Image retrieval

Axel Carlier, Axel.Carlier@enseeiht.fr

INP ENSEEIHT

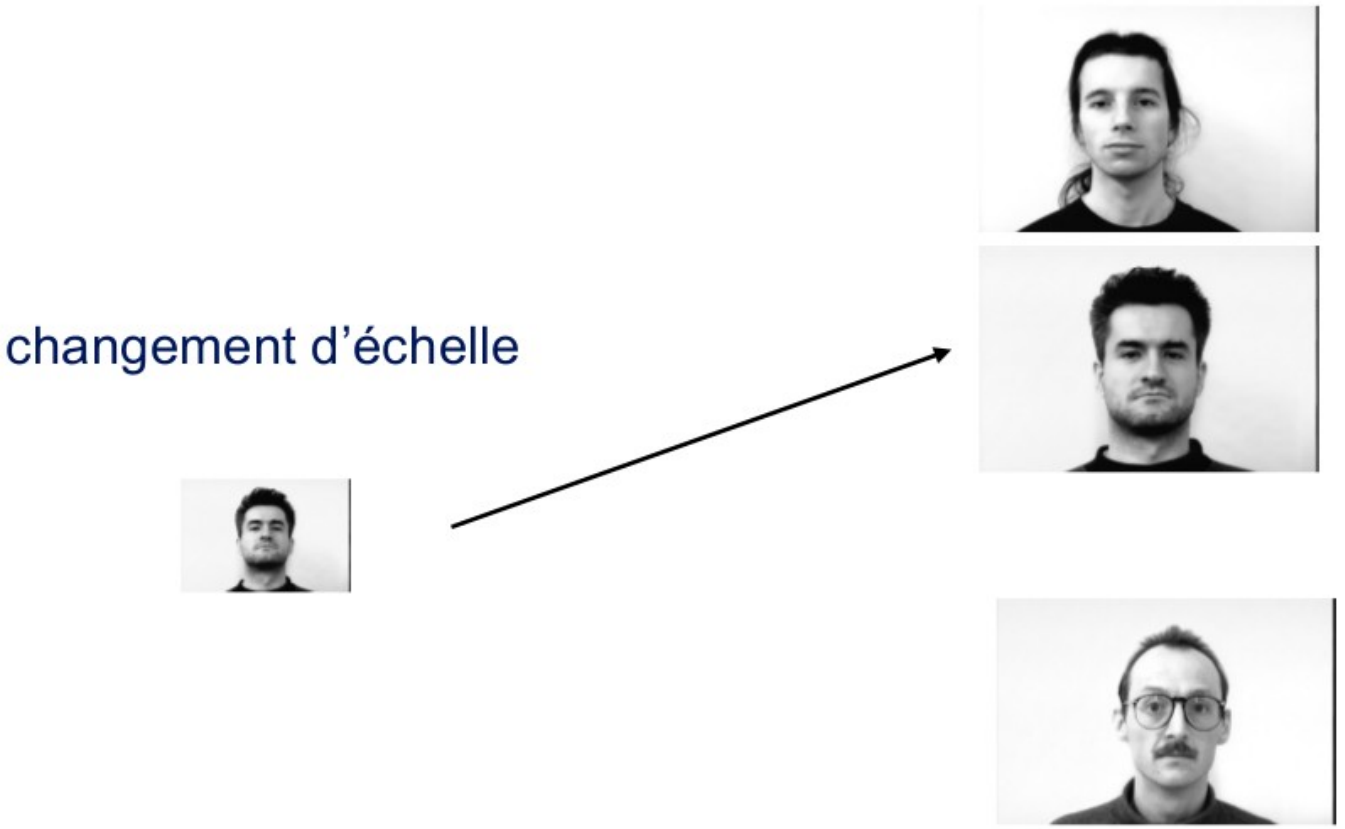
Lab preparation

- Build your own image dataset
  - 50 images
  - 10 queries, 5 results per query
  - Ground Truth for each query
  - Depict and explain in your report

