Shape Reconstruction and Camera Self-Calibration Using Cast Shadows and Scene Geometries

Hiroshi Kawasaki · Ryo Furukawa

Abstract Recently, various techniques of shape reconstruction using cast shadows have been proposed. These techniques have the advantage that they can be applied to various scenes, including outdoor scenes, without using special devices. Previously proposed techniques usually require calibration of camera parameters and light source positions, and such calibration processes limit the range of application of these techniques. In this paper, we propose a method to reconstruct 3D scenes even when the camera parameters or light source positions are unknown. The technique first recovers the shape with 4-DOF indeterminacy using coplanarities obtained by cast shadows of straight edges or visible planes in a scene, and then upgrades the shape using metric constraints obtained from the geometrical constraints in the scene. In order to circumvent the need for calibrations and special devices, we propose both linear and nonlinear methods in this paper. Experiments using simulated and real images verified the effectiveness of this technique.

Keywords Shape reconstruction · Shape from cast shadow · Camera calibration · Self-calibration

H. Kawasaki (✉)
Faculty of Engineering, Saitama University, 255, Shimo-okubo, Sakura-ku, Saitama, Japan
e-mail: kawasaki@cgv.ics.saitama-u.ac.jp

R. Furukawa
Faculty of Information Sciences, Hiroshima City University, 3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima, Japan
e-mail: ryo-f@cs.hiroshima-cu.ac.jp

1 Introduction

To date, various techniques for scene shape reconstruction using shadows have been proposed. One of the advantages of using shadows for scene shape reconstruction is that the information for 3D reconstruction can be acquired without the use of special devices, because shadows exist wherever light is present. Such techniques are applicable to outdoor poles on a sunny day or indoor objects under indoor lighting. Another advantage of shape reconstruction using shadows is that only a single camera is required.

However, most previously proposed methods assumed that the light sources were at known positions because unknown positions would result in ambiguities in the solution and Euclidean reconstruction cannot be achieved (Kriegman and Belhumeur 2001). If a shape can be reconstructed using unknown light source positions, then the technique can be used for a wider range of applications. For example, a remote web camera is usually used under such condition. In practical cases, the intrinsic parameters of a remote web camera are often unknown, thus, the application becomes more useful if those parameters can also be estimated at the same time. We propose such a method in this paper. Our technique is actually more general; that is, both the object that casts shadows and the light source can be freely moved while scanning because both of their positions do not need to be known and static. This is a great advantage for actual scanning processes, because moving the light source can drastically reduce unmeasured areas caused by self-shadows.

To actually realize this technique, we propose a novel formulation of simultaneous linear equations from planes created by the shadows of straight edges (shadow planes) and the real planes in the scene. These equations are extensions

of previous studies examining the shapes from planes (Bartoli and Sturm 2003; Kawasaki and Furukawa 2007) and interpretations of line drawings of polyhedrons (Sugihara 1986). Shadow planes and real planes are treated equally in our formulation, so various geometrical constraints of the planes can be utilized efficiently for Euclidean upgrade and camera calibration. In the proposed method, such geometrical constraints are assumed to be given as inputs. We developed a user interface for easy input of the constraints with little manual intervention required (see Sect. 4.1.2 for more details). Our method is especially advantageous in the case that the camera or the light source position is not known while some geometrical information of the scene such as orthogonalities or parallelisms are known.

In this paper, we assume two typical situations to reconstruct a scene. The first situation, which we call "*shadows of the static object*," assumes a fixed camera position, a static scene, and a static object of a straight edge which casts a moving shadow as the light source (*e.g.*, the sun or a point light source) moves. The second situation, which we call "*active scan by cast shadow*," assumes a fixed camera, a static scene, and arbitrary motion of both a light source and an object with a straight edge to generate shadows for conducting an active scan.

Both indoor and outdoor scenes are considered for the target scene. For the indoor case, since the light source and other conditions are easily controlled, the method can be used as a type of "active scanner." For the outdoor case, since the method does not require information about the camera's intrinsic and extrinsic parameters and the light source position, in general the method can be used as a type of "active scanner" using the sun or an artificial light source. Since outdoor scenes are usually very bright in the daytime and it is difficult to use artificial light sources in those situations, the sun is the most suitable light source in principle. Therefore, we conducted the experiment using the sun. In terms of the shadow sweeping method for outdoor scenes, moving the straight bar around manually is preferable for quick and efficient scanning, however, for a remote web-camera, the user cannot do anything about the captured scene and the light source, and thus, "a static bar with a moving light source" (the sun) is adopted. We also conducted an experiment under such a condition and successfully reconstructed the 3D shapes as described in detail in Sect. 5.4.

2 Related Work

2.1 Shape from Cast Shadows with Known Light Positions

Three-dimensional reconstruction using information from shadows has a long history. Shafer and Kanade presented the mathematical formulation of shadow geometries and derived constraints for surface orientation from shadow boundaries (Shafer and Kanade 1983). Hambrick et al. proposed a method for classifying the boundaries of shadow regions (Hambrick et al. 1987). Subsequently, several other methods for recovering 3D shapes up to Euclidean reconstruction based on geometrical constraints of cast shadows (*i.e.*, works of shape from cast shadows) have been proposed (Hatzitheodorou and Kender 1988; Raviv et al. 1989; Daum and Dudek 1998; Savarese et al. 2007). Yu and Chang (2005) reconstructed the height field of bumpy textures under the assumption of known light directions. All of these methods assume that objects that cast shadows are static and the light directions or positions are known. Yamazaki et al. (2007) proposed a method for 3D reconstruction by applying a shape-from-silhouette technique to cast shadows of a target object projected onto a single plane and was able to obtain accurate results. Although the light positions were not calibrated in advance, they used shadows of spheres captured with the shadows of the target object to estimate the light positions for each of the multiple images.

In some research, techniques of shape from cast shadows were used to develop active 3D scanning systems. Bouguet et al. proposed an active 3D scanning method based on shape from cast shadows in which users moved a straight-edged object freely so that the shadow generated by a fixed light source scanned the object (Bouguet and Perona 1998; Bouguet et al. 1999). However, the technique assumed that the camera parameters, the light source position, and a reference plane were all known.

2.2 Shape from Cast Shadows with Unknown Light Positions

If a Euclidean shape can be reconstructed with unknown light source positions, it may broaden the application of shape from cast shadow techniques. However, it was proven that scene reconstructions based on binary shadow regions have ambiguities for four degrees of freedom (DOFs) if the light positions are unknown (Kriegman and Belhumeur 2001). In the case of a perspective camera, these ambiguities correspond to the family of transformations called generalized projective bas-relief (GPBR) transformations.

Caspi and Werman proposed a method using two straight, parallel and fixed objects to cast shadows and a reference plane (*e.g.*, the ground) (Caspi and Werman 2006) to deal with unknown light source positions. To solve ambiguities caused by unknown light sources, they used parallelisms of shadows of straight edges by detecting vanishing points. Our method is more general when compared to theirs. For example, in our method, the camera may be partially calibrated or uncalibrated, the straight object and the light source can be moved, the light source can be a parallel or point light

source, and wider constraint types than parallelisms of shadows can be used to resolve ambiguities.

2.3 Geometric Reconstruction Using Planes

One of the other approaches proposed to recover geometric information is utilizing planes in the scene. We extended this approach not only by using real planes in the scene, but also by using the coplanarities of points (where coplanarity refers to a property of points being on a single plane). Examples of research utilizing planes in the scene are works of Sugihara (1986) and Criminisi et al. (2000). Sugihara analyzed line drawings of polyhedrons by deriving constraints between planar faces of the polyhedrons. His work assumed scenes of polyhedrons, while our method can densely reconstruct scenes with curved surfaces by utilizing shadow curves. Criminisi et al. assumed parallel planes in the scene and proposed a method for determining ratios of the distances between points that are on the planes. Their method requires parallel planes with a known line of infinity and parallel lines with a known vanishing point, whereas our method does not assume any constraints between planes. This property of our method can be achieved because relationships between planes that are not parallel can also be described as linear equations by using coplanar curves of shadow boundaries.

3 Shape Reconstruction from Cast Shadows

If a set of points exist on the same plane, they are said to be coplanar. A scene composed of plane structures has many coplanarities. In this paper, a coplanarity that is actually observed as a real plane in the scene is termed an *explicit coplanarity*; points or curves on the real plane are said to be *explicit-coplanar*; and a real plane is sometimes referred to as an *explicit plane*. All the points on a region of an image that correspond to a planar surface are explicit-coplanar. For example, in Fig. 1(a), the sets of points on a region of the same hatch pattern, except for white regions, are explicit-coplanar.

