The Bas-Relief Ambiguity

Peter N. Belhumeur * David J. Kriegman † Alan L. Yuille

Center for Computational Vision and Control
Yale University
New Haven, CT 06520-8267

Smith-Kettlewell Eye Research Institute
San Francisco, CA 94115

Abstract

Since antiquity, artisans have created flattened forms, often called "bas-reliefs," which give an exaggerated perception of depth when viewed from a particular vantage point. This paper presents an explanation of this phenomena, showing that the ambiguity in determining the relief of an object is not confined to bas-relief sculpture but is implicit in the determination of the structure of any object. Formally, if the object's true surface is denoted by $z_{true} = f(x,y)$, then we define the "generalized bas-relief transformation" as $z = \lambda f(x,y) + \mu x + \nu y$, with a corresponding transformation of the albedo. For each image of a Lambertian surface $f(x,y)$ produced by a point light source at infinity, there exists an identical image of a bas-relief produced by a transformed light source. This equality holds for both shaded and shadowed regions. Thus, the set of possible images (illumination cone) is invariant over generalized bas-relief transformations. When $\mu = \nu = 0$ (e.g. a classical bas-relief sculpture), we show that the set of possible motion fields are also identical. Thus, neither small unknown motions nor changes of illumination can resolve the bas-relief ambiguity. Implications of this ambiguity on structure recovery and shape representation are discussed.

1 Introduction

Throughout the millennia, artisans have created flattened forms, i.e. so-called "bas-reliefs," which when viewed from a particular vantage point are difficult, if not impossible, to distinguish from a full relief sculpture. As the sun moves through the sky, the shading and shadows change, yet the degree of flattening cannot be discerned on a well sculpted bas-relief. Even if an observer's head moves by a small amount, this ambiguity cannot be resolved. This paper not only presents an explanation of this phenomena, but also demonstrates that the ambiguity in determining the relief of an object is implicit in the determination of the structure of any object.

In particular, we show that the set of images produced by arbitrary illumination of an object (i.e. the object's illumination cone [1]) is the same as the set of images produced by what we call a "generalized bas-relief transformation" of the object. A generalized bas-relief transformation is a transformation of both the surface shape and the surface albedo for an arbitrary Lambertian surface. In particular, if an object's true surface is denoted by $f(x,y)$ where the (x,y) plane is perpendicular to the line of sight, a generalized bas-relief transformation of the surface shape is given by $\bar{f} = \lambda f(x,y) + \mu x + \nu y$, and the corresponding generalized bas-relief transformation of the surface albedo is given by Eq. 3. Classical bas-relief sculptures use a subset of the transformation on shape, with $0 < \lambda < 1$, $\mu = 0$, and $\nu = 0$, and, to the best of our knowledge, ignore the corresponding transformation on albedo.

The fact that a surface and a generalized bas-relief transformation of the surface produce the same set of images arises from an implicit duality. For each image of a Lambertian surface $f(x,y)$ produced by a point light source at infinity $\mathbf{s}$, there exists an identical image of the generalized bas-relief transformation $\lambda f(x,y) + \mu x + \nu y$ produced by a transformed light source $\bar{\mathbf{s}}$. This equality holds not only for the illuminated regions of the surfaces, but for the shadowed regions as well. Furthermore, due to superposition, the equality holds not only for a single light source, but for an arbitrary – possibly infinite – number of light sources.

Thus, from a single viewpoint, there is an ambiguity in the recovery of the surface: we can – at best – determine the relief of the surface up to a three parameter family of linear transformations. No information in either the shadowing or shading can resolve this. Yet, if the viewer moves relative to the surface, or the surface moves relative to the viewer, does this ambiguity vanish?

For infinitesimal motion under perspective projection, the structure estimates are sensitive to noise, producing an implicit error in the estimate of the relief λ of the surface [15, 21]. For infinitesimal motion under orthographic projection, there is a genuine bas-relief ambiguity: we can only recover the shape of the surface up to a scale factor in the direction of the camera's optical axis, i.e. a bas-relief transformation for which λ is unknown [9].

A summary of these and other results follow:

- For any light source direction, there exists another light source direction such that cast and attached shadows produced by a surface and a transformed surface are identical, irrespective of the material type.

*P. N. Belhumeur and A.L. Yuille were supported by ARO grant DAAH04-95-1-0494.

†D. J. Kriegman was supported by NSF under an NYI, IRI-9257990.

Figure 1: The figure shows two frontal views and a side view of a pair of bas-relief sculptures. Notice how the frontal views appear to have full 3-D depth, while the side view reveals the extent of the flattening.

