STEREOPSIS

PART I

computer vision

INTRODUCTION

- we refer to stereo vision as the problem of inferring 3D information (structure and distances) from two or more images taken from different viewpoints

- we consider an acquisition system with 2 cameras
  - "explicit" system: a stereo rig
  - "implicit": one moving camera

PROBLEMS IN STEREO VISION

- stereo covers two main problems:
    - finding **correspondences** between image pairs
    - **reconstructing** the 3D position of a point given its corresponding projections on the images
- step I: produces 2.5 disparity maps (or depth maps)
- step II: produces a 3D point cloud

before we do so, we need to understand the geometry of a stereo system…

A SIMPLE STEREO SYSTEM

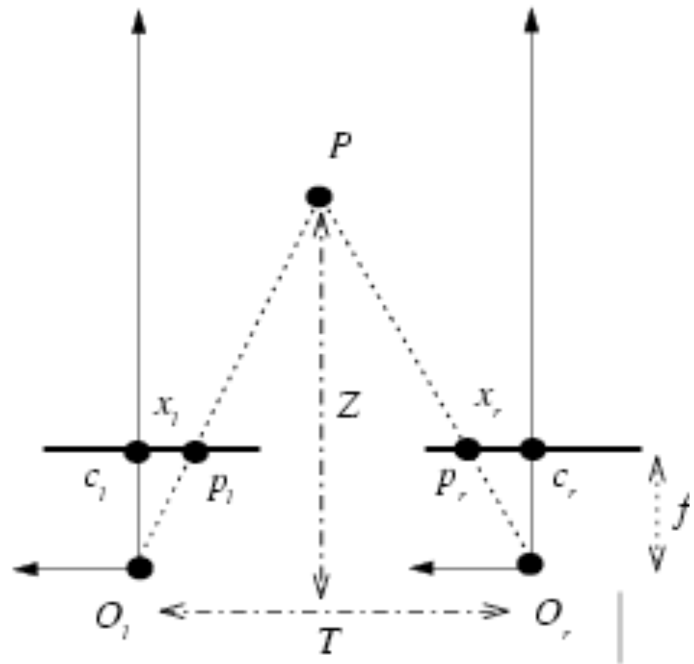
fixation point at infinity

- the solution of the problem is

$$Z = \frac{fT}{x_r - x_l} = \frac{fT}{d}$$

where d is the *disparity*

- depth is inversely proportional to disparity



CAUTION...

- the model described in the previous slide is not general, as it suggests that the disparity decreases as the distance between 3D point and cameras increases

- the truth is that *disparity increases with the distance of the 3D point from the fixation points*.
  - why? (hint: draw the model of a stereo system with a finite fixation point and study the projection of 3D points in different positions)

