ITMO UNIVERSITY

Image descriptors and machine learning

Natalia Khanzhina

Senior Researcher

Machine Learning Lab

nehanzhina@corp.ifmo.ru

Descriptors

- A representation is constructed by **detection** (deciding where to sample) followed by **description** (deciding how to sample)
- The resulting **descriptor** is a vector **v** that can be compared to memory
- Desirable properties of a **descriptor** vector:

1. invariance to nuisance parameters such as illumination, small shifts in position and scale of the region

2. **discriminative power** such that different objects can be told apart

Descriptors

Nomenclature for **descriptor** properties:

1. **Texture**
   - Fine details, e.g. wrinkles

2. **Colour**
   - Surface reflectance properties.

3. **Shape**
   - Coarse details, e.g. contours and depth boundaries

Intensity normalisation

- A very simple descriptor is the intensity normalized patch

$$v = \frac{\tilde{v} - \mu(\widetilde{v})}{\sigma(\widetilde{v})}$$

Where v =[f(x_1) … f(x_n)], $x_n \epsilon$ patch

HOG and SIFT - Crucial points

- Orientation based descriptors are very powerful
  - because robust to changes in brightness
- HOG feature
  - known window, make histogram of orientations
- SIFT feature
  - find domain
    - patch center and radius
  - compute descriptor
    - histogram of orientations
- Numerous powerful variants

Lowe’s SIFT features

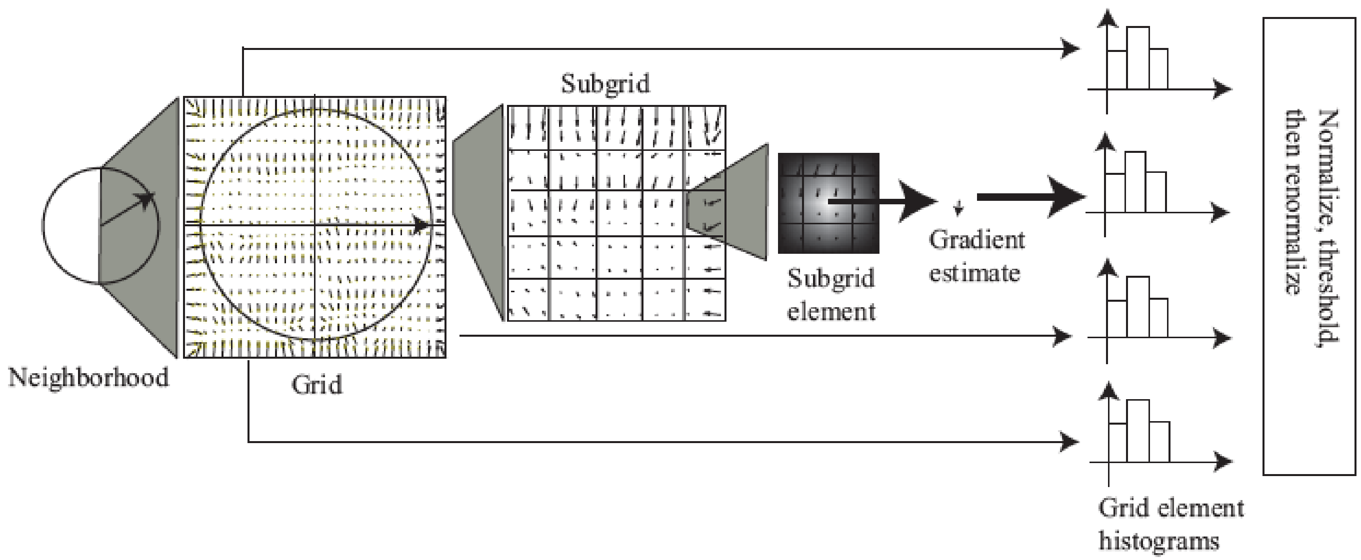
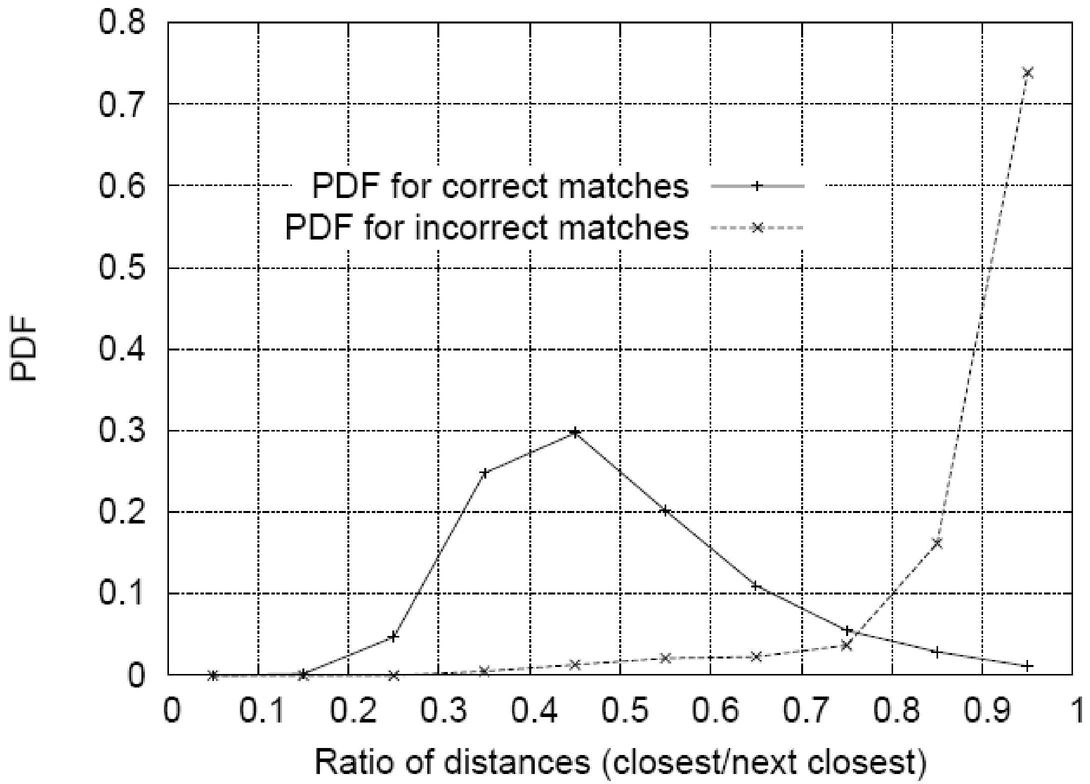


