Chapter 5

Precision: Effects of Foreshortening

Changes in latitude, changes in attitude.
— Jimmy Buffett

In this chapter we show how perspective foreshortening is manifest in the local spatial frequency representation of stereo images. Ours will be a forward-reasoning analysis, beginning with complete knowledge of the three-dimensional geometry of the scene and ending with its two-dimensional projection in the image plane. The primary result is the presentation (in Equation 5.10) of the Foreshortening Factor that allows us to compensate for arbitrary foreshortening effects without explicitly warping the images. This result makes no restrictions on the surface texture, and will not require the use of disparity derivatives. The complementary technique (starting with the projections to determine three-dimensional geometry) will be presented in Section 5.4.

To simplify the analysis, we assume the only object in the world is a textured flat plate that is either parallel to the image plane, or rotated about the vertical axis by some angle θ. We further assume that the stereo cameras have parallel optical (depth) and vertical (height) axes. Note that we can restrict our attention to the effects of foreshortening in one-dimensional image *scanlines*, rather than complete two-dimensional images, since all disparities will be horizontal under this assumption. Our world model will likewise be a two-dimensional slice through the three-dimensional scene. Figure 5.1 shows an overhead schematic of a horizontal slice through the world. We adopt the convention that parameters measuring distances in the world will be capitalized (e.g., X_S, Z_L), and those measuring

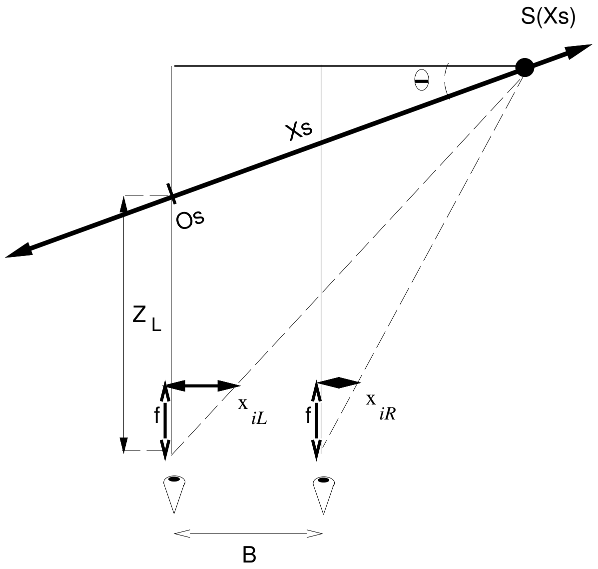


Figure 5.1: Overhead view of the foreshortening model. X_S is the distance from the point exactly in front of the left camera (the origin O_S at distance Z_L) to the point (S) on the plate being studied; x_{iL} and x_{iR} are the left and right pixel indices of the image of surface point S; the cameras are separated by baseline B and the surface tilts away from the cameras at angle θ.

pixel or camera distances will be lower case (e.g., x_{iL}, f).

Figure 5.2 illustrates the effect of this foreshortening in the frequency domain. To simplify the demonstration, a flat plate that has a surface texture with a single frequency component is used: a sine wave. The figure has two images of the plate on top, and the corresponding scalogram magnitude plots below. The head-on view of the plate on the left side of the figure has the expected scalogram; a horizontal line centered at the frequency of the sine wave, with some extra energy (dark regions) at the edges of the plate where it borders the plain white background. The rotated view also has a straight line in the scalogram, but it appears at higher frequencies and is no longer horizontal. This transformation will be quantified precisely in the closed-form *foreshortening factor* developed later in this chapter.

Although our ultimate goal is to find the disparity between two stereo images, we must first determine how the appearance of the object's surface texture will differ between them. Specifically, we want to know how the sampling rate varies between the two images. This is a geometric formulation; what matters is how much of the surface is being mapped to each pixel, not the actual surface texture (i.e., color intensity). So for each location X_S on the

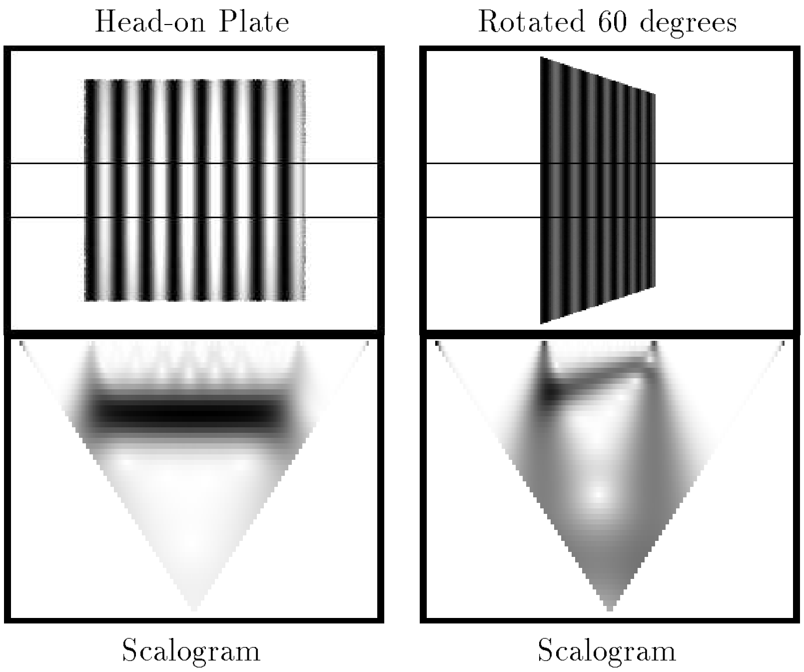


Figure 5.2: The effect of foreshortening on scalogram magnitude. Two views of a flat plate with a sinusoidal texture appear on top, and the scalogram magnitudes for their central scanlines appear below. The responses are similar, but are compressed to higher frequencies in the rotated view.

surface, we want to compare the pixel areas in the left and right images. Mathematically, we want to compare the left sampling rate $\frac{\delta X_S}{\delta x_{iL}}$ to the right sampling rate $\frac{\delta X_S}{\delta x_{iR}}$:

$$\text{Sampling ratio} \quad = \quad \frac{\frac{\delta X_S}{\delta x_{iL}}}{\frac{\delta X_S}{\delta x_{iR}}} = \frac{\delta x_{iR}}{\delta x_{iL}} \tag{5.1}$$

Simplifying the ratio in this way proves most useful. The resulting formula tells us we can compute the sampling ratio (which will be called *foreshortening factor* later) in *image space*, without having to explicitly model the distance X_S along the object. Unfortunately, it also implies that we need the disparity derivative (recall δdisparity$/\delta x_{iL}$ is simply $\delta(x_{iL} - x_{iR})/\delta x_{iL} = 1 - \text{Sampling Ratio}$). Since our ultimate goal is to estimate disparity, it would be best if we could avoid using both disparity and its derivative in our calculations (the derivative of a noisy signal will be even noisier). The remainder of this section will show how we can express this ratio with terms that do *not* require disparity derivatives.

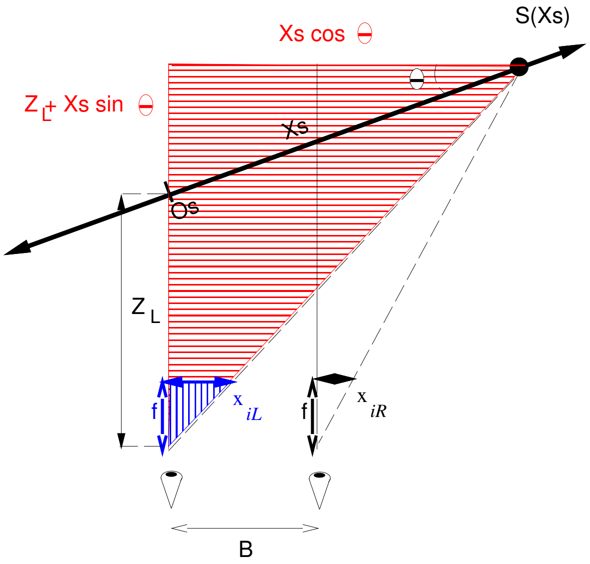


Figure 5.3: Overhead view of the foreshortening model. Similar triangles for the left camera geometry are highlighted (see Equation 5.3).

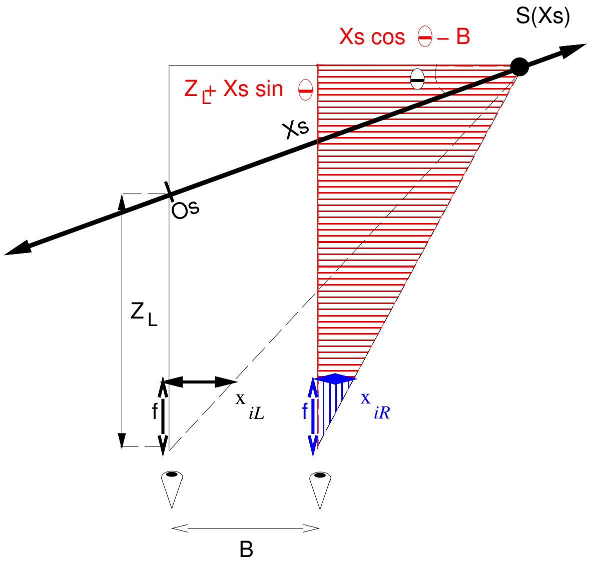


Figure 5.4: Overhead view of the foreshortening model. Similar triangles for the right camera geometry are highlighted (see Equation 5.4).

5.1 Relating Disparity to Surface Angle

How is disparity related to the surface angle? Equation 1.2 gives the disparity for an individual point, but we will now show how it varies across a surface. We will focus our attention on the distance from the left camera to the surface point immediately in front of it, expressing other depths in terms of this value Z_L.

