A Hierarchical Approach to Stereo Vision

Sabry F. El-Hakim
National Research Council of Canada, Robotics and Automation Section, 435 Ellice Avenue, Winnipeg, Manitoba R3B 1Y6, Canada

ABSTRACT: A stereo vision system for image metrology has been developed and targeted for applications in computerized manufacturing, quality control and robot guidance. Description of the approach, evaluation of the performance and the limiting factors of applying this technique are reported. The approach is designed to minimize the well known difficulties of digital and real-time photogrammetry, or stereo vision. The evaluation is based on the degree of the success of the feature extraction process and the matching technique as well as the achieved accuracy of the three-dimensional coordinate measurements.

INTRODUCTION

THREE-DIMENSIONAL INFORMATION is important for vision-based computerized manufacturing, quality control, and robot (or manipulator) guidance. To be effective in these applications, the system must be reliable, noise tolerant, cost effective, simple to operate, accurate (in consistency with application), fast (in consistency with application), and adaptable to various tasks (within the application). No one existing system satisfies all the requirements for all applications under all working and environmental conditions. Therefore, selection of the most suitable technology, or combination of technologies, depends entirely on the application at hand.

There are several competing technologies for determining three-dimensional (3-D) coordinates. One of them uses a line scanning laser and a light sensor, and the depth is computed by triangulation (Rioux, 1984). This technique, also called active 3-D vision is simple (for example, the matching of corresponding points is not needed), efficient, and does not require adding features (targets) to facilitate measurement of featureless objects. It also has the advantage of working under any lighting condition and has a larger depth of field than optical systems. However, because scanning is required, the technique is not suitable for large moving objects or when natural features or specific object points are to be identified and measured. Also, for obvious security reasons, it cannot be used for scene analysis for military purposes. Furthermore, the performance might require, for some applications, lasers with power levels unsafe for humans to be around.

Another approach, which is the most common in machine vision applications, uses one camera and a structured light source (Shirai and Suwa, 1971; Agin and Binford, 1973; El-Hakim, 1985; Murai *et al.*, 1986). Again, this approach avoids the matching problem, adds features to featureless surfaces, and usually is simple and efficient to use. It, however, suffers from a limited depth of field and most of the other limitations of the above mentioned approach.

Techniques which are based on two cameras and stereo disparity (mainly photogrammetry) have the advantage of being less restricted to the application as the above methods are, capturing the data instantaneously, which makes application to moving objects feasible, and have no mechanical components, which means less cost and better accuracy in the long run. However, these techniques have not usually been considered for most of the industrial applications mentioned above. The main reason has been, until recently, the requirement of photography, costly equipment, and highly trained operators. The time lag between data acquisition and results was also a limiting factor. Even now, when the all-digital fast photogrammetric techniques (also called "real-time" photogrammetric systems) are becoming available (e.g., El-Hakim, 1986; Haggren, 1986), many developments are required before these systems are widely accepted in industry. The main limiting factors are

- Illumination (because they rely on ambient light) and the shadowing effect (with large *b/h*, needed for high accuracy);
- Matching of corresponding points;
- Feature extraction, particularly in poor illumination and contrast conditions;
- Limited depth of field, a problem shared by all optical systems;
- Complexity of computations (for high speed applications). These limitations are unique to this technique, and there are other problems which are shared by all the above approaches;
- Camera metric quality (sensors, A/D, and lens); and
- Edge detection and target pointing to sub-pixel accuracy (although there has been a substantial effort, the problem is largely unsolved (Nalwa and Bindord, 1986)).

This paper describes an approach to stereo vision which is designed to minimize the effect of some of these limiting factors. A test procedure to evaluate the accuracy as well as results from various experiments are presented.

CHARACTERISTICS OF THE APPROACH

The approach presented here has been originally outlined in an earlier paper (El-Hakim, 1986). However, since then, and after several applications, the technique has been updated to improve on its performance. Figure 1 outlines the processing steps. The characteristics of this approach are

- It is *fully automated,* and no human intervention is expected except in the initial set up and calibration of the cameras and determining their orientation parameters;
- It is carried out in *separate steps.* The features are extracted first, then measured for each image, then the matching is performed on the image coordinates. The main reason for the stepwise solution is that once we reach the matching process, there will only be the desired features, and their locations, to deal with and thus much less chance of error;
- It is *hierarchical,* or a level by level (coarse to fine) within each of the separate steps mentioned above. Again, this is designed to reduce the chance of identification and matching errors to a minimum; and
- It is *flexible.* The user can elect to stop after any level of processing in any of the steps. For some applications where the speed is critical while the required accuracy is not high, it may be sufficient to stop after the first level of each step.

