The Dimensionality of Scene Appearance

Rahul Garg · Hao Du · Steven M. Seitz · Noah Snavely

Abstract Low-rank approximation of image collections (e.g., via PCA) is a popular tool in many areas of computer vision. Yet, surprisingly little is known justifying the observation that images of an object or scene tend to be low dimensional, beyond the special case of Lambertian scenes. This paper considers the question of how many basis images are needed to span the space of images of a scene under real-world lighting and viewing conditions, allowing for general BRDFs. We establish new theoretical upper bounds on the number of basis images necessary to represent a wide variety of scenes under very general conditions, and perform empirical studies to justify the assumptions. We then demonstrate a number of novel applications of linear models for scene appearance for Internet photo collections. These applications include image reconstruction, occluder-removal, and expanding field of view. Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords BRDF · Dimensionality

1 Introduction

Real world scenes vary in appearance as a function of viewpoint, lighting, weather and other effects. What is the *dimensionality* of this appearance space? More specifically, suppose you stacked all photos taken of a particular scene as rows in a matrix – what is the rank of that matrix?[1]

It is well known that certain types of image collections tend to be low-rank in practice, and can be spanned using linear combination of a small number of basis views computed using tools like Principle Component Analysis (PCA) or Singular Value Decomposition (SVD). First exploited in early work on eigenfaces (Kirby and Sirovich (1990); Turk and Pentland (1991)), these rank-reduction methods have become the basis for a broad range of successful applications in recognition (Pentland et al. (1994); Murase and Nayar (1995)), tracking (Hager and Toyama (1996)), background modeling (Oliver et al. (2000)), image-based rendering (Wang et al. (2001)), BRDF modeling (Hertzmann and Seitz (2003); Matusik et al. (2003)), compression and other domains.

In spite of the wide-spread use of rank-reduction on images, however, there is little theoretical justification that appearance space should be low-rank in general. An exception is the case of Lambertian scenes, for which a number of elegant results exist. Shashua (1992) proved that three images are sufficient to span the full range of images of a Lambertian scene rendered under distant lighting and a fixed viewpoint, neglecting shadows. Belhumeur and Kriegman (1998) considered the case of attached shadows, observing that the valid images lie in a restricted range of 3D subspace which they called the *illumination cone*. Basri and Jacobs (2003), and Ramamoorthi and Hanrahan (2001) independently showed that the illumination cone is well approximated with 9 basis images. Ramamoorthi more recently (Ramamoorthi (2002)) improved this bound to 5 images, bringing the theory in line with empirical studies on the dimensionality of face images (Epstein et al. (1995)).

Very little is known, however, about the dimensionality of images of *real-world scenes*, composed of real shapes, BRDFs, and illumination conditions. Consider, for example, the images of tourist sites on photo sharing websites like flickr.com, which exhibit vast changes in appearance. While it may seem difficult to prove strong results about such collections, a key property of real-world scenes is that they are not random. In particular, man-made scenes tend to be dominated by a small number of surface orientations. And while BRDFs can be very complex, many real BRDFs can be well-approximated by a low-rank linear basis (Matusik

Rahul Garg
Box 352350, Seattle, WA 98195-2350, USA
Computer Science and Engineering
University of Washington
E-mail: rahul@cs.washington.edu

[1] By dimensionality, we refer to linear dimensionality in this paper.

et al. (2003)). Similar considerations apply for illumination; for example, studies have shown that the space of daylight spectra is roughly two- or three-dimensional (Sunkavalli et al. (2008)). Based on these observations, this paper introduces new theoretical upper bounds on the dimensionality of scene appearance (improving on previous results by Belhumeur and Kriegman (1998)). While we make a few limiting assumptions (distant lighting, distant viewer, no cast shadows, interreflections or subsurface scattering), these results bring the theory to the point where it can capture much of the extreme variability in these Internet photo collections. Further, many of the results are still seen to hold empirically even when these assumptions are violated.

The highlights of this paper include a factorization framework for analyzing dimensionality questions, introduced in section 2. Using this framework, we prove new upper bounds on the number of basis images, allowing for variable illumination direction and spectra, viewpoint, BRDFs, and convolution effects (e.g., blur). Importantly, all prior low-rank results for Lambertian scenes (Shashua (1992); Belhumeur and Kriegman (1998); Basri and Jacobs (2003); Ramamoorthi and Hanrahan (2001); Ramamoorthi (2002)) do not apply under variations in light spectrum (even if the images are grayscale). We introduce new results that allow the light spectrum to vary in certain ways (Section 2.3), greatly broadening the scope of application (e.g., to outdoors). In Section 3, we perform experiments on BRDF databases to empirically verify some of the assumptions made. Finally in Section 4, we demonstrate that low rank linear models can be used to model the appearance of outdoor scenes in Internet Photo Collections and conclude by showing a number of interesting applications of low-rank linear models to problems in computational photography (Section 4.4).

2 Rank of the Image Matrix

In this section, we present our theoretical results. We first introduce a new framework to analyze the factorization of images (Section 2.1) which yields new insights and results in Section 2.2. Finally, we introduce wavelength (Section 2.3) bringing the theory closer to the real world images captured by cameras.

Throughout the paper, we assume that images are lit by distant light sources and observed from distant viewpoints. We ignore indirect illumination effects like transparent and translucent materials, interreflections, cast shadows and subsurface scattering. Our theory does account for attached shadows, however. Initially, we also make the assumption that images are taken from a fixed viewpoint, which we relax in Section 2.2.3. Similarly, we begin by considering grayscale images captured at a constant illumination spectrum across images and talk about more complex and practical cases in Section 2.3.

2.1 Four Factorizations of the Image Matrix

Suppose we are given a set of n-pixel images of a scene, $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_m$ taken under varying illumination conditions. Consider the $m \times n$ matrix $\mathbf{M}$ obtained by stacking $\mathbf{I}_1, \mathbf{I}_2, .., \mathbf{I}_m$ as rows of the matrix. Each row of $\mathbf{M}$ is an image, and each column describes the appearance of a single pixel, say x, under different illumination conditions, referred to as the *profile* of the pixel and denoted by $\mathbf{P}_x$, where $\mathbf{P}_x(i) = \mathbf{I}_i(x)$.

Consider a factorization of $\mathbf{M}$ into the product of two rank-k matrices:

$$\mathbf{M}_{m\times n} = \mathbf{C}_{m\times k}\mathbf{D}_{k\times n}. \tag{1}$$

Such a factorization may be obtained by PCA or SVD, for instance. We present four different interpretations of such a factorization, shown in Figure 1.

2.1.1 Basis Images

First, the rows of $\mathbf{D}$ can be interpreted as basis images, denoted by $\mathbf{B}^I$, and the rows of $\mathbf{C}$ can be interpreted as coefficients. This interpretation, shown in Figure 1(a), is commonly used. For instance, in work on eigenfaces Kirby and Sirovich (1990), the eigenvectors obtained from PCA comprise the basis images (assuming mean subtracted data). Here each image $\mathbf{I}_i$ is a linear combination of basis images:

$$\mathbf{I}_i = \sum_{j=1}^{k} a_{ij}\mathbf{B}^I_j. \tag{2}$$

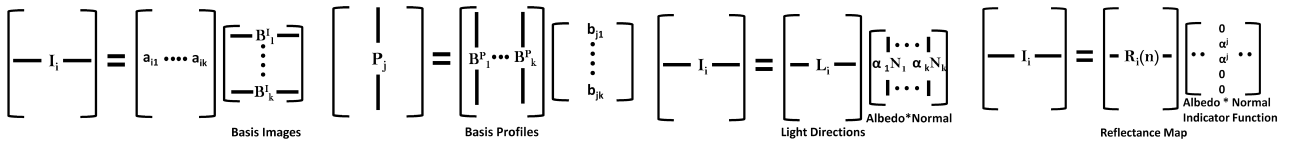
2.1.2 Basis Profiles

Another way to interpret this factorization is that each column (profile) $\mathbf{P}_j$ of $\mathbf{M}$ can be interpreted as a linear combination of columns of $\mathbf{C}$, with coefficients determined by columns of $\mathbf{D}$, as shown in Figure 1(b). In this interpretation, the columns of $\mathbf{C}$ form basis profiles, denoted by $\mathbf{B}^P$:

$$\mathbf{P}_j = \sum_{i=1}^{k} b_{ji}\mathbf{B}^P_i. \tag{3}$$

2.1.3 The Lambertian Case

For Lambertian scenes, neglecting *any* shadows, the rank of $\mathbf{M}$ is 3 Shashua (1992) , and the basis profiles and the basis images assume a special meaning, shown in Figure 1(c). $\mathbf{D}$ is a $3 \times n$ matrix, where the j^{th} column of $\mathbf{D}$ encodes the normal times the albedo at the j^{th} pixel in the scene. $\mathbf{C}$ is a $m \times 3$ matrix where the i^{th} row encodes the lighting direction times the light intensity for the i^{th} image. Hence, the basis images represent scene properties (normals and albedos) and the basis profiles encode illumination properties. In particular, each basis profile contains the light intensity along a coordinate axis for each image.



