Instance-level recognition

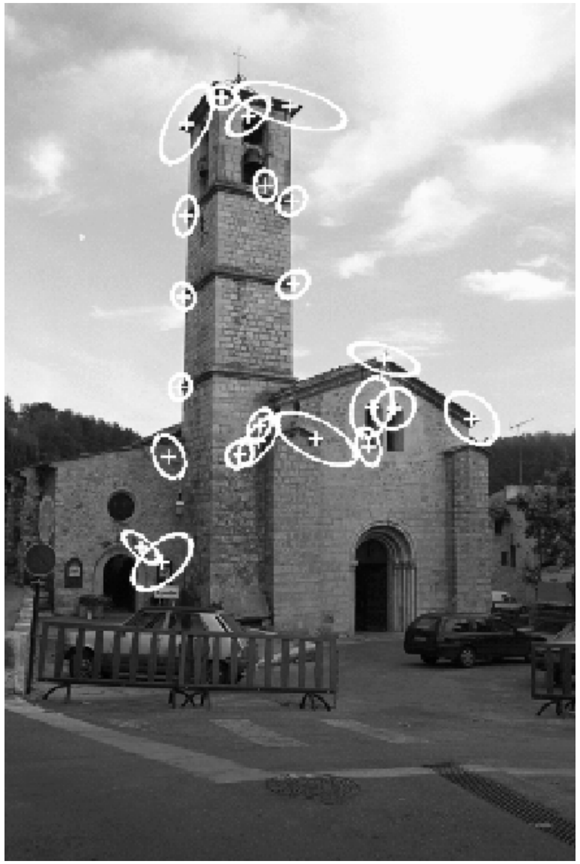
1) Local invariant features

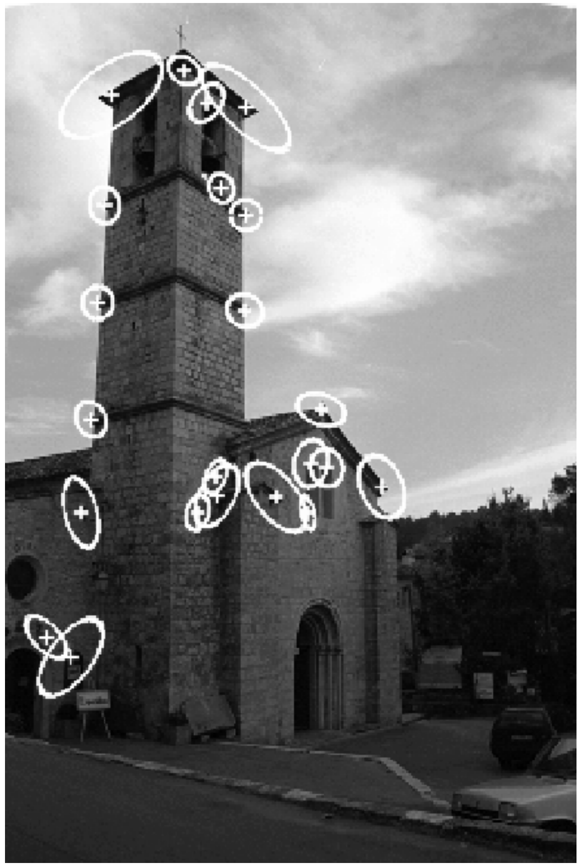
2) Matching and recognition with local features

3) Efficient visual search

4) Very large scale indexing

Matching of descriptors





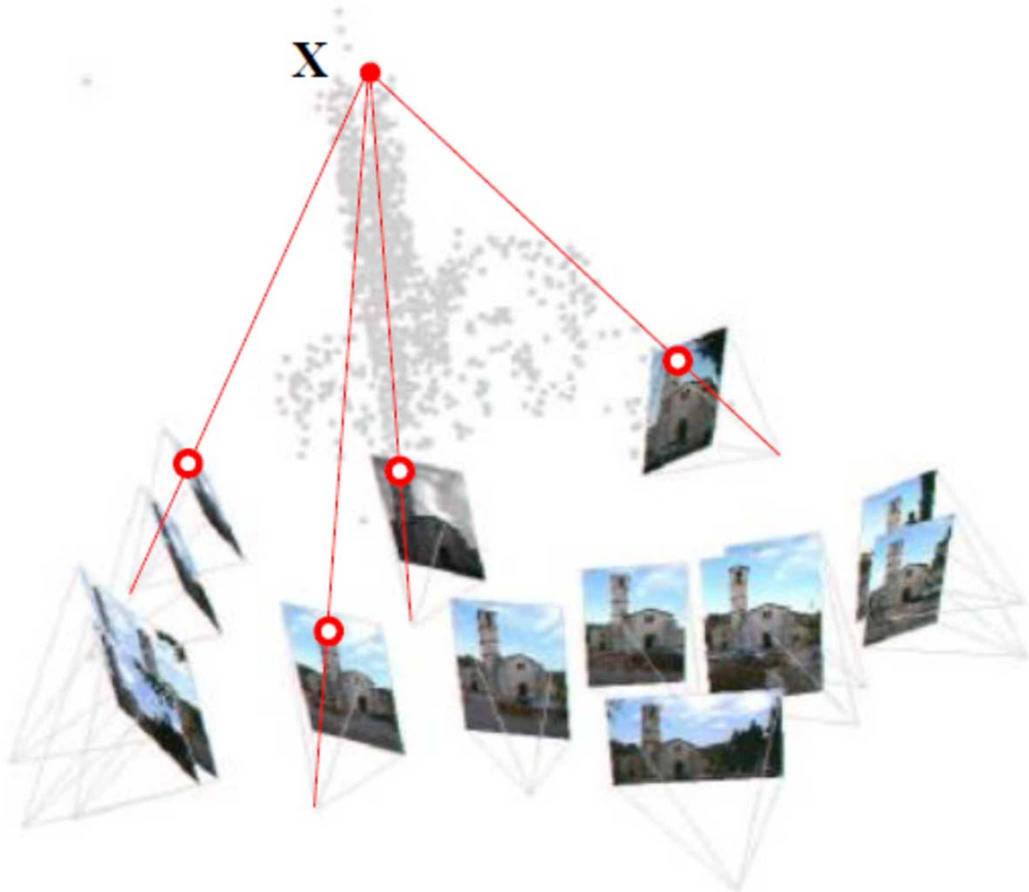
Matching and 3D reconstruction

- Establish correspondence between two (or more) images

[Schaffalitzky and Zisserman ECCV 2002]

Matching and 3D reconstruction

- Establish correspondence between two (or more) images



[Schaffalitzky and Zisserman ECCV 2002]

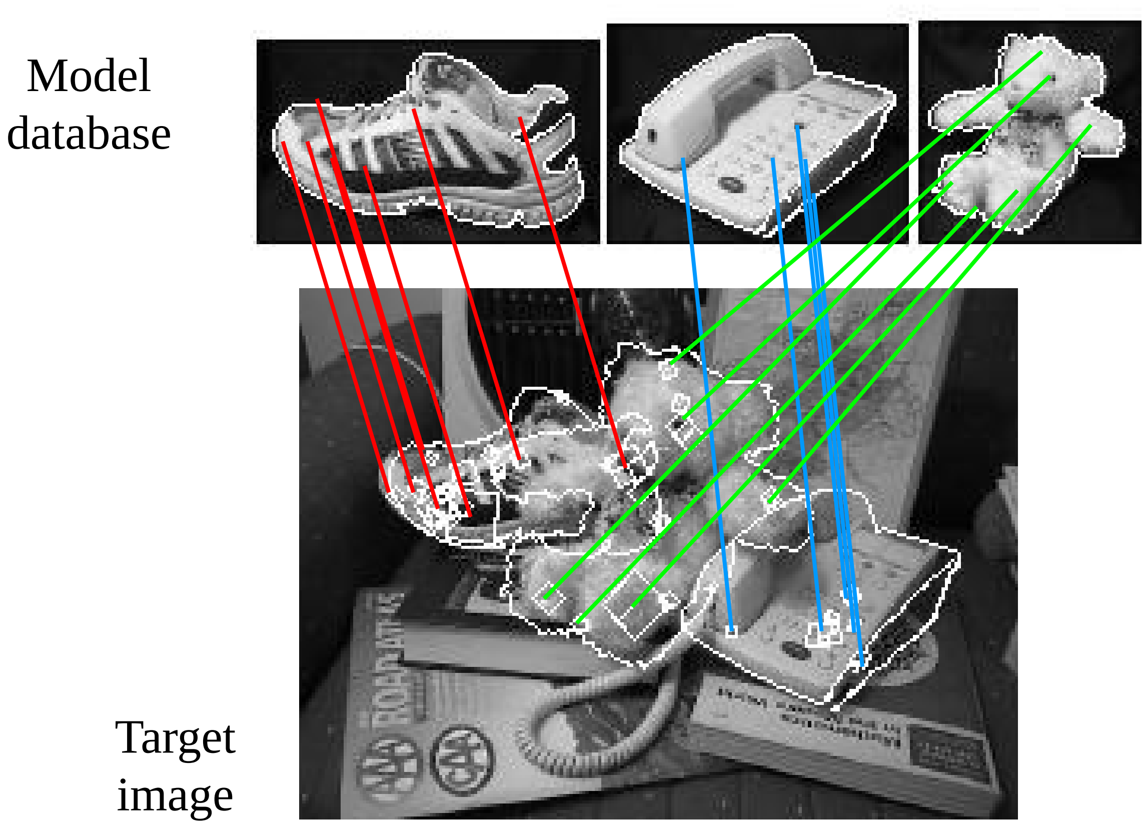
Building Rome in a Day

57,845 downloaded images, 11,868 registered images

[Agarwal, Snavely, Simon, Seitz, Szeliski, ICCV'09]

