The Generic Viewpoint Assumption and Planar Bias

A.L. Yuille, *Member*, *IEEE*, James M. Coughlan, and S. Konishi

Abstract—We show that generic viewpoint and lighting assumptions resolve standard visual ambiguities by biasing toward planar surfaces. Our model uses orthographic projection with a two-dimensional affine warp and Lambertian reflectance functions, including cast and attached shadows. We use uniform priors on nuisance variables such as viewpoint direction and the light source. Limitations of using uniform priors on nuisance variables are discussed.

Index Terms—Generic viewpoint, Bayesian inference, visual ambiguities.

1 Introduction

THE generic viewpoint assumption has been suggested as a way to resolve visual ambiguities [4], [11], [7] and has been used to explain perceptual phenomena, e.g., [1] and references cited. The idea is that some interpretations of an image correspond to accidental views and are unstable, in the sense that small changes in the viewing position would induce large changes in the image. The generic viewpoint assumption favors those image interpretations which are stable with respect to small changes of viewpoint. This generic assumption can be extended to apply to other variables such as lighting.

A precise mathematical formulation of the generic viewpoint assumption was proposed by Freeman [7]. He formulated the interpretation problem as probabilistic inference, where the viewpoint direction is treated as a nuisance parameter to be integrated out [12]. More recently, Weinshall and Werman [13] analyzed the generic viewpoint assumption using a different formulation than Freeman's. They represented objects as point features and showed that the assumption causes a bias toward planar objects. They hypothesized that this planar bias would also hold for Freeman's formulation.

In this paper, we examine the effect of the generic viewpoint assumption for resolving visual ambiguities using Freeman's formulation. In particular, we examine those shape and shading ambiguities which have been found by analyzing the geometry of point set features [10], [6], [9] and the photometric properties of objects with Lambertian reflectance functions [2], [17]. We use uniform priors on nuisance variables, such as viewpoint and lighting, and discuss later the limitation of these priors. Our results show that there is a bias toward planar surfaces when the generic assumption is used for viewpoint, lighting, or a combination of both. This proves Weinshall and Werman's hypothesis and goes further by including shading and shadowing effects.

Like Weinshall and Werman [13], we use orthographic projection and allow for two-dimensional affine warps on the image plane. We treat the affine warps either as nuisance parameters to be integrated out, or as quantities to be estimated (both approaches yield a planar bias if the warps have uniform priors). These two treatments correspond to alternative ways to think of the warps. The warps could, for example, correspond to the parameters of an affine camera [9], which motivates integrating them out. Alternatively, estimating the warps leads to an affine invariant measure of similarity between images as advocated by Werman and Weinshall [14]. In either case, the affine warp can be justified by: 1) assuming that the camera parameters are only approximately known and/or 2) modifying the orthographic projection equations to allow for perspective effects [9].

We first give the probabilistic formulation of the generic viewpoint assumption in Section 2, define the ambiguities we will be dealing with in Section 3, prove our results in Section 4, and then close with a discussion in Section 5.

• *The authors are with the Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115.*
E-mail: {yuille, coughlan, konishi}@ski.org.

Manuscript received 29 Apr. 2002; revised 10 Oct. 2002; accepted 13 Oct. 2002. Recommended for acceptance by W. Freeman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 116428.

2 Visual Ambiguities and the Generic Viewpoint Assumption

This section describes the mathematical framework for visual ambiguities and the generic viewpoint assumption. The framework is general and applies to any probabilistic estimation problem [12].

We assume that the image formation process is specified by a likelihood function $P(I|O,h)$, where I is the observed image, O is the object being viewed, and h is a nuisance variable (e.g., viewpoint or lighting).

Visual ambiguities arise when there are many different ways of generating the same image. For example, if $P(I|O,h) = P(I|\hat{O},\hat{h})$, then it seems difficult to distinguish between O,h and $\hat{O},\hat{h}$. A large class of visual ambiguities (see Section 3) correspond to a group of transformations that can be made on an object and the nuisance variable. For example, suppose we have $P(I|O,h) = P(I|f_t(O), f_t(h))$, where $f_t(.)$ is an element of a group of transformations on the object O and nuisance parameter h, and t indexes the group element. Then, we state that the likelihood function is *invariant* to the group of transformations $\{f_t(.)\}$. Much of the work on visual ambiguities, e.g., [10], [6], [2], [9], [17], assumes that the image formation model is purely deterministic. This special case can be obtained from our formulation by setting the likelihood function to be a delta function (e.g., $P(I|O,h) = \delta(I - F(O,h))$ for some function $F(.,.)$, where $\delta(.)$ is the Dirac delta function). But, these ambiguities will also remain even if we allow for noise in the imaging model, see Section 4.

