MODELING, IDENTIFICATION AND CONTROL, 1995, VOL. 16, NO. 2, 65–94

Approaches for stereo matching

A review†

TAKOUHI OZANIAN‡§

Keywords: *Computer vision, binocular stereo, feature selection, correspondence, constraints, matching, energy minimization, parallelism, trinocular stereo.*

This review focuses on the last decade's development of the computational stereopsis for recovering three-dimensional information. The main components of the stereo analysis are exposed: image acquisition and camera modeling, feature selection, feature matching and disparity interpretation. A brief survey is given of the well known feature selection approaches and the estimation parameters for this selection are mentioned. The difficulties in identifying correspondent locations in the two images are explained. Methods as to how effectively to constrain the search for correct solution of the correspondence problem are discussed, as are strategies for the whole matching process. Reasons for the occurrence of matching errors are considered. Some recently proposed approaches, employing new ideas in the modeling of stereo matching in terms of energy minimization, are described. Acknowledging the importance of computation time for real-time applications, special attention is paid to parallelism as a way to achieve the required level of performance. The development of trinocular stereo analysis as an alternative to the conventional binocular one, is described. Finally a classification based on the test images for verification of the stereo matching algorithms, is supplied.

1. Introduction

A major problem in computer vision systems is the recovery of geometric and physical properties of visible three-dimensional (3D) surfaces from two-dimensional (2D) intensity images, such as distance between the surfaces and the viewer, surface orientation and material properties (texture, reflectance, colour). The processes that provide such information about the 3D world are termed *early vision* (Poggio, Torre and Koch 1985, Bertero, Poggio and Torre 1988). Examples are edge detection, stereo matching, surface reconstruction, shape from shading and shape from texture. The results of the early vision processing are used for higher level tasks such as object recognition, object inspection, object manipulation, obstacle avoidance and navigation. The loss of information during the imaging process that projects the 3D world into 2D images often causes the solution of early vision problems to be non-unique (ambiguity), non-existing (occlusion), or not continuously dependent on the data. As a consequence early vision must rely on assumptions (constraints) about the physical world in order to lift the ambiguity in the inverse problem.

Received 5 April 1994.

† Review article invited by the Editor.

‡ Department of Engineering Cybernetics, Norwegian Institute of Technology, University of Trondheim.

§ Present address: Department of Computer Science, University of Hull, Hull HU6 7RX, United Kingdom.

The aim of this paper is to present a survey on the last decade's development of stereo matching as a passive means for inferring 3D information. It is to a large extent complementary to existing surveys (Barnard and Fischler 1982, Dhond and Aggarwal 1989) because of the following:

- known feature selection approaches are grouped and assessed according to certain criteria to elucidate the advantages and disadvantages of a particular choice.
- general methods for efficiently constraining the search for a correct and unambiguous solution are discussed, as are strategies for the whole matching process.
- difficulties in the matching process are considered and some approaches for the overcoming of possible errors are briefly mentioned.
- some new approaches for modeling the stereo correspondence problem in terms of energy minimization functions are described.
- attention is drawn to parallelism as an important way to achieve the required level of performance for real-time vision tasks, and therefore some parallel stereo algorithms as well as implementations are considered.
- trinocular stereo vision is discussed as an ongoing research trend and a powerful alternative to the conventional binocular stereo vision.
- classification of the test images for verifying the related algorithms is supplied.

Although all basic concepts and components of the stereo paradigm are briefly explained, some terms are introduced without explicit definition—for these, the reader is referred elsewhere (Barnard and Fischler 1982, Dhond and Aggarwal 1989). As for any scientific field, covering all aspects of the research in stereo vision is practically impossible (at least in a review like this). Reflecting the author's bias, the focus is put on the feature selection process and the matching process. The other components of computational stereo, such as image acquisition and disparity interpretation (reconstruction), in themselves constitute a wide research area, details of which go beyond the scope of this paper. The omission of biological stereo systems was motivated by similar reasons.

2. The computational stereo paradigm

Stereo vision has received a great deal of attention in the vision research area as a passive means for inferring 3D information from images. The 3D information can be computed by searching for corresponding items on a pair of images taken from slightly different viewpoints. The common area appearing in both images of the stereo pair is typically 40% to 80% of the total image area. The difference in position of corresponding items in the two images causes relative displacements called *disparities*. These disparities and the imaging geometry parameters enable depth to be calculated by triangulation. Typically the stereo approach is characterised by the following components:

- image acquisition and camera modeling—determining the level of resolution and precision, and the parameters of the imaging geometry sufficient to support the matching algorithm;
- feature selection—determining the primitive elements to be used as a basis for the matching algorithm;
- correspondence—matching those elements in the two views that are projections of the same element in the 3D world;
- disparity interpretation—converting the disparity in actual range information.

These components are described in more detail in the next sections.

2.1. *Image acquisition and camera modeling*

Key parameters associated with image acquisition are the absolute positioning accuracy and the relative alignment of the cameras (Barnard and Fischler 1982, Nishihara and Poggio 1984).

In general, camera systems can be modeled as transformations of three-dimensional coordinate systems. The transformations include translational, rotational, perspective, and scaling components. The translational component of the transformation is specified by the centre of projection (i.e. the optical centre of the camera) in the coordinate system. The rotational component is specified by a pan angle, a tilt angle and a roll angle. The distance from the origin of the camera coordinate system to the centre of projection is termed the focal length f.

The number (two or more) of cameras along with their position and orientation, i.e. the *imaging geometry*, are important factors for the determination of the disparity of corresponding items in the stereo images, and also for the computation of the epipolar lines (see the epipolar constraint in § 2.3.1).

Consider two laterally displaced cameras. There are two types of imaging geometries according to the orientation of the two cameras. In the first type, the cameras are oriented so that their optical axes converge at a certain point, called the fixation point. The disparities between the image features in this geometry can be either positive or negative. The sign of the disparity depends on the location of the physical surface point with respect to the fixation point (Fig. 1). Given a particular image feature F in the first image, its corresponding feature in the second image can lie to the left or to the right

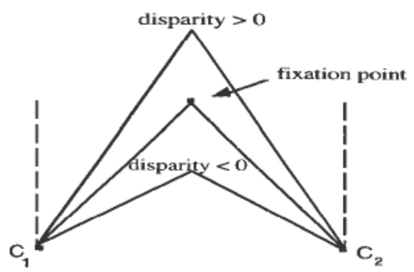

Figure 1. Convergent camera geometry.

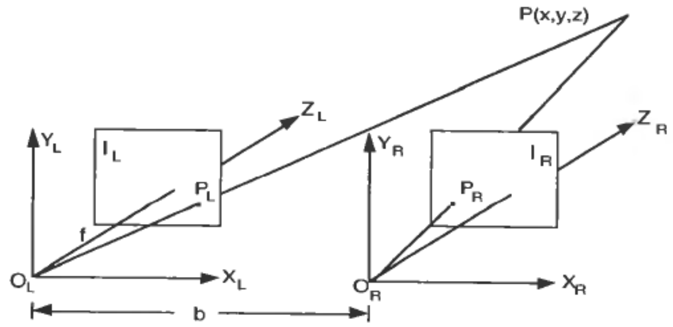

Figure 2. Conventional stereo geometry.

of F. Thus, a search for correspondence needs to be performed on both sides of F. Extra epipolar line computations become necessary in this case (Ayache 1991).

The second type of camera geometry is known as parallel axis geometry. The two cameras have parallel optical axes. This geometry can be regarded as having its fixation point at infinity. Here all the disparities must have the same sign, and hence, the search for correspondence can be restricted to only one side. Parallel axis geometry yields less overlap in both images of the scene compared to the convergent camera geometry, but it is the most frequently used one because of the simple epipolar line computation (see the epipolar constraint in § 2.3.1).

The geometry of a conventional stereo imaging system involves a pair of cameras with parallel optical axes and separated by a horizontal distance denoted as the stereo *baseline*. The optical axes are perpendicular to the stereo baseline, and the image scanlines are parallel to the baseline (Fig. 2).

The process of establishing the extrinsic (rotational and translational) and intrinsic (perspective and scaling) parameters of a camera system is called camera calibration. It gives the projection from 3D world coordinates to 2D image coordinates. This information is needed for two reasons: first, to determine the imaging geometry, and second, to reconstruct the 3D information from the 2D one after the matching process. Moreover, in a recent study Stewart (1992) has shown the importance of considering the image formation parameters such as focal length and baseline in the derivation of the matching constraints.

In general, there are two approaches to the computing of the camera parameters. One approach, called test-field calibration, assumes that 3D coordinates of some reference points are given (Faugeras and Toskani 1986, Tsai 1986). The other approach, called self-calibration, matches certain pairs of features in the two images (Faugeras, Luong and Maybank 1992, Olsen 1992).

2.2. *Feature selection*

An image array contains two pieces of information: light intensity changes and local geometry (position). A good feature selection approach must incorporate explicit representations of both pieces of information. The choice of matching features or primitives is very important because it can significantly reduce the combinatorial explosion of the search space associated with the stereo correspondence problem (see § 2.3.). Some estimation criteria for the selection of the primitives are (Ayache 1991):

- compactness: a representation of an image must be as concise as possible to reduce the complexity of the matching process;
- physical meaning: features should correspond to the projection into the image of physical phenomena; in particular, they should be invariant to change of viewpoint;
- stability: the representation should be insensitive to small photometric and geometric distortions arising from the stereo imaging;
- distinguishability: features should possess properties permitting discrimination amongst them to make the matching process easy; such properties can be geometric, intensity-based, or structure-based;
- precision: since the quality of localization of physical objects is dependent on the position of visual features the latter should be computable with precision;

- density: matching primitives should be sufficiently dense to represent all interesting surfaces in a scene;
- ease of computation: computational cost is especially important for real-time applications.

Some of the mentioned estimation parameters are mutually competing. For instance, very low-level features (features containing mainly intensity information) are dense but unstable and increase the complexity of the matching process. On the other hand, high-level features (features with rich description) are sparse but can be easily distinguished.

According to the choice of primitives the stereo matching algorithms can be divided into:

- intensity-based (area based) algorithms—the primitives produce a description at each pixel location, as for example, intensity values around a point;
- feature-based algorithms—more abstract, sparsely located features are extracted before the matching process.

A brief survey of the well known feature selection approaches is given below.

2.2.1. *Intensity-based approaches*

Pixel-based primitives can produce dense disparity measurements but are sensitive to distortions (illumination and contrast), and hence are not stable. They are easily measurable but their distinguishability is poor, especially in textureless or repetitive environments, or where depth discontinuities occur. Therefore, intensity-based algorithms usually suffer from slowness and ambiguous matches. However, taking advantage of directly generating dense disparity measurements, these algorithms have been successfully applied for analysis of aerial terrain images where the surface varies smoothly.

Recently suggested intensity-based methods try to overcome some of the above mentioned disadvantages (Barnard 1989, Fua 1993, Hannah 1989, Jones and Malik 1992, Jordan and Bovik 1991, Kanade and Okutomi 1991, Lee, Cho and Ha 1992, Matthies 1992, Okutomi and Kanade 1993, Sanger 1988, Vleeschauwer 1993. For example, Fua (1993) makes an attempt to preserve discontinuities. As primitives Jones and Malik (1992) use outputs of linear spatial filters tuned to a range of orientations and scales to make the description richer. Jordan and Bovik (1991) investigate the use of chromatic (colour) information. Kanade and Okutomi (1991) adaptively select neighbourhoods at each pixel position to obtain dense descriptions. Matthies (1992) presents stochastic models. Okutomi and Kanade (1993) utilize multiple baseline stereo pairs. Jenkin (1991) and Sanger (1988) use local phase difference between bandpass versions of the two images.

