Content-based Image Retrieval

Tutorial Image Retrieval

Thomas Deselaers, Henning Müller

Outline

- What is CBIR
- Approaches
  - Continuous approach
  - Discrete approach
- Features for content-based image retrieval
  - Global
  - Local
- Relationships between different approaches
- Available resources

Content-based Image Retrieval

- Two main questions
  - How are the images represented? ➔ features
  - How are the image descriptors compared ➔ similarity measures

- Two main views on the concept of a `feature':
  - Features are numerical values computed from each image
    - View connected to image classification
    - Ideas and methods from classification and machine learning
  - Features are image properties that are present or absent
    - View connected to textual information retrieval
    - Ideas and methods from text retrieval

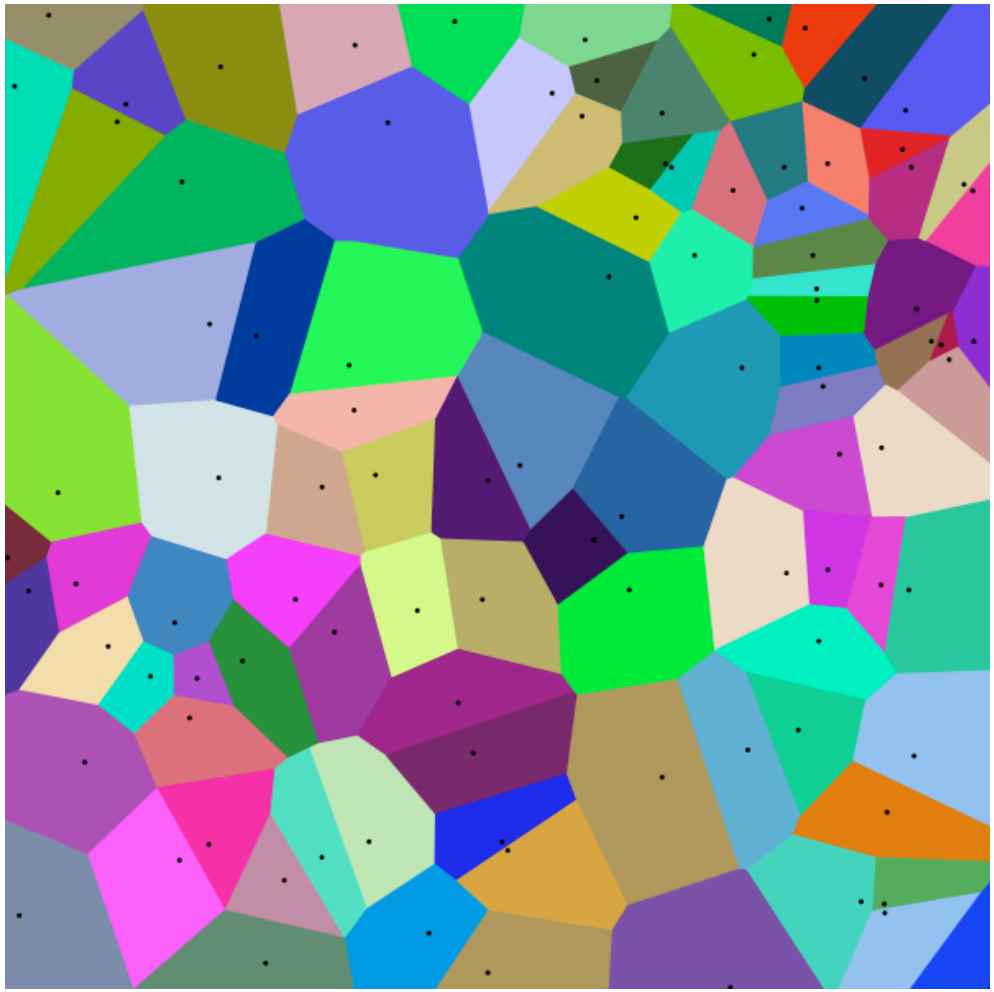
Continuous Approach

- Inspired by nearest-neighbour classification
- Each image is represented by a feature vector
- Feature vectors are compared using distance functions
- Images similar to a query image are assumed to be relevant
- Normally: query-by-visual example(s)

Nearest Neighbour

- Voronoi Diagram in a 2D space
- A set of points and the corresponding Voronoi cells
- A Voronoi cell is the area where a point's nearest neighbour is the seed of the cell

Source: Wikipedia

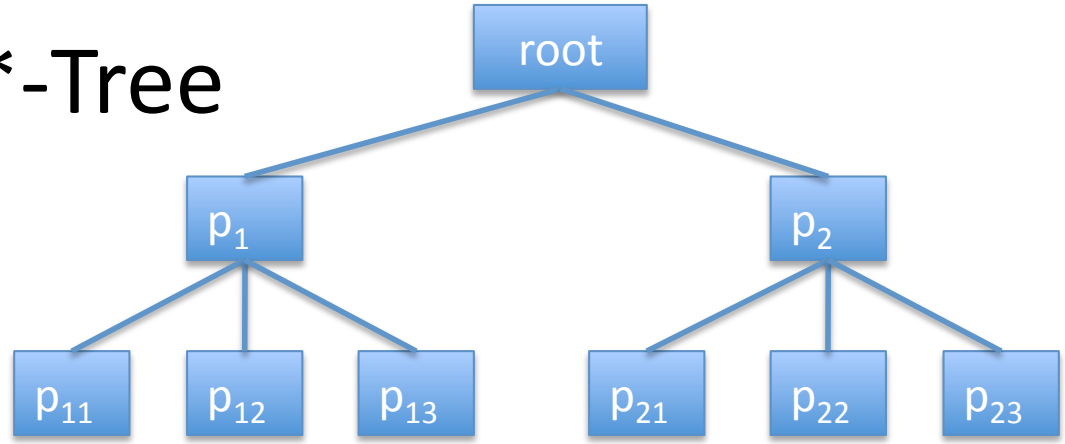


Nearest Neighbour

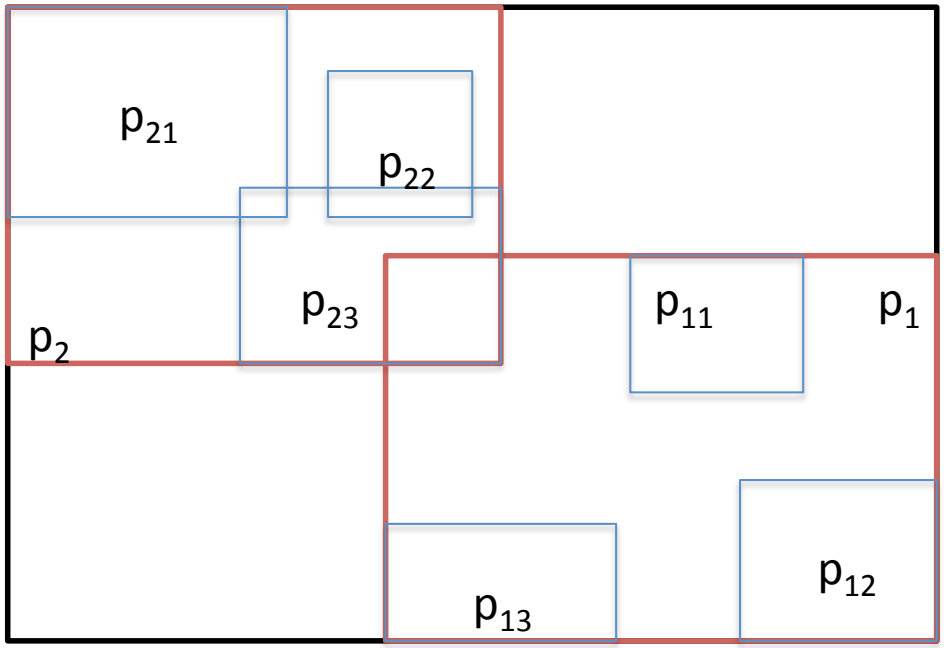
- Standard nearest neighbour search

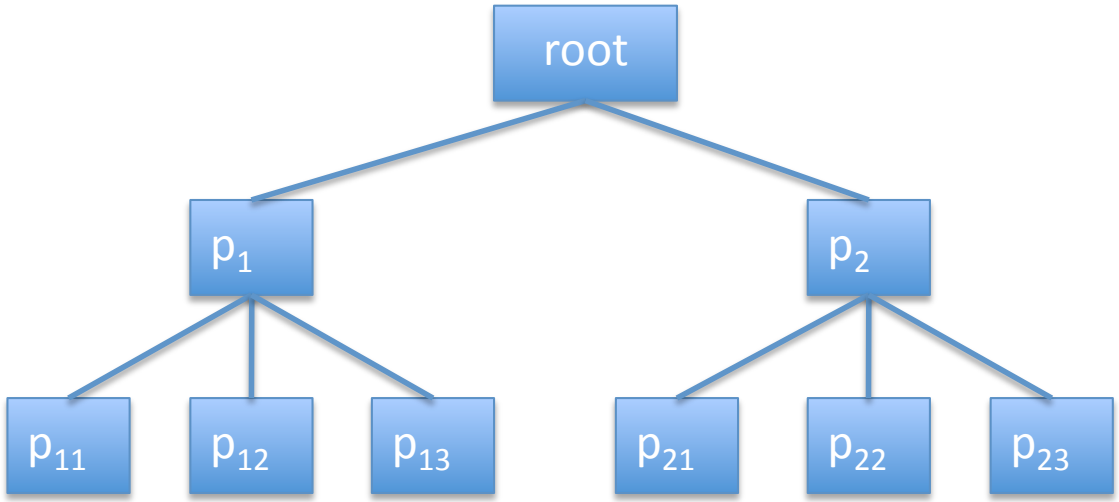
```
X[1] … X[N]: data set
Q: query
mindist=∞
bestN=-1

For n=1:N:
   d=dist(q,X[n])
   if d<mindist:
      mindist=d
      bestN=n
return bestN
```

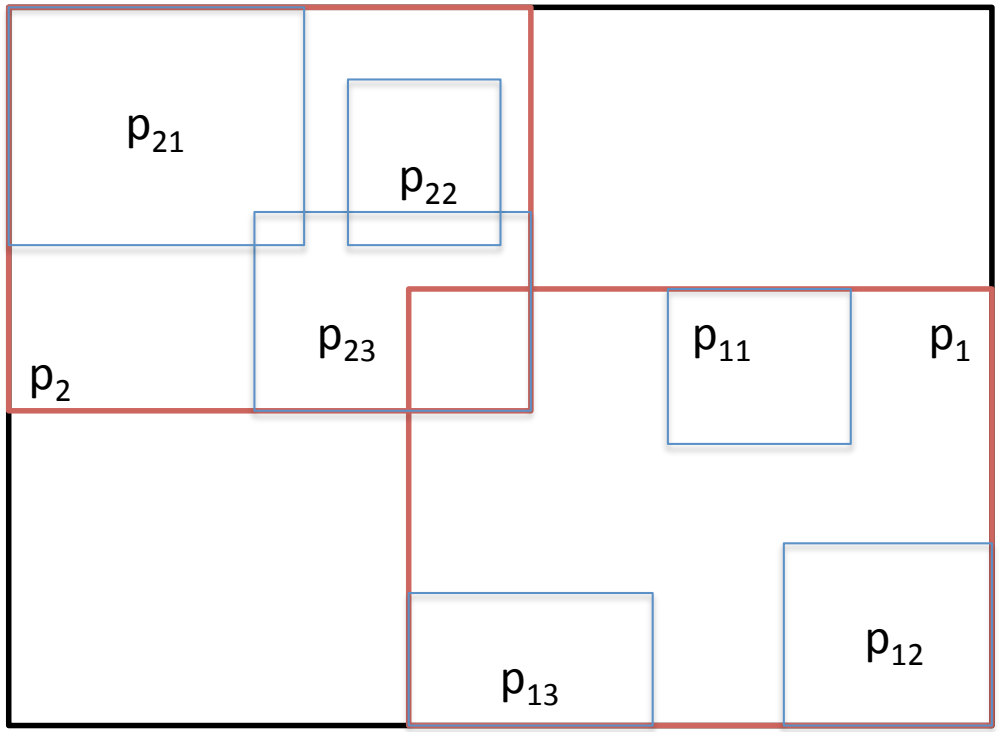
R*-Tree



- Data structure for multi-dimensional data
- Save data in a tree structure
- Directory nodes (blue)
- Data nodes (red)
- Save minimum bounding rectangles in directory nodes

