Computer Vision – TP11
Local Invariant Descriptors

Miguel Tavares Coimbra

Acknowledgement: Slides adapted from Kristen Grauman

U. PORTO FC

Outline

- Detection of interest points
- Local invariant descriptors
- Classification using visual words

Topic: Detection of interest points

- Detection of interest points
- Local invariant descriptors
- Classification using visual words

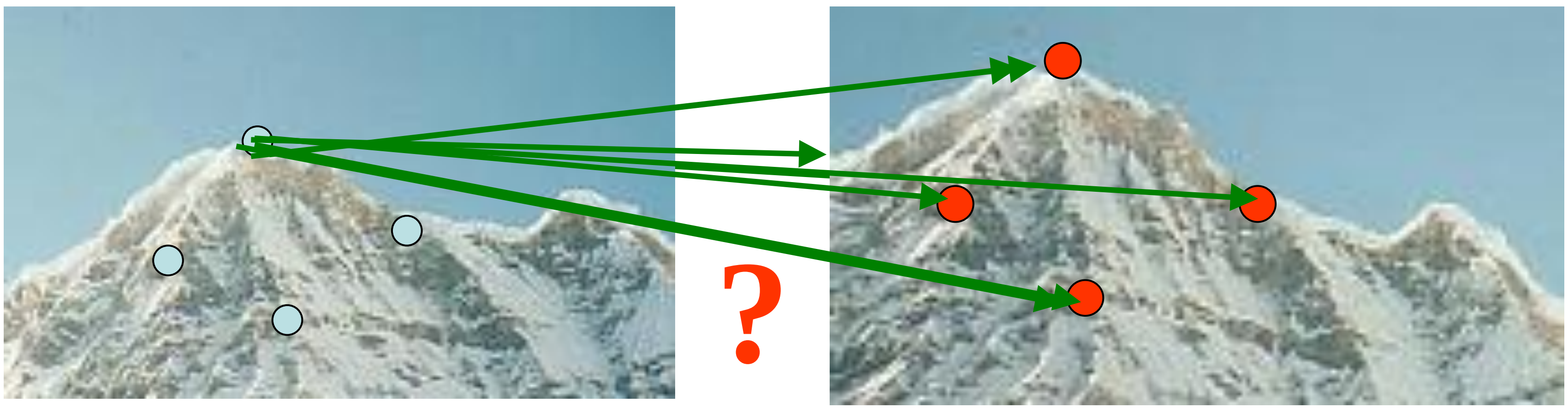
Motivation: Same interest points

- We want to detect the same points in both images

No chance to find true matches!

Motivation: ‘Unique’ descriptor per interest point

- We want to match the same interest points
- Need a descriptor invariant to geometric and photometric differences

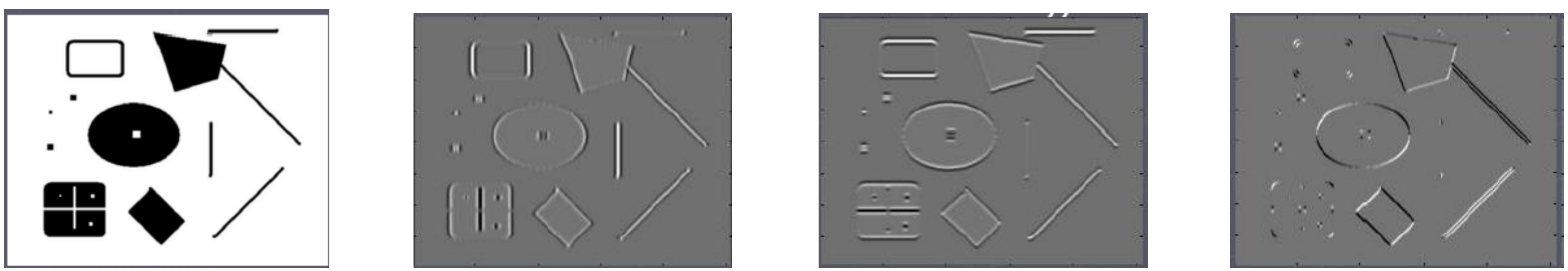


U. PORTO FC

Corners are distinctive interest points

$$M = \sum w(x,y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix}$$

2 x 2 matrix of image derivatives (averaged in neighborhood of a point)



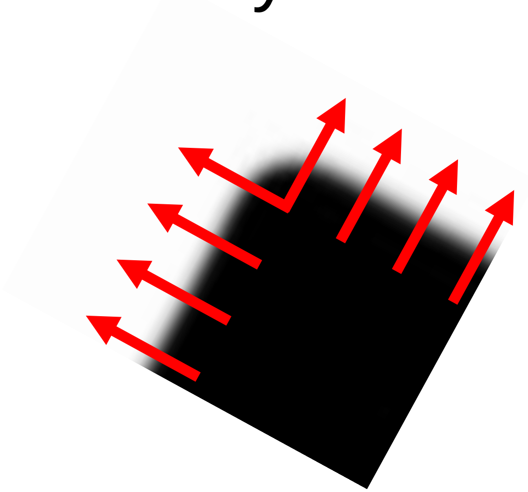
Notation: $I_x \Leftrightarrow \frac{\partial I}{\partial x}$ $I_y \Leftrightarrow \frac{\partial I}{\partial y}$ $I_x I_y \Leftrightarrow \frac{\partial I}{\partial x}\frac{\partial I}{\partial y}$

Gradient strength

Since *M* is symmetric, we have

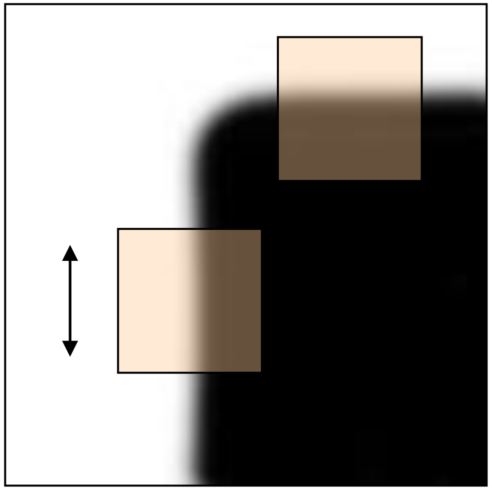
$$M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$$

$$Mx_i = \lambda_i x_i$$



The *eigenvalues* of *M* reveal the amount of intensity change in the two principal orthogonal gradient directions in the window

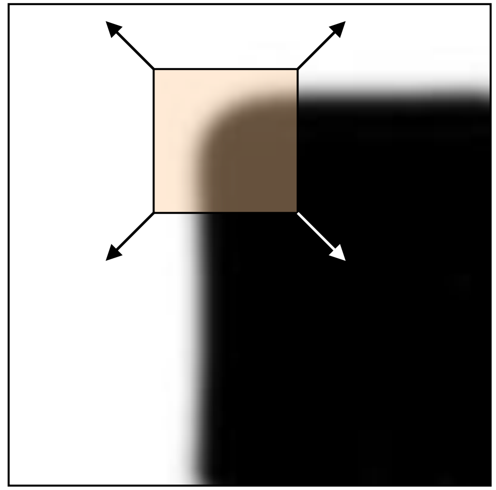
Scoring ‘cornerness’



“edge”:

$\lambda_1 >> \lambda_2$

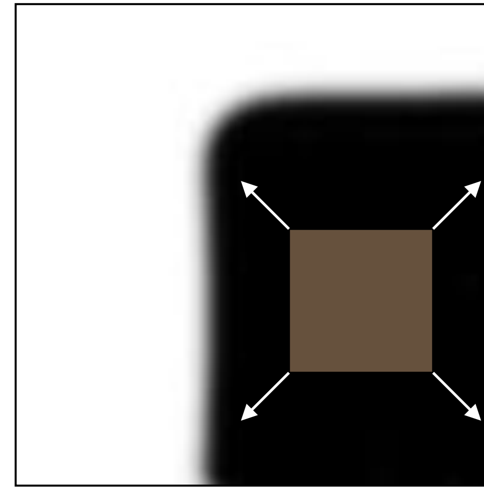
$\lambda_2 >> \lambda_1$



“corner”:

λ_1 and λ_2 are large,

$\lambda_1 \sim \lambda_2$;



“flat” region

λ_1 and λ_2 are small;

One way to score the cornerness:

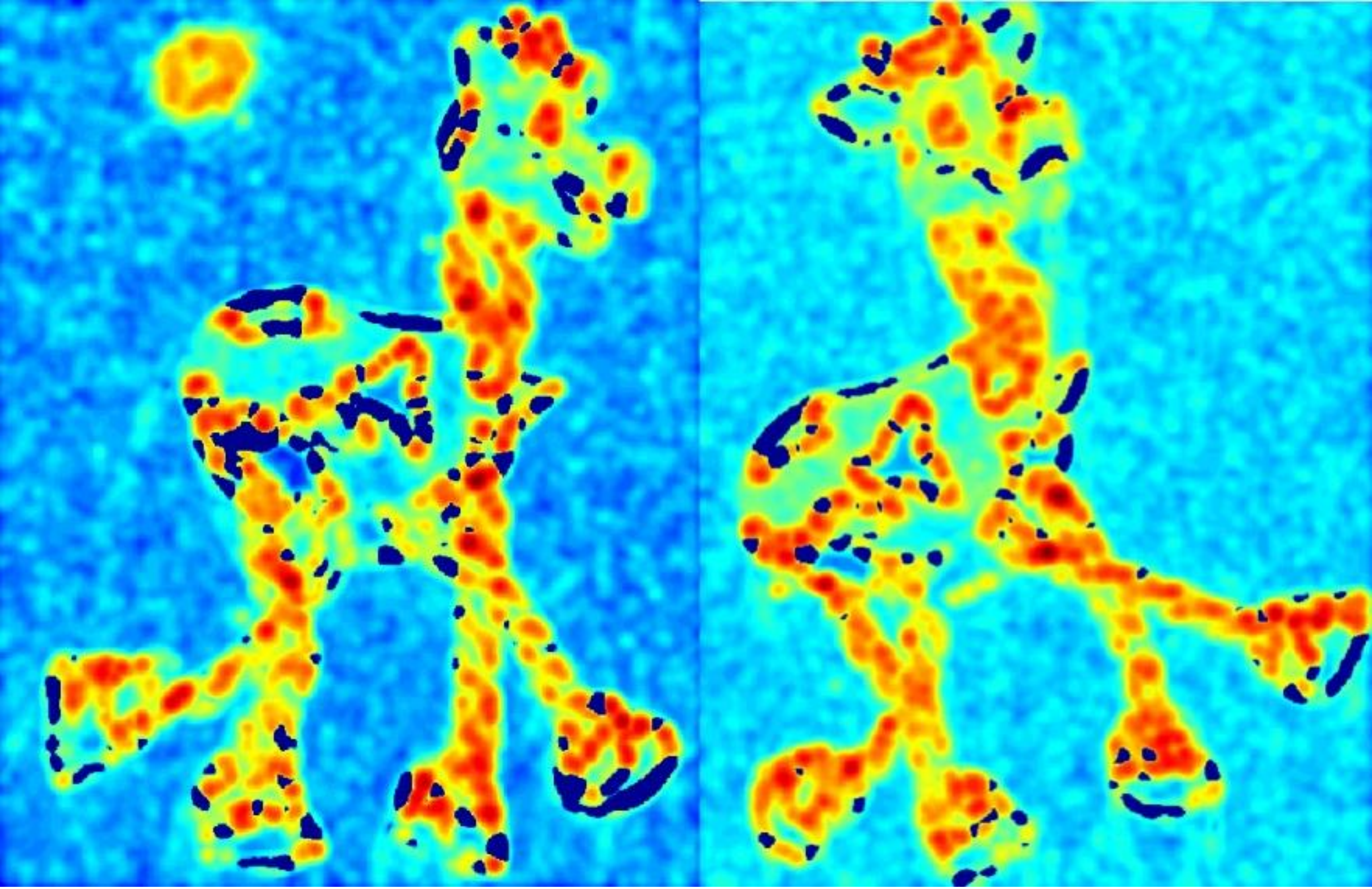
$$f = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$$

