

A Binocular Stereo Algorithm for Reconstructing Sloping, Creased, and Broken Surfaces in the Presence of Half-Occlusion

Peter N. Belhumeur
Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Abstract

In this paper we present a method for reconstructing the three dimensional scene geometry (i.e. depth, surface orientation, occluding contours, and surface creases) from a pair of stereo images. This reconstruction is not done as a post-processing step, but rather all of the above quantities are estimated simultaneously as part of the matching algorithm. We argue for an energy functional in which each of the quantities in the scene geometry is explicitly represented. For this energy functional we use a smoothness prior which, in addition to its ability to detect surface discontinuities and the accompanying half-occluded regions, is able to reconstruct steeply sloping surfaces with sharp creases. Experimental results are presented demonstrating the effectiveness of the algorithm.

1 Introduction

In the literature, binocular stereo vision algorithms construct a depth map of the 3-D surfaces captured in a pair of images taken from slightly different viewpoints. These surfaces are estimated by first finding matching pixels¹ in the images that correspond to the same point on a 3-D surface, and then computing the point's depth as a function of its displacement (or *disparity*) in the two images. The task of matching points between the two images is known as the *correspondence problem*.

This problem is made difficult by several well known complications. First, due to the inaccuracy of the measurements and the effects of variation in lighting between images, there are often many possible matches for a any given pixel. Second, many images contain large regions of constant luminance and, therefore, are effectively featureless in these regions. Even with near perfect measurements and minimal lighting variations between images, the matching is still ambiguous for a great number of pixels. Third, special

¹ We use the expression "matching a pixel or point" to mean matching a feature located at that point. The features might be image luminance, edges, or other filter responses.

attention must be paid to preserve the salient features in the scene geometry - namely the discontinuities in depth at the boundaries of objects, the discontinuities in the surface orientation, and steeply sloping surfaces. Finally, due to the phenomenon of occlusion, there are almost certainly whole regions of *half-occluded* points which appear in only one image and, consequently, have no match at all.

For years people have offered solutions to the correspondence problem without adequately addressing all of these complications. Many have convincingly argued that, by enforcing a smoothness constraint, the complications caused by variations in lighting and large featureless regions could be minimized. In some algorithms, the smoothness constraint was integrated into the matching process in the form of a smoothness "prior" which biased toward solutions with smooth surfaces. Here smoothness was imposed either as a soft constraint in an energy functional that biased toward depth maps with low order derivatives (see, for instance, Barnard [3]); or as a hard constraint by locally fitting the surface by low order polynomials (see Cernuschi et al. [8]). In many area-based algorithms, surfaces were assumed to be locally planar and the disparity was determined by a "winner take all" fit of a small planar patch at each point in the images.

Yet, while algorithms using smoothness constraints proved very effective in handling the first two complications, their performance deteriorated at salient features in the scene geometry. Discontinuities in depth at object boundaries ("breaks") or discontinuities in surface orientation ("creases") were either smoothed over or cause the algorithm to produce erratic results.

Recently Yuille [19] maintained that if a smoothness prior is used to influence the matching, there must be some mechanism for suspending the smoothing at the boundaries of objects. Here the suggestion was that "line processes" (i.e. binary random process) used to solve the segmentation problem (see Geman & Geman [10], Mumford & Shah [16], and Blake & Zisserman [6]) should be used to explicitly represent discontinuities in depth. However, what makes stereo different from the the segmentation problem is that in addition to identifying boundaries across which smoothing should be suspended due to a discontinuity, one must

also identify whole regions of half-occlusion caused by the discontinuity.

This paper argues that, in order to properly address these complications, a stereo algorithm should maintain, internally, a detailed map of the scene geometry. This map should include not simply depth, but also surface orientation, discontinuities in depth (also referred to throughout this paper as occluding contours), and creases. In the past, many algorithms have found these quantities by post-processing, as a second pass, the depth map obtained from a stereo matching algorithm (see again [6]). More recently, Wilde [21] proposed post-processing not the depth, but the disparity to obtain the scene geometry. In general, the problem with these approaches is that the matching process is separated from the process of identifying these quantities. For example, it is unclear how an algorithm using smoothing as a second pass is able to distinguish between discontinuities due to object boundaries and discontinuities due to false matches.

