RuSSIR
Russian Summer School in Information Retrieval 2008

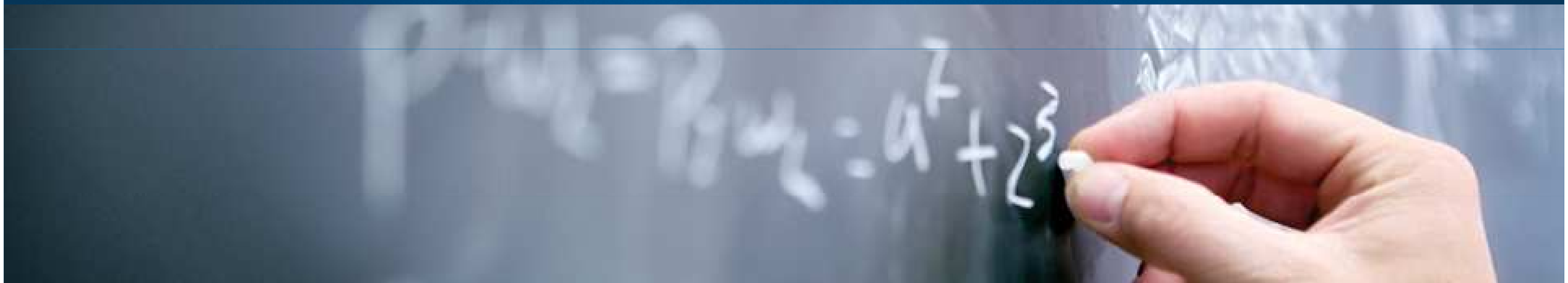
Content Based Image Retrieval

Natalia Vassilieva
nvassilieva@hp.com
HP Labs Russia

Tutorial outline

- Lecture 1
  - Introduction
  - Applications
- Lecture 2
  - Performance measurement
  - Visual perception
  - Color features
- Lecture 3
  - Texture features
  - Shape features
  - Fusion methods
- Lecture 4
  - Segmentation
  - Local descriptors
- Lecture 5
  - Multidimensional indexing
  - Survey of existing systems

Lecture 4
Segmentation
Local descriptors



Lecture 4: Outline

- Segmentation
  - Detection of discontinuities
  - Thresholding
  - Region-based segmentation
  - Watershed Segmentation
- Local descriptors
  - SIFT: Scale-invariant feature transform

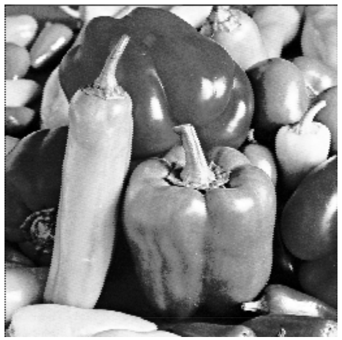
Introduction to segmentation

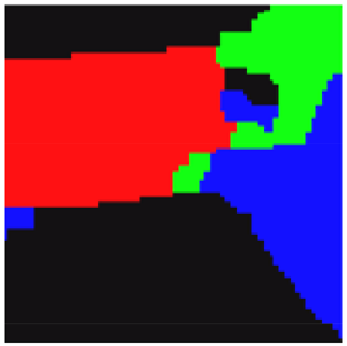
- The main purpose is to find meaningful regions with respect to a particular application
  - To detect homogeneous regions
  - To detect edges (boundaries, contours)
- Segmentation of non trivial images is one of the difficult task in image processing. Still under research
- Applications of image segmentation include
  - Objects in a scene (for object-based retrieval)
  - Objects in a moving scene *(MPEG4)*
  - Spatial layout of objects (Path planning for a mobile robots)

Principal approaches

- Edge based methods
  - Based on discontinuity: ex. to partition an image based on abrupt changes in intensity
- Region based methods
  - Based on similarity: to partition an image into regions that are similar according to a set of predefined criteria

Solution can be based on intensity, texture, color, motion, etc.



Lecture 4: Outline

- Segmentation
  - Detection of discontinuities
  - Thresholding
  - Region-based segmentation
  - Watershed Segmentation
- Local descriptors
  - SIFT: Scale-invariant feature transform

Detection of discontinuities

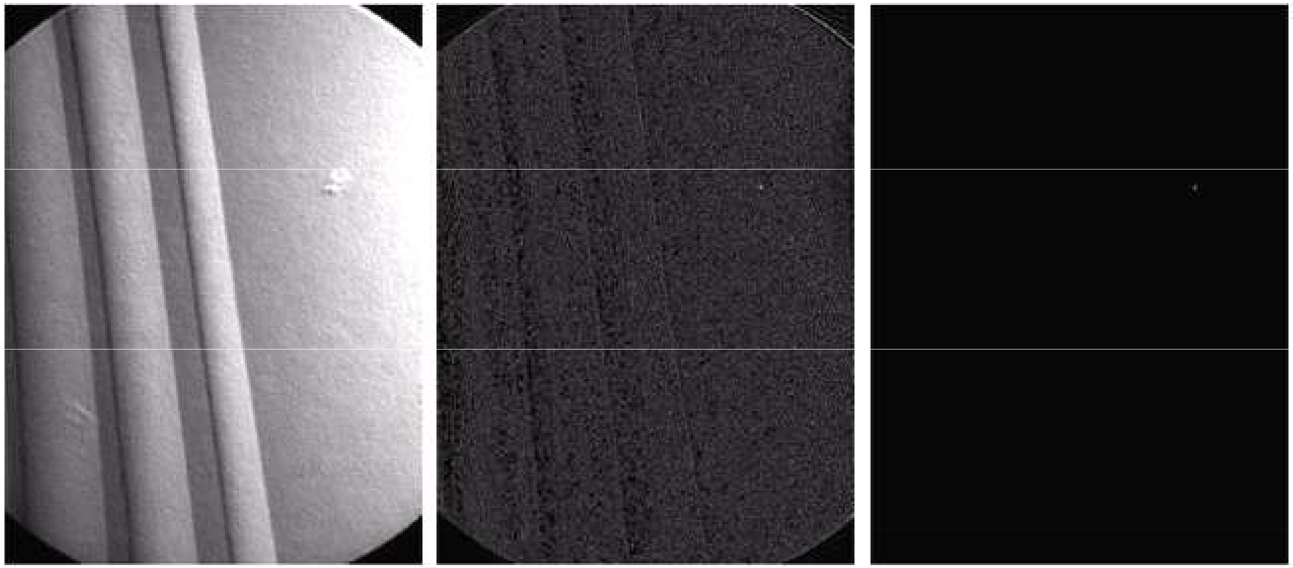
- 3 basic types of gray-level discontinuities:
  - points , lines , edges
- The common way is to run a mask through the image

