

# The Radiometry of Multiple Images

Q.-Tuan Luong<sup>†</sup>, Pascal Fua<sup>\*</sup>, and Yvan Leclerc<sup>†</sup>

<sup>†</sup> AI Center

<sup>\*</sup> DI-LIG

SRI International

EPFL

CA-94025 Menlo Park

CH-1015 Lausanne

USA

Switzerland

{luong,leclerc}@ai.sri.com    pascal.fua@epfl.ch

In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):19–33, January 2002.

## Abstract

We introduce a methodology for *radiometric reconstruction*, the simultaneous recovery of multiple illuminants and surface albedoes from multiple views, assuming that the geometry of the scene and of the cameras is known. We formulate the linear theory of multiple illuminants and show its similarities with the theory of geometric recovery of multiple views. Linear and non-linear implementations are proposed; simulation results are discussed; and, finally, results on real images are presented.

**keywords** reconstruction, multiple views, linear models, radiometric recovery, Lambertian reflection, point light sources, ambient illumination, surface albedo

## 1 Introduction

An important goal in computer vision is to be able to reconstruct the geometric and radiometric properties of a scene. In this paper, we assume that the geometry of the scene and of the cameras is known or has been recovered, and show how to recover simultaneously multiple illuminants and reflectances, a task that we designate as the *radiometric reconstruction* problem. The main insight of this work is that the radiometry can be reconstructed using multiple images, each of them taken with a different unknown illuminant, in much the same way as the geometry of a rigid scene can be reconstructed using multiple images with unknown but different viewpoints.

The image intensity measured at a given pixel depends on four world components: Illumination, reflectance, scene geometry, and camera geometry. If the geometric and radiometric variables are all unknown, the problem of recovering them clearly becomes under-constrained, no matter how many views are available. Here, we propose a paradigm that decouples *radiometric reconstruction* from *geometric reconstruction*. The latter involves determining both the geometry of the surfaces and of the cameras from multiple images of the scene. It is well understood, both in theory (see [6, 14] for recent accounts) and in practice (see [34] for a recent survey). Given the geometry, we infer all the radiometric

variables: varying albedoes, and one distinct point light source and ambient illumination for each image. Compared to the photometric stereo approaches, there is no need to use a constant viewpoint, thus allowing us to take advantage of existing solutions for the geometric reconstruction problem. These solutions have proven to be reliable and applicable in a variety of unconstrained environments. Then, in a second stage, radiometric parameters that are more sensitive are recovered using our proposed method. We do not require prior knowledge about either reflectance or illumination.

The decoupling of geometry and radiometry that we advocate makes sense for many applications. For example, this work was initially motivated by the need to perform change detection using images of the same geographic sites taken at different times of day. In such a case, digital terrain maps are available or can be computed. It then becomes natural to combine the maps and the actual images to estimate the illumination parameters, recover surface albedoes and check that they remain constant over time. A similar situation arises when constructing 3-D models of objects or people from video sequences for rendering or animation purposes. The illumination may vary from image to image, for example because the object or subject being modelled moves with respect to fixed illuminants. In this situation, it is often possible to use structure-from-motion techniques that are relatively insensitive to illumination changes to reconstruct the geometry, as will be shown in the result section of this paper. The next step is then to attenuate the lighting effects in the original images to create high-quality texture maps. More generally, provided that the images do not exhibit extreme lighting variations, geometric reconstruction techniques, such as stereo based on normalized cross-correlation [9], can be made to handle illumination effects. The only required prior knowledge is the internal geometric calibration of the cameras, which in most circumstances can be obtained using either photogrammetric or self-calibration techniques.

Previous work is summarized in table 1 and discussed in detail in the next section. It falls into two broad categories, depending on whether or not the geometry is known *a priori*. Approaches based on photometric stereo require that multiple images be taken from a fixed point of view with multiple illuminants, and aim at recovering the shape as well as the reflectance. In addition, to obtain an unambiguous solution, these methods require knowledge of the illuminants or of parts of the geometry. These approaches all rely on a Lambertian model with distant point light sources, and have been found to be sensitive to modeling deviations [8]. Inverse methods based on more sophisticated illumination and reflectance models require the knowledge of the geometry, and cannot recover simultaneously the reflectance and illuminant in a general situation. At least one of the two needs to be known accurately. Because the required prior measurements are difficult to obtain outside of a controlled environment, the applicability of all these methods has been limited.

To the best of our knowledge, no general algorithm or even formal framework for radiometric recovery from multiple views and multiple illuminants exists. The goal of this paper is to introduce such a framework, to propose algorithms to implement it, and to give experimental results. As a first step towards the general problem of radiometric reconstruction, we will investigate the particular case of Lambertian surfaces, which we believe will be the foundation for work with more general models.

In Section 2, we discuss previous work. While the fact that pixel intensities are in a bilinear relation with surface albedo and illumination radiances is well known, it has not been used before in the context of multiple illuminants and multiple views. In Section 3, we describe the linear radiometry of multiple illuminants, and compare this problem with the geometry of multiple views. Based on these parallels, we give a characterization of the minimal data required, and then propose in Section 4 a linear algorithm for the recovery of the radiometry of multiple views. Simulation shows this linear method to be quite sensitive to measurement noise. To obtain a more robust solution, we turn in Section 5 to a non-linear

method, which is the radiometric counterpart of the bundle adjustment approach to geometric recovery. Section 6 discusses limitations of the current system and areas for future developments.

## 2 Previous work

In the shape from shading paradigm ([19] includes many of the classical papers on the subject), it is assumed that the reflectance is Lambertian and uniform. This makes it possible to infer surface shape, as well as the direction of the illuminant from a single image [38] or from multiple images [28].

The main limitation of shape from shading is that reflectance has to be uniform or known. In the photometric stereo paradigm, by using multiple images, each taken with the same viewpoint, but a different illuminant, it is possible to recover a non-uniform albedo, as well as the surface orientation. In early work [33, 35, 18] illuminants needed to be known. A linear method to recover Lambertian surface albedo from three images, each taken with a different point light source, has been proposed [1]. Instead of relying on an explicit knowledge of the three illuminants, this method, in the spirit of the shape from shading methods, relies on the preliminary observation of an object of uniform reflectance to determine them.

Subsequently, it was shown [17] that even when the illuminants are unknown, the albedo and shape can still be recovered using SVD, which generalized photometric stereo. However, the price to pay is that both of albedo and shape are subject to an arbitrary linear transformation [17]. This was extended [37] to include an ambient illumination term in addition to the Lambertian reflectance, and the ambiguity result was corrected, resulting in a more accurate characterisation as a Generalized Bas-Relief ambiguity [3], which is intrinsic to the problem and can be entirely resolved only by sufficient prior knowledge about the object's geometry. If it is known that the object belongs to an instance of a class (such as human faces), the ambiguity can be resolved to a degree where novel synthetic images under different poses and lighting look realistic [12]. These methods are powerful and address a more difficult problem than we do. In particular, if a sufficient number of 3D measurement on the object are known, they can recover both the illuminant and reflectance. However their precision has not been documented, nor their applicability to unconstrained or large environments, where the requirement to vary the illumination while maintaining the camera at the same viewpoint can be difficult to meet.

The primary purpose of all the approaches previously cited was to recover the shape based on the Lambertian model. Radiometric parameters such as the illumination or the albedo were a byproduct. A second family of methods, more in the spirit of our approach, aims to recover radiometric parameters assuming that the camera and object geometry is available.

In inverse lighting [24, 25] it is assumed that the reflectance is Lambertian, and known or constant. The illuminant is modeled as a linear combination of a basis of predefined lights. There is a linear relationship between the coefficients of the linear combination and the pixel values, which makes it possible to recover the former from a single image.

With the dual approach, reflectance modeling, which consists in assuming that the point light source is known, it is possible to recover sophisticated reflectance properties with specular and diffuse parameters [20, 2, 31, 24]. A range imaging sensor was used by these author to recover the geometry. This has been extended to allow for multiple known point light sources, as well as indirect illumination effects, resulting in impressive re-renderings of controlled environments [36]. The geometry was modeled precisely by manual methods.

Framework	Camera geometry	Scene geometry	Reflectance	Illumination
Geometric reconstruction	recovered	recovered	reasonable variation	reasonable variation
Shape from shading	single	recovered	constant	recovered
Photometric stereo	constant	recovered	recovered	known
Generalized photometric stereo	constant	partly recovered	partly recovered	partly recovered
Inverse lighting	single	known	known or constant	recovered
Reflectance modeling	known	known	recovered	known
Radiometric reconstruction	known	known	recovered	recovered

Table 1: Radiometric reconstruction in relation with previous general frameworks

These approaches are able to recover rich models of either the illuminant or the reflectance, but unlike the method that we propose, they cannot recover both. One method which has this capability was recently introduced [30]. Using a linear relation between lighting coefficients and pixel values, reliable results were obtained. However this approach cannot be considered to be generic, as it depends on a specific situation where the shadow of an object of known shape is projected onto an object of known shape. It wouldn't be applicable in scenes where shadows cannot be measured with good accuracy.