In the following sections, the method for feature recognition, point location, and matching of corresponding points are described in some detail.

THE METHOD FOR FEATURE RECOGNITION

It is well known that it is much easier to use binary images for feature recognition. However, changing the image into adequate binary is not a simple task and, even when this is ac-

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 55, No. 4, April 1989, pp. 443–448.

0099-1112/89/5504–443$02.25/0
©1989 American Society for Photogrammetry
and Remote Sensing

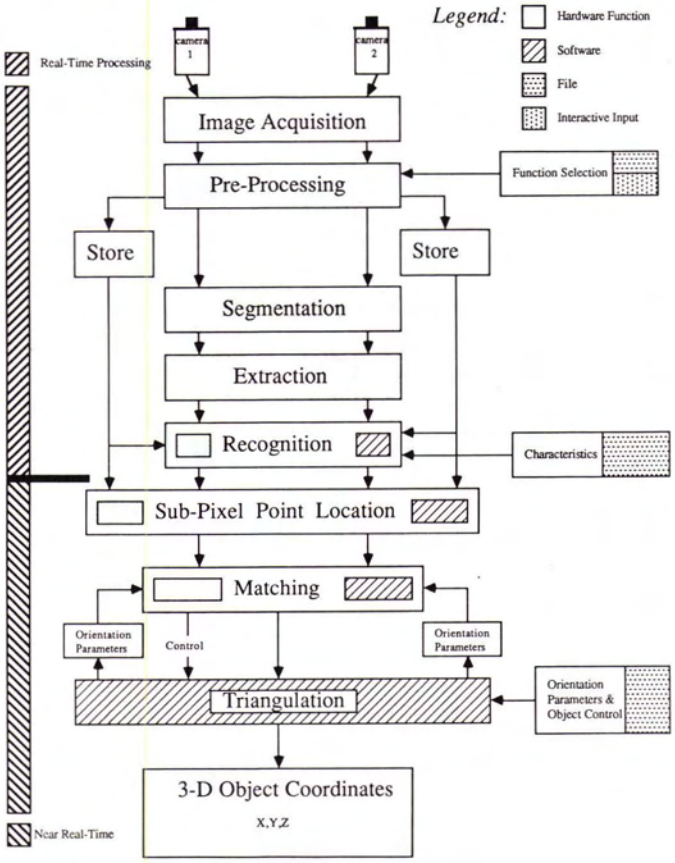


FIG. 1 Block diagram of the stereo measuring steps.

complished, the recognition process has many limitations because it can only work with the binary geometric shape of the feature. On the other hand, working with a full grey scale image and applying a template matching technique to find the desired feature can be time consuming and often unsuccessful if applied to the image as a whole. The following method utilizes the advantage of both binary and grey-scale image and minimizes their inadequacies.

LEVEL 1: FEATURE RECOGNITION FROM BINARY IMAGE

First, a sequence of image filters is applied to prepare the image for the change to binary form. These are usually a low pass followed by a high pass filter. The resulting image has fewer features and less grey-scale variation than the original image, but features like targets and edges are highly emphasized. A binary image is then easier to produce, in the form of white "blobs" on black background, or vice versa. These blobs are extracted, using the "connectivity analysis" technique, and parameters, representing the geometric shape, are computed and compared to a pre-taught set of parameters. The features fitting these parameters, with some tolerance, are labeled and their locations are stored for use in Level 2.

LEVEL 2: RECOGNITION USING FULL GREY-SCALE IMAGE

The locations of the recognized features from level 1 are entered into the original image and a template matching technique is applied at each position. This limits the search to the vicinity of the features already recognized by their geometric shape and results in highly successful recognition.