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

$$R = w_1 z_1 + w_2 z_2 + \ldots + w_9 z_9 = \sum_{i=1}^{9} w_i z_i$$

Point detection

- A point has been detected if $|R| \geq T$,
  - T is a nonnegative threshold



-1	-1	-1
-1	8	-1
-1	-1	-1

Line detection

−1	−1	−1
2	2	2
−1	−1	−1

Horizontal

−1	−1	2
−1	2	−1
2	−1	−1

+45°

−1	2	−1
−1	2	−1
−1	2	−1

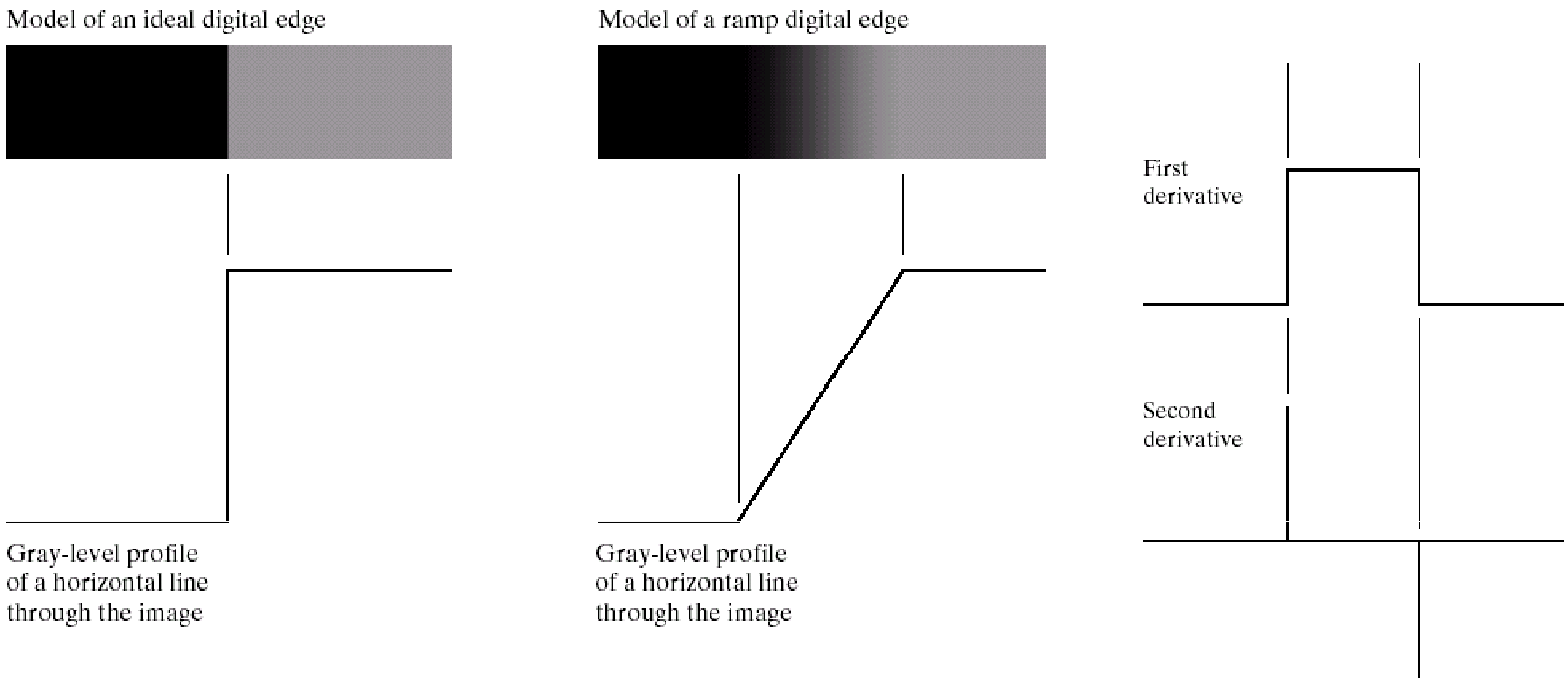
Vertical

2	−1	−1
−1	2	−1
−1	−1	2

−45°

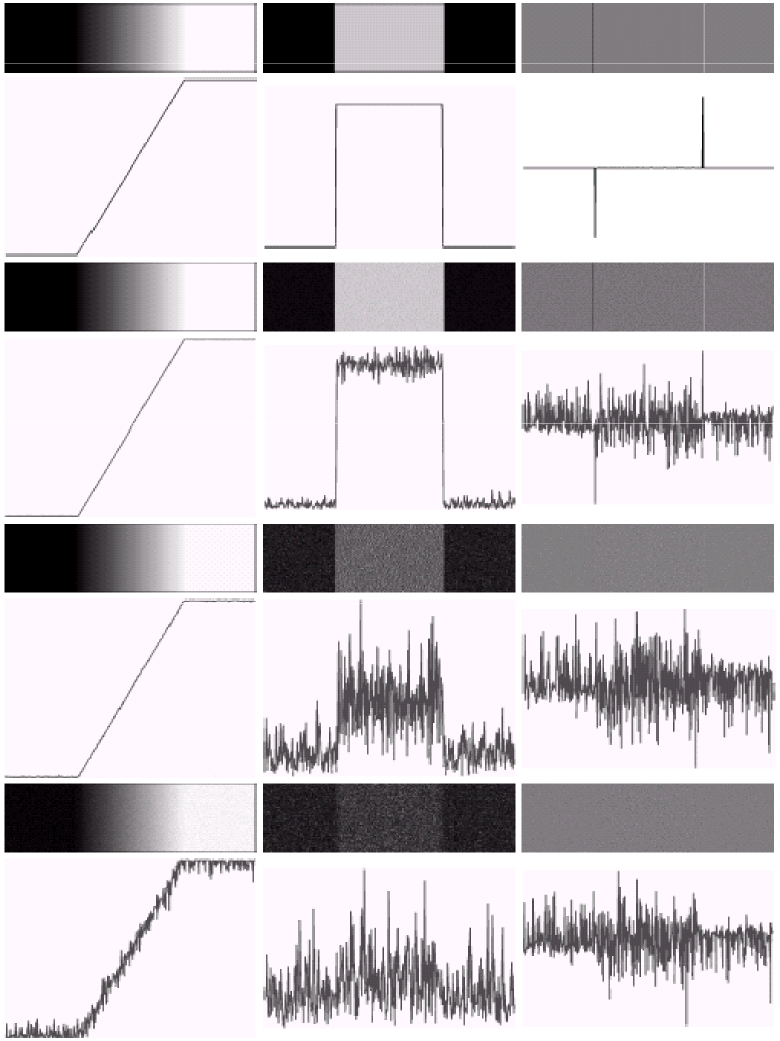
- If $|R_i| > |R_j|$ for all $j \neq i$ – the point is within line i.
- Use one mask to detect lines of a given direction

Edge Detection



- First derivative – detect if a point is on the edge
- Second derivative – detect the midpoint of the edge (zero-crossing property)

Edge detection in noisy images



- Examples of a ramp edge corrupted by random Gaussian noise of mean 0 and σ = 0.0, 0.1, 1.0 and 10.0.

Edge detection: calculating derivatives

- First derivative: magnitude of the gradient

$$\nabla \mathrm{f} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}, \quad |\nabla \mathrm{f}| = [G_x^2 + G_y^2]^{1/2} = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{1/2} \approx |G_x| + |G_y|$$

- To calculate: apply gradient masks

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

Roberts:

$$G_x = (z_9 - z_5)$$
$$G_y = (z_8 - z_6)$$

-1	0
0	1

0	-1
1	0

Prewitt:

$$G_x = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3)$$
$$G_y = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7)$$

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

Sobel:

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)$$
$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Example

a b
c d

FIGURE 10.10 (a) Original image. (b) $|G_x|$, component of the gradient in the x-direction. (c) $|G_y|$, component in the y-direction. (d) Gradient image, $|G_x| + |G_y|$.

Edge detection: calculating derivatives

- Second derivative: Laplacian

$$\nabla^2 f = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2}$$

- To calculate: apply laplacian masks

0	−1	0		−1	−1	−1
−1	4	−1		−1	8	−1
0	−1	0		−1	−1	−1

$$\nabla^2 f = 4z_5 - (z_2 + z_4 + z_6 + z_8)$$

$$\nabla^2 f = 8z_5 - (z_1 + z_2 + z_3 + z_4 + z_6 + z_7 + z_8 + z_9)$$

Edge detection: Laplacian of Gaussian

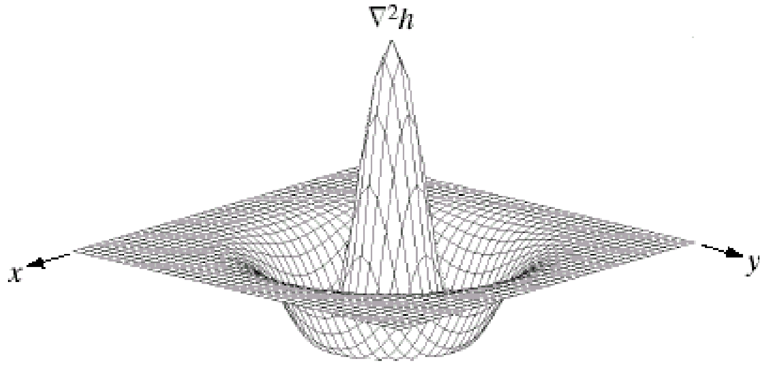
- Laplacian combined with smoothing as a precursor to find edges via zero-crossing.

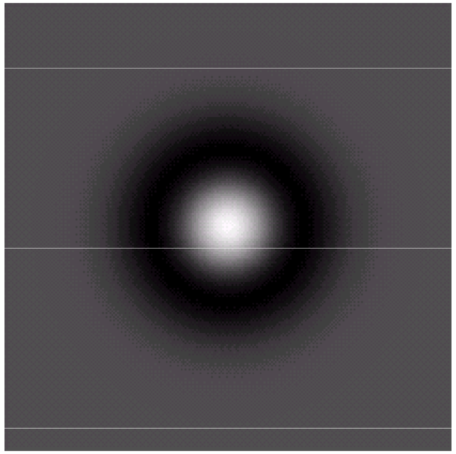
$$h(r) = -e^{-\frac{r^2}{2\sigma^2}}$$

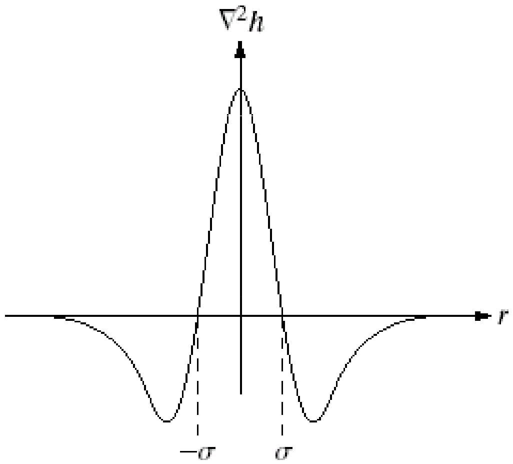
where $r^2 = x^2+y^2$, and
σ is the standard deviation

$$\nabla^2 h(r) = -\left[\frac{r^2 - \sigma^2}{\sigma^4}\right] e^{-\frac{r^2}{2\sigma^2}}$$

Mexican hat







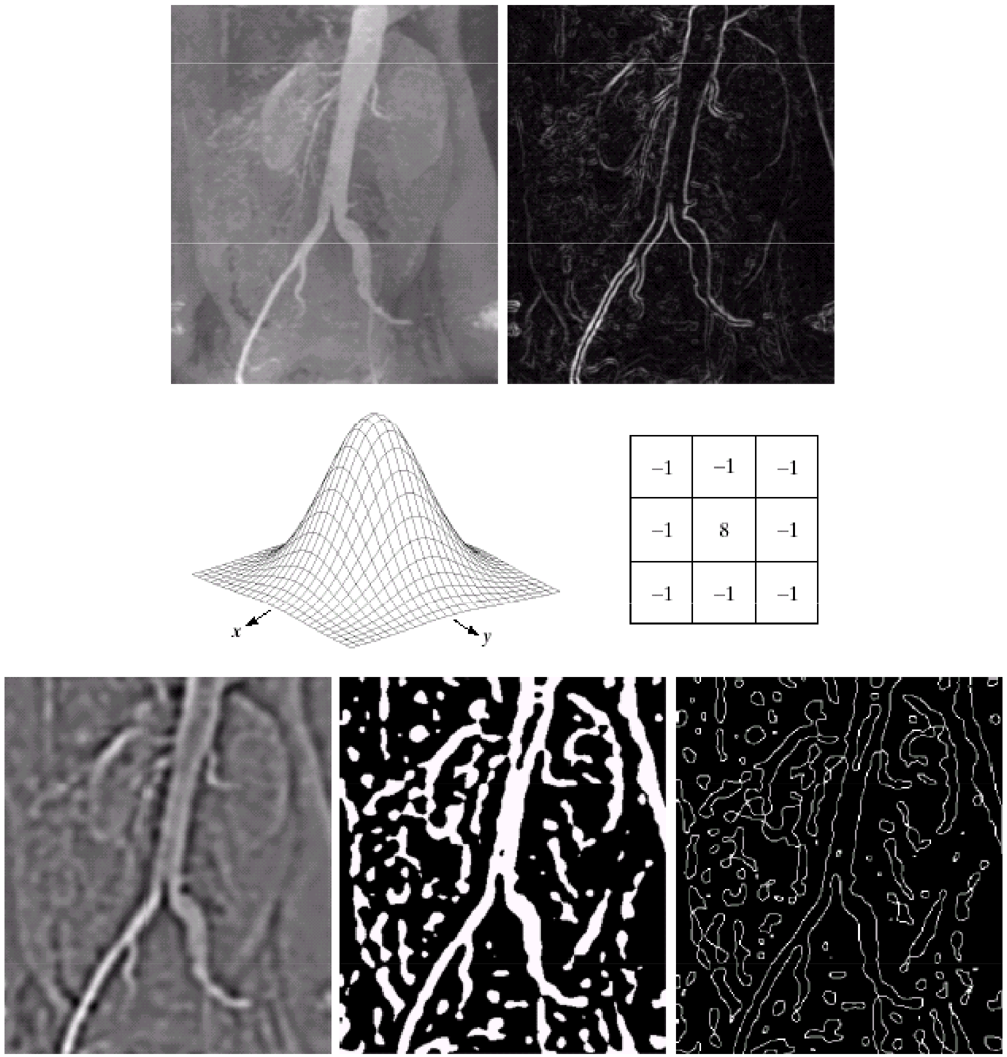
0	0	–1	0	0
0	–1	–2	–1	0
–1	–2	16	–2	–1
0	–1	–2	–1	0
0	0	–1	0	0

