Stereo Integral Equation

Grahame B. Smith

Artificial Intelligence Center,
SRI International
Menlo Park, California 94025

Abstract

A new approach to the formulation and solution of the problem of recovering scene topography from a stereo image pair is presented. The approach circumvents the need to solve the correspondence problem, returning a solution that makes surface interpolation unnecessary. The methodology demonstrates a way of handling image analysis problems that differs from the usual linear-system approach. We exploit the use of nonlinear functions of local image measurements to constrain and infer global solutions that must be consistent with such measurements. Because the solution techniques we present entail certain computational difficulties, significant work still lies ahead before they can be routinely applied to image analysis tasks.

1 Introduction

The recovery of scene topography from a stereo pair of images has typically proceeded by three, quasi-independent steps. In the first step, the relative orientation of the two images is determined. This is generally achieved by selecting a few scene features in one image and finding their counterparts in the other image. From the position of these features, we calculate the parameters of the transformation that would map the feature points in one image into their corresponding points in the other image. Once we have the relative orientation of the two images, we have constrained the position of corresponding image points to lie along lines in their respective images. Now we commence the second phase in the recovery of scene topography, namely, determining a large number of corresponding points. The purpose of the first step is to reduce the difficulty involved in finding this large set of corresponding points.

Because we have the relative orientation of the two images, we only have to make a one-dimensional search (along the epipolar lines) to find points in the two images that correspond to the same scene feature. This step, usually called solving the "correspondence" problem, has received much attention. Finding many corresponding points in stereo pairs of images is difficult. Irrespective of whether the technique employed is area-based correlation or that of edge-based matching, the resultant set of corresponding points is usually small, compared with the number of pixels in the image. The solution to the correspondence problem, therefore, is not a dense set of points over the two images but rather a sparse set. Solution of the correspondence problem is made more difficult in areas of the scene that are relatively featureless or when there is much repeated structure, constituting

local ambiguity. To generate the missing intermediate data, the third step of the process is one of surface interpolation.

Scene depth at corresponding points is calculated by simple triangulation; this gives a representation in which scene depth values are known for some set of image plane points. To fill this out and to obtain a dense set of points at which scene depth is known, an interpolation procedure is employed. Of late there has been significant interest in this problem and various techniques that use assumptions about the surface properties of the world have been demonstrated [1,3,5,8]. Such techniques, despite some difficulties, have made it possible to reconstruct credible scene topography.

Of the three steps outlined, the initial one of finding the relative orientation of the two images is really a procedure designed to simplify the second step, namely, finding a set of matched points. We can identify several aspects of these first two steps that suggest the need for an alternative view of the processes entailed in reconstructing scene topography from stereo image pairs.

The techniques employed to solve the correspondence problem are usually local processes. When a certain feature is found in one image, an attempt is made to find the corresponding point in the other image by searching for it within a limited region of that image. This limit is imposed not just to reduce computational costs, but to restrict the number of comparisons so that false matches can be avoided. Without such a limit many points may "match" the feature selected. Ambiguity cannot be resolved by a local process; some form of global postmatching process is required. The difficulties encountered in featureless areas and where repeated structure exists are those we bring upon ourselves by taking too local a view.

In part, the difficulties of matching even distinct features are self-imposed by our failure to build into the matching procedure the shape of the surface on which the feature lies. That is, when we are doing the matching we usually assume that a feature lies on a surface patch that is orthogonal to the line of sight – and it is only at some later stage that we calculate the true slope of the surface patch. Even when we try various slopes for the surface patch during the matching procedure, we rarely return after the surface shape has been estimated to determine whether that calculated shape is consistent with the best slope actually found in matching.

In the formulation presented in the following sections, the problem is deliberately couched in a form that allows us to ask the question: what is the shape of the surface in the world that can account for the two image irradiances we see when we view that surface from the two positions represented by the stereo pair? We make no assumptions about the surface shape to do the matching – in fact, we do not do any matching at all. What we are interested in is recovering the surface that explains simul-

taneously all the parts of the irradiance pattern that are depicted in the stereo pair of images. We seek the solution that is globally consistent and is not confused by local ambiguity.

