The Generic Viewpoint Constraint Resolves the Generalized Bas Relief Ambiguity

A. L. Yuille[1]
Smith-Kettlewell Eye Research Institute
2318 Fillmore Street
San Francisco, CA 94115
e-mail: yuille@ski.org

James M. Coughlan
Smith-Kettlewell Eye Research Institute
2318 Fillmore Street
San Francisco, CA 94115
e-mail: coughlan@ski.org

Scott Konishi
Smith-Kettlewell Eye Research Institute
2318 Fillmore Street
San Francisco, CA 94115
e-mail: konishi@ski.org

Abstract — **Recent work [1] has demonstrated the GBR ambiguity which is inherent in the perception of depth from shading and shadow cues. The purpose of this paper is firstly to extend the GBR transform to allow for different viewpoints. We demonstrate how this can be done by using image warping. This enables us to generalize the GBR to deal with situations where objects are viewed from different viewpoints (as well as under different lighting conditions). The second point is to analyze how the generic viewpoint assumption [4]interacts with the GBR. Our results show that generic viewpoint constraint disambiguates the GBR by biasing towards a flat surface.**

I. Introduction

From prehistoric times artists have carved portraits in bas relief (see the Assyrian art in the British Museum). In such portraits the relative depth of the object is compressed (mathematically there is a transformation $z \mapsto \lambda z$ where z is the depth and $\lambda < 1$ is the compression factor). The resulting sculptures, however, appear to be very realistic. Work by Koenderink and his collaborators [5], [6] have shown that the equations for structure from motion are insensitive to bas relief ambiguities. Thus an observer moving near the sculpture would only be able to estimate depth up to an unknown scale factor. Therefore motion cues would not enable an observer to distinguish between a full three-dimensional sculpture and a bas relief one.

Recent work by Belhumeur, Kriegman and Yuille [1], see also [8], has shown there is a similar ambiguity in the perception of shape through shading and shadows cues. In such cases, there is a generalized bas relief ambiguity (GBR), see figure (1), which includes the standard bas relief ambiguity as a special case. Thus in conjunctions with Koenderink's results this suggests that motion and shading/shadow cues taken independently seem unable to resolve the bas relief ambiguity.

This paper explores mechanisms for resolving the GBR ambiguity. In particular, we consider the generic view constraint first formulated in Bayesian terms by Freeman [4], [2]. The intuition behind this constraint is that human observers tend to avoid interpretations of the data which correspond to "accidental", or unusual, views. As we will prove, the generic viewpoint assumption is enough to resolve the GBR ambiguity completely. The solution biases towards flat surfaces.

We also explore the coupling between the ambiguities arising from shading/shadows and from the observer's egomotion (i.e. the object and the lighting stays still but the observer moves). First we demonstrate that changes of viewpoint can be modelled as two-dimensional spatial warps of the images, see [7]. We characterize how these warps change as the viewed object undergoes a GBR transformation. This enables us to generalize the GBR to allow for viewpoint changes. Suppose, for example, that we have two objects which are related by a GBR from a fixed viewpoint. Then for each possible novel viewpoint of the first object there exists a viewpoint of the second which is identical. These results assume that the projection equations are described by the affine camera ("a very good approximation" Dr. A. Zisserman. Private communication). We stress that these results do not apply to the changes at the object boundaries where certain regions become invisible as the viewpoint changes.

Figure 1: If the lighting conditions are unknown then it is impossible to distinguish between two objects related by a GBR (generalized bas relief ambiguity). Because for any image of the first object, under one illumination condition, we can always find a corresponding illumination condition which makes the second object appear identical (i.e. generate the identical image).

We briefly describe how our analysis of the generic viewpoint assumption to deal with this combined shading/shadows and viewpoint/egomotion ambiguity.

In section (II) we briefly summarize the GBR. Section (III) describes a new special case of the GBR ambiguity where the albedo of the surface is unchanged by the transformation. In section (IV) we extend the GBR to viewpoint changes by allowing for spatial warping of images.

