International Journal of Computer Vision 26(2), 91–106 (1998)

Stereo and Specular Reflection

DINKAR N. BHAT AND SHREE K. NAYAR
Department of Computer Science, Columbia University, New York, NY 10027

Received March 28, 1995; Revised October 30, 1996; Accepted October 30, 1996

Abstract. The problem of accurate depth estimation using stereo in the presence of specular reflection is addressed. Specular reflection, a fundamental and ubiquitous reflection mechanism, is viewpoint dependent and can cause large intensity differences at corresponding points, resulting in significant depth errors. We analyze the physics of specular reflection and the geometry of stereopsis which lead to a relationship between stereo vergence, surface roughness, and the likelihood of a correct match. Given a lower bound on surface roughness, an optimal binocular stereo configuration can be determined which maximizes precision in depth estimation despite specular reflection. However, surface roughness is difficult to estimate in unstructured environments. Therefore, trinocular configurations, independent of surface roughness are determined such that at each scene point visible to all sensors, at least one stereo pair can produce correct depth. We have developed a simple algorithm to reconstruct depth from the multiple stereo pairs.

Keywords: stereo, correspondence, specular reflection, image matching, depth estimation

1. Introduction

Stereo is a direct and passive method of obtaining three-dimensional structure of the visual world which makes it attractive for applications like autonomous navigation and surveying. The robustness of a stereo system is characterized to a large extent by its ability to obtain accurate depth estimates of scenes comprising objects with different reflectance properties.

The stereo correspondence problem (Barnard and Fischler, 1982) is inherently under-constrained. Therefore, constraints have to be imposed by making assumptions regarding scene reflectance and structure. A common assumption is that intensities at corresponding points in the images are identical. Based on this supposition, various search based strategies have been developed which correlate image regions (area-based), or image features (feature-based). However, this assumption is valid only when the surfaces in the scene are Lambertian. Corresponding point intensities are *not* identical in the presence of *specular reflection*, the specular intensity at any scene point being dependent on the viewing direction. This effect is more clearly manifest on smoother surfaces where highlights—bright regions due to specular reflection—shift on the surface even with slight changes in viewpoint. Thus, corresponding regions in stereo images can be poorly correlated, causing area-based schemes to compute incorrect depth. Similarly, when highlights are assumed to be real scene features and matched, feature-based schemes can fail. Figure 1 shows a stereo pair of a rendered cup, and depth obtained along two scanlines; one including a highlight and the other away from it. Depth was computed using a correlation-based algorithm, hence erroneous at points where corresponding intensities are vastly different.

Other methods have been developed for image matching which are not directly based on intensity values. Wolff and Angelopoulou (1994) developed a system which matches photometric ratios between stereo images. However, this scheme requires two illumination conditions which makes it unattractive for passive stereo. Furthermore, it does not extend to specular surfaces because the ratio loses its invariance to viewer

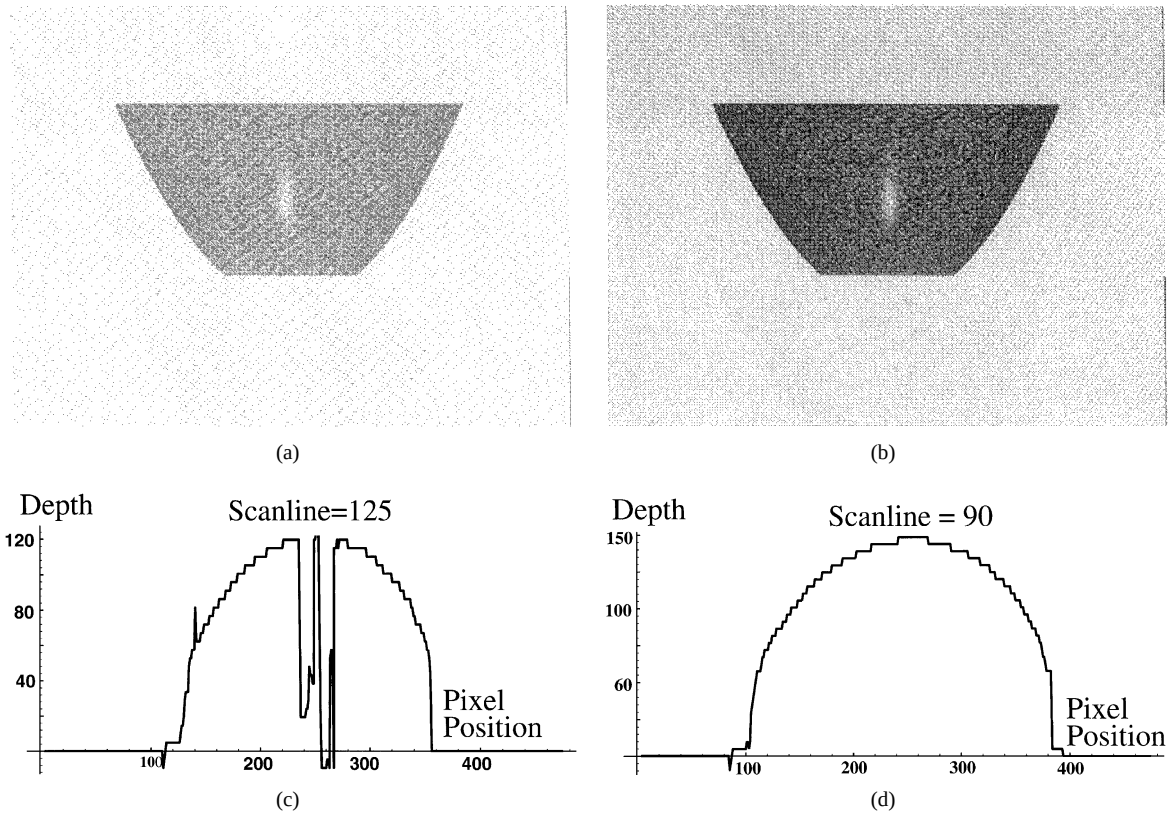


Figure 1. Rendered stereo pair and depth computed along two scanlines. (a) Left image; (b) right image; (c) depth along a scanline including the highlight; (d) depth along a scanline away from the highlight. Large depth errors can be seen in (c).

position. Smith (1986) proposed an elegant alternative to point correspondence. His approach attempts to find surface depth profiles that are consistent with image intensity profiles. Unfortunately, this too does not extend to handle specular reflection. Gennert (1988) suggested a method to relax the Lambertian assumption. It modeled intensities at corresponding points as being related by a multiplicative factor that varies smoothly. The model was developed using an assumption that relative change in intensity due to albedo variation was much greater than that due to changing surface orientation. This assumption does not seem valid for general scenes, or in the presence of specular reflection whence intensity change due to variation in surface normal can be large.

Multiple view systems have been designed to alleviate the problem of ambiguity in two-view matching. The redundant views provide strong constraints since correspondence must be established with respect to all images. In particular, trinocular stereo systems—systems using three views—have been employed to a measure of success (Ito and Ishii, 1986; Yachida et al., 1986). Here, the general strategy used for correspondence is to locate possible matches using one stereo pair and verify each potential match using the image from the third viewpoint. Only one matching triplet of points or features is expected to be consistent with respect to all three views. However, due to specular reflection, none of the matches in the first pair may correspond to the true match. Furthermore, even if the right match is located using the first stereo pair, it may not be possible to verify it in the third view as the corresponding point intensity can be greatly different. Dhond and Aggarwal (1991), in developing a feature-based system, analyzed the cost and benefit of adding an additional view for matching and concluded that the increased reliability of three view systems outweighs the added computational cost involved. However, they assumed diffuse surfaces and hence mismatches due to specular reflection were not considered in estimating reliability. A more complete analysis should deal with specular reflection in both binocular and trinocular stereo systems. Okutomi and Kanade (1993) developed a convincing, practical multiple view system. It uses

stereo pairs taken with different baselines to compute precise depth estimates overcoming ambiguities due to repetitive texture. The assumption is that the right match at any scene point is always found in each stereo pair along with other possibly incorrect matches. By combining matches from image pairs, only the correct correspondence is accentuated while the incorrect ones weaken. The above assumption is not correct, however, when specular reflection is present as the right match may not be found in one or many image pairs. Thus, all the above techniques must incorporate ways to detect and handle mismatches due to specular reflection.

To overcome the problem of depth errors due to strong highlights, Brelstaff and Blake (1988) suggested excising them from images before matching. Removal of highlights is difficult in images of real scenes and is an active area of research (Nayar et al., 1993). Ching et al. (1993) developed an empirical correlation-based technique to detect and avoid specular reflection when the camera is active. On a different note, Blake (1985) related the movement of a highlight to the Hessian of the surface which describes local surface geometry. The above techniques assume ideal specular reflection which is an extreme case as roughness tends to zero.

