CSE 252B: Computer Vision II

Lecturer: Serge Belongie
Scribe: Sameer Agarwal

LECTURE 1
Image Formation

1.1. The geometry of image formation

We begin by considering the process of image formation when a scene is viewed through a camera. The word camera has its origins in the Latin *camera* and the Greek *kamara*, both of which refer to a room or a chamber. In particular we will consider image formation through a **pinhole camera**. This is the dominant image formation model that is studied in computer vision.

A pinhole camera is a box in which one of the walls has been pierced to make a small hole through it. Assuming that the hole is indeed just a point, exactly one ray from each point in the scene passes through the pinhole and hits the wall opposite to it. This results in an inverted image of the scene, as can be seen in figure 1.

The inversion of the image is an annoyance and can be corrected for by instead considering a virtual image of the scene on a virtual plane parallel to the imaging plane but on the opposite side of the pinhole.

[1]Department of Computer Science and Engineering, University of California, San Diego.

March 29, 2004

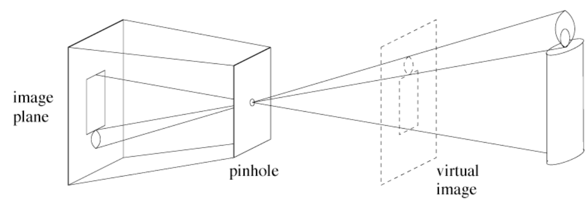

Figure 1. The pinhole imaging model, from *Forsyth & Ponce.*

Let us begin by considering a mathematical description of the imaging process through this idealized camera. We will consider issues like lens distortion subsequently.

The pinhole camera or the projective camera as it is known images the scene by applying a perspective projection to it. In the following we shall refer to scene coordinates with upper case roman letters, $\{X, Y, Z, \ldots\}$. Image coordinates will be referred to using lower case roman letters, $\{x, y, z, \ldots\}$. Vectors shall be denoted by boldfaced symbols, e.g., $\boldsymbol{X}$ or $\boldsymbol{x}$. (In class, when writing on the blackboard, I will put a tilde underneath the corresponding symbols to denote a vector.)

The scene is three dimensional, whereas the image is located in a two dimensional plane. Hence the perspective projection maps the 3D space to a 2D plane.

$$(X, Y, Z)^\top \xrightarrow{\textit{Projection}} (x, y)^\top$$

The equations of perspective projections are given by

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \tag{1.1}$$

here, f is the focal length of the camera, i.e., the distance between the image plane and the pinhole.

The process is illustrated in figure 2.

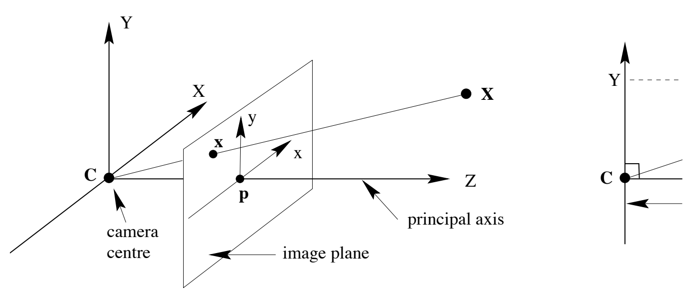

Figure 2. Image formation in a projective camera.

1.2. Homogeneous Coordinates

We now introduce a slightly modified representation of the points in the image plane. The usual representation of the image point

$$\boldsymbol{x} = (x, y)^\top$$

is referred to as the **inhomogeneous** representation of the point $\boldsymbol{x}$. The **homogeneous** representation of a point $\boldsymbol{x}$ is given by

$$\boldsymbol{x} = (x, y, 1)^\top$$

In fact, the homogeneous representation of a point maps it to an entire class of set of points:

$$(x, y) \leftrightarrow (\lambda x, \lambda y, \lambda), \qquad \forall \lambda \neq 0$$

in particular,

$$(x/z, y/z) \leftrightarrow (x, y, z)$$

Homogeneous coordinates encode the invariance of all points along a line and its projection.

