Computer Vision on Rolling Shutter Cameras

PART II: Rolling Shutter Geometry

Per-Erik Forssén, Erik Ringaby, Johan Hedborg

Computer Vision Laboratory
Dept. of Electrical Engineering
Linköping University

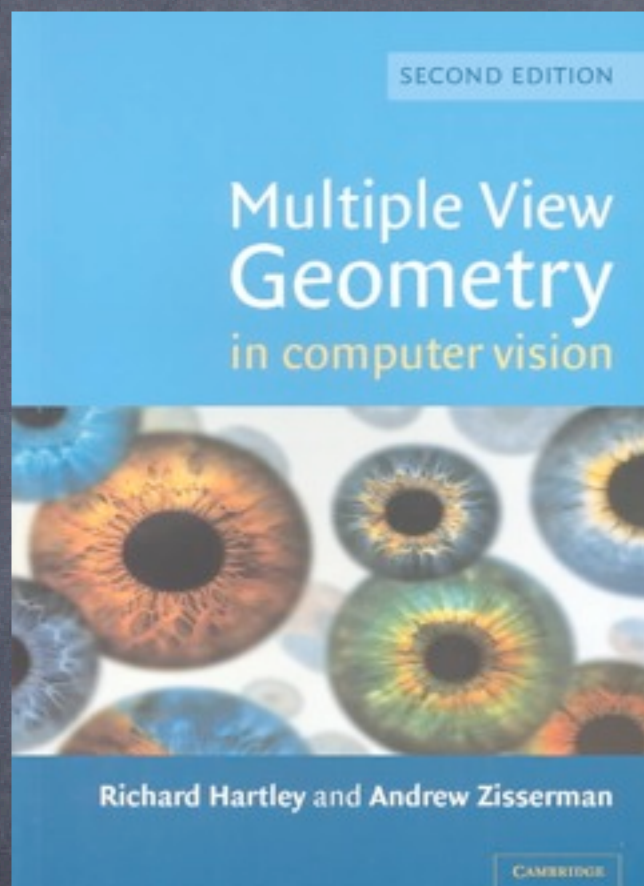
Linköping University
INSTITUTE OF TECHNOLOGY

Tutorial overview

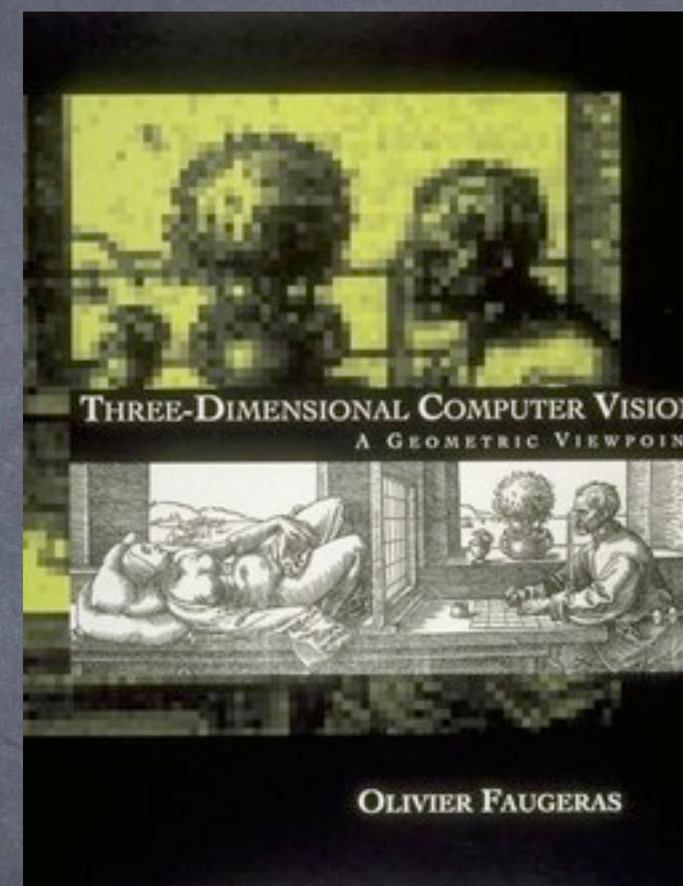
1:30-2:00pm	Introduction	Per-Erik
2:00-2:15pm	**Rolling Shutter Geometry**	**Per-Erik**
2:15-3:00pm	Rectification and Stabilisation	Erik
3:00-3:30pm	Break	
3:30-3:45pm	Rolling Shutter and the Kinect	Erik
3:45-4:30pm	Structure from Motion	Johan

Projective Geometry

- Textbook material



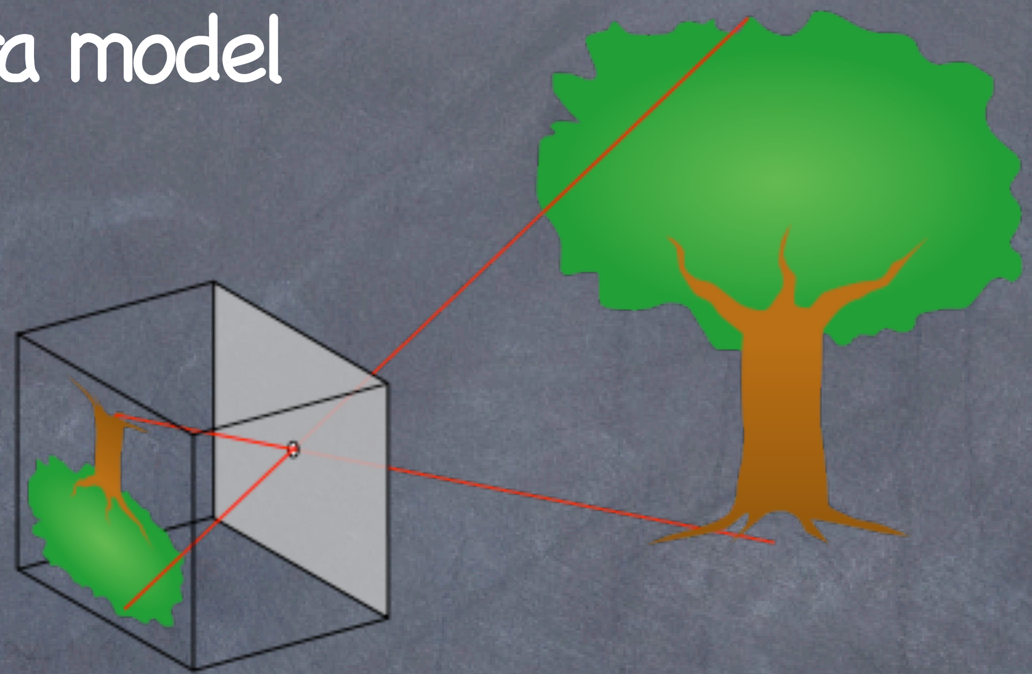
Hartley & Zisserman
Multiple View Geometry
2nd ed 2004



