Dense 3D Reconstruction from Specularity Consistency

Diego Nehab
Microsoft Research

Tim Weyrich
Princeton University

Szymon Rusinkiewicz
Princeton University

Abstract

In this work, we consider the dense reconstruction of specular objects. We propose the use of a specularity constraint, based on surface normal/depth consistency, to define a matching cost function that can drive standard stereo reconstruction methods. We discuss the types of ambiguity that can arise, and suggest an aggregation method based on anisotropic diffusion that is particularly suitable for this matching cost function.

We also present a controlled illumination setup that includes a pair of cameras and one LCD monitor, which is used as a calibrated, variable-position light source. We use this setup to evaluate the proposed method on real data, and demonstrate its capacity to recover high-quality depth and orientation from specular objects.

1. Introduction

A popular class of systems for determining the 3D shape of real-world objects relies on the notion of multi-view consistency. These systems may be thought of as hypothesizing 3D points, then evaluating whether their projections into two or more viewpoints are consistent with images taken from those views. In the simplest case, the consistency is evaluated based on color. More sophisticated systems might evaluate the consistency of windows of pixels (as is the case in many stereo systems), or might consider temporal variation (as in temporal active stereo or structured light systems). Any of these consistency criteria may be used to construct a matching cost function, which is then used during triangulation to find geometry (possibly augmented by an additional stage to enforce global consistency or smoothness). With few exceptions, however, such systems have focused on reconstructing diffuse objects and scenes.

In this paper, we consider the problem of reconstructing specular objects, positioned in a scene or environment of known geometry and appearance. Because the images of these objects consist entirely of reflections of the scene around them, simple consistency metrics based on color, pixel windows, or temporal variation will not be effective.

Instead, we rely on a criterion that considers the *reflection* of the scene on a surface element. Specifically, we consider the mirror reflection of a camera ray off a surface patch with some hypothesized position and normal; the proposed position/normal tuple exhibits *specularity consistency* if the observed pixel is consistent with the intersection of the reflected ray and the (known) scene.

Given only a single camera position, we find that an infinity of position/normal tuples will necessarily satisfy the consistency condition: for any proposed position, we can find a normal such that the reflected ray hits any desired scene point. Two camera positions, however, provide disambiguating information, and in most cases restrict us to a single allowed position/normal (exceptions are discussed later in the paper). Note that, in general, a point is reconstructed not because different cameras observe the same part of the scene reflected from it: they observe different reflections that are *both* consistent with the hypothesized position/normal.

This specularity consistency condition has been previously identified and exploited in other contexts [23, 5]. We propose to use the constraint to define a matching cost function for two camera views, and demonstrate that the condition may be used for dense stereo reconstruction of specular 3D objects. We analyze our system on synthetic imagery, and show real-world results. Our contributions include:

- defining a stereo matching cost function based on the specularity consistency constraint;
- proposing a novel normal-based anisotropic diffusion scheme for the matching cost, which strengthens matches lying on a continuous surface;
- presenting a theoretical analysis of the ambiguities that can arise from the specularity consistency constraint;
- obtaining dense and accurate 3D reconstructions of specular objects, by directly exploiting this constraint.

2. Related work

Specular surfaces have been widely studied in the literature. By imposing continuity and smoothness constraints, depth can be indirectly obtained from normal estimates [12, 29]. Normal estimates for glossy surfaces can be computed for example by fitting of explicit or parametric models [13, 9, 18], or by direct detection of highlights [8].

On the other hand, most work on stereo reconstruction in the presence of specularities attempts to avoid view dependent effects. Approaches that cope with specular highlights

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

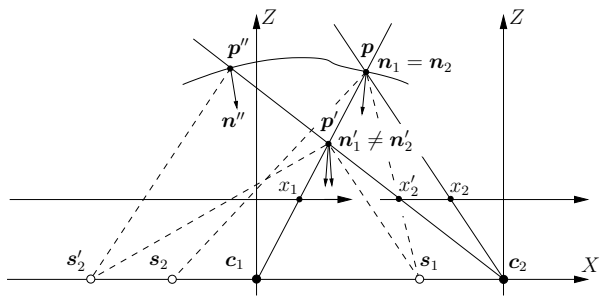


Figure 1. The specularity consistency. An incorrect match x'_2 is likely to produce normal hypotheses $\boldsymbol{n}'_1$ and $\boldsymbol{n}'_2$ that do not agree, allowing us to reject the match. In contrast, the correct match will always result in consistent normal estimates by the two cameras.

include capturing multiple views [2, 16], eliminating highlights with polarization filters [31, 19], treating them as occlusions [15], or removing them from the captured images prior to matching [28, 16, 34, 32].

Naturally, for the class of objects we are considering (i.e., smooth, glossy materials), view-dependent information provides the most useful of constraints. A wide range of methods derive local shape information from the identification and/or tracking of the distorted reflections of light-sources and special known features [3, 4, 36, 22, 24]. These methods tend to produce sparse reconstructions.

Dense measurements can be produced, for example, by the general framework of light-path triangulation [14]. Within this framework, our method can be described as a novel technique for $\langle 2, 1, 1\rangle$-triangulation. Conversely, Bonfort et al. [6] present a method for $\langle 1, 1, 2\rangle$-triangulation that requires just one viewpoint, but two known reference points per ray. Outside of this framework, shape has been recovered by rotating an object under known or unknown illumination [33, 1], and by exploring the Helmholtz reciprocity [35]. Recently proposed matching cost functions that are especially suited to specular surfaces include light transport constancy [10] and scatter-trace photography [17].

