Fundamentals of Image Processing

▶ Lecture 9: Characterization of image primitives ◀

Master of Computer Science
Sorbonne University

September 2022

Outline

Introduction

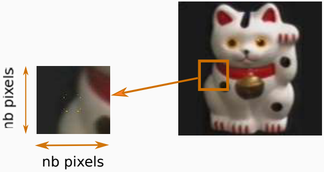
Back to classical detectors

- Several POI/ROI detectors are robust to changes in scale and point of view

  - MSER well suited for structured scenes
  - Harris and Hessian well suited to textured scenes
  - SIFT (DoG), and variants such as SURF, are very robust
- The algorithmic processing chain:
  - Detection of extrema in the space volume scale with possibly fine relocation of POI
  - Extraction of the area of interest (affine inv., rotation inv....)
  - Descriptor calculation

Introduction on feature descriptors

- Depending on the type of primitives
- Usual attributes:
  - Colorimetry
  - Calculations in a neighborhood: texture, curvature, shape
- Specialized attributes
  - Face images: distance between eyes...



Introduction on feature descriptors

- Why compute descriptors?
  - Registration: image alignment (homography, fundamental matrix)
  - 3D reconstruction
  - Motion tracking
  - Object recognition
  - Indexing and database retrieval
  - Navigation (robotics)
  - Classification
  - ...

Introduction on feature descriptors

- What are the expected properties?
  - Patches with similar content must have similar descriptors
  - The requirements of invariance (or robustness) in relation to transformations are multiple: image rotation, scale transformation, change of view point, change of brightness, change of objects in the category...

Issues

- Image transformations (rotation, scale)
- Brightness change
- Partial visibility / occlusions
- Additional objects, background
- Field of view for 3D objects

$\downarrow ?$

Introduction on feature descriptors

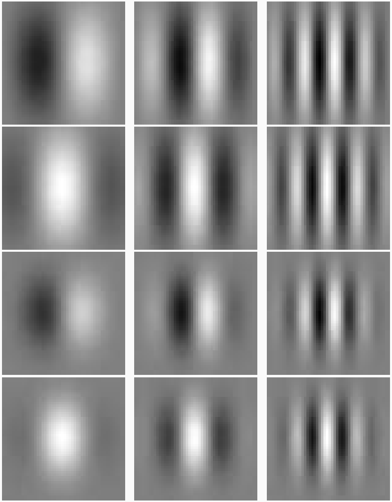
- What are the expected properties?
    - **Robustness** $\Rightarrow$ Invariance to changes in illumination, scale, rotation, affine transformations, perspective
    - **Distinctive** $\Rightarrow$ useful for inquiring large databases
    - **Efficient** $\Rightarrow$ easy and quick to calculate (real-time)
    - **Representative** $\Rightarrow$ many points/regions detected even on small areas/objects
- Local primitive descriptors (ROI, POI, edge):
    - **Local** $\Rightarrow$ therefore intrinsically robust to occlusions and (background) noise
    - **Modular** $\Rightarrow$ easily complemented by other types of descriptors to gain in robustness

Introduction on feature descriptors

- Typology of local descriptors:
  - Texture, filter banks
  - Derivatives of the gray level function
  - Differential invariants [van Gool et al., 1996]
  - SIFT descriptor [Lowe, 1999]
  - Moment invariant [van Gool et al., 1996]
  - Shape context analysis [Belongie et al., 2002]
  - Gray-scale - Centering, normalization (MOPS) [Brown et al., 2005]

Image descriptors

Descriptors based on Gabor filter banks

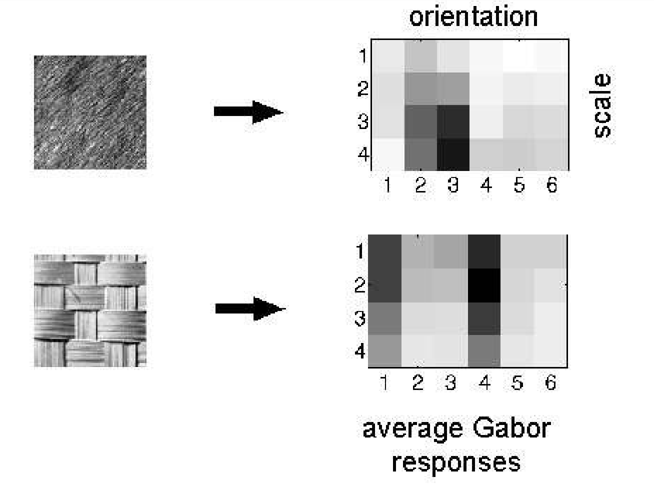


- Gabor filters at different spatial scales and frequencies
- Anti-symmetrical (odd) or symmetrical (even) filters