Faugeras
Three-Dimensional Computer Vision
1993

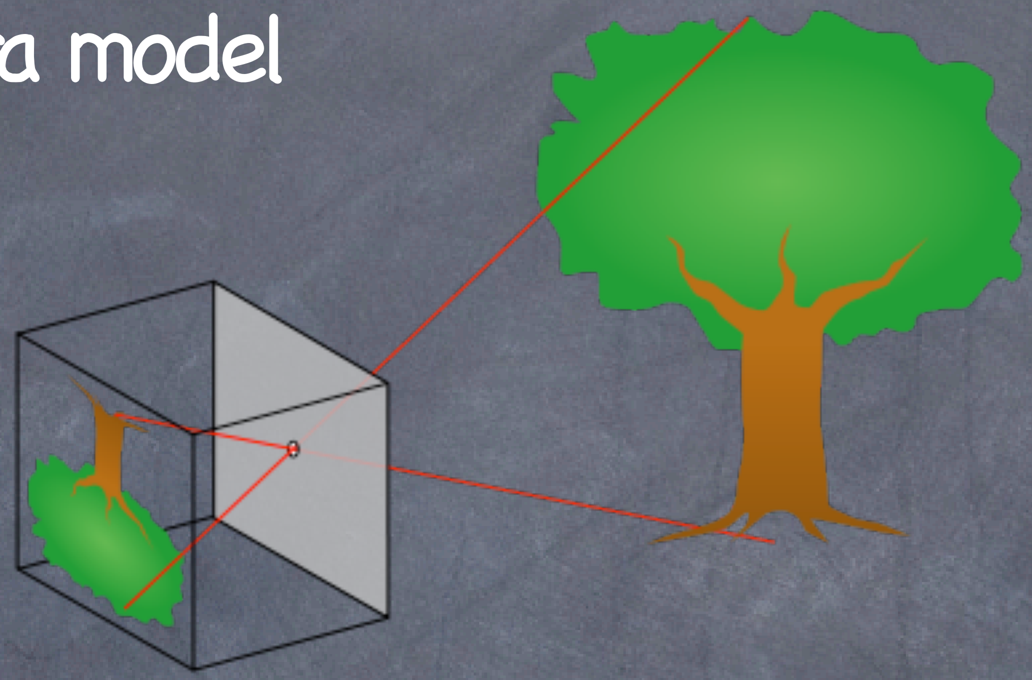
The pin-hole camera

- Pin-hole camera model



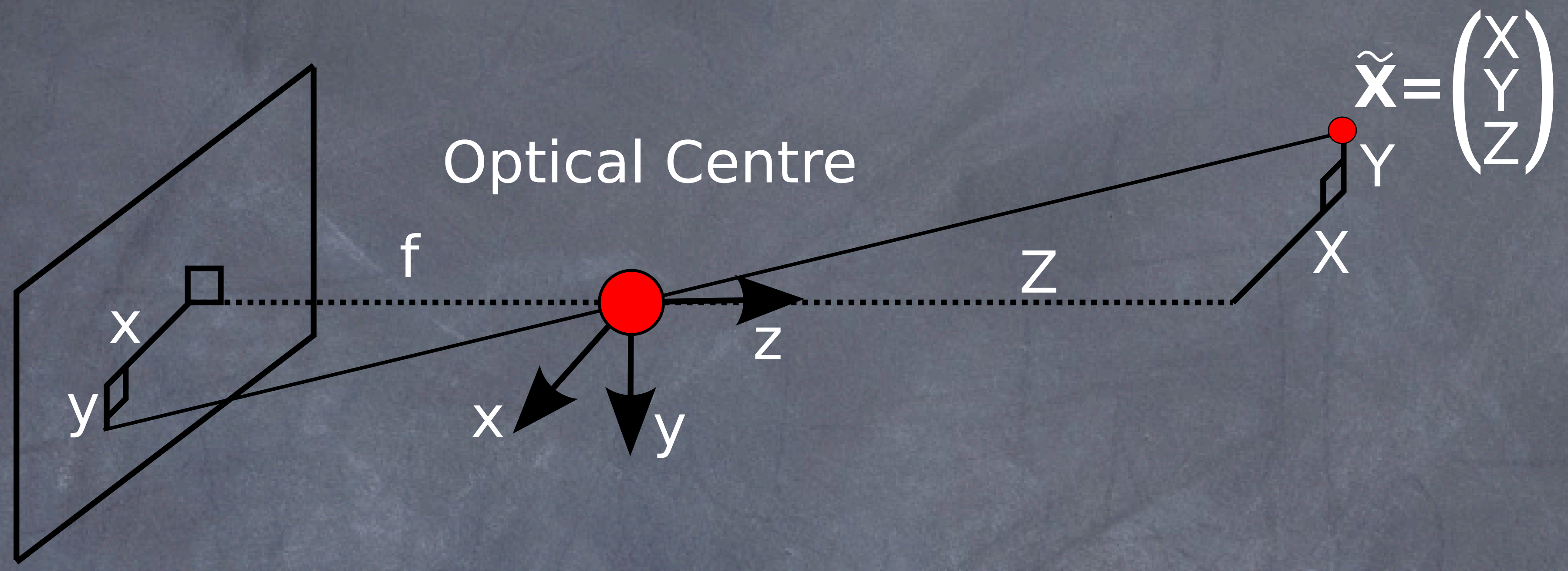
The pin-hole camera

- Pin-hole camera model



- A brightly illuminated scene will be projected onto a wall opposite of the pin-hole.
- The image is rotated 180°.

The pin-hole camera



- From similar triangles we get:

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

$$\gamma \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

The pin-hole camera

$$\gamma \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

- More generally we write:

$$\gamma \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & s & c_x \\ 0 & af & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

