Tr. J. of Engineering and Environmental Sciences
21 (1997) 325-333

Visual Control of a Heavy-Duty Robot Manipulator

Sadettin KAPUCU, Sedat BAYSEÇ
University of Gaziantep, Mechanical Engineering Department, Gaziantep-TURKEY

Received: 19 / 12 / 1996

Abstract: Robots can be made more flexible and adaptable by incorporating sensory information into the feedback loop. Vision-based robots can adapt quickly to the emerging requirements of an unknown task and can react appropriately to changes in the environment. Because image identification involves massive arithmetic and data analysis, the traditional approach to using visual feedback has been to separate the problems of obtaining information about the object identity and position according to the camera images from the manipulator control. In this paper, the amount of visual information required is reduced to a minimum so as to speed up image analysis, so that it can be done on-line and on-duty. A spot of definitive color is put onto the object to be identified and manipulated, and the control computer is programmed to search for this unique and definitive color spot in the image captured, relieving the system of the necessity for intricate image identification. The difference between the locations of the spot in stereo images produced by two static, parallel-axis cameras reveals the distance of the spot from the cameras. Three such points define the location and angular orientation of the object in space. The position of the characteristic spot is determined with respect to the manipulator coordinate frame and, via inverse kinematics, the robot is guided to reach the spot. Experimental results of the manipulator reaching a single spot, showing the duration of visual for vision analysis and manipulator motion are presented.

Key Words: Visual control, Stereo vision, Robot, Camera calibration

Görüntü Analizi ile Bir Ağır İş Robutunun Yönlendirilmesi

Özet: Robotlar değişik duyargalardan gelen bilgilerin geri beslenmesi ile daha esnek ve uyumlu hale getirilebilirler. Görüntü algılayabilen robotlar kendilerini bilinmeyen bir işte ortaya çıkan öngörülmemiş hareket ihtiyaçlarına yada çevrede oluşan değişikliklere kolayca adapte edebilirler. Görüntü tanımlama problemleri çok fazla aritmetik ve bilgi analizi gerektirdiğinden robotlarda görüntü algılamanın geleneksel yolu kamera görüntülerinden cismin tanımı yada konumu ile ilgili bilgilerin derlenmesi ile maniplatörün kontrolunun ayrı ayrı yapılması olagelmiştir. Bu çalışmada görüntü analizini hızlandırarak robotun çalışması sırasında uygulayabilmek için gerekli görsel bilgi ihtiyacı en aza indirgenmiştir. Tanımlanacak yada manipüle edilecek cismin üzerine önceden belirgin ve tanımlayıcı renkte bir benek konularak kontrol bilgisayarının alınan görüntü üzerinde karmaşık görüntü-algılama yöntemleri uygulamak yerine sadece bu belirgin renkteki beneği araması sağlanmıştır. Yanyana, eksenleri paralel duran iki statik kameradan gelen bir çift stereo görüntüde beneğin konumundaki farktan beneğin kameralara olan mesafesi belirlenmiştir. Cismin üzerindeki bu gibi üç beneğin koordinatları cismin konum ve açısal oryantasyonunu eksiksiz olarak belirler. Daha sonra cismin konumu manipülatör koordinatlarına çevrilmiş ve ters-kinematik denklemleri ile manipülatör beneğe ulaşmak üzere yönlendirilmiştir. Manipülatörün tek bir beneğe ulaşması sırasında görüntü-analiz ve manipülatör hareketinin zamanlaması deneysel sonuçları ile gösterilmiştir.

Anahtar Sözcükler: Görsel kontrol, Stereo Görüntü, Robot, Kamera kalibrasyonu.

Introduction

Many robots are designed in imitation of the human arm. The structure and capabilities of a manipulator arm are very basic. The most intricate part of the human motor system is the brain itself, which receives sense-data of the environment, compares these observations with knowledge already stored, decides what to do and gives commands to the appropriate muscles. While motion is continuing, the process of *observation* and *decision-making* continues. These mechanisms are dependent upon learning, the accumulation of knowledge and experience. Robot control operates on the same principles, i.e. sensing what is going on in the environment and within the manipulator, deciding *what to do, adapting and producing* actuation commands accordingly. Adaptive control systems can make use of any kind of *sensory device*, but of the five senses known to mankind, *vision* is by far the most important and valuable. From visual data information such as existence, position, color and shape can be obtained. The difference between positions recorded at two different times can be used to calculate the velocity and acceleration of an object, as explained in the paper of Roach and Aggarwal published in 1980. The aim of this study was to include vision in the sensory capabilities of a robot system working on the shop floor, enabling it to adapt itself to visually detectable changes in the environment.

In manufacturing systems, the additional cost of extra machinery required to place the object to be manipulated in the correct position and angular orientation is often overlooked. Jigs, fixtures, conveyors, indexers and other similar devices which are, to some extent automatic, are then added and are nearly the same price as the robot itself. More intelligent robots are equipped with more intricate sensory devices and with comparatively sophisticated software to replace this extra machinery. Vision provides the control system with vast amounts of data about the object to be manipulated. A machine vision system captures an image with a camera and analyzes the image to produce a description of the objects viewed. The input of a machine vision system is an image, while its output is a set of numbers describing the color and brightness levels of pixels. These numbers can be quickly determined from the image and they contain enough information for a decision to be made. An object does not usually have a unique description and hence it is impossible to describe it completely. A complete description of all the objects in the image is seldom necessary. Instead, features particular to the object under consideration are determined. The number of special features can vary according to the task for which image identification is used. The symbolic representation of an object must be known and this symbolic representation is interpreted and compared with that of the image by a computer.

