Computer Vision (ZDO)

Advanced Feature Description

Zdeněk Krňoul, Ph.D.

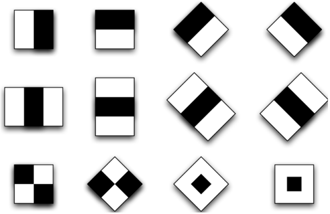
Department of Cybernetics
FACULTY OF APPLIED SCIENCES
University of West Bohemia

Outline:

- Texture Description
  - Haar-like
  - LBP
  - HoG

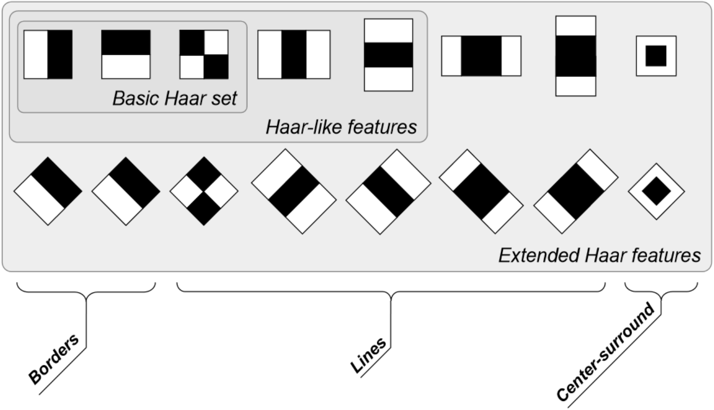
Haar-like features

- Based on Haar walelets
- They were used in the first **real-time** face detector (Viola – Jones, 2001)
- The principle lies in the adjacent rectangular areas in the **specific position** of the sliding window

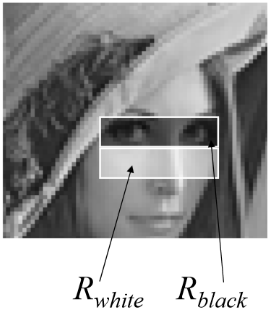


Set of features

- Haar-like features are used for image description similar to cosine transform



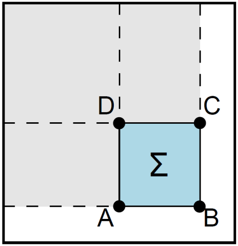
- the black regions are subtracted from the white regions
- the Haar-like filters can be computed efficiently by using integral image



$$F_{Haar} = E(R_{white}) - E(R_{black})$$

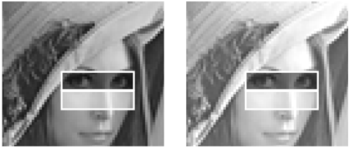
- integral image algorithm

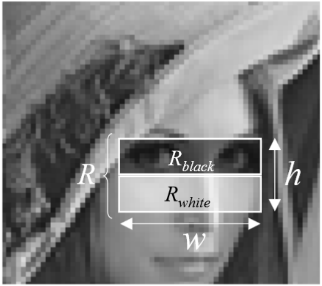
$$I_{\Sigma}(x,y) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i,j) \tag{1}$$



Σ = B - A - C + D

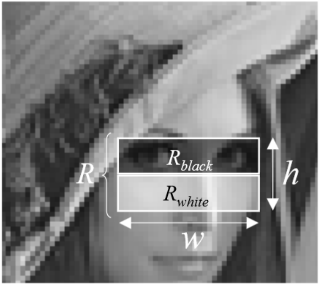
Normalization (monotonic illumination changes)





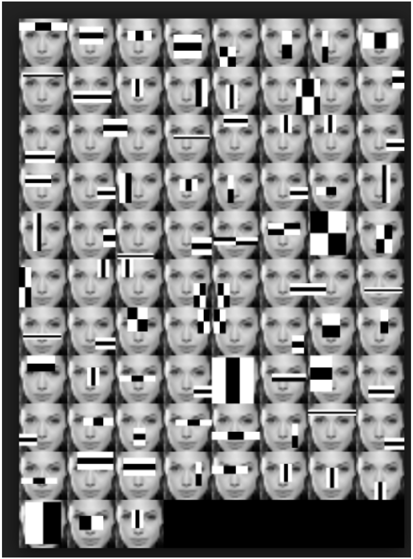
$$F_{Haar} = \frac{E(R_{black}) - E(R_{white})}{\underbrace{\sqrt{\left|E(R_{\mu})^2 - E(R_{\mu}{}^2)\right|}}_{\textit{Standard Deviation}}}$$

Normalization (monotonic illumination changes and scale)



$$F_{Haar} = \frac{E(R_{black}) - E(R_{white})}{w \cdot h \cdot \sqrt{\left| E(R_\mu)^2 - E(R_\mu{}^2) \right|}}$$