As opposed to this, in a 3D space there exist an infinite number of coplanarities that are not explicitly observed in ordinary situations, but could be observed under specific conditions. For example, the boundary of a cast shadow of a straight edge is the set of coplanar points shown in Fig. 1(b). This kind of coplanarity is not visible until the shadow is cast on the scene. In this paper, we refer to these coplanarities as *implicit coplanarities*; points or curves on the shadow boundary are said to be *implicit-coplanar*, and a plane that includes implicit-coplanar points is referred to as an *implicit plane*. Implicit coplanarities can be observed in various situations, such as in the case that buildings with straight
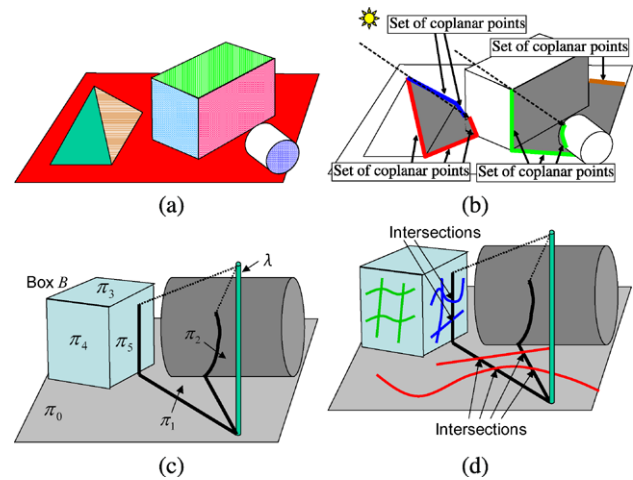
Fig. 1 Coplanarities in a scene: (**a**) Explicit coplanarities. Points on a region of the same hatch pattern except for white regions are a set of explicit-coplanar points. Note that the points on a region of a curved surface are not coplanar. (**b**) Implicit coplanarities. Four sets of implicit-coplanar points are shown. (**c**) Examples of metric constraints: π_1 and π_2 are implicit planes and π_0, π_3, π_4, and π_5 are explicit planes. $\pi_0 \perp \pi_1$ and $\pi_0 \perp \pi_2$ if $\lambda \perp \pi_0$. $\pi_3 \perp \pi_4$, $\pi_4 \perp \pi_5$, $\pi_3 \perp \pi_5$, and π_3 and π_0 are parallel if box B is rectangular and is placed on π_0. (**d**) Intersections between explicit-coplanar curves and implicit-coplanar curves in a scene. Curves drawn within a planar region are explicit-coplanar. Explicit-coplanar curves can be drawn by hand

edges are under the sun and cast shadows onto the scene. Although explicit-coplanar curves are observed only for planar parts of a scene, implicit coplanarities can be observed on arbitrary-shaped surfaces.

In this study, we first capture multiple images of the scene while the cast shadows of straight edges projected on the scene is moving (*i.e.*, changing the light source position or moving an object that casts shadows). The shadow boundaries are implicit-coplanar. Then we extract points on the intersections of multiple planes from the image of the scene. These planes may be explicit or implicit. From the locations of the points on the images, we can create simultaneous linear equations. By solving the simultaneous equations, the scene can be reconstructed up to a 4-DOF indeterminacy that is shown, in this paper, to be an essential property of the simultaneous equations, and another indeterminacy is caused by unknown camera parameters. Since the indeterminacies of the solution are represented as 3D homographies, we refer to the solution as the *projective reconstruction*. To obtain a Euclidean reconstruction from the solution, those indeterminacies must be eliminated. In this paper, the solution obtained from this process is termed *metric reconstruction*. To achieve a metric reconstruction, constraints other than coplanarities must be used. For many scenes, especially those that include artificial objects, we can find geometrical constraints among explicit and implicit planes. Examples of metric constraints are explained below.

(1) In Fig. 1(c), the ground is plane π_0, and linear object λ is standing vertically on the ground. If the planes corresponding shadows of λ are π_1 and π_2, $\pi_0 \perp \pi_1$, $\pi_0 \perp \pi_2$ can be derived from $\lambda \perp \pi_0$.

(2) In the same figure, the sides of box B are π_3, π_4, and π_5. If box B is rectangular, π_3, π_4, and π_5 are orthogonal to each other. If box B is on the ground, π_3 is parallel to π_0.

From the constraints available from the scene, such as the above examples, we can determine variables corresponding to the remaining indeterminacy and achieve metric reconstruction. With enough constraints, the camera parameters can be estimated at the same time. We call these constraints the *metric constraints*. In real cases, it is sometimes necessary to process numerous number of frames. For efficient processing in such cases, only selected frames are processed for the projective and metric reconstruction, and we reconstruct dense shapes using other methods. Based on this, the actual flow of the algorithms are as follows.

Step 1: Extraction of coplanarity constraints. From a series of images that are acquired from a scene with shadows captured by a fixed camera, shadow boundaries are extracted as implicit-coplanar curves. If the scene has plane areas, explicit-coplanar points are sampled from the area. In this paper, explicit-coplanar curves and points are acquired by simply drawing curves on a planar region of the image as shown in Fig. 1(d).

Step 2: Projective reconstruction using coplanarity constraints. Constraints are acquired as linear equations from a dataset of coplanarities. By numerically solving the simultaneous equations, projective reconstruction with indeterminacy of four or more DOFs can be acquired.

Step 3: Metric reconstruction using metric constraints. To achieve metric reconstruction, an upgrade process of the solution of Step 2 is required. The solution can be upgraded by solving the metric constraints.

Step 4: Dense shape reconstruction. The processes in Steps 2 and 3 are performed on selected frames. To realize dense shape reconstruction of a scene, implicit-coplanar curves from all the images are used to reconstruct 3D shapes using the results of the preceding processes.

4 Algorithm Details for Each Step

4.1 Data Acquisition

The data acquisition process of our method consists of two parts: the first is acquisition of coplanar shadow boundaries and the second is extraction of geometrical information from the scene. As already described, our method requires several manual interactions; and the manual processes are all related to the data acquisition process, and thus, we explain our data acquisition process including the manual process in detail in the following sections.

### 4.1.1 Shadow Boundary Acquisition

To detect coplanarities in a scene, extraction of the boundaries of cast shadows in a scene and labeling of each implicit plane are required. In terms of the extraction of a shadow area from a scene, several studies have been conducted over a long period of time (Jiang and Ward 1994; Salvador et al. 2004), and we have adopted the inter-frame subtraction-based method for our implementation. In terms of labeling shadow boundaries with each implicit plane, a simple image-processing algorithm would fail because shadow boundaries belonging to the same implicit plane are frequently segmented due to scene geometry. For a solution, we adopt a spatio-temporal based method to automatically label boundaries as follows.

1. Images are captured from a fixed camera under the condition that a cast shadow moves due to the motion of a light source or the motion of an object. Background subtraction is applied to the images to separate them into shadow and non-shadow regions, and a spatio-temporal image is created by stacking the images (Fig. 2(b)).

2. The spatio-temporal image is divided into shadow or non-shadow regions using 3D segmentation, which is achieved by applying a region-growing method to the spatio-temporal image. To deal with noise in real images, we merge small regions into the surrounding large regions.

3. After Step 2, there still remains small regions which have no large regions nearby. Therefore, we first delete all the regions that are smaller than a threshold value. Then, the shadow regions are selected manually from the remaining segmented regions (Fig. 2(c)).

4. The segmented regions are again divided into frames, and implicit-coplanar curves of shadow boundaries are extracted from each frame as boundaries of the divided regions (Fig. 2(d)). As shown in this figure, segmented shadow boundaries are successfully categorized into the same implicit planes.

For indoor environments, the proposed spatio-temporal based method usually works accurately and automatic shadow detection can be achieved; however, for an outdoor environments, we cannot obtain accurate results because of noise. In such cases, we remove the noise manually. Also, incorrect connections are segmented manually (*e.g.*, the right and left sides of the shadow of a thin rod are connected if it is not long enough). Note that because such manual interventions are applied to 3D segments, not to each frame, we are successful in drastically reducing the amount of human intervention required.

Fig. 2 (Color online) An example of shadow extraction using spatio-temporal analysis: (**a**) Input image. (**b**) Extracted shadow region by inter-frame subtraction. (**c**) Selected shadow region by using spatio-temporal analysis. (**d**) Extracted edges of selected shadow region. Each color belong to the same implicit planes, respectively

(a)   (b)

(c)   (d)
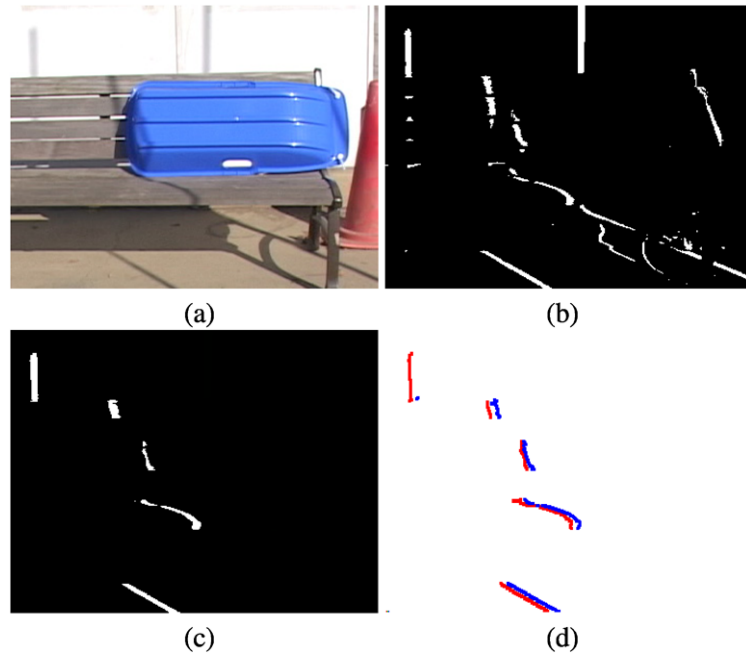
Self-shadows can be treated the same as other shadows. If a self-shadow is cast from a straight edge in a scene, its shadow boundaries can be used as the set of the coplanar points. In addition, in this case, not only the shadow boundaries, but also the straight edge itself can be used as the coplanar points and those are efficiently used to construct the equations described in the next section. (This is the case used for Fig. 1(c) and the simulated environment described in Sect. 5.1.)

