Universität Stuttgart

Pattern Recognition
Chapter 6: Features

Prof. Dr.-Ing. Uwe Sörgel
soergel@ifp.uni-stuttgart.de

ifp

Level model of model-based image analysis

top-down ↓ / bottom-up ↑	**Recognition/Evaluation** Features → Objects	high-level
	Segmentation Image → Features	mid-level
	Preprocessing Image → Image	low-level
	Image acquisition Scene → Image	

Feature extraction

- The term ”feature“ has two meanings:

1. **Geometric primitive:**
   - Point
   - Line
   - Segment

2. **Properties of a geometrical primitive, attribute**
   - Used for classification
   - Here: features as attributes of pixels or segments

Feature extraction

- Aim:
  - Reduction of redundant information
  - Qualitative and quantitative cues for image analysis

- Feature types:
  - Radiometric features: for pixels or segments
    - Densimetric features (probability density functions)
    - Texture Features
    - Structural Features
  - Geometric Features: only for segments

Content

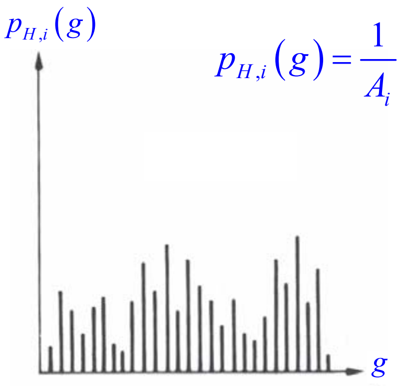
- Densimetric features
- Texture Features
- Structural Features
- Geometric Features
- Scaling of Features

Densimetric features for single pixels

- Grey values or color vectors
- Functions of the grey values, e.g.
  - Multispectral images: $NDVI = (IR - R) / (IR + R)$
  - Derivatives of the grey values
  - Color space transformations, e.g. intensity, hue, saturation.
- Features that are defined for segments can also be determined for individual pixels, using a square of side length s centred at the pixel.
- Multiscale feature vectors: determine features in square local neighborhoods of different side lengths s
  - Consider local image structure
  - s can be interpreted as scale parameter

Densimetric features: Approach

- Analysis of the grey value histograms.
- The histograms serve as approximation of underlying distribution.
- Based on the moment of the distribution we group or classify the segments.



$$p_{H,i}(g) = \frac{1}{A_i} \sum_{x,y \in Segment_i} \delta\left(g - g(x,y)\right)$$

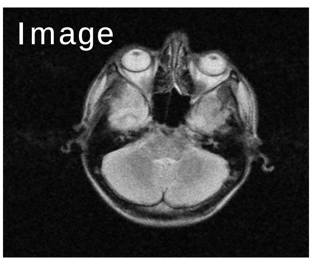
with A_i: Area of segment i

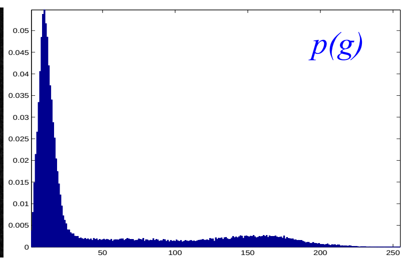
$\delta(z)$: Indicator function, i.e.

$$\delta(z) = \begin{cases} 1 \text{ for } z = 0 \\ 0 \text{ else} \end{cases}$$

Densimetric features: Approach

- The moments of the distributions can be easily derived from the histogram:
  - The mean represents the typical grey value for each channel $\quad \mu = M_1 = \sum_{g=0}^{255} g \cdot p(g)$
  - The variance is related to the radiometric homogeneity $\quad \sigma^2 = M_2 = \sum_{g=0}^{255} (g - \mu)^2 \cdot p(g)$





Densimetric features for images with k channels

- **Mean:** Mean vector $\boldsymbol{\mu}_i = \left(\mu_{i,1}, \mu_{i,2} \dots \mu_{i,k}\right)$
  - Components are the mean values of channel k