The idea of checking for normal consistency across different views has been originally proposed by Sanderson et al. [23], in the context of feature matching of specular highlights, and later used by Bonfort and Sturm [5], as part of a voxel carving method for recovering the shape of specular objects. In contrast, by rephrasing this constraint as a matching cost function, we can leverage decades of research on stereo reconstruction [25]. In particular, we can take advantage of efficient global matching algorithms that produce satisfactory dense reconstructions even when the matching cost functions are not particularly discriminant.

3. Triangulation by specularity consistency

The reconstruction method we propose relies on the ability to densely identify environmental scene points reflected by a specular surface. This can be achieved by illuminating the specular object with a dense set of controllable light sources. Various setups have been proposed to create a temporally encoded lighting environment using monitor pixels as light sources [37, 30, 6]. Using a suitable encoding scheme, the source position $\boldsymbol{s}$ of light reflected by a specular surface point $\boldsymbol{p}$ toward a camera can be decoded from the camera's intensity observations over time. In section 6, we present a similar prototype system to evaluate our approach.

With the light source position $\boldsymbol{s}$ known, the normal $\boldsymbol{n}$ at a point $\boldsymbol{p} \in \mathbb{R}^3$ can be easily computed as the bisector $\boldsymbol{n} = (\boldsymbol{l} + \boldsymbol{v})/\|\boldsymbol{l} + \boldsymbol{v}\|$ between the lighting and viewing directions: $\boldsymbol{l} = (\boldsymbol{s} - \boldsymbol{p})/\|\boldsymbol{s} - \boldsymbol{p}\|$ and $\boldsymbol{v} = (\boldsymbol{c} - \boldsymbol{p})/\|\boldsymbol{c} - \boldsymbol{p}\|$, with $\boldsymbol{c}$ the center of projection of the observing camera. In the context of stereo reconstruction, however, depths and hence surface point positions are not known in advance. Instead, a *candidate point* $\boldsymbol{p}$ corresponding to a disparity hypothesis has to be tested for consistency.

Consider figure 1. A point $\boldsymbol{p}$ on the specular surface projects to pixel x_1 as seen by camera 1. Assume we also know the light source position $\boldsymbol{s}_1$ that casts a highlight through x_1. We want to find the corresponding pixel x_2 to which $\boldsymbol{p}$ projects at camera 2, so we can triangulate for its position. Naturally, we must be able to distinguish the wrong candidate matches x'_2 from the correct match x_2.

The figure shows an example of incorrect match x'_2. By triangulation, this match determines an incorrect position $\boldsymbol{p}'$ along the line of sight through x_1. Given $\boldsymbol{p}'$ and $\boldsymbol{s}_1$, camera 1 hypothesizes a normal direction $\boldsymbol{n}'_1$. Camera 2, on the other hand, makes an independent normal direction hypothesis $\boldsymbol{n}'_2$. This hypothesis comes from $\boldsymbol{p}'$ and the light source position $\boldsymbol{s}'_2$ that casts a highlight through x'_2. Recall that, in the figure, $\boldsymbol{p}$ does not project to x'_2. Some other surface point $\boldsymbol{p}''$ does, with an independent normal direction $\boldsymbol{n}''$. This is the point responsible for the light source position $\boldsymbol{s}'_2$ that is visible through x'_2. The independence between points $\boldsymbol{p}$ and $\boldsymbol{p}''$ is likely to cause the normal direction hypotheses to disagree, in other words $\boldsymbol{n}'_1 \neq \boldsymbol{n}'_2$.

In contrast, the correct correspondence x_2 actually images the point $\boldsymbol{p}$. Triangulation produces the correct depth, and the light source position $\boldsymbol{s}_2$ visible through x_2 really is reflected from $\boldsymbol{p}$. Since both x_1 and x_2 observe the same surface point, the normal direction hypotheses $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ will agree. Our framework for dense stereo reconstruction exploits this fact.

4. Dense stereo framework

Dense stereo reconstruction generally requires a metric that assesses a disparity hypothesis d for a given image point (x, y). This is typically expressed by assigning a matching cost value $C(x, y, d)$ to this hypothesis. A stereo matching algorithm then assigns minimum-cost disparities to all pixels in a reference image, say, the image captured by

camera 1. Algorithms differ in whether disparity minimization takes place locally or globally, and in how an overall minimum is defined and computed [25]. However, matching costs are universal, and arbitrary cost functions can be directly used with most existing algorithms.

4.1. Matching Costs

Stereo matching cost functions are usually based on image intensities and consider squared or absolute differences between candidate matching pixels in the image pairs. That is, the cost of matching pixel (x, y) in image 1 with pixel $(x + d, y)$ in image 2 can be expressed, e.g., as

$$C(x, y, d) = ||I_1(x, y) - I_2(x + d, y)|| \ , \tag{1}$$

with $I_i(x, y)$ denoting the intensity at a pixel (x, y) in the image captured by camera i.