Object recognition

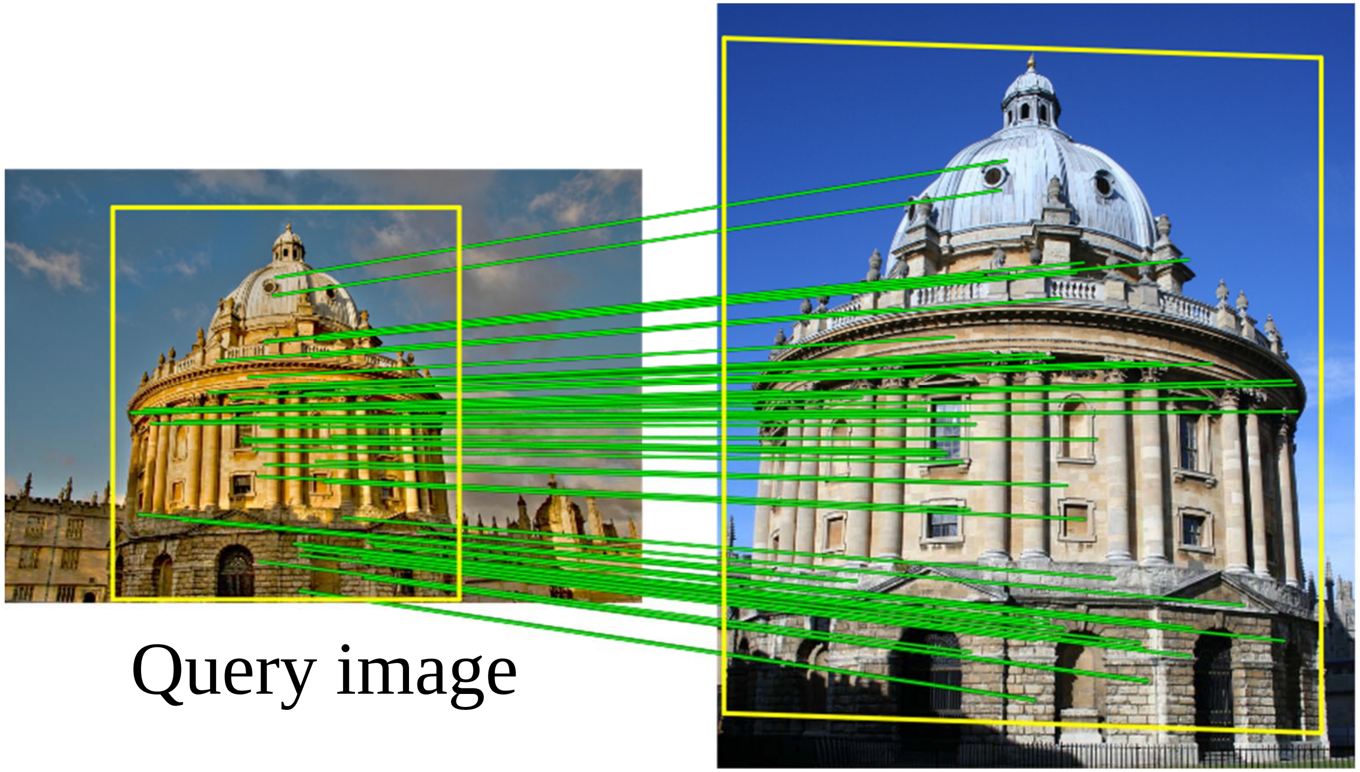
- Establish correspondence between the target image and (multiple) images in the model database



[D. Lowe, 1999]

Visual search

- Establish correspondence between the query image and all images from the database depicting the same object or scene

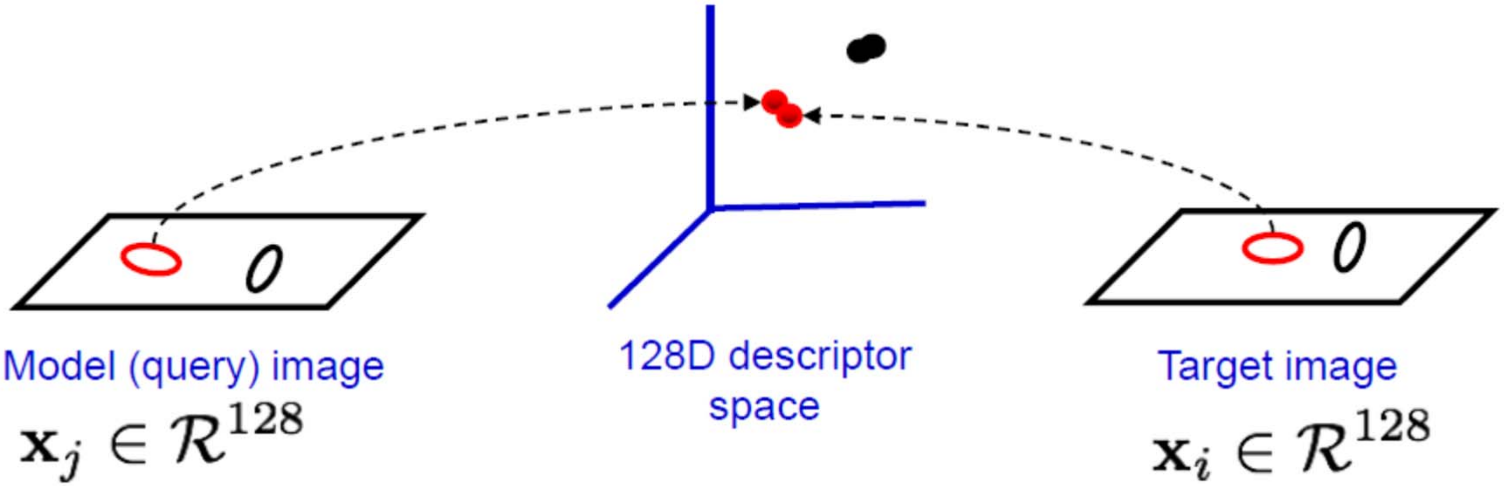


Query image

Database image(s)

Matching of descriptors

- Find the nearest neighbor in the second image for each descriptor, for example SIFT



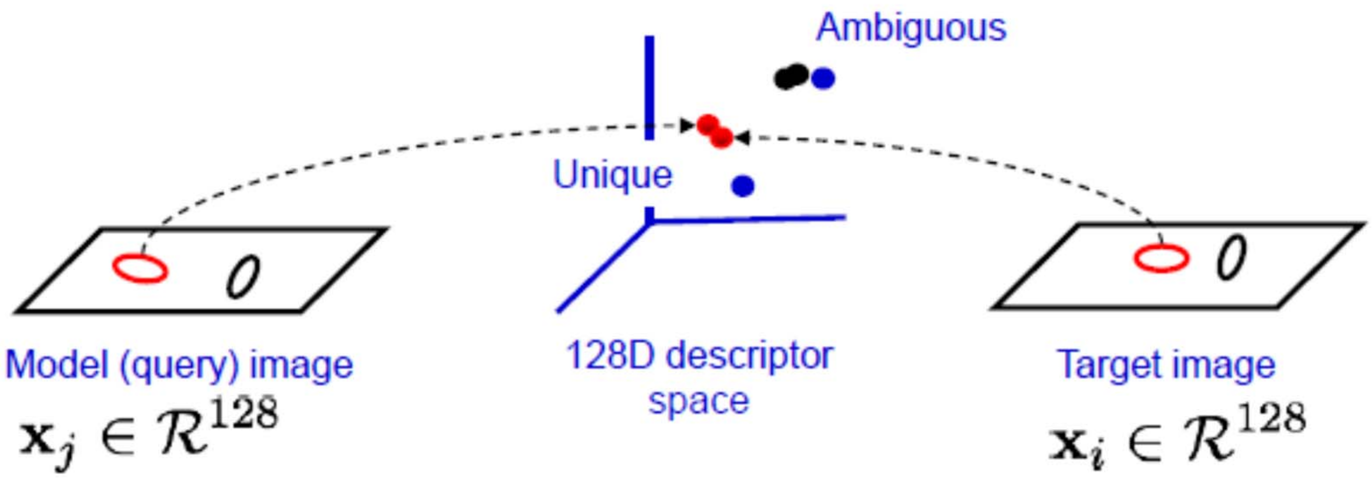
Need to solve some variant of the "nearest neighbor problem" for all feature vectors, $\mathbf{x}_j \in \mathcal{R}^{128}$, in the query image:

$$\forall j \; NN(j) = \arg\min_i ||\mathbf{x}_i - \mathbf{x}_j||,$$

where, $\mathbf{x}_i \in \mathcal{R}^{128}$, are features in the target image.

Matching of descriptors

- Pruning strategies
  - Ratio with respect to the second best match (d1/d2 << 1) [Lowe, '04]



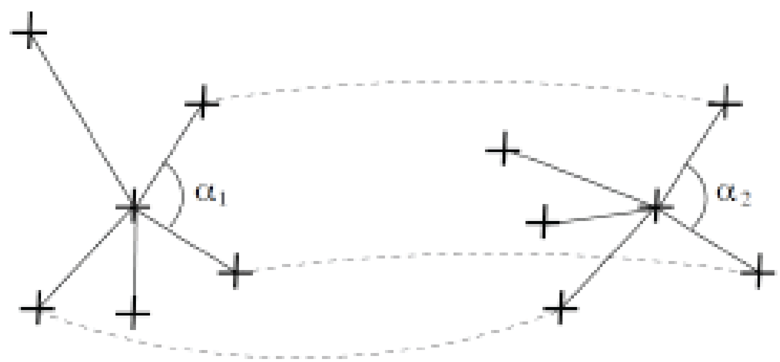
If the 2nd nearest neighbour is much further than the 1st nearest neighbour, the match is more "unique" or discriminative.

Measure this by the ratio: $r = d_{1NN} / d_{2NN}$

r is between 0 and 1
r is small the match is more unique.

Matching of descriptors

- Pruning strategies
  - Ratio with respect to the second best match (d1/d2 << 1)
  - Local neighborhood constraints (semi-local constraints)



Neighbors of the point have to match and angles have to correspond.
Note that in practice not all neighbors have to be matched correctly.

Matching of descriptors

- Pruning strategies
  - Ratio with respect to the second best match (d1/d2 << 1)
  - Local neighborhood constraints (semi-local constraints)
  - Backwards matching (matches are NN in both directions)

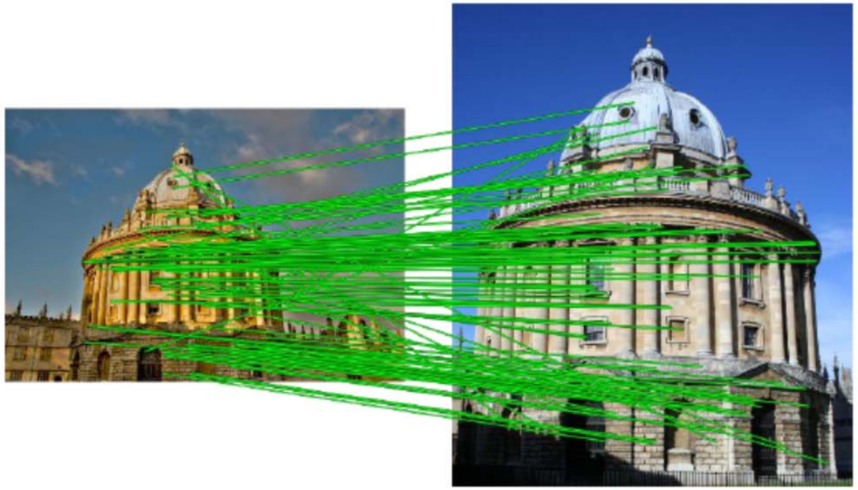
Matching of descriptors

- Pruning strategies
  - Ratio with respect to the second best match (d1/d2 << 1)
  - Local neighborhood constraints (semi-local constraints)
  - Backwards matching (matches are NN in both directions)

