Geometry of two or more views

Václav Hlaváč

Czech Technical University in Prague

Czech Institute of Informatics, Robotics and Cybernetics

160 00 Prague 6, Jugoslávských partyzánů 1580/3, Czech Republic

http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz

also Center for Machine Perception, http://cmp.felk.cvut.cz

Outline of the talk:

- Motivation = stereopsis.
- Epipolar constraint.
- Fundamental matrix.
- Essential matrix.
- Eight point algorithm.
- Trinocular constraint and transfer.

Stereopsis

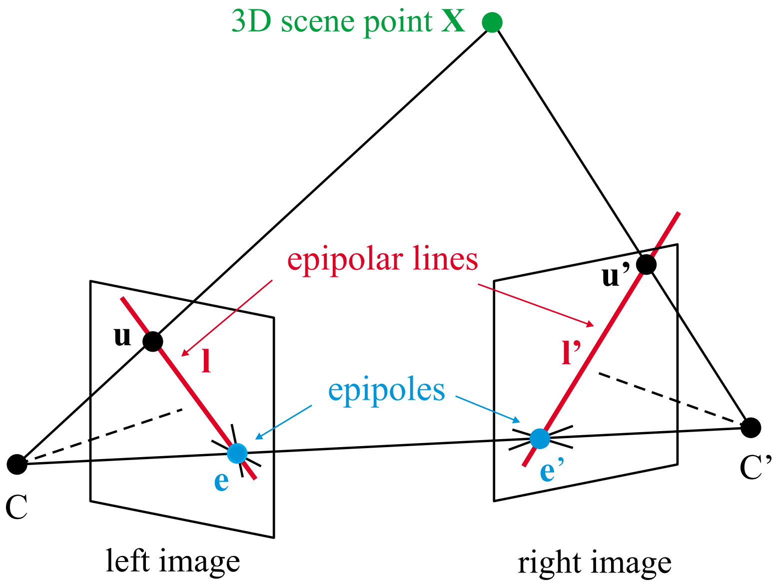
- Calibration of one camera and knowledge of the co-ordinates of one image point allows us to determine a ray in space uniquely.
- If two calibrated cameras observe the same scene point $\mathbf{X}$ then its 3D co-ordinates can be computed as the intersection of two such rays. This is the basic principle of stereo vision that typically consists of three steps:

  1. Camera calibration;
  2. Establishing point correspondences between pairs of points from the left and the right images;
  3. Reconstruction of 3D co-ordinates of the points in the scene.

