

# Building Illumination Coherent 3D Models of Large-Scale Outdoor Scenes

Alejandro Troccoli · Peter Allen

Received: 19 March 2007 / Accepted: 1 October 2007 / Published online: 13 November 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Systems for the creation of photorealistic models using range scans and digital photographs are becoming increasingly popular in a wide range of fields, from reverse engineering to cultural heritage preservation. These systems employ a range finder to acquire the geometry information and a digital camera to measure color detail. But bringing together a set of range scans and color images to produce an accurate and usable model is still an area of research with many unsolved problems. In this paper we address the problem of how to build illumination coherent integrated texture maps from images that were taken under different illumination conditions. To achieve this we present two different solutions. The first one is to align all the images to the same illumination, for which we have developed a technique that computes a relighting operator over the area of overlap of a pair of images that we then use to relight the entire image. Our proposed method can handle images with shadows and can effectively remove the shadows from the image, if required. The second technique uses the ratio of two images to factor out the diffuse reflectance of an image from its illumination. We do this without any light measuring device. By computing the actual reflectance we remove from the images any effects of the illumination, allowing us to create new renderings under novel illumination conditions.

**Keywords** 3D modeling · Inverse rendering · Texture acquisition

---

A. Troccoli (✉)  
2701 San Tomas Expy, Santa Clara, CA 95054, USA  
e-mail: atroccoli@acm.org

P. Allen  
Department of Computer Science, Columbia University,  
1214 Amsterdam Ave. MC 401, New York, NY 10027, USA

## 1 Introduction

This paper presents and compares two different methods for creating illumination coherent models of large scale outdoor scenes. We concentrate in the problem of texture fusion, i.e. how to combine multiple texture images with 3D range data to produce photo-realistic renderings in a physically plausible way. In particular, we analyze the problem of combining intensity images of outdoor environments captured under different (unknown) illumination conditions without the use of light probing devices. In this way, we keep model acquisition as simple as possible.

First, we will show how to solve this problem by applying a combination of relighting and de-shadowing operations. The relighting operation brings two images to the same illumination; the de-shadowing operation removes any shadows that are present in the image. This approach is a significant departure from the traditional methods of range and intensity image rendering that either use a weighted average of the input images (Pulli et al. 1997; Debevec et al. 1996; Buehler et al. 2001) or apply a global color correction matrix (Agathos and Fisher 2003; Bannai et al. 2004).

Secondly, we look into the problem of illumination and texture factorization, where our goal is to factor the illumination from the texture and solve for the shading of each image and the surface reflectance. By doing so, we obtain an illumination-free texture map from a pair of images and the object geometry without prior recording or calibration of the incident illumination. This technique falls into the category of inverse-rendering techniques, since we are measuring scene properties from images and objects of known geometry. While measuring surface reflectance of an object of known geometry under controlled and calibrated illumination has proved to produce very good results as shown by Debevec et al. (2000) and Lensch et al.

(2003), working with unknown illumination is yet an open problem. Typically, to handle unknown illumination it is assumed that the material properties of the object are homogeneous over the whole surface (Ikeuchi and Sato 1991; Ramamoorthi and Hanrahan 2001a). In addition, when dealing with textured objects the problem of recovering both texture and illumination becomes unconstrained. As noted by Ramamoorthi and Hanrahan (2001b) in the development of a signal-processing approach for inverse rendering, lighting and texture can not be factored without resorting to active methods or making prior assumptions of their expected characteristics. Our method achieves this factorization by assuming diffuse surface reflectance.

Both of the algorithms we present rely on the images overlapping each other. Our first algorithm computes a relighting operator by analyzing intensity values in the area of overlap of a pair of images: a source image to be relighted, and a target image whose illumination we want to match. By applying this operator over the non-overlapping region of the source image, we transform its color intensities in such a way that they are consistent with the colors observed in the target image. The operator we compute to perform the relighting is the ratio of intensity values per surface orientation, which under certain assumptions, is consistent over all points with the same orientation and only depends on the illumination of the scene. The texture factorization algorithm also relies on this orientation-consistency property. However, in this case we use the ratio image to solve for the irradiance map of each image.

Our main motivation is to push forward the automation of the 3D modeling pipeline for large-scale outdoor scenes. Our work has been done in collaboration with archaeologists, art historians and urban planners, to provide accurate and photorealistic 3D models that can serve as documentation for an archaeological site (Allen et al. 2004), teaching tools for the classroom or research material. During the past years we have worked in automatic methods and tools for range image registration (Allen et al. 2003), automatic image registration (Troccoli and Allen 2004), and more recently in texture normalization (Troccoli and Allen 2005, 2006). This paper is an extended version of these works with several additions:

- Extension of the relighting algorithm for improved penumbra handling and shadow removal.
- Addition of a “sun plus sky” illumination model to our previous factorization algorithm.
- Generation of new results on outdoor scenes showing the performance of our techniques in relighting, de-shadowing and texture factorization.

These extensions were developed for better handling of real world outdoor scenes.

## 2 Related Work

The creation of an illumination coherent texture map from multiple images can be a difficult problem for several reasons. First, the images are taken from different points of view, which means they show different degrees of foreshortening and view-dependent effects. Secondly, the illumination and camera parameters might have changed from one shot to the next, resulting in images with color mismatches. To cope with these difficulties, early work in model acquisition assumed all images were acquired under constant illumination, in which case the problem reduces to finding the optimal blending function (Debevec et al. 1996; Pulli et al. 1997; Buehler et al. 2001; Wang et al. 2001). When the constant illumination assumption is impractical, as it sometimes is for outdoor scenes, then some color adjustment is required. Agathos and Fisher (2003) compute a global  $3 \times 3$  color correction matrix from pixel intensity constraints obtained from the overlap region of two images, taking care to discard pixels in shadow or highlight areas by thresholding. Once the matrix is computed, one of the images is color-corrected. This pairwise adjustment is extended by Bannai et al. (2004) to multiple input views. However, the color transform approach is good for fixing camera parameters and small changes in the illumination, such as chromatic or intensity changes, but it is unclear the extent to which these methods can handle large illumination variations, since these methods do not make use of the geometry information. Marschner and Greenberg (1997) solve for the illumination of the scene by using a set of basis images synthetically created with a ray tracer. Once the illumination of the scene has been estimated, they relight the original photograph using the ratio image of a synthetic rendering of the scene under a new illumination divided by a similar rendering under the solved illumination. Beaudesne and Roy (2003) compute a relighting operator based on image ratios that they apply to a pair of input images to create a new set of images of faces with consistent lighting. Our relighting algorithm follows a similar idea to Beaudesne and Roy (2003), which we extend to outdoor scenes to handle shadows and perform de-shadowing. In recent work by (Xu et al. 2006), the authors solve for diffuse reflectance of large environments through analysis of the scanner’s returned intensity value in combination with color-balancing techniques on the input color images.

A different approach is to compute a surface reflectance map, obtaining an illumination-free texture representation that can be used to create new renderings under novel lighting conditions. Without any knowledge about the illumination, however, this problem can become intractable. In laboratory conditions, the position of the light sources can be measured; with such measurements (Debevec et al. 2000;

Lensch et al. (2003) show how to obtain a reflectance function once the light positions are known. To extend this idea to outdoor scenes, Yu and Malik (1998) measure the illumination of the scene by acquiring photographs of the sky and the surrounding environments. More recently, Debevec et al. (2004) introduced a novel lighting measurement apparatus that can record the high dynamic range of both, sunlit and cloudy environments, using a set of specular and diffuse calibrated spheres. With the captured illumination, they solve for spatially varying surface reflectance. Our texture factorization algorithm solves for a diffuse texture map without the need of a light probe.

In the domain of face recognition, Shashua and Riklin-Raviv (2001) define the quotient image of two different faces as the ratio of their albedos, which they use for face recognition under varying illumination. The main assumption in their work is that all faces have the same geometry and the variation in their appearance under the same point light source depends on the diffuse reflectance. Wang et al. (2004) take this method a step further and generalize it to faces illuminated by non-point light sources. Our methodology is different and built on the premise that we are imaging the same scene under changing illumination, thus the diffuse reflectance for the same scene points does not change, and the resulting ratio image is texture free.

### 3 Problem Definition

We define our problem assuming a 3D model of our scene has already been acquired, registered and merged into a triangle mesh. Then, given:

1.  $\mathcal{G}$  the geometry of the scene,
2.  $\mathcal{I} = \{I_1, I_2, \dots, I_n\}$  a set of overlapping photographs of the scene captured under illumination conditions  $\mathcal{L} = \{L_1, L_2, \dots, L_n\}$ ,
3.  $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$  the set of camera projection matrices for each image,

our goal is to create an illumination coherent photorealistic textured model of  $\mathcal{G}$ . We achieve this by normalizing the textures in one of two ways:

- Relighting all images in  $\mathcal{I}$  to a single illumination  $L_r$ . The output of our algorithm is in this case a new set of images  $\mathcal{I}^r = \{I_1^r, I_2^r, \dots, I_n^r\}$ . In addition, our technique can remove any shadows present in the images.
- Factoring illumination from the texture. In this case we solve for the relative illuminations  $L_i$  in the form of irradiance maps of each image and compute a diffuse albedo map of the scene. We solve the factorization problem by assuming one of three different illumination models: (1) an object illuminated by a point light source, (2) non-point light source illumination, (3) outdoor illumination

represented as a combination of a point light source and an ambient component.

