Chapter 18

Application of Stereo Vision and ARM Processor for Motion Control

The main idea of this chapter is to present the application of vision-based approach via stereo vision to determine position and orientation of mechanical manipulators. Stereo vision is a technique based on inferring the depth of an object with the aim of two cameras. In this case, a stereo vision setup that is made up of two parallel complementary metal oxide semiconductor (CMOS) cameras is provided and Universal Serial Bus (USB) port is used to transfer image data to a computer. Experimental tests are exerted for 6R mechanical manipulator in point-to-point motion. The image processing algorithm, matching algorithm and triangulation calculations are optimized to increase the speed of computations that are required for the matching procedure. Because of the rapid image processing algorithm, the proposed method can be used to detect the end-effector in dynamic trajectories. This stereo vision setup can be applied easily as a measurement system in closed-loop control systems or other applications.

18.1. Introduction

Determining the real end-effector position for mechanical manipulators is an important key point in industrial applications. This problem can be

Chapter written by Moharam HABIBNEJAD KORAYEM, Michal IRANI and Saeed RAFEE NEKOO.

considered as the process of determining the position and orientation of the end-effector of robotic manipulators with respect to a global frame. Various methods and solutions have been proposed and applied to this problem in robotic areas recently. The possible easiest solution to this problem by is doing relative position/orientation measurements using sensors such as odometer or an inertia measurement unit (IMU). A relative position measurement known as dead reckoning is the process of tracking the current position of the end-effector based on the integration of the path that has been previously traveled by the end-effector [HEE 07]. The most common sensors are IMU which is made of accelerometers and gyroscopic rate sensors. IMUs have the advantage of a high frame rate and relatively small latency, but only provide relative motion and they are subject to significant drift (or significant cost). IMU sensors are often complemented by global positioning system (GPS) sensors [JON 10]. In dead reckoning, odometry is probably the most well-known method that is used in mobile robots, as it provides easy and cheap real-time position information by the integration of incremental motion information over time. Unfortunately, this integration causes an accumulation of errors during the movement of the robot [GUT 08].

Another approach for an end-effector position measurement is by using active beacons. Measurement using active beacons has traditionally been used in the GPS for the localization of ships and airplanes [HEE 07]. There are two types of active beacon systems: trilateration and triangulation [KUC 06, TSU 09, SAN 09]. Trilateration is the determination of the end-effector position based on distance measurements to known beacon sources. In trilateration systems, there are usually three or more transmitters mounted at known locations in the workspace and one receiver fixed onto the end-effector. An example of the active beacons localization system that makes use of trilateration is given in [MOO 04].

Triangulation is another way of determination of the robot end-effector position that is based on the measurement of the angle from the beacon to the end-effector or robot heading. The distance of at least one of the beacons and its location must be known. Similar to the trilateration method, three or more beacon readings must be obtained to do triangulation. Note that it is also possible to do triangulation with two beacons if the angles from these beacons to the end-effector or robot heading, the distances from these beacons to the robot and the locations of these beacons are known [EVO]. An example of the triangulation with two beacons is the North Star kit [EVO]. Unlike dead reckoning, the errors in active beacons systems will not grow unbounded; however, accuracy is highly dependent on the size of its random errors and precise position of the beacons in the environment [HEE 07].

For compensating the disadvantages of relative measurement methods, sensors that provide measurements of absolute position are extremely important. These kind of methods are named absolute position measurements (reference-based systems). Magnetic compasses and vision systems [RAM 08, CHA 10, SPA 07] are two approaches of this category. Chang *et al.* [CHA 10] presented a vision-based navigation and localization system using two biologically inspired scene understanding models that are studied from human visual capabilities' point of view: (1) a Gist model that captures the holistic characteristics and layout of an image and (2) a saliency model that emulates the visual attention of primates to identify conspicuous regions in the image.

