RuSSIR
Russian Summer School in Information Retrieval 2008

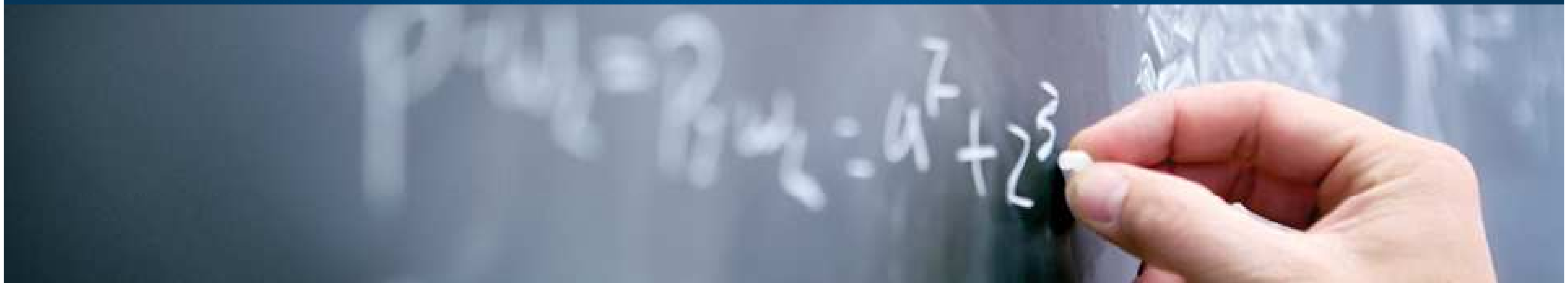
# Content Based Image Retrieval

Natalia Vassilieva
nvassilieva@hp.com
HP Labs Russia

# Tutorial outline

- Lecture 1
  - Introduction
  - Applications
- Lecture 2
  - Performance measurement
  - Visual perception
  - Color features
- Lecture 3
  - Texture features
  - Shape features
  - Fusion methods
- Lecture 4
  - Segmentation
  - Local descriptors
- Lecture 5
  - Multidimensional indexing
  - Survey of existing systems

# Lecture 3
# Texture features
# Shape features
# Fusion methods

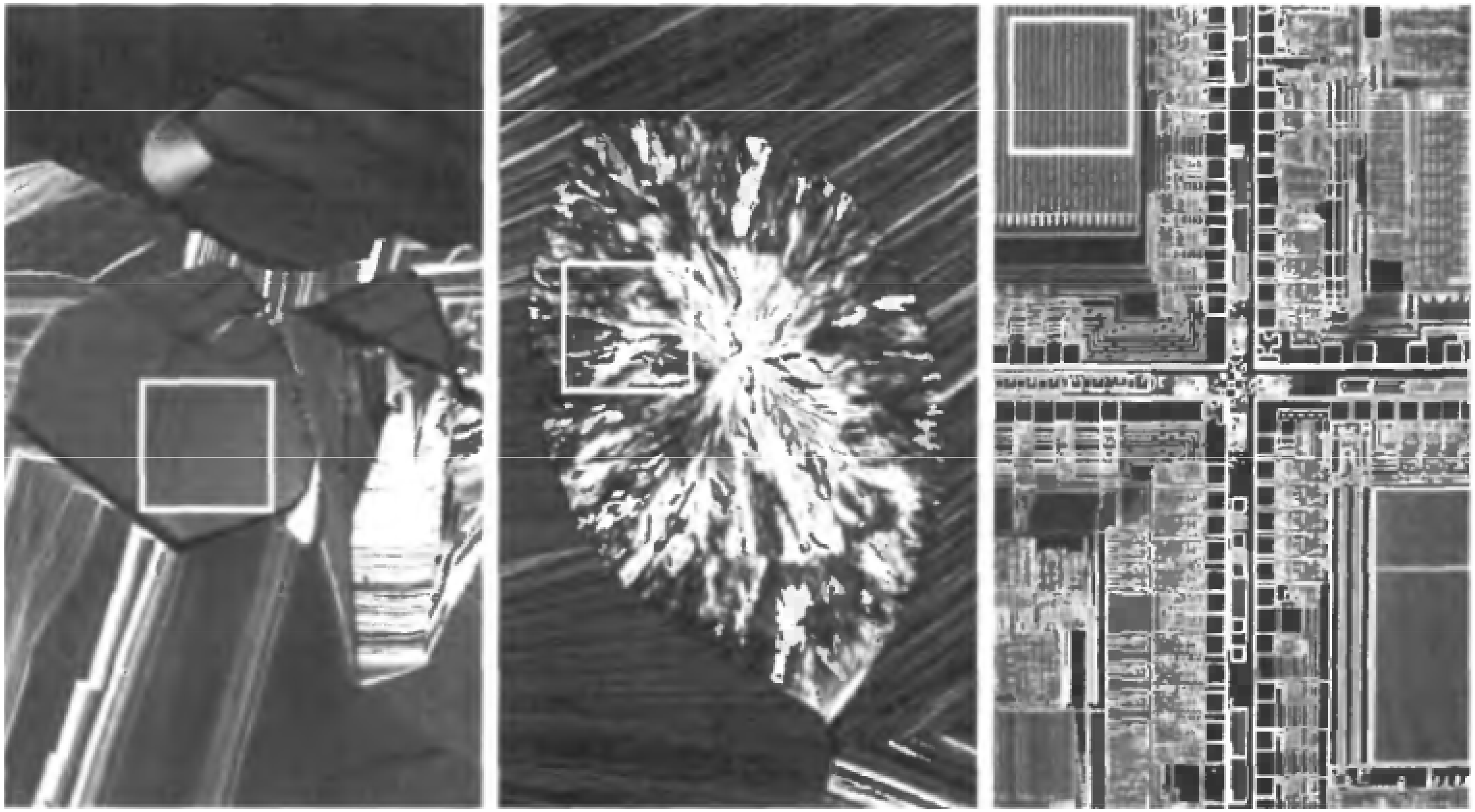


# Lecture 3: Outline

- Texture features
  - Statistical
  - Spectral
  - Comparison
- Shape features
  - Boundary based
  - Region based
  - Comparison
- Fusion methods

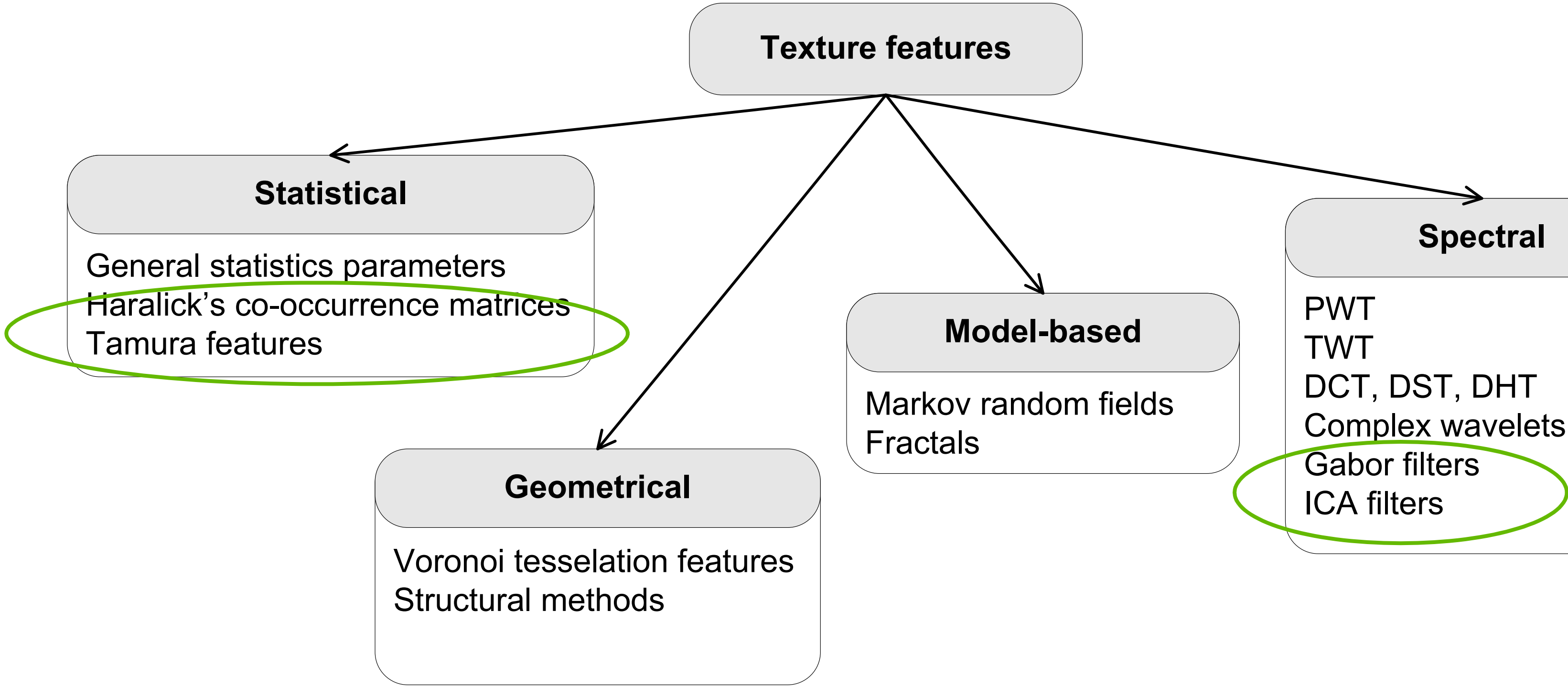
# Texture features

- What is texture?



Smooth Rough Regular

# Texture features



# Texture features

- ## General statistics

Based on intensity histogram of the whole image or its regions:

$p(z_i), i = 0, 1, 2, \ldots, L-1$ – histogram of intensity, L – number of intensity levels.

$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i)$ – central moment of order n.

$m = \sum_{i=0}^{L-1} z_i p(z_i)$ – average intensity.

---

$\sigma^2(z) = \mu_2(z)$ – variance, is a measure of contrast.

$R = 1 - \frac{1}{1+\sigma^2(z)}$ , R=0 where intensity is equal.

$\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$ – a measure of histogram assimetry.