We propose that depth, surface orientation, occluding contours, and creases should be estimated simultaneously.² To accomplish this we put forth an algorithm in the form of an energy functional in which the quantities in the scene geometry are explicitly represented. In [5] we introduced a method in which depth and occluding contours were represented. Unfortunately, the method relied too heavily on smoothing using the first derivative of depth and was, consequently, overly biased toward fronto-parallel surfaces. Furthermore, the previous method had no mechanism for suspending smoothing at the creases of objects. This paper significantly expands on this approach to incorporate surface orientation and creases in developing a more sophisticated smoothness prior - *one which preserves large depth gradients and creases common to most objects.*

2 Binocular Camera Geometry and Half-Occlusion

We must first define the terminology, notation, and basic equations for the geometry of our binocular stereo setup. These concepts have been described in detail in [5], but a few refinements have been added. To start, let us assume that we have two pinhole cameras whose optical axes are parallel and separated by some baseline distance. A point (or a small patch) p on the surface of an object in 3-D space is projected through the focal points and onto the image planes of the cameras. The brightness of each point projected onto the image planes creates image luminance functions I_l and I_r in the left and right planes, respectively. Next, let us create an imaginary cyclopean im-

age plane in the same manner, placing its focal point on the baseline half-way between the original two focal points. We look now at a horizontal plane through these focal points. It intersects the three image planes in what are called epipolar lines, which we denote X_l , X_r , and X , with coordinates $x_l \in X_l$, $x_r \in X_r$, and $x \in X$, respectively.

When the same point is visible from all three eyes it is easy to check that $x = (x_l + x_r)/2$. Thus, we can relate the coordinates of points projected onto all three image planes by a positive disparity function $d(x)$ via

$$x_l = x + d(x) \text{ and } x_r = x - d(x).$$

The distance $D(x)$ from the cyclopean focal point to a point p on the surface of an object can be related to the disparity $d(x)$ by $D(x) \simeq k/d(x)$ where k is some positive constant dependent on the focal length and baseline distance.

Now suppose a surface point is not visible to all three eyes. How are we to define $d(x)$ and $D(x)$? The simplest thing to do is to let $D(x)$ be the distance from the cyclopean focal point to the nearest surface point, and define $d(x) = k/D(x)$. But if this patch is occluded from the perspective of the left or right eye (or camera), the image values $I_l(x - d(x))$ and $I_r(x + d(x))$ will not be related to the light reflected off this patch.

To see when this patch is visible from both eyes, it is convenient to introduce a filtered version $d^*(x)$ of $d(x)$ as

$$d^*(x) = \max_a (d(x + a) - |a|).$$

Graphically, d^* is constructed by taking the graph of d , and letting each peak cast shadows at 45° to the left and right. Thus $|d^*(x) - d^*(y)| \leq |x - y|$, and $|(d^*)'(x)| \leq 1$. To interpret d^* in terms of occlusion, let us say that a point p visible to the cyclopean eye in direction x is *mutually visible* to the left and right eyes if and only if $d^*(x) = d(x)$.

We define the *half-occluded points* $O \subset X$ to be the closure of the set of x such that $d^*(x) > d(x)$, i.e. the set of points *not* mutually visible. Half-occluded regions are most commonly formed by a foreground surface partially occluding a background surface such that *there is a region on the background surface visible to both the left and the right eyes*.³ Half-occluded points will be in general the unmatched pixels, unless the ordering constraint is violated and a point p is visible from both eyes even though some smaller object lies in the triangle formed by p , the left focal point, and the right focal point.⁴ The most common way for unmatched pixels to arise is for $d(x)$ to jump discontinuously as it tracks visible points from points on one surface to points on a new surface. We define the

²Note that the quantities in our scene geometry are exactly those which Marr termed the $2\frac{1}{2}$ -D sketch (see Marr [14]). Yet Marr argued that low level modules for stereo, motion, shape contours, shading, and texture combine their output to form the $2\frac{1}{2}$ -D sketch. Here we argue that stereo algorithms should have these quantities explicitly represented.

³Half-occluded regions can be formed by two foreground surfaces partially occluding a background surface such that *there is no region on the background surface visible to both the left and the right eyes*. We do not consider this type of half-occlusion in this paper.

⁴This unusual possibility is usually referred to as the "double nail illusion."

break points $B \subset X$ to be the set of x where $d(x)$ is discontinuous. To the left or right of these points, we have half-occluded regions. To see this, note that at such a point $|d'(x)|$ is infinite, so nearby we must have $d''(x) > d(x)$.

3 The Algorithm in 1-D

This section introduces our stereo algorithm in 1-D for reconstructing a map of the scene geometry which includes depth, surface orientation, occluding contours, and creases. We present our algorithm in the form of an energy functional whose minimum is a maximum a posteriori (MAP) reconstruction of the quantities in the scene geometry. While there is too little space here to present derivations of these energy functionals as Bayesian MAP estimators, we refer the reader to [4] and [5]. In the two subsections to follow, we argue that smoothness priors using the first derivative of disparity are inadequate for reconstructing certain types of surfaces - namely surfaces with steep slopes or creases. We review our previous algorithm precisely to point out these shortcomings. We then introduce our matching algorithm with modifications in the smoothness prior which allow for long steeply sloped surfaces and sharp creases. Experimental results are presented comparing the effectiveness of these algorithms.

