Reconocimiento de Escritura

Daniel Keysers

Image Understanding and Pattern Recognition
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Mar-2007

UNIVERSIDAD POLITECNICA DE VALENCIA

Outline

Lecture Organization

Introduction to Text Recognition

Introduction to Image Processing

Outline

Lecture Organization

Introduction to Text Recognition

Introduction to Image Processing

Lecturers

- Daniel Keysers
  studied Computer Science in Aachen and Madrid
  PhD in Computer Science, Aachen
  research stays at UP Valencia and MSR, Redmond
  currently at DFKI, Germany
  - Statistical Pattern Recognition
  - Document Image Analysis
  - Modeling of Image Variability
  - Image Retrieval
- Alejandro Hector Toselli
  studied Electrical Engineering in Tucuman, Argentina
  PhD in Computer Science, UP Valencia
  currently at ITI, Valencia
  - On-line and Off-line Handwriting Recognition
  - Computer Vision
  - Image Preprocessing

Lectures

Times, Locations and Tentative Schedule:

by Daniel:

- Mar-27 — Introduction, Image Processing Basics (2.5h)
- Mar-30 — Isolated Character & Handwriting Recognition (4h)
- Apr-02 — Layout Analysis (2h)
- Apr-03 — OCR (2.5h)
- Apr-04 — Applications (3h)

by Alejandro (starting Apr-20):

- Off-Line Continuous Handwriting Recognition based on HMMs
- Use of Language Models
- On-Line Handwriting Recognition

Web-Site and Exercises

Web-Site:

- preliminary: http://www.iupr.org/~keysers/res/

Exercises:

- Exercise sheet for first part of lecture: today
- You can choose any combination of exercises you prefer.
- You should show your solutions before or on April-04

Literature

- Document Image Analysis
  - H. Bunke, P.S.P. Want: Handbook of Character Recognition and Document Image Analysis. World Scientific, 1997.
  - S.V. Rice, G. Nagy, T. A. Nartker: Optical Character Recognition: An Illustrated Guide to the Frontier. Kluwer, 1999.
- Image Processing
  - R.C. Gonzalez, Richard E. Woods: Digital Image Processing. Prentice Hall, 2nd edition, 2002.
  - B. Jähne: Digital Image Processing. Springer, 1997.
- Language Modeling
  - F. Jelinek: Statistical Methods for Speech Recognition. MIT Press, 1998.

Literature

- Pattern Recognition
  - Duda, Hart, Stork: Pattern Classification. Wiley, 2000.
  - Hastie, Tibshirani, Friedman: The Elements of Statistical Learning. Springer, 2003.
  - Bishop: Pattern Recognition and Machine Learning. Springer, 2006.
- Articles
  - G. Nagy: Twenty Years of Document Image Analysis in PAMI. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), 38-62, 2000.
  - S. Mori, S.Y. Ching, K. Yamamoto: Historical Review of OCR Research and Development. Proceedings of the IEEE 80(7), 1029-1058. 1992.
  - G. Nagy, T.A. Nartker, S.V. Rice: Optical Character Recognition: An Illustrated Guide to the Frontier. Procs. Document Recognition & Retrieval, SPIE 3967, 58-69.
- Also interesting:
  - A.J. Sellen, R.H.R. Harper: The Myth of the Paperless Office. MIT Press, 2001.

Prerequisites

We are assuming basic knowledge about pattern recognition theory including HMMs, FSTs, statistical methods, neural networks, etc.

We do not assume much knowledge in the area of image processing in general. Therefore, we will present an introduction to the relevant image processing topics.

