Digital Image Processing (CSE/ECE 478)

Lecture 18 : Representation and Description (2)

Ravi Kiran

Rajvi Shah

Recap : Image Representation & Description

- Defining Representation & Description
- Internal vs. External (Shape vs. Region)
- Boundary representation & description
- Texture Description

Modern Features / Descriptors

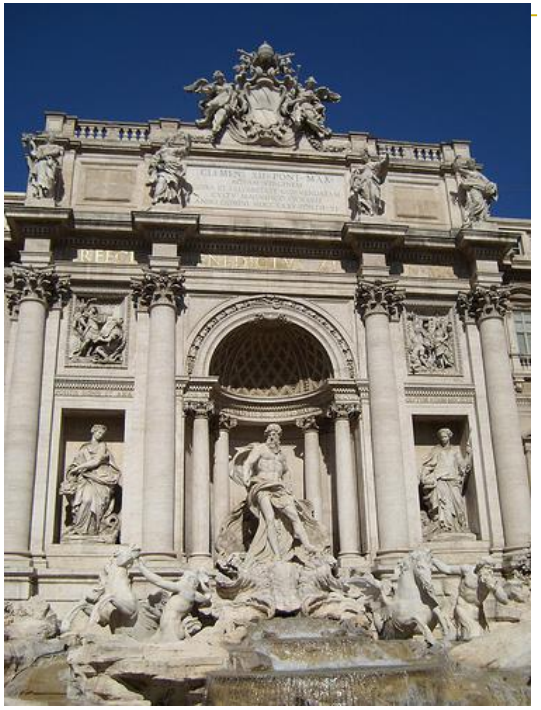
- Point Descriptors : SIFT, SURF, DAISY, LBP
- Region Descriptors : HOG, MSER
- Global Descriptors : Bag of Words, GIST
- Introduction to Learned Representation

Many slides borrowed from

UW Vision course by Steve Seitz
https://courses.cs.washington.edu/courses/cse576/09sp/

Image matching

by Diva Sian



by swashford

Harder case

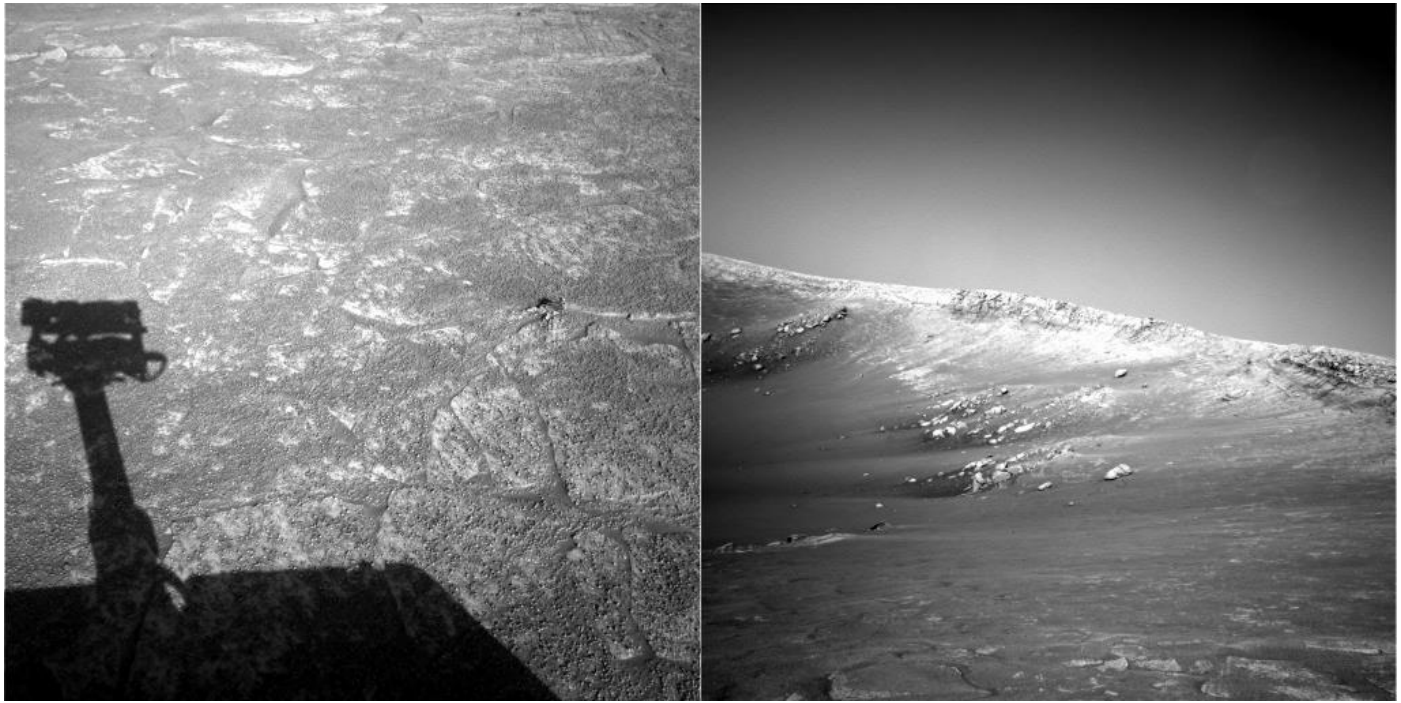
by Diva Sian

by scgbt

Even harder case

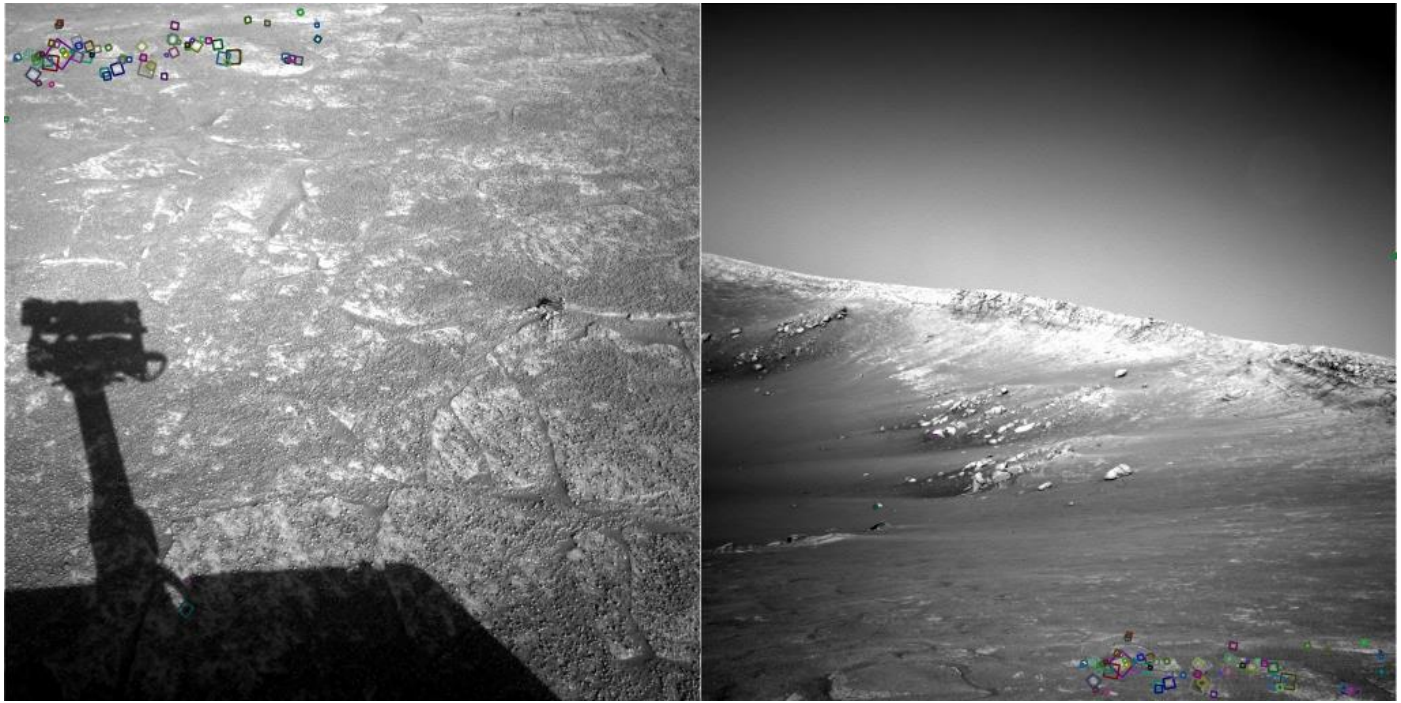
“*How the Afghan Girl was Identified by Her Iris Patterns*” Read the story

Harder still?



NASA Mars Rover images

Answer below (look for tiny colored squares…)



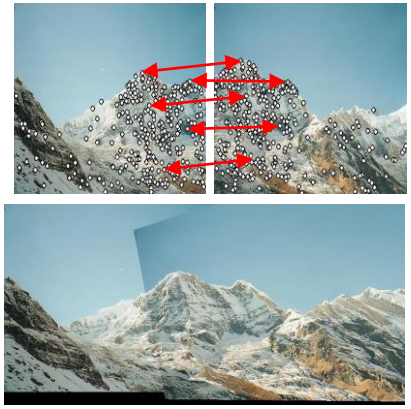
NASA Mars Rover images
with SIFT feature matches
Figure by Noah Snavely

Local vs. Global Features

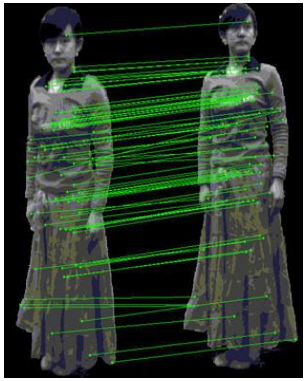
Object Recognition

Scene Classification

Image Registration

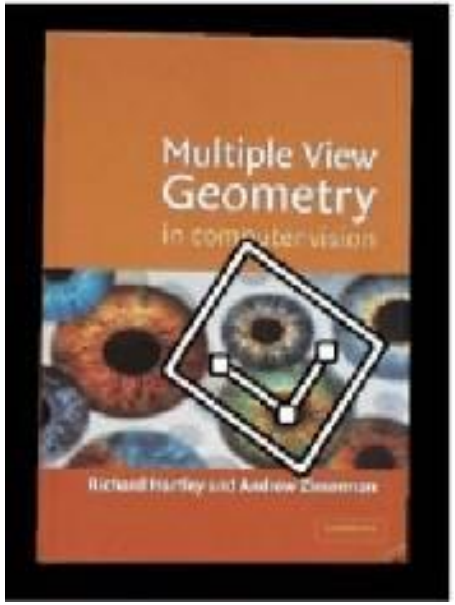


3D Reconstruction

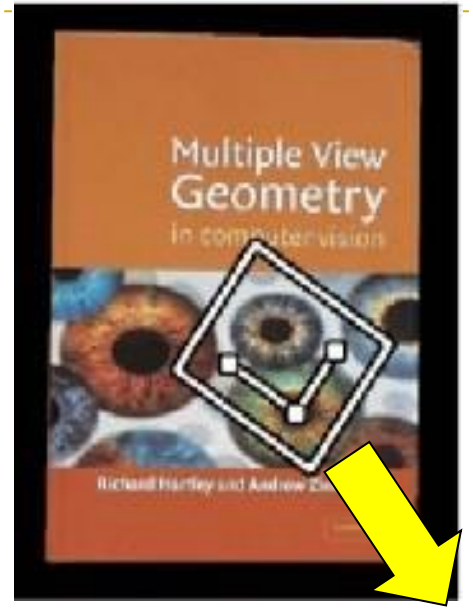


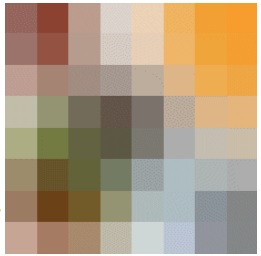
Landmark Recognition

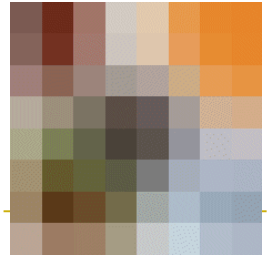
Local Features based Image Matching



Local Features based Image Matching



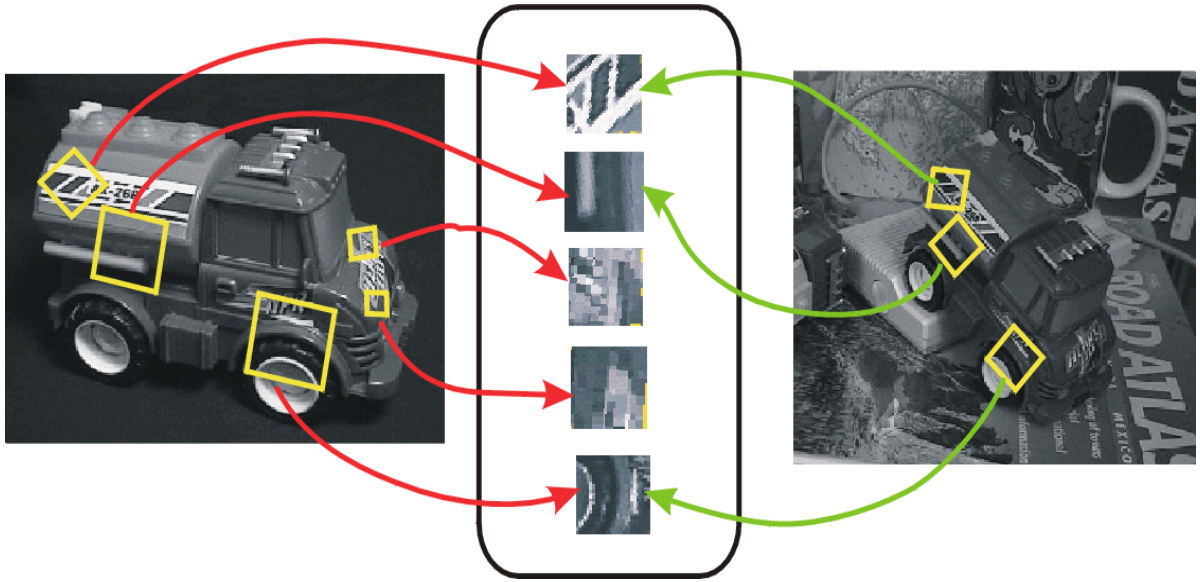




Invariant local features

Find features that are invariant to transformations

- geometric invariance: translation, rotation, scale
- photometric invariance: brightness, exposure, …



Feature Descriptors

Local Features based Image Matching

Applications that use …

- Image alignment (e.g., mosaics)
- 3D reconstruction
- Motion tracking
- Object recognition
- Indexing and database retrieval
- Robot navigation
- … other

Want uniqueness

Look for image regions that are unusual

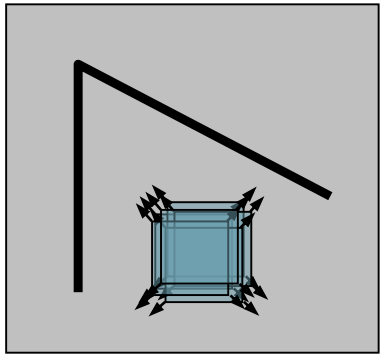
- Lead to unambiguous matches in other images

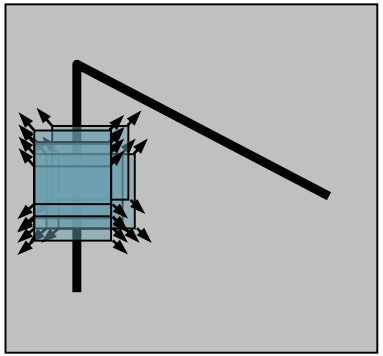
How to define “unusual”?