II. The Generalized Bas Relief Ambiguity for Shading and Shadows

The Lambertian lighting model determines the image intensity $I(x,y)$ in terms of the surface shape, represented by the surface normal $\vec{n}(x,y)$, the surface albedo $a(x,y)$, and the illumination sources $\{\vec{s}_i : i = 1, ..., N\}$.

$$I(x,y) = \sum_{i=1}^{N} \max\{\vec{b}(x,y) \cdot \vec{s}_i, 0\}, \quad (1)$$

[1]This work was supported by the National Institute of Health (NEI) with grant number RO1-EY 12691-01.

where $\vec{b}(x,y) = a(x,y)\vec{n}(x,y)$. The max operation is used to model attached shadows.

Often attached shadows are ignored in which case the Lambertian model can be simplified to be $I(x,y) = \vec{b}(x,y) \cdot \vec{s}$ (where removing the max operation means that we can sum all the light sources into a composite single source $\vec{s} = \sum_{i=1}^{N} \vec{s}_i$).

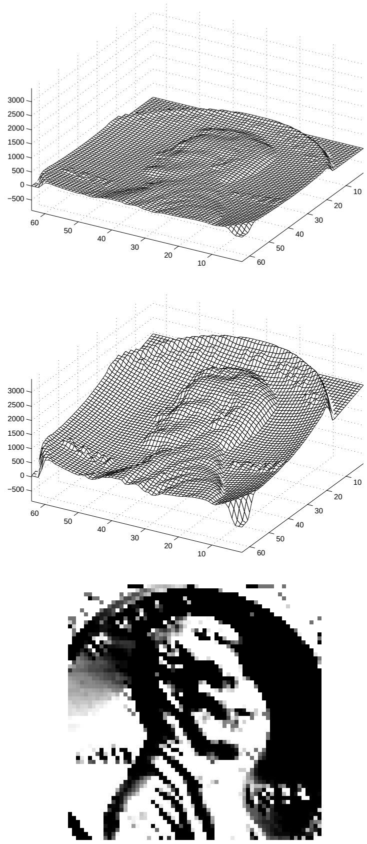

Figure 2: Two objects have their shape related by a GBR (top and centre). For any illumination of the first object we can always find an illumination of the second so that both objects generate the same image (bottom).

It has been shown [1] that there is an ambiguity in these equations. More precisely, there is a transformation $\vec{b}(x,y) \mapsto \mathbf{G}\vec{b}(x,y)$ and $\vec{s} \mapsto \mathbf{G}^{T,-1}\vec{s}$ where G is a matrix representing a *generalized bas relief* (GBR) transformation. If two objects $O, \hat{O}$ are related by a GBR, so that $\vec{\hat{b}}(x,y) = \mathbf{G}\vec{b}(x,y)$ for some GBR $\mathbf{G}$, then for any illumination of object O there will always be a corresponding illumination of object $\hat{O}$ so that the images of the two objects are identical, see figure (2). The Generalized Bas Relief (GBR) ambiguity holds for the full Lambertian model with attached shadows. It can also be shown to hold when cast shadows (cast by the object onto itself) are taken into account [1].

The GBR transformations form a group and can be represented in matrix form as:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & -\mu/\lambda \\ 0 & 1 & -\nu/\lambda \\ 0 & 0 & 1/\lambda \end{pmatrix}, \qquad (2)$$

where λ, μ, ν are constants. It is straightforward to check that matrices of this form satisfy the axioms of a group. The transformation $\mathbf{G}^{T,-1}$ on the light sources is of form:

$$\mathbf{G}^{T,-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix}. \qquad (3)$$

We can also express the object in terms of its surface shape $z = f(x,y)$ and albedo $a(x,y)$. These are related to the surface normals by the equations $\mathbf{n}(\mathbf{x}) = \frac{1}{\{\nabla\mathbf{f}\cdot\nabla\mathbf{f}+1\}^{(1/2)}}(\nabla\mathbf{f}, -1)$. In this representation the GBR transformation becomes:

$$f(x,y) \mapsto \lambda f(x,y) + \mu x + \nu y,$$
$$a(x,y) \mapsto \frac{a(x,y)}{\lambda} \frac{\{1 + (\lambda f_x + \mu)^2 + (\lambda f_y + \nu)^2\}^{1/2}}{\{1 + f_x^2 + f_y^2\}^{1/2}}. \qquad (4)$$

Geometrically, 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 $\mu x + \nu y$.

We illustrate the GBR for a Lincoln penny. The figure shows two objects related by a GBR. It is, however, impossible to tell them apart if the lighting condition is unknown. Because for any illumination of the first object there exists an illumination of the second object which gives identical images.

