

# Recovery of Reflectances and Varying Illuminants from Multiple Views

Q.-Tuan Luong<sup>1</sup>, Pascal Fua<sup>2</sup>, and Yvan Leclerc<sup>1\*</sup>

<sup>1</sup> AI Center, SRI International, CA-94025 Menlo Park, USA,  
{luong,leclerc}@ai.sri.com

<sup>2</sup> Computer Graphics Lab, EPFL, CH-1015 Lausanne, Switzerland  
pascal.fua@epfl.ch

**Abstract.** We introduce a new methodology for *radiometric reconstruction* from multiple images. It opens new possibilities because it allows simultaneous recovery of varying unknown illuminants (one per image), surface albedoes, and cameras' radiometric responses. Designed to complement geometric reconstruction techniques, it only requires as input the geometry of the scene and of the cameras. Unlike photometric stereo approaches, it is not restricted to images taken from a single viewpoint. Linear and non-linear implementations in the Lambertian case are proposed; simulation results are discussed and compared to related work to demonstrate the gain in stability; and results on real images are shown.

## 1 Introduction

Consider cdhaving several magazines with pictures of a mountain villa taken at different times of day and from different vantage points. Using modern bundle adjustment and geometric reconstructions techniques, it is possible to recover the geometry of the scene from the pictures. However, there is currently no way to automatically recover the unknown radiometric responses of the cameras/digitizers and the unknown illumination for each picture. Without these, it is not possible to create a photo-realistic rendering of the scene. Given the scene geometry and the camera's geometric parameters, we call the recovery of these unknowns the *radiometric recovery* problem.

Here, we focus on an important special case of the general problem in which the unknown camera/digitizer responses are modelled by an affine transformation (scale and offset), the unknown illuminants in each picture are modelled by a single point light sources at infinity plus ambient illumination, and the

---

\* This work was sponsored in part by the Defense Advanced Research Projects Agency under contract F33615-97-C-1023 monitored by Wright Laboratory and in part by the Swiss Federal Office for Education and Science. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency, the United States Government, SRI International or EPFL.

scene BRDF is textured Lambertian. This is reasonably close to the mountain villa scenario above. The main insight of this work is that these unknowns can be recovered from multiple images in much the same way as the geometry of a rigid scene can be recovered from multiple images with unknown but varying viewpoints. Compared to the photometric stereo approaches, there is no need to use a constant viewpoint, thus allowing us to take advantage of the reliable and flexible existing solutions for the geometric reconstruction problem.

The decoupling of geometry and radiometry that we advocate makes sense for many applications. In some cases, for example in an industrial setting, 3-D models of target objects may be available. In many other cases, they can be built using well understood vision-based techniques: When performing change detection using images of the same sites taken at different times of day, digital terrain maps are available or can be computed with geometric reconstruction techniques, such as stereo based on normalized cross-correlation, which can handle moderate lighting variations. It is then natural to combine maps and actual images to estimate the illumination parameters, recover surface albedoes and detect their changes over time. When creating 3-D models of objects or people from video sequences for rendering purposes, the illumination may vary from image to image because the subject moves with respect to fixed illuminants. After using structure-from-motion techniques that are relatively insensitive to illumination changes to recover geometry, one can attenuate the lighting effects in the original images to create high-quality texture maps.

**Table 1.** Radiometric reconstruction in relation with previous general frameworks

<b>Framework</b>	Camera geom.	Scene geom.	Reflectance	Illumination
Geom. reconstruction	recovered	recovered	some variation	some variation
Shape from shading	single	recovered	constant	recovered
Photometric stereo	constant	recovered	recovered	known
General. phot. stereo	constant	partly recov'd	partly recov'd	partly recov'd
Inverse lighting	single	known	known or constant	recovered
Reflectance modeling	known	known	recovered	known
<b>Our approach</b>	known	known	recovered	recovered

Previous work is summarized in Table 1 and discussed in detail in Section 2. 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 for radiometric recovery from multiple views and varying illuminants. Our goal is to introduce a framework for such a task and associated algorithms. As a first step towards the general problem of radiometric reconstruction, we investigate the case of Lambertian surfaces. The relevant lighting model consists of a single light source for each input image.

In Section 3, we describe the linear radiometry of multiple illuminants, 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. A more

detailed account of this theory, as well as a parallel between the radiometric reconstruction problem and the geometric reconstruction problem can be found in [12]. Simulations 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. The results of this approach compare very favorably with those of the SVD method [8,3,26]. Section 6 discusses limitations of the current system and areas for future developments.

## 2 Previous Work

Shape from shading [10] assumes that the reflectance is Lambertian and uniform. Surface shape, and the direction of the illuminant are recovered from a single [27] or multiple images [17].