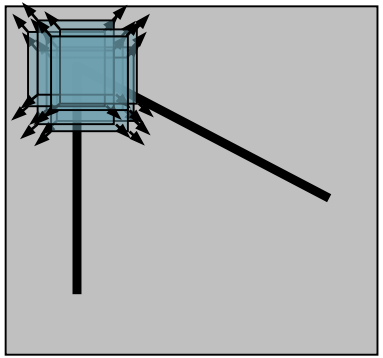
Recap : Harris Corner Detector

(Lec 15)

Harris Detector: Basic Idea







"flat" region:
no change in
all directions

"edge":
no change along
the edge direction

"corner":
significant
change in all
directions

Harris Detector: Mathematics

$$E(u,v) = \sum_{(x,y)\in W} [I(x+u, y+v) - I(x,y)]^2$$

"I'm sorry, Taylor, but Fourier had one of the best series of 1807."

For small shifts $[u,v]$ we have a *bilinear* approximation:

$$I(x+u, y+v) = I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$

$$I(x+u, y+v) \approx I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

$$\begin{aligned} E(u,v) &= \sum_{(x,y)\in W} [I(x+u, y+v) - I(x,y)]^2 \\ &\approx \sum_{(x,y)\in W} \left[I(x,y) + \begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} - I(x,y)\right]^2 \\ &\approx \sum_{(x,y)\in W} \left[\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}\right]^2 \end{aligned}$$

$$E(u,v) = \sum_{(x,y)\in W} \begin{bmatrix} u & v \end{bmatrix} \underbrace{\begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix}}_{H} \begin{bmatrix} u \\ v \end{bmatrix}$$

Moving the window in which directions will result in the largest and smallest E values?

Find out looking at the eigenvectors of ***H***

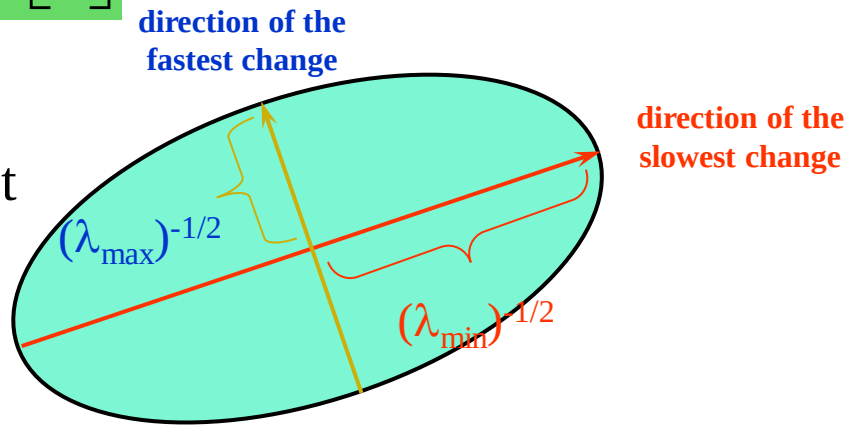
Harris Detector: Mathematics

Intensity change in shifting window: eigenvalue analysis

$$E(u,v) \cong \begin{bmatrix} u, v \end{bmatrix} \quad M \quad \begin{bmatrix} u \\ v \end{bmatrix}$$

λ_1, λ_2 – eigenvalues of M

Ellipse $E(u,v)$ = const

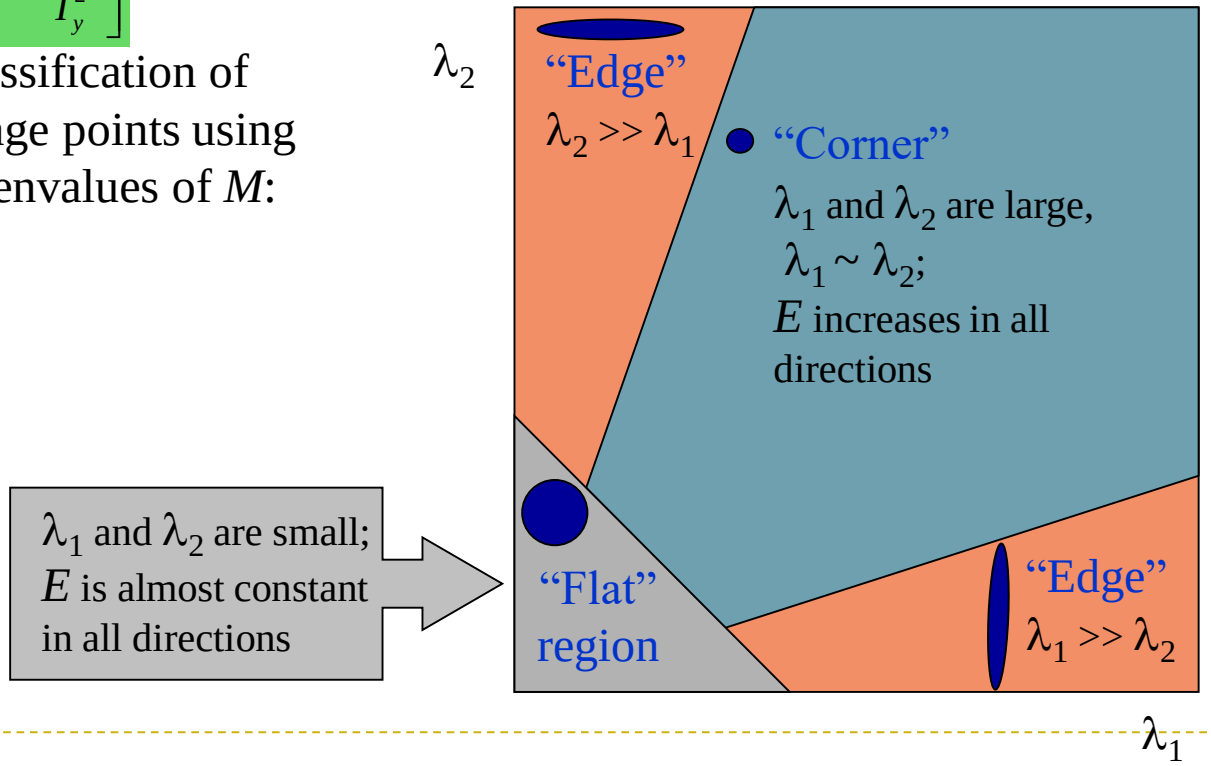


$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Harris Detector: Mathematics

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

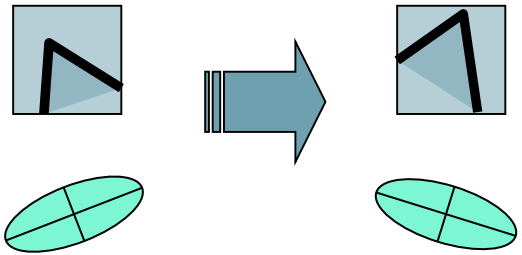
Classification of image points using eigenvalues of M:



Harris Detector: Workflow

Harris Detector: Some Properties

- Rotation invariance



Ellipse rotates but its shape (i.e. eigenvalues) remains the same

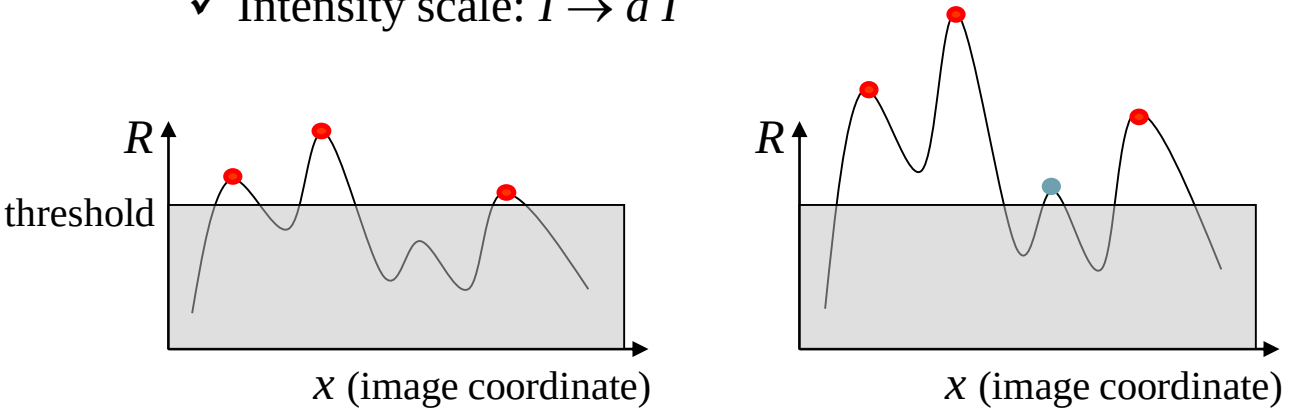
Corner response R is invariant to image rotation

Harris Detector: Some Properties

- Partial invariance to *affine intensity* change

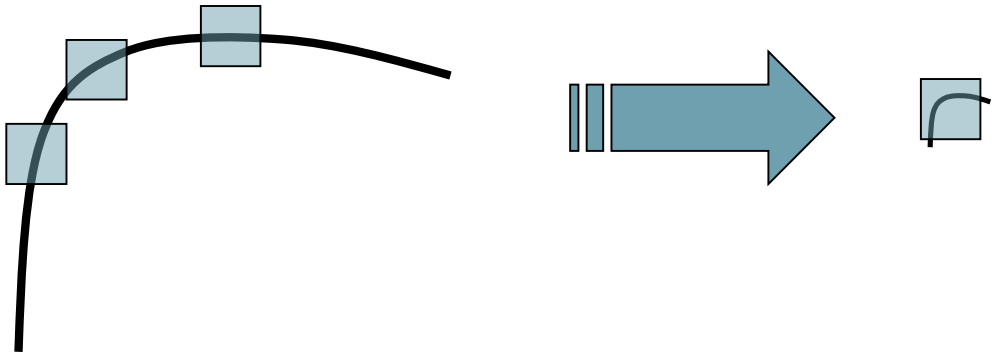
✓ Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$

✓ Intensity scale: $I \rightarrow a\, I$



Harris Detector: Some Properties

- But: non-invariant to *image scale*!

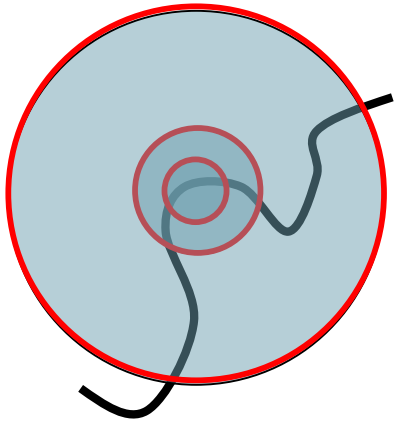


All points will be classified as edges

Corner !

Scale invariant detection

Suppose you're looking for corners



Key idea: find scale that gives local maximum of f

- f is a local maximum in both position and scale
- Common definition of f: Laplacian
  (or difference between two Gaussian filtered images with different sigmas)

Scale space for scale selection

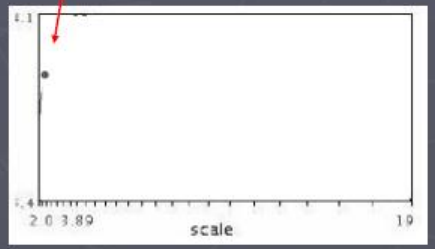
Lindeberg 1994

a major mechanism in algorithms for automatic scale selection, which automatically adapt the local scales of processing to image data. Let us hence generalize the above-mentioned observation to more complex signals and state the following principle for scale selection, to be applied in situations when no other information is available. In its most general form, it can be expressed as follows:

Principle for scale selection:
In the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data.