Recall that disparity is the difference of the left and right pixel indices. So let's see how each of the left and right indices (x_{iL} and x_{iR}) relates to the surface angle θ. A quick look at Figure 5.1 shows us the general answer using similar triangles:

$$\frac{\textit{pixel index}}{\textit{focal length}} = \frac{\textit{X World Coordinate}}{\textit{Z World Coordinate}} \tag{5.2}$$

Figures 5.3 and 5.4 highlight the similar triangles for the left and right scene geometries. Applying Equation 5.2 to those figures we obtain expressions for x_{iL} and x_{iR}:

$$\frac{x_{iL}}{f} = \frac{X_S \cos\theta}{Z_L + X_S \sin\theta} \tag{5.3}$$

$$\frac{x_{iR}}{f} = \frac{X_S \cos\theta - B}{Z_L + X_S \sin\theta} \tag{5.4}$$

Equations 5.3 and 5.4 give us expressions for x_{iL} and x_{iR} in terms of the focal length f, baseline B, distance in front of the left camera Z_L, surface angle θ, and location on the surface X_S. These equations represent projections of the same surface point X_S into two image planes, and we can find the relationship between them by solving Equations 5.3 and 5.4 for X_S and setting them equal.

$$\frac{x_{iL} Z_L}{f\cos\theta - x_{iL}\sin\theta} = \frac{x_{iR} Z_L + Bf}{f\cos\theta - x_{iR}\sin\theta} \tag{5.5}$$

Solving Equation 5.5 for the right pixel index gives us:

$$x_{iR} = x_{iL}\left(1 + \frac{B}{Z_L}\tan\theta\right) - \frac{Bf}{Z_L} \tag{5.6}$$

And finally, recalling that disparity is the difference of the two indices:

$$\textit{disparity} = x_{iL} - x_{iR} = \frac{Bf}{Z_L} - x_{iL}\frac{B}{Z_L}\tan\theta \tag{5.7}$$

Equation 5.7 is nearly the answer we want. It relates disparity to the scene parameters, and does not depend on knowing the actual surface location. It does require knowledge of Z_L (distance to the surface point in front of the left camera), unfortunately, but we will eliminate this restriction below.

Equation 5.7 has some interesting interpretations. When the surface is frontoplanar (i.e., $\theta = 0$ and thus $\tan\theta = 0$) it reduces to the familiar expression relating disparity to depth from Equation 1.2; this is correct since all surface points would lie at the same depth Z_L. And for an arbitrary fixed angle θ the disparity *derivative* is constant, i.e., the disparity varies linearly with respect to the image location x_{iL}. While we won't take advantage of this property of the derivative, it could prove useful to shape-recovery techniques.

5.2 Expressing the Foreshortening Factor using Image Parameters

Now that we know how the disparity and pixel locations relate to surface angle, let us return to the Foreshortening Factor (Equation 5.1) and eliminate the derivative by substituting for x_{iR}:

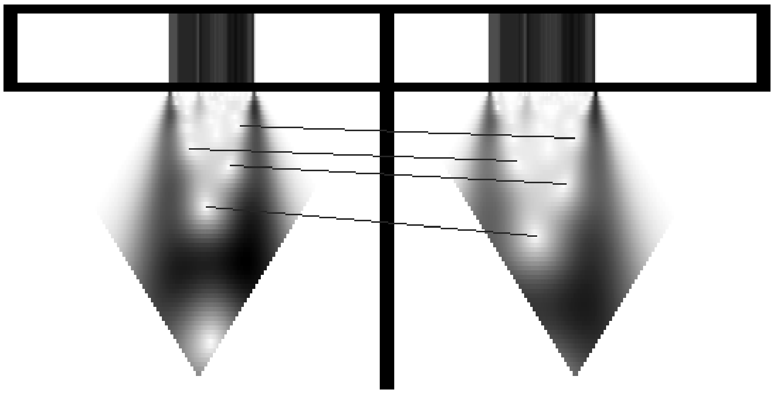


Figure 5.5: Left and right views of a surface tilted 65 degrees. Upper images are the central scanlines, lower images are their corresponding scalograms. You can see similar features in both scalograms: those in the left image are present at higher spatial frequencies because the left image is subject to greater foreshortening effects than the right image.

$$\begin{aligned} \text{Foreshortening Factor} &= \frac{\delta x_{iR}}{\delta x_{iL}} = \frac{\delta\left(x_{iL}\left(1+\frac{B}{Z_L}\tan\theta\right) - \frac{Bf}{Z_L}\right)}{\delta x_{iL}} && \textit{from Equation 5.6} \\ \textit{Geometric Form} &= 1 + \frac{B}{Z_L}\tan\theta && (5.8) \end{aligned}$$

This expression is very interesting. It tells us that for a given flat surface, the Foreshortening Factor is *constant* over both images of the surface. In other words, the local spatial frequencies of the left and right images are related by a simple constant scale factor. You can get a feel for this by visually tracking the low magnitude phase singularities (white spots) between the two image scalograms in Figure 5.5.

The fact that foreshortening causes frequency shifts has been noted in the literature (Fleet & Jepson, 1993), but no explicit model was given to explain it in the context of stereo vision (but see (Krumm, 1993; Malik & Perona, 1989) for single image texture-based models). Instead, the instantaneous frequency was recovered using a heuristic averaging technique. This technique yielded somewhat better results than the use of direct frequency, but did not take advantage of the scene geometry to compute the precise shift. This averaging technique also failed whenever the frequency shift caused the instantaneous frequency to fall outside the range of the filter in either of the images. Our model overcomes these problems by making use of all available frequency bands, rather than limiting attention to a small number.

The result in Equation 5.8 is useful for describing the form of the foreshortening effect (that of a constant scale factor), but it would be useless in a stereo matcher since it requires knowledge of the depth Z_L. A program that computed depth given depth would not be very impressive. So how can we eliminate the need to know Z_L? Consider the ratio $\frac{B}{Z_L}$. We can rewrite Equation 5.7 as:

$$\frac{B}{Z_L} = \frac{disparity}{f - x_{iL}\tan\theta} \tag{5.9}$$

and replace that in Equation 5.8, giving us this final expression for the projected form of the Foreshortening Factor:

$$Projected\ form = 1 + \frac{disparity\tan\theta}{f - x_{iL}\tan\theta} \tag{5.10}$$

This is what we want! Equation 5.10 relates parameters in the image plane to the surface slope θ, but does not require prior knowledge of the distance to the object or an estimate of the disparity derivative. It does require use of some known parameters (focal length f, image location x_{iL}) and variables being estimated (disparity, surface angle θ), but we will see how to manage these algorithmically in Section 5.4.

In this section we described the effect of perspective foreshortening in terms of local spatial frequency. We developed this theory in steps to demonstrate several properties: the frequency shift (aka *foreshortening factor*) between images of an oriented flat surface is constant, it is independent of the surface texture, and it can be expressed using only disparity and surface angle (without disparity derivatives). Section 5.4 will show how these results can be applied to a stereo matching system.

5.2.1 Verifying the Foreshortening Factor

Before continuing, we will verify the geometric form of this Foreshortening Factor using a simple example: a flat surface with a sinusoidal texture. If the model is correct, the surface's apparent spatial frequencies will be shifted between the two images by the amount given in Equation 5.8. Note that we're not solving the stereo problem yet, in fact this demonstration will use the *known* disparity to compare the left and right *image* frequencies at the same *surface* locations. What this will show is that Equation 5.8 accurately predicts the frequency