In all the previously cited work, the cameras are assumed to be radiometrically calibrated. The non-linear radiometric transfer function can be recovered using multiple images consisting of different exposures of a fixed scene and illumination, when the exposures are known [4] or even unknown [26].

### 3 Linear Model for the Radiometry of Multiple Illuminants

In this section, we develop the theory for Lambertian reflectance models, which leads to a linear model.

#### 3.1 The Lambertian model

We model the effect of illumination as the sum of a Lambertian term from a punctual distant light source, and an ambient term accounting for mutual illumination and sky light effects. For a surface element, such as a triangular facet in a surface mesh, of normal  $\mathbf{n}$  and albedo  $\alpha$  that projects into an image  $i$  at coordinates  $(u, v)$ , the light intensity striking the photographic (or CCD) plate (or radiance at the image plane) is given by:

$$\mathcal{L}_i(u, v) = \alpha(\mathbf{l}_i^T \mathbf{n} + \mu_i) \quad (1)$$

where  $\mathcal{L}_i(u, v)$  represents the radiance around the projection of the surface element (for example averaged over the facet's projection),  $\mathbf{l}_i$  is the light source vector,  $\mu_i$  is the ambient light. Each surface element is characterized by 3 parameters, the direction of  $\mathbf{n}$ , and the albedo  $\alpha$ . While  $\mathbf{n}$  is a unit vector,  $\mathbf{l}_i$  is not: its magnitude encodes the light source's "magnitude." The complete illumination model for each image is therefore given in terms of 4 parameters:

- The three components  $l_{i1}, l_{i2}, l_{i3}$  of  $\mathbf{l}_i$ .
- The ambient light  $\mu_i$ .

These parameters are described by the *illumination vector*, the vector of four components:

$$\mathbf{L}_i = [l_{i1}, l_{i2}, l_{i3}, \mu_i]^T$$

Using the illumination vector, and the notation  $\mathbf{N} = \begin{bmatrix} \mathbf{n} \\ 1 \end{bmatrix}$ , Eq. (1) can be written more compactly as:

$$\mathcal{L}_i(u, v) = \alpha \mathbf{L}_i^T \mathbf{N}$$

### 3.2 Radiometric calibration

The measures that we can access are pixel gray-level values. However, the quantity that has a physical meaning is the radiance at the image plane. The pixel gray-level bears a certain relationship (typically monotonic) to the radiance. That is, for each camera  $i$ ,

$$I_i(u, v) = F_i(\mathcal{L}_i(u, v)) \quad (2)$$

where  $\mathcal{L}_i(u, v)$  is the light intensity striking the image plane of camera  $j$  at coordinate  $(u, v)$ ,  $F_i$  is an unknown radiometric transformation (including digitization and other scanning effects), and  $I_i(u, v)$  is the resulting pixel value. Note that all of this generalizes to wavelength-dependent measures, but we will describe the scalar case only here for simplicity of exposition. A camera  $i$  is said to be *radiometrically calibrated* when the function  $F_i$  is known.

Images are typically not radiometrically calibrated, and unless the calibration is done beforehand, there is rarely sufficient information (such as multiple exposures of a same scene [26]) to recover exactly the unknown functions  $F_i$  from them. However, in practice, a first approximation is obtained by describing  $F_i$  as an affine transformation specified with a *radiometric scale factor*  $a_i$  and *radiometric offset*  $b_i$ :

$$F_i(\mathcal{L}) = a_i \mathcal{L} + b_i \quad (3)$$

A general, and more precise description of  $F_i$  could be obtained by adding non-linear parametric terms.

This parallels the geometric calibration problem, where the relationship between the projections on the physical image plane and the coordinates in terms of pixel are in a first approximation considered to be linear, and described by the camera intrinsic parameters (focal, aspect-ratio, etc.). Adding a few non-linear *geometric distortion* terms is sufficient, except for the most exacting of applications, where non-parametric models of distortion must be used.

Another simplification arises when it is known that the camera is the same, that is its calibration does not vary. Indeed for the purpose of detecting change, it oftentimes suffices to remove the dependency of  $F$  on  $i$ :

$$I_i(u, v) = F(\mathcal{L}_i(u, v))$$

where  $F$  is now the same unknown transformation applied to the light intensity.

In this paper, we will use only the affine model so that we can maintain the linearity of the whole model. It will be shown that even within this limitation, there is enough information to recover one of the parameters, the radiometric offset. When using a non-linear approach within our radiometric multiple image methodology, a sufficiently large amount of data would allow to recover more non-linear parametric terms. We don't claim to investigate thoroughly the issue of radiometric (self) calibration, but will just illustrate in a simple case how it can be solved in the radiometric reconstruction problem, much the same way as self-calibration [23] has been solved in the geometric reconstruction problem.

### 3.3 Multiple images

The key idea of our approach is to use multiple images of the same surface, taken with different illuminations. We assume that the geometry of the surfaces and cameras has been previously recovered, which makes it possible to express all the 3-D spatial relationship in a common coordinate system. Because of the Lambertian hypothesis, the image radiance does not depend on the position and orientation of the cameras, and therefore all the measurements made with a common illumination can be gathered into a single “image”, provided that the radiometric calibration is taken into account.

For simplicity’s sake, in this paper, we restrict ourselves to the case where only one point light source is used at a time. Each image  $i$  is therefore characterized by its illumination vector  $\mathbf{L}_i$ . Besides the minimum requirement that there are two non-proportional illumination vectors, the values of these vectors do not need to be all different (see Section 4). In this context, the only relevant geometric information are the normals  $\mathbf{n}_j$ , of each surface element  $j$ , which we assume to have been determined. The only radiometric parameter for each surface element is its albedo  $\alpha_j$ .

We have  $p$  surface elements, with  $n$  different illuminants, which give rise to the  $p \times n$  radiance values  $\mathcal{L}_{ij}$  and pixel gray-level values  $I_{ij}$ , according to Eq. (1) and Eq. (3) :

$$I_{ij} = a_i \mathcal{L}_{ij} + b_i = a_i \alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i \quad (4)$$

Since there are a total number of  $p + 4 \times n + 2 \times n$  unknowns:

- albedoes  $\alpha_j, j = 1 \dots p$ ,
- illumination vectors  $\mathbf{L}_i, i = 1 \dots n$ ,
- radiometric calibration parameters  $a_i, b_i, i = 1 \dots n$ ,

and a total number of  $p \times n$  equations, with a sufficient number of different surface elements and illuminations, we expect to be able to solve the problem of radiometric reconstruction. Note that if all the albedoes  $\alpha_j$  are known (as in [25]) the problem becomes linear, instead of bilinear, and can be solved with a single view.

This formulation shows that the radiometric reconstruction problem is similar to the geometric reconstruction problem, where we are given  $p$  3-D points  $\mathbf{M}_j$  seen in  $n$  different views of projection matrices  $\mathcal{P}_i$ , and try to recover the  $\mathbf{M}_j$  and  $\mathcal{P}_i$  from the  $2n \times p$  equations  $\mathbf{m}_{ij} \simeq \mathcal{P}_i \mathbf{M}_j$ . It is known [5, 16] that for this problem the solution can in general, for uncalibrated cameras, be recovered only up to a projective transformation  $\mathcal{H}$ , since if the previous equations are satisfied, so is the equation:  $\mathbf{m}_{ij} \simeq (\mathcal{P}_i \mathcal{H})(\mathcal{H}^{-1} \mathbf{M}_j)$ . In the case of calibrated cameras,  $\mathcal{H}$  is an Euclidean transformation  $\mathbf{Q}_E$ , a displacement and scaling.

### 3.4 Ambiguities and minimal data

Going back to the radiometric reconstruction problem, we notice from Eq. (4) that if:

$$a_i, b_i, \alpha_j, \mathbf{L}_i$$

is a solution, then another acceptable solution is obtained for any scalar  $s > 0$  and  $k_i, i = 1 \dots n$

$$\frac{a_i}{k_i}, b_i, \frac{\alpha_j}{s}, s k_i \mathbf{L}_i$$

It can be verified that this is the only ambiguity which can hold generically. The ambiguity due to the multiplier  $s$  is a uniform scale ambiguity, which means that we cannot distinguish between brighter surfaces lit by a dimmer illuminant or darker surfaces lit by a more intense illuminant. This ambiguity is similar to the depth-speed uniform scale ambiguity encountered in the geometric reconstruction problem. The fact that only the positive values of the global scale  $s$  are acceptable comes from the physical constraint that the albedoes be positive. Note that this is similar to the constraint that the points have to lie in front of the camera, yielding a positive depth in the geometric reconstruction problem. In the case of radiometrically calibrated cameras, the scale ambiguity is the only generic ambiguity of the problem.

The ambiguity due to the multipliers  $k_i$  means that for cameras which are not photometrically calibrated, it is not possible to distinguish between the intensity of the illuminants and the radiometric scale factor of the camera without further assumptions. Because of this essential indetermination, in this case we could arbitrarily reduce the ambiguity by setting  $a_i = 1, i = 1 \dots n$ . One way to resolve this ambiguity would be to use more constraints on the set of images, such as the fact that some calibration parameters, illuminants, or both are identical. This is analogous to the constraints on camera parameters used for geometric self-calibration [23].