U.PORTO FC

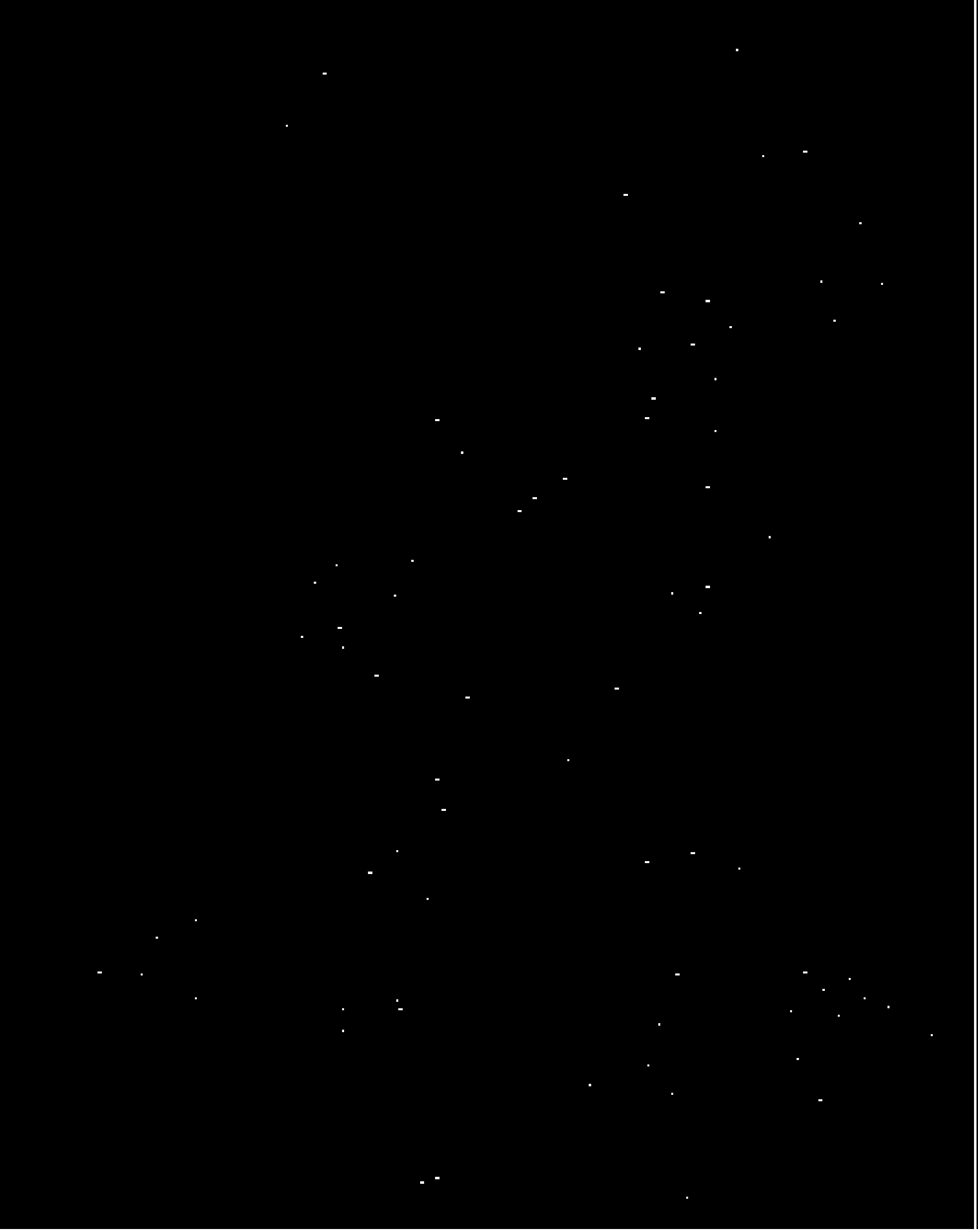
Harris corner detector

1) Compute *M* matrix for image window surrounding each pixel to get its *cornerness* score.
2) Find points with large corner response (*f* > threshold)
3) Take the points of local maxima, i.e., perform non-maximum suppression

U. PORTO FC



U.PORTO




U.PORTO FC

Properties of the Harris corner detector

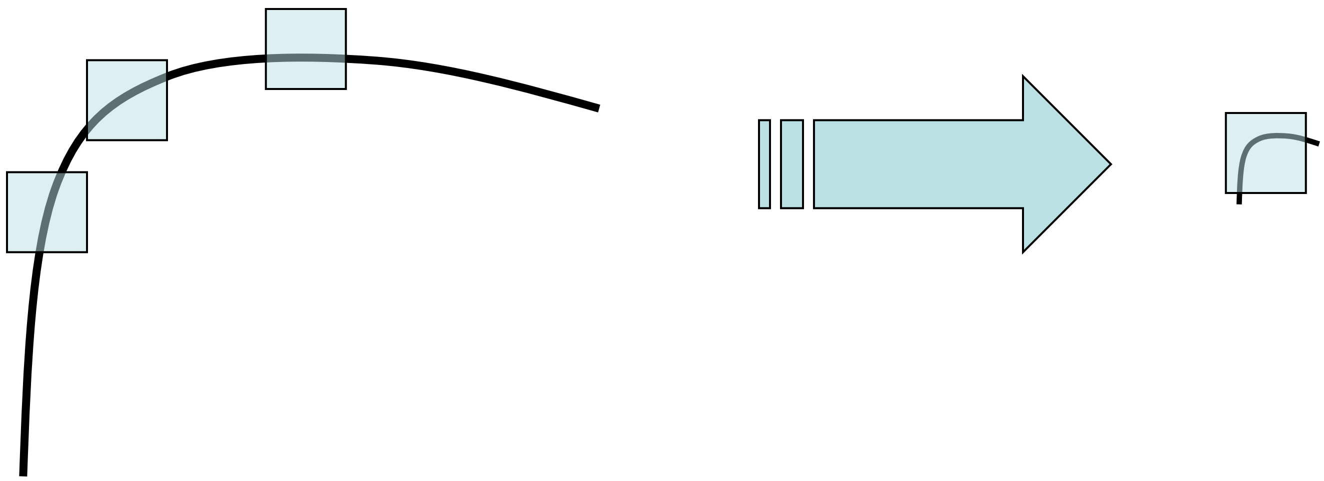
- Rotation invariant? Yes

$$M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$$

- Scale invariant?

Properties of the Harris corner detector

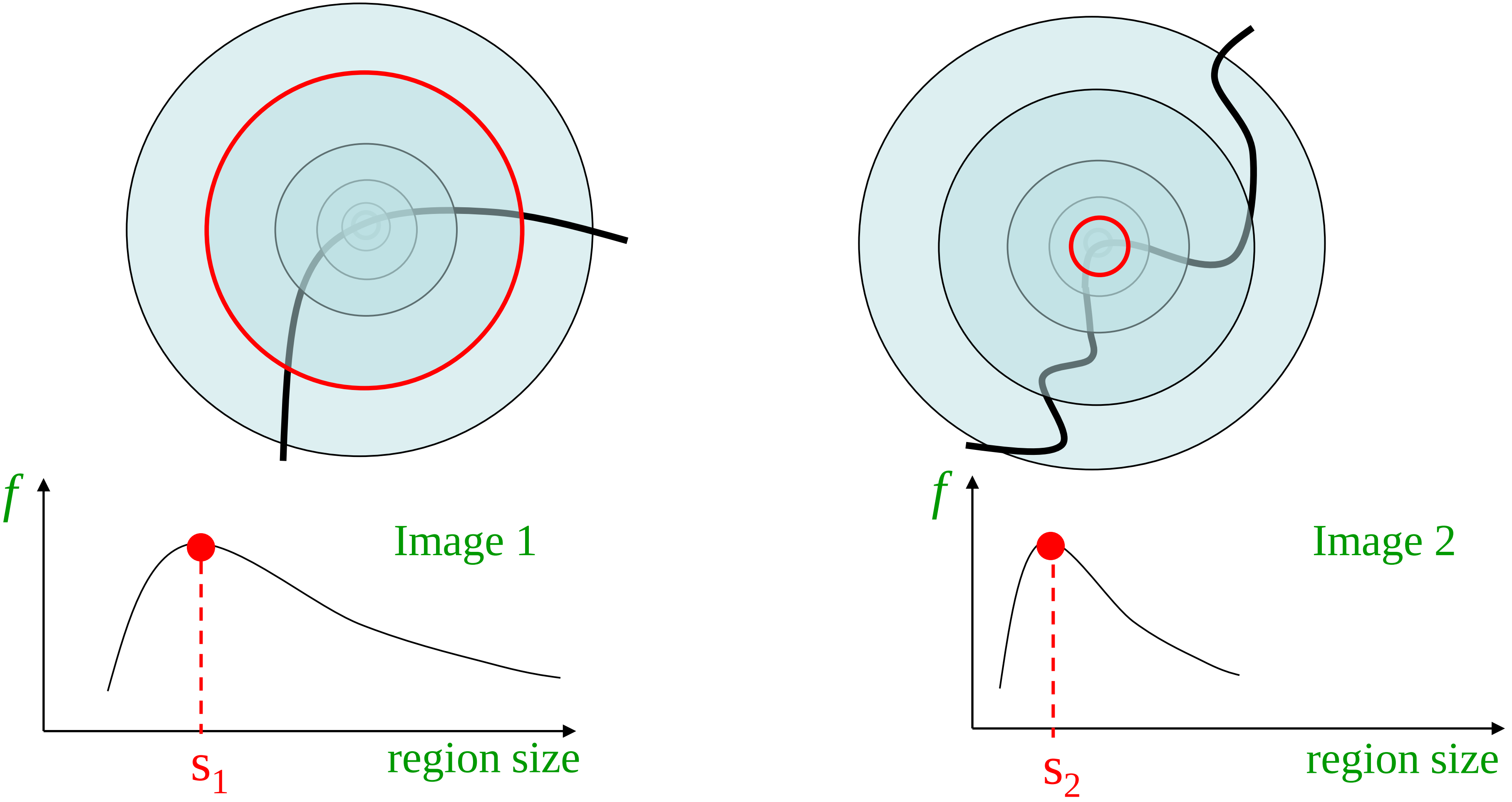
- Rotation invariant? Yes
- Scale invariant? No



All points will be classified as edges

Corner !

