THE VIEWING TRANSFORMATION

Kenneth I. Joy
Visualization and Graphics Research Group
Department of Computer Science
University of California, Davis

Overview

One of the most important operations in rendering is the projection of a a three-dimensional scene onto a two-dimensional screen from an arbitrary camera position. A fundamental part of this operation is the specification of a *viewing transformation*, a 4×4 matrix that transforms a region of space into image space.
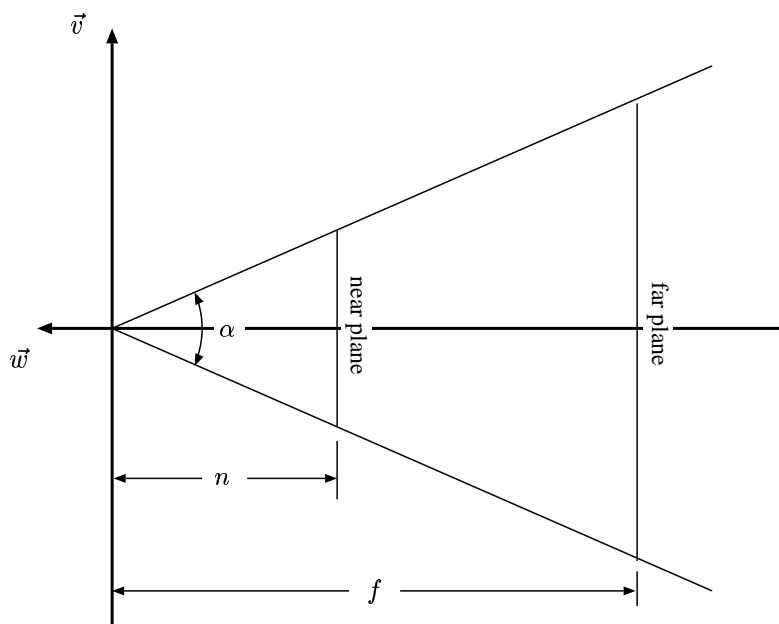

Specification of the Parameters

We will assume that the user has defined a camera transform which can be applied to the object in the Cartesian frame. This transform will convert the coordinates of the object into the local coordinate system of the camera. Assuming this, we can assume that we are viewing the object from the origin of the camera frame and the scene has been transformed to lie on the negative w-axis of the frame. We also assume that the user has defined the following parameters:

- A field-of-view angle, α, representing the angle subtended at the apex of the viewing pyramid, and

- The specification of "near" and "far" bounding planes. These planes are perpendicular to the z-axis at a distance of n and f from the camera, respectively.

A three-dimensional view of the camera and its viewing space is given in the following figure.
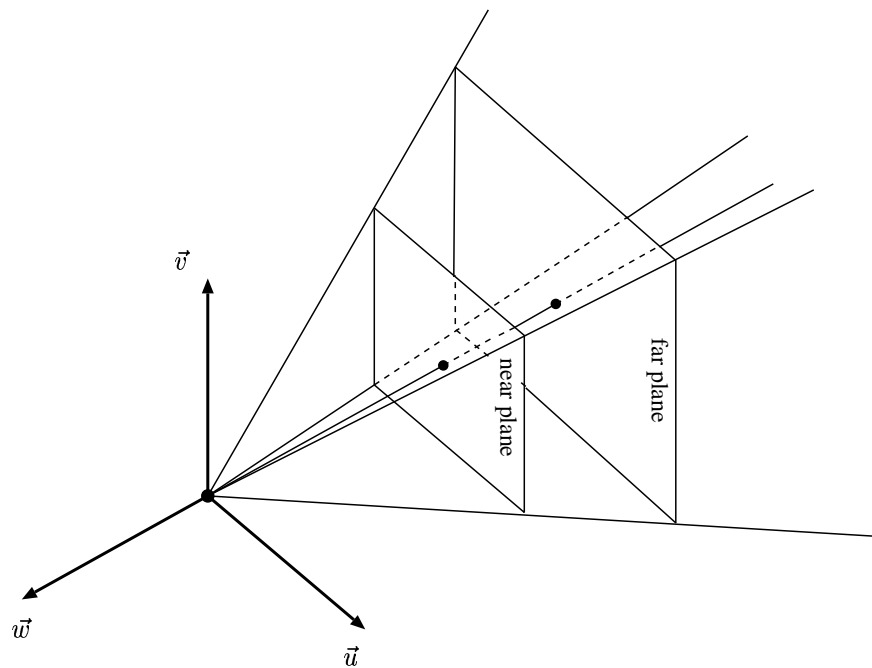
A side view of the this space, with the u-axis coming out of the paper, is given in the following figure. Note the field-of-view angle α.

