Structure from Stereo—A Review

UMESH R. DHOND, STUDENT MEMBER, IEEE, AND J. K. AGGARWAL, FELLOW, IEEE

***Abstract*—Major recent developments in establishing stereo correspondence for the extraction of the 3-D structure of a scene are reviewed. Broad categories of stereo algorithms are identified based upon differences in image geometry, matching primitives, and the computational structure used. Performance of these stereo techniques on various classes of test images is reviewed and the possible direction of future research is indicated.**

I. Introduction

A major portion of the research efforts of the computer vision community has been directed towards the study of the three-dimensional (3-D) structure of objects using machine analysis of images. Analysis of video images in stereo has emerged as an important passive method for extracting the 3-D structure of a scene. Earlier, Barnard and Fischler [6] presented a review covering the major steps involved in stereo analysis, the evaluation criteria for stereo algorithms, and a survey of the different approaches to computational stereo developed starting from the mid-70's up to 1981. In this paper we review the computational structure of the major schemes that have evolved in the past decade for recovering depth using stereo.

The basic principle involved in the recovery of depth using passive imaging is triangulation. Many active range sensing techniques are also based upon the triangulation principle. However, in active ranging techniques that use triangulation, the nature of the problem is different in that the triangle for recovering depth is predefined by three points—the light source, the illuminated spot in the scene, and its image point. Thus, in active methods that use triangulation, the correspondence problem has already been solved by using an artificial source of illumination.

In stereopsis, which is a passive technique, the triangulation needs to be achieved with the help of only the existing ambient illumination. Hence a correspondence needs to be established between features from two images that correspond to some physical feature in space. Then, provided the position of centers of projection, the effective focal length, the orientation of the optical axis, and the sampling interval of each camera are known, the depth can be reconstructed using triangulation. Based upon this basic correspondence problem, a particular matching paradigm can be constructed depending upon the specific matching features used, the number of cameras used, the positioning of the cameras, and the scene domain.

The problem of passive range sensing is important where there are overriding circumstantial constraints that prevent the use of artificial illumination or other active sources of radiation. Applications of stereo-based depth measurement include automated cartography, aircraft navigation, autonomous land rovers, robotics, industrial automation and stereomicroscopy.

In the following sections, we identify broad categories of matching algorithms depending upon various factors like the imaging geometry, the matching primitives, as well as the matching strategy used. Within each category, the implementation details of the contemporary approaches will be highlighted. Section II gives an overview of the major steps involved in the process of stereopsis, namely, preprocessing, stereo matching, and depth reconstruction. Section III examines various computational theories of stereopsis that have been motivated by the human visual system. Sections IV through X describe the major computational techniques that have been successfully tested in the past decade for solving the stereo correspondence problem. In Section IV we review area-based correlation schemes. Relaxation labeling processes have been used by many researchers to iteratively impose global consistency constraints on multiple matches for the purpose of disambiguation, which we describe in Section V. Many stereo algorithms use edge segments obtained from fitting piecewise linear curves to connected edges. Section VI describes two approaches for using edge segments in stereo matching. Stereo algorithms that utilize hierarchical computational structures are described in Section VII. In Section VIII, we examine the use of dynamic programming methods for stereo matching. The trinocular camera setup and the resulting matching paradigm, with both point- and segment-based matching algorithms, are reviewed in Section IX. Section X briefly describes the formulation of the correspondence problem using structural descriptions. Section XI deals with the aspect of performance evaluation of stereo algorithms and the various classes of test data used. Section XII contains concluding remarks.

II. The Process of Stereopsis

The major steps involved in the process of stereopsis are preprocessing, establishing correspondence, and recovering depth. In this section we shall briefly examine each of them.

Manuscript received October 17, 1988; revised March 30, 1989. This work was supported in part by a grant from the Army Research Office under Contract DAAL03-87-K-0089 and in part by a grant from the National Science Foundation under Contract NSF/ECS-8513123.

The authors are with the Computer and Vision Research Center, College of Engineering, ENS Building, University of Texas, Austin, TX 78712.

IEEE Log Number 8930349.

A. Preprocessing

Preprocessing of images is an important component of stereopsis. During this stage image locations satisfying certain well-defined feature characteristics are identified in each image. They have to be chosen carefully because the subsequent matching strategy shall make extensive use of these feature characteristics.

Some of the earlier stereo algorithms used area-based matching schemes in which area patches from two images were matched [18], [48]. Points of interest were located in one image using certain *interest operators*. Moravec [48] proposed one such interest operator that computed the local maxima of a directional variance measure over a 4×4 (or 8×8) window around a point. The sums of squares of differences of adjacent pixels were computed along all four directions (horizontal, vertical, and two diagonal), and the minimum sum was chosen as the value returned by the operator. The site of the local maximum of the values returned by the interest operator was chosen as a feature point whose stereo counterpart was to be found.

By and large most of the contemporary stereo algorithms match features directly rather than areas (in Section II-B we shall examine the issues regarding area-based and feature-based matching). Hence, the importance of good feature detectors has increased. Since physical discontinuities in a scene usually project as local changes in gray-level intensity in an image, edges have been increasingly used as matching primitives. A large number of edge operators have been proposed that compute the direction of orientation as well as the strength of an edge. Most of the edge operators currently in use can be classified [4] into three main categories:

1) Operators that approximate certain mathematical derivative operators (such as the Laplacian operator);
2) Operators that involve convolution of the image with a set of templates tuned to different orientations; and
3) Operators that fit local gray-level intensity values surrounding a point with (edge) surface models and extract edge parameters from the model.

The Marr–Hildreth edge operator [39] has been used by many algorithms for locating edge points during the feature extraction process. The operator convolves a mask approximating the Laplacian of Gaussian (∇^2G) function (see Section III-B) over the entire image and labels the zero-crossings of the convolution output as edge points. The edge orientation on a zero-crossing contour is given by the direction of the gradient of the convolution output. The edge strength is proportional to the magnitude of the gradient of the convolution output. Recently, Torre, and Poggio [73] have also studied the problem of using differential operators for edge detection.

Grimson [19]–[21], Mayhew and Frisby [44], [59], Kim and Aggarwal [35], and Ayache and Faverjon [1], Ayache and Lustman [2], among others, use the Marr–Hildreth operator (or a simplified version thereof) for feature point extraction. The edge detectors proposed by Canny [11], and Deriche [13] are also used fairly widely for low-level feature extraction. Baker and Binford [3] and Ohta and Kanade [54] locate peaks of the magnitude of the first derivative of the intensity profile along a scan line as feature points for matching. Some of the other popular gradient edge detectors are the Roberts, the Sobel, and the Prewitt operators [4]. Haralick [27] has proposed a step-edge detector based upon the second directional derivative, and compared its performance with the Marr–Hildreth zero-crossing detector and the Prewitt gradient operator. [22] and [26] contain interesting discussions about the comparison of the Marr–Hildreth and the Haralick edge operators.

Medioni and Nevatia [46] used a set of oriented step-edge masks (Type II) spaced at 30° intervals to extract edge points. Ito and Ishii [30], Harwood and Peitikäinen [57], and others have used Type (II) edge operators consisting of eight template masks tuned to the eight directions of the compass. The mask giving the maximum output decides the orientation and magnitude of the edge. The edges obtained using this type of operators need further processing in the form of edge thinning and edge linking.

Raju, Binford, and Shekhar [62] have used an operator of Type (III) described in [50] to detect an *edgel* (an edge element). The edgel operator fits a directional *tanh*-surface to a window in the image. Edgels are characterized by their position and orientation. Ballard and Brown [4], and Rosenfeld and Kak [68] contain a more in-depth treatment of edge detectors.

Linear edge segments have also been used as matching primitives for stereo by Medioni and Nevatia [46], Ayache and Faverjon [1], Ayache, and Lustman [2], Hansen, Ayache, and Lustman [25] and others. In the segment-based stereo algorithm of Medioni and Nevatia [46] edge points were extracted using Type (II) edge operators and fitted with piecewise-linear edge segments using the Nevatia–Babu algorithm [52]. Each edge segment description consisted of the coordinates of its endpoints, its orientation, and the average contrast (absolute value) in gray-level intensity along a direction normal to its orientation. Ayache and Faverjon [1] obtained edge points by using two methods. The first involved locating the zero-crossings in the output of the convolution of the intensity image with a difference-of-averages filter, and connecting them to obtain a chain of edge points. Then, the magnitude of intensity gradient along each chain was computed by the Sobel operator and portions of chains having connected points with the magnitude of intensity gradient below a certain threshold were discarded. The second method used a modified version of the Canny edge detector [11]. In both cases, the intermediate results were chains of connected edges. Each chain of connected edges was then approximated by a set of linear edge segments using a polynomial approximation algorithm [56]. Each of the resulting edge segments was described using the coordinates of its midpoint, its length, and its orientation.

Thus, two of the major types of features extracted from images are edge points and line segments.

B. Matching

Matching is perhaps the most important stage in stereo computation. Given two (or more) views of a scene, correspondence needs to be established among *homologous* features, that is, features that are projections of the same physical identity in each view. Matching strategies can be differentiated in the broadest sense according to the primitives used for matching as well as the imaging geometry. Differences in the matching primitives separate area-based matching from feature-based matching. Imaging geometry creates distinctions that separate parallel-axis stereo from nonparallel axis stereo, and binocular stereo from trinocular (and other multiocular) stereo paradigms. Local search procedures for possible matches are governed by the projection geometry of the imaging system, and are expressed in terms of the epipolar constraints. Various local properties of the features to be matched are used in order to achieve a reasonable amount of success in the *local matching* process. The *global consistency* of the local matches is then tested by figural continuity[1] or other similar constraints.

Area-based stereo techniques use correlation among brightness (intensity) patterns in the local neighborhood of a pixel in one image with brightness patterns in a corresponding neighborhood of a pixel in the other image. First, a point of interest is chosen in one image. A cross-correlation measure is then used to search for a point with a matching neighborhood in the other image. The area-based techniques have a disadvantage in that they use intensity values at each pixel directly, and are hence sensitive to distortions as a result of changes in viewing position (perspective) as well as changes in absolute intensity, contrast, and illumination. Also, the presence of occluding boundaries in the correlation window tends to confuse the correlation-based matcher, often giving an erroneous depth estimate.

Feature-based stereo techniques use symbolic features derived from intensity images rather than image intensities themselves. Hence, these systems are more stable towards changes in contrast and ambient lighting. The features used most commonly are either edge points or edge segments (derived from connected edge points) that may be located with subpixel precision. Also feature-based methods allow for simple comparisons between attributes of the features being matched, and are hence faster than correlation-based area matching methods.

Stereo matching paradigms are also characterized by the particular imaging geometry being used. Factors that could be changed include, but are not limited to, the mutual orientation of the optical axes of the cameras (either parallel or nonparallel) and the number of cameras used (either two or more than two). The imaging geometry of a conventional stereo imaging system involves a pair of cameras with their optical axes mutually parallel and separated by a horizontal distance denoted as the stereo baseline. The cameras have their optical axes perpendicular to the stereo baseline, and their image scanlines parallel to the baseline (horizontal). Since the displacement between the optical centers of the two cameras is purely horizontal, the position of corresponding points in the two images can differ only in the horizontal component. Fig. 1 shows the imaging geometry of a stereo pair of cameras. The two cameras are represented by their equivalent pinhole approximation models with their image planes, I_L and I_R, reflected about their centers of projections, O_L and O_R, respectively. The origin of the world coordinate system is at O_L, the effective focal length of each camera is f, and the stereo baseline is b. The world coordinate axes X_W, Y_W, and Z_W coincide with the coordinate axes of the left camera, X_L, Y_L, and Z_L, respectively. Let $P_L(x_L, y_L, z_L)$ and $P_R(x_R, y_R, z_R)$ be the projections of the 3-D scene point $P(x, y, z)$. The rays of projection $\overline{PO_L}$ and $\overline{PO_R}$ define the plane of projection of the 3-D scene point called the epipolar plane. For a given point P_L in the left image, its corresponding match point P_R in the right image must lie on the line of intersection of the epipolar plane and the image plane that is called the epipolar line. The epipolar line in the right image corresponding to a point P_L in the left image defines the search space within which the corresponding matchpoint P_R should lie in the right image. Thus the epipolar constraint is obtained as a result of the imaging geometry of the stereo camera system and helps limit the search space in the correspondence problem for stereo analysis. In the conventional parallel-axis geometry, all epipolar planes intersect the image planes along horizontal lines.

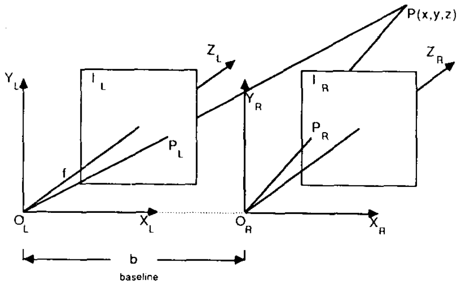

Fig. 1. Parallel axis stereo geometry.

However if the optical axis of any one of the cameras were not parallel to the world z-direction, then the epipolar lines in the image would appear inclined to the horizontal. Fig. 2 depicts a special case when the coordinate axes (Z_R and X_R) of only the right camera have been rotated by a pan angle ϕ (about the y-axis, Y_R) to Z_R' and X_R', respectively. Then the epipolar lines in the right image L_1, L_2, and L_3, corresponding to points P_{1L}, P_{2L}, and P_{3L} intersect at point E_R known as the epipole center of the right image. In general, the coordinate system of each

[1] The concept of figure continuity constraint and its various interpretations are discussed in Section III.

camera could have a pan angle ϕ (about the world y-direction), a tilt angle θ (about the world x-direction) as well as a roll angle α (about the world z-direction). Barnard and Fischler [4] contains a more detailed description of image acquisition and camera modeling.

Thus, extra epipolar line computations become necessary in the case of nonparallel imaging geometry. The advantage of nonparallel imaging geometry is that it allows for a greater overlap of the left and right images of the scene being observed. The epipolar search for matching edge points is usually aided by certain geometric similarity constraints like similarity of edge orientation or edge strength. This matching process is also referred to as local matching.