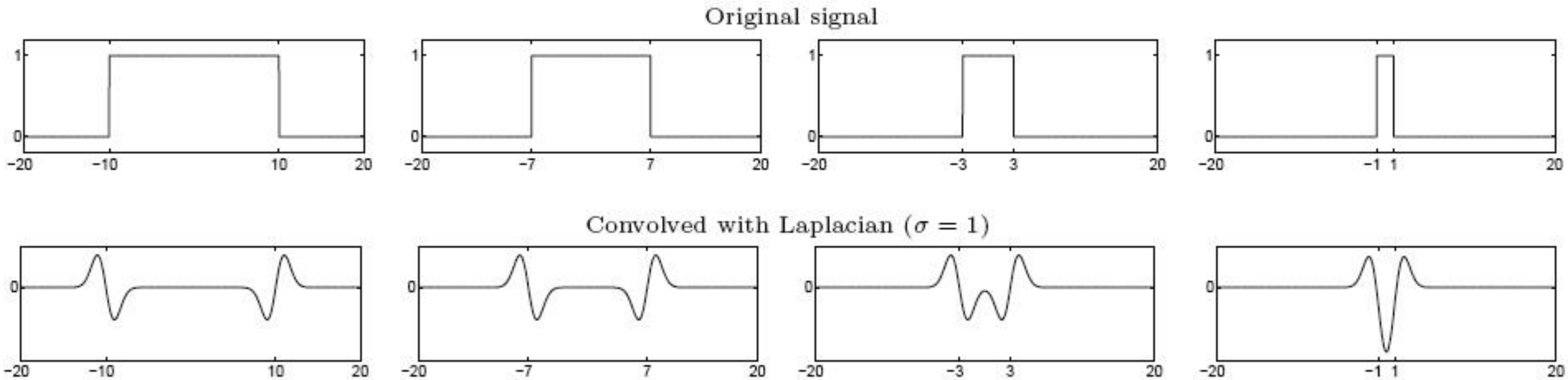
Automatic scale selection



U.PORTO FC

From edges to blobs

- Edge = ripple
- Blob = superposition of two ripples

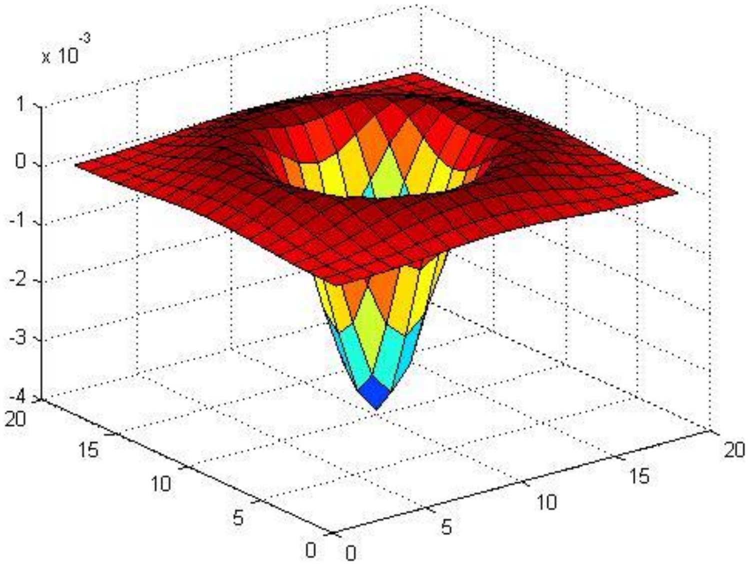


maximum

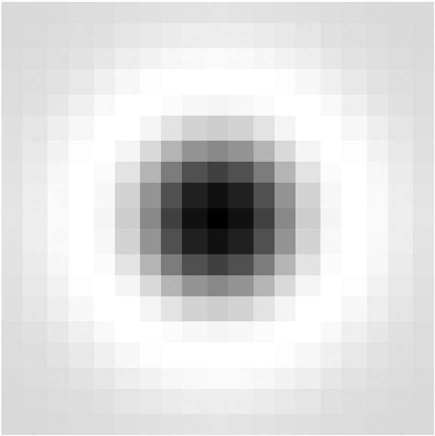
- Spatial selection: the magnitude of the Laplacian response will achieve a maximum at the center of the blob, provided the scale of the Laplacian is "matched" to the scale of the blob

U.PORTO FC

Blob detection in 2D



- Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D

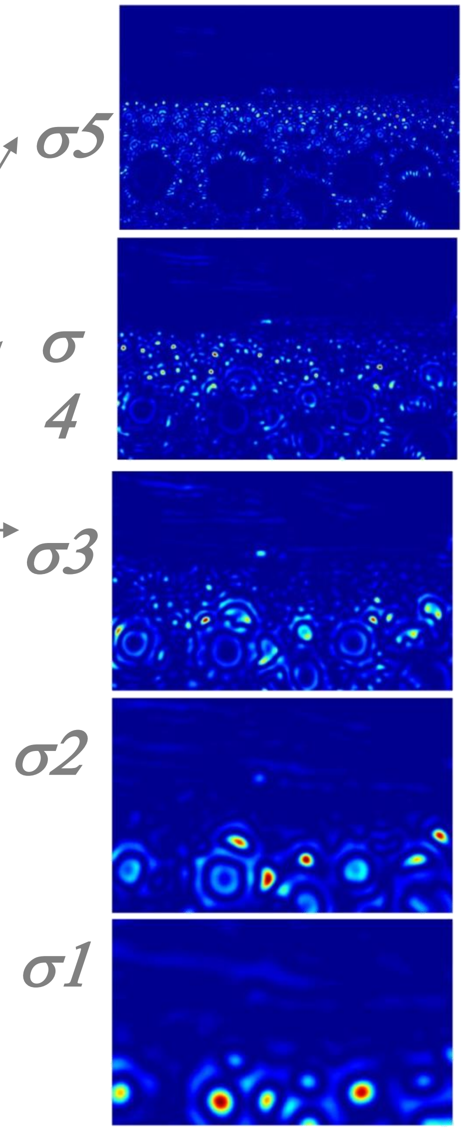


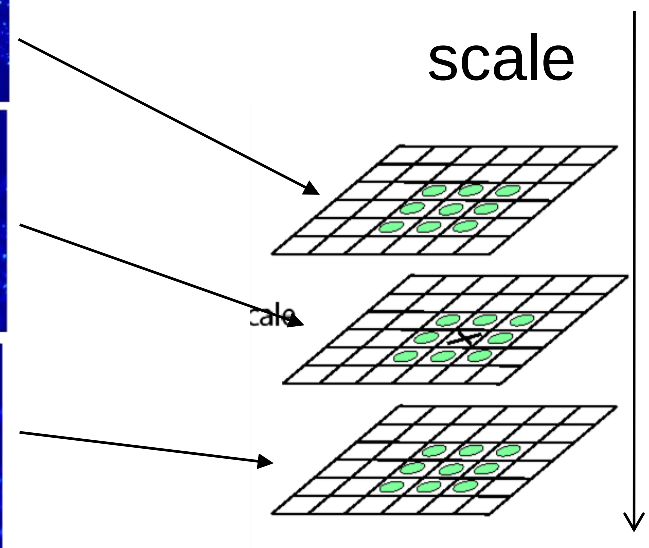
$$\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$$

