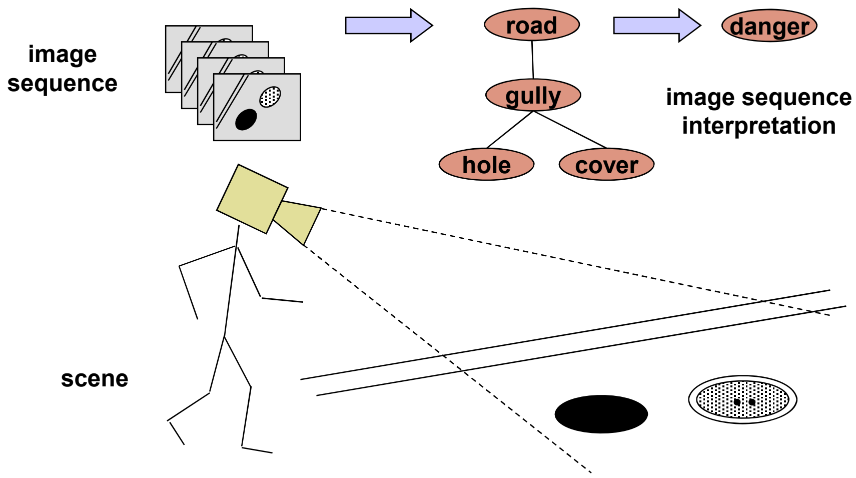
Definition of Image Understanding

Image understanding is the task-oriented reconstruction and interpretation of a scene by means of images

"scene":	section of the real world stationary (3D) or moving (4D)
"image":	view of a scene projection, density image (2D) depth image (2 1/2D) image sequence (3D)
"reconstruction and interpretation":	computer-internal scene description quantitative + qualitative + symbolic
"task-oriented":	for a purpose, to fulfil a particular task context-dependent, supporting actions of an agent

