CS 280 Computer Vision
Spring 2018 Efros, Malik, Yu Lecture 3

1 Lecture

Transformations

Orthogonal Transformations

"Orthogonal" means "perpendicular," i.e. it refers to *angles*. Orthogonal transformations are about preserving angles. Often we use inner products to talk about angles in a more generalized setting, and here is no exception: in order to preserve angles, we actually focus on preserving inner products.

An orthogonal transform A (as a matrix) takes vectors v to Av. (Clearly, this is also a linear transformation.) If we have another vector w, it will go to Aw. For orthogonal transformations, we want to say that the inner product is preserved, i.e. $v^T w = (Av)^T Aw$. This is the primary definition of an orthogonal transformation, and it must remain true for all (v, w). We must additionally have $A^T A = I$ and $A^T = A^{-1}$.

A's column vectors must be orthonormal: any of these vectors dotted with itself should be 1 (length 1), while any vector dotted with a different vector should be 0. Accordingly, the determinant must be ± 1. If $\det A = +1$, it corresponds to a rotation. If $\det A = -1$, it corresponds to a reflection.

$$\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

is a 2×2 reflection matrix (about the line with angle $\theta/2$). Its determinant is -1. (Recall that $\sin^2\theta + \cos^2\theta = 1$.) Composing two reflections gives us a rotation. Composing two rotations also gives us a rotation.

For a 3D rotation, there are three pairwise dot product constraints and three length constraints. Even though there are nine slots in the matrix, rotations in 3D really only have three parameters: two parameters for the first column vector, one for the second (*which* plane perpendicular to the first vector?), and none for the third.

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is a 3D rotation about the z-axis by θ. The z-coordinate stays constant.

An **isometry** (Euclidean transformation) is an orthogonal transformation followed by a translation (e.g. $\psi(a) = Aa + t$). Note that the composition of two isometries is also an isometry. If we have $\psi_1(a) = A_1 a + t_1$ and $\psi_2(a) = A_2 a + t_2$, then $\psi_1 \circ \psi_2(a) = (A_1 A_2)a + (A_1 t_2 + t_1)$.

Affine Transformations

An **affine transformation** is a nonsingular linear transformation followed by a translation: $\psi(a) = Aa + t$. It is similar to a Euclidean transformation; the difference is in the matrix A. In a Euclidean transformation, A is required to be an orthogonal matrix (falling into either the rotation or the reflection family). Affine transformations have no such constraint. The only restriction we impose is that A should be nonsingular, i.e. the determinant should be nonzero. Thus, the set of Euclidean transformations is a *subset* of the set of affine transformations.

If you compose two Euclidean transformations, you get a Euclidean transformation. Likewise, if you compose two affine transformations, you get an affine transformation. Each is a *group*. Incidentally, any affine transformation has an inverse.

There are six degrees of freedom for affine transformations (four for A, two for t). In 2D, we require $1+2=3$ parameters to specify an isometry and $4+2=6$ parameters to specify an affine transform. In 3D, we require $3+3=6$ parameters to specify an isometry and $9+3=12$ parameters to specify an affine transform. (The first part of these sums relates to the A matrix; the second relates to the translation.) Affine transformations are strictly more general.

Affine transformations include rotation, anisotropic scaling [scaling in one/a few direction(s), but not in all], and shearing.

Note: inner products (and by extension angles) are no longer being preserved. Areas and distances between pairs of points are also not being preserved. On the other hand, affine transformations do preserve parallelism (lines which are parallel remain parallel) and midpoints. These are examples of **invariants** – properties which are preserved.

As a result of these invariants, the concept of a [triangle's] median is an affine concept, because midpoints are an affine concept and intersections are an affine concept. (If lines intersect before an affine transformation, they will also intersect after it.) Since we can make any triangle into any other triangle via an affine transformation, we can turn any triangle into an equilateral triangle and show by symmetry that all three of its medians intersect. Thus, the three medians of any triangle must intersect.

Question: is there anything more general than affine transformations?

Projective Transformations

Of course. Remember the transformation of perspective, where lines which are parallel in the real world may no longer be parallel in an image (instead they can map to lines that intersect, e.g. at a vanishing point). Such perspective projections are instances of **projective transformations**, or linear transformations that use homogeneous coordinates. (Projective transformations also include affine transformations.)

Homogeneous coordinates: instead of using n coordinates for n-dimensional space, we use $n+1$ coordinates. The central idea is that we need a way to bring in infinities. Why? Because in perspective projection, parallel lines can converge in the projection to a vanishing point at infinity.

In general, a slight problem with Euclidean geometry is that there are two kinds of relationships between lines: some lines are parallel to each other and then meet (at infinity), and some lines simply intersect. To distinguish between the two, we want to make the use of infinity legitimate.

In $\mathbb{R}^1$, we will have $\mathbb{R}^1 \cup \{\text{point at } \infty\} = p^1$ the projective line. In $\mathbb{R}^2$, we will have $\mathbb{R}^2 \cup \{\text{line at } \infty\} = p^2$ the projective plane. And in $\mathbb{R}^3$, we will have $\mathbb{R}^3 \cup \{\text{plane at } \infty\} = p^3$ the projective 3-space. We want a way to work with these constructs mathematically, which is – for p^1, instead of using one coordinate we will use two. For p^2, instead of using two coordinates we will use three. And for p^3, instead of using three coordinates we will use four. In general, a projective n-dimensional space will require us to use $n+1$ coordinates.

The key rule is that if one vector is just a scalar multiple of the other, they correspond to the same point in p^{n-1}. Therefore there are only $n-1$ degrees of freedom. We can think of this as line(s) through the origin in n-dimensional space. To represent points in a line, we use lines through the origin in a plane. To represent points in a plane, we use lines through the origin in 3-dimensional space.

Given a set of homogeneous coordinates, we can canonicalize it (return to the Euclidean coordinates) by dividing by the final (homogeneous) coordinate. But if z is 0, what do we do? We realize that the canonical Euclidean point is *going off to infinity*. For example, an infinite point can be expressed as

$$\begin{bmatrix} x \\ 0 \end{bmatrix}$$

in 1D (on a projective line). Note that there is only one such point at infinity.

On the projective plane, any finite point can be represented as $\begin{bmatrix}\lambda_x & \lambda_y & \lambda\end{bmatrix}^T$, while any infinite point can be represented as $\begin{bmatrix}x & y & 0\end{bmatrix}^T$. The latter case represents a line at infinity. Different ratios $x : y$ give rise to different points.

A Euclidean line is written as $a_1 x + a_2 y + a_3 = 0$. That same line can be expressed in homogeneous coordinates as $a_1 \frac{x_1}{x_3} + a_2 \frac{x_2}{x_3} + a_3 = 0 \ \rightarrow \ a_1 x_1 + a_2 x_2 + a_3 x_3 = 0 \ \rightarrow \ a \cdot x = 0$.

(We can see that a point x lies on the line a if $a \cdot x = 0$.)

Where do the lines $\begin{bmatrix}a_1 & a_2 & a_3\end{bmatrix}^T$ and $\begin{bmatrix}b_1 & b_2 & b_3\end{bmatrix}^T$ intersect? (Note that in projective geometry, two lines will always intersect.) Given two vectors, we need a vector which is *perpendicular to both of them.* We acquire this from the cross-product: a point which lies on both lines must be in the cross-product of the lines' corresponding vectors.

In summary, we now have a rigorous mathematical way of dealing with points at infinity. Also, all lines intersect, and that intersection can be computed trivially.