Original Lecture #3:	8 October 1992
Topics:	Transformations
Scribe:	Patrick Teo

1 Transformation by Matrices

We can represent general transformations of homogeneous coordinates by matrices. This idea has been used widely in geometric modeling to describe the relationships between objects. Because the transitive closure of the transforms between multiple objects is simply the composition of the transformation matrices, the use of matrices to represent transformations is especially appealing. Also, matrix representation has been used to describe the viewer's perspective of a scene. Since the viewer's perspective is represented as a transformation matrix, it is indistinguishable from actual geometric relations between objects. Hence, we can manipulate the point of view in the same way as we do geometric relationships between objects.

2 2-D Transformation Matrices

We shall first consider the two dimensional case. The transformation of two dimensional homogeneous coordinates $(w;x,y)$ can be represented by a 3×3 matrix. We will treat the general transformation as a composition of the affine case and the perspective case.

2.1 2-D Affine Transformation

We have seen before that the rotation of a point about the origin by a counterclockwise angle θ can be written as the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \tag{1}$$

while the translation of a point can be described by the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ T_x & 1 & 0 \\ T_y & 0 & 1 \end{pmatrix} \tag{2}$$

Similarly, scaling[1] along the x and y axes can be represented by the matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & k_x & 0 \\ 0 & 0 & k_y \end{pmatrix} \tag{3}$$

More complex affine transformations may be described as a composition of these basic transformations. For example, the rotation of a point about an arbitrary point (x_0, y_0) by an angle θ can be described as a translation of the point (x_0, y_0) to the origin followed by a rotation about the origin and finally by a translation of the origin back to the point (x_0, y_0). This can be seen shown algebraically as:

$$R_{(x_0,y_0),\theta} = T^{-1} R_\theta T \tag{4}$$

where

$$R_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, T = \begin{pmatrix} 1 & 0 & 0 \\ -x_0 & 1 & 0 \\ -y_0 & 0 & 1 \end{pmatrix}$$

2.2 2-D Perspective Transformation

Having considered the affine case, we shall now treat the perspective case. Notice that the first row of the 3×3 matrix has been kept at $(1, 0, 0)$ throughout all the affine examples. This is because these are the parameters which determine the perspective part of a general transformation. In order to simplify understanding, we will explain the perspective transformation in one dimension first.

In the one-dimensional case, we have homogeneous coordinates of the form $(w; x)$ where $(1; x)$ is the specialized Cartesian case and correspondingly the 2×2 transformation matrix of the form:

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \tag{5}$$

Note that the above matrix has no translation or rotation components. This is only to simplify explanation. Rotation and translation components may be considered as compositions with the matrix subsequently.

Now, we can apply the perspective transformation onto the one- dimensional point by pre-multiplying the matrix by the homogeneous coordinates of the point:

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w \\ x \end{pmatrix} = \begin{pmatrix} w + ax \\ x \end{pmatrix}.$$

i.e. it can be thought of as a mapping

$$[w; x] \mapsto [w + ax; x]$$

[1]Enlargement, or uniform scaling, is the special case where the scale factor is the same along both axes.

in the homogeneous sense or a mapping

$$X \mapsto \frac{X}{1+aX}$$

in the Cartesian sense.

In order to visualize such a projection graphically, we have to jump up a dimension and consider it in two-dimensions. This can be seen in Figures 1(a)-(d). We can decompose the domain of the mapping into three regions.

The first region, the interval $(0,+\infty)$, is compressed into the interval $(0,\frac{1}{a})$. The second region, $(-\frac{1}{a},0)$, is expanded into the interval $(-\infty,0)$. Finally, the last region, $(-\infty,-\frac{1}{a})$, is mapped onto the interval $(\frac{1}{a},+\infty)$, which involves wrapping the region through infinity, over to the opposite side of the origin. In summary,

$$(0,+\infty) \to (0,\frac{1}{a})$$

$$(-\frac{1}{a},0) \to (-\infty,0)$$

$$(-\infty,-\frac{1}{a}) \to (\frac{1}{a},+\infty)$$

$$-(\frac{1}{a})^{-} \to +\infty$$

$$-(\frac{1}{a})^{+} \to -\infty$$

Now, we shall consider the two-dimensional perspective case, which can be represented by the following matrix, also with null translation and linear components.