The basic information needed from a machine vision system is the location and angular orientation of the object under consideration, disregarding all the other details embedded in the image. To extract this information, 6 independent quantities must be measured as the total degrees of freedom in space is 6. In practice, this is obtained from the coordinates of 3 non-coinciding points on the object, 9 numbers in total, 3 of which are redundant. The redundant data are eliminated by 3 constraint equations describing the rigidity of the object as the distances between any pair of points which are constant . 3 spots of distinct color, size and spacing can be put on an object at precise locations during manufacture. The x,y,z coordinates of these points are derived from the image. This method can define not only the location and orientation of an object, but its type as well, from the color, size and spacing of the spots. This method of identification has been considered by several researchers to be practical and useful on the shop floor, for example, Wijesoma et al., 1993. The coordinates of the points are obtained from two static, parallel-axis cameras. The displacement commands which move the manipulator to the correct position are then determined by processing the data with the inverse-kinematics routine. The commands are given to the actuators as step inputs. The process is explained in detail in the Ph.D. work of Kapucu, 1994.

Identification of a Characteristic Spot

The visual system used in identification presents images on a computer monitor, the color definition of each pixel being described by 3 independent integers. In a 24-bit system, these numbers range between 0 and 65535. The location of each pixel on the screen and the 3 different color definitions describe not only the external appearance of the object in view, but also its location and angular orientation, as well as all the other objects around it which are within the field of view. In robotics, this quantity of information cannot be analyzed in a reasonable amount of time and, hence, the identification of a definitive color spot on the image is proposed in this paper.

A subjective definition of color can be made by comparison with reference colors in terms of the attributes brightness, hue and saturation. Intensity, or brightness, is the attribute that is perceived in achromatic images. Hue represents dominant color as perceived by an observer. For example, red is of a different hue to green. Saturation refers to the relative purity or amount of white light mixed with a hue. The pure spectrum colors are fully saturated. The sensation of colour is stimulated by the wavelength of the emission. Pure red is defined as having a wavelength of 650 Nm, green 515 Nm and blue 480 Nm. Different objects may be expected to have different colors and, hence, color discontinuities are keys in the identification of color images. Nevatia, 1977, for example, uses color-edge discontinuities to identify different segments in an image. Brockelbank and Yang, 1989, used photogrammetry on color images of solid objects in order to describe their surface structure. It is widely accepted that a complete description of a point in a color image requires the identification of three mutually independent parameters. These are the primary colors, red (R), green (G) and blue (B). Any color of light can be produced by mixing these three primaries in the required proportions of energy. Similarly, the components of a color light can be defined, first, by separating it into its primaries using a prism, and then measuring the energy radiated in each primary color. On the screen of a digital computer images are formed by tiny pixels, each with a certain color. The energies of the primary components of a pixel can be determined with software. In a 24-bit system, for example, the component energies are represented by 3 integers ranging from 0 to 2^{24}. Healey, 1989, described how colors are differentiated on a digital computer according to their wavelengths. Different ways of defining and reproducing colors have been developed, all of which have become standard textbook material, for instance, Gonzales and Wintz, 1987. The most important are the Colorimetric System, Nevatia Transformation, and NTSC color systems. The colorimetric system is the basis of the objective definition of colors. It is shown in the chromaticity diagram shown in Figure 1, published by the Comission Internationale de l'Eclairage (CIE), 1931. This figure represents a 3-dimensional surface in a right-handed cartesian frame, in the form of a distorted tetrahedron. As seen in Figure 1, the base of the tetrahedron is on the x-y plane and at its corners are the pure primary colors. Towards the apex, the colors are close to white, the combination of all three primaries. This figure contains all the known colors and any color is represented by its x, y and z coordinates as:

$$x = X / (X+Y+Z)$$
$$y = Y / (X+Y+Z) \quad (1)$$
$$z = Z / (X+Y+Z)$$

where X,Y and Z are the red, green and blue energy levels, respectively, known as the *tristimulus values*. It is evident from these equations that :

$$x + y + z = 1 \quad (2)$$

Hue and saturation taken together are called chromaticity and, therefore, a color is defined by its brightness and chromaticity. Nevatia, 1977, proposed a new definition of color with a set of transformation formulae. The NTSC color system is explained in detail by Lim, 1990. The identification of a certain color spot and the distinction of the colour from false points of the same color definition is examined in the Ph.D. thesis of Kapucu, in which he concluded that the NTSC system is more versatile.