FIGURE 5.14: To construct a SIFT descriptor for a neighborhood, we place a grid over the rectified neighborhood. Each grid is divided into a subgrid, and a gradient estimate is computed at the center of each subgrid element. This gradient estimate is a weighted average of nearby gradients, with weights chosen so that gradients outside the subgrid cell contribute. The gradient estimates in each subgrid element are accumulated into an orientation histogram. Each gradient votes for its orientation, with a vote weighted by its magnitude and by its distance to the center of the neighborhood. The resulting orientation histograms are stacked to give a single feature vector. This is normalized to have unit norm; then terms in the normalized feature vector are thresholded, and the vector is normalized again.

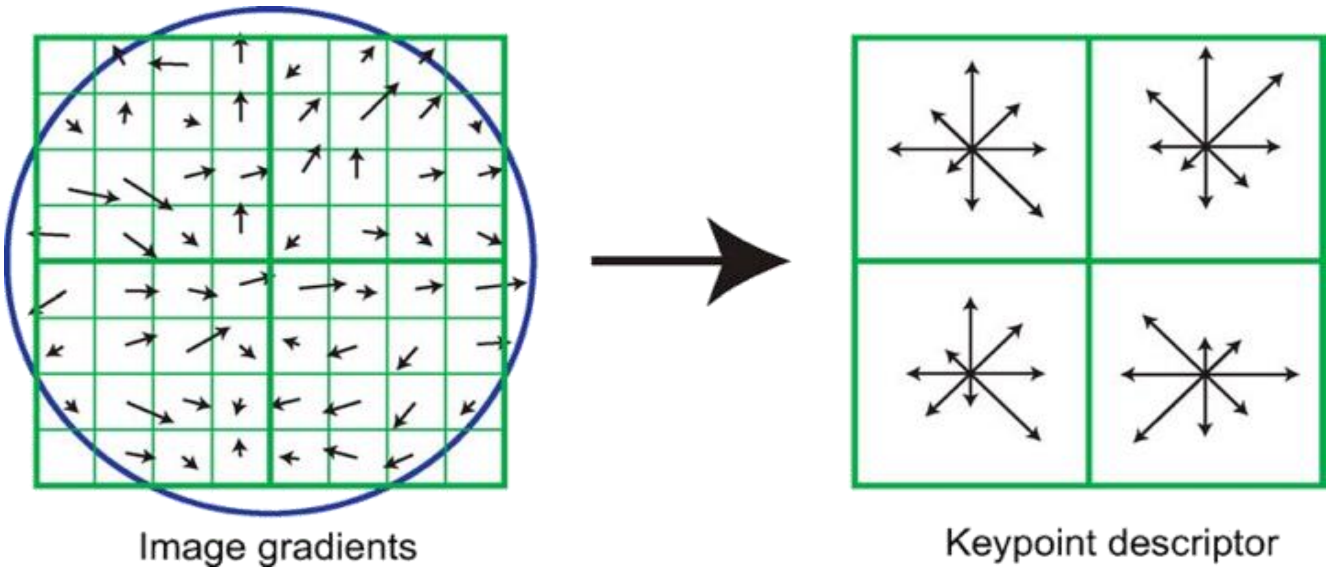
Matching SIFT features

- Can be compared with Euclidean distance
  - test: (dist to closest)/(dist to second closest)