- If the material can modeled as having Lambertian reflectance, then the set of possible images under any lighting condition (illumination cone [1]) including shadowing for a surface and its transformed surface are identical; therefore, these objects cannot be distinguished by any recognition algorithm.
- The generalized bas-relief transformation is the *only* transformation which has these first two properties.
- Under orthographic projection, the set of motion fields produced by a surface and its classical bas-relief are identical. Therefore, an object and its relief cannot be distinguished from small unknown camera motion.
- For photometric stereo where the light source directions are unknown, the structure can only be determined up to a generalized bas-relief transformation and shadows do not provide further information. Using prior information about the albedo and light source strength, the structure can be determined up to a reflection in depth. Cast shadows can be used to distinguish these two cases.

Thus, if an object is viewed orthographically, then neither illumination nor small motions of the viewer (or object) will resolve the object's depth relief. Figure 1 shows images of two bas-relief sculptures. Notice how the frontal views appear to have full depth, but the oblique views reveal the extent of the flattening.

2 Bas-Relief Ambiguity: Illumination

Consider a surface observed under orthographic projection and define a coordinate system attached to the image plane such that the $\mathbf{x}$ and $\mathbf{y}$ axes span the image plane. In this coordinate system, the depth of every visible point in the scene can be expressed as

$$z = f(x, y)$$

where f is a piecewise differentiable function. The graph of $f(x,y)$, i.e. $(x, y, f(x, y))$, defines a surface which will also be denoted by$f$.

The direction of the inward pointing surface normal $\mathbf{n}(x, y)$ can be expressed as

$$\mathbf{n}(x,y) = \begin{bmatrix} f_x \\ f_y \\ -1 \end{bmatrix} \tag{1}$$

where f_x and f_y denote the partial derivatives of f with respect to x and y respectively.

Consider transforming the surface f to a new surface $\bar{f}$ in the following manner. We first flatten (or scale) it along the $\mathbf{z}$ axis and then add a plane, i.e.

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

where $\lambda \neq 0$ [2]. We call this transformation the "generalized bas-relief transformation." See Figures 2 and 3. As will be seen in Section 4, this is the only linear transformation of the surface's normal field which preserves integrability. When $\mu = 0$ and $\nu = 0$, we call this transformation the classical "bas-relief transformation," since for $\lambda < 1$ the surface is flattened like classical bas-relief sculptures.

Note that for image point (x, y), the relation between the direction of the surface normal of $\bar{f}$ and f is given by $\bar{\mathbf{n}} = G\mathbf{n}$ where

$$G = \begin{bmatrix} \lambda & 0 & -\mu \\ 0 & \lambda & -\nu \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

Under the matrix product operation, the set $GBR = \{G\}$ forms a subgroup of $GL(3)$ with

$$G^{-1} = \frac{1}{\lambda}\begin{bmatrix} 1 & 0 & \mu \\ 0 & 1 & \nu \\ 0 & 0 & \lambda \end{bmatrix}.$$

Also, note that if $\mathbf{p} = (x, y, f(x, y))$ and $\bar{\mathbf{p}} = (x, y, \bar{f}(x, y))$, then $\bar{\mathbf{p}} = \lambda G^{-T}\mathbf{p}$ where $G^{-T} \equiv (G^T)^{-1} = (G^{-1})^T$.

Let the vector $\mathbf{s}$ denote a point light source at infinity, with the magnitude of $\mathbf{s}$ proportional to the intensity of the light source. (For a more general model

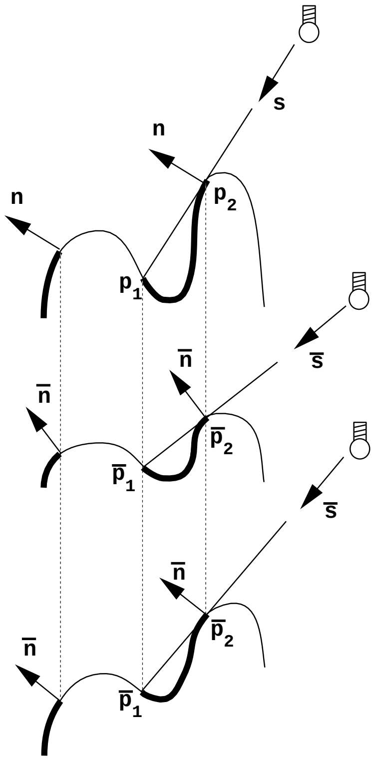

Figure 2: The image points that lie in a cast or attached shadow for a surface f under light source $\mathbf{s}$ are identical to those in shadow for transformed the surface $\bar{f}$ under light source $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$. The middle figure shows a classical bas-relief transformation of the upper figure while the bottom figure is a generalized bas-relief. For diagrammatic clarity, the surface normals are drawn outward.

of illumination, e.g. one that does not restrict light sources to be at infinity, see [12].) We first show that shadowing on a surface f for some light source $\mathbf{s}$ is identical to that on a bas-relief transformed surface $\bar{f}$ with an appropriate light source $\bar{\mathbf{s}}$; we then show that if the surfaces are Lambertian, the set of all possible images of both surfaces are identical.