a b
c d

FIGURE 10.14 Laplacian of a Gaussian (LoG). (a) 3-D plot. (b) Image (black is negative, gray is the zero plane, and white is positive). (c) Cross section showing zero crossings. (d) 5×5 mask approximation to the shape of (a).

the coefficient must be sum to zero

Example



a) Original image

b) Sobel Gradient

c) Spatial Gaussian smoothing function

d) Laplacian mask

e) LoG

f) Threshold LoG

g) Zero crossing

Edge linking

- Local processing

$$|\nabla f(x,y) - \nabla f(x_0,y_0)| \le E, \quad |\alpha(x,y) - \alpha(x_0,y_0)| \le A, \quad \alpha(x,y) = \mathrm{arctg}\left(\frac{G_y}{G_x}\right).$$

- Global processing via the Hough transform
  - Looking for lines between edge points

- Global processing via the Graph-based techniques
  - Edge points are graph vertexes
  - Looking for optimal path in graph

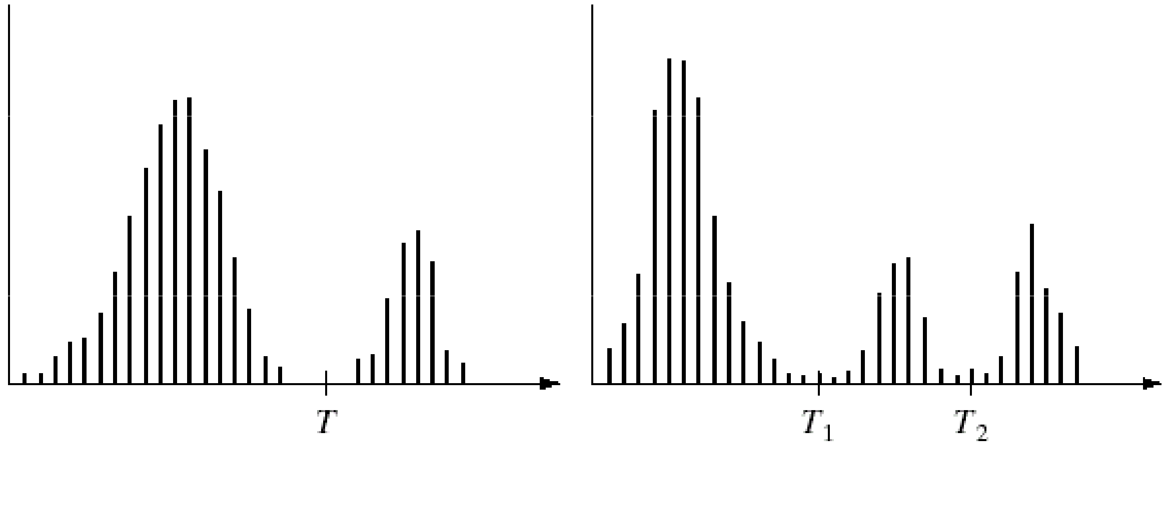
Lecture 4: Outline

- Segmentation
  - Detection of discontinuities
  - Thresholding
  - Region-based segmentation
  - Watershed Segmentation
- Local descriptors
  - SIFT: Scale-invariant feature transform

Thresholding

image with dark background and a light object

image with dark background and two light objects



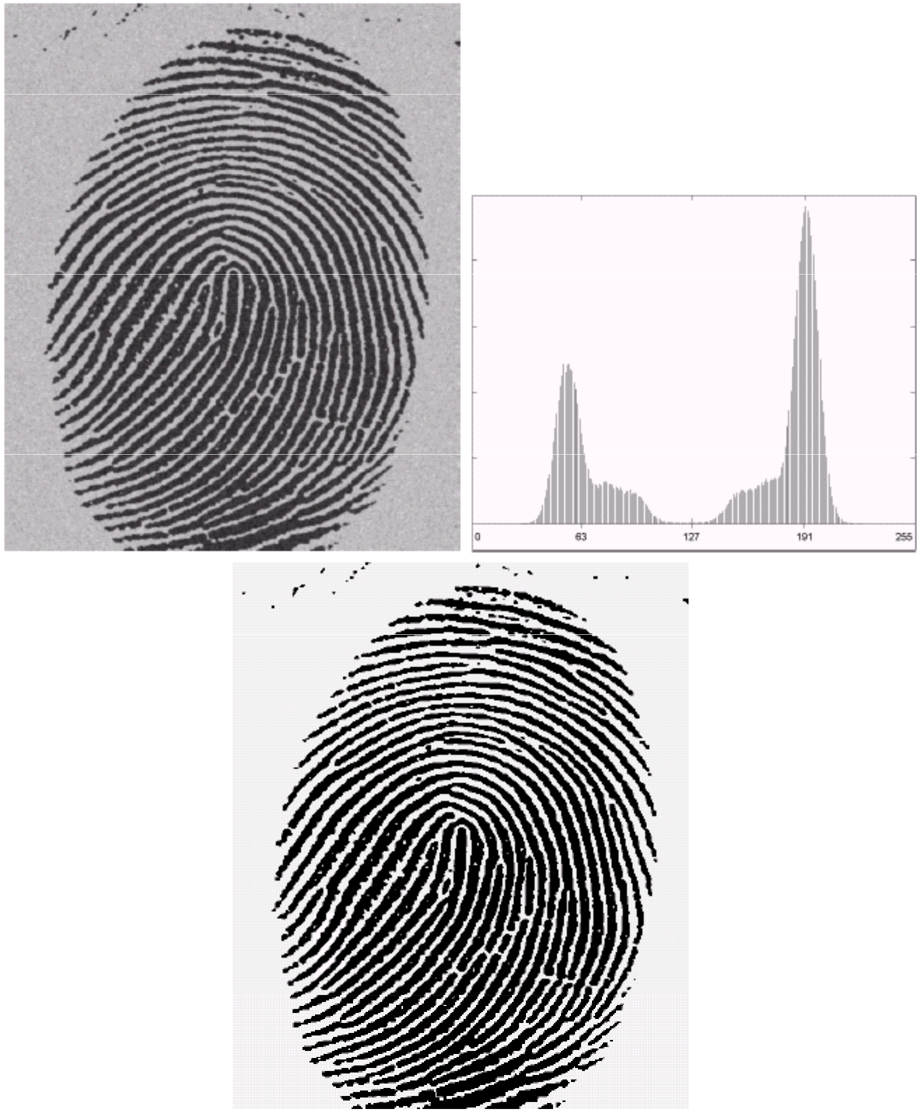
(a) Gray-level histograms that can be partitioned by (a) a single threshold, and (b) multiple thresholds.

- Global – when T is the same for all points of the image
- Local or Dynamic – when T depends on (x,y)
- Adaptive – when T depends on I(x,y)

Global thresholding

- Based on visual inspection of histogram
- Automatically
  - Select an initial estimate T_0.
  - Segment the image using T_0: regions G1 and G2 consisting of pixels with gray level values $>T_0$ and $\leq T_0$
  - Compute the average gray level values $\mu 1$ and $\mu 2$ for the pixels in regions G1 and G2
  - $T_1 = 0.5\ (\mu 1 + \mu 2)$
  - Repeat until $| T_i - T_{i+1}| < T_{th}$

Global thresholding: example



a b
c

FIGURE 10.29 (a) Original image. (b) Image histogram. (c) Result of segmentation with the threshold estimated by iteration. (Original courtesy of the National Institute of Standards and Technology.)

