Problems with template matching

- The template represents the object as we expect to find it in the image
- The object can indeed be scaled or rotated
- This technique requires a separate template for each scale and orientation
- Template matching become thus too expensive, especially for large templates
- Sensitive to:
  - noise
  - occlusions

Local template matching

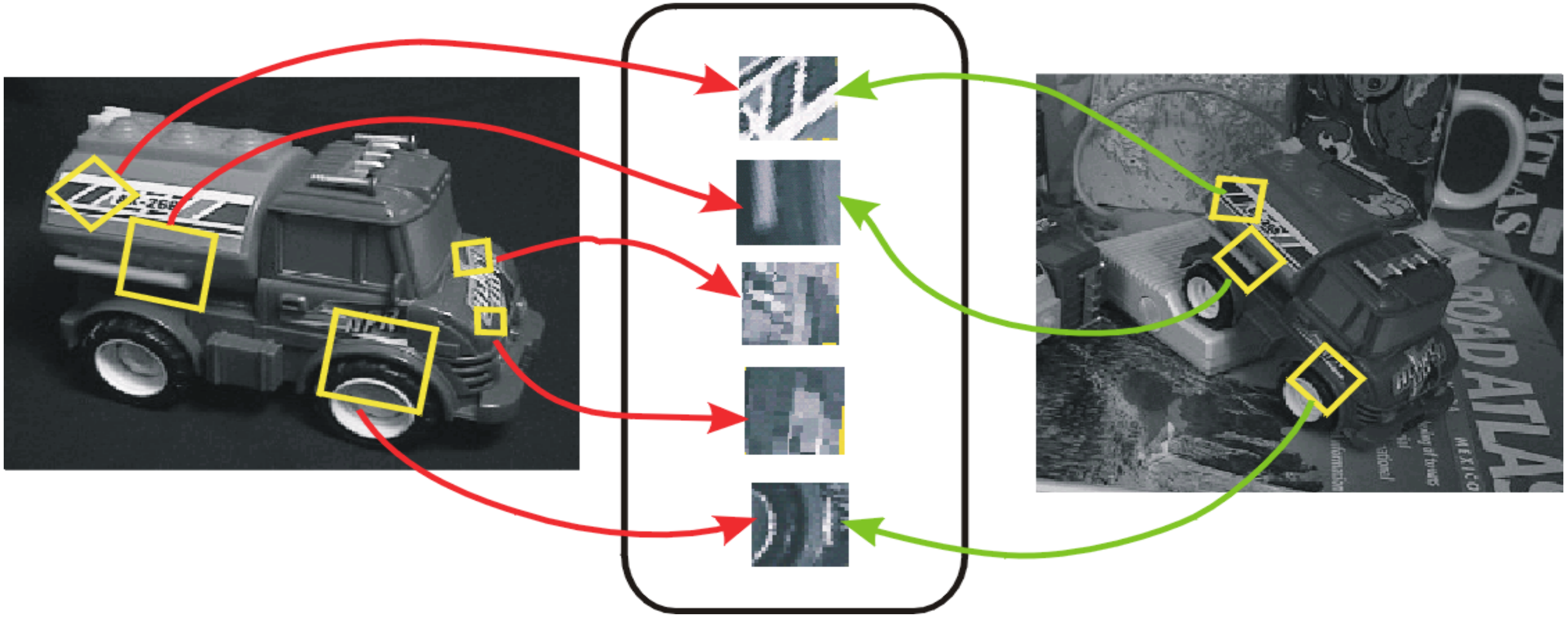
- A possible solution is to reduce the size of the templates, and detect *salient* features in the image that characterize the object we are interested in
- Extract a set of local features that are invariant to translation, rotation and scale…
- Perform matching only on these local features
- We then analyze the spatial relationships between those features

See for example:

Corner detector (Harris and Stephens,1988)

SIFT (Lowe, 1999)

