Fundamental principles of robot vision

Conference Paper *in* Proceedings of SPIE - The International Society for Optical Engineering · August 1993

DOI: 10.1117/12.150210

Ernest L. Hall
University of Cincinnati

Fundamental principles of robot vision

Ernest L. Hall

Center for Robotics Research, University of Cincinnati
ML 72, Cincinnati, Ohio 45221-0072

ABSTRACT

Robot vision is a specialty of intelligent machines which describes the interaction between robotic manipulators and machine vision. Early robot vision systems were built to demonstrate that a robot with vision could adapt to changes in its environment. More recently attention is being directed toward machines with expanded adaptation and learning capabilities. The use of robot vision for automatic inspection and recognition of objects for manipulation by an industrial robot or for guidance of a mobile robot are two primary applications. Adaptation and learning characteristics are often lacking in industrial automation and if they can be added successfully, result in a more robust system. Due to a real time requirement, the robot vision methods that have proven most successful have been ones which could be reduced to a simple, fast computation. The purpose of this paper is to discuss some of the fundamental concepts in sufficient detail to provide a starting point for the interested engineer or scientist. A detailed example of a camera system viewing an object and for a simple, two dimensional robot vision system is presented. Finally, conclusions and recommendations for further study are presented.

1. INTRODUCTION

The use of visual information to provide knowledge and guidance for a robot manipulator is desirable to aid humans. Vision measurements also add capabilities for control since we can control only what we can measure. Visualization for recognition of targets can provide a basis for the guidance of robots and permit global navigation, collision avoidance or other adaption or learning techniques. Such vision guided robots have now been used for aiding in brain surgery, delivering food to patients, playing a piano, mowing the lawn, and handling hazardous materials. However, the implementations, rather than being straightforward applications of technology, have been extremely difficult. In general this has resulted in machines which were not robust and were limited in performance. This has also uncovered research problems on topics which are being intensively studied today. The search for a truly intelligent robot is continuing. Robotics and machine vision are considered totally separate disciplines by many. However, the proper combination of vision sensors to a manipulator under computer control can provide an intelligent machine which can perform work useful to humans. The combination of these three components can perform much more than the individual robot, vision system or computer, a Gestalt accomplishment. From the onset of the use of robots in the industrial arena, the problem of integrating these new machines with existing processes has proven difficult. Some problems have arisen because of a lack of fundamental theoretical knowledge about intelligent systems. Most robots have been used in applications where major intelligence was not a necessity and little modification was needed with existing automation. As new tasks become considered for robot utilization the development of "intelligent" devices has emerged and rekindled interest in expert system algorithms and the employment of multiple sensory systems.

A major advantage of automated manufacturing systems is their ability to accommodate a range of production rates between single products and high rate products. Groover[1] points out two categories of mass production: quantity production and flow production. In both categories, material processing steps requiring manipulation are important. In quantity production of a single item, bin-picking and bin-packing are often the least automated processes in the entire plant. In flow production, the "flow" is only maintained if raw materials are available in sufficient amounts and if sufficient transfer devices are available. Again many of the operations required are labor intensive, especially in the packaging operations. Areas of investigation under this theme include automated packaging, palletizing, bin-packing, bin-picking, automated storage and retrieval, automated kitting of parts for assembly, and automated warehousing. Improvements are needed in all areas of technology which are better, cheaper, faster and safer.

An excellent text on vision techniques for robots is [3]. A recent text on robot manipulators and control is [4]. A recent text covering the vision aspect of mobile robots is [3]. A recent research study on mobile robots is given by Roning [2].

2. ROBOT VISION

It is easy to see intelligence or the lack of it in humans, but trying to define it in terms of machines conjures up science fiction themes. At present, robots cannot display original thought leading to complex solutions. However, with the use of expert system algorithms, a robot can appear to "think" in a logical manner and actually solve difficult problems which appear to require intelligence. The use of intelligent robots which consist of specialized computer programs, robot manipulators, and sensors has now been demonstrated in a variety of situations. For example, various game playing systems such as robot checkers, peg games, etc, have been developed as educational demonstrations. Game playing represents a form of intelligence which is easily understood. A game usually has a set of rules defining a competition, including the permissible set of actions. The elements of a game are defined and easily understood, and the set of actions required are also defined. The game rules specify the computer and sensor problems. The set of actions defines the robot manipulator requirements. The accomplishment of a game by an automated machine, especially for the first time, is neither commonplace nor lacking in importance, because it demonstrates the capabilities of robots for real world applications.

