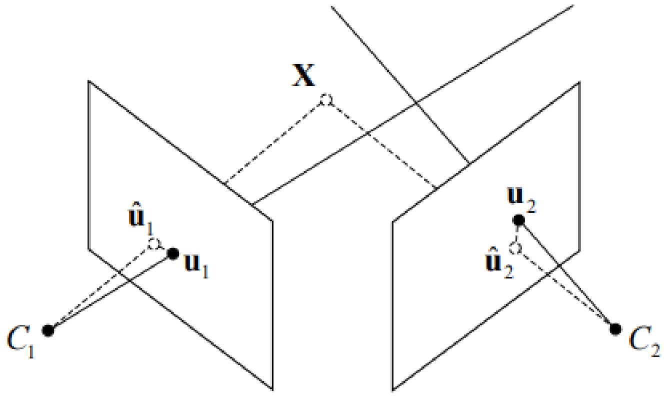
Structure from Motion

SfM focuses on how to recover information about the 3D scene from multiple 2D images.
We follow the same notation as in Epipolar Geometry.

Triangulation

One of the most fundamental problems in multiple view geometry is the problem of triangulation, the process of determining the location of a 3D point given its projections into two or more images.



Suppose that we have two cameras, and know the camera calibration matrices of each camera are $\mathbf{K}_1$, $\mathbf{K}_2$ and the relative orientation $\mathbf{R}$ and offsets $\mathbf{T}$ of these cameras w.r.t. each other. A point $\mathbf{X}$ in 3D, which can be found in the images of the two cameras at $\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{u}}_2$, has unknown 3D position. Because $\mathbf{K}_1, \mathbf{K}_2, \mathbf{R}, \mathbf{T}$ are known, we can compute two lines of sight l and l', which are defined by the camera centers C_1 and C_2. Therefore $\mathbf{X}$ can be computed as the intersection of l and l'.

Although this process appears both straightforward and mathematically sound, it does not work very well in practice. In the real world, because the observations $\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{u}}_2$ are noisy and the camera calibration parameters are not precise, finding the intersection point of l and l' may be problematic. In most cases, it will not exist at all, as the two lines may never intersect.

Linear Method for Triangulation

Given points in the images that correspond to each other $\tilde{\mathbf{u}}_i = \mathbf{P}_i\tilde{\mathbf{X}} = (u_i, v_i, 1)$. By definition of the cross product,

$$\tilde{\mathbf{u}}_i \times \mathbf{P}_i\tilde{\mathbf{X}} = [\tilde{\mathbf{u}}_i]_\times \mathbf{P}_i\tilde{\mathbf{X}} = 0$$

which is equivalent to

$$\begin{bmatrix} 0 & -1 & v_i \\ 1 & 0 & -u_i \\ -v_i & u_i & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_{i_1}^\top \\ \mathbf{p}_{i_2}^\top \\ \mathbf{p}_{i_3}^\top \end{bmatrix} \tilde{\mathbf{X}} = \begin{bmatrix} v_i\mathbf{p}_{i_3}^\top - \mathbf{p}_{i_2}^\top \\ u_i\mathbf{p}_{i_3}^\top - \mathbf{p}_{i_1}^\top \\ u_i\mathbf{p}_{i_2}^\top - v_i\mathbf{p}_{i_1}^\top \end{bmatrix} \tilde{\mathbf{X}} = \mathbf{0}$$

This equation has three rows but provides only two constraints on $\tilde{\mathbf{X}}$ since each row can be expressed as a linear combination of the other two.

Define

$$A_i = \begin{bmatrix} v_i\mathbf{p}_{i_3}^\top - \mathbf{p}_{i_2}^\top \\ u_i\mathbf{p}_{i_3}^\top - \mathbf{p}_{i_1}^\top \end{bmatrix}$$

and stack A_is to get

$$\mathbf{A} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}$$

All such constraints can be arranged into a matrix equation of the form

$$\mathbf{A}\tilde{\mathbf{X}} = 0$$

where $\mathbf{A}$ is a $3n \times 4$ matrix and n is the number of views in which the reconstructed point is visible.

The required solution for the homogenous 3D point $\tilde{\mathbf{X}}$ minimizes $\|\mathbf{A}\tilde{\mathbf{X}}\|$ subject to $\|\tilde{\mathbf{X}}\| = 1$ and is given by the eigenvector of $\mathbf{A}^\top\mathbf{A}$ corresponding to the smallest eigenvalue. It can be found by the singular value decomposition of the symmetric matrix $\mathbf{A}^\top\mathbf{A}$.

Nonlinear Method for Triangulation

The triangulation problem for real-world scenarios is often mathematically characterized as solving a minimization problem:

$$\min_{\hat{\mathbf{X}}} \sum_i \|\mathbf{P}_i\hat{\mathbf{X}} - \tilde{\mathbf{u}}_i\|^2$$

Affine Structure from Motion

Perspective Structure from Motion

Bundle Adjustment

Reference

04-stereo-systems.pdf (stanford.edu)