The Shading Isophotes: Model and Methods for Lambertian Planes and a Point Light

Damien Mariyanayagam Adrien Bartoli

ENCOV, IGT, Institut Pascal
UMR6602 CNRS / Université Clermont Auvergne

Corresponding author: Adrien Bartoli
adrien.bartoli@gmail.com

June 26, 2024

Abstract

Structure-from-Motion (SfM) and Shape-from-Shading (SfS) are complementary classical approaches to 3D vision. Broadly speaking, SfM exploits geometric primitives from textured surfaces and SfS exploits pixel intensity from the shading image. We propose an approach that exploits virtual geometric primitives extracted from the shading image, namely the level-sets, which we name shading isophotes. Our approach thus combines the strength of geometric reasoning with the rich shading information. We focus on the case of untextured Lambertian planes of unknown albedo lit by an unknown Point Light Source (PLS) of unknown intensity. We derive a comprehensive geometric model showing that the unknown scene parameters are in general all recoverable from a single image of at least two planes. We propose computational methods to detect the isophotes, to reconstruct the scene parameters in closed-form and to refine the results densely using pixel intensity. Our methods thus estimate light source, plane pose and camera pose parameters for untextured planes, which cannot be achieved by the existing approaches. We evaluate our model and methods on synthetic and real images.

1 Introduction

The existing approaches to 3D vision are either classical or learning-based. Classical approaches use explicit visual cues extracted from images. The most successful and widely used visual cues are probably visual motion and shading. Visual motion is typically captured by inter-image correspondences of geometric primitives such as keypoints. Its theoretical model is Multiple-View Geometry (MVG) and its practical implementations are Structure-from-Motion (SfM) [Ullman, 1979] and visual Simultaneous Localisation And Mapping (vSLAM) [Dissanayake et al., 2001]. In terms of usability, SfM and vSLAM make weak assumptions on lighting conditions but strong assumptions on the scene surface, which must be textured to allow the extraction and matching of the geometric primitives. Textureless surfaces do not provide visual motion cues and are hence not handled by SfM and vSLAM. In contrast, shading models are photometric in the sense that they model the pixel colour or intensity. They typically combine models of lighting, surface reflectance and camera. Their practical implementations are Shape-from-Shading (SfS) [Horn, 1975] for a single image and Photometric Stereo (PS) [Woodham, 1979] for multiple images. In terms of usability, SfS makes strong assumptions on lighting and surface reflectance and weak assumptions on surface geometry. Textured surfaces do not directly provide shading because their image mixes the effect of shading and texture; hence they are not handled well by SfS. While PS alleviates some of these assumptions, it still requires the images to be acquired in controlled conditions. Intrinsic Image Decomposition (IID) is a related problem, which originally consists in decomposing an images into an albedo and a shading components [Barrow et al., 1978], also alleviating some of the assumptions required by SfS. Table 1 summarises the characteristics of the classical approaches. In contrast, deep learning-based approaches do not generally model the relationship between the different elements forming the scene, including the surface depth, reflectance, shading and the camera. They thus capture several visual cues and scene priors implicitly from the data, for instance to estimate depth directly [Godard et al., 2017, Yu and Smith, 2019]. However, they were also used to solve IID [Baslamisli et al., 2021]. These methods thus alleviate some of the assumptions made by the classical approaches but do not provide a geometric understanding of the solution and its possible ambiguities.

We propose a new approach to exploit shading via geometric primitives. We choose these as the *shading isophotes* or simply *isophotes*, which we define as the level-set curves in the shading images. These curves have a rich and yet

Approaches	Texture	Primitives
SfM/vSLAM	✓	✓
SfS and PS	✗	✗
Proposed	✗	✓

Table 1: SfM and vSLAM use the surface texture to exploit visual motion via geometric primitives; SfS and PS favour uniform surfaces to exploit shading via photometric models; the proposed approach exploits shading via geometric primitives.

unexplored visual geometry, and concretely, have two major advantages. First, their visual geometry often leads to a well-defined solution and allows one to reason on the solution ambiguities. In contrast, SfS is ill-posed even for the simplest photometric models and does not allow one to understand the ambiguous solution space. Second, closed-form bottom-up solution methods can be found. In contrast, SfS is a nonconvex problem, even for the simplest photometric models, solved by iterative minimisation without guarantee of optimality. We focus on the case of a scene composed of untextured planes with Lambertian reflectance of unknown albedo lit by a single Point Light Source (PLS) of unknown position and intensity. The scene is captured by a camera of known intrinsics and unknown extrinsics providing a single view. We show that the plane poses, the PLS position and the camera extrinsics are all computable from isophotes. This is a strong achievement, as computing untextured plane pose is not possible for any previous method, even in simpler conditions.

Our contributions are four-fold. First, we introduce the notion of virtual geometric primitives, including the proposed isophotes which do not physically exist on the scene surface and hence do not have a fixed physical location. They can nonetheless be efficiently exploited to provide visual 3D information. Second, we propose a comprehensive geometric model for the Lambertian plane case and generic light fall-off. We show that the isophotes are conic sections, from which we construct closed-form solutions for the scene element poses. Third, we show that the isophote detection is highly noise sensitive and propose a stable photometric detection method. Fourth, we propose a complete method for scene reconstruction from isophote detection and initial closed-form solutions to a refined reconstruction based on our model. We study the general case and the special case where the PLS is colocated with the camera, which holds for instance in the endoscopic setting. We evaluate our theory and methods on simulated images, real images taken in controlled conditions with groundtruth and real images taken in uncontrolled conditions.

2 Existing Work and Contributions

2.1 Scope

Image irradiance may be explicitly modelled by involving models for scene elements, namely the light, surface reflectance and geometry, and the camera. This leads to image-based constraints on the scene elements. Inverse rendering is a general formulation using these constraints for 3D reconstruction [Marschner, 1998]. This formulation is generally ill-posed in the absence of strong additional priors. A very common assumption is that the image can be separated in components, which is known as the problem of Intrinsic Image Decomposition (IID) [Barrow et al., 1978, Garces et al., 2022, Ma et al., 2017]. In its simplest version, IID separates the image in albedo and shading. In contrast, SfS uses a shading image in conjunction with lighting priors to recover the surface geometry [Horn, 1975], where the shading image is provided by IID or by assuming constant surface albedo, meaning that the surface must be untextured and uniformly coloured. Advanced versions of IID separate the image in multiple components related to the material properties, scene geometry and lighting. We review the classical, learning-based and primitive-based approaches to SfS, IID and general inverse rendering.

2.2 Classical Approaches

The seminal work on IID is the retinex [Land, 1985], which recovers the albedo and shading by assuming that they have different frequency distributions. More recently, priors on the difference in chromaticity between albedo and shading were used, by assuming colour changes to be dominantly due to albedo variations [Tappen et al., 2002]. Texture analysis was also exploited, to find groups of pixels with constant albedo [Shen et al., 2008]. The shading image produced by these methods can be used as an input to SfS. Most SfS methods solve the image irradiance equation as a Partial Differential Equation (PDE) over the surface normal field. Since the pioneering work [Horn, 1975], which considers a single directional light, a Lambertian reflectance, constant albedo and an orthographic camera, much progress has been made to alleviate the assumptions and handle more general models. Regarding the light, spherical harmonics were shown to improve results under natural illumination [Quéau et al., 2017], while the PLS [Okatani and Deguchi, 1997, Visentini-Scarzanella et al., 2012] and the spot-light [Modrzejewski et al., 2020], chosen close to the camera optical center, were shown to improve results in endoscopy. Regarding surface reflectance, the advanced Oren–Nayar and Phong models were used to exploit specularities [Tozza and Falcone, 2016, Breuß and Ju, 2011]. Regarding surface geometry, a parameterised spline-based model was shown to improve robustness and stability [Courteille et al., 2008]. Regarding the camera, perspective projection was shown to improve performance and also leads to a well-posed formulation [Prados and Faugeras, 2003]. Finally, it was shown that a more general reflectance prior could be used [Barron and Malik, 2014] to perform IID jointly with SfS.

2.3 Learning-based Approaches

Learning-based approaches have been used for IID, taking the irradiance equation into account. A CNN was trained in a supervised manner which decomposes an image into albedo and shading, assuming a simple point-wise product between them re-composes the image [Narihira et al., 2015]. This was improved by introducing a fine-grained shading model, where several reflectance component are decoded separately [Baslamisli et al., 2021]. This allows one to use an image re-composition loss taking the irradiance equation into account, similarly to SfS. Their also exist methods addressing inverse rendering without explicitly performing IID. A multiscale CNN was trained in a supervised manner to simultaneously extract depth, normal and label maps from an image [Eigen and Fergus, 2015]. This method requires a very large training dataset with depth labels. This requirement was alleviated by using self-supervision from left-right stereo consistency [Godard et al., 2017]. Further steps to unsupervised learning consist in incorporating additional rendering priors. Classical rendering methods are generally not differentiable and efforts were recently put into creating differentiable rendering engines. This was used to train a network for silhouette-based mesh reconstruction without supervised 3D data except for camera extrinsics [Kato et al., 2018]. It was shown that a smooth rendered could also be exploited to learn 3D shape directly from an image-to-image loss [Petersen et al., 2019]. Finally, a network was trained which decomposes an image into lighting, normal and albedo maps, by using a differential renderer [Yu and Smith, 2019]. This exploits supervision from MVG and a statistical lighting prior, which does not require specific annotations or controlled conditions.

2.4 Primitive-based Approaches

Primitive-based approaches cast SfS and inverse rendering as geometric problems. They use geometric primitives chosen as special image points or curves, typically the critical points and level-sets. These solutions may be used to initialise an SfS method. It was shown that the surface curvature could be found in closed-form at the critical points in the shading image [Okatani and Deguchi, 2000, Healey and Binford, 1988]. More generally, it was shown that the surface could be constrained from the shading level-sets. Shape-from-Isophotes [Dragnea and Angelopoulou, 2005] estimates the surface normal along level-sets. It works under the assumptions of a directional light aligned with the optical axis, Lambertian reflectance, a smooth surface and an orthographic camera. The level-sets of specular reflection were also used. A simple elliptic model represented by a virtual 3D ellipsoid was used [Morgand et al., 2017] to predict specular highlights from new viewpoints. A more advanced model based on Phong reflectance [Bartoli, 2019] was derived for planar surfaces with a PLS and a perspective camera.

2.5 Contributions

Our method is part of the primitive-based approaches. It involves a complete geometric framework to exploit the shading level-sets as geometric primitives. It establishes their properties and reconstruction methods for a PLS and a piecewise planar surface. SfS methods directly model the pixel colour or intensity and use differential geometry within iterative non-convex solvers. In contrast, we model geometric primitives and use algebraic geometry, leading to closed-form solutions and a detailed geometric understanding of the problem solution sets and properties. The scene element models which our method rely on, especially the PLS and the perspective camera, were not studied in previous primitive-based methods, which use simpler light and camera models. Our method thus works on more realistic grounds. Finally, although existing learning-based methods use models which differ greatly from our method, they show that specific priors were strongly desirable to achieve training, which could be provided by our model. Importantly, both the SfS and the deep-learning approaches fail to provide closed-form solutions. Our method provides closed-form solutions, which have computational advantages and also allow us to develop a complete geometric reasoning about the solution uniqueness and possible ambiguities.

3 A Geometric Treatment of the Rendering Equation

We first introduce our assumptions and models for the light, the surface and the camera. We then show how they lead to the shading isophote model and derive its geometric properties. The mathematical symbols are illustrated in figure 1.

3.1 The Light, Surface and Camera Models

We give our main modelling assumptions. Most of them are very common in SfS. The main differences are that we do not assume the light and camera positions to be known and assume the scene surface to be piecewise planar. We represent a light ray traveling through a given 3D point by a 3D vector, whose direction is the ray's direction and whose norm is the ray's energy. In this model, we only consider a single wavelength and neglect the effect of dispersion.

The scene contains only one light source at position $\mathbf{S} \in \mathbb{R}^3$ and we use a PLS model with unknown intensity $\theta_0 > 0$. The surface irradiance, *i.e.*, the amount of light received, depends on a strictly decreasing positive function g of the distance between the surface point and the PLS. Function g models the phenomenon of light fall-off and is typically chosen as the inverse square distance. The incident light ray from the source to point $\mathbf{X} \in \mathbb{R}^3$ is thus:

$$\mathcal{L}_i(\mathbf{X}, \mathbf{S}) = \theta_0 \, g(\|\mathbf{S} - \mathbf{X}\|) \frac{\mathbf{S} - \mathbf{X}}{\|\mathbf{S} - \mathbf{X}\|}, \tag{1}$$

where $\mathbf{X} \neq \mathbf{S}$, *i.e.*, the surface point is not at the PLS. We neglect inter-reflections, meaning that any incident light ray hitting the surface comes from the primary light source. We use Lambertian reflectance with unknown constant albedo $k > 0$. We thus have the reflected light ray passing by point $\mathbf{O} \in \mathbb{R}^3$ and coming from surface point $\mathbf{X}$ as:

$$\mathcal{L}_r(\mathbf{O}, \mathbf{X}) = k \, \mathbf{N}^\top \mathcal{L}_i(\mathbf{X}, \mathbf{S}) \frac{\mathbf{O} - \mathbf{X}}{\|\mathbf{O} - \mathbf{X}\|}, \tag{2}$$

where $\mathbf{N} \in \mathbb{R}^3$ is the surface normal. We decompose camera projection in two parts: a geometric projection $\mathcal{G} : \mathbb{R}^3 \to \mathbb{R}^2$, which is a standard perspective projection with known intrinsic matrix $\mathsf{K} \in \mathbb{R}^{3\times3}$, and a radiometric Camera Response Function (CRF) C, which is a strictly increasing, hence bijective, scalar function. Our base image model is thus:

$$\mathcal{I}(\mathcal{G}(\mathbf{X})) = C\left(\|\mathcal{L}_r(\mathbf{O}, \mathbf{X})\|\right), \tag{3}$$

where $\mathcal{I} : \mathbb{R}^2 \to \mathbb{R}$ denotes the image function, defined as the continuous extension of the observed discrete image function, giving intensity from pixel coordinates. Combining equations (1), (2) and (3) we arrive at the image equation, similar to what SfS with the PLS model would use:

$$\mathcal{I}(\mathcal{G}(\mathbf{X})) = C\left(k \, \theta_0 \, g(\|\mathbf{S} - \mathbf{X}\|) \frac{|\mathbf{N}^\top (\mathbf{S} - \mathbf{X})|}{\|\mathbf{S} - \mathbf{X}\|}\right). \tag{4}$$

We constrain the scene surface to be composed of $n \geq 1$ planes and assume the image to be segmented in coplanar pixel sets $\mathcal{U}_i \subset \mathbb{N}^2$, with $i \in [1, n]$. Plane i is represented by its normal $\mathbf{N}_i$ and distance to camera d_i. These parameters are unknown. We omit the plane index i whenever a single plane is considered. Therefore, for a point $\mathbf{X}$ on plane i, we have $\mathbf{N}_i^\top \mathbf{X} - d_i = 0$, and when a single plane is considered, the notation reduces to $\mathbf{N}^\top \mathbf{X} - d = 0$. In addition, because $\|\mathbf{N}_i\| = 1$, we have that $\mathbf{N}_i^\top \mathbf{X} - d_i$ is the signed distance between point $\mathbf{X}$ and plane i.

3.2 The Shading Isophote

We define the shading isophote as a level-set, which is the set of points $\mathbf{u} \in \mathbb{R}^2$ where the image has the explicit intensity value $\rho > 0$:

$$\mathcal{I}(\mathbf{u}) = \rho. \tag{5}$$

We establish the relationship between this definition and the image model (4) by identifying their left and right hand sides, from which we obtain the following two equations:

$$\rho = C\left(k \, \theta_0 \, g(\|\mathbf{S} - \mathbf{X}\|) \frac{|\mathbf{N}^\top (\mathbf{S} - \mathbf{X})|}{\|\mathbf{S} - \mathbf{X}\|}\right) \tag{6}$$

$$\mathbf{u} = \mathcal{G}(\mathbf{X}) \text{ with } \mathcal{I}(\mathbf{u}) = \rho. \tag{7}$$

One can easily verify that recombining equations (6) and (7) by eliminating ρ and $\mathbf{u}$ gives equation (4). Equation (6) describes an implicit 3D surface, which, intersected with the scene surface, gives an implicit 3D curve. The solution of this equation is the isophote pre-image, which is the back-projection of the image isophote to the scene plane. Equation (7) represents the projection of the isophote pre-image to the image.

This breaks down the isophote for the image model into two parts, forming a key to understand the geometry in two steps. We first study the properties of the isophote pre-image from equation (6) and how its parameters are related to the geometric parameters of the scene. We then study how its reprojection from equation (7) leads to constraints exerted by its image on the geometric parameters of the scene.

3.3 The Shading isophote Pre-images are Circles

Our first result reveals the circular nature of the isophote pre-image, defined by the implicit curve emerging from equation (6) and the scene plane.

Lemma 1. *The pre-image of a shading isophote is the intersection of a sphere centred at the PLS with the scene plane.*

Proof. We work in spherical coordinates centred on $\mathbf{S}$, with the x-axis following the scene plane normal $\mathbf{N}$. We thus search for $\mathbf{X} = s[\cos\theta \ \ \sin\theta\cos\phi \ \ \sin\theta\sin\phi]^\top$. By substituting in equation (6) we have:

$$\rho = C\left(k\,\theta_0\,g(s)\frac{|s\cos\theta|}{s}\right). \tag{8}$$

We rewrite the equation, using the bijectivity of function C and positivity of function g, as:

$$\frac{C^{-1}(\rho)\,s}{g(s)} = k\,\theta_0\,|s\cos\theta|. \tag{9}$$

Recall that the isophote pre-image is the intersection of the above implicit surface with the scene plane $\mathbf{N}^\top\mathbf{X} - d = 0$, where d is the camera centre to scene plane distance, or $s\cos\theta = h$ in spherical coordinates. In this equation, h is the signed distance between the PLS $\mathbf{S}$ and the scene plane, which, using the scene plane equation, is directly given by:

$$h = \mathbf{N}^\top\mathbf{S} - d. \tag{10}$$

Substituting in equation (9) and rearranging, we have:

$$\frac{s}{g(s)} = \frac{k\,\theta_0\,|h|}{C^{-1}(\rho)}. \tag{11}$$

A point is on the isophote pre-image only if s is a solution of equation (11). Because g is a strictly decreasing positive function, the left-hand side is a function $t(s) = \frac{s}{g(s)}$ which is strictly increasing on $s > 0$, hence bijective, and the right-hand side is independent of the point. The equation thus has a single solution $\hat{s}$ with:

$$\hat{s} = t^{-1}\left(\frac{k\,\theta_0\,|h|}{C^{-1}(\rho)}\right). \tag{12}$$

The locus of points solving equation (8) is thus a sphere of radius $\hat{s}$ centred at the PLS. □

Lemma 1 allows us to establish our main result in the following proposition, which refines our understanding of the isophote pre-image and relates them to the scene plane's Brightest Point (BP). We introduce the BP as the surface point of maximal irradiance. Each scene plane has its own BP, which may or may not be visible in the image. Proposition 1 is illustrated by figure 1.

