Analysis of Disparity Gradient Based Cooperative Stereo

Ze-Nian Li, *Member, IEEE,* and Gongzhu Hu, *Member, IEEE*

Abstract—This paper argues that the disparity gradient subsumes various constraints for stereo matching, and can thus be used as the basis of a unified cooperative stereo algorithm. Traditionally, selection of the neighborhood support function (NSF) in cooperative stereo was left as a heuristic exercise. We present an analysis and evaluation of three families of NSF's based on the disparity gradient. It is shown that an exponential decay function with a conveniently selectable parameter is well behaved in that it yields the least error, converges steadily, and produces correctly located weak-winners. The discovery of the well-behaved function facilitates the success of the disparity gradient based approach. It is suggested that this function will help a two-pass algorithm in resolving the dilemma of surface continuity and discontinuity/occlusion. In our experiments, the unified cooperative stereo-matching algorithm is tested on random-dot stereograms containing opaque and transparent surfaces. It is also shown to be applicable to both area matching and contour matching in real-world images.

Fig. 1. Defining disparity gradient in stereo vision.

I. INTRODUCTION

MARR and Poggio [1] presented a cooperative stereo correspondence algorithm, which is essentially an iterative relaxation process where *uniqueness* and *continuity* constraints are employed. Their seminal experiment result from random dot stereograms (RDS's) remarkably resembles human stereopsis. An important advantage of the cooperative stereo is its suitability for massively parallel implementations, since only local interactions between nodes in a parallel network are required. The Drumheller–Poggio algorithm [2] is an attempt for parallel stereo. One of the major criticisms to the Marr–Poggio algorithm has been its reliance on the assumption of surface continuity. As a result, the constraint of *figural continuity* [3] is often viewed as a more acceptable alternative, especially in the domain of feature (contour) based stereo.

Burt and Julesz [4] pointed out that *disparity gradient* dictates binocular fusion when several objects occur near one another in the visual field. They argue that there is a disparity gradient limit of approximately 1 for most human subjects in their experiments. Furthermore, order reversal occurs in two potential matches when the disparity gradient is allowed to be ≥ 2. There have been numerous articles exploiting the
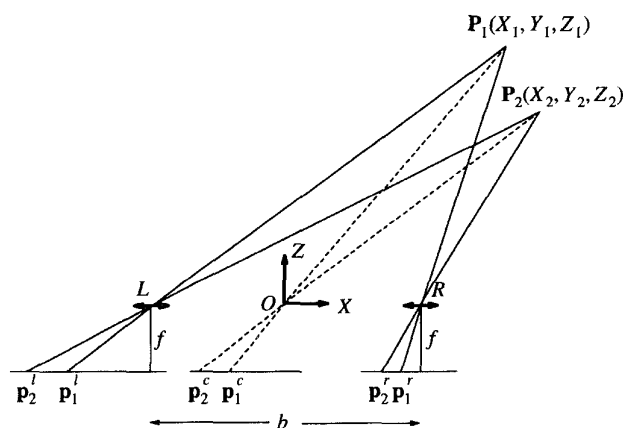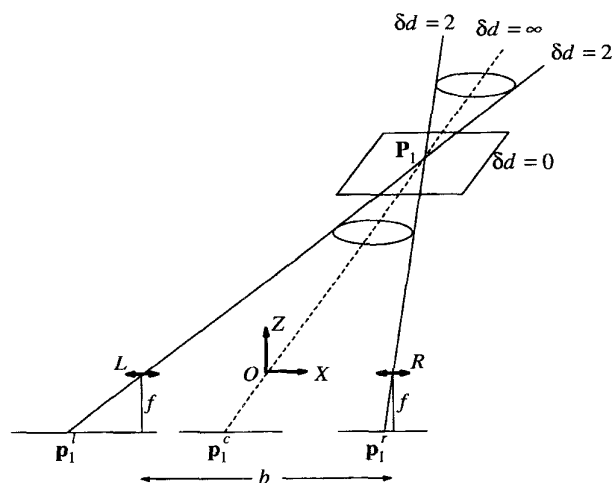

Fig. 2. Forbidden zone of point P_1 in 3-D scene.

constraints of disparity gradient limit and nonreversal ordering [5]–[7]. Pollard *et al.* [7] pointed out the intrinsic relationship between the disparity gradient and the surface orientation and depth in three-dimensional (3-D) scenes. The PMF algorithm by Pollard, Mayhew, and Frisby [6], [7] was one of the early stereo-matching algorithms that extended the Marr–Poggio algorithm by introducing the disparity gradient. Based on the observation [4] on human vision system, a disparity gradient limit of 1 was imposed. Only neighboring nodes with their disparity gradient within the limit were allowed to contribute

(and contribute homogeneously), although it was recognized [7] that the disparity gradient might vary in the range of 0 to 2. More recently, Stewart [8] proposed a new support equation for stereo matching based on the analysis of the probability of disparity changes and depth changes.

Although various constraints for stereo matching have been widely explored and proven powerful, most of them have been used in an ad hoc manner. We argue that, to various degrees, the disparity gradient subsumes the constraints of continuity, figural continuity, uniqueness, nonreversal ordering, and disparity-gradient limit. Therefore, a unified algorithm for the cooperative stereo matching based on the disparity gradient can be developed. What is needed is a systematic way of determining the neighbor support as a function of the disparity gradient.

This paper presents an analysis and evaluation of three families of neighborhood support functions (NSF's) based on the disparity gradient. For the purpose of comparison, the first family includes several ad hoc functions. The second family is adapted from the support function commonly used in simulated annealing [9]. The third family is a group of exponential decay functions. It will be shown that the exponential decay function with a single conveniently selectable parameter yields the least error.

Little and Gillett [10] studied the issue of surface discontinuity in the context of occlusion. They observed that the pixels near the occlusion boundaries receive conflicting (positive and negative) supports from neighbors on two sides. As a result, "weak-winners" are generated at locations of surface discontinuity. The relaxation process can be greatly improved if there is some knowledge of the potential surface discontinuity. They suggested that a two-pass algorithm could be designed to alleviate the dilemma of surface continuity and discontinuity/occlusion in cooperative stereo. During the second pass, the neighborhood would be carefully chosen as not to cross the surface discontinuity. The crucial point is thereby a reliable identification of the potential surface discontinuities after the first pass. Because the relaxation process with our exponential decay function converges steadily, the "weak-winners" will usually gain their strength at a slower pace. They can readily be identified and be exploited by the two-pass matching algorithm.

Transparent surfaces in RDS's are often used to illustrate the effectiveness of the cooperative stereo algorithm [11], [12]. In a way, the transparent situation is an extreme case of occlusion—the common assumption of smooth and opaque surfaces is no longer valid and occlusions can occur everywhere in the image. Prazdny [11] developed an algorithm that employs the *coherence principle*. Global support is sought according to a disparity similarity function. Unlike most conventional relaxation approaches, the algorithm is noniterative and excludes inhibitory inputs. It is reported to have nearly 100% accuracy for opaque surfaces. When transparent surfaces are present, its accuracy drops to about 75%. Szeliski and Hinton [12] present a similar support function that is the difference of two heat equations. The function can be implemented in a locally connected parallel network. The resulting accuracy is almost as good as Prazdny's. Our analysis shows that, with a more

(a)
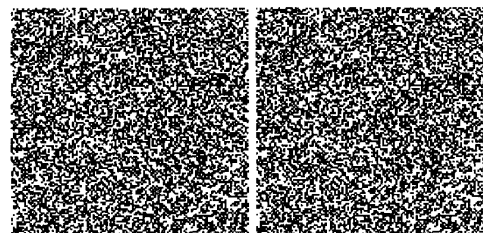
(b)
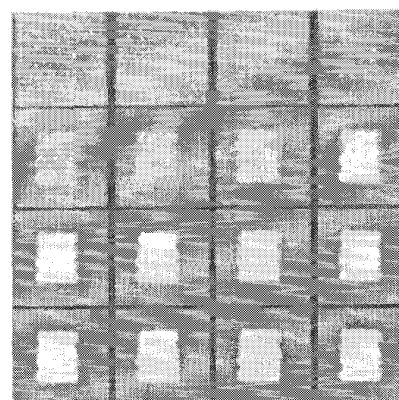
(c)
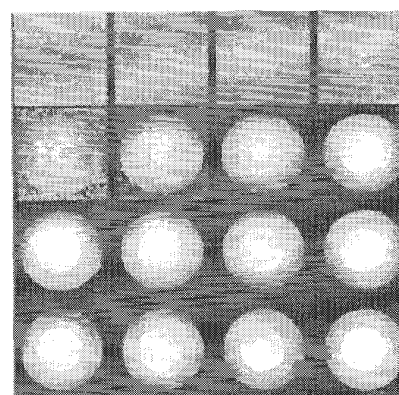
(d)

Fig. 3. (a) Random-dot stereogram (64 × 64) for a floating rectangle. Disparity value for the rectangle is $d_1 = 6$, for the background is $d_2 = 2$. (b) Random-dot stereogram (128 × 128) for a hemisphere on the background. Disparity value for the top of the sphere is 9, for the background is 1. (c) Iterative relaxation result for (a). Grey-level is used to represent disparity values. From iteration 0 to 45, every third iteration result is shown in order (row-major). (d) Iterative relaxation result for (b).

