

1100110011100100110100001100100111
0001000001100000001101100111101011
100110110100000111001001111111111
1101110100001110011001101000010100
0001101011001001101000111000110100
101011010000110111001110110001001
0111000000111101000101000000100111
110001101111101000111101101000000
0100100000000101001100000110001100
1000111100000000010000101110101000
0100110100100100100010111000010110
0110011010010010011000011101001111
1000001010000011011110000111011010
001000010111011111010111011110011
1111100010000010010011110011010000
0101000111001110110010110011110000
1011100000010000101110001111101010
110001101110000011011110001111101010
1000110100100100100010111000010110
1010101110000011011110001111101010
1000111010100100100010111000011110
0011011011000011011110001111101010
1101111100000011011110001111101010
010101100000011011110001111101010
0100100100011011110001111101010
1101111100000011011110001111101010
1101111100000011011110001111101010

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,...

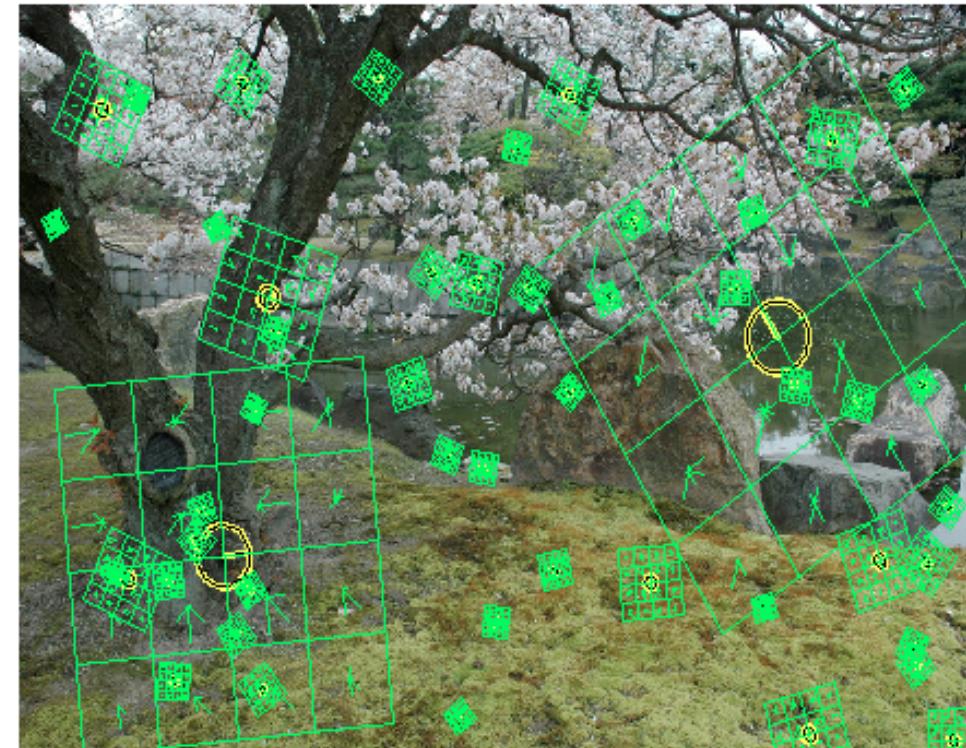


- ▶ Disadvantages
 - ▶ Cannot deal with occlusions, clutter, viewpoint changes.



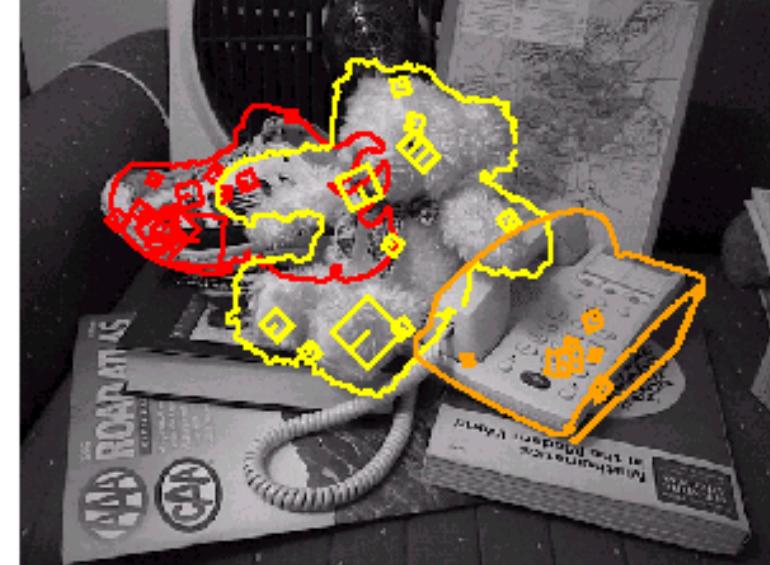
Image Representation: Local

- ▶ 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



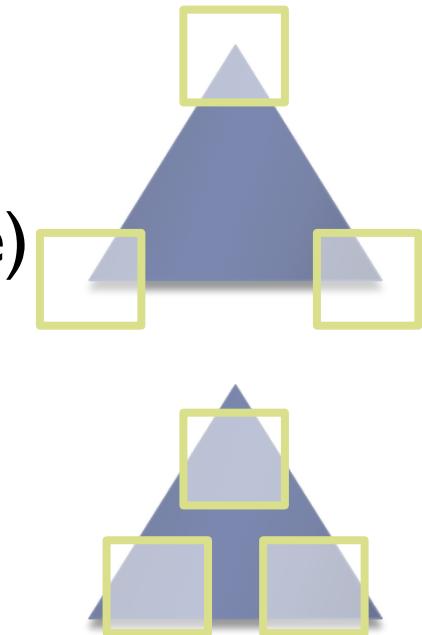
A Good Local Feature

- ▶ 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

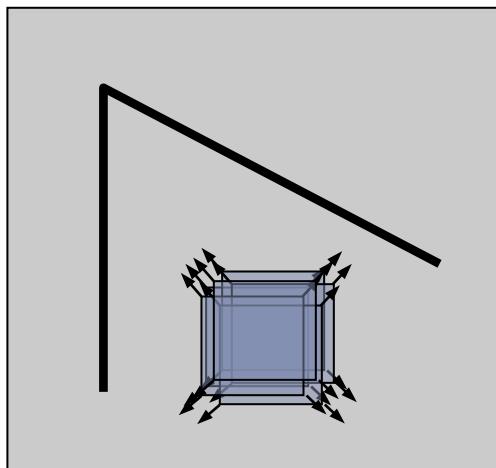
- ▶ We focus on interest points as local features
- ▶ Interest-point detector
 - ▶ Points on corners
 - ▶ Harris corners (first-order derivative)
 - ▶ Points on blob-like structures
 - ▶ SIFT (second-order derivative)
 - ▶ Interest-point descriptor
 - ▶ Local description of the neighborhood
 - ▶ Histogram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



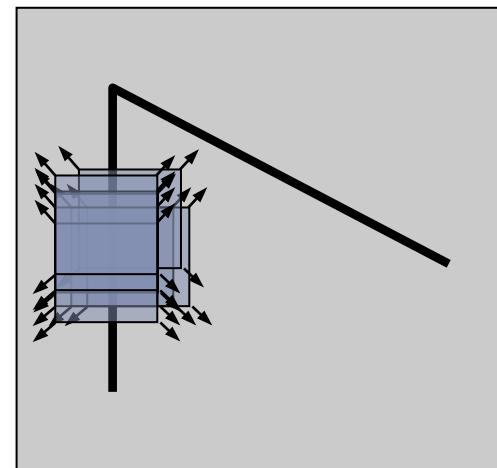
Harris-Corner Detector

▶ Intuition

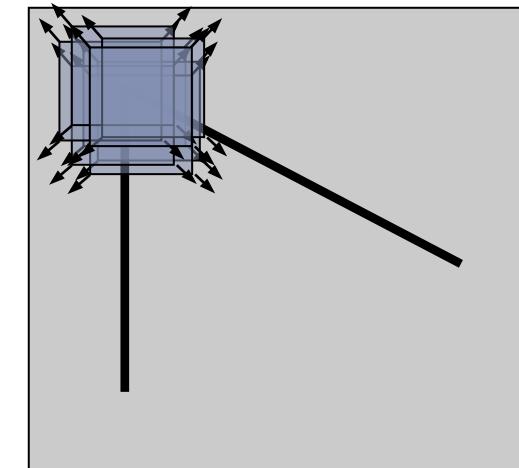
- ▶ Find points that different from their neighborhood



“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

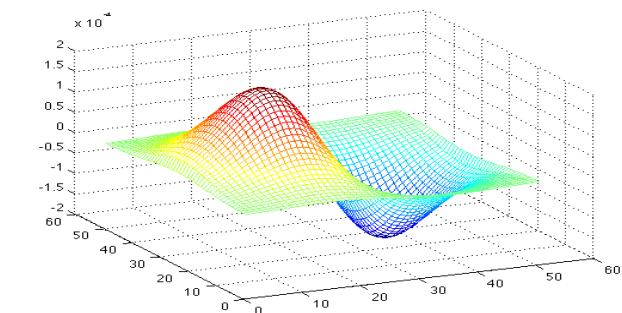
Smoothing

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

First-order derivatives

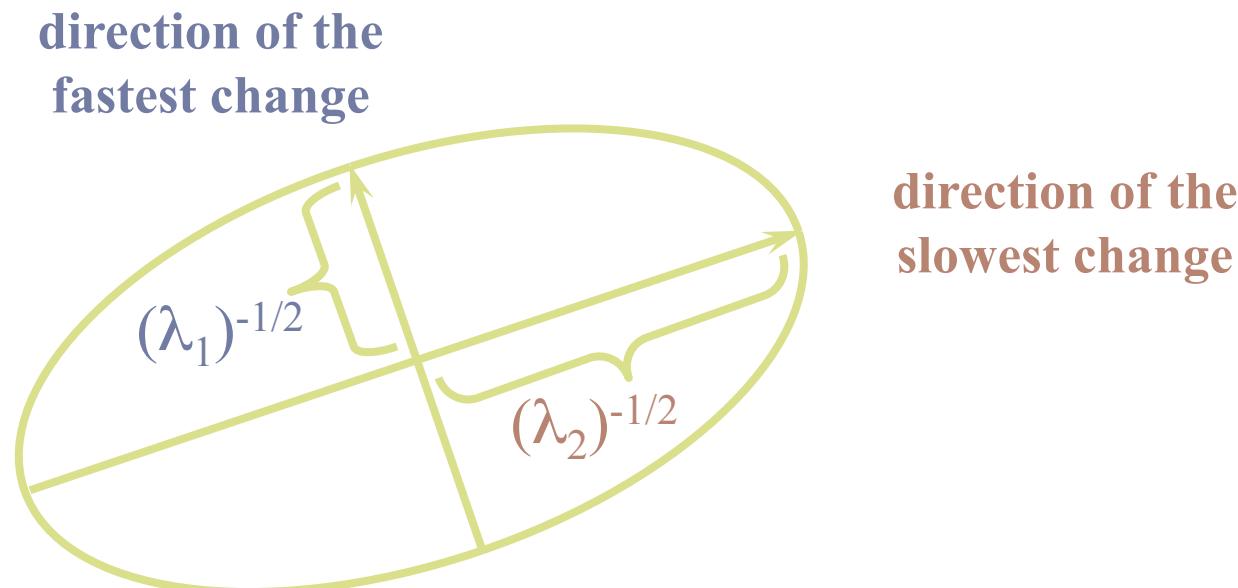
Derivatives computed with Gaussian kernels of scale σ_D .

