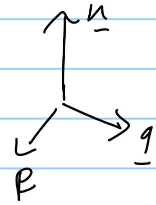
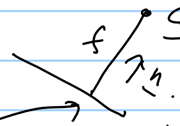
Lecture (17) Geometry and SVD.

Note Title 3/11/2010

Perspective Projection $\to$ Pinhole Camera.



right-handed coordinate system



Pinhole Camera

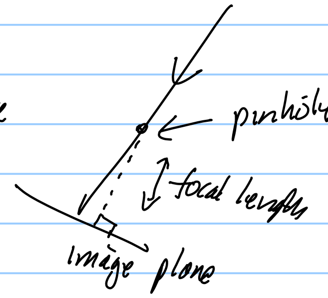


Image plane $\quad \underline{x}\cdot\underline{n} = k \quad k$ constant

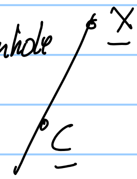
$\underline{c} - f\underline{n}$ lies on the plane.

Hence $\quad (\underline{c} - f\underline{n})\cdot\underline{n} = k \Rightarrow k = \underline{c}\cdot\underline{n} - f.$

Image Plane $\quad (\underline{x} - \underline{c})\cdot\underline{n} = -f$

Line in space through pinhole

$\underline{x}(\lambda) = \underline{c} + \lambda(\underline{x} - \underline{c})$

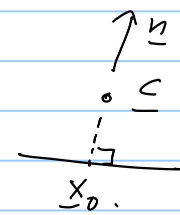


Line intersects image plane at λ^* s.t.

$$\lambda^* \underline{n}\cdot(\underline{x} - \underline{c}) + f = 0$$

$$\Rightarrow \lambda^* = \frac{-f}{\underline{n}\cdot(\underline{x} - \underline{c})}$$

Hence projection $\quad \underline{x}_p = \underline{c} - \frac{f(\underline{x} - \underline{c})}{(\underline{x} - \underline{c})\cdot\underline{n}}$

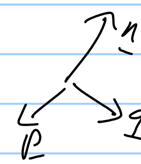


define center of projection $\quad \underline{x}_0 = \underline{c} - f\underline{n}$

then $\quad \underline{x}_p - \underline{x}_0 = f\underline{n} - \frac{f(\underline{x} - \underline{c})}{(\underline{x} - \underline{c})\cdot\underline{n}} = \frac{f}{(\underline{x} - \underline{c})\cdot\underline{n}}\{\underline{n}(\underline{x} - \underline{c})\cdot\underline{n} - (\underline{x} - \underline{c})\}$

Check that $(\underline{x}_p - \underline{x}_0)\cdot\underline{n} = 0$

so $(\underline{x}_p - \underline{x}_0)$ lies in space spanned by $\underline{p}$ & $\underline{q}$.



$$\underline{x}_p - \underline{x}_0 = \underline{p}\{\underline{p}\cdot(\underline{x}_p - \underline{x}_0)\} + \underline{q}\{\underline{q}\cdot(\underline{x}_p - \underline{x}_0)\}$$

$$x_p = \frac{f(\underline{x} - \underline{c})\cdot\underline{p}}{(\underline{x} - \underline{c})\cdot\underline{n}}, \quad y_p = \frac{f(\underline{x} - \underline{c})\cdot\underline{q}}{(\underline{x} - \underline{c})\cdot\underline{n}}$$

x_p, y_p position in image plane relative to image center.

Standard Coordinates: $\underline{n} = (0,0,1)$

$\underline{p} = (1,0,0)$, $\underline{q} = (0,1,0)$

$Z = (\underline{x} - \underline{c})\cdot\underline{n}$, $\quad X = (\underline{x} - \underline{c})\cdot\underline{p}$, $Y = (\underline{x} - \underline{c})\cdot\underline{q}$.

then $\quad x = f\frac{X}{Z}, \quad y = \frac{fY}{Z}$

(page 2)

$$\frac{b}{a+\epsilon} = \frac{b}{a(1+\frac{\epsilon}{a})} = \frac{b}{a} - \frac{b\epsilon}{a^2}$$

Linear Approximation

small expansion near $\underline{X}_0$.

