based on the optimized cross-ratio constraint.

The key contributions of this work are

- To the best of our knowledge, the first mirror surface reconstruction solution under an unknown motion and an uncalibrated camera.

- A closed-form (linear) solution for estimating the camera projection matrix from reflection correspondences.

- A cross-ratio based nonlinear formulation that allows a robust estimation of the camera projection matrix together with the mirror surface.

2. Related work

Great efforts have been devoted to the problem of mirror surface recovery [4, 12, 22]. Based on the assumed prior knowledge, shape recovery methods for mirror surfaces can be classified into those assuming an *unknown distant* environment and those assuming a *known nearby* environment.

Under an *unknown distant* environment, a set of methods referred to as shape from specular flow (SFSF) have been proposed. In [18], Oren and Nayar successfully recovered a 3D curve on the object surface by tracking the trajectory of the reflection of a light source on the mirror surface. However, it is difficult to track a complete trajectory since the reflected feature will be greatly distorted near the occluding boundary of an object. Roth and Black [23] introduced the concept of specular flow and derived its relation with the 3D shape of a mirror surface. Although they only recovered a surface with a parametric representation (e.g., sphere), their work provides a theoretical basis for the later methods. In [1, 2], Adato *et al.* showed that under far-field illumination and large object-environment distance, the observed specular flow can be related to surface shape through a pair of coupled nonlinear partial differential equations (PDEs). Vasilyev *et al.* [29] further suggested that it is possible to reconstruct a smooth surface from one specular flow by inducing integrability constraints on the surface normal field. In [9], Canas *et al.* reparameterized the nonlinear PDEs as linear equations that lead to a more manageable solution.

Although SFSF achieves a theoretical breakthrough in shape recovery of mirror surfaces, the essential issues in tracking dense specular flow and in solving PDEs still hinder their practical use. In [25], Sankaranarayanan *et al.*

developed an approach that uses sparse specular reflection correspondences instead of specular flow to recover a mirror surface linearly. Their proposed method is more practical than the traditional SFSF methods. Nevertheless, their method requires quite a number of specular reflection correspondences across different views, which are difficult to obtain due to the distorted reflections on the mirror surface.

Under a *known nearby* environment, a different set of methods for shape recovery of mirror surfaces can be derived. The majority of these methods are based on the smoothness assumption on the mirror surface. Under this assumption, one popular way is to formulate the surfaces into the problem of solving PDEs. In [26, 27], Savarese and Perona demonstrated that local surface geometry of a mirror surface can be determined by analyzing the local differential properties of the reflections of two calibrated lines. Following the same fashion, Rozenfeld *et al.* [24] explored the 1D homography relationship between the calibrated lines and the reflections using sparse correspondences. Depth and first order local shape are estimated by minimizing a statistically correct measure, and a dense 3D surface is then constructed by performing a constrained interpolation. In [15], Liu *et al.* proved that a smooth mirror surface can be determined up to a two-fold ambiguity from just one reflection view of a calibrated reference plane.

Another way to formulate the mirror surfaces is by employing normal consistency property to refine visual hull and/or integrate normal field. In [6], Bonfort and Sturm introduced a voxel carving method to reconstruct a mirror surface using a normal consistency criterion derived from the reflections of some calibrated reference planes. In order to get a better view for shape recovery, they further proposed that the camera may not need to face the reference plane, and the shape can be well recovered by using a mirror to calibrate the poses of the reference plane [7, 28]. In [17], Nehab *et al.* formulated the shape recovery as an image matching problem by minimizing a cost function based on normal consistency. In [30], Weinmann *et at.* employed a turntable setup with multiple cameras and displays, which enables the calculation of the normal field for each reflection view. The 3D surface is then estimated by a robust multi-view normal field integration technique. In [3], Balzer *et al.* deployed a room-sized cube consisting of six walls that encode/decode specular correspondences based on a phase shift method. The surface is then recovered by integration of normal fields.

Besides, instead of directly formulating the surfaces, another direction is to reconstruct the individual light paths based on the law of reflection. Kutulakos and Steger [14] showed that a point on a mirror surface can be recovered if the positions of two reference points are known in space and reflected to the same image point in a single view, or the positions of two reference points are known and are re-

flected by the same surface point to two different views. In [16], Liu *et al.* established reflection correspondences on the reference plane under three distinct poses, and derived a method for recovering the relative poses of the plane. Given the camera intrinsics, the camera pose can also be solved and the surface can be recovered by ray triangulation.

Note that calibration plays an important role in all the above methods that assume a known nearby environment. In this paper, we neither make assumption on the smoothness of the mirror surface, nor require the calibration of the camera. Our proposed approach can automatically calibrate the setup as well as reconstruct the mirror surface using the observed reflections of the reference plane.

Cross-ratio constraint has been used to estimate mirror position and camera pose for axial non-central catadioptric systems [19], and produce more point correspondences in the context of 3D reconstruction [20]. Our method also relies on a cross-ratio constraint to optimize the camera projection matrix as well as recovering the mirror surface. Unlike existing methods handling the case when both the mirror and the reference plane are simultaneously visible to the camera (e.g., [21]), we tackle a more challenging scenario where only the mirror surface is visible.
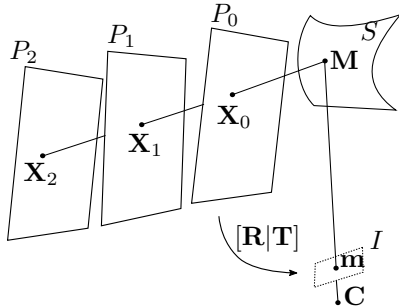
3. Acquisition Setup

Figure 2: Setup used for mirror surface reconstruction. Refer to Section 3 for notations and definitions.

