

Light Structure from Pin Motion: Simple and Accurate Point Light Calibration for Physics-based Modeling

Hiroaki Santo, Michael Waechter, Masaki Samejima, Yusuke Sugano, and
Yasuyuki Matsushita

Osaka University

{santo.hiroaki,waechter.michael,samejima,sugano,yasumat}@ist.osaka-u.ac.jp

Abstract. We present a practical method for geometric point light source calibration. Unlike in prior works that use Lambertian spheres, mirror spheres, or mirror planes, our calibration target consists of a Lambertian plane and small shadow casters at unknown positions above the plane. Due to their small size, the casters' shadows can be localized more precisely than highlights on mirrors. We show that, given shadow observations from a moving calibration target and a fixed camera, the shadow caster positions and the light position or direction can be simultaneously recovered in a structure from motion framework. Our evaluation on simulated and real scenes shows that our method yields light estimates that are stable and more accurate than existing techniques while having a considerably simpler setup and requiring less manual labor.

This project's source code can be downloaded from: <https://github.com/hiroaki-santo/light-structure-from-pin-motion>

Keywords: Light source calibration, photometric stereo, shape-from-shading, appearance modeling, physics-based modeling

1 Introduction

Estimating the position or direction of a light source *accurately* is essential for many physics-based techniques such as shape from shading, photometric stereo, or reflectance and material estimation. In these, inaccurate light positions immediately cause errors. Figure 1 shows the relation between light calibration error and normal estimation error in a synthetic experiment with directional light, a Lambertian sphere as target object, and a basic photometric stereo method [1, 2]. We can see the importance of accurate light calibration: Working on algorithmic improvements in photometric stereo that squeeze out the last few percent in accuracy is futile if the improvements are overshadowed by the calibration inaccuracy. 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. 2, that can be made

from off-the-shelf items for $< \$5$ within 1–2 min. Instead of specular highlights on spheres we use a planar board (shadow receiver) and pins (shadow casters) stuck on it 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 derive the light position/direction.

The accuracy with which we can localize the point shadows is the key factor in the overall calibration accuracy. We can control this through the pin head size. Ideally the shadows should be about 1 px wide, but even with off-the-shelf pins we can automatically localize their centers with an accuracy of about 1–2 px (Fig. 3, *right*), which is in marked contrast to how accurately we can detect specular highlights. Further, since our target is planar, it translates small shadow localization errors only into small light direction errors. This is in contrast to mirror sphere methods where the surface normal, which determines the light reflection angle, changes across the sphere and thus amplifies errors. We will come back to these advantages of our method in Sec. 2.

Geometrically, point lights are inverse pinhole cameras [3] (see Fig. 4). We can thus build upon past studies on projective geometry. In particular we show that, analogous to structure from motion (SfM) which jointly estimates camera poses and 3D feature positions, 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 pin) *structure from pin motion*.

In this paper we clarify the relationship between our problem and conventional SfM and develop a solution technique for our context. Interestingly, our method’s connection to SfM allows users to place the pins arbitrarily on the board because their positions will be estimated in the calibration process. This means that – in contrast to many previous works – we do not need to carefully manufacture or measure our calibration target. Further, in contrast to some previous methods, we require no hand annotations in the captured imagery. All required information can be inferred automatically with sufficient accuracy.

The primary contributions of our work are twofold. First, it introduces a practical light source calibration method that uses an easy-to-make calibration target. Instead of requiring a carefully designed calibration target, our method

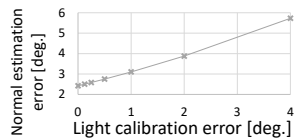


Fig. 1. Light calibration error vs. normal error in photometric stereo. Each data point is the mean of 100 runs.

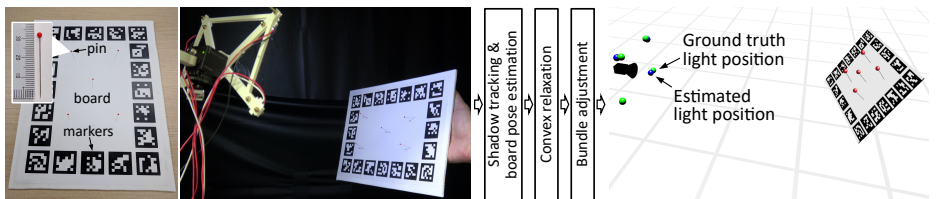


Fig. 2. Our calibration board, a camera observing the movement of shadows cast by a point light while the board moves, our algorithm’s workflow, and the estimation result.