Outline

Lecture Organization

Introduction to Text Recognition

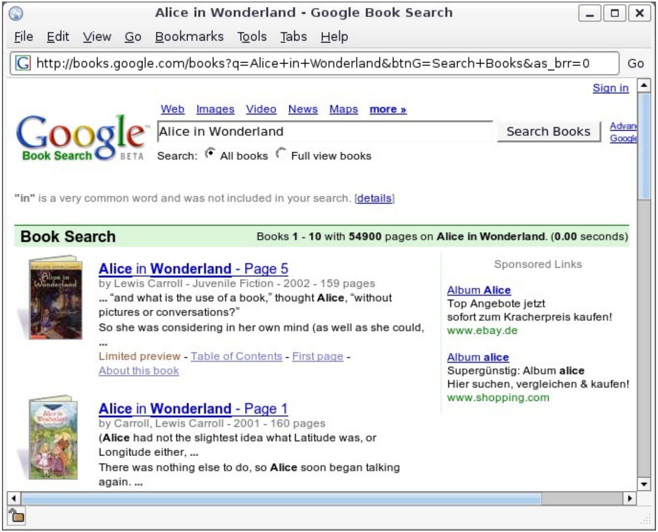
Introduction to Image Processing

Motivation

- WWW and e-mail *increase* amount of paper printouts
- paper will continue to play an important role in office life
- electronic library initiatives everywhere
- personal impression from recent CeBIT:
  document management systems and incoming mail
  automation is a huge business

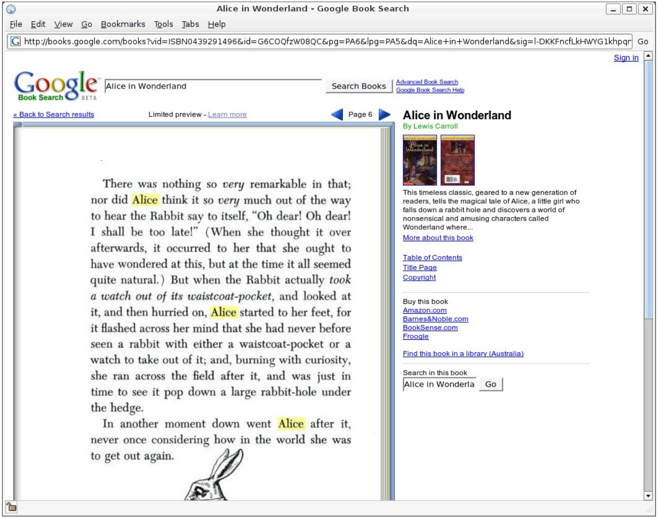
Google Book Search

Google Book Search

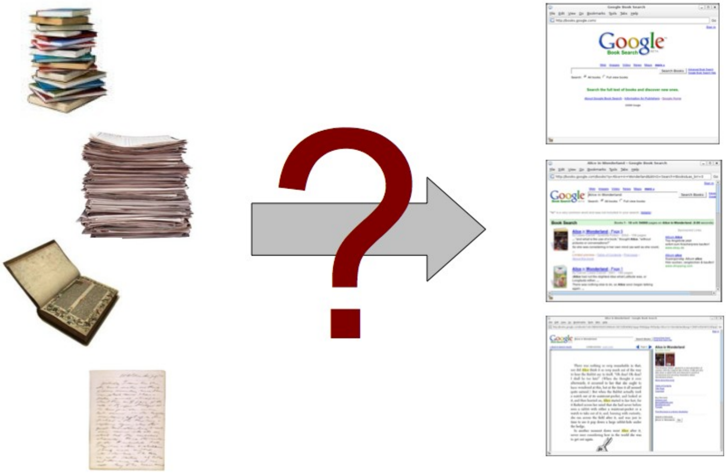


UNIVERSIDAD POLITECNICA DE VALENCIA

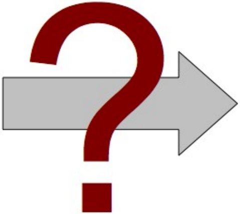
Google Book Search



Google Book Search



Google Book Search



machine learning, artificial intelligence, statistics
(+ excellent infrastructure)

Relation to Pattern Recognition

Text Recognition and Document Image Analysis are located in the intersection of pattern recognition and image processing

- handwriting recognition
- optical character recognition
- document understanding
- license plate recognition
- form reading

"Document image analysis (DIA) is the theory and practice of recovering the symbol structure of digital images scanned from paper or produced by computer." (Nagy, 2000)

There are some applications in text recognition that do not involve image processing directly, e.g. on-line handwriting recognition.