Typically, intensity-based algorithms perform some kind of correlation as a matching strategy (see § 2.3.2). Other algorithms use optimization (see § 3).

2.2.2. *Feature-based approaches*

Feature-based approaches have the advantage that features are less sensitive to noise and photometric distortions, they are fewer and more easily distinguishable compared to pixel intensity descriptions, and can be more precisely positioned. However, matches are sparse and interpolation is required, as well as some methods for modeling occlusion and side effects from the feature extraction process (see § 2.3.3.). Moreover, feature-based matching requires the existence of a sufficient number of features, which is why it is mostly suited for application in highly structured environments such as navigation indoors (Triendl and Kriegman 1987, Tsuji, Zheng and Asada 1986).

Many different features have been used as a basis for disparity measurements. Moravec (1977) introduced points in an image with high variance in all directions, and called them *points of interest*. These have been used lately by others (Nasrabadi and Choo 1992). The points of interest are places in an image that can be matched with relatively high confidence to corresponding points in the second image.

Since *edges* (large local intensity changes in an image) are more closely related to the physical properties of a scene, they have been used increasingly as matching primitives. A great number of edge operators have been proposed for computing the orientation (aligned with the direction of maximal gray-level change) and magnitude of an edge (quality of this change) (Ballard and Brown 1982, Canny 1985, Deriche 1987, Haralick 1984). A very popular one is the Marr–Hildreth edge detector (Marr and Hildreth 1980). Edges are extracted as zero-crossings of the convolved image with a mask approximating the Laplacian of a Gaussian (LoG). This edge detector has been widely used in stereo vision (Allen 1987, Braunegg 1990, Chakrapani, Khokhar and Prasanna 1992, Grimson 1981, Grimson 1985, Gu and Wu 1990, Hoff and Ahuja 1989, Kim and Aggarwal 1987, Lloyd, Haddow and Boyce 1987, Nasrabadi 1992, Olsen 1990, Pollard, Mayhew and Frisby 1985, Shirai 1992, Stewart and Dyer 1990). Among the properties that make it so popular are reduction of the effect of noise in the image, no need for thinning or thresholding procedures, easy interpolation of edges to subpixel precision, and separability of the convolution which simplifies the computation. The LoG operator is the best one for scale experiments. However, Lu and Jain (1989) have reported that neighbouring edges influence the response of LoG and can cause dislocation of edges, false edges, and merging of edges in the scale space.

Linear edge segments are higher level features that can capture more information from the intensity image compared to point features. In addition to geometrical description such as location, orientation and length, an edge line extractor can provide some of the attributes associated with the observed intensity images such as strength, contrast, width, straightness, and junctions, provided that the pixels around an edge are also analysed. Such is the line extraction algorithm proposed by Burns, Allen, Hanson and Riseman (1986) and its improved modifications (Kahn, Kitchen and Riseman 1990, Liu and Huang 1991, McIntosh and Mutch 1988). Further, line features are easier to extract from noisy images than are point features. The position and orientation of a line is easier to determine with a subpixel accuracy than the coordinates of a point. Also, the figural continuity constraint (see § 2.3.1.) is naturally incorporated. A disadvantage is that one needs to cope with problems such as fragmentation and fragility (see § 2.3.3.). Nevertheless, linear edge segments have quite often been used as features in stereo systems (Ayache 1991, Butler 1992, Chapron and Cocquerez 1992, Fornland, Jones, Matas, Kittler 1993, Horaud and Skordas 1989, Kim, Choi and Park 1992, Krotkov 1989, Liu and Huang 1991, Medioni and Nevatia 1985, Maravall and Fernandez 1992, McIntosh and Mutch 1988).

Kierkegaard (1993) states that *curves* rather than lines should form the basis for an edge-based stereo vision system, since information about curvature and smoothness can be used additionally for the feature description. Curve primitives have been used also by Brint and Brady (1990), Deriche and Faugeras (1990), Ma, Si and Chen (1992), Nasrabadi (1992), and Robert and Faugeras (1991).

Closed edge segments or *contours* represent the next level of abstraction. Kim and

Bovik (1988) use specific high information points (external points) along a contour to guide the matches. Sherman and Peleg (1990) also use contours as matching primitives.

While the edge-based feature extraction methods rely on the property of dissimilarity in images, the *region*-based ones rely on the property of similarity. The higher semantic content of a region makes it a good candidate for a matching primitive. Regions can be described by mean grey level, area, perimeter, principal axes, width-to-height (aspect) ratio, centroid, minimum enclosing rectangle, and holes. Region-based primitives are more stable and tolerant to noise then edge-based primitives. There are typically fewer regions in an image than edges or lines, and thus, fewer candidates to be evaluated by matching. Another advantage of using regions in stereo matching is that they make some of the matching constraints implicit or easier to incorporate. Moreover, the effects of occlusion are less severe when applied to regions than to points or segments. A great disadvantage, however, is the lack of capability to generate an accurate disparity map of fine resolution. Gagalowicz and Vinet (1989), Marapane and Trivedi (1989), Sander, Vinet, Cohen and Gagalowicz (1989), and Xu, Kondo and Tsuji (1989) have proposed region-based stereo matching algorithms.

Mohan and Nevatia (1992) concentrate on the geometrical relationships in images. Collated features are organized as description hierarchies representing much of the structural information in an image by grouping edges recursively. The grouping is based on relationships of continuity, proximity, symmetry and closure. The feature hierarchy is organized as follows: curves (grouped edges on continuity and proximity), points (terminations and junctions of curves), contours (ordered set of continuous curves), symmetries (pairs of mutually symmetric curves), and ribbons (regions enclosed by the symmetric curves).

It is noticeable that the higher the level of description the better the matching accuracy since high-level descriptions are likely to be invariant to photometric and geometric distortions and noise. Yet, high-level feature matching cannot generate fine disparity to assist surface interpretation in sufficient detail. Sander, Vinet, Cohen and Gagalowicz (1989) and Marapane and Trivedi (1992) come to the conclusion that to a certain extent, low-level features and high-level features are complementary and should be utilized together in order to facilitate accurate, robust and efficient stereo analysis. Marapane and Trivedi (1992) present a multiprimitive stereo approach that embeds a hierarchy of primitives, such as regions, edge segments, and edge pixels (edgels) in the matching process and attempts to overcome the limitations of purely edge-based or region-based approaches. Cohran and Medioni (1992) and Lim and Prager (1993) integrate intensity-based and feature-based matching in their research. Weng, Ahuja and Huang (1992) incorporate intensity, edgeness and cornerness attributes in their matching algorithm. Westman (1989) combines region- and contour-based approaches.

2.3. *Correspondence*

The stereo correspondence problem is referred to as the most important stage in the process of stereopsis. A feature P in one image and a feature Q in a second image are said to be corresponding or homologous, if P and Q are projections of the same physical entity. All pairs of corresponding features from two images of the same scene have to be matched.

The task of identifying corresponding locations in the two images is a difficult one. For an image of size $N \times N$ pixels, each pixel event in one image potentially has N^2

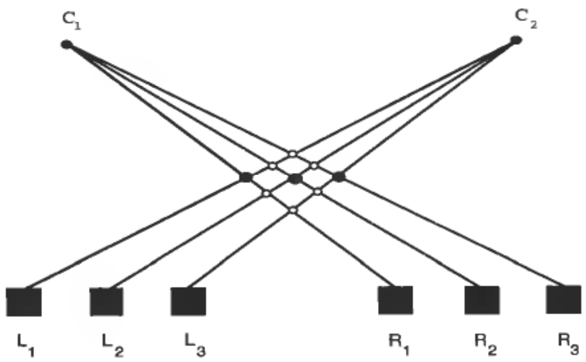

Figure 3. The false target problem.

possible matches in the other image. Consider an example (Fig. 3) which contains three elements in each of the left and right images. Each element in the one image is similar to any of the three in the other image. So there are nine possible matches, with only one being correct. The fact that a feature in one image may match equally well with a number of features in the other image is referred to as the *false target problem* (Grimson 1981, Marr and Poggio 1979). It arises from photometric differences between the two views (illumination or contrast), structured properties of the scene (repetitions), occlusion, blandness, or distortions in the feature extraction process (fragmentation and fragility of edge segments). The difficulties in the matching stage are described in detail in § 2.3.3.

When searching the space of possible correspondences, two interconnected issues have to be considered. The first one restricts the set of possible alternatives. The choice of appropriate features to be matched is one possibility. The more distinguishable those features are, the fewer the candidates to be evaluated by matching. In this context the above requirement is better met by high-level features, whereas local intensity values around a point are less appropriate. Another possibility for restricting the set of possible alternatives is to apply certain constraints to the number of possible matches, that used together with relevant matching strategies, can eliminate false matches and verify the match results. The known relationship between the two images imposes metric constraints on the relationship between the two projections of a space feature. In addition, certain descriptive properties of the scene are quasi invariant under perspective projection and therefore they do not change too much with the viewing position. In the next section some of the most popular constraints are described.

The second issue for finding the correct solution of the correspondence problem concerns the strategies for efficiently searching the space of alternatives. A matching strategy is the controlling rule for the whole matching process and the way the underlying constraints are incorporated into an algorithm to reduce the matching combinatorics. In general, the matching process consists of two stages: local matching and global matching. In the local matching stage, for every feature in the first image, an attempt is made to find a set of candidate match features in the second image that satisfy certain geometrical constraints and have similar local feature properties. In the global matching stage a scheme for imposing global consistency among the local matches is used to disambiguate multiple local match-feature candidates. In § 2.3.2 a brief survey of the most popular matching stategies is given.

2.3.1. *Binocular stereo matching constraints*

The matching process employs constraints to help select the valid matches among the candidates. Typically, the selection is made by supporting some matches and possibly inhibiting others. The constraints are derived from assumptions about the scene or about the imaging process. They are often implemented as relations between pairs of candidate matches. However, as will be shown later, it is possible that valid matches are not covered by the assumptions of a constraint. In general, when mismatches occur there is no way to decide whether this is due to violation (inapplicability) of the imposed constraints (Stewart and Dyer 1987).

Epipolar constraint

The dimensionality of the search space can be reduced from two dimensions to one dimension by observing that for any point in one of the images, its potentially matching points in the other image must lie along a known line: this is the epipolar constraint, a geometric property available whenever the stereo sensor is properly calibrated. Suppose that a point P in the scene is projected onto the left image. Let the two optical centres be C_1 and C_2 (Fig. 4). The line PC_1 and the baseline (the line passing through the focal points) determine a unique epipolar plane. Then the projection of P on the right image must lie on the intersection of the epipolar plane with the right image plane. The intersection of the epipolar plane with an image plane is called the epipolar line.

A very special situation arises where the cameras are registered so that the epipolar lines correspond to the scanlines in the images (conventional camera modeling). Under such conditions the computation of the epipolar lines is simple. Most of the presented stereo matching algorithms use such camera registration. Should this not be the case, it is possible to arrange such a configuration by a transformation of the images, called rectification (Ayache 1991, Ayache and Hansen 1988).