III. GBR transform which do not alter albedo

More understanding can be gained about the GBR by considering the following special case (which to the best of our knowledge has not been previously analyzed). The GBR usually involves altering the albedo $a(x,y)$ by a transformation that varies spatially (for example, a surface with spatially constant albedo is typically transformed into a surface with spatially varying albedo).

Suppose we have a surface $z = f(x,y)$ where $f_y(x,y) = K$ which is a constant (i.e. $f(x,y) = g(x) + Ky$ for some function $g(.)$). Now apply a GBR transform to it with $\mu = 0$ and the parameters λ, μ, ν chosen to satisfy $\{\lambda^2 - \nu^2 - 1\}/(2\lambda\nu) = K$. Then it is straightforward to derive the relations:

$$g(x) + Ky \mapsto \lambda g(x) + (\lambda K + \nu)y,$$
$$a(x,y) \mapsto a(x,y). \qquad (5)$$

In other words, if we consider surfaces which are flat in the y direction (i.e. f_y is spatially constant) then these surfaces are invariant to a subset of the GBR transforms which *preserve the spatial pattern of the albedo*, see figure (3).

IV. GBR and Viewpoint Changes

We now consider the effect of how the GBR transform is affected by change of viewpoint. It is known [1] that small changes do not affect the GBR. Moreover, from the work of Koenderink [1], [6] we know that for pointlike features in three-dimensions there is a bas relief ambiguity even when the viewpoint changes).

In this paper, we explore the interaction between the viewpoint and lighting changes. Our results enable us to generalize the GBR ambiguity to include both types of changes.

To allow for viewpoint changes, we define a world coordinate system $u, v, f(u,v)$ for representing an object in space. The coordinates of these points under projection are given by the equations:

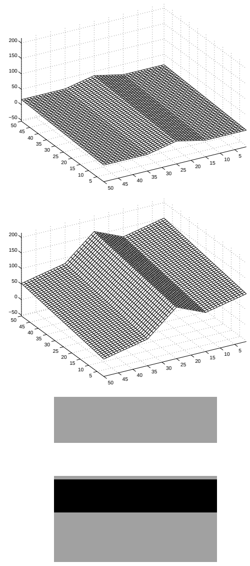

Figure 3: Two wedge objects have their shape related by a GBR (top and centre). For any illumination of the first object we can always find an illumination of the second so that both objects generate the same image (bottom). The light for the first object is $s = -(1,1,1)$ and for the second is $\hat{s} = -(0.3333, 0.3333, 1.6667)$.

$$x = a_{11}u + a_{12}v + a_{13}f(u,v),$$
$$y = a_{21}u + a_{22}v + a_{23}f(u,v). \qquad (6)$$

The coefficients $\vec{a}_1 = (a_{11}, a_{12}, a_{13})$ and $\vec{a}_2 = (a_{21}, a_{22}, a_{23})$ determine the projection. We will be concerned with two types of projection [3]. The first is the *affine camera* where these coefficients can take any value. The second is orthographic projection where $\vec{a}_1$ and $\vec{a}_2$ are constrained to be orthogonal unit vectors.

We first assume that we are dealing with the affine camera. Consider two objects which are related by a GBR in the world coordinate system (and the illumination is changed accordingly). The second object has shape $u, v, \lambda f(u,v) + \mu u + \nu v$. Let the second object be projected with coefficients $\vec{\hat{a}}_1 = (\hat{a}_{11}, \hat{a}_{12}, \hat{a}_{13})$ and $\vec{\hat{a}}_2 = (\hat{a}_{21}, \hat{a}_{22}, \hat{a}_{23})$. It will be projected to points:

$$\hat{x} = \hat{a}_{11}u + \hat{a}_{12}v + \hat{a}_{13}\{\lambda f(u,v) + \mu u + \nu v\},$$
$$\hat{y} = \hat{a}_{21}u + \hat{a}_{22}v + \hat{a}_{23}\{\lambda f(u,v) + \mu u + \nu v\}. \qquad (7)$$