Proposition 1. *The pre-image of shading isophotes are concentric circles centred at the BP. The radius of these circles increases as the isophote intensity ρ decreases. The BP is the orthogonal projection of the PLS on the scene plane. The coordinates of the BP are given by:*

$$\bar{\mathbf{X}} = \mathbf{S} + h\mathbf{N}. \tag{13}$$

Proof. We first recall that $\bar{\mathbf{X}}$ from equation (13) is the orthogonal projection of the PLS $\mathbf{S}$ on the scene plane. We then establish the nature of the isophote pre-image. From lemma 1, the pre-image of the shading isophote is the intersection of a sphere of radius $\hat{s}$ centred on the PLS with the scene plane lying at a distance $|h|$ of the PLS. This intersection may thus trivially be:

- Empty for $|h| > \hat{s}$
- A single point for $|h| = \hat{s}$, in which case the sphere is tangent to the plane and the intersection point is $\bar{\mathbf{X}}$

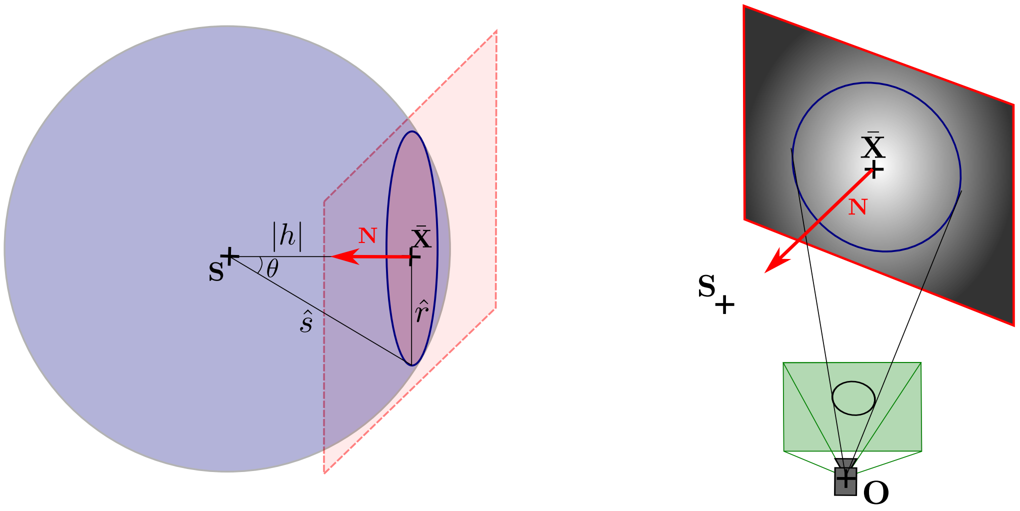

Figure 1: (left) The isophote pre-images (dark blue) are the intersection of spheres (blue) centred on the PLS $\mathbf{S}$ with the scene plane (red). They form concentric circles centred at $\bar{\mathbf{X}}$, the orthogonal projection of $\mathbf{S}$ on the scene plane, which is also the BP. (right) Their image by a perspective camera (green) are conics sections.

- A circle for $|h| < \hat{s}$, with radius $\hat{r}$ and centre $\bar{\mathbf{X}}$, with:

$$\hat{r} = \hat{s}\sqrt{1 - \frac{h^2}{\hat{s}^2}}. \tag{14}$$

Equation (14) stems from the Pythagorean theorem applied to a triangle containing $\mathbf{S}$, $\bar{\mathbf{X}}$ and any point of the circle. Figure 1 illustrates the intersection in the case $|h| < \hat{s}$.

Finally, we show that the BP lies at $\bar{\mathbf{X}}$ and that the circle radius decreases with the intensity ρ. We have from equation (12) that the sphere radius $\hat{s}$ depends on the isophote intensity ρ, which we formalise by defining an explicit function f giving $\hat{s}$ from ρ:

$$\hat{s} = f(\rho) = t^{-1}\left(\frac{k\,\theta_0\,|h|}{C^{-1}(\rho)}\right). \tag{15}$$

We have that f is strictly decreasing and positive, hence bijective. Geometrically, we can interpret it as a sphere of increasing radius for a decreasing intensity value ρ. Studying visible isophotes implies that the sphere-plane intersection must be non-empty and thus $|h| \leq \hat{s}$. By applying the reciprocal function f^{-1}, which is also positive and strictly decreasing to this condition, we obtain $f^{-1}(|h|) \geq \rho$. This implies the existence of a maximum observed intensity $\rho_{\max} = f^{-1}(|h|)$. For $\rho = \rho_{\max}$, the sphere is tangent to the plane and we have a single point of maximum intensity, which must be the BP. For $\rho < \rho_{\max}$, the sphere intersects the plane in a circle, whose radius increases as ρ decreases, following equation (14). □

The fact that the isophote pre-image is a circle is intuitive for the Lambertian reflectance and a PLS. It is however more surprising that the result still holds for a generic light fall-off, as modelled by function g. Our next result, given as a corollary of proposition 1, infers the nature of the shading isophotes via the perspective projection of their pre-image, which emerges from equation (7).

Corollary 1. *The shading isophotes are conics. The conic type is determined by the nature of the intersection between the isophote pre-image circle with the camera's principal plane as:*

- *An ellipse if the pre-image circle does not intersect the principal plane;*
- *A hyperbola if the pre-image circle intersects the principal plane in two real points;*
- *A parabola if the pre-image circle intersects the principal plane in a single real point.*

Proof. From proposition 1, the isophote pre-image is a circle. We have that the camera centre is not on the scene plane, as otherwise the scene plane would project to a single line the isophotes would not be observable. Thus the perspective projection of the circular pre-image is a full-rank conic of projective signature $(2, 1)$, which can be an ellipse, a hyperbola or a parabola [Hartley and Zisserman, 2000]. The nature of a full-rank conic depends on the number of

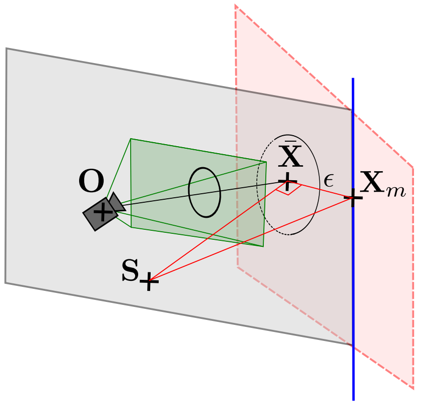

Figure 2: The principal plane (gray) and the scene plane (red) intersect in the line at infinity of the image plane (blue). For a visible BP, one can find circles centred at the BP, with radius lower than ϵ, which lie in front of the camera. These circles project to ellipses in the image plane (green).

intersections of the pre-image circle with the pre-image of the line at infinity of the image plane [Aleksandrov et al., 1999, *p*248], which is the intersection of the principal plane, *i.e.*, the plane of equation $z = 0$ in camera coordinates, with the scene plane. □

We give a final refinement on the nature of the shading isophotes in the image. This refinement, given by the next proposition, depends on the visibility of the Image of the Brightest Point (IBP).

Proposition 2. *If the BP lies in front of the camera, which occurs if the IBP is visible, the shading isophotes in its neighbourhood are nested ellipses. Conversely, if a shading isophote is elliptic, it encloses shading isophotes consisting of nested ellipses and the IBP.*

Proof. For the first part of the proposition, we have that the BP with coordinates $\bar{\mathbf{X}}$ lies in front of the principal plane $z = 0$. We distinguish two cases for the pose of the scene plane relative to the principal plane. The first case is for a scene plane parallel to the principal plane. In this case the circles that the scene plane supports do not intersect the principal plane. Hence, following corollary 1, the shading isophotes they induce are nested ellipses. The second case is for a scene plane skew to the principal plane. In this case, these planes intersect in a line, as shown in figure 2. We define point $\mathbf{X}_m$ as the closest point on this intersection line to the BP. The BP does not lie on the principal plane, as otherwise it would not be visible. We can thus define $\epsilon > 0$ so that $\|\mathbf{X}_m - \bar{\mathbf{X}}\| = \epsilon$. The circles centred at the BP with radii lower than ϵ do not interest the principal plane. Consequently, following corollary 1, the shading isophote they induce are ellipses. As the projective transformation preserves the intersections and the concentric circles do not self-intersect, their images do not intersect as well. Hence, the neighbourhood of the IPB are nested ellipses.

For the second part of the proposition, corollary 1 implies that the pre-image of an elliptic isophote is a circle that does not intersect the line at infinity. Therefore, any circle of small radius and the BP also does not intersect the line at infinity and lies in front of the camera. Their projections thus lie inside the isophote ellipse. □

4 Algebraic Scene Constraints from Shading isophotes

We study the algebraic constraints emerging on the scene parameters from the observation of one or several isophotes.

4.1 Scene Parameters

We first parameterise the light source and scene surface, then the isophotes. Without loss of generality, we align the world coordinates with the camera coordinates. We thus do not have explicit camera parameters and express all 3D entities in camera coordinates. For the light source and scene surface, following the model of section 3.1, we use the PLS position $\mathbf{S} \in \mathbb{R}^3$ and, for each scene plane, the signed distance h from the PLS to the plane, the plane normal $\mathbf{N} \in \mathbb{R}^3$ and the signed distance d from the camera centre to the plane. These parameters are redundant, as equation (10) must hold for each plane, which we systematically take into account into our derivations. For each isophote, we have from proposition 1 that the pre-image is a circle on the corresponding scene plane, which we explicitly parameterise by its centre, specifically the BP $\bar{\mathbf{X}}$, its support plane's normal $\mathbf{N}$ and its radius $\hat{r}$. The BP and plane normal are already part of the light source and scene surface parameters; we thus only add the radius parameter $\hat{r}$. The radius bundles many parameters of the scene model. Specifically, as can be seen from equation (14), it depends on the signed distance h and, via $\hat{s} = f(\rho)$, bundles all the other photometric parameters, namely k, C, ρ, g and θ_0. These parameters would require additional assumptions on the camera response or attenuation functions to be recoverable. They do not contribute geometric constraints and we can thus ignore their explicit connection to the radius. There are thus three groups of unknowns: global unknowns, plane-dependent unknowns and plane- and isophote-dependent unknowns. The next section derives the algebraic constraints.

4.2 Scene Constraints

Equation (7) describes the projection of an isophote pre-image to the image plane, where it is observed. Considering a calibrated camera, the projection function $\mathcal{G}$ it involves represents perspective projection, with a known intrinsic parameter matrix K. Without loss of generality, we use matrix K^{-1} to normalise the image coordinates. We have from corollary 1 that the isophote is a non-degenerate conic, whose matrix representation is a symmetric matrix $\mathsf{E} \in \mathbb{S}_3$. This conic is obtained algebraically by projecting the pre-image circle, which, following [Hartley and Zisserman, 2000, p37], is obtained as:

$$\mathsf{E} \sim \mathsf{H}^\top \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\hat{r}^2 \end{bmatrix} \mathsf{H} \qquad (16)$$

$$\mathsf{H} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{S} + h\mathbf{N} \end{bmatrix}^{-1}, \qquad (17)$$

with $\mathbf{R}_1$ and $\mathbf{R}_2$ two orthonormal vectors orthogonal to $\mathbf{N}$, and where $\sim$ represents equality up to scale. We thus have a total of 15 unknown parameters, broken down in the three above-described groups as follows. The first group are the global variables and includes the 3 parameters for the PLS in $\mathbf{S}$. The second group are the plane parameters and includes the signed distances h and d, the normal $\mathbf{N}$ and the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$. The number of unknowns from this second group scales with the number of scene planes. The third group are the isophote parameters and includes the radius $\hat{r}$. The number of unknowns from this third group scales with the number of isophotes.

In order to keep the notation uncluttered, we formulate a complete constraint system for a single plane and a single isophote. We expand the constraints available from the projection equations (16) and (17) with the orthonormality constraints, add an in-plane orientation constraint and group them with equation (10). This complete system is made of 13 polynomial constraints grouped in two sub-systems S_A and S_B for reasons to be shortly clarified:

$$S_A \begin{cases} d - \mathbf{N}^\top(\mathbf{S} + h\mathbf{N}) = 0 & (18.1) \\ \mathbf{R}_1^\top \mathsf{E} \mathbf{R}_1 + \hat{r}^2 (\mathbf{S} + h\mathbf{N})^\top \mathsf{E} (\mathbf{S} + h\mathbf{N}) = 0 & (18.2) \\ \mathbf{R}_1^\top \mathsf{E} (\mathbf{S} + h\mathbf{N}) = 0 & (18.3) \\ \mathbf{R}_2^\top \mathsf{E} (\mathbf{S} + h\mathbf{N}) = 0 & (18.4) \end{cases}$$

$$S_B \begin{cases} \mathbf{R}_1^\top \mathsf{E} \mathbf{R}_1 = \mathbf{R}_2^\top \mathsf{E} \mathbf{R}_2 & (18.5) \\ \mathbf{R}_1^\top \mathsf{E} \mathbf{R}_2 = 0 & (18.6) \\ \mathbf{R}_1^\top \mathbf{R}_1 = \mathbf{R}_2^\top \mathbf{R}_2 & (18.7) \\ \mathbf{R}_1^\top \mathbf{R}_2 = 0 & (18.8) \\ \mathbf{R}_1^\top \mathbf{R}_1 = \mathbf{N}^\top \mathbf{N} & (18.9) \\ \mathbf{R}_1^\top \mathbf{N} = 0 & (18.10) \\ \mathbf{R}_2^\top \mathbf{N} = 0 & (18.11) \\ \mathbf{N}^\top \mathbf{N} = 1 & (18.12) \\ R_{1,2} R_{2,2} = 0. & (18.13) \end{cases}$$

The first constraint, equation (18.1), is a simple rewriting of equation (10) with the normal vector $\mathbf{N}$. The next group of 5 constraints, equations (18.2) to (18.6), come from equation (16), after elimination of the unknown equality scale. The next group of 6 constraints, equations (18.7) to (18.12), represent the orthonormality of the matrix modelling the rotation between the scene plane and the camera coordinates. Finally, the last constraint, equation (18.13), arbitrarily fixes the in-plane orientation.

The two sub-systems S_A and S_B are defined by noticing that equations (18.5) to (18.13) only contain the unknowns $\mathbf{R}_1, \mathbf{R}_2$ and $\mathbf{N}$ and only depend on the known matrix E representing the observed isophote. We thus group them to define S_B, while we group the remaining equations (18.1) to (18.4) to define S_A, with $\mathbf{S}$, h, d and $\hat{r}$ as unknowns and dependencies on E, $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{N}$. This split simplifies the system study; we show in the next section that the sub-system S_B taken independently only admits a finite number of solutions. This allows us to study and solve the two sub-systems almost separately.

The next step is to study the dimension of the solution space more specifically and to understand under which conditions can it be brought to a well-constrained state to create a minimal problem.

4.3 The Minimal Problem for a General Configuration

We start by studying the system from its two sub-systems S_A and S_B, investigating the dimension of the solution space to each sub-system. We then show that the general problem is under-constrained for a single scene plane, even with multiple isophotes. Finally, we show that the general problem has a minimal configuration for two scene planes.

Lemma 2. *For a general configuration, the solution space of S_B is zero-dimensional (it has a finite set of solutions for $\mathbf{N}$, $\mathbf{R}_1$ and $\mathbf{R}_2$), whereas, given $\mathbf{N}$, $\mathbf{R}_1$ and $\mathbf{R}_2$, the solution space of S_A is two-dimensional (it has a two-way ambiguous solution sub-space for $\mathbf{S}$, h, d and $\hat{r}$).*

Proof. First, we study the sub-system S_B, constraining $\mathbf{N}$, $\mathbf{R}_1$ and $\mathbf{R}_2$, which can be assembled in the orthonormal matrix $\mathsf{R} = [\mathbf{R}_1\,\mathbf{R}_2\,\mathbf{N}]$. The rotation represented by R rectifies the observed isophote and the image of the absolute conic into circles, which is a metric rectification of the image plane [Hartley and Zisserman, 2000]. This problem has two ambiguous solutions [Gurdjos et al., 2006], given the image of two circles. In our case, one circle is the isophote and the other one is the absolute conic, as one easily verifies that, given equations (18.5) to (18.8), the circular points [Hartley and Zisserman, 2000, *p*52] belong to the rectified conics, respectively $\mathsf{R}^\top\mathsf{E}\mathsf{R}$ and $\mathsf{R}^\top\mathsf{R}$. The remaining equations determine $\mathbf{N}$ and fix the in-plane rotation. Second, we study the sub-system S_A, given $\mathbf{N}$, $\mathbf{R}_1$ and $\mathbf{R}_2$. Equations (18.1), (18.3) and (18.4) are linear in the unknowns d, $\mathbf{S}$ and h, which are thus constrained to lie in a linear sub-space of dimension $5 - 3 = 2$. Finally, equation (18.2) adds one constraint but introduces the unknown $\hat{r}$ and so does not change the dimension of the solution sub-space. □

Proposition 3. *From the observation of one isophote and without additional priors on the scene parameters, the general problem represented by system (18) is under-determined and its solution space has dimension 2.*

Proof. Using the two-way split of system (18), the global system's indeterminacy arises from the indeterminacy of S_A, which is of dimension 2 from lemma 2. □

Almost all 3D reconstruction settings, including SfM and SfS, do not allow one to recover an absolute scene scale. This is a fundamental limitation of image-based 3D vision with the perspective camera. The proposed setting does not escape this limitation and undergoes a scale ambiguity. Concretely, this scale ambiguity occurs in the unknown position parameters, including the PLS $\mathbf{S}$, the distances d and h, and the radius $\hat{r}$. Algebraically, these parameters are involved in quasi-homogeneous equations only, as shown in the following lemma.