Current stereo algorithms are therefore seriously deficient in dealing with specular reflection. In this paper, we address the problem of precise depth estimation in the presence of specular reflection from surfaces with macroscopic roughness. First, we seek an optimal binocular stereo configuration such that intensity differences at corresponding points is limited, while depth resolution is maximized. The optimal configuration is determined independent of surface normal and source direction, and its parameters are shown to be a function of surface roughness. Therefore, for a scene where the lower bound on roughness can be estimated—quite possible in structured environments—the two cameras can be positioned so as to minimize mismatches without losing depth precision. Next, we seek to avoid estimation of surface roughness since the measurement of surface roughness is impractical in general scenes. We determine trinocular configurations whose parameters are independent of surface roughness. The important characteristic of these configurations is that for each scene point in the common field of view of the sensors, *at least* one binocular pair provides the correct depth estimate. We have developed a practical correspondence algorithm to extract correct depth estimates of scene points from different pairs so as to yield an accurate and complete depth map of the scene.

Our approach considers specular reflection from rough surfaces in the context of stereo. Previous methods have implicitly (Ching et al., 1993) or explicitly assumed ideal specular reflection. We do not attempt to avoid or detect the immediate artifacts of specular reflection like strong highlights but rather perform accurate matching in their presence. Thus, preprocessing of images, like removal of highlights, is avoided. Our approach is general as it is not limited to any specific reflectance model, or correspondence scheme.

2. Reflection Mechanisms

Surfaces exhibit two forms of reflectance—diffuse and specular. Diffuse reflection occurs due to subsurface scattering of light. It is often assumed to be Lambertian, an assumption shown to be incorrect for surfaces with macroscopic roughness (Oren and Nayar, 1994). Nonetheless, the change in diffuse component with viewing direction is generally much less pronounced than that in the specular component.

2.1. Specular Reflection

Specular reflection occurs at the boundary between surface and medium. It comprises of two components—a spike and a lobe (Nayar et al., 1991). We do not deal with surfaces smooth in comparison to wavelength of incident light as they are rare in real scenes. Hence, specular reflection refers to the lobe only. The lobe spreads in directions other than and including the specular direction, the width of its distribution depending on surface roughness. This is described by the Torrance-Sparrow model (Torrance and Sparrow, 1967) as outlined below.

A surface is viewed as a collection of planar microfacets, each behaving like a perfect mirror. A rough surface can be modeled using a probability distribution for the slopes of the microfacets. The slope distribution model uses a parameter σ which represents surface roughness. A smoother surface is characterized by a lower value for σ. Using this surface model, the specular intensity I_s at any point was shown as:

$$I_s = \frac{K_s F\, G}{\hat{n} \cdot \hat{v}} \exp\left(-\frac{1}{2\sigma^2}(\cos^{-1}(\hat{h} \cdot \hat{n}))^2\right)$$
$$\hat{h} = \frac{\hat{v} + \hat{s}}{\|\hat{v} + \hat{s}\|} \qquad (1)$$

where $\hat{v}$, $\hat{s}$ and $\hat{n}$ are unit vectors pointing along the viewing, source and normal directions, respectively; $\hat{h}$ is the bisector of $\hat{v}$ and $\hat{s}$, G is the Geometrical Attenuation Factor, and F is the Fresnel's coefficient. K_s accounts for the gain of the sensor measuring intensity, the source strength, normalization factors of the exponential, and the reflectivity of the surface. From (1), it can be deduced that: (a) when the surface is smooth, the distribution of I_s is concentrated in a small region around the specular direction, and (b) as the surface becomes rougher, the peak value of I_s decreases and it widens.

2.2. Implications for Stereo

The total image intensity I_t for any point in the scene is given by the sum of diffuse and specular intensity components. Due to variation in each component with viewing direction, the total intensities of corresponding points in the stereo images are different. But, since the change in diffuse component is much smaller than the change in specular component, it follows that the overall intensity difference I_{diff} is approximately equal to the difference in specular intensities:

$$I_{\text{diff}} = \left| I_s^1 - I_s^2 \right| \tag{2}$$

where, I_s^1 and I_s^2 are the specular intensities of the point in the two stereo images. We assume the scene is illuminated by a light source whose direction is fixed but unknown, the gain of the stereo cameras are identical while obtaining the images, and the response of each camera is linear with respect to scene radiance. I_{diff} varies over the scene as the surface normal and roughness are generally not constant. Local variance of I_{diff} (in a window, for example) could be large if the viewing directions are chosen arbitrarily, resulting in wrong matches while computing stereo correspondence using linear correlation measures like the sum of squared differences or correlation.

Figure 2 shows a binocular stereo configuration operating at a point with some surface roughness. The question then is: How far apart can the viewing vectors be located beyond which I_{diff} exceeds a threshold? This upper limit is bound to be smaller for smoother surfaces since an equivalent change in viewing direction can cause a comparatively large change in I_s (Eq. (1)). We seek to ascertain this limit independent of surface normal and source direction since these are indeterminable except in highly structured environments.

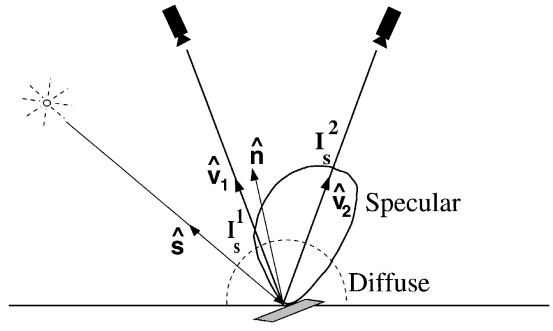


Figure 2. A binocular stereo configuration. Note that the specular intensity is different in the two sensors.

3. Vergence

We discuss how specular intensity difference at scene points can be affected by camera parameters. When points are projected orthographically, as shown in Fig. 3, corresponding rays are parallel to their respective optical axes. Thus, the angle between projected rays from all points in the scene can be simultaneously varied, by changing camera vergence β alone. θ_v is termed as *point vergence*. Point vergence is a controllable parameter, independent of surface normal, and affects specular intensity difference at scene points. The relation between point vergence and camera vergence for orthographic projection is simply, $\theta_v = \beta = \beta_1 + \beta_2$. In the case of perspective projection, viewing direction at each point in the scene, varies with respect to either viewpoint, i.e., point vergence varies across the scene. To define a single controllable parameter which affects specular intensity differences over the scene, point vergence can be averaged over

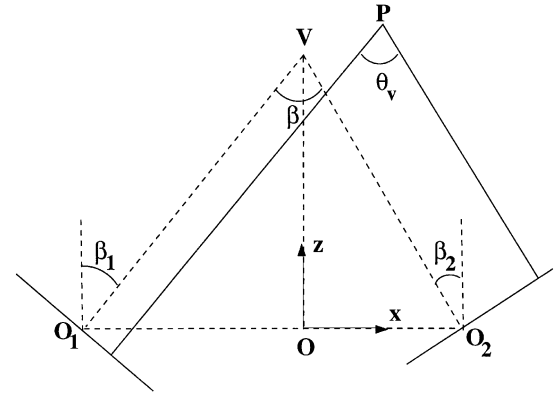


Figure 3. Point vergence and camera vergence under orthographic projection.

a workspace which could be the entire stereo field of view. If the workspace is defined explicitly in world coordinates, then a relation can be obtained between the mean value of point vergence and baseline. Therefore, the baseline indirectly controls I_{diff} over the workspace. Alternatively, the minimum value of the point vergence over the workspace could be chosen.

Vergence is related to depth resolution, an important design parameter. Depth accuracy, and hence resolution, are limited by spatial image quantization amongst other factors. The depth resolution attainable at any point is directly proportional to vergence (see Appendix A.1), assuming quantization is the primary cause for matching errors. Achieving maximum depth resolution therefore conflicts with the requirement of minimizing intensity difference over the scene.

4. Binocular Stereo

Determining maximum acceptable vergence in the presence of specular reflection can be formulated as a constrained optimization problem, as described below. We use a coordinate system (Fig. 4) with every scene point mapped to the origin O, and its surface normal is described by a unit vector $\hat{n}$ pointing away from O.

The aim is to attain maximum vergence in order to achieve best depth resolution. Hence, a suitable objective function f_{obj} is:

$$f_{\text{obj}} = \hat{v}_1 \cdot \hat{v}_2 \tag{3}$$

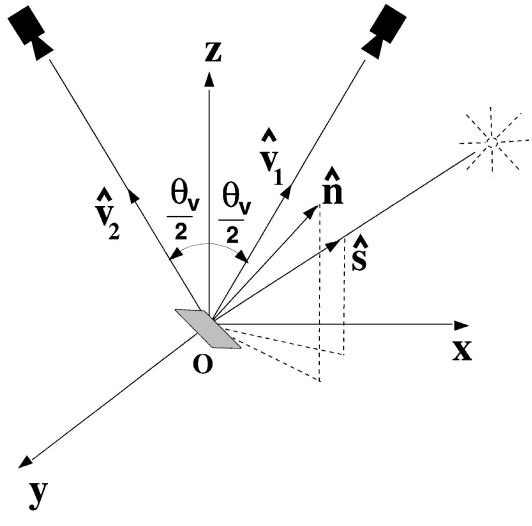


Figure 4. Coordinate system used for the stereo problem. Each scene point is mapped to its origin O.