Applications

application areas:

- postal automation
- incoming mail processing
- digital libraries
- forms processing
- bank check reading
- document management systems
- license plate reading
- bar-code reading
- reading machines for the blind
- processing of maps and technical drawings
- CAPTCHA analysis
- on-line handwriting recognition
- screen OCR

Origins

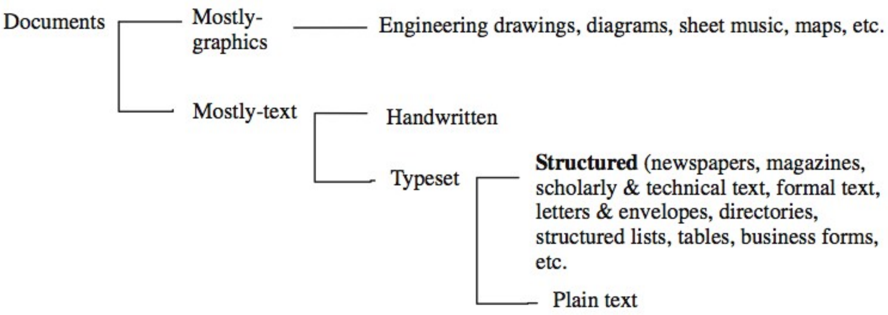
newspaper images transmitted over trans-atlantic cables and reproduced (1920s)

FIGURE 1.3 Unretouched cable picture of Generals Pershing and Foch, transmitted in 1929 from London to New York by 15-tone equipment. (McFarlane.)

specially designed printed digits used in business (1950s)

image processing development closely related to development of processing power of digital computers

Document Taxonomy

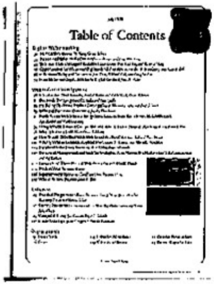


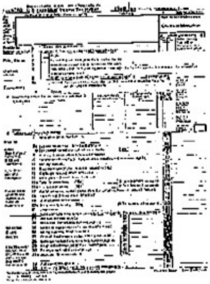
Document Taxonomy

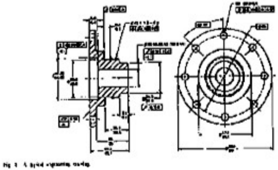

The New York Times

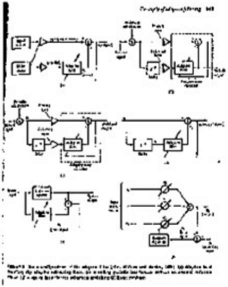

Wineries 463

Document Taxonomy









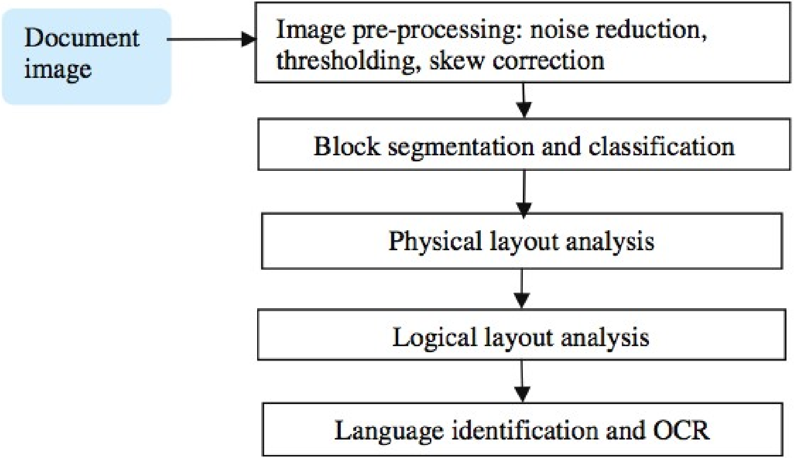
Document Taxonomy

Type	Example	DIA Task	Ancillary data
plain text (narrative or descriptive)	Moby Dick, Gettysburg Address	extract correct word order	English lexicon
newspaper, magazine	NY Times, Vogue	separate and reassemble articles; pointers to illustrations	publication-specific format
scholarly & technical text	IEEE-PAMI, Dr. Dobbs Journal	index: author, title, page; pointers to refs, figs, tables, footnotes, equations	abbreviations, acronyms, units
formal text	program listing, chess, bridge, recipe	extract executable, or compilable, form	program, chess, bridge syntax
letter, envelope	information request, complaint, recommendation	extract routing info; index: sender, date, subject	directories
directory	telephone directory, street index	extract name-attribute pairs	previous edition