The Generic Viewpoint Assumption (GVA) [4], [11], [7] is a method for resolving these ambiguities. First, the problem is expressed as Bayesian estimation by placing prior distributions $P(O), P(h)$ on O and h. The task of estimating O and h from I can be formulated as Bayesian inference using the posterior distribution:

$$P(O,h|I) = \frac{P(I|O,h)P(O)P(h)}{\int d\hat{h}\, d\hat{O}\ P(I|\hat{O},\hat{h})P(\hat{O})P(\hat{h})}. \tag{1}$$

Freeman's proposal is to estimate O alone after integrating out the nuisance parameter h. This corresponds to the standard procedure for dealing with nuisance variables in statistics [12]. It reduces to estimating O from:

$$P(O|I) = \int dh\ P(O,h|I). \tag{2}$$

Freeman's insight [7] is that *the integration over h is often sufficient to resolve many visual ambiguities even if the prior distributions on O and h are uniform*. (For these priors, the posterior distribution $P(O,h|I)$ is ambiguous if the likelihood is, i.e., $P(O,h|I) = P(f_t(O), f_t(h)|I)$ provided $P(I|O,h) = P(I|f_t(O), f_t(h))$). If uniform priors are assumed, then the generic viewpoint assumption reduces to simply integrating the likelihood function to obtain $P(O|I) = (1/Z)\int dh\ P(I|O,h)$, where Z is a constant and solving for:

$$O^* = \arg\max_O \int dh P(I|O,h). \tag{3}$$

To understand how the GVA works using uniform priors, suppose we have an ambiguity so that $P(I|O,h) = P(I|f_t(O), f_t(h))$. We calculate:

$$P(I|O) = \frac{1}{Z}\int dh\; P(I|O,h),$$
$$P(I|f_t(O)) = \frac{1}{Z}\int d\hat{h} P(I|f_t(O),\hat{h}) = \frac{1}{Z}\int dh\left|\frac{\partial f}{\partial h}\right| P(I|O,h), \quad (4)$$

where we have used the substitutions $\hat{h} = f_t(h)$ and $P(I|O,h) = P(I|f_t(O), f_t(h))$ to make the integrands of $P(I|O)$ and $P(I|f_t(O))$ as similar as possible. The only difference in the integrands is the factor $|\frac{\partial f}{\partial h}|$, which is the Jacobian of the transformation on the nuisance variable, $\hat{h} = f_t(h)$. This Jacobian term is the mechanism by which the GVA resolves the ambiguity.

This analysis has assumed uniform prior distributions. This is attractive because these priors embody complete ignorance about O and h. Moreover, this seems a natural choice of prior for the viewpoint, the light source, and other nuisance variables. However, the same analysis can be repeated if the priors are nonuniform. In this case, the posterior *before integration*, $P(O,h|I)$, will be unambiguous unless the prior *densities*$P(O), P(h)$ are invariant to the ambiguities (which, of course, the uniform distribution is). But, the posterior *after integration*, $P(O|I)$, will be unambiguous unless the prior *distributions* are invariant, $P(f(h))|\frac{\partial f}{\partial h}|dh = P(h)dh$.

In the next few sections, we will analyze the GVA for a large class of ambiguities assuming uniform priors on viewpoint and light source direction. For these ambiguities, we will show that the Jacobian factor $|\frac{\partial f}{\partial h}|$ will cause a bias to planar surfaces.

3 Geometric and Photometric Ambiguities

There is a well-known ambiguity in the estimation of shape corresponding to an affine transformation on the shape in conjunction with a transformation on the viewpoint and camera parameters [10], [6], [9]. More recently, this ambiguity was extended to include photometric effects by additional transformations on the object albedo and the lighting conditions [2], [17]. These photometric effects include shadows and shading modeled by Lambertian reflectance functions. We refer to the combined transformation on shape and albedo as a KGBR transformation [17]. These transformations form a group.

More formally, suppose we represent the surface and albedo of an object O by $\{\vec{r}(\vec{u}), a(\vec{u}) : \vec{u} \in U\}$, where $\vec{r}(\vec{u})$ represents surface position, $a(\vec{u})$ is the surface albedo, and $\vec{u} \in U$ are coordinates on the surface of the object. Let the viewpoint be specified by a unit vector $\vec{v}$ and the lighting by a point source $\vec{s}$.

The KGBR assumes that the reflectance function is Lambertian, which implies that the intensity of a surface point $\vec{u}$ is independent of the viewpoint and given by $I(\vec{u}) = \max\{0, a(\vec{u})\vec{n}(\vec{u})\cdot\vec{s}\}$. The key point of the KGBR is that we can keep the intensity $I(\vec{u})$ constant as we transform the geometry of the surface by an affine transformation provided we *also transform the albedo and the lighting* [2], [17]. The intensity can be thought of as being painted onto the surface (because Lambertian reflectance is independent of viewpoint) and, so, the standard affine ambiguities [10], [6], [9] can be extended to this photometric case.

