DIRECT LINEAR TRANSFORMATION (DLT)

1 Overview

Direct linear transformation (DLT) is a method of determining the three dimensional location of an object (or points on an object) in space using two views of the object. First, let's consider a few different ways of obtaining multiple views of an object:

1) Two cameras
+ The object can occupy the full image in each camera, thereby yielding a lot of pixels for high resolution.
+ Easy to adjust the angle between the cameras and the object for optimal viewing.
− Synchronization of the cameras can be difficult, and will usually involve separate (and often expensive) hardware when imaging a moving object.

2) One camera and one prism
+ This is a simple arrangement that only requires one camera.
+ No synchronization is necessary.
− The relative distances between the camera, prism, and object is limited, thus this is usually only good for small objects.
− The camera image is split between the two views, reducing the resolution.

3) One camera and an arrangement of mirrors
+ No synchronization necessary.
+ The relative distances between the camera, mirrors, and object can be varied over a wider range than that allowed using the prism approach.
− As in the prism approach, one image contains two views.
− High-quality mirrors and positioning optics can be moderately expensive.

4) One camera, two separate images
+ Very simple; only requires one camera.
+ As in the two camera approach, object can occupy the full image in each camera.
− Object must be stationary.
− Each image pair must be calibrated.

2 DLT Procedure

For the following discussion, let's assume that we have two cameras to obtain two images (method 1 from above).

First we need to define two coordinate systems and reference frames; these are shown in Fig. 1. Note that capital letters are used to denote coordinate systems and lower-case letters are used to locate points within coordinate systems. The object lies in what we call the "object space reference frame" and is referenced to as the *XYZ* coordinate system. It simply locates the object in real three-dimensional space. The *XYZ* coordinate system can have its origin anywhere you choose. One convenient choice is to choose a point on the object as the origin; another is to place another object in space from which to reference the coordinate system.

There is a two-dimensional reference frame associated with each camera image; these are called the "image plane reference frames" and are denoted using *U* and *R*. In DLT there will always be two views, which we will refer to as the "left" and "right" views. Thus the left and right image plane reference frames are referenced to the U_LV_L and U_RV_R coordinate systems, respectively. In Fig. 1 there is one image plane reference frame for the left camera, and one for the right camera.

Let's consider the point $[x, y, z]$ located on an object as shown in Fig. 1. This point appears in the left and right images, located by image coordinates $[u_L, v_L]$ and $[u_R, v_R]$, respectively. The point $[x, y, z]$ will have units of length (i.e., meters in SI units) and $[u_L, v_L]$ and $[u_R, v_R]$ will have units of pixels.

The goal of DLT is to determine the actual location of the point $[x, y, z]$ based on u_L, v_L, u_R, and v_R. Before this can be done using an object, the system must be calibrated using points of known location.
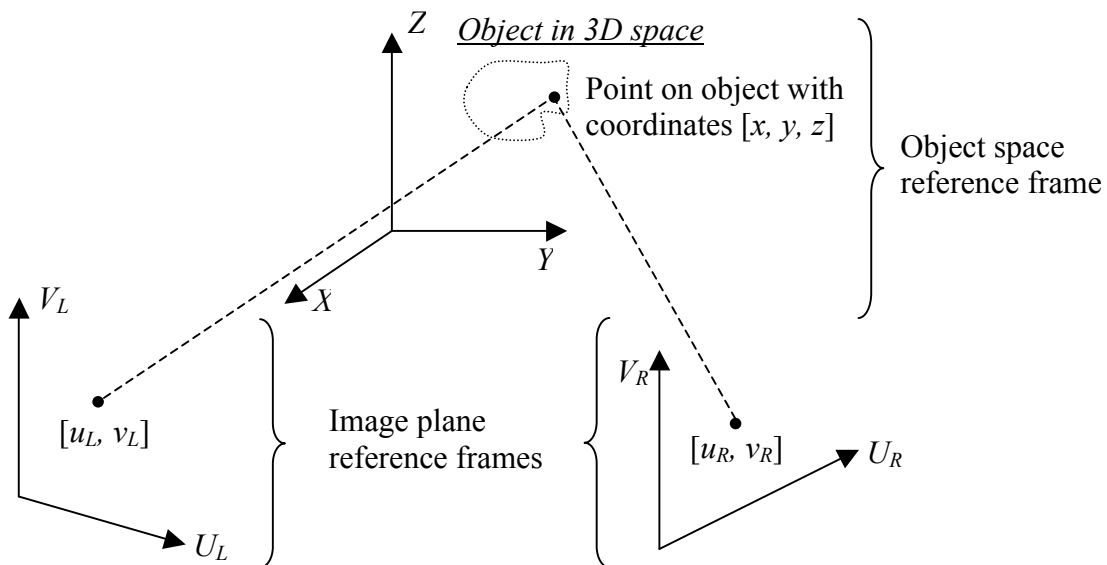