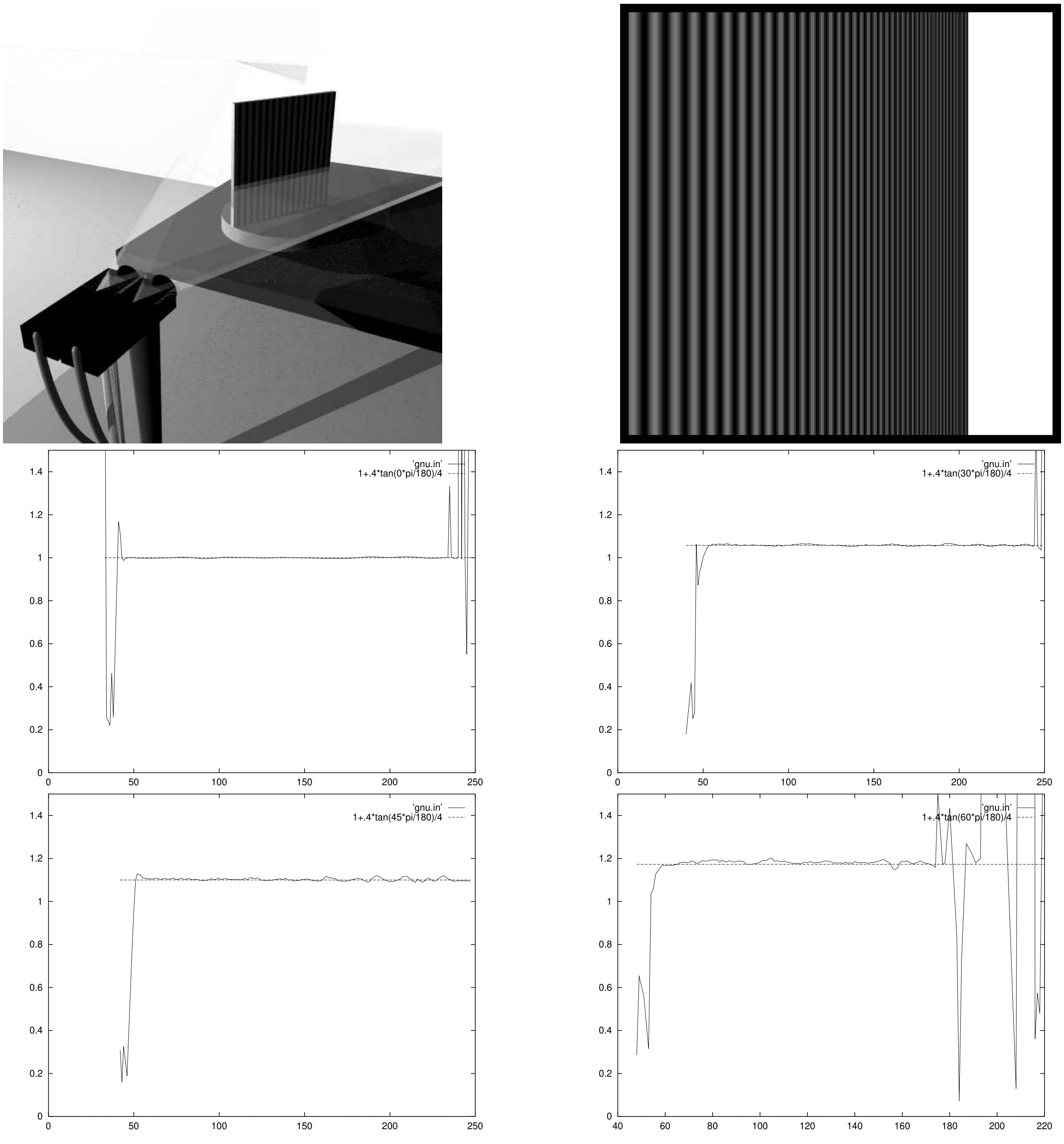


Figure 5.6: Verifying the Foreshortening Factor - These graphs compare the predicted foreshortening factor (dashed line) against that computed using only image information (solid line). The virtual lab setup (top left) and an example input image with surface angle of 60° (top right) are shown first. Next we have the results derived from a surface angled at 0° (middle left), 30° (middle right), 45° (bottom left), and 60° (bottom right). The virtual surface is 4.0 units from the left camera, both cameras have a field of view of 45° and are separated by a baseline of 0.4 (the surface in the actual images is larger than that shown in the top left rendering).

shift of a simple signal. We will use synthetic data so that our ground truth can be as precise as possible.

Recall the geometric form of the Foreshortening Factor from Equation 5.8:

$$\text{Foreshortening Factor} \quad = \quad 1 + \frac{B}{Z_L} \tan\theta$$

Just what is this foreshortening factor? It describes the relationship between the spatial frequencies at two image pixels representing the *same* surface point. How can we measure such frequencies, and how do we know they correspond to the same surface point?

Finding the frequency is easy, but imprecise; we will use an artificial surface texture that contains a single peak in the positive frequency domain, i.e., a sine wave. Its apparent frequency can be found simply by locating the filter output with highest magnitude.[1] As a further refinement, we will use the instantaneous frequency (phase derivative) of that filter output as our frequency estimate. Under the scalogram representation this corresponds to picking the maximum magnitude value in each column.

The procedure for finding corresponding points is somewhat complex, but simply stated involves using knowledge of the ground truth to give the disparity at each pixel (disparity is inversely related to depth, which is known from the 3D model). Remember, we are not trying to solve the stereo problem at this point, we are simply trying to verify a property of corresponding pixels.

Having established the correspondence in the 2D images, we extract the apparent frequency at each pixel using the method described above, linearly interpolating the instantaneous frequency measurements from the right image. Finally, we graph the ratio of the computed image frequencies values against the predicted ratio in Figure 5.6, for several surface angles. The computed ratio is quite accurate but gets progressively less precise as the angle increases. The loss of precision occurs from several factors, e.g., our use of simple linear interpolation to compute the frequencies, and our filter set which only samples the highest frequencies very sparsely.

[1] In practise our windowing scheme provides only high frequency info at the image borders, so our computed Foreshortening Factor will become inaccurate at the ends of the graph since the actual spatial frequency is lower than the lowest measured by the filters at that pixel.

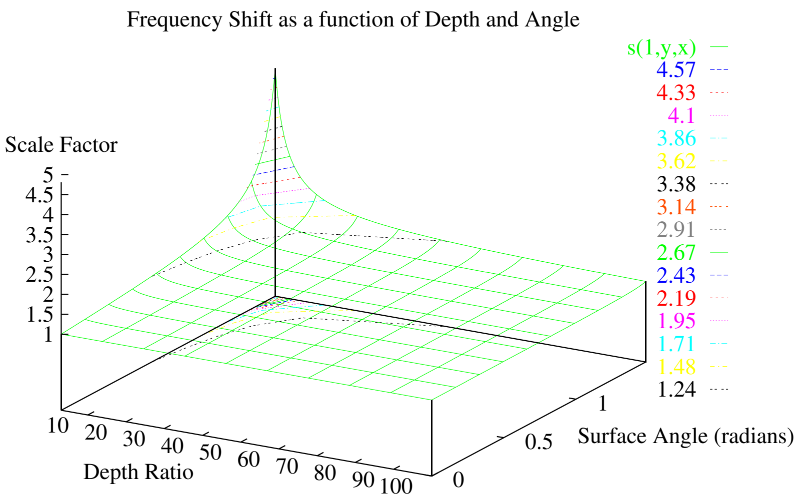


Figure 5.7: Foreshortening Factor as a function of Depth and Angle. Depth is unitless relative to the baseline, and varies from 3 to 100. Angle varies from zero to 85°.

5.3 Applicability

How important is this foreshortening analysis? More specifically, how often do situations arise in which the assumption that a surface is frontoplanar can cause problems for stereo systems? Intuitively the analysis would seem to be needed any time a surface is slanted at a sharp angle; but what if the surface is so far away the slant can't be measured? One might also think it only necessary for surfaces at the sharpest angles; but close up images can exaggerate even small angles. We will use the Foreshortening Factor to quantify these effects in the spatial domain.

Since we want to consider the scenery being imaged rather than the images themselves, we will use the geometric formulation of the Foreshortening Factor from Equation 5.8. Although this expression is a function of three variables, we can reduce it to two if we consider the ratio of depth over baseline $\frac{Z_L}{B}$ to be a single variable. In the rest of this chapter the word *depth* will denote this unitless version of depth, expressed relative to the camera baseline. For example, the distance between a person's eyes would be 1, the distance to their computer monitor 4-6, and the distance to the far wall in a typically small three-person graduate student office about 100. Figure 5.7 plots the near-complete Foreshortening Factor space for a person looking at objects in such an office.

Depth Range	P(≥10% effect)	Example Domain
0-100	0.210355	*Human in office*
5-20	0.354404	*Robot Vehicle*
30-100	0.0808227	*Inspection Robot*

Table 5.1: Probability that a surface exhibits ≥10% variation between images due to perspective foreshortening. The distribution of surfaces is assumed to be uniform within the range of orientation angles from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$, and depth ratios (distance divided by baseline) are as specified. A sample derivation can be found in Appendix B.

Figure 5.7 shows the Foreshortening Factor computed from many combinations of depth and orientation (except for the extreme values near the point at which it approaches infinity). The graph makes it clear that the Foreshortening Factor has its greatest impact when objects are sharply slanted and/or located near the cameras. We can quantify its influence using the contour lines that separate regions of large and smaller foreshortening effects. Suppose we assume that surface depth and orientation are uniformly distributed throughout a scene. Then we can compute the probability that a surface will require at least a 10% correction term by finding the area under the 1.1 Foreshortening Factor contour curve. The derivation follows in Appendix B, but the result is that given a uniform distribution of angles from 0° to 90° and depths from 0 to 100, the probability that a surface will require at least a 10% correction is 0.210355. Try it out; if you're sitting in an office, see if you can find one sharply foreshortened surface for each set of four nearly head-on surfaces in your immediate vicinity.

Of course the probability of finding foreshortened surfaces depends very much on the domain being studied. Robot vehicles like Carnegie Mellon's NAVLAB often use a very wide baseline, on the order of one meter. With the nearest visible ground point being about five meters away, depth ratios of 5 to 20 are common in this domain. In that range, under the same assumptions of uniform distribution, the probability of finding a foreshortened surface jumps to better than one in three (see Table 5.1). Inspection robots typically use much smaller baselines, with corresponding depth ratios from 30 to 100. Even in that range, the probability of finding a 10% foreshortened surface is significant (nearly one in twelve). These results suggest that a wide variety of stereo vision systems could benefit from an analysis that considers the effects of foreshortening.

5.4 Application

The analysis in Section 5 is not only theoretically interesting, it can also improve the performance of real stereo algorithms. Phase-based methods such as (Fleet et al., 1991; Sanger, 1988; Weng, 1993) as well as our method can benefit from this analysis. In this section we explain how to apply the Frequency Shift to these phase-based stereo matching algorithms and demonstrate how its application to our system increased the maximum matchable surface angle from 30 degrees to over 75 degrees.

5.4.1 Extending Phase-based Stereo Algorithms

Some have argued that a small number of Gabor filters are sufficient for stereo matching. (Fleet & Jepson, 1993; Weng, 1993) The idea is that although the phase may vary slightly across nearby frequencies, the amount of variation is small enough that the error introduced in measuring it at what might be the wrong frequency will be insignificant. But the assumption is made that the same filters can be applied to both images, i.e., that both images can be sparsely sampled at the same set of spatial frequencies. As was shown in the preceding section, that assumption is not true when perspective foreshortening occurs in the images; indeed, we have seen that frequency shifts of even 10% can occur often. Instead of introducing error by sampling at the wrong frequency, we would like to turn these perturbations to our advantage by using them to confirm hypotheses of surface slant.