- Symmetrical: $\cos(k_x x + k_y y)e^{-\frac{x^2+y^2}{2\sigma^2}}$
- Anti-symmetrical: $\sin(k_x x + k_y y)e^{-\frac{x^2+y^2}{2\sigma^2}}$

Descriptors based on Gabor filter banks

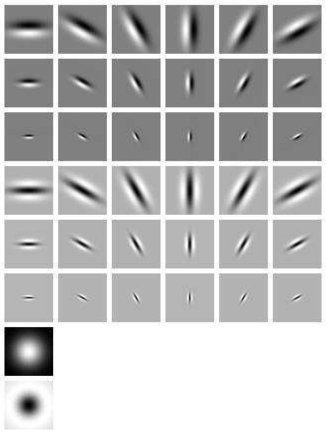
- Vector of dimension 6×4



darkest is the highest response

What does human use?

- Gabor filters!

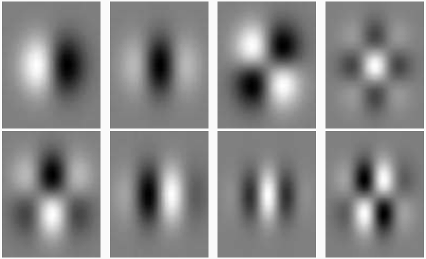


- ... and many other things.

Descriptors based on Gaussian derivative filter banks

- Steerable filters [Freeman and Adelson, 1991], dim 13.

$\star \Big\{$



Local descriptors based on image intensity derivatives

- Convolution with Gaussian derivatives

$$\mathbf{v}(x,y) = \begin{pmatrix} I \star G(\sigma) \\ I \star G_x(\sigma) \\ I \star G_y(\sigma) \\ I \star G_{xx}(\sigma) \\ I \star G_{xy}(\sigma) \\ I \star G_{yy}(\sigma) \\ \vdots \end{pmatrix}$$

- Reminder:
  - $I \star G(\sigma)(x,y) = \int_{\mathbb{R}} \int_{\mathbb{R}} G(x', y', \sigma) I(x - x', y - y') dx' dy'$
  - $G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$

Reminder: derivation and convolution commute

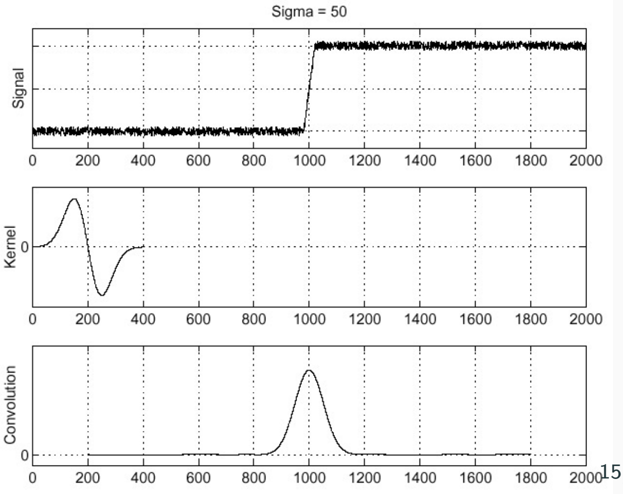
- As linear operators, derivation and convolution commute, i.e.:

$$\frac{\partial}{\partial x}(f \star g) = f \star \frac{\partial g}{\partial x}$$

f

$\frac{\partial g}{\partial x}$

$f \star \frac{\partial g}{\partial x}$



Local derivative descriptors

- Notation:

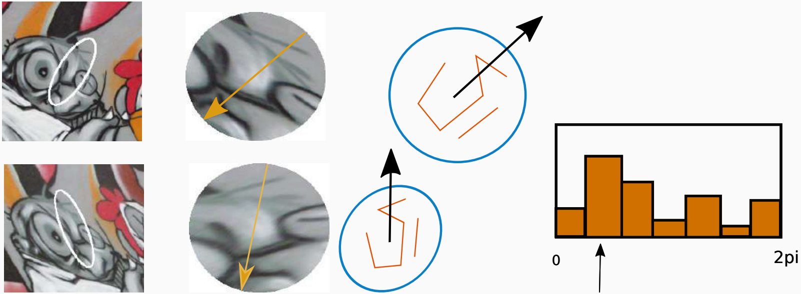
$$\mathbf{v}(x,y) = \begin{pmatrix} I \star G(\sigma) \\ I \star G_x(\sigma) \\ I \star G_y(\sigma) \\ I \star G_{xx}(\sigma) \\ I \star G_{xy}(\sigma) \\ I \star G_{yy}(\sigma) \\ \vdots \end{pmatrix} = \begin{pmatrix} L(x,y) \\ L_x(x,y) \\ L_y(x,y) \\ L_{xx}(x,y) \\ L_{xy}(x,y) \\ L_{yy}(x,y) \\ \vdots \end{pmatrix}$$

Local derivative descriptors: rotation invariance

- Rotation invariance: differential invariants

$$
\begin{array}{rl}
 & \\
\text{Gradient norm} \longrightarrow & \\
 & \\
\text{Laplacian} \longrightarrow & \\
 & \\
 &
\end{array}
\begin{pmatrix}
L \\
L_x L_x + L_y L_y \\
L_{xx} L_x L_x + 2 L_{xy} L_x L_y + L_{yy} L_{yy} \\
L_{xx} + L_{yy} \\
L_{xx} L_{xx} + 2 L_{xy} L_{xy} + L_{yy} L_{yy} \\
\vdots
\end{pmatrix}
$$

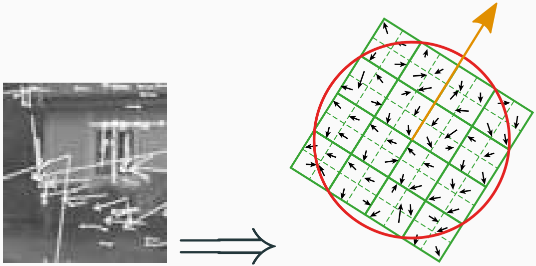
Alternative: normalization of the analysis window



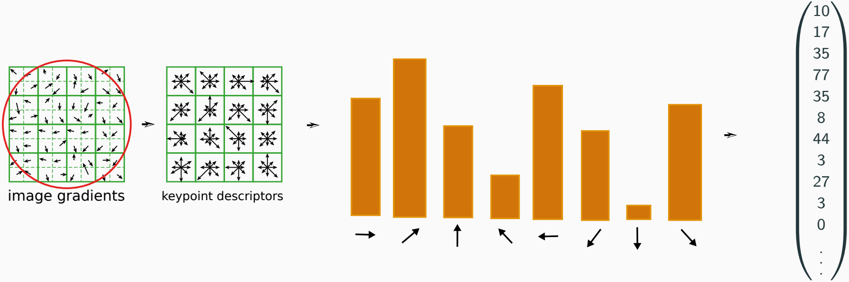
- Computation of orientations histogram
- Selection of the dominant orientation
- Normalization: rotation according to dominant orientation

SIFT descriptor

- Scale Invariant Feature Transform
- Principle:
  1. Square window shifted in the direction of the main gradient
  2. Gaussian weighting
  3. Histogram of gradient orientations image computed in each quadrant



- Creation of 16 gradient histograms (8 bins)
  - Gaussian weighting (centred on a POI, σ set to half of the window size)