$$f_2(\delta d) = \frac{2}{1 + e^{(\delta d - p)/T}} - 1.0$$

Fig. 4. Several $f_2(\delta d)$ functions: (a) f_{2a} ($p = 1.0, T = 0.1$), dotted; (b) f_{2b} ($p = 1.0, T = 0.3$), dashed; (c) f_{2d} ($p = 1.8, T = 0.5$), solid.
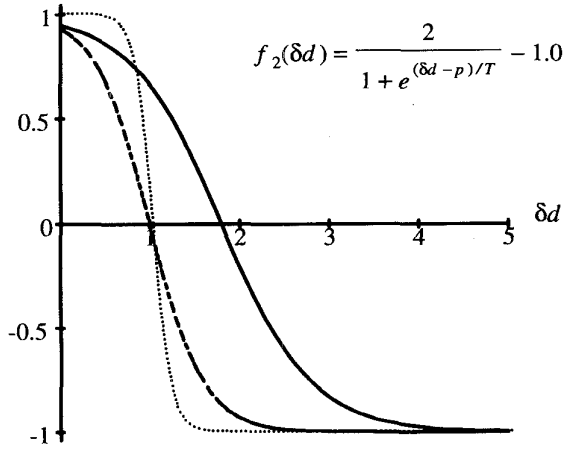
$$f_3(\delta d) = 2 e^{-\delta d /T} - 1.0$$

$T = 3.0$
$T = 1.5$
$T = 0.5$
$T = 0.1$

Fig. 5. Function $f_3(\delta d)$: the curves depict the function when $T = 0.1, 0.5, 1.5,$ and 3.0 respectively.

careful choice of the NSF and its parameter, the accuracy of matching in the presence of transparent surfaces can be maintained at a level comparable to that of opaque surfaces.

In recent years, there has seen a surge of interest in single-image random-dot stereograms (SIRD's) [13], [14]. The SIRD's are based on the same principle of stereopsis as discussed in this paper, except the left and right images of the RDS are now cleverly merged into a single image [14]. The reduced separation of the two images of the RDS makes fusion (viewing) of the stereogram much easier, so a vast majority (over 90%) of viewers are now able to appreciate the 3-D visual effect with little practice. The algorithm presented in this paper is readily applicable to the SIRD's, which has found its increasing use in scientific visualization and multimedia applications.

The organization of this paper is as follows. Section II reviews basics of the disparity gradient. Section III introduces the disparity gradient based cooperative stereo algorithm. Section IV describes the search for a good NSF, and the analysis and evaluation using RDS's. Section V presents results from RDS's containing transparent surfaces. Section VI describes experimental results using real-world images. Section VII concludes the paper.

II. BASICS OF DISPARITY GRADIENT

A. Definition

Fig. 1 depicts the camera geometry for stereo vision where the camera optical axes are parallel to each other and perpendicular to the baseline connecting the two cameras L and R. For a point $P(X, Y, Z)$ in the 3-D scene, its projections onto the left image and the right image are $p^l(x^l, y^l)$ and $p^r(x^r, y^r)$. Because of this simple camera geometry, $y^l = y^r$ and the disparity d is inversely proportional to the depth Z.

$$d = x^l - x^r = bf/Z \qquad (1)$$

where f is the focal length of the camera lens and b is the separation of the two cameras.
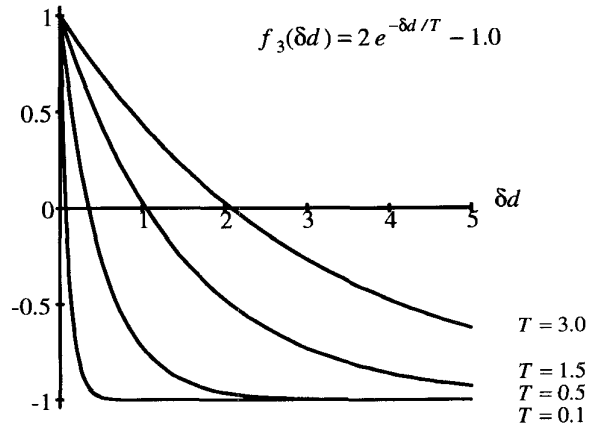
TABLE I
THREE FAMILIES OF NSF's

Function	Parameters
f_{1a}	$p = 1.5, \quad C_1 = 1.0, \quad C_2 = -1.0$
f_{1b}	$p = 1.5, \quad C_1 = 0.0, \quad C_2 = -1.0$
f_{1c}	$p = 1.5, \quad C_1 = 0.0, \quad C_2 = -0.5$
f_{1d}	$p = 2.0, \quad C_1 = 0.0, \quad C_2 = -1.0$
f_{2a}	$p = 1.0, \quad T = 0.1$
f_{2b}	$p = 1.0, \quad T = 0.3$
f_{2c}	$p = 1.5, \quad T = 0.3$
f_{2d}	$p = 1.8, \quad T = 0.5$
$f_{3(0.1)}$	$T = 0.1$
$f_{3(0.5)}$	$T = 0.5$
$\cdots$	$\cdots$
$f_{3(4.0)}$	$T = 4.0$

Given two points $P_1(X, Y, Z)$ and $P_2(X, Y, Z)$, their *disparity gradient* (δd) can be defined as $\delta d = difference\_in\_disparities/cyclopean\_separation$, where $cyclopean\_separation$ is the average distance between $p_1^l, p_2^l,$ and p_1^r, p_2^r [7].

$$\delta d = 2 \times |(x_2^l - x_2^r) - (x_1^l - x_1^r)|/\|(p_2^l - p_1^l) + (p_2^r - p_1^r)\| \qquad (2)$$

$$= 2 \times |(x_2^l - x_1^l) - (x_2^r - x_1^r)|/\|(p_2^l - p_1^l) + (p_2^r - p_1^r)\| \qquad (3)$$

where $\| \cdot \|$ denotes the vector norm. Note, from its definition δd is always a nonnegative number.

Suppose a virtual camera is placed in the middle of the cameras L and R, i.e., at the position of the origin. Since $p_1^c = (p_1^l + p_1^r)/2$ and $p_2^c = (p_2^l + p_2^r)/2$, it follows that

$$\delta d = |(x_2^l - x_1^l) - (x_2^r - x_1^r)|/\|(p_2^c - p_1^c)\| \qquad (4)$$

or

$$\delta d = |(d_2 - d_1)|/\|(p_2^c - p_1^c)\|. \qquad (5)$$

Equation (5) offers a clearer δd definition based on the view of the middle virtual camera.

(a) f_{1a} (b) f_{1b} (c) f_{1c} (d) f_{1d}

(e) f_{2a} (f) f_{2b} (g) f_{2c} (h) f_{2d}

(i) $f_{3(0.1)}$ (j) $f_{3(0.5)}$ (k) $f_{3(1.0)}$ (l) $f_{3(1.5)}$ (m) $f_{3(2.0)}$ (n) $f_{3(3.5)}$
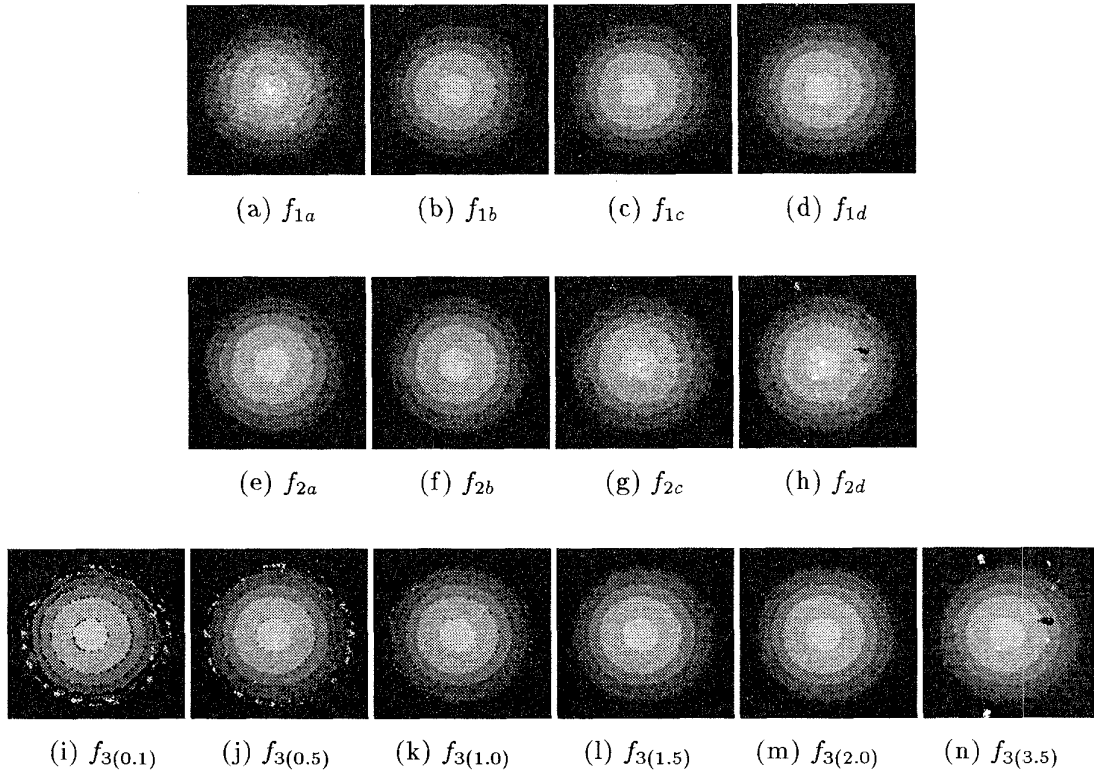
Fig. 6. Grey-level coded disparity maps from the cooperative stereo-matching algorithm. (a)–(n) Results from three families of NSF's.

