Determining Generative Models of Objects under Varying Illumination: Shape and Albedo from Multiple Images using SVD and Integrability.

A. L. Yuille, D. Snow, R. Epstein and P.N. Belhumeur

December 22, 2002

International Journal on Computer Vision. 35(3), pp 203-222. 1999.

Abstract

We describe a method of learning generative models of objects from a set of images of the object under different, and unknown, illumination. Such a model allows us to approximate the objects' appearance under a range of lighting conditions. This work is closely related to photometric stereo with unknown light sources and, in particular, to the use of Singular Value Decomposition (SVD) to estimate shape and albedo from multiple images up to a linear transformation [15]. Firstly we analyze and extend the SVD approach to this problem. We demonstrate that it applies to objects for which the dominant imaging effects are Lambertian reflectance with a distant light source and a background ambient term. To determine that this is a reasonable approximation we calculate the eigenvectors of the SVD on a set of real objects, under varying lighting conditions, and demonstrate that the first few eigenvectors account for most of the data in agreement with our predictions. We then analyze the linear ambiguities in the SVD approach and demonstrate that previous methods proposed to resolve them [15] are only valid under certain conditions. We discuss alternative possibilities and, in particular, demonstrate that knowledge of the object class is sufficient to resolve this problem. Secondly, we describe the use of surface consistency for putting constraints on the possible solutions. We prove that this constraint reduces the ambiguities to a subspace called the generalized bas relief ambiguity (GBR) which is inherent in the Lambertian reflectance function (and which can be shown to exist even if attached and cast shadows are present [3]). We demonstrate the use of surface consistency to solve for the shape and albedo up to a GBR and describe, and implement, a variety of additional assumptions to resolve the GBR. Thirdly, we demonstrate an iterative algorithm that can detect and remove some attached shadows from the objects thereby increasing the accuracy of the reconstructed shape and albedo.

1 Introduction

In the last few years there has been growing realization that accurate imaging (or lighting) models are needed to design vision systems which can recognize objects in complex lighting situations. Small changes in lighting conditions can cause large changes in appearance, often bigger than those due to viewpoint changes [21]. The amounts of these variations can be appreciated by looking at images of the same object taken under different, but calibrated, lighting conditions, see figure 1. Accurate lighting models are also required for the related reconstruction problem of photometric stereo. In both problems – learning object models and photometric stereo – the input is a set of images of the object, or scene, taken under different lighting conditions. The task is to estimate the shape of the object or scene, its reflection function, and its albedo.

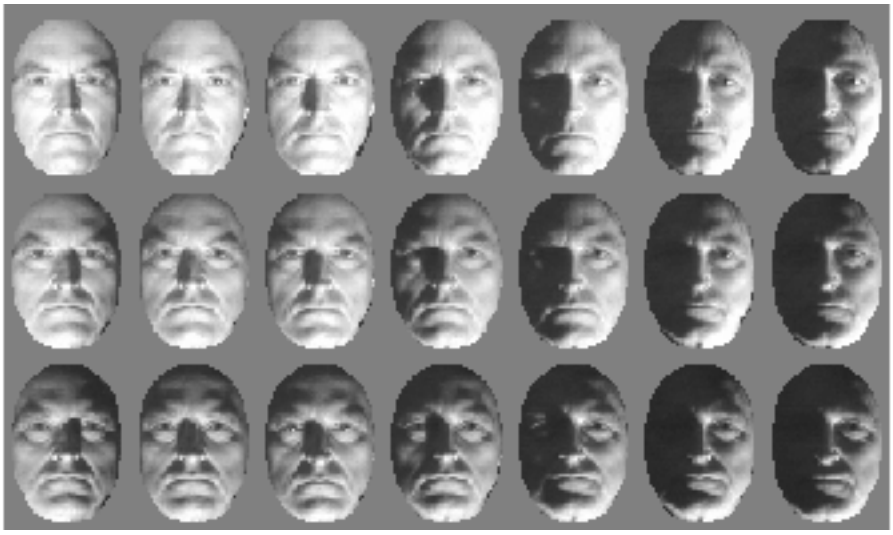


Figure 1: Examples of faces under different lighting conditions.

In recent years, two extreme strategies have been followed to deal with illumination variation. By far the most common is to build representations based on images features [20, 4], such as intensity edges or corners, which are belieeved to be somewhat insensitive, or invariant, to illumination changes. The idea being that object recognition and image understanding can then be performed using these "illumination invariant" representations as input. This approach has two significant drawbacks. First, when illumination variations are large, edges, and indeed all image features, are sensitive to the lighting conditions. Second, and perhaps more important, representations based on edges and corners are sparse and, consequently, throw out a large fraction of useful information.

A different strategy is to use what is often termed an image-based representation or an appearance model. This differs from the feature-based strategy mentioned above in that their representation is, in a least-squared sense, faithful to the original image [22, 24, 26, 14, 12, 32, 40]. (Such models have also been suggested by psychophysicists, see for example [34, 37].) An influential examples is the SLAM system [22] which simultaneously models variations due to pose and illumination by projecting the training images down into low-dimensional subspaces. Systems like these have demonstrated the power of appearance-based methods for certain visual tasks both in ease of implementation and in accuracy. These systems, however, confound the different factors (such as illumination, pose, albedo and geometry) which generate images in a non-transparent way. If the albedo of the viewed object was changed slightly, for example by drawing a red triangle on it, then the whole appearance model would need to be learnt from scratch. Moreover, in principle, appearance based models require that the object is seen under all possible viewing conditions and hence an enormous amount of training data is needed.

We argue that it is preferable to have a more transparent model which explicitly takes into account all the factors which generate the image. This approach requires isolating each factor in turn and modeling it with as simple a representation as possible. This can be thought of as a *generative* model. A big potential advantage of this approach is that from a small number of training images, one can model the object under all possible combinations of lighting and pose. For example, it has been shown [2] that the illumination cone of a convex Lambertian object can be reconstructed from as few as three images *taking into account shadow effects* which are notoriously hard to model. Thus, the representations generalize to novel conditions (requiring less learning data), see [2] for examples of images genereated in such a way. In addition, they can generalize to objects which have not been seen before but which are members of a known object

class [27].

In this paper we are specifically concerned with modeling the appearance of objects under a range of lighting conditions. Our approach makes use of Singular Value Decomposition (SVD) to split the images into components depending on surface properties (geometry and albedo) and lighting conditions. SVD, however, is only able to solve the problem up to an unknown constant linear transform in the shape and albedo of the viewed object(s) and further assumptions must be made to resolve this ambiguity. In addition, the SVD approach starts degrading when there are significant shadows and, as we will describe, robust estimation can be used to remove shadows as outliers.

Our approach has been strongly influenced by the linear lighting models suggested on both theoretical [28] and experimental [13, 5] grounds. A second important source has been the photometric stereo literature [30], [41],[33],[18]. The closest work, which we will refer to throughout the paper, is the application of SVD to this problem by Hayakawa [15] and the recent work by Fan and Wolff [8]. Both consider the situation with spatially varying albedo and unknown light sources. Hayakawa uses SVD in situations without shadows to recover the shape and albedo up to a linear ambiguity but, as we will show in section (2.4), his proposed method for solving for the ambiguity is only valid under limited situations. Fan and Wolff eliminate the albedo by considering ratios of images and demonstrate, on synthetic images (without shadows), that they can estimate properties such as the signs of the surface curvatures. Their results can be re-interpreted in light of the generalized bas relief (GBR) ambiguity, see section (3), which clarifies precisely which properties of the surface can be estimated. Some of the work described in this paper has appeared in conference articles [5, 6, 43, 3] and more details are available in PhD theses [14, 7]. Recent work on bilinear models [10] has applied SVD to a variety of vision problems including estimating shape and lighting. Other recent work includes [29].

In section (2) we describe, extend and analyze the SVD approach. We demonstrate in that it can be generalized to include a background ambient light source which causes the number of non-zero eigenvalues of the SVD to increase from three to four. Empirical analysis of images of real world objects under varying lighting shows that the first three or four eigenvalues typically describe most of the data. We then analyze the linear ambiguities in the SVD approach and demonstrate that additional assumptions proposed to solve them [15] are only valid under certain conditions and discuss alternative possibilities, such as knowledge of object class, to resolve the linear ambiguity. In section (3) we show that surface integrability can be used to reduce the ambiguity to a *generalized bas relief ambiguity* GBR provided that no shadows are present. (This ambiguity has been shown to exist even when cast and attached shadows are present [3] and under perspective projection [19].) In addition, we demonstrate the power of surface integrability for solving up to a GBR and apply other assumptions to remove the GBR. Finally, in section (4), we demonstrate an iterative algorithm which is able to improve the analysis by finding and rejecting shadows.

2 Lambertian Lighting Models and SVD

In this section we describe the SVD approach to lighting [15]. It builds on previous work on photometric stereo [30], [41], [33],[18] and on linear lighting models for objects [13], [28]. See also [31] for the use of linear re-rendering for interactive lighting design.

Suppose we have a set of images generated by a Lambertian model where the lighting conditions vary. We use $\mathbf{x}$ to label positions in the image plane $\mathbf{\Omega}$ and let $|\mathbf{\Omega}|$ be the number of these positions (we assume a finite grid). The light source directions are *unknown* and are labeled by $\mu = 1, ..., M$. This gives us a set of M images:

$$I(\mathbf{x}, \mu) = a(\mathbf{x})\mathbf{n}(\mathbf{x}) \cdot \mathbf{s}(\mu) \equiv \mathbf{b}(\mathbf{x}) \cdot \mathbf{s}(\mu), \tag{1}$$

where $a(\mathbf{x})$ is the albedo of the object, $\mathbf{n}(\mathbf{x})$ is its surface normal, $\mathbf{b}(\mathbf{x}) \equiv a(\mathbf{x})\mathbf{n}(\mathbf{x})$ (observe that $a(\mathbf{x}) = |\mathbf{b}(\mathbf{x})|$ and $\mathbf{n}(\mathbf{x}) = \hat{\mathbf{b}}(\mathbf{x})$), and $\mathbf{s}(\mu)$ is the light source direction. We will typically work with $\mathbf{b}(\mathbf{x})$ instead of $a(\mathbf{x})$ and $\mathbf{n}(\mathbf{x})$. Equation (1), however, has several limitations. It ignores shadows, ambient illumination, and specularities.

We wish to solve equation (1) for the albedo, shape, and light source directions. To do this, we define a least squares cost function:

$$E[\mathbf{b}, \mathbf{s}] = \sum_{\mu,\mathbf{x}} \{I(\mathbf{x}, \mu) - \sum_{i=1}^{3} b_i(\mathbf{x}) s_i(\mu)\}^2, \tag{2}$$

where the subscripts i denote the cartesian components of the vectors (i.e. $\mathbf{b} = (b_1, b_2, b_3)$).

It is possible to minimize this cost function to solve for $\mathbf{b}(\mathbf{x})$ and $\mathbf{s}(\mu)$ up to a constant linear transform using Singular Value Decomposition (SVD) [15],[6].

To see this, observe that the intensities $\{I(\mathbf{x}, \mu)\}$ can be expressed as a $M \times |\mathbf{\Omega}|$ matrix $\mathbf{J}$ where M is the number of images (light sources) and $|\mathbf{\Omega}|$ is the number of points $\mathbf{x}$. Similarly we can express the surface properties $\{b_i(\mathbf{x})\}$ as a $|\mathbf{\Omega}| \times 3$ matrix $\mathbf{B}$ and the light sources $\{s_i(\mu)\}$ as a $3 \times M$ matrix $\mathbf{S}$. SVD implies, see the Appendix, that we can write $\mathbf{J}$ as:

$$\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{3}$$

where $\mathbf{D}$ is a diagonal matrix whose elements are the square roots of the eigenvalues of $\mathbf{J}\mathbf{J}^T$. The columns of $\mathbf{U}$ correspond to the normalized eigenvectors of the matrix $\mathbf{J}^T\mathbf{J}$. The ordering of these columns corresponds to the ordering of the eigenvalues in $\mathbf{D}$. Similarly, the columns of $\mathbf{V}$ correspond to the eigenvectors of $\mathbf{J}\mathbf{J}^T$.

Observe that $(1/M)\mathbf{J}^T\mathbf{J}$ is the $|\mathbf{\Omega}| \times |\mathbf{\Omega}|$ the autocorrelation of the set of input images where $|\mathbf{\Omega}|$ is the image size. (Observe that the mean image is not subtracted – if it was, we would obtain the covariance, or Karhunen-Loeve, matrix used, for example, to compute the principal components of an image dataset). There is a direct relationship between the eigenvectors and eigenvalues of the two matrices $\mathbf{J}^T\mathbf{J}$ and $\mathbf{J}\mathbf{J}^T$. In fact, this relationship can be exploited to calculate the principal components in situations where the matrix $(1/M)\mathbf{J}^T\mathbf{J}$ is too large to calculate its eigenvectors directly (see, for example, [40]). This relationship will be important later in our theoretical analysis, see subsection (2.4).