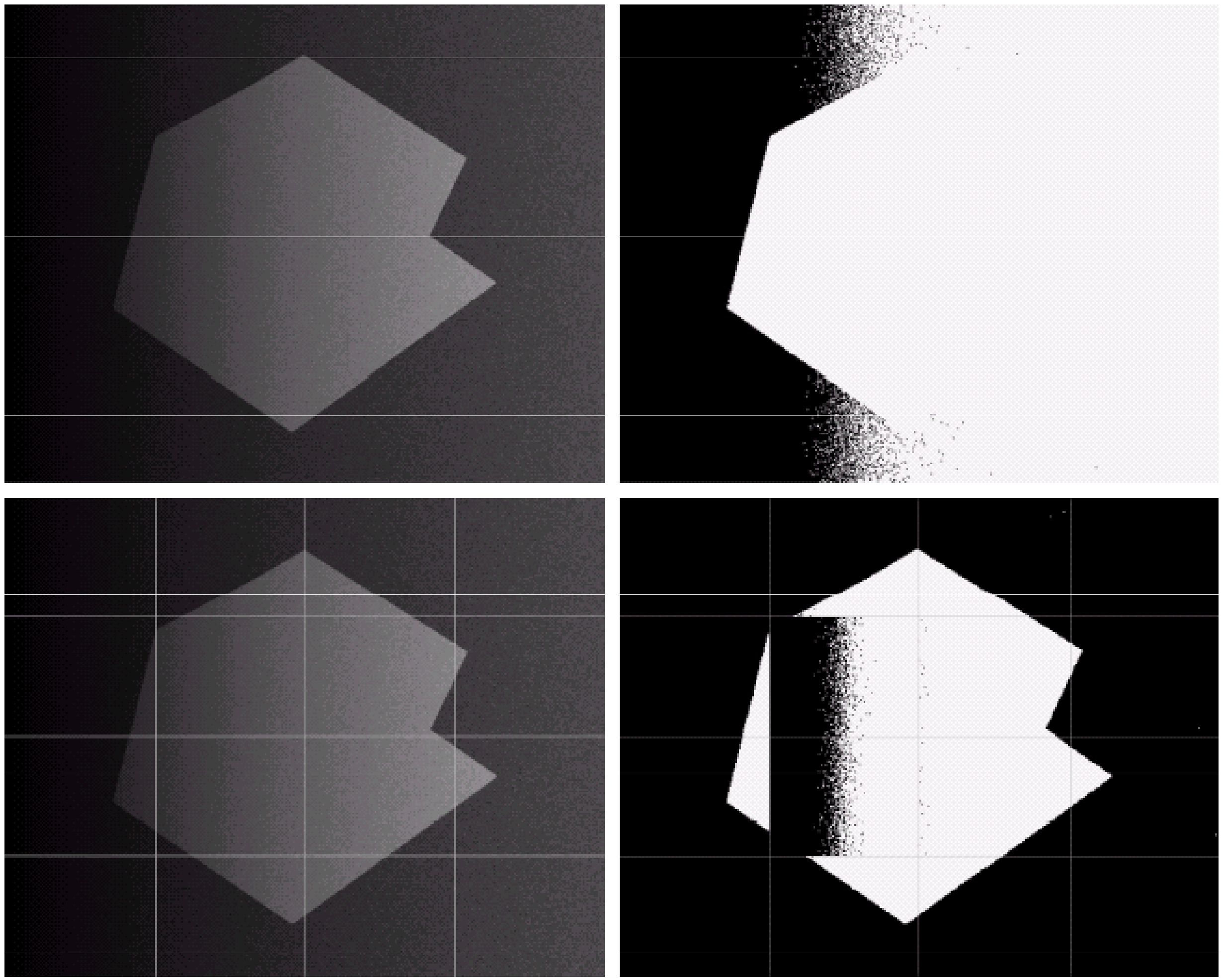
$T_{th} = 0$

3 iterations with result T = 125

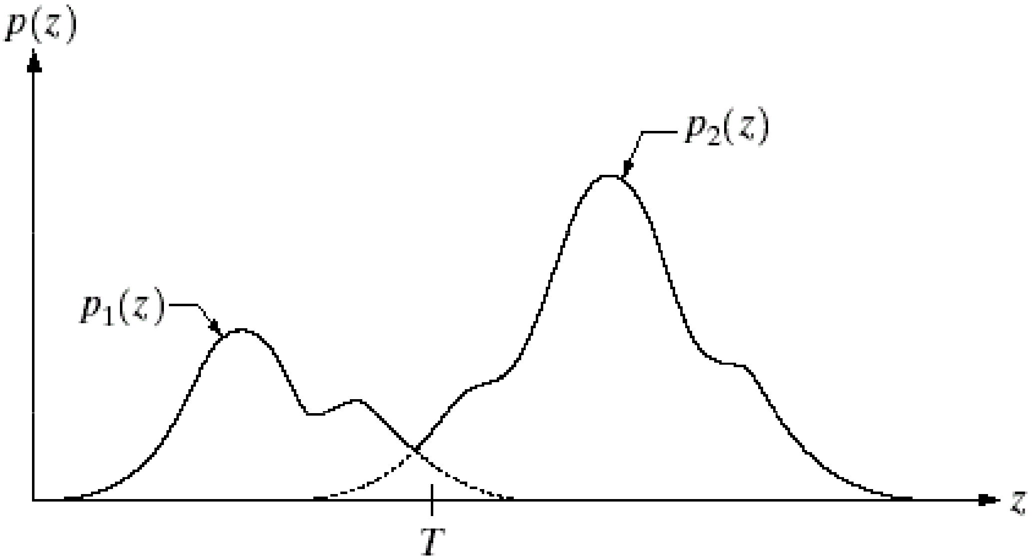
Adaptive thresholding

a b
c d

FIGURE 10.30 (a) Original image. (b) Result of global thresholding. (c) Image subdivided into individual subimages. (d) Result of adaptive thresholding.



Optimal thresholding



$$p(z) = P_1 p_1(z) + P_2 p_2(z)$$

$$P_1 + P_2 = 1$$

$$E_1(T) = \int_{-\infty}^{T} p_2(z)dz, \quad E_2(T) = \int_{-\infty}^{T} p_1(z)dz$$

$$E(T) = P_2 E_1(T) + P_1 E_2(T) \quad \Rightarrow \quad P_1 p_1(T) = P_2 p_2(T)$$

Multispectral thresholding

a b c

FIGURE 10.39 (a) Original color image shown as a monochrome picture. (b) Segmentation of pixels with colors close to facial tones. (c) Segmentation of red components.

Lecture 4: Outline

- Segmentation
  - Detection of discontinuities
  - Thresholding
  - Region-based segmentation
  - Watershed Segmentation
- Local descriptors
  - SIFT: Scale-invariant feature transform

Region-based segmentation

- A segmentation is the partition of an image R into sub-regions {Ri} such that

$$\bigcup_{i=1}^{n} R_i = R; \quad R_i \cap R_j = \varnothing$$

$$\text{s. t. } R_i \text{ is a connected region}$$

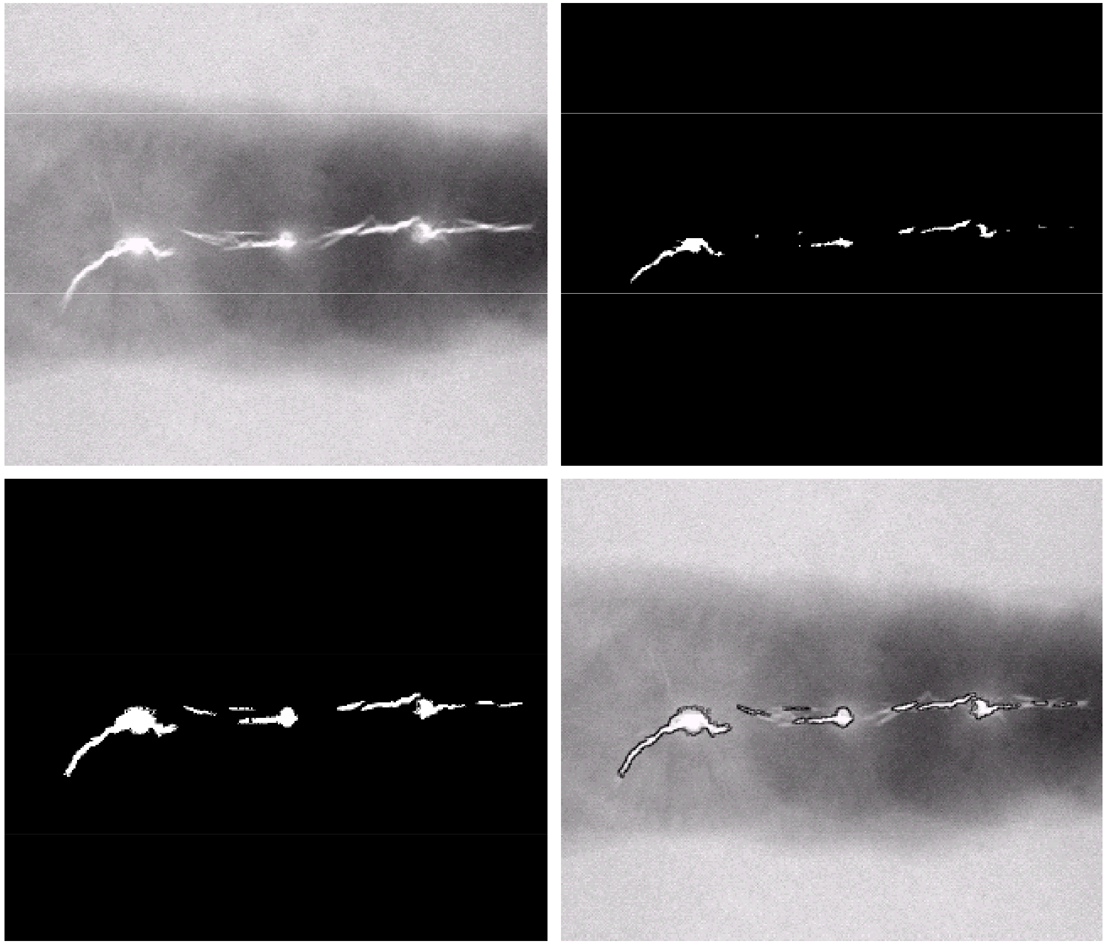
- A region can be defined by a predicate *P* such that *P*(*Ri*) = TRUE if all pixels within the region satisfy a specific property.
- *P*(*Ri* ∩*Rj*) = FALSE for *i* ≠*j*.