Consider that epipolar lines correspond to the scanlines in the images. A liberal interpretation of the epipolar constraint states that a point on a line y can be matched only to points on lines y' such that $y - v \leqslant y' \leqslant y + v$ for some constant v (Grimson 1985). When segments or regions are used as matching features, the epipolar constraint can be formulated as follows. A pair of features (S, S') satisfies the epipolar constraint if and only if the epipolar plane passing through the centroid of S intersects S'. A symmetric definition applies to the pair $(S' S)$ (Ayache 1991). This liberal interpretation of the epipolar constraint is more appropriate, since accurate camera calibration is difficult to achieve.

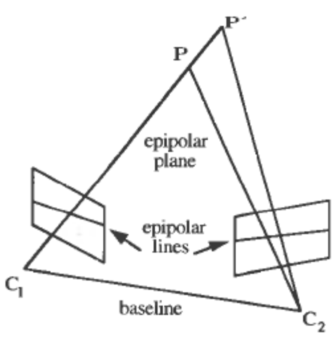

Figure 4. The epipolar constraint.

Geometric similarity constraint (*compatibility constraint*)

Some of the description properties of feature primitives are quasi invariant under perspective projection. Therefore, they can be used for a similarity measurement between features from two images in the local matching stage. For example, such properties can be the sign of zero-crossing, edge orientation, magnitude and contrast, area and perimeter of a region. For intensity-based matching the assumption is that matched pixels have similar intensities or similar values for certain simple functions of intensity (photometric constraint).

It should be noted that such similarity measurements can be made assuming that there are no photometric differences in the two views nor incorrect camera registration. Unfortunately, in general, this is not necessarily the case.

Uniqueness constraint

The uniqueness constraint was formulated by Marr and Poggio (1979) and it states that there is at most one valid match for each primitive, i.e. at most, one disparity value may be assigned to each feature from either image. This condition relies on the assumption that a feature corresponds to something that has a unique physical position. An exception can occur when two features lie along a single optical line for one of the cameras, but are separately visible from the other camera (Fig. 5). Assuming that the surfaces in the scene are opaque, only the physical point closest to the camera can be considered (point P_2 in Fig. 5). However, situations can be pointed out in which the spatial structure of the scene cannot be correctly recovered by imposing the uniqueness constraint (Fig. 6) (Ayache 1991).

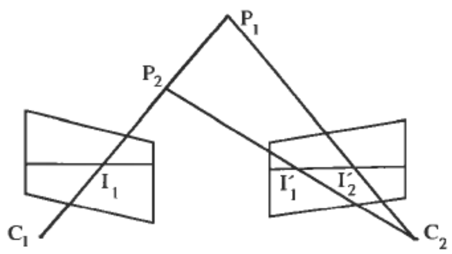

Figure 5. The uniqueness constraint.

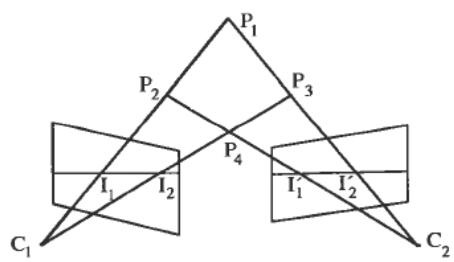

Figure 6. Violation of the uniqueness constraint. The pair of points (P_3, P_4) is aligned with the optical centre of the first camera whereas the pair (P_2, P_4) is aligned with the optical centre of the second camera.

Continuity constraints

The *surface continuity constraint* assumes that the physical world is composed of surfaces that are continuous almost everywhere. This suggests that disparities ought to vary smoothly, that is, the disparity of a point will be similar to the disparity of nearby points. This assumption does not hold true near object boundaries. The surface continuity constraint was formulated by Marr and Poggio (1979). They note that the greater the range of disparity over which a match is sought, the greater the number of false targets.

The gradient of the disparity is closely related to the smoothness of the scene surfaces. A *disparity gradient constraint* was formulated by Pollard *et al.* (1985). It defines support between nearby pairs of matches. Suppose that m_1 and m_2 are two candidate matches with disparities d_1 and d_2 respectively. Then the matches support each other if

$$\frac{|d_1 - d_2|}{D(m_1, m_2)} \leq 1$$

where $D(m_1, m_2) \leq K$ is the distance between the two matches, for constant K. More distant matches can have a greater disparity difference. Another close formulation is the coherence principle of Prazdny (1985), which also focuses on the similarity of disparity of nearby matches as a possible way for mutual support based on the assumption that objects occupy a well-defined 3D volume.

The *figural continuity* constraint is based on the assumption that contours appear similarly in the two images. Thus, disparity values can be restricted along edges in an image. This constraint was proposed by Mayhew and Frisby (1981). It is weaker than the surface continuity constraint and it overcomes the problems near object boundaries. In the cases when edge segments are considered as matching features, the figural continuity constraint is implicitly applied.

Ordering constraint

Since the image structure is largely preserved in both images, the relative position of two features should not differ much from the relative position of the corresponding features in the other image. The ordering constraint assumes the preservation of the order of matched points along corresponding epipolar lines unless the scene contains transparent or narrow occluding (thin) objects (Fig. 7).

General position constraint

The general position constraint relates to the observation that certain events occur quite infrequently in a statistical sense. It has been utilized by Arnold and Binford (1980). Horaud and Skordas (1989) extrapolate the results obtained by Arnold and Binford. Consider a camera geometry with optical centres behind the images planes. Let P be a scene point, and p and p' be the projections of P onto the left and right image planes. P lies along the line Op (Fig. 8). As P moves along this line, p' must lie

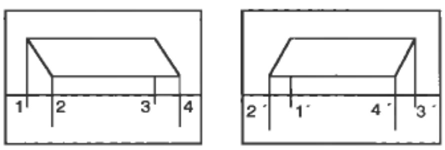

Figure 7. Violation of the ordering constraint.

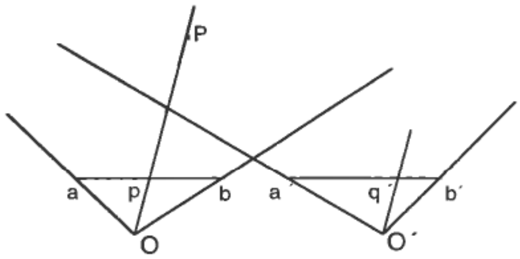

Figure 8. A top view of the epipolar plane defined by p, O and O'.

somewhere in between a' and q', where a' lies onto the left border of the right image and $O'q'$ is parallel to Op. Hence, the segment $a'q'$ can be regarded as the locus of possible positions of p'. When p is close to a, the segment $a'q'$ has its shortest length and when p is close to b, the segment $a'q'$ is as long as $a'b'$. This suggests that there are less potential matches for left image points that are close to the left border than for points that are close to the right border. This is equally true for right image points close to the right border.

2.3.2. *Matching strategies*

Correlation

Area-based stereo techniques mostly use correlation as a matching strategy. Intensity patches in one image are chosen and then a search is performed for the best matching location in the other image, typically using normalized cross-correlation as a measure of similarity, or mean-squared difference as a measure of dissimilarity (Fua 1993). Hence, correlation-based algorithms rely on the fact that the same texture can be found at corresponding points in the two images of a stereo pair. On the one hand, correlation is capable of providing dense disparity determination. On the other hand, it requires that the images have sufficient local texture at each point without repetitions or discontinuities. Furthermore, it is sensitive to noise, contrast, or lighting changes. So, the following problems can be encountered, as pointed out by Barnard (1989):

- A patch must be large enough to contain the information necessary to identify the corresponding patch unambiguously.
- At the same time, a patch must be small compared to the variation of the disparity.
- Correlation is unreliable in areas with uniform or slowly varying intensity.

Recently, a lot of trials have aimed to overcome some of these problems (Fua 1993, Hannah 1989, Jones and Malik 1992, Kanade and Okutomi 1991, Okutomi and Kanade 1993). Fua (1993) performs bidirectional correlation followed by interpolation to propagate across textureless areas preserving discontinuities. Jones and Malik (1992) suggest the use of filters at different scales and orientations to provide a rich description of an image patch. They introduce the notion of a visibility map to cope with occlusions, and the notion of a scale map to provide adaptive scale selection. Kanade and Okutomi (1991) present an algorithm that controls the size and shape of a patch (window) for the correlation process by applying a statistical model representing the uncertainty of disparity of points over the patch. Okutomi and Kanade (1993) provide a new approach for the removal of false matches that integrates the evidence for a final decision from multiple image pairs acquired with different baselines. The proposed method accumulates the measures of matching (sum of squared differences) from the different pairs into a single evaluation function that has a unique minimum at the correct matching position.

Correlation can be based not only on intensity patches but on features as well (Allen 1987, Singer 1987, Triendl and Kriegman 1987).

Relaxation

Relaxation labelling is a technique (proposed by Rosenfeld *et al.* (1976)) developed to deal with uncertainty in sensory data interpretation systems. It uses contextual information as an aid in classifying a set of interdependent objects by allowing interactions among the possible classifications of related objects. In the stereo paradigm the problem involves assigning unique labels (or matches) to a set of features in an image from a given list of possible matches. For each candidate pair of matches a matching probability is updated iteratively depending on the matching probabilities of neighbouring features. Therefore, stronger neighbouring matches improve the chances of weaker matches in a globally consistent manner. Marr and Poggio (1976) are pioneers in the use of relaxation for stereo vision. They associate weights to each match. At each iteration, and for each match, an 'inhibitory' process reduces the weights of conflicting matches (those which violate the uniqueness constraint), and an 'excitatory' process increases the weights of matches with similar disparities (the continuity constraint). Relaxation has also been used by Chakrapani, Khokhar and Prasanna (1992), Kim and Aggarwal (1987), Kittler, Christmas and Petrou (1993), Laine and Roman (1991), Lloyd *et al.* (1987), Medioni and Nevatia (1985), Nasrabadi (1992), Sherman and Peleg (1990), and Stewart and Dyer (1990). A big advantage of this technique is that it is well suited for parallel implementations.

Multiresolution

Multiresolution specifies relations between pairs of matches at adjacent resolutions. The traditional use of multiresolution matches features in a hierarchical, *coarse-to-fine* manner. That is, matching occurs at the coarsest resolution first, where the features are sparser. The results of the matches are then used to restrict the range of disparities at the next finer resolution prior to matching at that level. This continues to the finest resolution. Since this strategy can reduce significantly the computation many researchers have used it in stereo matching (Barnard 1989, Braunegg 1990, Cochran and Medioni 1992, Chen and Medioni 1990, Grimson 1981, Grimson 1985, Hoff and Ahuja 1989, Lew, Wong and Huang 1992, Marr and Poggio 1979, McKeown and Hsieh 1992, Olsen 1990, Terzopoulos, Witkin and Kass 1987, Vleeschauer 1993, Weng, Ahuja and Huang 1992). However, this form of multiresolution can cause certain errors in the matching. The main source of errors arises when false matches are made at coarse resolutions (for example, in complex scenes, edges detected from low level filters do not usually correspond to any true edges (Lu and Jain 1989). These errors can restrict the possible disparities for nearby edges at finer resolutions to a range that does not include the correct disparities. Thus, it is extremely difficult for finer resolution matching to recover from errors made in coarse resolution matching. Stewart and MacCrone (1990) propose another form of multiresolution that avoids this problem. All resolutions are considered simultaneously and *cross-channel consistency* between nearby pairs of matches at different resolutions are used as a basis for support. In addition, cross-channel consistency multiresolution allows parallel implementation. Lew, Wong and Huang (1992) also suggest an algorithm that works at multiple scales simultaneously. They introduce an error function defined as the summation of the scalar error over every scale and look for a match with the minimum error.