B. Relationship Between δd and Surface Orientation

It was pointed out [4] that order reversal occurs when $\delta d \geq 2$. In Fig. 2, imagine that the current fixation point is P_1 and there is another point P_2 somewhere in the scene. From (3), it can be readily shown that the order reversal occurs when P_2 is located in one of the two oblique circular cones tipped at point P_1. Assuming surfaces of objects in the scene are opaque, it can be said that the point P_1 creates a cone-shaped "forbidden zone" in the 3-D scene (XYZ space). The center line of the cone is the line of the sight, i.e., the line that connects O and P_1.

Similarly, for each potential match (x_1, y_1, d_1) in the xyd space as used by the Marr–Poggio algorithm [1], two right circular cones tipped at (x_1, y_1, d_1) also form a "forbidden zone" where the slope of the cone surface is 2 with respect to the xy plane.

It is important to remember that, in general, the shape of the forbidden zones in the xyd space and in the XYZ space are different. Nevertheless, these two "forbidden zones" are clearly related. The mapping between these two cones reflects an important relationship between the disparity gradient and the slope of the surface in the 3-D scene. From Fig. 1 and (4), it can be shown that

$$\delta d = b \cdot |Z_2 - Z_1| / \sqrt{(Z_1 X_2 - Z_2 X_1)^2 + (Y_2 Z_1 - Y_1 Z_2)^2}.$$
(6)

When $Y_1 = Y_2 = 0$, (6) degenerates to

$$\delta d = b \cdot |Z_2 - Z_1| / |Z_1 X_2 - Z_2 X_1| \approx \frac{b}{Z_1} |\tan \theta| \quad (7)$$

where $\tan \theta$ is the slope of the small surface patch connecting P_1 and P_2.

As shown in Fig. 2, P_2 on the $Z = Z_1$ plane yields $\delta d = 0$ (as $Z_1 = Z_2$). Also, other circular cone surfaces (tipped at P_1) between the forbidden zone and the $Z = Z_1$ plane represent the locations where $\delta d = c(0 < c < 2)$. In this sense, we have obtained the "δd map."

The δd map can be used to reveal various stereo-matching constraints. It is thus used as the basis of the unified cooperative stereo algorithm. A brief summary follows.

- $\delta d > 2$ — violation of nonreversal order constraint
- $\delta d = 2$ — violation of uniqueness constraint
- $\delta d < 1.1$ or 1.2 — disparity-gradient limit
- $\delta d \ll 1$ — continuity and figural continuity constraints

Finally, it is worthwhile to visualize what 3-D surface would yield $\delta d = 1$, since the earlier psychological study [4] seemed to suggest this to be approximately the limit of the disparity gradient for binocular fusion. We shall examine a simple situation in which $Y_1 = Y_2 = 0$. If $\delta d = 1, b = 2.5$ inches and $Z_1 = 25$ inches, then $|\tan \theta| \approx 10$, which indicates that the small surface patch is $84.3°$ to the X-axis—a surface almost along the line of sight.

C. Terrain Map

Since δd is defined from the view of the middle virtual camera, an ideal terrain map can be built from this (cyclopean) view. That is why the origin of the XYZ space is chosen in the middle, not at the right or left camera position. The x-

coordinates of matched features from the left and right images can simply be averaged to produce the mid-x-coordinate, and the disparity (or depth) value would be registered there. This terrain map is more natural to human observers and more useful for robotic navigators.

III. COOPERATIVE STEREO ALGORITHM BASED ON DISPARITY GRADIENT

A. The Original Marr–Poggio Algorithm

The framework of cooperative algorithms for computation of stereo disparity is defined by Marr and Poggio [1]. To eliminate false targets in the correspondence process, a network of several layers of neurons is constructed. Each neuron corresponds to a disparity value d at location (x, y), and has connections to its neighbors in the xyd space. The original Marr–Poggio algorithm suggests the examination of a neighborhood for each potential match in the xyd space in an iterative relaxation process. Based on the continuity constraint, neighboring neurons provide unit-positive (excitatory) support if they have the same d and their own strength is nonzero. Based on the uniqueness constraint, neurons associated with different disparity values at the same location (x, y) inhibit each other by providing $-\epsilon$ (inhibitory) support. In terms of δd, the former deals with $\delta d = 0$ and the latter $\delta d = 2$.

B. A Unified Disparity Gradient Based Algorithm

The disparity gradient will be used as the basis of our stereo-matching algorithm. A larger range of continuous δd values will be taken into account. The support from neighboring nodes is defined as a function of the disparity gradient $f(\delta d)$, whose value can be either positive or negative.

An initial matching of two stereo images will likely produce multiple candidate d values at (x, y) (on the virtual middle xy-image plane). Each candidate match can be assigned with some initial strength S^0 according to the quality of the match. During the iterative relaxation process, the amount of support

that node $r(x, y, d)$ receives from its neighbor $r'(x', y', d')$ in the xyd space is inversely proportional to the distance between (x, y) and (x', y') on the image plane (denoted as $r_{xy} = \sqrt{(x' - x)^2 + (y' - y)^2}$). The support is proportional to the neighbor's strength and the $f(\delta d)$ value (where $\delta d = |d' - d|/r_{xy}$).

A special treatment is given to the nodes where $r_{xy} = 0$. Since $\delta d = |d' - d|/0 = \infty$ and $f(\infty)$ is usually assigned to -1, the support will be proportional to $\eta \times (-1)$, where η is an adjustable coefficient whose value is >1. Moreover, as in [7], "winner-take-all" is implemented to enforce the uniqueness and convergence of the network.

The procedure found at the bottom of this page describes the unified algorithm.

RDS is used for testing the cooperative algorithm as in [1]. Fig. 3 shows the result of a floating rectangle and a hemisphere. The supporting neighborhood is chosen as a 7 × 7 × 7 cube (i.e., $a = b = c = 3$), $\eta = 8, S^{(0)} = 128$, and max_$strength = 255$. The relaxation process is iterated for 45 times.

IV. SEARCH FOR A WELL-BEHAVED NEIGHBORHOOD SUPPORT FUNCTION (NSF)

The behavior of the NSF is crucial to the performance and convergence of the cooperative stereo-matching process. As in [1], [7], and [8], the value of the NSF is usually a function of δd. The NSF by Marr and Poggio [1] has a value of 1 when $\delta d = 0$, and $-\epsilon$ when $\delta d = 2$. The support function by Pollard et al. [7] has a value of 1 when δd is less than a chosen limit and zero otherwise. These functions are basically step functions with some heuristic control parameters.

Determining values of the several parameters for the ad hoc functions is not an easy task. For example, some seemingly reasonable NSF's actually produce poor matching result. One such example is $f(\delta d) = 1$ if $\delta d \leq 1, -1$ otherwise. The problem is the excessive support assigned to the potential matches while δd is close or equal to 1.

PROCEDURE cooperative-stereo;
begin
(1) Initial match: potential matching candidate (x, y, d) is assigned strength S^0.
(2) Repeat until no change
    (a) for each $r = (x, y, d)$
        for each of its neighbor $r' = (x', y', d')$
            Apply the neighborhood support function, such that at $k + 1$ iteration

$$S^{(k+1)}(r) = S^{(k)}(r) + \sum_{r' \in \Re_1(r)} \frac{f(\delta d) \cdot S^{(k)}(r')}{r_{xy}} - \eta \cdot \sum_{r' \in \Re_2(r)} S^{(k)}(r'),$$