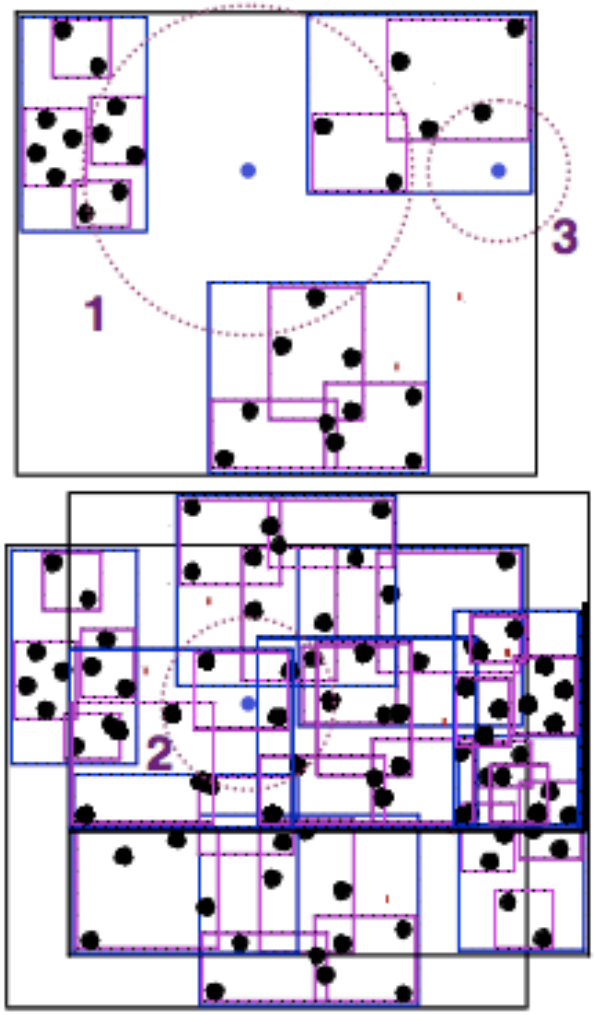



root
p_1
p_2
p_{11}
p_{12}
p_{13}
p_{21}
p_{22}
p_{23}


p_{21}
p_{22}
p_{23}
p_2
p_{11}
p_1
p_{13}
p_{12}

Index Structures: Complexity

- Linear search
  - Complexity O(n/C), very small overhead
- Bad situation for index structures
  - Large range
  - Strongly overlapping regions
  - Few regions are not accessed
  - Common with high dimensionalities
  - Complexity O(n/C), high overhead
- Good situation
  - Small range
  - Small overlaps
  - Many regions do not have to be read
  - Complexity $O(\log_C(n))$

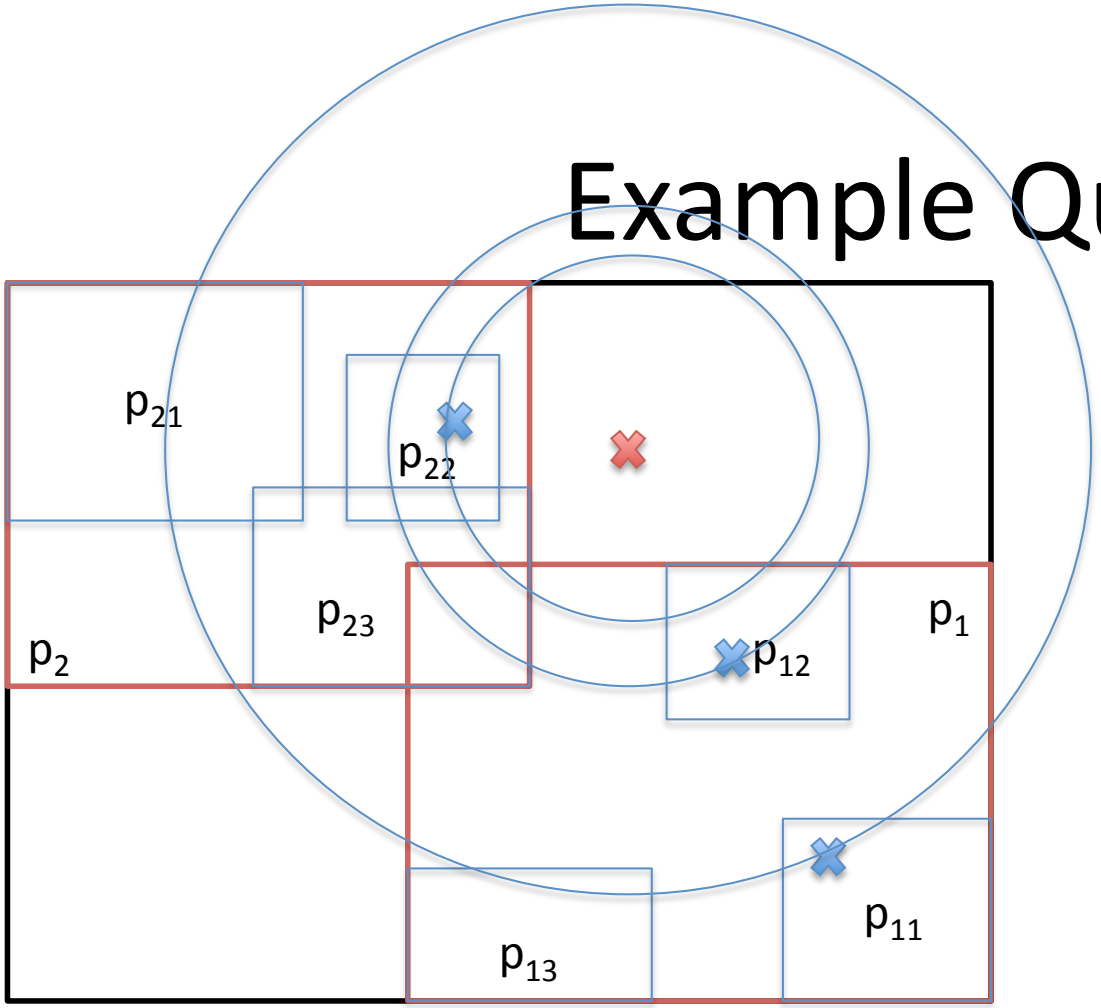


Nearest Neighbour Search with Index

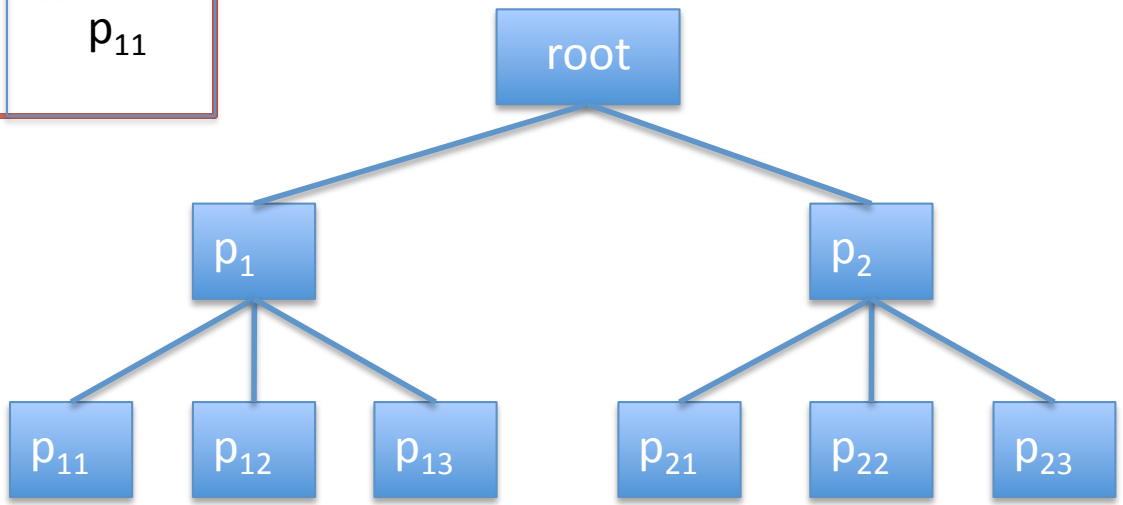
```
Init: resultdist=∞
Function SimpleNNQuery(Point q, Address: page):
    page.load()
    if isDataPage(page):
       for x in page.points:
           d=dist(q,x)
           if d<resultdist:
               resultdist=d
               result=x
    else:
       for p in page.childPages:
           if MINDIST(q,p)<resultdist:
               SimpleNNQuery(q,p);
```

- First path finds arbitrary point
- Search space only slowly reduced
- Many pages unnecessarily read

Example Query



- If the search had started with p_2 no page from p_1 would have been read
- Clearly non-optimal



Nearest Neighbour Search with Index

Best-First Search

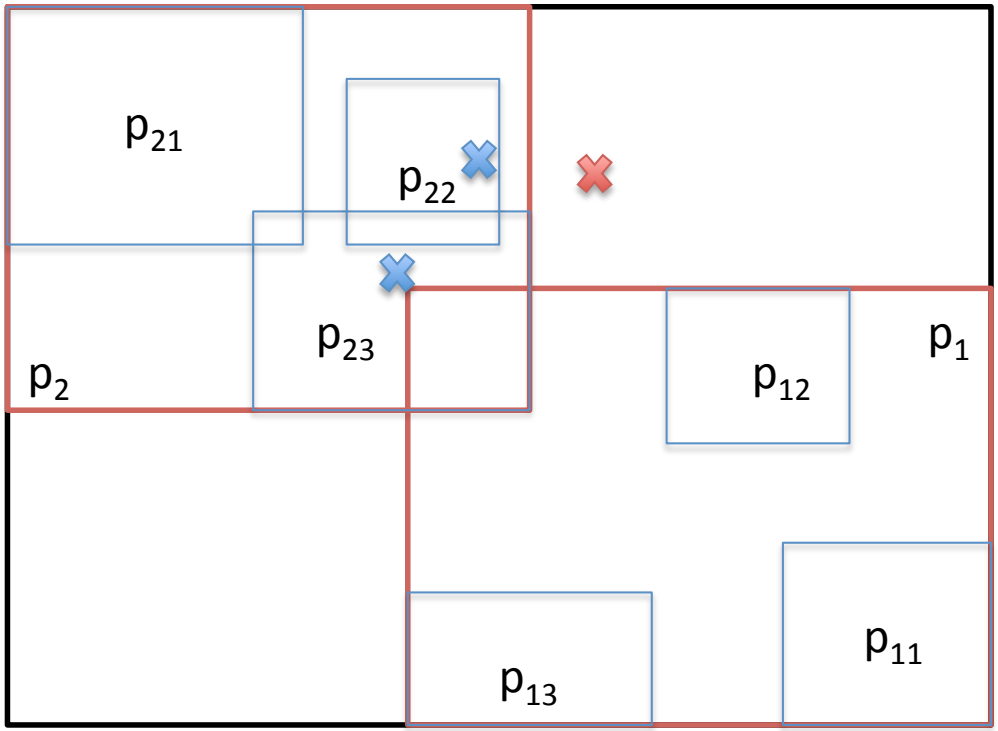
- Avoid recursion
- Instead use priority queue APL (active page list)
  - List which contains directory pages to be processed sorted by priority

- Definition: a page p is active, if and only if
  - p not yet processed
  - Parent of p processed
  - Minimum distance between p and q < current best distance
- Initialisation: APL = [root]
- In each step, process best page from APL
  - Data pages: as before
  - Directory pages: check minimum distance to query

Best First Nearest Neighbour Search with Index

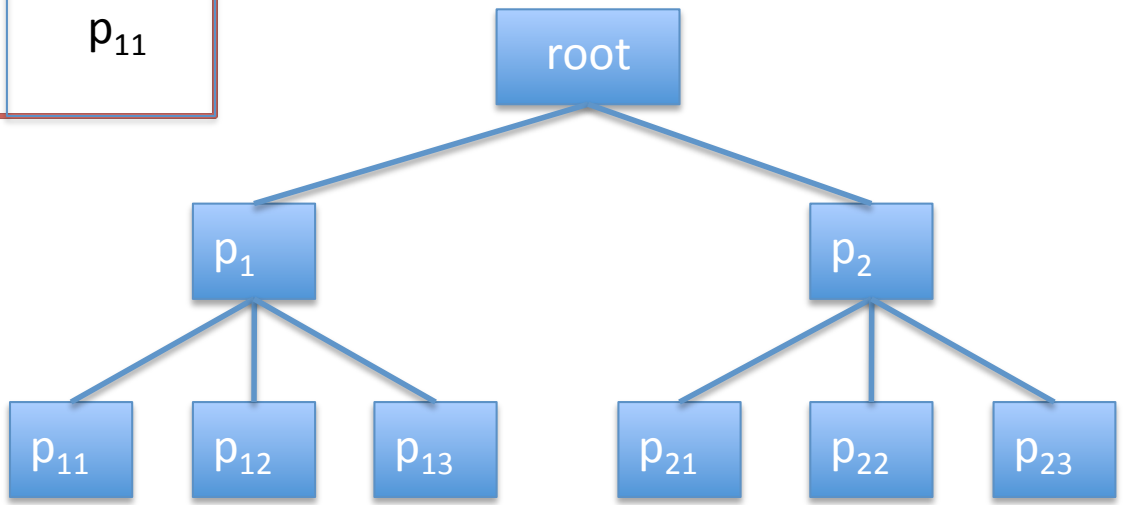
```
Init: apl = [(0.0,root)] // sorted by dist
      resultdist=∞
While apl.notEmpty() and apl[0].dist<resultdist:
    page=apl[0].load()
    delete(apl[0])
    if isDataPage(page):
       for x in page.points:
           d=dist(q,x)
           if d<resultdist:
               resultdist=d
               result=x
    else:
       for p in page.childPages:
           h=MINDIST(q,p)
           if h<resultdist:
               apl.insert((h,p))
```

Example Query with Best First Search



APL: [(6.8, p_{11}), (7.2, p_{12}), (8.2, p_{13})] [8.2, p_{21})]

result dist: 7.0



Best First Search is optimal

Here: a draft of the proof

1. Completeness: It will find the correct NN of a query
   - Every correct algorithm has to access all pages that intersect with the NN sphere of q
   - These pages have MINDIST < resultdist
2. It accesses pages in ascending order from the query
   - The APL is sorted by MINDIST and the algorithm will terminate once it is impossible to find any point closer to q than the current result
3. It will not access a single page with MINDIST larger than the true NN distance
   - Child pages cannot have a MINDIST smaller than its parents