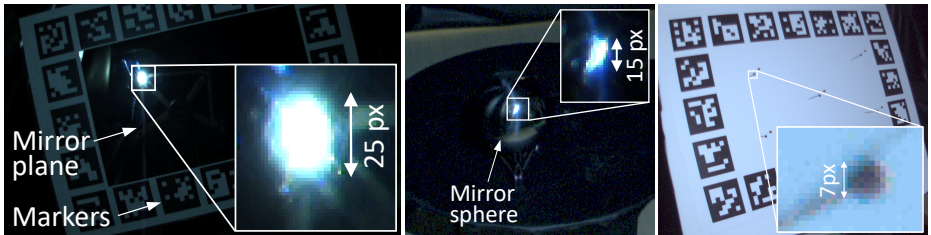


Fig. 3. *Left and center:* Specular highlights on a mirror plane [4] and a mirror sphere. *Right:* The pin head shadows on our planar board.

only uses needle pins that are stuck at unknown locations on a plane. Second, we show that the calibration of point light source positions/directions can be formulated as a bundle adjustment problem using the observations of cast shadows, and develop a robust solution technique for accurately achieving the calibration.

The benefits from the new calibration target and associated solution method are an extremely simple target construction process (shown in the supplemental video), a calibration process that requires no manual intervention other than moving the board, and improved accuracy compared to prior work.

2 Related work

Light source calibration methods can be roughly divided into three categories: (1) estimating light source directions for scenes with a distant point light source, (2) estimating illumination distributions in natural light source environments, and (3) estimating light source positions in scenes with a near point light source.

In the first category, Zhang and Yang [5] (and Wei [6] with a more robust implementation) proposed a method to estimate multiple distant light directions based on a Lambertian sphere’s shadow boundaries and their intensities. Wang and Samarasinghe [7] 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 [8] used stereo images of a reference sphere with specular reflection to estimate light directions. Cao and Shah [9] proposed a method for estimating camera parameters and the light direction from shadows cast by common vertical objects such as walls instead of special, precisely fabricated objects, which can be used for images from less restricted settings.

In the second category, Sato *et al.* [10, 11] 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.

In the third category, Powell *et al.* [12] triangulate multiple light positions from highlights on three specular spheres at known positions. Other methods also use reflective spheres [13–16] or specially designed geometric objects [17–19]. In contrast to some of these methods’ simple triangulation schemes, Ackermann *et al.* [20] model the light estimation as nonlinear least squares minimization of highlight reprojection errors, yielding improved accuracy. Park *et al.* [21] handle

non-isotropic lights and estimate light position and radiant intensity distribution from multi-view imagery of the shading and specular reflections on a plane. Further, some methods are based on planar mirrors [4, 22]. They model the mirror by perspective projection and infer parameters similar to camera calibration.

In highlight-based 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 we can still barely detect or annotate other parts of the calibration target (pose detection markers, sphere outline, *etc.*), the highlight is much bigger than the light source’s image itself (see Fig. 3, *left* and *center*). 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 mirror and the light cone’s axis) cannot be found by computing the highlight’s centroid, as for example Shen *et al.* [4] do. 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 small cast shadows for stable localization.

Mirror sphere-based calibration methods suffer from the fact that the sphere curvature amplifies highlight localization errors into larger light direction errors because the surface orientation, 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.* [20]). 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 annotating the sphere outline in the image which is highly tedious if done accurately, as anyone who has done it can confirm. This is because one has to fit a conic section into the sphere’s image and the sphere’s exact outline is hard to distinguish from the background, especially in dark images, since the sphere also mirrors the background.

The connection between pinhole cameras and point lights has already been shown by others: Hu *et al.* [3] use it in a theoretical setting similar to ours with point objects and shadows. Each shadow yields a line constraint for the light position and they triangulate multiple such constraints by computing the point closest to all lines. However, they only discuss the idea with geometric sketches.

We push the idea further by deriving mathematical solutions, extending it to distant light, embedding it in an SfM framework [23, 24] 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. Given noisy observations, minimizing reprojection error as we do gives better results (see Szeliski [25, Sec. 7.1] or Hartley and Sturm [26]) than the point-closest-to-all-rays triangulation used by Hu *et al.* [3], Shen and Cheng’s mirror plane method [4], and most sphere methods prior to Ackermann’s [20].

3 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. 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 positions of the shadow casters on the calibration board 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.

To illustrate the relationship between light source, shadow caster, and observed shadow, we begin with the shadow formation model which is connected to perspective camera geometry. We then describe our proposed calibration method that is based on bundle adjustment, as well as its implementation details.

We denote matrices and vectors with bold upper and lower case and the homogeneous representation of \mathbf{v} with $\hat{\mathbf{v}}$. Due to space constraints we show the exact derivations of Eqs. (2), (4), (7), and (9) in the supplemental material.

3.1 Shadow formation model

Let us for now assume that the pose of the shadow receiver plane Π is fixed to the world coordinate system’s x-y plane.