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



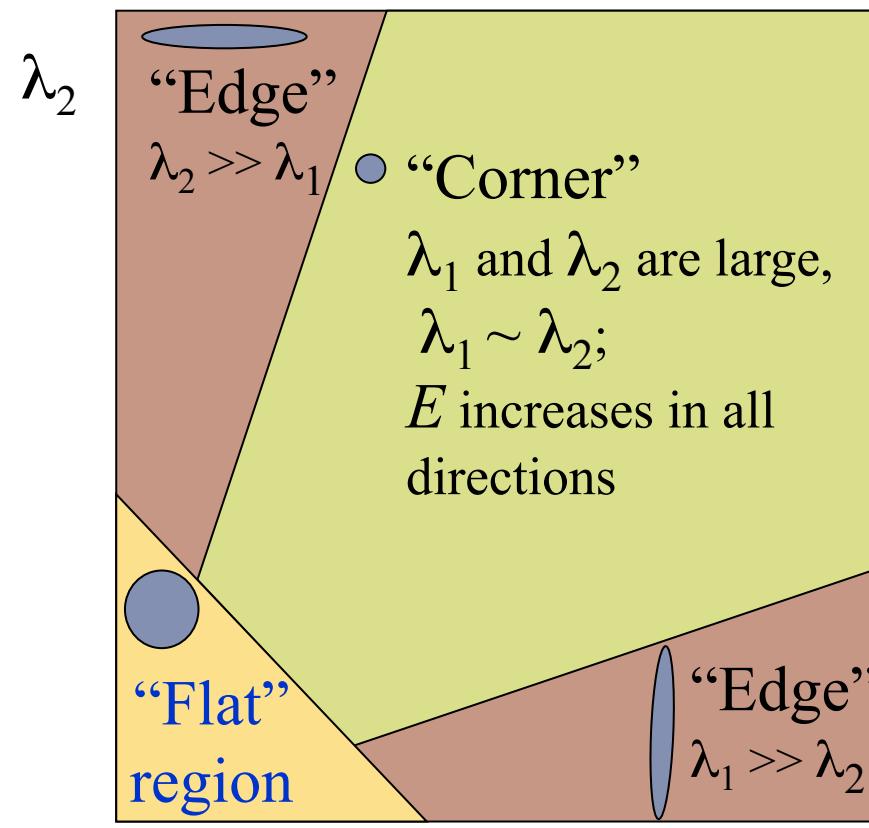
Eigenvalues

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



Eigenvalues

▶ Classification of image points



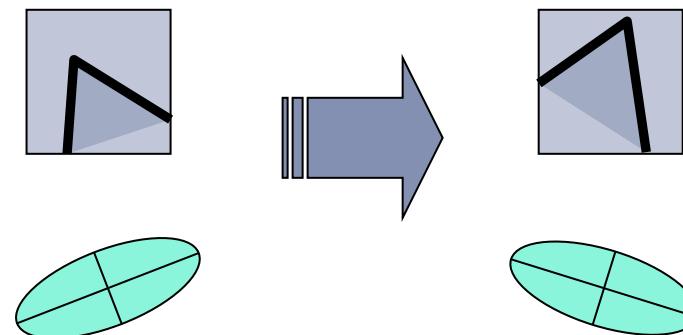
Determinant and Trace

- ▶ 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:
 - ▶ $\text{Det}(M) = ad - bc$
 - ▶ $\text{Trace}(M) = a + d$
 - ▶ $R = \det(M) - \kappa * \text{trace}^2(M)$
- ▶ Finding local maxima in the image

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

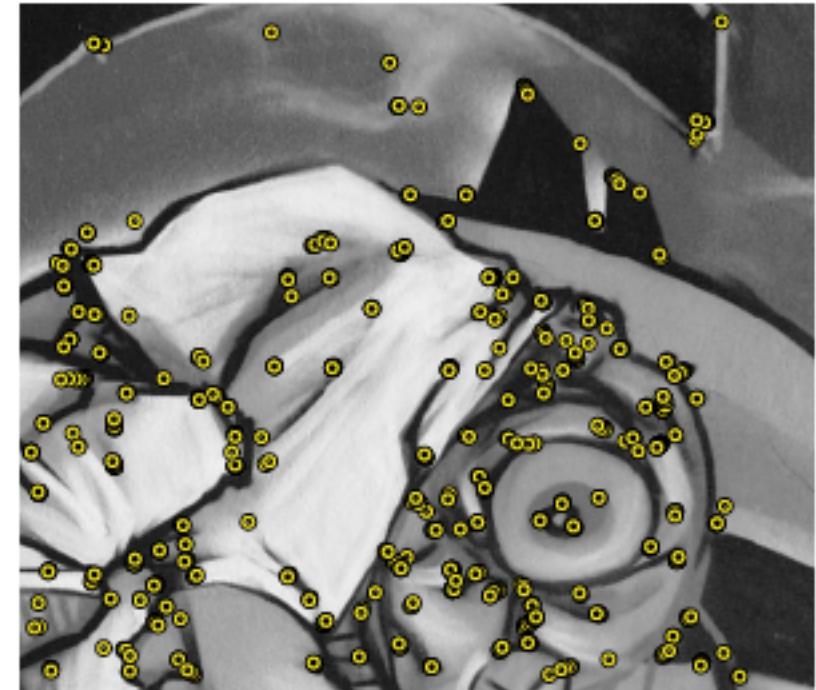
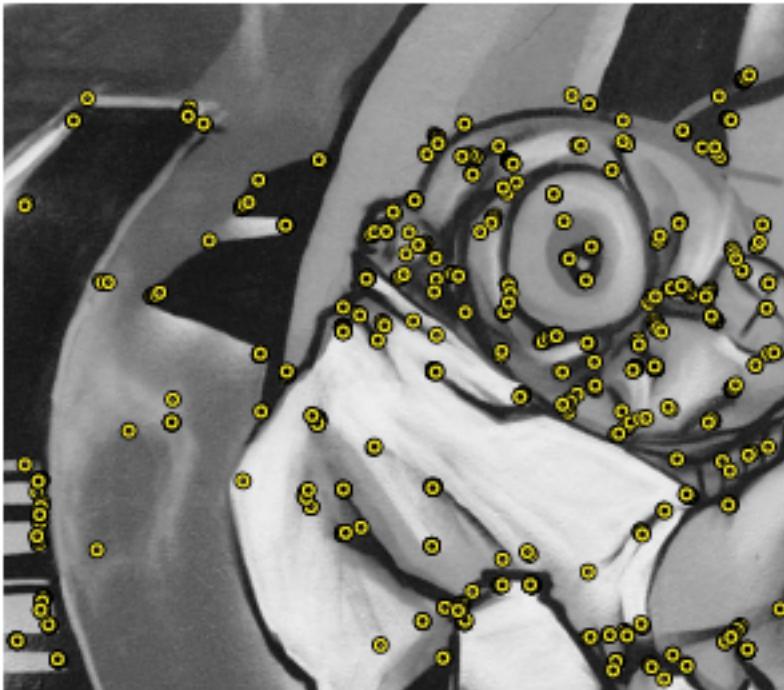
Rotational Invariance

- ▶ 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



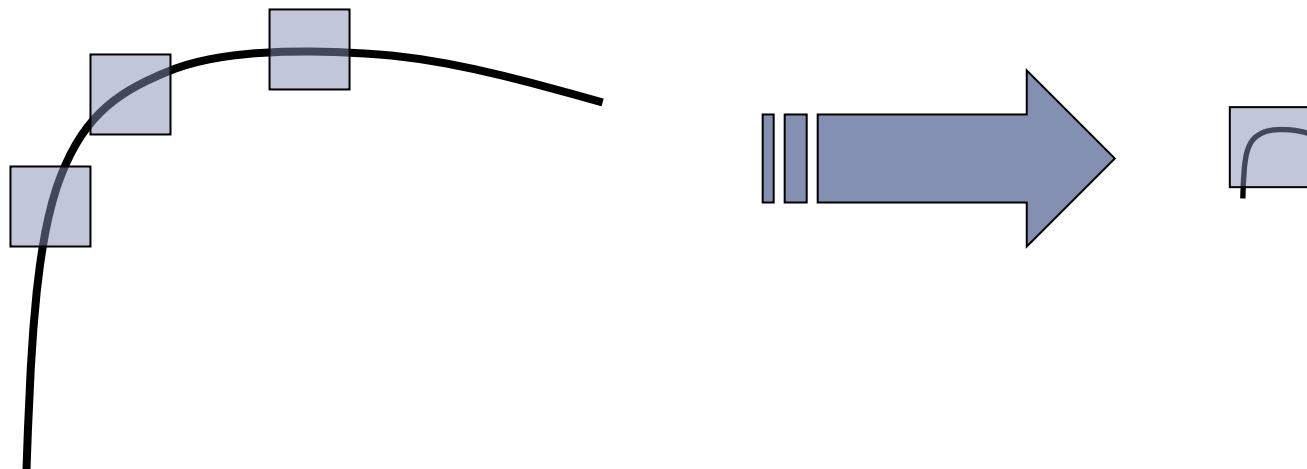
Gert Kootstra – Interest Points

Summer school, aug, 2010



Scale Invariance

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



All points will be
classified as edges

Corner !



