LINEAR IMAGE FEATURES IN STEREOPSIS

Michael Kass

Schlumberger Palo Alto Research
3340 Hillview Ave.
Palo Alto, CA 94304

ABSTRACT

Most proposed algorithms for solving the stereo correspondence problem have used matching based in some way on linear image features. Here the geometric effect of a change in viewing position on the output of a linear filter is modeled. A simple local computation is shown to provide confidence intervals for the difference between filter outputs at corresponding points. Examples of the use of the confidence interval are provided. For some widely used filters, the confidence intervals are tightest at isolated vertical step edges, lending support to the idea of using edge-like features in stereopsis. However, the same conclusion does not apply to image regions with more complicated variation on the scale of the filter support.

I Introduction

Most proposed algorithms for solving the stereo correspondence problem have used matching based in some way on linear image filters. The algorithms are usually based on the assumption that the filter outputs will be very similar at corresponding points in the two images. Differences in viewing geometry between the two views, can however, introduce fairly large distortions. For any given filter, there are some local image patterns for which even small changes in the viewing geometry will cause large changes in the filter output. Here, a simple computation will be developed to identify such points by placing confidence intervals on the difference between filter outputs at corresponding points. For some widely used filters, the confidence intervals are tightest at isolated vertical step edges, lending support to the idea of using edge-like features in stereopsis. However, the same conclusion does not apply to image regions with more complicated variation on the scale of the filter support. In general, the confidence interval is constructed from two linear filters, one measuring sensitivity of the filter output to horizontal compression and the other measuring sensitivity of the filter output to vertical skew.

The stereo correspondence problem is the problem of matching two images of the same scene from different viewing positions. Let $I_1(x,y)$ and $I_2(x,y)$ be the light intensity functions for two images whose correspondence is to be computed and let $T : \Re^2 \mapsto \Re^2$ be the mapping from points in the first image to corresponding points in the scond image. Then for any point $\mathbf{p}$ in the domain C of $\mathbf{T}$, $I_1(\mathbf{p})$ and $I_2(\mathbf{T}(p))$ are projections of the same physical point. The problem is to recover $\mathbf{T}$ from I_1 and I_2.

All solutions to the stereo correspondence problem are based on finding some type of similarity between the local image intensities surrounding corresponding points. Understanding in detail how image intensities change under a change of viewpoint is critical in constructing good measures of similarity for computing correspondence.

A large number of stereo algorithms use measures of similarity based in some way on the outputs of linear image filters. Multiresolution correlation-based algorithms [e.g Hannah, 1974; Gennery, 1977; Moravec, 1977; Barnard and Thompson, 1980] typically use the outputs of linear low-pass filters for coarse matching. Edge-based algorithms [e.g Marr and Poggio, 1979; Grimson, 1981a; Mayhew and Frisby, 1981; Baker and Binford, 1981; Medioni and Nevatia 1983; Ohta and Kanade 1983] typically use linear filters to identify the locations of edges. The combination of a large number of independent linear filters [Kass 1983a; Kass 1983b] has also been used effectively to compute correspondence.

Even if there is no important change in photometry between views, the outputs of these filters at corresponding image points will in general be different because of difference in the projection geometry of the two views. Since typical similarity measures make use of linear filters with local support, it is the local change in geometry between views which is of concern. To keep the analysis manageable, we will assume that locally (on the scale of the filter support), the transformation T can be accurately represented by a first order approximation. If $\mathbf{T} : (x,y) \mapsto (x',y')$, the assumption is that

$$\mathbf{T} \approx (\mathbf{DT}) \cdot (x,y)^t \tag{1}$$

where

$$\mathbf{DT} = \begin{pmatrix} \partial x'/\partial x & \partial x'/\partial y \\ \partial y'/\partial x & \partial y'/\partial y \end{pmatrix}$$

is the Jacobian matrix of $\mathbf{T}$ and the lower case superscript t denotes matrix transposition. The operator $\mathbf{D}$ will be used throughout to represent the Jacobian of a vector field. Equation 1 defines an affine transformation, which will be a good local approximation as long as $\mathbf{T}$ is smooth and continuous. For planar surfaces under orthogonal projection, affine mappings correctly describe the stereo transformation. For curved surfaces under perspective projection, affine mappings are the best linear approximations to the transformation. The spatial extent of the affine approximation will be the region of support of the linear filter in question.

Even when $\mathbf{T}$ is limited to locally affine transformations, arbitrarily large distortions are possible between I_1 and I_2. These distortions can change the output of linear filters by arbitrarily large amounts. Fortunately, large distortions by $\mathbf{T}$ occur rarely.

In the next section, we will develop confidence limits on the components of $\mathbf{DT}$. Then in section III, we will use these limits to construct a simple local computation which provides confidence limits on the difference between corresponding values of filtered images. The confidence intervals vary over the image, so some points can be identified as unusually good or bad points to try to match. Finally, In section IV, the relevance of these results to the Marr-Poggio-Grimson algorithm and to the Kass algorithm will be discussed.

