A PHOTOGRAMMETRIC VISION SYSTEM FOR ROBOTS

S.F. El-Hakim
National Research Council of Canada
Ottawa, Ontario, Canada, K1A 0R6
Commission V

Abstract

Most of the existing robots can follow only specific sets of instructions. The skills required of an intelligent robot will include the ability to recognize shapes and motion and to give accurate measurements of objects in the scene. This can be achieved if the robot is able to measure, in three dimensions, surface points in the scene and to interact with the objects. A proposed real-time photogrammetric measuring technique, acting as the robot's visual system to acquire scene information, is outlined.

Introduction

Robots can be classified into two main groups depending on their ability to interact with the external environment. The first group includes robots that can perform sequences of programmed operations but cannot detect nor adapt to changes in their working environment. The second group includes advanced, or intelligent, robots which have their own visual system to acquire scene information and to recognize shapes and motion. An intelligent robot can, for example, select the objects it needs from a group of randomly placed different objects. Most of the existing industrial robots are of the first type. The intelligent is still in an experimental stage and only capable of dealing with very limited arrangements of objects and scenes. However, intelligent systems are becoming increasingly used in industry. They have proved useful, even with their limitations, in applications such as automobile assembly lines and inspection for quality control [Callahan, 1982 and Taboada, 1981]. By reducing their limitations, a multitude of industrial applications can be expected to benefit.

Ideally, the system should be reprogrammable to perform a variety of tasks, which means that most operations are to be carried out by software rather than hardware. (The definition of a robot, according to the Robotics Institute of America, is "a reprogrammable, multifunction manipulator designed to move material parts, tools, or specialized devices through variable programmed motions for the performance of a variety of tasks".) However, until the speed of present general-purpose-computers increases drastically to allow all the necessary computations to be carried out in real-time, most of the operations will have to be performed through the hardware or at least by plug-in modules. The required computation time, in order to be called real-time processing, must be less than the data acquisition time of one video frame. This time is 1/60 s for a vidicon camera with a resolution of 512x512 pixels. Whether an operation is hardwired or processed by software depends on the state of art of the electronics technology. In the system described in this paper, most of the operations are performed by software, but in order to convert the system into a production-type system, many of them have to be hardwired or made into hardware modules. A hardware module is 10-100 times faster than regular software while special purpose hardware unit is 100-1000 times faster. Nedless to say, the cost also increases proportionally.

The vision system typically consists of an all-digital data acquisition, digital image processing unit, surface measurement from image, and object recognition facility. This results in "seeing" some objects in the field of view and paves the way for subsequent manipulation with any particular object in the scene. However, the seeing process will need a priori information about the objects expected to be in the view; otherwise a definite description of these objects could not be possible. The third component of the vision system, surface measurement from image, is certainly a photogrammetric operation. However, well-known photogrammetric techniques may not fulfill all the requirements for a vision system. The main requirement, which is real-time processing, will be very difficult to satisfy using a two-camera system due to the fact that an image correlator is needed. An alternative is to use one camera and a projected pattern. But even then, a time-consuming analysis of the line pattern is necessary in order to automatically determine absolute line identity which is required to convert the image coordinates into object coordinates [Frobin and Hierholzer, 1983]. Another difficulty is the limited depth-of-field that conventional optical systems suffer from. Moreover, proper illumination of the work area is essential. The distinction between object and background can be impossible to accomplish by a computer if neither is a priori known to be brighter or darker than the other.

The following is a proposed photogrammetric vision system that attempts to solve some of the above-mentioned problems.