In contrast to this, our reconstruction framework is based on differences in normal hypotheses rather than intensities. A given disparity hypothesis d for a pixel (x, y) corresponds to a candidate point $\boldsymbol{p}$ in space. As explained in section 3, $\boldsymbol{p}$ produces two normal estimates $\boldsymbol{n}_i(\boldsymbol{p})$ from the two cameras. We use the angular difference

$$\delta(x, y, d) = \cos^{-1}\big(\boldsymbol{n}_1(\boldsymbol{p})^\top \boldsymbol{n}_2(\boldsymbol{p})\big) \tag{2}$$

between the normal estimates as a correspondence measure for a hypothesis (x, y, d). If $\boldsymbol{p}$ exactly coincides with a specular surface, δ is expected to be zero. In a realistic setting, however, there will be a slight error in the normal estimate due to noise and to spatial and intensity quantization, that is, the measured δ will deviate from the ideal δ_0 by $\delta = \delta_0 + \delta^*$. Assuming the error term δ^* to be mean-free and normal-distributed with standard deviation σ, the likelihood of $\delta(x, y, d)$ denoting a match is

$$P(x, y, d) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\frac{\delta(x,y,d)^2}{2\sigma^2}} \ . \tag{3}$$

In our experiments, we choose σ to be between $5°$ and $8°$, depending on the data quality; using much different values leads to a poor discrimination of matches. We therefore set the matching cost function $C(x, y, d) = 1 - P(x, y, d)$. This cost function can now be used with a variety of existing stereo reconstruction algorithms. Section 7 shows example reconstructions using different alternatives.

4.2. Normal-aware cost aggregation

In traditional (intensity-based) stereo, the cost computation is often extended to a window region around the point of interest. This increases the reliability of the cost assessment for textured regions. Cost aggregation within a window can be expressed as a 2D or 3D convolution

$$C'(x, y, d) = (w * C)(x, y, d) \tag{4}$$

with a window function w [25]. This convolution can alternatively be expressed as a possibly anisotropic diffusion process [26].

When w is anisotropic, it introduces a bias for a certain slope during the surface reconstruction stage. In particular, if w is a 2-dimensional kernel over x and y (a common choice), the convolution corresponds to the integration over a window in the image domain, introducing a bias for image-parallel surface slopes.

We exploit the additional knowledge of normal directions to specifically bias the surface reconstruction toward the surface orientation corresponding to the normals. The intersection of the object surface with the epipolar (xd-)plane of y corresponds to a minimum-cost ridge in the cost function $C(x, y, d)$. Any surface reconstruction algorithm has to trace this ridge. In order to improve the precondition of the surface reconstruction, we perform anisotropic diffusion of the symmetrized cost function $F_y(x_1, x_2) \equiv C(x_1, y, x_2 - x_1)$, aggregating costs along the ridge direction predicted by the normal estimates. As shown in [20], given a pair of rectified cameras, the unit tangent direction $\boldsymbol{t}$ to the matching ridge at (x_1, x_2) can be expressed as

$$\boldsymbol{t} \propto (\boldsymbol{n}^\top \boldsymbol{p}_1, \boldsymbol{n}^\top \boldsymbol{p}_2)^\top \tag{5}$$

where $\boldsymbol{n}$ and $\boldsymbol{p}_i$ are, respectively, the 2D projections of the normal and position of the surface point at (x_1, x_2), on the epipolar plane of y relative to camera i.

Following equation (4), we convolve F_y with a spatially-varying, oriented Gaussian filter kernel g:

$$F'_y(x_1, x_2) = (g * F_y)(x_1, x_2) \ , \tag{6}$$

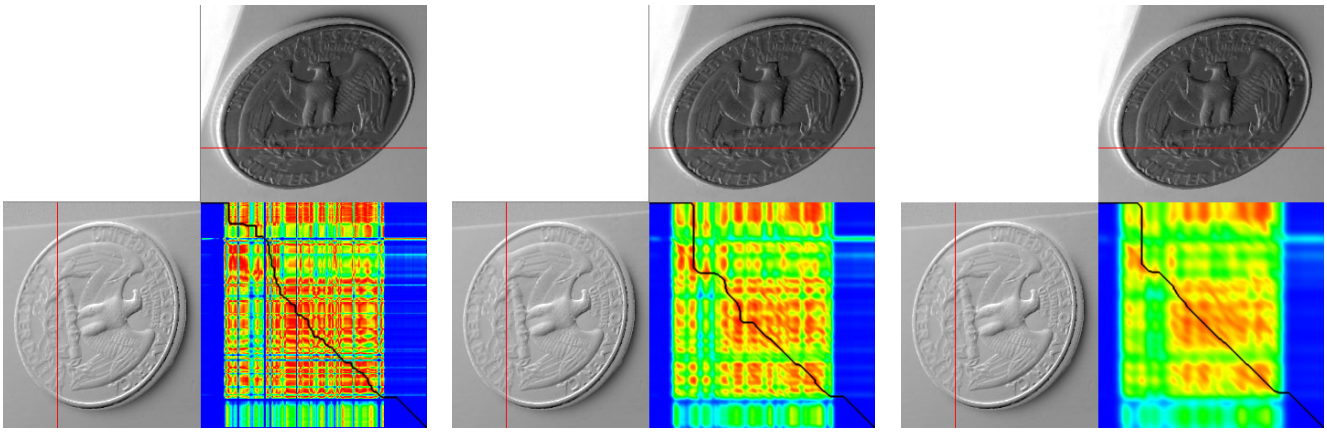
where

$$g(u, v) = \mathrm{e}^{-\boldsymbol{x}^\top \boldsymbol{V}^{-1} \boldsymbol{x}}, \quad \boldsymbol{x} = (u, v)^\top \ , \tag{7}$$

is controlled by a spatially-varying variance matrix $\boldsymbol{V} \in \mathbb{R}^{2\times 2}$, dependent on the location of the filter application (x_1, x_2):

$$\boldsymbol{V} = r_{\mathrm{aa}}\mathbf{I} \ + \ r_{\mathrm{d}} \sum_{i\in\{1,2\}} \boldsymbol{t}_i \boldsymbol{t}_i^\top + F_y(x_1, x_2)\bar{\boldsymbol{t}}_i \bar{\boldsymbol{t}}_i^\top \ . \tag{8}$$