4.1.2 Geometric Constraint Acquisition

In our method, geometric constraints are needed to achieve Euclidean reconstruction. To retrieve these constraints, two steps are necessary: the extraction of explicit planes in a scene and the description of the relationship between them.

In terms of the extraction of explicit planes, freely drawn curves on a coplanar area on an image, as shown in Fig. 1(d), are sufficient. These curves are explicit-coplanar. In terms of the description of the relationship between the explicit-coplanar curves, matrix representation is efficient. Although those two steps can be independently conducted, it is more efficient to reduce human interaction by integrating both processes into a single system. We have developed such an interface to conduct experiments for this paper. In reality, the minimum required number of geometric constraints is not large (*e.g.*, if the camera's intrinsic parameters are known, only three constraints are required), and so human intervention for this task is small.

By drawing all the implicit-coplanar curves detected as shadow boundaries and all the explicit-coplanar curves drawn manually on a single image, we can acquire many intersections. Since one intersection shares at least two planes,

we can construct simultaneous equations. The numerical solution of these equations is explained in the following section.

4.2 Projective Reconstruction

Suppose there is a set of N planes. The planes can be implicit or explicit. Let the j-th plane of the set be π_j. It is assumed that all of these planes do not include the optical center of the camera. Then, we can express the plane π_j in the camera coordinates system by the following equation:

$$a_j x + b_j y + c_j z + 1 = 0. \tag{1}$$

Suppose there is a set of points such that each point of the set exists on the intersections of multiple planes, and we assume that all the images of the points are observable. Let the i-th point of the set be represented as ξ_i and let this point exist on the intersection of π_j and π_k. Let coordinates $\mathbf{u}_i = [u_i, v_i]^\top$ be the image of ξ_i on the image plane, and the image's homogeneous coordinates be defined as $\tilde{\mathbf{u}}_i \equiv [\mathbf{u}_i^\top, 1]^\top$. Let the 3D coordinates of ξ_i be $\mathbf{x}_i = [x_i, y_i, z_i]^\top$, and its homogeneous coordinates be defined as $\tilde{\mathbf{x}}_i \equiv [\mathbf{x}_i^\top, 1]^\top$. In addition, let the upper triangular matrix that represents the intrinsic parameter of the camera be $\mathbf{K}$, so that the projection of ξ_i can be represented as

$$\zeta_i \tilde{\mathbf{u}}_i = \mathbf{K}[\mathbf{I}|\mathbf{0}_3]\tilde{\mathbf{x}}_i = \mathbf{K}[\mathbf{I}|\mathbf{0}_3]\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha & \sigma & p_u \\ 0 & \beta & p_v \\ 0 & 0 & 1 \end{bmatrix}[\mathbf{I}|\mathbf{0}_3]\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}, \tag{2}$$

where α, β, σ, p_u and p_v are intrinsic parameters of the camera, $\mathbf{I}$ is the 3×3 identity matrix, $\mathbf{0}_3$ is the 3×1 zero vector, and ζ_i is a scalar that is defined so that the third element of $\tilde{\mathbf{u}}_i$ is 1. Let the parameter vector of plane π_j be $\mathbf{a}_j = [a_j, b_j, c_j]^\top$. From form (2), $\zeta_i \tilde{\mathbf{u}}_i = \mathbf{K}\mathbf{x}_i$, thus, $\mathbf{x}_i = \mathbf{K}^{-1}\zeta_i \tilde{\mathbf{u}}_i$ since $\mathbf{K}$ is regular.

Therefore, from (1),

$$\mathbf{a}_j^\top \mathbf{x_i} + 1 = \zeta_i \mathbf{a}_j^\top \mathbf{K}^{-1}\tilde{\mathbf{u}}_i + 1 = \zeta_i \mathbf{a}_j^{*\top}\tilde{\mathbf{u}}_i + 1 = 0, \tag{3}$$

where $\mathbf{a}_j^{*\top} \equiv \mathbf{a}_j^\top \mathbf{K}^{-1}$. $\mathbf{a}_j^*$ can be regarded as a new parameter vector of plane π_j into which the intrinsic parameters are incorporated. Later in this section, we describe a method to obtain $\mathbf{a}_j^*$. By substituting $1/\zeta_i$ with t_i,

$$\mathbf{a}_j^{*\top}\tilde{\mathbf{u}}_i + t_i = 0. \tag{4}$$

Since the point ξ_i is also on the plane π_k,

$$\mathbf{a}_k^{*\top}\tilde{\mathbf{u}}_i + t_i = 0. \tag{5}$$

Equation (5) includes a variable t_i which depends on the depth of point ξ_i. In many cases, the number of points, M, becomes far more than the number of planes, N. In these cases, using t_i as a variable causes a large increase of the number of variables, which is undesirable from the viewpoint of computational costs. Therefore, we eliminate t_i from (4) and (5). Then, the linear homogeneous equation

$$\tilde{\mathbf{u}}_i^\top \mathbf{a}_j^* - \tilde{\mathbf{u}}_i^\top \mathbf{a}_k^* = \tilde{\mathbf{u}}_i^\top (\mathbf{a}_j^* - \mathbf{a}_k^*) = 0 \tag{6}$$

with variables $\mathbf{a}_j^*$ and $\mathbf{a}_k^*$ can be obtained for each point.

A geometrical interpretation of (6) is that the image of the intersection line of two planes (π_i and π_k) goes through point $\mathbf{u}_i$ on the image plane. A brief explanation of this follows: A line on the image plane can be represented as a 3D vector $\mathbf{l} = [l_x, l_y, l_z]^\top$, such that all the points $\tilde{\mathbf{x}}(\equiv [\mathbf{x}^\top, 1]^\top = [x, y, 1]^\top)$ on the line satisfy $\mathbf{x}^\top \mathbf{l} = 0$. Also, let the intersection line of π_j and π_k be ϕ_i. If the camera is a normalized camera ($\mathbf{K} = \mathbf{I}$), the image of line ϕ_i on the image plane can be represented as the 3D vector $\mathbf{a}_j - \mathbf{a}_k$ (Bouguet et al. 1999). In the case that the camera is not normalized ($\mathbf{K} \neq \mathbf{I}$), it can be proven that the image of the line ϕ_i on the image plane can be represented as $\mathbf{K}^{-1}(\mathbf{a}_j - \mathbf{a}_k) = \mathbf{a}_j^* - \mathbf{a}_k^*$. Therefore, if the image of line ϕ_i goes through point $\mathbf{u}_i$ in the image plane, (6) holds true.

Equation (6) for M points forms a set of simultaneous homogeneous linear equations. These can be represented by a matrix $\mathbf{L}$ with M rows and $3N$ columns whose elements can be described by u_i, v_i $(1 \le i \le M)$, and a $3N$-dimensional solution vector $\mathbf{p} \equiv [\mathbf{a}_1^{*\top}, \ldots, \mathbf{a}_N^{*\top}]^\top$. By using these expressions, the equations can be described in matrix form as

$$\mathbf{L}\mathbf{p} = \mathbf{0}_M. \tag{7}$$

The explicit form of $\mathbf{L}$ is

$$\mathbf{L} \equiv \begin{array}{c} \\ 1 \\ \vdots \\ i \\ \vdots \\ M \end{array} \overset{\displaystyle 1 \quad \cdots \quad j \quad \cdots \quad k \quad \cdots \quad N}{\begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \tilde{\mathbf{u}}_i^\top & \cdots & -\tilde{\mathbf{u}}_i^\top & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}}, \tag{8}$$

where the side and top indices specify rows and columns of 1×3 blocks.

Since (6) is a homogeneous equation and only the difference $\mathbf{a}_j^* - \mathbf{a}_k^*$ appears in it, there are indeterminacies of scaling with the same factor and addition of the same vector. That is, if a set of variables $\mathbf{a}_j^* = \mathbf{a}_j'$ and $\mathbf{a}_k^* = \mathbf{a}_k'$ is a solution of (6), the set of

$$\mathbf{a}_j^* = s\mathbf{a}_j' + \mathbf{b}, \qquad \mathbf{a}_k^* = s\mathbf{a}_k' + \mathbf{b} \tag{9}$$

is also a solution for an arbitrary scalar s and an arbitrary 3D vector $\mathbf{b}$. Thus, if there exists a solution $\mathbf{p}(\equiv [\mathbf{a}_1^{*\top}, \ldots, \mathbf{a}_N^{*\top}]^\top) = [\mathbf{a}_1'^\top, \ldots, \mathbf{a}_N'^\top]^\top$ for (7), then,

$$\mathbf{p} \equiv \begin{bmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_N^* \end{bmatrix} = \begin{bmatrix} s\mathbf{a}_1' + \mathbf{b} \\ \vdots \\ s\mathbf{a}_N' + \mathbf{b} \end{bmatrix} = s\begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_N' \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \vdots \\ \mathbf{b} \end{bmatrix} \tag{10}$$

is also a solution for an arbitrary scalar s and an arbitrary 3D vector $\mathbf{b}$. Thus, solution $\mathbf{p}$ has at least 4-DOF indeterminacy.