$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

As an example, consider the simple case where $b = 0$; we then have the following:

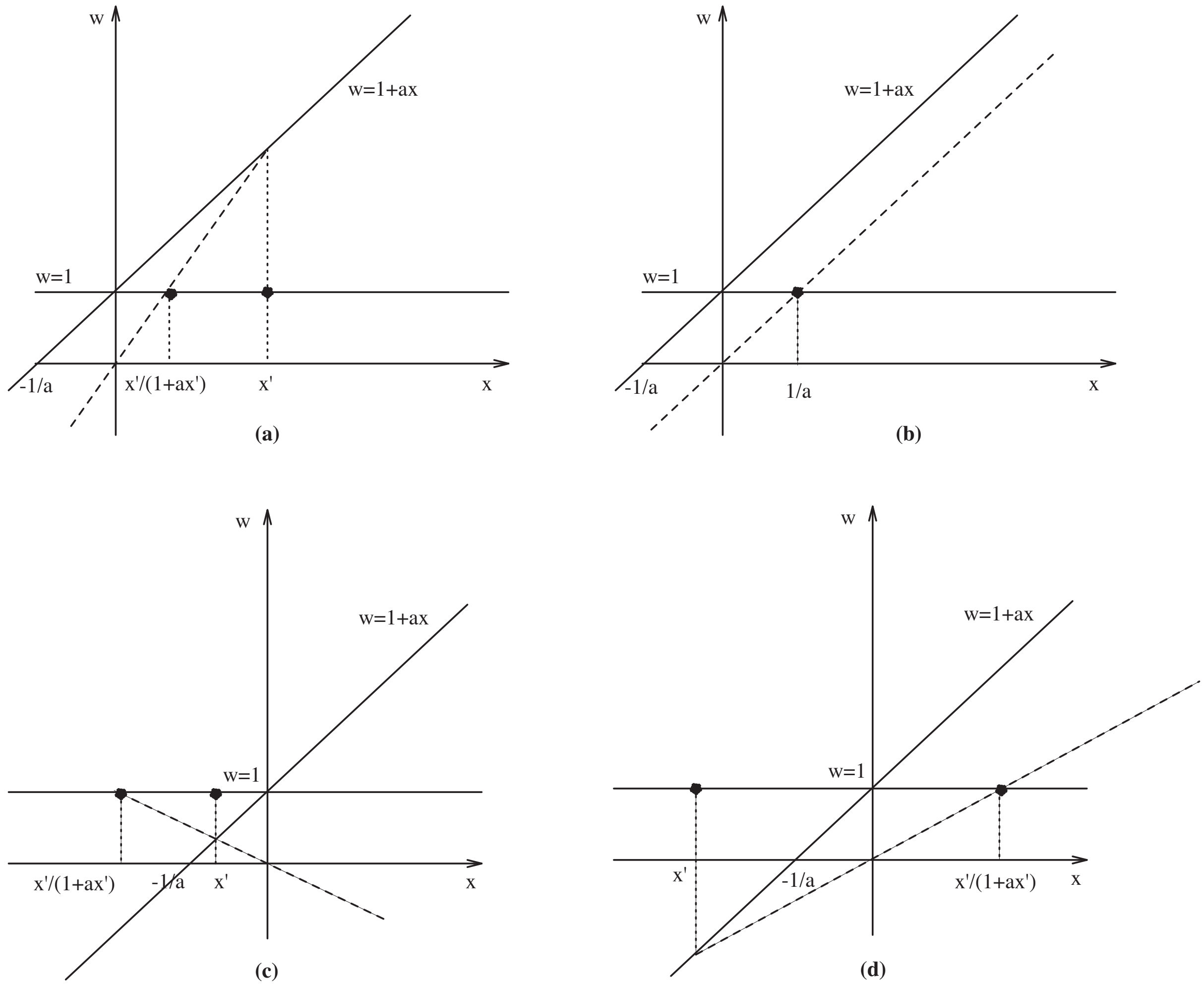
$$\begin{pmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \end{pmatrix} \begin{pmatrix} w+ax \\ x \\ y \end{pmatrix}$$

which can be thought of as a mapping, in cartesian coordinates, of the following:

$$X' = \frac{X}{1+aX}$$

$$Y' = \frac{Y}{1+aX}$$

We can observe that points with x-coordinates further away from the point $(0,-\frac{1}{a})$ are mapped closer to the y-axis and to the point $(0,-\frac{1}{a})$. Taking this to the limit, the point at infinity along the x-axis is mapped to the point $(\frac{1}{a},0)$. This behavior is known as *perspective foreshortening*, while the point $(0,-\frac{1}{a})$ can be thought to be the eye of the observer.