The main idea of this chapter is to use the stereo vision approach and propose a simple triangulation approach for determining position and orientation using a simple and inexpensive setup for mechanical manipulators. On the basis of stereo vision, a dynamical test is also implemented for a 6R manipulator to show the accuracy and speed of computations of the proposed approach. Stereo vision represents a way for reconstructing three-dimensional (3D) information from the surrounding environment and it is important for a large number of vision applications [BLE 08]. Unfortunately, the key step in stereo vision, that is the stereo matching problem, cannot be regarded as solved. Other factors that complicate the matching process are image noise, untextured regions and the occlusion problem [BLE 08]. Two methods are used for the matching process: local and global methods [HUM 10, YAN 06].

Stan and Carlo [STA 99] proposed an algorithm to detect depth discontinuities from a stereo pair of images. The algorithm matched individual pixels in corresponding scan line pairs while allowing occluded pixels to remain unmatched, and then propagated the information between scan lines by means of a fast post-processor. The algorithm handled large untextured regions, used the measurement of pixel dissimilarity that was insensitive to image sampling and pruned bad search nodes to increase the speed of dynamic programming. The computation was relatively fast, taking about 1.5 μ s per pixel per disparity on a workstation. Approximate disparity maps and precise depth discontinuities (along both horizontal and vertical boundaries) were shown for five stereo images.

18.2. Stereo vision

The three-dimensional measurement approach based on the human vision system is named stereo vision, but in this vision approach, human eyes are replaced with a pair of slightly displaced parallel cameras. Two-dimensional position information of an object is determined using one camera and a third coordinate can be computed by comparing two images by means of a technique based on stereo vision approach. This technique helps to infer the depth of an object from two cameras.

Using stereo vision, with two images, we can infer depth by means of the triangulation method if it is available to find corresponding (homologous) points in the two images [VIS]. According to Figure 18.1, the amount to which a single pixel is displaced in the two images is called disparity, so a pixel's disparity is inversely proportional to its depth in the scene. Before computing disparity, a rectification process must be applied to remove lens distortions and turn the stereo pair in to standard form.



Figure 18.1. Relation between depth and disparity [DAS 10]

The challenging part is to compute the disparity of each pixel and this task is known as the stereo matching problem. Two methods are used for the matching process: local and global methods and local and global/semi-global methods. Local algorithms use the simple winner-takes-all (WTA) disparity selection strategy but reduce the ambiguity (increasing the signal-to-noise

(SNR) ratio) by aggregating matching costs over a support window (aka kernel or correlation window). Global (and semi-global) algorithms search for disparity assignments that minimize an energy function over the whole stereo pair using a pixel-based matching cost (sometimes the matching cost is aggregated over a support) [VIS].

The Middlebury stereo evaluation site provides a framework and a data set for benchmarking novel algorithms [SCH]. An overview of the stereo vision approach is shown in Figure 18.2.



Figure 18.2. Overview of stereo vision approach

18.3. Triangulation

Given the disparity of any pixel, triangulation computes the position of the correspondence in the 3D space. Consider Figure 18.3 that shows two images, an object, the image frame (local) and the reference frame. Parameters that are shown in this figure are:

- OXYZ: reference frame;
- $-O_1x_1y_1z_1$: image frame of first camera;
- $-O_2x_2y_2z_2$: image frame of second camera;
- $-F_1, F_2$: focal lengths;
- $-\beta_1,\beta_2$: rotation angles of optical axis;
- $-u_1, v_1$: pixel coordinate in first image;
- $-u_2$, v_2 : pixel coordinate in second image.



Figure 18.3. Coordinates in stereo vision [DAS 10]

Transformation between the object position in the reference frame and each local frame can be stated as:

$$M^{T} = RTP^{T}$$
[18.1]

where P and M are the local and the global positions of the object and have the following forms:

$$P = \begin{bmatrix} x, y, z, 1 \end{bmatrix}$$
[18.2]

$$M = \begin{bmatrix} X, Y, Z, 1 \end{bmatrix}$$
[18.3]

and *R* and *T* matrices in equation [18.1] are rotation and translation matrixes, respectively, and have the following forms:

$$R = \begin{bmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
[18.4]

$$T = \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & Y_0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
[18.5]