A KGBR transformation [17] is specified by a three-dimensional matrix $\mathbf{K}$. It transforms an object O, $\{\vec{r}(\vec{u}), a(\vec{u}) : \vec{u} \in U\}$, to an object $\hat{O}$, $\{\vec{\hat{r}}(\vec{u}), \hat{a}(\vec{u}) : \vec{u} \in U\}$, and transforms the light source and viewpoint from $\vec{s}, \vec{v}$ to $\vec{\hat{s}}, \vec{\hat{v}}$, where:

$$\vec{\hat{r}}(\vec{u}) = \mathbf{K}\vec{r}(\vec{u}),\; \vec{\hat{n}}(\vec{u}) = \frac{\mathbf{K}^{-1,T}\vec{n}(\vec{u})}{|\mathbf{K}^{-1,T}\vec{n}(\vec{u})|},\; \forall\, \vec{u} \in U,\; \vec{\hat{v}} = \frac{\mathbf{K}\vec{v}}{|\mathbf{K}\vec{v}|}, \quad (5)$$

$$\hat{a}(\vec{u}) = a(\vec{u})|\mathbf{K}^{-1,T}\vec{n}(\vec{u})||\det\mathbf{K}|,\; \forall\, \vec{u} \in U,\; \vec{\hat{s}} = \frac{\mathbf{K}\vec{s}}{|\det\mathbf{K}|}. \quad (6)$$

Equation (5) gives the geometric transformation on the object shape and viewpoint (the transformation on the surface normals $\vec{n}$ is induced by the transformation on the surface points $\vec{r}$). Equation 6 gives the transformation on the albedo and light source. The transformed viewpoint $\vec{\hat{v}}$ is a unit vector. Observe that $\max\{0, a(\vec{u})\vec{n}(\vec{u})\cdot\vec{s}\} = \max\{0, \hat{a}(\vec{u})\vec{\hat{n}}(\vec{u})\cdot\vec{\hat{s}}\}$ and, so, the intensity at each point $\vec{u}$ is preserved by the transformation.

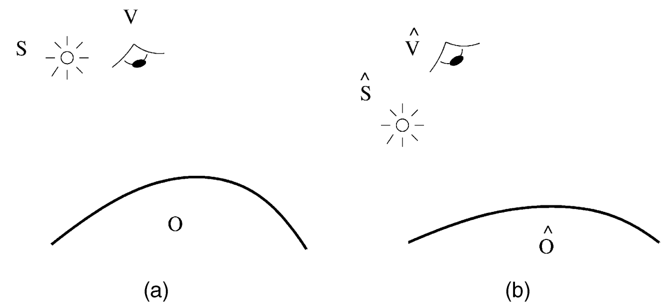

Fig. 1. Object O (left panel) with lighting conditions $\vec{s}$ seen from viewpoint $\vec{v}$ gives an identical image (up to affine warp) of object $\hat{O}$ (right panel) under lighting $\vec{\hat{s}}$ viewed from direction $\vec{\hat{v}}$.

We define $\vec{x}(\vec{u}; O, \vec{v})$ to be the orthographic projection of point $\vec{u}$ on the surface of object O to the image plane when viewed from viewpoint $\vec{v}$. Similarly, we define $\vec{u}(\vec{x}; O, \vec{v})$ to be the point $\vec{u}$ on the surface of object O, which projects to point $\vec{x}$ in the image under viewpoint $\vec{v}$.

Then, the geometry of orthographic projection implies that $\vec{x}(\vec{u}; \hat{O}, \vec{\hat{v}}) = \mathbf{A}\vec{x}(\vec{u}; O, \vec{v})$, provided that $O, \vec{v}$ and $\hat{O}, \vec{\hat{v}}$ are related by (5), where $\mathbf{A}$ is two-dimensional matrix, which is a function of $\mathbf{K}$. It can be shown [17] that $\det\mathbf{A} = \frac{|\det\mathbf{K}|}{|\mathbf{K}\vec{v}|}$ (intuitively: the transformation $\mathbf{K}$ on the space induces a change, $\det\mathbf{A}$, in the area of the image plane which equals the change in volume of space, $|\det\mathbf{K}|$, divided by the contraction, $|\mathbf{K}\vec{v}|$, in the viewing direction).