Whichever of the two approaches we take, the images need not to be taken from the same viewpoint. Since we have the geometry of the scene and the projection matrices, it is possible to warp any two overlapping images to the same view. For simplicity, in the discussions that follows we will consider a single pair of images acquired from the same view. The input to the texture normalization procedure is a ratio image  $R(x, y)$ , which we compute by taking the quotient of the two input images, and a normals image  $\mathbf{n}(x, y)$  which gives the normal for each pixel and that is generated by ray-tracing the geometry of the object.

### 4 The Relighting Approach

Our relighting algorithm is based on the fact that the ratio of two images of a diffuse object is texture-free. In this section we will justify this statement and establish the conditions under which it holds. We begin with the reflected radiance equation, which describes how a surface reflects the incoming light. The light reflected by a surface point  $\mathbf{x}$  in the direction  $(\theta_o, \phi_o)$  is given by:

$$B(\mathbf{x}, \theta_o, \phi_o) = \int_{\Omega_i} L(\mathbf{x}, \theta_i, \phi_i) f_r(\mathbf{x}, \theta_i, \phi_i, \theta_o, \phi_o) \cos \theta_i d\omega_i, \quad (1)$$

where  $L$  is the incident illumination and  $f_r$  the BRDF at surface point  $\mathbf{x}$ . For Lambertian surfaces, (1) becomes:

$$B(\mathbf{x}) = \rho(\mathbf{x}) \int_{\Omega_i} L(\mathbf{x}, \theta_i, \phi_i) \cos \theta_i d\omega_i, \quad (2)$$

where  $\rho(\mathbf{x})$  is the surface albedo at  $\mathbf{x}$ . For simplicity, define the irradiance at  $\mathbf{x}$  as

$$E(\mathbf{x}) = \int_{\Omega_i} L(\mathbf{x}, \theta_i, \phi_i) \cos \theta_i d\omega_i. \quad (3)$$

Then the reflected radiance (2) turns into:

$$B(\mathbf{x}) = \rho(\mathbf{x}) E(\mathbf{x}). \quad (4)$$

Note that there is still a dependence on the surface position, on both the albedo  $\rho$  and the irradiance  $E$ . Under the assumptions that the illumination of the scene is distant, the effects of interreflections are negligible and there are no shadows, then (4) can be reparametrized by the surface normal at  $\mathbf{x}$ , which we denote as  $\mathbf{n}_x$ . After reparametrization by surface orientation the above (4) becomes:

$$B(\mathbf{x}) = \rho(\mathbf{x}) E(\mathbf{n}_x). \quad (5)$$

Now consider two different observations of the same surface point  $\mathbf{x}$  under different and unknown illuminations  $L_i$

and  $L_j$ . The image irradiance values for these two observations are:

$$B_i(\mathbf{x}) = \rho(\mathbf{x})E_i(\mathbf{n}_\mathbf{x}),$$

$$B_j(\mathbf{x}) = \rho(\mathbf{x})E_j(\mathbf{n}_\mathbf{x}).$$

Taking the ratio of these two expressions we obtain:

$$R_{ij}[\mathbf{n}_\mathbf{x}] = \frac{E_i(\mathbf{n}_\mathbf{x})}{E_j(\mathbf{n}_\mathbf{x})}, \tag{6}$$

since the albedo terms cancel out.  $R$  is an irradiance ratio, and is only dependent on the surface orientation at  $\mathbf{x}$ . Thus, the irradiance ratio of a Lambertian object under the previously stated illumination assumptions is orientation-consistent, meaning that two surface points with the same orientation will share the same value. The union of the irradiance ratios over all possible orientations defines an irradiance ratio map (IRM). An irradiance ratio map can be built from the quotient of the overlapping region of two images.

### 4.1 Relighting Using Irradiance Ratios

It can be verified that given the image irradiance of a surface point  $\mathbf{x}'$  under illumination  $L_j$  we can compute its image irradiance under illumination  $L_i$  by taking the product with the corresponding irradiance ratio value, as follows:

$$B_i(\mathbf{x}') = B_j(\mathbf{x}')R_{ij}[\mathbf{n}_{\mathbf{x}'}]. \tag{7}$$

This is verified by substitution of (5) and (6) into (7). Therefore, this defines a relighting operator based on the orientation-consistency assumptions that allows us to relight an image  $I_j$  to the illumination of  $I_i$ . The only requirement is to have sufficient surface orientations in the area of overlap of  $I_i$  and  $I_j$  to be able to relight the non-overlapping region of  $I_j$ .

### 4.2 Relighting in the Presence of Shadows

Architectural scenes can contain structures that cast shadows. When shadows are present, orientation-consistency does no longer hold because there can exist two points  $\mathbf{x}$  and  $\mathbf{x}'$  with the same surface orientation that do not have the same intensity ratio. This will happen if in either of  $I_i$  or  $I_j$  one of these points is in shadow and the other one is not. However, it is possible to extend the relighting algorithm ratios to handle shadows. For a scene illuminated by a single source plus smooth ambient illumination (e.g. sun plus sky) there exists a shadow mapping function  $\mathcal{S} : \mathcal{G} \rightarrow [0, 1]$ , that assigns a value of 1 to those scenes points completely shadowed, a value of 0 to those scene points that are completely lit by the source, and an intermediate value in the

range (0, 1) to those points that are in the penumbra regions. Momentarily ignoring penumbra regions,  $\mathcal{S}$  is a binary function that partitions the scene into two sets: a set  $\mathcal{G}_0$  of points lit by the source and a set  $\mathcal{G}_1$  of shadowed points. In addition, in terms of surface orientations, points with the same surface orientation will also be partitioned in two sets. Therefore, when taking the ratio of two images  $I_i$  and  $I_j$ , the surface orientations in the scene will be partitioned into four different sets  $\{\mathcal{G}_{00}, \mathcal{G}_{01}, \mathcal{G}_{10}, \mathcal{G}_{11}\}$  according to the values of the shadow bits  $\mathcal{S}_i$  and  $\mathcal{S}_j$ . We can now redefine the orientation-consistency property for scenes with shadows: when orientation-consistency holds, two points with the same surface orientation and same shadow bit value will show the same intensity ratio. Under these new conditions, we can compute four different IRMs from the ratio image:  $R_{ij}^{00}, R_{ij}^{01}, R_{ij}^{10}$  and  $R_{ij}^{11}$ . The relighting equation for a point  $\mathbf{x}$  becomes:

$$B_i(\mathbf{x}) = B_j(\mathbf{x})R_{ij}^{s_i s_j}[\mathbf{n}_\mathbf{x}], \tag{8}$$

where  $s_i$  and  $s_j$  are the shadow bits of  $\mathbf{x}$  under illuminations  $L_i$  and  $L_j$ . This expression is almost identical to (7), with the exception of the index into one of the four computed IRMs, and is well defined for binary values of  $s_i$  and  $s_j$ . The points for which  $s_i$  or  $s_j$  are not in  $\{0, 1\}$  are points in the penumbra regions. Penumbra regions are transitions from shadow to non-shadow regions (or vice versa) in which the irradiance varies gradually. To deal with points in penumbra we can generalize the relighting operator for real values of  $s_i$  and  $s_j$  in the following manner: first, we define the four base cases for binary shadow values to be equal to the measured data:

$$R_{ij}[\mathbf{n}_\mathbf{x}, 0, 0] = R_{ij}^{00}[\mathbf{n}_\mathbf{x}], \tag{9}$$

$$R_{ij}[\mathbf{n}_\mathbf{x}, 0, 1] = R_{ij}^{01}[\mathbf{n}_\mathbf{x}], \tag{10}$$

$$R_{ij}[\mathbf{n}_\mathbf{x}, 1, 0] = R_{ij}^{10}[\mathbf{n}_\mathbf{x}], \tag{11}$$

$$R_{ij}[\mathbf{n}_\mathbf{x}, 1, 1] = R_{ij}^{11}[\mathbf{n}_\mathbf{x}]. \tag{12}$$

We recall now from the definition of the irradiance ratio, that we can write the IRMs in terms of its constituent components:

$$R_{ij}^{00}[\mathbf{n}_\mathbf{x}] = \frac{E_i^0(\mathbf{n}_\mathbf{x})}{E_j^0(\mathbf{n}_\mathbf{x})}, \quad R_{ij}^{01}[\mathbf{n}_\mathbf{x}] = \frac{E_i^0(\mathbf{n}_\mathbf{x})}{E_j^1(\mathbf{n}_\mathbf{x})},$$

$$R_{ij}^{10}[\mathbf{n}_\mathbf{x}] = \frac{E_i^1(\mathbf{n}_\mathbf{x})}{E_j^0(\mathbf{n}_\mathbf{x})}, \quad R_{ij}^{11}[\mathbf{n}_\mathbf{x}] = \frac{E_i^1(\mathbf{n}_\mathbf{x})}{E_j^1(\mathbf{n}_\mathbf{x})},$$

where  $E^0$  and  $E^1$  refer to non-shadow and shadow irradiance respectively. We first define the irradiance in the penumbra regions by linear interpolation of the respective shadow and non-shadow irradiance values:

$$E_i[\mathbf{n}_\mathbf{x}, s_i] = E_i^1(\mathbf{n}_\mathbf{x})s_i + E_i^0(\mathbf{n}_\mathbf{x})(1 - s_i). \tag{13}$$