Scale invariant interest points

Interest points are local maxima in both position and scale

$$L_{xx}(\sigma) + L_{yy}(\sigma)$$



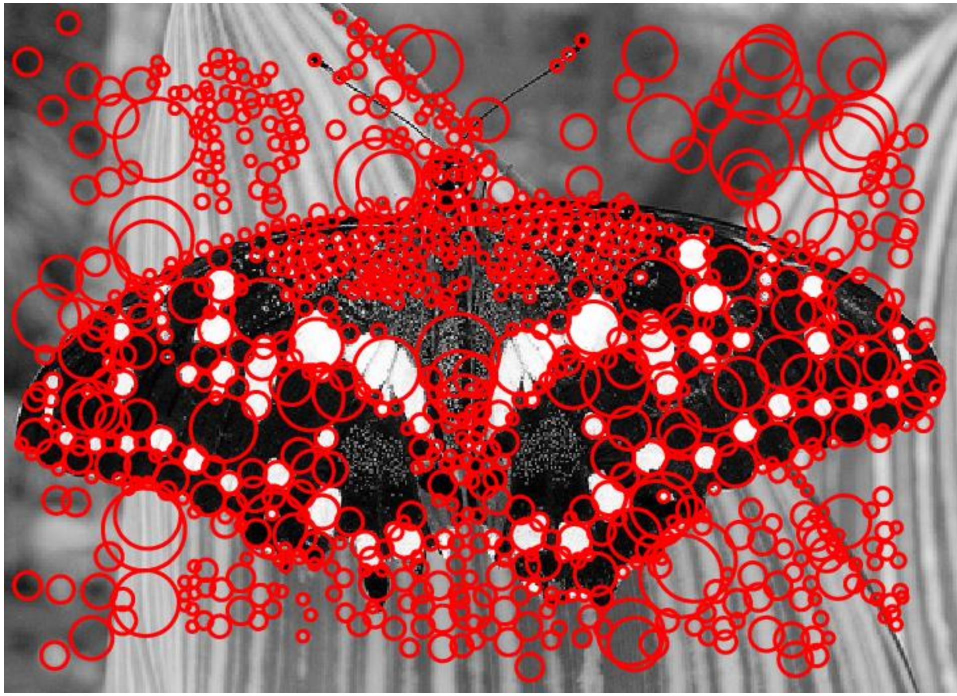


Squared filter response maps

⇒ **List of *(x, y,* σ*)***

U.PORTO FC

Example



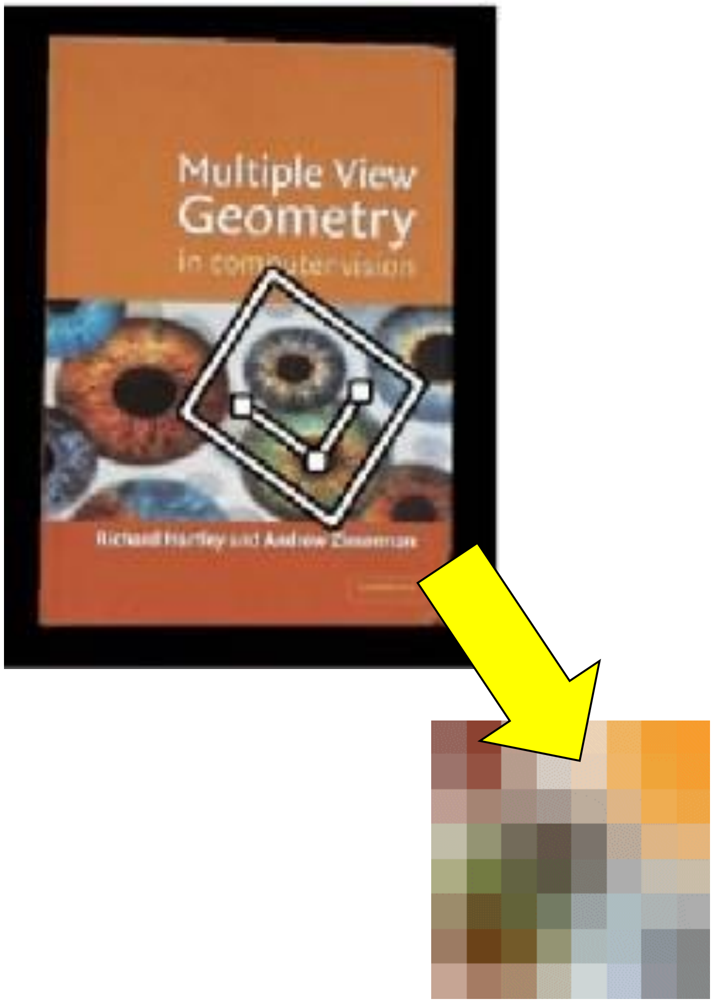
U. PORTO FC

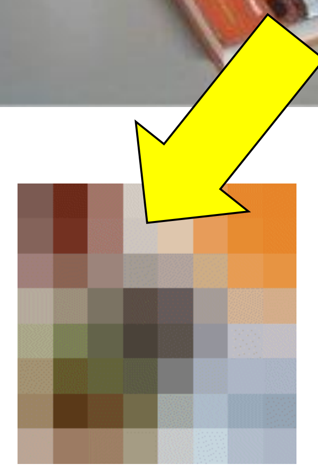
Topic: Local invariant descriptors

- Detection of interest points
- Local invariant descriptors
- Classification using visual words

U. PORTO FC

Geometric transformations

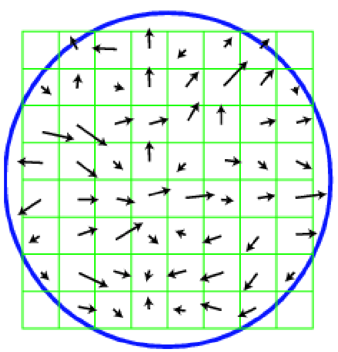


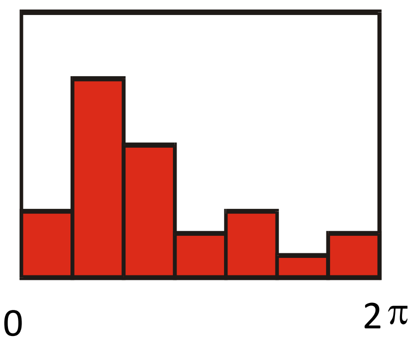


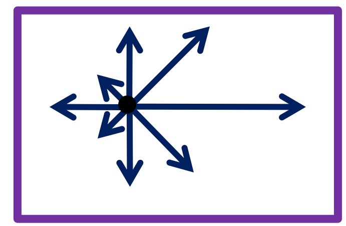
e.g. scale,
translation,
rotation

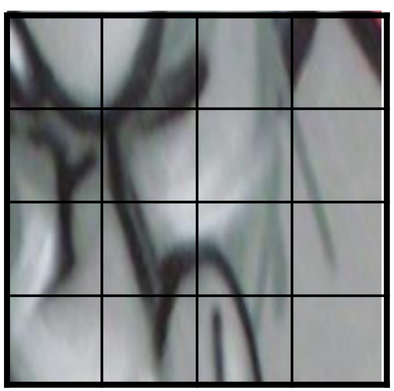
SIFT descriptor [Lowe 2004]

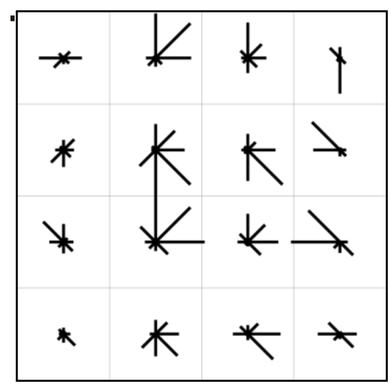
- Use histograms to bin pixels within sub-patches according to their orientation







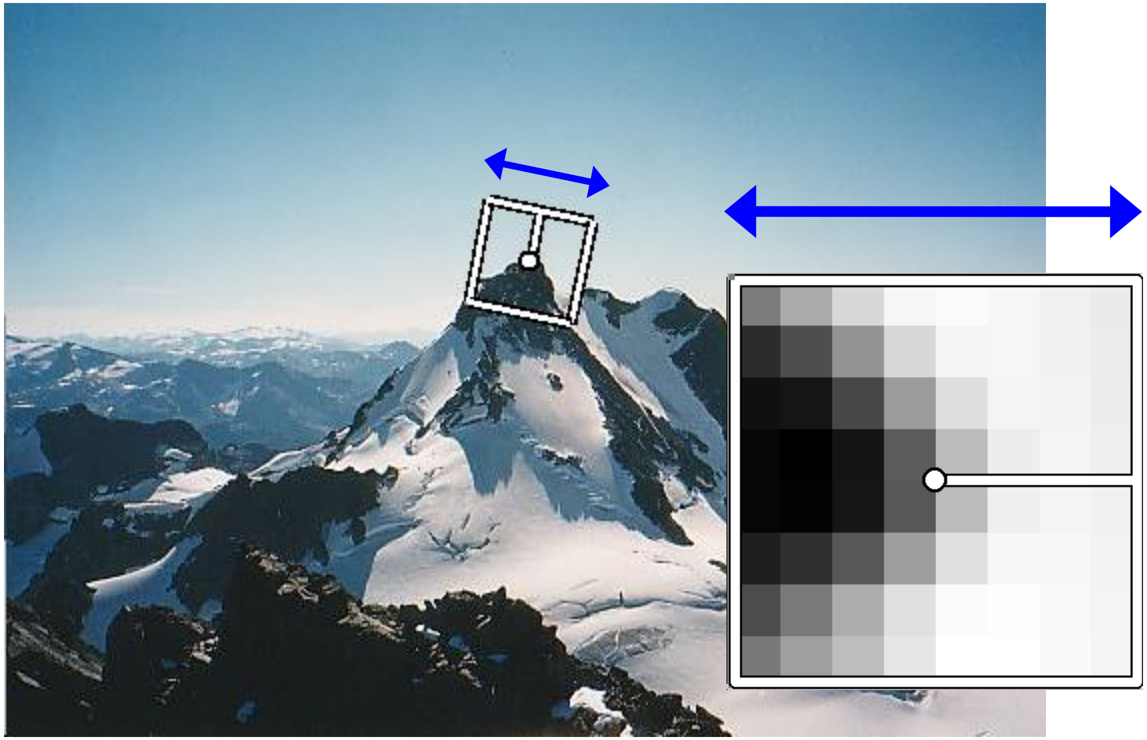




Why subpatches?

Why does SIFT have some illumination invariance?

