Coordinate conventions

Image space, camera space, homography and epipolar point Expressing planes, computing incidence relations

1 The epipolar constraint

Given two cameras with distinct optical centers O_1 and O_2, a point in the first camera's image plane x_1 has a line-preimage which (in general) maps to a line on the image plane of the second camera. This transformation from points to lines is a homography, and can be expressed by a 3×3 matrix $F \in PG(2, \mathbb{R})$. This is called the fundamental matrix of camera 1 and camera 2.

$$|x_1' \quad e_1' \quad x_2| = 0.$$

If the second camera is related to the first by a rigid body motion $M = (R, T)$, and has the same intrinsic parameters, then this constraint becomes

$$|Rx_1 \quad T \quad x_2| = 0,$$

as T is a vector parallel to e_1'. This constraint can be rearranged into a bilinear form in x_2, x_1 by the scalar triple product, where $\hat{T}$ is the skew-symmetric matrix such that $\hat{T}v = T \times v$.

$$|Rx_1 \quad T \quad x_2| = 0$$
$$\equiv \quad x_2^T (T \times Rx_1) = x_2^T \hat{T} Rx_1 = 0.$$

In this case the fundamental matrix is also called the *essential matrix*.

2 The trifocal tensor

$$P_i = \left[A_i e_i' \right].$$

$$\pi_i = l_i P_i.$$
$$\Pi = \left[\pi_1^T \pi_2^T \pi_3^T \right]^T.$$

We want the intersection of these three planes to be a unique line. Each of these planes corresponds to a 3-dimensional subspace in $\mathbb{R}^4$. Therefore we want the intersection of these subspaces in R^4 to be 2-dimensional. This is expressed as a constraint on the nullity,

$$\text{null} \, \Pi = 2.$$

$$\Pi = \begin{bmatrix} l_1 & l_2 A_2 & l_3 A_3 \\ 0 & l_2 e_2' & l_3 e_3' \end{bmatrix}$$

$$\Rightarrow l_1 = \alpha l_2 A_2 + \beta l_3 A_3,$$
$$0 = \alpha l_2 e_2' + \beta l_3 e_3' \quad \text{for some } \alpha, \beta \in \mathbb{R}$$
$$\Rightarrow \alpha = k l_3 l_3 e_3', \beta = -k l_2 e_2' \quad \text{for some } k \in \mathbb{R}$$
$$\equiv l_1 \propto l_3 e_3' l_2 A_2 - l_2 e_2' l_3 A_3 = l_2^T \left(A_2^T (e_3')^T - e_2' A_3^T \right) l_3^T.$$

— Factored-out tensor?

The trifocal tensor, analogous to the fundamental matrix, encodes "fundamental" homographies between linear elements of the image planes of the cameras.

Plucker coordinates

Geometric interpretation

Why a trilinear constraint?

Computing the trifocal tensor

These "spectral" methods do not minimize the reprojection error. Minimization of the reprojection error is the goal of bundle adjustment. N-view factorization gives an iterative method where each iteration is in some sense in "closed form".

Spectral methods are developed from the assumption that each camera has a complete set of perfectly matched points. From these point matches, equations of incidence are derived for various affine subspaces where the equations evaluate to zero exactly on incidence. These equations finally form a large linear system with a right-hand-side of zero, and are solved for a non-trivial vector in a one-dimensional nullspace. The central idea is that noisy data will in general give a full rank matrix – The nullspace problem is approximated by finding the minimal eigenspace. This works in principle, as perfect (non-degenerate) data will be solvable exactly. The hope is that, with noisy data, the minimal eigenspace will give a "close" solution. However, the eigenstructure of this matrix does not in general behave in a "nice" and stable way with respect to perturbations of the data. This major downside to these spectral methods stems from some initial assumptions – The method is developed, maintaining the clean exact properties of perfect data, and generalized in a non-principled way. No noise is modelled explicitly, and no explicit geometric error is minimized.

Bundle adjustment models the noise explicitly, and starts with an exact description of a "good" solution as the minimizer of a cost function.