Geometry of two cameras, epipolar constraint

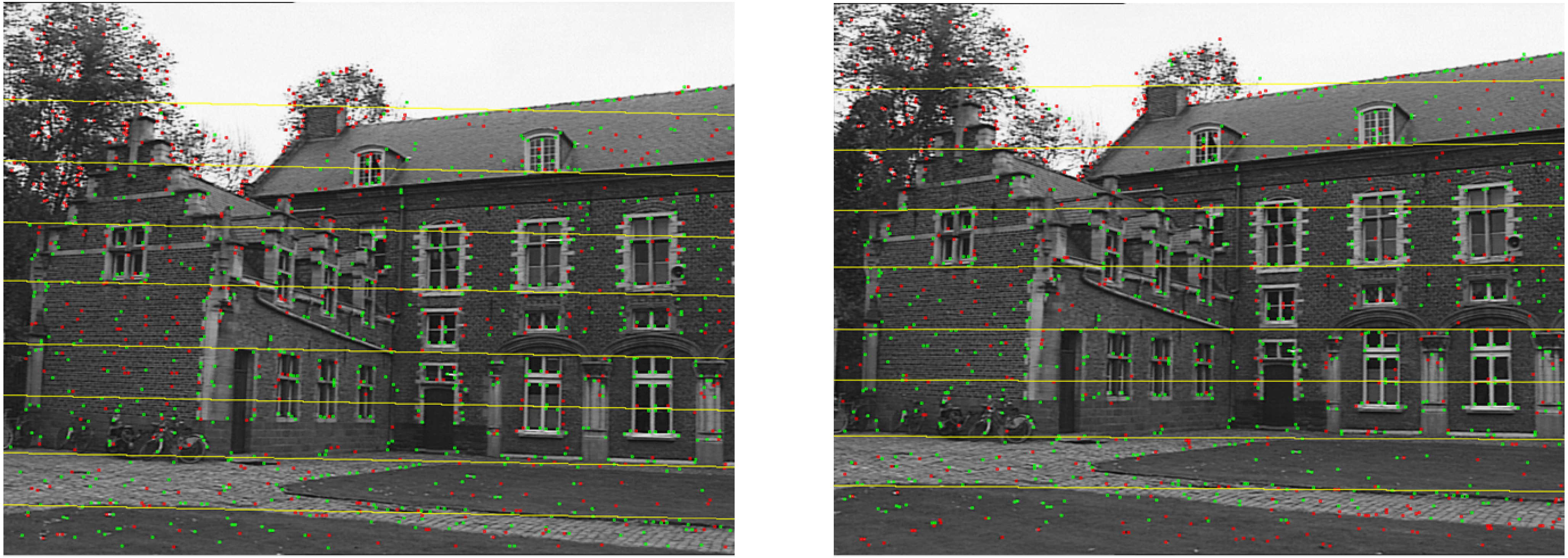
The epipolar constraint

- simplifies inter-image matching by bringing the additional constraint.
- encapsulates all information about relative position and orientation of two cameras in space.



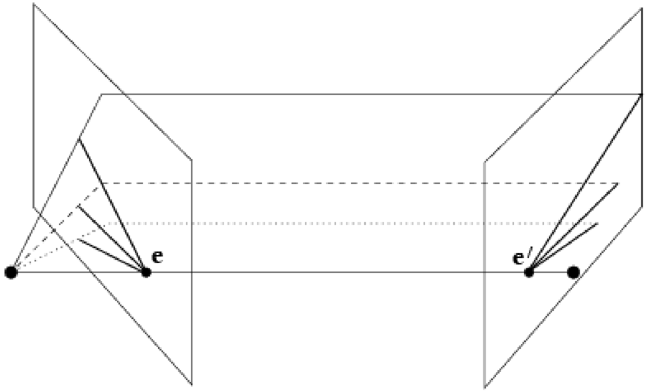
- Epipoles $\mathbf{e}$, $\mathbf{e}'$, epipolar lines $\mathbf{l}$, $\mathbf{l}'$.
- $\mathbf{e}$, $\mathbf{e}'$, $\mathbf{l}$, $\mathbf{l}'$, C, C', $\mathbf{X}$ lie in a single plane.
- Knowing epipolar geometry enables seeking correspondences as 1D task, i.e., between two 1D signals. It is expressed algebraically as a bilinear relation between $\mathbf{u}$, $\mathbf{u}'$.

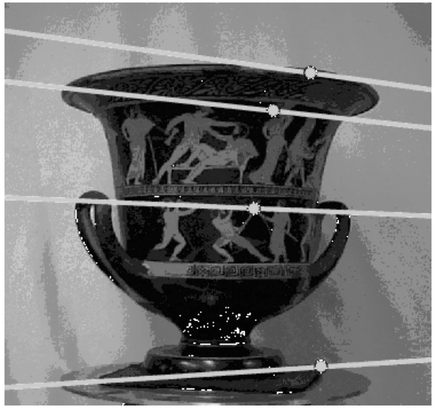
Epipolar lines example



Courtesy: M. Polefeys, ETH Zürich

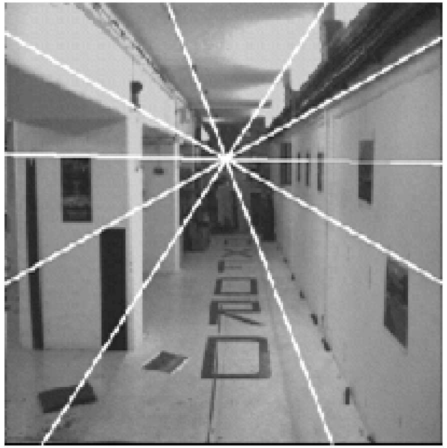
Epipolar lines example, converging views

