Geometry

Shira Kritchman

February 20, 2020

1 Camera Models

A camera is an optical device that transforms, or projects, a 3D scene to a 2D image. In this discussion, a 3D scene is simply a set of points in $\mathbb{R}^3$, and a 2D image is a set of points in 2D. The camera projection can involve complicated optical effects, but here we use simplified models.

1.1 Coordinates and Parameters

To describe the camera projection, we first need to agree on coordinate systems. We denote points in 3D by capital letters (usually P and Q), and their corresponding points in the image by their lowercase version (p and q).

The camera has an intrinsic 3D coordinate system, which is set by its focal center $O = (0, 0, 0)$ and the direction of its optical axis, Z (and arbitrary choice for X and Y). If we are given points in another 3D coordinate system, we usually should apply a linear transformation before we do anything else (as you did in question 3).

We identify the image with a 2D affine subspace embedded in $\mathbb{R}^3$, denoted by Π. This plane is perpendicular to the Z axis, and is at distance f from the focal center.

The 2D coordinates describing the image are chosen such that $(0, 0)$ corresponds to the 3D point $(0, 0, f)$, and the x and y directions correspond to these directions in 3D camera coordinates.

Later on we will introduce **homogeneous coordinates**.

1.2 Orthography

In this model, the camera's projection is simply

$$(X, Y, Z) \Rightarrow (x, y) = (X, Y)$$

For example, the point $(2, 3, 5)$ is projected to $(2, 3)$. The point $(2, 3, 6)$ is also projected to $(2, 3)$.

We also have **scaled orthographic projection**:

$$(X, Y, Z) \Rightarrow (x, y) = s(X, Y)$$

with $s = f/Z_0$, and Z_0 is some typical distance for the scene.

The advantage of this model is that it is simple. The disadvantage is that this is certainly not how a real camera works! But, in some cases, depending on the scene and on our application, this is good enough, and will result only in a slight error compared to the more complicated perspective model, which we discuss next. This happens when the variance of the depth of the points (i.e., their distance from Π, or their Z value) is small relative to the average depth. This can happen when the object is far from the camera, or if the object is very flat, and perpendicular to Π.

1.3 Perspective

This is a much more accurate model for a camera projection. It is derived from the *pinhole camera model*. In this model, a point $P = (X, Y, Z)$ is projected to the intersection of the ray OP with the plane Π. From simple geometrical consideration it follows that $p = \frac{f}{Z}P$, so:

$$(X, Y, Z) \Rightarrow (x, y) = \frac{f}{Z}(X, Y)$$

Note that if $Z \approx Z_0$, scaled orthographic projection gives a similar result.

2 Projective Geometry

Under the perspective model, the entire ray OP is mapped to the same point on Π. Therefore, there is a correspondence between points on Π and rays. And what about points $P = (X, Y, 0)$? In this case the ray OP is defined, but it doesn't intersect Π, so which point does it correspond to? To make this correspondence complete, we add to Π more points, which we call **ideal points**. They don't match any 2D points, but we can think of them as points that lie at the end of these infinite rays. To explore this set (the regular 2D points plus the ideal points) we use Projective geometry.

2.1 Homogeneous Coordinates

In Projective geometry we use homogeneous coordinates. A point is represented by a vector of length 3, and (x, y, z) and $\lambda(x, y, z)$ represent the same point. $(x, y, 0)$ is an ideal point.

A point $(2, 3)$ in the image, is now $(2, 3, 1)$ in homogeneous coordinates, and also $(4, 6, 2)$. And vice versa, a point $(4, 6, 2)$ in the Projective plane, is the point $(2, 3)$ in the image. The point $(4, 5, 0)$ is an ideal point in the Projective plane, and doesn't correspond to a point in the 2D image.

2.2 Lines

A straight line in the Projective plane is a set of points. Since points correspond to rays, a line corresponds to a set of rays, each going from O to the line (and beyond). Together these rays form a plane Γ. Γ goes through O and intersects Π at the line. We represent the line by the normal to Γ. This is again an homogeneous representation: ℓ and $\lambda\ell$ are the same line. To sum up, the line ℓ is the set of points p such that $p^T\ell = 0$. For example, if $\ell = (1, 1, 1)$, then $p = (1, 1, -2)$ is on the line. The 2D Cartesian location of this point in the image $(-0.5, -0.5)$.