Photometric stereo, by using multiple images, each taken with the same view-point but a different illuminant explicitly known [23,24,9] or implicitly determined [1], recovers a non-uniform albedo, as well as the surface orientation.

Subsequently, it was shown [8] that even with unknown illuminants, albedo and shape can be recovered using SVD, which generalized photometric stereo. However, both of them are subject to a Generalized Bas-Relief ambiguity [3,26]. This is intrinsic to the problem and can be entirely resolved only by sufficient prior knowledge about the object's geometry. For novel view generation, knowing that the object belongs to an instance of a class [7] provides an adequate approximation. These methods are powerful. If given a sufficient number of 3D measurement on the object, they can recover both the illuminant and reflectance. However their precision has not been documented, nor their applicability to unconstrained or large environments. In Sec. 5.2 we compare experimentally the SVD approach with our radiometric reconstruction approach.

The primary purpose of all the previous approaches is to recover the shape based on the Lambertian model. A second family of methods recovers radiometric parameters assuming that the camera and object geometry is available. Inverse lighting [13,14] assumes that the reflectance is Lambertian, and known or constant. The illuminant is modeled as a linear combination of a basis of pre-defined lights, whose coefficients are recovered from a single image using a linear relationship. The dual approach, reflectance modeling, consists in assuming that the point light source is known. It recovers sophisticated reflectance properties with specular and diffuse parameters [11,2,21,13]. This has been extended to multiple known point light sources, as well as indirect illumination effects [25].

These approaches recover rich models of either the illuminant or the reflectance, but unlike the method that we propose, they cannot recover both. By contrast, the approach of [20] can. It is, however, not generic because it depends on a specific situation where the shadow of an object of known shape is projected onto an object also of known shape. Recent work [18,19] done simultaneously to ours [12] demonstrated that a complex lighting model, including multiple light-sources and specular properties, can be recovered given several images of an

object for which a 3-D model is available but, unlike in our approach, requires assuming that the lighting is the same in all images.

### 3 Linear Model for Multiple-Illuminant Radiometry

#### 3.1 The Lambertian Model

We model the effect of illumination as the sum of a Lambertian term from a distant point light source, and an ambient term. For a surface element of normal  $\mathbf{n}$  and albedo  $\alpha$  that projects into an image  $i$  at coordinates  $(u, v)$ , the 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  $\mathbf{l}_i$  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 “intensity.” The complete illumination model for each image is therefore given in terms of 4 parameters described by the *illumination vector*:  $\mathbf{L}_i = [l_{i1}, l_{i2}, l_{i3}, \mu_i]^T$ . Using the illumination vector, and the notation  $\mathbf{N} = [\mathbf{n}, 1]^t$ , 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. A camera is said to be *radiometrically calibrated* when the relationship between these quantities is known. This (non-linear) relationship can be recovered using multiple different known [4] or unknown [15] exposures of a fixed scene.

In this paper, we illustrate in a simple case how the radiometric (self) calibration can be attacked, rather than investigating it thoroughly. We will approximate the relationship by an affine transformation specified with a *radiometric scale factor*  $a_i$  and *radiometric offset*  $b_i$ :

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

We will see that within this limitation, there is enough information to recover linearly one of the parameters, the radiometric offset. When using our non-linear approach, a sufficiently large amount of data could allow to recover non-linear parametric terms.

#### 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. 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  $\mathbf{L}_i$ . 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. (2) :

$$I_{ij} = a_i \mathcal{L}_{ij} + b_i = a_i \alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i \quad (3)$$

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 can expect to solve the radiometric reconstruction problem.

### 3.4 Ambiguities and Minimal Data

Going back to the radiometric reconstruction problem, we notice from Eq. (3) that if:

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

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

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

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. Only the positive values of the global scale  $s$  are acceptable because of the physical constraint that the albedoes be positive. 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 photometrically uncalibrated cameras, the intensity of the illuminants and the radiometric scale factor of the camera cannot be distinguished without further assumptions. In the reminder of this paper, we will therefore take  $a_i = 1$ .

Taking these ambiguities into account, 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. In the calibrated case, we have unique solution when  $np \geq 4n + p - 1$ : 7 surface elements for 2 views, 6 surface elements for 3 views, 5 surface elements for 4 views and more.