Here, we assume that the solutions for (7) have only the 4-DOF indeterminacy (we discuss degenerate conditions later, which cause 5-DOF or more indeterminacy). In this case, the general form of the solutions of (7) can be given in form (10). Because of the indeterminacy of adding vector $[\mathbf{b}^\top, \ldots, \mathbf{b}^\top]^\top$, one of the parameter vectors $\mathbf{a}_j^*$, $j = 1, \ldots, N$, for example, $\mathbf{a}_N^*$, can be treated as a free parameter vector without losing generality of form (10). So, we first calculate a particular solution $[\mathbf{a}_1'^\top, \ldots, \mathbf{a}_N'^\top]^\top$ such that $\mathbf{a}_N' = \mathbf{0}$. To do this, by defining $\mathbf{p}' \equiv [\mathbf{a}_1'^\top, \ldots, \mathbf{a}_{N-1}'^\top]^\top$, and by defining $\mathbf{L}'$ as the submatrix of the left $3N - 3$ columns of $\mathbf{L}$, we can obtain the equation $\mathbf{L}'\mathbf{p}' = \mathbf{0}$. Then, $\mathbf{p}'$ can be obtained as the basis of the 1-D null space of $\mathbf{L}'$ using singular value decomposition (SVD), and the general solution $\mathbf{p}$ can be represented by

$$\mathbf{p} \equiv \begin{bmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_N^* \end{bmatrix} = s\begin{bmatrix} \mathbf{p}' \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \vdots \\ \mathbf{b} \end{bmatrix} \equiv s\begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_{N-1}' \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \vdots \\ \mathbf{b} \end{bmatrix} \tag{11}$$

using an arbitrary scalar s and an arbitrary 3D vector $\mathbf{b}$.
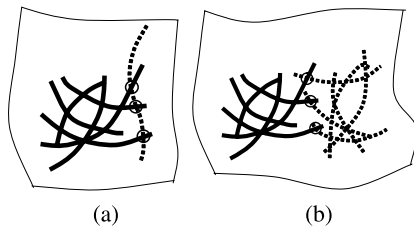
Fig. 3 Examples of solutions having more than 4-DOF indeterminacy

4.2.1 Degenerate Conditions

To obtain the solution of form (11), it was assumed that the degrees of freedom of solution **p** of (7) is 4. Here, we discuss the conditions under which this assumption holds true. If there are N planes in the scene, the number of parameters of the planes is $3N$. Since one point yields one question, at least $3N - 4$ points are required for the solution having 4-DOF indeterminacy. However, even if this condition is fulfilled, the solution may have indeterminacy of more degrees of freedom from degenerate conditions caused by the specific locations of the points or the specific shape of the scene. In the following are examples of such cases.

Figure 3(a) shows shadow curves drawn as bold solid curves and a single dashed curve. In the figure, "all" the intersection points (circles in the figure) between the dashed shadow curve and the others (solid curves) are on a single line (*i.e.*, collinear). In this case, the shadow plane that includes the dashed shadow curve has indeterminacy even if all the other curves are determined. Thus, it is a degenerate condition and the total indeterminacy is more than 4 DOFs.

Figure 3(b) shows another example, where the set of shadow curves are divided into either solid curves or dashed curves, and "all" the intersection points (circles in the figure) between the solid curves and the dashed curves are collinear on the image plane. In this case, the shapes of the dashed shadow curves have indeterminacy, even if all the shapes of the real shadow curves are determined, and the total indeterminacy is more than 4 DOFs.

4.2.2 Ambiguity of the Solution and GPBR Transformations

To conclude this section, we point out that the 4-DOF indeterminacy of the general solution of the planes can be described as 4-parameter 3D homographies that transform the 3D points and the plane parameters. Let the plane parameter $\mathbf{a}_j^{*\top}$ be described by 3D homogeneous coordinates $[\mathbf{a}_j^{*\top} 1]^\top$, and 4×4 matrix $\mathbf{B}(\mathbf{b}, s)$ be defined as

$$\mathbf{B}(\mathbf{b}, s) \equiv \begin{bmatrix} s\mathbf{I} & \mathbf{0} \\ \mathbf{b}^\top & 1 \end{bmatrix}. \tag{12}$$

Because of the indeterminacy described as form (10), if a set of planes $[\mathbf{a}_j^{*\top} 1]^\top$, $1 \le j \le N$ is a solution, then the set of planes

$$[\mathbf{a}_j^{*\top} 1]\mathbf{B}(\mathbf{b}, s), \quad 1 \le j \le N \tag{13}$$

is also a solution.

The plane parameter $\mathbf{a}_j^* \equiv \mathbf{K}^{-\top} \mathbf{a}_j$ can be regarded as a projective plane that is transformed by the intrinsic parameters of the camera. Let a point on the projective plane $\mathbf{a}_j^*$ be $\mathbf{x}^*$, then, $\mathbf{a}_j^{*\top} \mathbf{x}^* = 0$. Since the inverse of $\{\mathbf{B}(\mathbf{b}, s)\}$ is $\{\mathbf{B}(\mathbf{b}, s)\}^{-1} = \mathbf{B}(-(1/s)\mathbf{b}, 1/s)$,

$$[\mathbf{a}_j^{*\top} 1] \begin{bmatrix} \mathbf{x}^* \\ 1 \end{bmatrix} = 0$$

$$\implies \quad \left\{ [\mathbf{a}_j^{*\top} 1]\mathbf{B}(\mathbf{b}, s) \right\} \left\{ \mathbf{B}(-(1/s)\mathbf{b}, 1/s) \begin{bmatrix} \mathbf{x}^* \\ 1 \end{bmatrix} \right\} = 0 \tag{14}$$

holds. These forms mean that, if a plane parameter is transformed by form (13), points on the plane $\mathbf{x}^*$ are transformed by the homography $\mathbf{B}(-(1/s)\mathbf{b}, 1/s)$. This form represents a group of 4-parameter homographies that transform 3D points $\mathbf{x}^*$ and are termed generalized projective bas-relief (GPBR) transformations (Kriegman and Belhumeur 2001). It has already been pointed out in the reference (Kawasaki and Furukawa 2007) that solutions of shape reconstruction using line-lasers or shadows of straight edges inevitably have 4-DOF indeterminacy unless some metric constraints are used.

The indeterminacy caused by GPBR transformations and the indeterminacy caused by unknown camera parameters (elements of $\mathbf{K}$) are independent of each other. The indeterminacy caused by GPBR transformations corresponds to the 4-DOF family of 3D homographies by which all the lines of sight from the camera are fixed while the depth values along the lines of sight are changed. Therefore, this family includes scaling transformations. In contrast, the indeterminacy caused by unknown camera parameters corresponds to the 5-DOF family of 3D homographies by which the lines of sight from the camera are moved while the optical center of the camera and the focal plane are fixed.

The combined 3D homographies of the two groups can be represented as a 9-DOF matrix

$$\begin{bmatrix} s\mathbf{K}^{-1} & \mathbf{0} \\ \mathbf{b}^\top & 1 \end{bmatrix}. \tag{15}$$

This matrix is equivalent to general 3D homographies with constraints such that the origin, x-axis, and xy-plane are fixed.

4.3 Metric Reconstruction

As described in the previous section, the solutions of form (11) have indeterminacy of 4 DOFs. Moreover, if any

of the intrinsic parameters are unknown, $\mathbf{a}_j^*$, $j = 1, \ldots, N$ include variables of the parameters which cause additional indeterminacy.

