Toward a Stratification of Helmholtz Stereopsis[*]

Todd E. Zickler
Electrical Engineering
Yale University
New Haven, CT, 06520
zickler@yale.edu

Peter N. Belhumeur
Computer Science
Columbia University
New York, NY, 10027
belhumeur@cs.columbia.edu

David J. Kriegman
Computer Science and Engineering
University of California, San Diego
La Jolla, CA, 94143-1290
kriegman@cs.ucsd.edu

Abstract

Helmholtz stereopsis has been previously introduced as a surface reconstruction technique that does not assume a model of surface reflectance. This technique relies on the use of multiple cameras and light sources, and it has been shown to be effective when the camera and source positions are known. Here, we take a stratified look at uncalibrated Helmholtz stereopsis. We derive a new photometric matching constraint that can be used to establish correspondence without any knowledge of the cameras and sources (except that they are co-located), and we determine conditions under which we can obtain affine and metric reconstructions. An implementation and experimental results are presented.

1. Introduction

Numerous surface reconstruction techniques (structure from motion, binocular or multinocular stereopsis, photometric stereopsis, etc.) have been adapted to handle uncalibrated or weakly calibrated image and illumination equipment. What is equally important is that the community has established an understanding of what 3D information can be obtained under stratified levels of prior knowledge about the acquisition system. A stratified approach is useful because it tells us what can be obtained under a given calibration condition (e.g., projective, affine or metric reconstruction) and what assumptions are required in order to obtain more. This approach has been applied to the problems of binocular and multinocular stereo [3, 4, 6], structure from motion [8, 14] and photometric stereo [7, 13]. In this paper, we take a stratified look at a relatively new reconstruction technique, Helmholtz stereopsis [17].

Helmholtz stereopsis has the unique property of enabling the estimation of surface shape *without requiring an assumed model of reflectance.* Thus, unlike most existing methods, it enables the dense reconstruction of scenes that contain unknown and arbitrary surface reflectance functions (BRDFs.) In Helmholtz stereopsis, this is accomplished through the use of images that are collected in reciprocal pairs [17]. (A reciprocal pair is a set of two images for
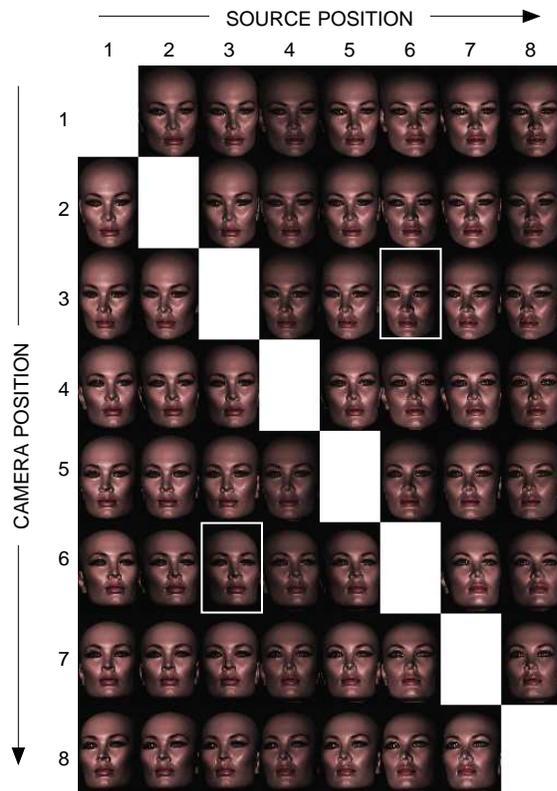

Figure 1. Reciprocal images from eight camera/source positions. Columns contain fixed illumination (stereo) images and rows contain fixed viewpoint (photometric stereo) images. One reciprocal pair is shown highlighted.

which the positions of the camera and a point light source are swapped — see Fig. 1.) Reciprocal pairs have the important property that the relationship between intensities at corresponding image points depends only on surface shape, and is independent of reflectance (a property that follows directly from the symmetry in surface reflectance.)

The use of reciprocal image pairs means that Helmholtz stereopsis requires images taken under multiple point-source illuminations and from multiple viewpoints. The changes in both viewpoint and illumination enable the method to combine the advantages of conventional

multinocular and photometric stereopsis. Figure 1 shows a typical Helmholtz stereo dataset. In this figure, rows correspond to fixed viewpoint (photometric stereo) images, and columns to fixed illumination (conventional stereo) images. Whereas reconstruction based on conventional or photometric stereo would use a single column or row from this dataset, respectively, Helmholtz stereopsis makes use of all of the images in this figure. Multiple viewpoints enable Helmholtz stereopsis to estimate depth (and therefore locate depth discontinuities), and multiple illuminations provide photometric information that enables the direct estimation of the field of surface normals.

Helmholtz stereopsis is related to a small set of fairly recent reconstruction techniques (others are [10, 11]) that utilize both changing viewpoint and illumination. Lu and Little [10] used the term *photogeometric* to describe their technique, which seems like an appropriate term for the entire class of methods. Although these methods are quite different (Helmholtz stereopsis being the only one that both provides direct surface normal estimates and can handle a BRDF that varies over the surface), they share the property of requiring calibrated cameras and light sources (at least in their original formulations.)

Generally, the calibration that is required by Helmholtz stereopsis and these other photogeometric techniques includes intrinsic and extrinsic camera parameters, as well as the radiometric response of the cameras, and the radiance map of the light sources (a measure of radiance as a function of output direction.) A question that has rarely been considered is[1]: What surface information can be obtained by photogeometric techniques in the uncalibrated case? This paper moves toward answering this question for Helmholtz stereopsis, and it does so in a stratified manner.