3.1 Representing only Depth and Occluding Contours

In an earlier work [5], we developed an algorithm for solving the correspondence problem that used a Bayesian treatment in deriving a measure of goodness fit and a prior based on a simplified model of objects in space. This approach led to an energy functional depending both on disparity as measured from a "cyclopean" perspective and on the implied half-occluded regions from the left and right eye perspectives. Explicit in this formulation are depth⁵ and occluding contours. In [4] we develop an approximation to the algorithm in [5] by placing a prior on the disparity function itself, not on the formation of objects and their surfaces in space. In 1-D along epipolar lines, the force of this algorithm is captured in the energy functional⁶

$$E[d, B] = E_D + E_P \quad (1)$$

where

$$E_D = \int_{X/O} (F_l(x+d) - F_r(x-d))^2 dx + \int_O \nu dx$$

$$E_P = \lambda^2 \int_{X/B} (d')^2 dx + \alpha_B |B|$$

⁵We use depth and disparity interchangeably since they are known functions of one another.

⁶Although the derivations and implementations are done in the discrete form, we have written down the continuous form to make the equations more readable.

and where: O is the set of half-occluded points; B is the set of discontinuities in d ; F_l and F_r are functions representing the features⁷ along the left and right epipolar lines; ν , λ , and α_B are preset constants whose choices are motivated in [4] and [5]; X/O is the region $X - O$, likewise for X/B ; and $|B|$ is the cardinality of the set B . In 1-D, our solution to the correspondence problem is simply the function $\hat{d}$ and the set $\hat{B}$ which minimizes the energy functional $E[d, B]$.

The strength of this algorithm is that the depth estimates, occluding contours, and half-occluded regions are inseparably linked. All of these quantities are found simultaneously. The E_D term forces the reconstruction to agree with the features in the data, but only for mutually visible points (X/O). For half-occluded points (O), we take a penalty ν which is proportional to variance of the noise in the images. The E_P term biases toward smooth reconstructions with the degree of the bias given by the constant λ , where λ is determined by the expected smoothness of the surfaces. Note, however, that the E_P term allows the smoothing to be suspended at the breaks (B) in the disparity. For break points, we take a penalty α_B which determined by the expected size of objects.

To demonstrate the effectiveness of this algorithm we introduce, at the end of this paper, some data and results. Fig. 1 shows a stereo pair of a cardboard R in front of a flat background. To the right of the R is a Rubik's cube sitting on an upside-down paper cup. Figure 2 contains the results⁸ produced by minimizing a 2-D variant of Eq. 1: Fig. 2a is an image of the depth map in which light corresponds to near and dark to far; Fig. 2b is a map of the occluding contours; and Fig. 2c is a wire frame view of the depth map. Notice that the occluding contours of the R and small bump of the Rubik's cube and paper cup are accurately reconstructed.

While this result is encouraging, it also deceptive in that the surfaces in the stereo pair are fronto-parallel and without creases. In other words, while the matching is made difficult by the large regions of constant luminance and the half-occlusion to the left and right of the foreground R , the scene does not contain steeply sloping horizontal surfaces or creases.

One might ask how this algorithm would perform on a more difficult stereo pair - in particular, one with large horizontal disparity gradients and creases present? Unfortunately, it would not perform as well. Figure 3 shows a stereo pair of a Q-tip box stood on end with its long vertical crease protruding toward the cameras. The box stands in front of a flat background. Figure 4 contains the results⁹ produced again by mini-

⁷By features we mean the output from filtering the left and right image luminance values with either a linear or nonlinear filter. This paper does not include any discussion of optimal choices for these filters.

⁸In this result, the functions F_l and F_r are taken to be the image luminance. The result took 10 minutes to generate on a DECstation 3100.

⁹In this result, the functions F_l and F_r are taken to be the convolution of image luminance with a DOG (difference of



Figure 1: (a) Cardboard R - left. (b) Cardboard R - right.

