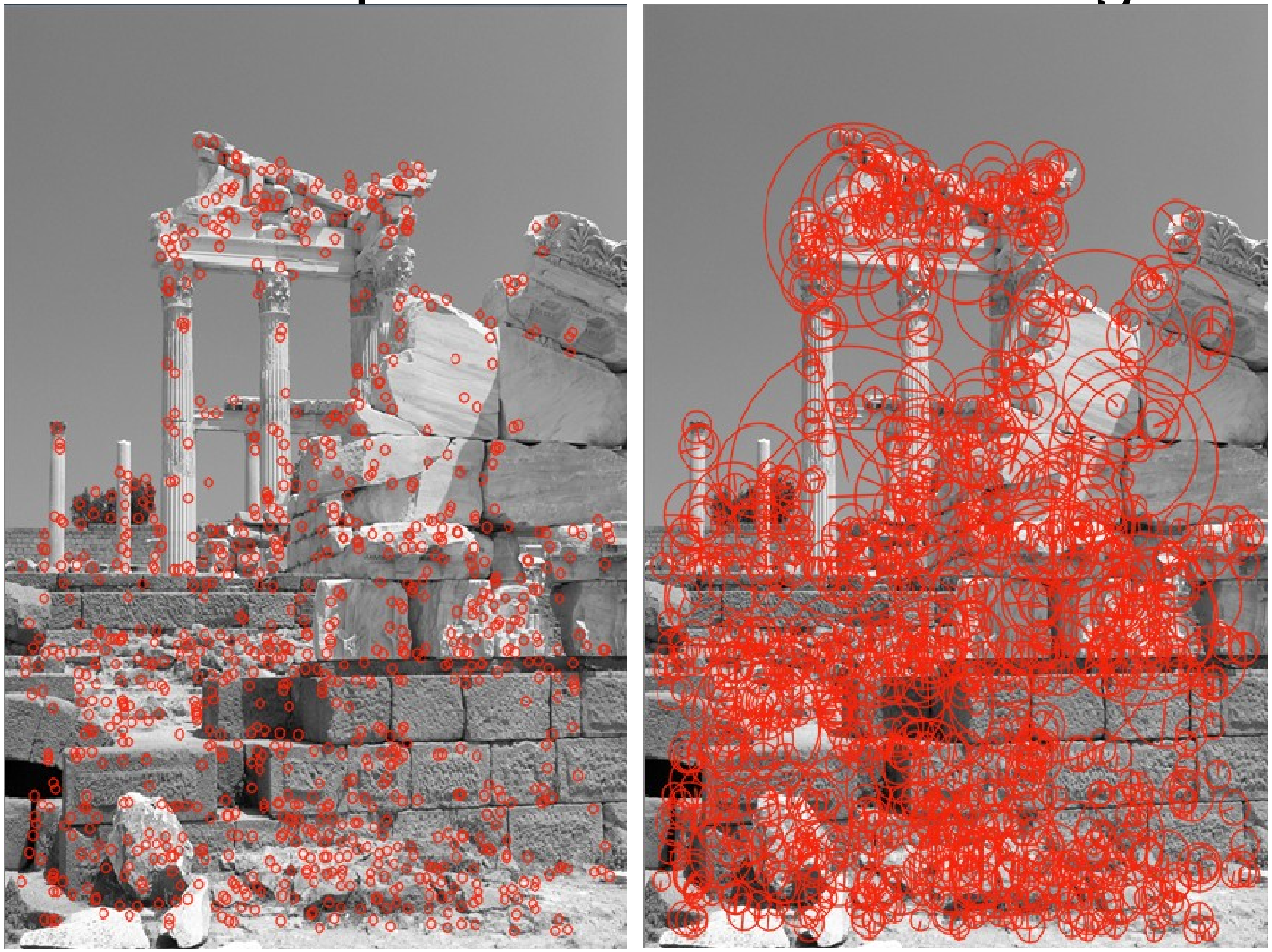
Interest points detection and description

Pavel Cagaš
395960

Interest points

- Points in an image suitable for subsequent processing
- Points with interesting surrounding
- Usually found in places with high contrast difference – edges, corners...
- Useful for many computer vision applications:
  - Object detection
  - Object recognition
  - Motion analysis
  - ...

