References

[1] C. M. Oman, "Sensory conflict in motion sickness: An observer theory approach," in *Pictorial Communication in Virtual and Real Environments*, S. R. Ellis, M. K. Kaiser, and A. J. Grunwald, Eds. Taylor and Francis, 1991.

[2] J. J. Gibson, *The Perception of the Visual World*. Boston, MA: Houghton Mifflin, 1950.

[3] J. J. Gibson, P. Olum, and F. Rosenblatt, "Parallax and perspective during aircraft landings," *Amer. J. Psychol.*, vol. 68, pp. 372–385, 1955.

[4] J. J. Gibson, "Visually controlled locomotion and visual orientation in animals," *British J. Psychol.*, vol. 49, pp. 182–194, 1958.

[5] D. A. Gordon, "Static and dynamic visual fields in human space perception," *J. Opt. Soc. Amer.*, vol. 55, pp. 1296–1303, 1965.

[6] ——, "Perceptual basis of vehicular guidance," *Public Roads*, vol. 34, pp. 53–68, 1966.

[7] R. Warren, "The perception of egomotion," *J. Experimental Psychol: Human Perception and Performance*, vol. 2, no. 3, pp. 448–456, 1976.

[8] ——, "Optical transformation during movement," AFOSR-TR-82-1028, NTIS AD-A122 275/1, Dep. Psychol., Ohio State Univ., Columbus, OH, Oct. 1982.

[9] G. L. Zacharias, "Flow-field cueing conditions for inferring observer self-motion," Cambridge, MA: Bolt Beranek and Newman, Rep. 5118, Sept. 1982.

[10] G. L. Zacharias, A. K. Caglayan, and J. B. Sinacori, "A visual cueing model for terrain-following applications," *J. Guidance, Contr. and Dynamics*, vol. 8, no. 2, pp. 201–207, Mar.–Apr. 1985.

[11] G. G. Denton, "The influence of visual pattern on perceived speed," *Perception*, vol. 9, pp. 393–402, 1980.

[12] D. H. Owen, L. Wolpert, and R. Warren, "Effects of optical flow acceleration, edge acceleration, and viewing time on the perception of egospeed acceleration," NASA Sci. Tech. Inform. Facility, 79-140, 1984.

[13] C. A. Awe, W. W. Johnson, and F. Schmitz, "Inflexibility in selecting the optical basis for perceiving speed," in *Proc. Human Factors Soc. 33rd Annu. Meeting*, 1989, pp. 1440–1444.

[14] J. F. Larish and J. M. Flach, "Sources of optical information useful for the perception of velocity of rectilinear selfmotion," *J. Experimental Psychol.: Human Perception and Performance*, vol. 16, no. 2, pp. 295–302, 1990.

[15] W. W. Johnson, C. T. Bennett, K. O'Donnell, and A. V. Phatak, "Optical variables useful in the active control of altitude," in *Proc. 23rd Annu. Conf. Manual Contr.*, Cambridge, MA, June 1988.

[16] K. R. Boff, L. Kaufman, and J. P. Thomas, *Handbook of Perception and Human Performance, Vol. 1*. New York: Wiley, pp. 16-18, 17-18–17-20, 1986.

[17] A. J. Grunwald and S. Kohn, "Flight-path estimation in passive low-altitude flight by visual cues," *AIAA J. Guidance, Contr. and Dynamics*, vol. 16, no. 2, pp. 363–370, Mar.–Apr. 1993.

Depth Map Construction from Range-Guided Multiresolution Stereo Matching

Kevin Tate and Ze-Nian Li

Abstract—This paper describes a multiresolution method for the acquisition of a complete, relatively noise-free, and high-resolution depth map from a low-resolution laser range image and a stereo pair of high-resolution intensity images. Depth information from the laser range data is used to constrain the initial search in stereo matching. The inter- and intralevel linkings of edges in the pyramid allow a process where the coarse laser depth information drives a multiresolution stereo-matching process to construct a high-resolution depth map. The motivation for this new approach to multi-sensor integration is to offset the advantages and disadvantages of traditional stereo matching and triangulation range finding approaches.

I. Introduction

The acquisition of accurate high-resolution 3-D data is an important problem in computer vision. Accurate and complete depth information is important for many tasks in computer vision such as robot navigation and object modeling and recognition.

Two of the most common techniques for acquiring 3-D data are stereo imagery and laser range finding. Both of these techniques have appealing aspects which make them attractive candidates for 3-D data acquisition. Unfortunately, each also has associated problems which make them unattractive as a definitive means of 3-D data acquisition.

There are two main methods for the implementation of laser range finding, one based on triangulation, the other on time-of-flight [3]. This paper is only concerned with triangulation range finders since they are cheaper and more easily constructed, and thus, more common.

The use of laser range finders for the acquisition of 3-D data is an increasingly common approach in computer vision research [13]. The primary reason for this is that 3-D data is directly acquired in a well-defined and easily understood manner without any negative effects due to lighting.

Triangulation range finding has a number of drawbacks [14], especially limited range, resolution, and speed. Range is limited by the power of the laser and the loss of accuracy in long range. Resolution of these devices is limited by the accuracy of the optical mechanism used to sweep the laser across the image. Moreover, most triangulation range finders have a built-in image processing software for locating the laser scan line, which usually involves detecting and thinning laser stripes in an intensity image. Naturally, the speed and accuracy of the software adversely affect the performance of the range finder.

Triangulation range finding also has the so-called *missing parts* problem, resulting from either: a) the triangulation geometry of the range finder, or b) the surface properties of the objects in the scene. Triangulation geometry problems occur when there are surfaces that either the laser or the camera cannot "see." The larger the separation between camera and laser, the more prevalent this problem. Surface properties of the objects in a scene can cause problems when they reflect or scatter the laser dot or stripe in an irregular fashion, such as surfaces with deep or irregularly shaped holes.

The primary advantage of stereo imagery is that image intensity data can be acquired rapidly (many frames per second) and at high resolution (512×512 pixels or better). Its main disadvantage is the so-called *correspondence* problem, which results from attempts to match features between the two images. Another disadvantage with stereo matching, and intensity imagery, in particular, is that it is subject to lighting and surface reflectance effects. These can lead to homogeneous intensity values across edges, making detection of actual edges difficult, and introducing anomalies such as false edges due to bright or dull spots.

Various constraints have been introduced to alleviate the correspondence problem in stereo matching. The most common are

compatibility, *uniqueness*, and *figural continuity* [17], [18]. These constraints have been shown to be important components of stereo matching [9]. However, these constraints are occasionally too weak to enforce surface smoothness and correctness of correspondence completely, since edges may not be present at major surface discontinuities where two adjacent but overlapping objects produce a homogeneous intensity, or where surface discontinuities are present [10]. This problem is most apparent when attempting to determine the final depth map.

Numerous efforts have been made in integrating multiple sensors; many have involved the use of stereo-intensity imagery or lasers. The NAVLAB [22] used intensity/texture low-resolution color images and laser range data for its autonomous vehicle navigation. Krotkov [15] demonstrated active depth recovery by cooperative focus and stereo. Ayache [2] developed a formalism for the combination of multisensory data, in particular, the combination of stereo and motion for dynamic robot vision. A multiresolution approach to integrating stereo, focus, and fixation was presented by Das and Ahuja [7]. In their approach, information in the peripheral, low-resolution parts of the current image is used for obtaining coarse surface estimates in the vicinity of the next fixation point. The coarse estimates are refined after the camera is refixated and images of higher resolutions are processed. Recently, Chu and Aggarwal [6] reported the integration of laser range (Ladar), intensity, velocity, and thermal data in an automatic image interpretation system.