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 Π at $\mathbf{s} \in \mathbb{R}^2$ in Π ’s 2D coordinate system which is $\bar{\mathbf{s}} = [\mathbf{s}^\top, 0]^\top$ in world coordinates (because Π coincides with the world’s x-y plane). Since \mathbf{l} , \mathbf{c} and $\bar{\mathbf{s}}$ are all on one line, the lines $\overline{\mathbf{c}\bar{\mathbf{s}}}$ and $\overline{\mathbf{l}\bar{\mathbf{s}}}$ are parallel:

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

From this it follows that the shadow formation can be written as

$$\lambda \hat{\mathbf{s}} = \begin{bmatrix} -l_z & 0 & l_x \\ 0 & -l_z & l_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -l_x \\ 0 & 1 & 0 & -l_y \\ 0 & 0 & 1 & -l_z \end{bmatrix} \hat{\mathbf{c}} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 1 & -l_z \end{bmatrix} \hat{\mathbf{c}} = \mathbf{L}\hat{\mathbf{c}}. \quad (2)$$

As such, point lights and pinhole cameras can be described by similar mathematical models with the following correspondences: (light source \Leftrightarrow camera), (shadow receiver plane \Leftrightarrow image plane), (shadow caster \Leftrightarrow observed point), and (first two matrices of Eq. (2) \Leftrightarrow camera intrinsics and extrinsics), see Fig. 4.

For *distant light* all light rays in the scene are parallel, $\mathbf{l} = [l_x, l_y, l_z]^\top$ is a light direction instead of a position, and the line $\overline{\mathbf{c}\bar{\mathbf{s}}}$ is parallel to \mathbf{l} :

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

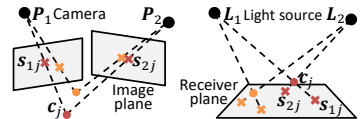


Fig. 4. Cameras vs. point lights. Camera matrix \mathbf{P}_i projects a scene point \mathbf{c}_j to \mathbf{s}_{ij} just like light matrix \mathbf{L}_i projects \mathbf{c}_j to \mathbf{s}_{ij} . SfM estimates \mathbf{P}_i and \mathbf{c}_j from $\{\mathbf{s}_{ij}\}$ and we estimate \mathbf{L}_i and \mathbf{c}_j .

From this follows an expression that resembles orthographic projection:

$$\lambda \hat{\mathbf{s}} = \begin{bmatrix} -l_z & 0 & l_x & 0 \\ 0 & -l_z & l_y & 0 \\ 0 & 0 & 0 & -l_z \end{bmatrix} \hat{\mathbf{c}} = \mathbf{L} \hat{\mathbf{c}}. \quad (4)$$

3.2 Light source calibration as bundle adjustment

Our goal is to determine the light source position or direction \mathbf{l} in Eq. (2) or (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

$$\mathbf{l}_i = [l_x^{(i)} \ l_y^{(i)} \ l_z^{(i)}]^\top = \mathbf{R}_i^\top \mathbf{l} - \mathbf{R}_i^\top \mathbf{t}_i.$$

With this index i the matrices $\{\mathbf{L}_i\}$ for nearby and distant light, resp., become

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

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 . Eqs. (2) and (4) then become

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

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 $\{\mathbf{c}_j\}$. We formulate this as a least-squares objective function of the reprojection error:

$$\min_{\mathbf{l}, \mathbf{c}_j, \lambda_{ij}} \sum_{i,j} \|\lambda_{ij} \hat{\mathbf{s}}_{ij} - \mathbf{L}_i \hat{\mathbf{c}}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{l} = \mathbf{R}_i \mathbf{l}_i + \mathbf{t}_i. \quad (5)$$

We solve this nonlinear least-squares problem with Levenberg-Marquardt [27]. For robust estimation we use RANSAC: We repeatedly choose a random observation set, estimate $(\mathbf{l}, \mathbf{c}_j, \lambda_{ij})$, and select the estimate with the smallest residual.

Initialization: Equation (5) 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.

In the near light case, the objective can be written analogous to Eq. (1) as $(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times (\mathbf{l}_i - \bar{\mathbf{s}}_{ij}) = \mathbf{0}$ and (using $\mathbf{l}_i = \mathbf{R}_i^\top \mathbf{l} - \mathbf{R}_i^\top \mathbf{t}_i$) rewritten as

$$(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times (\mathbf{R}_i^\top \mathbf{l} - \mathbf{R}_i^\top \mathbf{t}_i - \bar{\mathbf{s}}_{ij}) = \mathbf{0}. \quad (6)$$