- Be mindful of the challenges in image retrieval

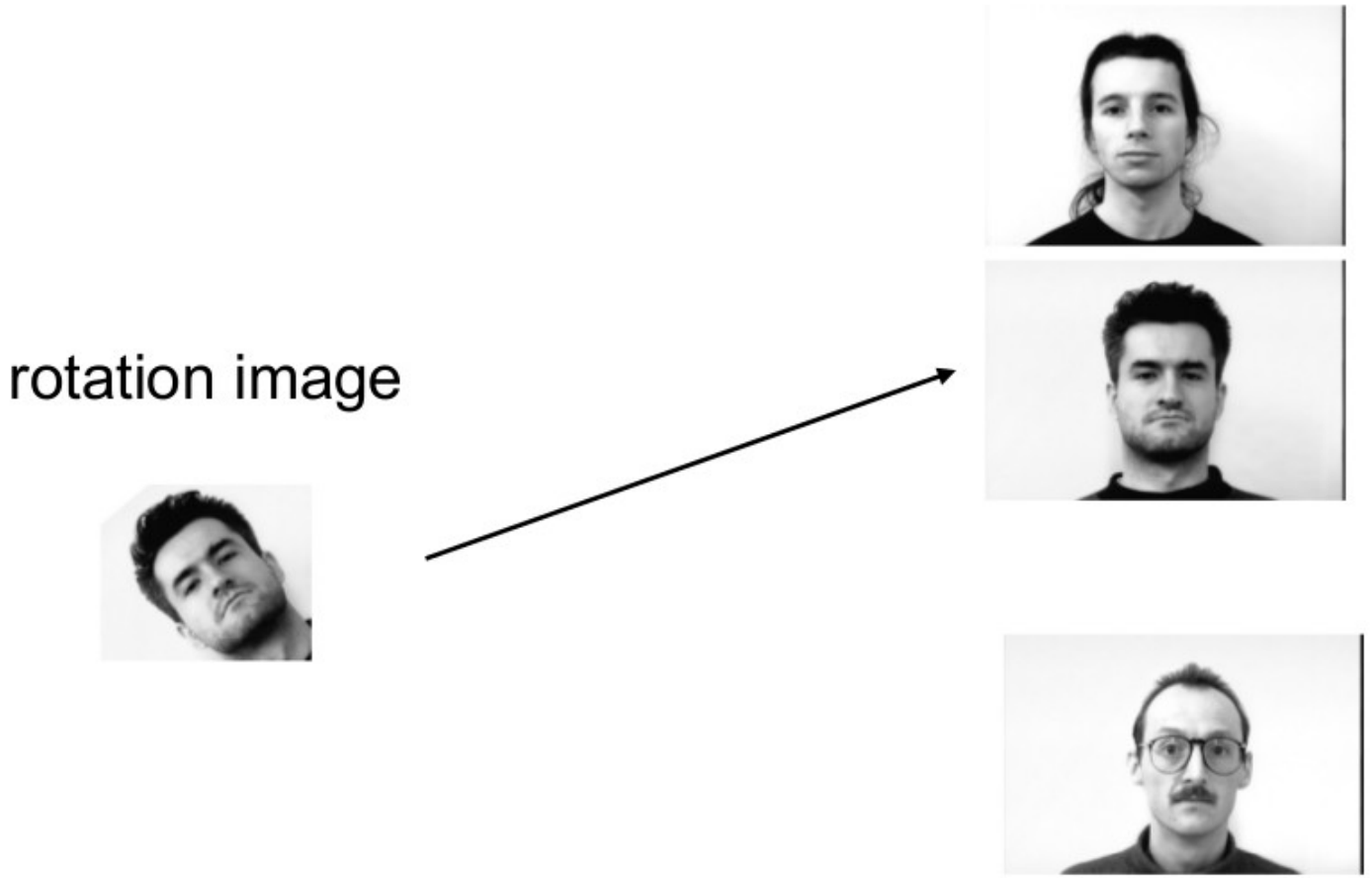
Challenges

Scaling



Challenges

Rotation

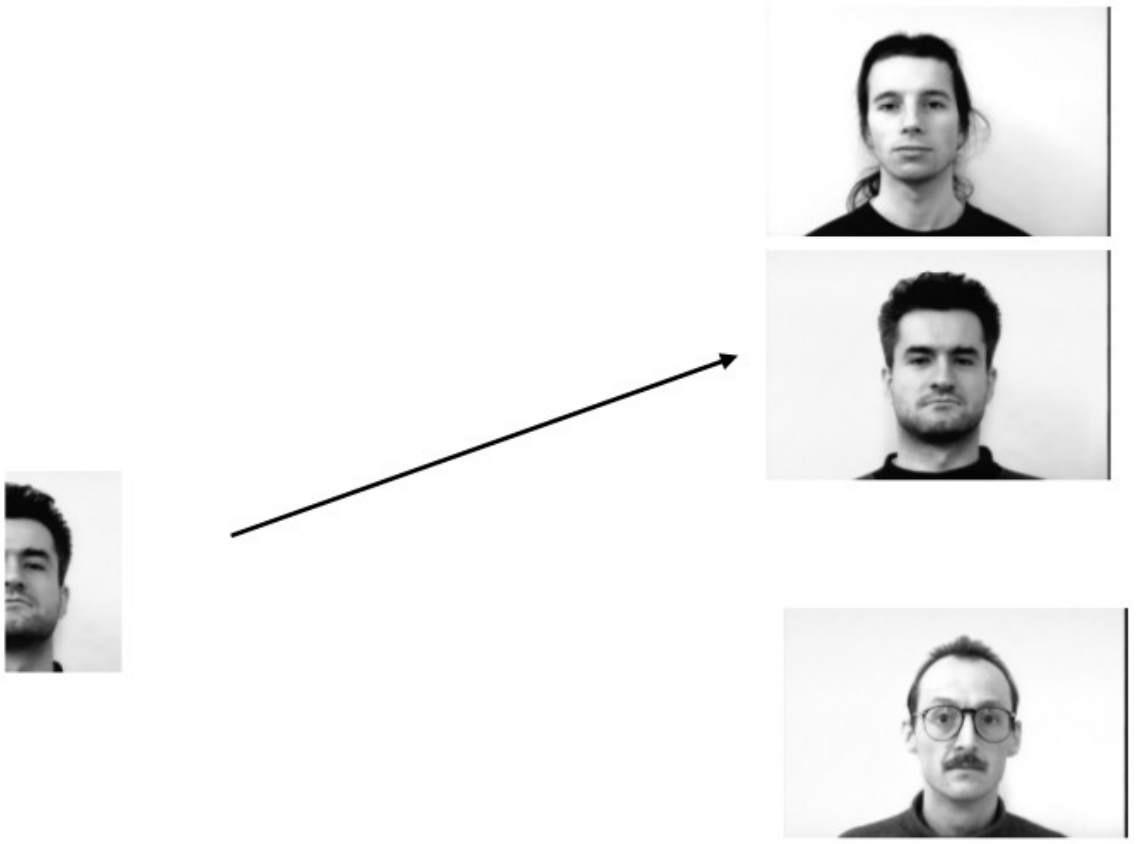


Challenges

Clutter

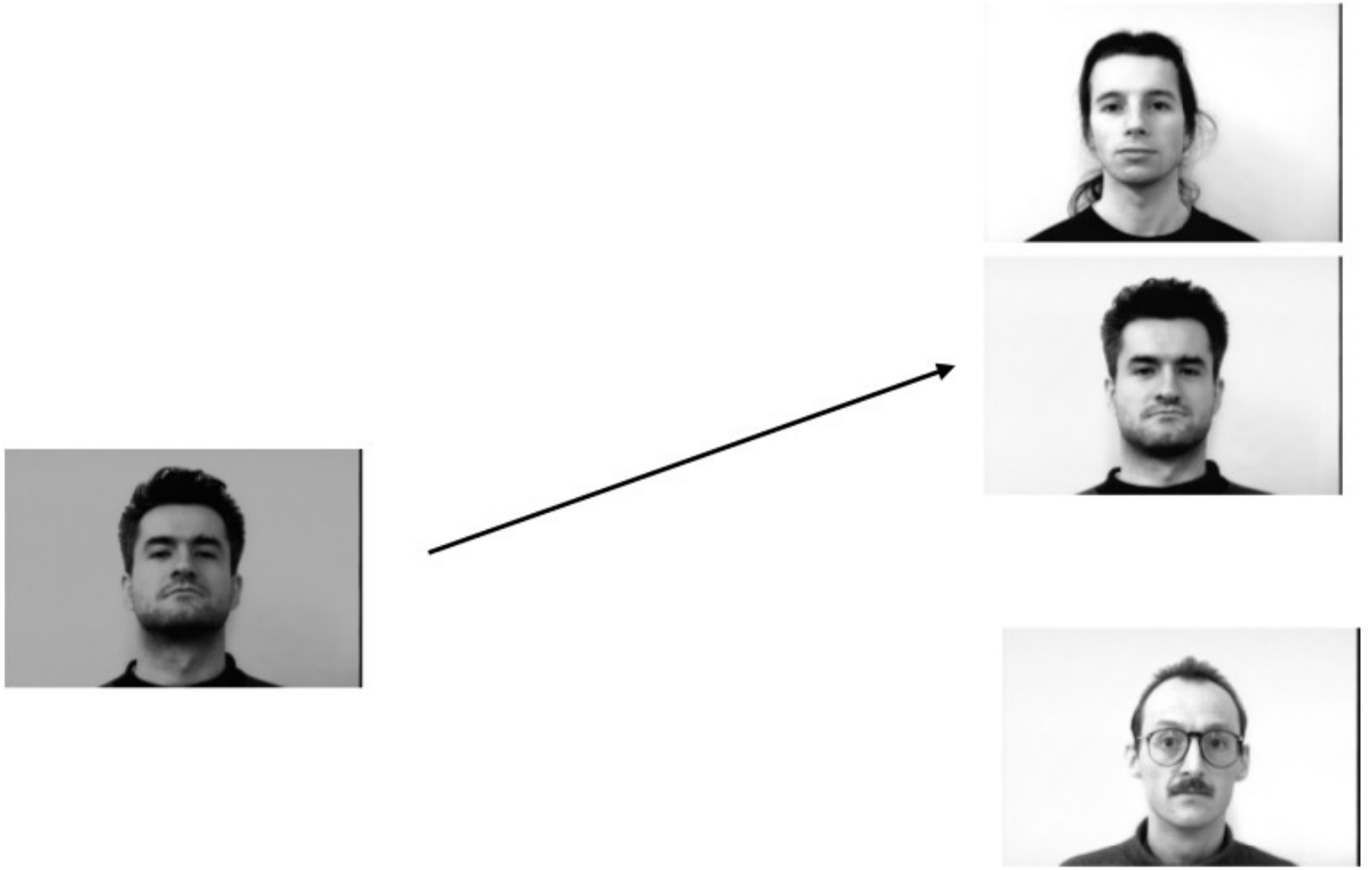
Challenges

Occlusion



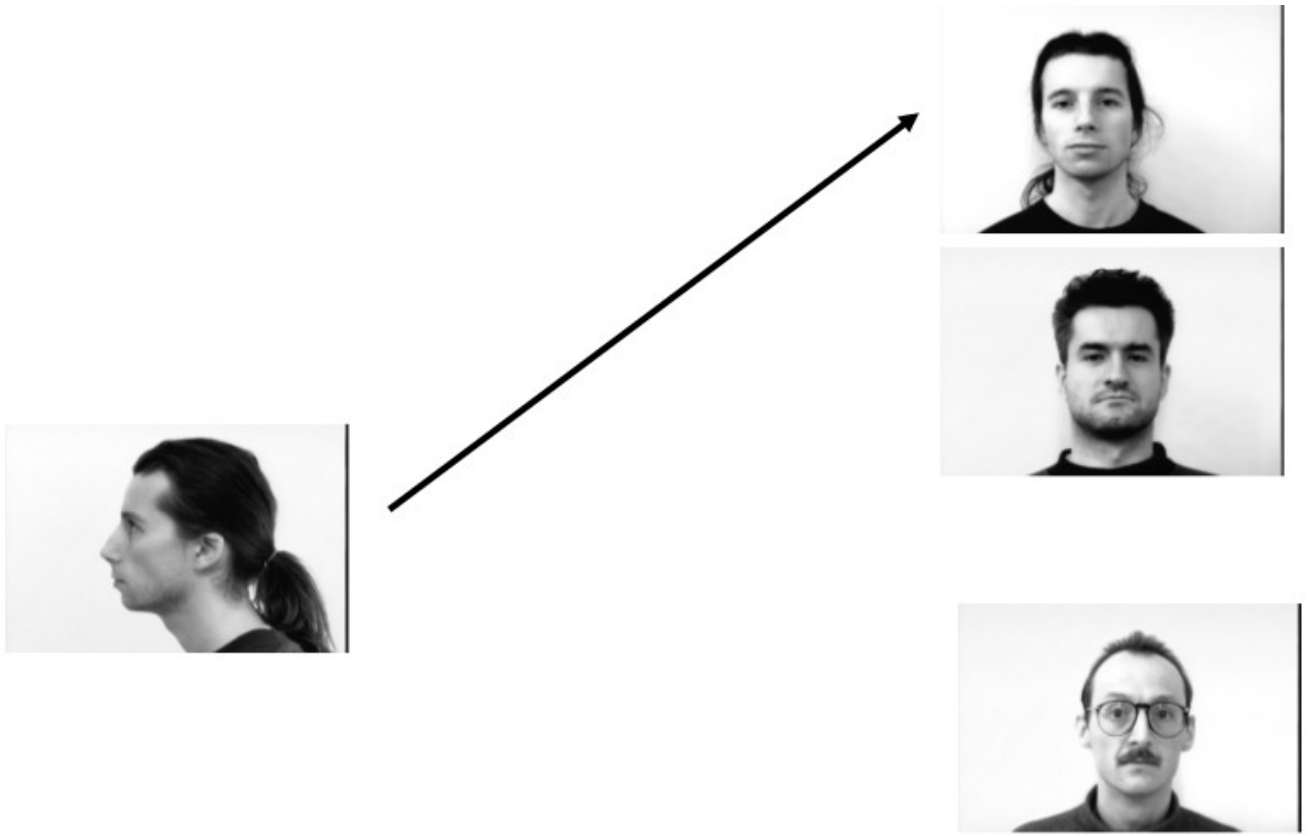
Challenges

Lightning



Challenges

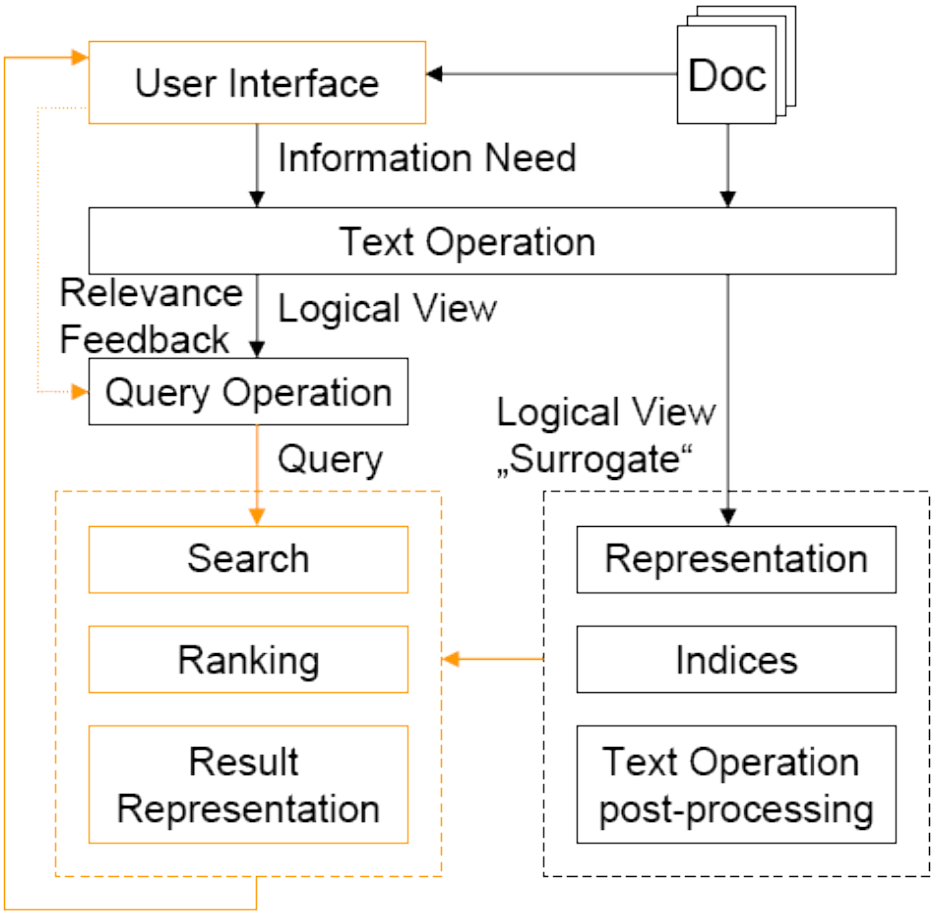
3D objects



Reminder

Aspects

- Query & languages
- IR models
- Documents
- Internal representation
- Pre- and post-processing
- Relevance feedback
- HCI



