Information Processing Technology of Internet of Things

Chapter 4
Visual Information Processing

Wu Liu

Beijing Key Lab of Intelligent Telecomm. Software and Multimedia
Beijing University of Posts and Telecommunications

4.1 Image & Its Representation

4.1.1 Introduction

Introduction

- Computer vision aims to duplicate the effect of human vision by electronically perceiving and understanding an image
- Computer vision is difficult
  - Loss of information in 3D ->2D
  - Interpretation of image
  - Noise
  - Too much data
  - Brightness
  - Local window vs. need for global view

Image representation

- Image understanding by a machine is to find a relation between input image(s) and previously established models of the observed world.
  - Include several steps and several levels representing the image
- Image representation can be roughly divided according to data organization into four levels,

Image representation

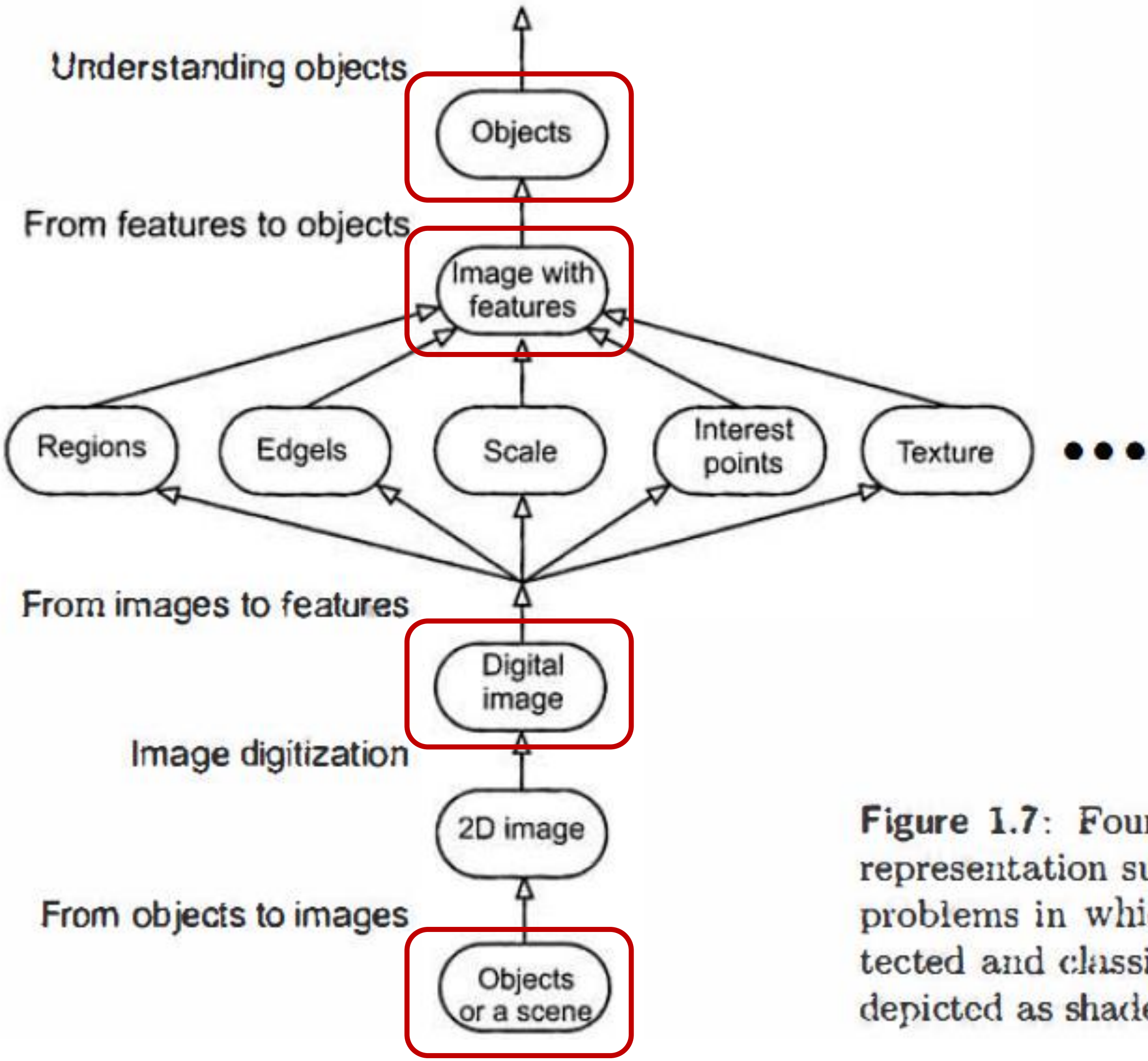


Figure 1.7: Four possible levels of image representation suitable for image analysis problems in which objects have to be detected and classified. Representations are depicted as shaded ovals.

