D E P A R T M E N T O F C O M P U T E R S C I E N C E
U N I V E R S I T Y   O F   C O P E N H A G E N

Features I:
Detecting Interest Points

Kim Steenstrup Pedersen

Plan for today and next couple of lectures

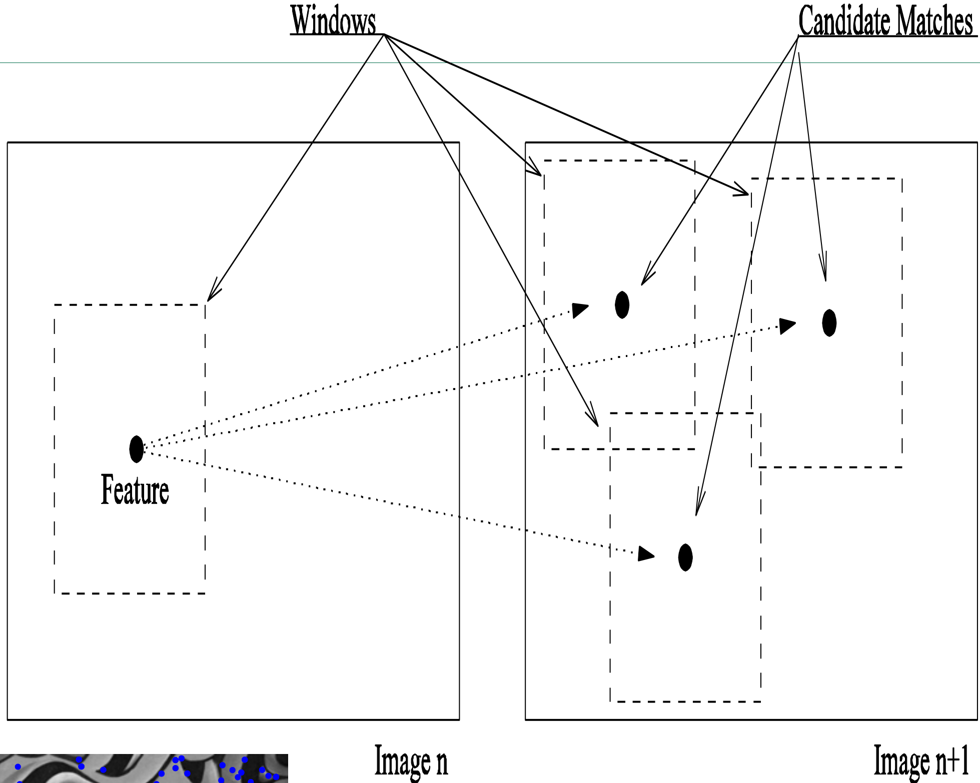
- We consider the problem of matching two or more images
- Today we focus on detecting good locations to match.
- Next week we focus on how to perform the actual matching.

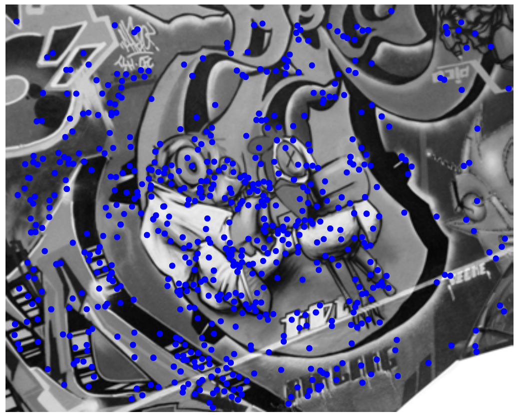
- Relevant for
  - 3D reconstruction
  - Content-based image retrieval
  - Object detection and recognition
  - Tracking

Fundamental Question:
Where did the underlying feature go?

Matching: Ideally, we want to find the location in the other image which corresponds to the same physical location on the building

Matching Strategy Illustrated





Not all image patches can be matched

To be able to figure out which points belongs to which we need a bit of structure:
Sky patches are impossible to match
Curves are also problematic

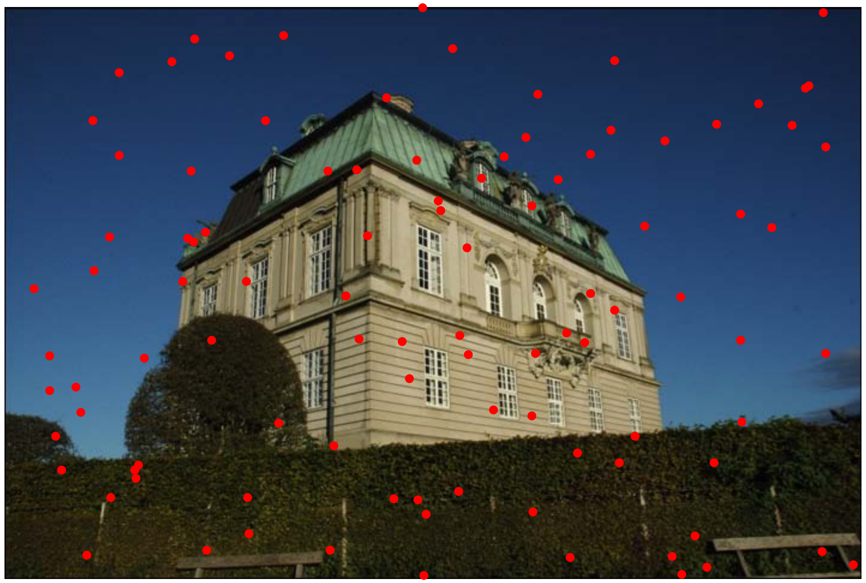
Dense versus sparse matching?