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters



Applications

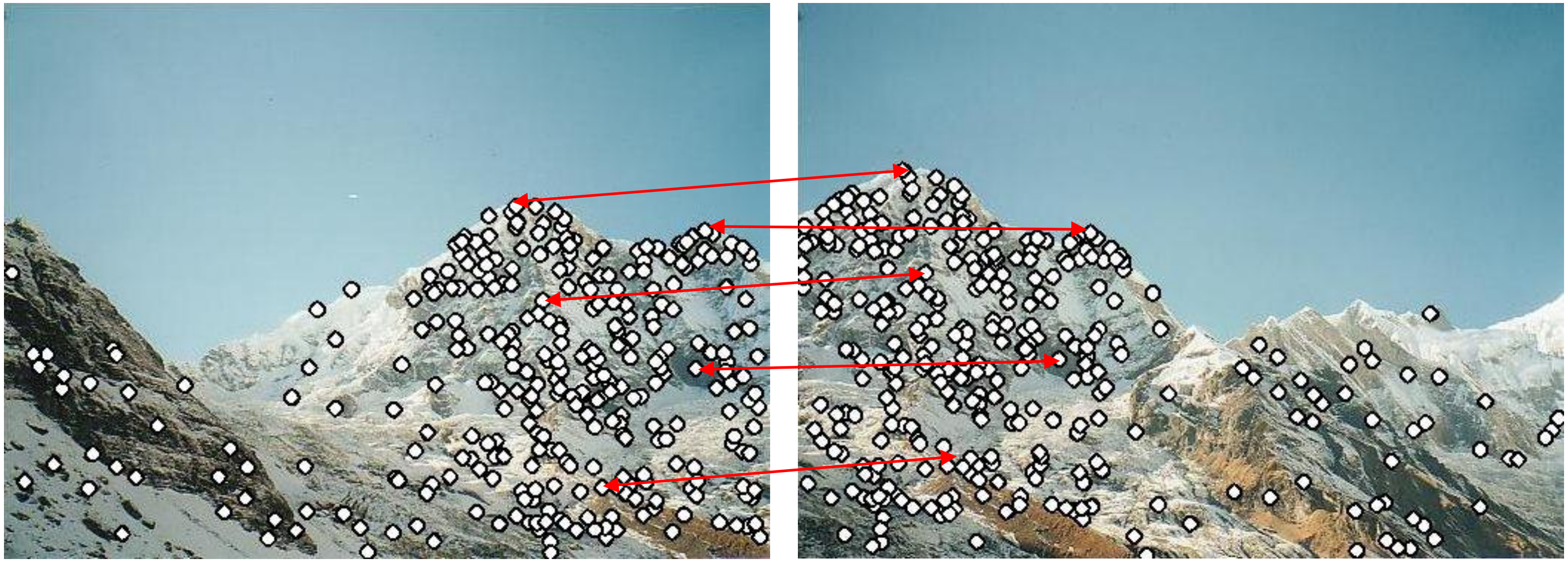
- Object recognition
- 3D reconstruction
- Motion detection
- Panorama reconstruction

How do we build a panorama?

- We need to match (align) images

Matching with Features

- Detect feature points in both images
- Find corresponding pairs



Matching with Features

- Detect feature points in both images
- Find corresponding pairs
- Use these pairs to align images

Matching with Features

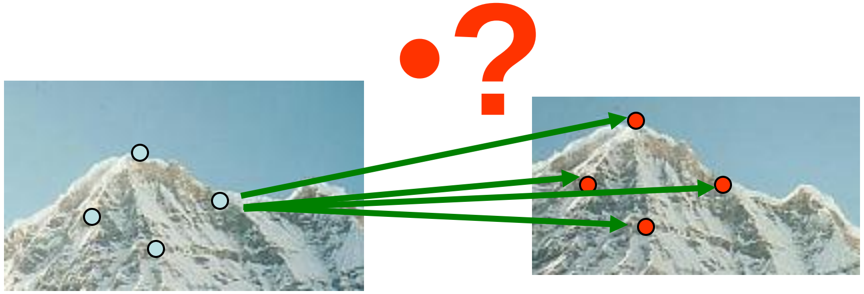
- Problem 1:
  – Detect the *same* point *independently* in both images

•no chance to match!