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

$$\begin{array}{c}
\begin{bmatrix}
r_6 s_y & -r_6 s_x & r_3 s_x - r_0 s_y \\
r_7 s_y & -r_7 s_x & r_4 s_x - r_1 s_y \\
r_8 s_y & -r_8 s_x & r_5 s_x - r_2 s_y \\
0 & t_z & s_y - t_y \\
-t_z & 0 & -s_x + t_x \\
-s_y + t_y & s_x - t_x & 0 \\
0 & r_6 & -r_3 \\
-r_6 & 0 & r_0 \\
r_3 & -r_0 & 0 \\
0 & r_7 & -r_4 \\
-r_7 & 0 & r_1 \\
r_4 & -r_1 & 0 \\
0 & r_8 & -r_5 \\
-r_8 & 0 & r_2 \\
r_5 & -r_2 & 0
\end{bmatrix}^{\top}
\begin{bmatrix}
l_x \\
l_y \\
l_z \\
c_{j,x} \\
c_{j,y} \\
c_{j,z} \\
l_x c_{j,x} \\
l_x c_{j,y} \\
l_x c_{j,z} \\
l_y c_{j,x} \\
l_y c_{j,y} \\
l_y c_{j,z} \\
l_z c_{j,x} \\
l_z c_{j,y} \\
l_z c_{j,z}
\end{bmatrix}
= \underbrace{\begin{bmatrix} -s_y t_z \\ s_x t_z \\ s_y t_x - s_x t_y \end{bmatrix}}_{\mathbf{b}_{ij}}.
\end{array} \quad (7)$$

$\underbrace{\hspace{15em}}_{\mathbf{A}_{ij}} \quad \underbrace{\hspace{1em}}_{\boldsymbol{\theta}_j}$

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

$$(\mathbf{c}_j - \bar{\mathbf{s}}_{ij}) \times \mathbf{R}_i^{\top} \mathbf{l} = \mathbf{0}. \quad (8)$$

Keeping the definitions of \mathbf{c}_j , $\bar{\mathbf{s}}_{ij}$, and \mathbf{R}_i^{\top} from above but setting $\mathbf{l} = [l_x, l_y, 1]^{\top}$ to reduce \mathbf{l} to two degrees of freedom, the system becomes

$$\begin{array}{c}
\begin{bmatrix}
r_6 s_y & -r_6 s_x & r_3 s_x - r_0 s_y \\
r_7 s_y & -r_7 s_x & r_4 s_x - r_1 s_y \\
0 & r_8 & -r_5 \\
-r_8 & 0 & r_2 \\
r_5 & -r_2 & 0 \\
0 & r_6 & -r_3 \\
-r_6 & 0 & r_0 \\
r_3 & -r_0 & 0 \\
0 & r_7 & -r_4 \\
-r_7 & 0 & r_1 \\
r_4 & -r_1 & 0
\end{bmatrix}^{\top}
\begin{bmatrix}
l_x \\
l_y \\
c_{j,x} \\
c_{j,y} \\
c_{j,z} \\
l_x c_{j,x} \\
l_x c_{j,y} \\
l_x c_{j,z} \\
l_y c_{j,x} \\
l_y c_{j,y} \\
l_y c_{j,z}
\end{bmatrix}
= \underbrace{\begin{bmatrix} -s_y r_8 \\ s_x r_8 \\ s_y r_2 - s_x r_5 \end{bmatrix}}_{\mathbf{b}_{ij}}.
\end{array} \quad (9)$$

$\underbrace{\hspace{15em}}_{\mathbf{A}_{ij}} \quad \underbrace{\hspace{1em}}_{\boldsymbol{\theta}_j}$

To make the estimation of $\boldsymbol{\theta}_j$ robust against outliers, we use ℓ_1 minimization:

$$\boldsymbol{\theta}_j^* = \operatorname{argmin}_{\boldsymbol{\theta}_j} \|\mathbf{A}_{ij} \boldsymbol{\theta}_j - \mathbf{b}_{ij}\|_1. \quad (10)$$

After obtaining $\boldsymbol{\theta}_j^*$ we disregard the second-order variables $l_x c_{j,x}$, *etc.* – making the problem convex – and use \mathbf{c}_j^* and \mathbf{l}^* as initialization for minimizing Eq. (5).

Minimal conditions for initialization: Let N_p and N_c be the number of target poses and casters. For solving Eqs. (7) or (9) we must fulfill

$$\underbrace{3N_p N_c}_{\text{\#equations}} \geq \underbrace{12N_c + 3}_{\text{\#variables}} \Leftrightarrow N_p \geq 4 + \frac{1}{N_c} \quad \text{or} \quad \underbrace{3N_p N_c}_{\text{\#equations}} \geq \underbrace{9N_c + 2}_{\text{\#variables}} \Leftrightarrow N_p \geq 3 + \frac{2}{3N_c}.$$

Thus, 5 and 4 poses suffice for nearby and distant light, resp., regardless of N_c .

3.3 Implementation

To obtain our target’s pose $\{[\mathbf{R}_i|\mathbf{t}_i]\}$, we print ArUco markers [28] on a piece of paper, attach it to the target (see Fig. 2, *left*), and use OpenCV 3D pose estimation [29]. Our shadow casters are off-the-shelf pins with a length of ~ 30 mm and a head diameter of ~ 3 mm, which is big enough to easily detect and small enough to accurately localize them. As mentioned, we can place the pins arbitrarily.

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 distinguishing between heads and head shadows.