The HOG descriptor

- Nearly identical to the SIFT-descriptor, but adapted to dense grids



The HOG descriptor

- Compute gradient with small filters

$$\nabla f = \left(f * \begin{bmatrix} dx \\ dy \end{bmatrix}\right)(x) \quad \begin{array}{l} dx = [-1\ 0\ 1] \\ dy = [-1\ 0\ 1]^T \end{array}$$

- Perform orientation binning with

$$h_k = \sum_{x \epsilon cell} |\nabla f(x)| B_K(tan^{-1} \nabla f(x))$$

The HOG descriptor

- Each cell now contains K values (K=9)

$$h_l = [h_{l,0} \dots h_{l,9}]^T$$

- These are grouped into 2x2 blocks

$$\tilde{b} = [h_1^T \; h_2^T \; h_3^T \; h_4^T]^T$$

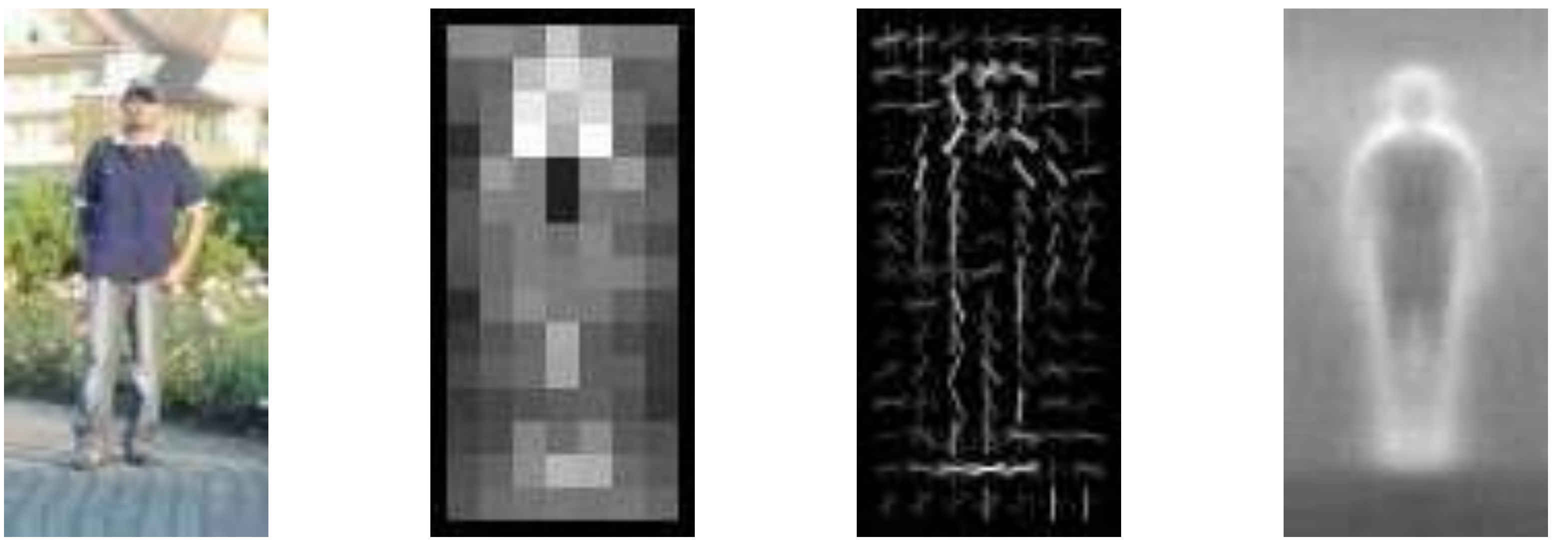
- and finally, the blocks are normalized

$$b = \tilde{b} / \left\| \tilde{b} + \epsilon \right\|$$

Blocks typically overlap, so each cell belongs to several blocks

The HOG descriptor

- The HOG descriptor was introduced in the paper ”Histograms of Oriented Gradients for Human Detection”, Dalal & Triggs, CVPR’05
- Still very common (>9500 citations in Google Scholar)