If our image formation model is correct then there will only be three nonzero eigenvalues of $\mathbf{J}\mathbf{J}^T$ and so $\mathbf{D}$ will have only three nonzero elements (this, of course, has been known in the photometric stereo and vision literature [28], [30], [41]). We do not expect this to be true for our dataset because of shadows, ambient background, specularities, and noise. But SVD is guaranteed to gives us the best least squares solution in any case. Visual study of these solutions [5], [7] suggests that the biggest three eigenvalues typically correspond to the Lambertian component of the reflectance function *provided the number of input images, and the variety of light sources, is large compared with the amount of specularity and shadows.* Intuitively, the positions of the specularities and shadows is highly sensitive to the lighting conditions and so they tend to get averaged out as we take images of the object under different lighting conditions. An object whose reflectance function is mostly Lambertian will only need a few images to ensure that the first three eigenvalues yield the Lambertian components. For a highly specular object, such as the helmet shown in figure (6), many more images are required, see [5], [7] for details. We will therefore assume that the biggest three eigenvalues of $\mathbf{\Sigma}$, and the corresponding columns of $\mathbf{U}$ and $\mathbf{V}$ represent the Lambertian part of the reflectance function of these objects. We define the vectors $\{\mathbf{f}(\mu) : \mu = 1, ..., M\}$ to be the first three columns of $\mathbf{U}$ and the $\{\mathbf{e}(\mathbf{x})\}$ to be the first three columns of $\mathbf{V}$.

This assumption enables us to use SVD to solve for $\mathbf{b}$ and $\mathbf{s}$ up to a linear transformation, see the Appendix. The solution is:

$$\begin{aligned} \mathbf{b}(\mathbf{x}) &= \mathbf{P}_3\mathbf{e}(\mathbf{x}), \ \forall\ \mathbf{x}, \\ \mathbf{s}(\mu) &= \mathbf{Q}_3\mathbf{f}(\mu), \ \forall\ \mu, \end{aligned} \tag{4}$$

where $\mathbf{P}_3$ and $\mathbf{Q}_3$ are 3×3 matrices which are constrained to satisfy $\mathbf{P}_3^T \mathbf{Q}_3 = \mathbf{D}_3$, where $\mathbf{D}_3$ is the 3×3 diagonal matrix containing the square roots of the biggest three eigenvalues of $\mathbf{J}\mathbf{J}^T$. There is an ambiguity $\mathbf{P}_3 \mapsto \mathbf{A}\mathbf{P}_3$, $\mathbf{Q}_3 \mapsto \mathbf{A}^{-1\ \mathbf{T}}\mathbf{Q}_3$ where $\mathbf{A}$ is an arbitrary invertible matrix.

This ambiguity is inherent in the original Lambertian equation (1), where it is clear that the equation is invariant to the transformation $\mathbf{b}(\mathbf{x}) \mapsto \mathbf{A}\mathbf{b}(\mathbf{x})$ and $\mathbf{s}(\mu) \mapsto \mathbf{A}^{-1^T}\mathbf{s}(\mu)$. As we will show later, see section (3), the requirement of surface consistency will reduce this ambiguity to a generalized bas relief transformation (GBR).

2.1 The Linear Ambiguity

For this subsection, we describe other methods besides surface consistency that have been proposed to solve for the linear ambiguity. These methods have their limitations but, as we will show in section (3), they can be used in conjunction with surface consistency to eliminate all ambiguity.

Hayakawa drew attention to the linear ambiguity and proposed certain methods for resolving it [15]. These methods place further restrictions about the application domain such as constant light source magnitudes or partially known reflectance functions. But it is unclear that these assumptions are powerful enough to resolve all the ambiguities.

In particular, Hayakawa proposed assuming that the light source had constant magnitude for six or more images. By equations (4), we have $\mathbf{s}(\mu) \cdot \mathbf{s}(\mu) = \mathbf{f}^T(\mu)\mathbf{Q}_3^T\mathbf{Q}_3\mathbf{f}(\mu)$. Imposing the constraint that this is constant only allows us to solve for $\mathbf{Q}_3^T\mathbf{Q}_3$ and hence there is a further ambiguity $\mathbf{Q}_3 \mapsto \mathbf{R}\mathbf{Q}_3$, where $\mathbf{R}$ is a rotation matrix satisfying $\mathbf{R}^T\mathbf{R} = \mathbf{I}$. Hayakawa makes the additional assumption that this rotation $\mathbf{R}$ is the identity. However, it is unclear why this is a reasonable assumption. It will be proven later, see subsection (2.4) and the Appendix, that this assumption will often be incorrect even in the ideal case where the reflectance function of the object is pure Lambertian with no shadows. Indeed, it is equivalent to making assumptions of symmetry about the dataset which are only appropriate for very special situations.

Another assumption suggested by Hayakawa – that the magnitude of the surface reflectance was known for six or more surface points – would lead to knowledge of $\mathbf{P}_3^T\mathbf{P}_3$ and hence to an equivalent ambiguity $\mathbf{P}_3 \mapsto \mathbf{R}\mathbf{P}_3$.

One method to resolve the linear ambiguity is to use knowledge about the class of the object to determine the linear transformations $\mathbf{P}$ and $\mathbf{Q}$, and hence determine the surface properties and the light sources uniquely [6]. We define a class to be a set of objects with similar albedos and surface normals. This method will only be appropriate if the variations between shape and albedos within the class is small. Below we illustrate this approach for faces.

To do this all we need is the shape and albedo $\{\mathbf{b}_{Pr}(\mathbf{x})\}$ for a prototype object Pr within the object class. This can obtained, for example, by applying SVD with a set of calibrated light source directions, see figure (2). Then when we get the data for a new face image we will estimate its $\mathbf{P}$ and $\mathbf{Q}$ matrices by assuming that it has the same surface properties as the prototype. Thus we estimate $\mathbf{P}$ by minimizing:

$$\sum_{\mathbf{x}} \left|\mathbf{b}_{Pr}(\mathbf{x}) - \mathbf{P}\mathbf{e}(\mathbf{x})\right|^2, \tag{5}$$

where the $\mathbf{e}(\mathbf{x})$ are computed from the new dataset. We are minimizing a quadratic function of $\mathbf{P}$ so the result, $\mathbf{P}^*$, can be obtained by linear algebra.

We now solve for the surface properties using:

$$\mathbf{b}(\mathbf{x}) = \mathbf{P}^*\mathbf{e}(\mathbf{x}), \quad \forall\ \mathbf{x}. \tag{6}$$

The results are shown in figure (3) where we obtain good surface properties even for a face of different shape to the prototype. Observe that the prototype is used merely in conjunction with the dataset to solve for the 3×3 matrix $\mathbf{P}$.

This result has used prior knowledge about object class in the simplest possible form – a prototype model. More sophisticated class knowledge, such as a prior probability distribution for

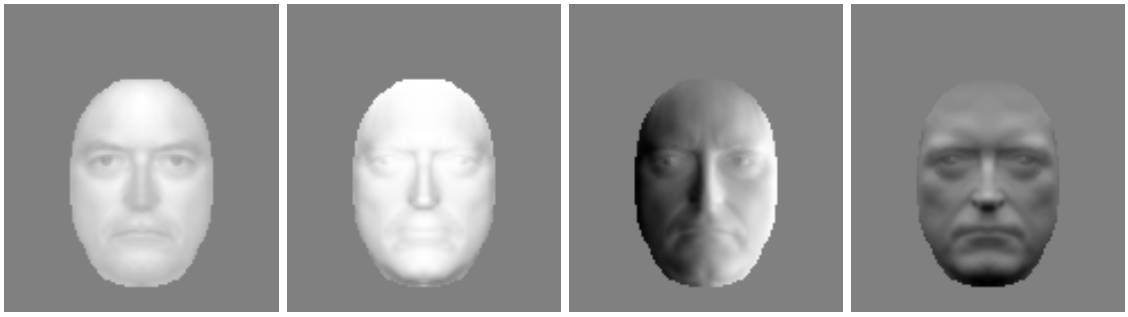


Figure 2: From left to right, the Albedo and z, x, y components of the surface normals calculated directly from SVD using known light source directions to estimate the linear transformations. Observe that the z component of the normal is large except at the sides of the face and near the eyes and nose. The x component of the surface normal is approximately asymmetric corresponding to the symmetry of the face.

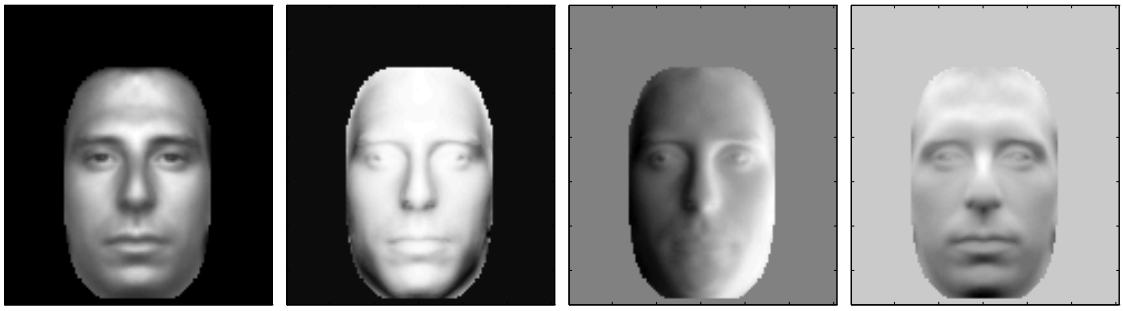


Figure 3: From left to right, the Albedo and z, x, y components of the surface normals calculated directly from SVD using the results shown in the previous figure as a prototype to resolve the linear ambiguity.

shapes and albedos, would lead to improved results. In some situations it is possible to use laser range data to put prior probability distributions on the three-dimensional shapes of object classes. For certain objects, like faces [1] this has been reported to generate accurate three-dimensional models from a single input image (i.e. not requiring an SVD stage).

2.2 Generalization to include ambient illumination

In this subsection we generalize the SVD approach to include a background ambient illumination term. This will mean that we can obtain the albedo, shape, and ambient term. It is standard to assume that ambient lighting corresponds to illuminating the object by a large number of point sources from a range of directions (e.g. illumination on a cloudy day). But another, less common, form of ambient lighting is when the object is illuminated by projecting a pattern onto it (e.g. projecting a slide onto a surface). We will demonstrate that our approach works for both cases.

This means we modify the equations to be:

$$I(\mathbf{x}, \mu) = \mathbf{b}(\mathbf{x}) \cdot \mathbf{s}(\mu) + \tilde{a}(\mathbf{x}), \tag{7}$$

where $\tilde{a}(\mathbf{x})$ is the ambient illumination which we assume to be independent of μ. (I.e. we assume that the ambient illumination stays constant while the Lambertian light sources vary).

We define a cost function for estimating $\mathbf{b}$, $\mathbf{s}$, and $\tilde{a}$:

$$E[\mathbf{b}, \mathbf{s}, \tilde{a}] = \sum_{\mathbf{x},\mu} \{I(\mathbf{x}, \mu) - (\mathbf{b}(\mathbf{x}) \cdot \mathbf{s}(\mu) + \tilde{a}(\mathbf{x}))\}^2. \tag{8}$$

It is straightforward to generalize our previous approach and apply SVD to estimate $\mathbf{b}$, $\mathbf{s}$, and $\tilde{a}$. The important difference is that we now rely on the first four eigenvalues of $\mathbf{J}\mathbf{J}^T$, see the Appendix. The additional eigenvalue is needed because of the ambient lighting term $\tilde{a}(x)$ (before the images could be expressed as the outer product of three-dimensional vectors $\mathbf{b}$ and $\mathbf{s}$ but now they are the outer product of four-dimensional vectors $(\mathbf{b}, \tilde{a})$ and $(\mathbf{s}, 1)$.). We can generalize equation (4) to:

$$\begin{aligned} \mathbf{b}(\mathbf{x}) &= \mathbf{P}_3\mathbf{e}(\mathbf{x}) + \mathbf{p}_4 e_4(\mathbf{x}), \\ \tilde{a}(\mathbf{x}) &= \mathbf{v} \cdot \mathbf{e}(\mathbf{x}) + v_4 e_4(\mathbf{x}), \\ \mathbf{s}(\mu) &= \mathbf{Q}_3\mathbf{f}(\mu) + \mathbf{q}_4 f_4(\mu), \\ 1 &= \mathbf{w} \cdot \mathbf{f}(\mu) + w_4 f_4(\mu). \end{aligned} \tag{9}$$

As before, there is a linear ambiguity. The difference is that it is now a four by four linear transformation instead of a three by three. It will turn out, however, that the surface consistency constraint in combination with assuming constant magnitude of the light source will be sufficient to remove this ambiguity, see section (3). Note that the last equation of (9) already gives conditions on $\mathbf{w}$ and w_4.