Here $\boldsymbol{t}_i$ is the unit vector in the direction of the slope, according to the normal estimate $\boldsymbol{n}_i$, and $\bar{\boldsymbol{t}}_i$ is unit length and orthogonal to $\boldsymbol{t}_i$. Each summand after the summation sign implements the variance matrix of an oriented Gaussian along $\boldsymbol{t}_i$, each converging to an isotropic Gaussian as the local matching cost reaches one. Adding the variance matrices corresponds to convolving the respective Gaussians. The additional summand $r_{\mathrm{aa}}\mathbf{I}$ provides an anti-aliasing prefilter to guarantee a faithful discretization of g. In our implementation we use an anti-aliasing radius of $r_{\mathrm{aa}} = 0.4$ and a diffusion speed of $r_{\mathrm{d}} = 1.41$; deviation from these values typically leads to a less stable cost aggregation, or to a



Source F_y | F'_y after 4 iterations | F'_y after 16 iterations

Figure 2. Consistency values within an epipolar plane (red line) through a problematic region with many false matches. High confidence values are displayed in red; the black line shows the result of a disparity optimization using dynamic programming. Anisotropic diffusion according to the normal estimates attenuates false positives. The images show (from left to right) unmodified consistency values $F_y(x_1, x_2)$ and $F'_y(x_1, x_2)$ after 4 and 16 diffusion steps, respectively.

slow convergence, respectively. Finally, in order to ensure energy preservation, each filter kernel is normalized after discretization to add up to one.

By iteratively applying equation (6), minimum-cost ridges are emphasized where both normal estimates agree with the matching ridge, while spurious ridges are attenuated. Figure 2 shows an example of a problematic epipolar plane through the quarter dataset: the complex leaf pattern on the coin, and the non-specular background object introduce many spurious matches that are gradually smoothed out by the diffusion process.

5. Ambiguities

Even disregarding interreflections, the specularity constraint can lead to ambiguities. An observation by camera i can be described by a pair (x_i, s_i), where x_i is the pixel coordinate and s_i the light source position that produces a highlight through x_i. Here, for simplicity, and without loss of generality, we assume that the light source lies along the baseline between the two cameras (as in figure 3). Ideally, for any surface, and for each observation (x_1, s_1) by camera 1, there would exist only *one* corresponding consistent observation (x_2, s_2) by camera 2. Unfortunately, this is not the case. We can distinguish between two types of ambiguity. A *week ambiguity* occurs when at least two different observations, (x_2, s_2) and (x'_2, s'_2), are consistent with (x_1, s_1). A *strong ambiguity* occurs when *every* observation (x_2, s_2) is consistent with (x_1, s_1).

5.1. Weak ambiguities

Consider a pixel x'_2 from camera 2 that does not correspond to x_1 (figure 3). By triangulation, we can compute the position $\boldsymbol{p}'$ of the non-existent corresponding surface point:

$$\boldsymbol{p}'(x_1, x'_2) = (X', Z')^\top = (\frac{x_1 T}{x_1 - x'_2}, \frac{T}{x_1 - x'_2})^\top, \qquad (9)$$

where T is the baseline distance between the two cameras.

From $\boldsymbol{p}'$ and s_1, we can compute the normal direction expected by camera 1 at $\boldsymbol{p}'$. This normal direction is uniquely determined by the intersection point n'_1 of the normal ray at point $\boldsymbol{p}'$ with the baseline between the two cameras. To find the intersection, we use the angle bisector theorem on the triangle formed by points $\boldsymbol{p}'$, $\boldsymbol{c}_1$, and s_1 to get

$$n'_1(x_1, s_1, x'_2) = \frac{s_1\sqrt{X'^2 + Z'^2}}{\sqrt{X'^2 + Z'^2} + \sqrt{(X' - s_1)^2 + Z'^2}}. \qquad (10)$$

If (x'_2, s'_2) is consistent with (x_1, s_1), then the light from s'_2 must reflect at $\boldsymbol{p}'$ towards the ray through x'_2. Using $\boldsymbol{p}'$ and n'_1, and the angle bisector theorem on the triangle formed by points $\boldsymbol{p}'$, $\boldsymbol{c}_2$, and s'_2, we get $s'_2(x_1, s_1, x'_2) =$

$$-\frac{{n'_1}^2(T - 2X') + (T - 2n'_1)(X'^2 + Z'^2)}{{n'_1}^2 + 2T(X' - n'_1) - (X'^2 + Z'^2)}. \qquad (11)$$

Even though the ray through x'_2 does not hit a surface point at $\boldsymbol{p}'$ (this point would occlude $\boldsymbol{p}$, which is by assumption visible through x_1), the ray may intersect the surface at *another depth* Z''. For each possible point $p'' = (X'', Z'')^\top = (x'_2 Z'' + T, Z'')^\top$ along the ray through x'_2, we can calculate a normal direction that reflects s'_2 towards x'_2. Once again, we use the angle bisector theorem on the triangle formed by points $\boldsymbol{p}''$, $\boldsymbol{c}_2$, and s'_2 to obtain $n''_2(x_1, s_1, x'_2, Z'') =$

$$\frac{T\sqrt{(s_2 - X'')^2 + Z''^2} + s_2\sqrt{(X'' - T)^2 + Z''^2}}{\sqrt{(s_2 - X'')^2 + Z''^2} + \sqrt{(X'' - T)^2 + Z''^2}}. \qquad (12)$$

