Biomedical Imaging & Analysis

Lecture 7, Part 1. Fall 2014

Image Segmentation (I)

*[**Text**: Ch: 10, Gonzalez and Woods, Digital Image Processing (3rd Edition) + Chapter 5, 8, & 9, Insight into Images, Terry Yoo, PhD]*

Prahlad G Menon, PhD

Assistant Professor

Sun Yat-sen University – Carnegie Mellon University (SYSU-CMU)

Joint Institute of Engineering

Outline

- Overview of Contemporary Segmentation Algorithms

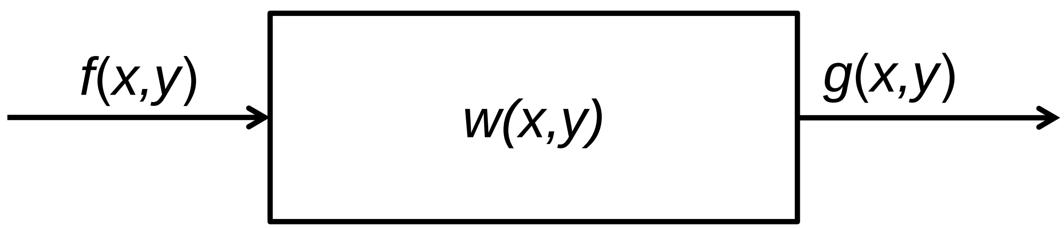
REFERENCES:

- Yoo, Terry (Editor) (2004), Insight into Images, A K Peters, ISBN 1-56881-217-5. Chapter 5, 8, & 9.
- Castleman, K. Digital Image Processing. Englewood Cliffs, NJ, Prentice Hall, 1996
- Oppenheim, A., and Willsky, A. Signals and Systems, (2nd Edition). Englewood Cliffs, NJ, Prentice Hall, 1996

Review: Spatial Filtering

- A spatial filter is often referred to as a mask, a kernel, a template, or a window.

$$\sum_{s=-a}^{a}\sum_{t=-b}^{b} w(s,t)\, f(x+s, y+t) = \sum_{s=-a}^{a}\sum_{t=-b}^{b} w(-s,-t)\, f(x+s, y+t) = w(x,y) \otimes f(x,y)$$



$$g(x,y) = w(x,y) \otimes f(x,y)$$

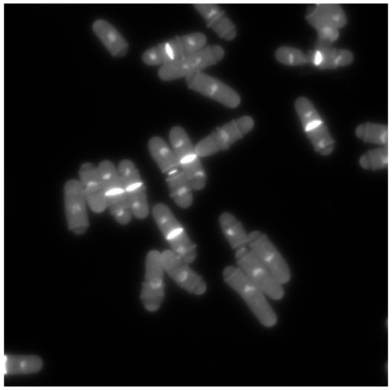
$$G(u,v) = W(u,v) \cdot F(u,v)$$

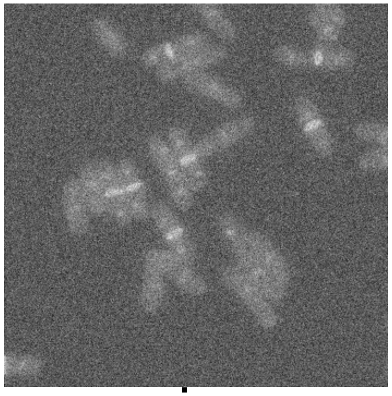
http://www.imageprocessingplace.com/

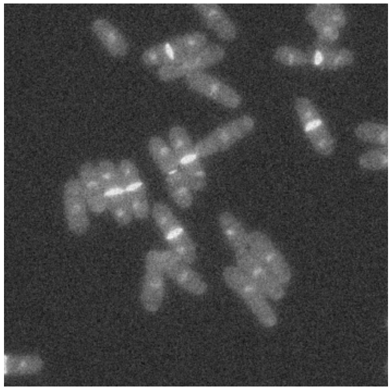
Oppenheim et al, Signals & Systems, 1997

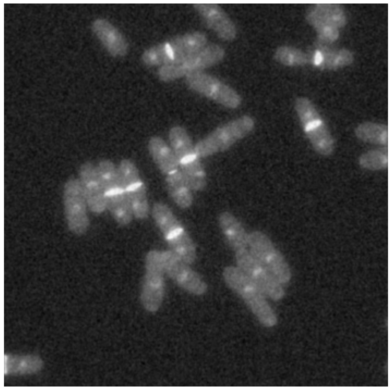
Gaussian Image Filtering

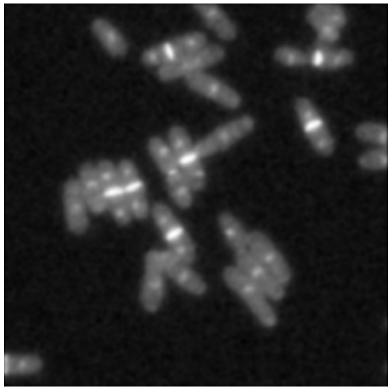
- Application I: noise suppression

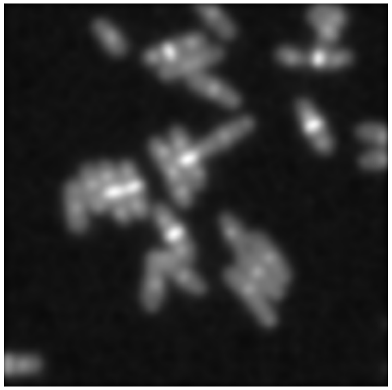

original


noise
added


σ=1


σ=2


σ=5


σ=10

Filtering by ***Anisotropic Diffusion:***

Better feature preserving noise elimination *using* Gradient Vectors

Isotropic Diffusion (Xu *et al.*, 1998)

$$\begin{cases} \dfrac{\partial u}{\partial t} = \mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) \\ \dfrac{\partial v}{\partial t} = \mu \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) \end{cases}$$

where:

(u(t), v(t)) stands for the evolving vector field;

μ is a constant;

f is the original image to be diffused;

(f_x, f_y) = (u(0), v(0)).

Anisotropic Diffusion (Yu & Bajaj ICPR'02)

$$\begin{cases} \dfrac{\partial u}{\partial t} = \mu \nabla (g(\alpha) \cdot \nabla u) - (u - f_x)(f_x^2 + f_y^2) \\ \dfrac{\partial v}{\partial t} = \mu \nabla (g(\alpha) \cdot \nabla v) - (v - f_y)(f_x^2 + f_y^2) \end{cases}$$

where:

(u(t), v(t)) stands for vector field;

μ is a constant; (f_x, f_y) = (u(0), v(0)).

f is the original image to be diffused;

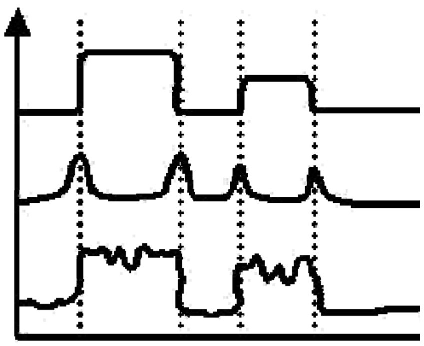
g(.) is the angle between two vectors

Z. Yu, C. Bajaj **Anisotropic Vector Diffusion in Image Smoothing** ***Proceeding of the 9th IEEE International Conference on Image Processing, vol. 1, 2002, 828-831***

Filtering by ***Anisotropic Diffusion:***

Better feature preserving noise elimination *using* Gradient Vectors

Removes noise without removing critical parts of the image.



(b) The relationship between the diffusion time parameter and the Gaussian cut-off wavelength

The relationship between the time parameter t of the diffusion process and the standard deviation σ of the equivalent Gaussian function is given by $\sigma = \sqrt{2t}$. From this relationship and the definition of the areal Gaussian weighting function (4.1), the required value of σ can be calculated for a given cut-off wavelength λ_c according to

$$\lambda_c = \pi\sigma\sqrt{\frac{2}{\ln 2}} \approx 5.3364\sigma. \quad (5.2)$$

From this relationship, the appropriate diffusion time parameter t is given by

$$t \approx 0.0176\lambda_c^2. \quad (5.3)$$

In the Euclidean case, the cut-off wavelength provides a 50 per cent transmission characteristic of a sine wave. In the case of freeform surfaces, since there is currently no clear definition of the equivalent to a sine wave on a freeform surface, this interpretation requires further exploratory research. It is sensible though to interpret the value of λ_c as a nesting index for data smoothing until a fully developed interpretation is available.

X. Jiang, P. Cooper and P. J. Scott. **Freeform surface filtering using the Diffusion equation.** ***Proc. R. Soc. A published online 9 September 2010***

General notes on discrete calculus

Operator	Vector calculus
Gradient	$G_{ij} = \langle \mathbf{x}_{u^i}, \mathbf{x}_{u^j} \rangle$ ∇
Divergence	$\nabla\cdot$
Curl	$\nabla \times \nabla$
Laplacian	$\nabla \cdot \nabla$
Beltrami	$\nabla C \cdot \nabla$

The Laplacian operator on $\mathcal{M}$ denoted by $\Delta_{\mathcal{M}}$ is then defined as (do Carmo 1993)

$$\Delta_{\mathcal{M}} f = \mathrm{div}_{\mathcal{M}}(\nabla_{\mathcal{M}} f), \tag{6.6}$$

which can be represented explicitly in terms of coordinates by

$$\Delta_{\mathcal{M}} f = \frac{1}{\sqrt{\det(G)}} \left[\frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2} \right] \left[\sqrt{\det(G)}\, G^{-1} [f_{u^1}, f_{u^2}]^{\mathrm{T}} \right]. \tag{6.7}$$

With equation (6.6), we can now model the isotropic diffusion of a quantity described by a function f defined over the surface $\mathcal{M}$ as

$$\frac{\partial f}{\partial t} - \Delta_{\mathcal{M}} f = 0. \tag{6.8}$$

- Calculus defined for images is applicable to Surfaces with a change in operator form.
- The Laplace–Beltrami operator is the generalization of the Laplacian operator to functions defined on surfaces or more generally Riemannian manifolds.
- When the manifold in question is Euclidean space, the Laplace–Beltrami operator simplifies to the standard Laplacian operator.

