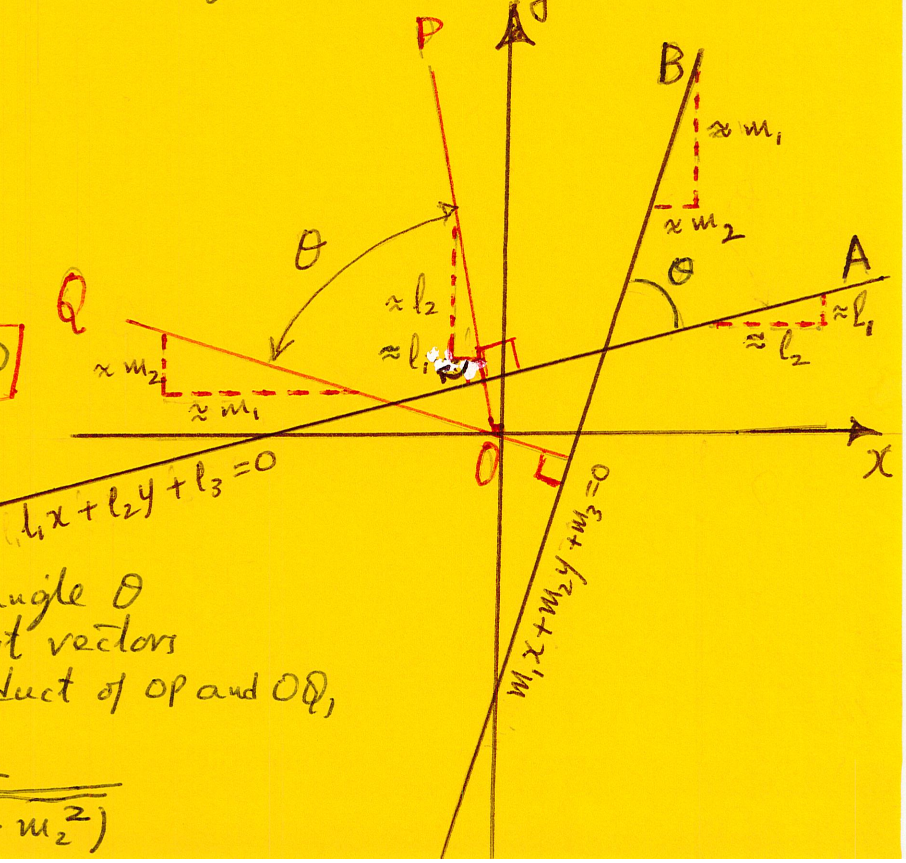
Estimating a Plane-to-Plane Homography with Angle-to-Angle and Point-to-Point Correspondences

Reference: "Multiple View Geometry in Computer Vision" by Hartley and Zisserman

- By this time you know that a typical camera image of a large planar scene can suffer from various types of distortions that correspond to the different transformation groups mentioned in Lecture 3.
- At the beginning of Lecture 4 I mentioned that if we can identify or calculate the vanishing line in an image, we can use the parameters of that line to get rid of that part of the distortion which is purely projective.
- Let's now talk about how we may get rid of the affine distortion that remains after we have rectified an image with respect to the purely projective distortion.
- The goal of the affine correction is to ensure that the angles between certain designated pairs of lines in an image are restored to what they are supposed to be in the original scene.
- As you know, affine distortion means that a shape like □ in the original scene will look like ▱ in the image. In other words, affine distortion turns a $90°$ angle into some value θ. The correction should take θ back to $90°$.
- We will now develop a formula for $\cos\theta$ where θ is the angle between two lines l and m in the original scene. Subsequently, we will express this formula in terms of the transformed forms for l and m as found in a recorded image. Assuming θ is $90°$ in the original planar scene, setting the expression for $\cos\theta$ to zero will give us the equations we need for solving for the correction homography.
- In the figure at right, we want a formula for $\cos\theta$, where θ is the angle between the lines A and B, in terms of the homogeneous representation l and m for the lines. Noting that l and m translate into $\boxed{l_1x + l_2y + l_3 = 0}$ and $\boxed{m_1x + m_2y + m_3 = 0}$ implicit forms, we see that the slope triangles for the two lines depend only on (l_1, l_2) for A and (m_1, m_2) for B. Also note that the angle θ is the same as the angle between the unit vectors OP and OQ to the lines. From the dot product of OP and OQ,