We will need a dense sampling of the phase space to get the most accurate results. We will also interpolate phase values between adjacent frequencies when possible. The image scalogram provides a useful framework for such computations, and will be used as the basis for our foreshortening-corrected stereo algorithm.

The method outlined in Section 3.7 uses a global minimization strategy to find the best disparity from a list of candidates. This framework makes it easy to include a foreshortening correction term: in addition to searching disparity space, we also search over surface angle. Recall from Figure 5.2 that a foreshortened object will generate a response at *different* frequencies (i.e., scalogram rows) in the two images. Searching over surface angle allows us to predict the corresponding frequency directly. Pseudocode for this revised algorithm is given in Table 5.2. The only difference between this and the original algorithm is the presence of the correction term on the right image phase measurements. This simple presentation of the

Given: A pair of greyscale images, lists of potential
disparities and surface angles, focal length f.

For each row
  Compute Left and Right Scalograms L and R
  For each column c
    For each disparity d
      For each angle a
        $correction = 1 + \frac{d \tan a}{c \tan a - f}$
        $error = \sum_{\lambda : \rho(\lambda) > threshold} \rho_L(c, \lambda) \cdot$
        $|\Delta\phi_{ideal}(d, \lambda) - (\phi_L(c, \lambda) - \phi_R(c + d, \lambda \cdot correction))|_{2\pi}$
    Return d (and a) that yield minimum $error$

Table 5.2: Pseudocode for the foreshortening-corrected algorithm. Column index c must be zero in the center of the image.

algorithm is only made feasible because of the large number of filters used in the scalogram. The large filter set gives us a dense set of phases at many scales from which to compute the appropriate subsampled phases.

There are several implementation details that arise from this simple correction factor. It depends on three variables: the currently hypothesized disparity, surface angle, and the current location within the image. Because these values vary at each pixel on the image scanline, it must be recomputed for each hypothesis. And as was mentioned above, the corrected frequency will almost certainly not be one of those already present in the scalogram; some method of interpolation will be required. These are not serious problems, but imply that their implementation will be very compute-intensive.

5.4.2 Results

We added the correction term to the algorithm presented in Section 3.7 using linear interpolation between adjacent phases. In this section we present the results of our method on real images that have been synthetically mapped onto planar surfaces. The use of synthetic data allows us to quantify its precision using perfect knowledge of the ground truth.

Consider the stereo pair in Figure 1.3. It shows a synthetic stereo image pair of a flat plate rotated 65 degrees from the image planes, with the image of a city scene texture-mapped onto the plate. The actual disparity map (known from the 3D world model) and differences between the ground truth and disparity computed by three stereo methods are presented in Figure 5.8. The figure shows disparity maps rendered as perspective surfaces; only the area known to have texture is shown since the plain white background makes depth recovery impossible in those areas.

For this demonstration of the foreshortening-corrected algorithm, a set of 501 potential disparities were considered (0 to 50 in steps of 0.1), and the angle was fixed at 65 degrees. The RMS error of this result was 0.38 pixels over the entire plate, with $\sigma = 0.63$. The bulk of this error can be attributed to two causes: the dark spots and a subtle systematic error over the surface. The spots most likely arise from an artifact of the rendering process which caused a few nearby pixels in one image to map to the same intensity. The more subtle effect is that the disparity error, while within measurement bounds at the ends and center of the plate, varies by as much as 0.5 pixels between the center and end of the plate (see Figure 5.8, upper right).

The Kanade-Okutomi variable-window refinement method (Kanade & Okutomi, 1990) uses a statistical analysis to grow the window from 3x3 to some maximum, stopping when an error criterion (based on local changes in intensity and disparity) is exceeded. For this test we let disparity vary between 0 and 50 pixels (as in our method), let the window size vary from 3 to 21 pixels, and ran the method for 10 iterations. It approximated the surface shape well, but produced many more outliers and quantized the flat tilted surface into several stair-step frontoplanar patches (see Figure 5.8, upper right). The RMS error of this method was 0.99 pixels over the entire plate, with $\sigma = 2.36$.

The uncorrected phase method results are also shown in Figure 5.8. The same 501 potential disparities were considered, but foreshortening correction was not applied. The RMS error of this result was 3.77 pixels over the plate, with $\sigma = 6.23$. The main source of error is a general flattening trend over the entire plate, most likely due to the larger windows used at lower frequencies. Like most traditional stereo matchers, the uncorrected method has a strong bias toward frontoplanar surfaces, but unlike Kanade/Okutomi this uncorrected phase method is unable to restrict its attention to the smallest-sized windows.

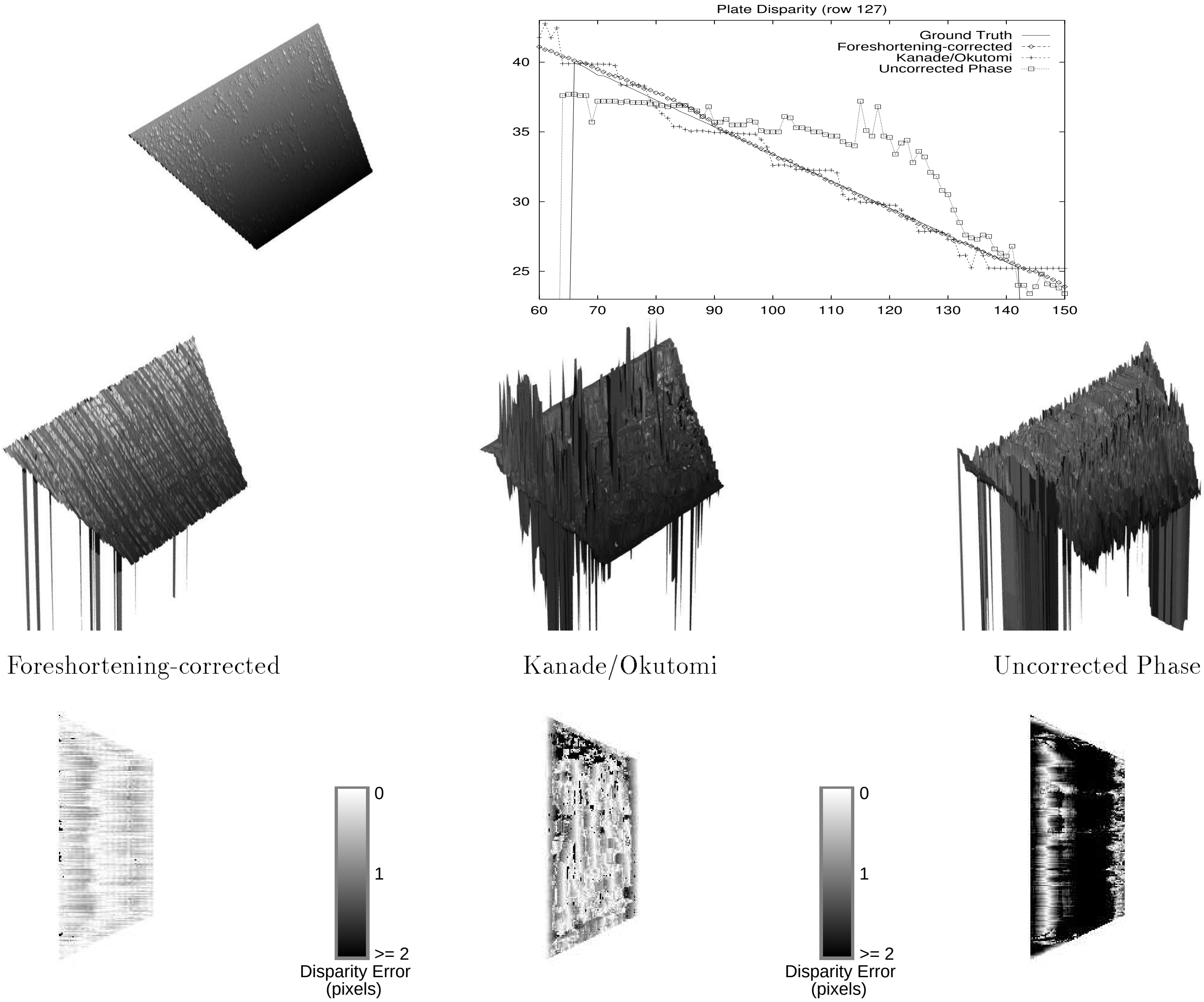


Figure 5.8: Ground Truth and computed disparity maps for a surface angled at 65°. The top row shows ground truth in perspective on the left, a graph of a representative scanline from all methods on the right. The middle row shows perspective views of the disparity maps computed by the foreshortening-corrected method, Kanade/Okutomi and the uncorrected phase method. The bottom row shows differences between actual disparities and those computed by the foreshortening-corrected method, Kanade/Okutomi and the uncorrected phase method, for pixels that image the plate; darker values denote larger errors. Only differences between 0 and 2 pixels are shown, errors larger than 2 pixels appear as a 2 pixel error. Actual plate disparities range from 25.3 to 39.9 pixels.

Other Rotation Angles A cross-section of results for different angles of rotation is presented in Figure 5.9. For these results only a representative scanline is shown, to demonstrate how closely the computed disparity matches the actual ground truth. Only the disparities on the plate itself are correct because the region behind it is a plain white background, and there is no way to distinguish the correct disparity of a featureless surface.

The uncorrected method does reasonably well with small angles, but at slants greater than 30° its performance degrades by several pixels. In contrast, the foreshortening-corrected method performs well even at 75°, though at 80° the systematic error becomes more apparent.