Now, we can define the relighting operator for real values of  $s_i$  and  $s_j$  as:

$$R_{ij}[\mathbf{n}_x, 0, s_j] = \frac{E_i^0(\mathbf{n}_x)}{E_j^1(\mathbf{n}_x)s_j + E_j^0(\mathbf{n}_x)(1 - s_j)}, \tag{14}$$

$$R_{ij}[\mathbf{n}_x, 1, s_j] = \frac{E_i^1(\mathbf{n}_x)}{E_j^1(\mathbf{n}_x)s_j + E_j^0(\mathbf{n}_x)(1 - s_j)}, \tag{15}$$

$$R_{ij}[\mathbf{n}_x, s_i, s_j] = \frac{E_i^1(\mathbf{n}_x)s_i + E_i^0(\mathbf{n}_x)(1 - s_i)}{E_j^1(\mathbf{n}_x)s_j + E_j^0(\mathbf{n}_x)(1 - s_j)}. \tag{16}$$

Finally, we write the above equations in terms of the measured data:

$$R_{ij}[\mathbf{n}_x, 0, s_j] = \frac{1}{\frac{1}{R_{ij}^{01}[\mathbf{n}_x]}s_j + \frac{1}{R_{ij}^{00}[\mathbf{n}_x]}(1 - s_j)}, \tag{17}$$

$$R_{ij}[\mathbf{n}_x, 1, s_j] = \frac{1}{\frac{1}{R_{ij}^{11}[\mathbf{n}_x]}s_j + \frac{1}{R_{ij}^{10}[\mathbf{n}_x]}(1 - s_j)}, \tag{18}$$

$$R_{ij}[\mathbf{n}_x, s_i, s_j] = R[\mathbf{n}_x, 1, s_j]s_i + R[\mathbf{n}_x, 0, s_j](1 - s_i). \tag{19}$$

We can now state the generalized relighting equation for scenes with shadows as:

$$B_i(\mathbf{x}) = B_j(\mathbf{x})R[\mathbf{n}_x, S_i(\mathbf{x}), S_j(\mathbf{x})]. \tag{20}$$

### 4.3 De-Shadowing

De-shadowing is a variation of the relighting problem, in which the resulting relighted image is shadow-free. It is straightforward to convert the relighting (20) into a de-shadowing equation by requiring all surface points in  $\mathcal{G}$  to be lit under the target illumination  $L_i$ . This is equivalent to re-defining  $S_i$  to map all points to 0. Hence, the de-shadowing equation is:

$$B_i(\mathbf{x}) = B_j(\mathbf{x})R[\mathbf{n}_x, 0, S_j(\mathbf{x})]. \tag{21}$$

We can also consider the problem of self de-shadowing, in which we remove the shadows of an image without recurring to relighting a second image. By using the computed IRMs over a pair of images, it is possible to define a de-shadowing IRM as:

$$R_i^0 = \frac{E_i^0}{E_j^0} \times \frac{E_j^0}{E_i^1} = \frac{R_{ij}^{00}}{R_{ij}^{10}}, \tag{22}$$

and the self de-shadowing equation is:

$$B_i^0(\mathbf{x}) = B_i(\mathbf{x})R_i^0[\mathbf{n}_x]S_i(\mathbf{x}) + B_i(\mathbf{x})(1 - S_i(\mathbf{x})). \tag{23}$$

### 4.4 Extending to Multiple Images

To work with multiple images, each acquired under a different illumination, we proceed chaining pairwise operations. First we select the image  $I_r$  whose illumination we will convert all other images to. Then, for every image that overlaps with  $I_r$  we can proceed as already explained. If there is an image  $I_j$  that does not overlap, then we apply a multi-step relighting. For example, if image  $I_i$  overlaps with both  $I_r$  and  $I_j$ , then we can first convert  $I_j$  to the illumination of  $I_i$  and then to the illumination of  $I_r$ , using the previously computed IRMs. The problem we might run into is that there might not be enough surface orientations in the respective areas of overlap to relight all points in the image.

If we have multiple images, but there exist a subset of these images that were all acquired under the same illumination, then we can compute the IRMs using the contributions from all images in this subset.

## 5 The Complete Relighting Algorithm

### 5.1 Data Acquisition

For the acquisition of the geometry, we have used a Leica HDS 3000 range finder. We acquire a set of range images and register them using a combination of fiduciary targets that are measured with a total station device. After the point clouds are registered, we build a triangular mesh using Vrip-Pack (Curless and Levoy 1996), which uses a 3D volume to accumulate a signed distance function to the surface. This 3D volumetric representation can require large amounts of memory. If we ran out of memory, we partition the 3D volume into two or more smaller volumes creating a triangle mesh for each. For example, for the model of the church of St. Marie at Chappes, we partitioned the volume in two.

Our input images are high-dynamic range (HDR) linearized images. These are built from a set of exposure-bracketed raw images, which are linear with respect to the light intensity. To build the final HDR image we compute the exposure ratios, and combine the input images using a weighted average. It is worth noting that the linearity assumption for the input images can be removed if the radiometric calibration of the camera is known, but since the camera we used produced uncompressed raw linear images, we found these more suitable than other compressed non-linear formats.

After we have created the HDR images, we compute the camera projection matrices using a point-and-click software tool that we have developed for that purpose (Troc coli 2006b). With the collection of high-dynamic range images taken under different illuminations, the geometry of the scene and the camera’s projection matrices we begin our relighting procedure. For a given pair of overlapping images, we carry out the following steps:

1. View alignment. If the images we are acquired from different view points, we warp one to the view of the other one so that same pixel locations correspond to the same scene point.
2. Shadow detection. We find the shadows using thresholding, as described next.
3. IRM computation. We iterate over all pixels and compute the 4 irradiance ratio maps. For each pixel, we obtain the surface normal of the corresponding scene point using ray-tracing.
4. Shadow map update. We refine the shadow function in the penumbra regions.
5. Relighting and de-shadowing.

## 5.2 Shadow Detection

The relighting algorithm requires a shadow map  $\mathcal{S}_i$  associated to each image  $I_i$ . We thus need a shadow detection method that assigns each pixel in  $I_i$  a real value in  $[0, 1]$ , where 0 means the corresponding point in the scene is lit, 1 means it is completely shadowed, and an intermediate value means the point is in a penumbra region.

It is well known that shadow detection from a single image is a difficult problem, and for this reason, many researchers perform shadow detection on image sequences or video (Chuang et al. 2003). In our application, knowledge of the scene geometry can help find the shadows if the position of the light source is known. Nevertheless, scanned geometry will not yield shadows that are correct to pixel level accuracy because of holes in the geometry and the effects of coarse sampling. For this reason, an image-based method for shadow detection is necessary. We employ a user-assisted method, but other shadow segmentation methods (Funkalea and Bajcsy 1995; Salvador et al. 2004) could be used instead. The user selects regions of the image and a pair of luminance thresholds  $s_0$  and  $s_1$ . All pixels with a luminance value that are below  $s_0$  are marked as shadowed, all pixels with luminance above  $s_1$  are marked as lit, and pixels in between are labeled as penumbra pixels with a value of  $(l - s_0)/(s_1 - s_0)$  where  $l$  is the pixel luminance. Before thresholding we filter the images using bilateral filtering (Tomasi and Manduchi 1998) to remove the high frequency effects of texture. The bilateral filter is designed to average spatially close pixels that are similar to each other. This similarity condition acts as an edge-stopping function and overcomes the traditional problem of edge-blurring that is common of Gaussian filtering.

## 5.3 IRM Computation

Values assigned to penumbra regions are ignored during the data collection for computing the IRMs. To compute the four IRMs we iterate over all pixels, computing the ratio of the

two images. For every pixel, we look up its surface normal and shadow bits. If the pixel has been labeled as penumbra in either of the images, we ignore it. Otherwise, we use the shadow bits to establish which of the four IRMs the pixel contributes to. Each IRM is stored in a 2D representation of the Gaussian sphere. The surface normal is used to index the corresponding entry to the current pixel. The final value stored in each IRM entry is the average of all values contributing to it. In addition, we compute one IRM for each of the three color channels.

## 5.4 Shadow Map Update

After the IRM data has been gathered, it is possible to update the shadow masks to find the correct values for the penumbra regions. This operation will work on all pixels in the area of overlap of the two images. The basic idea is to update the shadow map by comparing the true irradiance ratio at each pixel with the ratio obtained from the computed IRMs. Given images  $I_i$  and  $I_j$ , the current shadow masks  $\mathcal{S}_i$  and  $\mathcal{S}_j$ , and the computed IRMs  $\{R_{ij}^{00}, R_{ij}^{01}, R_{ij}^{10}, R_{ij}^{11}\}$ , we update both shadow masks  $\mathcal{S}_i$  and  $\mathcal{S}_j$ . To update  $\mathcal{S}_j$  we iterate over all the pixels applying the following rules:

1. Lookup  $\mathcal{S}_i(\mathbf{x})$  and  $\mathcal{S}_j(\mathbf{x})$  and the surface normal  $\mathbf{n}_x$ .
2. If  $\mathcal{S}_j(\mathbf{x})$  is either 1 (non-shadowed) or 0 (shadowed), skip the current pixel because only pixels labeled as penumbra will be updated.
3. Based on the value of  $\mathcal{S}_i(\mathbf{x})$  solve for the shadow mask  $s_j$  using (12) if  $\mathcal{S}_i(\mathbf{x})$  is 0, or (18) if  $\mathcal{S}_i(\mathbf{x})$  is 1. If the computed value of  $s_j$  is outside the range  $[0, 1]$ , we clamp it to the nearest value in the range. If  $\mathcal{S}_i(\mathbf{x})$  is neither 0 or 1, then  $\mathbf{x}$  has been labeled in penumbra in the two images. In this case, which is quite unlikely, we can not update the shadow mask unless we assume one of the two masks is correct.