The next section introduces a correspondence constraint that requires little knowledge about the cameras and sources and shows how this constraint could be used to obtain a projective reconstruction in the uncalibrated case. In Sect. 3, the case of distant sources and affine cameras is discussed in detail. In Sect. 4, we see that in the uncalibrated, distant source case, it is possible to obtain a reconstruction of the surface and its field of surface normals up to an unknown affine transformation. Sect. 5 discusses information that can be used to upgrade this affine reconstruction to a metric one; both geometric and photometric pieces of information are considered. Finally, Sect. 6 presents an implementation with results that lend empirical support to the discussion.

2. Helmholtz Stereopsis

This section is broken into two parts. First, we review calibrated Helmholtz stereopsis as presented in [17]. Second,

we derive a new photometric matching constraint that can be used to establish correspondence between views in the uncalibrated case.

Consider M isotropic point light sources co-located at the camera centers of M pinhole cameras. (This can be accomplished in numerous ways, such as by swapping cameras and sources, or by using half-mirrors or beam-splitters. Co-location can also be approximated by placing each light source near a camera.) Images are acquired in the following fashion. Light source i is turned on while the other sources are turned off, and $M - 1$ images are acquired from all cameras but camera i. This process is repeated M times, each with a different source turned on, until $M(M - 1)$ images are acquired. These images are comprised of $M(M - 1)/2$ reciprocal image pairs. Figure 1 shows a set of these images for $M = 8$. In this figure, the vertical direction corresponds to camera position, the horizontal direction to source position, and reciprocal pairs are in symmetric positions.

As shown in [17], given $M(M - 1)/2$ reciprocal pairs taken from camera/source positions $\mathbf{o}_1 \ldots \mathbf{o}_M$ we have $M(M-1)/2$ constraints on the intensities of corresponding image points. These constraints are of the form

$$\left(e_{ij} \frac{s_i \hat{\mathbf{v}}_i^\mathsf{T}}{|\mathbf{o}_i - \mathbf{p}|^2} - e_{ji} \frac{s_j \hat{\mathbf{v}}_j^\mathsf{T}}{|\mathbf{o}_j - \mathbf{p}|^2} \right) \cdot \hat{\mathbf{n}} = 0, \qquad (1)$$

where e_{ij} is the image irradiance at the projection of surface point $\mathbf{p}$ (with surface normal $\hat{\mathbf{n}}$) in camera $\mathbf{o}_i$ when illuminated by a point source at $\mathbf{o}_j$. The unit vector $\hat{\mathbf{v}}_i$ points in the direction from $\mathbf{p}$ to $\mathbf{o}_i$. We have also included the relative source strengths s_i. (All sources were assumed to be equal or of known strength in the original formulation— a non-trivial assumption.) Note that when the camera and source positions (and the source strengths) are known, everything inside the parentheses is determined by the value of a single parameter, the depth d of the surface. If we stack the constraints[2] (1) into rows of a matrix $\mathbf{M}$, we obtain the matrix constraint

$$\mathbf{M}(d)\hat{\mathbf{n}} = \mathbf{0} \qquad (2)$$

where $\mathbf{M} \in \mathbb{R}^{M(M-1)/2 \times 3}$, and following [17], $\mathbf{M}$ is written as a function of d to stress the fact that it is completely determined by depth.

The constraint (2) can be used to establish correspondence between the images and thus provide an estimate of the surface depth d. This is because, for correct values of d, rank $\mathbf{M} \leq 2$ (whereas rank $\mathbf{M} = 3$ in general.) In addition, the constraint provides an estimate of the surface normal, since for a correct depth d, the normal $\hat{\mathbf{n}}$ spans the 1D nullspace of $\mathbf{M}$.

2.1. Uncalibrated Correspondence

The original formulation of Helmholtz stereopsis as reviewed in the previous section assumes knowledge of the

---

[1] An exception is Rosenholtz and Koenderink [13], who took a stratified look at combining geometric and photometric information for polygonal surfaces

[2] For brevity, references to equations will be written in parentheses.

camera and source positions and the source strengths. It also implicitly assumes that the radiometric camera responses are known or are linear and equal (since (1) involves irradiance values and not pixel intensities) and that the M point sources are isotropic. Next, we show that by a simple rearrangement of the matrix constraint (2), we can eliminate the need for most of these assumptions.

By making the definition

$$\tilde{w}_i = \frac{s_i \hat{\mathbf{v}}_i \cdot \hat{\mathbf{n}}}{|\mathbf{o}_i - \mathbf{p}|^2} \tag{3}$$

we can write (1) as $e_{ij}\tilde{w}_i - e_{ji}\tilde{w}_j = 0$, and the matrix constraint (2) becomes

$$\mathbf{E}\tilde{\mathbf{w}}^\top = \mathbf{0} \tag{4}$$

where $\mathbf{E} \in \mathbb{R}^{(M(M-1)/2) \times M}$ and $\tilde{\mathbf{w}} \in \mathbb{R}^M$. For example, if $M = 4$, we have

$$\begin{bmatrix} e_{12} & -e_{21} & 0 & 0 \\ e_{13} & 0 & -e_{31} & 0 \\ e_{14} & 0 & 0 & -e_{41} \\ 0 & e_{23} & -e_{32} & 0 \\ 0 & e_{24} & 0 & -e_{42} \\ 0 & 0 & e_{34} & -e_{43} \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \tilde{w}_3 \\ \tilde{w}_4 \end{bmatrix} = \mathbf{0}.$$