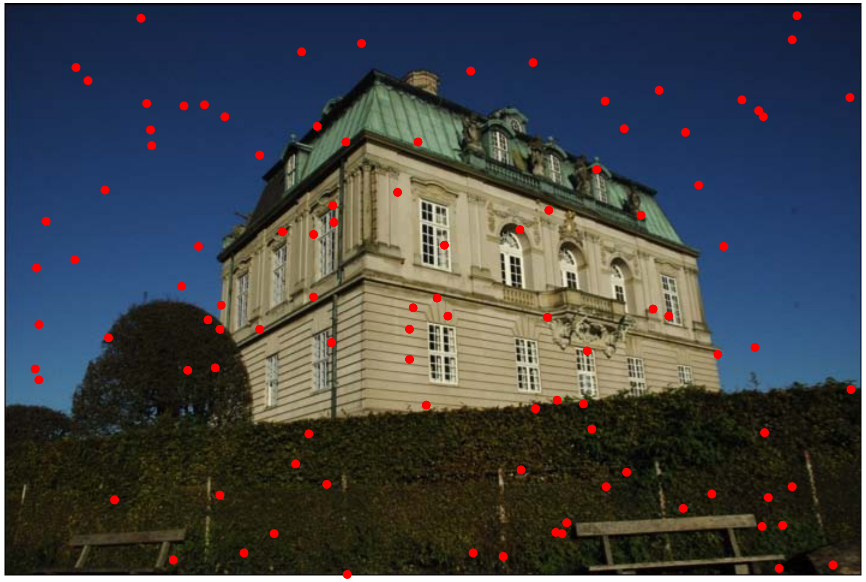
Dense : Match all pixels in the images
Sparse: Match a subset of positions in the images
Both approaches have merit, but dense is computationally expensive

Regular grid vs random points vs salient points

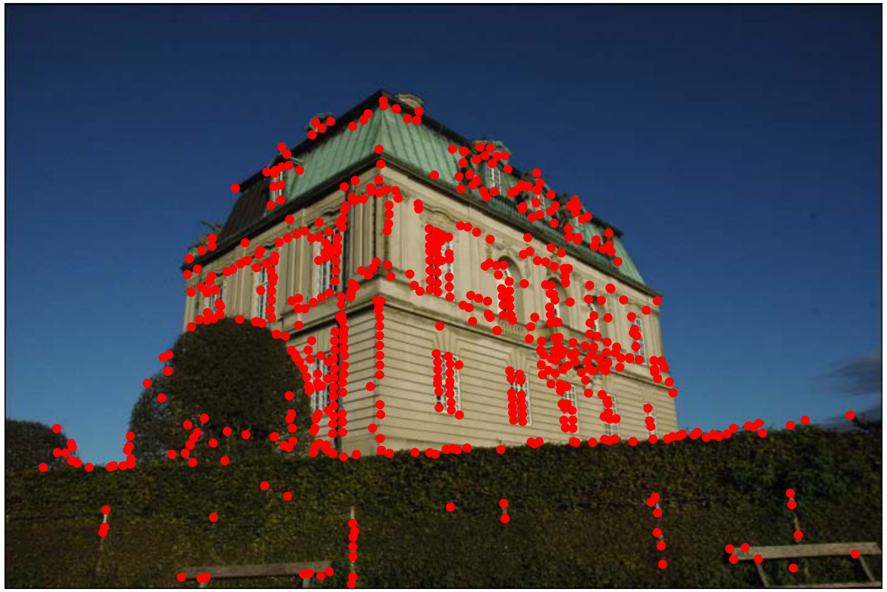
Regular grid vs random points vs salient points

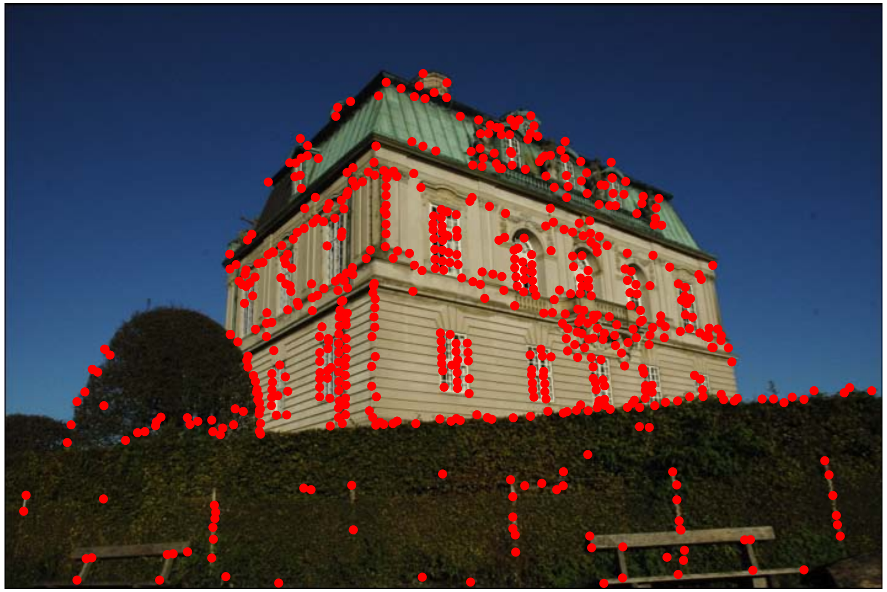
Regular grid vs random points vs salient points





Regular grid vs random points vs salient points





Salient points = Interests points

- **Salient points**: Local structure that appear distinct from the image in the surrounding region of the salient point.

