Lesson 03

Harris, SIFT, SURF

Ing. Marek Hrúz, Ph.D.

Katedra Kybernetiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

DEPARTMENT OF CYBERNETICS

Corner detection

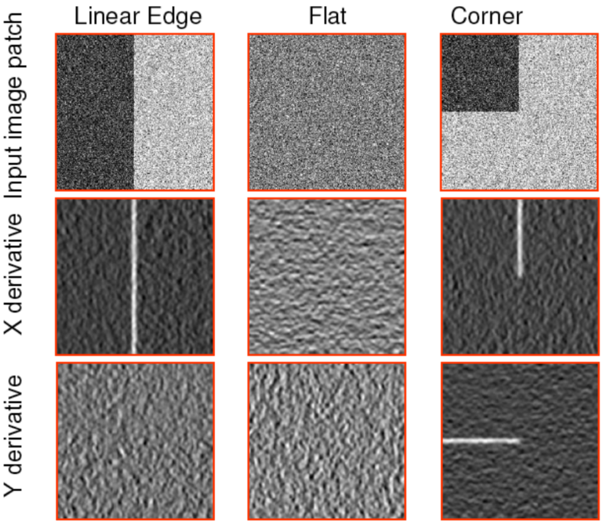
Interest Points

SIFT

SURF

Corner detection

- the corner can be defined as
  1. an intersection of two edges
  2. a (important) point where two dominant directions (gradients) exist
- every corner is an important point, but not the other way around
- a corner detection algorithm needs to be very robust



Obrázek: Different regions and their derivatives.

Moravec corner detection

- one of the first corner detection algorithm
- the alg. tests the similarity of a patch centered on the analyzed pixel with nearby patches
- the similarity is measured as a sum of absolute differences
- the corners are the pixels with a low similarity with its neighborhood - the local maxima of the SoAD

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u,y+v) - I(x,y)]^2. \quad (1)$$

- $(u,v) = \{(1,0),(1,1),(0,1),(-1,1)\}$

Harris corner detection

- reacts to weak points in Moravec algorithm
- the rectangular window $w(x, y)$ becomes a Gaussian window, which functions also as a filter
- the discretized directions (u, v) disappear and are replaced by Taylor expansion

$$I(x+u, y+v) \approx I(x, y) + I_u(x, y)u + I_v(x, y)v \qquad (2)$$

$$E(u, v) \approx \sum_{x,y} w(x, y)[I_u(x, y)u + I_v(x, y)v]^2. \qquad (3)$$

$$E(u, v) \approx \sum_{x,y} w(x, y)[u^2 I_u^2 + 2uvI_uI_v + v^2 I_v^2], \qquad (4)$$

DEPARTMENT OF CYBERNETICS

- which in matrix form can be written as

$$E(u,v) \approx \sum_{x,y} w(x,y) \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} I_u^2 & I_u I_v \\ I_u I_v & I_v^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{5}$$

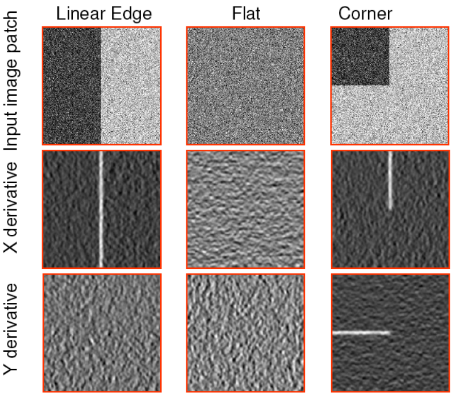
- next we define matrix M

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_u^2(x,y) & I_u I_v(x,y) \\ I_u I_v(x,y) & I_v^2(x,y) \end{bmatrix} \tag{6}$$