Equation (4) is satisfied for visible points on a surface in the scene. Therefore, similar to the use of (2) in the original formulation [17], the constraint $\text{rank}\,\mathbf{E} < M$ can be used to establish correspondence. When the epipolar geometry of the camera system is known (i.e., weak calibration), then $\mathbf{E}$ will be a function of depth, and a multinocular stereo search process can be performed using this constraint. When it is unknown, this constraint would have to be used within the context of a robust structure from motion algorithm (e.g., [1].) In either case, one obtains a set of corresponding points, and in turn, one can reconstruct the scene up to a projective transformation [3]. Also, notice that we do not require irradiance measurements from all $M(M-1)/2$ reciprocal pairs in order to use the correspondence constraint based on (4). We only require measurements from P pairs, where $P \geq M$.

The key advantage of this new photometric matching constraint is that it depends only on measured irradiance values, so we do not need to know the positions of the cameras and sources, nor do we need to know the source strengths. All that we require is: 1) that the radiometric responses of the cameras are linear and equal (or are known); and 2) that the light sources are isotropic.

In the following sections, we will discuss an uncalibrated reconstruction technique based on this constraint, even though it is a weaker correspondence constraint than that based on (2) (see the Appendix.) While this may seem disconcerting, note that the original constraint is itself only

a necessary condition for correspondence, and that the advantage of allowing for an uncalibrated system is an important one.

3. Distant Sources

In the previous section, we derived a correspondence constraint that does not require knowledge of the cameras and sources. This suggests that it is possible to establish a projective reconstruction of scene points from uncalibrated reciprocal pairs. This is similar to conventional uncalibrated stereo [3], except that since we have carefully varied the lighting between views, we are able to use a matching constraint that is independent of reflectance (as opposed to relying on brightness constancy.)

In this section, we show that when the cameras and sources are far from the scene, we can quite easily obtain more: the reciprocal image pairs provide accessible information about both the surface normals and the light sources.

When the cameras and sources are far from the scene, we can write (1) as

$$\left(e_{ij}\mathbf{s}_i^\top - e_{ji}\mathbf{s}_j^\top\right)\hat{\mathbf{n}} = 0, \tag{5}$$

where $\mathbf{s}_i$ is a product of the effective source strength s_i and direction $\hat{\mathbf{s}}_i$, both of which are the same for all points $\mathbf{p}$ in the scene. Accordingly, the vector $\tilde{\mathbf{w}}$ in (4) simplifies to

$$\mathbf{w} = \begin{bmatrix} \mathbf{s}_1^\top\hat{\mathbf{n}} & \mathbf{s}_2^\top\hat{\mathbf{n}} & \cdots & \mathbf{s}_M^\top\hat{\mathbf{n}} \end{bmatrix}^\top. \tag{6}$$

Now, suppose that we have established correspondence for N points. That is, we have corresponding observations of N unknown scene points $\mathbf{X}_1 \dots \mathbf{X}_N \in \mathbb{R}^3$ in each of M viewpoints. (This could be achieved using the constraint $\text{rank}\,\mathbf{E} < M$, for example.) For a given point $\mathbf{X}_k$, we have $M(M-1)$ irradiance observations, one for each source/camera pair, and we can form an irradiance matrix $\mathbf{E}_k$ for that point as in (4). Since these irradiance observations correspond to a single surface point, this matrix in general has rank $(M-1)$, and its 1D null space can be expressed as $c_k\mathbf{w}_k = c_k\begin{bmatrix} \mathbf{s}_1^\top\hat{\mathbf{n}}_k & \mathbf{s}_2^\top\hat{\mathbf{n}}_k & \cdots & \mathbf{s}_M^\top\hat{\mathbf{n}}_k \end{bmatrix}^\top$, $c_k \in \mathbb{R}$. (It may be possible for the rank of $\mathbf{E}$ to drop below $(M-1)$, but a discussion of these cases is outside the scope of this paper.) Letting $\mathbf{W}$ denote the $N \times M$ matrix (recall that N is the number of points and M is the number of sources) whose rows are the transposed null vectors $c_1\mathbf{w}_1^\top \dots c_N\mathbf{w}_N^\top$, we have

$$\mathbf{W} = \begin{bmatrix} c_1\mathbf{s}_1^\top\hat{\mathbf{n}}_1 & c_1\mathbf{s}_2^\top\hat{\mathbf{n}}_1 & \cdots & c_1\mathbf{s}_M^\top\hat{\mathbf{n}}_1 \\ c_2\mathbf{s}_1^\top\hat{\mathbf{n}}_2 & c_2\mathbf{s}_2^\top\hat{\mathbf{n}}_2 & & \vdots \\ \vdots & & \ddots & \\ c_N\mathbf{s}_1^\top\hat{\mathbf{n}}_N & \cdots & & c_N\mathbf{s}_M^\top\hat{\mathbf{n}}_N \end{bmatrix}. \tag{7}$$

The matrix $\mathbf{W}$ is of a familiar form; it is precisely this matrix that one considers in the bilinear calibration-estimation problem of uncalibrated photometric stereo [2,

7, 9]. The difference here is that the 'albedo values' are not due to surface reflectance properties, but are instead artifacts of the homogeneity of (4). Indeed, the reflectance of the surface has been effectively removed through the use of reciprocal images. Each column of $\mathbf{W}$ corresponds to a Lambertian image of the N points on the surface under a fixed source direction, and Fig. 2 shows what these images look like for dense points on a real surface.