•We need a repeatable detector

Matching with Features

- Problem 2:
  - For each point correctly recognize the corresponding one

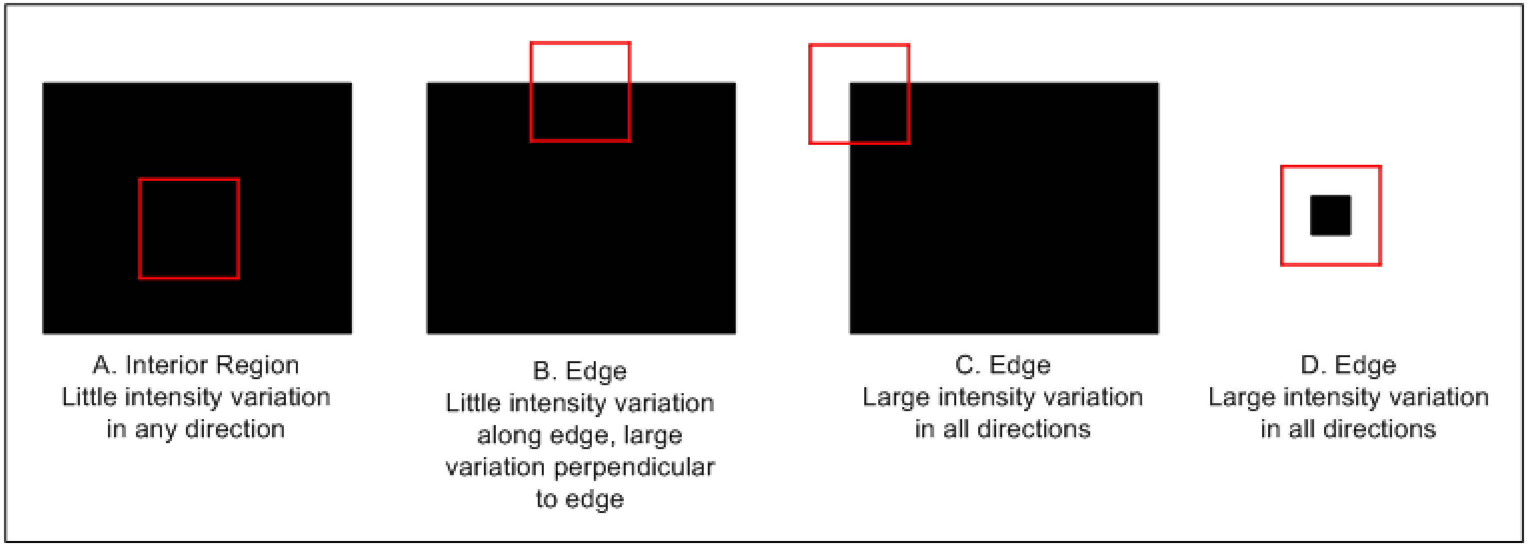


•We need a reliable and distinctive descriptor

Moravec interest operator

- This operator was developed back in 1977 for navigation
- Defines “points of interest” regions in the image that are good candidates for matching in consecutive image frames
- It is considered a “corner detection” since it defines interest points as points in which there is large intensity variation in every direction

- Place a small square window (3x3, 5x5, 7x7…) centered at a point P
- Shift this window by one pixel in each of eight principle directions (four diagonals, horizontal and vertical)
- Take the sum of squares of intensity differences of corresponding pixels in these two windows
- Define the intensity variation V as the *minimum* intensity variation over the eight principle directions


A. Interior Region
Little intensity variation
in any direction
B. Edge
Little intensity variation
along edge, large
variation perpendicular
to edge
C. Edge
Large intensity variation
in all directions
D. Edge
Large intensity variation
in all directions

Algorithm

for each x,y in the image

$$V_{u,v}(x,y) = \sum_{a,b} \left(I(x+u+a, y+v+b) - I(x+a, y+b) \right)^2$$

(u,v) are the considered shifts: $(1,0),(1,1),(0,1),(-1,1),(-1,0),(-1,-1),(0,-1),(1,-1)$