The match points obtained as a result of imposing the epipolar constraint on the local matching search could result in two or more candidate matches being judged as having almost equal possibility for getting matched. Or worse, an incorrect match point might satisfy the local matching constraints (epipolar constraint and geometric property constraint) and get chosen as a good match. The disparity obtained by computing the relative displacement of the matching feature points in the two images is used to extract the 3-D depth of the scene point that projects on the two matched points. Thus, if certain assumptions can be made regarding the nature of surfaces in the 3-D scene being observed, they could be used to determine the consistency of the disparities obtained as a result of the local matching, or guide the epipolar search so as to avoid inconsistent/false matching. An inherent assumption that is usually made about objects is that their surfaces are predominantly smooth. The smoothness in depth is expected to result in the smoothness of disparities obtained as a result of the matching process. This is formulated in the form of a regional disparity continuity constraint. Also the contours on the scene surface project on each image as continuous (or piecewise continuous) curves, which is the motivation behind the figural continuity constraint. Hence physical features on objects that satisfy the surface smoothness assumption and project on the stereo pair of images as image features would satisfy some form of the disparity continuity and figural continuity constraints. This is otherwise referred to as global matching. Thus, local matching and global matching can be regarded as two phases of the stereo matching process.

C. 3-D Structure Determination

The conventional parallel, axis stereo geometry provides a disparity value d for each matched pair of points $P_L(x_L, y_L)$ and $P_R(x_R, y_R)$ (see Fig. 1) as, $d = x_L - x_R$. By considering similar triangles, the world coordinates of the scene point $P(x, y, z)$ can be easily obtained as

$$x = \frac{bx_L}{d}, \qquad y = \frac{by_L}{d}, \quad \text{and} \quad z = \frac{bf}{d}$$

where b is the stereo baseline and f is the effective focal length of the camera.

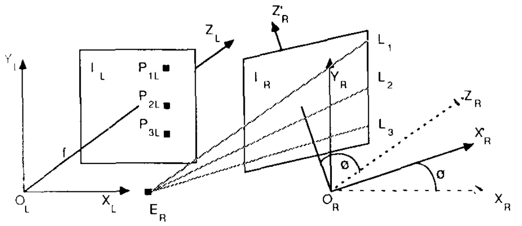

Fig. 2. Nonparallel axis stereo geometry.

The 3-D reconstruction process of nonparallel stereo systems ([1], [2], [30], [76]) requires a more general approach, since closed form solutions may not exist for many cases. The lines joining the center of projection and the image point in each of the stereo images are projected backwards into space. Then the point in space that minimizes the sum of its distance from each of the back-projected lines is chosen as the estimated 3-D position of the matched point. For nonparallel imaging systems using edge segments as matching features ([1], [2]), the end points of the matched edge segments are projected backwards in space and the 3-D position of the segment is determined using a similar minimization criterion.

III. Computational Theory of Stereopsis

Marr and Poggio [41] proposed a feature-point based computational model of human stereopis. Grimson [19], [21] developed a computer implementation of their algorithm and demonstrated the effectiveness of this model on standard psychological test images (random dot stereograms) as well as on natural images. A number of additional psychophysical predictions of the Marr–Poggio model have been tested and several modifications have been proposed [17], [44], [49], [59]. After extensive testing, Grimson [20] embodied these modifications in a newer version of his implementation. We shall briefly review the implementation of Marr–Poggio theory [19], examine the problems associated with it, and review the modifications which appear in Grimson's new implementation [20].

A. Marr-Poggio Theory

Marr and Poggio [41] based their computational structure of the stereo fusion problem upon biological evidence. Some of the early works by neurophysiologists and psychologists on subjects like existence of independent spatial-frequency-tuned channels [17], [31], [33], [45], cooperative processes [31], [32], [42], and vergence eye movements [63]–[65], [74], [75] in the human and other biological vision systems were used to formulate the outline of the theory.

The Marr–Poggio theory [41] proposed that the human visual processor solved the stereo matching problem in five main steps. 1) The left and right images are filtered at twelve different orientation-specific masks each approximated by the difference of two Gaussian functions with space-constants in the ratio 1:1.75. 2) Zero-crossings in the filtered images are found by scanning them along lines perpendicular to the orientation of the mask. 3) For each mask size, matching takes place between the zero-crossing segments extracted from each filtered image output that are of the same sign and roughly the same orientation. Local matching ambiguities are resolved by considering the disparity sign of nearby unambiguous matches. 4) Matches obtained from wider masks control vergence movements aiding matches among output of smaller masks; 5) The correspondence results are stored in a dynamic buffer called the 2.5-D sketch.

Marr and Poggio [41] formulate two basic rules for matching left- and right-image descriptions. Each item in an image can be assigned to one and only one disparity value (uniqueness). Secondly, matter is cohesive. Hence disparity varies smoothly almost everywhere, except where depth discontinuities occur at surface boundaries (continuity).

B. Grimson's Implementation

Grimson [19] implemented the computational theory of Marr and Poggio [41] and addressed certain implementation details that were not covered earlier by the Marr–Poggio theory.

1) Feature Extraction: Marr and Hildreth [39] have shown theoretically that, provided two simple conditions on the image intensity function in the neighborhood of an edge are satisfied, intensity changes occurring at a particular scale may be detected by locating the zero-crossings in the output of the $\nabla^2 G$ (Laplacian of Gaussian) filter. Instead of convolving each image with 12 directional *DoG* operators, each of which yield an approximation to the second directional derivative, Grimson [19] used the Laplacian of Gaussian ($\nabla^2 G$) operator and grouped the zero-crossing points in 12 directional bins. The precise form of the operator is given in polar coordinates (r, θ) by

$$\nabla^2 G(r,\theta) = \left[\frac{r^2 - 2\sigma^2}{\sigma^4}\right] \exp\left[\frac{-r^2}{2\sigma^2}\right] \tag{1}$$

where σ is the Gaussian space-constant. This is a rotationally symmetric function shaped like an inverted Mexican hat (Fig. 3). The width of the central negative region is given by $w_{2-D} = 2\sqrt{2}\,\sigma$. Grimson used three [20] or four [19] different sizes of filters for his images.

2) Matching: The algorithm begins with images filtered by the largest filters because the reduced density of zero-crossings makes matching easier. The overall matching strategy of Grimson [19] uses a coarse-to-fine iterative approach with disparities found at coarser resolutions used to guide match-point search at finer resolutions. Marr [38], [41] studied the probability distribution of the interval between adjacent zero-crossings of the same sign obtained from the convolution of random dot stereograms with the Laplacian of Gaussian filter. The results indicated that if the disparity between the images is less than $\pm(\omega/2)$, a search for matches within the range $\pm(\omega/2)$ will yield only the correct match with probability 0.95. However the alternate strategy of using a search space with range $\pm\,\omega$ is used by Grimson [19] since it allows one to search for matches over a larger disparity range and yet get unambiguous and correct matches with probability 0.5. In Grimson's implementation [19] for each zero crossing $P_L(x, y)$ in the left image, possible candidate matches $P_R'(x', y)$ are searched for along the epipolar line in the right image such that,

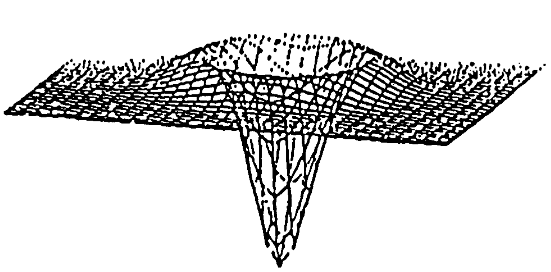

Fig. 3. 2-D Laplacian of Gaussian.

$$x + d_i - \omega \leqslant x' \leqslant x + d_i + \omega \tag{2}$$

as shown in Fig. 4(a), where d_i is the estimated disparity and ω ($= 2\sqrt{2}\,\sigma$) is the width of the *LoG* filter. Zero-crossings in the left and right images having the same contrast sign and approximately the same orientation (within ±30°) are matched. If only one match is found within the $\pm\,\omega$ region, then that match is accepted as unambiguous, and the disparity is recorded.

3) Disambiguation of multiple matches: If more than one match is found within the $\pm\,\omega$ region, then the one having disparity of the same type (convergent, divergent, or zero) as the dominant disparity in the neighborhood is accepted. Otherwise the match at that point is left ambiguous. This can be regarded as the pulling effect which is described in the psychophysical experiments of Julesz and Chang [32]. Each 2-D array of matched results is scanned and if the percentage of matched points is < 0.7 then all matches in that region are discarded.

C. Grimson's Modified Implementation of Marr–Poggio Theory

Grimson's earlier implementation [19] of the Marr–Poggio theory [41] imposes a regional continuity check on disparity. Later, Grimson [20] highlights some of the problems associated with the earlier implementation of the Marr–Poggio theory and presents a modified implementation.

1) Figural Continuity: Grimson's implementation [19] of the Marr–Poggio theory [41] used a regional continuity check on disparity in order to validate the matches. Grimson [20] observed that this caused difficulties in propagation of disparity at occluding boundaries between objects and along thin elongated surfaces. Elsewhere the matched feature points tended to form extended contours. Hence the figural continuity constraint of Mayhew and Frisby [44] that required continuity of disparity along contours was deemed more appropriate.

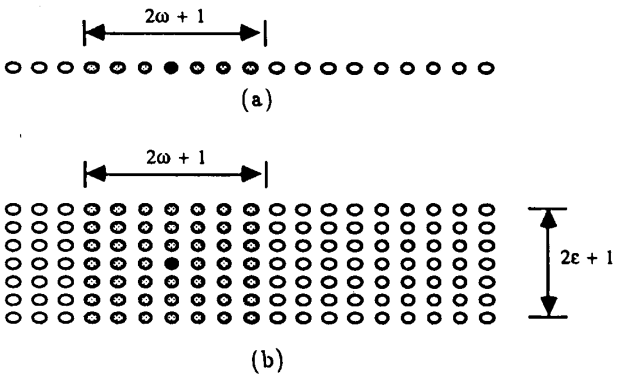

Fig. 4. Search space in Grimson's algorithm. $P_R'(x', y)$ is shown as black dot. Search space is shown as shaded dots around P_R'. (a) Original implementation. (b) Modified implementation.

Vertical Disparity: There is psychophysical evidence [15], [16], [53] to suggest that the human vision system does resort to eye movements in order to correct gross vertical misalignments in the images. Accordingly, the Marr–Poggio algorithm [41] uses a strict epipolar matching strategy (see Section III-B) after aligning the images in the vertical direction. However, local distortions due to perspective effects, noise in early processing, and discretization effects cause deterioration in matching performance at finer resolutions [20]. In the modified stereo algorithm, for a zero-crossing at a point $P_L(x, y)$ in the left image, Grimson [20] searches for the corresponding zero-crossing match points $P_R'(x', y')$ in the region

$$\{(x', y') \mid x + d_i - \omega \leqslant x' \leqslant x + d_i + \omega; \quad y - \epsilon \leqslant y' \leqslant y + \epsilon\} \qquad (3)$$

where ω and d_i denote quantities described in (2), and $(2\epsilon + 1)$ is the height of the search space in the vertical direction (see Fig. 4(b)).

D. Mayhew–Frisby Theory of Disparity Gradient

1) Figural Continuity: Mayhew and Frisby [44] propose a new interpretation of the surface continuity constraint, to include figural continuity. They extend the Marr–Poggio concept [41] of continuity to imply that edges of surfaces and surface markings would also be continuous resulting in continuity of disparity along figural contours. Baker and Binford [3], and Ohta and Kanade [54] also used a similar figural continuity constraint along with the added restriction of left-to-right ordering of edges for stereo matching.

2) Cross-Channel Activity: Mayhew and Frisby [44] postulate the existence of interaction between several spatial-frequency-tuned channels in parallel, as against the sequential coarse-to-fine process proposed by Marr and Poggio [41]. In simple terms, the rule for cross-channel correspondence requires that any feature attribute or pattern at a disparity location should be supported by a similar feature attribute or pattern in other spatial frequency channels within a certain disparity range, and that dissimilar cross-channel activity patterns should be rejected as figurally rivalrous.

In one stereo algorithm implemented by Mayhew and Frisby [44], the contrast-signed zero-crossings and peaks of the convolution of each image with the $\nabla^2 G$ operator are encoded at each location for three spatial-frequency tuned channels , as a triplet. Fig. 5 shows a schematic representation of using cross-channel activity according to Mayhew and Frisby [44]. The top row of Fig. 5 shows a triplet of measurement primitives found at a location in one image (say, left). Primitive values are marked +, − and · (nil) to signify positive, negative, or nonexistent zero-crossings, respectively. Beneath them are triplets at candidate match-points in the other image (say, right). The bottom row shows the result of the binocular cross-channel correspondence. Correct matches are marked M and incorrect (rivalrous) matches are marked R. If only one image has a primitive at a particular channel, the entry is marked U; and nil entries that match are ignored (marked ·).

3) Disparity Gradient Limit: Burt and Julesz [10] provide evidence supporting the claim that, for binocular fusion of random dot stereograms by the human visual system, the disparity gradient must not exceed 1. Pollard, Mayhew, and Frisby [59] suggest that for most natural scene surfaces, including jagged ones, the disparity gradients between correct matches is usually <1, whereas it is very rare among incorrect matches obtained for the same set of images.