To limit intensity difference at every scene point, the following constraint ($c1$) is imposed:

$$I_{\text{diff}} < T \tag{4}$$

where T is a threshold. From a statistical perspective, restricting I_{diff} amounts to limiting variance of specular intensity difference in any local region (see Appendix A.2). The cameras are restricted to lie in the positive x-z plane, and tilt symmetrically about the z-axis. These constraints ($c2$) can be expressed as:

$$\begin{aligned} \hat{v}_1 \cdot \hat{j} &= \hat{v}_2 \cdot \hat{j} = 0 \\ \hat{v}_1 \cdot \hat{k} &= \hat{v}_2 \cdot \hat{k} > 0 \end{aligned} \tag{5}$$

where $\hat{i}$, $\hat{j}$ and $\hat{k}$ are unit vectors along the x, y and z axes, respectively. To avoid grazing incidence and viewing angles, constraints ($c3$) are imposed:

$$\hat{v}_1 \cdot \hat{n}, \;\; \hat{v}_2 \cdot \hat{n}, \;\; \hat{s} \cdot \hat{n} > 0 \tag{6}$$

The optimization problem can now be stated:

$$\begin{aligned} &\text{Minimize:} \;\; f_{\text{obj}} \\ &\text{subject to constraints:} \;\; (c1, c2, c3) \end{aligned} \tag{7}$$

Note that the dot product of the two viewing vectors represents the cosine of point vergence. Therefore, minimizing the dot product amounts to maximizing vergence. The variables are $\hat{v}_1, \hat{v}_2, \hat{s}$ and $\hat{n}$. Solving the above problem, the optimal viewing directions $\hat{v}_1^{\text{opt}}$ and $\hat{v}_2^{\text{opt}}$ and hence the optimal vergence θ_v^{opt}, can be obtained independent of $\hat{s}$ and $\hat{n}$.

To demonstrate a particular solution, the expression for specular intensity given by (1) is used in constraint ($c1$). Dividing both sides by K_s, the constraint can be written as:

$$I_{\text{diff}}/K_s < T/K_s \tag{8}$$

It can be seen that T/K_s is an independent parameter which we call the *relative threshold*. Roughness σ is also unconstrained because surfaces in the scene are unknown. Thus, the *optimal vergence θ_v^{opt} is a function of surface roughness σ and relative threshold T/K_s*.

The optimization problem is solved numerically and the relationship obtained between θ_v^{opt}, σ and T/K_s is shown in Fig. 5. The salient features of this relationship are:

- The optimal vergence increases with roughness. The reason is that I_{diff} weakens with increasing roughness allowing larger vergence. The surface progressively

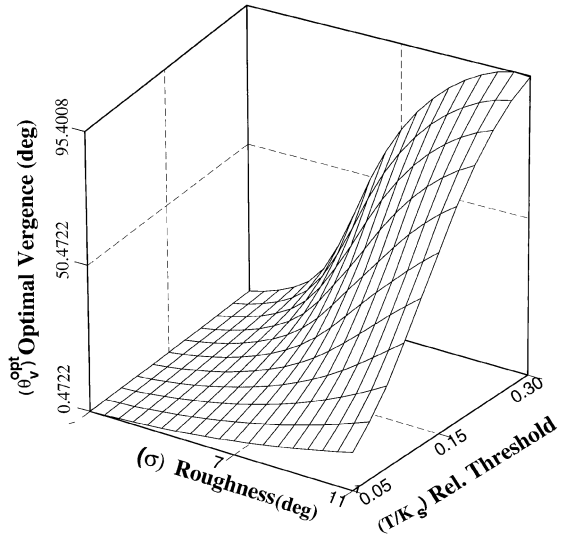


Figure 5. Graph illustrating the relationship between roughness, relative threshold and optimal vergence.

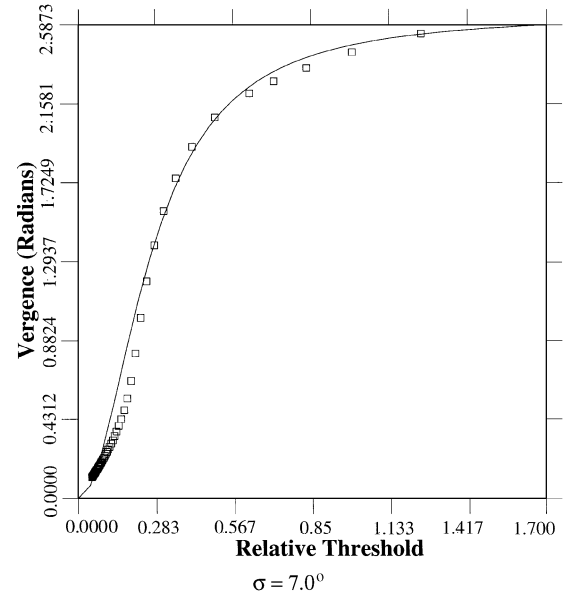


Figure 6. Graph illustrating fitting of vergence data to the approximating function. The solid line indicates the approximating function and the dots indicate vergence data.

behaves in a diffuse manner, and thus the effects of specular reflection on matching diminish.

- The optimal vergence also increases with relative threshold. This is perceivable because a larger threshold permits larger variation in I_s.

The monotonically increasing relationship of vergence with roughness implies that if the lowest roughness value in the scene is known, then the corresponding optimal vergence can be used for stereo. Arguing similarly, a conservative lower bound for the relative threshold is sufficient to configure a system that produces low intensity difference for all scene points. Variations to the general problem can now be considered by modifying the constraints, however, the approach for determining the optimal stereo configuration remains unchanged. For example, the normal vectors at all points could be constrained to lie in one plane, like those of a cylinder.

Since we do not have a closed-form expression for optimal vergence in terms of relative threshold and roughness, we pursued a functional approximation. If $\tilde{\theta}_v^{\text{opt}}$ approximates θ_v^{opt}, then

$$\tilde{\theta}_v^{\text{opt}} = \frac{a\,(T/K_s)^2\,\sigma^2}{(T/K_s)^2 + b\,\sigma^2} \tag{9}$$

where a and b are constants obtained numerically using the Levenberg-Marquardt algorithm. Figure 6 illustrates the fitting of vergence data to the approximating function for the given value of σ. The vergence data was obtained by solving the optimization problem (Eq. 7) numerically.

To make the relationship in Fig. 5 usable, a correspondence measure is required that is sensitive to changes in relative threshold and degrades gracefully. The normalized correlation coefficient (NCC) measures the degree of linear relationship between intensities in image windows. It is invariant to scaling of the intensities in the windows. With matching windows W in the two images, each containing N pixels and having intensities $I_1^{(i,j)}$ and $I_2^{(i,j)}$, NCC $= 1$ if $I_1^{(i,j)} = I_2^{(i,j)}$, $(i, j) \in W$, i.e., if the corresponding surface is Lambertian. However, due to specular reflection the intensities are not equal, hence NCC deviates from 1. The deviation is proportional to E, where:

$$E = \frac{1}{N} \sum_{(i,j)\in W} \left(\frac{I_1^{(i,j)} - I_2^{(i,j)}}{K_s} \right)^2$$

assuming that all points in each window are identically scaled in specular intensity by K_s. Since we limit intensity difference at all corresponding points by T, it follows that $E < (T/K_s)^2$; thus, NCC is sensitive to changing relative threshold. A closely related metric, the sum of squared differences (SSD), can also be made sensitive to variations in relative threshold, instead of

absolute threshold, by normalizing using the maximum intensity in the window.

The exact value of the relative threshold when mismatches begin to occur, the breaking threshold, depends on diffuse texture of the surface which are diverse, making its estimation a hard problem. Note that this problem is inherent to stereo matching, and it is only natural that the threshold appears in our formulation. Adopting a conservative lower bound for the relative threshold results in small vergence which in turn implies poor depth resolution. We will show that the problem is mitigated by using trinocular stereo.

5. Experiments

We illustrate the effect of vergence on stereo matching using surfaces with different roughness. For these experiments, we use a 5 degree of freedom SCARA (Adept) robot (see Fig. 7). The end-effector is equipped with a camera to obtain different viewing directions.

We use two uniformly rough cylindrical objects wrapped with different surfaces (see Fig. 8); a gift wrapper and a roughened xerox quality paper. Their surface roughness was measured (Bhat and Nayar, 1994), and the values obtained are $\sigma = 3.5^\circ$ and $\sigma = 6.3^\circ$, respectively. The differing roughness can be also be seen from the object images.

In order to use approximately the same relative threshold, similar random patterns on the surfaces were marked. Images obtained at equal angles about the z-axis are matched along scanlines containing texture. We have imposed the scanline epipolarity constraint by ensuring that the robot moves in the x-z plane only, and by using imaging optics that approximates orthographic projection. For each surface, depth obtained along a scanline at different vergences is shown in Fig. 7. It can be seen that for each surface large depth errors are computed at larger vergence: 8.0° and 11.0° respectively, although a higher vergence is acceptable for the rougher surface. For the smoother surface, mismatches are confined to the highlight region over which the variation of I_{diff} is large.

6. Trinocular Stereo

While binocular stereo as described earlier is viable in structured environments where surface roughness can be estimated, it is not practical in general scenes. Further, if the vergence corresponding to the lowest roughness estimate in the scene is used, then the depth resolution obtained for rougher surfaces is suboptimal. Thus, we seek an alternative scheme.