Fig. 2 shows the setup used for mirror surface reconstruction. Consider a pinhole camera centered at $\mathbf{C}$ observing the reflections of a moving reference plane on a mirror surface S. Let $\mathbf{X}_0$ be a point on the plane at its initial pose, denoted by P_0, which is reflected by a point $\mathbf{M}$ on S to a point $\mathbf{m}$ on the image plane I. Suppose the reference plane undergoes an unknown rigid body motion, and let P_1 and P_2 denote the plane at its two new poses. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be points on P_1 and P_2, respectively, which are both reflected by $\mathbf{M}$ on S to the same image point $\mathbf{m}$ on I. $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$ are referred to as *reflection correspondences* of the image point $\mathbf{m}$.

4. A Closed-form Solution

In this section, we first briefly review Plücker coordinates and the line projection matrix. We then derive a linear method for obtaining a closed-form solution to the line projection matrix of a camera from reflection correspondences of the image points.

4.1. Plücker Coordinates

A 3D line can be described by a skew-symmetric Plücker matrix $\mathbf{L} = \mathbf{Q}\mathbf{P}^{\mathrm{T}} - \mathbf{P}\mathbf{Q}^{\mathrm{T}} =$

$$\begin{bmatrix} 0 & q_1p_2 - q_2p_1 & q_1p_3 - q_3p_1 & q_1p_4 - q_4p_1 \\ q_2p_1 - q_1p_2 & 0 & q_2p_3 - q_3p_2 & q_2p_4 - q_4p_2 \\ q_3p_1 - q_1p_3 & q_3p_2 - q_2p_3 & 0 & q_3p_4 - q_4p_3 \\ q_4p_1 - q_1p_4 & q_4p_2 - q_2p_4 & q_4p_3 - q_3p_4 & 0 \end{bmatrix},$$

where $\mathbf{P} = [p_1\ p_2\ p_3\ p_4]^{\mathrm{T}}$ and $\mathbf{Q} = [q_1\ q_2\ q_3\ q_4]^{\mathrm{T}}$ are the homogeneous representations of two distinct 3D points. Since $\mathbf{L}$ is skew-symmetric, it can be represented simply by a Plücker vector $\mathcal{L}$ consisting of its 6 distinct non-zero elements

$$\mathcal{L} = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{bmatrix} = \begin{bmatrix} q_1p_2 - q_2p_1 \\ q_1p_3 - q_3p_1 \\ q_1p_4 - q_4p_1 \\ q_2p_3 - q_3p_2 \\ q_3p_4 - q_4p_3 \\ q_4p_2 - q_2p_4 \end{bmatrix}. \tag{1}$$

Dually, a matrix $\bar{\mathbf{L}}$ can be constructed from two distinct planes with homogeneous representations $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ as $\bar{\mathbf{L}} = \hat{\mathbf{Q}}\hat{\mathbf{P}}^{\mathrm{T}} - \hat{\mathbf{P}}\hat{\mathbf{Q}}^{\mathrm{T}}$. The *dual Plücker vector* can be constructed directly from $\bar{\mathbf{L}}$ or by rearranging the elements of $\mathcal{L}$ as

$$\bar{\mathcal{L}} = [l_5\ l_6\ l_4\ l_3\ l_1\ l_2]^{\mathrm{T}}. \tag{2}$$

Let $\mathbf{A} = [a_1\ a_2\ a_3]^{\mathrm{T}}$ and $\mathbf{B} = [b_1\ b_2\ b_3]^{\mathrm{T}}$ be two distinct 3D points in Cartesian coordinates. Geometrically, the line defined by these points can be represented by a direction vector $\boldsymbol{\omega} = (\mathbf{A} - \mathbf{B}) = [l_3, -l_6, l_5]^{\mathrm{T}}$ and a moment vector $\boldsymbol{\nu} = (\mathbf{A} \times \mathbf{B}) = [l_4, -l_2, l_1]^{\mathrm{T}}$, which define the line up to a scalar factor.

Two 3D lines $\mathcal{L}$ and $\mathcal{L}'$ can either be skew or coplanar. The geometric requirement for the latter case is that the dot product between the first direction vector and the second moment vector should equal the negative of the dot product between the second direction vector and the first moment vector. Let the two lines have direction vectors $\boldsymbol{\omega}, \boldsymbol{\omega}'$ and moment vectors $\boldsymbol{\nu}, \boldsymbol{\nu}'$, respectively. They are coplanar (i.e., either coincident or intersect) if and only if

$$\boldsymbol{\omega} \cdot \boldsymbol{\nu}' + \boldsymbol{\nu} \cdot \boldsymbol{\omega}' = 0 \Leftrightarrow \mathcal{L} \cdot \bar{\mathcal{L}}' = 0. \tag{3}$$

Note that a Plücker vector is not any arbitrary 6-vector. A valid Plücker vector must always intersect itself, i.e.,

$$\mathcal{L} \cdot \bar{\mathcal{L}} = 0 \Leftrightarrow det(\mathbf{L}) = 0. \tag{4}$$

4.2. Line Projection Matrix

Using homogeneous coordinates, a linear mapping can be defined for mapping a point $\mathbf{X}$ in 3D space to a point $\mathbf{x}$ in a 2D image, i.e.,

$$\mathbf{x} = \mathbf{P}\mathbf{X}, \tag{5}$$

where $\mathbf{P}$ is a 3×4 matrix known as the camera (point) projection matrix. Similarly, using Plücker coordinates, a linear mapping can be defined for mapping a line $\mathcal{L}$ in 3D space to a line $\mathbf{l}$ (in homogeneous coordinates) in a 2D image, i.e.,

$$\mathbf{l} = \mathcal{P}\bar{\mathcal{L}}, \tag{6}$$

where $\mathcal{P}$ is a 3×6 matrix known as the *line projection matrix*. Note that each row $\mathbf{P}_i^{\mathrm{T}}$ ($i \in \{1,2,3\}$) of $\mathbf{P}$ represents a *plane* (in homogeneous coordinates) that intersects at the optical center. Dually, each row $\mathcal{P}_i^{\mathrm{T}}$ ($i \in \{1,2,3\}$) of $\mathcal{P}$ represents a *line* that intersects at the optical center (see Fig. 3). It follows that a valid line projection matrix must satisfy