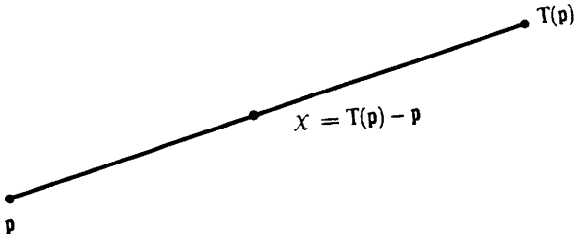

Figure 1: Definition of disparity field: Shift between $\mathbf{p}$ and $\mathbf{T}(\mathbf{p})$ is the disparity of the midpoint $(\mathbf{p} + \mathbf{T}(\mathbf{p}))/2$.

II Disparity Gradient Limits

An important observation about the transformation $\mathbf{T}$ was made by Arnold and Binford [1980]. Assuming a uniform distribution of surfaces on the Gaussian sphere, they were able to show that because of foreshortening, surfaces with steep depth gradients occupy only a small portion of most images. As a result, local transformations which cause extreme geometric distortion are rare.

In order to apply the Arnold and Binford results the problem of geometric distortion of linear filters, it is convenient to introduce the notion of a disparity field to represent the shift between the positions of corresponding points in a pair of images. Using a generalized version of the Burt-Julesz coordinate system, the disparity field $\chi(x, y)$ can be defined by the relationship:

$$\chi([\mathbf{p} + \mathbf{T}(\mathbf{p})]/2) = \mathbf{T}(\mathbf{p}) - \mathbf{p} \tag{2}$$

where $\mathbf{p}$ is a vector quantity (x, y). In the interest of symmetry, the shift $\mathbf{T}(\mathbf{p}) - \mathbf{p}$ is defined to be the disparity of the point halfway between $\mathbf{p}$ and $\mathbf{T}(\mathbf{p})$ (see figure 1). Other definitions of the disparity field have been used—the advantage of this coordinate system is that if I_1 and I_2 are exchanged, the disparity field merely changes its sign.

One problem with this definition of disparity is that under pathological conditions it can become multivalued. The problem is not very serious because for affine $\mathbf{T}$, χ is uniquely defined by the above formula except on a set of transformations of measure zero [Kass, 1984]. Thus the possible ambiguity in the definition of χ is not a major difficulty. Moreover, for affine $\mathbf{T}$, $\mathbf{D}\chi$ is constant.

Let H be the horizontal component of the disparity field and let V be its vertical component. Then if $\chi(0,0) = (0,0)$ and $\mathbf{T}$ is affine, the disparity field can be written

$$\chi(x, y) = \mathbf{D}\chi(x, y)^t = \begin{pmatrix} H_x & H_y \\ V_x & V_y \end{pmatrix} (x, y)^t. \tag{3}$$

In general, the non-translational component of a two-dimensional affine transformation can be decomposed into compression or expansion along two orthogonal axes, a rotation, and some skew. The compression and expansion components of $\mathbf{T}$ are determined by the diagonal elements of $\mathbf{D}\chi$ while the rotation and skew components are determined by its off-diagonal elements. A good discussion of the details of one possible decomposition as it relates to the disparity field can be found in Koenderink and Van Doorn [1976].

Under ordinary stereo viewing conditions, V_x and V_y are quite small so the geometric distortion is due primarily to H_x and H_y. As a consequence, the range of likely distortions is restricted to horizontal compression and vertical skew. Suppose $V_x = V_y = H_y = 0$ and $H_x \neq 0$. Then the two images are related by pure horizontal compression. Let $(x_2, y_2) = T((x_1, y_1))$. Since $\chi(x, y) = (H_x x, 0)$, we know $y_2 = y_1$ and

$$x_2 - x_1 = H_x(x_2 + x_1)/2. \tag{4}$$

Hence $x_2 = c_1(2 + H_x)/(2 - H_x)$. If we define $s = (2 + H_x)/(2 - H_x)$ then the transformation $\mathbf{T}$ can be described by the equation $\mathbf{T}((x_1, y_1)) = (sx_1, y_1)$ which describes horizontal compression by a factor of s^{-1}

If $H_x = 0$ and $H_y \neq 0$ then the two images are related by vertical skew. Consider a point (x_1, y_1) in the first image. Its corresponding point in the second image is $(x_1 + H_y y_1, y_1)$. Hence the line $ax = by$ in the first image will map to the line $ax = (b + aH_y)y$ in the second. Horizontal lines ($a = 0$) will be unchanged by the transformation, but all other lines are rotated by an angle that reaches a maximum of $\tan^{-1}(H_y)$ for vertical lines.

When neither H_x nor H_y is zero, compression and skew occur simultaneously. If the points (x_1, y) in the first image and (x_2, y) in the second image correspond, then since $H(x, y) = H_x x + H_y y$, we have

$$x_2 - x_1 = H_x(x_1 + x_2)/2 + H_y y \tag{5}$$

$$x_2 = x_1 s + y H_y/(1 - H_x/2). \tag{6}$$

Thus the horizontal compression is unaffected by the presence of skew, but the skew is adjusted by the factor $1/(1 - H_x/2)$.

Likely values of H_x and H_y are heavily constrained by foreshortening effects. Based on the Arnold-Binford assumption that surface orientations are uniformly distributed on the Gaussian sphere, complete distributions for H_x and H_y can be calculated. These distributions allow confidence limits on H_x and H_y to be established, so that the range of image compressions and skews to be considered can be suitably restricted.

