Interest Points

Gert Kootstra

Credits of some of the slides: Bahadir K. Gunturk and Fei-Fei Li

Overview

- Local Features for Image Representations
- Interest-Point Detection
  - Harris corners
  - Difference-of-Gaussians (SIFT)
- Interest-Point Description
  - Histogram-of-Gradients (SIFT)
- Bag of Words

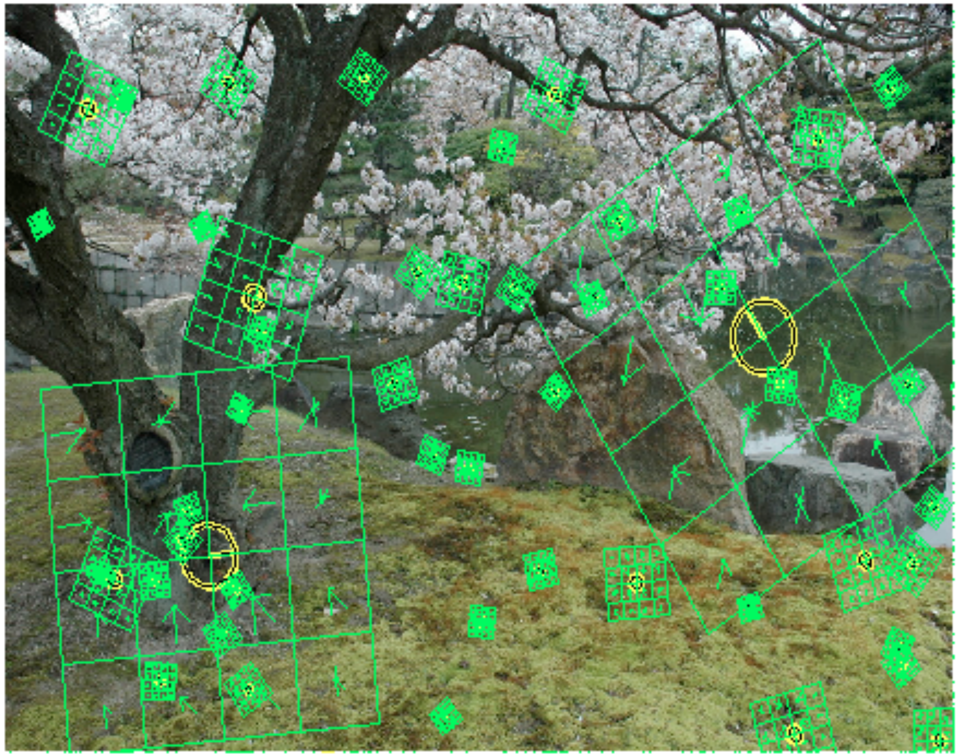
Image Representation: Global

- Global feature representation
  - Color histograms, Principle Component Analysis,…

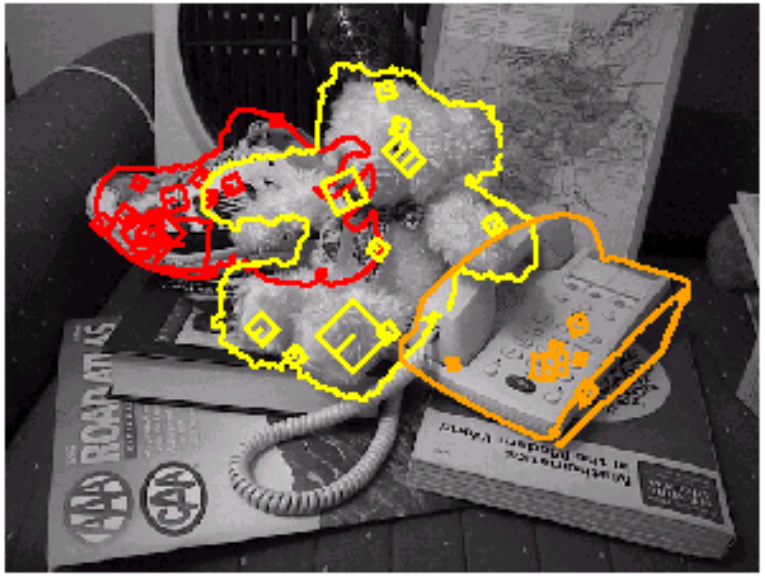
- Disadvantages
  - Cannot deal with occlusions, clutter, viewpoint changes.

Image Representation: Local

- Representation by a set of local features
  - Image points that differ from their surrounding
    - Well-localized points
  - The neighborhoods represent the image
    - Individually identifiable



Advantages of Local Features



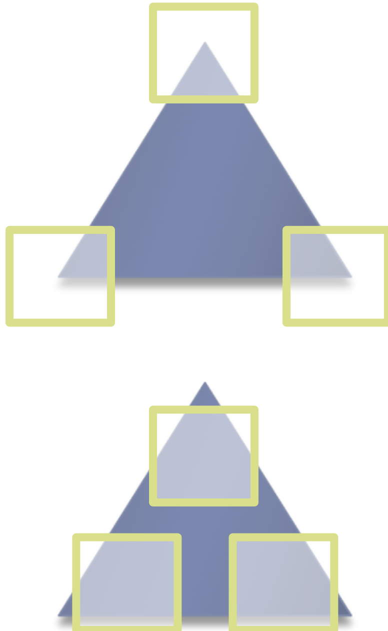
- Can deal with occlusions
- Can deal with clutter
- More invariant to image transformations
- More robust to noise
- Object recognition without segmentation
- Sparse representation of the image

A Good Local Feature

- Accurate and repeatable localization of the feature points
- Invariance to translation, rotation, scale, viewpoint
- Robustness to noise, lighting conditions, compression, blur.
- Distinctiveness of descriptor
- Efficiency

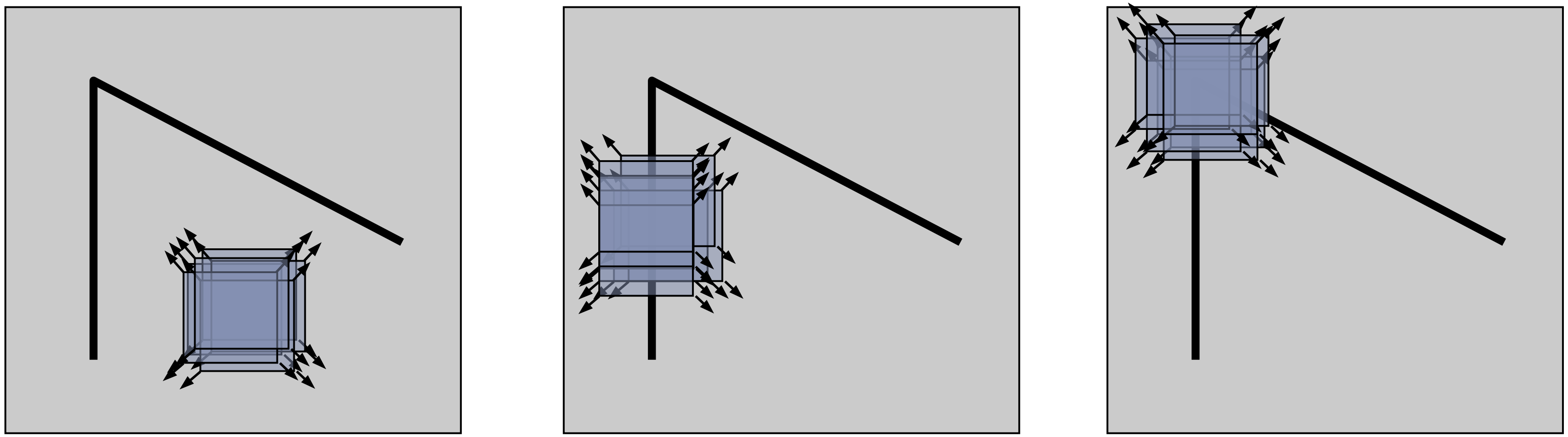
Interest points

- We focus on interest points as local features
- Interest-point detector
  - Points on corners
    - Harris corners (first-order derivative)
  - Points on blob-like structures
    - SIFT (second-order derivative)
- Interest-point descriptor
  - Local description of the neighborhood
  - Histogram of Oriented Gradients



Harris-Corner Detector

- Intuition
  - Find points that different from their neighborhood



“flat” region:
no change in all
directions

“edge”:
no change along the
edge direction

“corner”:
significant change in
all directions

The second-moment matrix

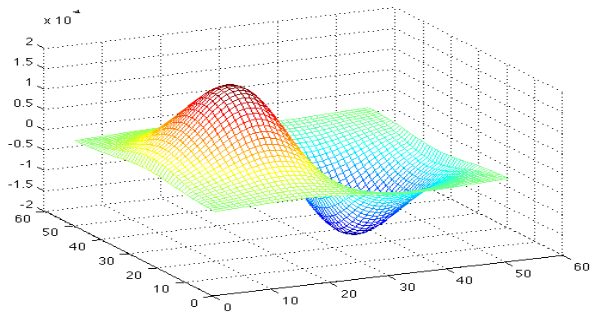
- The second-moment matrix

Smoothing — First-order derivatives

$$M = \sigma_D^2 g(\sigma_I) * \begin{bmatrix} I_x^2(\mathbf{x},\sigma_D) & I_x(\mathbf{x},\sigma_D)I_y(\mathbf{x},\sigma_D) \\ I_x(\mathbf{x},\sigma_D)I_y(\mathbf{x},\sigma_D) & I_y^2(\mathbf{x},\sigma_D) \end{bmatrix}$$

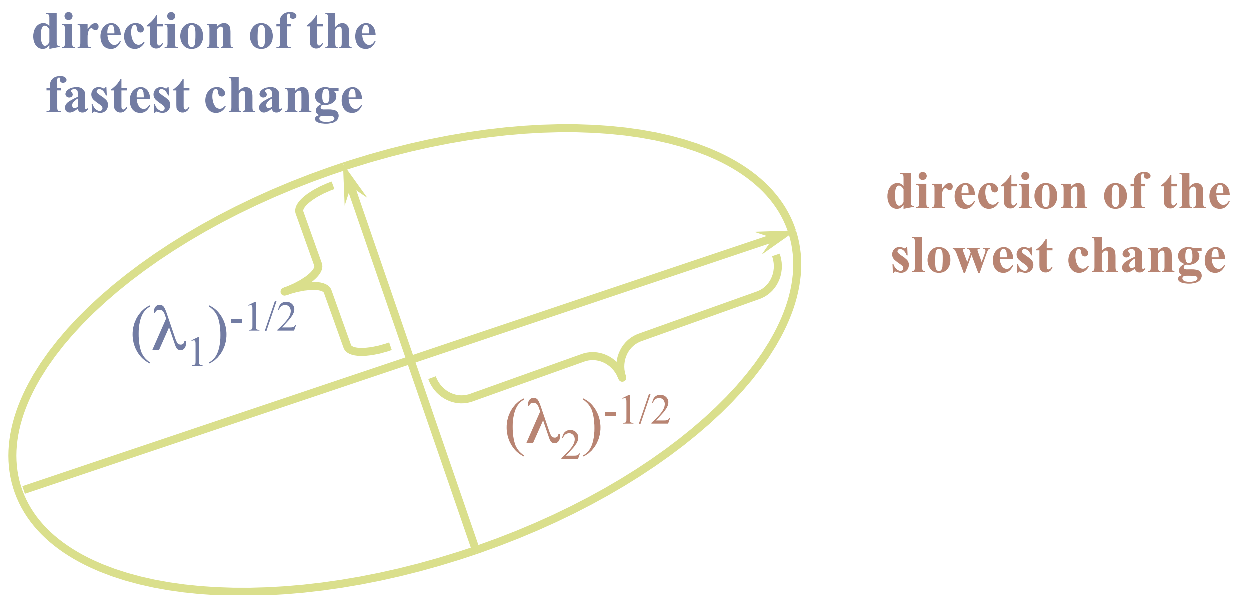
Derivatives computed with Gaussian kernels of scale σ_D.

$$I_x(\mathbf{x},\sigma_D) = \frac{\partial}{\partial x} d(\sigma_D) * I(x)$$



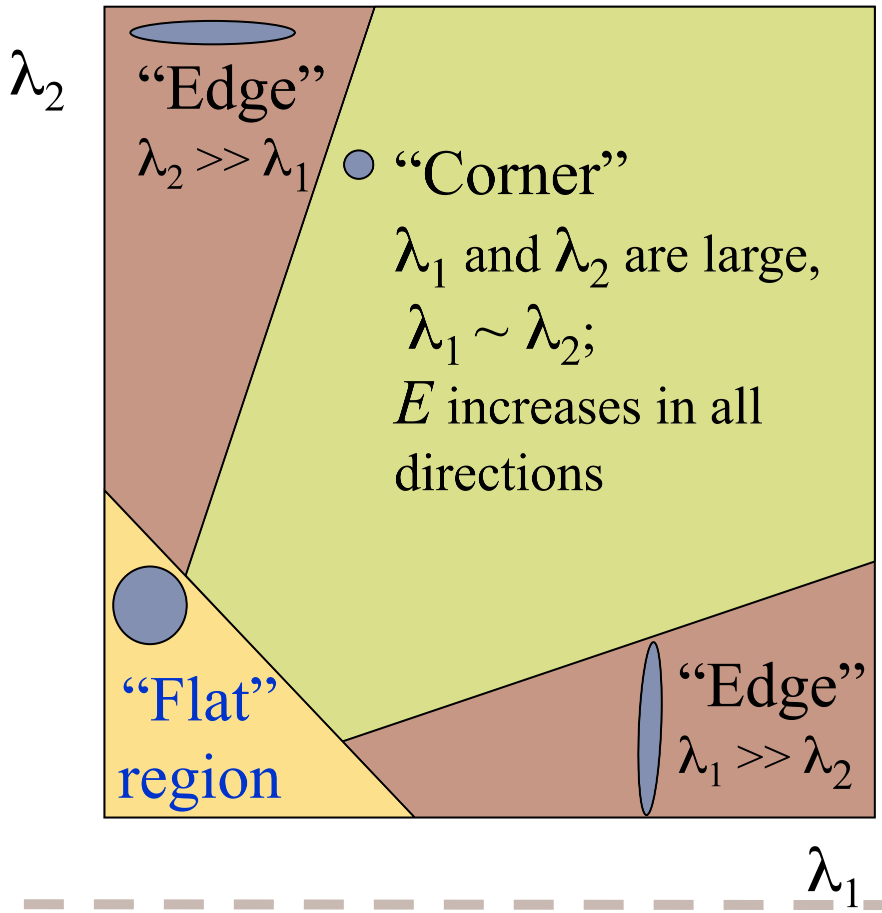
Eigenvalues

- The eigenvalues λ_1 and λ_2 of M represent the principal signal changes at **x**.