$$\mathcal{P}_i \cdot \bar{\mathcal{P}}_j = 0 \ \forall i,j \in \{1,2,3\} \Leftrightarrow \mathcal{P}\bar{\mathcal{P}}^{\mathrm{T}} = \mathbf{0}_{3,3}, \tag{7}$$

where $\bar{\mathcal{P}} = [\bar{\mathcal{P}}_1 \ \bar{\mathcal{P}}_2 \ \bar{\mathcal{P}}_3]^{\mathrm{T}}$.
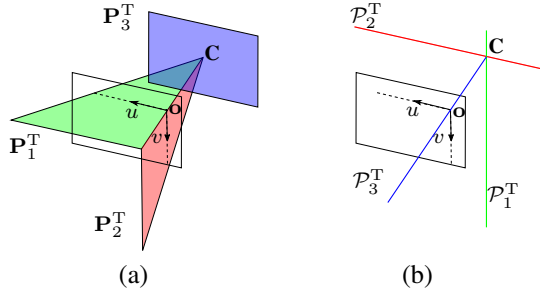

(a)   (b)

Figure 3: (a) Rows of a point projection matrix represent planes that intersect at the optical center $\mathbf{C}$ of the camera. (b) Dually, rows of a line projection matrix represent lines that intersect at the optical center.

4.3. Estimating the Line Projection Matrix

To estimate the line projection matrix of the camera, we first employ the method described in [16] to recover the relative poses of the reference plane under three distinct poses using reflection correspondences established across the images. We can then form a 3D Plücker line $\mathcal{L}$ from the reflection correspondences of each observed point $\mathbf{x}$ in the image. Note that, by construction, $\mathbf{x}$ must lie on the projection of $\mathcal{L}$, i.e.,

$$\mathbf{x}^{\mathrm{T}}\mathcal{P}\bar{\mathcal{L}} = 0. \tag{8}$$

Given a set of 3D space lines $\{\mathcal{L}_1, ..., \mathcal{L}_n\}$ constructed for a set of image points $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$, the constraint derived in (8) can be arranged into

$$\mathbf{A}\vec{\mathcal{P}} = \mathbf{0}, \tag{9}$$

where $\vec{\mathcal{P}} = [\mathcal{P}_1^{\mathrm{T}} \ \mathcal{P}_2^{\mathrm{T}} \ \mathcal{P}_3^{\mathrm{T}}]^{\mathrm{T}}$ and

$$\mathbf{A} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} \otimes \bar{\mathcal{L}}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \otimes \bar{\mathcal{L}}_n^{\mathrm{T}} \end{bmatrix}.^1 \tag{10}$$

The line projection matrix of the camera can then be estimated by solving

$$\underset{\vec{\mathcal{P}}}{\operatorname{argmin}} \|\mathbf{A}\vec{\mathcal{P}}\|^2 \tag{11}$$

subject to $\|\vec{\mathcal{P}}\| = 1$. The line projection matrix obtained thus can be transformed into a point projection matrix and vice versa. Note that, however, (11) minimizes only algebraic errors and does not enforce (7). The solution to (11) is therefore subject to numerical instability and not robust in the presence of noise.

Instead of solving (11), we can minimize the geometric distance from each image point to the projection of the corresponding 3D line. Let $\mathbf{l} = [a,b,c]^{\mathrm{T}} = \mathcal{P}\bar{\mathcal{L}}$ be the projection of the 3D line $\mathcal{L}$ corresponding to an image point $\mathbf{x} = [x_1, x_2, x_3]^{\mathrm{T}}$. $\mathcal{P}$ can be estimated by solving

$$\underset{\mathcal{P}}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{(\mathbf{x}_i^{\mathrm{T}}\mathcal{P}\bar{\mathcal{L}}_i)^2}{a_i^2 + b_i^2} \tag{12}$$

subject to $\|\mathcal{P}\| = 1$, where $\|\mathcal{P}\|$ is the Frobenius norm of $\mathcal{P}$. A straight-forward approach to enforce (7) is by incorporating it as a hard constraint in (12). However, experiments using a number of state-of-the-art optimization schemes show that such a solution often converges to local minima.

4.4. Enforcing Constraints

Given a proper camera projection matrix, the corresponding line projection matrix will automatically satisfy (7). However, given an improper 3×6 line projection matrix not satisfying (7), the corresponding camera projection matrix cannot be decomposed into one with proper intrinsic and extrinsic parameters. Based on this observation, we propose to enforce (7) by enforcing a proper decomposition of the camera projection matrix.

Consider a simplified scenario where the principal point (u_0, v_0) (which is often located at the image centre) is known. After translating the image origin to the principal point, the camera projection matrix can be expressed as

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \ \mathbf{T}] = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix},$$

and the corresponding line projection matrix can be expressed as

$$\mathcal{P} = \begin{bmatrix} f_y & 0 & 0 \\ 0 & f_x & 0 \\ 0 & 0 & f_x f_y \end{bmatrix} \mathcal{P}', \tag{13}$$

$^1\otimes$ stands for Kronecker product.

