On-Line Computer Graphics Notes

THE CAMERA TRANSFORM

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

Overview

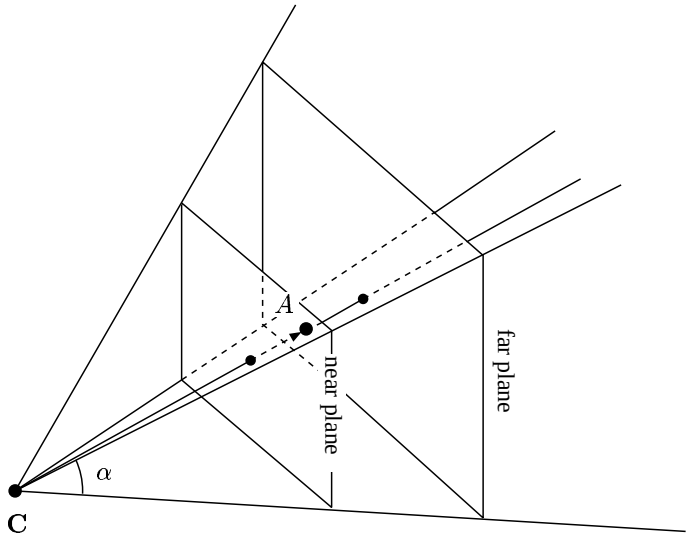
To understanding the rendering process, you must master the procedure that specifies a camera and then constructs a transformation that projects a three-dimensional scene onto a two-dimensional screen. This procedure has two several components: First, the specification of a camera model; second, the conversion of the scene's coordinates from Cartesian space to the space of the camera; and finally the specification of a viewing transformation that projects that scene into image space

---

The Camera Model

We specify our initial camera model by identifying the following parameters.

1. A scene, consisting of polygonal elements each represented by their vertices,
2. A point that represents the camera position — $\mathbf{C} = (x_c, y_c, z_c)$,
3. A point that represents the "center-of-attention" of the camera (i.e. where the camera is looking) — $\mathbf{A} = (x_a, y_a, z_a)$,
4. A field-of-view angle, α, representing the angle subtended at the apex of the viewing pyramid.
5. The specification of "near" and "far" bounding planes. These planes considered perpendicular to the direction-of-view vector at a distance of n and f from the camera, respectively.