Lemma 3. *Sub-system S_A is quasi-homogeneous of type* $(1, 1, 1, -1)$ *in the unknowns d, $\mathbf{S}$, h and $\hat{r}$.*

Proof. Equations (18.1), (18.3) and (18.4) are directly homogeneous. Equation (18.2) is quasi-homogeneous of type -1 in $\hat{r}$. This can be verified by multiplying all the unknowns but $\hat{r}$ by $\lambda > 0$ and $\hat{r}$ by $\frac{1}{\lambda}$, which leaves all 4 equations of S_A invariant. □

This result implies that regardless of the number of isophotes and planes being observed, only fixing one of the quasi-homogeneous unknowns can fix the scale of the 3D reconstruction. Therefore, in the absence of such an extra constraint, a problem setting can be considered minimal if the solution space for the quasi-homogeneous unknowns has dimension 1, and given an extra scale-fixing constraint, if it holds a finite number of solutions.

The next proposition gives a more general characterisation of the extra constraints brought by an additional isophote arising from the same scene plane as the initial one.

Proposition 4. *Observing an extra isophote on a plane for which an isophote is already observed does not reduce the dimension of the solution space.*

Proof. Two isophotes give two pairs of sub-systems (S_A, S_B) and (S'_A, S'_B). We identify the shared and specific unknowns, observations and constraints, to assemble a joint system (S_A, S_B, S'_A, S'_B). Because the isophotes share their support plane, the systems share the unknowns $\mathbf{N}$, $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{S}$, h and d. The observed conics are of course different, which we respectively write E and E', as well as the two unknown radii which we write $\hat{r}$ and $\hat{r}'$ respectively.

First, we consider the joint sub-system (S_B, S'_B) which, following lemma 2, gives a finite solution space for $\mathbf{N}$, $\mathbf{R}_1$ and $\mathbf{R}_2$. The solution is given by the intersection of the two solution spaces for S_B and S'_B. From the image of two concentric circles, the vanishing line, and so, because the camera is calibrated, the normal of the plane, can be recovered without ambiguity [Kim et al., 2005]. Consequently, the intersection of the two solution spaces for S_B and S'_B gives a single solution.

Second, we consider the joint sub-system (S_A, S'_A). This system is the union of two quasi-homogeneous systems of the same type, as per lemma 3, and is thus also quasi-homogeneous of that type. It means that the solution space is of dimension larger or equal to 1. A closer examination is required to show why the additional sub-system does not lower the dimension of the solution space. The expression $\mathbf{S} + h\mathbf{N}$ is shared by the two sub-systems. We substitute it by the BP $\bar{\mathbf{X}}$, following equation (13). We arrive at the following joint sub-system, with 10 constraints and 10 unknowns:

$$S_A \left\{ \begin{array}{rr} d - \mathbf{N}^\top \bar{\mathbf{X}} = 0 & \qquad (19.1) \\ \mathbf{R}_1^\top \mathsf{E} \mathbf{R}_1 + \hat{r}^2 \bar{\mathbf{X}}^\top \mathsf{E} \bar{\mathbf{X}} = 0 & \qquad (19.2) \\ \mathbf{R}_1^\top \mathsf{E} \bar{\mathbf{X}} = 0 & \qquad (19.3) \\ \mathbf{R}_2^\top \mathsf{E} \bar{\mathbf{X}} = 0 & \qquad (19.4) \end{array} \right.$$

$$S'_A \left\{ \begin{array}{rr} d - \mathbf{N}^\top \bar{\mathbf{X}} = 0 & \qquad (19.5) \\ \mathbf{R}_1^\top \mathsf{E}' \mathbf{R}_1 + (\hat{r}')^2 \bar{\mathbf{X}}^\top \mathsf{E}' \bar{\mathbf{X}} = 0 & \qquad (19.6) \\ \mathbf{R}_1^\top \mathsf{E}' \bar{\mathbf{X}} = 0 & \qquad (19.7) \\ \mathbf{R}_2^\top \mathsf{E}' \bar{\mathbf{X}} = 0 & \qquad (19.8) \end{array} \right.$$

$$\mathbf{S} + h\mathbf{N} = \bar{\mathbf{X}}. \qquad (19.9)$$

We show that the equations in S_A and S'_A are not independent, hence the unknowns remain under-constrained. Equations (19.1) and (19.5) are simply identical. Equations (19.3) and (19.4) are linear in $\bar{\mathbf{X}}$, constraining it to lie on a 3D line. As the system is quasi-homogeneous, this solution space is irreducible. Hence, equations (19.7) and (19.8) do not bring further constraints. Equation (19.6) adds a constraint but depends on the second unknown radius and so does not change the dimension of the solution space. Finally, equation (19.9) is the same for any isophote on the same scene plane. It adds 3 constraints but involves 4 unknowns in $\mathbf{S}$ and h, adding an extra dimension to the solution space. The joint sub-system (S_A, S'_A) for two isophotes is thus as under-determined in terms of the dimension of its solution space as the original sub-system S_A for one isophote. The reasoning is easily generalised to any number of isophotes on the same scene plane. Hence, extra coplanar isophotes may only reduce the number of discrete ambiguities in the solution space but not its dimension. □

Following proposition 4, the only possible ways to further constrain the problem are to introduce additional priors on the scene elements, which is discussed in the next section, or to increase the number of scene planes, which is considered directly below. Specifically, the next proposition considers the case of two scene planes, which each provide an observed isophote. We have that only the position $\mathbf{S}$ of the PLS is shared by the sub-systems given by each isophote. The unknown set is thus formed by the union of the unknowns of each individual system except $\mathbf{S}$. One can easily verify that the joint system remains quasi-homogeneous of the same type as in lemma 3.

Proposition 5. *For a general configuration, two isophotes on different scene planes lead to a minimal problem.*

Proof. We have from proposition 3 that the problem is under-constrained for a single isophote. We must thus consider the joint system (S_A, S_B, S'_A, S'_B) obtained from two isophotes. We have from lemma 2 that S_B and S'_B can be solved separately to resolve the orientation of the two scene planes. This leaves us with the joint sub-system (S_A, S'_A):

$$S_A \left\{ \begin{array}{rr} d - \mathbf{N}^\top (\mathbf{S} + h\mathbf{N}) = 0 & \qquad (20.1) \\ \mathbf{R}_1^\top \mathsf{E} \mathbf{R}_1 + \hat{r}^2 (\mathbf{S} + h\mathbf{N})^\top \mathsf{E} (\mathbf{S} + h\mathbf{N}) = 0 & \qquad (20.2) \\ \mathbf{R}_1^\top \mathsf{E} (\mathbf{S} + h\mathbf{N}) = 0 & \qquad (20.3) \\ \mathbf{R}_2^\top \mathsf{E} (\mathbf{S} + h\mathbf{N}) = 0 & \qquad (20.4) \end{array} \right.$$

$$S'_A \left\{ \begin{array}{rr} d' - {\mathbf{N}'}^\top (\mathbf{S} + h'\mathbf{N}') = 0 & \qquad (20.5) \\ {\mathbf{R}'_1}^\top \mathsf{E}' \mathbf{R}'_1 + (\hat{r}')^2 (\mathbf{S} + h'\mathbf{N}')^\top \mathsf{E}' (\mathbf{S} + h'\mathbf{N}') = 0 & \qquad (20.6) \\ {\mathbf{R}'_1}^\top \mathsf{E}' (\mathbf{S} + h'\mathbf{N}') = 0 & \qquad (20.7) \\ {\mathbf{R}'_2}^\top \mathsf{E}' (\mathbf{S} + h'\mathbf{N}') = 0. & \qquad (20.8) \end{array} \right.$$

Equations (20.3), (20.4), (20.7) and (20.8) form a homogeneous linear system of 4 equations for 5 homogeneous unknowns, $\mathbf{S}$, h and h'. The solution space is thus reduced to its minimal dimension of 1. The remaining equations each adds a new constraint and a new unknown, namely d, $\hat{r}$, d' and $\hat{r}'$, and can thus be solved separately. □

Configuration		PLS to Plane	Plane		PLS
		Distance (1 DoF)	Orientation	Distance	Position (3 DoF)
			(3 DoF)	(1 DoF)	
A	1-plane	known	known	known	known
B	1-plane	known	known	known	✓
C	1-plane	known	✓	✓	known
D	1-plane	known	✓	✗ (dim 1)	✗ (dim 1)
	2-plane	known	✓	✓	✓
E	1-plane	✓	known	known	known
F	1-plane	✗ (dim 1)	known	known	✗ (dim 1)
	2-plane	✓	known	known	✓
G	1-plane	✓	✓	✓	known
G^*	1-plane	✗ (dim 1)	✓	✗ (dim 1)	known
H	1-plane	✗ (dim 1)	✓	✗ (dim 1)	✗ (dim 2)
	2-plane	✗ (dim 1)	✓	✗ (dim 1)	✗ (dim 1)

Table 2: Summary of the solution ambiguities for various configurations defined from priors considered on the scene elements and numbers of scene plane. In the cases involving 2 scene planes, the priors are considered for each scene plane individually. For each configuration and scene parameter, 'known' means that the parameter is given as a prior, a check mark indicates that the solution space is finite and a cross mark indicates otherwise, followed by a number indicating the dimension of the solution space.

Our results might first appear counter-intuitive, as adding an isophote on an already modelled plane introduces new observations with a limited number of new unknowns, while adding an isophote on a new, yet unmodelled plane, introduces a larger number of unknowns. Yet, our results show that only by considering a new plane can one increase the number of redundant constraints. This is a strong result; it means that, up to a two-dimensional ambiguity, the observation of one isophote already contains all the geometric constraints that one has from one plane, while adding a second plane is enough to triangulate the position of the shared PLS.

4.4 Determinacy and Consistency for each Scene Parameters

We leave the general setup considered thus far by introducing priors on the scene elements.

4.4.1 Priors and Scene Configurations

We study the effect of the following four priors: knowledge of the PLS to plane distance h, knowledge of the plane orientation $\mathbf{N}$, knowledge of the plane to camera distance d and knowledge of the PLS position $\mathbf{S}$. We do not use priors on $\hat{r}$, the so-called radius parameter, as it depends on the radiometric imaging model, including the scene reflectance and camera response, and not on the scene geometry. We thus do not have radiometry priors beyond the general assumptions used to derive our model. Proposition 5 allows us to study the problem configurations for one or two isophotes on different scene planes only, as a larger set of isophotes on different scene planes would necessarily over-constrain the system and be equivalent to the two isophote cases. Table 2 summarises the possible configurations and our results.

The solution space for the quasi-homogeneous parameters, shown in columns 1, 3 and 4 of table 2, cannot be reduced beyond 1, regardless of the number of isophotes and planes, owing to the global scene scale ambiguity. However, some configurations such as G include priors on the quasi-homogeneous parameters, which resolve the global scale ambiguity. The radius $\hat{r}$ is considered an unknown in every configuration. It may be found from equation (18.2) if $\mathbf{S}$, $\mathbf{N}$, $\mathbf{R}_1$ and h are resolved, which specifically occurs in configurations A, B, C, D (2 planes), E, F (2 planes) and G. The next four sections study the solvability and solution space for each configuration from table 2 specifically.

4.4.2 Single-plane Over-constrained Configurations

We study configurations A, B, C, E and G, which involve a single-plane, and show that they are over-constrained.

Configuration A. All the geometric parameters are known from the priors. The system is strongly over-constrained. This configuration does not have a direct practical application. We use it experimentally in section 8.1 to study our model's empirical validity and expressiveness on real data.

Configuration B. There is a single unknown in the geometric parameters, namely the position $\mathbf{S}$ of the PLS. The global system is over-constrained, as the 4 unknown scalars in $\mathbf{S}$ and $\hat{r}$ appear in all 4 equations of S_A. These 4 equations are inhomogeneous because h and $\mathbf{N}$ are known from the priors, thus the global scene scale can be determined.

Configuration C. There are two unknowns in the geometric parameters, namely the orientation $\mathbf{N}$ of the plane and its distance d to the camera. Sub-system S_B determines $\mathbf{N}$, as per lemma 2, with a finite number of solutions. Using a solution for $\mathbf{N}$, equation (18.1) from S_A becomes linear and inhomogeneous, thus constrains d and fixes the global scene scale.

Configuration E. There is a single unknown in the geometric parameters, namely the distance h between the PLS and the scene plane. The global system is thus strongly over-constrained. The linear and inhomogeneous equations (18.1), (18.3) and (18.4) from S_A constrain h and fix the global scene scale.

Configuration G. There are three unknowns in the geometric parameters, namely the orientation $\mathbf{N}$ of the scene plane, its distance d to the camera and its distance h to the PLS. Sub-system S_B determines $\mathbf{N}$, as per lemma 2, with a finite number of solutions. Using a solution for $\mathbf{N}$, the remaining equations (18.1), (18.3) and (18.4) from S_A become linear, constrain d and h, and fix the global scene scale.

4.4.3 Single-Plane Under-constrained Configurations

We study configurations D, F and H specifically for a single plane and show that they are under-constrained.

Configuration D, 1-plane. There are three unknowns in the geometric parameters, namely the orientation $\mathbf{N}$ of the scene plane, its distance d to the camera and the position $\mathbf{S}$ of the PLS. This yields a total of 14 scalar unknowns, namely d, $\mathbf{S}$, $\hat{r}$, $\mathbf{N}$, $\mathbf{R}_1$ and $\mathbf{R}_2$, in a system of 13 equations, which is thus under-constrained. Sub-system S_B determines $\mathbf{N}$, as per lemma 2, with a finite number of solutions. Singling out equation (18.2) from S_A and the unknown $\hat{r}$, we obtain a system of 3 linear and inhomogeneous equations with 4 scalar unknowns in d and $\mathbf{S}$. The PLS position $\mathbf{S}$ is thus restrained to a line, h is unrecoverable, as well as the global scene scale.

Configuration F, 1-plane. There are two unknowns in the geometric parameters, namely the distance h of the scene plane to the PLS and the PLS position $\mathbf{S}$. This yields a total of 11 scalar unknowns in a system of 13 equations, which is thus over-constrained. Sub-system S_B does not contain unknowns, therefore the orientation of the scene plane is over-constrained. Singling out equation (18.2) from S_A and the unknown $\hat{r}$, we obtain a system of 3 linear and inhomogeneous equations with 4 scalar unknowns in h and $\mathbf{S}$. The PLS position $\mathbf{S}$ is thus restrained to a line, h is unrecoverable, as well as the global scene scale.

Configuration H, 1-plane. There are four unknowns in the geometric parameters, namely the distance h of the scene plane to the PLS, the orientation $\mathbf{N}$ of the scene plane, its distance d to the camera and the PLS position $\mathbf{S}$. This yields a total of 11 scalar unknowns in a system of 13 equations, which is thus over-constrained. Sub-system S_B determines $\mathbf{N}$, as per lemma 2, with a finite number of solutions. Singling out equation (18.2) from S_A and the unknown $\hat{r}$, we obtain a system of 3 linear and homogeneous equations with 5 scalar unknowns in h, d and $\mathbf{S}$. The PLS position $\mathbf{S}$ is thus restrained to a plane, h and d are unrecoverable, as well as the global scale.

4.4.4 Multiple-plane Configurations

We study configurations D, F and H specifically for two planes, as they were shown to be under-constrained for one plane, and show that most of the unknowns becomes well-constrained.

Configuration D, 2-plane. There are five unknowns in the geometric parameters, namely the orientation $\mathbf{N}$, $\mathbf{N}'$ and distance d, d' of each scene plane to the camera and the PLS position $\mathbf{S}$. Sub-systems S_B, S'_B determine $\mathbf{N}$, $\mathbf{N}'$, as per lemma 2, with a finite number of solutions. Singling out equation (18.2) from S_A, S'_A and the unknown $\hat{r}$, $\hat{r}'$, the remaining 6 equations are linear in 5 unknowns and the system is thus over-determined.

Configuration F, 2-plane. There are three unknowns in the geometric parameters, namely the distance h, h' of each scene plane to the PLS and the PLS position $\mathbf{S}$. Sub-systems S_B, S'_B do not contain unknowns. Singling out equation (18.2) from S_A, S'_A and the unknown $\hat{r}$, $\hat{r}'$, the remaining 6 equations are linear in 5 scalar unknowns and the system is thus over-determined.

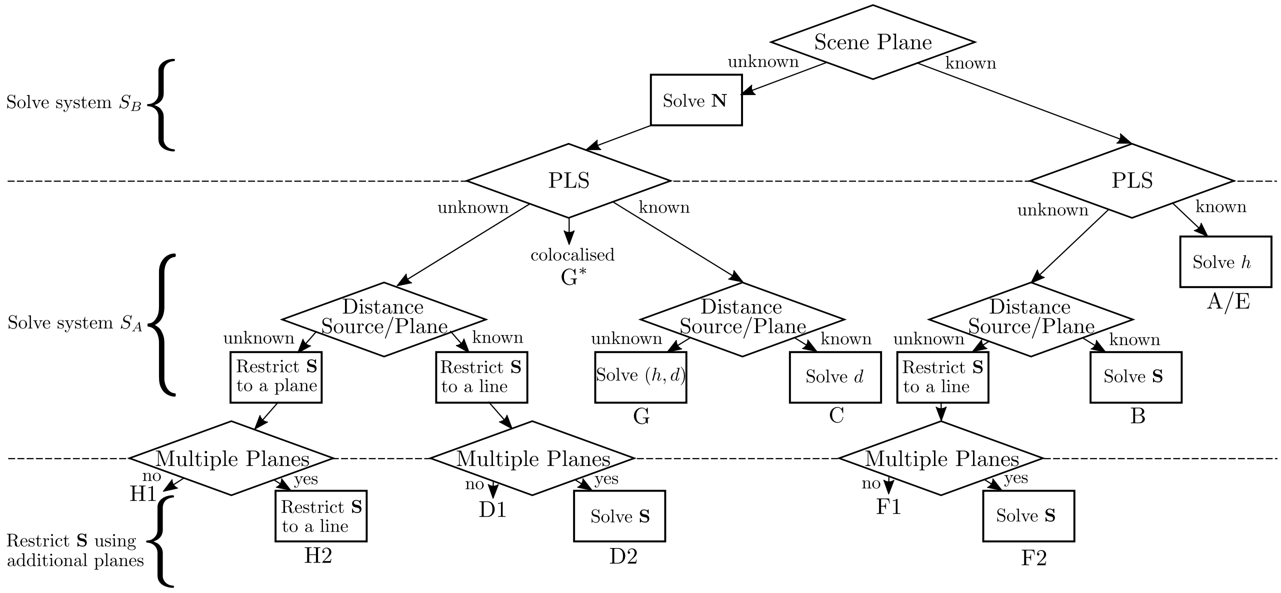

Figure 3: Steps taken for 3D reconstruction for all possible combinations of scene priors and numbers of planes. The correspondence with the configurations of table 2 is indicated on the leaf nodes, with the configuration letter and where 1 means a single plane and 2 means multiple planes.

