It is important to note that the error caused by center offset is not affected by the amount of image quantization [see (20)]. Therefore, the larger the image quantization is, the larger the 3-D error caused by quantization becomes, making the relative effect of center offset on the overall 3-D error smaller. To demonstrate this phenomenon, the same experiment as above was also done with the image feature coordinate data deliberately perturbed by rounding them to the nearest integer. This is equivalent to an additional random noise of $1/\sqrt{12} = 0.29$ pels in both coordinates. The center offset then increases the relative 3-D measurement error by a factor of only 2.

V. Conclusions

Simple and effective techniques have been presented for calibrating the horizontal scale factor and the image center for 3-D machine vision. Most of the techniques are accurate and efficient, as verified by the results of real experiments. The results indicate that image center calibration is important for high accuracy 3-D measurement applications where the image feature extraction can be done accurately due to the type of image features involved. This is also confirmed by the presented theory.

References

[1] Y. I. Abdel-Aziz and H. M. Karara, "Direct linear transformation into object space coordinates in close-range photogrammetry," in *Proc. Symp. Close-Range Photogrammetry*, Univ. Illinois at Urbana-Champaign, Jan. 26-29, 1971, pp. 1-18.

[2] Y. I. Abdel-Aziz and H. M. Karara, "Photogrammetric potential of non-metric cameras," Univ. Illinois at Urbana-Champaign, Civil Eng. Studies, Photogrammetry Series 36, Mar. 1974.

[3] D. C. Brown, "Close-range camera calibration," *Photogrammetric Eng.*, vol. 37, no. 8, pp. 855-866, 1971.

[4] R. R. Cohen and E. A. Feigenbaum, Eds., *The Handbook of Artificial Intelligence*, Vol. III. Los Altos, CA: Heuris Tech Press, William Kaufmann, 1982.

[5] A. Dainis and M. Juberts, "Accurate remote measurement of robot trajectory motion," in *Proc. Int. Conf. Robotics and Automation*, 1985, pp. 92-99.

[6] R. O. Duda and P. E. Hart, *Pattern Recognition and Scene Analysis*. New York: Wiley, 1973.

[7] W. Faig, "Calibration of close-range photogrammetry systems: Mathematical formulation," *Photogrammetric Eng. Remote Sensing*, vol. 41, no. 12, pp. 1479-1486, 1975.

[8] S. Ganapaphy, "Decomposition of transformation matrices for robot vision," in *Proc. Int. Conf. Robotics and Automation*, 1984, pp. 130-139.

[9] D. B. Gennery, "Stereo-camera calibration," in *Proc. Image Understanding Workshop*, Nov. 1979, pp. 101-108.

[10] E. L. Hall, M. B. K. Tio, C. A. McPherson, and F. A. Sadjadi, "Curved surface measurement and recognition for robot vision, conference record," in *Proc. IEEE Workshop Industrial Applications of Machine Vision*, May 3-5, 1982.

[11] H. Itoh, A. Miyauchi, and S. Ozawa, "Distance measuring method using only simple vision constructed for moving robots," in *Proc. 7th Int. Conf. Pattern Recognition*, vol. 1, Montreal, Canada, July 30-Aug. 2, 1984, p. 192.

[12] A. Isaguirre, P. Pu, and J. Summers, "A new development in camera calibration: Calibrating a pair of mobile cameras," in *Proc. Int. Conf. Robotics and Automation*, 1985, pp. 74-79.

[13] H. M. Karara, Ed., *Handbook of Non-Topographic Photogrammetry*, Amer. Soc. Photogrammetry, 1979.

[14] D. G. Lowe, "Solving for the parameters of object models from image descriptions," in *Proc. Image Understanding Workshop*, Apr. 1980, pp. 121-127.

[15] J. Y. Luh and J. A. Klaasen, "A three-dimensional vision by off-shelf system with multi-cameras," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 35-45, Jan. 1985.

[16] Malhotra, "A computer program for the calibration of close-range cameras," in *Proc. Symp. Cloase Range Photogrammetric Systems*, Urbana, IL, 1971.

[17] *Manual of Photogrammetry*, 4th ed., Amer. Soc. Photogrammetry, 1980.

[18] H. A. Martins, J. R. Birk, and R. B. Kelley, "Camera models based on data from two calibration planes," *Comput. Graphics and Image Processing*, vol. 17, pp. 173-180, 1981.

[19] H. Moravec, *Robot Rover Visual Navigation*. Ann Arbor, MI: UMI Research Press, 1981.

[20] A. Okamoto, "Orientation and construction of models, Part I: The orientation problem in close-range photogrammetry," *Photogrammetric Eng. Remote Sensing*, vol. 47, no. 10, pp. 1437-1454, 1981.

[21] A. Okamoto, "The model construction problem using the collinearity condition," *Photogrammetric Eng. Remote Sensing*, vol. 50, no. 6, pp. 705-711, 1984.

[22] I. Sobel, "On calibrating computer controlled cameras for perceiving 3-D scenes," *Artificial Intell.*, vol. 5, pp. 185-198, 1974.

[23] T. M. Strat, "Recovering the camera parameters from a transformation matrix," in *Proc. DARPA Image Understanding Workshop*, Oct. 1984, pp. 264-271.

[24] I. Sutherland, "Three-dimensional data input by tablet," *Proc. IEEE*, vol. 62, no. 4, pp. 453-461, 1974.

[25] R. Y. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IBM Res. Rep. RC 51342, May 8, 1985; also Best Paper Award for 1986 IEEE Int. Conf. Computer Vision and Pattern Recognition.

[26] R. Y. Tsai, "Review of RAC-based camera calibration," *Vision*, Nov. 1988, to be published.

[27] R. K. Lenz and R. Y. Tsai, "Calibrating a Cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship," *IEEE Trans. Pattern Anal. Machine Intell.*, to be published; see also *Proc. CVPR*, June 5-9, 1988.

[28] K. W. Wong, "Mathematical formulation and digital analysis in close-range photogrammetry," *Photogrammetric Eng. Remote Sensing*, vol. 41, no. 11, pp. 1355-1373, 1975.

[29] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Comput. Graphics Image Processing*, vol. 7, pp. 195-210, 1978.