We can identify two types of shadows: *attached shadows* and *cast shadows* [19]. See Figure 2. A surface point $\mathbf{p} = (x, y, f(x, y))$ lies in an *attached shadow* for light source direction $\mathbf{s}$ iff $\mathbf{n}(x, y)^T\mathbf{s} < 0$. This definition leads to the following lemma.

Lemma 2.1 *A point* $\mathbf{p} = (x, y, f(x, y))$ *lies in an attached shadow for light source direction* $\mathbf{s}$ *iff* $\bar{\mathbf{p}} = (x, y, \bar{f}(x, y))$ *lies in an attached shadow for light source direction* $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$.

Proof. If a point $\mathbf{p}$ on f lies in an attached shadow, then $\mathbf{n}^T\mathbf{s} < 0$. On the transformed surface, the point $\bar{\mathbf{p}} = (x, y, \bar{f}(x, y))$ also projects to (x, y) and for this point $\bar{\mathbf{n}}^T\bar{\mathbf{s}} = \mathbf{n}^T G^T G^{-T}\mathbf{s} = \mathbf{n}^T\mathbf{s}$. Therefore, $\bar{\mathbf{p}}$ is also in an attached shadow. The converse clearly holds as well. ▮

A necessary condition for a point $\mathbf{p}_1 = (x_1, y_1, f(x_1, y_1))$ on the surface to fall on the *cast shadow boundary* for light source direction $\mathbf{s}$ is that there exists another point $\mathbf{p}_2 = (x_2, y_2, f(x_2, y_2))$ on the surface such that the light ray in the direction $\mathbf{s}$ passing through $\mathbf{p}_2$ grazes the surface at $\mathbf{p}_2$ and intersects the surface at $\mathbf{p}_1$. The point $\mathbf{p}_2$ is the boundary of an attached shadow. For smooth surfaces, attached shadow and cast shadow boundaries can be distinguished in intensity images; along any image curve $(x(t), y(t))$ intersecting the shadow boundary transversally, the intensity $I(x(t), y(t))$ is continuous at an attached shadow boundary, whereas it is discontinuous at a cast shadow.

Lemma 2.2 *A point* $\mathbf{p} = (x, y, f(x, y))$ *satisfies the necessary condition for lying on a cast shadow boundary for light source direction* $\mathbf{s}$ *iff* $\bar{\mathbf{p}} = (x, y, \bar{f}(x, y))$ *satisfies the condition for light source direction* $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$.

Proof. The condition for a point $\mathbf{p}_1$ to be on a shadow boundary cast by $\mathbf{p}_2$ is that

$$\begin{cases} \mathbf{n}_2^T\mathbf{s} = 0 \\ \mathbf{p}_2 - \mathbf{p}_1 = \gamma\mathbf{s} \end{cases}$$

for some $\gamma < 0$. For the transformed surface, the first condition for a point to be on the shadow boundary is

$$\bar{\mathbf{n}}_2^T\bar{\mathbf{s}} = \mathbf{n}_2^T G^T G^{-T}\mathbf{s} = \mathbf{n}_2^T\mathbf{s} = 0.$$

Under the generalized bas-relief transformation $\bar{\mathbf{p}} = \lambda G^{-T}\mathbf{p}$, and the second condition can be expressed for the relief surface as

$$\begin{aligned} \bar{\mathbf{p}}_2 - \bar{\mathbf{p}}_1 - \bar{\gamma}\bar{\mathbf{s}} &= \lambda G^{-T}(\mathbf{p}_2 - \mathbf{p}_1) - \bar{\gamma} G^{-T}\mathbf{s} \\ &= \lambda(\mathbf{p}_2 - \mathbf{p}_1) - \bar{\gamma}\mathbf{s} = 0. \end{aligned}$$

This condition clearly holds when $\bar{\gamma} = \lambda\gamma$. The converse of this lemma can be similarly proven. ▮

This lemma becomes both necessary and sufficient for a point to lie on a shadow boundary when the ray from $\mathbf{p}_1$ passing through $\mathbf{p}_2$ does not intersect any other portion of the surface for both f and $\bar{f}$. In general, this is true when $\lambda > 0$.