Automatic scale selection

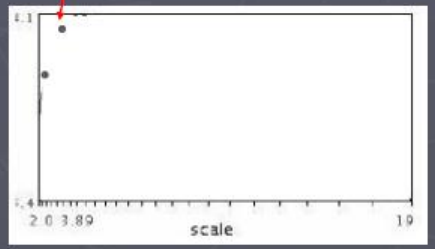
Lindeberg et al., 1996



$$f(I_{i_1 \cdots i_m}(x,\sigma))$$

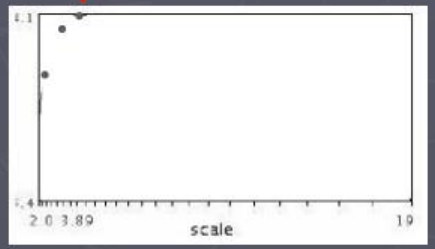
Slide from Tinne Tuytelaars

Automatic scale selection



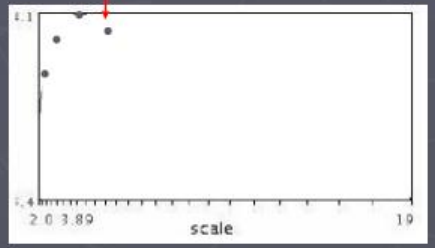
$$f(I_{i_1 \cdots i_m}(x,\sigma))$$

Automatic scale selection



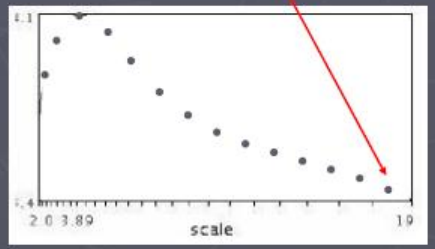
$f(I_{i_1 \cdots i_m}(x,\sigma))$

Automatic scale selection



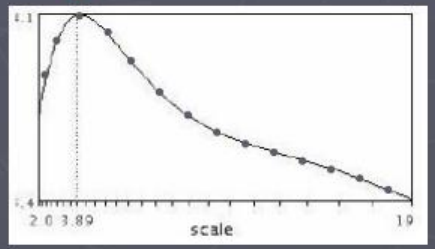
$$f(I_{i_1 \cdots i_m}(x,\sigma))$$

Automatic scale selection



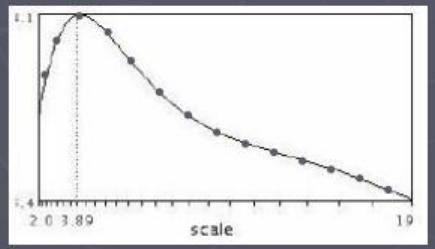
$$f(I_{i_1 \cdots i_m}(x,\sigma))$$

Automatic scale selection

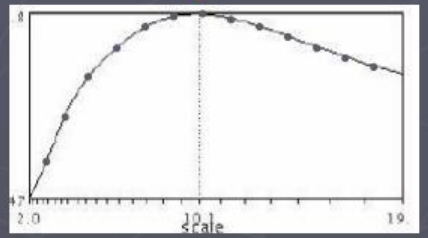


$$f(I_{i_1 \cdots i_m}(x,\sigma))$$

Automatic scale selection



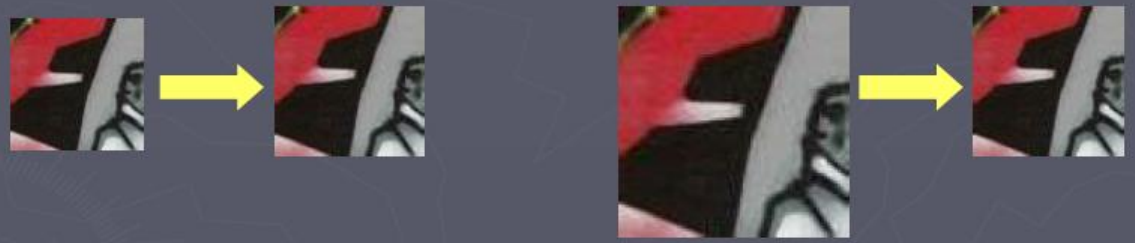
$f(I_{i_1 \ldots i_m}(x,\sigma))$



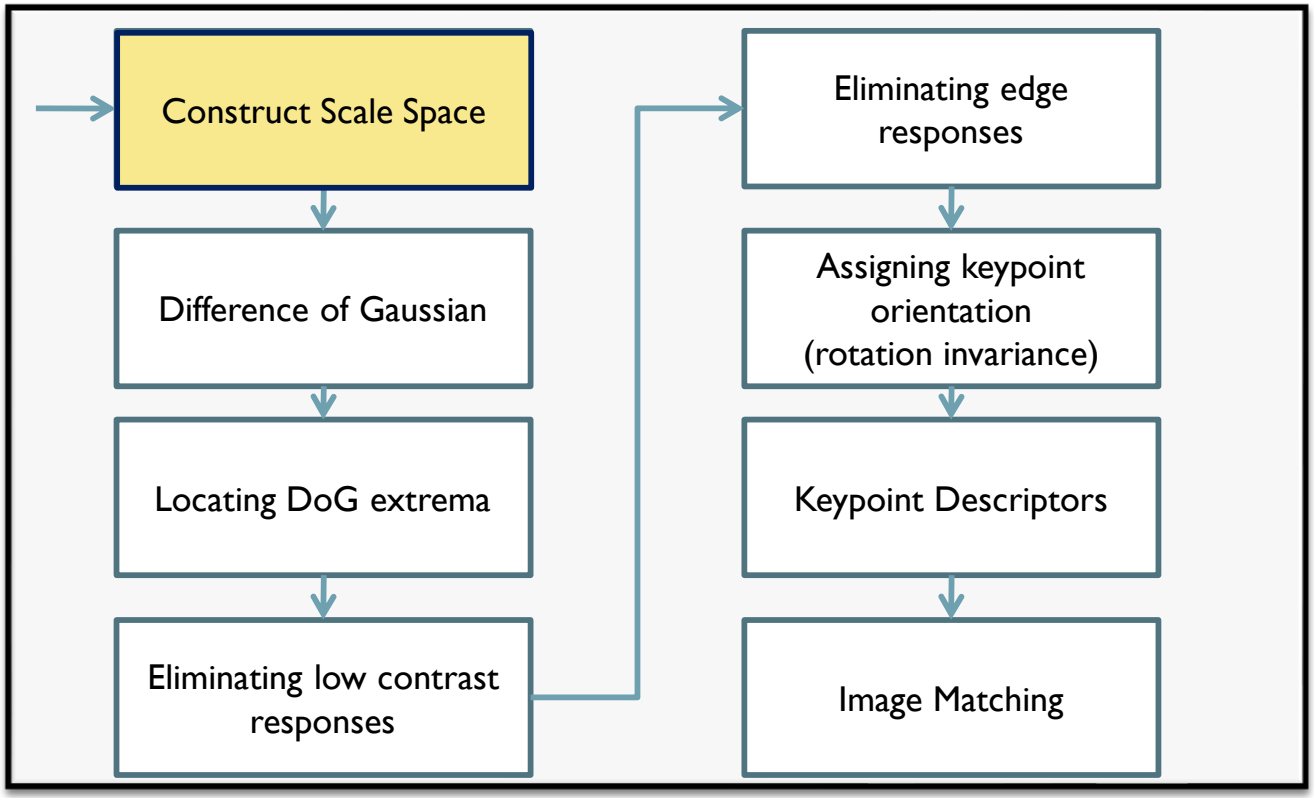
$f(I_{i_1 \ldots i_m}(x',\sigma'))$

Automatic scale selection

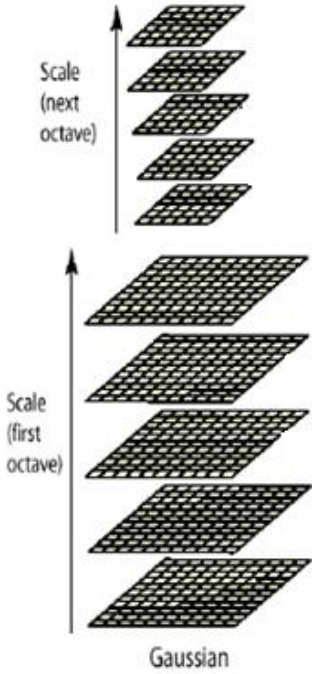
Normalize: rescale to fixed size



SIFT - Workflow

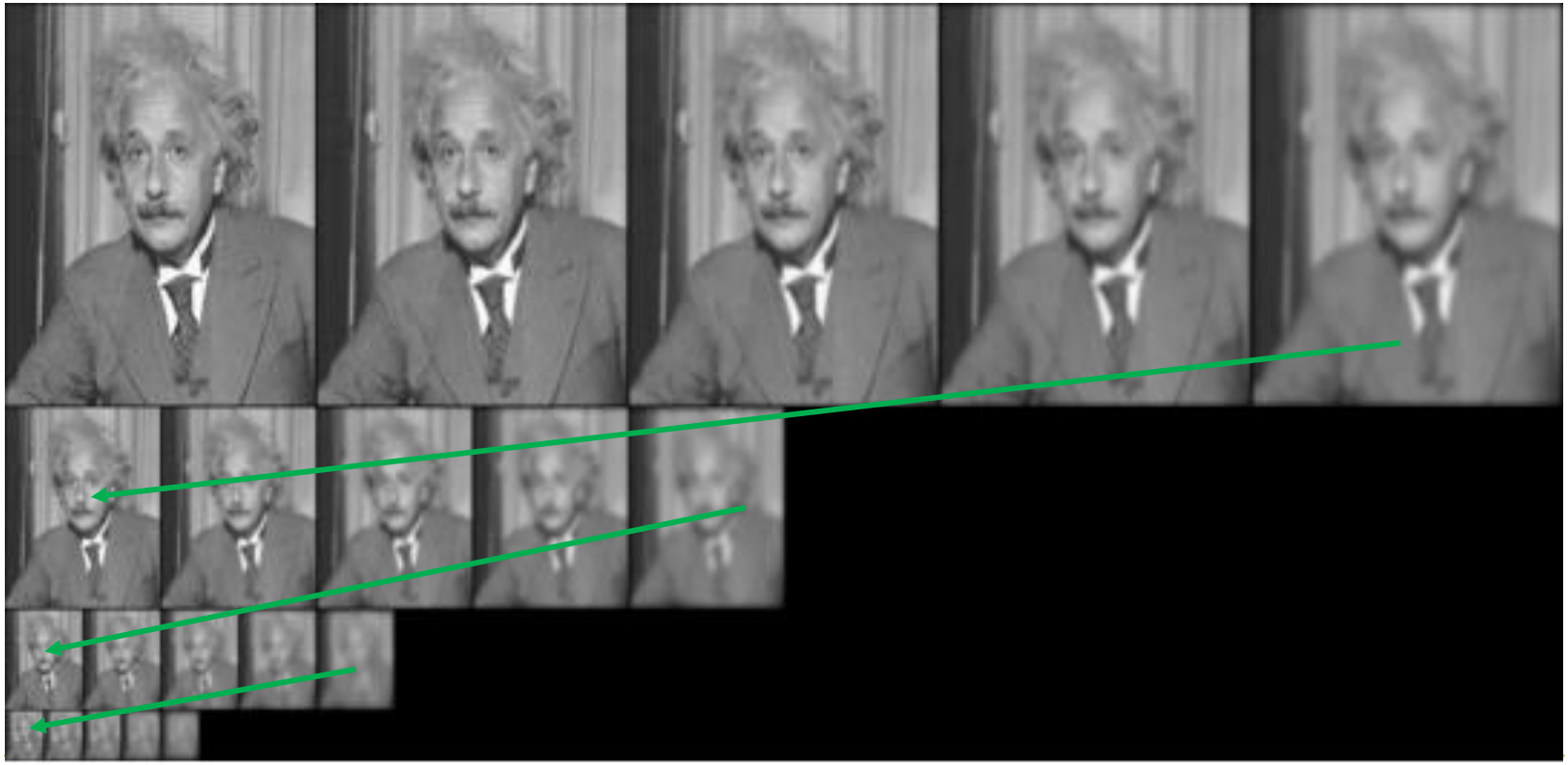


Scale-space Construction

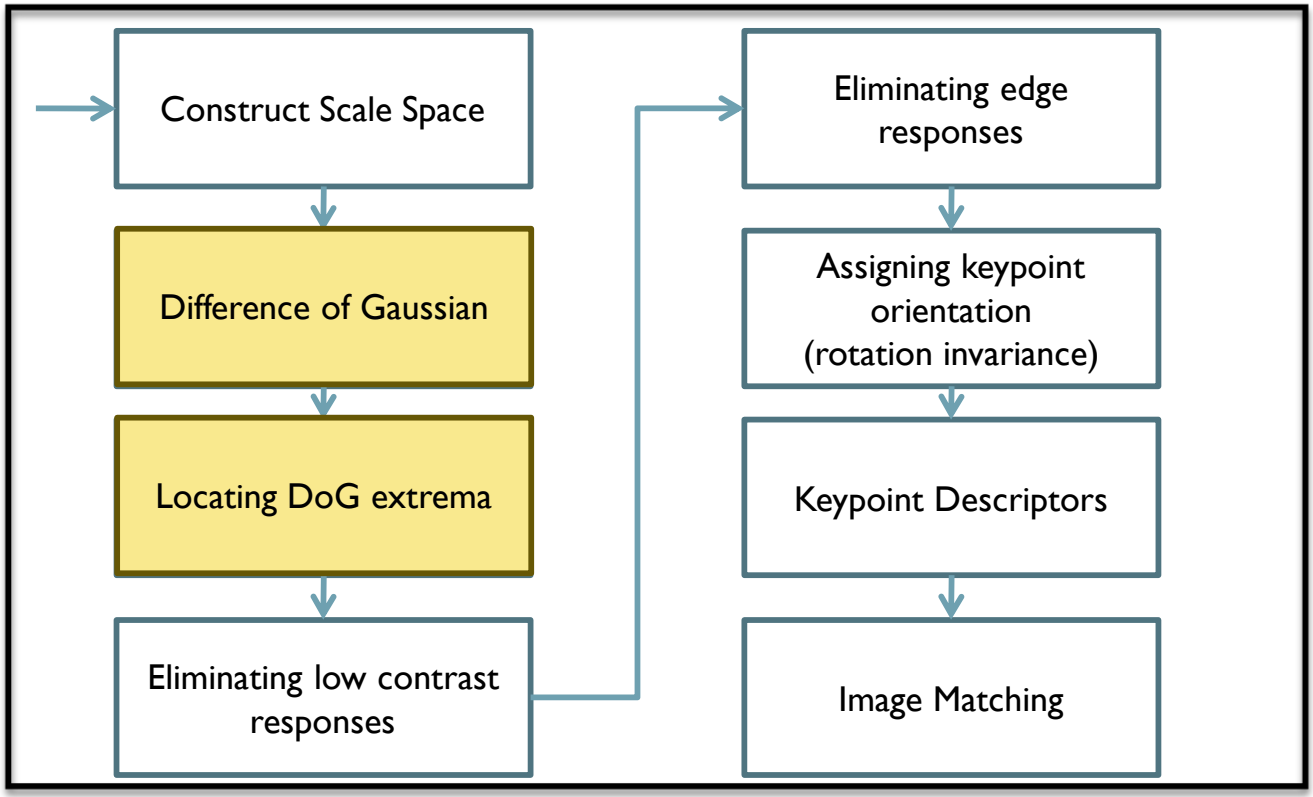


- To achieve scale invariance, image is represented at all scales
- Scale parameter σ is discritized in logarithmically across octaves
- Each octave is made up of S sublevels ,
  - at k^{th} sublevel value of σ is $\sigma_o \times 2^{k/s}$
- After each octave image is downsampled by 2
- To preserve highest frequencies, image is doubled prior to creating first octave