(a) (b) (c) (d)

Fig. 1: Four interpretations of factorization of image matrices. First, each image can be expressed as a linear combination of a set of basis images (a). Alternatively, the profile of each pixel can be expressed as a linear combination of a set of basis profiles (b). In the case of a Lambertian scene, the basis profiles and basis images assume special meaning (c). Finally, (d) shows the reflectance map interpretation.

2.1.4 The Reflectance Map Interpretation

The reflectance map Horn (1986), is defined for an image of a scene with a single BRDF as a function $R(\hat{\mathbf{n}})$ that maps scene normals to image intensity. $R(\hat{\mathbf{n}})$ can be encoded as an image of a sphere with the same BRDF as the scene and taken from the same viewpoint under identical illumination conditions.

We can alternately define $R_i(\hat{\mathbf{n}})$ using the rendering equation which under our assumptions, can be written as

$$\mathbf{I}_i(x) = \int_{\Omega} \alpha^x \rho^x(\omega', \omega) L_i(\omega')(-\hat{\omega'} \cdot \hat{\mathbf{n}}^x)_+ \mathbf{d}\omega' \tag{4}$$

where the integral is over a hemisphere of inward directions ω', ω is the viewing direction for point x, ρ^x is the reflectance function at point x (evaluated at ω', ω), $L_i(\omega')$ is the light arriving from direction ω' for image $\mathbf{I}_i$, and $\hat{\mathbf{n}}^x$ is the normal at x. The + subscript on the dot product indicates that it is clamped below to 0 to account for attached shadows.

Given this, we can define $R_i(\hat{\mathbf{n}})$ as

$$R_i(\hat{\mathbf{n}}) = \int_{\Omega} \rho(\omega', \omega) L_i(\omega')(-\hat{\omega'}.\hat{\mathbf{n}})_+ \mathbf{d}\omega' \tag{5}$$

where ρ^x has been replaced by ρ, as R_i represents a scene with a single BRDF.

Let us denote by $\mathbf{R}_i$ the image of the sphere when taken under identical illumination conditions as in image $\mathbf{I}_i$. Then we can write

$$\mathbf{I}_i{}^T = \mathbf{R}_i{}^T \mathbf{D} \tag{6}$$

where $\mathbf{D}$ is defined as:

$$\mathbf{D}(j,k) = \begin{cases} 1 \text{ if } \hat{\mathbf{n}}^k = \hat{\mathbf{m}}_j \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

where $\mathbf{D}(j,k)$ represent the value in the j^{th} row and k^{th} column of $\mathbf{D}$, $\hat{\mathbf{m}}_j$ is the normal at the j^{th} pixel of $\mathbf{R}_i$ and $\hat{\mathbf{n}}^k$ is the normal at the k^{th} pixel in the scene. The k^{th} column of $\mathbf{D}$ can be thought of as a *normal indicator function* $\mathbf{v}_k$.

It often happens that the BRDF is same across the scene save for a scaling factor (the albedo). The reflectance map factorization can also incorporate per pixel albedos if we define the k^{th} column of $\mathbf{D}$ as $\alpha^k \mathbf{v}_k$ where α^k is the albedo of the k^{th} pixel.

Now, observing that $\mathbf{D}$ does not depend on i, one can write $\mathbf{M} = \mathbf{CD}$ where $\mathbf{C}$ contains the reflectance maps $\mathbf{R}_i$'s stacked as rows.

2.2 Upper Bound on Rank of M

Belhumeur and Kriegman Belhumeur and Kriegman (1998) proved that, given an arbitrary scene with a single material and k_n distinct normals, the space of images of the scene taken from a fixed, distant viewpoint with distant lighting and no cast shadows is *exactly* k_n-dimensional. This result justifies the use of linear models for real-world scenes. For instance, many man-made scenes consist of large planar regions (such as walls and ground), and therefore contain only a small number of distinct normals. Curved surfaces may also be approximated by piecewise planar surfaces.

We first show how an upper bound of k_n can be seen to hold true for a scene with a single BRDF using the reflectance map interpretation of the factorization. Note that this upper bound is the same as that derived by Belhumeur and Kriegman Belhumeur and Kriegman (1998). However, we later extend our result under a number of different and more general settings.

From the reflectance map interpretation $\mathbf{M} = \mathbf{CD}$, it is easy to see that only k_n rows of $\mathbf{D}$ will be non-zero when the number of distinct normals in the scene is k_n. Hence, $rank(\mathbf{D}) \leq k_n$, which gives us an upper bound of k_n on $rank(\mathbf{M})$ as well. Belhumeur and Kriegman Belhumeur and Kriegman (1998) derive the same result but via a different route.

Now consider the more general case where there are k_ρ materials and k_n normals in the scene. In this case, we first define reflectance maps corresponding to every BRDF for each image, i.e.,

$$\mathbf{I}_i{}^T = \sum_{l=1}^{k_\rho} \mathbf{R}_{il}{}^T \mathbf{D}_l \tag{8}$$

where $\mathbf{D}_l$ now encodes the distribution of normals corresponding to the l^{th} material, i.e.,

$$\mathbf{D}_l(j,k) = \begin{cases} \alpha^k \text{ if } \hat{\mathbf{n}}^k = \hat{\mathbf{m}}_j \text{ and } \rho^k = \rho_l \\ 0 \text{ otherwise} \end{cases} \quad (9)$$

Hence, $\mathbf{M}=\sum_{l=1}^{k_\rho} \mathbf{C}_l\mathbf{D}_l$, which implies that $rank(\mathbf{M})\leq k_\rho k_n$. More precisely $rank(\mathbf{M})\leq\sum_{l=1}^{k_\rho} N(l)$ where $N(l)$ is the number of orientations corresponding to material l. Hence we have proven the following:

Theorem 1 *Consider a scene with k_ρ different BRDFs and k_n distinct normals. Consider the images $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_m$ of the scene obtained from a fixed distant viewpoint under different distant illuminations $L_{f_1}, L_{f_2}, \ldots, L_{f_m}$. Assuming that there are no cast shadows, the rank of the matrix M obtained by stacking $\mathbf{I}_1, \mathbf{I}_2..\mathbf{I}_m$ as rows is at most $k_\rho k_n$.*

It is also instructive to write $\mathbf{M} = \sum_{l=1}^{k_\rho} \mathbf{C}_l\mathbf{D}_l$ in the form $\mathbf{M} = \mathbf{CD}$ so that basis images and basis profiles can be explicitly defined. This can be done by stacking $\mathbf{C}_l$ side by side, i.e., $\mathbf{C} = [\mathbf{C}_1|\mathbf{C}_2|..|\mathbf{C}_{k_\rho}]$ and stacking $\mathbf{D}_l$ one over another. Finally, we can remove all zero rows from $\mathbf{D}$ and corresponding columns from $\mathbf{C}$ leaving at most $k_\rho k_n$ rows in $\mathbf{D}$, which correspond to basis images, and basis profiles are the remaining columns of $\mathbf{C}$. The columns of $\mathbf{D}$ are of the form $\mathbf{d}_k = \alpha^k \mathbf{v}_k$ where $\mathbf{v}_k$ is a $0,1$ vector that can be thought of as a *normal-material indicator function*.

The result may be modified to accommodate anisotropic BRDFs as well. For anisotropic materials, one needs to parametrize by both the *orientation* and the normal. Hence, one can derive the same bound where k_n now refers to the number of distinct orientations times normals in the scene.

In the following sections, we extend this result to a number of common scenarios.

2.2.1 Linear Families of BRDFs

While the world is composed of diverse materials, it has been argued Ramamoorthi and Hanrahan (2002); Matusik et al. (2003) that the space of BRDFs is low dimensional. We also verify this by conducting experiments on the CUReT Dana et al. (1999) database of BRDFs (Section 3.2).

Thus, we now generalize to the case when ρ^x is contained in the linear span of $\{\rho_1, \rho_2, \ldots, \rho_{k_\rho}\}$, i.e., $\rho^x = \Sigma_{l=1}^{k_\rho} c_l(x)\rho_l$. In this case, $\mathbf{I}_i$ can be represented as a sum of matrix products, ${\mathbf{I}_i}^T = \sum_{l=1}^{k_\rho} {\mathbf{R}_{il}}^T \mathbf{D}_l$, where

$$\mathbf{D}_l(j,k) = \begin{cases} \alpha^x c_l(x) \text{ if } \hat{\mathbf{n}}^k = \hat{\mathbf{m}}_j \\ 0 \text{ otherwise} \end{cases} \quad (10)$$

Hence the upper bound of $k_\rho k_n$ still holds, i.e., the rank is bounded by the dimensionality of BRDF family times the number of normals.

2.2.2 Low-dimensional BRDFs

Certain BRDFs tend to be *low-dimensional*. For example, three basis images suffice to span the images of a Lambertian scene captured under different lighting conditions, in the absence of shadows. Formally, we call a BRDF K-dimensional if the rank of the matrix $\mathbf{C}$ obtained by stacking reflectance maps obtained under arbitrary sampling of illumination conditions is always at most K. In the presence of such materials, the upper bound on dimensionality may be reduced to $\sum_{i=1}^{k_\rho} K(i)$, where $K(i)$ is the rank of the i^{th} BRDF.

We used the CUReT database for estimating the dimensionality of each material in the database and found that for 49 of the 61 material, the reconstruction error is less than 10% using 9 basis vectors (Section 3.1).