Dynamic programming

Dynamic programming is a non-linear optimization technique based on Bellman's principle of optimality (Bellman 1962), which states that the optimal path between two given points is also the optimum between any two points lying on that path. This means that the global optimization of a multistage problem can be split into series of single stage optimization problems. It has the advantage of reducing the number of searches for an N stage optimization problem with L possible values from $O(L^N)$ to $O(NL^2)$. Furthermore, it allows parallel implementations. This matching strategy implies the use of the ordering constraint. Dynamic programming technique has been employed lately by Belhumeur (1993), Boyer *et al.* (1990), Geiger, Ladendorf and Yuille (1992), Lloyd *et al.* (1987), Matthies (1992), Ohta and Kanade (1985), Ohta *et al.* (1988), and Wang and Pavlidis (1990).

For example, Ohta and Kanade (1985) decompose the correspondence search into two: *intra-scanline* search and *inter-scanline* search. Scanline intervals between edge points are matched in the intra-scanline search. A path-finding problem in a 2D search space is formulated, in which vertical and horizontal lines are the left and the right scanlines respectively. Intersections of these lines are referred to as nodes. With each partial path a cost function is defined, based on variances of gray-level intensities of the scanline intervals being matched. Edges are numbered from left to right on each scanline (the ordering constraint is applied). Suppose that there are M edges in the considered left scanline and N edges in the corresponding right scanline. Then the solution to the intra-scanline search can be represented as a path involving a sequence of straight lines from node $(0, 0)$ to node (M, N) with the optimum cost. This cost is computed recursively adding the cost of each new primitive path to the already existing partial optimal path. The inter-scanline search is performed in order to provide consistency constraints across the 2D search planes using edge connectivity (figural continuity). A 3D search space is formed. A 3D node is represented as collection of the 2D nodes connected across scanlines. An optimal path in the 3D search space is obtained in a manner similar to that in the intra-scanline search.

Sub-graph isomorphism

A description of each image in the form of a graph is built, which makes explicit certain characteristic groupings of primitives on a relational basis. Next, a correspondence graph is built where the nodes represent possible matches, and links are compatibilities between these matches. A search for maximal-clique (completely connected subgraph) is performed. Horaud and Skordas (1989) have proposed such an algorithm. Linear edge segments are extracted from both the left and the right images. Each segment is characterized by its position, orientation, length, and contrast in the image, as well as its relationships with the nearby segments. The relationships are *left-of*, *right-of*, *collinear-with* and *same-junction-as*. Thus, a relational graph is built from each image. For each segment in one image a set of potential assignments in the other image is determined employing the epipolar constraint, the position constraint, and the geometric similarity constraint. These assignments are represented as nodes in a correspondence graph. For each node a benefit is computed that sums up the differences in contrast, length, orientation, and number of relations. Arcs in this graph represent compatible assignments established on the basis of segment relationships.

Compatibilities and incompatibilities between nodes are taken into account. To overcome the difficulty with missing relations, a compatibility propagation technique is suggested. Sets of mutually compatible nodes are found, looking for maximal clique. Each maximal clique is evaluated and the best one with respect to this evaluation is selected as solution.

Fornland *et al*. (1993) and Kierkegaard (1993) also apply this technique in their stereo systems.

Structural matching

Boyer and Kak (1988) assert that a structural description of each image should be the first goal in the stereo matching process and develop a theory of inexact matching for such structures. The structural description is given by a set of primitives and a set of relations over the primitives. Let $DL = (P, R)$ be the structural description of the left image and $DR = (Q, S)$ be the structural description of the right image. The stereo matching problem is considered in the terms of a consistent labelling problem. The set P forms the object set, and the set Q forms the label set. Two sources of information are used in the labelling process: knowledge about the attributes of each label ($\mathfrak{I}$), and knowledge about relationships between different labels ($\mathfrak{R}$). $\mathfrak{I}$ 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 and therefore, $\mathfrak{I}$ reflects the primitive attribute distortion process between the two images. $\mathfrak{R}$ consists of a set of conditional probabilities, each item in the set being the probability that a relational parameter will take on a particular value in the right image for its given value in the left image. So $\mathfrak{R}$ reflects the changes in the values of relational constraint parameters. Given the information in $\mathfrak{I}$ and $\mathfrak{R}$, an optimal probability measurement (OPM) is defined with an $\mathfrak{I}$-term and an $\mathfrak{R}$-term, assuming that the information in $\mathfrak{I}$ is independent of the information in $\mathfrak{R}$. An information-theoretic interprimitive distance measure $DIST_h(P, Q)$ is formulated to represent the dissimilarity between the sets of primitives P and Q under a specific mapping $h{:}P \rightarrow Q$, and is related to the $\mathfrak{I}$-term in the OPM definition. A relational inconsistency measure $INC_h(R, S)$ is formulated to measure the distortion of relational parameters between the set of primitives, and is related to the $\mathfrak{R}$-term in the OPM definition. The solution of the consistent labelling problem is then formulated as

$$MIN(DIST_h(P, Q) + INC_h(R, S)).$$

The structural description of each image is derived from skeletons of objects in the scene. The edges of the skeletons form a set of primitives. The following relations over the set of primitives are chosen: pairwise-orientation (the mean orientation of the straight line joining the centroids of a pair of skeletal edges), distance (length of the straight line joining the two centroids), and end-distance (length of the line segment joining the closest pair of end points between two edges).

Another structural matching algorithm was presented by Pong, Haralick and Shapiro (1989), Topographic structures such as ridge, valley, hill-side, saddle and flat are extracted by the topographic primal sketch of Haralick, Watson and Laffey (1983) and matched. Gu and Wu (1990) also suggest the use of structural matching.

Prediction and verification

The matching proceeds by recursive prediction and verification. Reliable matches are hypothesized between feature primitives using epipolar constraint and geometric similiarity constraints. Then verification is performed for each hypothesis by recursive propagation to its neighbourhood by assigning new matches between neighbouring features using uniqueness and continuity constraint. The number of propagated matches is then used to discriminate between ambiguous matches and to validate final matches. Ayache (1991) and Krotkov (1989) also use such a strategy for stereo matching.

Bidirectional matching

This strategy is usually combined with another one. First matching is performed from the left image to the right image and then vice versa: from the right image to the left image and features that are not matched in both directions are not taken into consideration. Although such a strategy can sometimes eliminate correctly matched pairs it has been applied by Cochran and Medioni (1992), Fua (1993), Hannah (1989), Little and Gillett (1990), Liu and Huang (1991), and Pong, Haralick and Shapiro (1989).

2.3.3. *Difficulties in matching*

Stereo algorithms are no better than their ability to extract dense and accurate range information. Unfortunately, there are some difficulties in reliably finding the correct matches during the correspondence search.

False matches can occur due to

- photometric differences between two views, i.e. illumination or contrast;
- structural properties of the scene, such as periodic texture or lack of texture; a special case related to the practical implementation of edge-based algorithms is the matching of edges with almost the same orientation as the orientation of the epipolar lines;
- incorrect camera registration: small inaccuracies in measuring camera parameters can lead to large errors in computed depth;
- discrete nature of a digital image;
- noise;

Missing matches can arise due to

- occlusion: a point on a surface becomes invisible when the viewing position goes behind a surface or when the line of sight is interfered with by other objects (Fig. 9);
- part of the scene not being in the field of view of the other camera;
- blandness: a primitive is too weak in one of the images and, therefore, is discarded as noise in the feature extraction process.

A lot of stereo systems take into account occlusion effects (Belhumeur 1992, Belhumeur 1993, Dhond and Aggarwal 1992, Geiger, Ladendorf and Yuille 1992, Jones and Malik 1992, Little and Gillett 1990, Lloyd *et al.* 1987, Marroquin, Mitter and Poggio 1987, Ohta and Kanade 1985, Ohta *et al.* 1988, Olsen 1990, Sugimoto *et al.* 1988, Weng, Ahuja and Huang 1992, Yokoya 1992). For instance, the problem of modeling occlusion is addressed in the work of Geiger, Landendorf and Yuille (1992) by introducing a constraint that relates discontinuities in one image with occlusions in the other.

Some difficulties in the matching process can also be caused by distortions from the feature extracting stage. For example, by extracting linear edge segments, the following distortions can arise (Takahashi and Tomita 1988):

- fragmentation: a segment in one image appears as two segments in the second image (Fig. 10);

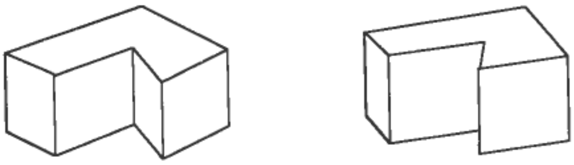

Figure 9. Occlusion.

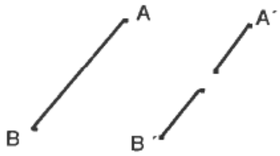

Figure 10. Fragmentation.

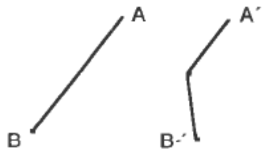

Figure 11. Fragile segmentation.

- fragile segmentation: a segment in one image appears broken in the second image (Fig. 11);

Some algorithms that cope with distortions from the linear edge segmentation process are those by Ayache (1991), Horaud and Skordas (1989), Medioni and Nevatia (1985), Ohta and Kanade (1985), and Takahashi and Tomita (1988). For example, Medioni and Nevatia (1985) overcome possible fragmentation of segments by allowing a segment in one image to be matched to two (or more) segments in the other image demanding that the candidates should not overlap each other.

Mohan *et al.* (1989) present a post processor to any edgel-based (features are edge pixels) stereo matching algorithm which, given a linear segment, first rejects the disparities violating figural continuity and then computes a linear function that describes the disparity change along that segment. This function is then used to correct disparity errors which can arise due to the fact that information along an epipolar line may not be sufficient for matching, or due to distortion of the real epipolar line.

2.4. *Disparity interpretation*

After the correct point matches have already been identified, the next step is to recover the 3D information. Assuming conventional parallel axis geometry, this can be done by a simple geometrical method known as triangulation. The disparity value d for each matched pair of points $P_L(x_L, y_L)$ and $P_R(x_R, y_R)$ is

$$d = x_L - x_R,$$

and the 3D coordinates of the imaged point $P(X, Y, Z)$ with respect to the left camera are

$$P(X, Y, Z) = \left(\frac{bx_L}{d}, \frac{by_L}{d}, \frac{bf}{d}\right)$$

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

Usually, some uncertainty considerations are needed. Verri and Torre (1986) have analysed the error in depth estimates. They separate the errors into two components. The first, precision of the setup geometry, affects absolute depth estimates. The second, image plane orientation and focal length, affects relative depth estimates. Blostein and Huang (1987) perform an analysis of the position error from image plane quantization. In general, there are several ways to improve depth measurements, each involving some additional computational cost as pointed out by Barnard and Fischler (1982), and these are subpixel estimation, increased stereo baseline, and statistical averaging over several views. The depth computation process of nonparallel stereo systems requires a more general approach (Grimson 1981, Krotkov 1989).

In most cases, the resulting depth points are sparse, whereas depth must be computed at every point in the scene. The most popular approach to interpreting stereo disparity is surface fitting. In its simplest form, the idea behind this approach is to interpolate (or approximate) the sparse depth values with a smooth surface. The surface fitting approach has two forms. The first one is minimization of a spline functional. The idea is to fit an elastic plate or membrane to the given data points and allow it to achieve equilibrium (Fua 1993, Grimson 1981). The resulting representation is of point by point distance. The second form is to fit polynomial based surface patches directly to the available distance data. The output representation is of an algebraic surface patch (Eastman and Waxman 1987, Hoff and Ahuja 1989).