        where $r_{xy} = \sqrt{(x' - x)^2 + (y' - y)^2}, \delta d = |d' - d| r_{xy}$,
            $\Re_1(r) = \{(x', y', d') | x - a \leq x' \leq x + a, y - b \leq y' \leq y + b,$
                   $d - c \leq d' \leq d + c, \neg(x' = x, y' = y)\}$,
            $\Re_2(r) = \{(x', y', d') | x' = x, y' = y, d' \neq d\}$.
    (b) /* Winner-take-all */
        If $S^{(k+1)}(r)$ reaches max_$strength$
            then $S^{(k+1)}(r') = 0|_{r' \in \Re_2(r)}$
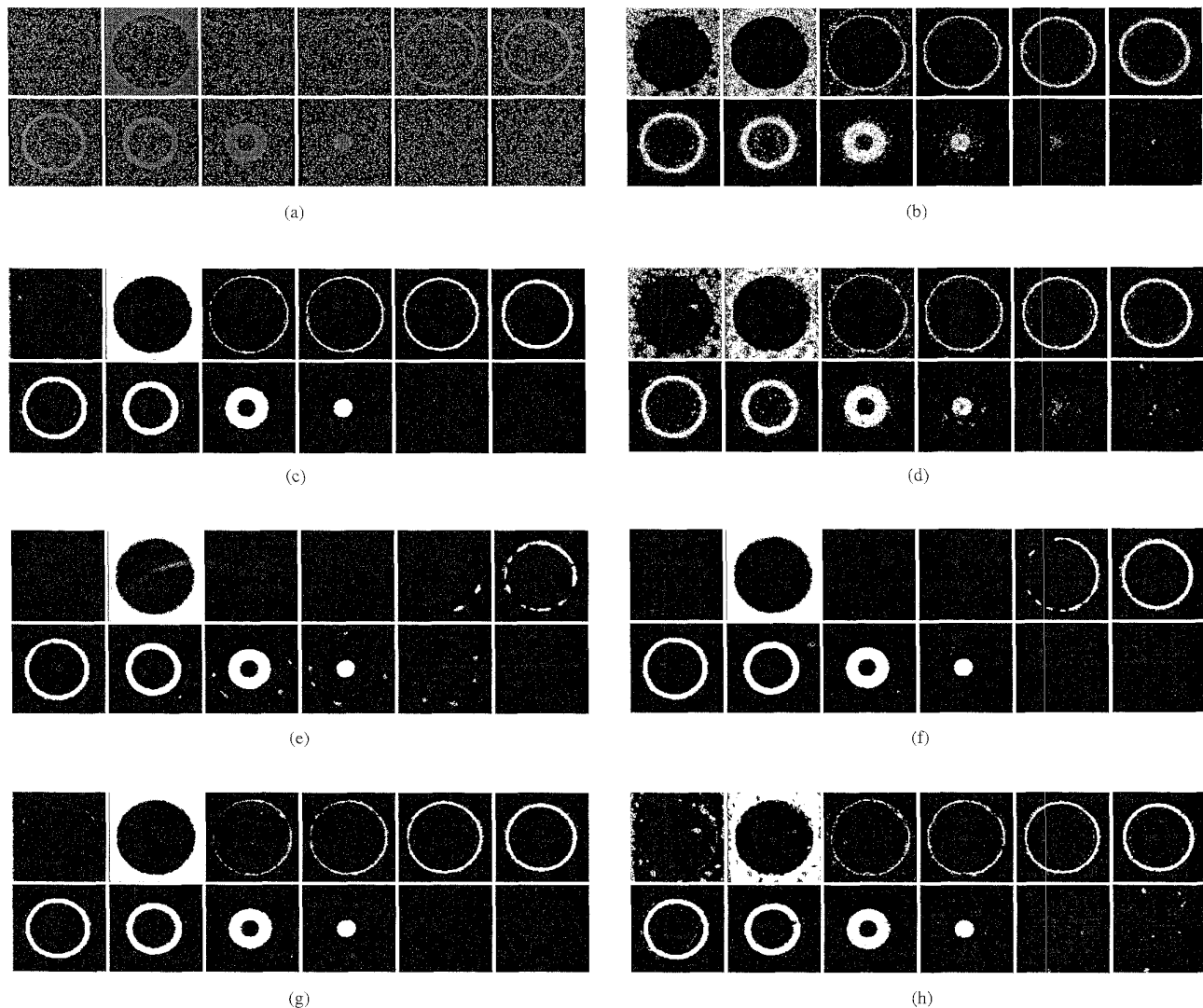end

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig. 7.   Solutions for the "hemisphere" RDS. (a) Initial match. (b) $f1_a$. (c) $f2_b$. (d) $f2_d$. (e) $f_{3(0.1)}$. (f) $f_{3(0.5)}$. (g) $f_{3(1.5)}$. (h) $f_{3(3.5)}$.

Since changes of the surface orientation and depth are continuous, it would suggest that a continuous $f(\delta d)$ is favorable over an ad hoc step function. As pointed out in Section II-B, the slope and depth of 3-D surfaces have an intrinsic relationship with δd. Given the maximum $f(\delta)$ to be at $f(0) = 1$, the following criteria can be used:

- $f(\delta d)$ should have a negative value when $\delta d \geq 2$, since this is a "forbidden" situation;
- $f(\delta d)$ should not be a significantly positive value when $1 \leq \delta d < 2$, since the related surfaces in the 3-D scene are severely away from the sight, and the probability of having these surfaces is not very high;
- $f(\delta d)$ should start to decline when δd approaches 1 (from 0), since the related surfaces are turning away from the sight, fusion becomes increasingly difficult and decreasingly useful.

In this study, the following three families of functions are tested. Table I lists the parameter values of the chosen test

functions. Figs. 4 and 5 show graphically the curves for some of the functions.

$$f_1(\delta d) = \begin{cases} 1 & 0 \leq \delta d < 1.0 \\ C_1 & \delta d = 1.0 \\ 0 & 1.0 < \delta d < p \\ C_2 & \delta d \geq p \end{cases} \qquad (8)$$

$$f_2(\delta d) = \frac{2}{1 + e^{(\delta d - p)/T}} - 1.0 \qquad (9)$$

$$f_3(\delta d) = 2 \cdot e^{-\delta d/T} - 1.0. \qquad (10)$$

The output grey-level coded disparity maps of the "hemisphere" RDS using three families of NSF's are shown in Fig. 6. The spherical surface is shown to be representative and challenging for testing arbitrary 3-D surfaces [15], since it offers a variety of surface orientations. In this case, it provides a whole range of disparity gradients.

Fig. 7(a) shows initial matches (including false targets) at the 12 layers ($d \in [0, 11]$, displayed in row major) of the xyd

TABLE II
COMPARISON OF MATCHING ERRORS FROM "HEMISPHERE" RDS

Function	Parameters	Percentage of Errors		Percentage of Correct Matches
		Type A	Type B	
f_{1a}	$p = 1.5, C_1 = 1.0, C_2 = -1.0$	29.17	0	70.83
f_{1b}	$p = 1.5, C_1 = 0.0, C_2 = -1.0$	9.01	0	90.99
f_{1c}	$p = 1.5, C_1 = 0.0, C_2 = -0.5$	9.53	0	90.47
f_{1d}	$p = 2.0, C_1 = 0.0, C_2 = -1.0$	9.52	0	90.48
f_{2a}	$p = 1.0, T = 0.1$	6.45	0	93.55
f_{2b}	$p = 1.0, T = 0.3$	1.86	0	98.14
f_{2c}	$p = 1.5, T = 0.3$	20.45	0	79.55
f_{2d}	$p = 1.8, T = 0.5$	20.00	0	80.00
$f_{3(0.1)}$	$T = 0.1$	3.24	10.25	86.51
$f_{3(0.5)}$	$T = 0.5$	2.47	4.30	93.23
$f_{3(1.0)}$	$T = 1.0$	2.33	0.08	97.58
$f_{3(1.2)}$	$T = 1.2$	1.81	0.01	98.18
$f_{3(1.5)}$	$T = 1.5$	1.48	0	98.52
$f_{3(1.8)}$	$T = 1.8$	1.51	0	98.49
$f_{3(2.0)}$	$T = 2.0$	1.75	0	98.25
$f_{3(2.5)}$	$T = 2.5$	2.52	0	97.48
$f_{3(3.0)}$	$T = 3.0$	4.06	0	95.94
$f_{3(3.5)}$	$T = 3.5$	6.37	0	93.63
$f_{3(4.0)}$	$T = 4.0$	9.19	0	90.81

Fig. 9. Unsettled pixels (in white) after 95% convergence. (a) f_1a. (b) f_2b. (c) f_2d. (d) $f_3(0.1)$. (e) $f_3(1.5)$. (f) $f_3(3.5)$.
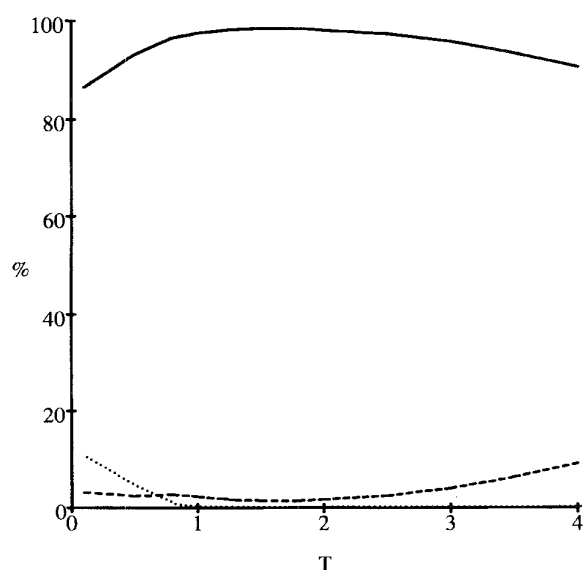

Fig. 8. Performance of various f_3 functions on the "hemisphere" RDS. The three curves indicate percentages of type A errors (dashed) and type B errors (dotted), and percentages of correct matches (solid), respectively.

Fig. 10. Random-dot stereogram (128 × 128) for two transparent surfaces. Disparity value for the flat surface is $d = 4$, and for the sloped surface ranges from $d = 1$ to 6.

space. Since all possible matches are initially assigned $S^{(0)} = 128$, they appear grey. Fig. 7(b)–(h) shows the convergence results using some of the above NSF's. Since false targets are expected to be eliminated and winners get the maximum strength (indicated by color white), a good convergence result for the "hemisphere" RDS should consist of layers of white rings and a clean background.