where

$$\mathcal{P}_i'^{\mathrm{T}} = \begin{bmatrix} \rho_{i1}' \\ \rho_{i2}' \\ \rho_{i3}' \\ \rho_{i4}' \\ \rho_{i5}' \\ \rho_{i6}' \end{bmatrix} = (-1)^{(i+1)} \begin{bmatrix} r_{j3}t_k - t_j r_{k3} \\ t_j r_{k2} - r_{j2}t_k \\ r_{j2}r_{k3} - r_{j3}r_{k2} \\ r_{j1}t_k - t_j r_{k1} \\ r_{j1}r_{k2} - r_{j2}r_{k1} \\ r_{j1}r_{k3} - r_{j3}r_{k1} \end{bmatrix}, \quad (14)$$

with $i \neq j \neq k \in \{1, 2, 3\}$ and $j < k$. (9) can then be rewritten as

$$\mathbf{A}\vec{\mathcal{P}} = \mathbf{A}\mathbf{D}\vec{\mathcal{P}'} = \mathbf{A}'\vec{\mathcal{P}'} = 0, \quad (15)$$

where $\mathbf{A}' = \mathbf{A}\mathbf{D}$ and $\mathbf{D}$ is a 18×18 diagonal matrix with $d_{ii} = f_y$ for $i \in \{1, ..., 6\}$, $d_{ii} = f_x$ for $i \in \{7, ..., 12\}$, and $d_{ii} = f_x f_y$ for $i \in \{13, ..., 18\}$.

With known f_x and f_y, $\vec{\mathcal{P}'}$ can be estimated by solving (15). Since $\mathcal{P}'$ only depends on the elements of $\mathbf{R}$ and $\mathbf{T}$, it can be converted to a point projection matrix in the form of $\lambda[\mathbf{R} \ \mathbf{T}]$. The magnitude of λ is determined by the orthogonality of $\mathbf{R}$, and its sign is determined by the sign of t_3. Hence, given the camera intrinsics, the camera extrinsics can be recovered using the reflection correspondences. [16] also provides another way for estimating $\mathbf{R}$ and $\mathbf{T}$ with given camera intrinsics. In Section 5, we tackle the problem of unknown camera intrinsics by formulating the problem into a nonlinear optimization by minimizing reprojection errors computed based on a cross-ratio formulation for the mirror surface. For initialization purpose, we assume (u_0, v_0) being located at the image center, and $f_x = f_y = f$. We choose a rough range of f and for each sample value of f within the range, we estimate $\mathbf{R}$ and $\mathbf{T}$ by solving (15). The point to line distance criterion in (12) is applied to find the best focal length f'. A camera projection matrix can then be constructed using f', (u_0, v_0), $\mathbf{R}$ and $\mathbf{T}$ that satisfies all the above mentioned constraints.

5. Cross-ratio Based Formulation

In this section, we obtain the camera projection matrix and the mirror surface by minimizing reprojection errors. We will derive a cross-ratio based formulation for recovering a 3D point on the mirror surface from its reflection correspondences. Note that minimizing point-to-point reprojection errors can provide a stronger geometrical constraint than minimizing the point-to-line distances in (12) (see Fig. 4).

Consider a point $\mathbf{M}$ on the mirror surface (see Fig. 5). Let $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$ be its reflection correspondences on the reference plane under three distinct poses, denoted by P_0, P_1 and P_2, respectively. Suppose $\mathbf{M}$, $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$ are projected to the image as $\mathbf{m}$, $\mathbf{x}_0$, $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively. We observe that the cross-ratios $\{\mathbf{M}, \mathbf{X}_0; \mathbf{X}_1, \mathbf{X}_2\}$ and $\{\mathbf{m}, \mathbf{x}_0; \mathbf{x}_1, \mathbf{x}_2\}$ are identical, i.e.,
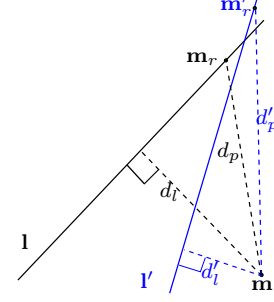

Figure 4: Minimizing point-to-line distance does not guarantee minimizing point-to-point distance. A 3D point $\mathbf{M}$ and a 3D line $\mathcal{L}$ passing through it are projected by $\mathcal{P}$ to a 2D point $\mathbf{m}_r$ and a 2D line $\mathbf{l}$, respectively. Let $\mathbf{m}$ denote the observation of $\mathbf{M}$. The distance between $\mathbf{m}$ and $\mathbf{m}_r$ is d_p, and the distance between $\mathbf{m}$ and $\mathbf{l}$ is d_l. Suppose the same 3D point $\mathbf{M}$ and 3D line $\mathcal{L}$ are projected by $\mathcal{P}'$ to $\mathbf{m}_r'$ and $\mathbf{l}'$, respectively. The distance between $\mathbf{m}$ and $\mathbf{m}_r'$ is d_p', and the distance between $\mathbf{m}$ and $\mathbf{l}$ is d_l'. Note that $d_l' < d_l$, but $d_p' > d_p$.
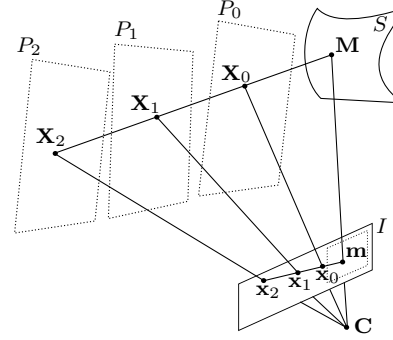

Figure 5: Camera projection matrix and mirror surface points are recovered by minimizing reprojection errors computed from the cross-ratio constraint $\{\mathbf{M}, \mathbf{X}_0; \mathbf{X}_1, \mathbf{X}_2\}$ $= \{\mathbf{m}, \mathbf{x}_0; \mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$ are the correspondences of $\mathbf{M}$ under three different pattern poses and $\mathbf{m}$, $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$ are their projection on image plane. Note that $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$ may not be visible by the camera.