We define the images of the two objects to be I_{syn} and $\hat{I}_{syn}$ using a Lambertian reflectance function with cast and attached shadows:

$$I_{syn}(\vec{x}; O, \vec{v}, \vec{s}) = f(\vec{x}; O, \vec{v}, \vec{s})\max\{0, a(\vec{u}(\vec{x}; O, \vec{v}))\vec{n}(\vec{u}(\vec{x}; O, \vec{v}))\cdot\vec{s}\},$$
$$\hat{I}_{syn}(\vec{x}; \hat{O}, \vec{\hat{v}}, \vec{\hat{s}}) = f(\vec{x}; \hat{O}, \vec{\hat{v}}, \vec{\hat{s}})\max\{0, \hat{a}\Big(\vec{u}(\vec{x}; \hat{O}, \vec{\hat{v}})\Big)\vec{\hat{n}}\Big(\vec{u}(\vec{x}; \hat{O}, \vec{\hat{v}})\Big)\cdot\vec{\hat{s}}\}, \quad (7)$$

where $f(\vec{x}; O, \vec{v}, \vec{s})$ is a binary-valued function which takes value 0 if point $\vec{x}$ is in shadow and takes value 1; otherwise, see analysis in [2], [17].

Then, it follows from (5) and (6) that $\hat{I}_{syn}(\mathbf{A}\vec{x}) = I_{syn}(\vec{x})$ for all points $\vec{x}$ in the image plane and, hence, the images of objects O and $\hat{O}$ are identical up to the affine transformation $\mathbf{A}$ [17]. This is illustrated in Fig. 1.

4 The Generic Viewpoint and Lighting Assumption

We express the problem of estimating the object as probabilistic inference. We define a generative model $P(I|O, \vec{v}, \vec{s}, \mathbf{W})$, where $\mathbf{W}$ represents the two-dimensional affine warp, by:

$$P(I|O, \vec{v}, \vec{s}, \mathbf{W}) = F(\{I(\vec{x})\}, \{I_{syn}(\mathbf{W}\vec{x}; O, \vec{v}, \vec{s})\}), \quad (8)$$

for some function $F(.,.)$. Equation (8) takes the synthesized image, see (7), then adds noise (linearly or nonlinearly), and relates it to the observed image I by an affine warp $\mathbf{W}$. One possibility is $F(\{I(\vec{x})\}, \{I_{syn}(\vec{x})\}) = \prod_{\vec{x}}\delta\big(I(\vec{x}) - I_{syn}(\vec{x})\big)$, which is the imaging model when there is no noise and no warp.

The affine image warp $\mathbf{W}$ is required in order for the affine geometry ambiguities to exist [10], [6], [9] and is also required for the generic viewpoint analysis of Weinshall and Werman [13]. Following these theories, we also make no prior assumptions about the warp and, in our probabilistic terminology, let it have a uniform prior distribution (or, equivalently, we put no prior on $\mathbf{W}$ and estimate it by maximum likelihood). The use of a uniform prior for variables such as $\mathbf{W}$ is discussed in Section 5.

We formulate two versions of the generic assumption which differ in the way they treat the affine warp $\mathbf{W}$. They correspond to estimating the quantities $Q_1(I|O)$ and $Q_2(I|O)$ defined by:

$$Q_1(I|O) = \int d\vec{s}\; d\vec{v}\; d\mathbf{W} P(I|O, \vec{v}, \vec{s}, \mathbf{W}), \tag{9}$$

$$Q_2(I|O) = \int d\vec{s}\; d\vec{v} \max_{\mathbf{W}} P(I|O, \vec{v}, \vec{s}, \mathbf{W}). \tag{10}$$

The definition of $Q_1(I|O)$, see (9), treats $\mathbf{W}$ as a nuisance parameter to be integrated out. For example, one could think of $\mathbf{W}$ as being the (unknown) parameters of an affine camera. $Q_1(I|O)$ is proportional to the probability $P(I|O)$ obtained by placing uniform priors on $\vec{s}, \vec{v}, \mathbf{W}$, computing $P(I, \mathbf{W}|O, \vec{v}, \vec{s})$, and then integrating out $\vec{s}, \vec{v}, \mathbf{W}$.

The definition of $Q_2(I|O)$, given by (10), is motivated by Werman and Weinshall's notion of affine invariant similarity [14], where two images are considered to be identical if their images are the same up to an affine transform $\mathbf{W}$. It can be checked that $Q_2(I|O)$ has this property and, to the author's knowledge, is the only way we can obtain such a quantity from our probabilistic formulation. (Suppose that images I and $\hat{I}$ are related by an affine transform $\hat{\mathbf{W}}$ so that $\hat{I}(\vec{x}) = I(\hat{\mathbf{W}}\vec{x})$. Then, $P(\hat{I}|O, \vec{v}, \vec{s}, \mathbf{W}) = P(I|O, \vec{v}, \vec{s}, \mathbf{W}\hat{\mathbf{W}}^{-1})$ and, so, $Q_2(\hat{I}|O) = Q_2(I|O)$.)

Before proceeding to our main results, we need an intermediate result on how the integration variables $d\vec{v}, d\vec{s}$ in (9) and (10) transform under a KGBR. The variable $\vec{v}$ is a solid angle (because the magnitude of $\vec{v}$ is set to unity). Their transformation is given by the following lemma.