Suppose $\underline{X} - \underline{C} = \underline{X}_0 + \Delta\underline{X}$

$$\frac{x_p}{f} = \frac{(\underline{X}_0+\Delta\underline{X})\cdot\underline{p}}{(\underline{X}_0+\Delta\underline{X})\cdot\underline{n}} = \frac{\underline{X}_0\cdot\underline{p}}{\underline{X}_0\cdot\underline{n}} + \frac{(\Delta\underline{X}\cdot\underline{p})}{(\underline{X}_0\cdot\underline{n})} - \frac{(\underline{X}_0\cdot\underline{p})(\Delta\underline{X}\cdot\underline{p})}{(\underline{X}_0\cdot\underline{n})^2} + O(\Delta^2)$$

$$\frac{y_p}{f} = \frac{(\underline{X}_0+\Delta\underline{X})\cdot\underline{q}}{(\underline{X}_0+\Delta\underline{X})\cdot\underline{n}} = \frac{\underline{X}_0\cdot\underline{q}}{\underline{X}_0\cdot\underline{n}} + \frac{(\Delta\underline{X}\cdot\underline{q})}{\underline{X}_0\underline{n}} - \frac{(\underline{X}_0\cdot\underline{q})(\Delta\underline{X}\cdot\underline{n})}{(\underline{X}_0\cdot\underline{n})^2} + O(\Delta^2).$$

Now $\Delta^2 \approx \frac{(\underline{X}_0\cdot\underline{p})}{(\underline{X}_0\cdot\underline{n})^3}(\Delta\underline{X}\cdot\underline{p})^2 + \frac{(\underline{X}_0\cdot\underline{q})}{(\underline{X}_0\cdot\underline{n})^3}(\Delta\underline{X}\cdot\underline{n})^2$

Important special case,
Suppose $\underline{X}_0$ lies directly ahead of the camera — i.e. $\underline{X}_0 \propto \underline{n}$

then $(\underline{X}_0\cdot\underline{p}) = (\underline{X}_0\cdot\underline{q}) = 0$

In this case $\frac{x_p}{f} \approx \frac{(\Delta\underline{X}\cdot\underline{p})}{|\underline{X}_0|}$, $\frac{y_p}{f} = \frac{(\Delta\underline{X}\cdot\underline{q})}{|\underline{X}_0|}$ scaled orthographic projection

In standard coordinates

$x = f\frac{X}{Z}$, $y = f\frac{Y}{Z}$ perspective

$x = kX$, $y = kY$ scaled orthograph $k = f/|X_0|$

More general approximation

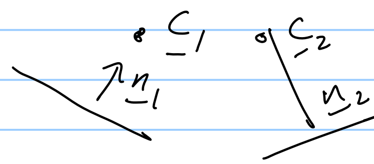
$$\begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

affine approximation

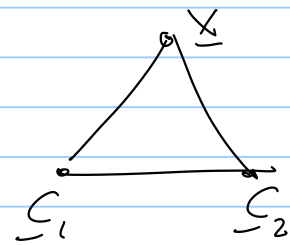
The affine approximation is often surprisingly good in practice.

Epipolar Line Constraint

<u>Two Cameras</u>:

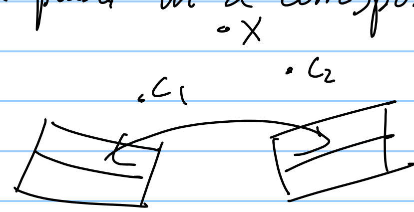


Any point $\underline{X}$ in space creates a plane passing through $\underline{c}_1, \underline{c}_2$ and $\underline{X}$.



i.e. $\underline{R}(\alpha,\beta) = \underline{X} + \alpha(\underline{c}_1 - \underline{X}) + \beta(\underline{c}_2 - \underline{X})$ for all α, β

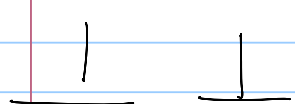
This plane intersects the first image plane in a line and the second plane in a corresponding line. These are epipolar lines.