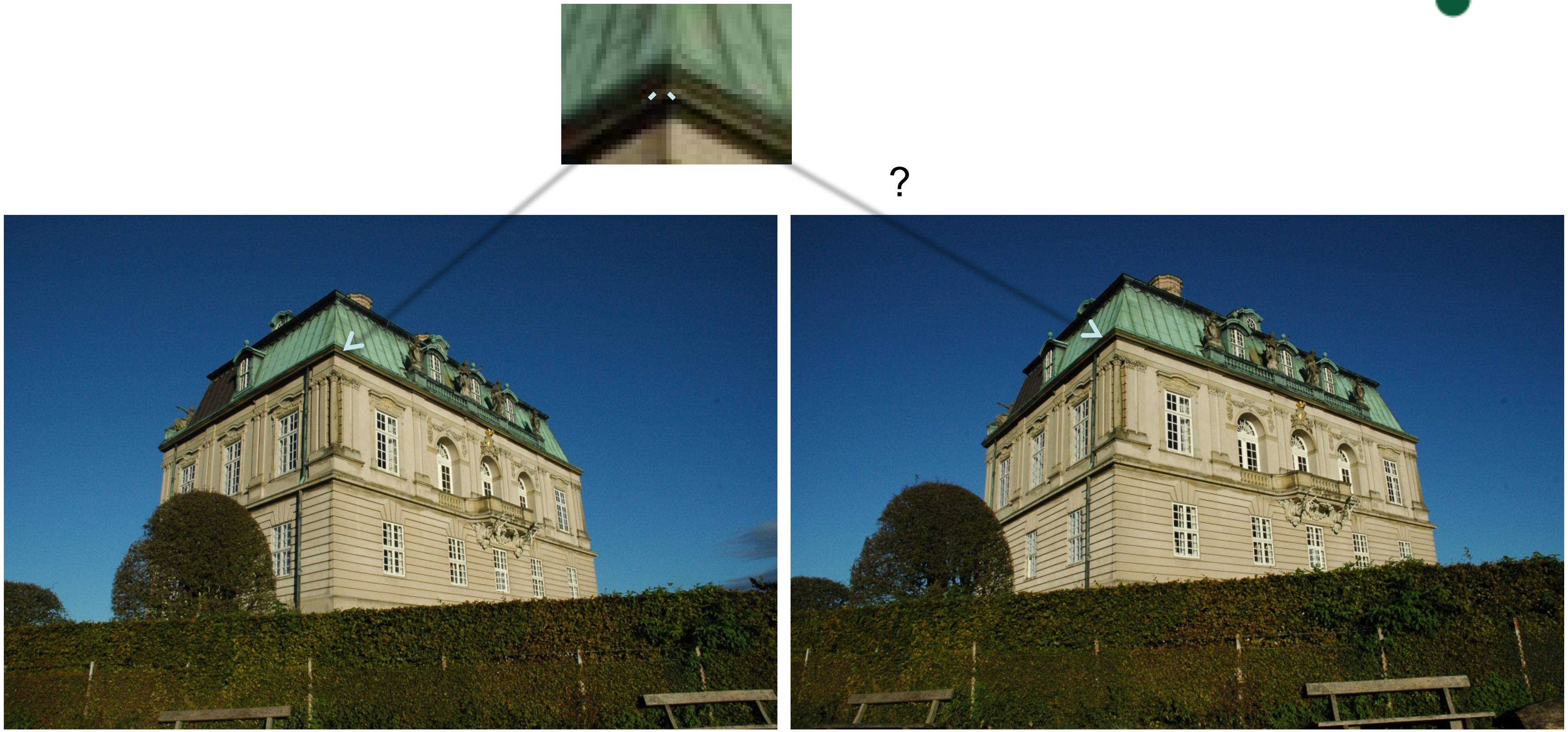
- We need an operational definition in order to detect such points!

- Terminology: Salient points aka interest points aka keypoints aka features (old confusing terminology)

How to automatically detect interesting points?

Detecting points by template matching

Template



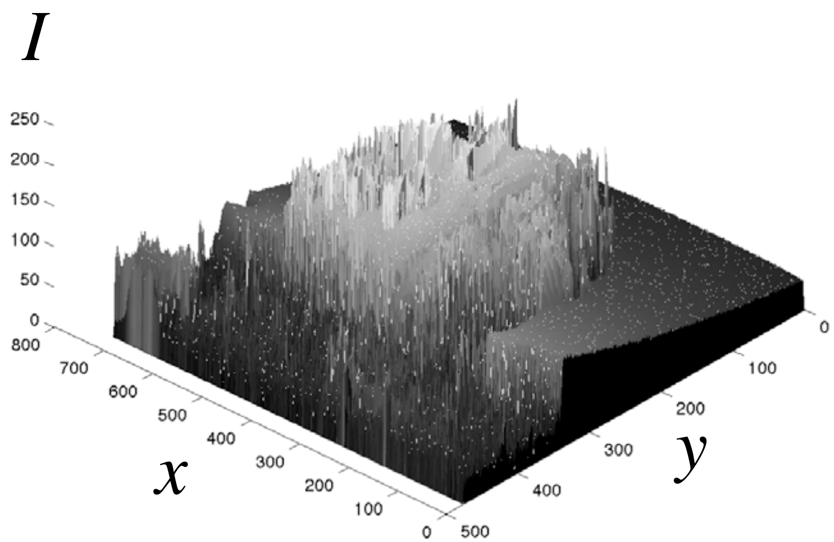
How do we construct a generic corner template?

Bad idea – won't work!

Viewing images as functions

$I(x,y)$

$I(x,y)$



Interest point detectors:
Detecting blobs

- Local intensity extrema (maxima and minima) are potential candidates for salient points.