Lemma 1. *If the viewpoint and lighting are related by $\vec{\hat{v}} = \mathbf{K}\vec{v}/|\mathbf{K}\vec{v}|$ and $\vec{\hat{s}} = \mathbf{K}\vec{s}/|\det \mathbf{K}|$, then the solid angles are related by $d\vec{\hat{v}} = \det \mathbf{K} d\vec{v}/|\mathbf{K}\vec{v}|^3$ and $d\vec{\hat{s}} = d\vec{s}/|\det \mathbf{K}|^2$.*

Proof. Let $\vec{v}(\alpha, \beta)$ to be a unit vector representing viewpoint where α, β are coordinates on the unit sphere. The solid angles on the original and transformed viewpoints are $d\vec{v} = |\vec{v}_\alpha \times \vec{v}_\beta| d\alpha d\beta$ and $d\vec{\hat{v}} = |\vec{\hat{v}}_\alpha \times \vec{\hat{v}}_\beta| d\alpha d\beta$, respectively, where subscripts $_{\alpha,\beta}$ denote derivatives with respect to α and β. Because $\vec{v}$ is a unit vector, it follows that $\vec{v}_\alpha \cdot \vec{v} = \vec{v}_\beta \cdot \vec{v} = 0$, so that $\vec{v}_\alpha \times \vec{v}_\beta \propto \vec{v}$ and, hence, $|\vec{v}_\alpha \times \vec{v}_\beta| = \vec{v} \cdot \{\vec{v}_\alpha \times \vec{v}_\beta\}$. Similarly, we compute $|\vec{\hat{v}}_\alpha \times \vec{\hat{v}}_\beta| = \vec{\hat{v}} \cdot \{\vec{\hat{v}}_\alpha \times \vec{\hat{v}}_\beta\}$. By setting $\vec{\hat{v}} = \mathbf{K}\vec{v}/|\mathbf{K}\vec{v}|$, we obtain $\frac{1}{|\mathbf{K}\vec{v}|^3}(\mathbf{K}\vec{v}) \cdot \{\mathbf{K}\vec{v}_\alpha \times \mathbf{K}\vec{v}_\beta\} = \frac{\det \mathbf{K}}{|\mathbf{K}\vec{v}|^3} \vec{v} \cdot \{\vec{v}_\alpha \times \vec{v}_\beta\}$, which proves the first statement. The result for $\vec{s}$ is a direct calculation. □

We now obtain our main result on generic viewpoint and lighting.

Theorem 1. *Let two objects O and $\hat{O}$ be related by a KGBR $\mathbf{K}$ where one eigenvalue of $\mathbf{K}$ is arbitrarily small (corresponding to flattening O in the direction of the corresponding eigenvector). Then $Q_1(I|\hat{O}) \to \infty$ and $Q_2(I|\hat{O}) \to \infty$ as $\det \mathbf{K} \to 0$. Therefore, $Q_1(I|O) << Q_1(I|\hat{O})$ and $Q_2(I|O) << Q_2(I|\hat{O})$, provided $\det \mathbf{K}$ is sufficiently small and, so, the flatter object is preferred and both criteria favor planar surface.*

Proof. Apply a KGBR transform $\mathbf{K}$ to O. Equations (5) and (6) imply that $I_{syn}(\mathbf{W}\vec{x}; O, \vec{v}, \vec{s}) = I_{syn}(\mathbf{A}\mathbf{W}\vec{x}; \hat{O}, \vec{\hat{v}}, \vec{\hat{s}})$. Hence, $P(I|O, \vec{v}, \vec{s}, \mathbf{W}) = P(I|\hat{O}, \vec{\hat{v}}, \vec{\hat{s}}, \mathbf{A}\mathbf{W})$. Using the KGBR transform to change variables, we express $Q_1(I|\hat{O})$ as:

$$\begin{aligned} Q_1(I|\hat{O}) &= \int d\vec{s}\; d\vec{v}\; d\vec{\mathbf{W}} \frac{1}{|\det \mathbf{K}|^2} \frac{\det \mathbf{K}}{|\mathbf{K}\vec{v}|^3} \frac{\det \mathbf{K}}{|\mathbf{K}\vec{v}|} P(I|O, \vec{v}, \vec{s}, \mathbf{W}) \\ &= \int d\vec{s}\; d\vec{v}\; d\vec{\mathbf{W}} \frac{1}{|\mathbf{K}\vec{v}|^4} P(I|O, \vec{v}, \vec{s}, W), \end{aligned} \tag{11}$$

where we have used Lemma 1 to obtain the transformation on the integrands, and we know that the Jacobian from $\mathbf{W}$ to $\hat{\mathbf{W}} = \mathbf{A}\mathbf{W}$ is given by $\det \mathbf{A}$, which equals $\det \mathbf{K}/|\mathbf{K}\vec{v}|$ as discussed in Section 3 and shown in [17].

