



**PROFESSUR
MEDIENINFORMATIK**

Content Based Image Retrieval (CBIR)
Lecture Media Retrieval
Maximilian Eibl, Medieninformatik, TU Chemnitz

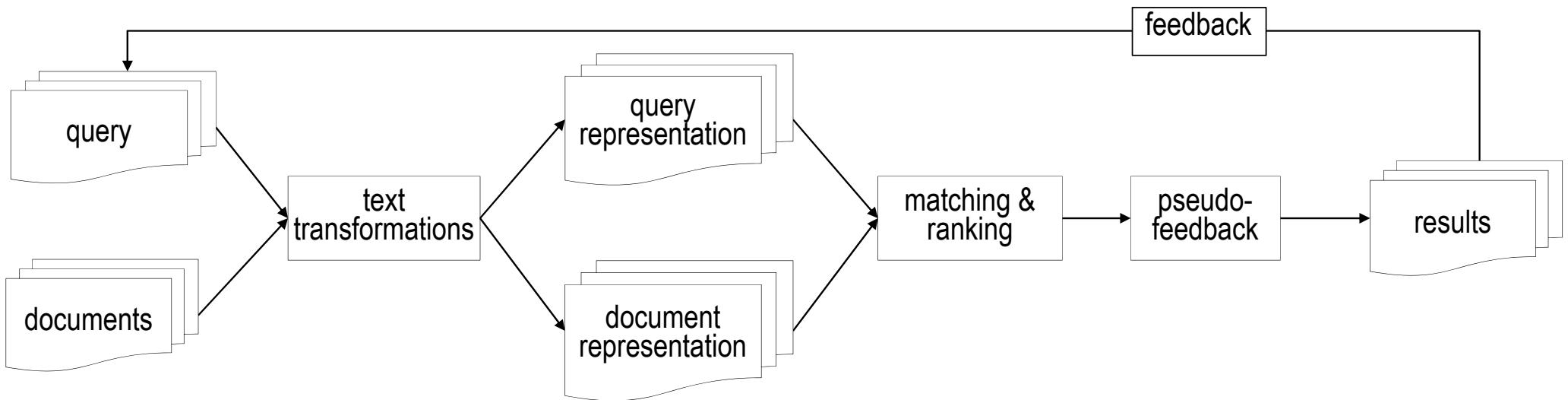


Image and Text: Different Modalities

Text	Image
Linear	2 dimensional
Syntactic structure	No implicit structure
Syntactic hierarchy	No implicit hierarchy
Semantic abstraction	Photorealism

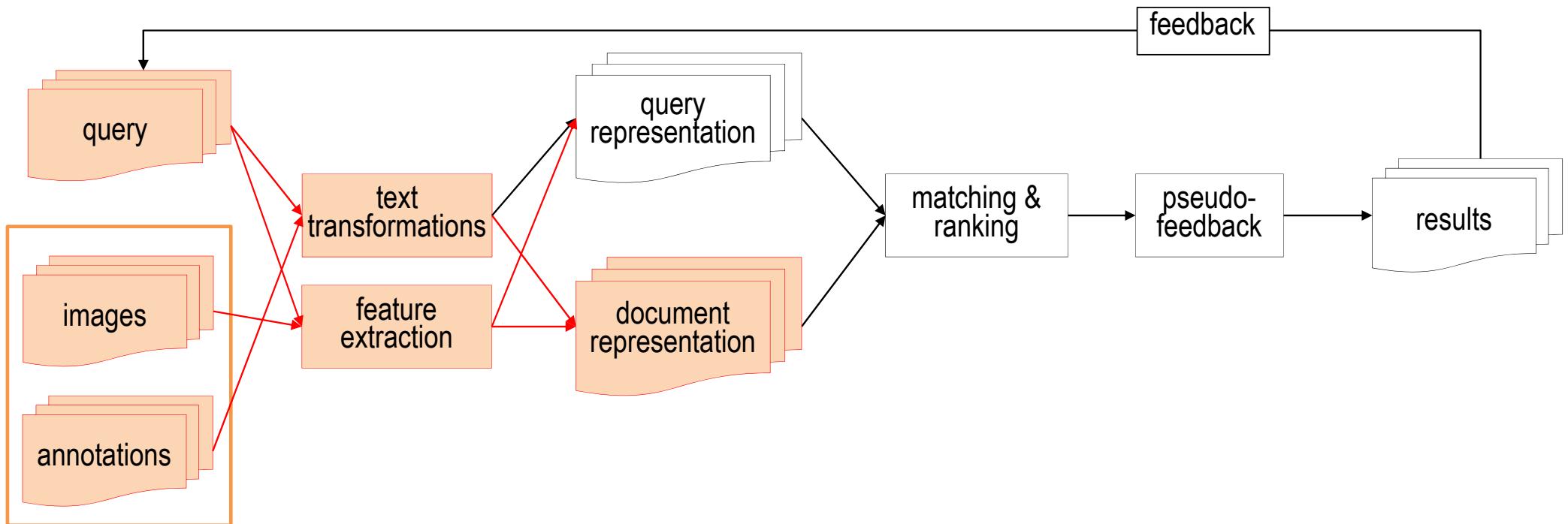


IRS Components: Text Retrieval



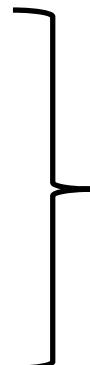


IRS Components: Image Retrieval





Types of Queries in Image Retrieval

- Query-by-text
 - Query-by-example
 - Query-by-sketch
 - Query-by-color
 - Query-by-texture
- 
- A large black brace is positioned to the right of the list, spanning from the second item to the fifth item. It has a vertical line on the left that connects to the middle of the list, and a horizontal line on the right that extends to the end of the list, grouping the five types of queries together.
- Content Based Image Retrieval CBIR



Types of Queries in Image Retrieval

- Query-by-text: “beach”
- Query-by-example
- Query-by-sketch
- Query-by-color
- Query-by-texture





Types of Queries in Image Retrieval

- Query-by-text
- Query-by-example:
- Query-by-sketch:
- Query-by-color
- Query-by-texture





Feature Types

Level 1: Primitive features

Level 2: Logical features

Level 3: Abstract features



„Wembley goal“: World Cup final, July 30th 1966, England – BRD, minute 101, 3:2, end result: 4:2, by: Geoff Hurst; Goal Keeper: Hans Tilkowski; the swiss referee Gottfried Dienst decides after discussion with sowjet line referee Tofiq Bəhramov



Semantic Gap

- Annotation
- Gamification



ImageNet

- Basis: WordNet (<http://wordnet.princeton.edu/>)
- Homepage: <http://www.image-net.org/>

WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **Persian cat** (a long-haired breed of cat)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
- S: (n) [domestic cat](#), [house cat](#), [Felis domesticus](#), [Felis catus](#) (any domesticated member of the genus Felis)
 - S: (n) [cat](#), [true cat](#) (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
 - S: (n) [feline](#), [felid](#) (any of various lithe-bodied roundheaded fissiped mammals, many with retractile claws)
 - S: (n) [carnivore](#) (a terrestrial or aquatic flesh-eating mammal) "terrestrial carnivores have four or five clawed digits on each limb"
 - S: (n) [placental](#), [placental mammal](#), [eutherian](#), [eutherian mammal](#) (mammals having a placenta; all mammals except monotremes and marsupials)
 - S: (n) [mammal](#), [mammalian](#) (any warm-blooded vertebrate having the skin

IMAGENET 14,197,122 images, 21841 synsets indexed

Not logged in. [Login](#) | [Signup](#)

Persian cat
A long-haired breed of cat

1662 pictures 59.56% Popularity Percentile Wordnet IDs

Treemap Visualization Images of the Synset Downloads

The Treemap Visualization shows the hierarchical structure of Persian cat synsets. The root node is 'ImageNet 2011 Fall Release (32326)', which branches into categories like 'plant, flora, plant life (4486)', 'geological formation, formation (175)', 'natural object (1112)', 'sport, athletics (176)', 'artifact, artefact (10504)', 'fungus (308)', 'person, individual, someone, somebody, mortal, soul (69)', 'animal, animate being, beast, brute, creature, fauna (399)', 'invertebrate (766)', 'homotherm, homoiotherm, homotherm (0)', 'work animal (4)', 'draft animal (0)', 'beast of burden, jument (2)', 'darer (0)', 'survivor (0)', 'range animal (0)', 'creepy-crawly (0)', 'domestic animal, domesticated animal (213)', 'domestic cat, house cat, Felis domesticus, Felis catus (1662)', 'Egyptian cat (0)', and 'Persian cat (0)'. The 'Images of the Synset' tab displays a grid of 1662 images of Persian cats, with one image highlighted in blue.