X. Jiang, P. Cooper and P. J. Scott. **Freeform surface filtering using the Diffusion equation.** ***Proc. R. Soc. A published online 9 September 2010.***

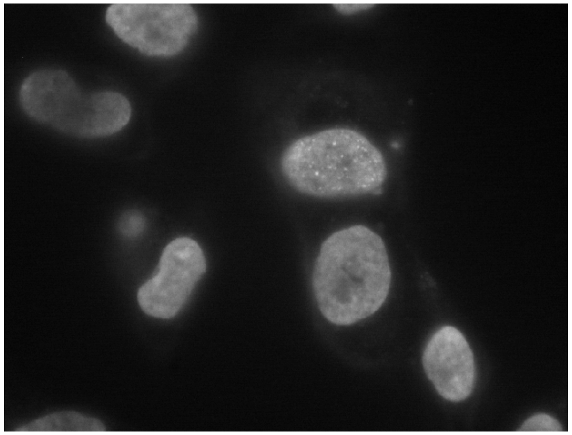
What is Image Segmentation?

- Definition

  Segmentation is the process of separating objects from background (Snyder & Qi in “Machine Vision)

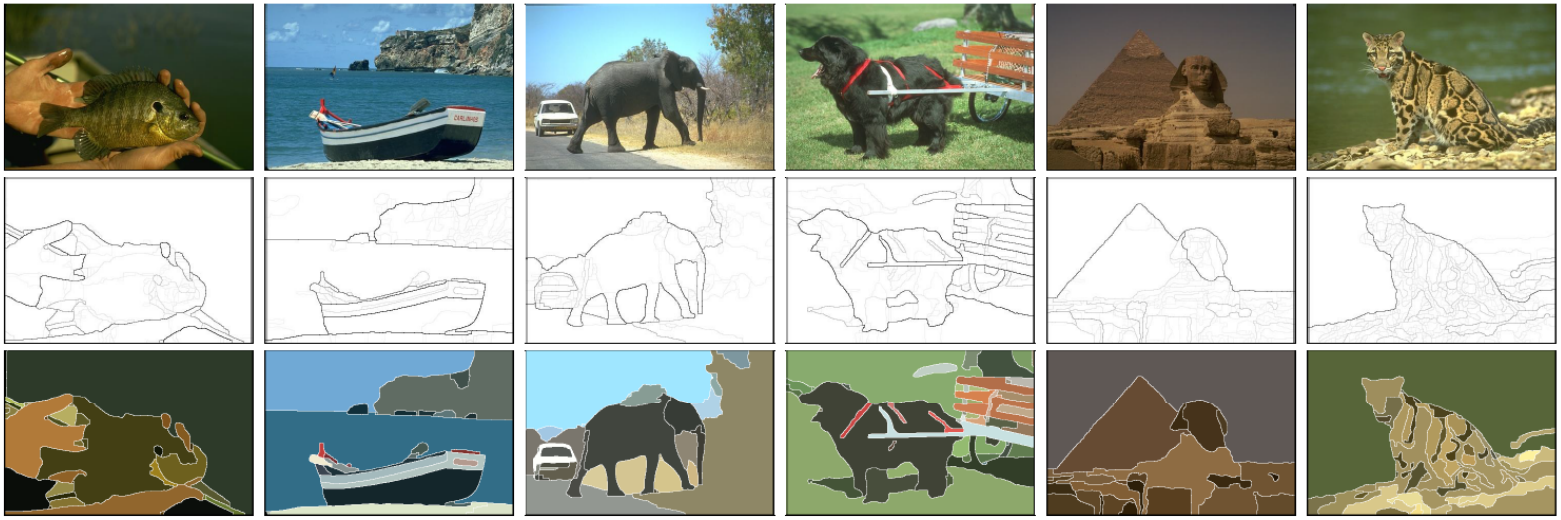
  Segmentation is the partitioning of a dataset into continuous regions (or volumes) whose member elements have common, cohesive properties (Yoo in "Insight into Images").

  Image segmentation is the task of finding groups of pixels that “go together” (Szeliski in “Computer Vision”).



- Segmentation is an essential process in bioimage analysis that is critical for many subsequent processes such as object recognition and shape analysis.

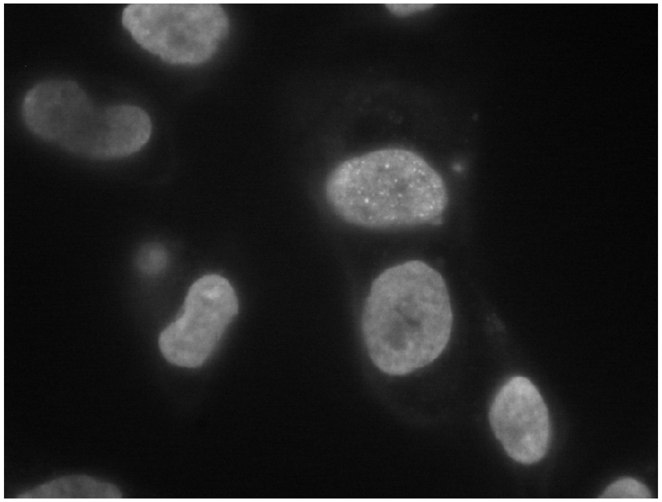
Image Segmentation of Natural Scenes

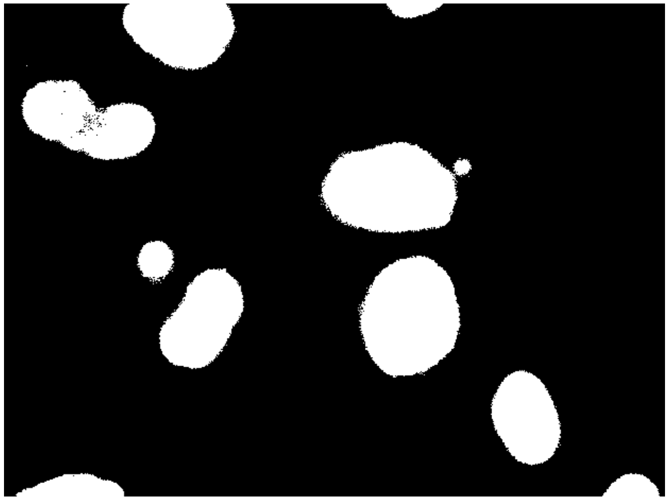


P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, “Contour Detection and Hierarchical Image Segmentation,” *IEEE Trans. on PAMI, 2010*

http://www.cs.berkeley.edu/~malik/

Segmentation of Biological Images





1. The main goal of bioimage informatics is to <u>measure</u>. Accuracy and precision are crucial.

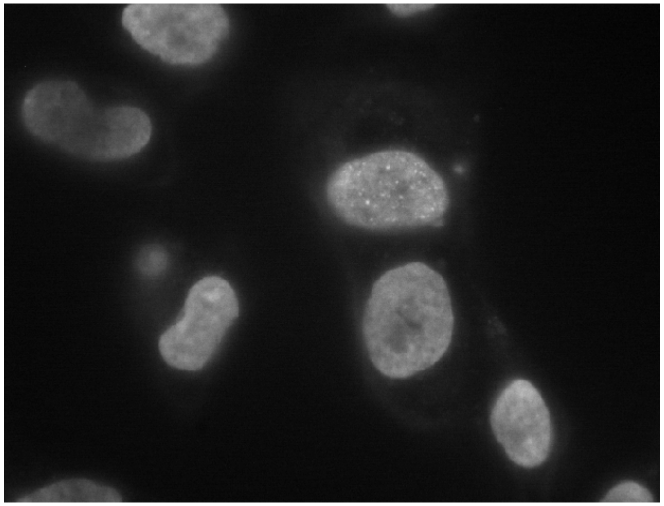
2. Image segmentation is challenging.

3. It is often possible to control image collection in biological imaging to make image segmentation more robust.

Overview of Image Segmentation Techniques

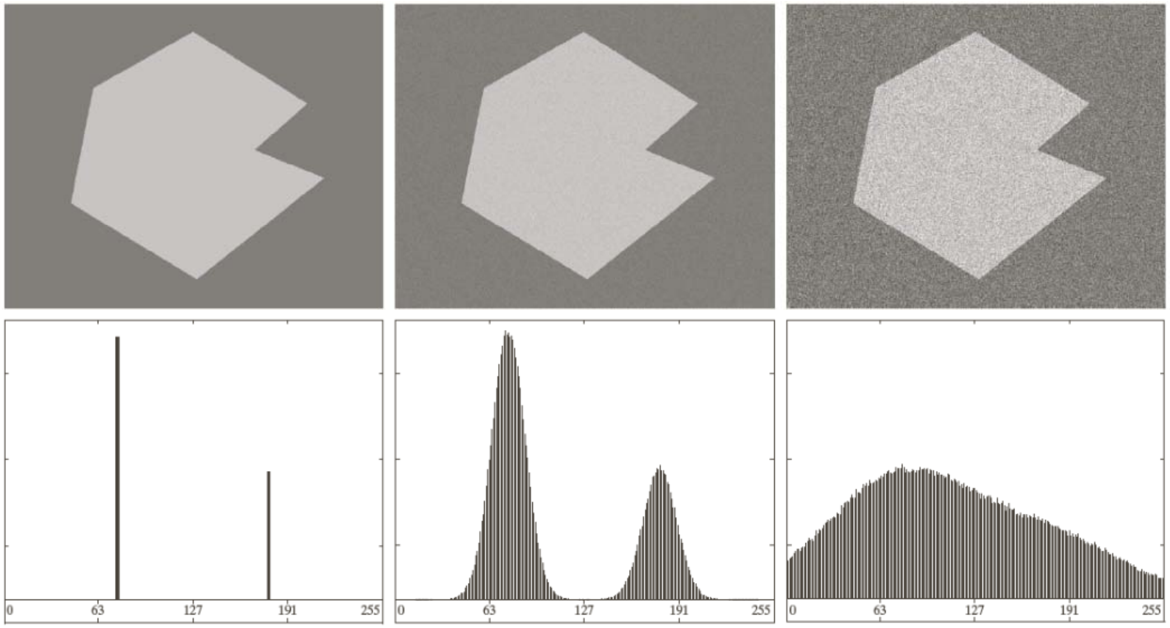
- There are many types of segmentation techniques:

  - Thresholding-based segmentation
  - Region-based segmentation
  - Boundary/surface-based segmentation
  - Motion-based segmentation
  - Color-based segmentation
  - Others…



- A commonly used strategy is to combine multiple techniques for image segmentation.