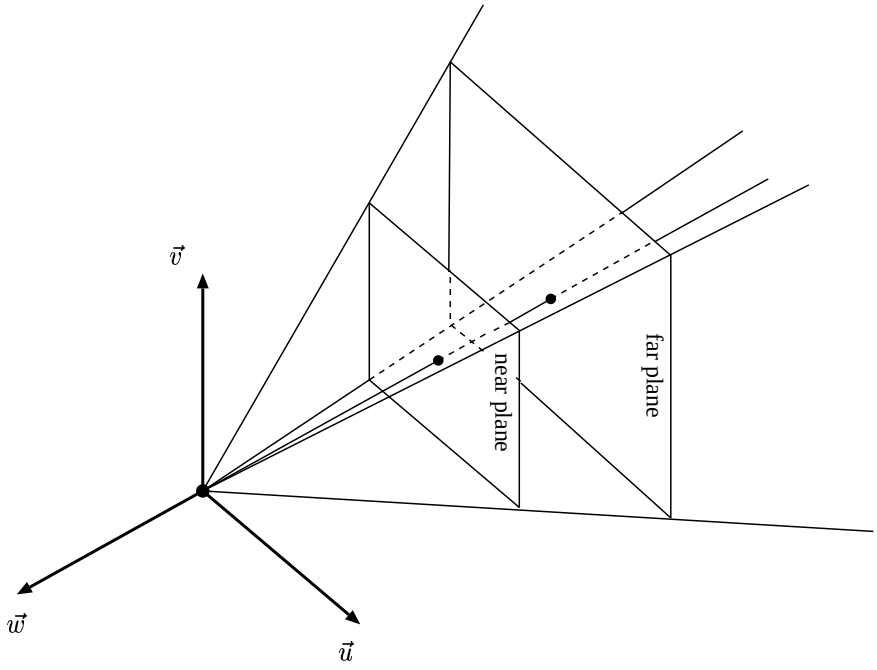


The specification of $\mathbf{C}$, $\mathbf{A}$ and α forms a viewing volume in the shape of a pyramid with the camera position $\mathbf{C}$ at the apex of the pyramid and the vector $\mathbf{A} - \mathbf{C}$ forming the axis of the pyramid. This pyramid is commonly referred to as *the viewing pyramid*. The specification of the near and far planes forms a truncated viewing pyramid which gives the region of space which contains the primary portion of the scene to be viewed[1]. The viewing transform, transforms this truncated pyramid onto the image space volume $-1 \leq x, y, z \leq 1$.

The Camera Transform

Given the definition of a camera $(\mathbf{C}, \mathbf{A}, \alpha, n, f)$, the camera transformation is a combination of a transform that first converts the coordinates of the Cartesian frame to the local coordinates of the camera's frame,

[1]We note that objects may extend outside the trunchated pyramid. In many situations polygons will lie between the near plane and the camera, or, in the distance beyond the far plane.



and second, applies the viewing transform. These two transformations are usually multiplied together to form a single 4×4 matrix that is applied to all points of the scene.

Defining a Frame at the Camera Position

The main idea here is to define a frame at the camera position. Given such a frame $\mathcal{F}_{\text{camera}} = (\vec{u}, \vec{v}, \vec{w}, \mathbf{C})$, we generate a transformation that converts the Cartesian Frame coordinates to the camera's frame.

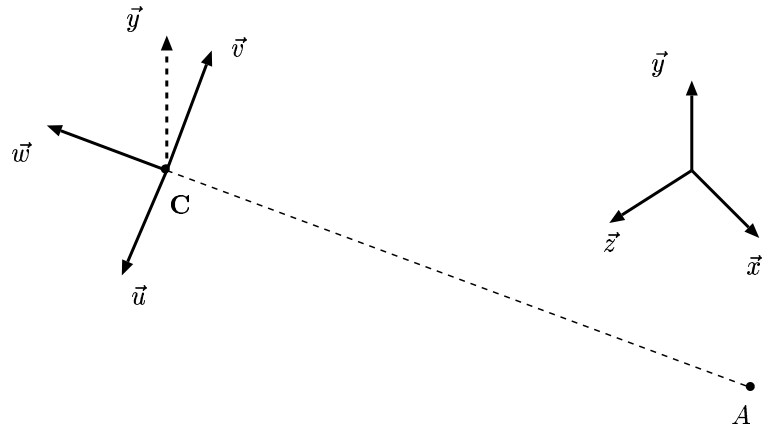
To define a frame at the camera position is easy – and there are actually an number of ways of doing this. One of the vectors is obvious – that is, we want

$$\vec{w} = \frac{\mathbf{C} - \mathbf{A}}{|\mathbf{C} - \mathbf{A}|}$$

(the transformed camera should be looking along the negative w axis).

In order to define the other vectors that make up the frame, we must make an assumption. We assume that the vertical direction of the camera must be in the plane defined by $\vec{w}$ and the vector $\vec{y} =< 0, 1, 0 >$. This frequently happens when you are taking a picture, if you think about it – and it actually fairly easy

to arrange. See the following figure for an illustration of this process. In the figure, the dotted line is the direction of view, and should be placed on the negative z axis by the transformation.



To define $\vec{u}$ and $\vec{v}$ we utilize the following steps

- define $\vec{u} = \frac{\vec{y}\times\vec{w}}{|\vec{y}\times\vec{w}|}$,
- define $\vec{v} = \vec{w} \times \vec{u}$. This insures that $\vec{v}$ is in the $\vec{y}$-$\vec{w}$ plane.