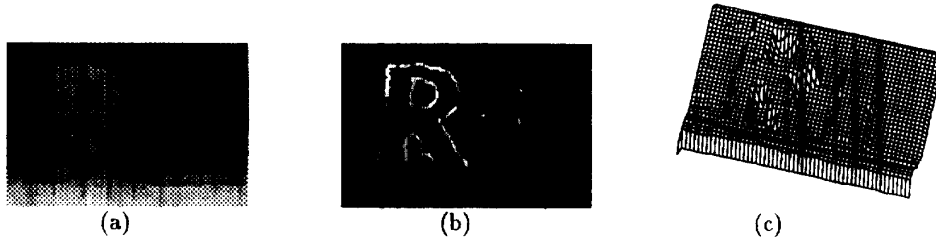


Figure 2: (a) Image of depth. (b) Occluding contours. (c) Wire frame of depth.

mizing a 2-D variant of Eq. 1. Although the occluding contour of the Q-tips box is correctly reconstructed, the algorithm smooths over the long vertical crease of the box and only partially captures the steep disparity gradient of the sides of the box. If the energy functional's smoothing parameter λ is decreased in an effort to better preserve the crease of the box and the disparity gradient, the results become more erratic.

3.2 Incorporating Surface Orientation and Creases

The weakness of the above algorithm is its inability to handle heavily sloping surfaces and to identify creases. The above algorithm can be significantly improved, however, by explicitly representing surface orientation and creases in devising a more sophisticated smoothness prior. In [4] we derive this prior in detail, in this paper we only present it.

Let us introduce the slope function m as a smoothed version of the derivative of the disparity d' . Let us also introduce the set C as the points x where the surface is creased, or, more precisely, as points at which there is a discontinuity in the slope m . With these new definitions, we rework Eq. 1 to create the 1-D energy functional¹⁰

$$E[d, m, B] = E_D + E_{P_d} + E_{P_m} \quad (2)$$

Gaussians) filter. The result took 10 minutes to generate on a DECstation 3100.

¹⁰This functional is similar to the one used by Harris [11] for the segmentation problem.

where

$$E_D = \int_{X/O} (F_l(x+d) - F_r(x-d))^2 dx + \int_O \nu dx$$

$$E_{P_d} = \lambda^2 \int_{X/B} (d' - m)^2 dx + \alpha_B |B|$$

$$E_{P_m} = \mu^4 \int_{X/B \cup C} (m')^2 dx + \alpha_C |C|$$

and where: O is the set of half-occluded points; B is the set of discontinuities in d ; C is the set of discontinuities in m ; F_l and F_r are functions representing the features along the left and right epipolar lines; ν , λ , μ , α_B , and α_C are preset constants whose choices are motivated in [4] and [5]; X/O is the region $X - O$, likewise for X/B and $X/B \cup C$; and $|B|$ is the cardinality of the set B , likewise for $|C|$. In 1-D, our solution to the correspondence problem is simply the functions $\hat{d}$ and $\hat{m}$ and the sets $\hat{B}$ and $\hat{C}$ which minimizes the energy functional $E[d, m, B]$. We can briefly summarize the differences of the two energy functionals in the following ways.

First, by incorporating the slope m we are able to create smoothing terms E_{P_d} and E_{P_m} that do not over bias toward fronto-parallel disparity. The E_{P_d} term biases toward reconstructions in which the derivative of the disparity d' is close to the slope m , with the degree of bias given by the parameter λ . The E_{P_m} term biases toward smooth reconstructions of the slope m , with the degree of bias given by the parameter μ . For any linear disparity function, $E_{P_d} + E_{P_m} = 0$,

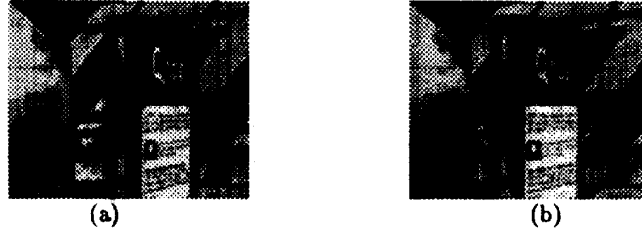


Figure 3: (a) Qtips box - left. (b) Qtips box - right.

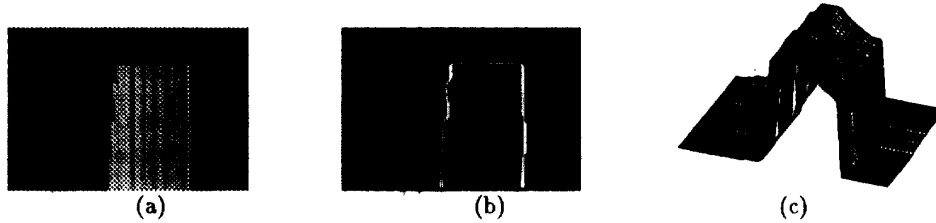


Figure 4: (a) Image of depth. (b) Occluding contours. (c) Wire frame of depth.