Singularity Theory and Phantom Edges in Scale Space

JAMES J. CLARK

***Abstract*—The process of detecting edges in a one-dimensional signal by finding the zeros of the second derivative of the signal can be interpreted as the process of detecting the critical points of a general class of contrast functions that are applied to the signal. We show that the second derivative of the contrast function at a critical point is related to the classification of the associated edge as being phantom or authentic. The contrast of authentic edges is shown to decrease with filter scale, while the contrast of phantom edges are shown to increase with scale. It is shown that as the filter scale increases, an authentic edge must either turn into a phantom edge or join with a phantom edge and vanish. The points in the scale space at which these events occur are seen to be the singular points of the contrast function. Using ideas from singularity, or catastrophe, theory one can show that the scale map contours near these singular points are either vertical or parabolic.**

***Index Terms*—Catastrophe theory, edge detection, phantom edges, scale space, singularity theory.**

I. Introduction

A common image representation that is used in computer vision is one based on image edges, loosely defined as a set of points

Manuscript received December 30, 1986; revised February 8, 1988. Recommended for acceptance by C. Brown. This work was supported in part by the Office of Naval Research under Grant N0014-84-K0504 and the Joint Services Electronics Program.

The author is with the Division of Applied Sciences, Harvard University, Cambridge, MA 02138.

IEEE Log Number 8823189.

where the image intensity function changes significantly. An edge based representation is sparser, and hence more computationally efficient than an intensity based representation, but still captures the important features of an image. Multiresolution edge based image representations are based on filtering an image with lowpass filters of various bandwidths, followed by an edge detection process, which localizes the position of changes in the intensity of the filtered images. Multiscale image representations have been used in a number of computer vision algorithms, one of the most notable being the multiresolution stereo vision algorithm developed by Marr and Poggio [13]. One of the first multiresolution edge detection methods was proposed by Rosenfield and Thurston [17]. More recently, Marr and Hildreth [12] have proposed a multiresolution edge detection method that is technically satisfying in many ways. They use a Gaussian lowpass filter to perform the smoothing, and use the zero crossings of the second derivative (Laplacian in two dimensions) to localize the edges. A Gaussian lowpass filter has the impulse response: $g(\vec{x}) = (\sigma\sqrt{2\pi})^{-n}e^{-(|\vec{x}|^2/2\sigma^2)}$ where n is the dimensionality of $\vec{x}$. The factor σ controls the degree of smoothing; a large σ results in more smoothing than a small σ.

One of the problems facing researchers in applying multiresolution image descriptions lies in how the information at different scales is to be integrated effectively. Marr [11] has suggested that edges that spatially coincide at different resolutions are somehow *physically significant*, but no real justification for this was given. Part of the problem with Marr's approach is that it is difficult to associate edges at widely different scales. Clearly, it would be easier to associate edges at different scales if the scales were fairly similar. This leads to the idea of the scale space [19], [22] representation of an image. The scale space is the Cartesian product of the image plane and the $\sigma \geq 0$ ray, and as such involves a continuum of scales. An edge at one scale can be associated with an edge at another scale if one can continuously follow the edge from the first scale to the second scale. The scale space concept is clearly a general one, in that any type of smoothing can be used, and any type of edge detection process can be used. However, the scale space obtained by using a Gaussian filter has special properties which make it the natural choice [2], [12], [23]. Thus many multiscale image representations used in computer vision are of this type, which we will call Gaussian scale space.

The scale space map $\mathfrak{M}$ of a function $i(x)$ is defined as the set of points for which the second derivative of the smoothed function is zero (this is a map in the sense of a contour map). Fig. 1 shows an example of the Gaussian scale space map of a one-dimensional function.

In this correspondence we introduce a new formalism for describing the edge detection process. This formalism is based on the fact that zeroes of the second derivative of a one dimensional function can be associated with the critical points of a contrast function. The singular critical points of this contrast function will be shown to play a critical role in the study of the properties of the scale space map. Using the machinery of catastrophe or singularity theory, we will show that the scale map contours can only be one of two types: either they are curves which extend, without bending back, towards infinite scale, or they are curves which bend over once and then return back to zero scale. All other possible contour shapes are either not allowed or are nongeneric. We also use the contrast function formalism to express how the classification of zero crossings as either authentic or phantom (as developed by Clark [4]) changes as the scale changes. This leads us to the concept of the reduced scale space map, which consists of only those zero crossings which are classified as authentic. The reduced scale space map is seen to consist solely of contours which originate at zero scale and either proceed, without bending back, towards infinite scale, or terminate at a singularity of the contrast function.

II. Phantom Edges in Scale Space

One of the nice properties of the Gaussian scale space map is its well behavedness, as can be seen with a cursory examination of

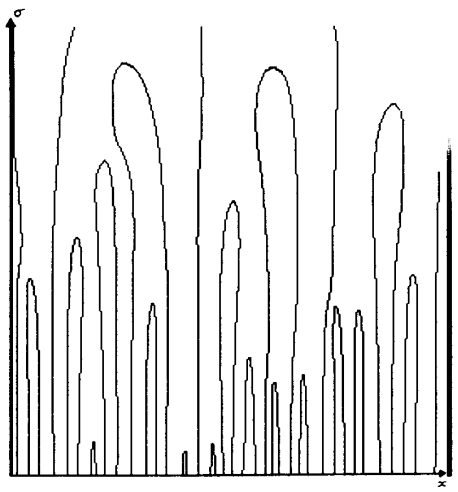

Fig. 1. A one-dimensional Gaussian scale space map.

Fig. 1, which depicts the scale map of a random function. The scale map contours are smooth and start at zero scale and either proceed forever towards infinite scale or gracefully turn back towards zero scale. The scale map contours can be shown to never have minima (i.e., pits or potholes) (proofs of this property can be found in many places [2], [9], [10], [23]; a proof based on the methods described in this correspondence can be found in [5]).

The rather well behaved nature of the Gaussian scale space map is somewhat deceptive, however, in that portions of the scale map may not have any direct relationship to features in the unfiltered image. Consider the step edge shown in Fig. 2(a). Suppose we smooth this signal with a Gaussian lowpass filter to yield the signal shown in Fig. 2(b). In this correspondence we will denote the unfiltered signal by $i(x)$ and the filtered function by $f(x)$. We can localize the position of the "edge" in the smoothed signal by letting the position of the edge be where the smoothed signal has maximum slope, as shown in Fig. 2(c). If we have a step edge in the opposite direction as shown in Fig. 3, it is clear that, to localize the edge in the smoothed signal, we should associate the edge with the point at which the smoothed signal has minimum slope. Thus, in general, we want to associate an edge with those points in the smoothed signal where the slope is a local extremum. We can find these extremal points by determining where the second derivative of the smoothed signal is zero. This is the basis of the zero crossing edge detection method.