MIT Label Me

LabelMe

IMG XML ALL ? ⟲ You are: eibl [change user](#)

Polygons in this image (12)

[Hide all polygons](#)

Drag a tag on top of another one to create a part-of relationship.

- [sky](#)
- [building occluded](#)
- [building occluded](#)
- [plants](#)
- [tree](#)
- [balustrade](#)
- [window occluded](#)
- [ground grass](#)
- [streetlight](#)
- [tree](#)
- [eaves](#)
- [railing](#)

The interface shows a photograph of a two-story house with a prominent tree in the foreground. Overlaid on the image are several polygons of different colors: yellow, green, and pink. These polygons are used to segment the image into different semantic regions. The yellow polygons cover the sky, the roof and upper walls of the house, and the trunk and branches of the tree. The green polygons are located on the lower part of the tree trunk and some foliage. The pink polygon is at the base of the tree. On the left side of the interface, there is a vertical toolbar with icons for different tools: a red crosshair icon (labeled 'Tool'), a red circle with a minus sign (labeled 'Mask Tool'), and a red eraser icon. The 'Tool' icon is currently selected. At the bottom of the toolbar, there are three small gray squares.



MIT Label Me - data



```
<?xml version="1.0"?>
- <annotation>
  <filename>123_2334.jpg</filename>
  <folder>static_outdoor_bay_area_submitted_alyosha_efros</folder>
  - <source>
    <sourceImage>The MIT-CSAIL database of objects and scenes</sourceImage>
    <sourceAnnotation>LabelMe Webtool</sourceAnnotation>
  </source>
+ <object>
- <object>
  <name>streetlight</name>
  <deleted>0</deleted>
  <verified>0</verified>
  <date>07-May-2007 16:53:13</date>
  <id>8</id>
- <polygon>
  <username>anonymous</username>
  - <pt>
    <x>196</x>
    <y>1956</y>
  </pt>
  - <pt>
    <x>227</x>
    <y>838</y>
  </pt>
```



Semantic Gap: Image Labeling – ESP Game

- a image
- b entered descriptions
- c taboo words
- d match
- e pass

https://www.interaction-design.org/images/encyclopedia/social_computing/fig3_social_computing_research_social_media_plawindow_illustration.jpg





Google Image Labeler

GoogleTM
Image Labeler BETA

Google Image Labeler

time left **label** **pass**

00:03 Your partner has suggested 3 labels.

score **400**

passes **1**

off-limits tie

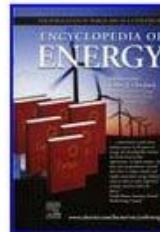
my labels
conference
speech



Images labeled so far:



passed
704 x 576 pixels
atlaseye.web.cern.ch



matched: energy
954 x 1390 pixels
www.social.mtu.edu



matched: sea
2928 x 1954 pixels
www.epa.gov



matched: airplane
432 x 177 pixels
www.airandspacemagazine.com



Models of CBIR

- Color-Based Modeling
- Texture-Based Modeling
- Shape-Based Modeling
- Spatial-Based Modeling

→ Level 1&2

→ Level 3: Semantic Gap



Color Space

- Different existing color spaces:
 - Physiological: CIE-RGB, CIE-XYZ
 - Psychological: CIE L*a*b*, CIE L*u*v*
 - Hardware oriented: RGB, CMY(K), YIQ
 - User oriented: HLS, HCV, HSV, HSB, MTM
- For Image Retrieval we need perceptual uniformity (*einheitliche Wahrnehmung*).
→CIE L*a*b*, CIE L*u*v* are hardware independent and well suited



Color Moments

- First order: mean (*Mittelwert*)

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij}$$

- Second order: variance (*Varianz*)

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{\frac{1}{2}}$$

- Third order: skewness (*Verzerrung*)

$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3 \right)^{\frac{1}{3}}$$

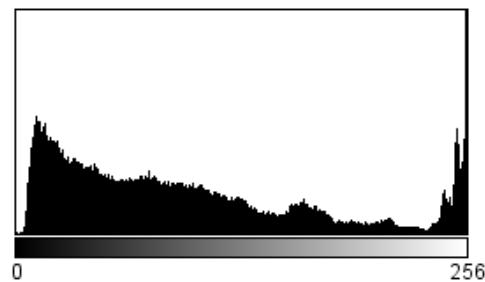
- where:

- f_{ij} : value of the i /color component in pixel j
 - N : number of pixels in image

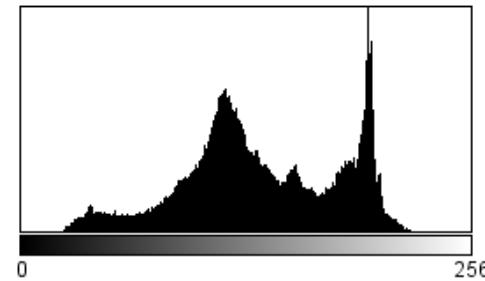


Color Histogram

- Effective representation of colors of an image when color distribution is unique
- Robust concerning translation and rotation
- Useful for global and local color distribution



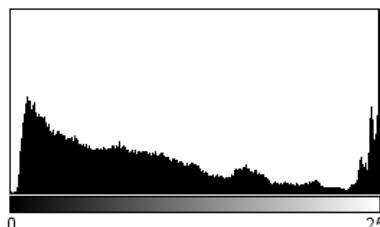
Count: 196608 Min: 0
Mean: 97.403 Max: 255
StdDev: 74.819 Mode: 255 (4182)



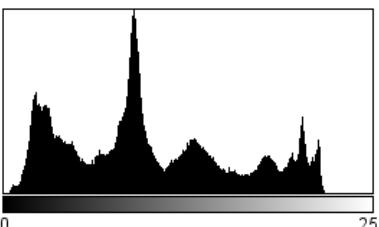
Count: 196608 Min: 3
Mean: 135.993 Max: 252
StdDev: 44.652 Mode: 197 (4179)



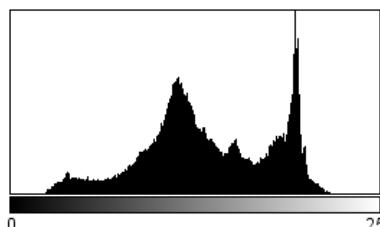
Color Histogram



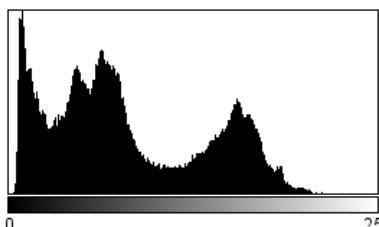
Count: 196608
Mean: 97.403
StdDev: 74.819



Count: 196608
Mean: 100.894
StdDev: 58.918



Count: 196608
Mean: 135.993
StdDev: 44.652



Count: 196608
Mean: 81.490
StdDev: 54.534



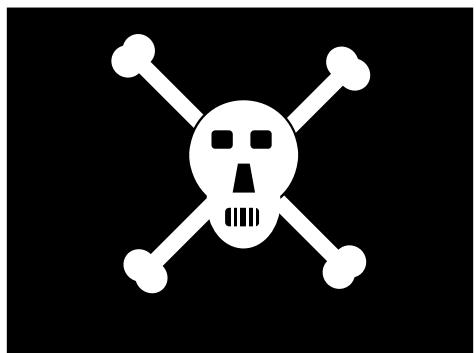


Color Histogram: Process

- Select color space
- Quantisize the three components of the color space to „bins“ (*Farbeimer*)
- Count absolute color distribution → fill the bins
- Calculate relative color distribution