$$C(x,y) = \min(V_{u,v}(x,y))$$

threshold, set C(x,y) to zero of C(x,y)<T

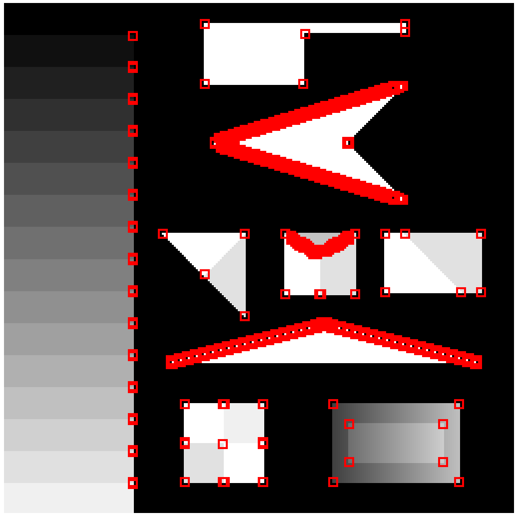
non-maximal suppression to find local maxima

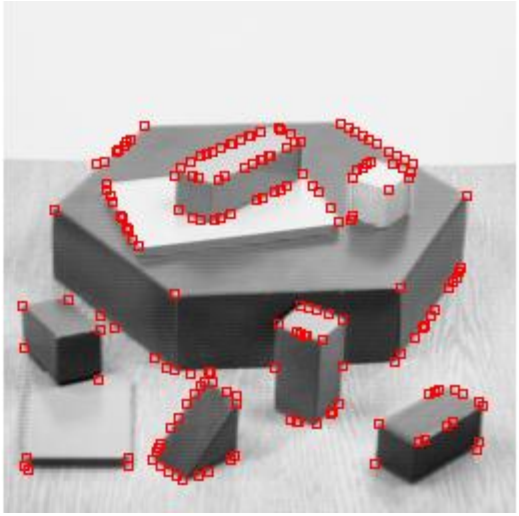
Non maximal suppression

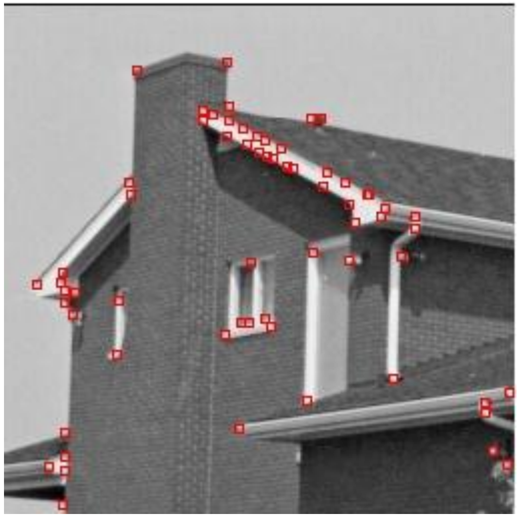
$$I(x,y)=\begin{cases}1 & if \quad V(x,y)\geq V(p,q), \forall (p,q)\in N(x,y)\\ 0 & otherwise\end{cases}$$

Source: http://www.cim.mcgill.ca/~dparks/CornerDetector/index.htm

Results







- Noise sensitivity
- Small imperfection in edges will be picked up as corners
- Edges not oriented along one of the height directions will also be candidate corners

Harris and Stephens

- Similarly to Moravec, consider now a continuous function:

$$E_h(\mathbf{x}) = \sum_{d \in W} \left(I(\mathbf{x}+\mathbf{d}) - I(\mathbf{x}+\mathbf{d}+\mathbf{h}) \right)^2$$

$$\mathbf{x} = (x, y) \in R^2$$

$$\mathbf{d}, \mathbf{h} \in R^2$$

- Perform local Taylor expansion:

$$I(\mathbf{x}+\mathbf{h}) \approx I(\mathbf{x}) + \nabla I(\mathbf{x})^T \mathbf{h} \Rightarrow I(\mathbf{x}+\mathbf{h}) - I(\mathbf{x}) = \nabla I(\mathbf{x})^T \mathbf{h}$$

$$\nabla I(\mathbf{x}) = \left[\frac{\partial I(\mathbf{x})}{\partial x}, \frac{\partial I(\mathbf{x})}{\partial y} \right]^T$$

$$E_h(\mathbf{x}) = \sum_{d \in W} \left(\nabla I(\mathbf{x}+\mathbf{d})^T \mathbf{h} \right)^2 = \sum_{d \in W} \mathbf{h}^T \nabla I(\mathbf{x}+\mathbf{d}) \nabla I(\mathbf{x}+\mathbf{d})^T \mathbf{h} =$$