However, one must be careful with this zero crossing edge detection method. Consider the "double-step" signal shown in Fig. 4(a). It is easily shown (see for example [16], [18]) that a Gaussian smoothed version of this signal produces not two but three zero crossings of its second derivative. That this is true is easily seen in Fig. 4. The two "authentic" step edges are detected, but, in addition, a third edge is found between them. It is evident from Figs. 2 and 3 that a positive going edge gives rise to a maximum of f_x (where the subscript indicates differentiation with respect to x) that has positive sign (or a "positive maximum"), while a negative going step edges gives rise to a minimum of f_x having a negative sign (or a "negative minimum"). It is obvious that a function must have a minimum between two maxima and vice versa. Therefore if one has two positive going edges, there must be a minimum of f_x between them. If this is not a negative minimum due to a negative going step, then it *must be* a positive minimum of f_x, which cannot be ascribed to any step edge at all. Similarly for two negative going steps. If there is not a positive going step between them, then a phantom edge, which does not correspond to any step in the unfiltered signal, will be inserted between them. Studies performed on

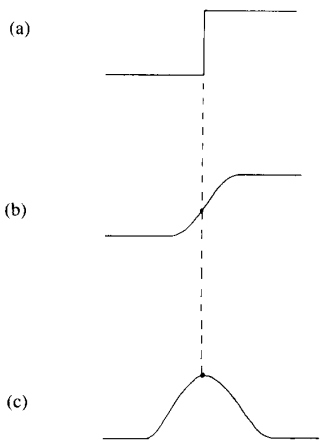

Fig. 2. (a) A step edge. (b) The smoothed step edge. (c) The slope function of the smoothed step edge.

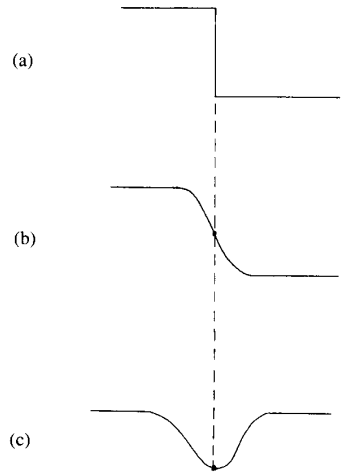

Fig. 3. (a) A step in the other direction. (b) The smoothed step. (c) The slope function of the smoothed step.

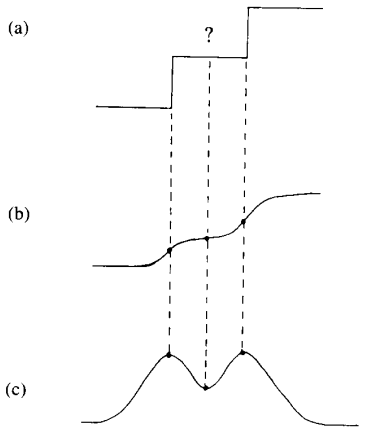

Fig. 4. (a) The double step. (b) The smoothed double step. (c) The slope function of the smoothed double step.

monkeys [16] have shown that there are cells in the monkey's visual cortex which respond to these phantom edges, indicating the presence of zero crossing detectors. A similar situation occurs in the two-dimensional case. That is, not all zero crossing surface patches can be associated with an intensity change in the unfiltered image (for the domain of *piecewise constant, or step* images).

A physically significant scale space representation should consist of only those zero crossings that directly correspond to actual features in the unfiltered image. In order to produce such a representation one must be able to classify zero crossings as to their authenticity. Clark [4] provides a method for performing such a classification. The ideas behind this classification procedure in the one dimensional case are as follows. As we have seen above an authentic step edge gives rise to either a positive maximum or a negative minimum of f_x. It can also be seen that a phantom edge will either correspond to a positive minimum or a negative maximum of f_x. Thus one can use the *classification variable* $\chi = (f_x f_{xxx})$ to determine the authenticity of a potential edge point, $f_{xx} = 0$. If $f_{xx} = 0$ and $\chi > 0$ the edge is classified as a phantom edge. If $f_{xx} = 0$ and $\chi < 0$ the edge is classified as an authentic edge. All other cases we classify as no edge. This latter classification includes the case where $f_x = f_{xx} = 0$ implies that there is no change in $f(x)$ and the case where $f_{xx} = f_{xxx} = 0$ which implies that $f_{xx} = 0$ but does not cross zero (these points correspond to the points in the scale map [22] of $f(x)$ at which the scale map contours are horizontal.)

III. The Reduced Scale Map

The removal of the phantom scale map contours from the scale space map results in what we will refer to as the *reduced scale space map*. An example of the reduced scale space map of a step type signal is shown in Fig. 5(a). The full scale space map of the signal is given in Fig. 5(b). Note that although zero crossing contours are still not created as the scale factor σ is increased, the contours can terminate at a finite and nonzero scale. An interesting property of the reduced scale space map is that its contours never turn over.

It is evident that the phantom zero crossing contours present in the full scale space map may introduce errors in some applications such as multiresolution stereo vision (e.g., the Marr–Poggio algorithm [13]). Errors arise in the case of stereo vision when phantom edges are used as matching primitives because their locations are only loosely coupled to actual physical events (by being constrained to lie between adjacent authentic edges which are coupled to physical events). In such a case the reduced scale map should be used to eliminate the disparity errors caused by the phantom edges.

IV. Singular Points of Contrast Functions

An interesting observation can be made about the classification of the zero crossings of the second derivative of the smoothed function f. Consider a *contrast* function, defined as follows:

Definition 1: A constrast function of a smoothed signal $f(x)$ is a smooth (C^∞) function $C(x) = h(f_x(x))$, where $h(\lambda) > 0$ for $\lambda \neq 0$, $h(0) = 0$, and $h'(\lambda) = 0$ only for $\lambda = 0$ (where $h'(\lambda) = dh/d\lambda$).