The specification of α forms a viewing volume in the shape of a pyramid with the camera (placed at the

origin) at the apex of the pyramid and the negative-w axis forming the axis of the pyramid. This pyramid is commonly referred to as *the viewing pyramid.*

The specification of the near and far planes forms a truncated viewing pyramid giving a region of space that contains the portion of the scene which is the "center of attention" of the camera. The viewing transform, defined below, will transform this truncated viewing pyramid onto the image space volume $-1 \leq u, v, w \leq 1$.

The Viewing Transformation Matrix

Given the specification of the parameters (α, n, f), we define a transformation that can be applied to all elements of a scene and takes the truncated viewing volume (bounded by the viewing pyramid and the planes $z = -n$ and $z = -f$) to the cube $-1 \leq u, v, w \leq 1$. This transformation is given by

$$
A_{\alpha,n,f} = \begin{bmatrix} \cot \frac{\alpha}{2} & 0 & 0 & 0 \\ 0 & \cot \frac{\alpha}{2} & 0 & 0 \\ 0 & 0 & \frac{f+n}{f-n} & -1 \\ 0 & 0 & \frac{2fn}{f-n} & 0 \end{bmatrix}
$$

The transformation $A_{\alpha,n,f}$ is commonly referred to as the *viewing transformation* and is developed below.

Development of the Matrix

The viewing transformation $A_{\alpha,n,f}$ is not a combination of simple translations, rotations, scales or shears: its development is more complex. First, we motivate the development of the projection portion of the matrix and then apply this knowledge to the construction of the actual matrix.

Motivation

As motivation, consider the case when the camera is at the origin, the viewing direction is along the negative w-axis, and points are to be projected along the line that passes through both the origin and the point, onto a plane defined by $w = -d$. The following figure illustrates the projection of a point (u, v, w) onto the plane.
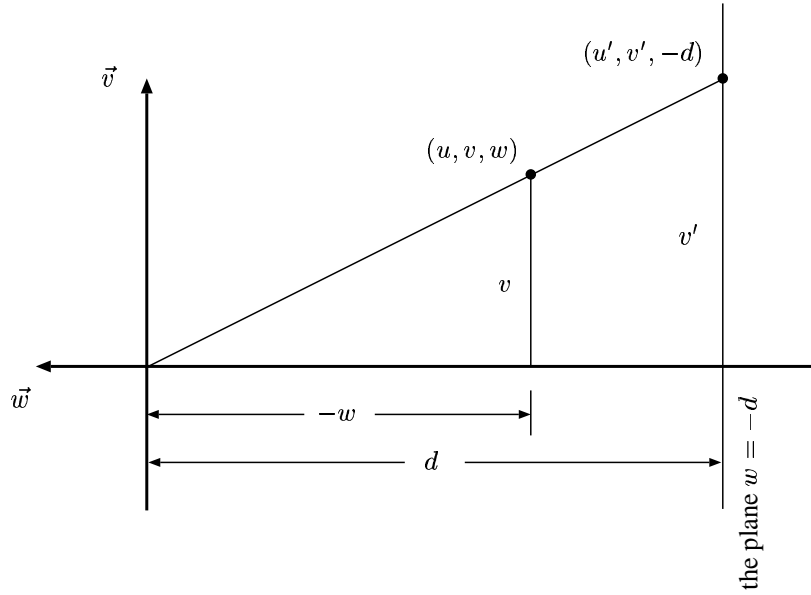
By a similar triangle argument

$$v' = \left(\frac{d}{-w}\right) v$$

(we note that the distance is $-w$ since the w coordinate of (u, v, w) is negative) and similarly

$$u' = \left(\frac{d}{-w}\right) u$$

A transformation that projects

$$(u, v, w) \longrightarrow \left(\frac{du}{-w}, \frac{dv}{-w}, -d\right) = \left(\frac{du}{-w}, \frac{dv}{-w}, \frac{dw}{-w}\right)$$

can be expressed in 4-dimensional homogeneous coordinates by

$$(u, v, w, 1) \longrightarrow (du, dv, dw, -w)$$

which can be expressed in a matrix form by

$$P_d = \begin{bmatrix} d & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

since

$$\begin{bmatrix} u & v & w & 1 \end{bmatrix} \begin{bmatrix} d & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} du & dv & dw & -w \end{bmatrix}$$