Because the two objects are related by a GBR transform the image intensities at $u, v, f(u,v)$ and $u, v, \lambda f(u,v) + \mu u + \nu v$ will have the same intensity. By the properties of Lambertian objects this intensity will be independent of the viewpoint. The two objects will therefore give rise to the same intensity image *provided* the projection coefficients are chosen so that $(\hat{x}, \hat{y}) = (x, y)$. It is straightforward algebra to determine that this condition is satisfied if, and only if, $\vec{a}_1 = \mathbf{G}^{-1}\vec{\hat{a}}_1$ and $\vec{a}_2 = \mathbf{G}^{-1}\vec{\hat{a}}_2$. In other words, the images are identical provided the projections are related by:

$$\begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \mu \\ 0 & 1 & \nu \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \hat{a}_{11} \\ \hat{a}_{12} \\ \hat{a}_{13} \end{pmatrix}$$
$$\begin{pmatrix} a_{21} \\ a_{22} \\ a_{23} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \mu \\ 0 & 1 & \nu \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \hat{a}_{21} \\ \hat{a}_{22} \\ \hat{a}_{23} \end{pmatrix}. \qquad (8)$$

This result can be summarized in the following statement: *If two objects are related by a GBR transform then for any viewpoint and illumination conditions of one object there exist viewpoint and illumination conditions for the second object so that the images of the (visible) portions of both objects have identical intensities.* The word "visible" is used to point out that as the observer's viewpoint changes there will be points on the object which are no longer visible (and our result will not apply to such points). Mathematically, the two objects are $\vec{b}(x,y), \vec{\hat{b}}(x,y)$ with $\vec{\hat{b}}(x,y) = \mathbf{G}\vec{b}(x,y)$. The light sources $\vec{s}, \vec{\hat{s}}$ are related by $\vec{\hat{s}} = \mathbf{G}^{T,-1}\vec{s}$ and the (affine) projection coefficients $\vec{a}_1, \vec{a}_2, \vec{\hat{a}}_1, \vec{\hat{a}}_2$ are related by $\vec{a}_1 = \mathbf{G}^{-1}\vec{\hat{a}}_1$ and $\vec{a}_2 = \mathbf{G}^{-1}\vec{\hat{a}}_2$.

We now obtain a related result which shows that the images of two objects (related by a GBR transform in world space) can always be made identical by affine warps in the image plane provided the projections satisfy certain limited requirements. In particular, this applies when the projections are orthographic.

Our goal is to show that we can always find affine parameters A, B, C, D such that:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}, \qquad (9)$$

for all $x, y, \hat{x}, \hat{y}$.

By using the previous equations, we can express these conditions as:

$$\vec{\hat{a}}_1 = A\mathbf{G}\vec{a}_1 + B\mathbf{G}\vec{a}_2$$
$$\vec{\hat{a}}_2 = C\mathbf{G}\vec{a}_1 + D\mathbf{G}\vec{a}_2. \qquad (10)$$

The point is that we can always solve these equations for A, B, C, D provided that $\vec{\hat{a}}_1, \vec{\hat{a}}_2$ both lie in the two-dimensional space spanned by $\mathbf{G}\vec{a}_1$ and $\mathbf{G}\vec{a}_2$.

In particular, we can solve these equations if the projections are restricted to being orthographic (i.e. $\vec{a}_1$ and $\vec{a}_2$ are orthogonal unit vectors as are $\vec{\hat{a}}_1$ and $\vec{\hat{a}}_2$).

V. The effect of Generic Views on the GBR Transform

How does the generic viewpoint assumption affect the conclusions of the previous sections?

Suppose our task is to estimate which of the objects is present. These objects are related by a GBR but the illumination conditions are unknown. We can therefore write the probability of the observed image $\{I(x,y)\}$ as $P(\{I(x,y)\}|O, \vec{s})$ where $\vec{s}$ is the light source direction (it is straightforward to generalize the argument to include multiple light sources). Then the Bayesian procedure for estimating the probability of the data conditioned on the object, $P(\{I(x,y)\}|O)$, requires specifying a prior distribution for the light source $\vec{s}$ and integrating it out, see [4]. Because it is unclear what prior distribution to place on a light source we will assume a uniform prior. This reduces to the equation:

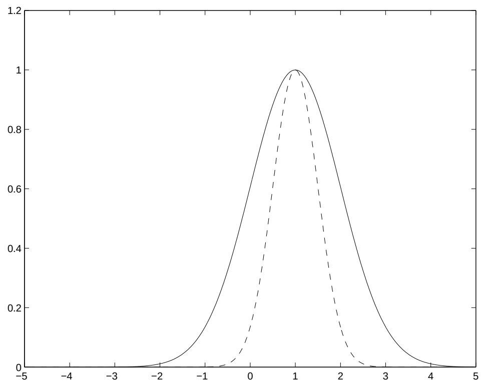

Figure 4: GBR symmetry breaking. $P(I|O,s)$ is shown in solid line, $P(I|\hat{O},s)$ in dashed line as function of s. (For simplicity I and s are scalars.) The image measurement I is equally likely for object O and $\hat{O}$ assuming that $s=1$, i.e. $P(I|O,s=1)=P(I|\hat{O},s=1)$. However, if we integrate over s we find that $P(I|O)>P(I|\hat{O})$, i.e. there is more evidence for object O if we assume only a uniform prior on s. In other words, the image measurement I is more stable for object O than $\hat{O}$.

$$P(\{I(x,y)\}|O)=\int d\vec{s}P(\{I(x,y)\}|O,\vec{s}). \quad (11)$$

As shown by Freeman [4] integrating over the light source direction helps prevent accidental views. Accidental views correspond to precise alignments of the light source and are very sensitive to small fluctuations in lighting direction. They will therefore contribute little to the integral. By contrast, views with a large amount of phase space contribute far more, see figure (4).

We write the illumination equation as $I(x,y)=\max\{\vec{b}(x,y)\cdot\vec{s},0\}$, where the maximum operation removes points x where $\vec{b}(x,y)\cdot\vec{s}\leq 0$ (these correspond to attached shadows). To allow for cast shadows, we also set the contribution from light source $\vec{s}$ to be zero at a point x if the light is blocked in reaching that point. We define a cast shadow function $f(x,y;\vec{s})$ which is zero if point x lies in a cast shadow under lighting condition $\vec{s}$ and equals 1 otherwise.

We assume a probability model for generating the image:

$$P(\{I(x,y)\}|O,\vec{s})=\frac{1}{Z}\,e^{-\sum_{x,y}G(I(x,y),f(x,y:\vec{s})\max\{\vec{b}(x,y)\cdot\vec{s}\})}, \quad (12)$$

where $G(.,.)$ is any function. One possibility to to set $G(a,b)=(a-b)^2$ in which case we obtain the standard Gaussian imaging model.

The probability, $P(\{I(x,y)\}|O)$, for model O is given by the integral:

$$K[\{I(x,y)\}]=\int d\vec{s}P(\{I(x,y)\}|O,\vec{s}) =\int d\vec{s}\frac{1}{(2\pi\sigma^2)^{(N/2)}} e^{-\sum_{x,y}G(I(x,y),f(x,y:\vec{s})\max\{\vec{b}(x,y)\cdot\vec{s},0\})}. \quad (13)$$

It is impossible to calculate this integral analytically. But we do not need to! We only need to compare its value to that of the evidence for model $\hat{O}$ which is related to O by a GBR transform. This can be done by observing that to compute the evidence for $\hat{O}$ we merely have to replace $\vec{b}(x,y)$ by $\hat{\vec{b}}(x,y)$ in the exponent. These are related by a GBR $\hat{\vec{b}}(x,y)=\mathbf{G}\vec{b}(x,y)$. Now we perform a change of variables so that $\hat{\vec{s}}=\mathbf{G}^{T,-1}\vec{s}$. With this change of variables the exponent is now the same whether we are computing the integral for model O or $\hat{O}$! But changing the variables means that we have to introduce a Jacobian factor in the integral. The factor, of course, is simply $|\det\mathbf{G}|=1/\lambda$. So, the difference in evidence between the two models is given only by this factor and we prefer surfaces $\hat{f}(u,v)=\lambda f(u,v)$ with λ as small as possible. Therefore there will be a tendency to favour "flatter" surfaces. Why? Well, intuitively if the object is flat then its appearance is based on its albedo and is largely independent of lighting conditions – so it is very stable under lighting changes.

It is straightforward to modify the previous results to allow for many illumination sources $\vec{s}_1,\vec{s}_2,...,\vec{s}_M$. The only difference is, after performing the change of variables, that we get a factor of $|\det\mathbf{G}|^M$ instead of $|\det\mathbf{G}|$.