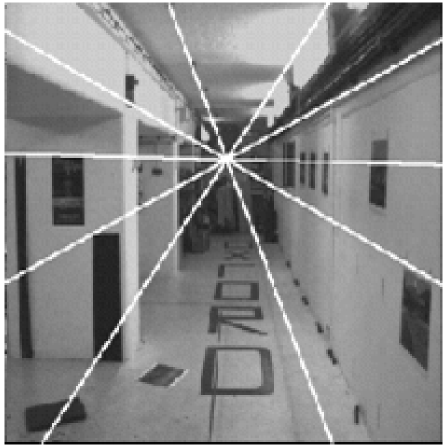


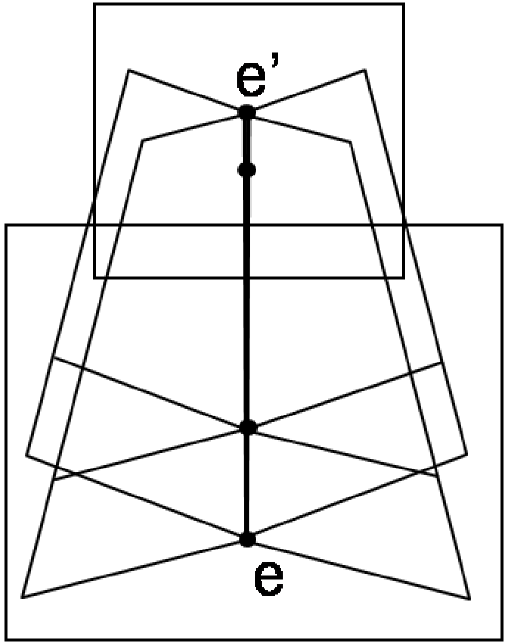


Courtesy: A. Zisserman, U of Oxford

Epipolar lines example, move to the front

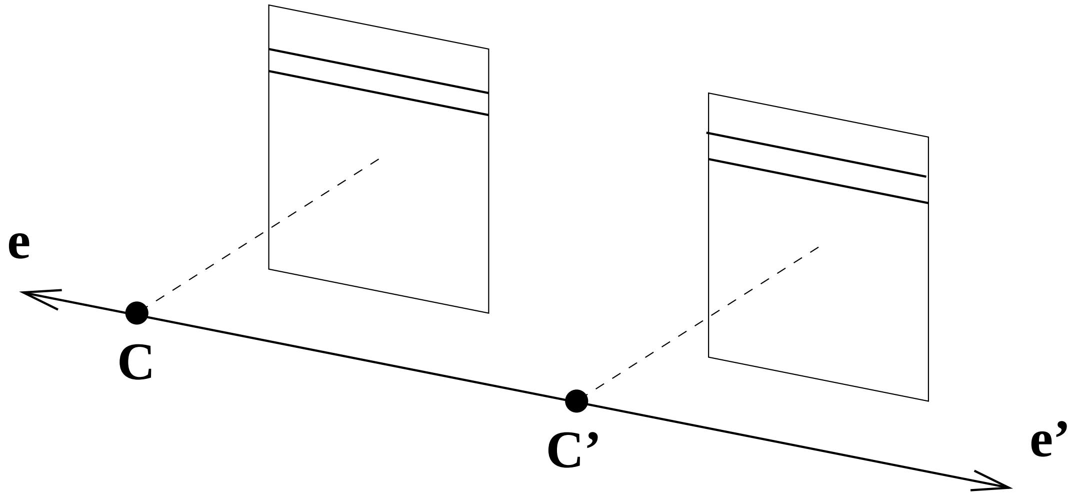






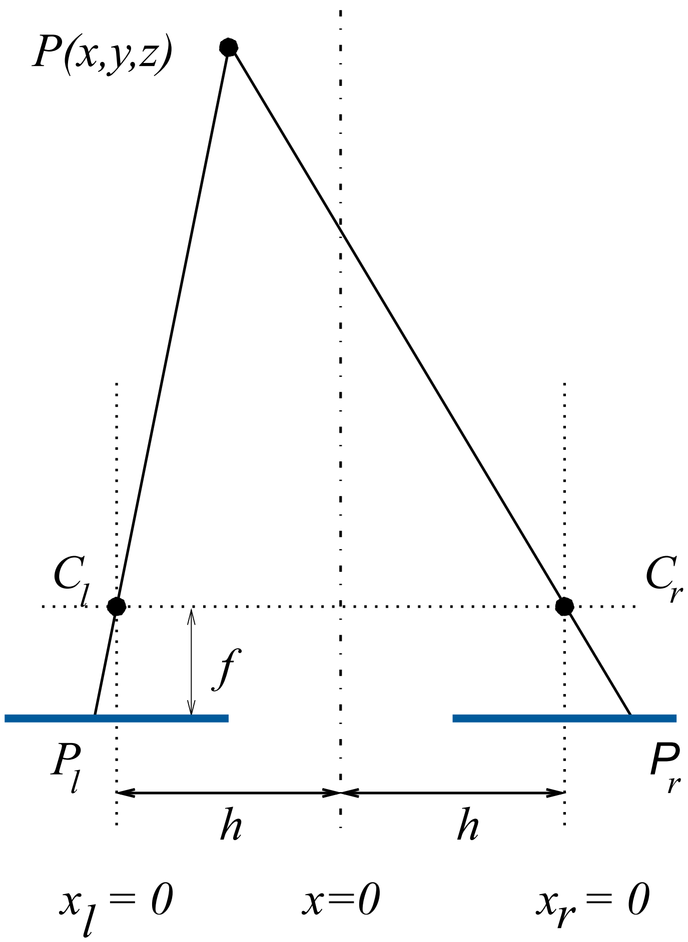
Courtesy A. Zisserman, U of Oxford

Canonical configuration of cameras



- Epipolar lines correspond to lines in images.
- It is often used when stereo correspondence is to be determined by a human operator who will find matching points linewise to be easier.
- Any pair of images with known epipolar geometry can be converted to canonical configuration by rectification.

Disparity and depth



baseline $2h$
disparity $|P_l - P_r| > 0$
focal length f

Calculation of depth (similar triangles)

$$\frac{P_l}{f} = -\frac{h+x}{z}\,, \qquad \frac{P_r}{f} = \frac{h-x}{z}$$

$$z\,(P_r - P_l) = 2hf$$
$$z = \frac{2hf}{P_r - P_l}$$

Note: if $P_r - P_l = 0$ then $z = \infty$.

Fundamental matrix (1)

Left projection $\mathbf{u}$ and right projection $\mathbf{u}'$ of the scene point $\mathbf{X}$.

$$
\begin{aligned}
\mathbf{u} &\simeq [K|\mathbf{0}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = K\,\mathbf{X}, \\
\mathbf{u}' &\simeq [K'R\,|-K'R\,\mathbf{t}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \\
&= K'(R\,\mathbf{X} - R\,\mathbf{t}) = K'\mathbf{X}'