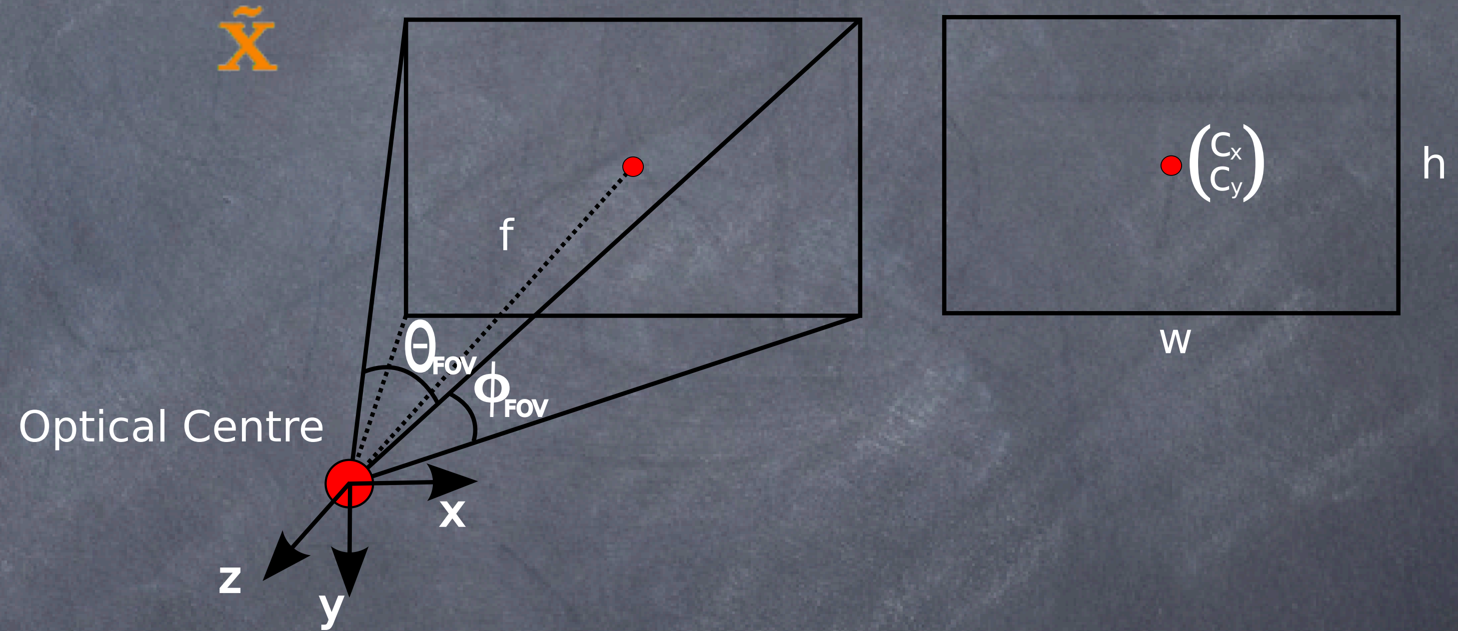
- f-focal length, s-skew, a-aspect ratio, **c**=[c_x c_y]-projection of optical centre

The pin-hole camera

$$\underbrace{\gamma\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} f & s & c_x \\ 0 & af & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}}_{\tilde{\mathbf{X}}}$$

$$\mathbf{x} \sim \mathbf{K}\tilde{\mathbf{X}}$$

- Motivation:



- f-focal length, s-skew, a-aspect ratio, **c**=[c_x c_y]-projection of optical centre

The pin-hole camera

- Projection of 3D points X_c in the camera coordinate system:

$$\mathbf{x} \sim \mathbf{K}\tilde{\mathbf{X}}_\mathbf{c}$$

- In order to relate several camera poses, we need to use a common world coordinate system (WCS) :

$$\tilde{\mathbf{X}}_\mathbf{c} = \mathbf{R}^T(\tilde{\mathbf{X}}_\mathbf{w} - \mathbf{d}) \quad \Rightarrow \quad \mathbf{x} \sim \mathbf{K}\mathbf{R}^T(\tilde{\mathbf{X}}_\mathbf{w} - \mathbf{d})$$

- d is a translation of the origin, and R is a rotation

The pin-hole camera

- We can simplify this to a single projection operation on the 3D points X_w:

$$\mathbf{x} \sim \mathbf{K}\mathbf{R}^T(\tilde{\mathbf{X}}_\mathbf{w} - \mathbf{d}) \quad \Rightarrow \quad \mathbf{x} \sim \mathbf{P}\mathbf{X}_\mathbf{w}$$

where P is a 3x4 matrix, and

$$\mathbf{X}_\mathbf{w} = \left[\tilde{\mathbf{X}}_\mathbf{w}^T \; 1\right]^T = [X \; Y \; Z \;\; 1]^T$$

The pin-hole camera

- We can simplify this to a single projection operation on the 3D points X_w:

$$\mathbf{x} \sim \mathbf{K}\mathbf{R}^T(\tilde{\mathbf{X}}_\mathbf{w} - \mathbf{d}) \quad \Rightarrow \quad \mathbf{x} \sim \mathbf{P}\mathbf{X}_\mathbf{w}$$

where P is a 3x4 matrix, and

$$\mathbf{X}_\mathbf{w} = \left[\tilde{\mathbf{X}}_\mathbf{w}^T \; 1\right]^T = [X \; Y \; Z \;\; 1]^T$$

- This matrix P has the explicit form:

$$\mathbf{P} = \mathbf{K}\left[\mathbf{R}^T | -\mathbf{R}^T\mathbf{d}\right]$$

The pin-hole camera

- We can simplify this to a single projection operation on the 3D points X_w:

$$\mathbf{x} \sim \mathbf{K}\mathbf{R}^T(\tilde{\mathbf{X}}_\mathbf{w} - \mathbf{d}) \quad \Rightarrow \quad \mathbf{x} \sim \mathbf{P}\mathbf{X}_\mathbf{w}$$

where P is a 3x4 matrix, and

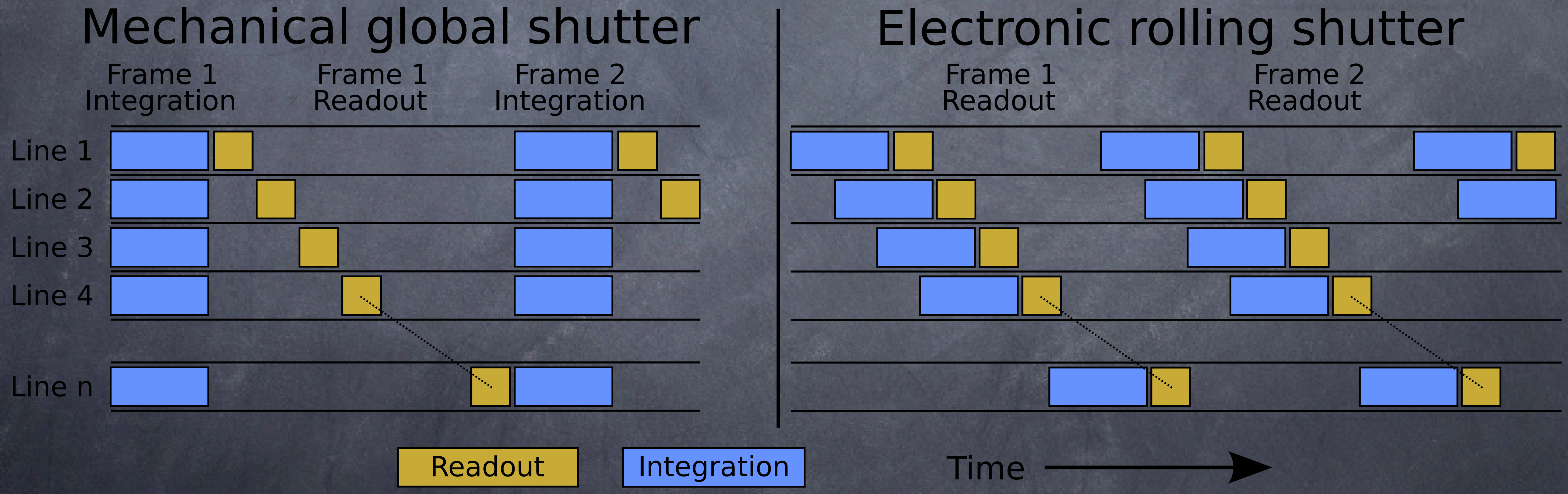
$$\mathbf{X}_\mathbf{w} = \left[\tilde{\mathbf{X}}_\mathbf{w}^T \; 1\right]^T = [X \; Y \; Z \;\; 1]^T$$

