Optimal Illumination for Image and Video Relighting

Francesc Moreno-Noguer **Shree K.Nayar** **Peter N.Belhumeur**
Department of Computer Science, Columbia University

Keywords: Video relighting, Illumination, Image-based rendering, Lighting representation

Abstract

It has been shown in the literature that image-based relighting of scenes with unknown geometry can be achieved through linear combinations of a set of pre-acquired reference images. Since the placement and brightness of the light sources can be controlled, it is natural to ask: what is the optimal way to illuminate the scene to reduce the number of reference images that are needed? We show that the best way to light the scene (i.e., the way that minimizes the number of reference images) is not using a sequence of single, compact light sources as is most commonly done, but rather to use a sequence of lighting patterns as given by an *object-dependent* lighting basis. While this lighting basis, which we call the optimal lighting basis (OLB), depends on camera and scene properties, we show that it can be determined as a simple calibration procedure before acquisition. We demonstrate through experiments on real and synthetic data that the optimal lighting basis significantly reduces the number of reference images that are needed to achieve a desired level of accuracy in the relit images. This reduction in the number of needed images is particularly critical in the problem of relighting in video, as corresponding points on moving objects must be aligned from frame to frame during each cycle of the lighting basis. We show, however, that the efficiencies gained by the optimal lighting basis makes relighting in video possible using only a simple optical flow alignment. We present several relighting results on real video sequences of moving objects, moving faces, and scenes containing both. In each case, although a single video clip was captured, we are able to relight again and again, controlling the lighting direction, extent, and color.

1 Introduction

Recently, much work has been done in relighting images of still objects [DHT*00, DWT*02, GBK01, LKG*03, LWS02, NN04, NSD94, RH01, SNB03, WHON97]. Most of the techniques use an image-based approach; the images under new lighting conditions are synthesized from a set of pre-acquired reference images that sample the plenoptic function [WFHL02]. Considering that the lighting process obeys the rules of superposition [Bus60, NSD94] new images are generated via linear combination of the set of reference images. These reference images are most often acquired by either moving a single compact light source over a sphere surrounding the scene, or by sequentially turning on one source at a time among an array of compact sources. Since the placement and brightness of the light sources can be controlled, it is natural to ask: what is the best way to illuminate the scene in each of the reference images? This, of course, depends on how the images will be used and the conditions under which they will be acquired. In [SNB03], it was shown that if one wanted to acquire n images of a scene lit by n distinct light sources, then one could reduce the signal to noise ratio (SNR) in these images by illuminating the scene with lighting patterns as dictated by Hadamard codes. The use of Hadamard codes was shown – in theory and in practice – to significantly reduce the image noise accumulating as a result of sensing and digitization.

For most applications of relighting, however, the reduction of image noise is not the first priority. Rather, the goal of relighting is to synthesize images of the scene under new illumination conditions such that the synthesized images are as close (in an L^2 sense) to real images as possible. The denser the sampling of the lighting directions for the reference images, the higher the quality of the synthesized images. And one could expect to achieve errorless relighting results (under the assumption of distant light sources), if one were to sample the space of lighting directions with infinite resolution. Yet, in no case is this practical or even feasible, thus one must settle with the implicit trade-off between quality and the number of reference images. (A study of this sampling problem was provided in [LWS02].)

Yet, we show in this paper that it is not simply the number of reference images that determine the quality of relighting, but also the way in which the scene is illuminated. In particular, we show that the best way (i.e., the way that minimizes the number of reference images) to light the scene is not using a sequence of compact, single point light sources as is most commonly done, but rather to use a sequence chosen from a family, or basis, of lighting patterns each composed of many compact light sources of varying brightness. Furthermore, we show that the optimal lighting basis can be determined as a simple calibration procedure before acquisition. We demonstrate through experiments on real and synthetic data that the optimal lighting basis significantly reduces the number of reference images that are needed to achieve a desired level of accuracy in the relit images. This reduction in the number of needed images is particularly critical in the problem of relighting in video as demonstrated in [GTW*04]. The reason for this is that corresponding points on moving objects must be aligned from frame to frame during each cycle of the lighting basis. We show that the optimal lighting basis can reduce the number of light patterns that are needed by a factor of $2-3$ as compared to the spherical harmonic basis used in [GTW*04].

We present several relighting results on real video sequences of moving objects, moving faces, and scenes containing both. In each case although a single video clip was captured, we are able to relight again and again, controlling the lighting direction, extent, and color. In addition, we show that lighting can be

changed over the course of the sequence to produce the effect of a moving source. The lighting can be specified by the user or by some pre-acquired measurement of natural illumination such as an environment map. In the examples presented here, we used a video camera with a wide angle lens to acquire a temporally dynamic measurement of the lighting in New York City's Times Square. This lighting map was then used to relight one of the video clips of a human face. Finally, we show that the lighting can even be controlled locally, so that different objects in a scene can be relit in different ways.

The rest of this paper is organized as follows: In Section 2, we review related work. In Section 3, we describe the relighting process of still and moving objects. In Section 4, we focus on the selection of the best illuminant basis for relighting. The performance of different light basis is analyzed for synthetic data in Section 5. The experiments are extended to real data in Section 6, where we show the relighting results of real video sequences. Conclusions are given in Section 7.

2 Related Work

A great deal of past work has focused on relighting scenes using pre-synthesized [DSG91] or pre-acquired [Hal94, GKB98, DHT*00, KBMK01, MGW01, MPN*02] reference images. In each of these, the reference images are gathered by systematically varying the lighting direction. If the sampling of the lighting directions is dense enough, then due to the linearity of scene radiance, images of the scene under a user specified illumination can be synthesized by superposition of the single light source images, see again [DSG91]. In nearly all of this work the reference images were acquired under single, compact source illumination. In [Hal94], incandescent spot lights were used to sample 66 lighting directions on a sphere as reference images of a human face were gathered. In [GKB98], xenon strobes were used to sample 64 lighting directions on a geodesic dome as reference images of a human face were gathered. In [DHT*00], a moving compact light source was used to sample 2048 lighting directions for the illumination of, yet again, a human face. Yet, here the density of the sampling allowed for impressively accurate results in the synthesis of effects such as specularities and cast shadows. In [MPN*02] compact light sources were used to sample 60 illumination directions per viewpoint for objects made of specular and fuzzy materials. And in [NBB04], a moving spotlight was used to gather 4096 images of a still-life.

One of the aims of this paper is to show that illumination using single, compact light sources is not the most efficient for relighting. In contrast to much of this past work, we will show that if properly chosen lighting patterns are used to illuminate the scene, then many fewer reference images need be gathered. This is not the first work to consider using light patterns for relighting. However, all of the lighting bases that have been used in the existing literature to date are pre-chosen and are not a function of the camera or scene properties. In [NSD94] natural skylight illumination is approximated by a set of *steerable functions*. In [SNB03] use a scheme based on *Hadamard* codes for reducing SNR in the images. [GTW*04] uses terms of the *spherical harmonic* basis. An analysis of the efficiency of spherical harmonics for relighting can be found in [RH01]; however, the optimality of a spherical harmonic basis holds only for objects with Lambertian reflectance and scenes without cast shadows.