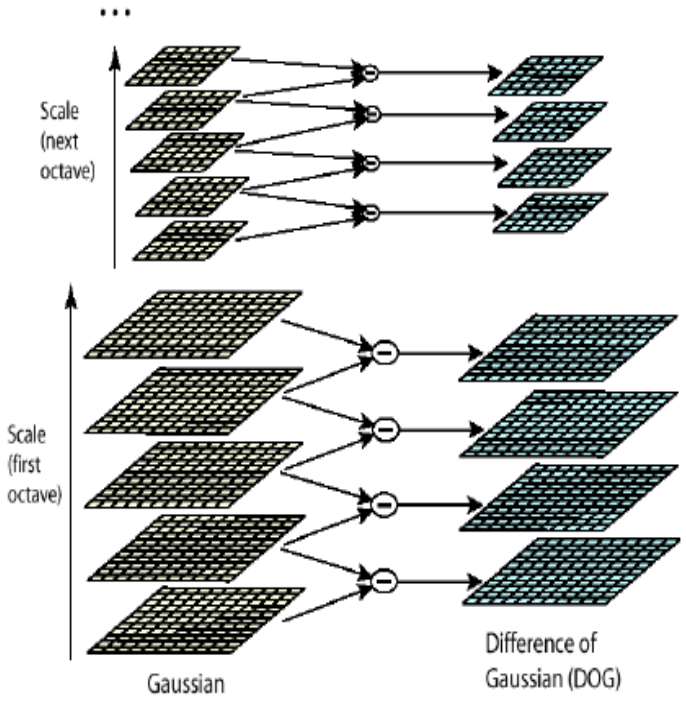
Scale-space Construction



SIFT - Workflow

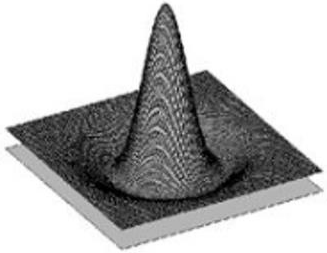


DoG and Extrema Detection



Laplacian ~ Difference of Gaussian

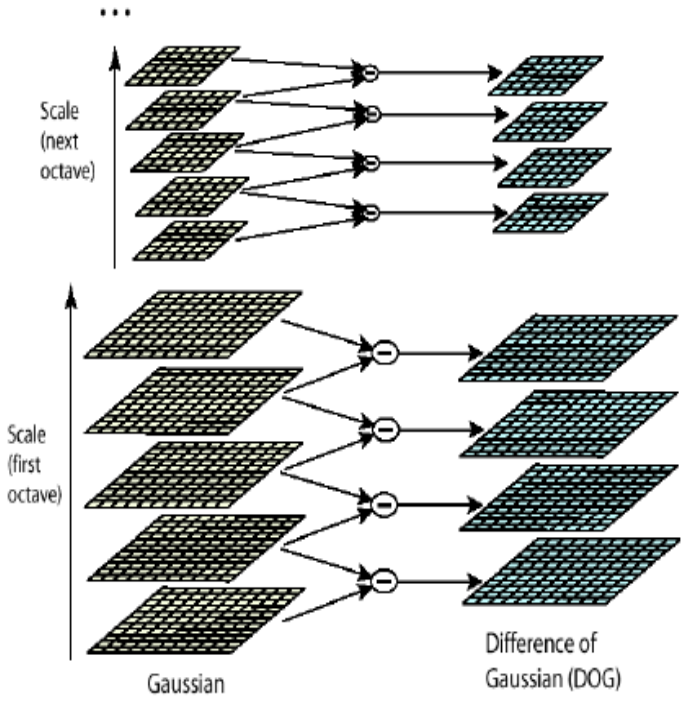
From Lec 15

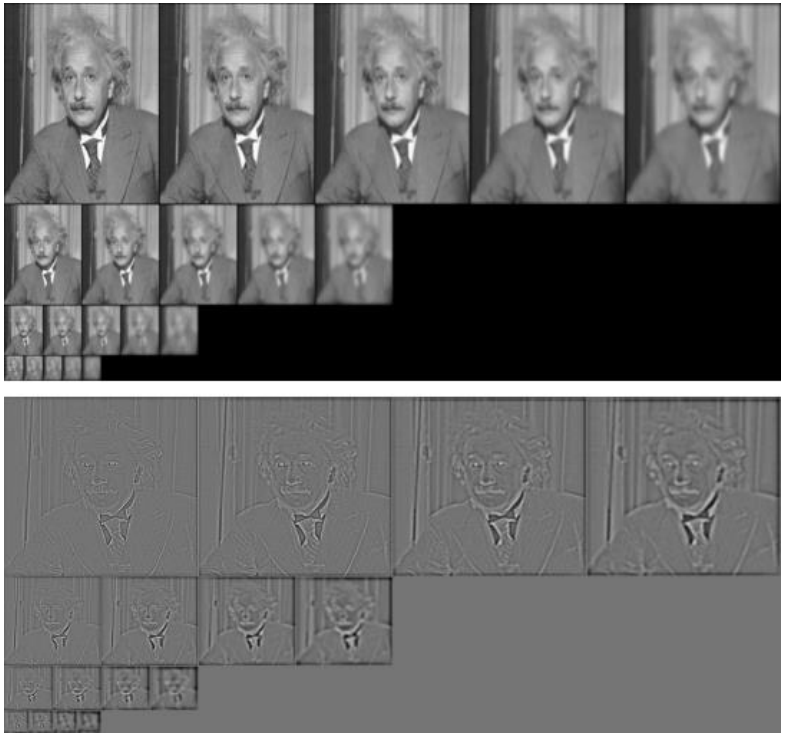


DoG = Difference of Gaussians

Cheap approximation - no derivatives needed.

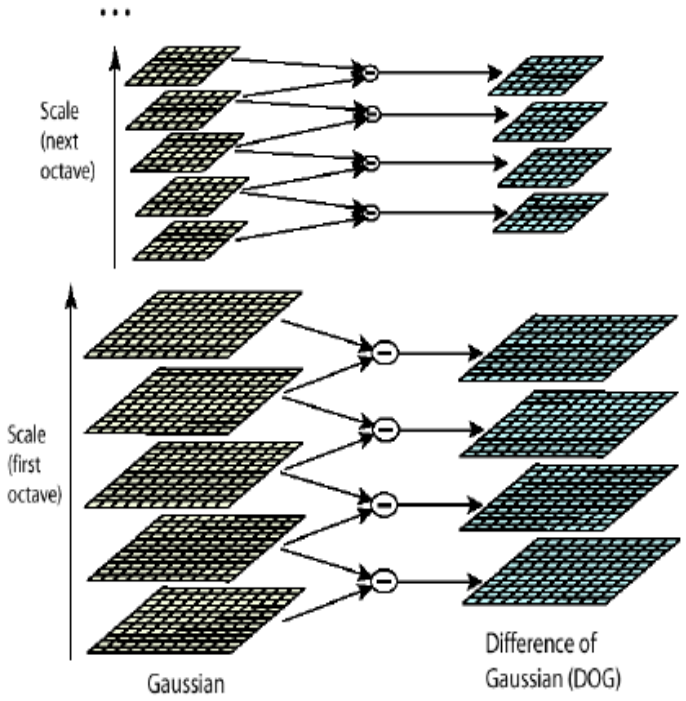
DoG and Extrema Detection

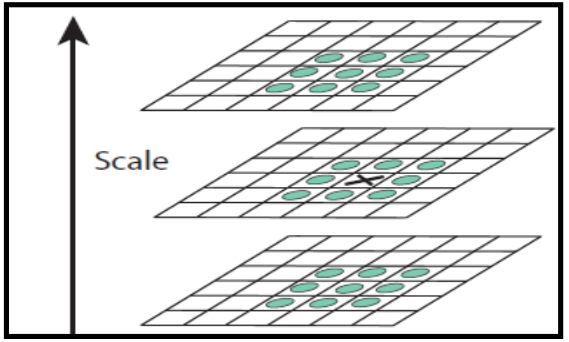




$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

DoG and Extrema Detection



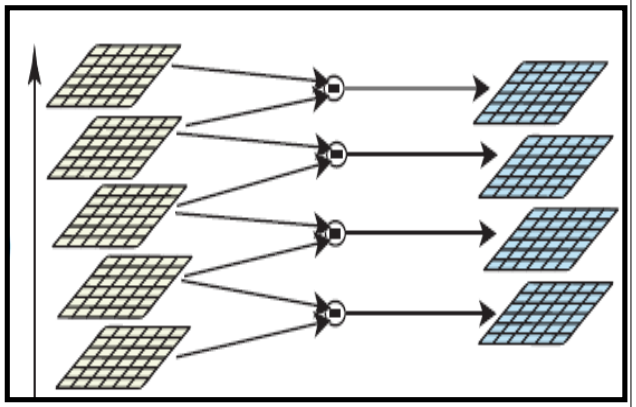


- Maxima and Minima of DoG images are detected by comparing a pixel to its 26 neighbors
- 8 on current scale and (9 + 9) on adjacent scales

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

Detecting Candidate Points

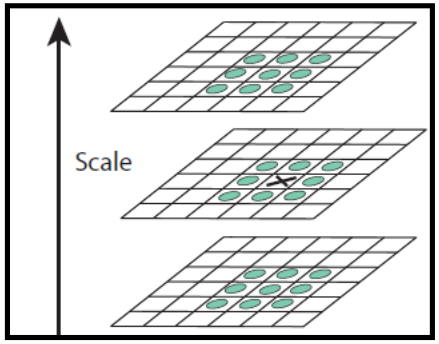
- Difference of Gaussian



- Stable Keypoints are detected in scale space using 'difference of Gaussian'

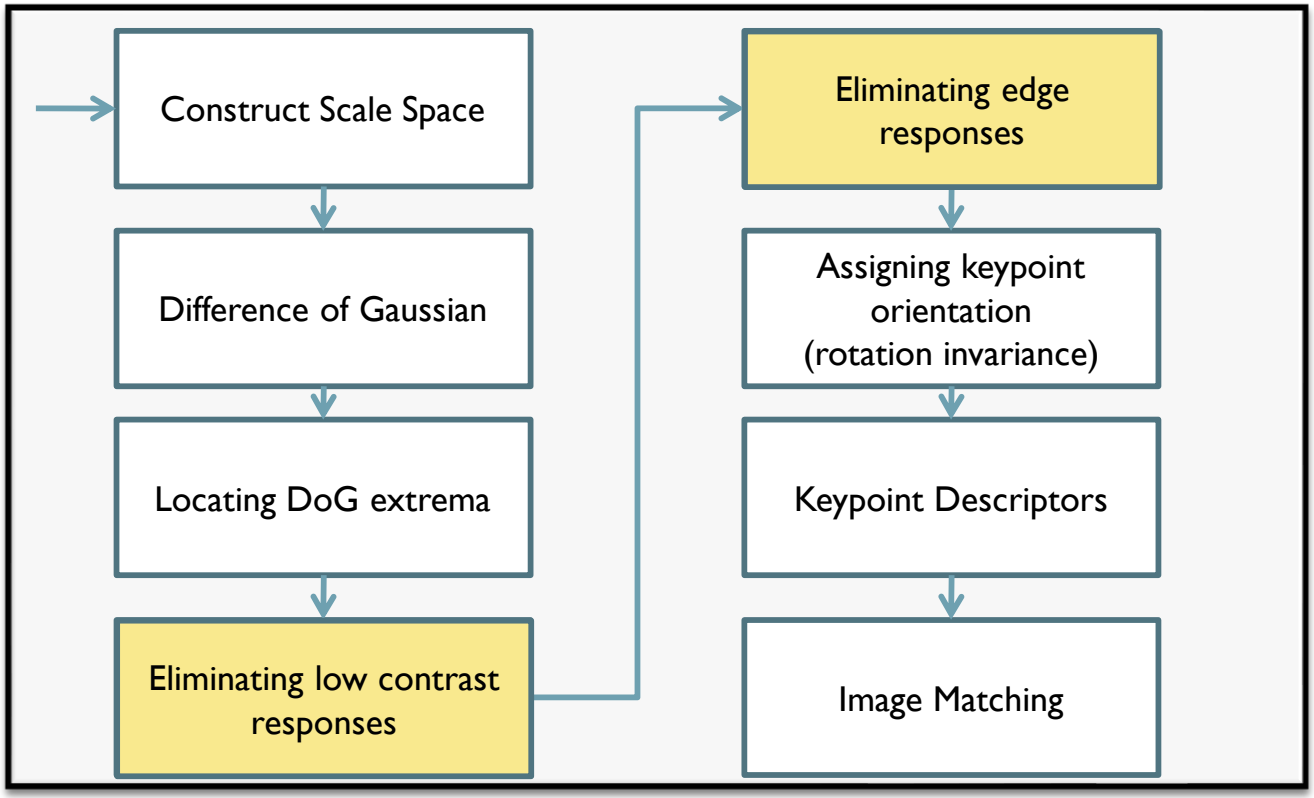
$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

- Detection of local extrema



- Maxima and Minima of DoG images are detected by comparing a pixel to its 26 neighbors
- 8 on current scale and (9 + 9) on adjacent scales