Figures 9(a)–(c) show schematics of a trinocular system configured such that the intensity difference at a point, with varying surface roughness, is constrained to a threshold in *at least* one pair of views. Therefore, depth of the point can be accurately computed in at least one stereo pair, regardless of surface roughness. While the configuration need not be limited to three sensors, increasing the number of sensors makes stereo implementation cumbersome.

We analyze a planar symmetric trinocular stereo system (Fig. 9(d)) with α as a single configurable parameter. Other configurations too can be used, for example, a non-planar system in which one sensor is placed at each corner of an equilateral triangle. But we choose the planar system since it is simple. Hence, the following geometrical constraints ($d1$) hold:

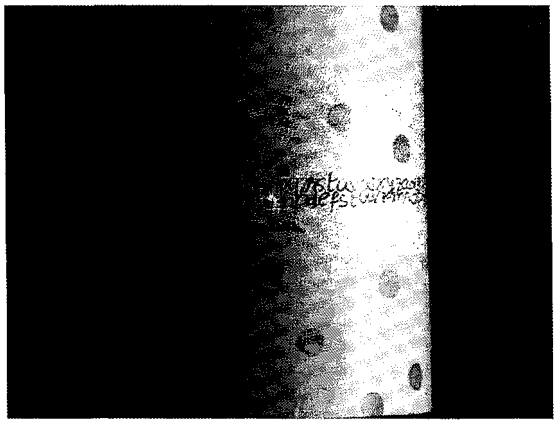
$$\begin{aligned} \hat{v}_1 \cdot \hat{v}_2 &= \hat{v}_2 \cdot \hat{v}_3 \\ \hat{v}_m \cdot \hat{k},\ \hat{v}_m \cdot \hat{n},\ \hat{s} \cdot \hat{n} &> 0 \\ \hat{v}_m \cdot \hat{j} &= 0, \quad m = 1, 2, 3 \end{aligned} \tag{10}$$

For any scene point, I_{diff} must not be too large in at least one stereo pair. This constraint ($d2$) can be expressed as:

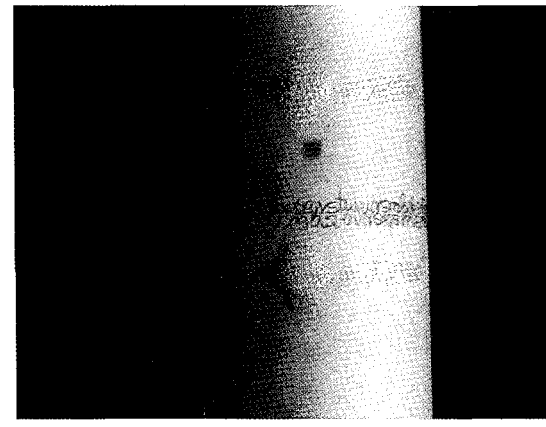
$$\exists (i,k) \left(\left| I_s^i - I_s^k \right| < T \right), \quad k \neq i,\ 1 \leq i, k \leq 3 \tag{11}$$

Note that the two views which satisfy the above constraint can change from one scene point to the next. Therefore, if the constraint is satisfied for all scene points, then conceivably an algorithm can be designed that switches between different stereo pairs to construct a complete depth map.

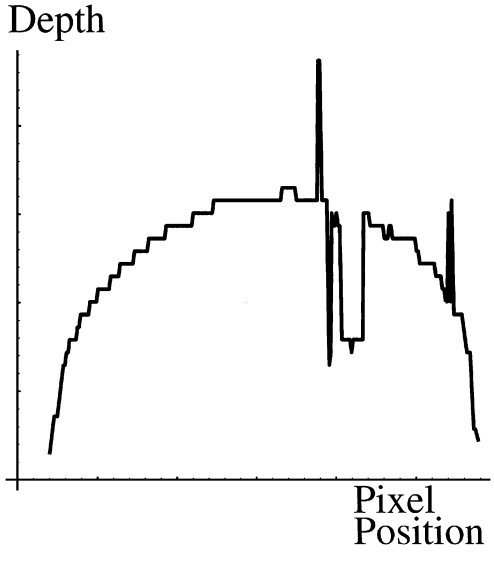
We analyze the following problem: Determine those values of the parameter α which satisfy the constraints $d1$ and $d2$. Like in the case of binocular stereo, the relative threshold T/K_s and the roughness σ are free parameters. The problem is solved numerically, and Fig. 10(a) illustrates the corresponding solution space (α vs. σ) for a given value of T/K_s. The unshaded region marked A denotes unacceptable vergences while the shaded region represents acceptable vergence values. Notice that all $\alpha > \alpha^{\text{opt}}$ (Region B) are acceptable values for any roughness value. In other words, α^{opt} denotes that vergence beyond which it is ensured that the intensity difference does not exceed the chosen value of threshold in at least one pair of views for any



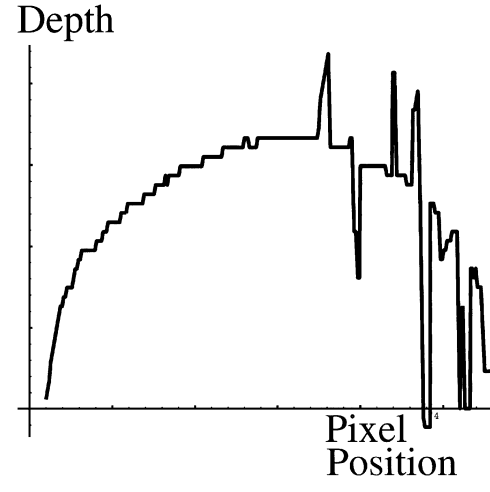
(a) $\sigma = 3.5^\circ$



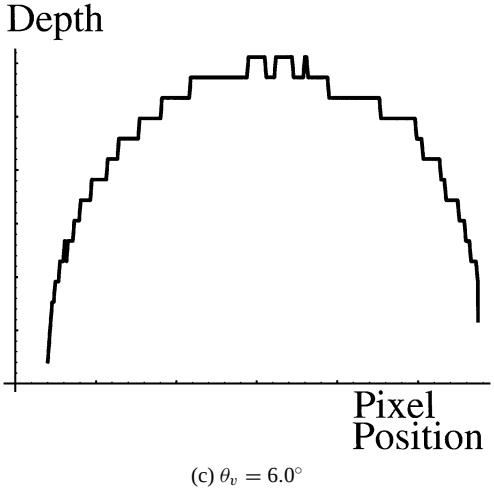
(d) $\sigma = 6.3^\circ$



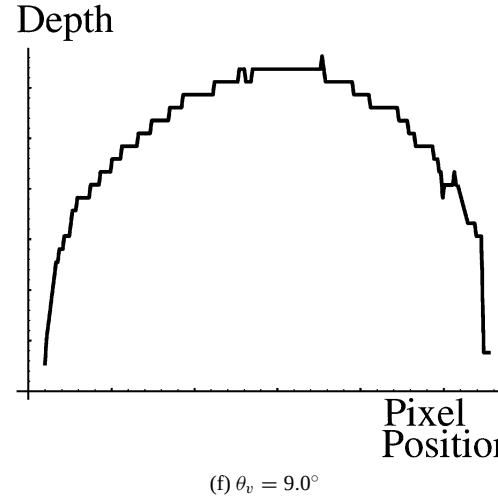
(b) $\theta_v = 8.0^\circ$



(e) $\theta_v = 11.0^\circ$

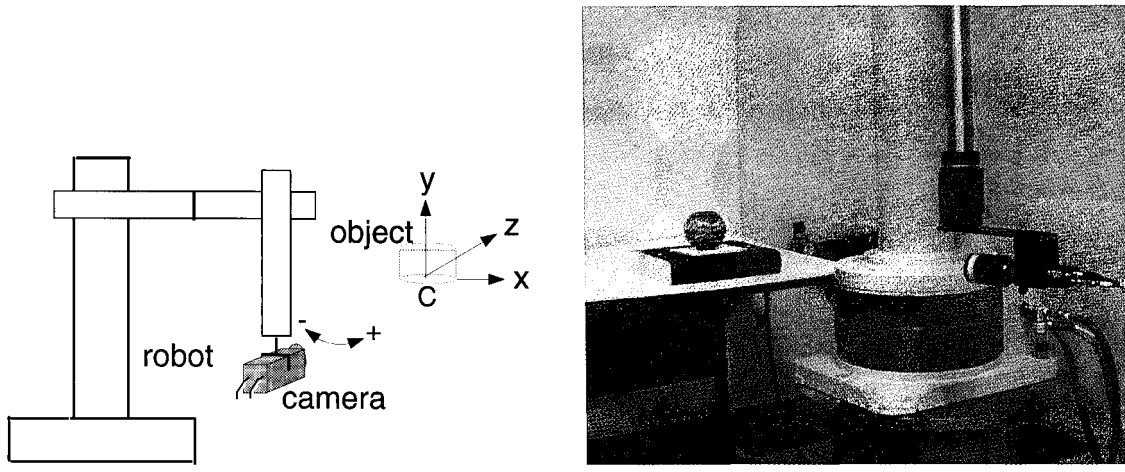


(c) $\theta_v = 6.0^\circ$



(f) $\theta_v = 9.0^\circ$

Figure 7. Effect of varying vergence on correspondence. (a)–(c) Image of the object with gift wrapper surface and depth obtained along a scanline using the vergence values shown. (d)–(f) Image of the object with rough xerox paper surface and depth obtained along a scanline using the vergence values shown. Notice that for both surfaces, depth is incorrectly recovered at larger vergence, although a relatively higher vergence is acceptable for the rougher surface.