The parameters employed in the above two levels are computed automatically beforehand in a "teach" routine. This is an interactive program in which the user answers a question by simply typing "yes" or "no" to every feature extracted and displayed on the monitor. For the features to which the answer was "yes," the characteristic parameters will be computed and stored (learned) to be used by the main program during future measurement of any similar feature. This technique functions particularly well when the teaching is carried out on a wide range of target and background variations.

EVALUATION

Several tests have been carried out to evaluate the performance of the process after each level. In almost every test, the system, due to the pre-set tolerance, recognized more features than those originally taught. After Level 1, an average of 12 percent extra features are recognized, but after Level 2 this drops to less than 5 percent. In virtually every case the extra features in one image were not the same as the extra ones in the second image, which means that they will not be matched and their coordinates will not be computed.

THE METHOD FOR POINT LOCATION

This is a crucial step because the accuracy of the object coordinate measurements depends directly on the pointing technique. Again, there are two levels of processing, the first of which acts as a set-up step for the second.

LEVEL 1: DETERMINING THE CENTROID AND SIZE OF THE TARGET FROM BINARY IMAGE

It is the function of this level to find the location of the target and the boundaries of the target area. At this level there is no need for high precision because this is achieved in the next level. All that is needed here is to define the area required for the precision pointing. The binary image used earlier is utilized here. The centroid of every recognized feature and the approximate dimensions of the target are computed and passed on to Level 2.

LEVEL 2: PRECISE DETERMINATION OF TARGET (EDGE) LOCATION FROM ORIGINAL GREY-SCALE IMAGE

The procedure is applied here to circular targets as an example. The dimensions of the target (in this case the diameter of the circle) as obtained from Level 1 are slightly enlarged (typically by two or three pixels) and a window is defined around the target. A number of profiles, symmetrically positioned from the center as shown in Figure 2, are extracted and analyzed to determine the precise location of edge points. Several techniques are available for the sub-pixel location of edges. The technique used here is based on least-squares adjustment, not unlike the one described by Mikhail *et al.* (1984). The ideal edge is convolved with the Gaussian image function to produce the actual edge, and the edge location is solved for.

Once all the coordinates of the edge points are determined, a circle is fitted, again using least squares, with the coordinates of its center, and its radius, as the unknowns. For convergent images, or objects largely unparallel to the camera, an ellipse is used instead. Two additional parameters, one for affine scale

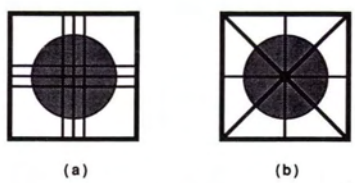


FIG. 2. Windowing and profiling for target edge measurement. (a) is for large targets and (b) is for small targets.

and one for non-perpendicularity of the two image axes (x and y), are also solved for.

Evaluation

The difference in accuracy between using the image coordinates immediately after Level 1 and using those after Level 2 has been studied extensively. The main factors affecting the difference have been the target size and the contrast between target and background. For high contrast targets of sizes larger than 100 pixels, Level 1 provided accuracies similar to Level 2. However, for target sizes less than 60 pixels, even at high contrast, Level 1 has been significantly inferior to Level 2, sometimes by a factor of 2.

THE MATCHING APPROACH

The approach assumes that the images have already been pre-processed and that the image coordinates of the points of interest have been measured in each image before matching is done. The technique is also somewhat influenced by modern theories of stereopsis (Marr and Poggio, 1979; Mayhew and Frisby, 1981) with some variations, particularly in the first and last level of matching.

Concept and Processing Steps

The hierarchical concept of this approach is outlined in Figure 3. At each level in this building process, the matching is under the control of several sets of constraints formed by geometric relationships, *a prior* knowledge about the scene and light intensity. The following are the design criteria for this matching approach:

- Within each level there are several sub-levels, or stages.
- At the early stages, only the features, or points, that follow the strictest constraints, with only very small tolerance, are matched. It should be kept in mind that an error at the early stages could have a serious effect on subsequent stages in a hierarchical solution.
- The coordinates of the successfully matched points after each stage, and level, are used to set-up the disparity constraint for subsequent stages. The expected range of disparity (or the difference between the x image coordinates from the two images) is computed from the depth (Z) of the already matched points, and this range is the disparity constraint for points within the corresponding image area.
- The process can be terminated after any stage, as dictated by the application. For example, if only targets are required, only Level 1 is selected.