Region-based segmentation

- Region growing

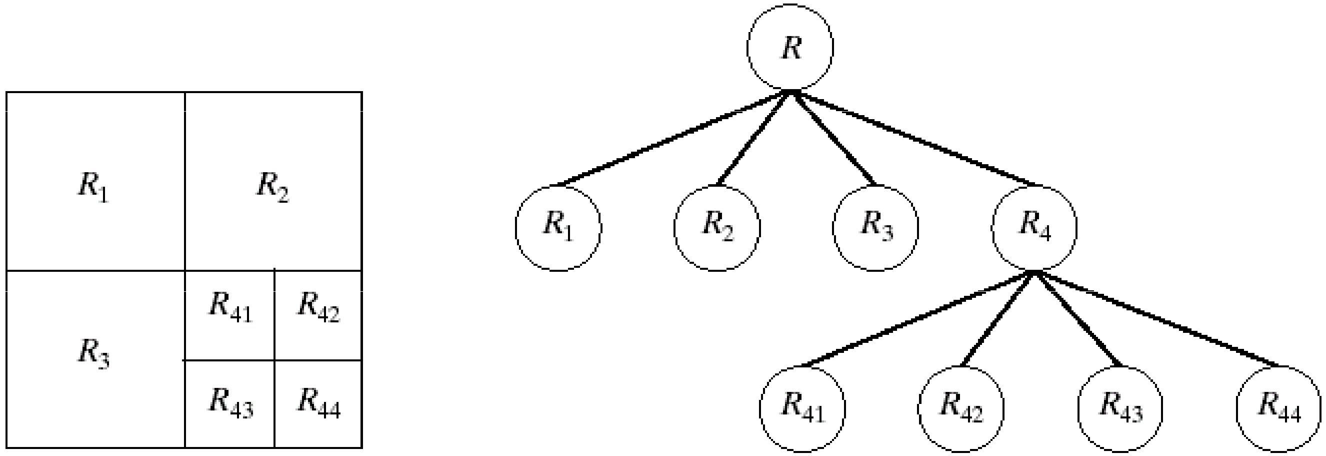
a b
c d

(a) Image showing defective welds. (b) Seed points. (c) Result of region growing. (d) Boundaries of segmented defective welds (in black). (Original image courtesy of X-TEK Systems, Ltd.).



Region-based segmentation

- Region splitting and merging



1. Split into 4 disjoint quadrants any region Ri for which P(Ri) = FALSE
2. Merge any adjacent region Rj and Rk for which P(Ri $\cup$ Rk) = TRUE
3. Stop when no further merging or splitting is possible.

Example

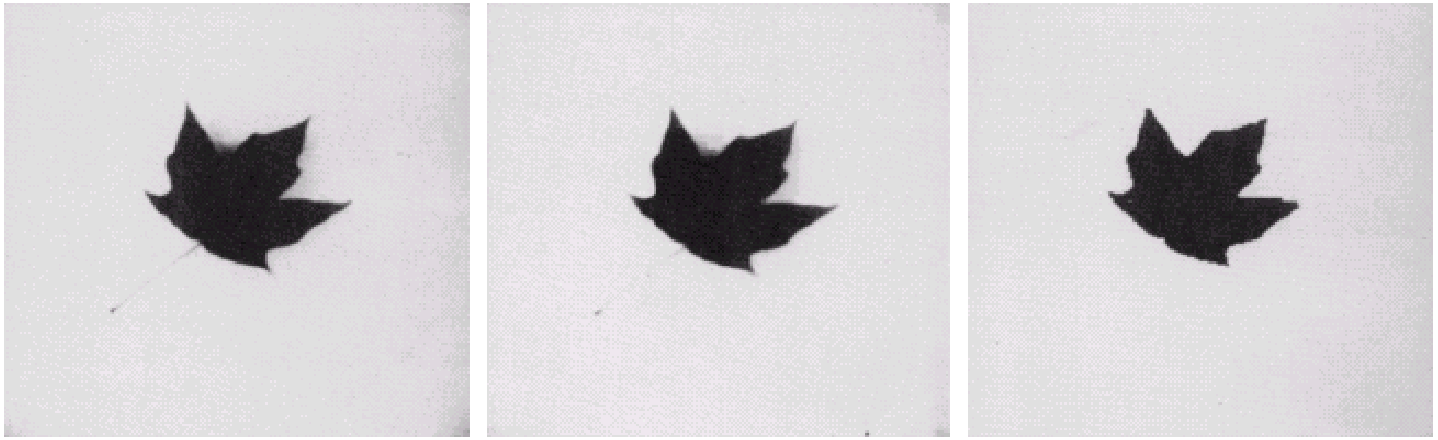


FIGURE 10.43 (a) Original image. (b) Result of split and merge procedure. (c) Result of thresholding (a).

$P(R_i)$ = TRUE if at least 80% of the pixels in R_i have the property $|z_j - m_i| \le 2\sigma_i$, where

z_j is the gray level of the j^{th} pixel in R_i
m_i is the mean gray level of that region
σ_i is the standard deviation of the gray levels in R_i

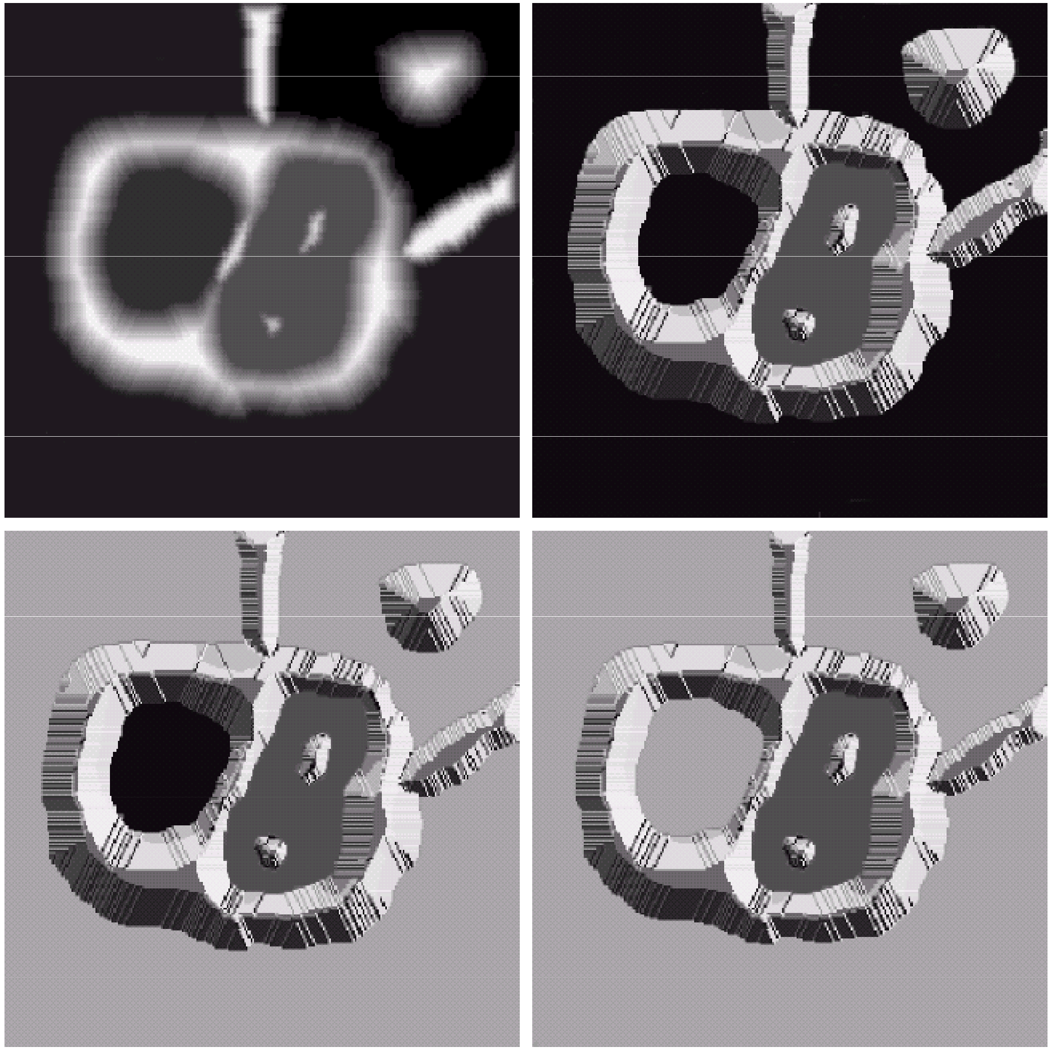
Lecture 4: Outline

- Segmentation
  - Detection of discontinuities
  - Thresholding
  - Region-based segmentation
  - Watershed Segmentation
- Local descriptors
  - SIFT: Scale-invariant feature transform

