HAL
open science

A full photometric and geometric model for attached webcam/matte screen devices

Yvain Quéau, Richard Modrzejewski, Pierre Gurdjos, Jean-Denis Durou

To cite this version:

Yvain Quéau, Richard Modrzejewski, Pierre Gurdjos, Jean-Denis Durou. A full photometric and geometric model for attached webcam/matte screen devices. Signal Processing: Image Communication, 2016, 40, pp.65-81. 10.1016/j.image.2015.11.006 . hal-01273010

HAL Id: hal-01273010

https://hal.science/hal-01273010v1

Submitted on 11 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 15049

To link to this article : DOI:10.1016/j.image.2015.11.006
URL : http://dx.doi.org/10.1016/j.image.2015.11.006

To cite this version :
Quéau, Yvain and Modrzejewski, Richard and Gurdjos, Pierre and Durou, Jean-Denis *A full photometric and geometric model for attached webcam/matte screen devices.* (2016) Signal Processing: Image Communication, vol. 40. pp. 65-81. ISSN 0923-5965

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

A full photometric and geometric model for attached webcam/matte screen devices

Yvain Quéau *, Richard Modrzejewski, Pierre Gurdjos, Jean-Denis Durou

Université de Toulouse, IRIT, UMR CNRS 5505, Toulouse, France

A B S T R A C T

We present a thorough photometric and geometric study of the multimedia devices composed of both a matte screen and an attached camera, where it is shown that the light emitted by an image displayed on the monitor can be expressed in closed-form at any point facing the screen, and that the geometric calibration of the camera attached to the screen can be simplified by introducing simple geometric constraints. These theoretical contributions are experimentally validated in a photometric stereo application with extended sources, where a colored scene is reconstructed while watching a collection of graylevel images displayed on the screen, providing a cheap and entertaining way to acquire realistic 3D-representations for, e.g., augmented reality.

Keywords:
Lighting
Camera calibration
Photometric stereo
Extended sources

1. Introduction

A lot of common multimedia devices (smartphones, tablets, etc.) are composed of both a screen and a webcam. If this is not the case, some cameras are designed to be easily clipped onto laptops or even monitors. Using such devices, a number of active vision applications attempt to use the camera as a photometric measuring device where the screen is used as a light source. One of the most appealing examples is 3D-reconstruction through photometric stereo [1]: different lightings can be obtained by successively displaying various patterns on the screen [2–7].

Assume that a user is watching a slideshow of images (cf. Fig. 1). A slideshow may correspond to a simple collection of white rectangles with varying locations, as suggested in [2,3,5,6], but also to circular patterns [7], or even to patterns with non-trivial geometry [4], as natural images. Can the light field emitted by the screen be treated as a light source for some applications e.g., 3D-reconstruction through photometric stereo? This question raises the key issue of this paper, which is that of efficiently estimating the light under realistic hypotheses. This introduces two key problems: modelling the emitted light field, and geometrically calibrating the device i.e., determining the camera pose w.r.t. the screen.

The most simple approximation of the screen as a light source is the infinitely distant point light source model. Such a model was considered in uncalibrated photometric stereo algorithms [3,5,7], though the discussion on the resolution of the underlying linear ambiguity (a generalized bas-relief ambiguity if integrability is imposed [8,9]) is rather limited. To avoid such ambiguities, the mean direction and the mean intensity of the light can be calibrated, as do Won et al. in [6]. Still, directional lighting seems rather unrealistic when modelling nearby screens: an anisotropic point light source model is considered instead in [2]. As we shall see in Section 2, this model is physically justified for modelling a single pixel, but does not account for the extended behavior of the screen. It is thus necessary to consider connected sets of pixels, as did Clark in [4] but without considering anisotropy. Our first contribution is to provide a general closed-form expression of the intensity and direction of light emitted by matte screens. As a particular case, we show how to compute the light emitted by a rectangular set of pixels, whatever its

* Corresponding author.
E-mail address: yvain.queau@enseeiht.fr (Y. Quéau).

http://dx.doi.org/10.1016/j.image.2015.11.006

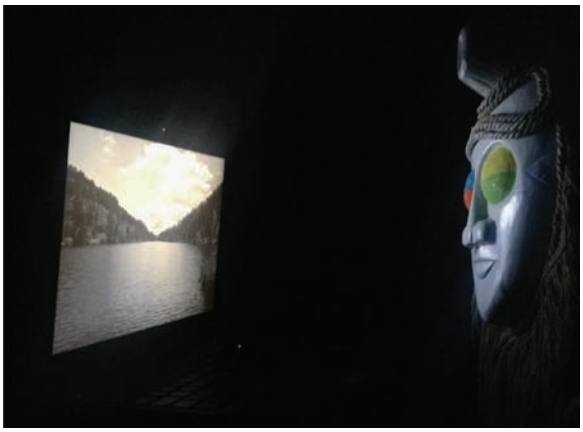

Fig. 1. Overview of our contributions. While he watches a slideshow of images, a user receives some incident light. In order to use such data in a photometric stereo application (Section 4), we need both to model the light field emitted by the screen (Section 2) and to estimate the pose of the camera (Section 3).

size and location. The light emitted by an image can then be approximated by a straightforward generalization of this model to quadtree-like [10] image decomposition.

The camera/screen calibration requires observing reference points on the screen plane. Due to our device configuration, calibration must be carried out without a direct view of the screen i.e., by using reflections in a mirror. Such a problem has been widely considered in relevant literature [11–16] and recently a solution from a single reflection in one spherical mirror has been reported [17]. In this paper, we aim at describing the most flexible *modus operandi* for users, and at providing a simple algorithm requiring few input data, such that we will consider one planar mirror and a minimal set of three reference points on the screen, as in [15]. Our contribution, initially presented in [18], is to take advantage of the constrained model of the device, as the camera pose only has four degrees of freedom: three for its location and one for its orientation restricted to a rotation around the horizontal axis of the screen. This hypothesis allows us to develop a more efficient calibration method, where as few as two mirror poses and three matched pairs of points are required.

The rest of this paper is organized as follows. After studying in Section 2 the light field emitted by a matte screen, we tackle in Section 3 the camera/screen calibration problem. As an application, we consider in Section 4 the photometric stereo problem using images displayed on the screen as light sources, and show that, using the proposed full photometric and geometric model for the device, reasonably accurate shape and reflectance can be recovered.

2. Light model for matte screens

We start by investigating the light field emitted by the screen: we show that a formal analysis of the problem allows one to reach a highly realistic closed-form model of the emitted light, which only relies on the definition of a Lambertian primary source. After briefly discussing the notion of matte screens, we will introduce a model for the light emitted by a single pixel, considered as an anisotropic point-wise source. Then, this elementary model will be extended into a new theoretical model for planar sources, holding anisotropy, spatially-varying luminance and partial occlusion. Afterwards, we will introduce a framework for simplifying this model when the luminance is uniform and occlusions are ignored. Eventually, we will provide closed-form approximations of the model for rectangular patterns and natural grayscale images, and experimentally validate them on real-world data.

2.1. Matte screens

Let us first introduce the class of monitors targeted in this work: for the sake of simplicity, we only deal with the so-called matte screens (both LCD and LED). Such screens are specifically designed by using an anti-glare coating to limit the apparition of shiny lighting effects, as opposed to bright screens which provide more vivid colors but also stronger reflections. We will also consider “small” viewing angles: when watching the screen from wide angles, brightness obviously tends to be vary much more, even for matte screens. Since we are overall interested in modelling the light emitted towards a user facing the screen, viewing angles can reasonably be assumed to be limited.

The problem tackled here is that of modelling in closed-form the luminous flux emitted by the screen. Considering the screen as a matrix of pixels, the total flux is the sum of the fluxes emitted by all pixels. An experimental study was conducted in [2], where it is demonstrated that the intensity of light emitted by a single pixel (or a small pattern) radially decreases according to a cosine law. We will show in the following that this is actually a consequence of Lambert's law, which states that in the ideal case, brightness must remain constant whatever the viewing point, leading to this anisotropic behavior.

2.2. Case of a single pixel

Our first contribution consists in showing that the empirical model of anisotropic punctual source proposed in [2] to characterize a pixel directly follows from Lambert's law. We define a pixel as a surface element $\mathrm{d}\Sigma_s$ around a point $\mathbf{x}_s = [x_s, y_s, z_s]^\top$ with unit normal $\mathbf{n}(\mathbf{x}_s)$. Let $\mathrm{d}^2\Phi$ denote the amount of luminous flux emitted by the pixel inside the elementary cone of vertex $\mathbf{x}_s$ with solid angle $\mathrm{d}\omega$ and direction $\mathbf{u}_e$ (see Fig. 2). By definition, the *luminance* of the pixel at $\mathbf{x}_s$ is the luminous flux per unit of apparent surface, seen from direction $\mathbf{u}_e$, and per unit solid angle; it is defined by

$$L_{\mathbf{x}_s}(\mathbf{u}_e) = \frac{\mathrm{d}^2\Phi}{\mathrm{d}\omega\,\mathrm{d}\Sigma'_s} \tag{1}$$

where $\mathrm{d}\Sigma'_s = \mathrm{d}\Sigma_s(\mathbf{n}(\mathbf{x}_s)\cdot\mathbf{u}_e)$ denotes the apparent surface of the pixel.

By assuming the pixels to be elementary *Lambertian primary sources*, it follows that $L_{\mathbf{x}_s}(\mathbf{u}_e)$ is independent of $\mathbf{u}_e$, so the luminance will now be referred to as:

$$L_{\mathbf{x}_s}(\mathbf{u}_e) = L(\mathbf{x}_s) \tag{2}$$

Let us now consider an elementary scene surface $\mathrm{d}\Sigma$ around a point $\mathbf{x} = [x, y, z]^\top$ at which the normal is $\mathbf{n}(\mathbf{x})$.

The solid angle $\mathrm{d}\omega$ of the cone of vertex $\mathbf{x}_s$, supported by $\mathrm{d}\Sigma$, writes:

$$\mathrm{d}\omega = \frac{\mathrm{d}\Sigma(-\mathbf{u}_e \cdot \mathbf{n}(\mathbf{x}))}{\|\mathbf{x}_s - \mathbf{x}\|^2} \tag{3}$$

On the other hand, the *irradiance* at $\mathbf{x}$:

$$\mathrm{d}I(\mathbf{x}) = \frac{\mathrm{d}^2\Phi}{\mathrm{d}\Sigma} \tag{4}$$

is the amount of luminous flux per unit of surface which is emitted by the pixel at $\mathbf{x}_s$ and received by the scene surface at $\mathbf{x}$. Using (1) and (3), the irradiance at $\mathbf{x}$ due to $\mathbf{x}_s$ is provided by the following proposition:

Proposition 1. *The irradiance (4) can be written in the dot product form*

$$\mathrm{d}I(\mathbf{x}) = \mathbf{n}(\mathbf{x}) \cdot \mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x}) \tag{5}$$

where $\mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x})$ *is the vector having the closed-form*:

$$\mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x}) = \frac{L(\mathbf{x}_s)\,\mathrm{d}\Sigma_s}{\|\mathbf{x}_s - \mathbf{x}\|^2}(\mathbf{n}(\mathbf{x}_s) \cdot \mathbf{u}_e)(-\mathbf{u}_e) \tag{6}$$

In the proposed model (6), the first factor represents the inverse of square falloff, the second stands for the cosine-like anisotropy, which has been experimentally validated in [2], and the third is the unit lighting direction (oriented towards the source). This model is nothing else than an anisotropic nearby pointwise source model, which has recently received some attention in the context of photometric stereo [19]. This justifies a posteriori the ability of the model to deal with LEDs-based screens, since LEDs can be realistically considered as anisotropic pointwise sources.

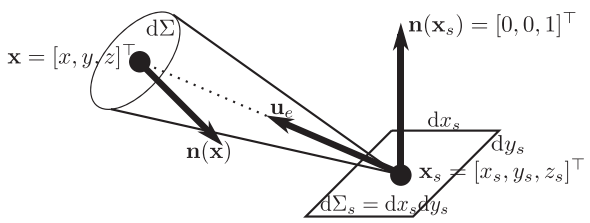

Fig. 2. Light emitted by a single pixel. Elementary scene surface $\mathrm{d}\Sigma$ with normal $\mathbf{n}(\mathbf{x})$, located around $\mathbf{x}$, is illuminated by pixel $\mathbf{x}_s \in \mathbb{R}^3$ with elementary surface $\mathrm{d}\Sigma_s$ and normal $\mathbf{n}(\mathbf{x}_s)$. Application of Lambert's law to this illumination configuration provides the closed-form anisotropic punctual source model (7).