\end{aligned}
$$

- Coplanarity of $\mathbf{X}$, $\mathbf{X}'$ and $\mathbf{t}$.
- Distinguish co-ordinates of the left and right cameras by the subscript $_L$, $_R$.
- Vector product $\times$.

Fundamental matrix (2)

- Coordinates rotation $\mathbf{X}'_R = R\,\mathbf{X}'_L$, and hence $\mathbf{X}'_L = R^{-1}\mathbf{X}'_R$.
- Coplanarity constraint $\mathbf{X}_L^\top(\mathbf{t} \times \mathbf{X'}_L) = 0$.
- Preparing for substitution $\mathbf{X}_L = K^{-1}\mathbf{u}$, $\mathbf{X}'_R = (K')^{-1}\mathbf{u}'$, and $\mathbf{X}'_L = R^{-1}(K')^{-1}\mathbf{u}'$.
- Epipolar constraint in the vector form

$$(K^{-1}\mathbf{u})^\top(\mathbf{t} \times R^{-1}\,(K')^{-1}\mathbf{u}') = 0\,.$$

- Equation is homogeneous with respect to $\mathbf{t}$, so the scale is not determined.
- Absolute scale cannot be recovered without ‘yardstick’.

Fundamental matrix (3)

Replacement of a vector product by a matrix multiplication, Rodriques' rotation formula, [O. Rodriques, 1840], see derivation in Wikipedia.

The translation vector is $\mathbf{t} = [t_x, t_y, t_z]^\top$, and a skew symmetric matrix $S(\mathbf{t})$ (i.e., $S^\top = -S$) can be created from it if $\mathbf{t} \neq \mathbf{0}$.

$$S(\mathbf{t}) = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Note that $\text{rank}(S) = 2$ if and only if $\mathbf{t} \neq \mathbf{0}$.