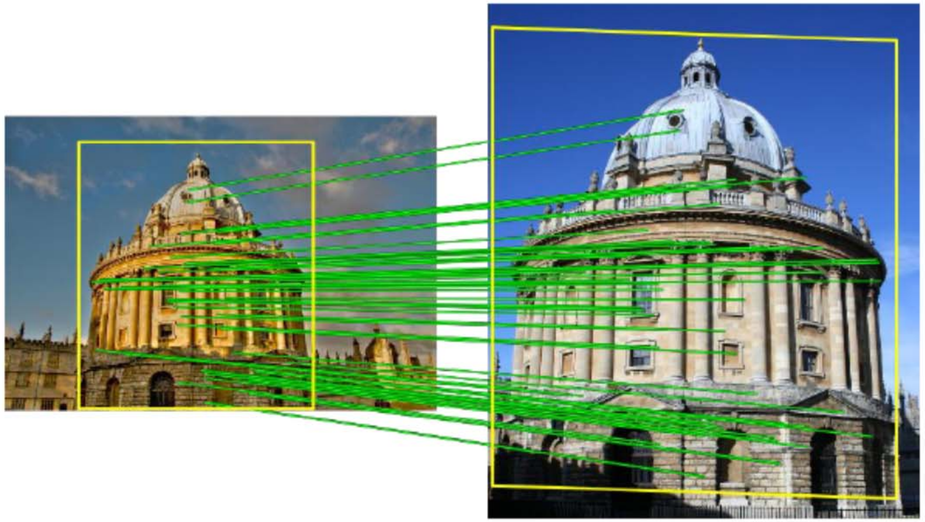
- Geometric verification with global constraint
  - All matches must be consistent with a global geometric transformation
  - However, there are many incorrect matches
  - Need to estimate simultaneously the geometric transformation and the set of consistent matches

Geometric verification with global constraint

- Example of a geometric verification

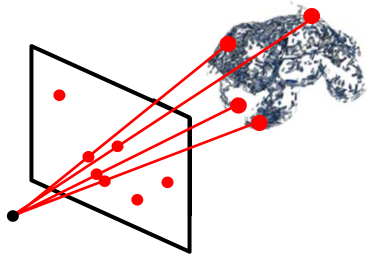


Tentative matches



Matches consistent with an affine transformation

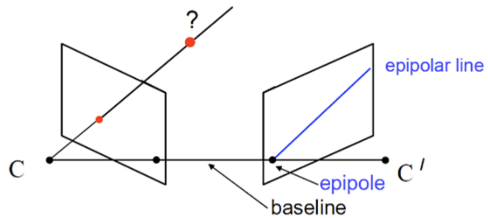
Examples of global constraints

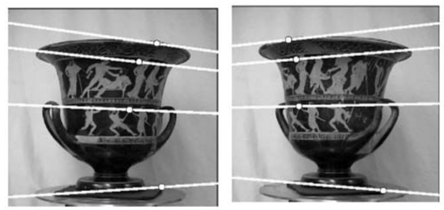


1 view and known 3D model.

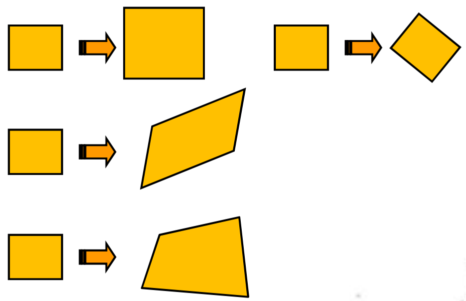
- Consistency with a (known) 3D model.

2 views





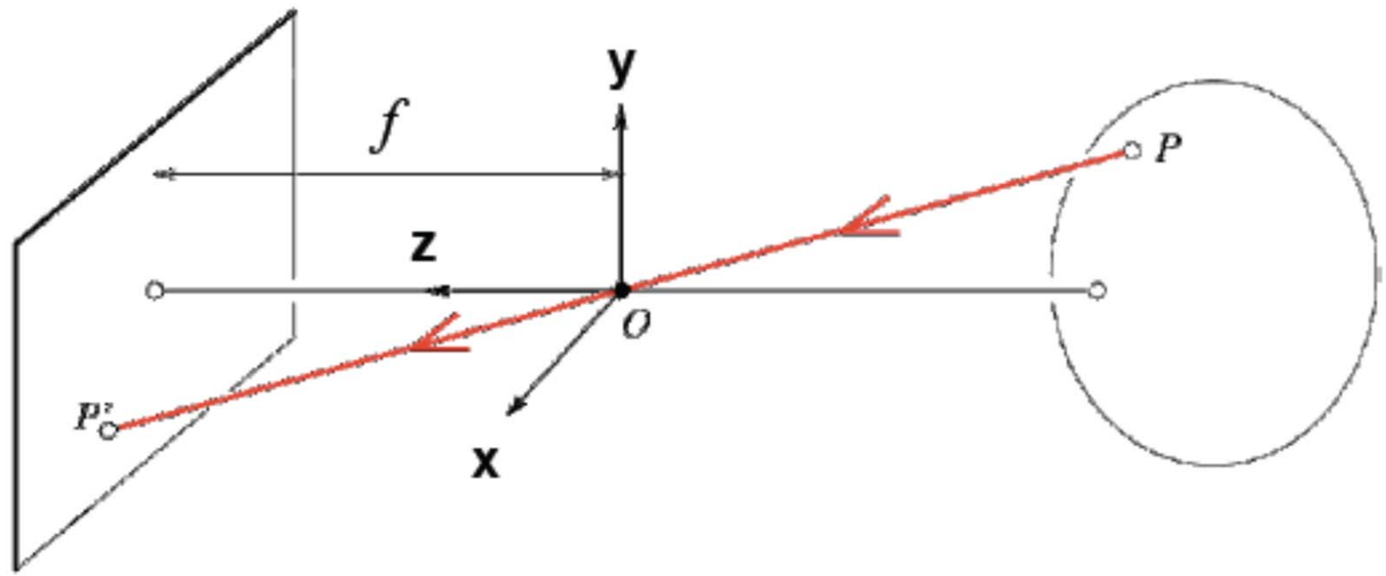
- Epipolar constraint
- **2D transformations**
    - Similarity transformation
    - Affine transformation
    - Projective transformation



N-views

Are images consistent with a 3D model?

Modeling projection



Projection equation: $(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z})$

Homogeneous coordinates

$$(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z})$$

Is this a linear transformation?

- no—division by **z** is nonlinear

Trick: add one more coordinate:

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

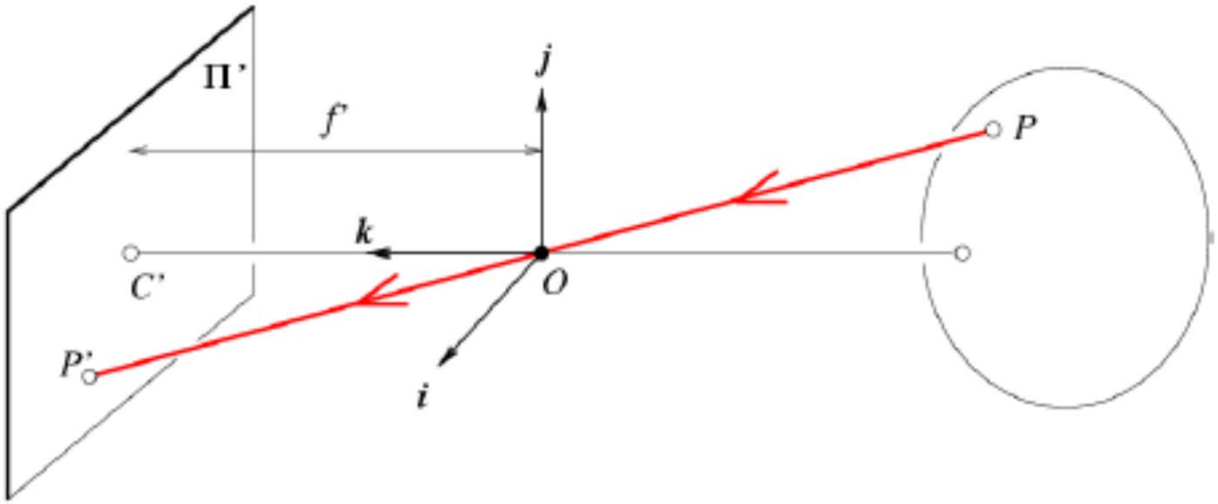
Converting *from* homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Slide by Steve Seitz

Projection matrix



Intrinsic Assumptions

- Unit aspect ratio
- Optical center at (0,0)
- No skew

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}\boxed{\begin{matrix}f & 0 & 0\\0 & f & 0\\0 & 0 & 1\end{matrix}}_{\mathbf{K}} & \begin{matrix}0\\0\\0\end{matrix}\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

Slide Credit: Saverese

Remove assumption: known optical center

Intrinsic Assumptions

- Unit aspect ratio
- No skew

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix}u \\ v \\ 1\end{bmatrix} = \begin{bmatrix}f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}x \\ y \\ z \\ 1\end{bmatrix}$$

Remove assumption: square pixels

Intrinsic Assumptions
- No skew

Extrinsic Assumptions
- No rotation
- Camera at (0,0,0)

$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}\alpha & 0 & u_0 & 0\\0 & \beta & v_0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

Remove assumption: non-skewed pixels

Intrinsic Assumptions

Extrinsic Assumptions
- No rotation
- Camera at (0,0,0)