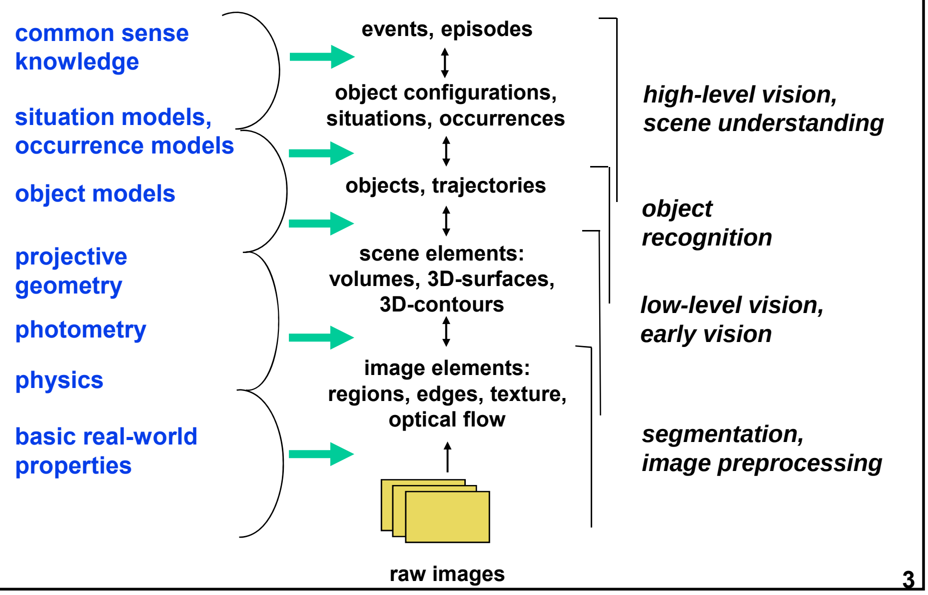
1

Illustration of Image Understanding



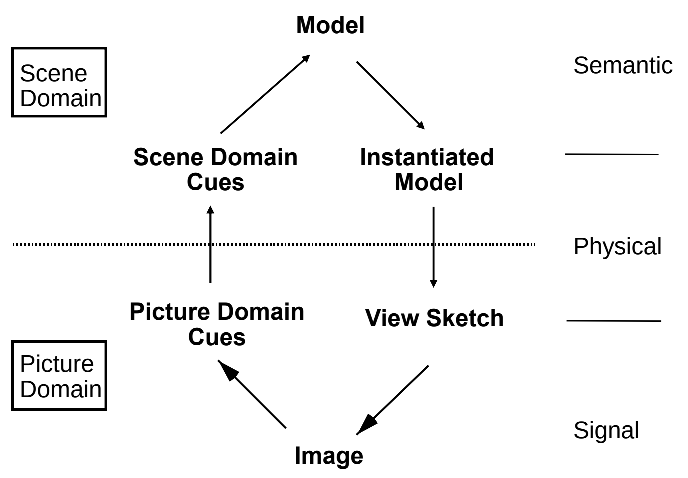
2

Image Understanding as a Knowledge-based Process



3

A Model of Scene Analysis (Kanade 78)



4

Abstraction Levels for the Description of Computer Vision Systems

Knowledge level

What knowledge or information enters a process? What knowledge or information is obtained by a process?

What are the laws and constraints which determine the behavior of a process?

Algorithmic level

How is the relevant information represented?

What algorithms are used to process the information?

Implementation level

What programming language is used?

What computer hardware is used?

5

Example for Knowledge-level Analysis

What knowledge or information enters a process? What knowledge or information is obtained by a process?

What are the laws and constraints which determine the behavior of a process?

Consider shape-from-shading:

In order to obtain the 3D shape of an object, it is necessary to

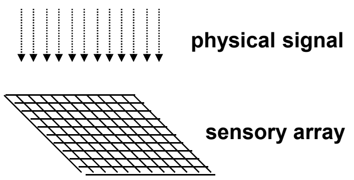
- **state what knowledge is available (greyvalues, surface properties, illumination direction, etc.)**
- **state what information is desired (e.g. qualitative vs. quantitative)**
- **exploit knowledge about the physics of image formation**

6

Image Formation

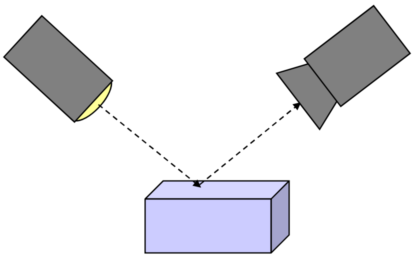
Images can be generated by various processes:

- illumination of surfaces, measurement of reflections ← "natural images"
- illumination of translucent material, measurement of irradiation
- measurement of heat (infrared) radiation
- X-ray of material, computation of density map
- measurement of any features by means of a sensory array



7

Formation of Natural Images



Intensity (brightness) of a pixel depends on

1. illumination (spectral energy, secondary illumination)
2. object surface properties (reflectivity)
3. sensor properties
4. geometry of light-source, object and sensor constellation (angles, distances)
5. transparency of irradiated medium (mistiness, dustiness)

8

Qualitative Surface Properties

When light hits a surface, it may be

- absorbed
- reflected
- scattered

} in general, all effects may be mixed

Simplifying assumptions:

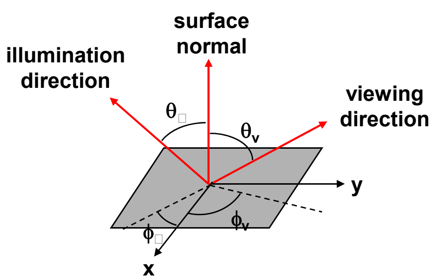
- Radiance leaving at a point is due to radiance arriving at this point
- All light leaving the surface at a wavelength is due to light arriving at the same wavelength
- Surface does not generate light internally

The "amount" of reflected light may depend on:

- the "amount" of incoming light
- the angles of the incoming light w.r.t. to the surface orientation
- the angles of the outgoing light w.r.t. to the surface orientation

9

Photometric Surface Properties



$\theta_\square$, θ_v polar (zenith) angles

$\phi_\square$, ϕ_v azimuth angles

In general, the ability of a surface to reflect light is given by the Bi-directional Reflectance Distribution Function (BRDF) r:

$$r(\theta_i, \phi_i; \theta_v, \phi_v) = \frac{\delta L(\theta_v, \phi_v)}{\delta E(\theta_i, \phi_i)}$$