To describe each level in detail along with all the constraints and filters involved is beyond the scope of this paper. Only few comments are given below.

- For Level 1, where a set of image coordinates of features (usually targets) have already been measured in each image, the epipolar line constraint is used in the first stage as described in El-Hakim (1986). In the subsequent stages, the disparity constraint is added as points are matched.
- For Levels 2 to n-1, the image coordinates to be measured are those of edge points. The edges are extracted with the zero-crossing operator (Marr and Oggio, 1979) shown in Figure 4. The width W_{2D} determines how many details are to be extracted. The smaller the width, the more the details. Therefore, at Level 2 the width is selected to produce the least number of edges, usually only the outer edges of the object, and the details are increased in the subsequent levels by decreasing W_{2D}. Again, for each level the z-coordinates computed for the edge points in the previous levels are used to determine the disparity constraints. The matching of edges is carried out by the intra- and inter-line approach (Lloyd, 1986; Ohta and Kanade, 1985). In this approach, the search for matching points is carried out along each pair of corresponding epipolar lines (intra-scanline search) and also along edges spanning several epipolar lines (inter-scanline search). The reason for this later search is to try to find consistency among scanlines by analyzing coordinates of edge points already matched by the epipolar line search. For example, if two edge points do not match in one pair of epipolar lines, their vertically connected edge points (within one or two pixels) should not match either. There are two different sets of constraints; each controls one of these two types of search. The method is well suited for dynamic programming and parallel processing, particularly for intra-scanline search.
- The last level [n] is reserved for least-squares correlation (Gruen, 1985) using the original or a noise reduced image. All the object coordinates of the successfully matched points from the previous levels are used here as constraints to improve the efficiency of the technique.

Evaluation

The evaluation is based on the degree of the success of the match after each level and each stage within the level. At Level

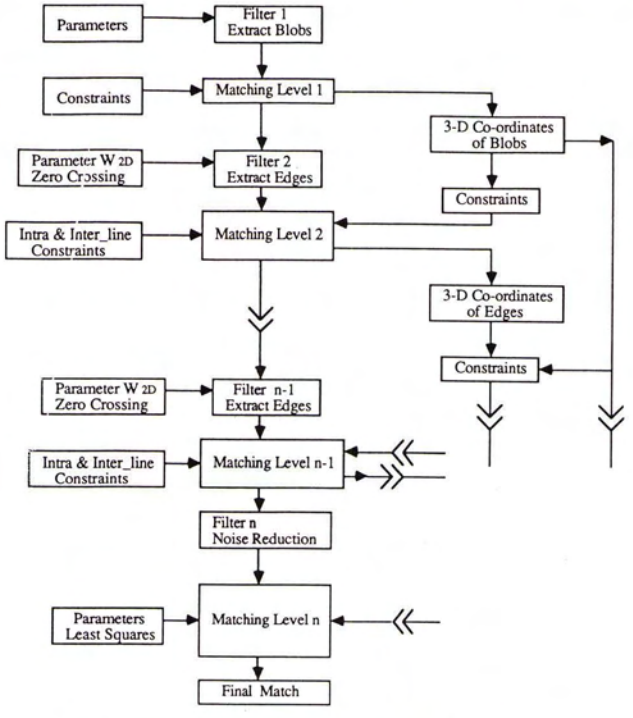


FIG. 3. Block diagram of the matching process.

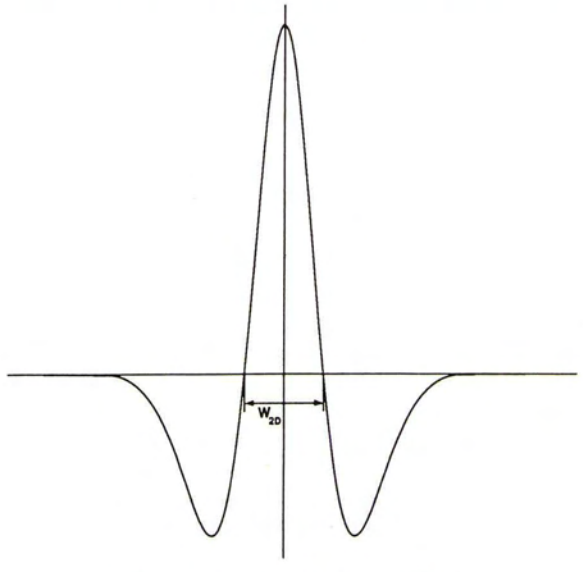


FIG. 4. The zero-crossing operator.