The research described in this paper is motivated by the analysis of the advantages and disadvantages of traditional intensity stereo matching and triangulation range finding approaches. It presents an efficient range-guided multiresolution stereo method for the acquisition of a complete, relatively noise-free, and high-resolution depth map from a low-resolution laser range image and a stereo pair of high-resolution intensity images. The result is a new approach to multisensor integration, which uses multiresolution pyramids as a framework for a system which incorporates data from stereo and laser range, while taking into account their advantages and disadvantages.

The remainder of this paper is organized as follows. Section II describes the interactions between the laser range and stereo-intensity data, and introduces the range-guided multiresolution stereo matching and depth map construction based on these interactions. Section III presents the implementation and experimental results. Section IV contains future research issues, and Section V is the conclusion.

II. RANGE-GUIDED MULTIRESOLUTION STEREO AND DEPTH MAP CONSTRUCTION

The key components of the approach outlined in subsequent sections are multiresolution processing and the complementary aspects of stereo-intensity image processing and laser range-finding techniques. Essentially, a coarse depth map from a laser range finder is used to guide a multiresolution stereo matching and depth map generation process, resulting in a high-resolution depth map. Coarse range data is acquired using an inexpensive triangulation laser range finder at a level of resolution sufficiently small to allow rapid data acquisition, while the resolution of the resulting fine depth map is determined by the resolution of the stereo images.

The key aspects of the described approach are:

- Multiresolution processing is performed using the constructs provided by the intensity pyramid [4]. Using the intensity pyramid model, a stereo pair of high-resolution intensity images are brought to the same level of resolution as a comparatively low-resolution range image. The common edges between the range and intensity images at this common level of resolution are used to start multiresolution stereo matching and depth map construction, progressing level-by-level to the bottom of the pyramid. Multiresolution pyramids allow iterative improvement in stereo matching and generated depth maps, while at the same time reducing errors due to depth estimation errors at low resolution levels.

- Stereo-intensity data is used as a supplementary source of depth information, providing smoother depth values at edges (where many laser reflectance problems occur) and finer detail (since in general, intensity data is capable of providing finer detail from an object than laser data). It is also used to provide depth information for surfaces or portions of surfaces the laser range finder is unable to view due to surface reflectance or absorption.

- The weaknesses of the traditional stereo-matching constraints are compensated for by the presence of depth information gathered without any lighting dependencies. The coarse range data is converted to estimated disparity values using known camera geometry properties and standard mathematical relationships at edges common to the intensity and range data to guide the search for a matching edge in stereo matching, thus reducing the stereo-correspondence problem.

- The acquisition of coarse range data reduces the time of its acquisition.

Fig. 1 shows the order of the processing. First, image intensity pyramids are constructed for the left and right intensity images to bring them to the same level of resolution as the range image. Then, edge detection is performed in each image in the intensity pyramid and in the range image. In both pyramids, bidirectional links defining a parent–child relationship are set up between corresponding edges at different levels of resolution [4], [11]. Also, edge segments are formed by linking each point on the edge with its primary and secondary predecessors and successors on the segment. Edge correspondences are then found between the range image and the top level of the pyramid for the registered intensity image (i.e., the image acquired by the same camera). The depth information from along these edges is then projected to the next level of the pyramid using the interlevel edge linking. Stereo matching is performed at this pyramid level using the depth values as guides to constrain the search for a potential match. Once all edges with associated depth have been matched, remaining edges are matched. Stereo matching is then followed by boundary improvement at the current level of resolution. New depth values are determined using the smoother depth values acquired through stereo matching and the depth values from the range image. These new depth values are projected to the next pyramid level for use as estimated depth values for stereo matching at that level. This process of depth projection, stereo matching, and boundary improvement is carried out to the bottom of the pyramid.

A. Intensity Pyramid Construction

The most important feature of the intensity pyramid for the purposes of this paper is the hierarchical linking of edges with common properties within the pyramid. As described in detail in Section III-A, two types of edge linking are used. The first, *intralevel edge linking*, is used to form edge segments within an individual pyramid level. The second, *interlevel edge linking*, links edges into a parent–child relationship between pyramid levels. Both types of linking play important roles in hierarchical stereo matching.
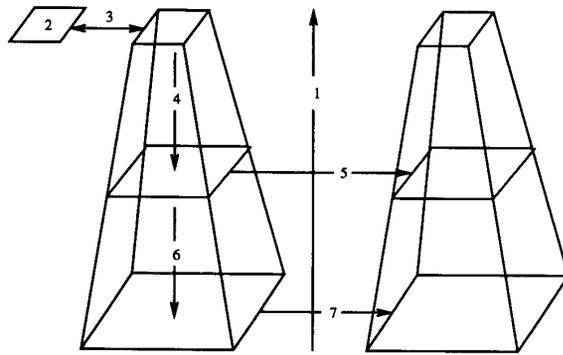
1. Intensity Pyramid Construction.
2. Range Image Edge Detection.
3. Edge Combination.
4. Project Depth Values.
5. Stereo Matching and Boundary Improvement.
6. Project Depth Values.
7. Stereo Matching and Boundary Improvement.

Fig. 1. Overview of the order of processing.

B. Hierarchical Stereo Matching

Stereo matching is performed hierarchically, and proceeds level-by-level in two passes. In the first pass, all possible matches using depth as a guide are made; these matches constrain the possible matches in the second pass. The second pass, which roughly corresponds to traditional stereo matching is used to complete the matches. All matches are performed using the epipolar and figural continuity constraints.

Two types of figural continuity are used. The first, *intralevel* figural continuity, applies within a pyramid level, and corresponds to the traditional definition of the constraint since it ensures that match disparities are smooth along an edge segment. It is enforced while searching along an edge segment. The second type, *inter-level* figural continuity, applies between pyramid levels, and ensures that matches along the same edge segment at consecutive levels of resolution agree with each other. It is enforced using the explicit interlevel linking between edge points in the image pyramid. The main purpose of interlevel figural continuity is to reduce the number of initial matching errors at low levels of resolution by ensuring that the same match at higher resolution is still possible. The use of interlevel figural continuity is an important departure from other stereo matching approaches [17], [18], [9], since it not only allows matches at low levels of resolution to guide matches at higher resolution, but also a greater interaction between levels of resolution.

Fig. 2 shows the steps involved in matching a single point. Essentially, matching proceeds from the top left to the bottom right of the source image. Each unmatched edge point encountered in this search is used as a starting point for matching the edge segment on which it lies. For each such edge point, a list of potential matches is made containing edges with compatible edge direction and magnitude, and which lie within a given search window as per the epipolar constraint. If the current pass over the image is using depth as a guide and the given point has an estimated depth value as determined from matching at a lower level of resolution or from the range image, then the list of potential matches is ordered by increasing difference with the estimated disparity value. (Estimated disparity values are derived from the estimated depth values using the mathematical relationship between depth and disparity and the known camera geometry parameters [1].)