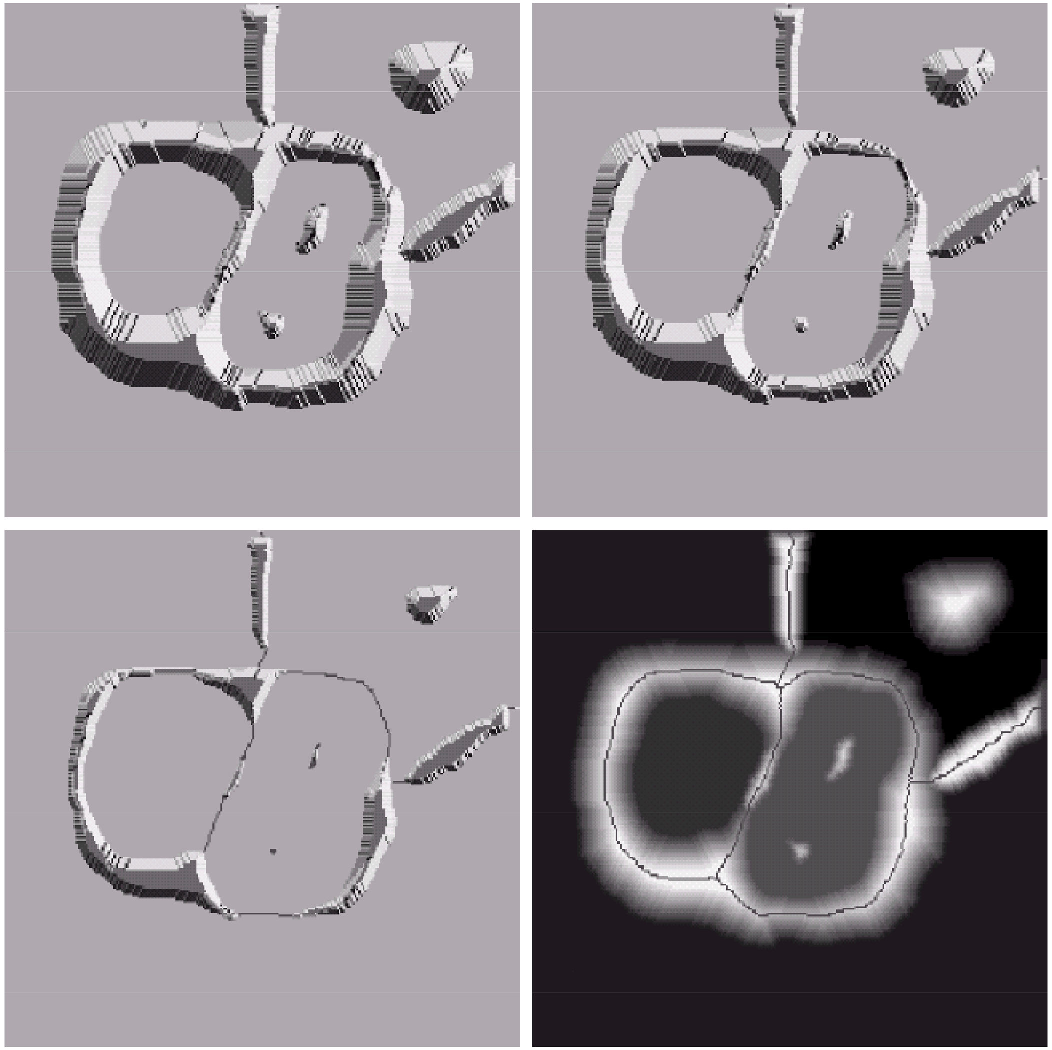
Watersheds Segmentation

a b
c d

FIGURE 10.44
(a) Original image.
(b) Topographic view. (c)–(d) Two stages of flooding.



Watersheds Segmentation



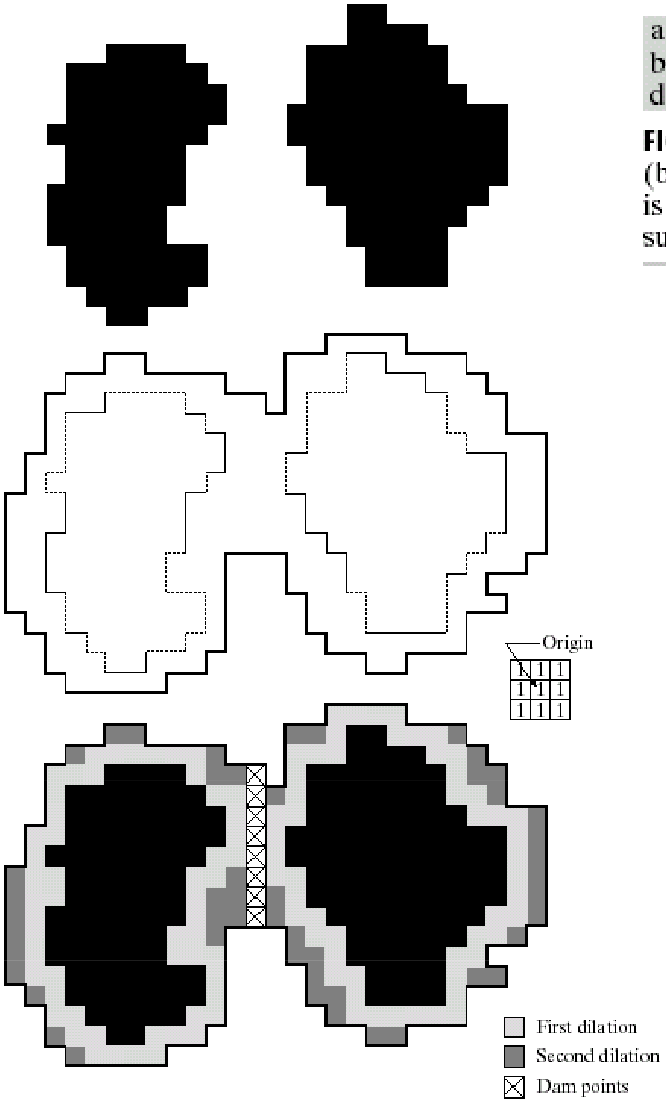
e	f
g	h

FIGURE 10.44 *(Continued)* (e) Result of further flooding. (f) Beginning of merging of water from two catchment basins (a short dam was built between them). (g) Longer dams. (h) Final watershed (segmentation) lines. (Courtesy of Dr. S. Beucher, CMM/Ecole des Mines de Paris.)

Watersheds Segmentation

- A morphological region growing approach.
- Seed points:
  - local minima points
- Growing method:
  - Dilation
- Predicates
  - Similar gradient values
- Sub-region boundary
  - Dam building
- To avoid over-segmentation
  - Use markers

Dam Building



a
b c
d

FIGURE 10.45 (a) Two partially flooded catchment basins at stage $n - 1$ of flooding. (b) Flooding at stage n, showing that water has spilled between basins (for clarity, water is shown in white rather than black). (c) Structuring element used for dilation. (d) Result of dilation and dam construction.

Watershed Segmentation Example

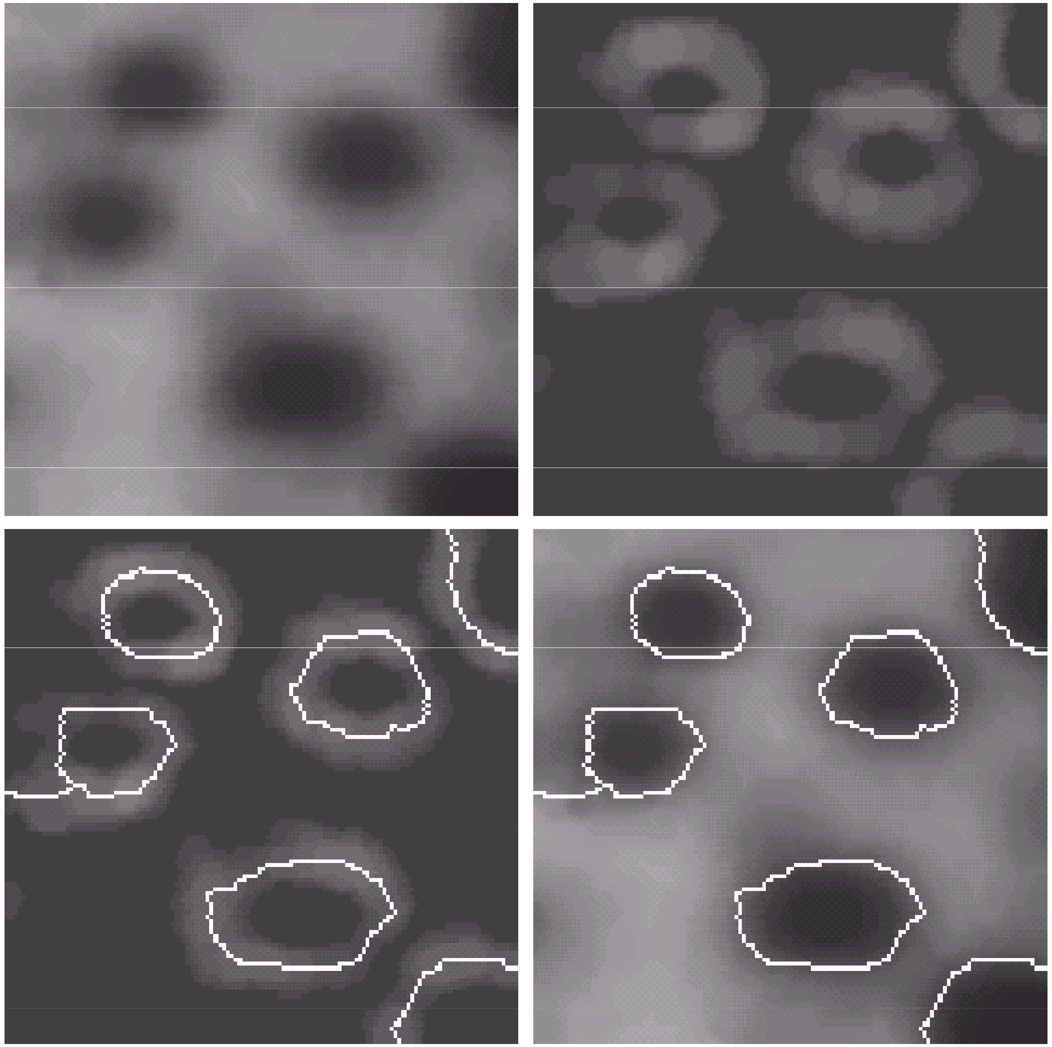
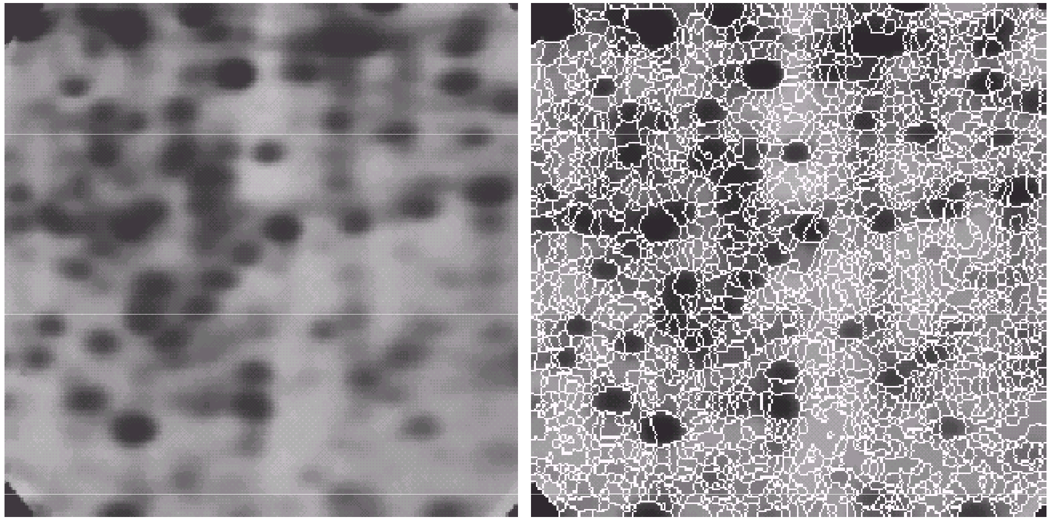