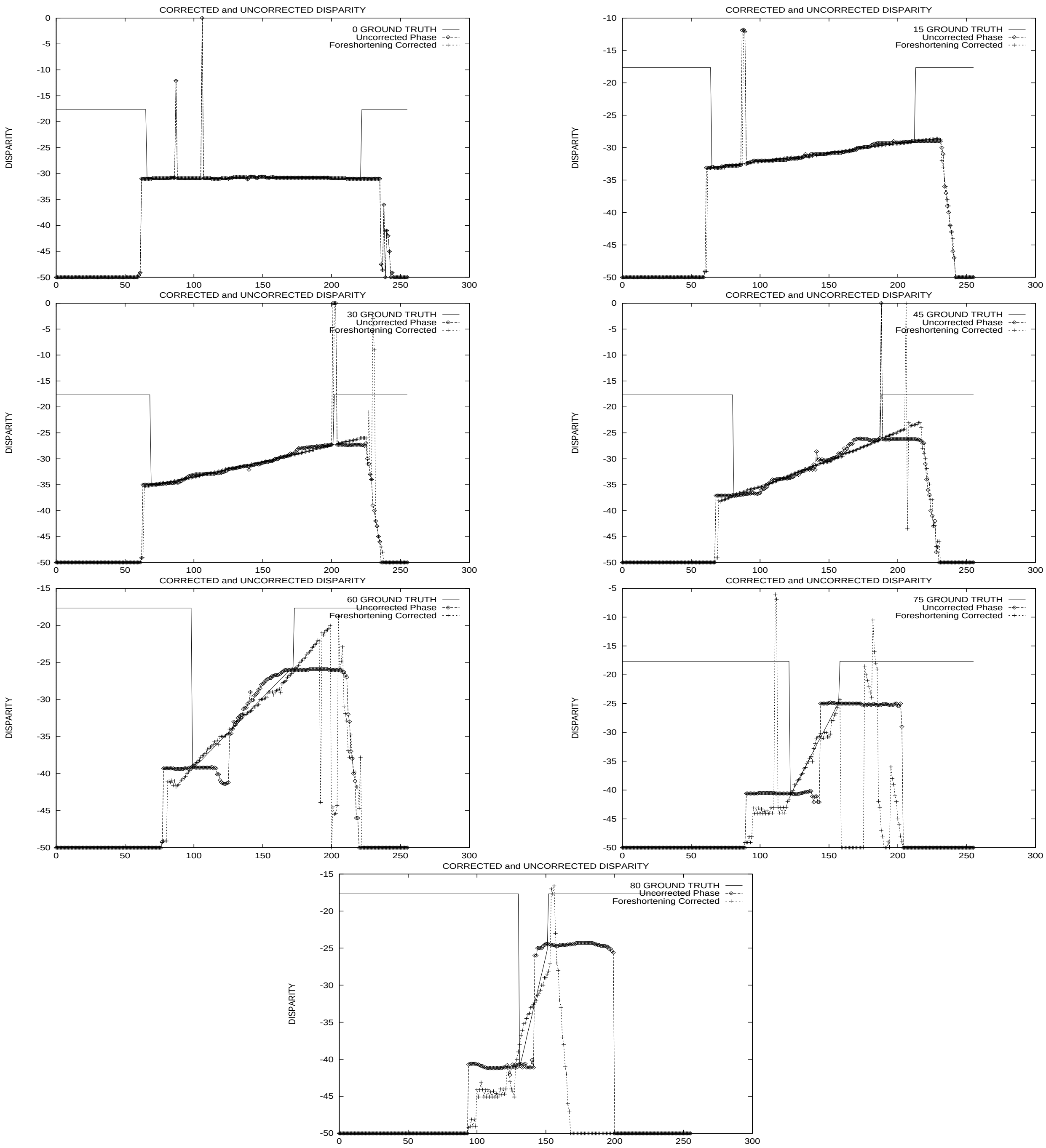


Figure 5.9: Ground truth and disparity (computed by both the uncorrected and foreshortening-corrected phase methods) for the center scanline of the city scene at various rotations. From left to right (and top to bottom): 0, 15, 30, 45, 60, 75, and 80 degrees.

Chapter 6

Contributions

And at the laste he took conclusioun.

— Chaucer, *the Knight's Tale*

This thesis has addressed several long-standing problems in stereo vision: perspective foreshortening, ambiguous matches, and the quantitative evaluation of stereo results. We also demonstrate by example the utility of the local spatial frequency representation in the context of stereo vision. Some particular contributions include:

Perspective Foreshortening We addressed the long-standing problem of perspective foreshortening in stereo vision. Ours is the first work to provide an analytical closed-form expression for the effect of perspective foreshortening on stereo matching in the frequency domain, and demonstrate that the model improves the results of a real stereo method. Results relating foreshortening to the frequency domain have appeared in the shape-from-texture literature (Krumm, 1993; Malik & Perona, 1989), but without a description of the use of disparity; instead, the mapping between two areas was left as a general affine matrix. Similarly, (Jones & Malik, 1991) uses an affine matrix to relate two corresponding image patches in the context of stereo, but does not provide the analytic relationship. Ours is the first presentation to unify these three themes (stereo, foreshortening, frequency domain) and the first to demonstrate good matching even at slants up to 75 degrees (the spatial-domain system of (Belhumeur, 1993) might perform as well, but this has not been demonstrated).

Adaptive Scale Selection We demonstrated the improvement of our scale-adaptive algorithm over traditional multiscale (e.g., coarse to fine) algorithms. Ours is the first multiscale phase-based method that is not confused by missing information at intermediate scales. Most of the prior stereo systems work either at a single scale or using a coarse-to-tine approach in a fixed order. Ours is the first system to invoke automatic scale selection (via filter magnitude weighting) in a nonrestrictive manner. (Sanger, 1988) also used magnitude weighting, but imposed a restrictive limit on candidate disparities. (Jones, 1991) also used some scale-space processing, but only to eliminate the coarsest scales at depth discontinuities. Ours is the first system that can handle missing information at *any* scales.

Phase Wraparound We eliminated the restriction of previous phase-based stereo algorithms on the maximum disparity range, and described a new stereo algorithm that eliminates the problem of phase wraparound. Ours is the first phase-based stereo method to overcome these limitations.

New Disparity Model We proposed a new error model for disparity that more accurately represents the inaccuracies that result from ambiguous matches. This new model characterizes the results of stereo processing more appropriately than the traditional two parameter model (disparity value with variance).

Datasets We proposed a new taxonomy for stereo vision experiments, and provided some of the first public datasets with piecewise dense ground truth. These datasets and the easy-to-use tools for creating them should encourage the community to use (and demand) more quantitative evaluations of stereo systems.

To summarize, our method uses an adaptive search through scale space. We combine estimates from the most reliable scales in a framework that can be used to evaluate the likelihood of *arbitrary* disparities at each pixel; we are not limited by the wavelength of any single filter, as are all previous phase-based methods. The method eliminates the need to perform explicit phase unwrapping, thereby improving accuracy, but at the cost of an additional search over candidate disparities.

Potential future extensions to this work include:

1. Addition of a precision parameter (variance) to the representation of ground truth in Chapter 2 to account for measurement errors and variation within a pixel.

2. Development of a synthetic data generator with better lighting models (e.g., using radiosity rather than ray tracing).

3. Addition of sequences of stereo images to the taxonomy in Chapter 2.

4. Experimentation with the shape of the evaluation function (as in Equation 3.14); a replacement for AbsDiffMod such as a cosine might yield smoother results, or be faster to implement.

5. Speed up the processing by incorporating the fast wavelet transform with appropriate interpolation in place of the complete scalogram computation, using a smaller set of filters, and a smaller set of foreshortening angle candidates.

6. Extend the occlusion and disparity models to account for multiple depths within a pixel.

7. Explore the use of nonsymmetric filters (in place of the Gaussian envelope used by Gabor filters) to better address the disparity spillover that occurs at depth discontinuities.

8. Develop search strategies over the binary "peaks-only" images of Chapter 4 to merge pixel disparities into potential surfaces.

Appendix A

Theodolite Error Analysis

Both stereo disparity and ground truth measurements have finite precision which should be made explicit. As a first step toward extending our notion of ground truth to include this precision, we present in this section an analysis of the resolution obtainable using surveyor's theodolites in their present configuration as part of the Calibrated Imaging Laboratory (CIL).

The Calibrated Imaging Laboratory theodolites (Sokkisha, 1984) can repeatably measure angles to within about 20 seconds of arc. That is, during a single test run, an individual can repeatedly aim the theodolite site at a target, unlock it, then aim again and be confident that the difference between successive measurements will never be more than 20 seconds. This is in spite of the fact that the instrument readout is apparently measured to the nearest tens of seconds (*tens* not tenths).

We would like to know how accurate subsequent **X-Y-Z** computations can be, under this limitation. A simple two-dimensional (**X-Z**) analysis will give us a rough idea of the magnitude of the precision in the horizontal plane. Figure A.1 shows the overall model: depending on the angles measured, the computed depth D might lie anywhere within the shaded region. Since that region is polygonal, we know that the largest possible error (i.e., the maximum distance between any two points in the region) will occur between the endpoints of one of its two diagonals. Just how long are the diagonals? To determine that, we need to derive equations for the horizontal and depth coordinates. We address the horizontal first.