In the conventional approach to stereo reconstruction, the final step involves some form of surface interpolation. This is necessary because the previous step – finding the corresponding points – could not perform well enough to obviate the need to fabricate data at intermediate points. Surface interpolation techniques employ a model of the expected surface to fill in between known values. Of course, these known data points are used to calculate the parameters of the models, but it does seem a pity that the image data encoding the variation of the surface between the known points are ignored in this process and replaced by assumptions about the expected surface.

In the following formulation we eliminate the interpolation step by recovering depth values at all the image pixels. In this sense, the image data, rather than knowledge of the expected surface shape, guide the recovery algorithm.

We previously presented a formulation of the stereo reconstruction problem in which we sought to skirt the correspondence problem and in which we recovered a dense set of depth values [6]. That approach took a pair of image irradiance profiles, one from the left image and its counterpart from the right image, and employed an integration procedure to recover the scene depth from what amounted to a differential formulation of the stereo problem. While successful in a noise-free context, it was extremely sensitive to noise. Once the procedure, which tracked the irradiance profiles, incurred an error recovery proved impossible. Errors occurred because there was no locally valid solution. It is clear that that procedure would not be successful in cases of occlusion when there are irradiance profile sections that do not correspond. The approach described in this paper attempts to overcome these problems by finding the solution at all image points simultaneously (not sequentially, as in the previous formulation) and making it the best approximation to an overconstrained system of equations. The rationale behind this methodology is based on the expectation that the best solution to the overconstrained system will be insensitive both to noise and to small discrepancies in the data, e.g., at occlusions. While the previous efforts and the work presented here aimed at similar objectives, the formulation of the problem is entirely different. However, the form of the input – image irradiance profiles – is identical.

The new formulation of the stereo reconstruction task is given in terms of one-dimensional problems. We relate the image irradiance along epipolar lines in the stereo pair of images to the depth profile of the surface in the world that produced the irradiance profiles. For each pair of epipolar lines we produce a depth profile, from which the profile for a whole scene may then be derived. The formulation could be extended directly to the two-dimensional case, but the essential information and ideas are better explained and more easily computed in the one-dimensional case.

We couch this presentation in terms of stereo reconstruction, although there is no restriction on the acquisition positions of the two images; they may equally well be frames from a motion sequence.

2 Stereo Geometry

As noted earlier, our formulation takes two image irradiance profiles – one from the left image, one from the right – and describes the relationship between these profiles and the corresponding
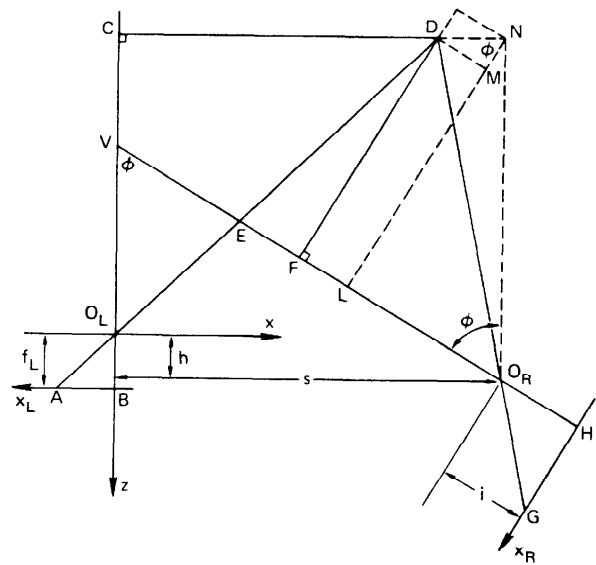
$$\frac{AB}{O_L B} = \frac{CD}{CO_L}$$

$$\frac{GH}{O_R H} = \frac{FD}{FO_R}$$

D is the point $(x, -z)$

$DN = (s - x)$

$O_R N = (h - z)$

Figure 1: Stereo Geometry. The two-dimensional arrangement in the epipolar plane that contains the optical axis of the left imaging system.