Finally, this is not the first work to consider relighting in video. [DWT*02] uses a sphere of controlled light sources to light a moving object during acquisition with illumination pre-acquired from a different environment. While any illumination can be specified during acquisition, the resulting video sequence cannot be relit. In order to relight a moving face, [PSS99, NN04, HGT*04] first fit 3-D models to the face shape and then used this to render frames under new illumination. The recent work in [GTW*04] is probably the closest to the video relighting component of this paper. Like ours, the goal of their work is to acquire a video sequence that can be relit again and again according to user specified illumination. To do this, [GTW*04] acquire and then process, as we do, a video sequence in which the lighting is systematically varied over the course of the sequence. In [GTW*04] use ten light patterns representing the first nine terms of spherical harmonics plus one directional light source to relight the face of an actor. In contrast, we develop and then use an object-dependent lighting basis that is significantly more efficient for video relighting. In our experiments, we have shown that we can reduce the number of light patterns that are needed by a factor of $2-3$. This is, as we will argue later, critical for the case of moving objects which require frame by frame alignment. A disadvantage of our method, however, is that it requires computation of the optimal lighting basis. Still, this can be accomplished within a few seconds prior to video capture.

3 Relighting with a Lighting Basis

In this section, we give a mathematical description of the relighting process using a lighting basis. We first define relighting for still objects and then introduce time dependence in the formulation in order to take into account the relighting of moving objects in video sequences.

3.1 Relighting in Static Scenes

Our setup is as follows. The scene is illuminated simultaneously by m single light sources, each of varying brightness; we call this illumination a light pattern. We assume that the light sources are distant, so they can be parameterized as a function of direction only. Let the m-dimensional array $\mathbf{L}_p = [L_p(\theta_1,\phi_1),\ldots,L_p(\theta_m,\phi_m)]^T$ be the vector of radiances of all the single light sources generating the p^{th} lighting pattern, where $L_p(\theta_l,\phi_l)$ is the radiance of the l^{th} light source of the p^{th} light pattern, and $\Phi_l = (\theta_l,\phi_l)$ are the global spherical coordinates of the l^{th} light source. In order to compute the image of a pixel $\mathbf{x}_i = [x_i, y_i]^T$ we make use of the properties of image superposition [Bus60, NSD94, WFHL02]:

1. The image resulting from multiplying each pixel by a

factor α is equivalent to an image resulting from a light source with intensity multiplied by the same factor.

2. An image of a scene illuminated by two light sources $L(\Phi_1)$ and $L(\Phi_2)$, equals the sum of an image illuminated with $L(\Phi_1)$ and another image illuminated with $L(\Phi_2)$.

From these properties, we compute the image of a pixel $\mathbf{x}_i$ under the light pattern $\mathbf{L}_p$ as follows

$$I_p(\mathbf{x}_i) = \sum_{l=1}^{m} R_{\mathbf{x}_i}(\Phi_l) L_p(\Phi_l) = \mathbf{R}_{\mathbf{x}_i}^T \mathbf{L}_p \tag{1}$$

where $\mathbf{R}_{\mathbf{x}_i} = [R_{\mathbf{x}_i}(\Phi_1), \ldots, R_{\mathbf{x}_i}(\Phi_m)]^T$ is an m-dimensional vector with the elements $R_{\mathbf{x}_i}(\Phi_l)$ being the *reflectance* of pixel $\mathbf{x}_i$ as a result of illumination from direction Φ_l, see again [DHT*00]. The above equation can be extended in order to consider all the n image pixels

$$\mathbf{I}_p = \mathbb{R}\mathbf{L}_p \tag{2}$$

where $\mathbf{I}_p = [I_p(\mathbf{x}_1), \ldots, I_p(\mathbf{x}_n)]^T$ is an n dimensional vector containing all image points, and $\mathbb{R} = [\mathbf{R}_{\mathbf{x}_1}, \ldots, \mathbf{R}_{\mathbf{x}_n}]^T$ is an $n \times m$ matrix of reflectance functions for all image points (we call it *reflectance matrix*). Note that the rows of $\mathbb{R}$ denote image pixels, while the columns correspond to different light source positions. Now the collection of p reference images of the scene illuminated by p lighting patterns can be expressed by

$$\mathbb{I}_{\mathbb{L}} = \mathbb{R}\mathbb{L} \tag{3}$$

where $\mathbb{I}_{\mathbb{L}} = [\mathbf{I}_1, \ldots, \mathbf{I}_p]$ is an $n \times p$ matrix containing the images of the object under different lights and $\mathbb{L} = [\mathbf{L}_1, \ldots, \mathbf{L}_p]$ is an $m \times p$ matrix representing the different lighting patterns used to illuminate the object. We now need to determine how the image of the scene illuminated by lighting patterns can be decoded into images of the scene illuminated by single light sources.

Consider acquiring a set of m reference images each of which is illuminated by a single point light source. The lighting in the i^{th} image can be represented by a vector $\mathbf{E}_i$ in which the i^{th} element has the value 1 and the remaining elements have the value 0. The matrix of all m single light source patterns can be written as $\mathbb{E} = [\mathbf{E}_1, \ldots, \mathbf{E}_m]$. Note that these light source patterns, as given by the columns of $\mathbb{E}$, form the standard basis; note also that $\mathbb{E}$ is the identity matrix. Now the images formed by these single light source patterns can be written as

$$\mathbb{I}_{\mathbb{E}} = \mathbb{R}\mathbb{E} \tag{4}$$

where $\mathbb{R}$ is the reflectance matrix described earlier. We need to find the linear transformation $\mathbb{D}$ that will decode the reference images acquired using the lighting patterns to recover the images that would be created under single light source illumination $\mathbb{E}$. By decode we mean we can find $\mathbb{I}_{\mathbb{E}}$ as

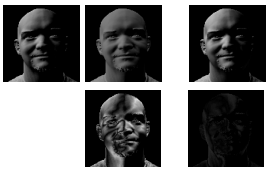
$$\mathbb{I}_{\mathbb{L}}\mathbb{D} = \mathbb{R}\mathbb{L}\mathbb{D} = \mathbb{R}\mathbb{E} = \mathbb{I}_{\mathbb{E}}. \tag{5}$$

If the lighting patterns are linearly independent and the number of patterns p is greater than or equal to the number of single light sources m, then we can decode the lighting patterns exactly using the decoder matrix $\mathbb{D} = \mathbb{L}^{-1}$. If the number of lighting patterns p is less than the number of single light sources m then the $rank(\mathbb{L}) < m$. In this case we cannot invert the matrix of lighting patterns $\mathbb{L}$ and must settle for and approximate decoding as given by $\mathbb{D} = \mathbb{L}^+ = (\mathbb{L}^T\mathbb{L})^{-1}\mathbb{L}^T$. In both cases, we write

$$\mathbb{I}_{\mathbb{D}} = \mathbb{I}_{\mathbb{L}}\mathbb{D} \approx \mathbb{R}\mathbb{E} \tag{6}$$

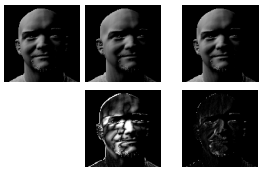
where $\mathbb{I}_{\mathbb{D}}$ is the matrix of decoded reference images. Relighting can then be achieved taking the desired linear combinations of the decoded images $\mathbb{I}_{\mathbb{D}}$. For example, imagine you want to relight a scene with user specified illumination $\mathbf{L}_{new}$, we get the

Sub-Basis Error



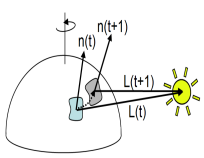
Sub-Basis Error: Produced when using a reduced number of lighting patterns. **Upper row**, left to right: Ground truth image; image decoded using only 3 lighting patterns; image decoded using 9 lighting patterns. **Lower row:** Error when using 3 lighting patterns and 9 lighting patterns, respectively.