Image analysis tasks

- Image is digitized first
  - Be represented by a rectangular matrix with elements corresponding to the brightness at appropriate image locations;
  - or be presented in color, implying (usually) three channels: red, green and blue
- Two levels are often distinguished:
  - Low-level image processing
  - High-level image understanding

Image analysis tasks

- Low-level processing methods: use very little knowledge about the content of images
  - image compression
  - pre-processing methods for noise filtering
  - edge extraction
  - image sharpening
  - …
- High-level processing is based on knowledge, goals, and plans of how to achieve those goals, and artificial intelligence methods are widely applicable.
  - High-level computer vision tries to imitate human cognition and the ability to make decisions according to the information contained in the image.

Image analysis tasks

- Low-level data
  - original images represented by matrices composed of brightness (or similar) values
- High-level data
  - originate in images as well, but only those data which are relevant to high-level goals are extracted, reducing the data quantity considerably.
  - represent knowledge about the image content, for example, object size, shape, and mutual relations between objects in the image

4.1.2 Image representations

The continuous image function

- Mathematical models are often used to describe images and other signals. A signal is a function depending on some variable with physical meaning
- The (gray-scale) image function values correspond to brightness at image points.
  - Brightness integrates different optical quantities.
  - The image is intrinsically two dimensional (2D)
- A monochromatic static image is represented by a continuous image function f (x, y) whose arguments are two co-ordinates in the plane.

The continuous image function

- Computerized image processing uses digital image functions which are usually represented by matrices, so co-ordinates are natural numbers.
  - The domain of the image function is a region R in the plane

$$R = \{(x, y),\ 1 \leq x \leq x_m,\ 1 \leq y \leq y_n\}$$

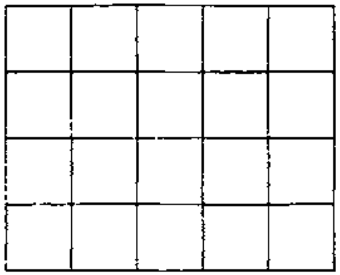
  - where xm , yn represent the maximal image co-ordinates. (horizontal x-axis, vertical y-axis , origin bottom-left)
- In monochromatic images, the lowest image function value corresponds to black and the highest to white. Brightness values bounded by these limits are gray-levels.

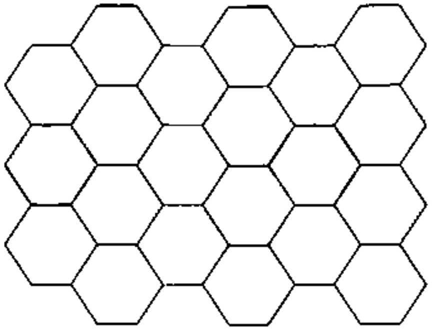
Image digitization

- Image digitization means that the function f (x, y) is sampled into a matrix with M rows and N columns.
- Image quantization assigns to each continuous sample an integer value——the continuous range of the image function f (x, y) is split into K intervals.
- The finer the sampling (i.e., the larger M and N) and quantization (the larger K) , the better the approximation of the continuous image function f(x, y) achieved.

Image digitization——Sampling

- A continuous image is digitized at sampling points. These sampling points are ordered in the plane, and their geometric relation is called the grid (which are usually square or hexagonal).





- One infinitely small sampling point in the grid corresponds to one picture element also called a pixel or image element in the digital image.

Image digitization——Quantization

- A value of the sampled image fs (jΔx, kΔy) is expressed as a digital value in image processing. The transition between continuous values of the image function (brightness) and its digital equivalent is called quantization.
- Most digital image processing devices use quantization into k equal intervals. If b bits are used to express the values of the pixel brightness then the number of brightness levels is $k = 2^b$. Eight bits per pixel per channel (one each for red, green, blue) are commonly used.