SIFT - Workflow



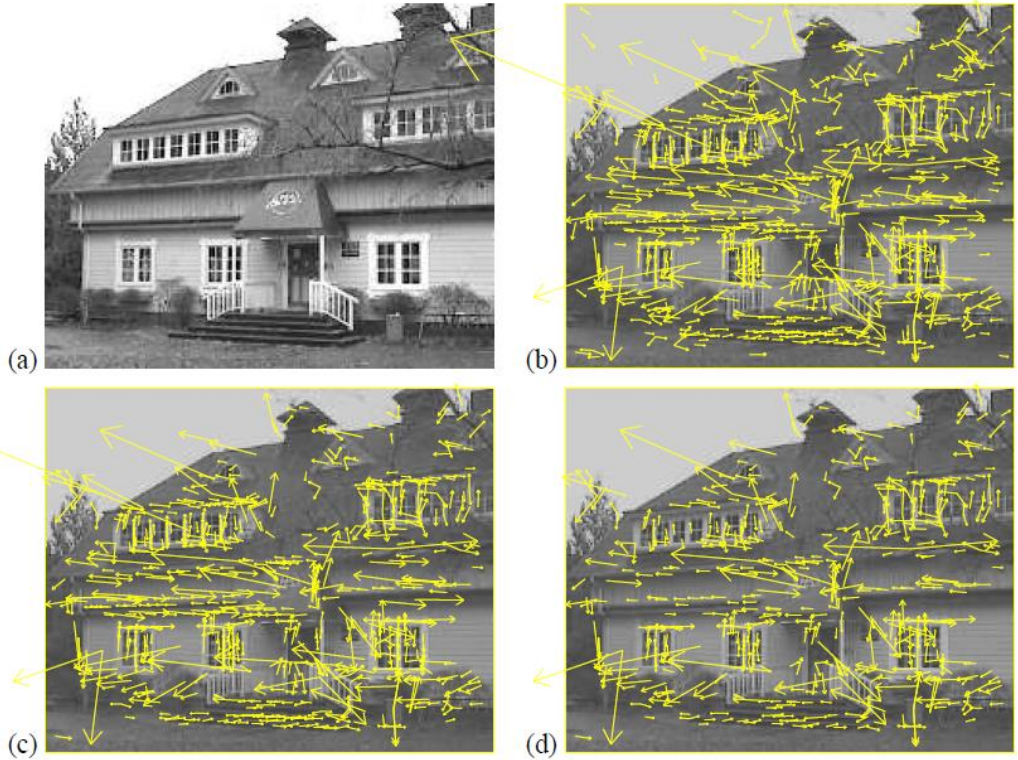
Rejecting Weak Candidate Points …

- Rejecting low contrast candidate points
  - The unstable extrema $D(\underline{x})$ with low contrast is rejected. $C < 0.03$ are discarded.

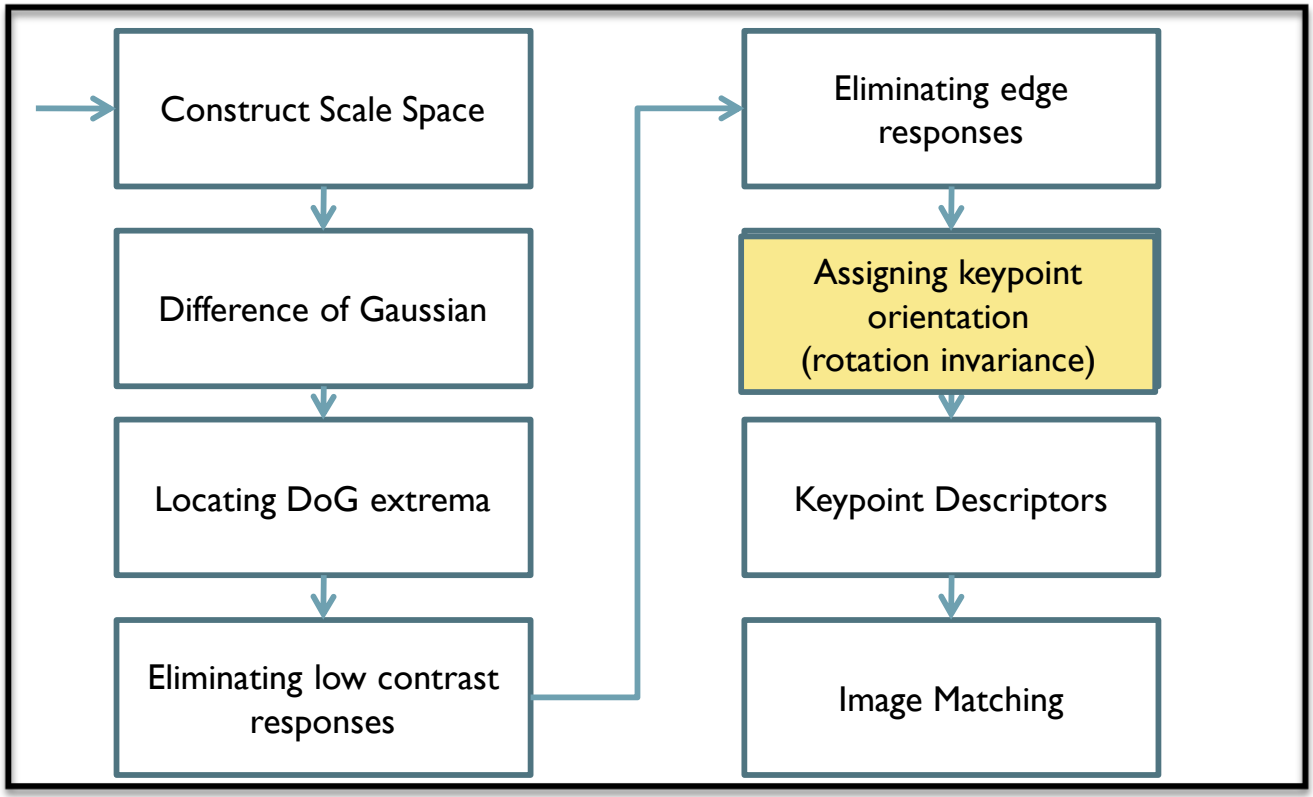
- Eliminating Edge Response
  - Corner-ness Detection using Hessian Matrix in DoG, akin to Harris Corner Operator

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

Rejecting Weak Candidate Points



SIFT - Workflow



Orientation Assignment

- Gradient Magnitude and orientation are calculated at each image point on given keypoint' s scale.

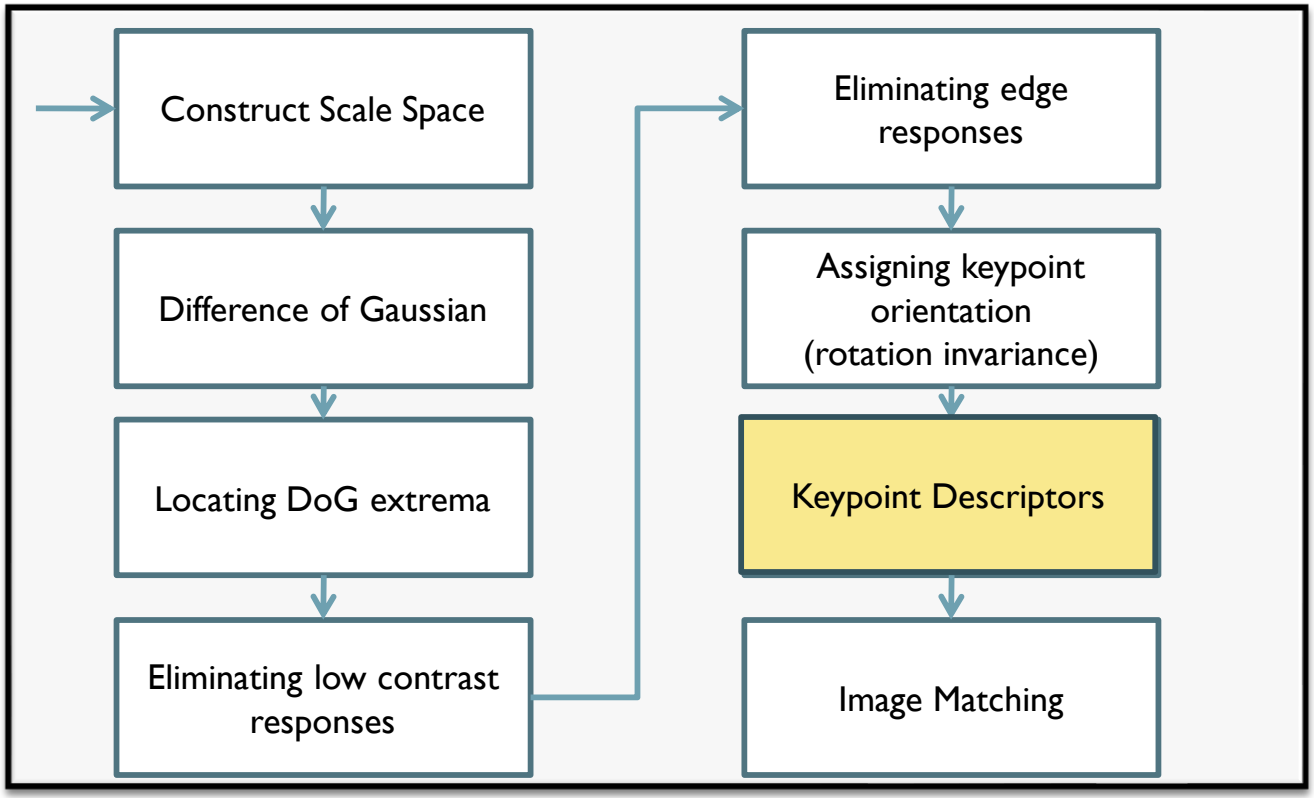
$$m(x,y) = \sqrt{((L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}((L(x+1,y) - L(x-1,y)) / (L(x,y+1) - L(x,y-1)))$$

- An orientation histogram of 36 bins covering 360° is formed from the gradient orientations of points in a region around keypoint

- Peaks in the histogram correspond to dominant directions of local gradients, these directions are assigned to keypoints

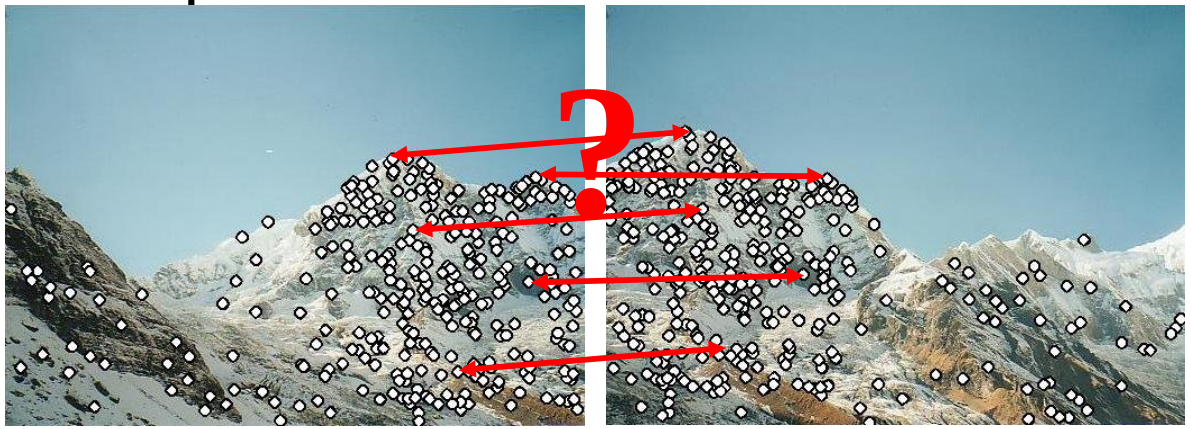
SIFT - Workflow



SIFT Feature descriptors

We know how to detect good points

Next question: **How to match them?**



Local Image Descriptor

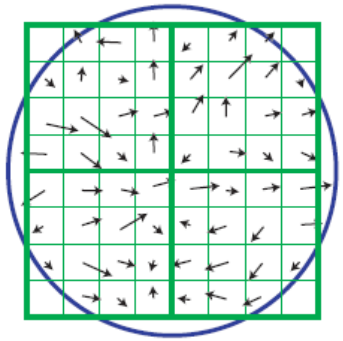
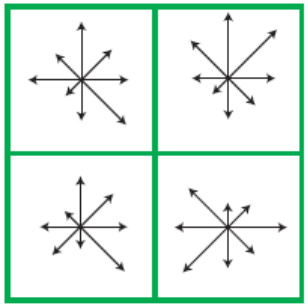
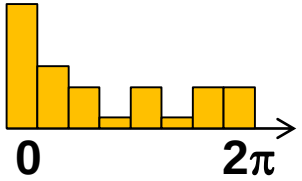


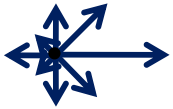
Image gradients



Keypoint descriptor



angle histogram



- Orientation information are accumulated into orientation histograms (of 8 bins) summarizing the content over 4x4 sub regions (shown for 2x2)
- This gives rise to 4x4x8 sized feature vector – 128.