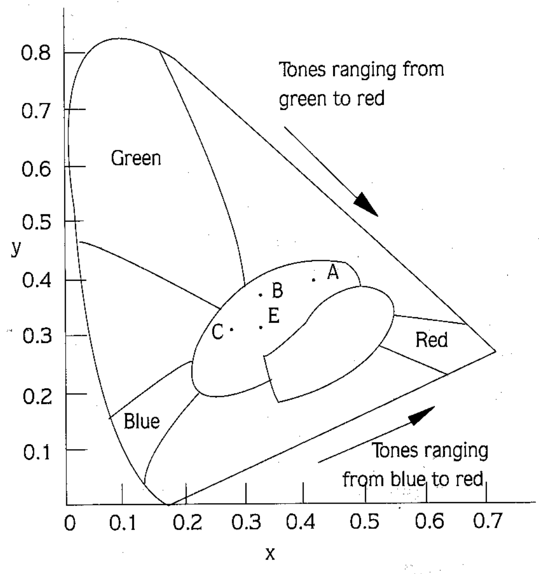

Figure 1. The CIE 1931 chromacity diagram with descriptions of the various color regions. A, B, and C represent the positions of standard sources, roughly equivalent to gas-filled incandescent lamps, noon sunlight, and average daylight, respectively. E represents the equal energy point corresponding to perfect white.

The NTSC (National Television System Committee) has described a new set of definitive quantities denoted by Y, I and Q, derived from the tristimulus values R, G and B, as

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.272 & -0.322 \\ 0.211 & -0.523 & -0.312 \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

and

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0.956 & 0.621 \\ 1.000 & -0.273 & -0.647 \\ 1.000 & -1.104 & 1.701 \end{bmatrix} = \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} \quad (4)$$

The Y component is called the *brightness component* since it roughly reflects the luminance of the color. It primarily accounts for the perception of the brightness of a color image and can be used to form a black and white image. The I and Q components are called the *chrominance components*, and they primarily account for the perception of the hue and saturation of a pixel. In the NTSC color system, the threshold values T' and T", representing the hue and saturation, respectively, can be arbitrarily defined for a certain color and the conditions

$$I > T' \text{ and } Q > T'' \quad (5)$$

are sought.

As daylight and white light produced by incandescent and fluorescent lamps contain all the color components, any one of the pixels forming the image may be found to contain the color combination with the limits set by equation 5, although it is not to be taken into consideration. Therefore, a sort of masking algorithm must be used to distinguish these false pixels from the pixels forming the definitive color spots. In this study 3x3 masking was used, which recognises a pixel as a part of the characteristic spot if all the 8 surrounding pixels are also of the same color definition.

Stereo vision

In a pinhole camera, light rays from an object at an infinite distance pass through the hole and form an inverted image of f distance from at the other side of the hole. f is called the focal length. It can be assumed that the non-inverted image occurs on a screen f distance from the pinhole, on the same side as the object itself. Effective photogrammetry requires two images taken from two different points. The focal lengths, camera directions, and the distances between the cameras and the object can all be different. To avoid the unnecessary complication of a large number of parameters, the general practice is to use two cameras of the same focal length in parallel-axis geometry. They are placed side by side at the same height, with their axes parallel. As seen in Figure 2, the cameras are represented by pinholes at O_L for the left and O_R for the right camera. It is assumed that non-inverted images are formed on screens f distance away from the pinholes. The light-sensitive imagers of the cameras are represented by the image plane I_L on the left, and I_R on the right. A point P with coordinates (X,Y,Z) appears as P_L on the left image plane and P_R on the right image plane. Two cartesian frames are attached, one to each camera X_L, Y_L, Z_L, for the left and X_R,Y_R,Z_R for the right camera, with origins at the corresponding pinholes. The separation between the camera axes or pinholes is called the baseline, denoted by b. In stereo vision, the observed point must appear in both image planes. From the geometry of Figure 2, the actual coordinates of point P(X,Y,Z) with respect to the coordinate frame of the right camera can be defined as:

$$X = bX_L/d \quad (6)$$

$$Y = bY_L/d \quad (7)$$

$$Z = bf_{eff}/d \quad (8)$$

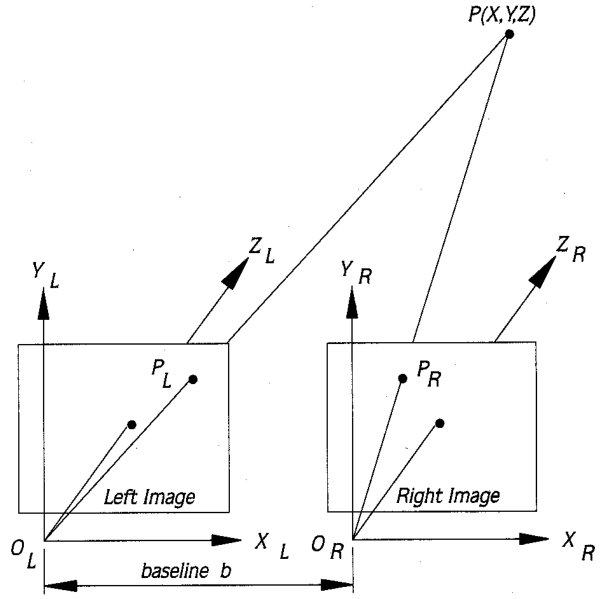

Figure 2. Perspective view of the parallel-axis-geometry stereo-vision camera arrangement.