For comparison, the first family (f_{1a} to f_{1d}) includes several ad hoc step functions. The parameters p, C_1 and C_2 are deemed to be reasonable, and most of them do deliver satisfactory
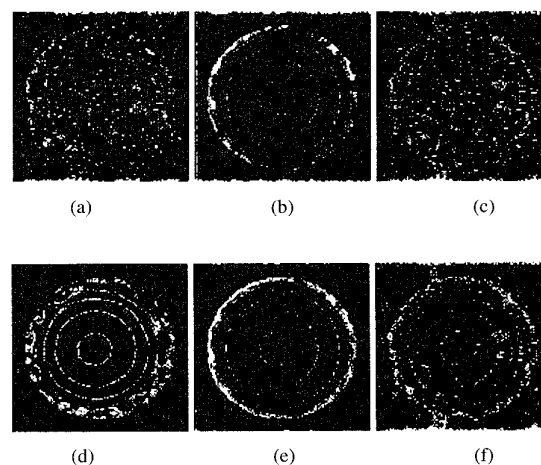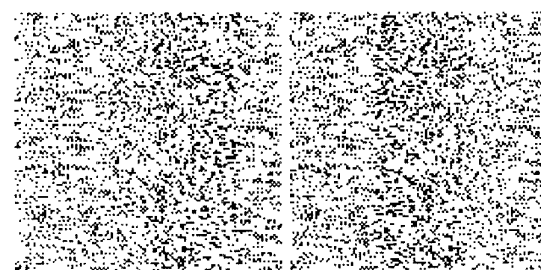
results. The variations of p and C_2 do not seem to have strong impact. The poor result from f_{1a} is due to the choice of
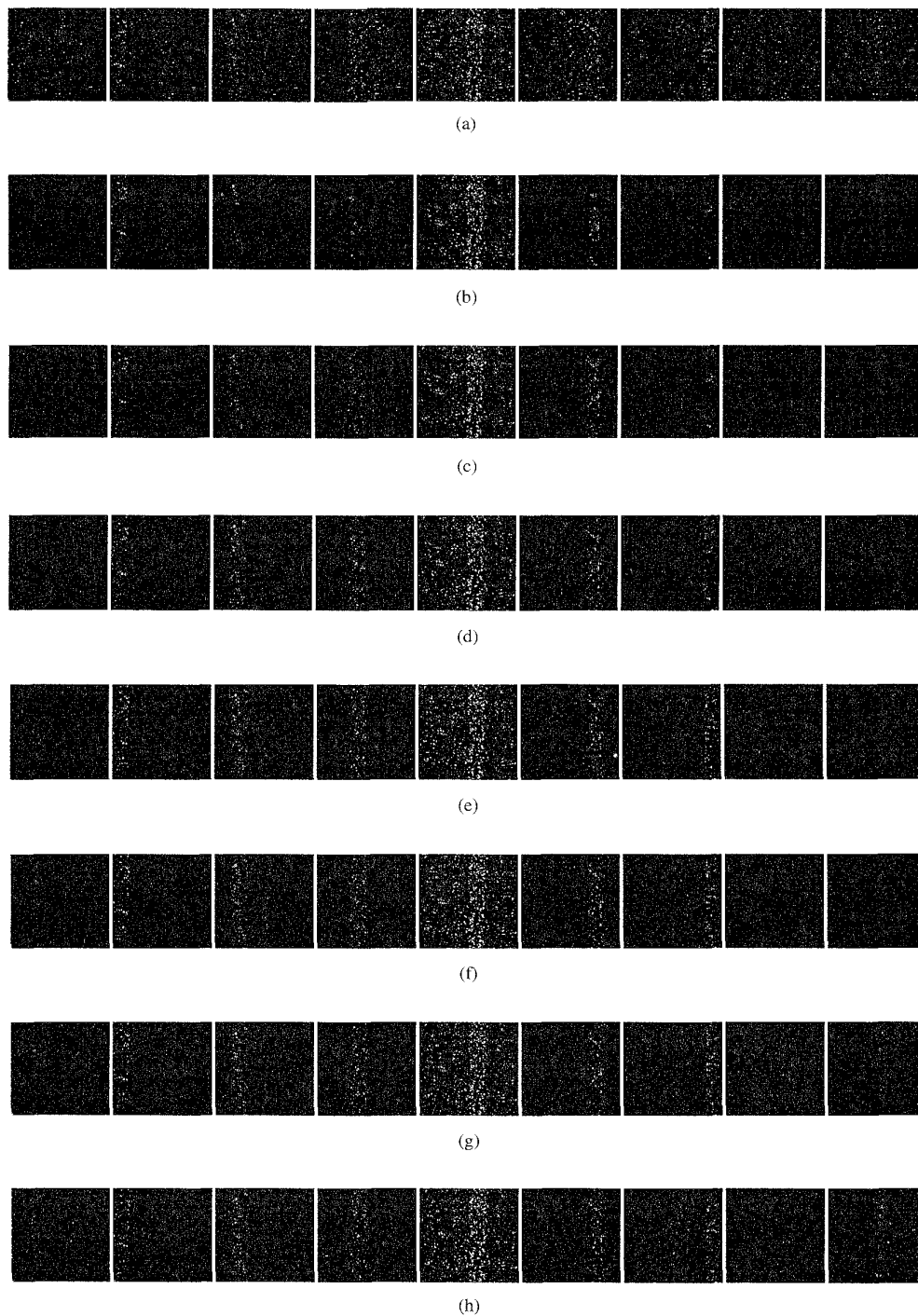
Fig. 11. Comparative results for transparent "slope" RDS using f_3 NSF's. (a) Initial match. (b) T = 0.1. (c) T = 0.5. (d) T = 0.8. (e) T = 1.1. (f) T = 1.5. (g) T = 2.0. (h) T = 3.0.

$C_1 = 1.0$. Because a fairly large portion of the neighbors (x', y', d') have $\delta d = 1$, the large positive value of C_1 is shown inappropriate.

The second family $f_2(\delta d)$ is inspired by the support function often used in the simulated annealing [9], in which the temperature parameter T controls the process. The function is varied to meet our criteria, and an additional parameter p

is introduced to provide variable turning points. Among the four tested functions in this family, f_{2b} yields the best result, and f_{2d} is the worst. The main reason is that f_{2d} provides too much positive support when $\delta d \geq 1$.

The relative success of f_{2b} leads to the study of the third family—exponential decay functions. The function peaks at 1 when $\delta d = 0$ and decays to -1 when $\delta d \rightarrow \infty$.
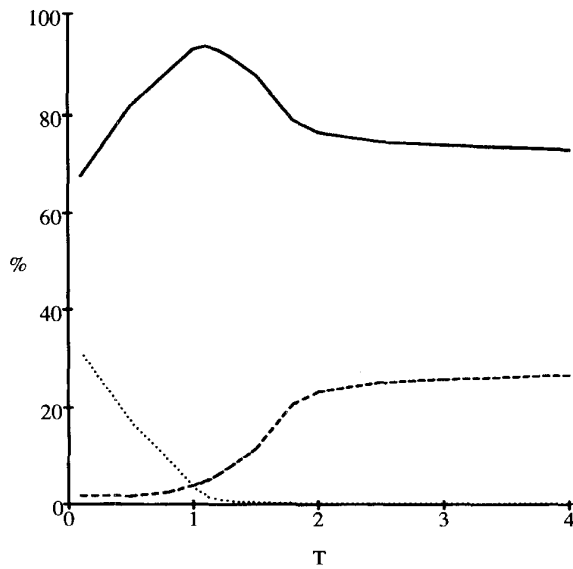
Fig. 12. Performance of f_3 NSF's on the transparent "slope" RDS. The three curves indicate percentages of type A errors (dashed) and type B errors (dotted), and percentages of correct matches (solid), respectively.

Moreover, it has a single control parameter T that facilitates a more conclusive analysis. It appears to have all the desirable characteristics.

As shown in Fig. 6(l), $f_{3(1.5)}$ apparently yields the best disparity map. (Interestingly, it can be seen that f_{2b} in Fig. 4 and $f_{3(1.5)}$ in Fig. 5 are somewhat similar in shape.) The convergence result of $f_{3(1.5)}$ is shown in Fig. 7(g) to be fairly clean. When T is too large, e.g., $T \geq 3.0$, the performance deteriorates due to the excessive positive support for large δd's. In fact, it starts to look like f_{2d}. When T is too small, e.g., 0.1, several rings at the sloped locations of the hemisphere are not preserved (Fig. 7(e)). It is because $f_{3(0.1)}$ basically only supports frontal surfaces where $\delta d = 0$, it overly suppresses surface discontinuities. As a result, the pixels at sloped surfaces lose strength S at all d layers (including the correct d). We call these type B errors as opposed to type A errors where winners emerge at erroneous d in the xyd space.

A. Quantitative Comparison of Errors

Because the synthetic terrain map for the RDS is known in advance, the exact number of type A error pixels can be counted when $\Delta d \neq 0$, where Δd is the difference between the output disparity and the correct disparity value. When there exists no winner/survivor at an (x, y) position, it is counted as a type B error. Errors are assessed when the relaxation process reaches its convergence state—no more changes for over 99% of the total pixels in the image.