In these equations, β is the rotation of an optical axis of the considered camera and $[X_0, Y_0, Z_0]$ are the position of the center of the local frame in the reference frame. Consider *F* as focal length and as a result, perspective relations for any image can be written as:

$$\frac{x}{u} = -\frac{z-F}{F}; \qquad \frac{y}{v} = -\frac{z-F}{F}$$
 [18.6]

Substituting equation [18.6] into equation [18.1], for two images, the 3D position of an object in the reference frame can be computed as:

$$X = \frac{u_1 F_1 - m_1 X_{01} - n_1 (Z - Z_{01})}{m_1}$$

$$Y = v_1 - Y_{01} - \frac{v_1}{F_1} \Big[(X - X_{01}) \sin \beta_1 + (Z - Z_{01}) \cos \beta_1 \Big]$$
[18.7]
$$Z = \frac{m_1 m_2 (X - X_{01}) + m_1 F_2 u_2 - m_2 F_1 u_1 + m_2 n_1 Z_{01} - m_1 n_2 Z_{02}}{m_1 n_2 - m_2 n_1}$$

in which

$$m_{1} = F_{1} \cos \beta_{1} + u_{1} \sin \beta_{1}$$

$$m_{2} = F_{2} \cos \beta_{2} - u_{2} \sin \beta_{2}$$

$$n_{1} = u_{1} \cos \beta_{1} - F_{1} \sin \beta_{1}$$

$$n_{2} = u_{2} \cos \beta_{2} + F_{2} \sin \beta_{2}$$
[18.8]

where $[u_i, v_i]$ are the coordinates of the projection of the object into *i*th image. If two cameras could be parallel with $\beta = 0$, then the following simple relations can be considered:

$$Y_{01} = Y_{02} = 0$$

$$\beta_1 = \beta_2 = 0$$

$$F_1 = F_2 = F$$

$$X_{01} = X_{02} = \frac{B}{2}$$

[18.9]

Finally, equation [18.7] can be rewritten as:

$$X = \frac{F - Z}{F} u_1 - \frac{B}{2}$$

$$Y = \frac{F - Z}{F} v_1$$

$$Z = F - \frac{BF}{u_2 - u_2}$$
[18.10]

In equation [18.10], $[u_1, u_2]$ are the differences between the X coordinate of two corresponding pixels that were named disparity. The disparity of an object that is in infinity is equal to zero. According to the above equations, parameters that affect the accuracy of measurements are:

- 1) accuracy of stereo correspondence;
- 2) distance between cameras and end-effector;
- 3) initial calibration.

18.4. End-effector orientation

According to the triangulation method and using the stereo vision setup, the position of the end-effector and coordinates can be calculated, which are shown in Figure 18.4 as P_1 , P_2 and P_3 . The direction vectors \vec{x}_b , \vec{y}_b and \vec{z}_b can be represented as:

$$\vec{x}_b = \frac{P_2 - P_1}{\|P_2 - P_1\|}$$
[18.11]

$$\vec{z}_{b} = \frac{(P_{2} + P_{1})/2 - P_{3}}{\left\| (P_{2} + P_{1})/2 - P_{3} \right\|}$$
[18.12]

$$\vec{y}_b = \vec{z}_b \times \vec{x}_b \tag{[18.13]}$$

Considering the rotation matrices $R_{Z,\phi}$, $R_{Y,\theta}$ and $R_{X,\psi}$ for roll, pitch and yaw angles, the rotation matrix between reference and body frames can be mentioned as:



Figure 18.4. Three points on end-effector

$$\begin{bmatrix} \vec{x}_{b} & \vec{y}_{b} & \vec{z}_{b} \end{bmatrix} = \begin{bmatrix} X_{x} & Y_{x} & Z_{x} \\ X_{y} & Y_{y} & Z_{y} \\ X_{z} & Y_{z} & Z_{z} \end{bmatrix}$$