For any regular matrix A, we have

$$\mathbf{t} \times A \; = S(\mathbf{t})\, A\,.$$

Fundamental matrix (4)

- The vector product can be replaced by the multiplication of two matrices.
- From previous slide, for any regular matrix A, we have $\mathbf{t} \times A \; = S(\mathbf{t})\, A$.
- Consequently we can rewrite the epipolar constraint from the vector form

$$(K^{-1}\mathbf{u})^\top(\mathbf{t} \times R^{-1}\,(K')^{-1}\mathbf{u}') = 0\,.$$

to a matrix form

$$(K^{-1}\mathbf{u})^\top\,(S(\mathbf{t})\,R^{-1}\,(K')^{-1}\mathbf{u}') = 0\,,$$
$$\mathbf{u}^\top(K^{-1})^\top S(\mathbf{t})R^{-1}(K')^{-1}\mathbf{u}' = 0\,.$$

Fundamental matrix (5)

The middle part can be concentrated into a single matrix F called the fundamental matrix of two views,

$$F = (K^{-1})^\top S(\mathbf{t}) R^{-1} (K')^{-1} \, .$$

With the substitution for F we finally get the bilinear relation (sometimes named after C. Longuet-Higgins) between any two views

$$\mathbf{u}^\top \, F \, \mathbf{u}' = 0 \, .$$

It can be seen that the fundamental matrix F captures all information that can be recovered from a pair of images if the correspondence problem is solved.

Relative motion of the camera
Essential matrix E

- A single camera moving in space, or two cameras with known calibration.
- Known calibration matrices K, K' allows us to normalize measurement in left and right images $\breve{\mathbf{u}}$, $\breve{\mathbf{u}}'$.

$$\breve{\mathbf{u}} = K^{-1}\mathbf{u}, \quad \breve{\mathbf{u}}' = (K')^{-1}\mathbf{u}'$$

- Substitute into

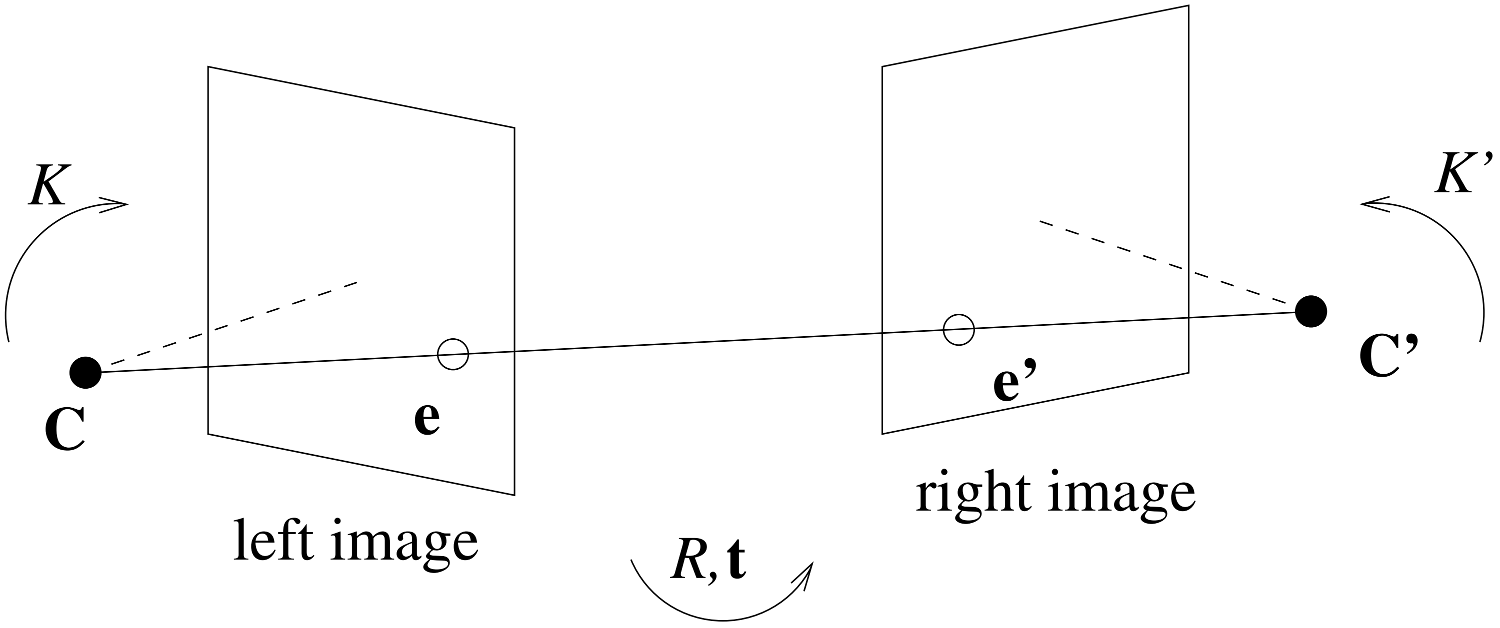
$$\mathbf{u}^\top (K^{-1})^\top S(\mathbf{t}) R^{-1} (K')^{-1} \mathbf{u}' = 0$$