Image digitization——Quantization

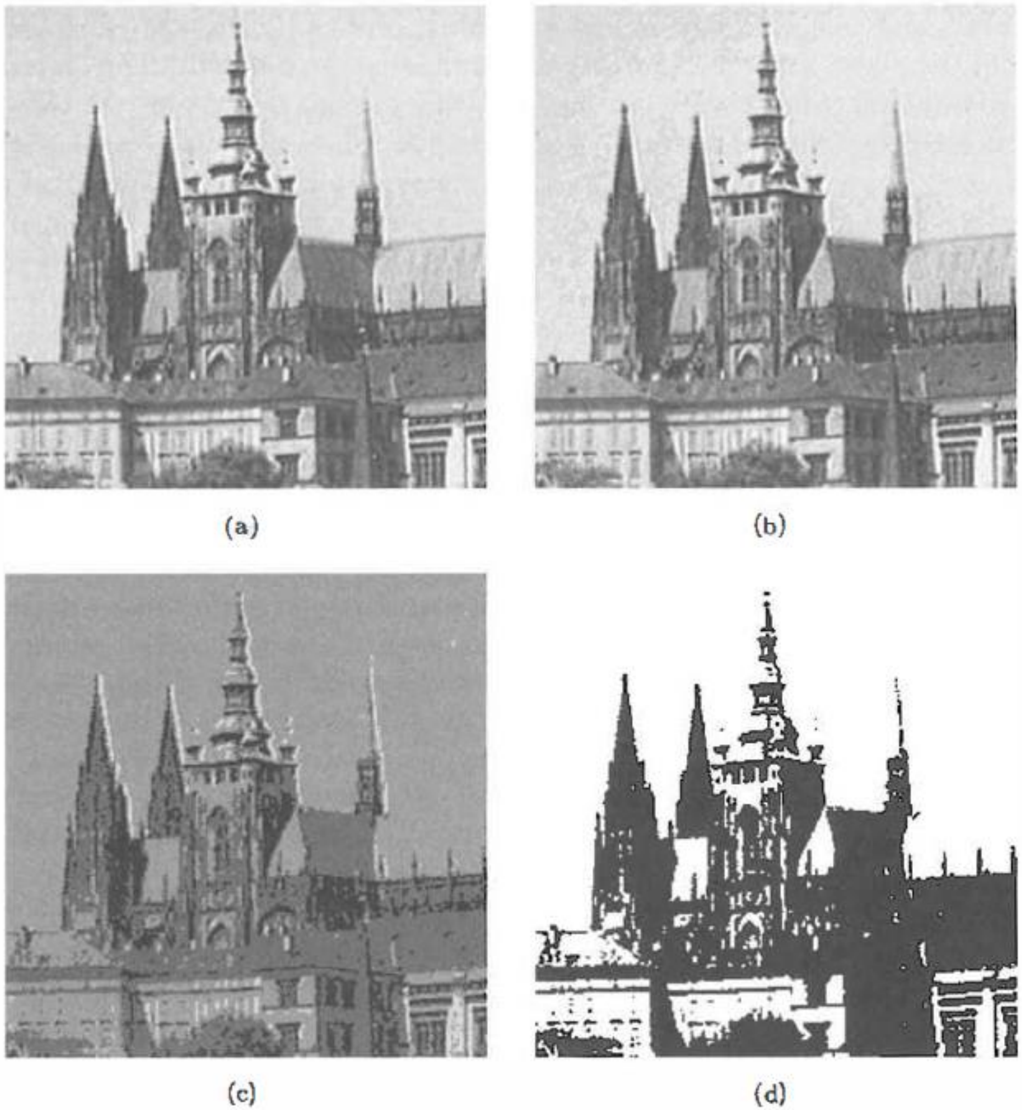


Figure 2.3: Brightness levels. (a) 64. (b) 16. (c) 4. (d) 2.

Metric and topological properties of digital images

- Distance
  - Any function D holding the following three condition is a 'distance' (or a metric)

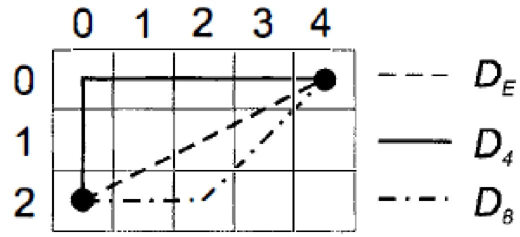
$$
\begin{aligned}
&D(p,q) \geq 0\,, \quad \text{for } (D(p,q) = 0 \text{ if and only if } p = q) && \textit{identity},\\
&D(p,q) = D(q,p)\,, && \textit{symmetry},\\
&D(p,r) \leq D(p,q) + D(q,r)\,, && \textit{triangular inequality}.
\end{aligned}
$$

- The distance between points with co-ordinates (i, j) and (h, k) may be defined in several different ways.
  - Euclidean distance D_E $\quad D_E\big((i,j),(h,k)\big) = \sqrt{(i-h)^2 + (j-k)^2}$
  - city block distance D_4 (also L_1 metric or Manhattan distance)

$$D_4\big((i,j),(h,k)\big) = |\,i-h\,| + |\,j-k\,|$$

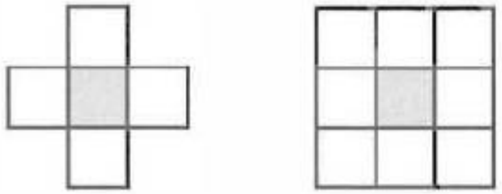
  - chessboard distance D_8

$$D_8\big((i,j),(h,k)\big) = \max\{\,|\,i-h\,|, |\,j-k\,|\,\}$$



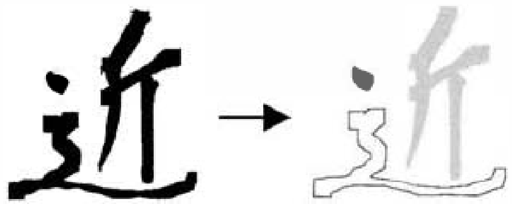
Metric and topological properties of digital images

- Pixel adjacency
  - Any two pixels (p, q) are called 4-neighbors if they have distance D_4 (p, q) = 1 .
  - 8-neighbors are two pixels with D_8(p, q) =1.