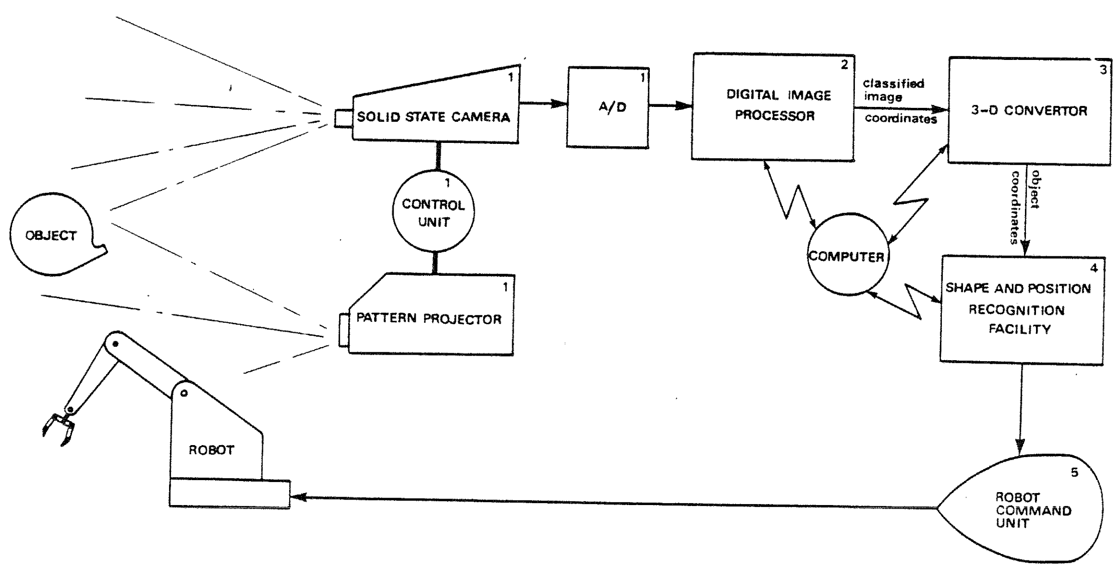

FIGURE 1: Elements of the Robot-Vision System

Elements of the Vision System

Figure 1 shows the elements of the proposed robot-vision system. There are five main elements:

1. data acquisition system which captures the scene data directly in digital form;
2. digital image processor to produce classified, or identifiable, image coordinates;
3. three-dimensional convertor to obtain object coordinates from image coordinates;
4. shape and position recognition facility to determine which of the measured objects is the one of interest; and
5. a robot command unit to manipulate the recognized object, from the previous step knowing its location with respect to the robot arm.

Some of these elements will be discussed in detail.

