

Reciprocal Image Features for Uncalibrated Helmholtz Stereopsis

Todd Zickler

Division of Engineering and Applied Sciences, Harvard University

zickler@eecs.harvard.edu

Abstract

Helmholtz stereopsis is a surface reconstruction method that exploits reciprocity for the recovery of 3D shape without an assumed BRDF model, and it has been shown to yield high quality results when the cameras and light sources are carefully calibrated. In many practical cases, however, accurate off-line calibration is difficult (or impossible) to achieve. We address this issue by exploring wide-baseline matching in Helmholtz stereo images. We identify two classes of local image interest regions ('features') that can be reliably detected and matched between views in a Helmholtz stereo dataset; and by exploiting reciprocity, we show how these regions can be used to recover both geometric and radiometric calibration information. When used in conjunction with existing methods for dense reconstruction, this provides an automated shape recovery pipeline that operates independent of reflectance and does not require the acquisition of additional calibration images off-line.

1. Introduction

Image-based modeling systems are designed to accurately recover the shape (and often reflectance) of a scene for tasks such as visual inspection, reverse engineering, digital object archival, virtual reality and enrollment for 3D face/object recognition systems. A large variety of suitable shape-recovery techniques exist, including view-based methods such as stereo and structure-from-motion and lighting-based methods such as photometric stereo. Helmholtz stereopsis [6, 16] is an image-based shape recovery method that combines the benefits of photometric reconstruction—direct estimation of surface normals—with those of traditional geometric methods like stereo and structure-from-motion (*e.g.*, identification of depth discontinuities.) The advantage of Helmholtz stereopsis is that it decouples the shape and reflectance information in images, enabling the recovery of surface shape without an assumed BRDF model. As a result, it can be used to recover accurate surface shape for a very broad class of surfaces.

Helmholtz stereopsis is based on the analysis of *reciprocal image pairs* (Fig. 2), and it has been shown to yield high

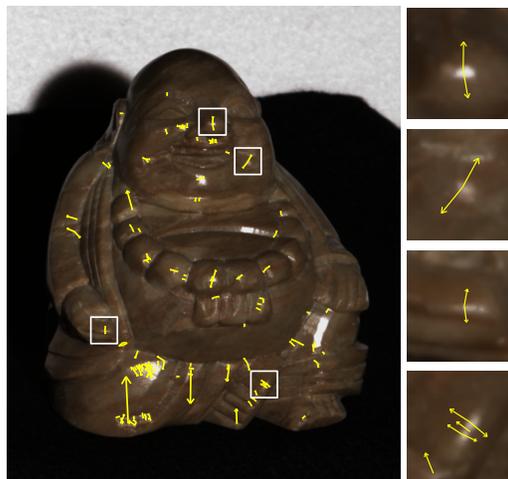


Figure 1. Induced image features for robust matching. Due to reciprocity, specular highlights are fixed to the surface and serve as stable image features in a Helmholtz stereo pair. They provide a means to recover the epipolar geometry from textureless surfaces which are void of ‘intrinsic’ features.

quality results from two [13, 18] or more [2, 16] such pairs when the cameras and light sources are carefully calibrated. But for many modeling applications, it is impossible—or at least impractical—to maintain a controlled acquisition environment. When digitizing cultural artifacts, for example, or for large-scale enrollment in face recognition systems, images often need to be acquired on-site. For Helmholtz stereopsis to be a practical modeling tool in these conditions, the need for off-line calibration must be addressed, and auto-calibration techniques must be developed.

Helmholtz stereo reconstruction with only partial calibration information has been explored in previous work [17], but it has been assumed that the epipolar geometry of the system is known. The pre-requisite step of finding the epipolar geometry has not yet been investigated. In this paper, we move closer to a complete auto-calibration system by investigating the use of interest regions—regions that can be reliably identified and matched across wide baselines—for recovering the epipolar geometry and radiometric calibration of a reciprocal pair of images.

We show that two types of interest regions can be dis-

tinguished. First, for surfaces with significant spatial reflectance variation (‘texture’), we can detect and match affine covariant regions using techniques similar to those used in conventional stereo (*e.g.*, [7, 14]). An important difference, however, is that due to reciprocity, intensities in corresponding regions are equal up to a scale factor—a property that does not hold in conventional stereo images unless the surface is Lambertian. (See Fig. 3.)

The second type of interest region is not texture-based, and is unique to reciprocal image pairs. For specular surfaces without texture, the specular highlights themselves provide stable interest regions, since these highlights remain fixed to the surface and can be used for matching. (See Fig. 1.) In addition to providing information for geometric calibration, this second type of interest region provides partial radiometric calibration information in the form of the relative light source strengths¹ This radiometric information is needed for dense matching [17], but until now, obtaining it has required an off-line calibration procedure [3].

When used in conjunction with existing methods for dense reconstruction [2, 13, 16–18], the techniques developed in this paper provide an automated system for image-based shape recovery using Helmholtz stereopsis.