The ambiguity due to the uniform scale  $s$  reduces the number of significant unknowns by 1, while the ambiguity due the multipliers  $k_i$  reduces this number by  $n$  in the uncalibrated case. Therefore, the total number of unknowns in the problem is  $5 \times n + p - 1$  in the uncalibrated case, and  $4 \times n + p - 1$  in the calibrated case. From this count, we can calculate the minimal amount of data needed to have a unique solution. The fact that these numbers are sufficient in the generic case will be confirmed by specific examples described in the next section. In the calibrated case, we expect a unique solution when  $np \geq 4n + p - 1$ , which we rewrite as  $p \geq \frac{4n-1}{n-1}$ . Therefore:

- with 2 views, a unique solution is expected for 7 surface elements,
- with 3 views, a unique solution is expected for 6 surface elements,
- with 4 views and more, a unique solution is expected for 5 surface elements.

The table 2 summarizes the parallel between the geometric reconstruction problem and the radiometric reconstruction problem.

## 4 A Linear Solution

### 4.1 Theory

In the geometric reconstruction problem, it is known [21, 22] that, given two views, any 2-D projections corresponding to the same 3-D point satisfy an algebraic constraint  $\mathbf{m}_k^T \mathbf{F}_{kl} \mathbf{m}_l = 0$  which is linear with respect to a representation of the motion between the views, and bilinear with respect to the measures. This gives a linear algorithm for motion recovery. We are going to see in this section that the situation is similar in the radiometric reconstruction problem, but somewhat simpler.

	Geometric reconstruction	Radiometric reconstruction
model	3-D points $M_j = (X_j, Y_j, Z_j), j = 1 \dots p \quad (3p)$ projection matrices $\mathcal{P}_i, i = 1 \dots n \quad (11n)$	albedoes $\alpha_j, j = 1 \dots p \quad (p)$ illumination vectors: $\mathbf{L}_i = (l_{i1}, l_{i2}, l_{i3}, \mu_i), i = 1 \dots n \quad (4n)$
calibration	camera intrinsic parameters $\mathbf{m}_{pixel} = \mathbf{A}_i \hat{\mathbf{m}} \quad (5n)$	radiometric scale and offset $I = a_i \mathcal{L} + b_i \quad (2n)$
measurements	2-D points, pixel coordinates $m_{ij} = (x_{ij}, y_{ij}) \quad (2pn)$ calibrated: $\hat{\mathbf{m}}$ , image plane coordinates	surface orientations $\mathbf{N}_j = (n_{j1}, n_{j2}, n_{j3}, 1), \ \mathbf{n}\  = 1 \quad (2p)$ gray-level values $I_{ij} \quad (pn)$ calibrated: $\mathcal{L}_{ij}$ , image plane radiance
equations	pinhole model $\mathbf{m}_{ij} \simeq \mathcal{P}_i \mathbf{M}_j \quad (2pn)$	Lambertian model $I_{ij} = a_i \alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i \quad (pn)$
ambiguity	$\mathcal{P}_j \leftrightarrow \mathcal{P}_j \mathbf{Q}_P \mathbf{Q}_E$ $\mathbf{Q}_P; \mathbf{Q}_E \quad (8 \text{ uncalibrated} + 7)$	$\mathbf{L}_i \leftrightarrow s k_i \mathbf{L}_i$ $k_i, i = 1 \dots n; s \quad (n \text{ uncalibrated} + 1)$
physical constraint	positive depths	positive albedoes
linear constraint for 2 views	rigidity of the scene $\mathbf{m}_k^T \mathbf{F}(\mathcal{P}_k, \mathcal{P}_l) \mathbf{m}_l = 0$	uniqueness of albedo $\begin{bmatrix} I_k - b_k \\ -I_l + b_l \end{bmatrix}^T \begin{bmatrix} a_l \mathbf{L}_l^T \\ a_k \mathbf{L}_k^T \end{bmatrix} \mathbf{N} = 0$

Table 2: Parallel between the geometry of multiple views and the radiometry of multiple illuminants

Given a surface element of albedo  $\alpha$  and normal  $\mathbf{n}$  that projects into two images with illuminants  $k$  and  $l$ , its albedo can be estimated using Eq. (4) as:

$$\alpha^k = \frac{I_k - b_k}{a_k \mathbf{L}_k^T \mathbf{N}}, \quad (5)$$

or as

$$\alpha^l = \frac{I_l - b_l}{a_l \mathbf{L}_l^T \mathbf{N}}, \quad (6)$$

where index  $k$  (resp.  $l$ ) denotes gray levels, light source vector parameters, and radiometric calibration parameters in the image  $k$  (resp.  $l$ ). Since the value of the albedo is supposed constant, we can write  $\alpha^k = \alpha^l$  and multiply this equality by the denominators, which gives:

$$\begin{bmatrix} I_k - b_k \\ -I_l + b_l \end{bmatrix}^T \begin{bmatrix} a_l \mathbf{L}_l^T \\ a_k \mathbf{L}_k^T \end{bmatrix} \mathbf{N} = 0 .$$

This can be expanded as:

$$I_l a_k \mathbf{L}_k^T \mathbf{n} + I_l a_k \mu_k - b_l a_k \mathbf{L}_k^T \mathbf{n} - b_l a_k \mu_k = I_k a_l \mathbf{L}_l^T \mathbf{n} + I_k a_l \mu_l - b_k a_l \mathbf{L}_l^T \mathbf{n} - b_k a_l \mu_l . \quad (7)$$

If we denote the illumination vectors as  $\mathbf{L}_k^T = [L_{k1}, L_{k2}, L_{k3}, L_{k4}] = [\mathbf{L}_k^T, \mu_k]$  and  $\mathbf{L}_l^T = [L_{l1}, L_{l2}, L_{l3}, L_{l4}] = [\mathbf{L}_l^T, \mu_l]$ , it can be seen that Eq. (7) is linear in the 12 variables:

$$a_k L_{ki}, a_l L_{li}, a_k L_{ki} b_l - a_l L_{li} b_k, 1 \leq i \leq 4 .$$

Note that, if we had used the 8 variables  $a_k L_{ki} b_l, a_l L_{li} b_k$  instead of the last 4 variables  $a_k L_{ki} b_l - a_l L_{li} b_k$ , we would have obtained a set of 16 non linearly independent variables.

This approach relies on the elimination of the albedoes in order to compute simply the illuminants. Once they are obtained, the albedo are easily computed using Eq. (5).

## 4.2 The calibrated case

In the calibrated case, because the  $a_i$  and  $b_i$  are known, the intensities can be transformed so that we can assume without loss of generality that  $a_i = 1$  and  $b_i = 0$ . Eq. (7) can be rewritten as:

$$\mathbf{U}^T \mathbf{f} = 0$$

$$\begin{aligned} \mathbf{U} &= [I_1 n_1, I_1 n_2, I_1 n_3, I_l, -I_k n_1, -I_k n_2, -I_k n_3, -I_k]^T \\ \mathbf{f} &= [l_{k1}, l_{k2}, l_{k3}, \mu_k, l_{l1}, l_{l2}, l_{l3}, \mu_l]^T \end{aligned} \quad (8)$$

where  $\mathbf{n}^T = [n_1, n_2, n_3]$ .

Combining the rows  $\mathbf{U}$  for each surface element provides a linear system of the form  $\tilde{\mathbf{U}} \mathbf{f} = 0$ . This formulation shows that in the calibrated case there is a unique solution up to a scale factor with 7 generic surface elements, provided that there is no rank loss for the matrix  $\tilde{\mathbf{U}}$ .

In practice, we have more than seven surface elements, but the resulting gray values are not exact. Therefore, we seek a least-squares solution:

$$\min_{\mathbf{f}} \|\tilde{\mathbf{U}} \mathbf{f}\| \quad \text{subject to} \quad \|\mathbf{f}\| = 1 \quad (9)$$

The constraint  $\|\mathbf{f}\| = 1$  is necessary because the illuminants are defined up to a global scale factor. What this means is that while the ratio of the norms of each illuminant vector is fixed (and therefore recovered correctly by solving the linear system), as discussed in Sec. 3.4, the global scale factor is arbitrary. However, its sign is not. One way to set the correct sign is to compute the albedoes using Eq. (5). If the albedoes turn out to be negative, then all the values of the illumination are multiplied by  $-1$ .