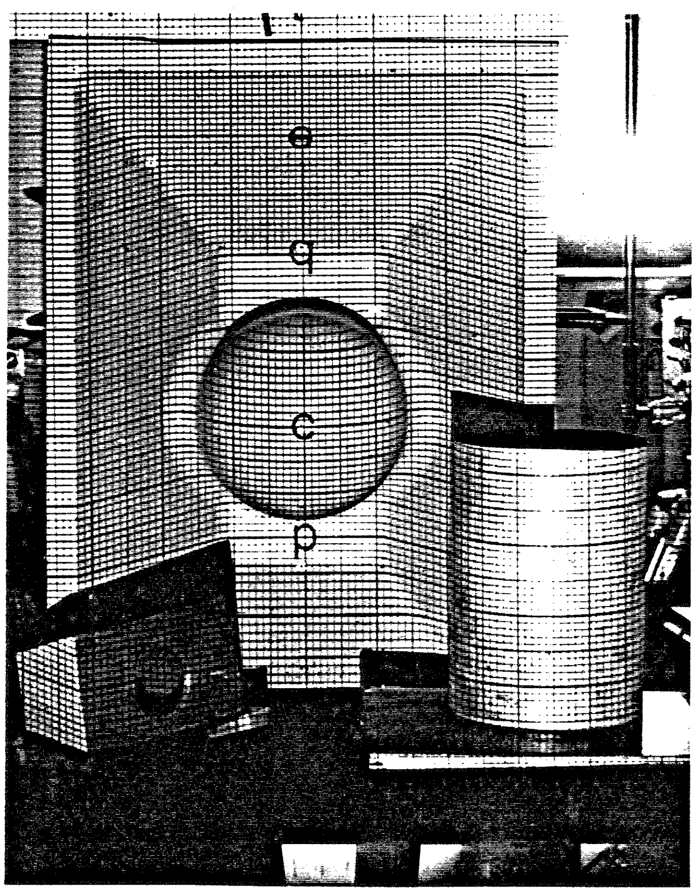

FIGURE 2: A Photograph of the Projected Grid

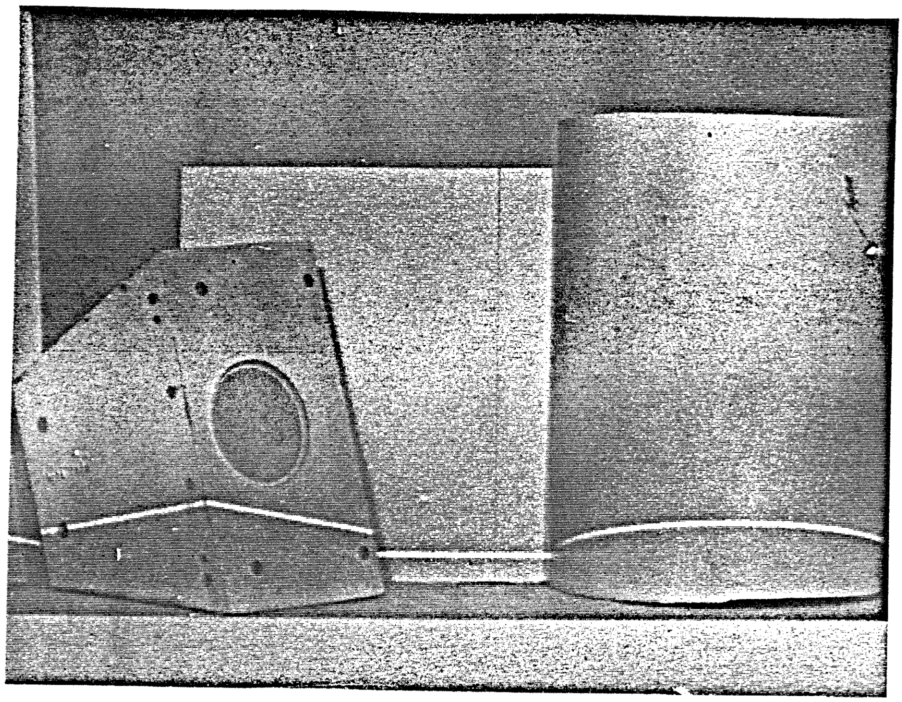

FIGURE 3: Single Line Projection

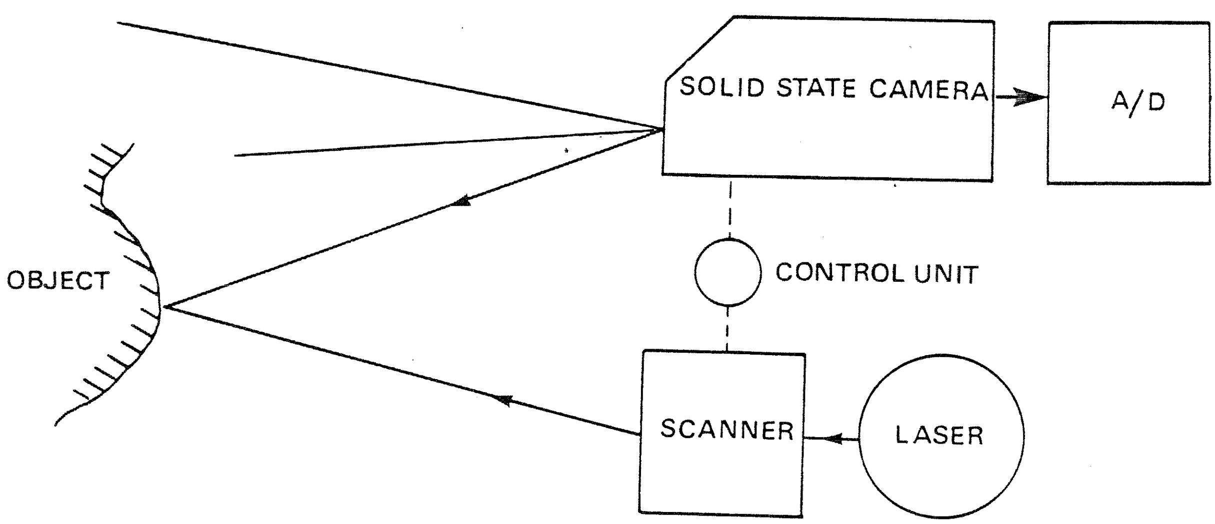

FIGURE 4: Data Acquisition System

The Data Acquisition System

For easier automatic image processing, a single camera and a projector with a raster diapositive are preferred to two cameras [Frobin and Hierholzer, 1983]. The former have the advantage of extracting all the information needed for three-dimensional measurement of a surface from a single photograph (Figure 2). This results in reduction of the time needed for processing and eliminates the need for correlation. The process can be further simplified by using a line raster instead of a grid. The three-dimensional object coordinates of any point on the raster line can be determined as shown later. However, this can be done only if each line is uniquely defined so that its coordinate, or number, on the raster image plane is known. An automatic procedure for unique line definition is given in [Frobin and Hierholzer, 1983] but being time consuming it is not suitable for real-time processing. Any other procedure using this type of raster would probably not be as fast as we wanted. Therefore, in the proposed data acquisition system, the projected pattern is that of only one straight line (Figure 3). This line scans the scene using a scanning device, such as an optical beam deflector, that can provide an accurate location (coordinate) of the line on the projector image plane, at a high speed (Figure 4). A control unit, to syncronize each line location with a camera frame is also needed. A laser beam is chosen for the light source to solve the problem of depth-of-field, at least for the projected pattern if not for the camera, and it also provides a high light intensity with very little noise and disturbance. The other component in the data acquisition system is the camera. A solid-state matrix or array camera with a high-speed analogue-to-digital (A/D) converter is suitable [Real, 1983]. Once the image has been digitized it will be stored in a frame-grab memory, and is ready for processing.