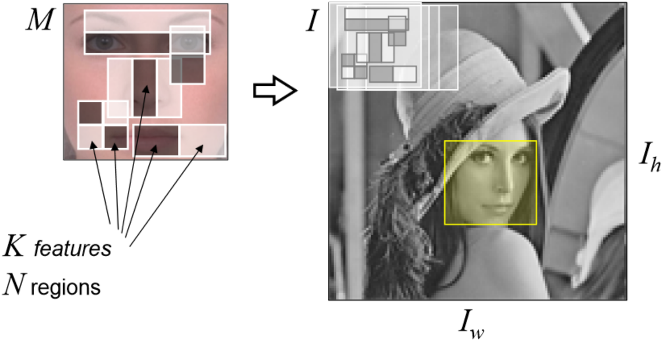
- A large number of Haar-like features are used at once for an accurate and robust description of a texture/object



DEPARTMENT OF CYBERNETICS

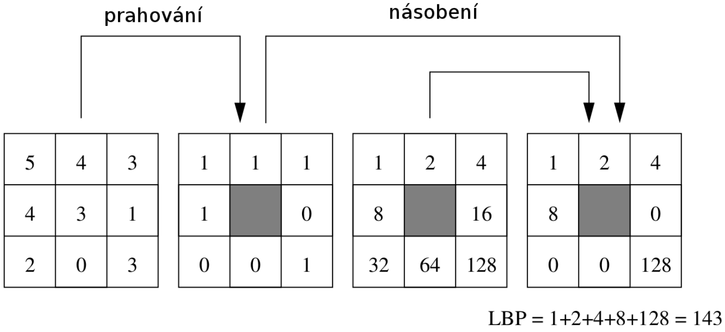
Face detection

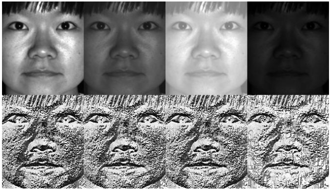
- a boosted classifier is used to train the right responses to certain (most informative) filters
- a sliding window in different scales is used to compute responses on different Haar filters



Local Binary Patterns - LBP

- Ojala 1996: *How to describe texture around a pixel with one scalar?*
- the basic version uses the 8-neighborhood of a pixel
- from this neighborhood a binary representation is build





for a given image patch, a **histogram of LBP** codes is constructed and used as a feature:

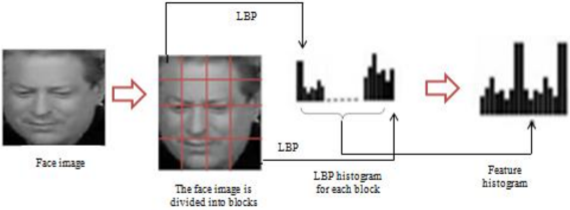


Fig 3: Face description using LBP operator

Rajeshwari et al. "Video Based Face Recognition Using Gabor Features and LBP under Varying Illumination or Pose." (2015)

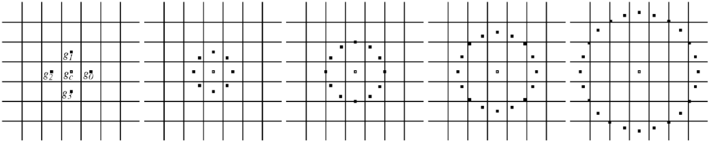
Extensions of LBP

- the basic LBP is invariant to brightness and contrast changes
- they are **variant with scale and rotation** - this is an issue
- texture description:

$$T = t(g_c, g_0, \dots, g_{P-1}) \tag{2}$$

- the position of pixels in the neighborhood is defined as:

$$g_p = (-R\sin(2\pi p/P), R\cos(2\pi p/P)) \tag{3}$$



(P=4,R=1.0) (P=8,R=1.0) (P=12,R=1.5) (P=16,R=2.0) (P=24,R=3.0)

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{4}$$

Rotation invariance and uniformity

- Rotation invariance is achieved by rotating the local neighborhood

$$LBP^{ri}_{P,R} = \min\left(ROR(LBP_{P,R}, i) | i = 0, 1, \ldots, P-1\right) \qquad (5)$$

- *ROR* is a bitwise rotation operator to find the minimal number
- **uniform patterns** - are patterns with at most 2 changes between 0 and 1
- there are a total of 58 uniform patterns, while the rest are put into 59^{th} bin