Inntelligent robots can be traced to research at the Massachusetts Institute of Technology in the early 1950's [5]. Shannon in 1955 implemented a maze-following electronic mouse. Approximately 10 years later, a robot equipped with vision was utilized by Minsky to stack blocks. Further refinements of this idea have led to the development of the expert system for palletizing random size and weight parcels [6]. In 1970, modern research in mobile robots was conducted at Stanford University with the design of Shakey by Nitzan. Mobile robotics is now a popular and growing research area. Some recent examples of game playing robots may be found in a book on robotics including a robot solution to Rubik's cube [7] and a checker playing robot.

2.1.1 Search for intelligent machines

Robot "intelligence" has been demonstrated by a variety of games and applications. In a game, perhaps a necessary ingredient is "uselessness." However, in the applications, a necessary justification of the machine is its "usefulness." What is it about these machines that makes their design such a challenge in both situations? Perhaps the underlying mystery is that we are not cognizant of human intelligence, although we may feel confident in being able to "know it when we see it." A characteristic of all the games is that they can be described by astoundingly simple arithmetic or logical relations. The most challenging games have such an enormous state space that exhaustive enumeration of all pòssible solutions is not practical, say on the order of 2^{64}. Since we have computers with 2^{32} memory locations, it seems likely that in just a few years, the games with 2^{64} states may become deterministic and, thus, lose their mystery. Examples of such games are chess, checkers, and a 64 ring tower of Hanoi. Simulated randomness can always be inserted into game decisions to retain their amusement value, but the problems they represent will be theoretically solved. The same cannot be said about human intelligence except perhaps by psychologists or sociologists. The important point is that in attempting to build intelligent robots, we may come closer to understanding human intelligence. We can, at least, present some interesting challenges to currently accepted practices.

Adaptability is another feature of intelligence in man and machine. The most intelligent robot today can be simply unplugged. One minor malfunction can incapacitate the entire machine. Redundancy of function is at least a partial solution to this. The human arm has more degrees of freedom which may be used to move the hand to any point in three dimensional space with any of three possible orientations which requires only six. Redundant degree of freedom machines are now being built. Self diagnosis and maintenance are also being studied. To accomplish these diagnostic decisions, sensors of the state of the machine are needed for full adaptability, sensors of the state of nature are required. The work in robot vision, tactile sensors, and other sensors is progressing; however, the problem called "multisensor fusion" has also been recognized. If more than one sensor is used to determine a state, then it is possible to have incongruence of the sensor outputs. A human in space may encounter little or no input from the balance sensors in the ear, be touching the wall as indicated by tactile and kinesthetic sensors in an arm, and see the interior of the spacecraft upside down. Vision, since humans receive over 70% of their information through their eyes, would probably be believed; however, unless at least one other sensor becomes congruent with the visual, the human might still have a difficult time moving about. For example, even an earth-bound human cannot be mobile using sight alone. Even though we know sensors are essential to adaptability, they alone are not enough, and confusion can arise when the sensor inputs are not congruent. The expedient solution for incongruent sensors is to not move. Learning is another feature of intelligence and a great deal of recent interest in "neural networks" can be observed. Learning has classically been divided into two types: supervised learning with a teacher and unsupervised learning from experience.

2.1.2 Fundamental theorem of robot vision

One may reasonably question the connection of the words (robot, vision). Since each noun supports a discipline of its own. Of the many answers possible, let us consider one as a rather fundamental idea.

Fundamental theorem of robot vision

The manipulation of a point in space, x_1, by either a robot manipulator which moves it to another point, x_2, or through a camera system which images the point onto a camera sensor at x_2, is described by the same matrix transformation which is of the form:

$$x_2 = T\ x_1$$

The transformation matrix, T, can describe the first order effects of translation, rotation, scaling, projective and perspective projections.

This theorem suggests that the sensing and manipulation of a point or collection of points on an object has some relation. Now the question is how to exploit this relation to build an intelligent robot system. We can proceed in either of the two logical approaches that have served science so well in the past, deduction or induction.