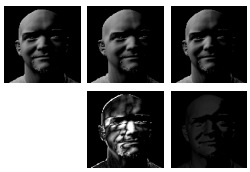
Alignment Error



Alignment Error: Produced by inaccuracies in the optical flow alignment. **Upper row:** $\mathbf{I}^{t_1}$; $\mathbf{I}^{t_2}$; aligned image $\mathbf{I}^{t_2}(\mathbf{w}(\mathbf{x};\mathbf{q}_{21}))$. **Lower row:** Error before the alignment $|\mathbf{I}^{t_1} - \mathbf{I}^{t_2}|$; Error after the alignment, $|\mathbf{I}^{t_1} - \mathbf{I}^{t_2}(\mathbf{w}(\mathbf{x};\mathbf{q}_{21}))|$. This error will increase as the displacement in the scene becomes larger; thus it will increase as more lighting patterns are used.

Orientation Error





Orientation Error: The displacement of a patch between frame t_1 and t_2 can cause a change in orientation with respect to the camera or light sources. This will cause errors as described below even when perfect geometric alignment is achieved before decoding the image. **Left:** Diagram showing how the movement between times t_1 and t_2 of a patch w.r.t. a light source effects its appearance. When frames t_1 and t_2 are used to decode image values at point $\mathbf{x}$ they induce error. **Right, upper row:** $\mathbf{I}^{t_1}$; $\mathbf{I}^{t_2}$; aligned image $\mathbf{I}^{t_2}(\mathbf{w}(\mathbf{x};\mathbf{q}_{21}))$. In this case, we assume a *perfect* alignment, so that the error is produced only by the change of appearance of the head because of its relative movement w.r.t. the light sources. **Right, lower row:** Error before alignment, $|\mathbf{I}^{t_1} - \mathbf{I}^{t_2}|$; error even after perfect alignment, $|\mathbf{I}^{t_1} - \mathbf{I}^{t_2}(\mathbf{w}(\mathbf{x};\mathbf{q}_{21}))|$. This error will also increase as the displacement in the scene becomes larger; thus it will increase as more lighting patterns are used.

Figure 1: *Sources of error when relighting video sequences.*

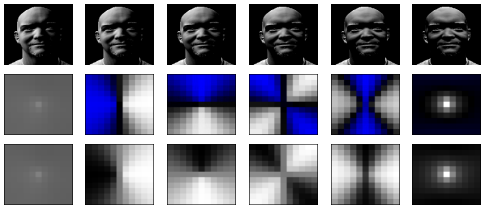


Figure 2: *Computing the optimal lighting basis using SVD. First row: Object illuminated a single light source in different positions (columns of matrix $\mathbb{I}_{\mathbb{E}}$). Second row: Lighting patterns from the optimal lighting basis (rows of matrix $\mathbb{L}^*$). They contain both positive values, shown in grey, and negative values, shown in blue. Third row: Offset and scaling of the optimal lighting basis in order to make all its values positive.*

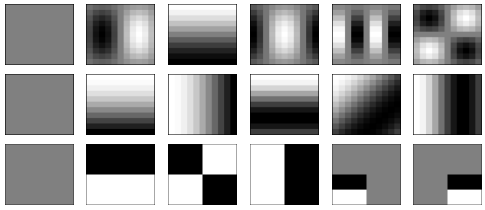


Figure 3: *The first 6 patterns of three object-independent lighting bases: spherical harmonics (first row), Fourier (second row) and Haar basis (third row). Compare these bases to the optimal lighting basis in Fig.2.*

image of the scene under this illumination as given by $\mathbf{I}_{new} = \mathbb{I}_{\mathbb{D}}\mathbf{L}_{new} \approx \mathbb{R}\mathbf{L}_{new}$.

3.2 Relighting in Video

For relighting moving objects, we illuminate the object with a sequence of p lighting patterns, synchronizing the lighting system with the camera, such that each image of the object is acquired with a single light pattern. To relight the video in a post-processing stage we first need to perform an optical flow alignment between consecutive frames. However, at this point we will delay the details of the optical flow alignment method until later in the paper.

Let $\mathbf{I}^{t_1}(\mathbf{x})$ denote a frame acquired at time t_1 under illumination given by lighting pattern $\mathbf{L}_{\mathrm{mod}\,(t_1,p)}$ where $\mathrm{mod}\,(t_1,p)$ is remainder of t_1 divided by p. This $\mathrm{mod}\,(\cdot)$ addresses the fact that we are cycling through the p patterns over the course of the sequence. Let $\mathbf{I}^{t_1}(\mathbf{w}(\mathbf{x};\mathbf{q}_{12}))$ denote the frame $\mathbf{I}^{t_1}(\mathbf{x})$ acquired at time t_1 but warped in such a way that it is aligned with $\mathbf{I}^{t_2}(\mathbf{x})$ acquired at time t_2. The warping function $\mathbf{w}(\mathbf{x};\mathbf{q}_{12})$ takes the pixel $\mathbf{x}$ in the time frame of $\mathbf{I}^{t_1}$ and maps it to the subpixel location $\mathbf{w}(\mathbf{x};\mathbf{q}_{12})$ in frame $\mathbf{I}^{t_2}$. Note that $\mathbf{q}_{12}$ is the vector of warping parameters needed for the mapping from t_1 to t_2.

In order to decode the lighting patterns at any given time t, we require a set of frames of the scene – in the pose of frame t – taken under p different lighting patterns. To do this, we take a window of p frames centered at frame t and align each of these $p-1$ frames to frame t. (We align only $p-1$ as frame t is already aligned.) Let the first frame in the window be called frame t_1, let the middle frame be t, and let the last frame be t_p This gives a matrix of p aligned frames that can be written as:

$$\mathbb{I}^t_{\mathbb{L}} = \left[\mathbf{I}^{t_1}(\mathbf{w}(\mathbf{x};\mathbf{q}_{1t})),\ldots,\mathbf{I}^{t_p}(\mathbf{w}(\mathbf{x};\mathbf{q}_{pt}))\right],$$

where $t_i = t - \lfloor \frac{p}{2} \rfloor + i - 1$. Now to relight the video sequence, we decode each frame using the same decoding matrix $\mathbb{D}$ to get $\mathbb{I}^t_{\mathbb{D}} = \mathbb{I}^t_L\mathbb{D}$. Finally every frame can be relit much like the static case. If the user specifies the lighting at time t as $\mathbf{L}^t_{new}$, we can compute the relit image in frame t as $\mathbf{I}^t_{new} = \mathbb{I}^t_{\mathbb{D}}\mathbf{L}^t_{new}$.