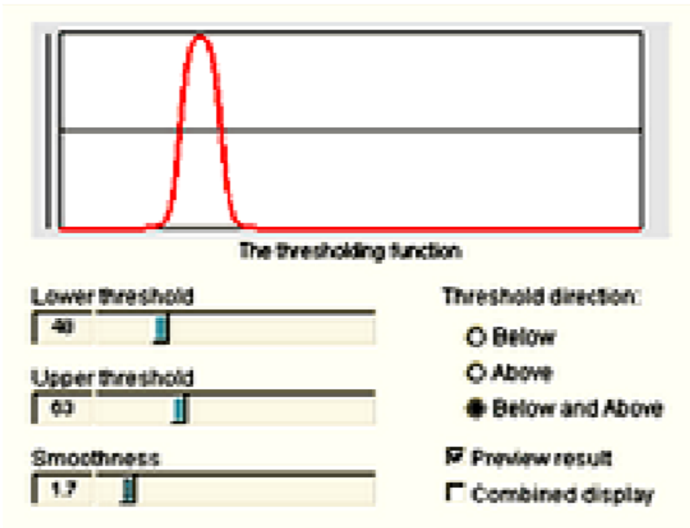
Basic Ideas of Thresholding-Based Segmentation (I)



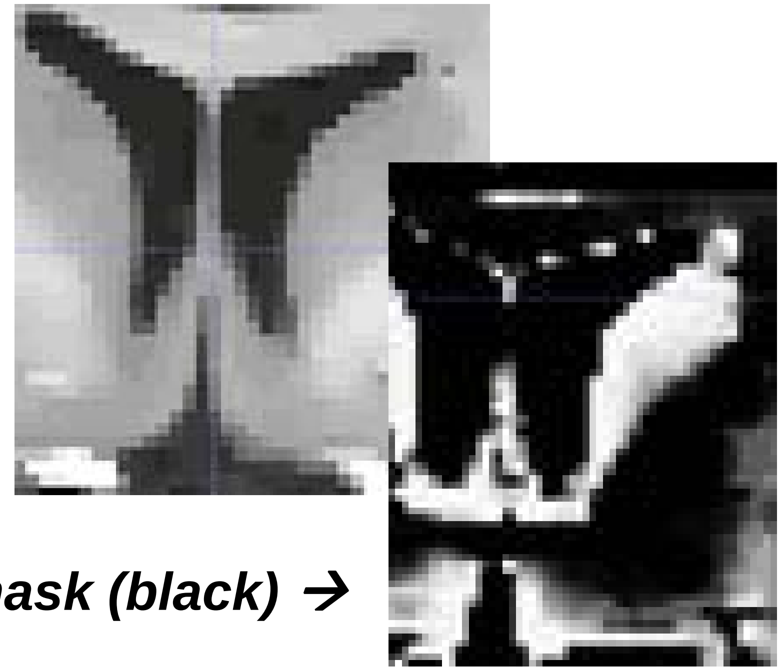
a b c
d e f

FIGURE 10.36 (a) Noiseless 8-bit image. (b) Image with additive Gaussian noise of mean 0 and standard deviation of 10 intensity levels. (c) Image with additive Gaussian noise of mean 0 and standard deviation of 50 intensity levels. (d)–(f) Corresponding histograms.

Pre-processing: Thresholding

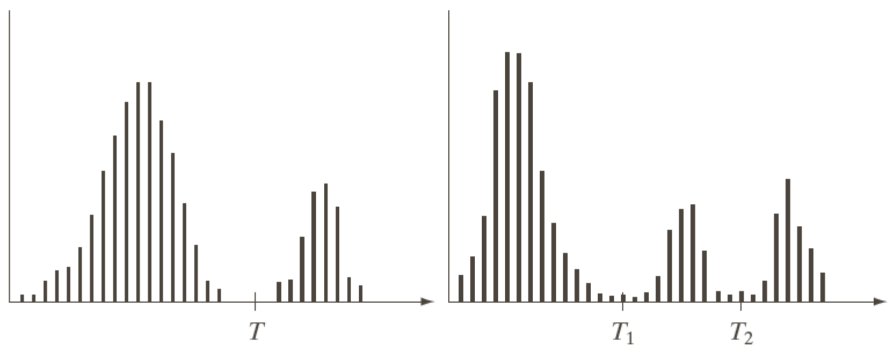


Bi-thresholded ***mask (black)*** **→**



- Provides boundaries within which more specific region-growing segmentation techniques can operate.
- Reduces the size of the problem in terms of image-space to be processed.

Basic Ideas of Thresholding-Based Segmentation (II)



a b

FIGURE 10.35 Intensity histograms that can be partitioned (a) by a single threshold, and (b) by dual thresholds.

$$g(x,y)=\begin{cases}1 & \text{if} \quad I(x,y)>T \\ 0 & \text{if} \quad I(x,y)\le T\end{cases}$$

$$g(x,y)=\begin{cases}a & \text{if} \quad I(x,y)>T_2 \\ b & \text{if} \quad T_1<I(x,y)\le T_2 \\ c & \text{if} \quad I(x,y)\le T_1\end{cases}$$

Thresholding-Based Segmentation (I)

- Revisit the definition

  "Segmentation is the partitioning of a dataset into continous regions (or volumes) whose member elements have common, cohesive properties".

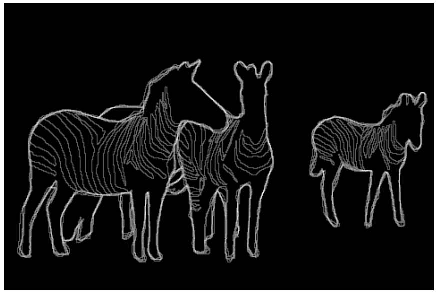
- Intensity is the most frequently used property.
- Multiple continuous regions of cohesive intensities will result in multiple peaks in intensity histogram.

Thresholding-Based Segmentation (II)

- Thresholding-based segmentation is usually among the first options to be considered.
  - Simple; can be quite reliable
  - Easy to implement.

- There are many refinements to the basic idea that work remarkably well.

- What are potential limitations?

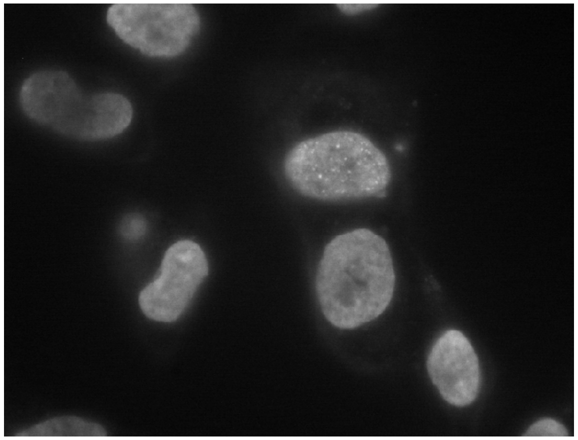


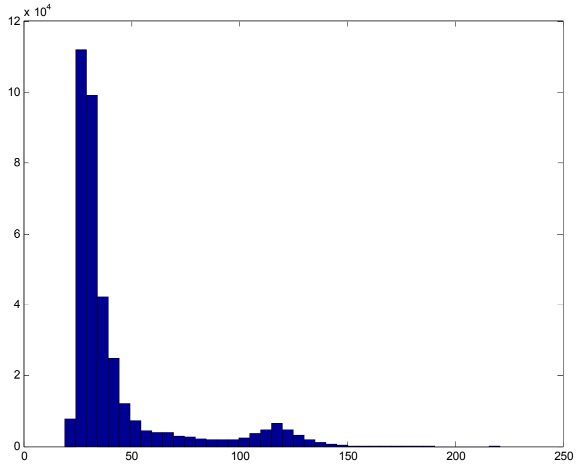
Berkeley Segmentation Dataset and Benchmark

How to Set Thresholds (I)

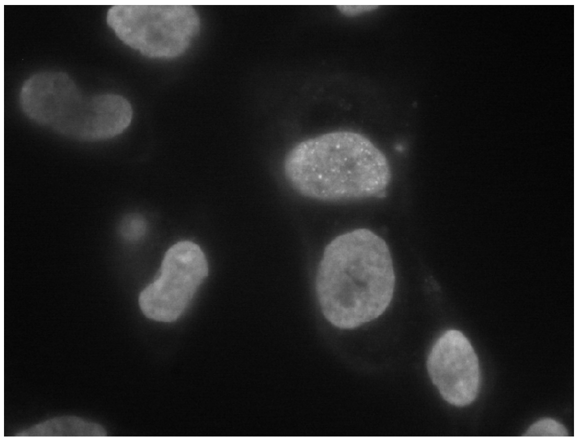
- There are several ways to set the thresholds.
  - Using local minima in the intensity histogram ← sensitive to noise
  - Use intensity histogram fitting with a mixture of Gaussians ← more robust

- Example:

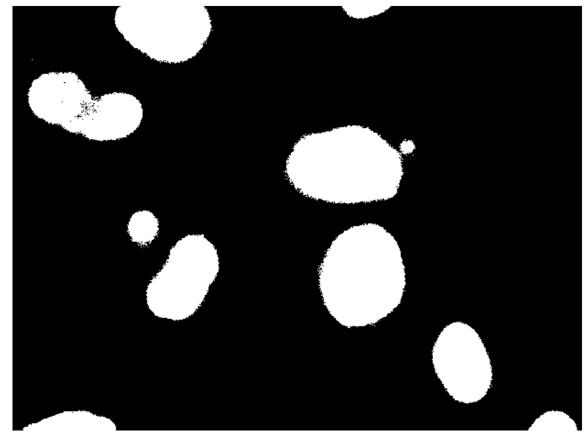




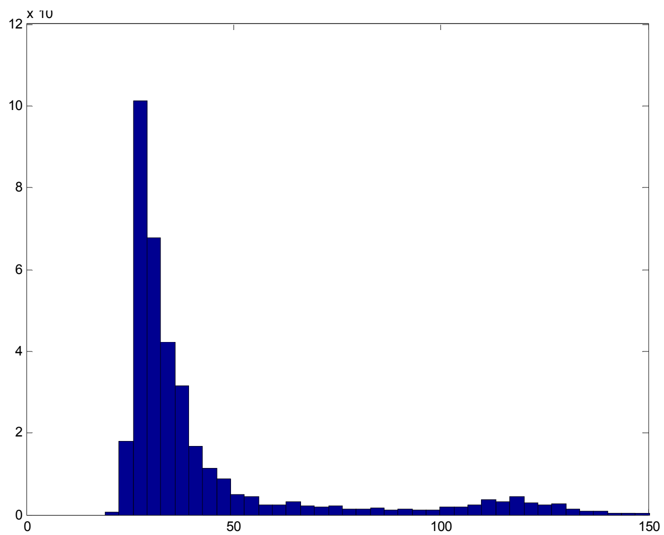
Example Results



Threshold = 70



Threshold = 60



How to Set Thresholds (II)

- One way to fit multiple Gaussians

  Reference: C. Fraley and A. E. Raftery. *Model-based clustering, discriminant analysis and density estimation*. Journal of the American Statistical Association, 97:611–631, 2002.

  Implementation in R : http://www.stat.washington.edu/mclust/

Outline of the Gaussian Mixture Fitting Algorithm

- A general formulation of the question

Given a collection of measurements: $Y = [y_1, y_2, \dots, y_n]$

and assuming that they are generated by a mixture of model with G components,

the likelihood is given as

$$L_{MIX}(\theta_1, \dots, \theta_G; \tau_1, \dots, \tau_G | Y) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(y_i | \theta_k)$$

- If we assume the models are Gaussian

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

$$L_{MIX}(\theta_1, \dots, \theta_G; \tau_1, \dots, \tau_G | Y) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k \frac{\exp\left\{-\frac{1}{2}(y_i - \mu_k)^T \Sigma_k^{-1} (y_i - \mu_k)\right\}}{\sqrt{\det(2\pi \Sigma_k)}}$$

[Discrete form!]

How to Set Thresholds (II)

- Determine the number of Gaussians

  - Bayesian information criterion (BIC)

$$\mathrm{BIC}_k \equiv -2\,log\; p\left(y \middle| \hat{\theta}_k, M_k\right) + N_k \cdot \log(n)$$

$M \rightarrow$ Model
$log\; p\left(y \middle| \hat{\theta}_k, M_k\right) \rightarrow$ maximized likelihood
$N_k \rightarrow$ Number of parameters in model M
$n \rightarrow$ Number of measurements

BIC is a likelihood criterion penalized by the model complexity: the number of parameters in the model. In detail, let $\mathcal{X} = \{x_i : i = 1, \cdots, N\}$ be the data set we are modeling; let $\mathcal{M} = \{M_i : i = 1, \cdots, K\}$ be the candidates of parametric models. Assuming we maximize the likelihood function separately for each model M, obtaining, say $L(\mathcal{X}, M)$. Denote $\#(M)$ as the number of parameters in the model M. The BIC criterion is defined as:

$$BIC(M) = \log L(\mathcal{X}, M) - \frac{1}{2}\#(M) \times \log(n) \quad (1)$$

- Reference on Blackboard…

  CLUSTERING VIA THE BAYESIAN INFORMATION CRITERION WITH APPLICATIONS IN SPEECH RECOGNITION. Scott Shaobing Chen & P.S. Gopalakrishnan

How to Set Thresholds (III)

- Discriminant analysis (supervised classification)

$$P\left(x \in \text{class } j\right) = \frac{w_j p_j\left(x\right)}{\sum_{k=1}^{M} w_k p_k\left(x\right)}$$

- Determine the threshold between two neighboring Gaussian

$$\frac{w_1}{\sigma_1\sqrt{2\pi}} exp\left(-\frac{\left(x-\mu_1\right)^2}{2\sigma_1^2}\right) = \frac{w_2}{\sigma_2\sqrt{2\pi}} exp\left(-\frac{\left(x-\mu_2\right)^2}{2\sigma_2^2}\right)$$

$$\left(\frac{1}{2\sigma_2^2}-\frac{1}{2\sigma_1^2}\right)x^2+\left(\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}\right)x+\left(\frac{\mu_2^2}{2\sigma_2^2}-\frac{\mu_1^1}{2\sigma_1^2}-log\frac{w_2\sigma_1}{w_1\sigma_2}\right)=0$$

Beyond Thresholding…

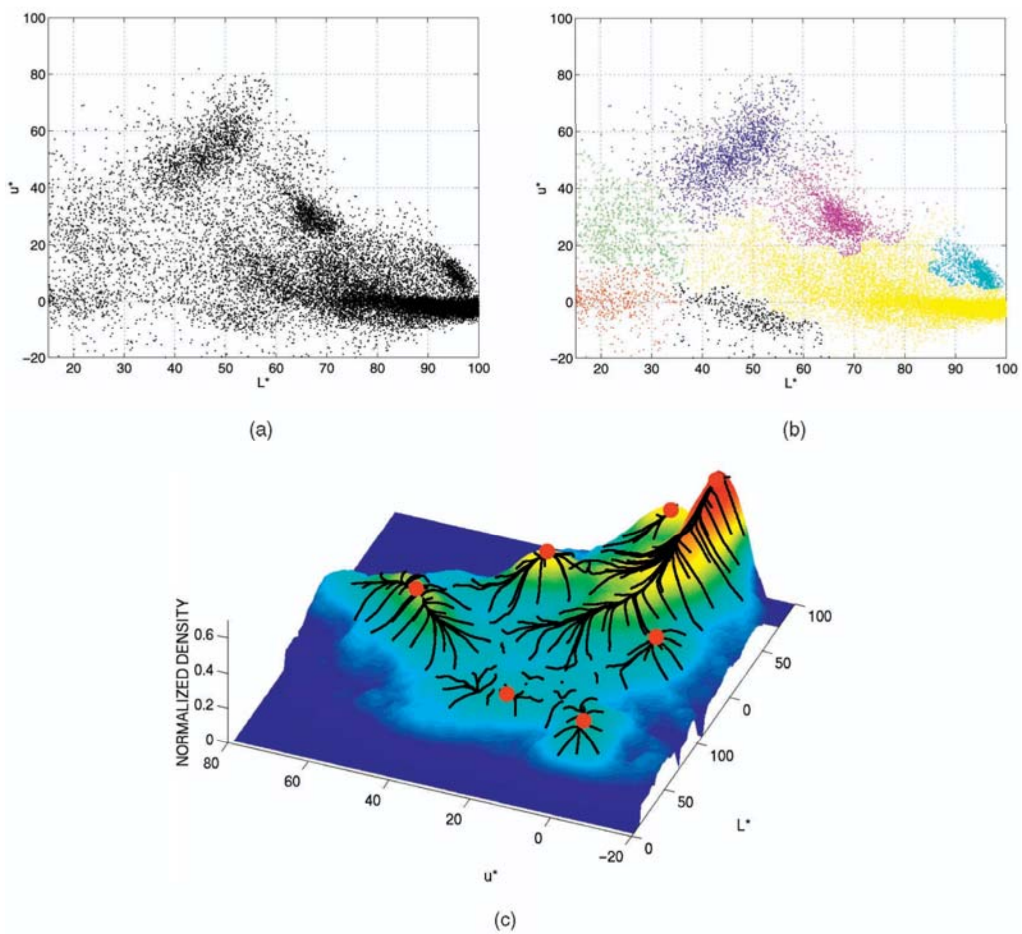
- A simple threshold becomes a linear discriminant on the histogram, dividing the distribution of image intensity values into two classes: “bone” and “not-bone”.
- **Thresholding is particularly useful for x-ray CT data where intensity values have an intuitive mapping to physical density.**
- More complex, multivalued data, such as MRI or color cryosections, **require more sophisticated statistical techniques** as simple thresholding may fail to capture the global boundary and shape properties of the object, leading to **noisy boundaries and holes** inside the object.

Statistical Pattern Recognition, Fuzzy Connectedness and Classification of Sub-Regions in a Segmentation

Statistical techniques to improve thresholding results (working in image space)

- **Region merging** is a basic means of accounting for spatial proximity and connectedness as well as alikeness of intensity values.
- A **connected-threshold approach** helps to recursively aggregate all voxels that are connected to the marked point by an unbroken lattice of voxels that appear within a user-defined interval of intensity values.
- **Fuzzy connectedness** allows capturing the spatio- topological concept of hanging-togetherness of image elements based on the notion of a “degree of connectedness”, even in the presence of a gradation of their intensities stemming from natural material heterogeneities, blurring and other artifacts.
- **Markov Random Field (MRF)** models are probabilistic models that uses the correlation between pixels in a neighborhood to decide the object region. Probabilistic segmentation techniques involve a mathematical technique of **Expectation Maximization (EM)** to confirm connectivity .

Mean-Shift Segmentation – Clustering Intensities in Feature Space



Comaniciu, D. & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.

Introduction to Mean-Shift Clustering

- First proposed by Fukunaga & Hosteler

  Fukunaga, K. & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1), 32–40.