This also insures that the vectors are all unit vectors, and that they are mutually perpendicular.

We note that this works well, except when you wish to have the camera look in the direction $< 0, 1, 0 >$ or $< 0, -1, 0 >$. In these cases, either $\vec{y} = \vec{w}$ or $\vec{y} = -\vec{w}$ and $\vec{y} \times \vec{w} = \vec{0}$, and we cannot calculate a frame in this manner. However, we can utilize another vector as the "up direction" to utilize with $\vec{w}$ to obtain $\vec{u}$.

Calculating the Matrix

To calculate the actual matrix that implements the transformation, we can write each of the vectors $< 1, 0, 0 >$, $< 0, 1, 0 >$ and $< 0, 0, 1 >$ as a linear combination of $\vec{u}$, $\vec{v}$, and $\vec{w}$ (Since the vectors defining $\mathcal{F}_{\text{camera}}$ are linearly independent). In addition, we can write the vector $(0, 0, 0) - \mathbf{C}$ as a linear combination of $\vec{u}$, $\vec{v}$ and $\vec{w}$. Thus we can calculate the values $e_{i,j}$, where

$$
\begin{aligned}
< 1, 0, 0 > &= e_{1,1}\vec{u} + e_{1,2}\vec{v} + e_{1,3}\vec{w} \\
< 0, 1, 0 > &= e_{2,1}\vec{u} + e_{2,2}\vec{v} + e_{2,3}\vec{w} \\
< 0, 0, 1 > &= e_{3,1}\vec{u} + e_{3,2}\vec{v} + e_{3,3}\vec{w} \\
(0, 0, 0) &= e_{4,1}\vec{u} + e_{4,2}\vec{v} + e_{4,3}\vec{w} + \mathbf{C}
\end{aligned}
$$

These equations can be solved by Cramers Rule,: To obtain $< 1, 0, 0 >= e_{1,1}\vec{u} + e_{1,2}\vec{v} + e_{1,3}\vec{w}$, we have

$$e_{1,1} = \frac{(< 1, 0, 0 > \times \vec{v}) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}},$$
$$e_{1,2} = \frac{(\vec{u} \times < 1, 0, 0 >) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}}, \text{ and}$$
$$e_{1,3} = \frac{(\vec{u} \times \vec{v}) \cdot < 1, 0, 0 >}{(\vec{u} \times \vec{v}) \cdot \vec{w}}$$

To obtain $< 0, 1, 0 >= e_{2,1}\vec{u} + e_{2,2}\vec{v} + e_{2,3}\vec{w}$, we have

$$e_{2,1} = \frac{(< 0, 1, 0 > \times \vec{v}) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}},$$
$$e_{2,2} = \frac{(\vec{u} \times < 0, 1, 0 >) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}}, \text{ and}$$
$$e_{2,3} = \frac{(\vec{u} \times \vec{v}) \cdot < 0, 1, 0 >}{(\vec{u} \times \vec{v}) \cdot \vec{w}}$$

And to obtain $< 0, 0, 1 >= e_{3,1}\vec{u} + e_{3,2}\vec{v} + e_{3,3}\vec{w}$, we have

$$e_{3,1} = \frac{(< 0, 0, 1 > \times \vec{v}) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}},$$
$$e_{3,2} = \frac{(\vec{u} \times < 0, 0, 1 >) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}}, \text{ and}$$
$$e_{3,3} = \frac{(\vec{u} \times \vec{v}) \cdot < 0, 0, 1 >}{(\vec{u} \times \vec{v}) \cdot \vec{w}}$$

In addition, if $\vec{t} = (0, 0, 0) - \mathbf{C}$, then we have

$$e_{4,1} = \frac{(\vec{t} \times \vec{v}) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}},$$
$$e_{4,2} = \frac{(\vec{u} \times \vec{t}) \cdot \vec{w}}{(\vec{u} \times \vec{v}) \cdot \vec{w}}, \text{ and}$$
$$e_{4,3} = \frac{(\vec{u} \times \vec{v}) \cdot \vec{t}}{(\vec{u} \times \vec{v}) \cdot \vec{w}}$$

to get $\vec{t} = e_{4,1}\vec{u} + e_{4,2}\vec{v} + e_{4,3}\vec{w}$.