(a) 4-neighborhood (b) 8-neighborhood

- Region
  - set consisting of several adjacent pixels
  - a region is a connected set



- Contiguous
  - if there is a path between two pixels in the set of pixels in the image, these pixels are called contiguous.
  - Alternatively, we can say that a region is a set of pixels in which each pair of pixels is contiguous.

Metric and topological properties of digital images

- Assume Ri are disjoint regions which do not touch the image bounds.
- Let R be the union of all regions Ri
- R^C be the set complement of R with respect to the image
- The subset of R^C which is contiguous with the image bounds is called the background
- the remainder of the complement R^C is called holes.
- A region is called simple contiguous if it has no holes.
- A region with holes is called multiple contiguous.

Metric and topological properties of digital images

- border of a region R (inner border)
  - the set of pixels within the region that have one or more neighbors outside R.
- outer border
  - the border of the background (i.e., it.s complement) of the region.

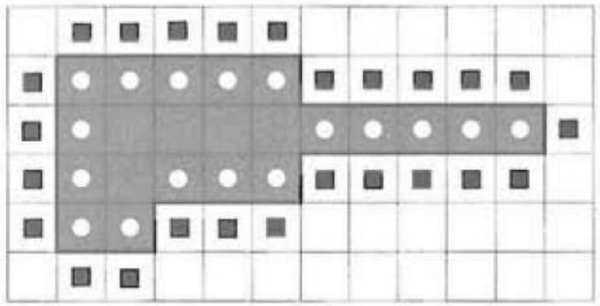


Figure 2.13: Inner borders of a region shown as white circles and outer borders shown as black squares. 4-neighborhood was considered.

Metric and topological properties of digital images

- Convex
  - If any two points within a region are connected by a straight line segment, and the whole line lies within the region, then this region is convex
- The property of convexity decomposes all regions into two equivalence classes: convex and non-convex.

Figure 2.14: A convex region (left) and non-convex region (right).

Metric and topological properties of digital images

- Convex hull
  - the smallest convex region containing the input region, possibly non-convex.
- The set inside the convex hull which does not belong to an object is called the deficit of convexity. This call be split into two subsets:
  - lakes (dark gray) are fully surrounded by the object;
  - bays (light gray) are contiguous with the border of the convex hull of the object.

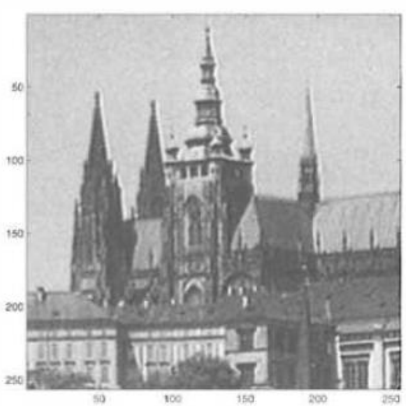
Figure 2.15: Description using topological components: An 'R' object, its convex hull, and the associated lakes and bays.

Histograms

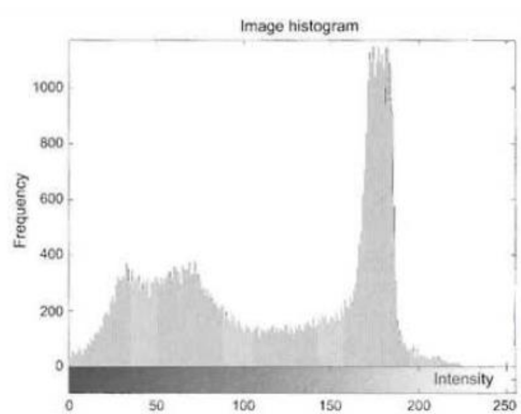
- The brightness histogram hf(z) of an image provides the frequency of the brightness value z in the image
  - the histogram of an image with L gray-levels is represented by a one-dimensional array with L elements.
  - We might want to find a first-order probability function p1(z; x, y) to indicate the probability that pixel (x, y) has brightness z.

Algorithm 2.2: Computing the brightness histogram

1. Assign zero values to all elements of the array h_f.
2. For all pixels (x,y) of the image f, increment $h_f(f(x,y))$ by 1.