1, using only the epipolar line constraint results in successfully matching only about 80 to 85 percent of the well defined targets selected in this level. By adding the disparity as a constraint in subsequent stages, all these targets are matched. At Levels 2 to n-1 the success is directly proportional to the number of levels. By trying to match all edges in one step (by having a very small W_{2D} to begin with), the success rate has been in the 60 to 65 percent range for objects with several inside edges. This has improved to about 90 percent when three levels were used. The final Level n has also been tested once without any constraint and once as described above. The success rate could be as low as 60 percent without constraints, while going through all the levels raised the rate to over 95 percent.

A PROCEDURE FOR ACCURACY EVALUATION

This section is a study based on hundreds of measurements of targets of various sizes in a variety of arrangements. It started out with the point of view that any strategy for accuracy evaluation of a fully automated measuring system, using live images, must take into account the repeatability of the results, in the presence of whatever noise, as well as the absolute accuracy. Unlike systems carrying out measurements on photographs, where repeatability is excellent, digital images are affected by some factors making it difficult to achieve repeatable and very highly accurate measurements. Assuming correct point identification and matching, the final accuracy is affected by

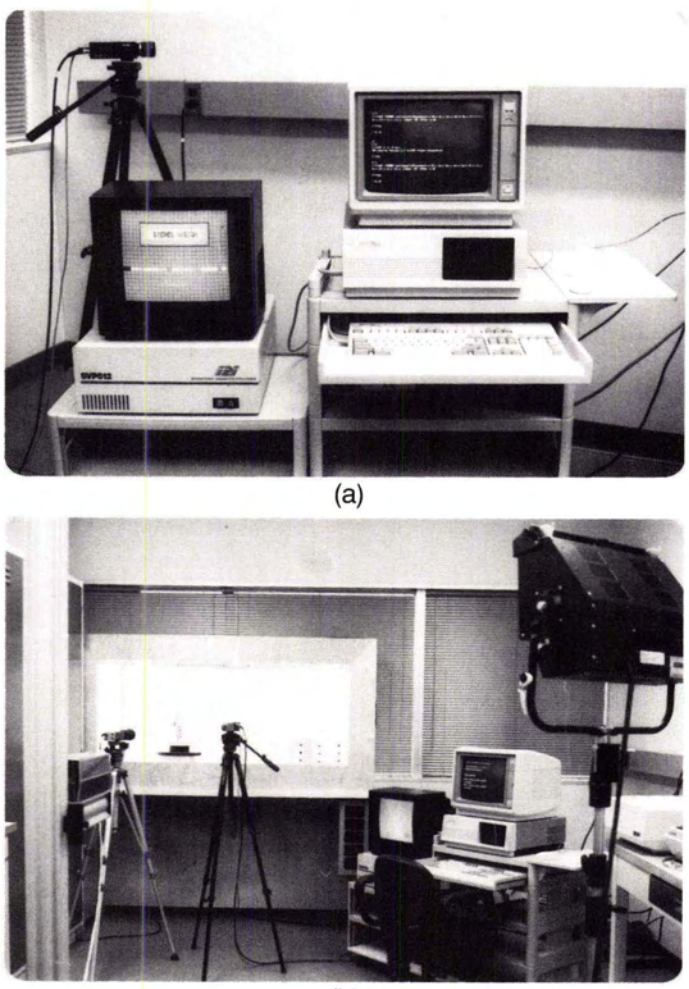
- Sensor resolution,
- Sensor geometric errors (especially the horizontal-line jitter and the warm up effect),
- Pointing accuracy (all existing methods have to make one or more approximate assumptions),
- Camera orientation (geometric strength), and
- Quality of the calibration.