$$\cos\theta = \frac{l_1 m_1 + l_2 m_2}{\sqrt{(l_1^2 + l_2^2)(m_1^2 + m_2^2)}}$$

- Let's now consider a camera image of the original scene that can be described by the affine homography H. So what we will see in the recorded image would be the lines $\boxed{l' = H^{-T} l}$ and $\boxed{m' = H^{-T} m}$ for A and B, respectively. We could try to substitute $l = H^T l'$ and $m = H^T m'$ in the formula for $\cos\theta$ and solve for H if we know θ, but that turns out to be too messy. Instead we will take the following elegent approach.

- We rewrite the formula for $\cos\theta$ as

$$\boxed{\cos\theta = \frac{l^T C_\infty^* m}{\sqrt{(l^T C_\infty^* l)(m^T C_\infty^* m)}}}$$

C_∞^* = Dual Degenerate Conic See Lecture 4

This works because $C_\infty^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

- We will now express l, m, and C_∞^* in the original planar scene in terms of the observed l', m', and $C_\infty^{*'}$ in the recorded image. Recall from Lecture 3 that conics tranform as $C' = H^{-T} C H^{-1}$. This result can be extended to show that dual conics tranform as $\boxed{C^{*'} = H C^* H^T}$. (You prove this by substituting in $l^T C^* l = 0$ the relation $l = H^T l'$.) Substituting these in the formula for $\cos\theta$:

$$\cos\theta\Big|_{\text{numerator}} = (l'^T H)(H^{-1} C_\infty^{*'} H^{-T})(H^T m') = \boxed{l'^T C_\infty^{*'} m'}$$

- Let's now assume that the angle θ in the original scene is $90°$. The above result then yields the following constraint for estimating H:

$$\boxed{l'^T H C_\infty^* H^T m' = 0} \Rightarrow \boxed{(l_1' \; l_2' \; l_3') \begin{bmatrix} A & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix} \begin{bmatrix} I & \vec{0} \\ \vec{0}^T & 0 \end{bmatrix} \begin{bmatrix} A^T & \vec{0} \\ \vec{t}^T & 1 \end{bmatrix} \begin{pmatrix} m_1' \\ m_2' \\ m_3' \end{pmatrix} = 0}$$

which collapses into

$$\boxed{(l_1' \; l_2' \; l_3') \begin{bmatrix} AA^T & \vec{0} \\ \vec{0}^T & 0 \end{bmatrix} \begin{pmatrix} m_1' \\ m_2' \\ m_3' \end{pmatrix} = 0} \xRightarrow{S = AA^T} \boxed{(l_1' \; l_2') \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{pmatrix} m_1' \\ m_2' \end{pmatrix} = 0}$$

AA^T is symmetric so $s_{12} = s_{21}$

which gives us the following equation for the elements of S:

$$\boxed{s_{11} l_1' m_1' + s_{12}(l_1' m_2' + l_2' m_1') + s_{22} l_2' m_2' = 0}$$

- If we have another angle-to-angle correspondence — between two lines forming a $90°$ angle in the original scene and the images of these lines forming some angle β in the recorded image — we can write another equation of the sort shown above for the elements of S. Note that although we have 3 unknowns — s_{11}, s_{12}, s_{22} — we only need to know them up to their ratios, which means that, in reality, we have only two unknowns since we can set one of the three unknowns to 1. So, two equations should be sufficient to solve for S. [As you know, affine distortion consists of unequal scaling of the scene along two orthogonal directions. This distortion will, in general, include a general stretching (or shrinkage) of the scene — we refer to this as the isotropic part of the affine distortion, the rest being the "purely" affine anisotropic part. In other words, the affine distortion will, in general, include similarity distortion. By only calculating S up to a scale value, the eventual correction to the image will only address the purely affine effect; the corrected image will still have similarity distortion.]

- Having calculated the 2x2 matrix S, we are still faced with the problem of estimating the 2x2 matrix A.