Let a be the ratio between the inter-ocular distance and the viewing distance. Arnold and Binford calculate the cumulative distribution function of s, the ratio of horizontal lengths in the two images, to be that of a Cauchy random variable. The cumulative distribution function can be written as follows

$$Pr[s < x] = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{2(x - 1)}{a(x + 1)}. \tag{7}$$

Since $H_x = 2(s - 1)/(s + 1)$, the cumulative distribution of H_x can be written

$$Pr[H_x < x] = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x/a). \tag{8}$$

Negative values of s are due to occlusion. Substituting $x = 0$ into equation 7, we see that occlusion occurs on the Gaussian sphere with probability

$$\frac{1}{2} - \frac{1}{\pi} \tan^{-1}(2/a). \tag{9}$$

The upper quartile of the distribution of H_x begins where $(1/\pi)\tan^{-1}(x/a) = 1/4$. Multiplying through by π and taking the tangent of both sides shows that the equation is satisfied when $x = a$. Hence

$$Pr[-a < H_x < a] = \frac{1}{2}. \tag{10}$$

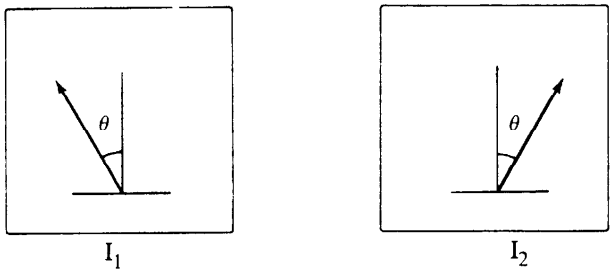

Figure 2: Line orientations for calculating the distribution of H_y

Differentiating the cumulative distribution function yields a Cauchy density for H_x

$$\frac{d}{dx} Pr[H_x < x] = \frac{a}{\pi(a^2 + x^2)}. \tag{11}$$

Thus the standard deviation of H_x is infinite. At $x = a$, the density falls to half its height at the origin, so in addition to marking the upper quartile, a is also the half-width at half-maximum (HWHM) for H_x.

The exact distribution of H_y is considerably more difficult to calculate. Suppose lines rotated counterclockwise from the y-axis by an angle θ in the first image correspond to lines rotated clockwise by θ in the second (see figure 2). Then $H_y = 2\tan\theta$. Arnold and Binford have calculated the joint distributions of line angles in stereo images assuming a uniform distribution of surface orientations on the Gaussian sphere. For $a = .07$ which is typical of human vision at a range of about one meter, Arnold and Binford find the HWHM of the angular difference between the two views (2θ) to be about 9 degrees. For $a = .7$ which is typical of wide angle aerial photography, the HWHM is 30 degrees. These values correspond to HWHM values for H_y of .16 and .54.

The following table summarizes probability of occlusion and HWHM values for H_x and H_y assuming values of a corresponding to human vision at a distance of about one meter and corresponding to typical wide angle aerial photography.

	a	H_x	H_y	Prob.
Human Vision	.07	.07	.16	.011
Aerial Photography	.70	.70	.54	.107

On the basis of psychophysical experiments with stereograms consisting of pairs of dots, Burt and Julesz [1980a, 1980b] have discovered that the human visual system seems unable to achieve fusion unless $|\nabla H| < 1$. The Burt-Julesz experiments are somewhat controversial (see Krol and van de Grind [1982] and the response by Burt and Julesz [1982]), but they are consistent with earlier work on sine wave gratings done by Tyler [1973, 1977] suggesting a disparity gradient limit for human stereopsis.

Based on the probability distributions calculated by Arnold and Binford, a visual system able to tolerate disparity gradients near the Burt-Julesz limit should have little difficulty with geometric distortion for conditions typical of human vision. Aerial photographs of mountainous regions could be expected to cause some problems, but for most other aerial photographs, a disparity gradient tolerance near one would probably be sufficient.

III Geometric Distortion Estimate

Given bounds on H_x and H_y, we can investigate limits on how much the outputs of linear filters can be distorted by the geometric differences between stereo images. Let $N_T(\mathbf{p})$ be the difference betwen the outputs of the filter f at corresponding points in the two images. Then

$$N_T(\mathbf{p}) = [f * I_2](T(p)) - [f * I_1](\mathbf{p}) \tag{12}$$

In general, the behavior of N_T is quite complex because it depends on both the transformation $\mathbf{T}(\mathbf{p})$ and the image $I(\mathbf{p})$. At some image points, large distortions between images caused by $T(\mathbf{p})$ will have only a small effect on the filter output. At these points, the filter outputs will be reliably preserved between views. At other image points, however, even a small amount of compression or skew induced by $\mathbf{T}(\mathbf{p})$ will have a large effect on the filter output. At these points, $[f * L_1](\mathbf{p})$ is a poor predictor of $[f * L_2](\mathbf{T}(\mathbf{p}))$.

The disparity gradient limits developed in section II provide a method of computing confidence limits on the transformation $\mathbf{T}(\mathbf{p})$. These confidence limits will be used here to develop confidence limits on $N_T(\mathbf{p})$. Since $N_T(\mathbf{p})$ depends on the local behavior of $I(\mathbf{p})$, the confidence limits will vary over the image.

By a simple change of variables, $N_T(\mathbf{p})$ can be rewritten as the convolution of $I(\mathbf{p})$ with a point-spread function that depends on the transformation T. This will make it easier to discuss the dependence of $N_T(\mathbf{p})$ on $I(\mathbf{p})$.