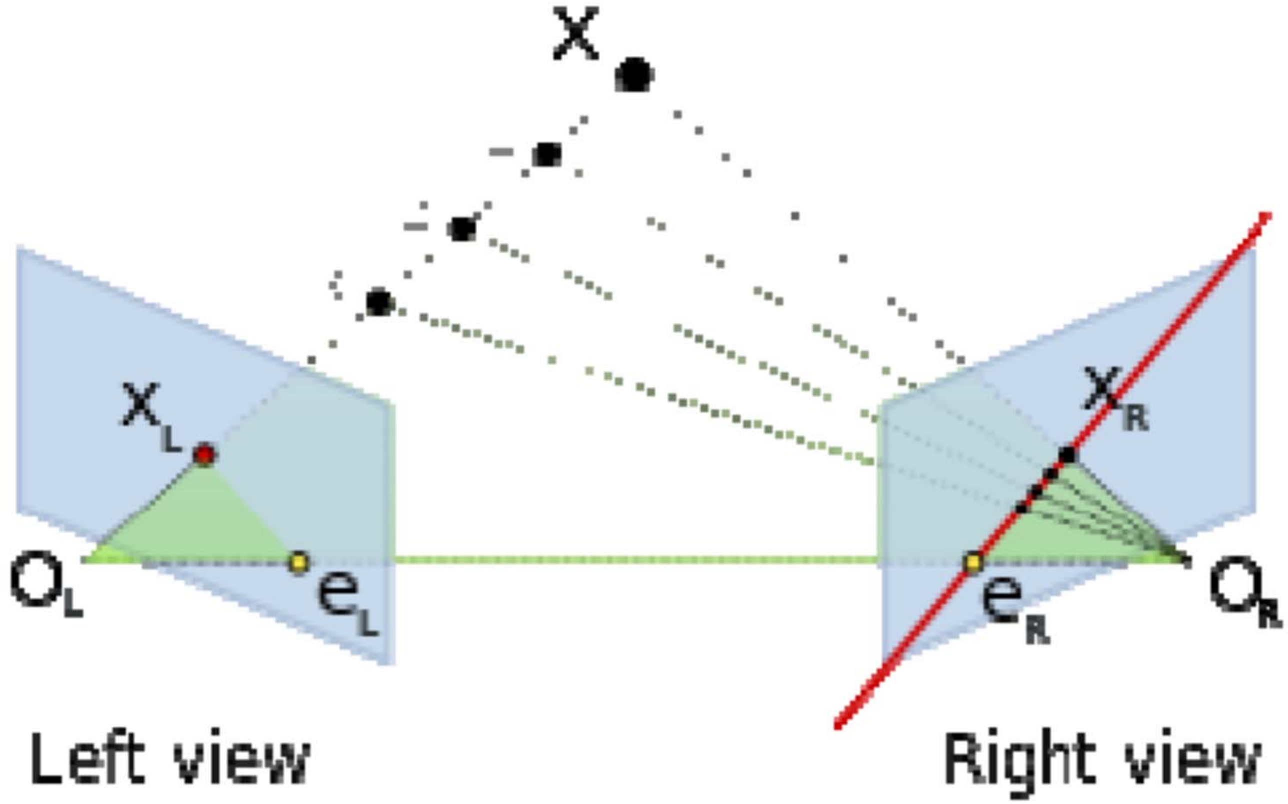
$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}\alpha & s & u_0 & 0\\0 & \beta & v_0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

Degrees of freedom

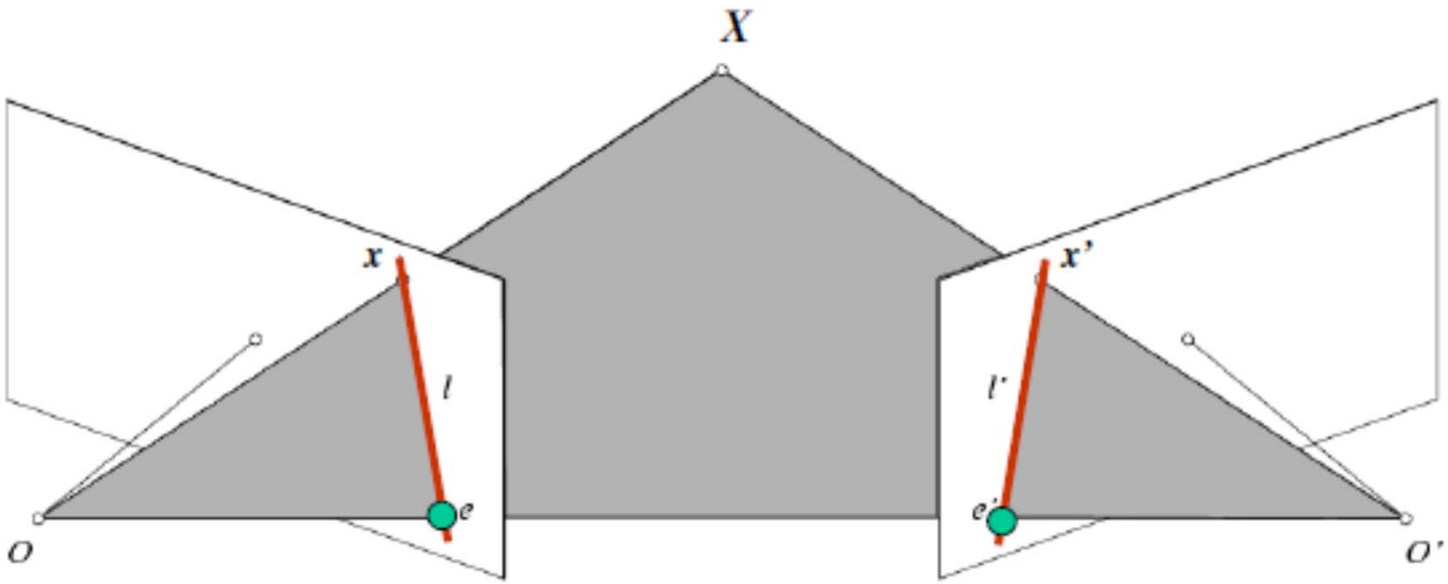
$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \overset{5}{\begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}} \overset{6}{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Epipolar geometry



Epipolar constraint: Uncalibrated case

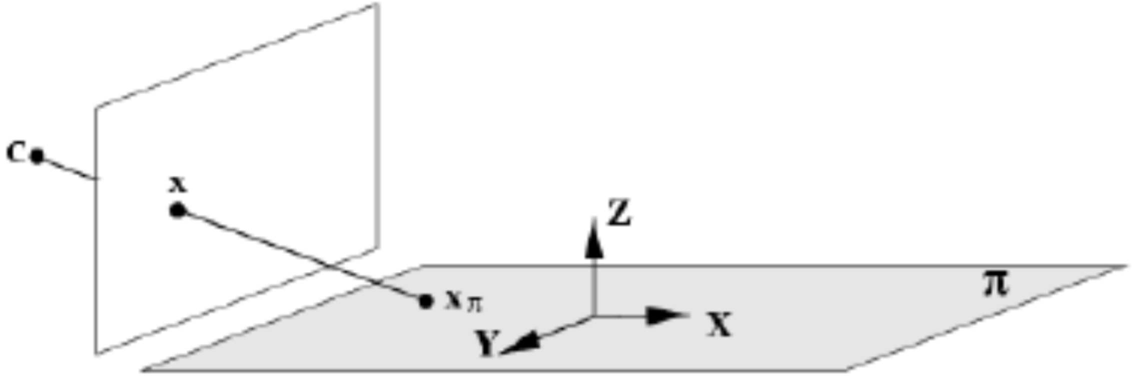


$$x'^T F x = 0$$

Fundamental matrix

- $\boldsymbol{F x}$ is the epipolar line associated with $\boldsymbol{x}$ ($\boldsymbol{l'} = \boldsymbol{F x}$)
- $\boldsymbol{F^T x'}$ is the epipolar line associated with $\boldsymbol{x'}$ ($\boldsymbol{l'} = \boldsymbol{F^T x'}$)
- $\boldsymbol{F e} = 0$ and $\boldsymbol{F^T e'} = 0$
- $\boldsymbol{F}$ is singular (rank two)
- $\boldsymbol{F}$ has *seven* degrees of freedom

Plane projective transformations



Choose the world coordinate system such that the plane of the points has zero z coordinate. Then the 3×4 matrix P reduces to

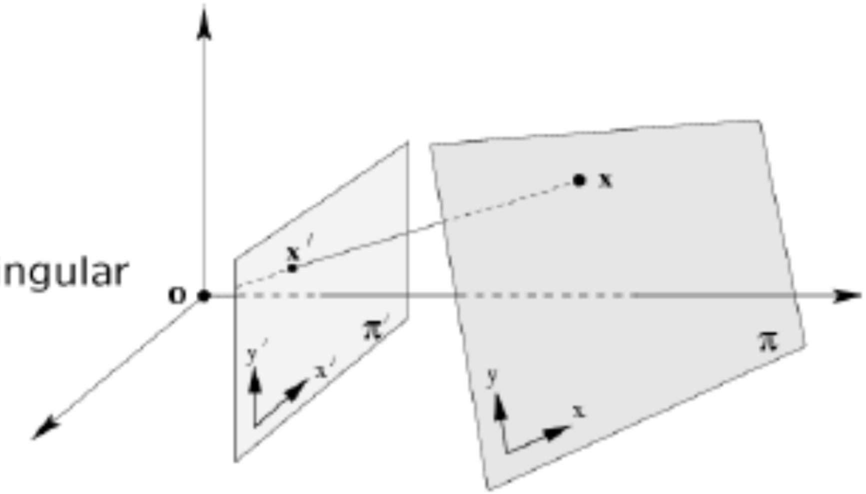
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{pmatrix} \mathrm{x} \\ \mathrm{y} \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{pmatrix} \mathrm{x} \\ \mathrm{y} \\ 1 \end{pmatrix}$$

which is a 3×3 matrix representing a general plane to plane projective transformation.

Projective transformations continued

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

or $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where $\mathtt{H}$ is a 3×3 non-singular homogeneous matrix.



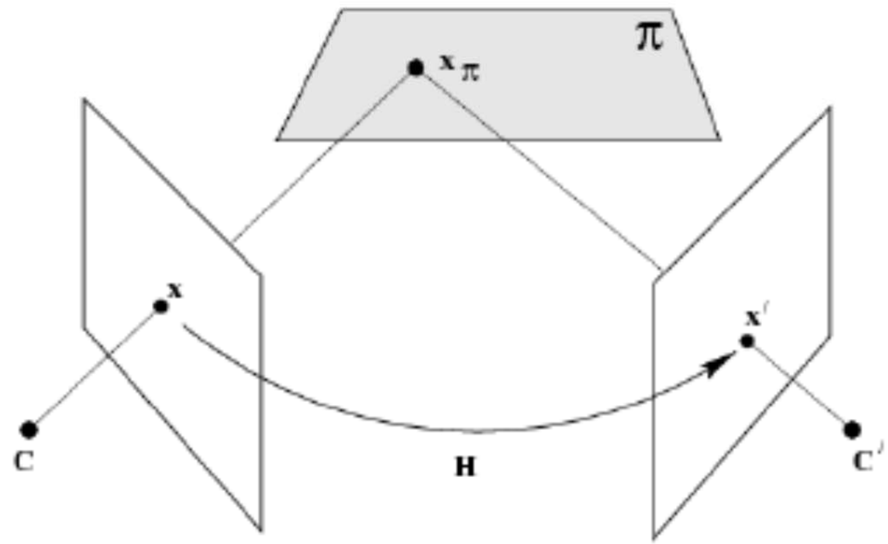
- This is the most general transformation between the world and image plane under imaging by a perspective camera.
- It is often only the 3 x 3 form of the matrix that is important in establishing properties of this transformation.
- A projective transformation is also called a ``homography'' and a ``collineation''.
- H has 8 degrees of freedom.