$$\begin{aligned} \breve{\mathbf{u}}^\top \, S(\mathbf{t}) R^{-1} \, \breve{\mathbf{u}}' &= 0 \\ \breve{\mathbf{u}}^\top \, E \, \breve{\mathbf{u}}' &= 0 \end{aligned}$$

- The essential matrix E captures all the information about the relative motion from the first to the second position of the calibrated camera.

Properties of the essential matrix E

- The essential matrix E has rank 2.
- Let $\mathbf{t}$ be the translational vector, and $\mathbf{t}' = R\,\mathbf{t}$.
  There holds $E\,\mathbf{t}' = 0$ and $\mathbf{t}^\top E = 0$.



Properties of the essential matrix E

SVD decomposes E as $E = UDV^\top$ for a diagonal D;

$$D = \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Rotation R and translation $\mathbf{t}$ from E

[Hartley 1992] We have seen $E = S(\mathbf{t})R^{-1}$.

$$\breve{\mathbf{u}}^\top S(\mathbf{t})R^{-1}\,\breve{\mathbf{u}}' = 0\,, \quad \breve{\mathbf{u}}'^\top RS(\mathbf{t})\,\breve{\mathbf{u}} = 0\,, \quad E = RS(\mathbf{t})$$

$$G = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The rotation matrix R can be calculated using SVD: $E = UDV^\top$.

$$R = UGV^\top \text{ or } R = UG^\top V^\top$$

Components of the translation vector can be derived from the matrix $S(\mathbf{t})$ expressed as 3×3 matrix.

$$S(\mathbf{t}) = VZV^\top$$

Properties of the fundamental matrix F

- rank$(E) = 2$. As $F = (K^{-1})^\top E K'^{-1}$ and the calibration matrices are regular $\Rightarrow$ F rank$(F) = 2$.
- Consider two epipoles $\mathbf{e}, \mathbf{e}'$.

$$\mathbf{e}^\top F = 0 \text{ and } F\,\mathbf{e}' = 0$$

- SVD of the fundamental matrix gives $F = UDV^\top$, where

$$D = \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; k_1 \neq k_2 \neq 0$$

Estimating F, 8-point algorithm

- Epipolar geometry has 7 degrees of freedom. Epipoles $\mathbf{e}$, $\mathbf{e}'$ have two co-ordinates each (giving four DOF), while another three come from the mapping of any three epipolar lines in the first image to the second image.
- Thus the correspondence of 7 points in left and right images enables the establishment of the fundamental matrix F using a nonlinear 7-points algorithm, numerically unstable.
- If there are eight non-coplanar corresponding points available then the linear 8-point algorithm is easier to use. We prefer to apply its overconstrained and robust version.
  - Least squares solution using SVD on (many) equations from 8 pairs of correspondences.
  - Enforce $\det(F) = 0$ constraint using SVD on the fundamental matrix F.