The equation of a line

The equation of a line

$$ax + by + c = 0$$

can be rewritten using homogeneous coordinates

$$\boldsymbol{x}^\top \boldsymbol{l} = 0, \quad \text{where} \quad \boldsymbol{l} = (a, b, c)^\top$$

The equation of a conic

The general conic in 3 dimensions is given by

$$ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0$$

which can be written using 2D homogeneous coordinates as

$$\boldsymbol{x}^\top C \boldsymbol{x} = 0$$

where

$$C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}$$

The matrix C has 6 unique entries. The equation remains invariant under a scaling of all coeffecients, hence a conic has $6 - 1 = 5$ degrees of freedom.

1.2.1. What if the third coordinate is zero?

Up till now we have considered homogeneous points of the form

$$\boldsymbol{x} = (x, y, z)^\top \quad z \neq 0$$

What about points for which $z = 0$? This issue goes to the heart of the fact that we are dealing with the projective plane $\mathbb{P}^2$ instead of the Euclidean plane $\mathbb{R}^2$. The primary distinction between the two is that in $\mathbb{R}^2$ all pairs of lines intersect except for the ones that are parallel; in $\mathbb{P}^2$ there is no such restriction, and all pairs of lines interesect. Parallel lines lines intersect in *points at infinity* (also known as *ideal points*) and these points have the form

$$(x, y, 0)^\top$$

Consider the two lines given by

$$\boldsymbol{l}_1 = (a_1, b_1, c_1)^\top \tag{1.2}$$

$$\boldsymbol{l}_2 = (a_2, b_2, c_2)^\top \tag{1.3}$$

The intersection of these two lines is given by their vector cross product,

$$\boldsymbol{x} = \boldsymbol{l}_1 \times \boldsymbol{l}_2 \tag{1.4}$$

$$= \begin{bmatrix} 0 & -c_1 & b_1 \\ c_1 & 0 & -a_1 \\ -b_1 & a_1 & 0 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = \widehat{l_1}\boldsymbol{l}_2 \tag{1.5}$$

Here, $\widehat{l_1}$ is a skew symmetric matrix that converts the vector cross product into a matrix-vector product. The symbol $\widehat{l}$ is referred to as the wedge or hat of $\boldsymbol{l}$ and matrices of this form constitute a group denoted by $so(3)$, the group of 3×3 skew symmetric matrices.

Now we can return to the intersection points of parallel lines in $\mathbb{P}^2$. Given a line

$$\boldsymbol{l}_1 = (a, b, c)^\top$$

a line parallel to it is given by

$$\boldsymbol{l}_2 = (a, b, c')^\top$$

and the intersection is now given by

$$\boldsymbol{l}_1 \times \boldsymbol{l}_2 = (bc' - cb, ac - ac', 0)^\top \tag{1.6}$$

$$= (c - c')(b, -a, 0)^\top \tag{1.7}$$

$$\sim (b, -a, 0)^\top \tag{1.8}$$

The symbol $\sim$ refers to projective equivalence and $(b, -a, 0)^\top$ is an ideal point. Thus the set of ideal points (i.e., points at infinity) is the set of points where parallel lines intersect.

Duality

Among the many fascinating properties that the projective plane has, perhaps the most important one is that of *duality.* More specifically, in $\mathbb{P}^2$ points and lines are duals of each other. By this we mean that for any statement involving points and/or lines that holds true in $\mathbb{P}^2$, a corresponding version of it in which the word 'point' has been substituted for 'line' and vice versa also holds true.

As an example of this, the point of intersection of two lines is their cross product, the dual of which states that the line passing through any two points is given by their cross product, i.e.,

$$\boldsymbol{l} = \boldsymbol{x}_1 \times \boldsymbol{x}_2$$

This result along with the definition of points at infinity leads us to the definition of the line at infinity $\boldsymbol{l}_\infty$.