Document Taxonomy

structured list	organization chart, table of contents, catalog	recover hierarchy; cross-references	previous edition
business form	order, invoice, subscription, survey, IRS-1040	link field content to dbms; convert to SGML or XML format;	formatted data, dbms, workflow system, lexicons
engineering drawing	assembly or part drawing; isometric view	convert to CAD format	part lists, drawing standards
schematic diagram	circuits, utility maps	extract net list or convert to CAD format	P-SPICE, manhole inventory
map	topographic quad, street map, road map	convert to GIS format	gazetteer, other maps, GIS
music score	Moonlight Sonata	recover MIDI representation	music syntax
table	airline schedules, stock quotes	construct formal model: headers <->entries	airline and stock abbreviations, previous edition

Typical Processing Chain



Outline

Lecture Organization

Introduction to Text Recognition

Introduction to Image Processing

Basic Definitions

An image is a two-dimensional function $f(x, y)$.

x, y: spatial coordinates
$f(x, y)$: intensity / gray level
$f(x, y) = (f_1(x, y), f_2(x, y), f_3(x, y))$: color

If the domain and the range of f are finite and discrete, we speak of a digital image.

The picture elements $(x, y, f(x, y))$ are called Pixels.

- Alternatives to a rectangular grid?
- Does the y-axis point up or down?
- Note: many similarities to 1-D, 3-D, (2+1)-D, (3+1)-D signal processing

Neighborhoods

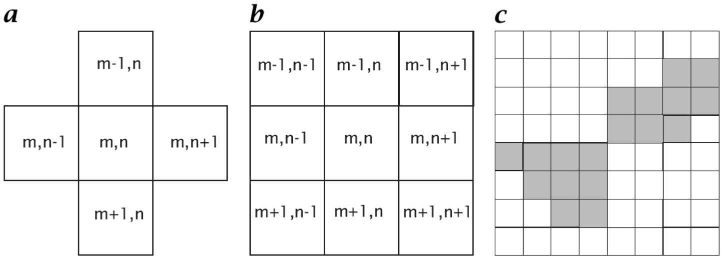


Figure 2.4: *Neighborhoods on a rectangular grid:* ***a*** *4-neighborhood and* ***b*** *8-neighborhood.* ***c*** *The black region counts as one object (connected region) in an 8-neighborhood but as two objects in a 4-neighborhood.*

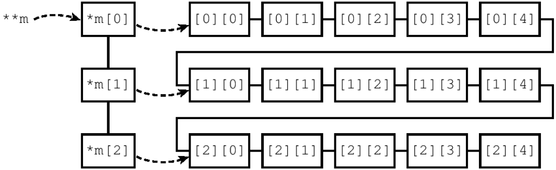
pixel adjacency

- 4-neighborhood
- 8-neighborhood

‘Programming’

representation of an image within the computer:

- 1-D array
- 2-D array (3-D array, 4-D array, ...)
- combined 1-D and 2-D access



Advantages and disadvantages of different dimensionalities?
Example Code:

```
template<typename T> struct Image {
  unsigned int width, height, channels;
  T *data;
}
```

Open Source Tools

- ImageMagick
- NetPBM
- ImageJ
- GIMP
- Octave
- Tesseract
- OCRopus
- FIRE
- MIT FST toolkit
- HTK
- ...

Building Blocks for Image Processing

- image acquisition
- image enhancement
- image restoration
- color image processing
- multiresolution processing
- compression
- morphological processing
- segmentation
- representation and feature extraction
- recognition

Image Formation Model

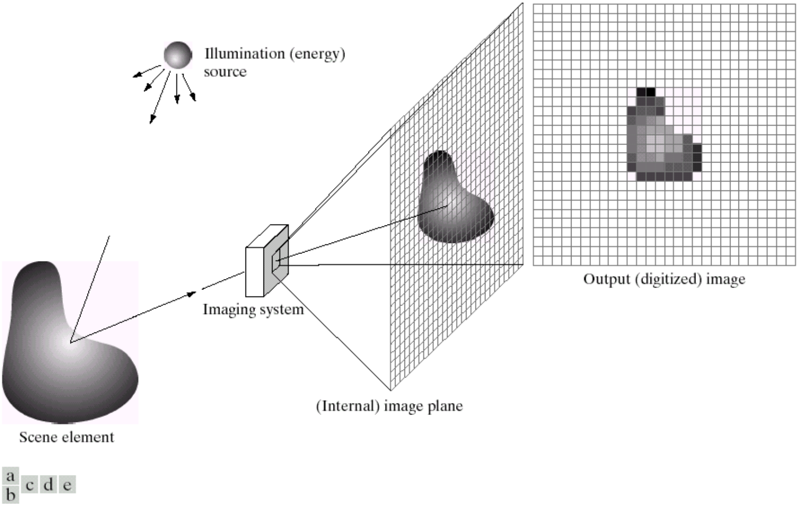


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

Image Formation Model

$$f(x, y) = i(x, y)r(x, y)$$

i: illumination

r: reflectance

typical values for r:

0.01 black velvet

0.80 white wall-paint

0.93 snow

effective gray scale is shifted to interval $[0, L - 1]$

typically $L = 2^k$ and we speak of a k-bit image

common values for k include 1, 8, 12, and 16

Digital Sensors

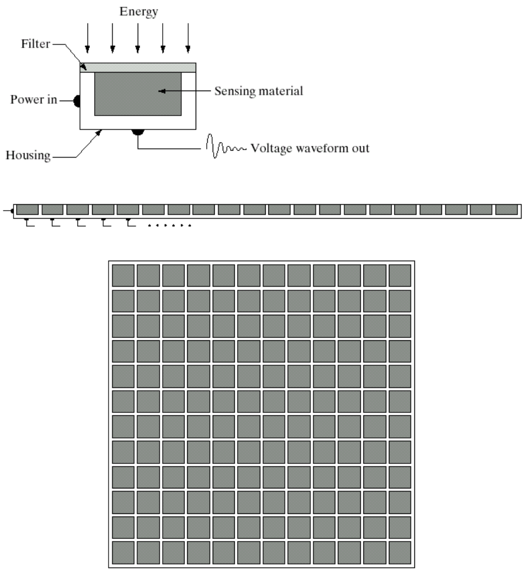
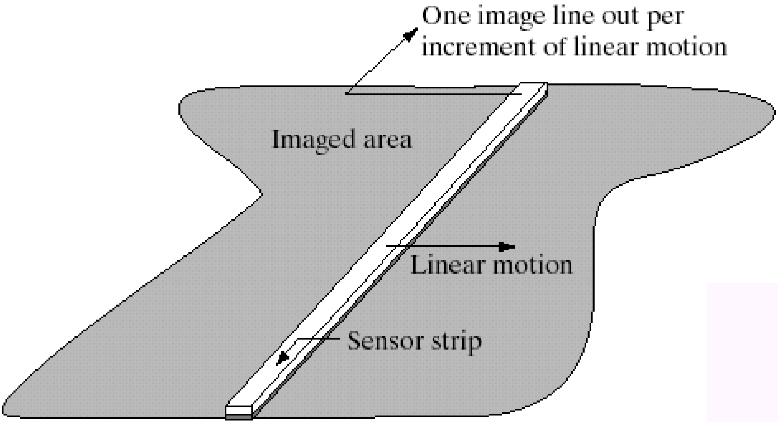


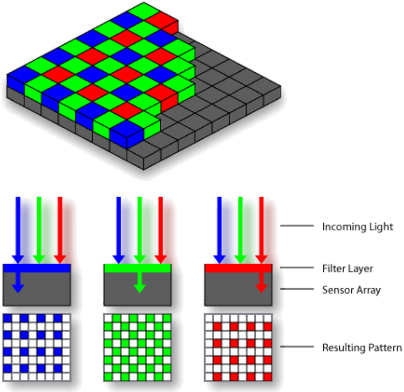
FIGURE 2.12
(a) Single imaging sensor.
(b) Line sensor.
(c) Array sensor.

Sensor Strips



→ most common use: scanners

Bayer Pattern



Bayer pattern

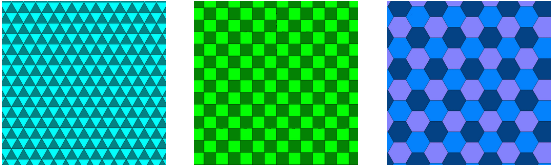
- mimic the human eye's greater resolving power for green light
- RAW data format
- interpolate final image (demosaicing)