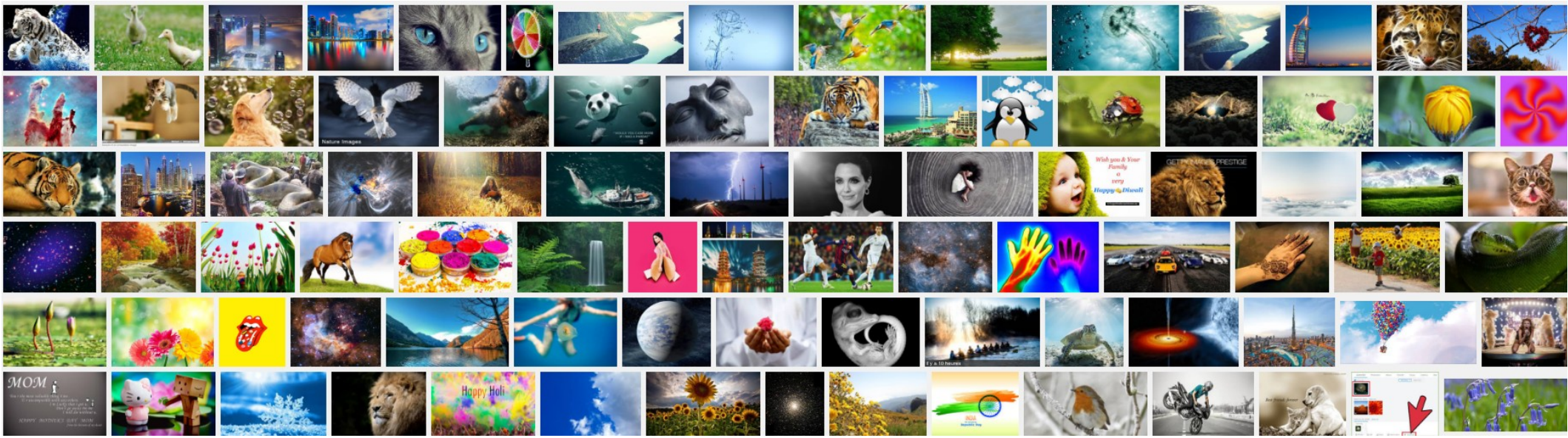
Idea

- Consider images as text document!
- Augment images with
  - Keywords
  - Metadata
  - Image description

Problem

« We now upload and share over 1.8 billion photos each day. »

(May 2014)



http://tech.firstpost.com/news-analysis/now-upload-share-1-8-billion-photos-everyday-meeker-report-224688.html

INP TOULOUSE ENSEEIHT

Possible Solutions

Possible solutions

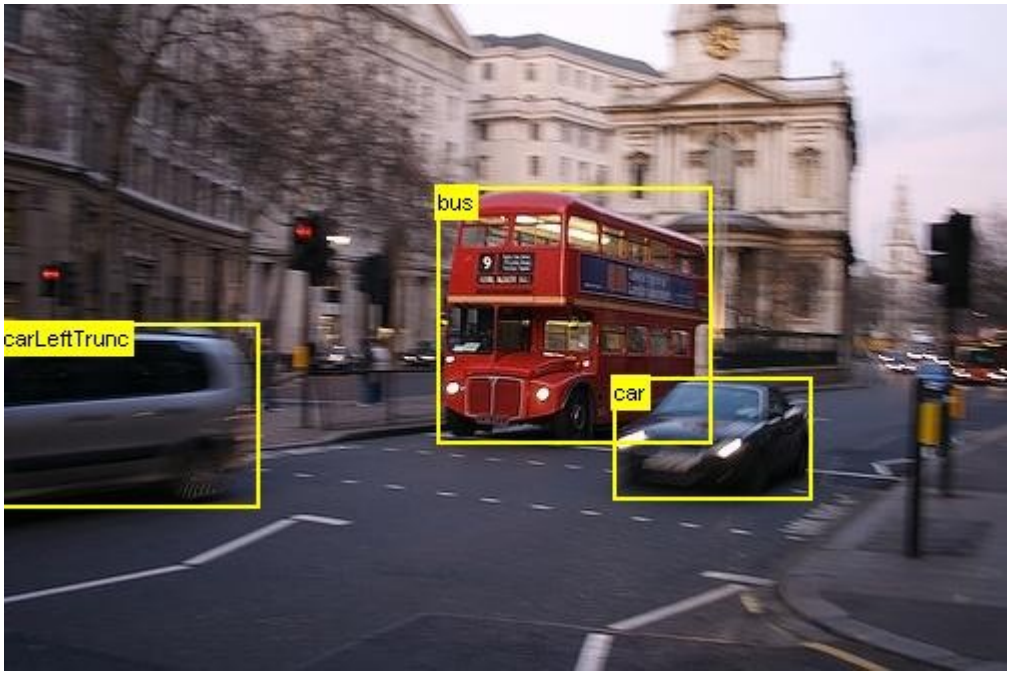
Literature

Vietnamese literature has a centuries-deep history. The country has a rich tradition of folk literature, based on the typical 6–to-8-verse poetic form named *ca dao*, which usually focuses on village ancestors and heroes.[162] Written literature has been found dating back to the 10th-century Ngô dynasty, with notable ancient authors including Nguyễn Trãi, Trần Hưng Đạo, Nguyễn Du and Nguyễn Đình Chiểu. Some literary genres play an important role in theatrical performance, such as *hát nói* in *ca trù*.[163] Some poetic unions have also been formed in Vietnam, such as the Tao Đàn. Vietnamese literature has in recent times been influenced by Western styles, with the first literary transformation movement – Thơ Mới – emerging in 1932.[164]

The Temple of Literature in Hanoi.

Other possibilities

- Object detection and recognition



PASCAL VOC Dataset

Other possibilities

- Image segmentation

Overall

Image annotation remains a challenge because

- **Big Data** (too much images to be annotated manually)
- Data **Heterogeneity** (too diverse images and objects to be annotated automatically)

Content Based Image Retrieval

- Motivation & Semantic Gap
- Local features based architecture
  - Feature detection
  - Feature description
  - Feature matching

INP ENSEEIHT

Motivation

Lots of good reasons ...

- Visual information overload
  - Devices (cameras, mobile phones, etc.)
  - Communication (email, mo-blogs, etc.)
- Metadata not available
  - Time consuming
  - No automation

Semantic Gap

- Defined as
  - Inability of automatic understanding
  - Gap between high- and low-level features / metadata
- Actually hard task for humans also



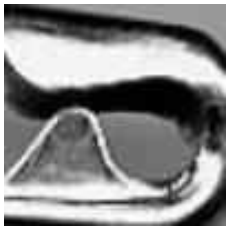



Image Similarity

Are these two images similar?

Applications

- Home User & Entertainment
  - Find picture of / from / at ....
  - Search & browse personal digital library
- Graphics & Design
  - Find picture representing something (Color in CD/CI, feeling, etc.)
- Medical Applications
  - Find images for diagnosis, documentation
- And many more (biology, advertisement, etc.)