Eigenvalues

- Classification of image points



Determinant and Trace

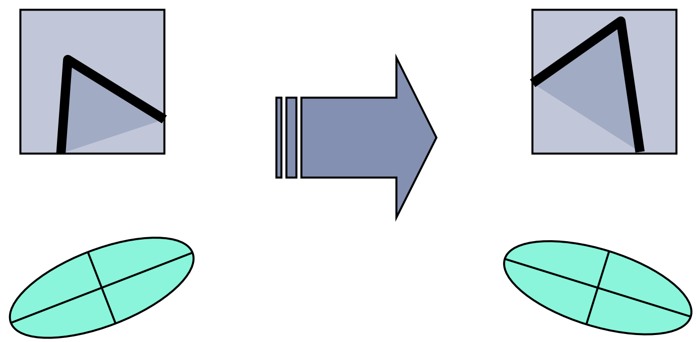
- No need to explicitly calculate the eigenvalues
  - Determinant of M is the product of λ_1 and λ_2
  - Trace of M is the sum of λ_1 and λ_2
- Harris cornerness:
  - Det(M) = ad – bc
  - Trace(M) = a + d
  - R = det(M) – κ*trace^2(M)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

- Finding local maxima in the image

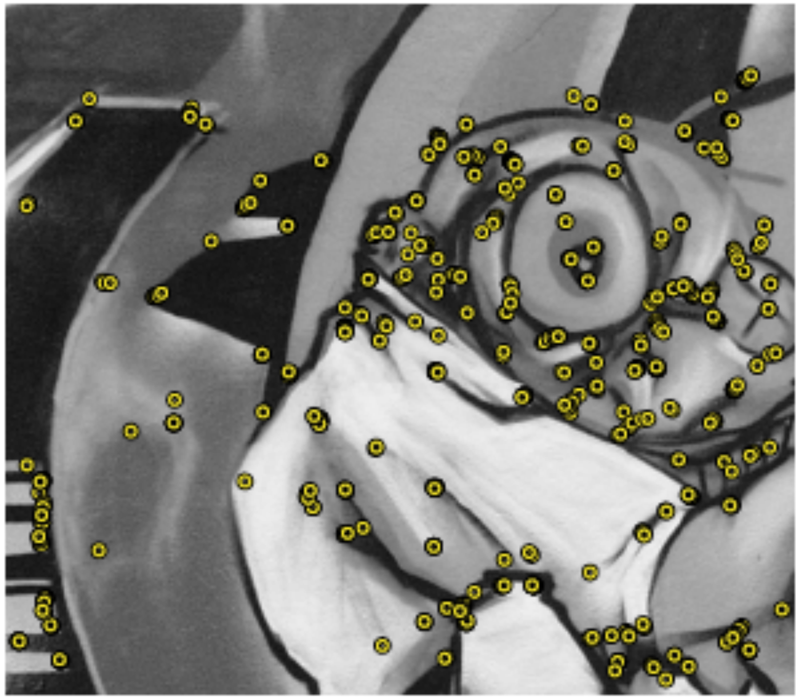
Rotational Invariance

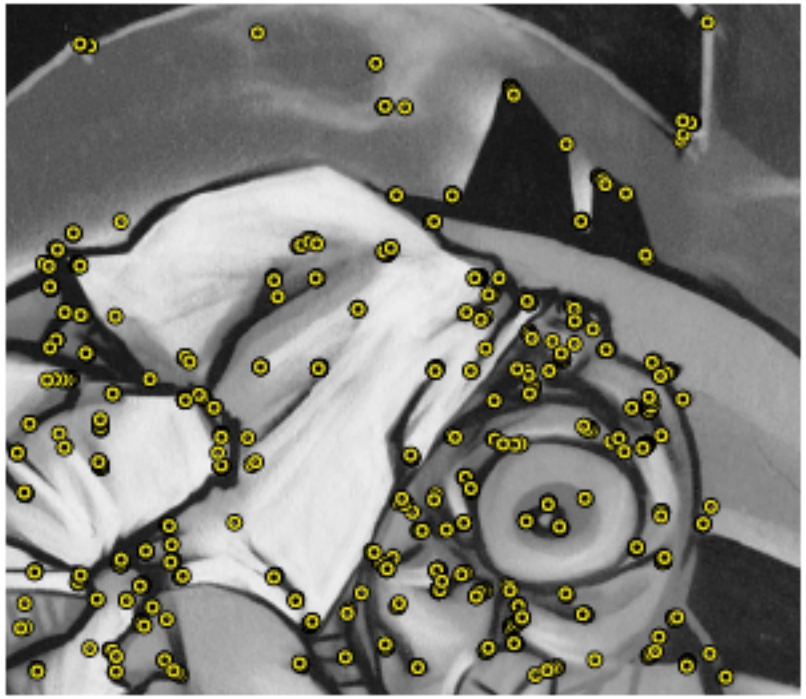
- Harris detector is rotational invariance
- Ellipse (defined by eigenvectors of M) rotates with the image, so cornerness value remains the same



Example

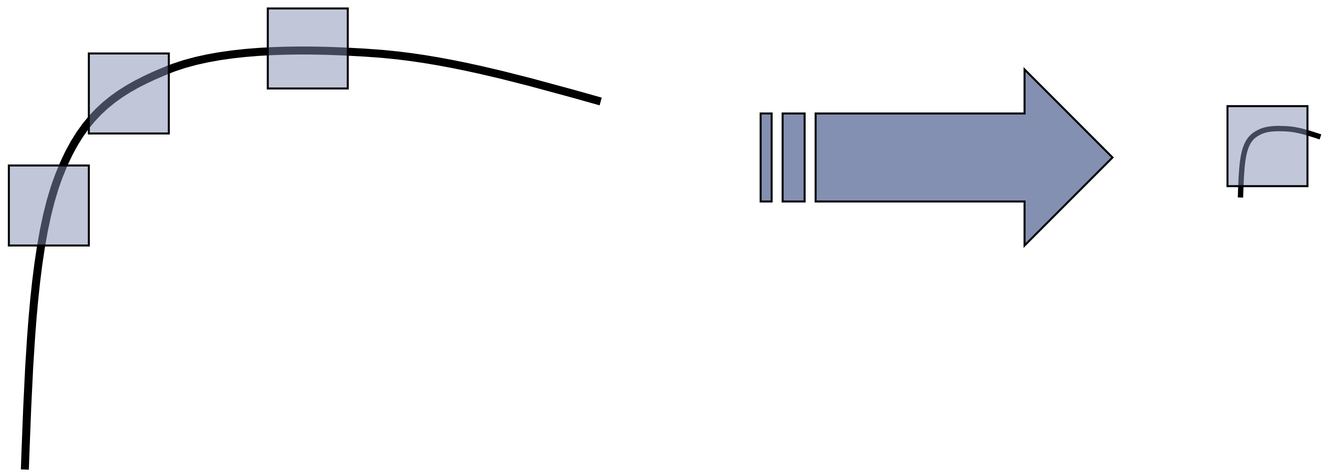
- Harris corners on rotated image





Scale Invariance

- The basic Harris detector is not invariant to changes in scale

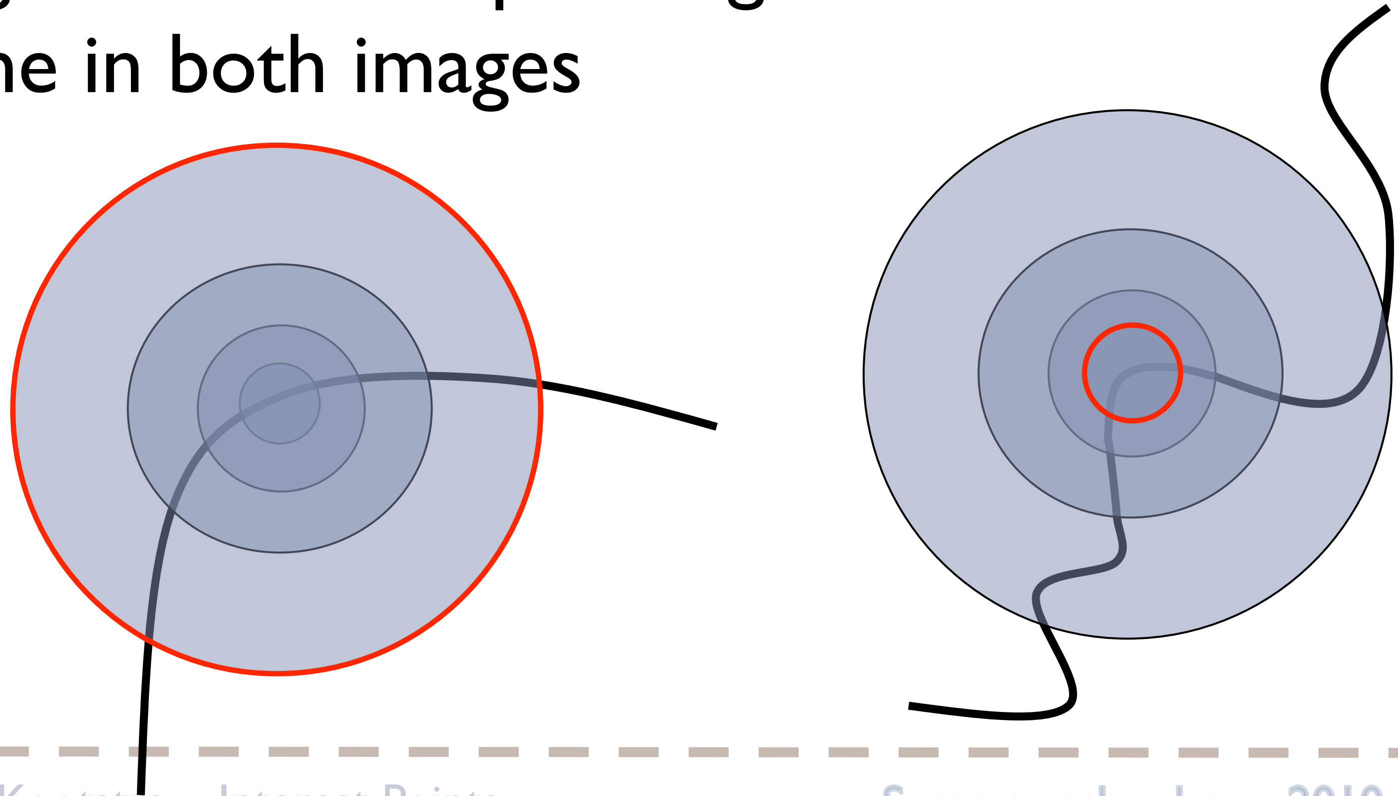


All points will be classified as edges

Corner !

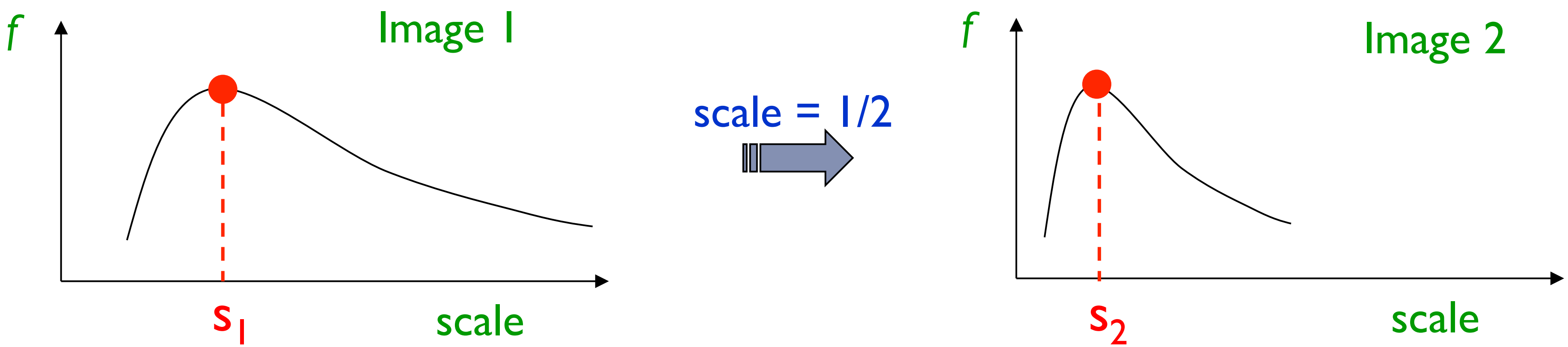
Scale-Invariance

- Consider regions (e.g. circles) of different sizes around a point
- Regions of corresponding sizes will look the same in both images