Images of Planes

Projective transformations between images induced by a plane

$$\mathbf{x} = \mathtt{H}_{1\pi}\mathbf{x}_\pi \qquad \mathbf{x}' = \mathtt{H}_{2\pi}\mathbf{x}_\pi$$

$$\begin{aligned} \mathbf{x}' &= \mathtt{H}_{2\pi}\mathbf{x}_\pi \\ &= \mathtt{H}_{2\pi}\mathtt{H}_{1\pi}^{-1}\mathbf{x} = \mathtt{H}\mathbf{x} \end{aligned}$$



- H can be computed from the correspondence of four points on the plane

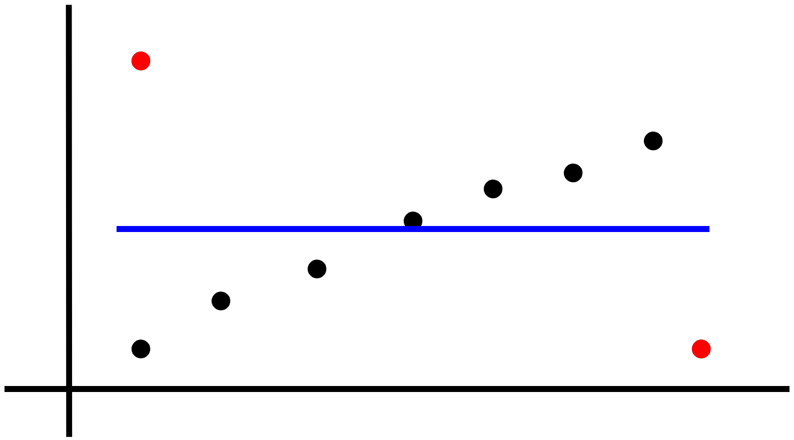
Matching of descriptors

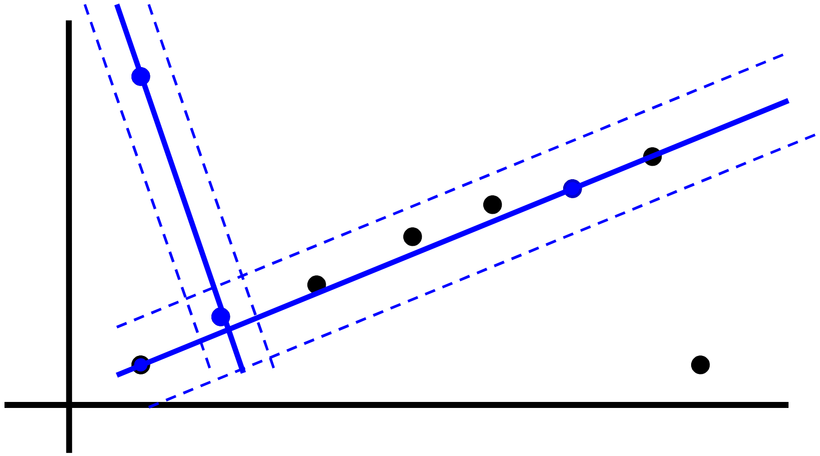
- Geometric verification with global constraint
  - All matches must be consistent with a global geometric transformation
  - However, there are many incorrect matches
  - Need to estimate simultaneously the geometric transformation and the set of consistent matches

- Robust estimation of global constraints
  - RANSAC (RANdom Sampling Consensus) [Fishler&Bolles'81]
  - Hough transform [Lowe'04]

RANSAC: Example of robust line estimation

Fit a line to 2D data containing outliers





There are two problems

1. a line fit which minimizes perpendicular distance
2. a classification into inliers (valid points) and outliers

Solution: use robust statistical estimation algorithm RANSAC

(RANdom Sample Consensus) [Fishler & Bolles, 1981]

Slide credit: A. Zisserman

RANSAC robust line estimation

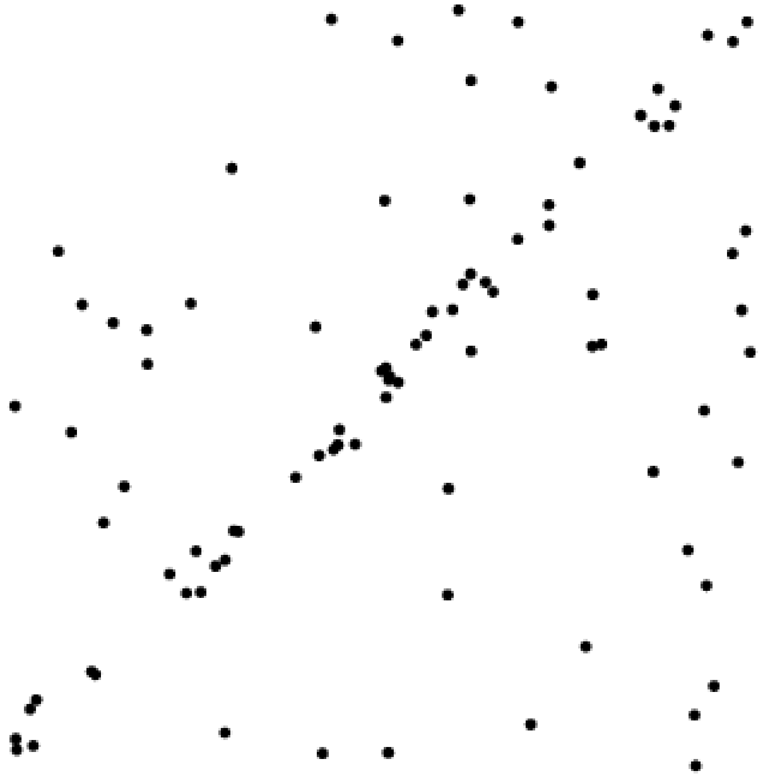
Repeat

1. Select random sample of 2 points
2. Compute the line through these points
3. Measure support (number of points within threshold distance of the line)

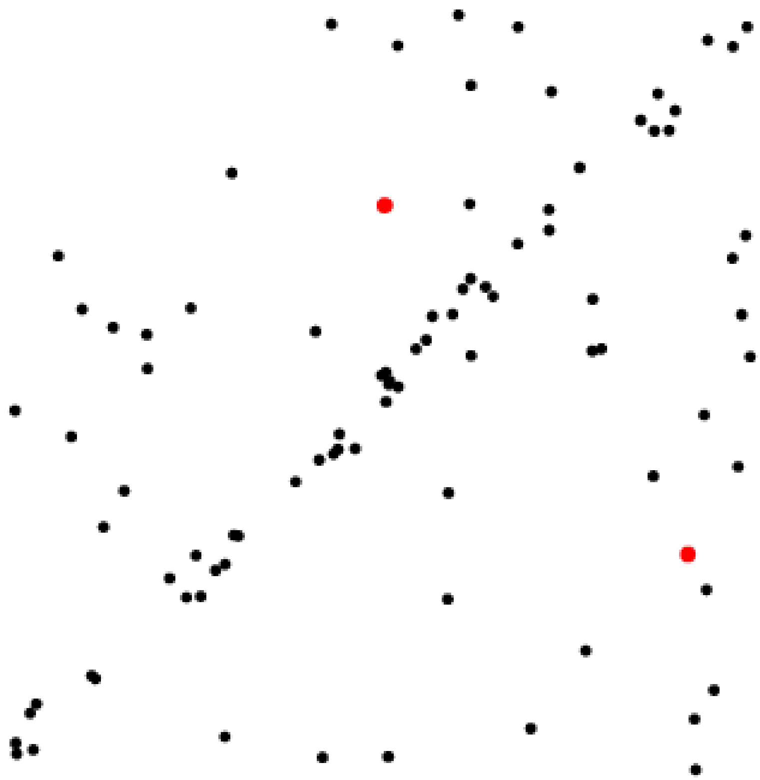
Choose the line with the largest number of inliers

- Compute least squares fit of line to inliers (regression)