8-point algorithm (2)

$$\mathbf{u}_i{}^\top F \, \mathbf{u}'_i = 0 \,, \quad \mathbf{u}^\top = [u_i, v_i, 1]$$

The 3×3 fundamental matrix F has only eight unknowns as it is only known up to scale $\Rightarrow$ 8 correspondences.

$$[u_i,\ v_i,\ 1] \, F \begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix} = 0$$

8-point algorithm (3)

- Rewriting the elements of the fundamental matrix as a column vector with nine elements $\mathbf{f}^\top = [f_{11}, f_{12}, \ldots, f_{33}]$, can be rewritten as a system of linear equations;
  Consider we have available n points in correspondence.

$$\begin{bmatrix} u_1 u'_1 & u_1 v'_1 & u_1 & v_1 u'_1 & v_1 v'_1 & v_1 & u'_1 & v'_1 & 1 \\ & & & & \vdots & & & & \\ u_n u'_n & u_n v'_n & u_n & v_n u'_n & v_n v'_n & v_n & u'_n & v'_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ \vdots \\ f_{33} \end{bmatrix} = A\,\mathbf{f} = 0$$

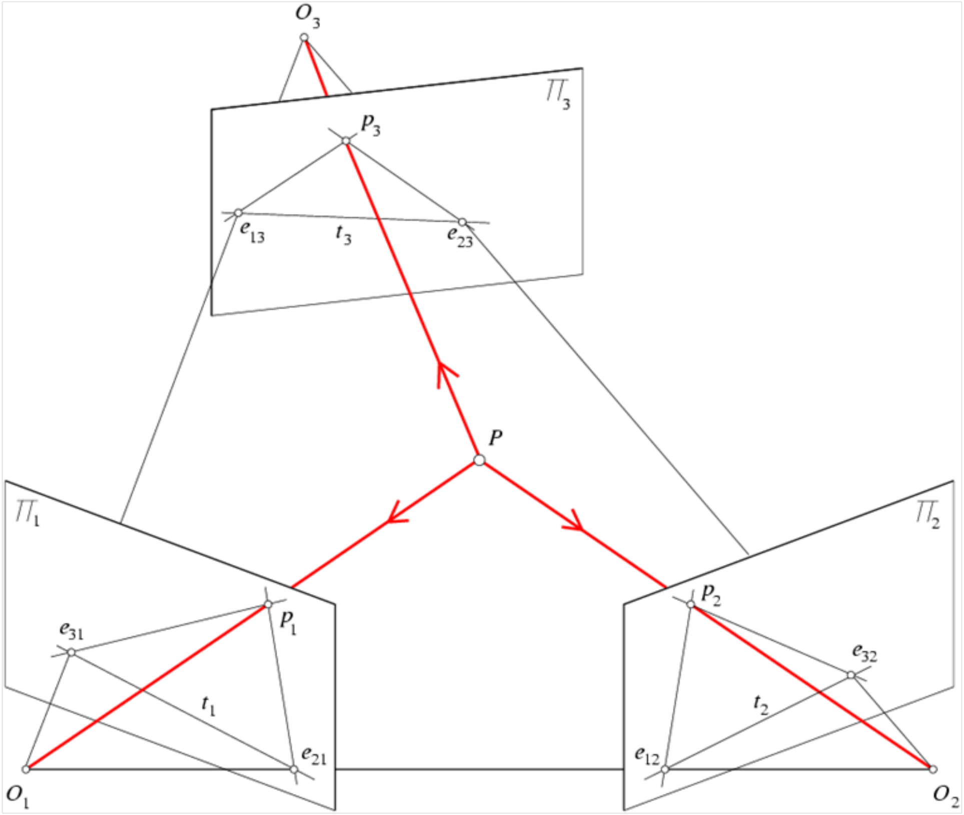
- Instead of solving $A\mathbf{f} = 0$, we seek f, which minimizes the algebraic error $||A\mathbf{f}||$.
- There is a better alternative, minimizing the geometric error (units $=$ pixels) $\min \sum_j ||x_1^j - F\, x_2^j||^2$.

8-point algorithm (4)

8-point algorithm used to determine parameters of the fundamental matrix F (or analogously of the essential matrix E)

1. Solve a system of (overconstraint) homogeneous linear equations
   (a) Write down the system of linear equations $A\mathbf{f} = 0$
   (b) Solve $\mathbf{f}$ from $A\mathbf{f} = 0$ using SVD
   MATLAB:
   ```
   [U, S, V] = svd(A);
   f = V(:, end);
   F = reshape(f, [3 3]);
   ```
2. Resolve $\det(F) = 0$ constraint using SVD
   MATLAB:
   ```
   [U, S, V] = svd(F);
   S(3,3) = 0;
   F = U * S * V';
   ```

Trinocular constraint (also trifocal constraint)



Trinocular constraint, transfer