$$\begin{bmatrix} X_{x} & Y_{x} & Z_{x} \\ X_{y} & Y_{y} & Z_{y} \\ X_{z} & Y_{z} & Z_{z} \end{bmatrix} = \begin{bmatrix} C_{\phi}C_{\theta} & C_{\phi}S_{\theta}S_{\psi} - S_{\phi}C_{\psi} & C_{\phi}S_{\theta}S_{\psi} + S_{\phi}S_{\psi} \\ S_{\phi}C_{\theta} & S_{\phi}S_{\theta}S_{\psi} - C_{\phi}C_{\psi} & S_{\phi}S_{\theta}C_{\psi} - C_{\phi}S_{\psi} \\ -S_{\theta} & C_{\theta}S_{\psi} & C_{\theta}C_{\psi} \end{bmatrix}$$
[18.14]

where $C_i = \cos(i)$ and $S_i = \sin(i)$. Therefore, the roll, pitch and yaw angles of the robot end-effector are determined as follows:

$$\phi = \arctan\left(\frac{X_y}{X_x}\right)$$

$$\theta = \arctan\left(\frac{-X_z}{\sqrt{I - X_z^2}}\right)$$

$$\Psi = \arctan\left(\frac{Y_z}{Z_z}\right)$$
[18.16]

Figure 18.3 may seem complicated, but with the above derivations the coordinates can be computed easily. One of the advantages of using measurements based on vision in different systems is the simple equations that determine the position of an object.

18.5. Experimental setup and results

To show the application of the vision system, an experimental test is conducted on a 6R manipulator. The robot arm is shown in Figure 18.5. This robot arm is an articulated six degree of freedom (DoF) arm that has the base of a programmable universal machine for assembly (PUMA) structure and a three DoF end-effector. The six DoFs of the 6R provide the motion in a 3D space with the path and the orientation of the gripper.



Figure 18.5. 6R robot arm

The schematic of the manipulator is shown in Figure 18.6. The coordinate frames on each joint are set based on the Denavit-Hartenberg algorithm to prepare the transformation matrices for calculation of direct and inverse kinematics. A distance is considered between fifth and sixth links for better presentation, which is in fact not the reality. The Denavit-Hartenberg parameters of the arm are presented in Table 18.1.

ARM Processor for Motion Control 493



Figure 18.6. Schematic of 6R robot arm

Joint	a _i (mm)	d _{<i>i</i>} (mm)	α_i^{o}	Motion	Home position
1	36.5	438	-90	Link 1	0
2	251.5	0	0	Link 2	0
3	125	0	0	Link 3	0
4	92	0	90	Yaw	0
5	0	0	-90	Pitch	-90
6	0	152.8	0	Roll	0

 Table 18.1. Denavit-Hartenberg of 6R robot arm

Transformation matrix, T, is used for forward kinematic computations based on multiplying six transformations coordinates to each other.

$$T = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
[18.17]

in which the elements of *T* are:

$$\begin{split} n_x &= -c_6 s_1 s_5 + c_1 (c_{234} c_5 c_6 - s_{234} s_6) \\ n_y &= c_{234} c_5 c_6 s_1 + c_1 c_6 s_5 - s_1 s_{234} s_6 \\ n_z &= -c_5 c_6 s_{234} - c_{234} s_6 \\ o_x &= s_1 s_5 s_6 - c_1 (c_6 s_{234} + c_{234} c_5 s_6) \\ o_y &= -c_6 s_1 s_{234} - (c_{234} c_5 s_1 + c_1 s_5) s_6 \\ o_z &= -c_{234} c_6 + c_5 s_{234} s_6 \\ a_x &= -c_5 s_1 - c_1 c_{234} s_5 \\ a_y &= c_1 c_5 - c_{234} s_1 s_5 \\ a_z &= s_{234} s_5 \\ p_x &= -d_6 c_5 s_1 + c_1 (a_1 + a_2 c_2 + a_3 c_{23} + c_{234} (a_4 - d_6 s_5)) \\ p_y &= d_6 c_1 c_5 + s_1 (a_1 + a_2 c_2 + a_3 c_{23} + c_{234} (a_4 - d_6 s_5)) \\ p_z &= d_1 - a_2 s_2 - a_3 s_{23} + s_{234} (-a_4 + d_6 s_5) \end{split}$$