Convolving the deformed image $I \circ \mathbf{T}^{-1}$ with the point spread function f is the same as convolving the original image I with the mask $f \circ T$ and multiplying by the Jacobian determinant of T. This follows easily from a change of variables. At the origin, we have

$$(I \circ \mathbf{T}^{-1}) * f = \int\int I \circ \mathbf{T}^{-1}(x', y') f(-x', -y') dx' dy'. \tag{13}$$

Transforming into the (x, y) coordinate system, we obtain

$$\begin{aligned}(I \circ \mathbf{T}^{-1}) * f &= \int\int I(x, y) f \circ \mathbf{T}(-x, -y) |\mathbf{DT}| dx dy \\ &= I * (f \circ \mathbf{T}) |\mathbf{DT}|\end{aligned} \tag{14}$$

where

$$\mathbf{DT} = \begin{pmatrix} \partial x'/\partial x & \partial x'/\partial y \\ \partial y'/\partial x & \partial y'/\partial y \end{pmatrix} \tag{15}$$

is the Jacobian matrix of $\mathbf{T}$ and $|\mathbf{DT}|$ is its determinant. Thus the geometric distortion noise N_g can be expressed as a single filter f_{NT} applied to the first image:

$$N_g = I * [(f \circ \mathbf{T})|\mathbf{DT}| - f] = I * f_{NT}. \tag{16}$$

Unfortunately, the filter f_{NT} is not known exactly until the correspondence problem is solved. However, away from depth discontinuities, it is restricted by the surface orientation constraint on $\mathbf{T}$. In order to apply the constraint, we need to represent f_{NT} in terms of $\mathbf{D}\chi$, the Jacobian of the disparity field. To do so, we first compute the Jacobian determinant $|\mathbf{DT}|$.

A. Jacobian Determinant

Let H be the horizontal component and V be the vertical component of the disparity field χ. If the origins of the two image coordinate systems correspond and $\mathbf{T}$ is affine then $\chi = (\mathbf{D}\chi)\cdot(x,y)^T$ where

$$\mathbf{D}\chi = \begin{pmatrix} H_x & H_y \\ V_x & V_y \end{pmatrix} \tag{17}$$

is the Jacobian matrix of the disparity field. The diagonal elements of $\mathbf{D}\chi$ are responsible for horizontal and vertical compression and expansion, while the off-diagonal elements cause rotation and skew.

Using a Burt-Julesz type coordinate system, (x,y) in the first image corresponds to (x',y') in the second if $H((x+x')/2,(y+y')/2) = x'-x$ and $V((x+x')/2,(y+y')/2) = y'-y$. Since $\chi = (H_x x + H_y y, V_x x + V_y y)$, the transformation $\mathbf{T}$ can be described by the equations

$$F_1(x,y,x',y') = H_x(x+x')/2 + H_y(y+y')/2 - x' + x = 0 \tag{18}$$

$$F_2(x,y,x',y') = V_x(x+x')/2 + V_y(y+y')/2 - y' + y = 0. \tag{19}$$

The Jacobian $\mathbf{DT}$ and its determinant can be computed by means of the implicit function theorem from the equations $F_1 = 0$ and $F_2 = 0$ that define $\mathbf{T}$:

$$|\mathbf{DT}| = -\begin{pmatrix} \partial F_1/\partial x' & \partial F_1/\partial y' \\ \partial F_2/\partial x' & \partial F_2/\partial y' \end{pmatrix}^{-1} \begin{pmatrix} \partial F_1/\partial x & \partial F_1/\partial y \\ \partial F_2/\partial x & \partial F_2/\partial y \end{pmatrix}. \tag{20}$$

Substituting in the partial derivatives, we obtain

$$|\mathbf{DT}| = -\begin{pmatrix} H_x/2-1 & H_y/2 \\ V_x/2 & V_y/2-1 \end{pmatrix}^{-1} \begin{pmatrix} H_x/2+1 & H_y/2 \\ V_x/2 & V_y/2+1 \end{pmatrix}. \tag{21}$$

Since the determinant of a product is the product of the determinants, we have

$$|\mathbf{DT}| = \frac{4 + 2H_x + 2V_y + H_xV_y - H_yV_x}{4 - 2H_x - 2V_y + H_xV_y - H_yV_x}. \tag{22}$$

B. First-Order Approximation

Equation 16 gives the total geometric distortion N_g as the convolution of the intensity $I(\mathbf{p})$ with the filter $f_{NT} = [f - (f \circ \mathbf{T})|\mathbf{DT}|]$. We already have an expression for $|\mathbf{DT}|$ in terms of the components of $\mathbf{D}\chi$. To express the entire geometric distortion noise in terms of $\mathbf{D}\chi$ and $I(\mathbf{p})$, we will make a first-order approximation to N_g.

Let $S = (x, y, H_x, H_y, V_x, V_y)$ and $S_0 = (0,0,0,0,0,0)$. The surface orientation constraint assures us that H_x, H_y, V_x, and V_y are usually quite small and since $N_g = I * f_{NT}$, we can approximate N_g at the origin as

$$N_g = N_{ge} \approx I * S \cdot \nabla f_{NT}|_{\mathbf{D}\chi=0} \tag{23}$$

where the gradient is in the variables x, y, H_x, H_y, V_x, and V_y.