Continuous Approach

- Simple Case:
- Database: $\{x_1, \ldots x_n, \ldots, x_N\}$
- Query: q
- Distance between q and x_n $d(q, x_n)$
- Sort images $(x_{n_1} \ldots, x_{n_i}, \ldots x_{n_N})$
- Such that $d(q, x_{n_i}) \leq d(q, x_{n_{i+1}})$
- Holds for all x_n and $x_{n_{i+1}}$

Continuous Approach

- How to use multiple features
- Each image x_n is represented by M descriptors x_{nm}
- Calculate score instead of distance to allow for different weights

$$S_q(x_n) = \exp\left(-\sum_{m=1\ldots M} w_m \frac{d_m(x_{nm}, q_m)}{\frac{1}{N}\sum_{n=1\ldots N} d_m(x_{nm}, q_m)}\right)$$

Discrete Approach

- Inspired by textual information retrieval
- Each image is represented by a set of binary features (features may be present (possibly multiple times) or absent)
- Feature is either present or absent
  - Similar to words being absent or present in a document
- Images containing the same (informative) features are assumed to be relevant to a query
- Example: GIFT – GNU Image Finding Tool

Discrete Approach

- GIFT uses TF-IDF (text frequency/inverse document frequency) ranking
  - Reduce the impact of features which occur frequently in the data (comparable to “the” in texts)
- TF: frequency a feature *i* has in a document d_j

$$tf(i, d_j) = \frac{n(i, d_j)}{\sum_k n(k, d_j)}$$

- IDF: measures importance of a term

$$idf(i) = \log\left(\frac{|\mathcal{B}|}{|\{d_j : n(i, j) > 0\}|}\right)$$

Discrete Approach

- *tf* captures how often a feature occurs in a document
  - Features that occur often in a document describe this document well
- *idf* captures how relevant a feature is
  - Features that occur rarely in the full database are important
- Important are those features which are often in one image, but seldom over the full data set
  - Images which share seldom features are relevant with respect to each other

Discrete Approach in GIFT

- In GIFT, 4 different feature sets are considered
  - Global colour
  - Local colour
  - Global texture
  - Local texture
- For all local features a tf-idf-like score is calculated and these are fused as a weighted sum
  - Global features are compared with a histogram intersection
- For many cases, the discrete and the continuous approach can be simulated in the respective other
- In GIFT
  - Images have about 1,500 to 2,000 features
  - Similar images share about 400-500

Discrete Approach: Inverted Files

- Store a mapping from content to location
  - E.g. for each feature a list of images that contain this feature

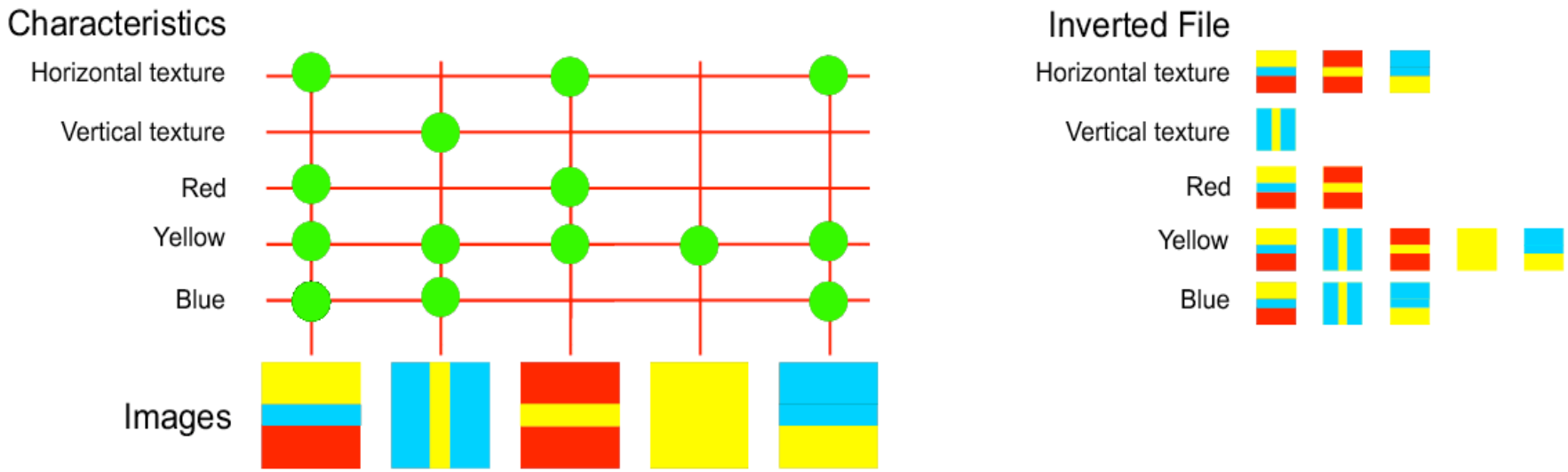
- Allow for efficient searches even for huge amounts of images
  - *idf* for each feature can be pre-calculated
  - *tf* for each feature is stored in the files
  - Allows for searching without accessing the images

- In the continuous approach searching for neighbours is linear to the amount of images, here it is at most linear to the number of features in an image
  - In practice, the features with high impact are processed first and the other features have less influence (Zipf's Law). Therefore, in practice, this approach is very fast

The inverted file

Inverted Files

- Access feature by feature instead of image by image
- Extremely fast access for rare features
- Efficient for sparsely populated spaces



Visual Properties of Images

- Colour
- Texture
- Shapes
- Image parts
- Complete image

- Meta data
- Textual labels/captions/annotations

Features for Content-based Image Retrieval

- Global Descriptors
  - Colour histograms
  - Texture Features
  - Shape Features

- Local Descriptors
  - Direct approach
  - Patch-histograms / bag-of-visual words
  - SIFT features

Global Descriptors

- Capture a property of the image with few values
- Describing the image in its entirety
- E.g.
  - Which colours occur in the image?
  - Is the image of high contrast?
  - Is the image bright/dark?

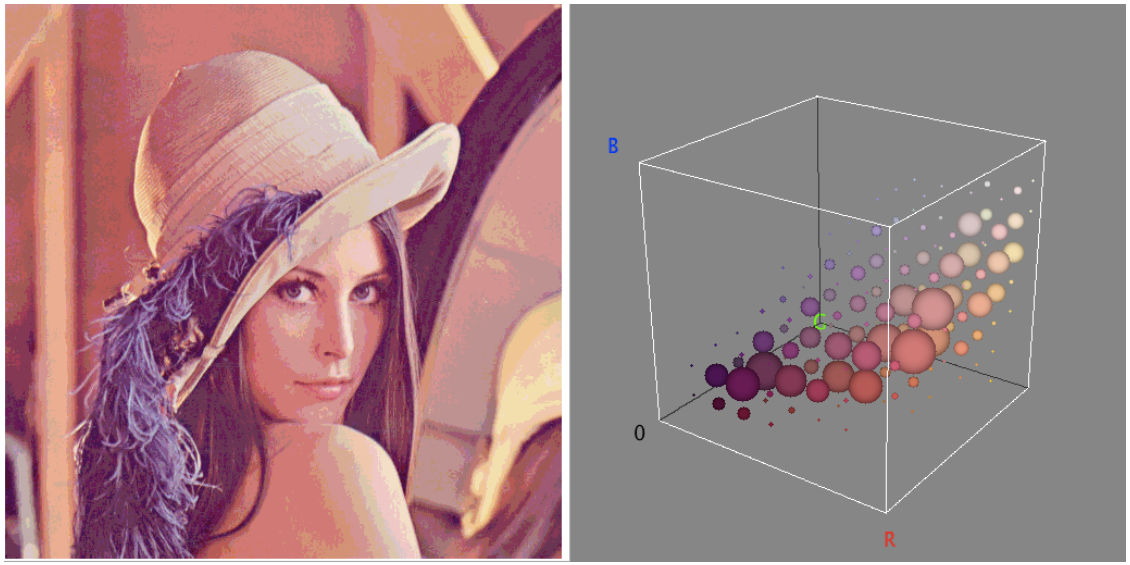
Colour histograms

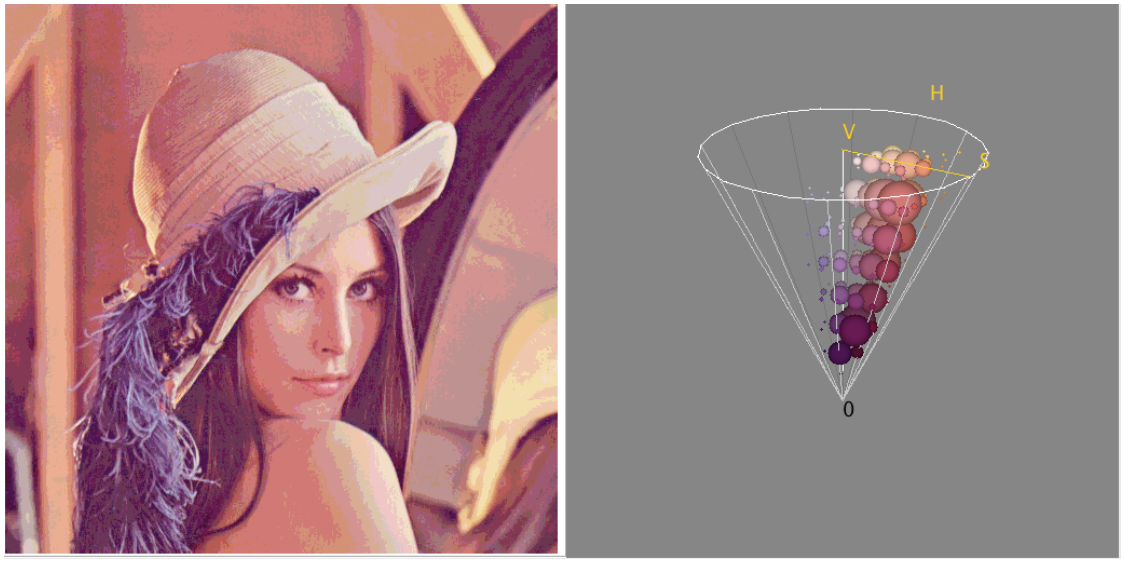
- Describe the distribution of colours in an image
- Discard spatial information