# Texture features

- General statistics (2)

$$U = \sum_{i=0}^{L-1} p^2(z_i)$$ – a measure of contrast of homogeneity (max for homogeneous areas ).

$$e = -\sum_{i=0}^{L-1} p(z_i)\log_2 p(z_i)$$ – entropy, a measure of variability (0 for homogeneous areas ).

| Texture | Average | Deviation | R | $\mu_3$ | U | Entropy |
|---|---|---|---|---|---|---|
| Smooth | 82,64 | 11,79 | 0,002 | –0,105 | 0,026 | 5,434 |
| Rough | 143,56 | 74,63 | 0,079 | –0,151 | 0,005 | 7,783 |
| Regular | 99,72 | 33,73 | 0,017 | 0,750 | 0,013 | 6,674 |

# Texture features

## Grey Level Co-occurrence Matrices (GLCM):

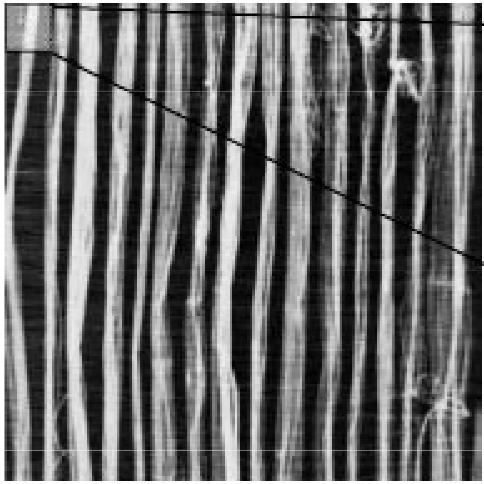
GLCM - matrix of frequencies at which two pixels, separated by a certain vector, occur in the image.

$$C(i,j) = \sum_{p=1}^{N} \sum_{q=1}^{M} \begin{cases} 1, \ if \ \ I(p,q) = i, \ I(p+\Delta x, q+\Delta y) = j \\ 0, \ otherwise \end{cases}$$

$(\Delta x, \Delta y)$ – separation vector;

I(p,q) – intensity of a pixel in position (p, q).

# GLCM – an example



original image

image pixel values

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 211 | 191 | 71 | 53 | 48 | 44 | 43 | 44 |
| 212 | 191 | 76 | 60 | 52 | 52 | 53 | 52 |
| 212 | 180 | 75 | 75 | 75 | 76 | 76 | 75 |
| 212 | 140 | 53 | 46 | 49 | 49 | 46 | 44 |
| 211 | 115 | 53 | 44 | 44 | 43 | 43 | 46 |
| 211 | 115 | 54 | 44 | 44 | 46 | 37 | 43 |
| 211 | 123 | 55 | 52 | 52 | 52 | 44 | 44 |
| 211 | 141 | 76 | 74 | 75 | 75 | 71 | 70 |

quantisation
4 levels

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |

quantised image

vector – length 1 pixel, direction horizontal

| | | | |
|---|---|---|---|
| 28 | 0 | 0 | 0 |
| 5 | 10 | 0 | 0 |
| 1 | 4 | 0 | 0 |
| 0 | 3 | 5 | 0 |

co-occurrence matrix

$$(\Delta x, \Delta y) = (1,0)$$

# GLCM – descriptors

## Statistical parameters calculated from GLCM values:

$$Energy = \sum_i \sum_j C^2(i,j)$$ – is minimal when all elements are equal

$$Entropy = -\sum_i \sum_j C(i,j) \log C(i,j)$$ – a measure of chaos, is maximal when all elements are equal

$$Contrast = \sum_i \sum_j (i-j)^2 C(i,j)$$ – has small values when big elements are near the main diagonal

$$Inverse\ Difference\ Moment = \sum_i \sum_j \frac{C(i,j)}{1+(i-j)^2}$$

– has small values when big elements are far from the main diagonal

# Texture features: Tamura features

## Features, which are important for visual perception:

- Coarseness
- Contrast
- Directionality
- Line-likeness
- Regularity
- Roughness

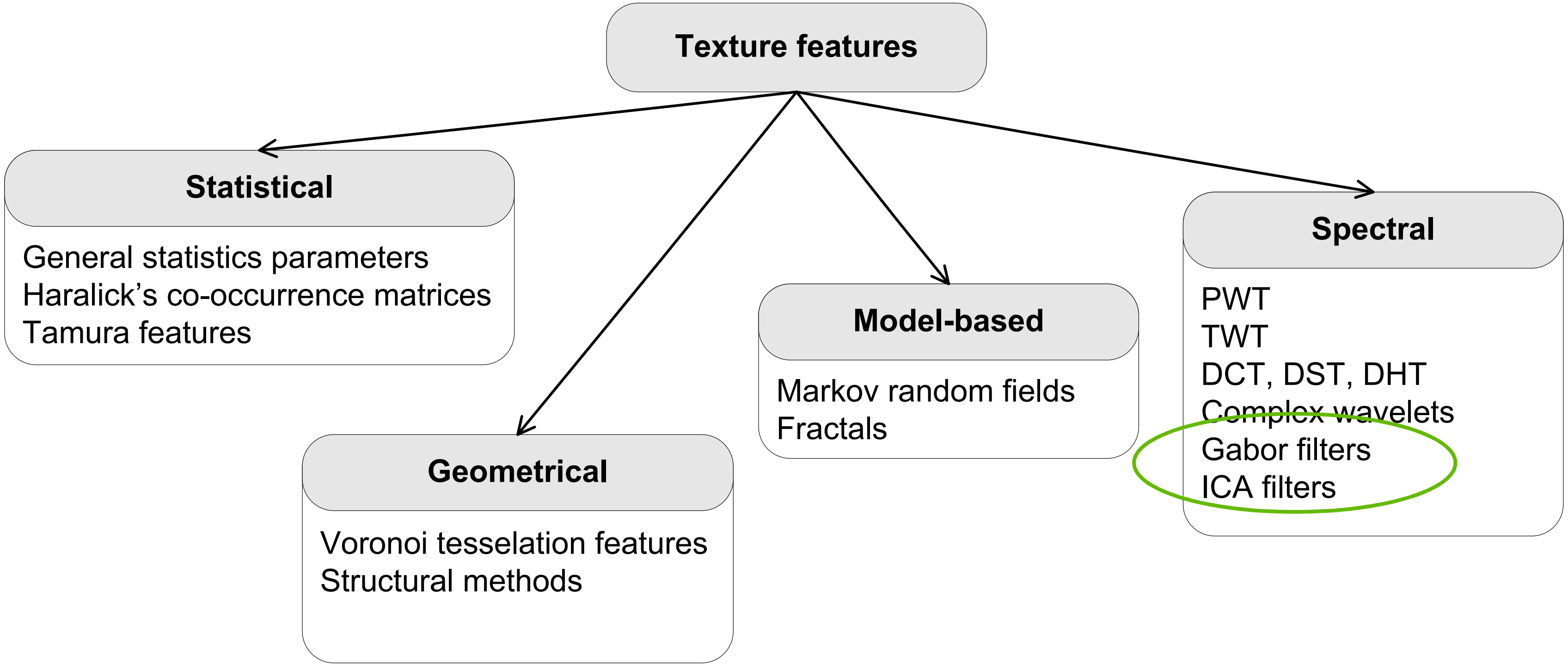
Tamura image:

Coarseness-coNtrast-Directionality – points in 3-D space CND

Features:

- Euclidean distance in 3D (QBIC)
- 3D histogram (Mars)

# Texture features: spectral



# Texture features: wavelet based

Wavelet analysis – decomposition of a signal:

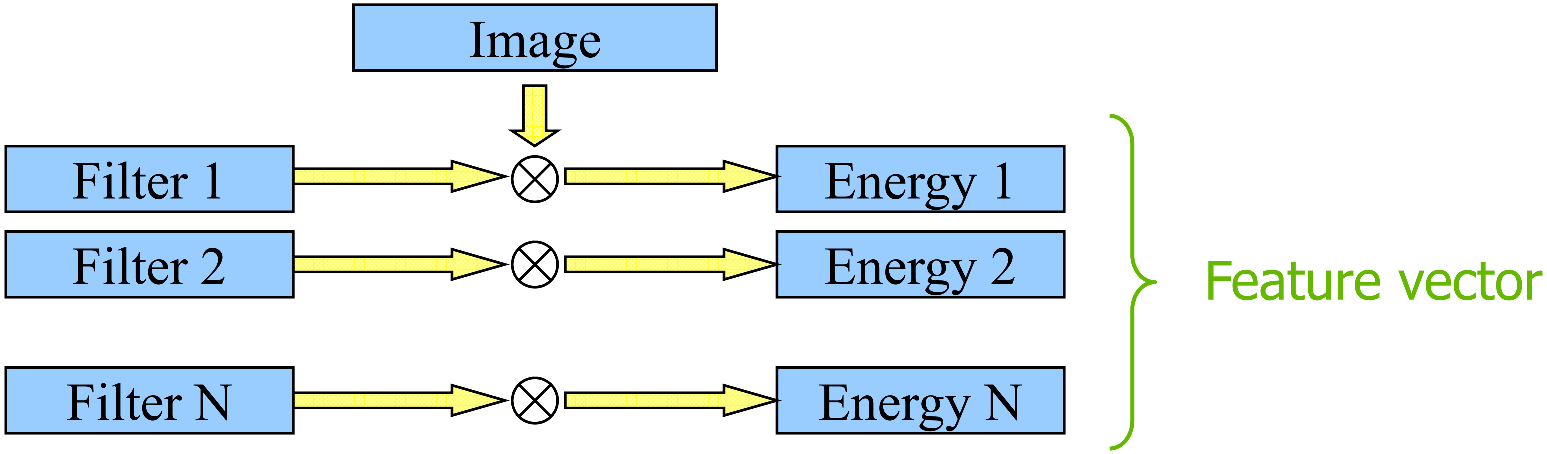
$$f(x) = \sum_{j,k} \alpha_k \psi_{j,k}(x)$$