In order to extract the surface normal and source information embedded in $\mathbf{W}$, we must find the decomposition $\mathbf{W} = \mathbf{BS}$, where $\mathbf{B}$ is the $N \times 3$ matrix of surface normals (each scaled by a constant c_k), and $\mathbf{S}$ is the $3 \times M$ matrix of source vectors (source directions scaled by source strength.) A common strategy is to find a preliminary rank 3 factorization of $\mathbf{W}$ using SVD, and then to correct that decomposition using additional constraints [7, 9]. That is, one computes $\mathbf{W} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}$ and defines $\tilde{\mathbf{B}} = \mathbf{U}\boldsymbol{\Sigma}^{1/2}$ and $\tilde{\mathbf{S}} = \boldsymbol{\Sigma}^{1/2}\mathbf{V}^\mathsf{T}$, keeping only the rows and columns of $\mathbf{U}, \boldsymbol{\Sigma}$, and $\mathbf{V}$ corresponding to the first three singular values. (Here, it is assumed that $N \geq 3$, $M \geq 3$ and that not all normals or sources lie in a plane.) This decomposition is not unique (since $\tilde{\mathbf{B}}\mathbf{Q}^{-1}\mathbf{Q}\tilde{\mathbf{S}} = \tilde{\mathbf{B}}\tilde{\mathbf{S}}$ for all $\mathbf{Q} \in GL(3)$), and the true decomposition $\mathbf{W} = \mathbf{BS}$ can be obtained by finding the matrix $\mathbf{Q}$ that satisfies

$$\begin{aligned} \mathbf{B} &= \tilde{\mathbf{B}}\mathbf{Q}^{-1} \\ \mathbf{S} &= \mathbf{Q}\tilde{\mathbf{S}}. \end{aligned} \qquad (8)$$

Note that $\mathbf{Q}$ can only be defined up to scale, which is an expression of the fact that we can apply a global scaling to the source strengths (and the inverse scaling to the c_k's) without affecting $\mathbf{W}$. Thus, $\mathbf{Q}$ has eight degrees of freedom. We examine relevant methods to determine $\mathbf{Q}$ in Sect. 4.

In the previous section, we derived a constraint that can be used to establish correspondence in the uncalibrated case. In the present section, we showed that in the case of distant sources, we can go further; we can use the available photometric information to estimate the surface normals at the points of observation as well as the strength and direction of the light sources. In making this statement, we are ignoring the problem of solving for $\mathbf{Q}$ in (8), but notice that we have not yet used the available geometric information. Since the source directions are equivalent to the viewing directions, they can alternatively be computed using established techniques of structure from motion. The interplay between the geometric and photometric constraints is discussed in the next two sections, leading to a number of ways in which we can establish a dense metric reconstruction in the uncalibrated case of distant-source Helmholtz stereopsis.

4. Affine Reconstruction

In this section we demonstrate that geometric information can be used to resolve the ambiguity in the calibration-

estimation problem. As a result, for $N \geq 4$ observed points and $M \geq 4$ camera/source positions we can obtain a dense affine reconstruction of the scene (including surface normals and source strengths) without making any assumptions beyond those of the previous section.

When the cameras are far from the scene, they can be accurately described using the parallel projection model. Here, we use the most general such model—that of an affine camera [12]. To represent an affine camera, we use the tuple $[\mathbf{P}, \mathbf{t}]$ where $\mathbf{P} \in \mathbb{R}^{2 \times 3}$ and $\mathbf{t} \in \mathbb{R}^2$. In this notation, the image point $\mathbf{x}_k^i \in \mathbb{R}^2$ that results from projecting scene point $\mathbf{X}_k \in \mathbb{R}^3$ into image i is given by

$$\mathbf{x}_k^i = \mathbf{P}_i\mathbf{X}_k + \mathbf{t}_i. \qquad (9)$$

The matrix $\mathbf{P}$ can be decomposed as

$$\mathbf{P} = \left[\begin{array}{cc} \alpha_x & s \\ 0 & \alpha_y \end{array} \right] \left[\begin{array}{c} \mathbf{r}_1^\mathsf{T} \\ \mathbf{r}_2^\mathsf{T} \end{array} \right] \qquad (10)$$

where $\mathbf{r}_1^\mathsf{T}$ and $\mathbf{r}_2^\mathsf{T}$ are the first two rows of a rotation matrix that describes the camera orientation in the world coordinate system, s is the pixel skew, and α_x and α_y are the horizontal and vertical scale factors. (The aspect ratio is given by α_x/α_y.) Also, given a matrix $\mathbf{P}$, the viewing direction (and in the present case, the source direction) in the world coordinate system is given by the unit vector in the negative direction of the cross product of the two row vectors $\mathbf{p}_1^\mathsf{T}$ and $\mathbf{p}_2^\mathsf{T}$. That is,

$$\hat{\mathbf{s}} = -\frac{\mathbf{p}_1^\mathsf{T} \times \mathbf{p}_2^\mathsf{T}}{|\mathbf{p}_1^\mathsf{T} \times \mathbf{p}_2^\mathsf{T}|}. \qquad (11)$$

It is well known that by observing $N \geq 4$ non-coplanar rigid[3] points over two or more unknown affine views, one can establish an affine coordinate frame, and thereby obtain the scene points and the cameras up to an unknown affine transformation [8, 14]. In the present case, this reconstruction includes the (affine) source directions, since they are equivalent to the viewing directions. We can show that given a sufficient number of camera/source positions, knowledge of these directions enables a unique solution to the calibration-estimation problem of Sect. 3.