where d = (X_L-X_R) and is called the *disparity*, and, f_{eff} the effective focal distance. The origin of the (X,Y,Z) coordinate system is coincident with the origin of the left camera coordinates. X_L and Y_L are the coordinates of P_L in the left image and X_R and Y_R are the coordinates of P_R in the right image. As the cameras are at the same height, $Y_L = Y_R$. A point in the environment visible from both camera stations produces a pair of image points called a conjugate pair. Note that a point in the right image and a corresponding point in the left image must lie somewhere on a particular line, because the two points have the same Y coordinates. This line is known as the *epipolar line.* A feature that appears in the left image may or may not have an equivalent in the right image, but if it does, it must appear on the corresponding epipolar line. It cannot appear elsewhere. Note that in parallel-axis-geometry stereo vision (Figure 2) all epipolar lines are parallel to the X axes of the camera coordinates.

Camera calibration

The manipulator under consideration is a heavy-duty hydraulic machine of Stanford configuration. The manipulator arm has three degrees of freedom, facilitated by two revolute joints and one prismatic joint (RRP), as shown in Figure 3. It is monitored by a pair of cameras fitted with zoom lenses whose focal lengths are controlled by a computer.

The cameras are positioned parallel at the same height. The distance between the camera axes is 12.6 cm. The cameras are intended to measure the physical dimensions and mobility of the manipulator by successively incrementing one degree of freedom while the others are kept stationary. At the same time, the cameras monitor the position of a miniature red-colored beacon placed at the tip of the manipulator. After the kinematic specifications of the manipulator have been determined, the cameras are zoomed by the computer so that only the area in which the manipulator operates is observed. Initially, the focal lengths are set to 1 cm (the widest angle position) so that the whole manipulator is seen by both cameras. It is assumed that there is a right-handed inertial cartesian frame whose origin is the pinhole of the right camera. The optical axis outward from the camera is chosen as the positive Z_c axis, while the vertical upward direction from the pinhole is the positive Y_c axis and the rightward direction is the positive X_c axis. This frame is fixed to the cameras and the ground and is called the frame of *camera coordinates*. A second right-handed cartesian inertial frame is assumed, as shown in Figure 4, whose origin is at the base of the manipulator. In this frame, the Z_W axis is vertically upward, in line with the revolute axis of the first joint, and the X_W axis Lags the Z axis of the camera coordinates by an arbitrary 126 degrees. The distance between the Y axis of the camera coordinates and the Z axis of the world coordinates is arbitrarily fixed at 454 cm. The origin of the world coordinates is 78.5 cm below that of the camera coordinates. The actual calibration of the cameras is done experimentally. During the calibration of the cameras, a red beacon placed at the tip of the manipulator is monitored. The parallelism of the cameras is ensured by observing the distance between two lines having a separation equal to that of the distance between the optical axis of the two cameras. The cameras see the plane on which these lines are drawn in actual size. The right camera sees the line on the right at its optical center while the left camera sees the other at its optical center. The plane containing the lines is translated along the Z_c axis to different locations and viewed through the cameras. If the camera axes are parallel, the distance between the images of the observed lines does not change. Adjustments are made until this criterion is satisfied.

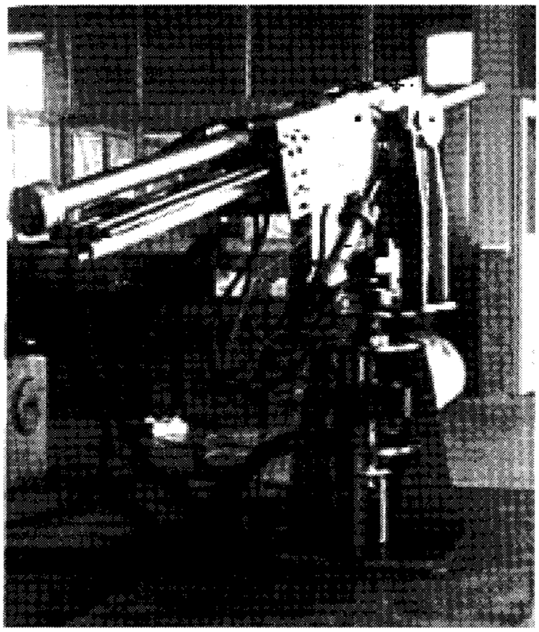

Figure 3. Heavy-duty hydraulic manipulator under visual control through images from a pair of video cameras used in the tests.

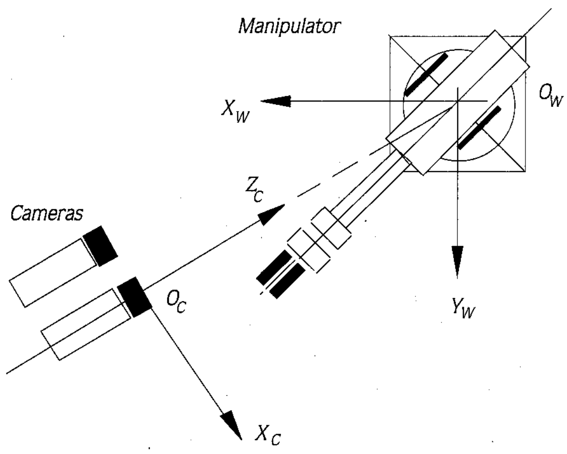

Figure 4. Arrangement of the cameras and the manipulator.