3.3 Sources of Error

If the object being relit remains static, error in relighting and decoding arises only if we use subset of the lighting basis, i.e., we use p m-dimensional vectors (where $p < m$) to span the m-dimensional space of lights. This error, which we call the *sub-basis error*, is reduced by using a higher number of light patterns and converges to zero when $p = m$. If the object moves, we need to consider two additional sources of error. There is intrinsic error in the alignment from the optical flow algorithm; we call this the *alignment error*. Since we need only to align the images that are inside a temporal window of length p, this error is reduced by decreasing the size of the temporal window, i.e., by decreasing the number of lighting patterns. This error can also be decreased by increasing the frame rate of the camera. In our experiments we used a camera capable of acquiring 30 frames per second (fps), but faster cameras are readily available. Finally, there is the error produced as an object in the scene changes its orientation with respect to the camera or light sources; we call this the *orientation error*. Even if the displacement in the scene is perfectly realigned with the alignment algorithm, the displacement itself can induce relative orientation change of surface points with respect to the camera and the light sources. A simple rotation of the object between frames will induce this error. And, if the camera and light source are close to the scene, a translation of the object can induce this error as well. As with the alignment error, the orientation error can again be reduced either by using a low number of lighting patterns p or by increasing the camera's frame rate.

In Fig. 1 we depict all three of these sources of error for an experiment using images generated from a synthetic 3-D head. As the data used was synthetic, we were able to isolate and separately display each source of error. A brief discussion of each is included within the figure.

4 Selecting the Optimal Lighting Basis

We now concentrate on the selection of the optimal lighting basis $\mathbb{L}$ for relighting video sequences. As we have shown at the end of the previous section, the alignment and orientation errors can be reduced with the use of fewer lighting patterns. However, the sub-basis error increases as the number of lighting patterns decreases. Therefore, we need to find the lighting basis that minimizes the sub-basis error.

Our goal is to synthesize images of the scene under new illumination conditions such that the synthesized images are as close (in an L^2 sense) to real images as possible. Equivalently,

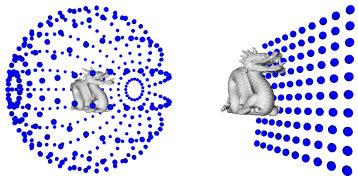


Figure 4: *The two different configurations of the light sources used in the synthetic experiments. Left: Light sources lying on a sphere. Right: Light sources lying on a plane.*

for a desired level of accuracy, we want to find the lighting basis that minimizes the number of reference images that need to be acquired – for reasons detailed in Section 3.3. It is important to note that this optimal lighting basis is a complex function of camera and scene properties and, thus, is what we call an *object-dependent* lighting basis. Yet, we will show subsequently that this optimal basis can be determined using singular value decomposition (SVD) on images gathered during a calibration step before acquisition. For a typical scene this calibration can be done in a matter of seconds before video capture. Consider again acquiring a set of m reference images each of which is illuminated by a single, compact point light source. The lighting in the i^{th} image can be represented by a vector $\mathbf{E}_i$ in which the i^{th} element has the value 1 and the remaining elements have the value 0. The matrix of all m single light source patterns can be written as $\mathbb{E} = [\mathbf{E}_1, \ldots, \mathbf{E}_m]$. Note that these light source patterns as given by the columns of $\mathbb{E}$ form the standard basis; note also that $\mathbb{E}$ is the identity matrix. Now the images formed by these single light source patterns can be written as

$$\mathbb{I}_{\mathbb{E}} = \mathbb{R}\mathbb{E} \tag{7}$$

where $\mathbb{R}$ is the reflectance matrix described earlier. Now let's say that instead of illuminating the scene with a sequence of single light sources as given by $\mathbb{E}$, we illuminate the scene with the optimal lighting basis denoted by $\mathbb{L}^*$. Under this illumination we get a different set of reference images $\mathbb{I}_{\mathbb{L}^*}$ as

$$\mathbb{I}_{\mathbb{L}^*} = \mathbb{R}\mathbb{L}^*. \tag{8}$$

Now there exits a linear transformation $\mathbb{D}^* = (\mathbb{L}^*)^{-1}$ that will decode the reference images acquired using the optimal lighting basis to recover the images that would be created under single light source illumination $\mathbb{E}$. By decode we mean we can find $\mathbb{I}_{\mathbb{E}}$ as

$$\mathbb{I}^*_{\mathbb{L}}\mathbb{D}^* = \mathbb{R}\mathbb{L}^*\mathbb{D}^* = \mathbb{R} = \mathbb{I}_{\mathbb{E}}. \tag{9}$$

But how is the optimal basis chosen given that we have measured $\mathbb{I}_{\mathbb{E}}$? Recall that we want to find the lighting basis that minimizes the number of reference images that need to be acquired. Our goal is to acquire many fewer than m reference images, yet still be able to decode these images to approximate, with the highest possible accuracy, the full set of reference images under point source illumination. If we perform a singular value decomposition (SVD), we can write

$$\mathbb{I}_{\mathbb{E}} = \mathbb{U}\mathbb{S}\mathbb{V}^T \tag{10}$$

where $\mathbb{U}$ is an orthogonal matrix; $\mathbb{S}$ is a diagonal matrix whose non-zero elements are singular values in decreasing order of

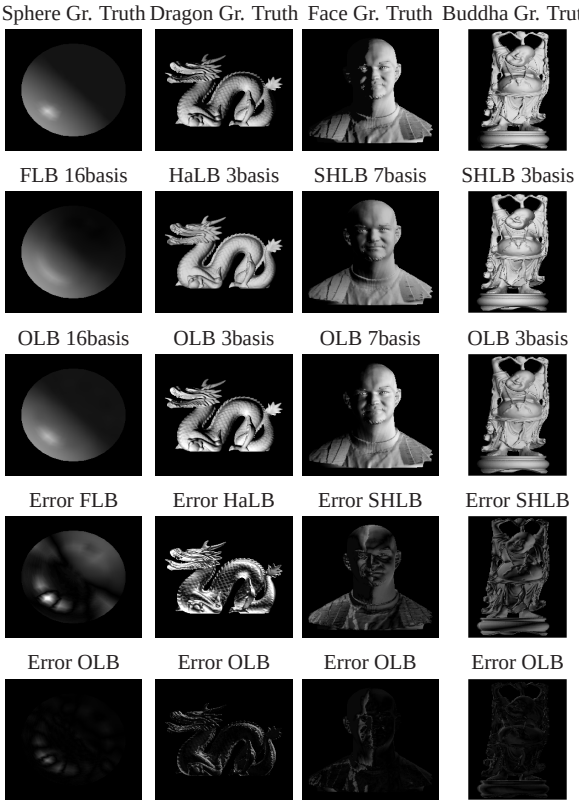


Figure 5: *Examples of reconstructed images, and reconstruction error for the synthetic static experiments (each column corresponds to a different experiment). For these examples we see that with the same number of basis images, the optimal lighting basis performs much better than the Fourier, Haar, and spherical harmonics lighting basis.*

$\mathbb{I}_{\mathbb{E}}$; $\mathbb{V}^T$ is an orthogonal matrix. We choose as the optimal basis $\mathbb{L}^* = \mathbb{V}$ and the optimal decoding matrix $\mathbb{D}^* = \mathbb{V}^T$; we claim they are optimal in the following sense. Let $\mathbb{L}$ be an $m \times p$ matrix formed from the first p columns of $\mathbb{L}^*$. Likewise let $\mathbb{D}$ be a $p \times m$ matrix formed from the first p rows of $\mathbb{D}^*$. Finally, let the p images acquired under illumination $\mathbb{L}$ be denoted again as $\mathbb{I}_{\mathbb{L}}$. We then write the following approximation

$$\mathbb{I}_{\mathbb{E}} \doteq \mathbb{I}_{\mathbb{D}} = \mathbb{I}_{\mathbb{L}}\mathbb{D}. \tag{11}$$

Now for any choice of p, the lighting patterns $\mathbb{L}$ extracted from the optimal lighting basis $\mathbb{L}^*$ minimize the sub-basis error in the above approximation, i.e., minimize the Frobenius norm $\|\mathbb{I}_{\mathbb{E}} - \mathbb{I}_{\mathbb{D}}\|$. Note that since the matrix $\mathbb{L}^*$ contains negative values in some elements, we must offset and scale each basis to range between 0 and 1 to make a physically feasible lighting basis. Fig. 2 shows an optimal lighting basis for a synthetic 3-D head computed from a superset of the images in the first row. Compare this optimal lighting basis to the bases shown in Fig. 3.