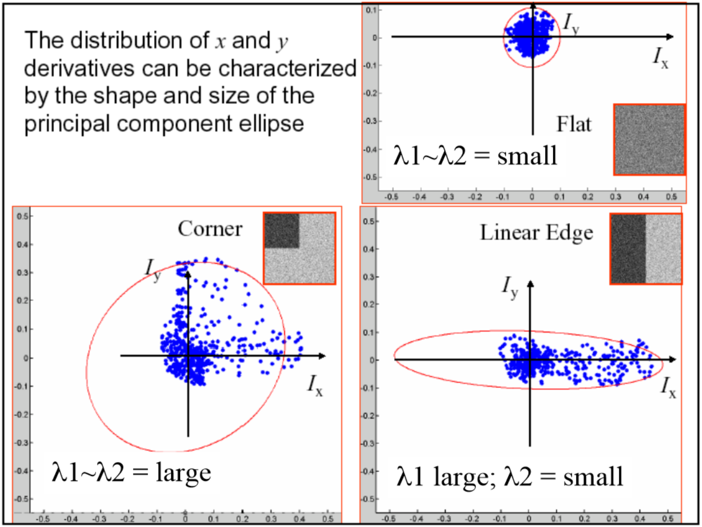
- and then we can write

$$E(u,v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix}. \tag{7}$$

- the matrix M is called a Harris matrix
- the derivations I_u, I_v can be approximated by gradient operators
- for every pixel we have a matrix M and we analyze their eigenvalues



Obrázek: Different regions and their derivatives.



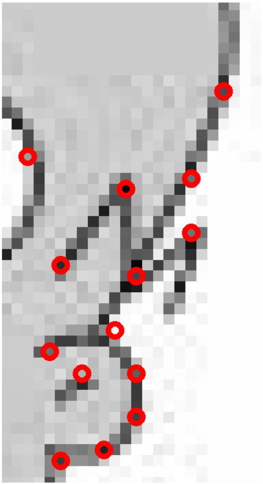
Obrázek: Body proložené elipsami.

- the computational cost of the eigenvalues is very high
- we want the eigenvalues to be relatively the same and also big

$$
\begin{aligned}
\lambda_1 \lambda_2 &= det(M) \\
\lambda_1 + \lambda_2 &= trace(M),
\end{aligned}
\tag{8}
$$

$$
R = det(M) - k(trace(M))^2, \tag{9}
$$

- big $R > 10000$ is a corner
- negative and big $R < -10000$ is an edge
- small $R \in (-10000; 10000)$ is a flat region



Obrázek: Detected corners.

Interest Points

- it has a clear, preferably mathematically well-founded, definition
- it has a well-defined position in image space
- the local image structure around the interest point is rich in terms of local information contents, such that the use of interest points simplify further processing in the vision system
- it is stable under local and global perturbations in the image domain as illumination/brightness variations, such that the interest points can be reliably computed with high degree of reproducibility
- optionally, the notion of interest point should include an attribute of scale, to make it possible to compute interest points from real-life images as well as under scale changes

Scale Invariant Feature Transform

- SIFT is an algorithm that finds interest point
- inspired by Harris corner detection
- the algorithm works the following way:

1. detection of extremes in scale-space representation
2. adjustment of the position of interest points
3. assignment of orientation to the interest points
4. construction of the descriptor of interest point

DEPARTMENT OF CYBERNETICS

Detection of Extremes in Scale-Space

- the scale-space representation is just the image in different resolutions, but with the same width and height
- the different resolution is achieved by convolving the image with a Gaussian kernel

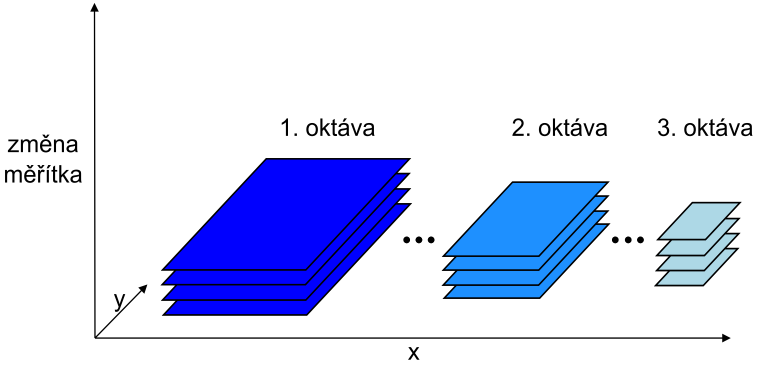
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y),$$
$$\textit{where} \quad G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\left(x^2+y^2\right)}{2\sigma^2}\right). \tag{10}$$

- the Gaussian is self-similar, we can apply it consecutively to obtain more blurred images

DEPARTMENT OF CYBERNETICS

Obrázek: Different scale representations

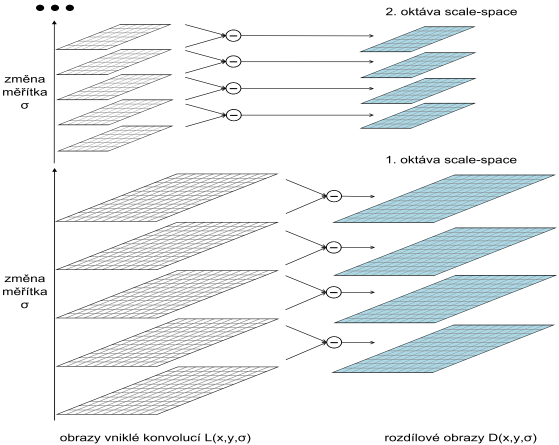
- such images compose an octave
- several octaves are built
- the octave is just the same representation only with smaller width and height