Figure 1: Object space and image plane reference frames and associated coordinate systems.

2.1 Calibration: Finding L and R Matrices

Let's assume that we know the location of the point $[x, y, z]$. We acquire an image pair, from which we can find u_L, v_L, u_R, and v_R. The image points $[u_L, v_L]$ and $[u_R, v_R]$ and the object point $[x, y, z]$ can be related through a series of constants:

$$u_L = \frac{L_1 x + L_2 y + L_3 z + L_4}{L_9 x + L_{10} y + L_{11} z + 1}, \tag{1a}$$

$$v_L = \frac{L_5 x + L_6 y + L_7 z + L_8}{L_9 x + L_{10} y + L_{11} z + 1}, \tag{1b}$$

$$u_R = \frac{R_1 x + R_2 y + R_3 z + R_4}{R_9 x + R_{10} y + R_{11} z + 1}, \tag{1c}$$

$$v_R = \frac{R_5 x + R_6 y + R_7 z + R_8}{R_9 x + R_{10} y + R_{11} z + 1}. \tag{1d}$$

From Eqns. (1a-1d) we can see that with one calibration point, we have seven knowns (u_L, v_L, u_R, v_R, x, y, and z), 22 unknowns ($L_1 \ldots L_{11}$ and $R_1 \ldots R_{11}$), and four equations. To find the 22 unknowns, we need at least 22 equations. This is done by choosing more than one calibration point. For each additional calibration point, we introduce four new equations, while the constants L and R remain the same. Six calibration points will yield 24 equations, thus we need to acquire at least six calibration points to determine L and R.

Once we have determined u_L, v_L, u_R, v_R, x, y, and z for at least six points, we assemble them in matrix form. To see this, let's consider two points as shown in Fig. 2.
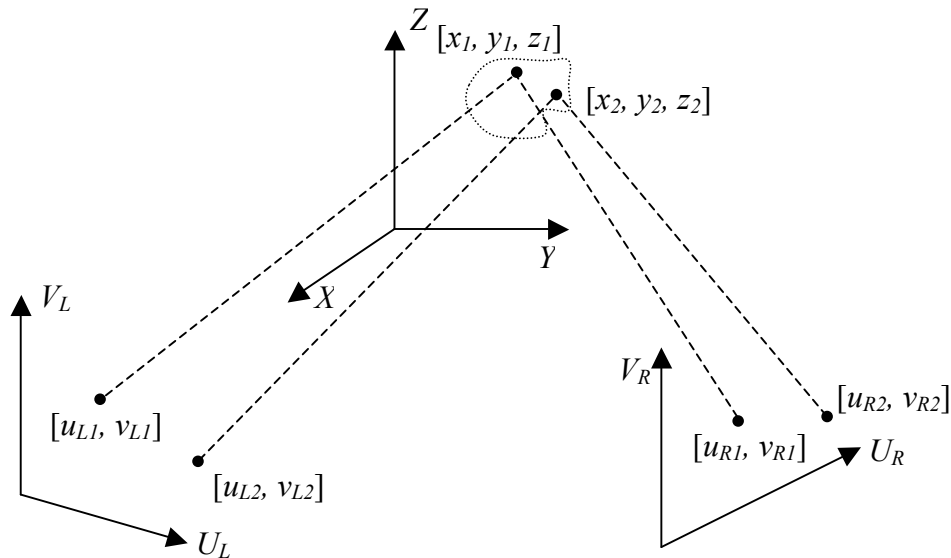

Figure 2: Imaging of two calibration points.