5 Experiments with Synthetic Data

We now show, using several experiments with synthetic data, that the scene-dependent optimal lighting basis (OLB) performs better than the Fourier lighting basis (FLB), Haar

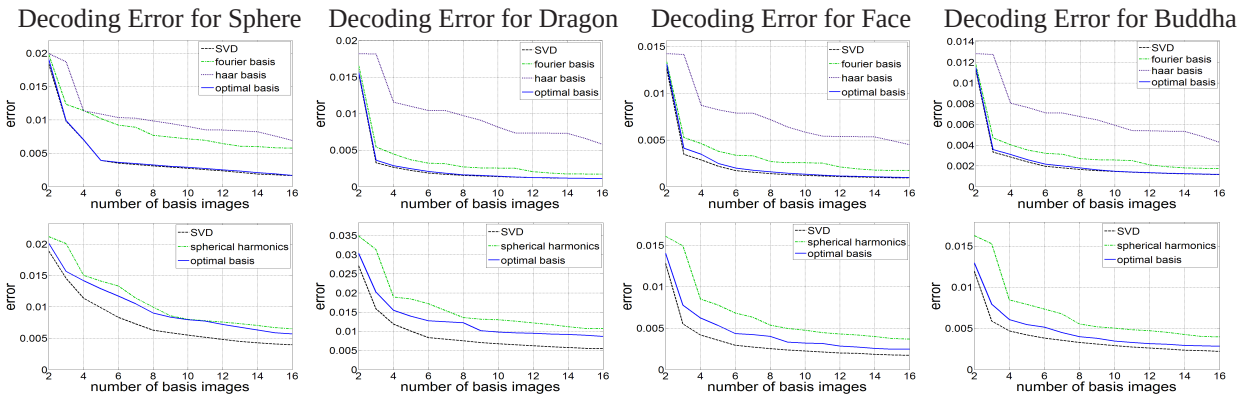


Figure 6: *The sub-basis errors in the synthetic experiments for the different types of lighting bases, plotted as a function of the number of basis images, for the synthetic sphere, dragon, face and buddha statue. Here, we have included plots for SVD approximation of the original data (see text for details).*

lighting basis (HaLB) and spherical harmonic lighting basis (SHLB). We present these results for both static as well as moving objects. In the case of moving objects, since we are using synthetic data, we can assume perfect alignment so as to focus on the sub-basis errors.

5.1 Performance Comparison for Static Objects

In the first experiment, we compare the four lighting bases: FLB, HaLB, SHLB, and OLB. We use the bases as illumination patterns to render images of static synthetic objects. Then, for each case, we recover (decode) single light source images from the rendered images. These recovered images are compared with rendered single light source images (ground truth) to compare the performances of the lighting bases. This experiment is done for several objects: a glossy sphere, and three non-convex objects: a human face, a buddha's statue and a dragon (courtesy of Cyberware). These models are assumed to have Lambertian reflectance with constant albedo.

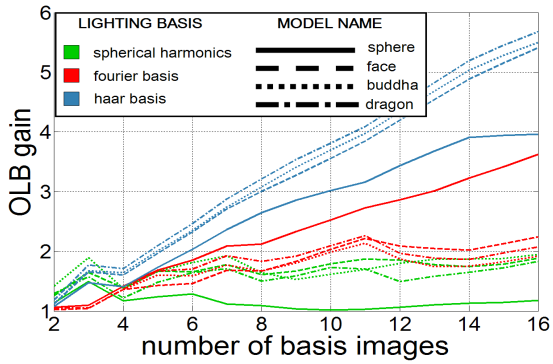
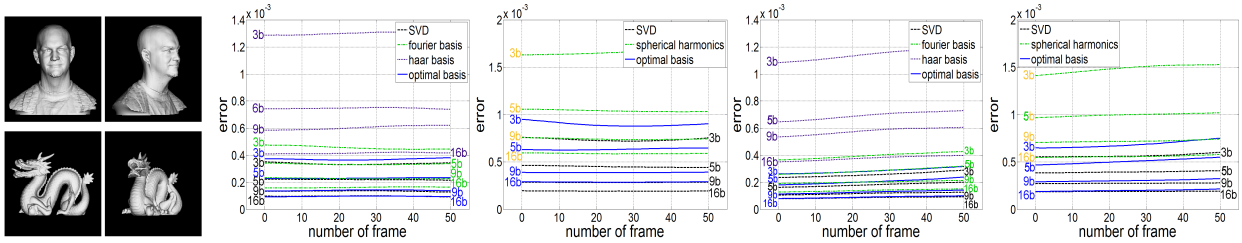


Figure 7: *Gains of the OLB w.r.t all other lighting basis, (for all synthetic and static experiments), plotted as a function of the number of basis images used. For any given number of optimal lighting basis images, the corresponding number of images of any other lighting basis that are needed to achieve the same reconstruction error equals the gain value. For instance, in the 'buddha' experiment instead of 6 OLB, we will need to use* $6 \times 1.8 \approx 11$ *SHLB images,* $6 \times 1.5 \approx 9$ *FLB images or* $6 \times 2.3 \approx 14$ *HaLB images.*

When comparing the OLB with the FLB and HaLB, we assume that there are 10×10 light sources that lie on a plane. The object is assumed to be placed in front of the plane. On the other hand, since spherical harmonics are suitable only for use on a sphere, to make our comparison fair, both OLB and SHLB patterns are represented through 20×20 light sources lying on a whole sphere (see Fig. 4). Our comparison is done using the following steps for each of the lighting bases (FLB, HaLB, SHLB, and OLB):

1. Render images of the object using single light sources.
2. Render images of the object using the lighting basis.
3. Decode single light source images of the object from the lighting basis images, as explained in Section 3.
4. Compute the sub-basis error (using the Frobenius norm) between the decoded and ground truth single light source images.

Fig. 5 shows several examples where the differences between the decoding results obtained using the OLB and the SHLB, HaLB or FLB are clearly visible, especially when the object has a 'complex' geometry. Fig. 6 shows the sub-basis errors for the different types of lighting bases in each of the experiments, plotted as a function of the number of basis images. We have also included the errors obtained when approximating the ground truth images using the most significative eigenvectors, resulting from the SVD decomposition of the single light source images. Note that these eigenvector images are not physically feasible through a relighting process because they contain negative values. We include these results just as a baseline case for comparison. Fig. 7 shows the gains of the OLB with respect all the other lighting basis plotted as a function of the number of basis images used. For any given number of optimal lighting basis images, the corresponding number of images of any other lighting basis that are needed to achieve the same reconstruction error equals the gain value. It is clear from this plot that the OLB is significantly more efficient than the others.