Image Retrieval classic Architecture

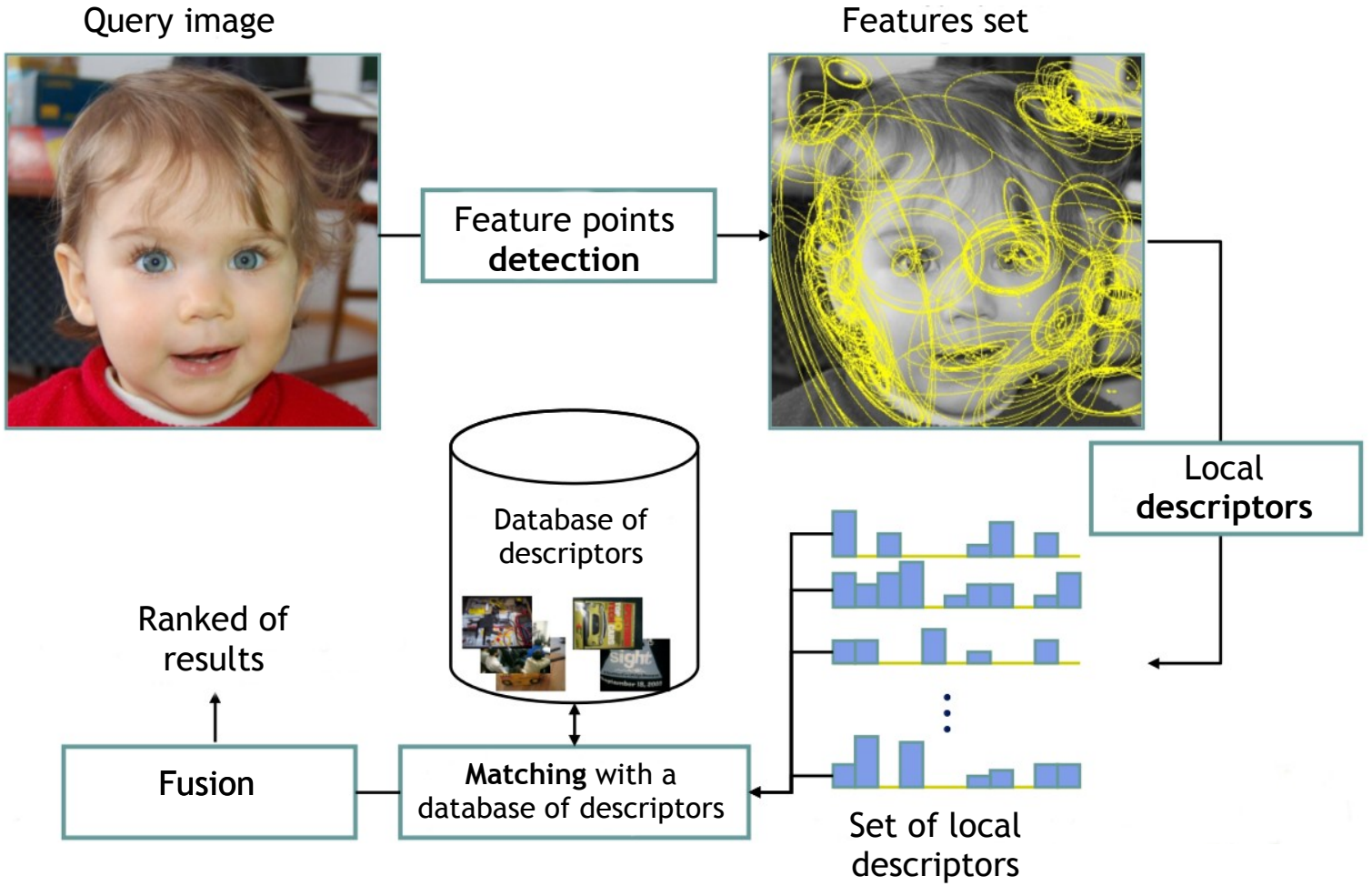
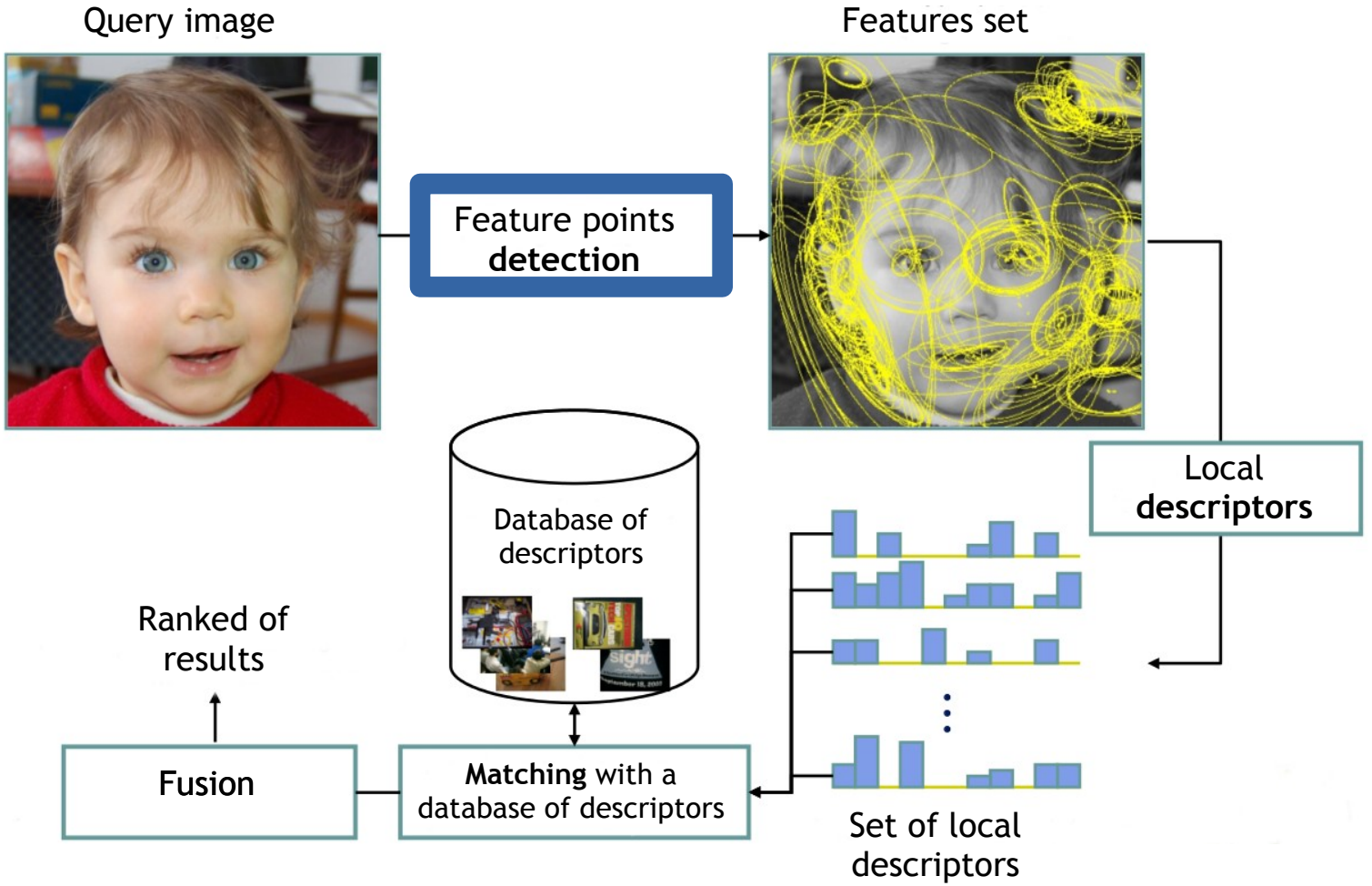


Image Retrieval classic Architecture



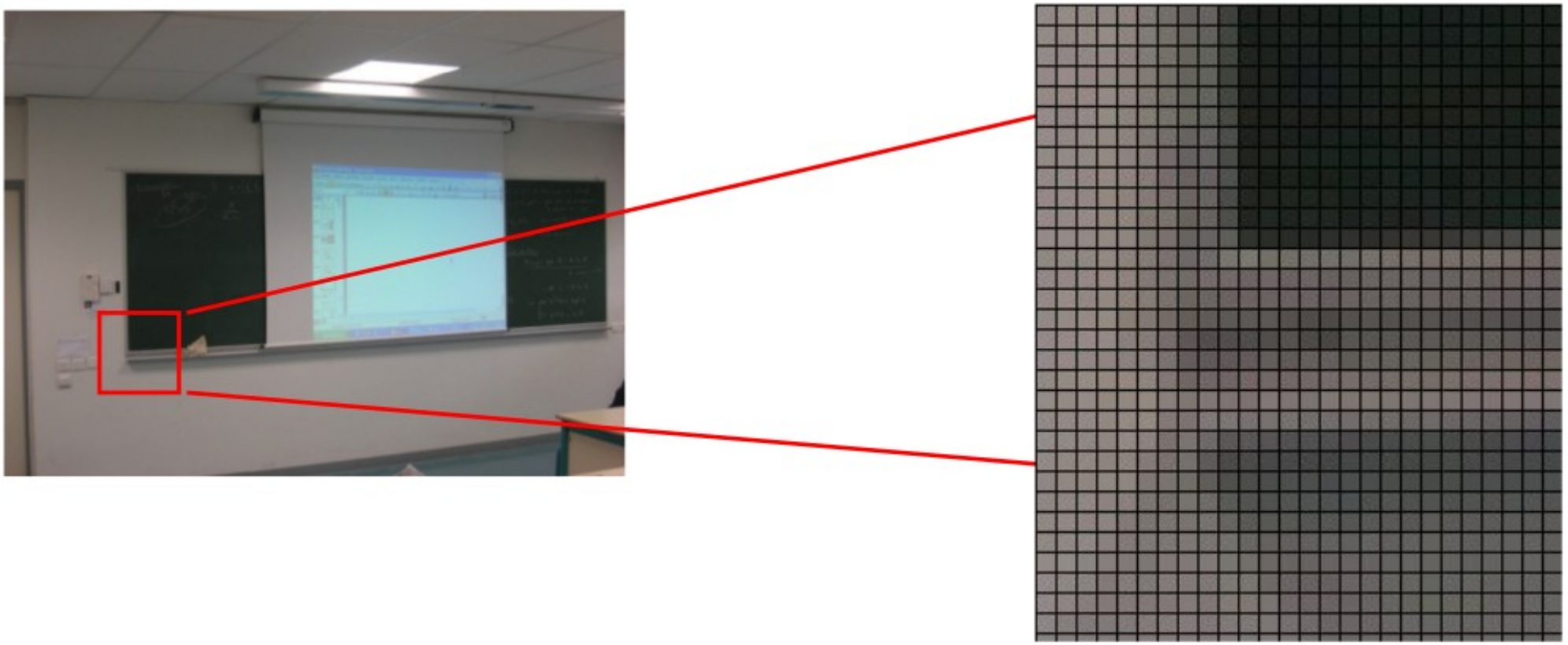
Feature points detector

Objective:

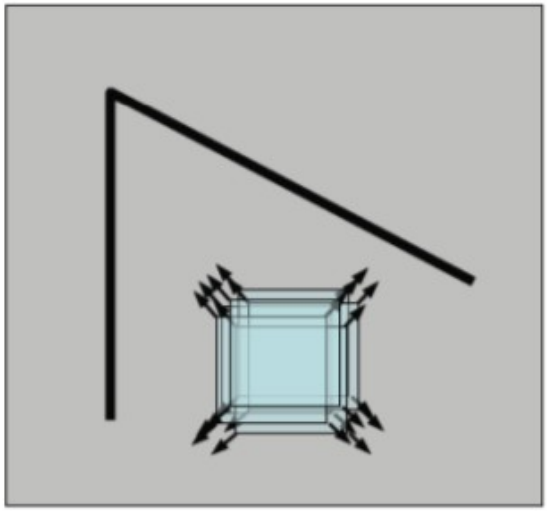
- Find points that have **particular** local properties, so that we can find them in several different images
- The properties shoud be robust to **geometric** transformations such as translations, rotations and **photometric** transformations such as lightning changes.

