LIMITING FACTORS OF REAL-TIME IMAGE METROLOGY

Sabry F. El-Hakim

Canadian Institute for
Industrial Technology (CIIT)
NRC, 435 Ellice Ave.,Winnipeg
Manitoba, Canada R3B 1Y6

Invited Paper, Commission II

ABSTRACT

A real-time 3-D image metrology system has been developed and targeted for applications mainly 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 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 measurements.

I. 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 those uses 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, since 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 can not 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, 1979, El-Hakim, 1985, and 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 are called "real-time" photogrammetric systems) are becoming available (e.g. El-Hakim, 1986 and Haggren, 1986), still many developments are required before these systems are widely accepted in industry. The main limiting factors are;

1. Illumination (since it relies on ambient light) and the shadowing effect (with large b/h, needed for high accuracy).

2. Matching of corresponding points.

3. Feature extraction, particularly in poor illumination and contrast conditions.

4. Limited depth of field, a problem shared by all optical systems.

5. Complexity of computations (for high speed applications).

These limitations are unique to this technique, and there are other problem which are shared by all the above approaches;

6. Camera metric quality (sensors, A/D and lens).

7. Edge detection and target pointing to sub-pixel accuracy (although there has been a substantial effort, the problem is largely unsolved, Nalwa and Binford, 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.