## 4 A Linear Solution

Given a surface element of normal  $\mathbf{n}$  that projects into two images, each with a different light source, its albedo can be estimated using Eq. (3) as:

$$\alpha^k = \frac{I_k - b_k}{\mathbf{L}_k^T \mathbf{N}} \quad \text{or as} \quad \alpha^l = \frac{I_l - b_l}{\mathbf{L}_l^T \mathbf{N}} , \quad (4)$$

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 presumed constant, by expanding  $\alpha^k = \alpha^l$  we obtain:

$$\begin{bmatrix} I_k - b_k \\ -I_l + b_l \end{bmatrix}^T \begin{bmatrix} \mathbf{L}_l^T \\ \mathbf{L}_k^T \end{bmatrix} \mathbf{N} = 0 , \quad (5)$$

linear in the 12 variables:  $L_{ki}, L_{li}, L_{ki}b_l - L_{li}b_k, 1 \leq i \leq 4$  .

### 4.1 The Calibrated Case

Intensities can be transformed so that we can assume without loss of generality that  $a_i = 1$  and  $b_i = 0$ . If  $\mathbf{n}^T = [n_1, n_2, n_3]$ , Eq. (5) can be rewritten as:

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

$$\begin{aligned} \mathbf{U} &= [I_l n_1, I_l n_2, I_l n_3, I_l, -I_k n_1, -I_k n_2, -I_k n_3, -I_k]^T \\ \mathbf{f} &= [l_{k1}, l_{k2}, l_{k3}, \mu_k, l_{l1}, l_{l2}, l_{l3}, \mu_l]^T \end{aligned} \quad (6)$$

Combining the rows  $\mathbf{U}$  for each surface element provides a linear system of the form  $\tilde{\mathbf{U}}\mathbf{f} = 0$ . In practice, this is solved in the least-squares sense, taking into account the global scale factor on the illuminant  $\mathbf{f}$ :

$$\min_{\mathbf{f}} \|\tilde{\mathbf{U}}\mathbf{f}\| \quad \text{subject to} \quad \|\mathbf{f}\| = 1 \quad (7)$$

With more than two views, we solve simultaneously for all the illuminants, by concatenating the equalities of Eq. (5).

### 4.2 Minimal Data and Degeneracies

A surface visible in  $n$  views yields  $n - 1$  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  $\max(p(n - 1), 4n - 1)$ . To prove that this data is sufficient when the illuminants and surface normals are fully general, we built a specific system of linear equations for each value of  $p = 1..7$  and  $n = 1..5$ . The Maple<sup>TM</sup> computer algebra system checked the rank of the system, and confirmed that there is no rank loss.

There are two types of degeneracies, one which is caused by illuminants, the other by the object's shape. When all  $N$  illumination vectors are identical the matrix  $\tilde{\mathbf{U}}$  has rank  $4 \times (n - 1)$ . The larger the distance between illuminants is, the more stable the solution, as will be confirmed by Sec. 4.4. Besides this obviously expected degeneracy, we found no case of degeneracy caused by illuminants with generic surface normals. In particular for  $n = 5$ , the system does not lose rank in the generic case, despite the fact [16,22] that the space of images (in the common image plane, or equivalently, taken from the same viewpoint) is of dimension four.

For the matrix  $\tilde{\mathbf{U}}$  to have full rank, it is necessary to have at least seven surface elements with different orientations, and that all these orientations vectors be non-planar. In particular planes and cylinders constitute degenerate surfaces.

### 4.3 Recovering the Offsets

According to Sec. 3.4, in the uncalibrated case, the only radiometric calibration parameters which can be recovered are the offsets  $b_i$ . Eq. (5) is rewritten as:

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

$$\begin{aligned} \mathbf{V} &= [I_1 n_1, I_1 n_2, I_1 n_3, I_l, -I_k n_1, -I_k n_2, -I_k n_3, -I_k, n_1, n_2, n_3, 1]^T \\ \mathbf{g} &= [L_{k1}, L_{k2}, L_{k3}, L_{k4}, L_{l1}, L_{l2}, L_{l3}, L_{l4}, m_{kl1}, m_{kl2}, m_{kl3}, m_{kl4}]^T \end{aligned}$$

where we have used the notations:

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

After solving these linear equations, to recover the  $n$  offsets, we notice that:

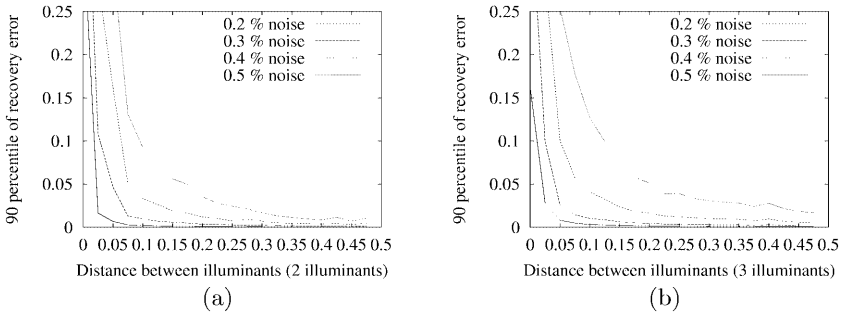
$$(L_{pj} L_{qi} - L_{pi} L_{qj}) b_k = m_{pqj} L_{ki} - m_{pqj} L_{ki} \quad (8)$$

After the  $L_{ij}$  and  $m_{ijk}$  are solved for, this equation can be used in a redundant way to recover the  $b_k$ .