1. Quantise colour space
2. Count which colour occurs how often

Colour Histograms: Example

RGB colour space



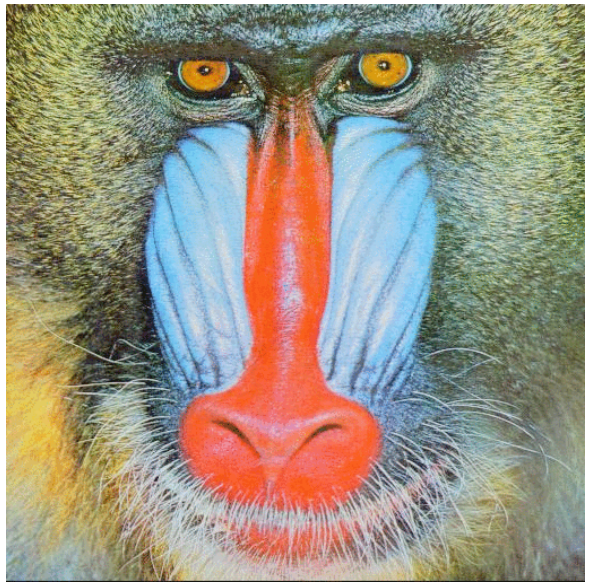


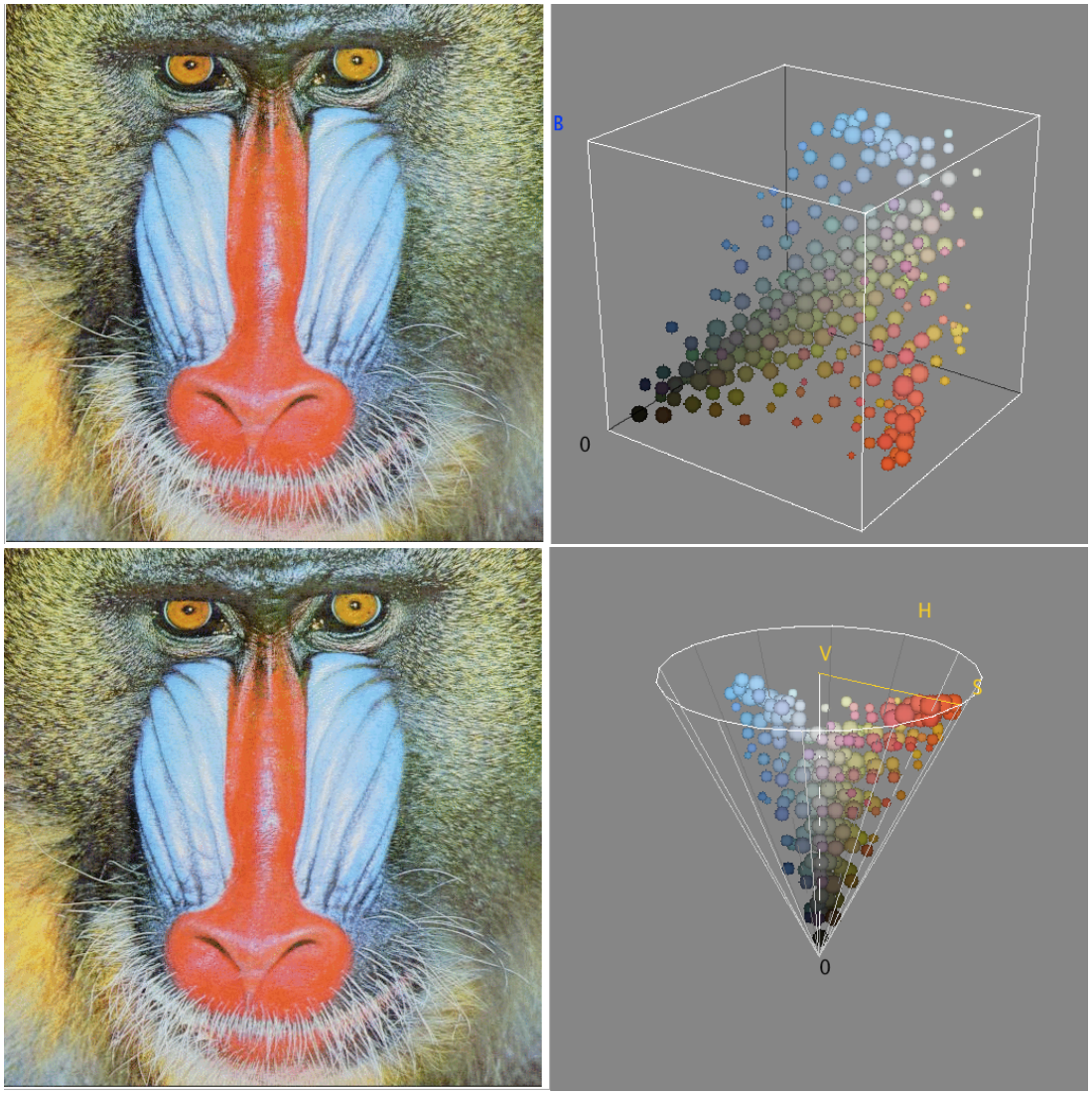
HSV colour space

Visualisation done with 3D Color Inspector: http://rsbweb.nih.giv/ij/plugins/color-inspector.html

Colour Histograms: Example

RGB colour space





HSV colour space

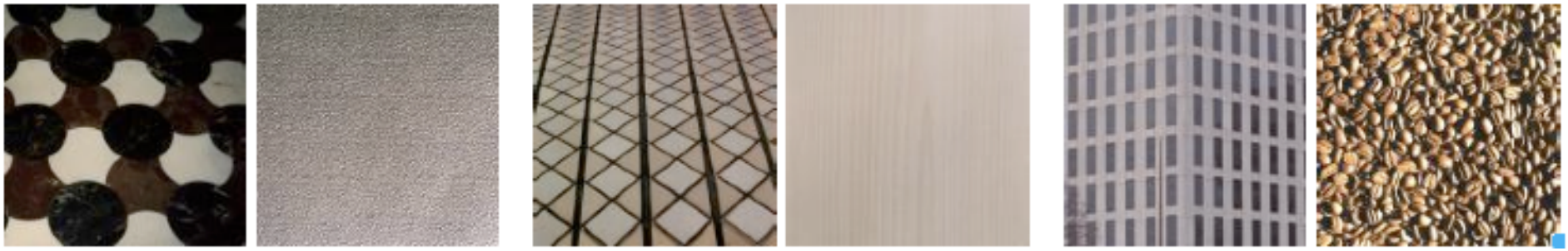
Visualisation done with 3D Color Inspector: http://rsbweb.nih.giv/ij/plugins/color-inspector.html

Texture Features

"**Texture** refers to the properties held and sensations caused by the external surface of objects received through the sense of touch."

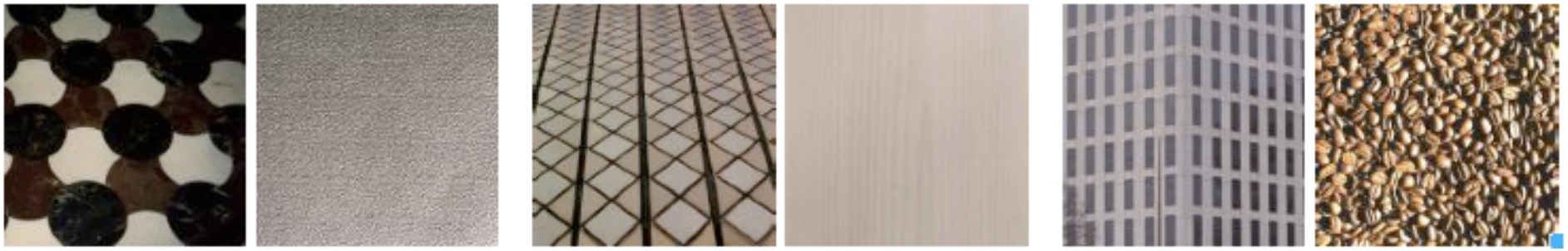
Texture in image processing:

- Different definitions
- Different representations



Tamura Features

- Proposed by Tamura [1978]
  - Features corresponding to human perception
  - Examined 6 different features, found 3 to correspond strongly to human perception
  - **Coarseness** – coarse vs. fine
  - **Contrast** – high vs. low
  - **Directionality** – directional vs. non-directional
  - Line-likeness – line-like vs. non-line-like
  - Regularity – regular vs. irregular
  - Roughness – rough vs. smooth

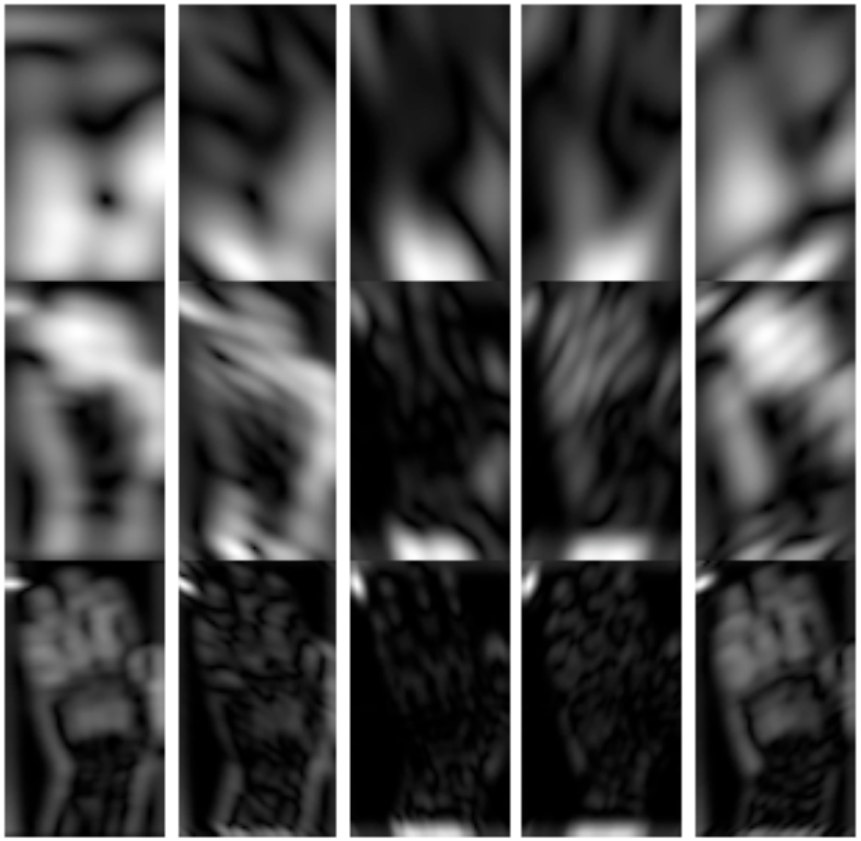


Tamura Features

- Originally one value per feature per image was determined
- Create a descriptor by
  1. Calculate coarseness, contrast, and directionality in a local neighbourhood for each pixel
  2. Create a joint histogram over these values

Gabor Features

- Obtain several values per pixel denoting spatial frequencies and directions



Gabor Features

- Windows Fourier transform with Gaussian $g_\alpha(t)$ as window function:

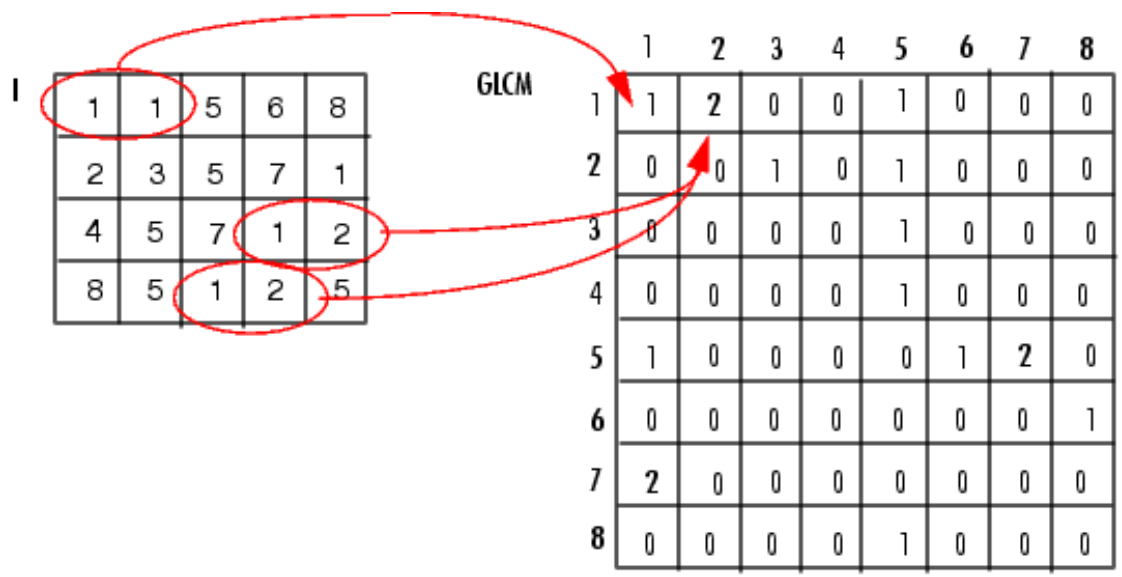
$$g_\alpha(t) = \frac{1}{2\sqrt{\pi\alpha}} e^{-\frac{t^2}{4\alpha}}$$

$$\begin{aligned} G_f(\omega, \tau) &= \int_{-\infty}^{+\infty} f(t) g_\alpha(t-\tau) e^{i\omega t} dt \\ &= e^{-i\omega\tau}(f(\tau) * h(\tau)) \end{aligned}$$

- Gabor features:
  - Yield one value per pixel (for grey value images)
  - For colour images use a transformed HSV space
- Compute global (histogram) or local features from Gabor responses

Gray-Level Co-Occurrence Matrices

- Statistical descriptor for texture properties of an image by comparing neighbouring pixels
  - Direction and distance
  - Extract features from matrix
    - Entropy
    - Contrast
    - Correlation
    - …



Histogram Comparison

- L_p - distances

$$d(h, h') = \left(\sum_i (h_i - h'_i)^p \right)^{\frac{1}{p}}$$

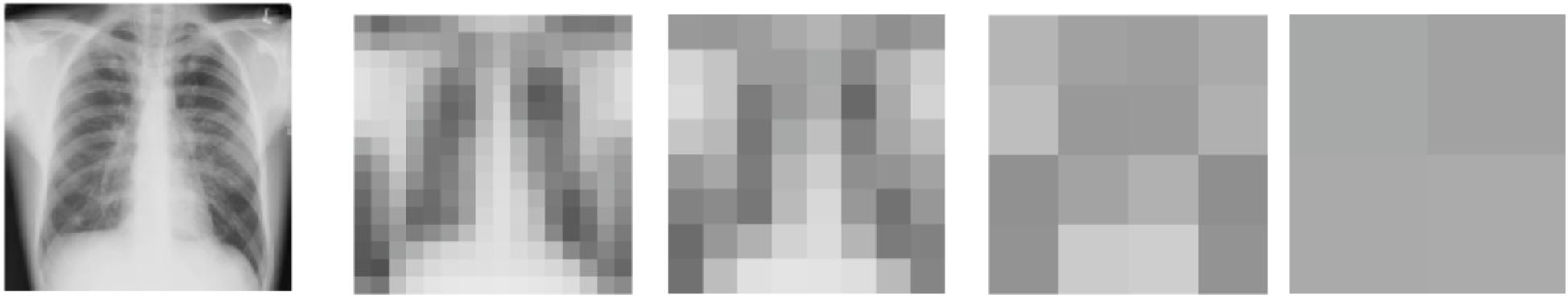
- Jensen Shannon Divergence

$$d_{JSD}(h, h') = \sum_i h_i \log \frac{2h_i}{h_i + h'_i} + h'_i \log \frac{2h'_i}{h'_i + h_i}$$

Pixel Values as Features

- Most straightforward
- Scale all images to a common size
- Compare pixels using e.g. Euclidean distance

- Multi-scale Representations:



Direct Comparison of Images

- Pixel-wise:

$$d(A,B) = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(A(i,j) - B(i,j)\right)^2}$$