2.2 Deductive approach

Proceeding by deduction to understand "intelligent machines" leads us into the study of human intelligence. Cognitive scientists study perception, cognition and action. Since human behavior is what we are attempting to understand, perhaps psychologists are our best teachers. This thought led us into a year long cross-disciplinary seminar course between our psychology and engineering colleagues. Everyone who attended this course learned a great deal. My impression at the end of the year was that human intelligence was tremendously superior to anything we have so far demonstrated in machine intelligence. The most promising technology seemed to be neural networks for learning and adaption. Nearly all the students from this course have gone on to study neural approaches but with the understanding that human behavior although exhibiting the highest forms of intelligence also from time to time exhibit actions that can only be described as "horrid." That is, even if we understood human intelligence, our problem of understanding intelligent machines would not be solved. We would still need work to make our robot obey Asimov's "Laws of Robotics." The inductive approach has led to a much greater appreciation of human intelligence and to the rise of neural networks as a major research area.

2.3 Inductive reasoning

The other logical approach is inductive reasoning. Here we attempt to understand specific intelligent machines in the hope of discovering the fundamental laws of robot vision. This approach is currently being followed around the world because even though the fundamental laws may not be discovered, something useful may be produced with the specific machine. If we can build a machine which accomplished "action X," we may has discovered a new, useful, and non-obvious solution to an old problem. The historical developments used the inductive approach.

2.4 An example - flatland robot

To provide a concrete example, let us consider a two dimensional, or flatland robot, in sufficient mathematical detail, to illustrate the three corners, machine vision, manipulators and controls, which box us in today's world of limited intelligent machines. Let us start by introducing a notation which lets us move objects about. We will concentrate on rigid body motion although the concepts can be easily extended. One of the simplest motions is a translation. Translation is a non-linear transformation. A linear transformation, $g = T(f)$, satisfies two properties.

1. $ag = T(af)$ (1)

2. $g_1 + g_2 = T(f_1 + f_2)$ if $g_1 = T(f_1)$, $g_2 = T(f_2)$ (2)

Let us try this with the translation: $x_2 = x_1 + h = T(x_1)$

1. $x_3 = ax_1 + h \neq ax_2$ (3)

2. $x_3 = (x_1 + x_2) + h \neq x_1 + h + x_2 + h$ (4)

Since neither of the conditions are satisfied, a dilemma is encountered at the beginning of our attempt to describe object motion mathematically. If we attempt a matrix description, we quickly see that translation cannot be represented by a 2 by 2 matrix operation.

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{5}$$

$$\begin{aligned} \text{or} \quad & x_1 = ax + by \\ & y_1 = cx + dy \\ \text{but} \quad & x_1 = x + h \\ & y_1 = y + k \end{aligned} \tag{6}$$

Now, translation may be represented by a 2 by 3 matrix operation.

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

However, since the matrix is not square, there is no inverse matrix. To alleviate these problems, we introduce the concept of homogeneous coordinates. The homogeneous coordinates of a two dimensional physical point are given by the three dimensional vector:

$$\begin{pmatrix} x \\ y \end{pmatrix} \ are \ \begin{pmatrix} wx \\ wy \\ w \end{pmatrix}. \tag{8}$$

Note that the conversion from homogeneous coordinates to physical coordinates simply requires division by w (called the scale term) and elimination of the third component. By using homogeneous coordinates, we homogenize the transformations (make them have the same structure). In homogeneous coordinates, translation is linear,

$$\begin{bmatrix} x_1 \\ y1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{9}$$

and the translation matrix has an easily computed inverse. If we move a point to the right and up with values (-h, -k), we can move it to the left and down with values (h, k).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -h \\ 0 & 1 & -k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{10}$$

The next common manipulation on an object is to rotate it. The rotation of a point clockwise or of the coordinate system counter clockwise may be described by:

$$\begin{aligned} x_2 &= x_1 \cos\theta + y_1 \sin\theta \\ y_2 &= -x_1 \sin\theta + y_1 \cos\theta \end{aligned} \tag{11}$$

Rotation is a linear transformation and may be represented by a 2 by 2 matrix operation:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \tag{12}$$

The homogeneous coordinate representation is:

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{13}$$

The inverse is easily computed. Rotations are orthonormal transformations.