### 4.4 Synthetic Experiments

To test our approach, we randomly draw values for the light source vectors, values for the surface normals that are such that the surface elements are front-lit, and values between 0 and 1 for the albedoes. We then compute the expected gray level values and add Gaussian noise. Finally, we use the normals and noisy gray-levels to recover estimated illuminants and compare them to the real values. To this end, we form "concatenated illuminant vectors" using the components in all images. The "distance" between two concatenated illuminant vectors  $\vec{V}_1$  and  $\vec{V}_2$  is:

$$D_{vect}(\vec{V}_1, \vec{V}_2) = 1 - \frac{\vec{V}_1 \cdot \vec{V}_2}{\|\vec{V}_1\| \|\vec{V}_2\|} \quad (9)$$



**Fig. 1.** Sensitivity of the linear algorithm to gray level noise. Each curve corresponds to a different amount of noise. (a) Using 2 images. (b) Using 3 images.

It is one when the two vectors are proportional to each other and zero when they are perpendicular.

Figure 1(a,b) shows the results of such experiments using 2 and 3 illuminants respectively. Each curve corresponds to a different value of the gray-level noise. For each noise value, we ran the algorithm 10000 times and computed both the true distance between illuminants<sup>1</sup> and the recovery error of Eq.(9). We then divided the range of distances between illuminants, that is the x-axis of the graphs 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 which is larger than 90% of these recovery errors. All these experiments were performed using 200 surface elements. We have verified experimentally that, as soon as the number of surface elements becomes larger than 50, the results become statistically independent of the actual number being used.

As expected, the recovery error increases with the amount of noise and also when the distance between the individual illuminants decreases. Note that increasing the number of illuminants from two to three, and beyond, does not bring any meaningful improvement.

## 5 Non-linear Solution: Radiometric Bundle Adjustment

The radiometric bundle adjustment method is the counterpart of the geometric bundle adjustment used in photogrammetry. 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:

<sup>1</sup> In the two-illuminant case, the distance is  $D_{vect}$  of Eq. (9) between the two individual illuminant vectors. In the three-illuminant case, it is the average distance of the three possible pairs of illuminants.



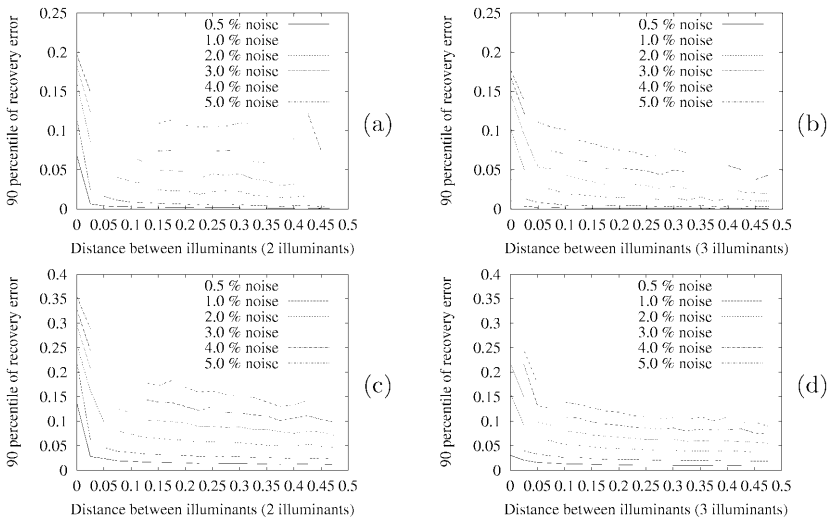
$$\sum_i \sum_j (I_{ij} - (\alpha_j \mathbf{L}_i^T \mathbf{N}_j + b_i))^2 \quad (10)$$

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.

The main drawback is that non-linear optimization is required to minimize the criterion of Eq. (10). However, as will be shown below, the increase in reliability is well worth the additional computational burden.

## 5.1 Synthetic Experiments

We ran experiments similar to those of Section 4.4, but using this non-linear approach and much higher noise levels. The results are depicted by Figure 2(a,b). As in Figure 1, the individual curves correspond to different levels of gray-level noise and 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.