The value of $C(x)$ at a zero crossing of f_{xx} gives a measure of the strength of the edge associated with the zero crossing. It is well known that any function that is a convolution of an arbitrary function with a Gaussian kernel is a solution of the following diffusion equation:

$$\sigma f_{xx}(x) = f_\sigma(x) \tag{1}$$

where σ is the space constant of the Gaussian filter. Using this relationship between x and σ derivatives one can show that:

$$C_\sigma(x) = \sigma h'(f_x) f_{xxx}(x). \tag{2}$$

This equation is valid if f_x, f_σ exist and are continuous. The smoothness requirement on $f(x)$ and $C(x)$ ensures this validity. Note that, because of the positivity of h and the fact that $h'(\lambda) = 0$ only at $\lambda = 0$, we have that sgn $(h'(f_x))$ is equal to sgn (f_x). Thus we

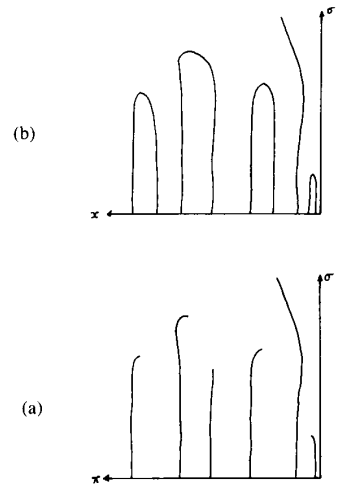

Fig. 5. (a) The reduced scale map of a one-dimensional signal. (b) The full scale space map of the signal.

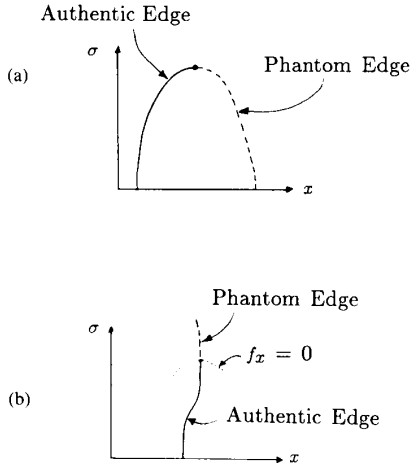

Fig. 6. (a) The pairwise annihilation of an authentic and a phantom zero crossing. (b) The transformation of an authentic zero crossing into a phantom zero crossing.

have, for positive σ, the following relation:

$$\operatorname{sgn}\left(C_\sigma(x)\right) = \operatorname{sgn}\left(f_x(x) f_{xxx}(x)\right) = \operatorname{sgn}(\chi). \tag{3}$$

This suggests the following observation.

Proposition 1: The contrast of a phantom zero crossing increases as σ increases, while the contrast of an authentic zero crossing decreases as σ increases. ■

Proof: Let $\vec{s}$ be the direction along a zero crossing contour in the direction of increasing σ (so that $ds/d\sigma > 0$). Then

$$\frac{\partial C}{\partial s} = \frac{\partial C}{\partial \sigma}\frac{\partial \sigma}{\partial s} + \frac{\partial C}{\partial x}\frac{\partial x}{\partial s} = \frac{\partial C}{\partial \sigma}\frac{\partial \sigma}{\partial s},$$

since $\partial C/\partial x = 0$ at a zero crossing. Thus we see that

$$\operatorname{sgn}\left(\frac{\partial C}{\partial s}\right) = \operatorname{sgn}\left(\frac{\partial C}{\partial \sigma}\right),$$

and hence, from (3) it is seen that the change in contrast as we move along the scale map contour towards higher σ has the same sign as the classification variable χ. ■

This observation implies that one can classify the zero crossings of $f_{xx}(x)$ based on the change in the contrast of the zero crossing with a change in the value of the filter scale constant σ. Let us examine the one-dimensional case in some detail. If we have a step type image (i.e., the image is piecewise constant) the strength (contrast) of a phantom edge will be zero at $\sigma = 0$. As the scale increases from zero the strength of a phantom edge which lies between two authentic edges will increase, while the strength of the two authentic edges will decrease. Now it is evident that the contrast of a phantom edge that lies between two authentic edges must always be less than the contrasts of both of the adjacent authentic edges. Thus as we increase σ the contrast of the phantom zero crossing will increase to a point, and the contrast of the authentic zero crossings will decrease to a point, where the contrast of the authentic and phantom zero crossings are equal. It can be seen that, since a phantom zero crossing corresponds to a minimum in the contrast and the authentic zero crossings corresponds to maxima of the contrast, the positions of the authentic zero crossings and the phantom zero crossings must approach one another as their contrasts become equal. Clearly such a state of affairs can not continue for higher σ, as the contrast of the phantom zero crossing can not increase, and the contrast of the authentic zero crossing decrease, without the contrast of the phantom zero crossing becoming larger than that of the authentic zero crossings, which is impossible. Thus the phantom zero crossing must merge with an authentic zero crossing and the pair disappear. This is illustrated in Fig. 6(a).

It is also possible for an authentic zero crossing to turn into a phantom zero crossing as σ increases. If the contrast of an authentic zero crossing decreases to zero, then it obviously cannot decrease any further. The fingerprint quality [23] of the full scale space map prohibits isolated zero crossing contours from disappearing at finite, nonzero σ values. Therefore this zero crossing must turn into a phantom zero crossing, and its contrast will then increase with σ until it meets up with an authentic zero crossing at which point it will vanish in concert with that authentic zero crossing (the fingerprint quality allows zero crossings to vanish pairwise at finite, nonzero σ values). This is illustrated in Fig. 6(b).

The above discussion of the life and death of authentic and phantom zero crossings in scale space is rather narrative and lacks a solid framework. One can provide a more solid framework by noting that the points at which the classification of a zero crossing contour changes correspond to the degenerate critical points of the contrast function. This observation allows us to utilize the machinery of Morse theory [14] and Catastrophe theory [15], [20] to characterize the types of scale space contours that one typically observes.

The critical points of the contrast function $C(x)$ are the points x at which the first derivative of C with respect to x vanishes. That is:

$$C_x(x) = h'(f_x) f_{xx} = 0. \tag{4}$$

Since $h'(f_x) = 0$ only when $f_x = 0$, the critical points of $C(x)$ can be categorized as one of two types: 1) Those points x where $f_{xx}(x) = 0$, which are the zero crossings which we associate with signal edges, and 2) Those points x for which $f_x(x) = 0$. For the moment it suffices to note that the zero crossings of $f_{xx}(x)$ correspond to a subset of the critical points of any contrast function $C(x)$ that fits the definition given earlier.