Table II shows the percentages of type A and type B errors and correct matches from the "hemisphere" RDS. Fig. 8 is a graphical display of the same result.

It should be obvious that $f_{3(1.5)}$ has the highest percentage (98.52%) of correct matches. Moreover, its errors are almost

entirely at $\Delta d = 1$. For $1.2 \leq T \leq 2.0$, all f_3 functions yield comparable results. As mentioned before, $f_{3(0.1)}$ and $f_{3(0.5)}$ do not give generous support even for small nonzero δd. Consequently, many pixels at sloped surfaces are oversuppressed. So far, these are the only times when the type B errors occur.

Similar results are also obtained from the RDS "floating rectangle." Not surprisingly, since both foreground and background are simply frontal planes, the functions $f_{3(0.1)}$ and $f_{3(0.5)}$, which almost exclusively favor $\delta d = 0$, also perform fairly well; they do not suffer from type B errors in this simple case.

B. Comparison in Resolving Occlusion

Occlusion usually accompanies a relatively large surface discontinuity (and, hence, large δd). As pointed out in [10], a single-pass matching usually generates "weak-winners" at the occlusion boundaries, and a two-pass algorithm can be designed to better recover the boundaries. This section compares the weak-winners produced by different NSF's to see whether they have the potential to be used in the two-pass algorithm.

The relaxation process is now terminated when 95% of the pixels are settled. Fig. 9 shows the remaining 5% unsettled pixels using different NSF's. The NSF's are selected to demonstrate "good" and "bad" results.

Apparently, $f_{3(1.5)}$ delivers the best result. After a couple dozen iterations, the interior of the relatively smooth foreground and background surfaces are almost all settled. The subsequent iterations are solely confined to the areas where large δd occurs. At the end, the steep portion of the hemisphere is correctly identified as "unsettled." It thus can be used in the proposed two-pass algorithm. Because the converging process is stable and monotonic, it does not need overly careful tuning. In fact, a deviation of T values in the range of 1.2 and 2.0 for the f_3 functions does not significantly affect the result.

The bad results come from two groups: One is represented by f_{3a}, the effect of oversuppression destroys the near-base portion of the hemisphere. The others are f_{1a}, f_{2d}, f_{3g}. These functions provide oversupport and are again losing by scattering unsettled pixels everywhere.

C. A Well-Behaved NSF

Our study concludes that the exponentially decay function is a well-behaved NSF for the unified cooperative stereo-matching algorithm, because (i) it has a single parameter T, which is easier to adjust, (ii) there exists a range of moderate T values that will yield good convergence results, (iii) it generates the least matching error, and (d) it has the potential to be used for resolving occlusions.

In the above evaluations, various values are also assigned to other parameters, e.g., η and the size of the neighborhood (a, b, c). The impact of η is somewhat insignificant. As long as it is significantly greater than 1 (e.g., in the range of 4 to 8), the convergence results are all comparable. The neighborhood
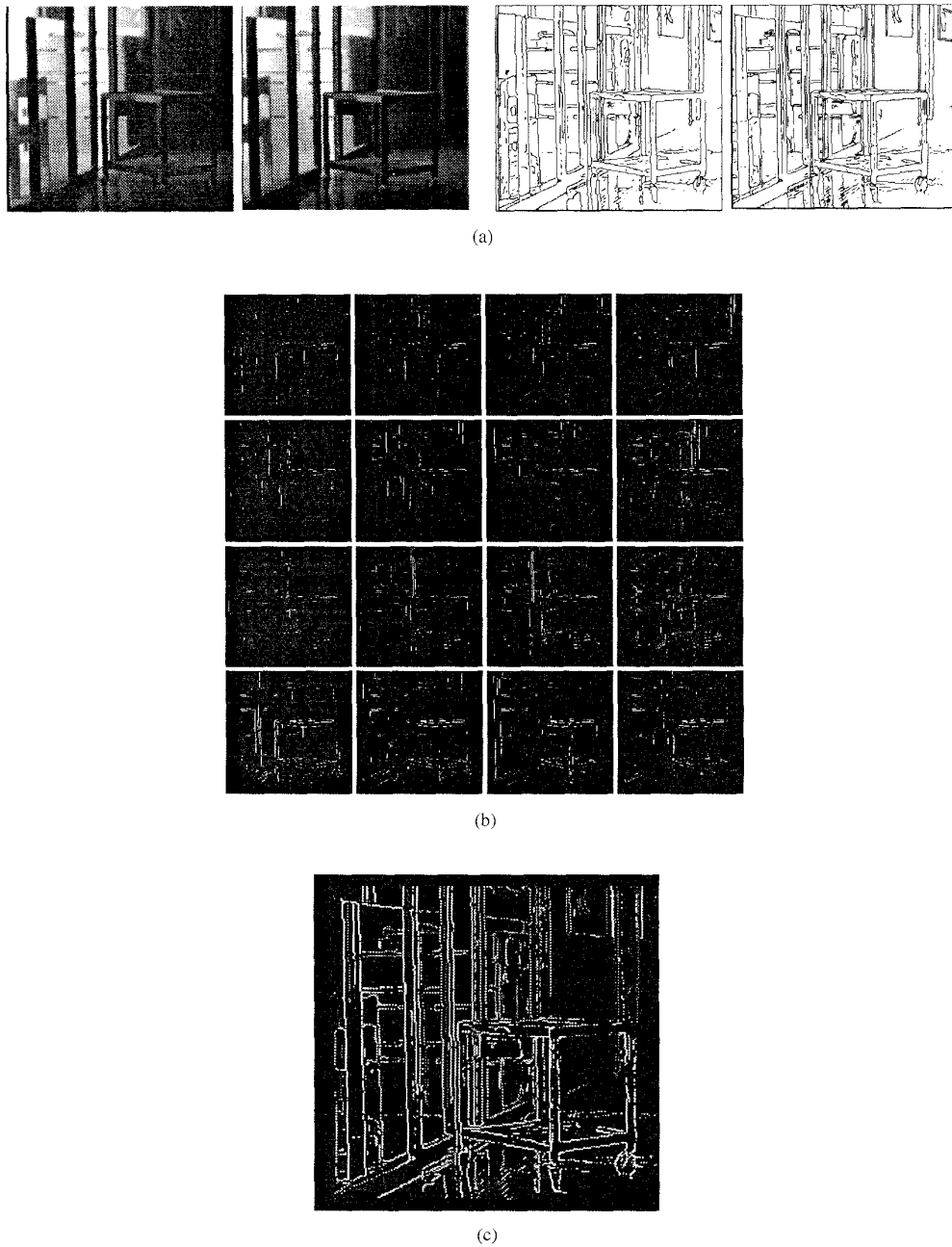
(a)

(b)

(c)

Fig. 13. (a) Left and right stereo corridor scene images and their edge maps. (b) Initial match at sixteen possible disparity values. (c) Disparity map (gray levels show disparities).

size is usually chosen according to the image resolution and the density of the features. In our tests, a moderate value for a, b, and c at 2, 3, 4, or 5 produces satisfactory (and similar) results. Our experiment does show that the converging speed of the relaxation process has a large impact on the quality of the result. If the amount of update is too large at each iteration, the quality of the result deteriorates. Unfortunately, using small updates at each iteration incurs an increase in the required number of iterations. In the above tests, the process

converges in the range of 30–50 iterations for most functions for both "hemisphere" and "floating rectangle."

V. TRANSPARENT RANDOM-DOT STEREOGRAMS

Transparent surfaces are often a challenge to stereo matching. The inclusion of transparency invalidates the concept of the "forbidden zone." Namely, order reversal of matching points becomes possible because there is no longer any opaque surface to prevent this from happening. In terms of disparity
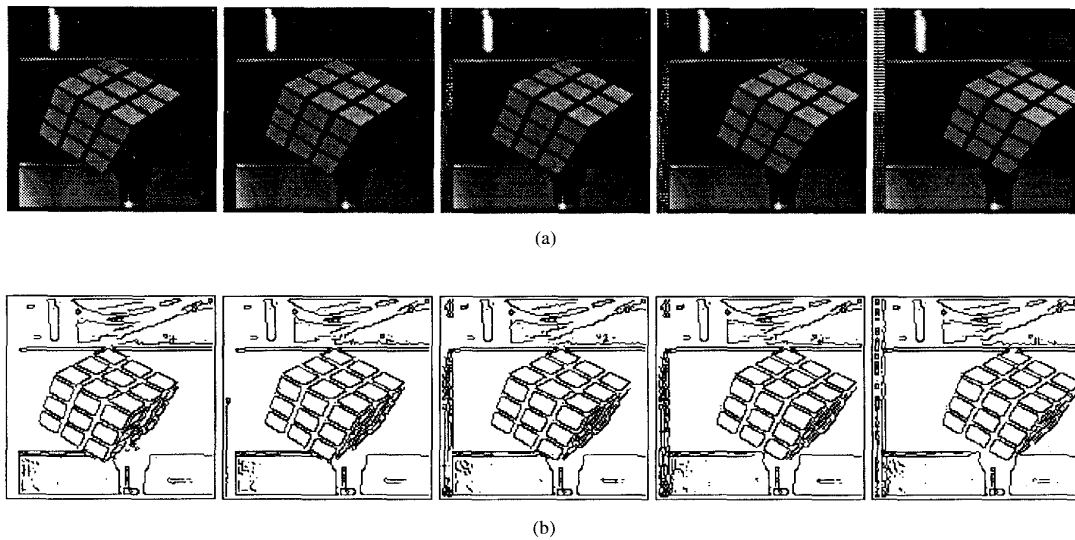
(a)

(b)

Fig. 14.   Sequence of motion stereo images. Nine images are used; five of them (at $t = t_0, t_2, t_4, t_6$, and t_8) are shown. (a) Gray-level images. (b) Edge maps.

gradient, large values ($\delta d > 2$) become legitimate. However, if the transparent surfaces are not extremely steep as is usually the case, then small δd should still be favored. In this sense, the disparity gradient as the basis of the algorithm is still valid, and the above good NSF's are expected to be applicable to the transparent RDS's.