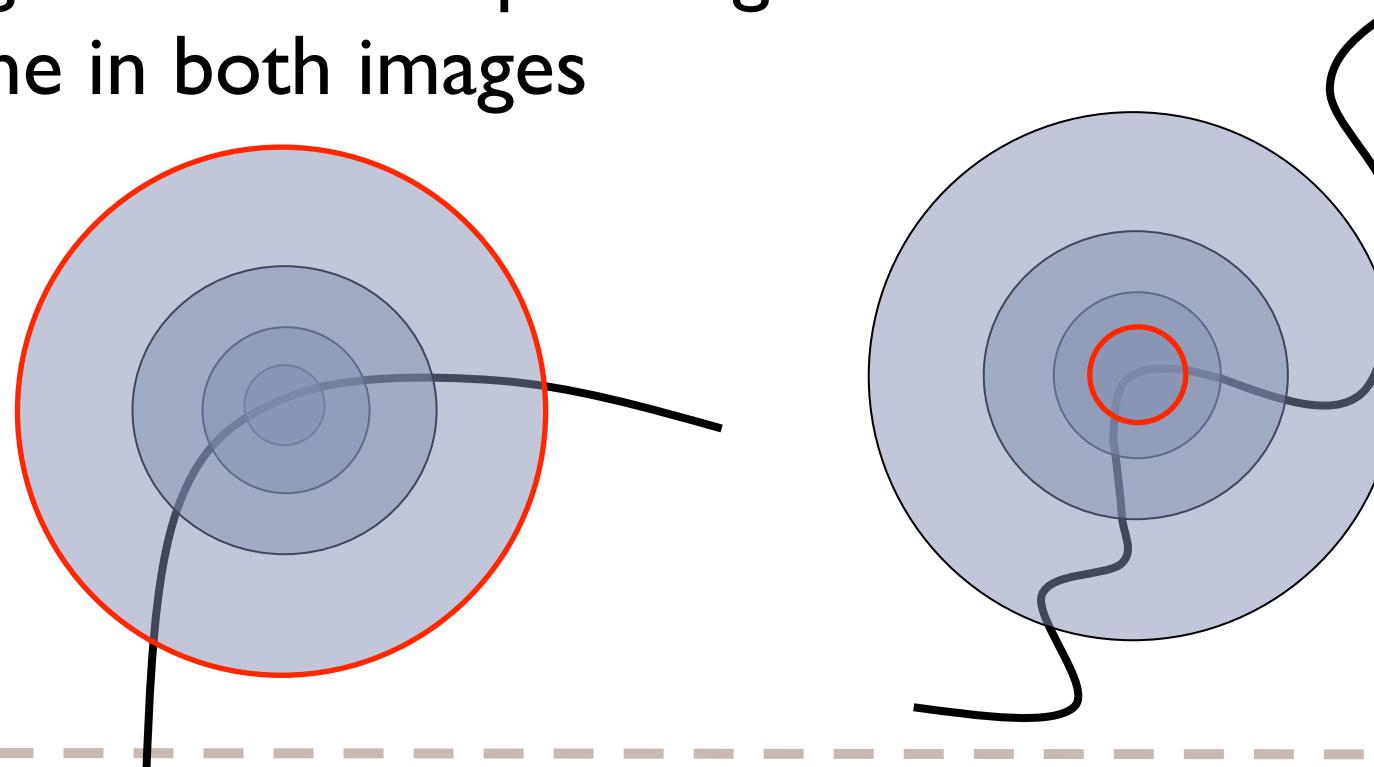
Gert Kootstra – Interest Points

Summer school, aug, 2010



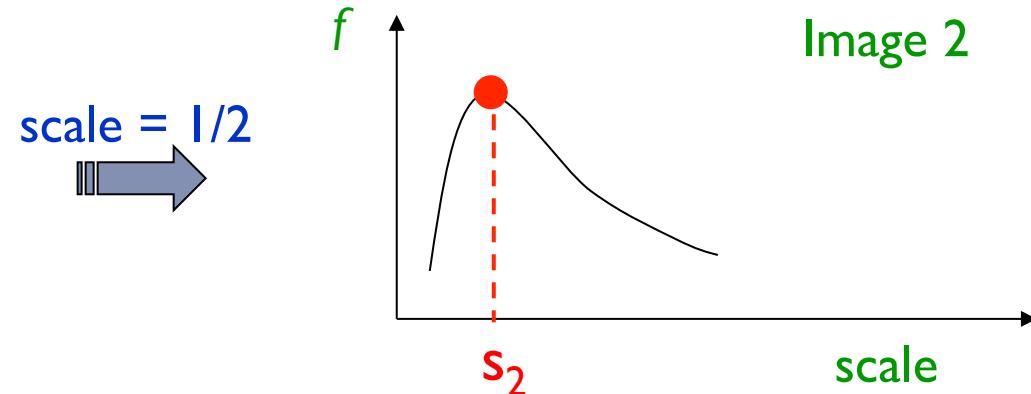
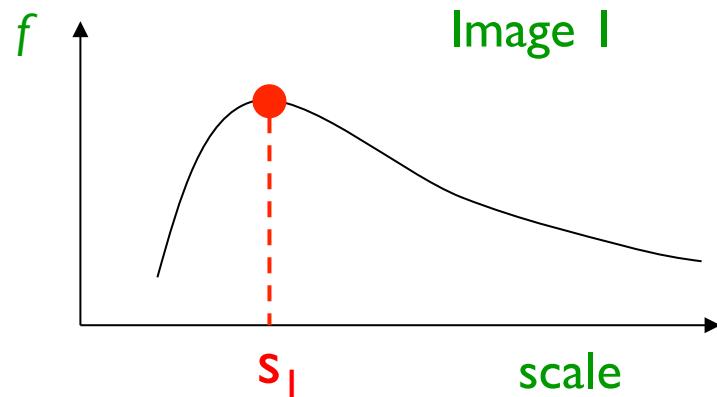
Scale-Invariance

- ▶ 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

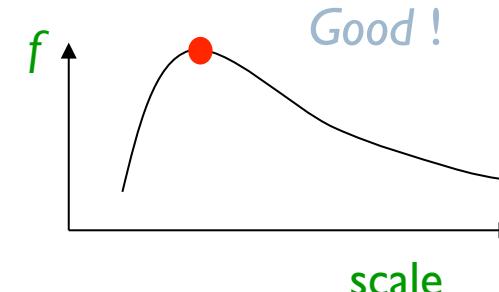
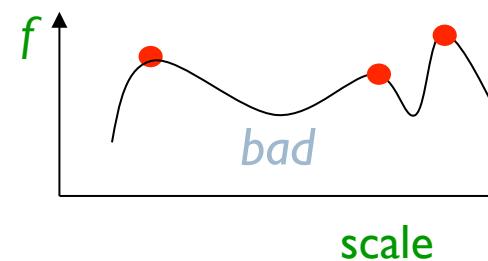
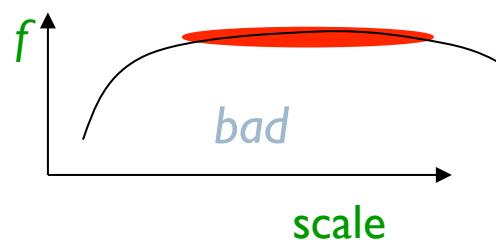
- ▶ Investigate the saliency (cornerness, ...) at different scales (T. Lindeberg).
- ▶ Characteristic scale: the scale that corresponds to the peak saliency



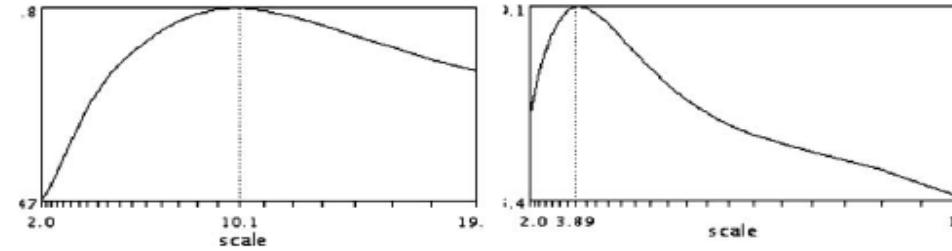
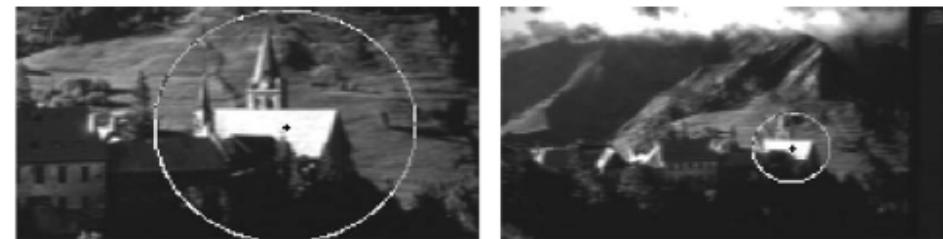
scale = 1/2
➡

Scale-Invariance Detection

- ▶ A good interest point corresponds to a unique scale



- ▶ Example



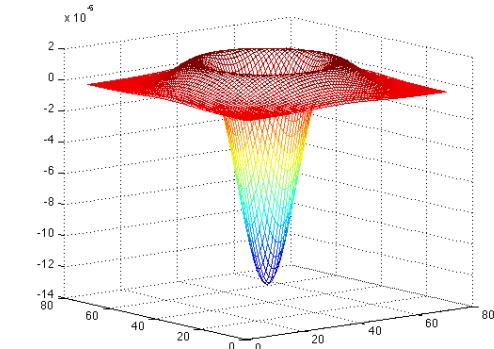
Gert Kootstra – Interest Points

Summer school, aug, 2010

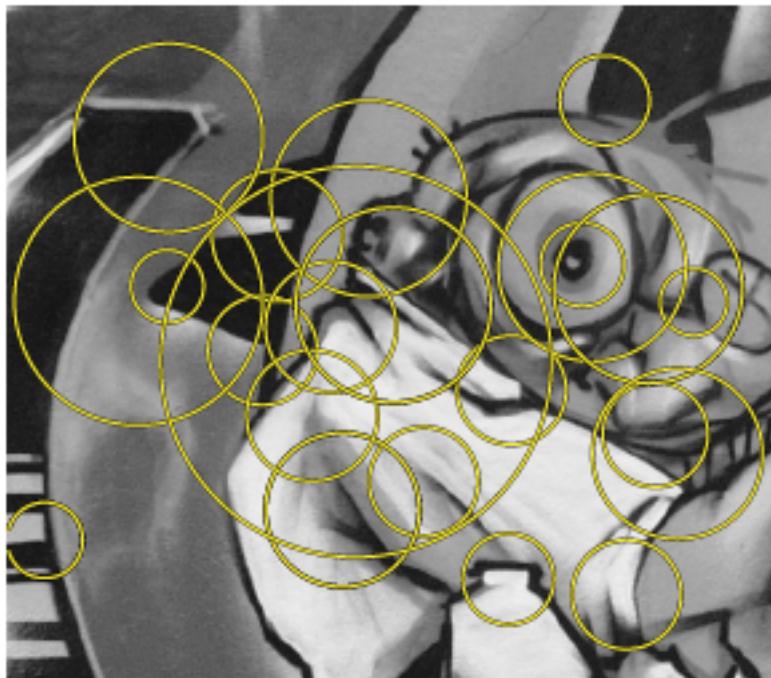


Harris-Laplace Detector (Mikolajczyk *et al* 2004)

- ▶ 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



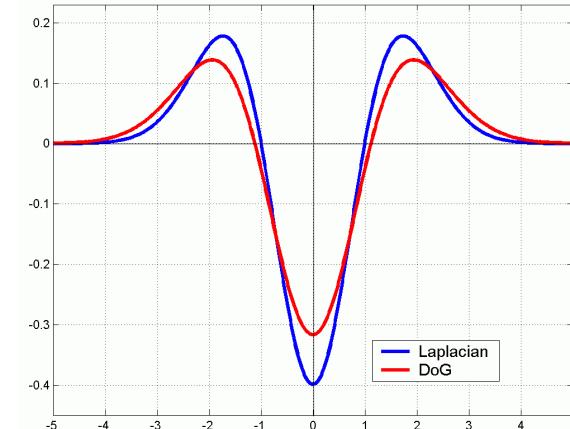
Gert Kootstra – Interest Points

Summer school, aug, 2010



Scale-Invariant Feature Transform

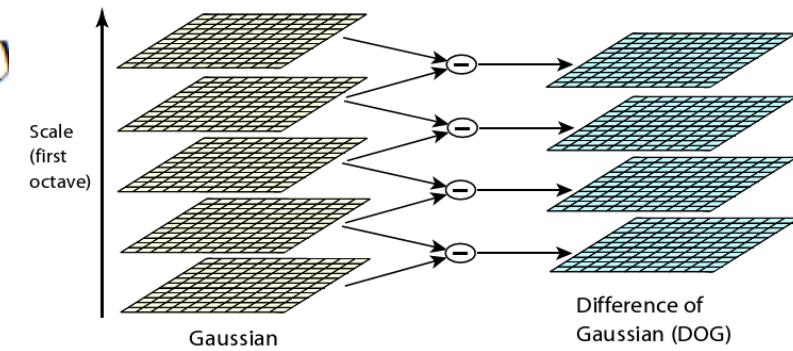
- ▶ **SIFT (Lowe 2004)**
 - ▶ Detects interest points on blobs
 - ▶ Invariant to scale and rotation
- ▶ **Based on Difference of Gaussians**
 - ▶ Approximation of Laplacian of Gaussians
 - ▶ Faster
 - ▶ Second-order derivative of image intensity