If, however, matching using depth as a guide is not being used, then any ordering of potential matches can be used; in this case, matches are ordered by decreasing disparity. Once the potential matches list has been ordered, the list is searched in order until a match is found with which at least one linked-to child agrees within the interlevel figural continuity threshold. The validity of this match
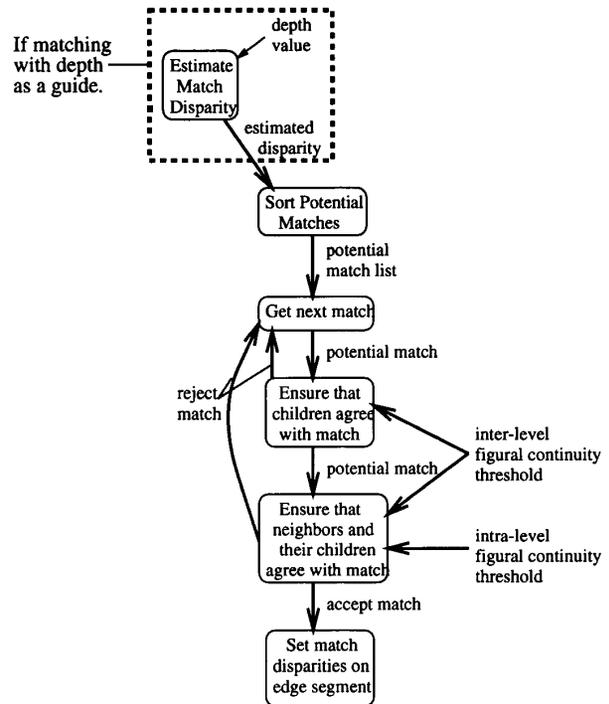
Fig. 2. The stereo-matching algorithm. Shown are the processing steps involved in attempting to match an individual edge point. The only difference between matching in the two passes is in the order in which the potential match list is sorted. If it is the first pass (matching with depth as a guide), the list is sorted by the difference in disparity between an element and the estimated disparity determined from previous stereo matching or the range image, depending on the current location in the image pyramid. Each potential match must first be approved by its children in the pyramid with respect to the interlevel figural continuity constraint, and then must be approved by its neighbors and their children on its edge segment with respect to intra- and interlevel figural continuity.

is tested by ensuring that a sufficient percentage of the edge points along the edge and their children agree with the match.

A match is evaluated sequentially along an edge segment using the primary predecessors and successors of the start point (the first edge point encountered in the image scan). As the edge is traversed, a current disparity measure is kept, recording the disparity of the last matched point on the edge. For each new neighbor, the

match with disparity closest to the current disparity is found with which at least one linked-to child agrees. A child is said to agree with a match if it is capable of making a match to the equivalent target-edge point within the interlevel figural continuity threshold (currently set at three plus two times the disparity from the candidate match at the previous level). If at least one linked-to child agrees with the match at the current disparity and the difference between the disparity of this match and the current disparity is less than the intralevel figural continuity threshold (currently set at 3), the edge point is said to agree with the match and the current disparity is set to the disparity of the current match. Then, the next neighbor is tested using this new disparity. If the new disparity does not agree, then the next neighbor is tested using the same current disparity. If some percentage (currently 60%) of the edge points (and their children) in an edge segment, agree with a match, then it is accepted and all edges in the segment and the edges matched to in the target pyramid are marked as matched. Edge points on the same edge segment which were unable to agree with the match are now marked as matched, and a disparity is calculated for them based on the average disparity of their primary and secondary predecessor and successor edge points. If an insufficient percentage of points on the edge agree with a match, then a new match is attempted, resulting in the same process of attempting to reach consensus along the edge at a new disparity. If there are no more potential edges, the edge segment remains unmatched and the image search for unmatched edges continues.

Once all possible matches using depth as a guide have been made, the image is searched again in order to match remaining edges using a nonguided matching process. This matching is performed in exactly the same fashion as matching by depth, except that the size of the potential match lists for individual edge points is constrained by the matches made in the first pass.

Two examples have been derived from actual data to illustrate the effectiveness of the hierarchical approach to stereo matching. Note that matching is performed from the left to the right image in both examples.

The first example, shown in Fig. 3, illustrates how the algorithm is capable of dealing with ambiguity and also how it can set the disparity of edge points after the edge segment has been matched. Matching begins at (96,72), where it has an associated depth of 62, producing an estimated disparity of 4. Its left neighbor (95,71) is unable to match since it has no compatible edges. To its right (97,73) matches at disparity 4, and (98,73) and (99,72) at disparity 5. For each of the next six edges (100,71) to (105,67), matches are made at disparity 4 or 5 since the current disparity measure and the figural continuity threshold do not allow matches to the other edge segment. (106,67) and (108,66) remain unmatched since they have no compatible edges, and the rest successfully match using the current disparity. Once matching is complete on this edge, three edge points are unmatched: (95,71), (106,67), and (108,66). These are all assigned disparity measures based on the average disparity of their predecessors and successors on the edge segment, which in this case gives them all disparity 4. These assigned disparity measures are important since they increase the density of the disparity map, thus providing more information to later processing steps which use the disparity map. This example can also be used to illustrate how matching would fail if it were to start at (100,71) with disparity 15. Although all its successors on the edge segment (101,71) to (107,66) can match at this disparity, the fact that the rest of the edge cannot will cause the match to be rejected, since in this case only 7 out 17 edge points on the segment successfully matched. Given that this match would be rejected, the second match, at disparity 4 would be attempted and would succeed. This
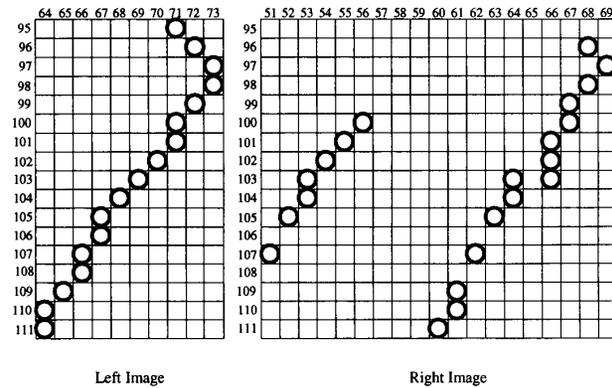
Fig. 3. Stereo matching, example 1. An illustration of dealing with ambiguity and in setting disparity values for unmatched edge points.

illustrates the point that the figural continuity constraint and the use of a current disparity measure along an edge segment ensure that the same match should be reached regardless of the starting point for matching.

The second example, shown in Fig. 4, illustrates the role played by interlevel figural continuity in stereo matching. The figure shows edge points present at two levels of the pyramid: Levels 6 and 7. The overlapped child sets for the edges in the left image at Level 6 are shown by arrows leading to 4 × 4 boxes at Level 7. The edges appearing within these boxes at Level 7 are all linked to by the edge points at Level 6. When matching the edge segment at Level 6, the depth estimate for (20,46) is large, indicating that a match should be performed to (20,40) in the right image even though (20,43) is the correct match. However, the match to (20,40) is rejected by the children, since none of them can match within the interlevel figural continuity threshold at this disparity. The threshold is violated since the proposed match at disparity 6 means that at least one of the children must be able to match at a disparity within 12 ± 3. (As outlined above, the interlevel figural continuity threshold is determined by doubling the disparity from the previous level and ensuring that the difference between this disparity and a child's match disparity is at most 3.) In this case, none of (40,91), (41,92), or (42,93) can make such a match so it is rejected. The other match for (20,46) to (20,43) is then attempted, and accepted by all of its children. This match at disparity 3 is then tested among all the successors of (20,46) along the edge segment at pyramid Level 6. In this case, all of the edges on this segment and their children agree with this match.

These examples illustrate many of the important features of this approach to range-guided stereo matching. As demonstrated, this approach is capable of dealing with poor depth estimates from the range data, allows matching of nearly horizontal line segments, and is not restricted to matching of straight edge segments. Also, the method of enforcing intralevel figural continuity along edge segments allows the disparity to gradually increase, decrease, or modulate along the edge. This intralevel figural continuity is enhanced by the interlevel figural continuity constraint, which ensures that matches are compatible with their linked-to children's.