And now, what about the line $\ell = (0, 0, 1)$? This corresponds to the plane $Z = 0$, which doesn't intersect Π. We call this the **line at infinity**, and it includes the set of all ideal points.

The intersection p of two lines ℓ_1 and ℓ_2 is given by their cross product, $p = \ell_1 \times \ell_2$. In the exercise you showed that parallel lines in the image (parallel in Euclidean geometry), intersect at ideal points. In projective geometry, any two lines intersect. Similarly, the line between two points p_1 and p_2 is given by $\ell = p_1 \times p_2$. You can verify that the line between two ideal points is the line at infinity.

3 Two-View Geometry

Suppose we take pictures of a 3D scene with two cameras, A and B, with different location and direction. Given a point in space, it is projected on both images, let's call the projections p and q. We say that p and q are a **matching pair**.

Each camera has its own intrinsic coordinate system, and the two are related by rotation and translation. If a point in space is P in camera A, then in B it will be $Q = RP + t$ (for some R and t).

Now, assume we get two images. We have a point p in image A, and we want to find its match q in image B. If we know P, then it is easy to infer both p and q. But if we don't have the 3D scene, what can we say about q? Knowing p means knowing that P is somewhere on the ray λp, therefore q is somewhere on the projection of λp on Π_B. This projection is a line in image B. Let's denote it by ℓ_p. This line is called an **epipolar line**.

q is on ℓ, so $q^T \ell_p = 0$. To find ℓ_p, all we need is to find two points on it, and compute their cross product. Well, any point on the 3D ray λp, will be projected to ℓ, so it's part of the plane Γ we mentioned before. Since we work in homogeneous coordinates and projective geometry, we say that such a point is on ℓ. One such point is p, and the other is O. But we have to be careful here - we need to translate them to the coordinate system of B. Thus p becomes $Rp + t$ and O becomes $RO + t = t$. Therefore $\ell_p = t \times (Rp + t) = t \times Rp = [t]_\times Rp$. Define the **essential matrix** $E = [t]_\times R$, then $q^T Ep = 0$. This is true for any matching pair of points in the images A

and B. Let's verify this:

$$EP = [t]_\times RP = [t]_\times (RP + t - t) = t \times Q - t \times t = t \times Q$$

and therefore

$$Q^T EP = Q^T (E \times P) = 0$$

and since $p \propto P$ and $q \propto Q$ we get

$$q^T Ep = 0.$$

Now, what about the point $O = (0, 0, 0)$, the focal point of camera A? This point belongs to all rays, so it should belong to all epipolar lines in image B. This is called the **epipole**. In B it is equal to $RO + t = t$.

To infer E, we find pairs of corresponding points between the images. Each pair gives us a scalar constraint on E, and given enough pairs (8/5, depending on our method) we can discover E.

3.1 Homography

An homography is an isomorphism between projective planes. An homography is (1) invertible and (2) transforms lines to lines. Two images are related by an homography iff $q \propto Hp$ for some *invertible* matrix H. Note that it still holds that $q^T Ep = 0$.

3.2 Planar Scene

Assume that all the points in the scene lie on a plane. This means that for some n and d they all satisfy $n^T P = d$. Therefore $1 = \frac{n^T P}{d}$. We know that $Q = RP + t$, and thus:

$$Q = RP + t = RP + t\frac{n^T P}{d} = \left(R + \frac{tn^T}{d}\right) P$$

Define $H = R + \frac{tn^T}{d}$, then we get $Q = HP$, and $q \propto Hp$. We conclude that two images of a planar scene are related by an homography.

Note that we also have to make sure that H is invertible. It is possible that it's not invertible. Imagine you have a scene of three points not on a line. From most points of view, a camera will see a triangle. But it is possible to point a camera such that it will see the three point collinear. In this case, the transformation between the two images is not invertible, and is not considered an homography.