International Journal of Computer Vision manuscript No.
(will be inserted by the editor)

Light structure from pin motion: Geometric point light source calibration

Hiroaki Santo · Michael Waechter · Wen-Yan Lin · Yusuke Sugano · Yasuyuki Matsushita

the date of receipt and acceptance should be inserted later

Abstract We present a method for geometric point light source calibration. Unlike prior works that use Lambertian spheres, mirror spheres, or mirror planes, we use a calibration target consisting of a plane and small shadow casters at unknown positions above the plane. We show that shadow observations from a moving calibration target under a fixed light follow the principles of pinhole camera geometry and epipolar geometry, allowing joint recovery of the light position *and* 3D shadow caster positions, equivalent to how conventional structure from motion jointly recovers camera parameters and 3D feature positions from observed 2D features. Moreover, we devised a unified light model that works with nearby point lights as well as distant light in one common framework. Our evaluation shows that our method yields light estimates that are stable and more accurate than existing techniques while having a much simpler setup and requiring less manual labor.

Keywords Light source calibration · photometric stereo · structure from motion

1 Introduction

Accurately estimating the position or direction of a light source is essential for many physics-based computer vision tasks, such as shape from shading [19], photometric stereo [37,47], or reflectance and material estimation [14]. In these tasks, inaccurate light positions cause errors. For example, Fig. 2 shows the relation between light calibration error and surface normal estimation error in a synthetic experiment with a directional light, a Lambertian sphere as target object, and a basic photometric stereo method [37,47]. We can clearly see the importance of accurate light calibration. Ideally, the error of a calibration method is so small that developers of physics-based modeling algorithms never need consider it. Although there are approaches to refine inaccurate light calibration [30] or bypass calibration altogether (uncalibrated photometric stereo [2, 36,8]), they do not make highly accurate calibration obsolete. Uncalibrated photometric stereo cannot overcome the generalized bas-relief ambiguity for Lambertian materials and even in favorable settings they do

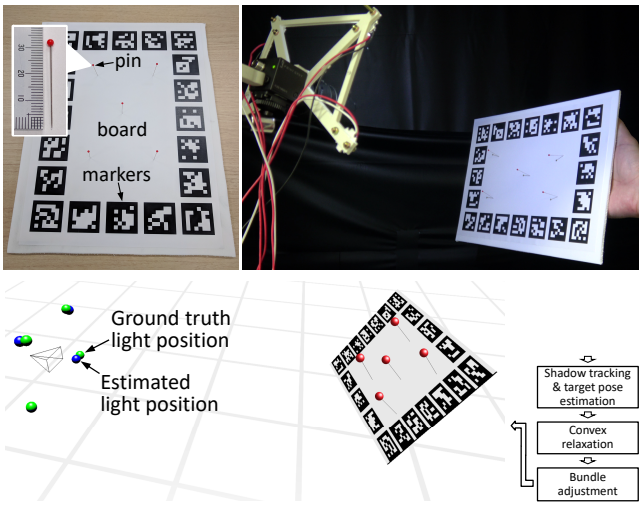

Fig. 1 *Clockwise from top left:* Our calibration target, a camera observing the movement of shadows cast by a point light while the target is moved, our algorithm's workflow, and the estimation result.

H. Santo, M. Waechter, W.-Y. Lin, Y. Matsushita
Osaka University, Grad. School of Inf. Sci. and Technology
E-mail: {santo.hiroaki,waechter.michael,lin.daniel,yasumat}@ist.osaka-u.ac.jp

Y. Sugano
The University of Tokyo, Institute of Industrial Science
E-mail: sugano@iis.u-tokyo.ac.jp

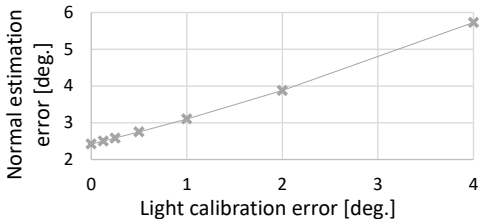

Fig. 2 Light calibration error vs. normal estimation error in photometric stereo. Each data point is the average of 100 independent runs.

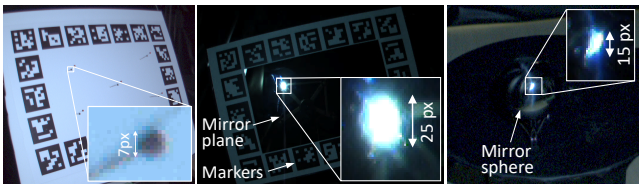

Fig. 3 *Left:* The pin head shadows on our planar target. *Center and right:* Specular highlights on a mirror plane [35] and a mirror sphere.

not reach the accuracy of accurate calibration. Despite the importance of accurate light calibration, it remains laborious as researchers have not yet come up with accurate *and* easy to use techniques.

This paper proposes a method for calibrating both distant and near point lights. We introduce a calibration target, shown in Fig. 1, that can be made within 1–2 min from off-the-shelf items for less than five dollars. Instead of specular highlights on spheres, we use a planar board (shadow receiver) and pins (shadow casters) that cast small point shadows on the board. Moving the board around in front of a static camera and light source and observing the pin head shadows under various board poses lets us determine the light position/direction.

The reasons why we operate with shadows on a planar target rather than with specular highlights or spherical targets are the following: A key factor in the overall calibration accuracy is the accuracy with which one can localize a calibration method's points of interest in the captured images. With the off-the-shelf pins that we use, we can automatically localize shadow centers with an accuracy of $\sim$1–2 px (Fig. 3, *left*), which is in marked contrast to how accurately we can detect specular highlights (Fig. 3, *center and right*). Moreover, our planar target translates small shadow localization errors only into small light direction errors. In contrast, mirror sphere methods amplify localization errors since the surface normal, which determines the light reflection angle, varies across the sphere.

From a geometric point of view, point lights are inverse pinhole cameras [20] (see Fig. 4). We can thus build upon past studies on multiview projective geometry. In particular, we show that shadows of static objects on a plane follow the principles of epipolar geometry. Further we show that, analogous to structure from motion (SfM) which jointly estimates camera poses and 3D point locations, we can jointly estimate light position/direction and shadow caster pin positions from moving our calibration target and observing the pin shadows, *i.e.*, we can estimate light and pins via *structure from* pin *motion*. Conveniently, this joint estimation of light and pins allows users to place the pins arbitrarily on the board and without needing to know their locations – in contrast to most previous works – our calibration target does not need to be carefully manufactured or measured.

To summarize, the primary contributions of our work are as follows. First, we show that shadow projection with a unified light model for both nearby and distant light follows the principles of pinhole camera projection and epipolar geometry. Second, using these principles we show how the joint estimation of light position/direction and 3D shadow caster positions based on shadow observations can be formulated as a bundle adjustment problem and we develop a robust solution technique for accurately achieving this estimation. Finally, we introduce a practical light source calibration method based on an easy-to-make calibration target. Instead of requiring a carefully designed calibration target, our method only uses needle pins that are stuck at unknown locations on a plane.

The benefits of the new calibration target and associated solution method are an extremely simple target construction process, a calibration process that requires no manual intervention other than moving the target since all required information is inferred automatically, and improved accuracy compared to prior work.

2 Related work

Light source calibration can be roughly divided into two tasks: geometric calibration and radiant intensity distribution (RID) calibration. This paper is solely concerned with geometric calibration but in this section, we introduce the prior works of both tasks and discuss the relationship to our work.

Geometric light source calibration: The goal of geometric light source calibration is to estimate

(a) light source directions in scenes with distant point light sources or
(b) light source positions in scenes with nearby point light sources.

In category (a), Zhang and Yang [48] (and Wei [45] with a more robust implementation) proposed a method to estimate multiple distant lights based on a Lambertian sphere's shadow boundaries and their intensities. Wang and Samaras [43] extended this to objects of arbitrary but known shape, combining information from the object's shading and the shadows cast on the scene. Zhou and Kambhamettu [49] estimated light directions from stereo images of a reference sphere with specular reflection. Cao and Shah [7] proposed a method for estimating camera parameters and the light direction from shadows of common vertical objects such as walls instead of special, precisely fabricated objects, which can be used for images from less restricted settings.

In category (b), Powell *et al.* [28] triangulated multiple light positions from highlights on three specular spheres at known positions. Other methods also used reflective spheres [1,15,33,41,46] or specially designed geometric objects [3,6,44]. Unlike these methods, some methods were based on planar mirrors [34,35]. They modeled the mirror by perspective projection and infer parameters similar to camera calibration. An interesting method, quite similar to ours in its simplicity and usage of shadows, is Bouguet and Perona's [4, Sec. 2.3]: They captured a pencil standing upright at multiple positions on a plane and triangulated all rays from pencil tip shadow to pencil tip. The core difference to our method is that it does not jointly estimate the calibration target (*i.e.*, pencil) with the light position.

In highlight-based geometric calibration methods, precisely localizing the light source center's reflection on the specular surface is problematic in practice: Even with the shortest exposure at which one can still barely detect or annotate other parts of the calibration target (pose detection markers, sphere outline, *etc.*), the highlight is much bigger than an image of the light source (such as a switched off LED seen in the mirror) would be; see Fig. 3, *center and right*. Lens flare, noise, *etc.* further complicate segmenting the highlight. Also, since the highlight is generally not a circle but a conic section on a mirror plane or an even more complicated shape on a mirror sphere, the light source center's image (*i.e.*, the intersection of the light cone's axis and the mirror) cannot be computed as the highlight's centroid, as for example Shen *et al.* [35] did. We thus argue that it is extremely hard to reliably localize light source centers on specular surfaces with pixel accuracy – even with careful manual annotation. Instead, we employ very small cast shadows for stable localization.

Mirror sphere-based geometric calibration methods suffer from the fact that the sphere curvature amplifies highlight localization errors into larger light direction errors since the surface normal, which determines the reflection angle, differs between erroneous and correct highlight location. Also, the spheres need to be very precise since "even slight geometric inaccuracies on the surface can lead to highlights that are offset by several pixels and markedly influence the stability of the results" (Ackermann *et al.* [1]). The prices of precise spheres ($\sim$ \$ 40 for a high-quality 60 mm bearing ball of which we need 3–8 for accurate calibration) rules out high-accuracy sphere-based calibration for users on a tight budget.

Further, sphere methods typically require accurate annotation of the sphere outline in the images. Although methods for automatic ellipse detection exist [26], accurately detecting the boundary of the mirror sphere is extremely difficult because the sphere's exact outline is hard to distinguish from the background, especially in dark images, since the sphere also mirrors the background.

Regarding the *triangulation* of multiple line constraints for the position of a light source, Hartley and Sturm [17] and Szeliski [40, Sec. 7.1] pointed out that, given noisy observations, reprojection error minimization is superior to finding the 3D point closest to each ray in a set of rays. The latter is a popular choice in many methods of category (b), for example Shen and Cheng's mirror plane method [35] or most sphere methods prior to Ackermann's [1]. By contrast, Ackermann *et al.* [1] and we follow Hartley and Sturm's suggestion and minimize reprojection error to obtain a more accurate prediction of light source positions.

The connection between pinhole cameras and point lights that we describe and exploit in the next section, has already been shown by others: Hu *et al.* [20] use it in a theoretical setting similar to ours with point objects and shadows. However, they do not turn it into a full mathematical formalism for the light estimation but only discuss it with geometric sketches (and suggest using the inferior triangulation method mentioned above).

We push the idea further by deriving mathematical solutions, extending it to a unified light model that includes distant light, embedding it in an SfM framework [38,42] that minimizes reprojection error, deriving an initialization for the non-convex minimization, devising a simple calibration target that leverages our method in the real world, and demonstrating our method's accuracy in simulated and real-world experiments.

Radiant intensity distribution calibration: While point light sources are often assumed to emit light uniformly in all directions, practical light sources such as LEDs actually have a non-isotropic lighting distribution (radiant intensity distribution; RID), which is described in

terms of light orientation, intensities, and an anisotropy function. Park *et al.* [25] and Ma *et al.* [22] handle non-isotropic lights and jointly estimate the light position and RID from imagery of shading and specular reflections on a planar calibration target. In the context of photometric stereo, Quéau *et al.* [29], Collins *et al.* [10], and Song *et al.* [39] proposed pipelines that first perform geometric calibration by specularity-based triangulation and then estimate the RID from shading on a planar target. Regarding the geometric calibration, these methods have shown to obtain a more accurate estimation compared to the plane-based methods of Park *et al.* [25] and Ma *et al.* [22]. Although this paper focuses on geometric calibration, we note that plane-based RID calibration could be combined with our method since it uses a similar planar target.

Another way to handle practical light sources in photometric stereo are "semi-calibrated" approaches [9, 21], which assume given light source positions/directions and estimate the non-uniform light intensities simultaneously with the scene shape. Their methods take care of part of RID, *i.e.*, the intensities, and assume isotropic lights or known lights anisotropy function.

More generally, incoming light may be not only from a single point source but also from a distribution of many points. Sato *et al.* [31,32] used shadows of an object of known shape to estimate illumination distributions of area lights while being restricted to distant light and having to estimate the shadow receiver's reflectance. Recently, Gardner *et al.* [12] proposed an estimation method of indoor illumination from a single image with a deep neural network. By limiting the scenes to indoor environments and training their model with a panorama image dataset, their method does not require any calibration target.

3 Shadow geometry

As foundation for the later sections, in this section we will lay out the mathematics behind shadow projection. Specifically, we analyze how a point light or distant, parallel light projects shadows of infinitesimal shadow casters on a shadow receiver plane. In Sec. 3.1, we will derive the shadow projection in an entirely static scene. In Sec. 3.2, we will then analyze shadows in a scene where the plane and the shadow casters remain static but the light source moves.

Throughout this paper, we will denote matrices and vectors with bold upper and lower case, respectively, and the homogeneous form of vector $\mathbf{v}$ with $\tilde{\mathbf{v}}$. Further, we will sometimes use parentheses and indices to refer to parts of a vector/matrix/tensor: $(\mathbf{v})_i$ denotes vector $\mathbf{v}$'s i^{th} element, $(\mathbf{L})_{i,j}$ is the element in row i and column j of matrix $\mathbf{L}$, $(\mathbf{L})_{i,:}$ refers to the entire row i of $\mathbf{L}$, and $(\mathbf{L})_{:,i:j}$ refers to all rows of columns i to j of $\mathbf{L}$.

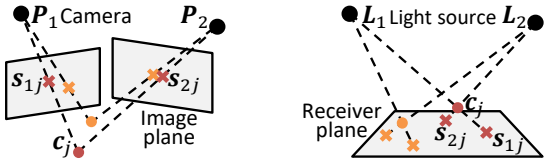

Fig. 4 Cameras vs. point lights. A camera matrix $\mathbf{P}_i$ projects a scene point $\mathbf{c}_j$ to an image point $\mathbf{s}_{ij}$ just like a light matrix $\mathbf{L}_i$ projects a scene point $\mathbf{c}_j$ to a shadow $\mathbf{s}_{ij}$. Conventional SfM estimates $\mathbf{P}_i$ and $\mathbf{c}_j$ from $\{\mathbf{s}_{ij}\}$ and in this paper we show how to estimate $\mathbf{L}_i$ and $\mathbf{c}_j$ from $\{\mathbf{s}_{ij}\}$.

3.1 Shadow formation model

In this section, we show the mathematical relationship between a light source, shadow casters, and their corresponding shadows on a shadow receiver plane $\mathit{\Pi}$. Let us for now assume that the pose of the plane $\mathit{\Pi}$ is fixed to the world coordinate system's x-y plane.