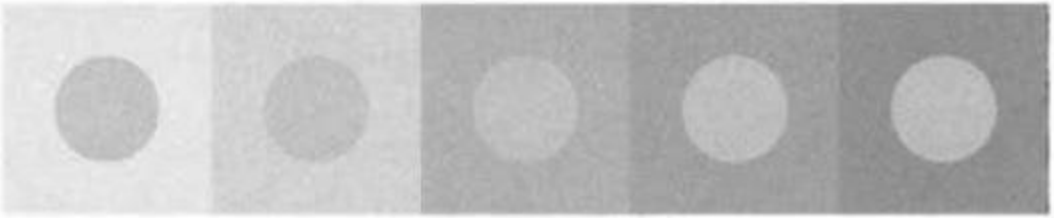
(a)



(b)

Visual perception of the image

- If an image is to be analyzed by a human the information should be expressed using variables which are easy to perceive; these are psycho-physical parameters such as contrast , border, shape, texture, color, etc.
- Contrast
  - the local change in brightness
  - defined as the ratio between average brightness of an object and the background
  - Apparent brightness depends very much on the brightness of the local surroundings; this effect is called conditional contrast.



Conditional contrast effect. Circles inside squares have the same brightness and are perceived as having different brightness values

Image quality

- An image might be degraded during capture, transmission, or processing, and measures of image quality can be used to assess the degree of degradation.
  - The quality required naturally depends on the purpose for which an image is used.
- Methods for assessing image quality can be divided into two categories
  - Subjective methods: the ultimate criterion is the perception of a selected group of professional and lay viewers, e.g., television
  - Objective methods: The quality of the image f(x, y) is usually estimated by comparison with a known reference image g(x, y), e.g., mean quadratic difference $\sum(g-f)^2$

Noise in images

- Real images are often degraded by some random errors- this degradation is usually called noise.
  - Noise can occur during image capture, transmission, or processing, and may be dependent on, or independent of, the image content .
- Noise is usually described by its probabilistic characteristics
  - Idealized noise, called white noise is often used. White noise has a constant power spectrum, meaning that all noise frequencies are present and have the same intensity.

Noise in images

- Gaussian noise
  - A random variable with a Gaussian (normal) distribution has its probability density function given by the Gaussian curve. In the 1D case the density function is

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

  - a very good approximation to noise that occurs in many practical eases.

Noise in images

- When an image is transmitted through some channel, noise which is usually independent of the image signal occurs.
- The signal-independent degradation is called additive noise and can be described by the model

$$f(x, y) = g(x, y) + \nu(x, y)$$

where the noise v and the input image g are independent variables.

Noise in images

- Signal-to-noise ratio (SNR)
  - computing the total square value of the noise contribution

$$E = \sum_{(x,y)} \nu^2(x,y)$$

  - computing the total square value of the observed signal

$$F = \sum_{(x,y)} f^2(x,y)$$

  - The signal-to-noise ratio is then $\mathrm{SNR} = \frac{F}{E}$
  - SNR represents a measure of image quality, with high values being 'good'.
  - often expressed in the logarithmic scale $\mathrm{SNR_{dB}} = 10 \log_{10} \mathrm{SNR}$

Noise in images

- Multiplicative noise $f = g\nu$
- Quantization noise occurs when insufficient quantization levels are used.
- Impulse noise means that an image is corrupted with individual noisy pixels whose brightness differs significantly from that of the neighborhood.
- Salt-and-pepper noise
  - used to describe saturated impulsive noise--an image corrupted with white and/or black pixels is an example. Salt-and-pepper noise can corrupt binary images.

Color spaces

- Color can be defined by almost any set of primary colors, e.g., red, green, and blue.
- Several different primary colors and corresponding color spaces are used in practice, and these spaces can be transformed into each other.
- Color spaces
  - RGB color space
  - CMY(Cyan, Magenta, Yellow) color model
  - HSV - Hue, Saturation, and Value (also known as HSB, hue, saturation, brightness)
  - HSL (hue, saturation, lightness/luminance), also known as HLS or HSI (hue, saturation, intensity) is similar to HSV.
  - …

4.1.3 Data structures for image analysis

Levels of image data representation

- The aim of computer visual perception is to find a **relation between an input image and models of the real world**.
- Several levels of visual information representation are defined on the way between the input image and the model
- Computer vision then comprises a design of the:
  - Intermediate representations (data structures) .
  - Algorithms used for the creation of representations and introduction of relations between them.

Levels of image data representation

- The visual information representations can be stratified in four levels –however, there are no strict borders between them.
- Four levels: from signals at a low level of abstraction to the description that a human can perceive.
  - iconic images: consists of images containing original data: integer matrices with data about pixel brightness
  - segmented images: Parts of the image are joined into groups that probably belong to the same objects
  - geometric representations: holding knowledge about 2D and 3D shapes
  - relational models: a higher level of abstraction, e.g., represented by semantic nets or frames

Traditional image data structures-Matrices

- A matrix is the most common data structure for low-level representation of an image.
- Elements of the matrix arc integer numbers corresponding to brightness, or to another property of the corresponding pixel of the sampling grid.
- Image information in the matrix is accessible through the co-ordinates of a pixel that correspond with row and column indices.
- The matrix contains spatial relations among semantically important parts of the image. The space is two-dimensional in the case of an image--a plane.