Consider the binocular parallel imaging geometry as shown in Fig. 6 with image centers $O_L(x_{OL}, y_O)$ and $O_R(x_{OR}, y_O)$, separated by baseline b. Let $A_L(x_{AL}, y_A)$, $B_L(x_{BL}, y_B)$ in the left image and $A_R(x_{AR}, y_A)$, $B_R(x_{BR}, y_B)$ in the right image be the projections of the world points A_P and B_P, respectively. Then a cyclopean space is defined such that the origin $OC(x_{OC}, y_{OC})$ is defined as,

$$x_{OC} = \frac{x_{OL} + x_{OR}}{2} \quad \text{and} \quad y_{OC} = y_O. \qquad (4)$$

Let A_P and B_P have disparities $d_A = x_{AL} - x_{AR}$ and $d_B = x_{BL} - x_{BR}$, respectively, and cyclopean images A_C and B_C, respectively. The disparity gradient D_g would then be

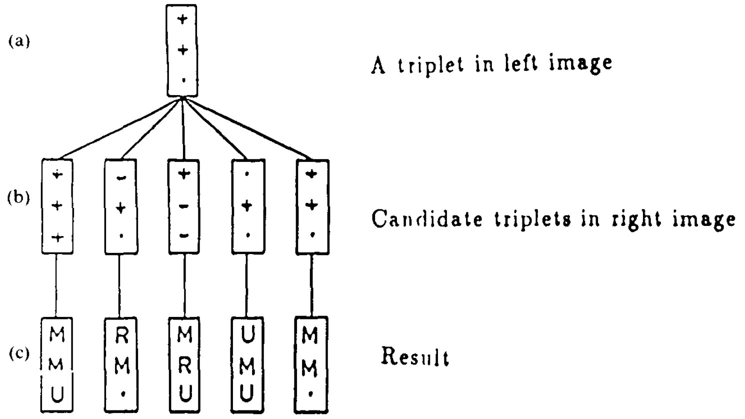

Fig. 5. Cross-channel correspondence.

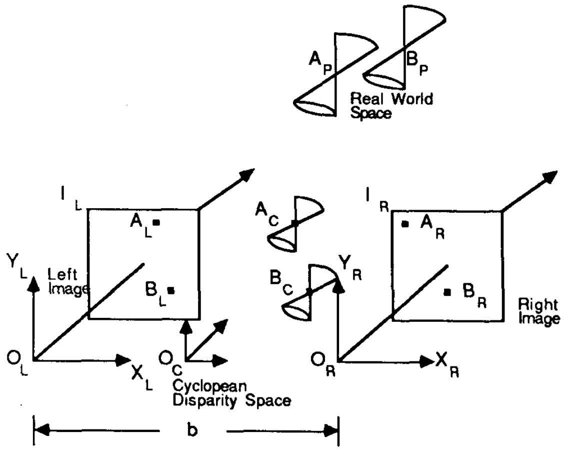

Fig. 6. Imaging geometry for PMF algorithm.

$D_g = (d_A - d_B)/d(A_C B_C)$, where $d(A_C B_C)$ is the cyclopean separation between A_C and B_C. A disparity gradient limit of 1 defines a cone-shaped forbidden zone for the point A_C in the cyclopean space, that is, any point within this forbidden zone violates the criterion for disparity gradient limit of 1.

Pollard, Mayhew, and Frisby [59] propose a new PMF algorithm that imposes a disparity gradient limiting constraint among correct matches. During the matching process, the matching strength of each potential match is evaluated as a sum of the support it receives from all potential matches in the neighborhood that satisfy the disparity limit criterion. In a match-pair of feature points (a_i, b_j), the support for the candidate match b_j is computed as a weighted sum of the number of potential matches in the neighborhood of b_j that have a disparity gradient less than 1, and vice-versa for the support of a_i. The PMF algorithm is interested only in the positive support to a potential match offered by surrounding matches that satisfy the within-disparity-gradient-limit criterion, and is unaffected by surrounding matches that exceed the disparity gradient limit. The uniqueness constraint is propagated using a discrete relaxation scheme such that if two image primitives a_i and b_j have the highest matching strength among their respective lists of candidate matches, then the match-pair (a_i, b_j) is considered as correct. Pollard, Mayhew, and Frisby [59] also show that since a disparity gradient limit in cyclopean space translates to a gradient limit in the real-world space it is possible even for planar surfaces to violate the disparity gradient limit criterion provided they have a sufficiently steep slope.

E. The Coherence Principle

Prazdny [61] has suggested the coherence principle to encompass the cohesiveness of matter [41] as well as the disparity continuity principles, which hold for opaque surface only. It recognizes the case of transparent objects. It allows the occurrence of a discontinuous disparity field if it is a result of several interlaced continuous disparity fields, each corresponding to a piecewise smooth surface. Two disparities facilitate each other if they possibly contain information about the same surface. When they do not interact at all they possibly contain information about different surfaces.

Prazdny [61] suggests one similarity function to quantify similarity between neighboring disparities. A Gaussian similarity function $s(i, j)$ is defined as

$$s(i, j) = \frac{1}{c|i - j|\sqrt{2\pi}} \exp\left[\frac{-|d_i - d_j|^2}{2c^2|i - j|^2}\right]. \tag{5}$$

This algorithm uses the quantity $|d_i - d_j|/|i - j|$ in the exponent of the Gaussian that is the disparity gradient used in Julesz [9]. However, in the Burt and Julesz algorithm increase in disparity difference results in inhibition of support, whereas in the coherence principle of Prazdny [61], there is no inhibition. The disparity gradient used by Prazdny [61] is also similar to that in the PMF algorithm [59].

F. Multifeature-Based Matching

Kass [34] proposes the use of matching coefficients obtained from a large number of uncorrelated (independent) measurements to contribute towards the local matching constraint. If the local matching constraint is chosen appropriately it is postulated that a large fraction of the points in the first image will have only one potential match in the second image making it unnecessary to use global consistency measures. Kass [34] has used a stochastic image model to substantiate this computational framework.

The local matching constraint proposed by Kass [34] relies on a representation of local intensity variation in the form of functionals $f_i(p, I)$, $1 < i < n$ for each point p in image I. No single image measurement (functional) is expected to contain all the information about the correspondence of a pair of image points. The functionals have been chosen to be orthogonal (low cross-correlation), linear, shift-invariant operators. At each point p, a vector $F(p, I) = (f_1(p, I), f_2(p, I), \cdots, f_n(p, I))$ is formed. Since each functional $f_i(p, I)$ in the representation defines a similarity measure for correspondence, $F(p_L, I_L) - F(p_R, I_R)$ is expected to be very small in each component, if p_L and p_R are truly matched. If p_L and p_R do not correspond, $F(p_L, I_L) - F(p_R, I_R)$ will most probably have at least one large component.

A predicate *matchp*(p_L, p_R) is defined such that *matchp*(p_L, p_R) is true if and only if,

$$|f_i(p_L, I_L) - f_i(p_R, I_R)| < k_i \sigma(f_i(p, I_L)), \quad \forall\, i \in \{1, 2, \cdots, n\}$$

where $\sigma(x)$ denotes the square root of the expected value of x^2 and k_i are appropriate scaling constants. (p_L, p_R) is considered a correct match if *matchp*(p_L, p_R) evaluates to true. In formulating matchp, Kass suggests one set of functionals $\mathscr{F}^* = \mathscr{F}_\sigma \cup \mathscr{F}_{\sigma s} \cup \mathscr{F}_{\sigma s^2}$ as the set of first and second partial derivatives of the Gaussian-smoothed images, with the sizes of the space constants being σ, σs, and σs^2, respectively. Each $\mathscr{F}_\sigma$ is the set of four partial derivatives with space constant σ as given by

$$\mathscr{F}_\sigma = \{\partial f_\sigma/\partial x, \partial f_\sigma/\partial y, \partial^2 f_\sigma/\partial x^2, \partial^2 f_\sigma/\partial y^2\}$$

f_σ being the Gaussian smoothing mask

$$f_\sigma = \frac{1}{s\pi\sigma^2}\exp\left[\frac{-(x^2+y^2)}{2\sigma^2}\right].$$

It is also shown that for synthetic stereo images derived from stationary Gaussian white noise if $s \geqslant 2.5$, the 12 functionals will have sufficiently low cross-correlations so that they can be regarded as approximately independent.

IV. Area-Based Stereo

Much of the earlier work done in stereo matching involves the use of correlation measures to match neighborhoods of points in the given images. Moravec [48] has used area-based correlation with a coarse-to-fine strategy to find corresponding match points. Initially feature points are identified in each image by the Moravec interest operator [48] that measures directional variance of image intensities in four directions surrounding a given pixel. Given a feature point P in one (source) image, the target image is searched at various resolutions ($\times 16$, $\times 8$, $\times 4$, and so on) starting from the coarsest. At each resolution the position in the target image that yields the highest correlation coefficient is enlarged to the next finer level of resolution. The process continues till the $\times 1$ resolution is reached. The same correlation process is applied to nine images taken two at a time to give 36 $({}^9C_2)$ possible stereo disparity values for each point of interest. The disparities and correlation coefficients are combined into a histogram, and a confidence measure is defined based upon the histogram peak. Matches with a confidence measure above a certain threshold are accepted.

Gennery [18] developed a high-resolution correlator that used the matches provided by the previous correlation matcher and produced an improved estimate of the matching point based upon the statistics of noise in the image intensities. This high-resolution correlator not only provided improved match points but also gave an estimate of the accuracy of the match in the form of variances and covariance of the (x, y) coordinates of the match in the second image.

Hannah [23] developed a correlation-based stereo system for an autonomous aerial vehicle. A modified Moravec operator is used to select control points. Autocorrelation in the neighborhood of a candidate match point is used to evaluate the goodness of a match. Subpixel matching accuracy is achieved through parabolic interpolation of correlation values. In Stereosys [24], Hannah has implemented a hierarchical correlation-based stereo system. Images of lower resolution (say, $n \times n$) are obtained by smoothing $2n \times 2n$ images by a Gaussian window and resampling. The points for which matches are to be searched for are picked by an interest operator as in [23]. A hill-climbing procedure is used to search for a match-point whose neighborhood results in a maximum in normalized cross-correlation with that of the original interesting point. Matches are propagated over the finer resolution images in the hierarchy. Matches found at the finest level are checked by reversing the role of left and right images, and repeating the hierarchical search starting from the just-found matching point. These initial matches are used to guide the match point search of neighboring points using the disparity continuity constraint. Finally the sparse density map is interpolated to construct a dense disparity map.

V. Relaxation Process in Stereo

Relaxation labeling is a fairly general model proposed earlier by Rosenfeld, Hummel, and Zucker [67] for scene labeling. In the paradigm of matching a stereo pair of images using relaxation labeling, a set of feature points (nodes) are identified in each image, and the problem involves assigning unique labels (or matches) to each node

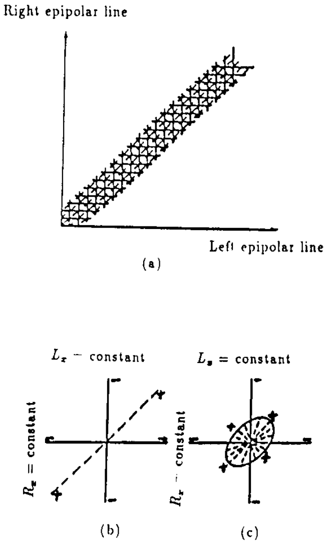

Fig. 7. Excitatory and inhibitory neighborhoods for Marr–Poggio cooperative algorithm. (a) Network of nodes for one scanline pair. (b) Linear excitatory neighborhood. (c) Disc-shaped excitatory neighborhood.

out of a discrete feature space (list of possible matches). For each candidate pair of matches, a matching probability is updated iteratively depending upon the matching probabilities of neighboring nodes so that stronger neighboring matches improve the chances of weaker matches in a globally consistent manner. This interaction between neighboring matches is motivated by the existence of cooperative processes in the biological vision systems postulated by Julesz [31], Julesz and Chang [32], Marr and Poggio [42], and others.

A. Marr–Poggio Cooperative Algorithm

Marr and Poggio [42] and Marr, Palm, and Poggio [40] have used the neighborhood information of matchable primitives in a simple iterative scheme. For each scanline pair in the stereo images (Fig. 7), a two-dimensional interconnected network of nodes (or cells) is set up. The horizontal and vertical connections are described as inhibitory, meaning all cells along each horizontal (or vertical) line inhibit each other, so that finally only one match remains on each horizontal (or vertical) line (uniqueness constraint). The diagonal connections are termed excitatory, meaning they favor diagonally adjacent matches to have the same disparity (disparity continuity). Fig. 7(b) shows the local disposition of the excitatory (+) and inhibitory (−) linkages in the neighborhood of a cell in the network. The bold lines (L_x = constant and R_x = constant) denote inhibitory interactions, and the dotted lines (diagonal, with slope = 1) denote excitatory interactions. A two-dimensional disparity continuity constraint can be effected by considering a disc-shaped excitatory neighborhood (Fig. 7(c)). In the cooperative process, let $C^t_{x,y;d}$ denote the state of a cell at time t corresponding to the coordinate (x, y) in the left-image matching $(x + d, y)$ in the right image. Initially the nodes which represent possible stereo match-points are loaded with 1's and all others are loaded with 0's. Thus, when the iterations begin, each cell adds the states of the neighboring excitatory potential matches in $S(x, y, d)$ (the excitatory neighborhood) to the previous state, and subtracts from it a weighted sum of the states of the neighboring inhibitory potential matches in $O(x, y, d)$ (the inhibitory neighborhood). This iterative updating can be represented by the relation

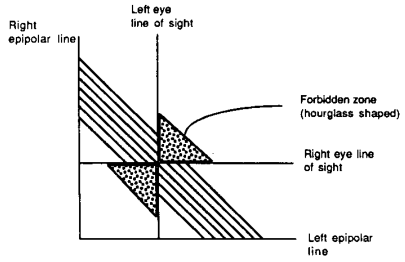

Fig. 8. Neighborhoods for Drumheller's implementation.

$$C^{t+1}_{x,y;d} = \sigma\left\{ \sum_{x',y';d' \in S(x,y,d)} C^t_{x',y';d'} - \epsilon \sum_{x',y';d' \in O(x,y,d)} c^t_{x',y';d'} + C^0_{x,y;d} \right\} \quad (6)$$

where, ϵ is the weighting factor for the inhibitor effect, and σ is the threshold function. This algorithm was shown to obtain stereo fusion for random dot stereograms successfully. It represents a very simple mechanism for the propagation of the uniqueness and the continuity constraints among neighboring match-points for disambiguation of multiple stereo matches in an iterative manner.