Book-Scanner 1

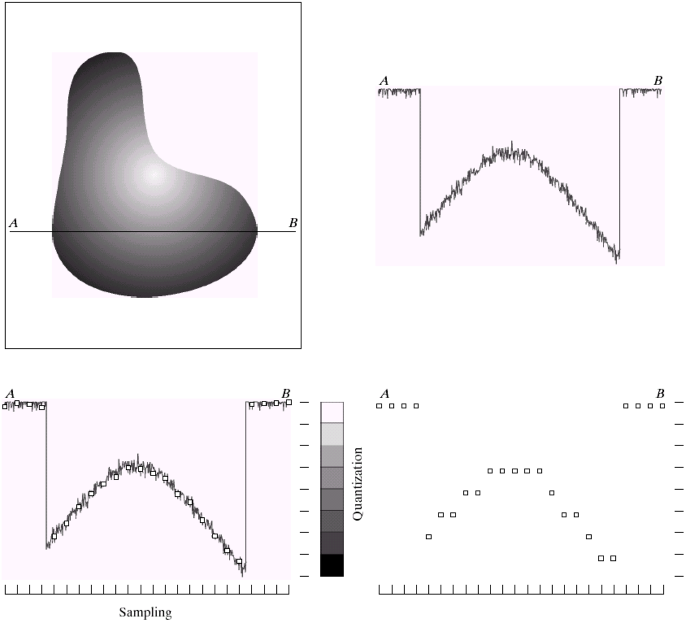
Book-Scanner 2

Digitization

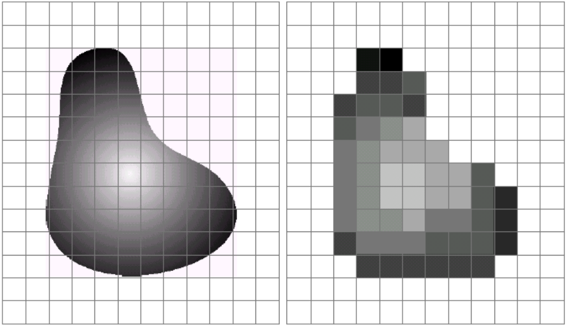
- sampling = digitizing coordinate values
- quantization = digitizing amplitude values
- image quality depends on number of samples and gray levels



Digitization



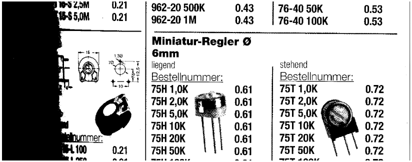
Digitization

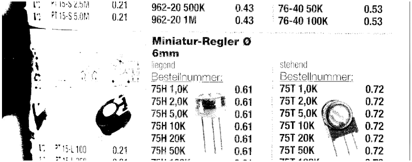


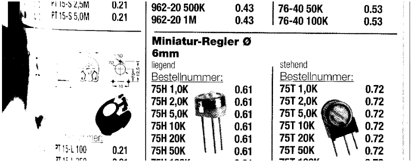
a b

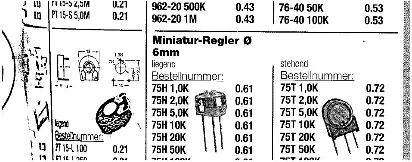
FIGURE 2.17 (a) Continuos image projected onto a sensor array. (b) Result of image sampling and quantization.

2-level Quantization — Binarization









- local/global methods
- different methods used to determine threshold

Image Matrix

use integer values for discrete coordinates

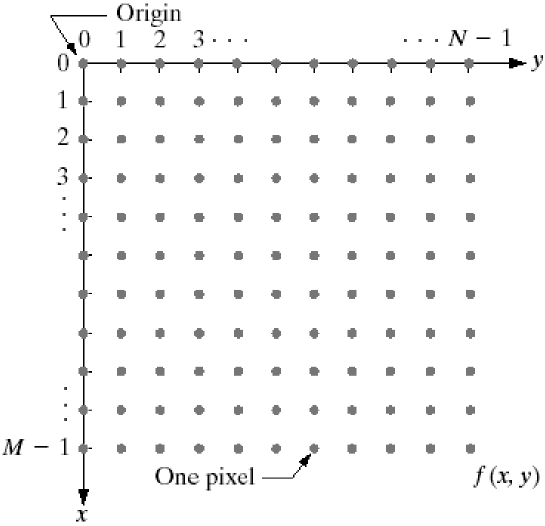


Image Matrix

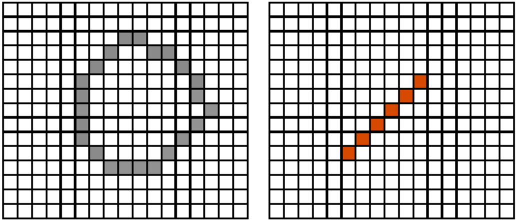
$$f(x,y) = \begin{pmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{pmatrix}$$