while for the earlier energy functional $E_P \neq 0$. This improvement allows the algorithm to reconstruct surfaces with strong disparity gradients. While there is some evidence that the human visual system is biased toward fronto-parallel surfaces (see, for instance, Bulthoff et al. [7]), this bias must be much more subtle than simply the (squared first derivative of the disparity) bias in E_P . Specifically, E_P favors surfaces which are globally fronto-parallel but locally very rough, over surfaces which are globally slightly slanted but locally very smooth (see again [6]). This bias is correctly reversed for the smoothing terms E_{P_d} and E_{P_m} . Note that, although we could achieve this property by smoothing using only the second derivative of disparity, in [4] we present detailed arguments for why this is, in fact, inferior.

Second, the smoothing term E_{P_m} is suspended at creases, while the term E_{P_d} is enforced. This allows the slope m to jump discontinuously at the creases, while keeping the disparity continuous. For creases, we take a penalty α_C which is determined by the expected frequency of creases. This improvement allows the algorithm to reconstruct not only discontinuities in disparity at object boundaries, but also the discontinuities in surface orientation at creases in objects.

Figure 5 shows the results¹¹ produced by minimizing a 2-D variant of Eq. 2 on the stereo pair shown in Fig. 3. In addition to depth and occluding contours, these results contain surface orientation and creases.

¹¹In this result, the functions F_l and F_r are taken to be the image luminance. The result took 60 minutes to generate on a DECstation 3100.

Fig. 5a is an image of the depth; Fig. 5b is an image of the horizontal slope; Fig. 5c is an image of the occluding contours (white) and the creases (grey); and Figure 5d is a wire frame of the depth. Here the sharp disparity gradients and the long vertical crease of the Qtips box are, for the most part, perfectly preserved.

4 The Algorithm in 2-D

In the previous sections we introduced a 1-D energy functional for finding of the scene geometry along the individual epipolar lines. In this 1-D formulation the solutions along the epipolar lines are obtained independently of one another. Clearly these solutions are not independent: There are strong smoothness constraints binding epipolar lines (see Marr & Poggio [15], Baker & Binford [2], and Otha & Kanade [18]). Here we extend the energy functional $E[d, m, B, C]$ to incorporate vertical smoothing, explicitly representing the vertical slopes. The matching will now be done the points (x, y) in a 2-D image plane $X \subset \mathcal{R}^2$. The disparity, horizontal slope, and vertical slope functions are $d(x, y)$, $m(x, y)$, and $n(x, y)$, respectively. For notational purposes, we define $\mathbf{m} = (m, n)$. With these definitions, we can write down the 2-D analog of Eq. 2 as the functional

$$E[d, \mathbf{m}, B, C] = E_D + E_{P_d} + E_{P_m} \quad (3)$$

where



Figure 5: (a) Image of depth. (b) Image of slope. (c) Occluding contours and creases. (d) Wire frame of depth.

$$\begin{aligned}
 E_D &= \iint_{X/O} (F_l(x+d, y) - F_r(x-d, y))^2 dx dy \\
 &\quad + \iint_O \nu dx dy \\
 E_{P_d} &= \lambda^2 \iint_{X/B} \|\nabla d - \mathbf{m}\|^2 dx dy + \alpha_B |B| \\
 E_{P_m} &= \mu^4 \iint_{X/B \cup C} \|\nabla m\|^2 + \|\nabla n\|^2 dx dy + \alpha_C |C|
 \end{aligned}$$

and where: $O \subset X$ are the 2-D regions of half-occluded points; B is a collection of contours across which d is discontinuous; C is a collection of contours across which either m or n is discontinuous; F_l and F_r are 2-D functions representing the features in the left and right images; ν , λ , μ , α_B , and α_C are pre-set constants; X/O is the region $X - O$, likewise for X/B and $X/B \cup C$; and $|B|$ is the total length of the contours of the set B , likewise for $|C|$. In 2-D, our solution to the correspondence problem is simply the functions $\hat{d}$ and $\hat{\mathbf{m}}$ and the sets $\hat{B}$ and $\hat{C}$ which minimizes the energy functional $E[d, \mathbf{m}, B, C]$. Note that these quantities are exactly those in Marr's $2\frac{1}{2}$ -D sketch (see again [14]).