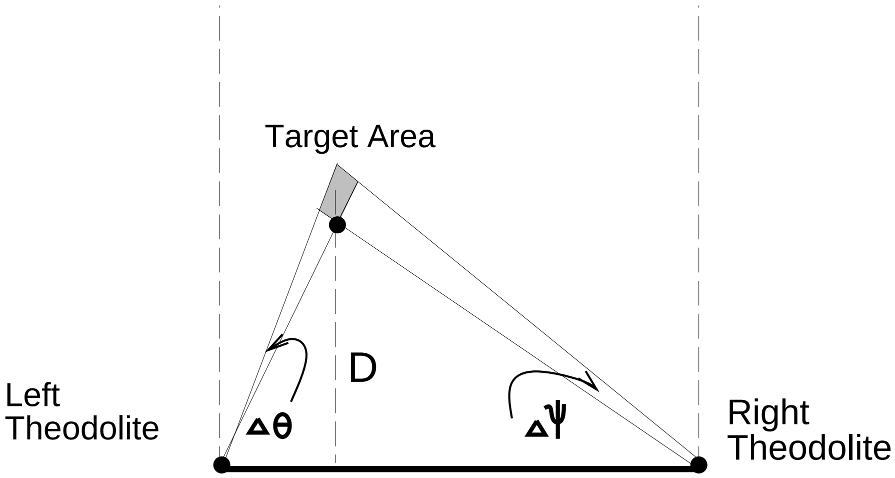


Figure A.1: Region of error. The greatest possible error occurs across one of the Target Area diagonals (see Figure A.3 for a close up).

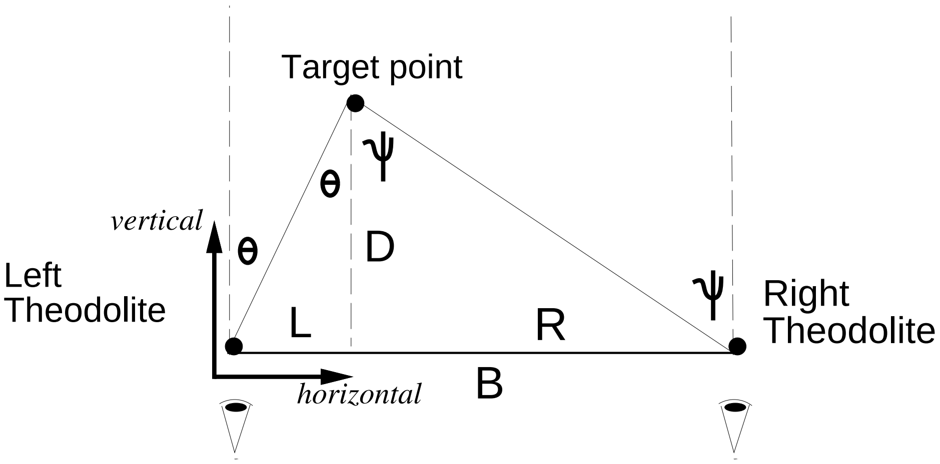


Figure A.2: Two dimensional view of the theodolite imaging process.

A.1 Deriving equations for the coordinate axes

Figure A.2 shows the geometry of the scene. We will treat the two theodolites (as well as the target) as points. θ is the angle measured by the left theodolite, ψ is that measured by the right. B is the length of the baseline between the two theodolites; the baseline is split in two at the projection of the target point: $B = L + R$. D is the distance from the baseline to the target point. If we define the left theodolite to be the origin of a coordinate system with horizontal axis along the baseline, we have D as the vertical *depth* coordinate, and L

as the horizontal coordinate.

Now we need to express D and L as functions of the two angles and baseline alone. The left and right angles bear a simple relationship to the two baseline parts:

$$\tan\theta = \frac{L}{D} \qquad \textit{and} \qquad \tan\psi = \frac{R}{D}$$

Solving for D and setting them equal, we have:

$$\frac{L}{R} = \frac{\tan(\theta)}{\tan(\psi)}$$

Recalling that L and R sum to the baseline B, we have by substitution:

$$L = \frac{B}{1 + \frac{\tan(\psi)}{\tan(\theta)}} \qquad \textit{and} \qquad R = \frac{B}{1 + \frac{\tan(\theta)}{\tan(\psi)}} \tag{A.1}$$

This solves our horizontal coordinate problem: we have an expression for L that depends only on angles θ, ψ and the baseline. Moving on, these equations give us two expressions for the distance D between the baseline and target point:

$$D = \frac{L}{\tan\theta} = \frac{R}{\tan\psi}$$

both of which reduce to the same symmetric result:

$$D = \frac{B}{\tan\theta + \tan\psi} \tag{A.2}$$

So now we have the depth coordinate D as well. Without loss of generality, we assume the baseline is a constant factor and write $D = d(\theta, \psi)$ and $L = l(\theta, \psi)$. Now we can compute the lengths of the two diagonals in the Target Area; the larger one will give us the maximum possible error.

To compute the lengths of the diagonals h and v (shown in Figure A.3), we find the Euclidean distance between their endpoints. Call the theodolite measurement error δ: for the CIL theodolites δ is $20''$. Then the length of the "horizontal" diagonal h (it's not really horizontal) is:

$$h(\theta, \psi) = \sqrt{(d(\theta, \psi - \delta) - d(\theta - \delta, \psi))^2 + (l(\theta, \psi - \delta) - l(\theta - \delta, \psi))^2} \tag{A.3}$$

The vertical diagonal v (it's not really vertical) is computed in the same way:

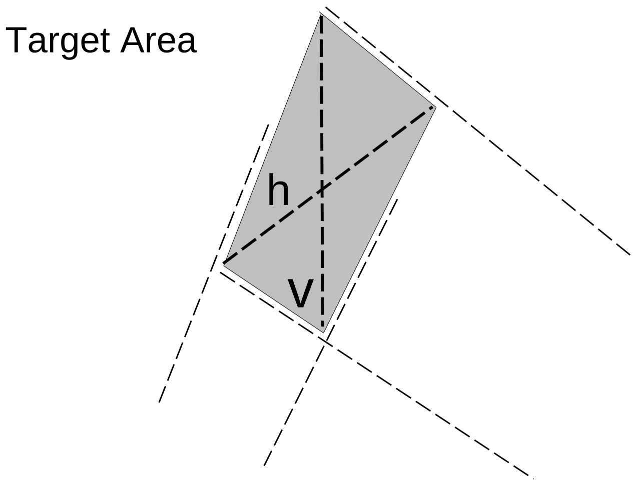


Figure A.3: Target Area Precision (zoom in on Figure A.1): the largest error is the length of one of the diagonals h and v.

$$v(\theta,\psi) = \sqrt{(d(\theta-\delta,\psi-\delta)-d(\theta,\psi))^2+(l(\theta-\delta,\psi-\delta)-l(\theta,\psi))^2} \tag{A.4}$$

Now we're done; the length of the horizontal diagonal is the maximum error in the horizontal direction, the vertical diagonal is the maximum error in depth. All that remains is to plug in the measurement error $\delta = 20''$.

A.2 Results

In the current laboratory configuration, the theodolite measurements will vary from 10 to 70 degrees, if all objects of interest lie on or above the object optical table. Figure A.4 shows the shape of the error surface for this configuration, with $\delta = 20''$ and assuming a unit baseline. How do we interpret this?

The largest error in Figure A.4 is about 0.0015, when both theodolites have angles of 10°. What does this really mean? Assuming a baseline of 86.1 inches (219cm), it means an object 6.21m away can only by measured to within 3.3mm.[1] However, the far end of the optical table is only about 3m away from the baseline, and according to the model the center point on the far edge gives angles of about 20° for each theodolite. The error for those angles is 0.0004, which means the best precision for the far end of the table is 0.876mm= 0.0004·219cm. Thus

[1] 6.21m= $\frac{2.19\text{m}}{\tan 10° + \tan 10°}$ and 3.3mm= 219cm·0.0015

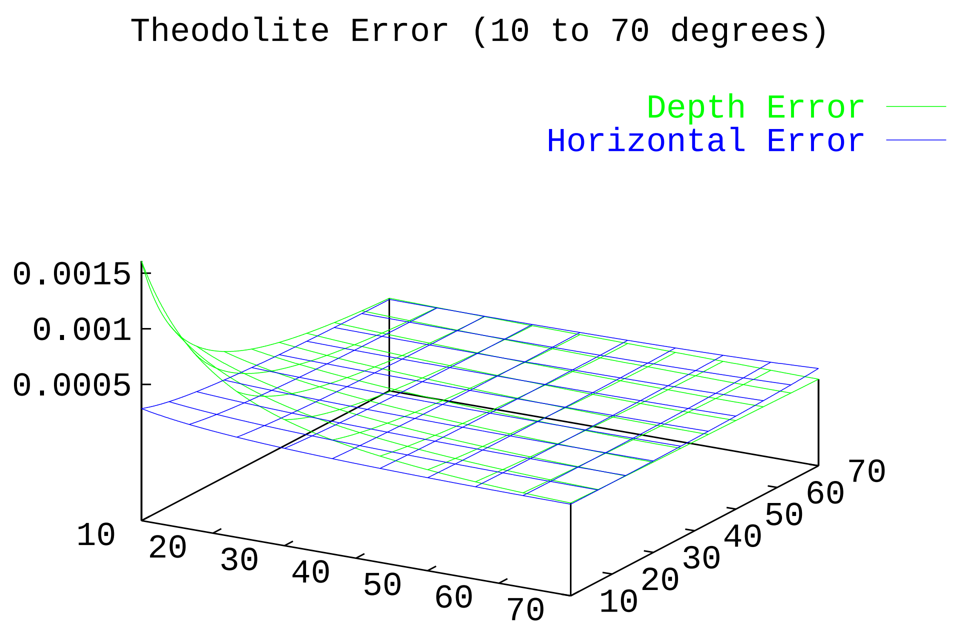


Figure A.4: Error Space for $\delta = 20''$ with unit baseline: depth error $v(\theta, \psi)$ is greater than horizontal error $h(\theta, \psi)$ for all but the nearest points (when the sum of the angles is greater than about 90 degrees)

it's safe to say our theodolite measurements over the optical table are in general accurate to within a millimeter.

Nearer measurements have better accuracy. For example, in the **CIL-0001** Stereo Dataset the left horizontal angles range from 41 to 50 degrees, and right angles from 31 to 42 degrees. The maximum error in that range is $0.296mm = 0.000135 \cdot 219cm$.