Once parallelism has been confirmed, pictures of the characteristic point are taken. The characteristic point is a distinctive color for easy identification. A definition of the color exists in the memory of the computer and the computer searches the image for a pixel containing this specific color definition using the method of bisection. Once such a pixel is detected, a window large enough to contain the characteristic spot is searched pixel by pixel. In this case, the window size is 90X90 pixels. After the colored patch is identified, the centroid is found. The same procedure is repeated by the other camera. At this stage, two sets of image coordinates, in terms of pixels, of the centroid of the characteristic point have been obtained. The image coordinates in pixels are transformed into camera coordinates, as described above. The camera coordinates describe the actual position of the characteristic point on the manipulator with respect to the camera coordinates, in centimeters. As the camera coordinates are not aligned with the manipulator, the calculation of the manipulator kinematics and position commands produces difficulties. Therefore, the creation of a second inertial cartesian frame, located at the base of the manipulator, is required. The conversion of the camera coordinates into world coordinates is done by multiplication by a displacement matrix. The elements of the displacement matrix are found experimentally by a procedure known as *hand-eye calibration.* One recent example of hand-eye calibration is described by Wijesoma et al. (1993). Brown presented a case study of the calibration of a close-range camera application in his 1971 paper. Grosky and Tamburino (1990) tried to bring about a unification in camera-calibration applications. As a result of defects in the material and geometry of lenses, a lens system may distort the image and the colors. In photogrammetry, distortion-of-image geometry is the most common error source and hence must be compensated for. Weng et al. (1992) presented distortion models and an accuracy analysis of them. Tsai (1987) stated that a simple radial distortion model is satisfactory for industrial-machine vision applications. The model he suggests generates the corrected coordinates x_{ic} and y_{ic} of a point from the measured coordinates x_d and y_d, which are distorted as follows:

$$x_{ic} = x_d + x_d\, kR$$
$$y_{ic} = y_d + y_d\, kR \qquad (9)$$

where R is the radial distance of the point from the center of the imager and k is a constant number called the *distortion coefficient.* The value of the distortion coefficient has to be determined experimentally as the origin of the camera coordinate frame is at the camera axis,

$$R = \sqrt{x_d^2 + y_d^2} \qquad (10)$$

This model was used in the work presented here to compensate for the *barreling* effect on the images. A thorough analysis of image technology can be found in Horn (1986).

Hand-eye Calibration

This is the determination of the transformation between the coordinate systems fixed to the camera and the manipulator. The transformation between the manipulator and camera coordinate systems can be treated as a rigid body motion and can thus be broken down into a rotation and a translation. If $r_c = (X_c, Y_c, Z_c)^T$ is the position of a point P in the camera coordinate system and $r_w = (X_w, Y_w, Z_w)^T$ is the position of the same point in the world coordinate system, then the following relationship exists between the two systems :

$$r_c = Rr_w + r_o \qquad (11)$$

where R is a 3X3 orthonormal matrix representing the rotational part and r_o is an offset vector corresponding to the translational part. An orthonormal matrix has columns that are mutually orthogonal unit vectors so that

$$R^T R = I \qquad (12)$$

where I is the 3X3 identity matrix. To find the elements of the rotation and translation matrices, world and camera coordinates from 4 points are required. These can be obtained by direct measurement and equation (5) introduces certain simplifications. In the hand-eye calibration done in this test, the rotation matrix and offset vector were found to be:

$$R = \begin{bmatrix} -0.82 & 0.58 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ -0.58 & -0.82 & 0.0 \end{bmatrix} \qquad (13)$$