These indeterminacies cannot be reduced by using coplanarities. However, depending on the situations or scenes, additional geometric constraints, which we already named "metric constraints," are available, and we may reduce the indeterminacies by using these constraints. For example, if only the focal length is unknown within the intrinsic parameters, we need at least 4 metric constraints to reconstruct the scene up to scale, since (number of conditions required) = 4 + (number of unknown intrinsic parameters) − (number of parameters that may remain) = $4 + 1 - 1 = 4$.[1]

4.3.1 Linear Method

If we have many (at least 9) orthogonalities between planes as metric constraints, it is possible to achieve metric reconstruction by estimating all the intrinsic parameters using a linear method. From the definition in Sect. 4.2, $\mathbf{a}_j^{*\top} = \mathbf{a}_j^\top \mathbf{K}^{-1}$. Thus, the plane parameter $\mathbf{a}_j^\top$ in the Euclidean space can be represented as

$$\mathbf{a}_j^\top = \mathbf{a}_j^{*\top} \mathbf{K} = (s\mathbf{a}_j' + \mathbf{b})^\top \mathbf{K}, \quad j = 1, \ldots, N \quad (16)$$

by using an arbitrary scalar s and an arbitrary 3D vector $\mathbf{b}$. The scalar s corresponds to scaling in the Euclidean space, which cannot be determined, in principle, from only the orthogonalities of planes. So, $s = 1$ is assumed. Then, the condition of orthogonality between π_j and π_k can be represented as

$$\mathbf{a}_j^\top \mathbf{a}_k = (\mathbf{a}_j' + \mathbf{b})^\top \mathbf{K} \mathbf{K}^\top (\mathbf{a}_k' + \mathbf{b})$$

$$= [1 \ \mathbf{a}_j'^\top] \begin{bmatrix} \mathbf{b}^\top \mathbf{K} \\ \mathbf{K} \end{bmatrix} \begin{bmatrix} \mathbf{K}^\top \mathbf{b} \ \mathbf{K}^\top \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{a}_k' \end{bmatrix}$$

$$= [1 \ \mathbf{a}_j'^\top] \mathbf{Q} \begin{bmatrix} 1 \\ \mathbf{a}_k' \end{bmatrix} = 0, \quad (17)$$

where $\mathbf{Q}$ is a 4×4 symmetric matrix defined as

$$\mathbf{Q} \equiv \begin{bmatrix} \mathbf{b}^\top \mathbf{K} \\ \mathbf{K} \end{bmatrix} \begin{bmatrix} \mathbf{K}^\top \mathbf{b} \ \mathbf{K}^\top \end{bmatrix} = \begin{bmatrix} \mathbf{b}^\top \mathbf{K} \mathbf{K}^\top \mathbf{b} & \mathbf{b}^\top \mathbf{K} \mathbf{K}^\top \\ \mathbf{K} \mathbf{K}^\top \mathbf{b} & \mathbf{K} \mathbf{K}^\top \end{bmatrix}. \quad (18)$$

Form (17) can be regarded as a linear equation with respect to the elements of $\mathbf{Q}$. Considering that $\mathbf{Q}$ is a symmetric matrix and that the right-bottom element of $\mathbf{Q}$ is always 1, 9 elements should be determined. If we have 9 independent conditions of orthogonalities, we can determine $\mathbf{Q}$. Since the right-bottom 3×3 submatrix of $\mathbf{Q}$ is $\mathbf{K} \mathbf{K}^\top$, $\mathbf{K}$ can be determined by Cholesky factorization of the submatrix. Once $\mathbf{K}$ is determined, $\mathbf{b}$ can be calculated by $\mathbf{b} = (\mathbf{K} \mathbf{K}^\top)^{-1} \mathbf{q}$, where $\mathbf{q}$ is the left-bottom 3×1 submatrix of $\mathbf{Q}$.

---

[1]The types of indeterminacies that are reducible depend on the types of metric constraints. For example, indeterminacy of scaling is not reducible using orthogonalities or parallelisms in the scene.

4.3.2 Nonlinear Method

The linear method described in Sect. 4.3.1 can be applicable only if at least 9 orthogonalities in the target scene are available. Even if this is true, these conditions are often dependent or nearly dependent on each other and thus the linear method tends to be unstable.

In real situations, some or all of the intrinsic parameters may be known. In such cases, it is often effective to achieve metric reconstruction by minimization of an error function that is defined to have a minimum value if the metric constraints are fulfilled. The plane parameter $\mathbf{a}_j$ in the Euclidean space can be described by using a set of unknown intrinsic parameters, represented as Θ, and variables s and $\mathbf{b}$ of form (11). Let this representation be described as $\mathbf{a}_j(\Theta, s, \mathbf{b})$. Then, for example, if we have a set of conditions of orthogonalities from the scene, which is represented as a set of pairs of indices of planes that are orthogonal $C_v = \{(i, j) \mid (\pi_i \perp \pi_j)\}$, we can define an error function $E(\mathbf{b}, \Theta)$ as

$$E(\mathbf{b}, \Theta) \equiv \sum_{(i,j) \in C_v} \cos^2 \theta_{i,j}(\mathbf{b}, 1, \Theta)$$

$$= \sum_{(i,j) \in C_v} \{N(\mathbf{a}_i(\mathbf{b}, 1, \Theta))^\top N(\mathbf{a}_j(\mathbf{b}, 1, \Theta))\}^2, \quad (19)$$

where $\theta_{i,j}(\Theta, s, \mathbf{b})$ is the angle between the planes π_i and π_j, and $N()$ refers to normalization of a vector. $E(\mathbf{b}, \Theta)$ does not have s as an argument because the scaling cannot be determined from only orthogonalities. Using the error function, we can obtain estimations of $\mathbf{b}$ and Θ, which are depicted as $\hat{\mathbf{b}}$ and $\hat{\Theta}$, by solving the nonlinear optimization problem of

$$(\hat{\mathbf{b}}, \hat{\Theta}) = \arg \min_{(\mathbf{b}, \Theta)} E(\mathbf{b}, \Theta). \quad (20)$$

Achieving metric reconstruction using nonlinear optimization has an advantage in that the types of usable metric constraints can be easily increased by representing new constraints as error functions. For example, if parallelisms of planes are available as a set $C_p = \{(i, j) \mid \pi_i \text{ and } \pi_j \text{ are parallel}\}$ in addition to the orthogonalities of planes, the error function can be defined as

$$E(\mathbf{b}, \Theta) \equiv \sum_{(i,j) \in C_v} \cos^2 \theta_{i,j}(\mathbf{b}, 1, \Theta)$$

$$+ \sum_{(i,j) \in C_p} \sin^2 \theta_{i,j}(\mathbf{b}, 1, \Theta)$$

$$= \sum_{(i,j) \in C_v} \{N(\mathbf{a}_i(\mathbf{b}, 1, \Theta))^\top N(\mathbf{a}_j(\mathbf{b}, 1, \Theta))\}^2$$

$$+ \sum_{(i,j) \in C_p} \|N(\mathbf{a}_i(\mathbf{b}, 1, \Theta)) \times N(\mathbf{a}_j(\mathbf{b}, 1, \Theta))\|^2. \quad (21)$$

Other types of metric constraints can be used as long as they can be represented as an equation described by using $\hat{\mathbf{b}}$, s, $\hat{\Theta}$, or constant values.

Many methods have been proposed for nonlinear optimization. We used an implementation of the Levenberg-Marquardt method in the experiments described later. For nonlinear optimization determining the initial values can sometimes be a problem. In the experiments, orthogonal projections of planes from the manually determined parameters into the solution spaces of form (11), which can be represented as $3N$-dimensional linear spaces, were calculated, and the projections were used as the initial solutions.

4.4 Dense Reconstruction

The above mentioned method is only applied to several selected curves, and results in a sparse 3D reconstruction. Therefore, to achieve a dense 3D reconstruction using all the captured frames, we first estimate the parameters of each implicit plane, and then reconstruct all the 3D points on the implicit-coplanar curves by applying a simple triangulation method. This triangulation method can be conducted by calculating the intersection between the implicit plane and the line of sight which goes through the point on a detected shadow boundary as shown in Fig. 4. The actual process of the dense reconstruction is as follows.

1. Up to this point, since only the plane parameters of the selected planes are estimated using the aforementioned method, recover all the 3D positions of the points on the coplanar curves (*i.e.*, shadow boundary) by using the estimated plane parameters.
2. Detect the intersections between an implicit-coplanar curve on an arbitrary frame and the implicit-coplanar curves on which 3D positions have already been estimated in step 1.
3. Estimate the plane parameters of the implicit-coplanar curve on an arbitrary frame by fitting the intersection
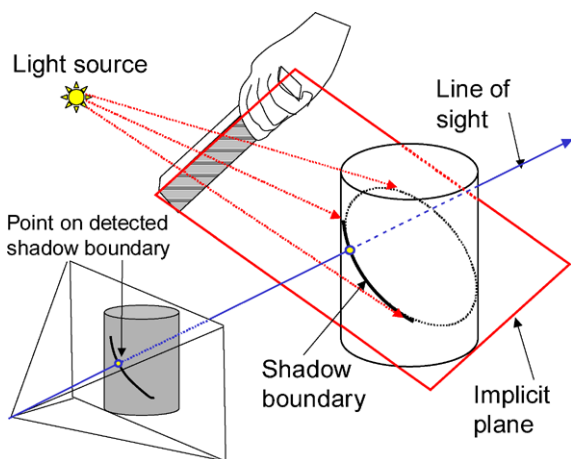

Fig. 4 Triangulation method

points to a plane by using principal components analysis (PCA).

4. Recover all the 3D positions of the points on the implicit-coplanar curve by using the estimated plane parameters from step 3.
5. Iterate steps 2 to 4 for all frames until all implicit-coplanar curves are reconstructed.

4.5 Optimality and Errors

In this section, we discuss the optimality and meaning of solving (6) using the proposed method. For the discussion in this section, we use fixed camera parameters of a normalized camera ($\mathbf{K} = \mathbf{I}$).