In these equations, a_i and d_i are shown in Figure 18.6, also s_i , c_i , s_{ij} and c_{ij} denote $\sin(\theta_i)$, $\cos(\theta_i)$, $\sin(\theta_i + \theta_j)$ and $\cos(\theta_i + \theta_j)$, respectively. P_x , P_y and P_z are the end-effector position of 6R. The relation between velocity of end-effector and angular velocity of joints is expressed by:

$$V = J\dot{q}$$
[18.19]

J is the Jacobian matrix of the 6R robot and can be obtained as:

$$J = \begin{bmatrix} j_{11} & j_{12} & j_{13} & j_{14} & j_{15} & 0\\ j_{21} & j_{22} & j_{23} & j_{24} & j_{25} & 0\\ 0 & j_{32} & j_{33} & j_{34} & d_6c_5s_{234} & 0\\ 0 & -s_1 & -s_1 & -s_1 & c_1s_{234} & j_{46}\\ 0 & c_1 & c_1 & c_1 & s_1s_{234} & j_{56}\\ 1 & 0 & 0 & 0 & c_{234} & s_{234}s_5 \end{bmatrix}$$
[18.20]

$$\begin{split} j_{11} &= -d_6c_1c_5 - s_1(a_1 + a_2c_2 + a_3c_{23} + c_{234}(a_4 - d_6s_5)) \\ j_{12} &= -c_1(a_2s_2 + a_3s_{23} + s_{234}(a_4 - d_6s_2)) \\ j_{13} &= -c_1(a_3s_{23} + s_{234}(a_4 - d_6s_5)) \\ j_{14} &= c_1s_{234}(-a_4 + d_6s_5) \\ j_{15} &= -d_6c_1c_{234}c_5 + d_6s_1s_5 \\ j_{21} &= -d_6s_1c_5 + c_1(a_1 + a_2c_2 + a_3c_{23} + c_{234}(a_4 - d_6s_5)) \\ j_{22} &= -s_1(a_2s_2 + a_3s_{23} + s_{234}(a_4 - d_6s_5)) \\ j_{23} &= -s_1(a_3s_{23} + s_{234}(a_4 - d_6s_5)) \\ j_{24} &= s_1s_{234}(-a_4 + d_6s_5) \\ j_{25} &= -d_6(c_{234}c_5s_1 + c_1s_5) \\ j_{32} &= -a_2c_2 - a_3c_{23} + c_{234}(-a_4 + d_6s_5) \\ j_{33} &= -a_3c_{23} + c_{234}(-a_4 + d_6s_5) \\ j_{34} &= c_{234}(-a_4 + d_6s_5) \\ j_{46} &= -c_5s_1 - c_1c_{234}s_5 \\ j_{56} &= c_1c_5 - c_{234}s_1s_5 \end{split}$$

The trajectory tracking cases need the Jacobian matrix to compute the desired angles from the end-effector path. The controller of the robot is proportional-integral derivative (PID), which is implemented on a board with an ARM microprocessor (Figure 18.7(a)). The vision system that consists of two high definition (HD) cameras, parallel to each other, is presented in Figure 18.7(b).



Figure 18.7. a) Digital board with ARM processor and b) vision system

By applying the PID controller, the result of the point-to-point motion is obtained. In Figure 18.8, the variations of angels of links are shown that are the results of a potentiometer measurement. Figure 18.9 shows the 3D path of the end-effector.



Figure 18.8. The variations of angles of links 1–6

In some cases, as it is for this case, the measurements based on the variation of angels of links and then the calculation of the end-effector of the arm is not the best way for an analysis of motion. The inaccuracy is due to improper modeling, backlash, etc. External measurements such as vision can provide an independent result that helps even to check the previous measurement system.



Figure 18.9. End-effector trajectory via potentiometer and vision measurements

18.6. Summary

The focus of this chapter was the application of a stereo vision approach on robotic systems to measure the 3D position of the end-effector during motion. Stereo vision is a method to determine the 3D position information of an object. A simple and inexpensive setup, containing two cameras, is made to implement stereo vision on any mechanical manipulator. Also a basic formulation of the triangulation process is mentioned to compute the 3D position and orientation of the end-effector based on two resulted images. Experimental results are provided for a 6R mechanical manipulator to show the accuracy of the proposed method for end-effector positioning. According to the experimental output, the following results can be mentioned:

1) The performance of the measurement system is related to the quality of the cameras and it can affect the accuracy and speed of the stereo vision system.