$$= \sum_{d \in W} \mathbf{h}^T \cdot \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \cdot \mathbf{h} =$$

$$= \mathbf{h}^T \begin{bmatrix} \sum_W I_x^2 & \sum_W I_x I_y \\ \sum_W I_x I_y & \sum_W I_y^2 \end{bmatrix} \mathbf{h}$$

$$I_x = \frac{\partial I(\mathbf{x}+\mathbf{d})}{\partial x}$$

$$I_y = \frac{\partial I(\mathbf{x}+\mathbf{d})}{\partial y}$$

It is possible to introduce a weighting window w() for each point. For example a Gaussian:

$$w(\mathbf{d}) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\|\mathbf{d}\|^2}{2\sigma^2}}$$

…and compute the filtered variation

$$\hat{E}_h(\mathbf{x}) = h^T \hat{C} h$$

$$\hat{C} = \begin{bmatrix} \sum I_x^2 w(\mathbf{d}) & \sum I_x I_y w(\mathbf{d}) \\ \ldots & \ldots \end{bmatrix}$$

the eigenvalues of C will be proportional to the principle curvatures of the image surface and form a rotationally invariant description of C

from linear algebra:

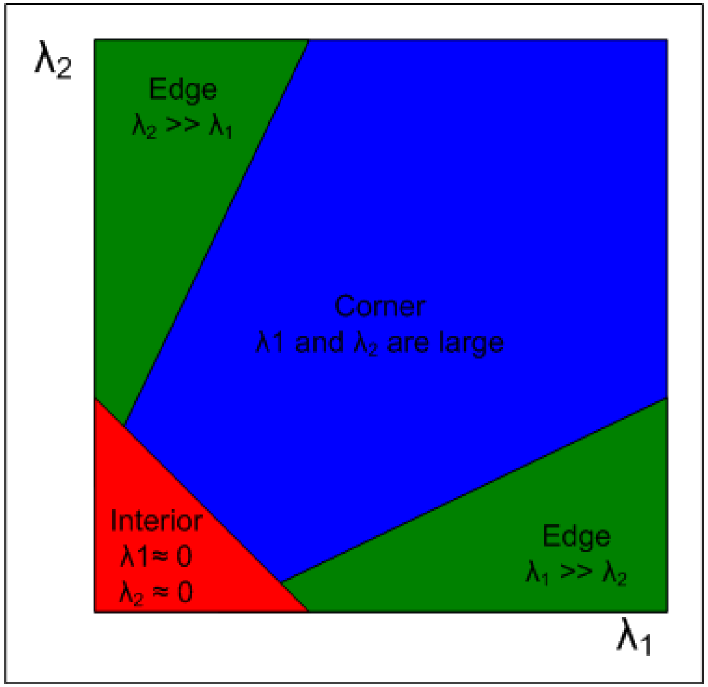
$$E_h(\mathbf{x}) = \lambda_1 v_1{}^2 + \lambda_2 v_2{}^2$$

Three cases

$$\lambda_1 \approx \lambda_2 \Rightarrow \text{no structure}$$

$$\lambda_1 >> \lambda_2 \approx 0 \text{ or } \lambda_2 >> \lambda_1 \approx 0 \Rightarrow \text{edge}$$