Since surfaces are now transparent, only black dots in the RDS will be treated as being on the surface; the white dots will be viewed as transparent holes to enable the visibility of any black dots behind them in the 3-D scene. The implication of this is a small change in the matching algorithm, i.e., only black-black matches are now allowed. Also, since features are sparser, a larger neighborhood is used in the algorithm, i.e., $a = b = c = 5$. The other parameters are not altered. Our first test transparent RDS is shown in Fig. 10.

Fig. 11 shows the result using various f_3 functions. The best result is obtained when $T = 1.1$, both the flat surface at $d = 4$ and a sloped surface ranging from $d = 1$ to 6 can be seen clearly. Again, when T is too small (e.g., $T = 0.1$), the surfaces are oversuppressed; whereas it is noisy when T is too large (e.g., $T = 3.0$). Fig. 12 is a graphical display of the matching result similar to Fig. 8. It is encouraging to see that the general shapes of the three curves are similar in both graphs. The peak percentage value of the correct matches is now lower (93.60% versus 98.52% for the opaque hemisphere surfaces) and the peak of the corresponding curve (solid) is much sharper. This reflects the fact that the matching process is more difficult in the transparency cases, and more sensitive to the choice of the T value. The smaller best T value (1.1 as opposed to 1.5 previously) seems to suggest that a slightly more suppressive NSF is needed under more severe situations of surface discontinuity/occlusion.

An additional test was performed to further test the robustness of the algorithm. The same two transparent surfaces are used, except twice as many black dots are assigned to the flat surface than to the sloped surface. Since the survival of any individual cell in the xyd space relies on

the aggregated support from its neighbors, this is to create an adversary situation for the sloped surface for reduced support as compared with the flat surface. Nevertheless, the result is quite satisfactory, except the percentage of correct matches is dropped to 91.69% for the best result, now at $T = 1.2$.

The same tests are performed on a "pyramid" (wedding-cake) transparent RDS that consists of a pyramidal surface penetrated by a flat surface in the middle. The success rate when $T = 1.1$ is actually slightly higher than the "slope" (95.51% as opposed to 93.60%) .

## VI.   TEST RESULTS FROM REAL-WORLD IMAGES

The above study employs random-dot stereograms in which features (dots) are perfectly matchable. However, for applications involving real-world images, noises are inevitable. In this section, real-world stereo images will be used. Also, edges are used as the feature to be matched. The algorithm is thus tested for contour matching instead of area matching as in the case of RDS's. Noises in edge features are inevitable during the image-acquisition and edge-detection processes. As pointed out by Grimson in [16], the success of contour matching algorithms relies on the constraint of figural continuity. In terms of disparity gradient, continuity and figural continuity are similar in favoring $\delta d \ll 1$. It will be shown that the disparity gradient based cooperative stereo algorithm is equally effective in enforcing the figural continuity in contour matching. Contrary to [10], the fact that features are quite sparse does not adversely affect the performance of the algorithm.

A.   Binocular Stereo

The disparity gradient based cooperative stereo algorithm has been tested for depth recovery from various binocular stereo images. As an example, a pair of such images from an ordinary corridor scene is shown in Fig. 13(a). The first room is an office with sliding glass doors. A cart is placed in the corridor to add a little more complexity to the scene.
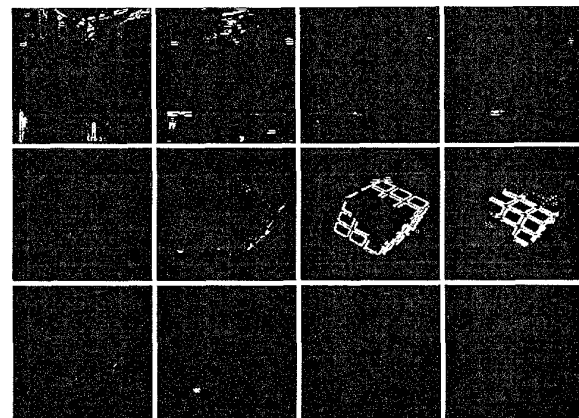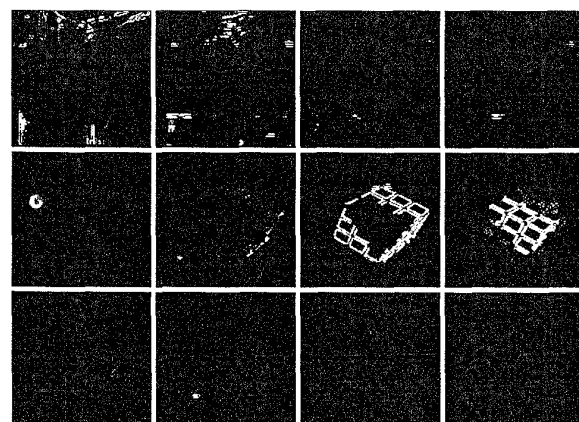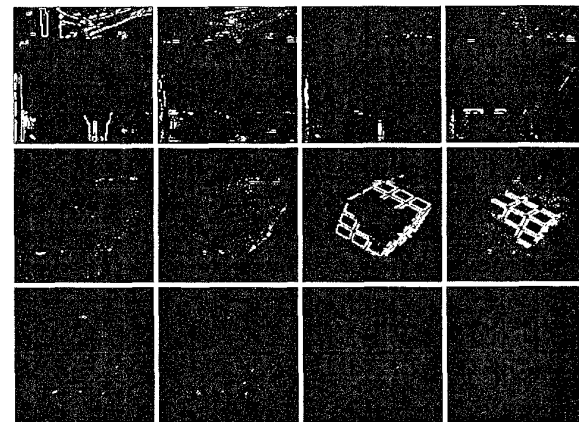
(a)

(b)
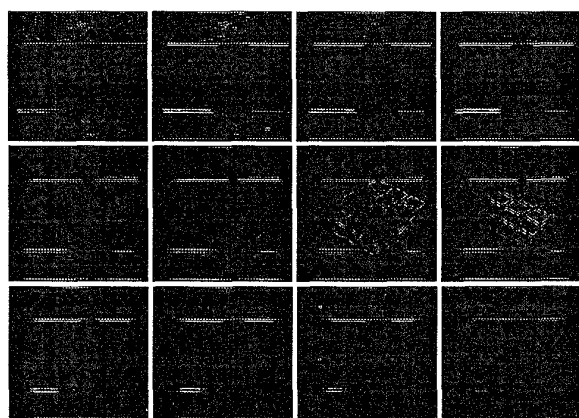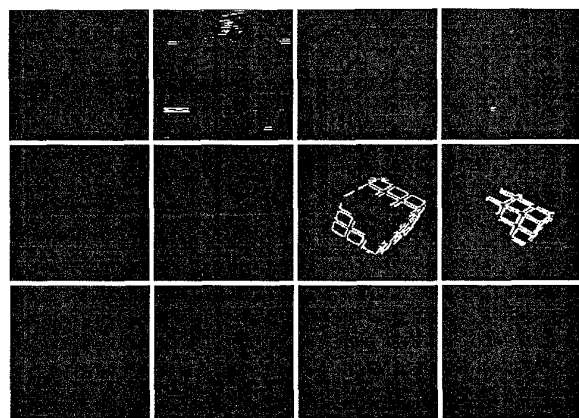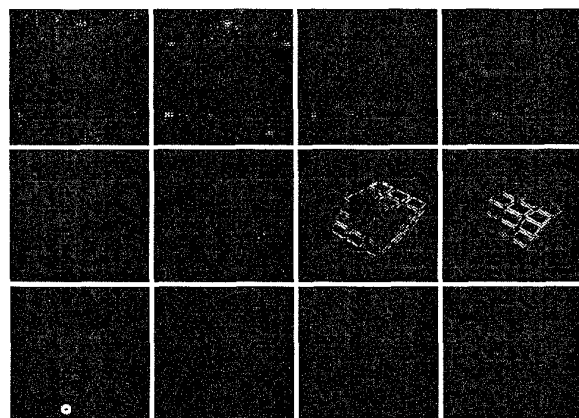
(c)

Fig. 15. Stereo matching results using the NSF $f_{3(1.3)}$. In each image, 12 layers of the xyd space are shown (in row major), gray levels are used to show the strength S. (a) Initial match. (b) After 12 iterations. (c) After 90 iterations.

(a)

(b)

(c)

Fig. 16. Comparative stereo matching results after 90 iterations using (a) $f_{3(0.5)}$, (b) $f_{3(1.3)}$, (c) $f_{3(3.0)}$.