2.2.3 Varying Viewpoint

Given images taken from different viewpoints, it is trivial to extend the upper bound to $k_v k_\rho k_n$ where k_v is the number of distinct viewpoints. However a much better bound of $k_\rho k_n$ holds true if we know the pixel corresponding to a point x' in the scene in every image. This correspondence can be found, for instance, if the camera parameters of each image and the 3D geometry of the scene are known. Using this correspondence, we can rearrange the pixels in each image so that the x^{th} pixel in every image corresponds to the same scene point. We assume that every scene point is seen by every image (We relax this assumption in Section 4.1). We again consider the rank of the matrix $\mathbf{M}$ obtained by stacking these rearranged images. The argument for Theorem 1 still holds, with R_{il} now defined as:

$$R_{il}(\hat{\mathbf{n}}) = \int_\Omega \rho_l(\omega', \omega_\mathbf{i}) L_i(\omega')(-\hat{\omega'}.\hat{\mathbf{n}})_+ \mathbf{d}\omega'. \quad (11)$$

where $\omega_\mathbf{i}$ is the viewing direction for image i.

2.2.4 Filtered Images

Many real-world images are blurry due to camera shake, or have been otherwise filtered (e.g., software sharpening). We extend the above result to filtered images.

Consider the family of images obtained by convolving image $\mathbf{I}(x,y)$ by an arbitrary $K\times K$ kernel $\mathbf{F}$. Each resulting image can be expressed as $\mathbf{I_{conv}}(x,y) = \sum_{i=1}^{K}\sum_{j=1}^{K} \mathbf{F}(i,j)\mathbf{I}(x-i, y-j)$. Since the space of each of the shifted images $\mathbf{I}(x-i, y-j)$ is at most rank $k_\rho k_n$, it follows that the space of all filtered images of the scene is at most rank $K^2 k_\rho k_n$.

An important special case is the family of radially symmetric filters (e.g., blur, sharpen). These filters can be spanned by a few *basis filters* (The basis filters are simply circles of varying radii.)

Suppose that the family of filters we are concerned with can be spanned by k_f basis filters. Consider convolving each of the $k_\rho k_n$ basis images with each of the k_f basis filters to yield $k_f k_\rho k_n$ images. Any filtered image can then be expressed as a linear combination of these filtered basis images. Hence, the bound reduces to $k_f k_\rho k_n$.

Note that it also takes into account images of different sizes. To see this, one can assume that all different sized images of the scene have been obtained by subsampling an appropriately blurred high resolution image of some fixed resolution. Even though we are applying different blur kernels to these images, the blurred images lie on a linear subspace. Subsampling them to some common lower resolution will also result in a linear subspace.

2.3 Light Spectra

Up until now, we assumed that all measurements are done at a particular wavelength of light, and that the spectrum of light is constant over all images. We now consider the case when the camera sensors and light spectra vary between images. Surprisingly, in general, the appearance space of a simple Lambertian scene with a single infinite plane can have unbounded dimension, even for grayscale images. That is because albedos, which were before treated as fixed scalars for every pixel, are now functions of wavelength, allowing the scene to have arbitrary appearance for different wavelengths. In the general case, using a linear response model, we have that

$$\mathbf{I}_i(x) = \int s_i(\lambda)\mathbf{I}_i(x,\lambda)d\lambda \tag{12}$$

where $s_i(\lambda)$ is the spectral response of the sensor i and $\mathbf{I}_i(x,\lambda)$ is the intensity of light of wavelength λ arriving at the sensor. We begin by analyzing the general case, then discuss results for some common special cases.

2.3.1 The General Case

Consider the matrix $\mathbf{M}$ obtained by stacking images $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_m$ captured by arbitrary sensors. We claim that the rank of $\mathbf{M}$ is bounded by $k_\rho k_n k_\alpha$, where k_α is the number of distinct albedos in the scene.

This result can again be derived from the reflectance map interpretation. We define a reflectance map corresponding to every pair of albedo and BRDF in the scene, with the incidence of normals encoded in the $\mathbf{D}$ matrix. More precisely, ${\mathbf{I}_i}^T = \sum_{h=1}^{k_\alpha}\sum_{l=1}^{k_\rho} \mathbf{R}_{ihl}^T\mathbf{D}_{hl}$ where $\mathbf{R}_{ihl}$ is the image of a sphere with BRDF ρ_l and albedo α_h captured under identical illumination conditions and by the same sensor as $\mathbf{I}_i$, and

$$\mathbf{D}_{hl}(j,k) = \begin{cases} 1 \text{ if } \hat{\mathbf{n}}^k = \hat{\mathbf{m}}_j \text{ and } \alpha^x = \alpha_h \text{ and } \rho^x = \rho_l \\ 0 \text{ otherwise} \end{cases} \tag{13}$$

Again, we can write $\mathbf{M} = \sum_{h=1}^{k_\alpha}\sum_{l=1}^{k_\rho} \mathbf{C}_{hl}\mathbf{D}_{hl}$ by stacking up the $\mathbf{R}_{ihl}$'s,. It follows that $rank(\mathbf{M}) \leq k_\rho k_n k_\alpha$.

More generally, the albedos in a scene (as a function of wavelength) may be spanned by k_α basis albedos. It can be shown in a fashion similar to Section 2.2.1 that the bound of $k_\rho k_n k_\alpha$ extends to this case as well.

2.3.2 Light Sources with Constant Spectra

Belhumeur and Kriegman Belhumeur and Kriegman (1998) showed that images of a Lambertian scene lit by light sources of identical spectra can be spanned by three basis images in the absence of shadows. We do a similar analysis in a more general setting.

Assume that (1) BRDFs do not depend on λ (save for a scale factor, the albedo), (2) all light sources within and across images have the same spectra (but with varying intensity and direction), and (3) all images are captured by identical sensors with spectral response $s(\lambda)$. Under these assumptions, the bound of $k_\rho k_n$ can be seen to hold true.

Under assumption (2), we can write $L_i(\omega',\lambda)$ as $K(\lambda)L_i'(\omega')$ and hence,

$$\mathbf{I}_i(x,\lambda) = K(\lambda)\int_\Omega \alpha^x(\lambda)\rho^x(\omega',\omega)L_i'(\omega')(-\hat{\omega'}.\hat{\mathbf{n}}^x)_+\mathbf{d}\omega' \tag{14}$$

We can write $\mathbf{I}_i(x,\lambda) = K(\lambda)\sum_{j,k} a_{jk}(i)\mathbf{B}_{jk}^I(x,\lambda)$ by invoking the basis image representation for the expression in the integral (Theorem 1), where the number of basis images $\mathbf{B}_{jk}^I$'s is at most $k_\rho k_n$. Note that the coefficients do not depend on λ as wavelength dependent albedos are encoded in the basis images. Substituting into Eq. (12), we get

$$\mathbf{I}_i(x) = \sum_{j,k} a_{jk}(i)\int s(\lambda)K(\lambda)\mathbf{B}_{jk}^I(x,\lambda)d\lambda \tag{15}$$

which implies that $\mathbf{I}_i(x) = \sum_{jk} a_{jk}(i)\mathbf{B'}_{jk}^I(x)$ where the new basis images are obtained by integrating over λ, i.e., $\mathbf{B'}_{jk}^I(x) = \int s(\lambda)K(\lambda)\mathbf{B}_{jk}^I(x,\lambda)d\lambda$. Hence, these images can also be spanned by at most $k_\rho k_n$ basis images.

At first, these assumptions might appear too restrictive. We tested assumption (a) using the CUReT database and found strong support for it (Section 3.3). If albedos and camera spectral responses are unconstrained, the scene may have an unbounded rank. However, if the camera responses are similar, assumption (c) is a reasonable approximation. Other assumptions may be relaxed by extending the result. For instance, consider the case where a scene is lit by k_L light sources, each with its own spectrum that stays constant across all images. This can model outdoor illumination, which is

often approximated as a combination of sunlight and skylight, each with its own spectrum Sunkavalli et al. (2008). Here, the bound can be seen to be $k_\rho k_n k_L$ by writing the illumination in the i^{th} image in the form $\sum_{l=1}^{k_l} K_l(\lambda) L_{li}(\omega', \lambda)$.

Similarly, consider the case when $K(\lambda)$ varies from image to image but lies in a linear subspace of dimension k_s. For illumination in outdoor scenes, the spectra is well approximated by a two or three-dimensional subspace Sunkavalli et al. (2008). The bound can be shown to be $k_\rho k_n k_s$ in this case, by writing $K_i(\lambda) = \sum_{l=1}^{k_s} c_l(i) K_l(\lambda)$.