Taking these two lemmas together, it follows that if some portion of the surface f is in a cast or attached shadow for a light source direction $\mathbf{s}$, then if the surface is subject to a generalized bas-relief transformation G, there exists a lighting direction $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$ such that the same portion of the transformed surface is also shadowed. Let us specify these shadowed regions – both attached and cast – through a binary function $\Psi_{f,\mathbf{s}}(x, y)$ such that

$$\Psi_{f,\mathbf{s}}(x, y) = \begin{cases} 0 & \text{if } (x, y) \text{ is shadowed} \\ 1 & \text{otherwise.} \end{cases}$$

Using this notation and the above two lemmas, we can then write $\Psi_{f,\mathbf{s}}(x,y) = \Psi_{\bar{f},\bar{\mathbf{s}}}(x,y)$.

We should stress that *shadowing* $\Psi_{f,\mathbf{s}}(x,y)$ *is a function of the object's geometry and light source direction – it is unaffected by the reflectance properties of the surface.* For any surface, any light source direction, and any generalized bas-relief transformation of that surface, there exists a light source direction such that the shadowing will be identical irrespective of the surface properties. Furthermore, the generalized bas-relief transformation is the *only* transformation for which this is true for any surface. (Space limitations preclude including a proof.)

We now show that if the surface reflectance is Lambertian [6, 11], then the sets of images produced by a surface (i.e. the surface's illumination cone [1]) and a transformed surface under all possible lighting conditions are identical. Letting the albedo of a Lambertian surface f be denoted by $a(x,y)$, the intensity image produced by a light source $\mathbf{s}$ can be expressed as

$$\mathbf{I}_{f,a,\mathbf{s}}(x,y) = \Psi_{f,\mathbf{s}}(x,y)\mathbf{b}(x,y)^T\mathbf{s}$$

where $\mathbf{b}(x,y)$ is the product of the albedo $a(x,y)$ of the surface and the inward pointing unit surface normal $\hat{\mathbf{n}}(x,y)$.

We now show that the set of images produced by f and $\bar{f}$ are identical when the albedo $\bar{a}(x,y)$ of $\bar{f}$ is

$$\bar{a} = a\sqrt{(\lambda n_1 - \mu n_3)^2 + (\lambda n_2 - \nu n_3)^2 + n_3^2} \qquad (3)$$

where $\hat{\mathbf{n}} = (n_1, n_2, n_3)^T$. The effect of applying Eq. 3 to a classical bas-relief transformation $0 < \lambda < 1$ is to darken points on the surface where $\mathbf{n}$ points away from the optical axis.

Lemma 2.3 *For each light source* $\mathbf{s}$ *illuminating a Lambertian surface* $f(x,y)$ *with albedo* $a(x,y)$*, there exists a light source* $\bar{\mathbf{s}}$ *illuminating a surface* $\bar{f}(x,y)$ *(a generalized bas-relief transformation of* f*) with albedo* $\bar{a}(x,y)$ *(as given in Eq. 3), such that* $I_{f,a,s}(x,y) = I_{\bar{f},\bar{a},\bar{\mathbf{s}}}(x,y)$.

Proof. The image of f is given by

$$I_{f,a,\mathbf{s}}(x,y) = \Psi_{f,\mathbf{s}}(x,y)\mathbf{b}^T(x,y)\mathbf{s}$$

For any 3×3 invertible matrix A, we have that

$$I_{f,a,\mathbf{s}}(x,y) = \Psi_{f,\mathbf{s}}(x,y)\mathbf{b}^T(x,y)AA^{-1}\mathbf{s}.$$

Since GBR is a subgroup of $GL(3)$ and $\Psi_{f,\mathbf{s}}(x,y) = \Psi_{\bar{f},\bar{\mathbf{s}}}(x,y)$,

$$\begin{aligned} I_{f,a,\mathbf{s}}(x,y) &= \Psi_{f,\mathbf{s}}(x,y)\mathbf{b}^T(x,y)G^TG^{-T}\mathbf{s} \\ &= \Psi_{f,\mathbf{s}}(x,y)\bar{\mathbf{b}}^T(x,y)\bar{\mathbf{s}} \\ &= I_{\bar{f},\bar{a},\bar{\mathbf{s}}}(x,y) \end{aligned}$$

where $\bar{\mathbf{b}}(x,y) = G\mathbf{b}(x,y)$ and $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$. ▮

With the above three lemmas in hand, we can now state and prove the central proposition of this section:

Proposition 2.1 *The set of images under all possible lighting conditions produced by a Lambertian surface* f *with albedo* $a(x,y)$ *and those surfaces* $\bar{f}$ *differing by any generalized bas-relief transformation with albedo* $\bar{a}(x,y)$ *given by Eq. 3 are identical.*