Fig. 13(b) shows the initial matching result, 16 layers ($d = 0..15$, row major) of the xyd space are shown. It can been seen that initially many edge pixels have potential matches at more than one disparity layer. After 90 iterations, over 95% edge pixels converge to a unique and correct disparity value. The final dispaity map is shown in Fig. 13(c), where gray-level intensity is used to illustrate the disparity. Errors on disparity values can be observed, for example, at the edges of the posters at the far end of the corridor. The quality of the disparity map

is comparable to the result presented in [17], where matching involves a higher level feature, i.e., line.

B. Motion Stereo Application

In this section, a sequence of motion stereo images are used. The reason for the additional test on motion stereo images is that they often incur more false targets and, hence, provide a more interesting test case.

A Brief Introduction to Motion Stereo: Consider an assembly line where a belt is moving at a relatively constant speed. Multiple snapshots of the moving objects on the belt can be taken in rapid succession by a single camera. The controlled belt movement provides the necessary stereo disparities. Moreover, it guarantees that the disparities occur only along the line of the belt movement, e.g., on the x-dimension, and thus provides well-defined epipolar lines. This method is called *motion stereo* [18]. Its greatest advantage is the simplicity in camera control and calibration.

Bolles, Baker, and Marimont [19] proposed a technique of epipolar-plane image analysis for determining structure from motion. It was pointed out that for straight-line camera motions, simple linear structures would be formed in the epipolar-planes—the xt-plane.

The correspondence problem in the motion stereo can be reduced to a problem of searching for collinear points in the epipolar planes. Similar to the procedures for the Hough transform [20], a voting algorithm for accumulating multiple evidence can be developed [21]. In general, in the xt-plane, any point at $t = t_i$ can be paired with any point at $t = t_j$ $(j > i)$ to form a hypothetical line segment. Its slope suggests a possible disparity value d. A 3-D xyd voting space can be created, and each hypothetical line will cast a vote at the position (x, y, d) in the voting space. Since $n + 1$ collinear points can form $O(n^2)$ hypothetical lines, and they will vote to the same (x, y, d), a peak will be formed in the xyd space that indicates the consensus on the correct disparity value for the point (x, y).

Since many pairs of points in the epipolar plane can form hypothetical lines and cast votes to the xyd voting space, the space is cluttered with false targets. Also, due to the errors in digitization, edge detection, and slope calculation, peaks are rarely distinct and sharp. A relaxation process is thus necessary following the voting.

Experimental Result of Motion Stereo: The scene of a Rubik's cube on a moving belt in our lab is used. The cube is propped up on one corner. A sequence of nine images are taken. Fig. 14 shows five of the nine images and their edge maps. Edges with similar orientations $(\Delta\theta < 45°)$ are matched. The correct disparity range for the cube pixels is $d \in [5.3, 6.9]^1$. Since the sensors and wires are not moving, they have $d = 0$. The initial search range for d in our program is from 0 to 11.

Fig. 15 shows the voting and relaxation results after 0, 12 and 90 iterations using the NSF $f_{3(1.3)}$. In each of the images (a), (b) and (c), 12 layers $(d = 0 \cdots 11$, row major) of the xyd

---

[1] This refers to the disparity in two consecutive frames. Subpixel precision is obtained by dividing the multiple disparity value by $(n - 1)$, where n is the number of frames in the motion stereo.

space are shown. Gray-level intensities are used to show the strength S. It can be seen from Fig. 15(a) that, initially, most candidates have intermediate to weak strengths, and many cube pixels show strength at more than one disparity layer. After 90 iterations (Fig. 15(c)), the cube pixels are shown to have correctly converged to one of the three disparity levels $d = 5, 6, 7$. The sensors and wires settle down to disparity 0 or 1.

For a comparison, Fig. 16 shows the results from three exponential decay functions with $T = 0.5, T = 1.3$ and $T = 3.0$, respectively. It confirms that $f_{3(1.3)}$ generates a better result. Fig. 16(a) shows an oversuppressed xyd space, and 16(c) is noisy and ambiguous—again a symptom of oversupport.

VII. CONCLUSION

This paper advocates a unified cooperative stereo approach based on the disparity gradient. We argue that the disparity gradient subsumes various constraints for stereo matching and, thus, forms a solid basis for the unified algorithm.

To seek for a better understanding of the intrinsic relationship between the NSF and the disparity gradient, a comprehensive analysis and evaluation of three families of NSF's is conducted. It is shown that the exponential decay function with a moderate T value is well behaved in that it yields the least error, converges steadily, and produces correctly located weak-winners. It is suggested that this function will help a two-pass algorithm in resolving the dilemma of surface continuity and discontinuity/occlusion.

Experimental results from random-dot stereograms (RDS's) containing both opaque and transparent surfaces are analyzed. Results from binocular stereo and motion stereo images from real-world scenes are also demonstrated. The accuracy of the unified cooperative stereo-matching algorithm compares well with the best-known previous results.

REFERENCES

[1] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283–287, Oct. 1976.
[2] M. Drumheller and T. Poggio, "On parallel stereo," in *Proc. IEEE Conf. Robot. Automat.*, 1986, pp. 1439–1448.
[3] J. E. W. Mayhew and J. P. Frisby, "Psychophysical and computational studies toward a theory of human stereopsis," *Artificial Intell.*, vol. 17, pp. 349–385, 1981.
[4] P. Burt and B. Julesz, "Modifications of the classical notion of Panum's fusional area," *Percept.*, vol. 9, pp. 671–682, 1980.
[5] H. P. Trivedi and S. A. Lloyd, The role of disparity gradient in stereo vision. *Percept.*, vol. 14, pp. 685–690, 1985.
[6] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Percept.*, vol. 14, pp. 449–470, 1985.
[7] S. B. Pollard, J. Porrill, J. E. W. Mayhew, and J. P. Frisby, "Disparity gradient, Lipschitz continuity, and computing binocular correspondence," in O. D. Faugeras and G. Giralt, Eds, *Proc. 3rd Int. Symp. Robot. Res.*, 1986, pp. 19–26.
[8] C. V. Stewart, "An analysis of the probability of disparity changes in stereo matching and a new algorithm based on the analysis," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 1991, pp. 670–671.
[9] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognit. Sci.*, vol. 9, pp. 147–169, 1985.
[10] J. Little and W. E. Gillett, "Direct evidence for occlusion in stereo and motion," *Image Vision Comput.*, vol. 8, no. 4, pp. 328–340, 1990.
[11] K. Prazdny, "Detection of binocular disparities. *Biol. Cybern.*, vol. 52, pp. 93–99, 1985.

[12] R. Szeliski and G. Hinton, "Solving random-dot stereograms using the heat equation," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 1985, pp. 284–288.

[13] C. W. Tyler and M. B. Clarke, "The autostereogram," *SPIE Stereoscop. Displays Applic.*, vol. 1258, pp. 182–196, 1990.

[14] H. W. Thimbleby, S. Inglis, and I. H. Witten, "Displaying 3-D images: Algorithms for single-image random-dot stereograms," *Computer,*vol. 27, no. 10, pp. 38–48, 1994.

[15] Z. N. Li and G. Hu, "On edge preservation in multiresolution images," *CVGIP: Graphic. Model. Image Processing*, vol. 54, no. 6, pp. 461–472, 1992.

[16] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 17–34, 1985.

[17] Z. N. Li, "Stereo correspondence based on line matching in Hough space using dynamic programming, *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 1, pp. 144–152, 1994.

[18] R. Nevatia, "Depth measurement by motion stereo," *Comput. Graphics Image Processing*," vol. 5, pp. 203–214, 1976.

[19] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vision*, vol. 1, pp. 7–55, 1987.

[20] R. O. Duda and P.E. Hart, "Use of the Hough transform to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[21] F. Tong and Z. N. Li, "Reciprocal-wedge transform for space-variant sensing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 5, pp. 500–511, 1995.

Gongzhu Hu (S'84–M'86–S-86–M'87) received the Dipl. in numerical analysis from Qing-Hua University, Beijing, China, in 1968, the M.S. degree in computer science from the University of Wisconsin, Madison, WI, in 1981, and the Ph.D. degreee in computer science from Michigan State University, East Lansing, MI, in 1988.

From 1968 to 1980, he worked as a software engineer in China. He joined the Computer Science Department at Central Michigan University, Mount Pleasant, MI, as an instructor in 1987, and became an Assistant Profesor in 1988, an Associate Professor in 1991, and a Professor in 1995. He currently serves as Department Chairperson. His research interests are image processing, computer vision, and object-oriented database systems with image data.

Dr. Hu is a member of the Association for Computing Machinery.

Ze-Nian Li (S'85–M'86) received the B.S. degree in electrical engineering from the University of Science and Technology of China in 1970, and the M.S. and Ph.D. degrees in computer science from the University of Wisconsin, Madison, WI, in 1981 and 1986, respectively.

From 1970–1979, he was an Electronic Engineer in charge of digital and analogical system design. He was an Assistant Professor at the University of Wisconsin, Milwaukee, WI, from 1986–1987. In 1988, he joined the School of Computing Science at Simon Fraser Univeristy, Burnaby, Canada, where he is currently an Associate Professor. His research interests include computer vision, pattern recognition, parallel vision machines and algorithms, and content-based search in multimedia systems.