Note: The discrete grid implies that

- even 'simple' transformations like scaling and rotation
- even 'simple' concepts like lines

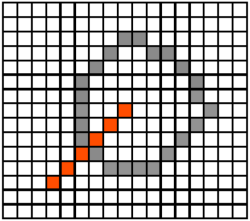
as known for the continuous plane have to be reconsidered carefully

Closed Curve Paradox



line 1 line 2

solid lines if
8-neighbourhood
is used



line 2 does not
intersect line 1
although it crosses
from the outside to the
inside

Effects of Digitization

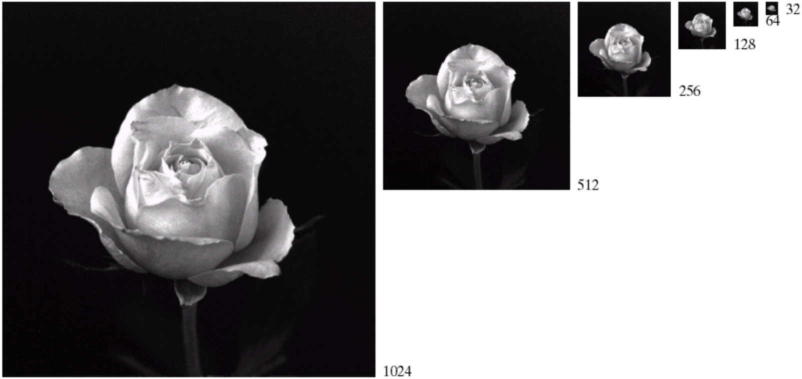
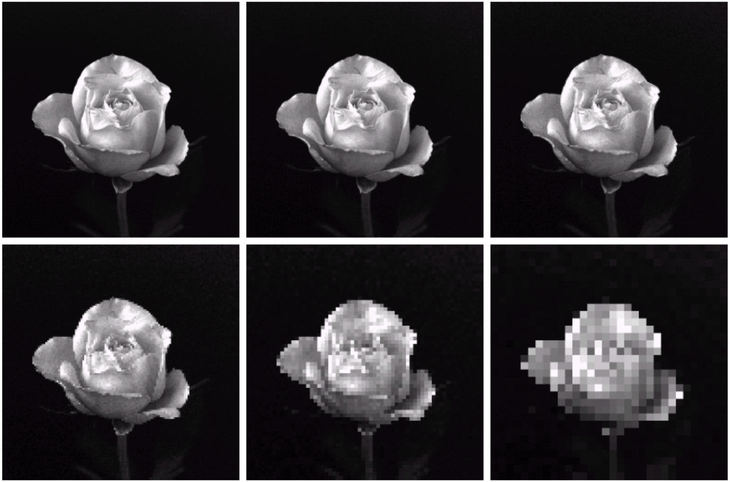


FIGURE 2.19 A 1024 × 1024, 8-bit image subsampled down to size 32 × 32 pixels. The number of allowable gray levels was kept at 256.

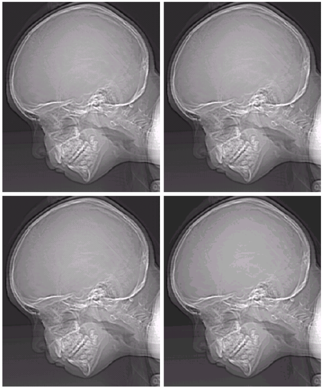
Effects of Digitization



a b c
d e f

FIGURE 2.20 (a) 1024 × 1024, 8-bit image. (b) 512 × 512 image resampled into 1024 × 1024 pixels by row and column duplication. (c) through (f) 256 × 256, 128 × 128, 64 × 64, and 32 × 32 images resampled into 1024 × 1024 pixels.

Effects of Digitization

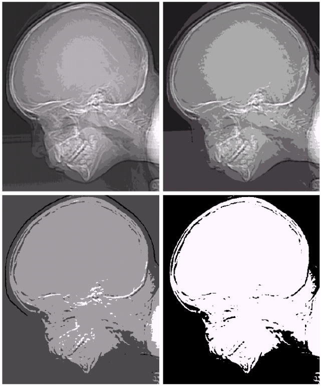


a b
c d

FIGURE 2.21 (a) 452 × 374, 256-level image. (b)–(d) Image displayed in 128, 64, and 32 gray levels, while keeping the spatial resolution constant.

e f
g h

FIGURE 2.21 *(Continued)* (e)–(h) Image displayed in 16, 8, 4, and 2 gray levels. (Original courtesy of Dr. David R. Pickens, Department of Radiology & Radiological Sciences, Vanderbilt University Medical Center.)