depth profile of the scene. The two irradiance profiles we consider are those obtained from corresponding epipolar lines in the stereo pair of images. Let us for the moment consider a pair of cameras pointed towards some scene. Further visualize the plane containing the optical axis of the left camera and the line joining the optical centers of the two cameras, i.e., an epipolar plane. This plane intersects the image plane in each camera, and the image irradiance profiles along these intersections are the corresponding irradiance profiles that we use. Of course, there are many epipolar planes, not just the one containing the left optical axis. Consequently, each plane gives us a pair of corresponding irradiance profiles. For the purpose of this formulation we can consider just the one epipolar plane containing the left optical axis since the others can be made equivalent. A description of this equivalence is given in a previous paper [6]. Figure 1 depicts the two-dimensional arrangement. AB and GH are in the camera image planes, while O_L and O_R are the cameras' optical centers. D is a typical point in the scene and AD and GD are rays of light from the scene onto the image planes of the cameras. From this diagram we can write two equations that relate the image coordinates x_L and x_R to the scene coordinates x and z. These are standard relationships that derive from the geometry of stereo viewing. For the left image

$$\frac{x}{-z} = \frac{x_L}{f_L} \quad ,$$

while for the right image

$$\frac{x}{-z} = g_R(x_R) - \frac{(s + g_R(x_R)h)}{z} \quad ,$$

where

$$g_R(x_R) = \frac{x_R \cos\phi - i\sin\phi}{x_R \sin\phi + i\cos\phi}$$

In addition, it should be noted that the origin of the scene coordinates is at the optical center of the left camera, and therefore the z values of all world points that may be imaged are such that

$$z < 0$$

3 Irradiance Considerations

From any given point in a scene, rays of light proceed to their image projections. What is the relationship between the scene radiance of the rays that project into the left and the right images? Let us suppose that the angle between the two rays is small. The bidirectional reflectance function of the scene's surface will vary little, even when it is a complex function of the lighting and viewing geometry. Alternatively, let us suppose that the surface exhibits Lambertian reflectance. The scene radiance is independent of the viewing angle; hence, the two rays will have identical scene radiances, irrespective of the size of the angle between them. For the model presented here, we assume that the scene radiance of the two rays emanating from a single scene point is identical. This assumption is a reasonable one when the scene depth is large compared with the separation distance between the two optical systems, or when the surface exhibits approximate Lambertian reflectance. It should be noted that there are no assumptions about albedo (i.e., it is not assumed to be constant across the surface) nor, in fact, is it even necessary to know or calculate the albedo of the surface. Since image irradiance is proportional to scene radiance, we can write, for corresponding image points,

$$I_L(x_L) = I_R(x_R)$$

I_L and I_R are the image irradiance measurements for the left and right images, respectively. It should be understood that these measurements at positions x_L and x_R are made at image points that correspond to a single scene point x.

While the above assumption is used in the following formulation, we see little difficulty in being less restrictive by allowing, for example, a change in linear contrast between the image profile and the real profile.

4 Integral Equation

Let us consider a single scene point x. For this scene point, we can write $I_L(x) = I_R(x)$. This equality relation holds for any function F of the image irradiance, that is, $F(I_L(x)) = F(I_R(x))$. If we let p select the particular function we want to use from some set of functions, we shall write

$$F(p, I_L(x)) = F(p, I_R(x))$$

The set of functions we use will be the set of all nonlinear functions for which $F(p_1, I) \neq \alpha(p_1, p_2) F(p_2, I)$ for all p. A specific example of such a function is $F(p, I) = I^p$.