Nearby light: Let a nearby point light be located at $\mathbf{l} = [l_x, l_y, l_z]^\top \in \mathbb{R}^3$ in world coordinates. An infinitesimally small caster located at $\mathbf{c} \in \mathbb{R}^3$ in world coordinates casts a shadow on the receiver plane $\mathit{\Pi}$ at $\mathbf{s} \in \mathbb{R}^2$ in $\mathit{\Pi}$'s 2D coordinate system, which is $\bar{\mathbf{s}} = \left[\mathbf{s}^\top, 0\right]^\top$ in world coordinates because $\mathit{\Pi}$ coincides with the world's x-y plane. Since $\mathbf{l}$, $\mathbf{c}$, and $\bar{\mathbf{s}}$ are all on the same line, the lines $\overline{\mathbf{c}\bar{\mathbf{s}}}$ and $\overline{\mathbf{l}\bar{\mathbf{s}}}$ must be parallel:

$$(\mathbf{c} - \bar{\mathbf{s}}) \times (\mathbf{l} - \bar{\mathbf{s}}) = \mathbf{0}. \tag{1}$$

Inserting $\mathbf{c} = [c_x, c_y, c_z]^\top$, $\bar{\mathbf{s}} = [s_x, s_y, 0]^\top$, and $\mathbf{l} = [l_x, l_y, l_z]^\top$ (all in non-homogeneous 3D global world coordinates) into Eq. (1) yields

$$(\mathbf{c} - \bar{\mathbf{s}}) \times (\mathbf{l} - \bar{\mathbf{s}}) = \begin{bmatrix} c_x - s_x \\ c_y - s_y \\ c_z - 0 \end{bmatrix} \times \begin{bmatrix} l_x - s_x \\ l_y - s_y \\ l_z - 0 \end{bmatrix} = \mathbf{0}.$$

Expanding the cross-product yields

$$\begin{cases} (c_y - s_y)l_z - c_z(l_y - s_y) = 0, \\ c_z(l_x - s_x) - (c_x - s_x)l_z = 0, \\ (c_x - s_x)(l_y - s_y) - (c_y - s_y)(l_x - s_x) = 0, \end{cases}$$
$$\Leftrightarrow \begin{cases} s_x = \dfrac{c_x l_z - c_z l_x}{l_z - c_z}, \\ s_y = \dfrac{c_y l_z - c_z l_y}{l_z - c_z}. \end{cases}$$

We can then rewrite $\mathbf{s}$ in homogeneous coordinates using scaling parameters γ and λ:

$$\gamma\tilde{\mathbf{s}} = \begin{bmatrix} \frac{c_x l_z - c_z l_x}{l_z - c_z} \\ \frac{c_y l_z - c_z l_y}{l_z - c_z} \\ 1 \end{bmatrix}$$

$$\underbrace{-(l_z - c_z)\gamma}_{\lambda}\tilde{\mathbf{s}} = \begin{bmatrix} -(c_x l_z - c_z l_x) \\ -(c_y l_z - c_z l_y) \\ -(l_z - c_z) \end{bmatrix}$$

$$\lambda\tilde{\mathbf{s}} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 1 & -l_z \end{bmatrix} \begin{bmatrix} c_x \\ c_y \\ c_z \\ 1 \end{bmatrix} = \mathbf{L}\tilde{\mathbf{c}}.$$

In the following, we will call $\mathbf{L}$ a light matrix. As we can see, point lights and pinhole cameras can be described by similar mathematical models with the following correspondences: (point light $\Leftrightarrow$ pinhole camera), (shadow receiver plane $\Leftrightarrow$ image plane), (shadow caster $\Leftrightarrow$ scene point), and (light matrix $\mathbf{L}$ $\Leftrightarrow$ camera projection matrix $\mathbf{P} = \mathbf{K}\,[\,\mathbf{R}\,|\,\mathbf{t}\,]$), as illustrated in Fig. 4. We can even decompose the light matrix $\mathbf{L}$ into "intrinsics" and "extrinsics" parameterized by the light location $\mathbf{l}$ as

$$\mathbf{L} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 1 & -l_z \end{bmatrix} = \underbrace{\begin{bmatrix} -l_z & 0 & l_x \\ 0 & -l_z & l_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{intrinsics } \mathbf{K}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & -l_x \\ 0 & 1 & 0 & -l_y \\ 0 & 0 & 1 & -l_z \end{bmatrix}}_{\text{extrinsics } [\mathbf{R}|\mathbf{t}]}. \quad (2)$$

Distant light: For distant light, all light rays in the scene are parallel, $\mathbf{l} = [l_x, l_y, l_z]^\top \in \mathcal{S}^2$ ($\mathcal{S}^2 = \{\mathbf{v} \in \mathbb{R}^3 : |\mathbf{v}| = 1\}$) is a light direction instead of a position, and the line $\overline{\mathbf{c}\bar{\mathbf{s}}}$ must be parallel to $\mathbf{l}$:

$$(\mathbf{c} - \bar{\mathbf{s}}) \times \mathbf{l} = \mathbf{0}. \quad (3)$$

Inserting $\mathbf{c}$, $\bar{\mathbf{s}}$, and $\mathbf{l}$ into Eq. (3) yields

$$(\mathbf{c} - \bar{\mathbf{s}}) \times \mathbf{l} = \begin{bmatrix} c_x - s_x \\ c_y - s_y \\ c_z - 0 \end{bmatrix} \times \begin{bmatrix} l_x \\ l_y \\ l_z \end{bmatrix} = \mathbf{0}.$$

By expanding the cross-product, we have

$$\begin{cases} (c_y - s_y)l_z - c_z l_y = 0, \\ c_z l_x - (c_x - s_x)l_z = 0, \\ (c_x - s_x)l_y - (c_y - s_y)l_x = 0, \end{cases}$$

$$\Leftrightarrow \begin{cases} s_x = \dfrac{c_x l_z - c_z l_x}{l_z}, \\ s_y = \dfrac{c_y l_z - c_z l_y}{l_z}. \end{cases}$$

We can then write $\mathbf{s}$ in homogeneous coordinates as:

$$\gamma\tilde{\mathbf{s}} = \begin{bmatrix} \frac{c_x l_z - c_z l_x}{l_z} \\ \frac{c_y l_z - c_z l_y}{l_z} \\ 1 \end{bmatrix}$$

$$\underbrace{-l_z\gamma}_{\lambda}\tilde{\mathbf{s}} = \begin{bmatrix} -(c_x l_z - c_z l_x) \\ -(c_y l_z - c_z l_y) \\ -l_z \end{bmatrix}$$

$$\lambda\tilde{\mathbf{s}} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 0 & -l_z \end{bmatrix} \begin{bmatrix} c_x \\ c_y \\ c_z \\ 1 \end{bmatrix} = \mathbf{L}\tilde{\mathbf{c}}.$$

The difference to the nearby light case is the entry $(\mathbf{L})_{3,3} = 0$. This light matrix resembles orthographic projection with a camera matrix $\left[\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{smallmatrix}\right]$.

Unifying nearby and distant light: Having two different models, a nearby and a distant light model, is a nuisance because it forces users to choose the one that better fits their scene, which can be hard especially for inexperienced users. Further, the real world does not exhibit a sharp transition from nearby to distant light (at a distance of, say, 3 m) but rather a smooth transition. It is thus desirable to have a unified model that also transitions smoothly. The intuition behind a unification is that orthographic projection can be seen as a special case of perspective projection with an infinite focal length.

Since in homogeneous coordinates we consider vectors equivalent if they are equal up to a constant, we can divide the nearby and distant light matrices by l_z:

$$\mathbf{L}_{\text{nearby}} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 1 & -l_z \end{bmatrix} \to \begin{bmatrix} -1 & 0 & l_x/l_z & 0 \\ 0 & -1 & l_y/l_z & 0 \\ 0 & 0 & 1/l_z & -1 \end{bmatrix},$$

$$\mathbf{L}_{\text{distant}} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 0 & -l_z \end{bmatrix} \to \begin{bmatrix} -1 & 0 & l_x/l_z & 0 \\ 0 & -1 & l_y/l_z & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

We can see that, if the light source moves towards infinity, $\mathbf{L}_{\text{nearby}}$ converges to $\mathbf{L}_{\text{distant}}$. Therefore, we use

$$\lambda\tilde{\mathbf{s}} = \begin{bmatrix} -1 & 0 & l_x/l_z & 0 \\ 0 & -1 & l_y/l_z & 0 \\ 0 & 0 & 1/l_z & -1 \end{bmatrix} \begin{bmatrix} c_x \\ c_y \\ c_z \\ 1 \end{bmatrix} = \mathbf{L}\tilde{\mathbf{c}}. \quad (4)$$

as unified shadow projection equation for representing both nearby and distant light. Later in this paper, we will see that we can use the unified projection model for determining light positions and directions in synthetic as well as real-world datasets.

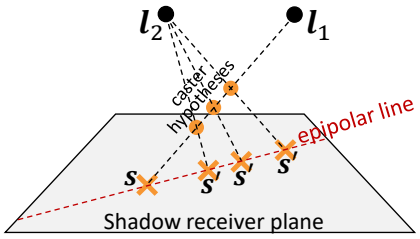

Fig. 5 A plane has a shadow caster above it at a fixed but unknown position. Given shadow $\mathbf{s}$ cast by light $\mathbf{l}_1$, the caster causing this shadow can be anywhere on the line $\overline{\mathbf{l}_1\mathbf{s}}$, yielding an infinite series of caster position hypotheses. When the light moves to position $\mathbf{l}_2$, this series results in a line of shadows, the equivalent of the epipolar line in camera geometry.

3.2 Epipolar geometry for shadow correspondences

We will now look at a scene with a moving light source, static shadow casters, and a static shadow receiver plane. We analyze shadow correspondences, *i.e.*, the relation between shadows belonging to the same caster but different light positions. As we saw, shadow projection is a special case of projective geometry. Thus, shadow correspondences in a static scene with a moving point light should follow the same principles as image point correspondences in pinhole camera projection with a static scene and a moving camera: epipolar geometry.

3.2.1 Fundamental shadow matrix

Figure 5 shows a shadow receiver plane Π and two point lights at positions $\mathbf{l}_1$ and $\mathbf{l}_2$. Shadow $\mathbf{s}$ from a shadow caster at an unknown position cast by light $\mathbf{l}_1$ has corresponding shadows $\mathbf{s}'$ cast by light $\mathbf{l}_2$. These can be found on an epipolar line arising from a fundamental matrix, which we will derive now analogous to the derivation of the standard fundamental matrix.

Let $\mathbf{c}$ be the caster position, $\mathbf{l}_1 \neq \mathbf{l}_2$ be the light positions in calibration target coordinates, $\mathbf{L}_i$ be $\mathbf{l}_i$'s light matrix, $\mathbf{L}_1^+ = (\mathbf{L}_1^\top \mathbf{L}_1)^{-1}\mathbf{L}_1^\top$ be $\mathbf{L}_1$'s pseudo inverse, $\boldsymbol{\emptyset}_{\mathbf{L}_1} \in \mathsf{null}(\mathbf{L}_1)$ be a non-zero vector in $\mathbf{L}_1$'s one-dimensional null space, and η be a scalar. We then have

$$\lambda_1 \tilde{\mathbf{s}}_1 = \mathbf{L}_1 \tilde{\mathbf{c}} \;\Rightarrow\; \tilde{\mathbf{c}} = \lambda_1 \mathbf{L}_1^+ \tilde{\mathbf{s}}_1 + \eta\, \boldsymbol{\emptyset}_{\mathbf{L}_1}$$
$$\lambda_2 \tilde{\mathbf{s}}_2 = \mathbf{L}_2 \tilde{\mathbf{c}} = \lambda_1 \mathbf{L}_2 \mathbf{L}_1^+ \tilde{\mathbf{s}}_1 + \eta\, \mathbf{L}_2\, \boldsymbol{\emptyset}_{\mathbf{L}_1}.$$

Multiplying $\tilde{\mathbf{s}}_2^\top [\mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times$ (with $[\cdot]_\times$ being the cross-product's matrix form) from the left, we obtain

$$\lambda_2 \underbrace{\tilde{\mathbf{s}}_2^\top [\mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \tilde{\mathbf{s}}_2}_{0} = \lambda_1 \tilde{\mathbf{s}}_2^\top [\mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{L}_2 \mathbf{L}_1^+ \tilde{\mathbf{s}}_1 + \eta\, \tilde{\mathbf{s}}_2^\top \underbrace{[\mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}}_{\mathbf{0}}$$
$$\Rightarrow \quad 0 = \tilde{\mathbf{s}}_2^\top \underbrace{[\mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{L}_2 \mathbf{L}_1^+}_{\mathbf{F}} \tilde{\mathbf{s}}_1. \tag{5}$$

Thus, corresponding shadows $\tilde{\mathbf{s}}_1$ and $\tilde{\mathbf{s}}_2$ fulfill a condition with some fundamental matrix $\mathbf{F}$ that is directly analogous to the regular correspondence condition. For $\mathbf{L}_i = \begin{bmatrix} -1 & 0 & l_x^{(i)}/l_z^{(i)} & 0 \\ 0 & -1 & l_y^{(i)}/l_z^{(i)} & 0 \\ 0 & 0 & 1/l_z^{(i)} & -1 \end{bmatrix}$, we obtain the null space vector $\boldsymbol{\emptyset}_{\mathbf{L}_1} \propto \left[l_x^{(1)}, l_y^{(1)}, l_z^{(1)}, 1\right]^\top$ and finally the fundamental shadow matrix

$$\begin{aligned}\mathbf{F} &= [\mathbf{L}_2 \boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{L}_2 \mathbf{L}_1^+ \\ &\propto \frac{1}{l_z^{(2)}} \begin{bmatrix} 0 & -l_z^{(1)}+l_z^{(2)} & -l_y^{(1)}l_z^{(2)}+l_y^{(2)}l_z^{(1)} \\ l_z^{(1)}-l_z^{(2)} & 0 & l_x^{(1)}l_z^{(2)}-l_x^{(2)}l_z^{(1)} \\ l_y^{(1)}l_z^{(2)}-l_y^{(2)}l_z^{(1)} & -l_x^{(1)}l_z^{(2)}+l_x^{(2)}l_z^{(1)} & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & f_1 & f_2 \\ -f_1 & 0 & f_3 \\ -f_2 & -f_3 & 0 \end{bmatrix}. \end{aligned} \tag{6}$$

Interestingly, this matrix is a special case of regular fundamental matrices: It is skew-symmetric. For corresponding shadows $\tilde{\mathbf{s}}_1 = [u, v, 1]^\top$ and $\tilde{\mathbf{s}}_2 = [u', v', 1]^\top$ the correspondence condition Eq. (5) becomes

$$\begin{aligned} 0 &= \begin{bmatrix} u & v & 1 \end{bmatrix} \begin{bmatrix} 0 & f_1 & f_2 \\ -f_1 & 0 & f_3 \\ -f_2 & -f_3 & 0 \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \\ &= (uv' - vu')f_1 + (u - u')f_2 + (v - v')f_3 \end{aligned} \tag{7}$$

We can thus estimate the parameters f_1, f_2, and f_3 of our fundamental shadow matrix up to scale by solving the homogeneous linear system

$$\begin{bmatrix} u_1 v_1' - v_1 u_1' & u_1 - u_1' & v_1 - v_1' \\ \vdots & \vdots & \vdots \\ u_n v_n' - v_n u_n' & u_n - u_n' & v_n - v_n' \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \mathbf{0}_{n \times 1} \tag{8}$$

with $n \geq 2$ correspondences using singular value decomposition. This is actually equivalent to estimating regular essential matrices for cameras that only undergo pure translation and no relative rotation [40, Eq. 7.27]. This makes sense since we saw in Eq. (2) that point lights act like cameras with an identity rotation. As a consequence, all properties discussed in the following also hold for essential matrix estimation of cameras with pure translation.

Conveniently, the fundamental shadow matrix has a rank of 2 as a direct result of estimating the parameters of a skew-symmetric matrix and the rank does not need to be enforced in a post-processing step. Further, the matrix can be estimated up to scale from 2 point correspondences. Moreover, in contrast to regular fundamental/essential matrix estimation, fundamental shadow matrix estimation does not suffer from the most common degenerate scene point configuration: all 3D scene points lying in a plane. We will show this now.