For the two points in the left image frame, the equations are

$$u_{L1} = \frac{L_1 x_1 + L_2 y_1 + L_3 z_1 + L_4}{L_9 x_1 + L_{10} y_1 + L_{11} z_1 + 1}, \tag{2a}$$

$$v_{L1} = \frac{L_5 x_1 + L_6 y_1 + L_7 z_1 + L_8}{L_9 x_1 + L_{10} y_1 + L_{11} z_1 + 1}, \tag{2b}$$

$$u_{L2} = \frac{L_1 x_2 + L_2 y_2 + L_3 z_2 + L_4}{L_9 x_2 + L_{10} y_2 + L_{11} z_2 + 1}, \tag{2c}$$

$$v_{L2} = \frac{L_5 x_2 + L_6 y_2 + L_7 z_2 + L_8}{L_9 x_2 + L_{10} y_2 + L_{11} z_2 + 1}. \tag{2d}$$

Additional equations similar to Eqns. (2a-2d) will result for each calibration point selected. These can each be rearranged as shown below, using Eq. (2a) as an example.

$$u_{L1}\left(L_9 x_1 + L_{10} y_1 + L_{11} z_1 + 1\right) = L_1 x_1 + L_2 y_1 + L_3 z_1 + L_4, \tag{3}$$

$$u_{L1} = L_1 x_1 + L_2 y_1 + L_3 z_1 + L_4 - u_{L1} L_9 x_1 - u_{L1} L_{10} y_1 - u_{L1} L_{11} z_1. \tag{4}$$

Similar equations can be obtained for each $u_{L1}...u_{LN}$, $v_{L1}...v_{LN}$, $u_{R1}...u_{RN}$, and $v_{R1}...v_{RN}$, where N is the number of calibration points (at least six, but can be more). For example, v_{L1} will yield:

$$v_{L1} = L_5 x_1 + L_6 y_1 + L_7 z_1 + L_8 - v_{L1} L_9 x_1 - v_{L1} L_{10} y_1 - v_{L1} L_{11} z_1. \tag{5}$$

Equations (4) and (5) can be assembled in matrix form as follows:

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -u_{L1}x_1 & -u_{L1}y_1 & -u_{L1}z_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -v_{L1}x_1 & -v_{L1}y_1 & -v_{L1}z_1 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11} \end{bmatrix} = \begin{bmatrix} u_{L1} \\ v_{L1} \end{bmatrix}. \tag{6}$$

We can similarly add to the matrix as we acquire up to N calibration points:

$$
\begin{array}{l}
\text{Point 1} \left\{\rule{0pt}{28pt}\right.\\[6pt]
\text{Point 2} \left\{\rule{0pt}{28pt}\right.\\[14pt]
\text{Point } N \left\{\rule{0pt}{28pt}\right.
\end{array}
\underbrace{
\begin{bmatrix}
x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -u_{L1}x_1 & -u_{L1}y_1 & -u_{L1}z_1 \\
0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -v_{L1}x_1 & -v_{L1}y_1 & -v_{L1}z_1 \\
x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 & -u_{L2}x_2 & -u_{L2}y_2 & -u_{L2}z_2 \\
0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 & -v_{L2}x_2 & -v_{L2}y_2 & -v_{L2}z_2 \\
& & & & & \vdots & & & & & \\
x_N & y_N & z_N & 1 & 0 & 0 & 0 & 0 & -u_{LN}x_N & -u_{LN}y_N & -u_{LN}z_N \\
0 & 0 & 0 & 0 & x_N & x_N & x_N & 1 & -v_{LN}x_N & -v_{LN}y_N & -v_{LN}z_N
\end{bmatrix}}_{2N \times 11}
\underbrace{
\begin{bmatrix}
L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11}
\end{bmatrix}}_{11 \times 1}
=
\underbrace{
\begin{bmatrix}
u_{L1} \\ v_{L1} \\ u_{L2} \\ v_{L2} \\ \vdots \\ u_{LN} \\ v_{LN}
\end{bmatrix}}_{2N \times 1}.
\tag{7}
$$