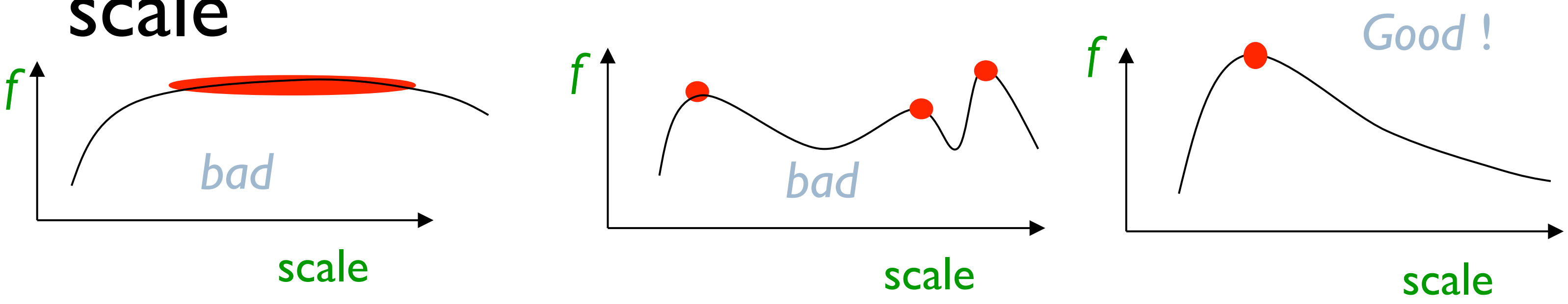
Scale-Invariance Detection

- Investigate the saliency (cornerness, …) at different scales (T. Lindeberg).
- Characteristic scale: the scale that corresponds to the peak saliency

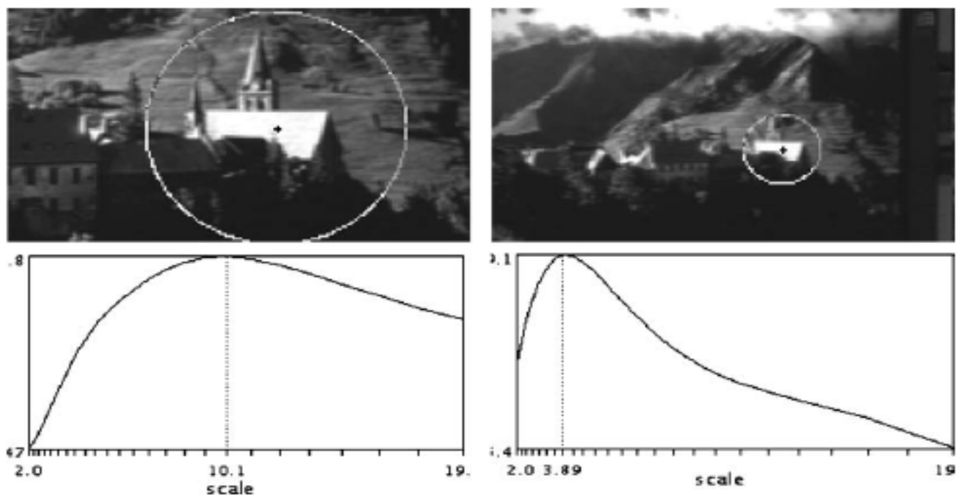


Scale-Invariance Detection

- A good interest point corresponds to a unique scale

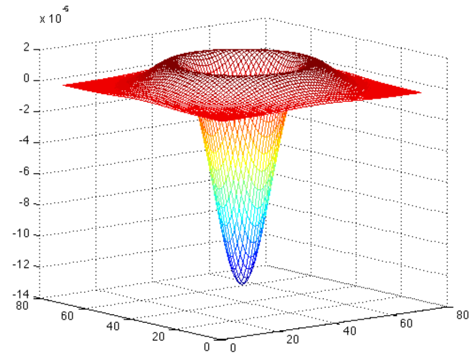


- Example



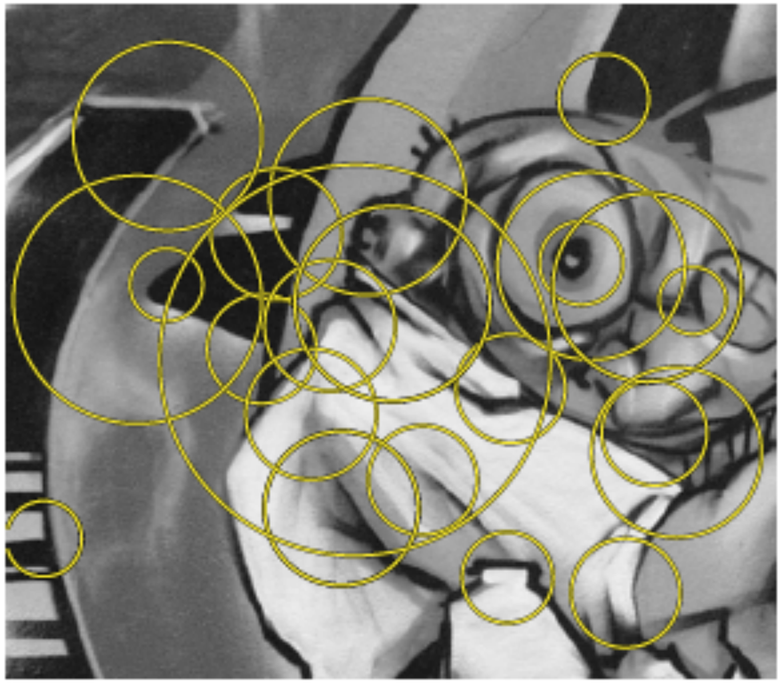
Harris-Laplace Detector (Mikolajczyk *et al* 2004)

- Using Laplacian of Gaussians for scale selection
  - Blob detection



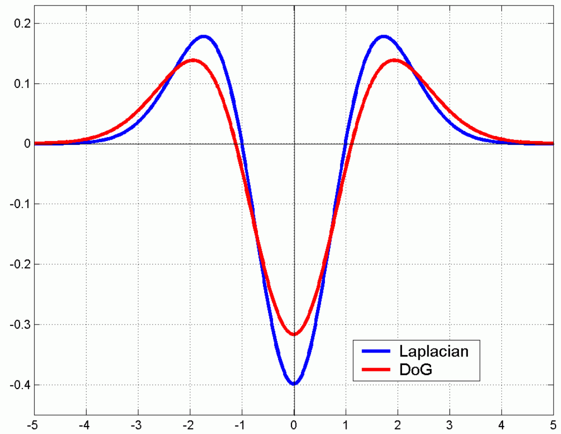
- Two steps
  - Finding Harris points at different scale
  - Finding characteristic scale iteratively
    - Find local extremum over scale σ^{k+1} in LoG for every Harris point $\mathbf{x}^k$.
    - Reposition point by find local maximum in Harris measure close to $\mathbf{x}^k$ for scale σ^{k+1}.
    - Continue until convergence

Harris-Laplace Detector



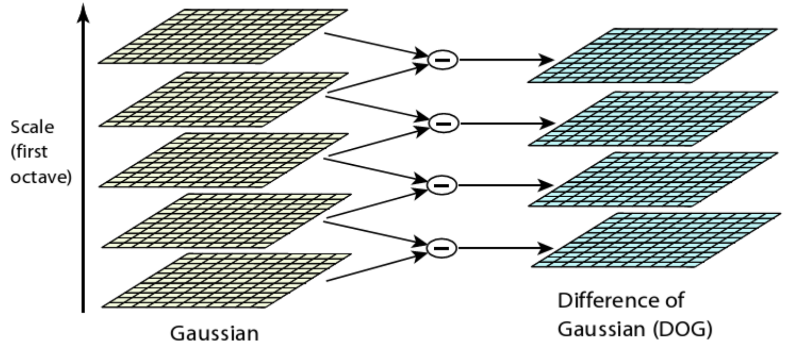
Scale-Invariant Feature Transform

- SIFT (Lowe 2004)
  - Detects interest points on blobs
  - Invariant to scale and rotation
- Based on Difference of Gaussians
  - Approximation of Laplacian of Gaussians
  - Faster
  - Second-order derivative of image intensity



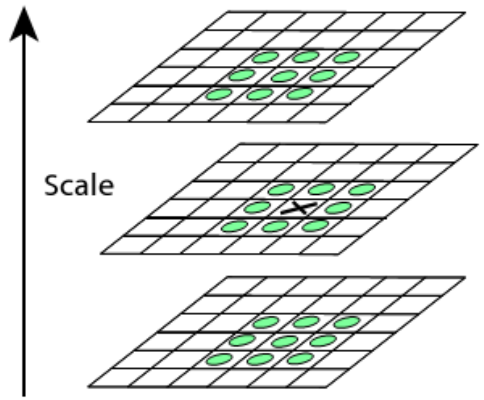
Scale-Invariant Feature Transform

- Pyramid of Gaussian images for different scales
  - $L(x,y,\sigma) \quad = \quad G(x,y,\sigma) * I(x,y$

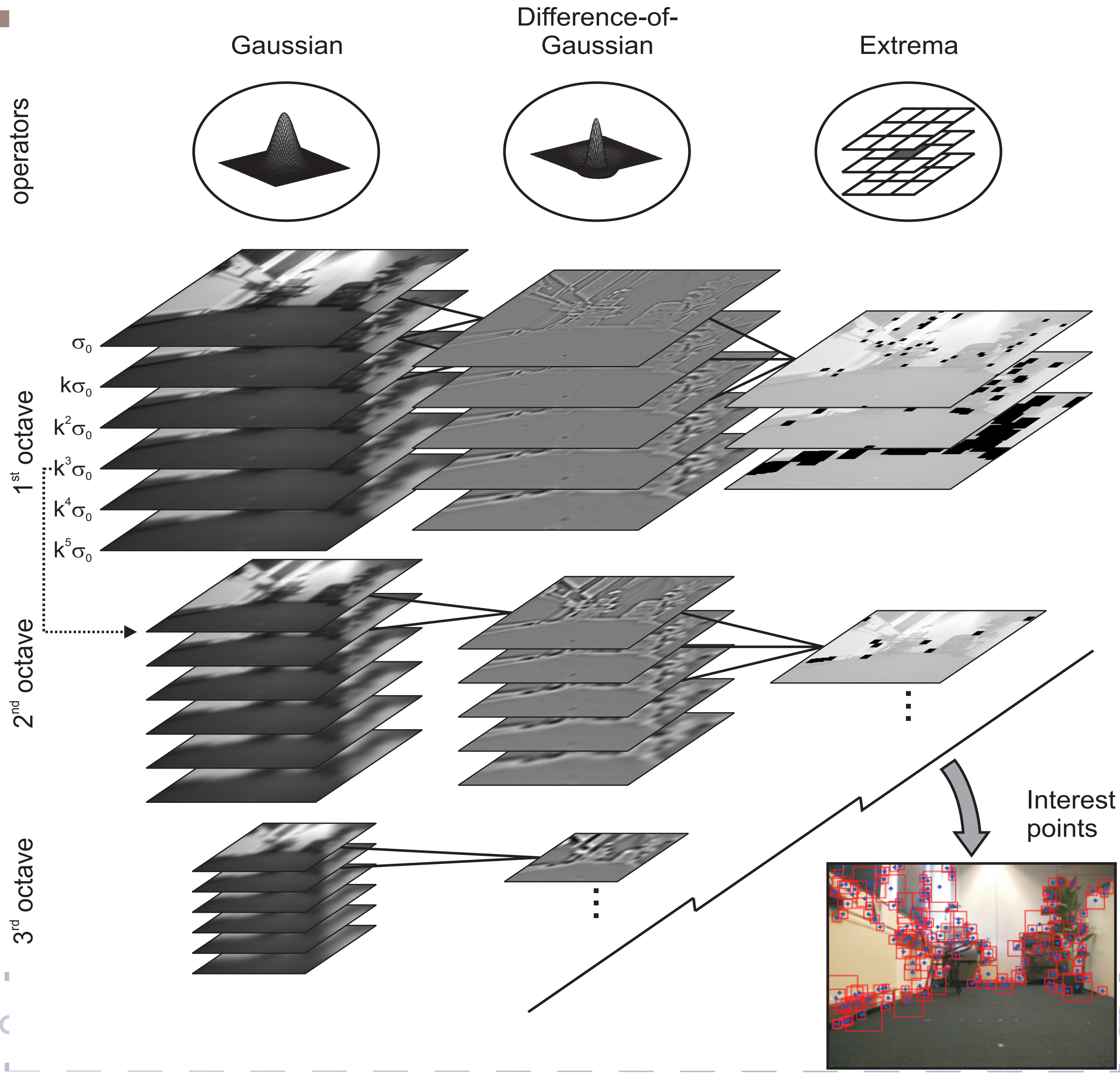


- Pyramid of DoG images
  - $D(x,y,\sigma) \quad = \quad L(x,y,k\sigma) - L(x,y,\sigma)$

- Local extrema detection
  - Minima and maxima in local 3x3x3 scale-space



Scale-Invariant Feature Transform

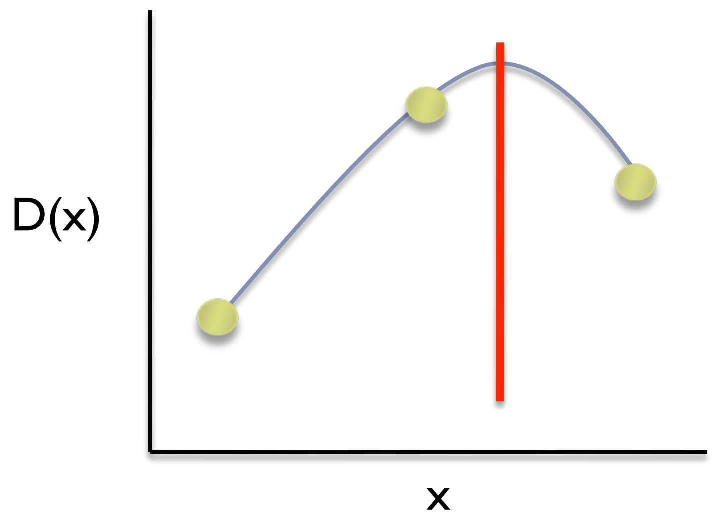


Accurate Localization of IP

- Sub-pixel localization of the interest point
  - Especially important for higher/coarser scales
- Fitting a quadratic function to the surrounding values using Taylor expansion