When we update the shadow masks we compute a new mask per color channel, as opposed of a single mask based on luminance. We have found that better results are obtained in this way. Finally, once  $\mathcal{S}_j$  has been updated, the process can be repeated to update  $\mathcal{S}_i$ , this time using IRMs  $R_{ji}$  and the newly computed shadow map  $\mathcal{S}_j$ , and interchanging  $\mathcal{S}_i$  and  $\mathcal{S}_j$  in the steps already outlined.

## 6 Relighting Results

We have tested the proposed algorithm on a variety of scenes. In this paper we present the most relevant ones, the complete set of results are available in (Troccoli 2006b). Our first results correspond to two buildings in the campus of Columbia University in New York City: a model of **Casa Italiana** and a model of **St. Paul's Chapel**.



**Fig. 1** Two images of Casa Italiana taken at different time of the day from slightly different view points. *Left*: image acquired at 3:22pm; *right*: image acquired at 1:28pm on the same day, under partly cloudy conditions



**Fig. 2** Input images to our relighting algorithm. *From left to right*: the two images of Casa Italiana warped to same view point and zoomed to show the region of interest; the corresponding shadow masks

The images used for the test on Casa Italiana are shown in Fig. 1. These images were taken from slightly different view points and registered manually. One of the images was acquired at 1:28pm and the other one at 3:22pm. In Fig. 2 we show the same images, now aligned to the same view point, and their computed shadow masks. Note that shadows are in different locations, and that in one case, one complete face of the building is shadowed. We do the view-warp process by back-projecting the pixels to the scene, finding the distance to the camera of the corresponding surface point and projection to the viewpoint of the corresponding image. Hence, only those pixels for which we have geometry can be warped. The remaining pixels are left black, and correspond to holes in the 3D model. After filtering and shadow detection we collected the IRM data and updated the shadow masks. Using these shadow masks and the computed IRMs, we ran the relighting and de-shadowing algorithms to turn the image that was acquired at 1:28pm to the illumination at 3:22pm. In Fig. 3 we show both, the relighted image and the de-shadowed image. Note that the relighted images preserves the shadows while the de-shadowed image has none.

One important aspect to note about the de-shadowed image is that regions that were in shadow in both of the input images have been successfully de-shadowed. For better visualization, Fig. 3 also shows two composite pictures with the image before and after the relighting operation. Each composite picture is divided in two regions. The left region shows the original image of Casa Italiana acquired at 3:22pm. The right region, shows the image acquired at 1:28pm. In one of the composite pictures we show the original image at 1:28pm and in the other one, the relighted image. Note how in the latter case it is hard to distinguish the boundaries between the actual and relighted image.

In a similar way we conducted the experiments on the images St. Paul's Chapel. Figure 4 shows the acquired images, one of which was taken at 11:22am and the other one at 12:35pm, with their corresponding shadow masks. In Fig. 5 we show the results of applying the de-shadowing operation. In this case, we completely removed the shadows from the image acquired at 12:35pm. The picture on the left of Fig. 5 shows the obtained de-shadowed image, and the picture on the right is a composite picture in which, the left half



**Fig. 3** Relighted and de-shadowed images of Casa Italiana. *Left*: image taken at 1:28pm relighted to illumination at 3:22pm; *middle-left*: same relighted image but with shadows removed. *Middle-right*: composite picture before relighting; the *left half* of the image is the origi-

nal input image acquired at 3:22pm, the *right half* is the image taken at 1:28pm. *Right*: Composite picture after relighting, where the right half is now replaced with the results of relighting the image taken at 1:28pm with the illumination at 3:22pm



**Fig. 4** Two images of St Paul's Chapel at Columbia University acquired at 11:22am and at 12:35pm, and their corresponding shadow masks

corresponds to the original image and the right-half to the de-shadowed image.

## 7 Irradiance and Texture Factorization

We now analyze the irradiance and texture factorization problem. In (Troccoli and Allen 2006) we showed how this could be achieved for point light sources or illumination models written as an expansion of spherical harmonics (SH) and PCA basis. In this paper we extend this idea to model natural illumination as a combination of point source plus and ambient component and taking shadows into account. These algorithms address different situations that arise in practice:

1. An object illuminated by a distant point light sources. Given two images of this object under different point source illumination and the surface normals at each pixel, the factorization algorithm solves for the direction of the lights and relative intensities.

2. An object illuminated by smooth ambient illumination expressed as an expansion in terms of spherical harmonics (SH) or PCA basis. Here, given two images, we solve for the terms of the SH or PCA expansion.
3. A scene illuminated by a point source and ambient illumination. Such is the case we encounter in outdoor scenes, where the sun acts as a point source and the sky and the surrounding environment as an ambient component.

### 7.1 Factorization Algorithm for Point Light Sources

When an object is illuminated by a directional point light source whose direction is described by a normalized 3-vector  $\mathbf{l}_1$ , the irradiance for a scene point  $\mathbf{x}$  with associated normal  $\mathbf{n}_x$  is:

$$E(\mathbf{n}_x) = L \max(\mathbf{n}_x \cdot \mathbf{l}_1, 0), \quad (24)$$

where  $L$  denotes the source intensity and  $\cdot$  the vector dot product. Given two images illuminated by point sources





**Fig. 5** De-shadowing results for the images of St Paul’s Chapel. *Left*: de-shadowed image. *Right*: Composite image that shows, *on the left*, the original image and *on the right*, the de-shadowed image. Note how the shadows are successfully removed

with directions  $\mathbf{l}_1$  and  $\mathbf{l}_2$ , respectively, the ratio image obtained is described by the following equation:

$$R(x, y) = \frac{L_1 \max(\mathbf{n}(x, y) \cdot \mathbf{l}_1, 0)}{L_2 \max(\mathbf{n}(x, y) \cdot \mathbf{l}_2, 0)}, \tag{25}$$

defined only for non-zero values of the denominator and numerator. Our goal is to solve for the direction of the light sources given the ratio image and the surface normals. It should be clear at this point that there will be an ambiguity in the light source intensities  $L_1$  and  $L_2$ , since multiplying the numerator and denominator by the same constant in the above expression does not affect the final result. Hence, we can fix  $L_1$  to unity and solve for  $\mathbf{l}_1$  and  $\mathbf{l}_2$  scaled by  $L_2$ . To simplify the notation we drop the  $(x, y)$  coordinates in  $R(x, y)$  and  $\mathbf{n}(x, y)$ . Instead we will use a single subindex to enumerate all pixels. Then, we can solve for the light direction and relative intensities from the following system of linear equations:

$$\begin{bmatrix} \mathbf{n}_0^T & -R_0 \mathbf{n}_0^T \\ \vdots & \vdots \\ \mathbf{n}_k^T & -R_k \mathbf{n}_k^T \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \\ L_2 \mathbf{l}_2 \end{bmatrix} = 0. \tag{26}$$

This is a linear system of the form  $\mathbf{Ax} = 0$ . The solution we are looking for is the one dimensional null-space of  $\mathbf{A}$ . Unfortunately, if the dimension of the null-space of  $\mathbf{A}$  is greater

than one it will not be possible to solve uniquely for the light directions. This condition will arise if the distribution of the imaged surface normals is small: e.g. if the scene is a flat wall. Given the null-space vector  $\mathbf{x} = (x_0, x_1, x_2, x_3, x_4, x_5)$ , we obtain  $\mathbf{l}_1, \mathbf{l}_2$  and  $L_2$  from:

$$\mathbf{l}_1 = \frac{(x_0, x_1, x_2)}{\|(x_0, x_1, x_2)\|}, \tag{27}$$

$$\mathbf{l}_2 = \frac{(x_3, x_4, x_5)}{\|(x_3, x_4, x_5)\|}, \tag{28}$$

$$L_2 = \frac{\|(x_3, x_4, x_5)\|}{\|(x_0, x_1, x_2)\|}. \tag{29}$$

To handle color images we could treat each channel separately and solve (26) per channel. However, this typically yields three slightly different positions for the light source. We can obtain a more robust solution if we convert the image to luminance space and use the luminance values, instead. After we recover the direction of the light sources, the relative scale  $L_2$  for each channel  $c$  is obtained from the original images by averaging the following expression over all pixels:

$$L_{2,c}(x, y) = \frac{\max(\mathbf{n}(x, y) \cdot \mathbf{l}_1, 0)}{R(x, y) \max(\mathbf{n}(x, y) \cdot \mathbf{l}_2, 0)}. \tag{30}$$

Also, note that nothing is known about the absolute chromaticity of the light sources. By fixing the intensity  $L_1$  to the same value for the all three channels, we assume that light to be white. This chromatic ambiguity can not be solved without further assumptions or resorting to a color calibration object.