2.3.3 RGB Images

Images captured by conventional cameras contain three color channels. Consider RGB images $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_m$, where we concatenate the channels together. Assume that each channel is captured by a separate sensor that is identical across all images,

$$\mathbf{I}_i^c(x) = \int s_c(\lambda) \mathbf{I}_i(x, \lambda) d\lambda \tag{16}$$

Consider the matrix $\mathbf{M}^c$ obtained by stacking channel c of all images. Under the assumptions of Section 2.3.2, we know that the rank of this matrix is bounded by $k_\rho k_n$ and it can be written as $\mathbf{M}^c = \mathbf{C}\mathbf{D}^c$ (the coefficients are embedded in the matrix $\mathbf{C}$ while the basis images $\mathbf{B'}_{jk}^I$ are stacked up in $\mathbf{D}$). Because $\mathbf{C}$ does not depend on c (From Eq. (15), we can see that $s_c(\lambda)$ is encoded in the basis images, i.e., $\mathbf{D}^c$), the rank of the matrix $\mathbf{M}$ obtained by concatenating the channels and stacking them is also bounded by $k_\rho k_n$ (we can write $[\mathbf{M}^1|\mathbf{M}^2|\mathbf{M}^3] = \mathbf{C}[\mathbf{D}^1|\mathbf{D}^2|\mathbf{D}^3]$).

In fact, we can go further and show that profiles corresponding to a particular pixel are identical across channels save for a scaling factor, i.e., there exists $k_c(x)$ for each channel such that $\mathbf{P}_x^c / k_c(x)$ is same for all c. This can be seen by substituting for $\mathbf{I}_i(x, \lambda)$ from Eq. (14) in Eq. (16) and writing:

$$\mathbf{P}_x^c(i) = k_c(x) \int_\Omega \rho^x(\omega', \omega) L_i'(\omega') (-\hat{\omega'}.\hat{\mathbf{n}}^x)_+ \mathbf{d}\omega' \tag{17}$$

where

$$k_c(x) = \int s_c(\lambda) K(\lambda) \alpha^x(\lambda) d\lambda \tag{18}$$

2.4 Summary

We started by proving an upper bound of $k_\rho k_n$ in Theorem 1 and then showed that the same bound holds for images taken from different viewpoints and for linear families of BRDFs. In Section 2.2.2, we showed that certain BRDFs allow the bound to be lowered. In Section 2.2.4, it was shown how filtered images can be handled in our theory. Finally, we introduced wavelength in Section 2.3. While in the most general case, the theoretical bound can shown to be $k_\rho k_n k_\alpha$, the bound of $k_\rho k_n$ holds under certain assumptions.

3 Experiments on BRDF Databases

To empirically validate the assumptions introduced in Section 2, we performed experiments on BRDF databases to ascertain the range of materials present in real world images. We looked at two different databases of BRDFs – the MERL BRDF database Matusik et al. (2003) which has 100 different materials, and the CUReT database Dana et al. (1999), which has 61 different materials. CUReT is more representative of real world materials (e.g., paper, grass, cloth, etc.) whereas MERL is restricted to machined and painted spheres. Since our goal is to understand the dimensionality of real-world scenes, we chose to focus on CUReT. Also, the prevalence of specularities in MERL in combination with high dynamic range (HDR) capture, makes approximating through linear models much more difficult, as the highlights alone capture the vast majority of image energy. We found, however, that converting the images to a standard 8-bit-per-channel dynamic range (LDR) (and clamping highlights to 255) yields a reasonable fit for MERL database (See Figure 2). We use the relative RMS error (here, and in all the subsequent results) to gauge the accuracy of the fit, and is measured as RMS value of the error divided by the RMS value of the original data. Here, to convert HDR images to LDR images, we quantized the range of each image so that 90% of the pixels fall in the range $[0, 255]$ and clamped the pixels outside this range to 255. These results imply that even for this dataset, linear models work reasonably well for real world LDR images. CUReT database is used for the results in the following sections.

3.1 Appearance Space of each Material

We first analyze how many basis images are required to span the appearance space of each material. For every material, a hundred 50×50 images of a sphere of that material were rendered. Each image was lit by a distant directional white light whose direction was randomly chosen (uniformly distributed over the front facing hemisphere). The rendered images were reduced to grayscale and SVD was used to compute the basis images corresponding to each material. The green (middle) curve in Figure 3 shows the fall of *relative RMS error* as the number of basis images used to model the appearance are increased. The curve is averaged over all 61 materials for CUReT. From the graph, we can see that the error falls to less than 10% after only 6 basis images.

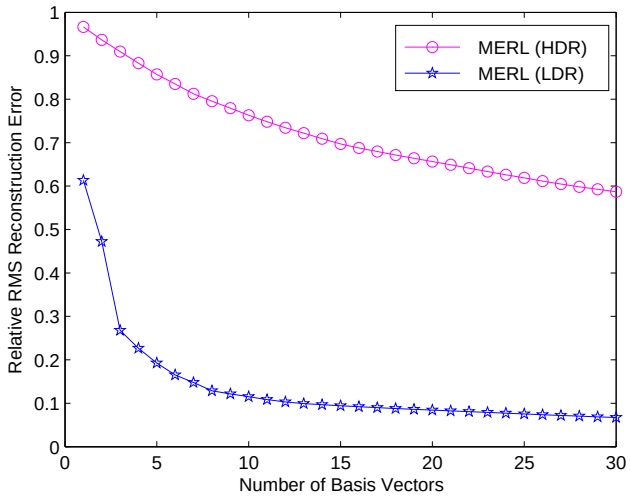


Fig. 2: Reconstruction Error vs Number of basis images for HDR and LDR images rendered using materials in the MERL database. A hundred 50×50 images of a sphere of each material were rendered under different illumination conditions. Basis images were computed using this collection of all $10,000$ images (100 materials, 100 images per material).

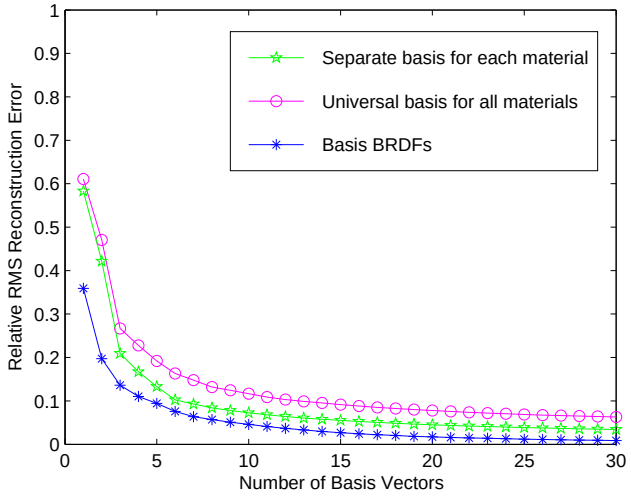


Fig. 3: The fall in relative RMS error vs number of basis vectors for CUReT database.

3.2 Appearance Space of all Materials

In Section 2.2.1, we showed that the upper bound on the dimensionality is reduced in case the BRDFs in a scene are contained in a low dimensional linear subspace. To gauge the range of materials present in the CUReT database, we computed *basis BRDFs* (treating each BRDF as a vector). The BRDFs corresponding to the three color channels were concatenated to form a single large vector. The blue (bottom) curve in Figure 3 shows the reconstruction error vs number of basis BRDFs. Note that here the reconstruction error refers to the error in reconstructing the original BRDFs, and not the rendered images.

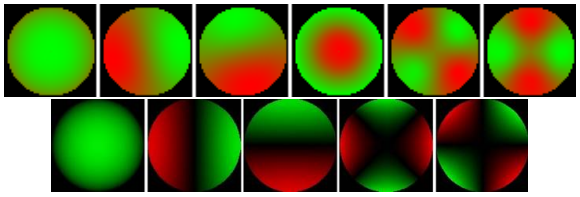


Fig. 5: The top row shows the first six *universal basis* images computed by performing a SVD over all sphere images rendered using materials in the CUReT database. Five of the six are remarkably similar to the 5 basis images used by Ramamoorthi Ramamoorthi (2002) to span the appearance space of a Lambertian sphere (bottom row).

We also run SVD over *all* 6100 *sphere images* rendered in the previous section (61 materials, 100 images per material) to calculate *universal basis images* and measure the reconstruction error versus the number of basis images (The pink (top) curve in Figure 3). The curve is marginally above the curve obtained by calculating a separate basis for each material indicating that the same basis can be shared across a large number of materials.

Reconstruction accuracy of each material is shown in Figure 4 (Materials are sorted according to the reconstruction accuracy using six basis images). The first six *universal basis* images are shown in the top row of Figure 5, five of which are very similar to the basis images used by Ramamoorthi Ramamoorthi (2002) to span the appearance of a Lambertian sphere. These results show that the Lambertian basis augmented with one additional basis gives an average reconstruction accuracy of 85% for the CUReT database. Some example reconstructions are shown in Figure 6. This result is remarkable as it suggests that the non-Lambertian component is relatively insignificant for a wide range of real world materials.

3.3 BRDF across Color Channels

We also test the assumption made in Section 2.3.2, i.e., the BRDF of a material does not depend on wavelength. For each material, we consider the $3 \times N$ matrix obtained by stacking the BRDFs of the 3 channels, where N is the number of samples in each BRDF for each channel. The mean ratio of first singular value to the second singular value was found to be 49.61 with a minimum of 4.55, indicating that for almost every material, the matrix is close to rank 1 and hence the BRDFs can be approximated as being the same across color channels save for a scaling factor.