2.3 Empirical Evidence for Approximate Lambertian Models

As described in previous sections, the SVD approach applies for Lambertian objects with a distant light source and an ambient term. However, real objects have significant specular lobes and spikes [23] and there are other complicating effects such as shadows, mutual inter-reflections, and occlusions [9]. In addition, how distant do the light sources need to be from the object? Hence, the issue of whether lighting variations can be approximated by Lambertian models requires empirical investigation. Can we assume, for example, that Lambertian is a good approximation (in the sense of least squares) for common objects such as faces and helmets? If this is true then we may be able to ignore non-Lambertian effects or, if necessary, treat them as residuals and remove them by an iterative algorithm, see section (4).

The most direct way to investigate this is by computing the eigenfunctions of $\mathbf{J}^T\mathbf{J}$ for each of a set of objects under a variety of lighting conditions. If only the first three eigenvalues are large then we have evidence for a Lambertian model. If the fourth eigenvalue is also big then this would require, at the least, an ambient background term. As discussed in the previous section, SVD gives an efficient way to compute the eigenvectors and eigenvalues of $\mathbf{J}^T\mathbf{J}$.

Such an empirical investigation was done for faces by Hallinan [13] and motivated his linear basis function model of lighting [14]. The work was extended by Epstein *al* [5] and applied to a range of objects. The conclusions of both studies were that only the first few eigenvalues were significant. Moreover, the light sources used in these studies were standard light bulbs at most six feet from the viewer [14]. Therefore the requirement that the light source be "distant" is very weak and will be satisfied in most environments. Similar results were also reported by Belhumeur and Kriegman in their work on illumination cones [2].

More precisely, it was concluded that for many objects: (a) 5 ± 2 eigenvectors will suffice to model the Lambertian and specular lobes, (b) specular spikes, small shadows and occluders can be treated as residuals and eliminated by projecting the original image onto the low dimensional eigenvector model, and (c) the sampling of lighting directions required in the training set increases with both the specularity and the complexity of the surface geometry.

The following methodology was used to construct the lighting models for each object. Seventy images were taken of each object under different lighting conditions. The dominant lighting in each image was from a small area source (floodlight) at a distance of about six feet. (See [14] for details.) This light could be moved along the surface of a sphere whose center was the object and whose North Pole was directly above the object. The light was moved along the sphere's lines of latitude and longitude in 15 degree intervals such that the lighting direction varied over the entire right front side of the object. These images formed the *dense* training set. The *sparse* training set was a subset of 20 images from the dense set. The lighting for the images of the sparse set varied in 30 degree intervals.

Two eigenimage models were constructed for each object by calculating the eigenvalues and eigenvectors of the autocorrelation matrix on the sparse and the dense data sets. (Tests were also run with the mean images subtracted but little change was noticed). Additional images of the objects were taken under ambient lighting conditions. These images were used to evaluate the models' ability to reconstruct novel images which were not in the training sets. In addition, when the sparse data alone was used, the remaining images of the dense set were used to test the sparse model.

The eigenvectors are of form shown in figure (4). The first three eigenvectors appear to correspond to the face illuminated from three orthogonal lighting conditions. The first three eigenvectors of many objects shared this property so we informally named it the *orthogonal lighting conjecture*. Mathematical analysis, see subsection (2.4) and the Appendix, suggests that this conjecture depends on the symmetry of the object being viewed and will not hold in general.

Figure 4: The eigenvectors calculated from the sparse set for the human face. Note that the images were only lit from the right so the eigenvectors are not perfectly symmetric. Observe also that the first three eigenvectors appear to be images of the face illuminated from three orthogonal lighting conditions in agreement with the orthogonal lighting conjecture.

The quality of the eigenimage models was measured by using a goodness of fit criterion. This is a measure of the difference between the image and its projection onto the space of eigenimages. More precisely, for an object o we construct the eigenvectors $\mathbf{e}_i^o(\mathbf{x})$ of the autocorrelation matrix indexed by $i = 1, ..., n$. For a specific image $\mathbf{I}^o(\mathbf{x})$ of the object we construct its projection $\mathbf{I}_p^o(\mathbf{x})$ to be:

$$\mathbf{I}_p^o(\mathbf{x}) = \sum_{i=1}^{n} \{ \sum_{\mathbf{z} \in \mathbf{\Omega}} \mathbf{I}^o(\mathbf{z}) \mathbf{e}_i^o(\mathbf{z}) \} \mathbf{e}_i^o(\mathbf{x}). \tag{10}$$

We then measured the quality of the projection of the *specific* image by the goodness of fit function:

$$\epsilon(o, \mathbf{I}^o) = 1 - \frac{\left| ||\mathbf{I}^o(\mathbf{x}) - \mathbf{I}_p^o(\mathbf{x})|| \right|^2}{||\mathbf{I}^o(\mathbf{x})||^2}, \tag{11}$$

where the norm $||\mathbf{I}(\mathbf{x})||^2 = \sum_{\mathbf{x} \in \mathbf{\Omega}} \{\mathbf{I}(\mathbf{x})\}^2$. Observe that the goodness of fit ranges from one (if the image and its projection are identical) to zero (if the image is infinitely different from its projection). We emphasize that this goodness of fit criterion does not necessarily account for

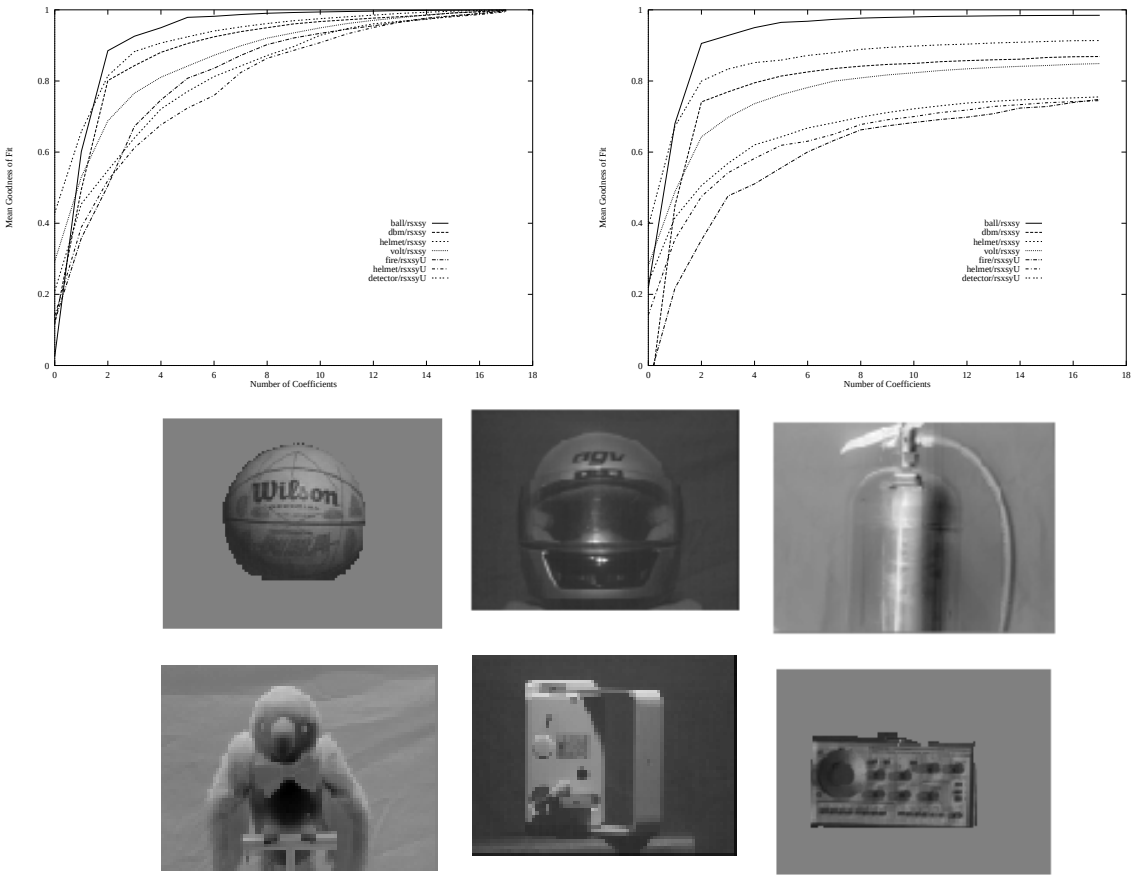


Figure 5: Plots of the mean goodness of fit vs the number of eigenvectors for training (top left) and test (top right) sets of several different objects. All models were constructed using a sparse data set. Examples images of the ball, helmet, fire-extinguisher,parrot,spectrometer, and voltmeter are shown below,

human perception – small specularities would give little contribution to the goodness of fit but might be visually salient.

In figure (5) we give graphs showing the mean goodness of fits for a variety of objects where the mean is with respect to either training or test data sets. The bases used were constructed from the sparse samples only. These graphs show clearly that, even for highly specular objects, the mean goodness rises very rapidly with the first three or four eigenvalues and improvements in performance begins to taper off at the fifth eigenvectors.

Table (1) shows the cumulative variances for different objects. Observe that even for highly specular objects such as the helmet, see figure (6), the variance only rises a little slower if the dense data set is used instead of the sparse set. The quality of the principal eigenvectors, however, does improve when the dense set is used. Not surprisingly (see [5, 7] for more details) the principal eigenvectors for the sparse dataset are influenced heavily by the precise location of the specularities in the sparse training set. When the dense dataset is used then the specularities average out and they appear not to influence the principal eigenvectors. This means that when images are projected the specularities are removed and appear nicely as residuals, see figure (6).

Eigen-vector	ball	parrot	phone	face	helmet	helmet	helmet	fire ext	function generator	infrared detector
	sparse (right)	dense (both)	dense (right)	sparse (right)	sparse (right)	dense (right)	dense (both)	dense (right)	dense (right)	dense (right)
#1	0.482	0.428	0.679	0.537	0.320	0.340	0.388	0.536	0.806	0.624
#2	0.844	0.697	0.832	0.752	0.569	0.540	0.474	0.687	0.879	0.805
#3	0.944	0.763	0.882	0.902	0.651	0.628	0.581	0.765	0.922	0.885
#4	0.965	0.815	0.920	0.921	0.728	0.746	0.655	0.816	0.936	0.915
#5	0.979	0.847	0.941	0.935	0.798	0.772	0.722	0.852	0.948	0.927
#6	0.989	0.872	0.952	0.945	0.845	0.794	0.750	0.882	0.956	0.939
#7	0.991	0.885	0.963	0.953	0.881	0.816	0.795	0.901	0.962	0.948
#8	0.993	0.897	0.968	0.958	0.905	0.833	0.811	0.913	0.968	0.954
#9	0.995	0.907	0.972	0.963	0.924	0.848	0.824	0.925	0.972	0.960
#10	0.996	0.917	0.975	0.966	0.943	0.861	0.837	0.933	0.975	0.965

Table 1: The variance (cumulative) accounted for by each eigenvector for several different objects, both for sparse and dense training sets.

2.4 Mathematical Analysis of the SVD approach

What can mathematical analysis tell us about the SVD method? As we will show, it is possible to theoretically analyze the eigenvectors that result from SVD provided we assume that the object has a pure Lambertian reflectance function with no shadows. In other words, we assume that the data is indeed generated by the model with which we analyze it. Of course, this is an ideal world assumption and so the theoretical results will start to degrade as shadows become significant.

We specifically investigate two issues arising earlier this section. The first concerns Hayakawa's assumption that a specific rotation ambiguity in SVD can be resolved by setting the rotation to be the identity. The second involves the orthogonal lighting conjecture – that the first three eigenvectors point along the axes of the cartesian coordinate system. We will show that these claims are closely related and depend on the symmetry of the dataset.

Let us assume that the data is generated by a true Lambertian surface. In other words, that the input image set $\{I(\mathbf{x}, \mu)\}$ can be expressed as:

$$I(\mathbf{x}, \mu) = \sum_{i=1}^{3} b_i(\mathbf{x}) s_i(\mu), \tag{12}$$

where $\{b_i(\mathbf{x})\}$ and $\{s_i(\mu)\}$ are the *true* albedo, shape and lighting.