The matrix that convets the coordinates of objects in the frame $\mathcal{F}_{\text{camera}}$ into coordinates for the frame

$\mathcal{F}_{\mathrm{C}}$ is given by

$$\begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} & 0 \\ e_{2,1} & e_{2,2} & e_{2,3} & 0 \\ e_{3,1} & e_{3,2} & e_{3,3} & 0 \\ e_{4,1} & e_{4,2} & e_{4,3} & 1 \end{bmatrix}$$

Any point $\mathbf{P} = (x, y, z)$ can be written in the frame $\mathcal{F}_{\mathrm{C}}$ by

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} < 1, 0, 0 > \\ < 0, 1, 0 > \\ < 0, 0, 1 > \\ (0, 0, 0) \end{bmatrix}$$

But by the above calculations this is equal to

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} & 0 \\ e_{2,1} & e_{2,2} & e_{2,3} & 0 \\ e_{3,1} & e_{3,2} & e_{3,3} & 0 \\ e_{4,1} & e_{4,2} & e_{4,3} & 1 \end{bmatrix} \begin{bmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \\ \mathbf{C} \end{bmatrix}$$

which implies that the coordinate

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} & 0 \\ e_{2,1} & e_{2,2} & e_{2,3} & 0 \\ e_{3,1} & e_{3,2} & e_{3,3} & 0 \\ e_{4,1} & e_{4,2} & e_{4,3} & 1 \end{bmatrix}$$

is the coordinate of the point in the frame $\mathcal{F}_{\mathrm{camera}}$.

We note, that by our construction, the frame $\mathcal{F}_{\mathrm{camera}}$ is an orthonormal frame (all vectors are unit vectors and are mutually perpendicular) and in this case the equations above simplify tremendously. In particular,

all the denominators $\vec{u} \cdot (\vec{v} \times \vec{w}) = 1$, and we can simplify the numerators utilizing the identities

$$\vec{u} \times \vec{v} = \vec{w}$$
$$\vec{v} \times \vec{w} = \vec{u}$$
$$\vec{w} \times \vec{u} = \vec{v}$$

to obtain

$$e_{1,1} = \vec{u} \cdot < 1, 0, 0 >$$
$$e_{1,2} = \vec{v} \cdot < 1, 0, 0 >$$
$$e_{1,3} = \vec{w} \cdot < 1, 0, 0 >$$
$$e_{2,1} = \vec{u} \cdot < 0, 1, 0 >$$
$$e_{2,2} = \vec{v} \cdot < 0, 1, 0 >$$
$$e_{2,3} = \vec{w} \cdot < 0, 1, 0 >$$
$$e_{3,1} = \vec{u} \cdot < 0, 0, 1 >$$
$$e_{3,2} = \vec{v} \cdot < 0, 0, 1 >$$
$$e_{3,3} = \vec{w} \cdot < 0, 0, 1 >$$
$$e_{4,1} = \vec{u} \cdot \vec{t}$$
$$e_{4,2} = \vec{v} \cdot \vec{t}$$
$$e_{4,3} = \vec{w} \cdot \vec{t}$$

The first few of these are extremely simple, as, for example $\vec{u} \cdot < 1, 0, 0 >$ is just the first coordinate of $\vec{u}$, etc.

Overview

The camera transform is a Cartesian-frame-to-frame transform. This is combined with the viewing transform to give a transformation that converts a scene into image space.