$$\lambda_1 >> 0 \text{ and } \lambda_2 >> 0 \Rightarrow \text{corner}$$



define the following index R

$$R = \det(\hat{C}) - k \cdot tr^2(\hat{C})$$

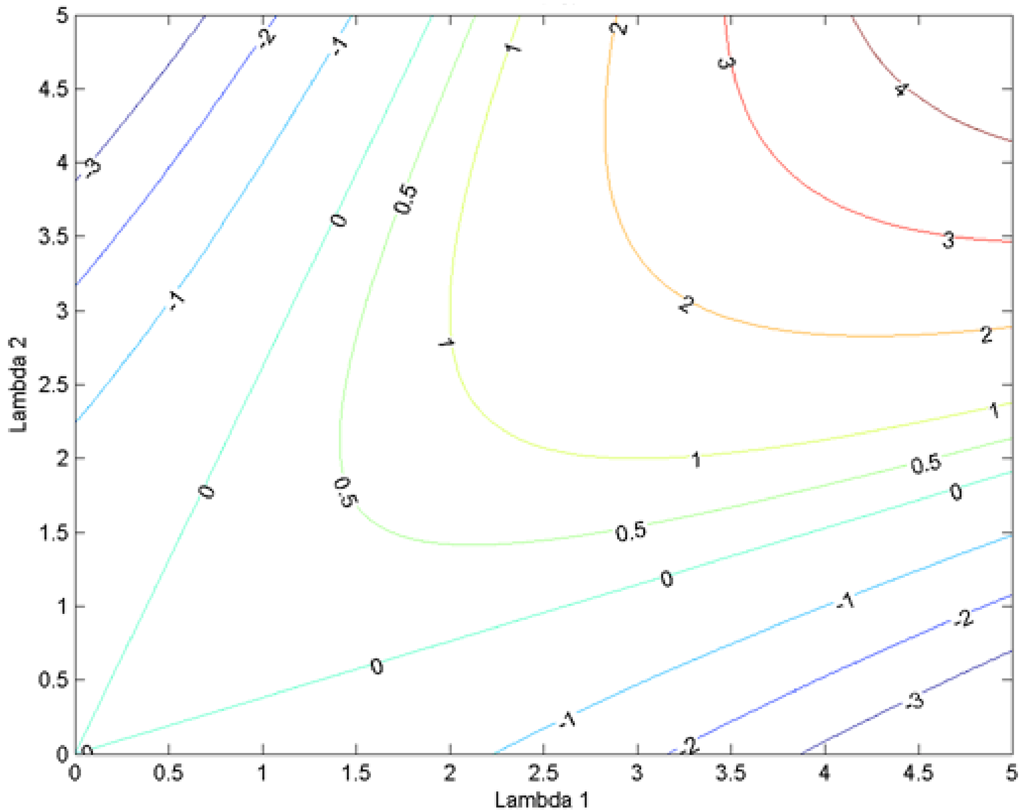
$$tr(\hat{C}) = \lambda_1 + \lambda_2 = \hat{I}_x^2 + \hat{I}_y^2$$

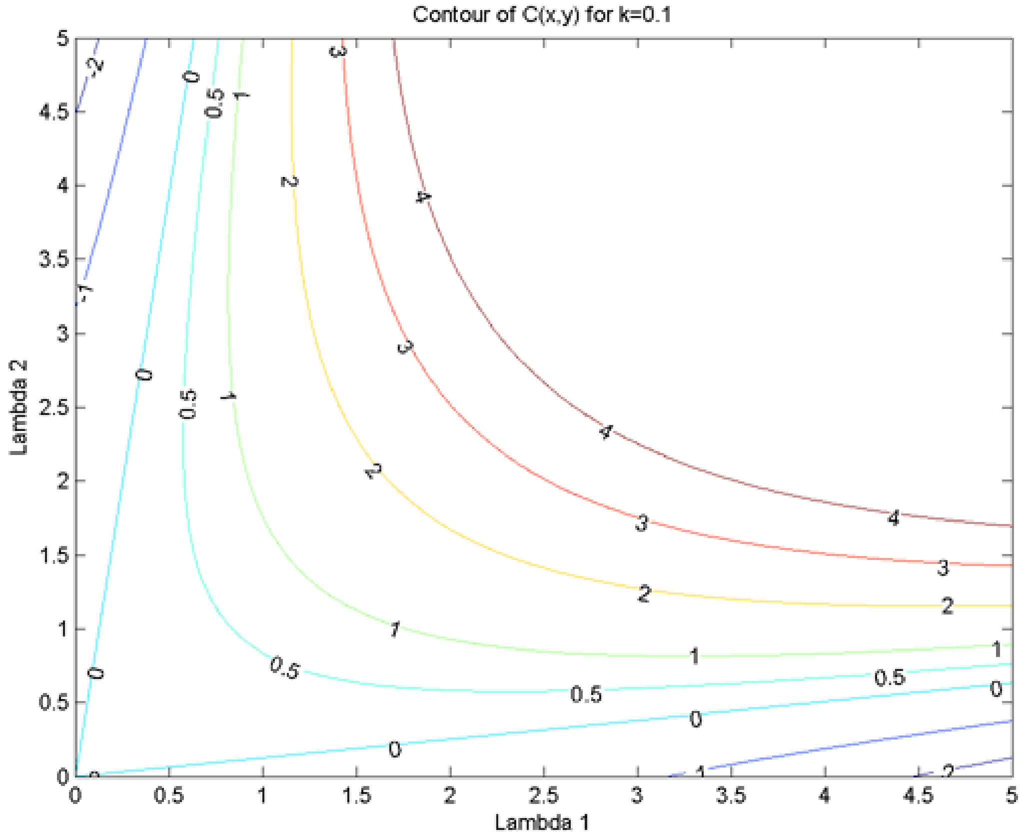
$$\det(\hat{C}) = \lambda_1 \lambda_2 = \hat{I}_x^2 \hat{I}_y^2 - \left(\widehat{I_x I_y} \right)^2$$