We can reformulate equation (4) in coordinate terms as:

$$\begin{aligned} b_i(\mathbf{x}) &= \sum_{j=1}^{3} P_{ij} e_j(\mathbf{x}) \quad \forall\, i, \mathbf{x}, \\ s_i(\mu) &= \sum_{j=1}^{3} Q_{ij} f_j(\mu), \quad \forall\, i, \mu, \end{aligned} \tag{13}$$

where the e and f obey the eigenvectors equations:

$$\begin{aligned} \sum_{\mu'} \{ \sum_{\mathbf{x}} I(\mathbf{x}, \mu) I(\mathbf{x}, \mu') \} f_i(\mu') &= \lambda_i f_i(\mu), \quad \forall\, i, \mu \\ \sum_{\mathbf{x}'} \{ \sum_{\mathbf{x}} I(\mathbf{x}, \mu) I(\mathbf{x}', \mu) \} e_i(\mathbf{x}') &= \lambda_i e_i(\mathbf{x}), \quad \forall\, i, \mathbf{x}, \end{aligned} \tag{14}$$

and the matrices $\mathbf{P}$ and $\mathbf{Q}$ are constrained to satisfy:

$$\mathbf{P}^T \mathbf{Q} = \mathbf{D}, \tag{15}$$

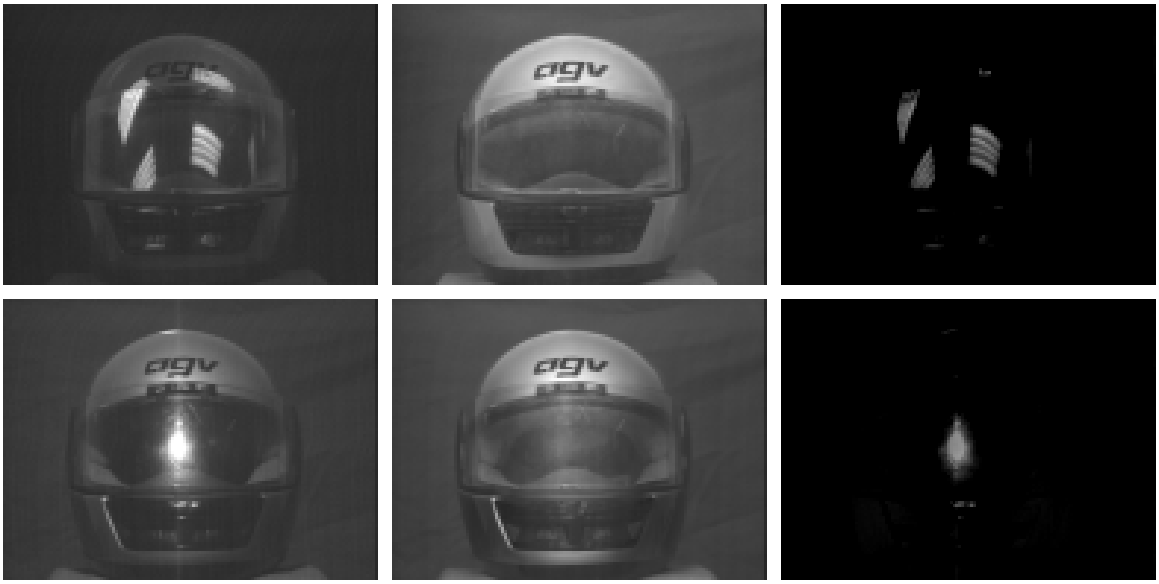


Figure 6: Top left is helmet under ambient lighting conditions. Top center is reconstruction from projection onto the first five eigenvectors. Note the specularities have been eliminated. Top right is the difference between original image and reconstruction. (Differences less than zero have been set to zero because the specularities are, by definition, positive.) The specularities are easily isolated. Bottom row shows the same effect for an image from the dense data set.

where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are $\lambda_1^{1/2}, \lambda_2^{1/2}, \lambda_3^{1/2}$.

We can now state the following theorems:

Theorem 1. *If the light sources in the dataset are such that* $\sum_\mu s_i(\mu)s_j(\mu) = \delta_{ij}$*, where* δ_{ij} *is the Kronecker delta, then Hayakawa's rotation matrix should be set equal to the identity if, and only if, the albedos and shapes in the data set satisfy* $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x}) = 0,\ i \neq j$.

Theorem 2. *The first three eigenvectors* $e_i(\mathbf{x}) : i = 1, 2, 3$ *point along the axes of the cartesian coordinate system if, and only if, both* $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x}) = 0,\ i \neq j$ *and* $\sum_\mu s_i(\mu)s_j(\mu) = 0,\ i \neq j$.

Both theorems show that interesting results occur if the input data is symmetric. More precisely, it corresponds to assuming that the off-diagonal terms of $\sum_p b_i(\mathbf{x})b_j(\mathbf{x})$ and $\sum_\mu s_i(\mu)s_j(\mu)$ vanish. This will be true if, for example, the light source directions sample the viewing hemisphere evenly and the object is an ellipsoid viewed head-on and with constant albedo. The off-diagonal terms will also be expected to vanish if the i and j components of the dataset are statistically independent (for then, by ergodicity, $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x}) \mapsto < b_i b_j >$ and $\sum_\mu s_i(\mu)s_j(\mu) \mapsto < s_i s_j >$). However, there will be many datasets for which these assumptions will be violated. Thus Hayakawa's assumptions and the orthogonal lighting conjecture will typically not be true.

The proofs of the theorems are long and involved. For reasons of space, we only give the broad outlines of the proofs here and refer the reader to the Appendix for more details.

Proof of Theorem 1. $\sum_\mu s_i(\mu)s_j(\mu) = \delta_{ij}$ *implies that* $\sum_\mu I(\mathbf{x}, \mu)I(\mathbf{x}', \mu) = \sum_i b_i(\mathbf{x})b_i(\mathbf{x}')$. *This implies that* $\mathbf{P}^T\mathbf{P} = \mathbf{D}^2$*, where* $\mathbf{D}$ *was defined above. Hayakawa's assumption involves setting* $\mathbf{P} = \mathbf{D}$*, but there are many other possible solutions of form* $\mathbf{P} = \mathbf{R}\mathbf{P}$ *where* $\mathbf{R}$ *is any rotation matrix. Observe, that if* $\mathbf{R} = \mathbf{I}$ *then* $b_i(\mathbf{x}) = \lambda_i^{1/2} e_i(\mathbf{x}),\ \forall\ i, \mathbf{x}$ *and so* $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x}) = 0$ *for* $i \neq j$*. Conversely, suppose that* $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x}) = \mu_i\delta_{ij},\ \forall\ i, j$ *for some values* $\{\mu_i\}$*. Then this implies that* $\mathbf{P}\mathbf{P}^T = \mathbf{D_1}$*, where* $\mathbf{D_1}$ *is diagonal with diagonal elements* $\{\mu_i\}$*. This is inconsistent with* $\mathbf{P}^T\mathbf{P} = \mathbf{D}^2$*, unless* $\mathbf{P}$ *is diagonal.*

Observe, in this proof, that there is a close connection between Hayakawa's assumption and

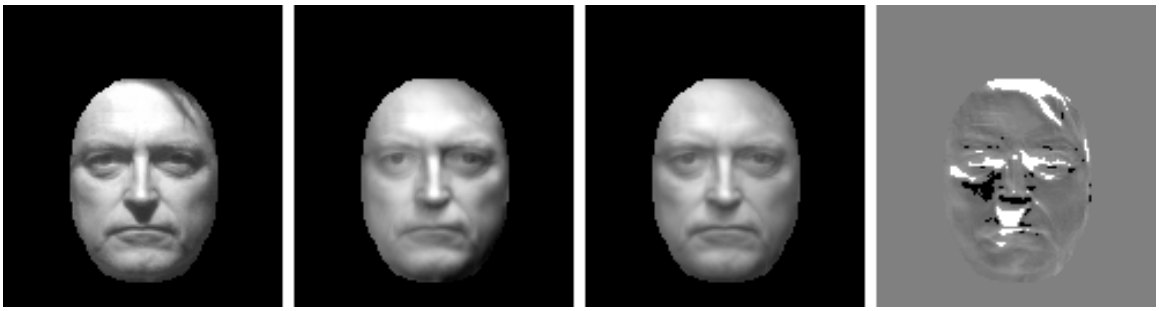


Figure 7: On the left is an image of a face from the dense set. The second image is the reconstruction from the projection onto the first 3 eigenvectors calculated from the sparse set. The third image is the reconstruction from its projection onto the first five eigenvectors calculated from the sparse set. The rightmost image shows the difference between the first and third images, where differences of greater than 30 greyscale levels have been highlighted in black (if the original image is darker than the reconstruction from the projection), or white (if the original image is brighter than the reconstruction from the projection).

the orthogonal lighting conjecture. In particular, there is a clear relation between the matrices $\sum_\mu s_i(\mu)s_j(\mu)$ and $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x})$ being diagonal and the relationship $b_i(\mathbf{x}) \propto e_i(\mathbf{x}),\ \forall\ i, \mathbf{x}$ and $s_i(\mu) \propto f_i(\mu),\ \forall i, \mu$. This motivated the second theorem.

Proof of Theorem 2. If $b_i(\mathbf{x}) = \mu_i e_i(\mathbf{x}),\ \forall\ i, \mathbf{x}$ *and* $s_i(\mu) = \lambda_i^{1/2}/\mu_i f_i(\mu),\ \forall i, \mu$ *then it is clear that* $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x})$ *and* $\sum_\mu s_i(\mu)s_j(\mu)$ *are diagonal. Conversely, suppose that* $\sum_{\mathbf{x}} b_i(\mathbf{x})b_j(\mathbf{x}) = \delta_{ij}\mu_i^2,\ \forall i, j$ *and* $\sum_\mu s_i(\mu)s_j(\mu) = \delta_{ij}/\mu_i^2,\ \forall i, j$*, for some* $\{\mu_i\}$*. Then it follows that* $\mathbf{P}^T\mathbf{D_2}\mathbf{P} = \mathbf{D}^2$ *and* $\mathbf{Q}^T\mathbf{D}_3\mathbf{Q} = \mathbf{D}^2$*, where* $\mathbf{D}_2$ *and* $\mathbf{D}_3$ *are diagonal matrices with diagonal elements* $\{\lambda_i/\mu_i^2\}$ *and* $\{\mu_i^2\}$ *respectively. The only solutions to these equations occur if* $\mathbf{Q}$ *and* $\mathbf{P}$ *are diagonal.*

3 Surface Integrability and the Generalized Bas Relief Ambiguity

From subsections (2.1,2.4) we see that existing assumptions, with the exception of class specific knowledge, are not sufficient to solve for the linear ambiguity.

There is, however, another constraint which can always be applied. This is the *surface consistency conditions* or *integrability constraints*, see various chapters such as Frankot and Chellappa in [16]. These integrability conditions are used to ensure that the set of surface normals forms a consistent surface. As we first showed in an earlier version of this work [6] these constraints are powerful enough, theoretically, to determine the surface and albedo up to a generalized bas relief transformation (GBR) (see [3] for a full treatment of the GBR). In this section we will introduce GBR and demonstrate a method, which we first presented in [43], for using it to solve for the shape and albedo up to a GBR. Additional assumptions can then be used to determine the full solution.

The integrability constraints are usually expressed in terms of the surface normals but, as shown in [6], they can be generalized to apply to the $\mathbf{b}(\mathbf{x})$ vectors. The constraints can be expressed in differential form:

$$\frac{\partial}{\partial x}\left(\frac{b_2(\mathbf{x})}{b_3(\mathbf{x})}\right) = \frac{\partial}{\partial y}\left(\frac{b_1(\mathbf{x})}{b_3(\mathbf{x})}\right) \tag{16}$$

Expanding this out we get

$$b_3\frac{\partial b_2}{\partial x} - b_2\frac{\partial b_3}{\partial x} = b_3\frac{\partial b_1}{\partial y} - b_1\frac{\partial b_3}{\partial y} \tag{17}$$

It is straightforward to check that these equations are invariant to the three-dimensional linear transformations given by:

$$\begin{aligned} b_1(\mathbf{x}) &\mapsto \lambda b_1(\mathbf{x}) + \alpha b_3(\mathbf{x}) \\ b_2(\mathbf{x}) &\mapsto \lambda b_2(\mathbf{x}) + \beta b_3(\mathbf{x}) \\ b_3(\mathbf{x}) &\mapsto \tau b_3(\mathbf{x}) \end{aligned} \tag{18}$$

This transformation, *the generalized bas relief transform* (GBR), has been further investigated [3] and shown to apply even when cast and attached shadows are present. It can also be shown to be the only linear transformation which preserves the integrability constraints. Integrability therefore means that, in theory, the only linear ambiguity is the three-dimensional GBR constraint.

To understand the GBR we re-express it in terms of the geometry of the surface being viewed. The surface can be locally parameterized as $z = f(x, y)$ with normals of form:

$$\mathbf{n}(\mathbf{x}) = \frac{1}{\{\nabla\mathbf{f}\cdot\nabla\mathbf{f} + 1\}^{(1/2)}}(\nabla\mathbf{f}, -1). \tag{19}$$

It follows directly that the transformed surface $z = \bar{f}(x, y)$ is given by:

$$\bar{f}(x, y) = \lambda f(x, y) + \mu x + \nu y. \tag{20}$$

The GBR therefore corresponds to the standard bas relief ambiguity $f(x, y) \mapsto \lambda f(x, y)$ with the (non-standard) addition of an arbitrary background plane $\mu x+\nu y$. See [3] for further discussion and results on GBR.

Observe that the GBR equation (20) implies that the eigenvalues of the Hessian stay the same under a GBR if λ is positive and both change sign if λ is negative (because the Hessian depends on the second order derivatives of $f(x, y)$). This implies that the principal curvatures of the surface either both stay the same or both change sign under a GBR. This throws light on the recent work of Fan and Wolff [8] who also dealt with shape from shading for Lambertian objects with multiple images and non-constant albedo. Their approach involved taking ratios of the intensities to eliminate the albedo dependence. They were then able to show that properties such as the signs of the principal curvatures of the surfaces could be determined for simulated data. From our perspective, they were reconstructing the surface up to a GBR.