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x},$$

- Find optimum of D(x)



Eliminate Edge Responses

- Using the DoGs some interest points will be found along strong edges in the image
- Edge point are not uniquely localizable
- Test 'blobness' using the Hessian
- The eigenvalues of H are proportional to the curvature of D

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix},$$

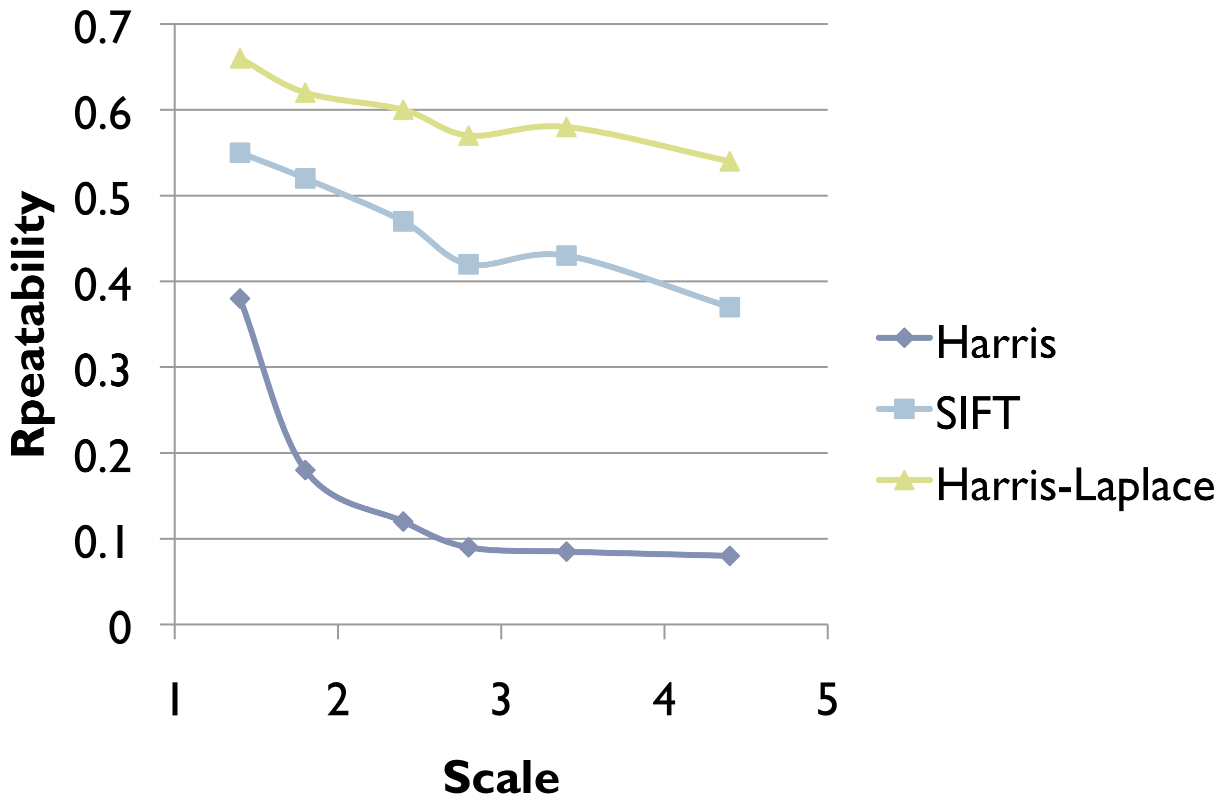
$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(\tau_r + 1)^2}{\tau_r};$$

- Only accept points with similar eigenvalues (ratio between the two is lower than τ_r)

SIFT Detector Example

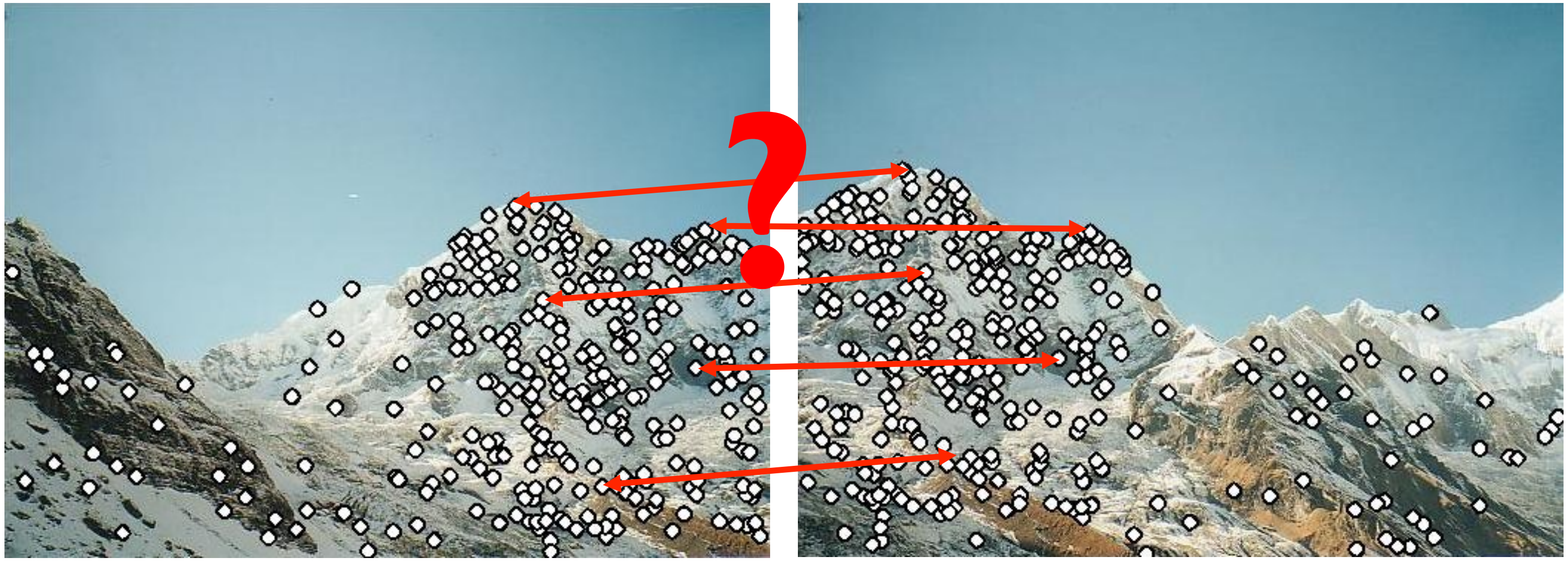
Interest-Point Detectors

- Repeatability



Interest-Point Descriptor

- We now know how to detect interest points
- Now we need to describe them, in order to recognize them later



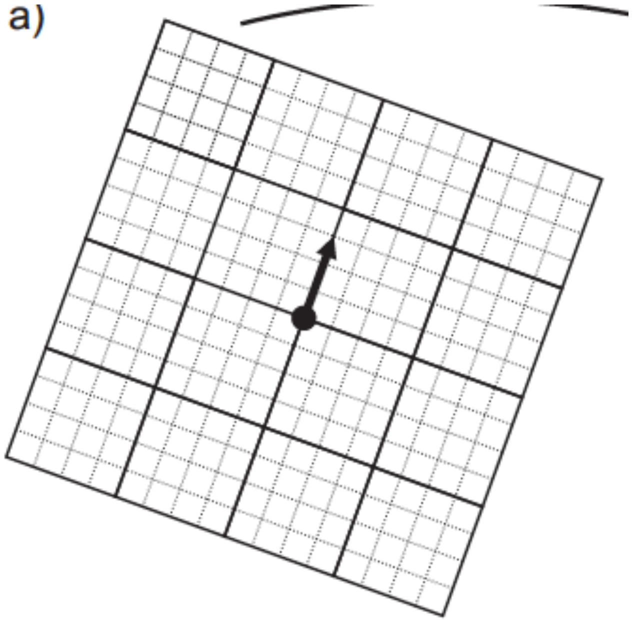
SIFT Descriptor

- The SIFT descriptor (Lowe 2004)
  - Currently most popular descriptor
  - Based on Histograms of Oriented Gradients
  - Describes the texture in the IP's neighborhood
  - Provides quite unique and identifiable descriptors

Scale and Rotational Invariant

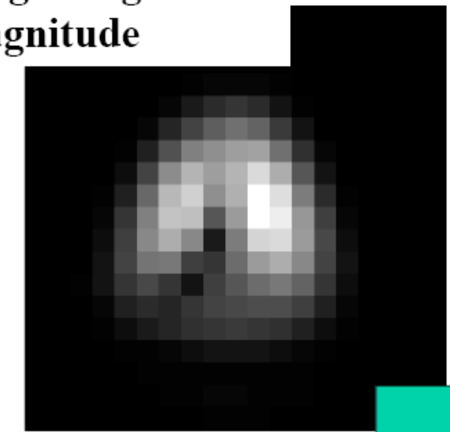
- Scale and Rotational Invariant
  - Size of window depending on the scale of the IP
  - Orientation based on dominant gradient orientation in the local surrounding of the IP
  - If multiple dominant orientations, then multiple descriptors

a)

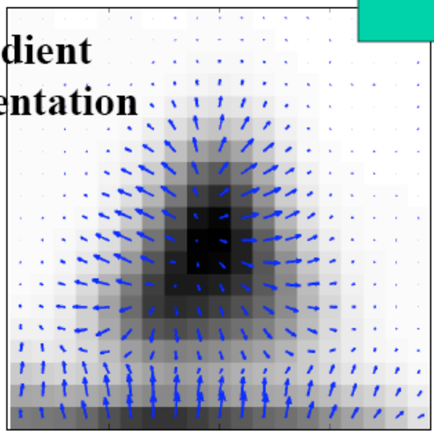


Orientation Assignment

weighted gradient magnitude

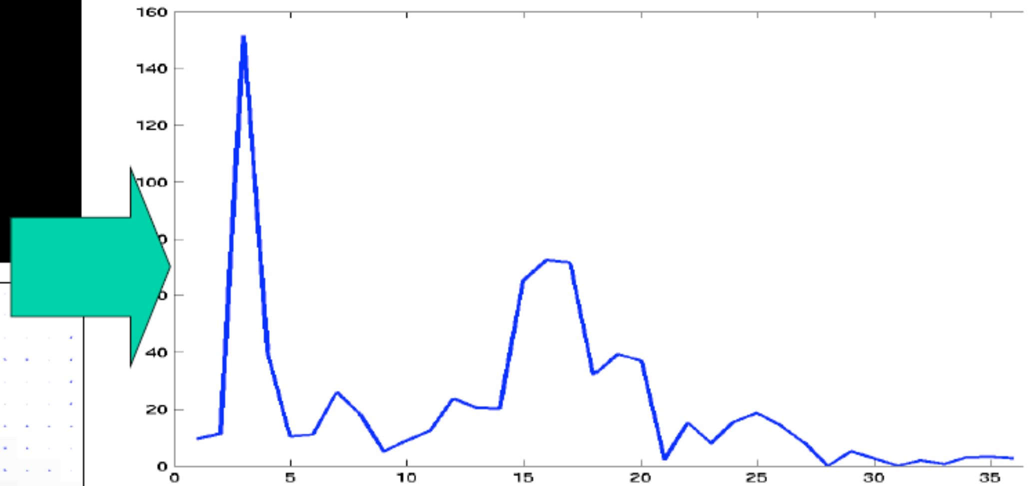


gradient orientation



weighted orientation histogram.

Each bucket contains sum of weighted gradient magnitudes corresponding to angles that fall within that bucket.



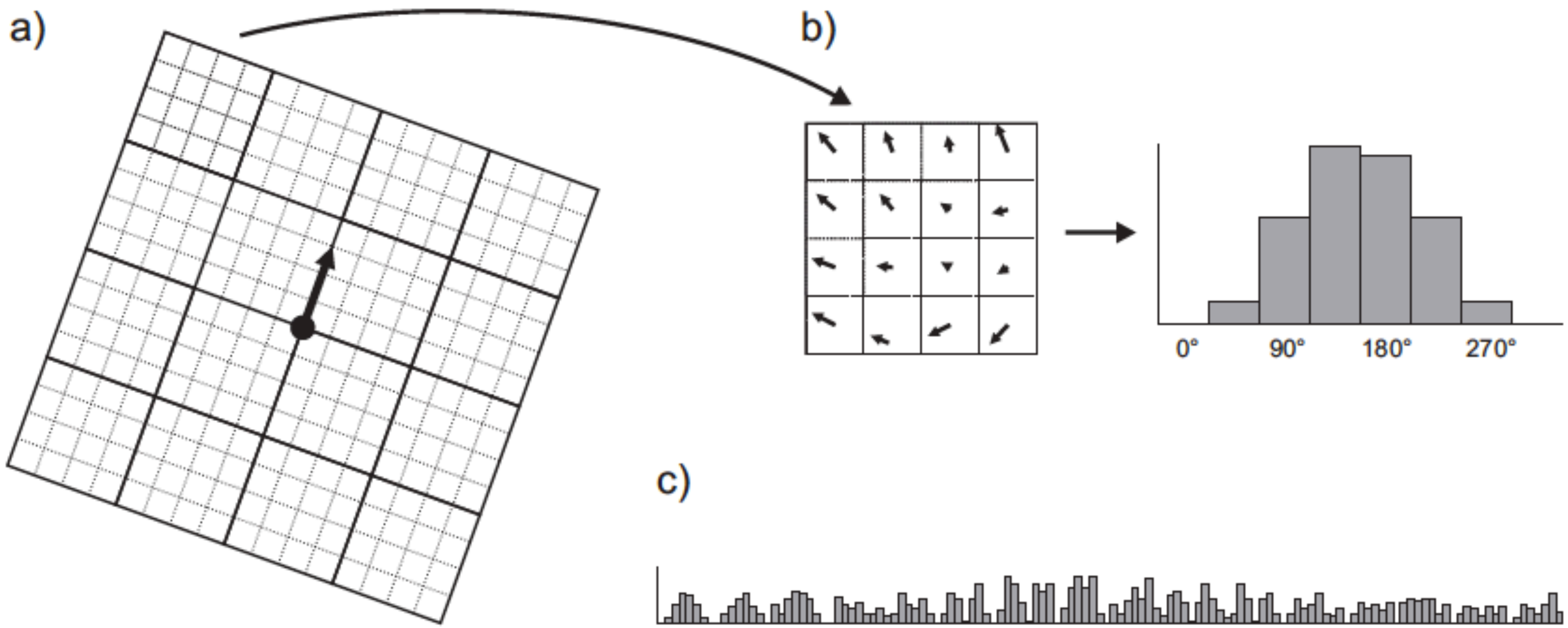
36 buckets

10 degree range of angles in each bucket, i.e.