First and last frames Sequence of the rotating face Sequence of the rotating dragon

Figure 8: *Experiment with synthetic moving objects. Left: the first and last images in the rendered sequences. Right: Reconstruction error for both experiments, plotted as a function of the frames of the sequence (horizontal axis), for the different lighting basis types and different numbers of basis images used for each type (the number next to each plot).*

5.2 Performance Comparison for Moving Objects

The second experiment relates to decoding rendered video sequences of the 'face' and 'dragon' models, that rotate around the vertical axis. The rendered sequences have 50 frames, and the rotation between consecutive frames is 1 degree. Just like in the experiment with static objects, when comparing the performance of the SHLB and OLB, the object is assumed to be lit by a spherical source, while when comparing the OLB with the HaLB and FLB, the sources are placed on a plane. The illumination pattern is varied from one frame to the next based on the chosen lighting basis.

In Fig. 8(left) we show the first and the last frames of the sequences. Since these are synthetic examples, the alignment between frames is known to us and the only sources of error in the decoded images is due to the use of a lower number of basis images (sub-basis error) and the change in surface orientation with respect to the camera and the light sources (orientation error). Fig. 8(right) shows the reconstruction error (average over reconstructions of all the single sources) plotted as a function of the frames of the sequence, for each of the experiments. In each experiment, the different plots correspond to different lighting bases types and different numbers of basis images used for each type. From the plots in Fig. 8(right), we can see that optimal lighting basis performs much better than SHLB, FLB and HaLB, even in video sequences. For instance, in the synthetic face example, with 3 images of the optimal lighting basis we obtain decoding results similar to using 6 spherical harmonic, 5 Fourier and 16 Haar basis images. Note that the optimal lighting basis was computed using rendered single source images of only the initial orientation (seen in the first frames) of the face and dragon. Even though these bases were used for all other orientations, we see that the reconstruction error does not increase noticeably as the objects rotate.

This efficiency of our OLB is critical in the context of video relighting. The smaller the number of required lighting patterns, the easier it is to align the images and the lower is the orientation error.

6 Experiments with Real Scenes

In the preceding section, we have shown that our lighting basis is optimal for relighting moving objects in an ideal setting: Lambertian objects having constant albedo, linear light sources with equal power, a linear image acquisition system and no errors in the alignment. We now report experiments with real scenes. We use a setup that is calibrated to satisfy many of the assumptions we have made. We then apply our relighting method to static and dynamic scenes that include non-Lambertian surfaces. In each case we show that the use of the OLB enables us to produce relit videos of high quality.

6.1 Experimental Setup

Our setup is based on the system described in [SNB03]. The components of our setup are a color camera (Dragonfly IEEE-1394, color, 640×480 pixels) running at 30fps and a PC-controlled projector (Infocus LP820). The projector projects the basis patterns on a white wall, which in turn illuminates the scene (see Fig. 9). The scene is captured using the camera which is synchronized with the projector. The complete system has frame rate of 22fps. Note that a significantly higher frame rate can be achieved using a high-speed camera and a projector with a higher refresh rate. However, even with the current system, the efficiency of the optimal lighting basis allows us to capture and relight scenes with objects that move at reasonable speeds. Since our setup uses a planar surface as the source area, we will only compare the results of using the optimal lighting basis with the Fourier and Haar lighting basis. Note that spherical harmonics are inappropriate for such a setup as they are defined over the sphere.

6.2 System Calibration

One of the key assumptions we have made is that the light sources are linear and that the camera has a linear response. To this end, we have measured the radiometric response functions of the projector and the camera and used these response functions to linearize our system. We also need to ensure that there is no angular variation in source brightness with respect to the center of the scene. The main reason for such a variation is that we are generating our sources on a plane rather than a sphere. Since the system has been radiometrically linearized, a single image of the plane taken with a uniform image projected by the projector reveals the fall-off function. With these simple calibrations done, our system satisfies the source and camera linearity assumption we have made.

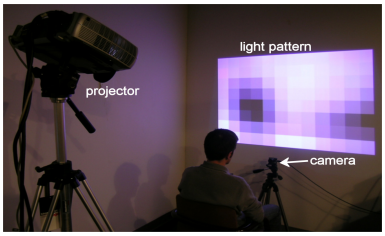


Figure 9: *Experimental setup used for the real experiments*

6.3 Relighting Static Objects

In order to validate our theoretical results and our empirical results with synthetic data, we conducted experiments with static scenes, before moving on to dynamic ones. Fig. 10 shows the objects used for this experiment, a mannequin head and a statue (bust of David). In both cases, a 6×8 grid of patches were used as the individual controllable sources. Note that the Lambertian assumption is not valid for both objects as they each have specular components in their reflectance. Even so, due to the additive nature of light [Bus60], the specular reflections can also be reproduced using a linear combination of measurements as long as these reflections do not saturate the camera or produce complex interreflections. In Fig. 10 we show examples of reconstructed images for each of the objects. We see that for the same number of basis, OLB perform better than FLB and HaLB. The differences in the reconstruction quality and error images are clearly visible. In Fig. 11, the errors in decoding are plotted as a function of the number of basis images used, for the Fourier, Haar and the optimal lighting bases. As expected, the optimal lighting basis is significantly more efficient than the other bases.

6.4 Relighting Real Moving Objects

The challenge when relighting moving objects is alignment. Specifically, we need to align points across a window of frames so that when we decode the light patterns we do not blur the information from different points on the object. To do this, we estimate the optical flow over the sequence of images. This problem is made more difficult by the fact that the illumination varies from one frame to the next due to the use of a lighting basis. We accomplish the alignment in two stages. First, we apply background subtraction to obtain the moving regions. We use the background subtraction approach presented in [MNSS04]. In this work, the object is parameterized by geometric and appearance features, which are adapted online using a particle filter formulation. We found this approach to be robust to the abrupt changes in illumination caused by the use of a lighting basis. Second, to align each segmented region in consecutive frames, we used an implementation of Baker et al.'s [BM04] illumination insensitive optical flow alignment algorithm.

Figures 12,13 and 14, show relighting results for different scenes. In Fig.12 we show the results of relighting a moving tennis ball. In an off-line procedure, we acquired images of the ball illuminated by single light sources where the sources form an 8×12 grid on the source plane. Using these images we computed the optimal lighting basis, and the first 3 of these

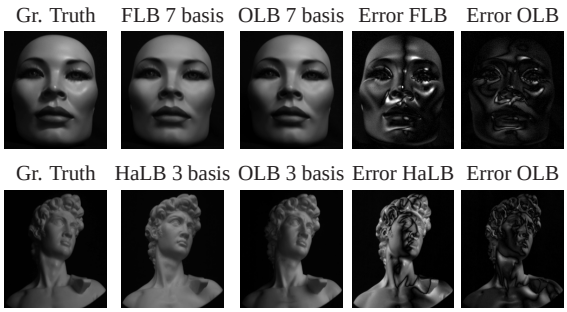


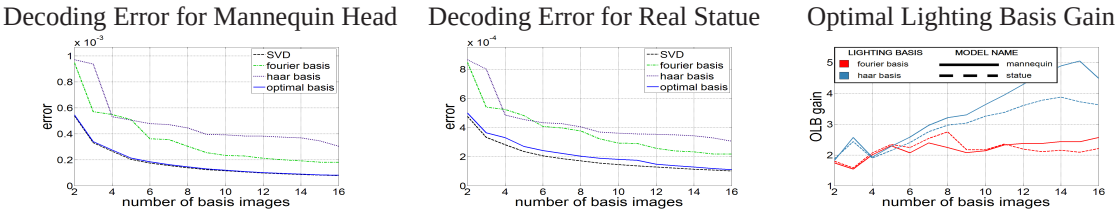
Figure 10: *Examples of reconstructed images for the mannequin head and the statue.*

light patterns were used to illuminate the ball while it was moved. The final sequence is acquired at 22fps, and contains 100 frames (the size of the light pattern grid, the frame rate, and the number of optimal lighting basis was the same for all of these experiments). In Fig.12, we show results of relighting the moving ball with a white point light source that moves smoothly across the horizontal axis. Similar results on the relighting of a human face are shown in Fig.13. In this case, the original sequence has 400 frames. Here, we have included the results of relighting the face with the illumination from New York City's Times Square, which was captured by simply panning a video camera with a wide-angle lens.