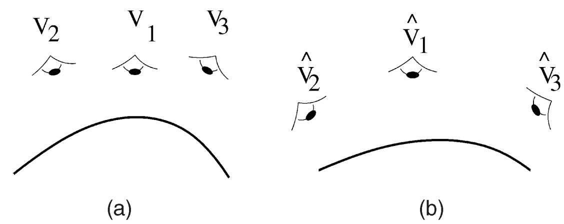

Fig. 2. Small changes of viewpoint of object O (left panel) correspond to large changes of viewpoint for object $\hat{O}$ (right panel). Here, $\vec{v}_1, \vec{v}_2, \vec{v}_3$ correspond to $\vec{\hat{v}}_1, \vec{\hat{v}}_2, \vec{\hat{v}}_3$, respectively.

The difference between the integrands of $Q_1(I|\hat{O})$ and $Q_1(I|O)$, see (9), is the term $1/|\mathbf{K}\vec{v}|^4$ in $Q_1(I|\hat{O})$. This term will be dominated by those values of $\vec{v}$ which are in the direction of the smallest eigenvalue of $\mathbf{K}$. As this eigenvalue tends to zero (i.e., the object gets flattened and $\det \mathbf{K} \to 0$), we see that $Q_1(I|\hat{O}) \to \infty$ and, so, $Q_1(I|\hat{O}) >> Q_1(I|O)$.

The argument is similar for $Q_2(I|O)$. The definitions imply that $\max_{\mathbf{W}} P(I|O, \vec{v}, \vec{s}, \mathbf{W}) = \max_{\hat{\mathbf{W}}} P(I|\hat{O}, \vec{\hat{v}}, \vec{\hat{s}}, \hat{\mathbf{W}})$. Then, we obtain

$$Q_2(I|\hat{O}) = \int d\vec{s}\; d\vec{v} \frac{1}{|\det \mathbf{K}|^2} \frac{\det \mathbf{K}}{|\mathbf{K}\vec{v}|^3} \max_{\mathbf{W}} P(I|O, \vec{v}, \vec{s}, \mathbf{W}), \tag{12}$$

which will also tend to infinity as $\det \mathbf{K}$ tends to zero. Hence, $Q_2(I|\hat{O}) >> Q_2(I|O)$ provided $\det \mathbf{K}$ is sufficiently small. □

Theorem 1 states that the flatter object is far more stable to changes in the lighting and viewpoint. Mathematically, this is due to the Jacobian factors which relate the variables $d\vec{s}, d\vec{v}, d\mathbf{W}$ to $d\vec{\hat{s}}, d\vec{\hat{v}}, d\hat{\mathbf{W}}$, see Section 2. These Jacobian factors dominate the integrals in (11) and (12). Small changes of the viewpoint $\vec{v}$ and lighting $\vec{s}$ for object O will correspond to large changes of the viewpoint $\vec{\hat{v}}$ and lighting $\vec{\hat{s}}$ of the flatter object $\hat{O}$, see Fig. 2. Therefore, the viewpoints and lighting of the second object are far more stable and are preferred by the GVA.

To understand this further, let the two smallest eigenvalues of $\mathbf{K}$ be λ_1, λ_2, with $0 < \lambda_1 << \lambda_2$, corresponding to eigenvectors $\vec{e}_1, \vec{e}_2$. Then, the integrals in (11) and (12) will be dominated by viewpoint $\vec{v}$ close to $\vec{e}_1$. To see what happens as we vary the viewpoint, let $\vec{v} = \cos\theta \vec{e}_1 + \sin\theta \vec{e}_2$ and calculate $\vec{\hat{v}} = \{\vec{e}_1 + (\lambda_2/\lambda_1) \tan\theta \vec{e}_2\}/\{1 + (\lambda_2/\lambda_1)^2 \tan^2\theta\}^{1/2}$. Because $\lambda_1/\lambda_2 << 1$, small changes of viewpoint $\vec{v}$ of the first object (i.e., small θ) correspond to large changes of viewpoint $\vec{\hat{v}}$ of the second object (particularly when $\tan\theta >> \lambda_1/\lambda_2$). Indeed, if θ is small, so that $\vec{v} \approx \vec{e}_1$, *but* $\tan\theta >> \lambda_1/\lambda_2$, then a Taylor series expansion in $(\lambda_1/\lambda_2 \tan\theta)$ yields the viewpoint of the second object to be $\vec{\hat{v}} \approx \vec{e}_2 + (\lambda_1/\lambda_2 \tan\theta)\vec{e}_1 + O(\lambda_1/\lambda_2 \tan\theta)^2$, which is almost perpendicular to $\vec{e}_1$. Therefore, changes in the viewing angle θ of object O correspond to large changes in the viewpoint of object $\hat{O}$.