Let us, for now, concentrate on those critical points of C for which $C > 0$ (and hence for which $f_x \neq 0$). These critical points are *nondegenerate* (or nonsingular) if $C_{xx} \neq 0$. Note that, since $f_{xx} = 0$ and $\operatorname{sgn}(h'(f_x)) = \operatorname{sgn}(f_x)$, we have that $\operatorname{sgn}[C_{xx}] = \operatorname{sgn}[\chi]$, the sign of the classification variable. Thus if a zero of f_{xx} corresponds to a nondegenerate point of C, then χ is either positive or negative, and the classification of the zero crossing is unambiguous and depends on the concavity of the contrast function. However, if the zero is a degenerate critical point, or *singular point*, of $C(x)$, then $\chi = 0$ and the classification is ambiguous (and the zero is typically not a zero crossing either). Furthermore, the singular points of $C(x)$ are seen to be the points at which the classification of a zero crossing contour changes from authentic to phantom as

one traces it out in the scale space map. Such points are known as *bifurcations* or *catastrophes*. These terms refer to the fact that the system changes state (from authentic to phantom edge) abruptly, or catastrophically, as the control parameter σ is smoothly varied. The set of σ values at which these bifurcations occur is called the *bifurcation set*. We can use the techniques of singularity theory [1] or catastrophe theory [15] to characterize the nature of the scale map contours near such singular points.

The nondegenerate critical points of $C(x)$ at a given scale σ_0 are stable, in that they typically (generically) remain nondegenerate when the function is perturbed slightly, even though their position may shift a little. Functions having *only* nondegenerate critical points are known as *Morse* functions. Functions are generically Morse functions; that is, the collection of all possible smooth Morse functions form an open dense subset of the set of all possible smooth functions. The complementary set, that containing non-Morse functions has measure zero. This means that a contrast function $C(x)$ will almost always contain no singular points, and if it does, a small perturbation of $i(x)$ will cause the singular points to become nondegenerate.

A *family* of functions can, however, generically contain singular points. In particular, the family of functions implicit in the scale space representation of a signal, where the filter scale σ is the parameter indexing the members of the family, typically contains degenerate critical points. These singularities are stable since perturbations of $C(x)$ will only shift the position of the singularities slightly in scale space and will not eliminate them. The scale space family of functions, $C^{(\sigma)}(x)$, is called a one-dimensional *unfolding* of $C(x)$ [15].

The only type of singular point (considering edges for which $C(x) > 0$ only) that is generically observed in any one parameter family of functions, such as the scale space family, is the so-called *fold catastrophe* [15] for which the scale space representation of $C(x)$ can be expanded in a Taylor's series about the singular point as follows:

$$C(y, u) = y^3 + uy + \mathcal{O}(y^4, u^2) \tag{5}$$

where y, u are obtained by a smooth coordinate transformation from x, σ. We have subtracted off the value of $C(x)$ at the singular point for convenience, as it does not affect the differential properties of $C(x)$. The singular point is located in these new coordinates at $(y, u) = (0, 0)$. For one parameter families, such as the scale space representation, the only generic singularity has the local form shown above. The locus of the critical points near such a singular point is shown in Fig. 7. This type of scale space contour is the *only* type of degeneracy observed in a scale space map generically. It can be observed that in the neighborhood of the singular point we have $\text{sgn}(C_{yy}(y)) = \text{sgn}(\chi) = \text{sgn}(y)$. Hence the classification of the edges on either side of the singular point are different. That is, at a singular point of the contrast function an authentic edge and a phantom edge come together. A result of this observation is that contours in the reduced scale space map can never turn over since they must terminate at a singular point.

This is not to say that other types of degeneracies are theoretically impossible only that that typically (generically) never appear. A example of a nongeneric degeneracy that has been described in the scale space literature (see, e.g., [18]) is the double step function, which was shown in the previous section, Fig. 4. At a certain scale it produces a singularity which locally is of the form:

$$C(y, u) = y^4 + uy^2 + \mathcal{O}(y^5, u^2). \tag{6}$$

The locus of critical points about $(y, u) = (0, 0)$ is shown in Fig. 8(a), and is the characteristic pitchfork shape observed in some bifurcation diagrams (such as for the Ginzburg–Landau description of second order theromdynamic phase transitions [6]) and is also known as the *cusp* catastrophe. This singularity, however, is nongeneric in the one parameter case, in that a small perturbation of the form ϵy, for any nonzero ϵ, of $C(y, u)$, will destroy the symmetrical nature of the degeneracy and one will instead have a standard fold type of singularity, as shown in Fig. 8(b). Thus, in practice, a scale space map contour of the form shown in Fig. 8(a) is never seen, but contours that are very close to that form, such as in Fig. 8(b), are often observed.

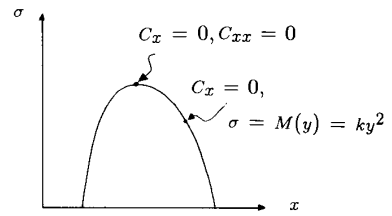

Fig. 7. The form of the scale space map near a fold type degenerate critical point.

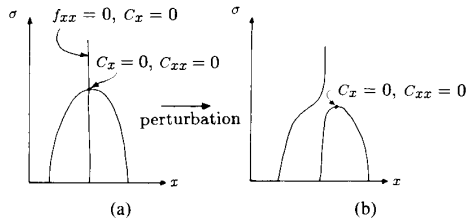

Fig. 8. (a) The form of the scale space map near a cusp type of singularity. (b) A slight perturbation results in the cusp catastrophe becoming a fold catastrophe.

The above analysis assumes that the critical points being examined are of the type for which $C(x) \neq 0$. However, if we have $C(x) = 0$ at the degenerate critical point we have an interesting phenomenon, one which helps explain why authentic zero crossings can generically vanish (by turning into phantom zero crossings) by themselves (i.e., not pairwise) at a nonzero, finite σ value. The case where two critical points of $C(x)$ having $C(x) = 0$ come together and annihilate each other (the analog of the fold catastrophe) can not happen since there must be an maximum of $C(x)$, and hence an authentic zero crossing, between them. Thus the simplest form of a degeneracy when $C = 0$ is given by:

$$C(y, u) = y^4 + uy^2 + \mathcal{O}(y^5, u^2). \tag{7}$$

We saw that for the case of $C(x) \neq 0$, this type of singular point was unstable in that it turned into a fold type of singularity when a small perturbation was added. The question now is, is this type of singularity stable when $C(x) = 0$? The answer is yes, because one cannot transform away the fact that $C(x) \geq 0$ when expanding $C(x, \sigma)$ about the singular point. Thus, in order to maintain the nonnegativity of $C(x)$, any perturbation of the function cannot have any odd order term with order less than 3. Thus the most general perturbation of (7) is:

$$C(y, u) = y^4 + uy^2 + wy^3 + \mathcal{O}(y^5, u^2). \tag{8}$$

It can be shown [20] that there is a smooth change of coordinates which results in:

$$C(z, r) = z^4 + rz^2 + \mathcal{O}(z^5, r^2). \tag{9}$$

Thus the singularity (7) is stable under a small perturbation. Hence the form of the scale space contour shown in Fig. 9 is generic, and they are often observed in practice (see, e.g., Fig. 1). In this case we have that $\text{sgn}(C_{yy}) = \text{sgn}(\chi) = \text{sgn}(12y^2 + u)$. The equation of the scale map contour through the above type of singularity is a vertical line. Since we have $y = 0$ we see that $\text{sgn}(\chi) = \text{sgn}(u)$.