The foregoing functions relate to the image irradiance. We can combine them with expressions that are functions of the stereo geometry. In particular, for the as yet unspecified function T of $\frac{x}{-z}$, we can write

$$F(p, I_L(x))\frac{d}{dx}T(\frac{x}{-z(x)}) = F(p, I_R(x))\frac{d}{dx}T(\frac{x}{-z(x)}) .$$

We have written z as $z(x)$ to emphasize the fact that the depth profile we wish to recover is a function of x. Should a more concrete example of our approach be required, we could select $T(\frac{x}{-z}) = \ln(\frac{x}{-z})$, which, when combined with the example for F above, gives us

$$I_L{}^p(x)\frac{d}{dx}\ln(\frac{x}{-z(x)}) = I_R{}^p(x)\frac{d}{dx}\ln(\frac{x}{-z(x)})$$

We now propose to develop the left-hand side of the above expression in terms of quantities that can be measured in the left stereo image, and develop the right-hand side in terms of quantities from the right stereo image. If we were to substitute x_L for x in the left-hand side of the above expression and x_R for x in the right-hand side, we would have to know the correspondence between x_L and x_R. This is a requirement we are trying to avoid. At first, we shall integrate both sides of the above expression with respect to x before attemping substitution for the variable x:

$$\int_a^b F(p, I_L(x))\frac{d}{dx}T(\frac{x}{-z(x)})dx =$$

$$\int_a^b F(p, I_R(x))\frac{d}{dx}T(\frac{x}{-z(x)})dx \quad ,$$

where a and b are specific scene points. Now let us change the integration variable in the left-hand side of the above expression to x_L, and the integration variable in the right-hand side to x_R:

$$\int_{a_L}^{b_L} F(p, I_L(x_L))\frac{d}{dx_L}T(\frac{x_L}{f_L})dx_L =$$

$$\int_{a_R}^{b_R} F(p, I_R(x_R))U(x_R)dx_R \quad , \quad (1)$$

where

$$U(x_R) = \frac{d}{dx_R}T(g_R(x_R) - \frac{(s + g_R(x_R)h)}{z(x_R)})$$

Equation (1) is our formulation of the stereo integral equation. Given that we have two image irradiance profiles that are matched at their end points – i.e., a_L and b_L in the left image correspond, respectively, to a_R and b_R in the right image – then Equation (1) expresses the relationship between the image irradiance profiles and the scene depth. It will be noted that the left-hand side of Equation (1) is composed of measurements that can be made in the left image of the stereo pair, while measurements in the right hand side are those that can be made in the right image. In addition, the right-hand side has a function of the scene depth as a variable. Our goal is to recover z as a function of the right-image coordinates x_R, not as a function of the world coordinates x. Once we have $z(x_R)$, we can transform it into any coordinate frame whose relationship to the image coordinates of the right image is known.

The recovery of $z(x_R)$ is a two-stage process. After first solving Equation (1) for $U(x_R)$, we integrate the latter to find $z(x_R)$ by using

$$T(g_R(x_R) - \frac{(s + g_R(x_R)h)}{z(x_R)}) =$$

$$T(g_R(a_R) - \frac{(s + g_R(a_R)h)}{z(a_R)}) + \int_{a_R}^{x_R} U(x'_R)dx'_R .$$

In this expression one should note that $z(a_R)$ is known, since a_R and a_L are corresponding points.

It is instructive as regards the nature of the formulation if we look at the means of solving this equation when we have discrete data. In particular, let us take another look at an example previously introduced, namely,

$$F(p, I) = I^p$$

$$T(\frac{x}{-z}) = \ln(\frac{x}{-z})$$

and hence

$$\int_{a_L}^{b_L} \frac{I_L{}^p(x_L)}{x_L} dx_L = \int_{a_R}^{b_R} I_R{}^p(x_R)U(x_R)dx_R \quad ,$$

and then

$$z(x_R) = \frac{(s + g_R(x_R)h)}{g_R(x_R) - K \exp \int_{a_R}^{x_R} U(x'_R)dx'_R} \quad ,$$

where

$$K = (g_R(a_R) - \frac{(s + g_R(a_R)h)}{z(a_R)})$$

Suppose that we have image data at points $x_{L1}, x_{L2}, ..., x_{Lq}$ that lie between the left integral limits and, similarly, that we have data from the right image, between its integral limits, at points $x_{R1}, x_{R2}, ..., x_{Rn}$. Further, let us approximate the integrals as follows:

$$\sum_{j=1}^{q} \frac{I_L{}^p(x_{Lj})}{x_{Lj}} = \sum_{j=1}^{n} I_R{}^p(x_{Rj})U(x_{Rj})$$