- **Covariance matrix:**

$$\mathrm{cov}_i = \begin{pmatrix} \sigma_{i,1,1} & \sigma_{i,1,2} & \cdots & \sigma_{i,1,k} \\ \sigma_{i,2,1} & \sigma_{i,2,2} & \cdots & \sigma_{i,2,k} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{i,k,1} & \sigma_{i,k,2} & \cdots & \sigma_{i,k,k} \end{pmatrix}$$

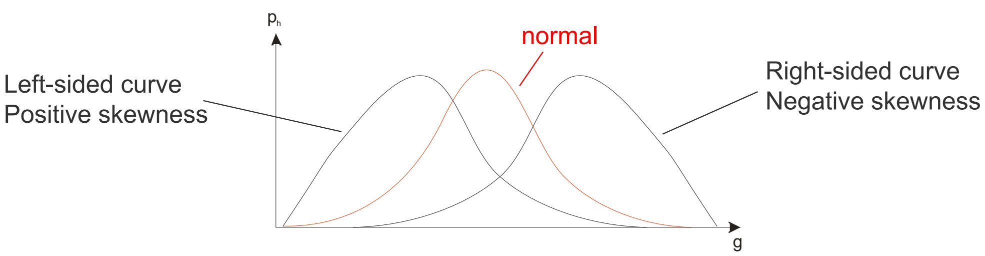
$$\sigma_{i,b_1,b_2} = \frac{1}{A_i - 1} \sum_{x,y \in Segment_i} \left(g_{b_1}(x,y) - \mu_{b_1}\right) \cdot \left(g_{b_2}(x,y) - \mu_{b_2}\right)$$

b_1, b_2: Indices of channels

Densimetric features: Skewness

- **Skewness:** $M_3 = \sum_{g=0}^{255} (g-\mu)^3 \cdot p(g)$ $\qquad M_3' = \frac{M_3}{\sqrt[3]{M_2}}$

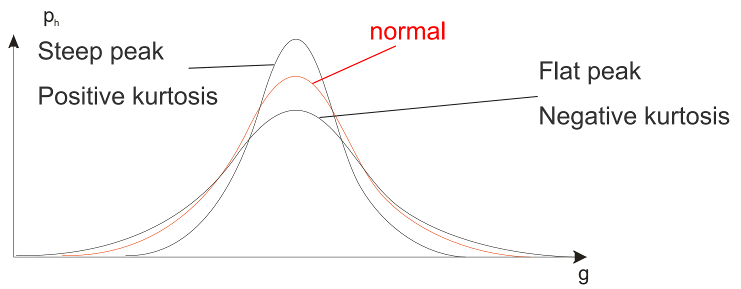
- The skewness is a measure of the degree of asymmetry of the distribution:
  - Deviation of histogram's shape from Gaussian normal form either to left or right side.



Densimetric features: Kurtosis

- Kurtosis: $M_4 = \sum_{g=0}^{255} (g-\mu)^4 \cdot p(g)$ $\qquad M_4' = \frac{M_4}{M_2^2} - 3$

- Deviation of the peak level of the histogram from the shape of a Gaussian.



Further densimetric features

- Range: $s_i = g_{max} - g_{min}$

- Most frequent grey value (mode)

- Energy: $E_i = \sum_{g=0}^{255} \left(A_i \cdot p_i(g)\right)^2$

- Variation coefficient: $VAR_i = \frac{\sigma}{\mu}$

Densimetric features: Entropy

- Entropy $$H = -\sum_{g=0}^{255} p(g) \cdot \log_2 p(g)$$
- Measure of
  - the mean information content of a message (here: a segment).
  - Average prior uncertainty per pixel.
  - Estimate for the mean number of bits per image point required for encoding the grey values of the image.
- Entropy for
  - Homogenous image: $H =$ 0
  - Binary image with equal probability of occurrence: $H =$ 1
  - 8bit image with equal probability of occurrence:

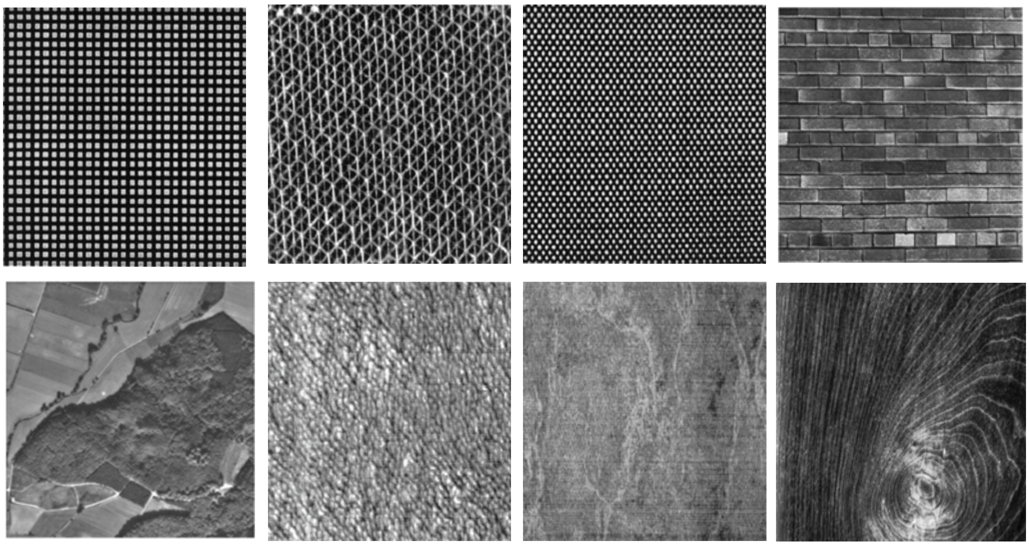
$$p(g_x) = \frac{1}{256} \Rightarrow H = -256 \cdot \frac{1}{256} \cdot \underbrace{\log_2 \frac{1}{256}}_{-8} = 8$$

Content

- Densimetric features
- Texture Features
- Structural Features
- Geometric Features
- Scaling of Features

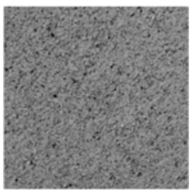
Texture: Repetitive pattern in images

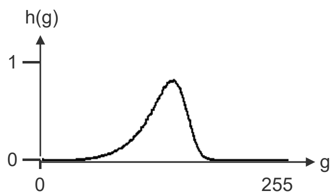
- Texture: characteristic structural pattern in an image
- Texture is a property of a certain area, not of a single pixel.



Characterization of textures: Moments of distribution

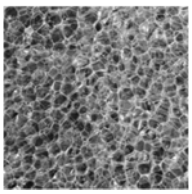
- Textures can to some extent be characterized by histograms:
  - Central moments: mean, variance, skewness,…

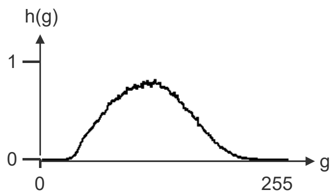




Mean: 123.66
Variance: 606.94
Skewness: -14834.70 << 0

source, FH Karlsruhe





Mean: 111.28
Variance: 1394.38
Skewness: 8561.81 >> 0

- Disadvantage:
  - No information about image structure (neighborhood relations)
  - Different textures may share the same histogram.

Locale texture analysis

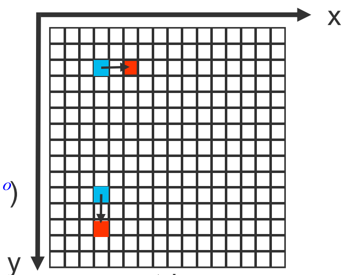
- The texture of a surface is related to the distribution and variation of grey levels in local areas (primitives) → texture element, Texel
  - Texture depends on scale, illumination, viewing direction
- Structural texture models
  Suitable for man-made objects, which feature high regularity.
- Statistical texture models
  Suitable for natural surfaces. The texture element are described according locale statistical features.
- Frequency-based texture models
  Superposition of different harmonic functions (e.g. sine curves)
- Texture analysis is carried out for segments or for local windows
  → window size has to be adapted to size of the texels

Texture analysis: Co-occurrence matrices

- Co-occurrence matrix: Frequency of occurrence of grey value pairs in a given offset.
- This offset is specified by a position operator φ(distance d, angle α).