In order to compute ∇f_{NT}, we can use equation 16 directly. Hence

$$(\nabla f_{NT})|_{\mathbf{D}\chi=0} = \nabla\left((f \circ T)|\mathbf{DT}| - f\right)|_{\mathbf{D}\chi=0}. \tag{24}$$

It is convenient here to extend $\mathbf{T}$ such that it maps vectors $(x, y, H_x, H_y, V_x, V_y)$ to vectors $(x', y', H_x', H_y', V_x', V_y')$ with $H_x = H_x', H_y = H_y'$ etc. Note that the Jacobian determinant $|\mathbf{DT}|$ is unaffected by the change. With this extension of $\mathbf{T}$, the multidimensional product rule can be applied to the gradient in equation 24 to yield

$$\begin{aligned}(\nabla f_{NT})|_{\mathbf{D}\chi=0} = {} & ((f \circ \mathbf{T})\nabla|\mathbf{DT}|)|_{\mathbf{D}\chi=0} \\ & + (|\mathbf{DT}|\mathbf{D}(f \circ \mathbf{T}))|_{\mathbf{D}\chi=0} - \nabla f|_{\mathbf{D}\chi=0}.\end{aligned} \tag{25}$$

where f and $|\mathbf{DT}|$ are regarded as functions of the six variables x, y, H_x, H_y, V_x, and V_y. When $\mathbf{D}\chi = 0$, the function $\mathbf{T}$ becomes the identity so $f \circ T = f$ and $|\mathbf{DT}| = 1$. Hence equation 25 can be re-written as

$$(\nabla f_{NT})|_{\mathbf{D}\chi=0} = f(\nabla|\mathbf{DT}|)|_{\mathbf{D}\chi=0} + \mathbf{D}(f \circ \mathbf{T})|_{\mathbf{D}\chi=0} - \nabla f|_{\mathbf{D}\chi=0}. \tag{26}$$

The Jacobian $\mathbf{D}(f \circ \mathbf{T})$ according to the chain rule is $\nabla f \cdot \mathbf{DT}$. Substituting into equation 26 leaves

$$\begin{aligned}(\nabla f_{NT})|_{\mathbf{D}\chi=0} = {} & f\,(\nabla|\mathbf{DT}|)|_{\mathbf{D}\chi=0} \\ & + (\nabla f)\cdot(\mathbf{DT})|_{\mathbf{D}\chi=0} - \nabla f|_{\mathbf{D}\chi=0}.\end{aligned} \tag{27}$$

The last term on the right in equation 27 is simply the vector $(-f_x, -f_y, 0, 0, 0, 0)$. Straightforward calculation of the derivatives from equation 22 shows that the first term is $(0, 0, f, 0, 0, f)$. Thus we have

$$(\nabla f_{NT})|_{\mathbf{D}\chi=0} = (-f_x, -f_y, f, 0, 0, f) + (\nabla f)\cdot(\mathbf{DT})|_{\mathbf{D}\chi=0}. \tag{28}$$

The extended Jacobian $\mathbf{DT}$ can be calculated exactly as before by adding the equations

$$F_3(x,y,x',y') = H_x - H_x' = 0 \tag{29}$$

$$F_4(x,y,x',y') = H_y - H_y' = 0 \tag{30}$$

$$F_5(x,y,x',y') = V_x - V_x' = 0 \tag{31}$$

$$F_6(x,y,x',y') = V_y - V_y' = 0. \tag{32}$$

The implicit function theorem states that $\mathbf{DT}$ is the product of two matrices. When $\mathbf{D}\chi = 0$, the first matrix becomes the identity and we have $x = x'$, $y = y'$. Thus $\mathbf{DT}$ is simply

$$\mathbf{DT}|_{\mathbf{D}\chi=0} = \begin{pmatrix} 1 & 0 & x & y & 0 & 0 \\ 0 & 1 & 0 & 0 & x & y \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{33}$$

Multiplying out $(\nabla f)\cdot(\mathbf{DT})|_{\mathbf{D}\chi=0}$ and substituting in the expression for $\nabla f_{NT}|_{\mathbf{D}\chi=0}$ leaves

$$\nabla f_{NT}|_{\mathbf{D}\chi=0} = (0, 0, f + xf_x, yf_x, xf_y, f + yf_y). \tag{34}$$

Hence the estimated geometric distortion noise is

$$\begin{aligned} N_{ge} &= I * S \cdot (0, 0, f + xf_x, yf_x, xf_y, f + yf_y) \\ &= I * (H_x(f + xf_x) + H_y y f_x + V_x x f_y + V_y(f + yf_y)). \end{aligned} \tag{35}$$

Under ordinary stereo viewing conditions, $\nabla V << \nabla H$ so the estimate of N_{ge} can be simplified to

$$N_{ge} = I * (H_x(f + xf_x) + H_y y f_x). \tag{36}$$

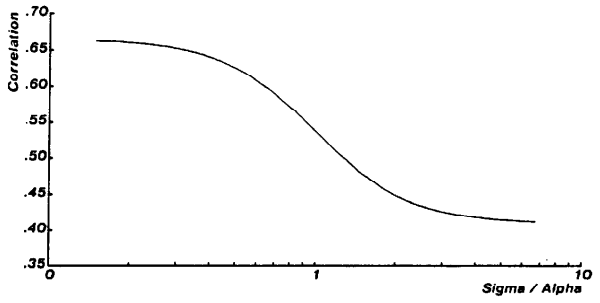

Figure 3: Correlation between $f * I$ and $f_H * I$ as a function of σ/α based on first-order Markov image model