Scale-Invariant Feature Transform

- ▶ **Pyramid of Gaussian images for different scales**

$$\triangleright L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

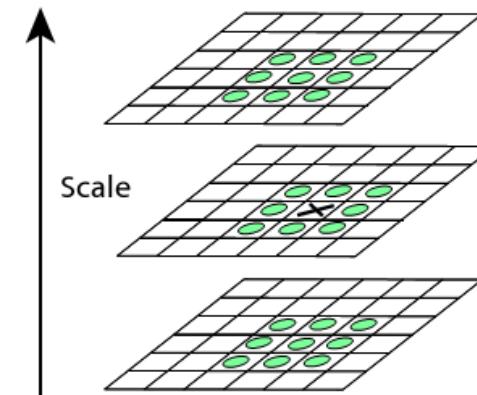


- ▶ **Pyramid of DoG images**

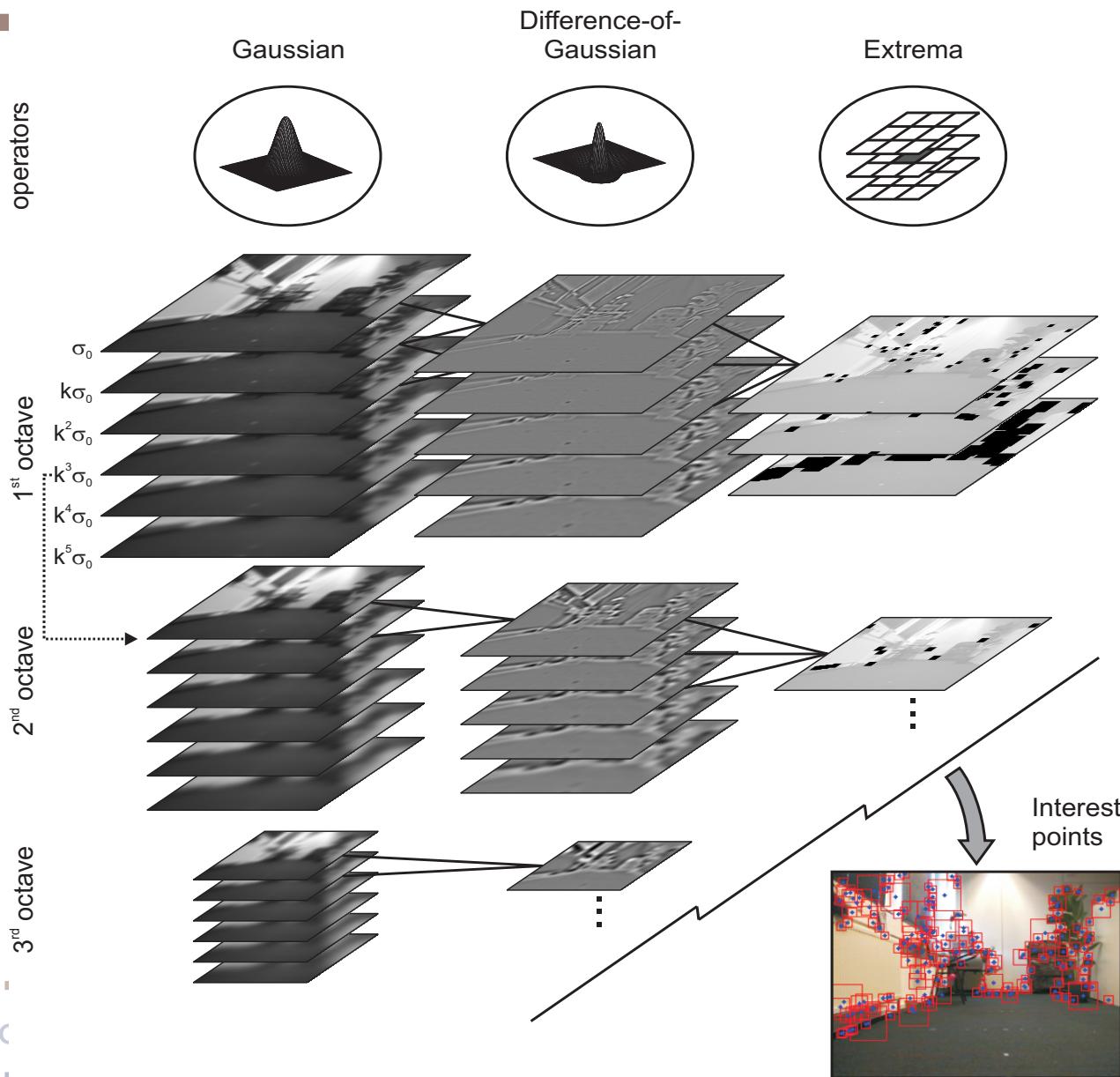
$$\triangleright D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

- ▶ **Local extrema detection**

- ▶ Minima and maxima in local $3 \times 3 \times 3$ scale-space



Scale-Invariant Feature Transform

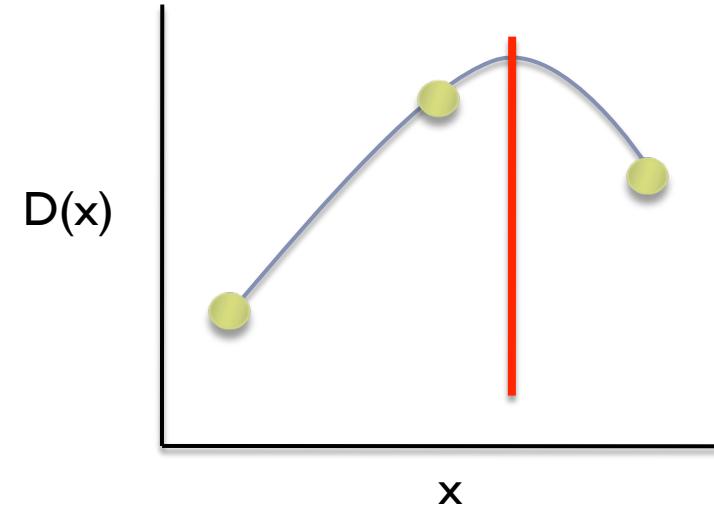


Accurate Localization of IP

- ▶ 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 \mathbf{H} are proportional to the curvature of D
- ▶ Only accept points with similar eigenvalues (ratio between the two is lower than τ_r)