So the projection induces a unique fourth column in the matrix. When the matrix is applied to a point (u, v, w) it returns the distance of the w coordinate from the xy plane ($-w$ because w is negative). Since we divide by the w coordinate, the result of the operation is inversely proportional to the distance of the point from the xy plane.

Now, consider the case where n, f and the field of view α are present as parameters, and it is necessary to transform the viewing pyramid defined by the angle α and the planes $w = -n$ and $w = -f$ into the cube $-1 \le u, v, w \le 1$.
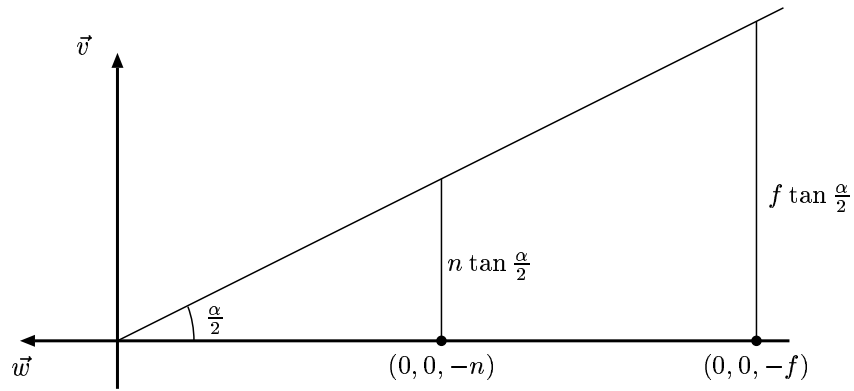
To transform the truncated viewing pyramid to the cube, we will start with a transformation P of the

5

form

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & a & -1 \\ 0 & 0 & b & 0 \end{bmatrix}.$$

Here we have incorporated the projection in the fourth column, and have recognized that we must scale the resulting z values (the z values in the pyramid range between $-n$ and $-f$ and the result in image space must lie between 1 and -1) and must translate the center truncated pyramid over to the origin. The values a and b are chosen so that the face of the truncated pyramid defined by the near plane ($z = -n$) goes to the face $w = 1$ of the image space cube, and the face defined by the far plane ($z = -f$) goes to the face $w = -1$ of the cube. At a minimum, this implies that

$$(0, 0, -n)P = (0, 0, 1), \text{ and}$$
$$(0, 0, -f)P = (0, 0, -1)$$

and these two equations will enable us to calculate a and b.

To calculate these, we apply the matrix to obtain

$$\begin{bmatrix} 0 & 0 & -n & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & a & -1 \\ 0 & 0 & b & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -an + b & n \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & 0 & -f & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & a & -1 \\ 0 & 0 & b & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -af + b & f \end{bmatrix}$$

Projecting these back to three dimensions, we obtain

$$(0,0,-n)P = (0,0,\frac{-an+b}{n}), \text{ and}$$

$$(0,0,-f)P = (0,0,\frac{-af+b}{f})$$

That is, in order that the values on the left map to $(0,0,1)$ and $(0,0,-1)$, respectively, we must have

$$-an+b = n, \text{ and}$$

$$-af+b = -f$$

Subtracting the equations, obtains

$$af - an = f + n$$

or

$$a = \left(\frac{f+n}{f-n}\right)$$

and by substitution,

$$b = n + an$$
$$= n(1+a)$$
$$= n(1 + \frac{f+n}{f-n})$$
$$= n(\frac{f-n+f+n}{f-n})$$
$$= \frac{2fn}{f-n}$$

Substituting for a and b, the transformation P becomes

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{f+n}{f-n} & -1 \\ 0 & 0 & \frac{2fn}{f-n} & 0 \end{bmatrix}$$