Interest points found in an image



SURF – Speeded-Up Robust Feature

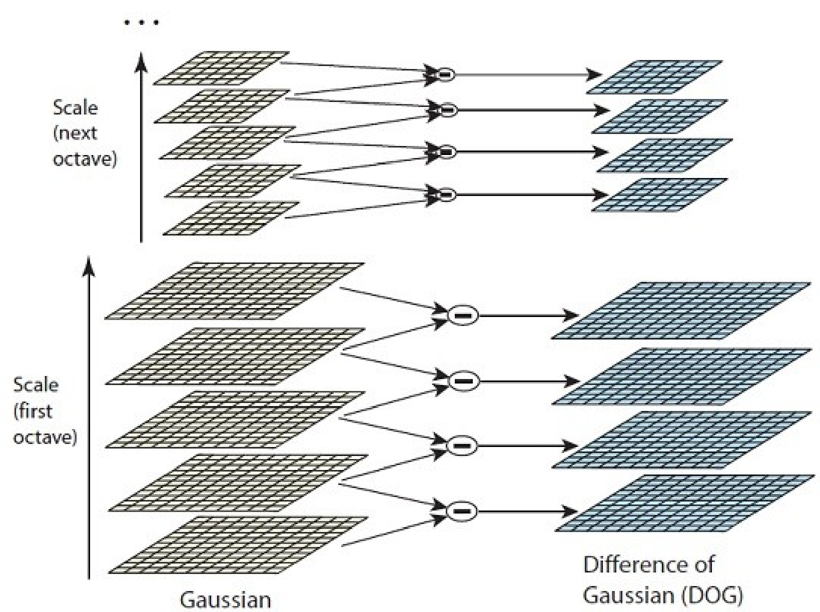
- Method for detecting and desribing interest points in an image
- Scale and rotation invariance
- Four parts needed for understanding:
  - Scale-Space theory
  - Integral image
  - Interest point detection
  - Interest point description

Scale-Space

- Real world objects are relevant only in certain scale
- Importance of internal representation
- Object as a whole or detailed structure?
- Need of representation in all scales
- Finding a way of getting rid of some image detail
- Frequency domain:

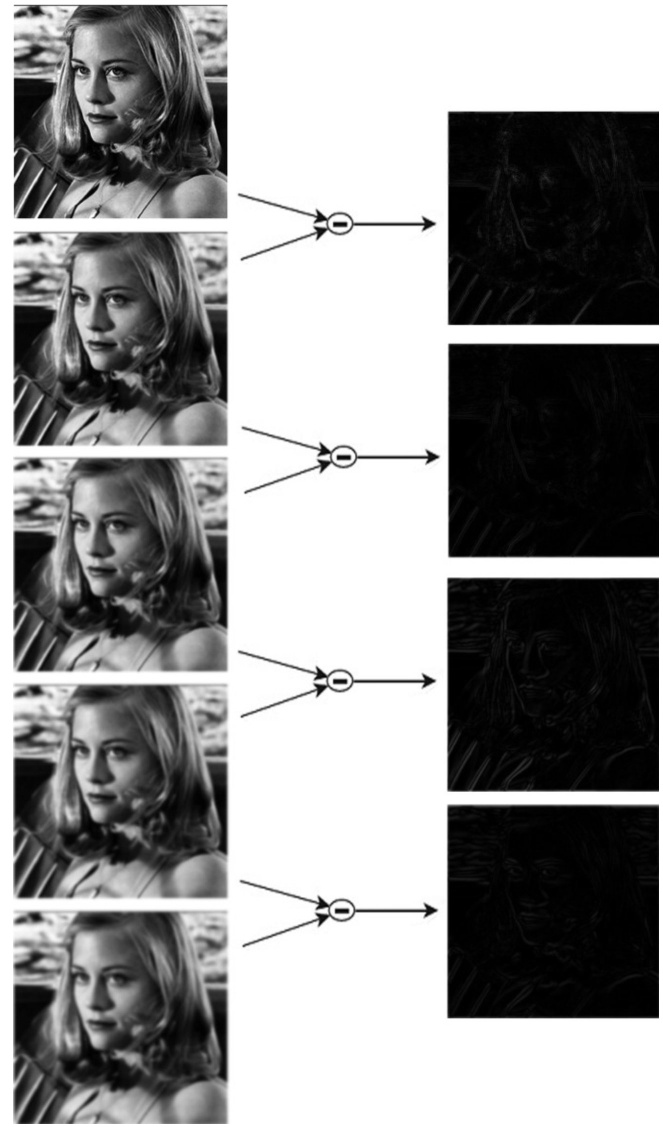
  high frequency – detailed structure

- Possible solution – Gaussian filter
- Greater σ parameter $\rightarrow$ greater blurring $\rightarrow$ detail suppression
- Scale-Space layer: $L(x,y,\sigma) = G(x,y,\sigma) * I(x,y)$
- Scale-Space pyramid
- DoG rather than LoG
- $D(x,y,\sigma) = L(x,y,k\sigma) – L(x,y,\sigma)$



Scale-space octaves

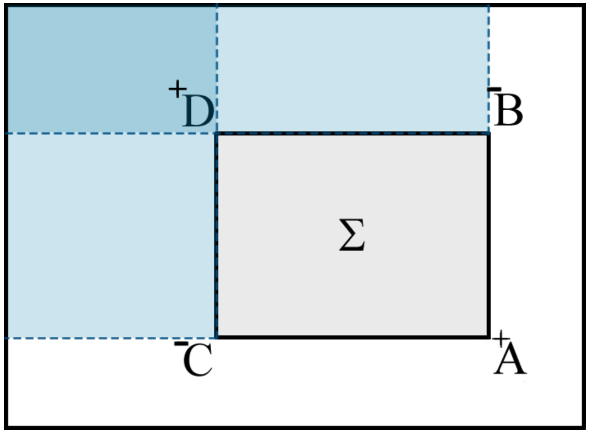
Difference of Gaussians



Integral image

- Structure generated from the input image for computing sum of any rectangular area in constant time
- Every pixel of integral image has value of sum of pixels in rectangular area of the input image defined by origin and given pixel

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



Interest point detection

- Hessian matrix determinant based detection
- Matrix of partial derivatives of function *f*

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

- The value of the determinant is used to clasify the results into local maxima or minima, based on second order derivative test

$$\mathcal{H}(x,y) = \det \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$$

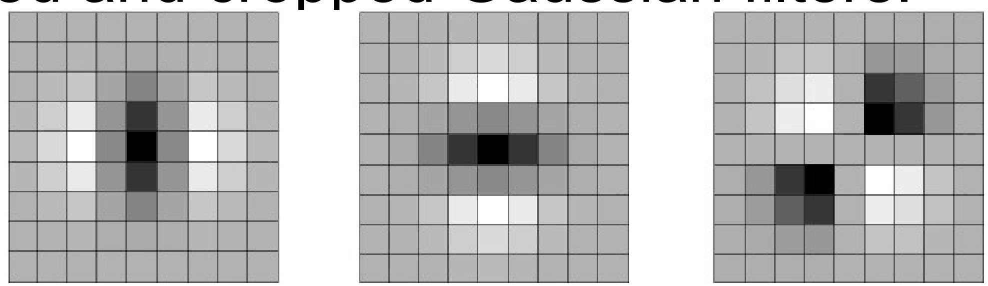
- Conversion into discrete image domain:

$$H(x, y, \sigma) = \begin{pmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{pmatrix}$$

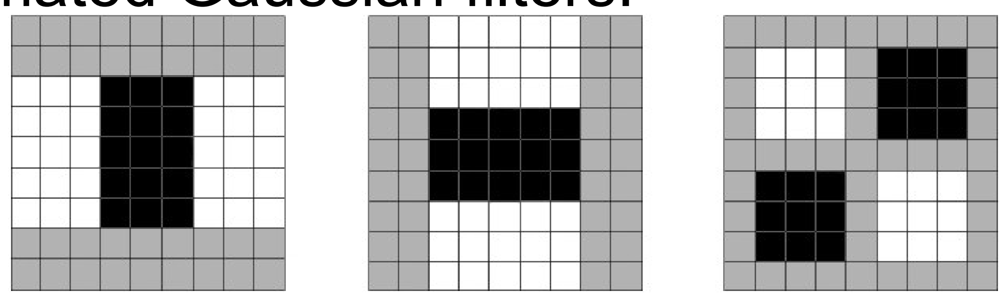
- Where L_{xx} is convolution of given point with Gaussian second order derivative at point x $\frac{\partial^2 g(\sigma)}{\partial x^2}$ d similarly for others
- Calculated for every point in am image