As we have previously mentioned, one of the advantages of the camera-projector setup that we are using is its scalability. Using this setup we can relight small objects as well as large scenes. In Fig.14, we present results for a room scene with a moving person. In this case, the original sequence has 400 frames. We also use this example to demonstrate the concept of local relighting, where different parts of the scene are lit by different sources. Notice the green light that is focused on the face of the person and the blue light that is shone on the cup. From the cast shadows, one can see that these sources illuminate their respective regions from different directions.

7 Conclusions

We have proposed the use of an *object-dependent* lighting basis for image and video relighting. This lighting basis is optimal in the sense that it minimizes the number of reference images that are needed for relighting. The basis is generated off-line, computing SVD over the set of images of the still object illuminated by single light sources. Once the light basis is computed, a subset of the light patterns is used for illuminating the objects in either still scenes or video. Our analysis shows that the lighting method used here is indeed more effective for video relighting. Yet, challenges remain for improving the performance of the system. In particular, we are investigating ways of using higher frame rate cameras to reduce the errors (sub-basis, alignment, and orientation) described in the text and allow for faster motion within the video sequences. Furthermore, we are considering the use of adaptive *motion-dependent* relighting bases which adapt based on the motion detected in the scene.



(a) (b) (c)

Figure 11: *(a) Reconstruction errors for the mannequin in Fig. 10 plotted as a function of the number of basis images used, for the Fourier, Haar and the optimal lighting bases. (b) Reconstruction errors for the statue in Fig. 10 plotted as a function of the number of basis images used, for the Fourier, Haar and the optimal lighting bases. (c) The gain of the OLB with respect FLB and HaLB for both experiments, plotted as a function of the number of basis images used.*

Acknowledgments

This work was partially supported by CICYT project DPI2004-05414, by a fellowship from the Spanish Ministry of Science and Technology, and by the National Science Foundation Award No. IIS-03-08185 'Complex Reflectance, Texture and Shape: Methods and Representations for Object Modeling'.

References

[BM04] BAKER S., MATTHEWS I.: Lucas-kanade 20 years on: A unifying framework. *IJCV 56(3)* (2004), 221–255.

[Bus60] BUSBRIDGE I.: *The Mathematics of Radiative Transfer*. Cambridge University Press, 1960.

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H., SAROKIN W.: Acquiring the reflectance field of a human face. In *SIGGRAPH 2000* (2000), ACM, pp. 145–156.

[DSG91] DORSEY J., SILLION F., GREENBERG D.: Design and simulation of opera lighting and projection effects. In *SIGGRAPH 1991* (1991), pp. 41–50.

[DWT*02] DEBEVEC P., WENGER A., TCHOU C., GARDNER A., WAESE J., HAWKINS T.: A lighting reproduction approach to live-action composing. In *SIGGRAPH 2002* (2002), vol. 21(3), ACM, pp. 547–556.

[GBK01] GEORGHIADES A., BELHUMEUR P., KRIEGMAN D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI 23(6)* (2001), 643–660.

[GKB98] GEORGHIADES A., KRIEGMAN D., BELHUMEUR P.: Illumination cones for recognition under variable lighting: Faces. In *Proc. CVPR* (1998), IEEE Computer Society, p. 52.

[GTW*04] GARDNER A., TCHOU C., WENGER A., HAWKINS T., DEBEVEC P.: Postproduction re-illumination of live action using time-multiplexed lighting. In *SIGGRAPH 2004 (Poster)* (2004), ACM.

[Hal94] HALLINAN P.: A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. CVPR* (1994), IEEE Computer Society, pp. 995–999.

[HGT*04] HAWKINS T., GARDNER A., TCHOU C., F.GORANSSON, DEBEVEC P.: Animatable facial reflectance fields. In *Proc. EGSR* (2004).

[KBMK01] KOUDELKA M., BELHUMEUR P., MAGDA S., KRIEGMAN D.: Image-based modeling and rendering of surfaces with arbitrary brdfs. In *Proc. ICCV* (2001), pp. 568–575.

[LKG*03] LENSCH H., KAUTZ J., GOESELE M., HEIDRICH W., SEIDEL H.: Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graphics 22* (2003), 234–257.

[LWS02] LIN Z., WONG T., SHUM H.: Relighting with the reflected irradiance field: Representation, sampling and reconstruction. *IJCV 49* (2002), 229–246.

[MGW01] MALZBENDER T., GELB D., WOLTERS H.: Polynomial texture maps. In *SIGGRAPH 2001* (2001), ACM, pp. 519–528.

[MNSS04] MORENO-NOGUER F., SANFELIU A., SAMARAS D.: Fusion of a multiple hypotheses color model and deformable contours for figure ground segmentation in dynamic environments. In *Proc. Work. ANM (in CVPR'04)* (2004).

[MPN*02] MATUSIK W., PFISTER H., NGAN A., BEARDSLEY P., ZIEGLER R., MCMILLAN L.: Image-based 3d photography using opacity hulls. In *SIGGRAPH 2002* (2002), ACM, pp. 427–437.

[NBB04] NAYAR S., BELHUMEUR P., BOULT T.: Lighting sensitive display. *ACM Trans. Graphics 23*, 4 (2004), 963–979.

[NN04] NISHINO K., NAYAR S.: Eyes for relighting. *ACM Trans. Graphics. 23*, 3 (2004), 704–711.

[NSD94] NIMEROFF J., SIMONCELLI E., DORSEY J.: Efficient re-rendering of naturally illuminated environments. In *Proc. EGSR* (1994), pp. 359–373.

[PSS99] PIGHIN F., SZELISKI R., SALESIN D.: Resynthesizing facial animation through 3d model-based tracking. In *Proc. ICCV* (1999), pp. 143–150.

[RH01] RAMAMOORTHI R., HANRAHAN P.: A signal-processing framework for inverse rendering. In *SIGGRAPH 2002* (2001), ACM, pp. 117–128.

[SNB03] SCHECHNER Y., NAYAR S., BELHUMEUR P.: A theory of multiplexed illumination. In *Proc. ICCV* (2003), vol. 2, pp. 808–815.

[WFHL02] WONG T., FU C., HENG P., LEUNG C.: The plenoptic illumination function. *IEEE Trans. Multimedia 4(3)* (2002), 361–371.