- Extrema are distinct from the neighboring pixels.

- We refer to these extrema as bright or dark blobs.

- How do we find intensity extrema?

  Compute image derivatives

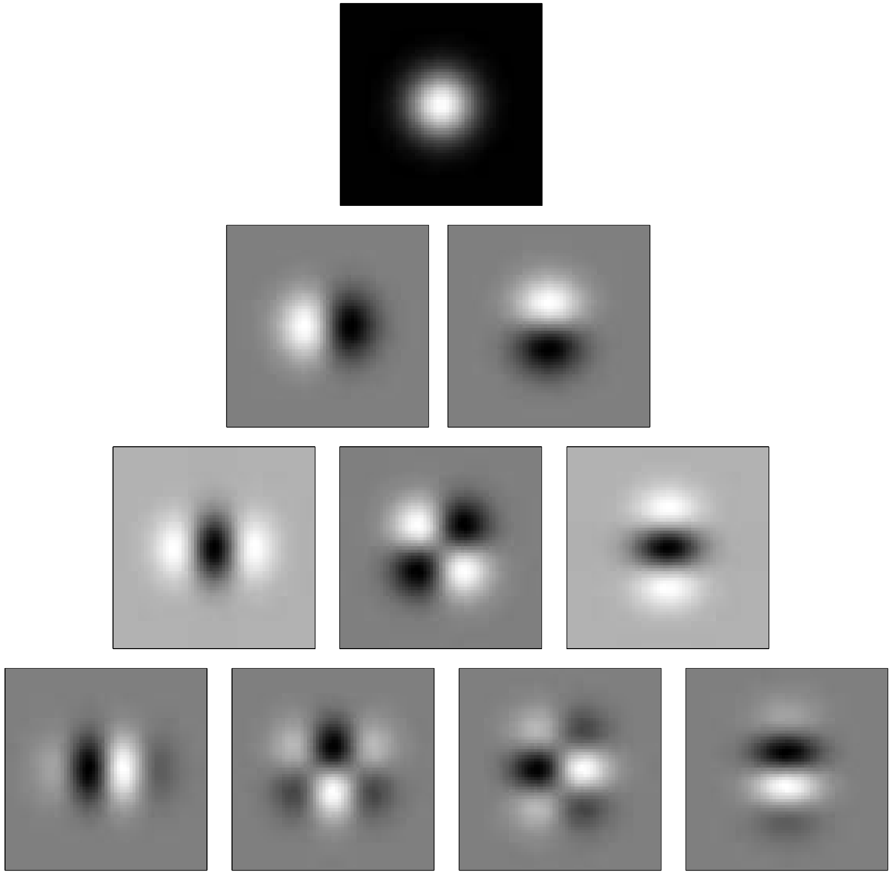
Linear Filtering 101

- Filtering by discrete convolution
  - A filter is defined by a filter kernel $h(x,y)$
  - Filtering the image $I(x,y)$ with the kernel $h(x,y)$ is defined as

$$R(x,y) = \sum_{u,v} I(u,v)h(x-u,y-v) \equiv I(x,y) * h(x,y)$$

- Consider filter kernels as images
  - They do not have to be of the same size as the image
  - Filtering slides the reverse filter kernel across the image and compute the product sum at each location
  - The result at (x,y) is called the filter response
  - *R* is an image of same size as the original image *I*

Gaussian derivative filter kernels



$$G(x,y;\sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

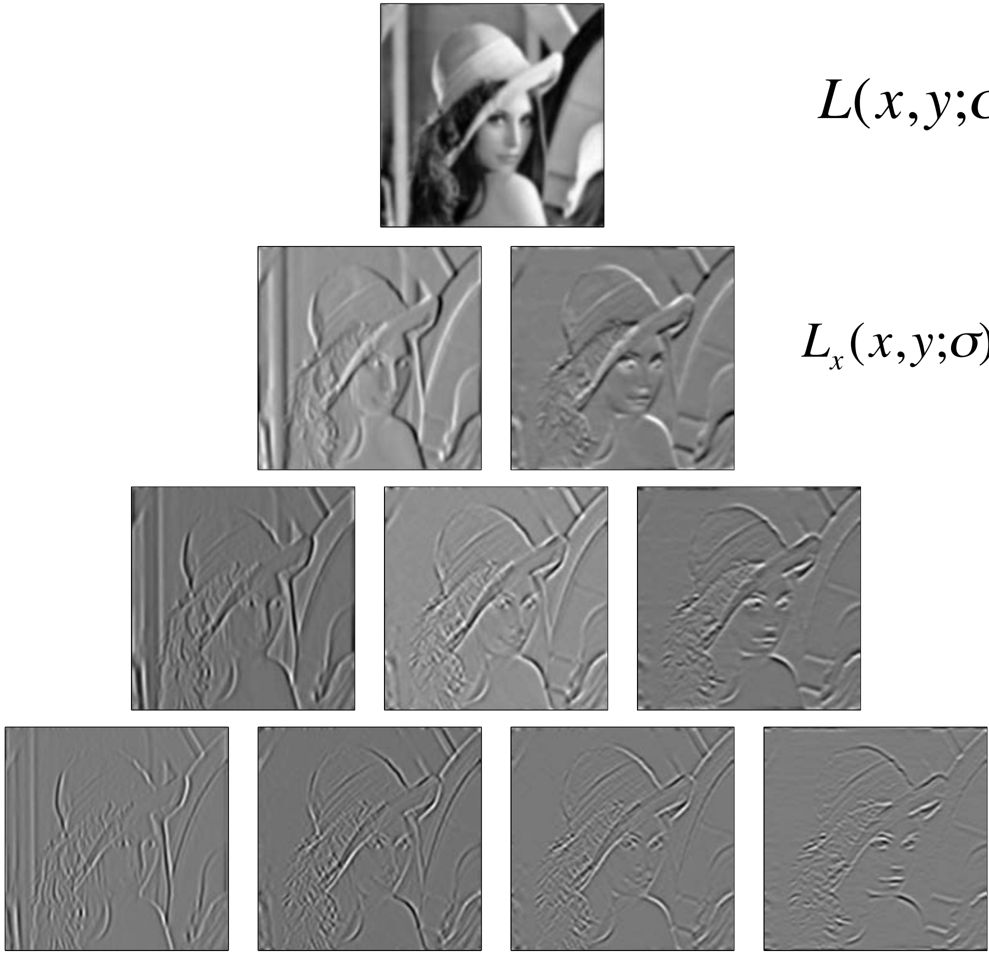
$$\frac{\partial}{\partial x}G\ ,\ \frac{\partial}{\partial y}G$$