Unfortunately, this is not quite correct, as we have mapped only the near and far faces that correspond to the near and far planes. We need to also adjust for the left, right, top and bottom faces so that the viewing pyramid is transformed to the image space cube. So consider points on the top plane that bounds the viewing pyramid.
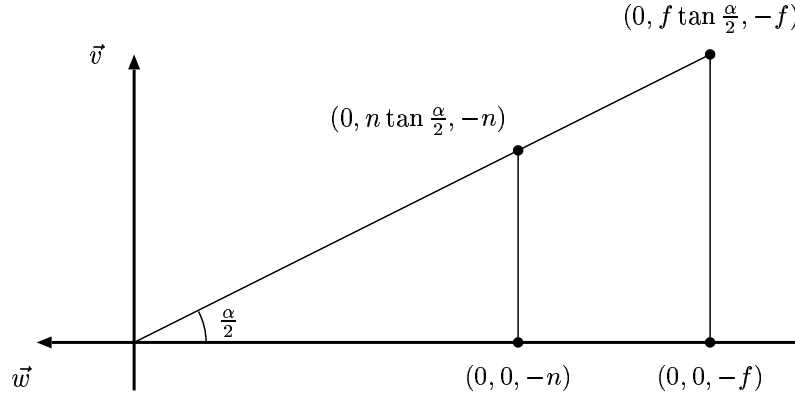
If we apply our transformation to these points, we obtain

$$\begin{bmatrix} 0 & n\tan\frac{\alpha}{2} & -n & 1 \end{bmatrix} P = \begin{bmatrix} 0 & n\tan\frac{\alpha}{2} & -n\frac{f-n}{f+n} + \frac{2nf}{f+n} & n \end{bmatrix}$$
$$= \begin{bmatrix} 0 & n\tan\frac{\alpha}{2} & n & n \end{bmatrix}$$

Similarly

$$\begin{bmatrix} 0 & f\tan\frac{\alpha}{2} & -f & 1 \end{bmatrix} P = \begin{bmatrix} 0 & f\tan\frac{\alpha}{2} & -f\frac{f-n}{f+n} + \frac{2nf}{f+n} & f \end{bmatrix}$$
$$= \begin{bmatrix} 0 & f\tan\frac{\alpha}{2} & f & f \end{bmatrix}$$

So after dividing by the fourth coordinate we see that the v coordinates of the points that lie on the line $\left\{ (0, -z\tan\frac{\alpha}{2}, z) \text{ for } w < 0 \right\}$ are all transformed to have a constant v value of

$$\tan\frac{\alpha}{2}$$

Since we wish v coordinates to be mapped to 1 (which is the v value of the top of the image space cube), we must multiply our projection matrix by a matrix that scales the x and y coordinates by

$$c = \frac{1}{\tan\frac{\alpha}{2}}$$

8

That is, we define

$$A_{\alpha,n,f} = S_{c,c,1}P$$

giving

$$A_{\alpha,n,f} = \begin{bmatrix} \cot\frac{\alpha}{2} & 0 & 0 & 0 \\ 0 & \cot\frac{\alpha}{2} & 0 & 0 \\ 0 & 0 & \frac{f+n}{f-n} & -1 \\ 0 & 0 & \frac{2fn}{f-n} & 0 \end{bmatrix}$$

This is the matrix that transforms the truncated viewing pyramid into image space.

The Inverse of the Viewing Transformation

The above transformation transforms the truncated viewing pyramid into image space. The inverse of this transformation does the opposite – transforms points from image space into the truncated viewing pyramid. This inverse is given by

$$A_{\alpha,n,f}^{-1} = \begin{bmatrix} \tan\frac{\alpha}{2} & 0 & 0 & 0 \\ 0 & \tan\frac{\alpha}{2} & 0 & 0 \\ 0 & 0 & 0 & +\frac{f-n}{2fn} \\ 0 & 0 & -1 & +\frac{f+n}{2fn} \end{bmatrix}$$

That this is actually the inverse can be easily checked.

Summary

We have developed a matrix that works in the local coordinates of the camera space and transforms the points of an object into image space. The matrix is applied in homogeneous space, so that the perspective divide must be done after the viewing matrix is applied.