- This matrix P has the explicit form:

$$\mathbf{P} = \mathbf{K}\left[\mathbf{R}^T | -\mathbf{R}^T\mathbf{d}\right] = \mathbf{K}\mathbf{R}^T\left[\mathbf{I} | -\mathbf{d}\right]$$

Rolling shutter model

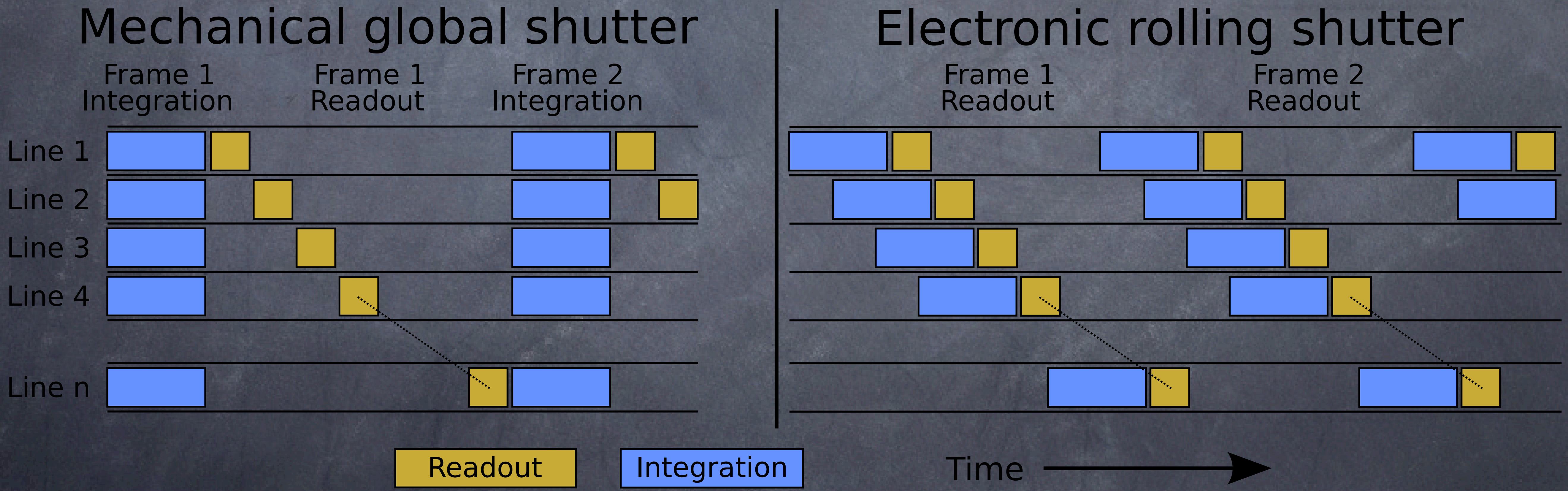
- Now recall the rolling shutter readout:



Rolling shutter model

- For a moving camera, projection in frame k becomes:

$$\mathbf{x}_k \sim \mathbf{K}\mathbf{R}_k^T \left[\mathbf{I} | -\mathbf{d}_k\right] \mathbf{X} \qquad \mathbf{x}_k \sim \mathbf{K}\mathbf{R}(x_2)^T \left[\mathbf{I} | -\mathbf{d}(x_2)\right] \mathbf{X}$$



Rolling shutter model

- For a moving camera, projection in frame k becomes:

$$\mathbf{x}_k \sim \mathbf{K}\mathbf{R}_k^T \left[\mathbf{I}| - \mathbf{d}_k\right] \mathbf{X} \qquad \mathbf{x}_k \sim \mathbf{K}\mathbf{R}(x_2)^T \left[\mathbf{I}| - \mathbf{d}(x_2)\right] \mathbf{X}$$

- In the global shutter case, we have one pose (R_k,d_k) per frame

- In the rolling shutter case, we instead get one pose ($R(x_2)$,$d(x_2)$) per image row, x_2

Time coordinate

- When interpolating the camera pose based on the image row, x_2 it is convenient to express time in number of rows, instead of seconds.

- Recall that the frame period T, is divided into readout time t_r and inter-frame delay t_d.

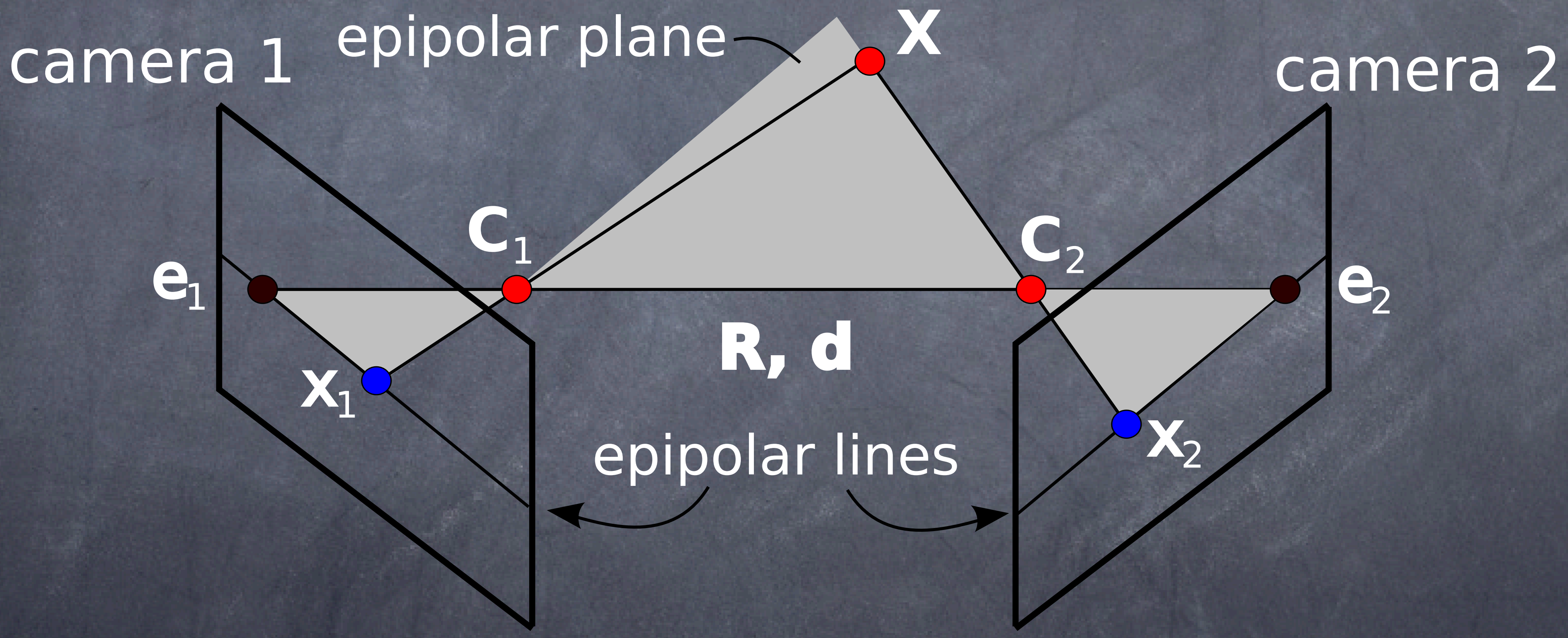
$$1/f = T = t_r + t_d$$

- t_r corresponds to number of image rows N_r, and t_d corresponds to number of blank-rows N_b.

$$N_b = N_r t_d / t_r = N_r (t_r / f - 1)$$

Triangulation

- Triangulation is the process of estimating a 3D point X from two projections x_1 and x_2.