$$\frac{|\overline{\mathbf{X}_1\mathbf{M}}||\overline{\mathbf{X}_2\mathbf{X}_0}|}{|\overline{\mathbf{X}_1\mathbf{X}_0}||\overline{\mathbf{X}_2\mathbf{M}}|} = \frac{|\overline{\mathbf{x}_1\mathbf{m}}||\overline{\mathbf{x}_2\mathbf{x}_0}|}{|\overline{\mathbf{x}_1\mathbf{x}_0}||\overline{\mathbf{x}_2\mathbf{m}}|}. \quad (16)$$

Let s be the distance between $\mathbf{X}_2$ and $\mathbf{M}$ (i.e., $s = |\overline{\mathbf{X}_2\mathbf{M}}|$), from (16)

$$s = \frac{|\overline{\mathbf{X}_2\mathbf{X}_1}||\overline{\mathbf{X}_2\mathbf{X}_0}||\overline{\mathbf{x}_1\mathbf{x}_0}||\overline{\mathbf{x}_2\mathbf{m}}|}{|\overline{\mathbf{X}_2\mathbf{X}_0}||\overline{\mathbf{x}_1\mathbf{x}_0}||\overline{\mathbf{x}_2\mathbf{m}}| - |\overline{\mathbf{X}_1\mathbf{X}_0}||\overline{\mathbf{x}_2\mathbf{x}_0}||\overline{\mathbf{x}_1\mathbf{m}}|}. \quad (17)$$

Given the projection matrix, a surface point $\mathbf{M}$ can be

recovered as

$$\mathbf{M} = \mathbf{X}_2 + s \frac{\overrightarrow{\mathbf{X}_2\mathbf{X}_0}}{|\overrightarrow{\mathbf{X}_2\mathbf{X}_0}|}, \tag{18}$$

where $\overrightarrow{\mathbf{X}_2\mathbf{X}_0}$ denotes the directed ray from $\mathbf{X}_2$ to $\mathbf{X}_0$.

We optimize the projection matrix by minimizing the reprojection errors, i.e.,

$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} (\mathbf{m}_i - \mathbf{m}_i')^2, \tag{19}$$

where $\mathbf{m}_i$ is the observation of $\mathbf{M}_i$, $\mathbf{m}_i' = \mathbf{P}(\theta)\mathbf{M}_i$, and $\theta = [f_x, f_y, u_0, v_0, r_x, r_y, r_z, t_x, t_y, t_z]^{\mathrm{T}}$ [2]. We initialize θ using the method proposed in Section 4, and solve the optimization problem using the Levenberg-Marquardt method. Given the estimated projection matrix, the mirror surface can be robustly reconstructed by solving (16)-(18).

6. Evaluation

To demonstrate the effectiveness of our method, we evaluate it using both synthetic and real data.
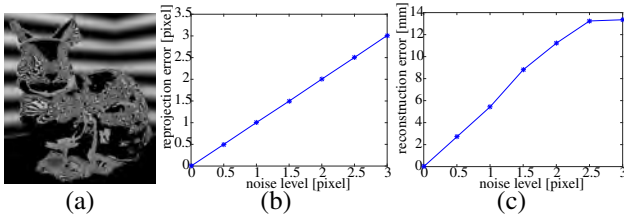
6.1. Synthetic Data

Figure 6: (a) An image of the mirror *Stanford bunny*. (b) RMS reprojection errors (computed against ground truth image points). (c) RMS reconstruction errors (computed against ground truth 3D surface points).

We employed a reflective *Stanford bunny* rendered by Balzer *et al.* [3] to generate our synthetic data. The bunny has a dimension of $880 \times 680 \times 870 \ mm^3$, and $208,573$ surface points. The images have a resolution of 960×1280 pixels. Fig. 6(a) shows the reflective appearance of the bunny. In their original data, the bunny is placed in a cubic room, with each side of the room working as a reference plane. The reference pattern has a dimension of $3048 \times 3048 \ mm^2$. The center of the room is defined as the world origin. A camera is placed in the room viewing the bunny. Since our method requires reflection correspondences under three

---

[2] We used angle-axis representation for rotation, i.e., $[r_x, r_y, r_z]^{\mathrm{T}} = \alpha \mathbf{e}$, where α is the rotation angle and $\mathbf{e}$ is the unit rotation axis.

distinct poses of a reference plane, we introduced two additional planes for each side of the room and obtained the reflection correspondences through ray tracing.

To evaluate the performance of our method, we added Gaussian noise to the image points with standard deviations ranging from 0 to 3.0 pixels. We initialized the projection matrix using the method described in Section 4. The optimized projection matrix together with the 3D surface points were obtained by minimizing reprojection errors computed based on our cross-ratio formulation. Our cross-ratio based formulation can effectively improve the initialization results. An example is given in Table 1. The error in the rotation matrix $\mathbf{R}$ is the angle of the rotation induced by $\mathbf{R}_{gt}\mathbf{R}^{\mathrm{T}}$, where $\mathbf{R}_{gt}$ denotes the ground truth rotation matrix. The error in the translation vector $\mathbf{T}$ is the angle ($\mathbf{T}_{deg}$) between $\mathbf{T}$ and $\mathbf{T}_{gt}$, where $\mathbf{T}_{gt}$ denotes the ground truth translation vector. In addition, we obtain $\mathbf{T}_{scale} = \|\mathbf{T}_{gt} - \mathbf{T}\|$ to estimate the error in $\mathbf{T}$.

	f_u	f_v	u_0	v_0	$\mathbf{R}[°]$	$\mathbf{T}_{deg}[°]$	$\mathbf{T}_{scale}$
L	1.18%	0.84%	1.08%	2.78%	0.63	0.89	1.76%
EL	1.32%	1.32%	0.07%	0.10%	1.01	1.26	2.12%
CR	0.14%	0.14%	0.18%	0.09%	0.06	0.07	0.16%

Table 1: Estimation error under noise lv $\sigma = 2.0 \ [pixel]$ on *bunny*. L: linear solution of Section 4.3; EL: constrained linear solution with strategy in Section 4.4; CR: estimation using cross-ratio formulation initialized with EL.