Basis functions:

$$\psi_{j,k} = 2^{j/2} \varphi(2^j x - k)$$ – scaling function

$$j, k \in Z, \quad \varphi(x) \in L^2(R)$$ – mother wavelet

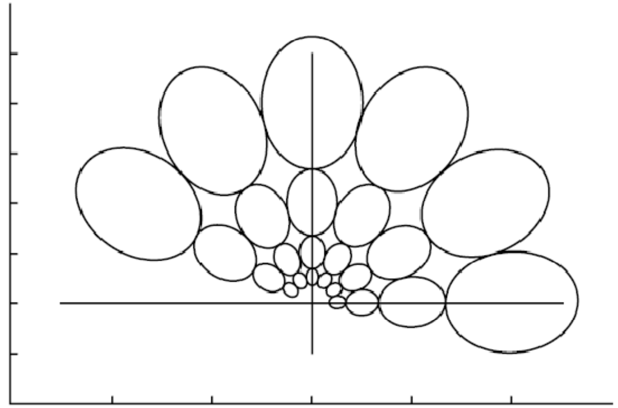
A set of basis functions – filters bank



# Texture features: Gabor filters

Mother wavelet: Gabor function

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right]$$



Filters bank:

$$g_{mn}(x,y) = a^{-m}g(x',y'), \quad a>1, \quad m,n = \text{integer}, \quad \mathrm{m} = 0,1,\ldots,\mathrm{S}-1,$$

$$x' = a^{-m}(x\cos\Theta + y\sin\Theta),$$

$$y' = a^{-m}(-x\sin\Theta + y\cos\Theta),$$

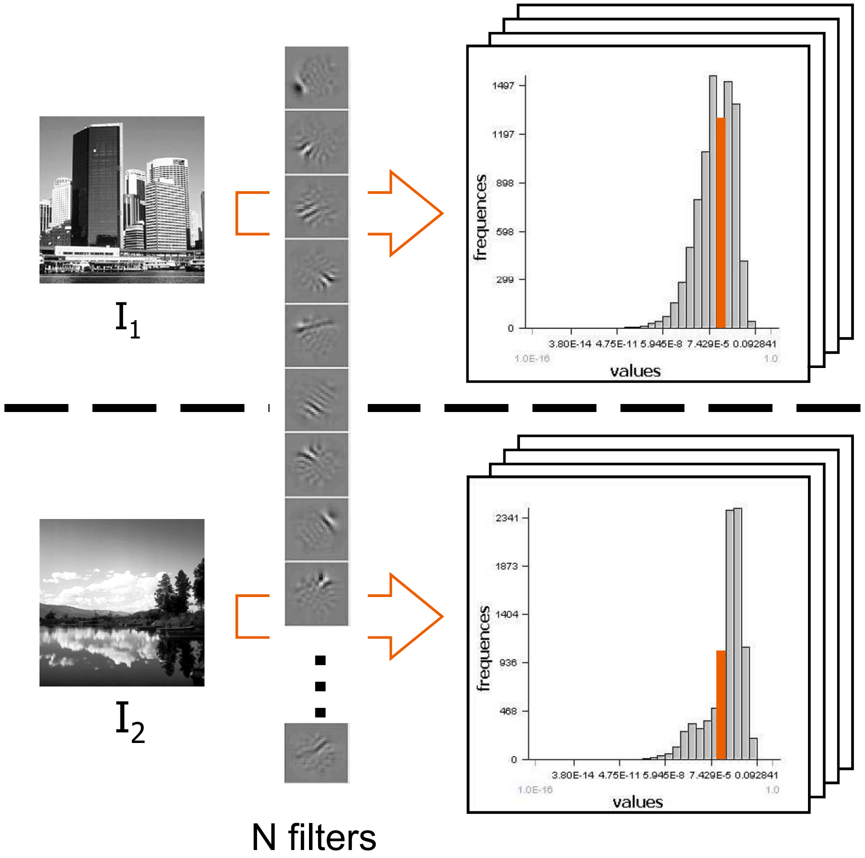
$$\Theta = n\pi / K$$

$$a = (U_h / U_l)^{-1/(S-1)}$$

K – a number of directions,
S – a number of scales,
$U_h$, $U_l$ – max and min of frequencies taken into consideration.

# Texture features: ICA filters

Filters are obtained using Independent Component Analysis

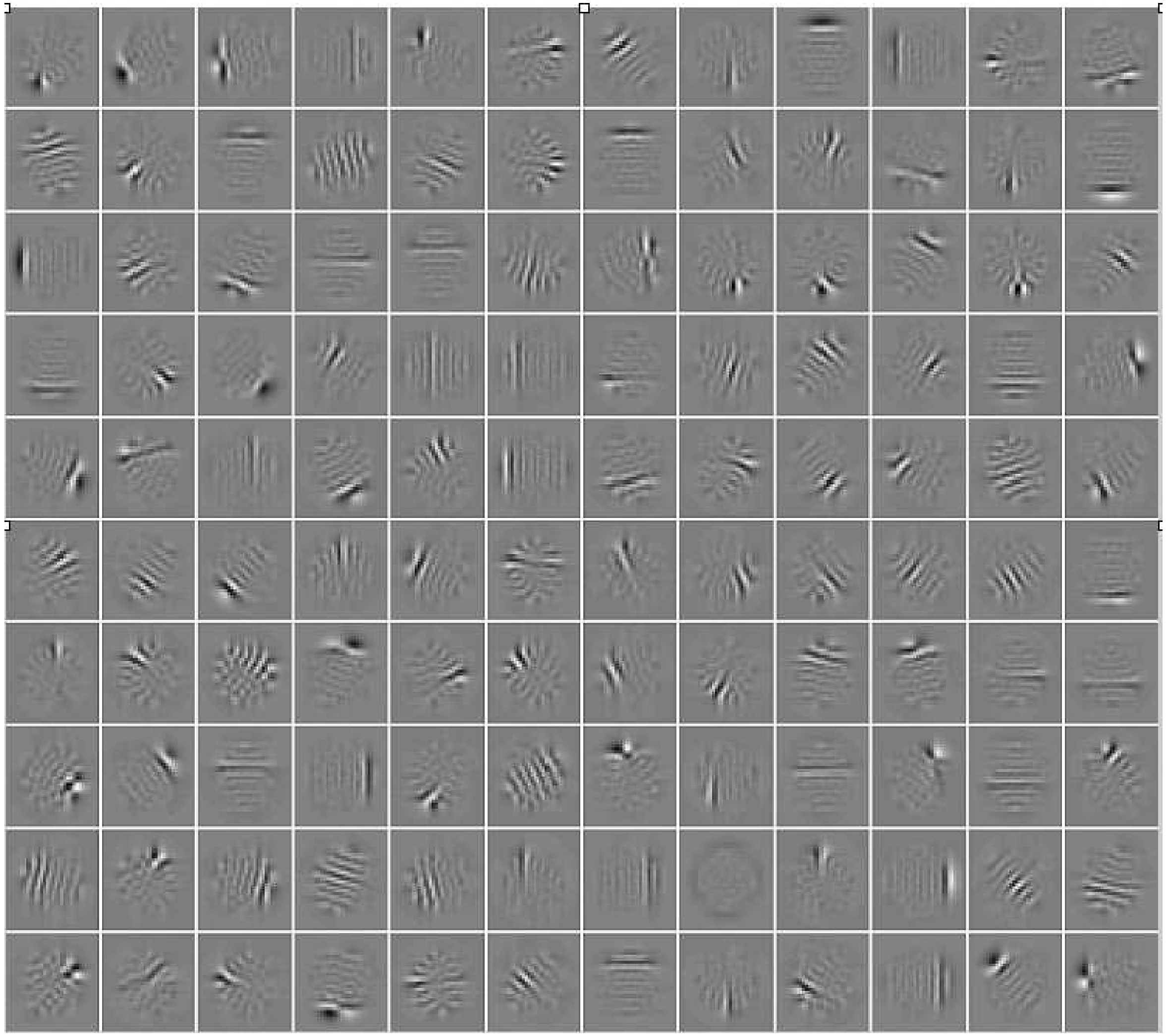



$$KL_H(H_1,H_2)=\sum_{b=1}^{B}(H_1(b)-H_2(b))\log\frac{H_1(b)}{H_2(b)}$$

$$dist(I_1,I_2) = \sum_{i=1}^{N} KL_H(H_{1i}, H_{2i})$$

H. Borgne, A. Guerin-Dugue, A. Antoniadis. Representation of images for classification with independent features. Pattern Recognition Letters, vol. 25, p. 141-154, 2004

# ICA Filters



# Lecture 3: Outline

- Texture features
  - Statistical
  - Spectral
  - Comparison
- Shape features
  - Boundary based
  - Region based
  - Comparison
- Fusion methods

# Texture features: comparison

**Table 6: TRECVID evaluation – mean average precision retrieval**

| Feature | Single | Combined with HSV |
|---|---|---|
| Gabor-2-4 | 3.93% | 4.31% |
| Co-occurrence homogeneity | 2.85% | 3.03% |
| Tamura standard all | 2.57% | 3.43% |
| Tamura CND | 1.65% | 2.72% |
| Tamura coarseness-2 | 0.97% | 2.49% |

**Table 7: ImageCLEF retrieval results**

| Feature | Mean average precision |
|---|---|
| Gabor-2-4 | 35.3% |
| Co-occurrence homogeneity | 19.8% |
| Tamura standard all | 20.7% |
| Tamura CND | 18.4% |
| Tamura coarseness-2 | 14.5% |

In the context of image retrieval!

P. Howarth, S. Rüger. Robust texture features for still image retrieval.
In Proc. IEE Vis. Image Signal Processing, vol. 152, No. 6, December 2006

# Texture features: comparison (2)

## Gabor filters v. s. ICA filters

Image classification task:

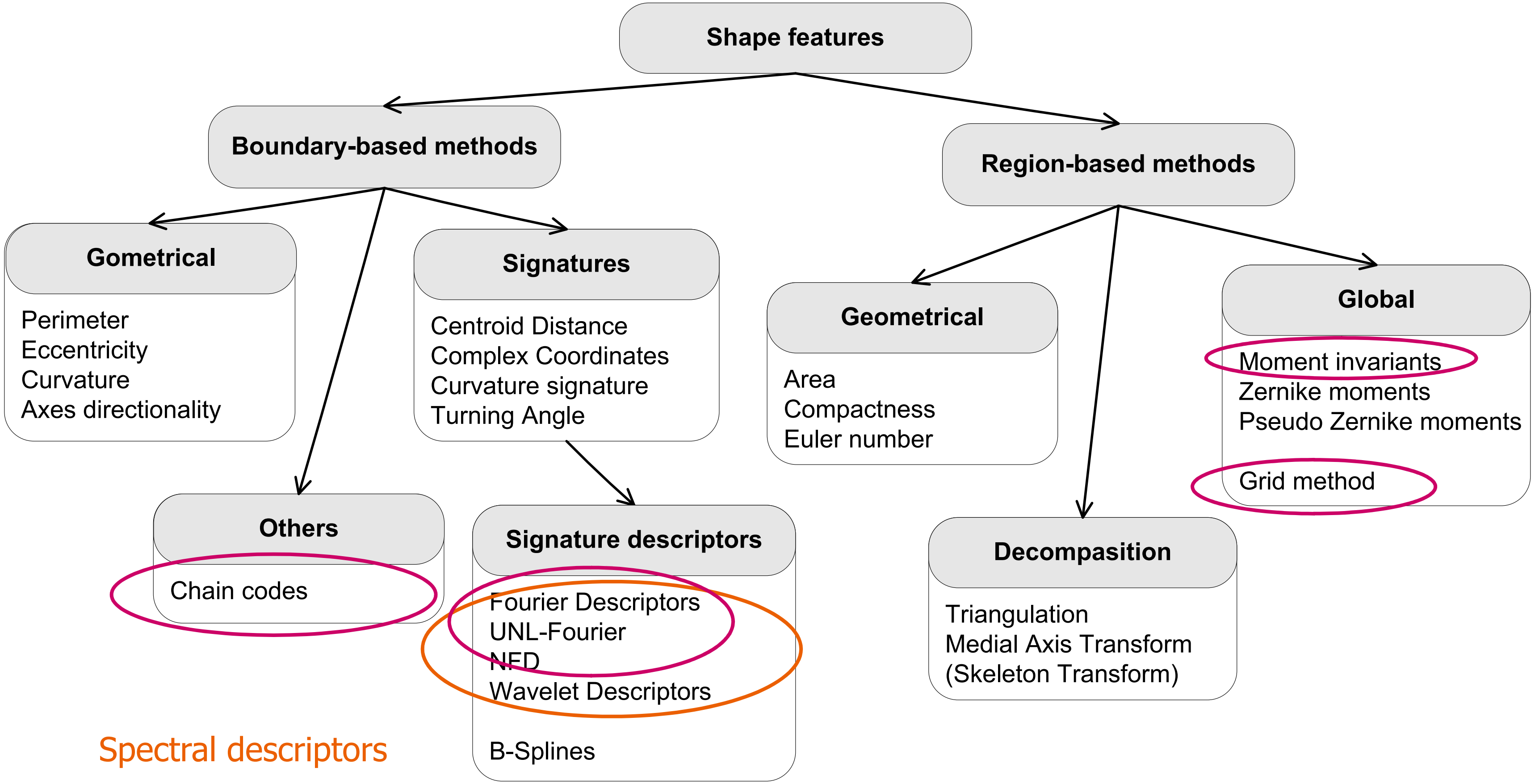
- Collection of angiographic images
  - ICA filters performs better by 13%
- Brodatz texture collection
  - ICA filters perform better by 4%

Snitkowska, E. Kasprzak, W. Independent Component Analysis of Textures in Angiography Images. Computational Imaging and Vision, vol. 32, pages 367-372, 2006.

# Lecture 3: Outline

- Texture features
  - Statistical
  - Spectral
  - Comparison
- Shape features
  - Boundary based
  - Region based
  - Comparison
- Fusion methods

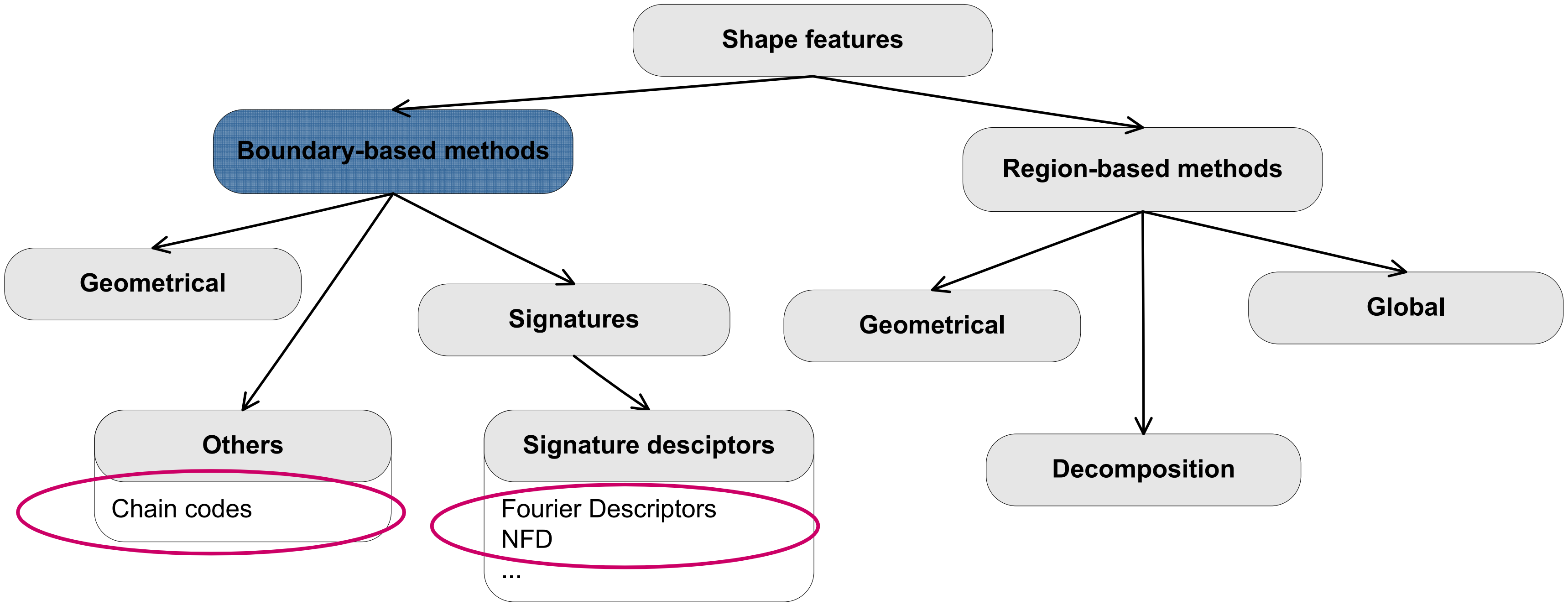
# Shape features



# Requirements to the shape features

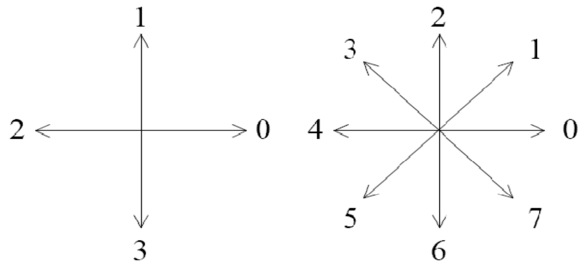
- Translation invariance
- Scale invariance
- Rotational invariance
- Stability against small form changes
- Low computation complexity
- Low comparison complexity

# Boundary-based features

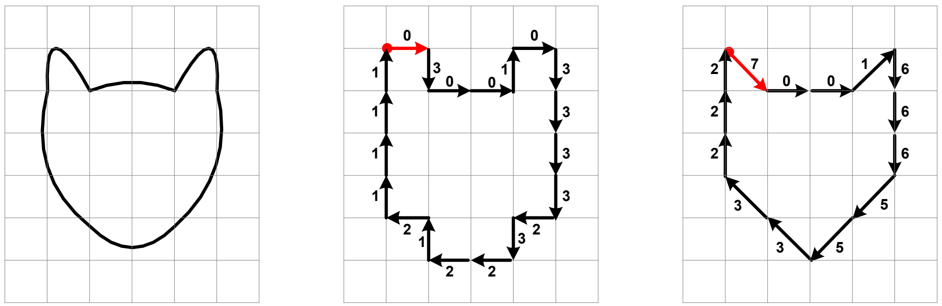


# Chain codes

Directions for 4-connected and 8-connected chain codes:




Example:



A

B

A: 03001033332322121111

B: 70016665533222

Starting point invariance: minimal code

70016665533222 -> 00166655332227

Rotation invariance: codes subtraction

00166655332227 -> 01500706070051

# Fourier descriptors

1. Signature calculation (2D -> 1D):
   - Centroid – contour distance
   - Complex coordinates: z(t) = x(t) + iy(t)
   - ...

2. Perform the discrete Fourier transform, take coefficients (s(t) – signature):

$$u_n = \frac{1}{N}\sum_{t=0}^{N-1} s(t)e^{-j2\pi nt/N}$$

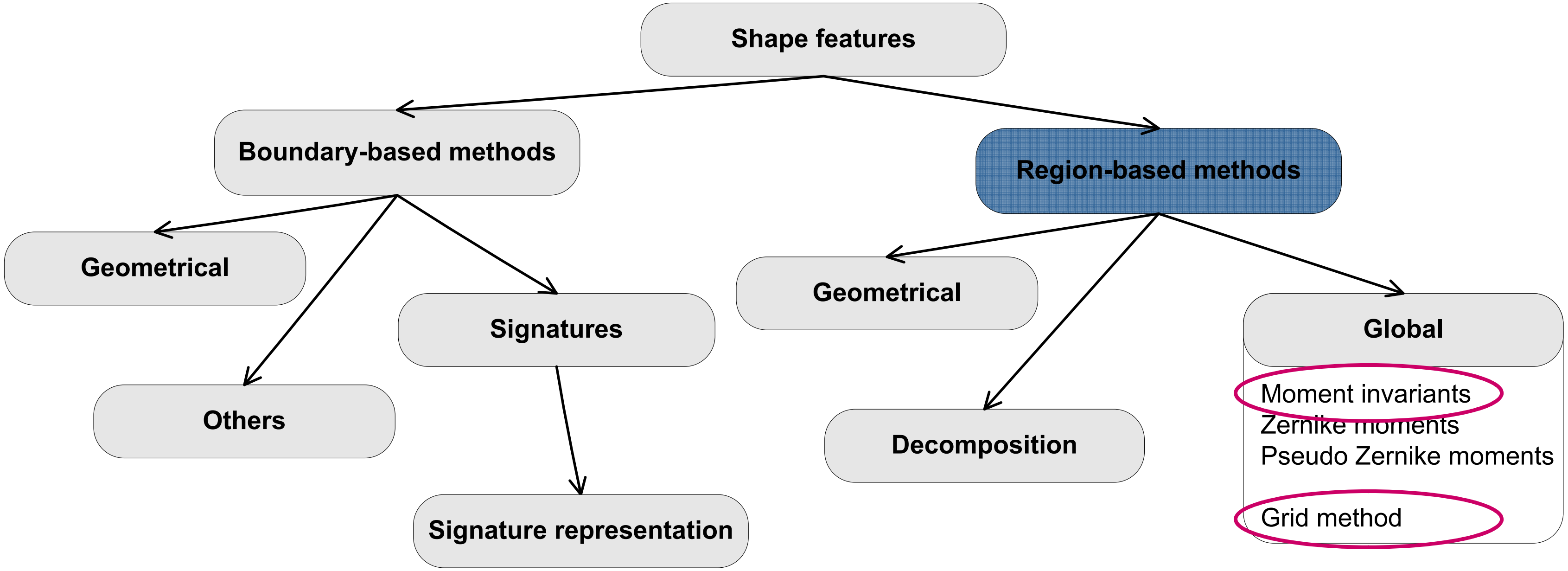
3. Normalization (NFD – Normalized Fourier Descriptors):

$$\frac{|u_1|}{|u_0|}, \frac{|u_2|}{|u_0|}, \ldots, \frac{|u_{N-1}|}{|u_0|}$$

4. Comparison:

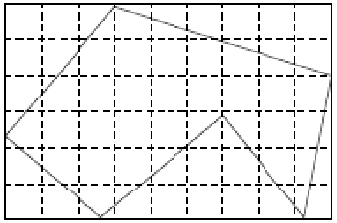
$$d = \left(\sum_{n=0}^{N_c} \left|f_I^n - f_J^n\right|^2\right)^{\frac{1}{2}}$$