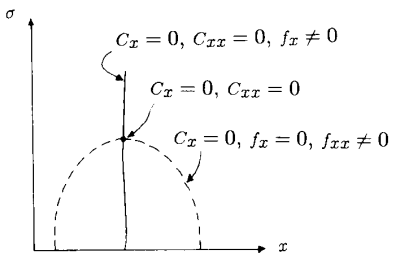

Fig. 9. The form of the scale space map near a singularity of the form (8).

This indicates that the part of the contour with a scale less than that of the singularity corresponds to an authentic edge, while the component above the singularity corresponds to a phantom edge. Thus we see that at this type of singularity an authentic edge changes into a phantom edge as the scale is increased.

V. The Shape of the Scale Map Contours

The analysis of the previous section allows us to determine the shape of the scale map contours near the singular points of the contrast functions. In [5] it is shown that the extrema of the scale map contours (which as we noted earlier are always maxima) occur at the singularities of the contrast function of the type defined by (5) (the *fold* catastrophe). The analysis that follows will thus describe the shape of the scale map contours near these extremal points.

We begin by parameterizing a connected component of the scale space map (i.e., a single contour) as $\sigma = M_i(x)$ whre the index i distinguishes which particular scale space map component is being referred to. Along the curve $\sigma = M(x)$ we have that $C_x(x) = 0$. Let us make the change of coordinates $u = \sigma - \sigma_0$, $y = x - x_0$, where (x_0, σ_0) is the position of the singular point in scale space. Let us first examine the case wherein $C(x) > 0$ at the singular point. For this case we can expand $C(x)$ in a neighborhood about the singular point as follows:

$$C = k_1 + k_2 y^3 + k_3 yu + k_4 u^2 + \text{higher order terms} \quad (10)$$

with k_1, k_2, k_3, and k_4 generically nonzero. Along $M(y)$, the scale map contour, we have that $C_y(y, u) = 0$. In a neighborhood U about the singularity for which the above expansion is valid we have:

$$C_y(y, u) = 3k_2 y^2 + k_3 u. \quad (11)$$

Thus we have:

$$u = M(y) = -3k_2 y^2/k_3. \quad (12)$$

Thus the scale map contour near this type of singular point, which corresponds to an extremum of a scale space map contour, is parabolic.

Similarly, for the other type of generic singularity we have the following expansion:

$$C = k_1 y^4 + k_2 uy^2 + \text{higher order terms}. \quad (13)$$

Along $M(y)$ we have that $C_y(y, u) = 0$ which leads to:

$$C_y(y, u) = 4k_1 y^3 + 2k_2 uy = 0 \quad (14)$$

which gives the two solutions for the locus of critical points $M(y)$:

$$u = M(y) = -2k_1 y^2/k_2 \quad (15)$$

and

$$y = 0. \quad (16)$$

However, of these two solutions, only the solution $y = 0$ corresponds to a scale map contour. To see this, recall that along a scale map contour we have $f_{xx} = 0$. Critical points, however, occur when either $f_{xx} = 0$ or $f_x = 0$. Thus only those critical points for which $f_{xx} = 0$ are taken to be part of a scale map contour. For Gaussian smoothing we can show that:

$$C_{x\sigma} = \sigma\left(h''(f_x) f_{xx} f_{xxx} + h'(f_x) f_{xxxx}\right) \quad (17)$$

and

$$C_{xxx} = \left(h'''(f_x) f_{xx}^3 + h''(f_x) f_{xx} f_{xxx} + h'(f_x) f_{xxxx}\right). \quad (18)$$

Along a scale map contour we must have $f_{xx} = 0$. Hence we have from the above two equations:

$$C_{x\sigma} = \sigma C_{xxx}. \quad (19)$$

Near a singularity of the type defined by (7) we have that

$$C_{yu} = 2k_2 y,\ C_{yyy} = 24k_1 y. \quad (20)$$

These are equal only when $y = 0$ (k_2 is generically $\neq 12k_1$). Hence the solution $y = 0$ corresponds to a scale map contour, while the solution $u = -2k_1 y^2/k_2$ is not a scale map contour, but is where $f_x = 0$. We can conclude that the scale map contour near a singularity of the type defined by (7) is a vertical line.

VI. Extension to Two Dimensions

The analysis presented above is for the one-dimensional case only. The two-dimensional case is naturally of more interest to computer vision researchers (although many computer vision algorithms, such as stereo vision, can be reduced to one dimensional problems). The question then arises whether or not this analysis can be extended to the two-dimensional case. Previous approaches to the analysis of two-dimensional scale space maps include the work of Torre and Poggio [21], and Koenderink [10]. Torre and Poggio show that closed zero crossing contours generically can disappear as scale is increased (an elliptic bifurcation) or split into two closed contours as the scale is increased or decreased (a hyperbolic bifurcation). Koenderink's analysis was done for image intensity functions and did not consider edges (or the scale map). His analysis did, however, describe how image detail is lost (by light or dark blobs merging together) as scale increases. The points in scale at which detail is lost is a bifurcation in the same sense as used by Torre and Poggio. Since Koenderink's analysis is concerned with image intensities and not edges it must be modified if the scale space map is to be investigated. Such a modification will result in an analysis similar to Torre and Poggio's.

Neither Koenderink or Torre and Poggio concern themselves with the distinction between authentic and phantom edges. Thus their analyses cannot provide the whole story about the *reduced* scale space map. We have seen that, in the one-dimensional case, the bifurcations of the reduced scale space map correspond to points where the classification of the edges as authentic or phantom is ambiguous ($\chi = 0$). These points include the bifurcations of Torre and Poggio as a special case, but in addition include those points analogous to the bifurcation depicted in Fig. 9. In order to say anything about the bifurcations of the reduced scale map we must extend the singularity theory approach used in the one-dimensional case to two dimensions. Clark [4] has shown that phantom edges do exist in two dimensions, and it is observed that authentic edges can change into phantom edges as the degree of smoothing of a two-dimensional image increases. In a two-dimensional scale map one obtains contours, which lie along the scale map surfaces, on which the classification of the edge as phantom or authentic is ambiguous (these contours are the places in the scale space map where phantom and authentic edges meet). Thus we could construct a two-dimensional reduced scale space map by deleting the phantom edges from the two-dimensional scale space map.