3.1 Using Surface Integrability

But we must demonstrate that surface integrability can be used in practice to reduce the linear ambiguity to a GBR. This subsection describes work, initially reported in [43], which demonstrates this. It enables us, for example, to estimate the shape and albedo even when the ambient lighting is completely different from the point illumination and might easily be confused with the albedo, see figure (8).

For simplicity of mathematics, we now define $\mathbf{P}$ to be a 3×4 matrix equal to $\left(\mathbf{P}_3, \mathbf{p_4}\right)$, where $\mathbf{P}_3$ and $\mathbf{p}_4$ are defined in equation (9). The three rows of $\mathbf{P}$ are three four-vectors which we denote by $\breve{\mathbf{p}}_1$, $\breve{\mathbf{p}}_2$, $\breve{\mathbf{p}}_3$.

Now we substitute the following values for $\mathbf{b}(\mathbf{x})$, $b_i(\mathbf{x}) = \sum_{\tau=1}^{4}\mathbf{P}_{i\tau}e_\tau(\mathbf{x})$, $\quad i = 1, 2, 3$.

$$\begin{aligned} &\textstyle\sum_{\mu<\nu}\{P_{3\mu}P_{2\nu} - P_{2\mu}P_{3\nu}\}\ \{e_\mu\frac{\partial e_\nu}{\partial x} - e_\nu\frac{\partial e_\mu}{\partial x}\} = \\ &\textstyle\sum_{\mu<\nu}\{P_{3\mu}P_{1\nu} - P_{1\mu}P_{3\nu}\}\ \{e_\mu\frac{\partial e_\nu}{\partial y} - e_\nu\frac{\partial e_\mu}{\partial y}\} \end{aligned} \tag{21}$$

This gives us $|\mathbf{\Omega}|$ linear equations for twelve unknowns (recall that $|\mathbf{\Omega}|$ is the size of the image). These equations are therefore over constrained but they can be solved by least squares to determine the $P_{3\mu}P_{2\nu} - P_{2\mu}P_{3\nu}$ and $P_{3\mu}P_{1\nu} - P_{1\mu}P_{3\nu}$ up to a constant scaling factor. These correspond to the cross products in four dimensions $\breve{\mathbf{p}}_1 \times \breve{\mathbf{p}}_3$, and $\breve{\mathbf{p}}_2 \times \breve{\mathbf{p}}_3$, . By inspection, the only transformation which preserves these cross products is:

$$\begin{aligned} \breve{\mathbf{p}}_1 &\rightarrow \lambda\breve{\mathbf{p}}_1 + \alpha\breve{\mathbf{p}}_3 \\ \breve{\mathbf{p}}_2 &\rightarrow \lambda\breve{\mathbf{p}}_2 + \beta\breve{\mathbf{p}}_3 \\ \breve{\mathbf{p}}_3 &\rightarrow \frac{1}{\lambda}\breve{\mathbf{p}}_3 \end{aligned} \tag{22}$$

which corresponds to the GBR [6],[3].

This means, consistent with the generalized bas relief ambiguity, that knowing these cross products will only allow us to solve for the $\mathbf{P}$ up to a generalized bas relief transformation. We now describe an explicit procedure to solve for the $\mathbf{P}$ in terms of the cross products.

First, recall that cross products arise in linear algebra as *co-factors* when calculating the inverses of matrices. In particular, the the cross product terms above are precisely the co-factors of the three by three submatrix $\mathbf{P}_3$. Recall that this matrix is related to co-factors by:

$$k\begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} = \begin{pmatrix} \Delta_{11} & \Delta_{21} & \Delta_{31} \\ \Delta_{12} & \Delta_{22} & \Delta_{32} \\ \Delta_{13} & \Delta_{23} & \Delta_{33} \end{pmatrix}^{-1} \tag{23}$$

where the co-factors are given by $\Delta_{11} = P_{22}P_{33} - P_{23}P_{32}$, etc. and k is a normalization constant.

In fact, the cross products determine the co-factors Δ_{11},Δ_{12},Δ_{13},Δ_{21},Δ_{22}, Δ_{23}. The remaining three co-factors $\Delta_{31}, \Delta_{32}, \Delta_{33}$ are unknown. These correspond to the parameters λ, α, β of the generalized bas relief transformation. Specific choices of them will correspond to specific transformations. We therefore select values for $\Delta_{31}, \Delta_{32}, \Delta_{33}$ which will later be modified as we solve for the generalized bas relief ambiguity.

We can now solve equation (23) to determine $\mathbf{P}_3$ up to GBR. To determine the remaining values of $\mathbf{P}$, the $\mathbf{p_4}$, we use the remaining cross product terms and least squares.

The results show that we can reconstruct the $\mathbf{b}$ up to a GBR. Figure (9) shows the result of the reconstruction in (a), but also shows the deformations that might arise due to a GBR in (b).

3.2 The Full SVD solution

The previous section has shown how we can use integrability to solve for $\mathbf{P}$, and hence $\mathbf{b}$, up to a GBR. This means that the full solution is given by $\mathbf{G}\mathbf{P}_3^*$, $\mathbf{G}\mathbf{p}_4^*$ where $\mathbf{G}$ is an arbitrary GBR and $\mathbf{P}_3^*$ and $\mathbf{p}_4^*$ are the output of our algorithm imposing integrability.

In this section, we show that additional assumptions can be used to determine the full solution. There are several possible choices. The one we describe here assumes that the magnitude of the light sources is constant. We emphasize that *we have not needed to make any assumptions about the magnitude of the light sources in order to estimate the surface up to a GBR.*

First, observe that we can directly solve for $\mathbf{w}$ and w_4 using the last equation of (9) and applying least squares. This gives:

$$w_4 = \sum_\mu f_4(\mu), \quad \mathbf{w} = \sum_\mu \mathbf{f}(\mu). \tag{24}$$

In addition, we can use the assumption that the light sources have constant magnitude which, without loss of generality, can be set equal to 1. Using equation (9) we find that we can express the magnitude squared of the $\mathbf{s}(\mu)\cdot\mathbf{s}(\mu)$ in terms of unknown quantities such as $\mathbf{Q}_3^T\mathbf{Q}_3$, $\mathbf{Q}_3^T\mathbf{q}_4$, and

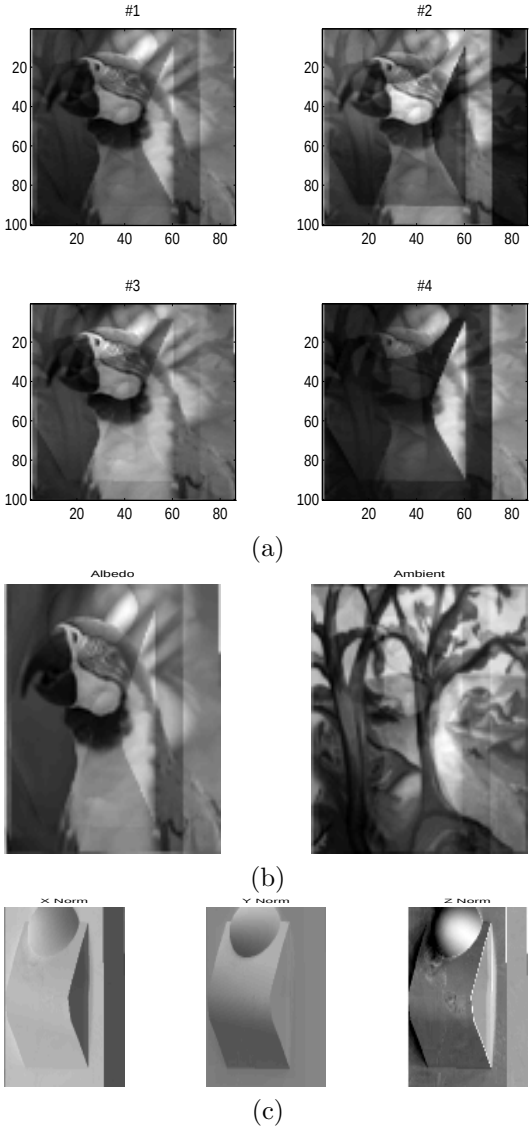
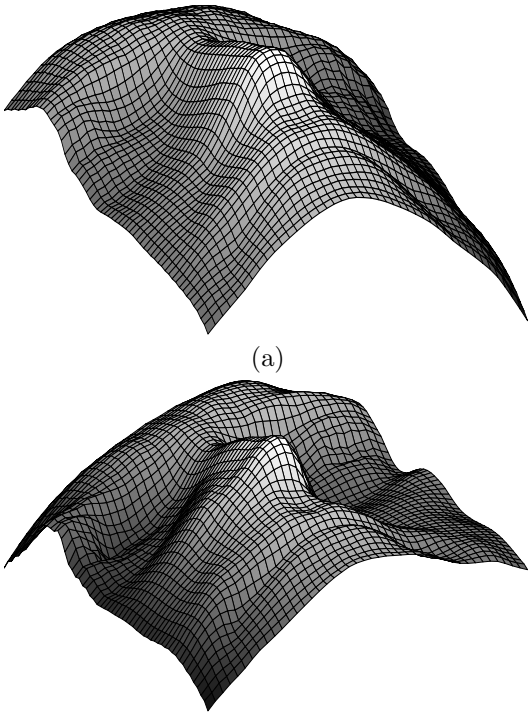


Figure 8: (a) shows four images of an object with the albedo of a parrot and ambient lighting of a tree. (b) shows that our approach manages to determine the albedo (left) and the ambient illumination (right). Moreover, the x,y and z components of the normals of the object are also found correctly. This nice separation between albedo and ambient will start degrading if there are many shadows in the images.

$\mathbf{q}_4^T\mathbf{q}_4$ and known quantities such as the eigenvectors. This, extending our analysis of Hayakawa, allows us to determine $\mathbf{Q}_3$ and $\mathbf{q}_4$ up to an arbitrary rotation matrix $\mathbf{R}$ (using a least squares cost function solved by a mixture of SVD and steepest descent).

By now, we have solved for the $\mathbf{P}$ up to a GBR $\mathbf{G}$, the $\mathbf{w}$ and w_4, and $\mathbf{Q}_3$ and $\mathbf{q}_4$ up to a rotation $\mathbf{R}$. We have, as yet, no knowledge of $\mathbf{v}$ and v_4.

But we still have the constraint that $\mathbf{P}^T\mathbf{Q} = \mathbf{D}$. We see that $\mathbf{G}$ and $\mathbf{R}$ only appear in this equation in the form $\mathbf{M} \equiv \mathbf{G}^T\mathbf{R}$. Indeed the equations $\mathbf{P}^T\mathbf{Q} = \mathbf{D}$ reduce to linear equations for



(a)

(b)

Figure 9: Figure (a) shows the face reconstructed up to an unknown GBR using integrability. Observe that the reconstruction appears to be accurate. In figure (b) we introduced a GBR by hand to demonstrate the type of deformations which might arise.

$\mathbf{M}, \mathbf{v}, v_4$. They can therefore be solved by least squares.

It now remains to determine $\mathbf{G}$ and $\mathbf{R}$ from $\mathbf{M}$. Recall that $\mathbf{M} = \mathbf{G}^T\mathbf{R}$, where $\mathbf{G}$ is a GBR and $\mathbf{R}$ is a rotation matrix. We therefore have that $\mathbf{M}\mathbf{M}^T = \mathbf{G}\mathbf{G}^T$ and so we can determine $\mathbf{G}\mathbf{G}^T$. From the form of a GBR it can be shown that $\mathbf{G}$ can be determined uniquely from $\mathbf{G}\mathbf{G}^T$ apart from a square root ambiguity (corresponding to the well-known concave/convex ambiguity). Now that $\mathbf{G}$ is known we can solve $\mathbf{M} = \mathbf{G}^T\mathbf{R}$ by least squares while imposing the condition that $\mathbf{R}$ is a rotation matrix. Figures (10) shows the results on the face.

4 Locating and Rejecting Shadows

So far, we have assumed that there are no shadows, or specularities, in the image. But it is clear from our dataset that this is a poor assumption. The least squares techniques we impose have given us some protection against the effect of shadows, but inevitably biases have been introduced.