# Region-based features



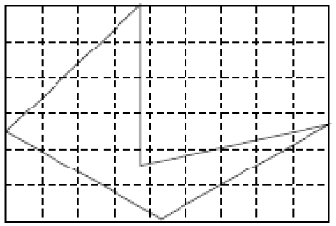
# Grid-method

А



А: 001111000 011111111 111111111 111111111 111110111 011100011

Б

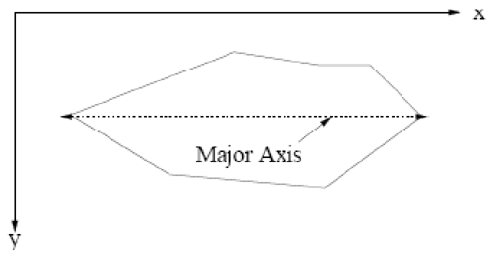


Б: 001100000 011100000 111100000 111101111 111111110 001111000

## Invariance:

Normalization by major axe:

- direction;
- scale;
- position.



# Moment invariants

The moment of order (p+q) for a two-dimension continuous function:

$$m_{pq} = \iint x^p y^q f(x, y)\,dxdy$$

Central moments for f(x,y) – discrete image:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y), \quad \bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

Feature vector:

Seven scale, translation and rotation invariant moments were derived based on central normalized moments of order p + q = 2; 3.

# Lecture 3: Outline

- Texture features
  - Statistical
  - Spectral
  - Comparison
- Shape features
  - Boundary based
  - Region based
  - Comparison
- Fusion methods

# Shape features comparison

Table 1. Average retrieval efficiency values for different methods. $T$ is the short list size of retrieved images

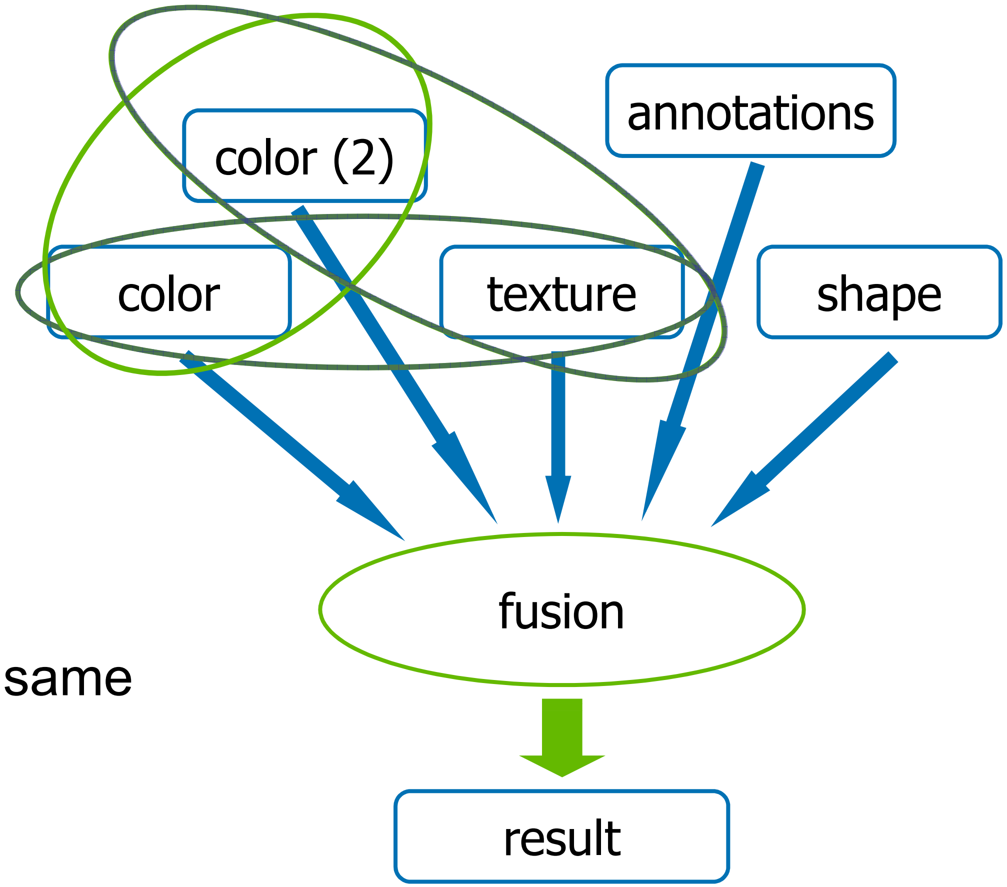
| Methods | T=5 | T=10 | T=15 | T=20 |
|---|---|---|---|---|
| Reduced chain code | 55.1% | 47.6% | 50.0% | 60.6% |
| Fourier descriptors (FD) | 72.2% | 76.9% | 75.9% | 74.9% |
| UNL features | 81.3% | 79.9% | 83.7% | 89.3% |
| Moment invariants (MI) | 84.7% | 86.3% | 86.8% | 87.7% |
| Zernike moments | 66.9% | 66.5% | 70.4% | 78.2% |
| Pseudo-Zernike moments | 66.9% | 66.5% | 70.4% | 78.2% |
| MI and FD | 93.8% | 87.3% | 87.1% | 89.6% |
| MI and UNL | 93.3% | 89.2% | 89.3% | 91.1% |

Mehtre B. M., Kankanhalli M. S., Lee W. F. Shape measures for content based image retrieval: a comparison. Inf. Processing and Management, vol. 33, No. 3, pages 319-337, 1997.

# Lecture 3: Outline

- Texture features
  - Statistical
  - Spectral
  - Comparison
- Shape features
  - Boundary based
  - Region based
  - Comparison
- Fusion methods

# Data fusion in CBIR




- Combined search
  (different features)
- Refine search results
  (different algorithms for the same feature)
- Supplement search results
  (different datasets)



LABS hp

# Fusion of retrieval result sets

## Fusion of weighted lists with ranked elements:

$$\left.\begin{array}{ll} \omega_1 & \boxed{(x^1_1, r^1_1), (x^1_2, r^1_2), \dots, (x^1_n, r^1_n)} \\ \omega_2 & \boxed{(x^2_1, r^2_1), (x^2_2, r^2_2), \dots, (x^2_k, r^2_n)} \\ & \dots \\ \omega_m & \boxed{(x^m_1, r^m_1), (x^m_2, r^m_2), \dots, (x^m_l, r^m_l)} \end{array}\right\} \quad ?$$

### Existing approaches in text retrieval:

- CombMax, CombMin, CombSum
- CombAVG
- CombMNZ = CombSUM * number of nonzero similarities
- ProbFuse
- HSC3D

# Fusion function: properties

1) Depend on both weight and rank
2) Symmetric
3) Monotony by weight and rank
4) MinMax condition /CombMin, CombMax, CombAVG/:

$$\min\{r_x^{(\alpha_1)}, r_x^{(\alpha_2)}, \ldots, r_x^{(\alpha_N)}\} \le r_x^{(0)} \le \max\{r_x^{(\alpha_1)}, r_x^{(\alpha_2)}, \ldots, r_x^{(\alpha_N)}\}$$

5) Additional property – "conic" property: non-linear dependency from weight and rank; high weight, high rank – influence bigger to the result than several inputs with low weight, low rank.

$$\forall\sigma \quad \exists\epsilon_1, \epsilon_2 : |1 - r_x^{(\alpha_1)}| < \epsilon_1 \wedge |1 - w^{(\alpha_1)}| < \epsilon_2 \quad \Longrightarrow \quad |1 - r_x^{(0)}| < \sigma$$

for merging any N rank lists

# Weighted Total with Gravitation Function

CombAVG as a base, but use gravitation function instead of weight:

$$r_x^{(0)} = \frac{\sum_i g\left(r_x^{(\alpha_i)}, w^{(\alpha_i)}\right) * r_x^{(\alpha_i)}}{\sum_i g\left(r_x^{(\alpha_i)}, w^{(\alpha_i)}\right)}$$

where

$$g\left(r_x^{(\alpha_i)}, w^{(\alpha_i)}\right) \; = \; \left(w^{(\alpha_i)}\right)^2 * \left(r_x^{(\alpha_i)} + \frac{1}{12}\right)^4$$

# WTGF: some results

- Experiments on search in semi annotated collections and of color and texture fusion (compare with CombMNZ)
- WTGF is good when:
  - There are a lot of viewpoints.
  - Viewpoints are very different (different opinions regarding the rank of the same element).
  - Viewpoints have different reliability.
- CombMNZ is good when:
  - Viewpoints have the same reliability.
  - Viewpoints have similar opinions.

Natalia Vassilieva, Alexander Dolnik, Ilya Markov. Image Retrieval. Combining multiple search methods' results. In "Internet-mathematics" Collection, 46—55, 2007.