Making descriptor rotation invariant

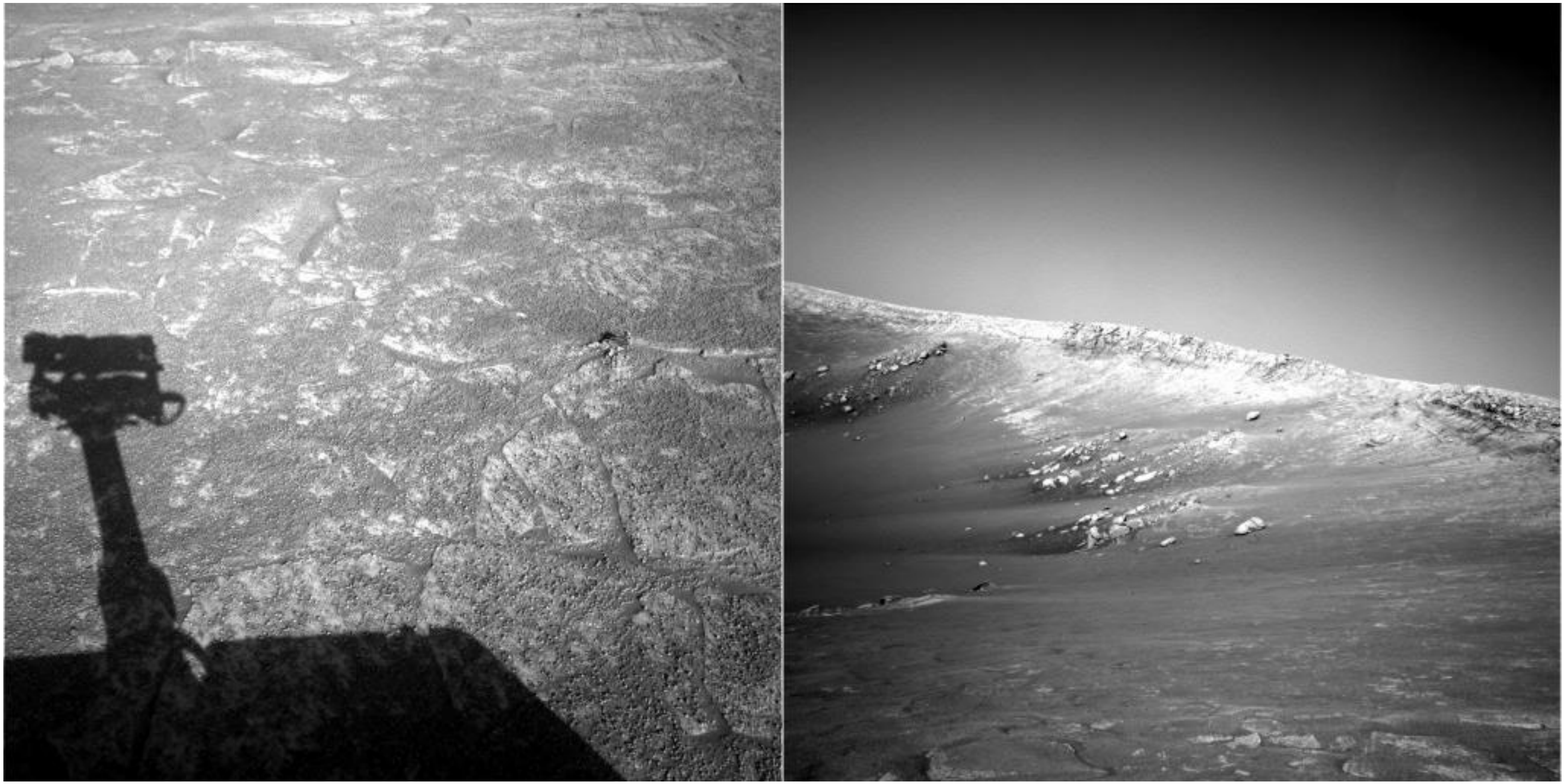


- Rotate patch according to its dominant gradient orientation
- This puts the patches into a canonical orientation

SIFT descriptor [Lowe 2004]

- Extraordinarily robust matching technique
  - Can handle changes in viewpoint
  - Can handle significant changes in illumination
  - Fast and efficient—can run in real time
  - Lots of code available

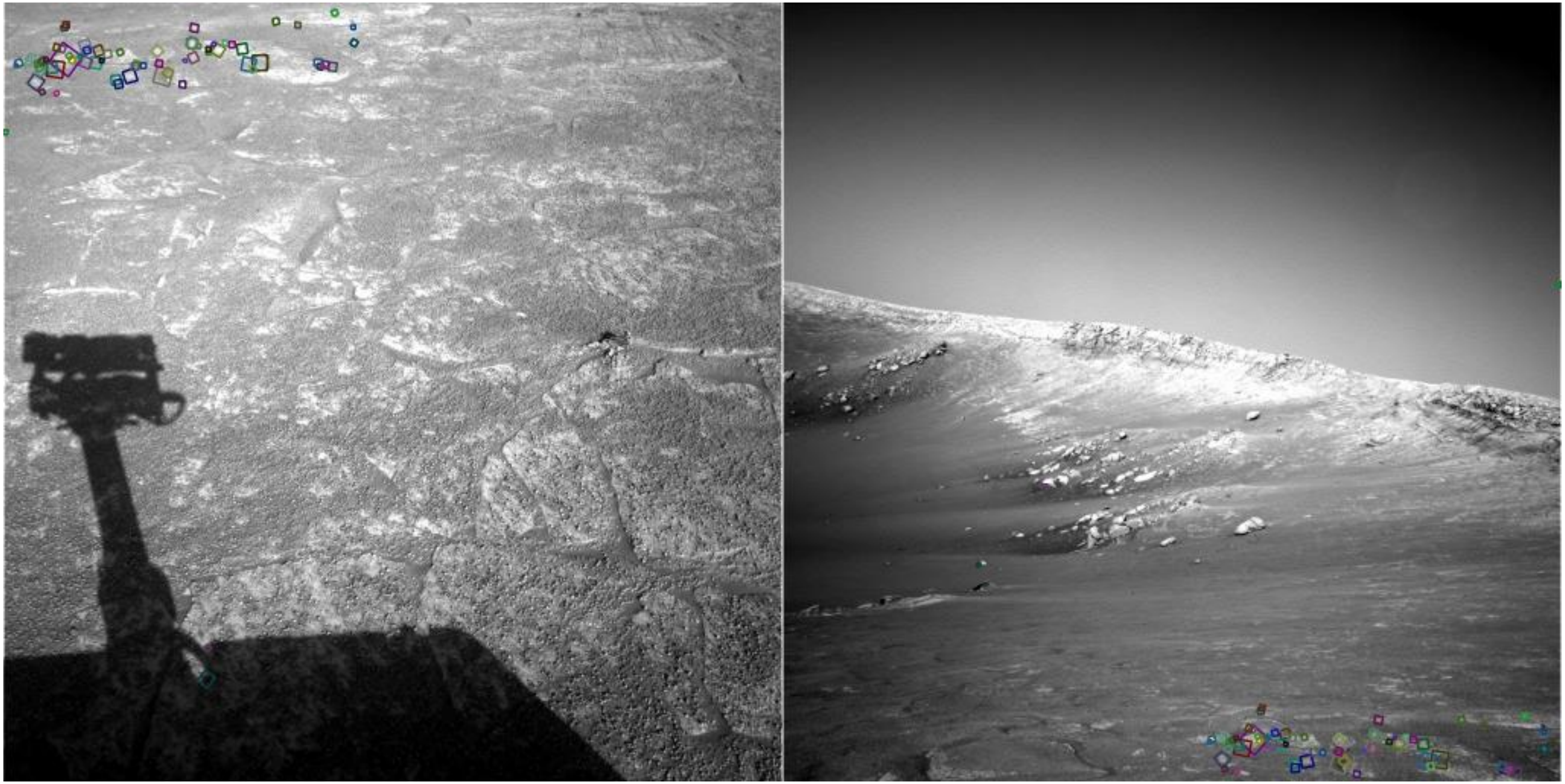
Example



NASA Mars Rover images

U. PORTO FC

Example



NASA Mars Rover images

U.PORTO FC

SIFT properties

- Invariant to
  - Scale
  - Rotation
- Partially invariant to
  - Illumination changes
  - Camera viewpoint
  - Occlusion, clutter

Summary

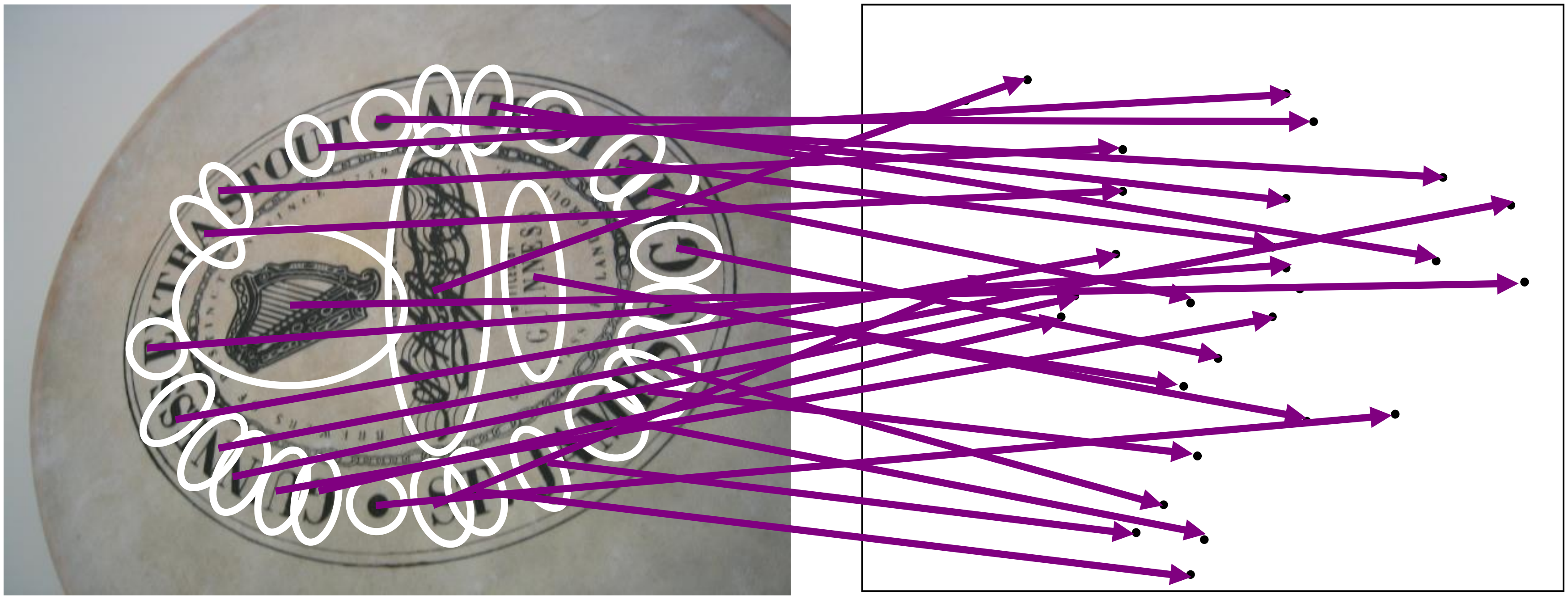
- Interest point detection
  - Harris corner detector
  - Laplacian of Gaussian, automatic scale selection
- Invariant descriptors
  - Rotation according to dominant gradient direction
  - Histograms for robustness to small shifts and translations (SIFT descriptor)

Topic: Classification using visual words

- Detection of interest points
- Local invariant descriptors
- Classification using visual words