Fig. 6(b) and (c) depict the root mean square (RMS) reprojection errors and reconstruction errors, respectively, under different noise levels. It can be seen that the reprojection errors are nearly identical to the noise level. While the reconstruction errors increase linearly with the noise level, the magnitude is relatively small compared to the size of the object. Fig. 7 shows the reconstructed point clouds and surfaces. Table 2 shows a quantitative comparison of our estimated projection matrices w.r.t the ground truth. Among all noise levels, the errors are below $\sim 2\%$ for f_u, f_v, u_0, v_0 and $\mathbf{T}_{scale}$, and angular errors are below $1°$ for $\mathbf{R}$ and $\mathbf{T}$.

Besides, we compared our method with state-of-the-art mirror surface reconstruction method [15] under smooth surface assumption and calibrated setup. Note that [15] assumes the mirror surface is C^2 continuous. In order to make fair comparison, we perform the experiment on a *sphere* patch under the same setup with the *bunny* dataset. Fig. 8 depicts the comparison between fully calibrated [15] and uncalibrated (proposed) methods. The overall reconstruction accuracy is similar. While our result is not as smooth as that from [15] due to our point-wise reconstruction, their result shows a global reconstruction bias due to the B-spline parameterization for the surface (see Fig. 8).

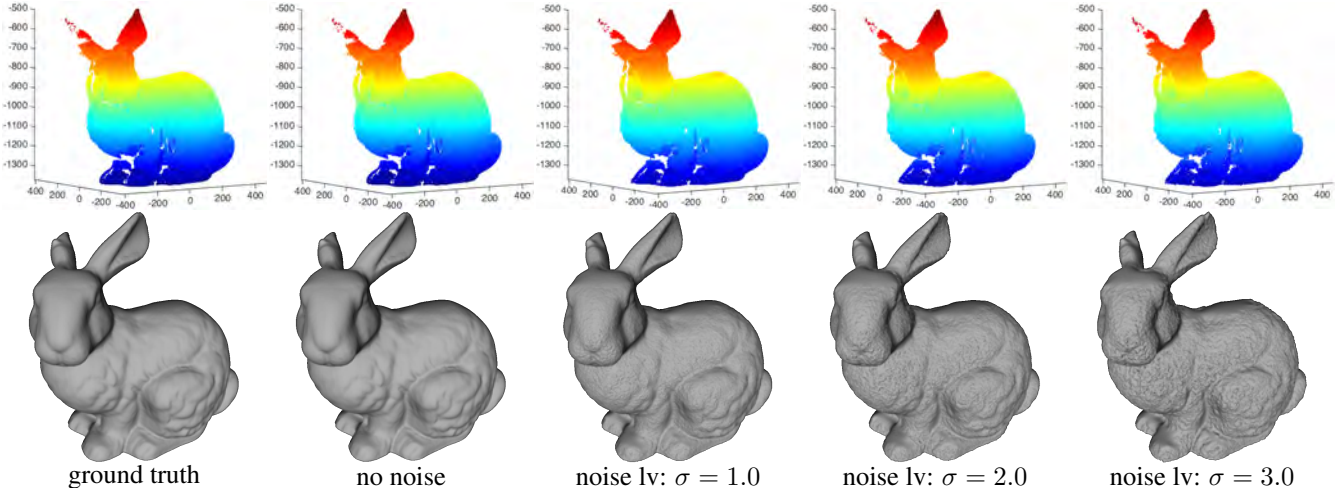| ground truth | no noise | noise lv: $\sigma = 1.0$ | noise lv: $\sigma = 2.0$ | noise lv: $\sigma = 3.0$ |

Figure 7: Top row: reconstructed point clouds under different noise levels. Coordinates are w.r.t world and colors are rendered w.r.t z coordinates. Note that the missing regions are due to the lack of correspondences in the original data set. Bottom row: surfaces generated using screened Poisson surface reconstruction method [13].

	f_u [pixel]	f_v [pixel]	u_0 [pixel]	v_0 [pixel]	$\mathbf{R}$ [°]	$\mathbf{T}_{deg}$ [°]	$\mathbf{T}_{scale}$ [mm]
$\sigma = 0.5$	0.31(0.02%)	0.31(0.02%)	0.49(0.08%)	0.38(0.08%)	0.03	0.03	0.90(0.05%)
$\sigma = 1.0$	0.22(0.02%)	0.22(0.02%)	0.57(0.09%)	0.63(0.13%)	0.04	0.03	0.93(0.05%)
$\sigma = 1.5$	0.62(0.04%)	0.62(0.04%)	0.63(0.10%)	0.15(0.03%)	0.03	0.03	0.92(0.05%)
$\sigma = 2.0$	2.02(0.14%)	2.02(0.14%)	1.17(0.18%)	0.43(0.09%)	0.06	0.07	2.91(0.16%)
$\sigma = 2.5$	7.22(0.52%)	7.22(0.52%)	5.18(0.81%)	2.03(0.42%)	0.22	0.28	11.24(0.62%)
$\sigma = 3.0$	19.11(1.36%)	19.11(1.36%)	13.11(2.05%)	5.01(1.04%)	0.57	0.72	28.79(1.59%)

Table 2: Camera intrinsic and extrinsic estimation error for the *Stanford bunny* dataset. The ground truth for the intrinsic parameters are $f_u = 1400$, $f_v = 1400$, and $(u_0, v_0) = (639.5, 479.5)$.
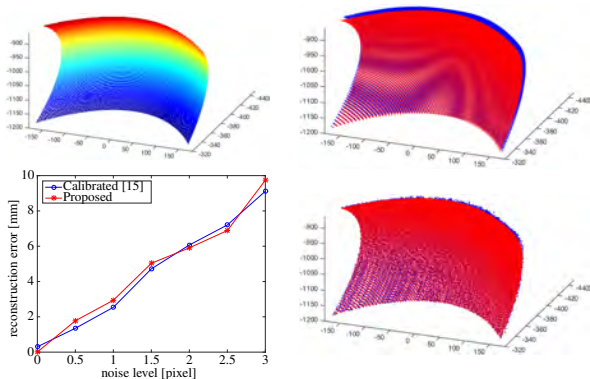

Figure 8: Upper left: ground truth. Lower left: RMS reconstruction errors. Upper right ([15]) & lower right (ours): reconstruction (blue) against ground truth (red) under $\sigma = 2.0$. Reconstruction accuracy is similar.