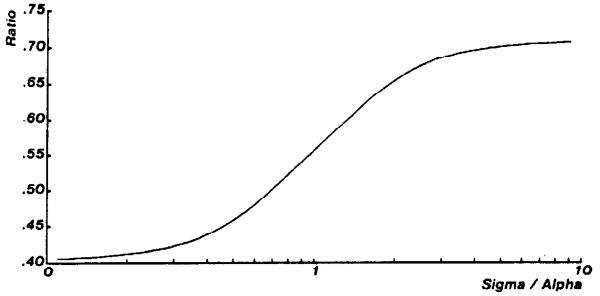

Figure 4: The ratio between the standard deviations of $f_V * I$ and $f * I$ as a function of σ/α based on the first-order Markov image model

C. Application of Orientation Constraint

Since H_x and H_y are unknown, equation 36 does not provide a method of calculating N_{ge} directly from an image. It does, however, give a method of translating constraints on H_x and H_y into constraints on N_{ge}. Using the Arnold-Binford analysis, for any ratio a between the inter-ocular distance and the viewing distance, confidence limits on H_x and H_y can be computed. If the limits are such that with probability q, $|H_x| < H_{xm}$ and $|H_y| < H_{ym}$ then with probability at least q, we have

$$|N_{ge}| < |I * H_{xm} f_H| + |I * H_{ym} f_V|. \tag{37}$$

For the Burt-Julesz psychophysical constraint that $|\nabla H| < 1$, the situation is much the same. Clearly $|H_x| < 1$ and $|H_y| < 1$ so $|N_{ge}|$ is bounded by the equation

$$|N_{ge}| < |I * f_H| + |I * f_V|. \tag{38}$$

In both cases, $|N_{ge}|$ is bounded by the sum of the absolute values of the outputs of two linear filters. The first filter measures the sensitivity of $I * f$ to small horizontal compression and the second measures the sensitivity of $I * f$ to small vertical skew. The computation is simple enough to be performed at every image point without excessive cost.

IV Applications of Distortion Estimate

A. Marr-Hildreth Special Case

An interesting special case of the estimate N_{ge} occurs when the filter used is that of the Marr-Hildreth edge-detector [Marr and Hildreth, 1980]. Then we have:

$$f = \nabla^2 G = \frac{1}{2\pi\sigma^2}\left(\frac{r^2}{\sigma^2} - 2\right) e^{-r^2/2\sigma^2} \tag{39}$$

$$f_H = \frac{1}{2\pi\sigma^2}\left(\frac{-x^2 r^2}{\sigma^4} + \frac{4x^2 + r^2}{\sigma^2} - 2\right) e^{-r^2/2\sigma^2} \tag{40}$$

$$f_V = \frac{1}{2\pi\sigma^2}\left(\frac{-xyr^2}{\sigma^4} + \frac{4xy}{\sigma^2}\right) e^{-r^2/2\sigma^2} \tag{41}$$

where $r^2 = x^2 + y^2$.

The Marr-Poggio theory argues that zeros of $I_1 * f$ reliably correspond to zeros of $I_2 * f$ because they will both be caused by physical edges or surface markings. At a vertical step edge, this view is easily confirmed by the preceeding analysis of geometric distortion. Let $I_1(x, y)$ be a step edge defined as follows:

$$I_1(x, y) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{42}$$

Direct calculation shows that

$$f * I_1 = \frac{-x}{\sqrt{2\pi}\sigma^3} e^{-x^2/2\sigma^2} \tag{43}$$

Since f_V is odd along the y-direction, $f_V * I_1 = 0$ everywhere.

The other component f_H of N_{ge} can be easily evaluated along the edge by noticing that $f_H = \partial x f / \partial x$. Hence

$$I * f_H = I * \frac{\partial}{\partial x} x f = x f \tag{44}$$

Along the edge $x = 0$, so $I * f_H = 0$. Since $I * f_V$ is also zero, the estimate N_{ge} of the geometric distortion noise must vanish along the edge. Thus zero-crossings in $I_1 * \nabla^2 G$ can be trusted to correspond to zero-crossings in $I_2 * \nabla^2 G$. The fact that N_{ge} vanishes along the edge should not be particularly surprising. A step edge under an affine transformation simply changes its orientation. For a radially symmetric filter like $\nabla^2 G$, the change in orientation does not introduce any geometric distortion noise.