## 4.3 Three views and more

One interesting feature of this linear algorithm is that if we have more than two views, we can solve simultaneously for all the illuminants, just by concatenating the equalities of Eq. (7). The unknowns are all the components  $l_{kr}$  and  $\mu_k$  of the illuminants, which are gathered in a vector  $\mathbf{f}$  of dimension  $4 \times n$ . This is possible because there is one illuminant attached to each image, whereas in the geometric recovery problem, the motion parameters are attached to pairs of images. For example, if we have three

views, a surface visible in each view yields the three equations, obtained by expanding respectively  $\alpha^1 = \alpha^2$ ,  $\alpha^1 = \alpha^3$ ,  $\alpha^2 = \alpha^3$ :

$$\begin{bmatrix} I_2 n_1 & I_2 n_2 & I_2 n_3 & I_2 & -I_1 n_1 & -I_1 n_2 & -I_1 n_3 & -I_1 & 0 & 0 & 0 & 0 \\ I_3 n_1 & I_3 n_2 & I_3 n_3 & I_3 & 0 & 0 & 0 & 0 & -I_1 n_1 & -I_1 n_2 & -I_1 n_3 & -I_1 \\ 0 & 0 & 0 & 0 & I_3 n_1 & I_3 n_2 & I_3 n_3 & I_3 & -I_2 n_1 & -I_2 n_2 & -I_2 n_3 & -I_2 \end{bmatrix} \mathbf{f} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (10)$$

with:

$$\mathbf{f} = [l_{11}, l_{12}, l_{13}, \mu_1, l_{21}, l_{22}, l_{23}, \mu_2, l_{31}, l_{32}, l_{33}, \mu_3]^T$$

Clearly, the two first equations are independent, and the last one is a linear combination of them. More generally, a surface visible in  $n$  views yields exactly  $n - 1$  independent equations. With  $p$  surfaces, the number of rows is  $p(n - 1)$ , and the number of columns is  $4n$ . Because of the scale factor, the number of independent equations is at most  $\max(p(n - 1), 4n - 1)$ , which is exactly consistent with the result found in Section 3.4.

All the analysis and counts of variables done previously gave us the minimum data which is necessary to obtain a unique solution. To prove that this data is sufficient when the illuminants and surface normals are generic (or, in other words, in general position) we built a specific system of linear equations for each value of  $(p, n)$  such that  $p = 1..7$  and  $n = 1..5$ . We then used the Maple<sup>tm</sup> computer algebra system to check the rank of the system, and confirmed that there is no rank loss. Table 3(a) shows the rank of the system of equations in the generic case.

images	2	3	4	5
surfaces				
1	1	2	3	4
2	2	4	6	8
3	3	6	9	12
4	4	8	12	16
5	5	10	<b>15</b>	<b>19</b>
6	6	<b>11</b>	<b>15</b>	<b>19</b>
7	<b>7</b>	<b>11</b>	<b>15</b>	<b>19</b>

Generic illumination vectors

images	2	3	4	5
surfaces				
1	1	2	3	4
2	2	4	6	8
3	3	6	9	12
4	4	8	12	16
5	5	9	13	17
6	6	10	14	18
7	<b>7</b>	<b>11</b>	<b>15</b>	<b>19</b>

Planar illumination vectors

(a)

(b)

Table 3: Number of linearly independent equations for small number of images and surface elements, obtained using computer algebra. Bold type indicates that there are enough equations to yield a unique solution (up to a global scale factor)

#### 4.4 Degeneracies

When the data is not in general position, which means that some relationships exist within it, two types of degeneracies can occur, one which is caused by illuminants, the other by the object's shape. In what follows, we give the results of a preliminary study of degenerate configurations, in which we limit ourselves to the case where either the distribution of illuminants or normals is generic.

Let us first examine the case of two illuminants. If the two illumination vectors are proportional, the four first columns of the matrix  $\tilde{\mathbf{U}}$  are proportional to the four last columns, The rank of the matrix is therefore at most 4, and so is the dimension of the space of solutions: As expected, the first illumination vector is unconstrained, the second one equal to the first. More generally, when all  $N$  illumination vectors are proportional, it can be seen that the matrix  $\tilde{\mathbf{U}}$  generalizing the form in Eq. (10) has rank at most  $4 \times (n - 1)$ : The last  $n - 1$  groups of four columns are independent, while the first four columns are a linear combination of them. In the geometric reconstruction case, this degeneracy corresponds to identical optical centers. The distance between illuminants plays a role similar to the baseline in geometrical reconstruction. It is expected that the larger it is, the more stable the solution will be. This will be confirmed by the experiments presented in Sec. 4.6.

Besides this obviously expected degeneracy, we found almost no case of degeneracy caused by illuminants with generic surface normals. Of particular interest is the fact that for  $n = 5$ , the system doesn't lose rank in the generic case, which implies that the information obtained by using more than four illuminants is not redundant. This is despite the fact (first noticed in [27, 32]) that the illumination vectors being of dimension four, any system of five or more illumination vectors is linearly dependent, and as a result, the space of images (in the common image plane, or equivalently, taken from the same viewpoint) is itself of dimension four. Moreover, as shown in Table 3(b), we verified that generically, there is no rank loss for  $p \geq 7$  as soon as there are at least two independent illumination vectors, even if all the other illumination vectors are linearly dependent on these two vectors (or, in other words, if all the illumination vectors are planar). As an application of this result, given ten views, only one of which has a different illuminant which is not yet identified, we are able to compute the ten illuminants and then classify nine images as having the same illuminant. Being able to use all ten views provides additional robustness against noise. Finally, the method is applicable even if all the light source vectors are identical—that is, when the point light sources are the same—but at least two illuminants have different ambient components. As expected, in such a situation, the algorithm becomes much more sensitive to noise.

It is known that a planar surface causes degeneracies in geometrical reconstruction. Likewise, a planar scene, for which all the normals are identical, is highly degenerate for radiometric reconstruction. Examining again Eq. (8), it is apparent that two surface elements which have the same normal yield two proportional rows  $\mathbf{U}$ , the proportionality factor being the ratio of the albedoes. Therefore, the rank of the system would be at most  $n - 1$ , no matter the number of surface elements. A more general degeneracy occurs when all the surface normals are coplanar (as in a cylindrical surface). It can be verified that then the rank of the system is generically at most  $3n$ . For each of the  $n$  light sources, the component along the direction orthogonal to all the normals is ambiguous. We conclude that for the matrix  $\tilde{\mathbf{U}}$  to have full rank, it is necessary that we have at least seven surface elements with different orientations, and that all these orientations be non-planar, which is a much stronger requirement than the non-coplanarity requirement. In practice, while this requirement might be problematic with simple polyhedral scenes, it is generally satisfied for smooth surfaces. It is always a good idea to watch the singular values of matrix  $\tilde{\mathbf{U}}$  to detect any possible rank loss.

## 4.5 Recovering the offsets

In the uncalibrated case, because of the ambiguity between the intensity of the illuminant and the radiometric scale described in Sec. 3.4, the only calibration parameters which can be recovered is the radiometric offsets  $b_i$ .

Eq. (7) can be rewritten as:

$$\mathbf{V}^T \mathbf{g} = 0$$

$$\begin{aligned} \mathbf{V} &= [I_l n_1, I_l n_2, I_l n_3, I_l, -I_k n_1, -I_k n_2, -I_k n_3, -I_k, n_1, n_2, n_3, 1]^T \\ \mathbf{g} &= [L_{k1}, L_{k2}, L_{k3}, L_{k4}, L_{l1}, L_{l2}, L_{l3}, L_{l4}, m_{kl1}, m_{kl2}, m_{kl3}, m_{kl4}]^T \end{aligned}$$

where we have used the notations:

$$\begin{cases} L_{ji} = l_{ji}, i = 1 \dots 3 \\ L_{j4} = \mu_j \end{cases} \quad \text{and} \quad \begin{cases} m_{jli} = b_l l_{ji} - b_j l_{li}, i = 1 \dots 3 \\ m_{jl4} = b_l \mu_j - b_j \mu_l \end{cases}$$

For  $n$  illuminants, the number of intermediate variables is  $4 \times (n + \frac{n(n-1)}{2}) = 2n(n+1)$ . We have verified that these variables are linearly independent in the generic situation.

In the uncalibrated case there is a unique solution up to the ambiguity described in Sec. 3.4 as soon as we have 11 surface elements.

The system of linear equations obtained by concatenating the previous equation is solved in the same way as for the calibrated case. Then to recover the offsets, we notice that:

$$\begin{aligned} (L_{pj}L_{qi} - L_{pi}L_{qj})b_k &= (b_p L_{qi} - b_q L_{pi})L_{kj} - (b_p L_{qj} - b_q L_{pj})L_{ki} \\ &= m_{pqi}L_{kj} - m_{pqj}L_{ki} \end{aligned} \quad (11)$$

Since this equation is valid for all the indices such that

$$\begin{aligned} k &= 1 \dots n \\ 1 &\leq i < j \leq 4 \\ 1 &\leq p < q \leq n \\ p &= k \text{ or } q = k \end{aligned} ,$$

we can estimate robustly the  $n$  offsets. In our implementation, we have used the simplest way to exploit the redundancy. We average all the values of  $b_k$  derived from Eq. (11).

## 4.6 Synthetic Experiments

The algorithm described below is very simple. It is the radiometric equivalent of the eight-point algorithm for geometric reconstruction. Unfortunately, we are going to see in this section that likewise, it is quite sensitive to image noise. We use synthetic data to assess this sensitivity.

There are three main parameters that we can vary: The amount of noise, the number of surface elements, and the number of illuminants. For each specific choice of these parameters, we randomly draw values for the light source vectors. We then randomly draw values for the surface normals that are such that the surface elements are front-lit, that is such that the dot products of the normal with all the illuminant orientations are positive. We then pick random values between 0 and 1 for the albedoes, compute the expected gray level values and add noise. It is taken to be Gaussian, and expressed in terms of relative error with respect to the maximum gray-level value. Finally, we run our procedure using the normals and noisy gray-levels to recover estimated illuminants and compare them to the real values.