Now, suppose we ignore the photometric effects and represent objects as isolated feature points and replace $\{\vec{r}(\vec{u}) : \vec{u} \in U\}$ by $\{\vec{r}_i : i = 1, \ldots, N\}$. We replace I_{syn} by an imaging model where we set $I_{syn}(\vec{x}) = \sum_{i=1}^{N} \delta(\vec{x} - \vec{x}(i; \vec{v}))$, with $\vec{x}(i; \vec{v})$ being the projection of the point $\vec{r}_i$ from view $\vec{v}$. We define $\tilde{Q}_1(I|O), \tilde{Q}_2(I|O)$ by modifying (8) to remove dependence on $\vec{s}$, and (9) and (10) to remove the integration with respect to $\vec{s}$.

Theorem 2. *Let two sets of point features O and $\hat{O}$ be related by a matrix $\mathbf{K}$ where one eigenvalue of $\mathbf{K}$ is arbitrarily small (corresponding to flattening O in the direction of the corresponding eigenvector). Then, $\tilde{Q}_1(I|\hat{O}) \to \infty$ and $\tilde{Q}_2(I|\hat{O}) \to \infty$ as $\det \mathbf{K} \mapsto 0$. Hence, $\tilde{Q}_1(I|O) << \tilde{Q}_1(I|\hat{O})$ and $\tilde{Q}_2(I|O) << \tilde{Q}_2(I|\hat{O})$, for sufficiently small $\det \mathbf{K}$, so the flatter object is preferred and the most probable interpretation of the image is a planar surface.*

Proof. Same as for Theorem 1, except we only need to change the variables $\vec{v}$ and $\mathbf{W}$. □

This result is the analog to Weinshall and Werman [13], but obtained using Freeman's formulation of the generic viewpoint assumption.

Now, suppose we fix the viewpoint to be $\vec{v} = (0, 0, 1)$ and allow the lighting to vary. In this case, the KGBR ambiguity [17] reduces to the GBR ambiguity [2]. This means that we can only transform objects by a matrix $\mathbf{K}$ determined by constants (λ, μ, ν) and of form

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda & \mu & \nu \end{pmatrix}.$$

We use the original imaging model with fixed viewpoint, see (7). We use the GBR transforms (i.e., restrict $\mathbf{K}$ to be of the form above), which are performed only on the object geometry, albedo, and light source direction. (There is no need for a two-dimensional affine warp for this case.) We define $\hat{Q}_1(I|O), \hat{Q}_2(I|O)$ by modifying (8) to remove dependence on $\vec{v}$ and $\mathbf{W}$, and (9) and (10) to remove the integration with respect to $\vec{v}$ and $\mathbf{W}$.

Theorem 3. *Let two objects O and $\hat{O}$ be related by a GBR matrix $\mathbf{K}$, where the eigenvalue ν of $\mathbf{K}$ in the direction $(0, 0, 1)$ is arbitrarily small (corresponding to flattening O in this direction). Then, $\hat{Q}_1(I|\hat{O}) \to \infty$ and $\hat{Q}_2(I|\hat{O}) \to \infty$ as $\det \mathbf{K} \to 0$. Hence, for sufficiently small $\det \mathbf{K}$, we have $\hat{Q}_1(I|O) << \hat{Q}_1(I|\hat{O})$ and $\hat{Q}_2(I|O) << \hat{Q}_2(I|\hat{O})$, so the flatter object is preferred.*

Proof. Same as for Theorem 1, except we only need to change the variable $\vec{s}$. □

The three theorems show that the generic assumption causes a bias toward planar interpretations when we take into account the standard ambiguities [10], [6], [2], [9], [17]. This result is fairly intuitive. If an object is flat, then its appearance is determined entirely by its albedo and is independent of the lighting conditions (up to a scale factor) or the viewpoint (up to two-dimensional affine warps). Hence, a flat planar interpretation (e.g., a painting) is the most stable percept.

What happens if we have a prior distribution $P(O)$ on the object shape or albedo? These priors can be taken outside the integrals in (9) and (10) and, after taking logarithms, we obtain $\log P(I|O) + \log P(O)$. The first term will, as we have shown, bias toward planar surfaces. The second term will bias toward surfaces which best satisfy the prior $P(O)$. Therefore, there will remain a tendency to bias perception toward objects that are flatter than the prior. This tendency is smaller the stronger the prior. For example, a prior on face shape will probably resist this flattening tendency better than a more generic prior on surface smoothness.

5 Discussion

Our analysis in this paper shows biases toward planar objects. Interestingly, psychophysical experiments show biases of this type [8]. It is unclear, however, whether such a bias is useful for a machine vision system. It would seem better to leave ambiguities unresolved.