- Recall that A is non-singular. If we also assume that A is positive-definite (meaning that $x^TAx > 0$ for all non-zero x), we can recover A by recognizing that a positive-definite A lends itself to the eigendecomposition $\boxed{A = VDV^T}$ where $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ with $\boxed{\lambda_1, \lambda_2 > 0}$ and where the columns of V are the eigenvectors of A. We can now write

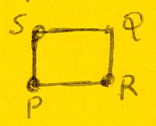
From Lecture 3: $A = UDV^T = (UV^T)(VDV^T)$. Let's leave the pure rotation UV^T to similarity correction

$$S = AA^T = VDV^TVDV^T = VD^2V^T = V\begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}V^T$$

where we used the fact that $V^TV = I$ because of the orthonormality of the eigenvectors. So, by doing an eigendecomposition of S, we get eigenvectors of A, with its eigenvalues given by the positive square-roots of the eigenvalues of S.

- Note that S is guaranteed to possess an eigendecomposition because it is positive-definite: $x^TSx = x^TA^TAx = (Ax)^T(Ax) = \|Ax\|^2 > 0$ for all nonzero x. For this assertion to hold, A must be nonsingular since it must be of full rank. If not of full rank, A will possess a null vector x for which $Ax = 0$, which will negate the assertion.

- Regarding positive-definiteness of A, note that whereas $A_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is positive definite, $A_2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is not. The former transforms



into

, the latter transforms the square into

. Notice how A_2 introduces a reflection about the PQ axis. Positive-definiteness of A means that we do not expect to see such reflections in distorted images.

- As already mentioned, for the matrix A to be positive definite, its eigenvalues must be positive. An interesting aside: For 2x2 matrices, the positive-definiteness condition can be checked algebraically as follows: The eigenvalues of A are the roots of the characteristic polynomial $\boxed{\det(A - \lambda I) = 0}$. With $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, this becomes $\boxed{\det\left(\begin{bmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{bmatrix}\right) = 0}$. Writing out the expressions for the roots λ_1 and λ_2 of the quadratic form and requiring that roots be positive gives us the conditions $\boxed{a_{11} > 0}$ and $\boxed{a_{11}a_{22} > a_{12}a_{21}}$ for A to be positive definite.

Point Correspondences For Estimating a Homography

- As you know already, given a point x in a planar scene and its corresponding pixel x' in the image plane, for most cameras we can write $\boxed{x' = Hx}$ assuming that x and x' are expressed using homogeneous coordinates: $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and $x' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix}$. With $H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$, we end up with →

$$\boxed{\begin{aligned} x_1' &= h_{11}x_1 + h_{12}x_2 + h_{13}x_3 \\ x_2' &= h_{21}x_1 + h_{22}x_2 + h_{23}x_3 \\ x_3' &= h_{31}x_1 + h_{32}x_2 + h_{33}x_3 \end{aligned}}$$

Denoting the physical scene coordinates by (x, y) and the physical pixel coordinates by (x', y'), we have $\boxed{x = x_1/x_3}$ $\boxed{y = x_2/x_3}$ and $\boxed{x' = x_1'/x_3'}$ $\boxed{y' = x_2'/x_3'}$

- So we can write for the physical coordinates of the image pixel:

$$x' = \frac{h_{11}x_1 + h_{12}x_2 + h_{13}x_3}{h_{31}x_1 + h_{32}x_2 + h_{33}x_3} \qquad y' = \frac{h_{21}x_1 + h_{22}x_2 + h_{23}x_3}{h_{31}x_1 + h_{32}x_2 + h_{33}x_3}$$

Dividing through by x_3 on the right-hand sides, we get purely in terms of just the physical coordinates on both sides:

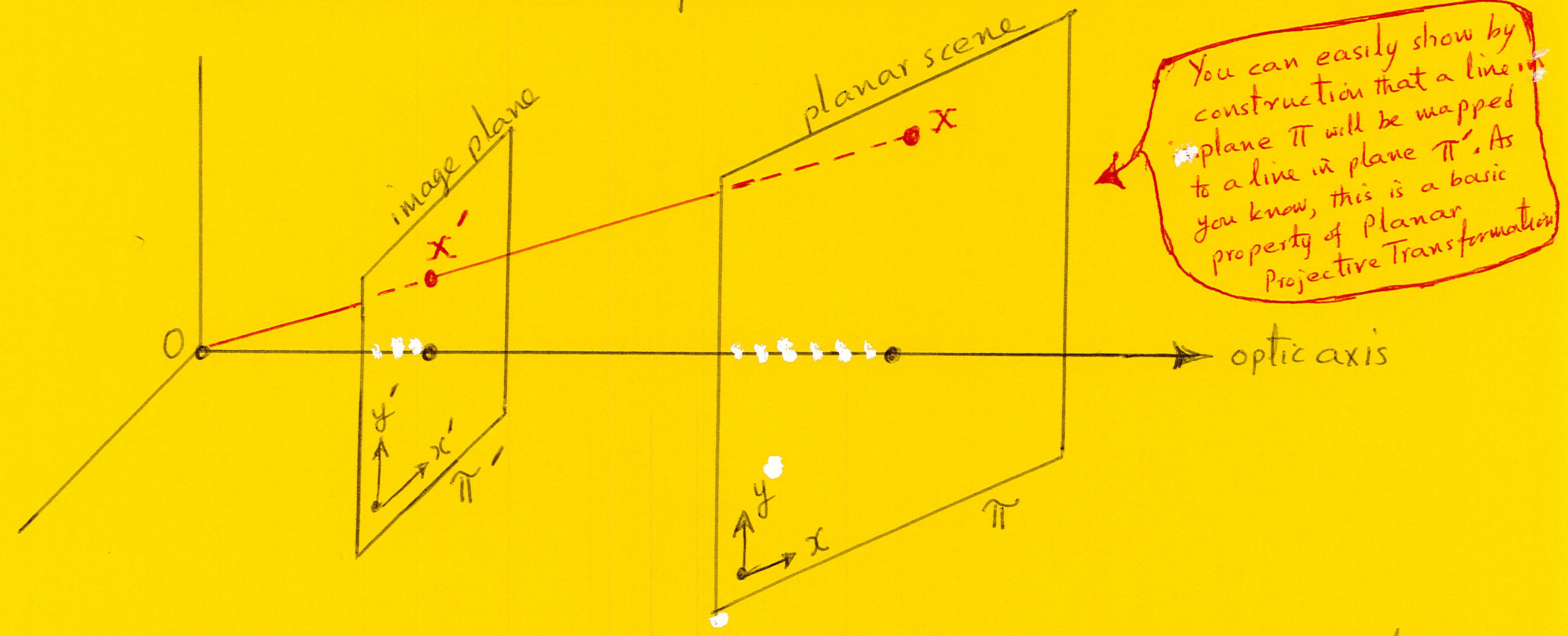
$$\boxed{x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}} \qquad \boxed{y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}}$$

- These expressions for the physical pixel coordinates can be rewritten in the following form:

$$\boxed{\begin{aligned} x h_{11} + y h_{12} + h_{13} \qquad\qquad\qquad - x x' h_{31} - y x' h_{32} - x' h_{33} &= 0 \\ x h_{21} + y h_{22} + h_{23} - x y' h_{31} - y y' h_{32} - y' h_{33} &= 0 \end{aligned}}$$

- Thus a single point correspondence between the original scene and the image gives us two linear equations for the elements of H.
- So with 4 point correspondences, we will get eight equations for the nine unknowns of H. However, since H is homogeneous (that is, since only the ratios of the elements of H are important), we only need to calculate H within a multiplicative constant. This can be done with four point-to-point correspondences with the proviso that the eight equations are linearly independent — a condition that is satisfied when no three of the four points fall on a single straight line.
- Later lectures will go into more robust methods for this approach to the estimation of a homography.

Planar Perspective Transformation



- This is a very special case of Planar Projective Transformation that is particularly suited to the modeling of most cameras.
- In Planar Perspective Transformation, all rays that join a scene point x with its corresponding image point x' must pass through the same point that is referred to as the Center of Projection or the Focal Center. In the depiction above, the origin is the CoP.
- Obviously, an image formed with a Planar Perspective Transformation will, in general, suffer from projective, affine, and similarity distortions.