II. 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.

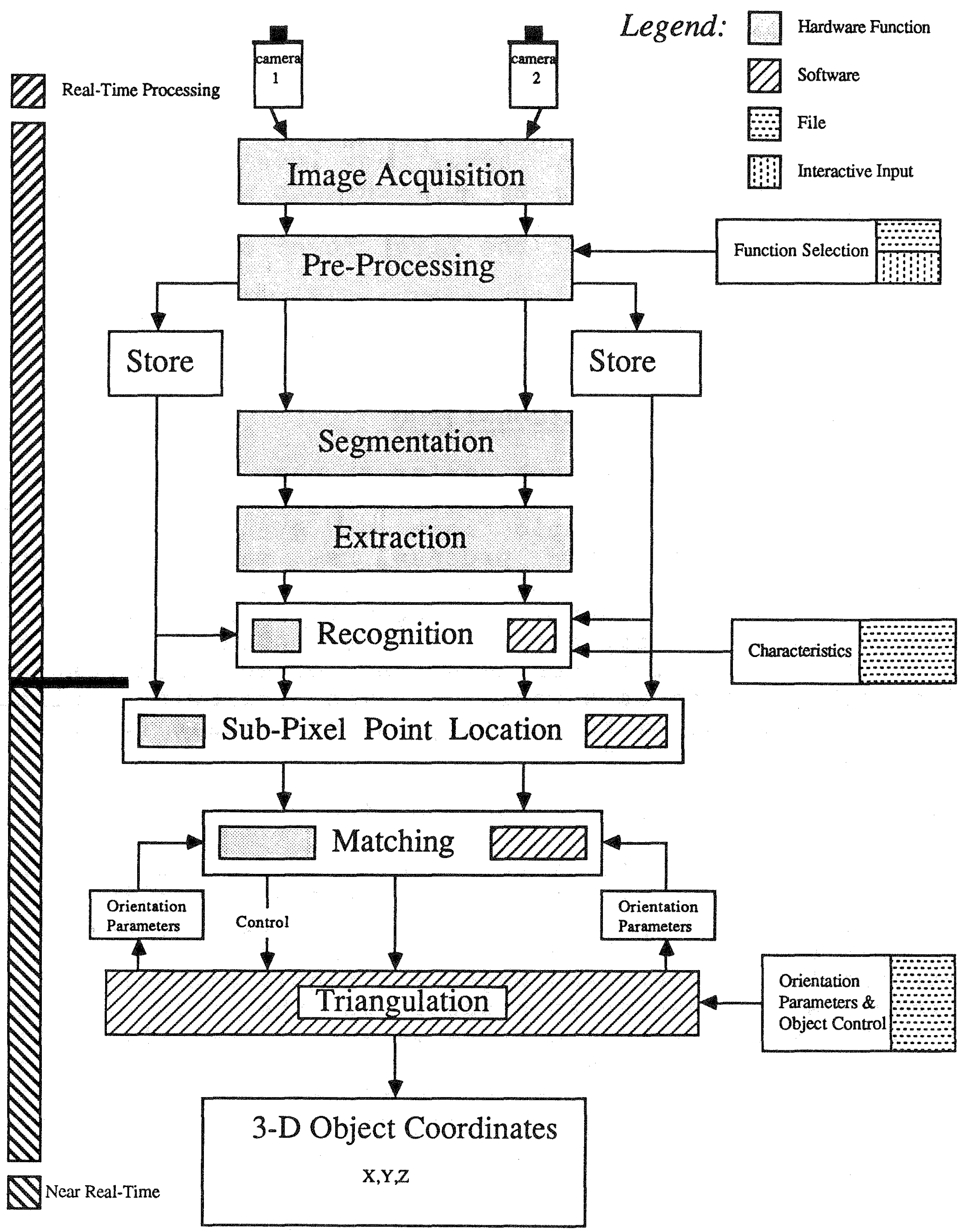


FIG. 1: Block Diagram of the Stereo Measuring System

The characteristics of this approach are;

1. 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.

2. 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.

3. 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.

4. 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 methods for feature recognition, point location and matching of corresponding points are described in some detail.

III. 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 accomplished, the recognition process has many limitations since 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.

3.1. Level 1: Feature Recognition from Binary Image:

First, a sequence of image filters are 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 less features and grey-scale variation than the original image with 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 vise 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 to these parameters, with some tolerance, are labeled and their locations are stored for use in level 2.

3.2. 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 limit 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.

3.3. 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% extra features are recognized, but after level [2] this drops to less than 5%. In virtually every case the extra features in one image were not the same extra ones in the second image, which means that these points will not be matched and there coordinated will not be computed.

IV. THE METHOD FOR POINT LOCATION

This is a crucial step since 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.

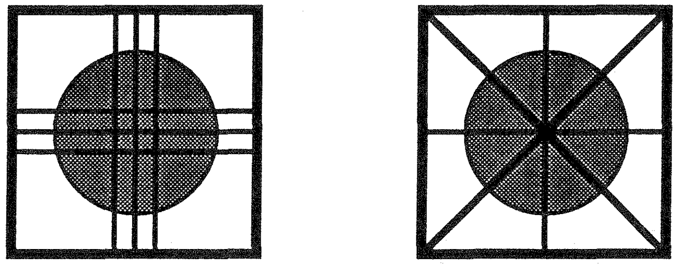
4.1. Level 1: Determining the Centroid and Size of the Target from Binary Image.

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 since this is achieved in the next level. All is needed here is to define the area required for the precision pointing. The binary image used in 3.1 is utilized here. The centroid of every recognized feature and the approximate dimensions of the target is computed and passed on to level 2.

4.2. 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 camera, ellipse shape is used instead.



(a) (b)

FIG.2: Windowing and Profiling for Target Edge Measurement. (a) is for large Targets and (b) is for Small Targets.

4.3. Evaluation:

The difference in accuracy between using the image coordinates immediately after level 1 and using those after level two 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 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.

V. THE MATCHING APPROACH

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

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

1. Within each level there are several sub-levels, or stages.
2. At the early stages, only the features, or points, that follows 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.
3. 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.
4. 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.