Consider two points at infinity

$$\boldsymbol{x}_1 = (x_1, y_1, 0)^\top \tag{1.9}$$

$$\boldsymbol{x}_2 = (x_2, y_2, 0)^\top \tag{1.10}$$

Now consider the line passing through these two points and denote it by $\boldsymbol{l}_\infty$.

$$\boldsymbol{l}_\infty = \boldsymbol{x}_1 \times \boldsymbol{x}_2 \tag{1.11}$$

$$= (0, 0, x_1 y_2 - y_1 x_2)^2 \tag{1.12}$$

$$\sim (0, 0, 1)^\top \tag{1.13}$$

The line $(0, 0, 1)^\top$ is the *line at infinity* and it passes through all the points at infinity.

A model for $\mathbb{P}^2$ in $\mathbb{R}^3$

As shown in Figure 3, points and lines on the image plane $\boldsymbol{\pi}$ can be identified with a line and planes, respectively, in $\mathbb{R}^3$.

Any point along the ray from the optical center through the point $\boldsymbol{x}$ on $\boldsymbol{\pi}$ projects to the same point on $\boldsymbol{\pi}$. Ideal points correspond to rays lying in the (x, y) plane.

Similarly, any line lying in the plane that intersects $\boldsymbol{\pi}$ in the line $\boldsymbol{l}$ projects to the same line on $\boldsymbol{\pi}$. The vector $\boldsymbol{l}$ is the normal vector of this plane in $\mathbb{R}^3$. Strictly speaking, a different symbol should be used for the image of the line on $\boldsymbol{\pi}$, but the intended meaning of $\boldsymbol{l}$ will be clear from the context. The line at infinity $\boldsymbol{l}_\infty$ corresponds to the plane with normal vector $(0, 0, 1)^\top$, i.e., the (x, y) plane.

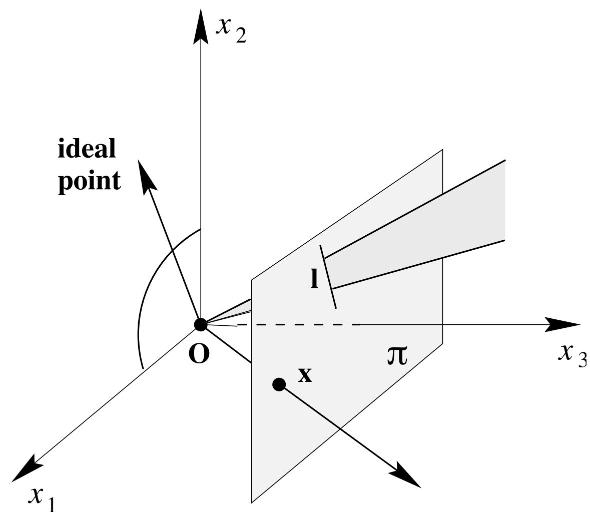

Figure 3. An illustration of how lines in $\mathbb{R}^3$ map to points in $\mathbb{P}^2$ and the projective line $\boldsymbol{l}$ corresponds to the plane in $\mathbb{R}^3$ with normal vector $\boldsymbol{l}$. (Figure from *Hartley & Zisserman.*)

Intrinsic/Extrinsic Parameters of a Camera

Let us now consider a more general image formation model that accounts for a number of factors affecting image formation in real cameras.

Let us begin by stating the general image formation model. The following equation maps the real world point $\boldsymbol{X}_0$ in homogeneous coordinates to its projection $\boldsymbol{x}'$ also in homogeneous coordinates.

$$\lambda \underbrace{\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}}_{\boldsymbol{x}'} = \underbrace{\begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Pi_0} \underbrace{\begin{bmatrix} R & \boldsymbol{T} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{g} \underbrace{\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix}}_{\boldsymbol{X}_0} \tag{1.14}$$

The matrix Π_0 is the canonical projection matrix. The matrix K consists of the intrinsic parameters of the camera. Here f is the focal length of the camera, s_x and s_y give the relative aspect of each pixel. o_x and o_y specify the coordinates of the image center. s_θ is the skew in the shape of the pixel, i.e., its deviation from an axis aligned rectangle.