If there are no errors in the positions of points $\mathbf{u}_i$, (6) should be strictly fulfilled. However, in real situations, there are errors in the coordinates of $\mathbf{u}_i$, and (6) may not be strictly fulfilled. In such a case, obtaining the general solution of form (11) by using SVD of the matrix $\mathbf{L}'$ is equivalent to obtaining the solution that minimizes the sum of the squares of the left side of (6) over all points $\mathbf{u}_i$.

Several geometric interpretations are possible for the solutions obtained in this way. One interpretation regards them as solutions such that the sum of the square of the inverse depth (z-coordinate) errors are minimized over all the intersections, as introduced by Bouguet et al. (1999). Suppose that point $\mathbf{u}_i$ on the image plane is a projection of a 3D point ξ_i on the intersection of π_j and π_k. Let the line of sight defined by $\mathbf{u}_i$ be denoted as μ_i, and let the depth (z-coordinate) of the intersection point between μ_i and π_j be denoted as $z_i^{(j)}$. The depth the intersection between μ_i and plane π_j, which is represented by a 3D parameter vector $\mathbf{a}_j$ ($= \mathbf{a}_j^*$ since $\mathbf{K} = \mathbf{I}$), can be calculated as $z_i^{(j)} = 1/(\tilde{\mathbf{u}}_i^\top \mathbf{a}_j)$. If there are no errors, $z_i^{(j)}$ (depth of the intersection between π_j and μ_i) and $z_i^{(k)}$ (depth of the intersection between π_j and μ_i) should coincide. However, they do not match because of errors. The difference of the inverse of the depths is

$$\frac{1}{z_i^{(j)}} - \frac{1}{z_i^{(k)}} = \tilde{\mathbf{u}}_i^\top \mathbf{a}_j - \tilde{\mathbf{u}}_i^\top \mathbf{a}_k = \mathbf{u}_i^\top (\mathbf{a}_j - \mathbf{a}_k) = \mathbf{u}_i^\top (\mathbf{a}_j^* - \mathbf{a}_k^*). \tag{22}$$

Thus, the difference is equal to the left side of (6), and the proposed method is optimum in the sense that the sum of the squares of the inverse depth is minimized.

There is another interpretation of the left side of (6). As described in Sect. 4.2, the 3D vector $\mathbf{a}_j^* - \mathbf{a}_k^*$ represents the image of line ϕ_i, which is the intersection line between π_j and π_k. The vector can be written using the elements as $\mathbf{a}_j^* - \mathbf{a}_k^* = [a_j^* - a_k^*, b_j^* - b_k^*, c_j^* - c_k^*]^\top$. On the image plane, the

distance between the point $\mathbf{u}_i = [u_i, v_i]^\top$ and the line $(\mathbf{a}_j^* - \mathbf{a}_k^*)$ can be calculated by

$$\frac{(a_j^* - a_k^*)u_i + (b_j^* - b_k^*)v_i + (c_j - c_k)}{\sqrt{(a_j^* - a_k^*)^2 + (b_j^* - b_k^*)^2}}. \tag{23}$$

The numerator of form (23) is the same as the left side of (6). Thus, the proposed method minimizes the sum of the "weighted" squared distances between the images of lines ϕ_i and points $\mathbf{u}_i$, with the weight $\{(a_j^* - a_k^*)^2 + (b_j^* - b_k^*)^2\}^{-1/2}$.

Equation (6) is derived from the constraint that the line ϕ_i goes through the point $\mathbf{u}_i$ on the image plane. In real cases, observation errors are caused by noise in the positions of points $\mathbf{u}_i$. If the distributions of the errors of the positions are isotropic and identical Gaussian distributions, solving problems by minimizing the sum of the squared distances between points $\mathbf{u}_i$ and lines ϕ_i achieves the maximum likelihood estimation. This can be realized by first obtaining metric reconstruction using (6) as it is, and then re-processing the proposed method using the modified matrix of $\mathbf{L}$ each of whose rows is divided by $\{(a_j^* - a_k^*)^2 + (b_j^* - b_k^*)^2\}^{1/2}$. Using this technique, all the weights of (6) for all the points are equalized, and the maximum likelihood estimation under isotropic and identical Gaussian distributions of $\mathbf{u}_i$ can be achieved.
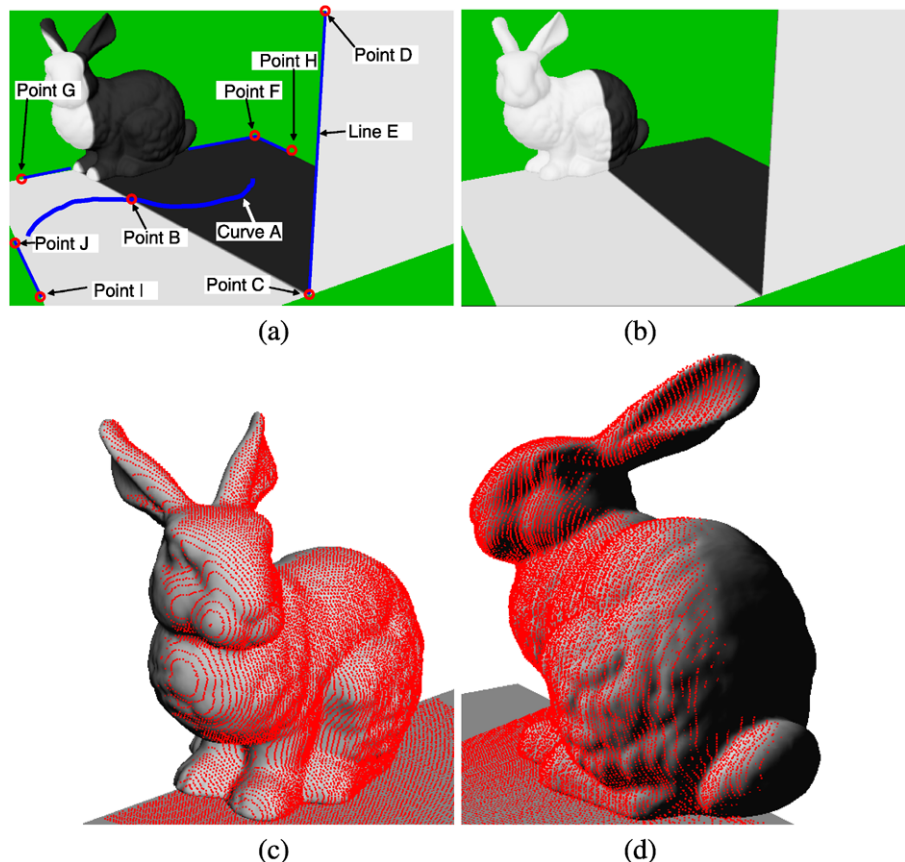
5 Experiments

5.1 Simulation Data (Shadow of Static Object)

Figures 5(a) and (b) show data synthesized by CG and include a square floor, a rabbit, and a perpendicular wall. While moving the light source, 160 images were generated, so that the edge of the shadow scanned the rabbit. The angle of view of the virtual camera was 30°, and the image size was 600 × 400 pixels. The parameters of the camera were the same for all the synthetic data used in this section. Although the implicit-coplanar curves of the shadows did not have intersections between each other, we were able to extract 160 intersections (point B in Fig. 5(a)) between the explicit-coplanar curve drawn on the floor (curve A) and the implicit-coplanar curves of the shadows. From the intersections, 160 linear equations were obtained. In addition, 160 linear equations and 159 linear equations were obtained from points C and D, respectively. The reason why the equations obtained from point D were less than those from point C by one is that point D is not included on the plane of the floor. As a result, 479 equations were obtained for 483 plane parameters of 161 planes, and the solution had 4-DOF indeterminacy.

Metric reconstruction was performed up to scale using the nonlinear method described in Sect. 4.3.2. The initial

Fig. 5 Reconstruction of simulation data (using nonlinear optimization): (**a**) An original image. Curve A is a curve for extracting intersections between planes; points B, C and D are points on intersections between planes, line E is an edge that generates shadow edges, and angle $\angle GFH$ and the angle between line segments IJ and GF are used as metric constraints. (**b**) Input images with shadows. (**c**), (**d**) Reconstruction results. In the figure, the shaded surfaces are ground truth and the points are reconstructed results

value of the nonlinear optimization was given by indicating whether the light source was located on the right or left. In this experiment, we also estimated the focal length, and thus needed four metric constraints. We used two metric constraints from the orthogonalities of the implicit planes of shadows with the floor, and another two constraints from the orthogonalities of the two corners of the floor (*i.e.*, the orthogonality of angle $\angle HFG$ and the orthogonality between line segments IJ and GF). Figures 5(c) and (d) show the result (points) and the ground truth (shaded surface). We can observe that the reconstruction result almost coincides with the correct shape. The root mean square (RMS) error of the z-coordinates of all the reconstructed points was 2.6×10^{-3}, where the average distance from the camera to the rabbit was scaled to 1.0. Thus, a highly accurate reconstruction of the technique was confirmed.