Example 1: $\varphi(2, 0^o)$
→ $\Delta x = 2, \Delta y = 0$

Example 2: $\varphi(2, 90^o)$
→ $\Delta x = 0, \Delta y = 2$



For N grey values, the dimension of the co-occurrance matrix C ($c_{ij} = p[i, j| d, \alpha]$) is $N \times N$ with:

$$p(i, j| d, \alpha) = \frac{\text{Number of instances for which } g(x,y) = i \text{ and } g(x+\Delta x, y+\Delta y) = j}{\text{Total number of such comparisons}}$$

Co-Occurrence Matrix: Examples

Gray value set $g=\{0,1,2,3\}$

Image:

$$\begin{matrix} 0 & 0 & 1 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 & 3 & 3 \\ 0 & 1 & 1 & 2 & 3 & 3 \\ 1 & 2 & 2 & 3 & 3 & 3 \\ 2 & 2 & 3 & 3 & 3 & 3 \end{matrix}$$

Co-occurrence matrix
$d=1,\ \alpha = 90°$ (lower neighbor)

0,1

0,0

1,0

$$\begin{matrix} 5 & 3 & 0 & 0 \\ 1 & 2 & 4 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 9 \end{matrix}$$

- The final values of $p(i,j \mid 1, 90°)$ result from so-occurrence matrix after normalization with total number of comparisons (here: divide by 30).

Co-Occurrance Matrix: Examples

- Same example using different angles, $d = 1$:

$$\begin{matrix} 0 & 0 & 1 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 & 3 & 3 \\ 0 & 1 & 1 & 2 & 3 & 3 \\ 1 & 2 & 2 & 3 & 3 & 3 \\ 2 & 2 & 3 & 3 & 3 & 3 \end{matrix}$$

Image with 2 bit
($g \in \{0,1,2,3\}$)

$a = 0°$ i j

$$\begin{matrix} 4 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 7 \end{matrix}$$ j, i

$p(i,j \mid 1, 0°)$

$a = 45°$ i j

$$\begin{matrix} 3 & 3 & 2 & 0 \\ 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 4 \end{matrix}$$ j, i

$p(i,j \mid 1, 45°)$

$a = 90°$ i j

$$\begin{matrix} 5 & 3 & 0 & 0 \\ 1 & 2 & 4 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 9 \end{matrix}$$ j, i

$p(i,j \mid 1, 90°)$

$a = 135°$ i j

$$\begin{matrix} 4 & 0 & 0 & 0 \\ 2 & 3 & 1 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & 7 \end{matrix}$$ j, i

$p(i,j \mid 1, 135°)$

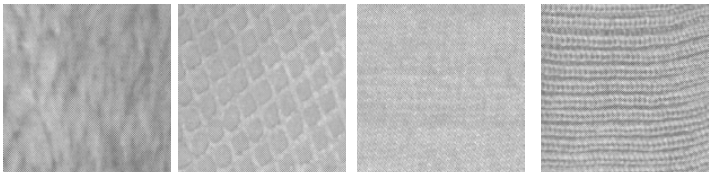
Distance and angle

- To obtain expressive texture features, it is important to use different distances d and angles α.
- The optimum set of distance / angle depends on the texture at hand
  - Distance d
    For good results, the maximum distance *d* has to correspond to the size of the texture primitives (the size of characteristic structures → *multi-scale analysis*
  - Angle α
    The co-occurrance matrix is mainly sensitive to textures whose edge directions are orthogonal to the angle α.
  - Coarse texture
    If the image contains large homogeneous areas, large values will mainly occur at the main diagonal.
  - Fine texture
    Fine texture with many steep edges will lead to high matrix entries in the lower left and upper right corners.

Haralick features derived from co-occurrence matrix

- Transition to meaningful features describing the distribution:
  - Energy distribution $$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j|d,\alpha)^2$$
  - Entropy $$-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j|d,\alpha)\cdot ld\left(p(i,j|d,\alpha)\right)$$
  - Contrast $$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} (i-j)^2\, p(i,j|d,\alpha)$$
  - Homogeneity $$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{p(i,j|d,\alpha)}{1+(i-j)^2}$$
- A Haralick feature is only valid for one matrix with given d and α!