The first three sources are the most responsible for the uncertainty of the computed coordinates and could produce errors, especially in the presence of noise, which are unpredictable and, so far, not possible to compensate for mathematically. In particular, the camera/digitizer synchronization produces horizontal-line jitter that accounts for noticeable error in this direction (Beyer, 1987; Dähler, 1987; Luchmann and Wester-Ebbinghaus, 1987) and affects the depth (Z) as well. This problem can be solved by a specially designed synchronization circuit (Havelock, personal communication; 1988). It is therefore essential to study the repeatability of the system and accept all the fluctuation in the results as part of the system accuracy limitations. In the approach outlined here, all data are used and none is rejected as an outlier or a blunder.

System Components

Vision systems consist of three major components: the computer architecture, the sensors, and the lighting. The results shown below have been obtained with the following system components (Figure 5):

- The computer architecture consists of three boards. The CPU, which is the 32-bit Motorola 68020, the frame buffer board with eight 480- by 512- by 8-bit frames, and a dedicated processor for operations on images. The three boards communicate with an IBM-PC-AT via a 64K gateway board. A C-language cross-compiler is imployed with MS-DOS operating system.
- Two high resolution frame transfer CCD cameras (the Pulnix TM-840 with an 490(v) × 800(H) pixel imager) are the system sensors.
- The light sources are one front soft-light floodlight with dimmer switch and one background light unit consisting of 15 independently controlled incandescent light sources covered with glass diffuser.



(a)

(b)

FIG. 5. System components [(b) shows the complete system set up including front and background lighting].

FIG. 6. XYZ-axes positioning system.

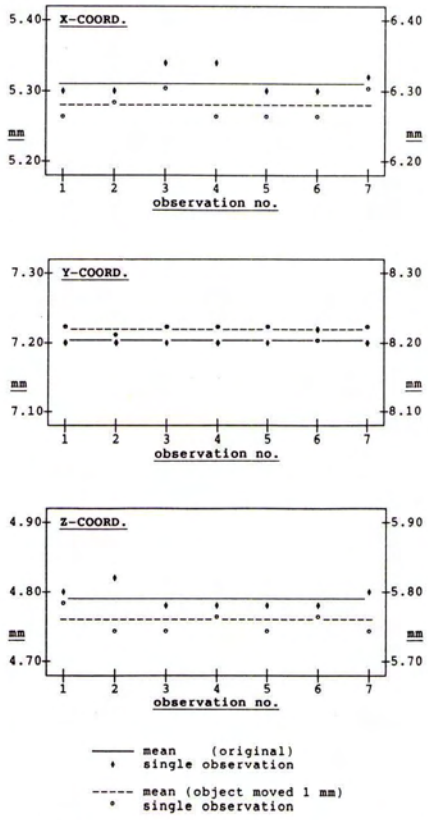


FIG. 7. Repeatability and absolute accuracy diagrams for one point. [The vertical scale on the left is for the original position and the one on the right is for after moving.]

ACCURACY TEST DEVICE

The three-axes positioning device shown in Figure 6 is used for the accuracy evaluation. The three micrometers shown are placed in *X*, *Y*, and *Z* directions parallel to the coordinate system of the control points used to determine the camera absolute orientation. Various targets are placed on the front vertical surface of the device and are measured by the system at an initial position. The micrometers are then moved by a certain amount in each direction and the same targets are remeasured. The difference between the coordinates in the two positions is compared to the actual movement and the result is an indication of the absolution accuracy.

CALIBRATION

The cameras are calibrated and their exterior orientation parameters are computed *a priori* to the actual measurements with a set of 15 control points located at three different vertical surfaces. The mathematical model used of systematic errors compensation had been developed in an earlier study (El-Hakim, 1986).