Triangulation

- For the two points, we have:

$$\mathbf{x}_1 \sim \mathbf{P}_1 \mathbf{X}$$
$$\mathbf{x}_2 \sim \mathbf{P}_2 \mathbf{X}$$

Triangulation

- For the two points, we have:

$$\mathbf{x}_1 \sim \mathbf{P}_1\mathbf{X}$$
$$\mathbf{x}_2 \sim \mathbf{P}_2\mathbf{X}$$

- Triangulation is typically solved by so called optimal triangulation [Hartley&Zisserman'04]

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \left[d^2(\mathbf{x}_1, \mathbf{P}_1\mathbf{X}) + d^2(\mathbf{x}_2, \mathbf{P}_2\mathbf{X})\right]$$

  The point $\mathbf{X}$ is sought, for which the squared re-projection error in both images is minimized.

- There exists a closed form solution, that is found by solving a 3rd degree polynomial.

Rolling shutter triangulation

- If we generalize triangulation to a (moving) rolling shutter rig, we get:

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \left[d^2(\mathbf{x}_1, \mathbf{P}_1\mathbf{X}(t_1)) + d^2(\mathbf{x}_2, \mathbf{P}_2\mathbf{X}(t_2)) \right]$$

- This has an unique solution, if, and only if $t_1 = t_2$, which happens if the point is projected in both images at the same time.

- That is, when the two points have the same y-coordinate. (Very rare!)

Rolling shutter SfM

- Suggestion from [Ait-Aider & Berry, ICCV'09]: Solve for triangulation of all points, and the object motion at the same time (structure-from-motion SfM).
- The projection constraints for a correspondence now assumes the form:

$$\mathbf{x}_1 \sim \mathbf{K}\left[\mathbf{R}(t_1)|\mathbf{d}(t_1)\right]\mathbf{X}$$
$$\mathbf{x}_2 \sim \mathbf{K}\mathbf{R}_2\left[\mathbf{R}(t_2)|\mathbf{d}_2 + \mathbf{d}(t_2)\right]\mathbf{X}$$

- The set of all such triangulation constraints can uniquely define the solution, if we assume a parametric form for R(t) and d(t). The most simple one is a linear motion(6dof).

Degeneracy

- The full optimization problem now looks like this:

$$\{\mathbf{X}_k^*\}_1^K, \mathbf{R}, \mathbf{d} = \arg \min_{\{\mathbf{X}_k\}_1^K, \mathbf{R}, \mathbf{d}} \left[\sum_{k=1}^{K} d^2(\mathbf{x}_{1,k}, \mathbf{P}_1 \mathbf{X}_k) + d^2(\mathbf{x}_{2,k}, \mathbf{P}_2 \mathbf{X}_k) \right]$$

where

$$\mathbf{P}_1 = \mathbf{K} \left[\mathbf{R}(t_1) | \mathbf{d}(t_1) \right]$$

$$\mathbf{P}_2 = \mathbf{K} \mathbf{R}_2 \left[\mathbf{R}(t_2) | \mathbf{d}_2 + \mathbf{d}(t_2) \right]$$

- This problem has a unique solution when the motion is different from a pure translation along the x-axis.

- That is, when $t_1 \neq t_2$ for most points. (Otherwise we never observe point motion.)

Degeneracy

- For a translation parallel to the line between the optical centra, there is an equivalent stationary structure

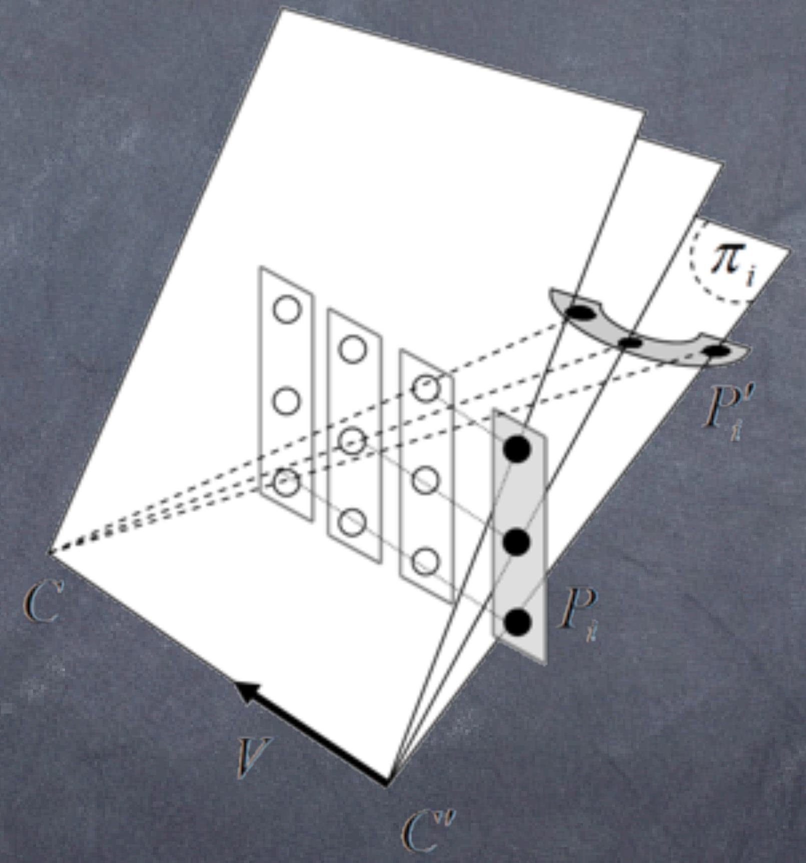
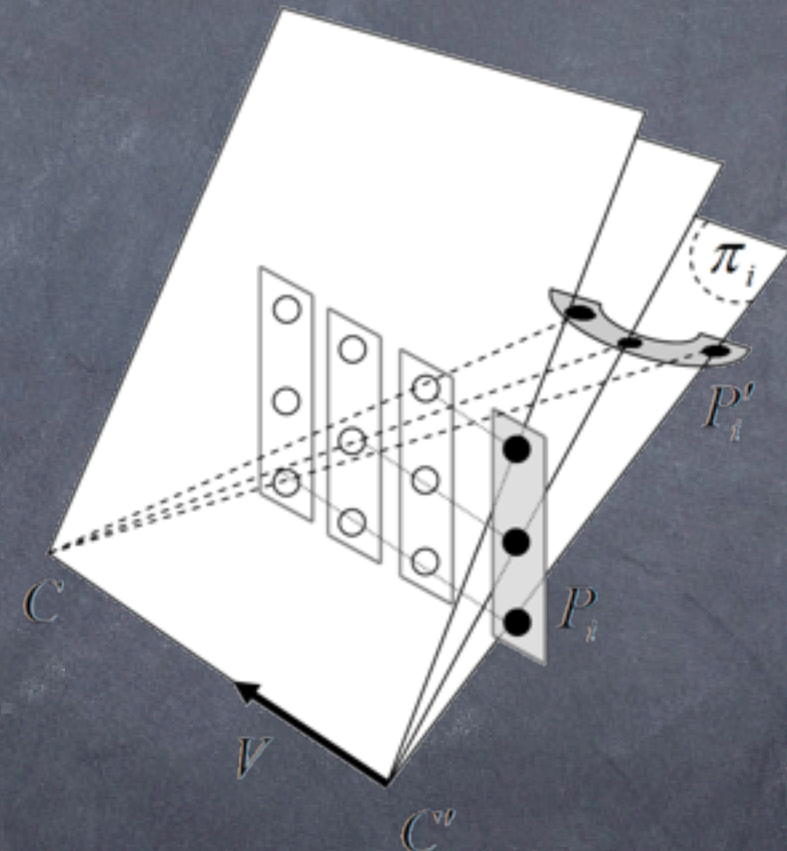