Haralick features: Example

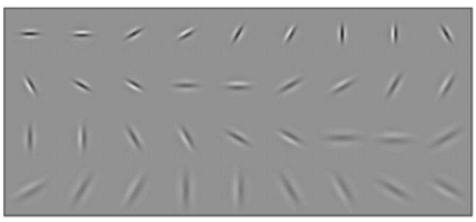


	Fur	Diamond	Silk	Rope
Energy	280,000	850,000	280,000	130,000
Contrast	1,100,000	410,000	7,000,000	2,200,000
Entropy	42,000	56,000	42,000	30,000
Homogeneity	3,200	5,300	1,400	2,800

Texture analysis: Filter banks (Textons)

General Principle:

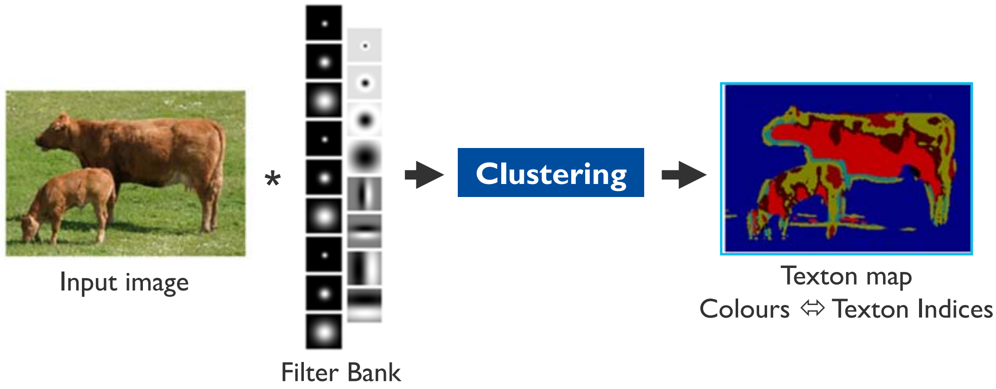
- Convolve an image with a set of different filters which correspond to different texture features
  - The filter set defines the filter bank, for example,"textons" shown here:



- For a pixel, each filter response is a texture feature
- For segments, the filter responses of all pixels inside the segment have to be combined → determine the energy for each filter response
- The resultant feature vector can be used for texture classification

Textons: Example

- Convolution with 17D filter bank (Gaussians, Difference of Gaussians)
- Clustering with K-means



Shotton et al.: TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. International Journal of Computer Vision, 2009

Further examples

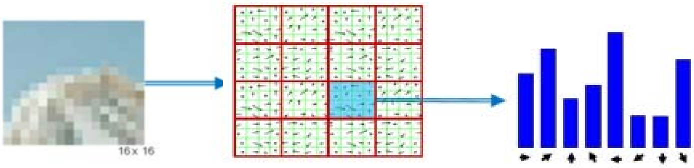
J. Malik, S. Belongie, T. Leung and J. Shi. "Contour and Texture Analysis for Image Segmentation". IJCV 43(1),7-27,2001.

Content

- Densimetric features
- Texture Features
- Structural Features
- Geometric Features
- Scaling of Features