Digital Image Processing

Research in digital image processing has been, and continues to be, growing very rapidly during the past two decades. There are publications, too numerous to mention, that cover the broad area of this field attempting to solve the large number of problems associated with processing images by a computer rather than by a human operator.

The digitized picture obtained, for example, by the above-mentioned data acquisition system and stored in the frame grabber, consists of n x m pixels, each having a location coordinate and a grey level value or a video signal associated with it. The digital image processor is thus needed to identify each pixel or to search for pixels of interest such as those belonging to edges, objects and targets (pattern recognition). It is obvious that the success of this step depends largely on the previous step (data acquisition). The planning for the different elements of the data acquisition system, particularly for illumination or object-background relative brightness, if not done properly, could result in an unreliable or improper digital image processing. Another consideration is the accuracy with which the location of edges, lines, and other targets is determined. There are many algorithms available for precisely locating such targets from the different grey levels in the image with an error much smaller than the pixel size [Mikhail, 1983].

We are interested here in locating the line projected on the scene. As shown in figure 5, the projected line can be located by analyzing the video signal-to-pixel-location relationship. For a line in x-direction, which is the case in this system, the analyzed pixel location is along the y-direction. Since the line is the most brightly illuminated target, the peak in this direction represents the line location. The exact pixel, or subpixel, that represents