The matrix g defines the pose of the camera. The elements of g constitute the extrinsic parameters of the camera. Here, R is a 3×3 rotation matrix and $\boldsymbol{T}$ is a vector in $\mathbb{R}^3$.

The task of camera calibration is to estimate the K and g matrices.

The Camera Pose

We consider the problem of specifying the rotation matrix R. R represents an arbitrary rotation in 3 dimensions. In fact, the set of such matrices form a group under matrix multiplication, and it is known as $SO(3)$, the Special Orthogonal group in 3 dimensions.

The group is orthogonal because

$$R^\top R = RR^\top = I \tag{1.15}$$

and special because

$$\det(R) = +1 \tag{1.16}$$

If the determinant of R is allowed to be -1, that would allow for mirror reflections. The pose of the camera $(R, \boldsymbol{T})$ is capable of representing arbitrary rigid motion in 3 dimensions. The set of poses of a camera also forms a group known as the Special Euclidean group in 3 dimensions, $SE(3)$. It is the cartesian product of the group $SO(3)$ with the Euclidean space $\mathbb{R}^3$:

$$SE(3) = SO(3) \times \mathbb{R}^3$$

There are many ways of computing the rotation matrix corresponding to a rotation, e.g., quarternions. Here we will use the Axis and Angle method. It allows for a particularly simple formalism.

Any arbitrary rotation in $\mathbb{R}^3$ can be represented as a unit vector $\boldsymbol{\omega}$ that represents the axis of rotation and a scalar θ that indicates the amount of rotation (in radians) around $\boldsymbol{\omega}$. The relation between the rotation matrix R and $(\boldsymbol{\omega}, \theta)$ can be stated as

$$R = e^{\widehat{\omega}\theta} \tag{1.17}$$

with

$$\widehat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

To exponentiate this matrix we use the following compact formula.

1.2.2. Rodrigues' Formula

Rodrigues' formula for $\theta \in \mathbb{R}$ and $\boldsymbol{\omega} \in \mathbb{R}^3$ with $\|\omega\| = 1$ is

$$e^{\widehat{\omega}\theta} = I + \widehat{\omega}\sin\theta + \widehat{\omega}^2(1 - \cos\theta) \tag{1.18}$$

For $\boldsymbol{\omega}$ of arbitrary (but nonzero) norm, the corresponding expression is:

$$e^{\widehat{\omega}} = I + \frac{\widehat{\omega}}{\|\omega\|}\sin\left(\|\omega\|\right) + \frac{\widehat{\omega}^2}{\|\omega\|^2}\left(1 - \cos(\|\omega\|)\right) \tag{1.19}$$

where $\|\omega\| = \theta$.

1.2.3. Appendix: Intuition about crossing lines and points

1.2.3.1. *Thoughts from Manmohan Chandraker, Spring 2004.* Several people asked about the intuition of crossing lines to get the intersection and crossing points to get the line passing through them when using homogeneous coordinates. Manmohan offered the following explanation.

First, consider the case of crossing two lines $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$. Since $\boldsymbol{l}_1 \times \boldsymbol{l}_2$ is orthogonal to both $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ (by definition of the cross product), we know $(\boldsymbol{l}_1 \times \boldsymbol{l}_2)^\top \boldsymbol{l}_i = 0, i = 1, 2$. If we let

$$\boldsymbol{x} = \boldsymbol{l}_1 \times \boldsymbol{l}_2$$

then we may write $\boldsymbol{x}^\top \boldsymbol{l}_i = 0, i = 1, 2$. This is in the form of the equation for a line – in particular, this expresses the set of all points $\boldsymbol{x}$ in the plane that have zero distance from lines $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$. If $\boldsymbol{x}$ has zero distance from both lines, then it must lie at their intersection.

The case of crossing two points to get the line passing through them follows a similar argument, with the roles of points and lines swapped.