Proof. From Lemmas 2.1, 2.2, and 2.3, we have that the image of a surface f produced by light source $\mathbf{s}$ is the same as the image of a generalized bas-relief transformed surface $\bar{f}$ produced by the transformed light source $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$, i.e. $I_{f,a,\mathbf{s}}(x,y) = I_{\bar{f},\bar{a},\bar{\mathbf{s}}}(x,y)$. When the object is illuminated by a set of light sources $\{\mathbf{s}_i\}$, then the image is determined by the superposition of those images that would be formed under the individual light sources. Similarly, the same image can be produced from the transformed surface if it is illuminated by the set of light sources given by $\{\bar{\mathbf{s}}_i\}$, where $\bar{\mathbf{s}}_i = G^{-T}\mathbf{s}_i$. ▮

This proposition says that the set of surfaces produced by a generalized bas transformation of a surface form an equivalence class. The sets of possible images produced by every surface in this equivalence class are identical. That is, the set of images is invariant over the equivalence class of surfaces formed under the generalized bas-relief transformation. An implication of this result is that given any number of images, it is impossible to distinguish two objects that differ only by a generalized bas-relief transformation. Additional information must be brought to bear to distinguish them.

In Figure 3, we have simulated bas-relief transformations of a human face. The middle row contains images produced by the true surface of the face. The top row contains images produced by a flattened form of the surface, and the bottom row contains images produced by an elongated form of the surface. The left column shows the surface of the face from a side view, orthogonal to the direction of the chosen bas-relief transformation. The middle column shows the faces from a frontal view, parallel to the direction of the transformation. By choosing the appropriate lighting directions for each surface all three images in the column appear identical. (The right column is explained in the next section.)

3 Bas-Relief Ambiguity: Motion

Consider again a surface observed under orthographic projection. We again define a coordinate system attached to the image plane such that the $\mathbf{x}$ and $\mathbf{y}$ axes span the image plane and the depth of every visible point in the scene can be expressed as $z = f(x,y)$ where f is again a continuous function.

If the surface undergoes a rigid motion and is viewed under perspective projection, the object's Euclidean structure can be determined from as few as two images [13, 16, 23]. If the object is viewed orthographically, the object's structure can only be determined up to a one parameter family of affine distortions [9]. To

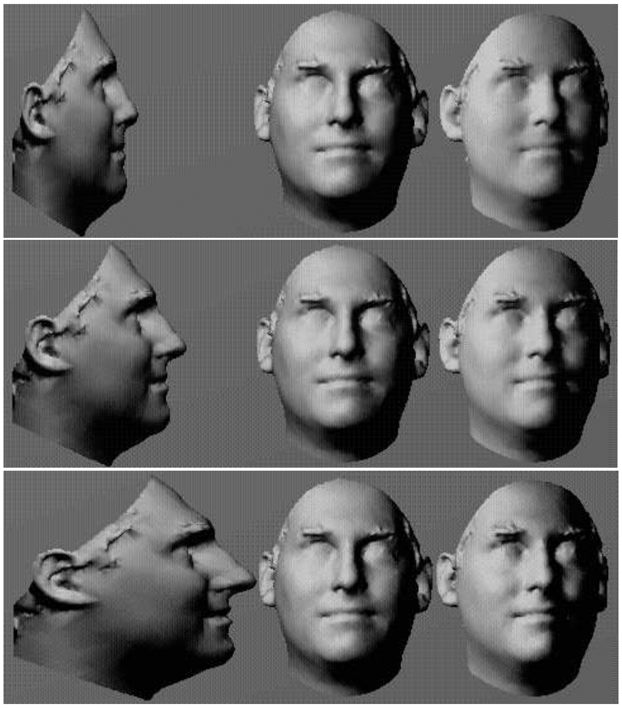

Figure 3: This figure shows images of three human faces each differing by a classical bas-relief transformation. The 3-D data for the true head (the middle row) was obtained with a 3-D scanner (Cyberwave) and rendered assuming a Lambertian surface with constant albedo. The heads in the first and third row were obtained by scaling the z coordinate with $\lambda < 1$ and $\lambda > 1$ respectively. The left column shows the faces from a side view, orthogonal to the direction of the chosen bas-relief transformation. The middle column shows the faces from a frontal view, parallel to the direction of the transformation. We have chosen lighting directions to illuminate the middle images so that all three images will appear identical. The right column shows images of the faces after being rotated. We have chosen rotations angles (7, 5, and 3.5 degrees from top to bottom) to make the images appear nearly identical.

determine the Euclidean structure under orthographic projection, at least three images are needed.