2. Background

To review Helmholtz stereopsis, we follow [6, 16] and write the *reciprocity constraint* relating radiance observations at corresponding image points in a reciprocal pair of images:

$$\left(e_l(\mathbf{x}) \frac{s_l \hat{\mathbf{v}}_l(\mathbf{x})}{|\mathbf{o}_l - \mathbf{x}|^2} - e_r(\mathbf{x}) \frac{s_r \hat{\mathbf{v}}_r(\mathbf{x})}{|\mathbf{o}_r - \mathbf{x}|^2} \right) \cdot \hat{\mathbf{n}}(\mathbf{x}) = 0. \quad (1)$$

Here, $e_l(\mathbf{x})$ and $e_r(\mathbf{x})$ are the scene radiance values measured at the left- and right-image projections of a surface point $\mathbf{x} \in \mathbb{R}^3$. (These are abbreviations for the composite functions $e_i(\pi_i(\mathbf{x}))$ where $\pi_i: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is the projection of the i^{th} camera, and $e_i: E_i \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is the scene radiance measured on the finite domain E_i of the image plane.) As shown in Fig. 2, $\hat{\mathbf{n}}(\mathbf{x})$ is the surface normal at \mathbf{x} , and the unit vectors $\hat{\mathbf{v}}_l$ and $\hat{\mathbf{v}}_r$ are the directions from \mathbf{x} to the camera/source positions \mathbf{o}_l and \mathbf{o}_r . The scalar values s_r and s_l are the strengths of the two isotropic point light sources. (Anisotropic illumination can also be modeled [3].)

The two main observations regarding Eq. 1 are: 1) due to Helmholtz reciprocity, it does not depend on the surface BRDF, and 2) it provides direct information about the surface normal in addition to the depth at each surface point.

When the distance from the scene to the cameras/sources is large relative to the relief of the scene, the illumination approaches that of directional sources, and the cameras can be represented using a parallel-projection model.

¹Complete radiometric calibration includes the radiometric response functions of the cameras. Here, we assume these functions to be either linear or known.

Under these conditions, the values $s_l \hat{\mathbf{v}}_l(\mathbf{x})/|\mathbf{o}_l - \mathbf{x}|^2$ and $s_r \hat{\mathbf{v}}_r(\mathbf{x})/|\mathbf{o}_r - \mathbf{x}|^2$ are independent of \mathbf{x} , and Eq. 1 reduces to the *far-field reciprocity constraint* [17, 18],

$$(e_l(\mathbf{x})s_l\hat{\mathbf{v}}_l - e_r(\mathbf{x})s_r\hat{\mathbf{v}}_r) \cdot \hat{\mathbf{n}}(\mathbf{x}) = 0. \quad (2)$$

(We have abused our notation slightly by appropriately re-defining s_l and s_r .)

Since they relate the observed intensities to surface information, Eqs. 1 and 2 can be used for surface reconstruction [2, 6, 13, 16, 18]. But all of these methods require that the cameras are geometrically and radiometrically calibrated *a priori*. Geometric calibration provides the intrinsic and extrinsic camera parameters, and radiometric calibration provides: 1) the radiometric response functions of the cameras (not required if they are linear and equal), 2) the optical fall-off and vignetting effects in the cameras, 3) the relative strength of the light sources, and 4) the light source anisotropy. Techniques for target-based geometric calibration are generally the same as those used for conventional stereo systems (*e.g.*, [15]), and an elegant technique for radiometric calibration is presented by Janko *et al.* [3]; but these methods require precise calibration targets and the careful acquisition of additional calibration images off-line.

Zickler *et al.* [17] show that the constraints in Eqs. 1 and 2 can be adapted so that dense correspondence can be achieved with limited radiometric calibration provided that the epipolar geometry is known. (The pre-requisite recovery of the epipolar geometry was not explored.) In the far-field case, this approach yields a dense reconstruction—including surface normals and relative source strengths—that is defined up to an affine transformation of the world.

2.1. Regions of Local Interest

Most methods for recovering the epipolar geometry from a set of two or more images of a static scene are based on the detection and matching of local interest regions or ‘features’. In addition to robust matching across large variations in viewpoint, these techniques can be useful for object recognition, image databases, and a variety of other applications. There is a large body of work in this area, only a small portion of which is discussed here. For more information, the reader is referred to a recent article by Lowe [5] and the survey articles by Mikolajczyk *et al.* [9, 11].

There are typically three stages in estimating epipolar geometry from interest regions. The first step is to detect these regions in a manner that is invariant under similarity (or affine) transformations, for example, by locating extrema of a derivative-based operator in scale-space [4, 5]. The output of this detection stage is typically the location of the region center as well as its ‘characteristic scale’. Once candidate regions have been detected, the second stage consists of establishing correspondence between regions in one

image and those in another. This is accomplished by computing a descriptor that summarizes the region in a similarity or affine invariant manner. In experiments conducted by Mikolajczyk and Schmid [9], the SIFT descriptor [5] has been shown to outperform a number of others. It is a 128-element vector computed from the spatial distribution of image gradients in the region of interest. (To make this descriptor invariant under rotations, the region is first rotated according to its dominant rotation.)