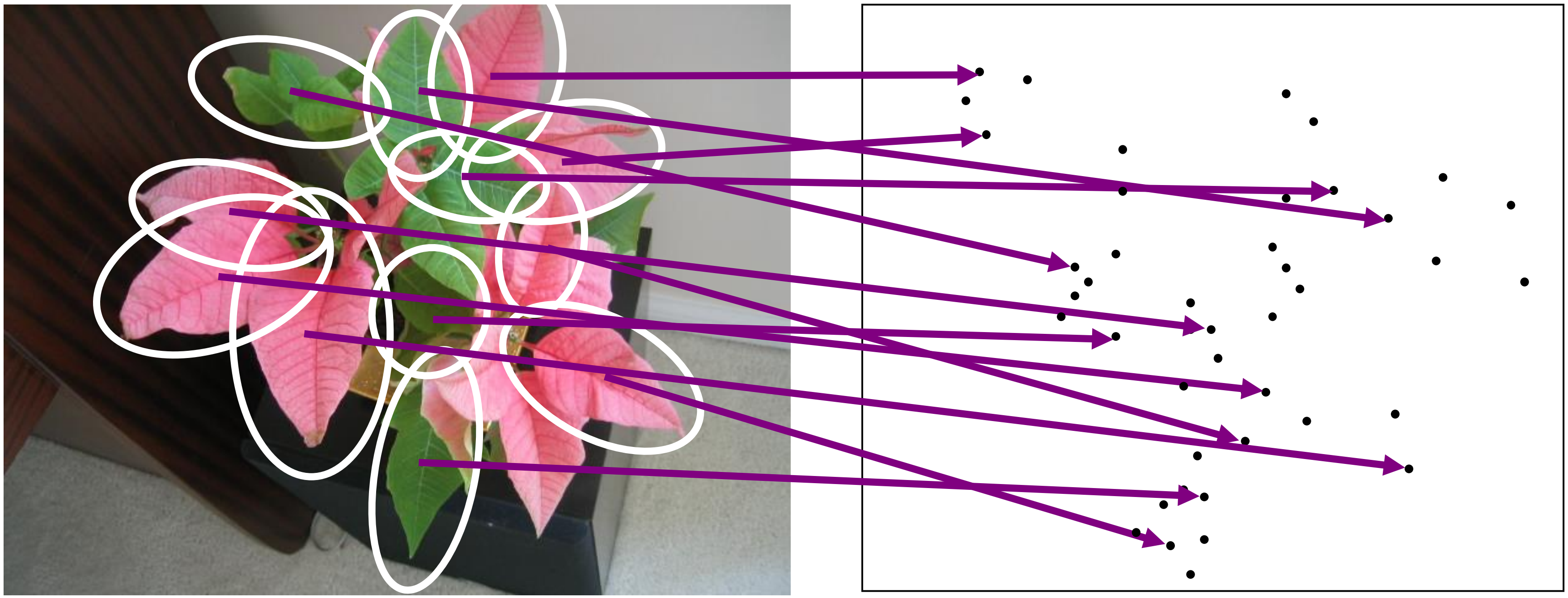
U. PORTO FC

Visual words: main idea

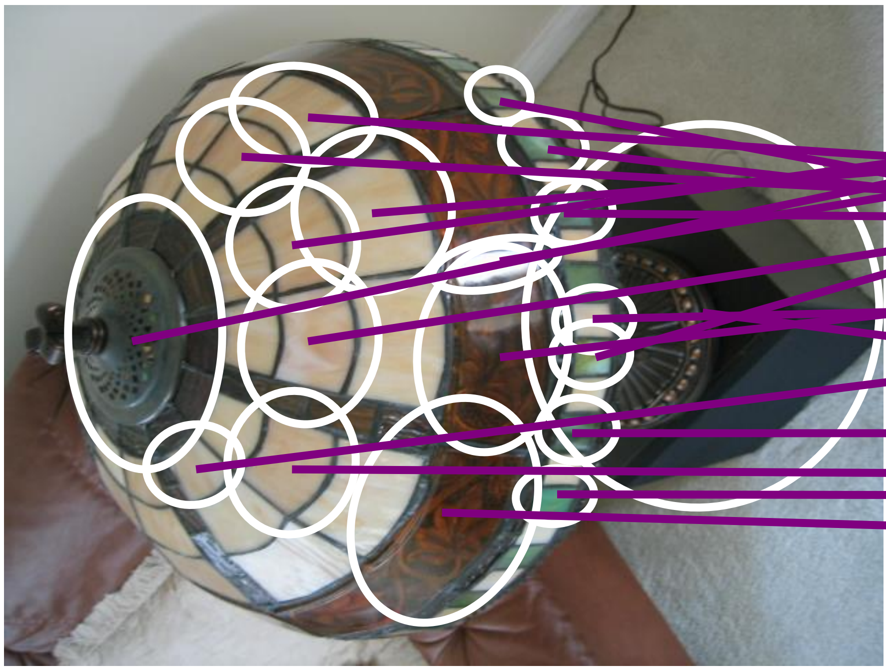


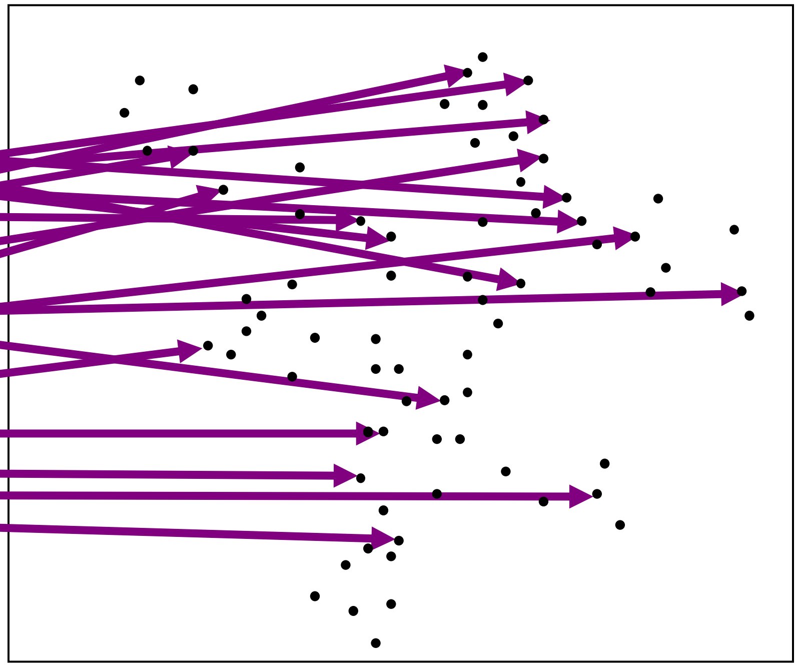
- Extract some local features from a number of images …