In this section, we show that we can modify our method and eliminate shadows by an iterative process starting with the results given by the SVD method. Our strategy is to treat the shadows as outliers which can be removed by techniques from Robust Statistics [17]. We introduce a binary indicator variable $\mathbf{V}(\mathbf{x}, \mu)$ which can be used to indicate whether a point $\mathbf{x}$ is in shadow when the illuminant is $\mathbf{s}(\mu)$. This variable $\mathbf{V}(\mathbf{x}, \mu)$ must be estimated. To do so, we can use our current estimates of $\mathbf{b}(\mathbf{x})$ and $\mathbf{s}(\mu)$ to determine whether $\mathbf{x}$ is likely to be in shadow from light source $\mathbf{s}(\mu)$. We set $\mathbf{V}(\mathbf{x}, \mu) = 0$ if $\mathbf{n}(\mathbf{x}) \cdot \mathbf{s}(\mu) \leq T$, where T is a threshold. We then re-estimate $\mathbf{b}(\mathbf{x})$ and $\mathbf{s}(\mu)$

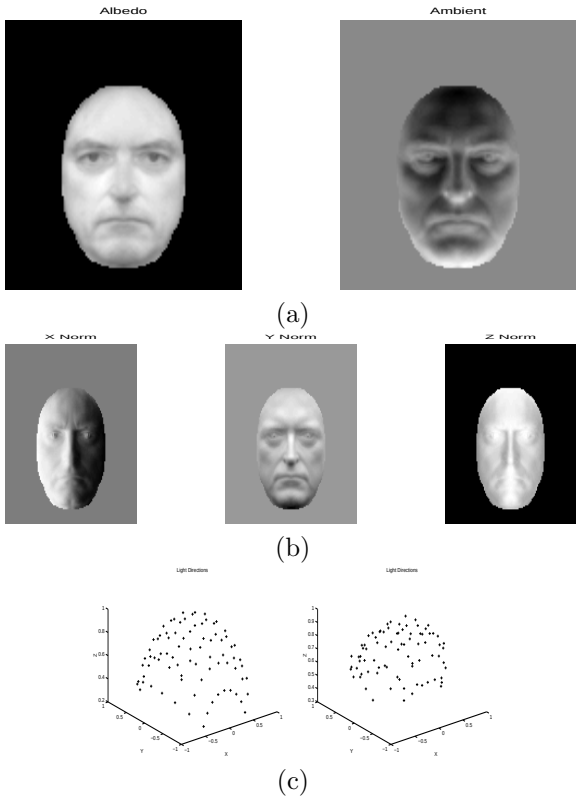


Figure 10: The full SVD on the face. (a) shows the estimated albedo and ambient term. (b) shows the x,y and z components of the surface normals. (c) shows the true lighting positions (left) and the estimated light source positions (right).

and repeat.

More precisely, we define a modified energy function:

$$E[\mathbf{V}(\mathbf{x},\mu),\mathbf{b}(\mathbf{x}),\tilde{a}(\mathbf{x}),\mathbf{s}(\mu)] = \sum_{\mathbf{x},\mu}\{\mathbf{I}(\mathbf{x},\mu) - \mathbf{V}(\mathbf{x},\mu)\mathbf{b}(\mathbf{x})\cdot\mathbf{s}(\mu) - \tilde{a}(\mathbf{x})\}^2$$
$$+c_1\sum_x\{(b_3\frac{\partial b_1}{\partial y} - b1\frac{\partial b_3}{\partial y}) - (b_3\frac{\partial b_2}{\partial x} - b_2\frac{\partial b_3}{\partial x})\}^2 + c_2\sum_\mu\{1-\mathbf{s}(\mu)\cdot\mathbf{s}(\mu)\}^2 \quad (25)$$

where c_1 and c_2 are constants.

We set

$$\mathbf{V}(\mathbf{x},\mu) = 0, \quad if\ \mathbf{b}(\mathbf{x})\cdot\mathbf{s}(\mu) \le T$$
$$\mathbf{V}(\mathbf{x},\mu) = 1, \quad if\ \mathbf{b}(\mathbf{x})\cdot\mathbf{s}(\mu) > T. \quad (26)$$

Then we minimize with respect to the variables $\mathbf{b}(\mathbf{x})$, $\mathbf{s}(\mu)$, and $\tilde{a}(\mathbf{x})$.

The energy is quadratic in $\tilde{a}(\mathbf{x})$ and can be minimized directly,

$$\frac{\partial E}{\partial \tilde{a}(\mathbf{x})} = -2\sum_\mu\{\mathbf{I}(\mathbf{x},\mu) - \mathbf{V}(\mathbf{x},\mu)\mathbf{b}(\mathbf{x})\cdot\mathbf{s}(\mu) - \tilde{a}(\mathbf{x})\}$$
$$\tilde{a}(\mathbf{x})^* = \frac{1}{N}\sum_\mu\{\mathbf{I}(\mathbf{x},\mu) - \mathbf{V}(\mathbf{x},\mu)\mathbf{b}(\mathbf{x})\cdot\mathbf{s}(\mu)\} \quad (27)$$

Where N is the number of images.

To minimize with respect to $\mathbf{s}(\mu)$ requires steepest descent.

$$\frac{\partial E}{\partial \mathbf{s}(\mu)} = \sum_{\mathbf{x}} \{\mathbf{I}(\mathbf{x},\mu) - \mathbf{V}(\mathbf{x},\mu)\mathbf{b}(\mathbf{x}) \cdot \mathbf{s}(\mu) - \tilde{a}(\mathbf{x})\}(-2\mathbf{V}(\mathbf{x},\mu)\mathbf{b}(\mathbf{x})) - 4c_2\{1 - \mathbf{s}(\mu)\cdot\mathbf{s}(\mu)\}\mathbf{s}(\mu) \quad (28)$$

For $\mathbf{b}(\mathbf{x})$ we also need steepest descent. Because of the derivative in $\mathbf{b}(\mathbf{x})$ we need to discretize the integrability terms.

$$\frac{\partial E}{\partial \mathbf{b}(\mathbf{x})} = \sum_{\mu} \{\mathbf{I}(\mathbf{x},\mu) - \mathbf{V}(\mathbf{x},\mu)\mathbf{b}(\mathbf{x}) \cdot \mathbf{s}(\mu) - \tilde{a}(\mathbf{x})\}(-2\mathbf{V}(\mathbf{x},\mu)\mathbf{s}(\mu)) + \text{ integrability terms} \quad (29)$$

The integrability energy terms are

$$(b_3\frac{\partial b_1}{\partial y} - b_1\frac{\partial b_3}{\partial y}) \rightarrow \{b^3_{i,j}(b^1_{i,j+1} - b^1_{i,j}) - b^1_{i,j}(b^3_{i,j+1} - b^3_{i,j})\}$$
$$(b_3\frac{\partial b_2}{\partial x} - b_2\frac{\partial b_3}{\partial x}) \rightarrow \{b^3_{i,j}(b^2_{i+1,j} - b^2_{i,j}) - b^2_{i,j}(b^3_{i+1,j} - b^3_{i,j})\}$$

The derivatives are given by:

$$\frac{\partial I_c}{\partial b^1_{i,j}} = -2c_1 \sum_{i,j} \{b^3_{i,j}(b^1_{i,j+1} - b^1_{i,j}) - b^1_{i,j}(b^3_{i,j+1} - b^3_{i,j}) - (b^3_{i,j}(b^2_{i+1,j} - b^2_{i,j}) - b^2_{i,j}(b^3_{i+1,j} - b^3_{i,j}))\}(b^3_{i,j+1} - b^3_{i,j}) \quad (30)$$

and similarly for the derivatives with respect to $b^2_{i,j}$ and $b^3_{i,j}$.

We repeat this process several times and obtain the shadows, see figure (11) and the albedo, ambient, and light source direction, see figure (12).

5 Conclusion

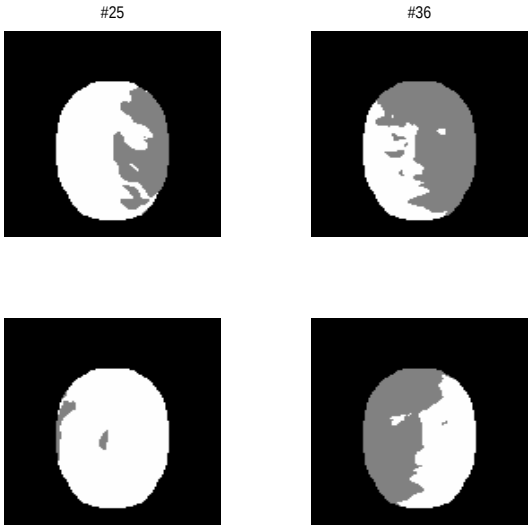
In this paper we have argued for generative models of object appearance and have stressed the necessity of modeling each factor – shape, albedo, illumination – independently. In particular, we have considered the special case of determining generative models when the input consists of multiple views of the same object, at similar poses, under different illumination conditions.

We have shown that the SVD approach for Lambertian models [15] can be generalized to include ambient illumination and we have argued, on empirical grounds, that such models are a good approximation to real world objects and that effects such as shadows and specularities can be treated as outliers.

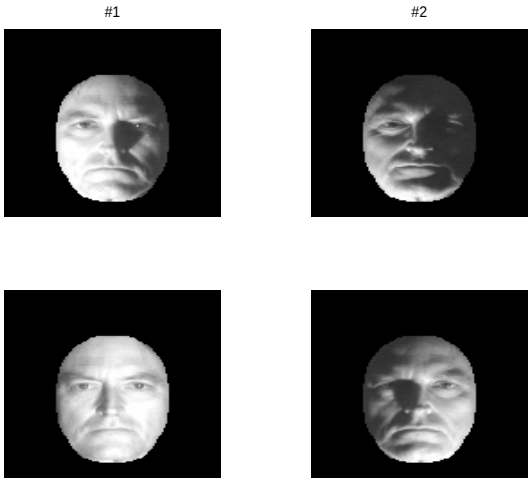
We analyzed the linear ambiguity remaining after SVD and showed that it could be reduced to a generalized bas relief ambiguity [3] by imposing surface consistency. Other methods for resolving the ambiguity were discussed and it was shown that some of them only worked under very special conditions. Our methods were implemented on both real and artificial data. Finally, we developed an iterative algorithm which, when combined with SVD, was able to detect shadows as outliers and remove them thereby improving the quality of the results.

Our approach assumed that the images of the object were viewed from the same position. This assumption has been relaxed, see [7], to allow structure from motion to aid shape from shading to obtain more accurate results.

In conclusion, we note that recent results by Georghiades, Kriegman and Belhumeur [11] are very encouraging. They demonstrate that models constructed using similar techniques to those in this paper can be highly effective for recognizing faces under different illumination conditions.



(a)



(b)

Figure 11: The top four images show the shadows extracted by our approach for the corresponding four input images at the bottom.

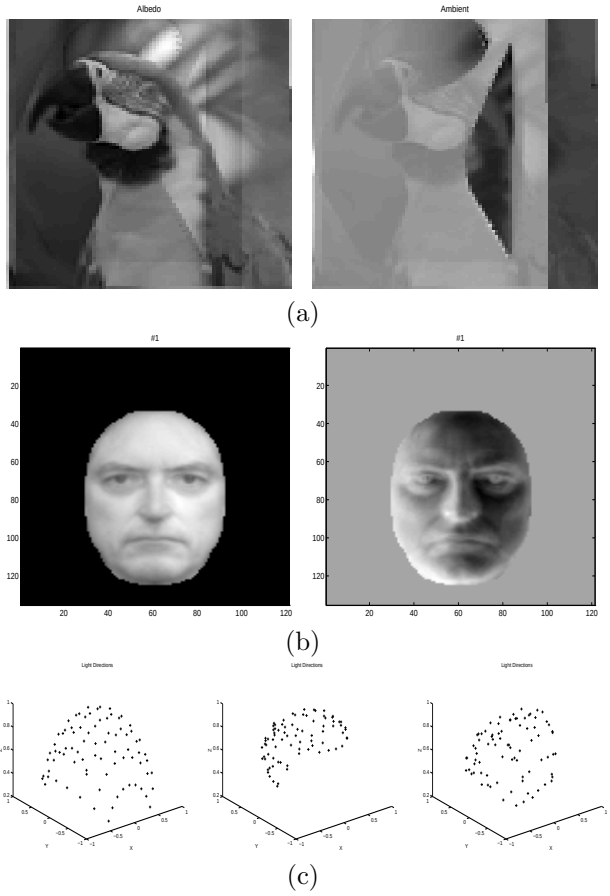


Figure 12: The albedo and ambient of the parrot image and the face image are obtained using our shadow rejection technique. They appear more accurate that those shown in figure 8. The bottom row shows the true lighting (left), the estimated lighting for the parrot (center) and the estimated lighting for the face (right). Again these appear more accurate than in figure 8.

Acknowledgements

It is a pleasure to acknowledge many helpful conversations with David Kriegman. James Coughlan also gave useful feedback on a draft of this paper. Detailed comments by three anonymous reviewers were also appreciated. Support for this work was provided by the NSF Grant IRI 9696107 and ARPA/ONR Contract N00014-95-1-1022.

Appendix I: SVD and Ambiguities.

This appendix introduces SVD and then gives details for the proofs sketched in section (2.4). More precisely. we will address: (i) Hayakawa's assumption that a specific rotation ambiguity in SVD can be resolved by setting the rotation to be the identity, (ii) the observation that the first three eigenvectors typically point along the axes of the cartesian coordinate system. We will show that these claims are closely related and depend on the symmetry of the dataset.