# Adaptive merge: color and texture

$Dist(I, Q) = \alpha * C(I, Q) + (1 - \alpha) * T(I, Q)$,

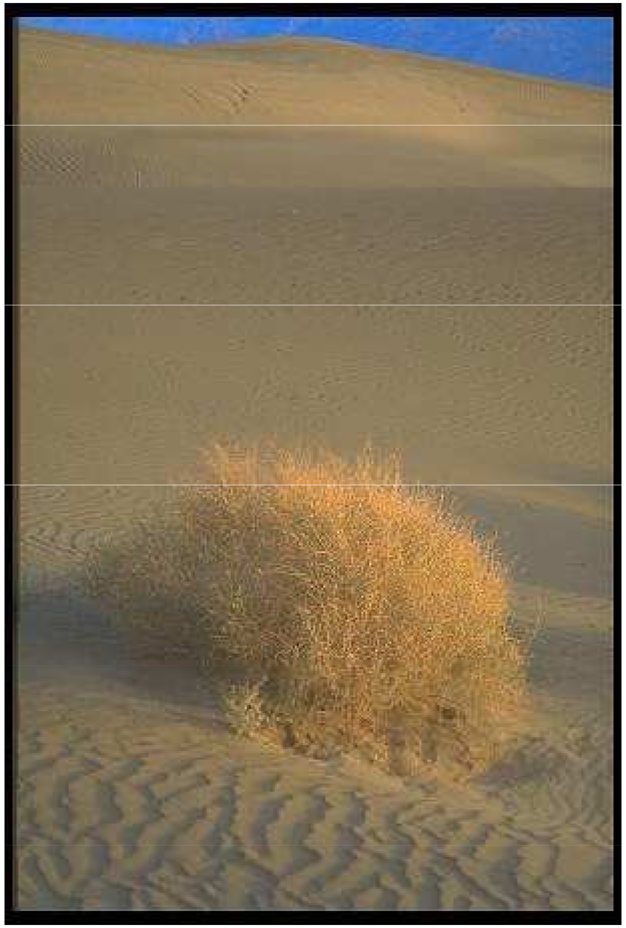
$C(I, Q)$ – color distance between I and Q;
$T(I, Q)$ – texture distance between I and Q;
$0 \leq \alpha \leq 1$

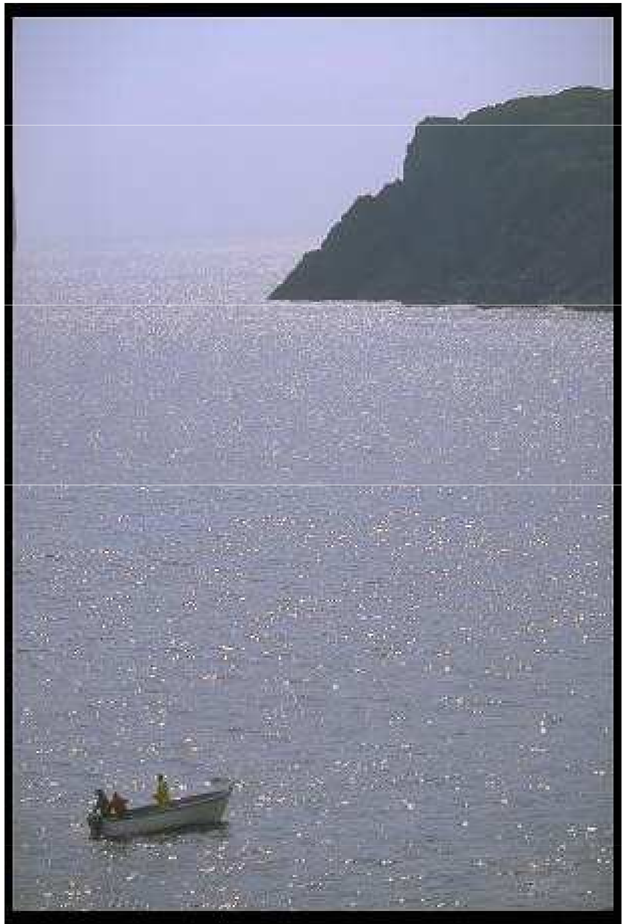
## Hypothesis:

Optimal α depends on features of query Q. It is possible to distinguish common features for images that have the same "best" α.

Ilya Markov, Natalia Vassilieva, Alexander Yaremchuk. Image retrieval. Optimal weights for color and texture fusion based on query object. In Proceedings of the Ninth National Russian Research Conference RCDL'2007