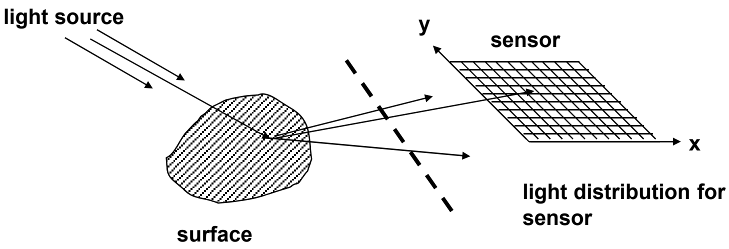
δE = irradiance of light source received by the surface patch

δL = radiance of surface patch towards viewer

For many materials the reflectance properties are rotation invariant, in this case the BRDF depends on $\theta_\square$, θ_v, ϕ, where□ϕ □ $\phi_\square$ - ϕ_v.

10

Intensity of Sensor Signals



Intensities of sensor signals depend on
- location x, y on sensor plane
- instance of time t
- frequency of incoming light wave λ
- spectral sensitivity of sensor

$$f(x,y,t) = \int_0^\infty C(x,y,t,\lambda) S(\lambda) d\lambda$$

$S(\lambda)$: sensitivity function of sensor

$C(x,y,t,\lambda)$: spectral energy distribution

11

Multispectral Images

Sensors with separate channels of different spectral sensitivities generate multispectral images:

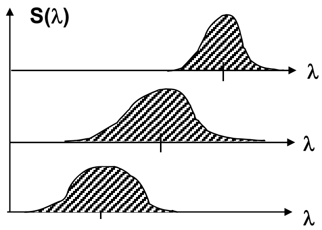
$$f_1(x,y,t) = \int_0^\infty C(x,y,t,\lambda) S_1(\lambda) d\lambda$$

$$f_2(x,y,t) = \int_0^\infty C(x,y,t,\lambda) S_2(\lambda) d\lambda$$

$$f_3(x,y,t) = \int_0^\infty C(x,y,t,\lambda) S_3(\lambda) d\lambda$$

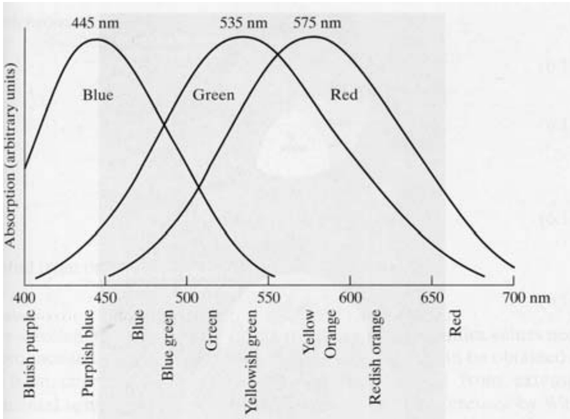
Example:

R (red)	650 nm center frequency
G (green)	530 nm center frequency
B (blue)	410 nm center frequency



12

Spectral Sensitivity of Human Eyes



Standardized wavelengths:
red = 700 nm, green = 546.1 nm, blue = 435.8 nm

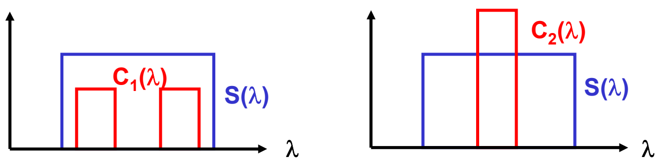
Non-unique Sensor Response

Different spectral distributions may lead to identical sensor responses and hence cannot be distinguished

$$f(x,y,t) = \int_0^\infty C_1(x,y,t,\lambda)S(\lambda)d\lambda = \int_0^\infty C_2(x,y,t,\lambda)S(\lambda)d\lambda$$

different spectral energy distributions

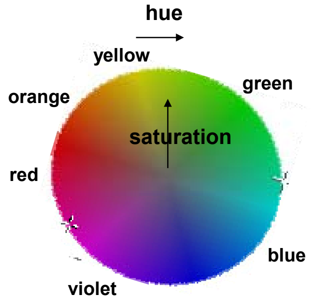
Example:



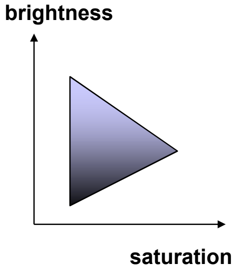
Dimensions of Colour

Human perception of colour distinguishes between 3 dimensions:
- hue
- saturation
- brightness

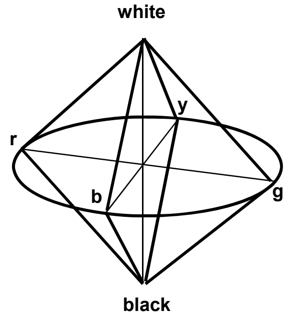
(hue and saturation } chromaticity)



colour circle



NCS* colour spindle



* Swedish Natural Colour System

15

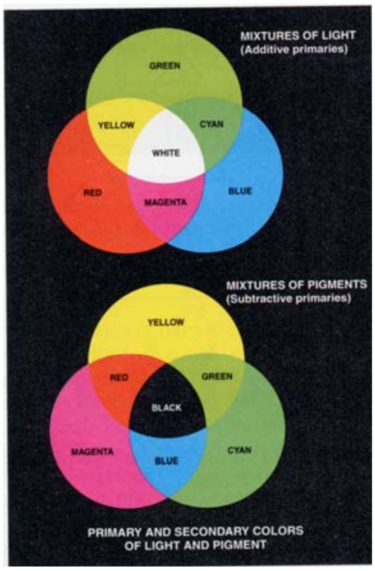
RGB Images of a Natural Scene

Here, single colour images are rendered as greyvalue intensity images:
stronger spectral intensity = more brightness

R+G+B | R | G | B

16

Primary and Secondary Colours



Primary colours:

red, green, blue

Secondary colours:

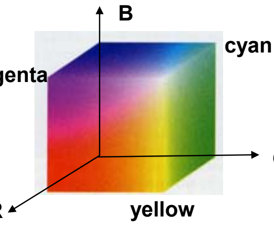
magenta = red + blue
cyan = green + blue
yellow = red + green

aus: Gonzales & Woods
Digital Image Processing
Prentice Hall 2002

17

Technical Colour Models

RGB colour model



Typical discretization:

8 bits per colour dimension
=> 16.777.216 colours

CMY colour model

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

HSI colour model

Hue:

$$H = \begin{cases} \Theta & \text{if } B \leq G \\ 360 - \Theta & \text{if } B > G \end{cases}$$

$$\Theta = \cos^{-1} \frac{1/2\,[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}}$$

Saturation:

$$S = 1 - \frac{3}{(R + G + B)}\,[\min(R, G, B)]$$

Intensity:

$$I = 1/3\,(R + G + B)$$

18

Discretization of Images

Image functions must be discretized for computer processing:

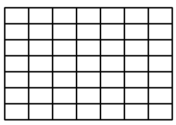
- **spatial quantization**
  the image plane is represented by a 2D array of picture cells
- **greyvalue quantization**
  each greyvalue is taken from a discrete value range
- **temporal quantization**
  greyvalues are taken at discrete time intervals

$$
\begin{aligned}
f(x,y,t) \Rightarrow \{\, & f_s(x_1, y_1,t_1), f_s(x_2, y_2,t_1), f_s(x_3, y_3,t_1), \ldots \\
& f_s(x_1, y_1,t_2), f_s(x_2, y_2,t_2), f_s(x_3, y_3,t_2), \ldots \\
& f_s(x_1, y_1,t_3), f_s(x_2, y_2,t_3), f_s(x_3, y_3,t_3), \ldots \,\}
\end{aligned}
$$

A single value of the discretized image function is called a pixel (picture element).