In the sequel, without loss of generality, it is assumed that $z_s = 0$ and $\mathbf{n}(\mathbf{x}_s) = [0, 0, 1]^\top$. As a consequence, and knowing that $\mathbf{u}_e = (\mathbf{x} - \mathbf{x}_s)/\|\mathbf{x} - \mathbf{x}_s\|$, Eq. (6) simplifies so the light received at $\mathbf{x} = [x, y, z]^\top$ from a pixel at $\mathbf{x}_s$ is written, in intensity and direction:

$$\mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x}) = L(\mathbf{x}_s)\,\mathrm{d}\Sigma_s\,\frac{z\,(\mathbf{x}_s - \mathbf{x})}{\|\mathbf{x}_s - \mathbf{x}\|^4} \tag{7}$$

2.3. Case of a connected set of pixels

Since it is not realistic to illuminate a scene by a single pixel (the elementary pointwise sources composing the screen have very low intensities), we need to consider connected sets of pixels. In most relevant papers [2,3,5,6], rectangular patterns are considered, with size small enough to allow approximation by a punctual source model, yet large enough to provide sufficient lighting. As a consequence, such models are empirical: our second contribution is to derive from the infinitesimal model (7), holding for an infinitely small pattern, a general expression of the light emitted by any (extended) planar domain.

Now considering the screen $\mathcal{S}$ as a regular grid of pixels, vector $\mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x})$ describes the light emitted by an infinitesimal part of the screen, of size $\mathrm{d}\Sigma_s = \mathrm{d}x_s\,\mathrm{d}y_s$. Summing the elementary contributions (5) of all the pixels, the total irradiance at a point $\mathbf{x}$ of a surface facing the screen, with outward normal $\mathbf{n}(\mathbf{x})$, is written:

$$I(\mathbf{x}) = \iiint_{\mathbf{x}_s \in \mathcal{T}(\mathbf{x})} \mathbf{n}(\mathbf{x}) \cdot \mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x}) = \mathbf{n}(\mathbf{x}) \cdot \underbrace{\iiint_{\mathbf{x}_s \in \mathcal{T}(\mathbf{x})} \mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x})}_{\mathbf{s}(\mathbf{x})} \tag{8}$$

where the domain $\mathcal{T}(\mathbf{x}) \subset \mathbb{R}^3$, which models *penumbra*, refers to the set indicating which pixels $\mathbf{x}_s$ of the screen $\mathcal{S}$ are visible from $\mathbf{x}$. Causes of penumbra actually include *self-shadowing* effects such as described in Fig. 3b ($\mathbf{n}(\mathbf{x}) \cdot \mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x}) < 0$) as well as *cast-shadowing* effects such as described in Fig. 3c ($\mathbf{n}(\mathbf{x}) \cdot \mathrm{d}\mathbf{s}_{\mathbf{x}_s}(\mathbf{x}) \geq 0$, but the screen is partly occluded by the surface itself).

Eventually, let us recall that $z_s = 0$ for a point $[x_s, y_s, z_s]^\top \in \mathcal{S}$, so that $\mathcal{T}(\mathbf{x})$ is characterized by a 2D-domain $\Omega(\mathbf{x})$ in screen coordinates. Considering Eqs. (7) and (8), we obtain:

Proposition 2. *The light received at* $\mathbf{x} = [x, y, z]^\top$ *from a set* $\Omega(\mathbf{x})$ *of visible elementary Lambertian sources* (x_s, y_s), *lying*

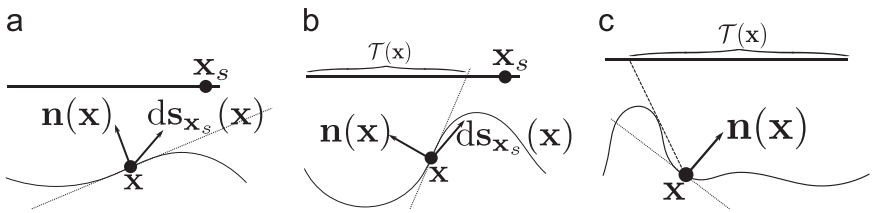

Fig. 3. Partial occlusion of the screen. (a) When there is no occluding object and when the tangent plane to the surface at $\mathbf{x}$ does not intersect the screen, then all pixels $\mathbf{x}_s \in \mathcal{S}$ contribute to the total irradiance. This is not the case in (b), where only the pixels inside $\mathcal{T}(\mathbf{x})$ are visible from $\mathbf{x}$. It is possible to explicitly define this set by intersecting the screen with the tangent plane to the surface at point $\mathbf{x}$, since it only depends on the local description of the surface through its normal. In the situation (c), the whole screen is not visible either, because of an occlusion: this type of partial visibility is much harder to deal with, since it involves global knowledge of the surface.

within a plane with normal $[0,0,1]^\top$, *is given by*:

$$\mathbf{s}(\mathbf{x}) = z \iint_{(x_s,y_s)\in\Omega(\mathbf{x})} L(\mathbf{x}_s)\frac{(\mathbf{x}_s-\mathbf{x})}{\|\mathbf{x}_s-\mathbf{x}\|^4}\,\mathrm{d}x_s\,\mathrm{d}y_s \tag{9}$$

where $L(\mathbf{x}_s)$ *is the luminance of pixel* $\mathbf{x}_s=[x_s,y_s,0]^\top$.

Adapting this result to planar sources with arbitrary orientation is straightforward.

2.4. Some remarks on partial visibility

The result from Proposition 2 can be used to model any kind of extended planar Lambertian light source, provided the luminance $L(\mathbf{x}_s)$ is known, which is the case here, since $L(\mathbf{x}_s)$ is proportional to the displayed graylevel. The $\mathbf{x}_s$ locations of the pixels being known as well, a discrete approximation of the integral in (9) by a finite sum over the pixels can be numerically computed. This is sufficient for the rendering of synthetic images: the geometry $\mathbf{x}=[x,y,z]^\top$ of the scene being perfectly known, the visibility subspace $\mathcal{T}(\mathbf{x})$, and hence $\Omega(\mathbf{x})$ can be computed by raytracing techniques.

On the contrary, in 3D-reconstruction applications, the geometry of the scene is the main unknown. Visibility should thus be estimated within an iterative process, by considering the previous estimates of $\mathbf{x}$ and of $\mathbf{n}(\mathbf{x})$ to approximate the current visibility. Yet, proceeding so is not reasonable, because this process has to be repeated for every point $\mathbf{x}$ of the scene, resulting in an extremely slow process, even on modern computers. Indeed, in real-world scenarios such as photometric stereo, $\mathbf{x}$ is one point inside a dense point cloud containing as many points as the camera has pixels, and the size of $\Omega(\mathbf{x})$ can be up to the resolution of the screen: the computation time required to evaluate the light field becomes prohibitive when considering HD devices, and computation of the visibility makes things even worse.

Hence, for the sake of simplicity, we now wish to find a closed-form approximation of the integral in (9) that can provide a fast, yet reasonably realistic, estimation of the lighting. For this purpose, we ignore the visibility issue in the following, and leave it as an interesting perspective, as in other state-of-the-art large sources models for photometric stereo [4]. As a consequence, the proposed model will be accurate for surfaces with relatively small slopes and no occlusion, but approximate in the presence of shadows or penumbra effects.

2.5. Domains with arbitrary shape and uniform luminance

To further simplify the integral in (9), we need to explicit the dependency of the emitted luminance $L(\mathbf{x}_s)$ in terms of screen coordinates, so as to obtain a closed-form expression. Let us start with the simplest case of a uniform luminance.

Let $\mathcal{S}'$ be a subset of $\mathcal{S}$ over which the luminance is uniform i.e., $L(\mathbf{x}_s)=L_0$, and let Ω be the corresponding 2D-domain. According to Eq. (9), the light field received in $\mathbf{x}=[x,y,z]^\top$ from the pixels $\mathbf{x}_s=[x_s,y_s,0]^\top\in\mathcal{S}'$, with $(x_s,y_s)\in\Omega$, is written:

$$\mathbf{s}(\mathbf{x}) = z\,L_0 \iint_{(x_s,y_s)\in\Omega} \frac{\mathbf{x}_s-\mathbf{x}}{\|\mathbf{x}_s-\mathbf{x}\|^4}\,\mathrm{d}x_s\,\mathrm{d}y_s \tag{10}$$

or, equivalently:

$$\mathbf{s}(\mathbf{x}) = -\frac{L_0}{2}[F_1(\mathbf{x}),\ F_2(\mathbf{x}),\ F_3(\mathbf{x})]^\top \tag{11}$$

where:

$$\begin{cases} F_1(\mathbf{x}) = -2z\iint_{(x_s,y_s)\in\Omega} \dfrac{x_s-x}{\left[(x_s-x)^2+(y_s-y)^2+z^2\right]^2}\,\mathrm{d}x_s\,\mathrm{d}y_s \\ F_2(\mathbf{x}) = -2z\iint_{(x_s,y_s)\in\Omega} \dfrac{y_s-y}{\left[(x_s-x)^2+(y_s-y)^2+z^2\right]^2}\,\mathrm{d}x_s\,\mathrm{d}y_s \\ F_3(\mathbf{x}) = 2z^2\iint_{(x_s,y_s)\in\Omega} \dfrac{1}{\left[(x_s-x)^2+(y_s-y)^2+z^2\right]^2}\,\mathrm{d}x_s\,\mathrm{d}y_s \end{cases} \tag{12}$$

Denoting $r=x_s-x$ and $s=y_s-y$, these functions are rewritten:

$$\begin{cases} F_1(\mathbf{x}) = -2z\iint_{(r,s)\in\Omega-(x,y)} \dfrac{r}{(r^2+s^2+z^2)^2}\,\mathrm{d}r\,\mathrm{d}s \\ F_2(\mathbf{x}) = -2z\iint_{(r,s)\in\Omega-(x,y)} \dfrac{s}{(r^2+s^2+z^2)^2}\,\mathrm{d}r\,\mathrm{d}s \\ F_3(\mathbf{x}) = 2z^2\iint_{(r,s)\in\Omega-(x,y)} \dfrac{1}{(r^2+s^2+z^2)^2}\,\mathrm{d}r\,\mathrm{d}s \end{cases} \tag{13}$$

Let $\mathcal{C}\subset\mathbb{R}^2$ be a planar domain whose contour $\partial\mathcal{C}$ is positively oriented and piecewise C^1. For any pair (P,Q) of continuous functions $\mathcal{C}\to\mathbb{R}$, the Green–Riemann formula writes:

$$\iint_{(r,s)\in\mathcal{C}} \left(\frac{\partial Q}{\partial r}-\frac{\partial P}{\partial s}\right)\mathrm{d}r\,\mathrm{d}s = \oint_{(r,s)\in\partial\mathcal{C}} (P\,\mathrm{d}r+Q\,\mathrm{d}s) \tag{14}$$

Using the following identities:

$$\begin{cases} \dfrac{\partial}{\partial r}\left(\dfrac{1}{r^2+s^2+z^2}\right) = -\dfrac{2r}{(r^2+s^2+z^2)^2} \\ -\dfrac{\partial}{\partial s}\left(\dfrac{1}{r^2+s^2+z^2}\right) = \dfrac{2s}{(r^2+s^2+z^2)^2} \\ \dfrac{\partial}{\partial r}\left(\dfrac{r}{r^2+s^2+z^2}\right) - \dfrac{\partial}{\partial s}\left(\dfrac{-s}{r^2+s^2+z^2}\right) = \dfrac{2z^2}{(r^2+s^2+z^2)^2} \end{cases} \tag{15}$$

we easily deduce from Eqs. (13)–(15):

$$\begin{cases} F_1(\mathbf{x}) = z\oint_{(r,s)\in\partial\Omega-(x,y)} \dfrac{\mathrm{d}s}{r^2+s^2+z^2} \\ F_2(\mathbf{x}) = -z\oint_{(r,s)\in\partial\Omega-(x,y)} \dfrac{\mathrm{d}r}{r^2+s^2+z^2} \\ F_3(\mathbf{x}) = \oint_{(r,s)\in\partial\Omega-(x,y)} \dfrac{r\,\mathrm{d}s-s\,\mathrm{d}r}{r^2+s^2+z^2} \end{cases} \tag{16}$$

As soon as contour $\partial\Omega$ is "simple", closed-formed expressions of the three curvilinear integrals in Eq. (16) can be found. Example of "simple" contours include the case of rectangular patterns such as those considered in [2,3,5–7], circular shapes [7], and even partly self-occluded sources (Fig. 3b), since the set Ω can be expressed in

closed-form by intersecting the tangent plane to the surface with the screen (though, as discussed in the previous paragraph, effective handling of occlusions is left for future prospect). The originality of the expressions (16) is thus their generality, as they provide a framework for handling arbitrary extended planar illuminants. As an example, let us now provide the explicit form of the integrals in (16) for a rectangular set Ω.