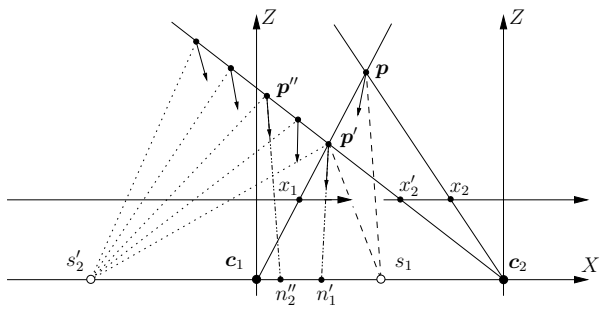


Figure 3. Weak ambiguities. Given a surface point $\boldsymbol{p}$, observed as (x_1, s_1) by camera 1, and a false candidate match x_2', the surface cannot go through the intersection point $\boldsymbol{p}'$. However, for every point $\boldsymbol{p}''$ along the ray through x_2', there is a surface orientation that results in an observation (x_2', s_2'), consistent with (x_1, s_1).

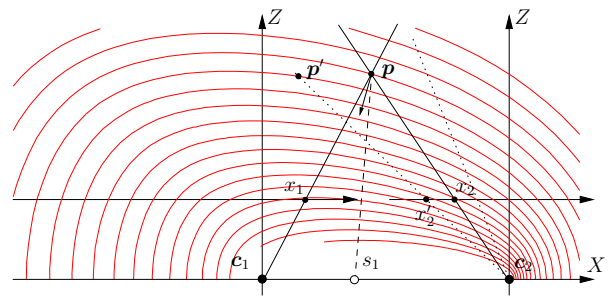


Figure 4. Strong ambiguity curves. An observation (x_1, s_1) by camera 1 imposes a normal field on the XZ-plane. If a surface conforms to this normal field, every observation (x_2, s_2) by camera 2 will be consistent with (x_1, s_1). In other words, we have a strong ambiguity.

In summary, the rays through x_1 and x_2' observe independent surface points, and intersect at a non-existent hypotetical point $\boldsymbol{p}'$ (equation (9)). Camera 1 expects $\boldsymbol{p}'$ to have normal n_1' (equation (10)). For each x_2', we can compute a light source position s_2' that causes (x_2', s_2') to be consistent with (x_1, s_1) (equation (11)). Then, for each point $\boldsymbol{p}''$ along the ray through x_2', we can compute a normal direction n_2'' that reflects light from s_2' towards the ray through x_2' (equation (12)). Whenever the surface goes through a point $\boldsymbol{p}''$ with the appropriate normal direction n_2'', there is weak ambiguity.

5.2. Strong ambiguities

Consider figure 4. A point $\boldsymbol{p}$ generates an observation (x_1, s_1) by camera 1. Take now an arbitrary point $\boldsymbol{p}' = (X', Y')^\top$ being observed by camera 2 through pixel x_2'. The ray through x_2' intersects the ray through x_1 at a point that we can obtain from equation (9). Following equations (11) and (12), we can obtain a normal direction at $\boldsymbol{p}'$ that makes the observation at x_2' consistent with the observation at x_1.

In other words, each observation by camera 1 imposes a normal field on the XZ-plane. Starting from any point in the XZ-plane (i.e., an initial condition), any trajectory that respects this normal field (i.e., solves the associated differential equation) will generate a strong ambiguity curve. The curves in figure 4 were generated by numerical integration. If the surface follows one of these curves, *every* observation by camera 2 will be consistent with (x_1, s_1) by camera 1. These are the strong ambiguity curves.

In theory, weak and strong ambiguities are extremely unlikely. In practice, finite precision on the light source position estimation, as well as noise and discretization artifacts allow the phenomenon to manifest itself as wrong matches. These can be particularly problematic if the surface resembles a strong ambiguity curve.

6. Acquisition

Specular triangulation requires a camera pair and an apparatus to illuminate an object with a dense set of calibrated light sources. To that end, the cameras acquire images of the object being lit by a temporally encoded light pattern. In principle, we can use any light pattern that allows us to find the identity of a single light source L_i in a specular reflection observed by a camera at $\boldsymbol{c}_i$.

Efficient encoding schemes, such as gray-code patterns, exist; however, specular reflection is subject to a convolution with a specular lobe. Accordingly, the chosen pattern has to be robust under convolution. A very robust encoding would be to trigger one light source at a time. For each surface point, the light source L^i_{max} is determined that produces a maximum intensity response under observation from camera i. The temporal location of the maximum response is comparatively stable even under convolution with a specular lobe. For a dense set of many light sources, however, this procedure is highly inefficient.

In order to reduce acquisition times, we use multiple linear light sources instead of a dense set of single point light sources [11]. During acquisition, the linear light sources are swept through space. For each surface point, the maximum response during each sweep is determined. In a good approximation, these maxima occur when the linear light source covers L^i_{max}. Accordingly, we intersect the locations of the corresponding illuminating lines to obtain the position $\boldsymbol{s}_i$ of L^i_{max}.