5 Implementation

In section 4 we claimed that our solution is simply the reconstructions $\hat{d}$, $\hat{\mathbf{m}}$, $\hat{B}$, and $\hat{C}$ which minimize the energy functional $E[d, \mathbf{m}, B, C]$ of Eq. 3. Unfortunately, solving variational problems of this type is not such a simple task - there are no known methods for finding optimal solutions. Possible options for finding approximate solutions would be to first discretize the energy in Eq. 3 to define the energy for a Markov random field (MRF), and then apply a stochastic annealing (see again [10]) or a deterministic annealing algorithm (see Yuille [20]). Yet, while these methods have proven effective for simpler energy functionals, they are often too slow or unreliable for more complex energies.

In this section we present an alternative two stage method for finding approximate solutions. In the first

stage we find solutions along epipolar lines by optimizing a discrete analog of the 1-D energy functional in Eq. 2. Because the matching and smoothing in $E[d, \mathbf{m}, B, C]$ is done only in 1-D, we are able to optimize the energy functional by applying the formalism of dynamic programming (as done for stereo by Henderson et al. [12] and Baker & Binford [2]). Naturally, the advantage of this approach is that we are guaranteed of finding the optimum solution. The disadvantage is that the solutions along epipolar lines are found independently so there is no vertical smoothing. The second stage of the algorithm uses the solutions found in the first stage as a starting point for minimizing a discrete analog of the 2-D energy functional in Eq. 3. To do this, we use what we call “*iterated stochastic dynamic programming*.” The subtleties in our algorithm warrant the discussion in the following subsections.

5.1 In 1-D

Before applying dynamic programming, we must cast our problem in a discrete setting. To do this, let us take the fixed interval X of the cyclopean epipolar line and sample it at n evenly spaced points represented by the vector $\mathbf{x} = (x_1, \dots, x_n)$. Let the disparities and slopes at the sampled points be represented by the vectors $\mathbf{d} = (d_1, \dots, d_n)$ and $\mathbf{m} = (m_1, \dots, m_n)$. Recall from section 2 that $x \in O$ if $d^*(x) \neq d(x)$. In the discrete setting, this condition translates as $x_i \in O$ if $d_j - d_i > |x_j - x_i|$ for any j . Furthermore, recall that $x \in B$ if d is discontinuous at x . While there is no discrete equivalent of a function being discontinuous, we define that $x_i \in B$ if $|d_{i+1} - d_i| > 1$, the point at which the surface is so steeply sloped that it becomes half-occluded. With these definitions, we write the discrete analog of Eq. 2 as

$$E[\mathbf{d}, \mathbf{m}, B] = E_D + E_{P_d} + E_{P_m} \quad (4)$$

where

$$\begin{aligned}
E_D &= \sum_{x/O} (F_l(x_i + d_i) - F_r(x_i - d_i))^2 + \sum_O \alpha_O \\
E_{P_d} &= \lambda^2 \sum_{x/B} (d_{i+1} - d_i - m_i)^2 + \alpha_B |B| \\
E_{P_m} &= \mu^4 \sum_{x/B \cup C} (m_{i+1} - m_i)^2 + \alpha_C |C|.
\end{aligned}$$

To explain the dynamic programming, the following notation is helpful. For simplicity, let us rename $d = (d_1, m_1, \dots, d_n, m_n)$. Let $d_j^k = (d_j, m_j, \dots, d_k, m_k)$ be the subset of disparities and slopes from points j to k . Let $d_j = (d_j, m_j)$ be the disparity and slope at j .

For the present, let us ignore breaks and half-occlusions. Notice the optimal set for C is $\hat{C} = \{x_i \mid \mu^4(\hat{m}_{i+1} - \hat{m}_i)^2 > \alpha_C\}$. Therefore, if we constrain our search fixing $C = \{x_i \mid \mu^4(m_{i+1} - m_i)^2 > \alpha_C\}$, then we can write

$$\begin{aligned}
\min_{d, C} E[d, C] &= \min_d E[d] \\
&= \min_{d_1^n} E[d_1^n] \\
&= \min_{d_n} (\min_{d_1^{n-1}} E[d_1^{n-1}]).
\end{aligned} \tag{5}$$

We can break down the above minimization by applying the following recursive step

$$\min_{d_1^{j-1}} E[d_1^j] = \min_{d_{j-1}} (\min_{d_1^{j-2}} E[d_1^{j-1}] + \Psi_P^{j-1}) + \Psi_D^j \tag{6}$$

where

$$\begin{aligned}
\Psi_D^j &= (F_l(x_j + d_j) - F_r(x_j - d_j))^2 \\
\Psi_P^j &= \lambda^2 (d_{j+1} - d_j - m_j)^2 + \min(\mu^4 (m_{j+1} - m_j)^2, \alpha_C).
\end{aligned}$$

In fact, this trick can be repeated over and over until we are left, at the depth of the recursion, with

$$\min_{d_1} E[d_1^2] = \min_{d_1} (\Psi_D^1 + \Psi_P^1) + \Psi_D^2.$$

If we discretize, with *sub-pixel fineness*, the disparity and slope values as $d_i \in \{0, \dots, d_{\max}\}$ ¹² and $m_i \in \{-1, \dots, 1\}$ ¹³ $\forall i$ with M_d and M_m being the number of possible disparity and slope values, then we can use this recursive procedure to find the optimal solution in $O(NM_d^2M_m^2)$ time.