Color Histogram: Challenges

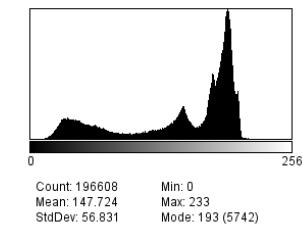
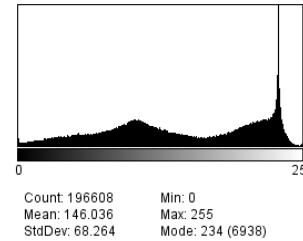
- Huge amount of images in a data base
 - Simple comparison of histograms is ineffective
- Spatial distribution of pixels is not considered
 - Different solutions for spatial distribution





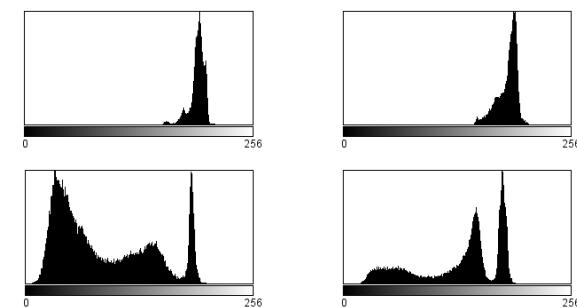
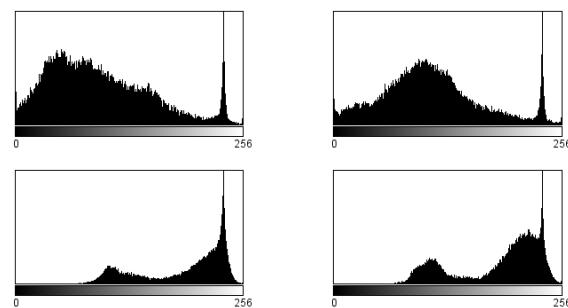
Color Histogram: Spatial Distribution

- Grid
- Annular Histogram
- Angular Histogram
- Color Coherence Vector (CVV)
- Color Correlogram
- Segmenting
- ...





Color Histogram: Grid





Color Coherence Vector (CVV)

- Devide the histograms in coherent and incoherent pixels:

CVV: $\langle(a_1, \beta_1), (a_1, \beta_1), \dots, (a_N, \beta_N)\rangle$

where

a_i : number of coherent pixels of color i of an image

β_i : number of incoherent pixels of color i of an image

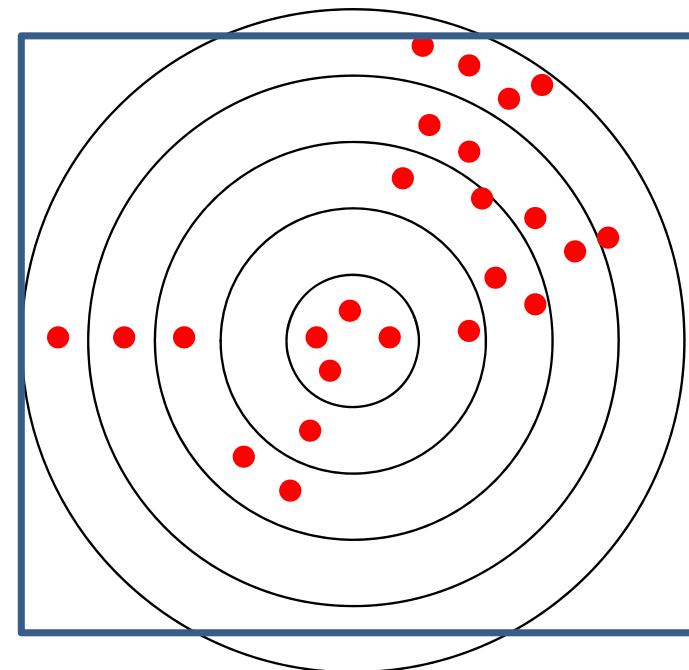
(simple histogram: $\langle(a_1+\beta_1), (a_1+\beta_1), \dots, (a_N+\beta_N)\rangle$)

- Especially for images with large cohrent color regions.
- HSV performs better than CIE L*u*v* oder CIE L*a*b*

Color Histogramm: annular

1. Divide the image in circular regions
 2. Count colors from the center to the outer regions
- annular distribution density vector :

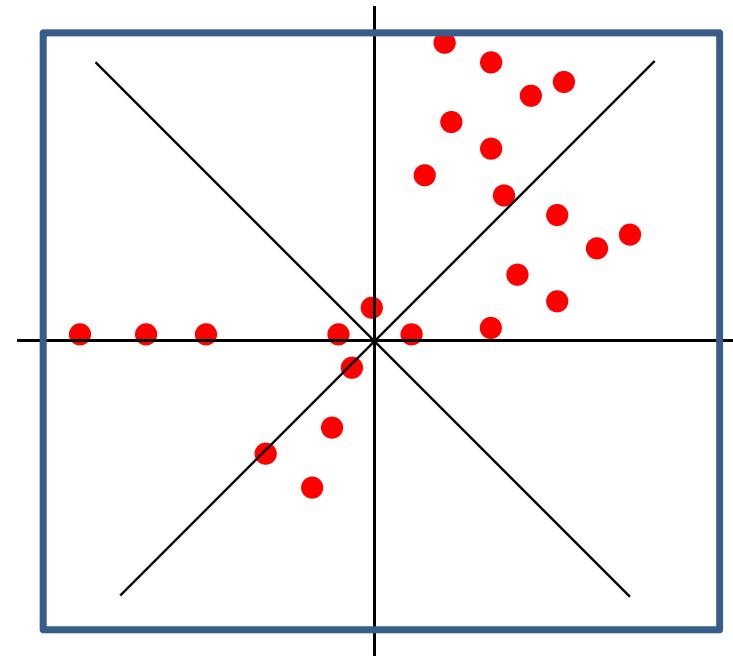
$\langle 4, 2, 7, 5, 6 \rangle$



siehe: Rao et al. 1999

Color Histogramm: angular

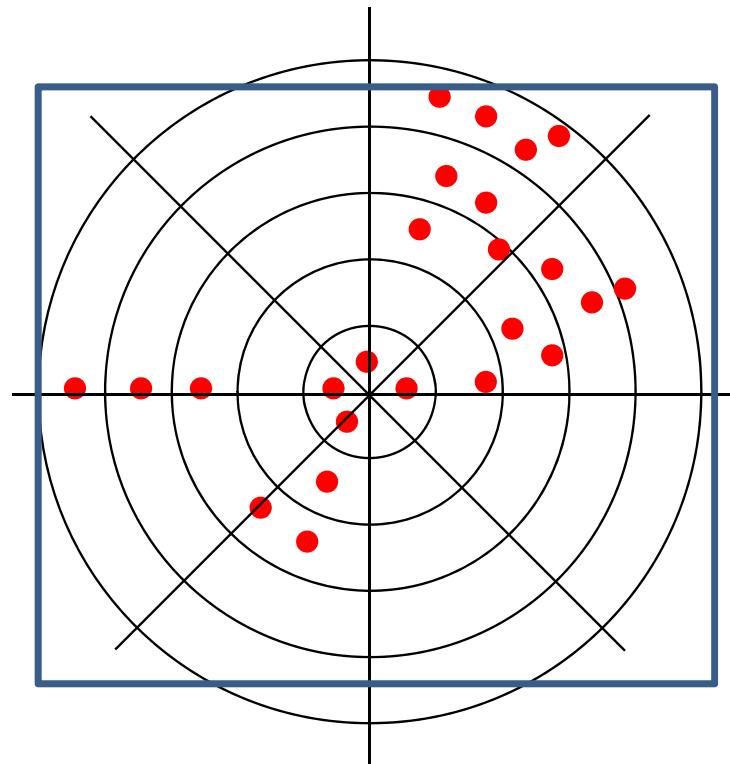
1. Divide the image in angles
2. Count the colors counter clockwise
3. → *angular distribution density vector:*
 $<7,8,1,4,0,4,0,0>$



siehe: Rao et al. 1999

Color Histogramm: hybrid

1. Divide the image in angles and circles
2. Count the colors counter clockwise from the center to the outer regions
3. → *hybrid distribution density vector:*
 $<1,0,1,1,0,1,0,0,1,0,0,0,0,1,0,0,2,2,0,1,0,2,0,2,2,0,1,0,0,0,1,4,0,1,0,0,0,0>$



siehe: Rao et al. 1999



Color Histogram: Test

- Test von Rao, Srihari&Zhang (1999)
- Test setting: 500 images in 41 similarity groups
- 5 types of histograms: „normal“, CCV, annular, angular, hybrid
- Results - performance:
 1. Annular
 2. Angular
 3. Hybrid
 4. CCV
 5. „normal“

