Raytracing and the camera matrix – a connection.

Sean Borman
Laboratory for Image and Signal Analysis
Department of Electrical Engineering
University of Notre Dame, Notre Dame, IN 46556, U.S.A.

June 2, 2003

1 Introduction

In this tutorial paper we illuminate (for lack of a better word) the relationship between the raytracing equations used in computer graphics rendering and the camera matrix used in the computer vision community. We derive the equations for each from first principles so that the equivalences among the parameters is clear.

2 A camera model for raytracing

In this section we derive the algebraic equations commonly used in the computer graphics technique of raytracing, where the simulation of geometric propagation of rays in an illuminated scene is used to produce highly realistic renderings of three-dimensional virtual scenes. Somewhat counter-intuitively, but for reasons of efficiency, in practical raytracer implementations, ray trajectories originate not from the illumination source but rather from the camera center, proceeding into the scene where the ray may encounter objects or illumination sources.

For the purpose of raytracing, it is convenient to describe the properties of the camera in a manner that simplifies the problem of projecting rays from the camera's center of projection (COP) into the three-dimensional scene. In particular, the following parameters are sufficient to describe the geometric projection characteristics of the pinhole camera: (1) the COP or camera center $\boldsymbol{C}$, (2) the principal, or optical axis $\hat{\boldsymbol{p}}$, (3) the horizontal axis of the image plane $\hat{\boldsymbol{u}}$, (4) the vertical axis of the image plane $\hat{\boldsymbol{v}}$, and (5) the focal length f. Vectors are typeset in bold face, and unit vectors are denoted with a hat. We make the simplifying assumption, which is true for most cameras, that $\hat{\boldsymbol{p}}$, $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ form an orthonormal set. The camera geometry and parameters are illustrated in Fig. (1), where the camera's COP is positioned in the three-dimensional world-coordinate systems denoted by the $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ axes. The principal axis $\hat{\boldsymbol{p}}$ intersects the focal-plane at the *principal point* at the origin of the (u, v)

focal-plane coordinate system, a distance f from the camera center $\boldsymbol{C}$ along the optical axis.

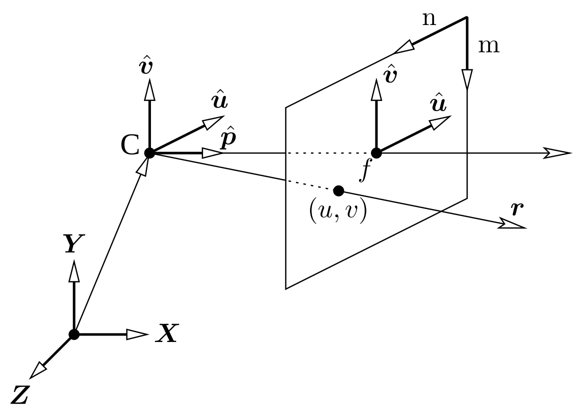

Figure 1: Pinhole camera geometry.

Raytracing involves casting rays originating at the center of projection $\boldsymbol{C}$ toward locations in the focal plane and on through to the scene where the ray interacts with objects. Let $\boldsymbol{r}$ denote a ray originating at the camera COP $\boldsymbol{C}$ and which passes through the location (u, v) in the focal-plane coordinate system. The trajectory of the ray is linear and thus may be easily described using it's origin $\boldsymbol{C}$, a scalar parameter $\gamma \geq 0$ and a direction vector $\boldsymbol{d}$ as,

$$\boldsymbol{r} = \boldsymbol{C} + \gamma \boldsymbol{d}. \tag{1}$$

Referring to Fig. (1), the direction of the ray may be conveniently described in terms of the camera focal-plane axes $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$ and the optical axis $\hat{\boldsymbol{p}}$ as,

$$\boldsymbol{r} = \boldsymbol{C} + \gamma \left[\, u\hat{\boldsymbol{u}} + v\hat{\boldsymbol{v}} + f\hat{\boldsymbol{p}} \,\right], \tag{2}$$