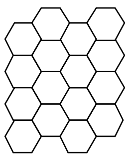
Spatial Quantization

Rectangular grid

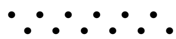


Greyvalues represent the quantized value of the signal power falling into a grid cell.

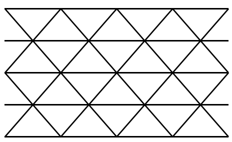
Hexagonal grid



Note that samples of a hexagonal grid are equally spaced along rows, with successive rows shifted by half a sampling interval.



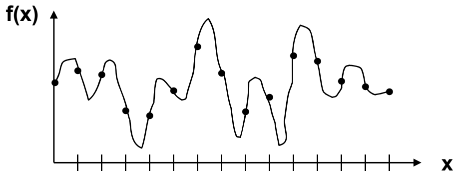
Triangular grid



Reconstruction from Samples

Under what conditions can the original (continuous) signal be reconstructed from its sampled version?

Consider a 1-dimensional function f(x):



Reconstruction is only possible, if "variability" of function is restricted.

21

Sampling Theorem

Shannon´s Sampling Theorem:

A bandlimited function with bandwidth W can be exactly reconstructed from equally spaced samples, if the sampling distance is not larger than $\frac{1}{2\,W}$

bandwidth = largest frequency contained in signal

(=> Fourier decomposition of a signal)

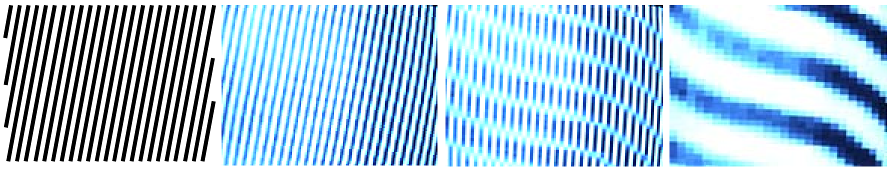
Analogous theorem holds for 2D signals with limited spatial frequencies W_x **and** W_y

22

Aliasing

Sampling an image with fewer samples than required by the sampling theorem may cause "aliasing" (artificial structures).

Example:



original | 143 x 128 | 71 x 64 | 35 x 32

To avoid aliasing, bandwidth of image must by reduced prior to sampling.

(=> low-pass filtering)

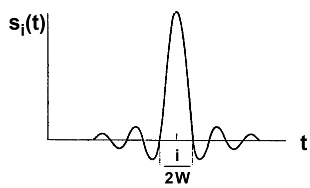
23

Reconstructing the Image Function from Samples

Formally, a continuous function f(t) with bandwidth W can be exactly reconstructed using <u>sampling functions</u> $s_i(t)$:

$$s_i(t) = \sqrt{2W}\,\frac{\sin 2\pi W[t - i/(2W)]}{2\pi W[t - i/(2W)]}$$

$$x(t) = \sum_{i=-\infty}^{\infty} \sqrt{\frac{1}{2W}} \times \underbrace{\left(\frac{i}{2W}\right)}_{\text{sample values}} S_i(t)$$



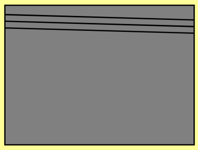
An analogous equation holds for 2D.

In practice, image functions are generated from samples by interpolation.

24

Sampling TV Signals

PAL standard:
- picture format 3 : 4
- 25 full frames (50 half frames) per second
- interlaced rows: 1, 3, 5, ... , 2, 4, 6, ...
- 625 rows per full frame, 576 visible
- 64 µs per row, 52 µs visible
- 5 MHz bandwidth



Only 1D sampling is required because of fixed row structure.

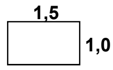
Sampling intervals of $\Delta t = 1/(2W) = 10^{-7}$s = 100 ns give maximal possible resolution.

With Δt = 100 ns, a row of duration 52 µs gives rise to 520 samples.

In practice, one often chooses <u>512 pixels per TV row</u>.