1.2.3.2. *Further explanation from Will Chang, Spring 2007.* We learned in class that we can obtain the intersection of two lines $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ in the projective plane $\mathbb{P}^2$ simply by taking the cross product $\boldsymbol{l}_1 \times \boldsymbol{l}_2$. Similarly, for two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we can obtain the join of the two points (the line that goes through the points) by taking the cross product $\boldsymbol{x}_1 \times \boldsymbol{x}_2$. Why is the cross product involved to calculating these quantities? Here is a geometric explanation.

There is a useful connection between $\mathbb{P}^2$ and $\mathbb{R}^3$: each point $p \in \mathbb{P}^2$ is identified with a line passing through the origin in $\mathbb{R}^3$, and a line $\boldsymbol{l} \in \mathbb{P}^2$ is identified with a plane passing through the origin in $\mathbb{R}^3$. The line $\boldsymbol{l}$ in this case is simply represented by a normal vector $\boldsymbol{l} = (a, b, c)^\top$ which is orthogonal to the plane, and, without loss of generality, we can represent each point p in $\mathbb{P}^2$ with a unit vector along the direction of the corresponding line in $\mathbb{R}^3$.

With this in mind, consider the join operation between two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{P}^2$. This is illustrated in Figure 4. Viewing the situation in $\mathbb{R}^3$, the join operation takes the unit vectors corresponding to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to produce a new vector $\boldsymbol{n}$ which is orthogonal to both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Here, $\boldsymbol{n}$ is actually orthogonal to all vectors spanned by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, i.e., the plane in $\mathbb{R}^3$ passing through $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Thus, $\boldsymbol{n}$ is the normal vector corresponding to this plane and thus describes the line $\boldsymbol{l}$ in $\mathbb{P}^2$ that passes through both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Now consider the intersection between two lines $\boldsymbol{l}_1, \boldsymbol{l}_2 \in \mathbb{P}^2$. This is illustrated in Figure 5. Again, the situation in $\mathbb{R}^3$ is that we have two planes in $\mathbb{R}^3$ corresponding to $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$, and the point of intersection $\boldsymbol{x} \in \mathbb{P}^2$ corresponds to the line formed by intersecting the two planes. Here, $\boldsymbol{x}$ must be orthogonal to both the normals for $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$, since $\boldsymbol{x}$ corresponds to a line

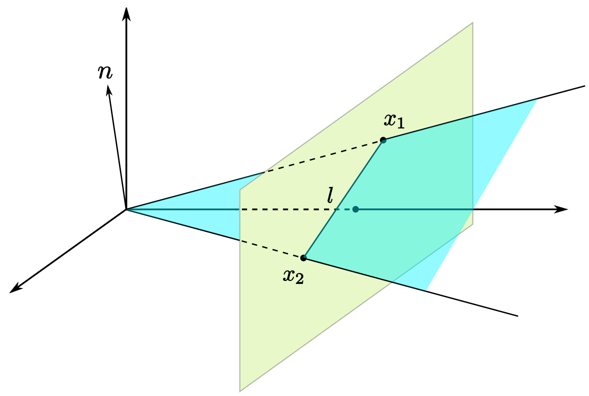

Figure 4. Illustrating the join of two points in $\mathbb{P}^2$.

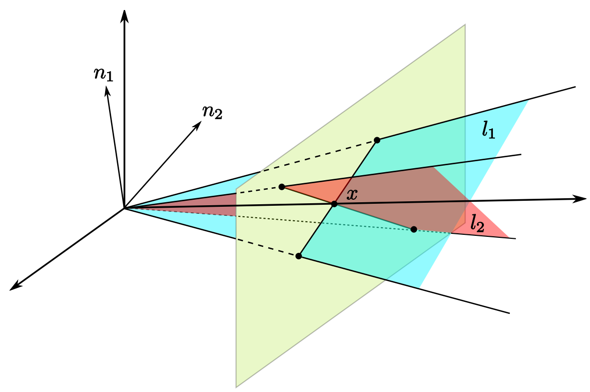

Figure 5. Illustrating the intersection of two lines in $\mathbb{P}^2$.

that lies in both planes. Therefore, $\boldsymbol{x}$ is obtained by taking the cross product of $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$, producing a direction that is orthogonal to both normals.