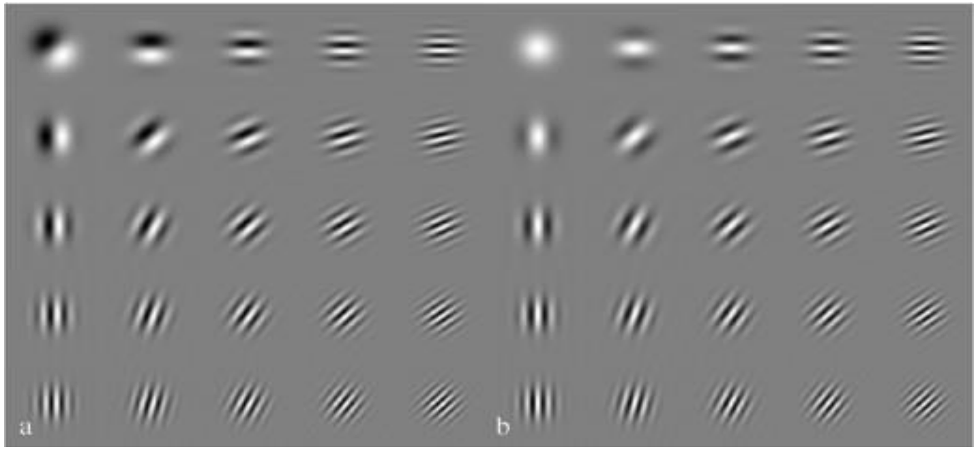


Detector+descriptor pairs

- SIFT, Scale Invariant Feature Transform [D. Lowe ICCV’99, IJCV’04]
- An interest point detector (DoG) + a descriptor
- 4x4 HOG blocks, with a single common normalization
- Other common detector+descriptor features: SURF, BRISK, ORB, SFOP, FREAK (Covered in LE4)

Gabor Jet

- A set of responses from filters that are oriented and localized wavelets



- A filter bank. Other filter banks include e.g. derivative filters in multiple scales, and wavelets.

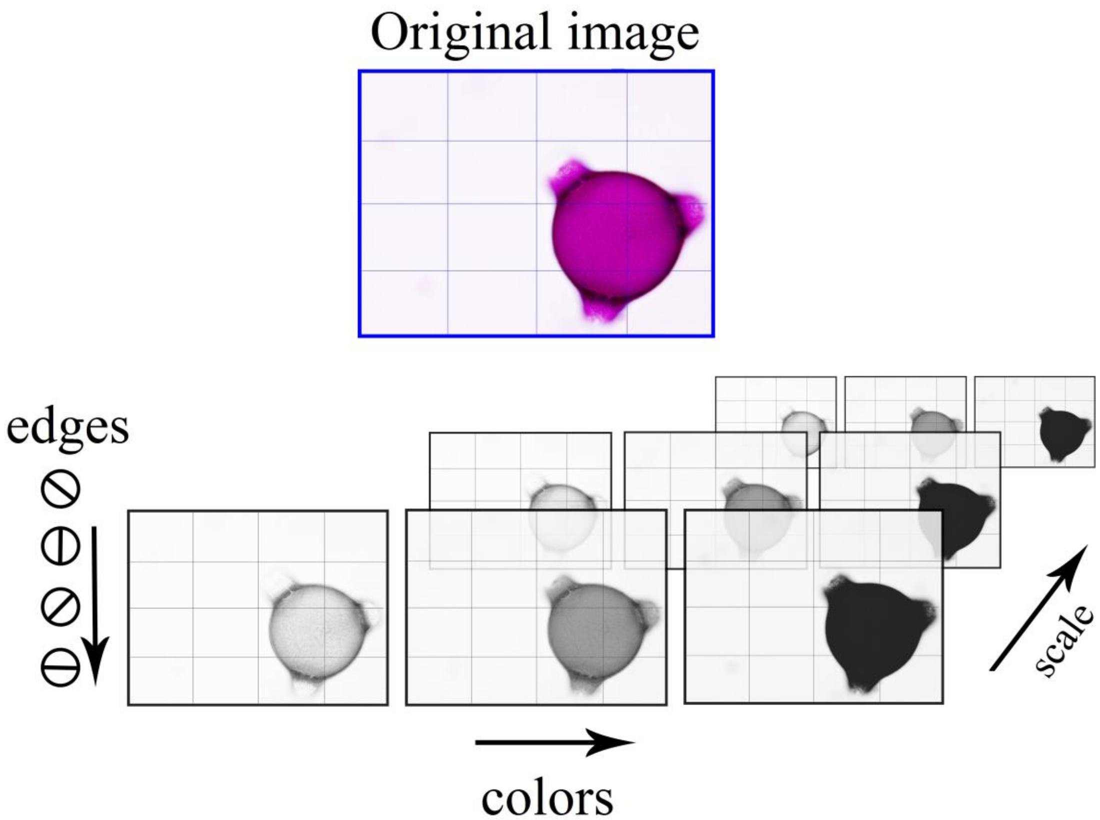
Gabor Jet

- Filter banks are typically used to classify texture, e.g. E. Hayman et al. ”On the Significance of Real-World Conditions for Material Classification”, ECCV’04