w
w=1+ax
w=1
-1/a
x'/(1+ax')
x'
x
(a)
w
w=1+ax
w=1
-1/a
1/a
x
(b)
w
w=1+ax
w=1
x'/(1+ax')
-1/a
x'
x
(c)
w
w=1+ax
w=1
x'
-1/a
x'/(1+ax')
x
(d)

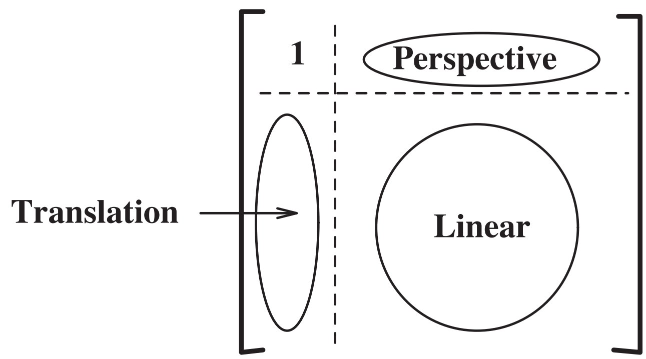


Figure 2: Visualizing the Transformation Matrix.

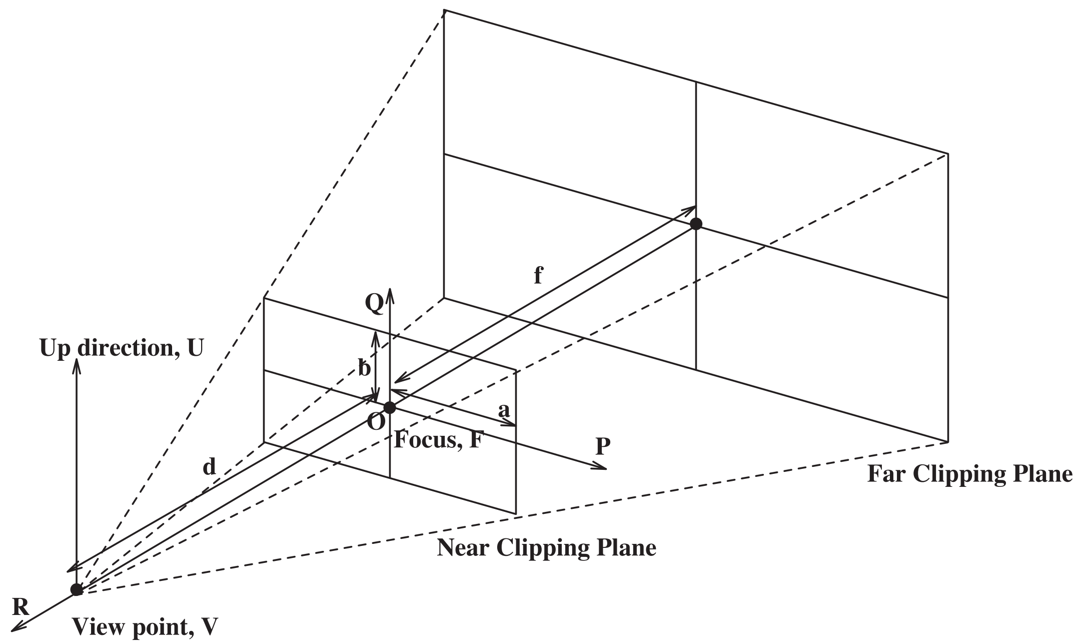


Figure 3: Viewing Coordinate System.

2.3 2-D General Transformation

In general, we can view the general transformation of homogeneous coordinates as an $(n+1) \times (n+1)$ matrix (where $n = 2$ in the two dimensional case) with a translation component, a linear component and a perspective component as shown in Figure 2.

3 The General Viewing Transform

The viewing transformation is involved with mapping a region of the world coordinate system onto a local coordinate system. In general, to specify a viewing transform in three dimensions, we need to specify the following parameters:

View Position This is essentially the position of the eye and affords three degrees of freedom.

Focus Point This can be thought of as the direction of gaze and affords two degrees of freedom (represented as a normalized three-dimensional vector).

Vertical Direction This specifies the up-direction. Given the view position and the focus point, this is just an angle and affords only one extra degree of freedom. Although this is an angle, it is often represented as a vector.