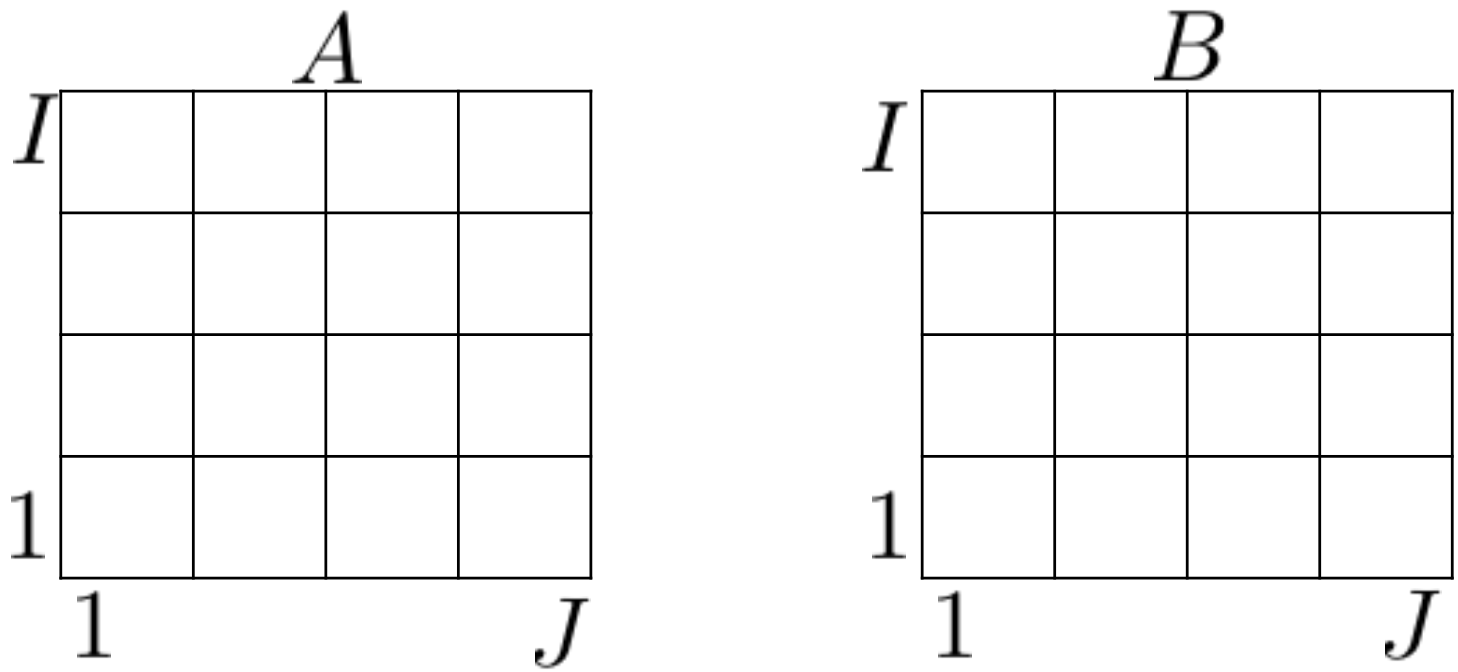
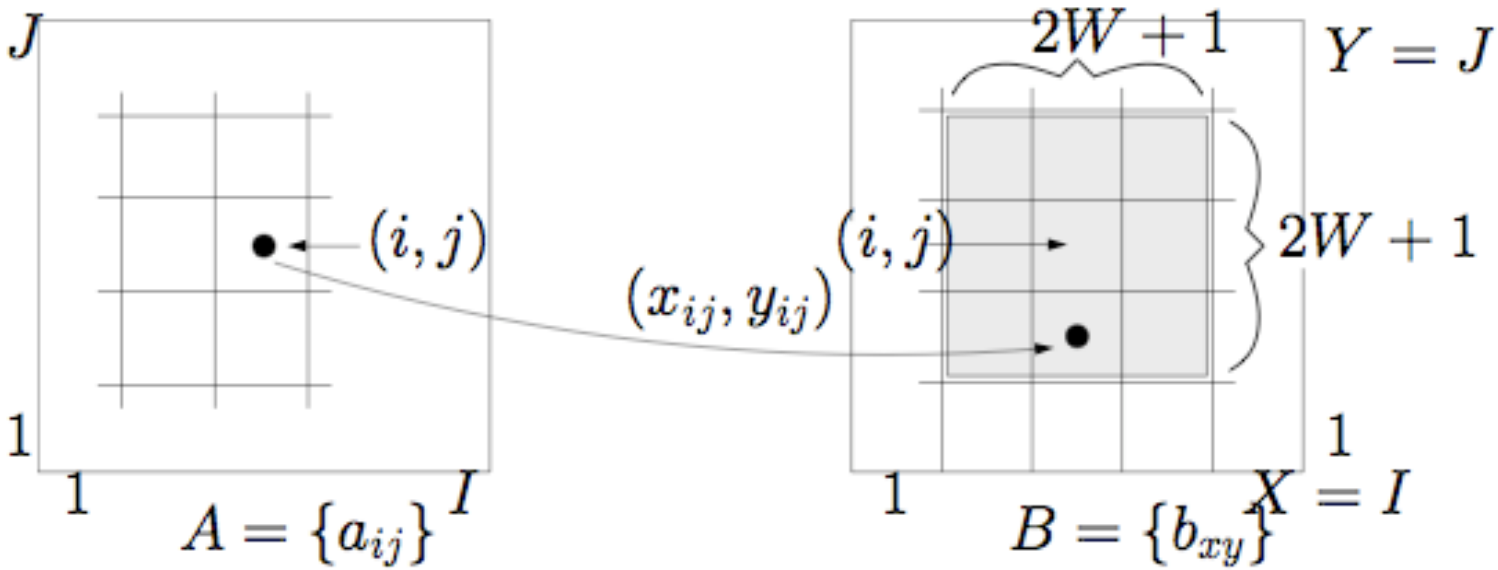


Image Distortion Model

- Allow for small local displacements
- Computationally efficient

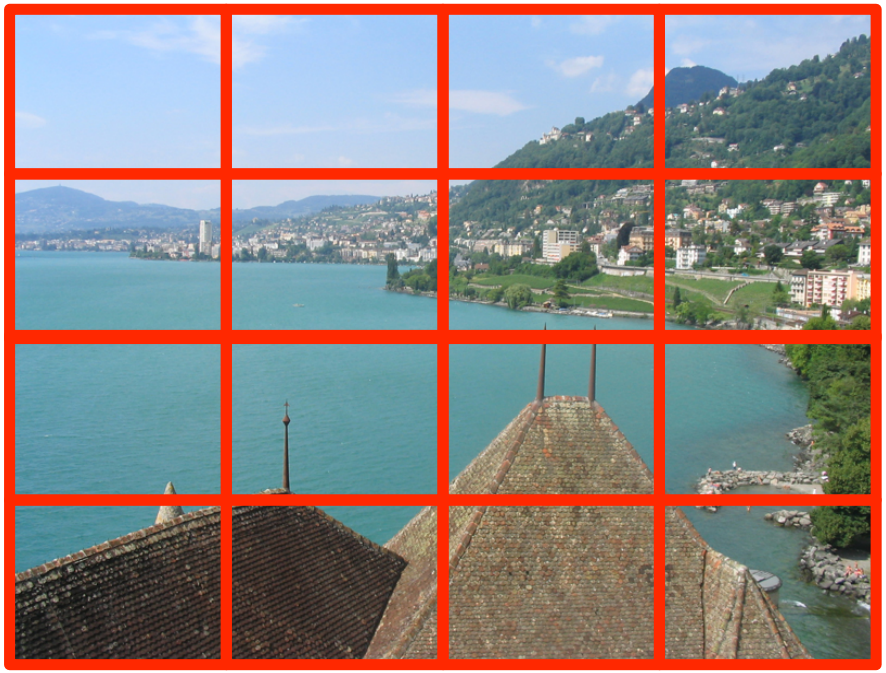


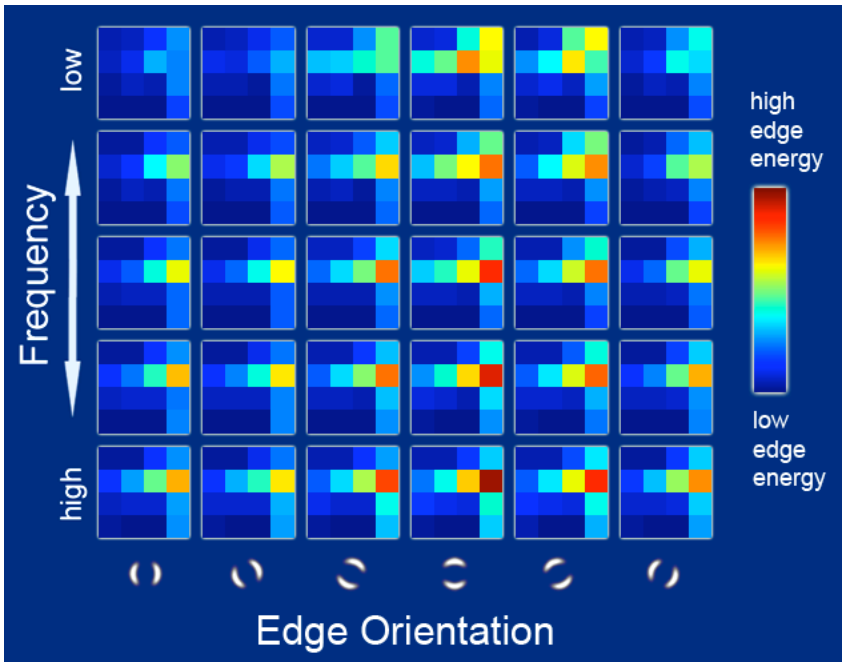
Shape: GIST descriptor

- Describe the shapes occurring in an image with one descriptor
  - Subdivide image in 4×4 sub images
  - Calculate Gabor responses in each of these
  - Create histograms of Gabor responses in each sub image

  - Has been shown to be helpful to distinguish
    - Naturalness
    - Openness
    - Roughness
    - Expansion
    - Ruggedness

Shape: GIST descriptor





GIST descriptor
Oliva and Torralba, IJCV 2001

Slide by James Hays and Alexei Efros

Local Descriptors

- Definition:
  - Features extracted from local regions from the image
  - E.g. patches, SIFT features, local colour histograms, …
  - Extraction position determined by interest points
  - Known to achieve good results in many tasks
- Active field of research in object recognition, detection, scene classification, image annotation, and more recently: image retrieval

Local Descriptors: Interest Points

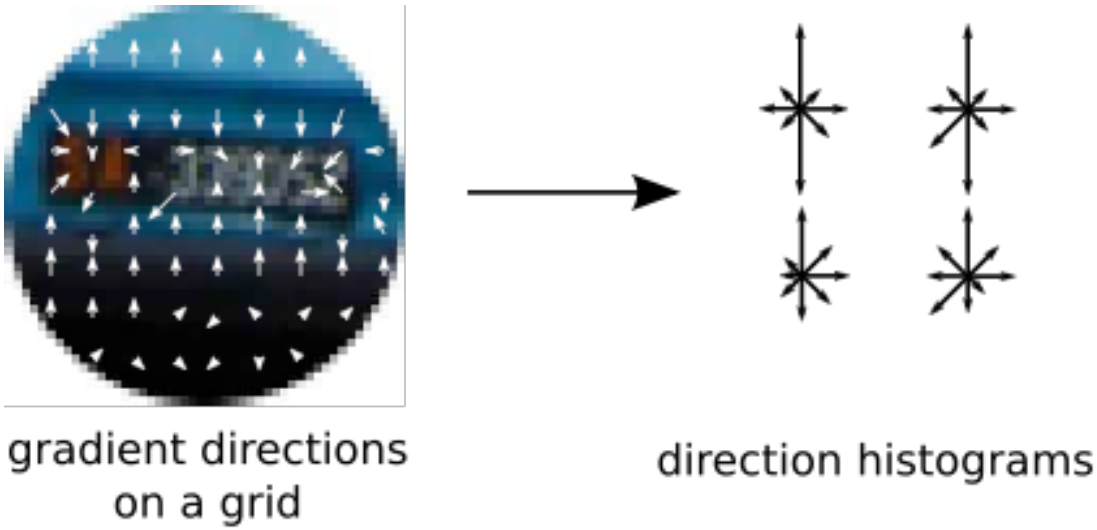
Local Descriptors: Patches

- Extract patches from the image
- Apply a PCA transformation to reduce dimensionality

- Can easily handle colour and gray value images
- All methods from invariant image comparison can be applied at patch level