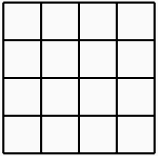


$\Rightarrow$ vector of size 128

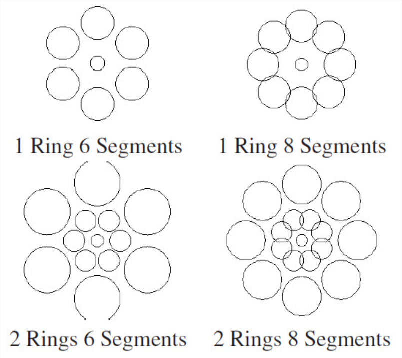
SIFT descriptor

- Invariance to affine illumination changes: $I \mapsto a \times I + b$:
  - Gradients unaffected by a bias ($I \mapsto I + b$, gain change)
  - Normalization of the histogram $\Rightarrow$ insures robustness to contrast change ($I \mapsto a \times I$)
- Compared to non-linear changes:
  - Saturation affects more amplitude than orientation
  - Thresholds on amplitude (empirically set to 0.2) and normalization

Other methods: Daisy



SIFT



DAISY

Other methods: BRIEF

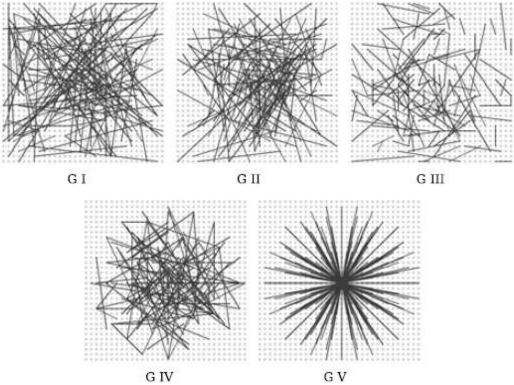
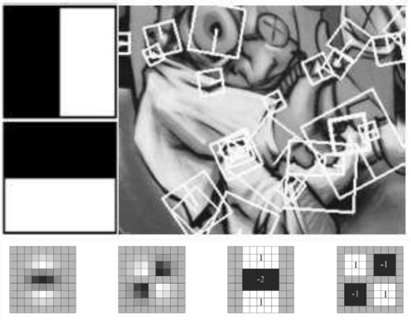


Fig. 2. Different approaches to choosing the test locations. All except the righmost one are selected by random sampling. Showing 128 tests in every image.

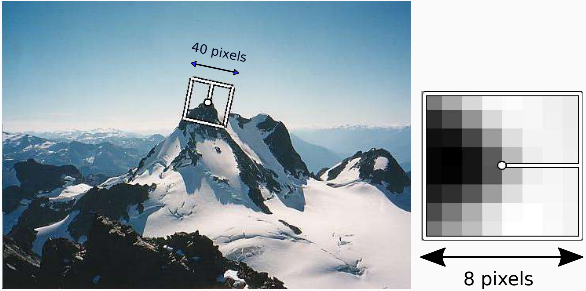
SURF: a faster SIFT



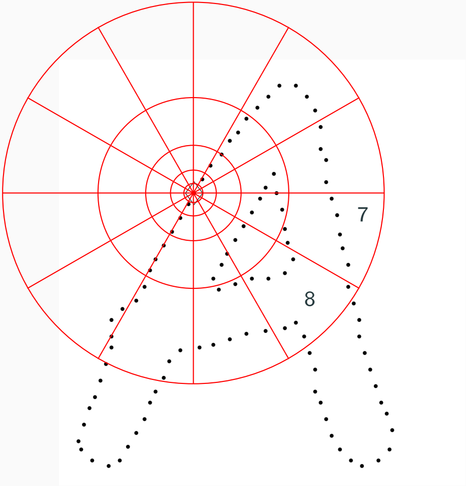
- Fast approximation of SIFT
    - Efficient computation with 2D filters and integral images
    - 6 times faster than SIFT with similar quality for object identification

MOPS: normalized patch type descriptor

- Compute multiscale Harris corners, first eigenvector gives orientation.
- Build 8x8 patches with orientation normalization (64 values)
  - Sampled with step 5 and using anti-aliasing filtering
- Renormalization (bias/gain): $I' = (I - \mu)/\sigma$



Shape-based descriptors [Belongie et al., 2002]



- Log-polar histogram
  - 5 bins in distance
  - 12 bins in orientation
- Count edges for all sub-domains

Conclusion on descriptors

- Many descriptors
- Many evaluations/comparisons
    - Presentation of top performers: SIFT
- Best choice often depends on the application
- Often better results by combining them
- How do we use these descriptors?