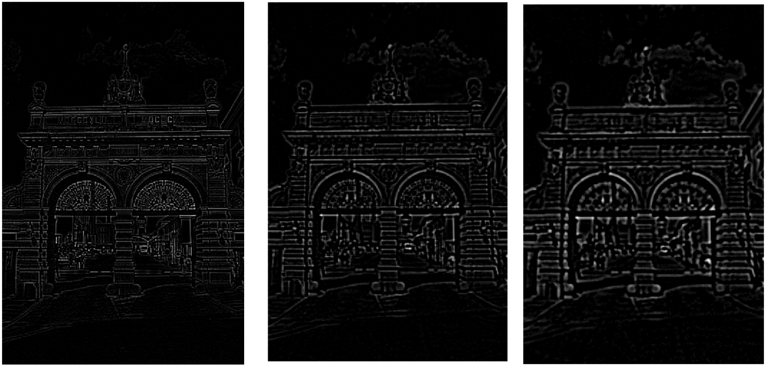


Obrázek: Scale-space representations

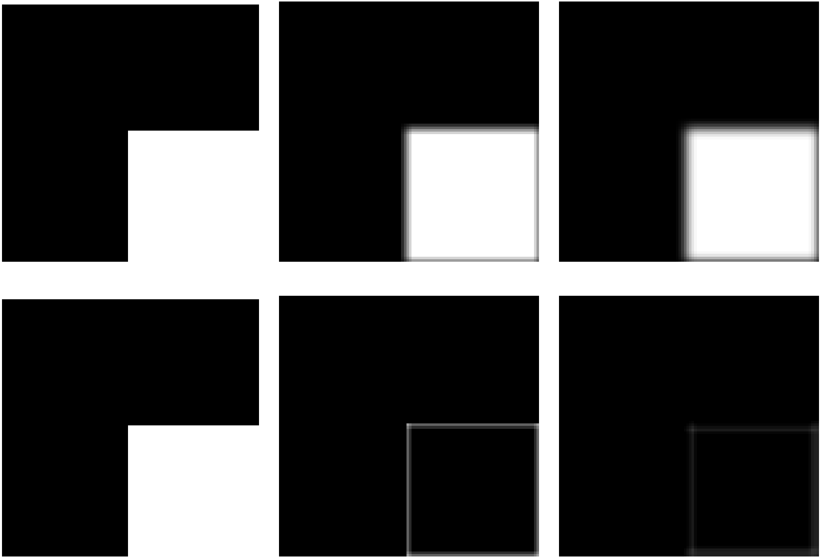
- difference images are constructed by using the octave scale-space representation

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{11}$$



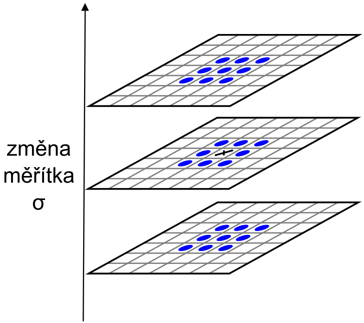


Obrázek: Difference images computed as Difference of Gaussians



Obrázek: Difference images computed as Difference of Gaussians on a corner

- local maxima and minima are detected using non-maxima suppression
- the size of the window is 3x3x3 which means 26 values are compared with the center pixel
- the detected extremes are considered candidates of the interest points



Adjustment of the position of interest points

- the candidate points are fixed on the raster and can be adjusted
- the Taylor expansion is used

$$\tilde{D}(\mathbf{x}) = D + \frac{\partial D^\top}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^\top \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x} \tag{12}$$

- the extreme of the expansion is found by derivation and setting the derivative to zero

$$\frac{\partial \tilde{D}}{\partial \mathbf{x}} = \frac{\partial D}{\partial \mathbf{x}} + \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x} \tag{13}$$

$$\hat{\mathbf{x}} = -\left(\frac{\partial^2 D}{\partial \mathbf{x}^2}\right)^{-1} \frac{\partial D}{\partial \mathbf{x}} \tag{14}$$