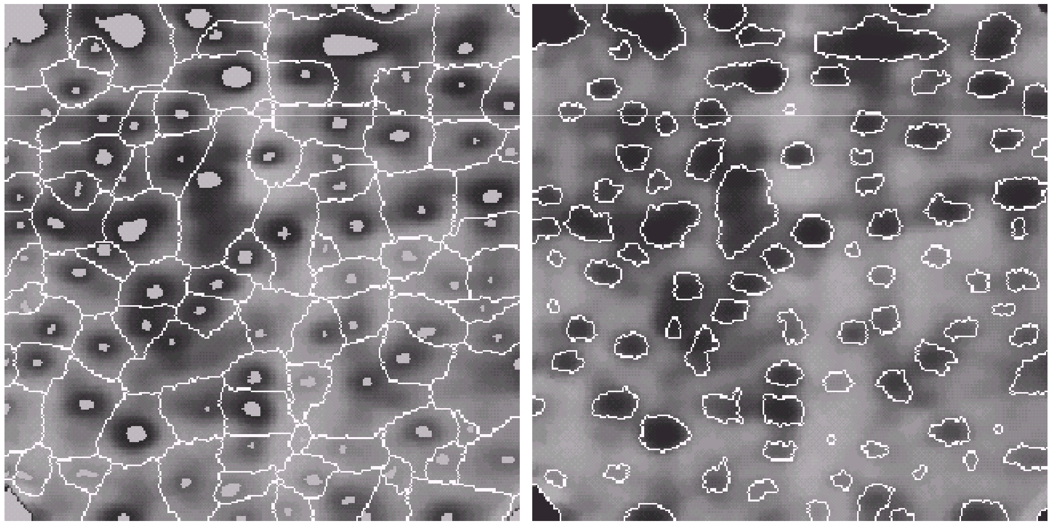
FIGURE 10.46 (a) Image of blobs. (b) Image gradient. (c) Watershed lines. (d) Watershed lines superimposed on original image. (Courtesy of Dr. S. Beucher, CMM/Ecole des Mines de Paris.)

Over-Segmentation and Use of Marker



a b

FIGURE 10.47 (a) Electrophoresis image. (b) Result of applying the watershed segmentation algorithm to the gradient image. Oversegmentation is evident. (Courtesy of Dr. S. Beucher, CMM/Ecole des Mines de Paris.)



a b

FIGURE 10.48 (a) Image showing internal markers (light gray regions) and external markers (watershed lines). (b) Result of segmentation. Note the improvement over Fig. 10.47(b). (Courtesy of Dr. S. Beucher, CMM/Ecole des Mines de Paris.)

Lecture 4: Outline

- Segmentation
  - Detection of discontinuities
  - Thresholding
  - Region-based segmentation
  - Watershed Segmentation
- Local descriptors
  - SIFT: Scale-invariant feature transform

Local descriptors

- Features for local regions in the image
  - Regions obtained by segmentation
  - Regions of interest (RoI) – around interest points (keypoints)

- Interest points: corners, edges and others

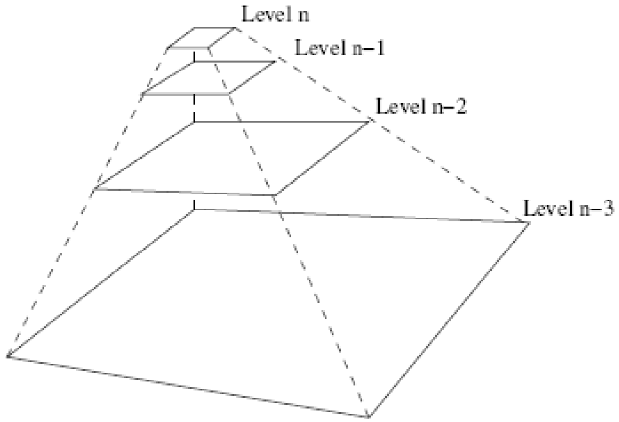
- Keypoints: points in images, which are invariant to image translation, scale and rotation, and are minimally affected by noise and small distortions

- Scale-invariant feature transform (SIFT) by David Lowe

SIFT: main steps

1. Scale-space peak selection
   - Using Difference-of-Gaussians (DoG)
2. Keypoint localization
   - Elimination of unstable keypoints
3. Orientation assignment
   - Based on keypoint local image patch
4. Keypoint descriptor
   - Based upon the image gradients in keypoint local neighbourhood

Scale space



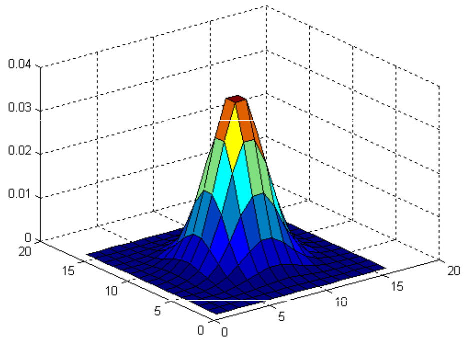
Build an image pyramid with resampling between each level

Difference-of-Gaussian

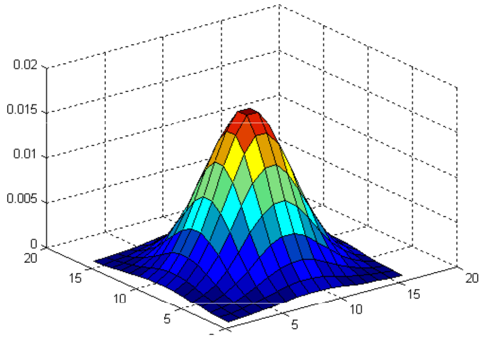
The input image is convolved with Gaussian function:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

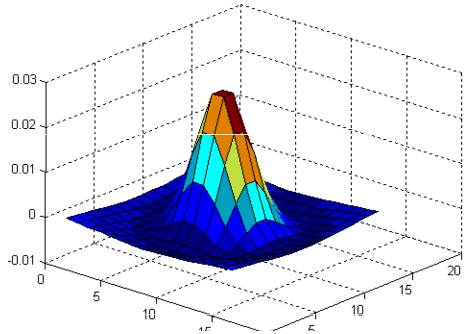
$\sigma_1 = 2$



$\sigma_2 = 4$



$\sigma_1 - \sigma_2$



Difference-of-Gaussian

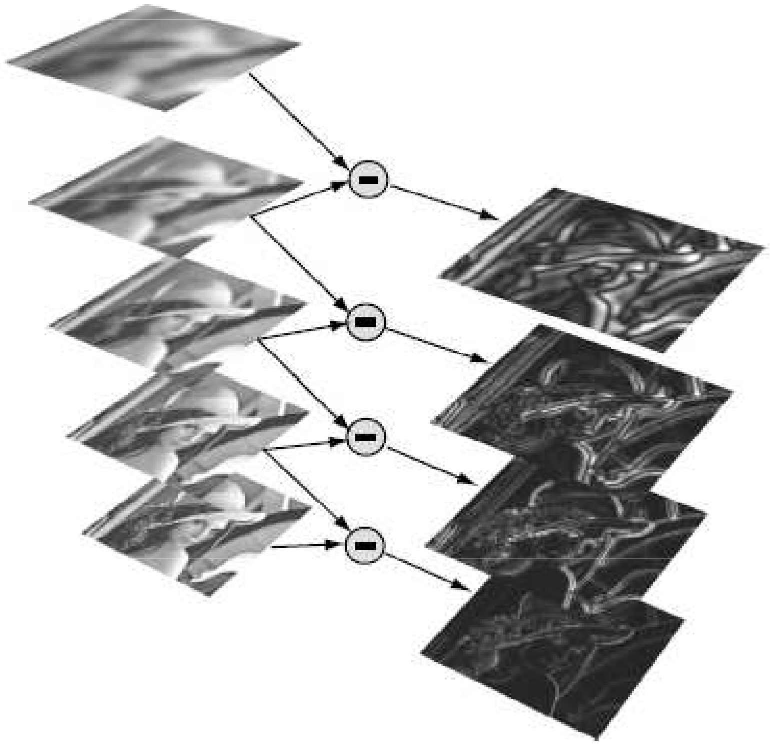
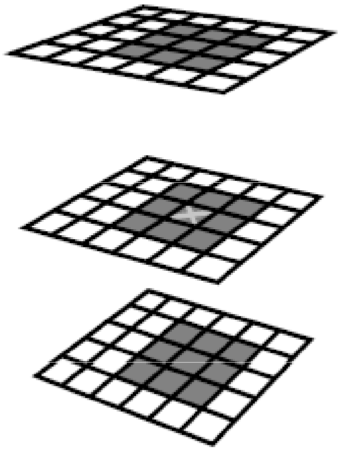


Figure 9.1: A Difference-of-Gaussian octave. The five images in the left stack are incrementally smoothed versions of the input image. The right stack shows the resulting DoG.