Usually, the interpolation (reconstruction) step is performed as postprocessing, after the matching is finished and sparse disparity measurements are obtained. Eastman and Waxman (1987), Hoff and Ahuja (1989) and Olsen (1990) integrate the feature-based correspondence analysis and the reconstruction. Eastman and Waxman (1987) and Hoff and Ahuja (1989) fit planar and quadratic surface patches to the measurements. Olsen (1990) presents multipass attribute matching and reconstruction to achieve dense disparity estimation. During the matching the requirement for attribute similarity is relaxed from pass to pass, while the requirement of close agreement between predicted and measured disparity is strengthened as more measurements are accumulated.

Wildes (1991) presents an analysis of the direct recovery of 3D scene geometry from binocular stereo disparity. The analysis makes explicit the geometric relations between a stereo disparity field and a differentially projected scene. It shows how three-dimensional surface geometry can be recovered through first-order (i.e. distance and orientation of a surface relative to an observer) and binocular viewing parameters from stereo disparity.

3. Energy minimization approaches to stereopsis

Many tasks in computer vision, including stereo matching, have been defined recently in terms of minimization of so-called global energy functions, describing nonlinear interactions between the observed data and features to be extracted from the images (Yuille 1989). Regularization theory and stochastic modeling are the major approaches for mapping the stereo correspondence problem into an optimization problem, and will be described briefly in this section. Some of the mostly used optimization techniques will be covered as well.

Regularization

Stereo matching is an ill-posed problem in the sense of Hadamard (Bertero, Poggio and Torre 1988, Poggio, Torre and Koch 1985). That is, the solution may not be unique (ambiguity), or it may not exist (occlusions), or it may not depend continuously on the measurements (discontinuities). Under some restrictive conditions, it can be regularized (made well-posed) by a variational principle that contains a term measuring the dissimilarity in the two images, and a stabilizer that penalizes large disparity gradients (disparity gradient limit). Assuming conventional camera geometry this results in minimization of the following energy functional:

$$\|L(x) - R(x + d(x))\| + \lambda \|d'(x)\|,$$

where $L(x)$ and $R(x)$ are continuous measures of the left and right images, and $d(x)$ is the disparity. λ is a so-called regularization parameter, and it controls the compromise between the degree of regularization of the solution and its closeness to the data.

The minimization of the above functional provides a physically plausible solution. However, it can be inconsistent since it requires $d(x)$ to be continuous and differentiable, which is not valid in the presence of occlusions and discontinuities. Furthermore, simple formulation of the first term in the functional such as direct brightness matching of pixels is a poor solution. If $L(x)$ and $R(x)$ are feature measures rather than intensity measures, they are not continuous functions of x. To deal with discontinuities and occlusions, a binary occlusion process (marking locations of occlusion) and a binary line process (marking locations of discontinuities) can be introduced in the energy functional as follows:

$$\|L(x) - R(x + d(x))\|(1 - o) + \lambda \|d'(x)\|(1 - l) + V_L + V_O,$$

where o is the binary occlusion process, l is the binary line process, and V_L and V_O are cost terms associated with the line and occlusion processes (Belhumeur 1993, March 1988, Terzopoulos, Witkin and Kass 1987, Toborg and Hwang 1991, Yokoya 1992).

Another problem is that there is no reliable way to estimate the parameters that can be introduced in the functionals which control the relative weight of the various terms in the energy function. They often depend on the individual image on view.

Stochastic modeling

While in standard regularization use of *a priori* knowledge leads to restriction of the solution space, in stochastic modeling *a priori* knowledge is represented in terms of appropriate probability distributions. Bayes theory and Markov random fields (MRF) are used to compute the posterior distribution $P(f/g)$ which represents the likelihood of a solution f given the observations g. In this way the stereo correspondence problem can be solved by finding the estimate f that either maximizes the *a posteriori* probability (MAP estimate), or minimizes the expected value of an appropriate error function (Marroquin, Mitter and Poggio 1987). The MRF formulation makes the value on a discrete location dependent only on values within a given neighbourhood and the prior probability can be expressed in the form of a Gibb's distribution. Thus, it allows the capturing of many properties of the system of interest just by adding appropriate terms in the cost function. Such an approach for modeling the stereo matching problem has been used by Belhumeur (1992), Chang and Chatterjee (1990), Chang and Chatterjee

(1992), Geiger, Ladendorf and Yuille (1992), Lim and Prager (1993), Marroquin, Mitter and Poggio (1987), Matthies (1992), and Yuille, Geiger and Bulthoff (1991). Dynamic programming, simulated annealing and mean field annealing are applied for the optimization process. Disadvantages of stochastic modeling are the increase of computational complexity and the difficulty of estimating the parameters in the cost function.

Optimization

Minimizing a global energy function for stereo matching is a complicated task since the energy functions are usually highly nonconvex, so that ordinary optimization becomes trapped in local minima. Computationally demanding stochastic optimization algorithms, such as simulated annealing (described briefly below), are generally necessary to compute globally optimal solutions. Faster deterministic algorithms, such as mean field annealing (described briefly below), graduated non-convexity (Blake and Zisserman 1987), and Hopfield neural networks (Lee, Choo and Ha 1992, Mousavi and Schalkoff 1991, Nasrabadi and Choo 1992, Zhou and Chellappa 1988), can often be used instead, when a good initial guess is available. Multigrid methods can significantly improve the convergence rate of the iterative schemes (Barnard 1989, Chang and Chatterjee 1990, Jordan and Bovik 1991, Terzopoulos, Witkin and Kass 1987, Yokoya 1992).

Simulated annealing is a stochastic iterative improvement strategy very suitable for finding the global minimum of highly nonconvex functions. The basic idea is that controlled uphill steps are incorporated, thus providing a mechanism for escaping from local minima (van Laarhoven 1987). The frequency and severity of these uphill movements is reduced by slowly decreasing a parameter T (referred to as temperature) so that global or nearly global minimum is found. Barnard (1989), Belhumeur (1992), Chang and Chatterjee (1990), and Jordan and Bovik (1991) utilize simulated annealing in the matching process. A major drawback of this optimization technique is that the convergence process is very slow.

Mean field annealing is a new optimization technique that has been developed to speed up the convergence process in simulated annealing (Bilbro *et al.* 1989, Chang and Chatterjee 1992, Lim and Prager 1993, Yuille, Geiger and Bulthoff 1991). While in simulated annealing the global minimum is achieved through many probabilistic transitions of every site according to the local characteristics of the site, in mean field annealing these transitions are replaced by direct evaluation of the average field at each site, given the local probability density function. This leads to an overall decrease in the computational time.

4. Parallelism in stereo

Computer vision tasks are computationally very expensive and therefore they require high-performance computers for practical real-time applications. Requirements for navigation systems, machine inspection, and other robot vision tasks are still difficult to meet, although recent stereo matching techniques have speeded up the computation without compromising reliability. Parallel computing is currently the major tool in pursuit of the required level of performance. This is why some parallel implementations of stereo matching algorithms are covered briefly in this section.

Chakrapani, Khokhar and Prasanna (1992) provide a parallel algorithm on fixed size mesh array using zero-crossings of the LoG operator applied to the images. $O(nm/P)$

time complexity is shown for a $\sqrt{P}x\sqrt{P}$ processor mesh array, where n is the number of the zero-crossing points in the left image, m is the set of possible candidate points in the right image for a given zero-crossing, and $1 \leqslant P \leqslant n$.

Drumheller and Poggio (1986) present a stereo algorithm for the Connection Machine, using the north-east-west-south (NEWS) mechanism for parallel execution of convolution. Zero-crossings of the convolution of the images with an LoG operator are used as features. Then a 3D set of potential matches are obtained. The amount of local support for each potential match is determined through 3D convolution. Correct matches are chosen on the basis of local support, uniqueness and ordering.

Laine and Roman (1991) propose an algorithm for incremental stereo matching on a SIMD machine. Several existing techniques dealing with the classification and evaluation of matches, ordering constraint and relaxation, have been integrated and reformulated in terms of parallel execution on a theoretical SIMD machine. The machine model assumes a 2D array of pipelined processors and a set of memory arrays that may be read and/or updated during each matching cycle. Stage processors are capable of performing four kinds of operations: logical, integer arithmetic, min/max, and function of one variable. The feasibility of the algorithm is shown by implementation on a typical commercial SIMD pipelined processor. Reported machine time for a synthetic image is 54 s and for a real urban one is 62 s.

Williams and Anandan (1986) present two hierarchical algorithms designed for a mesh-connected computer, using the NEWS mechanism for near-neighbour communications. The first is a mesh-computer implementation of the Marr–Poggio–Grimson stereo matcher (Grimson 1981) which runs in $O(d \log d)$ time where d is the maximal disparity considered. The second is the mesh-computer implementation of a motion correspondence algorithm which uses a correlation based technique requiring $O(d^2 \log d)$ time. The algorithms were simulated on a sequential machine.

Stewart and Dyer (1990) describe a parallel simulation of the General Support Algorithm (GSA), a connectionist network for stereo matching. The features used in GSA are oriented zero-crossings of the LoG smoothed images. Measures are used to define consistency among nearby edges (disparity gradient constraint), along edge contours (figural continuity constraint), and across resolution levels (simultaneous multiresolution). The nodes in the network represent individual candidate matches, and the connections between nodes are defined by the consistency measures and the uniqueness constraint. The simulation involves two phases of processing: the first phase builds three large data structures, and the second phase simulates the iterations of the network (relaxation). The data structures include the edge images, the list of candidate matches, and the list of connections between the candidate matches. During each iteration, new activations of all the nodes are computed based on their prior activations and on the lists of incoming connections. For both of the phases, computation may be partitioned among a number of processors with little need for synchronization or mutual exclusion.

Watanabe and Ohta (1990) propose a new scheme to integrate multiple stereo algorithms into a cooperative framework. Each algorithm is implemented in a separate module and executed in parallel. There are three stereo matching modules: point-based, interval-based and segment-based matching modules. A disparity map generation module fuses the results obtained by the three modules. These four modules are activated in parallel and communicate with each other. When a stereo matching module encounters a situation which needs assistance from another module, it sends a query to the module. The module which received a query then generates an answer and returns it to the sender.

5. Trinocular stereo

Trinocular stereo analysis has been proposed recently in computer vision as an alternative to the conventional binocular stereo analysis. The matching is found to reduce the uncertainty regarding the goodness of local matching by more than a half, while the computational cost of local matching is increased over binocular stereopsis by only about one quarter (Dhond and Aggarwal 1990).

The basic advantage of the third camera is the extra epipolar geometry constraint explained below. Disadvantages are the appearance of a greater number of occluding edges and the smaller common field of view compared to ordinary stereo. Also, the requirements on the precision and linearity of the camera system are stricter.

Consider three cameras with optical centres C_1, C_2, C_3 before image planes P_1, P_2, P_3 respectively. Given a physical point A, its image A_i on camera i is defined as the intersection of line AC_i with an image plane P_i (Fig. 12). Points A_1, A_2, and A_3 form a triplet of homologous image points. Any triplet (A_1, A_2, A_3) is such that A_i lies at the intersection of the epipolar lines L_{ij} and L_{ik} defined by the two other image points A_j and A_k. Thus, the search for homologous points between two of the images can be reduced by a simple checking for a point at the intersection of two epipolar lines in the third image. For example, checking that (A_1, A_2) form a pair of homologous image points, consists in verifying the presence of A_3 at the intersection of L_{31} and L_{32}. This condition is referred to as the trinocular epipolar constraint.