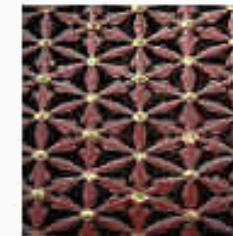
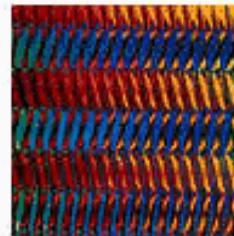
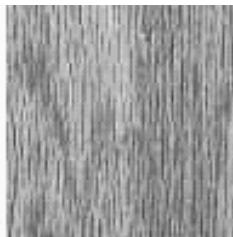
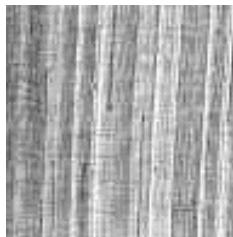


Textur

- „...the way a fabric or other substance feels to the touch, its thickness or firmness or solidity.“ (Oxford American Dictionary)
- „...something composed of closely interwoven elements. Frequent repetition of similar elements (*texemes*).“ (*Webster's Dictionary*)
- „In very general terms, it is reasonable to say that a texture is a combination of a relatively large number of objects in which the individuality of each of object is not regarded as important, and in which the structural properties of the set of objects is predominant.“ (Santini 2001)



Examples of Textures





Texture Representation Methods

Tamura Features

- *coarseness* (Granularität)
- *contrast* (Kontrast)
- *directionality* (Richtung)
- *linelikeness* (Linienhaftigkeit)
- *regularity* (Regelmäßigkeit)
- *roughness* (Rauhigkeit)

Rao & Lohse Features

- *repetitiveness* (Wiederholung)
- *directionality* (Richtung)
- *granularity & complexity* (Granulatität & Komplexität)

Wold Decomposition

- *harmonic* (harmonisch)
- *evanescent* (flüchtig)
- *indeterministic* (unbestimmt)

Amundson & King Features

- *coarseness* (Einfachheit, Granulatität)
- *contrast* (Kontrast)
- *busyness* (Hektik)
- *complexity* (Komplexität)
- *texture strength* (Wirksamkeit)



Shape-Based Modeling: simple measures

- Circumference (Umfang)
- Area (Fläche)
- Compactness
- Center of gravity (Schwerpunkt)



Edge Detection

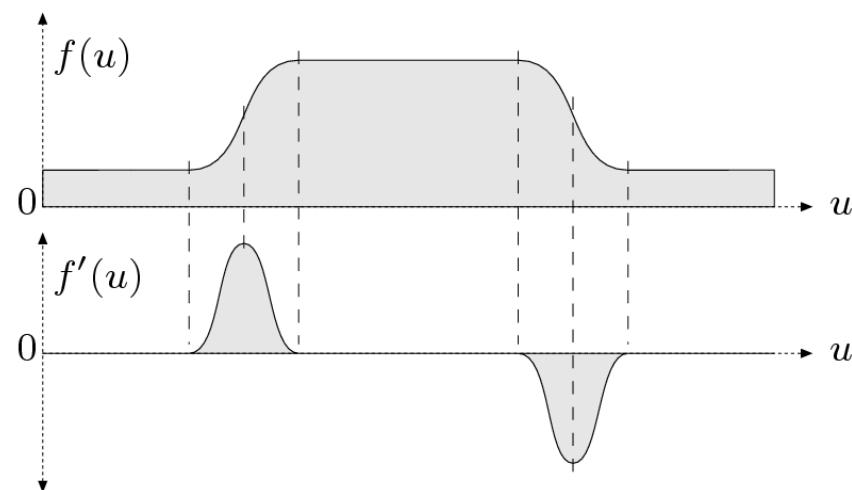




Gradient-based Edge Detection



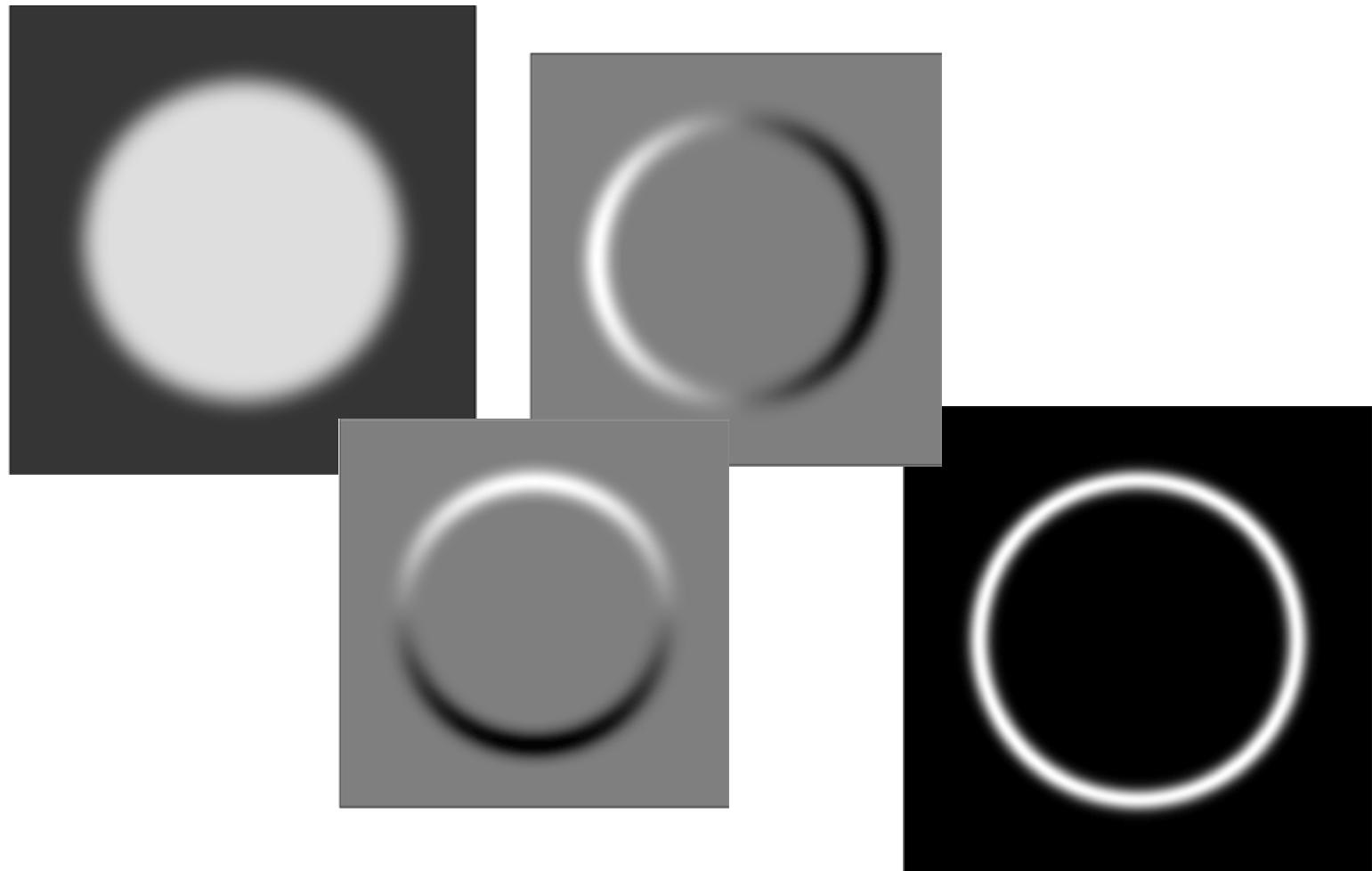
(a)



(b)