Drumheller and Poggio [14] mapped the previous cooperative stereopsis model of Marr, Palm and Poggio [40] on the Connection Machine [28], using the north–east–west–south (NEWS) mechanism for near-neighbor communication. The uniqueness constraint in Drumheller and Poggio's implementation [14] imposed an hour-glass shaped forbidden zone (see Fig. 8) and did not allow more than one match in the entire forbidden zone, unless the scene contained transparent or narrow occluding objects. Other variations of the cooperative model are proposed by Prazdny [61], Pollard, Mayhew, Porrill, and Frisby [60] and Marroquin [43]. Barnard and Thompson [7] and Kim and Aggarwal [35] have used the principle of cooperative processing to formulate the relaxation-based algorithms which incorporate more complicated disambiguating constraints.

B. Barnard–Thompson Algorithm

1) Computation of Feature Attributes: Barnard and Thompson [7] extract feature points (nodes) from each image using the Moravec interest operator. Each node a_i, at position $Z_l(x_i, y_i)$ in the left image L is assigned a set of labels L_i that represent the possible candidate matches

$Z_R(x_j, y_j)$ in the right image R within a disparity range. Every label set also contains a label l^* in the initial stage, which denotes undefined disparity. A node a_i has undefined disparity if point $Z_l(x_i, y_i)$ in image L does not correspond to any point in image R. Each label l of node a_i is assigned a weight function $w_i(l)$ that reflects the degree of similarity of intensity values in the neighborhoods of the candidate pair. An initial probability estimate $p_i^0(l)$ that the point $Z_l(x_i, y_i)$ in image L has a disparity l is then derived from the weight function $w_i(l)$.

2) Relaxation Process: The initial probabilities $p_i^k(l)$ computed from similarity in gray-level intensity values surrounding the match points are now updated iteratively to impose global consistency. That is, the probability $p_i^k(l)$ is increased if the neighbors of a_i have high probability values for disparities close to l. In particular, at the kth iteration, for a node a_i at (x_i, y_i) having neighbors a_j at (x_j, y_j), a quantity $q_i^k(l)$ is defined as

$$q_i^k(l) = \sum_{\|l - l'\| \leqslant 1} p_j^k(l')$$

where only those neighbors a_j are considered whose disparity label l' differs from l by $\leqslant 1$ in both $x-$ and $y-$ directions. The $q_i^k(l)$ serves as a measure of consistency of disparity in the neighborhood because it increases if more neighbors of $a(i)$ have disparities closer to l. The probabilities $p_i^k(l)$ are updated at the kth iteration as

$$\hat{p}_i^{k+1}(l) = p_i^k(l) * \left(a + b * q_i^k(l)\right), \; l \neq l^*$$

and

$$\hat{p}_i^{k+1}(l^*) = \hat{p}_i^k(l^*) \qquad (7)$$

where a is the rate constant to delay the suppression of unlikely labels (prevents $\hat{p}_i^{k+1}(l)$ from going to 0 if $q_i^k(l) = 0$), and b controls the speed of convergence.

The iterative procedure is continued either until the probabilities reach a steady state or a predetermined number of iterations are completed. This relaxation-based algorithm essentially imposes a disparity continuity constraint in the neighborhood of each matchpoint, favoring labels (disparities) consistent with the strong labels (disparities) occurring in the immediate neighborhood. This constraint is similar to the disparity continuity constraint over a region proposed by Marr–Poggio [41].

C. Kim–Aggarwal Algorithm

Kim and Aggarwal [35] propose a relaxation scheme that combines three disambiguating constraints, namely, continuity of disparity, figural continuity, and smoothness of probability (certainty) of matching. A conventional parallel-axis binocular setup is considered. Edge points are extracted by convolving each image with the *LoG* operator and locating the zero-crossings in the output.

1) Matching Primitives: A novel set of matching primitives is used. Depending upon the connectivity of the surrounding zero-crossings, 16 zero-crossing patterns are identified (see Fig. 9). A similarity measure is defined between two zero-crossing points depending upon the zero-crossing pattern surrounding each of them. The relaxation process is set up on lines similar to that explained in Barnard and Thompson [7]. The collection of all zero-crossings in the left image, which do not have horizontal patterns, form the set of nodes $\{a_i\}$. Each node is assigned a set of labels $L_i = \{l_j\}$ and a probability $p_i(l_j)$ that node a_i at point $Z_l(x_i, y_i)$ in left image L matches $Z_r(x_j, y_j)$ in right image R. A weight function $w_i(l_j)$ for node a_i with disparity l_j is computed based upon the similarity of the zero-crossing patterns as well as the difference in the intensity gradients. The initial probability $p_i^0(l_j)$ that node a_i has disparity l_j is computed using the weight functions $w_i(l_j)$.

2) Relaxation Process: A three-dimensional probability array is constructed on the zero-crossing map. The probability of matching of node a_i at $Z_l(x_i, y_i)$ to a point in R at disparity value l_j is stored in the point $(Z_l(x_i, y_i), l_j)$ in the 3-D array. In effect, it is a collection of 2-D arrays of probabilities corresponding to the zero-crossing map, with each 2-D array representing probabilities for one disparity value. $Z_l(x_f, y_f)$ and $Z_l(x_s, y_s)$ are the first and second neighboring zero-crossing points of $Z_l(x_i, y_i)$ (among the total 8 neighborhood points). The $p_i^k(l_j)$, $p_f^k(l_j)$, and $p_s^k(l_j)$ represent the entries in the 3-D array at positions $(Z_l(x_i, y_i), l_j)$, $(Z_l(x_f, y_f), l_j)$ and $(Z_l(x_s, y_s), l_j)$, respectively. The procedure for updating the matching probabilities is given by

$$p_i^{k+1}(l_j) = p_i^k(l_j) + c * F\left(p_i^k(l_j)\right) * \left(p_S^k\right) - d * p_i^k(l_j) * I(P_{FS}) \qquad (8)$$

where,

$$p_F^k = \max\left[p_f^k(l_j - 1), p_f^k(l_j), p_f^k(l_j + 1)\right]$$

$$p_S^k = \max\left[p_s^k(l_j - 1), p_s^k(l_j), p_s^k(l_j + 1)\right]$$

$$F\left(p_i^k(l_j)\right) = \begin{cases} \left[p_i^k(l_j)\right]^2 & 0 \leqslant p_i^k(l_j) \leqslant 0.5 \\ p_i^k(l_j) * \left(1 - p_i^k(l_j)\right) & 0.5 < p_i^k(l_j) < 1 \end{cases}$$

$$I(P_{FS}) = \begin{cases} 0 & p_F^k + p_S^k \neq 0 \\ 1 & p_F^k + p_S^k = 0. \end{cases}$$

The foregoing formula combines three constraints—disparity continuity, figural continuity, and smoothness of probability of matching. The function $F(p_i^k(l_j))$ controls the rate of convergence in two ways: 1) It reduces the tendency to converge fast to the most probable disparity value so that less-probable values may still have chances to compete; and 2) If all other conditions are the same, the magnitudes of increases of higher probabilities are higher. In other words, the third term in (8) implements the figural continuity criterion proposed by Mayhew and Frisby [44]. The p_F^k and p_S^k check the existence of nonzero probabilities for a match with disparity l_j in the connected neighborhood of $Z_l(x_i, y_i)$. If the connected zero-crossings do not have disparities within the disparity gradient limit ±1 of l_j, $I(P_{FS})$ is set and the probability $\hat{p}_i^{k+1}(l_j)$ is decre-

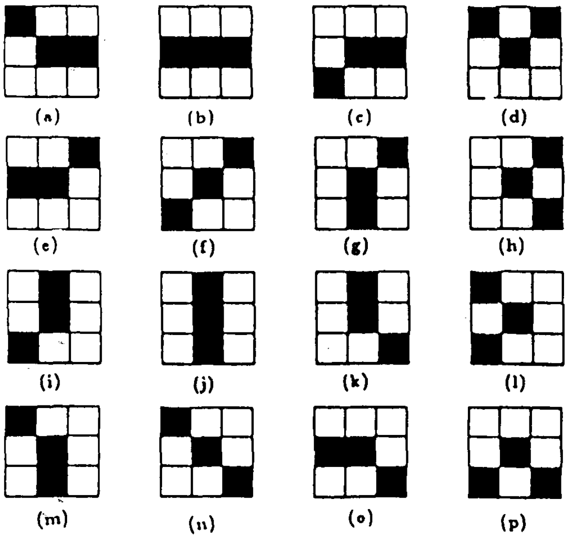

Fig. 9. Zero-crossing patterns.

mented. The second term in (8) reinforces the probability at $(Z_l(x_i, y_i), l_j)$ if the points in the neighborhood have a nonzero probability for disparities close to l_j by ± 1. This is similar to the area-based disparity continuity constraint which checks for surface smoothness and is in common with the implementation of Barnard and Thompson [7]. However the control of the rate of convergence using $F(p_i^k(l_j))$ is more flexible rather than a set of constants as used by Barnard and Thompson [7].

VI. Stereo Matching Using Edge Segments

The use of piecewise-linear approximations to connected edge points as matching primitives has been shown to be a viable alternative to matching of individual edge points ([1], [2], [25], [46]). Linear edge segments have certain advantages over single-edge points in the matching process. Firstly when edge points are grouped into a piecewise-linear segment, positional error at an isolated point has little effect on the position and orientation of the edge segment and most of the remaining edge points lie very close to the best fit. Secondly the edge connectivity constraint, which states that connected edge points in one image must match to connected edge points in the other image, must be imposed as an explicit disambiguating constraint while matching point-like features as against while matching line segments. On the other hand, due to possible fragmentation of edge segments during preprocessing, allowance has to be made for matching a single segment in one image with two or more segments in the other image, and vice versa.

A. Minimum Differential Disparity Algorithm

Medioni and Nevatia [46] describe a segment-based matching algorithm that uses a disparity continuity constraint called the minimal differential disparity criterion applied over neighboring edge segments.

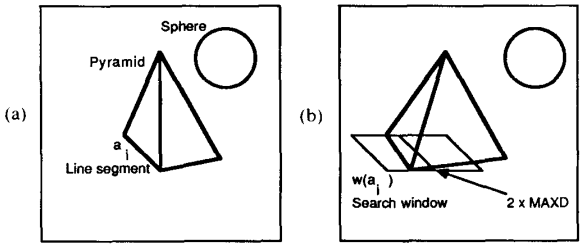

Fig. 10. Parallelogram-shaped search window. (a) Left image. (b) Right image.

1) Feature Extraction: The two stereo images, L and R, are brought into vertical alignment and a parallel-axis imaging geometry is assumed such that the epipolar lines run along horizontal scan lines. The feature-extraction stage described by Nevatia and Babu [52] is used to extract linear edge segments. Each edge segment is described by the coordinates of its end points, its orientation, and the average contrast in gray-level intensity (absolute value) along a direction normal to its orientation.

2) Matching Algorithm: Let $\mathscr{A} = \{a_i\}$ be the set of line segments in L, and $\mathscr{B} = \{b_j\}$ be the set of line segments in R. For each segment a_i in L, the search space is defined by a parallelogram-shaped window $w(a_i)$ in R whose one side is a_i and the other side is a horizontal vector of length $2 \times \text{MAX}\,D$, where $\text{MAX}\,D$ is the upper limit on the expected disparity (see Fig. 10). Similarly for each segment b_j in R, a window $w(b_j)$ is defined in L. Thus for a match (a_i, b_j), a_i lies in $w(b_j)$ and b_j lies in $w(a_i)$. Two segments x and y are said to be overlapping if by sliding either one of them along a direction parallel to the epipolar line, they can be made to intersect. Segments a_i in L and b_j in R can match only if a_i and b_j overlap, they have similar contrast in gray-levels, and have similar orientations. $\mathscr{S}_p(a_i) \subseteq w(a_i)$ denotes the set of all possible matches for a_i of L. A segment a_i in one image can be matched to two (or more) segments $b_{i1}, b_{i2}, \cdots, b_{in}$ in the other image provided none of the candidates $b_{i1}, \cdots, b_{in}$ overlap with each other.

An evaluation function $v^t(i, j)$ is computed iteratively to determine the merit of each match (a_i, b_j) as

$$v^{t+1}(i, j) = \sum_{a_h \in w(b_j)} \min_{b_k \text{ verifies } C_1(a_h)} \lambda_{ijhk}|d_{hk} - d_{ij}|/\text{card}(b_j) + \sum_{b_k \in w(a_i)} \min_{a_h \text{ verifies } C_2(b_k)} \lambda_{ijhk}|d_{hk} - d_{ij}|/\text{card}(a_i) \quad (9)$$

where λ_{ijhk} denotes the smaller of the overlap lengths for the match-pairs (a_i, b_j) and (a_h, b_k). The card (a_i) and card (b_j) are the number of segments in $w(a_i)$ and $w(b_j)$, respectively. Condition $C_1(a_h)$ allows for a_i and a_h to be matched to the same segment $b_j(= b_k)$ only if a_i and a_h do not overlap, and vice versa for condition $C_2(b_k)$. This allows for the possibility that if a_i and a_h are parts of a fragmented segment, they can get mapped to a single (unfragmented) segment b_j. The evaluation function $v(i, j)$ is updated during each iteration depending upon the dis-

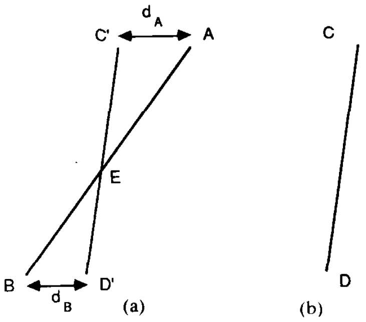

Fig. 11. Disparity change across linear segments. (a) Left image. (b) Right image.

parities between the segments neighboring a_i and b_j, and their respective preferred matches. For each segment a_h in the window $w(b_j)$ (recall, $w(b_j)$ defines a neighborhood of a_i) a preferred match b_k is found such that, $|d_{hk} - d_{ij}|$ is minimized. During the first iteration, the selection of b_k for each a_h is done from among the complete set $\mathscr{S}_p(a_h)$ since the set of preferred matches is empty. For each a_i, the match which yields the lowest $v(i, j)$ is chosen as the preferred match.