Fast Hessian

- Discretized and cropped Gaussian filters:



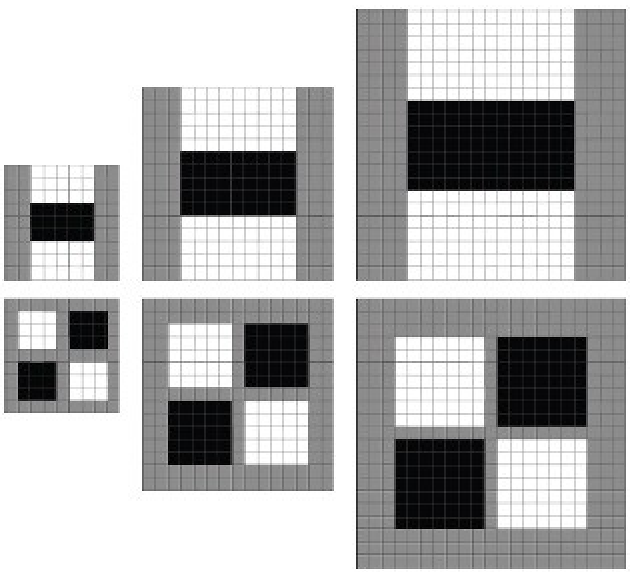
- Approximated Gaussian filters:

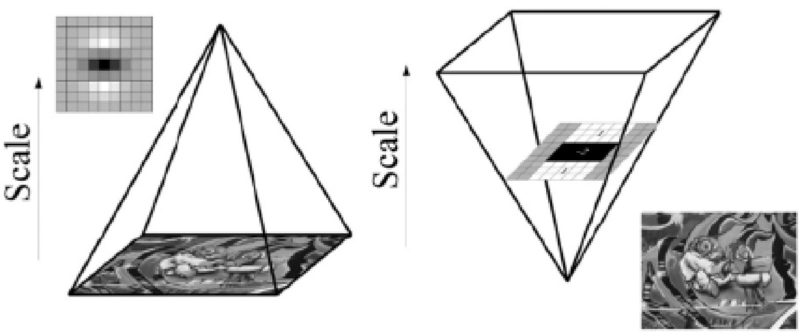


- Deviation correction: $\det(H_{approx}) = DxxDyy - (0,9*Dxy)^2$

Scale Invariance

- In order to detect points on different scales, we increase the size of Box Filters rather than progressively blurring and sub-sampling the input image

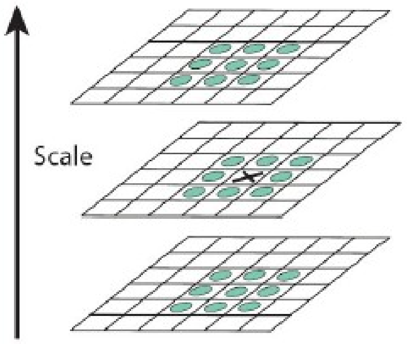




$$\sigma_{approx} = \text{Current Filter Size} \cdot \frac{\text{Base Filter Scale}}{\text{Base Filter Size}}$$
$$= \text{Current Filter Size} \cdot \frac{1.2}{9}$$

Interest point localization

- Response map thresholding
- Determining local maxima in areal surronding of a point
  - comparing each point with 26 points in total of three scales of the same octave



- Interpolating data into sub-pixel position
  - Hessian determinant $\rightarrow$ Taylor expansion up to quadratic terms

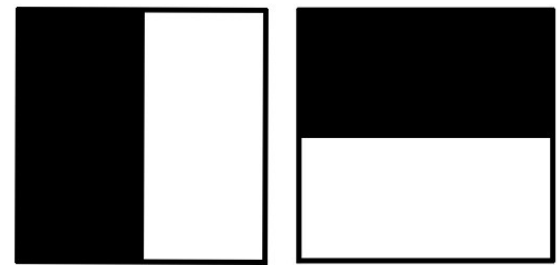
$$H(\mathbf{x}) = H + \frac{\partial H}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 H}{\partial \mathbf{x}^2}\mathbf{x}$$

$$\hat{x} = -\frac{\partial^2 H}{\partial \mathbf{x}^2}^{-1} \frac{\partial H}{\partial \mathbf{x}}$$