GIST

- A. Olivia and A. Torralba, ”Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope”, IJCV’01
- A global feature for images that is useful in scene categorization.
- Motivation: Perceptual studies indicate that scene category is recognized before semantic information such as objects and their relations.

GIST descriptors



-960 descriptors

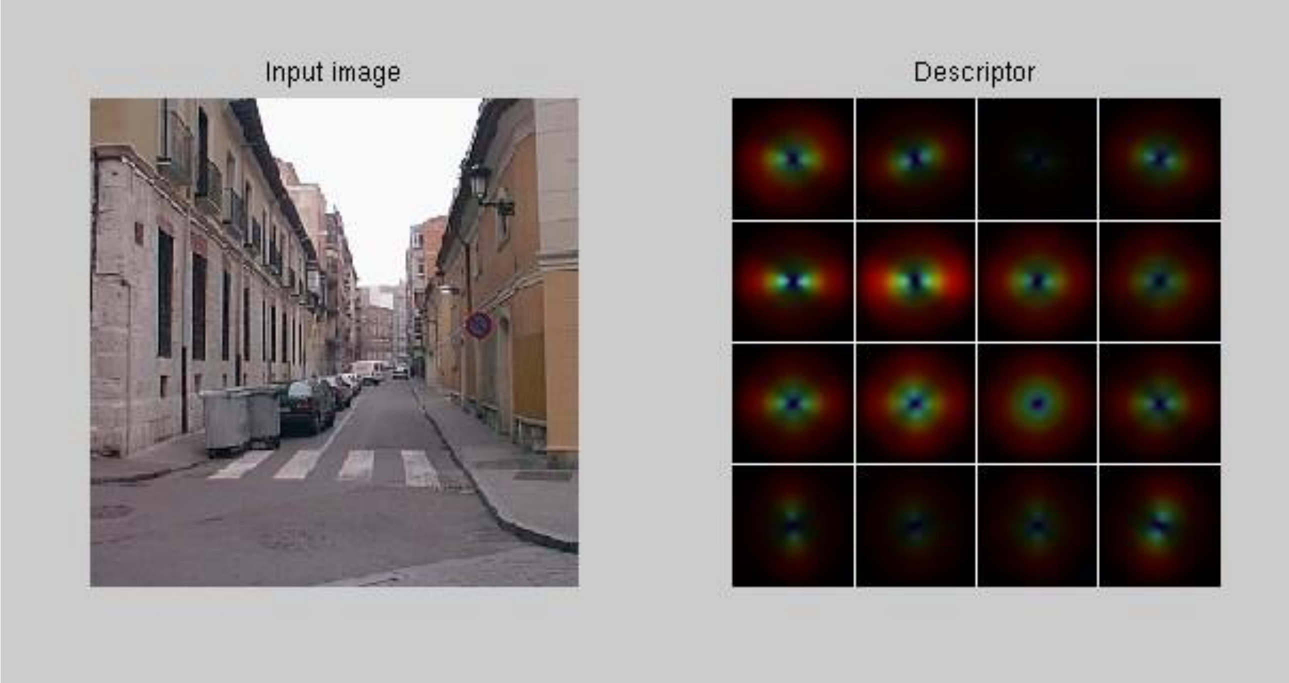
GIST

- Scene categorization dataset

coasts rivers/lakes forests plains mountains

GIST examples

- Gabor jets in 4x4 grid (4 scales, 8 directions) on downsampled images (128x128) 512 element descriptor.



Colour histograms

- Many different variants. E.g. from C. Carson et al. ”Blobworld: A system for region-based image indexing and retrieval”, ICVIS’99
- Transform region of interest into La*b* colour space. Use coarse binning of Lab space, 5x10x10 bins. Select the 218 bins that fall within the RGB gamut.
- Spatial position is discarded. Shift insensitive, scale insensitive.

Colour histograms

- Colour Names, J. van de Weijer et al. ”Learning Color Names for Real-world Applications”, TIP’09
- Label pixels as one of 11 different colours:

- Non-uniform decision regions in Lab space.
- Descriptor by histogramming.