Since this matching algorithm minimizes the disparity difference among matched line segments in a neighborhood it is termed as the minimum differential disparity algorithm. This, in effect, imposes a condition that the matched line segment pairs, when reconstructed in space, form 3-D contours of surfaces that are smooth almost everywhere. Thus this matching algorithm has implemented the surface continuity constraint proposed by Marr and Poggio [41] for a paradigm of stereo matching that uses line segments as matching primitives.

Recently, Mohan, Medioni, and Nevatia [47] have proposed a scheme to detect and correct local segment-matching errors based upon disparity variation across linear segments. Let AB and CD (Fig. 11) be matching linear segments (or linear approximations to segments), and let $C'D'$ be the position of CD when superposed such that pixels with zero disparity coincide. Then it can be shown that disparity varies linearly along the length of the matched segments, and

$$\frac{d_A}{|AE|} = \frac{d_B}{|EB|} = \text{constant} \tag{10}$$

where d_A and $-d_B$ are the disparities associated with the points A and B, respectively.

Next we examine the matching algorithm of Ayache and Faverjon [1] that implements the disparity gradient limit approach for imposing a surface smoothness constraint on the reconstructed scene, and for disambiguation of false matches within the framework of segment-based matching.

B. Ayache–Faverjon Algorithm

Ayache and Faverjon [1] use descriptions of edge segments with the coordinates of the midpoint, the length of the segment, and its orientation for stereo matching. Unlike the minimum differential disparity algorithm [46], this method utilizes a generalized nonparallel axis imaging geometry and uses disparity between midpoints of matching line segments rather than average disparity between corresponding points that lie on matching line segments. A neighborhood graph is used to store the information regarding the adjacency of line segments in each image and a disparity gradient limit criterion (defined for line segments) is used to guide the global correspondence search. A neighborhood graph is constructed for each image using nodes to represent edge segments, and links to connect the nodes satisfying certain neighborhood relationships. Thus, each segment s_j has a list of neighbors that is obtained as a union of buckets of segments $\{b_k\}$ attached to windows $\{w_k\}$ that it intersects (see Fig. 12). The global matching stage uses a specialized representation of potential matches called the disparity graph. The idea is to use the disparity graph to propagate these matches within their neighborhoods to recover subsets of 3-D segments lying on a smooth surface patch.

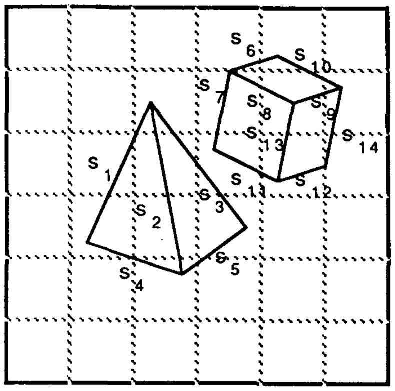

Fig. 12. Partitioning of segments into buckets. Examples of neighbors of segments are S_2: $\{S_1, S_3, S_4, S_5\}$ and S_3: $\{S_1, S_2, S_5, S_7, S_{11}\}$.

1) Local Matching Constraints: A pair of line segments a_i and b_j in the left and right images, respectively, constitutes a pair of potential matches if they satisfy the geometrical similarity constraint for line segments and their midpoints satisfy the epipolar constraint. A pair of edge segments whose length ratio and orientation difference lies below a preset threshold satisfies the geometrical similarity constraint. For the midpoint I_L of an edge segment a_i, a corresponding point I_R is searched for along the corresponding epipolar line near an expected disparity value. Ayache and Faverjon [1] compute disparity in the case of a pair of edge segments as follows (Fig. 13). If $\overline{P_R'Q_R'}$ (part of segment b_j) with midpoint I_L be a candidate match-segment for $\overline{P_LQ_L}$ (part of segment a_i) with center I_L, then the disparity d_{ij} between a_i and b_j is defined by

$$d_{ij} = E_R I_R - E_L I_L \tag{11}$$

where, E_L and E_R are the epipole centers in the left and right images, respectively.

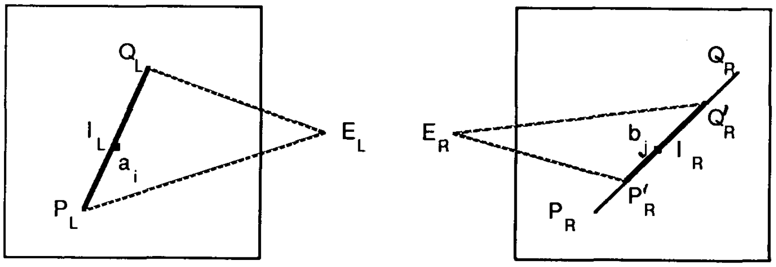

Fig. 13. Disparity for matched segment pair.

2) Global Matching Constraints: The global matching scheme of the Ayache–Faverjon algorithm [1] consists of a prediction and recursive propagation process. A disparity graph is constructed with nodes as pairs of potential matches (a_i, b_j) between the left and right images, and edges that connect pairs of nodes $(a_i, b_j), (a_i', b_j')$ that are adjacent segments in their respective neighborhood graphs. The allowable difference in disparity among neighboring nodes of matched pairs in the disparity graph is called the disparity gradient limit and corresponds to an ϵ variation in depth. For each node of the disparity graph (a_i, b_j), the neighborhood graphs of (a_i) and (b_j) are recursively explored for potential matched pairs that have disparities within the allowable disparity interval. Out of the potential matches the one with disparity closest to the predicted disparity is chosen. This favors those matches of line segments that make the 3-D scene maximally smooth in the sense of surface continuity as proposed by Marr and Poggio [41].

VII. Hierarchical Approaches to Stereo Matching

In this section we consider algorithms that utilize a hierarchical computational structure for stereo matching. The hierarchical structure of the algorithms allows matching information to be interchanged amongst various levels of matching computations, thus imposing global consistency in the disparity map. Apart from the Marr–Poggio-Grimson algorithm [19], [41] considered in Section III, the computational models of Terzopoulos [71], Hoff and Ahuja [29], and Lim and Binford [37] are some examples of hierarchical approaches to stereo matching.

A. Concurrent Multilevel Relaxation

Terzopoulos [71], [72] has developed an efficient multilevel relaxation computational model for low-level visual processing in concurrent mode. Conventional multigrid schemes employ recursive coordination of computations and flow of intermediate results starting from the coarsest level and proceeding successively to the finest level. Results at any level are used as approximations for the next level. With the advent of massively parallel architectures, such sequential algorithms result in inefficient use of hardware because most of the time is spent performing relaxations on only a single level, while processors at other levels (if configured in a multilevel architecture) remain idle. The concurrent strategy of Terzopoulos [71] maintains processors on all levels busy performing simultaneous relaxation operations. The concurrent strategy seeks to optimize a multilevel objective functional, with each term having three components: 1) A discrete version of the given functional at each level of a multigrid hierarchy; 2) an additive functional coupling each level (except the finest) to its next finer level; and 3) an additive functional coupling each level (except the coarsest) to the next coarser level. A concurrent multigrid algorithm for the problem of computing visible surface representations as formulated in [70] has been implemented.

B. Surfaces from Stereo: An Integrated Approach

Hoff and Ahuja [29] have argued in favor of integrating the steps of stereo matching and surface interpolation. Objects have faces that have a smooth variation in surface normals. Object surface meet on ridges that are smooth (or piecewise smooth) curves in 3-D space. They propose an integration of the matching and surface fitting processes in a way that the correctness of the choice of matches could be judged by the type of surface it produces.

Consider a stereo pair of $4n \times 4n$ images. Edge points are extracted using the Laplacian of Gaussian $(\nabla^2 G)$ operator at three resolutions $- n \times n$, $2n \times 2n$, and $4n \times 4n$. Initial matching is performed in both left-to-right and right-to-left directions. For each feature point P_i in, say, the left image a set of candidate match points $\{Q_i\}$ is selected from the right image according to similarity of local (or geometric) properties of feature points. A set of parameterized functions, planar and quadratic, are fitted to circular image regions centered at each grid point (x, y) in sequence. First, up to two planar patches are chosen at each grid point (x_i, y_i) that give the best least-squares fit-rating with the observed disparity z_i. Secondly, quadratic patches are fitted at each grid point to the above combinations of matches. The quadratic surface containing the most points is kept as the best fit for that grid point. Next, depth and orientation contours are detected by fitting bipartite planar patches and detecting discontinuities between the two halves. The bipartite planar patches are actually circular patches divided into two halves by a diameter with a given 3-D orientation. Finally a smooth surface is interpolated away from contours to yield a piecewise-smooth surface map at each resolution. Matching of edges at finer resolutions is guided by the interpolated surface at the coarser resolution.

C. From Objects to Surfaces to Edgels

In the hierarchical stereo algorithm proposed by Lim and Binford [37], matching begins at the highest level (objects). Results of matching are propagated to each successive lower level (surface boundaries, junctions, and edgels) and are used to guide the matching of lower-level features.

Edgels are detected using the Nalwa operator [50] that fits a tanh surface to each window in the image. Edges are linked into connected edges and curves (straight lines or conic sections) are fitted using best-fit criteria of Nalwa and Pauchon [51]. Surfaces are identified by tracing the boundaries of connected curves using both left-wall following as well as right-wall following strategies. Bodies are identified as groups of surfaces that share edges. The ordering information of surfaces in a body in the left-to-right as well as the top-to-down directions is saved to be used later as a matching constraint.

Matching of bodies is attempted at the highest level. Bodies that lie within the limits of corresponding upper and lower epipolar lines (i.e., having the same extent) are candidate matches. Multiple candidate matches are disambiguated using the left-to-right ordering of bodies along epipolar lines, the number of surfaces in the bodies, and the ordering of surfaces in the bodies. Next the system attempts to match surfaces that have the same extent and so on, down to edge segments and edgels. The advantage of this hierarchical stereo system is that the depth map obtained is already segmented and ready for surface interpolation.

D. Hierarchical Stochastic Optimization

Barnard [5] has implemented a solution to the stereo matching problem using a stochastic optimization technique called microcanonical annealing. Poggio, Torre, and Koch [58] have posed the stereomatching problem as imposing a regularization criterion on the stereo images,

$$\min:\ \epsilon = \int\!\!\int \Big\{ \big[\nabla^2 G \circ \big(I_L(x,y) - I_R(x + D(x,y), y) \big) \big]^2 + \lambda (\nabla D)^2 \Big\}\, dx\, dy \qquad (12)$$

where $I_L(x,y)$ and $I_R(x,y)$ are continuous intensity functions in the left and right images, respectively, $\nabla^2 G$ is the *LoG* operator, ∇D is the gradient of disparity, and λ is a constant. The first term of the integrand in (12) can be understood as a measure of the difference in image brightness values of corresponding points, and the second term as a measure of disparity gradient.

In the discrete version, (12) can be represented as minimizing the total potential energy $E = \Sigma E(i, j)$. Finding a disparity map $D(x, y)$ that results in the minimal energy constitutes a solution to the stereo correspondence problem. A stochastic optimization technique called microcanonical annealing using the Creutz algorithm [12] is used to control the combinatorial explosion of the search involved. Actually the Creutz algorithm [12] (microcanonical annealing) is a variation of the standard simulated annealing technique [36] used for solving combinatorial optimization problems. A coarse-to-fine method of computation speeds up the convergence process. At the coarser level, the number of pixel positions as well as the range of disparity is small. Hence a ground state can be reached quickly which can serve as an initial estimate for the next finer scale.

VIII. Stereo Matching by Dynamic Programming

Baker and Binford [3] use the Viterbi algorithm, a dynamic programming technique, to partition the stereo matching problem recursively based upon the constraint that a left-to-right ordering of edges is preserved along a scanline in a stereo image pair. In this edge-based technique, each edge is treated as a doublet, with a left half-edge and a right half-edge. The dynamic programming procedure is repeatedly applied for matching edge points on each scanline pair. The first and second passes of the Viterbi algorithm (preliminary edge correlation) match half-edges in the left image to those in the right image, and vice-versa. Next, a cooperative procedure uses an edge connectivity constraint to identify surface contours that are not continuous in disparity. That is, a connected sequence of edges in one image should match a connected sequence of edges in the other (both L-to-R and R-to-L). Finally, an intensity-based Viterbi correlation performed between intensity pixels from scanline intervals lying between the paired edges in the two images yields a denser depth map.

Ohta and Kanade [54] have also used pixel intensities of scanline intervals (delimited by edge points) to guide the *intrascanline* matching search by dynamic programming. This intrascanline search is formulated as a path-finding problem in a 2-D search space in which vertical and horizontal axes are the right and left scanlines, respectively. This is achieved by defining a cost function associated with each partial path based upon variances of gray-level intensities of the scanline intervals being matched. The edges are numbered from left to right on each scanline, with two ends of each scanline being also treated as nodes. If there are M nodes in the left scanline and N nodes in the right scanline, the solution to the intra-scanline search could be represented as a path comprised of a sequence of straight lines form node (0,0) to node (M, N) with the optimum cost. The cost of the optimal path from node (0,0) to node m is denoted by $D(m)$, and is the sum of the costs of its primitive paths. A primitive path between nodes k and m is a partial path that contains no vertices as in Fig. 14. The cost of the optimal path $D(m)$ is obtained by recursively adding the cost of each newly added primitive path to the already existing partial optimal path. The results of this intrascanline search are used to establish global consistency among matches achieved in neighboring scanlines using an *interscanline* search. The interscanline search is aimed at imposing a consistency

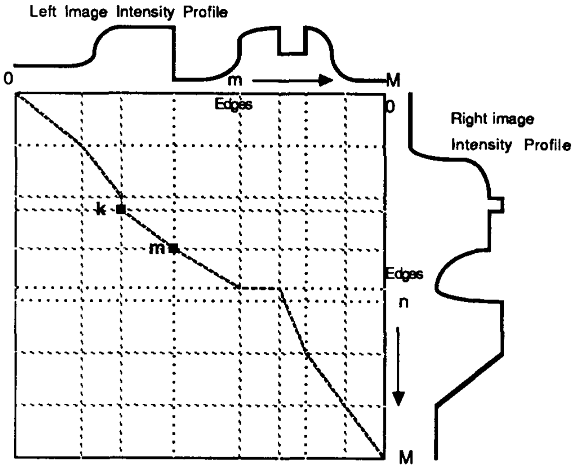

Fig. 14. 2-D search plane for intra-scanline search.

constraint among matches obtained at each scanline using edge connectivity. The problem is posed as that of finding the least-cost path between 3-D nodes in a 3-D search space. Each 3-D node is formed as a collection of the 2-D nodes connected across scanlines. The optimal path in the 3-D search space is obtained by recursively adding an optimal 3-D primitive path to the existing optimal partial path.