6.2. Real Data

We evaluated our method on a *sauce boat* and *two spheres* respectively (see Fig. 9). We captured images us-

Figure 9: Top row: *sauce boat* and *two spheres* in real experiments. Bottom row: a sweeping line is reflected by *two spheres* under three distinct positions of the LCD monitor while the camera and mirror surfaces are stationary.

ing a $Canon\ EOS\ 40D$ digital camera with a 24-70 mm lens. A 19 $inch$ LCD monitor was used as a reference plane and was placed at three different positions. For each position, we captured an image sequence of a thin bright stripe sweeping across the screen in vertical direction and then in horizontal direction [14, 10]. For each direction, we examined the intensity value sequence for each image point, and established the reflection correspondence by identifying the image in which the intensity attained a peak value. To improve the accuracy, quadratic approximation was applied to
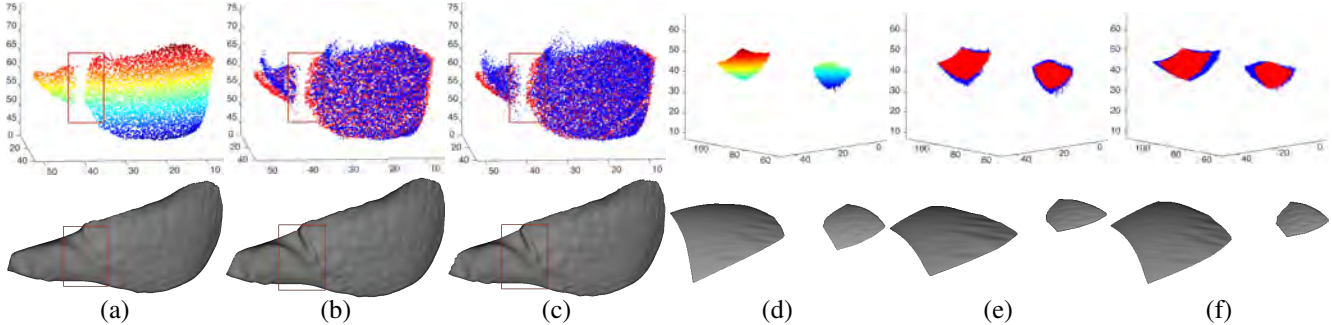
Figure 10: (a)-(c): reconstructions of *sauce boat*. Results are obtained under (a) calibrated camera with calibrated plane poses (this result is treated as ground truth and overlaid in (b) and (c) for comparison (red)); (b) uncalibrated camera with calibrated plane poses (blue); (c) uncalibrated camera with uncalibrated plane poses (ours, blue). Note the missing regions (in red rectangle) in the reconstructed point clouds are filled by the mesh generation algorithm and should be ignored in comparing the surface meshes. (d)-(f): reconstructions of *two spheres*.

	f_u [pixel]	f_v [pixel]	u_0 [pixel]	v_0 [pixel]	$\mathbf{R}$ [°]	$\mathbf{T}_{deg}$ [°]	$\mathbf{T}_{scale}$ [mm]	S_{rms} [mm]
B_{uc}	36.70(0.63%)	21.99(0.38%)	99.10(5.03%)	100.00(8.13%)	9.12	1.00	19.16(8.23%)	2.55
B_{uu}	101.70(1.75%)	86.90(1.49%)	112.10(5.69%)	113.00(9.19%)	9.86	1.99	17.02(7.34%)	2.71
S_{uc}	63.38(1.09%)	68.01(1.17%)	61.49(3.18%)	42.7(3.47%)	6.67	1.78	33.83(8.96%)	1.78
S_{uu}	81.38(1.40%)	86.02(1.48%)	81.67(4.14%)	56.70(4.61%)	7.17	2.13	37.69(9.98%)	2.03

Table 3: Real experiments evaluation. B and S denote the results of *sauce boat* and *two spheres*, respectively. The subscripts uc and uu stand for experiments under an uncalibrated camera with calibrated plane poses and under an uncalibrated camera with uncalibrated plane poses, respectively. The ground truth for the intrinsic parameters are $f_u = 5812.86$, $f_v = 5812.82$, and $(u_0, v_0) = (1971.95, 1230.02)$. S_{rms} stands for the RMS reconstruction error.

the intensity profile in the neighborhood of the peak value.

After establishing reflection correspondences, we first estimated the relative poses of the reference plane using the method in [16]. We then formed 3D lines from the reflection correspondences on the reference plane under the two poses that are furthest apart (e.g., P_0 and P_2 in Fig. 5). These 3D lines were used to obtain a preliminary solution of projection matrix using the method in Section 4. This was used to initialize the nonlinear optimization described in Section 5.

To evaluate our method, we calibrated the camera and reference plane poses using [8]. We used the calibration result to estimate the surface and treated it as the ground truth. This result was compared against the result obtained using uncalibrated camera but calibrated plane poses, and our result using uncalibrated camera and uncalibrated plane poses. Fig. 10 shows the reconstructed surfaces and Table 3 shows the numerical errors. We aligned each estimated surface with the ground truth by a rigid body transformation before computing the reconstruction error [5]. The RMS reconstruction errors are below 3 mm. f_u and f_v errors are below 2%. u_0, v_0 and $\mathbf{T}_{scale}$ errors are below 10%. The angular errors are below 10° for $\mathbf{R}$ and below 3° for $\mathbf{T}$. The errors in intrinsics and extrinsics are larger than those in the synthetic experiments. This is reasonable since accurate specular correspondences in real case are difficult to obtain due to the large and complex distortion caused by the mirror

surface and varying lighting condition. The qualitative and quantitative results suggest the accuracy of our method.