2.6. Closed-form expressions for rectangular sets

If Ω is the rectangle $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, the integrals in (16) are easily simplified, and one obtains:

Proposition 3. *If the illuminant is the rectangle $\Omega = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, and the luminance is uniform ($L(\mathbf{x}_s) = L_0$), then the emitted light field received in $\mathbf{x}$ is given by Eq. (11), with:*

$$\begin{cases} F_1(\mathbf{x}) = \left[\left[\dfrac{z}{\sqrt{r^2+z^2}} \tan^{-1}\left(\dfrac{s}{\sqrt{r^2+z^2}}\right)\right]_{s=y_{\min}-y}^{y_{\max}-y}\right]_{r=x_{\min}-x}^{x_{\max}-x} \\ F_2(\mathbf{x}) = \left[\left[\dfrac{z}{\sqrt{s^2+z^2}} \tan^{-1}\left(\dfrac{r}{\sqrt{s^2+z^2}}\right)\right]_{s=y_{\min}-y}^{y_{\max}-y}\right]_{r=x_{\min}-x}^{x_{\max}-x} \\ F_3(\mathbf{x}) = \left[\left[\dfrac{r\tan^{-1}\left(\dfrac{s}{\sqrt{r^2+z^2}}\right)}{\sqrt{r^2+z^2}} + \dfrac{s\tan^{-1}\left(\dfrac{r}{\sqrt{s^2+z^2}}\right)}{\sqrt{s^2+z^2}}\right]_{s=y_{\min}-y}^{y_{\max}-y}\right]_{r=x_{\min}-x}^{x_{\max}-x} \end{cases} \tag{17}$$

We believe that the closed-form model above might help improving the results obtained by photometric stereo techniques using rectangular patterns [2,3,5–7], since it is physically motivated. Compared to Clark's model [4], our model considers anisotropy, while Clark considers that pixels emit light in an isotropic way, limiting the applications of his model to small objects, as stated in the sentence: "We assume that the LCD pixels are isotropic illuminants, which is not the case [...]. The assumption of isotropy is made more palatable [...] in our experimental setup, where the object is small [...]". As we shall see in Section 2.7, considering this anisotropic falloff dramatically increases the accuracy of the model.

The other simple case we study is that of image approximation by a non-uniform rectangular partition, as suggested by Clark in [4] using his simplified (isotropic) light model. This case is illustrated in Fig. 4. It trivially follows from the previous proposition that:

Proposition 4. *If the illuminant is a non-uniform rectangular partition $\cup_{i=1}^{n} \Omega^i$, where $\Omega^i = [x^i_{\min}, x^i_{\max}] \times [y^i_{\min}, y^i_{\max}]$, and the luminance is uniform inside each Ω^i, with value L^i_0, then the light field received in $\mathbf{x}$ is given by:*

$$\mathbf{s}(\mathbf{x}) = -\sum_{i=1}^{n} \frac{L^i_0}{2}\left[F^i_1(\mathbf{x}), F^i_2(\mathbf{x}), F^i_3(\mathbf{x})\right]^\top \tag{18}$$

with the same notations as in Proposition 3.

2.7. Experimental validation

We now experimentally assess the accuracy of the proposed light model for rectangular patterns Ω with varying size $|\Omega|$, and by natural images, using a white sheet of paper located on a plane parallel to the screen (Fig. 5).

Methodology: We assume that the sheet of paper is Lambertian, and that its pose is known (we stuck the sheet on a chessboard). Its normal will be denoted $\mathbf{n}$. According to Lambert's law, the luminance emitted by the sheet is given, in every point $\mathbf{x}$ of its surface, by:

$$l(\mathbf{x}) = -\frac{\rho(\mathbf{x})}{\pi}\,\mathbf{n}\cdot\mathbf{s}(\mathbf{x}) \tag{19}$$

where $\rho(\mathbf{x})$ is the albedo, which is a scalar: we assume for simplicity in this experimental part that the camera captures graylevel images, and that the screen also displays graylevel images. The albedo of the paper sheet being uniform, we denote $\rho(\mathbf{x}) = \rho$. It follows that:

$$\sqrt{\sum_{\mathbf{x}} l(\mathbf{x})^2} = \frac{\rho}{\pi}\sqrt{\sum_{\mathbf{x}} (\mathbf{n}\cdot s(\mathbf{x}))^2} \tag{20}$$

when summing over all the points $\mathbf{x}$ of the sheet. Thus, for every point $\mathbf{x}$:

$$\frac{l(\mathbf{x})}{\sqrt{\sum_{\mathbf{x}} l(\mathbf{x})^2}} = \frac{-\mathbf{n}\cdot s(\mathbf{x})}{\sqrt{\sum_{\mathbf{x}} (\mathbf{n}\cdot \mathbf{s}(\mathbf{x}))^2}} \tag{21}$$

This normalization basically eliminates the unknown albedo ρ of the sheet of paper. The luminance $l(\mathbf{x})$ is proportional to the graylevel of the image captured by the camera, up to the $\cos^4\alpha$ factor of the image irradiance equation [20], which is also removed by the normalization. Thus, the left hand side of Eq. (21) can be directly measured as data, and compared to its right hand side (model), both qualitatively and quantitatively, for our model (Proposition 3) and both the other physics-based ones [2,4].

Rectangular patterns: We first consider rectangular patterns with varying sizes (Figs. 6 and 7). As expected, a punctual light source model [2] is accurate enough for small patterns, while an extended model [4] well describes large patterns. On the other hand, our model performs as good in both cases.

Natural images: Now, the light emitted by a graylevel image is approximated by that emitted by its quadtree decomposition [10], which provides a non-uniform rectangular approximation of the image with n rectangles (see Fig. 8), and we use Proposition 4, setting each emitted luminance L^i_0 to the mean graylevel of each rectangle. We compare qualitatively and quantitatively our model with Clark's [4] in Fig. 8 and Table 1 (we cannot include a comparison with the anisotropic point light source model here, since only the case of homogeneous luminance is considered in [2], preventing one from using real images as illuminants).

These experiments prove that our model can simulate the behavior of a screen in various conditions. The number n of rectangles is set to 64 in the following, which experimentally seems to offer a good compromise between accuracy ($n=1$ corresponds to the gross approximation of the image by its mean graylevel) and speed (if n is equal to the number of pixels, we get the discrete

version of (9), which is untractable when working with large images).

3. Camera/screen calibration

Besides photometric 3D-reconstruction techniques modelling the light emitted by the screen, several other computer vision applications, such as gaze tracking, need to refer to pixels w.r.t. the three-dimensional Euclidean coordinate system attached to the screen. On the other hand, when referring to the 3D-geometry of the scene (or the gaze), coordinates are usually expressed in the camera coordinates system. Hence, in such applications requiring to handle both systems, the camera pose (location and orientation) w.r.t. the screen needs to be estimated beforehand.

In this section, we provide a theoretical study of this pose calibration problem. In particular, we show that, by constraining the orientation of the camera (its location is still unconstrained), the calibration problem is unambiguous from four matched pairs of points and two mirror poses. Regarding the minimal case with only three pairs, as a finite number of solutions exist, we furthermore describe a geometric heuristic for determining the good one which is favorably compared against state-of-the-art.

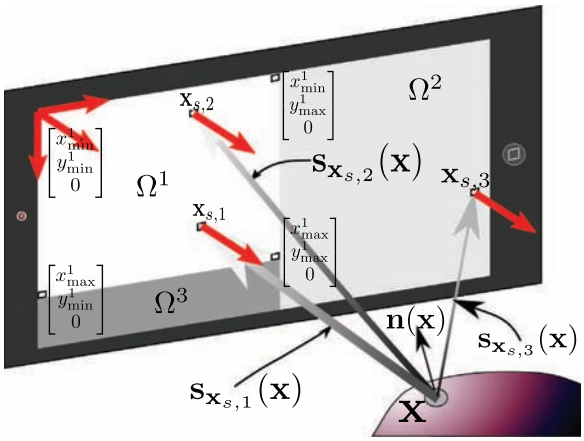

Fig. 4. Modelling of the light field emitted by three rectangular patterns Ω^1, Ω^2 and Ω^3. The resulting field (extended source) is the sum of the contributions of each infinitesimal source $\mathbf{x}_s$. The width of the gray arrows represents the intensity for three infinitesimal sources $\mathbf{x}_{s,1}$, $\mathbf{x}_{s,2}$ and $\mathbf{x}_{s,3}$, which is a function of both the pixel–object distance (inverse-of-square falloff), the angle between the lighting direction and the direction $[0,0,1]^\top$ (cosine-like anisotropy), and the luminance.

3.1. Scope of this study

Our problem is that of estimating the camera pose w.r.t. to the three-dimensional screen coordinate system, from images of known reference 3D-points. As the reference points should lie on the plane supporting the screen, a solution must be sought without a direct view of these points by using their reflections in a moving planar mirror e.g., as in [11]. In the relevant literature, the input data consist in $n=4$ reference points [11,12,14] and $k \geq 5$ [12] or $k \geq 3$ [11,14] mirror poses. Recently, the problem with only $n=3$ reference points and $k \geq 3$ mirror poses was solved by predicting all the possible solutions, and selecting the best one according to the reprojection error [13] or an orthogonality-based criterion [15].

Algorithms for solving the pose problem from minimal cases are widely reported in the literature [21]. A minimal case is a set of equations where the solution set generally is finite. Both the approach in [15] and the proposed approach

Fig. 5. Experimental setup. Top: a screen (laptop HP EliteBook8570w) displaying a rectangular pattern (left) or a graylevel image (right) in front of a white planar sheet of paper. Bottom: real images used in the experiments. All images are of size 1600×900 (screen resolution).

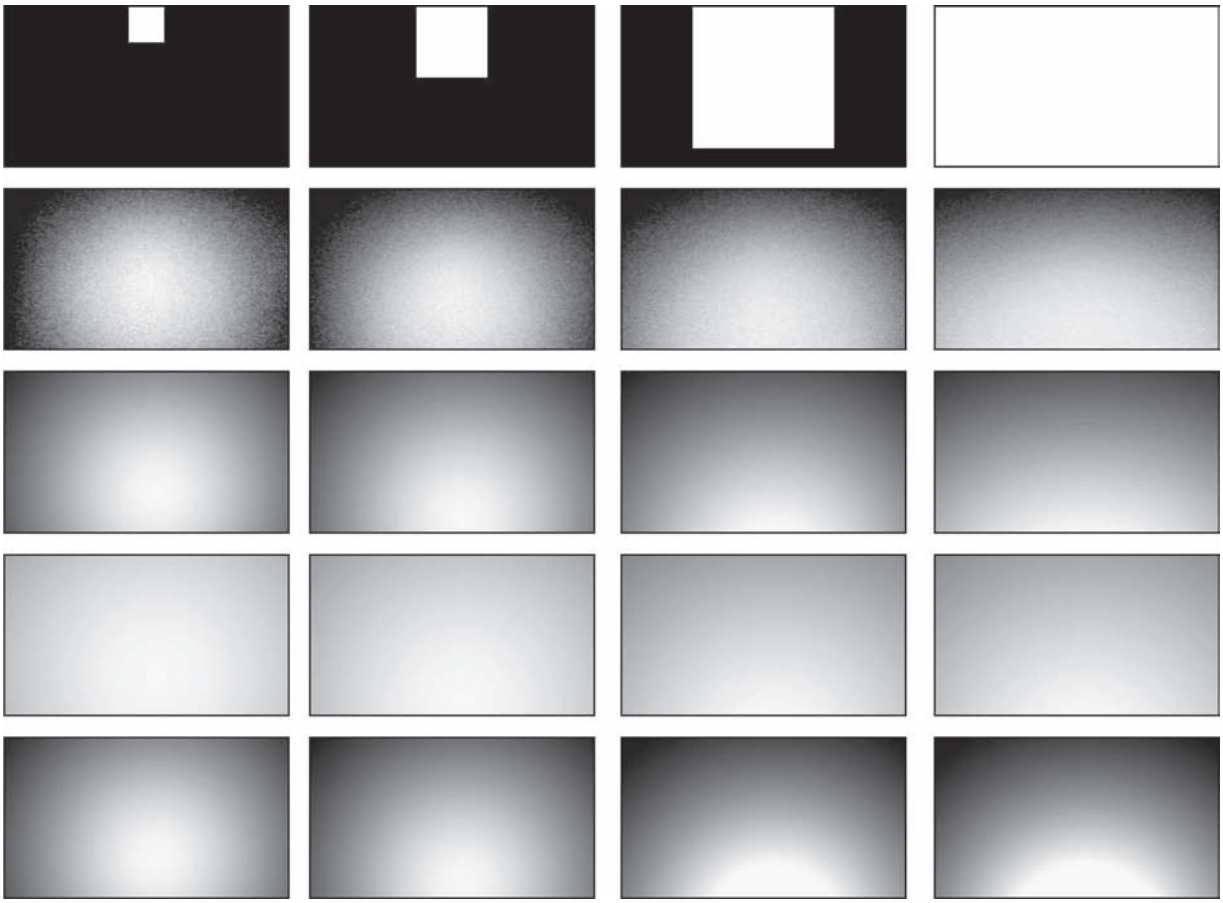

Fig. 6. Qualitative evaluation of light models for rectangular patterns Ω with varying sizes. From top to bottom: displayed patterns, with respective sizes 200×200, 400×400, 800×800 and 1600×900 (full size); data (left hand side of (21)); model (right hand side of (21)) using, respectively, our model (Proposition 3), an isotropic extended model [4], or an anisotropic point–source model [2].