The approaches of Baker and Binford [3], and Ohta and Kanade [54] are based on the assumption that the ordering of edges remains unchanged for a stereo pair. The ordering of corresponding edges does not remain intact in scenes having large differences in depth, especially if these features were derived from thin, ribbon-like overlapping objects. Such a system is also liable to get confused in case of scenes with repetitive features especially if some of the features are missing in one of the images.

IX. Trinocular Stereo

The trinocular approach to the stereo problem has been proposed recently as an alternative means to conduct the correspondence search. The basic advantage of the third camera has been the extra epipolar geometry constraints offered by the three cameras. Provided the centers of projection of the three cameras are noncolinear, the true match points in the three images satisfy the condition that they must lie on the conjugate epipolar lines of the other two cameras. This allows for disambiguation of the multiple candidate matches that are found during local binocular-type correspondence search.

A. Edge-Based Trinocular Stereo

Yachida, Kitamura, and Kimachi [76] use an edge-based trinocular algorithm to obtain 3-D information about objects. Consider three cameras with centers of projection O_B, O_H, and O_V at known positions and with their optical axes having known orientations (Fig. 15). The *trinocular epipolar constraint* works as follows: For any point P_B in the image plane I_B, let there be multiple match point candidates $\{P_{H_1}, P_{H_2}, \cdots, P_{H_m}\}$ along the epipolar line l_{BH}. A set of epipolar lines $\{l_{HV_i}\}$ is constructed in I_V for each candidate $P_{H_i} \in \{P_{H_1}, P_{H_2}, \cdots, P_{H_m}\}$. At the intersection P_{V_i} of each l_{HV_i} and l_{BV}, the presence of an edge point P_{V_i} is tested. Each triplet (P_B, P_{H_i}, P_{V_i}) is tested for local similarity of feature attributes, and the best match is considered. In case some matching ambiguities still persist, the matchpoint candidate that yields a disparity closest to that of the points in the surrounding neighborhood is considered the best match. Ito and Ishii [30] have proposed a trinocular algorithm that uses a similar epipolar search procedure and a matching coefficient based upon the difference in gray-level intensity values in a 5 × 5 neighborhood of the candidate points. If any of the edge points do not get matched in the first pass due to occlusion, special one-sided matching coefficients are used in a second pass to match occluded points.

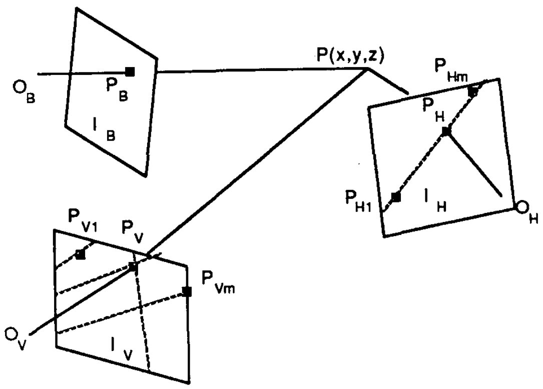

Fig. 15. Trinocular imaging geometry.

Ohta, Watanabe, and Ikeda [55] use a third camera and a relaxation procedure to improve the depth map obtained from binocular stereo. The camera geometry involves a left (L), a right (R), and an upper (U) camera, all having axes parallel to each other. The two image pairs $L-U$ and $L-R$ are processed independently using binocular stereo [54] to give two separate depth maps, H-depth and V-depth, respectively. The H-depth and V-depth values thus obtained are then combined into one depth image using a relaxation process. In this scheme, the trinocular geometry is used only to provide additional depth values that would be available from using two simultaneous binocular matching processes operating on mutually orthogonal epipolar lines.

Peitikäinen and Harwood [57] have used a three-view system with a parallel-axis geometry. The camera geometry involves a base camera (B) and two other cameras, H and V, displaced in the horizontal and vertical directions, respectively. Local features of the edge points like edge orientations and intensity contrast are used as local similarity attributes. In addition to the trinocular epipolar constraints, a postprocessing algorithm using connectivity of contours is also used to disambiguate multiple matches.

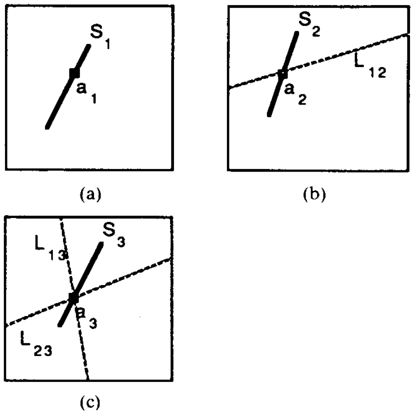

Fig. 16. Trinocular segment matching. (a) Image 1. (b) Image 2. (c) Image 3.

B. Segment-Based Trinocular Stereo

Ayache and Lustman [2], and Hansen, Ayache, and Lustman [25] have applied the segment-based (binocular) matching technique of Ayache and Faverjon [1] to three views.

In [2], Ayache and Lustman employ a prediction and verification scheme using neighborhood graphs of linear segments in three images that is an extension of the earlier binocular algorithm [1]. For any segment S_1 in image 1, if a triplet (S_1, S_2, S_3) can be found to satisfy the trinocular epipolar constraint of lines L_{12}, L_{23}, and L_{13} (see Fig. 16), and have sufficient similarity in local geometric properties then it is retained as a potential triplet.

Subsequently Hansen, Ayache, and Lustman [25] made further developments in the trinocular matching system using image rectification in the preprocessing stage. The original images are reprojected, as shown in Fig. 17 (for a binocular system), on a new image plane that is parallel to the plane containing the centers of projection of the cameras. As a result, the conjugate epipolar lines become parallel to each other in an image, and align themselves with the image coordinate frame. This reduces the search for matches to the horizontal and vertical lines, thus speeding up the matching process. A problem occurs when, due to some noise in preprocessing, a single segment (say, S_1) in the image 1 gets broken up into two (or more) segments, say S_2 and S_2' in image 2 (see Fig. 18). Hansen, Ayache, and Lustman [25] handle the problem by allowing flexibility in the order in which images are traversed for hypothesis generation. The problem of broken segments was also mentioned earlier by Peitikäinen and Harwood [57].

X. Structural Stereopsis

Boyer and Kak [8] have proposed the use of *structural descriptions* of image primitives and certain *information theoretic measures* defined on the basis of the structural descriptions to formulate the stereo matching problem.

The structural description of each image is derived from a skeletal or *stick-figure* representation of objects in the scene. The edges of the skeletons form a set of primitives over which the following binary scalar parametric relations are defined. *1*) *Pairwise Orientation*: The mean orientation of the straight line segment joining the centroids of a pair of skeletal edges. *2*) *Distance*: A function of the length of the straight line distance joining the two centroids of skeletal edge. *3*) *End-distance*: A function of the length of the straight line distance joining the closest pair of end points between two edges.

Boyer and Kak [8] have modified the exact matching approach developed by Shapiro and Haralick [69] in favor of an information theoretic approach for achieving inexact structural matching by defining interprimitive distance measures and relational inconsistency measures. The stereomatching problem is framed as a consistent labeling problem. The set of primitives in the left image $\boldsymbol{P} = \{p_i\}$ form the object set and the set of primitives in the right image $\boldsymbol{Q} = \{q_j\}$ form the label set. The labeling process utilizes two kinds of information: knowledge about the attributes of each label ($\mathscr{L}$) and knowledge about the relationship between labels ($\mathscr{R}$). $\mathscr{L}$ captures the information regarding the primitive distortion process (due to perspective effects as well as noise effects) and consists of a set of conditional probabilities of an attribute taking on a specific value in the right image, given its value in the left image. $\mathscr{R}$ captures the changes in the values of relational constraint parameters. It consists of a set of conditional probabilities, each item in the set being the probability that a relational parameter would take on a particular value in the right image having known its value in the left image. An event $p_i^{q_j}$ is defined as the left image primitive p_i being assigned to the right image primitive q_j in the stereo mapping h. The solution to the consistent labeling problem is considered optimal if the $\text{prob}[p_1^{k_1}, p_2^{k_2}, \cdots, p_n^{k_n}]$ is maximum, given the information in $\mathscr{L}$ and $\mathscr{R}$. That is

$$\max \text{OPM}: \text{prob}\left[p_1^{k_1}, p_2^{k_2}, \cdots, p_n^{k_n} | \mathscr{L}, \mathscr{R}\right] = \text{prob}\left[h | \mathscr{L}, \mathscr{R}\right]. \quad (13)$$

where, OPM is the optimal probability measure. The following basic assumptions are made in this probabilistic model: 1) Information in $\mathscr{L}$ is independent of the information in $\mathscr{R}$. This is based upon the idea that relational information is perceived by higher-level cognition processes that may be independent of lower-level processes required for the perception of primitive attributes. 2) The *a priori* probabilities of any particular event $p_i^{k_i}$ is constant, which translates to the fact that no advance information is available about the correct mapping function. Based upon these assumptions (13) becomes

$$\text{OPM} = \left(\text{prob}\left[\bigcap_i p_i^{k_i} \varepsilon \mathscr{L}\right]\right) \cdot \left(\text{prob}\left[\bigcap_i p_i^{k_i} \varepsilon \mathscr{R}\right]\right). \quad (14)$$

The two terms in the right hand side expression are referred to as the $\mathscr{L}$-term and $\mathscr{R}$-term. An information-theoretic interprimitive distance measure $\text{DIST}_h(\boldsymbol{P}, \boldsymbol{Q})$ is formulated to represent the dissimilarity between the sets of primitives $\boldsymbol{P}$ and $\boldsymbol{Q}$ under a specific mapping h: $\boldsymbol{P} \to \boldsymbol{Q}$,

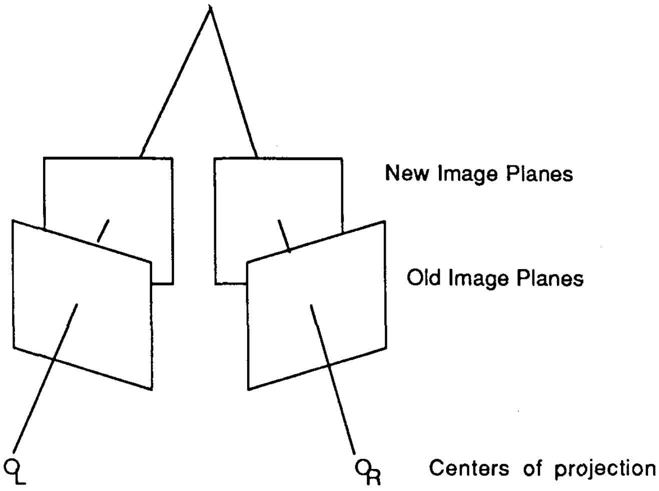

Fig. 17. Rectification of two images by reprojection.

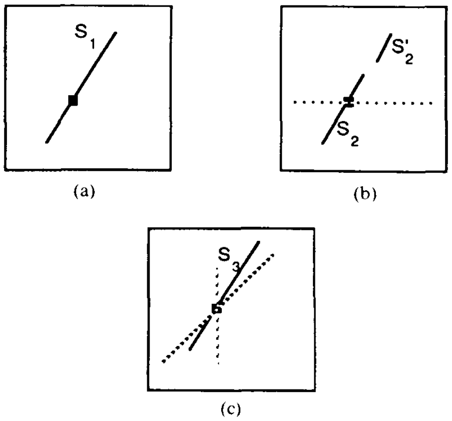

Fig. 18. Matching broken segments. (a) Image 1. (b) Image 2. (c) Image 3.

given the information $\mathscr{L}$ about the primitive attribute distortion process between the two images. Also, a relational inconsistency measure $\text{INC}_h(\boldsymbol{R},\boldsymbol{S})$ is formulated to measure the distortion of relational parameters between the sets of primitives ($\boldsymbol{R}$ and $\boldsymbol{S}$ represent parametric relations between elements of the sets of primitives $\boldsymbol{P}$ and $\boldsymbol{Q}$, respectively). The DIST_h and INC_h are defined to be $\text{DIST}_h(\boldsymbol{P},\boldsymbol{Q}) = -\log[\mathscr{L}$ term] and $\text{INC}_h(\boldsymbol{R},\boldsymbol{S}) = -\log[\mathscr{R}$ term]. Then (14) becomes

$$\text{Min: DIST}_h(\boldsymbol{P},\boldsymbol{Q}) + \text{INC}_h(\boldsymbol{R},\boldsymbol{S}). \tag{15}$$

The matching of the structural descriptions of the two images is performed in the following steps. For each primitive p_i in the left image, a match pool of potential primitives $\{q_j\}$ is obtained. This is achieved by accessing a look-up table of attributes and computing the distance between the two primitives. Any right primitive whose distance from a left primitive lies within a certain threshold is included in the match pool. The match pool for each left primitive is then stored in a best-first order. A *nilmap entry* is added as the last entry of any match pool if the cost associated with the best-fitting primitive in that pool were to exceed a certain threshold value, which signifies that the particular primitive in the left image may not have a matching primitive in the right image. Finally the consistent labeling problem is solved using a backtracking tree search. Out of the resulting list of possible mappings between the two primitive sets, the one that has the lowest value for $\text{INC}_h(\boldsymbol{R},\boldsymbol{S}) + \text{DIST}_h(\boldsymbol{P},\boldsymbol{Q})$ is chosen as the solution to the consistent labeling formulation of the stereo matching problem.