Planar degeneracy: A degeneracy in fundamental matrix estimation means that a correct fundamental matrix $\mathbf{F}$ exists and could be computed from the projection matrices, but the 3D scene points are in a configuration such that we can find an $\mathbf{F}' \neq \mathbf{F}$ that also fulfills the correspondence condition $\tilde{\mathbf{s}}_2^\top \mathbf{F}' \tilde{\mathbf{s}}_1 = 0$.

In regular SfM, a well-known degeneracy are coplanar scene points. In this case, their projections in both views are related by a homography: $\mathbf{x}_i' = \mathbf{H}\mathbf{x}_i$, the correspondence condition $0 = \mathbf{x}_i'^\top \mathbf{F}\mathbf{x}_i = \mathbf{x}_i'^\top \underbrace{\mathbf{F}\mathbf{H}^{-1}}_{\mathbf{S}}\mathbf{x}_i'$ is true for any skew-symmetric $\mathbf{S}$, and thus any fundamental matrix $\mathbf{F} = \mathbf{S}\mathbf{H}$ (with any skew-symmetric $\mathbf{S}$ and the homography $\mathbf{H}$) is a valid solution [18, Sec. 11.9.2]. The planar degeneracy is relevant in practice because a camera may have only captured scene points from a wall, floor, or table surface.

We will now show that fundamental shadow matrices have no *general* planar degeneracy, *i.e.*, a degeneracy from coplanar casters independent of their configuration towards the lights and image plane: fundamental shadow matrices are skew-symmetric, are thus essential matrices [18, Result 9.17] and can have at most the degeneracies of essential matrices. Further, since essential matrices have the form $[\mathbf{t}]_\times \mathbf{R}$, our skew-symmetric shadow matrices are a special case of essential matrices where we have $\mathbf{R} = \mathbf{I}$. From Negahdaripour [23] we know that when observing the projection of a 3D plane, there are 2 sets of translation, rotation, and 3D plane coordinates that satisfy the observations. Let $\mathbf{R}$ be the true and $\mathbf{R}'$ be the alternative rotation. Negahdaripour's lemma [23, p. 5] states that

$$\begin{aligned} &\mathbf{R}' = \mathbf{V}\mathbf{R} \text{ with} \\ &\mathbf{V} = (1-\cos\bar{\theta})\,\bar{\mathbf{n}}\bar{\mathbf{n}}^\top + \sin\bar{\theta}\,\mathbf{N} + \cos\bar{\theta}\,\mathbf{I}, \end{aligned} \tag{9}$$

where $\mathbf{N}$ is skew-symmetric and $\bar{\mathbf{n}}$ is a unit vector. Here the precise meanings of $\bar{\mathbf{n}}$, $\bar{\theta}$, and $\mathbf{N}$ do not matter, only the form of Eq. (9) does. Recall that the rotations of fundamental shadow matrices are identities. Inserting $\mathbf{R}' = \mathbf{R} = \mathbf{I}$ into Eq. (9) yields

$$\begin{aligned} &\mathbf{I} = \mathbf{V}\mathbf{I} = (1-\cos\bar{\theta})\,\bar{\mathbf{n}}\bar{\mathbf{n}}^\top + \sin\bar{\theta}\,\mathbf{N} + \cos\bar{\theta}\,\mathbf{I} \\ \Rightarrow\ &\mathbf{I} = \bar{\mathbf{n}}\bar{\mathbf{n}}^\top + \frac{\sin\bar{\theta}}{1-\cos\bar{\theta}}\mathbf{N}. \end{aligned} \tag{10}$$

As $|\bar{\mathbf{n}}| = 1$, $\bar{\mathbf{n}}\bar{\mathbf{n}}^\top$ has at most one diagonal element that is 1. As $\mathbf{N}$ is skew-symmetric, the diagonal elements of $\frac{\sin\bar{\theta}}{1-\cos\bar{\theta}}\mathbf{N}$ are all 0. Thus, Eq. (10) cannot be true, implying that constraining the essential matrix to a skew-symmetric matrix removes the planar degeneracy.

In our application, not having a general planar degeneracy is important as we work with approximately coplanar shadow casters (the red pin heads in Fig. 1). Further, it means we can estimate $\mathbf{F}$ with 3 casters even though 3 casters are necessarily coplanar.

Apart from the lack of a general planar degeneracy, there do exist specialized ones. Most of them are fairly obvious and rather irrelevant in practice, *e.g.*, casters lying on the image plane (their shadows thus always keeping their positions) or all lights and casters being collinear (all shadows thus being projected to the same point). Beside these, the only degeneracy we could find is all casters and both lights being coplanar, in which case all shadows are projected to the line where the image plane and the casters-and-lights plane intersect. However, this configuration is unlikely in reality. To make it non-degenerate, it suffices that 1 caster or 1 light is not in the casters-and-lights plane.

We further found empirically that the estimation of $\mathbf{F}$ from 2 casters, which are necessarily collinear, is not generally degenerate. Both lights and one or both casters or both casters and one or both lights being collinear is degenerate. And again, both casters and both lights being coplanar is degenerate but one light or caster not being in the casters-and-lights plane is non-degenerate.

Hartley normalization: Hartley [16] proposed normalizing the points $[u_i, v_i]^\top$ and $[u_i', v_i']^\top$ to zero mean and unit variance in x- and y-directions to increase the algorithm's numerical stability in applications with large x and y image coordinates. In regular Hartley normalization, it is admissible to normalize the points of the left and right images independently (with scaling parameters s_x, s_y, s_x', s_y' and shifting parameters d_x, d_y, d_x', d_y') to obtain the normalized point coordinates $(u_\mathrm{n}, v_\mathrm{n})$ and $(u_\mathrm{n}', v_\mathrm{n}')$:

$$\begin{bmatrix} u_\mathrm{n} \\ v_\mathrm{n} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s_x} & 0 & -\frac{d_x}{s_x} \\ 0 & \frac{1}{s_y} & -\frac{d_y}{s_y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} u_\mathrm{n}' \\ v_\mathrm{n}' \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s_x'} & 0 & -\frac{d_x'}{s_x'} \\ 0 & \frac{1}{s_y'} & -\frac{d_y'}{s_y'} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}.$$

For fundamental shadow matrices, we need to be more cautious because we want to maintain their special structure. Normalizing both images identically, *i.e.*,

$$\begin{bmatrix} u_\mathrm{n} \\ v_\mathrm{n} \\ 1 \end{bmatrix} = \mathbf{N} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} u_\mathrm{n}' \\ v_\mathrm{n}' \\ 1 \end{bmatrix} = \mathbf{N} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \text{ with } \mathbf{N} = \begin{bmatrix} \frac{1}{s_x} & 0 & -\frac{d_x}{s_x} \\ 0 & \frac{1}{s_y} & -\frac{d_y}{s_y} \\ 0 & 0 & 1 \end{bmatrix}$$

preserves $\mathbf{F}$'s skew-symmetry:

$$\mathbf{N}^\top \begin{bmatrix} 0 & f_1 & f_2 \\ -f_1 & 0 & f_3 \\ -f_2 & -f_3 & 0 \end{bmatrix} \mathbf{N} = \frac{1}{s_x s_y} \begin{bmatrix} 0 & f_1 & -f_1 d_y + f_2 s_y \\ -f_1 & 0 & f_1 d_x + f_3 s_x \\ f_1 d_y - f_2 s_y & -f_1 d_x - f_3 s_x & 0 \end{bmatrix}.$$

Using different matrices for the left and right image would result in a non-skew-symmetric matrix which would require more than 2 point correspondences for estimation.

Analogous to regular Hartley normalization, from the matrix $\mathbf{F}_{\mathrm{n}}$ for normalized points, we can compute the matrix $\mathbf{F}$ for unnormalized points as

$$\mathbf{F} = \mathbf{N}^\top \mathbf{F}_{\mathrm{n}} \mathbf{N}.$$

3.2.2 Trifocal shadow tensor

For shadow correspondences in 3 images, we have a trifocal shadow tensor ("tritensor" or "shadow tritensor" in the following). The tritensor's general form for 3 general projection matrices $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ is [18, Eq. (17.12)]

$$(\mathcal{T})_{p,q,r} = (-1)^{p+1} \det \begin{bmatrix} (\mathbf{P}_1)_{1:p-1,:} \\ (\mathbf{P}_1)_{p+1:n,:} \\ (\mathbf{P}_2)_{q,:} \\ (\mathbf{P}_3)_{r,:} \end{bmatrix} \tag{11}$$

(recall that the $(\cdot)_{i,j}$ notation extracts parts of a matrix). Inserting 3 light matrices $\mathbf{L}_i = \begin{bmatrix} -1 & 0 & l_x^{(i)}/l_z^{(i)} & 0 \\ 0 & -1 & l_y^{(i)}/l_z^{(i)} & 0 \\ 0 & 0 & 1/l_z^{(i)} & -1 \end{bmatrix}$ for $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ yields

$$\mathcal{T} = \frac{1}{l_z^{(1)} l_z^{(2)} l_z^{(3)}} \cdot \left[\begin{bmatrix} l_z^{(1)}\left(l_x^{(2)} l_z^{(3)} - l_x^{(3)} l_z^{(2)}\right) & l_z^{(2)}\left(l_y^{(1)} l_z^{(3)} - l_y^{(3)} l_z^{(1)}\right) & l_z^{(2)}\left(l_z^{(3)} - l_z^{(1)}\right) \\ l_z^{(3)}\left(l_y^{(2)} l_z^{(1)} - l_y^{(1)} l_z^{(2)}\right) & 0 & 0 \\ l_z^{(3)}\left(l_z^{(1)} - l_z^{(2)}\right) & 0 & 0 \end{bmatrix} \right.$$
$$\begin{bmatrix} 0 & l_z^{(3)}\left(l_x^{(2)} l_z^{(1)} - l_x^{(1)} l_z^{(2)}\right) & 0 \\ l_z^{(2)}\left(l_x^{(1)} l_z^{(3)} - l_x^{(3)} l_z^{(1)}\right) & l_z^{(1)}\left(l_y^{(2)} l_z^{(3)} - l_y^{(3)} l_z^{(2)}\right) & l_z^{(2)}\left(l_z^{(3)} - l_z^{(1)}\right) \\ 0 & l_z^{(3)}\left(l_z^{(1)} - l_z^{(2)}\right) & 0 \end{bmatrix}$$
$$\left. \begin{bmatrix} 0 & 0 & l_z^{(3)}\left(l_x^{(2)} l_z^{(1)} - l_x^{(1)} l_z^{(2)}\right) \\ 0 & 0 & l_z^{(3)}\left(l_y^{(2)} l_z^{(1)} - l_y^{(1)} l_z^{(2)}\right) \\ l_z^{(2)}\left(l_x^{(1)} l_z^{(3)} - l_x^{(3)} l_z^{(1)}\right) & l_z^{(2)}\left(l_y^{(1)} l_z^{(3)} - l_y^{(3)} l_z^{(1)}\right) & l_z^{(1)}\left(l_z^{(3)} - l_z^{(2)}\right) \end{bmatrix} \right].$$

Thus, similar to the fundamental shadow matrix, the shadow tritensor also has a special structure as

$$\mathcal{T} = [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] = \left[\begin{bmatrix} t_1 & t_2 & t_3 \\ t_4 & 0 & 0 \\ t_5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & t_6 & 0 \\ t_7 & t_8 & t_3 \\ 0 & t_5 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & t_6 \\ 0 & 0 & t_4 \\ t_7 & t_2 & t_9 \end{bmatrix} \right]. \tag{12}$$

It can further be verified that $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{T}_3$ always have rank 2 each, as required [18, p. 373].

For the shadow tritensor, the correspondence condition for shadows $\tilde{\mathbf{s}}_1 = [u, v, 1]^\top$, $\tilde{\mathbf{s}}_2 = [u', v', 1]^\top$, and $\tilde{\mathbf{s}}_3 = [u'', v'', 1]^\top$ is [18, Eq. (15.7)]:

$$[\tilde{\mathbf{s}}_2]_\times \Big(\sum\nolimits_{i \in \{1,2,3\}} (\tilde{\mathbf{s}}_1)_i \, \mathbf{T}_i \Big) [\tilde{\mathbf{s}}_3]_\times = \mathbf{0}_{3\times 3},$$

which can be reformulated into the system shown in Eq. (13) on the next page. Its row echelon form, which has the same solution as the original system, is shown in Eq. (14) on the next page, revealing the system's rank of 4. We thus need at least 2 shadow correspondences and stack their matrices to estimate the 9 unknowns of the shadow tritensor up to scale. In contrast, general tritensor estimation requires at least 7 correspondences for the linear algorithm [18, Algorithm 16.1].

Hartley normalization: When normalizing input points for better numerical stability, we again need to normalize all images identically to maintain the tritensor's special structure (Eq. (12)) and keep its parameters to 9:

$$\begin{bmatrix} u_{\mathrm{n}} \\ v_{\mathrm{n}} \\ 1 \end{bmatrix} = \mathbf{N} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad \begin{bmatrix} u'_{\mathrm{n}} \\ v'_{\mathrm{n}} \\ 1 \end{bmatrix} = \mathbf{N} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} u''_{\mathrm{n}} \\ v''_{\mathrm{n}} \\ 1 \end{bmatrix} = \mathbf{N} \begin{bmatrix} u'' \\ v'' \\ 1 \end{bmatrix}$$
$$\text{with } \mathbf{N} = \begin{bmatrix} \frac{1}{s_x} & 0 & -\frac{d_x}{s_x} \\ 0 & \frac{1}{s_y} & -\frac{d_y}{s_y} \\ 0 & 0 & 1 \end{bmatrix}.$$

3.2.3 Quadrifocal shadow tensor

The quadrifocal shadow tensor $\mathcal{Q}$ (or "shadow quadtensor" or "quadtensor" in the following) describes 4-view shadow correspondences. The quadtensor's form for 4 general projection matrices $\mathbf{P}_i$ is [18, Eq. (17.21)]

$$(\mathcal{Q})_{p,q,r,s} = \det \begin{bmatrix} (\mathbf{P}_1)_{p,:} \\ (\mathbf{P}_2)_{q,:} \\ (\mathbf{P}_3)_{r,:} \\ (\mathbf{P}_4)_{s,:} \end{bmatrix}. \tag{15}$$

Inserting light matrices for the projections yields $\mathcal{Q}$ as shown in the left of Fig. A.1 in the appendix. Its exact content is less relevant here but we note that it has the following shape with 18 unknowns:

$$\mathcal{Q} = \begin{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -q_1 \\ 0 & q_1 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & q_2 \\ 0 & 0 & q_3 \\ q_4 & q_5 & q_6 \end{bmatrix} & \begin{bmatrix} 0 & -q_2 & 0 \\ -q_4 & q_7 & q_8 \\ 0 & q_9 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & q_{10} \\ 0 & 0 & q_{11} \\ q_{12} & q_{13} & -q_6 \end{bmatrix} & \begin{bmatrix} 0 & 0 & -q_{14} \\ 0 & 0 & 0 \\ q_{14} & 0 & 0 \end{bmatrix} & \begin{bmatrix} q_{15} & -q_{13} & -q_8 \\ -q_{11} & 0 & 0 \\ -q_9 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & -q_{10} & 0 \\ -q_{12} & -q_7 & -q_{16} \\ 0 & q_{17} & 0 \end{bmatrix} & \begin{bmatrix} -q_{15} & -q_5 & q_{16} \\ -q_3 & 0 & 0 \\ -q_{17} & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & -q_{18} & 0 \\ q_{18} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{bmatrix} \tag{16}$$

The correspondence conditions [18, Eq. (17.20)] for 4 shadows $\tilde{\mathbf{s}}_1 = [u, v, 1]^\top$, $\tilde{\mathbf{s}}_2 = [u', v', 1]^\top$, $\tilde{\mathbf{s}}_3 = [u'', v'', 1]^\top$, $\tilde{\mathbf{s}}_4 = [u''', v''', 1]^\top$ are (with the Levi-Civita symbol ε)

$$\sum_{\substack{i,j,k,l, \\ p,q,r,s}} (\tilde{\mathbf{s}}_1)_i (\tilde{\mathbf{s}}_2)_j (\tilde{\mathbf{s}}_3)_k (\tilde{\mathbf{s}}_4)_l \, \varepsilon_{ipw} \varepsilon_{jqx} \varepsilon_{kry} \varepsilon_{lsz} (\mathcal{Q})_{p,q,r,s} = 0.$$