C. Hierarchical Dense Depth Map Generation

There are two primary components for depth map generation. The first component is based on hierarchical averaging of depth values from the original laser range map with depth values resulting from stereo matching. The primary motivation behind this tech-
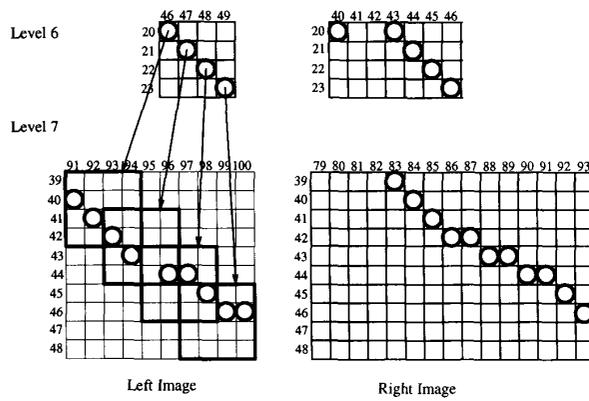
Fig. 4. Stereo matching, example 2. An illustration of the role of inter-level figural continuity.

Fig. 5. Proposed camera geometry. This camera system consists of a traditional triangulation range finder and an additional camera orthogonal to this system. The bottom camera collects both laser range and intensity data, while the added camera collects intensity data only.

nique is that the laser range depth values are generally reliable on the interior of regions and not reliable on region boundaries. Therefore, the averaging technique relies heavily on stereo-matching depth values at boundaries and laser range depth values in surface interiors. The second component of depth map generation is a surface interpolation method used to create depth values for surfaces or portions of surfaces missing from the laser range map yet present in the stereo-intensity data. This method is only capable of interpolating planar or simply curved surfaces.

1) Hierarchical Dense Depth Map Generation with Boundary Improvement: After the completion of stereo matching at a pyramid level, a depth map is generated. Depth maps are generated using an averaging technique with *boundary improvement*, which is the process of ensuring that object boundaries are reliably followed. This method of depth map generation is also hierarchical in nature, since it combines depth information from the previous pyramid level with the depth values obtained at the current pyramid level from stereo matching.

The primary problem in hierarchical depth map generation is determining the best method for combining the various sources of depth information. One source of depth is the laser range depth map, which is coarse and cannot be relied upon for a complete surface description, not only due to its resolution, but also due to noise in the range image, especially at the edges. The other source of depth information are the disparity measures produced from stereo matching. These values are smoother than along the edges of the range image due to the enforcement of figural continuity, but unfortunately are sparse since they correspond only to the edge locations.

At each pyramid level, a depth map is generated from two depth maps. The first depth map is the depth map from the previous pyramid level, called the *estimated depth map*. These are the same depth values used to drive stereo matching at this pyramid level. The estimated depth map is constructed using the nonoverlapped pyramid scheme, where each of a parent's 2 × 2 children are assigned to the same depth value. Note that at the first pyramid level where stereo matching takes place, the estimated depth map is constructed from the coarse laser range data. On all subsequent pyramid levels, it is constructed from the previous level's generated depth map. The second depth map is the *stereo-depth map*, which is the sparse map of depth values resulting from stereo matching.

Depth map construction by averaging is performed on a point-by-point basis on a single pyramid level. For each point, k-nearest averaging is used within a 3 × 3 neighborhood centered on the point. For each neighbor within this neighborhood, there are two possible depth values—an estimated depth value and a stereo-depth value. If a neighbor has only one of these, then this value is used for the k-nearest averaging. If it has both, then the stereo-depth value is used. The k-closest values chosen in this manner are averaged together and assigned to the given point.
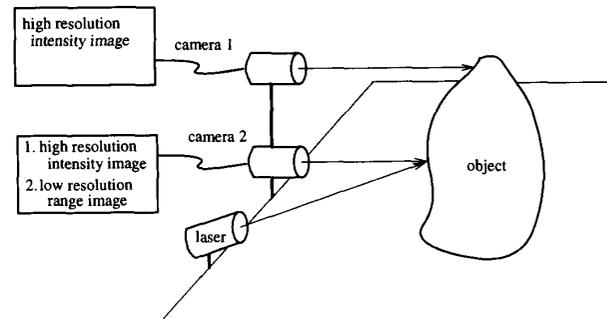
Boundary improvement is performed by thresholding the calculated depth values. At each pyramid level, a threshold value is chosen and all depth values less than this threshold are zeroed. This enforces object boundaries by clearly defining outer boundaries and by removing "noisy" depth values created by averaging.

2) Depth Map Generation for Missing Parts: Attempting to derive depth information for missing parts is a difficult problem. Other researchers have found single-camera intensity data useful for the recovery of missing parts [5]. The major purpose of the method presented in this paper is to show how the availability of stereo-intensity data makes it possible to develop a simple approach for depth map generation of missing parts. The novel technique presented in this paper combines various existing algorithms for this purpose. This technique performs reasonably well when the missing part meets the following criteria: a sufficiently large portion of the surface is visible, and it is simply curved or planar. As described below, missing parts are discovered at a low level of resolution, then labeled as a missing part at the highest resolution using the hierarchical edge linking.

The camera geometry shown in Fig. 5 is sufficient to detect missing parts lying above and to the left of the region visible from the traditional laser range finding system. Missing parts in this system are those visible in both intensity cameras, but not visible from the position of the laser. These surfaces are discovered by finding edge segments enclosing regions with no corresponding laser range data. Such regions are found after stereo matching and boundary improvement on a pyramid level using the compact object extraction algorithm of Hong and Shneier [12]. The hierarchical nature of their algorithm and the interlevel linking at edge boundaries of these regions, allows missing part region detection at low resolution and labeling at the highest level of resolution, the bottom of the pyramid.

Given a region labeled as a missing part, the depth for that region is filled in using a linear Coons surface model [20]. In this technique, four edge segments and four corner points on opposing sides of the region are chosen (currently randomly) as representative of the region boundary. A spline curve is fit to each chosen edge segment. Depth points internal to the region are generated using the linear blending function for Coons surfaces. This method works reasonably well for surfaces which have no radical changes in
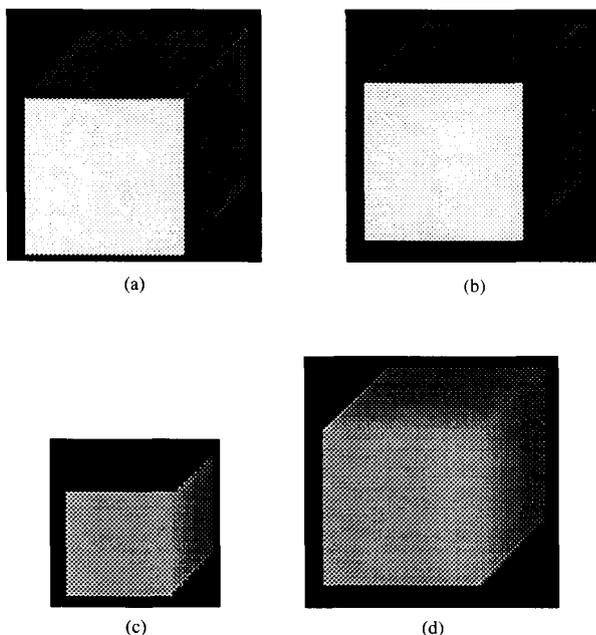
Fig. 6. Missing part extraction example. This example shows a synthetic image containing a single cube. The range image in (c) shows that the top surface of the cube is a missing part. The synthetic intensity images in (a) and (b) are the top and bottom images, respectively, for this cube. The final range map, shown in (d), shows the result after surface interpolation.

shape, since the accuracy of the resulting surface is dependent on the accuracy of the B-splines fit to the boundary segments, and which have only minor disparity errors, since if large disparity errors are present, the derived surface will not conform to the desired surface and thus does not match surrounding surface regions (i.e., it may appear to have been peeled away from or indented into a surrounding region). The selection of the edge segments and corner points is also crucial to the resulting surface; poor choices will result in inaccurate surfaces.