Yet, complications arise when the object's motion is small. For infinitesimal motion under perspective projection, the structure estimates are sensitive to noise, producing an implicit error in the estimate of the relief of the surface [15, 21]. For small (infinitesimal) unknown motion under orthographic projection, there is a genuine bas-relief ambiguity: the shape of the surface can only be recovered up to a scale factor in the direction of the camera's optical axis, i.e. a classical bas-relief transformation ($\lambda > 0, \mu = \nu = 0$).

To see this, let us assume that the surface does, in fact, undergo and arbitrary infinitesimal motion. The velocity $(\dot{x}, \dot{y}, \dot{z})$ of a point (x, y, z) on the surface f induces a velocity $(\dot{x}, \dot{y})$ in the image plane. The collection of velocities for all points in the image plane is often called the motion field. In the following proposition, we show the motion fields of any surface and a classical bas-relief transformation (not a generalized bas-relief transformation) of the surface are identical.

Proposition 3.1 *The set of motion fields induced by all 3-D infinitesimal motions of a surface f is the same, under orthographic projection, as the set of all motion fields of a surface differing by a bas-relief transformation $\bar{f}(x, y) = \lambda f(x, y)$ where $\lambda \neq 0$.*

Proof. Any rigid motion of the surface can be decomposed into a rotation about an axis through the origin and a translation. The overall motion field is the sum of the motion fields produced by rotation and translation. For translation, the motion field is independent of depth, i.e. constant for all (x, y) and, consequently, equivalent for both f and $\bar{f}$. For rotation, the motion can be further decomposed into a rotation about the camera's optical axis and a rotation about an axis in the image plane. Rotations in the image plane create motion fields which are again independent of depth. Thus, the only motion field in this decomposition that is dependent on depth is a rotation about an axis in the image plane.

Without loss of generality, let us choose the **y** axis. Denoting the angular velocity of the surface by $\Omega = (0, \dot{\theta}, 0)$, the 3-D velocity of a point $\mathbf{p} = (x, y, z)$ is

$$\dot{\mathbf{p}} = \Omega \times \mathbf{p} = \begin{bmatrix} -\dot{\theta} z \\ 0 \\ \dot{\theta} x \end{bmatrix}$$

Under orthographic projection, the motion field is given by

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\dot{\theta}\, z \\ 0 \end{bmatrix} = \begin{bmatrix} -\dot{\theta}\, f(x, y) \\ 0 \end{bmatrix}$$

for all (x, y). Since the angular velocity is unspecified, the surfaces f and $\bar{f} = \lambda f$ have the same motion vector fields. ▮

This proof follows the results in [9, 17]. An implication of Proposition 3.1 is that under orthographic projection, a small motion of either the object or the observer cannot resolve the bas-relief ambiguity. Furthermore, since the motion field is linear in $f(x, y)$, the classical bas-relief transformation is the *only* transformation of f that will be preserve the set of motion fields.

Consider the third column in Figure 3. The image produced by a "normal" relief for a viewing direction of 5° from frontal is nearly identical to the images produced by a motion of 7° for the flattened head and by a motion of 3.5° for the elongated head.

4 Integrability, Reconstruction, and the Bas-Relief Ambiguity

In this section, we investigate the role of the generalized bas-relief ambiguity on surface reconstruction using photometric stereo. Let us assume that a Lambertian surface is illuminated by a point light source at infinity. When there is no shadowing (i.e. $\Psi_{f,\mathbf{s}}(x,y) = 1$), the intensity image produced by a light source $\mathbf{s}$ can be expressed as

$$\mathbf{I}_{f,a,\mathbf{s}}(x,y) = \mathbf{b}(x,y)^T\mathbf{s} \quad (4)$$

where $\mathbf{b}(x,y)$ is the product of the albedo $a(x,y)$ of the surface and the inward pointing unit surface normal $\hat{\mathbf{n}}(x,y)$. From multiple images of the object seen from a fixed viewpoint but with different light source direction, we can solve Eq. 4 for $\mathbf{b}$ when the light source strengths and directions are known. This, of course, is the standard photometric stereo technique, see [6, 20, 24].

However, if the light source strengths and directions are *not known*, then we can only determine the vector field $\mathbf{b}(x,y)$ of surface normals and albedos up to a 3×3 linear transformation. For any invertible 3×3 linear transformation $A \in GL(3)$ [2, 5, 17]

$$\mathbf{b}^T\mathbf{s} = \mathbf{b}^T A^T A^{-T}\mathbf{s}. \quad (5)$$

If $\mathbf{b}(x,y)$ is the true vector field of surface normals then the recovered vector field $\mathbf{b}^*(x,y)$ is any vector field in the orbit of $\mathbf{b}(x,y)$ under the group $GL(3)$. For a pixelated image with no surface point in shadow, $\mathbf{b}^*$ can be estimated from a collection of images using singular value decomposition; when some of the surface points are shadowed, Jacobs' method can be used to estimate $\mathbf{b}^*$ [8]. Note, however, that not all vector fields $\mathbf{b}^*(x,y)$ correspond to continuous (or even piecewise continuous) surfaces. We will use this observation to restrict the group of allowable transformations on $\mathbf{b}(x,y)$ [2].