### 7.2 Factorization Algorithm for Generalized Illumination

We can extend the previous case to a more general form of illumination. To do so, we define irradiance as an expansion in terms of spherical harmonic basis functions. Ramamoorthi and Hanrahan (2001a) and Basri and Jacobs (2003) have established that the image of a diffuse convex object under general illumination is well approximated by a low dimensional spherical harmonic expansion. Spherical harmonics are orthonormal basis defined over the sphere. Using this framework, we can approximate the incident irradiance as:

$$E(\mathbf{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_l L_{lm} Y_{lm}(\mathbf{n}). \tag{31}$$

In (31) above,  $Y_{lm}$  are the spherical harmonic functions,  $L_{lm}$  are the spherical harmonic coefficients of the incident illumination, and  $A_l$  is a constant that represents the effects of multiplying the incident light with the half-cosine function. In this context, we want to solve for  $L_{lm}$ . Since  $A_l$  decays very rapidly, a very good approximation can be obtained by limiting  $l \leq 2$ . A first order spherical harmonic approximation (up to  $l = 1$ ) has a total of four terms and a second order approximation has a total of nine. Before writing down the irradiance ratio in terms of spherical harmonics, we do one more notation change for clarity purposes. We replace the double-indexed  $Y_{lm}$  functions and  $L_{lm}$  coefficients by their single-index equivalent  $Y_s$  and  $L_s$ , where  $s = l^2 + l + m$ . Also, since we have to solve for two different illuminations  $L_s$ , we will denote these with  $L_{1s}$  and  $L_{2s}$ . Using this new notation, we can substitute (31) into our irradiance ratio expression to obtain:

$$R_i = \frac{\sum_{s=0}^n A_s L_{1s} Y_s(\mathbf{n}_i)}{\sum_{s=0}^n A_s L_{2s} Y_s(\mathbf{n}_i)} \tag{32}$$

where  $n = 4$  or  $n = 9$  depending on the order of the desired approximation. We can now derive a system of linear equations similar to (26) on the unknown lighting coefficients  $L_{1s}$  and  $L_{2s}$ .

$$\begin{bmatrix} A_0 Y_0(\mathbf{n}_0) \dots A_n Y_n(\mathbf{n}_0) - R_0 A_0 Y_0(\mathbf{n}_0) \dots \\ \vdots \\ A_0 Y_0(\mathbf{n}_k) \dots A_n Y_n(\mathbf{n}_k) - R_k A_0 Y_0(\mathbf{n}_k) \dots \end{bmatrix} \begin{bmatrix} L_{10} \\ \vdots \\ L_{1n} \\ L_{20} \\ \vdots \\ L_{2n} \end{bmatrix} = 0. \tag{33}$$

The solution to (31) is once more the null-space of  $A$  and the coefficients  $L_{1s}$  and  $L_{2s}$  will be defined up to a scale factor. This scaling factor can be fixed by setting  $L_{10} = 1$ , which fixes both the relative scale and chromaticity of the illumination.

The well-conditioning of the system (33) will depend on the distribution of surface normals. For better results and higher robustness against noise, it is possible to re-cast the problem in terms of principal components. This means replacing the spherical harmonic basis by lower dimensional orthogonal basis obtained using principal component analysis (PCA). The rationale behind this change of basis is that the principal components are vectors in the direction of greater variability (in this case due to illumination changes). Ramamoorthi (2002) derived an analytic expression for the principal components of the image of an object under all possible point light sources and showed that these are related to the spherical harmonic basis  $Y_s$ . In particular, Ramamoorthi shows that the eigenvectors obtained from PCA of the image space of an object illuminated under all possible point light sources can be well approximated as a linear combination of spherical harmonic functions up to order two. Let  $V$  be the matrix with the principal eigenvectors as columns, then there exists a matrix  $U$  such that  $V \approx YU$ , where  $Y$  is a matrix whose columns are the first nine spherical harmonics  $Y_0 \dots Y_8$ . The matrix  $U$  can be computed analytically from the geometry of the object and details on how to do this are given in (Ramamoorthi 2002). Using the eigenvectors  $V_i$  as the new basis, we can now write the incident irradiance as:

$$E(\mathbf{n}) = \sum_{i=0}^n e_i V_i(\mathbf{n}), \tag{34}$$

where  $e_i$  are the coefficients of the irradiance in principal components basis. The number of terms to employ in this new approximation will depend on the object geometry, but by looking at the eigenvalues associated to each vector it is possible to determine a good cut-off point. Now, we can

write (33) as:

$$\begin{bmatrix} V_0(\mathbf{n}_0) \dots V_n(\mathbf{n}_0) - R_0 V_0(\mathbf{n}_0) \dots \\ \vdots \\ V_0(\mathbf{n}_k) \dots V_n(\mathbf{n}_k) - R_k V_0(\mathbf{n}_k) \dots \end{bmatrix} \begin{bmatrix} e_{10} \\ \vdots \\ e_{1n} \\ e_{20} \\ \vdots \\ e_{2n} \end{bmatrix} = 0. \quad (35)$$

Once we find the coefficients  $e_{1i}$  and  $e_{2i}$  we can find the corresponding  $L_{1s}$  and  $L_{2s}$  coefficients by substitution into:

$$L_{1s} = \frac{\sum_i^n U_{si} e_{1i}}{A_s}, \quad L_{2s} = \frac{\sum_i^n U_{si} e_{2i}}{A_s}, \quad (36)$$

where  $U_s$  is the  $s^{th}$  row of  $U$ . To handle color images we treat each of the RGB channels separately and solve for three sets of coefficients  $L_{1i}$  and  $L_{2i}$ . Once again, as before, there is an inherent chromatic ambiguity that we can only solve for if we have an image of a color calibration object.

### 7.3 The Point Light Plus Ambient Illumination Algorithm

Consider the case of outdoor illumination, where the sun is a distant point light source and the sky a hemispherical area source. We can model this situation as a sum of a point light source plus an ambient component. However, the sun is not just any directional light source. Its daily trajectory over the sky has been very well studied, and its position in the sky dome can be computed from the time of the day and the geolocation (latitude and longitude) of the scene (Reda and Andreas 2003). Since these two pieces of information are easily available, the time being provided by the camera and the geolocation by a GPS unit or any of today’s web-based mapping engines, we can assume the position of the sun to be known. Hence, our model of outdoor irradiance can be expressed as the combination of the sky irradiance plus a half-cosine term. To model sky irradiance, we can use a spherical harmonics or PCA basis expansion. Then, we can express outdoor irradiance as:

$$E(\mathbf{n}) = P \max(\mathbf{n} \cdot \mathbf{s}, 0) + \sum_{s=0}^n L_s Y_s(\mathbf{n}). \quad (37)$$

Here  $L_s$  are the coefficients for the basis expansion of sky irradiance in SH or PCA basis,  $P$  is the relative intensity of the sun and  $\mathbf{s}$  the sun direction. Since the ambient and direct components are modeled separately in (37), we can work with images with shadows, as we did with the relighting technique. All we need is the shadow mapping function  $S$ , and (37) becomes:

$$E(\mathbf{x}) = S(\mathbf{x}) P \max(\mathbf{n}_x \cdot \mathbf{s}, 0) + \sum_{s=0}^n A_s L_s Y_s(\mathbf{n}_x). \quad (38)$$

For two images taken at different times of the day, with the sun at directions  $\mathbf{s}_1$  and  $\mathbf{s}_2$  and relative intensities  $P_1$  and  $P_2$ , the irradiance ratio of point  $\mathbf{x}$  is given by:

$$R(\mathbf{x}) = \frac{S_1(\mathbf{x}) P_1 \max(\mathbf{n}_x \cdot \mathbf{s}_1, 0) + \sum_{s=0}^n A_s L_{1s} Y_s(\mathbf{n}_x)}{S_2(\mathbf{x}) P_2 \max(\mathbf{n}_x \cdot \mathbf{s}_2, 0) + \sum_{s=0}^n A_s L_{2s} Y_s(\mathbf{n}_x)}. \quad (39)$$

Our goal is to solve for the relative sun intensities  $P_1$  and  $P_2$  and the sky irradiance coefficients  $L_{1s}$  and  $L_{2s}$ , which we do as before, linearizing the equations into the form  $A\mathbf{x} = 0$  and solving for the null-space of  $A$ .

### 7.4 Extracting the Albedo Map

After we have solved for the relative irradiance using one of the three models presented, we compute an albedo map for the scene. However, note that the chromatic and scale ambiguity in the estimated irradiance will translate to the estimation of the albedo map, which will also be defined up to scale. From the image pair  $I_1$  and  $I_2$  with estimated irradiance  $E_1$  and  $E_2$  we compute the albedos at each pixel:

$$\rho_1(x, y) = \frac{I_1(x, y)}{E_1(\mathbf{n}(x, y))}, \quad (40)$$

$$\rho_2(x, y) = \frac{I_2(x, y)}{E_2(\mathbf{n}(x, y))}, \quad (41)$$

$$\rho(x, y) = \frac{I_1(x, y)\rho_1(x, y) + I_2(x, y)\rho_2(x, y)}{I_1(x, y) + I_2(x, y)}. \quad (42)$$

In other words, for each pixel  $(x, y)$  we set its albedo  $\rho(x, y)$  to a weighted average of the albedos we obtain from  $I_1$  and  $I_2$ . The weights are set to the pixel intensities, so that dark pixels, which are more dominated by noise, are down-weighted.

## 8 Texture Factorization Results

We tested the three factorization algorithms on different kinds of scenes under different types of illumination conditions. In this section, we report the obtained results.