It is true that the CUReT database contains samples of real world materials and it does not have many specular materials. With specular BRDFs, it is often the case that the Lambertian part of the BRDF is responsible for the the

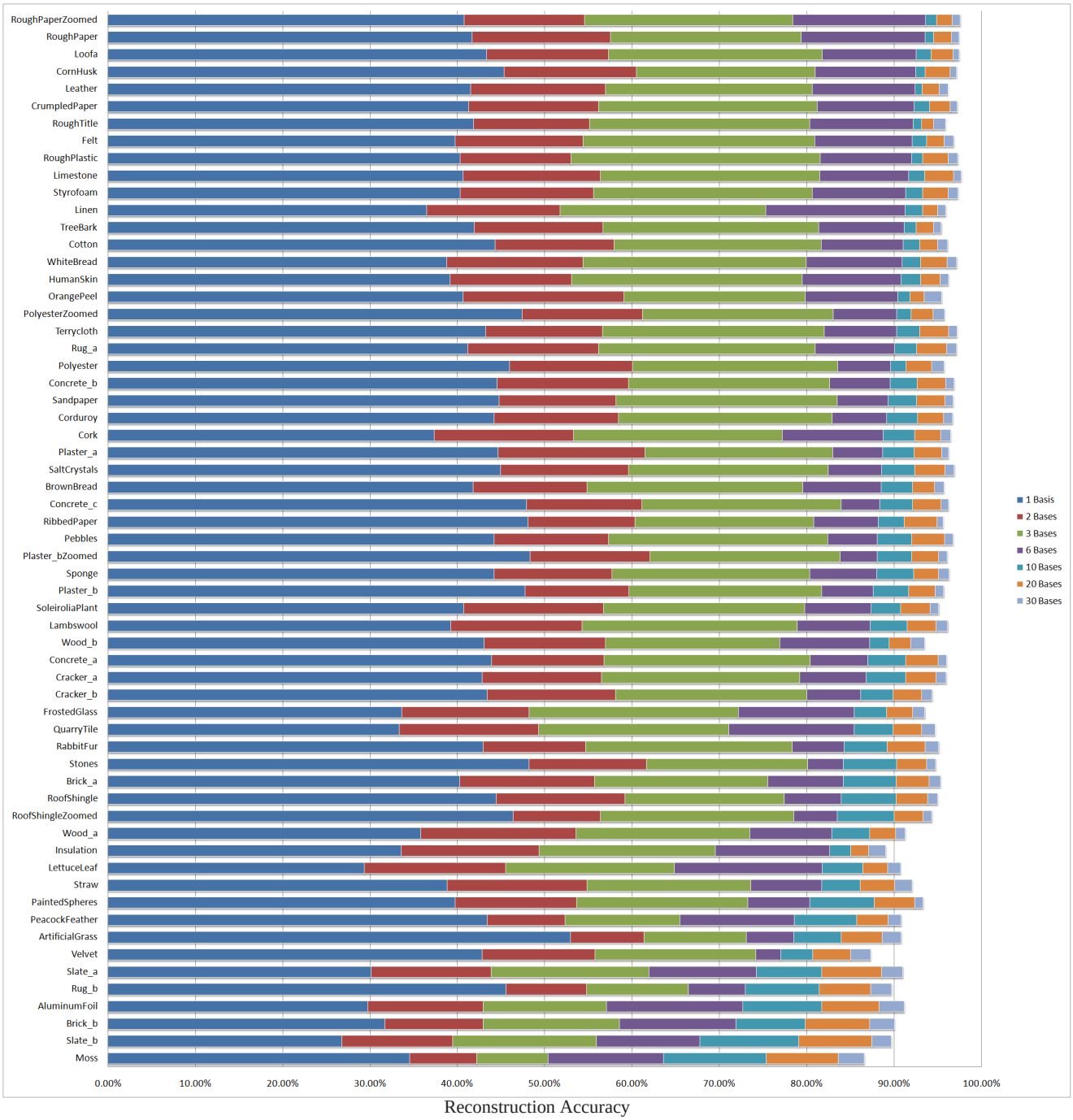


Fig. 4: Reconstruction accuracy achieved for each material in the CUReT database using $1, 2, 3, 6, 10, 20$ and 30 basis images respectively. The basis was computed using the set of all images of all materials (universal basis). The first six basis images can be seen in Figure 5. The materials have been sorted according to the reconstruction accuracy using six basis images. Best seen in color.

color of the object while the specular lobe has a much wider spread across different wavelengths. However, the theory can still explain it by treating the BRDF as a linear combination of a specular and a Lambertian BRDF and they are then allowed to scale independently with wavelength.

4 Linear Modeling of Internet Photo Collections

The results in Section 2 show that linear models well approximate a broad range of real world images. Much of the previous application of linear models has been to images captured in the lab under controlled conditions. Here, we

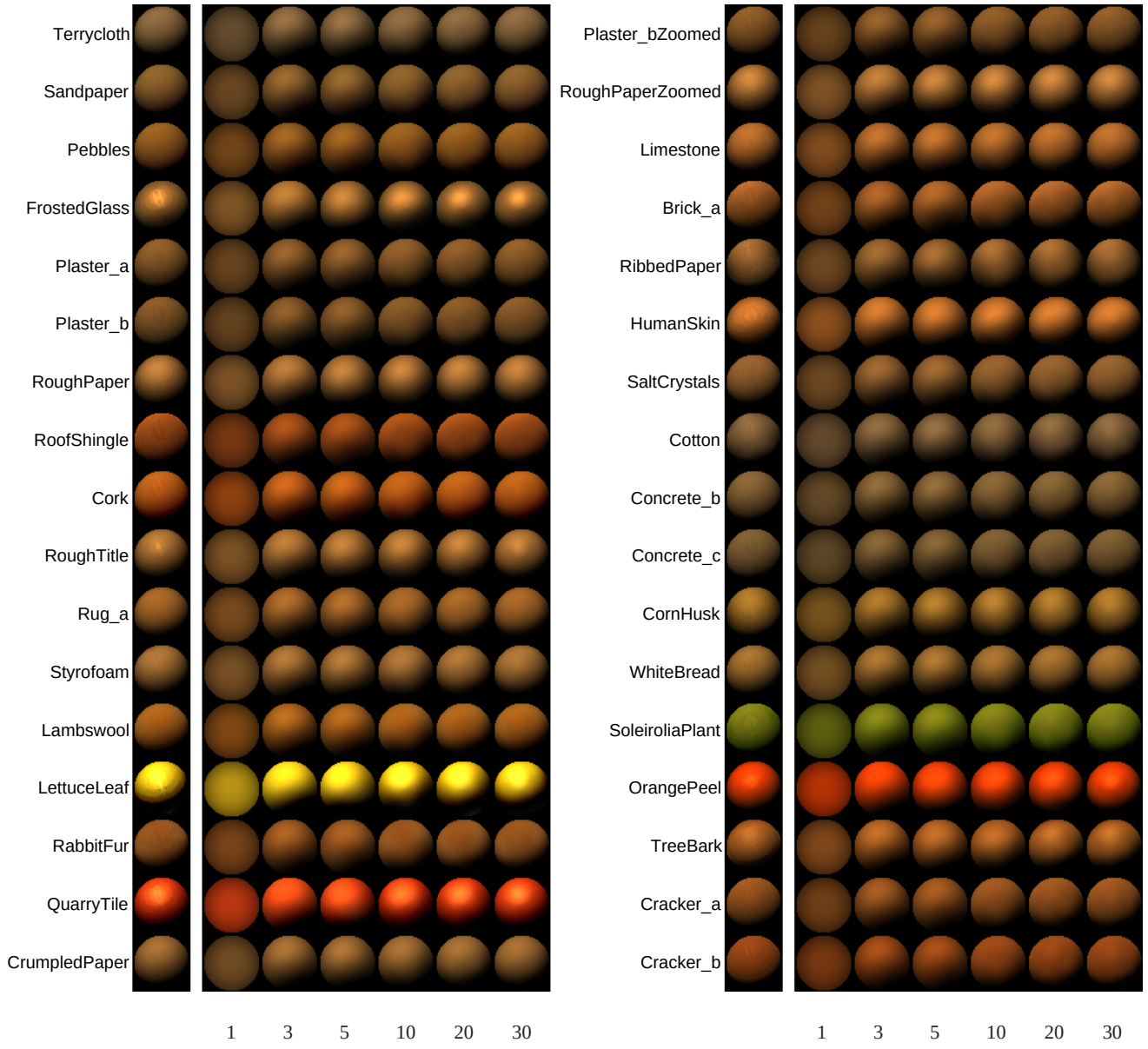


Fig. 6: Some samples from the CURET BRDF database and corresponding reconstructions using $1, 3, 5, 10, 20$ and 30 basis images. A single basis was learnt from images of all the materials. The reconstructions look visually similar (except for smoothing of highlight in some cases) indicating that the same basis may be shared across multiple materials.

apply linear models to the more challenging case of photos of popular locations downloaded from photo sharing websites. The difficulties here stem from the wide variation in the scene appearance. Moreover, the images are captured using many different cameras and viewpoints.

The organization of this section is as follows. In Section 4.1, we give an overview of how we compute the linear model and what representation we use. In Section 4.2, we describe in detail how we process the three channels of color images and what assumptions we make. In Section 4.3, we give quantitative and qualitative evaluation of the results. Finally in Section 4.4, we demonstrate a couple of novel applications of these linear models.