If $\mathbf{b}$ is transformed by an arbitrary $A \in GL(3)$ (i.e. any vector field $\mathbf{b}^*(x,y)$ in the orbit of $\mathbf{b}$ under $GL(3)$), then in general, there will be no surface $f^*(x,y)$ with unit normal field $\hat{\mathbf{n}}^*(x,y)$ and albedo $a^*(x,y)$ that could have produced the vector field $b^*(x,y)$. For $f^*(x,y)$ to be a surface, it must satisfy the following integrability constraint [7]:

$$f^*_{xy} = f^*_{yx}$$

which, in turn, means $\mathbf{b}^*(x,y)$ must satisfy

$$\left(\frac{b_1^*}{b_3^*}\right)_y = \left(\frac{b_2^*}{b_3^*}\right)_x \quad (6)$$

Proposition 4.1 *If $\mathbf{b}(x,y)$ corresponds to a surface $f(x,y)$ with albedo $a(x,y)$, then the set of linear transformation $\mathbf{b}^*(x,y) = A\mathbf{b}(x,y)$ which satisfy the integrability constraint in Eq. 6 are the generalized bas-relief transformations G given in Eq. 2.*

Proof. The integrability constraint given in Eq. 6 can be written as $(b^*_{1_y} - b^*_{2_x})b^*_3 + b^*_{3_x}b^*_2 - b^*_{3_y}b^*_1 = 0$. Letting A_{ij} be the i,j-th element of A, and recalling that $\mathbf{b}^* = A\mathbf{b}$, the left hand side is a function of $b_i(x,y), b_{i_x}(x,y), b_{i_y}(x,y)$ for $i = 1,2,3$. Since these functions are generally independent, the coefficients of these function must all vanish for the integrability constraint to hold for all (x,y). This leads to the following algebraic constraints on the elements of A.

$$\begin{cases} A_{22}A_{31} - A_{21}A_{32} = 0 \\ A_{21}A_{33} - A_{23}A_{31} = 0 \\ A_{12}A_{33} - A_{13}A_{32} = 0 \\ A_{12}A_{31} - A_{11}A_{32} = 0 \\ A_{22}A_{33} - A_{11}A_{33} + A_{13}A_{31} - A_{32}A_{23} = 0 \end{cases}$$

Since this system is homogeneous, for any A satisfying this system, ρA also satisfies the system; varying ρ corresponds to changing the light source intensity while making a corresponding global scaling of the albedo function. It can be shown that if $A_{33} = 0$, the matrix A satisfying the constraints is singular. So we can let $A_{33} = 1$, and solve for the remaining coefficients. The only nonsingular solution is $A_{11} = A_{22}$ and $A_{12} = A_{21} = A_{31} = A_{32} = 0$. That is, A must be a generalized bas-relief transformation. ∎

The choice of $\mathbf{b}^*(x,y)$ is, of course, not unique since $\mathbf{b}^*(x,y) = G\mathbf{b}$ satisfies the integrability constraint for any $G \in GBR$. Yet, every $\mathbf{b}^*$ has a corresponding surface f^* with a corresponding albedo $a(x,y)$, and these surfaces differ by a generalized bas-relief ambiguity. Thus, if we have at least three images – each acquired under different light source directions – of a surface $f(x,y)$ with Lambertian reflectance and albedo $a(x,y)$, then by imposing the integrability constraint in Eq. 6, we can recover the surface $f(x,y)$ up to a generalized bas-relief transformation $\bar{f}(x,y) = \lambda f(x,y) + \mu x + \nu y$. Note that no information given in the image shadows can resolve this ambiguity, as Section 2 showed that the set of all possible images of a surface $f(x,y)$ is invariant under the generalized bas-relief transformation. If, however, we have additional information about the albedo or the strength of the light sources we can further restrict the ambiguity.