Local Descriptors: SIFT

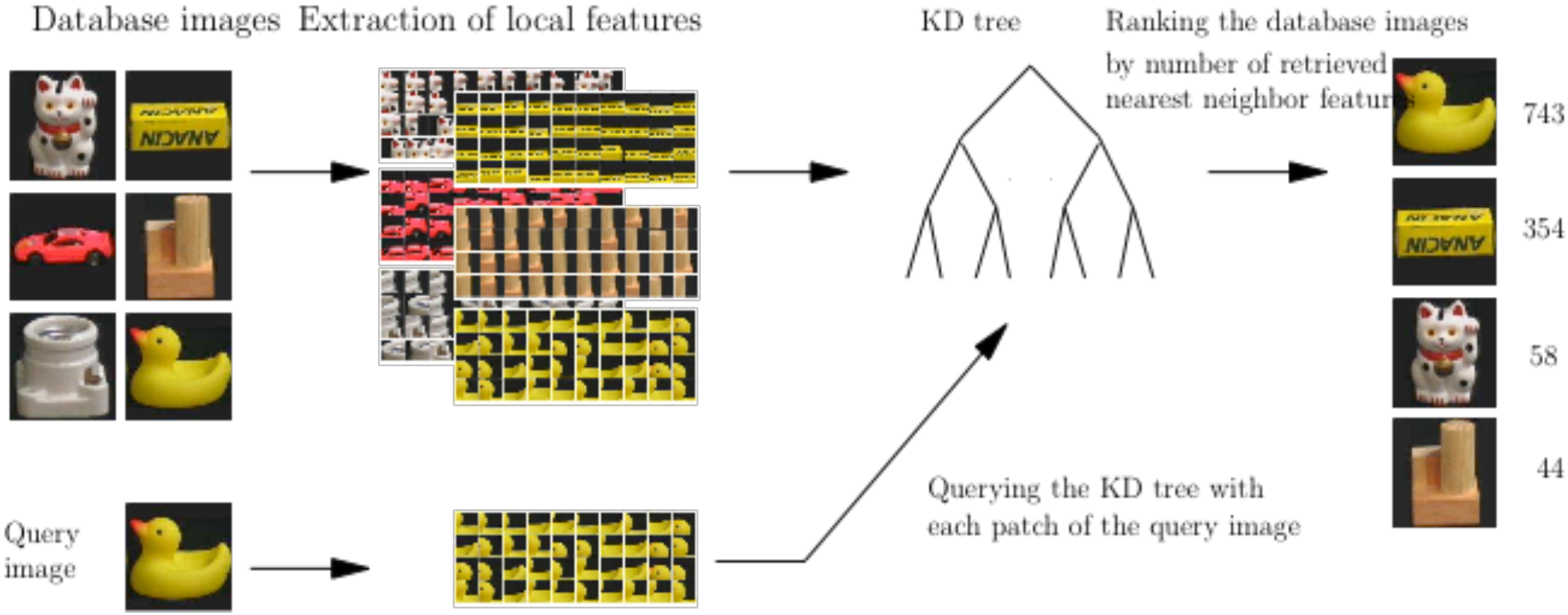
- Store a histogram of gradients in local areas
- SIFT = Scale Invariant Feature Transform



- Leading to 128-dimensional feature vectors
- Have been shown to perform well in many tasks

Figure by T. Weyand

Local Descriptors: Direct Retrieval



Histograms of Local Descriptors

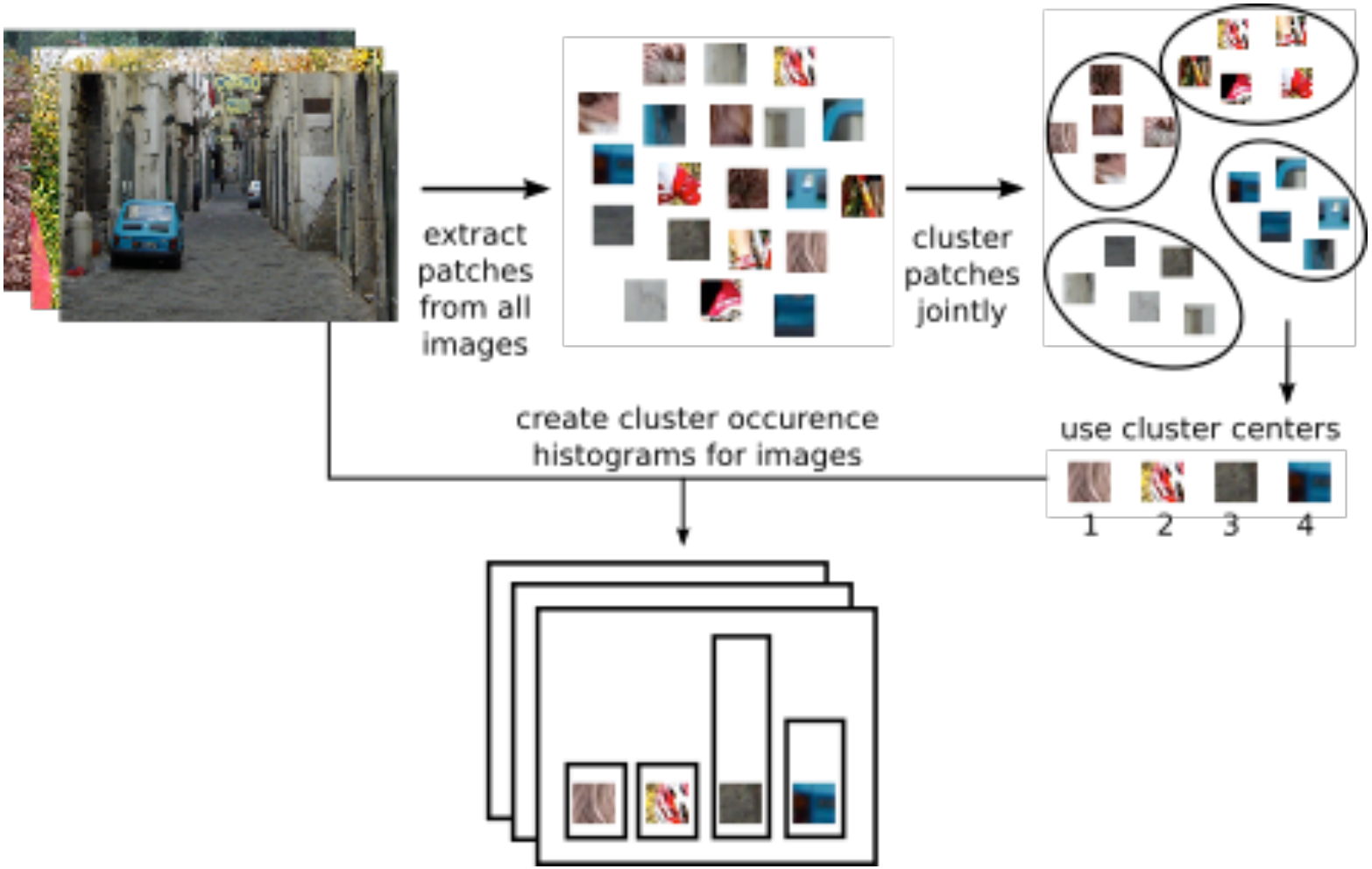
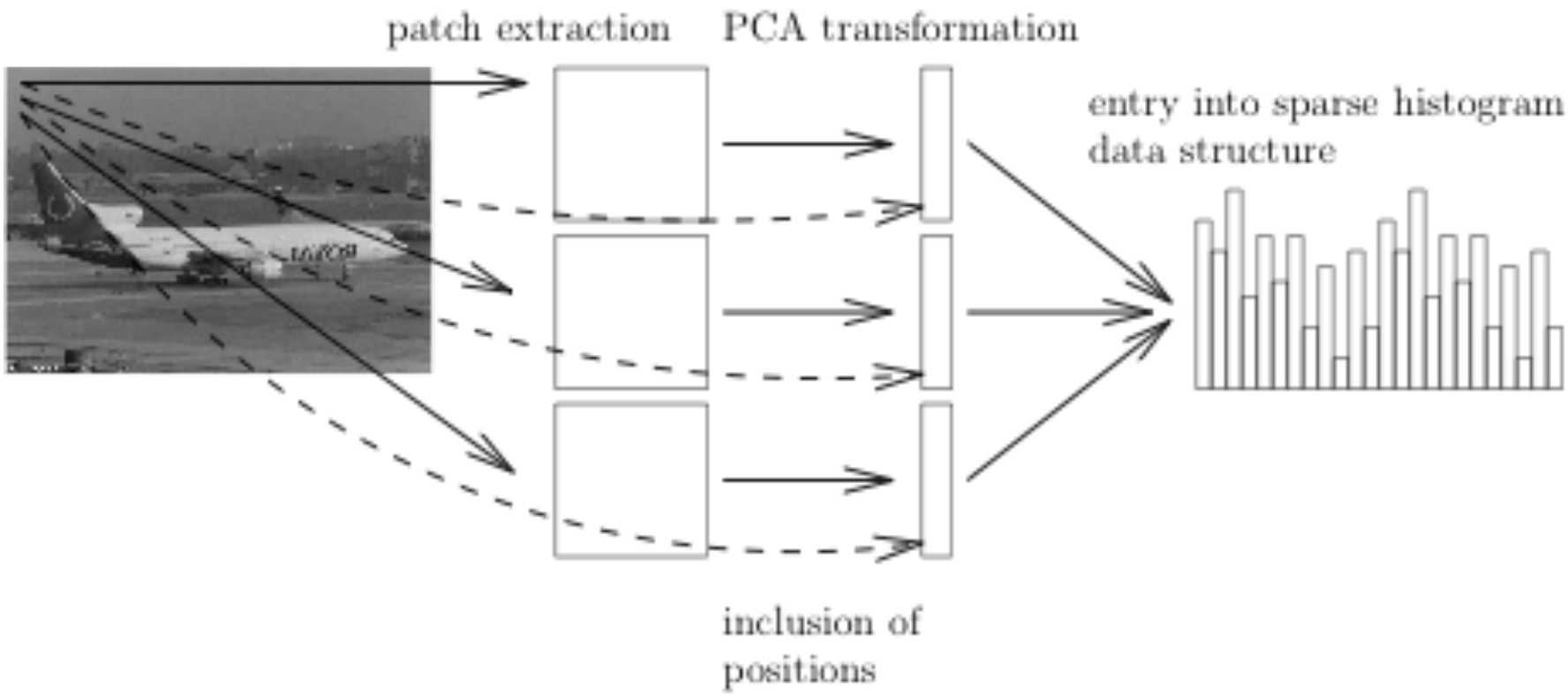
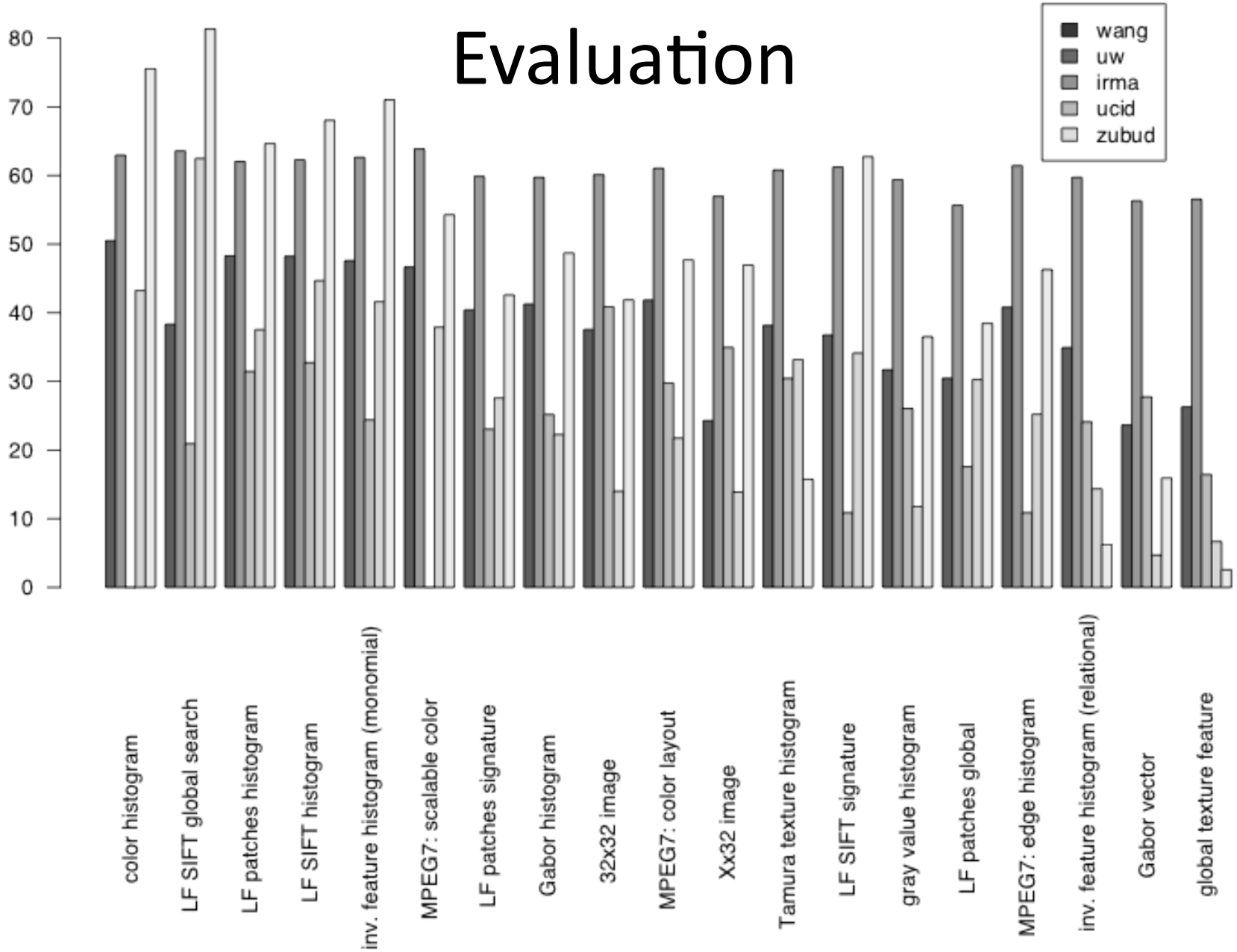