Visual words: main idea

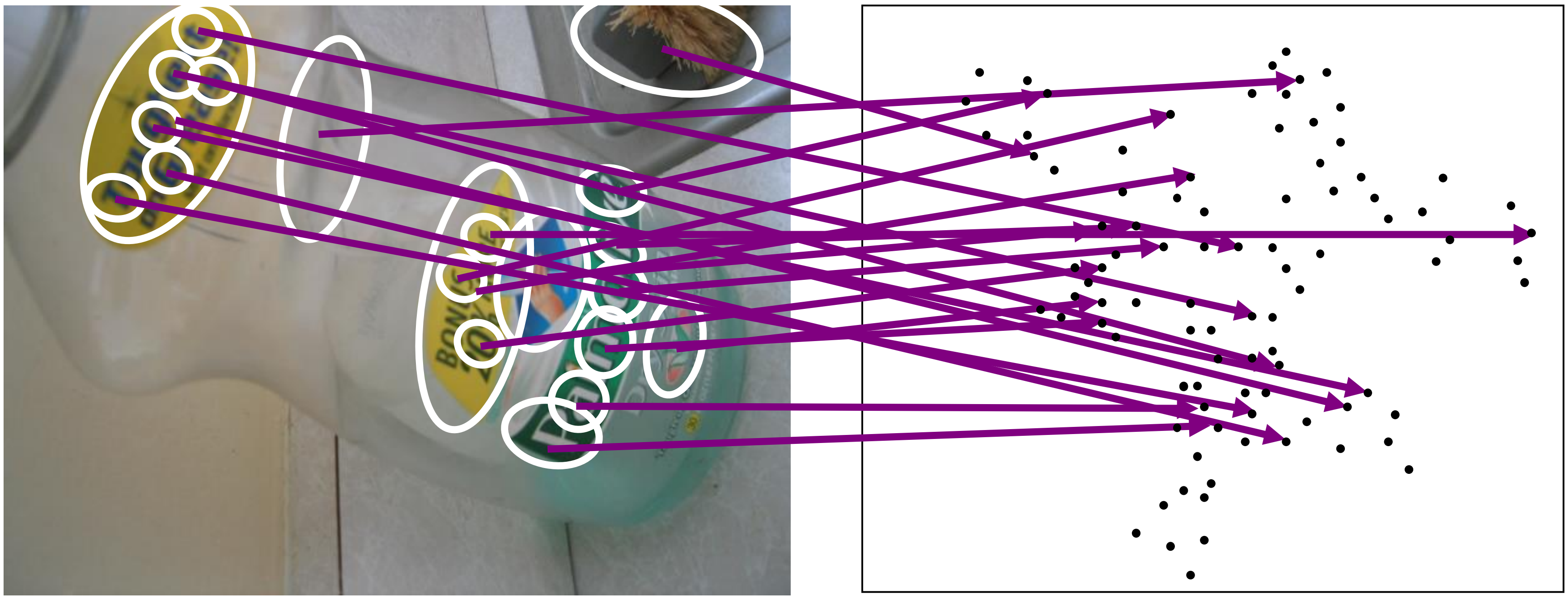


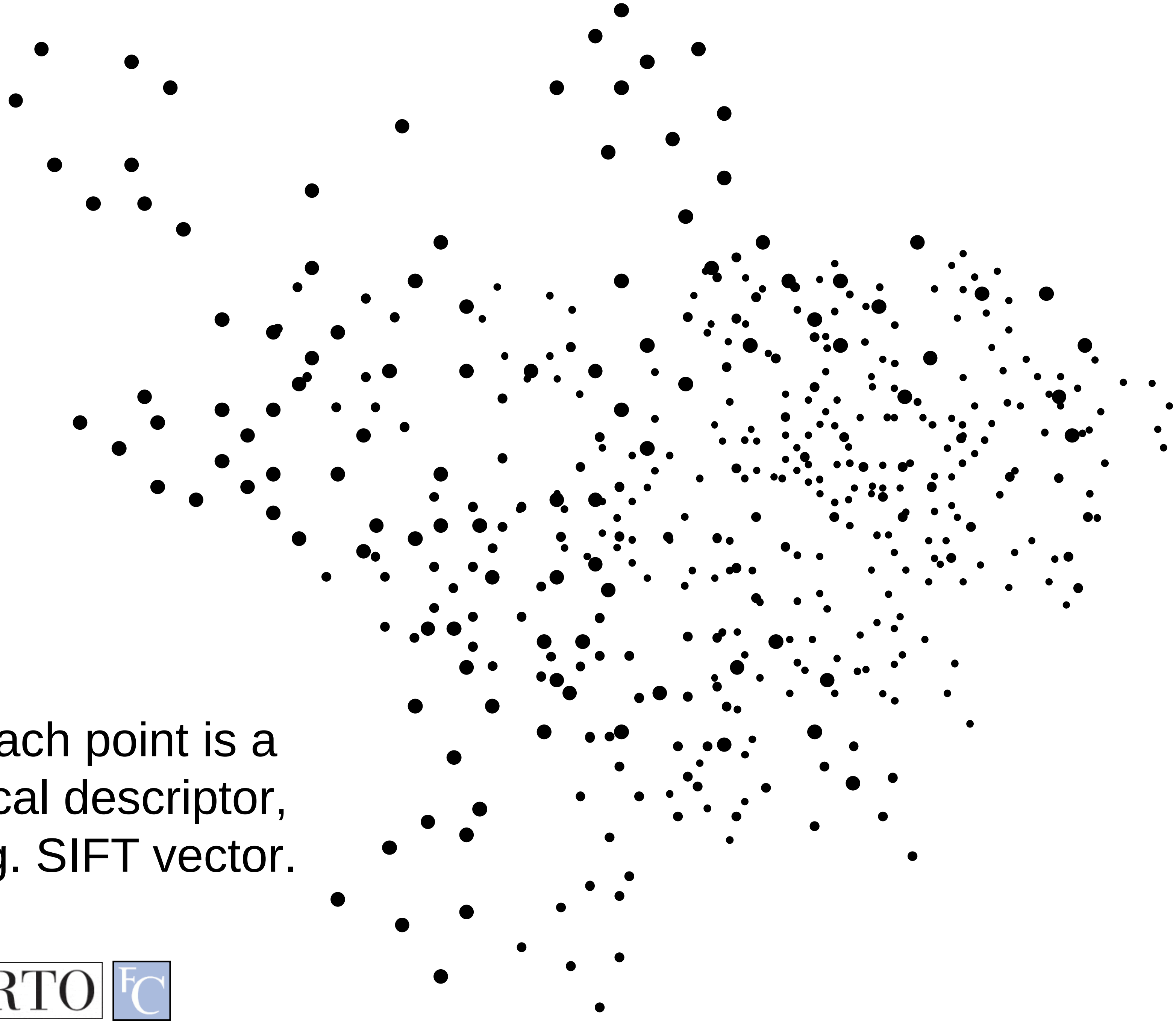
Visual words: main idea



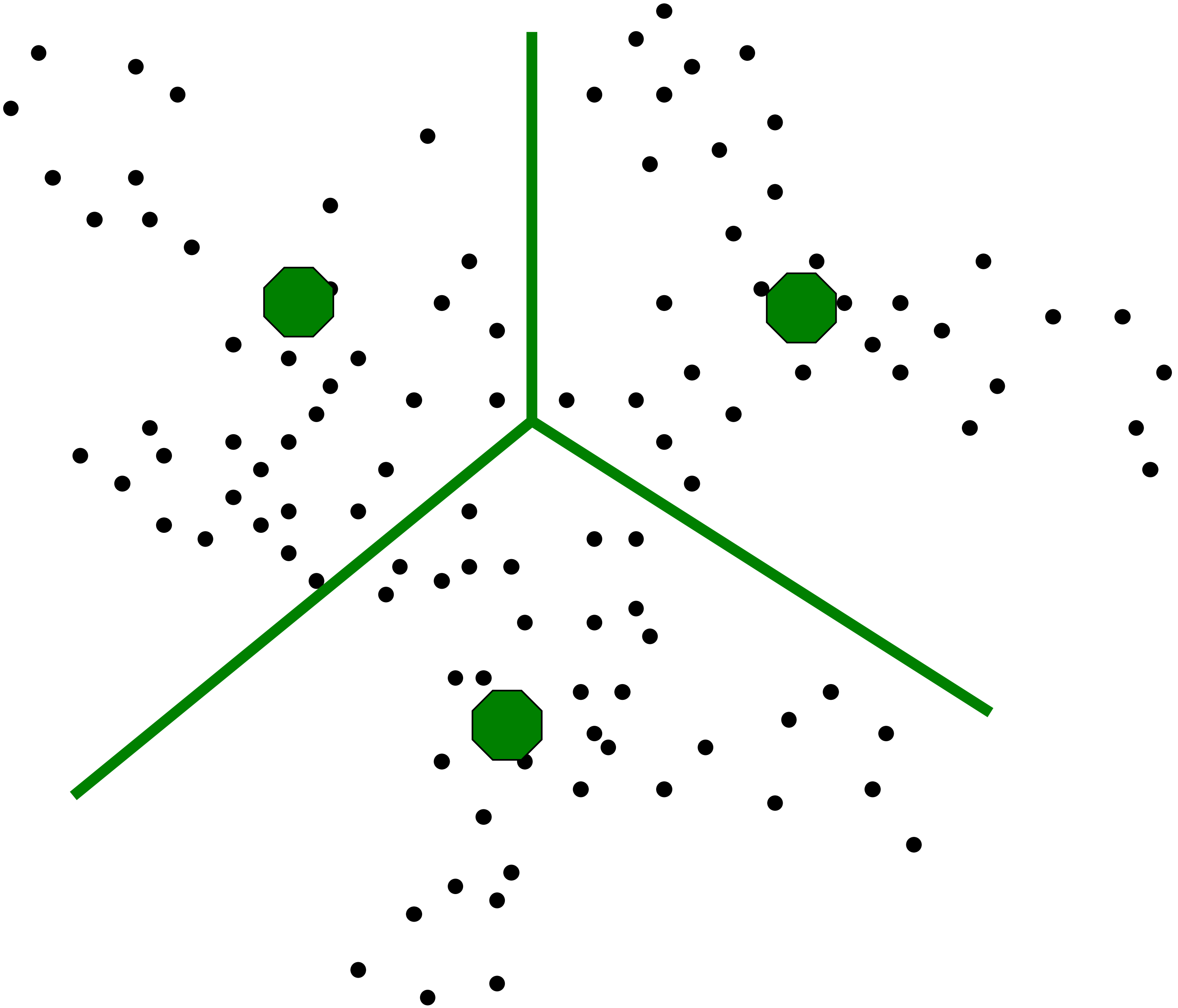


Visual words: main idea



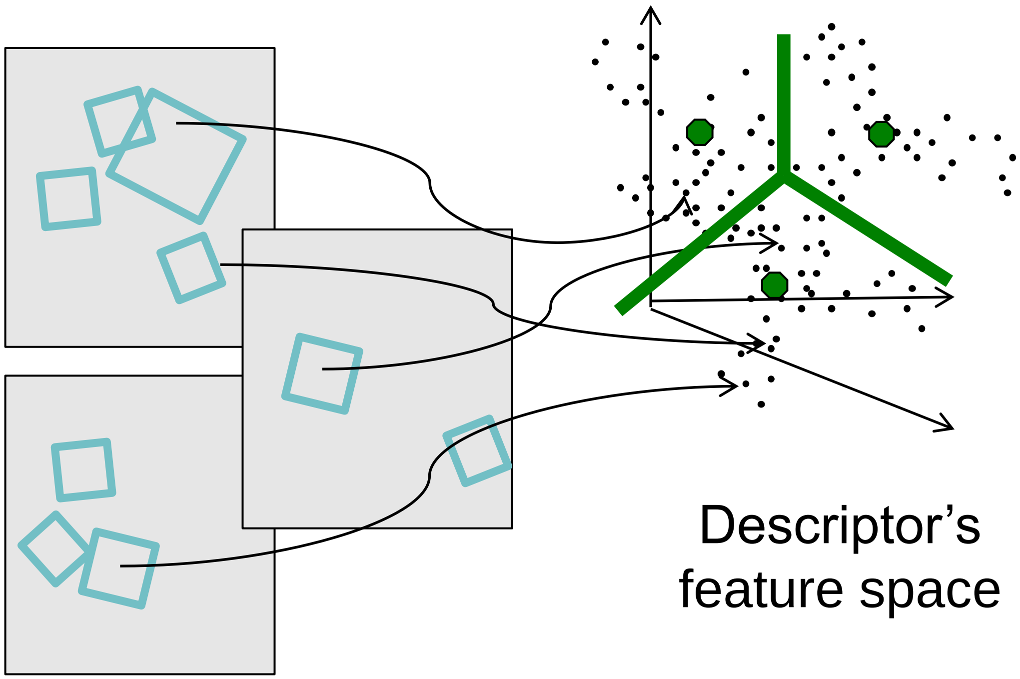


Each point is a local descriptor, e.g. SIFT vector.



Visual words

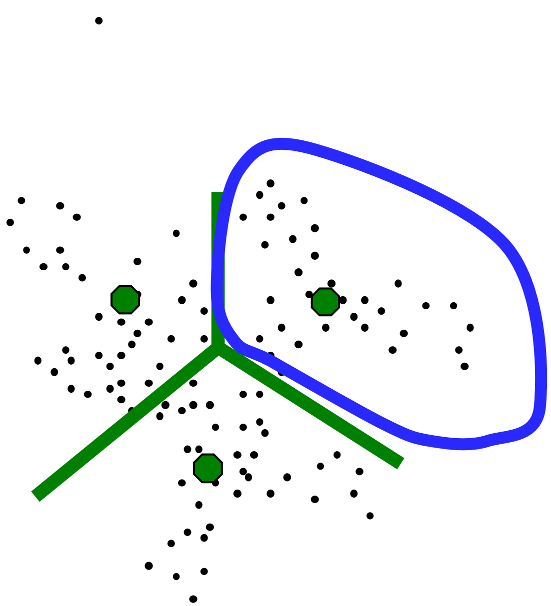
- Map high-dimensional descriptors to tokens by quantizing the feature space