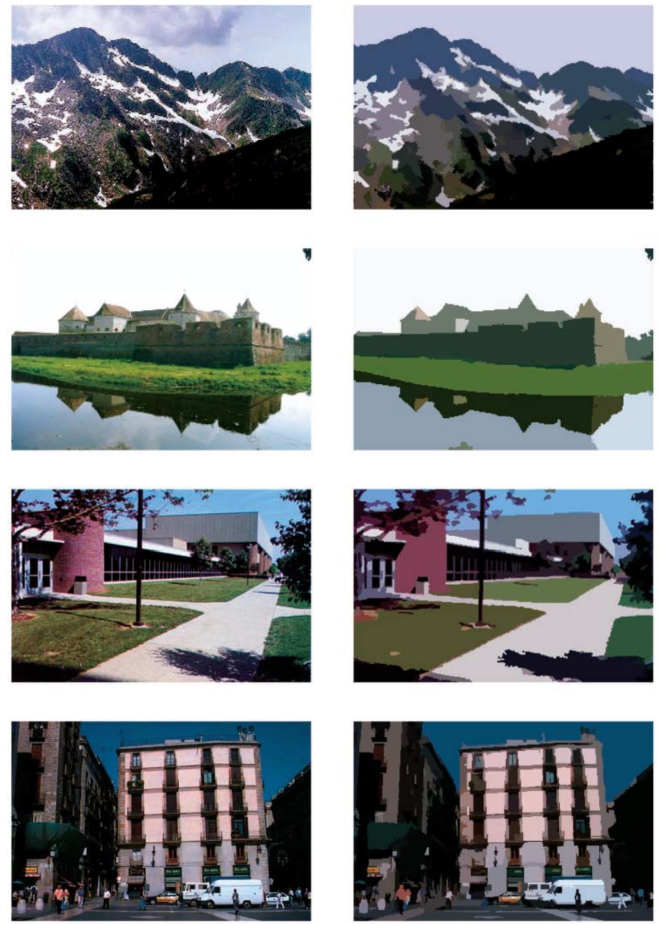
- Rediscovered by Cheng

  Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), 790–799.

- Further developed by Comaniciu & Meer

  Comaniciu, D. & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.

Mean-Shift Segmentation Results



Comaniciu, D. & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.

Morphological Processing After Thresholding

- Thresholded pixels need to be connected into regions, often through recursive region growth.

- Morphological image processing is often required to remove noise-related irregularities.

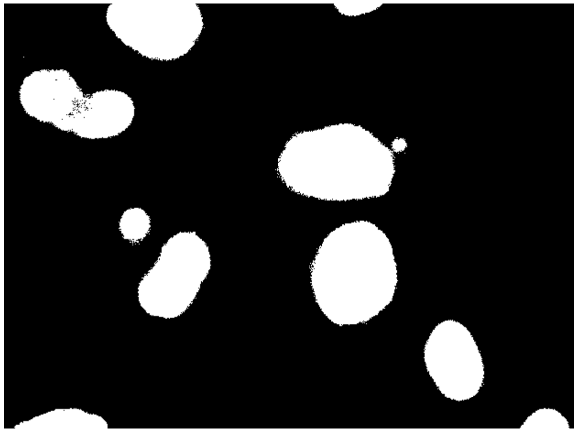


Image Morphology
Altering Operations

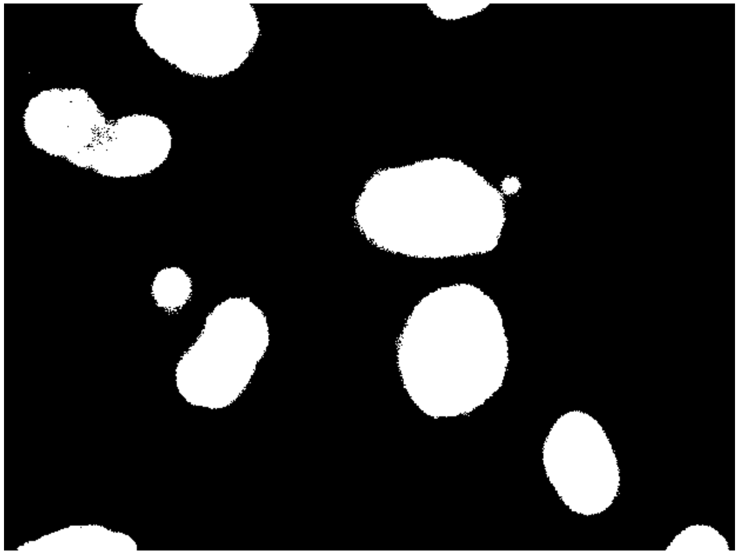
Morphological Image Processing

- Some typical applications:

  - To remove small isolated regions generated by noise

  - Contour smoothing

  - Region filling and connection



Mathematical Background

- Denote the 2D integer space as Z^2, which is the space of image pixel coordinates.

- Basic relations: Let A be a set in Z^2

$$a = (a_1, a_2) \quad a \in A \quad \text{or} \quad a \notin A$$

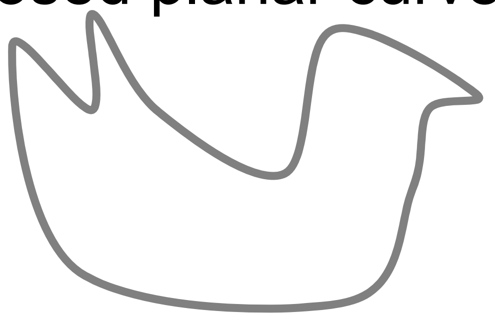
$$A \subseteq B$$

- Basic operations

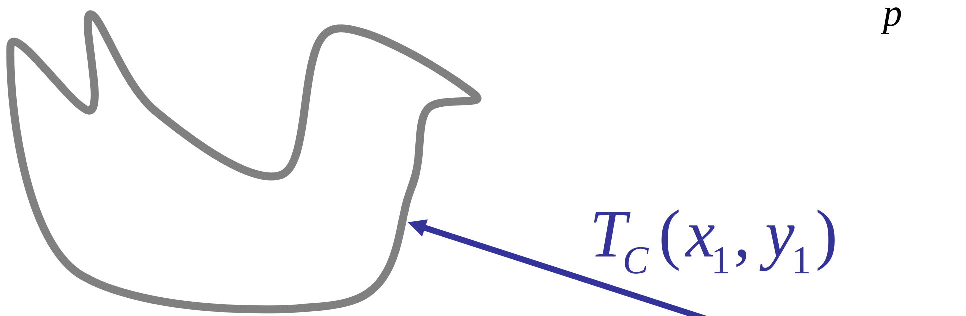
$$C = A \bigcup B$$
$$D = A \bigcap B$$
$$A - B = \{ w \mid w \in A, w \notin B \}$$

Distance map

Given a closed planar curve $C(p): \mathbf{S}^1 \rightarrow \mathbf{R}^2$



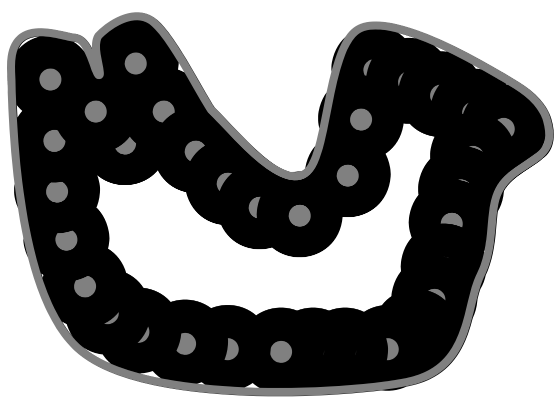
Define the distance map $T_C(x,y) = \arg\min_p \left\| C(p) - \{x,y\} \right\|$



Distance Map Properties

By Huygens principle a **level set** of $T(x, y)$, is given by the envelope of all disks of radius c centered on the curve C.

The new shape is also known as `dilation' with a circular `structuring element' of the shape.