What happens if we have a prior distribution on the objects? This does not alter the conclusions greatly. We integrate $P(\{I(x,y)\}|O,\vec{s})P(\{\vec{b}(x,y))\})$ with respect to $\vec{s}$. The point is that the prior $P(\{\vec{b}(x,y))\})$ is independent of $\vec{s}$ and so can be taken outside the integral. This gives:

$$\log\{P(\{I(x,y)\}|O)P(\{\vec{b}(x,y))\})\}=-\log|\det\mathbf{G}| +\log K[\{I(x,y)\}]+\log P(\{\vec{b}(x,y))\}). \quad (14)$$

Recall that the term $\log K[\{I(x,y)\}]$ is independent of the GBR. So the two important terms are the first term which encourages the object to be flat and the final term which pulls it towards the prior. This means that the interpretation is pulled towards the most probable *a priori* interpretation desired by the prior and the flat interpretation determined by the generic factor. The conclusion is that there is an overall bias towards flatter objects unless the prior is incredibly strongly peaked (i.e. almost a delta function).

Overall, we see that integrating over the lighting conditions "breaks" the GBR ambiguity. Observe, moreover, that it induces a bias towards surfaces which are flat which is against the spirit of bas relief in art where one tries to use a pattern with only small relief to substitute for a pattern with large relief. (Of course, this effect is not very strong if the scaling λ of the transformation is close to one). We suggest that the effectiveness of bas relief is because of prior expectations on shape and albedo. Such priors can only be partially effective, however.

VI. A Special Case

To further illustrate our approach we now consider a special case which is motivated by Brainard and Freeman's analysis [2] where they study the use of the generic view assumption to resolve the ambiguity for estimating b from data I when there is a secondary variable s. Mathematically, they studied the distribution $P(I|s,b)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(I-sb)^2/(2\sigma^2)}$ (where I,s,b are scalars), see figure (5). This equation would appear to have an ambiguity given by $s\mapsto\lambda s$ and $b\mapsto(1/\lambda)b$ but Brainard and Freeman show that there is a unique solution for b when the secondary variable s is integrated out.

We apply a generalization of their analysis to the GBR ambiguity. In both cases there is a symmetry which is broken by the generic view assumption (i.e. when one integrates out the secondary variables). We ignore cast and attached shadows and hence use the Lambertian equation $I(x,y) = \vec{b}(x,y) \cdot \vec{s}$.

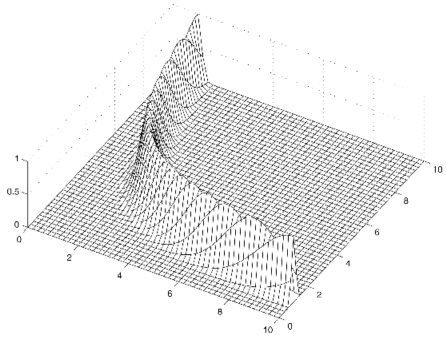

Figure 5: Brainard and Freeman analyzed the use of the generic assumption for the probability distribution $P(I|s,b) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(I-sb)^2/(2\sigma^2)}$. We plot this as a function of s, b, shown for $I = 10.0$. If the secondary variable s is integrated out then there becomes a unique best estimate for b.

We assume that the imaging model introduces independent Gaussian noise. The probability models are therefore:

$$P(\{I(x,y)\}|O,\vec{s}) = \frac{1}{(2\pi\sigma^2)^{(N/2)}} e^{-\sum_x \{I(x,y) - \vec{b}(x,y)\cdot\vec{s}\}^2/(2\sigma^2)}, \quad (15)$$

$$P(\{I(x,y)\}|\hat{O},\vec{\hat{s}}) = \frac{1}{(2\pi\sigma^2)^{(N/2)}} e^{-\sum_x \{I(x,y) - \vec{\hat{b}}(x,y)\cdot\vec{\hat{s}}\}^2/(2\sigma^2)}. \quad (16)$$

To determine the evidence for each model we must integrate out over the lighting configurations $\vec{s}$ and $\vec{\hat{s}}$. Each of the likelihood functions can be re-expressed in form:

$$P(\{I(x,y)\}|O,\vec{s}) \frac{1}{(2\pi\sigma^2)^{(N/2)}} e^{-\{\sum_{i,j=1}^{3} T^{ij} s^i s^j - 2\sum_{i=1}^{3} s^i \phi^i + \psi\}/(2\sigma^2)}, \quad (17)$$

where $T^{ij} = \sum_x b^i(x,y) b^j(x,y)$, $\phi^i = \sum_x b^i(x,y) I(x,y)$, and $\psi = \sum_x \{I(x,y)\}^2$. We are, of course, using the indices i,j to label the three spatial components of the vectors $\vec{b}(x,y)$.