We refer to matrices either in bold face – i.e. $\mathbf{A}$ – or in coordinate form A_{ij}, where i labels the rows and j the columns. Similarly, we can refer to a vector as $\vec{e}$ or as e_i, where i labels the coordinates. Rotation matrices will be expressed as $\mathbf{R}$ and $\mathbf{W}$ and satisfy $\mathbf{W}^T\mathbf{W} = \mathbf{W}\mathbf{W}^T = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Singular Value Decomposition SVD

SVD implies that we can write any matrix $\mathbf{J}$ in form $\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ where $\mathbf{D}$ is a diagonal matrix with non-negative entries and $\mathbf{U}$ and $\mathbf{V}$ obey $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. It can then be verified that the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{J}\mathbf{J}^T$ and the eigenvalues are the squares of the diagonal elements of $\mathbf{D}$. Similarly, the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{J}^T\mathbf{J}$ and the eigenvalues are, once more, the squares of the diagonal elements of $\mathbf{D}$.

One way to obtain the SVD is to recall that any positive semi-definite matrix $\mathbf{M}$ can be expressed as $\mathbf{M} = \sum_i \lambda_i \mathbf{e}_i \mathbf{e}_i^T$ where $\{\lambda_i\}$ and $\mathbf{e}_i$ are the eigenvectors and eigenvalues of $\mathbf{M}$ respectively. This can be rewritten as $\mathbf{M} = \mathbf{E}\mathbf{D_1}\mathbf{E}^T$, where $\mathbf{D_1}$ is a diagonal matrix with diagonal terms $\{\lambda_i\}$ (non-negative because $\mathbf{M}$ is positive semi-definite) and the matrix $\mathbf{E}$ has $\mathbf{e}_i$ as its ith column. Because $\mathbf{D_1}$ is diagonal we can write it as $\mathbf{D_1} = \mathbf{D_2}\mathbf{D_2}$ (where $\mathbf{D_2}$ is a diagonal matrix whose diagonal terms are the square roots of the diagonal terms of $\mathbf{D_1}$). Therefore we can write $\mathbf{M} = \{\mathbf{E}\mathbf{D_2}\}\{\mathbf{E}\mathbf{D_2}\}^T$ (note that $\mathbf{D_2}^T = \mathbf{D_2}$ because the matrix is diagonal). More generally, we can write $\mathbf{M} = \{\mathbf{E}\mathbf{D_2}\mathbf{F}\}\{\mathbf{E}\mathbf{D_2}\mathbf{F}\}^T$ where $\mathbf{F}$ is an arbitrary matrix such that $\mathbf{F}\mathbf{F}^T = \mathbf{I}$. Now $\mathbf{J}\mathbf{J}^T$ is a positive semi-definite square matrix so it can be expressed in form $\{\mathbf{E}\mathbf{D_2}\mathbf{F}\}\{\mathbf{E}\mathbf{D_2}\mathbf{F}\}^T$ and hence we get $\mathbf{J} = \mathbf{E}\mathbf{D_2}\mathbf{F}$ where $\mathbf{D_2}$ is diagonal, $\mathbf{F}\mathbf{F}^T = \mathbf{I}$ and $\mathbf{E}\mathbf{E}^T = \mathbf{I}$ (by properties of eigenvectors).

SVD and Energy Minimization

We assume that the data $I(p, \mu)$ has been generated by a Lambertian model plus some additive noise. We define a cost function which allows us to obtain the lighting and surface properties:

$$E[b, s] = \sum_{p,\mu} \{I(p, \mu) - \sum_{i=1}^{3} b_i(p) s_i(\mu)\}^2. \tag{31}$$

Extremizing these equations with respect to $\mathbf{b}$ and $\mathbf{s}$ gives us two coupled equations to solve for $\mathbf{b}^*, \mathbf{s}^*$. It can be shown that $E[b, s]$ has a single minimum (though it has many saddle points). This solution can be attained using SVD. This gives:

$$\begin{aligned} b_i^*(p) &= \sum_j P_{ij} e_j(p), \quad \forall\, i, p, \\ s_i^*(\mu) &= \sum_j Q_{ij} f_j(\mu), \quad \forall\, i, \mu, \end{aligned} \tag{32}$$

where the $e(.)$ and $f(.)$ obey the eigenvectors equations:

$$\sum_{\mu'}\{\sum_{p} I(p,\mu)I(p,\mu')\}f_i(\mu') = \lambda_i f_i(\mu), \quad \forall\ i,\mu$$
$$\sum_{p'}\{\sum_{p} I(p,\mu)I(p',\mu)\}e_i(p') = \lambda_i e_i(p), \quad \forall\ i,p, \tag{33}$$

where the eigenvectors $e(.)$ and $f(.)$ are orthogonal (like all eigenvectors), so $\sum_p e_i(p)e_j(p) = \delta_{ij}$, and $\sum_\mu f_i(\mu)f_j(\mu) = \delta_{ij}$. The matrices $\mathbf{P}$ and $\mathbf{Q}$ are constrained to satisfy:

$$\mathbf{P}^T\mathbf{Q} = \mathbf{D_3}, \tag{34}$$

where $\mathbf{D_3}$ is a diagonal matrix whose diagonal elements are $\lambda_1^{1/2}, \lambda_2^{1/2}, \lambda_3^{1/2}$. These are the biggest three eigenvalues of the SVD decomposition of $\mathbf{J} = I(p,\mu)$ and correspond to the three dimensions of the vectors $\mathbf{b}$ and $\mathbf{s}$. For the more general case when we seek to minimize $\sum_{p,\mu}\{I(p,\mu) - \sum_{i=1}^{r} b_i(p)s_i(\mu)\}^2$, then we will need to take the r biggest eigenvalues of the SVD expansion (four the ambient case $r = 4$ and so we need four eigenvalues).

It can be seen that the error of the best solution $E[\mathbf{b}^*, \mathbf{s}^*]$, see equation (31), is simply the sum of the squares of the remaining eigenvalues of the SVD of $\mathbf{J}$ (i.e. all those not in $\mathbf{D_3}$).

Observe that $\mathbf{P}$ and $\mathbf{Q}$ are only determined up to an arbitrary linear transformation $\mathbf{P} \mapsto \mathbf{AP}$ and $\mathbf{Q} \mapsto \mathbf{A^{-1T}Q}$. As described in section (3), part of this ambiguity can be removed by requiring surface integrability and the remaining ambiguity is the Generalized Bas Relief (GBR) ambiguity.

The Orthogonal Lighting Conjecture and Hayakawa's Assumption

The orthogonal lighting conjecture states that the first three lighting eigenvectors correspond to the object illuminated from three orthogonal directions. The Hayakawa assumption specifies a way of choosing $\mathbf{P}$ assuming that $\mathbf{P}^T\mathbf{P}$ is known (it can be known if the albedos of the object are approximately known, alternatively we can estimate $\mathbf{QQ}^T$ if we assume that the light sources are of constant strength).

We analyze this conjecture and assumption assuming that the data has been generated by the naive Lambertian model (i.e. ignoring shadows). Our empirical experiments [13], [5] suggest that this assumption is fairly good as a first approximation. Moreover, the nature of singular value decomposition approach (SVD) means that, even if the reflectance is not Lambertian, the first three eigenvectors are likely to be the Lambertian components.

We will prove theorems which give necessary and sufficient conditions that the orthogonal lighting conjecture and Hayakawa's hidden assumption are true.

Let us define:

$$B_{ij} = \sum_{p} b_i(p)b_j(p), \quad S_{ij} = \sum_{\mu} s_i(\mu)s_j(\mu), \tag{35}$$

where $\{b_i(p)\}$ and $\{s_i(\mu)\}$ are the surface properties and the lighting conditions of the naive Lambertian model which generated the data.

Now we formally state the orthogonal lighting conjecture and Hayakawa's assumption.

Orthogonal Lighting Conjecture. *There exist three orthogonal unit vectors $\vec{k}_1, \vec{k}_2, \vec{k}_3$ such that $\vec{e}_1(p) \propto \vec{k}_1 \cdot \vec{b}(p)$, $\vec{e}_2(p) \propto \vec{k}_2 \cdot \vec{b}(p)$, and $\vec{e}_3(p) \propto \vec{k}_3 \cdot \vec{b}(p)$.*

Comment. By equation (32) this condition is equivalent to saying we can find a coordinate system in which $\mathbf{P}$ is diagonal.

Hayakawa's Assumption. *If we express $\mathbf{P}^T\mathbf{P}$ in diagonal form as $\mathbf{W}^T\mathbf{MW}$, where $\mathbf{W}$ is an orthogonal matrix and $\mathbf{M}$ is diagonal, then $\mathbf{P} = \mathbf{M}^{1/2}\mathbf{W}$.*

Comment I. $\mathbf{P}^T\mathbf{P}$ can always be expressed in diagonal form as $\mathbf{W}^T\mathbf{MW}$. But the general solution is $\mathbf{P} = \mathbf{RM}^{1/2}\mathbf{W}$, where $\mathbf{R}$ is an arbitrary rotation matrix.

Comment II. The diagonal form $\mathbf{W}^T\mathbf{M}\mathbf{W}$ is not unique because we can always re-order the diagonal elements of $\mathbf{M}$ by applying a permutation matrix $\mathbf{\Pi}$[1]. More precisely, $\mathbf{\Pi}\mathbf{M}\mathbf{\Pi}^T$ is also a diagonal matrix whose diagonal elements are a re-ordering of those in $\mathbf{M}$. Thus we can send $\mathbf{M} \mapsto \mathbf{\Pi}\mathbf{M}\mathbf{\Pi}^T$ and $\mathbf{W} \mapsto \mathbf{\Pi}\mathbf{W}$. The only way to make the diagonal form unique is by putting a condition on the ordering of the diagonal elements of $\mathbf{M}$.

This means that Hayakawa's assumption is ill-defined. We can, however, generalize it by incorporating an arbitrary permutation matrix. Observe that this only causes slight ambiguity in the output of the algorithm (i.e. we know what the shape and surface properties are, but only up to a permutation on the x, y, z axes).

Hayakawa's Generalized Assumption. *If we express* $\mathbf{P}^T\mathbf{P}$ *in diagonal form as* $\mathbf{W}^T\mathbf{M}\mathbf{W}$*, where* $\mathbf{W}$ *is an orthogonal matrix and* $\mathbf{M}$ *is diagonal, then* $\mathbf{P} = \mathbf{\Pi}\mathbf{M}^{1/2}\mathbf{W}$*, where* $\mathbf{\Pi}$ *is an arbitrary permutation matrix.*

To prove our theorems, we first need some intermediate results.

Lemma 1. *If the data has been generated by a naive Lambertian model with lighting* $s_i(\mu)$ *and surface properties* $b_i(p)$*, then we have* $\mathbf{B} = \mathbf{P}\mathbf{P}^T$ *and* $\mathbf{S} = \mathbf{Q}\mathbf{Q}^T$*.*

Proof. By equation (35), $S_{ij} = \sum_\mu s_i(\mu)s_j(\mu)$. *Now, equation (32) gives us* $s_i(\mu) = \sum_j Q_{ij}f_j(\mu)$. *Substituting for* $s_i(\mu)$ *and using the orthogonality of the* $f(.)$*'s we obtain* $\mathbf{S} = \mathbf{Q}\mathbf{Q}^T$. *Similarly,* $\mathbf{B} = \mathbf{P}\mathbf{P}^T$.

This Lemma allows us to understand when Hayakawa's generalized assumption is valid.

Theorem 1. *If the data is generated by a naive Lambertian model, then Hayakawa's generalized assumption is valid if, and only if,* $\mathbf{B}$ *is diagonal.*

Proof. *Set* $\mathbf{P} = \mathbf{R}\mathbf{M}^{\mathbf{1/2}}\mathbf{W}$*, where the rotation matrix* $\mathbf{R}$ *is unknown. If the data is naive Lambertian, then, by Lemma 1,* $\mathbf{B} = \mathbf{P}\mathbf{P}^T = \mathbf{R}\mathbf{M}\mathbf{R}^T$ *(recalling that* $\mathbf{W}$ *is a rotation and* $\mathbf{M}$ *is diagonal). Thus* $\mathbf{B}$ *is diagonal if, and only if,* $\mathbf{R}$ *is a permutation matrix – i.e. if, and only if, Hayakawa's generalized assumption is true.*

Comment. The constraint that $\mathbf{B}$ is diagonal can be thought of as a symmetry assumption about the surface properties of the viewed object. It will be true, for example, if we look at an ellipsoidal object, with constant albedo, from directly in front. More importantly, it will also be true (statistically speaking) if the individual components of the surface properties can be modelled as independent random variables with zero mean. This seems close to the data used in Hayakawa's experiments, and may be a good approximation for some scenes.

Lemma 2. *The orthogonal lighting conjecture is independent of the choice of coordinate system.*

Comment. This result is not surprisingly, but because this problem contains special coordinate systems, defined by the viewer direction and defined by the eigenvectors, it is wise to check it.