Iterating over the free variables w, x, y, z yields 81 equations of which 3 are zero independent of the data. The remaining 78 put in a homogeneous system are shown in Fig. A.1, right, in the appendix. The system has rank 13 and stacking observations from ≥ 2 correspondences suffices to estimate the 18 unknowns up to scale.

$$\begin{bmatrix} 0 & v' & vv'' & v'' & vv' & 0 & 0 & -v & -v'v'' \\ 0 & 0 & -u''v & u-u'' & -uv' & 0 & v-v' & 0 & u''v' \\ 0 & -u''v' & 0 & -uv'' & v'(uv''-u''v) & 0 & v''(v'-v) & u''v & 0 \\ 0 & u-u' & -uv'' & 0 & -u'v & v-v'' & 0 & 0 & u'v'' \\ -u & 0 & uu'' & 0 & uu' & u'' & u' & 0 & -u'u'' \\ uv'' & u''(u'-u) & 0 & 0 & u'(u''v-uv'') & -u''v & -u'v'' & 0 & 0 \\ 0 & -uv' & v''(uv'-u'v) & -u'v'' & 0 & v'v''-vv' & 0 & u'v & 0 \\ uv' & 0 & u''(u'v-uv') & u'(u''-u) & 0 & -u''v' & -u'v & 0 & 0 \\ -uv'v'' & uu''v' & 0 & uu'v'' & 0 & u''vv' & u'vv'' & -u'u''v & 0 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \\ t_9 \end{bmatrix} = \mathbf{0}_{9\times 1} \tag{13}$$

$$\begin{bmatrix} -u & 0 & uu'' & 0 & uu' & u'' & u' & 0 & -u'u'' \\ 0 & -u''v' & 0 & -uv'' & v'(uv''-u''v) & 0 & v''(-v+v') & u''v & 0 \\ 0 & 0 & -u''v & u-u'' & -uv' & 0 & v-v' & 0 & u''v' \\ 0 & 0 & 0 & -uu''v''(v(u-u')+v'(u-u')) & uu''v'(uvv''+uv'v''-u'vv''-u''v^2) & u''^2vv'(v-v'') & -u''v''(v-v')(uv'+v(u-u')) & u''^2v^2(u-u') & u''^2v'v''(-uv'+u'v) \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \\ t_9 \end{bmatrix} = \mathbf{0}_{4\times 1} \tag{14}$$

3.2.4 Distant light

In our calibration method that we describe in Sec. 4, we always use the unified light model for which we derived the shadow matrices/tensors above. Nevertheless, it is interesting to see what properties shadow matrix/tensor estimation has for scenes where we can assume distant light with certainty.

Inserting distant light matrices into our derivations reveals that the shadow matrices/tensors take on specialized shapes: Compared to their counterparts for the unified light model (Eqs. (6), (12), and (16)), they have the same pattern of repeated entries and zero entries, but a few additional entries become zero. For the fundamental shadow matrix we have $f_1 = 0$ and the remaining 2 unknowns can be estimated from 1 shadow correspondence. For the tritensor, we have $t_3 = t_5 = t_9 = 0$, and the system (Eq. (13)) stays rank 4; thus, we need 2 correspondences to estimate the remaining 6 unknowns. For the quadtensor, we have $q_6 = q_8 = q_9 = q_{16} = q_{17} = q_{18} = 0$, the rank of the estimation system (Fig. A.1, right) drops from 13 to 11 and we need only 1 correspondence to estimate the remaining 12 unknowns.

3.2.5 Shadow correspondences through an uncalibrated camera

If the shadow observations for which we try to find a shadow matrix/tensor or shadow correspondences, are not given in the coordinate system of the shadow receiver plane but only in the image coordinates of a static, uncalibrated pinhole camera that observes the receiver plane, the complexity of shadow matrix/tensor estimation does interestingly not increase: We can still estimate them from just 2 shadow correspondences. A static pinhole camera observing coplanar shadows can be modeled with a homography (invertible 3×3 matrix) $\mathbf{H}$ in the shadow projection equations:

$$\lambda_i \tilde{\mathbf{s}}_i = \mathbf{H}\mathbf{L}_i\tilde{\mathbf{c}}.$$

Analogous to Eq. (5), we can then derive the fundamental shadow matrix for shadows observed through an uncalibrated camera:

$$\begin{aligned}
\mathbf{F}_\mathrm{c} &= [\mathbf{H}\mathbf{L}_2\boldsymbol{\emptyset}_{\mathbf{H}\mathbf{L}_1}]_\times \mathbf{H}\mathbf{L}_2(\mathbf{H}\mathbf{L}_1)^+ \\
&= [\mathbf{H}\mathbf{L}_2\boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{H}\mathbf{L}_2 \overbrace{(\mathbf{H}^+\mathbf{H}\mathbf{L}_1)^+(\mathbf{H}\mathbf{L}_1\mathbf{L}_1^+)^+}^{\text{see [27, Eq. (214)]}} \\
&= [\mathbf{H}\mathbf{L}_2\boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{H}\mathbf{L}_2(\mathbf{H}^{-1}\mathbf{H}\mathbf{L}_1)^+\mathbf{H}^+ \\
&= [\mathbf{H}\mathbf{L}_2\boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times \mathbf{H}\mathbf{L}_2\mathbf{L}_1^+\mathbf{H}^{-1} \\
&= \overbrace{\det(\mathbf{H})\,\mathbf{H}^{-\top}\,[\mathbf{L}_2\boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times\mathbf{L}_2}^{\substack{\text{follows from} \\ (\mathbf{H}\mathbf{a})\times(\mathbf{H}\mathbf{b})=\det(\mathbf{H})\mathbf{H}^{-\top}(\mathbf{a}\times\mathbf{b})}}\ \mathbf{L}_1^+\mathbf{H}^{-1} \\
&\propto \mathbf{H}^{-\top}\,[\mathbf{L}_2\boldsymbol{\emptyset}_{\mathbf{L}_1}]_\times\mathbf{L}_2\mathbf{L}_1^+\mathbf{H}^{-1} \\
&= \mathbf{H}^{-\top}\mathbf{F}\mathbf{H}^{-1},
\end{aligned}$$

where $\mathbf{F}$ is the original fundamental matrix of Eq. (5) for the shadows given in shadow receiver plane coordinates. $\mathbf{F}_\mathrm{c}$ is also skew-symmetric and can thus be estimated from ≥ 2 correspondences using Eq. (8). Analogous to Eqs. (11) and (15), the shadow tensors become

$$(\mathcal{T})_{p,q,r} = (-1)^{p+1}\det\begin{bmatrix} (\mathbf{H}\mathbf{P}_1)_{1:p-1,:} \\ (\mathbf{H}\mathbf{P}_1)_{p+1:n,:} \\ (\mathbf{H}\mathbf{P}_2)_{q,:} \\ (\mathbf{H}\mathbf{P}_3)_{r,:} \end{bmatrix} \quad \text{and}$$

$$(\mathcal{Q})_{p,q,r,s} = \det\begin{bmatrix} (\mathbf{H}\mathbf{P}_1)_{p,:} \\ (\mathbf{H}\mathbf{P}_2)_{q,:} \\ (\mathbf{H}\mathbf{P}_3)_{r,:} \\ (\mathbf{H}\mathbf{P}_4)_{s,:} \end{bmatrix}.$$

These have the same patterns of zeros and identical entries as the original tensors (Eqs. (12) and (16)) and can thus be estimated from ≥ 2 correspondences using

Table 1 Minimal number of shadow correspondences required to estimate shadow matrices/tensors.

coordinate system of the shadows	light model	shadow F-matrix	shadow tritensor	shadow quadtensor
shadow receiver	nearby/unified	2	2	2
shadow receiver	distant	1	2	1
observing camera	nearby/unified	2	2	2
observing camera	distant	2	2	2

Eqs. (13), (14), or (A.1), respectively. The counterparts of these matrices/tensors for distant light have the same shape as those for nearby light and they all need to be estimated from ≥ 2 correspondences.

Table 1 sums up the minimal number of correspondences required for shadow matrix/tensor estimation.

4 Proposed method

Our method estimates a nearby point light's position or a distant light's direction using a simple calibration target consisting of a shadow receiver plane and shadow casters above the plane (see Fig. 1). Our method automatically achieves the point light source calibration by observing the calibration target multiple times from a fixed viewpoint under a fixed point light source while changing the calibration target's pose. The 3D positions of the shadow casters relative to the calibration target are treated *unknown*, which makes it particularly easy to build the target while the problem remains tractable as we will see later in this section. We now describe our proposed calibration method.

4.1 Light source calibration as bundle adjustment

Our goal is to determine the light $\mathbf{l}$ in Eq. (4) by observing the shadows cast by unknown casters. A single shadow observation $\mathbf{s}$ does not provide sufficient information to solve this. We thus let the receiver plane undergo multiple poses $\{[\mathbf{R}_i|\mathbf{t}_i]\}$. In pose i, the light position $\mathbf{l}_i$ in receiver plane coordinates is related to $\mathbf{l}$ in world coordinates as

$$\mathbf{l}_i = \begin{bmatrix} l_x^{(i)} & l_y^{(i)} & l_z^{(i)} \end{bmatrix}^\top = \mathbf{R}_i^\top \mathbf{l} - \mathbf{R}_i^\top \mathbf{t}_i.$$

With this index i the matrices $\{\mathbf{L}_i\}$ read

$$\mathbf{L}_i = \begin{bmatrix} -1 & 0 & l_x^{(i)}/l_z^{(i)} & 0 \\ 0 & -1 & l_y^{(i)}/l_z^{(i)} & 0 \\ 0 & 0 & 1/l_z^{(i)} & -1 \end{bmatrix}.$$

If we use not only multiple poses $\{[\mathbf{R}_i|\mathbf{t}_i]\}$ but also multiple shadow casters $\{\mathbf{c}_j\}$ (to increase the calibration accuracy as we show later), we obtain shadows $\{\mathbf{s}_{ij}\}$ for each combination of pose i and caster j. Equation (4) then becomes

$$\lambda_{ij}\tilde{\mathbf{s}}_{ij} = \mathbf{L}_i\tilde{\mathbf{c}}_j.$$

Assuming that the target poses $\{[\mathbf{R}_i|\mathbf{t}_i]\}$ are known, our goal is to estimate the light position $\mathbf{l}$ in world coordinates and the shadow caster locations $\{\mathbf{c}_j\}$ in calibration target coordinates. We formulate this as a least-squares objective function of the reprojection error:

$$\min_{\mathbf{l},\mathbf{c}_j,\lambda_{ij}} \sum_{i,j} \|\lambda_{ij}\tilde{\mathbf{s}}_{ij} - \mathbf{L}_i\tilde{\mathbf{c}}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{l} = \mathbf{R}_i\mathbf{l}_i + \mathbf{t}_i. \tag{17}$$

We solve this nonlinear least-squares problem with Levenberg-Marquardt [24]. For robust estimation we use RANSAC [11]: We repeatedly choose a random observation set, estimate $(\mathbf{l}, \mathbf{c}_j, \lambda_{ij})$, and select the estimate with the smallest residual.

4.2 Initializing the bundle adjustment

Equation (17) is non-convex and thus affected by the initialization. To find a good initial guess, we relax our problem into a convex one as follows.

Nearby light: For nearby light, we can write the objective analogous to Eq. (1) as $(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times (\mathbf{l}_i - \bar{\mathbf{s}}_{ij}) = \mathbf{0}$ and then, using $\mathbf{l}_i = \mathbf{R}_i^\top\mathbf{l} - \mathbf{R}_i^\top\mathbf{t}_i$, as

$$(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times (\mathbf{R}_i^\top\mathbf{l} - \mathbf{R}_i^\top\mathbf{t}_i - \bar{\mathbf{s}}_{ij}) = \mathbf{0}.$$

With $\mathbf{c}_j = [c_{j,x}, c_{j,y}, c_{j,z}]^\top$, $\bar{\mathbf{s}}_{ij} = [s_x, s_y, 0]^\top$, $\mathbf{l} = [l_x, l_y, l_z]^\top$, $\mathbf{R}_i^\top = \begin{bmatrix} r_0 & r_1 & r_2 \\ r_3 & r_4 & r_5 \\ r_6 & r_7 & r_8 \end{bmatrix}$, and $-\mathbf{R}_i^\top\mathbf{t}_i = [t_x, t_y, t_z]^\top$, we can rewrite this as

$$\begin{aligned} \mathbf{0} &= (\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times (\mathbf{R}_i^\top\mathbf{l} - \mathbf{R}_i^\top\mathbf{t}_i - \bar{\mathbf{s}}_{ij}) \\ &= \begin{bmatrix} c_{j,x} - s_x \\ c_{j,y} - s_y \\ c_{j,z} \end{bmatrix} \times \begin{bmatrix} [r_0, r_1, r_2]\,\mathbf{l} + t_x - s_x \\ [r_3, r_4, r_5]\,\mathbf{l} + t_y - s_y \\ [r_6, r_7, r_8]\,\mathbf{l} + t_z \end{bmatrix}. \end{aligned}$$

Expanding the cross-product yields

$$\begin{cases} 0 = (c_{j,y} - s_y)([r_6, r_7, r_8]\,\mathbf{l} + t_z) - c_{j,z}([r_3, r_4, r_5]\,\mathbf{l} + t_y - s_y), \\ 0 = c_{j,z}([r_0, r_1, r_2]\,\mathbf{l} + t_x - s_x) - (c_{j,x} - s_x)([r_6, r_7, r_8]\,\mathbf{l} + t_z), \\ 0 = (c_{j,x} - s_x)([r_3, r_4, r_5]\,\mathbf{l} + t_y - s_y) \\ \qquad - (c_{j,y} - s_y)([r_0, r_1, r_2]\,\mathbf{l} + t_x - s_x), \end{cases}$$

which we can rewrite as

$$\begin{cases} -s_y t_z = -c_{j,y}([r_6, r_7, r_8]\,\mathbf{l} + t_z) + s_y[r_6, r_7, r_8]\,\mathbf{l} \\ \qquad\qquad + c_{j,z}([r_3, r_4, r_5]\,\mathbf{l} + t_y - s_y), \\ s_x t_z = -c_{j,z}([r_0, r_1, r_2]\,\mathbf{l} + t_x - s_x), \\ \qquad\qquad + c_{j,x}([r_6, r_7, r_8]\,\mathbf{l} + t_z) - s_x[r_6, r_7, r_8]\,\mathbf{l} \\ s_y t_x - s_x t_y = -c_{j,x}([r_3, r_4, r_5]\,\mathbf{l} + t_y - s_y) + s_x[r_3, r_4, r_5]\,\mathbf{l} \\ \qquad\qquad + c_{j,y}([r_0, r_1, r_2]\,\mathbf{l} + t_x - s_x) - s_y[r_0, r_1, r_2]\,\mathbf{l}. \end{cases}$$

and then write it in matrix form:

$$\begin{bmatrix} r_6 s_y & -r_6 s_x & r_3 s_x - r_0 s_y \\ r_7 s_y & -r_7 s_x & r_4 s_x - r_1 s_y \\ r_8 s_y & -r_8 s_x & r_5 s_x - r_2 s_y \\ 0 & t_z & s_y - t_y \\ -t_z & 0 & -s_x + t_x \\ -s_y + t_y & s_x - t_x & 0 \\ 0 & r_6 & -r_3 \\ -r_6 & 0 & r_0 \\ r_3 & -r_0 & 0 \\ 0 & r_7 & -r_4 \\ -r_7 & 0 & r_1 \\ r_4 & -r_1 & 0 \\ 0 & r_8 & -r_5 \\ -r_8 & 0 & r_2 \\ r_5 & -r_2 & 0 \end{bmatrix}^\top \begin{bmatrix} l_x \\ l_y \\ l_z \\ c_{j,x} \\ c_{j,y} \\ c_{j,z} \\ l_x c_{j,x} \\ l_x c_{j,y} \\ l_x c_{j,z} \\ l_y c_{j,x} \\ l_y c_{j,y} \\ l_y c_{j,z} \\ l_z c_{j,x} \\ l_z c_{j,y} \\ l_z c_{j,z} \end{bmatrix} = \begin{bmatrix} -s_y t_z \\ s_x t_z \\ s_y t_x - s_x t_y \end{bmatrix}. \tag{18}$$

Note the matrix transpose used for space reasons. This equation captures one observation, *i.e.*, one combination of pose i and caster j, but we need to stack equations from multiple observations. To simplify the following step, let us first split Eq. (18) into sub-matrices:

$$\begin{bmatrix} \mathbf{Q}_{ij} \in \mathbb{R}^{3\times 3} & \mathbf{W}_{ij} \in \mathbb{R}^{3\times 12} \end{bmatrix} \begin{bmatrix} \mathbf{l} \in \mathbb{R}^3 \\ \boldsymbol{\theta}_j \in \mathbb{R}^{12} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{ij} \in \mathbb{R}^3 \end{bmatrix}.$$

Note that here the matrix is not transposed.