The contours in the scale space map are presumably analogous to the critical points of the contrast function in the one-dimensional

case. This suggests that a reasonable approach for extension of our one dimensional analysis to two dimensions might be to find an analogous two-dimensional contrast function and find its singularities. This approach has its problems, though, in defining what we mean by a critical point of the contrast function, and by what constitutes a singularity. We know that zero crossings in two dimensions form closed curves. Hence our "critical" points in the two dimensional analysis can not be isolated points but instead must form one dimensional closed loci. This implies that the criticality condition result in a single constraint on the coordinates of the critical point (x_c, y_c) of the form $F(x_c, y_c) = 0$. The usual notion of a critical point of a two-dimensional function, however, yields two constraints: $\nabla f(x_c, y_c) = (0, 0)$. Since this equation imposes two constraints on (x_c, y_c) these critical points will be isolated, and not form one-dimensional contours in general. One must therefore extend this notion of criticality. One can do this by devising some measure of "flatness" or of "stationarity" of a contrast function (such as $|\nabla f|$). Different measures will yield different edge detectors.[1] For example, using as a flatness measure $F(x, y) = \partial C/\partial n = 0$ where C is the first directional derivative of f along the gradient direction (so that $C = \partial f/\partial n$), yields the second directional derivative edge detector [3], [8]. Using a stationarity condition like $F(x, y) = \text{div}(\nabla f) = 0$ yields the Laplacian edge detector [12], where div is the divergence operator.

The idea of a singularity can be similarly extended by thinking of a critical point as degenerate when the function $F(x, y)$ is "flat" in the neighborhood of the critical point (where $F(x_c, y_c) = 0$). A measure of flatness must involve a first order derivative of $F(x, y)$. However, in two dimensions, there are many directions in which a first order derivative can be taken. We must choose some direction, or combination of directions. In the case of the second directional derivative edge detection method it makes sense to take this derivative along the gradient ∇f. Thus an edge ("critical point") may be said to be "singular" if

$$\frac{\partial^3 f}{\partial n^3} = \nabla F(x, y) \cdot \frac{\nabla f}{|\nabla f|} = 0.$$

In the one-dimensional case we had that the classification of an edge as authentic or phantom depended on the sign of $dF(x)/dx = d^2C/dx^2$. Extending this to two dimension we would then expect that we could classify edges in two dimensions by the sign of $\partial F(x, y)/\partial n$. In fact, this is the rule proposed in [4] for classifying edges in two dimensions. One can also show that for the case of Laplacian edge detection that the change of the contrast of an edge with scale obeys the rule that $\text{sgn}(C_\sigma) = \text{sgn}(\partial \nabla^2 f/\partial n)$. Hence in two dimensions, as in one, the contrast of authentic edges decreases as the degree of smoothing is increased.

To use the machinery of singularity theory as we did in the one dimensional case to make statements about the shape of the scale map contours requires that we be more rigorous in defining what we mean by critical and singular points. A more rigorous definition of a critical point is supplied by the theory of differential topology (see for example [7]):

Definition 2: Let X and Y be smooth manifolds and let $f: X \to Y$ be smooth map. Then a point $x \in X$ is a critical point of f if the mapping $df|_x: T_x(X) \to T_y(Y)$ is not onto (not surjective), where $T_x(X)$ is the tangent space to the manifold X at the point x, and $T_y(Y)$ is the tangent space to the manifold Y at the point y where $y = f(x)$.

[1]It is for this reason that we use a general contrast function in the one dimensional case instead of just $f_x(x)$. While it is true that the results in one dimension are independent of what contrast function is used, and that $f_x(x)$ may just as well be used throughout, it is not true in two dimensions. The type of edge detector you get depends on the form of the contrast function used. So to maintain generality we use as general contrast function, even in the one-dimensional case.

A point x is a degenerate critical point or singular point if x is a critical point of $df|_x$ as well as of f (where df in this case can be thought of as the Jacobian matrix of f with respect to x). The measures of criticality that we have introduced above to generate our edge detection methods (second directional derivative and Laplacian) cannot be put into the above definition of critical points. One can, however, generalize the notion of criticality through the idea of (non)-transversal mappings and manifolds (see [15], [7] for discussions of transversality theory). The use of transversality theory in investigating the shape of the two-dimensional scale map is beyond the scope of this correspondence, however, and we will not attempt it here.

VII. Summary

The results presented above contain a number of results of interest to computer vision researchers. First, the singular points of any contrast function as defined in this paper are the points in scale space at which authentic zero crossings vanish (either by turning into phantom zero crossings, or by performing an *electron-positron* like annihilation with a phantom zero crossing). These are significant events as they localize the scale at which detail in the image is lost as the image is smoothed.

Secondly, the manner in which the authentic zero crossings vanish is typically limited to only the two cases shown in Figs. 8 and 9. The shape of the scale space map contours near a singular point of the form shown in Fig. 6(a) is seen to be parabolic while the scale map contours that pass through a singular point of the form shown in Fig. 6(b) are vertical near the singularity. These findings strengthen the notion of the well-behavedness of the Gaussian scale space map by showing that the shape of the scale map contours are highly constrained near the singularities of the contrast function.

With the aid of the contrast function formalism we were able to demonstrate the intriguing property that the strength (contrast) of authentic zero crossings (for the case of Gaussian smoothing) *decreases* as the scale increases, while the strength of phantom zero crossings *increases* as the scale increases.

We have introduced in this paper a modified scale space signal representation, the reduced scale space map, which is composed of the loci of the authentic zero crossings of the second derivative of the smoothed image. The reduced scale map contours are seen to be curves which start at the $\sigma = 0$ line and terminate at a singularity of a contrast function of the smoothed signal. If one is using a multiresolution representation for a vision algorithm such as in image matching for stereopsis or motion analysis, one should use a reduced scale space representation so that errors due to the nonphysicality of the phantom edges do not arise.