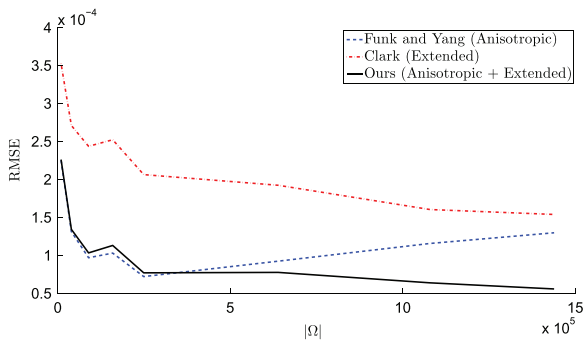

Fig. 7. Evolution of the RMSE between the data and different lighting models, according to the size $|\Omega|$ of the patterns.

deal with minimal cases. Why dealing with minimal cases is important? In addition to the theoretical interest one can find in studying minimal cases, solutions from minimal cases are of practical interest because they allow one to exhaustively evaluate all the possible poses [15]. In a sense, this yields the foundations to robust approaches as, by computing minimal solutions from a large number of mirror poses and reference points, one can check all possible poses by minimizing for each one a strongly nonlinear energy, as in the recent approach [16], and select the one with the smallest residual. Each possible pose can then be interpreted, in terms of the nonlinear energy, as a point inside each local convergence basin, so that selecting one convergence basin comes down to selecting one of the possible poses. This pose is naturally a good candidate for nonlinear optimization. In addition to lie within the convergence basin of the global minimum, the ideal candidate for initialization should obviously be quick to compute.

In this work we show that limiting the rotation of the camera to a single angle, as described in the next paragraph, reduces the minimal case to $n=3$ reference points and $k=2$ mirror poses. A unique solution can be found by introducing a new simple geometric criterion based on line intersection.

3.2. Assumptions

A study of the camera/screen calibration problem for smartphones was recently conducted by Delaunoy et al. in [16]. The authors report experimental results which indicate that the camera orientation is basically exactly the same as that of the screen.

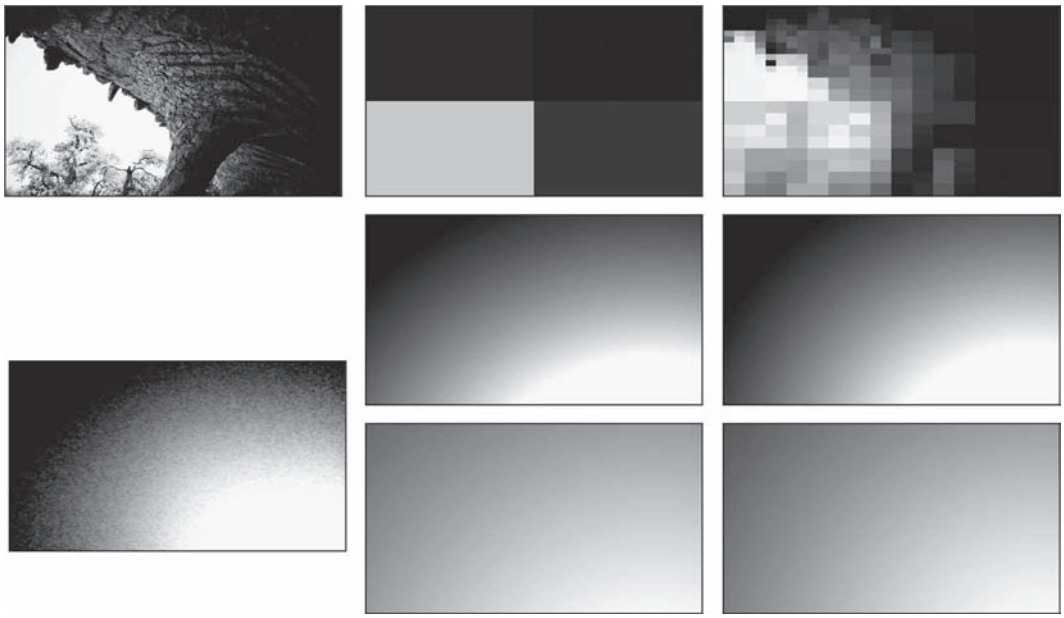

Fig. 8. Qualitative evaluation of the model on natural images. Left: image displayed on the screen (top), and normalized luminance measured on the sheet (bottom). Middle: image approximation by $n=4$ rectangles (top); luminance simulated using our model (middle), luminance simulated using Clark's model [4] (bottom). Right: same with $n=256$. Note that, due to the configuration of the device, a bright area in the left of the displayed image results in a bright area in the right of the sheet, as captured by the webcam.

Table 1
RMSE (multiplied by 10^4) between the measured normalized luminances and those simulated according to an extended isotropic light model [4] or to ours, for the 10 images shown in Fig. 5. Our model systematically outperforms Clark's, confirming the importance of considering anisotropy.

Real image	$n=4$		$n=64$		$n=256$	
	[4]	Ours	[4]	Ours	[4]	Ours
Dog	1.14	**0.109**	1.12	**9.8**	1.10	**0.98**
Snow	2.20	**0.78**	2.19	**0.67**	2.19	**0.66**
Sun	1.46	**1.17**	1.55	**1.14**	1.56	**1.15**
Lake	1.76	**1.11**	1.62	**0.97**	1.62	**0.95**
Wall	1.34	**0.94**	1.53	**0.94**	1.55	**0.98**
Graveyard	1.05	**1.04**	1.15	**0.82**	1.17	**0.79**
Building	1.36	**0.96**	1.42	**0.88**	1.43	**0.88**
Street	1.43	**1.13**	1.66	**0.57**	1.70	**0.55**
Flower	1.90	**1.03**	1.95	**0.97**	1.96	**0.98**
Cave	3.05	**1.26**	2.76	**1.02**	2.76	**1.01**

We deal with the case where the camera of the considered multimedia devices is assumed to be integrated or clipped onto the screen so it can be moved around an axis parallel to the screen's x-axis (or equivalently, the y-axis). Thus, we allow the webcam to have one degree of freedom, as shown in Fig. 9. With this hypothesis, the problem tackled in the following is that of estimating the camera location $\mathbf{t}$ in screen 3D-coordinates, and the angle θ characterizing the rotational part.

As in recent previous works [15,16], we assume that the intrinsic parameters of the camera are known in advance and that the distortion is already corrected. Typically, we can run any publicly available algorithm to estimate these parameters. From a practical point of view, we intrinsically calibrate the camera and compute undistorted images using the plane-based approach in [22], from multiple images of a chessboard.

3.3. Geometric model

3.3.1. Change of coordinates

In this work, the three-dimensional world coordinate system is the Euclidean coordinate system attached to the screen. If $\mathbf{x} \in \mathbb{R}^3$ represents the Cartesian coordinates of a 3D-point in the screen coordinate system then the coordinates $\mathbf{x}'$ of the same point in the Euclidean coordinate system attached to the camera are given by:

$$\mathbf{x}' = \mathbf{R}^\top (\mathbf{x} - \mathbf{t}) \tag{22}$$

where $\mathbf{R} \in \mathbb{R}^{3\times 3}$ is the rotation matrix describing the orientation of the camera in the screen coordinate system, and $\mathbf{t} \in \mathbb{R}^3$ is the camera center. It is necessary to use this relation whenever a variable described in the screen system (such as the light flux in Section 2) needs to be referred to in the camera system (as in the photometric stereo application described in Section 4). Hence, both $\mathbf{R}$ and $\mathbf{t}$ need to be estimated.

The 3D-points $\mathbf{x}$ are furthermore projected onto image 2D-points $\mathbf{x}_p$ of the camera according to the pinhole projection equation:

$$\left[\mathbf{x}_p^\top, 1\right]^\top \sim \mathbf{K}\mathbf{R}^\top [\mathbf{I}|-\mathbf{t}]\left[\mathbf{x}^\top, 1\right]^\top \tag{23}$$

where $\sim$ denotes the projective equality, and $\mathbf{K} \in \mathbb{R}^3$ is the (upper-triangular) calibration matrix of the camera intrinsic parameters commonly defined as in [23, p. 163]. Yet, the points $\mathbf{x}$ lying on the screen plane are not directly visible from the camera. Thus, a planar mirror should be

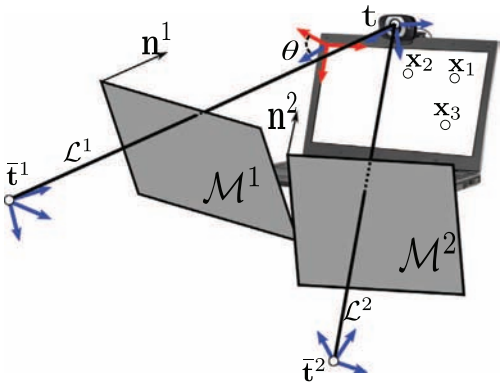

Fig. 9. Geometric setup. To estimate the location $\mathbf{t}$ of the camera and the angle θ describing its orientation w.r.t. the screen (red system), we use $k \geq 2$ poses $\mathcal{M}^i$, $i=1\ldots k$, of a planar mirror. By matching at least $n=3$ reference points ($\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ here) with their images captured by the webcam after reflections in the mirror, one estimate of θ can be obtained for each mirror pose, and the camera center can be estimated as the intersection of the lines $\mathcal{L}^i$, passing through the images $\overline{\mathbf{t}}^i$ of $\mathbf{t}$ by reflections, and oriented along the normals $\mathbf{n}^i$ characterizing the mirror poses. See also Figs. 10 (2D-view) and 11 (real-world example).

introduced in the setup, and one should use k poses of this mirror in order to estimate $\mathbf{R}$ and $\mathbf{t}$.

3.3.2. *Reflections in a planar mirror*

We now consider the ith pose of the mirror. It is characterized by the unit normal $\mathbf{n}^i \in \mathbb{R}^3$ of its supporting plane, oriented towards the camera, and the orthogonal distance $d^i \in \mathbb{R}$ from this plane to the origin of the screen coordinate system. Actually, the mirror acts like a "generator" of virtual 3D-points by reflecting (in the geometrical sense) real 3D-points in the mirror plane. Keep in mind that these virtual points will be seen from the camera viewpoint when there is no direct view of the corresponding real points. The virtual point $\overline{\mathbf{x}}^i$ (Fig. 10), the so-called mirrored point, which is the reflection of a real point x in the mirror plane w.r.t. its ith pose, satisfies the reflection equation:

$$\overline{\mathbf{x}}^i = \underbrace{\left[\mathbf{I} - 2\mathbf{n}^i\mathbf{n}^{i\top}\right]}_{\overline{\mathbf{U}}^i \in \mathbb{R}^{3\times 3}} \mathbf{x} - 2d^i\mathbf{n}^i \tag{24}$$

where $\overline{\mathbf{U}}^i$ is orthogonal with determinant -1.

The images captured by the camera are then obtained by applying the projection Eq. (23) to the mirrored point $\overline{\mathbf{x}}^i$, that is to say:

$$\left[\mathbf{x}_p^{i\top}, 1\right]^\top \sim \mathbf{K}\mathbf{R}^\top[\mathbf{I}|-\mathbf{t}]\mathbf{U}^i\left[\mathbf{x}^\top, 1\right]^\top \tag{25}$$

where $\mathbf{U}^i \in \mathbb{R}^{4\times 4}$ is the reflection matrix in homogeneous coordinates w.r.t. the ith pose of the mirror:

$$\mathbf{U}^i = \begin{bmatrix} \overline{\mathbf{U}}^i & -2d^i\mathbf{n}^i \\ \mathbf{0}^\top & 1 \end{bmatrix} \tag{26}$$

Let us now introduce a dual reformulation of this setup in terms of "virtual cameras", which are those obtained by considering reflections of the real camera.