(a) (b)

Figure 8. The experimental setup. (a) Diagram of a robot with a camera fixed to its end-effector. (b) A photograph of the setting.

scene point. This is true for arbitrary surface roughness. α^{opt} is termed as the *minimum acceptable vergence*. Figure 10(b) illustrates the variation of α^{opt} with T/K_s. The monotonically decreasing relationship suggests that a conservative lower bound for T/K_s will provide good depth resolution unlike binocular stereo. Another advantage of this approach is that the depth resolution obtainable is greater than its binocular counterpart for smoother surfaces since α^{opt} is much higher than the corresponding θ_v^{opt}. Finally, note that the binocular stereo solution is subsumed herein.

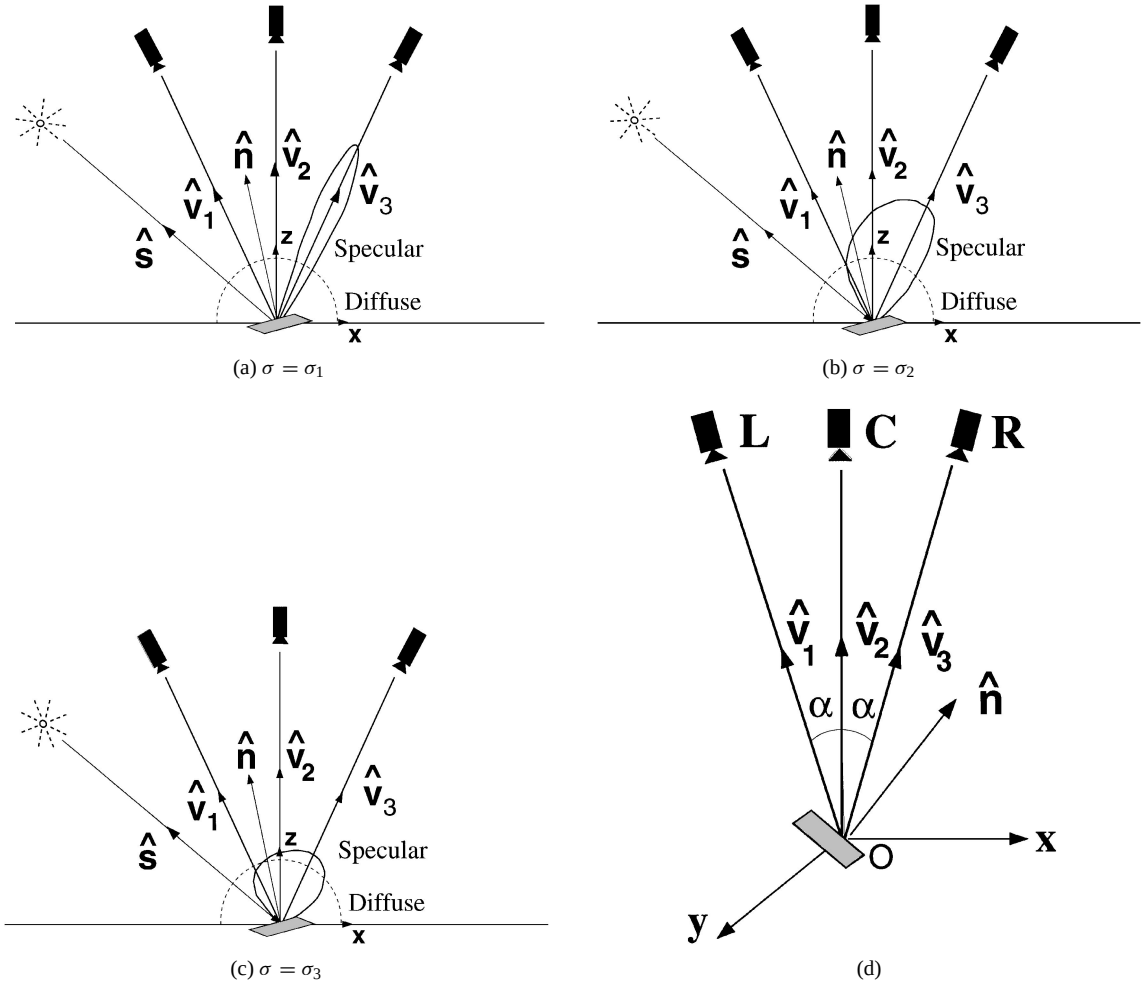
7. Reconstruction

We describe an algorithm for matching three views obtained using a configuration with $\alpha > \alpha^{\text{opt}}$. The three designated stereo image pairs are, (L, R), (C, R) and (L, C). The essence of the algorithm lies in determining which stereo pair provides a "good" depth estimate for any point in the scene.

To evaluate goodness of a match, the following confidence tests are used: (i) $C1$: Compare the NCC value obtained with a predefined threshold. Only if the normalized correlation value is higher, accept the match. At a wrong match, texture and shading between the windows being different, similarity is expected to be poor. (ii) $C2$: Let I_1 and I_2 denote two stereo images. If x_b is the current match in image I_2 for pixel x_a in I_1, then reverse the search and find the corresponding pixel for x_b by searching in I_1. This match must coincide with x_a if x_b and x_a are corresponding points. However, both tests will work only when there is sufficient albedo variation on the surface since they are based on window similarity. If there is no significant albedo variation, then a specular region of one image would match the specular region of the other with high NCC value, though they are truly corresponding. Hence, it would be not be possible to reject it using test $C1$. Similarly, test $C2$ would not be able to reject a match between two specular regions. Therefore, an alternate test $C3$ (Ching et al., 1993) using the monotonicity constraint could be adopted: (iii) $C3$: The relative ordering of projections of scene points along the epipolar lines in the two images must be preserved. To keep the algorithm simple, we discuss it using tests $C1$ and $C2$, but is easy to incorporate $C3$ (or for that matter any other confidence test).

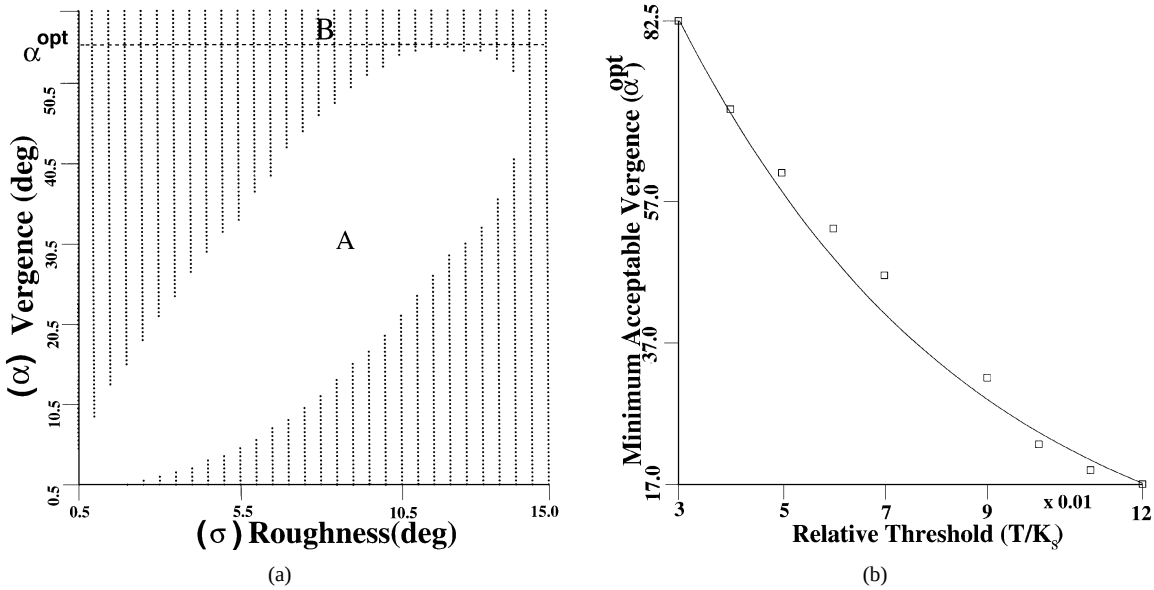
Algorithm

- *Step 1.* Initialize the current stereo pair to (L, R). The reason for choosing this pair is that it yields maximum vergence thereby providing good depth resolution.
- *Step 2.* Choose pixel x_L in L with adequate surrounding texture.
- *Step 3.* For the pixel x_L, do
  - — Find a matching pixel in R. Using confidence tests $C1$ and $C2$, evaluate the goodness of match. If the match is good, compute depth and go to Step 2. If not, the current stereo pair (L, R) cannot be used for matching pixel x_L, and hence perform the following step.