4.1 Basis Computation

Because the input photos are taken from different viewpoints, we first find pixel correspondences. We use the Structure from Motion (SfM) system of Snavely et al. Snavely et al. (2006) to recover the camera parameters. The 3D reconstruction uses the multi view stereo method of Goesele et al. Goesele et al. (2007). The 3D models are simplified using qslim Garland and Heckbert (1997) to a mesh with $\sim$

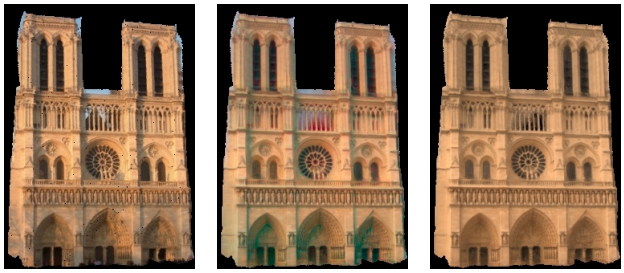


Fig. 7: The image on the left is the original image. The middle image shows the reconstruction obtained using 5 basis images when processing the color channels independently. Notice the false color in regions of shadow or incorrect geometry (for instance, the doors and the columns above the rose window). The figure on the right is the reconstruction obtained by the described approach.

$300,000$ faces. We use a simple representation where we associate a color corresponding to each mesh vertex. Images in this representation (which can be thought of as a texture map), can be treated in a fashion similar to images taken from a fixed viewpoint with mesh vertices assuming the role of pixels. However, a single image covers only a part of the scene, i.e., there is missing data in each texture map. To compute basis vectors with missing data, we use the EM based method of Srebro and Jaakkola Srebro and Jaakkola (2003) to compute SVD. However, the algorithm was found to be sensitive to initialization when the amount of missing data is large. We use the method of Roweis Roweis (1998) which fills the missing data using EM based sensible PCA, to initialize.

Internet photo collections are often dominated by people and other occluders who block the background scene. As our focus is modeling the scene and not the people, we start by manually identifying clean occluder-free images from which we compute a clean basis. We will show later how to handle occlusions in other images using this basis in Section 4.4.2.

4.2 Processing Color Images

We cannot directly apply the ideas in Section 2.3.3 to these color images as the assumption of identical spectra and identical sensors does not hold for Internet photo collections. One option is to process each color channel independently and compute a separate basis for each channel. The selected clean set still has some outliers (e.g. cast shadows) and processing the three channels independently produces *rainbow* artifacts shown in Figure 7 due to inconsistent fits between color channels. Instead, we make some simplifying assumptions that allow us to reconstruct the other channels given the reconstruction of one. Hence, we choose to process only the green channel of these images.

First, under the assumptions made in Section 2.3.2 and 2.3.3, the profile of a pixel x corresponding to, say the red channel, $\mathbf{P_x^{red}}$ can be written as $\mathbf{P_x^{red}} = f^{red}(x)\mathbf{P_x^{green}}$ where

$$f^{red}(x) = \frac{k_{green}(x)}{k_{red}(x)} \tag{19}$$

with $k_{red}(x)$ and $k_{green}(x)$ defined by Eq. 18. However, Internet photos are not captured by identical cameras and the spectrum of light is also different for different photos (for instance, the spectrum of sunlight in the evening is very different from the spectrum at noon). We give a more rigorous argument later, but intuitively, the combined effect of these two factors (camera and light spectra) can be thought of as individual scaling applied to the entire channels of an image, i.e., $\mathbf{P_x^{red}}(i) = g^{red}(i)f^{red}(x)\mathbf{P_x^{green}}(i)$ where $g^{red}(i)$ is the scaling applied to the red channel of the i^{th} image and depends upon the camera sensor and the light spectrum of that particular image. $f^{red}(x)$ can be thought of as a measure of the red color of the pixel (relative to the green channel of the image).

Now, if we can recover g^{red} for all images and f^{red} for all pixels, then we can recover the red channel of any image given the reconstruction of the green channel. We found that the following simple method works well in practice for recovering g^{red} and f^{red}. We assume that there exists a dominant value of $g^{red}(i)$ across images, say g^{red}_{dom}. Note that $g^{red}(i)$ depends on the interaction of the camera sensor and the illumination. So, this assumption translates to saying that a large number of photos are taken by similar cameras under lighting conditions with similar spectra. This is plausible as the spectrum of natural illumination for a given scene remains largely constant for a large part of the day. Also, one can safely assume that a large number of cameras will have similar response curve. Hence, we can recover $f^{red}(x)$ by

$$f^{red}(x)g^{red}_{dom} = median_i\left(\frac{\mathbf{P_x^{red}}(i)}{\mathbf{P_x^{green}}(i)}\right) \tag{20}$$

where the median is taken over i, i.e, along the profile. Once, we have recovered $f^{red}(x)$ (up to the factor g^{red}_{dom}), one can also recover $g^{red}(i)$ by

$$\frac{g^{red}(i)}{g^{red}_{dom}} = median_x\left(\frac{\mathbf{I_i^{red}}(x)}{f^{red}(x)g^{red}_{dom}\mathbf{I_i^{green}}(x)}\right) \tag{21}$$

where now the median is taken over x, i.e., over the pixels of image i. The blue channel can also be processed similarly.

Figure 8 shows the relative RMS error in reconstructions of the three color channels for the Notre Dame and Arc De Triomphe datasets where the basis was only computed for the green channel and the other channels were reconstructed using the method outlined above. The curves corresponding to the red and blue channels are seen to be only marginally

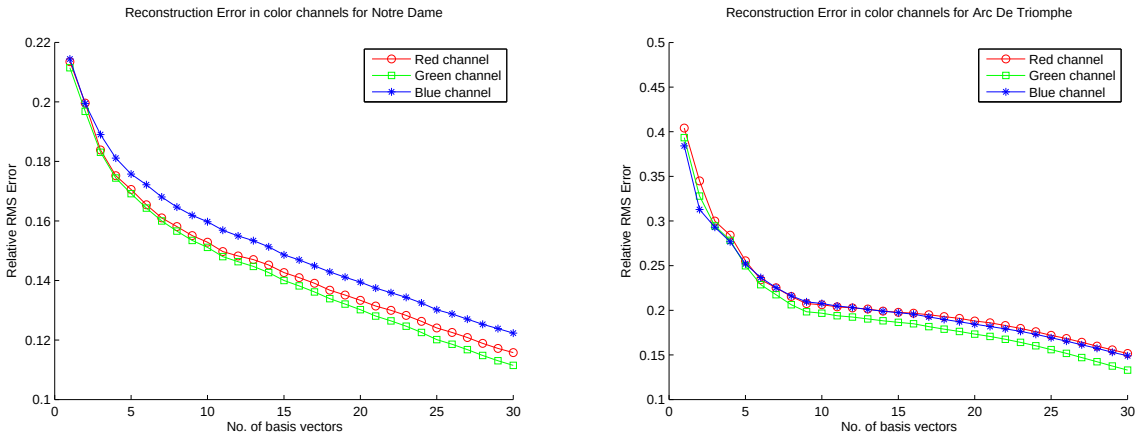


Fig. 8: Relative RMS error in red, blue and green channels for the Notre Dame and Arc De Triomphe datasets. The blue and red curves are only marginally above the green curve indicating that the accuracy achieved in the reconstruction of the red and blue channel by the described approach is similar to the accuracy in the green channel (via SVD).

above the green channel curve, supporting the case of our approximation.

Let us now formally see under what physical conditions the above intuition is correct. We still make the assumption that in a *particular* image, the spectrum of light is the same for all light sources, which allows us to write $L_i(\omega', \lambda)$ as $K^i(\lambda)L'_i(\omega')$ (the illumination can still vary across images). Even though illumination in outdoor scenes is often modeled using two distinct light sources – sunlight and skylight which have different spectra, one can assume that one of them dominates, i.e., the intensity of one is much stronger than the other. E.g., sunlight will dominate on a clear sunny day while skylight will dominate on an overcast day. Repeating the analysis of Section 2.3.2 (where we now have $K^i(\lambda)$ instead of $K(\lambda)$), it can be seen that $\mathbf{P^{c_1}_x(i)}/k_{c_1}(x,i) = \mathbf{P^{c_2}_x(i)}/k_{c_2}(x,i)$, where

$$k_{c_l}(x,i) = \int s_{c_l}(\lambda)K^i(\lambda)\alpha^x(\lambda)d\lambda \tag{22}$$

i.e., k_{c_l} depends on i as well, unlike in Equation (18). We want to be able to write the above integral in the form $f^{c_l}(x)g^{c_l}(i)$. An assumption under which this holds true is when the albedos remain constant over the range of λ over which the support of spectral responses varies and hence can be taken out of the integral. This essentially translates to saying that the support of the spectral response of, say the red sensor, stays in a small neighborhood of a particular wavelength across different cameras.