PARAMETERS OF A STEREO SYSTEM

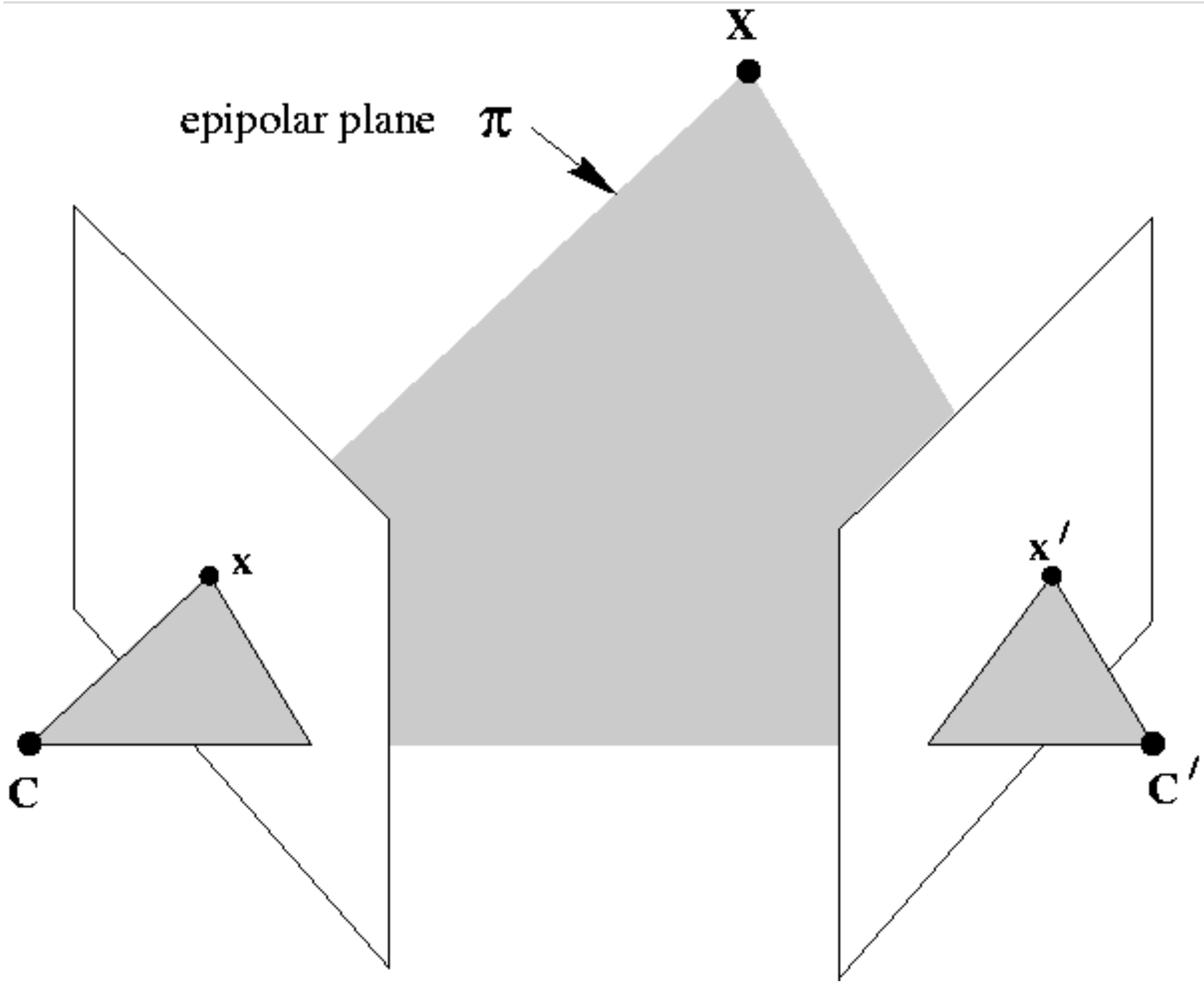
- intrinsic parameters
  - characterize the mapping of an image point from camera to pixel coordinates in each camera
  - two full-rank 3x3 upper triangular matrices K_l and K_r
- extrinsic parameters
  - describe the relative position and orientation of the two cameras (R, **T**)

$$\mathbf{X}_r = R(\mathbf{X}_l - \mathbf{T})$$

  - without loss of generality we may assume the origin of the world ref frame corresponds to the left camera

EPIPOLAR GEOMETRY

- the projective geometry of 2 views



EPIPOLAR CONSTRAINTS

- let us consider ***x*** How is ***x'*** constrained?