6.1. Acquisition Procedure

We have implemented a system that performs the linear light source sweeps by displaying horizontal and vertical stripes on an LCD monitor. The stripes are 0.5 cm wide and are sequentially shifted by one stripe-width on each frame. Scanning times could be improved by multiplexing the stripes with Hadamard patterns [27], but we give preference to quality over speed in the acquisition system in order

Figure 5. Photograph of our experimental setup. The object is placed in front of the monitor, facing the camera pair. The cameras capture image pairs while the monitor displays sweeping linear light sources. The small projector visible above the cameras is used for comparison with structured-light reconstructions.

to avoid additional sources of error. Figure 5 shows our experimental setup. Both the camera pair and the monitor's location with respect to the cameras are calibrated. Knowing the light source locations $\boldsymbol{s}_i$, specular triangulation can be performed as described in section 3.

In practice, the crucial part is a reliable estimation of L^i_{max} from the intensity variation over time of each observed point. Camera noise and the discretization of the line sweep prevent us from directly picking the stripe location with the maximum intensity response. Instead, we obtained good results by fitting a Gaussian mixture model

$$I(t) = \sum_i^n a_i \mathrm{e}^{-\frac{(t-\mu_i)^2}{\sigma_i^2}} \tag{13}$$

to the data, and taking the maximum intensity point in time as μ_i of the narrowest Gaussian lobe (i.e., for i with minimum σ_i). The use of multiple Gaussians allows the fit to approximate diffuse and ambient reflectance contributions by wider lobes, with the specular peak corresponding to a single, narrow lobe.

A common choice for fits of Gaussian mixture models is the Expectation Maximization algorithm. In our case, however, it is not applicable, as the support of the measurements is truncated. Instead, we use a general non-linear fit to equation (13). In experiments with real data, we found that, for most specular materials, the peak can reliably be detected with even a single lobe. See figure 6 for sample temporal reflectance profiles of measurements of a coin.

In the proposed setup, the maximum-response peaks μ_h and μ_v of the horizontal and vertical sweep, respectively, directly correspond to a monitor location $(x, y) \propto (\mu_\text{h}, \mu_\text{v})$, which in turn allows us to determine $\boldsymbol{s}_i$.

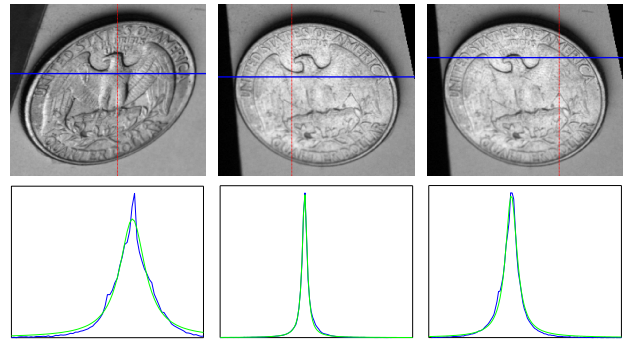


Figure 6. (Top) Temporal response measurements at different points of a coin. (Bottom) Corresponding measured intensity profiles. Measurements are shown in blue, and the fit of a single Gaussian lobe is shown in green.

6.2. Properties

The approach requires free lines of sight between the surface point and the two cameras, as well as from the point to the two light sources in the reflection direction. This imposes constraints on the object geometry. Similar constraints are also present with other photometric reconstruction techniques, such as photometric stereo. Unlike photometric stereo, which suffers from normal bias in the case of (partial) self-shadowing, specularity consistency only requires a narrow free cone of sight to the lighting environment, and is hence more robust against such bias.

A practical limitation of the proposed setup, however, is the relatively small solid angle covered by the monitor. In order for a point to be reconstructed, the reflection of (potentially different) areas of the monitor must be visible by both cameras observing the point. Only points with normals falling within a narrow range satisfy this requirement. Normals outside of this range result in gaps in the reconstruction. We are currently investigating alternative layouts that will reduce this problem by covering a larger solid angle.

7. Results

Using the setup described in the previous section, we captured two real datasets: a quarter and a mirrored sphere. In order to evaluate our method in ideal conditions, we also generated synthetic data for a Greek panel and for an analytic sphere, using the same calibration parameters as the real scanner. The simulated datasets were processed by the same pipeline as the real datasets, starting with simulated camera images instead of real images.

In order to demonstrate the versatility of the proposed cost function, we use two different stereo reconstruction algorithms. Figure 7 shows a reconstruction of the simulated sphere dataset using Markov Random Fields stereo reconstruction [7]. All other reconstructions were obtained using dynamic programming. Results are shown in figure 8.

In general, we have found that our triangulation stage

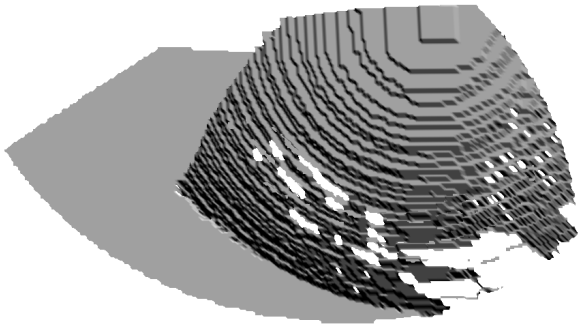


Figure 7. Rendering from discrete matches produced by Markov Random Fields stereo reconstruction [7], evaluated on a simulated dataset of a sphere using our specularity constraint.

consistently produces high-quality normals, but can generate incorrect depth estimates (compare the first and second columns of figure 8). Due to the geometry of the capturing setup, an incorrect match can result in a large error in the reconstructed depth, whereas the normal estimate computation tends to be more robust to such errors.