4.3 Evaluation

We present results on 6 datasets: Notre Dame Cathedral (212 images), Statue of Liberty (318 images), Arc De Triomphe (268 images), Half Dome, Yosemite (95 images) Orvieto Cathedral (228 images), and the Moon (259 images). An image from each of these datasets is shown in the first column of Figure 9. The Moon presents an interesting case due to its retro-reflective BRDF. We are able to register the Moon images using SfM (There exists sufficient parallax for SfM to work Kaula and Baxa (1973)) and then fit a sphere to the 3D points obtained.

All images were gamma corrected assuming $\gamma = 2.2$. As was mentioned in Section 4.1, we use a manually selected clean set of images for computing the basis. Also, as described in Section 4.2, we only need to compute the basis for a single color channel and rest of the color channels can be reconstructed from that. We used the green channel of the images to compute a basis for each dataset. We observed that the reconstructions look reasonably good visually even with three or four basis vectors. With ten basis vectors, some of the finer details such as specularities and self shadowing are also modeled well (we use a basis of size ten to generate results in Section 4.4). There is little visual improvement in the reconstructions after 10 bases though the numerical error continues to fall, but the numerical error stays at 12% even for 30 basis vectors. Figure 11 shows the fall of relative RMS error vs the number of basis images (for the green channel of images). This error can be explained by the fact that even the *clean* set of images have a lot of noise. E.g., Half Dome's view is almost always partially occluded by trees.

Figure 9 shows an example image from these datasets and the corresponding reconstruction for $1, 3, 5, 10$ and 20 basis vectors respectively. The top row (Notre Dame), shows that it becomes possible to model the appearance of night scenes using a larger basis. While such scenes violate our assumptions of distant lighting, as the night-time illumination consists of light sources placed close to the scene, the configuration of light sources is fixed across all night im-

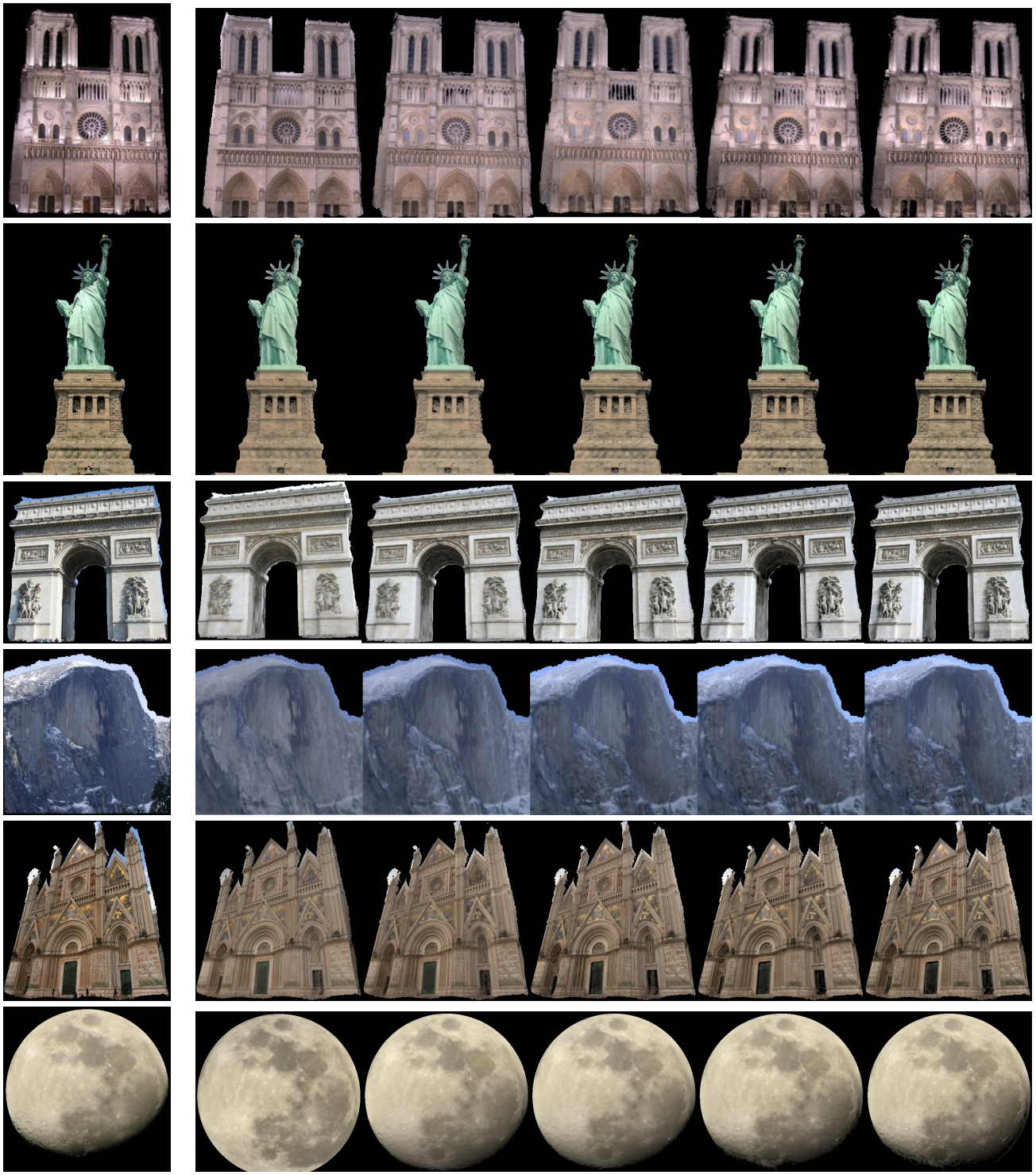


Fig. 9: Internet datasets and reconstructions. The first columns shows an image from the dataset. The following columns show corresponding reconstructions using $1, 3, 5, 10$ and 20 basis images

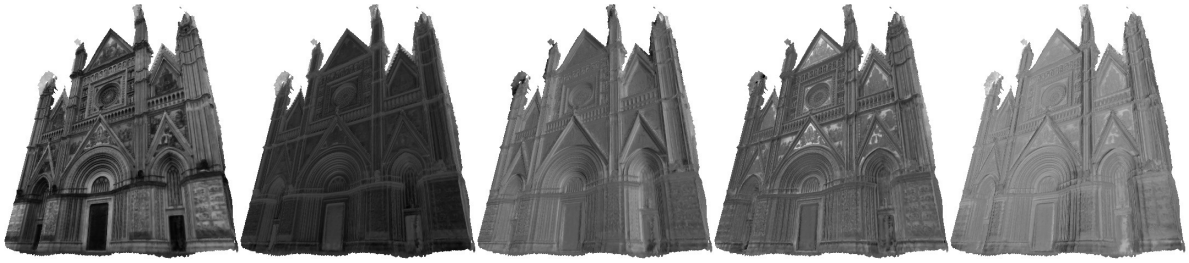


Fig. 10: First 5 basis images for Orvieto. Basis 1 resembles the mean. Bases 2 and 3 model shading, and Bases 4 and 5 specularities.

ages and hence can be modeled by a single additional basis. The results in the second and third rows (Statue of Liberty and Arc De Triomphe) demonstrate that it is possible to approximate cast shadows using a larger basis even though the shadow boundaries appear blurry in the reconstruction. The linear model does not work well for the Half Dome dataset (the fourth row), as there are drastic appearance changes (such as seasonal snow). An image of Orvieto Cathedral, whose facade is highly specular, is approximated in the fifth row. Figure 10 shows the first 5 basis images. While the first basis image simply looks like the mean image, the second and the third model diffuse shading. The fourth and fifth bases seem to model view dependent effects (highlights). Note that the facade of the cathedral is planar and under the assumptions of distant viewer and distant lighting, the specular highlight should ideally cover the whole facade. The presence of specular highlights only on a portion of the facade implies that the viewer is close to the scene which is a violation of our assumption of distant viewer. But as was the case in night scenes, a particular configuration of near-viewpoint and the lighting direction can be modeled by a single additional basis image. For the Moon in the last row, the appearance is modeled well using the first basis, while subsequent bases explain the shadows and the *texture at the terminator* Koenderink and Pont (2002).

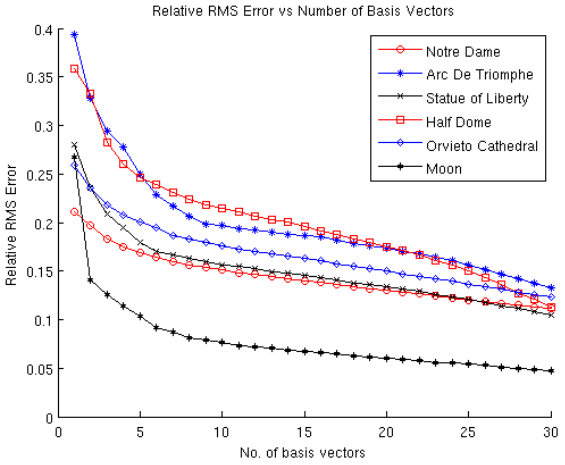


Fig. 11: Relative RMS Error vs number of basis images for different datasets with a clean set of manually selected images. Even though the reconstruction error is around 12% (due to noise, occluders, etc.) for most datasets even with 30 basis images, reconstructions with only 10 basis images look visually similar.