Each known source direction $\hat{\mathbf{s}}_i$ gives two linear constraints on the matrix $\mathbf{Q}$ in (8), since for each we have $\hat{\mathbf{s}}_i = \alpha_i\mathbf{Q}\tilde{\mathbf{s}}_i$ for some $\alpha_i > 0$. As noted in Sect. 3, $\mathbf{Q}$ is only defined up to scale, and thus has eight degrees of freedom. It follows that in general, $\mathbf{Q}$ can be uniquely recovered given $M \geq 4$ camera/source positions by solving the corresponding constrained linear system of equations.

To summarize, given $N \geq 4$ observed points over $M \geq 4$ cameras/sources in general position, we can obtain the

---

[3]The general affine camera model allows smooth non-rigid transformations of the scene points between views (see [8].) Here we assume fixed source positions, however, and in order to make use of the photometric constraints in our system, we require that the scene be rigid.

cameras (and source directions), the scene points, the source strengths, and the surface normals at the observed points. All of this information is in an affine coordinate frame. (See Fig. 3 for an example.) In Sect. 5, we discuss ways in which we can upgrade to a metric reconstruction.

Note that if we have fewer views (if $M = 3$) we can still establish correspondence using (4), and we can still establish an affine reconstruction of the cameras and the observed points. We cannot, however, determine the source strengths or the surface normals without further information.

5. Metric Reconstruction

The reconstruction obtained in the previous section differs from a metric reconstruction (i.e., up to scale) by an unknown affine transformation $\mathbf{A} \in GL(3)$. The problem of "upgrading" the reconstruction to a metric one is thus the problem of estimating the nine parameters of this transformation, commonly termed the autocalibration problem. In order to solve this problem, we require more information, either geometric or photometric, about the system.

Autocalibration is well studied, and numerous techniques exist for obtaining a metric reconstruction using constraints on the intrinsic parameters of the cameras (see [6].) For example, we can obtain a metric reconstruction if we know the aspect ratio of the cameras (e.g., the pixels are square) and there is no pixel skew (valid for CCD cameras.) With this knowledge in hand, metric reconstruction follows directly from the methods in Sect. 4 with no extra work, since in this case, structure from motion yields a metric reconstruction of the points and cameras [8]. Then, the Euclidean source directions can be used to resolve $\mathbf{Q}$, yielding the Euclidean surface normals and the relative source strengths. This is perhaps the most practical method, and the one we use in Sect. 6.

In addition to knowledge about that camera parameters, we can also make use of photometric information about the system. For example, knowledge of the relative strength of the light sources was used by Hayakawa [7] to partially resolve the ambiguity in the calibration-estimation problem for uncalibrated Lambertian photometric stereo. In that paper, it was shown that knowledge of the relative strength of six sources was enough to resolve the surface normals up to a unitary transformation [7, 13]. Similar analysis could be applied here. As another example, if the BRDF of the surface is known to be highly peaked in the specular direction, we can use detected specularities to constrain $\mathbf{A}$. If a specularity is detected at the projection of the scene point $\mathbf{X}$ in the reciprocal images corresponding to camera/source directions $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{s}}_j$, it follows that the normal at that point must bisect these two directions, or $\hat{\mathbf{n}} = (\hat{\mathbf{s}}_i + \hat{\mathbf{s}}_j)/2$. Detected specularities have already been used to reduce ambiguities in photometric stereo and affine binocular stereo [2].

Any of these geometric or photometric constraints can be combined to determine the unknown affine transformation and thereby upgrade to a metric reconstruction. All of these techniques and their combinations are relevant to this paper, but an exhaustive list of the possibilities is outside of its scope. In the next section, we provide a demonstration of one method of resolving $\mathbf{A}$ and obtaining a metric reconstruction.

6. Implementation and Results

For the results in this section, we acquired reciprocal pairs of images using a single 12-bit Kodak DCS760 camera (whose radiometric response function is linear) and multiple Lowel Pro-light sources with 250-watt halogen bulbs. M separate sources were placed at each of M camera/source positions which were far (~ 2.5 meters) from the scene, and the collection process was as follows. First, the camera replaced the source at position 1, and $M-1$ images were captured, each with illumination by one of the other sources. Then, the original source was returned to position 1, and the camera replaced the source at position 2. The process was repeated, resulting in a total of $M(M-1)$ images, or $M(M-1)/2$ reciprocal pairs. An example of a typical dataset for $M = 8$ is shown in Fig. 1.

6.1. Affine

The first step of the reconstruction is the estimation of the epipolar geometry. Currently, this is done by clicking a small number of points, although it could be achieved using a robust structure from motion technique (e.g., [1].) Given $N \geq 4$ correspondences over all M views, we compute the points $\mathbf{X}_1 \dots \mathbf{X}_N$ and cameras $[\mathbf{P}_1, \mathbf{t}_1], \dots, [\mathbf{P}_M, \mathbf{t}_M]$ using the Tomasi-Kanade factorization algorithm [15] generalized to affine cameras as in [14, 16]. The corresponding source directions $\hat{\mathbf{s}}_i$ are computed using (11).