where u and v are the coordinates of the ray intersection with respect to the $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}})$ focal-plane coordinate system. In Eqn. (2) the ray intersects the focal-plane (at the principal point) for $\gamma = 1$. It is more convenient to have the parameter denote distance from $\boldsymbol{C}$. This is easily achieved by introducing the parameter $\alpha \geq 0$ and rescaling so that,

$$\boldsymbol{r} = \boldsymbol{C} + \alpha \left[\, \frac{u}{f}\hat{\boldsymbol{u}} + \frac{v}{f}\hat{\boldsymbol{v}} + \hat{\boldsymbol{p}} \,\right]. \tag{3}$$

Now with $\alpha = f$, the focal distance, the ray intersects the focal-plane as desired. Typically rays are cast through a discrete set of locations on the focal plane in form of a regular lattice consisting of M rows by N columns of pixels. It is convenient therefore to specify the ray intersection with the focal-plane in this pixel coordinate system. The ordering of the pixels is chosen so that the origin in pixel indices is located at the top left of the array when viewed from the camera center (see Fig. (2)). Note that the principal point, the location where the principal axis intersects the focal-plane, need not be located at the center of the sampling array. The (u, v)-coordinates of the center of the pixel in row m and column n for $m \in \{0, 1, \ldots, M-1\}$ and $n \in \{0, 1, \ldots, N-1\}$ are then given by,

$$u = (n_0 - n)S_x \quad \text{and} \quad v = (m_0 - m)S_y, \tag{4}$$

where m_0 and n_0 are respectively the row and column locations of the principal point in pixel coordinates and S_x and S_y are the sample spacing in the horizontal and vertical directions respectively. Combining Eqn. (3) with Eqn. (4) yields,

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{C} + \alpha\left[\frac{n_0 - n}{f}S_x\hat{\boldsymbol{u}} + \frac{m_0 - m}{f}S_y\hat{\boldsymbol{v}} + \hat{\boldsymbol{p}}\right] \\ &= \boldsymbol{C} + \alpha\left[-n\frac{S_x}{f}\hat{\boldsymbol{u}} - m\frac{S_y}{f}\hat{\boldsymbol{v}} + \left\{n_0\frac{S_x}{f}\hat{\boldsymbol{u}} + m_0\frac{S_y}{f}\hat{\boldsymbol{v}} + \hat{\boldsymbol{p}}\right\}\right]. \end{aligned} \tag{5}$$

Note that the terms surrounded by $\{\}$ in Eqn. (5) are constant with respect to the pixel indices (n, m). The rays resulting from a raster scan over the focal-plane pixel array may thus be efficiently computed using the increments $\frac{S_x}{f}\hat{\boldsymbol{u}}$ and $\frac{S_y}{f}\hat{\boldsymbol{v}}$.

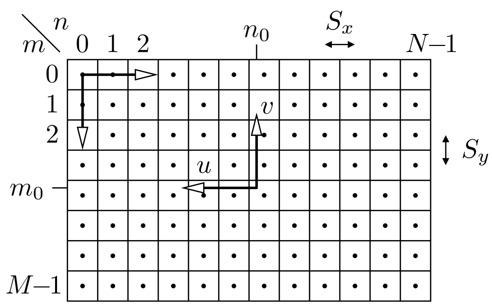

Figure 2: Focal-plane sensor array geometry.

2.1 Representing the raytracing equations in matrix form

It is instructive to restate the ray in Eqn. (3) in matrix-vector notation. For a given value of α, let the location of the ray $\boldsymbol{r}$ in world coordinates be given by $\boldsymbol{P} = [P_x, P_y, P_z]^T$ and the camera COP be $\boldsymbol{C} = [C_x, C_y, C_z]^T$. We may then write,

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix} + \alpha \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} u/f \\ v/f \\ 1 \end{bmatrix}, \quad \text{and using Eqn. (4),}$$

$$= \begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix} + \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} -S_x/f & 0 & n_0 S_x/f \\ 0 & -S_y/f & m_0 S_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha\, n \\ \alpha\, m \\ \alpha \end{bmatrix}. \tag{6}$$

This representation will be useful in light of the material in the following section.