Given this procedure, if we allowed the illuminant orientation to span the whole space, the set of available orientations for the surface elements would become increasingly small as the number of illuminants increased. It would decrease the algorithm's robustness and bias the results against large number of cameras. To avoid this, we restrict the illuminants' orientation to be within 45 degrees on either side of the  $z$  axis, thereby guaranteeing that the maximum angle between them is always less than 90 degrees.

To compare true and recovered illuminants, we form "concatenated illuminant vectors" by concatenating all the components of illuminations in all images, first using the true illumination values and second using the recovered one. We take the "distance" between two such concatenated illuminant vectors  $\vec{V}_1$  and  $\vec{V}_2$  to be

$$D_{vect}(\vec{V}_1, \vec{V}_2) = 1 - \frac{\vec{V}_1 \cdot \vec{V}_2}{\|\vec{V}_1\| \|\vec{V}_2\|} . \quad (12)$$

It is one when the two vectors are proportional to each other and zero when they are perpendicular. This distance is a measure of the recovery error.

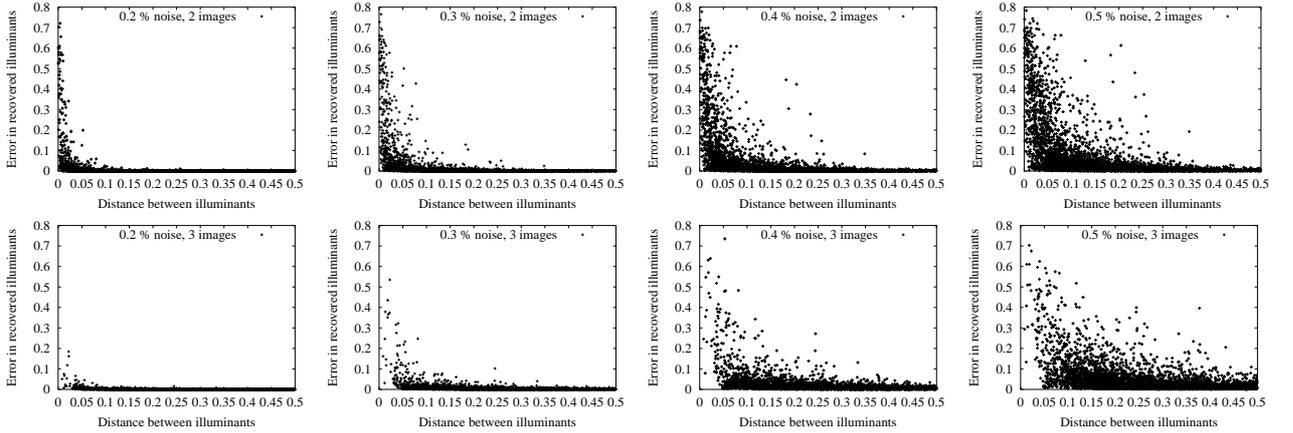


Figure 1: Sensitivity of the linear algorithm using 2 or 3 images and from 0.2% to 0.5% gray-level noise. Each point in the scatter plots depicts one run of the algorithm. Its x-coordinate is the distance between the illuminants and its y-coordinates the recovery error.

In the first line of Figure 1, we show the results of four such experiments using two illuminants and noise values of 0.2%, 0.3%, 0.4% and 0.5%. In the second line, we show the corresponding results using three illuminants instead. Each point of each scatter plot corresponds to one run of the algorithm. Its y-coordinate is the recovery error of Eq.(12). In the two-illuminant case, its x-coordinate is the distance  $D_{vect}$  of Eq. (12) between the two individual illuminant vectors. In the three-illuminant case, its x-coordinate is the average distance of the three possible pairs of illuminants.

These results are presented in a different fashion in Figure 2. We divide the range of distances between illuminants, that is the x-axis of the graphs of Figure 2, into 20 bins. Within each bin, for all runs whose randomly chosen illuminants have a distance that falls into that bin, we compute and plot the 90 percentile value of all the recovery errors, that is, the smallest value that is larger than 90% of these recovery errors. Again in the three-illuminant case, we use the average distance between pairs.

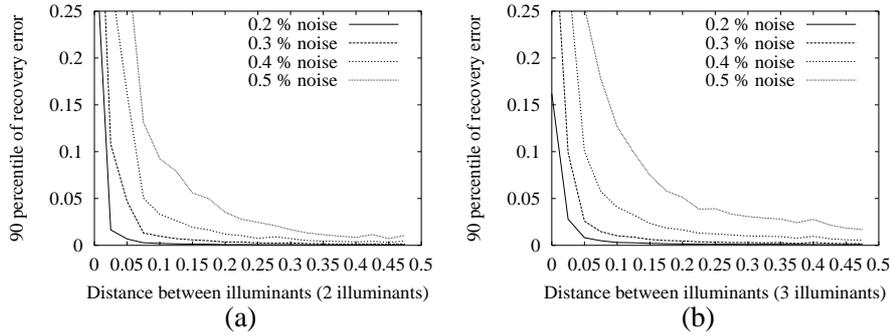


Figure 2: Sensitivity of the linear algorithm to gray-level noise in terms of the 90 percentile of recovery errors. Each curve corresponds to a different amount of gray-level noise. (a) Using two images. (b) Using three images.

Naturally, the recovery error increases with the amount of noise. It also increases when the distance between the individual illuminants decreases. This also is to be expected since, as this distance approaches zero, the illuminants become closer to being collinear and the systems of equations closer to being degenerate.

Note that increasing the number of illuminants from two to three, and beyond, does not bring any meaningful improvement. The usefulness of the extra equations appears to be severely restricted by the additional number of unknowns.

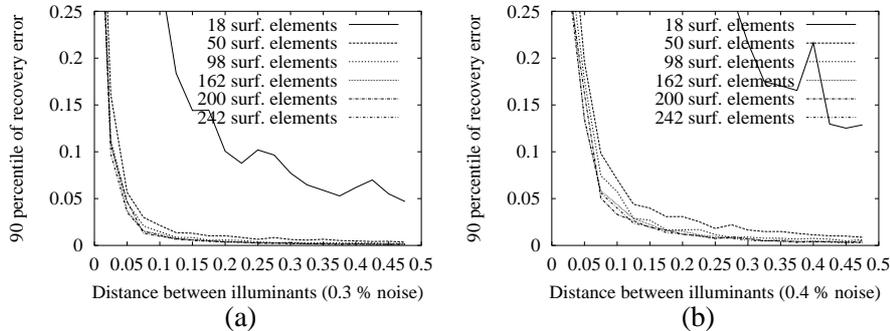


Figure 3: Sensitivity of the linear algorithm to the number of surface elements. Each 90 percentile curve corresponds to a different number of elements. (a) 2 images and 0.3% gray-level noise. (b) 2 images and 0.4% gray-level noise.

All the experiments depicted by Figures 1 and 2 were performed using 200 surface elements for each run of the algorithm. To illustrate the method's sensitivity to this number, we reran our experiments using different numbers of surface elements. Figure 3 depicts the results. In practice, as soon as the number of surface elements becomes larger than 50, the results become statistically indistinguishable.

These simulations show that the linear algorithm is quite sensitive to noise. This sensitivity is actually comparable to that of the eight-point algorithm [22, 15]. To derive satisfactory results with it, points must be located with an accuracy of one pixel in  $512 \times 512$  images, which corresponds to a relative error of 0.002 (0.2%). Practice has shown that such a geometric precision is quite attainable,

and sometimes improved upon by using sub-pixel detectors. Furthermore, it is often possible to use a higher resolution image. On the other hand, the difficulty with radiometric reconstruction is that the values we measure are digitized gray levels, which are typically numbers between 0 and 255. In this case, an error of one gray-level is 0.004 (0.4%) and improving upon this precision is problematic. Unfortunately, as seen above, at this level of precision the linear algorithm is already starting to make significant errors. We therefore turn to a non-linear approach to solving the problem that will prove to be much less noise-sensitive.

## 5 Non-linear Solution: Radiometric Bundle Adjustment

The radiometric bundle adjustment method is the counterpart of the geometric bundle adjustment used in photogrammetry [13]. It is promising because the latter one has proved to be one of the most robust methods for geometric reconstruction. In this approach, we seek to find the values of the  $5 \times n + p - 1$  unknowns  $\alpha_j, j = 1 \dots p$ , and  $\mathbf{L}_i, b_i, 1 \dots n$ , which minimize the error function:

$$\sum_i \sum_j (I_{ij} - (\alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i))^2 \quad (13)$$

where  $I_{ij}$  are the measured-gray levels in the images, and the sums are taken over all the images and for each projection of a surface element. In, practice, the minimization is performed using the Levenberg-Marquardt algorithm [29]. We use the linear algorithm previously described to get a starting point.