**Fig. 2.** Sensitivity of the non-linear algorithm to gray level noise. Each curve corresponds to a different amount of noise. (a,b) 90 percentiles of recovery errors of illuminant direction using 2 and 3 images. (c,d) Root-mean square errors between true and recovered albedoes using 2 and 3 images.

The non-linear algorithm also recovers an estimated albedo value for each facet. In Figure 2(c,d), 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 2(a,b). Note that the graphs of both figures have exactly the same shapes, indicating that our distance measure between illuminants is also a good indicator of the quality of albedo recovery.

In practice, normals will only be known up to some precision. We have therefore verified in a similar manner that the results are 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 that do not at all conform to the Lambertian model, for example because some surface elements are specular or in shadow. To handle these, we adopt a RANSAC style approach [5] to robustly minimizing the criterion of Eq. (10). We have checked that it tolerates up to 15% of outliers [12].

## 5.2 Comparison with the SVD Approach

As discussed earlier, when enough images are available, albedoes and light source directions can be recovered, up to a Generalized Bas-Relief (GBR) ambiguity, using an SVD based-approach that does not require the use of normals [3]. Following this approach, we form the  $p \times n$  matrix  $X$  whose rows are the average gray levels of the  $p$  facets in each of the  $n$  images. We then use SVD to decompose it as

$$X = BS \quad , \quad (11)$$

where  $B$  is a  $p \times 5$  matrix in which the first three elements of all rows are proportional to the normals multiplied by the facet albedoes and  $S$  is a  $5 \times n$  matrix whose columns are the parameters of the light sources associated to each image. In an ideal no-noise case, for our model with ambient light source and additive camera offset, the matrix  $B$  can be taken to be the matrix with rows

$$\alpha_j n_{x_j}, \alpha_j n_{y_j}, \alpha_j n_{z_j}, \alpha_j, 1 \quad , \quad (12)$$

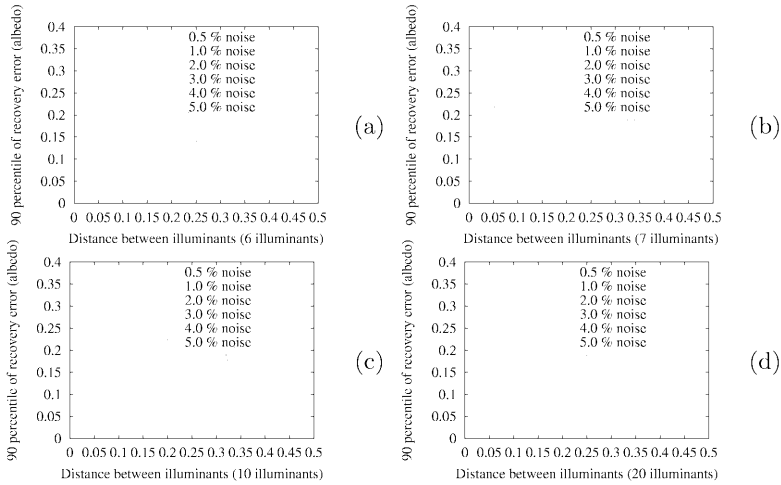
where  $\alpha_j$  is the albedo of facet  $j$  and  $\{n_{x_j}, n_{y_j}, n_{z_j}\}$  is its normal vector.<sup>2</sup> The GBR ambiguity comes from the fact that, for any invertible  $5 \times 5$  matrix  $T$ ,  $BT$  and  $T^{-1}S$  is also a solution of equation 11.

In our specific case, we do not know the albedoes but we do know the normals. We can therefore use them to resolve the ambiguity by computing a matrix  $T$  and the  $\alpha_j$  that jointly minimize:

$$\|BT - A\|^2 \quad (13)$$

---

<sup>2</sup> In the original SVD work [3],  $B$  is  $p \times 3$  or a  $p \times 4$  matrix. Here we use a  $p \times 5$  to account for the additive offset associated to each illuminant. This approach thus requires at least 5 images.



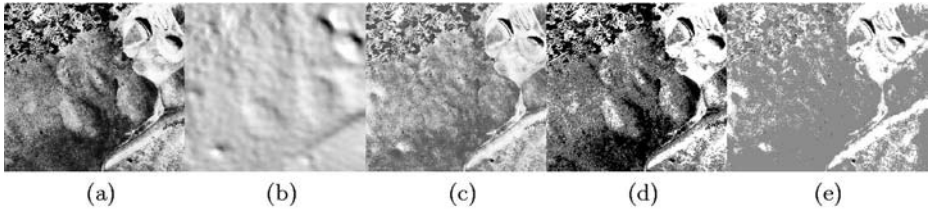
**Fig. 3.** Sensitivity of the SVD algorithm to gray level noise in terms of the 90 percentile of the root-mean square error between true and recovered albedoes. Each curve corresponds to a different amount of gray-level noise. (a) Using 6 images. (b) Using 7 images. (c) Using 10 images. (d) Using 20 images.

where  $A$  is the matrix whose rows are those defined by Equation 12. Because this system is homogeneous, this involves fixing one of the albedoes and finding the 25 unknown coefficients of  $T$  and the  $p - 1$  unknown albedoes. We then take the  $\alpha_j$  to be the recovered albedoes and  $T^{-1}S$  to be the matrix whose columns are the recovered illuminants  $L_i$ .