DEPARTMENT OF CYBERNETICS

Eliminating low-contrast and edge points

- when we use the $\hat{\mathbf{x}}$ to compute the value of $D(\hat{\mathbf{x}})$ we get

$$D\left(\hat{\mathbf{x}}\right) = D + \frac{1}{2}\frac{\partial D^\top}{\partial \mathbf{x}}\hat{\mathbf{x}} \tag{15}$$

- we use the value of $D\left(\hat{\mathbf{x}}\right)$ to eliminate low contrast key-points $(\ < 0.03\ )$
- we also want to eliminate unstable key-points - edge points
- we use similar algorithm as in Harris corner detector - the analysis of eigenvalues of Hess (not Harris) matrix

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \tag{16}$$

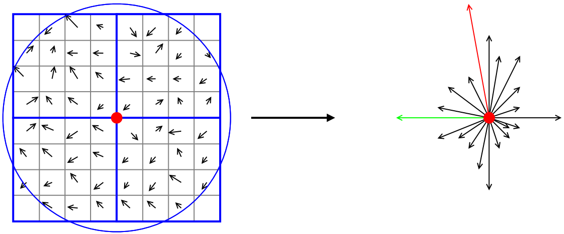
$$\frac{\mathrm{Tr}\left(\mathbf{H}\right)^2}{\mathrm{Det}\left(\mathbf{H}\right)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \tag{17}$$

$$\frac{\mathrm{Tr}\left(\mathbf{H}\right)^2}{\mathrm{Det}\left(\mathbf{H}\right)} < \frac{(r+1)^2}{r} \tag{18}$$

DEPARTMENT OF CYBERNETICS

Assigning the orientation to the key-points

- to make the key-points independent on rotation we have to find their "main"orientation
- in the image $L(x, y, \sigma)$ in the key-point we find the magnitudes and directions of the image gradient
- the directions are quantified into bins of 36°



gradienty v okolí

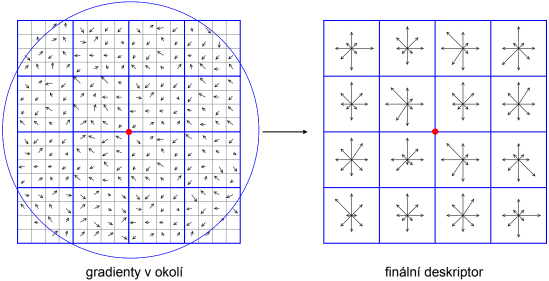
histogram orientací

- if there are more important directions (at least 80% of the biggest) then new key-points are established in the same pixel

DEPARTMENT OF CYBERNETICS

The Key-point descriptor

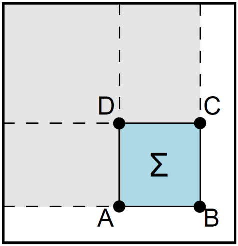
- a description should be independent on geometric and brightness transformations
- the neighborhood of the key-point is divided into 4x4 regions
- in each region the gradients are computed
- the orientations of the gradients are then rotated to align with the dominant direction
- they are concatenated into a 128-dimensional feature vector



SURF - Speeded Up Robust Features

- inspired by SIFT with real-time capabilities
- the DoG images and computing of Hess matrix is integrated into computing the determinant of Hess matrix
- this approach is using the integral image

$$I_{\Sigma}(x,y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i,j) \tag{19}$$



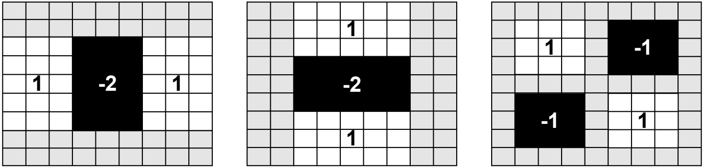
Σ = B - A - C + D

Hess matrix approximation

- the Hess matrix can be written as

$$H(x,y,\sigma) = \begin{bmatrix} L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\ L_{yx}(x,y,\sigma) & L_{yy}(x,y,\sigma) \end{bmatrix} \tag{20}$$