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

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



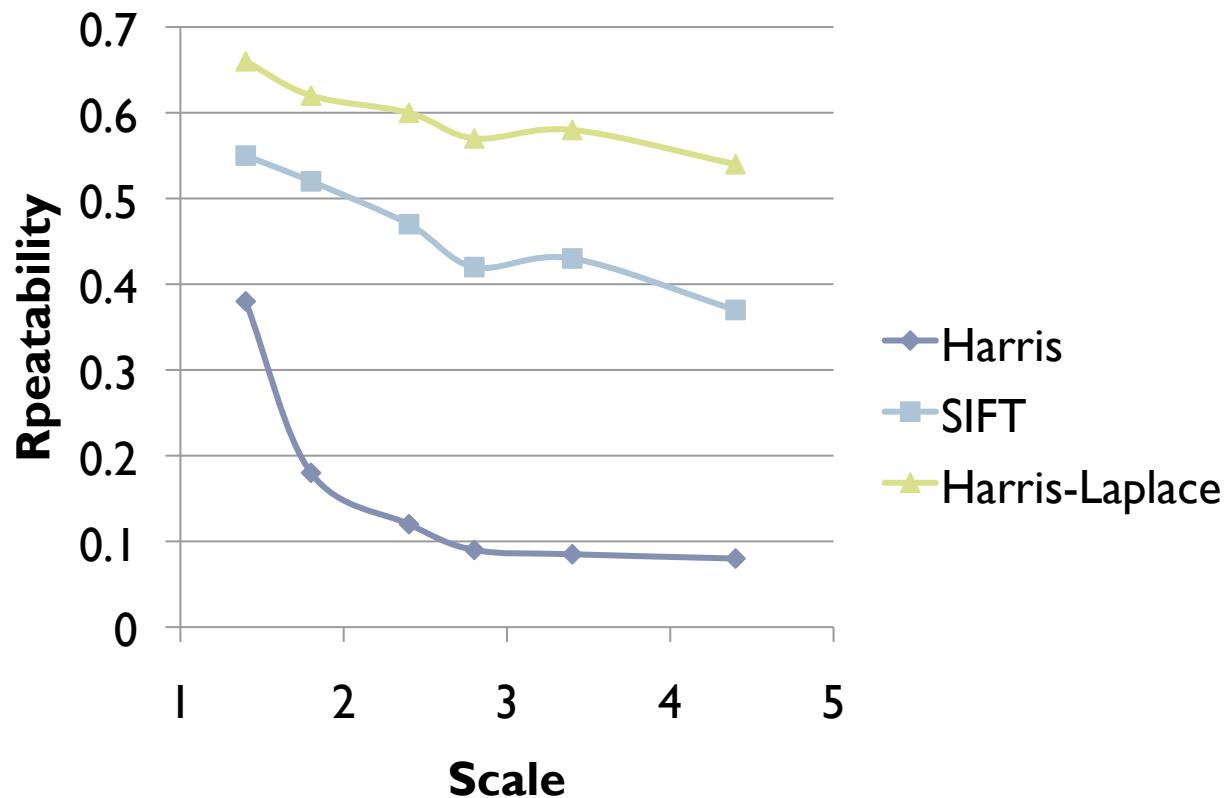
SIFT Detector Example



► G

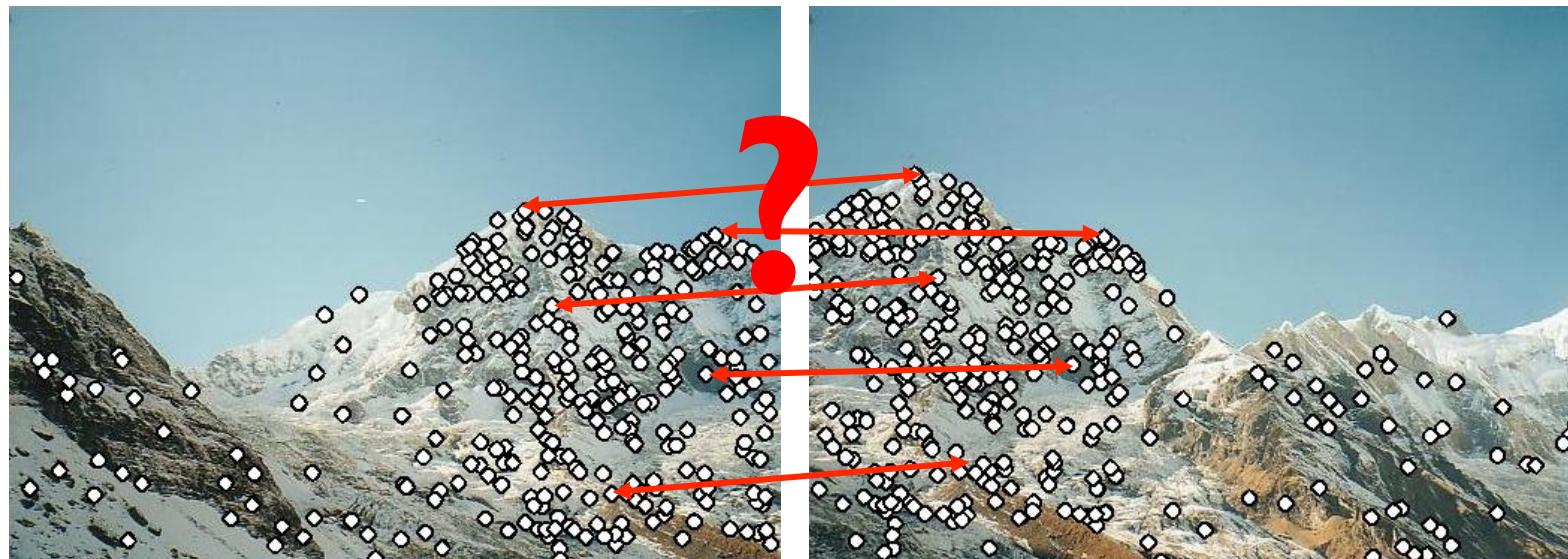
Interest-Point Detectors

▶ Repeatability



Interest-Point Descriptor

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



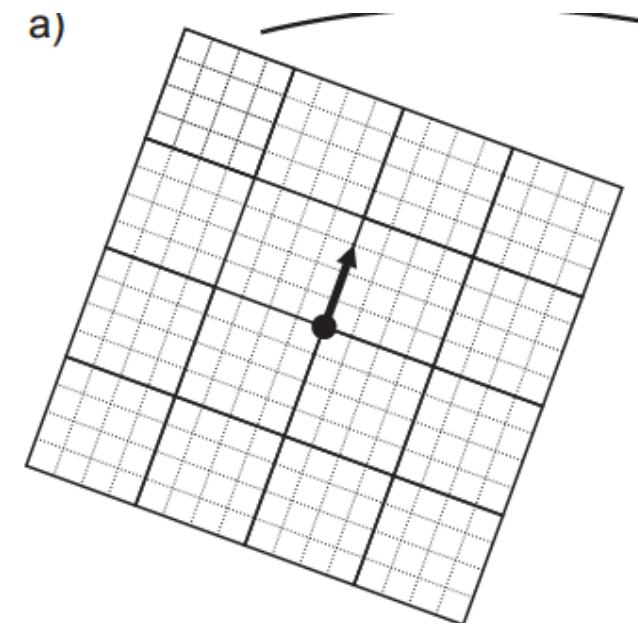
SIFT Descriptor

- ▶ 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



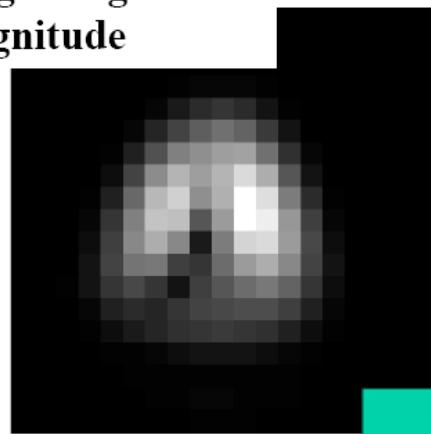
Scale and Rotational Invariant

- ▶ 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

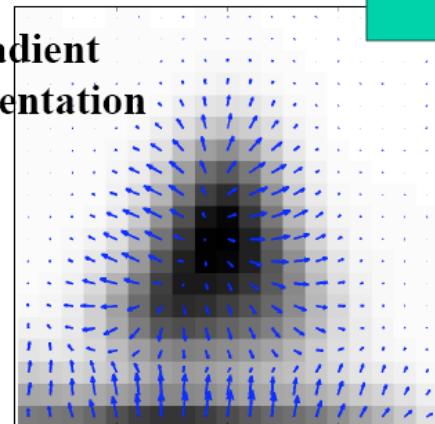


Orientation Assignment

weighted gradient magnitude

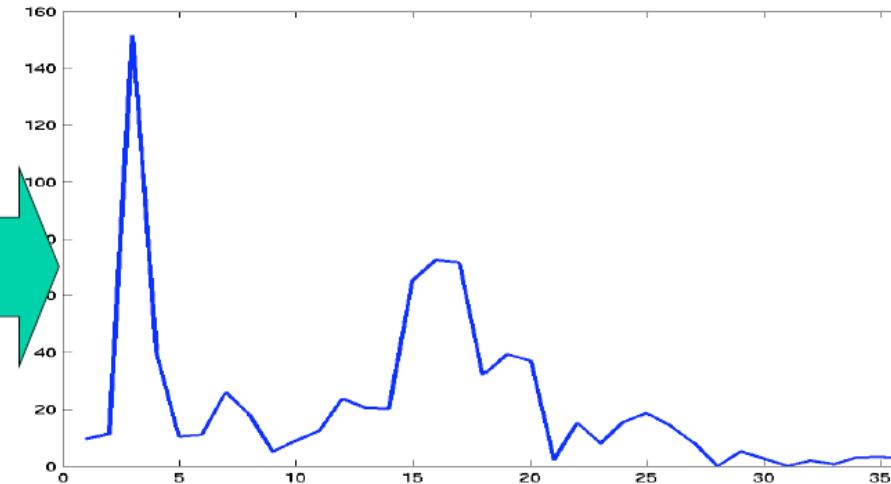


gradient orientation



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.

$0 \leq \text{ang} < 10$: bucket 1

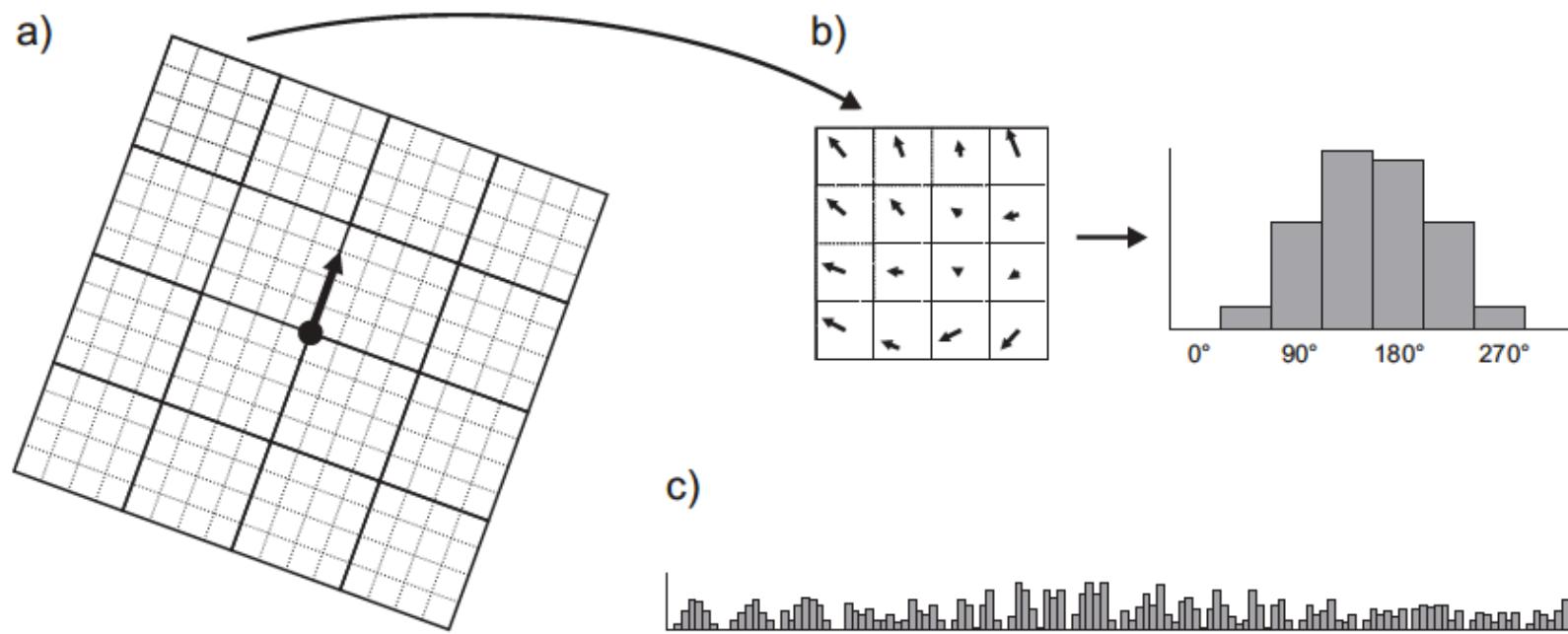
$10 \leq \text{ang} < 20$: bucket 2

$20 \leq \text{ang} < 30$: bucket 3 ...



Histograms of Oriented Gradients

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

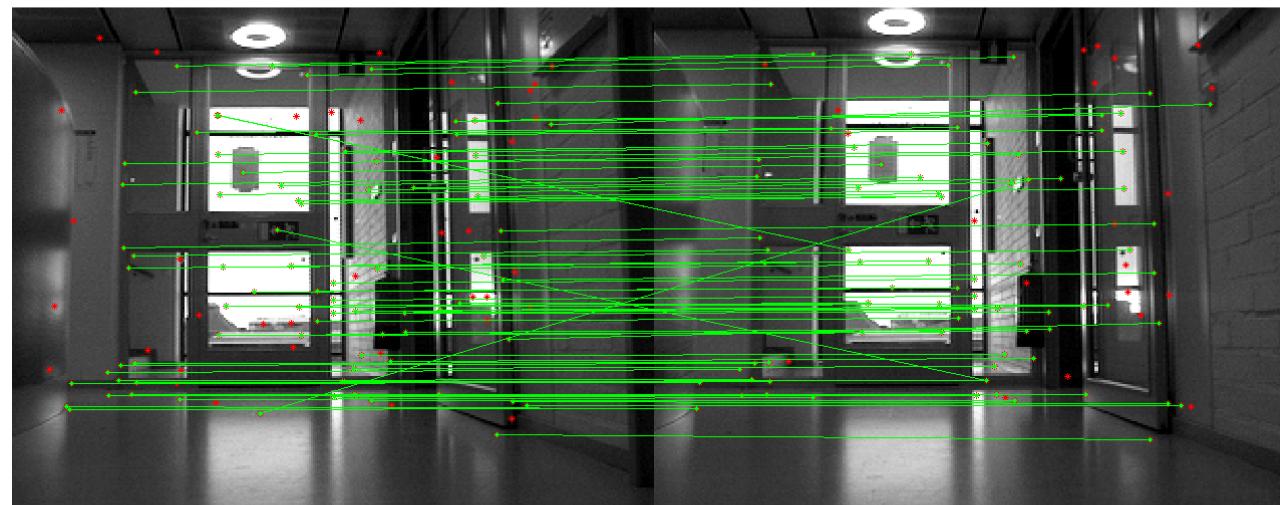
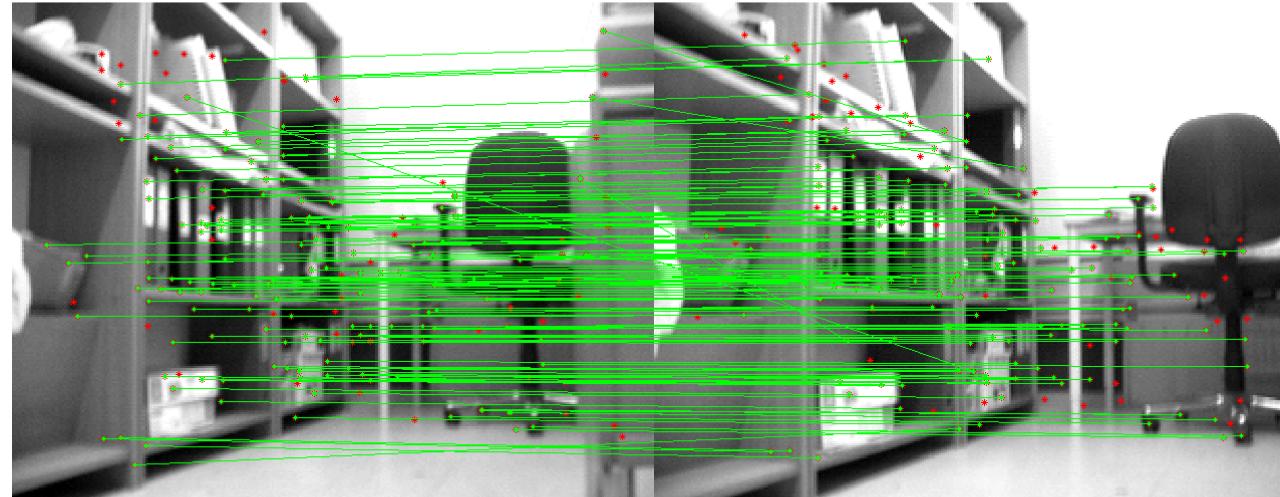


SIFT Descriptor

- ▶ 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



Gert Kootstra – Interest Points

Summer school, aug, 2010



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

► ...