Figure by T. Weyand

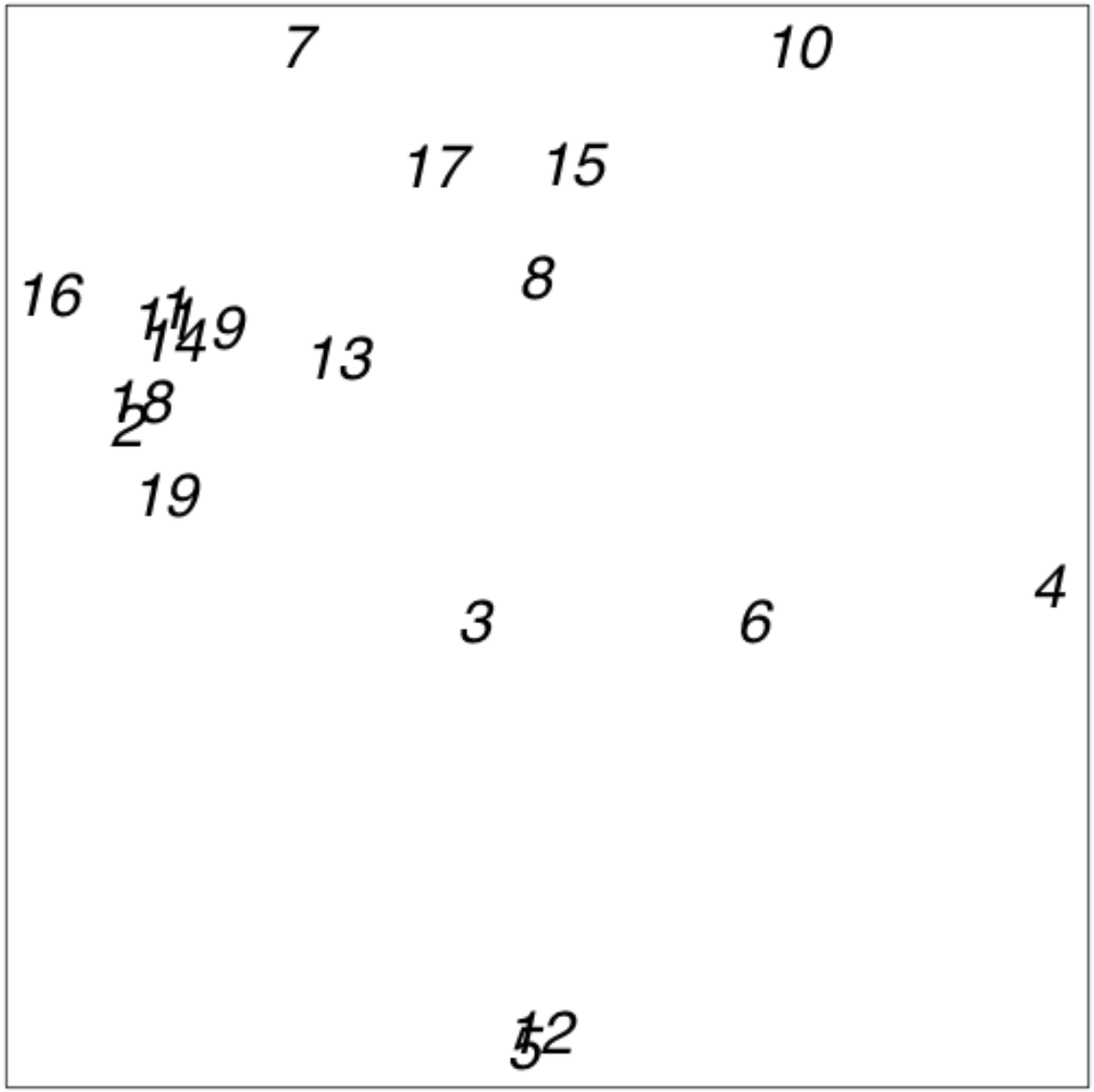
Histograms of Local Descriptors



Features for CBIR: Performance Evaluation

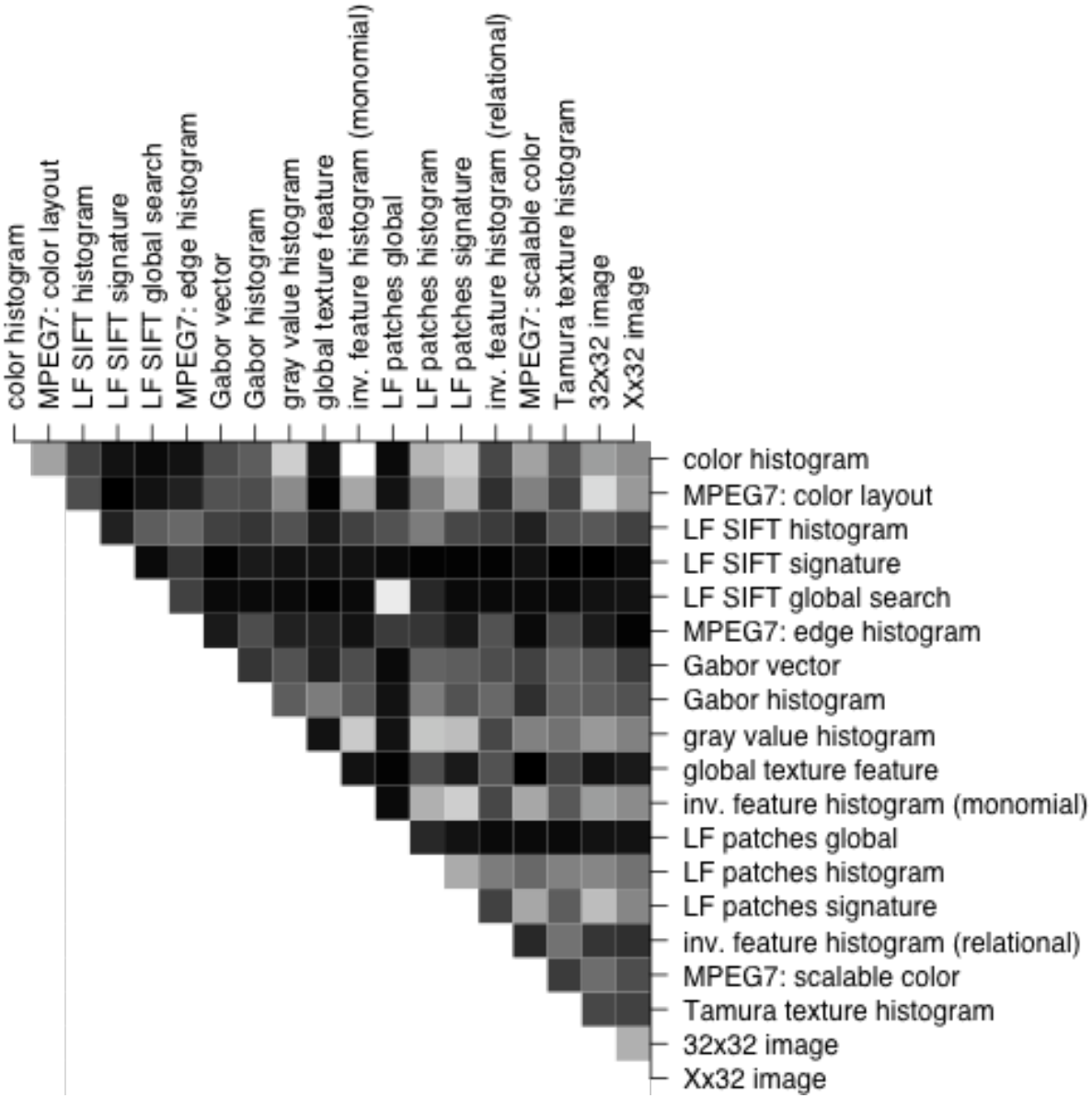


Correlation Between Features



1: colour histogram
2: MPEG7: colour layout
3: LFSIFThistogram
4: LFSIFTsignature
5: LFSIFTglobal search
6: MPEG7: edgehistogram,
7: Gaborvector
8: Gaborhistograms
9: grayvaluehistogram
10: global texturefeature,
11: inv. Feat histo color
12: Lfpatchesglobal
13: LFpatcheshistogram
14: LFpatchessignature,
15: inv. feat histo rel
16: MPEG7: scalablecolor,
17: Tamura
18: 32x32 image
19: Xx32image.

Correlation Between Features



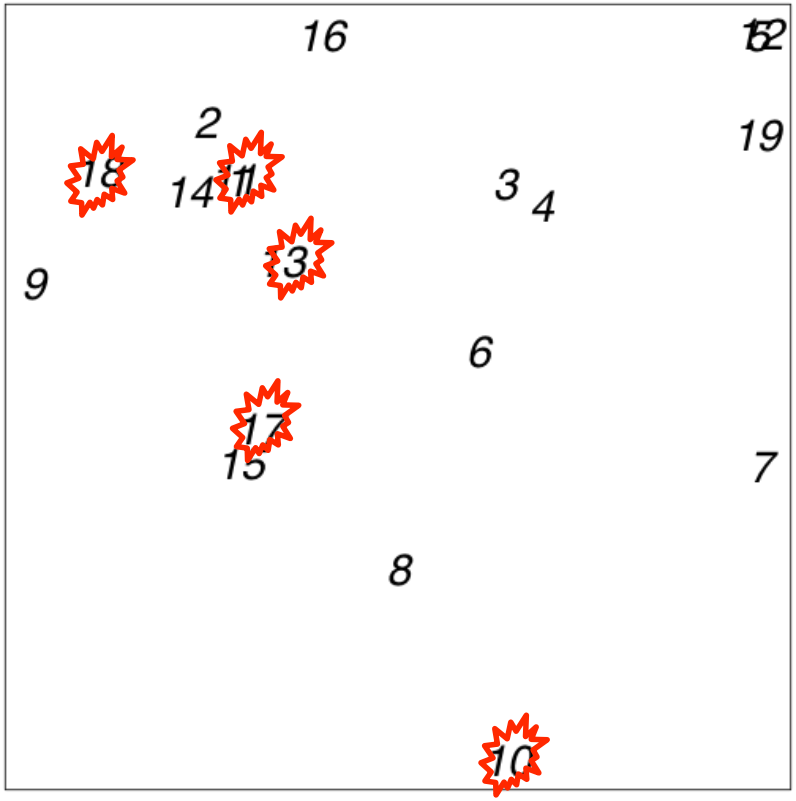
Combining Features

- Manually tuned
  - Have an `expert' find a proper set of parameters
- Heuristic to capture different image properties
- Combination to reflect human perception
- Combination to obtain optimal performance

  (given a set of training queries)

Combining Features to Capture Different Image Properties

- Given the result from the correlation analysis, first choose a simple feature
- Then add features which have low correlation



Color Histogram: 50.5% MAP

+ Global Texture Features: 49.5 % MAP

+ Tamura Texture Histogram: 51.2% MAP

+ Image Thumbnails: 53.9% MAP

+ Patch Histograms: 55.7% MAP

Combining Features Reflecting Human Perception

- Comparison of Human perception of image similarity and low-level image descriptors:

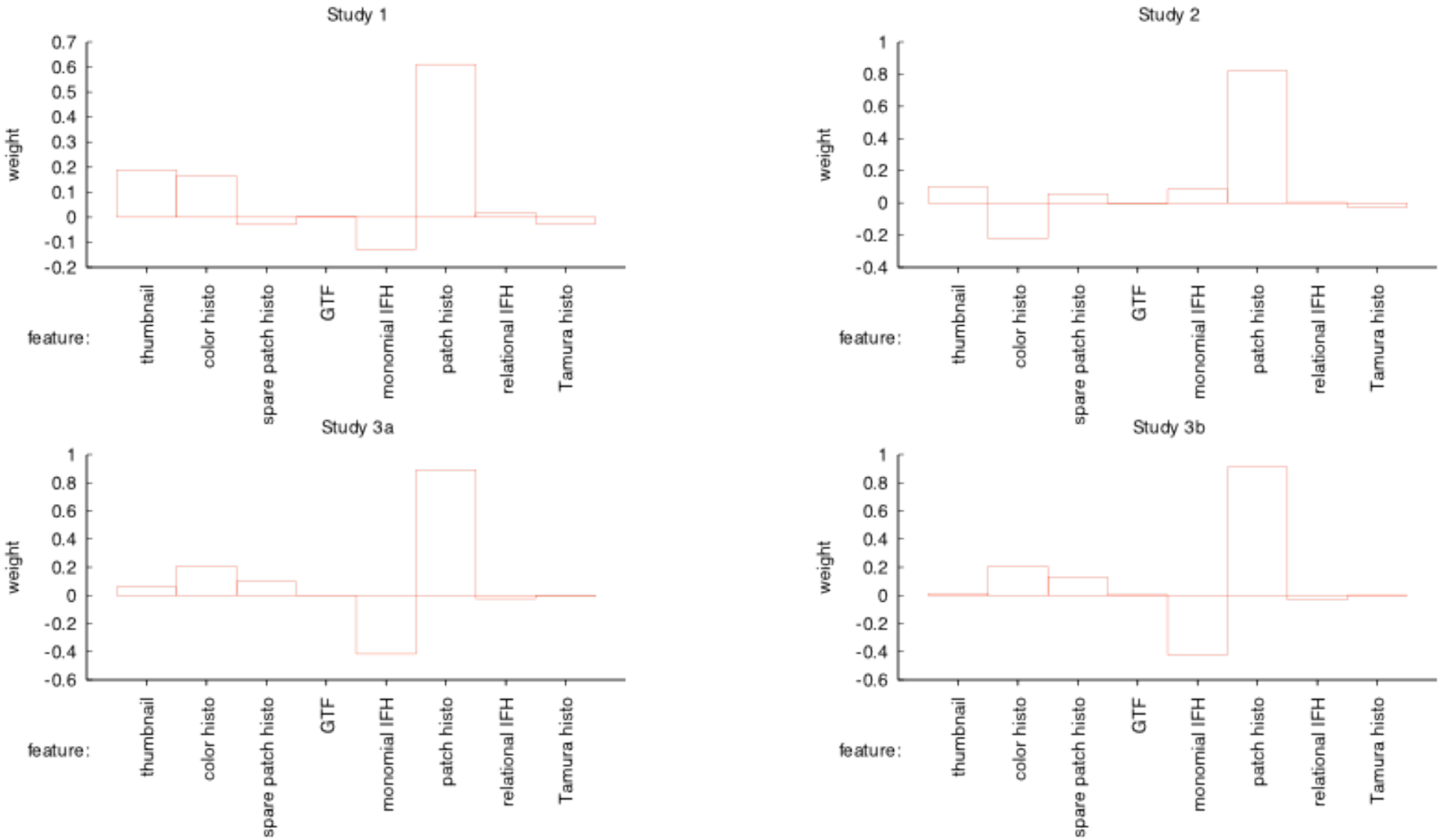
	Feature	Study 1		Study 2		Study 3 (AB)		Study 3 (BA)	
		r	p	r	p	r	p	r	p
1	32×32 image	0.230*	0.0009	0.162*	0.0009	0.103	0.066	0.070	0.148
2	Colour histogram	0.269*	0.0009	0.160*	0.0009	0.062	0.168	0.076	0.118
3	GTF	0.015	0.3027	0.047	0.0569	0.055	0.199	0.079	0.128
4	Monomial IFH	0.232*	0.0009	0.162*	0.0009	0.043	0.2478	0.055	0.1888
5	Relational IFH	0.185*	0.0009	0.090**	0.0029	-0.037	0.315	-0.047	0.271
6	Tamura histogram	0.184*	0.0009	0.107*	0.0009	0.021	0.343	0.042	0.262
7	4,096 bin patch histo	0.295*	0.0009	0.237*	0.0009	0.214*	0.0009	0.267*	0.0009
8	65,536 bin patch histo	0.281*	0.0009	0.245*	0.0009	0.121	0.015	0.143	0.008