Having established the epipolar geometry, we use the uncalibrated distant-source reciprocity constraint (4) to establish dense correspondence in the multinocular system. This is accomplished by an exhaustive search over disparity space as follows. One camera is selected to serve as the principle view, and the scene space is parameterized by the disparity $d \in [0, d_{max}]$ between this and one other view. For a given point $\mathbf{x}$ in the principle view, and for a given disparity d, we sample all $M(M - 1)$ images at the corresponding points in order to build $\mathbf{E}_\mathbf{x}(d)$ as in (4). As a measure of the likelihood that $\text{rank}\,\mathbf{E}_\mathbf{x}(d) = M-1$ we use the ratio of its two smallest singular values,

$$r_\mathbf{x}(d) = \frac{\sigma_{M-1}}{\sigma_M}. \tag{12}$$

Note that there are a number of possible measures of numerical rank that could be used and that could be effective.

Since the constraint (4) is necessary for a valid disparity value but not sufficient, we use simple windowing to make

Figure 2. The columns of the $\mathbf{W}$ matrix defined in (7) displayed as images. These correspond to images of the surface with the reflectance removed.

(a)   (b)

Figure 3. Surface normals seen from the viewpoint of the principle camera. (a) Normal field that results from the affine reconstruction as described in Sect. 4. (b) That which results from enforcing known aspect ratio and zero skew in all cameras.

the disparity search more robust. For a given point $\mathbf{x}_\circ$ in the principle view and a disparity value d, we compute the ratio $r_\mathbf{x}(d)$ at this point as well as at points in a small window W_r around $\mathbf{x}_\circ$. Then, the estimated disparity at this point is given by

$$d^\star = \arg \max_d \sum_{\mathbf{x} \in W_r} r_\mathbf{x}(d). \qquad (13)$$

Once we have estimated the disparity $d^\star$, the corresponding null vector $\mathbf{w}_{\mathbf{x}_\circ}$ can be obtained as the linear least-squares estimate,

$$\mathbf{w}_{\mathbf{x}_\circ}^\star = \arg \min_\mathbf{w} \|\mathbf{E}_{\mathbf{x}_\circ}(d^\star)\mathbf{w}\|^2, \quad \|\mathbf{w}\| = 1, \qquad (14)$$

which is simply given by the right singular vector corresponding to the smallest singular value of $\mathbf{E}_{\mathbf{x}_\circ}(d^\star)$. However, due to sampling noise, and noise in the estimation of the epipolar geometry and the depth estimate, the irradiance measurements in $\mathbf{E}_{\mathbf{x}_\circ}(d^\star)$ are in general noisy. This affects the accuracy of the estimated null vector $\mathbf{w}_{\mathbf{x}_\circ}^\star$, especially when these measurements are made near discontinuities in scene radiance (e.g., at albedo edges.) In order to mitigate these effects, we use a second (weighted) windowing scheme. Given the disparity estimates at and near $\mathbf{x}_\circ$, we construct the $\mathbf{E}$ matrices at $\mathbf{x}_\circ$ and in a neighborhood $W_\mathbf{E}$ around it. The null vector at $\mathbf{x}_\circ$ is estimated using

$$\mathbf{w}_{\mathbf{x}_\circ}^\star = \arg \min_\mathbf{w} \sum_{\mathbf{x} \in W_\mathbf{E}} r_\mathbf{x} \|\mathbf{E}_\mathbf{x}\mathbf{w}\|^2, \quad \|\mathbf{w}\| = 1, \qquad (15)$$

where $r_\mathbf{x}$ is given by (12). Again, this calculation can be done using SVD, since $\mathbf{w}_{\mathbf{x}_\circ}^\star$ is simply the null vector of an augmented matrix created by stacking the weighted $\mathbf{E}_\mathbf{x}$ matrices vertically.

To summarize, we use a two-pass technique to establish correspondence and to estimate the null vector $\mathbf{w}_\mathbf{x}$ for each
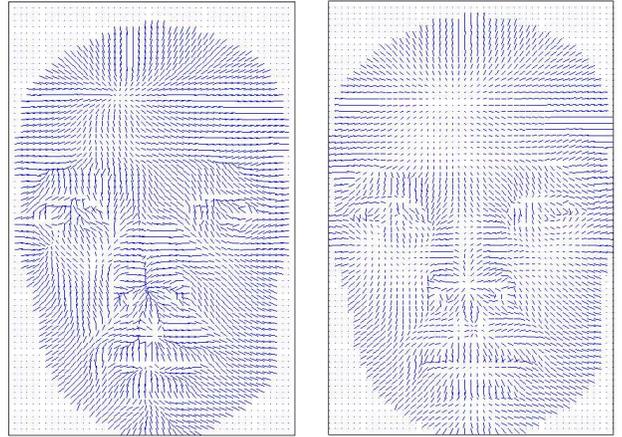
point $\mathbf{x}$ in the principle view. In the first pass, the ratio (12) is computed and stored for each $\mathbf{x}$ and for each possible disparity. By maximizing (with respect to disparity) this ratio summed over a window in the principle view, we estimate the true disparity associated with each point. In the second pass, we use these established disparity estimates and the associated ratios to compute an estimate of the null vectors over a second weighted window. Note that these two steps need not be performed at the same resolution.