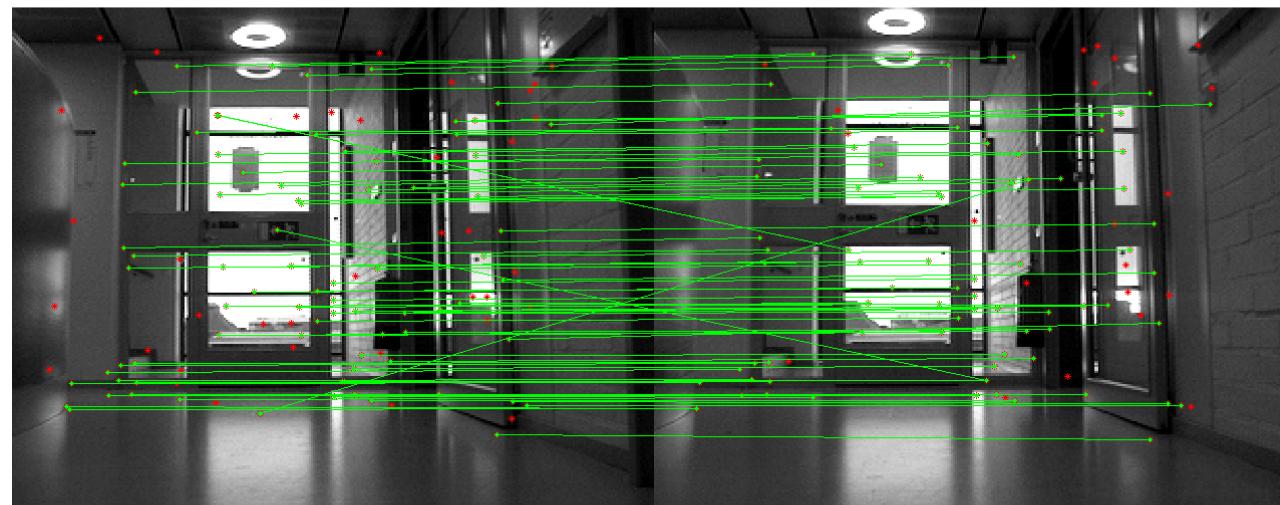


Gert Kootstra – Interest Points

Summer school, aug, 2010



Robotic Localization and Mapping

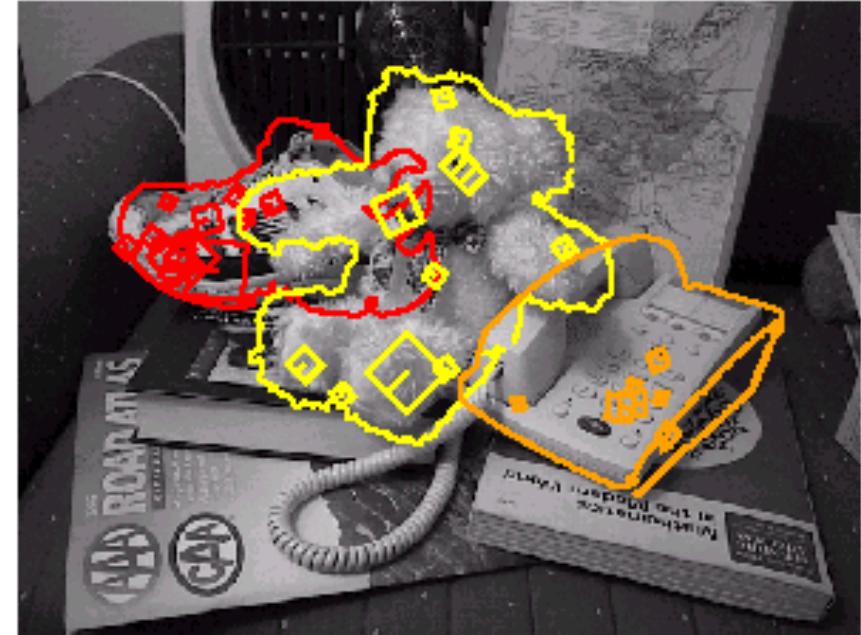
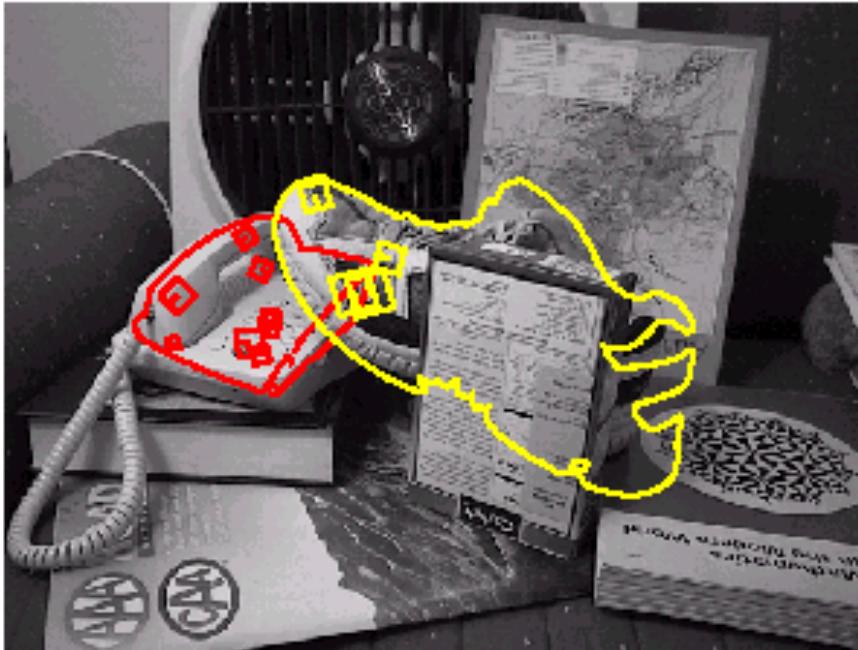


Gert Kootstra – Interest Points

Summer school, aug, 2010



Object Recognition

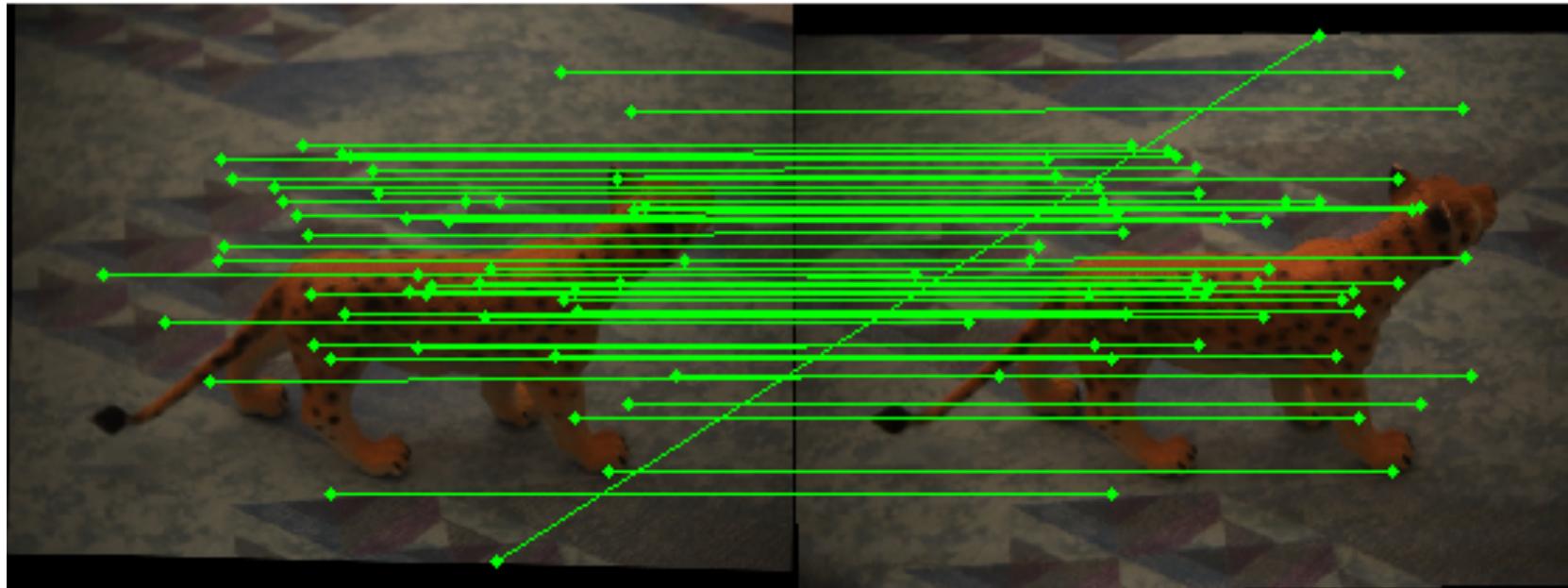


Gert Kootstra – Interest Points

Summer school, aug, 2010



Stereo Matching



Gert Kootstra – Interest Points

Summer school, aug, 2010



Panorama Stitching

Panorama stitching



(a) Matier data set (7 images)



(b) Matier final stitch

Brown, Szeliski, and Winder, 2005



Gert Kootstra – Interest Points

Summer school, aug, 2010



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’



Gert Kootstra – Interest Points

Summer school, aug, 2010



Analogy to documents

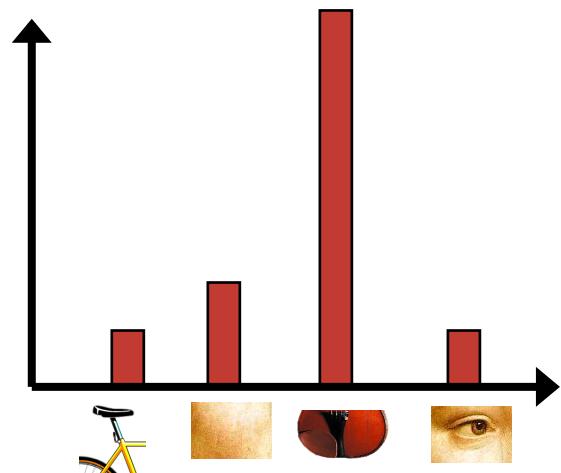
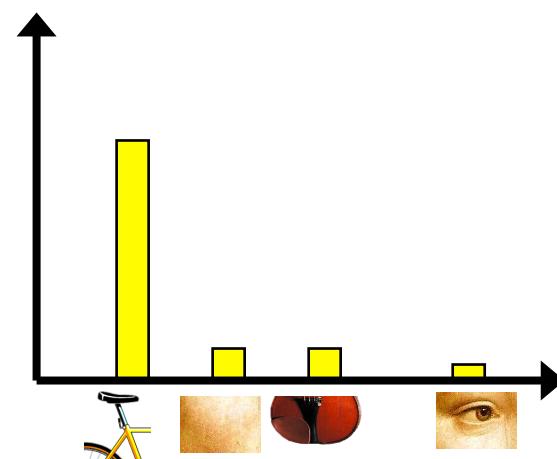
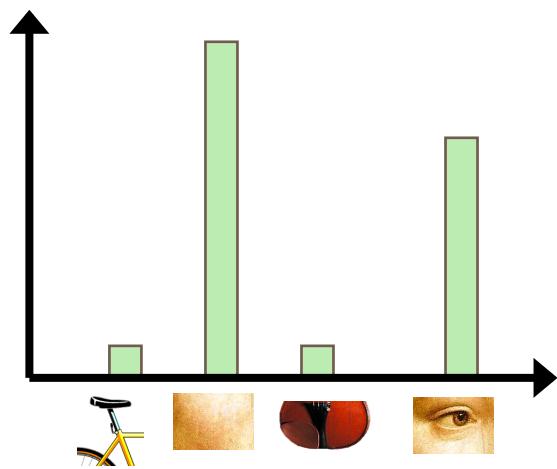
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially upon the way in which each the brain from outside. In this way we have thought the problem from the point of view of the point by point analysis of the nervous system. The cerebrum receives the sensory information upon which it acts. Through this process we have now known that the mechanism of perception is more complicated than was at first thought. The visual impulse is transmitted through various cell layers of the optic nerve to the cerebral cortex. Wiesel and Hubel have been able to demonstrate that the brain has a detailed knowledge about the image falling on the retina under investigation. They have shown that the image is analyzed in a step-wise analysis in a system of nerve cells situated in the cerebral cortex. These cells are arranged in columns. In this system each cell has its specific function and is responsible for a specific detail in the overall pattern of the retinal image.