XI. Results and Discussion

In this section we shall review the experimental results of some of the stereo matching algorithms and the characteristics of the test images used therein. Testing of stereo algorithms has not been standardized as yet in the research community. Different algorithms have each been tested on different sets of stereo images. Without standardized test procedures, it is difficult to comment on the relative merits of stereo algorithms. However, one can identify the classes of images that have been used to test the algorithms, examine their performance, and know more about the domain of applicability of the algorithms.

The scene domains used for testing stereo algorithms have ranged from simple blocks world images to outdoor/aerial scenes. The block world images ([8], [30], [46], [54], [55], [76]) are typically scenes depicting an assortment of objects with polyhedral, cylindrical, conical, or spherical surfaces characterized by sharp physical boundaries and/or surface markings, all laid against a sharply contrasting background. Since the features being matched are few and most of them correspond to object boundaries, these images serve well as test images. Indoor (laboratory) scenes ([1], [2], [25], [57]) represent a higher degree of complexity in that the background of objects is no longer controlled. This makes the matching task more complicated. Also, many straight line edges (provided by doors, windows, and furniture) with their repetitive structure add to the complexity of the correspondence problem. The outdoor/aerial scenes are by far the most unstructured of scene domains and pose more complex matching problems.

Secondly, the task of computing the accuracy of depth estimates and the correctness of matches is plagued with the problem of lack of reliable ground truth measurements. For example, in the case of random dot stereograms, exact knowledge is available about the disparity value at each pixel in the stereo pair. Such exactness is seldom available in natural outdoor scenes or indoor laboratory scenes. Hence accuracy of the depth estimates is, at best, determined at a few selected points by actual measurement and compared with the results obtained from the stereo algorithm. Finally, stereopsis being a passive method, it suffers from the additional drawbacks namely, the problem of false matches and the sparseness of the resulting depth maps.

The test images used can be broadly classified into the following categories:

1) Psychophysical test patterns,
2) Indoor scenes,
3) Synthetic scenes,
4) Outdoor/aerial scenes.

We shall discuss, in brief, the experimental results obtained in each category.

A. Psychophysical Test Patterns

Grimson [19] used random dot stereograms to compare the performance of the computer implementation of the Marr–Poggio theory [41] of stereo fusion with the results of numerous psychophysical experiments conducted on the human vision system. Random dot stereograms are pairs of images, each consisting of two or more planar patches of random dots such that when a stereo pair is fused by humans a 3-D structure can be perceived. Since the disparity value is known at each pixel position for random dot stereograms, they can be used to test the correctness of the algorithm's performance. Typical 3-D structures used by Grimson [19], [21] include a square block rising out of a planar background, a series of square planes arranged on top of each other like a wedding cake, and a rectangular staircase pattern. For most random dot stereograms, a 50 percent dot density was used. Each stereogram was analyzed at four spatial channels with $w = 4$, 9, 17 and 35 pixels. Disparities obtained at coarser channels were used to guide the fusion at finer channels. In case of the pattern with a central square separated in depth from a second plane, out of 11 847 zero-crossing points only three (roughly 0.03 percent) were wrongly matched. Similar patterns with dot densities 25 percent, 10 percent, and 5 percent gave percent mismatch errors of 0.07 percent, 0.04 percent, and 0.06 percent, respectively. The wedding cake pattern (at 50 percent dot density) gave 0.06 percent mismatches. Almost all of the mismatches occurred at the boundary between the planes.

Grimson also found that the computer implementation [19] of the Marr–Poggio theory was in agreement with other psychophysical test results. Julesz [31] found earlier that the human vision system could perform binocular fusion even when one of the images of the stereo pair was blurred. The blurring caused the flat surfaces to be perceived by humans as slightly warped; nevertheless the 3-D structure was preserved. Grimson used Gaussian smoothing to blurr one image of the random dot stereogram before running the algorithm on the computer. The resulting disparity map was consistent with the 3-D structure, but it had a slightly higher number of errors in the reconstructed depth. In addition, Grimson studied [19] the effect of adding low-frequency as well as high-frequency noise to the random dot patterns. The results were in agreement with the psychophysical evidence found by Julesz and Miller [33] that stereo fusion is possible for noisy stereograms if the spectrum of the noise is sufficiently far from the spectrum of the pattern. In one example, high-frequency noise was added to one image such that the maximum magnitude of the added noise was twice that of the maximum magnitude of the original image. Results showed that matching was severely impaired for the smallest ($w = 4$) channel (17 percent wrong matches) whereas the next larger ($w = 9$) channel was only marginally affected (6 percent wrong matches).

Mayhew and Frisby [44] used stereograms of textured patterns in order to support the role of the figural continuity constraint in their computational theory of stereopsis. They report a significant (factor of 35) reduction in the ratio of potential false matches to the number of matchable points, after making explicit use of the figural continuity constraint.

Pollard, Mayhew, and Frisby [59] have tested the disparity gradient limit approach for imposing global consistency among matches using random dot stereograms. They report 98 percent correct matches for random dot stereograms that have disparity gradients up to 1.0. The matching performance degrades to about 50 percent correct matches for a disparity gradient of 1.8.

B. Indoor Scenes

The simplest of the indoor scenes are composed of a few objects scattered against a featureless (usually dark) background. Several stereo algorithms ([20], [30], [35], [46], [54], [55], [57], and [76]) have been tested on blocks world images. Grimson [20] presents results of matching stereo images of dark blocks placed against a bright background. In a typical blocks world scene, out of 2703 zero-crossing points as many as 1780 (65.9 percent) are reported to have been matched. The difficulty of matching blocks world images increases in the presence of occluding objects and repetitive features on the object surfaces. Medioni and Nevatia [46] and Ohta and Kanade [54] each report a matching example of a blocks scene (containing the Rubik cube) that has the aforementioned characteristics. Ohta and Kanade [54] compared the number of mismatches (or inconsistencies in matching) before and after the interscanline search for the blocks world scene. The global constraint imposed by the interscanline search was shown to reduce the number of mismatches by more than a factor 4. Ayache and Faverjon [1] and Medioni and Nevatia [46] have reported the matching result of a stereo image pair depicting an industrial part. The objects being viewed are essentially similar and provide a means for comparing the performance of stereo algorithms on common ground.

Indoor scenes of real-life laboratory environments have also been used for verifying stereo matching algorithms. Moravec's [48] stereo algorithm was used by a mobile cart to navigate its way around obstacles. The stereo system identified world position, the height of each obstacle, and the associated positional error caused by the pixel resolution of the camera and built an internal map of its immediate surroundings. The cart made successive short runs punctuated by halts during which the internal map was updated. The cart made successful runs in both indoor and outdoor environments. Kim and Aggarwal [35] tested their algorithm on indoor room scenes. Estimated depth was checked against actual (measured) depth at a few selected points. Percent error in depth varied between 0.17 percent and 3.7 percent. Percentage of false matches was as low as 2 percent for an optimum choice of parameters in the relaxation process. Ayache and Faverjon [1], and Hansen, Ayache, and Lustman [25] used indoor scenes for segment-based matching. The results report a maximum of 2 percent mismatches after applying global consistency validation.

C. Synthetic Scenes

Barnard and Thompson [7], Medioni and Nevatia [46], and Ohta and Kanade [54] have used a synthetic image (obtained from Control Data Corporation) for testing their algorithms. The correctness of matches was checked manually. Ito and Ishii [30] have tested their trinocular matching algorithm using a synthetic image of a pyramid-shaped block. Accuracy of depth estimates was found at selected points and compared with the actual depth. The process was then repeated with an actual block placed under similar conditions. In the experimental results with synthetic as well as real blocks world images of Ito and Ishii [30], the estimated maximum percent positional error of the selected 3-D points was within ±0.5 percent with the maximum percent measured error.

D. Outdoor Scenes

Grimson [20] tested his implementation of the Marr–Poggio theory on a number of aerial terrain images ("Phoenix" and "Ft. Sill" images). An interesting case depicts a highway interchange scene ("Boeing" image) that consists of a number of thin, elongated, and closely-spaced contours, each at different depths. The difficulty caused by such "spaghetti" contour scenes is evident by comparing the percentage matching errors of the highway aerial scene with that of the urban aerial scene (obtained from the Univ. of British Columbia, Vancouver) consisting mostly of buildings and a few roads. The same matching algorithm that resulted in 0.07 percent matching errors for the urban scene gave as high as 2.53 percent errors in the highway interchange scene. Ohta and Kanade [54] have also tested their algorithm on aerial scenes of the Washington D.C. area ("Pentagon" and "White House" images).

E. Discussion

One of the major differences among the different stereo algorithms discussed in this paper is the way they handle the global consistency of the matches obtained. As was mentioned earlier in Section II-B, a stereo algorithm can detect false positive matches obtained as a result of the local matching procedure by looking for other matches in the neighborhood that are consistent in disparity with that particular match. The disparity value for a given match is easily translated to a depth estimate by inverting the perspective projection equations. The prime motivation for imposing some sort of continuity constraint on the disparity values is that a mismatch would result in a disparity value that would translate into a strikingly discontinuous depth estimate as compared to the other neighboring points.

Marr and Poggio [41] have proposed a coarse-to-fine approach for propagation of the disparity continuity in the neighborhood of the matches. A purely region-based approach, as in [41], for imposing disparity continuity does not work very well when the scene is composed of a large number of thin, ribbon-like overlapping objects at various depths that partially overlap each other. This was recognized by many researchers like Grimson [20], Mayhew and Frisby [44], Kim and Aggarwal [35], Baker and Binford [3], and Barnard and Thompson [7], who among others have also included a figural continuity constraint in their stereo implementations. In the figural continuity constraint, a potential match (i, j) is favored if all their connected, neighboring matches (h, k) also have similar disparities. Figural continuity is used in segment-based matching in an implicit manner by which connected edge points are grouped together into segments and are matched as a group. Medioni and Nevatia [46] apply the minimal-differential-disparity rule for edge segments (a_i, b_j) by taking into account the disparity of each of the edge points in the segments and then taking an average disparity d_{ij}. In the segment-based matching of Ayache and Faverjon [1], for a match pair (a_i, b_j), the disparity d_{ij} does not explicitly take into account the disparities of the individual points in the edge segments but is computed using the positions of midpoint of a_i (I_L) and its corresponding potential match in segment b_j (I_R). The edge segments are treated as one unit, and the disparity of neighboring edge segment matches is constrained by a disparity gradient limit (defined specially for edge segments).

Kim and Aggarwal [35] have used a relaxation (cooperative) algorithm that uses a smoothness constraint on the probability of matching, in addition to the aforementioned constraints of regional continuity of disparity and figural continuity. Also, rather than using individual edge points or edge segments, they use 16 distinct edge (zero-crossing) patterns as matching primitives.

A disparity gradient limit approach is proposed in the PMF algorithm by Pollard, Mayhew, and Frisby [59] as an alternative to the figural continuity criterion, and it allows for matching of smooth as well as jagged surfaces.

The left-to-right (L-to-R) ordering of edges has also been used as a global matching constraint to disambiguate multiple matches and identify false positive matches. Baker and Binford [3], and Ohta and Kanade [54] have used the L-to-R ordering constraint in their dynamic programming algorithm to do the intrascanline search. In the Ayache–Faverjon algorithm [1], an L–R ordering relationship is used in building the disparity graph of edge segments from the left and right images. The disparity graph guides the prediction of matching hypotheses and thus controls the matching search. However, it must be noted that an L–R ordering constraint is not universally valid for guiding a binocular search. In the presence of transparent objects and/or thin, ribbon-like objects (also called spaghetti contours), differences in depth could result in the reversal of the L–R ordering of feature primitives on a scanline.

Apart from the factors discussed previously, the performance of various stereo algorithms can be dependent upon a lot of implementation details like the choice of threshold factors and the rate constants used to control the convergence of iterative algorithms. Also the behavior of a stereo algorithm in widely different scene domains needs to be understood carefully before choosing any one algorithm to be used in an application.

In this paper the authors have presented a broad review of the major recent developments in stereo algorithms. Experimental results of a variety of computational techniques have been grouped according to the scene domains on which the tests were conducted, and comparisons are made wherever possible. However it must be noted that the matching statistics like percentage error in depth, and percentage of mismatches mentioned in this paper appear as they were quoted in the respective technical publications of the said author(s) and were not observed under strictly identical conditions. Hence, if the reader is interested in building an application of a stereovision depth finder for a specific scene domain, a certain degree of caution needs to be exercised in the interpretation of the performance statistics and in understanding the trade-offs between various approaches.

XII. Conclusion

In this paper we have presented a review of the major techniques developed in the recent past for recovering the 3-D structure of a scene from analysis of stereo images. We have outlined the three main stages of stereo analysis, namely, preprocessing, establishing correspondence, and recovering depth. Based upon the differences in matching primitives as well as the imaging geometry being used, distinctions were made between area-based and feature-based matching, between parallel-axis and nonparallel axis stereo, between point-based and segment-based matching, and between binocular and trinocular matching.

We described the computational theory of stereopsis formulated by Marr and Poggio [41], which is motivated by a model of the human stereo vision system, and that formulates the basic constraints of uniqueness and regional continuity. Mayhew and Frisby developed further upon the figural continuity [44] and disparity gradient limit [59] criteria that impose global consistency constraints in order to disambiguate false matches. In the successive sections, we describe the different approaches developed for solving the stereo correspondence problem: area-based matching [18], [48], relaxation labeling [7], [35], dynamic programming [54], hierarchical approaches [5], [29], [37], [71], segment-based matching [1], [46], trinocular matching (edge-based [30], [55], [57], [76], as well as segment-based [2], [25]), and structural matching [8]. The performance of various approaches was discussed for different classes of test images and the difficulties involved in the evaluation of stereo algorithms were addressed.