Traditional image data structures-Matrices

- Some special images that are represented by matrices
  - A binary image (an image with two brightness levels only) is represented by a matrix containing only zeros and ones.
  - Several matrices can contain information about one multispectral image. Each of these matrices contains one image corresponding to one spectral band.
  - Matrices of different resolution are used to obtain hierarchical image data structures.

Traditional image data structures-Matrices

- integral image
  - A matrix representation that holds global image information
  - its values ii(i, j) in the location (i, j) represent the sums of all the original image pixel-values left of and above (i, j):

$$ii(i,j) = \sum_{k \le i, l \le j} f(k,l)$$

where f is the original image.

Algorithm 4.2: Integral image construction

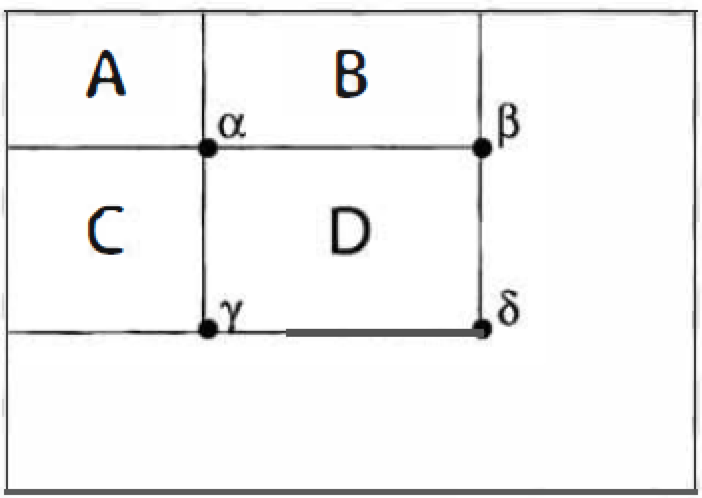
1. Let $s(i,j)$ denote a cumulative row sum, let $s(i,-1) = 0$.
2. Let $ii(i,j)$ be an integral image, let $ii(-1,j) = 0$.
3. Using a single row-by-row pass through the image, for each image pixel (i,j) calculate the cumulative row sums $s(i,j)$ and the integral image value $ii(i,j)$ using the recurrences

$$s(i,j) = s(i,j-1) + f(i,j) , \tag{4.2}$$
$$ii(i,j) = ii(i-1,j) + s(i,j) . \tag{4.3}$$

4. After reaching the lower right image corner pixel after a single pass through the image, the integral image ii is constructed.

Traditional image data structures-Matrices

- The main use of integral image data structures is in rapid calculation of simple rectangle image features at multiple scales.
  - Used for rapid object identification and for object tracking
  - As shown in the Figure, any rectangular sum can be computed using four array references.



Calculation of rectangle features from an integral image. The sum of pixels within rectangle D can be obtained using four array references.

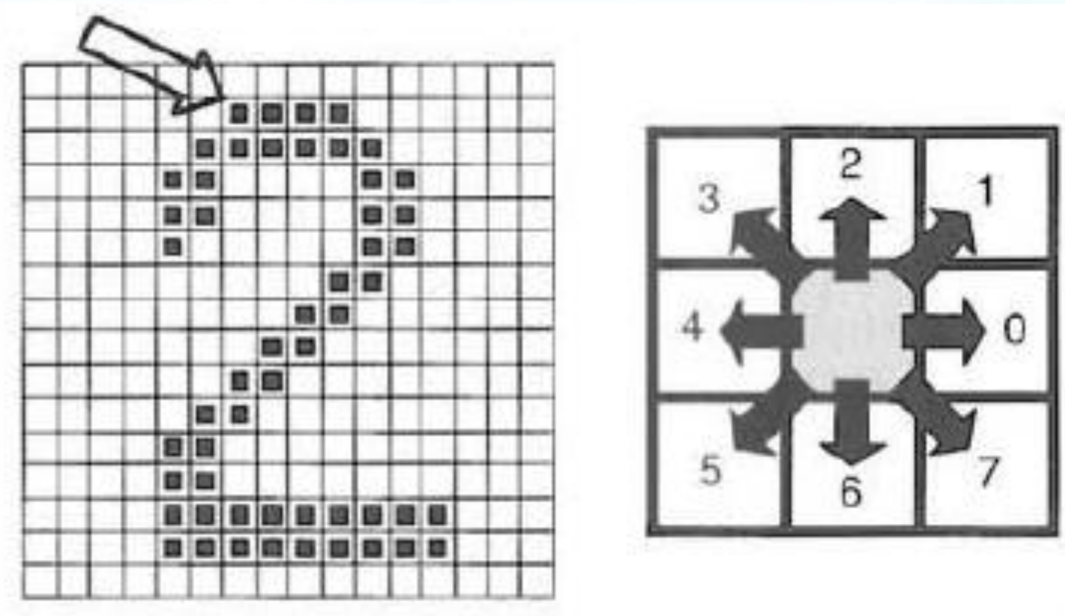
$$D_{sum} = ii(\delta) + ii(\alpha) - (ii(\beta) + ii(\gamma))$$