=> 576 x 512 = 294912 pixels per full frame

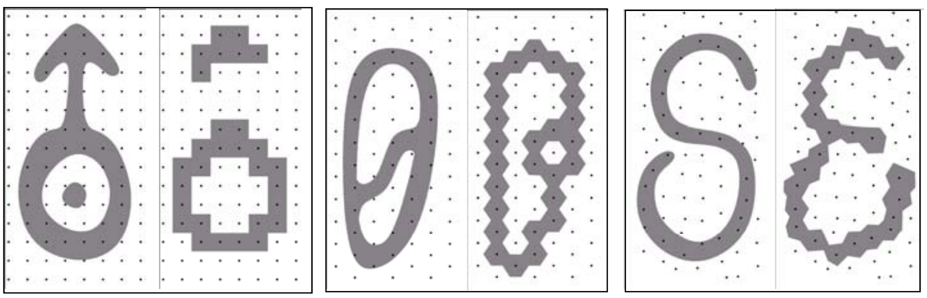
=> rectangular pixel size with width/height $=(\frac{4}{512}) / (\frac{3}{576}) = 1{,}5$



25

Sampling of Binary Images (1)

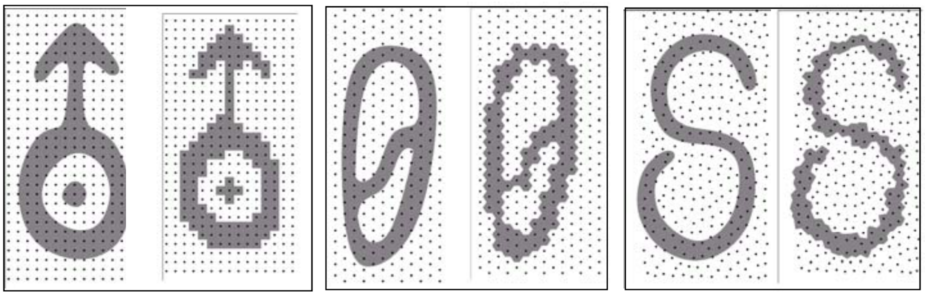
Problem: Shapes may change under digitization



26

Sampling of Binary Images (2)

Problem: Shapes may change under digitization



This cannot be solved by using Shannon´s Theorem since binary images are not bandlimited.

Shape Preserving Sampling Theorem (1)

Shape Preserving Sampling Theorem:
The shape of an r-regular image can be correctly reconstructed after sampling with any sampling grid, if the grid point distance is not larger than r.

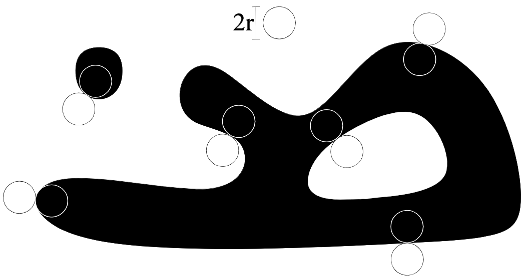
Stelldinger, Köthe 2003

"grid point distance": maximal distance from arbitrary sampling point to the next sampling point

"r-regular binary image":

osculating r-discs at each boundary point of the shape

⇒ curvature bounded by 1/r
⇒ bounded thinness of parts
⇒ minimal distance between parts



Shape Preserving Sampling Theorem (2)

What does correct reconstruction mean?

Topological and geometric similarity criterion:

One shape can be mapped onto the other by twisting the whole plane, such that the displacement of each point is smaller than r.

29

Sampling of Shapes in Arbitrary Images (1)

The previous sampling theorem also holds for greyvalue images, if each level set is an r-regular shape.

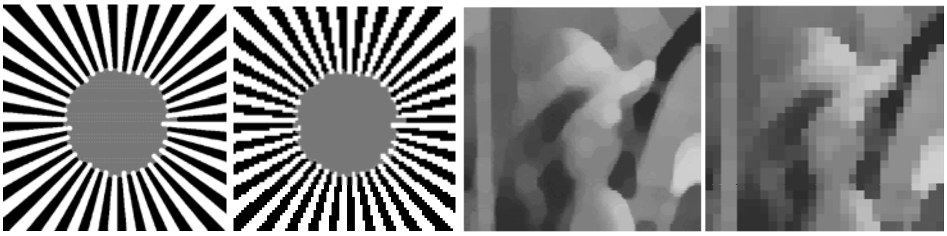
A "level set" is the set where the image is brighter than a given threshold value.

sampling + reconstruction

30

Sampling of Shapes in Arbitrary Images (2)

Reconstruction after sampling from r-regular originals



There exist generalizations to more complex cases (e.g. higher dimensions, blurring, colors, noise).