The Harris detector

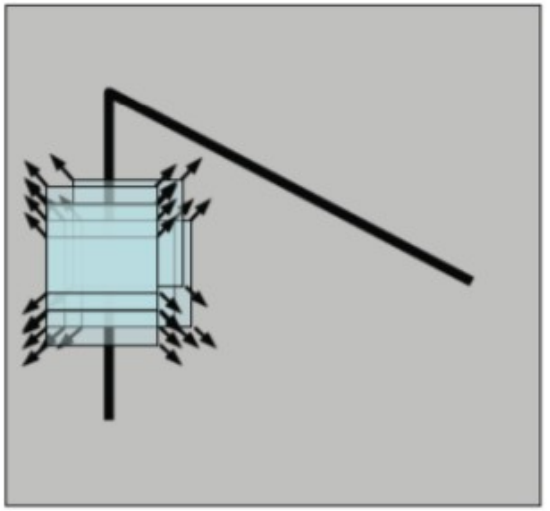
- Finds « corners »



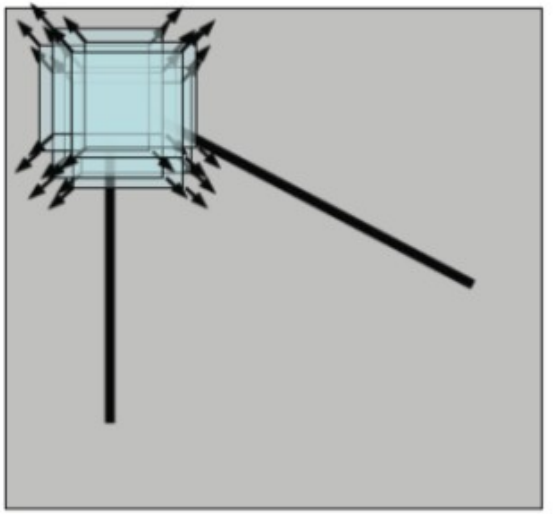
Corner detection



« Flat » region:
No change in
all directions

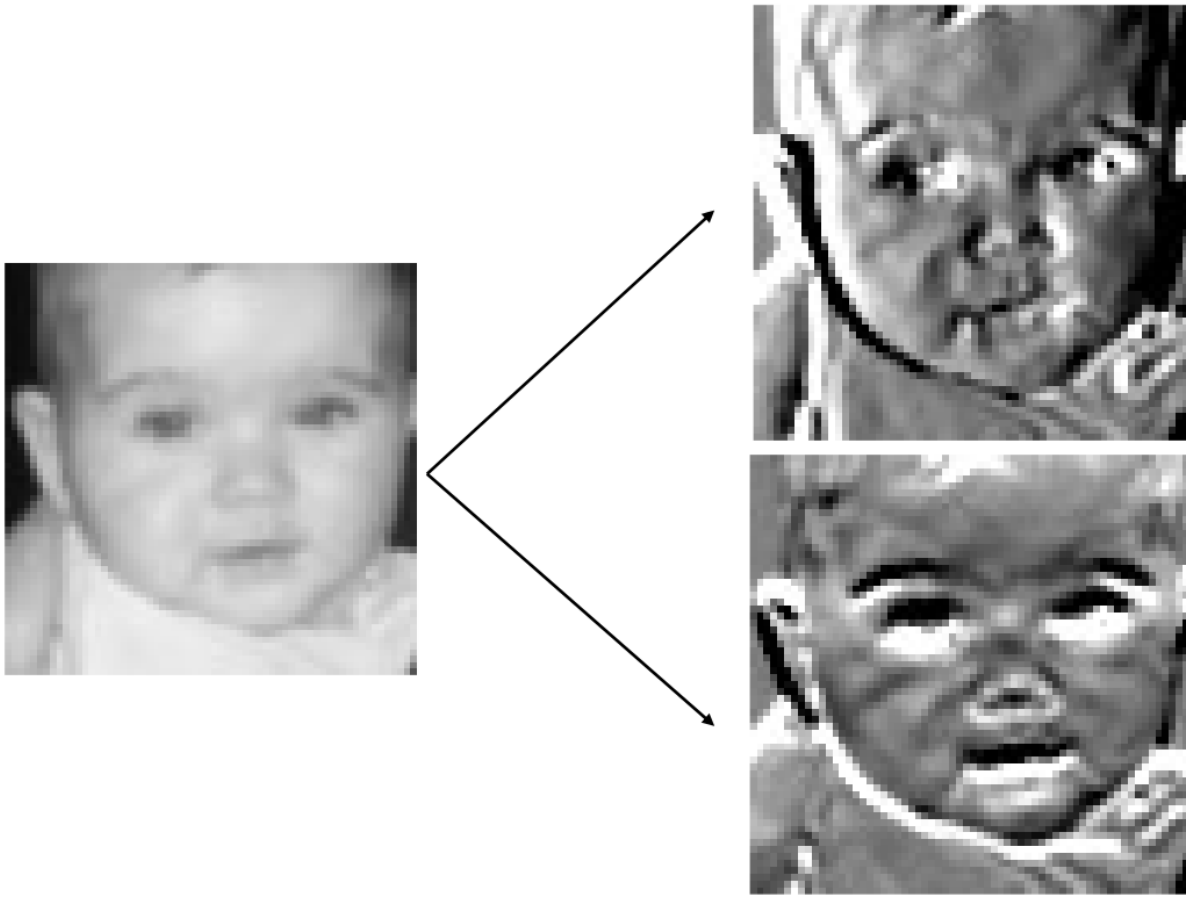


«Edge» region:
No change along
the edge direction



«Corner» region:
Significant change in
all directions

Gradients



Horixontal gradient
(x-axis)

Vertical gradient
(y-axis)

Harris algorithm

1. Compute x and y derivatives of image

$$I_x = G_\sigma^x * I \quad I_y = G_\sigma^y * I$$

2. Compute products of derivatives at every pixel

$$I_{x2} = I_x.I_x \quad I_{y2} = I_y.I_y \quad I_{xy} = I_x.I_y$$

3. Compute the sums of the products of derivatives at each pixel

$$S_{x2} = G_{\sigma\prime} * I_{x2} \quad S_{y2} = G_{\sigma\prime} * I_{y2} \quad S_{xy} = G_{\sigma\prime} * I_{xy}$$

4. Define at each pixel (x, y) the matrix

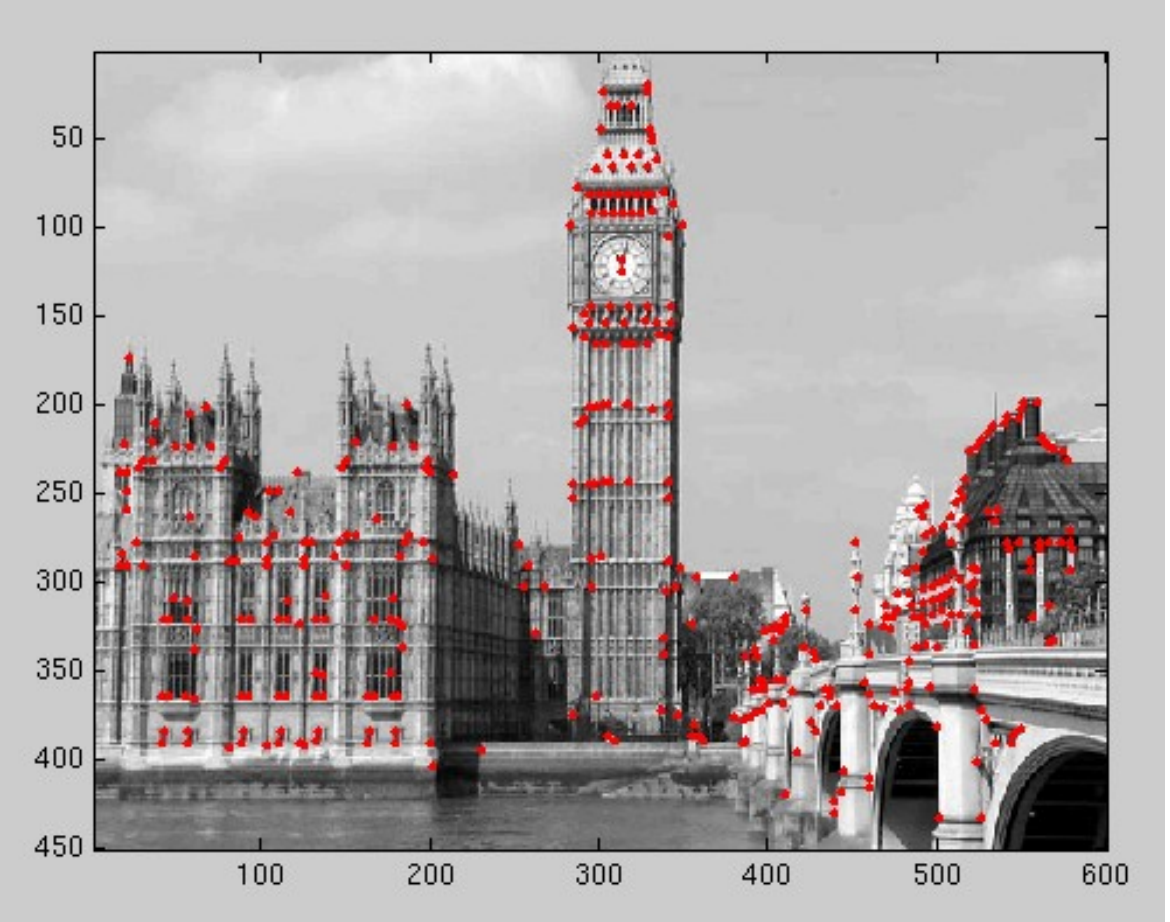
$$H(x,y) = \begin{bmatrix} S_{x2}(x,y) & S_{xy}(x,y) \\ S_{xy}(x,y) & S_{y2}(x,y) \end{bmatrix}$$