Thus, for images which consist entirely of sparse, straight, step-edges, geometric distortion is not a problem for Marr-Poggio based correspondence algorithms except perhaps near edge intersections. Nevertheless, for more general images, geometric distortion can pose severe problems. If the image spectrum is symmetric and representable at least locally as a Gaussian process, then $f * I_1$ and $f_V * I_1$ are independent random variables. Thus points where $f * I = 0$ (edges in the Marr-Hildreth theory) are no less susceptible to geometric distortion due to skew than are other image points. With geometric distortion due to compression, the situation is more favorable for the Marr-Poggio approach. Suppose for example that the image is a stationary first-order Gaussian Markov process with autocorrelation $\exp(-|r/\alpha|)$ [Kass 1984]. Then $f * I$ and $f_H * I$ are somewhat correlated, so points where $f * I = 0$ should on the average have somewhat less geometric distortion due to compression than randomly selected image points. Figure 3 shows the correlation between $f * I$ and $f_H * I$ as a function of the ratio σ/α between the space constants of the filter and the image. As a result of these correlations, the Marr-Poggio approach to stereopsis should be slightly more tolerant of horizontal disparity gradients than vertical ones.

When complex scenes are viewed, $f_V * I$ and $f_H * I$ can take on reasonably large values with high probability even at zero-crossings of $\nabla^2 G * I$. Attempting to match such points on the

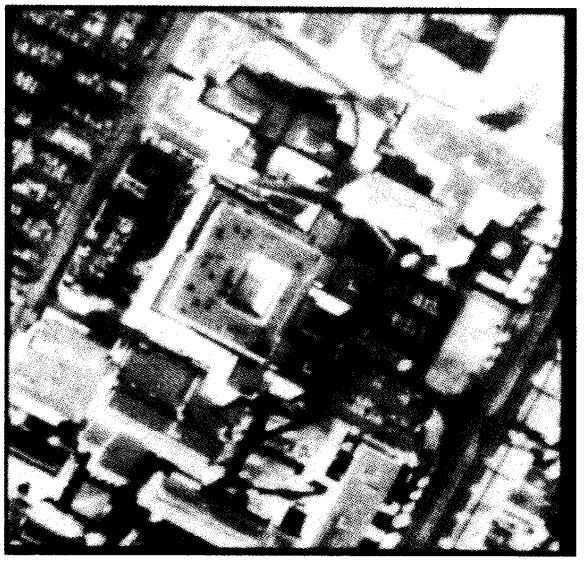

Figure 5: University of British Columbia Acute Care Center from the air (right image)

Figure 6: University of British Columbia Acute Care Center from the air (left image)

basis of the value of $\nabla^2 G * I$ seems imprudent. The ratio of the standard deviation of $f_V * I$ to the standard deviation of $f * I$ is shown in figure 4 again as a function of σ/α based on the first-order Markov model. Assuming a vertical disparity gradient at the Burt-Julesz limit of 1 and a large σ/α ratio, up to 48 percent of the zero-crossings will have geometric distortion in excess of $\sigma(\nabla^2 G * I)/2$ and up to 15 percent will have geometric distortion in excess of $\sigma(\nabla^2 G * I)$. Under such conditions, geometric distortion is clearly a major problem.

Equations 37 and 38 provide a simple method of modifying the Marr-Poggio approach to substantially improve its immunity to geometric distortion. At each zero-crossing in $\nabla^2 G * I$, the bound on $|N_{ge}|$ can be computed from $f_H * I$ and $f_V * I$. If the bound on $|N_{ge}|$ is too large, the zero-crossing should not be matched since any match would be very unreliable. Empirical investigations of this approach are planned.

B. Gaussian Filter Special Case

Another interesting special case for the geometric distortion analysis is when f is a Gaussian filter. Then

$$f = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{45}$$

$$f_H = f + x f_x = \left(1 - \frac{x^2}{\sigma^2}\right) f = -\sigma^2 f_{xx} \tag{46}$$

$$f_V = y f_x = x f_y = \frac{-xy}{\sigma^2} f = -\sigma^2 f_{xy}. \tag{47}$$

Thus the estimate N_{ge} becomes

$$N_{ge} = H_x \sigma^2 I * f_{xx} + H_y \sigma^2 I * f_{xy}. \tag{48}$$

Once again, there is a connection with the Marr-Hildreth theory of edge detection. Not all intensity edges are guaranteed to give rise to zero-crossings in $I * \nabla^2 G$. However, Marr and Hildreth [1980] showed that under a condition known as *linear variation* edges of all orientations cause zero-crossings in $I * \nabla^2 G$ (for a detailed discussion, see Torre and Poggio [1986]). At edge points, the condition of linear variation states roughly that the image intensities are locally linear so the Hessian matrix vanishes. At a zero-crossing in $I * \nabla^2 G$ where the condition of linear variation holds, both $I * f_{xx}$ and $I * f_{xy}$ are zero so the estimate N_{ge} goes to zero. Hence, under the condition of linear variation, edges in the Marr-Hildreth theory are points where $I * f$ is best preserved between views when the filter f is a Gaussian. Note that this does not imply anything about how well $I * \nabla^2 G$ is preserved between views.

C. Application to Kass stereo algorithm

The Kass stereo algorithm [Kass 1983a,b 1984] computes correspondence based on combining indications from a set of nearly independent linear filters at each point. A decision about whether point $\mathbf{p}_1$ in the first image can match $\mathbf{p}_2$ in the second is made by comparing vectors of linear filters at the two points. If the output of any linear filter differs in the two images by more than the threshold for that operator, then the potential match is rejected. Using the geometric distortion estimate N_{ge}, these thresholds can be adjusted dynamically across the image so that the relative weighting of the different filters depends on how invariant they are with respect to changes in geometry.