3 The camera matrix model

In raytracing, a ray cast through a given location on the focal plane is associated with a location (typically on an object) in the world coordinate system. Raytracing may be thought of as a mapping from 2-space to 3-space given the scene.

The opposite problem, that of determining the intersection on the focal plane of a ray which intersects an object at a given location in the world coordinate system is addressed using a *camera matrix* and homogeneous coordinates. The camera matrix is ubiquitous (or dare we say *seen* everywhere) in the computer vision literature.

We shall initially assume that the camera's COP is located at the origin of the world coordinate system $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, with the camera coordinate system axes coincident with the world coordinate system axes. Thus the camera points in the positive $\boldsymbol{Z}$ direction. We also assume that the camera focal plane is located a distance f from the camera center along the camera principal axis. This arrangement is illustrated in Fig. (3). Consider the point $\boldsymbol{P}_C = [P_x, P_y, P_z]^T$ in the camera coordinate system. Using similar triangles we have,

$$\frac{P_x}{P_z} = \frac{u}{f} \quad \text{and} \quad \frac{P_y}{P_z} = \frac{v}{f} \tag{7}$$

or equivalently,

$$P_z u = f P_x \quad \text{and} \quad P_z v = f P_y. \tag{8}$$

This may be written using homogeneous coordinates using a linear *projection matrix* as,

$$\begin{bmatrix} P_z u \\ P_z v \\ P_z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix}_C , \tag{9}$$

where the subscript C draws attention to the fact that the point is located in the camera coordinate system. Using Eqn. (4) we may rewrite this in terms of the row and column focal-plane coordinates as,

$$\begin{bmatrix} P_z S_x (n_0 - n) \\ P_z S_y (m_0 - m) \\ P_z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix}_C , \tag{10}$$

which may in turn be rearranged into a more succinct form,

$$\begin{bmatrix} P_z\, n \\ P_z\, m \\ P_z \end{bmatrix} = \begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix}_C . \tag{11}$$

Next we discuss how to relax the assumption that the camera and world-coordinate systems are coincident and consider the case where the camera is positioned and oriented arbitrarily with respect to the world coordinate system. In this case, the camera and world coordinate systems are related by at most a translation and a rotation. In particular, let the camera center be located at position $\boldsymbol{C}$ in the world coordinate frame. Let the orientation of the orthonormal camera coordinate axes be given by the unit vectors $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{p}}$ in the world coordinate system, so that the camera axis vectors define the 3×3 rotation matrix $\boldsymbol{R} \doteq [\ \hat{\boldsymbol{u}}\ \hat{\boldsymbol{v}}\ \hat{\boldsymbol{p}}\ ]$. Under these assumptions, a point $\boldsymbol{P}_W$ in the world coordinate system is located at $\boldsymbol{P}_C$ in the camera coordinate system where,

$$\boldsymbol{P}_C = \boldsymbol{R}^{-1}[\boldsymbol{P}_W - \boldsymbol{C}]. \tag{12}$$

Note that since the rotation matrix $\boldsymbol{R}$ is unitary, $\boldsymbol{R}^{-1} = \boldsymbol{R}^T$. Eqn. (12) can be represented using matrices and homogeneous coordinates as,

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix}_C = \begin{bmatrix} -\ \hat{\boldsymbol{u}}^T\ - \\ -\ \hat{\boldsymbol{v}}^T\ - \\ -\ \hat{\boldsymbol{p}}^T\ - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix}_W . \tag{13}$$