$$\frac{\partial^2}{\partial x^2}G\ ,\ \frac{\partial^2}{\partial x\partial y}G\ ,\ \frac{\partial^2}{\partial y^2}G$$

$$\frac{\partial^3}{\partial x^3}G\ ,\ \frac{\partial^3}{\partial x^2\partial y}G\ ,\ \frac{\partial^3}{\partial x\partial y^2}G\ ,\ \frac{\partial^3}{\partial y^3}G$$

Aside: Neurophysiological experiments by Young (1987) show that the receptive field profiles in the human retina and visual cortex can be modelled by Gaussian derivatives.

Image derivatives with Gaussian filters



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

$$L_x(x,y;\sigma) = I * \frac{\partial}{\partial x} G \;,\; L_y(x,y;\sigma) = I * \frac{\partial}{\partial y} G$$

$$L_{x^2} = I * \frac{\partial^2}{\partial x^2} G \;,\; L_{xy} = I * \frac{\partial^2}{\partial x \partial y} G \;,$$

$$L_{y^2} = I * \frac{\partial^2}{\partial y^2} G$$

$$L_{x^3} = I * \frac{\partial^3}{\partial x^3} G \;,\; L_{x^2 y} = I * \frac{\partial^3}{\partial x^2 \partial y} G \;,$$

$$L_{xy^2} = I * \frac{\partial^3}{\partial x \partial y^2} G \;,\; L_{y^3} = I * \frac{\partial^3}{\partial y^3} G$$

Interest point detectors: Detecting blobs by Laplacian of Gaussian filter

- Finding extrema in 2D:
  - Find points that are simultaneous extrema in the x and y direction.
  - We can solve this in one go by looking for extrema of the Laplacian of Gaussian filter $\nabla^2 L(x,y;\sigma) = L_{x^2} + L_{y^2}$

    $$\nabla(\nabla^2 L(x,y;\sigma)) = 0$$

  - Bright blob: $\nabla^2 L(x,y;\sigma) < 0$ Dark blob: $\nabla^2 L(x,y;\sigma) > 0$

- Discrete implementation: Extrema search in 2D
  - Bright blob: $\nabla^2 L(x,y;\sigma) <$ than all neighbor pixels
  - Dark blob: $\nabla^2 L(x,y;\sigma) >$ than all neighbor pixels
  - 4-neighbors or 8-neighbors
  - Keep only blobs where $|\nabla^2 L(x,y;\sigma)| >$ threshold value

Interest point detectors: Detecting blobs

Interest point detectors: Detecting corners with Harris corner detector

- Harris corners are defined as points of local maxima of the Harris corner measure:

$$R(x,y;\sigma) = \det(\mathbf{A}) - \alpha\,\mathrm{trace}(\mathbf{A})^2$$

- Where

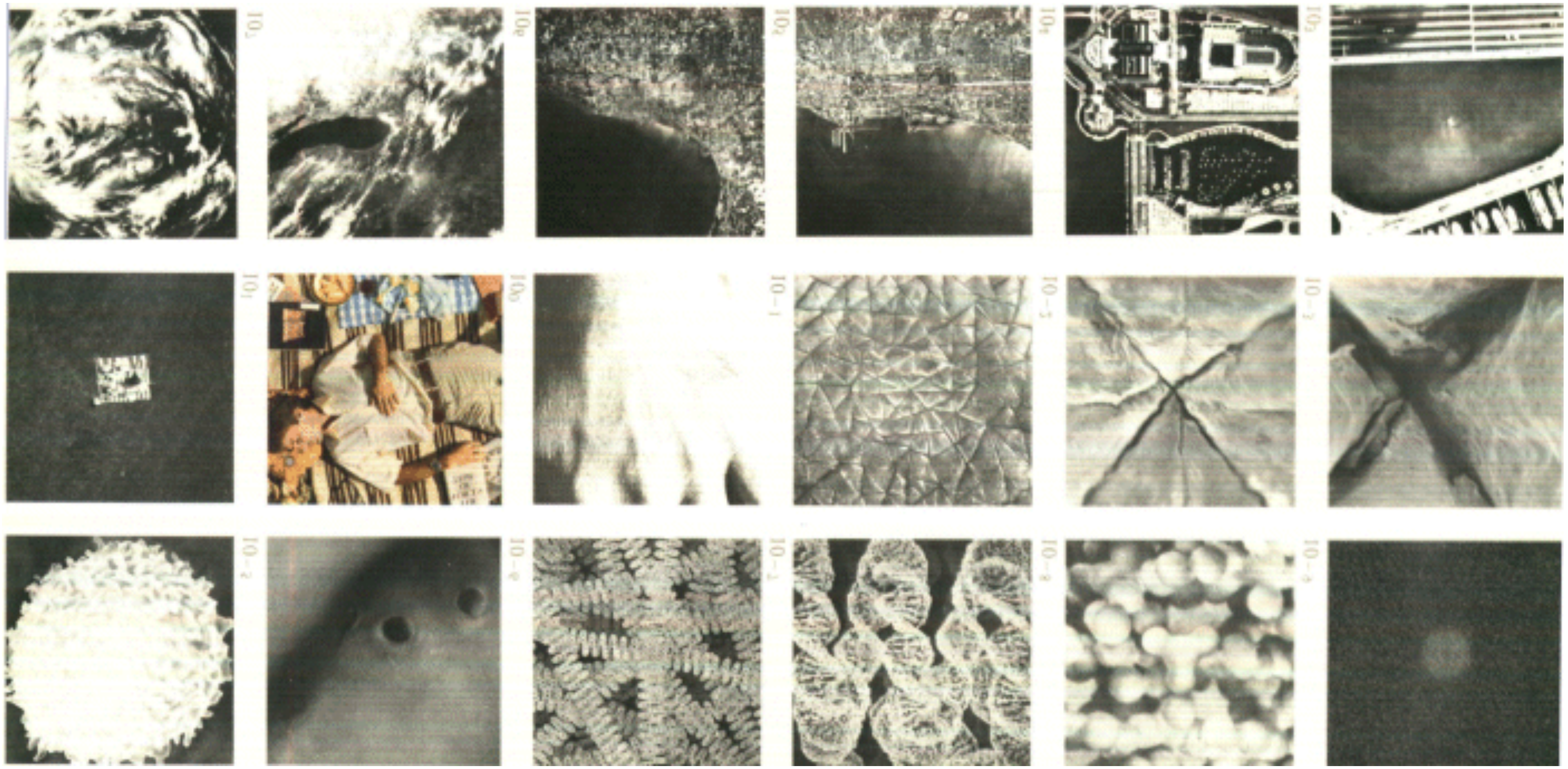
$$\mathbf{A}(x,y;\sigma) = \begin{bmatrix} G(x,y;k\sigma) * L_x^2(x,y;\sigma) & G(x,y;k\sigma) * (L_x(x,y;\sigma)L_y(x,y;\sigma)) \\ G(x,y;k\sigma) * (L_x(x,y;\sigma)L_y(x,y;\sigma)) & G(x,y;k\sigma) * L_y^2(x,y;\sigma) \end{bmatrix}$$

- It encode gradient information under the window G.
- *R* obtain maxima at intensity corners, but also for other structures.
- This version is referred to as the improved or fixed scale Harris corner detector.

Interest point detectors:
Detecting Harris corners (fixed scale)

Multi-Scale analysis
An Introduction to Scale Space Theory

Measurements
What do we measure?



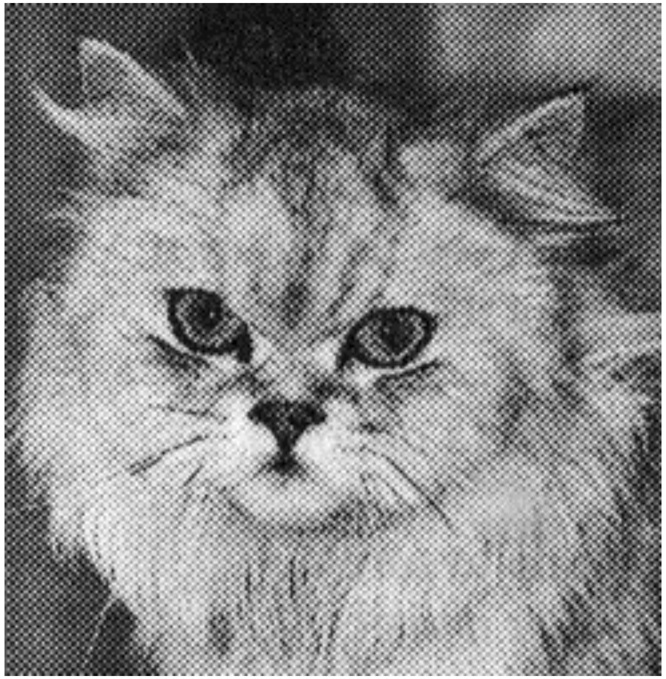
Slide by Arjan Kuijper, 2003

Observations

- Objects have a size / scale.
- Objects consists of objects of various sizes.
    - They contain several scales.
- Objects are measured by some device.
    - Cameras, the eye, …
- Devices are finite.
    - They have a minimum and a maximum detection range: the inner and outer scale. They determine the spatial resolution.
- The device must allow multi-scale structures.
    - It has to respect the various sizes of the object. The inner scale isn't always the best scale.