In actual calculation, we may wish to use a better integral formula than that above, (particularly at the end points), but this approximation enables us to demonstrate the essential ideas without being distracted by the details. Although the above approximation holds for all values of p, let us take a finite set of values, $p_1, p_2, ..., p_m$, and write the approximation out as a matrix equation, namely,

$$\begin{bmatrix} I_R{}^{p_1}(x_{R1}) & I_R{}^{p_1}(x_{R2}) & ... & I_R{}^{p_1}(x_{Rn}) \\ I_R{}^{p_2}(x_{R1}) & I_R{}^{p_2}(x_{R2}) & ... & I_R{}^{p_2}(x_{Rn}) \\ ... & ... & ... & ... \\ I_R{}^{p_n}(x_{R1}) & I_R{}^{p_n}(x_{R2}) & ... & I_R{}^{p_n}(x_{Rn}) \\ ... & ... & ... & ... \\ I_R{}^{p_m}(x_{R1}) & I_R{}^{p_m}(x_{R2}) & ... & I_R{}^{p_m}(x_{Rn}) \end{bmatrix} \begin{bmatrix} U(x_{R1}) \\ U(x_{R2}) \\ ... \\ U(x_{Rn}) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{q} \frac{I_L{}^{p_1}(x_{Lj})}{x_{Lj}} \\ \sum_{j=1}^{q} \frac{I_L{}^{p_2}(x_{Lj})}{x_{Lj}} \\ ... \\ \sum_{j=1}^{q} \frac{I_L{}^{p_n}(x_{Lj})}{x_{Lj}} \\ ... \\ \sum_{j=1}^{q} \frac{I_L{}^{p_m}(x_{Lj})}{x_{Lj}} \end{bmatrix}$$

Let us now recall what we have done. We have taken a set of image measurements, along with measurements that are just some non-linear functions of these image measurments, multiplied them by a function of the depth, and expressed the relationship between the measurements made in the right and left images. Why should one set of measurements, however purposefully manipulated, provide enough constraints to find a solution with almost the same number of variables as there are image measurements? The matrix equation helps in our understanding of this. First, we are not trying to find the solution for the scene depth at each point independently, but rather for all the points simultaneously. Second, we are exploiting the fact that, if the functions of image irradiance used by us are nonlinear, then each equation represented in the above matrix is linearly independent and constrains the solution. There is another way of saying this: even though we have only one set of measurements, requiring that the one depth profile relates the irradiance profile in the left image to the irradiance profile in the right image, and also relates the irradiance squared profile in the left image to the irradiance squared profile in the right image, and also relates the irradiance cubed profile etc., provides constraints on that depth profile.

The question arises as to whether there are sufficient constraints to enable a unique solution to the above equations to be found. This question really has three parts. Does an integral equation of the form of Equation (1) have a unique solution? This is impossible to answer when the irradiance profiles are unknown; even when they are known an exceedingly difficult problem confronts us [2,4]. Does the discrete approximation, even with an unlimited number of constraints, have the same solution as the integral equation? Again this is extremely difficult to answer even when the irradiance profiles are known. The final question relates to the finite set of constraint equations, such as those shown above. Does the matrix equation have a unique solution, and is it the same as the solution to the integral equation? Yes, it does have an unique solution – or at least we can impose solution requirements that makes a unique answer possible. But the question that asks whether the solution we find is a solution of the integral equation remains unanswered. From an empirical standpoint, we would be satisified if the solution we recover is a believable depth profile. Issues about sensitivity to noise, function type, and the form of the integral approximation will be discussed later in the section on solution methods.