Ayache and Lustman (1991) present an original trinocular stereo technique. It can be summarized as follows. A graph-based description of a polygonal approximation of the contours is extracted from each image. Rectification is performed to simplify the epipolar geometry. A hypotheses prediction–verification matching strategy is used. Triplets of potential matches are derived from the previously constructed graphs by simple geometric verification. The orientation and length of segments are compared, and epipolar constraints verified. Consistency checks utilizing continuity constraint are performed to remove ambiguous matches.

Ohta, Yamamoto and Ikeda (1988) describe a trinocular stereo algorithm with three collinearly arranged cameras (Fig. 13). The correspondence search is based on two-level dynamic programming. Intervals between two edges on each scanline are considered as matching primitives. Let (l_i, c_j, r_k) be edge positions on each scanline. A node (l_i, c_j, r_k) in the search represents the correspondence of the three edges. Three kinds of nodes are distinguished: LCR nodes, LC nodes and RC nodes. An LCR node

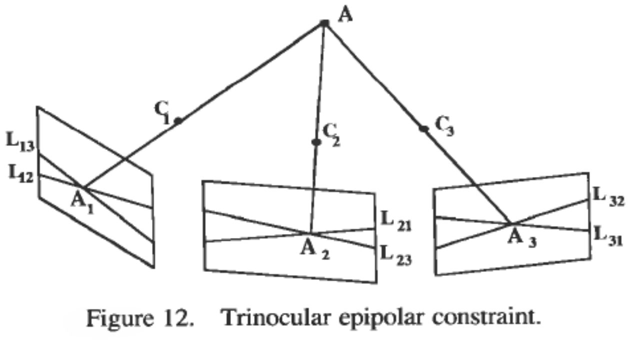

Figure 12. Trinocular epipolar constraint.

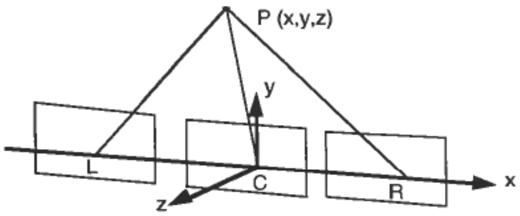

Figure 13. Collinearly arranged three cameras.

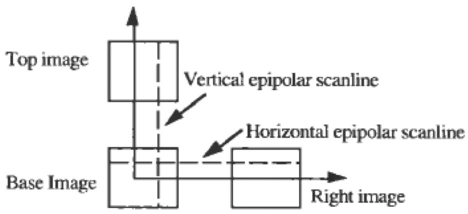

Figure 14. TGSA imaging geometry.

is the intersection of three lines each of which corresponds to an edge position on the scanlines *l*, *c* and *r*. An LC (RC) node is the intersection of two lines each of which corresponds to an edge position on the scanlines *l*(*r*) and *c*, and occluded on scanline *r*(*l*). In the first level only LCR (visible) nodes are processed and search is done on the whole scanline. In the second level, LC or RC (occluded) nodes are processed and search is applied to each of the local intervals between two adjacent LCR nodes.

Steward and Dyer (1988) proposed the so-called Trinocular General Support Algorithm (TGSA). The TGSA is a relaxation algorithm that integrates binocular and trinocular constraints to select valid matches. It uses cameras positioned at the vertices of an isosceles right triangle (Fig. 14). Candidate matches are located between the horizontally-aligned and vertically-aligned pairs of images. Binocular matching constraints are implemented within each pair of matching images. New trinocular constraints are used in addition, called trinocular uniqueness and trinocular disparity gradient. Trinocular uniqueness constrains vertical and horizontal matches for an edge in the base image so that they have nearly identical disparities. The trinocular disparity gradient defines support relations between nearby orthogonal matches based on the relative distance between them and on the difference in their disparities. Stewart and Dyer assert that the last constraint resolves ambiguity in periodic regions in many cases.

Ohta, Watanabe and Ikeda (1986) 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 with axes parallel to each other. Correspondence search using dynamic programming is performed on the two stereo pairs L-U and L-R independently. The two obtained depth maps are then merged using a relaxation method. Since relations between pairs of images during matching are not employed, errors from ambiguities in both pairs cannot be recovered.

Other trinocular stereo algorithms are described in Ito and Ishii (1986), Pietikäinen and Harwood (1986), Robert and Faugeras (1991), Yashida *et al.* (1986).

6. Test images

The following classification can be made about the test images used for the verification of the stereo matching algorithms:

(1) *Random dot stereograms* (RDS) are a popular kind of imagery for testing computational models of stereopsis. An RDS consists of a pair of similar structural images filled with randomly generated black and white dots, with some regions of one of the images shifted to either the left or the right relative to the other image. Thus, the disparity is known at each pixel location. They were suggested by Julesz (1960) for the investigation of monocular and binocular depth cues in the perception of depth. For example, such tests have been performed by Chang and Chatterjee (1990), Grimson (1981), Mayhew and Frisby (1981), Olsen (1986), Pollard *et al.* (1985), Stewart and Dyer (1990).

(2) *Synthetic scenes* have been used, for instance, by Chang and Chatterjee (1990), Hoff and Ahuja (1989), Ito and Ishii (1986), Medioni and Nevatia (1985), Ohta and Kanade (1985), and Stewart and Dyer (1988). The correctness of matches was checked manually.

(3) *Block world scenes* typically contain objects with polyhedral, cylindrical, or spherical surfaces characterized by sharp physical boundaries and/or surface markings and dispersed against contrasting background. The complexity of such images increases because of the presence of occluding objects and repetitive features on the object surfaces (an example is a scene containing a Rubik cube). Such tests have been carried out by Grimson (1985), Ito and Ishii (1986), Kim and Bovik (1988), Medioni and Nevatia (1985), Ohta and Kanade (1985), Ohta *et al.* (1986), Ohta *et al.* (1988), Pietikäinen and Harwood (1986), Sugimoto *et al.* (1988), Takahashi and Tomita (1988), Watanabe and Ohta (1990), Xu *et al.* (1989), and Yashida *et al.* (1986).

(4) *Indoor scenes* usually represent real-life laboratory environments consisting of many straight line edges with repetitive structure (doors, windows, furniture). Such scenes have been used by Ayache (1991), Braunegg (1990), Horaud and Skordas (1989), and Watanabe and Ohta (1990). Indoor scenes are a typical environment for mobile robots for solving obstacle avoidance and navigation tasks (Triendl and Kriegman (1987), Tsuji and Zheng (1986)).

(5) *Industrial scenes* contain one or more industrial parts. Such tests have been performed by Ayache (1991), Hoff and Ahyuja (1989), Marapane and Trivedi (1989), Medioni and Nevatia (1985), and Watanabe and Ohta (1990).

(6) *Outdoor scenes* consist of aerial, natural or man-made terrain images or pictures of trees. Tests on outdoor scenes have been performed by Ayache (1991), Barnard (1987), Butler (1992), Fua (1993), Grimson (1985), Gu and Wu (1990), Hoff and Ahuja (1989), Ohta and Kanade (1985), Pietikäinen and Harwood (1986), and Singer (1987).

SRI International has recently compiled stereo pairs and sequences for testing and evaluation (Bolles, Baker and Mannah (1993)). The image sets contain synthetic, indoor and outdoor scenes that are a real challenge for any stereo matching algorithm. Some of the provided traps are photometric differences between the two views (brightness or contrast), repeated patterns, highly ambiguous featureless backgrounds, data sets with increasing baselines, and slightly incorrect rectification of the images (vertical misalignment or rotation).

The Calibrated Imaging Laboratory at CMU also provides data sets with multiple images of static indoor scene with accurate information about object locations in 3D and camera calibration parameters. The type of objects in the images varies from simple polyhedra to complex model train sets.

7. Conclusion

Stereo vision offers a passive means for deriving detailed 3D information for a scene without using specialized illumination or energy resources. The correspondence problem is a key item in stereo. In order to solve it accurately, reliably and efficiently, the following must be taken into account:

- The choice of features to be matched, the matching rules (constraints) and strategy, is of great importance. The availability of abstract information pertaining to the features yields tremendous advantages in restricting the search space of possible correspondences. Suitable selection of matching rules and strategy can eliminate false matches and verify obtained results.
- Stereo vision is task dependent. A given task can influence the choice of features, matching rules and strategy, as well as the output sufficiency. For example, a detailed 3D surface profile of an object is required for object inspection and assembly tasks, whereas for object recognition and navigation tasks, it may be sufficient to detect features that are located roughly at the same depth, i.e., do not have large variations in disparities (Grimson 1993). Also, for object inspection and manipulation tasks, the work space consists typically of man-made objects with smooth and finished surface details. On the other hand, vision systems for navigation tasks can function in outdoor environments normally characterized as having rich, textured surface details.
- Parallel implementations can contribute towards making stereo vision an attractive tool for practical real-time applications.

ACKNOWLEDGMENTS

This research work was begun at the Institute for Real-Time Computer Control Systems and Robotics, University of Karlsruhe, Germany, within a TEMPUS individual mobility project, and finished at the Department of Engineering Cybernetics, Norwegian Institute of Technology, University of Trondheim, supported by the MOBATEL project and by the Research Council of Norway.

The author wishes to thank Prof. Dr.-Ing. R. Dillmann for introducing her to the subject and Prof. J. Balchen for his continuous support and interest. She is grateful also to T. Melen and M. Jenkin for their helpful comments on this paper.

REFERENCES

ALLEN, P. (1987). *Robotic Object Recognition Using Vision and Touch* (Boston Kluwer).

ARNOLD, R. and BINFORD, T. (1980). Geometric constraints in stereo vision. In *Image Processing for Missile Guidance, Proc.* SPIE, **238,** 281–292.

AYACHE, N. (1991). *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception* (MIT Press, Cambridge, MA, USA).

AYACHE, N. and HANSEN, C. (1988). Rectification of images for binocular and trinocular stereovision. In *Proc. 9th Int. Conf. Pattern Recognition*, Rome, pp. 15–20.

AYACHE, N. and LUSTMAN, F. (1991). Trinocular stereo vision for robotics. *IEEE Trans. Pattern Anal. Machine Intell.*, **13,** 73–85.

BALLARD, D. and BROWN, C. (1982). *Computer Vision* (Prentice-Hall: Englewood Cliffs, New Jersey).

BARNARD, S. (1989). Stochastic stereo matching over scale. *International Journal of Computer Vision*, **3,** 17–32.

BARNARD, S. and FISCHLER, M. (1982). Computational stereo. *ACM Computing Surveys*, **14,** 553–572.

BELHUMEUR, P. (1992). A bayesian treatment of the stereo correspondence problem using half-occluded regions. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Urbana, IL, pp. 506–512.

BELHUMEUR, P. (1993). A binocular stereo algorithm for reconstructing sloping, creased and broken surfaces in the presence of half-occlusion. In *Proc. 4th Int. Conf. on Computer Vision*, Berlin, pp. 431–438.

BELLMAN, R. and DREYFUS, S., (1962). *Applied Dynamic Programming* (Princeton Univ. Press).

BERTERO, M., POGGIO, T. and TORRE, V. (1988). Ill-posed problems in early vision. *Proc. of the IEEE*, **76,** 869–889.