3.3.3. *A geometric interpretation of the reflections as virtual cameras*

Another interpretation of Eq. (25) can be obtained by remarking that:

$$[\mathbf{I}|-\mathbf{t}]\mathbf{U}^i = \left[\overline{\mathbf{U}}^i|-(\mathbf{t}+2d^i\mathbf{n}^i)\right] \tag{27}$$

$$[\mathbf{I}|-\mathbf{t}]\mathbf{U}^i = \overline{\mathbf{U}}^i[\mathbf{I}|-\underbrace{\left(\overline{\mathbf{U}}^i\mathbf{t}-2d^i\mathbf{n}^i\right)}_{\overline{\mathbf{t}}^i}] \tag{28}$$

$$[\mathbf{I}|-\mathbf{t}]\mathbf{U}^i = \overline{\mathbf{U}}^i\left[\mathbf{I}|-\overline{\mathbf{t}}^i\right] \tag{29}$$

where we used the identities $(\overline{\mathbf{U}}^i)^2 = \mathbf{I}$ and $\overline{\mathbf{U}}^i\mathbf{n}^i = -\mathbf{n}^i$ to go from (27) to (29). Hence, $\mathbf{x}_p^i$ can also be seen as the image of real 3D-point $\mathbf{x}$ by a "virtual" camera:

$$[\mathbf{x}_p^{i\top}, 1]^\top \sim \mathbf{K}\underbrace{\mathbf{R}^\top\overline{\mathbf{U}}^i}_{\overline{\mathbf{R}}^{i\top}}\left[\mathbf{I}|-\overline{\mathbf{t}}^i\right]\left[\mathbf{x}^\top, 1\right]^\top \tag{30}$$

where $\overline{\mathbf{t}}^i$ is the virtual camera location, the indirect orthogonal matrix $\overline{\mathbf{R}}^i = \overline{\mathbf{U}}^i\mathbf{R}$ is the virtual camera orientation (it is not a rotation matrix since $\det \overline{\mathbf{R}}^i = -1$), while the intrinsic parameters are the same as the real camera. Obviously, this camera is nothing else than that obtained by reflection of the real camera while both cameras produce exactly the same images for a real point and its corresponding virtual reflection.

3.4. *Estimation of **R** from a single mirror pose*

For each mirror pose, as the virtual camera intrinsic parameters are *de facto* known, the virtual camera pose $(\overline{\mathbf{R}}^i, \overline{\mathbf{t}}^i)$ can be unambiguously estimated using perspective-n-points (PnP) algorithms, from at least $n=4$ matched pairs $\{(\mathbf{x}, \mathbf{x}_p^i)\}$, for a cost of $O(n)$ in most recent approaches [24]. When $n=3$, the problem is known as perspective-3-points (P3P) [25], and there is a fourfold ambiguity i.e., four possible solution-pairs $(\overline{\mathbf{R}}^i, \overline{\mathbf{t}}^i)$ exist (cf. Section 3.6).

Let us first assume that $n \geq 4$. The solution $(\overline{\mathbf{R}}^i, \overline{\mathbf{t}}^i)$ is unique, yet we are interested in recovering $(\mathbf{R}, \mathbf{t})$ from $(\overline{\mathbf{R}}^i, \overline{\mathbf{t}}^i)$. The equations to be solved are hence:

$$\begin{cases} \overline{\mathbf{U}}^i\mathbf{R} = \overline{\mathbf{R}}^i \\ \overline{\mathbf{U}}^i\mathbf{t} - 2d^i\mathbf{n}^i = \overline{\mathbf{t}}^i \end{cases} \Longleftrightarrow \begin{cases} \mathbf{R} = \overline{\mathbf{U}}^i\overline{\mathbf{R}}^i \\ \mathbf{t} = \overline{\mathbf{U}}^i\overline{\mathbf{t}}^i - 2d^i\mathbf{n}^i \end{cases} \tag{31}$$

but they cannot be directly used to recover $\mathbf{R}$ and $\mathbf{t}$, since $\overline{\mathbf{U}}^i$, d^i and $\mathbf{n}^i$ are unknown.

3.4.1. *Estimation of **R***

Even if we use the constraints saying that $\mathbf{R}$ is a direct orthogonal matrix, $\overline{\mathbf{R}}^i$ is an indirect orthogonal matrix, and $\overline{\mathbf{U}}^i$ is a symmetry matrix, the first equation of (31) admits an infinity of solutions $(\overline{\mathbf{U}}^i, \mathbf{R})$. For instance, denoting $\overline{\mathbf{R}}^i = [\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3]^\top$, the (trivial) solution:

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, [\mathbf{l}_1, \mathbf{l}_2, -\mathbf{l}_3]^\top\right)$$

is valid. Furthermore, if we post-multiply the solution for $\overline{\mathbf{U}}^i$ by an arbitrary rotation matrix and we pre-multiply the

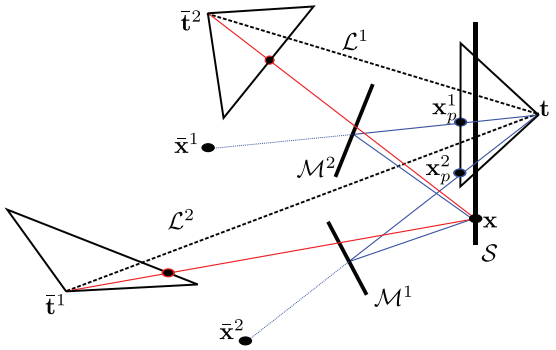

Fig. 10. 2D-representation of the geometric model. A point $\mathbf{x}$ on the screen $\mathcal{S}$ is reflected by the mirror $\mathcal{M}^i$, $i=1,2$, and projected on the pixel $\mathbf{x}_p^i$ of the real camera (blue line), as if the camera was directly observing the mirrored point $\bar{\mathbf{x}}^i$. On the other hand, the reflection by the mirror $\mathcal{M}^i$ defines a mirrored (virtual) camera which directly observes the real point $\mathbf{x}$ (red line).

solution for $\mathbf{R}$ by the transpose of this rotation matrix, the obtained solutions remain valid. Yet, such solutions are not consistent with the assumption discussed in Section 3.2, which constrains the rotation to a single angle around the x-axis:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{32}$$

where $\theta \approx 0$ for multimedia devices with integrated camera [16], and $\theta \neq 0$ for cameras clipped onto a screen, as in the example of Fig. 9.

This constraint reduces the number of possible solutions, but we know that there exists at least one exact solution, corresponding to the real rotation. Is this solution unique? From the equation $\overline{\mathbf{U}}^i\overline{\mathbf{R}}^i = \mathbf{R}$, and using (32), we obtain a linear system of equations for the six independent coefficients of the symmetric matrix $\overline{\mathbf{U}}^i$:

$$\begin{bmatrix} \overline{\mathbf{U}}^i_{11} & \overline{\mathbf{U}}^i_{12} & \overline{\mathbf{U}}^i_{13} \\ \overline{\mathbf{U}}^i_{12} & \overline{\mathbf{U}}^i_{22} & \overline{\mathbf{U}}^i_{23} \\ \overline{\mathbf{U}}^i_{13} & \overline{\mathbf{U}}^i_{23} & \overline{\mathbf{U}}^i_{33} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{R}}^i_{11} & \overline{\mathbf{R}}^i_{12} & \overline{\mathbf{R}}^i_{13} \\ \overline{\mathbf{R}}^i_{21} & \overline{\mathbf{R}}^i_{22} & \overline{\mathbf{R}}^i_{23} \\ \overline{\mathbf{R}}^i_{31} & \overline{\mathbf{R}}^i_{32} & \overline{\mathbf{R}}^i_{33} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{33}$$

which provides us with nine equations. Keeping only the four ones where we have a 0 value in the right-hand side, and both those obtained by remarking that $\mathbf{R}_{22} = \mathbf{R}_{33}$ and $\mathbf{R}_{23} = -\mathbf{R}_{32}$, those six equations can be synthetized into a homogeneous 6×6 linear system:

$$\mathbf{C}\,\mathbf{y}^i = \mathbf{0} \tag{34}$$

where:

$$\mathbf{C} = \begin{bmatrix} \overline{\mathbf{R}}^i_{12} & \overline{\mathbf{R}}^i_{22} & \overline{\mathbf{R}}^i_{32} & 0 & 0 & 0 \\ \overline{\mathbf{R}}^i_{13} & \overline{\mathbf{R}}^i_{23} & \overline{\mathbf{R}}^i_{33} & 0 & 0 & 0 \\ 0 & \overline{\mathbf{R}}^i_{11} & 0 & \overline{\mathbf{R}}^i_{21} & \overline{\mathbf{R}}^i_{31} & 0 \\ 0 & 0 & \overline{\mathbf{R}}^i_{11} & 0 & \overline{\mathbf{R}}^i_{21} & \overline{\mathbf{R}}^i_{31} \\ 0 & \overline{\mathbf{R}}^i_{12} & -\overline{\mathbf{R}}^i_{13} & \overline{\mathbf{R}}^i_{22} & \overline{\mathbf{R}}^i_{32} - \overline{\mathbf{R}}^i_{23} & -\overline{\mathbf{R}}^i_{33} \\ 0 & \overline{\mathbf{R}}^i_{13} & \overline{\mathbf{R}}^i_{12} & \overline{\mathbf{R}}^i_{23} & \overline{\mathbf{R}}^i_{22} + \overline{\mathbf{R}}^i_{33} & \overline{\mathbf{R}}^i_{32} \end{bmatrix} \tag{35}$$

and

$$\mathbf{y}^i = \left[\overline{\mathbf{U}}^i_{11},\ \overline{\mathbf{U}}^i_{12},\ \overline{\mathbf{U}}^i_{13},\ \overline{\mathbf{U}}^i_{22},\ \overline{\mathbf{U}}^i_{23},\ \overline{\mathbf{U}}^i_{33}\right]^\top \tag{36}$$

When the system (34) is well-determined, its solution in $\mathbf{y}^i$ is unique and can be computed by solving a total least-squares problem through the singular value decomposition of $\mathbf{C}$. As the unicity of solution $\overline{\mathbf{U}}^i$ depends on the rank of $\mathbf{C}$, it can be proven that, in general, the solution is unique. Nevertheless, degenerate configurations exist: for instance, if $(\overline{\mathbf{R}}^i_{12}, \overline{\mathbf{R}}^i_{13}) = (0,0)$ i.e., the axis of rotation of the mirror is the same as that of the true camera, then the solution is not unique. A simple way to detect such degenerate cases is to check the singular values of $\mathbf{C}$.

From the solution for $\mathbf{y}^i$ in Eq. (34), we construct the candidate reflection matrix $\overline{\mathbf{U}}^i$, which is normalized by observing that $\overline{\mathbf{U}}^i$ is a symmetry matrix with $\det \overline{\mathbf{U}}^i = -1$. It is then straightforward to obtain θ using:

$$\cos\theta = (\overline{\mathbf{U}}^i\overline{\mathbf{R}}^i)_{22}; \quad \sin\theta = (\overline{\mathbf{U}}^i\overline{\mathbf{R}}^i)_{32} \tag{37}$$

3.4.2. *Estimation of* $\mathbf{n}^i$

Before estimating $\mathbf{t}$, it is necessary to deduce $\mathbf{n}^i$ from $\overline{\mathbf{U}}^i$, knowing from (24) that:

$$\overline{\mathbf{U}}^i = \mathbf{I} - 2\mathbf{n}^i\mathbf{n}^{i\top} \Longleftrightarrow \mathbf{n}^i\mathbf{n}^{i\top} = \tfrac{1}{2}\left(\mathbf{I} - \overline{\mathbf{U}}^i\right) \tag{38}$$

Writing the singular value decomposition of the (symmetric) second member leads to:

$$\mathcal{U}^i\Sigma^i\mathcal{U}^{i\top} = \tfrac{1}{2}\left(\mathbf{I} - \overline{\mathbf{U}}^i\right) \tag{39}$$

where $\mathcal{U}^i$ is an order-3 orthogonal matrix, and Σ^i is the diagonal matrix of the singular values, sorted by descending order. Then, $\mathbf{n}^i = \pm\mathcal{U}^i[1,0,0]^\top$, where we solve the residual ambiguity on the sign of $\mathbf{n}^i$, assuming that the mirror is oriented towards the camera.

Knowing from (31) that:

$$\mathbf{t} = \overline{\mathbf{U}}^i\overline{\mathbf{t}}^i - 2d^i\mathbf{n}^i \tag{40}$$

the value of $\mathbf{t}$ depends on d^i, which is still unknown. Hence, this shows that it is not possible to estimate $\mathbf{t}$ from a single mirror pose.

3.5. Estimation of $\mathbf{t}$ from $k \geq 2$ mirror poses

Given k mirror poses, we obtain a system of $3k$ equations (40) in $3+k$ unknowns $(\mathbf{t}, \{d^i\})$:

$$\begin{cases} \mathbf{t}+2d^1\mathbf{n}^1 = \overline{\mathbf{U}}^1\overline{\mathbf{t}}^1 \\ \vdots \\ \mathbf{t}+2d^k\mathbf{n}^k = \overline{\mathbf{U}}^k\overline{\mathbf{t}}^k \end{cases} \tag{41}$$

(in the case $k=2$, we obtain a system of 6 linear equations in 5 unknowns $(\mathbf{t}, d^1, d^2)$).