Interest point description

- Desriptor is a 64-dimensional vector used for describing each detected interest point
- In order to keep rotation invariance, we need to determine each interest point orientation
- Descriptor extraction is performed relative to this orientation

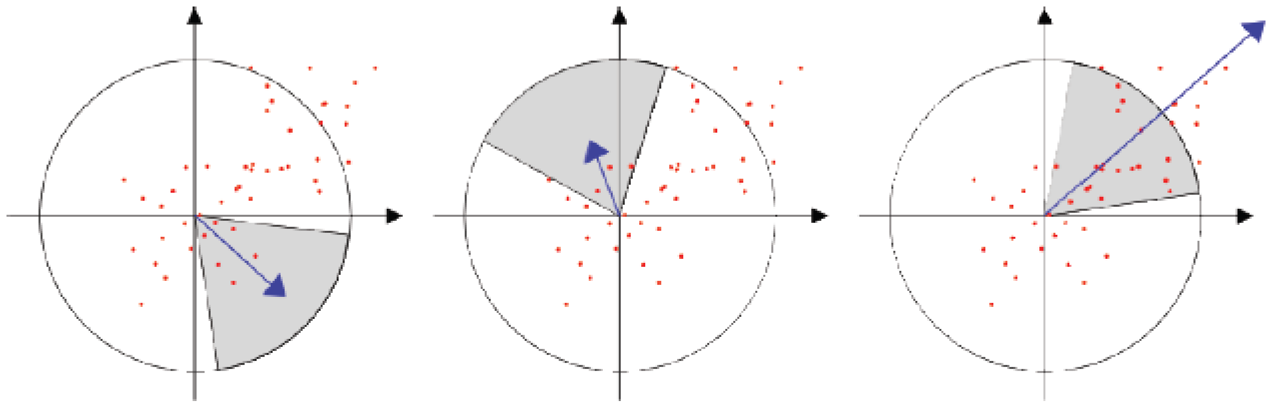
Assigning orientation



- Using Haar wavelets for *x* and *y* direction
- We calculate responses in a circular area around the interest point and weight it with Gaussian
- All sizes relative to

  Responses are represented as points in space with their own x and y value

- Orientation vectors are calculated in a circular sector of $\pi/3$ rotating around the iterest point
- Dominant orientation is determined as the greatest calculated vector

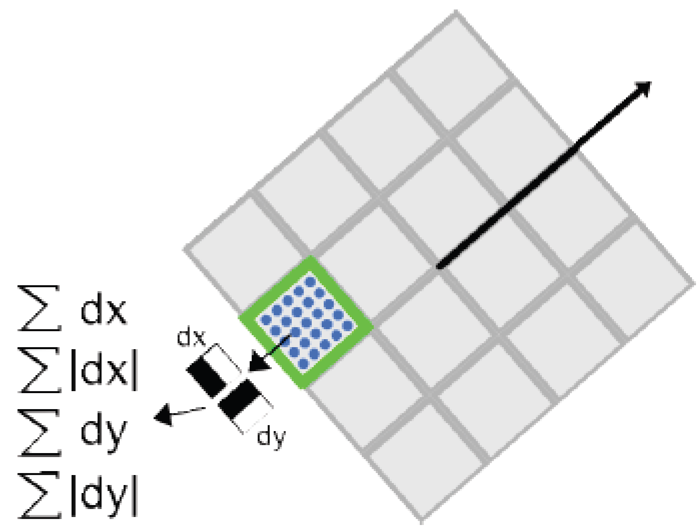


Descriptor extraction

- Rectangular window around interest point is created
- Haar wavelets in every of 16 subregions are calculated for 25 evenly distributed

  sample points

$$v_{subregion} = \left[\sum dx, \sum dy, \sum |dx|, \sum |dy|\right]$$



Thank you for attention

Feel free to ask any question