Difficult cases for Descriptors

- Background clutter in 3D scenes Label pixels as one of 11 different colours:
- Patches cut out around features will have varying background.

SURF

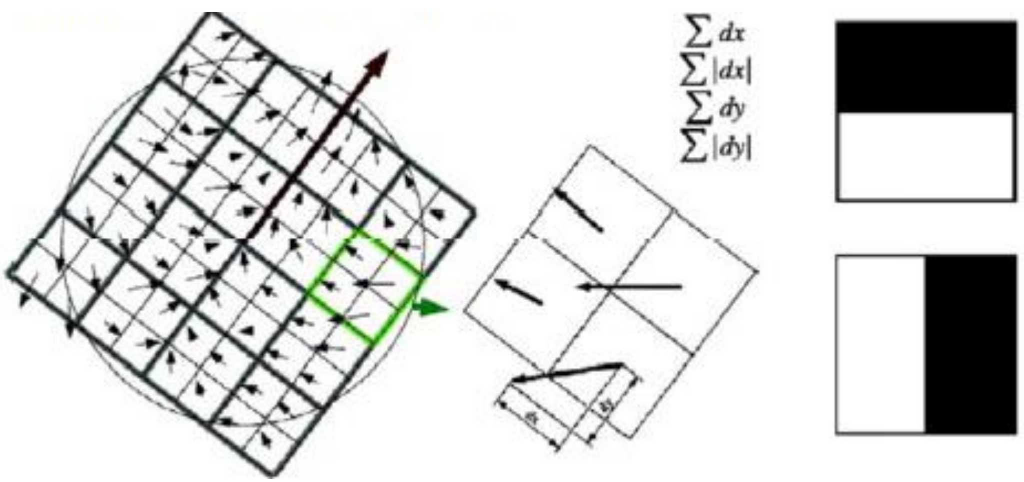
- Speed-Up Robust Features (SURF)
  - Simplified version of SIFT
  - Faster computation but comparable performance
- Characteristics Fast interest point detection
  - Distinctive interest point description
  - Speeded-up descriptor matching
  - Invariant to common image transformations:
    - Image rotation
    - Scale changes
    - Illumination change
    - Small change in viewpoint

Orientation Assignment

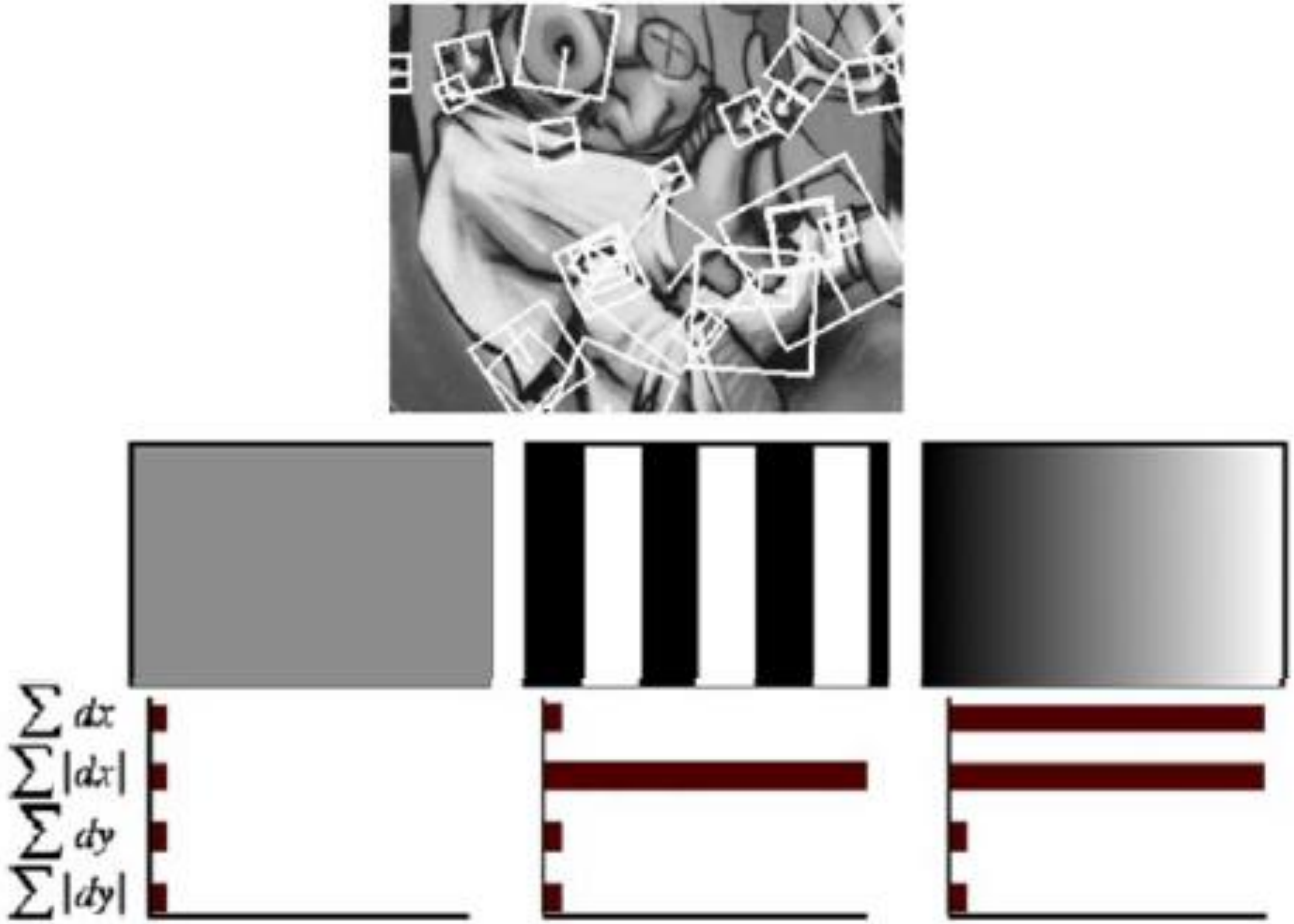
- Methodology
  - The Haar wavelet responses are represented as vectors.
  - Sum all responses within a sliding orientation window covering 60 degree
  - The two summed response yield a new vector
  - The longest vector is the dominant orientation