The missing part algorithm has been tested on a synthetic image of a cube with one missing part. Fig. 6 shows the range image, the corresponding intensity images, and the resulting depth map after stereo matching, boundary improvement, and missing part extraction.

III. IMPLEMENTATION AND EXPERIMENTAL RESULTS

Test data were created from registered range and intensity images obtained from the Pattern Recognition and Image Processing Lab at Michigan State University. Coarse range images were generated using a tailor-made program, which uses the same method of image reduction as in pyramid construction. These coarse images were then used to test the range-guided stereo matching. The second image in the stereo pair were created from the original intensity image and the original range image using the "stereo" program from the University of British Columbia's IFF (Image File Format) Library. The original coarse range data and stereo-intensity data used to generate the results shown in the remainder of this section are shown in Fig. 7.

A. Pyramid Construction and Edge Linking for Stereo Images

Pyramid construction is performed using a noniterative, nonoverlapped scheme, where four gray-level values reduce to one

value in the next higher pyramid level. Gray-level reduction is performed using the best image-preserving technique described by Li and Hu [16], where the four values are sorted, and the average of the middle two is assigned as the gray-level value to the parent of the four elements. Once the gray-level pyramid is established, edge detection is performed on all pyramid levels independently, using the Sobel operator. Edges are thinned using nonmaximum suppression, where two neighbors along the direction of the edge gradient are chosen for each edge point. If the point's edge magnitude is less than either of these two neighbors having similar gradient direction, then it is removed.

Intralevel edge linking is used to form edge segments at each pyramid level. Each edge point records its primary and secondary predecessors and successors on the segment, thus allowing forks [19]. For each edge point in the image, six neighbors from the surrounding eight are chosen, based on its edge gradient, three as candidates for the predecessor and three for the successor, as in Fig. 8. For both candidate predecessors and successors, neighbors are assumed compatible if the difference between the direction of their edge gradient and the center point's is within 90°. If there is only one compatible candidate, then it is listed as the primary predecessor or successor (as appropriate). If there are two or more compatible candidates, the candidate whose edge gradient is closest to the center point's is kept as the primary predecessor or successor, and the candidate with the next closest edge gradient is kept as the secondary predecessor or successor.

Interlevel edge linking is based on the Pyramid Linking concept where an overlapped pyramid is employed [4], [11]. In this approach, each edge point has up to four parents and sixteen children. Interlevel linking begins at the top of the pyramid. Each edge checks its sixteen potential children for compatible edges, where an edge is compatible if the difference between their edge gradients is within 45°. For efficiency, bidirectional edge linking is used, where each edge element records its compatible parents and children. Afterwards, pyramid relinking is performed iteratively to ensure the best possible hierarchical linking. This relinking is made necessary by the fact that edge gradients are different between consecutive levels of resolution in the pyramid due to a decrease in edge gradient accuracy, where the edge gradients may be sufficiently different to create a situation where edges which should link to each other do not.

An example intensity image pyramid for the toy image is shown in Fig. 10. A binary edge map for the left image pyramid for this image is shown in Fig. 11. Since the right image pyramids are very similar, they are not shown.

B. Range Image Edge Detection

In order to perform range image edge detection, a traditional, yet novel approach, is used to reduce the number of passes over the image. In the first pass, jump edges are detected using the Sobel intensity edge detection operator, and the first-order surface derivates f_x and f_y are calculated from the Sobel operator and used to compute the surface normal for each point. In the second pass, roof-edge detection is performed by comparing the angular differences between the surface normal of a single point with the surface normals of its eight surrounding points [24]:

$$M_{\text{roof}} = \max \{\cos^{-1} (\vec{n}(x, y)$$
$$\cdot \vec{n} (x + k, y + l)), -1 \le k, l \le 1\}$$

In the third and final pass, jump and roof edges are thinned using nonmaximum suppression with respect to two neighbors chosen for each edge point based on its edge gradient.

(a)   (b)   (c)

(d)   (e)   (f)
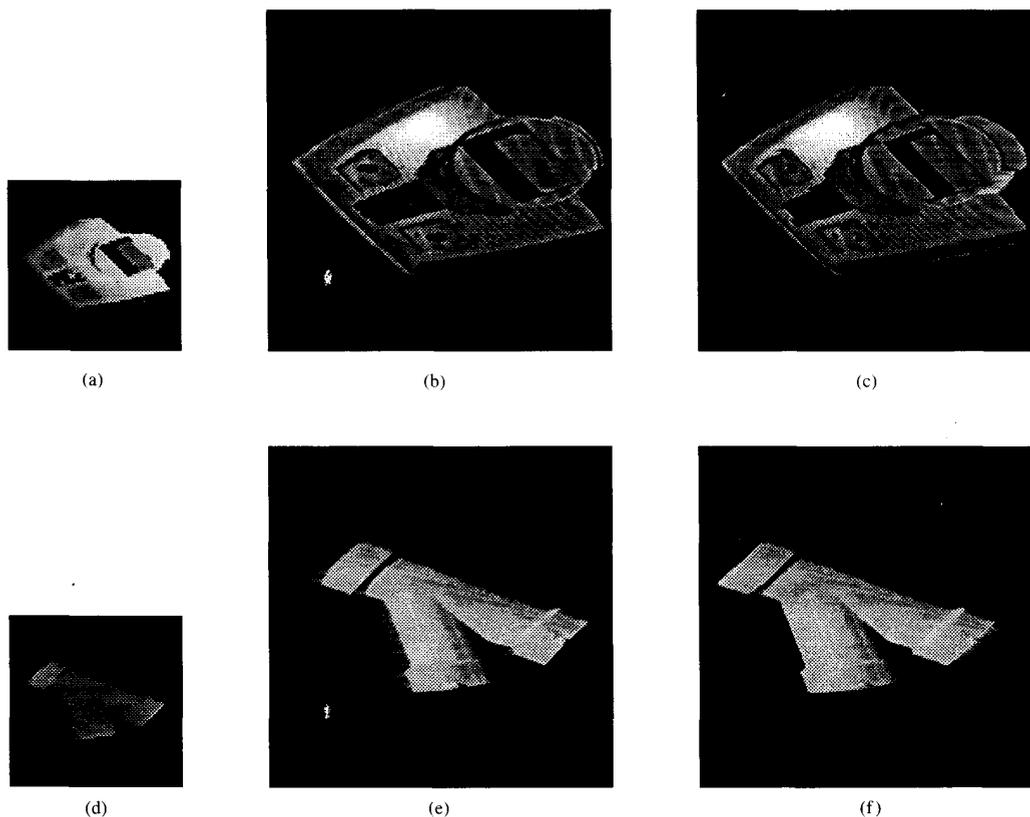
Fig. 7. Original data. (a) The laser range depth map for the "toy" image. (b) and (c) The corresponding stereo pair of intensity images. (d) The laser range depth map for the "wye1" image. (e) and (f) The corresponding stereo pair of intensity images. Both depth maps are 64 × 64 pixels; all intensity data is 256 × 256 pixels. Depth is shown as a gray-level intensity, where white indicates a region closer to the viewer.
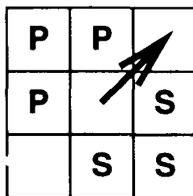
Fig. 8. Intralevel edge linking. Three potential predecessors and successors are chosen for the center point based on its edge gradient whose direction is indicated by the arrow in the figure. Forks in a segment are allowed by linking to, at most, two compatible predecessors and successors.
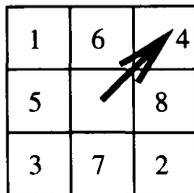
Fig. 9. Prioritized local AND. The neighbors of a point are searched in a given order determined from its edge gradient, thus maximizing the likelihood of the best possible edge equivalence.