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{14}$$

Scaling operations such as magnification, minification, shearing, etc. may also be represented by matrix operations.

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \tag{15}$$

If $s_1 > 1$, magnification occurs; if $s_1 < 1$, minification occurs; is s_1 is not equal to s_2, shearing occurs. Scaling may also be accomplished using the w term in homogeneous coordinates.

$$\begin{bmatrix} x_2 \\ y_2 \\ s \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{16}$$

In converting these homogeneous coordinates to physical coordinates an inverse scaling is accomplished.

$$\begin{bmatrix} x_{2/s} \\ y_{2/s} \end{bmatrix} \tag{17}$$

The non-linear perspective transformation induced by a camera system can also be described by a linear transformation in homogeneous coordinates. The transformation from an object located at $(x_o, y_o)^T$, to an image at $(0, y_i)^T$ through a lens with center on the x axis at location x = f, is often presented through the use of similar triangles.

$$y_i = y_o \left(\frac{f}{f-x_o} \right) \tag{18}$$

In homogeneous coordinates, the perspective transformation is:

$$\begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{f} & 0 & 1 \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ 1 \end{bmatrix} \tag{19}$$

The reduction in dimensionality caused by projecting an object point onto an image plane may also be represented by a linear transformation in homogeneous coordinates. For projection onto the y axis or x = 0 plane, the x coordinate is essentially discarded by the following transformation.

$$\begin{bmatrix} 0 \\ y_i \\ w_i \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ w_1 \end{bmatrix} \tag{20}$$

Note that this matrix is singular and therefore has no inverse. Recovering this lost information is the goal of many three dimensional techniques such as stereo and shape from shading. Let us now consider a specific example which not only shows the use of the transformation matrices but also shows that the extension to three dimensions is relatively easy.

Example 1. Perspective transformation

Consider a unit cube in three dimensions. It is described by the vertex coordinates:

Vertex	x	y	z
a	0	0	0
b	0	0	1
c	0	1	0
d	0	1	1
e	1	0	0
f	1	0	1
g	1	1	0
h	1	1	1

(21)

Let us determine the perspective, projective transformation on the plane defined by z = 4 with the camera lens centered at $(1/2, 1/2, 2)^T$. To determine the transformed image, we will first translate the coordinate system to one with the lens center on the z axis. Considering the original object coordinate system as the global coordinates, we may transform to lens centered coordinates by the following translation of the coordinate system.

$$\begin{bmatrix} x_l \\ y_l \\ z_l \\ w_l \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -0.5 \\ 0 & 1 & 0 & -0.5 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_g \\ y_g \\ z_g \\ w_g \end{bmatrix} \tag{22}$$

This transformation moves the global origin to the point $(-0.5, -0.5, -4)^T$. The perspective transformation is now simplified since the optical axis is along the z axis and the lens center is located at z = -2. The perspective transformation may now be written as:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ w_c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ w_l \end{bmatrix} \tag{23}$$

This transformation produces the scaling relation:

$$w_c = \frac{z_l}{2} + w_l \tag{24}$$

The projection of the transformed object points onto the image plane, $z_i = 0$, is accomplished with the following transformation.

$$\begin{bmatrix} x_z \\ y_z \\ 0 \\ w_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ w_c \end{bmatrix} \tag{25}$$

The product of these matrices may now be computed to simplify the computation. The resulting matrix is:

$$\begin{bmatrix} x_z \\ y_z \\ 0 \\ w_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -0.5 \\ 0 & 1 & 0 & -0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ w_c \end{bmatrix} \quad (26)$$

The original object vertex points in homogeneous coordinates are:

$$\begin{bmatrix} x_g \\ y_g \\ z_g \\ w_g \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (27)$$

The new image coordinates are given by the following.

$$\begin{bmatrix} x_z \\ y_z \\ 0 \\ w_z \end{bmatrix} = \begin{bmatrix} -0.5 & -0.5 & -0.5 & -0.5 & -0.5 & -0.5 & -0.5 & -0.5 \\ -0.5 & -0.5 & 0.5 & 0.5 & -0.5 & -0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -0.5 & -1 & -0.5 & -1 & -0.5 & -1 & -0.5 \end{bmatrix} \quad (28)$$

The physical image coordinates determined with respect to the image coordinate system are as follows.