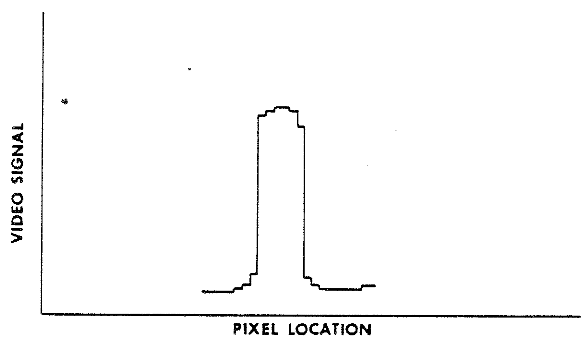

FIGURE 5: Video Signal at and Near Projected Line

the precise location has to be determined by some thinning techniques [see Kreis, et al, 1982]. In addition, any noise or disturbance to be eliminated. After searching through all the columns of pixels in the stored frame, a string of line point coordinates become available. However, the number of pixels in a frame could be too large to be stored and analyzed in real-time. Therefore, only the range in which the line is expected to appear is stored and processed. This range can easily be determined from the line location in the projector system and the expected smallest and largest depth in the scene.

The next step is to identify the edge points between different objects or between objects and background or edges within the same object. Edges between different objects or object and background can be defined as points where sharp discontinuity is detected (Figure 6). On the other hand, edges between different surfaces of the same object can be defined as points with highest variation in slope without discontinuity in the line. Determination of these edge points is thus possible by analyzing the image coordinates of line points in succession from one end of the line to the other. The result is the classification of points into groups belonging to certain objects, different surfaces of the same object or background. Image coordinates of object points only are transferred to the next processing step.

Three-Dimensional Surface Measurement

The object coordinates of any point on the projected line can be determined from the image coordinates (x', y') obtained from the digital image processor and from the known location of the line in the projector system (y"), provided the camera and the projector relative positions and orientations are known.

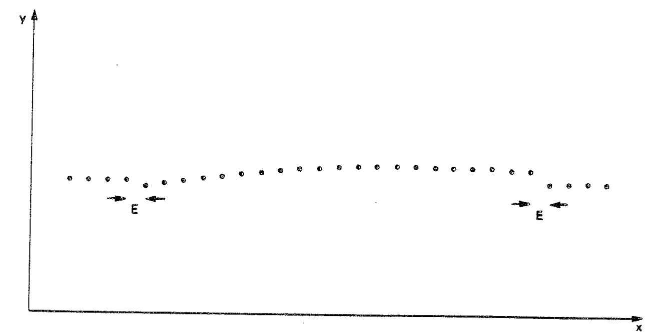

FIGURE 6: Edge Detection

This requires a priori calibration of the data acquisition system, using control points. The mathematical model is the collinearity equation:

$$\frac{T_1}{T_3} + \frac{x}{f} = 0, \frac{T_2}{T_3} + \frac{y}{f} = 0$$

where x and y are image coordinates of a point, f is camera focal length and T_1, T_2, T_3 are transformed object coordinates given by

$$(T_1, T_2, T_3)^{-1} = R^{-1} (X - X_c, Y - Y_c, Z - Z_c)^{-1}$$

where R is a rotation matrix, X, Y and Z are object coordinates in the object coordinate system, and X_c, Y_c and Z_c are coordinates of camera projection centre in the same coordinate system. The origin of this coordinate system can be chosen at the camera or the projector. R, X_c, Y_c, Z_c and f are determined from the calibration step before the real-time process. The mathematical model is then used with only X, Y, and Z of surface object points as unknowns. This means that we have three equations (one for each: x', y', y") and three unknowns, and thus only a simple direct solution without adjustment. Although not a rigorous solution in an operator-performed measurement, it can be accepted in an automated measurement due to the extremely high precision and the much smaller chance of a blunder.

The above procedure is applied to each object point on the projected line which gives a profile of the objects in the scene. For each picture frame, we have a new profile at a different interval. All the coordinates are relative to a coordinate system located at the camera or the projector and are now ready for the next phase of processing.

Object Shape and Position Recognition

Before being able to recognize certain objects, the computer has to learn about and memorize them. This can be achieved by placing each object in the view of the data acquisition system and applying all the previous steps. The computed coordinates, or cross sections, are then used to extract all the necessary information about the object. This includes fitting a function, s(X,Y,Z; a) = o, usually a polynomial or a simpler function, to the surface, or each surface, of the object. The values of parameters a of the function are determined and stored. Other useful information such as the perimeter area and the center of gravity are also determined and stored.