In Eq. (7) the values $L_1...L_{11}$ are the only unknowns. A similar matrix system involving u_R, v_R, and $R_1...R_{11}$ can be written. From here, we'll denote the left-hand matrix of (7) as $\mathbf{F_L}$ (using bold, non-italicized font to signify a matrix), the $L_1...L_{11}$ matrix as $\mathbf{L}$, and the RHS matrix as $\mathbf{g_L}$. The corresponding matrices for the right image will be $\mathbf{F_R}$, $\mathbf{R}$, and $\mathbf{g_R}$, so that Eq. (7) and its right image counterpart can be expressed as

$$\mathbf{F_L L} = \mathbf{g_L}, \tag{8}$$

$$\mathbf{F_R R} = \mathbf{g_R}. \tag{9}$$

Calibration is achieved by solving for $\mathbf{L}$ and $\mathbf{R}$. Since $\mathbf{F_L}$ and $\mathbf{F_R}$ are not square, they cannot be inverted, and $\mathbf{L}$ and $\mathbf{R}$ must instead be calculated using the method of least squares. A simple way to do this is using the "Moore-Penrose pseudo-inverse" method. This is shown here for Eq. (8). The first step is to pre-multiply both sides by $\mathbf{F_L}^{\mathbf{T}}$:

$$\mathbf{F_L^T F_L L} = \mathbf{F_L^T g_L}. \tag{10}$$

Since $\mathbf{F_L}^{\mathbf{T}}\mathbf{F_L}$ is square, it can be inverted. When we pre-multiply both sides by the product $(\mathbf{F_L^T F_L})^{-1}$,

$$\left(\mathbf{F_L^T F_L}\right)^{-1}\left(\mathbf{F_L^T F_L}\right)\mathbf{L} = \left(\mathbf{F_L^T F_L}\right)^{-1}\mathbf{F_L^T g_L}, \tag{11}$$

the identity matrix is formed on the left-hand side, so that the solution for $\mathbf{L}$ is obtained:

$$\mathbf{L} = \left(\mathbf{F_L^T F_L}\right)^{-1}\mathbf{F_L^T g_L}, \tag{12}$$

and similarly for **R**:

$$\mathbf{R} = \left(\mathbf{F_R^T F_R}\right)^{-1}\mathbf{F_R^T g_R} \ . \tag{13}$$

A few notes about this process:

1. At least six points are necessary to determine **L** and another six to find **R**. It is not necessary to use the same six points for both **L** and **R**, but it is often convenient to do so.

2. More than six points can improve the estimate of **L** and **R**.

3. The calibration points should span the object region of interest. The test object that will eventually be imaged (after calibration) should be contained within regions where calibration points were located. A smaller calibration region will not provide a good estimate, and an excessively large calibration region may be more work and will be less accurate.

4. Once **L** and **R** have been obtained, no further calibration is necessary as long as the camera positions and settings do not change and the object to be imaged is in the calibrated domain.

2.2 Implementation: Using L *and* R *to Locate Points of Unknown Position*

Once we have calibrated the imaging system by finding **L** and **R**, we can now find the location of other points in the calibrated space. This is illustrated in Fig. 3, where the knowns are now u_L, v_L, u_R, v_R, **L**, and **R**, and the only unknowns are x, y, and z. These can found using Eqns. (1a-1d) (copied here for convenience):

$$u_L = \frac{L_1 x + L_2 y + L_3 z + L_4}{L_9 x + L_{10} y + L_{11} z + 1}, \tag{1a}$$

$$v_L = \frac{L_5 x + L_6 y + L_7 z + L_8}{L_9 x + L_{10} y + L_{11} z + 1}, \tag{1b}$$

$$u_R = \frac{R_1 x + R_2 y + R_3 z + R_4}{R_9 x + R_{10} y + R_{11} z + 1}, \tag{1c}$$