Applications

Image matching

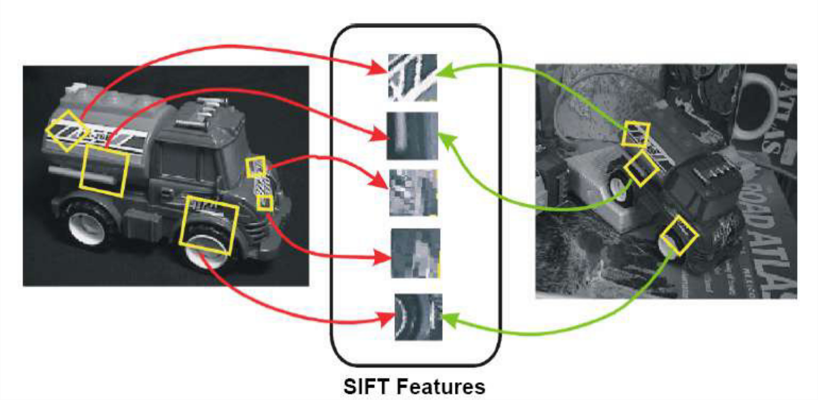
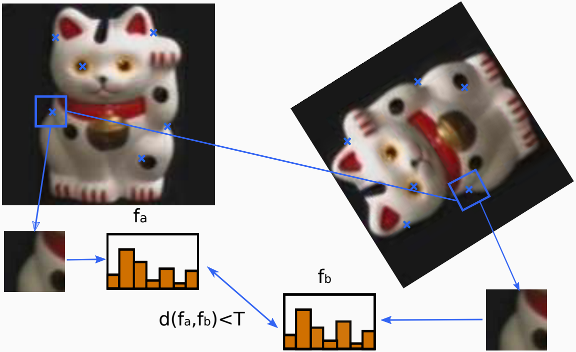


Image matching



1. Detection
2. Define a region of interest
3. Extraction / normalization
4. Descriptor computation
5. Matching

Image matching

Image matching

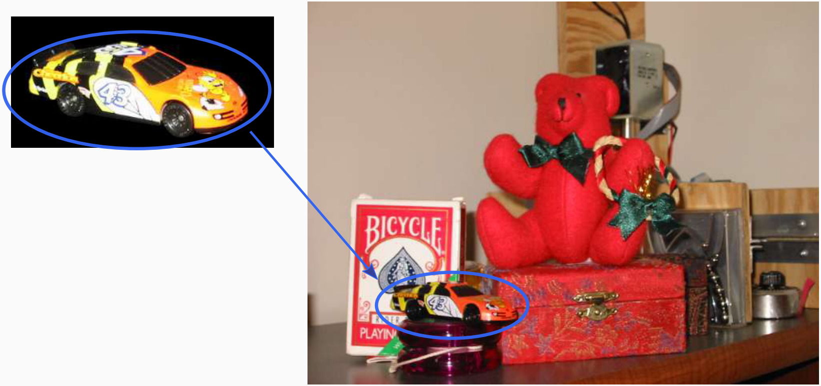
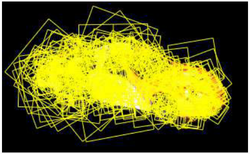