2) The other main factor is the distance between setup and the end-effector. If this distance is decreased, then the accuracy of the positioning system will be increased.

3) The sampling frequency of the measurement is a function of the quality and speed of the cameras.

4) A 3D position that is determined via stereo vision can be used in a closed-loop control system. But high sampling frequency is needed.

18.7. Bibliography

- [BLE 08] BLEYER M., GELAUTZ M., "Simple but effective tree structures for dynamic programming-based stereo matching", *Computer Vision Theory and Applications*, *VISAPP*, Funchal, Madeira, Portugal, pp. 415–422, 2008.
- [CHA 10] CHANG C.K., SIAGIAN C., ITTI L., "Mobile robot vision navigation and localization using gist and saliency", *IEEE International Conference on Intelligent Robots and Systems*, Taiwan, pp. 4147–4154, 2010.
- [DAS 10] DASTJERDI A.H., Optimizing tracking algorithm for contact probe of CMM by non contact probe based on image processing, MSc Thesis, Amir Kabir University, Iran, 2010.
- [EVO] EVOLUTION ROBOTICS INC., North Star Projector Kit User Guide, 1.0 ed., CA. https://www.irobot.com/
- [GUT 08] GUTIERREZ A., CAMPO A., L. SANTOS F.C., et al., Social Odometry in Populations of Autonomous Robots, Lecture Notes in Computer Science, vol. 5217, pp. 371–378, 2008.
- [HEE 07] HEE L.J., Sensor based localization of a mobile, PhD Thesis, National University of Singapore, 2007.
- [HUM 10] HUMENBERGER M., ENGELKE T., KUBINGER W., "A census-based stereo vision algorithm using modified semi-global matching and plane fitting to improve matching quality", *IEEE Computer Society Conference CVPRW*, San Francisco, CA, pp. 77–84, 2010.
- [JON 10] JONES E., SOATTO S., "Visual-inertial navigation, mapping and localization: a scalable real-time causal approach", *International Journal of Robotics Research*, vol. 30, pp. 407–430, 2010.
- [KUC 06] KUCSERA P., "Sensors for mobile robot systems", *Academic and Applied Research in Military Science*, vol. 5, no. 4, pp. 645–658, 2006.
- [MOO 04] MOORE D., LEONARD J., RUS D., et al., "Robust distributed network localization with noisy range measurements", Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, ACM, Baltimore, MD, USA, 2004.
- [RAM 08] RAMISA A., TAPUS A., MANTARAS R.L., et al., "Mobile robot localization using panoramic vision and combinations of feature region detectors", *IEEE International Conference on Robotics and Automation*, Pasadena, CA, pp. 538–543, 2008.
- [SAN 09] SANTOLARIA J., GUILLOMÍA D., CAJAL C., et al., "Modeling and calibration technique of laser triangulation sensors for integration in robot arms and articulated arm coordinate measuring machines", Sensors, vol. 9, pp. 7374–7396, 2009.

- [SCH] Middlebury Stereo Evaluation, available at http://vision.middlebury.edu/ stereo/eval/
- [SPA 07] SPACEK L., BURBRIDGE C., "Instantaneous robot self-localization and motion estimation with omnidirectional vision", *Robotics and Autonomous Systems*, vol. 55, pp. 667–674, 2007.
- [STA 99] STAN B., CARLO T., "Depth discontinuities by pixel-to-pixel stereo", International Journal of Computer Vision, vol. 35, no. 3, pp. 269–293, 1999.
- [TSU 09] TSUKIYAMA T., "Mobile robot localization from landmark bearings", *World Congress Fundamental and Applied Metrology*, 6–11 September, Lisbon, Portugal, 2009.
- [VIS] University of Bologna, http://www.vision.deis.unibo.it/smatt/
- [YAN 06] YANG Q., "Real-time global stereo matching using hierarchical belief propagation", *British Machine Vision Conference*, Edinburgh, Scotland, pp. 989–998, 2006.