At this point, we have a null vector $\mathbf{w}_\mathbf{x}$ for every point in the principle view[4], and we can construct the matrix $\mathbf{W}$ as in (7). The columns of this matrix for the dataset in Fig. 1 are shown in Fig. 2. As mentioned previously, these columns of $\mathbf{W}$ correspond to images of the surface with the original reflectance removed. In order to recover the surface normals and the source strengths as described in Sect. 4, we compute a preliminary factorization of $\mathbf{W}$ and then resolve the ambiguity in the factorization using the known affine source directions $\hat{\mathbf{s}}_i$. As a result of this procedure, we obtain the field of surface normals as shown in Fig. 3(a). For display purposes, we have transformed the world coordinate system so that the source direction corresponding to the principle view is $[0 \; 0 \; -1]$. (There is an entire family of such transformations, each leading to a different representation of the surface. We simply chose one.) The result of integrating this normal field (after enforcing integrability [5]) is shown in the top row of Fig. 4.

For this result, we used the images shown in Fig. 1. The camera corresponding to the second row of images in that figure was used as the principle view. Twenty correspon-

---

[4]For the results in this section, the background was manually identified so that points $\mathbf{x}$ backprojecting to rays that did not intersect the surface were ignored.
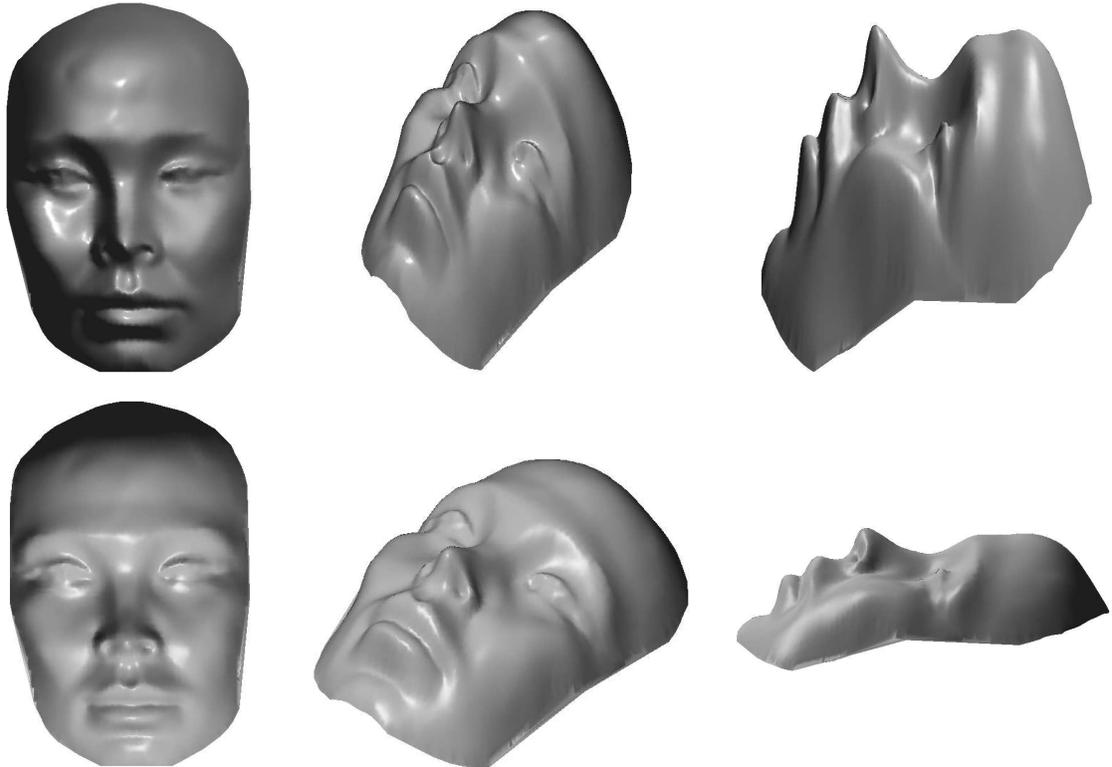
Figure 4. Three views of the two surfaces that result from integrating the normal fields in Fig. 3. The top row corresponds to the affine reconstruction (see Fig. 3(a)), and the bottom row the metric reconstruction (see Fig. 3(b)).

dences were used to establish the epipolar geometry, the resolution of the reconstruction was 153×105, and we used square windows with $W_r = 9 \times 9$, and $W_{\mathbf{E}} = 3 \times 3$.

6.2. Metric

In order to obtain a metric reconstruction, we make the assumption that the cameras have unit aspect ratio and zero skew, and the procedure of the previous section is repeated with only minor changes. We take the set of affine points $\mathbf{X}_k$ and cameras $[\mathbf{P}_i, \mathbf{t}_i]$ and enforce known aspect ratio and zero skew by finding $\mathbf{G} \in GL(3)$ that satisfies the set of quadratic constraints [14, 16]

$$\mathbf{p}_{i1}^{\mathsf{T}}\mathbf{G}\mathbf{G}^{\mathsf{T}}\mathbf{p}_{i1} = \mathbf{p}_{i2}^{\mathsf{T}}\mathbf{G}\mathbf{G}^{\mathsf{T}}\mathbf{p}_{i2}$$
$$\mathbf{p}_{i1}^{\mathsf{T}}\mathbf{G}\mathbf{G}^{\mathsf{T}}\mathbf{p}_{i2} = 0,$$

where $\mathbf{p}_{i1}^{\mathsf{T}}$ and $\mathbf{p}_{i2}^{\mathsf{T}}$ are the rows of the ith camera matrix $\mathbf{P}_i$. By applying the transformation $\mathbf{G}$ to the system (the cameras transform as $\mathbf{PG}$ and the points as $\mathbf{G}^{-1}\mathbf{X}$), we obtain the points and cameras in a metric coordinate frame, and we can compute the true Euclidean source directions $\hat{\mathbf{s}}_i$.