The likelihood function is now quadratic in the variables $\vec{s}^i$ that we wish to integrate over. This integral can therefore be done by standard methods of completing the square. The result is given by:

$$\int d\vec{s} P(\{I(x,y)\}|O,\vec{s}) = \frac{1}{(2\pi\sigma^2)^{(N/2)}} \frac{(2\pi\sigma^2)^{3/2}}{|\det \mathbf{T}|^{1/2}} e^{-\{\psi - \sum_{i,j=1}^{3} T^{-1\ ij} \phi^i \phi^j\}/(2\sigma^2)}. \quad (18)$$

A similar result can be obtained for integrating out $P(\{I(x,y)\}|\hat{O},\vec{\hat{s}})$ with respect to $\vec{\hat{s}}$. It yields a similar formula with T, ϕ replaced by $\hat{T}, \hat{\phi}$ where $\hat{T}^{ij} = \sum_x \hat{b}^i(x,y)\hat{b}^\mu(x,y)$, $\hat{\phi}^i = \sum_x \hat{b}^i(x,y) I(x,y)$ (the number $\psi = \sum_x \{I(x,y)\}^2$ is the same for both cases). To relate these results we recall that $b^i(x,y) = \sum_{j=1}^{3} G^{ij}\hat{b}^j(x,y),\ \forall\ x$ (G^{ij} are the components of the GBR transform matrix $\mathbf{G}$). This leads to the relations $\hat{\phi}^i = \sum_{j=1}^{3} G^{ij}\phi^j$ and $\hat{T}^{ij} = \sum_{\rho,\tau=1}^{3} G^{i\rho} G^{j\tau} T^{i\rho}$. It is then straightforward algebra to check that

$$\sum_{i,j=1}^{3} T^{-1\ ij} \phi^i \phi^j = \sum_{i,j=1}^{3} \hat{T}^{-1\ ij} \hat{\phi}^i \hat{\phi}^j |\det \mathbf{T}| = |\det \mathbf{G}|^2 |\det \hat{\mathbf{T}}|. \quad (19)$$

It is also straightforward algebra to determine that $\psi - \sum_{i,j=1}^{3} T^{-1\ ij}\phi^i\phi^j = \min_{\vec{s}} \sum_x \{I(x,y) - \vec{b}(x,y)\cdot\vec{s}\}^2$. We define this to be $E_{min}[\{I(x,y)\}]$. (Similar results apply for the second model). This gives:

$$\int d\vec{s} P(\{I(x,y)\}|O,\vec{s}) = \frac{1}{(2\pi\sigma^2)^{(N/2)}} \frac{(2\pi\sigma^2)^{3/2}}{|\det \mathbf{T}|^{1/2}} e^{-E_{min}[\{I(x,y)\}]/(2\sigma^2)}$$

$$\int d\vec{\hat{s}} P(\{I(x,y)\}|\hat{O},\vec{\hat{s}}) = \frac{1}{(2\pi\sigma^2)^{(N/2)}} \frac{(2\pi\sigma^2)^{3/2}}{|\det \mathbf{G}||\det \mathbf{T}|^{1/2}} e^{-E_{min}[\{I(x,y)\}]/(2\sigma^2)}. \quad (20)$$

So we see that, after integration, the two hypotheses are not equally likely. The difference is the factor $|\det \mathbf{G}|$ in the denominator. This says that of two hypotheses $\vec{b}(x,y), \vec{\hat{b}}(x,y)$ related by $\vec{\hat{b}}(x,y) = \mathbf{G}\vec{b}(x,y)$ we prefer $\vec{b}(x,y)$ if $|\det \mathbf{G}| > 1$ and $\vec{\hat{b}}(x,y)$ otherwise. Now the determinant of a GBR is given by λ where the transformation scales the z-axis by λ. So if $\hat{O}$ is enlarged in the z direction (i.e. $\lambda > 1$ relative to O, then we prefer O. So of the two possible hypotheses we prefer the most flattened one.