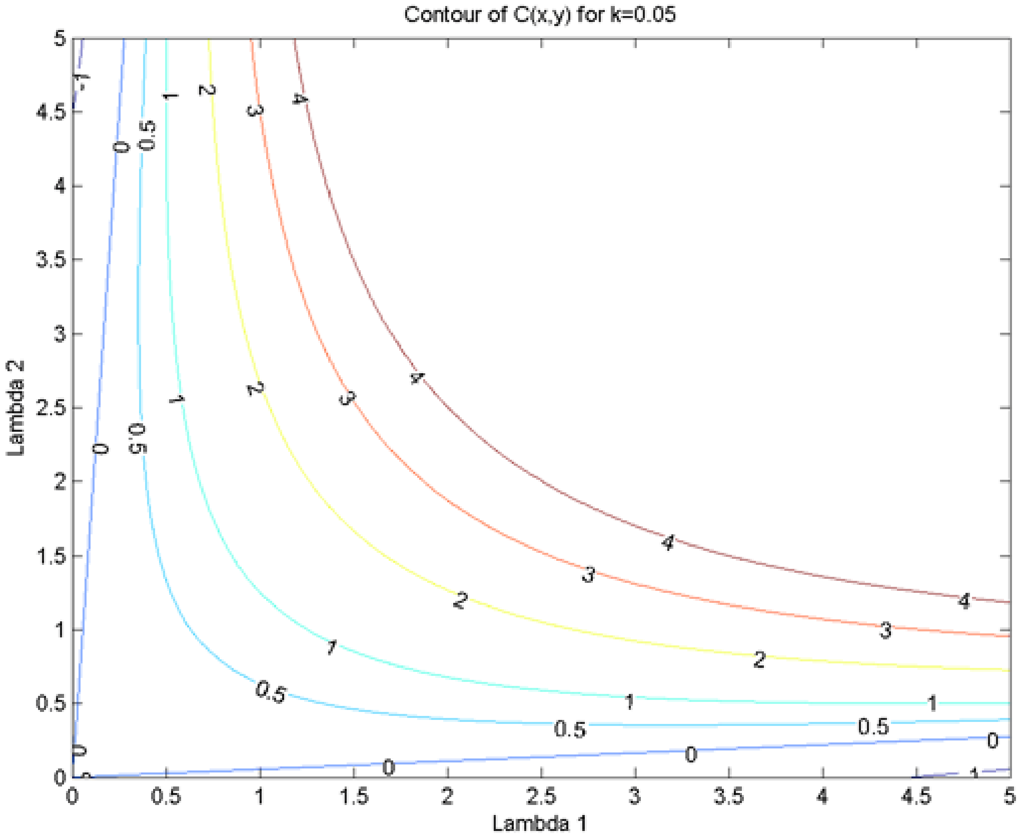
$C = C^t \Rightarrow$ trace and determinant are preserved