Let N_p and N_c be the number of target poses and casters. The whole system of stacked equations then is

$$\underbrace{\begin{bmatrix} \mathbf{Q}_{1,1} & \mathbf{W}_{1,1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{Q}_{N_p,1} & \mathbf{W}_{N_p,1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{Q}_{1,2} & \mathbf{0} & \mathbf{W}_{1,2} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{Q}_{N_p,2} & \mathbf{0} & \mathbf{W}_{N_p,2} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{Q}_{1,N_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{1,N_c} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{Q}_{N_p,N_c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{N_p,N_c} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{l} \\ \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_{N_c} \end{bmatrix}}_{\boldsymbol{\theta}} = \underbrace{\begin{bmatrix} \mathbf{b}_{1,1} \\ \vdots \\ \mathbf{b}_{N_p,1} \\ \mathbf{b}_{1,2} \\ \vdots \\ \mathbf{b}_{N_p,2} \\ \vdots \\ \mathbf{b}_{1,N_c} \\ \vdots \\ \mathbf{b}_{N_p,N_c} \end{bmatrix}}_{\mathbf{b}}. \tag{19}$$

We have $3+12N_c$ unknowns since all observations share $\mathbf{l} = [l_x, l_y, l_z]^\top$ and each $\boldsymbol{\theta}_j$ has 12 unknowns. We solve

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\mathbf{A}\boldsymbol{\theta} - \mathbf{b}\|_1 \tag{20}$$

using ℓ_1 minimization to be robust against outliers. For solving this equation we must fulfill

$$\underbrace{3N_pN_c}_{\#\text{equations}} \geq \underbrace{12N_c + 3}_{\#\text{variables}} \quad \Leftrightarrow \quad N_p \geq 4 + \frac{1}{N_c}.$$

Thus, observations from $N_p = 5$ poses suffice to derive a solution regardless of the number of casters, if the problem has a non-degenerate, unique solution. After obtaining $\boldsymbol{\theta}^*$, we disregard the second-order variables, such as $l_x c_{j,x}$ and $l_x c_{j,y}$, and use $\mathbf{c}_j^*$ and $\mathbf{l}^*$ as initialization for minimizing Eq. (17).

Distant light: For distant light, $\mathbf{A}$ is rank deficient: $\operatorname{rank}(\mathbf{A}) = 3 + 12N_c - 1$, because we modeled the light with 3 degrees of freedom (DoF) but distant light only has 2 DoF and the shadow observations can thus be explained by an infinite set of light vectors with same directions but different lengths. Thus, we can unfortunately not use Eq. (18) for distant light but we can *automatically* detect this case, switch to equations for distant light, solve, and switch back to the unified bundle adjustment of Eq. (17). So, again users do not have to choose a light model for their scene.

We detect distant light by constructing matrix $\mathbf{A}$ for nearby light and detecting $\mathbf{A}$'s rank degeneracy: If the ratio between $\mathbf{A}$'s largest and smallest singular value is larger than $4 \cdot 10^4$ (in Sec. 5.1.4 we will give an analysis of this threshold), we switch to the following distant light equations.

For distant light we can write the objective analogous to Eq. (3) (using $\mathbf{l}_i = \mathbf{R}_i^\top \mathbf{l}$):

$$(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times \mathbf{R}_i^\top \mathbf{l} = \mathbf{0}.$$

Keeping the definitions of $\mathbf{c}_j$, $\bar{\mathbf{s}}_{ij}$, $\mathbf{R}_i^\top$, and $-\mathbf{R}_i^\top \mathbf{t}_i$, we can write this as

$$(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times \mathbf{R}_i^\top \mathbf{l} = \begin{bmatrix} c_{j,x} - s_x \\ c_{j,y} - s_y \\ c_{j,z} \end{bmatrix} \times \begin{bmatrix} [r_0, r_1, r_2]\,\mathbf{l} \\ [r_3, r_4, r_5]\,\mathbf{l} \\ [r_6, r_7, r_8]\,\mathbf{l} \end{bmatrix} = \mathbf{0}.$$

Expanding the cross-product yields

$$\begin{cases} 0 = (c_{j,y} - s_y)\,[r_6, r_7, r_8]\,\mathbf{l} - \ c_{j,z}\,[r_3, r_4, r_5]\,\mathbf{l}, \\ 0 = c_{j,z}\,[r_0, r_1, r_2]\,\mathbf{l} - (c_{j,x} - s_x)\,[r_6, r_7, r_8]\,\mathbf{l}, \\ 0 = (c_{j,x} - s_x)\,[r_3, r_4, r_5]\,\mathbf{l} - (c_{j,y} - s_y)\,[r_0, r_1, r_2]\,\mathbf{l}. \end{cases}$$

Setting $\mathbf{l} = [l_x, l_y, 1]^\top$ to reduce $\mathbf{l}$ to two DoF yields

$$\begin{cases} 0 = (c_{j,y} - s_y)(r_6 l_x + r_7 l_y + r_8) - c_{j,z}(r_3 l_x + r_4 l_y + r_5), \\ 0 = c_{j,z}(r_0 l_x + r_1 l_y + r_2) - (c_{j,x} - s_x)(r_6 l_x + r_7 l_y + r_8), \\ 0 = (c_{j,x} - s_x)(r_3 l_x + r_4 l_y + r_5) - (c_{j,y} - s_y)(r_0 l_x + r_1 l_y + r_2), \end{cases}$$

which we can rewrite as

$$\begin{cases} -s_y r_8 = -\,c_{j,y}(r_6 l_x + r_7 l_y + r_8) + s_y(r_6 l_x + r_7 l_y) \\ \qquad\qquad + c_{j,z}(r_3 l_x + r_4 l_y + r_5), \\ s_x r_8 = -\,c_{j,z}(r_0 l_x + r_1 l_y + r_2) \\ \qquad\qquad + c_{j,x}(r_6 l_x + r_7 l_y + r_8) - s_x(r_6 l_x + r_7 l_y), \\ s_y r_2 - s_x r_5 = -\,c_{j,x}(r_3 l_x + r_4 l_y + r_5) + s_x(r_3 l_x + r_4 l_y) \\ \qquad\qquad + c_{j,y}(r_0 l_x + r_1 l_y + r_2) - s_y(r_0 l_x + r_1 l_y). \end{cases}$$

This can be rewritten in matrix form as

$$\underbrace{\begin{bmatrix} r_6 s_y & -r_6 s_x & r_3 s_x - r_0 s_y \\ r_7 s_y & -r_7 s_x & r_4 s_x - r_1 s_y \\ 0 & r_8 & -r_5 \\ -r_8 & 0 & r_2 \\ r_5 & -r_2 & 0 \\ 0 & r_6 & -r_3 \\ -r_6 & 0 & r_0 \\ r_3 & -r_0 & 0 \\ 0 & r_7 & -r_4 \\ -r_7 & 0 & r_1 \\ r_4 & -r_1 & 0 \end{bmatrix}^\top}_{\mathbf{A}_{ij}} \underbrace{\begin{bmatrix} l_x \\ l_y \\ c_{j,x} \\ c_{j,y} \\ c_{j,z} \\ l_x c_{j,x} \\ l_x c_{j,y} \\ l_x c_{j,z} \\ l_y c_{j,x} \\ l_y c_{j,y} \\ l_y c_{j,z} \end{bmatrix}}_{\boldsymbol{\theta}_j} = \underbrace{\begin{bmatrix} -s_y r_8 \\ s_x r_8 \\ s_y r_2 - s_x r_5 \end{bmatrix}}_{\mathbf{b}_{ij}} . \tag{21}$$

Again, note the matrix transpose for space saving. Split into sub-matrices, this reads

$$\begin{bmatrix} \mathbf{Q}_{ij} \in \mathbb{R}^{3\times 2} & \mathbf{W}_{ij} \in \mathbb{R}^{3\times 9} \end{bmatrix} \begin{bmatrix} \mathbf{l} \in \mathbb{R}^2 \\ \boldsymbol{\theta}_j \in \mathbb{R}^9 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{ij} \in \mathbb{R}^3 \end{bmatrix}.$$

Multiple observations get stacked in a similar manner to Eq. (19) and solved using Eq. (20). For solving, we must fulfill

$$\underbrace{3N_p N_c}_{\#\text{equations}} \geq \underbrace{9N_c + 2}_{\#\text{variables}} \quad \Leftrightarrow \quad N_p \geq 3 + \frac{2}{3N_c}.$$

Thus, 4 poses suffice regardless of the number of casters. Since the formulation for distant light (Eq. (20) in conjunction with Eqs. (19) and (21)) returns a light direction for $\mathbf{l}^*$ but the unified bundle adjustment requires a light position, we need to convert from a direction to a position. We start at one of the casters and move the light source out very far in space:

$$\mathbf{l}^*_{\text{position}} = \mathbf{c} + \kappa\, h_{\mathbf{c}}\, \mathbf{l}^*_{\text{direction}},$$

where $\mathbf{c}$ is an arbitrary caster's position in world coordinates (obtained through the relaxation method), κ is a large constant (we use $\kappa = 10^{10}$), and $h_{\mathbf{c}}$ is the caster's height above the shadow receiver plane (which takes the scale of the scene into account).

4.3 Shadow correspondence search

For Eqs. (17), (18), and (21) we need to assign the same index j to all shadows $\bar{\mathbf{s}}_{ij}$ that belong to the same caster $\mathbf{c}_j$ in different images $\{\mathbf{I}_i\}$. This correspondence problem is easier to solve if the input is structured, *i.e.*, we have information about the relation between the input images. If the input is a video for example, then we can track shadows over consecutive frames. However, it is clearly desirable to also be able to handle unstructured input just like regular SfM. Unstructured input occurs if separate images or multiple videos are captured or if we record a video but some tracks break (due to lens flare, noise, shadows leaving the field of view, *etc.*).

In Sec. 3.2, we developed the basis for finding shadow correspondences in unstructured input: Shadows on a plane from a moving point light obey correspondence conditions with specialized fundamental matrices, tritensors, and quadtensors, which we will use for finding shadow correspondences and rejecting shadow misdetections. Since quadtensors are impractical because correspondence search in four views is very costly (unless some strong prior knowledge about possible correspondences narrows the search down), we will only cover fundamental matrices and tritensors in the following.

Formally, shadow correspondence search means we need to find permutations that match corresponding shadows between images. Let $\mathbf{S}$ be the shadows $\{\tilde{\mathbf{s}}_i\}$ and $\mathbf{S}'$ be the shadows $\{\tilde{\mathbf{s}}'_i\}$ stacked horizontally into matrices. For fundamental matrices we seek to find

$$\underset{\mathbf{P}',\mathbf{F}}{\operatorname{argmin}} \sum_{i\in\{1,\dots,N_c\}} \left\| ((\mathbf{S}'\mathbf{P}')_{:,i})^\top \mathbf{F}\, (\mathbf{S})_{:,i} \right\| \tag{22}$$

s.t. $\mathbf{P}'$ is a permutation matrix.

(Recall that $(\cdot)_{:,i}$ extracts a matrix's i^{th} column.) For tritensor estimation we need to solve

$$\underset{\substack{\mathbf{P}',\mathbf{P}'',\\ \mathcal{T}=[\mathbf{T}_1,\mathbf{T}_2,\mathbf{T}_3]}}{\operatorname{argmin}} \sum_{i\in\{1,\dots,N_c\}} \Bigg\| [(\mathbf{S}'\mathbf{P}')_{:,i}]_\times \cdot \Big(\sum_{j\in\{1,2,3\}} (\mathbf{S})_{j,i} \mathbf{T}_j \Big) [(\mathbf{S}''\mathbf{P}'')_{:,i}]_\times \Bigg\|$$

s.t. $\mathbf{P}'$ and $\mathbf{P}''$ are permutation matrices. (23)

We cannot use feature descriptors to narrow the search down since we want the shadows to be very small (to make the shadow "center" precisely localizable) and thus cannot vary a caster's shape enough to make its shadows clearly distinguishable from other casters' shadows. We thus find the minimizer of Eq. (23) with branch and bound on the detected shadow points without descriptors. This procedure returns a significant fraction of wrong correspondences. Inspired by the removal of inconsistent feature tracks in SfM (see, *e.g.*, PhotoTourism [38, Sec. 4.1]) we check the consistency of correspondences across multiple images.

Correspondence consistency check: We work on two image pools; images with established shadow correspondences ("established pool") and images with unknown correspondences ("unknown pool"). The established pool is initialized with a random unknown image and

the goal is to move as many images as possible from the unknown to the established pool.

For *fundamental shadow matrices* we work in two phases: In phase 1 we randomly pick an image $\mathbf{I}_e$ from the established pool and k images $\mathbf{I}_{i_1}, ..., \mathbf{I}_{i_k}$ from the unknown pool. Let ${}^{a\to b}m(\mathbf{s}_i, \mathbf{s}_j)$ be a binary function that is true iff Eq. (22)'s minimizer for images $\mathbf{I}_a$ and $\mathbf{I}_b$, matches shadow $\mathbf{s}_i$ in $\mathbf{I}_a$ with shadow $\mathbf{s}_j$ in $\mathbf{I}_b$. Then, if

$$\forall\, \mathbf{s}_{j_0} \text{in } \mathbf{I}_e,\ \mathbf{s}_{j_1} \text{in } \mathbf{I}_{i_1},\ \mathbf{s}_{j_2} \text{in } \mathbf{I}_{i_2}, \ldots,\ \mathbf{s}_{j_k} \text{in } \mathbf{I}_{i_k}:$$
$${}^{e\to i_1}m(\mathbf{s}_{j_0}, \mathbf{s}_{j_1}) \wedge {}^{i_1\to i_2}m(\mathbf{s}_{j_1}, \mathbf{s}_{j_2}) \wedge {}^{i_2\to i_3}m(\mathbf{s}_{j_2}, \mathbf{s}_{j_3}) \wedge \ldots \wedge$$
$${}^{i_{k\text{-}2}\to i_{k\text{-}1}}m(\mathbf{s}_{j_{k\text{-}2}}, \mathbf{s}_{j_{k\text{-}1}}) \wedge {}^{i_{k\text{-}1}\to i_k}m(\mathbf{s}_{j_{k\text{-}1}}, \mathbf{s}_{j_k}) \Rightarrow {}^{e\to i_k}m(\mathbf{s}_{j_0}, \mathbf{s}_{j_k})$$

holds (*i.e.*, correspondences through the chain of images are consistent with the direct correspondences from the first to the last image), we move $\{\mathbf{I}_{i_1}, ..., \mathbf{I}_{i_k}\}$ to the established pool. To make the constraint relatively strict, our implementation uses $k = 3$. Once half of all images are in the established pool, we switch to phase 2.