### 8.1 Point Light Source Model

First, we ran the point light source algorithm on synthetic and real data. The images used in our tests are shown in Fig. 6. The first two renderings are synthetic scenes for which we used a model of a sphere and a model of the Armadillo<sup>1</sup> textured with a synthetic wooden pattern and rendered using a ray-tracer. We then included three objects that

<sup>1</sup>The Armadillo model was downloaded from the Stanford scanning repository.



**Fig. 6** Objects used for testing our algorithm. Starting from the left, first come the synthetic renderings: a sphere and the Armadillo; followed by three objects with their geometry acquired using photomet-

ric stereo: the Buddha, the cat and the owl; and finally two scanned objects: the chicken and the girl. Each row shows the objects with a different illumination

**Table 1** Ground truth and recovered light directions and relative intensities for the synthetic images of the sphere and the Armadillo

	Point source 1			Point source 2			Rel. intensity (R, G, B)
	$x$	$y$	$z$	$x$	$y$	$z$	
Actual position	-0.58	0.36	0.73	0.28	-0.28	0.92	(5.00, 10.00, 20.00)
Sphere	-0.58	0.36	0.73	0.27	-0.28	0.92	(5.03, 10.10, 20.48)
Armadillo	-0.58	0.35	0.73	0.27	-0.28	0.92	(5.11, 10.27, 20.88)

**Table 2** Ground truth and recovered light directions and intensities for the Buddha, cat and owl models

	Point source 1			Point source 2			Rel. intensity (R, G, B)
	$x$	$y$	$z$	$x$	$y$	$z$	
Actual position	0.40	0.48	0.78	-0.32	0.49	0.92	(1.00, 1.00, 1.00)
Buddha	0.44	0.47	0.77	-0.32	0.47	0.82	(1.03, 1.03, 1.04)
Cat	0.39	0.49	0.78	-0.33	0.47	0.82	(1.09, 1.09, 1.05)
Owl	0.39	0.48	0.78	-0.31	0.44	0.84	(1.02, 1.01, 1.00)

had been imaged under known point-light source:—a Buddha, a cat and an owl.<sup>2</sup> The geometry and normals-map for these objects were obtained using photometric stereo. Finally, we included in our testing a set two objects that had been scanned using a Polhemus hand-held scanner:<sup>3</sup> a chicken and a figure of an Asian girl. The synthetic renderings and photometric stereo models are good for ground-truth comparisons, because we know the position of the light sources and do not require image registration. For the chicken and girl models, we captured several images varying the position of a point light source but leaving the view-point fixed. We manually registered these images with the 3D model.

<sup>2</sup>The Buddha, cat and owl data sets were generously provided by Dan Goldman and Steve Seitz from the University of Washington, Seattle, WA.

<sup>3</sup>These models were scanned for us using the Polhemus scanner by Michael Reed of Blue Sky Studios.

We ran our point-light source estimation model of Sect. 7.1 on all of the image pairs shown in Fig. 6. Tables 1 and 2 show the ground truth and recovered light source positions and relative scales for the synthetic and photometric stereo models. For the synthetic scenes, the recovered light source directions and scaling factors shown in Table 1 are almost identical to the actual directions. Likewise, the computed light source directions for the Buddha, cat and owl models listed in Table 2 are very close the ground truth data. We had no ground truth data for the chicken and Asian girl models. Nevertheless, we ran our algorithm to obtain the position of the light sources and obtain the factorization.

As a second step, we used the computed light direction to factor the input images into their corresponding texture (albedo maps) and shading (irradiance) components. The results are shown in Fig. 7—the top row shows the albedo map, and the second and third rows the irradiance maps for each of the input images. Note that, with the exception of a few minor artifacts, the albedo maps do not con-



**Fig. 7** Results obtained using the point light source model for the images in Fig. 6. The *top row* shows the recovered albedo, the *middle row* shows the factored irradiance for the first illumination, and the *last row* the factored irradiance for the second illumination. Notice how the factorization de-couples texture from shading

tain any shading effects. This is certainly true for the synthetic models: both the sphere and the armadillo albedo maps look completely flat. For the real data sets, some artifacts can be seen where the surfaces are not purely Lambertian. For example, the owl shows some specular components in the albedo map. Other artifacts are brighter spots in non-convex regions, in particular at the junction of the head and body of the cat and owl models, the junction of the arm and body in the chicken model, and the junction of the hair and face in the girl model. The convexity assumption fails here and inter-reflections influence the final result. The visible effect is a brightening of the albedo map, since the pure Lambertian model can not explain the increase in irradiance due to inter-reflections. As a final comment, the pants in the chicken model are not in the albedo map since the luminance value of those pixels falls below the shadow threshold, and hence ignored by the algorithm.

When ground truth was available, we also computed a quantitative measure of the quality of the factorization by comparing the obtained irradiance images with the irradiance images generated using the ground truth light source position. Since there is a scale ambiguity that is inherent to our method, we normalized all images before computing the error metric. This normalization was achieved by setting the norm  $\|I\|^2 = \sum_{x,y} I(x,y)^2$  equal to one. Then, for a given pair of normalized ground truth image  $I^0$  and reconstructed irradiance image  $I^1$ , we computed the relative squared error

of the reconstruction:

**Table 3** Normalized reconstruction error for the irradiance images. The first column indicates the object, the second one the method used (PL = point light, PCA  $n$  = generalized illumination using PCA of size  $n$ ), and the last two columns show the normalized reconstruction error for the two irradiance images

Model	Method	Error 1	Error 2
Sphere	PL	<0.1%	<0.1%
	PCA 5	0.40%	0.20%
Armadillo	PL	<0.1%	<0.1%
	PCA 3	3.50%	4.30%
Buddha	PL	0.40%	0.50%
	PCA 3	0.10%	1.60%
Cat	PL	<0.1%	<0.1%
	PCA 3	4.40%	4.50%
Owl	PL	<0.1%	<0.1%
	PCA 3	3.80%	3.50%

of the reconstruction:

$$\text{err}(I^1, I^0) = \frac{\|I^1 - I^0\|^2}{\|I^0\|^2}. \tag{43}$$

The resulting reconstruction errors are reported as percentages in Table 3. It can be observed that the reconstructed irradiance images using the point light source algorithm are very accurate.

<sup>4</sup>The relative squared error is frequently used in the literature (Basri and Jacobs 2003; Frolova et al. 2004) as a metric for evaluating the goodness of image reconstructions.



**Fig. 8** Results obtained using the generalized illumination model the images in Fig. 6. The *top row* shows the recovered albedo, the *middle row* shows the factored irradiance for the first illumination, and the *last row* the factored irradiance for the second illumination

## 8.2 Generalized Illumination Model

We tested the generalized light model algorithm on the same set of images we used for testing the point-light source factorization, shown in Fig. 6. In all cases, we first analytically computed the PCA basis. We used a basis of dimension 3 for all of the models except for the sphere, for which we used a PCA basis of dimension 5. We found these dimensions empirically, by testing with different basis size. The resulting factorizations into irradiance images and albedo maps are shown in Fig. 8 and the reconstruction errors for the irradiance images are tabulated in Table 3. It can be seen that the model approximates well the irradiance and produces a good factorization. The reported reconstruction errors are less than 4.5%. For the sphere, the quality of the approximation is as good as for the point-light source model. For the remaining models, the irradiance reconstruction error varies between 0.5% for the Buddha to 4.50% for the cat.

## 8.3 Point Plus Ambient Illumination Model

We ran the factorization algorithm on the images of Casa Italiana and images of the church of St. Marie, in Chappes, France. For Casa Italiana, the input images are the same we used for our relighting test (Fig. 2). For comparison purposes, we modeled the ambient light with two different approximations: a constant term and a PCA basis expansion of order 3. The results are shown in Figs. 9 and 10, respectively. Each figure shows the computed albedo and irradiance maps for each of the illuminations. In the absence of

ground truth, we restrict ourselves to a qualitative analysis of the results. Comparing the albedo maps we can see the best results were obtained from the constant term ambient approximation. The albedo map in Fig. 9 shows almost none of the illumination effects, except for a few brightening in regions where the sampling of the scanner was not high enough to approximate correctly the geometry variations. In contrast, the results obtained using a 3 term expansion do show some artifacts for surfaces looking down. Regions with downward pointing normals only receive indirect illumination from inter-reflections, so it is very unlikely that orientation consistency will hold for those points, and the higher order expansion attempts to account for these effects.