- the approximation uses discrete convolution with kernels



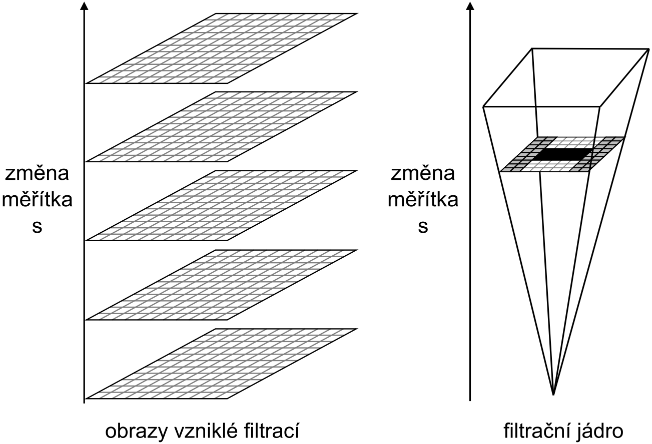
- the determinant of Hess matrix is then computed as

$$\mathrm{Det}(\mathbf{H}_{aprox}) = D_{xx}D_{yy} - (wD_{xy})^2 \tag{21}$$

DEPARTMENT OF CYBERNETICS

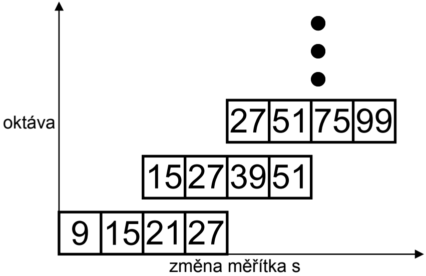
Scale-space approximation

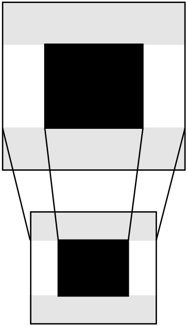
- the scale-space does not need to be constructed explicitly
- different sizes of the kernels fulfill this operation



DEPARTMENT OF CYBERNETICS

- the different octaves are constructed by using different combinations of sizes of the kernels



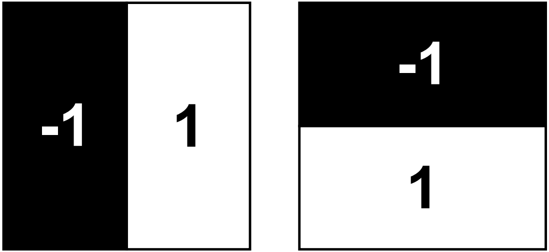


Obrázek: Změny rozměrů filtračních jader pro jednotlivé oktávy scale-space (vlevo) a názorná ukázka změny rozměru jádra (vpravo). Poznamenejme, že krok l_0 je vždy sudý (6, 12, 24) tak, aby při zvyšování měřítka nedocházelo ke změně struktury filtračních jader.

- again, the key-points are local extremes of the determinants of Hess matrix

Orientation of the Key-points

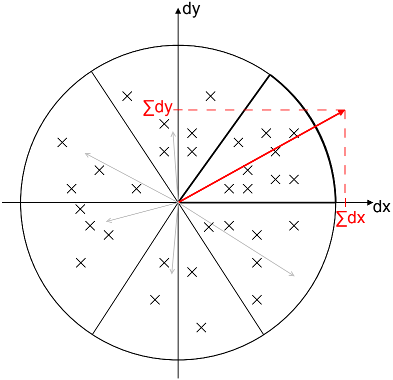
- the Haar filters are used to approximate the orientation of the gradients
- the size of the filters is relative to the scale (4σ) at which the key-point is detected



Obrázek: Haarova vlnka aproximovaná obdélníkovými filtry ve směru osy x a y.

- the responses are filtered with a Gaussian

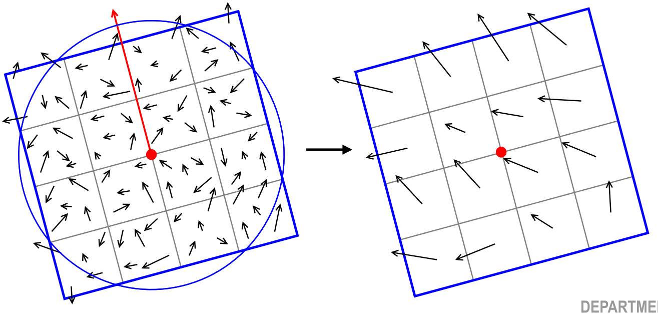
- the space of the responses (d_x, d_y) is divided into several segments



- the dominant direction is the one with the biggest sum of vectors inside it

The SURF descriptor

- a neighborhood around the key-point is constructed and rotated by the angle of the dominant direction
- the neighborhood is of size 20σ
- this patch is divided into 4×4 segments
- for each segment the responses of the Haar filter is computed - (d_x, d_y)
- the descriptor is then a vector $(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$



Application of key-points

- https://www.youtube.com/watch?v=-r9J1eO4qg4