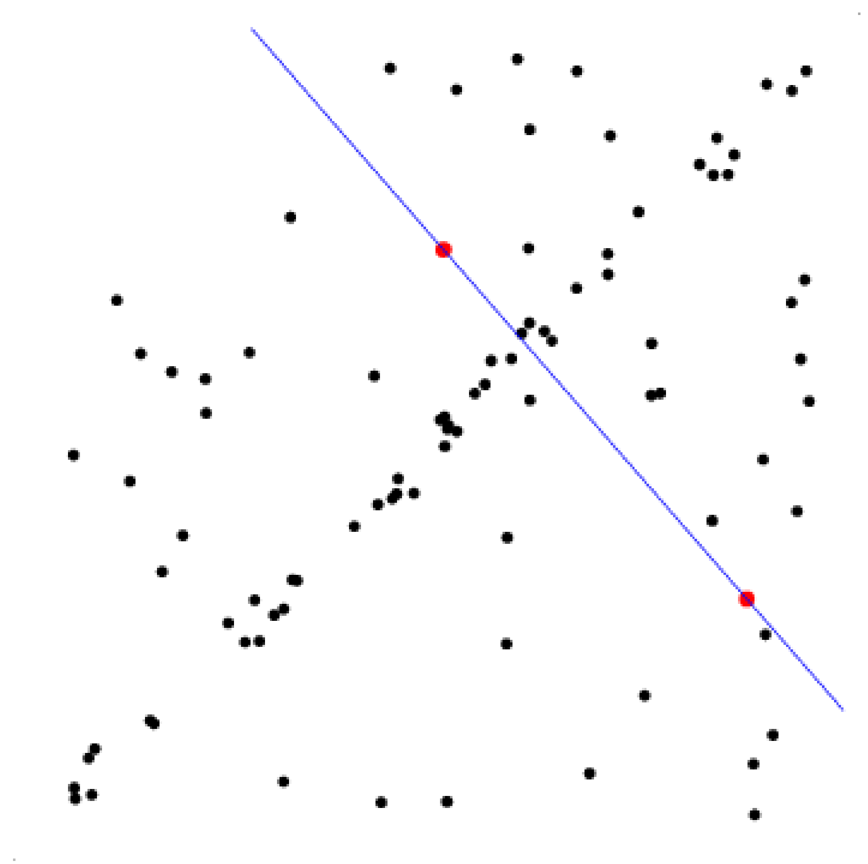
Slide credit: A. Zisserman



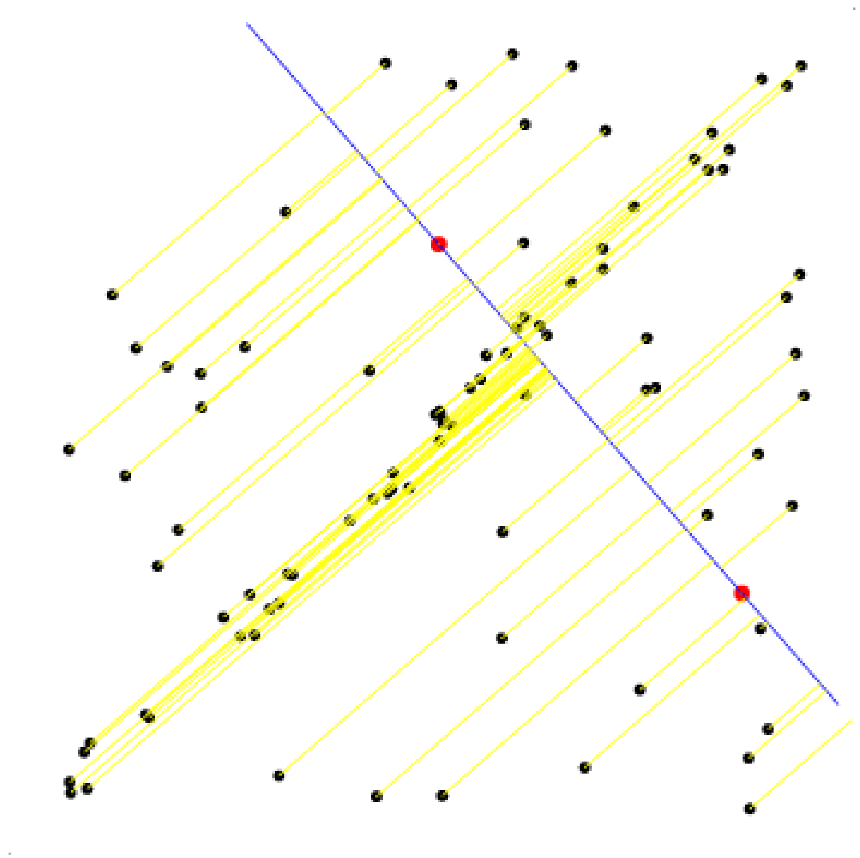
Slide credit: O. Chum



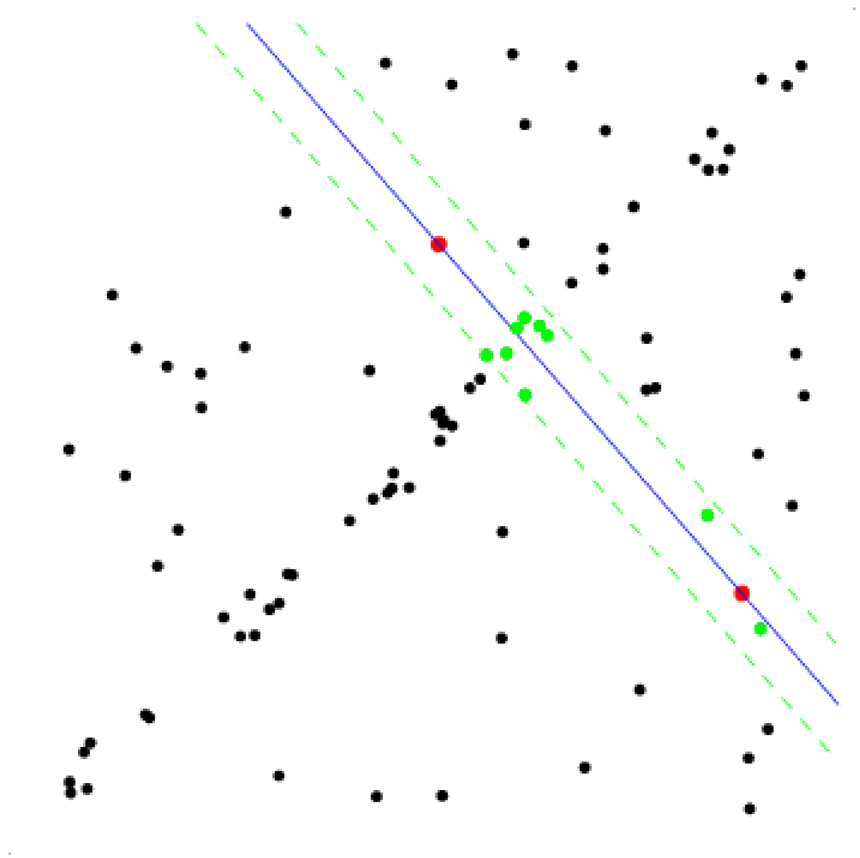
Slide credit: O. Chum



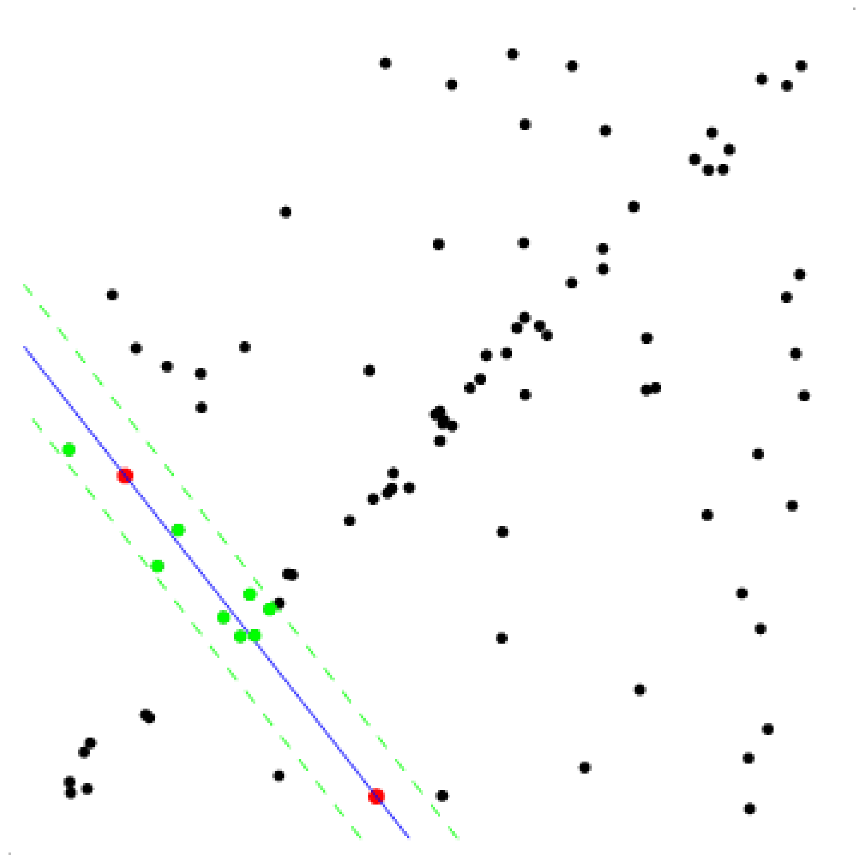
Slide credit: O. Chum



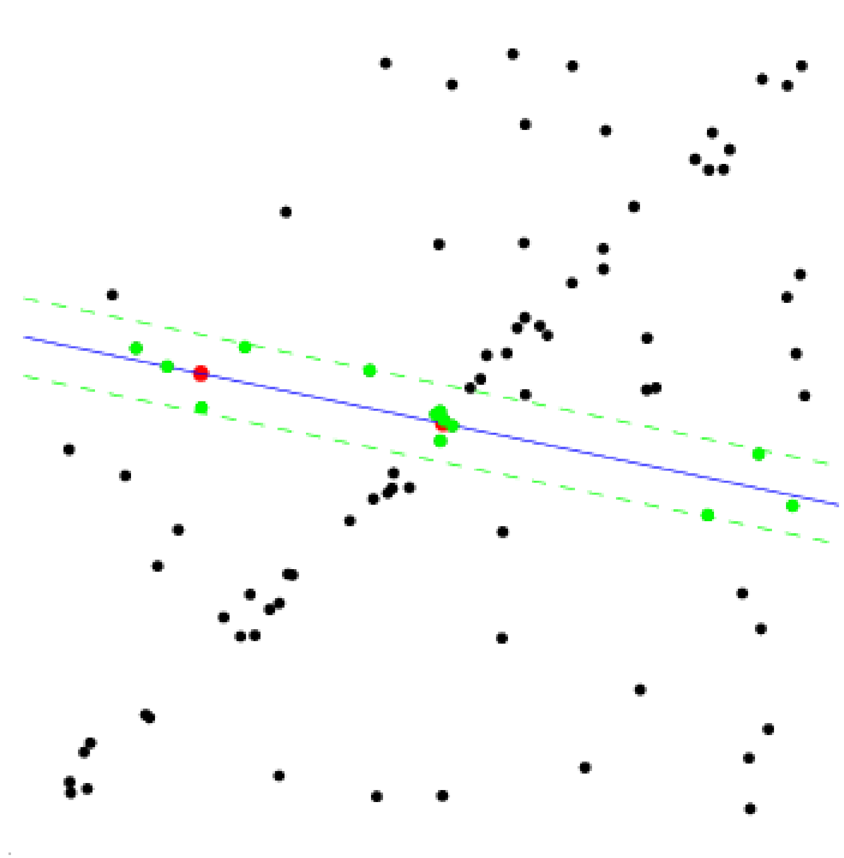
Slide credit: O. Chum



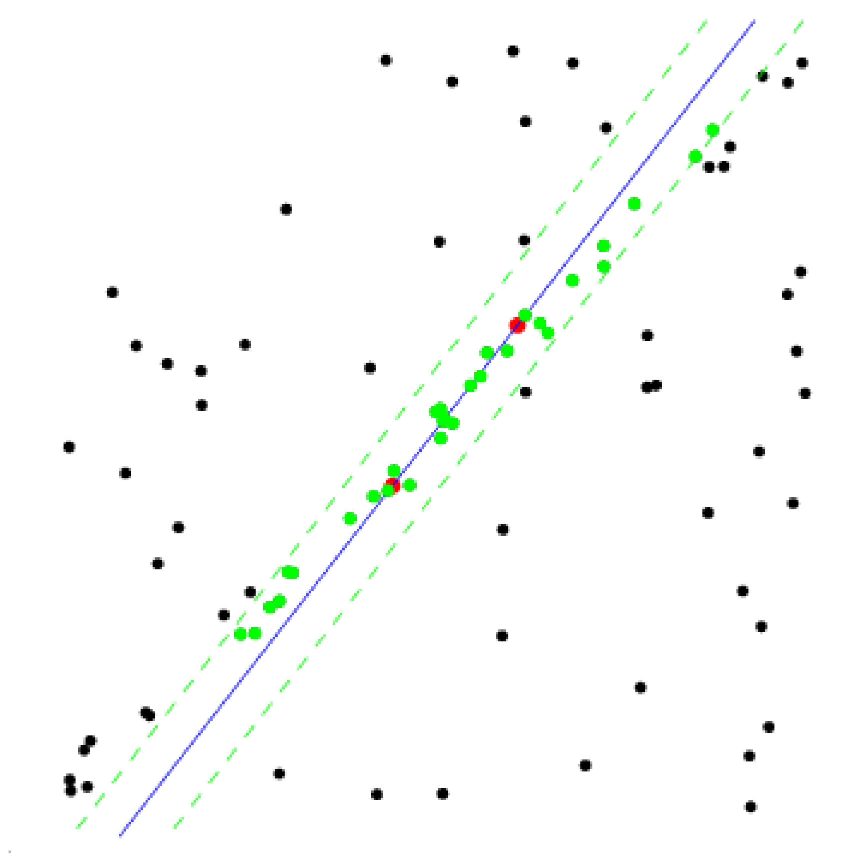
Slide credit: O. Chum



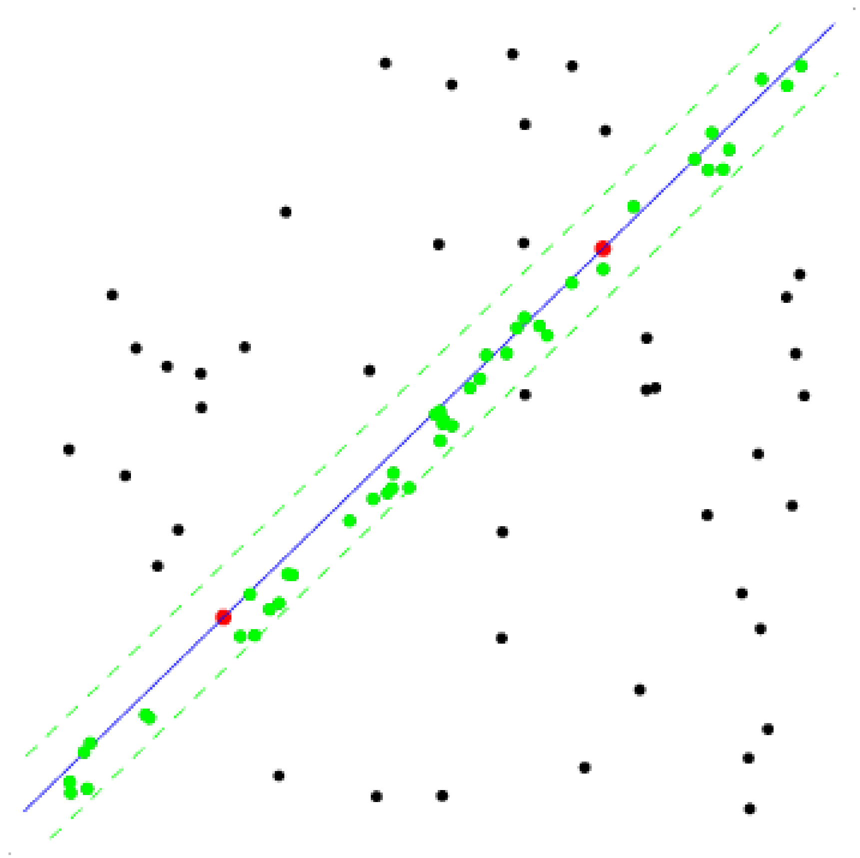
Slide credit: O. Chum



Slide credit: O. Chum



Slide credit: O. Chum

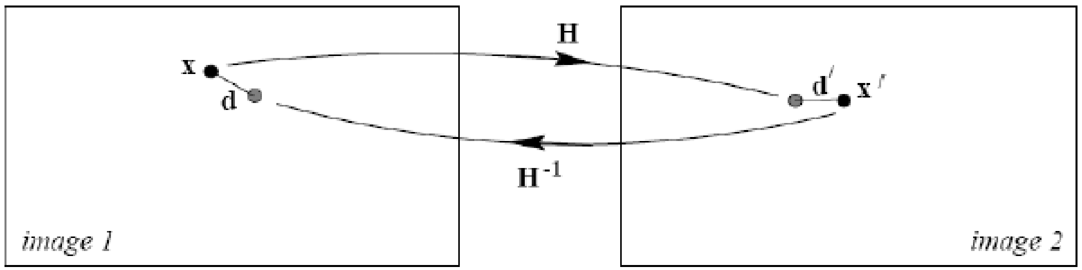


Slide credit: O. Chum

Algorithm RANSAC

- Robust estimation with RANSAC of a homography
  - Repeat
    - Select 4 point matches
    - Compute 3x3 homography
    - Measure support (number of inliers within threshold, i.e. $d^2_{transfer} < t$)

$$d^2_{\mathrm{transfer}} = d(\mathbf{x}, \mathtt{H}^{-1}\mathbf{x}')^2 + d(\mathbf{x}', \mathtt{H}\mathbf{x})^2$$



  - Choose (H with the largest number of inliers)
  - Re-estimate H with all inliers