5. Compute the response of the detector at each pixel

$$R = Det(H) - k(Trace(H))^2$$

6. Threshold on value of R. Compute nonmax suppression.

Harris – typical result



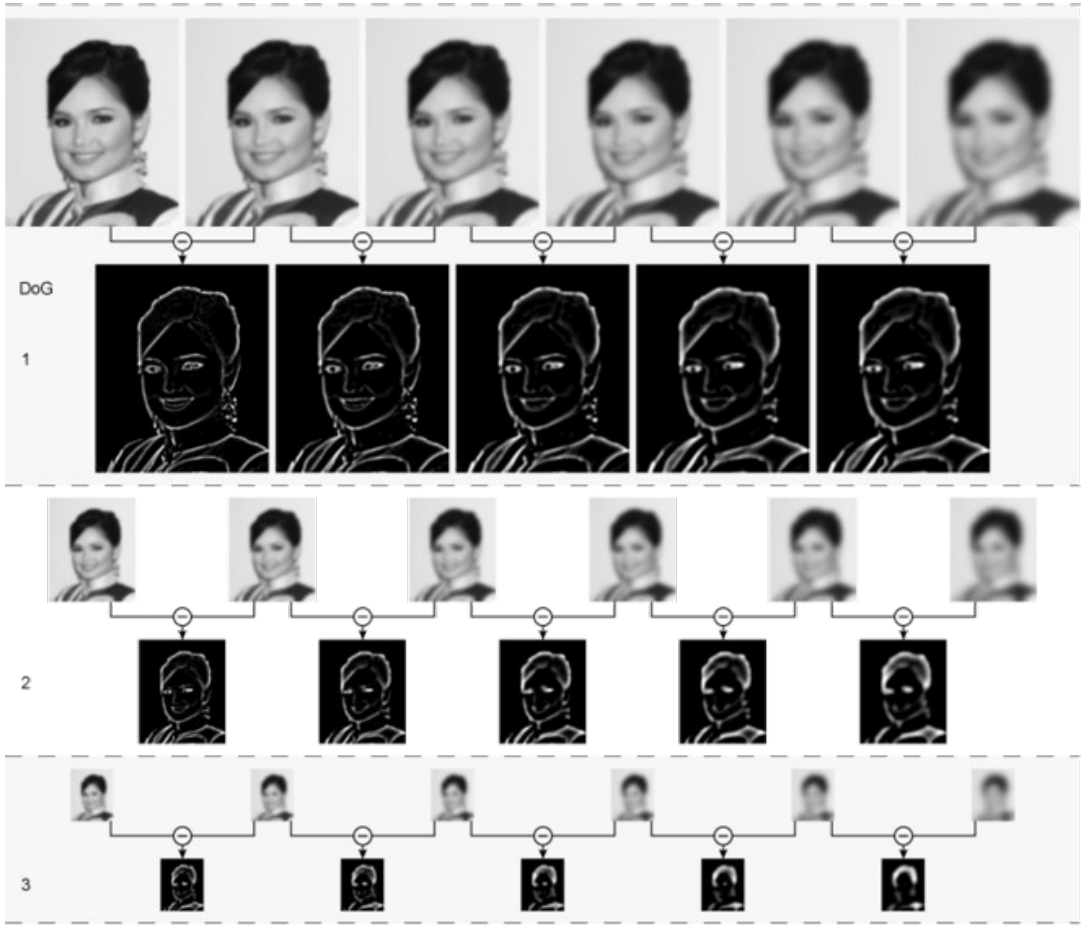
Harris detector

- Very simple to implement
- Has good properties (invariant to rotation)
- But limited:
  - Not invariant to scale

SIFT detector

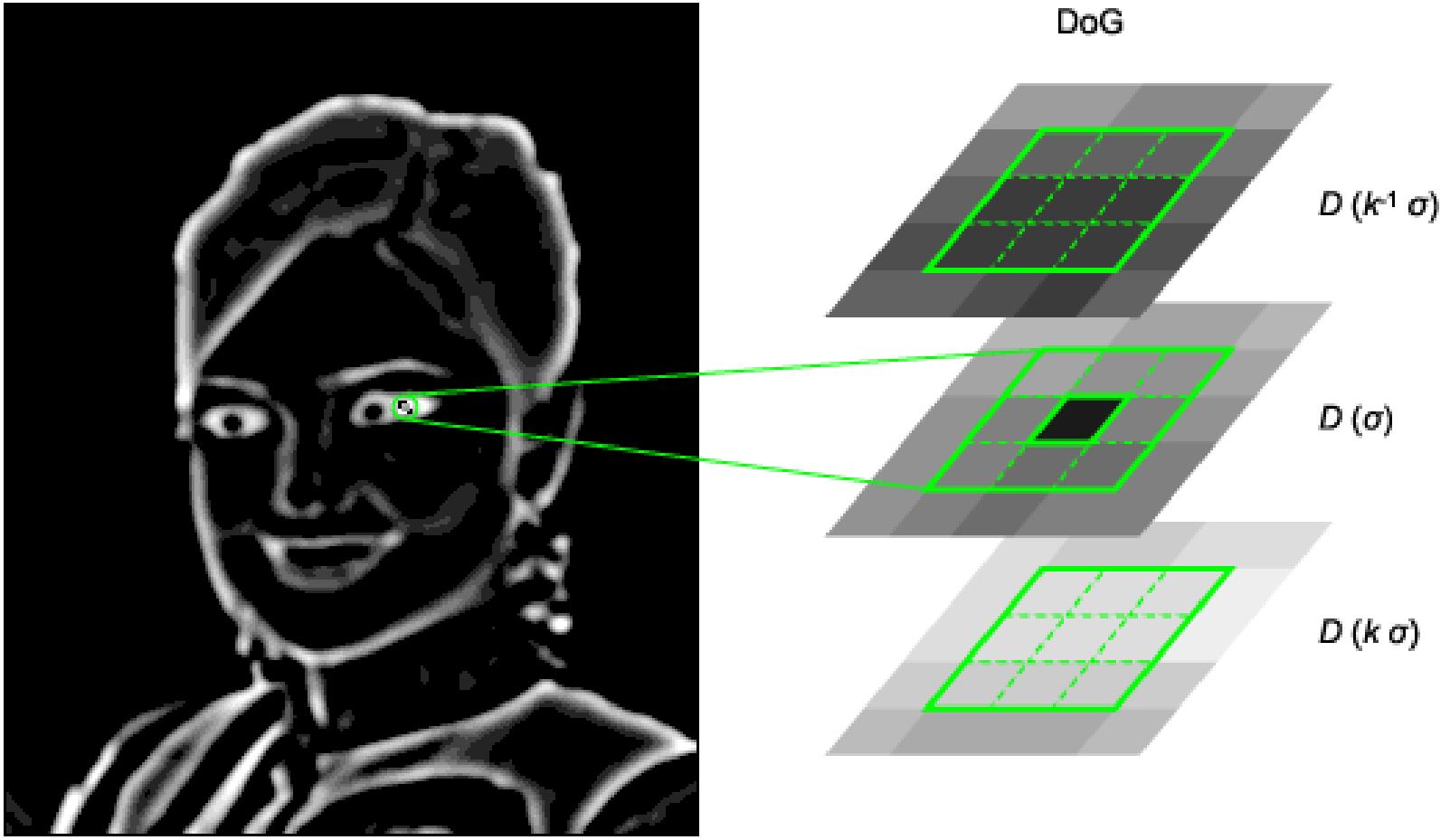
- Widely used
- Invariant to scale by construction
- Scale Invariant Feature Transform