BILBRO, G., MANN, R., MILLER, T., SNYDER, W., VAN DEN BOUT, D. and WHITE, M. (1989). Optimization by mean field annealing. In *Advances in Neutral Information Processing Systems I* (ed. D. S. Touretzky), pp. 91–98.

BLAKE, A. and ZISSERMANN, A. (1987). *Visual Reconstruction* (MIT Press, Cambridge, MA).

BLOSTEIN, S. and HUANG, T. (1987). Error analysis in stereo determination of 3D point positions. *IEEE Trans. Pattern Anal. Machine Intell.*, **9,** 752–765.

BOLTES, R., BAKER, H. and HANNAH, M. (1993). The JISCT stereo evaluation. In *Proc. ARPA Image Understanding Workshop*, pp. 263–274.

BOYER, M. and KAK, A. (1988). Structural stereopsis for 3D vision. *IEEE Trans. Pattern Anal. Machine Intell.*, **10,** 144–166.

BOYER, K., WUESCHER, D. and SARKAR, S. (1990). Dynamic edge warping: Experiments in disparity estimation under weak constraints. In *Proc. 3rd Int. Conf. Comp. Vision*, Osaka, pp. 471–475.

BRAUNEGG, D. (1990). Stereo feature matching in disparity space. In *Proc. IEEE Int. Conf. Robotics and Automation*, Cincinnati, OH, USA, pp. 796–803.

BRINT, A. and BRADY, M. (1990). Stereo matching of curves. *Image and Vision Computing*, **8,** 50–56.

BURNS, J., HANSON, A. and RISEMAN, E. (1986). Extracting straight lines. *IEEE Trans. Pattern Anal. Machine Intell.*, **8,** 424–455.

BUTLER, N. (1992). Matching stereo satellite images. In *Proc. Int. Conf. Pattern Recognition*, vol. 1, Hague, The Netherlands, pp. 716–719.

CANNY, J. (1985). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, **8,** 679–689.

CHAKRAPANI, P., KHOKHAR, A. and PRASANNA, V. (1992). Parallel stereo on fixed size arrays using zero-crossings. In *Proc. Int. Conf. on Pattern Recognition*, vol. 4, Hague, The Netherlands, pp. 79–82.

CHANG, C. and CHATTERJEE, S. (1990). Multiresolution stereo—A bayesian approach. In *Proc. 10th Int. Conf. Pattern Recognition*, Atlantic City, NJ, pp. 908–912.

CHANG, C., and CHATTERJEE, S. (1992). A deterministic approach for stereo disparity calculation. In *Proc. European Conference Computer Vision*, Italy, pp. 420–424.

CHAPRON, M., COCQUEREZ, N. and LIU, N. (1992). Precision of camera calibration and stereo vision performed by standard cameras and image digitizer. In *Proc. Int. Conf. Pattern Recognition*, vol. 1, Hague, The Netherlands, pp. 704–707.

CHEN, J.-S. and MEDIONI, G. (1990). Parallel Multiscale Stereo Matching Using Adaptive Smoothing. In *Proc. First European Conference Computer Vision*, Antibes, France, pp. 99–103.

COCHRAN, S. and MEDIONI, G. (1992). 3D surface description from binocular stereo. *IEEE Trans. Pattern Anal. Machine Intell.*, **14,** 981–994.

DERICHE, R. (1987). Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. Journal of Computer Vision*, **2,** 167–187.

DERICHE, R. and FAUGERAS, O. (1990). 2D curve matching using high curvature points: application to stereo vision. In *Proc. 10th Int. Conf. Pattern Recognition*, Atlantic City, NJ, pp. 240–242.

DHOND, U. and AGGARWAL, J. (1992). Analysis of the stereo correspondence process in scenes with narrow occluding objects. In *Proc. Int. Conf. Pattern Recognition*, vol. 1, Hague, The Netherlands, pp. 470–473.

DHOND, U. and AGGARWAL, J. (1992). Computing stereo correspondences in the presence of narrow occluding objects. In *Proc. Int. Conf. Comp. Vision and Pattern Recognition*, Champaign, IL, USA, pp. 758–760.

DHOND, U. and AGGARWAL, J. (1990). Binocular versus trinocular stereo. In *Proc. IEEE Int. Conf. Robotics and Automation*, Cincinnati, OH, USA, pp. 2045–2050.

DHOND, U. and AGGARWAL, J. (1989). Structure from stereo. *IEEE Trans. on Systems, Man, and Cybernetics*, **19,** 1489–1510.

DRUMHELLER, M. and POGGIO, T. (1986). On parallel stereo. In *Proc. IEEE Int. Conf. Robotics and Automation*, San Francisco, California, pp. 1439–1448.

EASTMAN, R. and WAXMAN, A. (1987). Using disparity functionals for stereo correspondence and surface reconstruction. *Comp. Vision Graph. Image Proc.*, **39,** 73–101.

FAUGERAS, O., LUONG, Q.-T., and MAYBANK, S. (1992). Camera self-calibration: theory and experiments. In *Proc. European Conference Computer Vision*, Santa Margherita, Ligure, Italy, pp. 321–334.

FAUGERAS, O. and TOSCANI, G. (1986). The calibration problem for stereo. In *Proc. Conf. Computer Vision and Pattern Recognition*, Miami Beach, Florida, USA, pp. 15–20.

FORNLAND, P., JONES, G., MATAS, G., and KITTLER, J. (1993). Stereo correspondence from junctions. In *Proc. 8th Scandinavian Conference on Image Analysis*, Tromsoe, Norway, vol. 1, pp. 449–455.

FUA, P. (1993). A parallel stereo algorithm that produces dense depth maps and preserves image and features. *Machine Vision and Applications*, **6,** 35–49.

GAGALOWICZ, A. and VINET, L. (1989). Region matching for stereo pairs. In *Proc. 6th Scandinavian Conf. on Image Analysis*, Oulu, Finland, vol. 1, pp. 63–70.

GEIGER, D., LADENDORF, B. and YUILLE, A. (1992). Occlusions and binocular stereo. In *Proc. European Conf. Comp. Vision*, Santa Margherita Ligure, Italy. pp. 425–433.

GRIMSON, W. (1993). Why stereo vision is not always about 3D reconstruction? MIT, AI Lab, A. I. Memo N 1435.

GRIMSON, W. (1985). Computational experiments with a feature-based stereo algorithm. *IEEE Trans. Pattern Anal. Machine Intell.*, **76,** 17–34.

GRIMSON, W. (1981). *From Images to Surfaces: A Computational Study of the Human Early Visual System* (Cambridge, MA, MIT Press).

GU, C. and WU, L. (1990). Structural matching of multiresolution for stereo vision. In *Proc. 10th Int. Conf. Pattern Recognition*, Atlantic City, NJ. pp. 243–245.

HANNAH, M. (1989). A system for digital stereo image matching. *Photogrammetric Engineering and Remote Sensing*, **55,** 1765–1770.

HARALICK, R. (1984). Digital step edges from zero-crossing of second directional derivatives. *IEEE Trans. Pattern Anal. Machine Intell.*, **6,** 58–68.

HARALICK, R., WATSON, L. and LAFFEY, T. (1983). The topographic primal sketch. *Int. Journal Robotic Research*, **2,** 50–72.

HOFF, W. and AHUJA, N. (1989). Surfaces from stereo: Integration feature matching, disparity estimation, and contour detection. *IEEE Trans. Pattern Anal. Machine Intell.*, **11,** 121–136.

HORAUD, R. and SKORDAS, T. (1989). Stereo correspondence through feature grouping and maximal cliques. *IEEE Trans. Pattern Anal. Machine Intell.*, **11,** 1168–1180.

ITO, M. and ISHII, A. (1986). Range and shape measurement using three-view stereo analysis. In *Proc. Conf. Comp. Vision and Pattern Recognition*, Miami Beach, Florida, USA, pp. 9–14.

JENKIN, M., JEPSON S. and TSOTSOS, J. (1991). Technique for disparity measurement. *CVGIP: Image Understanding*, **53,** 14–30.

JONES, D. and MALIK, J. (1992). A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proc. European Conf. Comp. Vision*, Santa Margherita Ligure, Italy. pp. 395–410.

JORDAN, J. and BOVIK, A. (1991). Dense stereo correspondence using color. *Proc. SPIE*, **1382,** 111–122.

JULESZ, B. (1960). Binocular depth perception of computer-generated patterns. *Bell Systems Technical J.*, **39,** 1125–1162.

KAHN, P., KITCHEN, L. and RISENMAN, E. (1990). A fast line finder for vision-guided robot navigation. *IEEE Trans. Pattern Anal. Machine Intell.*, **12,** 1098–1102.

KANADE, T. and OKUTOMI, M. (1991). A stereo matching algorithm with an adaptive window: theory and experiment. In *Proc. IEEE Int. Conf. on Robotics and Automation*, Sacramento, CA, USA, pp. 1088–1095.

KIERKEGAARD, P. (1993). Stereo matching of curved segments. In *Proc. 8th Scandinavian Conference on Image Analysis*, Tromsoe, Norway, vol. 1, pp. 457–464.

KIM, D. H., CHOI, W. Y. and PARK, R.-H. (1992). Stereo matching technique based on the theory of possibility. *Pattern Recognition Letters*. **13,** 735–744.

KIM, N. and BOVIK, A. (1988). A contour-based stereo matching algorithm using disparity continuity. *Pattern Recognition*, **21,** 505–514.

KIM, Y. and AGGARWAL, J. (1987). Positioning 3D objects using stereo images. *IEEE Journal of Robotics and Automation*, **3,** 361–373.

KITTLER, J., CHRISTMAS, W. and PETROU, M. (1993). Probabilistic relaxation for matching problems in computer vision. In *Proc. 4th Int. Conf. on Computer Vision*, Berlin, pp. 666–673.

KROTKOV, E. (1989). *Active Computer Vision by Cooperative Focus and Stereo*, (Springer Verlag).

VAN LAARHOVEN, P. and AARTS, E. (1987). *Simulated Annealing: Theory and Applications* (D. Riedel Publishing Co., Holland).

LAINE, A. and ROMAN, G. (1991). A parallel algorithm for incremental stereo matching on SIMD machines. *IEEE Trans. on Robotics and Automation*, **7,** 123–134.

LEE, J., CHO, S. and HA, Y. (1992). Neural network modelling of new energy function for stereo matching. *Proc. SPIE*, **1608,** 490–499.

LEW, M., WONG, K. and HUANG, T. (1992). Multi-scale stereo matching. In *Proc. Int. Conf. Pattern Recognition*, vol. 1, Hague, The Netherlands, pp. 620–623.

LIM, K. and PRAGER, R. (1983). Using Markov Random Fields to integrate stereo modules. In *Proc. Scandinavian Conference Image Analysis*, Troemso, Norway, pp. 435–440.

LITTLE, J. and GILLETT, W. (1990). Direct evidence for occlusion in stereo and motion. *Image and Vision Computing*, **8,** 328–340.

LIU, Y. and HUANG, T. S. (1991). Determining straight line correspondence from intensity images. *Pattern Recognition*, **24,** 489–504.

LLOYD, S., HADDOW, E. and BOYCE, J. (1987). A parallel binocular stereo algorithm utilizing dynamic programming and relaxation labelling. *Comp. Vision, Graphics, Image Processing*, **39,** 202–225.

LU, Y. and JAIN, R. (1989). Behavior of edges in scale space. *IEEE Trans. Pattern Anal. Machine Intell.*, **11,** 337–356.