Traditional image data structures-Chains

- Chains are used for the description of object borders in computer vision. One element of the chain is a basic symbol.
- Chains are appropriate for data that can be arranged as a sequence of symbols, and the neighboring symbols in a chain usually correspond to the neighborhood of primitives in the image.
- Chain codes (and Freeman codes) are often used for the description of object borders, or other one-pixel-wide lines in images.
- The border is defined by the co-ordinates of its reference pixel and the sequence of symbols corresponding to the line of the unit length in several pre-defined orientations.

Traditional image data structures-Chains

- An example of a chain code is shown in the Figure , where 8-neighborhoods are used.



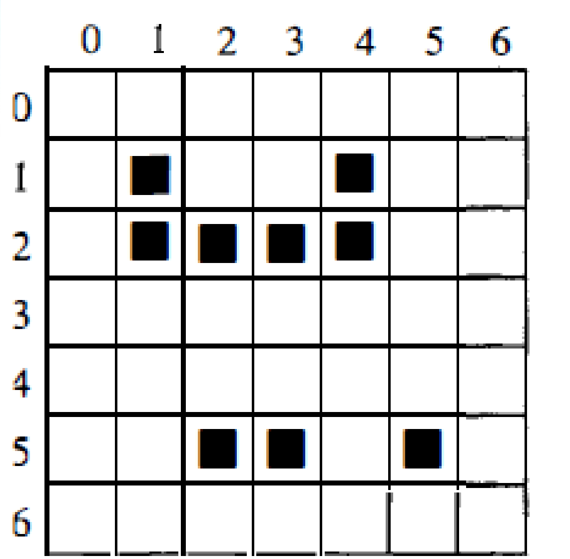
the reference pixel starting the chain is marked by an arrow:
000776655555566000000064444444422211111122344445652211.

Traditional image data structures-Chains

- Run length coding is quite often used to represent strings of symbols in an image matrix (for instance, FAX machines use run length coding).
- Consider a binary image first. Run length coding records only areas that belong to the object in the image; the area is then represented as a list of lists.
  - a representative one describes each row of the image by a sublist
  - The first element of which is the row number
  - Subsequent terms are co-ordinate pairs; the first element of a pair is the beginning of a run and the second is the end (the beginning and the end are described by column coordinates).
  - there can be several such sequences in the row.

Traditional image data structures-Chains

- Run length coding is illustrated in the Figure



Run length coding; the code is ((11144) (214) (52355))

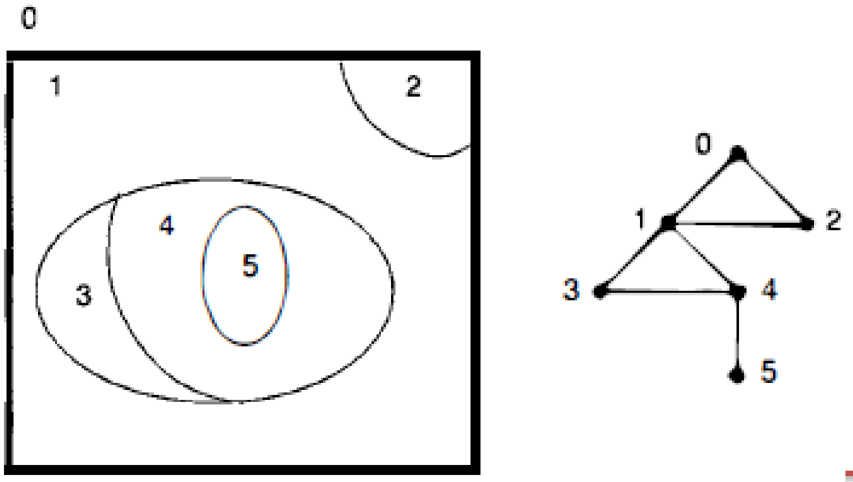
- The main advantage of run length coding is the existence of simple algorithms for intersections and unions of regions in the image.