At this point we allow for the possibility of breaks and the implied half-occlusions. If a point is in O , then there is no penalty for matching; instead, there is a fixed penalty ν . Therefore, the disparities in regions O are simply those that minimize the smoothness prior. Section 3.2 pointed out that, for linear

fits, the smoothness penalties E_{P_d} and E_{P_m} are zero. Therefore, we take the disparity values in the half-occluded regions O to be linear extensions of X/O .

This observation is key to our algorithm, essentially allowing us to jump over the half-occluded regions in the minimization. If we allow breaks and half-occlusion, then the recursive step in Eq. 6 can be broken down into two disjoint possibilities. The first possibility is that *the pixel to the left of x_j is mutually visible, $x_{j-1} \notin O$* . The second possibility is that *the pixel to the left of x_j is not mutually visible, $x_{j-1} \in O$* . In the first case the minimization is, for the most part, as before. In the second case, we must find the best possible previously visible point. We can write down this step as

$$\min_{d_1^{j-1}} E[d_1^j] = \min(\Theta_{nb}, \Theta_b) \tag{7}$$

where

$$\begin{aligned}
\Theta_{nb} &= \min_{d_{j-1}} (\min_{d_1^{j-2}} E[d_1^{j-1}] + \Psi_P^{j-1}) + \Psi_D^j \\
\Theta_b &= \min_{l \geq 2} (\min_{d_1^{j-l}} E[d_1^{j-l}] + (l-1)\nu) + \alpha_B + \Psi_D^j
\end{aligned}$$

and where: the subscripts of Θ_{nb} and Θ_b stand for no break and break; the l index indicates the size of the jump, so $l-1$ is the number of half-occluded pixels; the $*$ in the expression Θ_{nb} indicates that the constraint $|d_j - d_{j-1}| \leq 1$ is enforced; and the $**$ in the expression Θ_b indicates that the constraint $l-1 < |d_j - d_{j-l}| \leq l$ is enforced. Using these procedures, we find the optimal solution in $O(NM_d^2M_m^2)$ time. However, since many of these calculations can be done in advance, we can dramatically reduce the number of operations to roughly $O(NM_d^2M_m^2/d_{\max}^2)$.

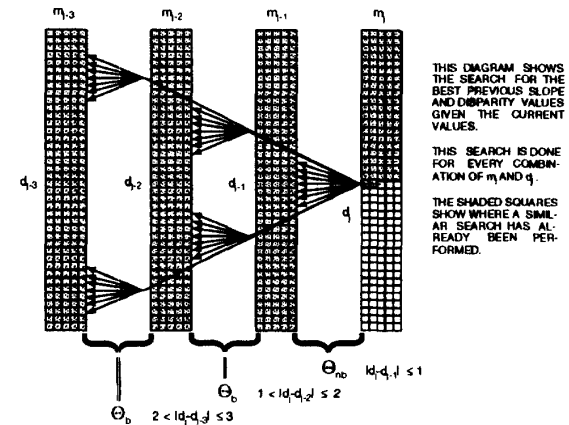


Figure 6: Illustration of $\min_{d_1^{j-1}} E[d_1^j]$.

This search is illustrated in Fig. 6. The array elements represent, for each point, the different possible combinations of the disparity and slope. The rows

¹²As in the human visual system, we put a limit on the maximum possible disparity - $d_{\max}$.

¹³The magnitude of the slope, $|m_i|$, can not exceed 1 if the surface at the point x_i is mutually visible to both the left and right cameras.

correspond to the different possible disparities ranging from 0 to $d_{\max}$. The columns correspond to different possible slopes ranging from -1 to 1 . Recorded in each array element is the extremal energy up to that point and the path followed to get there. The arrows indicate the search at a particular d_j and m_j over the space of possible previous values. The spray of arrows furthest to the right represent the search Θ_{nb} assuming there is no break. The rest of the arrows to the left of these represent the search Θ_b assuming there was a break and pixels were half-occluded.