Image matching



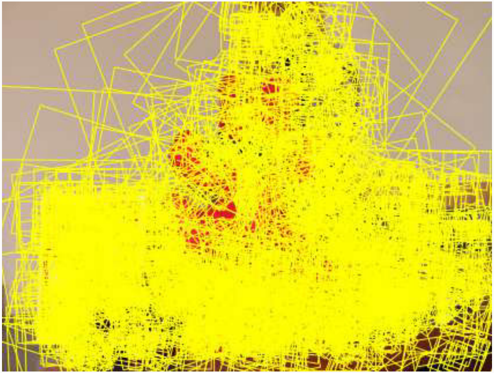


Image matching

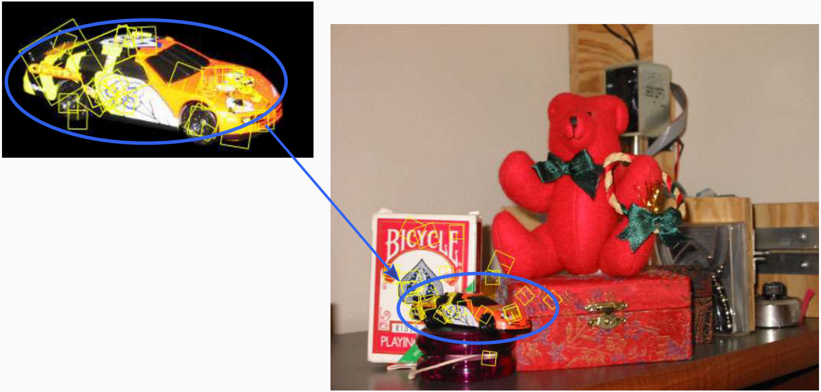


Image matching: localization of patches

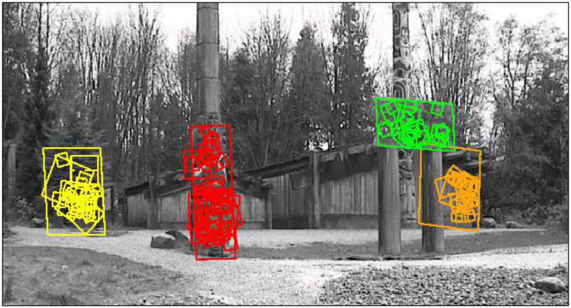


Image matching: change of point of view

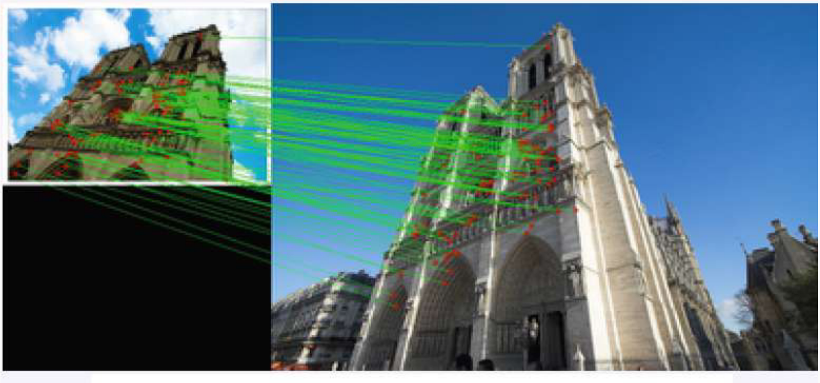
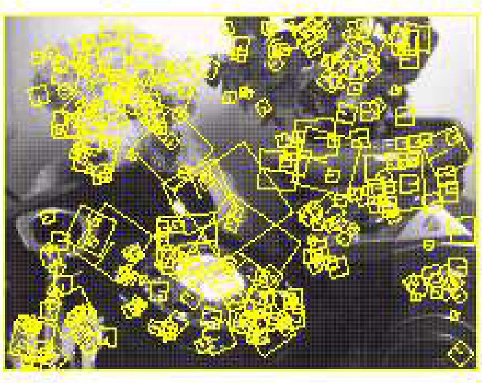