The main drawback of this method is that involves non-linear optimization to solve the least-squares problem of Eq. (13). Its size is proportional to the number of surface elements, and can therefore become large. However, as will be shown below, the increase in reliability is well worth the additional computational burden. Furthermore, as shown by Figure 3, 200 surface elements are more than sufficient to constrain the problem. Therefore, for large problems, not all surface elements need to be used.

### 5.1 Synthetic Experiments

We ran experiments similar to those depicted by Section 4.6, but using this non-linear approach and much higher noise levels. The results are depicted by Figure 4. As in Figure 2, the individual curves correspond to different levels of gray-level noise. They are the 90 percentile of recovery errors shown as functions of the distance between illuminants. In the case of two images, the results are excellent at the 0.5% and 1.0% noise levels, levels for which the linear algorithm's results are already becoming incorrect. The precision then degrades gracefully as the noise level increases. Also, unlike in the linear case, using three illuminants instead of two yields a marked improvement. The method is still sensitive to noise but, because a 1.0% noise level corresponds to about 3 gray-levels in a typical 8-bit image, it becomes much more applicable.

The non-linear algorithm also recovers an estimated albedo value for each facet. In Figure 5, we divide the x-axis of the graphs in the same 20 bins as before. We compute and plot the 90 percentile of the root mean-square difference between recovered albedoes and real ones, instead of the 90 percentile of the distance between true and recovered illuminants as in Figure 4. Note that the graphs of both

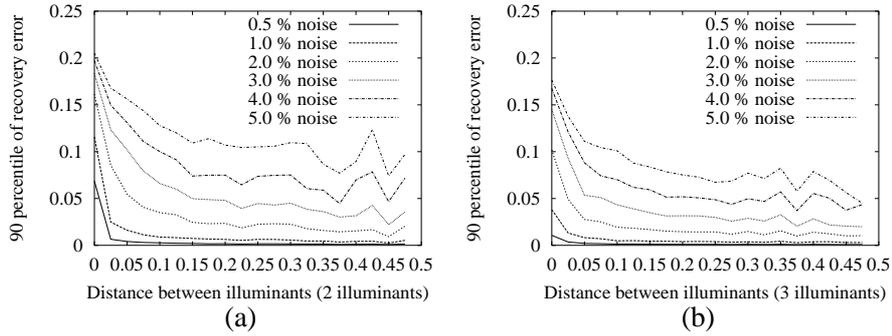


Figure 4: Sensitivity of the non-linear algorithm to gray level noise in terms of the 90 percentile of recovery errors. Each curve corresponds to a different amount of gray-level noise. (a) Using two images. (b) Using three images.

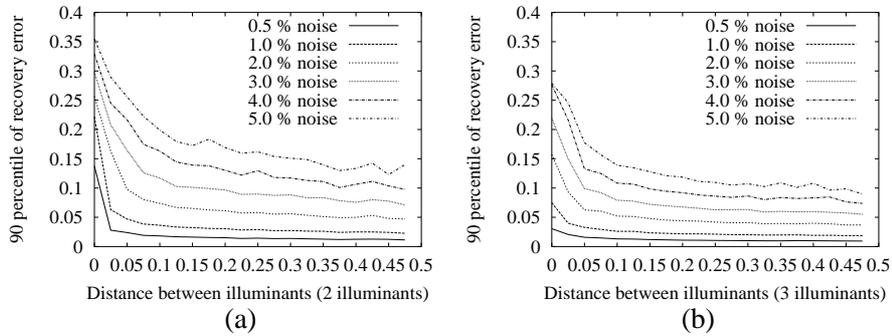


Figure 5: Sensitivity of the non-linear algorithm to gray level noise in terms of the 90 percentile of the root-mean square error between true and recovered albedoes. Each curve corresponds to a different amount of gray-level noise. (a) Using two images. (b) Using three images.

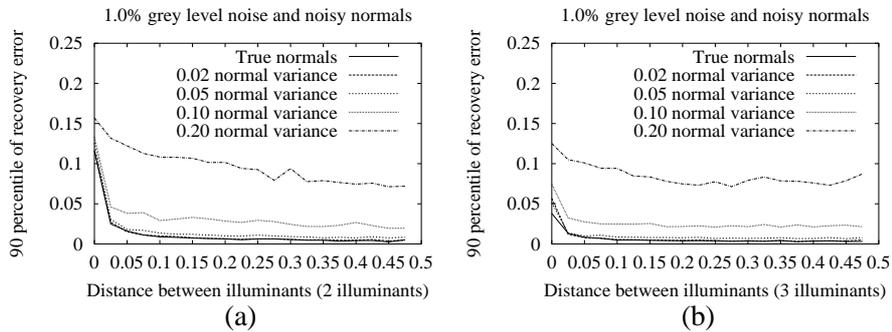


Figure 6: Sensitivity of the non-linear algorithm to errors in the normals. Each 90 percentile curve corresponds to a different amount of randomization of the normals. (a) 2 images and 1.0% gray-level noise. (b) 3 images and 1.0% gray-level noise.

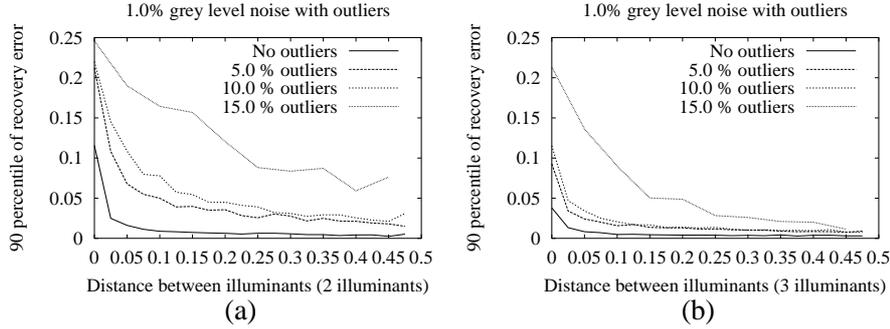


Figure 7: Sensitivity of the non-linear algorithm to outliers. Each 90 percentile curve corresponds to a different proportion of outliers. (a) 2 images and 1.0% gray-level noise. (b) 3 images and 1.0% gray-level noise.

figures have exactly the same shapes, indicating that our distance measure between illuminants is also a good indicator of the quality of albedo recovery.

So far we have assumed that the normals were precisely known. In practice, this will of course never be the case. Normals will only be known up to some precision. To test the influence of this uncertainty, we ran additional experiments by also randomizing the normal vectors. We added Gaussian noise to each of their 3 components and renormalized the vectors to make them unit vectors again. Figure 6 depicts the results when randomizing the normals using noise of variance ranging from 0.02 to 0.20 *and* introducing 1.0% of gray-level noise. Note that the result is relatively insensitive to fairly large imprecisions in the normals.

Another source of errors that must be taken into account are gray-levels that are not merely imprecise but actually completely wrong, "blunders" in photogrammetric terms. They are guaranteed to arise in real scenes, for example because some surface elements are specular or in shadow. To handle these, we replace the simple Levenberg-Marquardt minimization of Eq. (13) by a RANSAC style approach [7]. It takes the following steps:

1. Randomly draw, with replacement, a subset  $S$  of surface elements. Use it to estimate the illuminants parameters by minimizing the corresponding sum of residuals

$$\sum_i \sum_{j \in S} (I_{ij} - (\alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i))^2 .$$

Using these illumination parameters, estimate the albedo  $\alpha_j$  of each surface elements assuming that it is Lambertian and not in shadow. For each surface element, form the element's residual:

$$res_j = \sum_i (I_{ij} - (\alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i))^2$$

If the model's assumptions hold for a particular surface element, this residual should be small. We take their median value for all surface elements to be a measure of the quality of the illumination estimates.

2. Perform this computation several times and keep the illumination estimate that yields the smallest median residual. We denote it  $res_{med}$ .

3. Improve this estimate by performing iteratively reweighted least squares as follows. For each surface element  $j$ , compute a weight  $w_j = \exp(-res_j/res_{med})$  that is inversely proportional to the corresponding residual and minimize

$$\sum_i \sum_j w_j (I_{ij} - (\alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i))^2 .$$

Perform this minimization several time and recompute the  $res_j$  residuals and corresponding  $w_j$  weights every time. After a small number of iterations, the outliers have very small weights and are effectively discounted.

Figure 7 shows the results of this procedure, still using a 1.0% gray level noise and adding up to 15% of outliers.