For Eqs. (5), (7), and (9) we need to assign the same index j to all shadows \bar{s}_{ij} from the same caster \mathbf{c}_j in different images. Normal SfM solves this correspondence problem using the appearance of feature points. We want our shadows to be very small and can therefore not alter their shape enough to make them clearly distinguishable. Instead, we track them through a video of the calibration process. To facilitate the tracking, we place the pins far apart from each other.

3.4 Estimating multiple lights simultaneously

We can even calibrate multiple lights simultaneously to a) save time by letting the user capture data for multiple lights in one video and b) increase the estimation accuracy by constraining each caster’s position by multiple lines from a light through the caster to a shadow.

Above we discussed that tracking helps us find corresponding shadows from the same caster across all images. Multiple lights entail another correspondence problem: finding all shadows from the same light to couple the correct shadows $\bar{s}_{i,j,k}$ and lights $\mathbf{l}_{i,k}$ in our equations. To solve this we first put each shadow track *separately* into our algorithm. For each of the N_l lights we get N_c position estimates which vary slightly due to noise. We then cluster the $N_l \times N_c$ estimates into N_l clusters, each corresponding to one light. Finally we solve the bundle adjustment, Eq. (5), with an additional summation over all lights. We initialize Eq. (5) with the mean of the first step’s N_c duplicate estimates.

This solution for the correspondence problem only requires users to provide N_l and in contrast to, *e.g.*, Powell *et al.*’s [12] ordering constraint it does not fail in certain configurations of camera, lights and target. The clustering might, however, produce wrong clusters if two lights are so close that their clusters overlap, but this may be practically irrelevant since in physics-based modeling two lights need to be far enough apart to give an information gain over one light.

Interestingly, we can even simultaneously calibrate lights whose imagery has not been captured simultaneously. This is possible since applying the board poses transforms all shadow positions – no matter whether they were captured simultaneously or not – into the same coordinate system, namely the target’s.

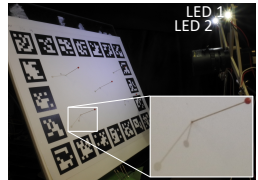


Fig. 5. Two lights casting two shadows per pin.

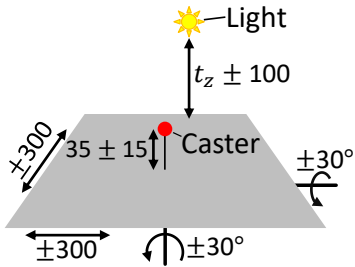


Fig. 6. Arrows show value ranges for our simulation experiments.

	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.

Table 1. Estimation error (mean of ten random trials) in a synthetic, noise-free setting.

4 Evaluation

We now assess our method’s accuracy using simulation experiments (Sec. 4.1) and real-world scenes (Sec. 4.2).

4.1 Simulation

We randomly sampled board poses, caster positions and light positions (the latter only for near light conditions) from uniform distributions within the ranges shown in Fig. 6. The casters were randomly placed on a board of size 200×200 . For distant light, we sampled the light direction’s polar angle θ from $[0^\circ, 45^\circ]$.

We evaluated the absolute/angular error of estimated light positions/directions while varying the distance t_z of the light to the calibration board and the number of casters N_c . Table 1 shows that the mean error of each configuration is 14 or more 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 practice, light source estimates will be deteriorated by two main error sources: (1) Shadow localization and (2) the marker-based board pose estimation.

Shadow localization errors: To analyze the influence of shadow localization, we perturbed the shadow positions with Gaussian noise. Figure 7 shows the estimation accuracy obtained as solutions from the convex relaxation (Eq. (7) or (9)) compared to the full bundle adjustment (Eq. (5) after initialization with convex relaxation) for near and distant light in various settings. Figure 7’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 board poses makes Eqs. (10) and (5) more overconstrained and should thus reduce the error from noisy shadow locations. Figure 7’s middle and bottom row confirm decreasing errors for larger N_p or N_c .

Board pose estimation errors: To simulate errors in the board pose estimation, we performed an experiment where we added Gaussian noise to the board’s