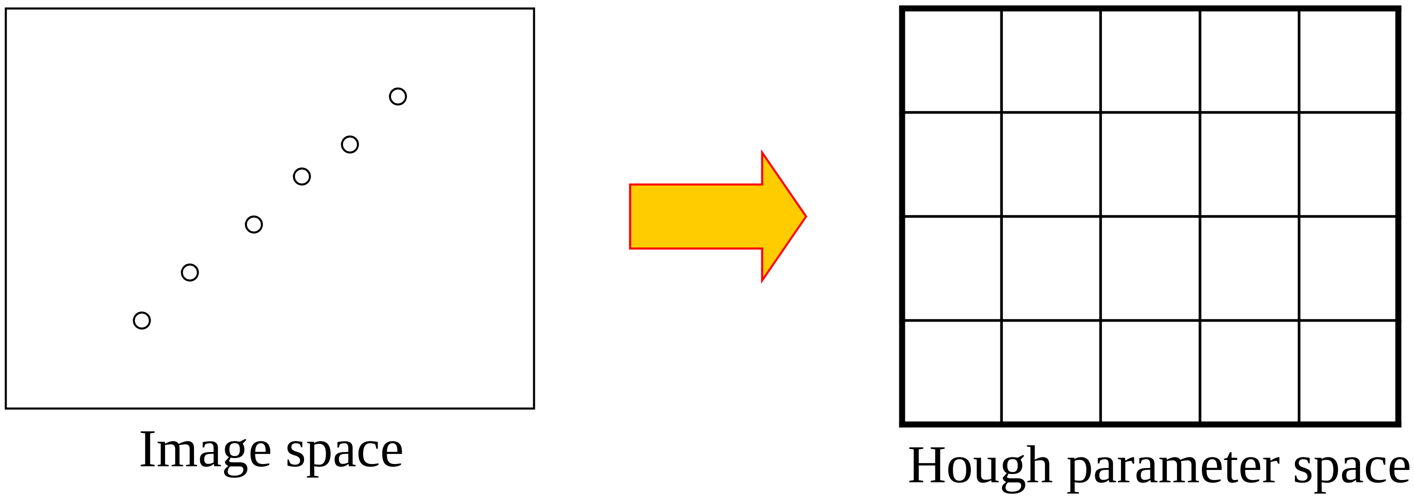
Matching of descriptors

- Geometric verification with global constraint
  - All matches must be consistent with a global geometric transformation
  - However, there are many incorrect matches
  - Need to estimate simultaneously the geometric transformation and the set of consistent matches

- Robust estimation of global constraint
  - RANSAC (RANdom Sampling Consensus) [Fishler&Bolles'81]
  - **Hough transform [Lowe'04]**

Strategy 2: Hough transform

- General outline:
  - Discretize parameter space into bins
  - For each feature point in the image, put a vote in every bin in the parameter space that could have generated this point
  - Find bins that have the most votes



P.V.C. Hough, *Machine Analysis of Bubble Chamber Pictures*, Proc. Int. Conf. High Energy Accelerators and Instrumentation, 1959

Hough transform for lines

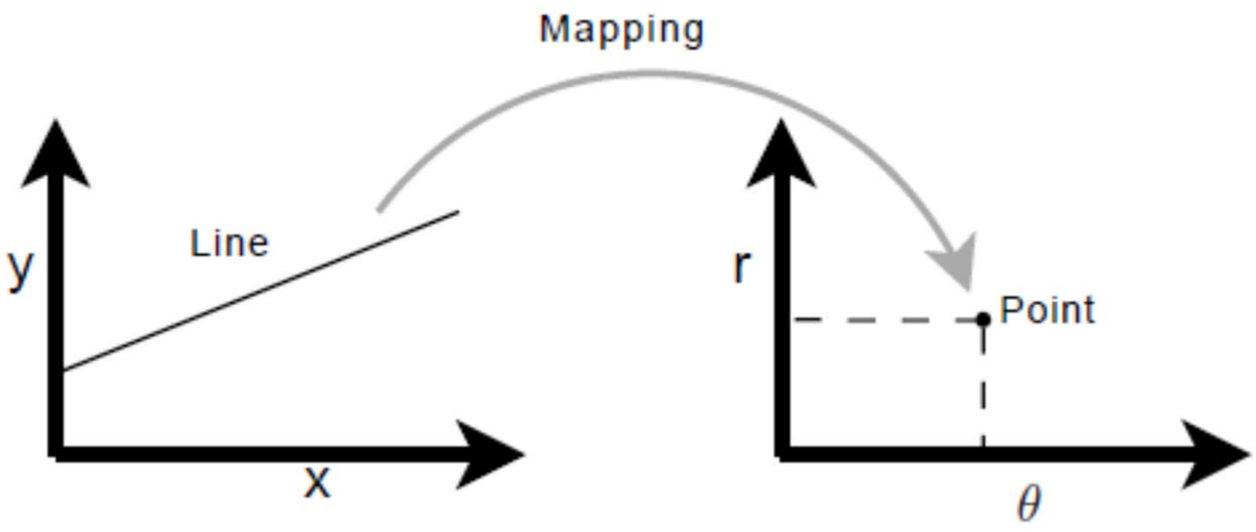
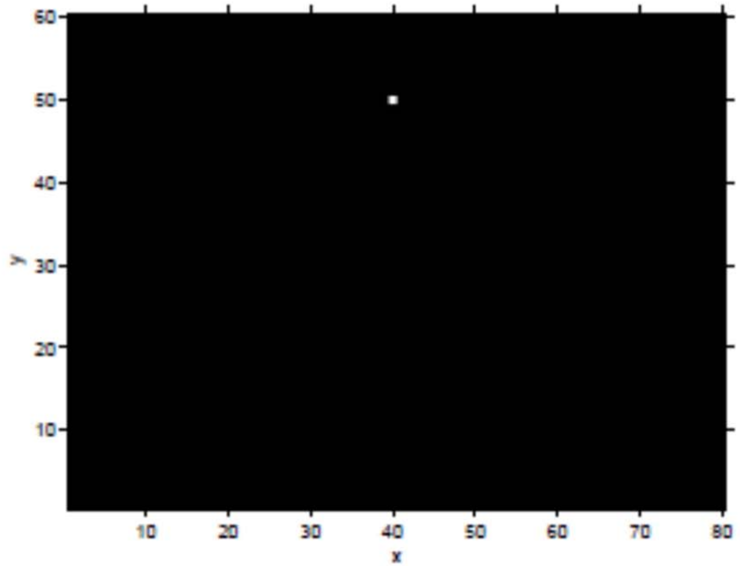
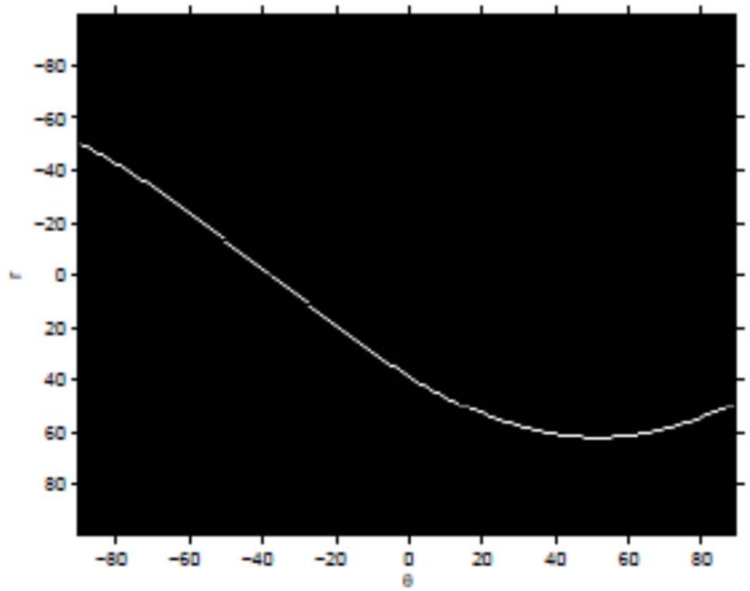