C. Edge Combination

The goal of edge combination is to find edges in the same relative positions in the registered range and intensity image pair. The overall success of this process is vital to stereo matching, since the depth value from the equivalent edge is used to find a stereo-match disparity estimate when matching is attempted at that edge point.

The method used is based on the local AND method [8]. In this method, a 3 × 3 neighborhood in the range image is searched for a roof or jump edge, where the center of the neighborhood is in the same row and column as in the intensity map. In this implementation, jump edges are considered to correspond to an intensity edge only if the difference between their orientations lies within a threshold (currently set at 30°). Roof edges are considered to correspond to an intensity edge when they lie within the corresponding neighborhood of an intensity edge.

The use of orientation for finding common jump and intensity edges leads to a critical observation about the choice of edge detection methods for the range and intensity images. Namely, different methods place edges differently and give varying degrees of accuracy in their estimate of orientation. This observation also applies to edge thinning, since a simple difference in method between that used for range and intensity images can reduce the number of correspondences. For all these reasons, the Sobel operator is used for edge detection and the same method for edge thinning is used on all edge types (roof, jump, and intensity).

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

(e)   (f)   (g)   (h)
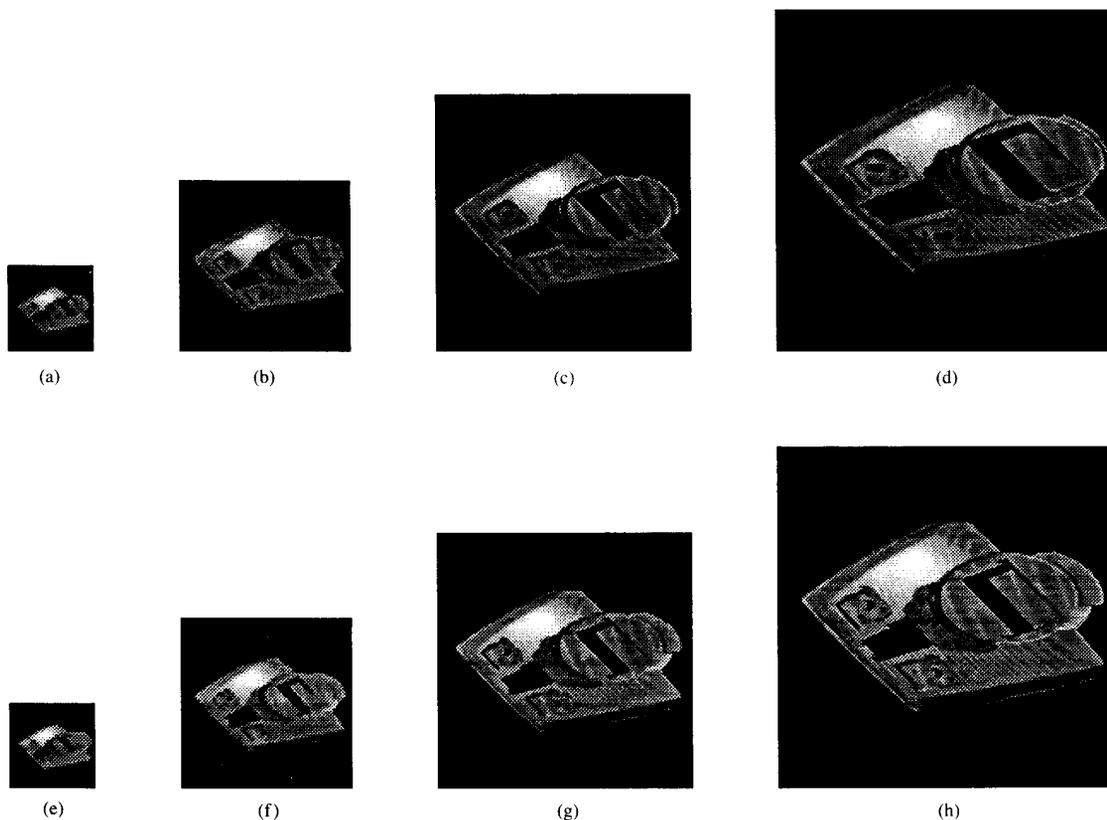
Fig. 10. Intensity pyramids for the "toy" image. (a) to (d) The left image pyramid. (e) to (h) The right image pyramid. Image resolution at the top of the pyramid is 32 × 32 pixels, and at the bottom 256 × 256 pixels. The images at the bottom of the pyramid correspond to the original intensity images.

(a)   (b)   (c)   (d)
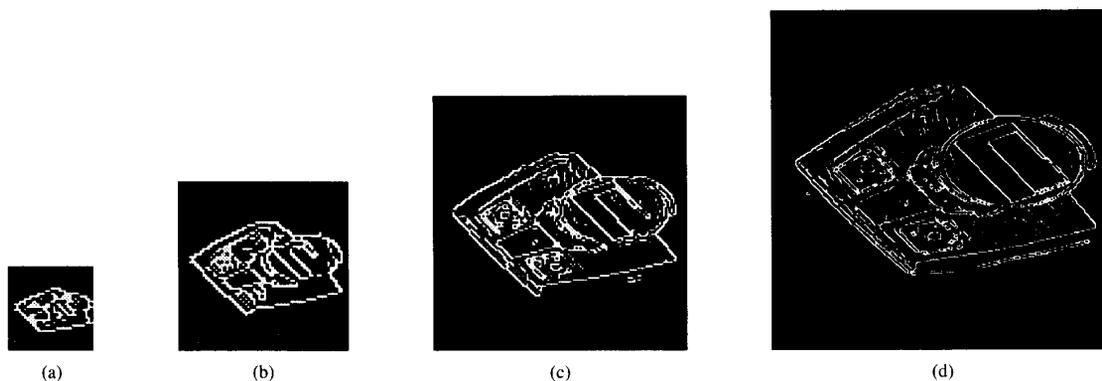
Fig. 11. Intensity pyramid edges. (a) to (d) The left image edge pyramid for the "toy" image. Image resolution at the top of the pyramid is 32 × 32 pixels, and at the bottom 256 × 256 pixels.

One problem with the original local AND method [8] for this implementation, is that no order of search is specified for the neighborhood of a point. This is a problem when more than one possible correspondence lies within the neighborhood. For this implementation, where the best possible correspondence (and thus depth estimate) must be found, it is necessary to use a *prioritized* local AND. The order of search in the neighborhood is determined by the section of the intensity edge orientation, as shown in Fig. 9. The first nonzero depth encountered in this ordered search is used as the initial depth estimate for that point.

Fig. 12 shows the jump and common edges for two sample images. Jump edges are shown in binary, and common edges are shown as the depth found at that point in the range image as a gray-level intensity value.