We briefly discussed the two-dimensional case where it was pointed out that many of the interesting aspects of the one dimensional analysis carry over to two dimensions. For example, phantom edges do exist in two dimensions, and the contrast of authentic edges in the case of the $\nabla^2 G$ edge detector decreases as the level of smoothing is increased. We were not able to obtain results as to the shape of the scale map surfaces in the two-dimensional case due to the complexities involved in extending the notion of critical and singular points from one to two dimensions. The use of transversality as a generalization of regularity (the opposite of criticality) promises a way of obtaining the scale map shape information.

References

[1] V. I. Arnold, *Singularity Theory*. Cambridge, MA: Cambridge University Press, 1981.

[2] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale space filtering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 1, pp. 26–33, 1986.

[3] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679–689, 1986.

[4] J. J. Clark, "Authenticating edges produced by zero crossing algorithms," Harvard Robotics Laboratory Tech. Rep., 1986; also *IEEE Trans. Pattern Anal. Machine Intell.*, to be published.
[5] ——, "Singularities of contrast functions in scale space," in *Proc. 1st Int. Conf. Computer Vision*, London, England, 1987, pp. 491–496.
[6] R. Gilmore, *Catastrophe Theory for Scientists and Engineers*. New York: Wiley-Interscience, 1981.
[7] V. Guillemin and A. Pollack, *Differential Topology*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[8] R. M. Haralick, "Digital step edges from zero crossings of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 1, pp. 58–68, 1984.
[9] R. A. Hummel and B. C. Gidas, "Zero crossings and the heat equation," Courant Inst. Math. Sci., Comput. Sci. Division, New York Univ., Tech. Rep. 111, 1984.
[10] J. J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.
[11] D. Marr, *Vision*. San Francisco, CA: W. H. Freeman, 1982.
[12] D. Marr and E. C. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London*, series B, vol. 207, pp. 187–217, 1980.
[13] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Roy. Soc. London*, series B, vol. 207, pp. 187–217, 1980.
[14] J. Milnor, *Morse Theory*, Annals of Mathematics Studies Number 51. Princeton, NJ: Princeton University, 1963.
[15] T. Poston and I. Stewart, *Catastrophe Theory and its Applications*. London: Pitman, 1978.
[16] J. Richter and S. Ullman, "Non-linearities in cortical simple cells and the possible detection of zero crossings," *Biol. Cybern.*, vol. 53, pp. 195–202, 1986.
[17] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. 20, pp. 562–569, 1971.
[18] M. Shah, A. Sood, and R. Jain, "Pulse and staircase edge models," *Comput. Vision, Graphics, Image Processing*, vol. 34, pp. 321–343.
[19] J. L. Stansfield, "Conclusions from the commodity expert project," Massachusetts Inst. Technol., AI Lab. Memo 601, 1980.
[20] R. Thom, *Structural Stability and Morphogenesis* (translated D. H. Fowler). New York: Benjamin-Addison Wesley, 1975.
[21] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 2, pp. 147–163, 1986.
[22] A. Witkin, "Scale space filtering," in *Proc. 8th Int. Joint Conf. Artificial Intelligence*, Karlsruhe, 1983, pp. 1019–1022.
[23] A. L. Yuille and T. Poggio, "Scaling theorems for zero crossings," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 1, pp. 15–25, 1986.

Sensor Roll Angle Error for a Mobile Robot Using a Navigation Line

K. C. DRAKE, E. S. McVEY, AND R. M. IÑIGO

***Abstract*—An investigation of the effects of various error components in a mobile robot system which uses a navigation line is available in the literature. An analysis of the effects of these error components on the navigation line width determination and obstacle range determination resulted in equations which may be used to predict the error resulting from changes in various system parameters. One component of error which has not been investigated is the change in the rotation angle of the robot sensor about its optical axis, called the roll angle. The effects of changes in the sensor roll angle on both navigation line width determination and obstacle range determination are presented to complete this phase of robot error analysis. Numerical examples showing the characteristics of the derived error equations are also given.**

***Index Terms*—Autonomous navigation, discrete sensor, machine vision, navigation line, sensor roll angle, visual navigation.**

Manuscript received October 14, 1985; revised November 3, 1987. Recommended for acceptance by R. Bajcsy. This work was supported by the National Science Foundation under Grant ECS-82-15443.

K. C. Drake is with the U.S. Air Force Flight Dynamics, Wright-Patterson AFB, OH.
E. S. McVey and R. M. Iñigo are with the School of Engineering and Applied Sciences, University of Virginia, Charlottesville, VA 22901.

IEEE Log Number 8821775.

I. Introduction

The use of a contrasting line on the pathway floor of a mobile robot has been proposed as a navigational guide [1]. A system which uses a single discrete sensor and a continuous, straight-edged line was proposed in [2]. This analytical investigation revealed the effects of variations in system parameters in determining the best width of the navigation line. Equations were developed which yield the percent error caused by each component of system error. These components of error included sensor resolution error, sensor height error, sensor pan angle error, and sensor tilt angle error. These equations enable the system designer to predict system error given values for system parameters and the variations that may be expected in these parameters. Line width determination seems an appropriate basis for this error investigation because the width of the line as it appears in the image plane is the fundamental characteristic of the robot's view of its environment. The width of the line as it appears in the sensor's image plane is the basis for the width ranging method to be discussed. Determining the line width also leads to segmentation of the entire line in the image plane, therefore aiding in the navigation of the robot (i.e., maintaining the lateral offset of the robot from the navigation line). Variations in system parameters which cause the width of the line in the image to change are ultimately detrimental to the entire system performance.

Two methods of determining the range to objects in the sensor's field of view were developed in [3]. A similar error analysis was performed which gave the expected error in determining the range to objects using each ranging method. The analytical results given in [2] and [3] were tested extensively [4]. The methods developed in [2] and [3] were found to be useful from two standpoints. They were found to be both accurate (actual error was often less than was predicted by the error analysis equations) and practical from a real time implemenation point of view.

Fig. 1 represents the system geometry considered in [2]–[4] where θ is the pan angle and ϕ is the tilt angle of the sensor. Two coordinate systems are shown: the unprimed coordinates (x, y, z) representing the global coordinate system and the primed coordinates (x', y') representing the coordinate system of the image plane.

One component of system geometry error not considered in [2]–[4] is the roll angle of the sensor α or the rotation of the sensor about its own optical axis, as shown in Fig. 1. The effect of changes in the roll angle of the sensor on determining the width of the line and on range determination will be considered in this paper. This will complete the error analysis which is of fundamental importance in the design of this class of robots.

II. Direct Perspective Transformation

As derived in [5], the direct perspective transformation between the global coordinate system and the image plane coordinate sys-