VII. Coupled GBR and Viewpoint Transformations

We now consider integrating out over viewpoint as well as over lighting conditions. The analysis is similar to that for the case where we only considered illumination. We have two objects related by a GBR transform in the world coordinate system and we do not know the viewpoint or lighting. To get evidence for the first object O we should integrate over the viewpoint vectors $\vec{a}_1, \vec{a}_2$ as well as over the light source directions. Similarly for object $\hat{O}$ we must integrate over the viewpoint directions $\vec{\hat{a}}_1, \vec{\hat{a}}_2$ as well as the lighting. As before, we use change of variables to factor out the integration over the $\vec{a}_1, \vec{a}_2$ variables. Because these are related to the $\vec{\hat{a}}_1, \vec{\hat{a}}_2$ by a GBR the effect is to introduce the determinant of the GBR transform squared, i.e. $\det \mathbf{G}^2$.

However integrating over the viewpoint directions presupposes a uniform prior distribution on these variables. It is not clear that this is a reasonable prior to have. In the case of orthographic projection the "natural" prior is to assume that all views of the object are equally likely. But it is not clear how to generalize this to the affine camera case. This is a topic for further research.

VIII. Conclusion

The purpose of this paper was firstly to extend the GBR transform to allow for different viewpoints. We demonstrated how this could be done by using image warping. This enables us to generalize the GBR to deal with situations where objects are viewed from different viewpoints (as well as under different lighting conditions).

The second point was to analyze how the generic viewpoint assumption interacted with the GBR. Our results show that generic viewpoint constraint disambiguates the GBR by biasing towards a flat surface. This is perhaps unfortunate since it suggests that the surface should be perceived as being as flat as possible! We observe, however, that the bias towards a flat plane induced by the generic viewpoint assumption is fairly weak except for surfaces which have large depth variations. Moreover, we suggest that this increases the need for prior probabilities in order to estimate depth. Unlike the hopes expressed in [4] we think it unlikely that use of generic view constraint can significantly reduce the dependence on priors.

In a more technical sense, we have extended the work of Brainard and Freeman [2] by demonstrating that symmetries in visual perception can be disambiguated by using the generic viewpoint assumption or, more abstractly, by integrating out secondary variables (such as the light source directions). Our results make use of the underlying symmetries of the problem, which lead to changes of variables in the integrands, and do not require that the integration be explicitly computed.

Acknowledgments

We thank Suzanne McKee, Christopher Tyler and Andrew Zisserman for helpful discussions.

References

[1] P.N. Belhumeur, D. Kriegman, and A.L. Yuille. "The Generalized Bas Relief Ambiguity". *International Journal of Computer Vision.* 35 (1), pp 33-44. 1999.

[2] D.H. Brainard and W.H. Freeman. "Bayesian Method for Recovering Surface and Illuminant Properties from Photosensor Responses". *Proceedings of SPIE.* Vol. 2719, San Jose, CA. 1994.

[3] O. Faugeras. "Stratification of 3-D vision: Projective, affine, and metric representations". *Journal of the Optical Society of America* JOSA-A. Vol 12(7), pp 465-484. 1995.

[4] W.T. Freeman. "The Generic Viewpoint Assumption in a Framework for Visual Perception". *Nature.* Vol 368. pp 542-545. 1994.

[5] J.J. Koenderink and A. Van Doorn. "Affine Structure from Motion". *Journal of the Optical Society of America* JOSA-A. Vol. 8(2), pp 377-385. 1991.

[6] R. Rosenholtz and J. Koenderink. "Affine Structure and Photometry". In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition.* CVPR'96. San Francisco. 1996.

[7] A.L. Yuille, M. Ferraro, and T, Zhang. "Image Warping for Shape Recovery and Recognition". *Computer Vision and Image Understanding.* Vol. 72. No. 3. pp 351-359. 1998.

[8] A.L. Yuille, D. Snow, R. Epstein, P. Belhumeur. "Determining Generative Models for Objects Under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability". *International Journal on Computer Vision.* 35(3), pp 203-222. 1999.