Let us return to considerations of the general equation, Equation (1). We have just remarked upon the difficulty of solving this equation, so any additional constraints we can impose on the solution are likely to be beneficial. In the previous section on geometrical constraints, we noted that an acceptable solution has $z < 0$ and hence $z(x_R) < 0$. Unfortunately, solution methods for matrix equations (that have real coefficients) find solutions that are usually unrestricted over the domain of the real numbers. To impose the restriction of $z(x_R) < 0$, we follow the methods of Stockham [7]; instead of using the function itself, we formulate the problem in terms of the logarithm of the function. Consequently, in Equation (1) we usually set $T(\frac{x}{-z}) = \ln(\frac{x}{-z})$, just as we have done in our example. It should be noted that use of the logarithm also restricts $x > 0$ if $z < 0$. To construct the $x < 0$ side of the stereo reconstruction problem, we have to employ reflected coordinate systems for the world and image coordinates. Use of the logarithmic function ensures $z < 0$ and allows us to use standard matrix methods for solving the system of constraint equations. Once we have found the solution to the matrix equation, we can integrate that solution to find the depth profile.

In our previous example, we picked $F(p, I) = I^p$. In our experiments, we have used combinations of different functions to establish a particular matrix equation. For example we have used functions such as

$$\begin{aligned} F(p, I) &= |\cos pI|^p \\ &= (p + I)^{\frac{p}{2}} \\ &= p^I \\ &= \sin pI \\ &= (p + I)^{\frac{1}{2}} \end{aligned}$$

and we often use image density rather than image irradiance.

The point to be made here is that the form of the function F in the general equation is unrestricted, provided that it is nonlinear.

Equation (1) provides a framework for investigating stereo reconstruction in a manner that exploits the global nature of the solution. This framework arises from the realization that nonlinear functions provide a means of creating an arbitary number of constraints on that solution. In addition, the framework provides a means of avoiding the correspondence problem, except at the end points, for we never match points. Solutions have the same resolution as the data and this allows us to avoid the interpolation problem.

5 Solution Methods

Equation (1) is an inhomogeneous Fredholm equation of the first kind whose kernel function is the function $F(p, I_R(x_R))$. To solve this equation, we create a matrix equation in the manner previously shown in our example. We usually approximate the integral with the trapezoidal rule, where the sample spacing is that corresponding to the image resolution. Typically we use more than one functional form for the function F, each of which is parameterized by p. We have noticed that the sensitivity of the solution to image noise is affected by the choice of these functions, although we have not yet characterized this relationship. In the matrix equation, we usually pick the number of rows to be approximately twice the number of columns. However, owing to the rank-deficient nature of the matrix and hence to the selection of our solution technique, the solution we recover is only marginally different from the one obtained when we use square matrices.

Unfortunately, there are considerable numerical difficulties associated with solving this type of integral equation by matrix methods. Such systems are often ill-conditioned, particularly when the kernel function is a smooth function of the image coordinates. It is easy to see that, if the irradiance function varies smoothly with image position, each column of the matrix will be almost linearly dependent on the next. Consequently, it is advisible to assume that the matrix is rank-deficient and to utilize a procedure that can estimate the actual numerical rank. We use singular-valued decomposition to estimate the rank of the matrix; we then set the small singular values to zero and find the pseudoinverse of the matrix. Examples of results obtained with this procedure are shown in the following section.

An alternative approach to solving the integral equation is to decompose the kernel function and the dependent variable into orthogonal functions, then to solve for the coefficients of this decomposition, using the aforementioned techniques. We have used Fourier spectral decomposition for this purpose. The Fourier coefficients of the depth function were then calculated by solving a matrix equation composed of the Fourier components of image irradiance. However, the resultant solution did not vary significantly from that obtained without spectral decomposition.

While the techniques outlined can handle various cases, they are not as robust as we would like. We are actively engaged in overcoming the difficulties these solution methods encounter because of noise and irradiance discontinuities.