Building the Descriptor

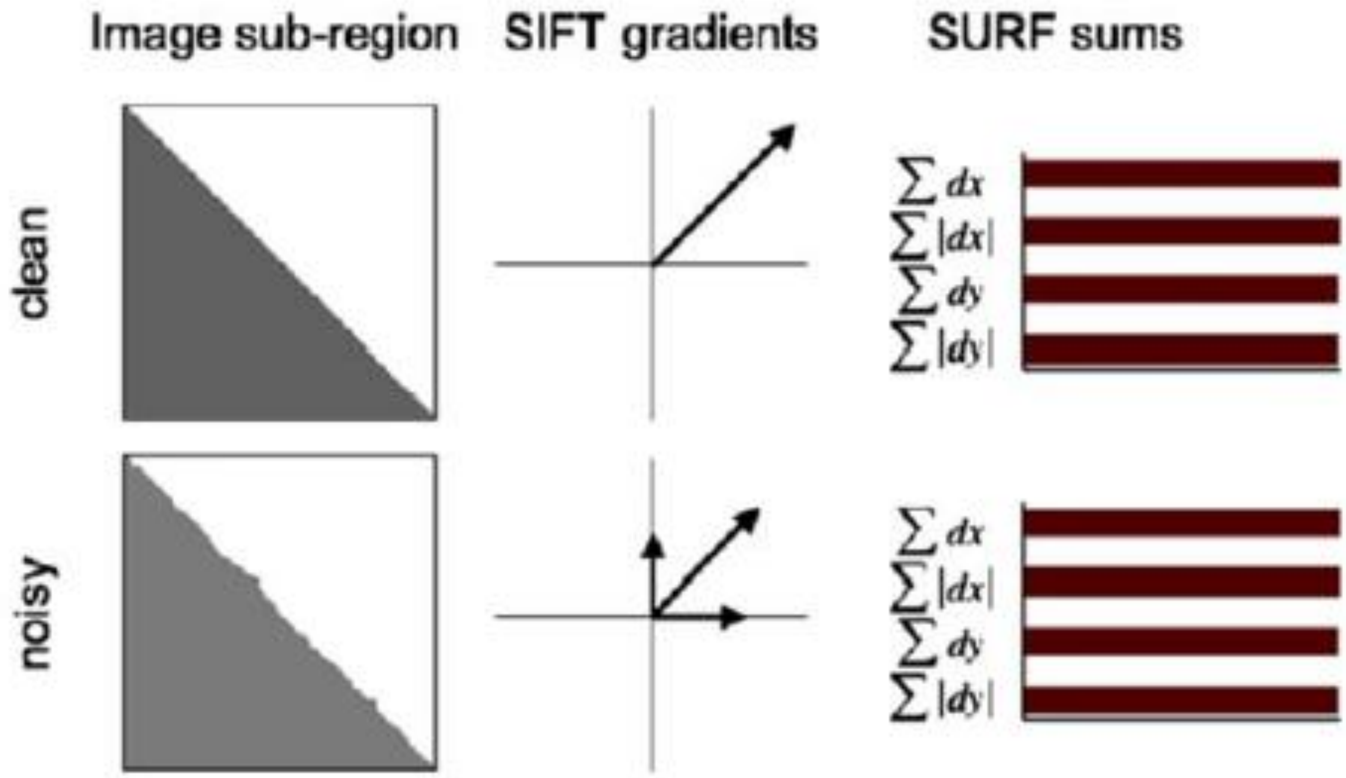
- Split the interest region up into 4 x 4 square sub-regions
- Compute gradients by applying Haar-like features
- Compute *∑dx, ∑|dx|, ∑dy, ∑|dy|:* altogether
- Normalize the vector into unit length