In the third stage, incorrect matches are discarded using a robust estimation scheme such as RANSAC, using a seven-point computation of the fundamental matrix [12].

3. Reciprocal Image Interest Regions

As discussed in Sect. 2, the relation between intensities at corresponding points in a reciprocal pair of images is independent of the surface reflectance. In this section, we show how this property can be exploited to reliably detect and accurately match interest regions. We distinguish between two types of regions and discuss each separately.

3.1. Spatial Variation (Type I)

Consider a planar region $\Omega \subset \mathbb{R}^2$ on a surface, and suppose we acquire a reciprocal image pair of this surface as shown in Fig. 2. When it is sufficiently small, the relation between the left and right projections of this region is well represented by an affinity. That is, there exists an affine transformation $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ satisfying $A(\pi_l(\mathbf{u})) = \pi_r(\mathbf{u}) \forall \mathbf{u} \in \Omega$. Letting $\omega_l \subset E_l = \{\pi_l(\mathbf{u}), \mathbf{u} \in \Omega\}$ and $\omega_r \subset E_r = \{\pi_r(\mathbf{u}), \mathbf{u} \in \Omega\}$ represent the left and right projections of Ω , we write this geometric relationship as

$$\omega_l \leftrightarrow A(\omega_r). \quad (3)$$

Since Ω is small, the directions ($\hat{\mathbf{v}}_l$ and $\hat{\mathbf{v}}_r$) from the surface to the camera/source positions are constant over the region, and from Eq. 1 we have

$$e_l(\omega_l) = \beta_1 e_r(\omega_r), \quad (4)$$

where

$$\beta_1 = \frac{s_r \hat{\mathbf{v}}_r^\top \hat{\mathbf{n}}}{s_l \hat{\mathbf{v}}_l^\top \hat{\mathbf{n}}} \text{ or } \beta_1 = \frac{s_r \hat{\mathbf{v}}_r^\top \hat{\mathbf{n}} \|\mathbf{o}_l - \mathbf{x}\|^2}{s_l \hat{\mathbf{v}}_l^\top \hat{\mathbf{n}} \|\mathbf{o}_r - \mathbf{x}\|^2},$$

for the far-field or near-field case, respectively. Here, $\hat{\mathbf{n}}$ is the normal to the planar region Ω .

Equation 4 states that the radiance measured at each point in the region ω_r of the right image is equal to the corresponding point in the left image up to a common scale factor. By reciprocity, this is true for any planar region, even those with arbitrary BRDFs that vary—perhaps discontinuously—from point to point. Note that this is not

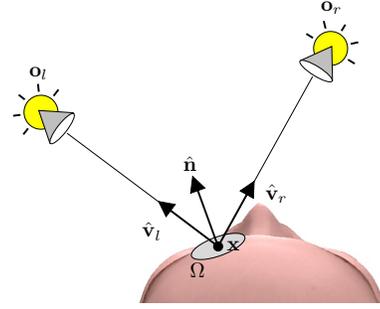


Figure 2. A reciprocal pair of images of a small planar region. Radiance measured at the left and right projections of the planar region Ω are related by an affine transformation and a single scale factor. (See Eq. 4.) This relationship is independent of the BRDF.

true for a conventional stereo pair unless the BRDF at each point is constant (*i.e.*, the surface is Lambertian.)

The distinction between reciprocal and conventional stereo images is demonstrated by an extreme example in Fig. 3, which shows images of a planar surface with a spatially-varying BRDF. The top-left shows a conventional (fixed-illumination) stereo pair, and the top-right shows a reciprocal image pair captured from the same two viewpoints. For each pair, the bottom row shows the difference $e_l(\omega_l) - \beta_1 e_r(A(\omega_r))$. (Here, the affine transformation A is computed from manually identified correspondences, and β_1 is chosen to minimize the difference.)

In the conventional stereo pair in the left of Fig. 3, we see that the painted regions of the surface are nearly Lambertian, and their emitted radiance does not change significantly with viewpoint. The unpainted regions, however, are highly specular, and extreme variation is observed between views. Detecting and matching regions in cases of such extreme spatial variation is difficult—if not impossible—under the fixed illumination conditions of conventional stereo, even when using methods that are invariant to uniform changes in brightness and contrast.

The right of Fig. 3 demonstrates that reciprocal images do not suffer from this same limitation, since the radiance values observed in corresponding regions ω_l and ω_r are related by Eq. 4, regardless of the reflectance. For this reason, most existing region detectors and descriptors that have been developed for conventional stereo and SFM (*e.g.*, [5, 7, 10, 14]) can be used directly on reciprocal images without change. Indeed, as this example suggests, we can only expect their performance to improve relative to that under fixed-illumination conditions. All that is required is that they be invariant to changes in contrast (*i.e.*, a global scaling of illumination), which is generally the case.

As mentioned in Sect. 2, there is a vast body of existing work in this area, and we refer the reader to those references for further details. The next section discusses a second type of interest region that is unique to reciprocal images.