Viewing Pyramid This is essentially the angles of view along both the $\mathscr{P}$ and $\mathscr{Q}$ axes.[2]

Front and Back Clipping Planes These planes place constraints on the depth at which viewing is possible.[3]

A typical viewing coordinate system is shown in Figure 3. By convention, the general viewing matrix is defined over the region $-1 \leqslant p \leqslant 1$, $-1 \leqslant q \leqslant 1$ and $-1 \leqslant r \leqslant 0$. Hence we need to map the truncated viewing pyramid between the near and far clipping planes onto this box. This can be characterized by the following mappings of the $\mathscr{R}$-axis:

$$d \to +\infty$$

$$0 \to 0$$

$$-f \to -1$$

Considering only the $\mathscr{R}$-axis, we can deduce the transform required to be:

$$\begin{pmatrix} 1 & -\frac{1}{d} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ r \end{pmatrix} = \begin{pmatrix} 1 - \frac{r}{d} \\ r \end{pmatrix}$$

$$\mathscr{R} \to \frac{\mathscr{R}}{1 - \frac{\mathscr{R}}{d}}$$

This takes the coordinate $\mathscr{R} = d$ to infinity and maps $\mathscr{R} = 0$ to zero, but doesn't take $\mathscr{R} = -f$ to -1 as we would like it to. In order to achieve this, we need to apply a scaling factor to scale the interval $\mathscr{R} \in [-f, 0]$ to $[-1, 0]$. This gives us the following matrix:

$$\begin{pmatrix} 1 & -\frac{1}{d} \\ 0 & \frac{1}{d} + \frac{1}{f} \end{pmatrix}$$

[2]The $\mathscr{R}$-axis here is taken to be the depth of the scene.

[3]By convention, the front clipping plane is taken to be the plane at $\mathscr{R} = 0$ and the back clipping plane is the plane at $\mathscr{R} = -f$.

Extending this to three dimensions we get,

$$\begin{pmatrix} 1 & 0 & 0 & -\frac{1}{d} \\ 0 & \frac{1}{a} & 0 & 0 \\ 0 & 0 & \frac{1}{b} & 0 \\ 0 & 0 & 0 & \frac{1}{d}+\frac{1}{f} \end{pmatrix} \tag{7}$$

where the $\frac{1}{a}$ and $\frac{1}{b}$ terms scale the $\mathscr{P}$ and $\mathscr{Q}$ axes to the intervals $-1 \leqslant p \leqslant 1$ and $-1 \leqslant q \leqslant 1$ respectively.

Up till now, we have only considered the perspective component of the transform. In order to map the world $[X,Y,Z]$ axes to our local $[P,Q,R]$ axes of our viewing coordinate system, we utilize the vectors, $\mathscr{V}$, the point of view, $\mathscr{F}$, the focal point (which is the origin of our $[P,Q,R]$ axes) and $\mathscr{U}$, the vector representing the up-direction. From these, we derive:

$$\mathscr{R} = \frac{\mathscr{V}-\mathscr{F}}{\|\mathscr{V}-\mathscr{F}\|}$$

$$\mathscr{Q} = \frac{(\mathscr{U}-\mathscr{F}) - [(\mathscr{U}-\mathscr{F})\cdot\mathscr{R}]\mathscr{R}}{\|(\mathscr{U}-\mathscr{F}) - [(\mathscr{U}-\mathscr{F})\cdot\mathscr{R}]\mathscr{R}\|}$$

$$\mathscr{P} = \mathscr{Q}\times\mathscr{R}$$

Since we can consider the linear component of a transformation as a change of bases, we can represent the change of coordinate system simply as the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathscr{P}_x & \mathscr{Q}_x & \mathscr{R}_x \\ 0 & \mathscr{P}_y & \mathscr{Q}_y & \mathscr{R}_y \\ 0 & \mathscr{P}_z & \mathscr{Q}_z & \mathscr{R}_z \end{pmatrix} \tag{8}$$

Finally, in order to completely describe the viewing transform, we need to move the origin of the $[X,Y,Z]$ axes to our $[P,Q,R]$ axes. This can be accomplished by the following translation matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \mathscr{T}_x & 1 & 0 & 0 \\ \mathscr{T}_y & 0 & 1 & 0 \\ \mathscr{T}_z & 0 & 0 & 1 \end{pmatrix} \tag{9}$$

The general transform matrix can then be computed as the composition of the perspective, linear and translation components described above.