6 Results and Discussion

Our examples make use of synthetic image profiles that we have produced from known surface profiles. The irradiance profiles were generated under the assumptions that the surface was a Lambertian reflector and that the source of illumination was a
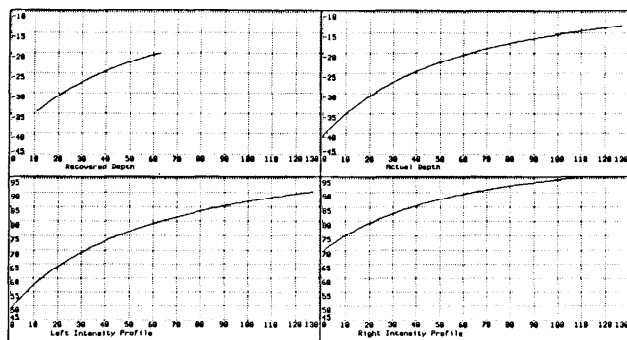

Figure 2: Planar Surface. At the upper left is depicted the recovered depth from the two irradiance profiles shown in the lower half. For comparison, the actual depth is shown in the upper right.

point source directly above the surface. This choice was made so that our assumption concerning image irradiance, namely, that $I(x_L) = I(x_R)$ at matched points, would be complied with. In addition, synthetic images derived from a known depth profile allow comparison between the recovered profile and ground truth. Nonetheless, our goal is to demonstrate these techniques on real-world data. It should be noted that the examples used have smooth irradiance profiles; they therefore represent a worst case for the numerical procedures, as the matrix is most ill-conditioned under these circumstances.

Our first example, illustrated in Figure 2, is of a flat surface with constant albedo. In the lower half of the figure, the left and right irradiance profiles are shown, while in the upper right, ground truth – the actual depth profile as a function of the image coordinates of the right image, x_R – is shown. The upper left of the figure contains the recovered solution. The limits of the recovered solution correspond to our selection of the integral end points. This solution was obtained from a formulation of the problem in which we used image density instead of irradiance in the kernel of the integral equation, and for which the function T was $\ln\left(\frac{x}{-z(x)}\right)$.

The second example, Figure 3, shows a spherical surface with constant albedo, except for the stripe we have painted across the surface. The recovered solution was produced from the same formulation of the problem as in the previous example. The ripple effects in the recovered profile appear to have been induced by the details of the recovery procedure; the attendant difficulties are in part numerical in nature. However, any changes made in the actual functions used in the kernel of the equation do have effects that cannot be dismissed as numerical inaccuracies.

As we add noise to the irradiance profiles, the solutions tend to become more oscillatory. Although we suspect numerical problems, we have not yet ascertained the method's range of effectiveness. This aspect of our approach, however, is being actively investigated.

In the formulation presented here, we have used a particular function of the stereo geometry, $\frac{x}{-z}$, in the derivation of Equation (1) but we are not limited to this particular form. Its attractiveness is based on the fact that, if we use this particular function of the geometry, the side of the integral equation related to the left image is independent of the scene depth. We have used other functional forms but these result in more complicated integral equations. Equations of these forms have been subjected to relatively little study in the mathematical literature. Consequently, the effectiveness of solution methods on these forms remains un-
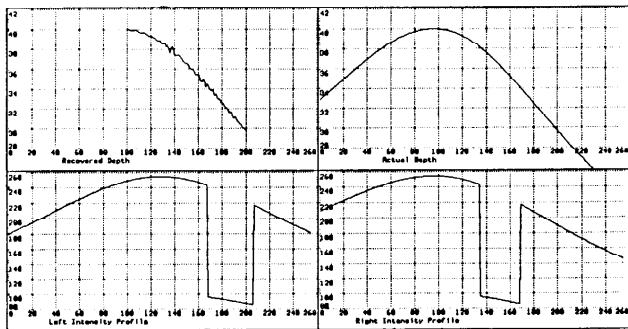
Figure 3: Spherical Surface with a Painted Stripe.

known.

In most of our study we have used $T(\frac{x}{-z})$ to be $\ln(\frac{x}{-z})$ and the properties of this particular formulation should be noted. It is necessary to process the right half of the visual field separately from the left half. The integral is more sensitive to image measurements near the optical axis than those measurements off-axis. In fact, the irradiance is weighted by the reciprocal of the distance off-axis. If we were interested in an integral approximation exhibiting uniform error across the extent of that integral, we might expect measurements that had been taken at interval spacings proportional to the off-axis distance to be appropriate. While it is obvious that two properties of a formulation that match those of the human visual system do not in themselves give cause for excitement it is worthy of note that the formulation presented is at least not at odds with the properties of the human stereo system.