It is in fact possible to compute the precision at each point. Simply plug the angles measured into the error terms defined above. Or to estimate the precision for a whole region, run the code in Table A.1 through the GNUplot package with appropriate limits (instead of $10°$ to $70°$), and visually pick out the largest error.

Limitations: The acquisition of ground truth requires that static objects be imaged in a laboratory environment. However there are many applications for which dynamic imagery is required, and stereo systems must be tested using comparable data. The acquisition of such data implies that no ground truth will be available.

```
d2r = pi / 180.0
hor(x,y) = 1/(1+tan(x * d2r)/tan(y * d2r))
ver(x,y) = 1/(tan(x * d2r) + tan(y * d2r))
d = 20.0/3600
splot [10:70] [10:70] \
      sqrt((hor(x-d,y-d)-hor(x,y))**2 + \
           (ver(x-d,y-d)-ver(x,y))**2) title "Depth Error", \
      sqrt((hor(x-d,y)-hor(x,y-d))**2 + \
           (ver(x-d,y)-ver(x,y-d))**2) title "Horizontal Error"
```

Table A.1: GNUplot commands that generated Figure A.4

Appendix B

Derivation of Foreshortening Probabilities

We present here the derivation of one of the probabilities from Table 5.1. Figure 5.7 showed the Foreshortening Factor computed from many combinations of depth and orientation (except for the extreme values near the point at which it approaches infinity). The graph makes it clear that the Foreshortening Factor has its greatest impact when objects are sharply slanted and/or located near the cameras. We can quantify its influence using the contour lines that separate regions of large and smaller foreshortening effects. Suppose we assume that surface depth and orientation are uniformly distributed throughout a scene. Then we can compute the probability that a surface will require at least a 10% correction term by finding the area under the 1.1 Foreshortening Factor contour curve. This derivation assumes the depth range begins at zero (the more general results require a little more work).

We want to find:

$$P(\text{Foreshortening Factor} \geq 1.1 \text{ or} \leq 0.9) = P\left(\left|\frac{\tan\theta}{d}\right| \geq 0.1\right)$$

for $d = \frac{Z_L}{B} \in [0:100]$ and $\theta \in [-\frac{\pi}{2} : \frac{\pi}{2}]$, each uniformly distributed. Since tan is symmetric we can eliminate the absolute value by restricting the angle θ to $[0 : \frac{\pi}{2}]$. Continuing:

$$\begin{aligned} P\left(\left|\frac{\tan\theta}{d}\right| \geq 0.1\right) &= P\left(\frac{\tan\theta}{d} \geq 0.1\right) = \frac{\int_0^{\frac{\pi}{2}} \min\left(\frac{\tan\theta}{\text{Foreshortening Factor}-1}, 100\right) d\theta}{\int_0^{\frac{\pi}{2}}\int_0^{100} dd\, d\theta} \\ &= \frac{\int_0^{\frac{\pi}{2}} \min\left(\frac{\tan\theta}{0.1}, 100\right) d\theta}{50\pi} \end{aligned}$$

To eliminate the min from the integral we must find the minimum angle requiring a 10% correction at distance 100:

$$\theta_{\mathrm{min}} = \arctan 100(\text{Foreshortening Factor} - 1) = \arctan 10 = 84.2894^\circ$$

Now we can split up the integral into two parts and evaluate it:

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \min\left(\frac{\tan\theta}{0.1}, 100\right) d\theta &= \int_0^{\theta_{\mathrm{min}}} \frac{\tan\theta}{0.1} d\theta + \int_{\theta_{\mathrm{min}}}^{\frac{\pi}{2}} 100 d\theta \\
&= \left[\frac{\ln\sec\theta}{0.1}\right]_0^{\theta_{\mathrm{min}}} + 9.96688 \\
&= 23.0756 + 9.96688
\end{aligned}$$

This brings us to the final result:

$$P\left(\left|\frac{\tan\theta}{d}\right| \geq 0.1\right) = \frac{33.0425}{50\pi} = 0.210355$$

So under the assumption of uniform distribution on depth ratio from 0 to 100 and angle from -90° to 90°, the probability of a surface exhibiting at least a 10% foreshortening effect is 0.210355.

Appendix C

Numbers

Table C.1 shows the default parameter values used to generate the images in this thesis. Additional details can be found in the Matlab source code used to generate many of the actual images, especially those in Chapters 3 and 4. This code is available on the web from *Mark Maimone's Index Page.*[1]

[1] `http://www.cs.cmu.edu/~mwm/`

Parameter	Value	Comments
Disparity Range d	$0, 1, \ldots, 50$	The range of candidate disparities can usually be inferred by figure axes, and the step size by the visible quantization in disparity space images. Often the step size will be less than 1, which indicates subpixel precision.
Wavelengths W	$2, 3, \ldots, n/4$	Unless otherwise stated, we typically use the image scalogram sampling which is linear in wavelength.
Wavelengths per window m	4	The Gaussian envelope of a Gabor filter will be truncated outside this many wavelengths. This determines the window size.
Sigma fraction σ_f	1/6	Given a fixed window size (m), the σ parameter of the Gaussian will be this fraction of the window.
Fleet threshold	.05	Cutoff for Equation 3.10.

Table C.1: Default parameter values used to generate images in this thesis.

Appendix D

Application of Visual Reconstruction

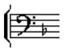
XXXX XXXXXXX XXXXX
Pittsburgh, PA XXXXX
(412) XXX - XXXX
work: (412) 268 - 7698
email: mwm@cs.cmu.edu
October 3, 1995

Prof. Peter Schickele
XXXXX XXXXXXXXXXX / NY Campus of U of SND at H
XXX XXXX XXXX XXXXXX
New York, NY XXXXX

Dear Colleague:

Greetings! I am writing to inform you of a most remarkable discovery: the visual appearance of Herr P.D.Q. Bach's headstone within the mausoleum on the outskirts of Baden-Baden-Baden. Although said mausoleum was destroyed in the 1840s ([**1**] pp. 112-113), computer analysis of the data collected by your esteemed research team ([**1**] pp. 131-135) has made possible visualization of this artifact whose historical significance cannot be underestimated.

As you are no doubt aware, my own area of expertise lies in techniques for reconstruction of three-dimensional shapes from oblique imagery [**2**]. Application of these techniques to those passages of text referring to Herr Bach's exploits at Madame Höllender's ([**1**] pp. 100-101) resulted in images of a variety of organic experience [**3**] unsuitable for publication in any *reputable* medium of scientific discourse. I remained *firmly committed* to the project, however (anybody who says I'm not is a liar), and following a, hunch, applied the principles of super-resolution [**4**] to your original images graciously made available to me by a random house-keeper for a remarkably small fee.

I shan't bore you with the details of the reconstruction, but merely outline the process. Having accepted our dough, a dear (a female dear) house-keeper started us off with the data. Following the principles of laser holography [**5**], a specially-modified computer filtered the light ray (a drop of gold on Sun interface was required) passed by me through a Headpiece to the Staff of Fa ([**6**] ) over a distance of some kilometers, a long long way to run. So like a needle pulling thread[1], the beam was fed through the IID/NM T-junction [**7**] while I retired for a drink with jam and bread (that will bring us back to dough). Having scaled these obstacles – naturally filtering out any non-diatons we found floating around – we achieved our final result, an image of which is enclosed herewith.

Should you find yourself with access to the World Wide Web I encourage you to examine:

`http://www.cs.cmu.edu/~mwm/pdq/smix.html`

for background.

Anxiously awaiting your reply I remain...

Your Honor's and my most especially Highly
Honored Sir's most obedient servant,

Mark W. Maimone

[1] La

References

[1] P. P. Schickele. *The Definitive Biography of P.D.Q. Bach.* Random House, 1976.

[2] M. W. Maimone and S. A. Shafer. Modeling foreshortening in stereo vision using local spatial frequency. In *Intelligent Robotics and Systems*, pages 519–524, 1995.

[3] P. Schickele. Schickele Mix no. 40. Public Radio International (formerly APR).

[4] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson. Super-resolved surface reconstruction from multiple images. Technical Report FIA-94-12, NASA Ames AI Research, December 1994.

[5] S. A. Benton. Photographic holography. In *SPIE Proc. Optics in Entertainment*, volume 391, pages 2–9, January 1983.

[6] I. Jones, M. Brody, and Sallah. Scaling the tannis map room. [recently declassified Top Secret Army Intelligence Report 9906753], 1936.

[7] D. Adams. *The Restaurant at the End of the Universe*, pages 11–13. Random House, 1980.


PDQ
Bach

Bibliography

Bani-Hashemi, A. (1993). A Fourier Approach to Camera Orientation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1197–1202.

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

Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77. `ftp://csd.uwo.ca/pub/vision`.

Belhumeur, P. N. (1993). A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In *International Conference on Computer Vision*, pages 431–438.

Belhumeur, P. N. (1995). A bayesian approach to binocular stereopsis. To appear in IJCV.

Besl, P. J. (1988). Active, optical range imaging sensors. *Machine Vision and Applications*, 1:127–152.

Bhat, D. N. & Nayar, S. K. (1995). Stereo in the presence of specular reflection. In *International Conference on Computer Vision*, pages 1086–1092.

Bolles, R. C., Baker, H. H., & Hannah, M. J. (1993). The JISCT Stereo Evaluation. In *ARPA Image Understanding Workshop*, pages 263–274.

Boufama, B. & Mohr, R. (1995). Epipole and fundamental matrix estimation using virtual parallax. In *International Conference on Computer Vision*, pages 1030–1036.

Boyer, E. & Berger, M. O. (1995). 3D Surface Recostruction Using Occluding Contours. In *International Conference on Computer Analysis of Images and Patterns.*

Brown, L. G. (1992). A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376.