- 0 <=ang<10 : bucket 1
- 10<=ang<20 : bucket 2
- 20<=ang<30 : bucket 3 …

Histograms of Oriented Gradients

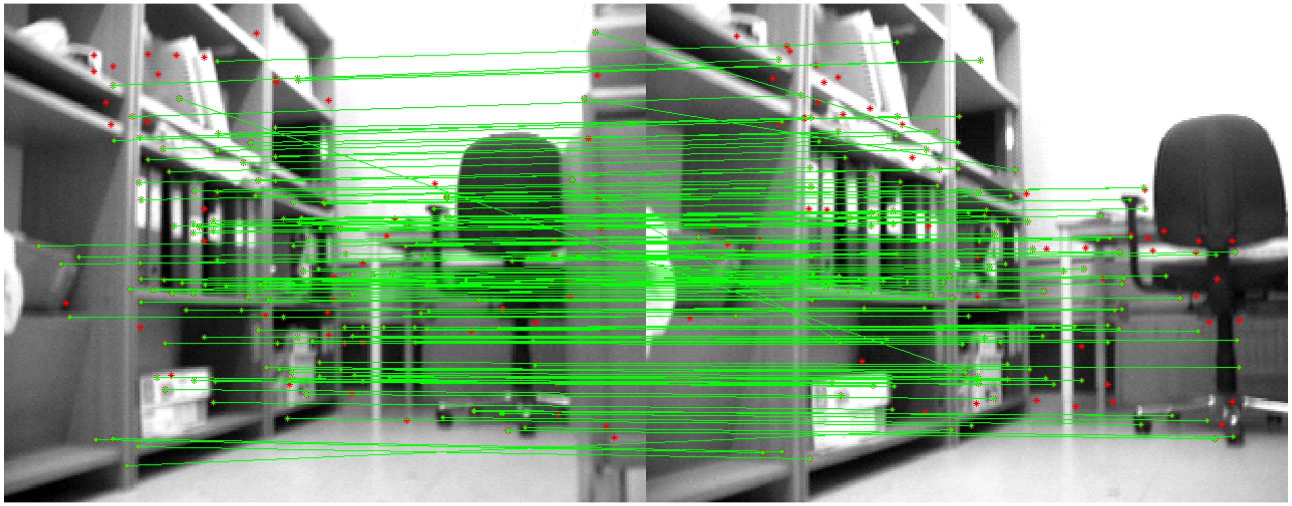
- HOGs
  - 4x4 histograms, 8 bins per histogram = 128 features

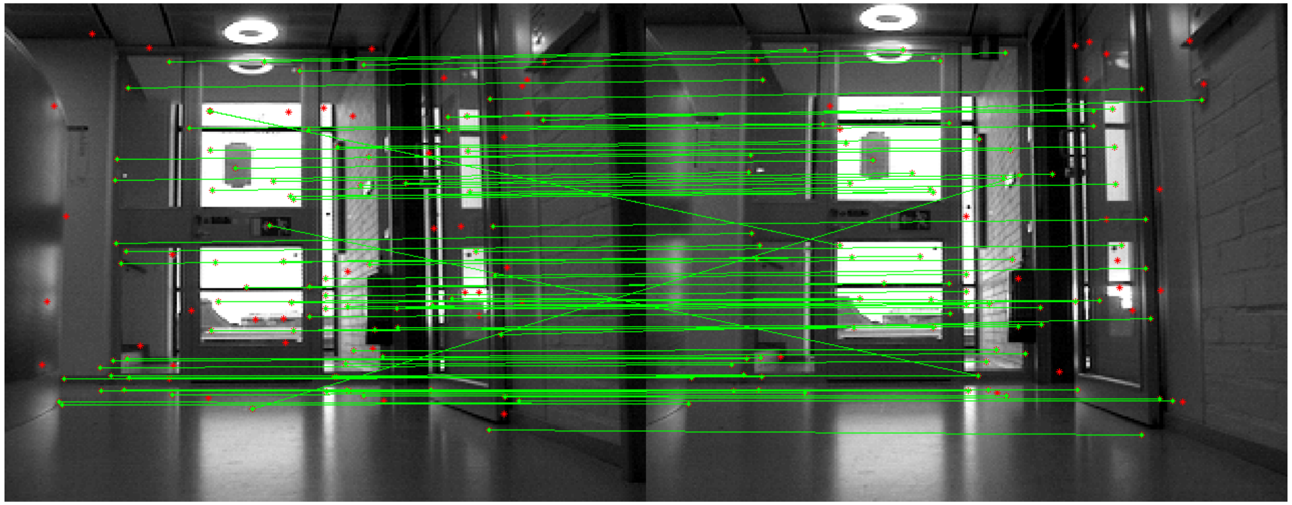


SIFT Descriptor

- Robust to illumination
  - Changes in illumination have little effect on the orientation of the image gradients
  - Might have some effect on the gradient magnitudes, but therefore the histograms are normalized.

Matching example





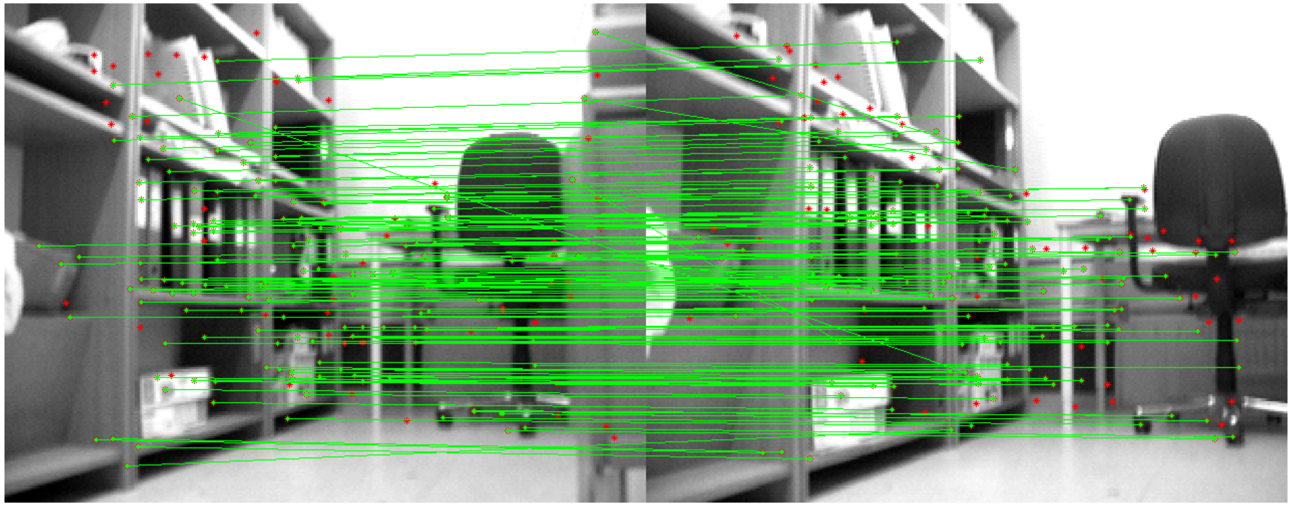
SIFT Results

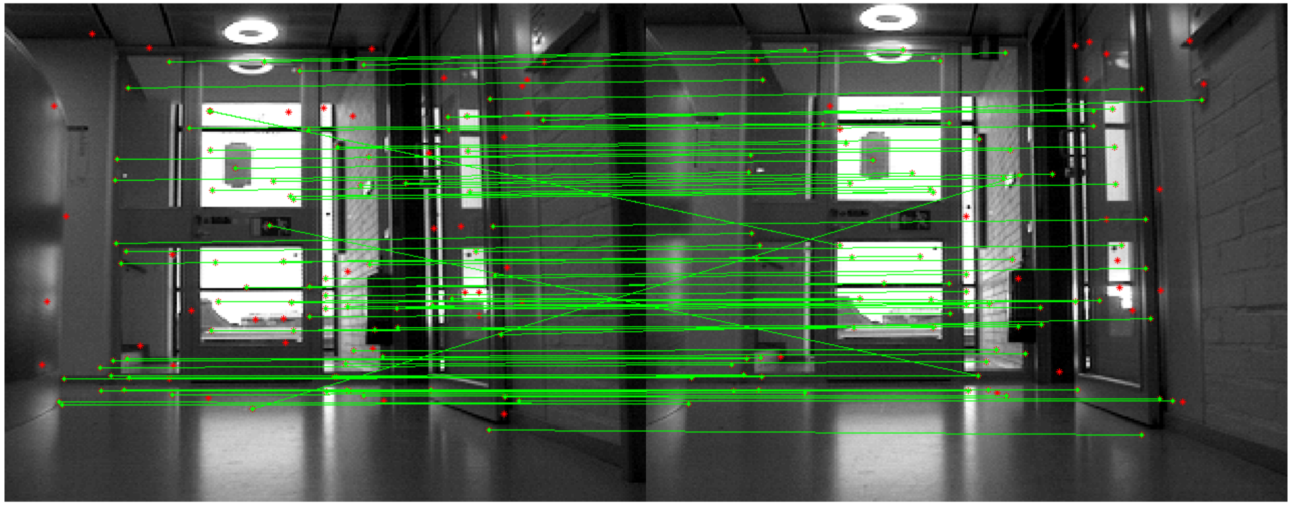
- Best performance in the study of Mikolajczyk & Schmid 2005
- Among best for all tests:
  - Viewpoint changes
  - Scale changes
  - Image rotation
  - Image blur
  - JPEG compression
  - Illumination changes