Vertex	x	y
a'	0.5	0.5
b'	1	1
c'	0.5	-0.5
d'	1	-1
e'	-0.5	0.5
f'	-1	1
g'	-0.5	-0.5
h'	-1	-1

(29)

These image points would be recorded by an image sensor. Several vision interpretation problems can now be posed. How does one recognize the object from the image data? How does one determine the centroid position of the object? How does one determine the orientation of the object? How does one recover the three dimensional coordinates of the object vertices? For completeness we may also compute the image vertex points with respect to the global coordinate system.

Vertex	x	y	z
a"	1	1	4
b"	1.5	1.5	4
c"	1	0	4
d"	1.5	-0.5	4
e"	0	1	4
f"	-0.5	1.5	4
g"	0	0	4
h"	-0.5	-0.5	4

(30)

If the camera position and orientation are known, then absolute object coordinates may be determined. These coordinates may be used to guide the motion of a robot.

Example 2. The flatland robot

To illustrate that the matrix transformations are also useful for robot manipulation, let us consider a two dimensional example with a manipulator called the flatland robot. This example will let us see the concepts while avoiding many of the complexities of robots with more degrees of freedom. We will consider four related problems associated with this robot. The first is simply to describe where the robot is in space, the kinematics, given the parameters of the robot and the Cartesian coordinates of the point. The kinematic descriptions involve joint angle variables for rotary joints or linear variables for prismatic joints. The flatland robot has rotational variables. The Jacobian matrix which relates small changes between joint and Cartesian spaces is also important to relate linear and angular velocities in the two spaces. The second problem is called the inverse kinematic solution and involves determining the inverse transformation from Cartesian to joint space. The next two problems deal with explaining rather than describing the motion of the robot. The dynamic equations include both static and dynamic forces and torques and relate the joint accelerations to the torques which produce the motion. The inverse dynamic equations let us determine the accelerations which will be produced by given torques. The dynamic equations also indicate the characteristics of the system which are needed for automatic control of the robot. Again, we need a notation to permit a mathematical description of the motion. One important concept is to impose, at each joint of the robot, a coordinate system or frame. For our example, we place one at the base, another at the elbow and another at the hand. Consider a point, $\mathbf{P} = (x_1, y_1)^T$, defined with respect to the hand coordinate frame. We would like to develop the transformations to describe this same point with respect to the base coordinate frame. This may be accomplished by a series of matrix transformation which correspond to the translations and rotations required to move the coordinate system from the wrist to the base. We will again use homogeneous coordinates since they permit the same size matrices to be used for all the motions.

The point is described in homogeneous coordinates by:

$$P = \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{31}$$

The transformations required to describe this point in base or global coordinates are simply:

$$\mathbf{x} = \mathbf{A}_4\ \mathbf{A}_3\ \mathbf{A}_2\ \mathbf{A}_1\ \mathbf{P} = \mathbf{T}\ \mathbf{P} \tag{32}$$

where the individual matrices are defined by the following relationships:

$$A_1 : \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ell_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{33}$$

$$A_2 : \begin{bmatrix} x_3 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta_2 & \sin\theta_2 & 0 \\ -\sin\theta_2 & \cos\theta_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \tag{34}$$

$$A_3 : \begin{bmatrix} x_4 \\ y_4 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ell_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ y_3 \\ 1 \end{bmatrix} \tag{35}$$

$$A_4 : \begin{bmatrix} x_5 \\ y_5 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_4 \\ y_4 \\ 1 \end{bmatrix} \tag{36}$$

At this point we must compute the matrix products and the equations become rather long. For simplicity let us introduce the notation, si = $\sin(\Theta_i)$ and ci = $\cos(\Theta_i)$. The matrix products can be computed in a variety of ways. Here are some examples:

$$(37) \quad A_2A_1 = \begin{bmatrix} c2 & -s2 & \ell_2 c2 \\ s2 & c2 & \ell_2\, s2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A_4A_3 = \begin{bmatrix} c1 & -s1 & \ell_1 c1 \\ s1 & c1 & \ell_1\, s1 \\ 0 & 0 & 1 \end{bmatrix} \tag{38}$$

and

$$T = A_4A_3A_2A_1 = \begin{bmatrix} c1c2-s1s2 & -c1s2-s1c2 & l_2c1c2-l_2s1s2+l_1c1 \\ s1c2+c1s2 & -s1s2+c1c2 & l_2\; s1c2+l_2c1s2+l_1s1 \\ 0 & 0 & 1 \end{bmatrix} \tag{39}$$