Carson, J. R. & Fry, T. C. (1937). Variable-frequency electric circuit theory. *Bell System Technical Journal*, 16:513–540.

Dhond, U. R. & Aggarwal, J. K. (1989). Structure from stereo — a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1489–1510.

Faugeras, O. D. (1993). *Three-Dimensional Computer Vision*. MIT Press.

Faugeras, O. D. & Toscani, G. (1986). The calibration problem for stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 15–20.

Fleet, D. & Jepson, A. (1993). Stability of phase information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1253–1268.

Fleet, D. J., Jepson, A. D., & Jenkin, M. R. M. (1991). Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210.

Gabor, D. (1946). Theory of communication. *Journal of the Institution of Electrical Engineers*, pages 429–457.

Ghiglia, D. C. & Romero, L. A. (1994). Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *Journal of the Optical Society of America A: Optics, ..., Vision*, 11(1):107–117.

Grant, C. W. (1992). *Visibility Algorithms in Image Synthesis*. PhD thesis, University of California Davis Computer Science.

Gülch, E. (1991). Results of test on image matching of ISPRS WG III/4. *ISPRS Journal of Photogrammetry and Remote Sensing*, 46:1–18.

Haines, E. (1993). Free ray tracer summary. *Ray Tracing News*, 6(3):article 4. `ftp://ftp-graphics.stanford.edu/pub/Graphics/RTNews/html/rtnv6n3.html#art4`.

Hartley, R. I. (1994). Self-calibration from multiple views with a rotating camera. In *European Conference on Computer Vision*, pages 471–478.

Hebert, M. & Krotkov, E. (1992). 3D measurements from imaging laser radars: How good are they? *Intl. Journal of Image and Vision Computing*, 10(3):170–178.

Hlawatsch, F. & Boudreaux-Bartels, G. F. (1992). Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, pages 21–67.

Horn, B. K. (1986). *Robot Vision*. MIT Press.

Horner, J. L. & Gianino, P. D. (1984). Phase-only matched filtering. *Applied Optics*, 23(6):812–816.

Jenkin, M. R. M. & Jepson, A. D. (1994). Recovering local surface structure through local phase difference measurements. *Computer Vision, Graphics and Image Processing: Image Understanding*, 59(1):72–93.

Jones, D. G. (1991). *Computational Models of Binocular Vision*. PhD thesis, Stanford University.

Jones, D. G. & Malik, J. (1991). Determining three-dimensional shape from orientation and spatial frequency disparities part ii - using corresponding image patches. Technical Report UCB/CSD 91/657, University of California Berkeley Computer Science Department.

Kanade, T., Kano, H., Kimura, S., Yoshida, A., & Oda, K. (1995). Development of a video-rate stereo machine. In *International Robotics and Systems Conference (IROS)*, volume 3, pages 95–100.

Kanade, T. & Okutomi, M. (1990). A stereo matching algorithm with an adaptive window: Theory and experiment. In *DARPA Image Understanding Workshop Proceedings*, pages 383–398.

Kanade, T. & Okutomi, M. (1991). A stereo matching algorithm with an adaptive window: Theory and experiment. In *Intl. Conference on Robotics and Automation*, pages 1088–1095.

Kanade, T., Okutomi, M., & Nakahara, T. (1992). A multiple-baseline stereo method. In *ARPA Image Understanding Workshop*, pages 409–426.

Krumm, J. (1993). *Shape from Texture and Segmentation using Local Spatial Frequency.* PhD thesis, Carnegie Mellon Robotics Institute.

Kuglin, C. D. & Hines, D. C. (1975). The phase correlation image alignment method. In *Proceedings of the IEEE Int. Conference on Cybernetics and Society*, pages 163–165.

Maimone, M. W. (1995). Watch the Birdie: A Guide to Imaging in the Calibrated Imaging Laboratory. Unpublished CMU Calibrated Imaging Laboratory (CIL) Tech Report.

Maimone, M. W. & Shafer, S. A. (1995a). Modeling foreshortening in stereo vision using local spatial frequency. Technical Report CMU-CS-95-104, Carnegie Mellon University Computer Science Department.

Maimone, M. W. & Shafer, S. A. (1995b). Modeling foreshortening in stereo vision using local spatial frequency. In *International Robotics and Systems Conference (IROS)*, pages 519–524. IEEE Computer Society Press. `http://www.ius.cs.cmu.edu/project/cil/fore/tr.html`.

Maimone, M. W. & Shafer, S. A. (1996). A taxonomy for stereo computer vision experiments. ECCV'96 *Workshop on Performance Characteristics of Vision Algorithms* `http://www.ius.cs.cmu.edu/project/cil/tax/`.

Malik, J. & Perona, P. (1989). A computational model of texture segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 326–332.

Matthies, L. (1989). *Dynamic Stereo Vision.* PhD thesis, Carnegie Mellon University Computer Science Department.

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

Mori, K.-I., Kidode, M., & Asada, H. (1973). An iterative prediction and correction method for automatic stereocomparison. *Computer Graphics and Image Processing*, 2(3/4):393–401.

Nakamura, Y., Matsuura, T., Satoh, K., & Ohta, Y. (1996). Occlusion detectable stereo — occlusion patterns in camera matrix. In *IEEE Conference on Computer Vision and Pattern Recognition.*

Nayar, S. K., Watanabe, M., & Noguchi, M. (1995). Real-time focus range sensor. In *International Conference on Computer Vision*, pages 995–1001.

Nishihara, H. K. (1984). Prism: A practical real-time imaging stereo matcher. Technical Report AI Memo 780, Massachusetts Institute of Technology AI Laboratory.

Okutomi, M. & Kanade, T. (1991). A multiple-baseline stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 63–69.

Oppenheim, A. V. & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541.

Photometrics (1990). *Charge-Coupled Devices for Quantitative Electronic Imaging.* Photometrics, Ltd., 3440 E. Britannia Drive, Tucson, AZ 85706.

Reeves, W. T., Salesin, D. H., & Cook, R. L. (1987). Rendering antialiased shadows with depth maps. In *ACM Computer Graphics SIGGRAPH*, pages 283–291.

Rioul, O. & Vetterli, M. (1991). Wavelets and signal processing. *IEEE Signal Processing Magazine*, pages 14–38.

Robert, L., Buffa, M., & Hebert, M. (1994). Weakly-calibrated stereo perception for rover navigation. In *ARPA Image Understanding Workshop.*

Ross, B. (1993). A practical stereo vision system. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 148–153.

Sanger, T. D. (1988). Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59:405–418.

Shaw, M. (1990). Prospects for an engineering discipline of software. Technical Report CMU-CS-90-165, Carnegie Mellon University Computer Science Department. Also IEEE Software, Nov 1990.

Shum, H.-Y., Hebert, M., Ikeuchi, K., & Reddy, R. (1995). An integral approach to free-form object modeling. Technical Report CMU-CS-95-135, Carnegie Mellon University Computer Science Department.

Smith, P. W. & Nandhakumar, N. (1996). An improved power cepstrum based stereo correspondence method for textured scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):338–348.

Sokkisha (1984). *Electronic Digital Theodolite DT20E Operation Manual*. Sokkisha Co., Ltd., Keio Yoyogi Building 5th Floor, No. 1, 1, 1-chome, Tomigaya, Shibuta-ku, Tokyo, 151 Japan.

Stein, G. P. (1995). Accurate internal camera calibration using rotation, with analysis of sources of error. In *International Conference on Computer Vision*, pages 230–236.

Stevenson, D. E. & Fleck, M. M. (1995). Robot aerobics: Four easy steps to a more flexible calibration. In *International Conference on Computer Vision*, pages 34–39.

Szeliski, R. (1994). Image mosaicing for tele-reality applications. Technical Report 94/2, DEC Cambridge Research Lab.

Tada, S., Gruss, A., & Kanade, T. (1993). CMU very fast range-imaging system. Technical Report CMU-CS-93-179, Carnegie Mellon University Computer Science Department.

Tribolet, J. (1977). A new phase unwrapping algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(2):170–177.

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344. `http://www.cs.cmu.edu/~rgw/TsaiCode.html`.

Wang, Z. & Jepson, A. (1994). A new closed-form solution for absolute orientation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 129–134.

Webb, J. (1993). Implementation and performance of fast parallel multi-baseline stereo vision. In *Computer Architectures for Machine Perception*. New Orleans, LA. Also appeared in ARPA Image Understanding Workshop, Washington, D.C.

Weng, J. (1993). Image matching using the windowed Fourier phase. *International Journal of Computer Vision*, 11(3):211–236.

Westelius, C.-J. (1995). *Focus of Attention and Gaze Control for Robot Vision*. PhD thesis, Linköping University. Dept. of Electrical Engineering Dissertation 379.

Williams, L. (1978). Casting curved shadows on curved surfaces. In *ACM Computer Graphics SIGGRAPH*, pages 270–274.

Willson, R. G. (1994). *Modeling and Calibration of Automated Zoom Lenses.* PhD thesis, Carnegie Mellon Electrical and Computer Engineering.

Xiong, Y. (1995). *High Precision Image Matching and Shape Recovery.* PhD thesis, Carnegie Mellon Robotics Institute.

Yeshurun, Y. & Schwartz, E. L. (1989). Cepstral filtering on a columnar image architecture: A fast algorithm for binocular stereo segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):759–767.

Zhang, Z., Faugeras, O., & Deriche, R. (1995). Calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment. In *Europe-China Workshop on Geometrical modelling and Invariance for Computer Vision*, pages 253–260.

Zitnick, C. L. & Webb, J. A. (1995). The reformulation of stereo vision. Unpublished Tech Report by Visual Interface, Inc.

Index