D. Stereo Matching and Depth Map Generation

To show the effectiveness of matching by depth, the disparity maps between the first pass (matching by depth) and the second pass (matching remaining edges) were written out. The results of
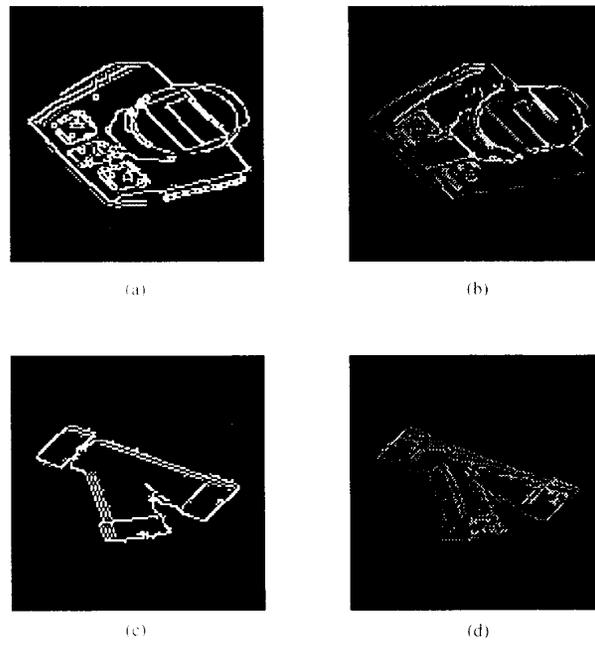
Fig. 12. Range and combined image edges. (a) Jump edges, and (b) common edges for the "toy" image. (c) Jump edges, and (d) common edges for the "wyet" image. All images are 128 × 128 pixels.
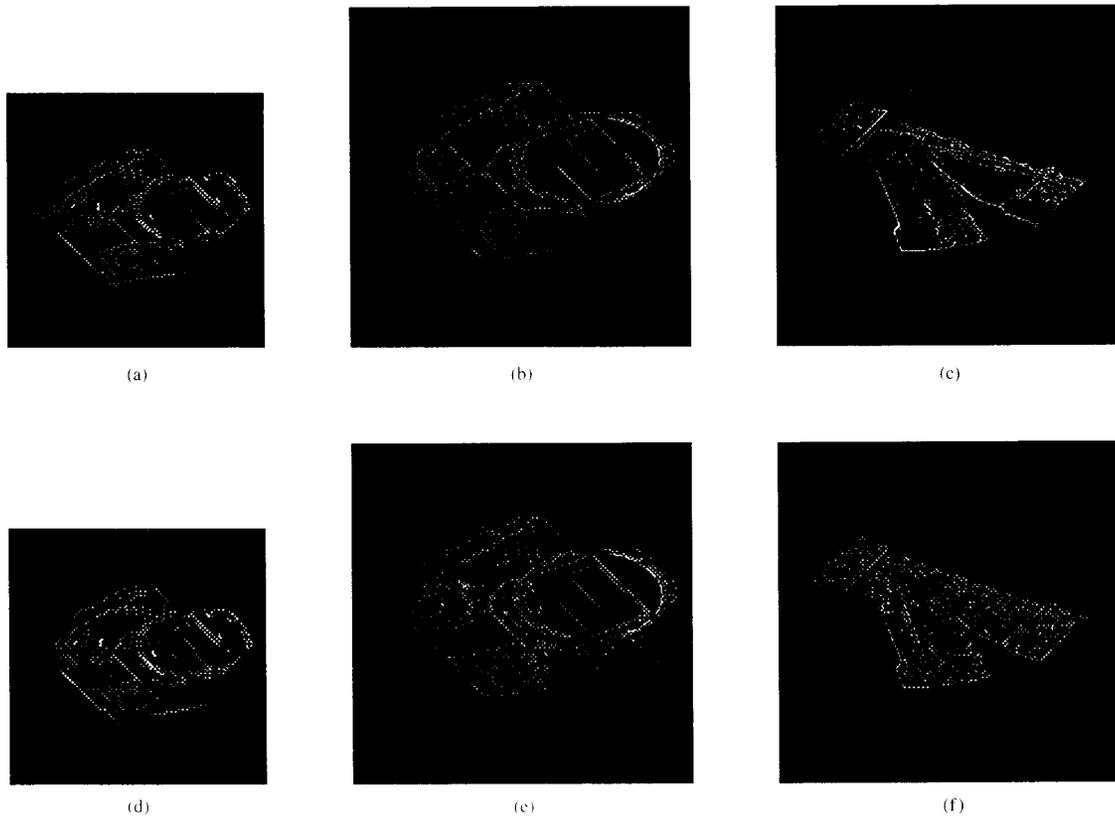
Fig. 13. Stereo-matching results. (a), (b), and (c) The disparities after matching by depth. (d), (e), and (f) The complete disparity maps after all edges have been matched. (a) and (d) are 128 × 128 pixels; all others are 256 × 256 pixels.

TABLE I
STEREO-MATCHING RESULTS

Image	Pyramid Level	Total Number of Edges	Number Matched by Depth (%)	Number Matched Without Depth (%)	Total Matched (%)	Processing Time (s)
toy	7:	1137	983 (91.78)	88 (8.22)	1071 (94.2)	
	8:	3018	2230 (78.47)	612 (21.53)	2842 (94.17)	
	totals:	4155	3213 (82.11)	700 (17.89)	3913 (94.18)	12.02
wye1	7:	986	806 (90.7)	81 (9.13)	887 (89.96)	
	8:	4096	2392 (65.48)	1261 (34.52)	3653 (89.18)	
	totals:	5082	3198 (70.44)	1342 (29.56)	4540 (89.33)	13.01

(a) (b) (c)
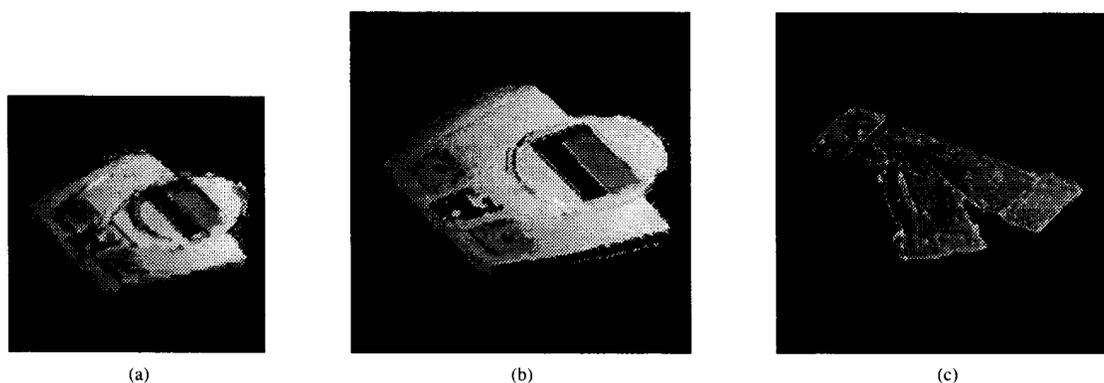
Fig. 14. Final depth maps. (a) 128 × 128 hierarchically constructed depth map for the "toy" image. (b) Final 256 × 256 depth map. (c) Final 256 × 256 depth map for the "wye1" image.

stereo matching are shown in Fig. 13. Related results are shown in tabular form in Table I. The table illustrates the results of hierarchical matching, showing the total number of edges and number of edges matched with and without depth (i.e., on the second matching pass) for each pyramid level. Also shown is the total processing time measured on a SPARCstation 1. Note that the percentage of edges matched by depth on subsequent pyramid levels, is lower than that matched at upper levels, which is a negative effect of performing averaging to obtain depth values.

The final depth maps resulting from hierarchical boundary improvement are shown in Fig. 14.