Configuration H, 2-plane. There are seven unknowns in the geometric parameters, namely the distance h, h' of each scene plane to the PLS, their orientation $\mathbf{N}$, $\mathbf{N}'$, their distance d, d' to the camera, and the PLS position $\mathbf{S}$. Subsystems S_B, S'_B determine $\mathbf{N}$, $\mathbf{N}'$, as per lemma 2, with a finite number of solutions. Singling out equations (18.2) from S_A, S'_A and the unknown $\hat{r}$, $\hat{r}'$, the remaining 6 equations are linear and homogeneous equations in 7 scalar unknowns. The parameters h, h', d and d' are thus unrecoverable, as well as the global scene scale.

4.4.5 Configuration G^*: Co-located Light and Camera Configuration

Configuration G^* is a special case of configuration G, where the known PLS is specifically placed at the camera centre. This configuration commonly appears in medical and industrial endoscopes. At first, one may think that it should behave similarly to configuration G, which uses the same priors, the unknowns and the equations thus remaining identical. Indeed, similarly to configuration G, the plane orientation $\mathbf{N}$ can be determined from S_B. However, the similarity does not hold for S_A, which does not share the same degrees of freedom. Algebraically, as $\mathbf{S}$ is the zero vector, $\mathbf{N}$ and $\mathbf{S}$ are always co-linear, making h unrecoverable from the quasi-homogeneous equations of S_A. Consequently, the scene plane to camera distance d becomes unrecoverable.

Unlike configuration H, where multiple planes give systems that share a common inhomogeneous variable, the position of the PLS, each system in configuration G^* is independently quasi-homogeneous. Consequently, fixing the scale for one scene plane does not fix the scale for the others.

5 Closed-form Solutions from Algebraic Geometry

Given a single or multiple isophotes and priors on the scene elements, 3D reconstruction is achieved by solving the polynomial system (18). While the previous section analyses the solution space, the current section gives solution methods for the different cases.

5.1 Overview

The proposed solution methods for the different cases share some of their components, as shown in the flowchart of figure 3. The flowchart also relates each solution method to a case in table 2. The methods rely on three layers. The first two layers deal with a single scene plane and are thus repeated for each of the scene planes. The third layer combines the outputs of the first two layers for multiple planes. Specifically, the first layer deals with the scene plane normal by solving sub-system S_B, except if the scene priors would already include this normal. The second layer deals with the rest of the unknowns, essentially the PLS, but also the PLS to plane distance and the camera to plane

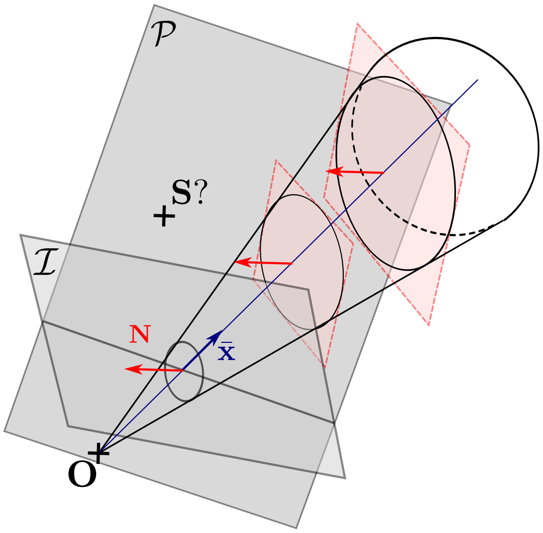

Figure 4: The PLS is located on the plane $\mathcal{P}$, which is formed from the isophote image.

distance, by solving sub-system S_A, taking the scene priors into account. The next three sections present the steps taken in each of the three layers in turn. Each step of each layer involves a closed-form solution, which guarantees that all possible solutions are found and avoids iterative nonconvex numerical optimisation, hence local minima.

5.2 Scene Plane Normal Reconstruction

We use the circle-based plane-pose method [Mariyanayagam et al., 2018] to estimate each scene plane normal $\mathbf{N}$ from an isophote given as a point conic matrix E. Matrix E is symmetric, full-rank and defined up to scale. We can thus scale it so that $\det(\mathsf{E}) = 1$. Defining $\mathbf{V}_j$, $j \in [1, 3]$, as the eigenvectors of E in descending order of their eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, we have a two-fold ambiguous solution:

$$\mathbf{N}^{\pm} = \sqrt{\lambda_1 - \lambda_2}\frac{\mathbf{V}_1}{\|\mathbf{V}_1\|} \pm \sqrt{\lambda_2 - \lambda_3}\frac{\mathbf{V}_3}{\|\mathbf{V}_3\|}, \tag{21}$$

where, with $q = \text{trace}(\mathsf{E})$ and $h = \text{trace}(\mathsf{E}^{-1})$, we have:

$$\alpha = \frac{1}{3}\arccos\left(\frac{1}{2}\frac{2q^3 - 9qh + 27}{(q^2 - 3h)^{\frac{3}{2}}}\right) \tag{22}$$

$$\lambda_1 - \lambda_2 = \sqrt{q^2 - 3h}\left(\cos\alpha - \frac{\sqrt{3}}{3}\sin\alpha\right) \tag{23}$$

$$\lambda_2 - \lambda_3 = \frac{2\sqrt{3q^2 - 9h}}{3}\sin\alpha. \tag{24}$$

For the specific case where $\lambda_2 = \lambda_3$, the second term of equation (21) vanishes. This case occurs if the line joining the BP and the camera centre is parallel to the scene plane normal. It specifically happens in configuration G^*, where the PLS and the camera centre are co-localised.

5.3 Point-light Source Reconstruction from a Single Scene Plane

We suppose that one or two solutions are available for the scene plane normal, denoted $\mathbf{N}^{\pm}$. We handle them separately, using equations (18.1), (18.3) and (18.4) from sub-system S_A. Equation (18.2) is solved separately as it is the only equation which involves the unknown radius. We introduce the homogeneous coordinates of the BP, denoted $\bar{\mathbf{x}}$, given by:

$$\bar{\mathbf{x}} \sim \mathsf{E}^{-1}\mathbf{N} \quad \text{with } \bar{\mathbf{x}} \sim \bar{\mathbf{X}} = \mathbf{S} + h\mathbf{N}. \tag{25}$$

We can easily verify that $\bar{\mathbf{x}}$ is a solution to equations (18.3) and (18.4) if $\mathbf{N}$ is a solution to equations (18.10) and (18.11). Figure 4 illustrates the geometry, by showing the scene plane normal, $\bar{\mathbf{x}}$ and the plane $\mathcal{P}$, which we define to contain both these vectors.

Configurations C, E. The PLS and one of the distances h or d between the scene plane and the other scene elements are known; equation (18.1) thus directly gives the other distance.

Configuration B. The distance between the scene plane and the PLS, and the distance between the scene plane and the camera centre are known; the sub-system can thus be reformulated as an inhomogeneous linear system in $\mathbf{S}$ as:

$$\mathbf{S}^\top \begin{bmatrix}\mathsf{E}^\top\mathbf{R}_1 & \mathsf{E}^\top\mathbf{R}_2 & \mathbf{N}\end{bmatrix} + \begin{bmatrix} h\mathbf{R}_1^\top\mathsf{E}\mathbf{N} & h\mathbf{R}_2^\top\mathsf{E}\mathbf{N} & -d-h\end{bmatrix} \;=\; 0, \tag{26}$$

which gives, in the general case, a single solution to $\mathbf{S}$.

Configuration G. The PLS is known; we can thus use one of the equations (18.3) and (18.4) to obtain a solution to h and use it to solve for d from equation (18.1). This is valid provided that $\mathbf{S}$ and $\mathbf{N}$ are not co-linear. In particular, in configuration G^* where the PLS and the camera centre are co-localised, hence $\mathbf{S} = \mathbf{0}$, none of these equations can constrain h.

Configuration D. The distance between the scene plane and the PLS h is known; we can thus use the homogeneous coordinates of the BP from equation (25) to restrict the PLS. We then have:

$$\mathbf{S} = \lambda\bar{\mathbf{x}} - h\mathbf{N} \quad \text{with } \lambda \neq 0. \tag{27}$$

The PLS $\mathbf{S}$ is then restricted to a line whose Plücker coordinates are given by $(\bar{\mathbf{x}}, -h\mathbf{N} \times \bar{\mathbf{x}})$, where we use × to denote the vector cross-product. The distance between the scene plane and the camera cannot be recovered.

Configuration F. The scene plane to camera distance d is known and non zero; we can thus estimate the BP from equation (18.1), $\mathbf{N}^\top\bar{\mathbf{X}} - d = 0$. Using its homogeneous coordinate $\bar{\mathbf{x}}$ equation (25) gives the following general solution:

$$\bar{\mathbf{X}} = -\frac{d\mathsf{E}^{-1}\mathbf{N}}{\mathbf{N}^\top\mathsf{E}^{-1}\mathbf{N}}. \tag{28}$$

The PLS $\mathbf{S}$ is then restricted to a line whose Plücker coordinates are given by $(\mathbf{N}, \bar{\mathbf{X}} \times \mathbf{N})$. The distance between the scene plane and the PLS cannot be recovered.

Configuration H. In the absence of other priors, we can only restrict the PLS to plane $\mathcal{P}$. Indeed, introducing a scalar λ to handle homogeneity in equation (25), we have:

$$\mathbf{S} = \lambda\bar{\mathbf{x}} - h\mathbf{N}. \tag{29}$$

The supporting plane $\mathcal{P}$ contains both the BP's backprojection sightline and the scene plane normal. Its coordinates are thus given by:

$$\bar{\mathbf{P}} = \begin{bmatrix}\bar{\mathbf{x}} \times \mathbf{N} \\ 0\end{bmatrix}. \tag{30}$$

Importantly, the supporting plane is not affected by the scene plane normal ambiguity $\mathbf{N}^\pm$, as shown in the following proposition.

Proposition 6. *The supporting plane $\mathcal{P}$ does not depend on the ambiguous solution obtained for the scene plane normal.*

Proof. We form the supporting plane equations $\mathbf{P}^+$ and $\mathbf{P}^-$ corresponding to both scene plane normals $\mathbf{N}^+$ and $\mathbf{N}^-$, and then show that they are equal. We have:

$$\mathbf{P}^+ = \bar{\mathbf{x}}^+ \times \mathbf{N}^+ = (\mathsf{E}^{-1}(\mathbf{p}_1 + \mathbf{p}_3)) \times (\mathbf{p}_1 + \mathbf{p}_3) \tag{31}$$

$$\mathbf{P}^- = \bar{\mathbf{x}}^- \times \mathbf{N}^- = (\mathsf{E}^{-1}(\mathbf{p}_1 - \mathbf{p}_3)) \times (\mathbf{p}_1 - \mathbf{p}_3), \tag{32}$$

where $\mathbf{p}_1 = \sqrt{\lambda_1 - \lambda_2)}\frac{\mathbf{V}_1}{\|\mathbf{V}_1\|}$ and $\mathbf{p}_3 = \sqrt{\lambda_2 - \lambda_3)}\frac{\mathbf{V}_3}{\|\mathbf{V}_3\|}$. Consequently, we have:

$$\begin{aligned}
\mathbf{P}^+ - \mathbf{P}^- &= 2((\mathsf{E}^{-1}\mathbf{p}_1) \times \mathbf{p}_1 + (\mathsf{E}^{-1}\mathbf{p}_3) \times \mathbf{p}_3) && (33)\\
&= 2(\lambda_1\sqrt{\lambda_1 - \lambda_2)}\frac{\mathbf{V}_1}{\|\mathbf{V}_1\|} \times \sqrt{\lambda_1 - \lambda_2)}\frac{\mathbf{V}_1}{\|\mathbf{V}_1\|} + \lambda_3\sqrt{\lambda_2 - \lambda_3)}\frac{\mathbf{V}_3}{\|\mathbf{V}_3\|} \times \sqrt{\lambda_2 - \lambda_3)}\frac{\mathbf{V}_3}{\|\mathbf{V}_3\|}) && (34)\\
&= 0. && (35)
\end{aligned}$$

The normal vectors $\mathbf{P}^+$ and $\mathbf{P}^-$ thus represent the same orientation. □

5.4 Point-light Source Reconstruction Refinement from Multiple Scene Planes

Observing a single scene plane leaves ambiguities on its pose and on the PLS in cases D, F and H, whereas it does not for all the other cases. We study how the combination of multiple scene planes can reduce the ambiguities. In short, observing at least two scene planes fully determine the PLS in cases D and F and restricts it to a line in case H.

Configurations D, F. Each isophote gives a line or two lines containing the PLS $\mathbf{S}$ as discussed in section 5.3 for configurations D and F. These lines are denoted $\mathcal{L}_i$, $i \in [1, n]$, and their Plücker coordinates are denoted $(\mathbf{D}_i, \mathbf{M}_i)$ with $\|\mathbf{D}_i\| = 1$. In configuration D, we have $\mathbf{D}_i = \mathsf{E}^{-1}\mathbf{N}_i$ and $\mathbf{M}_i = -h_i\mathbf{N}_i \times \mathbf{D}_i$. In configuration F, we have $\mathbf{D}_i = \mathbf{N}_i$ and $\mathbf{M}_i = \bar{\mathbf{X}} \times \mathbf{N}_i$ with $\bar{\mathbf{X}}$ given by equation (28). The distance between the PLS $\mathbf{S}$ and the line $\mathcal{L}_i$ is given by:

$$d(\mathbf{S}, \mathcal{L}_i) = \|\mathbf{S} \times \mathbf{D}_i - \mathbf{M}_i\|_2. \tag{36}$$

We can thus find the PLS which minimises the distance to both lines by solving:

$$\min_{\mathbf{S}} \sum_{i=1}^{n} \|[\mathbf{D}_i]_\times \mathbf{S} - \mathbf{M}_i\|^2,$$

where $[\cdot]_\times$ is the skew-symmetric cross-product matrix. This is a linear-least squares minimisation problem, which is solved in closed-form. Once a solution for the PLS is found, the remaining scene parameters (the h_i for configuration D and the d_i for configuration F) are computed from equation (18.1).

Configuration H. Each isophote gives a plane $\mathcal{P}_i$, $i \in [1, n]$, containing the PLS $\mathbf{S}$ as discussed in section 5.3 for configuration H. These planes all contain the camera centre and the PLS; consequently they all intersect in a single line corresponding to the backprojected sightline of the PLS. The direction $\mathbf{L}$ of this line may be found that minimises the distance to a unitary point $\bar{\mathbf{P}}_i$ of all planes by solving:

$$\min_{\mathbf{L}} \sum_{i=1}^{n} \|\bar{\mathbf{P}}_i^\top \mathbf{L}\|^2 \text{ s.t. } \|\mathbf{L}\| = 1.$$

This is a homogeneous linear-least squares problem, which is solved in closed-form. We obtain the direction of the PLS, related to the actual PLS as $\mathbf{S} = \|\mathbf{S}\|\mathbf{L}$. This remains valid even if the PLS is located on the camera's principal plane. However this fails as the PLS approaches the camera centre. This is because $\mathbf{P}_i$ vanishes as $\bar{\mathbf{x}}_i$ and $\mathbf{N}_i$ become collinear; the linear least-squares problem then becomes undetermined.

In order to obtain a complete reconstruction, we may fix the overall signed scale and then the position of the PLS in camera coordinates. It then becomes possible in some cases to resolve the algebraic ambiguity on the scene plane normal if a simple sufficient condition is met, which we now develop. By convention, we use a normal vector pointing towards the camera, denoted with a top bar, which gives $\bar{\mathbf{N}}^\top\mathbf{X} < 0$ for any point $\mathbf{X}$ on the plane. A plane is only shaded and visible if the PLS and the camera center lie on the same side of the plane, giving the algebraic condition $(\bar{\mathbf{N}}^\top\mathbf{X})(\bar{\mathbf{N}}^\top(\mathbf{S} - \mathbf{X})) < 0$. From this condition, the twofold ambiguity may be resolved if each ambiguous normal is associated with a distinct partition of the space where the PLS may be located. Given the PLS, the right solution is then naturally revealed by identifying the space partition containing it. This idea is illustrated in figure 5 and formalised in the following proposition.

Proposition 7. *The scene plane normal ambiguity can be resolved given the PLS and a visible point* $\mathbf{X}$ *if:*

$$(\mathbf{X}^\top\mathbf{N}^+)(\mathbf{X}^\top\mathbf{N}^-) > 0. \tag{37}$$

where the normals' signs are chosen so that they both points towards or away from the camera centre.