Figures 5 and 6 show a stereo pair of the University of British Columbia Acute Care Center from the air. Using the Kass stereo algorithm and the geometric distortion estimate N_{ge}, these stereo images were matched. Figure 7 shows the results plotted as contours of constant height above the ground. Note that the buildings are accurately separated from the ground. Matching without the geometric distortion estimate resulted in far noisier results.

V Conclusions

The geometric distortion estimate N_{ge} is applicable to any stereo algorithm which uses a linear filtering step. In particular, this includes coarse-to-fine techniques which blurr the images prior to matching, as well as edge-based techniques which detect edges using linear filters. The estimate makes it possible to identify the points in an image where geometric distortion is

Figure 7: Contours of constant height above the ground for the stereo pair of figures 5 and 6

likely to pose a large problem. The cost of doing so is very minimal because the estimate N_{ge} can be computed trivially from the outputs of two linear filters. Computational experiments with the Kass stereo algorithm have shown that the theoretical advantages of using the distortion estimate are easily attainable.

Acknowledgements

Tomaso Poggio and Keith Nishihara provided important guidance. R.J. Woodham and Eric Grimson made the stereo photographs available.

References

[1] Arnold, R.D. and Binford, T.O. "Geometric constraints in stereo vision," *Proc. SPIE, San Diego* **238**, (1980), 281-292.

[2] Baker, H. H. and Binford, T. O. "Depth from Edge and Intensity Based Stereo," *Seventh International Joint Conference on Artificial Intelligence*, August (1981), 631-636.

[3] Barnard, S. T. and Thompson, W. B. "Disparity analysis of images," *IEEE Pattern Analysis and Machine Intelligence* **PAMI-2**, 4, (1980), 333-340.

[4] Burt, P., Julesz, B., "A disparity gradient limit for binocular fusion," *Science* **208** (1980a) 615-617.

[5] Burt, P., Julesz, B., "Modifications of the classical notion of Panum's fusional area," *Perception*, **9** (1980b) 671-682.

[6] Burt, P., Julesz, B. "The disparity gradient limit for binocular fusion: an answer to J. D. Krol and W. A. van de Grind," *Perception*, **11** (1982) 621-624.

[7] Gennery, D. B. "A system for stereo computer vision with geometric models," *Fifth International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, (1977), 576-582.

[8] Grimson, W.E.L. "A computer implementation of a theory of human stereo vision," *Phil. Trans. Roy. Soc. Lond.*B **292** (1981a), 217-253.

[9] Grimson, W.E.L. **From Images to Surfaces: A computational study of the human early visual system** MIT Press, Cambridge, Ma., (1981b).

[10] Hannah, M.J., "Computer Matching of areas in stereo images," Stanford Artificial Intelligence Laboratory memo, AIM-239, July (1974).

[11] Kass, M. "Computing visual correspondence," *Proc. ARPA Image Understanding Workshop*, Washington D.C., (1983a), 54-60. Also re-printed in Pentland, A. (ed), **From pixels to predicates.** Ablex, Norwood, NJ 1986.

[12] Kass, M. "A computational framework for the visual correspondence problem," *Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, W. Germany, August (1983b), 1043-1045.

[13] Kass, M. "Computing visual correspondence," S.M. Thesis, Department of Computer Science and Electrical Engineering, Massachusetts Institute of Technology, 1984.

[14] Koenderink, J. and van Doorn, A. "Geometry of binocular vision and a model for stereopsis," *Biol. Cybernetics* **21** (1976), 29-35.

[15] Krol, J. D. and Grind, W. A. van de, "Rehabilitation of a classical notion of Panum's fusional area," *Perception* **11** (1982) 621-624.

[16] Marr, D., Hildreth, E. "Theory of edge detection," *Proc. R. Soc. Lond. B*, **207**, (1980) 187-217.

[17] Marr, D. and Poggio, T. "A theory of human stereo vision," *Proc. Roy. Soc. Lond.*B **204** (1979), 301-328. (an earlier version appeared as MIT AI Lab Memo 451, 1977).

[18] Mayhew, J.E.W. and Frisby, J.P. "Psychophysical and computational studies towards a theory of human stereopsis," *Artificial Intelligence* **17** (1981), 349-385.

[19] Medioni, G.G. and Nevatia, R. "Segment-based stereo matching" *Proceedings of the DARPA Image Understanding Workshop*, Washington, D.C., June, (1983) 128-136.

[20] Moravec, H.P. "Towards automatic visual obstacle avoidance," *Fifth International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts (1977), 584.

[21] Nishihara, H.K. "PRISM: A practical realtime imaging stereo matcher," Proceedings, Third International Conference on Robot Vision and Sensory Controls, SPIE Cambridge Symposium on Optical and Electro-Optical Engineering, November (1983) 449.

[22] Ohta, Y. and Kanade, T. "Stereo by intra- and inter-scanline search using dynamic programming," Carnegie-Mellon University Technical Report CMU-CS-83-162, 1983.

[23] Torre, V. and Poggio, T., "On Edge Detection," *IEEE Pattern Analysis and Machine Intelligence*, **PAMI-8**, 2 (1986).

[24] Tyler, C. W., "Stereoscopic vision: cortical limitations and a disparity scaling effect", *Science* **181** (1973) 276-278.

[25] Tyler, C. W. "Spatial limitations of human stereoscopic vision," *Proceedings, SPIE* **120** (1977).