The major issue involved in the stereo analysis of images is the correspondence problem. Algorithms need to be improved to give a lower percentage of false matches as well as better accuracy of depth estimates. Performance of algorithms needs to be evaluated over a broad range of image types in order to test their robustness. Most of the stereo work done so far has been limited to developing basic stereo matching capabilities for working with simplistic images. A great deal of research in stereo is needed in order to not only overcome the abovementioned difficulties but also to apply stereo techniques to solve more real-world problems.

Acknowledgment

The authors wish to thank E. Grimson and various anonymous reviewers for several insightful suggestions that enhanced the usefulness of this paper. They are grateful to J. J. Rodríguez and F. Arman for their helpful comments and for careful proofreading of this paper.

References

[1] N. Ayache and B. Faverjon, "Efficient registration of stereo images by matching graph descriptions of edge segments," *Int. J. Comput. Vision*, pp. 107–131, 1987.

[2] N. Ayache and F. Lustman, "Fast and reliable trinocular stereovision," in *Proc. 1st Int. Conf. Comput. Vision*, June 8–11, 1987, pp. 422–427.

[3] H. H. Baker and T. O. Binford, "Depth from edge and intensity based stereo," in *Proc. 7th Int. Joint Conf. Artificial Intell.*, Vancouver, Canada, Aug. 1981, pp. 631–636.

[4] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[5] S. T. Barnard, "Stochastic stereo matching over scale," in *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, pp. 769–778, Apr. 6–8, 1988.

[6] S. T. Barnard and M. A. Fischler, "Computational stereo," *Comput. Surveys*, vol. 14, no. 4, pp. 553–572, Dec. 1982.

[7] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 4, pp. 333–340, July 1980.

[8] K. L. Boyer and A. C. Kak, "Structural stereopsis for 3-D vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-10, no. 2, pp. 144–166, Mar. 1988.

[9] P. Burt and B. Julesz, "A disparity gradient limit for binocular fusion," *Science*, vol. 208, pp. 615–617, 1980.

[10] ——, "Modifications of the classical notion of Panum's fusional area," *Perception*, vol. 9, pp. 671–682, 1980.

[11] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Jan. 1985.

[12] M. Creutz, "Microcanonical Monte Carlo simulation," *Physical Rev. Lett.*, vol. 50, pp. 1141–1414, 1983.

[13] R. Deriche, "Using Canny's criteria to derive a recursively implemented optimal edge detector," *Int. J. Computer Vision*, vol. 1, no. 2, May 1987.

[14] M. Drumheller and T. Poggio, "On parallel stereo," in *Proc. IEEE Int. Conf. Robotics and Automation*, Apr. 7-10, 1986, pp. 1439–1448.

[15] A. L. Duwaer and G. van den Brink, "Diplopia thresholds and the initiation of vergence eye movements," *Vision Res.*, vol. 21, pp. 1727–1737, 1981.

[16] ——, "What is the diplopia threshold?," *Perception Psychophys.*, vol. 29, pp. 295–309, 1981.

[17] J. P. Frisby and J. E. W. Mayhew, "The role of spatial frequency tuned channels in vergence control," *Vision Res.*, vol. 20, pp. 727–732, 1981.

[18] D. B. Gennery, "Object detection and measurement using stereo vision," in *Proc. ARPA Image Understanding Workshop*, College Park, MD, Apr. 1980, pp. 161–167.

[19] W. E. L. Grimson, "A computer implementation of a theory of human stereo vision," *Phil. Trans. Royal Soc. London*, vol. B292, pp. 217–253, 1981.

[20] ——, "Computational experiments with a feature-based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 17–34, Jan. 1985.

[21] ——, *From Images to Surfaces: A Computational Study of the Human Early Visual System*. Cambridge, MA: M.I.T. Press, 1981.

[22] W. E. L. Grimson and E. C. Hildreth, "Comments on digital step edges from zero-crossings of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 121–126, Jan. 1985.

[23] M. J. Hannah, "Bootstrap stereo," in *Proc. ARPA Image Understanding Workshop*, College Park, MD, Apr. 1980, pp. 201–208.

[24] ——, "SRI's baseline stereo system," in *Proc. DARPA Image Understanding Workshop*, Miami Beach, FL, Dec. 1985, pp. 149–155.

[25] C. Hansen, N. Ayache, and F. Lustman, "High-speed trinocular stereo for mobile-robot navigation," in *Proc. NATO Adv. Res. Workshop Highly Redundant Sensor Systems*, Il Chiocco, Italy, May 16–20, 1988.

[26] R. M. Haralick, "Author's reply," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 126–128, Jan. 1985.

[27] ——, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 1, pp. 58–68, Jan. 1984.

[28] D. Hillis, "The connection machine," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., M.I.T., Cambridge, MA, 1985.

[29] W. Hoff and N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, no. 2, pp. 121–136, Feb. 1989.

[30] M. Ito and A. Ishii, "Three view stereo analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 524–532, July 1986.

[31] B. Julesz, *Foundations of Cyclopean Perception*. Chicago, IL: Univ. of Chicago Press, 1971.

[32] B. Julesz and J. J. Chang, "Interaction between pools of binocular disparity detectors tuned to different disparities," *Biol. Cybern.*, vol. 22, pp. 107–120, 1976.

[33] B. Julesz and J. E. Miller, "Independent spatial frequency tuned channels in binocular fusion and rivalry," *Perception*, vol. 4, pp. 125–143, 1975.

[34] M. Kass, "Computing visual correspondence," in *Proc. DARPA Image Understanding Workshop*, Arlington, VA, June 1983, pp. 54–60.

[35] Y. C. Kim and J. K. Aggarwal, "Positioning 3-D objects using stereo images," *IEEE J. Robotics and Automation*, vol. RA-3, no. 4, pp. 361–373, Aug. 1987.

[36] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Dordrecht, Holland: D. Riedel Publishing Co., 1987.

[37] H. S. Lim and T. O. Binford, "Stereo correspondence: A hierarchical approach," in *Proc. DARPA Image Understanding Workshop*, Los Angeles, CA, pp. 234–241, Feb. 1987.

[38] D. Marr, *Vision*. San Francisco, CA: Freeman, 1982.

[39] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Soc. London*, vol. B207, pp. 187–217, 1980.

[40] D. Marr, G. Palm, and T. Poggio, "Analysis of a cooperative stereo algorithm," *Biol. Cybern.*, vol. 28, pp. 223–229, 1978.

[41] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Royal Soc. London*, vol. B204, pp. 301–328, 1979.

[42] ——, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283–287, 1976.

[43] J. L. Marroquin, "Design of cooperative networks," AI Lab, Mass. Inst. Technol., Cambridge, MA, working paper 253, 1983.

[44] J. E. W. Mayhew and J. P. Frisby, "Psychophysical and computational studies towards a theory of human stereopsis," *Artificial Intell.*, vol. 17, pp. 349–385, 1981.

[45] ——, "Rivalrous texture stereograms," *Nature*, vol. 264, pp. 53–56, 1976.

[46] G. Medioni and R. Nevatia, "Segment-based stereo matching," *Comput. Vision, Graphics, Image Processing*, vol. 31, pp. 2–18, 1985.

[47] R. Mohan, G. Medioni, and R. Nevatia, "Stereo error detection, correction, and evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, no. 2, pp. 113–120, Feb. 1989.

[48] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, 1977, p. 584.

[49] P. Mowforth, J. E. W. Mayhew and J. P. Frisby, "Vergence eye movements made in response to spatial frequency filtered random dot stereograms," *Perception*, vol. 10, pp. 299–304, 1981.

[50] V. S. Nalwa and T. O. Binford, "On detecting edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 699–714, Nov. 1986.

[51] V. S. Nalwa and E. Pauchon, "Algorithms for edgel aggregation and edge description," in *Proc. DARPA Image Understanding Workshop*, Miami Beach, FL, Dec. 1985, pp. 176–185.

[52] R. Nevatia and K. Babu, "Linear feature extraction and description," *Comput. Graphics, Image Processing*, vol. 13, pp. 257–269, 1980.

[53] K. R. K. Nielsen and T. Poggio, "Vertical image registration in human stereopsis," AI Lab, Mass. Inst. Technol., Cambridge, MA, Memo 743, 1983.

[54] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 2, pp. 139–154, Mar. 1985.

[55] Y. Ohta, and M. Watanabe, and K. Ikeda, "Improving depth map by trinocular stereo," in *Proc. 8th Int. Conf. Pattern Recognition*, Paris, France, Oct. 27–31, 1986, pp. 519–521.

[56] T. Pavlidis, *Structural Pattern Recognition*. New York: Springer-Verlag, 1977.

[57] M. Peitikäinen and D. Harwood, "Depth from three camera stereo," in *Proc. IEEE CS Conf. Pattern Recognition*, Miami Beach, FL, June 22–26, 1986, pp. 2–8.

[58] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.

[59] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449–470, 1981.

[60] S. B. Pollard, J. E. W. Mayhew, J. Porrill, and J. P. Frisby, "Disparity gradient, Lipschitz continuity, and computing binocular correspondences," U. of Sheffield, Artificial Intelligence Vision Research Unit, Tech. Rep. 010, 1985.

[61] K. Prazdny, "Detection of binocular disparities," *Biol. Cybernetics*, vol. 52, pp. 93–99, 1985.

[62] G. V. S. Raju, T. O. Binford, and S. Shekhar, "Stereo matching using Viterbi algorithm," in *Proc. DARPA Image Understanding Workshop*, Los Angeles, CA, Feb. 23–25, 1987, pp. 766–776.

[63] C. Rashbass and G. Westheimer, "Disjunctive eye movements," *J. Physiology*, vol. 159, pp. 339–360, 1961.

[64] C. Rashbass and G. Westheimer, "Independence of conjunctive and disjunctive eye movements," *J. Physiology*, vol. 159, pp. 361-364, 1961.

[65] L. A. Riggs and E. W. Niehl, "Eye movements recorded during convergence and divergence," *J. Opt. Soc. Amer.*, vol. 50, pp. 913–920, 1960.

[66] G. Robinson, "Edge detection by compass gradient mask," *Comput. Graphics Image Processing*, vol. 6, pp. 492–572, 1977.

[67] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operation," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, pp. 420–423, June 1976.

[68] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic Press, 1976.

[69] L. G. Shapiro and R. M. Haralick, "Structural descriptions and inexact matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, no. 5, pp. 504–519, Sept. 1981.

[70] D. Terzopoulos, "Computing visible-surface representations," AI Lab, Mass. Inst. Technol., Cambridge, MA, Memo 800, 1985.

[71] ____, "Concurrent multilevel relaxation," in *Proc. DARPA Image Understanding Workshop*, Miami Beach, FL, Dec. 1985, pp. 156–161.

[72] ____, "Multilevel computational processes for visual surface reconstruction," *Comput. Vision Graphics Image Processing*, vol. 24, pp. 52–96, 1983.

[73] V. Torre and T. A. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 2, pp. 147–163, Mar. 1986.

[74] G. Westheimer and D. E. Mitchell, "The sensory stimulus for disjunctive eye movements," *Vision Res.*, vol. 9, pp. 749–755, 1969.

[75] R. H. Williams and D. H. Fender, "The synchrony of binocular saccadic eye movements," *Vision Res.*, vol. 17, pp. 303–306, 1969.

[76] M. Yachida, Y. Kitamura, and M. Kimachi, "Trinocular vision: New approach for correspondence problem," in *Proc. 8th Int. Conf. Pattern Recognition*, Paris, France, Oct. 27–31, 1986, pp. 1041–1044.

Umesh R. Dhond (SM'89) was born in Bombay, India, on March 13, 1964. He received the B.Tech. degree in electrical engineering from Indian Institute of Technology, Bombay, in 1985, and the M.S. degree in electrical and computer engineering from Louisiana State University, Baton Rouge, in 1987.

He is currently a Research Assistant at Computer and Vision Research Center, The University of Texas at Austin, and is working towards the Ph.D. degree. His research interests include computer vision, image processing, and artificial intelligence.

J. K. Aggarwal (S'62–M'65–SM'74–F'76) received the B.S. degree in mathematics and physics from the University of Bombay, India, in 1956, the B.Eng. degree from the University of Liverpool, England, in 1960, and the M.S. and Ph.D. degrees from the University of Illinois, Urbana, in 1961 and 1964, respectively.

He joined the University of Texas in 1964 as an Assistant Professor and has since held positions as Associate Professor (1968) and Professor (1972). Currently he is the John J. McKetta Energy Professor of Electrical and Computer Engineering and Computer Sciences at the University of Texas, Austin. Further he was a Visiting Assistant Professor at Brown University, Providence, RI (1968), and a Visiting Associate Professor at the University of California, Berkeley, during 1969–70. He has published numerous technical papers and several books, *Notes on Nonlinear Systems* (1972), *Nonlinear Systems: Stability Analysis* (1977), *Computer Methods in Image Analysis* (1977), *Digital Signal Processing* (1979), and *Deconvolution of Seismic Data* (1982). His current research interests are image processing and computer vision.

Dr. Aggarwal is an active member of IEEE Computer Society, ACM, AAAI, the International Society for Optical Engineering, the Pattern Recognition Society, and Eta Kappa Nu. He was Co-Editor of the Special Issue on Digital Filtering and Image Processing of the IEEE Transactions on Circuits and Systems, March 1975, and on Motion and Time Varying Imagery, IEEE Transactions Pattern Analysis and Machine Intelligence, November 1980, and Editor of the two-volume Special Issue on Motion of *Computer Vision, Graphics and Image Processing*, January and February 1983. He was the General Chairman for the IEEE Computer Society Conference and Pattern Recognition and Image Processing, Dallas, TX, 1981, and was the Program Chairman for the First Conference on Artificial Intelligence Applications sponsored by the IEEE Computer Society and AAAI, Denver, CO, 1984. Currently he is an Associate Editor of the journals *Pattern Recognition*, *Image and Vision Computing*, and *Computer Vision, Graphics and Image Processing*. Further, he is a member of the IEEE Transnational Relations Committee, member of the Editorial Board of *IEEE Press*, and the Chairman of the IEEE Computer Society Technical Committee on PAMI.