Corollary 4.1 *If the albedo $a(x,y)$ is constant (or known), or the light sources $\mathbf{s}_i$ all have the same intensity, then the generalized bas-relief ambiguity G is restricted to the binary subgroup given by $\lambda = \pm 1, \mu = 0$, and $\nu = 0$.*

Proof. If $a(x,y) = |\mathbf{b}(x,y)|$ is constant (or known), then for $|\mathbf{b}(x,y)| = |\mathbf{b}^*(x,y)| = |A\mathbf{b}(x,y)|$, A must preserve length for any $\mathbf{b}$. The only matrices that preserve length are the orthonormal matrices. The only orthonormal matrices that are also generalized bas-relief transformations correspond to $\lambda = \pm 1, \mu = 0$, and $\nu = 0$. A similar argument holds about G^{-T} when the light source intensities are known. ∎

Thus, we can determine the true surface up to a sign, i.e. $\bar{f}(x, y) = \pm f(x, y)$. This is the classical in-out ambiguity that occurs in shape from shading [6, 14]. Note however, that the shadowing configurations change when λ changes sign, and if shadowing is present, this ambiguity can be resolved.

5 Discussion

We have shown that under any lighting condition, the shading and shadowing on an object is identical to the shading and shadowing on any generalized bas-relief transformation of the object. The generalized bas-relief transformation is unique in that it is the only transformation of the surface having this property. Thus, from a single viewpoint, there is an ambiguity in the recovery of the surface: we can – at best – determine the relief of the surface up to a three parameter family of linear transformations. No information in either the shadowing or shading can resolve this. This result supports the recent psychophysical findings of [10] that for a variety of surfaces this ambiguity exists and is often unresolved in the human visual system. Furthermore, the motion fields produced by small camera motions cannot be used to resolve the surface relief.

In shape recovery, the generalized bas-relief transformation arises because the recovered surface must be piecewise integrable. While it has been thought that photometric stereo with unknown light source direction could be solved by first estimating the light source directions and then estimating the surface structure, this paper has shown that these estimates are coupled through an unresolvable generalized bas-relief transformation. Taken together, these results suggest that the aim of structure recovery should be a weaker non-Euclidean representation, such as an affine representation [9, 17, 18, 22], a projective representation [3], or an ordinal representation [4]; object recognition should not depend on resolution of these ambiguities.

References

[1] 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.

[2] R. Epstein, A. Yuille, and P. N. Belhumeur. Learning and recognizing objects using illumination subspaces. In *Proc. of the Int. Workshop on Object Representation for Computer Vision*, 1996.

[3] O. Faugeras. Stratification of 3-d vision: Projective, affine, and metric representations. *J. Opt. Soc. Am. A*, 12(7):465–484, 1995.

[4] C. Fermuller and Y. Aloimonos. Ordinal representations of visual space. In *Proc. Image Understanding Workshop*, pages 897–904, 1996.

[5] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *JOSA-A*, 11(11):3079–3089, Nov. 1994.

[6] B. Horn. *Computer Vision.* MIT Press, Cambridge, Mass., 1986.

[7] B. Horn and M. Brooks. The variational approach to shape from shading. In *Computer Vision, Graphics and Image Processing*, volume 33, pages 174–208, 1986.

[8] D. Jacobs. Linear fitting with missing data: Applications to structure from motion and characterizing intensity images. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1997.

[9] J. Koenderink and A. Van Doorn. Affine structure from motion. *JOSA-A*, 8(2):377–385, 1991.

[10] J. J. Koenderink, A. J. Van Doorn, and C. Christon. Shape constancy in pictorial relief. In J. Ponce, A. Zisserman, and M. Hebert, editors, *Object Representation in Computer Vision II*, pages 151–164. Springer, 1996.

[11] J. Lambert. *Photometria Sive de Mensura et Gradibus Luminus, Colorum et Umbrae.* Eberhard Klett, 1760.

[12] M. Langer and S. Zucker. What is a light soruce? In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1997.

[13] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[14] J. Oliensis. Uniqueness in shape from shading. *Int. J. Computer Vision*, 6(2):75–104, June 1991.

[15] J. Oliensis. Structure from linear or planar motions. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 335–342, June 1996.

[16] J. Roach and K. Aggarwal, J. Computer tracking of objects from a sequence of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1(2):127–135, 1979.

[17] R. Rosenholtz and J. Koenderink. Affine structure and photometry. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 790–795, 1996.

[18] L. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *Int. J. Computer Vision*, 16(2):147–182, October 1995.

[19] A. Shashua. *Geometry and Photometry in 3D Visual Recognition.* PhD thesis, MIT, 1992.

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

[21] R. Szeliski and S. Kang. Shape ambiguities in structure from motion. In *European Conf. on Computer Vision*, pages I:709–721, 1996.

[22] S. Ullman and R. Basri. Recognition by a linear combination of models. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 13:992–1006, 1991.

[23] A. Waxman and S. Ullman. Surface structure and three-dimensional motion from image flow kinematics. *Int. J. Robotics Research*, 4:72–92, 1985.

[24] R. Woodham. Analysing images of curved surfaces. *Artificial Intelligence*, 17:117–140, 1981.