Combining Features Reflecting Human Perception

- Combinations of image descriptors to achieve best compliance with human perception

		r values ($p = 0.000999$)			
		Study 1	Study 2	Study 3 (AB)	Study 3 (BA)
Weights from study	1	0.393648	0.244613	0.201997	0.207113
	2	0.257257	0.269415	0.265245	0.294700
	3 (AB)	0.256781	0.167512	0.298475	0.350045
	3 (BA)	0.245353	0.163617	0.290779	0.355902

Combining Features Reflecting Human Perception



Combining Features to Obtain Optimal Performance

- Consider retrieval to be a two-class problem
- Classify images as `relevant' or `irrelevant'
- Use trained classifiers
- **Requires training data**
  - Can be obtained from experts
  - Or relevance feedback (more on this later)

Combining Features: Maximum Entropy/Log-Linear Models

- Learn a model to predict probabilities for an image to be relevant

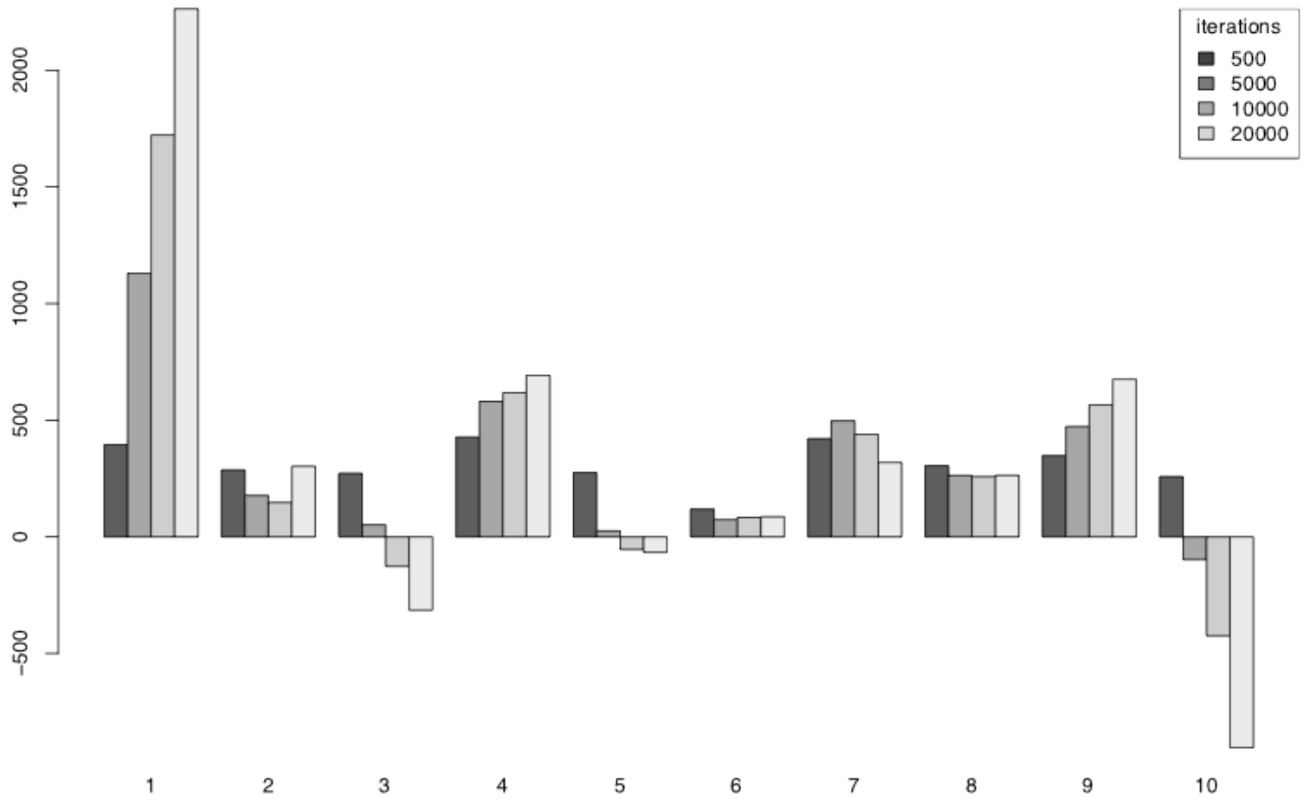
$$p(\oplus|Q,X) \quad = \quad \frac{\exp\left[\sum_i \lambda_{\oplus i} f_i(Q,X)\right]}{\sum\limits_{k\in\{\oplus,\ominus\}} \exp\left[\sum_i \lambda_{ki} f_i(Q,X)\right]}$$

- $f_i(Q,X)$ is a function describing some similarity/dissimilarity of the query Q and image X
- $\lambda_{\oplus i}$ are the parameters to be trained

Trained Lambdas

for medical retrieval

1. En
2. Fr
3. Ge
4. Colour
5. Gray
6. GTF
7. IFH
8. Tamura
9. 32x32
10. PatchHisto



Combining Features: Support Vector Machines

- Train a two-class SVM, +1= relevant, -1=irrelevant

$$X \mapsto \hat{c}(X) \quad = \quad \mathrm{sgn}\left\{ \sum_{v_i \in \mathcal{S}} \alpha_i K(X, v_i) + \beta \right\}$$

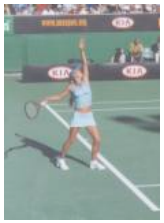
- Since SVMs sometimes have problems with non-uniformly distributed classes, subsample training samples to have approximately the same amount of positive and negative samples

Available Resources

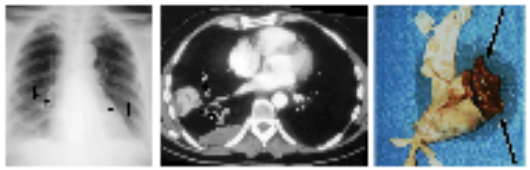
- Image Retrieval Systems:
  - GIFT – GNU Image Finding Tool
    - http://www.gnu.org/software/gift/
    - Full image retrieval system
    - Following the discrete approach
  - FIRE – Flexible Image Retrieval Engine
    - http://www-i6.informatik.rwth-aachen.de/~deselaers/fire/
    - Research image retrieval system
    - Developed to allow for easy extension
    - Following the Continuous approach
  - openCV – computer vision library
    - http://sourceforge.net/projects/opencvlibrary/
    - Implements may image processing operations
    - E.g. face detection and recognition, feature extraction

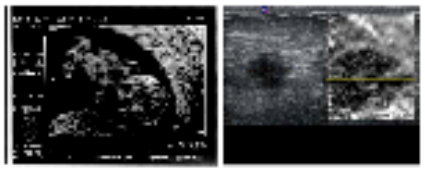
Available Resources

Data sets



- IAPR TC 12 Dataset
  - Used in ImageCLEF since 2006
  - 20,000 images with text in English, German, Spanish
  - Available from `www.imageclef.org/photodata`





- ImageCLEFmed Datasets
  - Nearly 70,000 images
  - From 50,000 medical cases
  - Images and text

Available Resources

Data sets

- Corel dataset
  - Widely used but not freely available
  - Questionable for evaluation
  - Different sets are in use in the CBIR community
    - Between 1,000 and 200,000 images
- MSRC dataset
  - Provided by Microsoft Research, Cambridge, UK
  - 4320 images from 45 classes
- Flickr
  - Difficult to use for evaluation although large number of images

References

- Overview
  - Smeulders, A. W., Worring, M., Santini, S., Gupta, A., and Jain, R. 2000. Content-Based Image Retrieval at the End of the Early Years. IEEE Trans. Pattern Anal. Mach. Intell. 22, 12 (Dec. 2000), 1349-1380.
  - Rui, Y., Huang, T., & Chang, S. (1999). Image retrieval: Current techniques, promising directions and open issues. Journal of Visual Communication and Image Representation, 10(4), 39–62.
  - Datta, R., Li, J., & Wang, J. Z. (2005). Content-based image retrieval—approaches and trends of the new age. In ACM Intl. Workshop on Multimedia Information Retrieval, ACM Multimedia. Singapore.
  - Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications and Applications, 2(1), 1–19.
  - Henning Müller, Nicolas Michoux, David Bandon, Antoine Geissbuhler, A review of content-based image retrieval systems in medicine - clinical benefits and future directions, International Journal of Medical Informatics, volume 73, pages 1-23, 2004
- Features:
  - Deselaers T., Keysers D., Ney H., "Features for Image Retrieval: An Experimental Comparison", Information Retrieval, vol. 11, issue 2, The Netherlands, Springer, pp. 77-107, 03/2008.
  - SIFT Descriptors
    - Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2: 1150–1157.
    - Lowe, David G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision 60 (2): 91–110.
  - Patch Histograms
    - Deselaers T., Keysers D., Ney H., "Discriminative Training for Object Recognition using Image Patches", CVPR, vol. 2, San Diego, CA, USA, IEEE, pp. 157-162, 20/06/2005.
  - GIST
    - Modeling the shape of the scene: a holistic representation of the spatial envelope.Aude Oliva, Antonio Torralba. International Journal of Computer Vision, Vol. 42(3): 145-175, 2001

References

- Continuous/Discrete
  - FIRE: see features
  - GIFT: Squire, D. M., Müller, W., Müller, H., & Raki, J. (1999) Content-Based query of image databases, inspi- rations from text retrieval: Inverted files, frequency-based weights and relevance feedback. In Scandinavian Conference on Image Analysis (pp. 143–149). Kangerlussuaq.
  - de Vries, A. P., & Westerveld, T. (2004). A comparison of continuous vs. discrete image models for probabilistic image and video retrieval. In Proc. International Conference on Image Processing (pp. 2387–2390). Singapore.
- Datasets
  - Grubinger M., Clough P.D., Müller H., Deselaers T., "The IAPR Benchmark: A New Evaluation Resource for Visual Information Systems", International Conference on Language Resources and Evaluation, Genoa, Italy, 24/05/2006.
  - Müller, H., Marchand-Maillet, S., and Pun, T. 2002. The Truth about Corel - Evaluation in Image Retrieval. In Proceedings of the international Conference on Image and Video Retrieval (July 18 - 19, 2002). M. S. Lew, N. Sebe, and J. P. Eakins, Eds. Lecture Notes In Computer Science, vol. 2383. Springer-Verlag, London, 38-49.
  - www.imageclef.org
- Learning Feature Combinations
  - Deselaers T., Weyand T., Ney H., "Image Retrieval and Annotation Using Maximum Entropy", CLEF Workshop 2006, vol. 4730, Alicante, Spain, Springer, pp. 725-734, 20/09/2006, 2007.
  - Gass T., Weyand T., Deselaers T., Ney H., "FIRE in ImageCLEF 2007: Support Vector Machines and Logistic Regression to Fuse Image Descriptors in for Photo Retrieval", Advances in Multilingual and Multimodal Information Retrieval 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, vol. 5152, Budapest, Hungary, Springer, 19/09/2007, 2008.
- Efficient NN search
  - HjaltasonG. R., SametH.: *Ranking in Spatial Databases, Int. Symp. on Large Spatial Databases(SSD), 1995.*
  - Berchtold, Böhm, Keim, Kriegel: *A Cost Model forNearest Neighbor Searchin High-Dimensional Space, ACM Symposium on Principlesof Database Systems, 1997.*