For the test on a model of the church of St. Marie, we used two sets of four images each: one set taken at 10:56am, and the other taken at 4:55pm. We put the images together into a composite picture for each illumination, shown in Fig. 11. Since each set of images is acquired under constant illumination with the same camera parameters, these blended images do not show artifacts. The two composites, plus the shadows masks shown in Fig. 12, the normals map, and the directions of the sun, were given to our illumination and texture factorization algorithm. The resulting albedo and illumination maps are shown in Figs. 13 and 14, respectively. The resulting albedo map is illumination free; the only artifacts are some regions in which the shadow boundaries are noticeable, which is a result of inaccuracies in the shadow detection. Unfortunately, this is a limitation of our method: it requires very accurate shadow masks which are very difficult to compute. Nevertheless, the computed albedo map is good for generat-



Fig. 9 Casa Italiana—Computed albedo and irradiance maps with constant ambient component



Fig. 10 Casa Italiana—Computed albedo and irradiance maps using an order 3 PCA basis for the ambient component

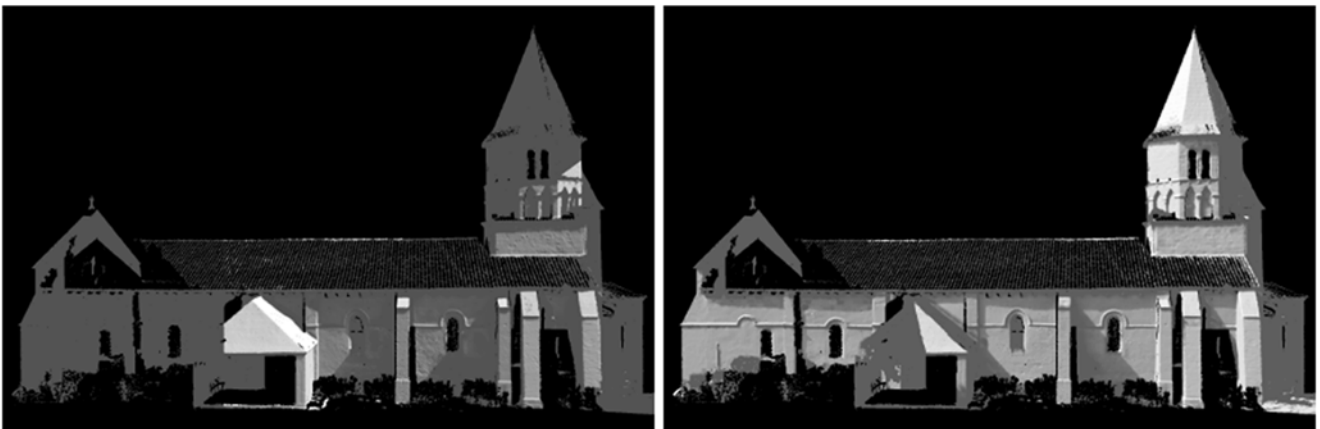


Fig. 11 Real images of Saint Marie, Chappes, France. The *top image* is a composite made of images acquired at 10:56am on May 26th 2005. The *bottom* is a composite made of images taken at 4:55pm on that same day



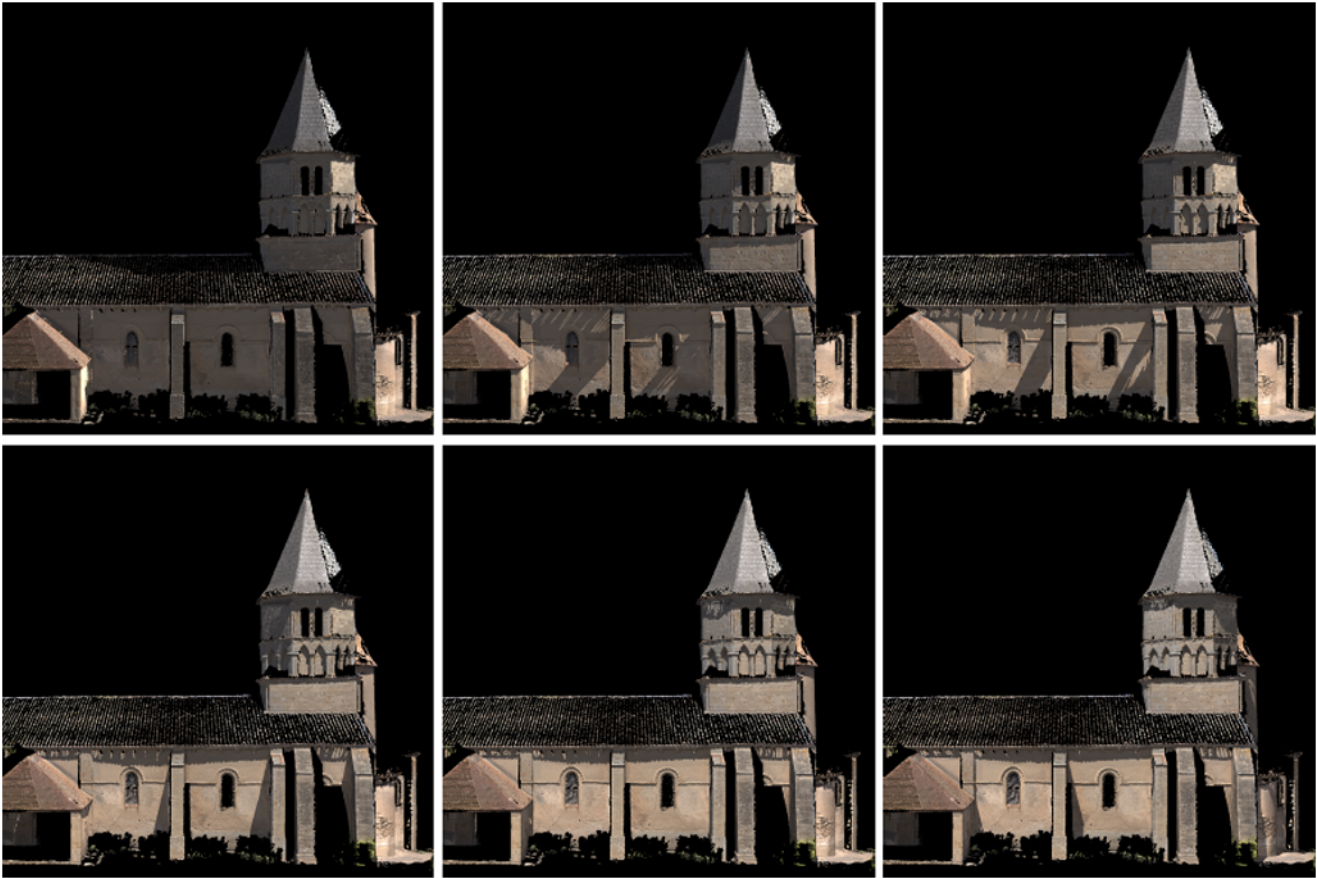
**Fig. 12** Shadow masks for the input images of Saint Marie, Chappes, France

**Fig. 13** Albedo map for the south facade of Saint Marie, Chappes, France



**Fig. 14** Illumination images computed for the images of Saint Marie, Chappes, France





**Fig. 15** Renderings of Saint Marie at Chappes under novel illumination conditions. These renderings simulate a day-time sequence with the sun at 10am, 11am, 12pm, 1pm, 2pm and 3pm

ing new renderings under novel illumination conditions. To show this, we created a set of renderings of St. Marie as the sun traverses the sky from morning to afternoon. The results, which were created using a ray tracer, are shown in Fig. 15. Additionally, an animation is available from (Troccoli 2006a).

## 9 Discussion

Our algorithms produce high quality results, as shown in the previous section and in (Troccoli 2006b). It is important, however, to understand the possible sources of error that can affect the quality of the reconstructed texture map. These possible sources of error are:

*Inadequate Geometry Sampling* The geometric models we use are obtained using a time-of-flight laser scanner that samples the scene in a discrete manner. Using discrete sampling we are only able to reconstruct geometry variations at half the frequency of the sampling rate. Rapid variations in the geometry are lost. Since surface normals are computed

by triangulation of the measured points, the algorithm is expected to work better in scenes with smooth varying surfaces than in rapidly changing scenes.

*Registration Errors* These are errors that are introduced when computing the camera parameters, either intrinsic or extrinsic. Registration errors can result in the incorrect mapping between a pixel and a surface normal, and in incorrect pixel correspondences between two images. Image-geometry registration is a difficult problem, and we have tried in our experiments to reduce registration errors as much as possible. Registration errors can influence the final result more or less depending on the scene. For example, small registration errors have less impact on geometrically smooth surfaces, where a small misalignment will still map a pixel to the correct surface normal. This is the case of flat walls, for example. Also, for smooth textures, small registration errors can be tolerated. It is for this reason that we apply a bilateral filter and smooth high frequency textures before computing the IRMs: we smooth out the texture and hence reduce the effects of image misalignments. Note that this smoothing is not a theoretical requirement but a practical one to make the algorithm more robust.

*Shadow Detection* The shadow labels define which of the four IRMs a sample contributes to. If a shadow label is incorrectly assigned, then the sample will incorrectly be attributed to the wrong IRM. For large number of samples per surface normal, a small number of outliers will not affect the results; but if there are only a few samples for a given surface normal, then the computed IRM could contain errors.

*Non-Lambertian Reflectance* The algorithms presented were developed assuming objects with Lambertian surface reflectance, but it is known that real-world surfaces are hardly pure Lambertian. As we have shown, however, the Lambertian model can serve as a good approximation for mostly diffuse surfaces. In our experiments, the only truly Lambertian objects we used for testing were the synthetic renderings of the sphere and the armadillo; while the remaining objects and images showed non-Lambertian behavior and still produced good results. Nevertheless, a big deviation from diffuse reflectance will cause incorrect results. In architectural scenes windows are a problem since they behave like mirrors. For this reason, we masked windows out, completely ignoring them during the IRM computation stage.

*Interreflections and Spatially Varying Ambient Occlusion* When a scene is not perfectly convex, the effects of interreflections and spatially varying ambient occlusion can affect the results of the relighting and de-shadowing algorithms. Interreflections are indirect contributions of light bouncing off a surface and reaching another one. Spatially varying ambient light occlusions are variations in the cone of sky<sup>5</sup> (or extended distant light sources) visible by a scene point. If two scene points with the same surface normal see a different cone of sky, then the irradiance at each of these points will be different, violating the orientation-consistency assumption. Also, the effects of interreflections and spatially varying ambient light occlusion are more noticeable in shadowed regions. In our algorithm, we do handle these indirectly and in an ad-hoc manner by allowing the shadow masks to take real values in the range [0, 1].