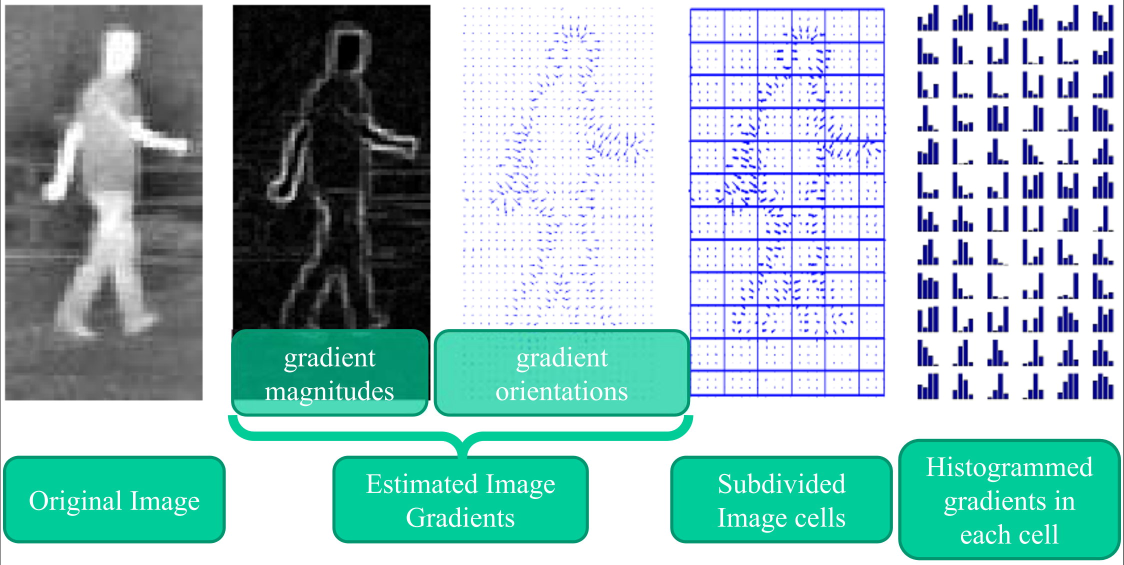
HOG features

- The histogram itself can serve as a feature vector
- Classical HOG features [Dalal & Triggs, 2005]:
  - Compute histograms of orientation angles in cells of size $n \times n$.
  - Contrast normalization: The histograms are normalized using overlapping blocks consisting of $m \times m$ neighboring cells.



  - Goal of Dalal & Triggs HOG: Detection of objects of a given size
    - → Concatenate histograms of all blocks inside a window 
    - → Explore scale space to detect objects of different size
    - → Supervised classification based on HOG features

HOG features: Example



Histograms of Oriented Gradients (HOG) features

[Dalal & Triggs, '06]

Content

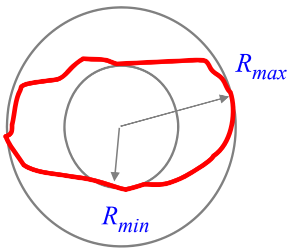
- Densimetric features
- Texture Features
- Structural Features
- Geometric Features
- Scaling of Features

Geometric features

- Geometric features to describe shape of objects:
  - Area: $A = \sum_{x=1}^{nx} \sum_{y=1}^{ny} b(x,y)$ for $b(x,y) = \{0,1\}$
  - Perimeter: $P = \sum_{x=1}^{nx} \sum_{y=1}^{ny} b_U(x,y)$ $b_U = G \setminus (G \ominus S_b)$

- Form factors, which rate roundness of object's shape:

$$FF_1 = \frac{P^2}{4\pi A} \qquad FF_2 = \frac{R_{\max}}{R_{\min}}$$



  - R_{max}: Radius of outer circle around center of gravity
  - R_{min}: Radius of inner circle

- Both form factors FF_1 and FF_2 score circular objects with value 1. Deviation of round shape result in values >1

 $FF_1 = 1$  $FF_1 = 1{,}27$

Geometric features of segments: Moments

Central moments

$$\mu_{pq} = \sum_{x=1}^{nx} \sum_{y=1}^{ny} b(x,y)(x - x_S)^p (y - y_S)^q$$

Determination of the mass center (center of gravity):

$$x_S = \frac{1}{I} \sum_{i=1}^{I} x_i = \frac{\mu_{10}}{\mu_{00}}$$

$$y_s = \frac{1}{I} \sum_{i=1}^{I} y_i = \frac{\mu_{01}}{\mu_{00}}$$

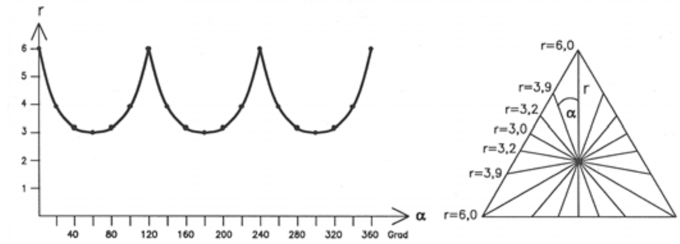
$\{x_i, y_i\}$: Coordinates of object pixel

I : Number of object pixels

2. moment: mass centroid axis of an ellipse around mass center.