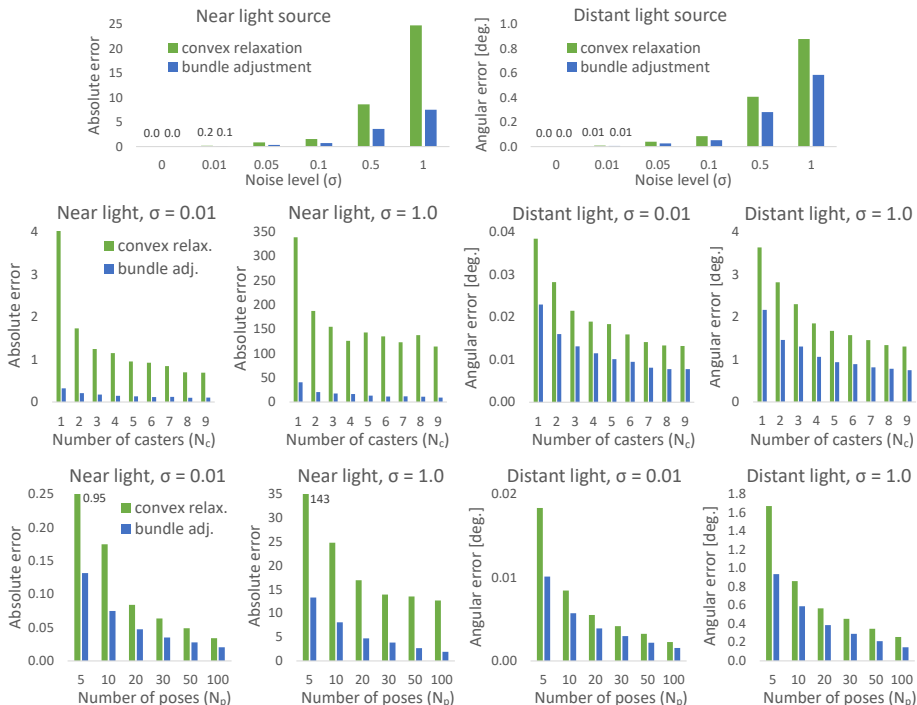


Fig. 7. Estimation error for synthetic near and distant light with Gaussian noise added to the shadow positions. Each data point is the median of 500 random trials. *Top row:* $N_p = 10$ and $N_c = 5$. The noise’s standard deviation σ is on the x-axis. *Middle row:* $N_p = 5$ and N_c is on the x-axis. *Bottom row:* $N_c = 5$ and N_p is on the x-axis.

roll, pitch, and yaw. Figure 8’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. 8’s middle and bottom row we increase the number of casters and board poses again. Bundle adjustment and increasing the number of poses reduce the error, but increasing the number of casters does not. However, this is not surprising since adding constraints to our system only helps if the constraints have independent noise. Here, the noises for all shadows $\bar{s}_{i,j}$ of the same pose i stem from the same pose noise and are thus highly correlated. Therefore, increasing the number of poses is the primary method of error reduction.

4.2 Real-world experiments

We created 4 real-world environments, see Fig. 9. 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

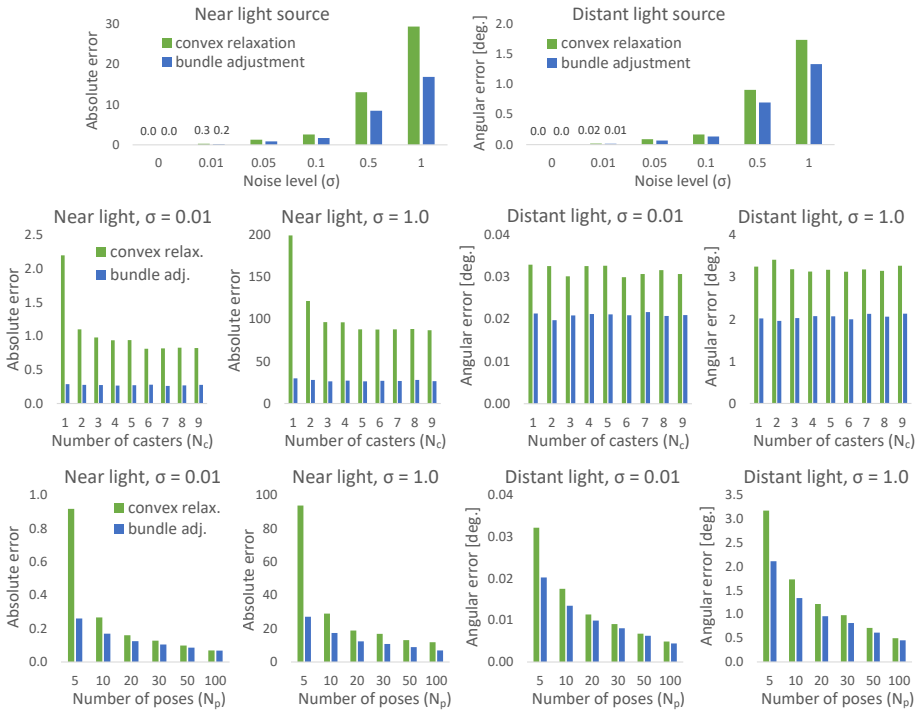


Fig. 8. Estimation error for synthetic near and distant light with Gaussian noise added to the board orientation (in deg.). Each data point is the median of 500 random trials. *Top row:* $N_p = 10$ and $N_c = 5$. The noise’s standard deviation σ is on the x-axis. *Middle row:* $N_p = 5$ and N_c is on the x-axis. *Bottom row:* $N_c = 5$ and N_p is on the x-axis.