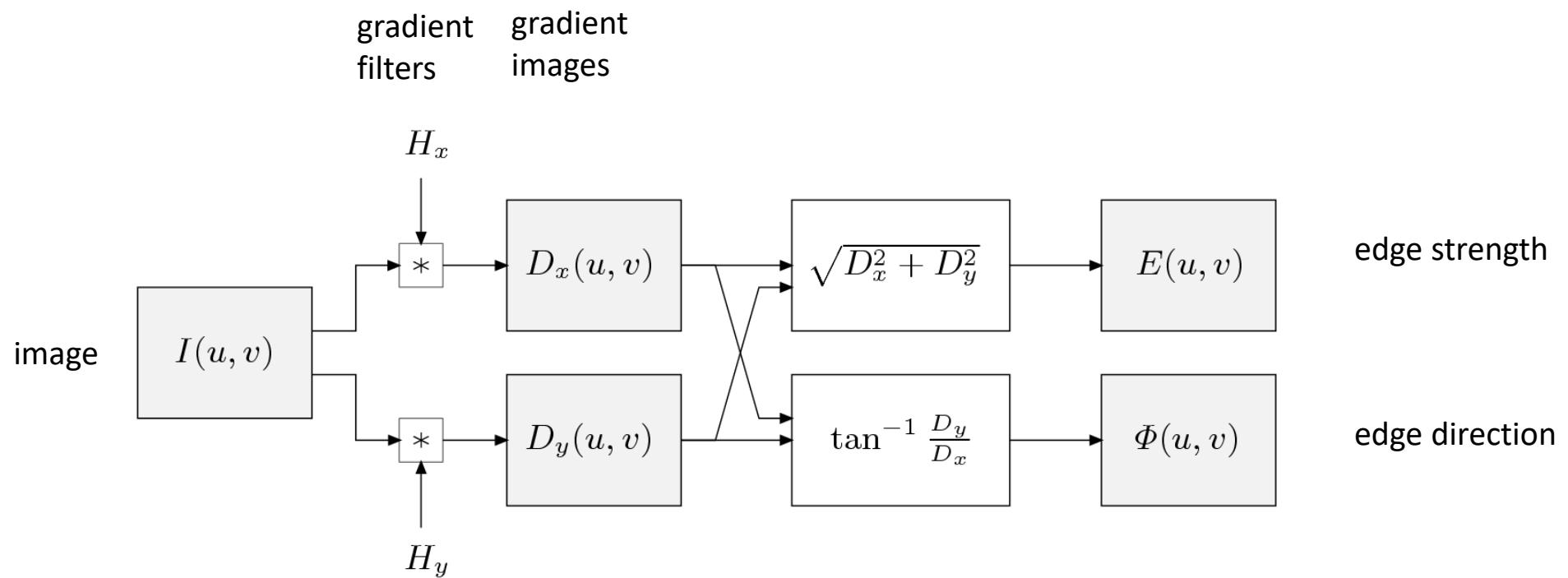


Gradienten-basierte Kantendetektion IV

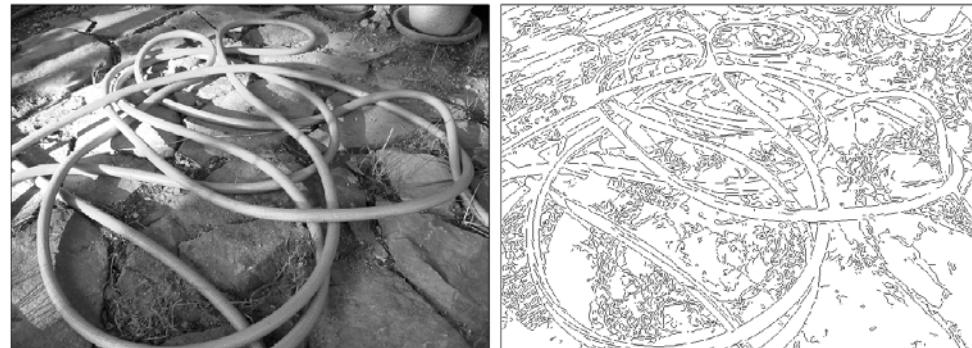




Process Edge Detection



Canny-Filter



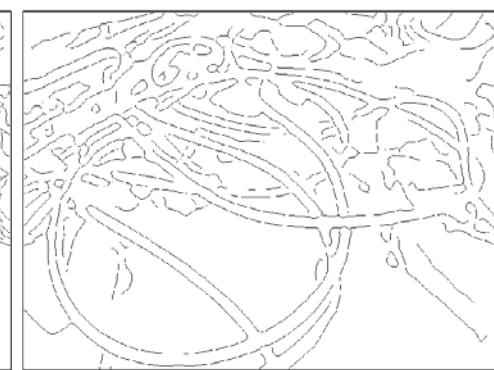
Original



$\sigma = 1.0$



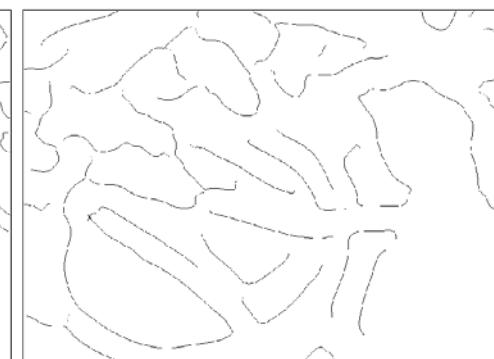
$\sigma = 2.0$



$\sigma = 4.0$

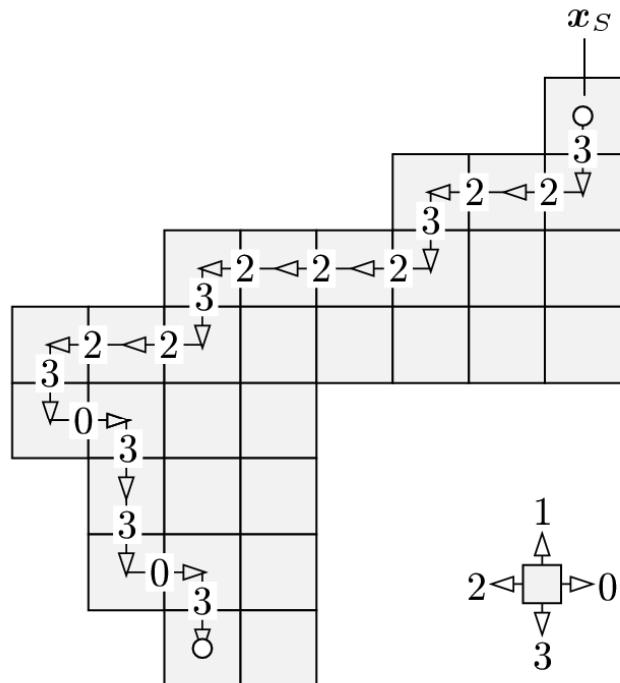


$\sigma = 8.0$



$\sigma = 16.0$

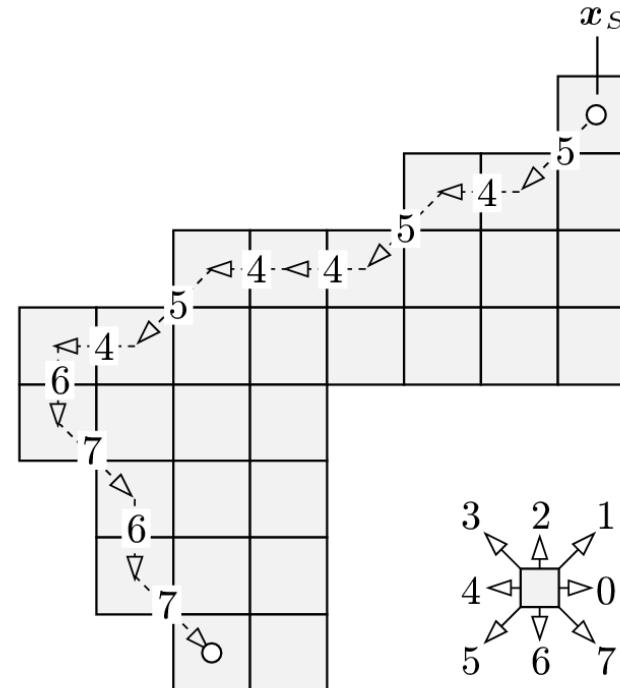
Codierung: Chain Codes



4-Chain Code

32223222322303303...111

length = 28



8-Chain Code

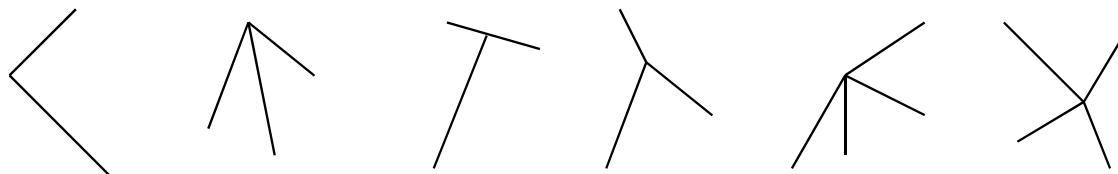
54544546767...222

length = $18 + 5\sqrt{2} \approx 25$



Codierung: LISG

- Labeled Image Structure Graph



L

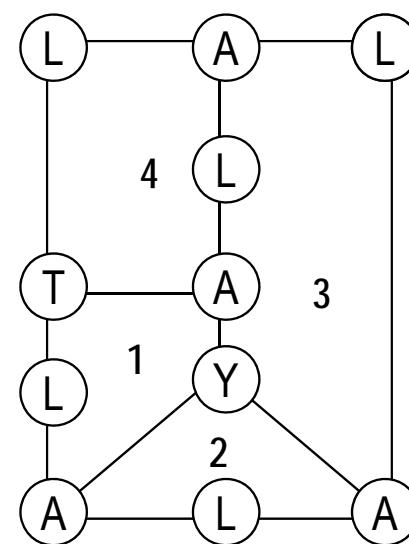
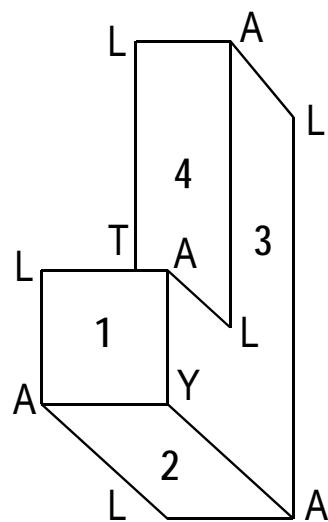
A

T

Y

K

X



Outer Shape:

LALALT

Inner Shapes:

1: AYALT

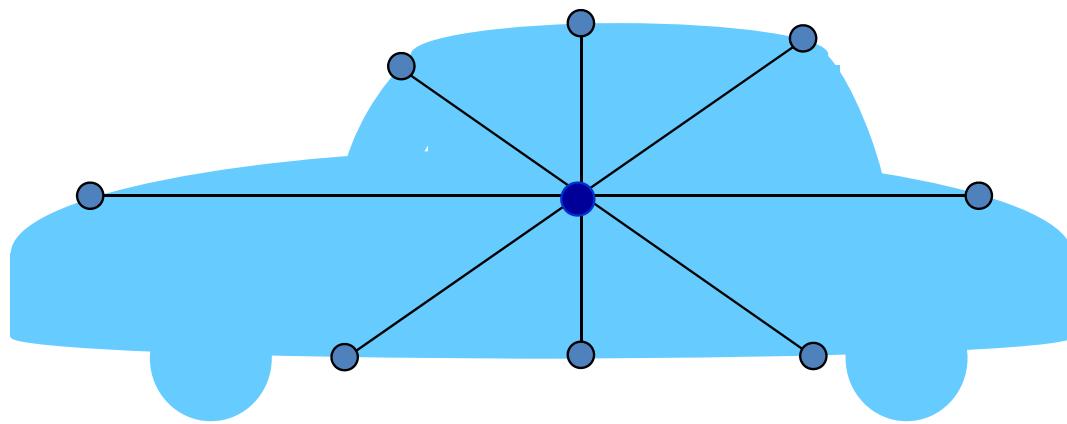
2: YALA

3: LAYALA

4: ALATL

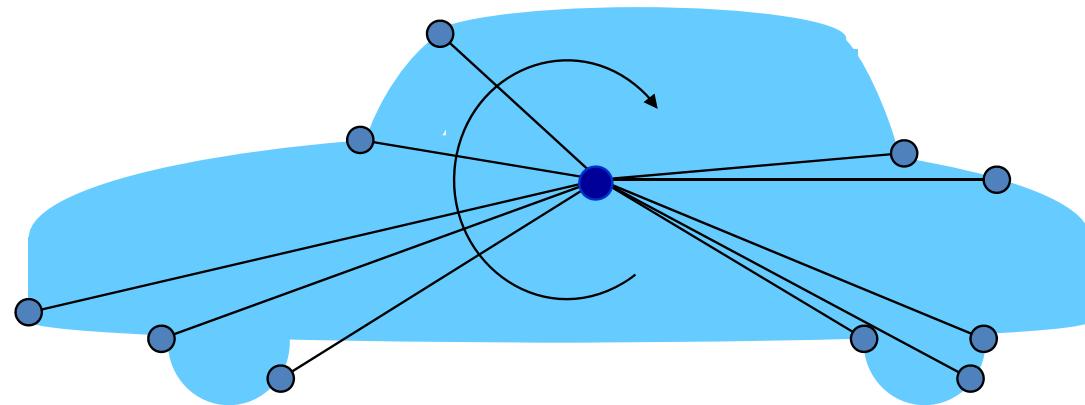


Shape Radii Representation



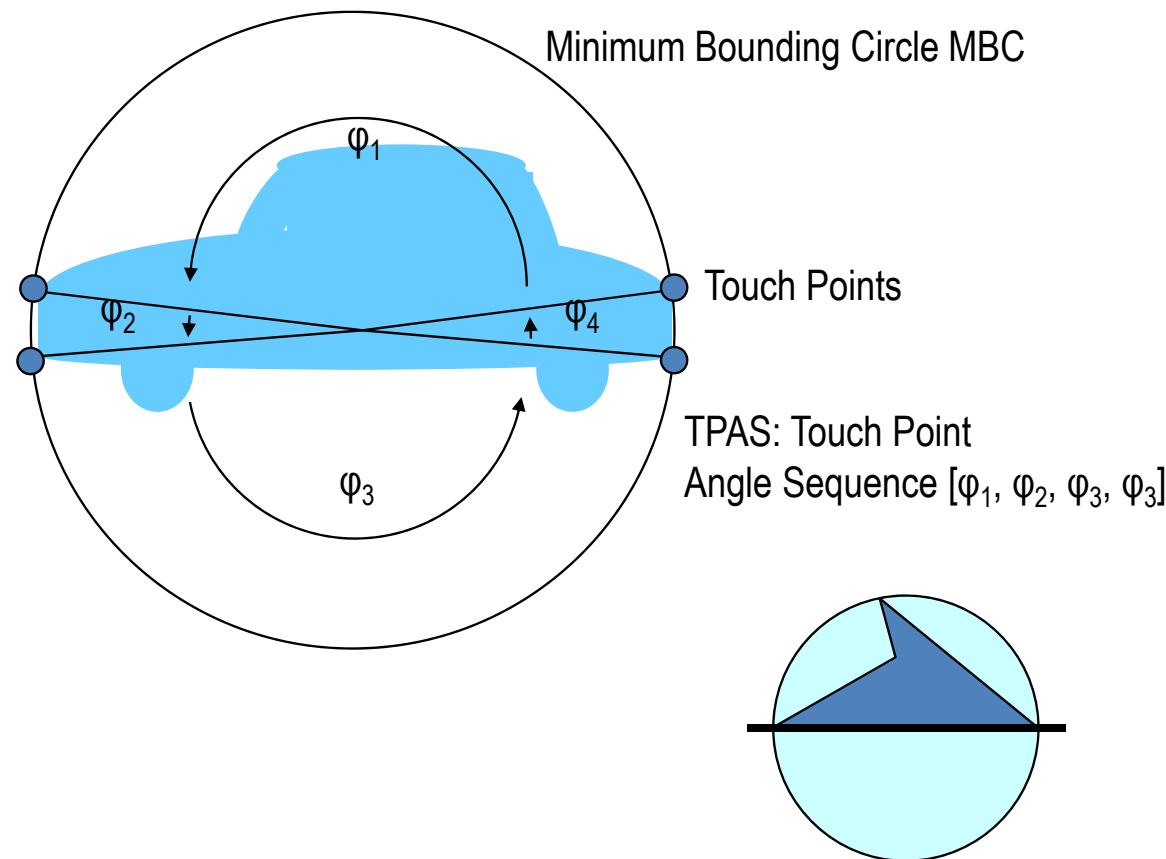


Turning Point Representation



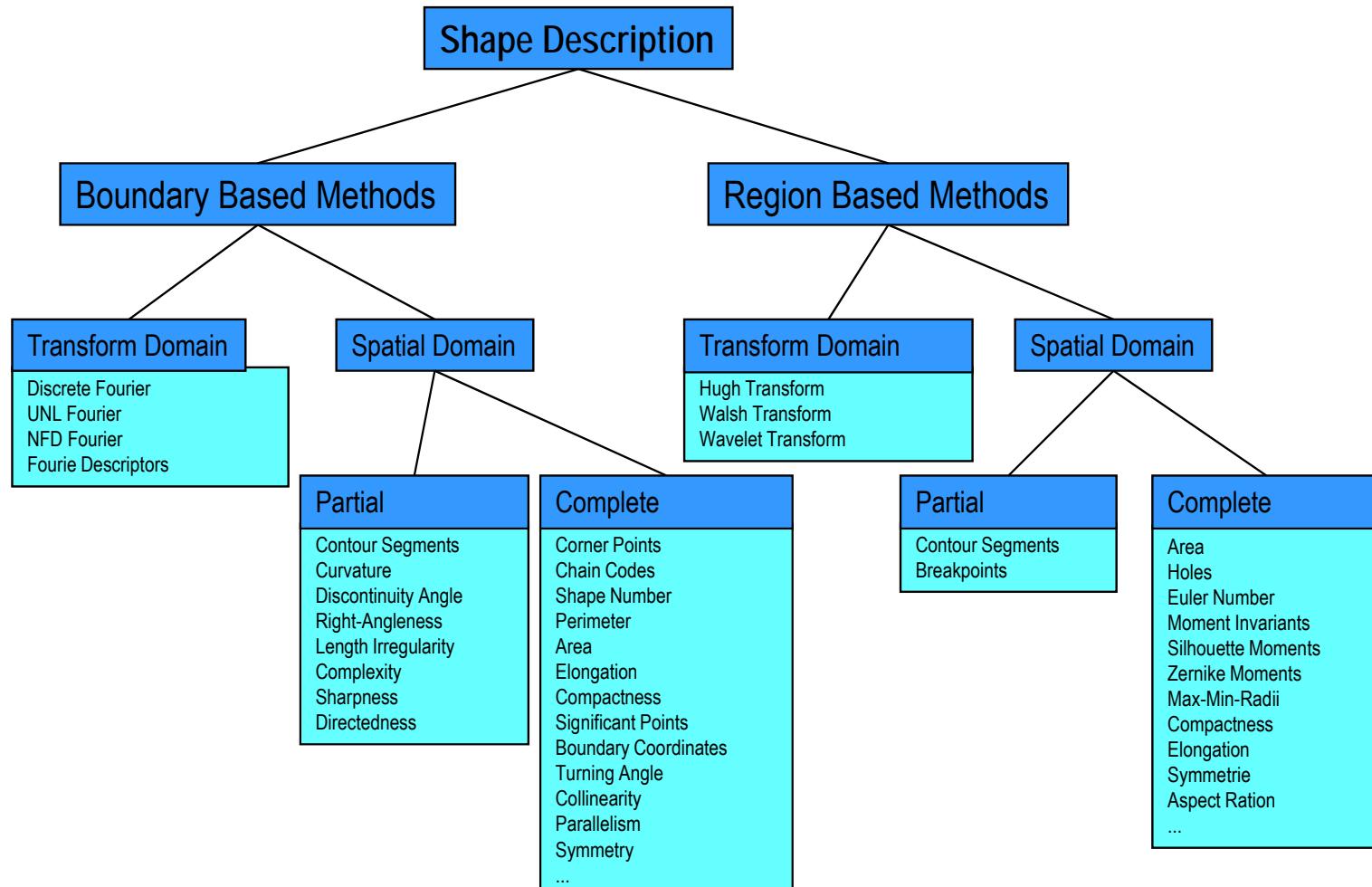


Minimum Bounding Circle (MBC)





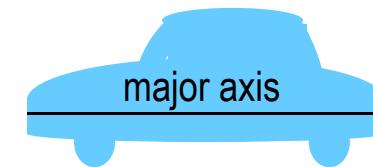
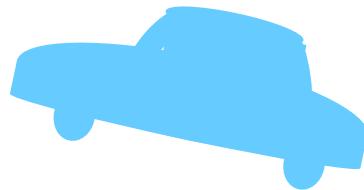