SIFT - Principle

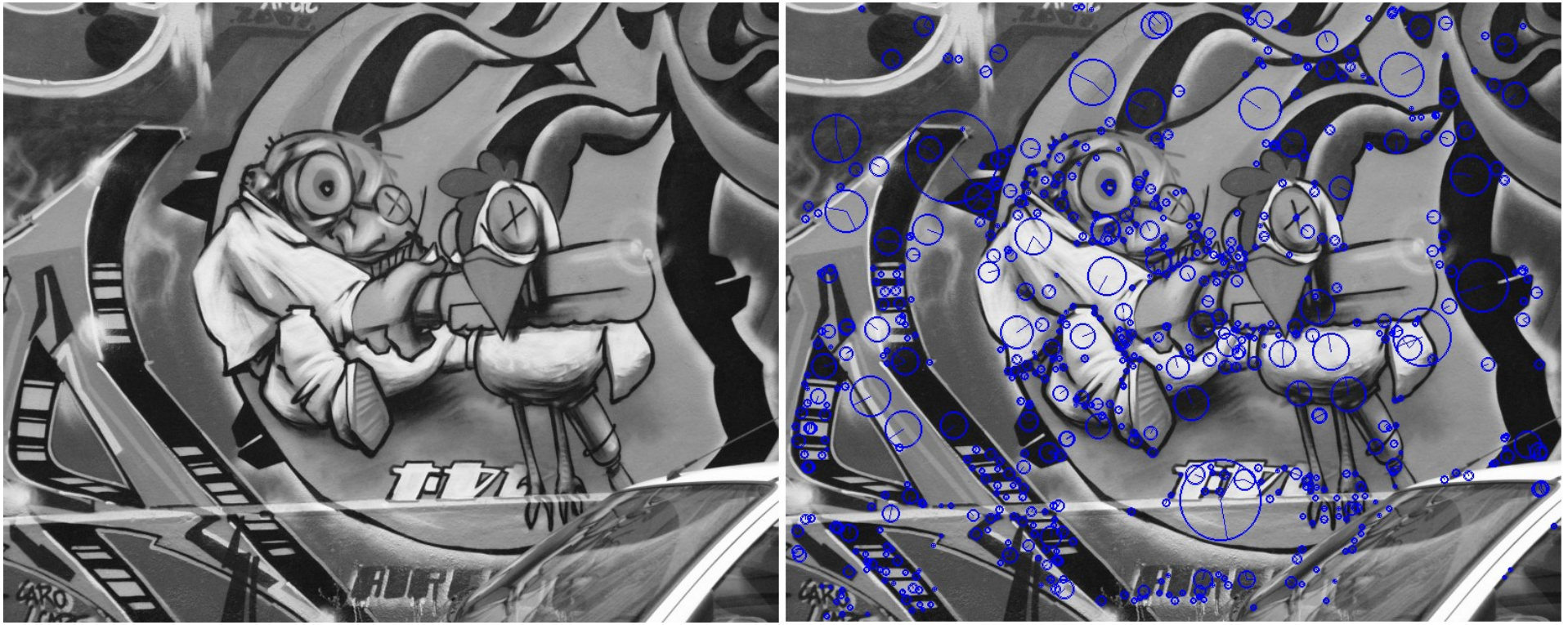


- Different scales
- Different level of details

SIFT - Principle



SIFT - output



Points coordinates + scale

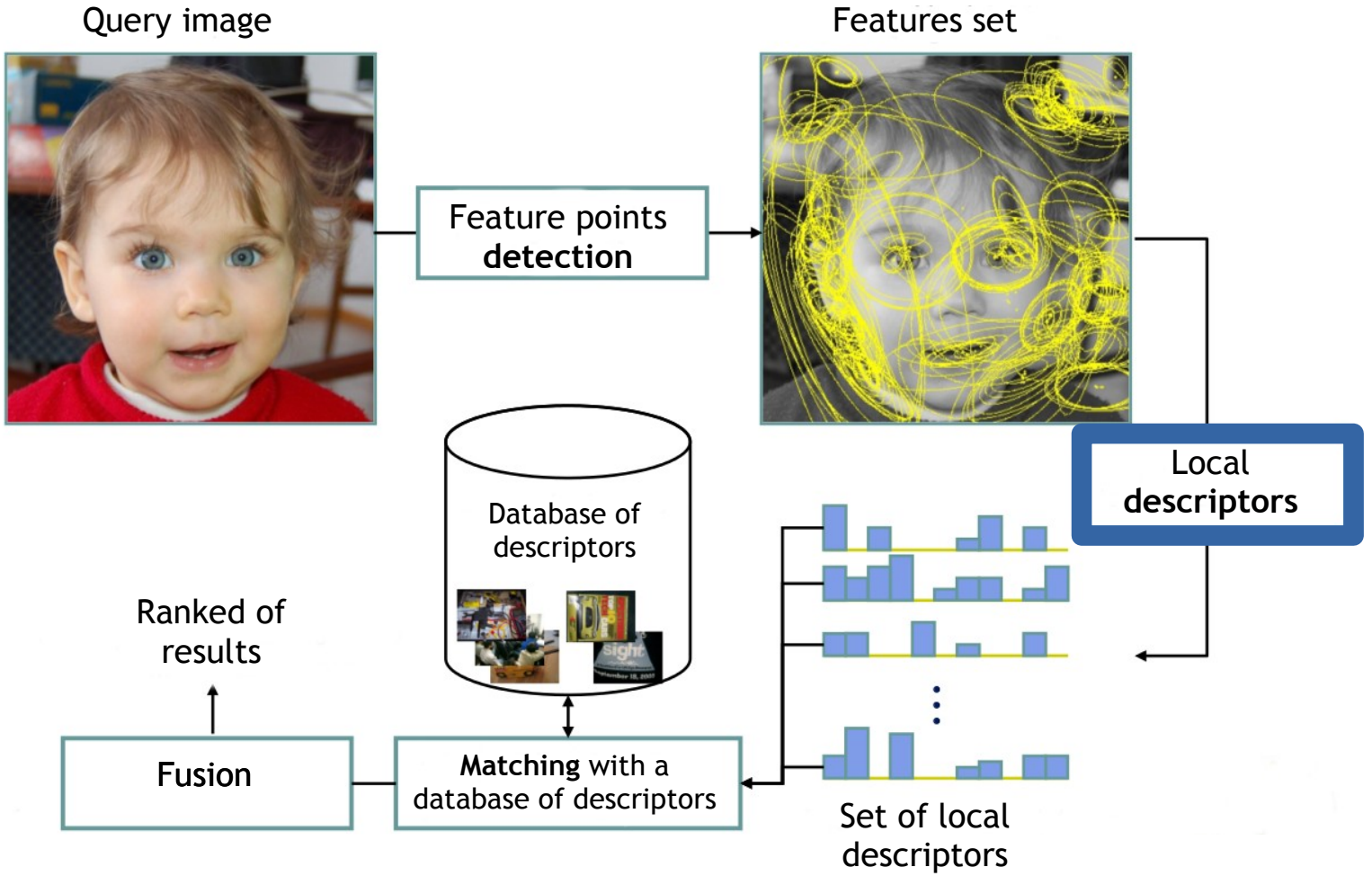
SIFT – Pros and Cons

- Pros
  - Robust to most of the existing transformations
- Cons
  - Lot of results (need to filter points)
  - Runtime

Other feature detectors

- SURF: Speeded-Up Robust Features
  - Similar to SIFT but lower runtime
- Hessian Affine
- EBR: Edge-Based Region Detector
- IBR: Intensity Extrema-Based Regions Detector
- Etc.

Image Retrieval classic Architecture

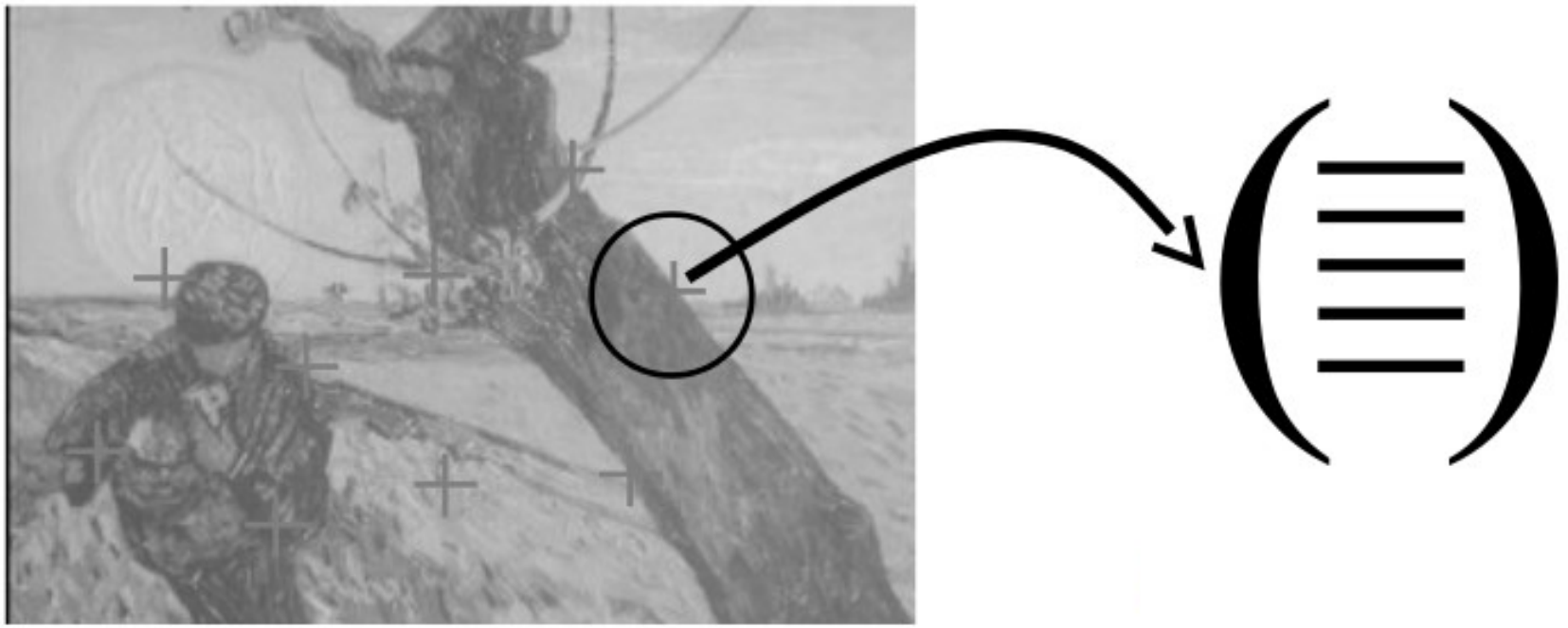


Local features descriptors

Goal:

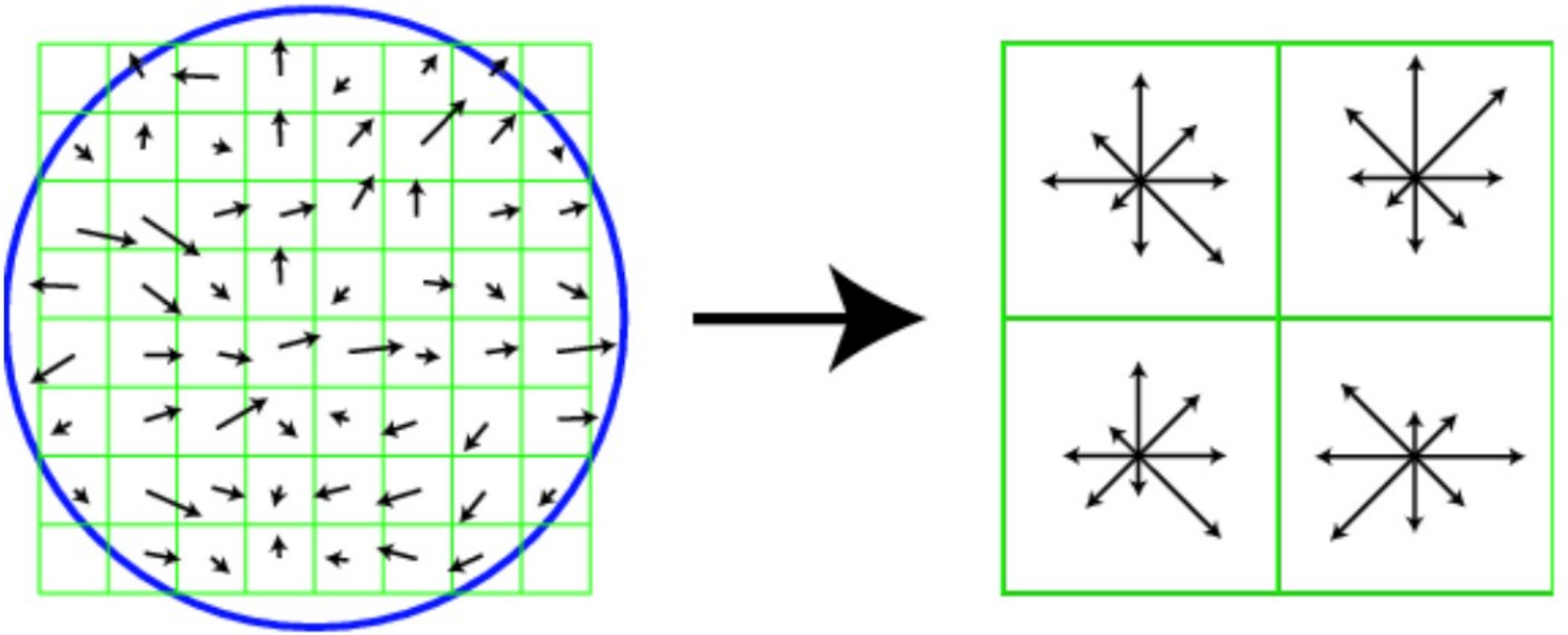
- For a detected feature, describe it so that it can be found again in another image
- Has to be robust to photometric and geometric transformations

The simplest descriptor



Dimension?