When the actual real-time operation starts, the objects in the field of view are processed as mentioned earlier to determine their three-dimensional coordinates. It is required now to identify which of these objects match a certain stored, or memorized, object. This requires fitting X, Y and Z coordinates to the function S(X,Y,Z,a) = o using the previously determined a-parameters for the object of interest. However, additional unknowns must be added, particularly the orientation parameters and shifts in the centre of gravity between the object to be recognized and the stored or reference object. A match is declared if the RMS of the residuals is within a preset tolerance, otherwise another object is processed.

The efficiency and reliability of this approach depends on the complexity of the object and its visibility to the data acquisition system. A significant improvement can be achieved if several data acquisition systems cover the objects from different angles. These systems will have a common recognition facility into which the information from all the units is entered for processing.

Conclusion

The vision system outlined above is able to measure surfaces in real-time and to recognize shapes and positions of objects for robot manipulation. At the present time, many of the elements are in a feasibility study stage. The elements which have been tested so far are the three-dimensional surface measurements and the recognition of simple objects. The main purpose of this was to test the data acquisition system and the efficiency and workability of the algorithms. Although all the computations were carried out by software on a general purpose computer, the surface measurements took only about one second while object recognition took about five seconds. This time will increase for more complex scenes and object arrangements and thus, depending on the time limit set by the system's eventual-user, some of the software must be converted into modular hardware.

Acknowledgements

The author wishes to thank Manfred Paulun and Al Way-Nee of his affiliation for their help in implementing some of the ideas presented in this paper at NRC laboratory.

Bibliography and References

Callahan, J.M. 1982, "The State of Industrial Robotics", BYTE, October, pp. 128-142.

Frobin, W. and E. Hierholzer, 1983, "Automatic Measurement of Body Surfaces Using Rasterstereography - Part 1: Image Scan and Control Point Measurement", Photogrammetric Engineering and Remote Sensing, Vol. 49, No. 3, pp. 377-384.
Part II: Analysis of the Rasterstereographic Line Pattern and Three-Dimensional Surface Recognition", Photogrammetric Engineering and Remote Sensing, Vol. 49, No. 10, pp. 1463-1452.

Geisler, W. (1982), "A Vision System for Shape and Position Recognition of Industrial Parts", Proceedings of the 2nd International Conference on Robot Vision and Sensory Control, Nov. 2-4, Stuttgart, W. Germany, SPIE, Vol. 392.

Hall, E., J.B.K. Tio, C.A. McPherson and F.A. Sadjadi, 1982, "Measuring Curved Surfaces for Robot Vision", Journal of IEEE Computer Society, Vol. 15, No. 12, pp. 42-54.

Kreis, Th., H. Kreitlow and W. Juptner, 1982, "Detection of Edges by Video Systems", Proceedings of the 2nd International Conference on Robot Vision and Sensory Control, Nov. 2-4, Stuttgart, W. Germany, SPIE Vol. 392.

Mikhail, E.M., 1983, "Photogrammetric Target Location to Subpixel Accuracy in Digital Images", Proceedings of the 39th Photogrammetric Week, Stuttgart, W. Germany, Sept. 19-24.

Pugh, A., 1983 (Editor), "Robot Vision", IFS publications Ltd., U.K.

Real, R.R., 1983, "Matrix Camera with Digital Image Processing in Photogrammetric Applications", Proceedings of the Annual Meeting of ASP, Washington, D.C., pp. 255-266.

Sucui, R.E. and A.P. Reeves, "A Comparison of Differential and Moment Based Edge Detectors", IEEE, 82CH1761-6, pp. 97-106.

Taboada, J., 1981, "Coherent-Optical Methods for Applications in Robot Visual Sensing", SPIE Vol. 283, 3-D Machine Perception.