In phase 2 we assume all images in the established pool to be consistent. Thus, if we consider one unknown image and one shadow caster, all shadows in all established images that correspond to that particular caster should match the same shadow in the unknown image; and this should hold for all shadow casters. We pick a random unknown image, verify this criterion, and if more than half of all established images agree on their correspondences to the unknown image, it is moved to the established set. Elsewise, it is discarded. Phase 2 ends when the unknown pool is empty.

For *shadow tritensors*, we also work in two phases. In phase 1 we randomly select one unknown image as target, one established image, and l unknown images for testing ($l = 15$ in our implementation), iterate over the test images, and use Eq. (23) to compute the trifocal tensors for the target image, the established image and the current test image. If more than $\frac{l}{2}$ of the tensors agree on the correspondences between target and established image, we move the target image to the established pool. Once half of all images are in the established pool, we switch to phase 2 which works equivalent to phase 2 for fundamental matrices.

4.4 Implementation details

To obtain our target's pose $\{[\mathbf{R}_i|\mathbf{t}_i]\}$, we print ArUco markers [13] on a piece of paper, attach it to the target (see Fig. 1, *left*), and use OpenCV 3D pose estimation [5]. Our shadow casters are off-the-shelf pins with a length of $\sim 30\,\text{mm}$ and a head diameter of $\sim 3\,\text{mm}$, which is big enough to easily detect them and small enough to accurately localize them. We can place the pins arbitrarily without measuring their position since the bundle adjustment estimates their position.

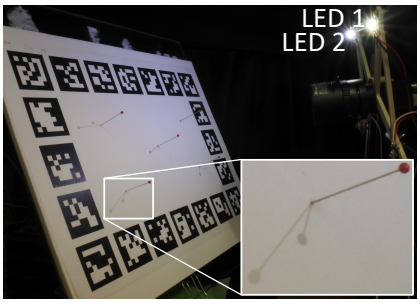

Fig. 6 Two lights casting two shadows per pin.

For shadow detection we developed a simple template matching scheme. For the templates we generated synthetic images of shadows consisting of a line with a circle at the end. To deal with varying projective transformations we use 12 rotation angles with 3 scalings each. We match the templates after binarizing the input image to extract shadowed regions more easily. Further we use the color of the pin heads to distinguish between heads and head shadows.

4.5 Estimating multiple lights jointly

Our method can estimate multiple lights jointly. It can work with multiple lights that were captured

a) simultaneously (see Fig. 6) or
b) separately, *i.e.*, we switch on one light at a time and capture its shadows. It is, of course, necessary to use the same calibration target for all lights so that all caster positions stay constant in calibration target coordinates.

From the bundle adjustment's viewpoint both cases are equivalent since applying the calibration target poses transforms all shadow positions into the same coordinate system, namely the target's – no matter whether they were captured simultaneously or separately.

For both cases the benefit over single light calibration is improved accuracy: More captured data puts more constraints on the shadow caster positions, therefore the caster position estimates will be more accurate and as a consequence the light position estimates will also be more accurate. An additional benefit of case a) is that simultaneous light capturing saves time. Note that, although our equations set no theoretical limit for the number of lights to be simultaneously estimated, there are very strong practical limits: Since we need to reliably detect each shadow point in the imagery, we are limited to very few lights in practice due to the low contrast and overlap of shadows under too many lights.

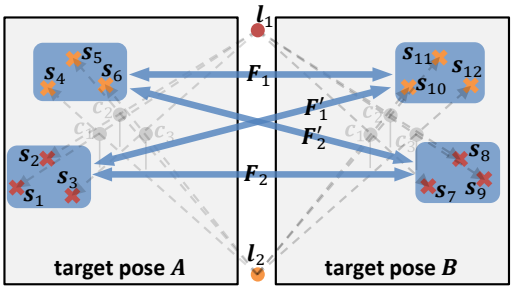

Fig. 7 A scene with 2 lights and 3 casters. In each pose, fundamental matrices enable us to split the 6 shadows into 2 sets (blue rectangles) of 3 shadows, each set corresponding to one of the lights. However, the fundamental matrices cannot match these shadow sets across poses due to an ambiguity that allows connecting any pair of sets (blue arrows).

In Sec. 4.3 we discussed a correspondence problem: Finding all shadows $\bar{\mathbf{s}}_{ij}$ that belong to the same caster $\mathbf{c}_j$ in different images i. Multiple lights entail another correspondence problem: finding all shadows from the same light to couple the correct shadows $\bar{\mathbf{s}}_{i,j,k}$, casters $\mathbf{c}_j$ and lights $\mathbf{l}_k$ in our equations. For separately captured lights this is trivially to solve since we know which light was on when a particular shadow was captured.

For simultaneously captured light this is harder. Let us consider the example in Fig. 7: We have a scene with 2 lights and 3 casters. The shadow projection process is hinted at with transparent casters and arrows. In each of the two target poses we can use fundamental shadow matrix estimation to separate the 6 shadows into 2 sets of 3 shadows, each corresponding to one of the lights. We, however, cannot match these sets of 3 shadows *across images*. We can find fundamental shadow matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ that connect the shadow set $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ with $\{\mathbf{s}_7, \mathbf{s}_8, \mathbf{s}_9\}$ and $\{\mathbf{s}_4, \mathbf{s}_5, \mathbf{s}_6\}$ with $\{\mathbf{s}_{10}, \mathbf{s}_{11}, \mathbf{s}_{12}\}$ but we can also find matrices $\mathbf{F}'_1 \neq \mathbf{F}_1$ and $\mathbf{F}'_2 \neq \mathbf{F}_2$ that connect $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ with $\{\mathbf{s}_{10}, \mathbf{s}_{11}, \mathbf{s}_{12}\}$ and $\{\mathbf{s}_4, \mathbf{s}_5, \mathbf{s}_6\}$ with $\{\mathbf{s}_7, \mathbf{s}_8, \mathbf{s}_9\}$. This is because fundamental shadow matrices cannot distinguish whether the shadow movement resulted from changing the calibration target pose or from changing the light position.

When using only fundamental matrices, tritensors or quadtensors, this is a fundamental ambiguity and not an implementation problem. To overcome this, we require users to capture a video and track each shadow from frame to frame. Thereby we can follow the set $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ transitioning into $\{\mathbf{s}_7, \mathbf{s}_8, \mathbf{s}_9\}$ and then assign the same light index k to them.

To sum up, estimating multiple lights jointly increases the calibration accuracy. If multiple lights are captured separately, the data can be completely unstructured. If the lights are captured simultaneously, additional information is needed to resolve an ambiguity. In this case we expect a video as input.

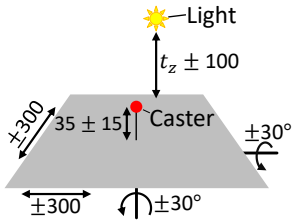

Fig. 8 The arrows show the value ranges of our simulation experiments.

Table 2 Estimation error (mean of ten random trials) in a synthetic, noise-free setting.

	t_z	N_c	mean absolute/angular error of light source positions/directions
near light	500	2	6.4×10^{-14}
	500	5	9.5×10^{-14}
	500	10	5.4×10^{-14}
	1000	2	3.5×10^{-13}
	1000	5	7.0×10^{-14}
	1000	10	2.6×10^{-13}
distant light	∞	2	1.2×10^{-12} deg.
	∞	5	2.4×10^{-15} deg.
	∞	10	1.4×10^{-12} deg.

5 Evaluation

We now assess our method's accuracy using simulation experiments (Sec. 5.1) and real-world scenes (Sec. 5.2).

5.1 Simulation

For all following simulated experiments we randomly sampled target poses, caster positions and light positions (the latter only for near light conditions) from uniform distributions within the ranges shown in Fig. 8. The casters were randomly placed on a target of size 200×200. For distant light, we sampled the light direction's polar angle θ from $[0°, 45°]$.

We evaluated the absolute/angular error of estimated light positions/directions while varying the distance t_z between light and calibration target and the number of casters N_c. Table 2 shows that each configuration's mean error is ≥ 14 orders of magnitude smaller than the scene extent, confirming that our method solves the joint estimation of light position/direction and shadow caster positions accurately in an ideal setup. In our experience, the difference between using the unified light model and using one of the specialized models (nearby or distant) is negligible.

In practice, light source estimates will be deteriorated by two main error sources: (1) Shadow localization and (2) the marker-based target pose estimation.

5.1.1 Shadow localization errors

To analyze the influence of shadow localization, we perturbed the shadow positions with Gaussian noise. In this and the following experiments we set $t_z = 500$ so that our synthetic scenes' proportions match those of our main real-world scene `E1` (see Fig. 13) that we introduce later, with 1 synthetic length unit corresponding to 1 mm. Figure 9 shows the estimation accuracy of the convex relaxation (Eq. (20)) compared to the accuracy of full bundle adjustment (Eq. (17) after initialization with convex relaxation) for near and distant light, varying N_p and N_c, and varying standard deviation σ for the shadow position noise.

Figure 9's top row confirms that larger shadow position noise results in larger error and full bundle adjustment mitigates the error compared to solving only the convex relaxation. Increasing the number of casters or target poses makes Eqs. (20) and (17) more overconstrained and thus reduces the error from noisy shadow locations as Figure 9's middle and bottom row confirm.

5.1.2 Target pose estimation errors

To simulate errors in the target pose estimation, we performed an experiment where we added Gaussian noise to the target's roll, pitch, and yaw. Figure 10's top row shows that the error is again higher for stronger noise and the bundle adjustment mitigates the error of the convex relaxation. In Fig. 10's middle and bottom row we increased the number of casters and target poses again. Bundle adjustment and increasing the number of poses reduce the error, but increasing the number of casters does not. This is not surprising since adding constraints to our system only helps if the constraints have independent noise. Here, the noises for all shadows $\bar{\mathbf{s}}_{i,j}$ of the same pose i stem from the same pose noise and are thus highly correlated. Thus, increasing the number of poses is more important for improving the accuracy than increasing the number of casters.

5.1.3 Combined shadow localization and target pose estimation errors

Previously, we studied the effect of shadow localization errors and target pose errors separately. In this section we show simulation results where we added both types of noise at the same time. Comparing the top rows of Figs. 9 and 10, we can see that shadow localization noise causes errors roughly twice as big as those from target pose noise. In this experiment we thus set the standard deviation for shadow localization noise to $\sigma_{\text{shadows}} = 0.01$ and for target pose noise to $\sigma_{\text{pose}} = 0.005$.

For the number of shadow casters varying from 1 to 9 and the number of poses varying from 5 to 100, Fig. 11 shows color-coded (log-scale) median error in the *top row* and standard deviation in the *bottom row*. Again, bundle adjustment and more poses and casters decrease the error. If the application at hand dictates one of the two parameters, *e.g.*, if time restrictions forbid increasing N_p beyond 20, this can always be countered by increasing the other parameter. Even though the minimal conditions for solving the calibration are 1 caster and 4 or 5 poses, the data suggests that users should probably always use 3 or more casters and 20 or more poses in practice.

5.1.4 Discerning nearby from distant light

As discussed in Sec. 4.2, we can discern nearby and distant light based on the condition number of $\mathbf{A}$. In a noise-free setup, the condition number becomes larger than 10^{15} for distant light and smaller than 10^5 for near light even in the hardest setting: $N_p = 5$. Thus, near and distant light can easily be discerned.

For noisy input, our method requires more poses for a clear distinction: Figure 12 shows histograms of the condition numbers for $N_p = 5$, 20, and 50. We can see that for $N_p = 20$ and 50, near and distant light can clearly be separated using a threshold of $4 \cdot 10^4$ while for $N_p = 5$ the blue histogram of near light extends well beyond $4{\cdot}10^4$ and thus cannot be discerned from distant light. Based on this, we set the threshold to $4 \cdot 10^4$ and suggest working with $N_p \geq 20$ target poses. As already discussed, we recommend increasing N_p as the primary way of error reduction. 20–50 poses are captured rather quickly.

5.2 Real-world experiments

We created 4 real-world environments, see Fig. 13. In all experiments we calibrated the intrinsic camera parameters beforehand and removed lens distortions.

Environments `E1` and `E2` have near light, and `E3` and `E4` have distant light. In `E1` we fixed four LEDs to positions around the camera with a 3D printed frame and calculated the LED's ground truth locations from the frame geometry. We used a `FLIR FL2G-13S2C-C` camera with a resolution of 1280×960. In `E2` we separately calibrated two smartphones (Sony Xperia XZ1 and Huawei P9) to potentially open up the path for inexpensive, end user-oriented physics-based modeling with phones. Both phones have a 1920×1080 px camera and an LED light. We assumed that LED and camera are in a plane orthogonal to the camera axis and through the optical center, and measured the distance

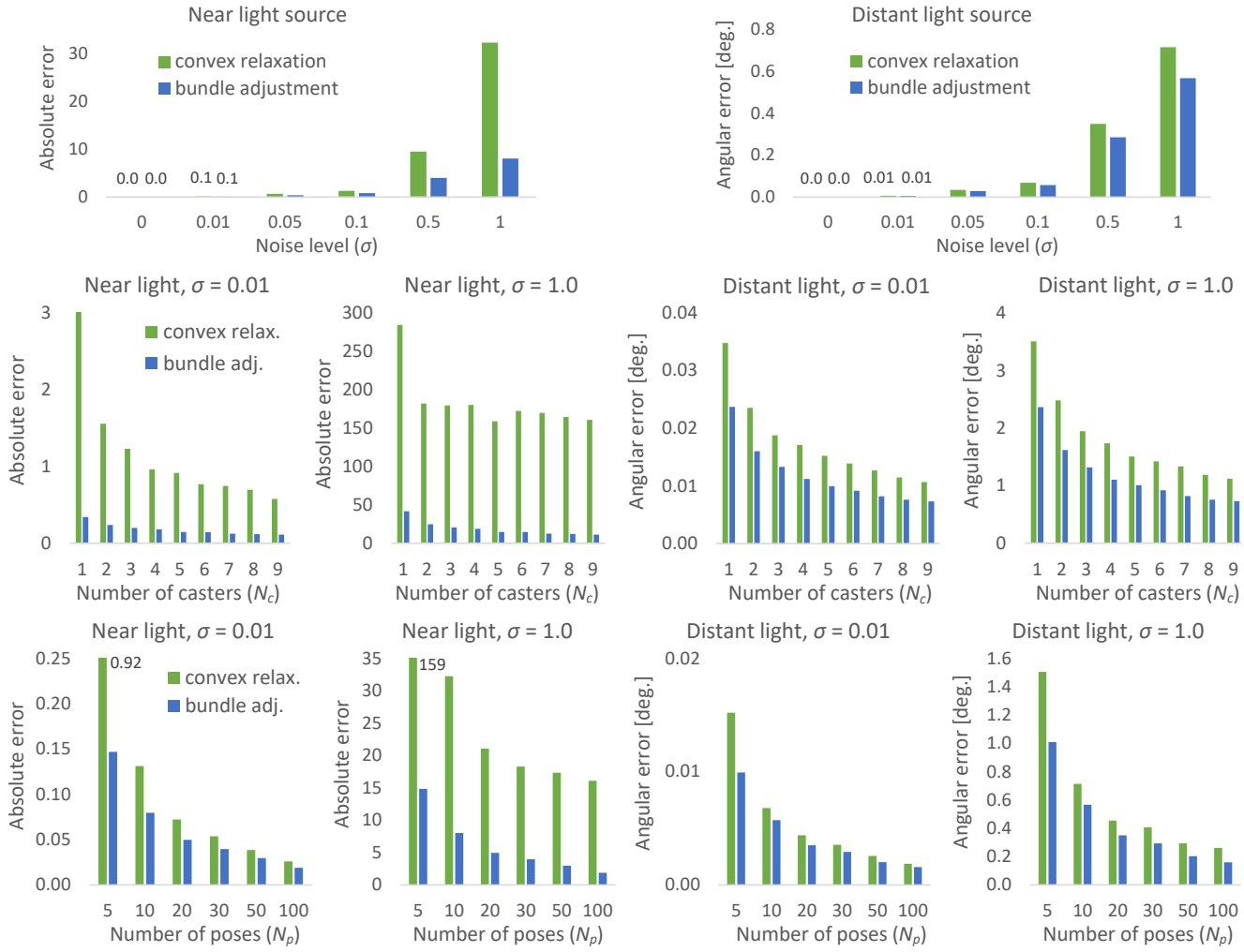

Fig. 9 Estimation error for synthetic near and distant light with Gaussian noise added to the shadow positions. Each data point is the median of 500 random trials. *Top row:* $N_p = 10$ and $N_c = 5$. The noise's standard deviation σ is on the x-axis. *Middle row:* $N_p = 5$ and N_c is on the x-axis. *Bottom row:* $N_c = 5$ and N_p is on the x-axis.

between LED and camera to obtain the ground truth. In `E3` we placed the target under direct sun light and took datasets at three different times to obtain three light directions. In `E4` a floodlight was fixed about 3 m away from the target to approximate distant lighting. In both `E3` and `E4` we used a `Canon EOS 5D Mark IV` with a 35 mm single-focus lens and a resolution of 6720×4480 and obtained the ground truth light directions from measured shadow caster positions and hand-annotated shadow positions. In all environments, we used an A5-sized calibration target with $N_c = 5$ pins.