Some Examples

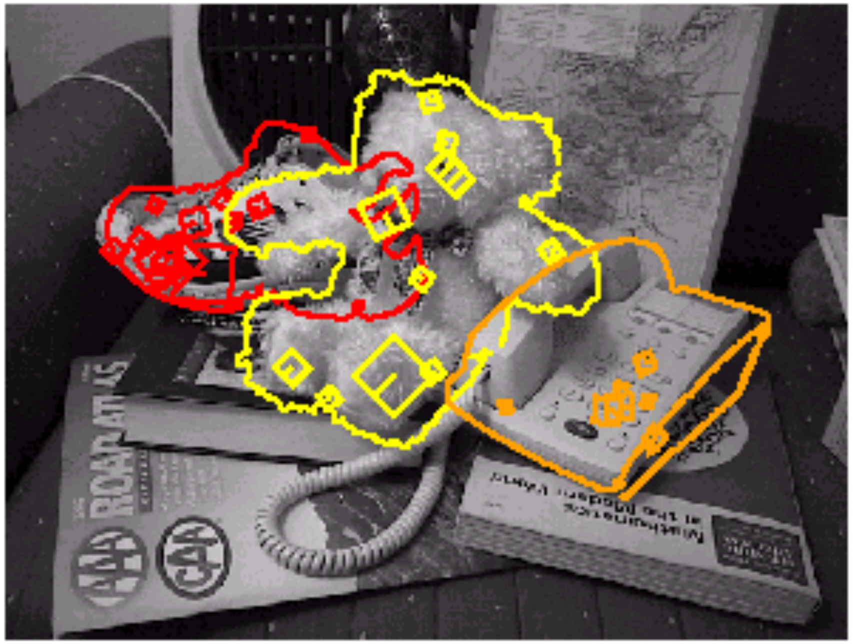
- …

Robotic Localization and Mapping

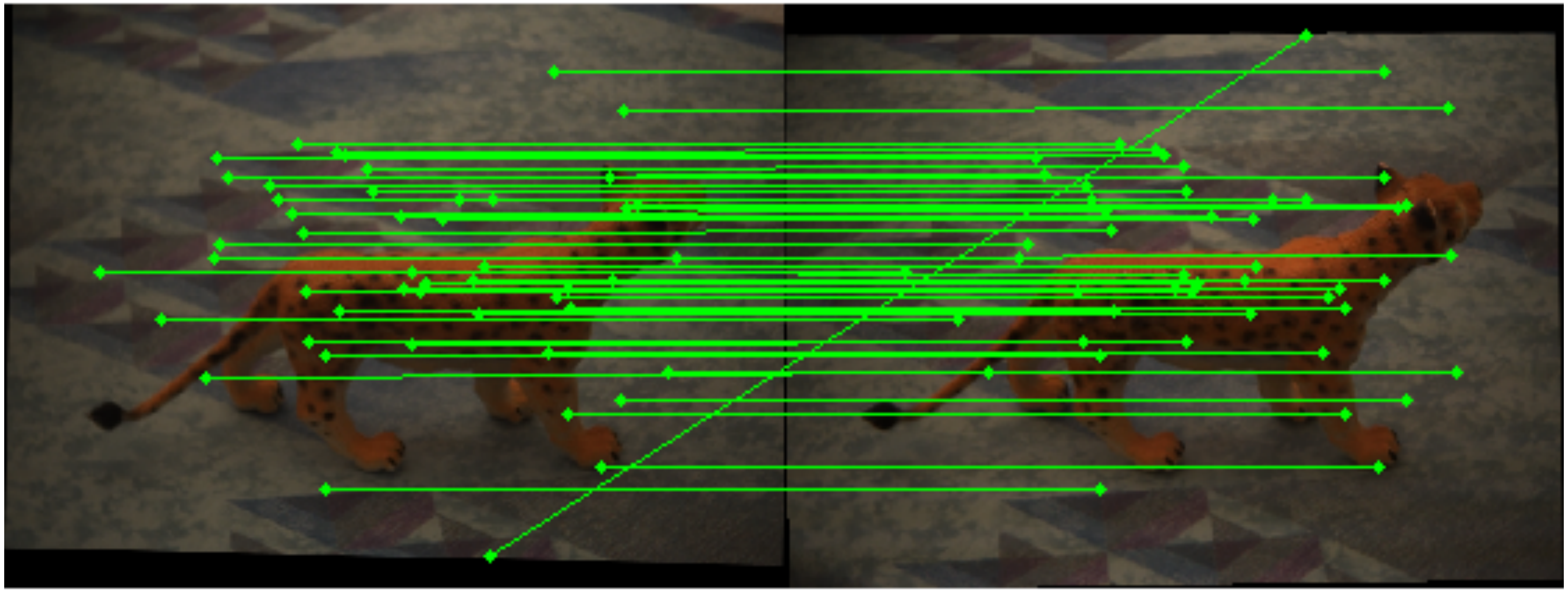




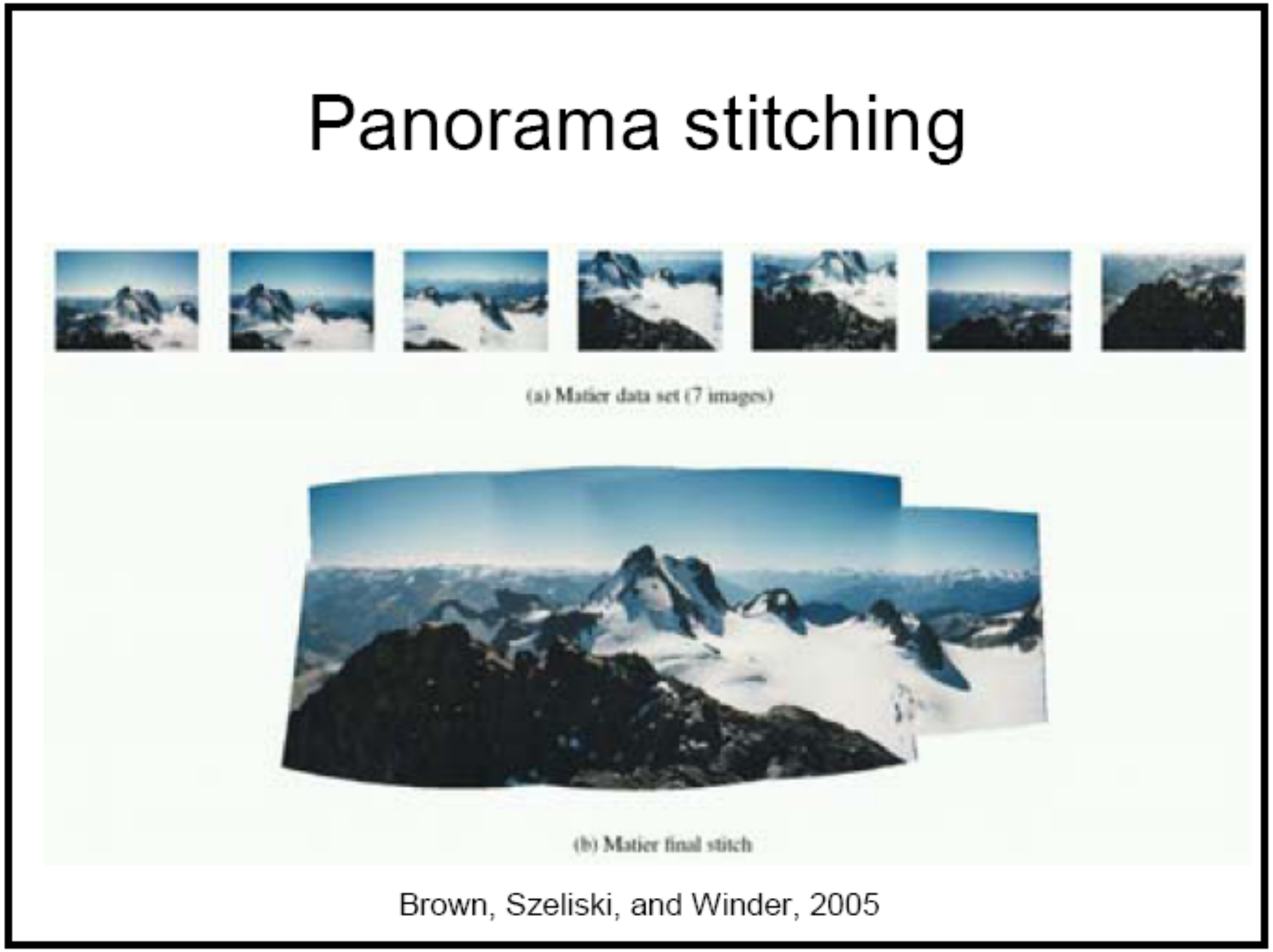
Object Recognition



Stereo Matching



Panorama Stitching

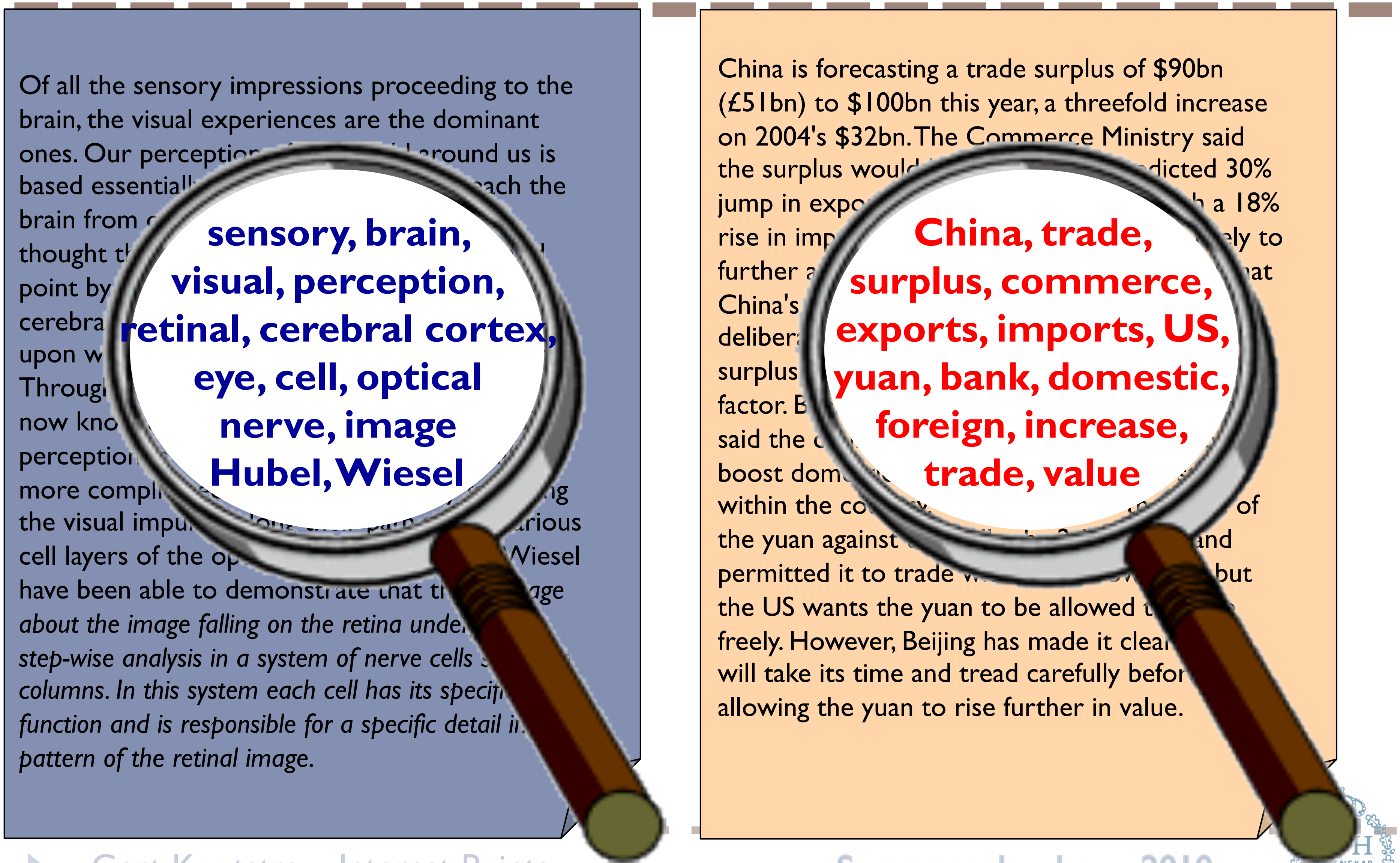


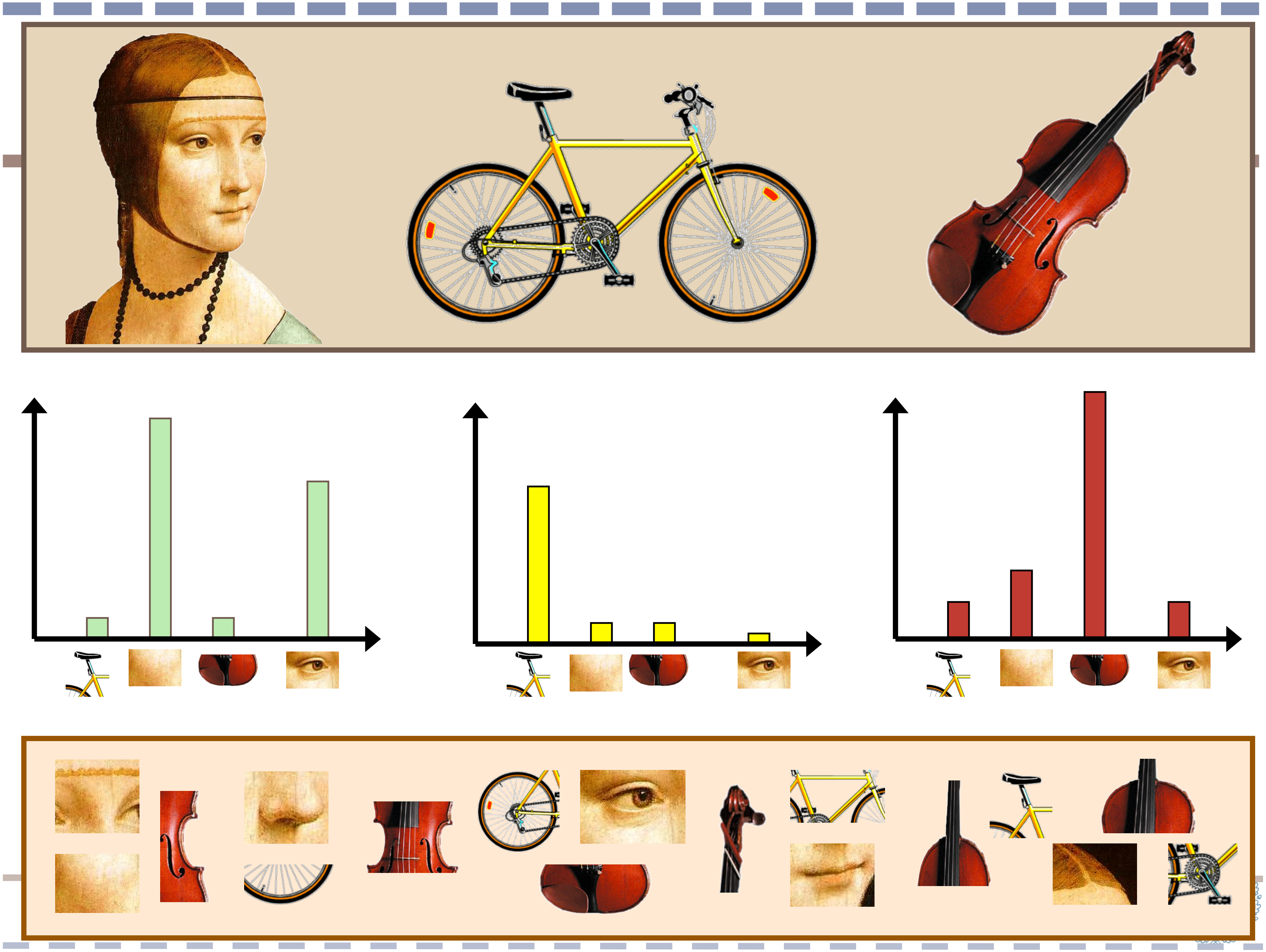
Bag of Features

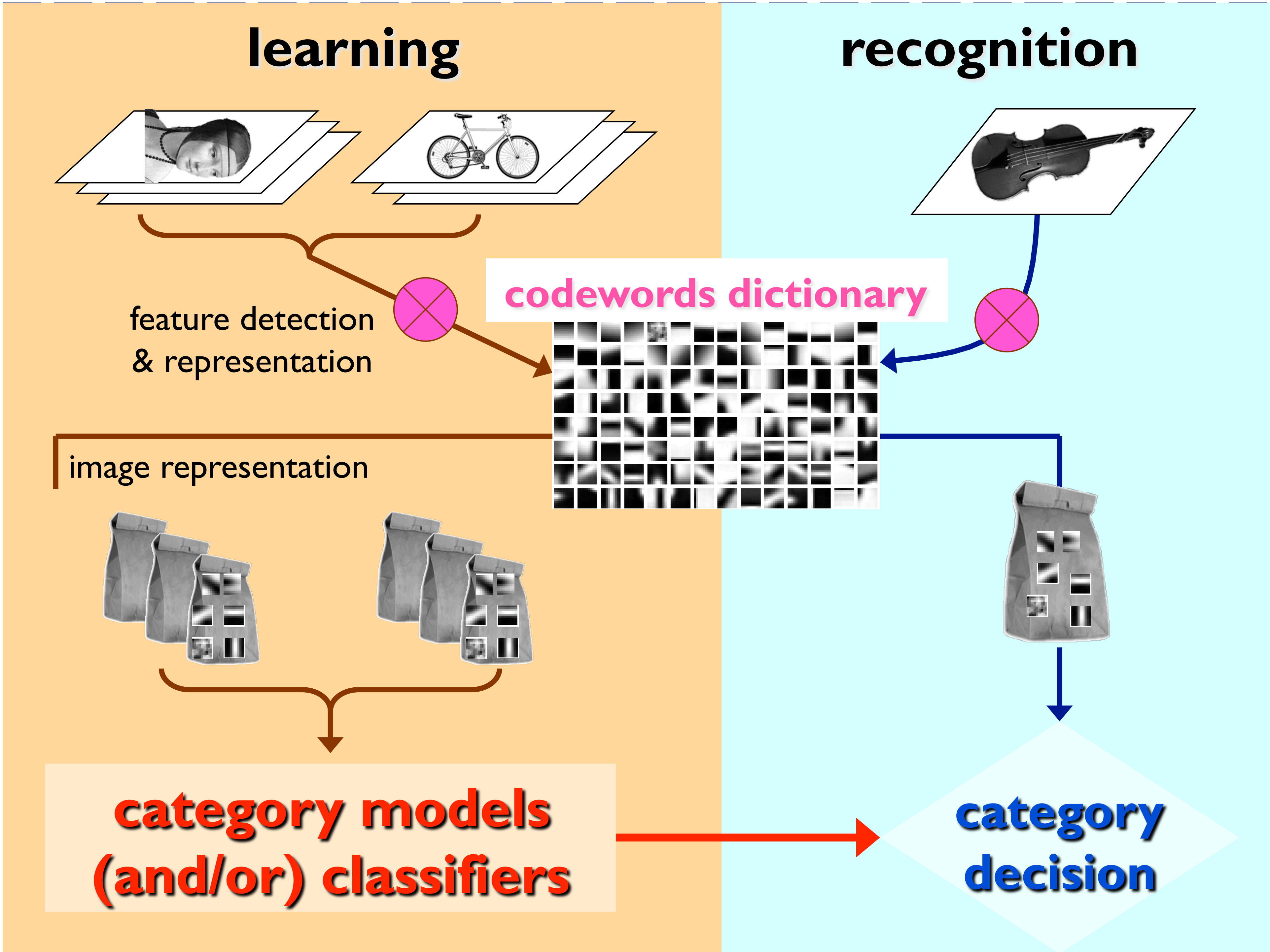
- In the presented form interest points are very suitable for object recognition
- Not so good for object/image classification and retrieval
  - SIFT points and descriptor are too specific