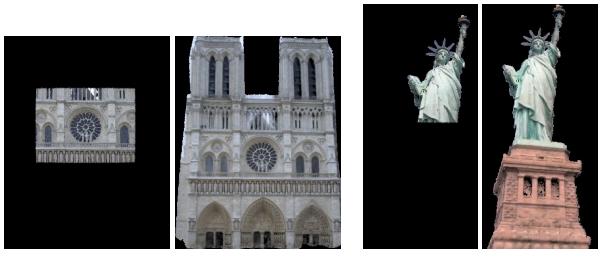


Fig. 12: View Expansion: The left image in each pair shows the input image with limited viewing area. The right image shows the reconstructed image with an expanded field of view.

4.4 Applications

We now show a few novel and interesting applications of linear scene appearance modeling. For all the results shown in the paper, we empirically chose a basis of size 10.

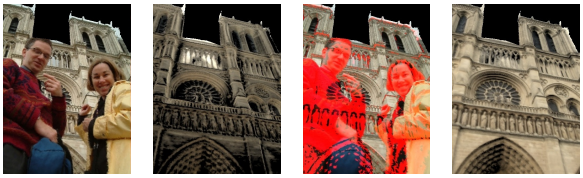
4.4.1 View Expansion

As was mentioned in section 4.1, a single image might cover only a part of the scene. However, since the basis computation method can interpolate missing data, the derived basis images (and hence the reconstructions) cover the entire scene allowing us to hallucinate how the parts of the scene, not visible in the original image, would have appeared under similar illumination conditions. The results of this view expansion approach are shown in Figure 12.

4.4.2 Occluder Removal

Given a basis, we can solve for the coefficients given a new image. We choose a projection approach that is robust to outliers in the image. This allows us to handle occluders. More

Fig. 13: Occluder removal, where the occluder is removed and the scene behind is rendered under the same illumination conditions by robustly solving for basis image coefficients. In each pair, the left image is the input image while the right image shows the corresponding reconstruction with the occluder removed.



(a) (b) (c) (d)

Fig. 14: (a) shows the input image. (b) shows the result of robust projection. It does not work well as the number of outlier pixels is large. (c) shows the precomputed outliers (in red). (d) shows the result of robust projection with precomputed outliers.

precisely, in order to project a new image onto k basis images, we use a RANSAC approach where k pixels are sampled randomly and k coefficients are computed. The number of pixels that lie within a threshold of the original pixel values in the reconstruction obtained using these k coefficients are counted as *inliers*. Finally, the sample with the largest number of inliers is chosen and the estimate of coefficients is refined using all the inliers. Again, we first reconstruct the green channel, and then reconstruct red and blue channels from it (as explained in Section 4.2). The reconstruction constructed from the basis using these robustly computed coefficients will be free of occluders. Some results are shown in Figure 13.

Using the color information, we can also compute outliers. Assuming we have $f^{red}(x)$ computed for the scene (from the set of images used to compute the basis), given a new image, we can compute $g^{red}(i)$ using Equation 21. With this information, one can mark pixels where $|\mathbf{I_i^{red}}(x) - f^{red}(x)g^{red}(i)\mathbf{I_i^{green}}(x)|$ (and a similar expression for the blue channel) is beyond a certain threshold as outliers. Figure 14 shows an example.

5 Conclusion

In this paper, we made the following theoretical contributions.

- A simple factorization framework for analyzing dimensionality of image collections.
- New upper bounds on the number of basis images, allowing for variable illumination direction and spectra, viewpoint, BRDFs, and convolution effects (e.g., blur). As in prior work, we assume distant viewers, distant illumination and ignore cast shadows. The results are motivated by models of shape (particularly for man-made scenes), BRDFs, and light spectra that approximate real world scenes.
- Bounds that take into account the illumination spectrum. Prior low-rank results for Lambertian scenes Shashua (1992); Belhumeur and Kriegman (1998); Basri and Jacobs (2003); Ramamoorthi and Hanrahan (2001); Ramamoorthi (2002) do not apply under variations in light spectrum (even if images are grayscale). Hence prior results are applicable under very controlled conditions.

These results bring the theory much closer to the point where it applies to uncontrolled, real-world scenes. We also verified the assumptions and results empirically using the CUReT BRDF database. Further, we demonstrated the application of low-dimensional models to several large photo collections from the Internet, and showed compelling results for image reconstruction, view expansion, and occluder removal.

While linear models can represent appearance space of scenes under a number of different conditions, we would like to explore how efficient they are in representing the appearance. For example, linear models do not model specularities very efficiently (a large number of basis images is required) and linear models augmented with more complex non linear models may perform better. It would also be interesting to explore other applications of basis texture maps. Linear combination of basis texture maps provides us with a parameterized representation of the scene appearance and one can try to map these parameters onto more interesting physical parameters like the time of the day, cloudiness, etc.

Acknowledgements We wish to thank Ryan Kaminsky for his invaluable help with this project. This work was supported in part by National Science Foundation grant IIS-0811878, the Office of Naval Research, the University of Washington Animation Research Labs, and Microsoft. We are thankful to Flickr users whose photos we used.

References

Basri, R. and D. Jacobs: 2003, 'Lambertian Reflectance and Linear Subspaces'. *PAMI* **25**(2), 218–233.

Belhumeur, P. N. and D. J. Kriegman: 1998, 'What Is The Set Of Images Of An Object Under All Possible Lighting Conditions'. *IJCV* **28**, 270–277.

Dana, K., B. Van-Ginneken, S. Nayar, and J. Koenderink: 1999, 'Reflectance and Texture of Real World Surfaces'. *ACM Trans. on Graphics* **18**(1), 1–34.

Epstein, R., P. Hallinan, and A. Yuille: 1995, '5+/-2 Eigenimages Suffice: An Empirical Investigation of Low-Dimensional Lighting Models'. *IEEE Workshop on Physics-Based Modeling in Computer Vision*.

Garland, M. and P. Heckbert: 1997, 'Surface Simplification Using Quadric Error Metrics'. *Proc. SIGGRAPH* pp. 209–216.

Goesele, M., N. Snavely, B. Curless, H. Hoppe, and S. Seitz: 2007, 'Multi-View Stereo for Community Photo Collections'. *Proc. ICCV* pp. 1–8.

Hager, G. and K. Toyama: 1996, 'X Vision: Combining Image Warping And Geometric Constraints For Fast Visual Tracking'. *Proc. ECCV* pp. 507–517.

Hertzmann, A. and S. M. Seitz: 2003, 'Shape and materials by example: a photometric stereo approach'. *Proc. CVPR* pp. 533–540.

Horn, B. K.: 1986, *Robot Vision*. McGraw-Hill Higher Education.

Kaula, W. M. and P. A. Baxa: 1973, 'The physical librations of the moon, including higher harmonic effects'. *Earth, Moon and Planets* **8**(3), 287–307.

Kirby, M. and L. Sirovich: 1990, 'Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces'. *PAMI* **12**(1), 103–108.

Koenderink, J. J. and S. C. Pont: 2002, 'Texture at the Terminator'. *3D Data Processing Visualization and Transmission, Int. Symp. on* pp. 406–415.

Matusik, W., H. Pfister, M. Br, and L. Mcmillan: 2003, 'A data-driven reflectance model'. *ACM Trans. on Graphics* **22**, 759–769.

Murase, H. and S. Nayar: 1995, 'Visual Learning And Recognition Of 3-D Objects From Appearance'. *IJCV* **14**(1), 5–24.

Oliver, N., B. Rosario, and A. Pentland: 2000, 'A Bayesian Computer Vision System for Modeling Human Interactions'. *PAMI* **22**(8), 831–843.

Pentland, A., B. Moghaddam, and T. Starner: 1994, 'View-based and Modular Eigenspaces for Face Recognition'. *Proc. CVPR* pp. 84–91.

Ramamoorthi, R.: 2002, 'Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object'. *PAMI* **24**(10), 1322–1333.

Ramamoorthi, R. and P. Hanrahan: 2001, 'A signal-processing framework for inverse rendering'. *Proc. SIGGRAPH* pp. 117–128.

Ramamoorthi, R. and P. Hanrahan: 2002, 'Frequency space environment map rendering'. *Proc. SIGGRAPH* pp. 517–526.

Roweis, S.: 1998, 'EM Algorithms for PCA and SPCA'. *Proc. NIPS* **10**, 626–632.

Shashua, A.: 1992, 'Geometry and Photometry in 3D Visual Recognition'. Technical report, MIT AI Lab.

Snavely, N., S. M. Seitz, and R. Szeliski: 2006, 'Photo tourism: exploring photo collections in 3D'. *Proc. SIGGRAPH* pp. 835–846.

Srebro, N. and T. Jaakkola: 2003, 'Weighted low-rank approximations'. *Proc. NIPS* pp. 720–727.

Sunkavalli, K., F. Romeiro, W. Matusik, T. Zickler, and H. Pfister: 2008, 'What do color changes reveal about an outdoor scene?'. *Proc. CVPR* pp. 1–8.

Turk, M. and A. Pentland: 1991, 'Face Recognition Using Eigenfaces'. *Proc. CVPR* pp. 586–591.

Wang, L., S. Kang, R. Szeliski, and H. Shum: 2001, 'Optimal Texture Map Reconstruction from Multiple Views'. *Proc. CVPR* pp. 347–354.