This problem comes down to intersecting k straight lines in $\mathbb{R}^3$. The interpretation of Section 3.3.3 in terms of virtual cameras provides us with a geometric interpretation of this line intersection problem. Indeed, remembering that $\overline{\mathbf{U}}^i = \mathbf{I} - 2\mathbf{n}^i\mathbf{n}^{i\top}$, Eq. (40) is rewritten as:

$$\mathbf{t}+2\left(\underbrace{\mathbf{n}^{i\top}\overline{\mathbf{t}}^i+d^i}_{\alpha^i}\right)\mathbf{n}^i = \overline{\mathbf{t}}^i \tag{42}$$

which indicates that, in the absence of noise on data, the true camera center is exactly located at the intersection of the k lines $\mathcal{L}^i$, $i=1\ldots k$, passing through the centers $\overline{\mathbf{t}}^i$ of the virtual cameras, and oriented by the vectors $\mathbf{n}^i$ (mirror normals), as illustrated in Figs. 9 and 10. Hence, as long as the vectors $\mathbf{n}^i$ are not collinear, this problem should admit exactly one solution in the ideal case where the lines $\mathcal{L}^i$ actually intersect each others. Due to noisy measurements and numerical approximations, this is obviously wrong in real-world scenarios, but an approximate solution can be found by solving the intersection problem in the least-squares sense.

Eq. (42) gives rise to a new linear system with $3k$ equations and $3+k$ unknowns $(\mathbf{t}, \{\alpha^i\})$, thus overdetermined as soon as $k \geq 2$. When $k=2$, it is written as:

$$\underbrace{\begin{bmatrix} \mathbf{I} & 2\mathbf{n}^1 & \mathbf{0}_{3\times1} \\ \mathbf{I} & \mathbf{0}_{3\times1} & 2\mathbf{n}^2 \end{bmatrix}}_{\mathbf{A}\in\mathbb{R}^{6\times5}} \underbrace{\begin{bmatrix} \mathbf{t} \\ \alpha^1 \\ \alpha^2 \end{bmatrix}}_{\mathbf{u}\in\mathbb{R}^5} = \underbrace{\begin{bmatrix} \overline{\mathbf{t}}^1 \\ \overline{\mathbf{t}}^2 \end{bmatrix}}_{\mathbf{b}\in\mathbb{R}^6} \tag{43}$$

This system usually admits no exact solution, but its (ordinary) least-squares solution can be obtained by solving the associated normal equations:

$$\begin{bmatrix} \mathbf{I} & \mathbf{n}^1 & \mathbf{n}^2 \\ \mathbf{n}^{1\top} & 2 & 0 \\ \mathbf{n}^{2\top} & 0 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \alpha^1 \\ \alpha^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ 2\mathbf{n}^{1\top} & \mathbf{0}_{1\times3} \\ \mathbf{0}_{1\times3} & 2\mathbf{n}^{2\top} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{t}}^1 \\ \overline{\mathbf{t}}^2 \end{bmatrix} \tag{44}$$

It can be shown that the determinant of the pseudoinverse of $\mathbf{A}$, defined in Eq. (43), is equal to $\|\mathbf{n}^1 \times \mathbf{n}^2\|^2$. Hence, as predicted, as long as both mirror poses are not parallel and non-degenerate (so that $\overline{\mathbf{U}}^i$ and $\mathbf{n}^i$, $i=1,2$, are unambiguously determined), the (approximate) solution in $(\mathbf{t}, \alpha^1, \alpha^2)$ is unique.

Clearly, robustness is improved when ensuring the angle between $\mathbf{n}^1$ and $\mathbf{n}^2$ is high enough, and when considering more than $k=2$ poses. In this case, removing outliers according to any outlier detection heuristic such as RANSAC [26] may be worthwhile. Since each pose provides us with an estimation of θ, a robust estimation of this angle can also be performed at this step.

Eventually, let us remark that the system $\mathbf{Au}=\mathbf{b}$, defined in Eq. (43), satisfies

$$\|\mathbf{Au}-\mathbf{b}\|^2 = \sum_{i=1}^{k} d_\perp(\mathbf{t}, \mathcal{L}^i)^2 \tag{45}$$

which provides a geometric interpretation of its residuals in terms of the orthogonal distances $d_\perp(\mathbf{t}, \mathcal{L}^i)$ from the estimate $\mathbf{t}$ to the lines $\mathcal{L}^i$.

We can illustrate the proposed approach by applying our algorithm to images captured by the laptop considered

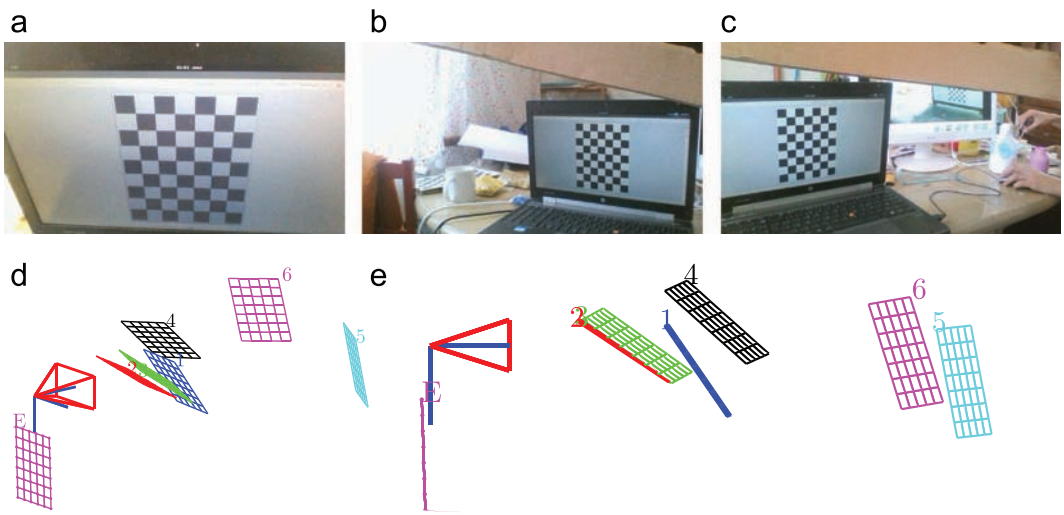

Fig. 11. Example result of geometric calibration of a laptop with integrated camera. (a–c) Three out of the six calibration images captured by the webcam. A chessboard pattern is displayed on the screen, and a mirror provides 48 correspondences per image through reflection (the mirror is partly visible on the right images). (d–e) Reconstructed geometric setup (the images (a–c) correspond to the poses 1, 5 and 6 of the reconstruction, and E describes the calibration pattern located on the screen). For this device, we estimated $\theta \approx -3°$.

in the experiments of Section 2, as seen in Fig. 11. In such real-world applications, in order to have at ones disposal a large number of point matches between image-pairs, we use a chessboard pattern whose corners are easily detected by standard algorithms.

3.6. Estimation of $(\mathbf{R}, \mathbf{t})$ in the ambiguous case where $n=3$

When only $n=3$ matched pairs $\{(\mathbf{x}, \mathbf{x}_p^i)\}$ are available, the pose $(\overline{\mathbf{R}}^i, \overline{\mathbf{t}}^i)$ of the ith virtual camera can still be estimated by any P3P algorithm, but only up to a fourfold ambiguity. Thus, if k poses of the mirror are considered, the previous rationale provides us with 4^k possible solutions for $(\overline{\mathbf{R}}^i, \overline{\mathbf{t}}^i)$. We can exhaustively compute all these solutions, and select the one with the lowest geometric residual (45). This heuristic is similar to the state-of-the-art method of Takahashi et al. [15], who assess all their candidates so as to keep the one which fits best an algebraic constraint. As stated earlier, such minimal case solutions can be considered "good enough" initial estimates for most accurate iterative calibration methods e.g., like in [16].

To evaluate our approach taking into account constraints on the orientation of the camera, we will thus compare against [15] in this minimal case, keeping in mind that both calibration methods can obviously be made increasingly accurate by using more pairs of matched points and more mirror poses.

3.7. Quantitative evaluation

As in [15], we assume to have at ones disposal a minimal input data consisting of a set of $n=3$ reference points and $k=3$ mirror poses. Given the matched pairs $\{(\mathbf{x}, \mathbf{x}_p^i)\}$, we computed the four admissible poses of each virtual camera using a standard P3P algorithm [26], and estimated the camera pose using the state-of-the-art approach from [15] and ours. For both methods, we measured the mean Riemannian distance $E_{\mathbf{R}}$ between the estimated $\mathbf{R}$ and the ground truth matrix, and the RMSE $E_{\mathbf{t}}$ on $\mathbf{t}$, for several levels of noise on the 2D-observations (zero-mean Gaussian noise with standard deviation σ) and different camera-mirror distances $d^i = d$, $i = 1 \ldots k$.

In these synthetic experiments, we used the same intrinsic parameters as in [15], and similar values for the

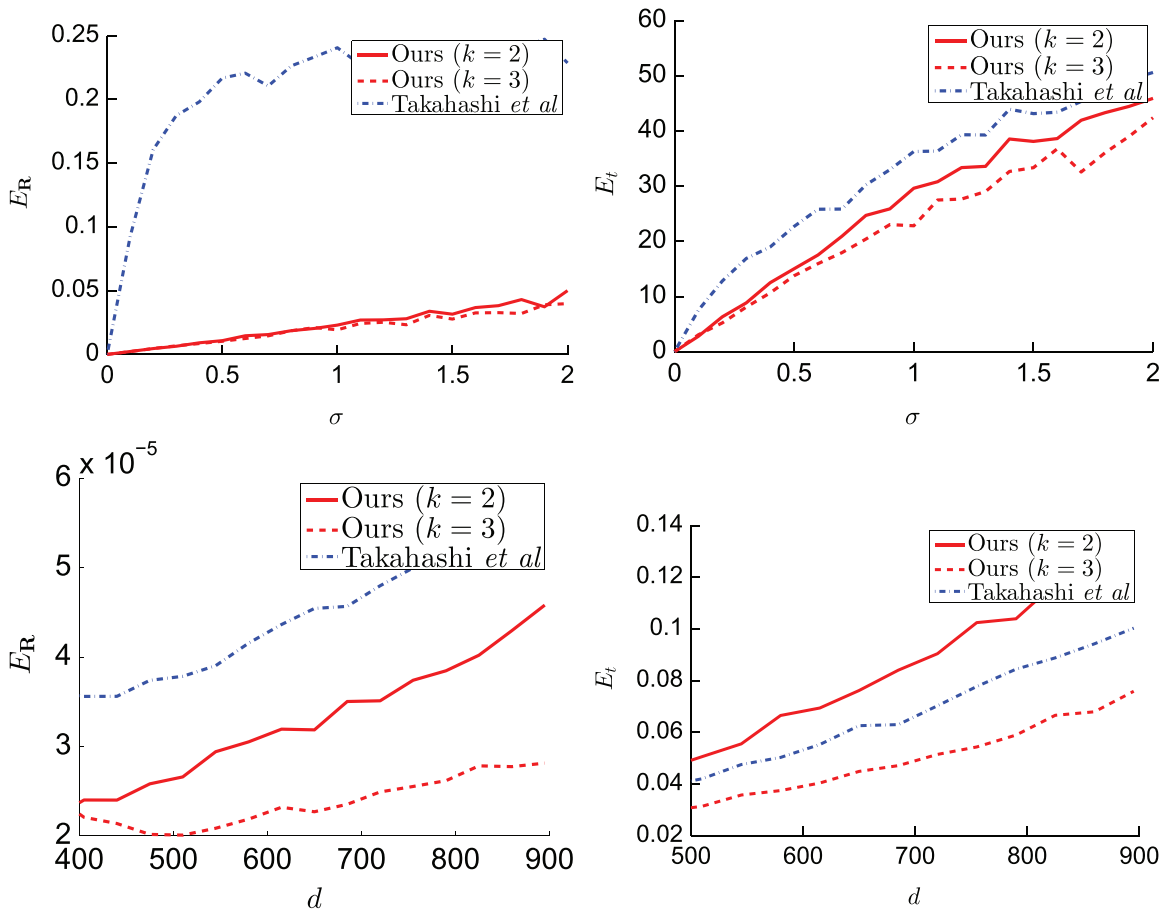

Fig. 12. Quantitative evaluation of the proposed extrinsic calibration method. We show the error rate $E_{\mathbf{R}}$ on the rotation matrix and the error rate $E_{\mathbf{t}}$ on the camera center location, against the noise level σ added to the 2D-observations (in pixels, top), and the orthogonal distance d to the mirrors (in pixels, bottom). Our method outperforms the state-of-the-art method from [15] with $k=3$, and reaches comparable results even in the case $k=2$, which is impossible to consider in [15].