$$r_o = \begin{bmatrix} 0.0 \\ -78.5 \\ 454.0 \end{bmatrix} \qquad (14)$$

Experimental Results

To demonstrate the positional accuracies attainable with the NTSC system, identifying a color beacon, and using pixel masking and radial distortion in the visual control of a manipulator, a red-colored beacon was placed within the work space of the manipulator. The cameras and lenses were set to observe the whole manipulator and its work space. A red light bulb was used as the characteristic spot to be located by the robot. Normally, the control program is capable of identifying a color patch when there are conditions of visibility. The world coordinates of the centroid of the red light bulb was calculated from the left and right images taken by two parallel-axis Pulnix TMC 74 color cameras. These coordinates were used to calculate the joint variables of the manipulator. The joint variables were then converted into command signals and fed to the appropriate actuators. The response of each joint to these command signals was recorded by the computer. Figure 5 shows the action of detecting a beacon located at the world coordinates (80, 32.5, 69.5) centimeters and the manipulator's reaching that point in joint coordinates. Successive image processings were performed 10 times to improve accuracy. The mean camera coordinates of the point were then calculated and transformed into world coordinates. The world coordinates of the point were calculated as (81.46, 35.13, 69.72) centimeters. The errors in the calculated X, Y, and Z coordinates were found to be 1.83 %, 8.1 % and 0.32 %, respectively. It can be seen from Figure 5 that there were no steady-state errors between the manipulator response and the calculated and commanded joint variables in the revolute joints, but there was an offset error between the commanded and actual position of the prismatic joint variable. Position errors, stemming mainly from image processing, in the first and second joints were 1.75° and 1.09°, respectively. However, in the third joint, there was a steady-state error of 0.5 cm, in addition to the image-processing error of 2.54 cm. During the experiment, the lenses were at the widest angle settings and the relative size of the point in the image window was small. Such conditions generally cause the loss of a portion at the boundary completely if it is less than half a pixel. This, of course, distorts the actual shape of the spot and introduces an error into the determination of the coordinates of the centroid. Larger size spots can be processed with greater accuracy. To improve accuracy and reduce lens distortion, normally, the lenses should be zoomed so that the spot is as large as possible in the image. Successive analyses of the same view gave different, but similar data. The calculated coordinates gave a distribution around the correct value. Therefore, taking images several times and averaging the resulting coordinates produced a better result. Searching a 400X400 pixel window in both images for a red pixel using the method of bisection took about 9.2 seconds. Once a red color pixel was found, opening a secondary window of 50X50 pixels around the identified pixel in both images and defining the colors of all the pixels inside them took about 8.4 seconds. In the first experiment, the lenses were kept at the widest angle position so that the

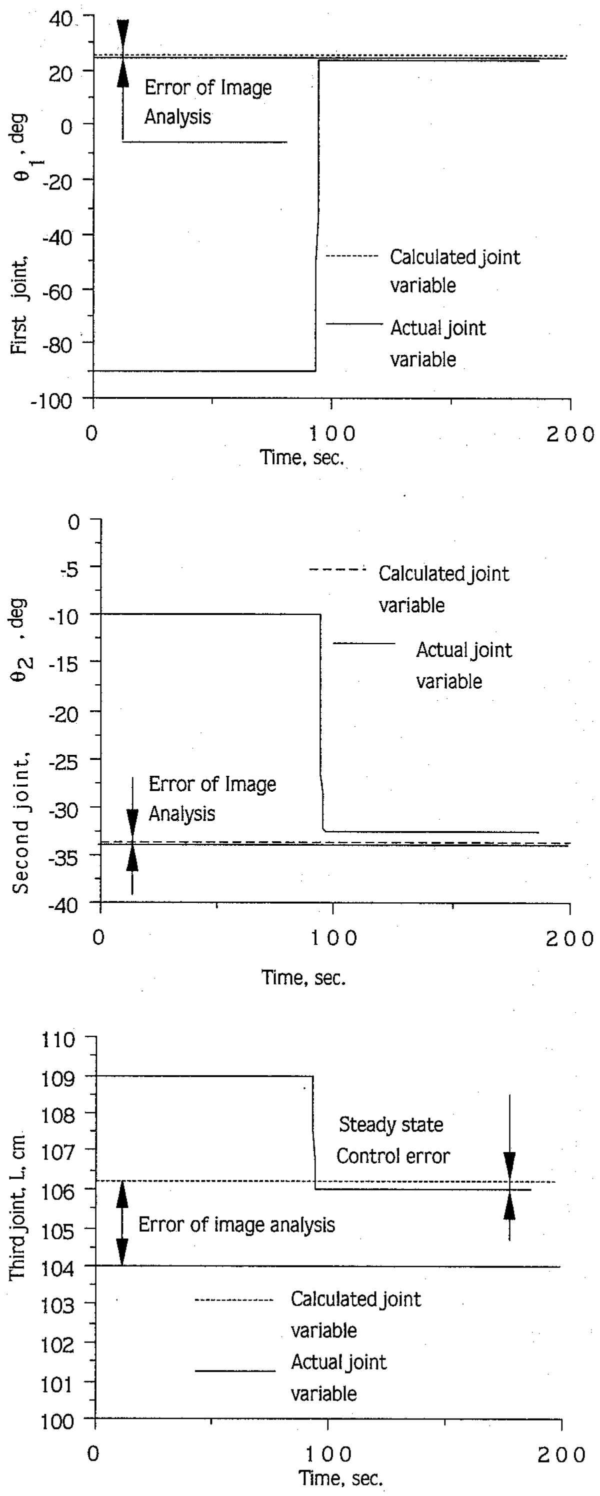

Figure 5. Joint response of the manipulator to reach the point placed at (80, 32.5, 69.5) cm coordinates from a parked position. To improve accuracy, the image processing was performed 10 times and the mean of the measurements was used in the guidance of the manipulator.

test of i*dentifying a spot and commanding the robot to touch it* could be repeated several times with points at different locations. The time which elapsed during the change in position can be obtained from the control computer and, hence, it is possible to calculate the average velocity defined as the ratio of *change in position to the corresponding change in time*. Once the average velocities at certain precision points/positions are found, a curve can be fitted to get the velocity profile through time. From this curve, higher order derivatives can also be worked out, but generally the most intricate conditions one can ask from a robot are velocity constraints.