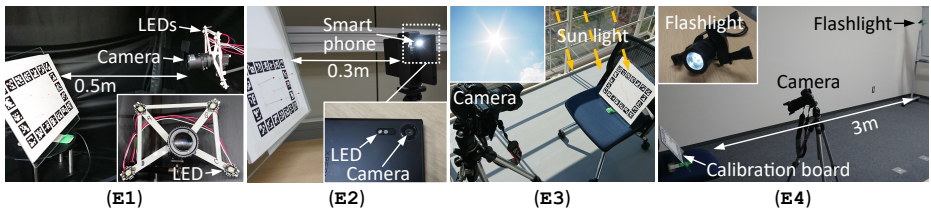


Fig. 9. Our real-world experiment environments. **E1** has four LEDs fixed around the camera. In **E2** we use a smartphone’s camera and LED. In **E3** we observe the board under sun light. **E4** has a flashlight fixed about 3m away from the board.

E2 we separately calibrated two smartphones (Sony Xperia XZs and Huawei P9) to potentially open up the path for simple, inexpensive, end user oriented photometric stereo 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 camera-LED distance to obtain the ground truth. In **E3** we placed the board under direct sun light and took datasets at three different times to obtain three light

Table 2. Estimation errors in our four real-world scenes.

Scene	number of experiments	abs./ang. light error		abs. caster position error	
		mean	stdev.	mean	stdev.
E1	4 lights	6.4 mm	2.6 mm	1.8 mm	0.44 mm
E2	2 phones	2.5 mm	0.30 mm	1.2 mm	0.61 mm
E3	3 sun positions	1.1 deg.	0.34 deg.	2.0 mm	0.52 mm
E4	1 light	0.5 deg.	–	1.7 mm	0.51 mm

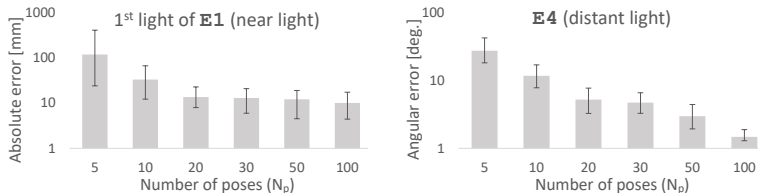


Fig. 10. Estimation error for the first light of scene E1 and for scene E4. For each scene we captured 200 images, randomly picked N_p images from these, and estimated the light and caster positions. The gray bars and error bars represent median and interquartile range of 100 random iterations of this procedure.

directions. In E4 a flashlight was fixed about 3m away from the board to approximate distant lighting. In both E3 and E4 we used a Canon EOS 5D Mark IV with a 35mm 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 E1, E3, and E4 we used the A4-sized calibration board shown in Fig. 2. In E2, since LED and camera are close together, our normal board’s pins occlude the pin shadows in the image as illustrated in Fig. 11(a). We thus used an A6-sized board with four pins with 2mm heads and brought it close to the camera to effectively increase the baseline (see Fig. 11(b)).

Table 2 shows the achieved estimation results. The light position errors are ~ 6.5 mm for E1 and ~ 2.5 mm for E2 (whose scene extent and target is smaller), the light direction errors are $\sim 1^\circ$, and the caster position errors are ≤ 2 mm. Figure 10 shows how increasing the number of board poses monotonously decreases the estimation error on two of our real-world scenes.

4.3 Estimating multiple lights simultaneously

Capturing and estimating scene E1’s two top lights simultaneously (reliably detecting shadows of > 2 lights requires a better detector) as described in Sec. 3.4 reduces the mean light and caster position errors from 7.3 and 1.8 mm to 3.5 and 1.5 mm respectively. As mentioned, we can also jointly calibrate lights whose imagery was captured separately. For E1 this decreases errors as shown in Tab. 3.

4.4 Comparison with existing method

To put our method’s accuracy into perspective, on scenes 2–3 times as big as ours Ackermann *et al.* [20] achieved accuracies of about 30–70 mm despite also min-

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

Calibration	2 lights				3 lights				4 lights			
	light err.		caster err.		light err.		caster err.		light err.		caster err.	
	mean	stdev.	mean	stdev.	mean	stdev.	mean	stdev.	mean	stdev.	mean	stdev.
separate	7.34	1.00	1.81	0.39	7.82	2.39	1.80	0.39	6.40	2.63	1.79	0.44
simultaneous	6.84	1.74	1.74	0.38	6.92	1.08	1.78	0.39	6.23	0.94	1.77	0.46

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

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

imizing reprojection error (thus being more accurate than earlier mirror sphere methods based on simpler triangulation schemes [26]) with very careful 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 reimplementaion of a state-of-the-art method based on a planar mirror [4] – 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 mirrors pose and we annotated the highlight positions manually. For a fair comparison we also annotated the shadow positions for our method manually.

In both methods we 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 4 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.

5 Discussion

With our noise-free simulation experiments we verified 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 board 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 board 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.

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 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.