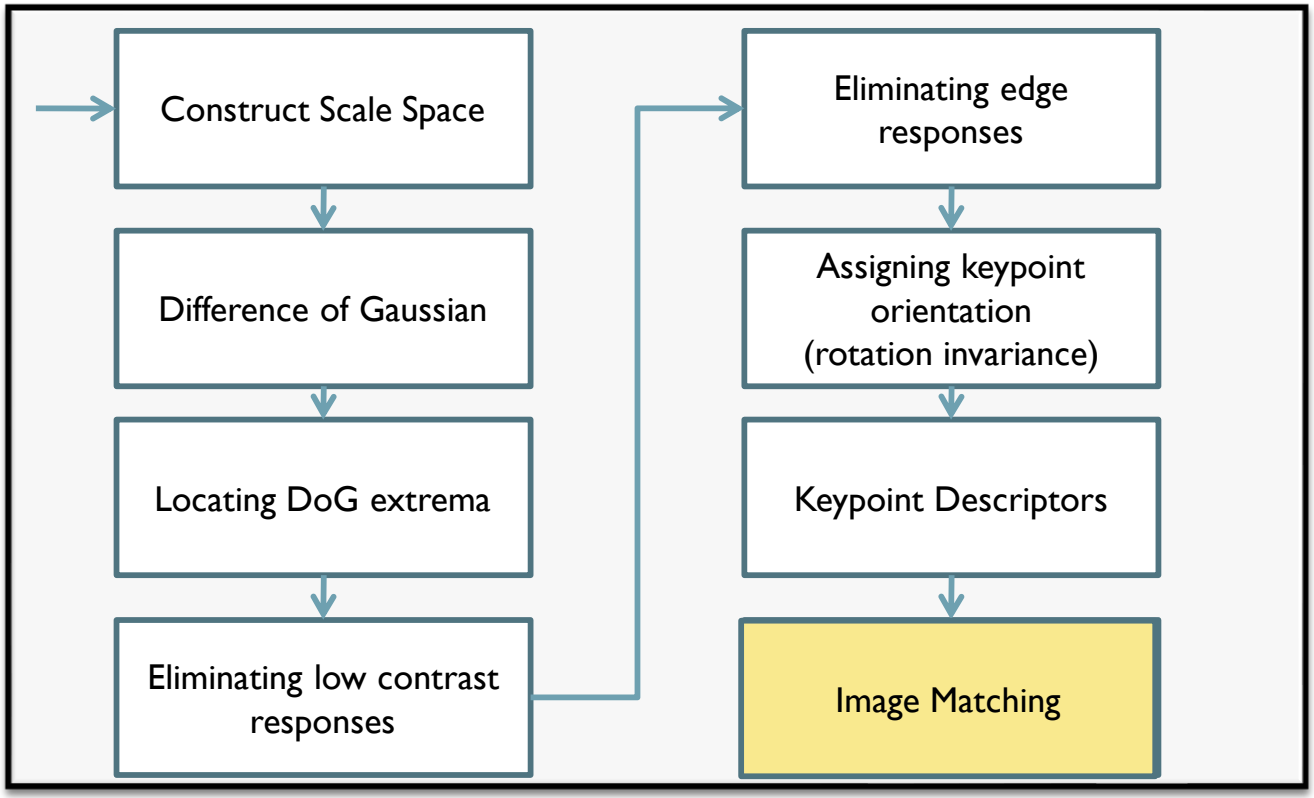
Illumination Invariance

- Feature vector is normalized to unit length to achieve invariance to affine changes in illumination

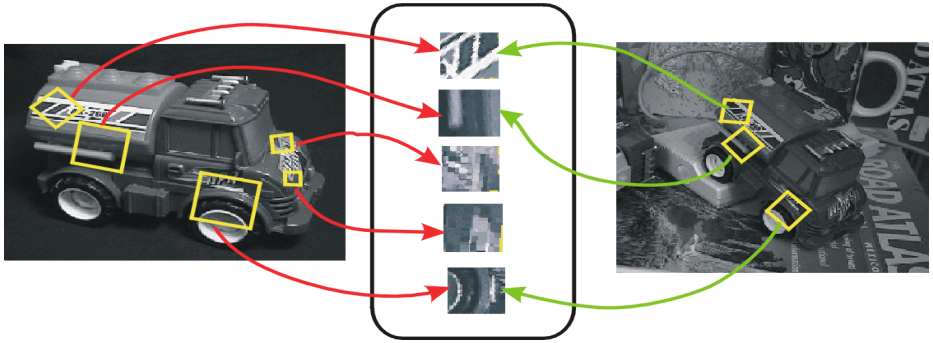
  - Contrast change – Each Pixel Value is multiplied by a constant and hence gradients are also multiplied by same constant which gets cancelled by vector normalization

  - Brightness change – A constant is added or subtracted globally from all pixel values, which does not affect gradient

SIFT - Workflow

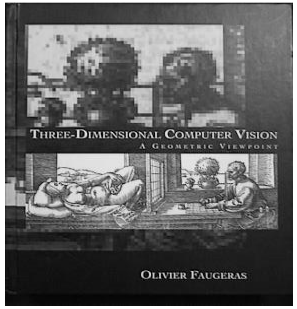


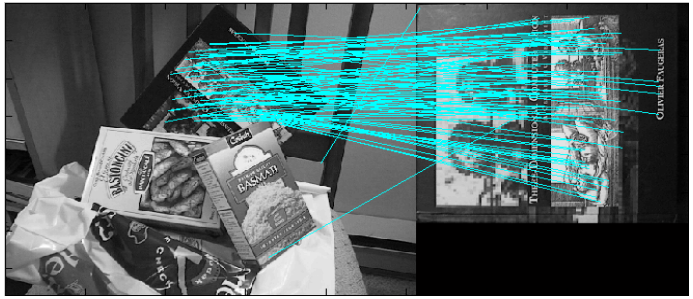
Keypoint Matching

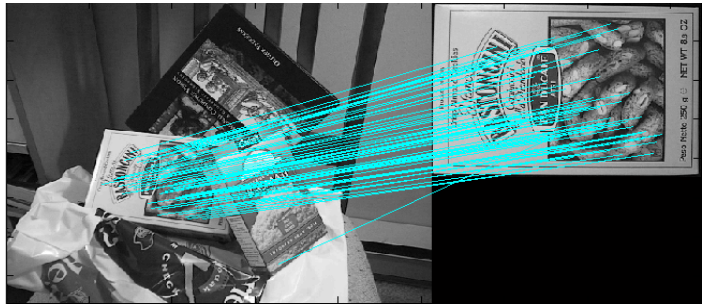


- The best candidate match for each keypoint is found by identifying its Nearest Neighbor in the database of keypoints.
- A global threshold on the distance to the closest feature does not perform well to discard features not having good match
- A comparison of the distance of the closest neighbor with the second-closest neighbor makes a good threshold parameter.

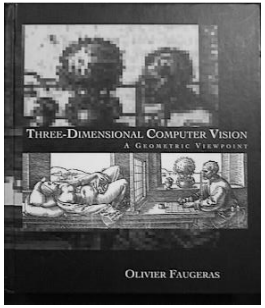
David Lowe's Implementation

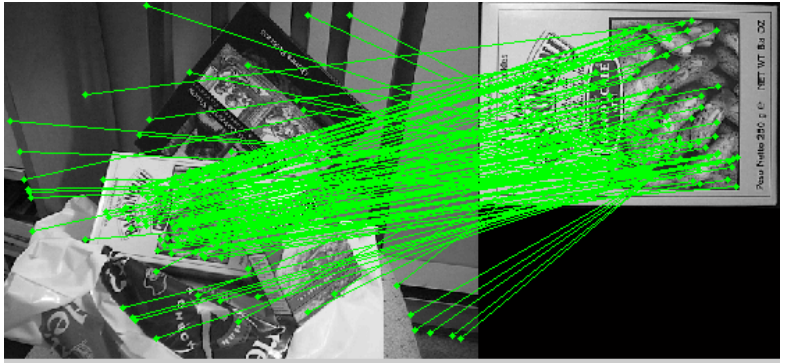


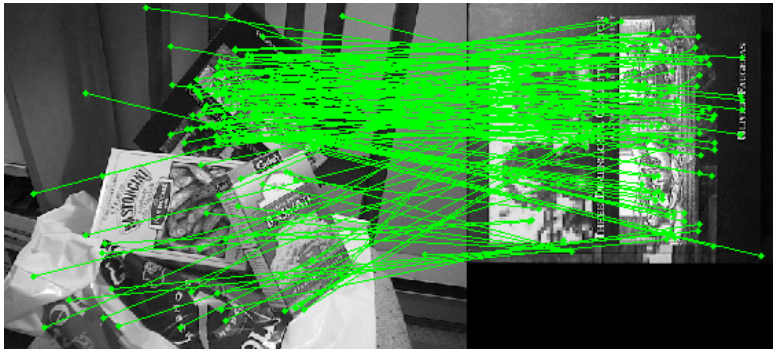




A. Vedaldi's Implementation







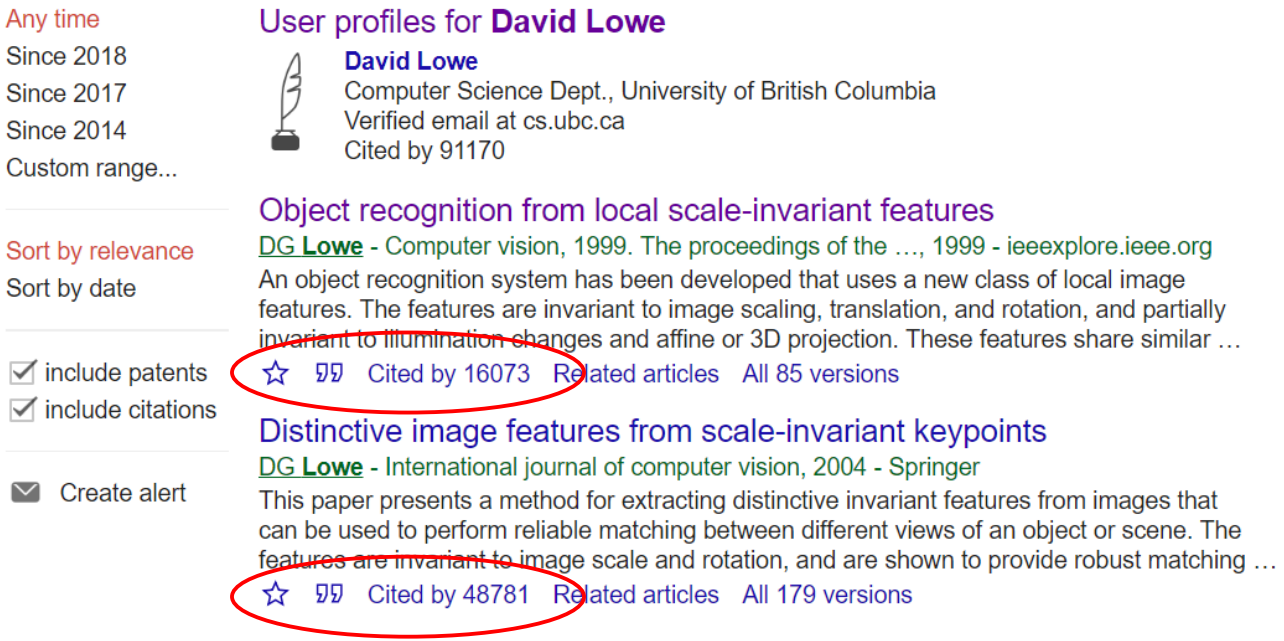
SIFT Paper

Object recognition from local scale-invariant features, ICCV 1999

Distinctive Image Features from Scale-Invariant Keypoints , IJCV 2004

SIFT is commonly used as a benchmark against which other vision methods are compared. The original SIFT research paper by author David Lowe was initially rejected several times for publication by the major computer vision journals, and as a result Lowe filed for a patent and took a different direction. According to Lowe, "By then I had decided the computer vision community was not interested, so I applied for a patent and intended to promote it just for industrial applications."[1] Eventually, the SIFT paper was published

SIFT Paper



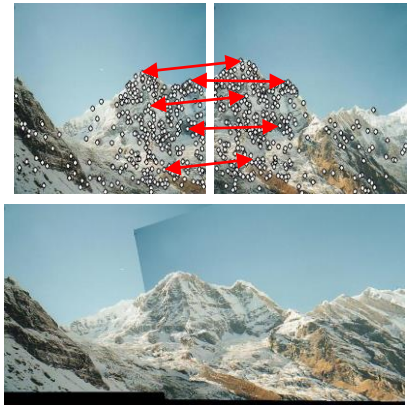
- Wins ICCV test of time award in 2011

Local vs. Global Features

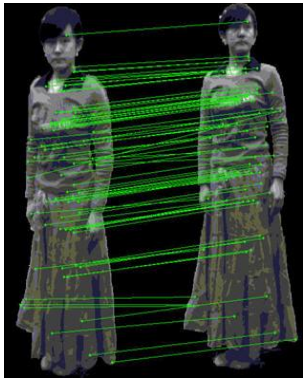
Object Recognition

Scene Classification

Image Registration



3D Reconstruction



Landmark Recognition

Global Features

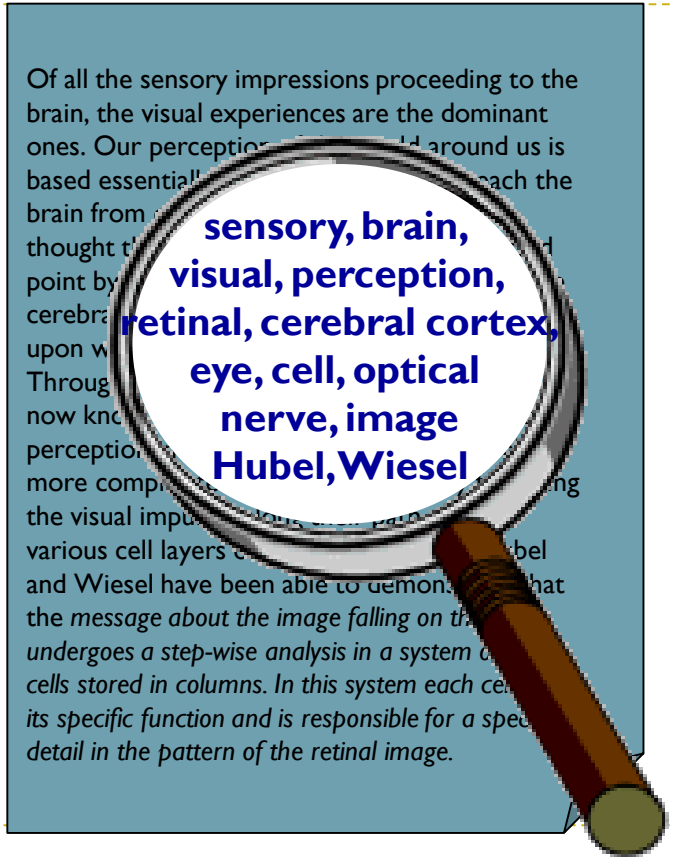
Bag of Visual Words (BoW)