  - Variable number of points, so total feature vector of image has unknown size
- The Bag-of-Features approach

Image → Bag of ‘words’

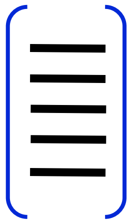
Analogy to documents





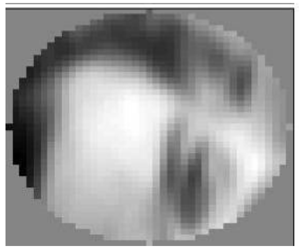

learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

Interest Point Features

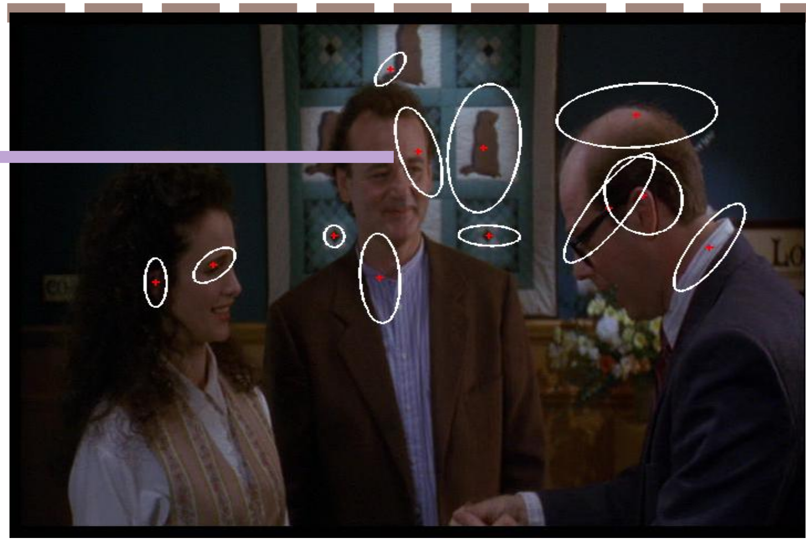


Compute SIFT descriptor

[Lowe'99]



Normalize patch



Detect patches

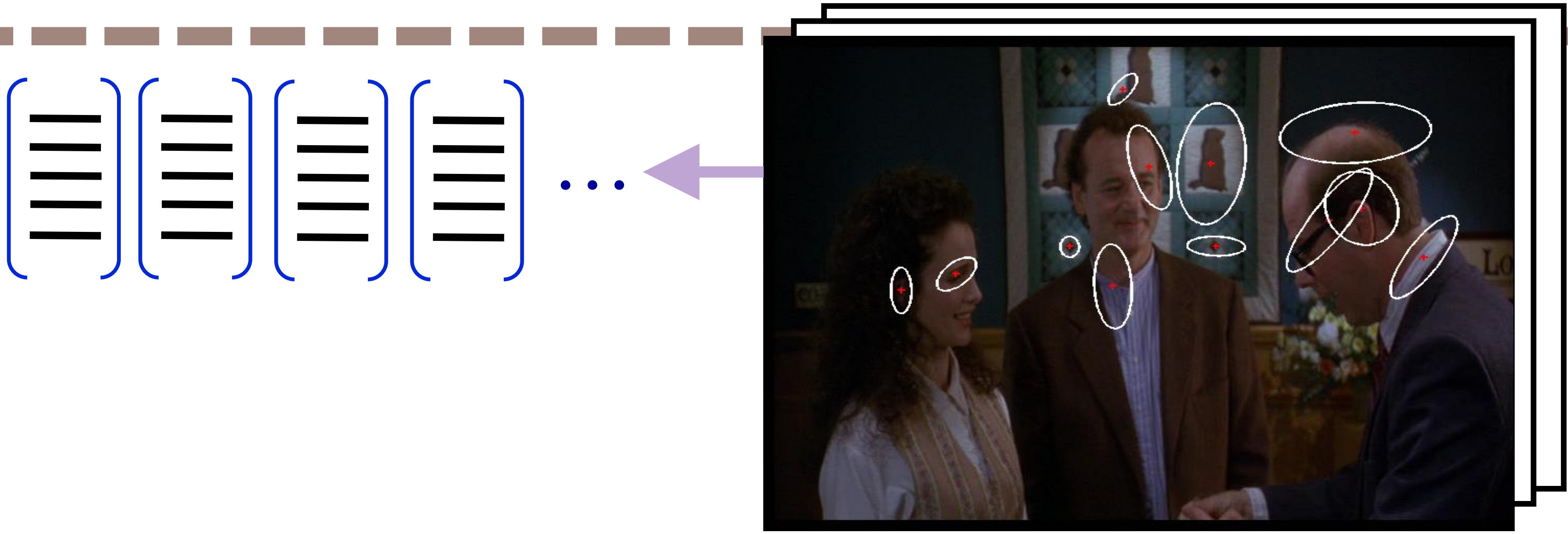
[Mikojaczyk and Schmid '02]

[Matas et al. '02]

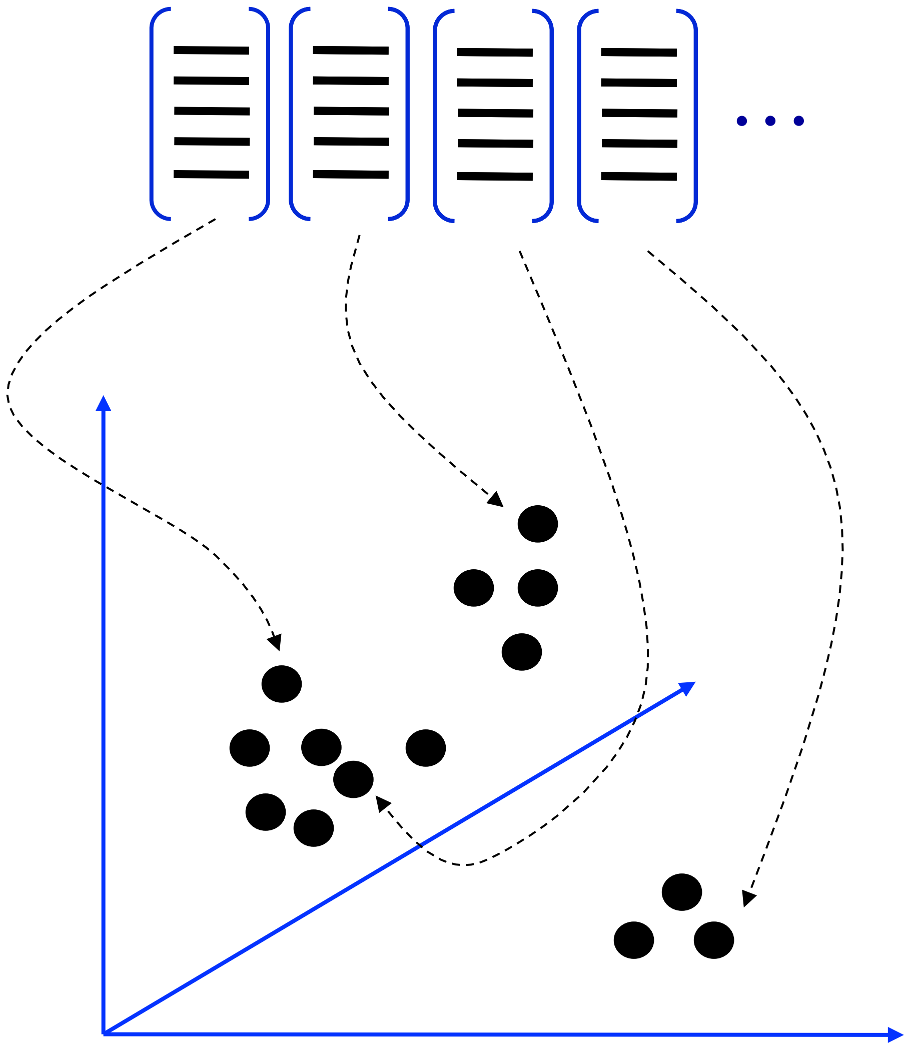
[Sivic et al. '03]