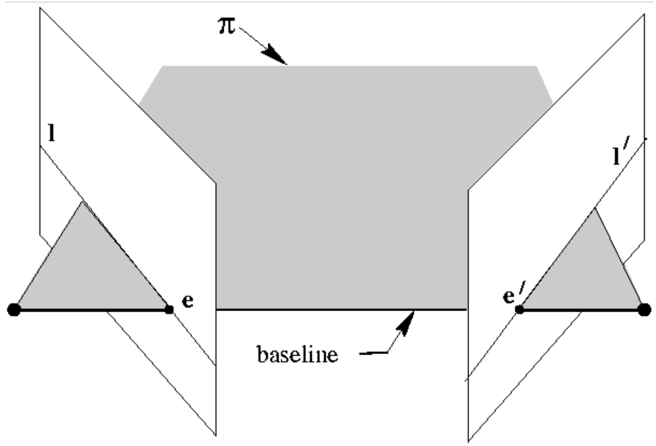
- ***x'*** lies on the line ***l'*** intersection between the epipolar plane and the (right) image plane

- ***l'*** is the projection on the (right) image plane of the ray passing through **C** and ***x***

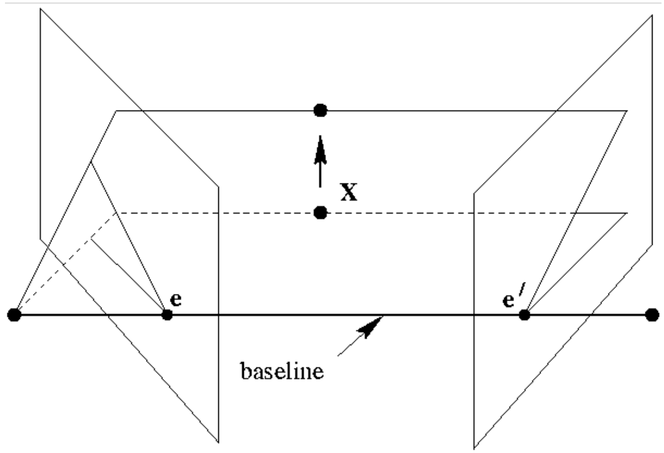
- In practice, when we look for the corresponding point ***x'*** we can limit our search to a line, ***l'***

TERMINOLOGY

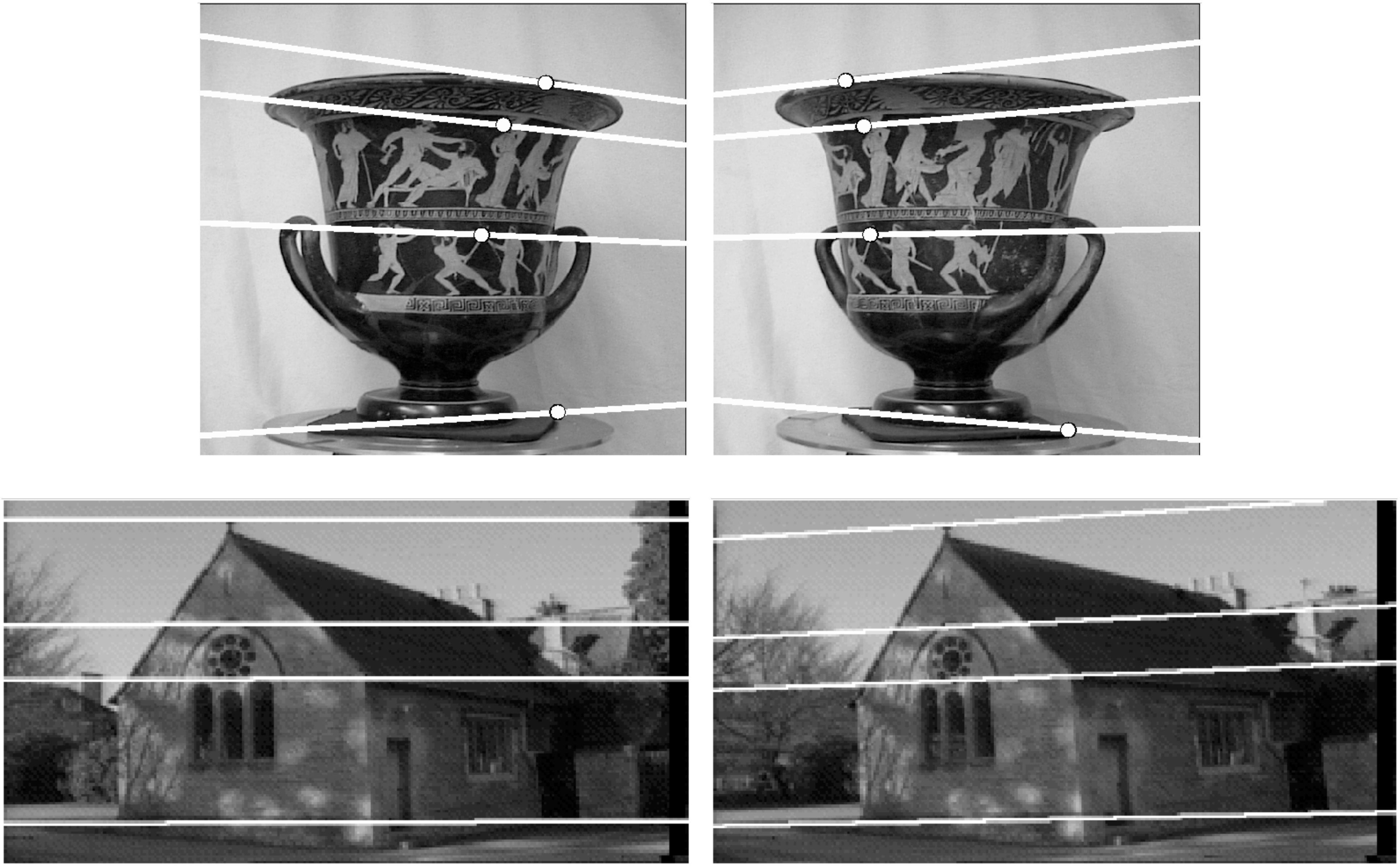
- **epipole:** the intersection of the baseline with the 2 image planes. Also: the projection of one camera centre to the other image plane