Table 3 shows the achieved estimation results. The light position errors are 1.5 % of the target-camera distance for `E1` and 1.2 % for `E2`, the light direction errors are $\sim 1°$, and the caster position errors are < 2.5 mm. Figure 14 shows how increasing the number of target poses monotonously decreases the estimation error on two of our real-world scenes.

Table 3 Estimation errors in our four real-world scenes. Percentages in the mean light error are relative errors compared to the target-to-camera distance.

Scene	number of experiments	light error		caster error	
		mean (relative)	std. deviation	mean	std. deviation
`E1`	4 lights	7.7 mm (1.5 %)	1.1 mm	1.1 mm	0.46 mm
`E2`	2 phones	3.7 mm (1.2 %)	0.27 mm	0.87 mm	0.39 mm
`E3`	3 sun positions	1.2 deg.	0.42 deg.	1.4 mm	0.42 mm
`E4`	1 light	0.88 deg.	–	2.4 mm	0.48 mm

5.3 Fundamental shadow matrix vs. shadow tritensor

We now analyze the shadow correspondence search and verification based on fundamental shadow matrices and trifocal shadow tensors described in Sec. 4.3. We picked 200 images from `E1-1` and obtained ground truth correspondences through video tracking. The fundamental matrix-based method returned correspondences for 145 images of which 143 were correct and the tritensor-

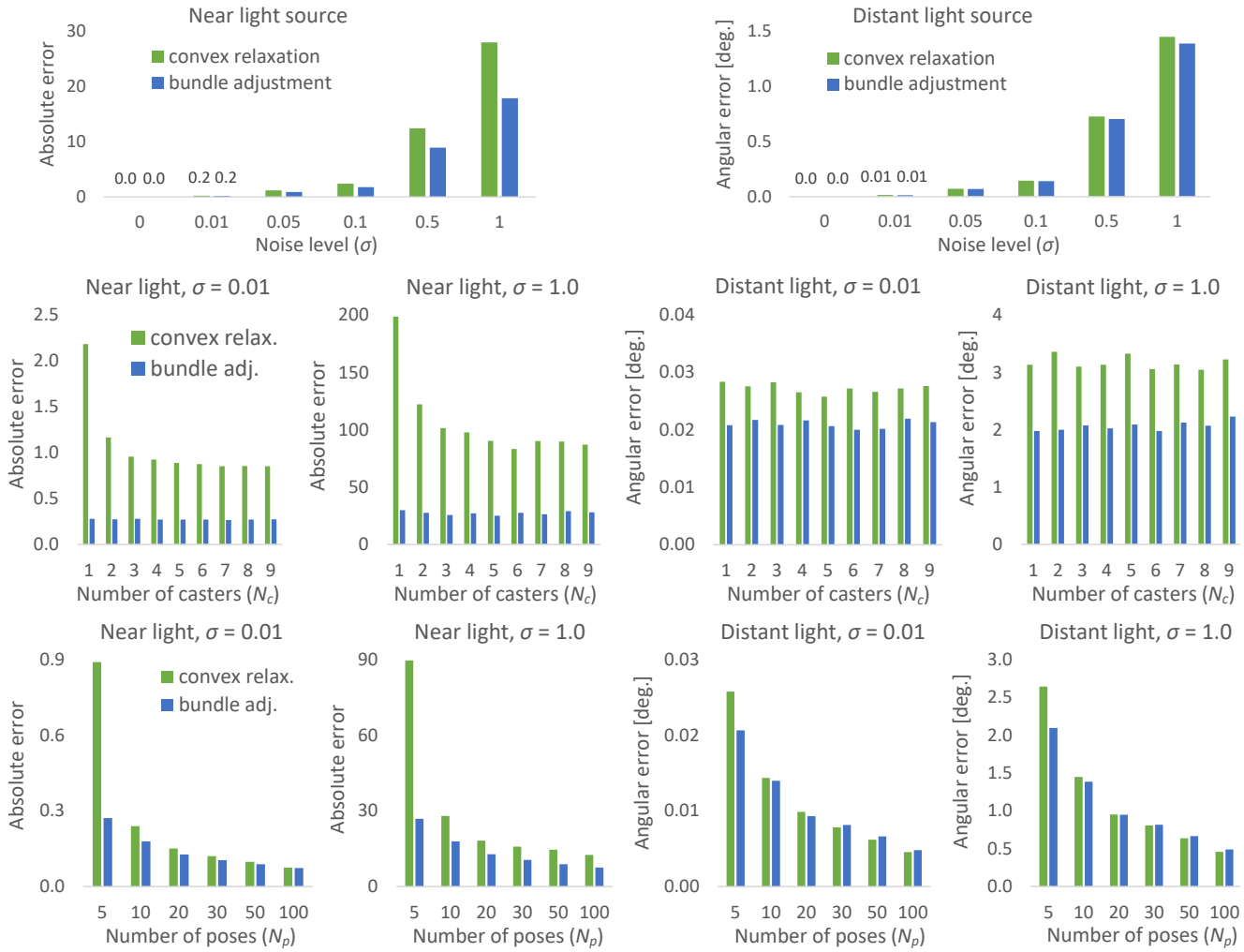

Fig. 10 Estimation error for synthetic near and distant light with Gaussian noise added to the target orientation (in deg.). Each data point is the median of 500 random trials. *Top row:* $N_p = 10$ and $N_c = 5$. The noise's standard deviation σ is on the x-axis. *Middle row:* $N_p = 5$ and N_c is on the x-axis. *Bottom row:* $N_c = 5$ and N_p is on the x-axis.

based method returned 152 images of which 151 were correct. Thus, the found correspondences are almost perfect (because we chose all consistency check criteria and parameters to be rather strict to discard a lot of frames and prefer to capture more frames instead), the number of matched images was by far sufficient for the subsequent calibration steps, and both consistency check methods performed almost identical. Note that we cannot deduce from this that fundamental shadow matrices and shadow tritensors themselves perform equally, because they are embedded in different consistency checks. We prefer working with fundamental matrices because their correspondence search runs an order of magnitude faster.

In certain situations users may, however, prefer tritensors over fundamental matrices or vice versa: The tritensor restricts the positions of shadows more and thus works well if the shadow detector detects the correct shadow but also has misdetections in the correct shadow's vicinity. The tritensor can rule out almost all of those misdetections. If there are no misdetections but the correct shadow's detected position has large noise, the tritensor may be too unforgiving and a user may prefer the fundamental matrix instead.

5.4 Estimating multiple lights simultaneously

Capturing and estimating `E1`'s two top lights (reliably detecting shadows of more than two lights requires a better detector) simultaneously as described in Sec. 4.5 reduces the mean light and caster position errors from 6.9 mm and 1.1 mm to 5.1 mm and 0.6 mm, respectively.

As mentioned, we can also jointly calibrate lights whose image sets were captured separately, as long as they were all captured with the same calibration target. This necessitates the use of fundamental shadow

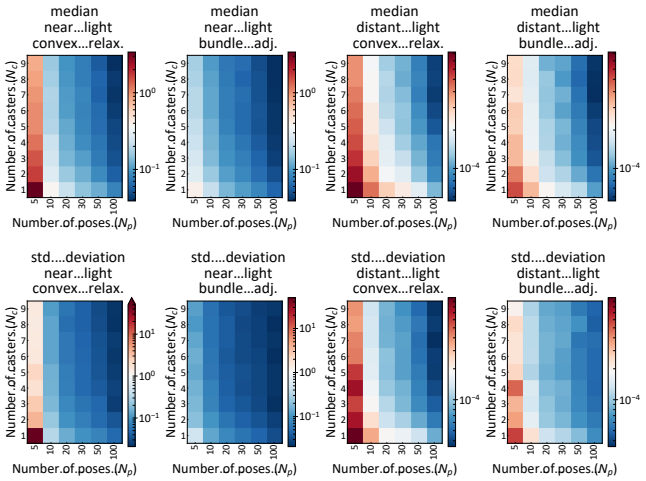

Fig. 11 Estimation error for synthetic near and distant light with Gaussian noise added to shadow positions ($\sigma_{\text{shadows}} = 0.01$) and target orientation in degrees ($\sigma_{\text{pose}} = 0.005$). N_p is on the x-axis and N_c is on the y-axis. For each data point we performed 500 random trials. *Top row:* The median of the absolute error (near light) / angular error in degrees (distant light). *Bottom row:* The error's standard deviation.

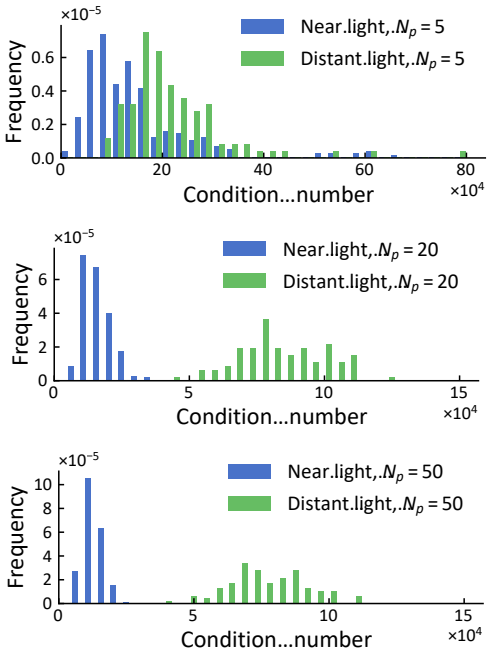

Fig. 12 Histograms of **A**'s condition number for $N_p = 5$ (*top*), $N_p = 20$ (*middle*), and $N_p = 50$ (*bottom*). Each histogram is the aggregate of 100 random trials with $\sigma_{\text{shadows}} = 0.01, 0.1, 1.0$. As in the previous experiment of Sec. 5.1.3 we set $\sigma_{\text{pose}} = \frac{1}{2}\sigma_{\text{shadows}}$.

matrices / shadow tritensors since there is (currently) no other method for matching shadows across image sets. For `E1`, joint calibration decreased the errors as shown in Tab. 4.

We can even jointly estimate *nearby and distant* light that was captured separately. We put the data from nearby and distant light separately through the convex relaxation, run bundle adjustment on them separately, and then run joint bundle adjustment. Picking one light from `E1` and one light from `E3`, this procedure reduced the light position and direction errors and mean caster position error from 7.1 mm, 0.71 deg. and 1.21 mm to 3.8 mm, 0.53 deg. and 1.15 mm, respectively.

Table 4 Average estimation errors (all units in mm) in scene `E1` for 2, 3, and 4 lights captured separately and calibrated separately or simultaneously.

number of lights	calibration	light error mean	light error stdev.	caster error mean	caster error stdev.
2	separately	6.91	**0.20**	1.13	0.51
2	simultaneously	**6.29**	1.23	**0.79**	**0.25**
3	separately	7.11	**0.33**	1.18	0.47
3	simultaneously	**5.2**	0.54	**1.03**	**0.29**
4	separately	7.72	1.09	1.10	0.46
4	simultaneously	**7.40**	**0.58**	**1.06**	**0.30**

Table 5 Estimation error in scene `E1` (averaged over `E1`'s 4 lights) for *Ours* and *Shen*.

Method	mean of light error	stdev. of light error
Ours, shadows hand-annotated	9.45 mm	1.06 mm
Ours, shadows detected	15.4 mm	7.45 mm
Shen, highlights hand-annotated	18.6 mm	5.33 mm

5.5 Comparison with existing method

To put our method's accuracy into perspective, Ackermann *et al.* [1] achieved accuracies of about 30–70 mm on scenes 2–3 times as big as ours despite also minimizing reprojection error (thus being theoretically more accurate than methods based on simpler triangulation schemes according to Hartley [17]) with very careful experiment execution. We believe this is at least partially due to their usage of spheres.

In this section we compare our calibration method – denoted as *Ours* – with an existing method. Because of Ackermann's achieved accuracy we ruled out spheres and compared to a reimplementation of a state-of-the-art method based on a planar mirror [35] – denoted as *Shen*. Their method observes the specular reflection of the point light in the mirror, also models the mirror with perspective projection and infers parameters similar to camera calibration. In our implementation of *Shen* we again used ArUco markers to obtain the mirror's pose and we annotated the highlight positions manually. For a fair comparison we also annotated the shadow positions for our method manually. For all hand annotations, we zoomed in on the spec-

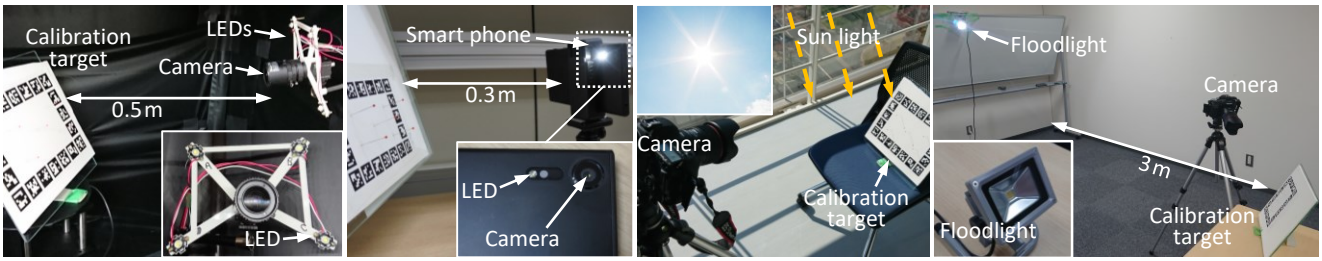

Fig. 13 Our real-world environments. `E1` has 4 LEDs fixed around the camera. In `E2` we use a smartphone's camera and LED. In `E3` we observe the target under sun light. `E4` has a floodlight fixed about 3 m away from the target.

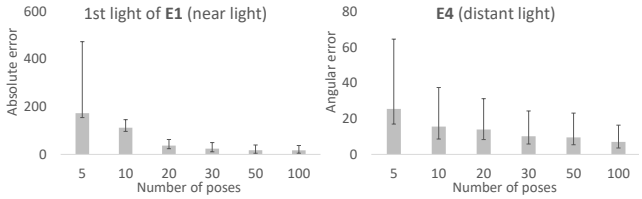

Fig. 14 Estimation error for the first light of scene `E1` and for scene `E4`. For each scene we captured 200 images, randomly picked N_p images from these, and estimated the light and caster positions. The gray bars and error bars represent median and interquartile range of 100 random iterations of this procedure.

ular/shadow area of an image and carefully picked up the specularity's/shadow's centroid.