Taxonomie Shape Based Retrieval



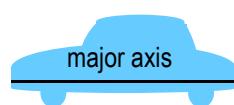


Normalisierung

- Rotation



- Scale

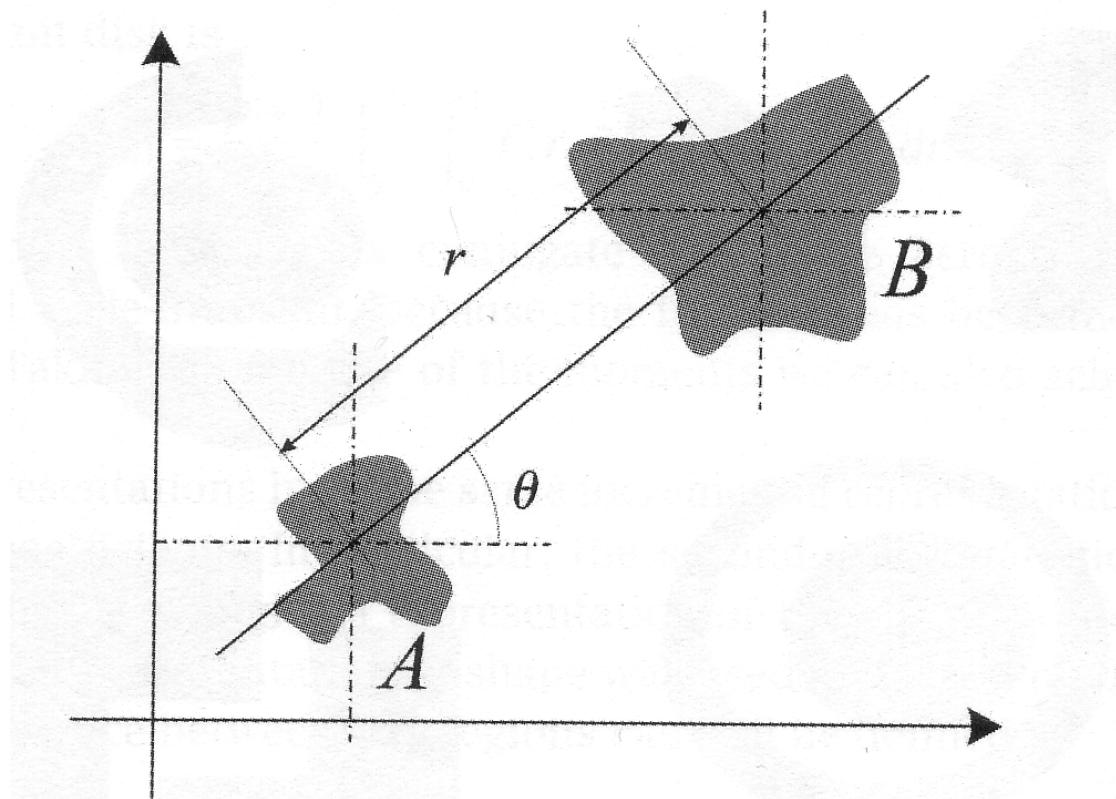




Spatial Relations

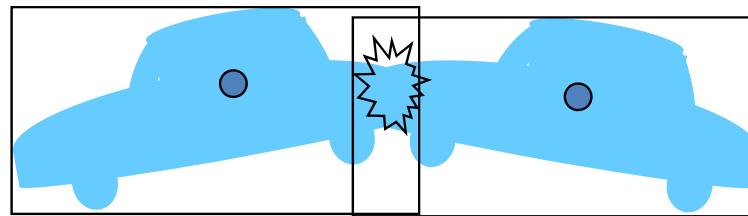
Type of Relation	Invariant to...	Example
Absolute	None	at position x,y
Distance	Translation	close to
Directional	Scale	left of
Topological	Continuous	Overlapping