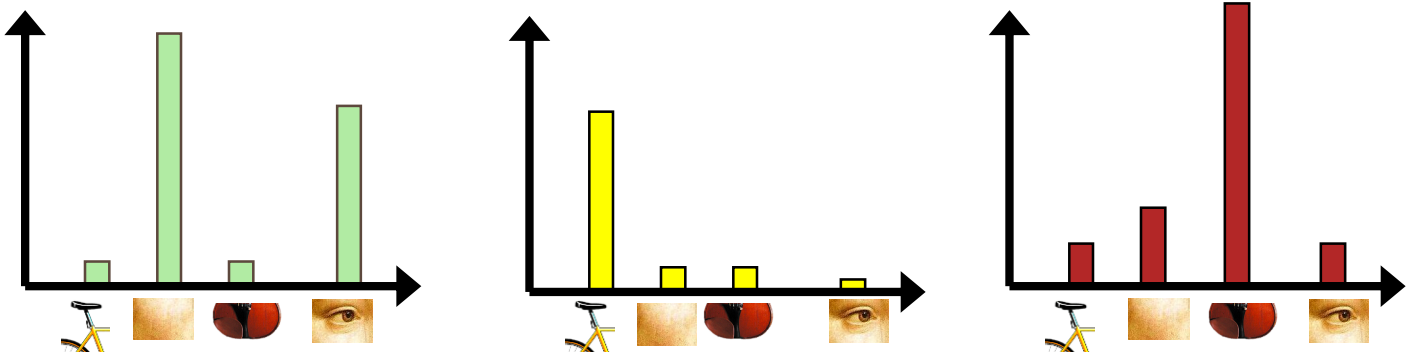
Video Google: A Text Retrieval Approach to Object Matching in Videos, ICCV '03

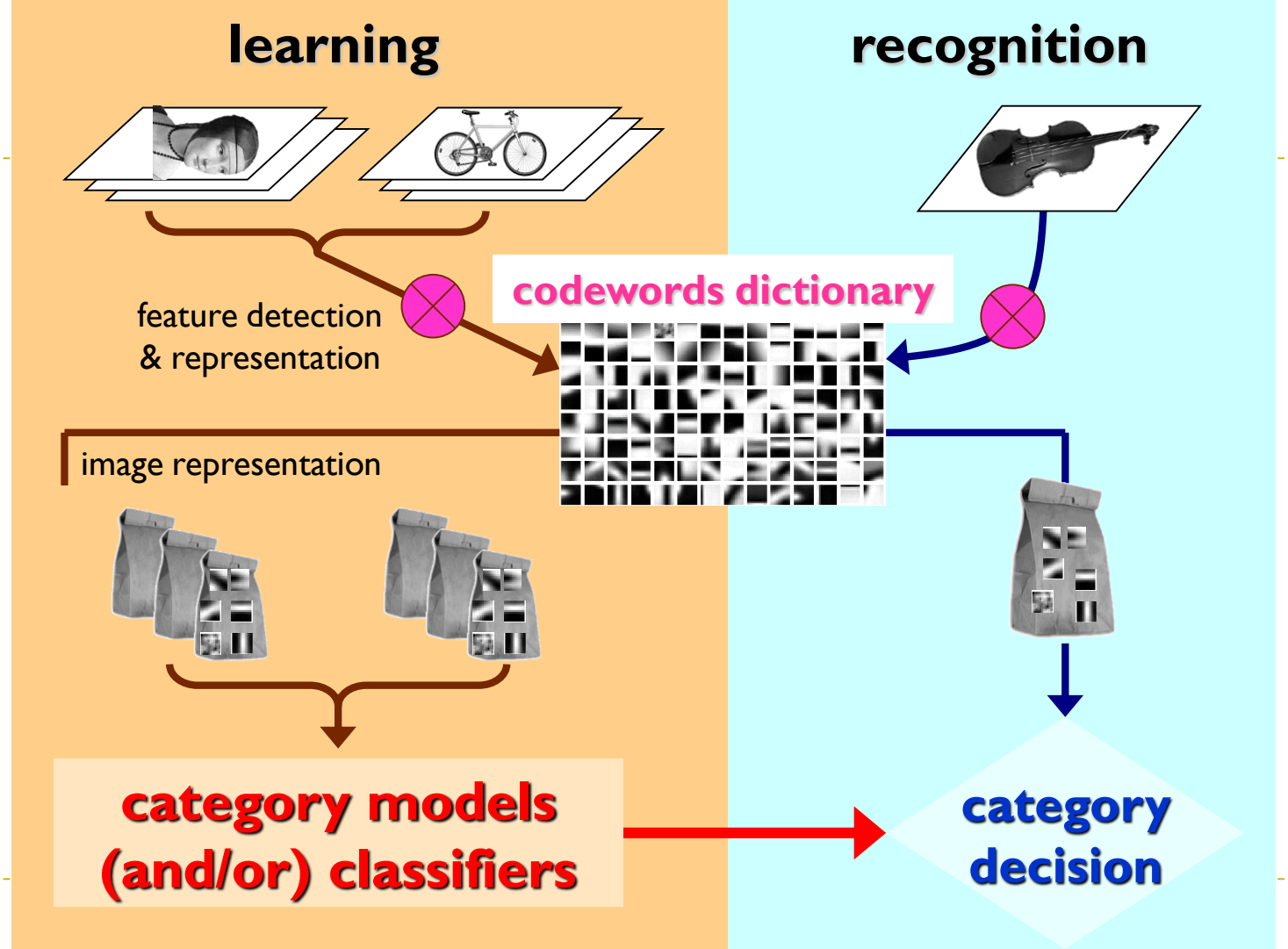
http://www.robots.ox.ac.uk/~vgg/research/vgoogle/

Object → Bag of 'words'

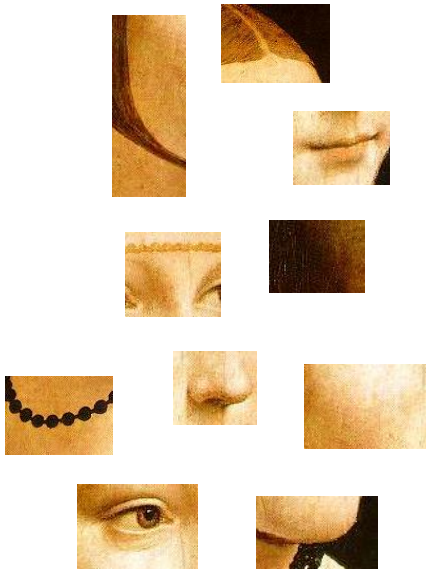
Analogy to documents






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

1.Feature detection and representation



Feature detection

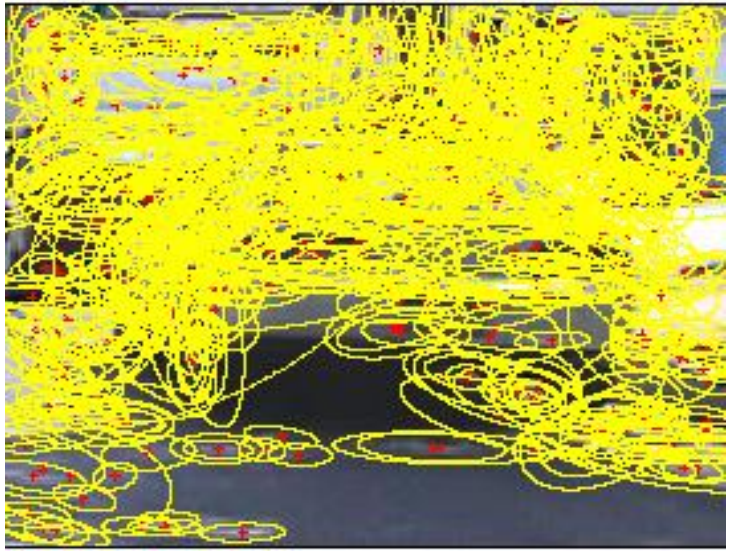
- Sliding Window
  - Leung et al, 1999
  - Viola et al, 1999
  - Renninger et al 2002

Feature detection

- Sliding Window
  - Leung et al, 1999
  - Viola et al, 1999
  - Renninger et al 2002
- Regular grid
  - Vogel et al. 2003
  - Fei-Fei et al. 2005

Feature detection

- Sliding Window
  - Leung et al, 1999
  - Viola et al, 1999
  - Renninger et al 2002
- Regular grid
  - Vogel et al. 2003
  - Fei-Fei et al. 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei et al. 2005
  - Sivic et al. 2005



Feature Representation

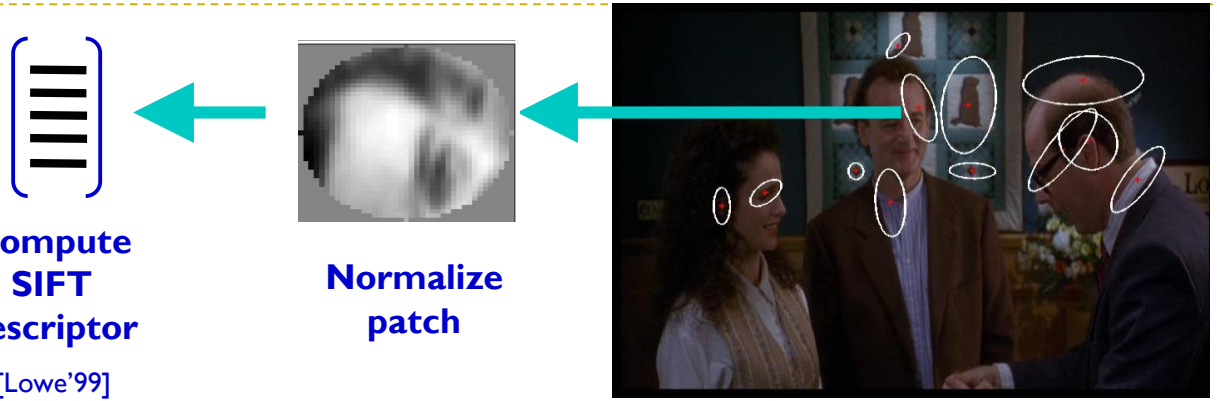
Visual words, aka textons, aka keypoints:

K-means clustered pieces of the image

- Various Representations:
  - Filter bank responses
  - Image Patches
  - SIFT descriptors

All encode more-or-less the same thing…

Interest Point Features



Compute SIFT descriptor

[Lowe'99]

Normalize patch

Detect patches

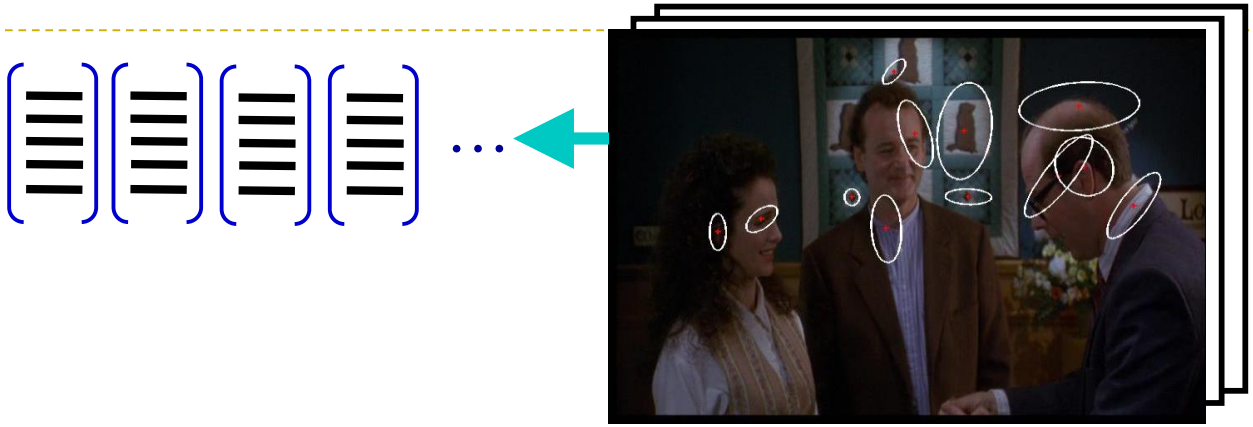
[Mikojaczyk and Schmid '02]

[Matas et al. '02]

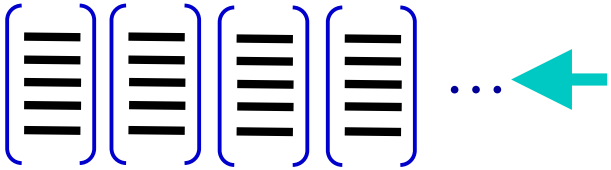
[Sivic et al. '03]