(a) $\sigma = \sigma_1$ (b) $\sigma = \sigma_2$ (c) $\sigma = \sigma_3$ (d)

Figure 9. A trinocular configuration illustrated on three surfaces with $\sigma_1 < \sigma_2 < \sigma_3$. (a) $|I_s^1 - I_s^2| < T$, (b) $|I_s^2 - I_s^3| < T$, and (c) $|I_s^1 - I_s^2| < T$, $|I_s^2 - I_s^3| < T$ and $|I_s^1 - I_s^3| < T$, (d) layout of the system with sensors labeled L, C, R.



(a) (b)

Figure 10. The trinocular stereo solution space. (a) $\alpha - \sigma$ plot with relative threshold $(T/K_s) = 0.06$. Region B denotes the region $\alpha > \alpha^{\text{opt}}$ where all vergences are acceptable. (b) Variation of α^{opt} with T/K_s.

— Set (L, C) as the current stereo pair, and find the corresponding pixel for x_L in C. Evaluate the confidence of matching using $C1$ and $C2$. If the match is good, then compute the corresponding pixel x_R in image R by transformation. Compute depth using x_L and x_R and go to Step 2. If the match is not good, then the current stereo pair too has failed to establish correspondence, and hence perform the following step:
— Set C, R as the current stereo pair. If S is the search range, find that pixel x_C in C within the range $(x_L - S, x_L + S)$ which matches well with x_R in R, and together map onto x_L when transformed into the image coordinate system of L. The mapping under orthographic projection is given in Appendix 10. Thus, we establish consistent correspondence for x_L in the three images. Compute depth using x_L and x_R and go to Step 2. If no such consistent correspondence can be established, then depth cannot be computed at point x_L, hence go back to Step 2 for processing the next pixel.

The computational complexity is proportional to the number of switches between stereo pairs that can take place in evaluating the complete depth map. If the surface geometry is known, then the switching sequence and the total complexity can actually be evaluated analytically.

8. Experiments

We present trinocular stereo experiments with objects of different roughness. Here we do not estimate surface roughness as required in the case of binocular stereo. Figure 8 shows the photograph of the experimental stereo setup used. As with the earlier experiments on binocular stereo, different vergence values are obtained by moving the camera in a circle about a center close to which objects are placed.

Figure 11 shows trinocular stereo images of an egg-shaped object. The object is relatively rough, as is perceivable from the spread out highlight region. Notice that the specular region shifts in the image space differently from the neighboring texture. The images were obtained using $\alpha = 7.5^\circ$, i.e., the binocular vergence with the left and right images is 15.0°. This value was chosen to keep search ranges relatively small. A large value for α will necessitate a coarse to fine matching strategy which we have not implemented currently. We used a single distant light source in order to keep the experiments consistent with the theory. The performance of the reconstruction algorithm is first illustrated on one scanline. Figure 12 compares our algorithm with naive binocular stereo matching (using views L and R). It can be seen that our algorithm works well demonstrating robustness to specularities. A complete depth map is shown in Fig. 13.

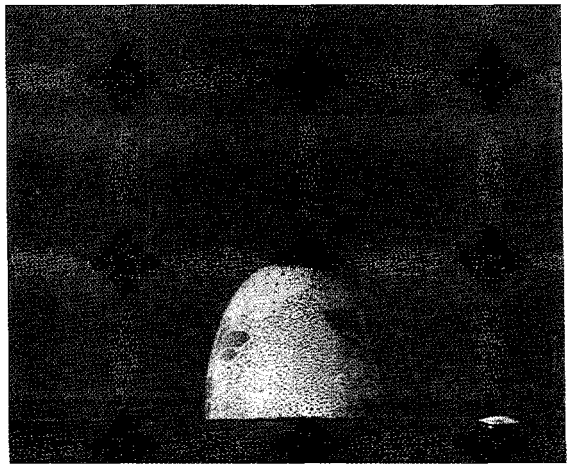
The second scene (Fig. 14) contains two objects with different surfaces: a vase shaped object whose roughness varies over the surface, and a cylindrical object with unknown roughness. Again, $\alpha = 7.5^\circ$ was used to capture the trinocular images. Figure 15 illustrates the depth map of the scene produced by the reconstruction algorithm. The experiments demonstrate that the algorithm works reasonably well in the case of objects with different reflectance characteristics, an essential requirement for a practical stereo algorithm. There are a few points at which depth is incorrectly estimated especially at depth discontinuities like the top of the vase, where window measures like NCC are not robust.

9. Conclusion

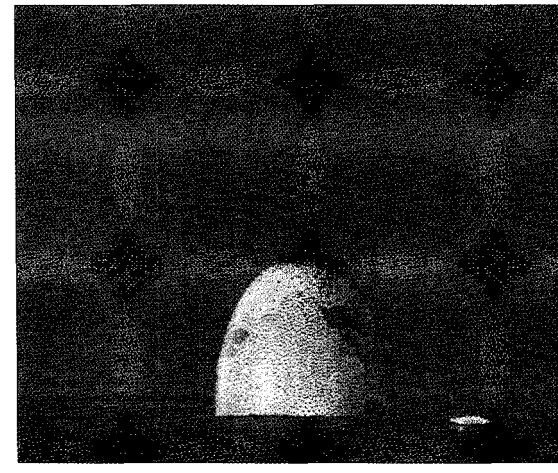
We summarize the main results and contributions of the paper:

- We developed a physically based approach for reliable stereo in presence of specular reflection.
- A scene-independent binocular stereo solution was obtained by minimizing intensity differences at corresponding points while maximizing depth resolution. The solution was shown to be a function of surface roughness. Hence, this configuration is usable in structured environments where roughness can be assessed.
- Trinocular stereo configurations were derived to obviate the need for surface roughness measurement. These configurations can be used in scenes containing unknown objects with possibly varying reflectance properties.
- We have developed a simple algorithm for reconstructing accurate depth maps from three views of a scene that include specular reflections from surfaces of unknown roughness.

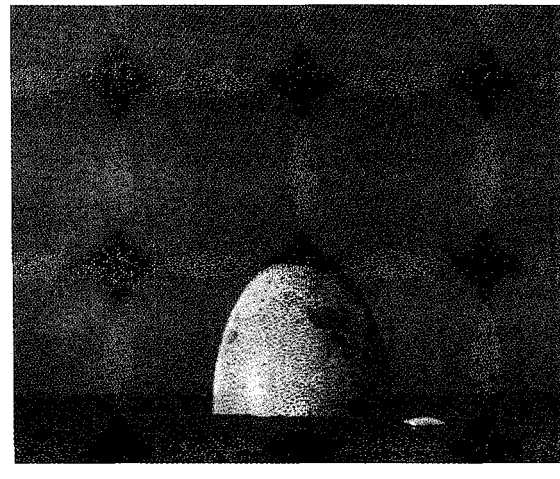
There is a close relationship between the relative threshold and the measure used for correspondence. Consider a robust correlation measure like in (Bhat and Nayar, 1996) that is relatively insensitive to relative threshold, i.e., when the relative threshold



(a) Left (L)

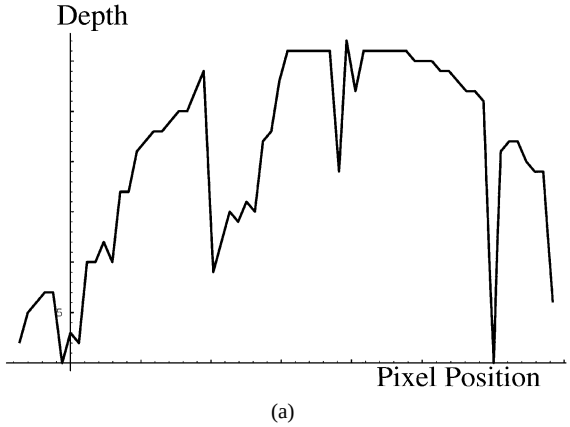


(b) Center (C)

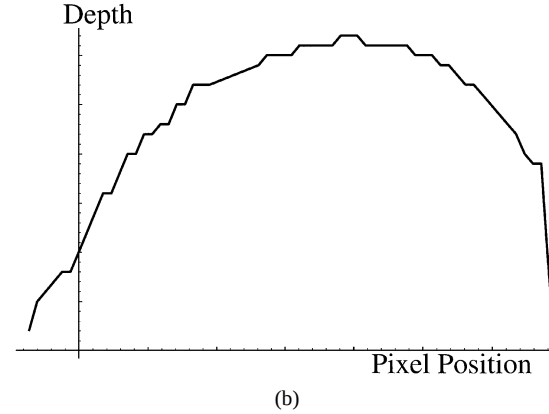


(c) Right (R)

Figure 11. Trinocular stereo images of an egg-shaped object, obtained using $\alpha = 7.5^\circ$. The images are gamma corrected to enhance contrast for display.