Nevertheless, normal and depth are not independent, and we therefore use the measured normals to improve the quality of the depth estimates. To this end, we use the optimization method described in [21]. Recall the idea is to formulate an energy minimization problem whose solution attempts to reconcile both measurements. Since the method expects to correct relatively small depth variations, we iterate the process to produce larger corrections. The third column of figure 8 shows the result of such optimization, which produces great improvements.

The mirrored sphere shows the entire range of orientations that can be reconstructed by our system. As discussed in section 6.2, reconstruction is only possible for points within the overlapping images of the monitor as seen by both cameras. For further illustration, figure 8(h) shows the reflection of the monitor in the sphere as seen from one of the cameras.

For the quarter dataset, we also attempted to obtain geometry with a structured light scanner. The projector, used exclusively for this experiment, is visible in figure 5. Figure 8(d) shows the resulting reconstruction. The artifacts demonstrate the difficulties arising from scanning specular objects with active triangulation. Attempts to obtain the shape of the mirrored sphere using structured light projection consequently failed completely.

8. Conclusions

We have demonstrated the use of a specularity consistency criterion, based on consistency of position/normal tuples, for dense stereo matching in specular scenes. The fact that normals are closely related to the geometry being reconstructed allows us to aggregate matching costs in a meaningful manner, biasing the stereo reconstruction toward the measured surface orientation. In addition, the normal information can be used to refine the output of the stereo algorithm, yielding low noise and detailed reconstructions. We have demonstrated the applicability of the proposed criterion, using a simple acquisition setup based on two cameras and a monitor.

References

[1] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *ICCV*, 2007.

[2] D. N. Bhat and S. K. Nayar. Stereo and specular reflection. *IJCV*, 26(2):91–106, 1998.

[3] A. Blake. Specular stereo. In *IJCAI*, volume 2, pages 973–976, 1985.

[4] A. Blake and G. Brelstaff. Geometry from specularities. In *ICCV*, pages 394–403, 1988.

[5] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *ICCV*, pages 591–596, 2003.

[6] T. Bonfort, P. Sturm, and P. Gargallo. General specular surface triangulation. In *ACCV*, pages 872–881, 2006.

[7] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *CVPR*, pages 648–655, 1998.

[8] T. Chen, M. Goesele, and H-P. Seidel. Mesostructure from specularity. In *CVPR*, volume 2, pages 1825–1832, 2006.

[9] E. N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *CGIP*, 18:309–328, 1982.

[10] J. Davis, R. Yang, and L. Wang. BRDF invariant stereo using light transport constancy. In *ICCV*, pages 436–443, 2005.

[11] A. Gardner, C. Tchou, T. Hawkins, and P. Debevec. Linear light source reflectometry. *SIGGRAPH 2003*, 22(3):749–758, 2003.

[12] M. A. Halstead, B. A. Barsky, S. A. Klein, and R. B. Mandell. Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals. In *SIGGRAPH 1996*, pages 335–342, 1996.

[13] K. Ikeuchi and B. K. P. Horn. Numerical shape from shading and occluding boundaries. *AI*, 17(1–3):141–184, 1981.

[14] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3D shape by light-path triangulation. In *ICCV*, pages 1448–1455, 2005.