$$v_R = \frac{R_5 x + R_6 y + R_7 z + R_8}{R_9 x + R_{10} y + R_{11} z + 1}. \tag{1d}$$
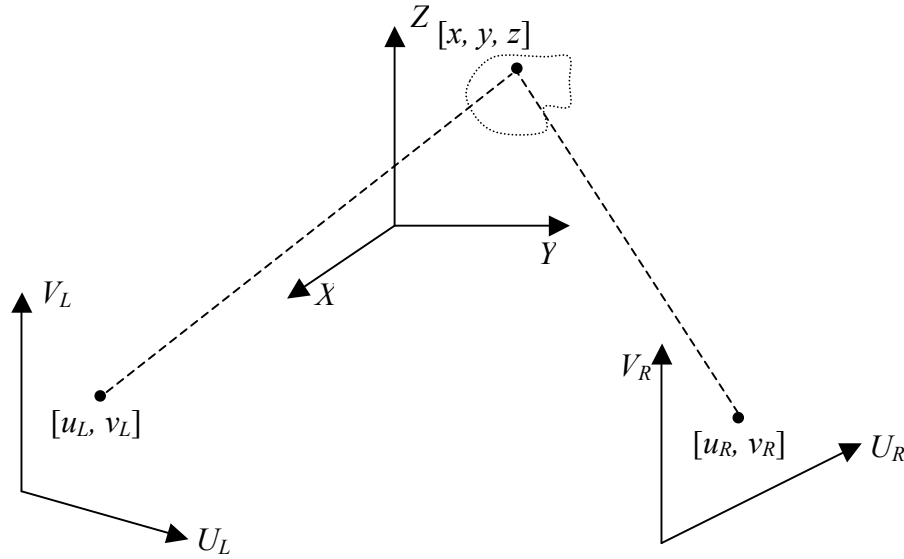
Figure 3: Imaging an unknown point.

In Eqns. (1a-1d) after calibration, $L_1 \ldots L_{11}$, $R_1 \ldots R_{11}$ are known, and u_L, v_L, u_R, and v_R, are known by inspecting the images. There are thus three unknowns (x, y, and z) and four equations. Rearrange the equations and combining into a matrix system yields

$$
\begin{bmatrix}
L_1 - L_9 u_L & L_2 - L_{10} u_L & L_3 - L_{11} u_L \\
L_5 - L_9 v_L & L_6 - L_{10} v_L & L_7 - L_{11} v_L \\
R_1 - R_9 u_R & R_2 - R_{10} u_R & R_3 - R_{11} u_R \\
R_5 - R_9 v_R & R_6 - R_{10} v_R & R_7 - R_{11} v_R
\end{bmatrix}
\begin{bmatrix} x \\ y \\ z \end{bmatrix}
=
\begin{bmatrix}
u_L - L_4 \\
v_L - L_8 \\
u_R - R_4 \\
v_R - R_8
\end{bmatrix}. \tag{14}
$$

If we denote the first matrix on the left-hand side as **Q** and the right-hand side matrix as **q**, Eq. (14) can be written as:

$$
\mathbf{Q} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{q}, \tag{15}
$$

from which [x, y, z] can be found using the Moore-Penrose pseudo-inverse method:

$$
\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \left(\mathbf{Q}^{\mathrm{T}} \mathbf{Q} \right)^{-1} \mathbf{Q}^{\mathrm{T}} \mathbf{q}. \tag{16}
$$

3 Summary of the DLT Procedure

In summary, the steps necessary to use DLT to locate the position of an unknown point (or points) in space are as follows:

1. Calibrate the system

   a. Find $[x, y, z]$, $[u_L, v_L]$, and $[u_R, v_R]$ for at least six points, making sure that they span the physical domain of interest.

   b. Use Eqns. (12) and (13) to find the calibration matrices **L** and **R**.

2. Find the position of unknown points

   a. Take images of the object or points to be located and determine $[u_L, v_L]$ and $[u_R, v_R]$ for each point of interest.

   b. Using calibration matrices **L** and **R**, use Eq. (16) to find $[x, y, z]$ for each point of interest.