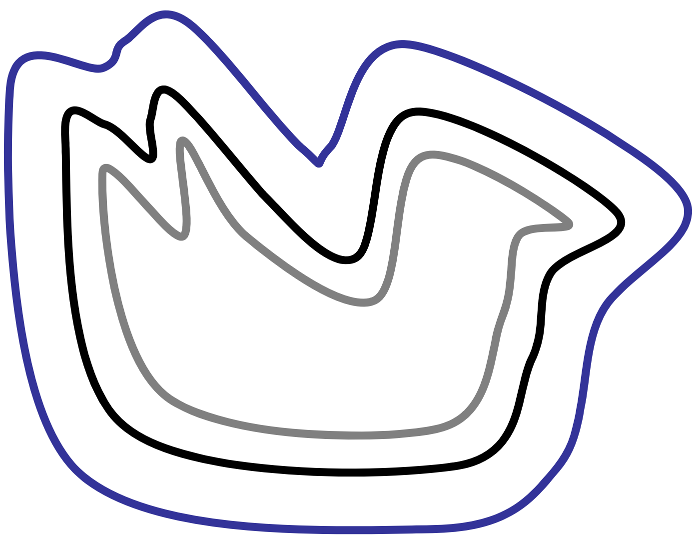
Distance Map Properties

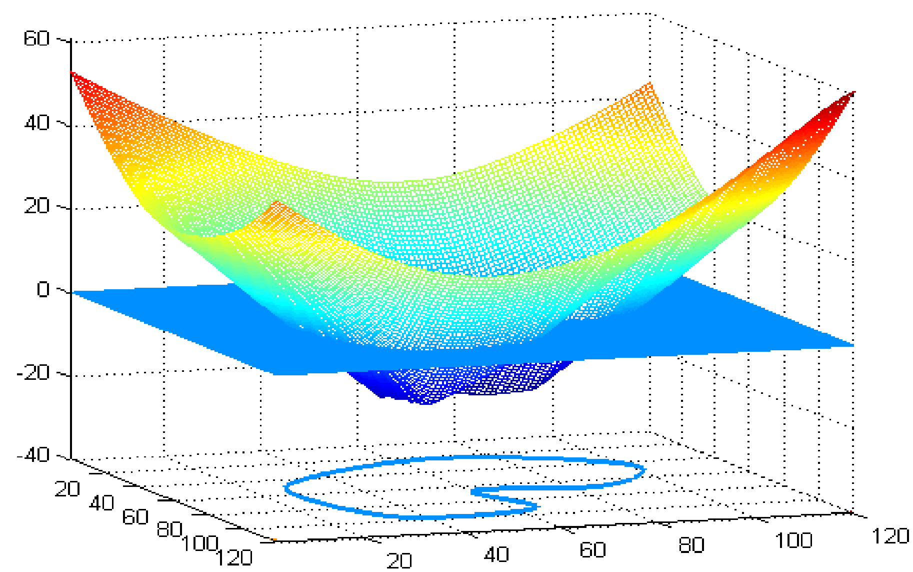
Almost everywhere $|\nabla T| = 1$

The level sets of $T(x, y)$ given by

$$T^{-1}(\mathrm{c}) = \{(x, y) : T(x, y) = \mathrm{c}\}$$

are the offsets of C



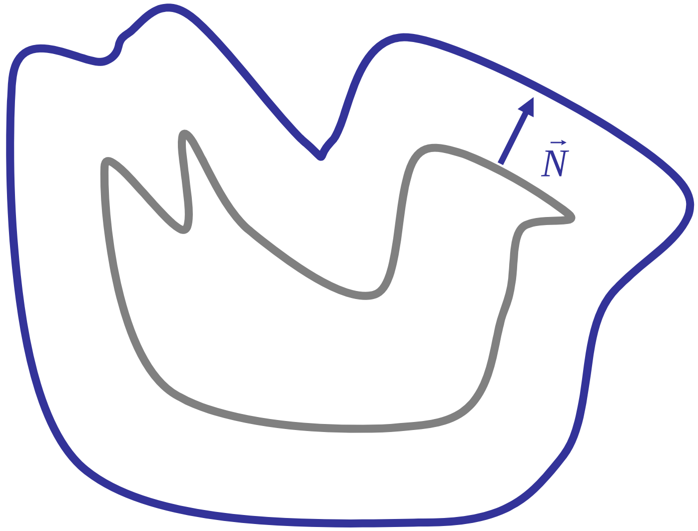


Distance Map Properties

The distance map $T(x,y)$ represents the set $C(p,t)$, generated by the curve evolution

$$C_t = \vec{N}$$

with the right `entropy condition'



How to Compute a Distance Map?

Can it be computed in *O(N)* ?

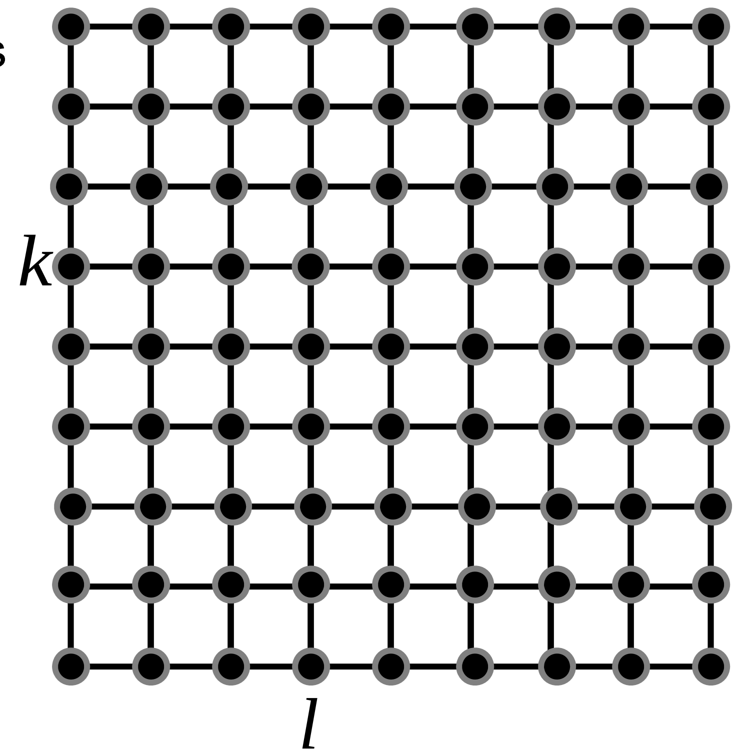
Solution: **Danielson algorithm.**

4 scans algorithm with alternating directions (up/down left/right)

Ask your 4 neighbors their coordinate offset to the closest detected source point.

Compute your offset to that point and decide if you change your choice of closest source point

Initial offset is $\{\infty, \infty\}$



Solution to Eikonal Equation

We can also use a numeric scheme to solve the 'eikonal' equation $\boxed{|\nabla T| = 1}$ $\quad |\nabla T(x,y)|^2 \equiv \left(\frac{d}{dx}T\right)^2 + \left(\frac{d}{dy}T\right)^2$

Initialize all source points $T_{kl} \leftarrow 0$

and all non-source points $T_{ij} \leftarrow \infty$

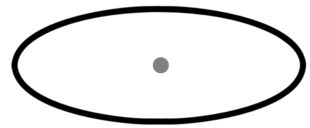
Solve the quadratic equation for T_{ij} given by the upwind monotone approximation of the eikonal equation

$$\boxed{\left(\max\left(-D_-^x T_{ij}, D_+^x T_{ij}, 0\right)\right)^2 + \left(\max\left(-D_-^y T_{ij}, D_+^y T_{ij}, 0\right)\right)^2 = 1}$$

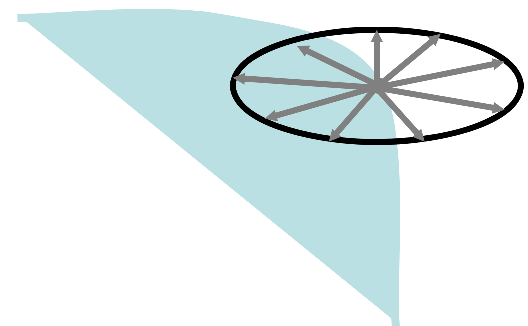
Again, use the 4 scans with alternating directions…

Continuous Morphology

Structuring element

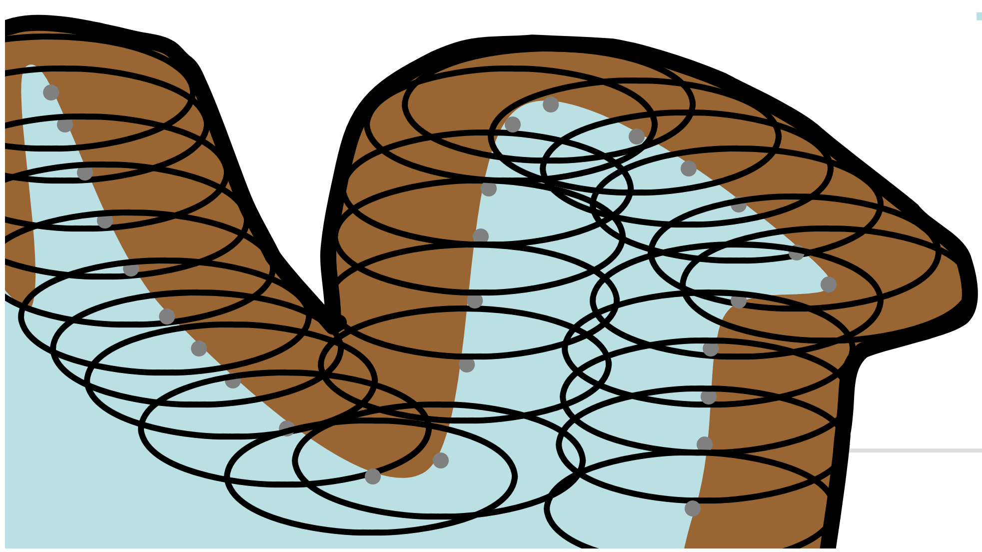


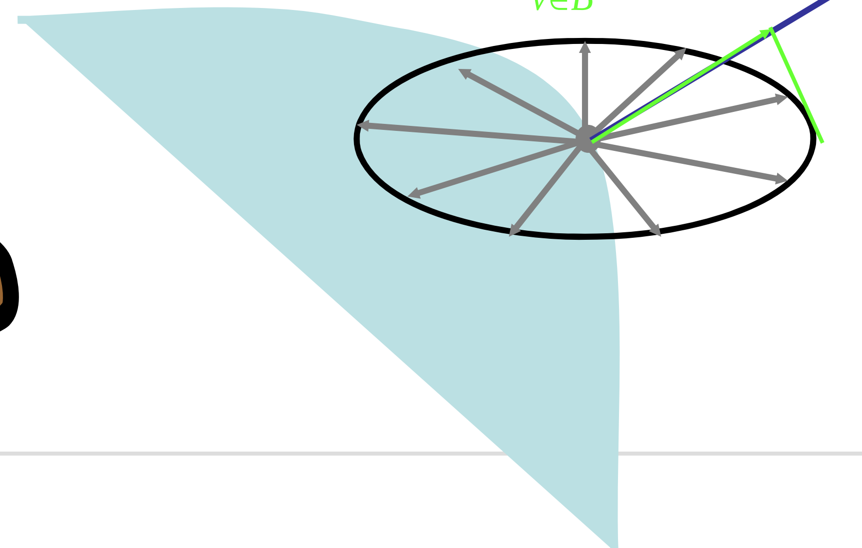
$$C_t = \sup_{v \in B} \left\langle \vec{N}, v \right\rangle \vec{N} \quad \Leftrightarrow \quad \phi_t = \sup_{v \in B} \left\langle \nabla \phi, v \right\rangle$$



where "sup" denotes the supremum.

The **supremum** (sup) of a subset *S* of a totally or partially ordered set *T* is the least element of *T* that is *greater than or equal to* all elements of *S*.





Mathematical Background

- Morphological operations are based on structural elements.

- There are many forms of structural elements.