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. Our target’s construction is simple, fast and cheap and most calibration steps (*e.g.*, board pose estimation and shadow detection/matching) run automatically. The only manual interaction – moving the board and recording a video – is simple. To our knowledge no other method combines such simplicity and accuracy.

We want to add a thought on calibration target design: Ackermann [20] pointed out that narrow baseline targets (*e.g.*, Powell’s [12]) have a high light position uncertainty along the light direction. This uncertainty can be decreased by either building a big, static target such as two widely spaced spheres, or by moving the target in the scene as we do. So, again our method is strongly connected to SfM where camera movement is the key to reducing depth uncertainty.

Limitations: One limitation of our method is that it requires the camera to be able to capture sequential images (i.e., a video) for tracking the shadows to solve the shadow correspondence problem. If the casters are sufficiently far apart, a solution would be to cluster the shadows in the board coordinate system. The second limitation are scenes where light and camera are so close together that the caster occludes the image of the shadow (see Fig. 11(a)). This occurred in our smartphone environment E2 and the solution is to effectively increase the baseline between camera and light as shown in Fig. 11(b).

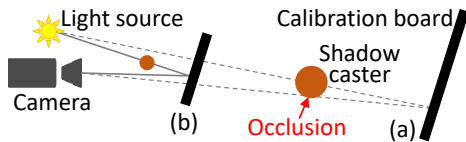


Fig. 11. With a small camera-to-light baseline the caster may occlude the shadow as seen from the camera (a). To solve this, we use a small caster, bring the board close to the camera (b) and make the board smaller so the camera can capture it fully.

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 board with one caster being moved such that its shadow stays on the same spot. Finally, we want to solve the correspondences between shadows from multiple lights (Sec. 4.3) more mathematically principled with equations that describe the co-movement of shadows from one light and multiple casters on a moving plane.

Acknowledgments: This work was supported by JSPS KAKENHI Grant Number JP16H01732. Michael Waechter is grateful for support through a postdoctoral fellowship by the Japan Society for the Promotion of Science.

References

1. Silver, W.M.: Determining shape and reflectance using multiple images. Master's thesis, Massachusetts Institute of Technology (1980)
2. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. *Optical Engineering* **19**(1) (1980) 139–144
3. Hu, B., Brown, C.M., Nelson, R.C.: The geometry of point light source from shadows. Technical Report UR CSD / TR810, University of Rochester (2004)
4. Shen, H.L., Cheng, Y.: Calibrating light sources by using a planar mirror. *Journal of Electronic Imaging* **20**(1) (2011) 013002-1–013002-6
5. 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) (2001) 915–920
6. Wei, J.: Robust recovery of multiple light source based on local light source constant constraint. *Pattern Recognition Letters* **24**(1) (2003) 159–172
7. 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*. (2002) 38–47
8. Zhou, W., Kambhamettu, C.: Estimation of illuminant direction and intensity of multiple light sources. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. (2002) 206–220
9. 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)*. (2005) 918–923
10. 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)*. (2001) II-400–II-407
11. Sato, I., Sato, Y., Ikeuchi, K.: Illumination from shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **25**(3) (2003) 290–300
12. 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) (2001) 1022–1027
13. 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) (2005) 493–505
14. 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)*. (2008) 631–642
15. 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) (2009) 966–978
16. Schnieders, D., Wong, K.Y.K.: Camera and light calibration from reflections on a sphere. *Computer Vision and Image Understanding Journal (CVIU)* **117**(10) (2013) 1536–1547
17. Weber, M., Cipolla, R.: A practical method for estimation of point light-sources. In: *Proceedings of the British Machine Vision Conference (BMVC)*. Volume 2. (2001) 471–480
18. 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)*. (2012) 3721–3724

19. Bunteong, A., Chotikakamthorn, N.: Light source estimation using feature points from specular highlights and cast shadows. *International Journal of Physical Sciences* **11**(13) (2016) 168–177
20. Ackermann, J., Fuhrmann, S., Goesele, M.: Geometric point light source calibration. In: *Proceedings of Vision, Modeling, and Visualization*. (2013) 161–168
21. 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)*. (2014) 2267–2274
22. Schnieders, D., Wong, K.Y.K., Dai, Z.: Polygonal light source estimation. In: *Proceedings of the Asian Conference on Computer Vision (ACCV)*. (2009) 96–107
23. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. In: *Proceedings of SIGGRAPH*. (2006) 835–846
24. 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*. (2000) 298–372
25. Szeliski, R.: *Computer Vision: Algorithms and Applications*. Springer (2010)
26. Hartley, R.I., Sturm, P.: Triangulation. *Computer Vision and Image Understanding Journal (CVIU)* **68**(2) (1997) 146–157
27. Nocedal, J., Wright, S.J.: *Numerical optimization*. Springer (2006)
28. 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) (2014) 2280–2292
29. Bradski, G.: *The OpenCV Library*. Dr. Dobbs’s Journal of Software Tools (2000)