Figure 3 depicts the results of experiments similar to those of Figure 2(c,d), but using the SVD approach instead of our proposed method. Note that, even using 20 illuminants instead of the 2 or 3 of Figure 2(c,d), the recovery errors are higher for identical values of the distance between illuminants. Furthermore, the precision quickly degrades as the number of illuminants decreases. This is not surprising since the SVD method, unlike ours, does not use the normal information early in the computation.

### 5.3 Using Real Images

In Figure 4, we show one of 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 shape using a deformable surface technique, resulting in 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 4(b) shows the mesh shaded using the illuminant recovered for the first image. In Figure 4(c), we show the albedo at each pixel location that would produce the observed image gray-levels, assuming that the illuminant we computed is correct. The resulting



**Fig. 4.** Cartographic Modeling: (a) One of two images of a site taken at two different times of day. (b) The corresponding terrain, shaded using the illuminant recovered by our system. (c) The corresponding de-lighted image, stretched so that it has the same dynamic range as the original one. (d,e) Thresholded versions of the original and de-lighted image: The pixels whose gray level is inferior to 50 appear in black, those whose gray level is greater than 150 in white, and the rest in gray. In (e), unlike in (d), most of the grassy areas appear in solid gray, which is correct since the physical properties are roughly similar. This would make the task of a region-based segmentation algorithm much easier.

de-lighted image exhibits much less structure than the original one, even though both images have the same dynamic range: Shading effects have been greatly attenuated because they are accounted for by changes in shape. To illustrate this point, in Figure 4(d,e), we show versions of the original and de-lighted images coded using three gray-levels only: The pixels whose gray level is inferior to 50 are painted in black, those whose gray level is greater than 150 in white, and the rest in gray. In (e) unlike in (d), most of the grassy areas appear in gray, the shadows in black and the road pixels in white. In other words, a gray-level based segmentation algorithm would have a much easier time of the de-lighted image than of the original one.

In Figure 5, we show results in a case where we can use a CAD/CAM 3-D model of a target object but where there are strong specularities: We acquired four different images of an O2 computer by using a static camera and moving the computer before each exposure. We aligned the geometric model with each image by running a bundle-adjustment algorithm. This method yields the motion parameters from one view to the other and, thus, allows us to reduce the number of degrees of freedom: Only one illuminant is free and the others are taken to be rotated versions of it. Again, as shown in Figure 5(e), the recovered illumination estimate allows us to de-light the original image. And, as depicted by the graphs of Figure 5(f,g,h), the de-lighted gray-levels corresponding to object points having the same physical material, such as those on the vertical sides of the computer tower, are much more uniform than the original ones, except for specular pixels. As shown in Figure 5(i,j), it becomes much easier to segment meaningful regions from the de-lighted image than from the original one: In reality, the computer's base and the rest of the tower have two different and roughly constant albedoes. As a result, we can pick out these two regions in the de-lighted image using simple gray-level thresholds, which is impossible in the original images.

In Figure 6, we present a similar behavior in the more difficult case where we do not have an a-priori shape-model: The three images to the left were acquired using a static camera watching a moving head. We computed the head motion and reconstructed its shape using a stereo-based facial reconstruction technique [6] that is relatively insensitive to illumination changes. Because our geometric reconstruction method also yields motion parameters, as in the case of the images of Figure 5, we can assume that illuminants are rotated versions of each other. 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 Figure 6(d), the resulting surface is shaded using the illuminant we recover while Figure 6(e) shows the de-lighted image corresponding to image (b). For comparison purposes, in Figure 6(f), we show the corresponding portion of the original image, stretched to have the same dynamic range and with two scan-lines overlaid. In Figure 6(g), we plot the gray levels along both scan-lines. Each graph corresponds to one scan-line and contains two curves: One for the original gray levels in (f) and one for the de-lighted gray-levels in (e). Note that the de-lighted gray levels are again much flatter than the original ones: Most of the gray-level variations are accounted for by changes in shape and we are left with the almost constant albedoes one would expect on a face. This is true for most of the skin surface, except on the right side the nose and the cheeks: The recovered illuminant comes from the left side of the face, resulting in exceedingly bright pixels at places where cast shadows should be, but are not, modeled. As a result, the estimated albedoes are too large. Nevertheless, these de-lighted images can then be averaged to produce de-lighted texture-mapped models such as the one of Figure 6(h,i,j). These, in turn, could be re-lighted from a different direction and inserted in an animation.