DEPARTMENT OF CYBERNETICS

Histogram of Oriented Gradients - HoG

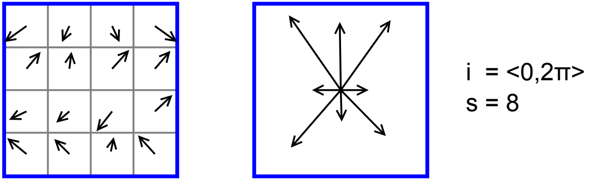
- method for describing images by histogram analysis
- the results of the method are directly dependent on the gradient operator; many had been tested
- the best results were obtained for simple gradient approximation

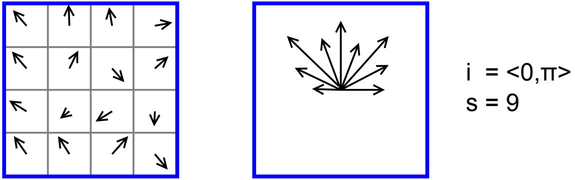
$$\begin{aligned} I_x &= I * [-1, 0, 1] \\ I_y &= I * [-1, 0, 1]^T \end{aligned} \tag{6}$$

- for every pixel, the size and orientation of the gradient are computed
- a histogram is constructed from these values

DEPARTMENT OF CYBERNETICS

- the histogram is parametrized by **interval** i and number of **sectors** s
- the interval i is mostly $i =< 0, \pi >, i =< 0, 2\pi >$
- the magnitude of the histogram is added to each bin and, moreover bilinearly distributed into neighbouring bins





Histogram Normalization

- the normalization is useful to cope with brightness transformations
- the most used normalizations are

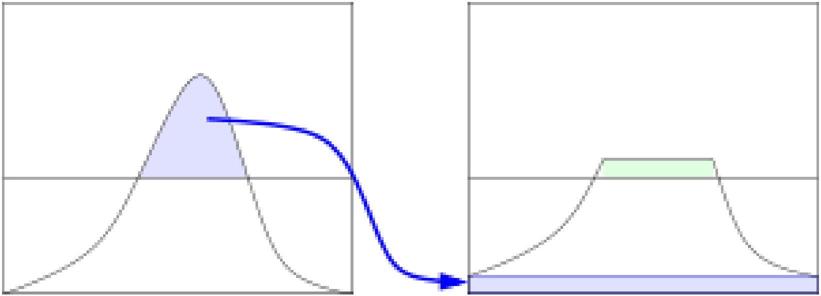
$$L^1 - norm = \frac{v}{(\|v\|_1 + e)}, \quad (7)$$

$$L^1 - sqrt = \sqrt{\frac{v}{(\|v\|_1 + e)}}, \quad (8)$$

$$L^2 - norm = \frac{v}{\sqrt{\left(\|v\|_2^2 + e^2\right)}}, \quad (9)$$

- v is the histogram to be normalized and e is a small constant
- a special case of normalization $L_2 - Hys$ - the normalized vector is clipped as in CLAHE (Contrast Limited Adaptive Histogram Equalization)[1]

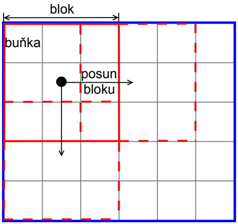
[1] více v předmětu MPV



- the values (after normalization) above a given threshold are distributed into all the bins;
- the process is repeated until no value is above the threshold

Algorithm:

- the image is divided into blocks of size (k, k)
- individual blocks are divided into cells of size (l, l)



- for each cell, the normalized histograms of gradients are computed
- for each block, the cell histograms are averaged
- , then the block shifts by some pixels, and the process is repeated
- the averaged block histograms are concatenated to obtain the descriptor

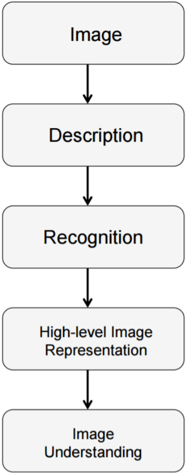
And what else?

- There are additional descriptors based on the brightness properties of the areas
- often blends with the segmentation technique, so I use the detector + descriptor, e.g., for key point matching
- BRIEF, DAISY, ORB, BRISK, SIFT, SURF, KAZE, AKAZE[2]
- CNN an end-to-end method (next time)

[2] more in MPV

Summary:

The usual procedure for computer vision tasks:



Summary:

We know that there are two basic approaches to area representation, i.e. the description of the projection of a real object in a 2D image:

- external descriptions (boundary) - description focused only on the shape
- internal descriptions (area) - description focused on color, texture, etc.

Some descriptions may be invariant to shift rotation and change the scale, and in terms of classification, we must decide whether:

- this property is required
- on the other hand, needs to capture these properties in the description

Descriptions often combine with each other, including a combination of shape and area description (mentioned in MPV)