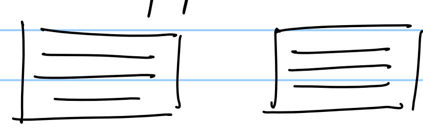
<u>Epipolar line</u> correspondence constraint: <u>points</u> on one epipolar line can only match points on the corresponding epipolar line.

The epipolar lines are determined by the camera geometry $(\underline{c}_1, \underline{n}_1, f_1)$ & $(\underline{c}_2, \underline{n}_2, f_2)$.

In the cameras point straight ahead $\underline{n}_1 = \underline{n}_2$



then the epipolar lines are straightforward.



How to determine epipolar lines?

<u>Theoretical Result</u> → If we can match N points in M images then we can solve for the camera parameters.

Which points to use? Find sparse interest points which can be matched unambiguously.

More about epipolar lines in the next lecture – or in the handout by Szeliski.

(4) Estimate Depth by Singular Value Decomposition (SVD)

Let $\Pi(\underline{X}^3_\mu, \underline{K}^m)$ be the projection of a point $\underline{X}^3_\mu$ for a camera with parameters $\underline{K}^m$.

Assume a linear model $\underline{\underline{K}}^m =$

$$\text{projected} \quad \underline{X}^P_{\mu,m} = \underset{\substack{2\times 3 \\ \text{matrix}}}{\underline{\underline{K}}^m} \; \underset{\substack{3\times 1 \\ \text{matrix}}}{\underline{X}^3_\mu}.$$

Formulate the problem as least squares estimation

$$E[\{X^3_\mu\} : \{\underline{\underline{K}}^m\}] = \sum_\mu \sum_m |\underline{X}^P_{\mu,m} - \Pi(\underline{X}^3_\mu, \underline{K}_m)|^2$$
$$= \sum_\mu \sum_m |\underline{X}^P_{\mu,m} - \underline{\underline{K}}^m \underline{X}^3_\mu|^2$$

This is a bilinear problem.

It can be solved by SVD up to an ambiguity.

– same as for lighting, see earlier lecture.

Ambiguity $\quad \underline{\underline{K}}^m \underline{X}^3_\mu \to \underline{\underline{K}}^m \underline{\underline{A}}\, \underline{\underline{A}}^{-1} \underline{X}^3_\mu$.

$\underline{\underline{A}}$ any invertible 3x3 matrix.

So ambiguity is an affine transformation on the geometry in 3 dimension.

SVD methods, Tomasi & Kanade, Kontsevich, Kontsevich.

This can be extended to some other types of projection

More generally, Sturm-Triggs method for perspective projection (but see modifications by Oliensis & Hartley).

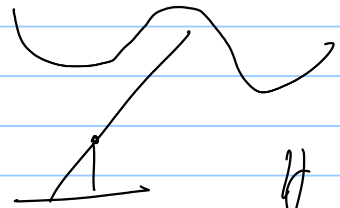
Note that full affine ambiguity is unlikely.

In practice, there are practical constraints on the values of the projection coefficients $\underline{\underline{K}}$.

If we have scaled orthographic, then the ambiguity reduces to $\underline{\underline{A}} = a\underline{\underline{I}}$, where $\underline{\underline{I}}$ is the identity and a is the scaling (unknown).

(5) Lighting and Viewpoint Ambiguity

Assume affine projection $(x, y) = \underline{\underline{K}}\begin{pmatrix}X\\Y\end{pmatrix}$ $\quad \underline{\underline{A}}$ 3x3 matrix. ambiguity



object + albedo $a(x)$
$\underline{n}(x)$

If two objects are related by affine transformation on the geometry and corresponding transformation on the albedo, then for any viewpoint and lighting conditions of the first object there is a corresponding viewpoint and lighting condition of the second object such that the images are identify (for affine camera)

Note: this ambiguity also includes shadows, both cast and attached.

The Generalized Bas Relief ambiguity (from lighting notes) is a special case if you constrain the viewpoint to be identical.

There is some evidence (Koenderink et al) that humans only perceive shapes up to affine transformation.