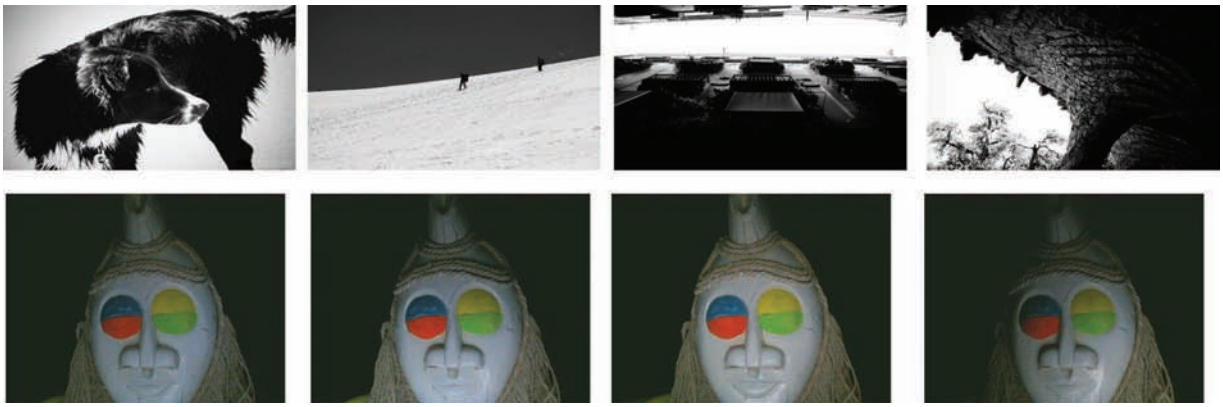

Fig. 13. As a person watches a slideshow of images (first row, we show 4 out of 40 images), pictures are recorded (second row). The pictures displayed on the screen serving as light sources, we employ the photometric stereo approach to recover the geometry and photometry of the scene.

other parameters. The distances d^i between the camera and the mirrors are set to $d^i = d = 500$, $\forall i = 1 \ldots k$. The reference points are $\mathbf{x}_1 = [0, 0, 0]^\top$, $\mathbf{x}_2 = [225, 0, 0]^\top$ and $\mathbf{x}_3 = [0, 225, 0]^\top$. The mirror normals are set to $\mathbf{n}^i = -[\cos a^i \sin b^i, \sin a^i \sin b^i, \cos b^i]^\top$, with $(a^1, a^2, a^3) = (\pi/4, -\pi/5, \pi/6)$ and $(b^1, b^2, b^3) = (\pi/7, -\pi/7, \pi/9)$. A Gaussian noise with zero-mean and standard deviation $\sigma = 0.01$ is added to the 2D-measurements. The rotation $\mathbf{R}$ is set to the identity, and $\mathbf{t}$ is generated by assigning a random value within $[0, 20]$ to each of its components. To obtain the results in Fig. 12, we performed 100 trials before meaning the error rates $E_{\mathbf{R}}$ and $E_{\mathbf{t}}$.

Results shown in Fig. 12 prove that the proposed method offers better performances against noise and distant mirrors than state-of-the-art. We also notice that using only $k=2$ mirror poses (we chose both first poses) offers acceptable performances on such synthetic data.

4. Application to 3D-reconstruction

Up to this point, we have introduced an explicit closed-form model for the light emitted by the screen, considered as an extended anisotropic source, and provided a thorough geometrical study of the camera pose estimation problem, using a minimal amount of inputs. Let us now describe, as an example application, a classical computer vision problem involving both these photometric and geometric constraints.

In the photometric stereo context [1], the 3D-reconstruction of a surface is obtained by successively illuminating the surface from various directions. Considering realistic lighting models for photometric stereo has recently become an important research direction [27,28], since neglecting radial and distance attenuation of light causes a strong low-frequency bias in real-world applications ("potato chip"-like 3D-reconstructions [29], see Fig. 17). Up to now, extended sources have not been really considered: apart from Clark's work [4], most photometric stereo approaches using such sources have considered infinitely distant [3,5–7] or pointwise [2] approximations.

In this section, we show how to use images displayed on the screen as extended light sources for photometric stereo (Fig. 13).

4.1. Photometric stereo setting

The screen successively displays m different images in front of a still person (Fig. 13). For simplicity, we consider here graylevel images, no additional lighting (black room setting), and assume that the luminance emitted by these images is the same in every channel. The case of color images illuminating a colored scene being way more complicated, it is left for future prospect.

Those m images behaving as m light sources, as described in Section 2, each graylevel image produces a light field $\mathbf{s}^i$ which is given by Proposition 4: at each surface point $\mathbf{x}$, we can thus define a light matrix $\mathbf{S}(\mathbf{x}) \in \mathbb{R}^{3\times m}$, by concatenating all the light vectors: $\mathbf{S}(\mathbf{x}) = [\mathbf{s}^1(\mathbf{x}), \ldots, \mathbf{s}^m(\mathbf{x})]$. In the same way, the $3m$ RGB values collected at pixel $\mathbf{x}_p$ are stacked in the matrix $\mathbf{I}(\mathbf{x}_p) \in \mathbb{R}^{m\times 3}$ defined by:

$$\mathbf{I}(\mathbf{x}_p) = \begin{bmatrix} I_R^1(\mathbf{x}_p) & I_G^1(\mathbf{x}_p) & I_B^1(\mathbf{x}_p) \\ & \vdots & \\ I_R^m(\mathbf{x}_p) & I_G^m(\mathbf{x}_p) & I_B^m(\mathbf{x}_p) \end{bmatrix} \tag{46}$$

According to Lambert's law, the image formation model is given by:

$$\mathbf{I}(\mathbf{x}_p) = \mathbf{S}(\mathbf{x})^\top \mathbf{n}(\mathbf{x}) \rho(\mathbf{x})^\top \tag{47}$$

where $\rho(\mathbf{x}) = [\rho_R(\mathbf{x}), \rho_G(\mathbf{x}), \rho_B(\mathbf{x})]^\top$ is the albedo vector, representing the percentage of light re-emitted by the surface in each channel, and $\mathbf{n}(\mathbf{x})$ is the unit outward normal to the surface. In Proposition 4, the light vectors are given in screen coordinates: they need to be converted into camera coordinates as described in Section 3.

4.2. Iterative resolution

The color photometric stereo model (47) is similar to that considered by Barsky and Petrou in [30], with the major difference that in our case the light matrix $\mathbf{S}(\mathbf{x})$ depends on $\mathbf{x}$, which increases the accuracy of the model, but also prevents us from obtaining a closed-form 3D-

reconstruction. To deal with this problem, we follow an iterative procedure which alternatively estimates the surface and updates the lights. Such iterations were already proposed in recent works dealing with near-light photometric stereo [31,27]. Given the current estimate $\mathbf{x}^q$ of the 3D-points representing the surface (in camera coordinates) at iteration q, a typical update writes:

1. use the camera pose to compute the 3D-points $\mathbf{x}^q$ in screen coordinates;
2. deduce the light vectors $\mathbf{s}^i(\mathbf{x}^q)$ in screen coordinates;
3. use the camera pose to compute these light vectors $\mathbf{s}^i(\mathbf{x}^q)$ in camera coordinates;
4. solve Eq. (47) by least-squares to estimate $\mathbf{n}(\mathbf{x}^q)$ and $\rho(\mathbf{x}^q)$ [30];
5. integrate the normals $\mathbf{n}(\mathbf{x}^q)$ into new 3D-points $\mathbf{x}^{q+1}$ [32].

In the experiments, we used as initial guess $\mathbf{x}^0$ a plane parallel to the screen at distance d, with d being an *a priori* estimate of the mean screen–object distance. The solution of the integration subproblem providing a solution only up to a global scale (perspective ambiguity), disambiguation was performed, as advised in [27], by setting to d the mean screen–object distance. This prevents any drift in the iterative process, which typically converges after 5–10 iterations [27]. In all our experiments, the algorithm was run until convergence, defined as a mean relative change of the 3D-points $\mathbf{x}^k$ smaller than 10^{-6}. Each iteration is around 10 s on a I7 processor, with non-optimized Matlab code.

4.3. Qualitative results

We finally present some qualitative 3D-reconstruction results using the same HP EliteBook laptop as in the

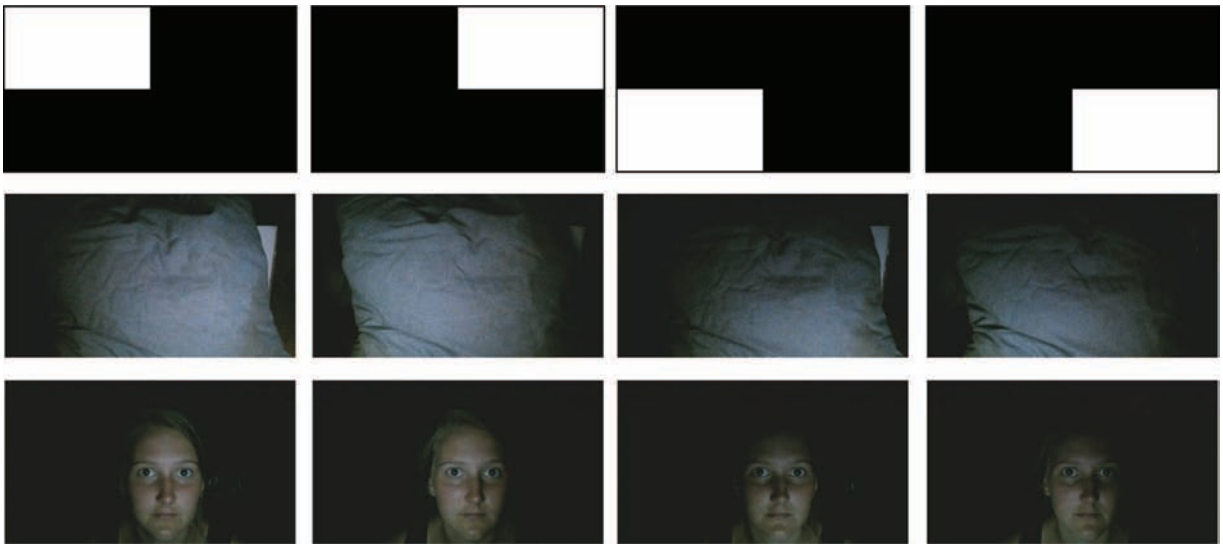

Fig. 14. Rectangular patterns used as illuminants. Top: the $m=4$ rectangular patterns displayed on the screen. Middle and bottom: the corresponding images captured by the webcam.

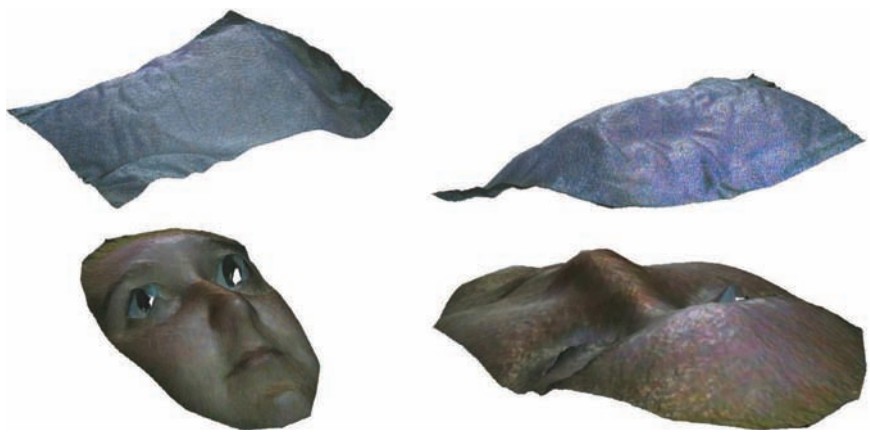

Fig. 15. 3D-reconstructions by the proposed photometric stereo technique, for the two series of $m=4$ images shown in Fig. 14. For the face dataset, we manually removed the eyes from the reconstruction domain, since they are purely specular. Also note that the reconstruction of the face appears more "noisy" than that of the pillow: this is the consequence of slight displacements of the person during the acquisition.

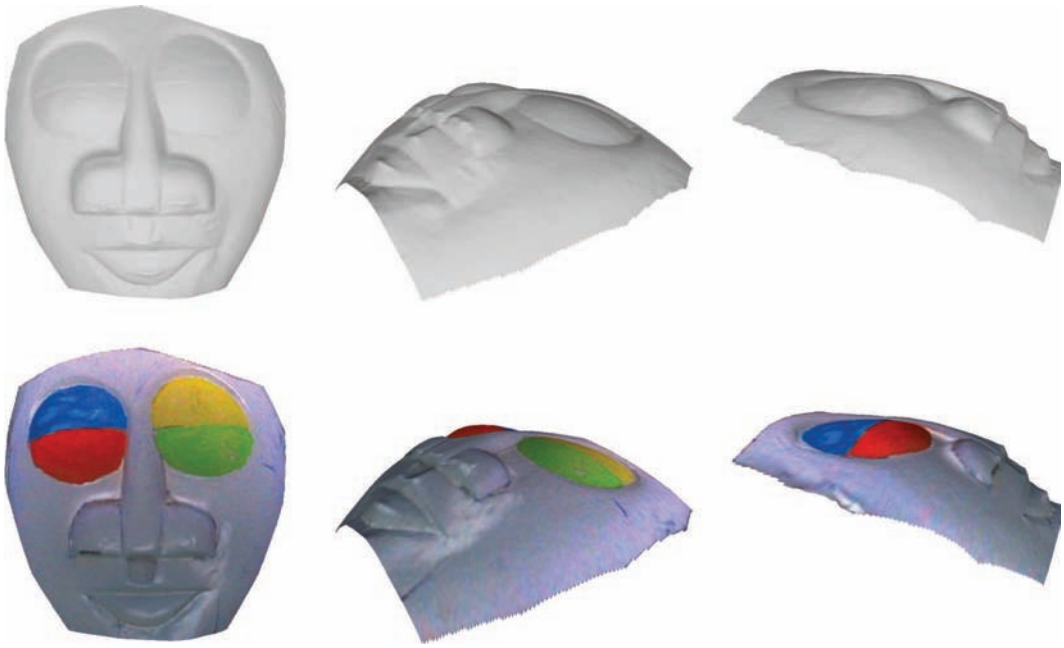

Fig. 16. Three views of a relighting of the 3D-reconstruction, with or without color albedo warped onto the surface, obtained from the images of Fig. 13. Note that shadows, highlights and discontinuities create some artifacts, since all the potential outliers to Lambert's law were ignored. Comparing this 3D-reconstruction to the one obtained using naive models (Fig. 17) confirms the importance of considering a full geometric and photometric model rather than naive simplifications.