Proof. Following the algebraic condition on the visibility of a shaded surface point, we have in particular:

$$\mathbf{N}^\top(\mathbf{S} - \bar{\mathbf{X}}) > 0 \tag{38}$$

$$\mathbf{N}^\top\bar{\mathbf{X}} < 0. \tag{39}$$

We derive the BP from equation (25) by introducing $\mu \neq 0$ as:

$$\bar{\mathbf{X}}^\pm = \mu\mathsf{E}^{-1}\mathbf{N}^\pm \tag{40}$$

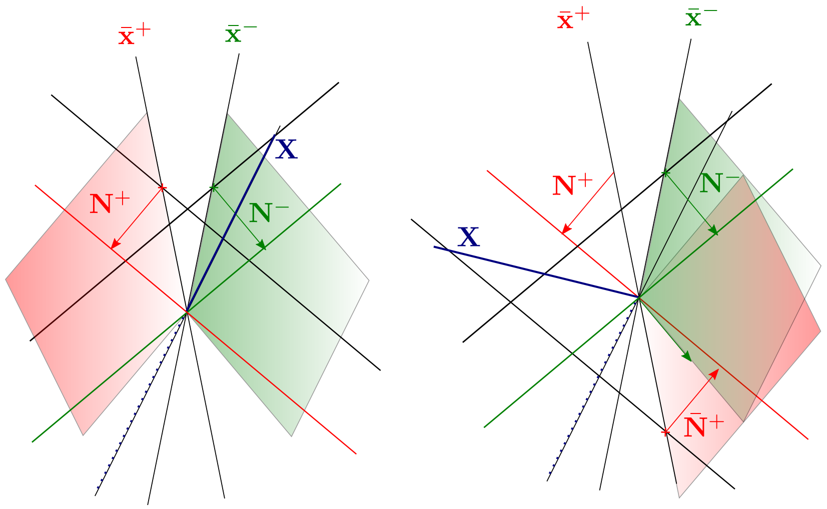

Figure 5: The supporting plane $\mathcal{P}$, representing the possible locations of the PLS, can be separated in two parts shown in red and green if both the estimated normals $\mathbf{N}^+$ and $\mathbf{N}^-$ face the camera at a visible point $\mathbf{X}$ (left). The two parts are however not disjoint as the sign of only one ambiguous normal must be inverted to follow the orientation convention (right).

Therefore, $\bar{\mathbf{X}}^\pm$ is given by:

$$\bar{\mathbf{X}}^\pm = \mu \left(\frac{\sqrt{\lambda_1 - \lambda_2}}{\lambda_1} \mathbf{V}_1 \pm \frac{\sqrt{\lambda_2 - \lambda_3}}{\lambda_3} \mathbf{V}_3 \right). \tag{41}$$

Equations (39) and (21) then give:

$$(\mathbf{N}^\top \mu \mathsf{E}^{-1} \mathbf{N}) > 0 \tag{42}$$

$$(\mathbf{N}^\top \mu \mathsf{E}^{-1} \mathbf{N}) = \mu \left(\frac{\lambda_1 - \lambda_2}{\lambda_1} - \frac{\lambda_2 - \lambda_3}{\lambda_3} \right). \tag{43}$$

The sign of the eigenvalues can be determined from $\det(\mathsf{E}) = 1$ and the Sylvester's law of inertia:

$$\lambda_1 > 0 > \lambda_2 \geq \lambda_3. \tag{44}$$

We obtain the sign of μ from equation (43) as $\mu < 0$.

Equation (38) reveals the sign of h, by substituting $\bar{\mathbf{X}}$ with its definition from equation (13), leading to $-\mathbf{N}^\top h \mathbf{N} > 0$, hence $h < 0$.

Finally, we can estimation the PLS $\mathbf{S}$ from equations (13) and (41) as:

$$\mathbf{S} = \left(\mu \frac{\sqrt{\lambda_1 - \lambda_2}}{\lambda_1} - h\sqrt{\lambda_1 - \lambda_2} \right) \mathbf{V}_1 \pm \left(\mu \frac{\sqrt{\lambda_2 - \lambda_3}}{\lambda_3} - h\sqrt{\lambda_2 - \lambda_3} \right) \mathbf{V}_3. \tag{45}$$

Using the signs of μ and h shows that the sign of the second coefficient, for $\mathbf{V}_3$, only depends on which of the ambiguous solutions is chosen. As vectors $\mathbf{V}_i$ are orthogonal, the plane can be separated in two parts. Therefore knowing the PLS removes the ambiguity. The coefficient can only be 0 if $\lambda_2 = \lambda_3$, for which there is no ambiguity, or if $\mu = 0$, where the camera centre is on the scene plane, forming an impossible degenerate case where the scene plane would project to a line. □

6 Isophote Detection

We discuss the practical aspects of isophote detection and estimation in an input image. These isophotes are image curves forming geometric primitives, whose theoretical modelling and usage for 3D reconstruction are studied in the previous section. Isophote detection is thus an early and crucial step for our 3D reconstruction method, similarly

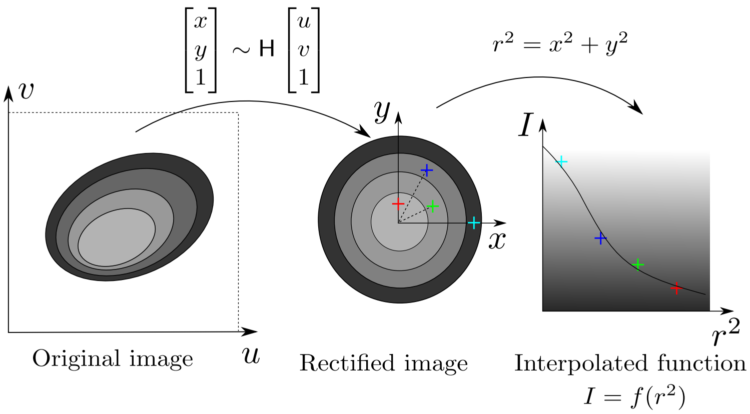
$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \mathsf{H} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$
$r^2 = x^2 + y^2$
v
u
y
x
I
r^2
Original image
Rectified image
Interpolated function
$I = f(r^2)$

Figure 6: Proposed top-down, generative method for isophote detection, by rectifying homography and intensity profile optimisation. The method fits a generative image model by estimating the rectifying homography H, mapping the isophote ellipses to concentric circles, and the monotonic intensity profile function f, mapping the circle radius to the observed intensity.

to keypoint detection and matching in classical SfM. Importantly, we have established with our theoretical model that the isophotes are imaged as non-degenerate conics, a property which we explicitly use in our detection methods. Several factors affect the stability of isophote detection in real images, including sensor noise, modelling errors (albedo variation, surface curvature, *etc*) and visibility conditions such as occlusions, creating open isophote curves. We propose a bottom-up and a top-down methods. Without loss of generality, we select two isophotes, for ρ values linearly spanning the intensity value interval from $\mathcal{I}(\mathbf{u})$ with $\mathbf{u} \in \mathcal{U}_i$ for each plane i.

6.1 Naive Bottom-up Detection

We consider a small patch centred around the area of maximum intensity on an observed surface where, according to proposition 2, the isophotes are elliptic. The use of elliptic primitives based on edge points is known to be robust and efficient in fiducial marker detection [Calvet et al., 2016]. However, the isophote points are not edge points. In contrast, from our image model in equation (7), we should in theory simply search for pixels whose intensity value is $\mathcal{I}(\mathbf{u}) = \rho$. As opposed to classical ellipse detection, these pixels do not lie at sharp intensity transitions, making their detection highly sensitive to noise. The use of noise removal filters is thus of utmost importance. Concretely, we use a Wiener filter followed by a Gaussian filter. The Wiener filter has been chosen for the good experimental results on real images where high frequency noise was correclty removed. The Gaussian filter is used to slightly blur the image so that the intensity can be threshold adequately on a continuous domain. The isophote points are then detected by intensity thresholding within $[\rho - \epsilon, \rho + \epsilon]$, where ϵ is a tolerance which we set empirically to $\epsilon = \frac{1}{600}$. We finally fit an ellipse to the isophote points following [Szpak et al., 2015]. In practice, this naive method has only worked on synthetic data with limited noise. This is because it detects a single isophote at a time, without exploiting the strong structure of nested ellipses that the multiple isophotes follow.

6.2 Advanced Top-down Detection

We propose to fit a generative shading image model, drawn from the concentric circle structure established in proposition 1. In our model, the homography H of equation (17) transforms the image isophote ellipses into the concentric circles lying on the scene plane, which are the solutions of equation (6). Our key idea is illustrated in figure 6: we estimate the rectifying homography H, along with a one-dimensional intensity profile for the isophotes, which we parameterise by the pre-image circle radius. This approach can alternatively be seen as solving for ρ from equation (9).

The intensity profile is represented by a function f of the square radius. Function f is strictly decreasing, since, as shown in proposition 1, the pre-image circles are centred on the BP. As function f is also smooth, we model it using a monotonic cubic spline [Wolberg and Alfy, 1999] with N control points. Function $f : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ thus predicts the image intensity $\mathcal{I}(u, v)$ for some square radius $x^2 + y^2$ obtained from normalised coordinates u, v by homography transfer as $[x\, y\, 1]^\top \sim \mathsf{H}\, [u\, v\, 1]^\top$. It is controlled by squared radii values $\mathbf{p} \in \mathbb{R}^N$ at the N control points, which are to be estimated. The corresponding intensity values at the control points are fixed following a uniform sampling of the maximum to the minimum intensity values within the image patch.

We use a specific reduced parameterisation of the homography. Indeed, a general homography has 8 DoF, hence 8 parameters at least. We can reduce this number to only 4 parameters because a) we use a calibrated camera, b) we only use the scene plane normal (as opposed to the full rotation matrix, as the concentric isophote circles are rotational symmetric) and c) the scale can be arbitrarily fixed in the translation vector. We thus end up with 2 rotational and 2 translational parameters. In addition, in the specific case where the PLS and the camera are co-localised, the translation vanishes and we end up with only 2 rotational parameters to estimate.

We design an optimisation problem whose goal is to minimise the discrepancy between the image and the model-generated intensity values over the image area $\mathcal{U}$ showing the scene plane. This leads to the following nonlinear least-squares problem:

$$\min_{\mathsf{H},\mathbf{p}} \sum_{(u,v)\in\mathcal{U}} \left\| \mathcal{I}(u,v) - f(x^2+y^2,\mathbf{p}) \right\|^2 ,$$

where H is our homography parameterisation with 4 or 2 parameters and $\mathbf{p}$ is the intensity at the spline's control points. The homography parameters are initialised using the maximum and minimum intensity points in the patch: specifically, we form the homography which rectifies the point of minimum intensity to $[0\,0\,1]^\top$ and the point of maximum intensity to $[1\,0\,1]^\top$. We empirically found that using $N = 5$ control points was enough, which we initialise using a linear distribution spanning the interval between the minimum and maximum intensity and their corresponding squared radii obtained by applying the initial homography to extrema points.

The number of unknown parameters to estimate depends on the number of control points and the homography parameterisation, and is thus $4 + N = 4 + 5 = 9$ for the general case and $2 + N = 2 + 5 = 7$ for the co-localised PLS and camera case. We use the Levenberg-Marquardt method to minimise the cost function. Upon completion, one or several isophotes can be easily extracted within the patch. This is done by estimating the radius values for each image pixel with the inverse homography. One or several radii are selected in the working range, then their corresponding isophote is obtained from equation (16), from which the scene plane normal is estimated.

We use isophotes rather than the homography to estimate the scene plane normal for two reasons. First, the same method applies to both the bottom-up and the top-down approaches. Second, a single isophote contains all and only the useful information for our model. Specifically, using only one isophote gives two ambiguous solutions for the normal, whereas using the homography would theoretically give a single normal. This may seem like a drawback of the approach at first; nonetheless, it is a strong advantage because the homography solution is unstable, especially for smaller patches, often corresponding to a spurious normal solution, without any possibility to recover. This corresponds to the known two-way affine ambiguity studied in plane pose estimation [Collins and Bartoli, 2014]. In contrast, using the isophote to obtain the scene plane normal guarantees that the true normal is within the two ambiguous normal solutions.

7 Refinement from Geometric and Photometric Discrepancy

We show how to refine an initial 3D reconstruction of the scene found from one of the methods of the previous section. We describe two criteria measuring the quality of fit to the observed data, a geometric one and a photometric one, which can be iteratively minimised. We then discuss the relationship between the two criteria.

7.1 Geometric Criterion

Our geometric criterion is based on the reprojection error computed by predicting the equation of each observed isophote primitive with the model being reconstructed. Specifically, we consider that the observed data is obtained from the bottom-up isophote detection method and is a set of image points on the different isophotes, for specific values of the intensity. Concretely, we denote as $\mathbf{Y}_{i,j,k}$ the image coordinates of the k^{th} point for the j^{th} isophote of the i^{th} scene plane. We parameterise each plane by its pose $\mathbf{h}_i = [\alpha_i\,\beta_i\,\gamma_i, d_i]$ in camera coordinates, namely two angles α_i, β_i defining its normal, an angle γ_i defining the in-plane orientation and its distance d_i to the camera centre. We associate to each observed image point a 3D point on the corresponding scene plane, which we parameterise in polar coordinates by an angle $\theta_{i,j,k}$ and the radius $r_{i,j}$ of its associated isophote circle on the scene plane. We gather all the parameters as $\mathcal{H} = \{\mathbf{h}_i\}$, $\mathcal{R} = \{r_{i,j}\}$ and $\mathcal{O} = \{\theta_{i,j,k}\}$ with $i \in [1,n]$, $j \in [1,m]$ and $k \in [1,N_{i,j}]$ and formulate the following nonlinear least-squares problem:

$$\min_{\mathbf{S},\mathcal{H},\mathcal{R},\mathcal{O}} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{N_{i,j}} \|\mathbf{P}_{i,j,k} - \mathbf{Y}_{i,j,k}\|^2.$$

The cost function requires one to evaluate the reprojection $\mathbf{P}_{i,j,k}$ of the points, which is done in three steps. First, we compute the normal of each scene plane as $\mathbf{N}_i = [\cos\alpha_i \cos\beta_i \; \sin\alpha_i \cos\beta_i \; \sin\beta_i]^\top$ and the orthogonal projection of the PLS on the plane, which is the BP, as $\mathbf{X}_i = \mathbf{S} - (d_i + \mathbf{S}^\top \mathbf{N}_i)\mathbf{N}_i$. Second, we compute the pre-image of each point on the scene plane as $\mathbf{W}_{i,j,k} = \mathbf{X}_i + r_{i,j}[\cos\theta_{i,j,k}, \sin\theta_{i,j,k}]^\top \bar{\mathsf{R}}_i$ with:

$$\bar{\mathsf{R}}_i = \begin{bmatrix} -\sin\alpha_i & -\sin\beta_i\cos\alpha_i \\ \cos\alpha_i & -\sin\alpha_i\sin\beta_i \\ 0 & \cos\beta_i \end{bmatrix}.$$

Third, we reproject each point, which is given in homogeneous coordinates by $[\mathbf{P}_{i,j,k}^\top\, 1]^\top \sim \mathsf{K}\mathbf{W}_{i,j,k}$.

The geometric criterion has latent parameters which model the position of each observed isophote point on the scene plane, namely, their pre-images. These latent parameters simplify the evaluation of the criterion. Alternatively, they could be cancelled out. This would be less practical in terms of evaluation but allows one to understand the criterion as a measure of distance between the observed and the model predicted isophotes, as:

$$\min_{\mathbf{S},\mathcal{H},\mathcal{R}} \sum_{i}^{n}\sum_{j=1}^{m}\sum_{k=1}^{N_{i,j}} \min_{\theta_{i,j,k}} \|\mathcal{G}(\mathbf{X}_i + r_{i,j}[\cos\theta_{i,j,k}, \sin\theta_{i,j,k}]^\top) - \mathbf{Y}_{i,j,k}\|^2 = \min_{\mathbf{S},\mathcal{H},\mathcal{R}} d(\mathcal{G}(\mathcal{C}(r_{i,j}, \mathbf{X}_i, \mathbf{N}_i)), \mathbf{Y}_{i,j,k})^2, \tag{46}$$

where $\mathcal{C}(r, \mathbf{X}, \mathbf{N})$ is the circle of radius r and centre $\mathbf{X}$ on the plane with normal $\mathbf{N}$ and $d(\cdot,\cdot)$ gives the Euclidean distance between the circle projected by function $\mathcal{G}$ and the observed isophote points.

7.2 Photometric Criterion

Our photometric criterion is based on the dense intensity difference computed by predicting the pixel intensity for the image area of each scene plane. This can be seen as an extension of our top-down isophote detection method to a full 3D scene generative optimisation with n planes. It uses a similar intensity profile parameterisation with monotonic splines for each plane $i \in [1, n]$, parameterised by control points $\mathbf{p}_i \in \mathbb{R}^N$. Similarly to the geometric criterion, we parameterise each plane by $\mathbf{h}_i = [\alpha_i\, \beta_i\, d_i]$. We gather all the parameters as $\mathcal{H} = \{\mathbf{h}_1, \ldots, \mathbf{h}_n\}$ and $\mathcal{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ and formulate the following nonlinear least-squares problem:

$$\min_{\mathbf{S},\mathcal{H},\mathcal{P}} \sum_{i=1}^{n} \sum_{(u,v)\in\mathcal{U}_i} \left\| I(u,v) - f(s^2, \mathbf{p}_i) \right\|^2.$$

The cost function requires one to evaluate the square distance s^2 to the BP for image point (u, v), which is done in three steps. First, we compute the normal $\mathbf{N}_i$ and the BP $\mathbf{X}_i$ as for the geometric criterion. Second, we compute the intersection of each image point's backprojected sightline $\bar{\mathbf{x}} = [u\, v\, 1]$ with the scene plane as $\mathbf{Y} = -d_i/(\mathbf{N}_i{}^\top \bar{\mathbf{x}})\bar{\mathbf{x}}$. Third, we compute the sought distance as $s^2 = \|\mathbf{Y} - \mathbf{X}_i\|^2$.

7.3 Comparing the Criteria

The Figure 7 shows the distribution (grey area) of points in a patch containing isophotes as a function of their distance from the BP in the rectified coordinate system and their intensity in a similar way to Figure 6. Ideally, one expects to have a distribution that follows only the green line, the theoretical model. If we consider a rectification error on a rectangular patch, at the corners the pixels are slightly offset, creating "branches" on the distribution. The aim of the optimisation is to correct this deviation, we compare on this diagram how the proposed criteria penalise it.

The photometric criterion corresponds to measuring the difference in intensity between that measured on a point and that of the isophote associated with the same distance s to the BP on the rectified image. In other words, it is the vertical distance (in blue) on the diagram between a point and the model (green spline).