# Example: texture search





# Example: color search

# Mixed metrics: semantic groups

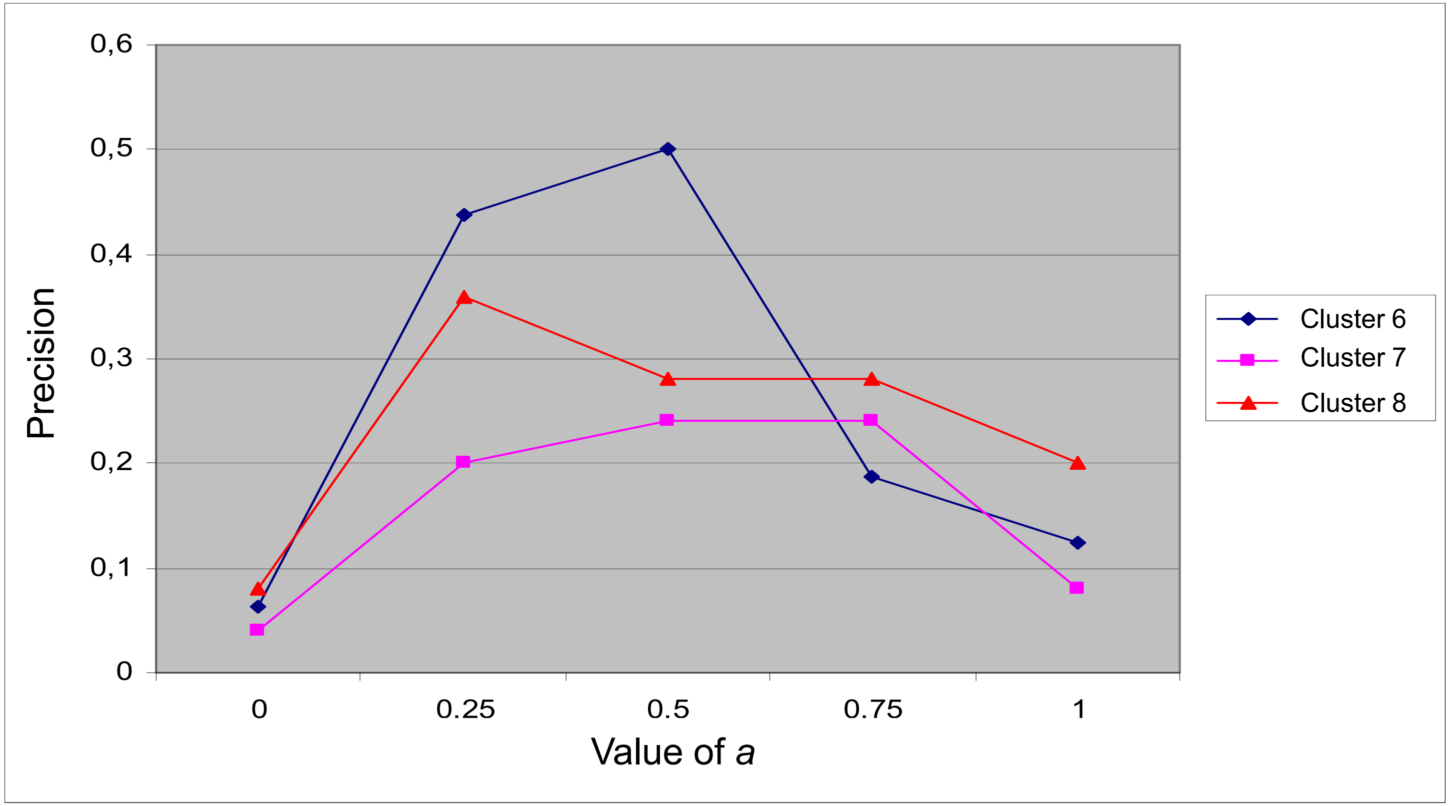
$a_1$

$a_2$

$a_3$

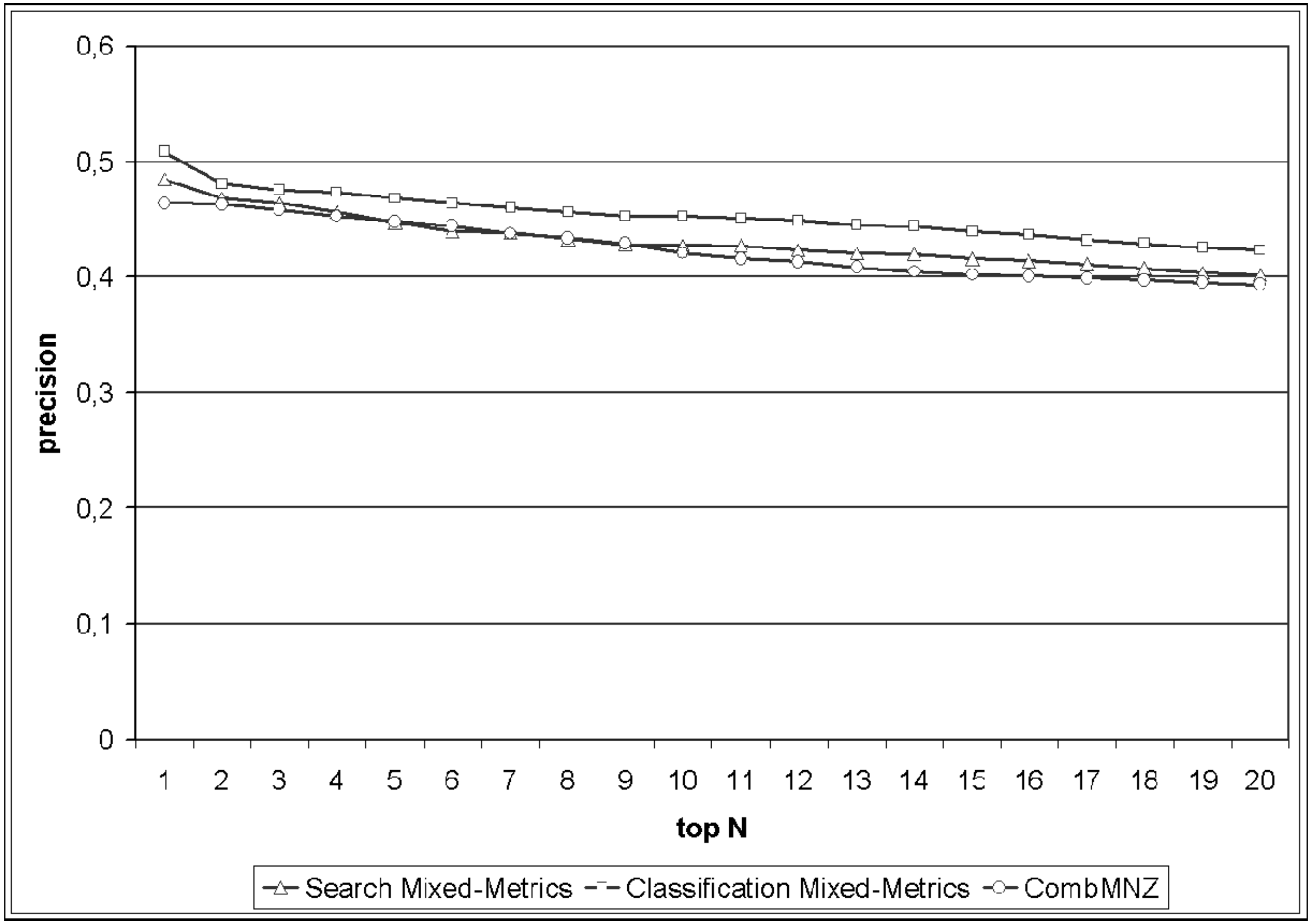
# Experimental results 1

- It is possible to select the best value of *a*

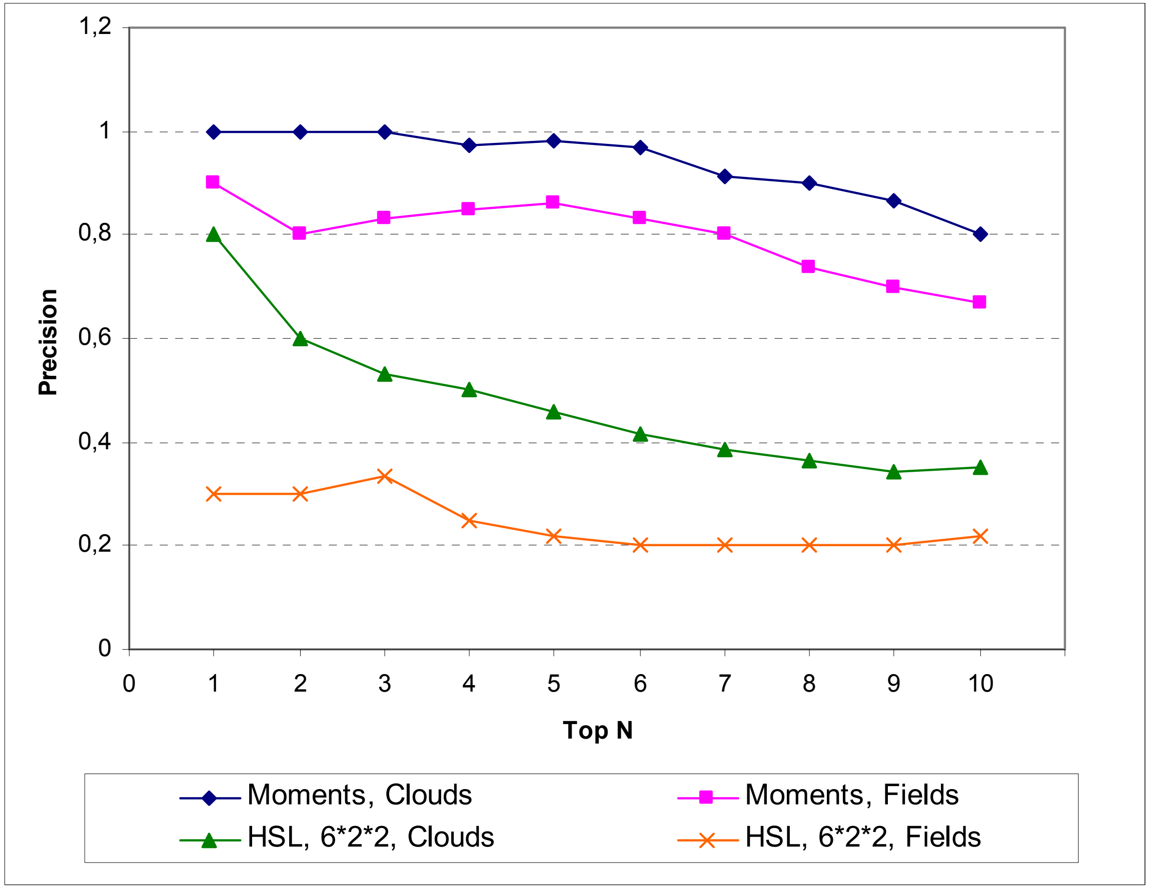


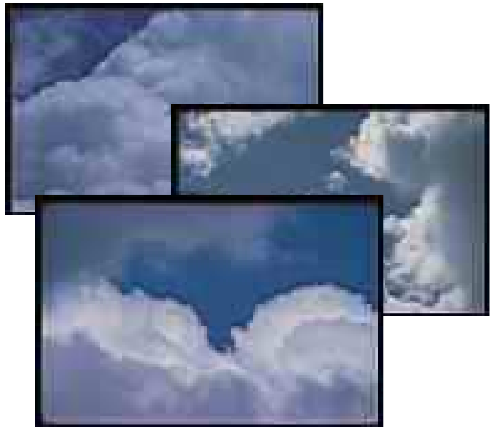
# Experimental results 2

- Adaptive mixed-metrics increase precision

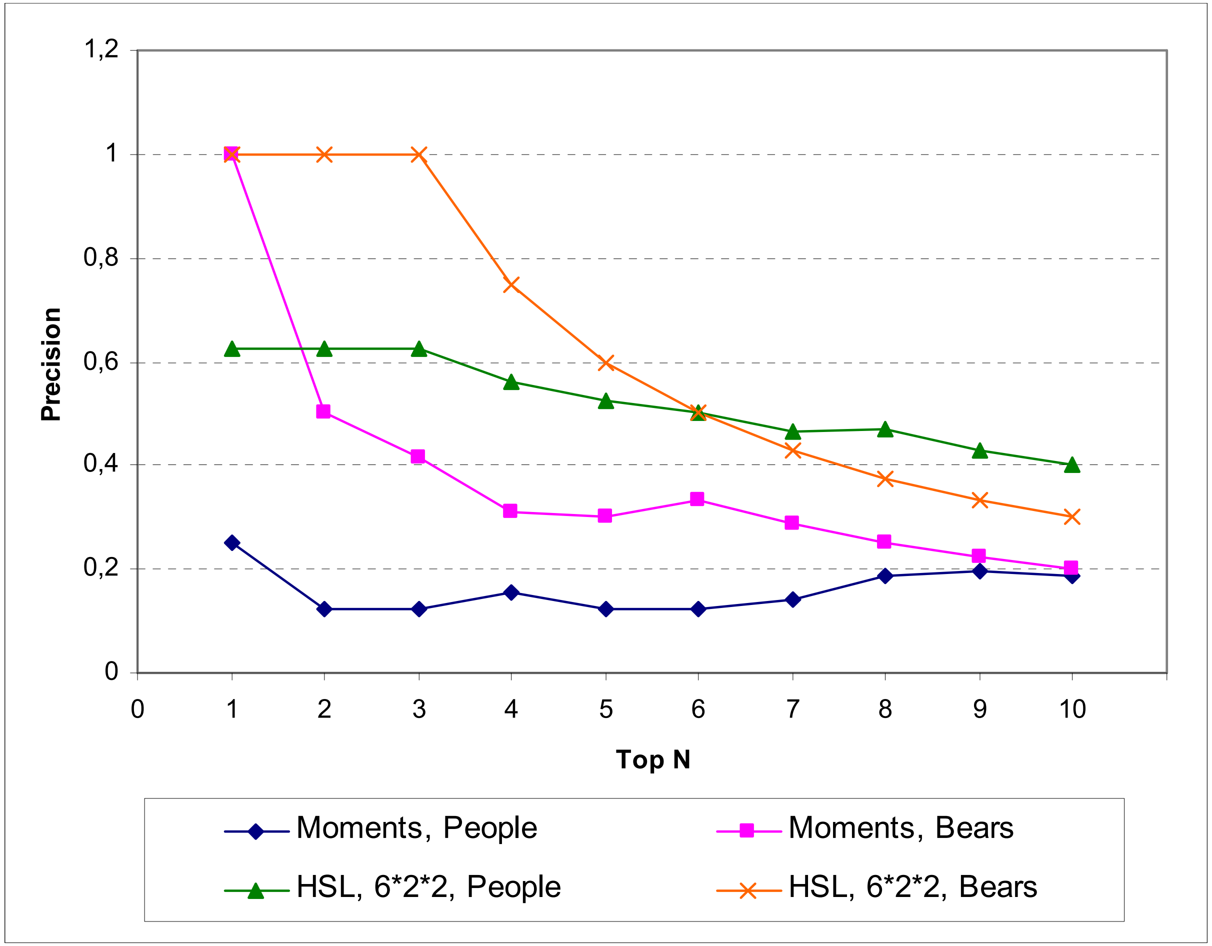


# Adaptive merge: color and color



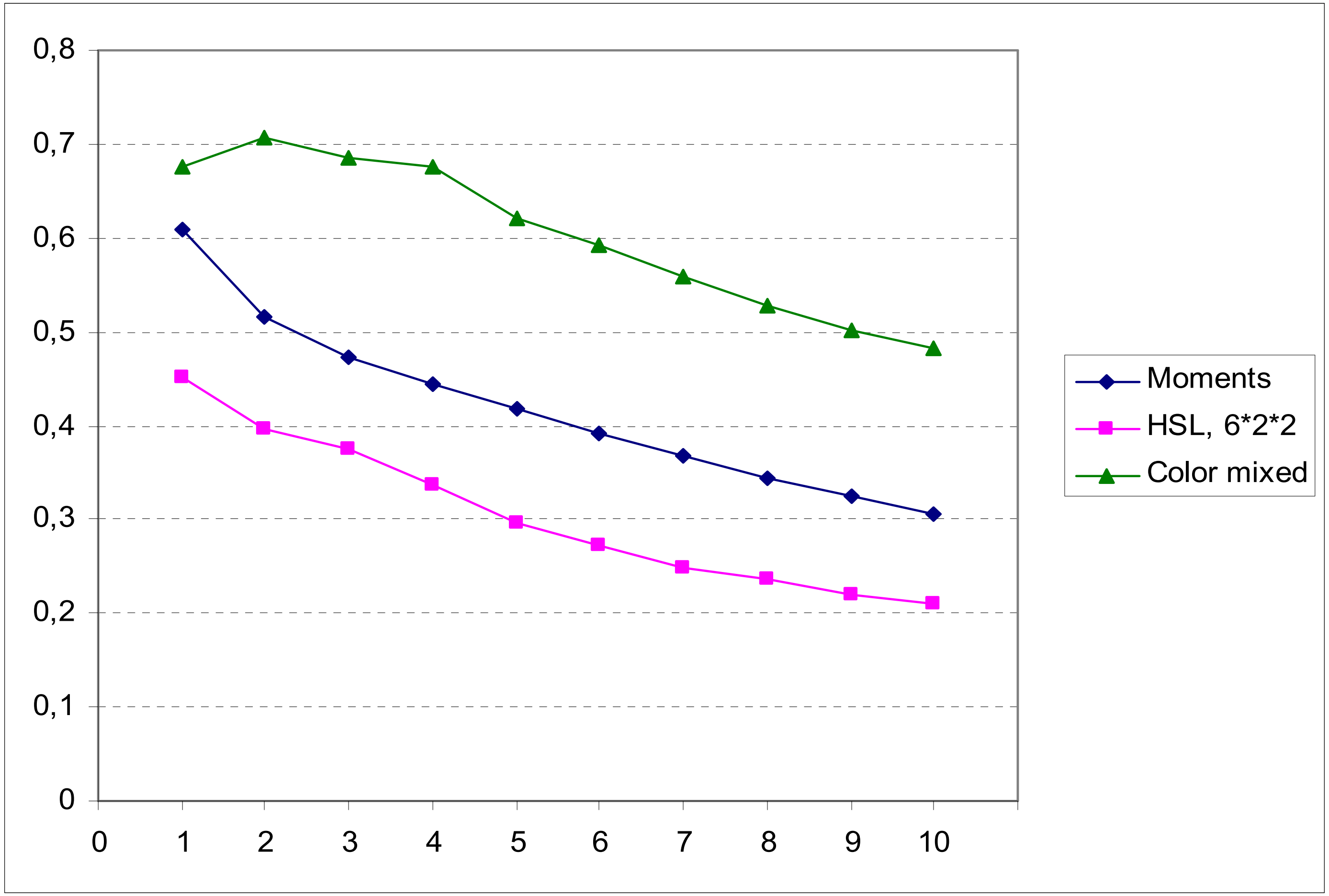



# Adaptive merge: color and color



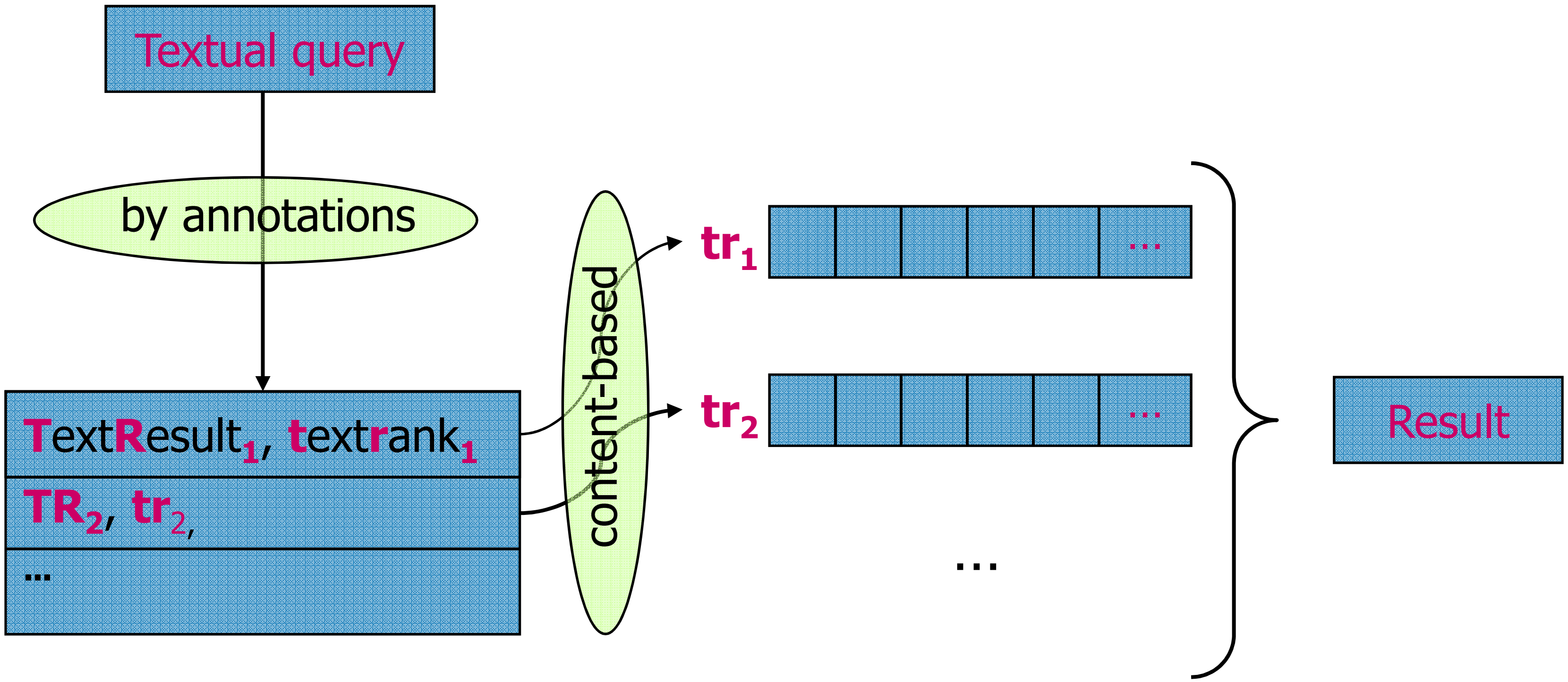

# Color fusion

## CombMNZ (Moments + HSL histogram)

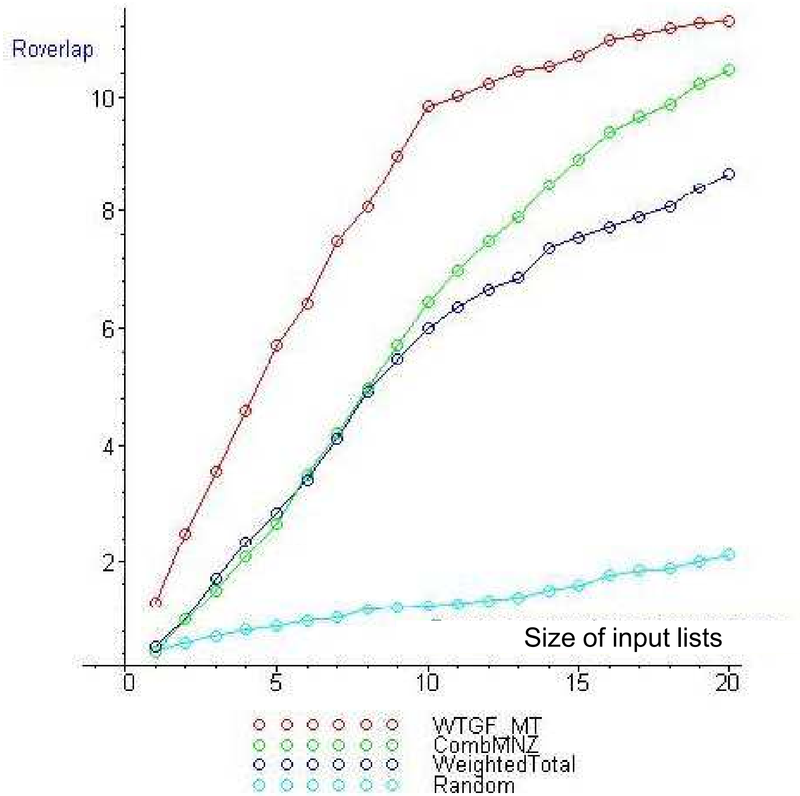


# Ranked lists fusion: application area

- Search by textual query in semi annotated image collection



# Retrieve by text: fusion results




$$R_{overlap}(x) = \frac{M \cdot R^{(0)}(x)}{\sum_{i}^{M} R^{(\alpha_i)}(x)}$$

$$N_{overlap}(x) = \frac{M \cdot N^{(0)}(x)}{\sum_{i}^{M} N^{(\alpha_i)}(x)}$$

# Lecture 3: Resume

- Texture features
  - Statistics (Haralik's co-occurance matrices, Tamura features)
  - Spectral features are more efficient (Gabor filters, ICA filters)

- Shape features
  - Boundary-based (Fourier descriptors)
  - Region-based (Moment invariants)

- Fusion methods
  - Are very important
  - Need to choose based on a particular fusion task

# Lecture 3: Bibliography


- Haralick R. M., Shanmugam K., Dienstein I. Textural features for image classification. In IEEE Transactions on Systems, Man and Cybernetics, vol. 3(6), pp. 610 – 621, Nov. 1973.
- Tamura H., Mori S., Yamawaki T. Textural features corresponding to visual perception. In IEEE Transactions on Systems, Man and Cybernetics, vol. 8, pp. 460 – 472, 1978.