Figure 3. Top row: conventional (left) and reciprocal (right) stereo pairs of a planar surface with spatially-varying BRDF. Bottom: Difference images after a global scaling of intensity and geometrical alignment by an affine transformation. In the reciprocal case, the aligned images are equal up to scale. This is not true in the conventional case unless the surface is Lambertian.

3.2. Specular Highlights (Type II)

It has been observed that highlights due to specular reflection appear ‘fixed to the surface’ in a reciprocal image pair [16]. This section explores this idea, showing that specular highlights are indeed stable interest regions.

In reciprocal images of a curved, specular surface, highlights are observed near points at which the surface normal bisects the view and illumination directions $\hat{\mathbf{v}}_l$ and $\hat{\mathbf{v}}_r$. In the neighborhood of such a point, a smooth surface can be locally approximated by a patch $\mathbf{x}: \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$ of the form $\mathbf{x}(x, y) = (x, y, z(x, y))$, where

$$z(x, y) = -\frac{a}{2}x^2 - \frac{b}{2}y^2. \quad (5)$$

As shown in Fig. 4, we can choose the coordinate system such that the origin is at the point with bisecting surface normal, the plane $z = 0$ is aligned with the tangent plane at that point, and the x - and y -axes are aligned with the directions of principal curvature.

As in the previous section, the region Ω is assumed to be small so that the far-field approximation can be made locally, even if the cameras are perspective. In this case, the reciprocal directions are constant over the patch, and in the coordinate system of Fig. 4, they can be expressed as

$$\begin{aligned} \hat{\mathbf{v}}_l &= (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi) \\ \hat{\mathbf{v}}_r &= (-\cos \theta \sin \phi, -\sin \theta \sin \phi, \cos \phi), \end{aligned} \quad (6)$$

where 2ϕ is the vergence angle of the reciprocal pair.

Unlike the Type I case of the previous section, the surface is not planar, and there is no simple geometric relationship between corresponding sets of image points analogous to Eq. 3. As shown on the right of Fig. 4, however, a symmetry does exist because the imaging configurations are equivalent up to a 180° rotation about the z -axis. The surface point \mathbf{x} , for example, when illuminated from $\hat{\mathbf{v}}_l$ and viewed from $\hat{\mathbf{v}}_r$, emits the same radiance—assuming equal source strengths and an isotropic BRDF—as the point \mathbf{x}' in the reciprocal case. As a result, when the reflectance is spatially-invariant and isotropic² over the patch, the two im-

²In fact, anisotropic BRDFs are allowed provided that the orientations of the tangent vectors are consistent at symmetric surface points.

ages of this patch will ‘look the same’, even though points of equal radiance in the two images are not projections of the same surface point. This is true for arbitrary BRDFs, including those with ‘off-specular peaks’ for which the maximum intensity does not occur at the bisecting normal.

To state this relationship more precisely, let

$$\begin{aligned} \omega_l \subset E_l &= \{\pi_l(\mathbf{x}(x, y)), (x, y) \in \Omega\}, \\ \omega_r \subset E_r &= \{\pi_r(\mathbf{x}(x, y)), (x, y) \in \Omega\} \end{aligned}$$

represent the left and right projections of the surface patch. Then there exists a similarity transformation $S: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that

$$e_l(\omega_l) = \beta_2 e_r(S(\omega_r)), \quad (7)$$

where

$$\beta_2 = \frac{s_r}{s_l} \text{ or } \beta_2 = \frac{s_r |\mathbf{o}_l - \mathbf{x}|^2}{s_l |\mathbf{o}_r - \mathbf{x}|^2},$$

is the ratio of the (distance-weighted) source strengths for the far-field or near-field case, respectively. Note that in the strict sense, for a similarity transformation to be sufficient in Eq. 7, the cameras must be well-approximated locally by scaled-orthographic cameras as opposed to arbitrary affine cameras. But when this is true (as it typically is), the transformation S allows for changes in scale (e.g., due to zoom) and for rotations of the two cameras about their optical axes.

3.2.1 Region Correspondence

Figure 5 provides an example of how this relationship can be used to determine correspondence between detected highlight regions in a reciprocal image pair. The left-most panel in the figure shows highlights that are detected using a generic interest point detector [5]. For many of these regions, the surface behaves locally like the quadratic patch in Fig. 4, and as a result, corresponding regions in the two images will be similar in appearance as described by Eq. 7.

To determine the correct correspondence between regions, we can use any existing method that is invariant to similarity transformations and contrast variation. Here, we use the SIFT descriptor [5]. For each detected region, a 128-dimensional SIFT descriptor is computed using the appropriate orientation and scale, and for each descriptor, the

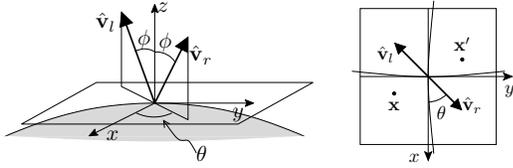


Figure 4. Side and top views of a surface patch in the neighborhood of a specular highlight. Reciprocal images of this patch are acquired from directions $\hat{\mathbf{v}}_l$ and $\hat{\mathbf{v}}_r$, and due to symmetry and BRDF isotropy, the two images are equal up to similarity transformation.

nearest neighbor (in term of Euclidean distance in descriptor space) in the other image is determined through an exhaustive search. Regions that do not yield a symmetric match are discarded, and the results of this procedure are shown in the center panel of Fig. 5. False matches are rare because the images of each highlight region depend on the local curvature of the surface which is often quite distinct.