The geometric criterion corresponds to measuring the reprojection distance in the original image between the position of the point and its associated isophote. This geometric distance is not exactly equivalent to the one in the rectified image, but it can be related. In the rectified image, it corresponds to the horizontal distance (in red) between a point and its associated isophote of similar intensity (red dot). The radius of this isophote is estimated from all the points of the same intensity (red zone).

It can be seen that both criteria penalise the deviation of the data from our model. The geometric criterion is however discrete since the isophotes are considered individually with a thresholding on the intensity levels.

8 Experimental Results

We present experimental results on both synthetic and real images. Our objectives are twofold. First, we validate our geometric and photometric models on realistic scenes. Second, we measure the reconstruction ability and accuracy of our pose estimation method in two concretes configurations. The first configuration considers scene planes and a

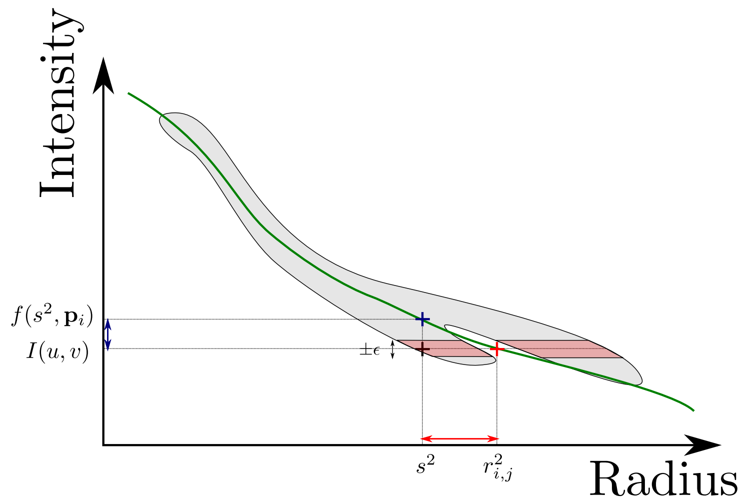

Figure 7: The relationship between the geometric and the photometric criteria. We consider the pixels of a patch, spread according to their intensity and their distance to the BP. The geometric criterion corresponds to the horizontal distance in red and the photometric criterion corresponds to the vertical distance in blue.

PLS with unknown poses, which is configuration H in table 2. This configuration demonstrates the ability of our method to work even in settings where classical SfS approaches are not applicable and fail to obtain results. The second configuration considers the endoscopic case, with a PLS co-located with the camera and a scene plane with unknown pose, which is configuration G^* in table 2.

8.1 Model Validation

We created a lab setup using a candle for the PLS, lighting three boards with diffuse reflectance, shown left in figure 8. Board A is covered by a fine-grained matte garment and boards B and C by grainy white papers. We chose to use a candle for three main reasons: its small size relative to the boards, its isotropic nature due to the absence of optics and its sufficient intensity to capture images with a reasonable level of noise. We attempted prior experiments using an LED and an incandescent bulb as light sources but they did not comply with the PLS model as consistently. While the LED is small, it has the main disadvantage of being strongly anisotropic; with a principal direction where the illumination is greater, it is probably better modelled by a spotlight. The incandescent bulb is closer to the PLS but its illumination is not uniform, depending on the shape of the glass, this effect can be seen on surfaces close to it. We used a high-end DSLR camera which we calibrated with the Matlab toolbox. We carefully acquired pose groundtruth using ARTag markers. In particular, the candle stick was precisely positioned at the centre of the four markers from its support plane and the height of the brightest point of the flame was measured.

We validate our model by measuring the compliance of the observed image intensities with the expected radial distribution. Figure 9 shows the pixel intensities from the board images as red dots, distributed according to their square distances on the plane to the brightest point. On the first row, we used the groundtruth to calculate these distances (configuration A), on the second row we optimised the PLS position to decrease the photometric discrepancy (configuration B), on the third row we optimised a perspective rectification for each scene plane individually and on the fourth row we performed a complete pose optimisation based on the photometric criterion (configuration H). The fitted intensity model is shown as a blue curve in figure 9. Table 3 shows statistics on the distance between the measured pixel intensities and the fitted curve for the three configurations.

First row, using groundtruth to calculate the radius of each pixel. We observe that the measured intensities follow a similar trend to our model, namely a decreasing curve. However we can see multiple branches, indicating a deviation from our model. This is caused by an offset existing between the true pose and the measured groundtruth.

Second row, optimising the PLS position with the photometric criterion. We observe that the measured intensities follow the model curve more consistently than on the first row. The remaining discrepancies are due to the flame not being a single point. While its measured groundtruth position corresponds to the wick, its ideal position should takes into account the entire distribution of the light rays, causing a possible offset of up to 3 cm for this image. However we can still see that the fitting is not perfect, data points are scattered and two branches appear on the board C (leftmost image).

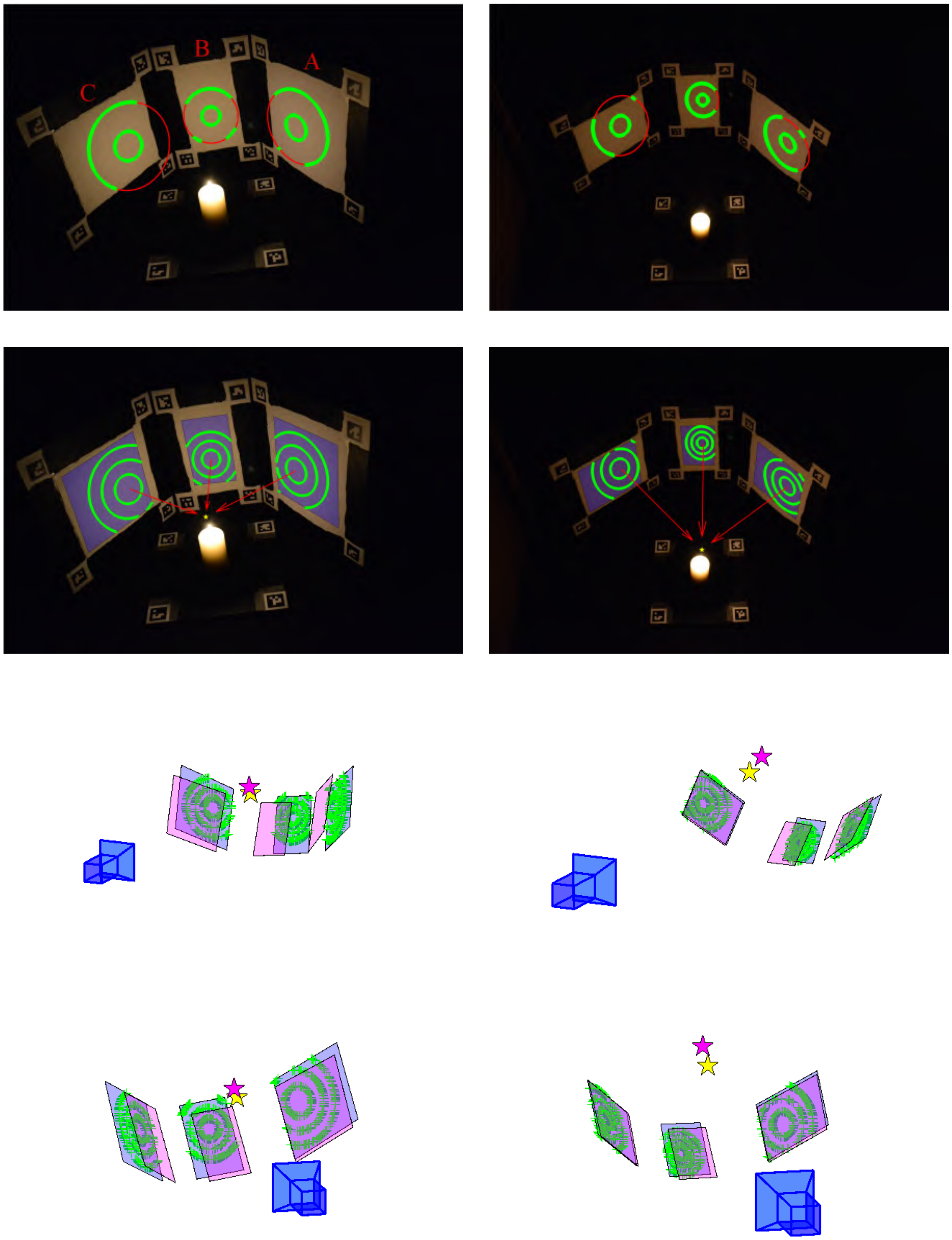

Figure 8: Real data experiments in lab conditions. Example images are shown with detected isophotes and reconstruction. The reconstructed planes are in blue with the detected isophote pre-images in green. The reconstructed PLS is the yellow star. The groundtruth planes and PLS are in magenta. The left and right columns follow the close and far setups respectively.

Third row, optimising a perspective transformation for each board individually with the photometric criterion. This case uses our top-down detection method. We observe that the branches disappear for all boards, thus giving more consistent observations and clearly showing the benefit of photometric refinement.

Fourth row, optimising all parameters, including the plane poses. We observe that for boards A, B and C, there is no significant change in the trend, indicating that the model fits the observed data well. However the pose has been adjusted and now deviate slightly from the ground-truth, this can be seen later in figure 8. It indicates that our model deviates from the observed reality, but this does not prevent it from adapting to it.

	Board C			Board B			Board A		
	Avg.	Avg. Abs.	Stdev	Avg.	Avg. Abs.	Stdev	Avg.	Avg. Abs.	Stdev
PLS only	0.0386	2.1036	2.7598	-0.0620	1.7380	2.2497	0.0082	2.1365	2.7863
Indiv. Persp.	0.0629	2.2177	2.8612	0.1473	1.5858	2.0344	0.0875	1.7908	2.3190
Global Pose	0.0299	1.7945	2.3260	0.0248	1.5404	1.9877	0.0270	1.7757	2.3046

Table 3: Statistics from model fitting for model validation. The three fitted models are shown as the three rows and described in the main text. For each board, 'Avg.' is the average of the algebraic distance, 'Avg. Abs.' is the average of the absolute distance and 'Stdev' is the standard deviation.

8.2 Pose Estimation

We experimentally evaluate our pose estimation methods in configuration H, which is the most complete one, with three types of experiments. We start with synthetic data, to validate against a great variety of poses and against varying image noise. We follow with lab-condition real images. We finish with natural real images, which validate the possible use of our methods in cases where previous methods fail.

8.2.1 Synthetic Data Experiments

Data generation, varied parameters and measured errors. We simulate a synthetic scene with parameters shown in figure 10. It is composed of a PLS and two planes with 1 m side length, placed 5 m away from the camera. We use Blender to achieve realistic rendering, with image intensities in $[0, 255]$, using hard thresholding to simulate saturation, and add additive Gaussian noise to the rendered images. The planes are fully visible with the default camera using 35 mm focal length and HD images. In order to avoid degenerate cases, we fix an angle of 30° between the principal line and the line joining the PLS and the planes connection point. We independently vary three key parameters: the angle between the planes (in red on the figure), the distance between the planes and the PLS source (in green on the figure) and the noise magnitude, whose default values are respectively 90°, 1 m and 1 unit. We compute statistics from 10 samples for each parameter setting, offsetting the plane connection point randomly from a uniform distribution within $[-0.1, 0.1]$ m. Some images of synthetic data are shown in figure 12. In these experiments, the isophote detection is based on the naive bottom-up approach which, recall, consist in selecting isopoints directly based on their intensity, and the global refinement is based on the geometric criterion. The impact of the top-down approach and the photometric criterion are studied at the end of the section. In order to make metric error measurements, we rescale the reconstruction to match the true reconstruction scale by fixing the distance between the camera centre and the estimated PLS to its real value with all other metrics scaled accordingly to this reference. We break down the reconstruction error as the following three measurements:

- The plane position error, which is the absolute difference in m between the true and estimated distances to camera.
- The plane orientation error, which is the angle in degrees between the true and estimated plane normals.
- The PLS position error, which is the distance in m between the true and estimated PLS.

Varying the between-plane angle. We vary the angle between the two scene planes, shown in red in figure 10, between 15° and 160°. The results are shown in figure 11. The figure shows the errors and a measurement of the average isophote ellipse's eccentricity and visibility. The individual samples are in blue and the median curve is in red. The errors are reasonably low in most settings. More precisely, with an angle chosen between 40° and 120°, the plane position error is lower than 0.05 m and the plane orientation error is lower than 0.5°. The method is, as expected, less accurate when the detection of the isophotes increases in difficulty. This happens when the BP moves close to the boundary of the surface: for an angle greater than 120°, the isophote visibility rate decreases, causing the error to rise. This also happens when the eccentricity of the ellipses increases, which occurs for angles lower than 40°.

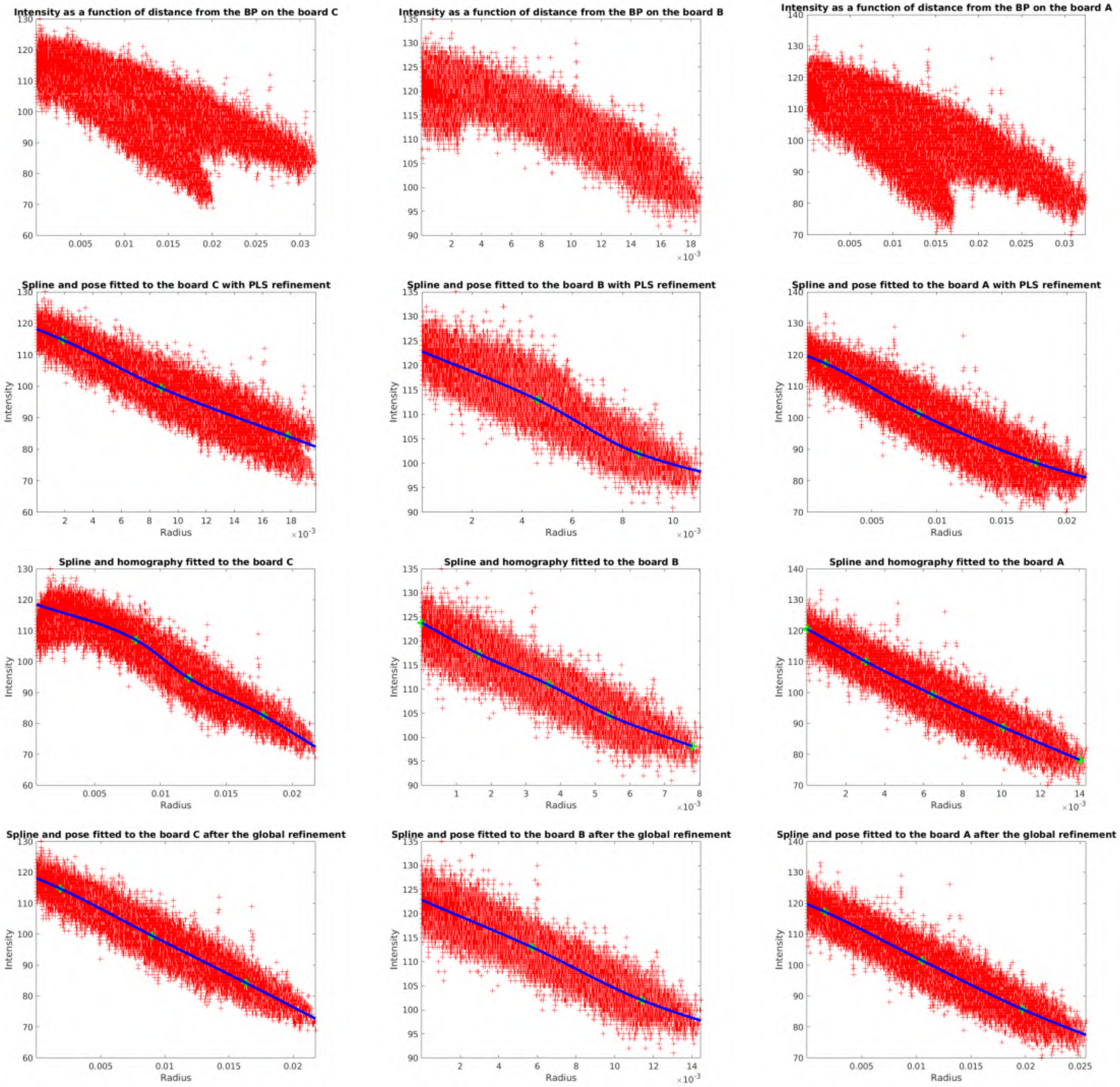

Figure 9: Radial intensity profile for the three boards C, B and A from left to right. Each red dot is a pixel, the green dots are the control points of the spline model shown in blue. Different rectifications are shown from top to bottom, with radii calculated respectively as ground-truth distance to BP, optimal PLS with ground-truth plane poses, optimal perspectivities fitted individually and optimal global pose following the photometric criterion.

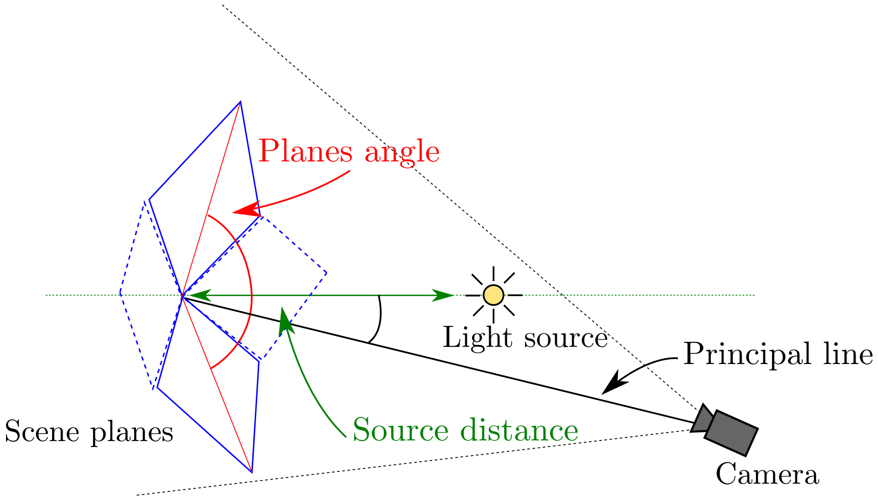

Figure 10: Synthetic scene generation parameters.