## 5.2 Using Real Images

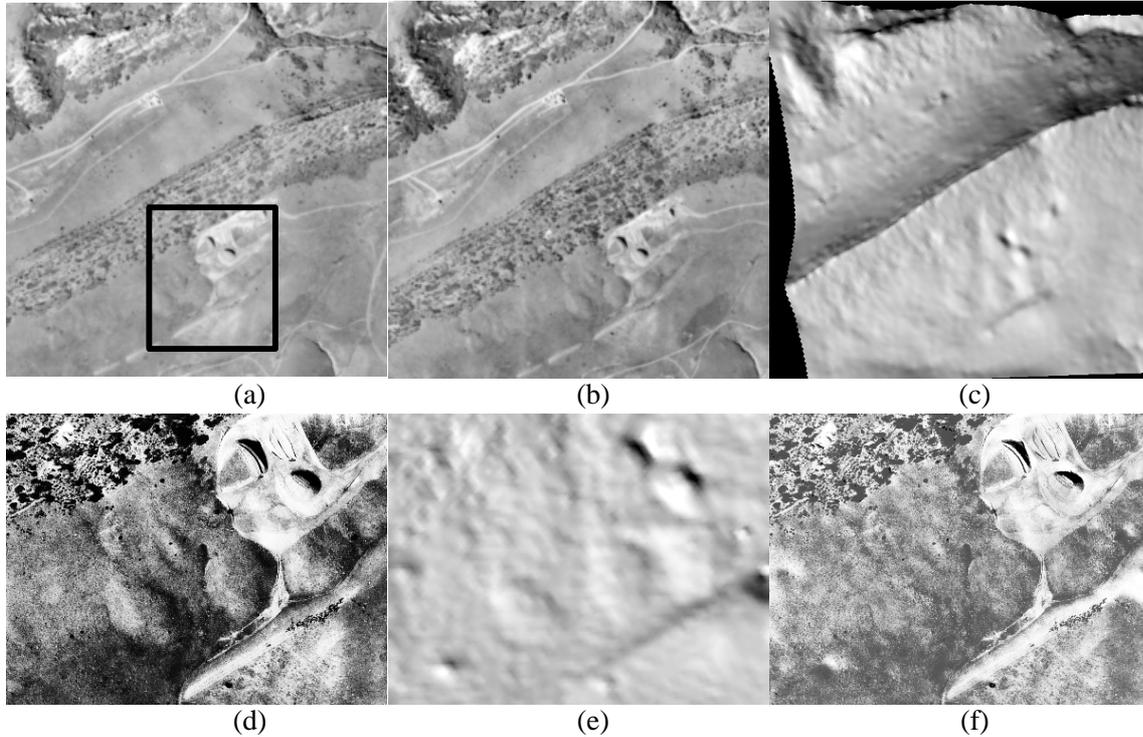


Figure 8: Cartographic Modeling: (a,b) Two aerial images of a site taken at two different times of the day. (c) The recovered terrain model, shaded using the illuminant computed by our system for image (a). (d) A higher resolution image of the area enclosed by the rectangular frame in (a). (e) A shaded view of the corresponding terrain, again using the illuminant recovered by our system. (f) The corresponding de-lighted image.

In Figure 8, we show two aerial images of the same site, but taken at different times of day and from slightly different viewpoints. We treated them as a stereo pair. We reconstructed the terrain's

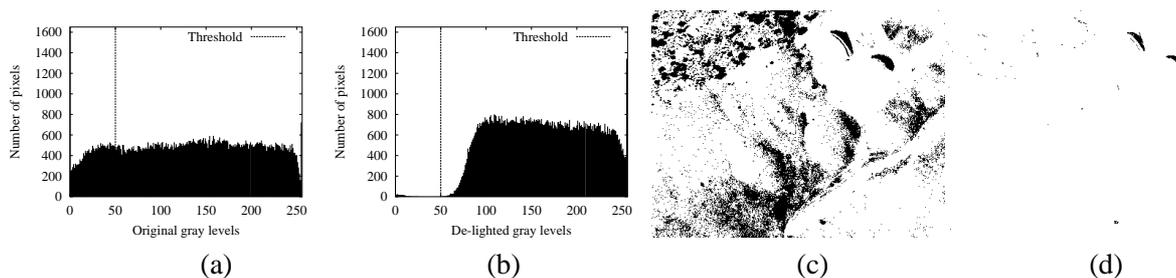


Figure 9: Dynamic range before and after illumination correction: (a) Histogram of the original image of Figure 8(d). (b) Histogram of the de-lighted image of Figure 8(f). (c,d) The pixels whose gray level is smaller than 50, in the original and de-lighted image respectively, are shown in black. Note that the two images have the same overall dynamic range but that, in the case of the de-lighted image, only a small number of pixels are truly dark. These are the pixels that belong to the shadows in the upper-right corner of the image.

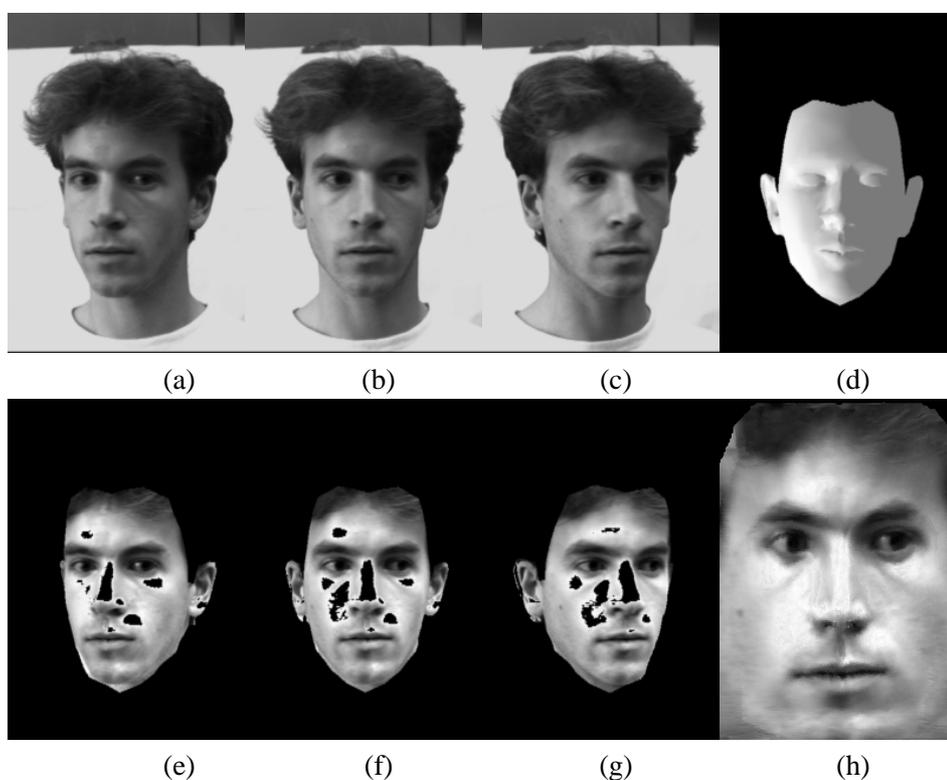


Figure 10: Facial Reconstruction: (a,b,c) Three image of a moving head acquired using a static camera. The non-diffuse component of the lighting comes from the upper left corner of the image and the right side of the face appears darker. (d) The reconstructed head model shaded using the illuminant computed by our system. (e,f,g) The corresponding de-lighted images. The black areas correspond to pixels whose gray level is very different from the one predicted by the Lambertian model. (h) Cylindrical texture map produced by combining the (e),(f) and (g) images and filling the holes using a texture filling algorithm.

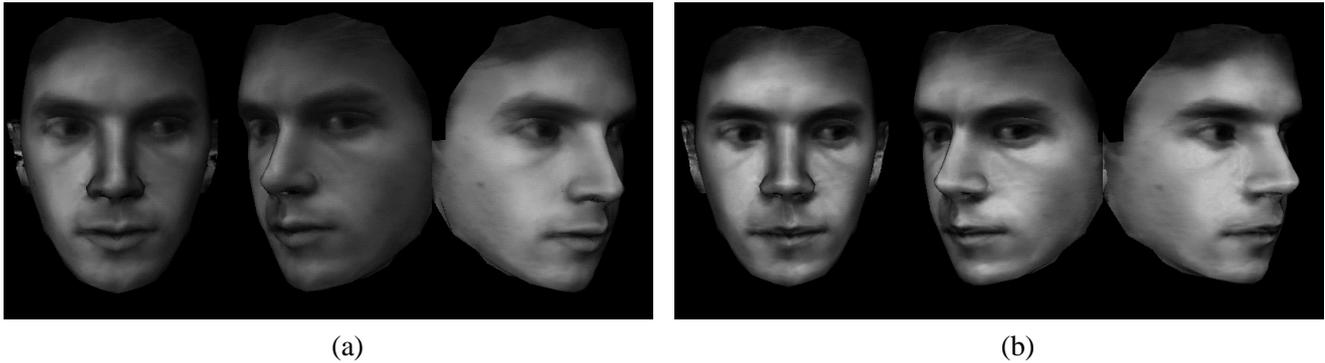


Figure 11: Three views of the 3-D model of Figure 10(d): (a) Texture-mapped using the original images. (b) Texture-mapped using the image of Figure 10(h) that has been computed using the de-lighted images.

shape using a deformable surface technique we developed in earlier work [11]. Note that, because the illuminant is different for the two scenes, we first ran a difference of Gaussian operator on the images before performing the reconstruction. The result is a triangulated mesh. We computed the average gray-level of each mesh facet's projection in each image. We then fed these gray-levels along with the facet normals to our algorithm. Figure 8(c) shows the mesh shaded using the illuminant recovered for the first image. Note that the dark parts of the shaded mesh are on the same side of the relief as the shadows, indicating that the illuminant's orientation is, at least roughly, correct. In the second row of Figure 8, we show a higher resolution detail of the first image, the corresponding shaded model and the albedo at each pixel location that would produce the observed image gray-levels, assuming that the illuminant we computed is correct. The resulting de-lighted image exhibits much less structure than the original one, even though it has the same dynamic range as the original one as shown in Figure 9: There are still some very dark pixels in the shadowed areas of the upper right corner, but not elsewhere. The shading effects have been greatly attenuated because they are accounted for by changes in shape.