Illustration by Ait-Aider and Berry

Degeneracy

- For a translation parallel to the line between the optical centra, there is an equivalent stationary structure



⟵Translating structure

Illustration by Ait-Aider and Berry

Degeneracy

- For a translation parallel to the line between the optical centra, there is an equivalent stationary structure

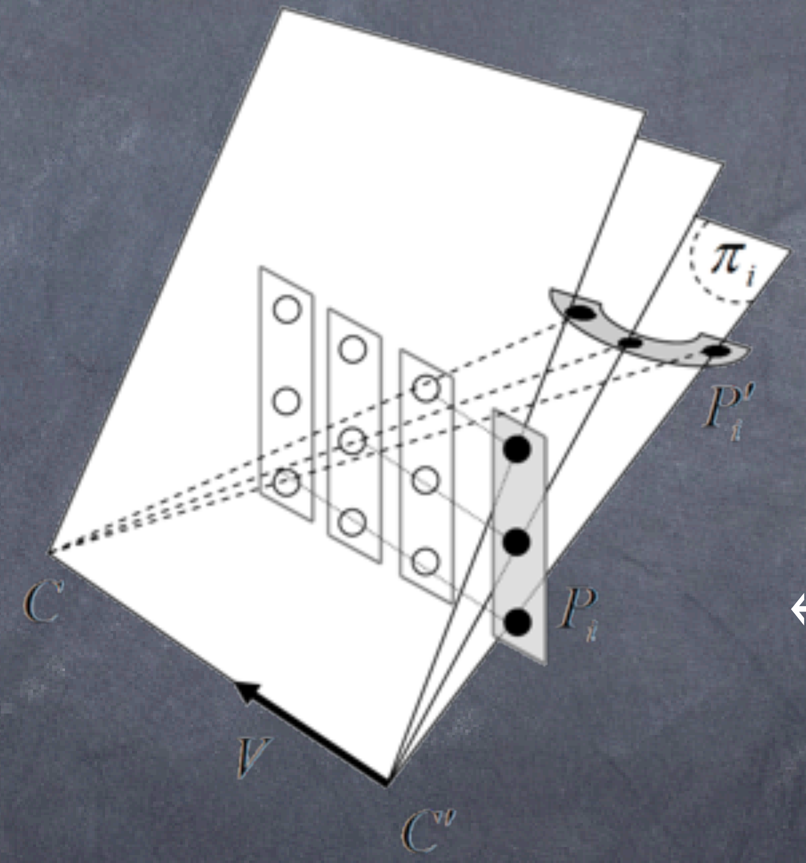


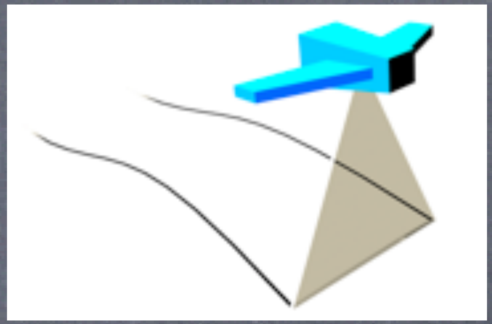
Illustration by Ait-Aider and Berry

Degeneracy

- Structure and motion (SfM) from two frames is unstable for sideways motion, when both cameras have the same readout speed (or are the same).
- If one of the cameras has a global shutter, both structure and motion can be obtained [Ait-Aider&Berry ICCV'09]
- If multiple frames are used, rolling shutter structure from motion (SfM) becomes stable again [Hedborg et al. CVPR'12].

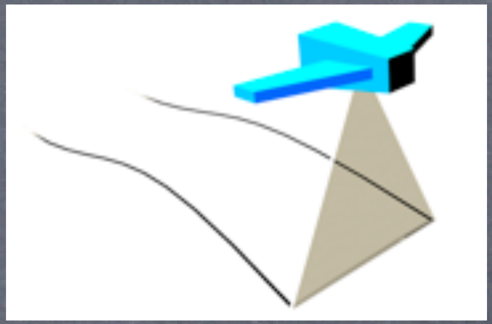
Related scanning cameras

- Pushbroom [Gupta and Hartley PAMI'97]
  Single line camera that is moving.

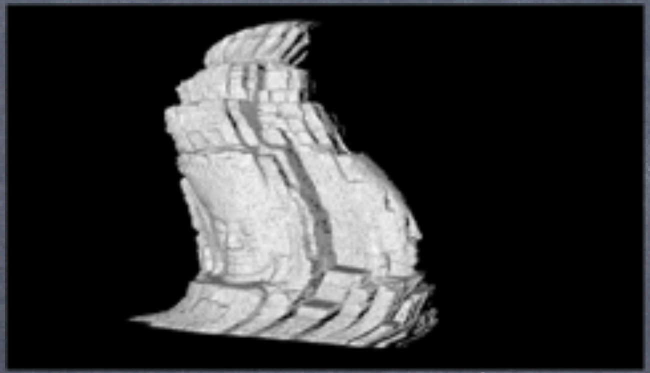


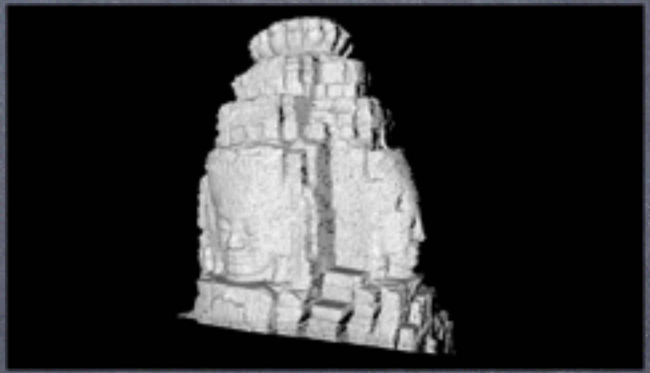
Related scanning cameras

- Pushbroom [Gupta and Hartley PAMI'97] Single line camera that is moving.