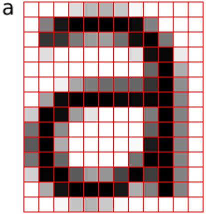
Storing Images

- storing in main memory: representation
- storing on disk: image formats

Bitmaps

names: array, matrix, bitmap, raster image

most common representation format



1.0	1.0	1.0	0.9	0.6	0.6	0.6	1.0	1.0	1.0	1.0	1.0
1.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0	1.0	1.0
1.0	0.2	0.2	0.5	0.6	0.6	0.5	0.0	0.0	0.5	1.0	1.0
1.0	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.0	0.0	0.9	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
1.0	1.0	1.0	0.5	0.5	0.5	0.5	0.5	0.4	0.0	0.5	1.0
1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	1.0
0.9	0.0	0.0	0.6	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
0.5	0.0	0.6	1.0	1.0	1.0	1.0	1.0	0.5	0.0	0.5	1.0
0.5	0.0	0.7	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.5	1.0
0.6	0.0	0.6	1.0	1.0	1.0	1.0	0.5	0.0	0.0	0.5	1.0
0.9	0.1	0.0	0.6	0.7	0.7	0.5	0.0	0.5	0.0	0.5	1.0
1.0	0.7	0.1	0.0	0.0	0.0	0.1	0.9	0.8	0.0	0.5	1.0
1.0	1.0	1.0	0.8	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0

possible problem:
hold 100 5-MPixel 24-bit color images in memory?

Pixel Neighborhoods

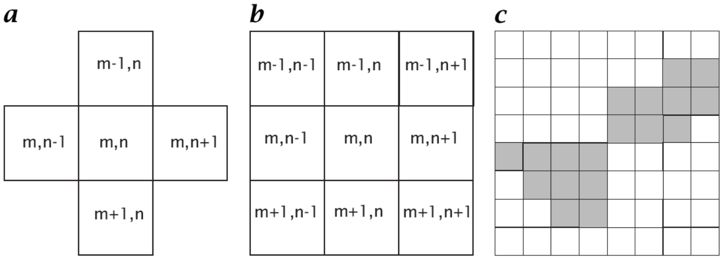


Figure 2.4: *Neighborhoods on a rectangular grid:* ***a*** *4-neighborhood and* ***b*** *8-neighborhood.* ***c*** *The black region counts as one object (connected region) in an 8-neighborhood but as two objects in a 4-neighborhood.*

Pixel Adjacency

pixel adjacency for (x, y):

- 4-neighborhood
  $N_4((x,y)) = \{(x+1,y),(x,y+1),(x-1,y),(x,y-1)\}$
- 8-neighborhood
  $N_8((x,y)) = N_4((x,y)) \cup$
  $\{(x+1,y+1),(x-1,y+1),(x-1,y-1),(x+1,y-1)\}$

p and q (with $f(p) = f(q)$) are

- 4-adjacent if $q \in N_4(p)$
- 8-adjacent if $q \in N_8(p)$

Pixel Connectivity

A path from p to q is a sequence of pixels $p_0, \dots, p_N$ with $p_0 = p$, $p_N = q$ and $p_{n+1} \in N(p_n)$ for $n = 0, \dots, N-1$.

path length: N
closed path: $p = q$

p and q are connected if a path exists between them.

A maximal set S of pixels for which any two $p, q \in S$ are connected is called a connected component.

Computation of Connected Components

compute connected components on binary image f

$\rightarrow$ label each foreground pixel with a label $l(p)$

- with p scan image (in reading order)
  - if $f(p) = 0$ continue
  - let S be the set of pixels q that have already been visited and for which $q \in N(p)$ and $f(q) = 1$
  - if $|S| = 0$ assign new label to p and continue
  - if $|l(S)| = 1$ assign to p the label in $l(S)$ and continue
  - if $|l(S)| > 1$ assign to p one of the labels in $l(S)$ and make note of the equivalence of all labels in $l(S)$
- sort labels into equivalence classes and assign unique labels