31

Comparison of the Sampling Theorems

	Shannon´s Sampling Theorem	Shape Preserving Sampling Theorem
necessary image property	bandlimited with bandwidth W	r-regular
equation	$\left(\frac{r'}{\sqrt{2}} =\right) d < \frac{1}{2W}$	$r' < r$
reconstructed image	identical to original image	same shape as the original image
prefiltering	band-limitation: efficient algorithms (but shapes may change!)	regularization: unsolved problem
2D sampling grid	rectangular grid	arbitrary grids
dimension of definition	1D (generalizable to n-D)	2D (partly generalizable to n-D)

32

Quantization of Greyvalues

Quantization of greyvalues transforms continuous values of a sampled image function into digital quantities.

Typically 2 ... 2^{10} quantization levels are used, depending on task.

Less than 2^9 quantization levels may cause artificial contours for human perception.

<u>Example</u>:

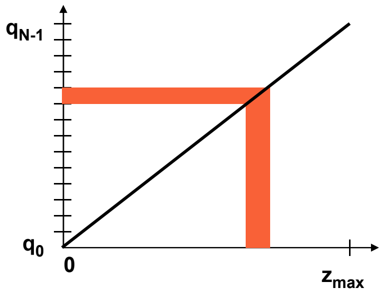
256 16 8 4 2

33

Uniform Quantization

Equally spaced discrete values $q_0 \dots q_{N-1}$ represent equal-width greyvalue intervals of the continuous signal.

Typically $N = 2^K$ for K = 1 ... 10



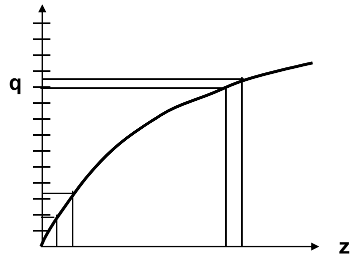
Uniform quantization may "waste" quantization levels, if greyvalues are not equally distributed.

34

Nonlinear Quantization Curves

E.g. fine resolution for "interesting" greyvalue ranges, coarse resolution for less interesting greyvalue ranges.

<u>Example</u>:
Low greyvalues are mapped into more quantization levels than high greyvalues.



<u>Note</u>:

Subjective brightness in human perception depends (roughly) logarithmically on the actual (measurable) brightness.

To let the computer see brightness as humans, use a logarithmic quantization curve.

35

Optimal Quantization (1)

<u>Assumption</u>:

Probability density p(z) for continuous greyvalues and number of quantization levels N are known.

<u>Goal</u>:

Minimize mean quadratic quantization error d_q by choosing optimal interval boundaries z_n and optimal discrete representatives q_n.

$$d_q^2 = \sum_{n=0}^{N-1} \int_{z_n}^{z_{n+1}} (z - q_n)^2 p(z) dz$$

Minimizing by setting the derivatives zero:

$$\frac{\delta}{\delta z_n} d_q^2 = (z_n - q_{n-1})^2 p(z_n) - (z_n - q_n)^2 p(z_n) = 0 \quad \text{for } n = 1 \ldots N-1$$

$$\frac{\delta}{\delta q_n} d_q^2 = -2 \int_{z_n}^{z_{n+1}} (z - q_n) p(z) dz = 0 \quad \text{for } n = 0 \ldots N-1$$

36

Optimal Quantization (2)

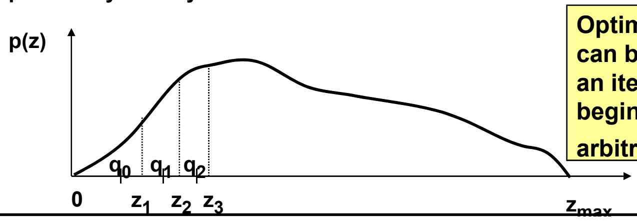
Solution for optimal quantization:

$$z_n = \frac{1}{2}(q_{n-1} + q_n) \quad \text{for } n = 1 \ldots N-1 \text{ when } p(z_n) > 0$$