Although this equation looks quite complicated, there is a special case in which it simplifies. The location of the tip of the arm is at the origin of the hand coordinate system, $(x_1, y_1)^T = (0,0)^T$. At this point the matrix transformation simplifies to:

$$\begin{aligned} x &= \ell_2 c1c2 - \ell_2 s1s2 + \ell_1 c1 \\ y &= \ell_2 s1c2 + \ell_2 c1s2 + \ell_1 s1 \end{aligned} \tag{40}$$

This solution may be verified from a geometrical viewpoint. The generalization of this techniques is called the Denavit-Hartenberg notation and is easily applied to any multi-link manipulator (open kinematic chain). Programs are also available to perform symbolic manipulation of the matrices.

Inverse kinematics

A derivation of the inverse kinematics is much more difficult than the forward kinematic equations. Examples of solutions and an expert system program is described in [8]. When one realizes that the inverse dynamic solution is essential before the construction of an industrial robot, it shows the importance of these solutions. For the flatland robot the solution is given below.

$$\cos\,\theta_2 = \frac{x^2 + y^2 - \ell_1^2 - \ell_2^2}{2\,\ell_1\,\ell_2} \tag{41}$$

There are two solutions for Θ_2 which are of equal magnitude and opposite signs corresponding to the two arm positions which can reach the same point.

$$\tan\,\theta_1 = \frac{\dfrac{y}{x} - \dfrac{\ell_2 \sin\,\theta_2}{(\ell_1 + \ell_2 \cos\,\theta_2)}}{1 + \left(\dfrac{y}{x}\right)\left(\dfrac{\ell_2 \sin\,\theta_2}{\ell_1 + \ell_2 \cos\,\theta_2}\right)} \tag{42}$$

The inverse kinematic equations may be used to determine the joint angles required to position the robot at a given Cartesian space point. They may also be used to determine practical facts such as the workspace of the robot or to simulate its motion. The kinematic equations may also be written as:

$$\begin{aligned} x &= \ell_1 \cos\,\theta_1 + \ell_2 \cos\,(\theta_1 + \theta_2) \\ y &= \ell_1 \sin\,\theta_1 + \ell_2 \sin\,(\theta_1 + \theta_2) \end{aligned} \tag{43}$$

$$or \;\; \underline{dx} = \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} \dfrac{\partial x}{\partial\theta_1} & \dfrac{\partial x}{\partial\theta_2} \\ \dfrac{\partial y}{\partial\theta_1} & \dfrac{\partial y}{\partial\theta_2} \end{bmatrix} \begin{bmatrix} d\theta_1 \\ d\theta_2 \end{bmatrix} = \underline{J}\;\underline{d\theta} \tag{44}$$

Since we have an explicit non-linear relation between the Cartesian and joint variables, we may easily compute the Jacobian, **J**, which describes the infinitismal relationships between the two spaces.

$$where \;\; \underline{J} = \begin{bmatrix} -\ell_1 \sin\,\theta_1 - \ell_2 \sin\,(\theta_1 + \theta_2) & -\ell_2 \sin\,(\theta_1 + \theta_2) \\ \ell_1 \cos\,\theta_1 + \ell_2 \cos\,(\theta_1 + \theta_2) & \ell_2 \cos\,(\theta_1 + \theta_2) \end{bmatrix} \tag{45}$$

The Jacobian may be used to relate velocities and accelerations in the two spaces or for linear approximations about an operating point. The robot is more than a position generator. It may also exert a force on an object. If the force, F, exerted by the tip is:

$$F = \begin{bmatrix} u \\ v \end{bmatrix} \tag{46}$$

Then the joint torques, **T**, required to generate this force may be determined. Let us first determine this force when the robot is static. By forming force and torque balances for each link, the following relationship may be determined:

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} \ell_1\cos(\theta_1) + \ell_2\cos(\theta_1 + \theta_2) & -(\ell_1\sin(\theta_1) + \ell_2\sin(\theta_1 + \theta_2)) \\ \ell_2\cos(\theta_1 + \theta_2) & -\ell_2\sin(\theta_1 + \theta_2) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{47}$$