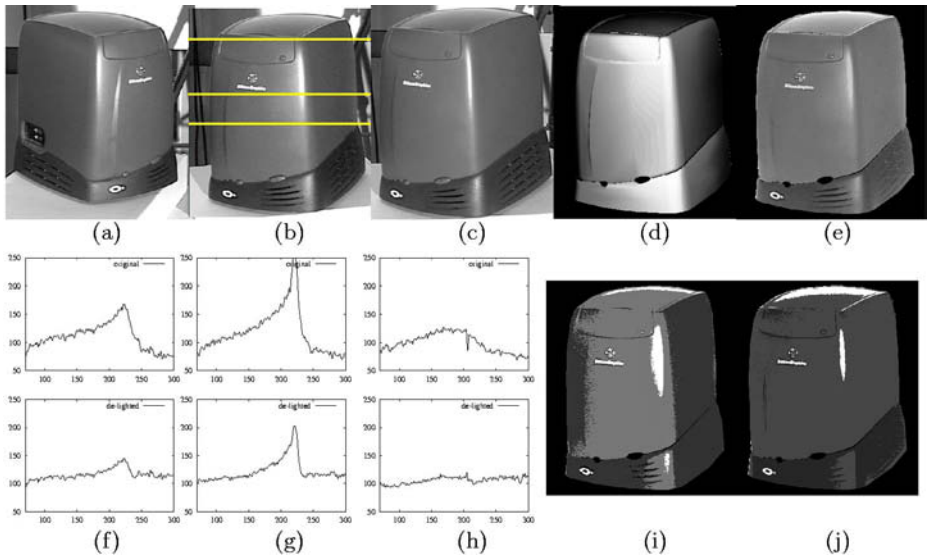
## 6 Conclusion

We have introduced a new methodology for recovering the illuminants and surface albedoes from multiple views taken with varying illuminants in the Lambertian case. 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.

The new theory indicates the minimal data as well as degenerate configurations and sets the limits for radiometric reconstruction. We proposed two different algorithmic implementations:

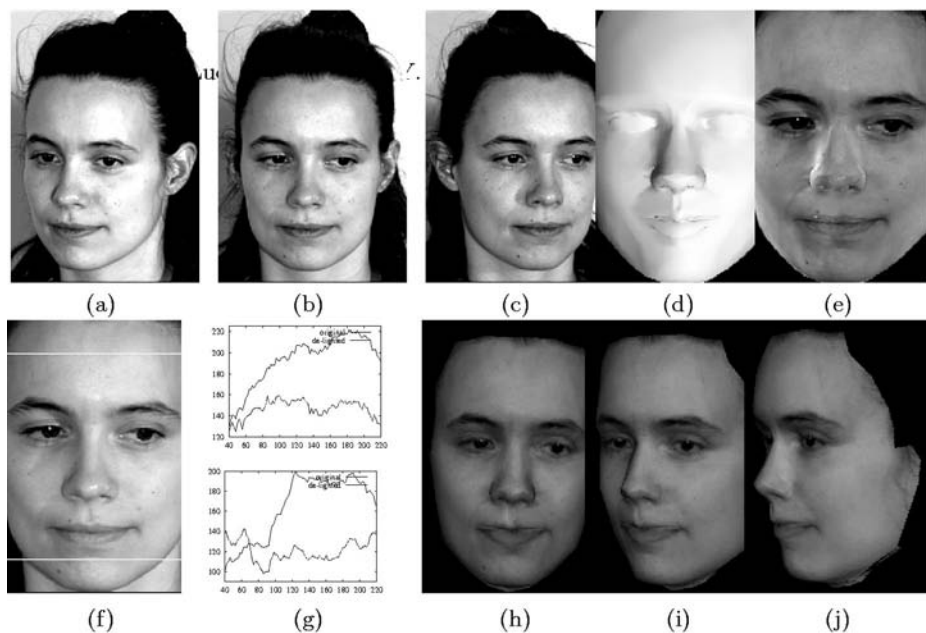
1. **A linear algorithm** It allows fast computation, but is sensitive to noise.
2. **A radiometric bundle-adjustment algorithm.** It involves non-linear optimization that makes it slower but also much more robust.