Figure 1: *Mapping of one unique line to the Hough space.*

Hough transform for lines

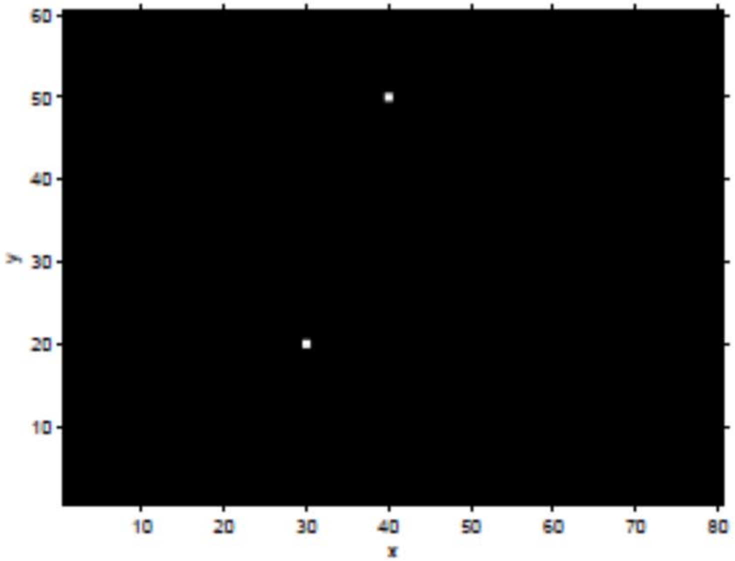


(a) Point p_0.

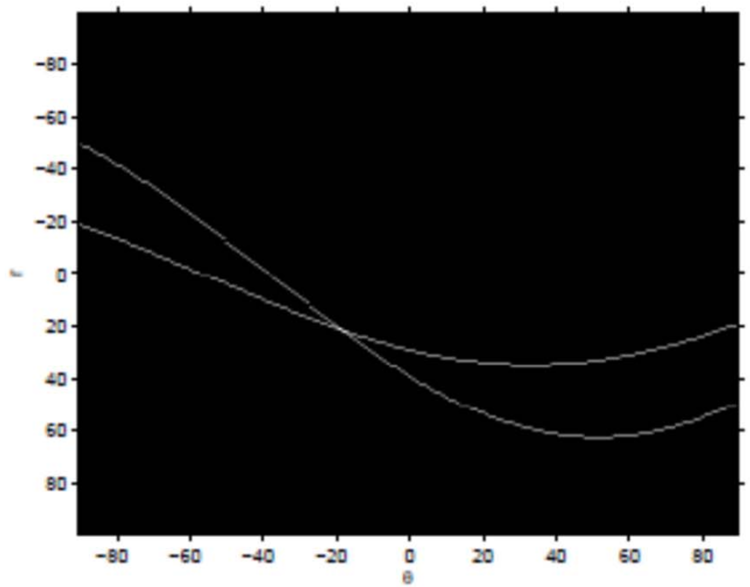


(b) All possible lines through p_0 represented in the Hough space.

Hough transform for lines



(a) Points p_0 and p_1.



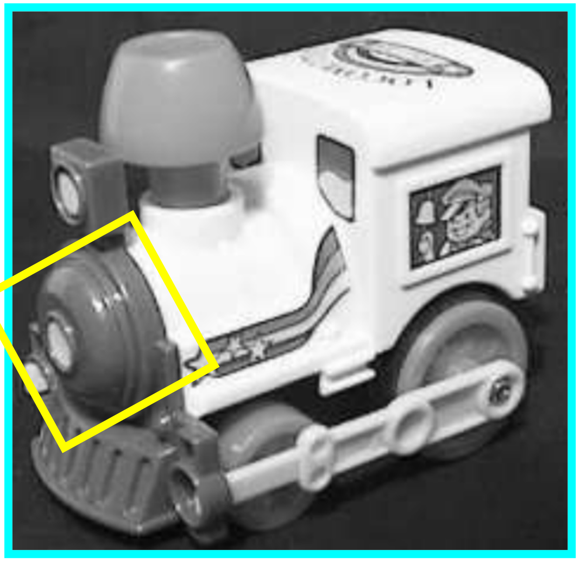
(b) All possible lines through p_0 and/or p_1 represented in the Hough space.

Hough transform for object recognition

Suppose our features are scale- and rotation-covariant

- Then a single feature match provides an alignment hypothesis (translation, scale, orientation)

model



Target image

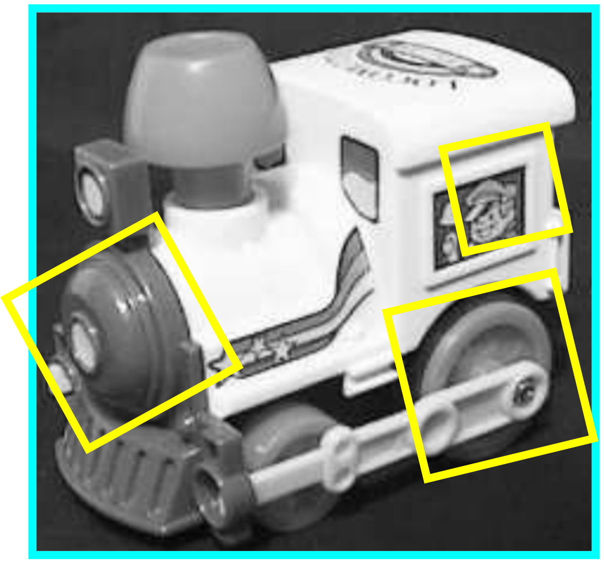
David G. Lowe. **"Distinctive image features from scale-invariant keypoints",** *IJCV* 60 (2), pp. 91-110, 2004.

Hough transform for object recognition

Suppose our features are scale- and rotation-covariant

- Then a single feature match provides an alignment hypothesis (translation, scale, orientation)
- Of course, a hypothesis obtained from a single match is unreliable
- Solution: Coarsely quantize the transformation space. Let each match vote for its hypothesis in the quantized space.

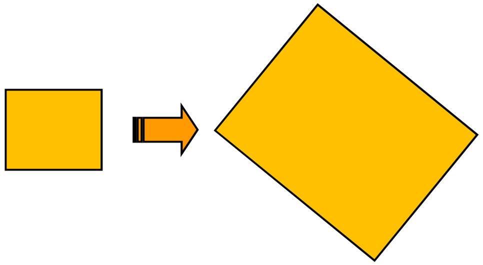
model



David G. Lowe. **“Distinctive image features from scale-invariant keypoints”,** *IJCV* 60 (2), pp. 91-110, 2004.

Similarity transformation is specified by four parameters: scale factor s, rotation θ, and translations t_x and t_y.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = sR(\theta) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$



Recall, each SIFT detection has: position (x_i, y_i), scale s_i, and orientation θ_i.

How many correspondences are needed to compute similarity transformation?

Compute similarity transformation from a single correspondence:

$$(x_A, y_A, s_A, \theta_A) \leftrightarrow (x'_A, y'_A, s'_A, \theta'_A)$$

$$\theta = \theta'_A - \theta_A$$

$$s = s'_A / s_A$$

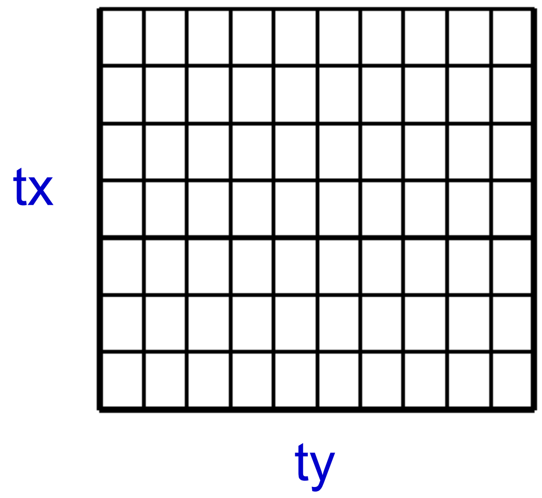
$$t_x = x'_A - sR(\theta)x_A$$

$$t_y = y'_A - sR(\theta)y_A$$

Basic algorithm outline

1. Initialize accumulator H
   to all zeros
2. For each tentative match
   compute transformation
   hypothesis: tx, ty, s, θ
   H(tx,ty,s,θ) = H(tx,ty,s,θ) + 1
   end
   end
3. Find all bins (tx,ty,s,θ) where H(tx,ty,s,θ) has at least three votes

H: 4D-accumulator array
(only 2-d shown here)



- Correct matches will consistently vote for the same transformation while mismatches will spread votes.
- Cost: Linear scan through the matches (step 2), followed by a linear scan through the accumulator (step 3).

Comparison

Hough Transform

- Advantages
  - Can handle high percentage of outliers (>95%)
  - Extracts groupings from clutter in linear time
- Disadvantages
  - Quantization issues
  - Only practical for small number of dimensions (up to 4)
- Improvements available
  - Probabilistic Extensions
  - Continuous Voting Space
  - Can be generalized to arbitrary shapes and objects

RANSAC

- Advantages
  - General method suited to large range of problems
  - Easy to implement
  - “Independent” of number of dimensions
- Disadvantages
  - Basic version only handles moderate number of outliers (<50%)
- Many variants available, e.g.
  - PROSAC: Progressive RANSAC [Chum05]
  - Preemptive RANSAC [Nister05]

Summary

Finding correspondences in images is useful for

- Image matching, panorama stitching
- Object recognition
- Large scale image search: next part of the lecture

Beyond local point matching

- Semi-local relations
- Global geometric relations:
  - Epipolar constraint
  - 3D constraint (when 3D model is available)
  - 2D tnfs: Similarity / Affine / Homography
- Algorithms:
  - RANSAC
  - Hough transform

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

$$\boxed{\mathrm{x} = \mathrm{PX}}$$

$$\boxed{\mathrm{x}' = \mathrm{Hx}}$$