Image matching: robustness to occlusions

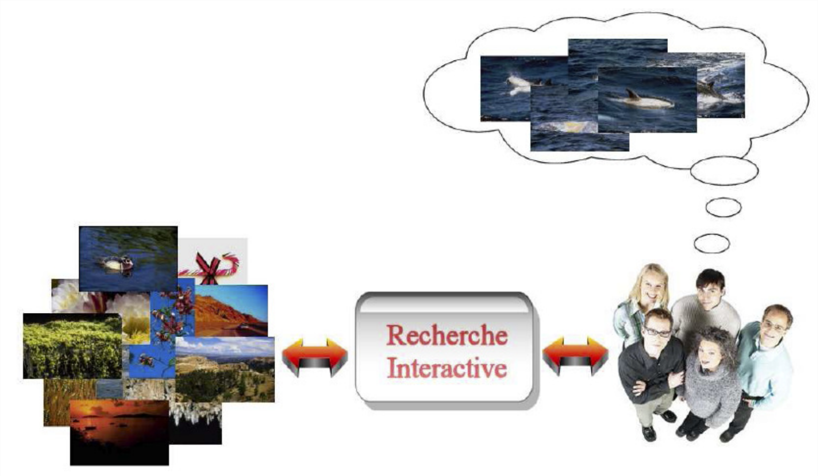


Image matching: robustness to change of illumination



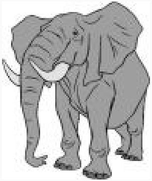
Generation of panoramas by registration

Indexing and searching in a database

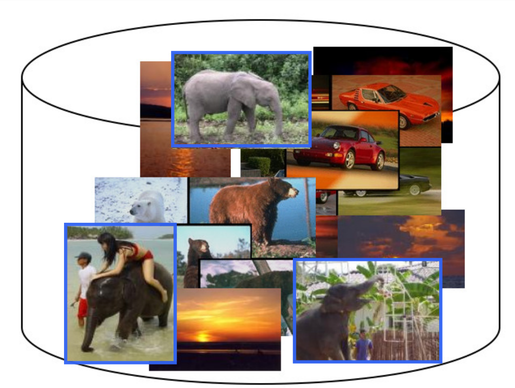


Search for image classes

- Query a category of objects



Query



Database

Search for image classes

- Query by semantics
  - Search for high level scene types or descriptors

Query

Database

Search engine architecture

- Off-line:
  1. Base of images
  2. Indexation (computation of image signatures)
  3. Indexed base
- On-line:
  1. User query
  2. Computation of image signature
  3. Find in the indexed base the most similar signatures (define a metric)
  4. Display results
  5. Optional: user feedback and back to 3.

Indexing: image signatures

- local signature:
  - set of descriptors (bag of vectors)
  - typical choice for images: SIFT
- global signature: dictionary-based approaches
  - construction of a dictionary of visual patterns (words)
  - projection on the dictionary and count words

Example of global signature: color histogram

- Codebook color

- Image signature: histogram computed on codebook

$[8, 27, 35, 48, 1, 10, 32, 9, 1, 4, 26, 6, 19, 7, 3, 8, 3, 7, 2, 0, 0, 0, 3, 0, 2]$

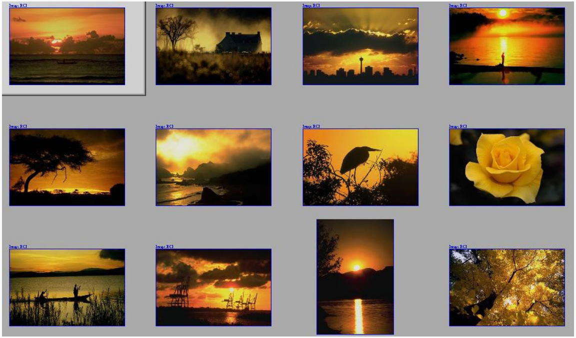
Measure of similarity

Measure of similarity: scalar product, L_1, L_2 norms...
Principle:

1. Pre-processing: centering, normalization...
2. Fusion of features: for a model i, compute on I a vector of features $F_i(I)$
3. Final vector of features: $F(I) = (F_1(I), F_2(I), \cdots, F_n(I))$
4. Similarity between I_A and I_B: $\mathrm{Sim}(F(I_A), F(I_B))$

Example (practical work)

- Example of semantic search based on similarity between color histograms



Belongie, S., Malik, J., and Puzicha, J. (2002).
Shape matching and object recognition using shape contexts.
Trans. Pattern Analysis and Machine Intelligence, 2(4):509–522.

Brown, M., Szeliski, R., and S., W. (2005).
Multi-image matching using multi-scale oriented patches.
In *International Conference on Computer Vision and Pattern Recognition*, pages 510–517.

Freeman, W. and Adelson, E. (1991).
The design and use of steerable filters.
pami, 13(9):891–906.

Lowe, D. (1999).
Object recognition from local scale-invariant features.
In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1150–1157.

van Gool, L., Moons, T., and Ungureanu, D. (1996).
Affine/Photometric Invariants for Planar Intensity Patterns.
In *European Conference on Computer Vision*.