5.2 Simulation Data (Active Scan by Cast Shadow)

Next, we reconstructed 3D shapes by sweeping the cast shadows on the objects by moving both a light source and a straight object. We synthesized images that included 20 implicit coplanarities and three visible planes (*i.e.*, explicit planes). The metric reconstruction was performed up to scale using the nonlinear method. There were three metric constraints of orthogonalities and parallelisms between the three explicit planes visible in the scene. Figure 6(a) shows an example of the synthesized images, and Fig. 6(b) shows all the implicit-coplanar curves as borders of the grid patterns. Figures 6(c) and (d) show the results. The scaling of
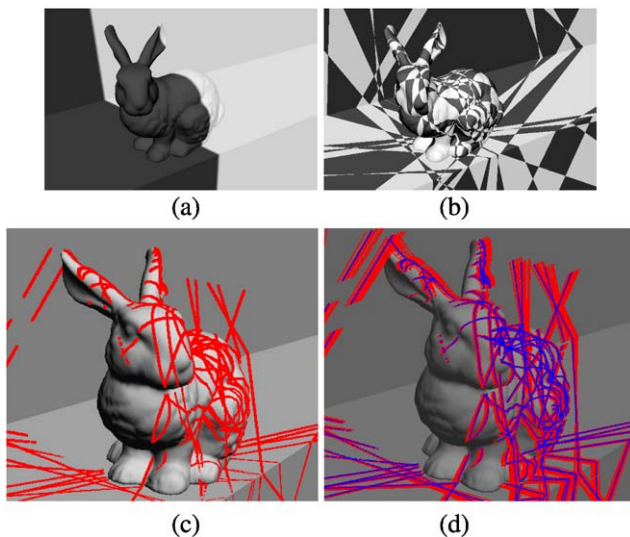
Fig. 6 Reconstruction of simulation data (active scanning): (**a**) Input image. (**b**) Explicit and implicit coplanarities. (**c**) The reconstruction result (curves) and the ground truth (shaded surfaces). (**d**) Reconstruction results disturbed by noise. *Dark lines* (*blue lines* in online version) are the ground truth and *bright lines* (*red lines* in online version) are reconstructed results of disturbed inputs

the results was adjusted so that the average depth was normalized to 1.0. The RMS error of the z-coordinates of all the reconstructed points was 4.6×10^{-3}. Thus, we can confirm the high accuracy of the results.

We also conducted an experiment to evaluate the depth accuracy when noise was added to the detected positions of the intersection points. We made 10 test inputs with independent noise patterns by adding noise to the original inputs of Fig. 6. The noise was generated independently with the normal distribution $N(0, 1)$ in pixels. Reconstruction was performed using the nonlinear method with known intrinsic parameters. Figure 6(d) shows that the reconstruction results and the RMS error of the depth values from the ground truth normalized by the average distance was 1.3×10^{-2}.

5.3 Simulation Data (Metric Reconstruction Using the Linear Method)

Next, an experiment of simultaneously reconstructing the scene and all the intrinsic parameters using the linear method described in Sect. 4.3.1 was conducted using simulated data. Since many orthogonalities are required for this method, three rectangular boxes were located in the scene. To enhance the independences of (17), two of the boxes were rotated so that they were not aligned with each other. We assumed that the scene included two poles that were vertical to the floor, and the shadows of the poles for various light directions were calculated in the simulation. Figure 7(a) shows one of the input images. Since no shadow curve passes through face I in Fig. 7(b), we added points on the intersection between faces G and I, and on the intersection between G and H (black dots in Fig. 7(b)). Face L was also handled similarly. Figure 7(c) shows the locations of the points used in the coplanarity conditions. The metric conditions used were orthogonalities between the following pairs of faces, A–B, A–C, A–D, A–E, B–C, C–D, D–E, B–F, C–F, G–H, H–I, G–I, J–K, K–L, and J–L, and orthogonalities between face A (floor) and all the shadow planes. Figure 7(d) shows the reconstructed shape and Table 1 shows the estimated intrinsic parameters. These results indicate that appropriate estimations were obtained for the shape of the scene and the intrinsic parameters.

5.4 Real Outdoor Scene (Shadow of Static Object)

We conducted a shape reconstruction from images acquired by an outdoor uncalibrated fixed camera. Images from a fixed outdoor camera were captured for two hours in the morning with periodic intervals of 20 seconds, and the shape and focal length of the camera were reconstructed. Since the scene also contained many shadows generated by non-straight edges, automatic extraction of complete shadows was difficult. In this experiment, these noises were eliminated manually. The metric reconstruction was processed up

Fig. 7 Reconstruction of
simulation data (estimating all
intrinsic parameters): (**a**) The
target scene. (**b**) Cast shadows
in the scene. (**c**) Points used to
describe the coplanarity
constraints. (**d**) Reconstructed
points
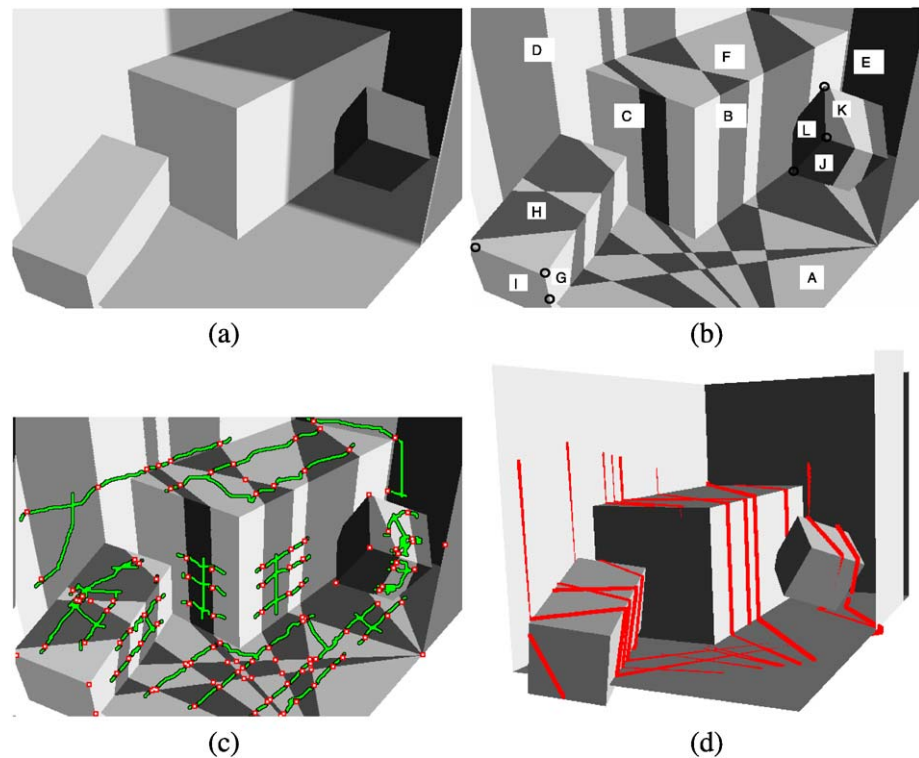
Table 1 Estimation of intrinsic parameters from simulated data

	α	β	u_p	v_p	σ
The true values	746.3	746.3	300.0	200.0	0.0
The estimated values	731.8	744.4	295.4	199.5	4.622

to scale using the nonlinear method. For metric constraints, we used the orthogonality of the pole with the ground, the orthogonality of the wall with the ground, and the parallelism of the bench (seat) and the ground. Figure 8(a) shows one of the input frames, (b) shows the detected coplanar shadow curves of (a), (c) shows all the coplanar curves and their intersections of (a), and (d) through (f) show the reconstruction result. The proposed technique could correctly reconstruct the scene by using images from a fixed remote camera.

5.5 Real Indoor Scene (Active Scan by Cast Shadow)

We conducted an indoor experiment of an actual scene by using a point light source. A video camera was directed toward a target object and multiple boxes, and the scene was captured while the light source and the bar for shadowing were moved freely. From the captured image sequence, several images were selected and the shadow curves of the bar were detected from the images. By using the detected coplanar shadow curves, we performed 3D reconstruction up to

4-DOF indeterminacy. The metric reconstruction was performed up to scale using the nonlinear method. For the metric reconstruction, orthogonalities of the faces of the boxes were used. Figure 9 shows the captured scenes and the reconstruction result. In this case, since there were only small noises extracted because of the indoor environment, shadow detection was stable and no human interaction was required. These results show that the dense shape was correctly reconstructed.

We also reconstructed a scene of a box (dimensions: 0.4 m × 0.3 m × 0.3 m) and a cylinder (height: 0.2 m, diameter: 0.2 m) to evaluate the accuracies of the proposed method. The process of reconstruction was conducted in the same way as in the previous experiment, except that we also measured the 3D scene by an active measurement method using coded structured light (Sato and Inokuchi 1987) as the ground truth. The reconstruction result was scaled to match the ground truth using the average distance to the points. Figures 10(a) and (b) show the captured scenes, (c) shows the coplanar curves and their intersections, and (d) through (f) show both the scaled reconstruction (shaded polygon mesh) and the ground truth (point cloud). Although there were small differences between the reconstruction and the ground truth, the shape was correctly recovered. The RMS error of the reconstruction from the ground truth normalized by the average distance was 1.80×10^{-2}.