SIFT descriptor

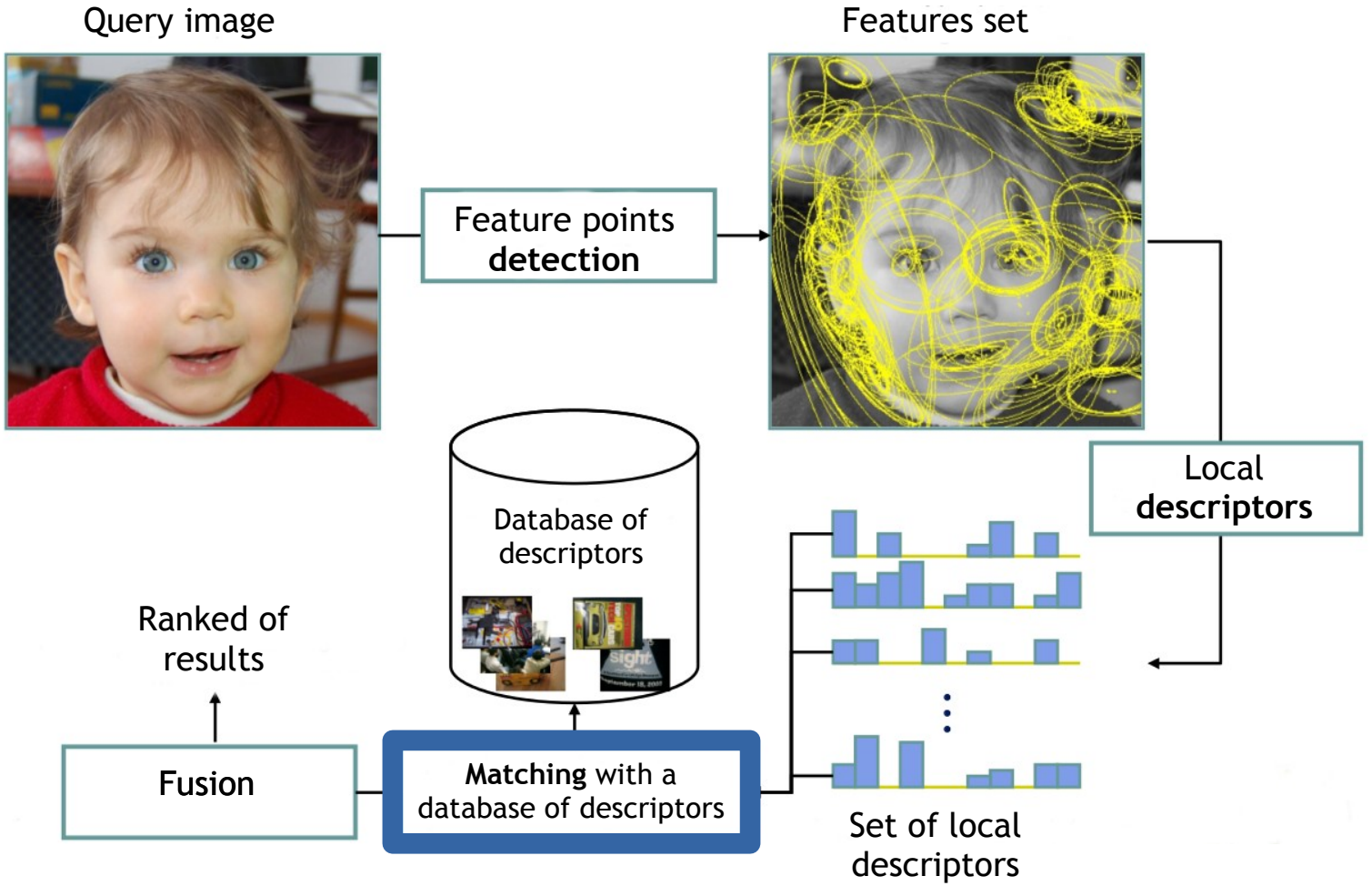


Dimension?

Many more descriptors

- Mostly differential descriptors
  - HOG: Histogram Of Gradients
  - GLOH: Gradient Location and Orientation Histogram
  - SURF
  - Etc.

Image Retrieval classic Architecture



Matching descriptors

$x = (x_1,\ldots,x_i,\ldots,x_n) \in R^n$

Euclidean distance (distance L2)

$$d(x,y)=\sqrt{\sum_i (x_i-y_i)^2}$$

Manhattan distance (distance L1)

$$d(x,y)=\sum_i |x_i-y_i|$$

Minkowski distance (p-distance)

$$d(x,y)=\sqrt[p]{\sum_i (x_i-y_i)^p}$$

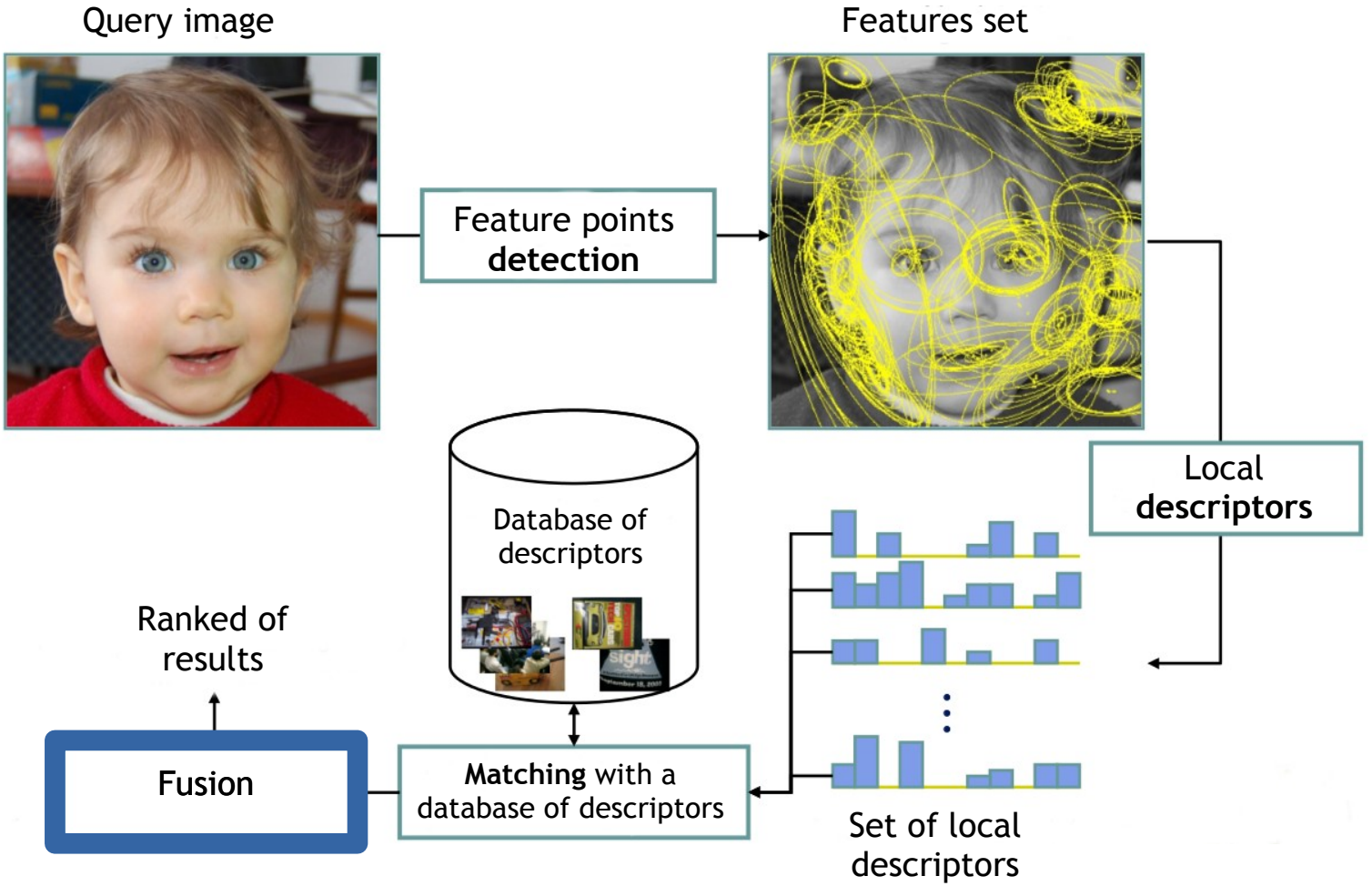
Infinite distance

$$d(x,y)=\max_i |x_i-y_i|$$

Matching descriptors

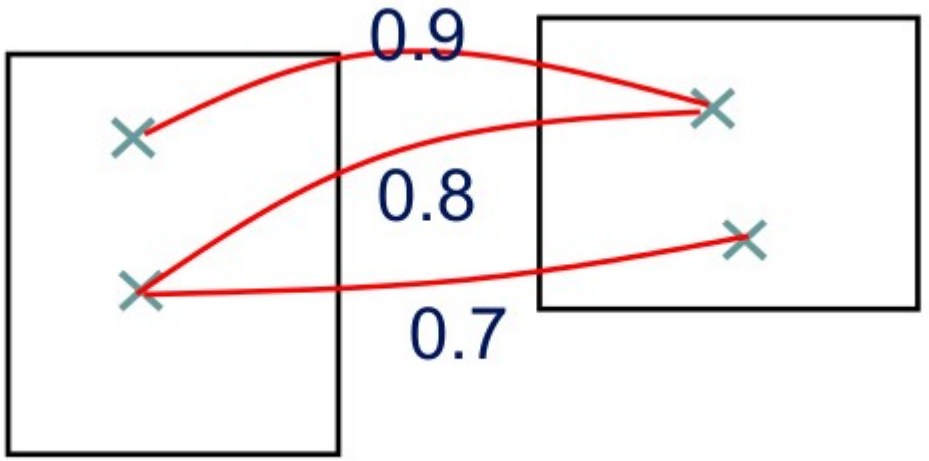
- Need Ground Truth matched points on pairs of images
- Learn the best threshold to classify a pair into match / no match

Image Retrieval classic Architecture



Fusion

- Matching will typically get results such as:



Solution: Winner-takes-all

Fusion

- Matching will typically get results such as:

Solution: Epipolar geometry

Fusion

- Matching will typically get results such as:

Solution: Epipolar geometry

Conclusion

- Local descriptors can be very precise
- Image recognition work well with the proper descriptors and filtering techniques
- **BUT**
  - **Painfully slow**