7. Discussions and Conclusions

A novel method is introduced for mirror surface reconstruction. Our method works under an uncalibrated setup and can recover the camera intrinsics and extrinsics, along with the surface. We first proposed an analytical solution for camera projection matrix estimation, and then derived a cross-ratio based formulation to achieve a robust estimation. Our cross-ratio based formulation does not encounter degeneracy. However, degenerate cases (e.g., a planar mirror, a spherical mirror, etc) may occur to the system due to the application of [16] to estimate relative poses of the reference plane. Employing methods without degeneracy to estimate the relative poses will help handle these cases.

The proposed method only needs reflection correspondences as input and removes the restrictive assumptions of known motions, C^n continuity of the surface, and calibrated camera that are being used by other existing methods. This greatly simplifies the challenging problem of mirror surface recovery. We believe our work can provide a meaningful insight towards solving this problem. In the future, we would like to extend the proposed method to recover complete surfaces and investigate inter-reflection cases.

References

[1] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *ICCV*, pages 1–8, 2007.

[2] Y. Adato, Y. Vasilyev, T. Zickler, and O. Ben-Shahar. Shape from specular flow. *PAMI*, 32(11):2054–2070, 2010.

[3] J. Balzer, D. Acevedo-Feliz, S. Soatto, S. Höfer, M. Hadwiger, and J. Beyerer. Cavlectometry: Towards holistic reconstruction of large mirror objects. In *International Conference on 3D Vision (3DV)*, pages 448–455, 2014.

[4] J. Balzer and S. Werling. Principles of shape from specular reflection. *Measurement*, 43(10):1305–1317, 2010.

[5] P. J. Besl and H. D. McKay. A method for registration of 3-d shapes. *PAMI*, 14(2):239–256, 1992.

[6] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *ICCV*, pages 691–696, 2003.

[7] T. Bonfort, P. Sturm, and P. Gargallo. General specular surface triangulation. In *ACCV*, pages 872–881, 2006.

[8] J.-Y. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/.

[9] G. D. Canas, Y. Vasilyev, Y. Adato, T. Zickler, S. Gortler, and O. Ben-Shahar. A linear formulation of shape from specular flow. In *ICCV*, pages 191–198, 2009.

[10] K. Han, K.-Y. K. Wong, and M. Liu. A fixed viewpoint approach for dense reconstruction of transparent objects. In *CVPR*, pages 4001–4008, 2015.

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[12] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction. *Computer Graphics Forum*, 29:2400–2426, 2010.

[13] M. Kazhdan and H. Hoppe. Screened Poisson surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32:29:1–29:13, 2013.

[14] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3D shape by light-path triangulation. *IJCV*, 76:13–29, 2008.

[15] M. Liu, R. Hartley, and M. Salzmann. Mirror surface reconstruction from a single image. *PAMI*, 37(4):760–773, 2015.

[16] M. Liu, K.-Y. K. Wong, Z. Dai, and Z. Chen. Pose estimation from reflections for specular surface recovery. In *ICCV*, pages 579–586, 2011.

[17] D. Nehab, T. Weyrich, and S. Rusinkiewicz. Dense 3d reconstruction from specularity consistency. In *CVPR*, pages 1–8, 2008.

[18] M. Oren and S. K. Nayar. A theory of specular surface geometry. *IJCV*, 24(2):105–124, 1996.

[19] L. Perdigoto and H. Araujo. Calibration of mirror position and extrinsic parameters in axial non-central catadioptric systems. *CVIU*, 117:909–921, 2013.

[20] S. Ramalingam, M. Antunes, D. Snow, G. Hee Lee, and S. Pillai. Line-sweep: Cross-ratio for wide-baseline matching and 3d reconstruction. In *CVPR*, pages 1238–1246, 2015.

[21] S. Ramalingam, P. Sturm, and S. K. Lodha. Towards complete generic camera calibration. In *CVPR*, pages 1093–1098, 2005.

[22] I. Reshetouski and I. Ihrke. Mirrors in computer graphics, computer vision and time-of-flight imaging. *Lect. Notes Comput. Sc.*, 8200:77–104, 2013.

[23] S. Roth and M. J. Black. Specular flow and the recovery of surface structure. In *CVPR*, pages 1869–1876, 2006.

[24] S. Rozenfeld, I. Shimshoni, and M. Lindenbaum. Dense mirroring surface recovery from 1d homographies and sparse correspondences. *PAMI*, 33(2):325–337, 2011.

[25] A. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal. Specular surface reconstruction from sparse reflection correspondences. In *CVPR*, pages 1245–1252, 2010.

[26] S. Savarese and P. Perona. Local analysis for 3d reconstruction of specular surfaces. In *CVPR*, pages 738–745, 2001.

[27] S. Savarese and P. Perona. Local analysis for 3d reconstruction of specular surfaces – part ii. In *ECCV*, pages 759–774, 2002.

[28] P. Sturm and T. Bonfort. How to compute the pose of an object without a direct view. In *ACCV*, pages 21–31, 2006.

[29] Y. Vasilyev, T. Zickler, S. Gortler, and O. Ben-Shahar. Shape from specular flow: Is one flow enough? In *CVPR*, pages 2561–2568, 2011.

[30] M. Weinmann, A. Osep, R. Ruiters, and R. Klein. Multi-view normal field integration for 3d reconstruction of mirroring objects. In *ICCV*, pages 2504–2511, 2013.