3.2.2 Radiometric Calibration

Since the relationship between corresponding Type II regions (Eq. 7) is independent of the surface normal and camera/source directions, it provides a constraint for recovering the relative strengths of the light sources for radiometric calibration. In the general case, this relationship depends on the relative distances to the cameras/sources, and source strengths must be recovered iteratively in conjunction with the surface depth. In the far-field case, however, this dependence disappears, and each pair of matched Type II features provides a direct constraint on the source strengths.

For far-field cameras and sources, Eq. 7 states that the ratio of the source strengths is simply the inverse of the ratio of the radiance measurements of the two aligned regions. In practice, given a region correspondence as described in Sect. 3.2.1, the value of β_2 can be recovered by minimizing

$$C(\mathbf{p}, \beta_2) = \sum_{\mathbf{u} \in \omega_l} (e_l(\mathbf{u}) - \beta_2 e_r(S(\mathbf{u}; \mathbf{p})))^2, \quad (8)$$

where \mathbf{p} is a vector of four parameters defining a similarity transformation. This optimization can be computed very efficiently using established methods for alignment [1].

Each matched Type II region provides an estimate of the relative source strength, and if the sources are assumed to be isotropic (or uniform directional sources), one can simply take the average of these estimates. More general models of source variation are possible, however, and by choosing a basis for the source anisotropy, more accurate radiometric calibration information can be recovered. Janko *et al.* [3] show that both anisotropic sources and optical fall-off in the cameras can be modeled using functions $s_l(\mathbf{u}_l)$ and $s_r(\mathbf{u}_r)$ defined on the image planes. Based on this observation, if these functions are chosen to be suitably smooth (e.g., polynomials of low order), they could be uniquely re-

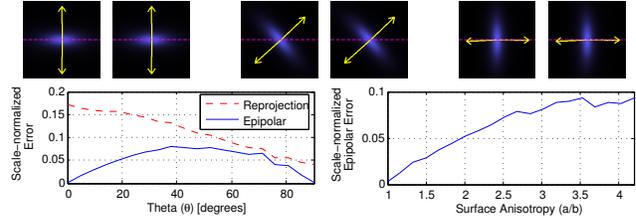


Figure 6. Error in point matches for synthetic quadratic patches (Eq. 5) rendered with the measured metallic-blue BRDF [8]. Top: Three of the reciprocal pairs ($a/b = 3$, $\theta = 0^\circ, 45^\circ, 90^\circ$) used as input. Left: reprojection and epipolar errors (normalized by feature scale) for patch with $a/b = 3$ at different orientations relative to the epipolar plane. Right: epipolar error for patches with $\theta = 45^\circ$ and increasing surface anisotropy a/b .

covered from the relative source strength measurements obtained from a small number of matched highlight regions.

For the data used in this paper, we found constant source strengths to be sufficiently accurate, and leave the analysis of more general bases for source anisotropy for future work.

4. Evaluation

The previous section introduced two types of interest regions, one that is similar to regions used for conventional stereo and another that is unique to Helmholtz stereopsis. Since Type I regions have been well-studied in the context of conventional stereo, we restrict the evaluation in this paper to Type II regions, and we examine their utility for both geometric and radiometric calibration.

4.1. Synthetic Data

As shown in Sect. 3.2.1, for matched Type II regions, left and right image points of equal radiance do not necessarily correspond to projections of the same surface point. These ‘region correspondences’ are sufficient for the purpose of radiometric calibration, but in order to estimate epipolar geometry, we are forced to select a unique pair of corresponding points within matched regions. A correct choice cannot be made without additional knowledge about the surface BRDF, but such knowledge is rarely available. Instead, we investigated the efficacy of one particular choice that does not require this knowledge.

We choose points within each Type II region using the extremum of the difference-of-Gaussian function in scale space, and as suggested by Lowe [5], we localize the extrema by fitting a 3D quadratic function. To assess the geometric calibration error induced by this choice, we performed a synthetic experiment with rendered far-field reciprocal pairs of quadratic patches using the metallic-blue BRDF from the MIT/MERL database [8] and a vergence angle (2ψ) of 40° . Three example pairs are in the top of Fig. 4 with the extrema points (base of the arrows), extrema scales (one-sixth of arrow lengths), and epipolar lines all

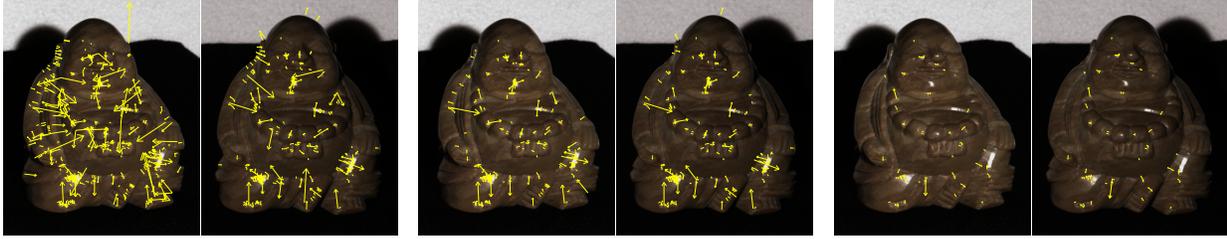


Figure 5. Specular highlight matching. Left: Initial interest regions detected as extrema in scale space. Middle: Region correspondence obtained using SIFT descriptors. Right: Inlier point matches obtained following local maxima detection and RANSAC.

shown.