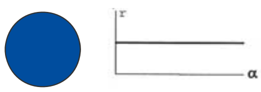
Geometric features of segments: Polar distance

Polar distance:

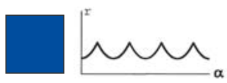


$$P_i = \frac{r_{i,max}}{r_{i,min}}$$ with $r_{min,max}$: Minimal, maximal distance to mass center.

i: Segment number



$P_i = 1{,}00$

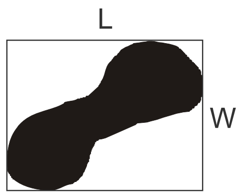


$P_i = 1{,}41$

Geometric features: Bounding rectangles I

Minimum bounding rectangle parallel to coordinate axes:

$$L_i = x_{i,max} - x_{i,min}$$
$$W_i = y_{i,max} - y_{i,min}$$



- Fill factor:

$$Fg_i = \frac{A_i}{L_i \cdot W_i}$$

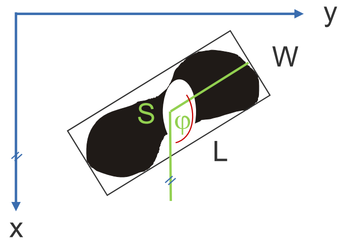
with

L, W:	Length, width
A:	Area
x,y:	Points of segment
i:	Segment number

Geometric features: Bounding rectangles II

Minimum bounding rectangle (MBR):

Orientation: $$\varphi_i = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{i,11}}{\mu_{i,20}-\mu_{i,02}}\right)$$



Dimension:

$$L_i = a_{i,max} - a_{i,min}$$
$$W_i = b_{i,max} - b_{i,min}$$

$$a_i(x_i,y_i) = x_i \cdot cos(\varphi_i) + y_i \cdot sin(\varphi_i)$$
$$b_i(x_i,y_i) = -x_i \cdot sin(\varphi_i) + y_i \cdot cos(\varphi_i)$$

with

x_i,y_i: Contour points

μ_{pq}: Central moments with order pq

i: Segment number

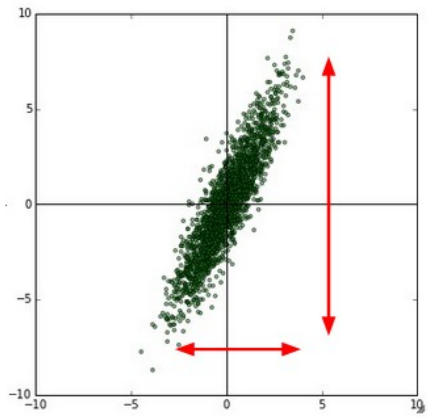
Content

- Densimetric features
- Texture Features
- Structural Features
- Geometric Features
- Scaling of Features

Scaling of features

- Frequently, the features will have different units and, thus different numerical values, e.g.
  - Age, annual income, number of children of an adult
- Features with large numerical values dominate the distance between feature vectors

➔ scaling required



Scaling of features

- Typically, the features are shifted and scaled.
  - Examples:
    - Shift and scale to the interval [0…1]

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

- Problem? Outlier

- An alternative (normalization):
  - Determine mean μ and standard deviation σ of all features from training data
  - Shift by μ, scale by $1 / \sigma$

  → Features are mapped into the same range

$$x' = \frac{x - \mu}{\sigma}$$

- Important: Transformation determined in training must be applied at test time!

Example for scaling by normalization

- Determine mean μ and standard deviation σ of all features from training data
- Shift by μ, scale by $1 / \sigma$

→ Features are mapped into the same range

$$x' = \frac{x - \mu}{\sigma}$$