In both methods we used an A4-sized target and observed the target while varying its pose ~ 500 mm away from the light source. We captured 30 poses for each method and annotated the shadows/reflections. Table 5 shows the estimation error of light source positions for *Ours* and *Shen* in scene `E1`. *Ours* with hand-annotated shadows as well as detected shadows outperforms *Shen* with annotated highlights.

6 Conclusions

Theoretic contributions: In this paper we explored the connections between point lights and pinhole cameras: Single view shadow projection from point lights follows the same principles as pinhole camera projection but with more specialized projection matrices. We devised a unified light model that smoothly interpolates between projection from a nearby light and distant light and thereby spares users having to choose between light models. As a consequence of point lights behaving like pinhole cameras, we saw that multi-view shadow correspondences follow the principles of epipolar geometry. Their fundamental matrices, trifocal and quadrifocal tensors have specialized shapes that allow estimating them from as few as 2 correspondences, and there is no general degeneracy in estimating fundamental shadow matrices from coplanar scene points. Shadow matrices/tensors allow us to establish point correspondences from unstructured sets of images without the use of tracking or feature matching. We further saw, that point lights and shadow caster positions can be simultaneously estimated using structure from motion and bundle adjustment.

We want to add a thought on calibration target design: Ackermann [1] pointed out that, analogous to the large depth uncertainty in narrow-baseline stereo, narrow baseline calibration targets such as Powell's [28] have a large light position uncertainty along the light direction. This can be decreased by either building a static wide-baseline calibration target, or by moving the target in the scene as we do. So, again our method is strongly connected to SfM where camera movement is key to reducing depth uncertainty.

Experimental results: Our noise-free simulation experiments showed that our formulation is correct and the solution method derives accurate estimates with negligible numerical errors. Thus, the solution quality is rather governed by the inaccuracy of target pose estimation and shadow detection. We showed on synthetic and real-world scenes that even with these inaccuracies our method accurately estimates light source positions/directions with measurements from a sufficient number of shadow casters and (more importantly) target poses, which can easily be collected by moving the proposed calibration target in front of the camera. Further, we showed that we can increase the calibration accuracy by estimating multiple lights simultaneously. Regarding the choice between fundamental shadow matrices and trifocal shadow tensors, we saw that both yield approximately equally accurate results.

A comparison with a state-of-the-art method based on highlights on a mirror plane showed our method's superior accuracy. We believe the reason lies in our pin shadows' accurate localizability. As discussed in Sec. 2, highlights are hard to localize accurately. In contrast, our pin shadows do not "bleed" into their neighborhood

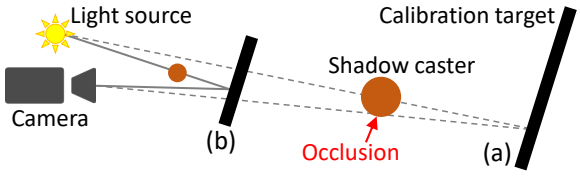

Fig. 15 (a) With a small baseline between camera and light, the caster may occlude the shadow as seen from the camera. (b) To solve this, we use a smaller shadow caster, bring the target closer to the camera and make the target smaller so the camera can capture it fully.

and we can easily control their size through the pin head size. If higher localization accuracy is required, one can choose pins smaller than ours.

Practical implications: In contrast to related work, our method requires no tedious, error-prone hand annotations of, *e.g.*, sphere outlines, no precisely fabricated objects such as precise spheres, and no precise measurements of, *e.g.*, sphere positions. Users need not even choose between the two different light models (nearby and distant) since our calibration method infers this automatically. Further, shadow matrices/tensors even allow unstructured image sets and not just videos to be used as input for the calibration. The construction of our calibration target is simple, fast and cheap and most calibration steps (*e.g.*, target pose estimation and shadow detection/matching) run automatically. The only manual interaction – capturing images or a video while moving the target – is simple. To our knowledge no other method combines such simplicity and accuracy.

Limitations: Our method cannot be used for scenes where light and camera are so close together that the caster occludes the image of the shadow (see Fig. 15(a)). The solution is to effectively increase the baseline between camera and light by using a smaller target and bringing it closer to the camera, as shown in Fig. 15(b).

Future work: It may be possible to alleviate the occlusion problem above with a shadow detection that handles partial occlusions. Further, we want to analyze degenerate cases where our equations are rank deficient, *e.g.*, a target with one caster being moved such that its shadow stays on the same spot.

The source code for this project can be downloaded from: `github.com/hiroaki-santo/light-structure-from-pin-motion`

Acknowledgements This work was supported by JSPS KAKENHI Grant Number JP19H01123. Hiroaki Santo is grateful for support through a JSPS research fellowship for young scientists by the Japan Society for the Promotion of Science (JP19J10326). Michael Waechter is grateful for support through a JSPS postdoctoral fellowship by the Japan Society for the Promotion of Science (JP17F17350).

References

1. Ackermann, J., Fuhrmann, S., Goesele, M.: Geometric point light source calibration. In: Proceedings of Vision, Modeling, and Visualization, pp. 161–168 (2013)
2. Alldrin, N.G., Mallick, S.P., Kriegman, D.J.: Resolving the generalized bas-relief ambiguity by entropy minimization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–7 (2007)
3. Aoto, T., Taketomi, T., Sato, T., Mukaigawa, Y., Yokoya, N.: Position estimation of near point light sources using a clear hollow sphere. In: Proceedings of the International Conference on Pattern Recognition (ICPR), pp. 3721–3724 (2012)
4. Bouguet, J.Y., Perona, P.: 3D photography using shadows in dual-space geometry. International Journal of Computer Vision (IJCV) **35**(2), 129–149 (1999)
5. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
6. Bunteong, A., Chotikakamthorn, N.: Light source estimation using feature points from specular highlights and cast shadows. International Journal of Physical Sciences **11**(13), 168–177 (2016)
7. Cao, X., Shah, M.: Camera calibration and light source estimation from images with shadows. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 918–923 (2005)
8. Chen, G., Han, K., Shi, B., Matsushita, Y., Wong, K.Y.K.: Self-calibrating deep photometric stereo networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8739–8747 (2019)
9. Cho, D., Matsushita, Y., Tai, Y.W., Kweon, I.S.: Semi-calibrated photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2018)
10. Collins, T., Bartoli, A.: 3D reconstruction in laparoscopy with close-range photometric stereo. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 634–642 (2012)
11. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
12. Gardner, M.A., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C., Lalonde, J.F.: Learning to predict indoor illumination from a single image. ACM Transactions on Graphics (TOG) (2017)
13. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition **47**(6), 2280–2292 (2014)
14. Goldman, D.B., Curless, B., Hertzmann, A., Seitz, S.M.: Shape and spatially-varying BRDFs from photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **32**(6), 1060–1071 (2010)
15. Hara, K., Nishino, K., Ikeuchi, K.: Light source position and reflectance estimation from a single view without the distant illumination assumption. IEEE Transactions

on Pattern Analysis and Machine Intelligence (PAMI) **27**(4), 493–505 (2005)
16. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **19**(6), 580–593 (1997)
17. Hartley, R.I., Sturm, P.: Triangulation. Computer Vision and Image Understanding Journal (CVIU) **68**(2), 146–157 (1997)
18. Hartley, R.I., Zisserman, A.: Multiple view geometry in computer vision, second edn. Cambridge University Press (2004)
19. Horn, B.K.: Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Tech. Rep. AITR-232, MIT (1970)
20. Hu, B., Brown, C.M., Nelson, R.C.: The geometry of point light source from shadows. Tech. Rep. UR CSD / TR810, University of Rochester (2004)
21. Logothetis, F., Mecca, R., Cipolla, R.: Semi-calibrated near field photometric stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 941–950 (2017)
22. Ma, L., Liu, J., Pei, X., Hu, Y., Sun, F.: Calibration of position and orientation for point light source synchronously with single image in photometric stereo. Optics Express **27**(4), 4024–4033 (2019)
23. Negahdaripour, S.: Closed-form relationship between the two interpretations of a moving plane. Journal of the Optical Society of America **7**(2), 279–285 (1990)
24. Nocedal, J., Wright, S.J.: Numerical optimization. Springer (2006)
25. Park, J., Sinha, S.N., Matsushita, Y., Tai, Y., Kweon, I.: Calibrating a non-isotropic near point light source using a plane. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2267–2274 (2014)
26. Pătrăucean, V., Gurdjos, P., von Gioi, R.G.: Joint a contrario ellipse and line detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **39**(4), 788–802 (2016)
27. Petersen, K.B., Pedersen, M.S.: The Matrix Cookbook (2012). URL `http://www2.imm.dtu.dk/pubdb/p.php?3274`. Version 20121115
28. Powell, M.W., Sarkar, S., Goldgof, D.: A simple strategy for calibrating the geometry of light sources. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **23**(9), 1022–1027 (2001)
29. Quéau, Y., Durix, B., Wu, T., Cremers, D., Lauze, F., Durou, J.D.: Led-based photometric stereo: modeling, calibration and numerical solution. Journal of Mathematical Imaging and Vision **60**(3), 313–340 (2018)
30. Quéau, Y., Wu, T., Lauze, F., Durou, J.D., Cremers, D.: A non-convex variational approach to photometric stereo under inaccurate lighting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
31. Sato, I., Sato, Y., Ikeuchi, K.: Stability issues in recovering illumination distribution from brightness in shadows. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. II-400–II-407 (2001)
32. Sato, I., Sato, Y., Ikeuchi, K.: Illumination from shadows. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **25**(3), 290–300 (2003)
33. Schnieders, D., Wong, K.Y.K.: Camera and light calibration from reflections on a sphere. Computer Vision and Image Understanding Journal (CVIU) **117**(10), 1536–1547 (2013)
34. Schnieders, D., Wong, K.Y.K., Dai, Z.: Polygonal light source estimation. In: Proceedings of the Asian Conference on Computer Vision (ACCV), pp. 96–107 (2009)
35. Shen, H.L., Cheng, Y.: Calibrating light sources by using a planar mirror. Journal of Electronic Imaging **20**(1), 013002-1–013002-6 (2011)
36. Shi, B., Matsushita, Y., Wei, Y., Xu, C., Tan, P.: Self-calibrating photometric stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1118–1125 (2010)
37. Silver, W.M.: Determining shape and reflectance using multiple images. Master's thesis, Massachusetts Institute of Technology (1980)
38. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. In: Proceedings of SIGGRAPH, pp. 835–846 (2006)
39. Song, Z., Nie, Y., Song, Z.: Photometric stereo with quasi-point light source. Optics and Lasers in Engineering **111**, 172–182 (2018)
40. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer (2010)
41. Takai, T., Maki, A., Niinuma, K., Matsuyama, T.: Difference sphere: An approach to near light source estimation. Computer Vision and Image Understanding Journal (CVIU) **113**(9), 966–978 (2009)
42. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – A modern synthesis. In: ICCV Workshop on Vision Algorithms: Theory and Practice, pp. 298–372 (2000)
43. Wang, Y., Samaras, D.: Estimation of multiple directional light sources for synthesis of mixed reality images. In: Proceedings of the Pacific Conference on Computer Graphics and Applications, pp. 38–47 (2002)
44. Weber, M., Cipolla, R.: A practical method for estimation of point light-sources. In: Proceedings of the British Machine Vision Conference (BMVC), vol. 2, pp. 471–480 (2001)
45. Wei, J.: Robust recovery of multiple light source based on local light source constant constraint. Pattern Recognition Letters **24**(1), 159–172 (2003)
46. Wong, K.Y.K., Schnieders, D., Li, S.: Recovering light directions and camera poses from a single sphere. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 631–642 (2008)
47. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineering **19**(1), 139–144 (1980)
48. Zhang, Y., Yang, Y.H.: Multiple illuminant direction detection with application to image synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **23**(8), 915–920 (2001)
49. Zhou, W., Kambhamettu, C.: Estimation of illuminant direction and intensity of multiple light sources. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 206–220 (2002)

A Estimating the quadrifocal tensor $\mathcal{Q}$

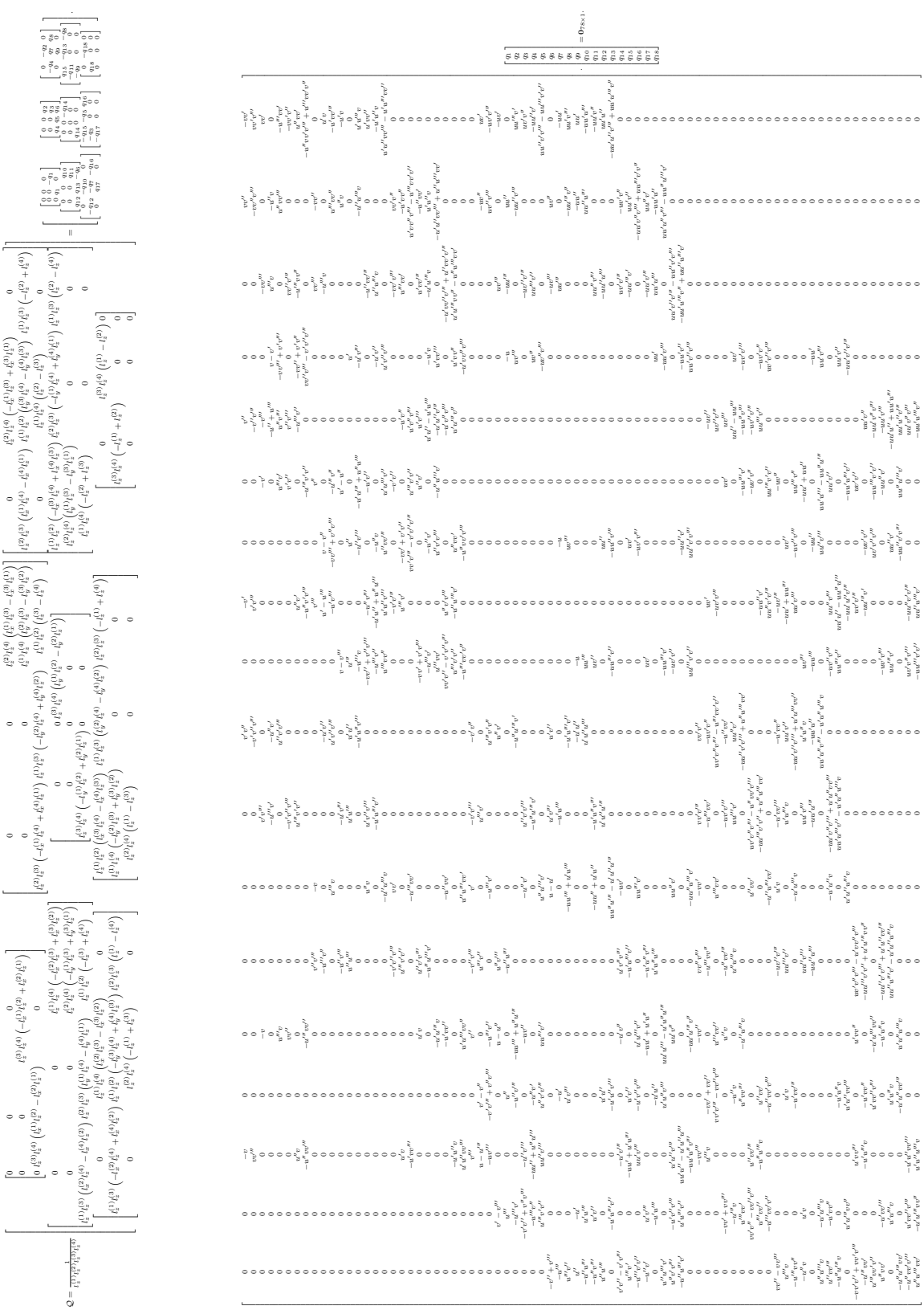

Fig. A.1 *Left:* The quadrifocal tensor $\mathcal{Q}$. *Right:* The equation system whose solution gives the parameters $q_1, \ldots, q_{18}$ of $\mathcal{Q}$.