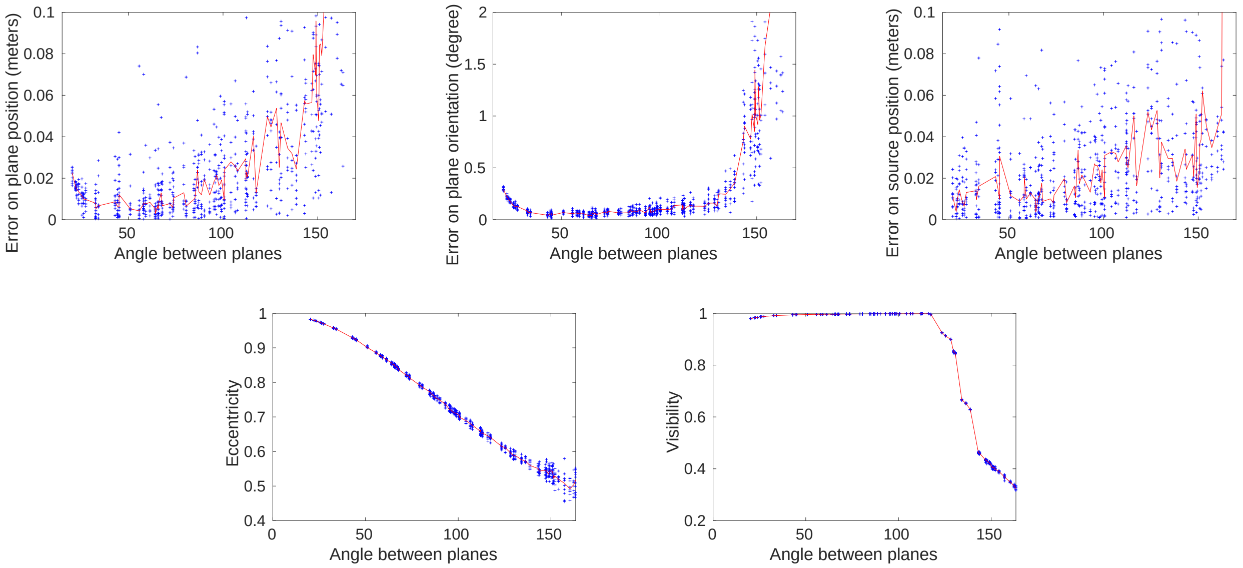

Figure 11: Evaluation with varying between-plane angle in synthetic data.

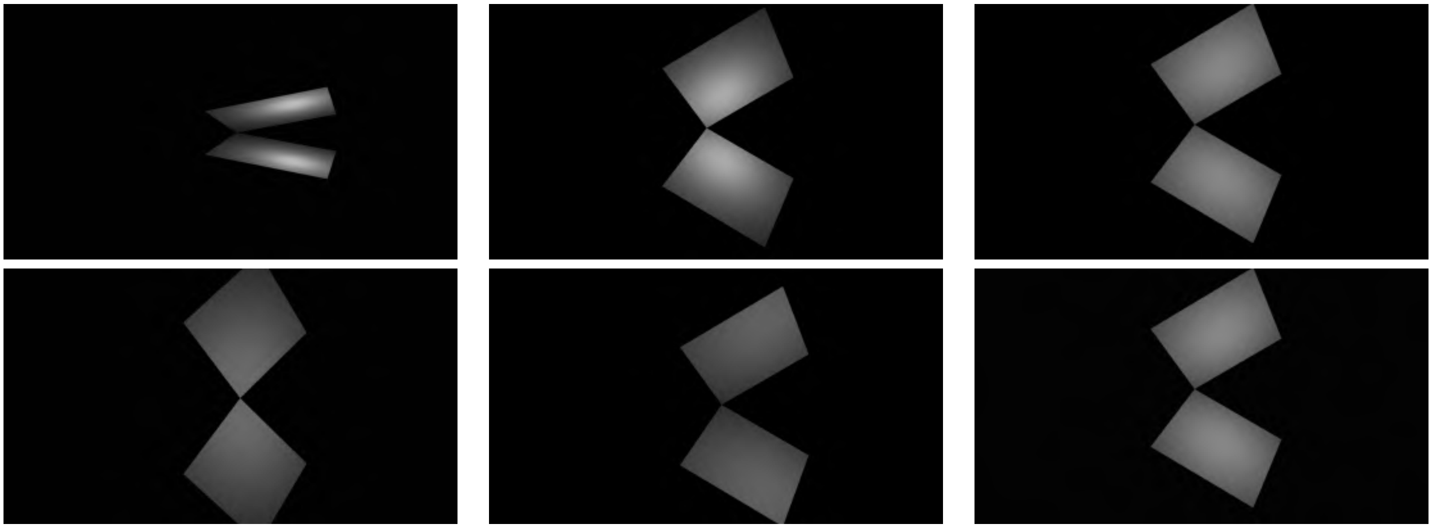

Figure 12: Representative synthetic images, with extreme values for the simulation parameters. Left: minimal (top) and maximal (bottom) inter-plane angle. Middle: minimal (top) and maximal (bottom) PLS to plane distance. Right: minimal (top) and maximal (bottom) noise level.

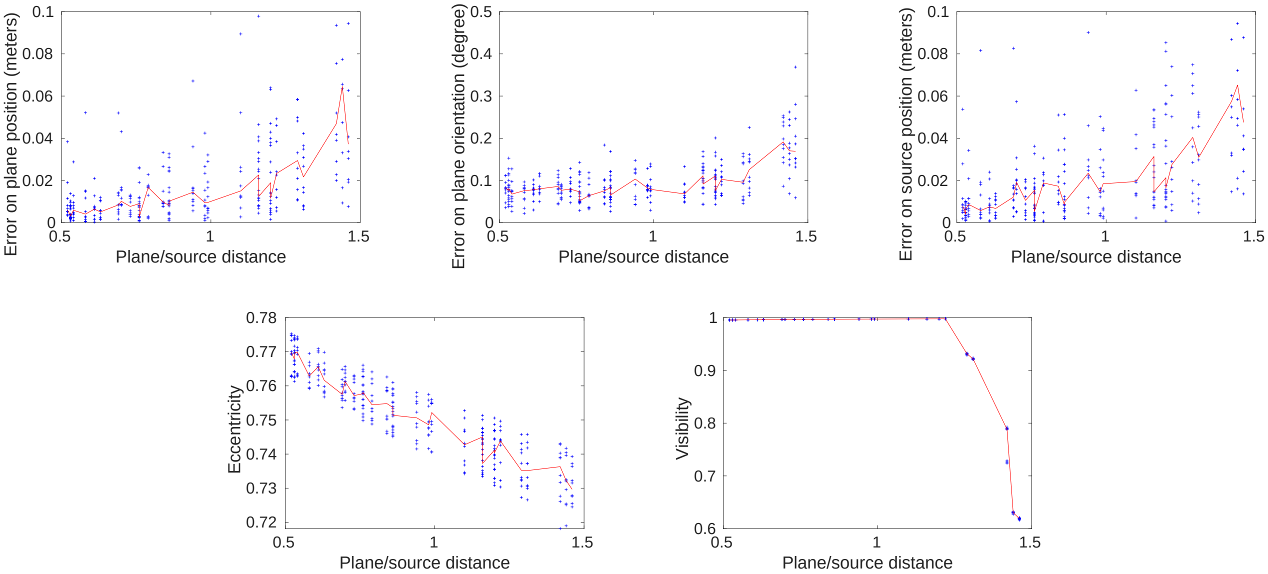

Figure 13: Evaluation with varying source-plane distance in synthetic data.

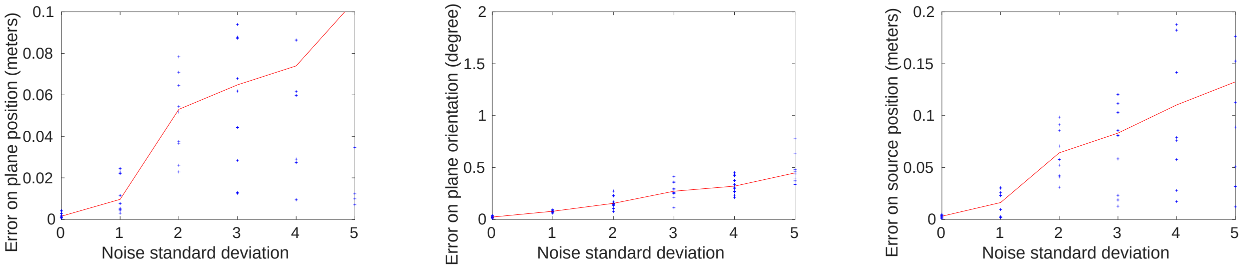

Figure 14: Evaluation with varying image intensity noise magnitude in synthetic data.

Varying the distance between plane and light source. We vary the angle between the two scene planes, shown in green in figure 10, between 0.5 m and 1.5 m. The results are shown in figure 13. The figure shows the errors and a measurement of the average isophote ellipse's eccentricity and visibility. The individual samples are in blue and the median curve is in red. We observe that, as for the angle variation, the error is reasonably low. Similarly as well, it is strongly correlated to the visibility of the isophote, which decreases when the distance increases or when the BP moves away from the visible parts of the planes.

Varying image noise magnitude. We vary the noise added to the rendered image intensities between 0 and 5. The results are shown in figure 14. The plane orientation is the most stable against increasing noise. This is because the plane normals are estimated independently of each other and of the distance and PLS position. The plane and PLS positions are more sensitive to noise; the error magnitude however remains lower than 15 cm for a noise magnitude of 5, which is reasonable given the scene size, lying at a distance of 5 m from the camera.

8.2.2 Impact of the Refinement

Table 4 shows the results obtained with variation of both isophote detection and reconstruction refinement methods on synthetic images. We observe that the bottom-up method outperforms the more advanced top-down method. We give two main reasons that explain this observation. The first reason is that the generated data follow the theoretical model; the only source of error is thus the added Gaussian noise. Therefore, the image smoothing filter responds very well to this disturbance, facilitating an accurate detection of isophotes in the image. The second reason is that the detection by the top-down approach relies on local optimisation with naive initialisation. The solution may thus in some cases converge far from the global minimum, which is supported by the observed difference between the median and the mean.

	Scene planes orientation error (deg)			**PLS position error (cm)**		
Refinement	**None**	**Geometric**	**Photometric**	**None**	**Geometric**	**Photometric**
Bottom-up, mean	0.1325	0.1485	0.0335	0.6702	0.2125	0.0755
Top-down, mean	2.9010	2.9010	0.5401	9.7074	9.4469	6.9052
Bottom-up, median	0.1092	0.1324	0.0328	0.6119	0.1646	0.0718
Top-down, median	1.8288	1.8288	0.0328	1.1907	1.0007	0.0727

Table 4: Reconstruction errors measured in synthetic data for combinations of isophote detection (rows) and refinement methods (columns).

8.2.3 Real Data in Lab Conditions

We show results for experiments with real images using the general setup described in section 8.1. We use two specific setups: the *close setup*, where the candle is about 20 cm away from the planes, and the *far setup*, where it is about 40 cm away. Figure 8 shows qualitative results for both setups. We observed that the naive bottom-up isophote detection method generally fails and used the advanced top-down method. The first row of the figure shows the resulting isophotes. The second row shows the estimated normals and the resulting estimated PLS. Finally, the third and fourth rows show two views of the reconstructed scene. We took 20 images of each setup to compute error statistics.

	Scene planes						**PLS**
	Orientation error (deg)			**Position error (m)**			**Position error (m)**
Board	C	B	A	C	B	A	Candle
Close setup, mean	09.26	07.61	08.72	00.20	00.17	00.18	00.02
Far setup, mean	04.39	19.11	09.27	00.14	00.63	00.34	00.07
Close setup, median	09.28	06.67	08.64	00.19	00.13	00.17	00.02
Far setup, median	04.58	14.33	07.54	00.07	00.57	00.17	00.03

Table 5: Reconstruction errors measured in real data taken in lab conditions with groundtruth.

We observe from table 5 that the mean and median errors are reasonably close to each other in most cases, indicating the absence of gross errors. The plane orientations are estimated with an error below 15° for the close setup and for boards B and C in the far setup. As the plane positions depend on the full reconstruction, their estimates are not as consistent, as a small orientation error can create an important position offset. This occurred with board B in the far setup, where the error raised beyond 50 cm. This is because board B is smaller and the intensity variation is thus narrower than for boards A and C in the images. In spite of this, we observe that the estimated PLS remains relatively accurate, with an error lower than 2 cm in the close setup and lower than 7 cm in the far setup. This is a sound achievement, given that the light source is a flame with diameter lower than 2 cm. These results were obtained from our complete pipeline including nonlinear refinement.

8.2.4 Impact of the Refinement

We measure the impact of the refinement in Table 6. The photometric criterion is the most effective and reduces the orientation error of the scene planes by 30% while reducing the average error on the PLS by a factor of 7. The geometric criterion does not improve the plane orientations but reduces the error on the PLS. This was expected as both criteria consider all parameters globally, which improves the PLS localisation.

	Scene planes orientation error (deg)			**PLS position error (cm)**		
Refinement	**None**	**Geometric**	**Photometric**	**None**	**Geometric**	**Photometric**
Top-down, mean	11.8437	11.7718	8.6020	12.6062	1.8669	1.7617
Top-down, median	12.0342	11.870	8.3739	11.9670	1.5486	1.6800

Table 6: Measured errors on real images in lab conditions with respect to groundtruth depending on isophote detection and reconstruction refinement criteria.

8.2.5 Real Data in the Endoscopic Configuration

We evaluate our model in the special case where the source is colocalised with the camera, which is configuration G^* in table 2. This configuration naturally appears when working with endoscopic cameras. As the light is emitted by the extremity of the device and forms a halo around the optics, the light source and the optical centre can be reasonably approximated by a single point. A scene captured by an endoscopic camera therefore comes close to the proposed model. We can however point out some deviations, the impact of which will be discussed in the results. First, the light coming from an optical fibre, the direction of the rays is not isotropic but rather concentrated along the optical axis, which makes the model closer to a spotlight than a PLS. Second, the image is processed internally by the endoscope hardware with embedded filters, which affect the intensity balance. This breaks our assumptions that the intensity received by a pixel, which is a camera sensor unit, does only depend on the radiance of the incoming light rays. Our evaluation setup consists of white boards placed in a box to represent flat surfaces and made several acquisitions with a Karl Storz medical endoscope. Figure 15 shows the qualitative results obtained on two example images with respectively 3 and 2 white boards.

Table 7 sums up the results obtained on the 6 images with various poses of the boards. The average angle error on the planes is 16˚. The reconstruction is thus not highly accurate ; it however gives a consistent reconstruction of the scene. Given the deviation of the observed case from our model this result is not surprising. We observe that the reconstructed normals are all more fronto-parallel than the groundtruth. This is because, as shown in the right image from figure 15, while the BP should be under the camera in a non visible part of the image, the spotlight effect moves it close to the image centre, skewing the reconstructed normals. We also notice that in the colocalised configuration, a closed contour isophote is less likely to appear, as it would require that the visible part of the plane face the camera orthogonally. Yet, as shown in the synthetic data experiments, open isophotes rapidly decrease reconstruction accuracy. The effect of the internal filters also certainly add up to the reconstruction errors. As their implementation details are unknown to us, we cannot however study their precise effects. Generally speaking, the results are very encouraging and validate the proposed model.

	Scene planes orientation error (deg)
Median	16.22°
Mean	18.63°
Minimum	1.05°
Maximum	36.61°

Table 7: Measured orientation error with the endoscopic images.

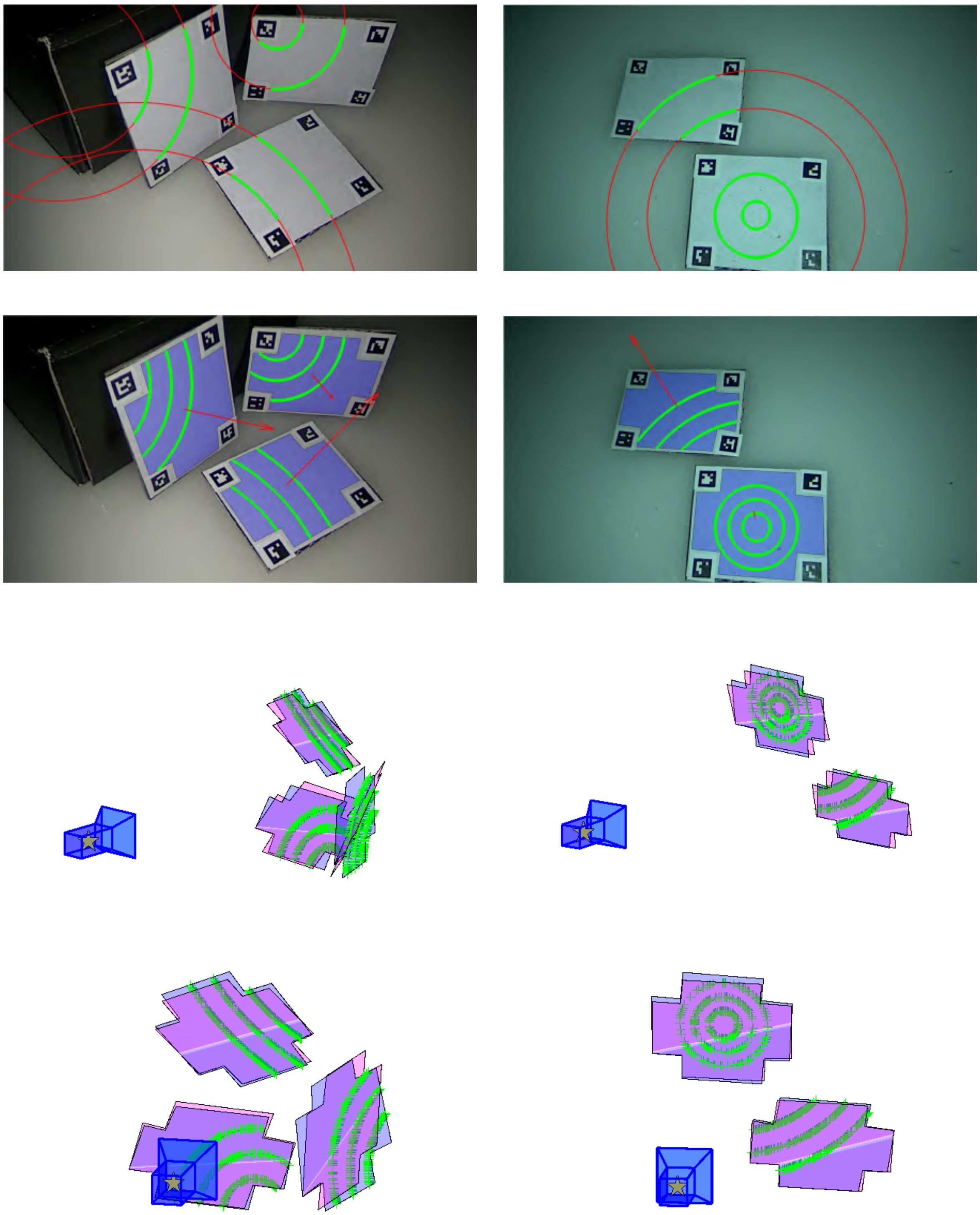

Figure 15: Images captured in the endoscopic configuration, with the detected isophotes (first row), the reconstructed plane normals (second row) and two views of the reconstruction. The reconstructed planes are in blue with the detected isophote pre-images in green. The groundtruth planes are in magenta.