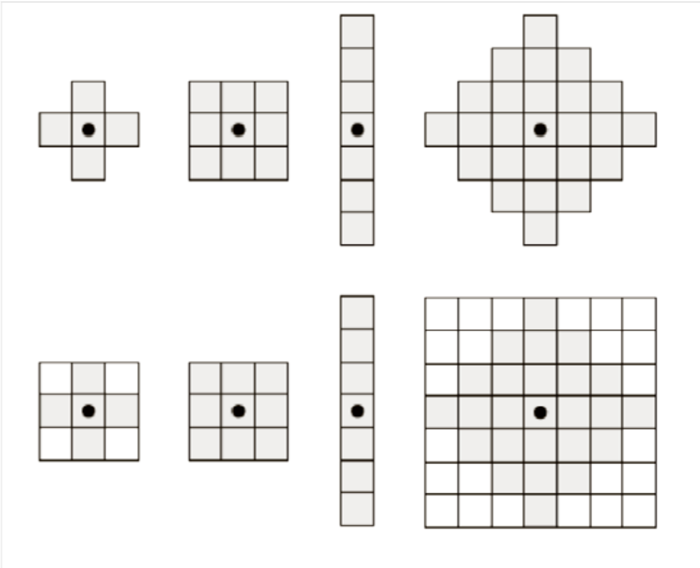
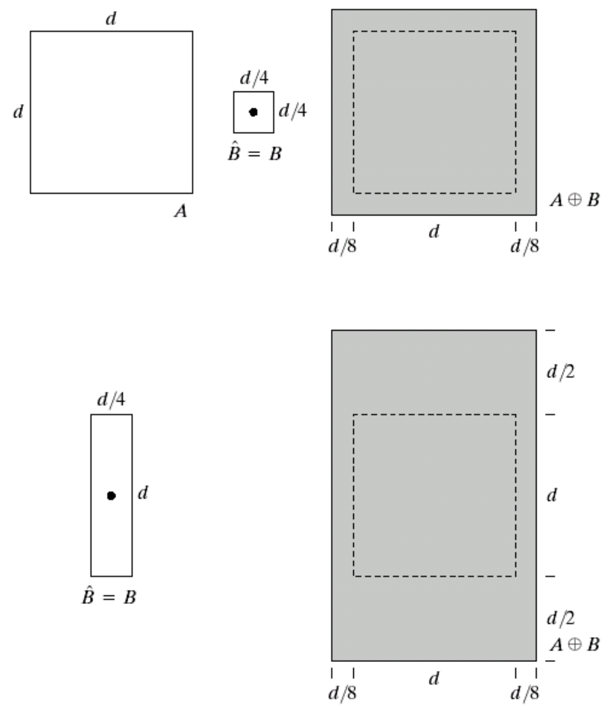


FIGURE 9.2 First row: Examples of structuring elements. Second row: Structuring elements converted to rectangular arrays. The dots denote the centers of the SEs.

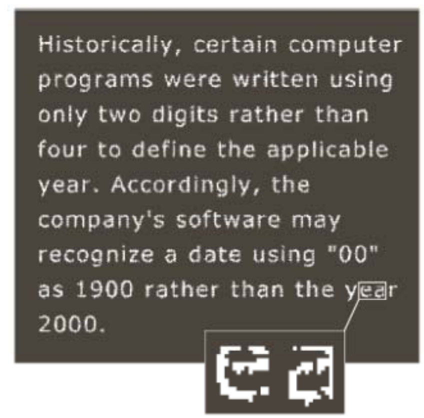
Image Morphology: Dilation



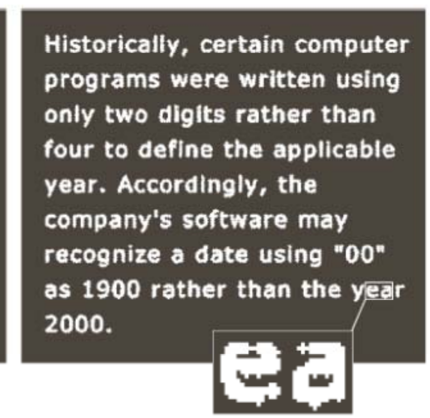
a b c
d e

FIGURE 9.4 (a) Set A. (b) Square structuring element (dot is the center). (c) Dilation of A by B, shown shaded. (d) Elongated structuring element. (e) Dilation of A using this element.

$$A \oplus B = \left\{ z \,\middle|\, \left[\left(\hat{B} \right)_z \cap A \right] \neq \varnothing \right\}$$



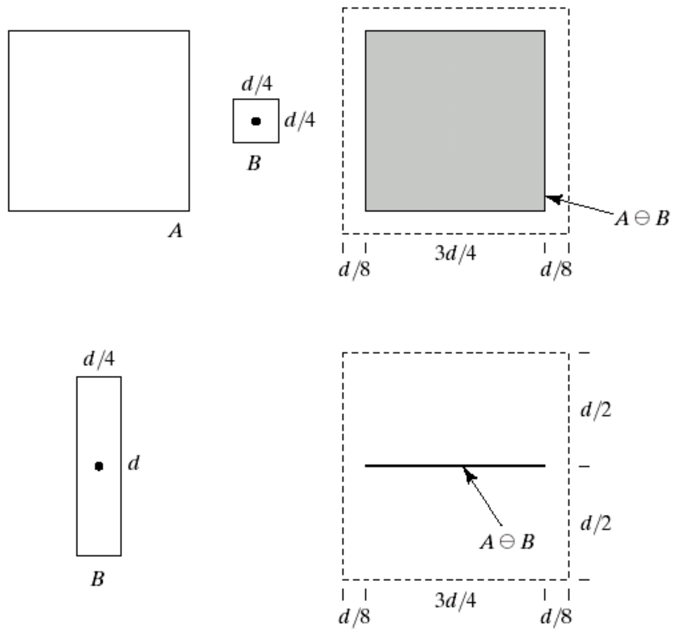
0	1	0
1	1	1
0	1	0



a c
b

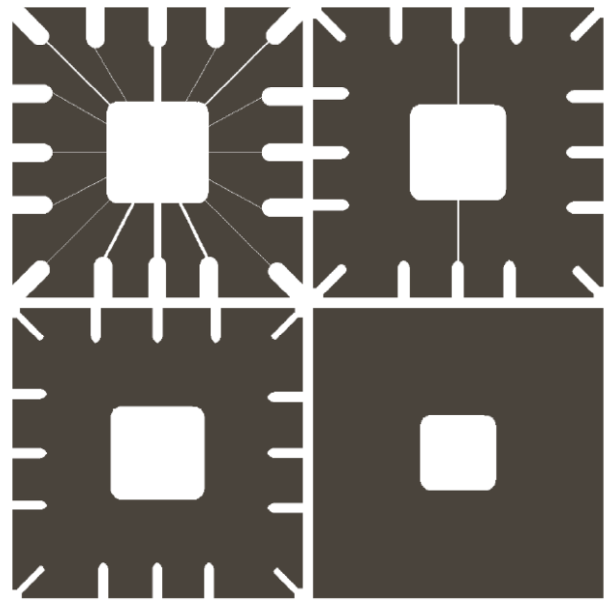
FIGURE 9.7 (a) Sample text of poor resolution with broken characters (see magnified view). (b) Structuring element. (c) Dilation of (a) by (b). Broken segments were joined.

Image Morphology: Erosion



a b c
d e

FIGURE 9.6 (a) Set A. (b) Square structuring element. (c) Erosion of A by B, shown shaded. (d) Elongated structuring element. (e) Erosion of A using this element.

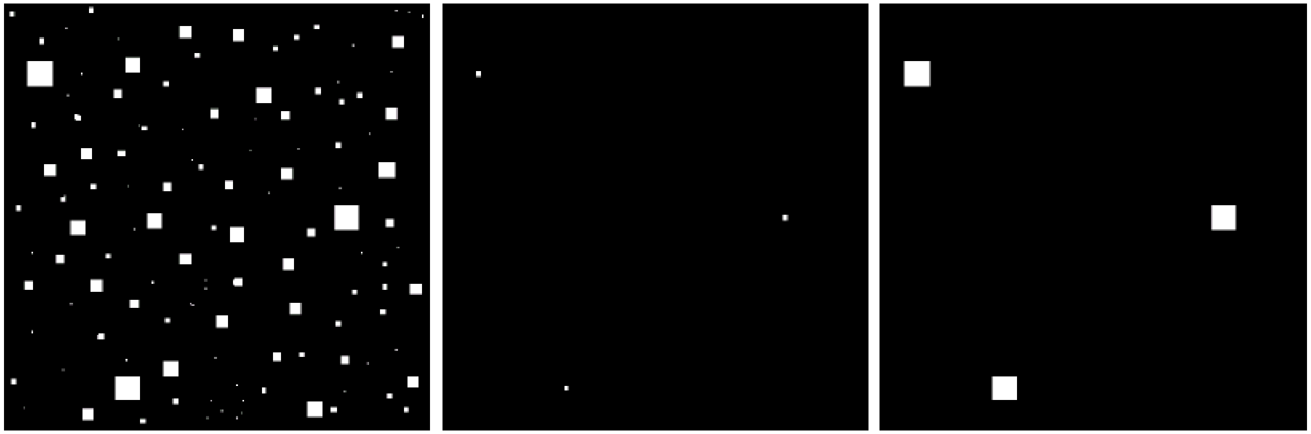


a b
c d

FIGURE 9.5 Using erosion to remove image components. (a) A 486×486 binary image of a wire-bond mask. (b)–(d) Image eroded using square structuring elements of sizes 11×11, 15×15, and 45×45, respectively. The elements of the SEs were all 1s.

$$A \ominus B = \{ z \mid (B)_z \subseteq A \}$$

Erosion+Dilation



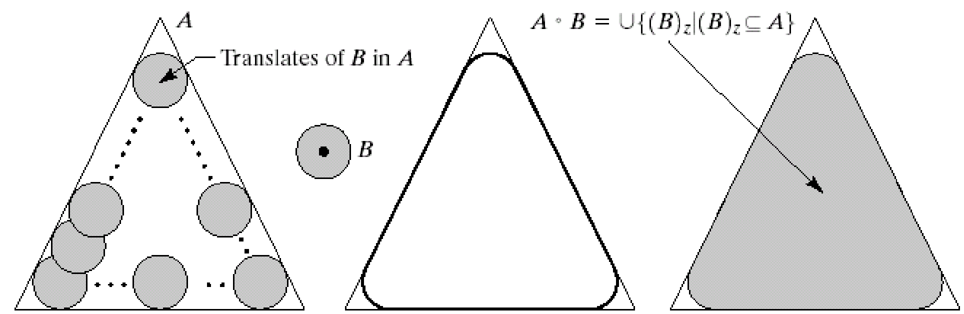
a b c

FIGURE 9.7 (a) Image of squares of size 1, 3, 5, 7, 9, and 15 pixels on the side. (b) Erosion of (a) with a square structuring element of 1's, 13 pixels on the side. (c) Dilation of (b) with the same structuring element.