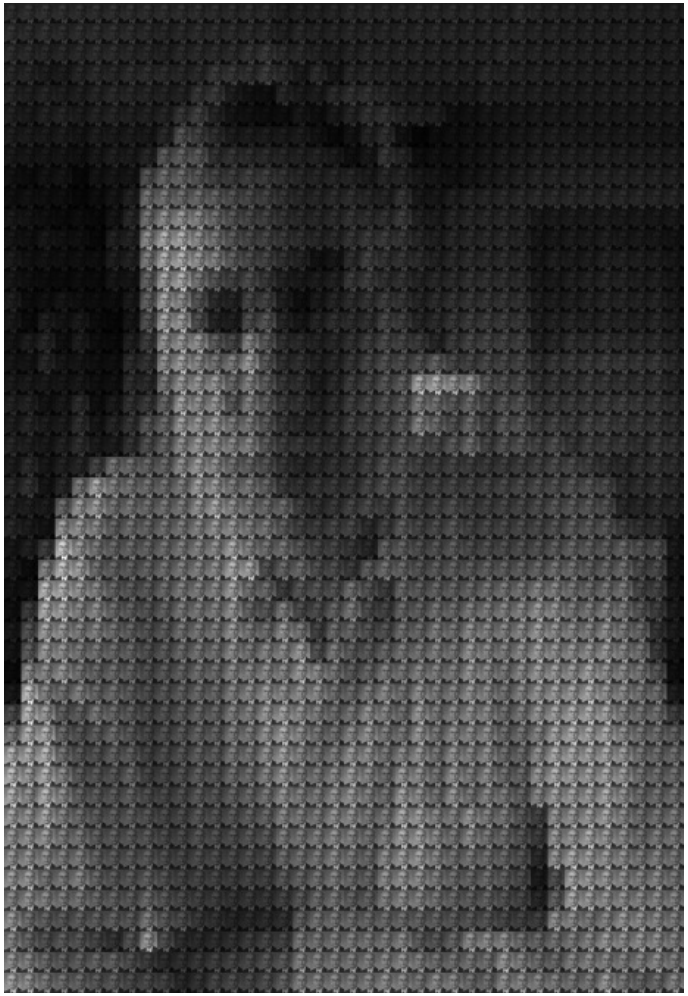
The visual system

- We see multi-scale:
  - The images only contain two values (black and white).
  - We regard them as grey level images, or see structure.

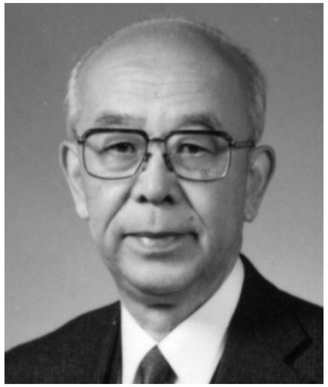


Slide by Arjan Kuijper, 2003

The visual system
(Founding fathers of scale space theory)



Jan J. Koenderink

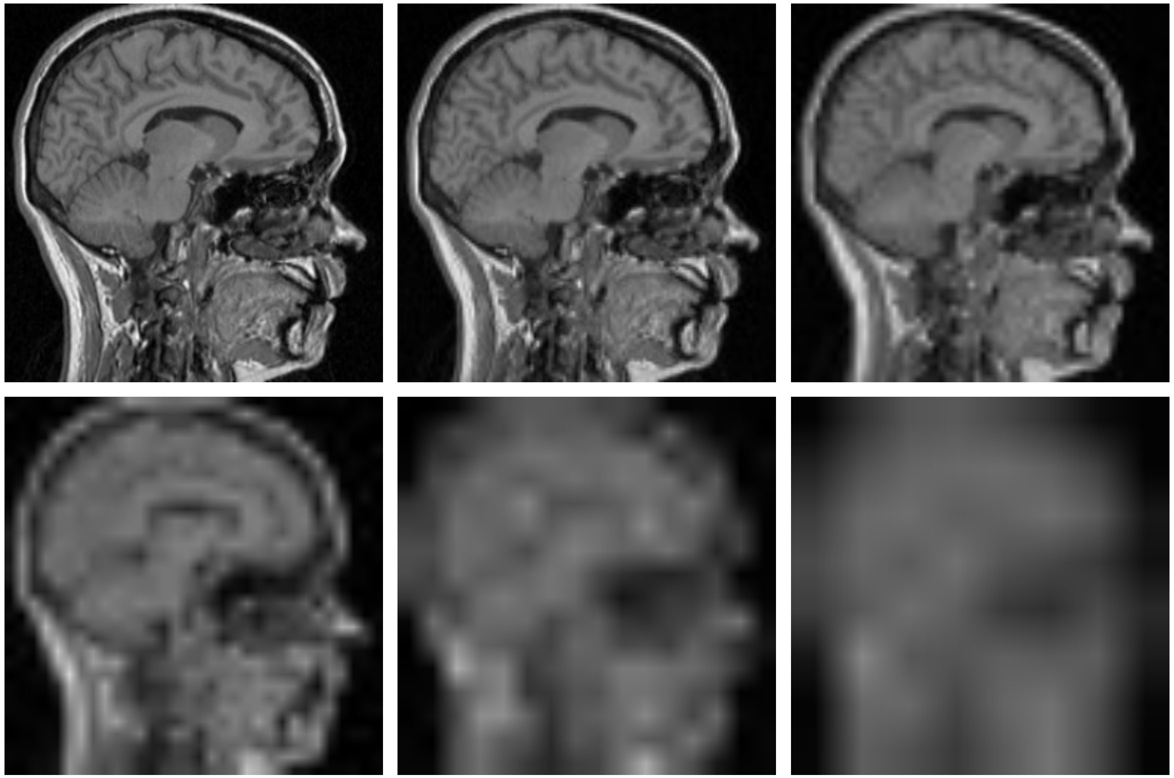


Taizo Iijima

Slide by Arjan Kuijper, 2003

To model (II)

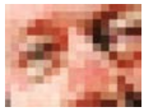
- Don’t trust the resolution.
  - What does a detector of a 3 pixels circular size detect?



Slide by Arjan Kuijper, 2003

To model (III)

- Don’t trust the grid.