In Figure 10, we present a similar behavior in a different context. The three images of the first row were acquired using a static camera watching a moving head. The lighting is not diffuse and one side of the head appears to be much darker than the other. Furthermore, there are many specularities that make our task more difficult. We computed the head motion and reconstructed its shape using a facial reconstruction technique also developed in earlier work [10]. As before, because the illuminant changes from image to image, we extracted 3-D shape information using a technique, normalized cross-correlation [9], that is relatively insensitive to such changes. Again, the result is a triangulated mesh. In Figure 10(d), it is shaded using the illuminant recovered by our system. In this case there is only one light source and the head is moving with respect to it. Because our geometric reconstruction method also derives the motion parameters [10], we can reduce the number of degrees of freedom: Only one of illuminant is free and the others are taken to be rotated versions of it. To overcome the influence of specularities, we further reduce the number of degrees of freedom: We constrain the facets that are symmetric with respect to a vertical plane that goes through the middle of the head to have the same albedoes.

In the second row of Figure 10, we show de-lighted images derived as before. In these images, pixels whose gray level is very different from the one predicted by the Lambertian model are shown in black. Again, the effects of shading have been greatly attenuated. Of course, ignoring specularities has led to the incorporation of specular components into diffuse albedoes. As a result, the estimated

albedoes are too large on the forehead and cheeks. This is a clear indication that specularities should be explicitly modeled. Nevertheless, these de-lighted images can then be used to produce the texture-map of Figure 10(h). This is achieved by first generating a cylindrical projection of the head model and then, for each projected point, finding the images in which it is visible and averaging the corresponding gray-levels. This texture-map can in turn be used to synthesize 3-D models such as those of Figure 11. For comparison purposes, we also show the 3-D model texture-mapped using the original images. Note that, when the images are not corrected for lighting, one side of the face appears much darker than the other.

## 6 Conclusion and discussion

We have introduced a new methodology for recovering the illuminants and surface albedoes from multiple views taken with multiple illuminants. When coupled with a method for shape reconstruction, it allows us to recover both the geometric and radiometric attributes of a scene. We have also shown that in a linear setting, the only observable camera radiometric parameter, the offset, can be recovered.

We have developed the theory, which indicates the minimal data as well as degenerate configurations and sets the limits for radiometric reconstruction. We proposed two different algorithmic implementations:

1. **A linear approach.** It allows fast computation, but is sensitive to noise.
2. **A radiometric bundle-adjustment algorithm.** It involves non-linear optimization, that makes it slower but also much more robust.

Using synthetic data, we have quantified the sensitivity of the algorithms to image noise, outliers, imprecision of surface normals, distance between illuminants, and number of surface elements. Because reliable results can be obtained using relatively few surface elements, we have not found the extra computational expense involved by the non-linear method to be a severe drawback. The usefulness of this algorithm has been illustrated by two experiments that show that, even with a minimal number of real images, we can obtain 3-D models and plausible estimates of the light source direction and albedoes. The implementation represents a first step towards a solution of the radiometric reconstruction problem, rather than a complete solution, and could be combined with existing approaches.

This work, like many, is based on the Lambertian reflectance model with ambient light and distant point sources. The surface elements that do not conform to it are treated as outliers and discarded as such by the bundle-adjustment algorithm. Note, however, that the recovery algorithm is implemented using a general least-squares solver in which the constraints derived using the Lambertian model could easily be replaced by ones computed using more complex illumination models such as those that take specularities into account.

Similarly, we do not explicitly model shadows and treat shadow areas as outliers. There are, however, shadow extraction techniques that do not depend on geometry and could be used. Moreover, given a known surface geometry, the shadow boundaries provide clues as to the location of point light sources. In an iterative scheme, after a first estimation of the light sources has been obtained, it would be possible to use the geometry to predict potential shadowed areas and take that information into account in the next iteration.

In the absence of shadows and specularities, any combination of light sources could be accounted for by a single point source and an ambient term. Understanding their roles is therefore crucial to extending our approach from the simple case where there is a single point-light source per image to more realistic situations that involve multiple light-sources per image. In future work, we will therefore focus on introducing shadow and specular models into the proposed framework.

**Acknowledgements** The authors would like to thank the reviewers for encouraging and useful comments.

## References

- [1] E. Angelopoulou and J. Williams. Photometric surface analysis in a tri-luminal environment. In *International Conference on Computer Vision*, pages 442–447, 1999.
- [2] R. Baribeau, M. Rioux, and G. Godin. Color reflectance modeling using a polychromatic laser range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):263–269, February 1992.
- [3] P.N. Belhumeur, D.J. Kriegman, and A.L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, November 1999.
- [4] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *Computer Graphics, SIGGRAPH Proceedings*, 31(Annual Conference Series):369–378, 1997.
- [5] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *European Conference on Computer Vision*, pages 563–578, Santa-Margherita, Italy, 1992.
- [6] O.D. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. MIT Press, 2001.
- [7] M.A. Fischler and R.C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications ACM*, 24(6):381–395, 1981.
- [8] D.A. Forsyth and A. Zisserman. Reflections on shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):671–679, July 1991.
- [9] P. Fua. A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features. *Machine Vision and Applications*, 6(1):35–49, Winter 1993.
- [10] P. Fua. Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data. *International Journal of Computer Vision*, 38(2), July 2000.
- [11] P. Fua and Y. G. Leclerc. Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading. *International Journal of Computer Vision*, 16:35–56, September 1995.
- [12] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting. In *IEEE Workshop on multiple-view modeling and analysis of visual scenes*, pages 47–54, 1999.
- [13] A. Gruen and H.A. Beyer. System Calibration through Self-Calibration. In *Calibration and Orientation of Cameras in Computer Vision*, Washington D.C., August 1992.
- [14] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [15] R.I. Hartley. In Defense of the Eight-Point Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997.
- [16] R.I. Hartley, R. Gupta, and T. Chang. Stereo from Uncalibrated Cameras. In *Conference on Computer Vision and Pattern Recognition*, pages 761–764, 1992.

- [17] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *JOSA-A*, 11(11):3079–3089, November 1994.
- [18] B.K.P. Horn. *Robot Vision*. MIT Press, 1986.
- [19] B.K.P. Horn and M.J. Brooks. *Shape from Shading*. MIT Press, 1989.
- [20] K. Ikeuchi and K. Sato. Determining reflectance properties of an object using range and brightness images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1139–1153, November 1991.
- [21] H.C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature*, 293:133–135, 1981.
- [22] Q.-T. Luong and O.D. Faugeras. The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.
- [23] Q.-T. Luong and O.D. Faugeras. Self Calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1997.
- [24] S. Marschner. *Inverse rendering for computer graphics*. PhD thesis, Cornell University, 1998.
- [25] S. Marschner and D. Greenberg. Inverse lighting for photography. In *IST/SID Fifth Color Imaging Conference*, pages 262–265, 1997.
- [26] T. Mitsunaga and S.K. Nayar. Radiometric self calibration. In *Conference on Computer Vision and Pattern Recognition*, pages I:374–380, 1999.
- [27] Y. Moses. *Face recognition: generalization to novel images*. PhD thesis, The Weizmann Institute of Science, Israel, 1993.
- [28] N. Mukawa. Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *International Conference on Computer Vision*, pages 507–512, 1990.
- [29] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes, the Art of Scientific Computing*. Cambridge U. Press, Cambridge, MA, 1986.
- [30] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from brightness in shadows: Adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In *International Conference on Computer Vision*, pages 875–883, 1999.
- [31] M. Sato, Y. Wheeler and K. Ikeuchi. Object shape and reflectance modeling from observation. In *Computer Graphics, SIGGRAPH Proceedings*, pages 379–387, 1997.
- [32] A. Shashua. On photometric issues in 3d visual recognition from a single 2d image. *International Journal of Computer Vision*, 21(1-2):99–122, January 1997.
- [33] W. Silver. *Determining shape and reflectance using multiple images*. PhD thesis, MIT, Cambridge, MA, 1990.
- [34] R. Szeliski. Stereo Algorithms and Representations for Image-Based Rendering. In *British Machine Vision Conference*, pages II:314–328, 1999.
- [35] R.J. Woodham. Photometric method for determining surface orientation from multiple images. *OptEng*, 19(1):139–144, January 1980.
- [36] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. *Computer Graphics, SIGGRAPH Proceedings*, pages 215–224, August 1999.
- [37] A.L. Yuille, D. Snow, R. Epstein, and P.N. Belhumeur. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *International Journal of Computer Vision*, 35(3):1–20, December 1999.
- [38] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):680–702, July 1991.