Examples of Descriptors

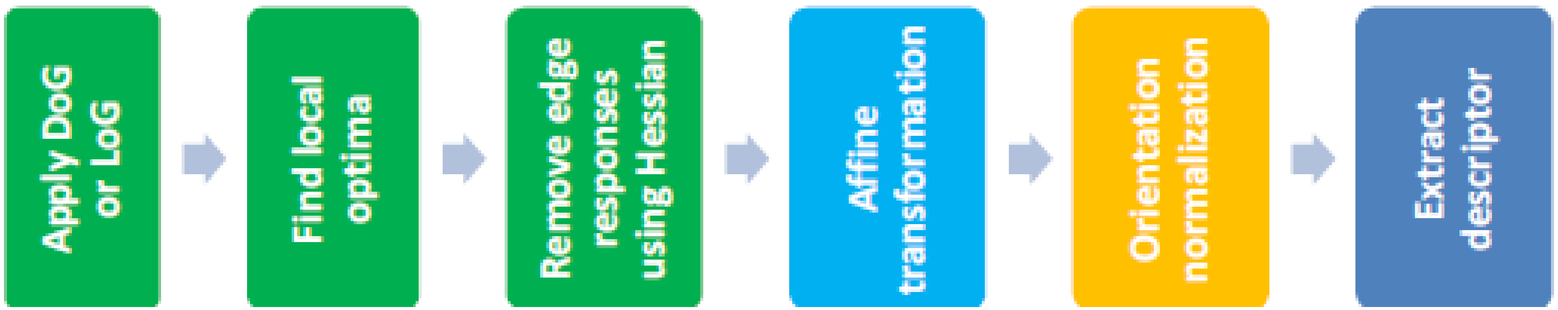


Robustness of SURF

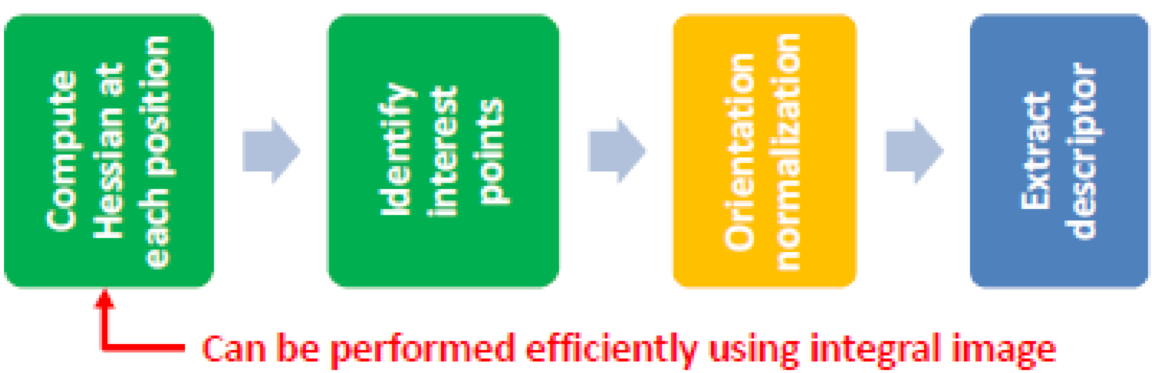


SIFT vs. SURF

- SIFT



- Surf



Sources

- http://cvlab.postech.ac.kr/~bhhan/class/cse441_2015s/csed441_lecture6.pdf
- https://www.cvl.isy.liu.se/education/graduate/VOR14/Lecture3.pdf
- http://luthuli.cs.uiuc.edu/~daf/courses/ComputerVisionTutorial2012/EdgesOrientationHOGSIFT-2012.pdf

Task

- Use libraries to extract image descriptors (GIST)
- Use Pickle module to load dataset
- [may be] Use dimension reduction methods (at least 2)
- Classify images using ML techniques (at least 2)