- Tuceryan M., Jain A. K. Texture analysis. The Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co., 1998.
- Tuceryan M., Jain A. Texture segmentation using Voronoi polygons. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, No 2, pp. 211 – 216, February 1990.
- Walker R., Jackway P., Longstaff I. D. Improving co-occurrence matrix feature discrimination. In Proc. of DICTA'95, The 3rd Conference on Digital Image Computing: Techniques and Applications, pp. 643 – 648, 6-8 December, 1995.

# Lecture 3: Bibliography


- Li B., Ma S. D. On the relation between region and contour representation. In Proc. of the IEEE International Conference on Pattern Recognition, vol. 1, pp. 352 – 355, 1994.
- Lin T.-W., Chou Y.-F. A Comparative Study of Zernike Moments for Image Retrieval. In Proc. of 16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003), pp. 621 – 629, 2003.
- Loncaric S. A survey of shape analysis techniques. In Pattern Recognition, vol. 31(8), pp. 983 – 1001, 1998.
- Luren Y., Fritz A. Fast computation of invariant geometric moments: A new method giving correct results. In Proc. of IEEE International Conference on Image Processing, 1994.
- Zakaria M. F., Vroomen L. J., Zsombor-Murray P. J. A., van Kessel J. M. H. M. Fast algorithm for the computation of moment invariants. In Pattern Recognition, vol. 20(6), pp. 639 – 643, 1987.
- Zernike polynomials. Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Zernike_polynomials
- Zhang D., Lu G. Shape-based image retrieval using generic Fourier descriptor. In Signal Processing: Image Communication, vol. 17, pp. 825 – 848, 2002.
- Zhang D., Lu G. A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures. In Proc. of the International Conference on Multimedia, 2001.