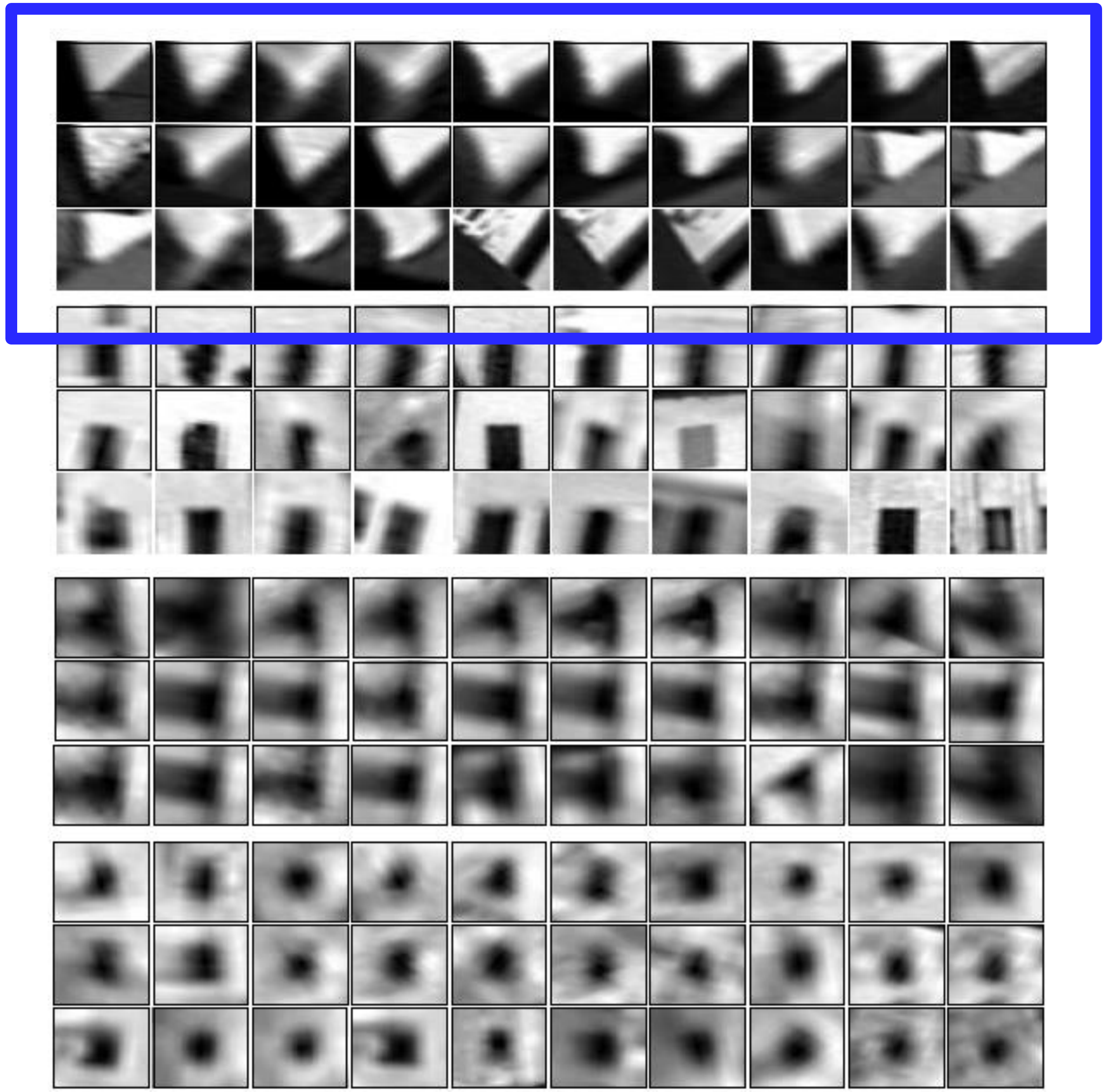


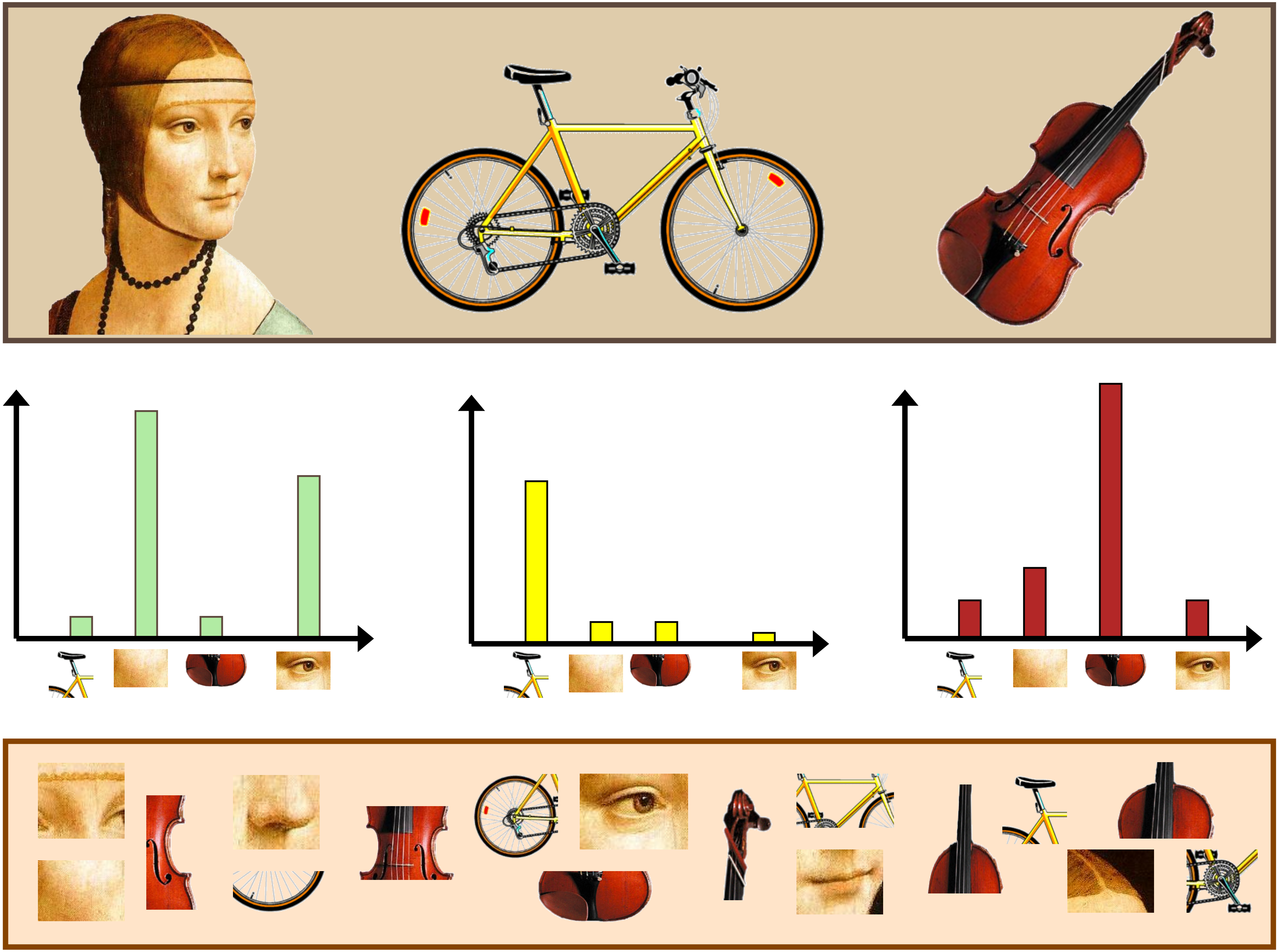
- Quantize via clustering, let cluster centers be the prototype "words"
- Determine which word to assign to each new image region by finding the closest cluster center

Visual words

Example: each group of patches belongs to the same visual word



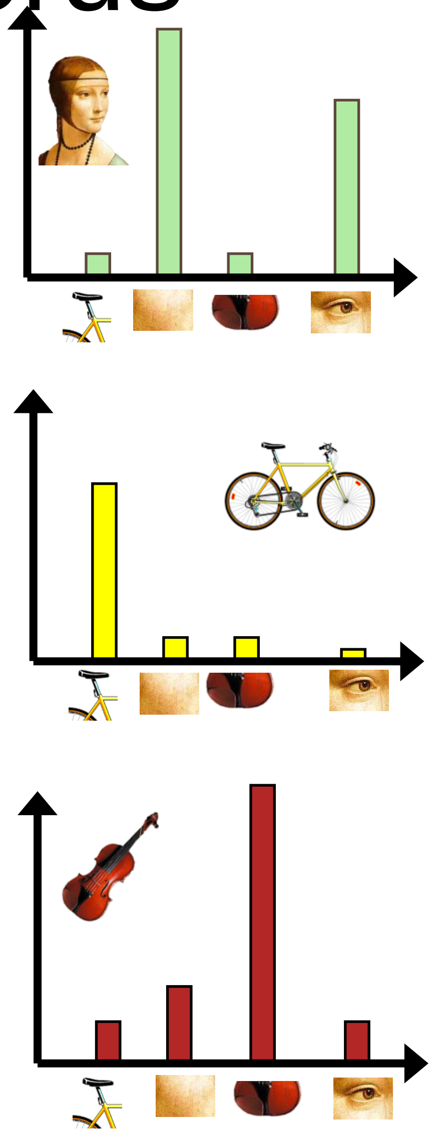




Bags of visual words

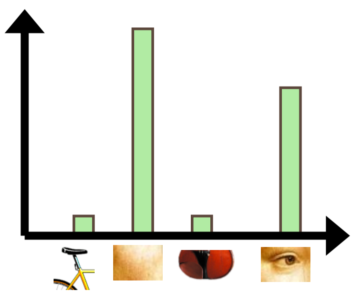
- Summarize entire image based on its distribution (histogram) of word occurrences

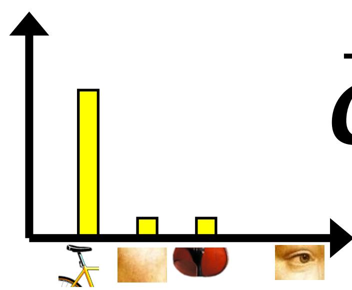
- Analogous to bag of words representation commonly used for documents



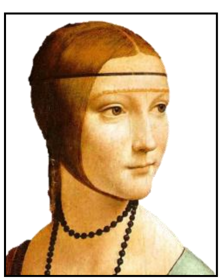
Comparing bags of words

- Rank frames by normalized scalar product between their (possibly weighted) occurrence counts---*nearest neighbor* search for similar images

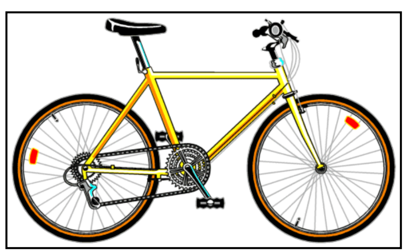




$\vec{q}$



$\vec{d}_j$



$$sim(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \, \|q\|}$$

$$= \frac{\sum_{i=1}^{V} d_j(i) * q(i)}{\sqrt{\sum_{i=1}^{V} d_j(i)^2} * \sqrt{\sum_{i=1}^{V} q(i)^2}}$$

for vocabulary of *V* words

U.PORTO FC

Bags of words: pros and cons

+ flexible to geometry / deformations / viewpoint
+ compact summary of image content
+ provides vector representation for sets
+ very good results in practice

- basic model ignores geometry – must verify afterwards, or encode via features
- background and foreground mixed when bag covers whole image
- optimal vocabulary formation remains unclear

Resources

- Szeliski, “Computer Vision: Algorithms and Applications”, Springer, 2011
  – Chapter 4 – “Feature Detection and Matching”