Proof. *Let us start in the coordinate system defined by the viewer position (i.e. with the z axis pointing along the direction of sight). By equation (32):*

$$\begin{aligned} \vec{b}(p) &= \mathbf{P}\vec{e}(p), \\ \vec{s}(\mu) &= \mathbf{Q}\vec{f}(\mu). \end{aligned} \tag{36}$$

[1]The permutation matrices form a discrete subgroup of the rotation group. Geometrically their function is to relabel the x, y and z axes. This means there are $3! = 6$ of them.

Now let us rotate the coordinate system by a rotation $\mathbf{R}$*. This gives:*

$$\begin{aligned}\vec{\hat{b}}(p) &= \mathbf{R}\vec{b}(p) = \mathbf{RP}\vec{e}(p),\\ \vec{\hat{s}}(\mu) &= \mathbf{R}\vec{s}(\mu) = \mathbf{RQ}\vec{f}(\mu).\end{aligned} \tag{37}$$

Thus we can define:

$$\vec{\hat{b}}(p) = \hat{\mathbf{P}}\vec{e}(p), \quad \vec{\hat{s}}(\mu) = \hat{\mathbf{Q}}\vec{f}(\mu). \tag{38}$$

This gives:

$$\hat{\mathbf{P}} = \mathbf{RP}, \quad \hat{\mathbf{Q}} = \mathbf{RQ} \tag{39}$$

and we can confirm that $\hat{\mathbf{P}}^T\hat{\mathbf{Q}} = \mathbf{P}^T\mathbf{Q}$*. Therefore equations (32,34) retain the same form if we rotate the coordinate system.*

We are now ready to prove our main theorem.

Theorem 2. *If the data has been generated by the naive Lambertian model, then the orthogonal lighting conjecture holds if, and only if, the two matrices* $S_{ij} = \sum_\mu s_i(\mu)s_j(\mu)$ *and* $B_{ij} = \sum_p b_i(p)b_j(p)$ *commute.*

We have:

Proof. *First observe that the orthogonal lighting conjecture is true if, and only if, we can find a coordinate system in which* $\mathbf{P}$ *is diagonal which, by equation (34)* $\mathbf{P}^T\mathbf{Q} = \mathbf{D}$*, also implies that* $\mathbf{Q}$ *is diagonal. If both* $\mathbf{P}$ *and* $\mathbf{Q}$ *are diagonal, then the eigenvectors* $e(.)$ *and* $f(.)$ *will be proportional to the cartesian components of the* $b(.)$ *and* $s(.)$*.*

We consider the eigenvalue equation (33) for $e(.)$ *and* $f(.)$*. Now*

$$\sum_p I(p,\mu)I(p,\mu') = \sum_p b_i(p)s_i(\mu)b_j(p)s_j(\mu') = s_i(\mu)s_j(\mu')B_{ij},$$
$$\sum_\mu I(p,\mu)I(p',\mu) = b_i(p)b_j(p'S_{ij}. \tag{40}$$

The eigenvalue equation (33) now becomes:

$$\begin{aligned}\sum_{\mu'}\{B_{ij}s_i(\mu)s_j(\mu')\}f_j(\mu') &= \lambda_i f_i(\mu),\\ \sum_{p'}\{S_{ij}b_i(p)b_j(p')\}e_j(p') &= \lambda_i e_i(p).\end{aligned} \tag{41}$$

Now, from equation (32), we substitute $\vec{f}(\mu') = \mathbf{Q}^{-1}\vec{s}(\mu')$ *and* $\vec{e}(p') = \mathbf{P}^{-1}\vec{b}(p')$ *into these equations, recalling the definitions, equation (35), of* $\mathbf{S}$ *and* $\mathbf{B}$*, and comparing coefficients of* $\vec{s}(\mu)$ *and* $\vec{b}(p)$ *gives:*

$$\mathbf{Q}^T\mathbf{BS} = \mathbf{D}^2\mathbf{Q}^T, \quad \mathbf{P}^T\mathbf{SB} = \mathbf{D}^2\mathbf{P}^T, \tag{42}$$

where $\mathbf{D}$ *is a diagonal matrix with elements* $\lambda_1^{1/2}, \lambda_2^{1/2}, \lambda_3^{1/2}$*.*

Using the results of Lemma 1 gives:

$$\mathbf{Q}^T\mathbf{BQ} = \mathbf{D}^2, \quad \mathbf{P}^T\mathbf{SP} = \mathbf{D}^2. \tag{43}$$

Now let us assume that $\mathbf{S}$ *and* $\mathbf{B}$ *commute. This means that we can rotate the coordinate system until they are both diagonal. This will allow us, using Lemma 1, to express:*

$$\mathbf{Q} = \mathbf{S}^{1/2}\mathbf{R}_1, \quad \mathbf{P} = \mathbf{B}^{1/2}\mathbf{R}_2, \tag{44}$$

where $\mathbf{R}_1$ *and* $\mathbf{R}_2$ *are (unknown) rotation matrices.*

Substituting into equation (43), and using the fact that $\mathbf{S}$ *and* $\mathbf{B}$ *are diagonal (and hence commute), gives:*

$$\mathbf{R}_1^T\mathbf{S}\mathbf{B}\mathbf{R}_1 = \mathbf{D}^2 = \mathbf{R}_2^T\mathbf{S}\mathbf{B}\mathbf{R}_2. \tag{45}$$

But both $\mathbf{SB}$ *and* $\mathbf{D}$ *are diagonal, and so the only solution to these equations is to set* $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{\Pi}$*, where* $\mathbf{\Pi}$ *is a permutation matrix. This implies that* $\mathbf{Q}$ *and* $\mathbf{P}$ *are diagonal, and hence the result follows.*

Conversely, suppose that the orthogonal lighting conjecture is true. Then we can find a coordinate system in which $\mathbf{P}$ *and* $\mathbf{Q}$ *are diagonalizable. Then Lemma 2 implies that* $\mathbf{S}$ *and* $\mathbf{B}$ *are both diagonal, and hence commute.*

References

[1] J.J. Atick, P.A. Griffin, and A.N. Redlich. "Statistical Approach to Shape from Shading: Reconstruction of 3-Dimensional Face Surfaces from Single 2-Dimensional Images". *Neural Computation.* 8, No. 6, pp 1321-1340. 1996.

[2] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible lighting conditions. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 270–277, 1996.

[3] P. Belhumeur, D. Kriegman, and A.L. Yuille. "The Bas-Relief Ambiguity". In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition.* CVPR97. Puerto Rico. 1997.

[4] I. Biederman. "Recognition-by-components: A theory of human image understanding". *Psychological Review*, Vol 94, pp. 115-147. 1987.

[5] R. Epstein, P.W. Hallinan and A.L. Yuille. "5 ± Eigenimages Suffice: An Empirical Investigation of Low-Dimensional Lighting Models". In *Proceedings of IEEE WORKSHOP ON PHYSICS-BASED MODELING IN COMPUTER VISION.* pp 108-116. 1995.

[6] R. Epstein, A.L. Yuille, and P.N. Belhumeur. "Learning object representations from lighting variations".In **Object Representation in Computer Vision II**. Eds. J. Ponce, A. Zisserman, and M. Hebert. Springer Lecture Notes in Computer Science 1144. 1996.

[7] R. Epstein. **Learning Object Representations from Greyscale Images.** PhD Thesis. Division of Applied Sciences. Harvard University. 1996.

[8] J. Fan and L.B. Wolff. "Surface Curvature and Shape Reconstruction from Unknown Multiple Illumination and Integrability". *Computer Vision and Image Understanding.* Vol. 65. No. 2. pp 347-359. 1997.

[9] D. Forsyth and A. Zisserman. "Reflections on Shading". *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 7. 1991.

[10] W. Freeman and J. Tennenbaum. "Learning Bilinear Models for Two-Factor problems in Vision". In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition.* CVPR97. Puerto Rico. 1997.

[11] A. Georghiades, D. Kriegman, and P. Belhumeur. "Illumination cones for recognition under variable lighting". In *Proceedings Computer Vision and Pattern Recognition.* CVPR'98. 1998.

[12] U. Grenander, Y. Chou, and K.M. Keenan. **HANDS: A Pattern Theoretic Study of Biological Shapes**. Springer-Verlag. New York. 1991.

[13] P.W. Hallinan. "A low-dimensional lighting representation of human faces for arbitrary lighting conditions". In. *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pp 995-999. 1994.

[14] P.W. Hallinan. **A Deformable Model for Face Recognition under Arbitrary Lighting Conditions.** PhD Thesis. Division of Applied Sciences. Harvard University. 1995.

[15] K. Hayakawa. "Photometric Stereo under a light source with arbitrary motion". *Journal of the Optical Society of America* A, 11(11). 1994.

[16] B.K.P. Horn and M. J. Brooks, Eds. **Shape from Shading**. Cambridge MA, MIT Press, 1989.

[17] P.J. Huber. **Robust Statistics**. John Wiley and Sons. New York. 1980.

[18] Y. Iwahori, R.J. Woodham and A Bagheri "Principal Components Analysis and Neural Network Implementation of Photometric Stereo". In *Proceedings of the IEEE Workshop on Physics-Based Modeling in Computer Vision.* pp117-125 (1995).

[19] D. Kriegman and P.N. Belhumeur. "What Shadows Reveal About Object Structure". To appear in Proceedings of the European Conference on Computer Vision. ECCV'98. 1998.

[20] D. Marr. Vision. Freeman: New York. 1982.

[21] Y. Moses, Y. Adini, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. In *European Conf. on Computer Vision*, pages 286–296, 1994.

[22] H. Murase and S. Nayar. "Visual learning and recognition of 3-D objects from appearance". *Int. Journal of Computer Vision*. 14. pp 5-24. 1995.

[23] S.K. Nayar, K. Ikeuchi, and T. Kanade. "Surface reflection: physical and geometrical perspectives". *IEEE Trans. Pattern Analysis and Machine Intelligence*. Vol. 13, pp 611-634. 1991.

[24] A. Pentland, B. Moghaddam, and Starner. View-based and modular eigenspaces for face recognition. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[25] Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, **343**, 263-266.

[26] T. Poggio and K. Sung. Example-based learning for view-based human face detection. In *Proc. Image Understanding Workshop*, pages II:843–850, 1994.

[27] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, P. Boyes-Braem. "Basic objects in natural categories". Cogn. Psychol. 8:382-439. 1976.

[28] A. Shashua. **Geometry and Photometry in 3D Visual Recognition**. PhD Thesis. MIT. 1992.

[29] H-Y Shum, M.H. Hebert, K. Ikeuchi and R. Reddy. "An Integral Approach to free-Form Object Modeling". *IEEE Transactions of Pattern Analysis and Machine Intelligence*. Vol. 19. No. 12. December. 1997.

[30] W. Silver. *Determining Shape and Reflectance Using Multiple Images.* PhD Thesis. MIT, Cambridge, MA. 1980.

[31] P.C. Teo, E.P. Simoncelli and D.J. Heeger. "Efficient Linear Re-rendering for Interactive Lighting Design". Report No. STAN-CS-TN-97-60. Dept. Computer Science. Stanford University. October 1997.

[32] L. Sirovitch and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Optical Soc. of America A*, 2:519–524, 1987.

[33] F. Solomon and K. Ikeuchi. "An Illumination Plannner for Convex and Concave Objects". In *Proceddings of the IEEE Workshop on Physics-Based Modeling in Computer Vision*, pp 100-107.. June 18-19. 1995.

[34] M. J. Tarr, & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, (28), 233-282.

[35] M. J. Tarr and H. H. Bülthoff. "Is human object recognition better described by geon-structural-descriptions or by multiple-views?" *Journal of Experimental Psychology: Human Perception and Performance*, Vol 21, No. 6. 1995.

[36] M. J. Tarr, Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, *8*(4), 282-289.

[37] M. J. Tarr and D. Kersten and H. H. Bülthoff. "Why the visual system might encode the effects of illumination". Submitted to *Vision Research*. 1997.

[38] M. J. Tarr. "Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects". *Psychonomic Bulletin and Review*, Vol. 2, No. 1, pp. 55-82. 1995.

[39] G. Taubin, F. Cukierman, S. Sullivan, J. Ponce, and D. Kriegman. Parameterized families of polynomials for bounded algebraic curve and surface fitting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 16(3):287–303, 1994.

[40] M. Turk and A. Pentland. "Eigenfaces for recognition." *J. of Cognitive Neuroscience*, 3(1), 1991.

[41] R. Woodham. "Analyzing Images of Curved Surfaces". *Artificial Intelligence*, 17, pp 117-140. 1981.

[42] A.L. Yuille. "Mathematical Results on SVD for Lighting and Warps". Technical Report. Smith-Kettlewell Eye Research Institute. San Francisco. California 94115.

[43] A.L. Yuille and D. Snow. "Shape and Albedo from Multiple Images using Integrability." In *Proceedings of Computer Vision and Pattern Recognition* (CVPR'97). Puerto-Rico. 1997.