Slide by Arjan Kuijper, 2003

To model a device

- It has finite resolution.
  - Infinite resolution is impossible.
- Take uncommitted observations
  - There is no bias, no knowledge, no memory.
- We know nothing.
  - At least, at the first stage. Refine later on.
- Allow different scales.
  - There’s more than just pixels .
- View them simultaneously.
  - There is no preferred size.
- Noise is part of the measurement.
  - Beware!

Slide by Arjan Kuijper, 2003

Deep structure

The challenge is to understand the image
really on all the levels simultaneously,
and not as an unrelated set of derived images
at different levels of blurring.

Jan Koenderink (1984)

Our choice of model: Linear scale space

- The scale space of *I* is a 1-parameter family

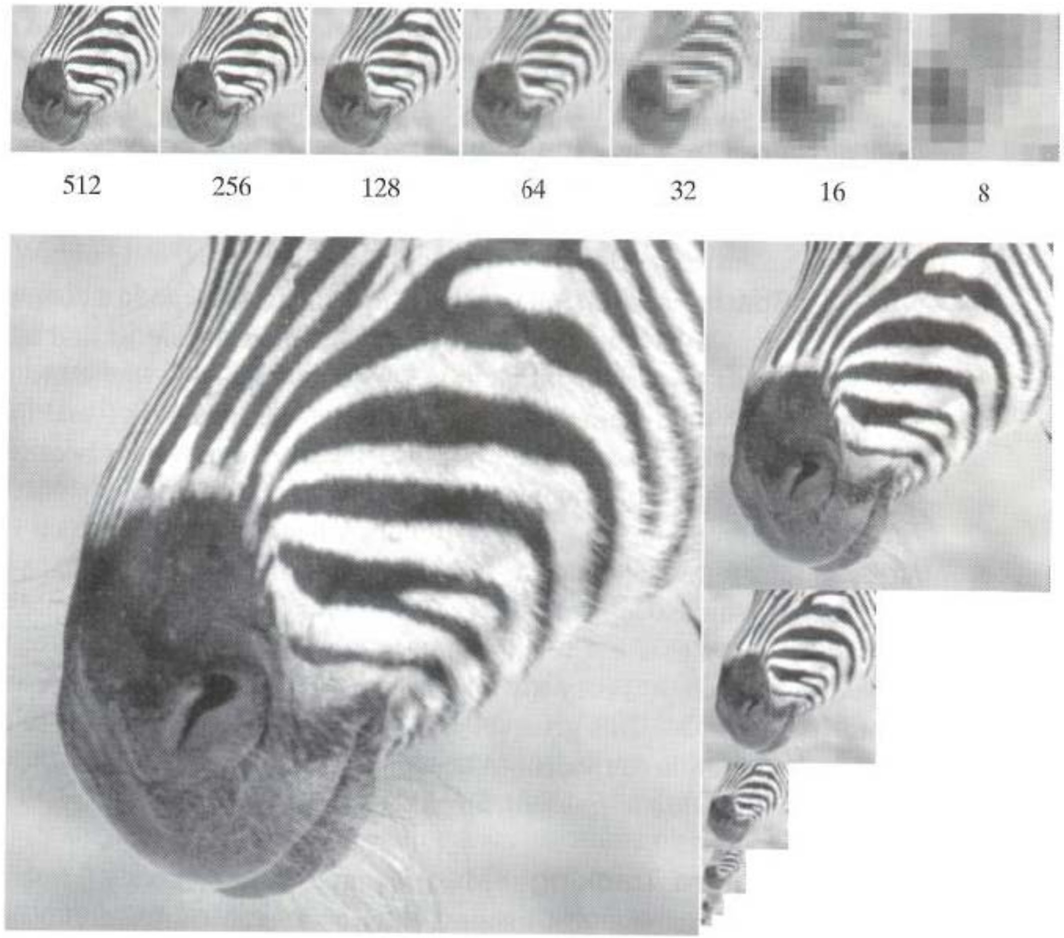
  $$L(x,y;\sigma) = (I * G)(x,y;\sigma)$$

  $$L(x,y;\sigma = 0) \equiv I(x,y)$$

- Scale is given by σ and is an important parameter in Computer Vision algorithms

- As the scale increases details in the image disappear and we focus on the large scale structures that are left.

Multi-scale versus multi-resolution: The Gaussian pyramid



Interest point detectors: Detecting blobs by Difference of Gaussians (DoG)

- Difference of Gaussians (DoG) is an approximation to the Laplacian of Gaussian filter

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

$$= L(x,y;k\sigma) - L(x,y;\sigma) \approx \nabla^2 L(x,y;\sigma)$$

- Either subtract Gaussian filters prior to filtering or subtract filtered images.
- As before look for extrema in $D(x,y;\sigma)$
- In a multi-scale setting compute the scale space images with factor *k* increase in scale and search for extrema in scale space. Can be combined with a Gaussian pyramid for increased processing speed.

Summary

- Matching salient points
- Filtering 101
- Interest point detection
  - Blobs (DoG, Laplacian)
  - Corners (Harris corners)
- Scale invariance: We need to consider structure at multiple scales (Scale space theory)

Literature

Reading material:

- Forsyth and Ponce: Ch. 7.1 – 7.2, 7.7 (Filtering basics)
- Lowe IJCV 2004, Sec. 1 – 3 (DoG, Laplacian)
- Schmid, Mohr and Bauckhage IJCV 2000, Sec. 1 – 2 (Harris corner)
- Lindeberg 1996 (Scale-space theory)

Additional material:

- C. Harris and M. Stephens: A Combined Corner and Edge Detector. 4th Alvey Vision Conference, 147—151, 1988.

Lets have a break and then work with the assignment

Mandatory assignment 1:
Feature extraction

- Build and experiment with interest point detectors
- Find matching points between two images