MA, S., SI, S. and CHEN, Z. (1992). Quadric curve-based stereo. In *Proc. Int. Conf. Patt. Recognition*, vol. 1, Hague, The Netherlands, pp. 1–4.

MARAPANE, S. and TRIVEDI, M. (1992). Multi-primitive hierarchical stereo system. In *Proc. Int. Conf. Computer Vision and Pattern Recognition*, Champaign, IL, USA, pp. 499–505.

MARAPANE, S. and TRIVEDI, M. (1989). Region-based stereo analysis for robotic applications. *IEEE Trans. Systems, Man, and Cybernetics*, **9,** 1447–1464.

MARCH, R. (1988). Computation of stereo disparity using regularization. *Pattern Recognition Letters*, **8,** 181–187.

MARR, D. and POGGIO, T. (1976). Cooperative computation of stereo disparity. *Science*, **194,** 283–287.

MARR, D. and POGGIO, T. (1979). A computational theory of human stereo vision. *Proc. Royal Soc. London*, **B204,** 301–328.

MARR, D. and HILDRETH, E. (1980). Theory of edge detection. *Proc. Royal Soc. London*, **B207,** 187–217.

MARROQUIN, J., MITTER, S. and POGGIO, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, **82,** 76–89.

MATTHIES, L. (1992). Passive stereo range imaging for semi-autonomous land navigation. *Journal of Robotic Systems*, **9,** 787–816.

MAYHEW, J. and FRISBY, J. (1981). Psychophysical and computation studies towards a theory of human stereopsis. *Artif. Intell.*, **17,** 349–385.

MCINTOSH, J. and MUTCH, K. (1988). Matching straight lines. *Comp. Vision, Graphics, Image Processing*, **43,** 386–408.

McKEOWN, D. and HSIEH, Y. (1992). Hierarchical waveform matching: a new feature-based stereo technique. In *Proc. Int. Conf. Comp. Vision and Patt. Recognition*, Champaign, IL, USA, pp. 513–519.

MEDIONI, G. and NEVATIA, R. (1985). Segment-based stereo matching. *Comp. Vision, Graphics, Image Processing*, **31,** 2–18.

MOHAN, R., MEDIONI, G. and NEVATIA, R. (1989). Stereo error detection, correction, and evaluation. *IEEE Trans. Pattern Anal. Machine Intell.*, **11,** 113–120.

MOHAN, R. and NEVATIA, R. (1992). Perceptual organization for scene segmentation and description. *IEE Trans. Pattern Anal. Machine Intell.*, **14,** 616–635.

MORAVEC, H. (1977). Towards automatic visual obstacle avoidance. In *Proc. 5th Int. Joint Conf. Artif. Intelligence*, Massachusetts, pp. 584.

MOUSAVI, M. and SCHALKOFF, R. (1991). Stereo vision: a neural network application to constraint satisfaction problem. *Proc. SPIE*, **1382,** 228–239.

NASRABADI, N. (1992). A stereo vision technique using curve segments and relaxation matching. *IEEE Trans. Pattern Anal. Machine Intell.*, **14,** 566–572.

NASRABADI, N. and CHOO, C. (1992). Hopfield network for stereo vision correspondence. *IEEE Trans. on Neural Networks*, **3,** 5–13.

NISHIHARA, H. and POGGIO, T. (1984). Stereo vision for robotics. In *Robotics Research* (ed. M. Brady and R. Paul) (MIT Press), pp. 490–505.

OHTA, Y. and KANADE, T. (1985). Stereo by intra- and inter-scanline search. *IEEE Trans. Pattern Anal. Machine Intell.*, **7,** 139–154.

OHTA, Y., WATANABE, M. and IKEDA, K. (1986). Improving depth map by right-angled trinocular stereo. In *Proc. 8th Int. Conf. Pattern Recognition*, Paris, pp. 519–521.

OHTA, Y., YAMAMOTO, T. and IKEDA, K. (1988). Collinear trinocular stereo using two-level dynamic programming. In *Proc. 9th Int. Conf. Pattern. Recognition*, Rome, pp. 658–662.

OKUTOMI, M. and KANADE, T. (1993). A multiple-baseline stereo. *IEEE Trans. Pattern Anal. Machine Intell.*, **15,** 353–363.

OLSEN, S. (1990). Stereo correspondence by surface reconstruction. *IEEE Trans. Pattern Anal. Machine Intell.*, **12,** 309–315.

OLSEN, S. (1992). Epipolar line estimation. In *Proc. European Conf. Comp. Vision*, Santa Margherita Ligure, Italy, pp. 307–311.

PIETIKÄINEN, M. and HARWOOD, D. (1986). Depth from three camera stereo. In *Proc. Conf. Comp. Vision and Pattern Recognition*, Miami Beach, Florida, pp. 2–8.

POGGIO, T., TORRE, V. and KOCH, C. (1985). Computation vision and regularization theory. *Nature*, **317,** 314–319.

POLLARD, S., MAYHEW, J. and FRISBY, J. (1985). PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, **14,** 449–470.

PONG, T., HARALICK, R. and SHAPIRO, L. (1989). Matching topographic structures in stereo vision. *Pattern Recognition Letters*, **9,** 127–36.

PRAZDNY, K. (1985). Detection of binocular disparities. *Biol. Cybernetics*, **52,** 93–99.

ROBERT, L. and FAUGERAS, O. (1991). Curve-based stereo: figural continuity and curvature. In *Proc. IEEE Conf. Comp. Vision and Patt. Recognition*, Lahaina, Hawaii, pp. 57–62.

ROSENFELD, A., HUMMEL, R. and ZUCKER, S. (1976). Scene labelling by relaxation operation. *IEEE Trans. Systems, Man, and Cybernetics*, **6,** 420–423.

SANDER, P., VINET, L., COHEN, L. and GAGALOWICZ, A. (1989). Hierarchical region-based stereo matching. In *Proc. 6th Scandinavian Conf. on Image Analysis*, Oulu, Finland, vol. 1, pp. 71–78.

SANGER, T. (1988). Stereo disparity computation using Gabor filters. *Biological Cybernetics*, **59,** 405–418.

SHERMAN, D. and PELEG, S. (1990). Stereo by incremental matching of contours. *IEEE Trans. Pattern Anal. Machine Intell.*, **12,** 1102–1106.

SHIRAI, Y. (1992). 3D computer vision and applications. In *Proc. Int. Conf. Pattern Recognition*, vol. 1, Hague, The Netherlands, pp. 236–245.

SINGER, M. (1987). Significant feature detection and matching in image pairs. In *Proc. Int. Joint Conf. Artif. Intelligence*, Milan, pp. 829–831.

STEWART, C. (1992). On the derivation of geometric constraints in stereo. In *Proc. Int. Conf. Comp. Vision and Patt. Recognition*, Champaign, IL, USA, pp. 769–772.

STEWART, C. and DYER, C. (1990). Simulation of a connectionist stereo algorithm on a shared-memory multiprocessor. In *Parallel Algorithms for Machine Intelligence and Vision* (ed. V. Kumar), (Springer Verlag), pp. 341–359.

STEWART, C. and DYER, C. (1988). The trinocular general support algorithm: a three camera stereo algorithm for overcoming binocular matching errors. In *Proc. 2nd Int. Conf. on Comp. Vision*, Tampa, Florida, pp. 134–138.

STEWART, C and DYER, C. (1987). A connectionnist model for stereo vision. *Proc. IEEE 1st Int. Conf. on Neural Networks*, San Diego, USA, vol. 4, pp. 215–223.

STEWART, C., and MACCRONE, J. (1990). Experimental analysis of a number of stereo matching components using LMA. In *Proc. 10th Int. Conf. Pattern Recognition*, Atlantic City, NJ, pp. 254–258.

SUGIMOTO, K., TAKAHASHI, H. and TOMITA, F. (1988). Scene interpretation based on boundary representations of stereo images. In *Proc. 9th Int. Conf. Pattern Recognition*, Rome, pp. 155–159.

TAKAHASHI, H. and TOMITA, F. (1988). Planarity constraint in stereo matching. In *Proc. 9th Int. Conf. Pattern Recognition*, Rome, pp. 446–449.

TERZOPOULOUS, D., WITKIN, A. and KASS, M. (1987). Stereo matching as constraint optimization using scale continuation methods. *Proc. SPIE, vol. 754*, pp. 92–99.

TOBORG, S., and HWANG, K. (1991). Cooperative Vision Integration through data-parallel neural computations. *IEEE Transactions on Computers*, **40,** 1368–1379.

TRIENDL, E., and KRIEGMAN, D. (1987). Stereo vision and navigation within buildings. In *Proc. IEEE Int. Conf. Robotics and Automation*, Raleigh, pp. 1725–1730.

TSAI, R. (1986). An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. Conf. on Comp. Vision and Pattern Recognition*, Miami Beach, Florida, pp. 364–374.

TSUJI, S., ZHENG, J. and ASADA, M. (1986). Stereo vision of a mobile robot: world constraints for image matching and interpretation. In *Proc. IEEE Int. Conf. Robotics and Automation*, San Francisco, California, pp. 1594–1599.

VERRI, A. and TORRE, V. (1986). Absolute depth estimates in stereopsis. *Journal Opt. Soc. Amer.*, **3,** 297–299.

VLEESCHAUWER, D. (1993). An intensity-based, coarse-to-fine approach to reliably measure binocular disparity. *CVGIP: Image understanding*, **57,** 204–218.

WANG, Y. and PAVLIDIS, T. (1990). Optimal correspondence of string subsequences. *IEEE Trans. Pattern Anal. Machine Intell.*, **12,** 1080–1087.

WATANABE, M. and OHTA, Y. (1990). Cooperative integration of multiple stereo algorithms. In *Proc. 3rd Int. Conf. on Comp. Vision*, Osaka, pp. 476–480.

WENG, J., AHUJA, N. and HUANG, S. (1992). Matching two perspective views. *IEEE Trans. Pattern Anal. Machine Intell.*, **14,** 806–825.

WESTMAN, T. (1989). A combined region- and contour-based stereo vision system. In *Proc. 6th Scandinavian Conf. on Image Analysis*, Oulu, Finland, vol. 1, pp. 79–87.

WILDES, R. (1991). Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(8): 761–774.

WILLIAMS, L. and ANANDAN, P. (1986). A coarse-to-fine control strategy for stereo and motion on a mesh-connected computer. In *Proc. of IEEE Conf. on Comp. Vision Pattern Recognition*, Miami Beach, Florida, pp. 219–226.

XU, G., KONDO, H. and TSUJI, S. (1989). A region-based stereo algorithm. In *Proc. 11th Int. Joint Conf. Artif. Intelligence*, Detroit, pp. 1661–1666.

YASHIDA, M., KITAMURA, Y. and KIMACHI, M. (1986). Trinocular vision: new approach for correspondence problem. In *Proc. 8th Int. Conf. Pattern Recognition*, Paris, pp. 1041–1044.

YOKOYA, N. (1992). Surface reconstruction directly from binocular stereo images by multiscale multistage regularization. In *Proc. Int. Conf. Patt. Recognition*, vol. 1, Hague, The Netherlands, pp. 642–646.

YUILLE, A. (1989). Energy functions for early vision and analog networks. *Biological Cybernetics*, **61,** 115–123.

YUILLE, A., GEIGER, D. and BULTHOFF, H. (1991). Stereo integration, mean field theory and psychophysics. *Network*, **2,** 423–442.

ZHOU, Y. and CHELLAPPA, R. (1988). Neural network approach to stereo matching. *Proc. SPIE*, **974,** 243–250.