Fig. 8 Reconstruction of a real outdoor scene: (**a**) Input image. (**b**) Example frame of the 3D segmentation result. (**c**) Implicit (*bright gray*; *green* in online version) and explicit (*dark gray*; *red* in online version) coplanar curves. (**d**) Reconstructed result of coplanar curves (curved lines) and dense 3D points (shaded surfaces). (**e**), (**f**) The textured reconstructed scenes
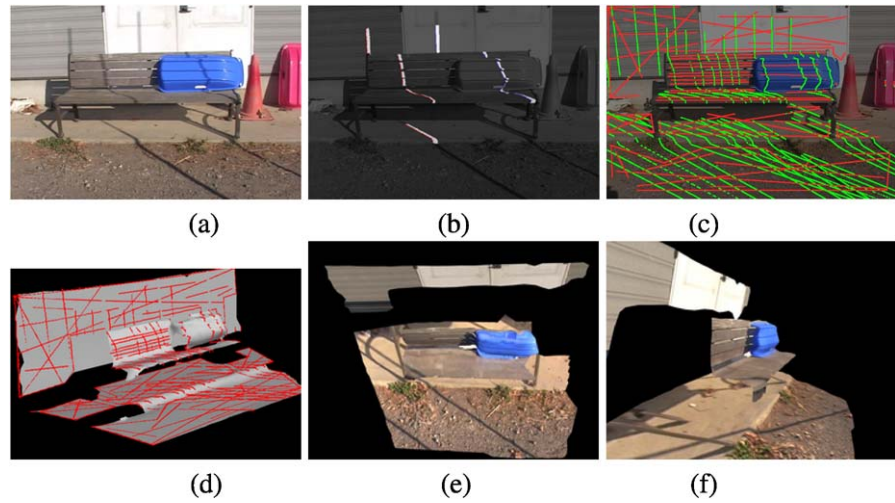
Fig. 9 Reconstruction of a real indoor scene: (**a**), (**b**) Captured frames. (**c**) Implicit and explicit coplanar curves and their intersection points. (**d**) Reconstructed coplanar shadow curves (curved lines) with dense reconstructed model (shaded surfaces). (**e**), (**f**) The textured reconstructed model
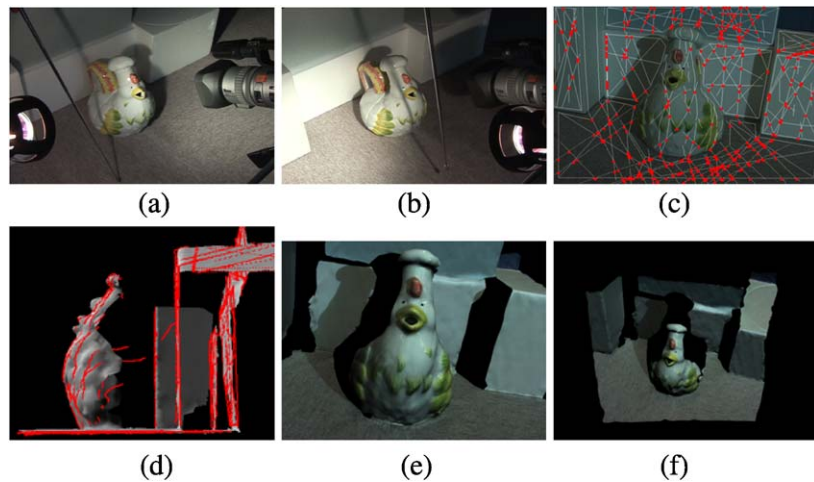
Fig. 10 Reconstruction and evaluation of a real indoor scene: (**a**), (**b**) Captured frames. (**c**) Implicit and explicit coplanar curves and their intersection points. (**d**) Reconstructed model. (**e**), (**f**) Reconstructed model displayed with the ground truth data (shaded model: reconstructed model, point cloud: ground truth)
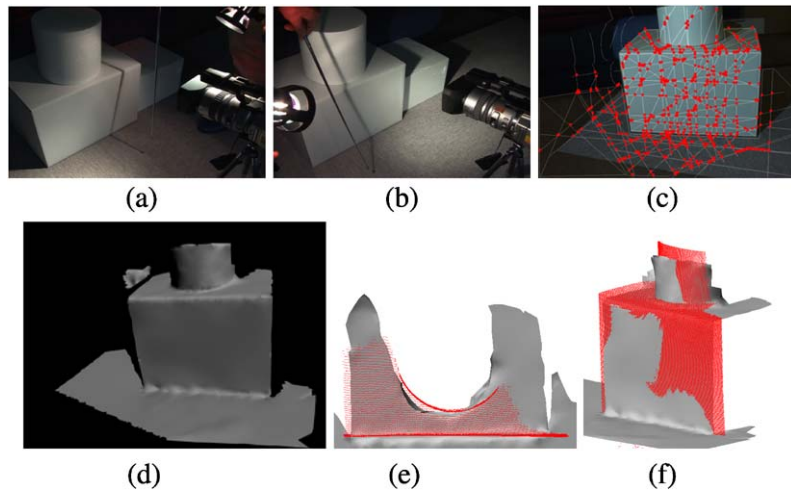
6 Conclusion

This paper proposed a technique capable of reconstructing a shape if only multiple shadows of straight linear objects or straight edges are available from a scene, even when the light source position is unknown and the camera is not calibrated. This technique can be understood as an extension of the conventional method, which reconstructs a polyhe-

dron from its visible edges, to general curved surfaces. Since reconstruction from coplanarities can be solved only up to projective reconstruction that has at least 4-DOF indeterminacy, we proposed a technique to upgrade it to metric reconstruction by adding metric constraints. Two types of methods to achieve metric reconstruction were proposed: a linear method and a nonlinear method. We conducted experiments using these methods and showed that the 3D shapes of the scenes were successfully reconstructed. For the stable extraction of shadow areas from a scene, we developed a spatio-temporal image processing technique. By implementing the technique and conducting an experiment using simulated and real images, accurate and dense shape reconstruction was verified.

References

Bartoli, A., & Sturm, P. (2003). Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal of Computer Vision*, *52*(1), 45–64.

Bouguet, J. Y., & Perona, P. (1998). 3D Photography on your desk. In *International conference on computer vision* (pp. 129–149).

Bouguet, J.-Y., Weber, M., & Perona, P. (1999). What do planar shadows tell about scene geometry? In *CVPR* (Vol. 01, pp. 514–520).

Caspi, Y., & Werman, M. (2006). Vertical parallax from moving shadows. In *CVPR* (pp. 2309–2315), Washington, DC, USA. Los Alamitos: IEEE Comput. Soc.

Criminisi, A., Reid, I. D., & Zisserman, A. (2000). Single view metrology. *International Journal of Computer Vision*, *40*(2), 123–148.

Daum, M., & Dudek, G. (1998). On 3-D surface reconstruction using shape from shadows. In *CVPR* (pp. 461–468).

Hambrick, L. N., Loew, M. H., & Carroll, J. R. L. (1987). The entry exit method of shadow boundary segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *9*(5), 597–607.

Hatzitheodorou, M., & Kender, J. (1988). An optimal algorithm for the derivation of shape from shadows. In *CVPR* (pp. 486–491).

Jiang, C., & Ward, M. O. (1994). Shadow segmentation and classification in a constrained environment. *Computer Vision, Graphics, and Image Processing, Image Understanding*, *59*(2), 213–225.

Kawasaki, H., & Furukawa, R. (2007). Dense 3D reconstruction method using coplanarities and metric constraints for line laser scanning. In *International conference on 3D digital imaging and modeling*.

Kriegman, D. J., & Belhumeur, P. N. (2001). What shadows reveal about object structure. *Journal of the Optical Society of America*, *18*(8), 1804–1813.

Raviv, D., Pao, Y., & Loparo, K. A. (1989). Reconstruction of three-dimensional surfaces from two-dimensional binary images. *IEEE Transactions on Robotics and Automation*, *5*(5), 701–710.

Salvador, E., Cavallaro, A., & Ebrahimi, T. (2004). Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, *95*(2), 238–259.

Sato, K., & Inokuchi, S. (1987). Range-imaging system utilizing nematic liquid crystal mask. In *Proc. of first international conference on computer vision* (pp. 657–661).

Savarese, S., Andreetto, M., Rushmeier, H., Bernardini, F., & Perona, P. (2007). 3D reconstruction by shadow carving: theory and practical evaluation. *International Journal of Computer Vision*, *71*(3), 305–336.

Shafer, S. A., & Kanade, T. (1983). Using shadows in finding surface orientations. *Computer Vision, Graphics, and Image Processing*, *22*(1), 145–176.

Sugihara, K. (1986). *Machine interpretation of line drawings*. Cambridge: MIT Press.

Yamazaki, S., Narasimhan, S. G., Baker, S., & Kanade, T. (2007). Coplanar shadowgrams for acquiring visual hulls of intricate objects. In *International conference on computer vision*.

Yu, Y., & Chang, J. T. (2005). Shadow graphs and 3D texture reconstruction. *International Journal of Computer Vision*, *62*(1-2), 35–60.