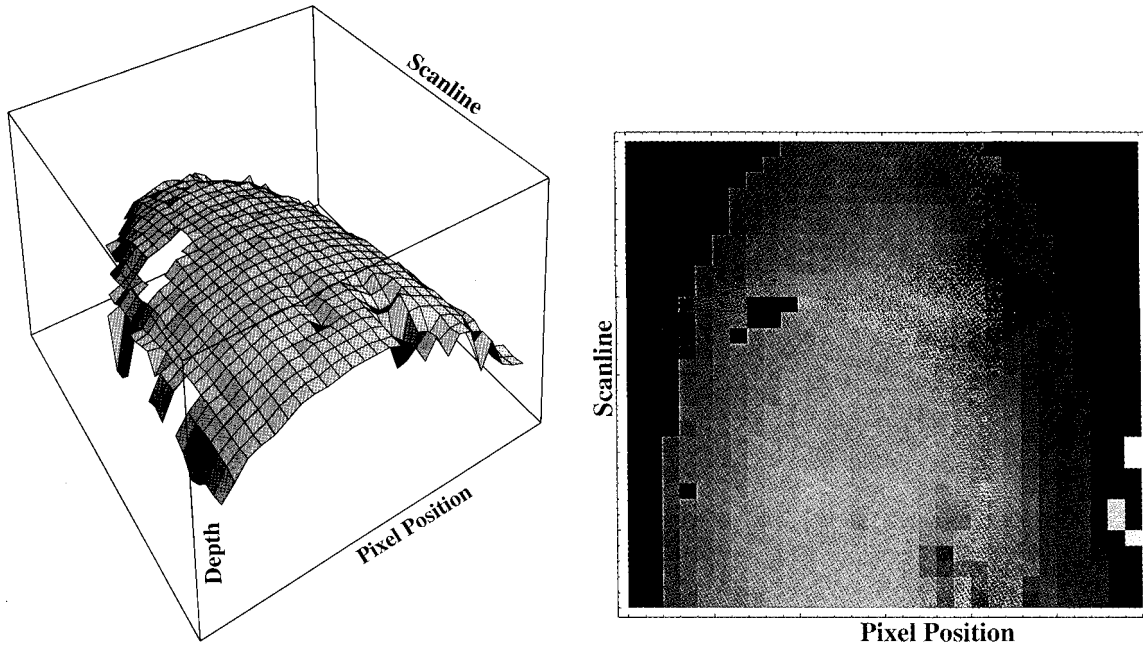


(a)



(b)

Figure 12. Depth computed for the egg-shaped object along a scanline, (a) using views L and R; and (b) using the proposed reconstruction algorithm which uses all three views.



(a) (b)

Figure 13. Depth map of the egg-shaped object scene computed using our algorithm, shown as a (a) surface plot, (b) density plot.

varies in a certain range, the measure remains unchanged. This is unlike normalized correlation which changes continuously with relative threshold. Consequently, the breaking threshold value corresponding to the robust measure would be larger. Therefore, in the case of binocular stereo, even for smoother surfaces a larger range of vergence values would be acceptable with a robust operator. Similarly, in the case of trinocular stereo, the optimal vergence would be lower implying a larger range of usable configurations which is desirable.

Appendix

A.1. Depth Resolution and Vergence

Below, we express the depth z of any point $P(x, z)$ in the scene with respect to the world coordinate system when the cameras (Fig. 3) tilt equally about the z axis. Let x_l and x_r be the projections of the point on the left and right image planes. The distance O_1O and O_2O are represented by L_1 and L_2, respectively. The depth z of P is:

$$z = \frac{(L_1 + L_2) - (x_l - x_r)\left(\sec\frac{\beta}{2}\right)}{\left(2\tan\frac{\beta}{2}\right)}, \quad \beta < \pi \qquad \text{(A1)}$$

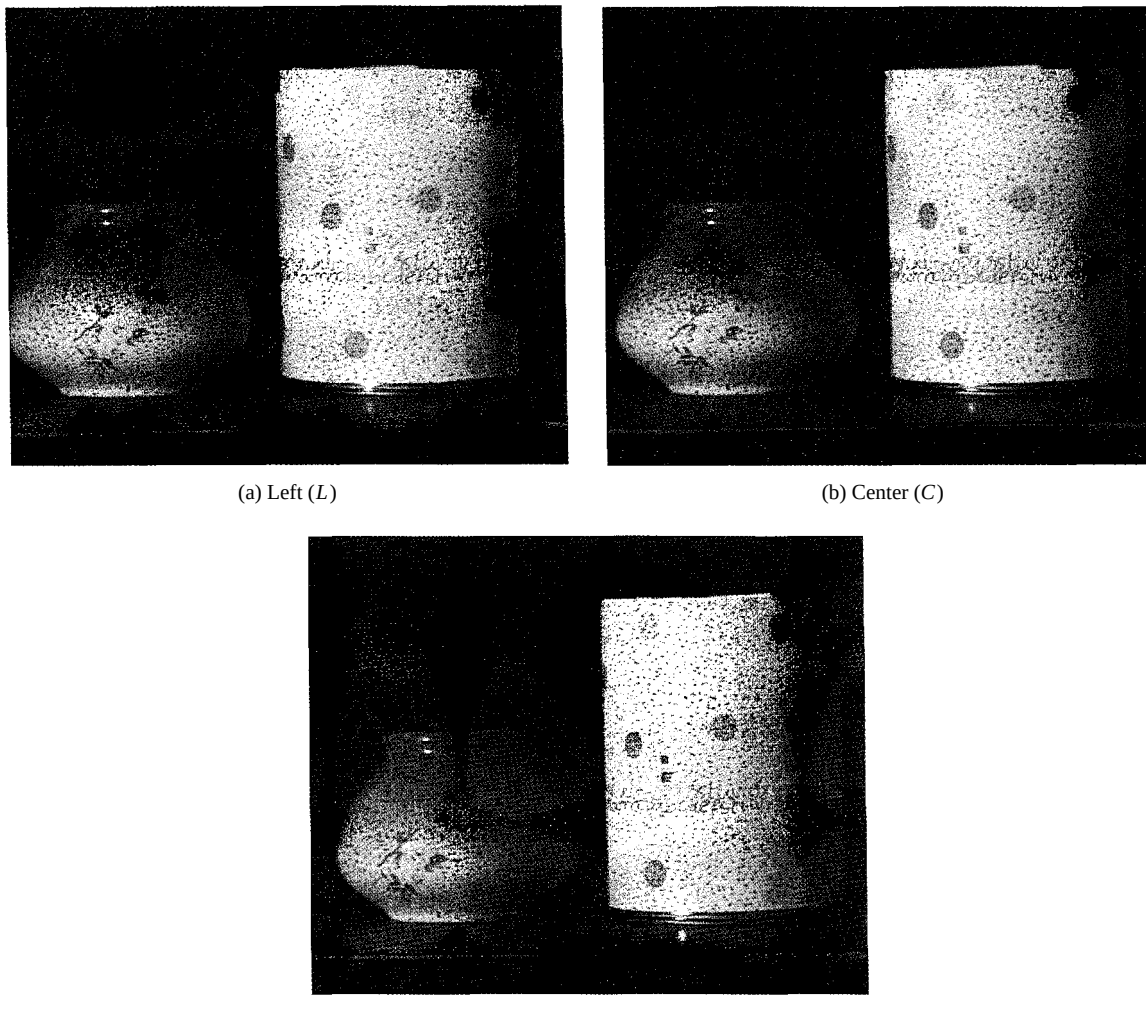
Two quantities to estimate depth resolution are the absolute range error (Verri and Torre, 1986) and the expected range error (Rodriguez and Aggarwal, 1988). Here, the absolute error Δz is used that is given by:

$$\Delta z = \left|\frac{\partial z}{\partial x_l}\right|\Delta x_l + \left|\frac{\partial z}{\partial x_r}\right|\Delta x_r + \left|\frac{\partial z}{\partial \beta}\right|\Delta\beta \qquad \text{(A2)}$$

Δx_l and Δx_r represent errors due to matching inaccuracies and quantization. $\Delta\beta$ represents error in camera vergence due to mechanical defects and improper calibration of the cameras. If we assume that error in camera vergence is negligible, then $\Delta\beta = 0$. Further, if matching is achieved to pixel accuracy, then depth errors are primarily due to quantization. Using (A1) and (A2), the depth error can be expressed as:

$$\frac{\Delta z}{(\Delta x_l + \Delta x_r)} = \frac{1}{2\sin\frac{\beta}{2}} = \frac{1}{2\sin\frac{\theta_v}{2}} \qquad \text{(A3)}$$

where the term, $(\Delta x_l + \Delta x_r)$, represents correspondence error due to quantization. From (A3), it follows that the absolute depth error is inversely proportional to vergence. In other words, depth resolution increases with increasing vergence.



(a) Left (L) (b) Center (C)

(c) Right (R)

Figure 14. Trinocular stereo images of a scene with two objects of different roughness, obtained using $\alpha = 7.5^\circ$. The center of the cylinder and the vase-shaped object exhibit significant specular reflection.