- Work on scanning LIDARs, e.g. archeological reconstruction work by [Ikeuchi et al.]



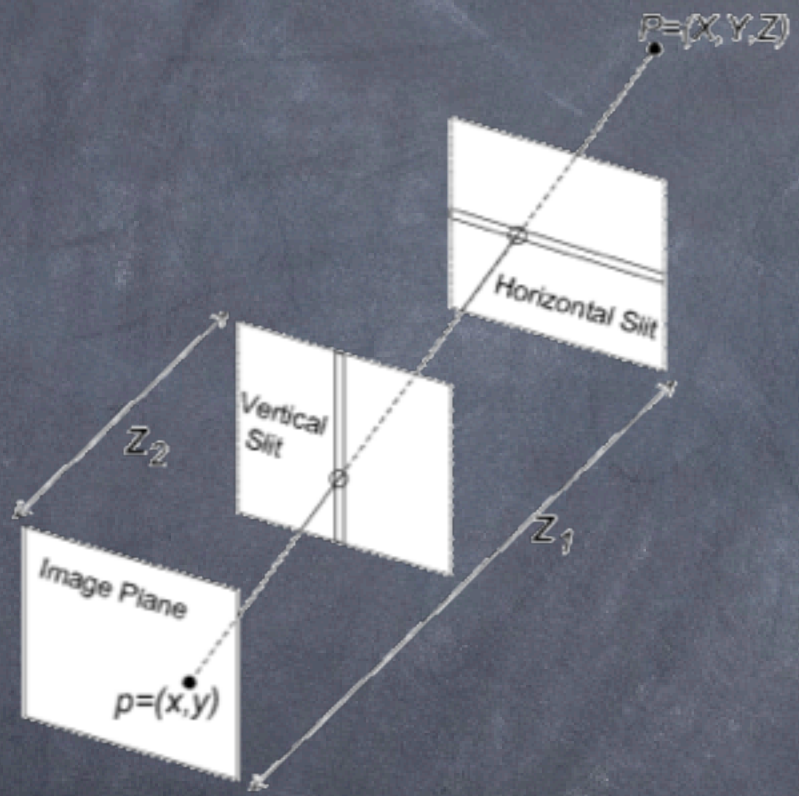


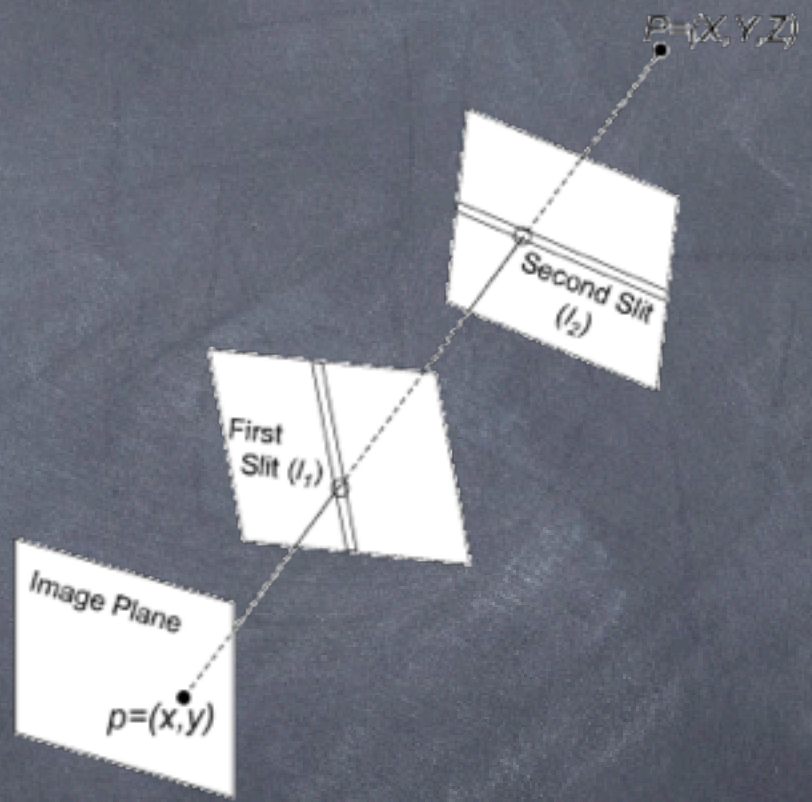
Images from lab of Katsushi Ikeuchi

- [Bosse ICRA'09] e.g. rotating, and bouncing LIDARs.

Related scanning cameras

- Crossed slits [Zomet et al. PAMI'03]





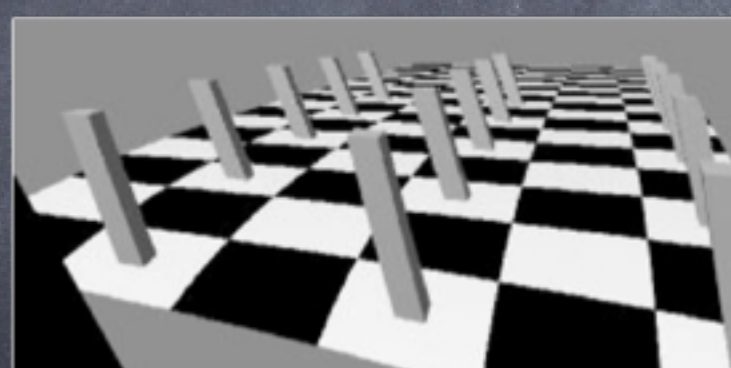
- projection rays from 3D points to the image plane are defined as intersections of two "slit planes".

Related scanning cameras

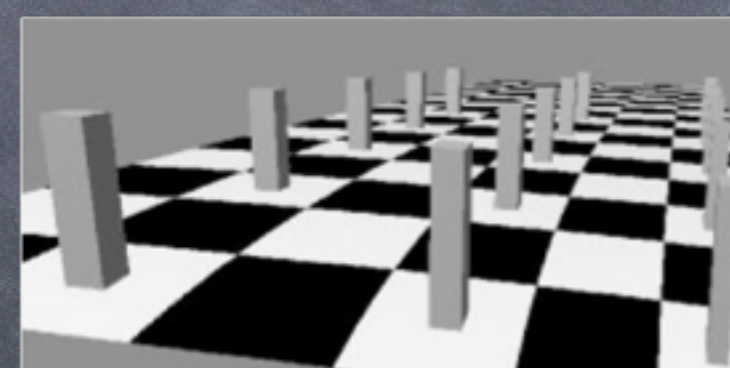
- Crossed slits [Zomet et al. PAMI'03]