Substituting Eqn. (13) into Eqn. (11) we find the projection of points in the world coordinate system for an arbitrarily positioned and oriented camera as,

$$\begin{bmatrix} P_z\, n \\ P_z\, m \\ P_z \end{bmatrix} = \begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\ \hat{\boldsymbol{u}}^T\ - \\ -\ \hat{\boldsymbol{v}}^T\ - \\ -\ \hat{\boldsymbol{p}}^T\ - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix}_W . \tag{14}$$

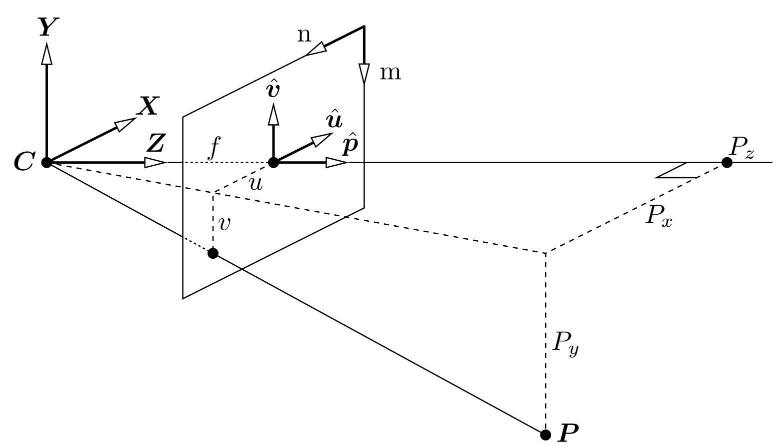

Figure 3: Camera in standard position.

4 Relating the raytracing and projection matrix representations

In Eqn. (6) we derived an expression for the ray from the camera center through the pixel in row m, column n in the focal-plane. We may rewrite this equation as,

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} - \begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix} = \alpha \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} u/f \\ v/f \\ 1 \end{bmatrix}. \tag{15}$$

Using homogeneous coordinates the translation by $-\boldsymbol{C}$ may be achieved using matrix multiplication as,

$$\begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} \alpha u/f \\ \alpha v/f \\ \alpha \end{bmatrix}. \tag{16}$$

Multiplying by the inverse of the rotation matrix yields,

$$\begin{bmatrix} - & \hat{\boldsymbol{u}}^T & - \\ - & \hat{\boldsymbol{v}}^T & - \\ - & \hat{\boldsymbol{p}}^T & - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha u/f \\ \alpha v/f \\ \alpha \end{bmatrix}. \tag{17}$$

Changing coordinate systems using using Eqn. (4) we have,

$$\begin{bmatrix} - & \hat{\boldsymbol{u}}^T & - \\ - & \hat{\boldsymbol{v}}^T & - \\ - & \hat{\boldsymbol{p}}^T & - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} -S_x/f & 0 & n_0 S_x/f \\ 0 & -S_y/f & m_0 S_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha\, n \\ \alpha\, m \\ \alpha \end{bmatrix}. \tag{18}$$

Noting that

$$\begin{bmatrix} -S_x/f & 0 & n_0 S_x/f \\ 0 & -S_y/f & m_0 S_y/f \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{19}$$

we finally have,

$$\begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} - & \hat{\boldsymbol{u}}^T & - \\ - & \hat{\boldsymbol{v}}^T & - \\ - & \hat{\boldsymbol{p}}^T & - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\, n \\ \alpha\, m \\ \alpha \end{bmatrix}. \tag{20}$$

Eqn. (20) is identical to Eqn. (14). Starting with the matrix-vector representation of the ray in Eqn. (3), we have shown the relationship between the raytracing camera model based on linear combinations of the camera axis vectors, as is typically used in the computer graphics community [1], and the standard projection matrix representation of the general pinhole camera in Eqn. (14) as is commonly found in the computer vision literature [2]. The two approaches are, not surprisingly, exact inverses of each other.

References

[1] J. Foley, A. van Dam, S. Feiner, and J. Hughes, *Computer Graphics – Principles and Practice*, Addison-Wesley Publishing Company, Second Edition in C, 1996.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.