Comparison of centroids (*Schwerpunkte*)



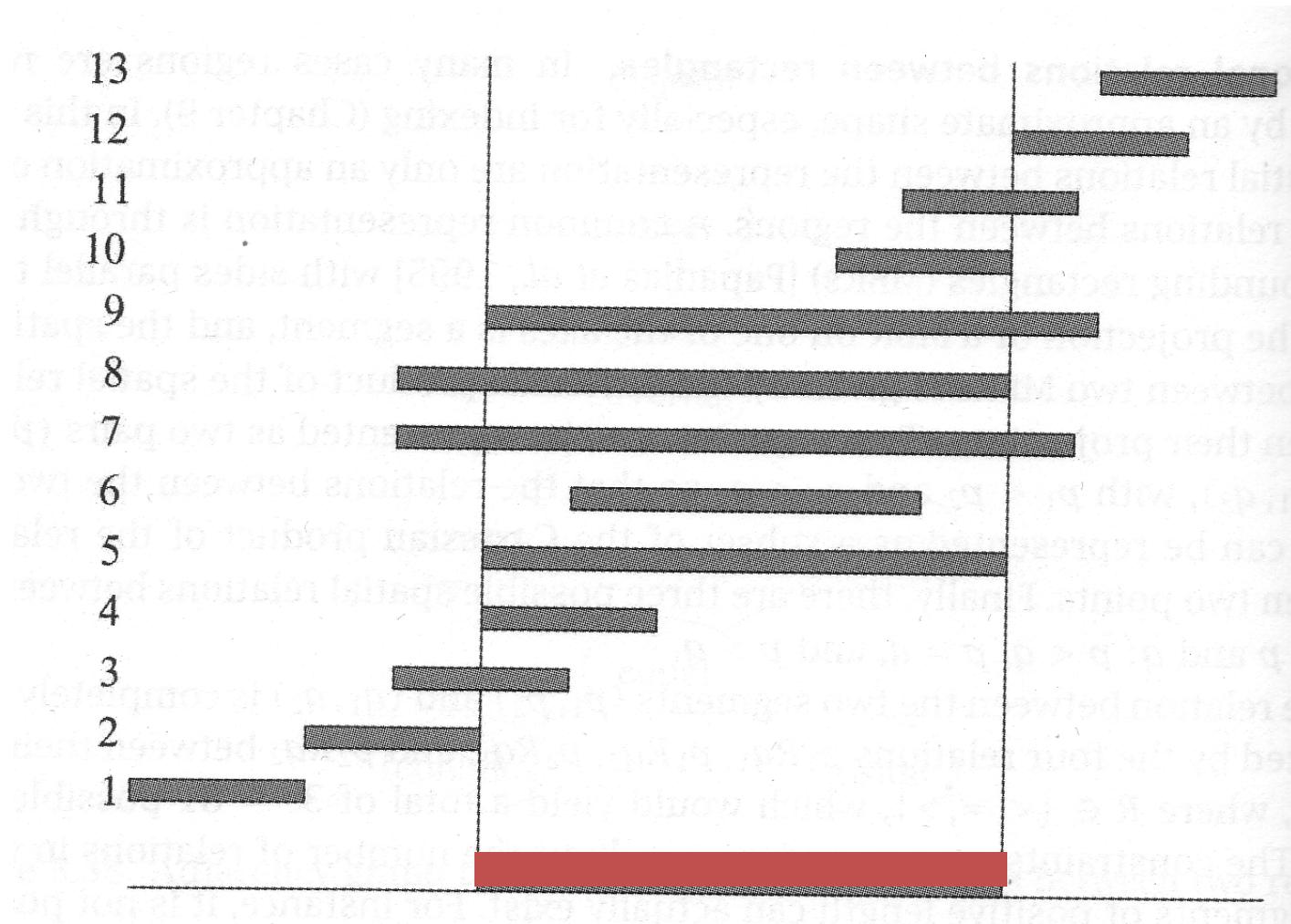


Topological Relations

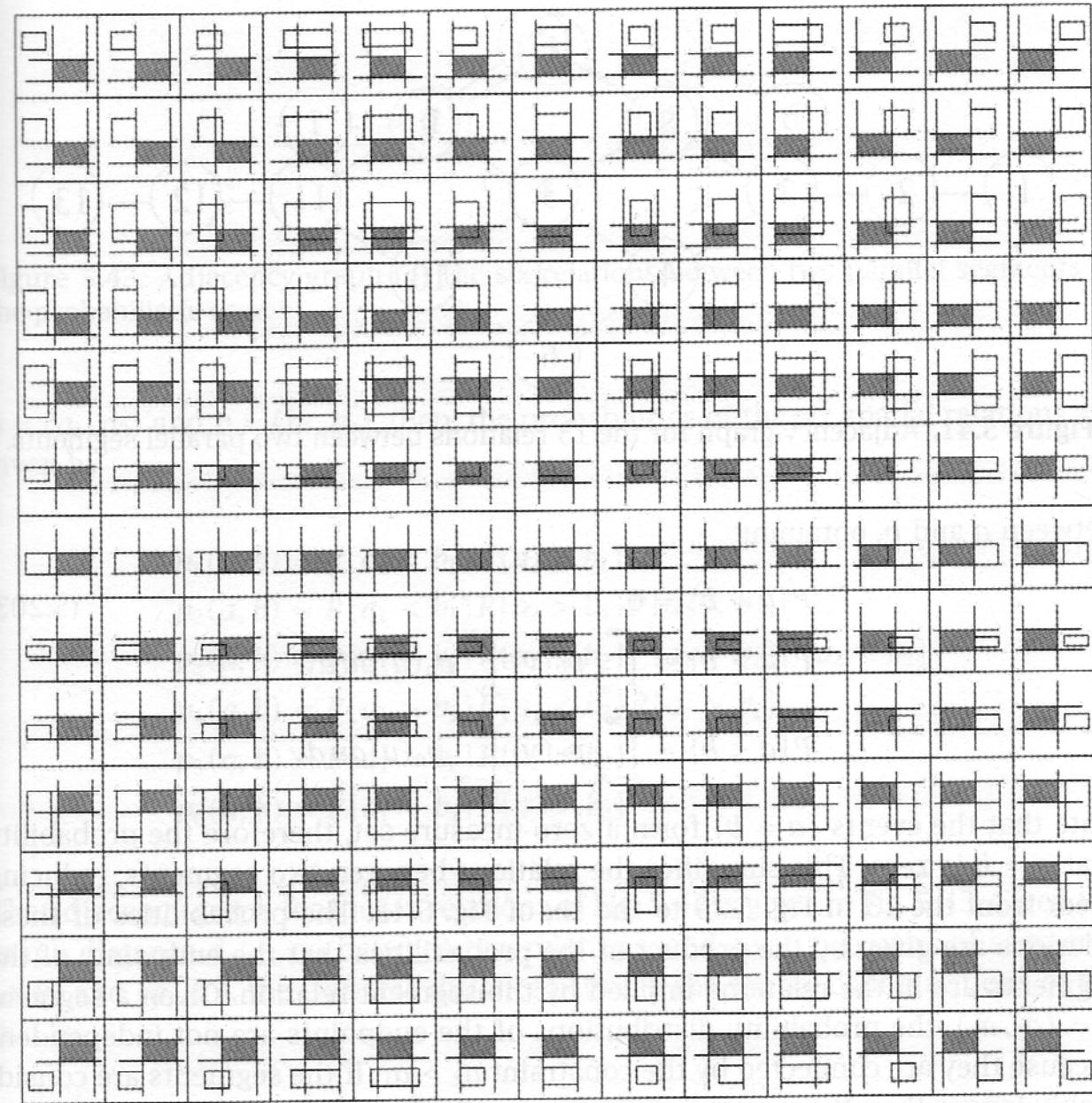




Spatial Relations



Spatial F



169 possible
spatial relations
between two
rectangles





Topological Relations