The graphs in Fig. 6 are errors that show the effects of orientation (θ in Eq. 6 and Fig. 4) and surface anisotropy (a/b , with a and b as in Eq. 5). For each rendered pair, left and right image points (\mathbf{u}_l and \mathbf{u}_r) corresponding to scale space extrema were found, and two measures of geometric error for the match $\mathbf{u}_l \leftrightarrow \mathbf{u}_r$ were computed. Symmetric epipolar error $(d(\mathbf{u}_r, \mathbf{F}\mathbf{u}_l)^2 + d(\mathbf{u}_l, \mathbf{F}^\top \mathbf{u}_r)^2)^{1/2}$, where \mathbf{F} is the known fundamental matrix and $d(\cdot, \cdot)$ is the image-plane distance between a point and a line, measures deviation from the epipolar constraint; and symmetric reprojection error $(d(\mathbf{u}_l, \pi_l(\pi_r^{-1}(\mathbf{u}_r)))^2 + d(\mathbf{u}_r, \pi_r(\pi_l^{-1}(\mathbf{u}_l)))^2)^{1/2}$, where $\pi^{-1}(\cdot)$ is the back-projection of an image point onto the known quadratic patch, measures ‘total error’. These two errors are scale-normalized by dividing by the scale of the extremum, making them independent of both the image resolution and the absolute magnitudes of a and b .

The left graph in Fig. 6 shows the error for a quadratic patch with $a/b = 3$ at a range of orientations. At $\theta = 0^\circ$ and 90° , one of the directions of principal surface curvature is contained in the epipolar plane, and even though the match is inaccurate (as evidenced by the reprojection error), the error is confined to the epipolar plane. As a result, correspondences $\mathbf{u}_l \leftrightarrow \mathbf{u}_r$ from these patches can be used to estimate the fundamental matrix without negatively affecting the result. Note that the maximum epipolar error occurs when the directions of principal surface curvature are oriented at 45° angles to the epipolar plane.

The right graph in Fig. 6 shows epipolar error for a patch with ‘worst case’ orientation ($\theta = 45^\circ$) and increasing surface anisotropy. When $a = b$, the error is confined to the epipolar plane, and as above, a correspondence from the patch satisfies the epipolar constraint almost exactly. The epipolar error increases linearly with the surface anisotropy.

To provide a sense of how these scale-normalized errors translate to pixel errors, consider the example in the left of Fig. 7. The average scale of the inlying Type II features for this pair was found to be 3.54. A scale-normalized error of 0.05 translates to 0.18 pixels at this scale.

4.2. Real Data

For evaluation with real data, reciprocal pairs were acquired using a 12-bit Canon D10 camera (with linear radiometric response function) and multiple Lowel Pro 250W halogen light sources. The camera and sources were placed at a distance of 2m from the object and were approximately swapped to simulate reciprocal imaging conditions. Different light sources (of the same make and model) were used for the left and right images of each reciprocal pair, and the source strengths were seen to differ by as much as 25%.

Figure 5 illustrates the stages of the matching process. As described in Sect. 3.2.1, the first and second panels show the regions initially detected as extrema in scale space and those surviving the SIFT-based symmetric nearest-neighbor matching process. In the final stage of geometric calibration, symmetric SIFT matches are used as input to a 7-point RANSAC procedure (10 000 iterations; inlier threshold = 0.01 pixels) for estimation of the fundamental matrix [12]. The results are shown in the third panel of Fig. 5. To obtain radiometric calibration information, each of these final matches is used to compute an estimate of the relative source strengths (Sect. 3.2.2), and these estimates are averaged to obtain global relative source strength values.

Reciprocal Pair	Detected L/R	Symmetric Matches	Inliers	RMS Epi. Error	Source L:R
Fig. 7 (left)	573/361	276	139	0.093	1.21:1
Fig. 7 (right)	513/322	231	132	0.083	1.30:1

Table 1. Columns 2–5: Number of detected interest regions and error with respect to estimated epipolar geometry. Column 6: Estimated relative strengths of the two light sources.

Figure 7 shows the recovered epipolar lines and inlying matches for two reciprocal image pairs, and summaries of the detection numbers and source strengths are shown in Table 1. For both reciprocal pairs, the accuracy is quite high, with the RMS symmetric epipolar distance being less than one-tenth of a pixel.