TABLE 1: ABSOLUTE ACCURACY AND STANDARD DEVIATION [MM]

Max. Object Size [cm]	Approx. Target Size [px]	Absolute Accuracy			Stand. Deviation		
		X	*Y*	*Z*	*X*	*Y*	*Z*
30 by 23	160	0.020	0.013	0.030	0.006	0.005	0.011
	300	0.017	0.011	0.029	0.005	0.005	0.010
40 by 30	65	0.037	0.037	0.050	0.007	0.005	0.012
	120	0.025	0.017	0.034	0.006	0.005	0.010
60 by 45	30	0.054	0.054	0.072	0.008	0.006	0.014
	60	0.033	0.029	0.053	0.007	0.006	0.012

TEST PROCEDURE AND RESULTS - TARGETS

The following are important considerations which must be observed if high accuracy is demanded:

- The cameras should be turned on several hours before the actual measurements are started to minimize warm-up effects; and
- The images should be snapped at least ten times and averaged to produce the image used for measurements. This reduces noise and, possibly, the effect of line jitter.

Table 1 displays the absolute accuracy, as indicated by the difference between the measured movement of the targets on the positioning device and the actual movement measured by the micrometers. The procedure has been repeated several times at each position for each target and the standard deviation of these are shown in the table. This has been carried out at different image scales (object sizes) and for different sizes of targets.

Figure 7 illustrates the repeatability of measurements on one point (measured seven times) for the *X*, *Y*, and *Z* coordinates as well as the mean of the measurements before and after moving the point by 1 mm in each direction. Taking the difference between the two means as a measure of the absolute accuracy gives the values 0.03 mm in *X*, 0.02 mm in *Y*, and 0.03 mm in *Z*. However, these values are uncertain by as much as 0.03 mm in *X*, 0.02 mm in *Y*, and 0.04 mm in *Z*, as shown in the figure (the maximum uncertainty is computed by taking the furthest coordinate from the mean in each case). It is therefore very important to provide these repeatability figures along with absolute accuracy estimates.

In order to plan for a specific application, it might be useful to express the accuracy in relative terms. In terms of the pixel size, the achievable accuracy ranges from 0.03 to 0.05 pixel in the *X*-*Y* plane and from 0.06 to 0.07 pixel in depth (*Z*). Using 512 by 512 resolution, this translates to ranges of 1:15,000 to 1:9000 of object size in the *X*-*Y* plane and of 1:7,500 to 1:6,500 of object size in depth.

TEST PROCEDURE AND RESULTS - EDGES

Accuracy of edge measurement, using the matching approach as described above, was evaluated using a solid wood block with well defined edges (Figure 8). The distance between the edges were measured using a precise micrometer. The same distances were then measured by the stereo vision system by measuring three dimensional coordinates of several edge points.

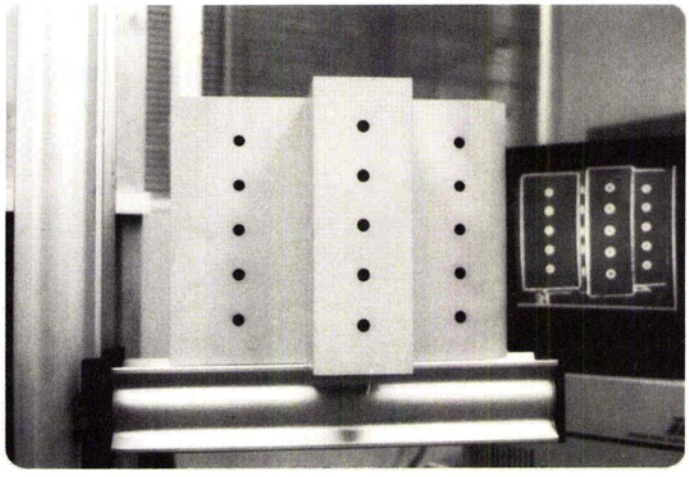


FIG. 8. Test object for edge accuracy evaluation.

The RMS of the differences between micrometer measurements and the vision system measurements was 0.04 mm on an object size of 30 cm, thus giving an accuracy of 1:7,300 of the object size. The standard deviation of ten measurements was 0.014 mm.

CONCLUDING REMARKS

The limitations of the feature recognition and the matching of corresponding points are drastically reduced by applying a hierarchical solution. Geometric and grey scale constraints and *a priori* knowledge about the object and the scene are essential to the success of these operations. This represents a limiting factor in case of dealing with unknown scenes or features. It is therefore fortunate that most industrial applications deal with known objects and predictable scenes. In some applications, such as general robot vision, however, this may not be the case and stereo vision alone will not function properly and other sensors may be needed.

the main limiting factor of achieving high accuracy, say higher than 1:20,000, is the quality of the sensors. There are several errors, not behaving in a systematic predictable manner, such as the horizontal synchronization error (x-jitter), which are difficult to model mathematically. These problems are well recognized in the vision community and the most desirable solution, which will eventually come in the near future, is to design and build better sensors and circuits.