Fig. 17. 3D-reconstruction by photometric stereo, using the standard directional approximation [3,5,7]. Neglecting the extended behavior of the screen creates a large-scale bias, preventing realistic applications.

experiments of Sections 2 and 3, which has a 1600×900 matte screen and a 640×480 integrated camera, whose pose was calibrated as described in Section 3. Experiments were conducted using both uniform rectangular patterns and natural images, considered as light sources as described in Section 2.

Let us first consider the usual case or rectangular illumination patterns [2,3,5,6]. In the experiments of Fig. 14, we used $m=4$ rectangles to illuminate a pillow and a human. Using the proposed photometric stereo method, we obtained the 3D-reconstructions shown in Fig. 15. Note that the locations of the rectangles are very different in each case, which maximizes the condition number of the illumination matrix $\mathbf{S}$ in every point of the scene and thus allows us to obtain very satisfactory 3D-reconstructions using few illumination patterns.

When dealing with natural images instead of homogeneous rectangles, we lose the ability to control this condition number, and more illumination conditions need to be introduced. We used as source images the 10 natural images shown in Fig. 5, that were flipped around the horizontal axis, the vertical axis and both axes, so as to obtain a total of 40 images with reasonable variations in the lighting directions (this "trick" was proposed by Clark in [4]). Results shown in Fig. 16 are qualitatively satisfactory: the reconstructed shape and reflectance are sufficiently realistic to be used for instance in augmented reality applications. Let us emphasize that metrological accuracy is not the objective here: for such applications of photometric stereo, much more controlled environments are usually considered [33,34], and the outliers to the model have to be treated. Shadows and highlights [30,35,36], as well as depth discontinuities [32], represent well-known difficulties. In the case of extended sources, the penumbra effects discussed in Section 2 would represent additional difficulties that are less studied: neglecting them causes the surface slope and the albedo to

be over-estimated in penumbra areas, as can be seen in Fig. 16.

Even though we neglected all these outliers, taking into account the extended behavior of the source already improved considerably, at least qualitatively, the accuracy of the 3D-reconstruction, compared to more naive models considered for instance in [3,5,7], as illustrated in Fig. 17 where the mean light directions and intensities were estimated using [37] and considered as models for the light instead of the proposed extended anisotropic model.

5. Conclusion and perspectives

We have tackled both the problems of realistically modelling the light field emitted by a graylevel image displayed on the screen of a multimedia device, and of geometrically calibrating an attached camera with respect to this screen. We first showed that a very general closed-form expression of an extended anisotropic planar illuminant with spatially-varying luminance could be obtained without empirical approximation, providing an accurate model for the light emitted by the screen. Then, we proposed a theoretical study of the pose estimation problem for multimedia devices, which incorporates the natural geometric constraints induced by such devices. Finally, we introduced a cheap and ludique 3D-reconstruction application, where a 3D-model of a person is reconstructed while watching a collection of images.

Up to this point, the main drawback of the proposed photometric stereo application is the CPU time required to iteratively refine the 3D-model. Yet, this is not much of an issue, since the relevant literature already offers two ways of accelerating screen-based photometric stereo applications. First, the process can be made multi-scale and ported onto a GPU, as did Nozick in [5]. Second, a single image can provide 3D-reconstruction, if one considers the screen is displaying color images: this is what Schindler studied in [3], following the approach of Hernandez et al. in [38]. Yet, the latter requires the observed scene to have uniform reflectance: studying the case of both colored illuminants and colored scene remains, to the best of our knowledge, an open problem which is an interesting future direction for research.

We also mentioned in Section 2.3 the problem of partial occlusion of the screen, resulting in penumbra effects. This is very easy to model, and to use in rendering through the raytracing technique. Yet, it is a much more complicated issue in the 3D-reconstruction framework, since visibility of a pixel from a point on the surface depends on the location of this point (and on the local orientation of the surface), which is precisely the unknown. Theoretically, this could be naturally handled using an iterative framework, by computing the visibility at each iteration, based on the previous estimation of the shape. Yet, this would require an efficient raytracer, and hence porting the whole application to GPU. We believe that such an extension would be a very interesting perspective, and open the door to efficient photometric 3D-reconstruction under a wide variety of extended sources, including natural indoor illumination (windows, neon lightings etc.).

We also plan to study how the proposed models can be used in other computer vision applications. For instance, accurately locating the screen w.r.t. the camera would be useful for gaze-tracking applications, which usually require introducing an additional device [39]. As shown in [40], the detection of one person's eye provides important 3D-clues: we believe that coupling such a technique with the proposed photometric and geometric models would result in an improved gaze-tracking system which would involve nothing but a computer screen and an integrated webcam.

References

[1] R.J. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.
[2] N. Funk, Y.-H. Yang, Using a raster display for photometric stereo, in: Fourth Canadian Conference on Computer and Robot Vision (CRV), 2007, pp. 201–207.
[3] G. Schindler, Photometric stereo via computer screen lighting for real-time surface reconstruction, in: Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), 2008.
[4] J.J. Clark, Photometric stereo using LCD displays, Image Vis. Comput. 28 (4) (2010) 704–714.
[5] V. Nozick, Pyramidal normal map integration for real-time photometric stereo, EAM Mechatron. (2010) 128–132.
[6] J.H. Won, M.H. Lee, I.K. Park, Active 3D shape acquisition using smartphones, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012, pp. 29–34.
[7] L. Bi, Z. Song, L. Xie, A novel LCD based photometric stereo method, in: 2014 IEEE International Conference on Information Science and Technology (IST), 2014, pp. 611–614.
[8] P. Belhumeur, D. Kriegman, A.L. Yuille, The bas-relief ambiguity, Int. J. Comput. Vis. 35 (1) (1999) 33–44.
[9] A.L. Yuille, D. Snow, R. Epstein, P.N. Belhumeur, Determining generative models of objects under varying illumination: shape and albedo from multiple images using SVD and integrability, Int. J. Comput. Vis. 35 (3) (1999) 203–222.
[10] R.A. Finkel, J.L. Bentley, Quad trees: a data structure for retrieval on composite keys, Acta Inform. 4 (1) (1974) 1–9.
[11] P. Sturm, T. Bonfort, How to compute the pose of an object without a direct view? in: Computer Vision—ACCV 2006, Lecture Notes in Computer Science, vol. 3852, 2006, pp. 21–31.
[12] R. Kumar, A. Ilie, J.-M. Frahm, M. Pollefeys, Simple calibration of non-overlapping cameras with a mirror, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
[13] J. Hesch, A. Mourikis, S. Roumeliotis, Mirror-based extrinsic camera calibration, in: Algorithmic Foundation of Robotics VIII, Springer Tracts in Advanced Robotics, vol. 57, 2010, pp. 285–299.
[14] R. Rodrigues, J. Barreto, U. Nunes, Camera pose estimation using images of planar mirror reflections, in: Computer Vision—ECCV 2010, Lecture Notes in Computer Science, vol. 6314, 2010, pp. 382–395.
[15] K. Takahashi, S. Nobuhara, T. Matsuyama, A new mirror-based extrinsic camera calibration using an orthogonality constraint, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1051–1058.
[16] A. Delaunoy, J. Li, B. Jacquet, M. Pollefeys, Two cameras and a screen: How to calibrate mobile devices? in: 2nd International Conference on 3D Vision (3DV), 2014, pp. 123–130.
[17] A. Agrawal, Extrinsic camera calibration without a direct view using spherical mirror, in: Computer Vision (ICCV), 2013 IEEE International Conference on, 2013, pp. 2368–2375.
[18] Y. Quéau, R. Modrzejewski, P. Gurdjos, J.-D. Durou, Transformation d'un dispositif multimédia webcam-écran en un scanner 3D, in: COmpression et REprésentation des Signaux Audiovisuels (CORESA), 2014 (in french).
[19] R. Mecca, A. Wetzler, A.M. Bruckstein, R. Kimmel, Near field photometric stereo with point light sources, SIAM J. Imaging Sci. 7 (4) (2014) 2732–2770.
[20] B.K.P. Horn, Robot Vision, MIT Press, Cambridge, Massachushetts, 1986.

[21] H. Stewénius, Gröbner basis methods for minimal problems in computer vision (Ph.D. thesis), Lund University, 2005.
[22] J.-Y. Bouguet, Camera Calibration Toolbox for Matlab.
[23] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, United Kingdom, 2003.
[24] V. Lepetit, F. Moreno-Noguer, P. Fua, Epnp: an accurate o(n) solution to the pnp problem, Int. J. Comput. Vis. 81 (2) (2009) 155–166.
[25] B. Haralick, C.-N. Lee, K. Ottenberg, M. Nölle, Review and analysis of solutions of the three point perspective pose estimation problem, Int. J. Comput. Vis. 13 (3) (1994) 331–356.
[26] M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[27] T. Papadhimitri, P. Favaro, Uncalibrated near-light photometric stereo, in: Proceedings of the British Machine Vision Conference (BMVC), 2014.
[28] R. Mecca, A. Tankus, A. Wetzler, A.M. Bruckstein, A direct differential approach to photometric stereo with perspective wiewing, SIAM J. Imaging Sci. 7 (2) (2014) 579–612.
[29] X. Huang, M. Walton, G. Bearman, O. Cossairt, Near light correction for image relighting and 3d shape recovery, in: International Conference on Digital Heritage, 2015.
[30] S. Barsky, M. Petrou, The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows, IEEE Trans. Pattern Anal. Mach. Intell. 25 (10) (2003) 1239–1252.
[31] T. Migita, S. Ogino, T. Shakunaga, Direct bundle estimation for recovery of shape, reflectance property and light position, in: Computer Vision—ECCV 2008, Lecture Notes in Computer Science, vol. 5304, 2008, pp. 412–425.
[32] J.-D. Durou, J.-F. Aujol, F. Courteille, Integrating the normal field of a surface in the presence of discontinuities, in: Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), Lecture Notes in Computer Science, vol. 5681, 2009, pp. 261–273.
[33] D. Vlasic, P. Peers, I. Baran, P. E. Debevec, J. Popovic, S. Rusinkiewicz, W. Matusik, Dynamic shape capture using multi-view photometric stereo, ACM Trans. Graph. 28 (5) (2009).
[34] M.K. Johnson, F. Cole, A. Raj, E.H. Adelson, Microgeometry capture using an elastomeric sensor, ACM Trans. Graph. 30 (4) (2011) 46:1–46:8.
[35] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, Y. Ma, Robust photometric stereo via low-rank matrix completion and recovery, in: Computer Vision—ACCV 2010, Lecture Notes in Computer Science, vol. 6494, 2011, pp. 703–717.
[36] S. Ikehata, D. Wipf, Y. Matsushita, K. Aizawa, Photometric stereo using sparse Bayesian regression for general diffuse surfaces, IEEE Trans. Pattern Anal. Mach. Intell. 36 (9) (2014) 1816–1831.
[37] Y. Quéau, F. Lauze, J.-D. Durou, Solving uncalibrated photometric stereo using total variation, J. Math. Imaging Vis. 52 (1) (2015) 87–107.
[38] C. Hernandez, G. Vogiatzis, G. Brostow, B. Stenger, R. Cipolla, Non-rigid photometric stereo with colored lights, in: 2007 IEEE International Conference on Computer Vision (ICCV), 2007.
[39] B. Noris, J.-B. Keller, A. Billard, A wearable gaze tracking system for children in unconstrained environments, Comput. Vis. Image Underst. 115 (4) (2011) 476–486.
[40] L. Calvet, P. Gurdjos, An enhanced structure-from-motion paradigm based on the absolute dual quadric and images of circular points, in: 2013 IEEE International Conference on Computer Vision (ICCV), 2013, pp. 985–992.