A.2. A Statistical Interpretation

The reason for minimizing I_{diff} can be viewed in a statistical sense. We use the following model for the images in one dimension (Matthies, 1992):

$$I_1(x) = I_d(x) + I_{s1}(n(x))$$
$$I_2(x) = I_d(x + d_m(x)) + I_{s2}(n(x))$$

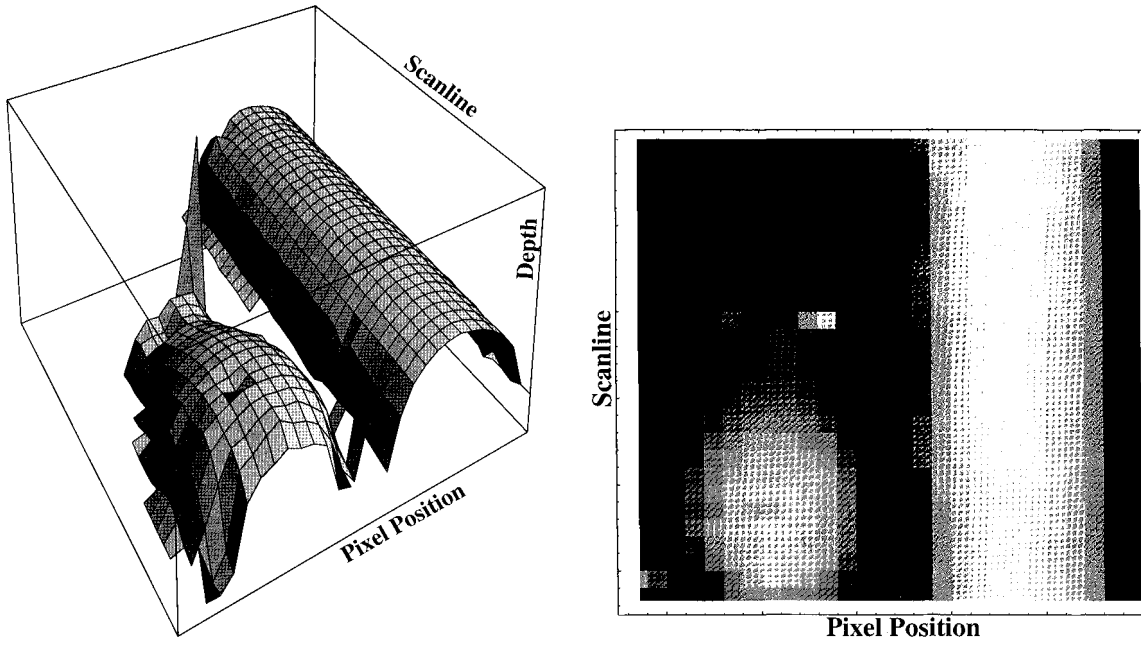
where $I_d(x)$ corresponds to the diffuse component, $I_s(n(x))$ refers to the specular component, $n(x)$ is the surface normal of the scene point projected as x, d_m is the disparity at point x. Because the surface normal varies randomly, we model the specular intensity difference between corresponding points as a random variable, and for simplicity we choose a uniform distribution as given by:

$$g(I_{\text{diff}} = |I_{s1} - I_{s2}|) = \frac{1}{T}, \quad 0 \le I_{\text{diff}} \le T \qquad \text{(A4)}$$

where the chosen viewing directions decide the threshold T.

To compute disparity at any point x_i in the left image, we use the absolute values of the differences between point intensities in windows, for each candidate disparity. The distance (error) between two windows being matched is expressed as:

$$\hat{e}(x_i; d) = [e(x_i + \Delta x_1; d), \ldots, e(x_i + \Delta x_k; d), \ldots]$$
$$e(x_i + \Delta x_k; d) = |I_1(x_i + \Delta x_k) - I_2(x_i + \Delta x_k + d)|$$



(a) (b)

Figure 15. Depth map of the two object scene computed using our algorithm, shown as a (a) surface plot, (b) a density plot.

where Δx_k represents pixels in the windows, d is the candidate disparity assumed constant around x_i. Under the distribution model adopted (Eq. (A4)) for intensity differences, the conditional p.d.f. of $\hat{e}$ (the likelihood function) at the correct correspondence estimated by d_0 is:

$$f(\hat{e} \mid d = d_0) = \frac{1}{T^N}$$

where N is the number of pixels in a window.

Note that the true disparity d_m could be different from d_0 because the latter is being estimated from discrete images. To increase the probability of obtaining the correct estimate d_0, T must be minimized, i.e., minimizing T amounts to maximizing the likelihood function f.

A.3. Correspondence in Trinocular Stereo

Here we relate the x-coordinates of the projections of a point on the three stereo images. Figure A.1 shows orthographic projections of a point $P(x, z)$ on the images L, R and C. The projections are denoted by x_L, x_R and x_C, respectively.

The depth of P, the z component, in the world coordinate system (which coincides with the image coordinate system C) is given below, with the stereo pair used being indicated in superscripts:

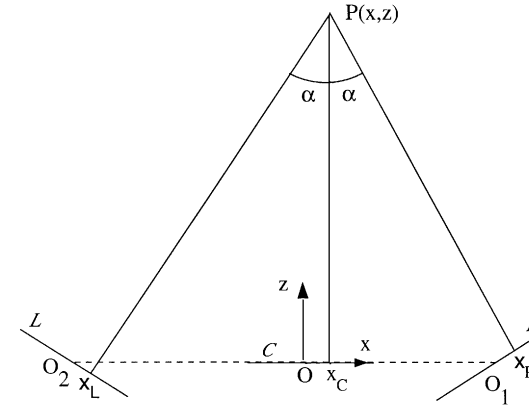


Figure A.1. Correspondence of a scene point in the three images under orthographic projection.

$$z^{(L,R)} = \frac{B + \frac{x_R - x_L}{\cos\alpha}}{2\tan\alpha}$$

$$z^{(L,C)} = \frac{B + 2\left(x_C - \frac{x_L}{\cos\alpha}\right)}{2\tan\alpha}$$

$$z^{(R,C)} = \frac{B + 2\left(\frac{x_R}{\cos\alpha} - x_C\right)}{2\tan\alpha} \qquad \text{(A5)}$$

where B is the distance O_1O_2. Equating z in the above relations,

$$x_L = 2x_C \cos\alpha - x_R \tag{A6}$$

Acknowledgments

This research was conducted at the Center for Research in Intelligent System, Department of Computer Science, Columbia University. It was supported in parts by ARPA contract DACA-76-92-C-007, a David and Lucile Packard Fellowship, and a NSF National Young Investigator Award. We wish to thank Sameer Nene for his help with the robot.

References

Barnard, S.T. and Fischler, M.A. 1982. Computational stereo. *ACM Computing Surveys*, 14(4):553–572.

Bhat, D. and Nayar, S.K. 1994. Stereo and specular reflection. Technical Report CUCS-030-94, Columbia Univ.

Bhat, D.N. and Nayar, S.K. 1996. Ordinal measures for visual correspondence. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 351–357.

Blake, A. 1985. Specular stereo. In *Proc. of the Ninth International Joint Conference on Artificial Intelligence*, pp. 973–976.

Brelstaff, G. and Blake, A. 1988. Detecting specular reflections using lambertian constraints. In *Proc. of the IEEE Computer Society International Conference on Computer Vision*, pp. 297–302.

Ching, Wee-Soon, Toh, Peng-Seng, Kap-Luk, and Er, Meng-Hwa 1993. Robust vergence with concurrent detection of occlusion and specular highlights. In *Proc. of the IEEE Computer Society International Conference on Computer Vision*, pp. 384–394.

Dhond, U.R. and Aggarwal, J.K. 1991. A cost-benefit analysis of a third camera for stereo correspondence. *International Journal of Computer Vision*, 6:39–58.

Gennert, M. 1988. Brightness-based stereo matching. In *Proc. of the IEEE Computer Society International Conference on Computer Vision*, pp. 139–143.

Ito, M. and Ishii, A. 1986. Range and shape measurement using three-view stereo analysis. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 9–14.

Matthies, L. 1992. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *International Journal of Computer Vision*, 8(1):71–91.

Nayar, S.K., Ikeuchi, K., and Kanade, T. 1991. Surface reflection: Physical and geometrical perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):611–634.

Nayar, S.K., Fang, Xi-Sheng, and Boult, T. 1993. Removal of specularities using color and polarization. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 583–590.

Okutomi, M. and Kanade, T. 1993. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363.

Oren, M. and Nayar, S.K. 1994. Generalization of the Lambertian model and implications for machine vision. In *Proc. of the European Conference on Computer Vision*, pp. 269–280.

Rodriguez, J.J. and Aggarwal, J.K. 1988. Quantization error in stereo imaging. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 153–158.

Smith, G.B. 1986. Stereo integral equation. In *Proc. of the AAAI*, pp. 689–694.

Torrance, K.E. and Sparrow, E.M. 1967. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57:1105–1114.

Verri, A. and Torre, V. 1986. Absolute depth estimate in stereopsis. *Journal of the Optical Society of America*, 3:297–299.

Wolff, L.B. and Angelopoulou, E. 1994. 3-d stereo using photometric ratios. In *Proc. of the European Conference on Computer Vision*, pp. 247–258.

Yachida, M., Kitamura, Y., and Kimachi, M. 1986. Trinocular vision: New approach for correspondence problem. In *Proc. of the Eighth International Conference on Pattern Recognition*, pp. 27–31.