In Figure 6, the motion profiles of the 3 joints of the manipulator touching 3 different points successively are shown. The points were all identical red lightbulbs and were lit one at a time to simulate a moving point. In the first trial, only one image was processed to define the coordinates of a point. With 3 points, approximately 54 seconds were required. 52.8 seconds of this duration was used in 3 image analyses. If image analysis is carried out only once, the possibility of errors is great and random. To improve accuracy, the experiment was repeated with 10 samples of image analyses for each point and the corresponding motion profiles are presented in Figure 7. It was observed that increasing the number of samples improved the accuracy of the calculated values. Naturally, this also increased the total time of identification.

Conclusion

Positioning 3-dimensional objects using stereo images has been the concern of numerous researchers and has potential in many fields of application, as indicated by Kim and Aggarwal (1987). Generally, time on the shop floor is extremely expensive. Therefore, sources of errors in image analysis must be examined in more detail to optimize the process so that a reliable position search can be performed via image analysis. This paper has presented a case study in which the location of a single point is determined by stereo images. The problem can easily be extended to 3 points which together define the position and angular orientation of a 3-dimensional object. It is also possible to determine the coordinates of a moving body at discrete positions at equal time intervals, which can lead to the *traction of a moving point,* at least for objects moving slowly. These durations can further be reduced by writing the programs in Assembler instead of C, the language of the control program in this

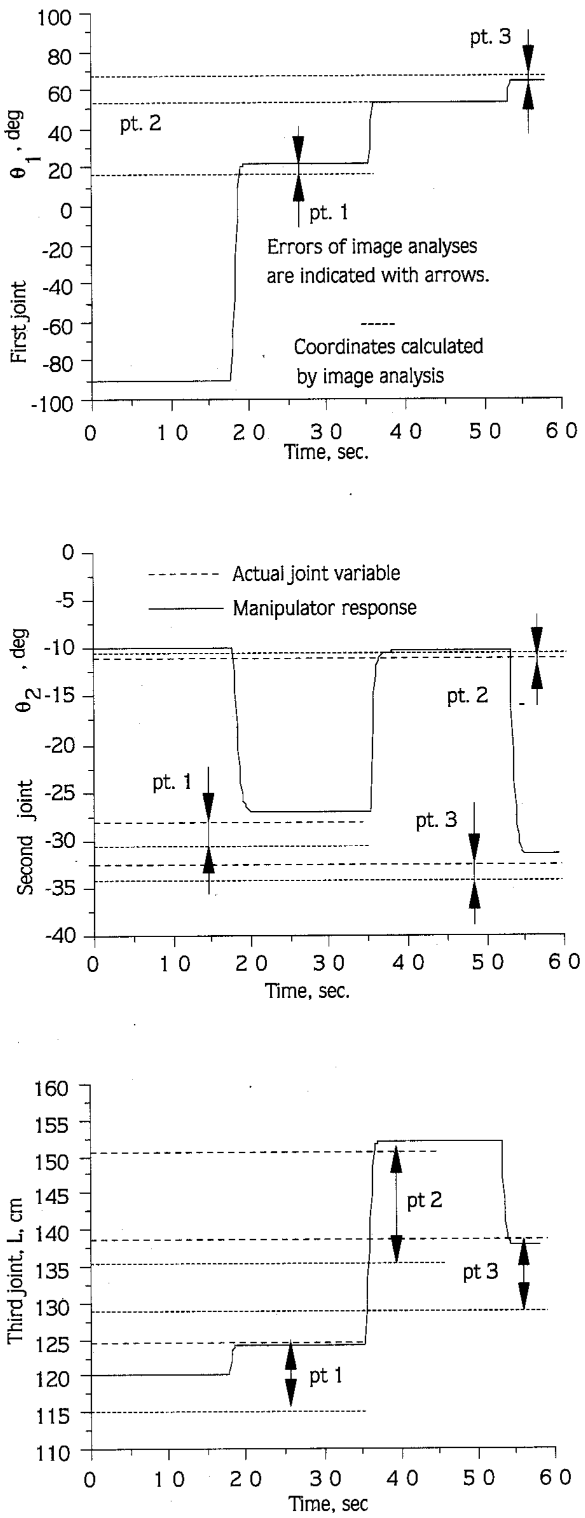

Figure 6. Response of the manipulator at joint coordinates in reaching 3 different points in succession. Image processing was performed only once for each point.