Slide credit: Josef Sivic

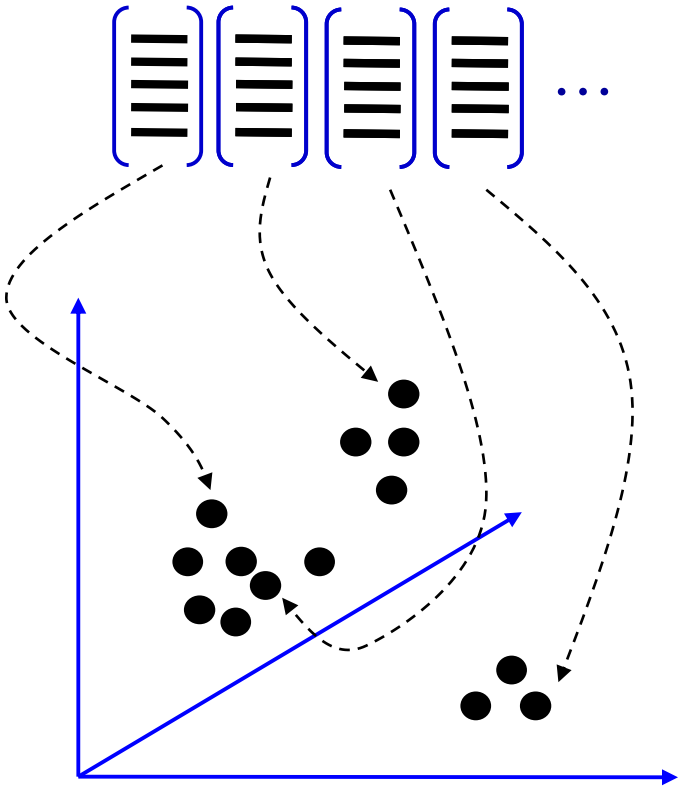
Interest Point Features



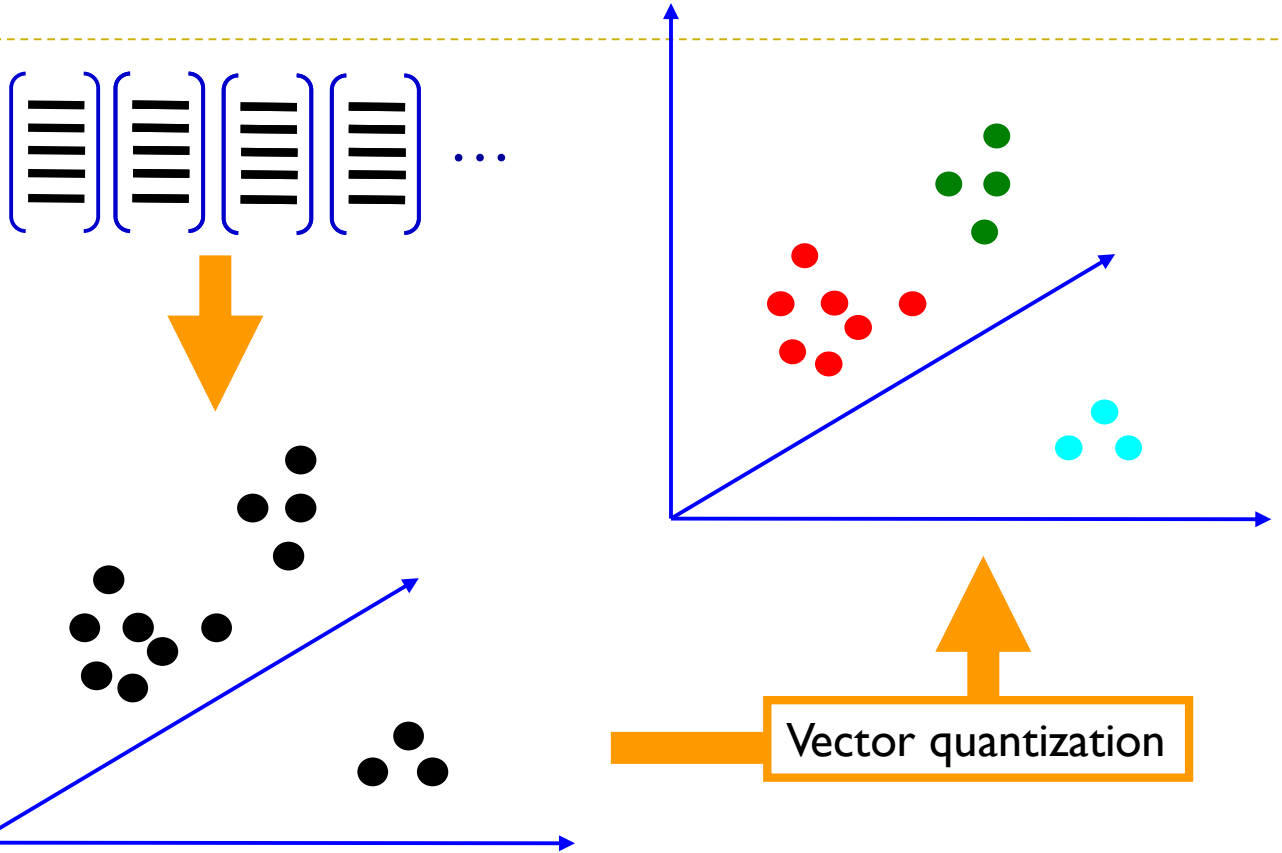
Patch Features



dictionary formation

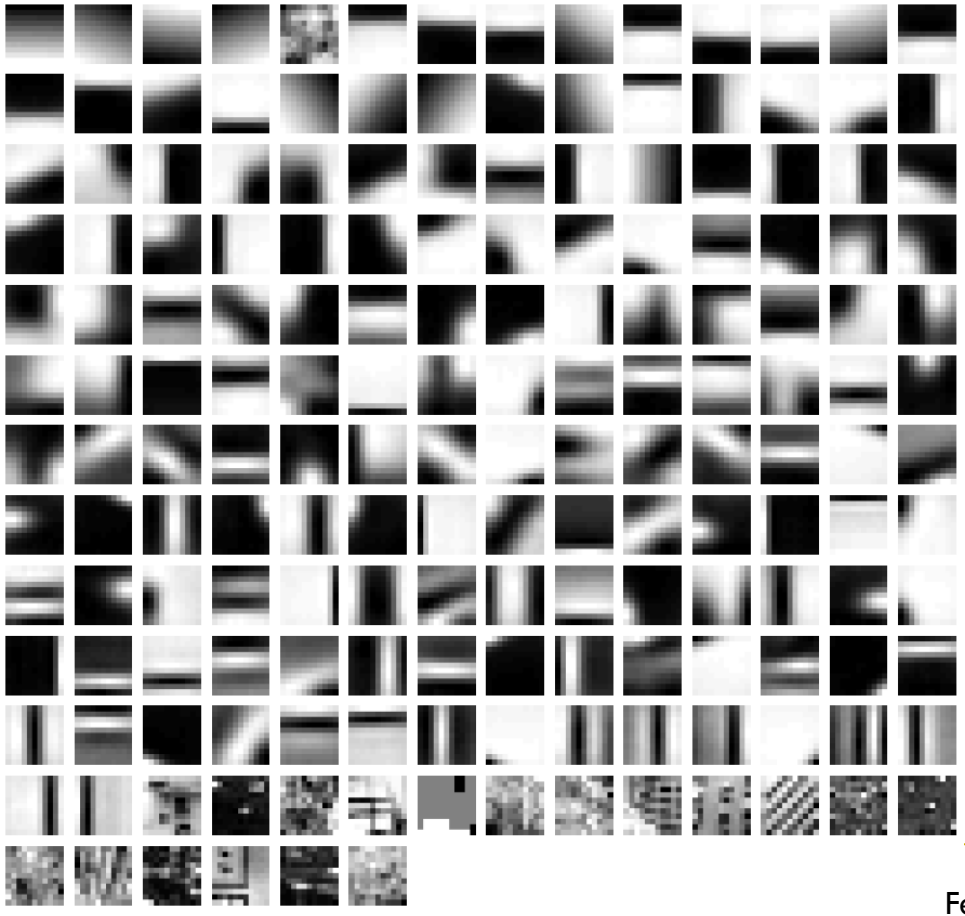


Clustering (usually k-means)



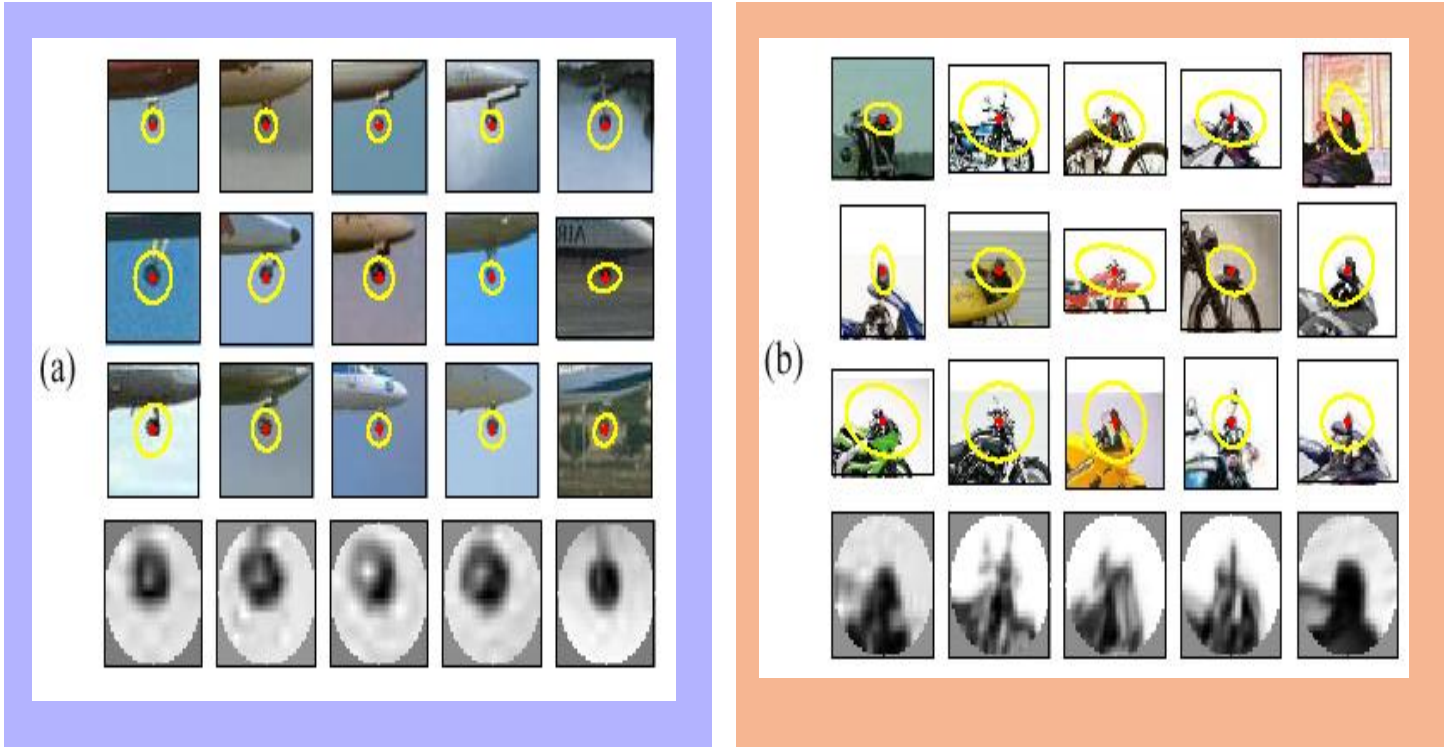
Slide credit: Josef Sivic

Clustered Image Patches



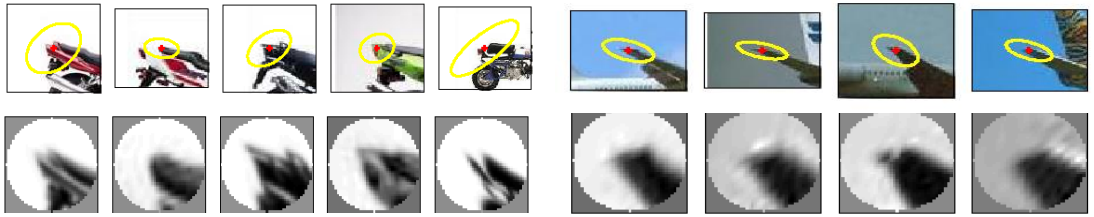
Fei-Fei et al. 2005

Image patch examples of codewords

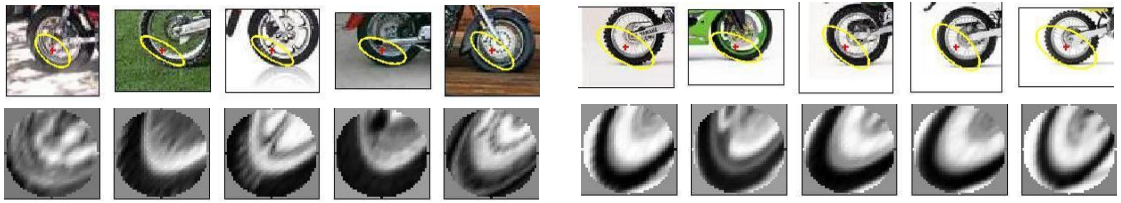


Sivic et al. 2005

Visual synonyms and polysemy

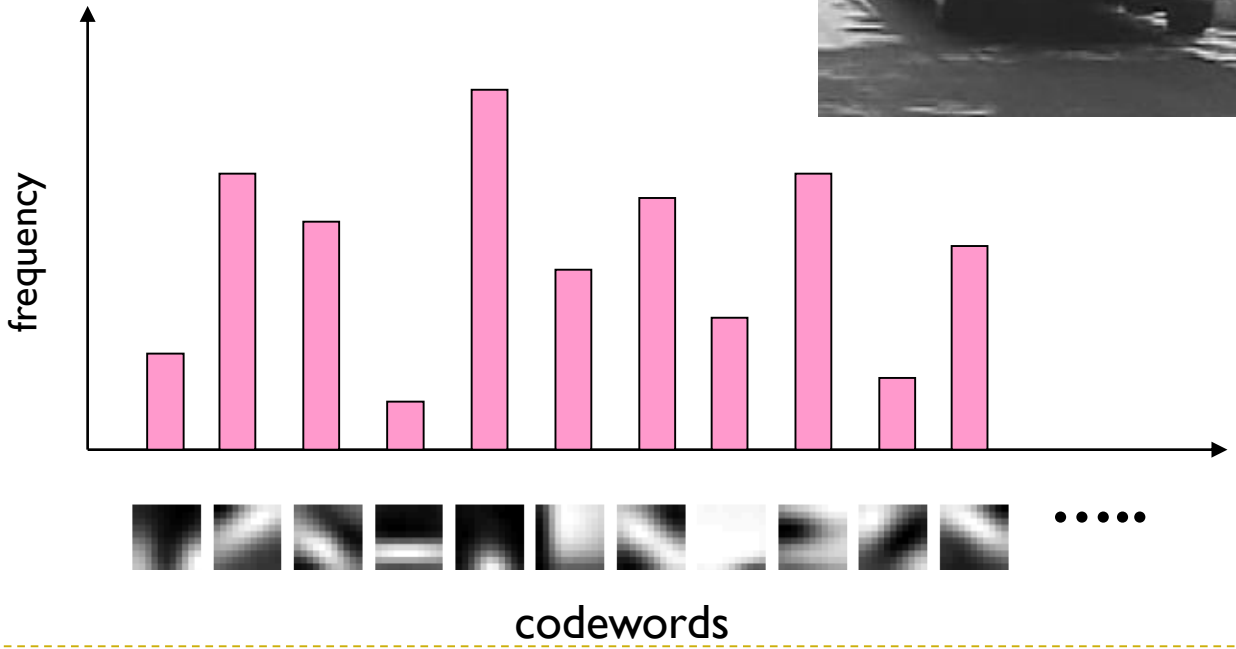


Visual Polysemy. Single visual word occurring on different (but locally similar) parts on different object categories.



Visual Synonyms. Two different visual words representing a similar part of an object (wheel of a motorbike).

Image representation



Scene Classification (Renninger & Malik)






Feature Selection

- Which feature to use for which task?
- Can I select features automatically?
- Can I design features automatically?