5. Application to Dense Reconstruction

While a typical object or scene includes both Type I and Type II features, in this section we consider the problem of dense reconstruction in the extreme case where only Type

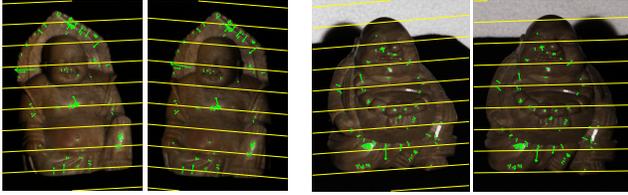


Figure 7. Epipolar geometry recovered using specular highlights as stable interest points. Superimposed on each of the two reciprocal pairs are representative epipolar lines (yellow) and corresponding specular feature points deemed as inliers (green). In each case, the RMS epipolar error is below one-tenth of a pixel.

II features are available. For a multi-camera reciprocal system, calibration using only Type II features requires special consideration. While specular highlights remain fixed to the surface in a single reciprocal image pair and can be used to estimate a fundamental matrix, they generally shift across the surface from one reciprocal pair to the next. As a result, for the purposes of geometric calibration, we are limited by the fact that there are generally no detected surface points observed by more than two views.

To solve this problem, we make use of the fact that there is a four-parameter family of camera matrix pairs $\{\mathbf{P}_1, \mathbf{P}_2\}$ corresponding to a given fundamental matrix \mathbf{F}_{12} , and that we can fix a projective coordinate system (and obtain a projective reconstruction) simply by choosing a member of this family. Then, to recover a third camera \mathbf{P}_3 in the same projective coordinate frame, we can solve the system of linear equations resulting from the constraint that the matrices $\mathbf{P}_3^\top \mathbf{F}_{13} \mathbf{P}_1$ and $\mathbf{P}_3^\top \mathbf{F}_{23} \mathbf{P}_2$ must be skew-symmetric. The only restriction in using this procedure is that the three camera centers cannot be collinear. In this way, a three-camera projective reconstruction can be obtained even though none of the world points are explicitly identified.

In a multi-camera reciprocal acquisition system, one can obtain fundamental matrices for all possible camera/source pairs using only Type II features as described in the previous section. By selecting an initial pair of cameras, the procedure outlined above can be used to successively recover all of the cameras in the system. The collinearity constraint is generally not a severe restriction, since the camera/source positions can be chosen appropriately.

An example of an object reconstructed using this approach is shown in Fig. 8. This surface was obtained from ten reciprocal pairs (one of which are shown in the left of Fig. 7) obtained from five camera/source positions using the same equipment as described in Sect. 4.

Type II features were detected and matched between pairs, and they were used to estimate the fundamental matrices and the pair-wise relative source strengths as described in Sect. 3.2. The ‘global’ relative strengths $s_1 \dots s_5$ of the five light sources were recovered from the pair-wise relations by solving a homogeneous linear system (the ratio of source strengths from each reciprocal pair give one linear

constraint on these five values), and these were found to be 0.74, 0.89, 0.89, 0.92 and 1.

The five cameras $\mathbf{P}_1 \dots \mathbf{P}_5$ were recovered from the collection of fundamental matrices using the procedure outlined above. Unlike Sect. 4, for the result in Fig. 8, the calibration problem was simplified by assuming affine cameras (and affine fundamental matrices), and the affine ambiguity in the reconstruction was resolved using so-called ‘metric constraints’ on the cameras (*i.e.*, that they have unit aspect ratio and zero pixel skew). Finally, a dense reconstruction was obtained using the method described by Zickler *et al.* [16] adapted for variable source strengths.

The importance of a radiometric calibration is demonstrated in Fig. 8(d), which compares the surface obtained using the estimated relative source strengths to that obtained assuming all sources are equal. To generate the latter, the surface from Fig. 8(a-c) was projected into the images, the surface normals were re-estimated using Eq. 2 with $s_l = s_r = 1$ for all pairs, and these new surface normals were integrated to recover the new surface. When the sources are erroneously assumed to be constant, each surface normal is biased, which results in the skewed surface on the bottom of Fig. 8(d). In contrast, by making use of the radiometric information from matched Type II features, this ‘skewing’ is avoided.

6. Conclusion

Despite its ability to provide accurate reconstructions for a broad class of surfaces, the utility of Helmholtz stereopsis has been limited by its supposed need for off-line geometric and radiometric calibration. This paper demonstrates that stable interest regions exist in Helmholtz image pairs, and that these interest regions can be used to obtain much of the necessary calibration information directly from the data without an off-line calibration.

We identify two distinct classes of features. The first is analogous to interest regions used in conventional stereo, while the second class is based on the use of specular highlights and is unique to Helmholtz stereopsis. In addition to providing geometric information, the symmetry of a reciprocal imaging system means that specular features provide essential radiometric information in the form of constraints on the relative strengths of the light sources.