Each interval boundary must be in the middle of the corresponding quantization levels.

$$q_n = \frac{\int_{z_n}^{z_{n+1}} z\,p(z)\,dz}{\int_{z_n}^{z_{n+1}} p(z)\,dz} \quad \text{for } n = 0 \ \ldots \ N-1$$

Each quantization level is the center-of-gravity coordinate of the corresponding probability density area.



Optimal quantization can be determined by an iterative algorithm beginning with an arbitrary choice of z_1

37

Binarization

For many applications it is convenient to distinguish only between 2 greyvalues, "black" and "white", or "1" and "0".

<u>Example</u>: Separate object from background

Binarization = transforming an image function into a binary image

<u>Thresholding</u>:

$$g(x, y) \Rightarrow \begin{cases} 0 & \text{if } g(x, y) < T \\ 1 & \text{if } g(x, y) \geq T \end{cases}$$

T is threshold

Thresholding is often applied to digital images in order to isolate parts of the image, e.g. edge areas.

38

Threshold Selection by Trial and Error

A threshold which perfectly isolates an image component must not always exist.

Selection by trial and error:

Select threshold until some image property is fulfilled, e.g.

$$q = \frac{\#\ \text{white pixels}}{\#\ \text{black pixels}} \Rightarrow q_0$$

line strength $\Rightarrow d_0$

number of connected components $\Rightarrow n_0$

Number of trials may be small if logarithmic search can be used.

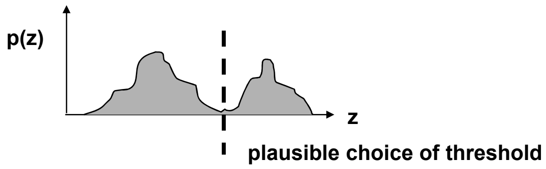
Example:

At most 8 trials are needed to select a threshold $0 \leq T \leq 255$ which best approximates a given q_0.

39

Distribution-based Threshold Selection

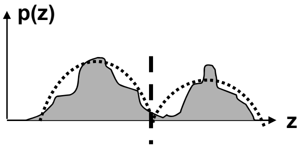
The greyvalue distribution of the image function may show a bimodality:



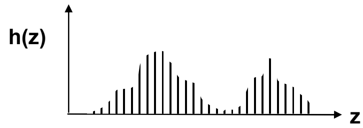
Two methods for finding a plausible threshold:

1. Find "valley" between two "hills"

2. Fit hill templates and compute intersection



Typically, computations are based on histograms which provide a discrete approximation of a distribution.

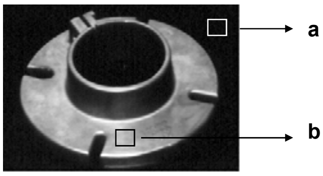


40

Threshold Selection Based on Reference Positions

In many applicatons, the approximate position of image components is known a priori. These positions may provide useful reference greyvalues.

<u>Example</u>:



possible choice of threshold:

$$T = \frac{a+b}{2}$$

Threshold selection and binarization may be decisively facilitated by a good choice of illumination and image capturing techniques.

41

Image Capturing for Thresholding

If the image capturing process can be controlled, thresholding can be facilitated by a suitable choice of

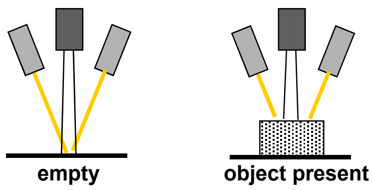
- illumination
- camera position
- object placement
- background greyvalue or colour
- preprocessing

<u>Example</u>: Backlighting

Illumination from the rear gives bright background and shadowed object

<u>Example</u>: Slit illumination

On a conveyor belt illuminated by a light slit at an angle, elevations give rise to displacements which can be recognized by a camera.



42