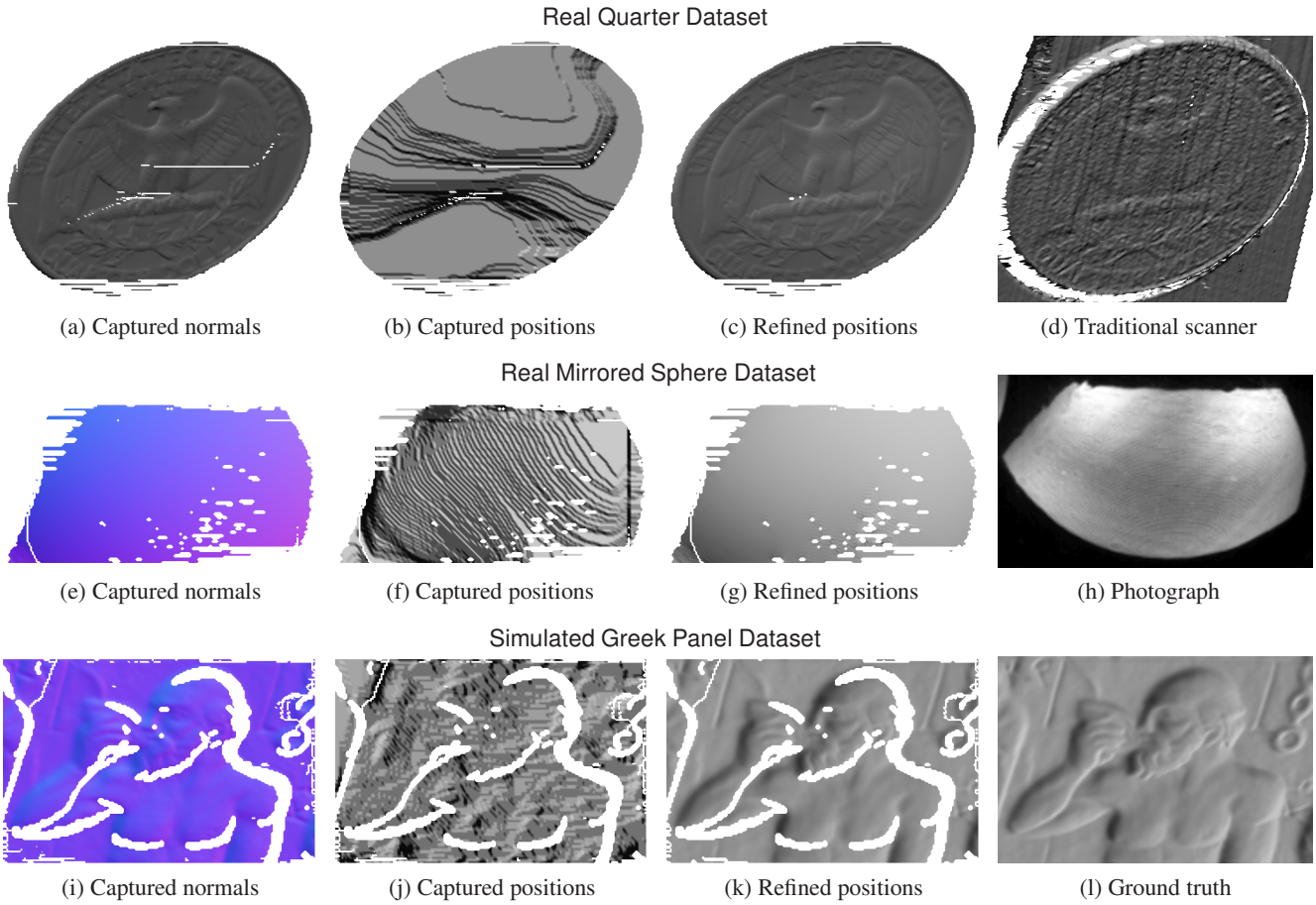
[15] Y. Li, S. Lin, H. Lu, S. B. Kang, and H-Y. Shum. Multibaseline stereo in the presence of specular reflections. In *ICPR*, volume 3, pages 573–576, 2002.

[16] S. Lin, Y. Li, S. B. Kang, X. Tong, and H-Y. Shum. Diffuse-specular separation and depth recovery from image sequences. In *ECCV*, volume 3, pages 210–224, 2002.

[17] N. J. W. Morris and K. N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *ICCV*, 2007.

[18] S. K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE TRA*, 6(4):418–431, 1990.

[19] S. K. Nayar, X-S. Fang, and T. Boult. Stereo and specular reflection. *IJCV*, 21(3):163–186, 1997.



(a) Captured normals (b) Captured positions (c) Refined positions (d) Traditional scanner

(e) Captured normals (f) Captured positions (g) Refined positions (h) Photograph

(i) Captured normals (j) Captured positions (k) Refined positions (l) Ground truth

Figure 8. Final reconstruction results.

[20] D. Nehab, S. Rusinkiewicz, and J. Davis. Improved sub-pixel stereo correspondences through symmetric refinement. In *ICCV*, pages 557–563, 2005.

[21] D. Nehab, S. Rusinkiewicz, J. E. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *SIGGRAPH 2005*, 24(3):536–543, 2005.

[22] M. Oren and S. K. Nayar. A theory of specular surface geometry. *IJCV*, 24(2):105–124, 1997.

[23] A. C. Sanderson, L. E. Weiss, and S. K. Nayar. Structured highlight inspection of specular surfaces. *PAMI*, 10(1):44–55, 1988.

[24] S. Savarese, M. Chen, and P. Perona. Local shape from mirror reflections. *IJCV*, 64(1):31–67, 2005.

[25] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1/2/3):7–42, 2002.

[26] D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. *IJCV*, 28(2):155–174, 1998.

[27] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur. A theory of multiplexed illumination. In *ICCV*, volume 2, pages 808–815, 2003.

[28] S. A. Shafer. Using color to separate reflection components. *Color research & applications*, 10(4):210–218, 1985.

[29] J. E. Solem, H. Aanaes, and A. Heyden. A variational analysis of shape from specularities using sparse data. In *3DPVT*, pages 26–33, 2004.

[30] M. Tarini, H. P. A. Lensch, M. Goesele, and H-P. Seidel. 3D acquisition of mirroring objects using striped patterns. *Graphical Models*, 67(4):233–259, 2005.

[31] L. B. Wolff and T. E. Boult. Constraining object features using a polarization reflectance model. *PAMI*, 13(7):635–657, 1991.

[32] K.-J. Yoon and I.-S. Kweon. Correspondence search in the presence of specular highlights using specular-free two-band images. In *ACCV*, pages 761–770, 2006.

[33] J. Y. Zheng and A. Murata. Acquiring a complete 3D model from specular motion under the illumination of circular-shaped light sources. *PAMI*, 22(8):913–920, 2000.

[34] W. Zhou and C. Kambhamettu. Binocular stereo dense matching in the presence of specular reflections. In *CVPR*, pages 2363–2370, 2006.

[35] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *IJCV*, 49(2/3):215–227, 2002.

[36] P. Zisserman, A. Giblin and A. Blake. The information available to a moving observer from specularities. *Image and Vision Computing*, 7(1):38–42, 1989.

[37] D. E. Zongker, D. M. Werner, B. Curless, and D. H. Salesin. Environment matting and compositing. In *SIGGRAPH 1999*, pages 205–214, 1999.