On balance, we must say that significant work still lies ahead before this method can be applied to real-world images. While the details of the formulation may be varied, the overall form presented in Equation (1) seems the most promising. Nonetheless, solution methods for this class of equations are known to be difficult and, in particular, further efforts towards the goal of selecting appropriate numerical procedures are essential.

In formulating the integral equation, we took a function of the image irradiance and multiplied it by a function of the stereo geometry. To introduce image measurements, we changed variables in the integrals. If we had not used the derivative of the function of the stereo geometry, we would have had to introduce terms like $\frac{dz}{dz_R}$ and $\frac{dz}{dz_L}$ into the integrals. By introducing the derivative we avoided this. However, we did not really have to select the function of the geometry for this purpose; we could equally well have introduced the derivative through the function of image irradiance. Then we would have exchanged the calculation of irradiance gradients for the direct recovery of scene depth (thus eliminating the integration step we now use). Our selection of the formulation presented here was based on the belief that irradiance gradients are quite susceptible to noise; consequently we prefered to integrate the solution rather than differentiate the data. In a noise-free environment, however, both approaches are equivalent (as integration by parts will confirm).

7 Conclusion

The formulation presented herein for the recovery of scene depth from a stereo pair of images is based not on matching of image features, but rather on determining which surface in the world is consistent with the pair of image irradiance profiles we see. The solution method does not attempt to determine the nature

of the surface locally; it looks instead for the best global solution. Although we have yet to demonstrate the procedure on real images, it does offer the potential to deal in a new way with problems associated with albedo change, occlusions, and discontinuous surfaces. It is the approach, rather than the details of a particular formulation, that distinguishes this method from conventional stereo processing.

This formulation is based on the observation that a global solution can be constrained by manufacturing additional constraints from nonlinear functions of local image measurements. Image analysis researchers have generally tried to use linear-systems theory to perform analysis; this has led them, consequently, to replace (at least locally) nonlinear functions with their linear approximation. Here we exploit the nonlinearity; "What is one man's noise is another man's signal."

While the presentation of the approach described here is focussed upon stereo problems, its essential ideas apply to other image analysis problems as well. The stereo problem is a convenient problem on which to demonstrate our approach; the formulation of the problem reduces to a linear system of equations, which allows the approach to be investigated without diversion into techniques for solving nonlinear systems. We remain actively interested in the application of this methodology to other problems, as well as in the details of the numerical solution.

References

[1] Boult, T.E., and J.R. Kender, "On Surface Reconstruction Using Sparse Depth Data," *Proceedings: Image Understanding Workshop*, Miami Beach, Florida, December 1985.

[2] Courant, R., and D. Hilbert, *Methods of Mathematical Physics*, Interscience Publishers, Inc., New York, 1953.

[3] Grimson, W.E.L., "An Implementation of a Computational Theory of Visual Surface Interpolation," *Computer Vision, Graphics, and Image Processing*, Vol. 22, pp 39-69, April 1983.

[4] Hildebrand, F.B., *Methods of Applied Mathematics*, 2nd ed., Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.

[5] Smith, G.B., "A Fast Surface Interpolation Technique," *Proceedings: Image Understanding Workshop*, New Orleans, Louisiana, October 1984.

[6] Smith, G.B., "Stereo Reconstruction of Scene Depth," *IEEE Proceedings on Computer Vision and Pattern Recognition*, San Francisco, California, June 1985, pp 271-276.

[7] Stockham, T.G., "Image Processing in the Context of a Visual Model," *Proceeding of IEEE*, Vol. 60, pp 828-842, July 1972.

[8] Terzopoulos, D., "Multilevel Computational Processes for Visual Surface Reconstruction," *Computer Vision, Graphics, and Image Processing*, Vol. 24, pp 52-96, October 1983.