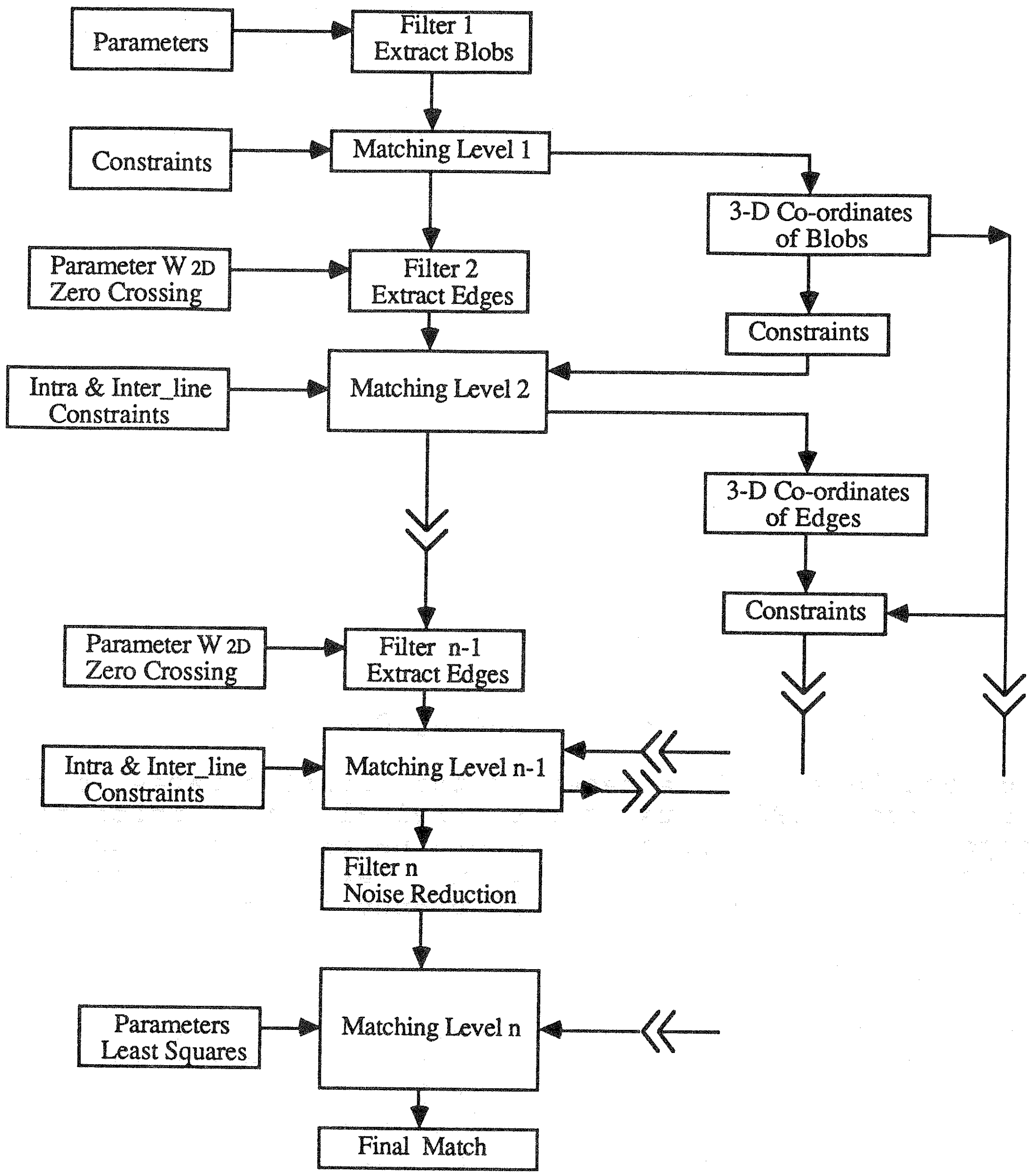


FIG. 3: Block Diagram of the Matching Process

To describe each level, and all the constraints and filters involved, in detail is beyond the scope of this paper. Only few comments are given below.
1. 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.

2. 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 Poggio, 1979) shown in figure 4. The width W_{2D} determines how much details are to be extracted. The smaller the width the greater 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 deceasing W_{2D}. Again, for each level the Z-coordinates as 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 and Ohta and Kanade, 1985).

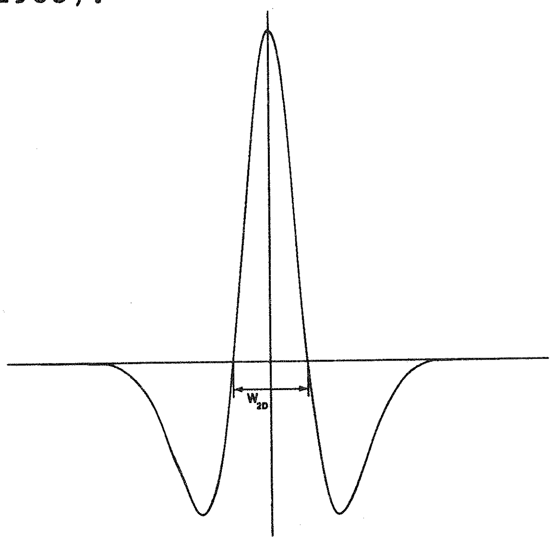


FIG. 4: The Zero-Crossing Operator

3. 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.