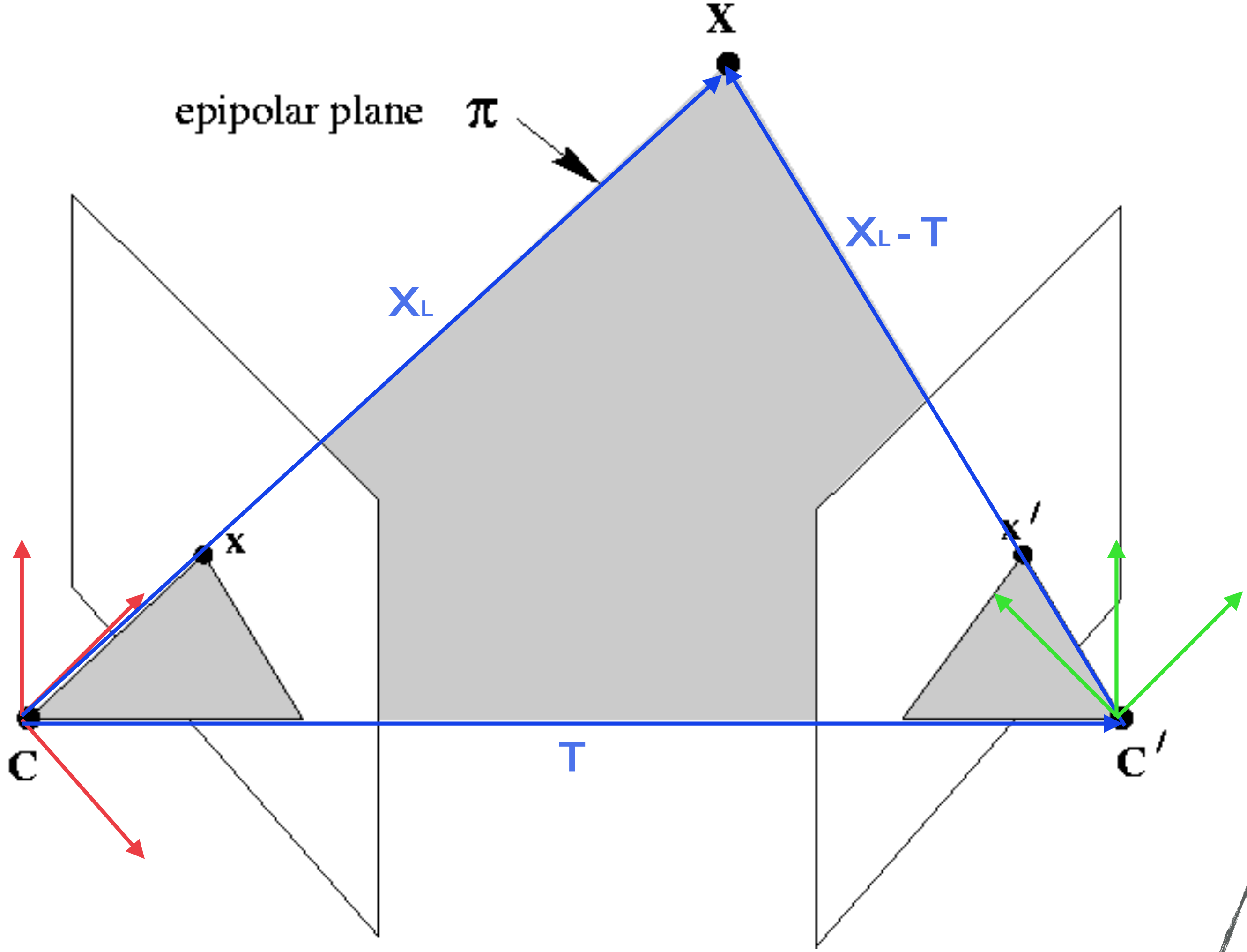
- **epipolar plane:** a plane containing the baseline and passing through a 3D point ***X***. It is a *one parameter family (stencil)* of planes