IV. FUTURE RESEARCH ISSUES

The viability of the proposed approach as an alternative for the rapid acquisition of high-resolution range data is an issue requiring further exploration. Clearly, fast, and perhaps custom hardware is required. Parallel implementations, in particular, seem to provide great promise, especially given the efforts towards parallelizing pyramid implementations [23].

The largest issue open to further research is that of deriving more complete surface descriptions, especially for missing parts. Clearly, some method not totally dependent on the range data is required to deal with surfaces and finer details on existing surfaces introduced by the stereo-intensity data. Linear Coons surfaces produce acceptable results for planar or simple curved surfaces. However, they require good choices for bounding edge segments and corner points and can produce results inconsistent with the desired surface. It might also be desirable to use a multiresolution approach to surface interpolation, where a surface fit is progressively refined to the bottom of the pyramid instead of using a single resolution similar to Terzopoulos' method for stereo-intensity images [21].

V. CONCLUSIONS

This paper presented a unique approach to the combined processing of laser range data and stereo-intensity data. Stereo matching serves as a complementary source of depth information, attempting to deal with the noise inherent in laser range data, especially along the edges of an object. Laser range data is used to guide a stereo-matching process where known depth values from the laser range map are converted to estimated disparity values to reduce the search space for candidate matches, thus reducing the correspondence problem. The hierarchical nature of this approach allows the creation of a high-resolution depth map from a low-resolution range image and a stereo pair of high-resolution intensity images.

Arguably the most important feature of this research is the multiresolution approach to stereo matching and depth map construction. This paper has presented a method where edge relationships between different levels of resolution of the edge map are explicitly represented using interlevel edge linking in an image pyramid data structure. Furthermore, the image pyramid is used to provide intralevel edge linking to form edge segments. These edge segments form the basis of stereo matching and are not restricted to being a straight line or any particular orientation. Edge segments also provide an effective means of enforcing the traditional figural continuity constraint. The interlevel edge linking along edge segments allows the enforcement of the interlevel figural continuity constraint. This constraint, not present in other approaches, provides a powerful method of enforcing surface similarity between levels of resolution of the image, thus reducing the chances of an incorrect match made at low resolution propagating to the bottom of the pyramid.

Depth maps can be constructed for missing parts which are planar or simply curved surfaces. Missing parts are discovered in a hierarchical fashion, and interpolated using the linear Coons surface model blending function at the bottom of the pyramid.

This paper has also explored the range and intensity edge correspondence issue further than previous researchers. In particular, it was found that the most reliable edge correspondences are achieved when the same edge detection and thinning method is used for all edges. Also, a more general-purpose method of finding edge correspondences has been presented which records the actual region of equivalence.

ACKNOWLEDGMENT

REFERENCES

[1] N. Alvertos, B. Dragana, and R. Gonzalez, "Camera geometries for image matching in 3-D machine vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 9, pp. 897-915, 1989.

[2] N. Ayache, *Artificial Vision of Mobile Robots: Stereo Vision and Multisensory Perception*. Cambridge, MA: MIT Press, 1991.

[3] P. Besl, "Active optical range imaging," in *Advances in Machine Vision*, J. L. C. Sanz, Ed. Berlin: Springer-Verlag, 1989, pp. 1-63.

[4] P. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. Berlin: Springer-Verlag, 1984, pp. 6-35.

[5] W. Chen, W. Lie, and Y. Chen, "Segmented description of 3-D objects with range and intensity data," in *Int. Conf. Image Processing*, 1989, pp. 212-216.

[6] C. C. Chu and J. K. Aggarwal, "Image interpretation using multiple sensing modalities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 8, pp. 840-847, 1992.

[7] S. Das and N. Ahuja, "Integrating multiresolution image acquisition and coarse-to-fine surface reconstruction from stereo," in *Proc. IEEE Workshop Interpretation 3D Scenes*, 1989, pp. 9-15.

[8] B. Gil, A. Mitchie, and J. Aggarwal, "Experiments in combining intensity and range edge maps," *Comput. Vision Graphics Image Processing*, vol. 21, no. 3, pp. 395-411, 1983.

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

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

[11] T. Hong, M. Shneier, and A. Rosenfeld, "Border extraction using linked edge pyramids," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-12, no. 5, pp. 660-668, 1982.

[12] T. Hong and M. Shneier, "Extracting compact objects using linked pyramids," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 2, pp. 229-237, 1984.

[13] R. C. Jain and A. K. Jain, Ed. *Analysis and Interpretation of Range Images*. Berlin: Springer-Verlag, 1990.

[14] R. Jarvis, "A perspective on range finding techniques for computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, no. 2, pp. 122-139, 1983.

[15] E. P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*. Berlin: Springer-Verlag, 1989.

[16] Z. N. Li and G. Hu, "On edge preservation in multiresolution images," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 6, pp. 461-472, 1992.

[17] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Royal Soc. London (Series B)*, vol. 204, pp. 301-328, 1979.

[18] J. Mayhew and J. Frisby, "Psychophysical and computational studies towards a theory of human stereopsis," *Arti. Intell.*, vol. 17, pp. 349-385, 1981.

[19] R. Nevatia and K. Babu, "Linear feature extraction and description," *Comput. Graphics Image Processing*, vol. 23, no. 1, pp. 67-91, 1980.

[20] D. Rogers and J. Adams, *Mathematical Elements for Computer Graphics*, 2nd Ed. New York: McGraw-Hill, 1990.

[21] D. Terzopoulos, "Multilevel reconstruction of visual surfaces: Variational principles and finite-element representations," in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. Berlin: Springer-Verlag, 1984, pp. 237-310.

[22] C. Thorpe, M. H. Hebert, T. Kanade, and S. A. Shafer, "Vision and navigation for the Carnegie-Mellon Navlab. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 3, pp. 362-373, 1988.

[23] L. Uhr, "Parallel, hierarchical software/hardware pyramid architectures," in *Pyramid Systems for Computer Vision*, V. Cantoni and S. Levialdi, Eds. Berlin: Springer-Verlag, 1986, pp. 1-20.

[24] N. Yokoya and M. Levine, "Range image segmentation based on differential geometry," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 6, pp. 643-649, 1989.

Stereo Correspondence Based on Line Matching in Hough Space Using Dynamic Programming

Ze-Nian Li

Abstract—This paper presents a method of using Hough space for solving the correspondence problem in stereo vision. It is shown that the line-matching problem in image space can readily be converted into a point-matching problem in Hough (ρ-θ) space. Dynamic programming can be used for searching the optimal matching, now in Hough space. The combination of multiple constraints, especially the natural embedding of the constraint of figural continuity, ensures the accuracy of the matching. The time complexity for searching in dynamic programming is $O(pmn)$, where m and n are the numbers of the lines for each θ in the pair of stereo images, respectively, and p is the number of all possible line orientations. Since m and n are usually fairly small, the matching process is very efficient. Experimental results from both binocular and trinocular matchings are presented and analyzed.

I. INTRODUCTION

The recovery of depth information is important for 3-D image analysis. One method for depth recovery is *stereo vision*, in which pairs of images from horizontally and/or vertically displaced cameras are used. One of the most difficult problems in stereo vision is *correspondence* [1]. Once corresponding points in the pair of images are identified, their *disparity* values can be calculated and used to recover the depth.

We are developing a vision system for mobile robots. The proposed domain is an office environment. The robot is planned to walk in corridors and rooms to fetch and deliver simple objects. Both stereo and laser range data will be used. Since most of the objects of interest will be man-made, it is natural to think of straight lines as the main feature for stereo matching. The stereo algorithm described in this paper, can quickly render a nondense depth map

Z. N. Li is with the School of Computing Science, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada.
IEEE Log Number 9212949.