5.1. 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 [1], using only the epipolar line constraint results in successfully matching only about 80-85% 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 on step, (by having a very small W_{2D} to begin with) the success rate has been in the 60-65% range for objects with several inside edges. This has improved to about 90% 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% without constraints while going through all the levels raised the rate to over 95%.

VI. A PROCEDURE FOR ACCURACY EVALUATION

This section is a study based on hundreds of measurements of 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 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;

1. Sensor resolution
2. Sensor geometric errors (especially the horizontal-line jitter and the warm up effect)
3. Pointing accuracy (all existing methods have to make one or more approximate assumptions)
4. Camera orientation (geometric strength)
5. 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, Dahler,1987 and Luhmann and Wester-Ebbinghaus,1987) and affects the depth (Z) as well. This problem can be solved by a specially designed synchronization circuit (Havelock, 1988). It is therefore essential to study the repeatability of the system and accept all the fluctuation

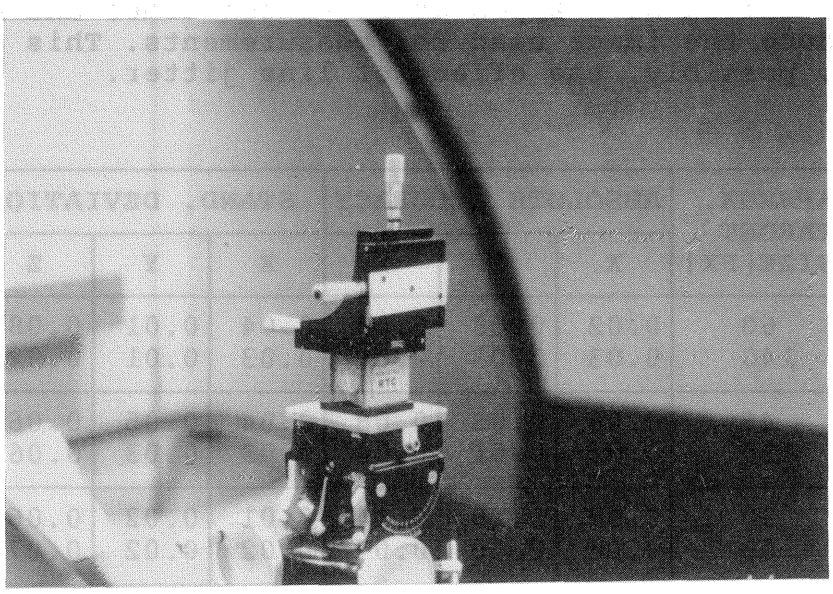


FIG. 5: XYZ-Axes Positioning System

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.

6.1. Test Device:

The three-axes positioning device shown in figure 5 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 absolute accuracy.

6.2. 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 for systematic errors compensation has been developed in an earlier study (El-Hakim, 1986).