Image Morphology: Opening

- Erosion followed by dilation

- Functions
  - Smooth contours
  - Break thin connections
  - Remove thin protrusions



a b c d

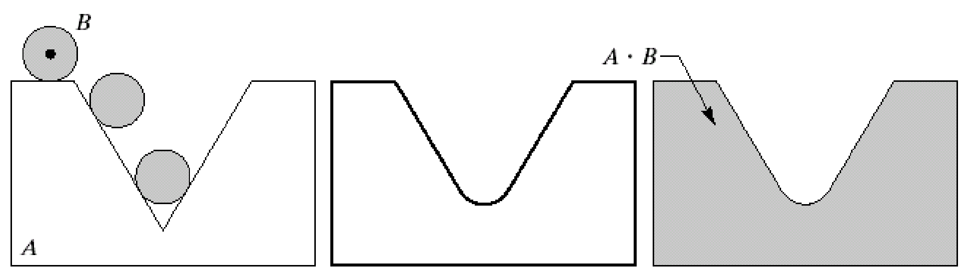
FIGURE 9.8 (a) Structuring element B "rolling" along the inner boundary of A (the dot indicates the origin of B). (c) The heavy line is the outer boundary of the opening. (d) Complete opening (shaded).

$$\text{Opening}(A,B) = (A \ominus B) \oplus B$$

Image Morphology Operation: Closing

- Dilation followed by erosion

- Functions
  - Connects narrow breaks
  - Fills small holes

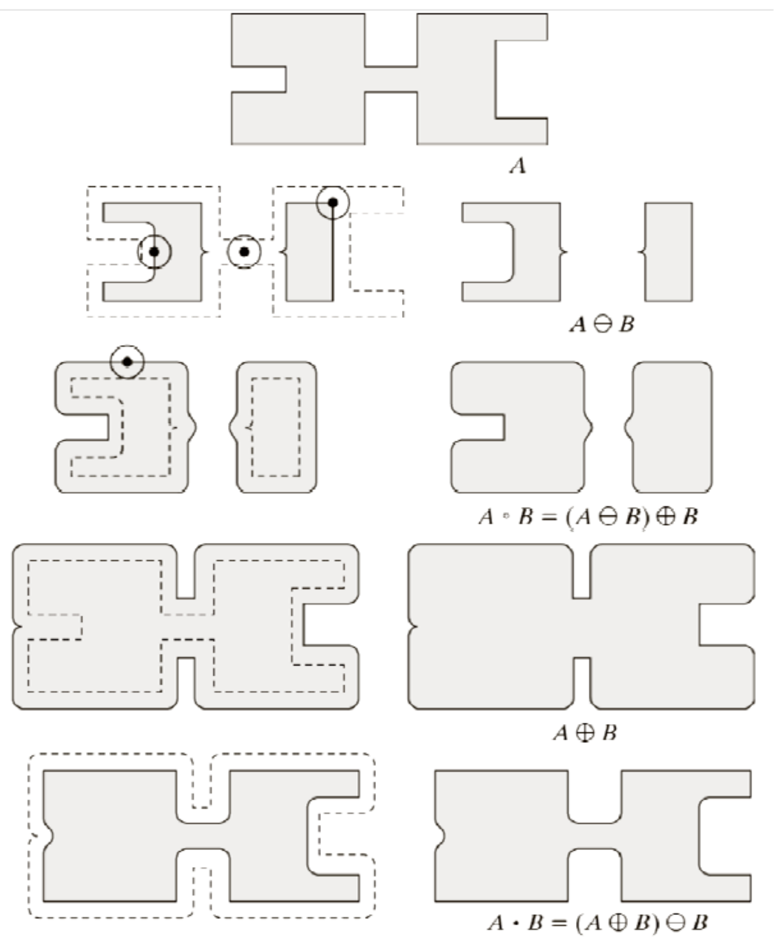


a b c

FIGURE 9.9 (a) Structuring element B "rolling" on the outer boundary of set A. (b) Heavy line is the outer boundary of the closing. (c) Complete closing (shaded).

$$\text{Closing}(A,B) = (A \oplus B) \ominus B$$

An Example with More Complex Geometry



a
b c
d e
f g
h i

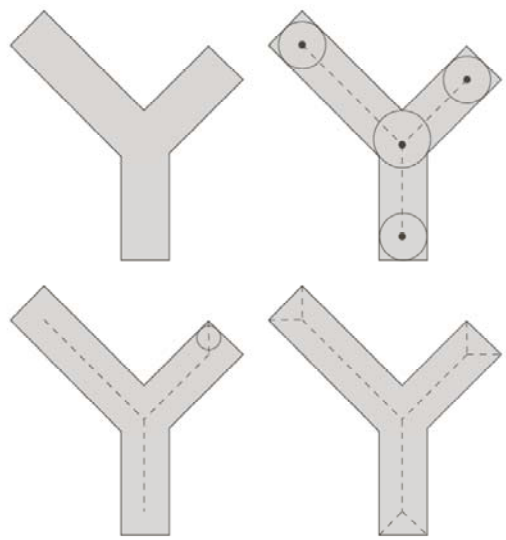
FIGURE 9.10 Morphological opening and closing. The structuring element is the small circle shown in various positions in (b). The SE was not shaded here for clarity. The dark dot is the center of the structuring element.

Common Applications of Image Morphology

- Boundary extraction

- Hole filling

- Skeleton extraction
  - More on this next class
  - Fast Marching Algorithm



a b
c d

FIGURE 9.23 (a) Set A. (b) Various positions of maximum disks with centers on the skeleton of A. (c) Another maximum disk on a different segment of the skeleton of A. (d) Complete skeleton.

Region-Based Segmentation

- Segmentation based on region growth

  Starting with seeds, recursive grow from the seeds based on

  1) Similarities; 2) Connectivity or adjacency

- Seed selection

  - Strong evidence
  - Random sampling

- Criteria for stopping the iteration

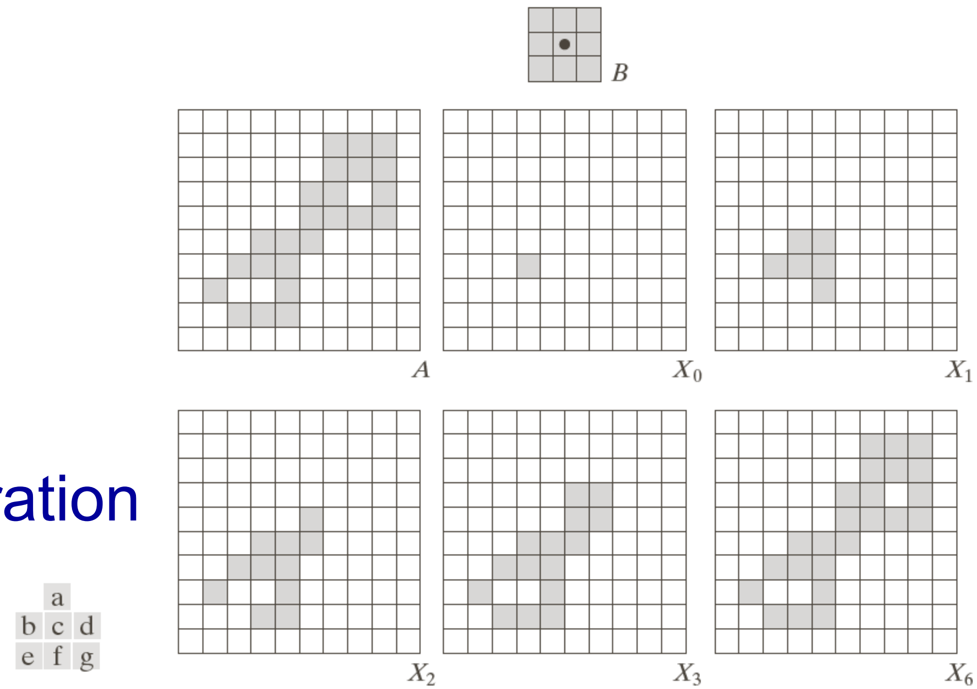
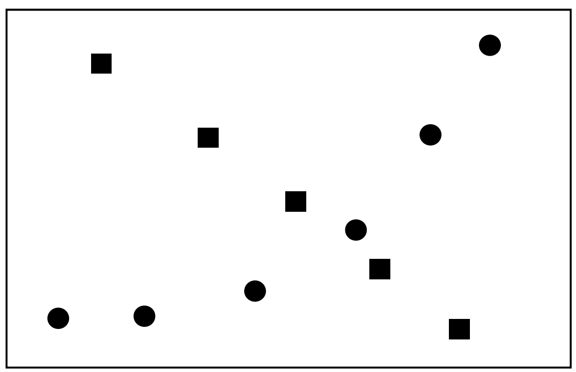


FIGURE 9.17 Extracting connected components. (a) Structuring element. (b) Array containing a set with one connected component. (c) Initial array containing a 1 in the region of the connected component. (d)–(g) Various steps in the iteration of Eq. (9.5-3).

Region-Based Segmentation

- Fundamental limitations of region-based segmentation
  - Decision is typically based on local measures; No global information



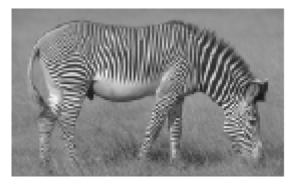


Image Morphology References

[1] Soille, *Morphological image analysis: principles and applications*, Springer, 2002.

[2] Dougherty & Lotufo, *Hands-on morphological image processing*, SPIE, 2003.