**sensory, brain,
visual, perception,
retinal, cerebral cortex,
eye, cell, optical
nerve, image
Hubel, Wiesel**

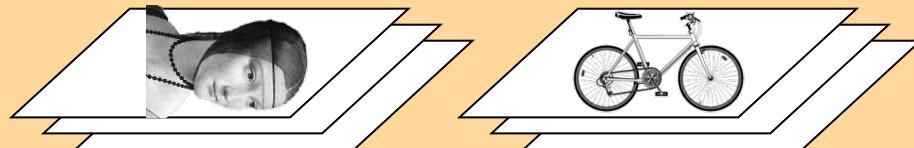
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be driven by a predicted 30% jump in exports and a 15% rise in imports, with a 18% rise in imports. The ministry also predicted further a 10% increase in imports. China's foreign exchange reserves have been deliberately built up over the last few years. The surplus is seen as a factor in the appreciation of the yuan. Beijing has said the central bank will boost domestic interest rates to encourage investment within the country. The Chinese government has allowed the yuan against the US dollar to appreciate and permitted it to trade with other currencies freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

**China, trade,
surplus, commerce,
exports, imports, US,
yuan, bank, domestic,
foreign, increase,
trade, value**





learning



feature detection
& representation

codewords dictionary

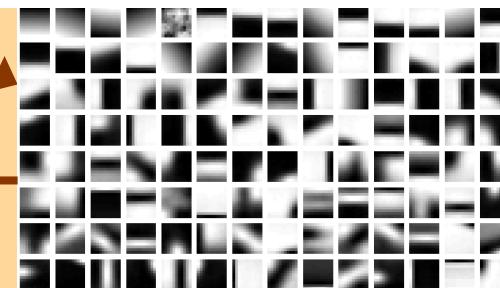
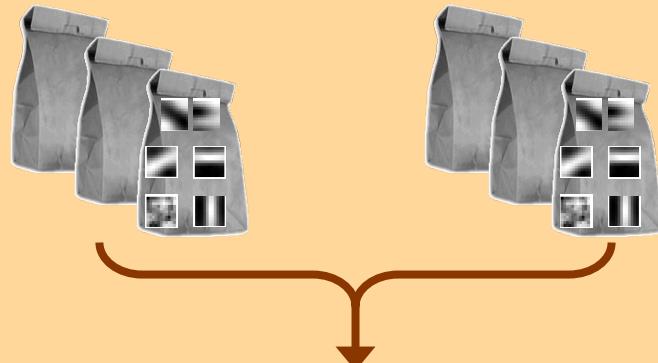


image representation

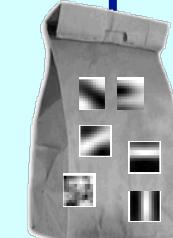
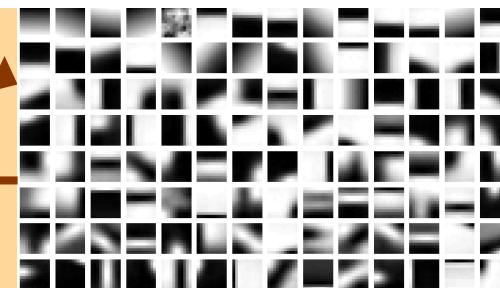


**category models
(and/or) classifiers**

recognition



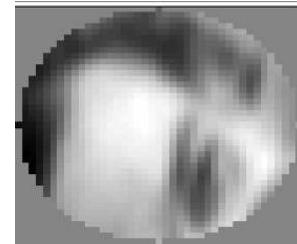
feature detection
& representation



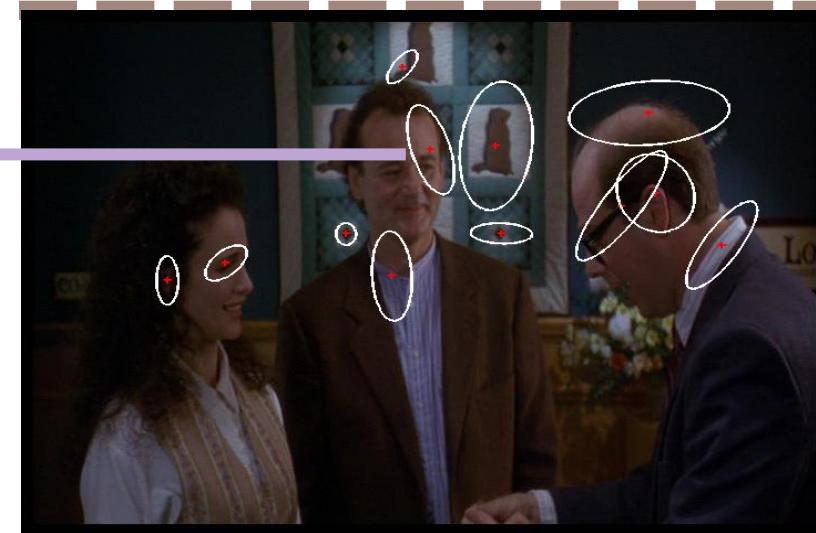
**category
decision**

Interest Point Features

Compute
SIFT
descriptor
[Lowe'99]



Normalize
patch



Detect patches

[Mikojaczyk and Schmid '02]

[Matas et al. '02]

[Sivic et al. '03]



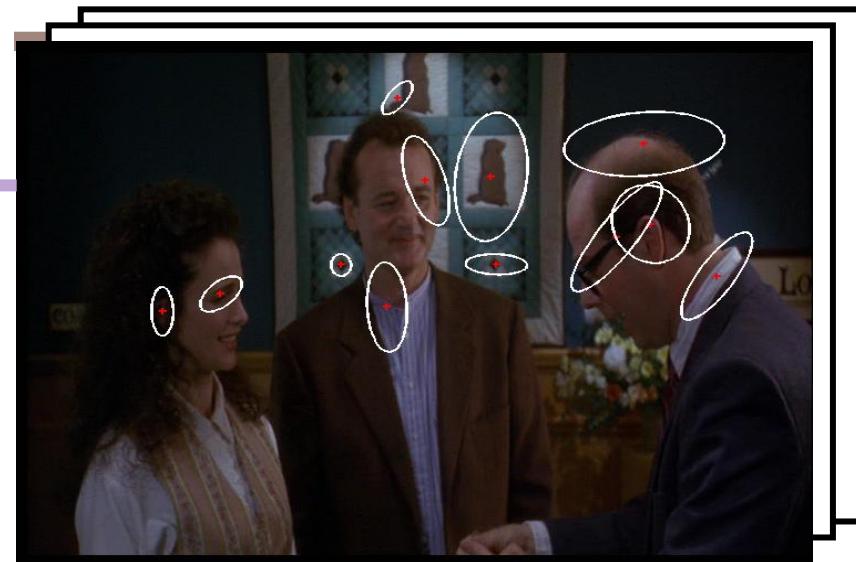
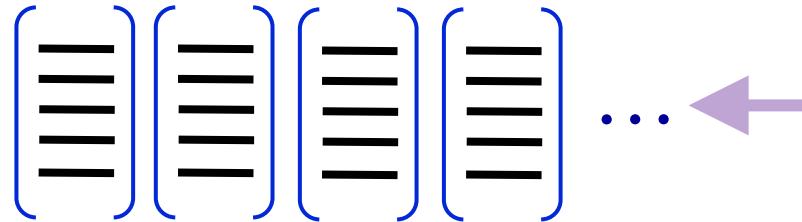
Gert Kootstra – Interest Points

Summer school aug 2010

Slide credit: Josef Sivic



Interest Point Features

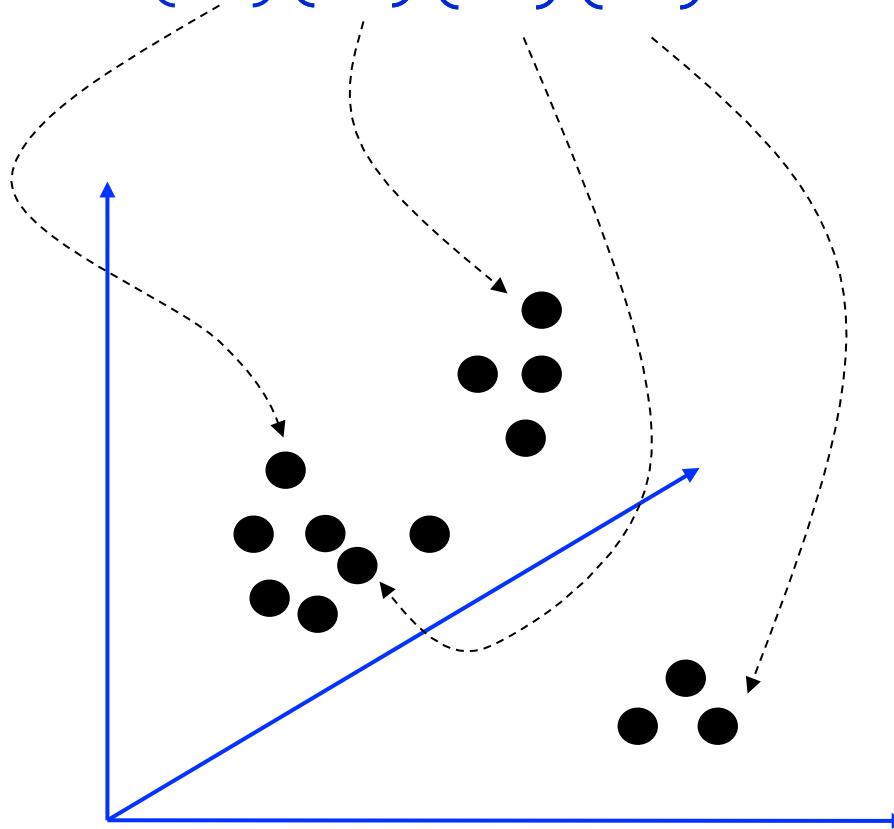
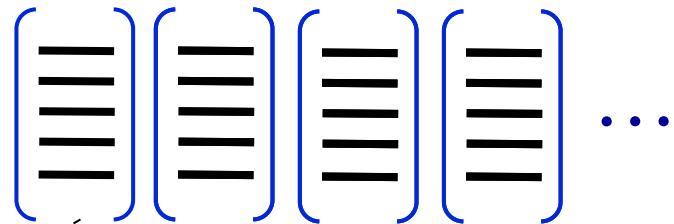


Gert Kootstra – Interest Points

Summer school, aug, 2010



dictionary formation

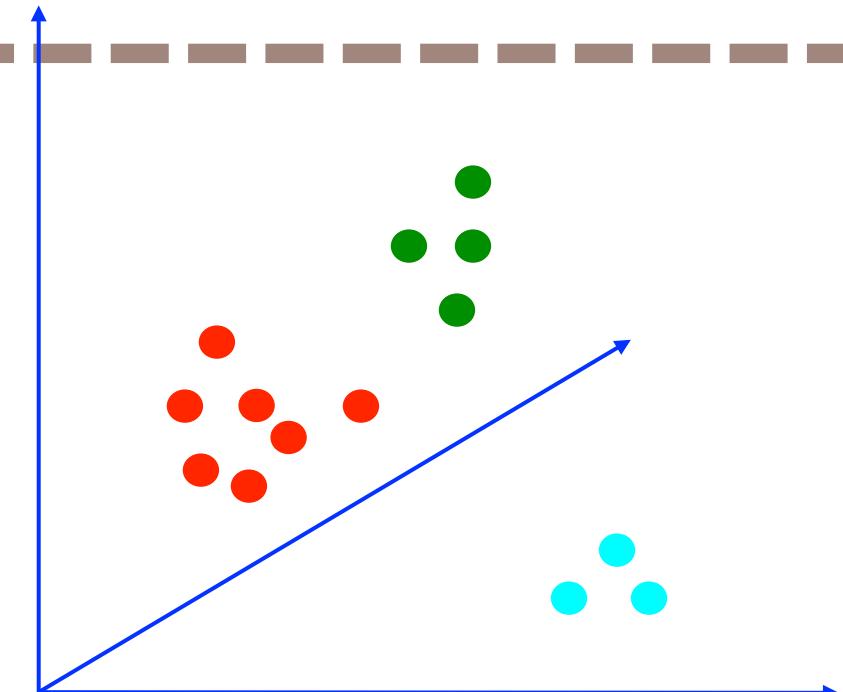
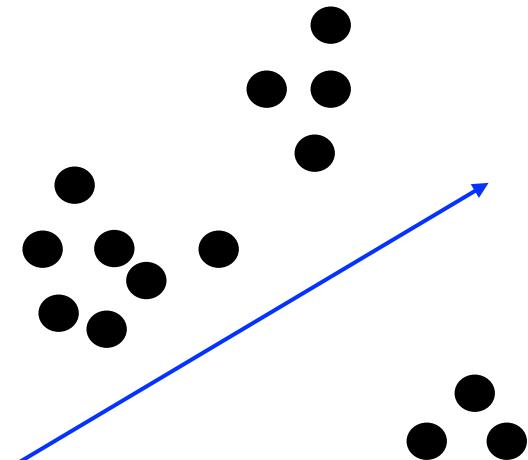
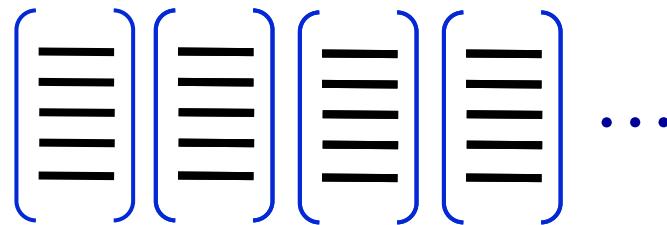