6.3. Test Procedure and Results

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

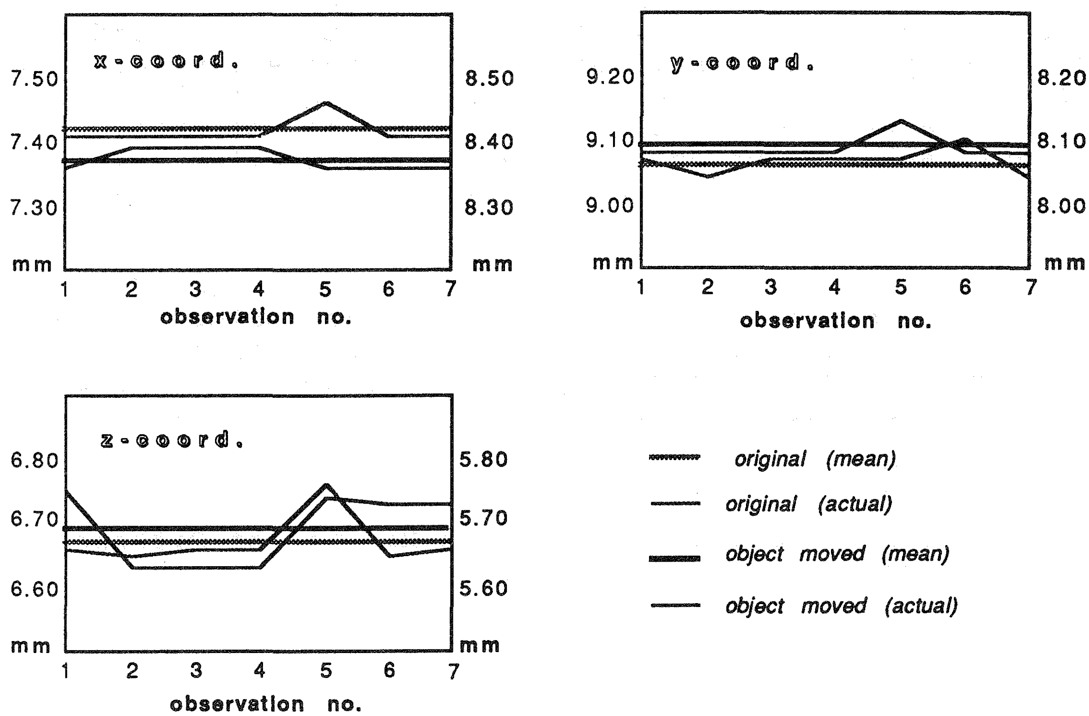
a. The cameras should be turned on several hours before the actual measurements started to minimize warm-up effects.
b. The images should be snapped at least ten times and averaged to produce the image used for measurements. This reduces noise and, possibly, the effect of line jitter.

MAX. OBJECT SIZE[CM]	**APPROX. TARGET SIZE[PX]**	**ABSOLUTE ACCURACY**			**STAND. DEVIATION**		
		X	**Y**	**Z**	**X**	**Y**	**Z**
25 X 19	60 140	0.02 0.03	0.03 0.03	0.08 0.01	0.04 0.03	0.01 0.01	0.08 0.02
40 X 30	45 100	0.06 0.06	0.03 0.01	0.08 0.04	0.04 0.03	0.05 0.03	0.06 0.06
60 X 45	30 60	0.07 0.06	0.10 0.10	0.14 0.07	0.01 0.02	0.02 0.02	0.06 0.07

TABLE 1: Absolute Accuracy and Standard Deviation [mm]

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 6 displays the repeatability of 7 measurements of one point 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 the measure of absolute accuracy gives the value;0.05 mm in X, 0.03 mm in Y and 0.01 mm in Z. However, these values are uncertain by as much as 0.09 mm in X, 0.08 mm in Y and 0.11 mm in Z, as shown in the figure (these values are computed by taking the furthest observation from the mean in each case). It is therefore very important to provide these repeatability figures along with absolute accuracy estimates.



[The vertical scale on the left is for the original position and the one on the right is for after moving]

FIG.6: Repeatability and Absolute Accuracy Diagrams for One Point.

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 0.06 pixel in X-Y plane and from 0.05 to 0.10 pixel in depth (Z). Using 256x256 resolution, this translates to a range of 1:7500 to 1:4600 of object size in XY plane and of 1:6700 to 1:4000 of object size in depth. A proportionally better accuracy is expected to be achieved for 512x512 resolution (such a system is now being tested at CIIT).

VII. 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 that 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.

ACKNOWLEDGMENT

The author wishes to thank Mr. Curtis Hillier of Diffracto Ltd. for his assistance with the tests.

REFERENCES

Agin,G. and Binford,T.O. (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.

Dahler,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", PERS, Vol.51(5), pp.545-552.

El-Hakim, S.F. (1986), "Real-Time Image Metrology with CCD Cameras", PERS, Vol.52(11), pp.1756-1766.

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

Havelock,D.,(1988), Personal communications, Division of Physics, NRC, Ottawa, Canada.

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 Wester-Ebbinghaus,W.,(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 Poggio,T.,(1979),"A theory of human stereo vision", Proc. of the Royal Society of London, Vol.B204, pp.301-328.

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

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

Murai,S.,Otomo,F.and Ohtani,H.,(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 Binford,T.O. (1986),"On Detecting Edges", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No.6.

Ohta, Y. and Kanade,T. (1985),"Stereo by intra- and inter-scanline search using dynamic programming", IEEE Trans. on Pattern Recognition and Machine Intelligence, 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.