Using synthetic data, we have quantified the sensitivity of the algorithms to image noise, outliers, imprecision of surface normals, distance between illuminants,



**Fig. 5.** Modeling an object with known geometry and strong specularities: (a,b,c) Three image of an O2 computer acquired using a static camera. The computer was moved before each shot. (d) CAD model shaded using the recovered illuminant for image (b). (e) The corresponding de-lighted image, stretched so that it has the same dynamic range as the original one. (f,g,h) Top: Gray-levels along the three horizontal scan-lines overlaid in (b). Because of illumination effects, the gray levels vary a lot, even though, in reality, the albedo remains almost constant. Bottom: Gray-levels along the same scan-lines in the de-lighted image (e). They are almost constant, except for the specular pixels that produce a peak. (i,j) Thresholded versions of the original and de-lighted image: The background appears in black, the foreground pixels whose gray level is inferior to 80 appear in dark gray, those whose gray level is greater than 80 and smaller than 140 in light gray, and the rest in white.

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 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 Lam-



**Fig. 6.** Facial Reconstruction: (a,b,c) Three image of a moving head acquired using a static camera. (d) Close-up view of the reconstructed head model, shaded using the recovered illuminant. (e) The corresponding de-lighted image. (f) Corresponding part of image (b) stretched so that it has the dynamic range (e). (g) The gray-levels along the two scan-lines overlaid in (f). Each graph, contains two curves: One for the original gray levels in (f) and one for the de-lighted gray-levels in (e). (h,i,j) De-lighted texture-mapped view of the 3-D model.

bertian 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 but treat them 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, focus will be on introducing shadows, specular models, and multiple light sources into the proposed framework.

## References

1. E. Angelopoulou and J. Williams. Photometric surface analysis in a tri-luminal environment. In *ICCV*, pages 442–447, 1999.
2. R. Baribeau, M. Rioux, and G. Godin. Color reflectance modeling using a polychromatic laser range sensor. *PAMI*, 14(2):263–269, 1992.
3. P.N. Belhumeur, D.J. Kriegman, and A.L. Yuille. The bas-relief ambiguity. *IJCV*, 35(1):33–44, November 1999.
4. P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *SIGGRAPH*, 31:369–378, 1997.
5. 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.
6. P. Fua. Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data. *International Journal of Computer Vision*, 38(2):153–171, July 2000.
7. A.S. Georghiadis, 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.
8. H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *JOSA-A*, 11(11):3079–3089, November 1994.
9. B.K.P. Horn. *Robot Vision*. MIT Press, 1986.
10. B.K.P. Horn and M.J. Brooks. *Shape from Shading*. MIT Press, 1989.
11. K. Ikeuchi and K. Sato. Determining reflectance properties of an object using range and brightness images. *PAMI*, 13(11):1139–1153, 1991.
12. Q.T. Luong, P. Fua, and Y. G. Leclerc. The Radiometry of Multiple Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):19–33, January 2002.
13. S. Marschner. *Inverse rendering for computer graphics*. PhD thesis, Cornell University, 1998.
14. S. Marschner and D. Greenberg. Inverse lighting for photography. In *IST/SID Fifth Color Imaging Conference*, pages 262–265, 1997.
15. T. Mitsunaga and S.K. Nayar. Radiometric self calibration. In *CVPR*, pages I:374–380, 1999.
16. Y. Moses. *Face recognition: generalization to novel images*. PhD thesis, The Weizmann Institute of Science, Israel, 1993.
17. N. Mukawa. Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *ICCV*, pages 507–512, 1990.
18. K. Nishino, Z. Zhang, and K. Ikeuchi. Determining Reflectance Parameters and Illumination Distribution from a Sparse Set of Images for View-Dependent Image Synthesis. In *International Conference on Computer Vision*, pages 599–606, Vancouver, Canada, July 2001.
19. R. Ramamoorthi and P. Hanrahan. A signal processing framework for inverse rendering. In *SIGGRAPH*, pages 117–128, 2001.
20. 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 *ICCV*, pages 875–883, 1999.
21. M. Sato, Y. Wheeler and K. Ikeuchi. Object shape and reflectance modeling from observation. In *SIGGRAPH*, pages 379–387, 1997.



22. A. Sashua. On photometric issues in 3d visual recognition from a single 2d image. *IJCV*, 21(1-2):99–122, 1997.
23. W. Silver. *Determining shape and reflectance using multiple images*. PhD thesis, MIT, Cambridge, MA, 1990.
24. R.J. Woodham. Photometric method for determining surface orientation from multiple images. *OptEng*, 19(1):139–144, 1980.
25. Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. *SIGGRAPH*, pages 215–224, August 1999.
26. 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. *IJCV*, 35(3):1–20, 1999.
27. Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *PAMI*, 13(7):680–702, 1991.