Gert Kootstra – Interest Points

Summer school, aug, 2010



Clustering (usually k-means)

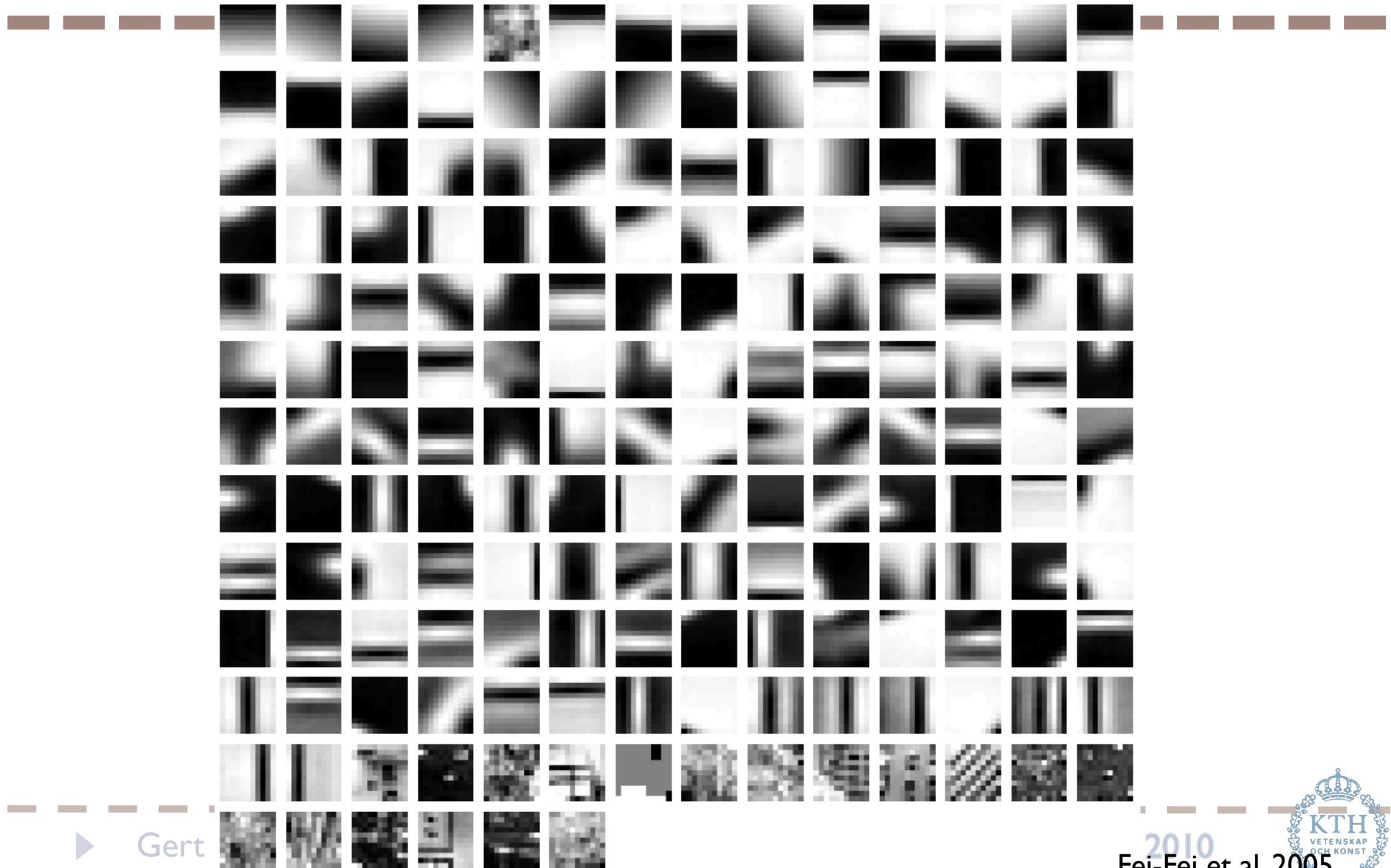


Gert Kootstra – Interest Points

Summer school aug 2010

Slide credit: Josef Sivic

Clustered Image Patches

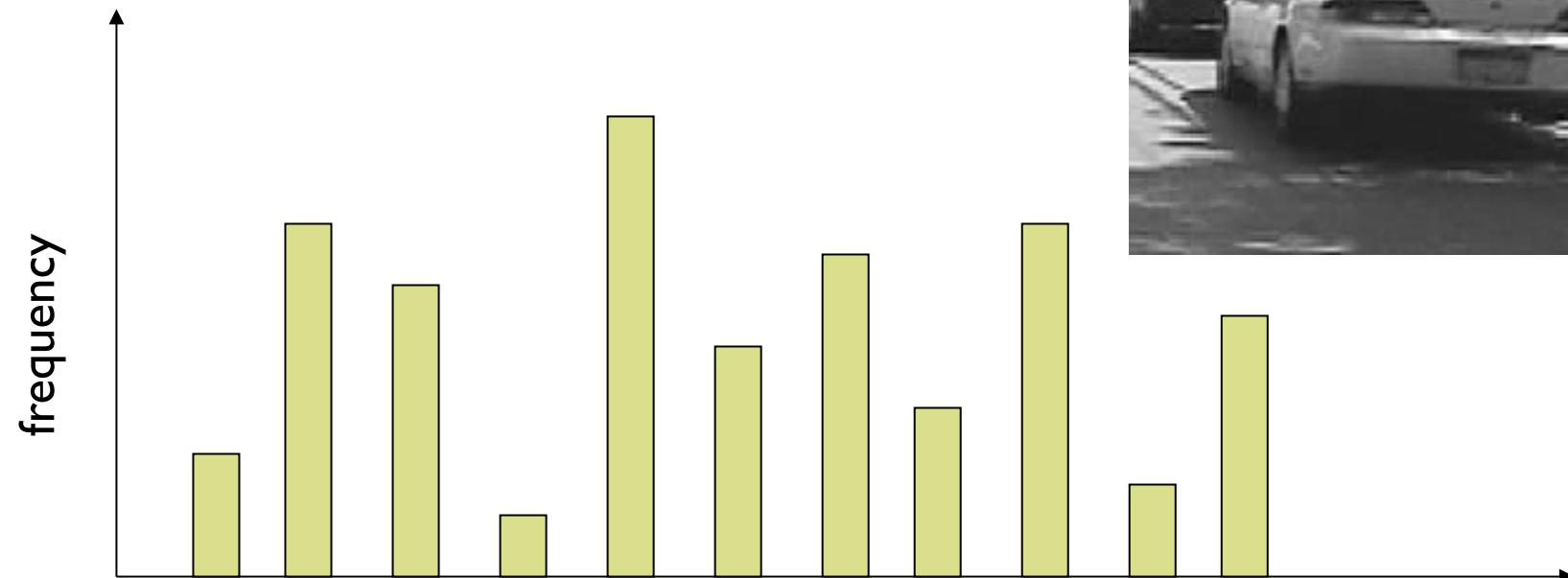


► Gert

2010
Fei-Fei et al. 2005



Image representation



codewords



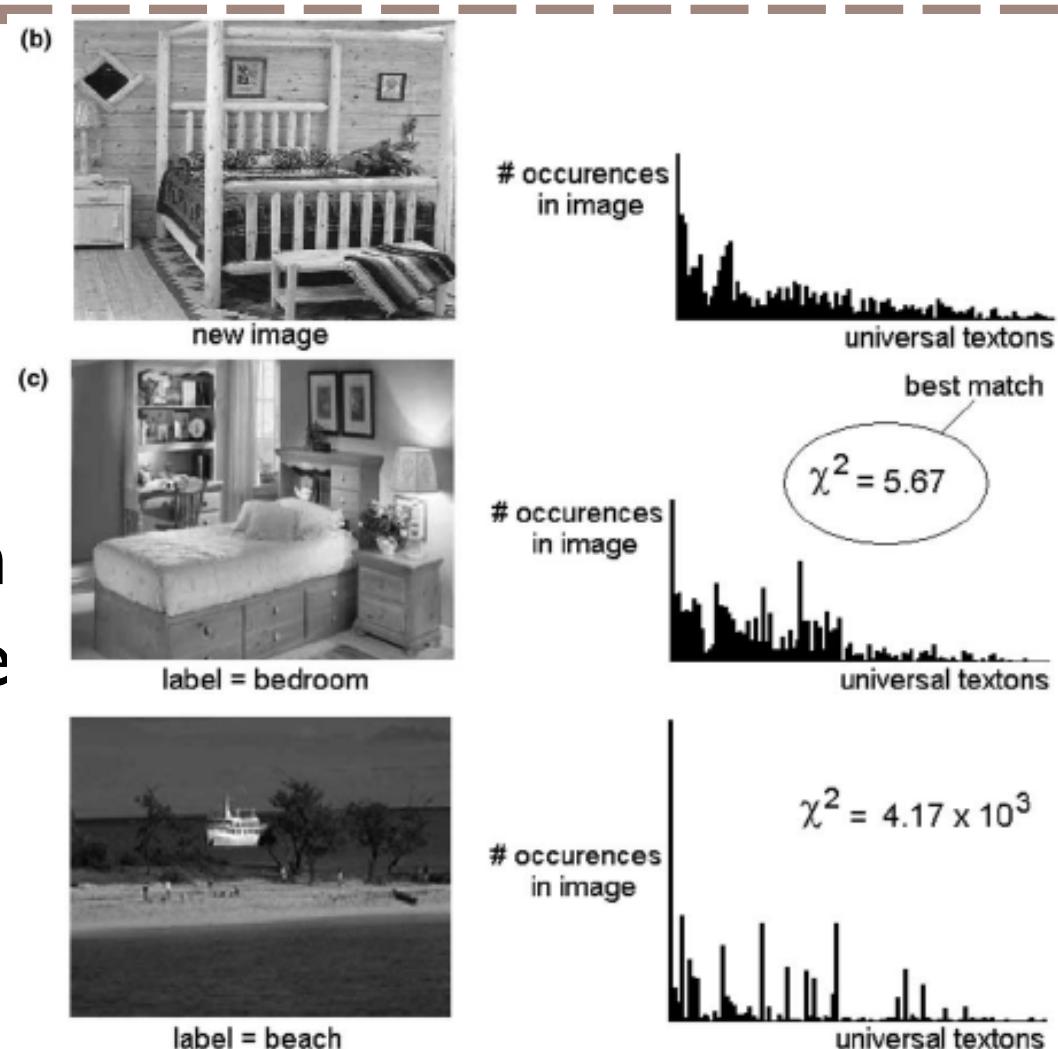
Gert Kootstra – Interest Points

Summer school, aug, 2010



Image Matching

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



Bag of Words

- ▶ Works well for image/object classification
- ▶ Reduces the number of features
 - ▶ Standard SIFT
 - ▶ $\pm 1,000$ 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