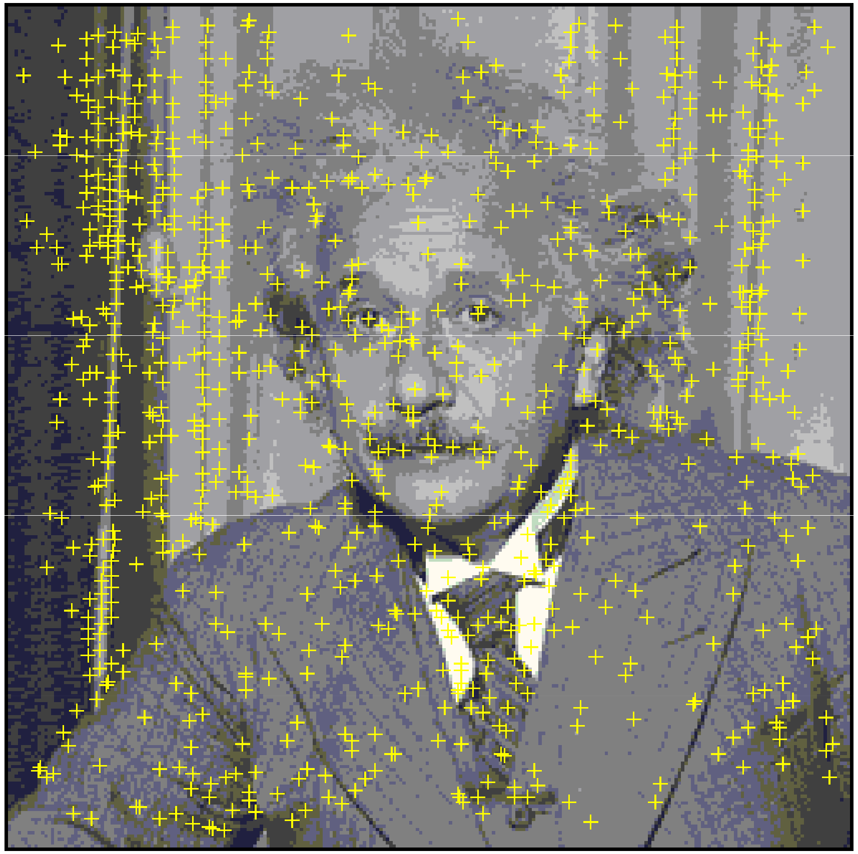
SIFT keypoints

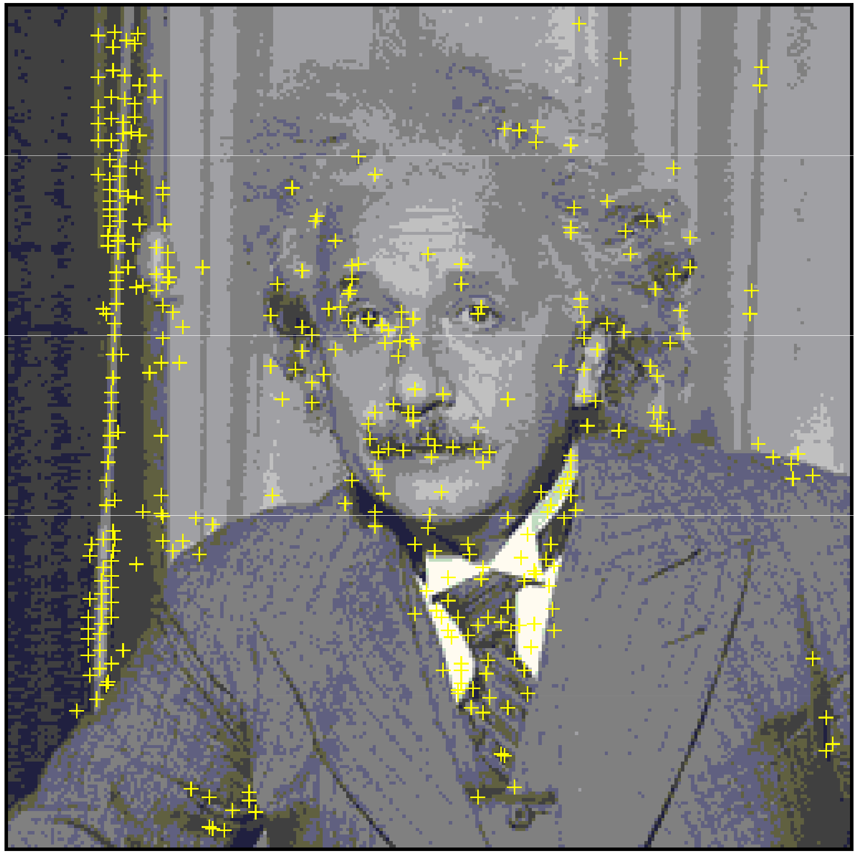


Maxima and minima of DoG applied in scale-space:

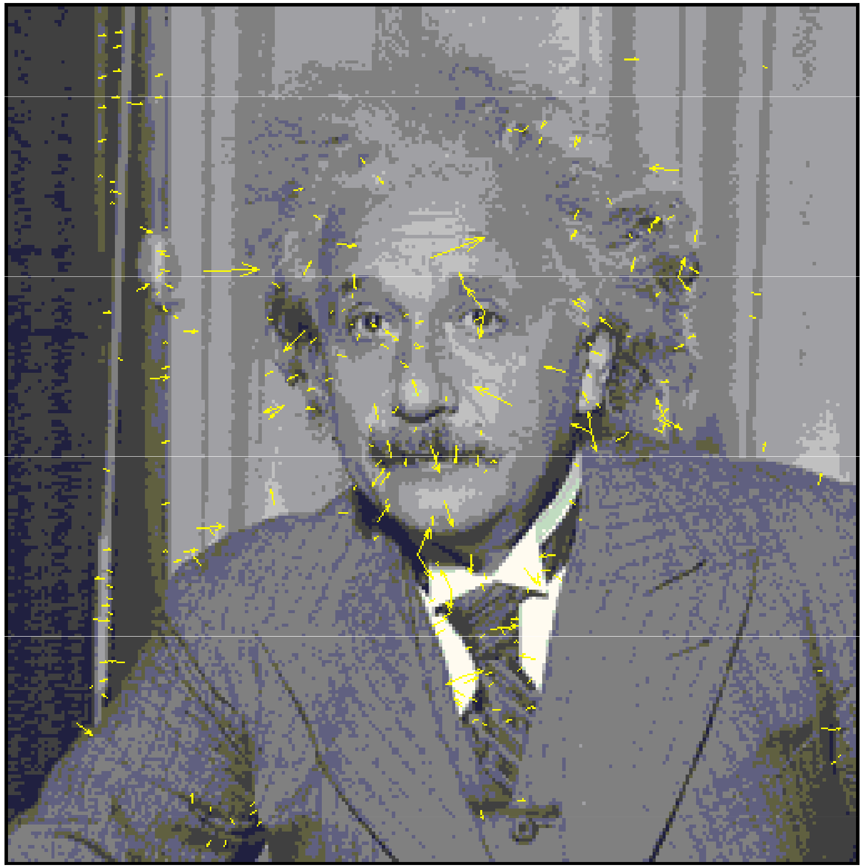
1) Extrema detection for the same scale
2) Check if it is stable for different scales

Scale-space extrema detection





Keypoints orientation and scale



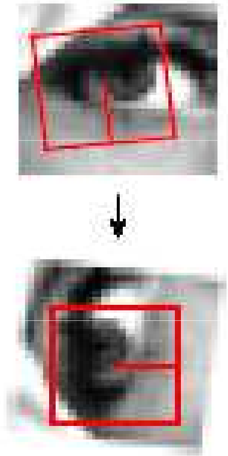
$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2}$$

$$R_{ij} = \text{atan2}\left(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij}\right)$$

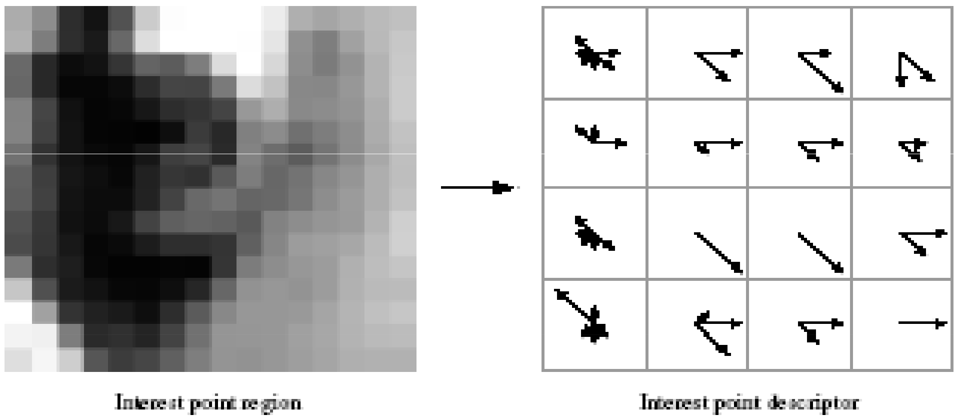
- Extract image gradients and orientations at each pixel
- Each key location is assigned a canonical orientation
- The orientation is determined by the peak in a histogram of local image gradient orientations

Example

(a) A subset of the extracted interest points, and the associated regions used to create descriptors.



(b) An interest point region covering Lena's eye before and after rotation in respect to the reference orientation of the point of interest.



(c) Computation of a descriptor by determining a 4×4 gradient orientation histogram array from a 16×16 pixels region around the interest point location.

Lecture 4: Resume

- Image segmentation
  - Is necessary for many image processing tasks (shape features, object detection)
  - The optimal methods depends on application

- Local descriptors
  - Necessary for image/object matching, sub image retrieval, near duplicates detection
  - SIFT is a very powerfull method for keypoints detection building local descriptors

Lecture 4: Bibliography

- Gonzalez R, Woods R. Digital Image Processing, published by Pearson Education, Inc, 2002.
- Lowe, David. Distinctive Image Features from Scale Invariant Keypoints. International Journal of Computer Vision, 2004.
- Ke Yan, Sukthankar Rahul. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors.
- Mikolajczyk Krystian, Schmid Cordelia. A performance evaluation of local descriptors.