- **epipolar line:** the intersection of an epipolar plane with one image plane. All epipolar lines contain the epipole.



EPIPOLAR GEOMETRY



$$S = \begin{bmatrix} 0 & -T_3 & T_2 \\ T_3 & 0 & -T_1 \\ -T_2 & T_1 & 0 \end{bmatrix}$$

$\mathbf{X}_l, \mathbf{T}, \mathbf{X}_l - \mathbf{T}$ are coplanar $\Rightarrow$ $(\mathbf{X}_l - \mathbf{T})^\top \mathbf{T} \times \mathbf{X}_l = 0$

$$(R^\top \mathbf{X}_r)^\top \mathbf{S}\mathbf{X}_l = 0$$

$$\mathbf{X}_r = R(\mathbf{X}_l - \mathbf{T})$$

$$\mathbf{X}_r^\top R\mathbf{S}\mathbf{X}_l = 0$$

THE ESSENTIAL MATRIX

- The essential matrix is a 3x3 matrix E=SR $\mathbf{X}_r^\top E \mathbf{X}_l = 0$
- by projecting the point on the two image planes we obtain the equation $\mathbf{x_m}'^\top E \mathbf{x_m} = 0$
  which is *satisfied by each (x_m x_m') pair of corresponding points in mm coordinates*
  - E has rank 2, since S has rank 2 and R is full rank
  - E has 5 d.o.f. : 3 from R 3 from **T** <u>but with a scale ambiguity</u>
    3+3-1=5
  - A 3x3 matrix is an essential matrix iff 2 of its singular values are equal, and the third is 0

THE FUNDAMENTAL MATRIX

- we now derive an equivalent equation relating points in pixel coordinates:

$$\mathbf{x}_m = K_l^{-1}\mathbf{x}$$

$$\mathbf{x}'_m = K_r^{-1}\mathbf{x}'$$

- Then

$$\mathbf{x}'^{\top} \underbrace{K_r^{-T} E K_l^{-1}}_{F} \mathbf{x} = 0$$

- The fundamental matrix satisfies the equation

$$\mathbf{x}'^{\top} F \mathbf{x} = 0$$

for each pair of (x, x') of corresponding points in pixel coordinates

WRAP UP ON F: ITS DEFINITION

- Suppose we have two images acquired by cameras with no coinciding centres.