The static forces and torque balance equations may be easily inverted to obtain the following relationships which permit one to calculate the joint forces generated by given torques.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \ell_2\cos(\theta_1+\theta_2) & -(\ell_1\cos(\theta_1+\ell_2\cos(\theta_1+\theta_2)) \\ \ell_2\sin(\theta_1+\theta_2) & -(\ell_1\sin(\theta 1)+\ell_2\sin(\theta_1+\theta_2)) \end{bmatrix} \begin{bmatrix} T_1/\Delta \\ T_2/\Delta \end{bmatrix} \tag{48}$$

$$\Delta = \ell_1\ell_2\sin(\theta_2) \tag{49}$$

When $\sin\theta_2 = 0$, the links are parallel, and the joint torques have no control over the force component along the length of the links. When the robot moves, we must use the dynamic equations. To determine the dynamic equation we may use the Lagrangian, L, or kinetic potential which is equal to the difference between kinetic and potential energy. For each degree of freedom, the generalized momentum, p_i, can be expressed as:

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \tag{50}$$

where q_i is a generalized coordinate. The generalized force, Q_i, is given by:

$$\frac{d}{dt}\left[\frac{\partial L}{\partial q_i}\right] - \frac{\partial L}{\partial q_i} = Q_i \tag{51}$$

For the special case of unity link lengths, the dynamic equations are:

$$\begin{bmatrix} 2\left[\frac{5}{3}+\cos(\theta_2)\right] & \left(\frac{2}{3}+\cos(\theta_2)\right) \\ \left(\frac{2}{3}+\cos(\theta_2)\right) & \left(2\left(\frac{1}{3}\right)\right) \end{bmatrix} \begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{bmatrix} = \begin{bmatrix} T_1/\mathrm{E}+\sin(\theta_2)\dot{\theta}_2(2\dot{\theta}_1+\dot{\theta}_2) \\ T_2/\mathrm{E}-\sin(\theta_2)\dot{\theta}^2{}_1 \end{bmatrix} \tag{52}$$

The dynamic equations permit us to determine joint torques given the arm state. The state variables are:

$$\{\theta_1\ \dot{\theta}_1\ \ddot{\theta}_1\ \theta_2\ \dot{\theta}_2\ \ddot{\theta}_2\} \tag{53}$$

We could also use this dynamic equation for a state space description for a control system for the robot.

$$\mathrm{E} = \frac{1}{2}m\ell^2 \tag{54}$$

This brief example may serve to illustrate some of the problems involved in intelligent robot design.

3. CONCLUSIONS

A brief introduction to robot vision has been presented to encourage a new investigations of fundamental challenges in intelligent machines. A mathematical connection between robot manipulation and visual transformations is presented along with detailed examples which illustrate the use of the transformations for both vision and manipulation. Both topics are needed to master vision control, hand-eye coordination or more complex applications.

REFERENCES

1. Groover, M., **Automation, Production Systems, and Computer-Aided Manufacturing**, pp. 22-23, Prentice-Hall, Inc., New Jersey (1980).

2. Klaus, B. and Horn, B., **Robot Vision**, The MIT Press, Massachusetts (1980).

3. Ayache, N., **Artificial Vision for Mobile Robots**, MIT Press, Cambridge, 1991.

4. Lewis, F. L., C. T. Abdallah and D. M. Dawson, **Control of Robot Manipulators**, Macmillan, New York, 1993.

5. Roning, J., "Model-Based Visual Navigation of a Mobile Robot," Ph.D. Dissertation, Department of Electrical Engineering, University of Oulu, Finland, 1992.

6. Hall, E. and Hall, B., **Robotics: A User-Friendly Introduction**, Holt, Rhinhart, and Winston, New York (1985).

7. Mazouz, K., Hall E., and Shell, R., "Expert System for Flexible Palletizing of Mixed Size and Weight Parcels," Proceedings, Society of Photo-Optical Instrumentation Engineers, Intelligent Robots and Computer Vision Conference, Cambridge, MA, November 6-10, 1987.

8. Mu, E., Cain, J.T., and Herrera-Bendezu, L.G., "Computer Generation of Symbolic Kinematics for Robot Manipulators," **Journal of Robotic Systems**, 9(3), 385-410, 1992.