In spite of the limitations, the results obtained from this system were very encouraging, and it is now being applied for several industrial applications such as computerized manufacturing, on-machine inspection and quality control, and even in some robot guidance applications.

REFERENCES

Agin, G., and T. O. Binford, 1973. Computer description of curved objects, *Proc. of the 3rd Int. Joint Conf. on A.I.*, pp. 629–640.

Beyer, H. A., 1987. Some aspects of the geometric calibration of CCD cameras, *Proc. of Intercomm. Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switz., pp. 68–81.

Dähler, J., 1987. Problems in digital image acquisition with CCD cameras, *Proc. of Intercomm. Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switz., pp. 48–59.

El-Hakim. S.F., 1985. A Photogrammetric Vision System for Robots, *Photogrammetric Engineering and Remote Sensing* Vol. 51(5), pp. 545–552.

———, (1986). Real-Time Image Metrology with CCD Cameras, *Photogrammetric Engineering and Remote Sensing*, Vol. 52(11), pp. 1756–1766.

Haggfen, H. (1986). Real-time photogrammetry as used for machine vision applications, *Proc. of ISPRS Comm. V Symp.*, pp. 374–382.

Gruen, A. W., (1985), Adaptive least squares correlation: a powerful image matching technique, *S. African J. of Photogrammetry, Remote Sensing and Cartography*, 14(3), pp. 175–187.

Lloyd, D., 1986. Stereo matching using intra- and inter-row dynamic programming, *Pattern Recognition Letters*, 4(4), pp. 273–277.

Luhmann, T., and W. Wester-Ebbinghous, 1987. On Geometric Calibration of Digitized Video Images of CCD Arrays, *Proc. of Intercomm. Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switz., pp. 35–47.

Marr, D.C., and T. Poggio, 1979. A theory of human stereo vision, *Proc. of the Royal Society of London*, Vol B204, pp. 301–328.

Mayhew, J. E. W., and J. P. Frisby, 1981. Psychophysical and computational studies toward a theory of human stereopsis, *Artificial Intelligence*, Vol. 17 (Special issue on Computer Vision).

Mikhail, E. M., M. L. Akey, and O. R. Mitchell, 1984. Detection and sub-pixel-location of photogrammetric targets in digital images, *Photogrammetria*, 39(3), pp. 63–83.

Murai, S., F. Otomo, and H. Ohtani, 1986. Automated three-dimensional measurements using stereo CCD camera in the application to close range photogrammetry, *Proc. of ISPRS Comm. V Symp.*, Ottawa, pp. 409–413.

Nalwa, V. S., and T. O. Binford, 1986. On Detecting Edges, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6.

Ohta, Y., and T. Kanade, 1985. Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. on Pattern Recognition and Machine Intellignece*, Vol. PAMI-7, No. 2, pp. 139–154.

Rioux, M., 1984. Laser range finder based on synchronous scanners, *Applied Optics*, Vol. 23 (21), pp. 3837–3844.

Shirai and Suwa, 1971. Recognition of polyhedrons with a rangefinder, *Proc. of the 2nd Int. Joint Conf. on A. I.*, pp. 80–87.

(Received 25 September 1988; accepted 25 October 1988; revised 8 November 1988)

FOR SALE

SANTONI G-6 — with DAT/EM Digital Mapping System — Wyse 286 12 MHZ
Hard Disk — 19" Monitor — DOS — AutoCAD —
Everything included for immediate Digital Mapping System in AutoCAD
(602) 258-6471

WANTED

WILD A-8 - late model (serial number 4000 or above)
RC-8 AERIAL CAMERA - late model (serial number above 1000)
PHOTO LAB EQUIPMENT - copy camera, vacuum frames, enlargers

Employment opportunities available for experienced and entry level stereoplotter operators.

EASTERN TOPOGRAPHICS, Route 28, Ossipee, New Hampshire 03864
(603) 539-5055