8.2.6 Real Data in Uncontrolled Conditions

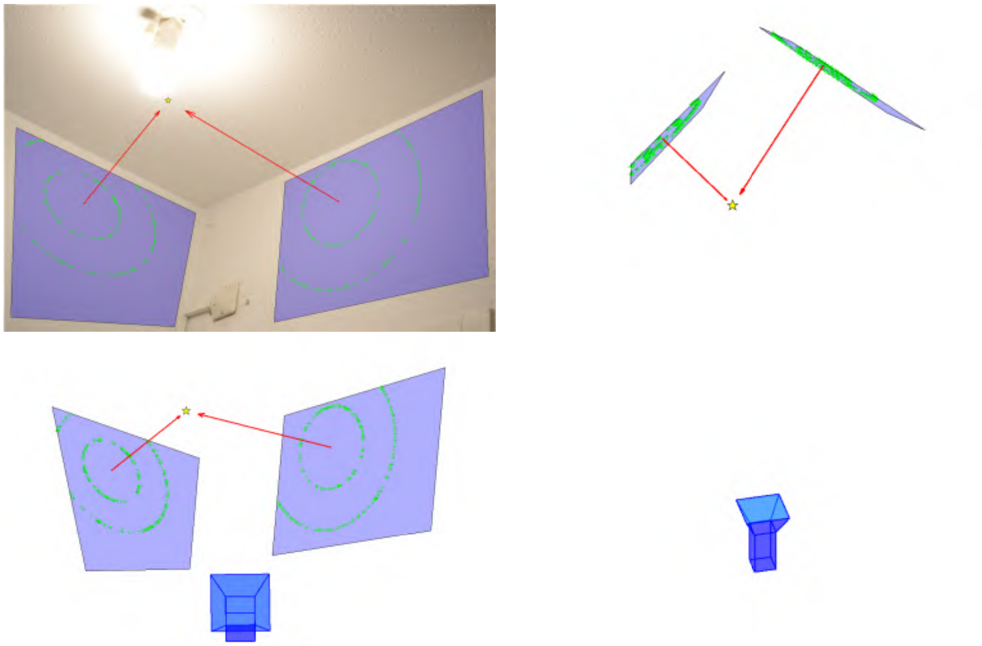

Figure 16: Real data experiments in uncontrolled conditions, qualitative results. The estimated light source position is shown by the yellow star and the scene planes as blue quadrilaterals.

Figure 16 now shows an example of the use of our method under uncontrolled conditions. The estimated light source position seems accurate, as it reprojects reasonably close to the light bulb. The angle between the two reconstructed scene planes is measured as 79°, which gives an error of 11°, given that the walls are supposedly perpendicular. This is a satisfying result, given that our input data are formed of a single image, and that both the light source position and 3D scene structure are estimated. The proposed model uses strongly simplifying assumptions of the real physical process; this result represents a positive validation, albeit preliminary, on real-world data. Existing methods do not directly cope with this type of data.

8.3 Baseline Comparison

We have used two baseline methods to compare with, using 20 real images acquired in lab condition with accurate groundtruth. An example image is shown in figure 8. For both methods, we use an external implementation to infer the normal map. We then fit planes using linear least-squares, averaging the per-pixel normals to retrieve a single normal over each planar region. We have chosen two methods representing the state of the art of classical and neural methods respectively. The first method is an optimisation-based advanced method extending Shape-from-Shading named SIRFS [Barron and Malik, 2014]. The second method is a neural method with uncertainty reasoning which we denote as CNN [Bae et al., 2021].

We ran both methods on the 20 images and evaluate their performance against groundtruth. We report statistics in table 8 and a visual example in figure 17 for the image shown in the left column of figure 8. These results can be directly compared with the results obtained for the proposed method in table 5. We observe that the mean and median statistics are close and consistent for all methods. Roughly speaking, while the proposed method's orientation errors are lower than 10 degrees for all planes, SIRFS' orientation errors are about 60 degrees and CNN's orientation errors are about 30 degrees. SIRFS' subpar performance may be explained by the assumptions it relies on. In particular, it assumes constant directional lights in the scene, modelled by spherical harmonics. While the proposed model explains image intensity using the change of the light's incident direction, SIRFS explains it using the variation of the scene normal, which is less adapted to the type of setup used in this experiment. Consequently, SIRFS fails to estimate the normals, leading to invalid scene plane poses. In contrast, CNN performs surprisingly well given its genericity. This may be explained by its training on indoor images and its uncertainty reasoning mechanism, discarding the pixels at which it is least confident of the normal estimate. In addition, recall that our implementation combines the normals from all pixels of each plane, which certainly has a beneficial effect in averaging out some of the estimation noise. Overall, we thus observe that the proposed method, using a detailed geometric model of the scene and optimisation-based reconstruction, outperforms general purpose methods, as expected.

Board	C	B	A
SIRFS, mean	75.19	50.21	58.21
SIRFS, median	80.92	46.65	55.79
CNN, mean	42.81	35.12	26.52
CNN, median	37.49	34.53	28.18

Table 8: Orientation reconstruction errors (deg) for the baseline methods measured in real data taken in lab conditions with groundtruth.

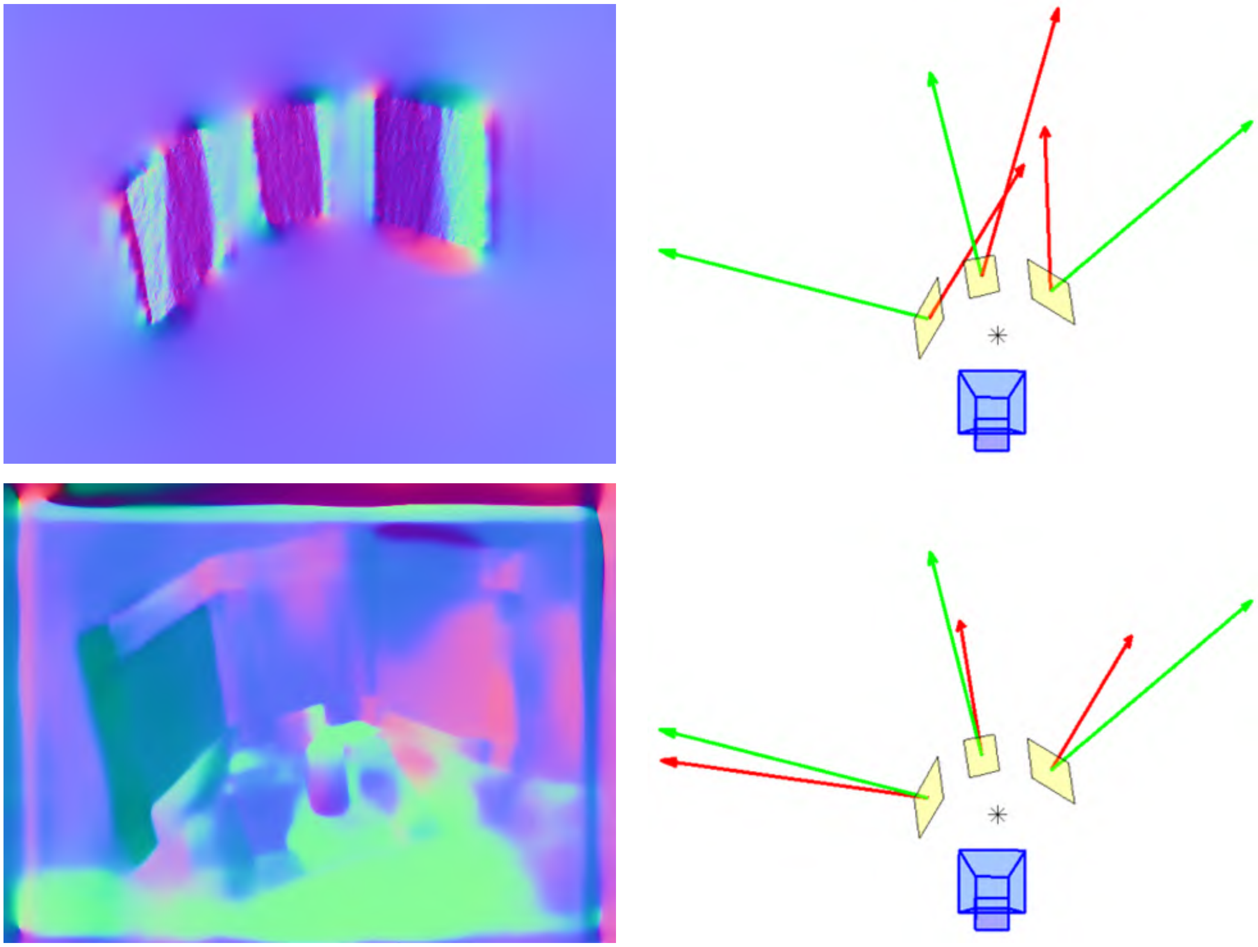

Figure 17: Reconstruction results for the baseline methods SIRFS (left) and CNN (right) on real data taken in lab conditions with groundtruth. The left column shows the normal maps and the right column shows the estimated and groundtruth normals in red and green respectively, along with the camera, light source and scene planes.

9 Conclusion

We have introduced a new framework to exploit reflectance image-based via geometric primitives. These primitives are isophotes, which are image curves formed of points of equal intensity. The framework is general and we have used the typical Lambertian reflection assumption from SfS to instantiate it concretely for a Point-Light Source, a piecewise planar scene and a calibrated perspective camera. We have shown that explicit image isophote curves could be established from the physics-based model, which specifically are conic sections, and that their pre-images on the scene are concentric circles, whose radii depend on the image intensity. We have proposed an extensive study of the different scene configurations and a complete closed-form solution to reconstruct the 3D scene elements, namely the plane poses and the Point-Light Source, from isophotes detected in a single image. The concept of using geometric primitives to model image intensity is promising, both from a theoretical standpoint, as it opens new geometric problems, allows one to study problem well-posedness, and, from a practical standpoint, as it allows one to resolve scene reconstruction in setups which cannot be handled by previous methods. A major strength of the proposed approach is that it can extract geometric constraints from the shading cues without the need to calibrate the camera radiometrically or to estimate the complex radiometric parameters of the scene.

Leads for future work include two main directions. First, we have focused on a specific instance of the proposed framework but many other ones can be studied. These instances are obtained by changing the reflection model (for instance, using a specular one), the light model (for instance, using a spot-light one) and the surface model (for instance, using a B-spline one). For each model, one should find out whether the isophotes can be reliably detected, whether they allow one to reconstruct the geometric parameters and find computational methods for doing so. Second, we have found that 3D reconstruction in medical endoscopy could form a particularly interesting application of our framework, for the light source is well approximated by a point or a spot-light located at the camera centre. Endoscopic images are typically rich in specularities appearing as white dots owing to camera saturation, which can thus be reliably detected.

Acknowledgments. This work was funded by the FET-Open grant 863146 Endomapper.

References

A. D. Aleksandrov, M. A. Lavrent'ev, et al. *Mathematics: its content, methods and meaning.* Courier Corporation, 1999.

G. Bae, I. Budvytis, and R. Cipolla. Estimating and exploiting the aleatoric uncertainty in surface normal estimation. In *ICCV*, 2021.

J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014.

H. Barrow, J. Tenenbaum, A. Hanson, and E. Riseman. Recovering intrinsic scene characteristics. *Comput. Vis. Syst*, 2(3-26):2, 1978.

A. Bartoli. The highlight ovals. *Journal of Mathematical Imaging and Vision*, 61(7):919–943, 2019.

A. S. Baslamisli, P. Das, H.-A. Le, S. Karaoglu, and T. Gevers. Shadingnet: image intrinsics by fine-grained shading decomposition. *International Journal of Computer Vision*, pages 1–29, 2021.

M. Breuß and Y. C. Ju. Shape from shading with specular highlights: Analysis of the phong model. In *2011 18th IEEE International Conference on Image Processing*, pages 9–12. IEEE, 2011.

L. Calvet, P. Gurdjos, C. Griwodz, and S. Gasparini. Detection and accurate localization of circular fiducials under highly challenging conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 562–570, 2016.

T. Collins and A. Bartoli. Infinitesimal plane-based pose estimation. *International journal of computer vision*, 109(3): 252–286, 2014.

F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos. 3d-spline reconstruction using shape from shading: Spline from shading. *Image and Vision Computing*, 26(4):466–479, 2008.

M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (slam) problem. *IEEE Transactions on robotics and automation*, 17(3):229–241, 2001.

V. Dragnea and E. Angelopoulou. Direct shape from isophotes. In *Proceedings of the ISPRS Workshop Ben-COS05*, pages 45–50, 2005.

D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision*, pages 2650–2658, 2015.

E. Garces, C. Rodriguez-Pardo, D. Casas, and J. Lopez-Moreno. A survey on intrinsic images: Delving deep into lambert and beyond. *International Journal of Computer Vision*, 130(3):836–868, 2022.

C. Godard, O. Mac Aodha, and G. J. Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 270–279, 2017.

P. Gurdjos, P. Sturm, and Y. Wu. Euclidean structure from $n \geq 2$ parallel circles: theory and algorithms. In *European Conference on Computer Vision*, pages 238–252. Springer, 2006.

R. Hartley and A. Zisserman. Multiple view geometry in computer vision, cambridge uni. *Pr., Cambridge, UK*, 1:2, 2000.

G. Healey and T. O. Binford. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, 42 (1):62–86, 1988.

B. K. Horn. Obtaining shape from shading information. *The psychology of computer vision*, pages 115–155, 1975.

H. Kato, Y. Ushiku, and T. Harada. Neural 3d mesh renderer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3907–3916, 2018.

J.-S. Kim, P. Gurdjos, and I.-S. Kweon. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4): 637–642, 2005.

E. H. Land. Recent advances in retinex theory. *Central and peripheral mechanisms of colour vision*, pages 5–17, 1985.

Y. Ma, X. Feng, X. Jiang, Z. Xia, and J. Peng. Intrinsic image decomposition: A comprehensive review. In *International Conference on Image and Graphics*, pages 626–638. Springer, 2017.

D. Mariyanayagam, P. Gurdjos, S. Chambon, F. Brunet, and V. Charvillat. Pose estimation of a single circle using default intrinsic calibration. In *Asian Conference on Computer Vision*, pages 575–589. Springer, 2018.

S. R. Marschner. *Inverse rendering for computer graphics.* Cornell University, 1998.

R. Modrzejewski, T. Collins, A. Hostettler, J. Marescaux, and A. Bartoli. Light modelling and calibration in laparoscopy. *International journal of computer assisted radiology and surgery*, 15(5):859–866, 2020.

A. Morgand, M. Tamaazousti, and A. Bartoli. A geometric model for specularity prediction on planar surfaces with multiple light sources. *IEEE transactions on visualization and computer graphics*, 24(5):1691–1704, 2017.

T. Narihira, M. Maire, and S. X. Yu. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *Proceedings of the IEEE international conference on computer vision*, pages 2992–2992, 2015.

T. Okatani and K. Deguchi. Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer vision and image understanding*, 66(2):119–131, 1997.

T. Okatani and K. Deguchi. Closed form solution of local shape from shading at critical points. *International Journal of Computer Vision*, 40(2):169–178, 2000.

F. Petersen, A. H. Bermano, O. Deussen, and D. Cohen-Or. Pix2vex: Image-to-geometry reconstruction using a smooth differentiable renderer. *arXiv preprint arXiv:1903.11149*, 2019.

E. Prados and O. D. Faugeras. "perspective shape from shading" and viscosity solutions. In *ICCV*, 2003.

Y. Quéau, J. Mélou, F. Castan, D. Cremers, and J.-D. Durou. A variational approach to shape-from-shading under natural illumination. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 342–357. Springer, 2017.

L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7. IEEE, 2008.

Z. L. Szpak, W. Chojnacki, and A. van den Hengel. Guaranteed ellipse fitting with a confidence region and an uncertainty measure for centre, axes, and orientation. *Journal of Mathematical Imaging and Vision*, 52(2):173–199, 2015.

M. Tappen, W. Freeman, and E. Adelson. Recovering intrinsic images from a single image. *Advances in neural information processing systems*, 15, 2002.

S. Tozza and M. Falcone. Analysis and approximation of some shape-from-shading models for non-lambertian surfaces. *Journal of mathematical imaging and vision*, 55(2):153–178, 2016.

S. Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, 1979.

M. Visentini-Scarzanella, D. Stoyanov, and G.-Z. Yang. Metric depth recovery from monocular images using shape-from-shading and specularities. In *2012 19th IEEE International Conference on Image Processing*, pages 25–28. IEEE, 2012.

G. Wolberg and I. Alfy. Monotonic cubic spline interpolation. In *Computer Graphics International*, pages 188–195, 1999.

R. J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image understanding systems and industrial applications I*, volume 155, pages 136–143. SPIE, 1979.

Y. Yu and W. A. Smith. Inverserendernet: Learning single image inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2019.