One can question the use of uniform distributions as priors for the generic viewpoint assumption. Uniform distributions are attractive choices to model ignorance. But, probabilists have debated whether lack of knowledge of a variable is best expressed by placing a uniform distribution on it [3]. For example, suppose we wish to estimate depth using binocular stereo. The depth d is inversely proportional to the disparity α and assume that we lack knowledge of both. We can attempt to express this ignorance by placing a uniform prior either on the depth or on the disparity. But, placing a uniform prior on the depth d is not equivalent to placing a uniform prior on the disparity α, see Fig. 3. Implementing the generic viewpoint assumption by placing a uniform distribution on the nuisance variable will give different results depending on how the problem in question is represented.

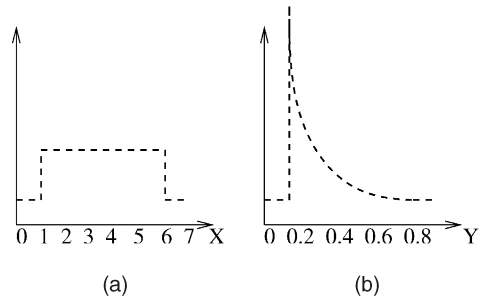

Fig. 3. (a) The uniform prior for a variable x and (b) the corresponding prior for $y = 1/x$ (b). A uniform prior on x is not equivalent to a uniform prior on $1/x$.

Alternative results will be obtained if we use another formulation of the generic view assumption based on decision theory [5], [15]. This introduces the loss function as an additional factor of choice. Results in this case seem to depend on the specific choice of loss function. This goes against the apparent simplicity of the generic viewpoint assumption.

Although the generic viewpoint assumption is very intuitive, there seem to be limitations on what it can achieve. Prior distributions on object shape and properties seems a more reasonable way to resolve ambiguities. Nevertheless, the generic viewpoint assumption has considerable intuitive plausibility and remains a useful guide for choices of priors.

Acknowledgments

This work was supported by the National Institute of Health (NEI) with grant number RO1-EY 12691-01. The authors thank Dan Kersten for helpful discussions.

References

[1] M.K. Albert and D.D. Hoffman, "The Generic-Viewpoint Assumption and Illusory Contours," *Perception,* vol. 29, no. 3, pp. 303-312, 2000.
[2] P.N. Belhumeur, D. Kriegman, and A.L. Yuille, "The Generalized Bas Relief Ambiguity," *Int'l J. Computer Vision,* vol. 35, no. 1, pp. 33-44, 1999.
[3] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis,* second ed. New York: Springer-Verlag, 1985.
[4] T.O. Binford, "Inferring Surfaces from Images," *Artificial Intelligence,* vol. 17, pp. 205-244, 1981.
[5] D.H. Brainard and W.H. Freeman, "Bayesian Method for Recovering Surface and Illuminant Properties from Photosensor Responses," *Proc. SPIE,* vol. 2719, 1994.
[6] O. Faugeras, "Stratification of 3-D Vision: Projective, Affine, and Metric Representations," *J. Optical Soc. Am.,* vol. 12, no. 7, pp. 465-484, 1995.
[7] W.T. Freeman, "The Generic Viewpoint Assumption in a Framework for Visual Perception," *Nature,* vol. 368, pp. 542-545, 1994.
[8] J.J. Gibson, "The Perception of Visual Surfaces," *Am. J. Experimental Psychology,* vol. 63, pp. 367-384, 1950.
[9] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2000.
[10] J.J. Koenderink and A. Van Doorn, "Affine Structure from Motion," *J. Optical Soc. Am.,* vol. 8, no. 2. pp. 377-385, 1991.
[11] D.G. Lowe and T.O. Binford, "The Recovery of Three-Dimensional Structure from Image Curves" *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 7, no. 3, pp. 320-326, 1985.
[12] B.D. Ripley, *Pattern Recognition and Neural Networks.* Cambridge Univ. Press, 1996.
[13] D. Weinshall and M. Werman, "On View Likelihood and Stability," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 2, pp. 97-108, 1997.
[14] M. Werman and D. Weinshall, "Similarity and Affine Invariant Distance Between Point Sets," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 8, pp. 810-814, Aug. 1995.
[15] A.L. Yuille and H.H. Bülthoff, "Bayesian Decision Theory and Psychophysics," *Bayesian Approaches to Perception,* D. Knill and W. Richards, eds., Cambridge Univ. Press, 1996.
[16] A.L. Yuille, J.M. Coughlan, and S. Konishi, "The Generic Viewpoint Constraint Resolves the Generalized Bas Relief Ambiguity" *Proc. Conf. Information Sciences and Systems,* Mar. 2000.
[17] A.L. Yuille, J.M. Coughlan, and S. Konishi, "The KGBR Viewpoint-Lighting Ambiguity and Its Resolution by Generic Constraints," *Proc. Int'l Conf. Computer Vision,* 2001.