Slide credit: Josef Sivic

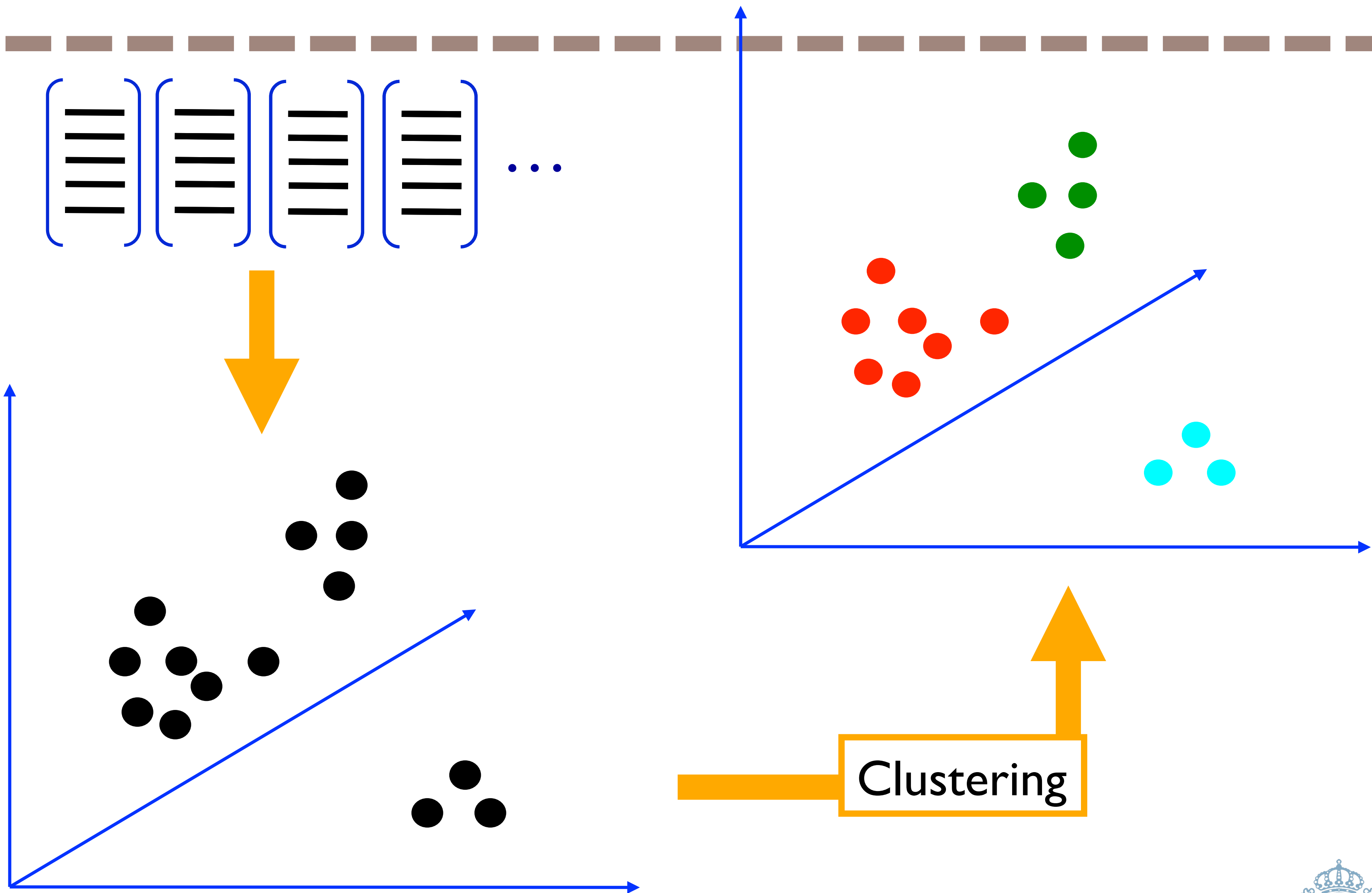
Interest Point Features



dictionary formation

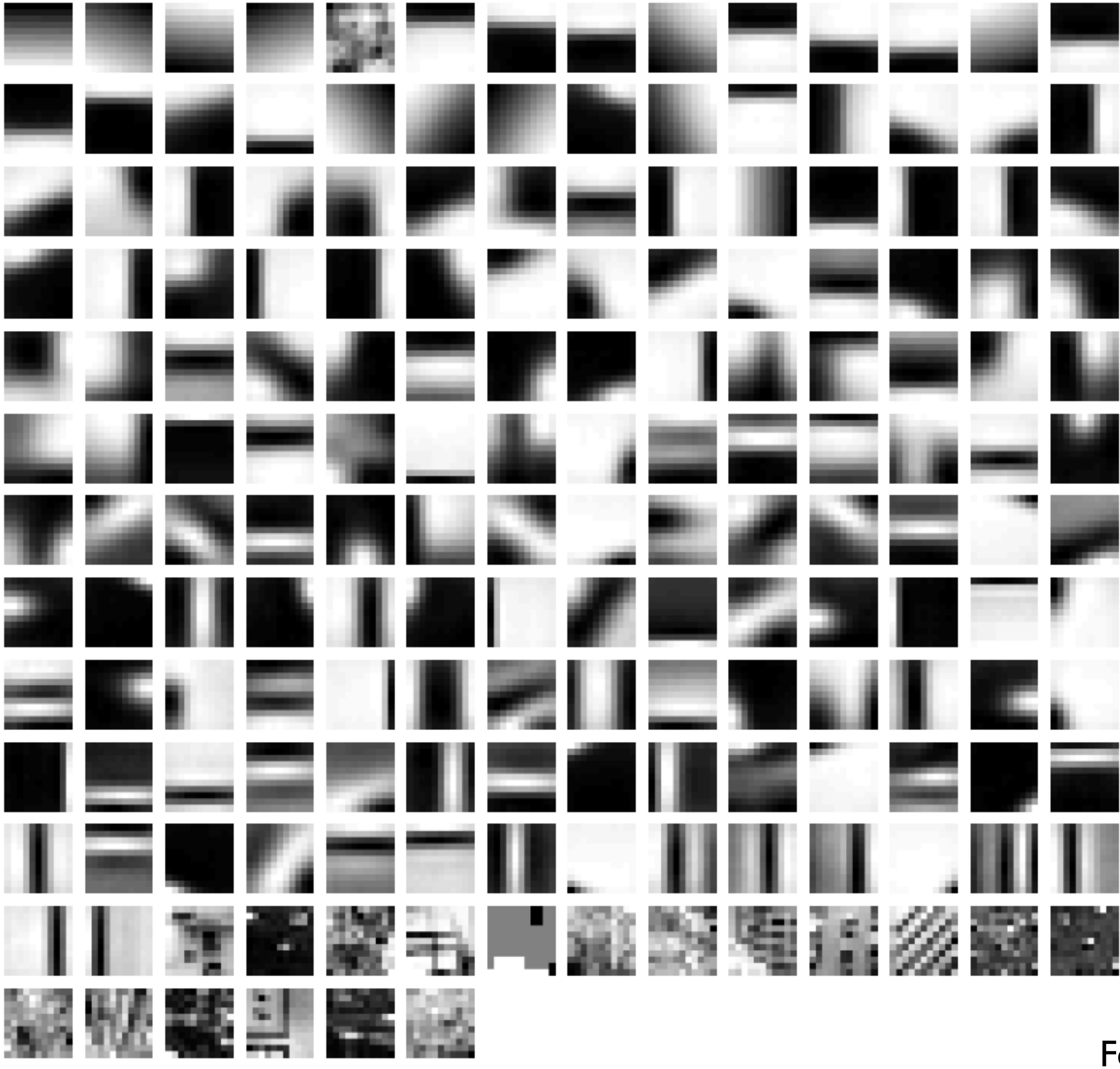


Clustering (usually k-means)



Slide credit: Josef Sivic

Clustered Image Patches



Fei-Fei et al. 2005

Image representation

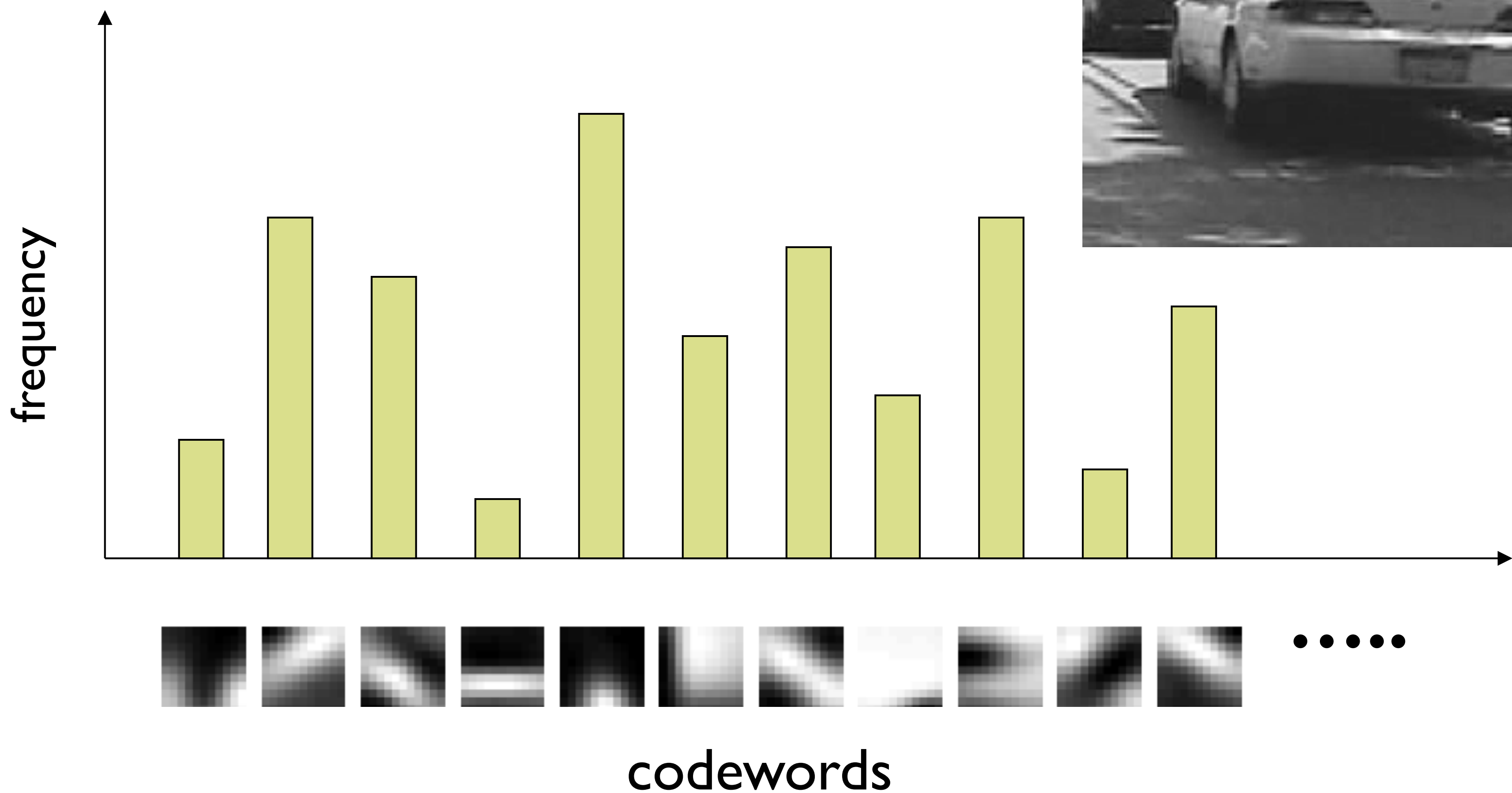
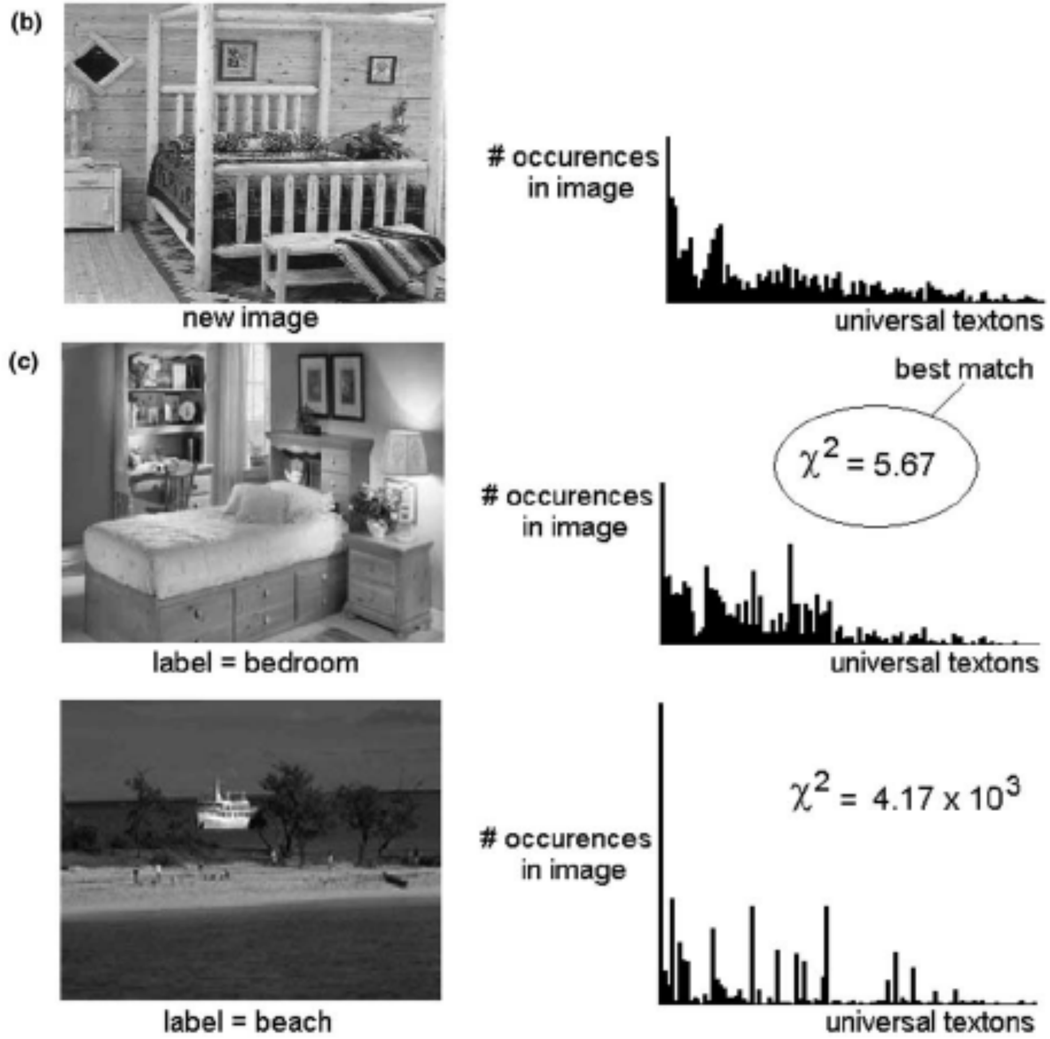


Image Matching

- Interest points
- IP descriptors
- Make visual-word histogram
- Compare histogram to histograms in the databse



Bag of Words

- Works well for image/object classification
- Reduces the number of features
  - Standard SIFT
    - ±1,000 IPs per image, 128 D feature vector
  - Bag of Words
    - 1,000-10,000 words
- But loss of geometric information

Summary

- Local features
- Interest-point detectors
  - Harris / Harris-Laplace
  - SIFT detector (DoG)
- Interest-point descriptors
  - SIFT descriptor (HOG)
- Bag of words