[WHON97] WONG T., HENG P., OR S., NG W.: Image-based rendering with controllable illumination. In *Proc. EGSR* (1997), pp. 13–22.

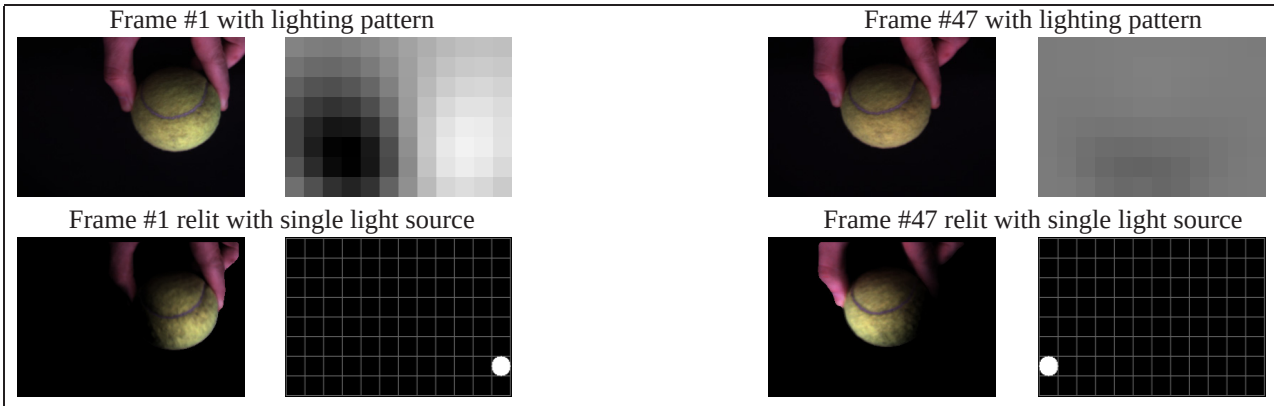


Figure 12: *Relighting a tennis ball. Upper row: two basis image frames of a tennis ball and the lighting pattern used to produce them. Lower row: Relighting results with a single white light source moving in the horizontal axis. The grid in the lighting pattern shows the distribution of the original light sources used to generate the lighting pattern.*

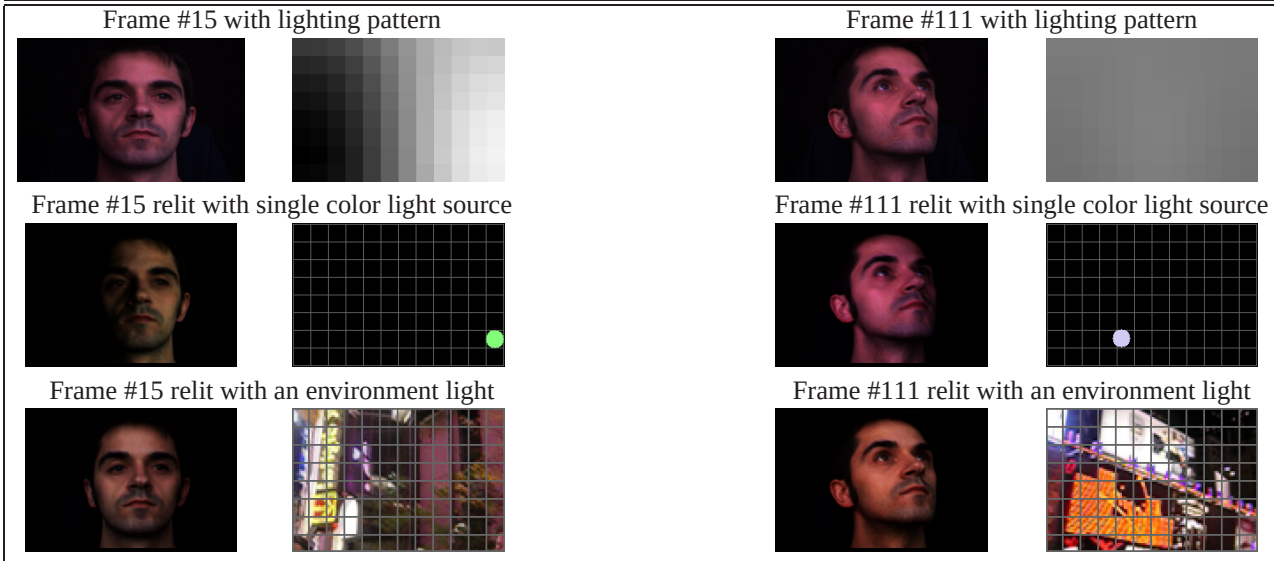


Figure 13: *Relighting a face. Upper row: two basis image frames of a face and lighting patterns used to produce them. Middle row: Relighting results with a single color light source moving in the horizontal axis. Lower row: Relighting results with lighting from Times Square.*

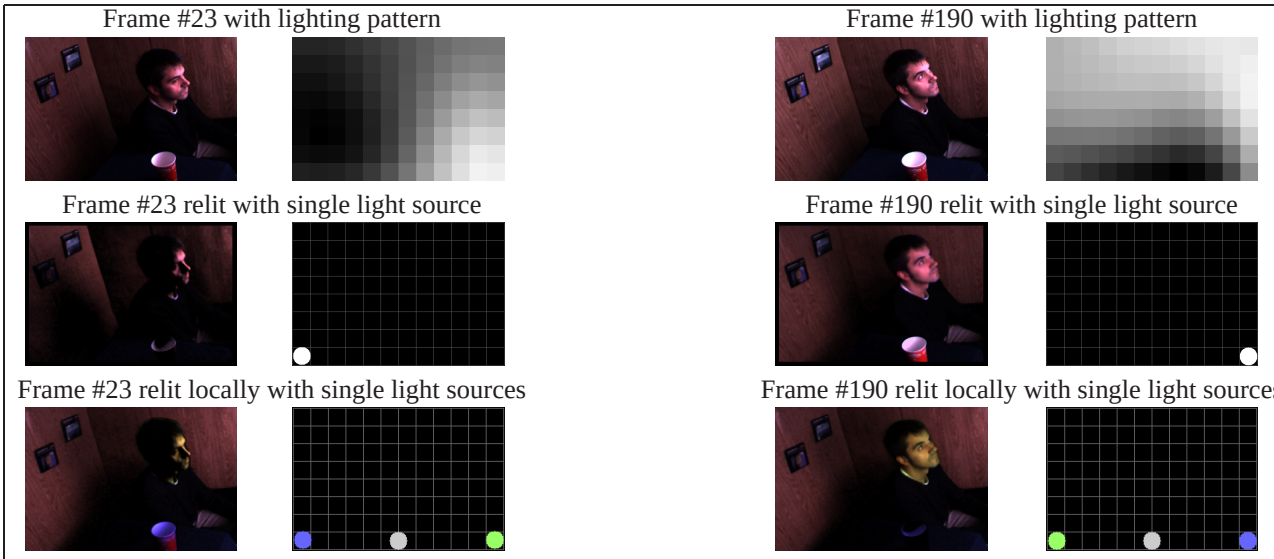


Figure 14: *Relighting a corner of a room. Upper row: two basis image frames and the lighting patterns used to produce them. Middle row: Relighting results with a single white light source moving in the horizontal axis. Lower row: Local relighting of the scene. The gray light source is a frontal light illuminating the whole scene. The blue light locally relights the cup while the green light focuses on the face.*