The methods for identifying and matching interest regions described in this paper provide a necessary stepping stone from a Helmholtz stereo dataset to a sparse, projective reconstruction. This opens the door to questions regarding how this reconstruction—along with the associated radiometric constraints—is best upgraded to a dense, metric reconstruction. This paper provides an example in which an affine to metric upgrade is accomplished using established tools from multi-view geometry; and reciprocity-specific auto-calibration for the far-field (affine) case has received

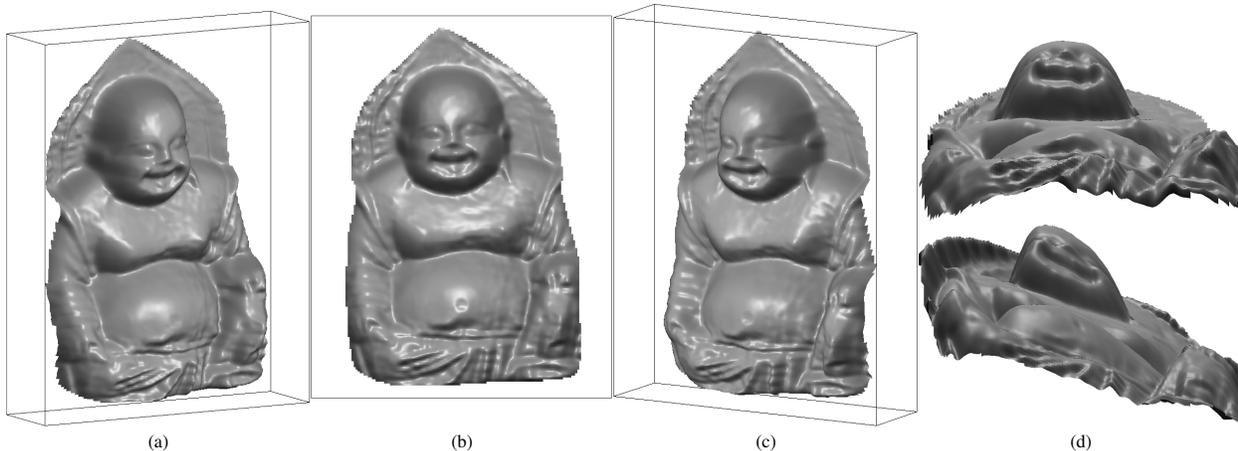


Figure 8. Dense reconstruction obtained by a self-calibrating reciprocal imaging system. Five camera/source positions capture ten reciprocal image pairs—one of which is in the left of Fig. 7—and the system self-calibrates both geometrically and radiometrically (in terms of relative source strengths) using specular highlights as features. (a–c) Surface recovered using complete calibration information. (d) Profile comparison with the surface obtained by a naive approach (bottom) that ignores radiometric calibration and erroneously assumes all sources to be of equal strength.

some attention in past work [17]. However, reciprocity-specific auto-calibration for dense reconstruction in the projective case remains largely unexplored.

There is a growing demand for three-dimensional visual content, and in order to fill this need, practical and accurate reconstruction systems must be developed. In this context, tools for self-calibration, such as those presented here, will likely play an important role.

Acknowledgment

We thank D. Lowe for making his SIFT code available and P.H.S. Torr for his implementation of RANSAC for fundamental matrices. This work was funded in part by a National Science Foundation CAREER award.

References

- [1] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Tech. Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon Univ., 2003.
- [2] J.-Y. Guillemaut, O. Drbohlav, R. Šára, and J. Illingworth. Helmholtz stereopsis on rough and strongly textured surfaces. In *Proc. Int. Symp. on 3D Data Processing, Visualization and Transmission*, pages 10–17, 2004.
- [3] Z. Janko, O. Drbohlav, and R. Šára. Radiometric calibration of a Helmholtz stereo rig. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 166–171, 2004.
- [4] T. Lindeberg. Feature detection with automatic scale selection. *Int. Journal of Computer Vision*, 30(2):79–116, 1998.
- [5] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, 60(2):91–110, 2004.
- [6] S. Magda, D. Kriegman, T. Zickler, and P. Belhumeur. Beyond Lambert: Reconstructing surfaces with arbitrary BRDFs. In *Proc. Int. Conf. on Computer Vision*, pages II: 391–398, 2001.
- [7] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. British Machine Vision Conf.*, pages 384–393, 2002.
- [8] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, 2003.
- [9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 2, pages 257–263, 2003.
- [10] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Int. Journal of Computer Vision*, 60(1):63–86, 2004.
- [11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. Journal of Computer Vision*, 65(1/2):43–72, 2005.
- [12] P. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int. Journal of Computer Vision*, 24(3):271–300, 1997.
- [13] P. Tu and P. Mendonça. Surface reconstruction via Helmholtz reciprocity with a single image pair. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 541–547, June 2003.
- [14] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *Int. Journal of Computer Vision*, 59(1):61–85, 2004.
- [15] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
- [16] 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.
- [17] T. Zickler, P. Belhumeur, and D. Kriegman. Toward a stratification of Helmholtz stereopsis. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 548–555, June 2003.
- [18] T. Zickler, J. Ho, D. Kriegman, J. Ponce, and P. Belhumeur. Binocular Helmholtz stereopsis. In *Proc. Int. Conf. on Computer Vision*, pages 1411–1417, 2003.