Traditional image data structures-Topological data structures

- Topological data structures describe the image as a set of elements and their relations; these relations are often represented using graphs.
- A graph $G = (V, E)$ is an algebraic structure which consists of a set of nodes $V = \{vl, v2, \ldots, vn\}$ and a set of arcs $E = \{e1, e2, \ldots, em\}$.
- Each arc ek is incident to an unordered (or ordered) pair of nodes $\{vi, vj\}$ which are not necessarily distinct.
- The degree of the node is equal to the number of incident arcs of the node.
- An evaluated graph (also weighted graph) is a graph in which values are assigned to arcs, to nodes, or to both--these values may, for example, represent weights, or costs.

Traditional image data structures-Topological data structures

- Region adjacency graph: nodes correspond to regions and neighboring regions are connected by an arc
- The segmented image consists of regions with similar properties (brightness, texture, color, . . .) that correspond to some entities in the scene, and the neighborhood relation is fulfilled when the regions have some common border.
- An example of region adjacency graph



the label 0 denotes pixels out of the image

Traditional image data structures-Topological data structures

- The region adjacency graph has several attractive features. If a region encloses other regions, then the part of the graph corresponding with the areas inside can be separated by a cut in the graph. Nodes of degree 1 represent simple holes.

Hierarchical data structures

- Computer vision is by its nature very computationally expensive
- One of the solutions is to use parallel computers. Unfortunately there are many computer vision problems that are very difficult to divide among processors, or decompose in any way.
- Hierarchical data structures make it possible to use algorithms which decide a strategy for processing on the basis of relatively small quantities of data.
  - pyramids
  - quad trees

Hierarchical data structures--Pyramids

- Pyramids are among the simplest hierarchical data structures.
  - M-pyramids (matrix-pyramids)
  - T-pyramids (tree-pyramids)
- A Matrix-pyramid (M-pyramid)
  - a sequence $\{M_L, ., M_{L-1}, \ldots, M_0\}$ of images
  - M_L has the same dimensions and elements as the original image,
  - M_{i-1} is derived from the M_i; by reducing the resolution by one-half.
  - M_0 corresponds to one pixel only.
  - used when it is necessary to work with an image at different resolutions simultaneously. (An image having one degree smaller resolution in a pyramid contains four times less data, so it is processed approximately four times as quickly.)

Hierarchical data structures--Pyramids

- Tree-pyramids: a tree structure
  - Let 2^L be the size of an original image (the highest resolution) . A tree-pyramid (T-pyramid) is defined by:

1. A set of nodes $P = \{P = (k, i, j)$ such that level $k \in [0, L]$; $i, j \in [0, 2^k - 1]\}$.
2. A mapping F between subsequent nodes P_{k-1}, P_k of the pyramid

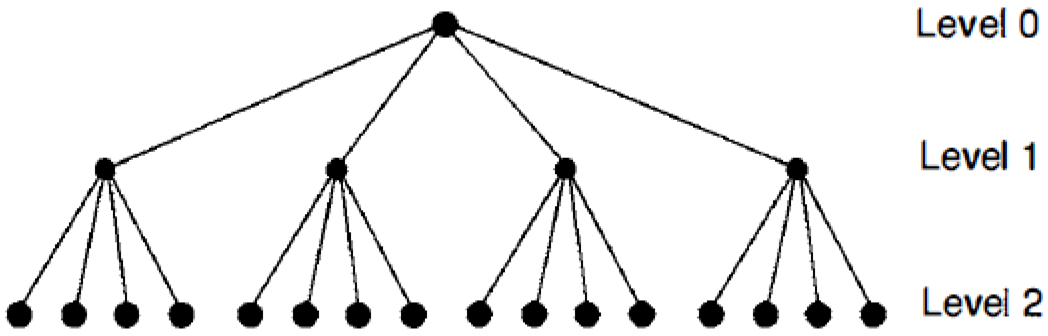
$$F(k, i, j) = (k - 1, i \text{ div } 2, j \text{ div } 2),$$

   where 'div' denotes whole-number division.

3. A function V that maps a node of the pyramid P to Z, where Z is the subset of the whole numbers corresponding to the number of brightness levels, for example, $Z = \{0, 1, 2, \ldots, 255\}$.

Hierarchical data structures--Pyramids

- Tree-pyramids:
  - elements of the set of nodes P = {(k, i, j)} correspond with individual matrices in the M-pyramid (k is called the level of the pyramid). An image P = {(k, i, j)} for a specific k constitutes an image at the kth level of the pyramid.
  - F is the so-called parent mapping, which is defined for all nodes Pk of the T-pyramid except its root (0, 0, 0) .
  - Every node of the T-pyramid has four child nodes except leaf nodes, which are nodes of level L that correspond to the individual pixels in the image.



Hierarchical data structures--Pyramids

- Tree-pyramids:
  - Values of individual nodes of the T-pyrarmid are defined by the function V.
  - Values of leaf nodes are the same as values of the image function (brightness) in the original image at the finest resolution
  - values of nodes in other levels of the tree are either an arithmetic mean of four child nodes or they are defined by coarser sampling, meaning that the value of one child (e.g. , top left) is used.