- Then the *fundamental matrix* F is the unique 3x3 rank 2 matrix so that, for each corresponding pair (x,x')

$$\mathbf{x}'^{\top} F \mathbf{x} = 0$$

- Fundamental (and essential) matrix is a map between points and (epipolar) lines

$$\mathbf{l}' = F\mathbf{x}$$

PROPERTIES OF F

- the transpose of F, $F^\top$, models the cameras pair in the opposite order

$$\mathbf{x}^\top F^\top \mathbf{x}' = 0$$

$\mathbf{l} = F^\top \mathbf{x}'$ epipolar line corresponding to x'

- the epipole e_R is the left null space of F (similarly e_L is its right null space) $\mathbf{e}_R{}^\top F = 0$

$$\mathbf{e}_R{}^\top \mathbf{l}' = \mathbf{e}_R{}^\top F \mathbf{x} = 0 \;\; \forall \mathbf{x}$$

- F has 7 d.o.f. since it is a 3x3 homogeneous matrix but has the further constraint of det(F)=0

THE 8 POINTS ALGORITHM

$$\mathbf{x} = (x, y, 1)^\top \quad \mathbf{x}' = (x', y', 1)^\top$$

- we may obtain a **linear algorithm** from 8 points correspondences (notice this is **not** the minimal problem)

$$\begin{bmatrix} {x_1}'x_1 & {x_1}'y_1 & {x_1}' & {y_1}'x_1 & {y_1}'y_1 & {y_1}' & x_1 & y_1 & 1 \\ x'_2x_2 & x'_2y_2 & x'_2 & y'_2x_2 & y'_2y_2 & y'_2 & x_2 & y_2 & 1 \\ & . & . & . & . & . & . & . & \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ . \\ . \\ . \\ f_{33} \end{pmatrix} = \mathbf{0}$$

- It is a homogeneous system $A\mathbf{f}=\mathbf{0}$, overdetermined, which may be solved with SVD

-

THE 8 POINTS ALGORITHM

Rank enforcement

- the rank of the estimated F in general will be 3
- we then look for a matrix F' "close" to the one we estimated but with det(F')=0

$$
\begin{aligned}
F &= UDV^\top \\
D &= \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}
\end{aligned}
\qquad
D' = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & 0 \end{bmatrix}
$$

$$F' = UD'V^\top$$

POINT NORMALIZATION

- the F estimated with this approach is not invariant to point transformations.
  - for this reason (and also to control the conditioning number of the matrix A) it is advisable to normalize points as follows
- **translate** points so that the image coordinate center corresponds to the points centroid
- **scale** points so that their average distance from the origin is $\sqrt{2}$

$$T = \begin{bmatrix} s & 0 & -sc_x \\ 0 & s & -sc_y \\ 0 & 0 & 1 \end{bmatrix}$$

FUNDAMENTAL MATRIX ESTIMATION SUMMARY

- let us consider $$\tilde{\mathbf{x}} = T\mathbf{x}$$ $$\tilde{\mathbf{x}}' = T'\mathbf{x}'$$
- compute *T* and *T'* for points in the left and right image
- normalize points obtaining $(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}')$
- compute $\tilde{F}$ with the 8 points algorithm starting from $(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}')$
- denormalize the matrix as follows $F = T'^{\top}\tilde{F}T$

ROBUST ESTIMATION

- what if some of our correspondences are wrong?
- we may want to find the inliers (correct matches) and compute the fundamental matrix on these points only
- We may consider the RANSAC method (RANdom SAmpling Consensus)