horizontal/vertical slits

pin-hole camera rendering



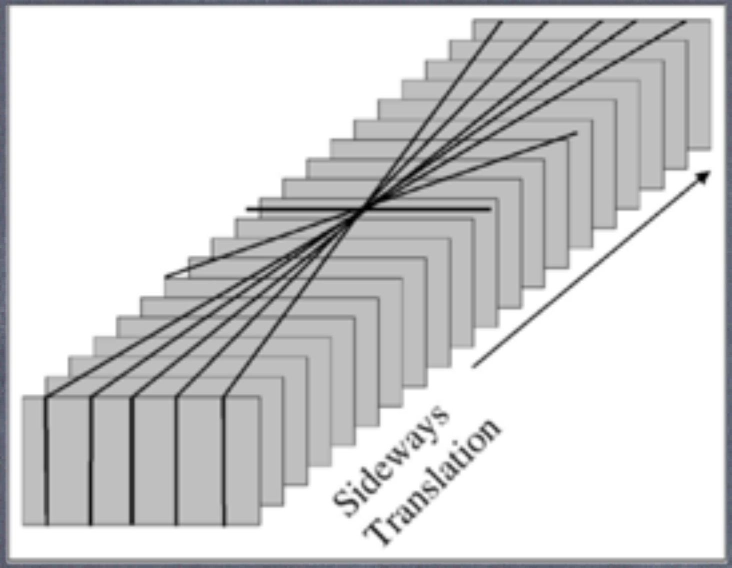
Z rotation of vertical slit



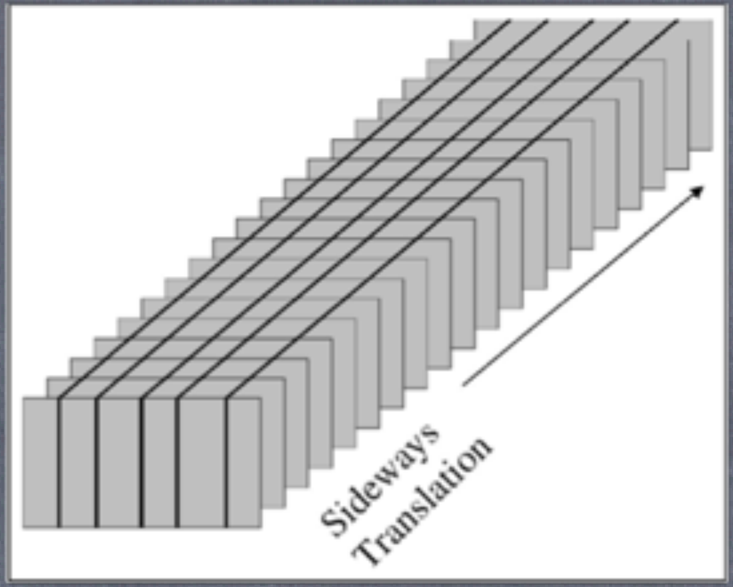
X rotation of vertical slit

Related scanning cameras

- Crossed slits [Zomet et al. PAMI'03]



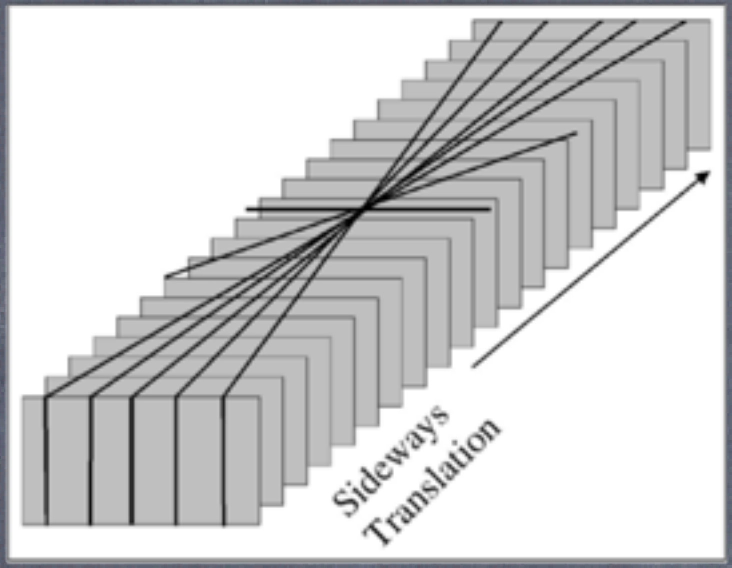
sweeping push-broom mosaic



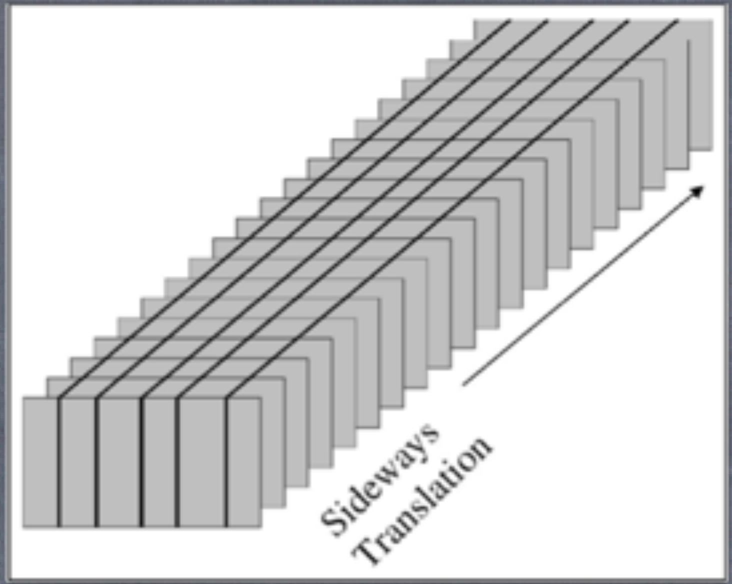
standard push-broom mosaic

Related scanning cameras

- Crossed slits [Zomet et al. PAMI'03]



sweeping push-broom mosaic



standard push-broom mosaic

rendered image

pin-hole camera image

Related scanning cameras

- Crossed slits [Zomet et al. PAMI'03]
- [Geyer et al. OMNIVIS'05] demonstrate that a rolling shutter camera is equivalent to a crossed-slits camera for a pure translation parallel to the image plane. (but not in general)
- A crossed-slits camera can thus be seen as a special case of the rolling shutter camera.

Summary

- In rolling shutter geometry, the camera trajectory is best modelled as continuous
- There is an ambiguity between structure and sideways motion for two-frame rolling-shutter geometry.
- Other types of scanning geometries (push-broom and moving LIDAR) do not have the temporal regularity of rolling shutter geometry.

References

- Zomet, Feldman, Peleg, Weinshall, "Mosaicing New Views: The Crossed-Slits Projection", PAMI'03
- Hartley, Zisserman, "Multiple View Geometry for Computer Vision"
- Geyer, Meingast, Sastry, "Geometric Models of Rolling Shutter Cameras", OMIVIS'05
- Gupta, Hartley, "Linear Pushbroom Cameras", PAMI'97
- Ait-Aider, Berry, "Structure and Kinematics Triangulation from a Rolling Shutter Stereo Rig", ICCV'09
- Bosse, Zlot "Continuous 3D Scan-Matching with a Spinning 2D Laser", ICRA'09
- Hedborg, Forssén, Felsberg, Ringaby, "Rolling Shutter Bundle Adjustment", CVPR'12