5.2 In 2-D

As mentioned, there is not an obtainable optimum solution to the 2-D energy functional of Eq. 3. However, we propose using what we call "iterated stochastic dynamic programming." The algorithm incorporates the vertical smoothing by performing dynamic programming on an epipolar line in the discrete form of Eq. 3 given the values along neighboring epipolar lines are fixed. This algorithm can be described in the following steps:

1. Use, as an initial starting point, the solutions to the 1-D energy functional of Eq. 2, so that initially there is no vertical smoothing.
2. Use the disparity values to estimate the vertical slopes.
3. Randomly select three adjoining epipolar lines and fix the variables along the two outside epipolar lines. Use dynamic programming to find the optimum solution for the middle epipolar given the values along the outside epipolar lines are fixed. Do this minimization for every epipolar line.
4. Repeat steps 1-3 until there is no significant change in the overall 2-D energy (about 5 times).

All of the results in this paper use this procedure to find the reconstructions shown in the figures. However we have not, as yet, allowed for a vertical slopes.

6 Conclusion

In this paper we have presented a new stereo algorithm for reconstructing not only depth, but also occluding contours, surface orientation, and creases. These quantities are all computed simultaneously. We do not claim that our method is a general purpose solution, but rather, we argue that our method considers important complications common to the matching problem. We have presented results that demonstrate that our method is quite effective at handling the considered complications. As always, questions still remain. For instance: Are there adaptive methods for choosing the preset parameters? Would this method benefit from a multi-resolution approach? What are better methods for optimizing the 2-D energy functional? What features would yield optimal results?

Dedication

For my mother Helena Una Belhumeur.

References

- [1] H. H. Baker, *Depth from edge and intensity based stereo*, PhD thesis, University of Illinois, Urbana, IL, 1982.
- [2] H. H. Baker and T. O. Binford, "Depth from edge and intensity based stereo," in *Proc. of 7th IJCAI 1981*, vol. 2, pp. 631-636.
- [3] S. Barnard, "Stochastic stereo matching of scale," in *Int. J. Computer Vision*, vol. 3, pp. 17-33.
- [4] P. N. Belhumeur, *A Bayesian Approach to the Stereo Correspondence Problem*, PhD dissertation, Division Applied Sciences, Harvard Univ., 1993.
- [5] P. N. Belhumeur and D. Mumford, "A Bayesian treatment of the stereo correspondence problem using half-occluded regions," *Proc. IEEE Conf. CVPR*, Urbana, IL, June 1992.
- [6] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.
- [7] H. H. Bulthoff, M. Fahle, M. Wegmann, "Disparity gradients and depth scaling," *Perception*, vol. 20, pp. 145-153.
- [8] B. Cernuschi-Frias, D. B. Cooper, Y. P. Hung, and P. N. Belhumeur, "Toward a model-based Bayesian theory for estimating and recognizing parameterized 3-D objects using two or more images taken from different positions," *IEEE Trans. Pattern Anal. Machine Intell.*, November 1989, pp. 1028-1052.
- [9] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and binocular stereo," *EECV*, 1991.
- [10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions, AMI* 6, pp. 721-741, 1984.
- [11] J. G. Harris, "The coupled depth/slope approach to surface reconstruction," *Artificial Intelligence Lab*, no. TR-908, MIT, Cambridge, MA.
- [12] R. L. Henderson, W. J. Miller, C.B. Grosch, "Automatic stereo recognition of man-made targets," *Soc. Photo-Optical Instrumentation Engineers*, vol. 186, August 1979.
- [13] D. Jones, *Computational Models of Binocular Vision*, PhD dissertation, Dept. of Computer Science, Stanford Univ., 1991.
- [14] D. Marr, *Vision*, Freeman, San Francisco, 1982.
- [15] D. Marr and T. Poggio, "A theory of human stereo vision," *MIT AI Lab Memo* 451, 1979.
- [16] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," *Proc. IEEE CVPR Conf.*, vol. 22, 1985.
- [17] K. Nakayama and S. Shimojo, "Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points," *Vision Res.* Vol. 30, No. 11, 1990, pp. 1811-1825.
- [18] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 139-154, March 1985.
- [19] A. L. Yuille, "Energy functions for early vision and analog networks," *Biological Cybernetics*, vol. 61, pp. 115-123, 1989.
- [20] A. L. Yuille, D. Geiger, and Heinrich H. Bulthoff, "Stereo integration, mean field theory and psychophysics," *Network*, vol. 2, pp. 423-442, 1991.
- [21] R. P. Wilde, "Direct Recovery of three-dimensional scene geometry from binocular stereo disparity," *IEEE Trans. Pattern Anal. Machine Intell.*, August 1991, pp. 761-774.