Since the correspondences remain unchanged, the matrix $\mathbf{W}$ need not be recomputed. We simply resolve the ambiguity in the factorization using the Euclidean source directions in place of the affine directions. The resulting normal field is shown in Fig. 3(b) and the integrated surface (after enforcing integrability [5]) is shown in the bottom row of Fig. 4, and in Fig. 5.

7. Conclusion

This paper takes a stratified look at uncalibrated Helmholtz stereopsis (i.e., where we have little or no information about the cameras and light sources.) We derive a photometric matching constraint that can be used to establish correspondence without knowledge of the cameras and sources and thus obtain a projective reconstruction of the observed points. As in the original formulation of Helmholtz stereopsis, this new constraint has the important property of not assuming a reflectance model for the surfaces in the scene being reconstructed.

Next, we show that when the distance from the scene to the cameras/sources is large, we can obtain an affine reconstruction that (in addition to the observed points) includes the surface normal at each observed point and the relative strength of the light sources.

Finally, we discuss ways in which further information about the cameras and sources can be used to upgrade from this affine reconstruction to a metric reconstruction.

This paper leaves a number of open questions. First, we do not explore the near-source (perspective) case in detail; it is likely that more information about the scene is obtainable in this case, and this remains a topic for future work. Second, although we discuss some techniques for achieving the affine to metric upgrade, an exploration of the many possible methods is outside the scope of this paper.

Figure 5. The same surface shown in the bottom row of Fig. 4, but interpolated and texture mapped with one input image.

Appendix

In this appendix, we compare the two correspondence constraints discussed in Sect. 2. Consider a reciprocal imaging system consisting of P reciprocal pairs captured from $M \leq P$ positions. If the system is calibrated, the matrix equation (2) can be used to establish correspondence, since for a valid correspondence, the matrix $\mathbf{M}$ will satisfy rank $\mathbf{M} < 3$. In the uncalibrated case, we cannot compute $\mathbf{M}$, and instead we factor it as

$$\mathbf{M}_{P \times 3} = \mathbf{E}_{P \times M} \mathbf{U}_{M \times 3}, \qquad (16)$$

where $\mathbf{U}\hat{\mathbf{n}} = \tilde{\mathbf{w}}$, with $\tilde{\mathbf{w}}$ as defined in (3). Then, the uncalibrated constraint that we use for correspondence is rank $\mathbf{E} < M$. Here we show that for $M \geq 4$ this is a weaker constraint than that based on (2).

There are two relevant rank inequalities for a general matrix product,

$$\mathrm{rank}(\mathbf{A}\mathbf{B}) \leq \min(\mathrm{rank}\,\mathbf{A}, \mathrm{rank}\,\mathbf{B})$$
$$\mathrm{rank}\,\mathbf{A}_{n \times k}\mathbf{B}_{k \times m} \geq \mathrm{rank}\,\mathbf{A} + \mathrm{rank}\,\mathbf{B} - k.$$

Applying these to the present case (and noting that rank $\mathbf{U} = 3$ for non-coplanar sources) we get

$$\mathrm{rank}\,\mathbf{E} + 3 - M \leq \mathrm{rank}\,\mathbf{M} \leq \min(\mathrm{rank}\,\mathbf{E}, 3),$$

which tells us the following. First, if $M = 3$, rank $\mathbf{E} < M \iff$ rank $\mathbf{M} < 3$, and the two correspondence constraints are equivalent. They are not equivalent, however, for $M \geq 4$. In this case, rank $\mathbf{M} < 3 \implies$ rank $\mathbf{E} < M$, but the converse does not hold.

References

[1] P. A. Beardsley, P. H. S. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In *Proc. European Conf. on Computer Vision*, pages 683–695, 1996.

[2] O. Drbohlav and R. Šára. Specularities reduce ambiguity of uncalibrated photometric stereo. In *Proc. European Conf. on Computer Vision*, pages II: 46–60, 2002.

[3] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. European Conf. on Computer Vision*, pages 563–578, 1992.

[4] O. Faugeras. Stratification of 3-D vision: Projective, affine, and metric representations. *J. Optical Society of America A*, 12(7):465–484, 1995.

[5] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[7] K. Hayakawa. Photometric stereo under a light source with arbitrary motion. *J. Opt Soc. Am.*, 11(11), 1994.

[8] J. Koenderink and A. van Doorn. Affine structure from motion. *J. Optical Society of America A*, 8(2):377–385, February 1991.

[9] J. J. Koenderink and A. J. van Doorn. The generic bilinear calibration-estimation problem. *Int. Journal of Computer Vision*, 23(3):217–234, 1997.

[10] J. Lu and J. Little. Reflectance and shape from images using a collinear light source. *Int. Journal of Computer Vision*, 32(3):1–28, 1999.

[11] S. Magda, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond Lambert: Reconstructing surfaces with arbitrary BRDFs. In *Proc. Int. Conf. on Computer Vision*, pages II: 391–398, 2001.

[12] J. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[13] R. Rosenholtz and J. J. Koenderink. Affine structure and photometry. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 790–795, June 1996.

[14] L. S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *Int. Journal of Computer Vision*, 16:147–182, 1995.

[15] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *Int. Journal of Computer Vision*, 9(2):137–154, 1992.

[16] D. Weinshall. Model-based invariants for 3-D vision. *Int. Journal of Computer Vision*, 10(1):27–42, 1993.

[17] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *Proc. European Conf. on Computer Vision*, pages III: 869–884, May 2002.