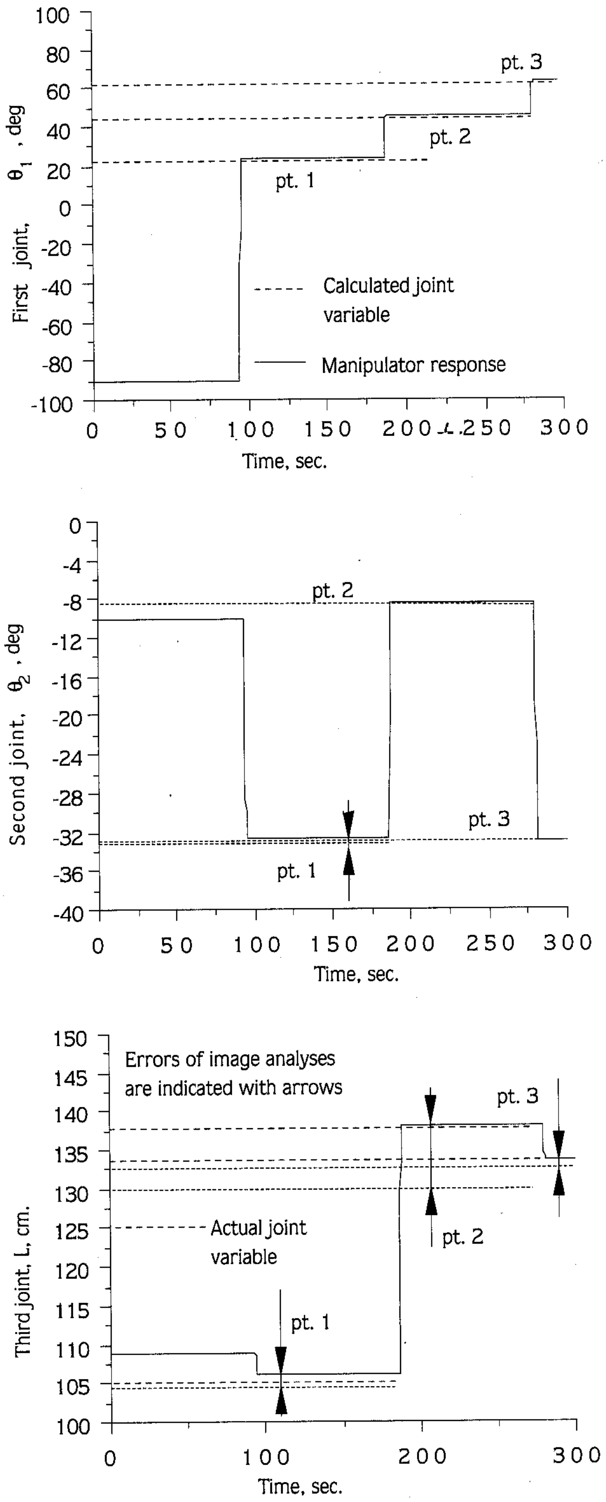

Figure 7. Response of the manipulator at joint coordinates in reaching 3 different points in succession. Image processing was performed for each point 10 times.

study, and by using *region growing* around the first detected pixel rather than opening a fixed-size window and checking all the pixels inside.

Acknowledgments

The material presented in this paper was obtained using an experimental rig purchased via a grant from the State Planning Organization of $ 67480, with grant number 91 K 120 830.

References

Brockelbank, D.C., Yang, Y.H.,"An experimental investigation into the use of color in computational stereopsis" IEEE Trans. Syst. Man. Cybern., Vol. 19, No.6, pp. 1365-1383, 1989.

Brown, D.C., "Close range camera calibration", Photogrammetry Eng. Vol 37, pp 855-866, 1971

Dhond, U.R., Aggarwal, J.K., "Structure from Stereo-A Review". IEEE Trans. Syst., Man. Cybern., Vol. 19, no 6, pp. 1489-1509, 1989.

Gonzales, R.C., Wintz, P., "Digital Image Processing", Addison-Wesley Publishing Company, Inc. pp. 190-200, 1987.

Grosky, W.I., Tamburino L.A., "A unified approach to the linear camera calibration problem", IEEE Trans. Patt. Anal. Mach. Intell. Vol 12, No 7, pp 663-671, 1990.

Healey, G.,"Color discrimination by computer", IEEE Trans. Syst. Man. Cybern., Vol. 19, No.6, pp. 1613-1617, 1989.

Horn, B.K.P., "Robot Vision", The MIT Press, USA, 1986.

Kapucu, S.,. " Adaptive Control of a Robot Manipulator by Visual Data", PhD Thesis, University of Gaziantep, Turkey, 1994.

Kim, Y.C., Aggarwal J.K., "Positioning 3-dimensional objects using stereo images", IEEE J. of Robotics and Automation, RA-3, No 4, pp. 361-373, 1987.

Lim, J.S., "Two-Dimensional Signal and Image Processing", Prentice Hall, Englewood Cliffs New Jersey, 1990.

Nevatia, R., "A color edge detector and its use in scene segmentation" IEEE Trans. Syst. Man. Cybern., Vol. SMC-7, No.11, pp. 820-826, 1977.

Roach, J.W., Aggarwal J.K., "Determining the movement of objects from a sequence of images", IEEE Trans. Patt. Anal. Mach. Intell. PAMI-2, pp 554-555, 1980.

Tsai, R.Y., "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", IEEE J. of Robotics and Automation, Vol RA3, No 4, 1987.

Weng, J., Cohen, P., Herniou, M., "Camera calibration with distortion models and accuracy evaluation", IEEE Trans. Patt. Anal. Mach. Intell. Vol 14, No 10, pp 965-980, 1992.

Wijesoma, S.W., Wolfe, D.F.H., Richards, R.J., "Eye to hand coordination for vision guided robot control applications", Int. J. of Robotic Res. Vol.12, no.1, pp 65-78, Feb. 1993.