Of the above enumerated potential sources of error, the first two: inadequate geometry sampling and image registration errors, are not inherent to our methodology, but will be present in any system used to recover illumination-coherent texture maps from range and color images. To better account for the effects of interreflections and ambient occlusions, the contributions from these sources could be approximated using a ray-tracer; however, ray-tracing requires hole-free geometry with accurate normals, which is not always available from range scans. Acquiring a hole-free model is possible if unlimited time is available, but that is not typically the case during field work.

<sup>5</sup>Also defined as sky-aperture in (Narasimhan and Nayar 2001).

## 10 Conclusions and Future Work

In this paper we have presented two different methods for building illumination coherent photorealistic models of outdoor scenes from images acquired under different illumination conditions. We did this without introducing additional overhead to the acquisition process. This required that all the information needed for the creation of these models in a physically plausible way would need to come from the photographs themselves and the geometry of the object. Under these conditions, we developed two different algorithms: an algorithm that normalizes the textures using a relighting operator and an algorithm that does it by solving the texture and factorization problem, for which we have presented results on different real scenes. Comparing one method against the other, it is worth mentioning that the relighting algorithm does not assume any illumination model, and thus can be more general than the factorization approach, which does assume a particular parametric model for the illumination and uses the input data to find the parameters of this model. However, in those cases where the parametric model adjusts closely to real world scenarios, the factorization approach has the advantage of generating illumination-free textures which can then be relighted under novel illumination settings.

The results we have presented are for a number of outdoor scenes using a variety of lighting conditions. For more complex scenes and illumination settings, our assumptions might not hold, and our method may not produce the expected results. Inverse rendering, in particular for outdoor scenes, is indeed a difficult problem and even more elaborate techniques that make use of a light-probing device are limited to mostly diffuse like surfaces. Debevec et al. (2004) fit image data to a pre-measured BRDF model based on the Lafortune representation and assume that the colors of the Lambertian component and retroreflective lobes both follow the Lambertian color.

In our method, by assuming materials to be Lambertian-like, we guaranteed that the ratio image is texture-free, allowing us to derive a simple model for relighting and for factoring irradiance and texture. It seems unlikely that this restriction can be relaxed for the general case, but under certain circumstances, we might be able to. For example, if we had a small database of materials representative of the materials in the scene, and assumed point light source illumination, we could search the space of materials and source positions until we find the parameters that best fit the observed ratio image. This will likely require proper segmentation and materials clustering, as in (Hertzmann and Seitz 2005).

Yet to be explored is the use of active methods for reflectance modeling in outdoor environments. The range finder is, in fact, an active device which does report a returned intensity value. In recent work (Xu et al. 2006), pure

diffuse reflectance of large environments is obtained by analyzing the scanner's returned intensity value in combination with color-balancing techniques on the input color images.

**Acknowledgements** This research was funded in part by NSF grant IIS-0121239 and a grant from the Andrew Mellon Foundation. Research in France was done in collaboration with Prof. Stephen Murray from the department of Art History and Archeology at Columbia University, NY.

## References

- Agathos, A., & Fisher, R. (2003). Colour texture fusion of multiple range images. In *Proceedings of the 4th international conference on 3D digital imaging and modeling* (pp. 139–146).
- Allen, P. K., Troccoli, A., Smith, B., Murray, S., Stamos, I., & Leordeanu, M. (2003). New methods for digital modeling of historic sites. *IEEE Computer Graphics and Applications Magazine*, 23(6), 32–41.
- Allen, P., Feiner, S., Troccoli, A., Benko, H., Ishak, E., & Smith, B. (2004). Seeing into the past: creating a 3D modeling pipeline for archaeological visualization. In *Proceedings of 2nd international symposium on 3D data processing, visualization and transmission*.
- Bannai, N., Agathos, A., & Fisher, R. B. (2004). Fusing multiple color images for texturing models. In *Proceedings of 2nd international symposium on 3D data processing, visualization and transmission* (pp. 558–565). Los Alamitos: IEEE Computer Society.
- Basri, R., & Jacobs, D. W. (2003). Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2), 218–233.
- Beauchesne, E., & Roy, S. (2003). Automatic relighting of overlapping textures of a 3D model. In *Proceedings of computer vision and pattern recognition*.
- Buehler, C., Bosse, M., McMillan, L., Gortler, S., & Cohen, M. (2001). Unstructured lumigraph rendering. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques* (pp. 425–432). New York: ACM.
- Chuang, Y.-Y., Goldman, D. B., Curless, B., Salesin, D. H., & Szeliski, R. (2003). Shadow matting and compositing. *ACM Transactions on Graphics*, 22(3), 494–500.
- Curless, B., & Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on computer graphics and interactive techniques* (pp. 303–312). New York: ACM.
- Debevec, P. E., Taylor, C. J., & Malik, J. (1996). Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *Proceedings of the 23rd annual conference on computer graphics and interactive techniques* (pp. 11–20). New York: ACM.
- Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., & Sagar, M. (2000). Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on computer graphics and interactive techniques* (pp. 145–156). New York: ACM/Addison-Wesley.
- Debevec, P., Tchou, C., Gardner, A., Hawkins, T., Stumpfel, J., Jones, A., et al. (2004). *Estimating surface reflectance of a complex scene under natural captured illumination* (Technical report). University of Southern Californian, Institute for Creative Technologies.
- Frolova, D., Simakov, D., & Basri, R. (2004). Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. In *Proceedings of the European conference on computer vision* (Vol. 1, pp. 574–587).
- Funk-Lea, G., & Bajcsy, R. (1995). Combining color and geometry for the active, visual recognition of shadows. In *ICCV '95: proceedings of the fifth international conference on computer vision* (p. 203). Washington: IEEE Computer Society.
- Hertzmann, A., & Seitz, S. M. (2005). Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1254–1264.
- Ikeuchi, K., & Sato, K. (1991). Determining reflectance properties of an object using range and brightness images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11), 1139–1153.
- Lensch, H. P. A., Kautz, J., Goesele, M., Heidrich, W., & Seidel, H.-P. (2003). Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics*, 22(2), 234–257.
- Marschner, S. R., & Greenberg, D. P. (1997). Inverse lighting for photography. In *Proceedings of the 5th color imaging conference*.
- Narasimhan, S. G., & Nayar, S. K. (2001). Removing weather effects from monochrome images. In *IEEE conference on computer vision and pattern recognition (CVPR)* (Vol. 2, p. 186).
- Pulli, K., Cohen, M., Duchamp, T., Hoppe, H., Shapiro, L., & Stuetzle, W. (1997). View-based rendering: visualizing real objects from scanned range and color data. In *Rendering techniques '97* (pp. 23–34). New York: Springer Wien.
- Ramamoorthi, R. (2002). Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10), 1322–1333.
- Ramamoorthi, R., & Hanrahan, P. (2001a). On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. *Journal of the Optical Society of America A*, 18(10), 2448–2459.
- Ramamoorthi, R., & Hanrahan, P. (2001b). A signal-processing framework for inverse rendering. In *SIGGRAPH '01: proceedings of the 28th annual conference on computer graphics and interactive techniques* (pp. 117–128). New York: ACM.
- Reda, I., & Andreas, A. (2003). *Solar position algorithm for solar radiation applications* (Technical report). National Renewable Energy Laboratory, Golden, CO.
- Salvador, E., Cavallaro, A., & Ebrahimi, T. (2004). Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2), 238–259.
- Shashua, A., & Riklin-Raviv, T. (2001). The quotient image: class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 129–139.
- Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. In *ICCV '98: proceedings of the sixth international conference on computer vision* (p. 839). Washington: IEEE Computer Society.
- Troccoli, A. (2006a). Animation of relighted image sequence of the church of St. Marie at Chappes, France. <http://www.cs.columbia.edu/atroccoli/jcvcv07>.
- Troccoli, A. (2006b). *New methods and tools for 3D-modeling of large scale outdoor scenes using range and color images*. PhD thesis, Department of Computer Science, Columbia University.
- Troccoli, A., & Allen, P. K. (2004). A shadow based method for image to model registration. In *2nd IEEE workshop on image and video registration (IVR 04)*.
- Troccoli, A., & Allen, P. (2005). Relighting acquired models of outdoor scenes. In *Proceedings of 3DIM'05*.
- Troccoli, A., & Allen, P. (2006). Illumination and textures factorization using ratio images of an object of known geometry. In *Proceedings of the 3rd international symposium 3D data processing, visualization, and transmission (3DPVT 06)*. Los Alamitos: IEEE Computer Society.

- Wang, L., Kang, S. B., Szeliski, R., & Shum, H.-Y. (2001). Optimal texture map reconstruction from multiple views. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition* (pp. 347–354).
- Wang, H., Li, S. Z., & Wang, Y. (2004). Generalized quotient image. In *IEEE conference on computer vision and pattern recognition (CVPR)* (Vol. 2, pp. 498–505).
- Xu, C., Georghiades, A., Rushmeier, H., & Dorsey, J. (2006). A system for reconstructing integrated texture maps for large structures. In *Proceedings of 3rd international symposium on 3D data processing, visualization and transmission*.
- Yu, Y., & Malik, J. (1998). Recovering photometric properties of architectural scenes from photographs. In *Proceedings of the 25th annual conference on computer graphics and interactive techniques* (pp. 207–217). New York: ACM.