usually, k=0.04

R(x,y) for k=0.2



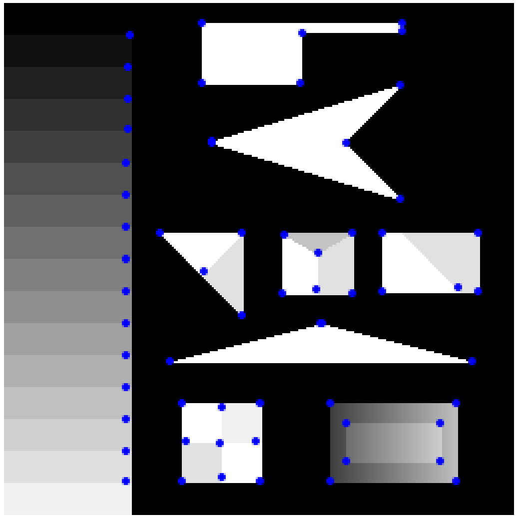

Contour of C(x,y) for k=0.1
Lambda 1
Lambda 2

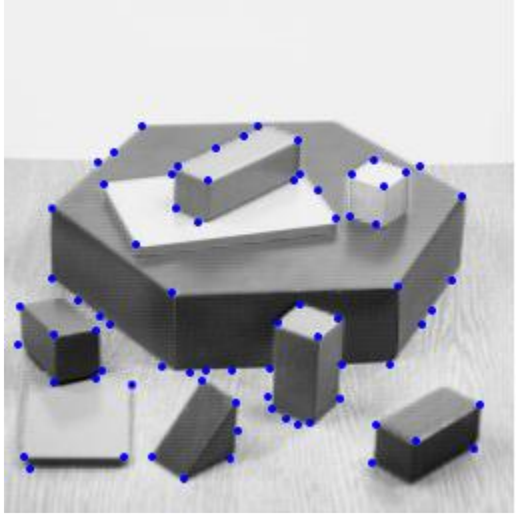


Contour of C(x,y) for k=0.05
Lambda 2
Lambda 1

Source: http://www.cim.mcgill.ca/~dparks/CornerDetector/index.htm

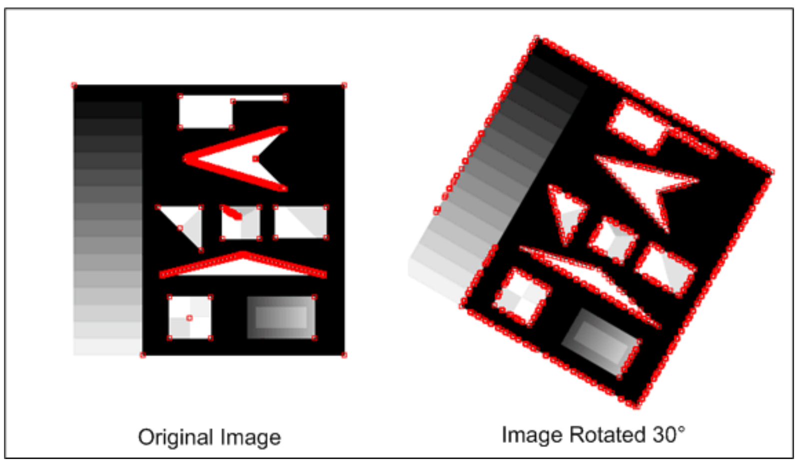
Results

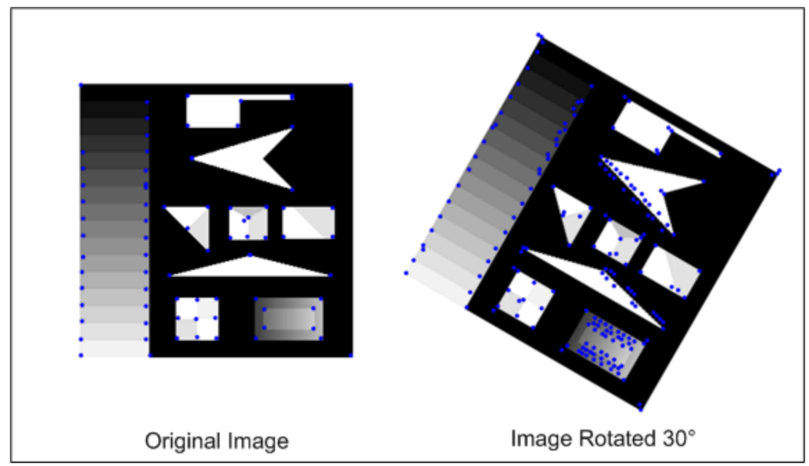




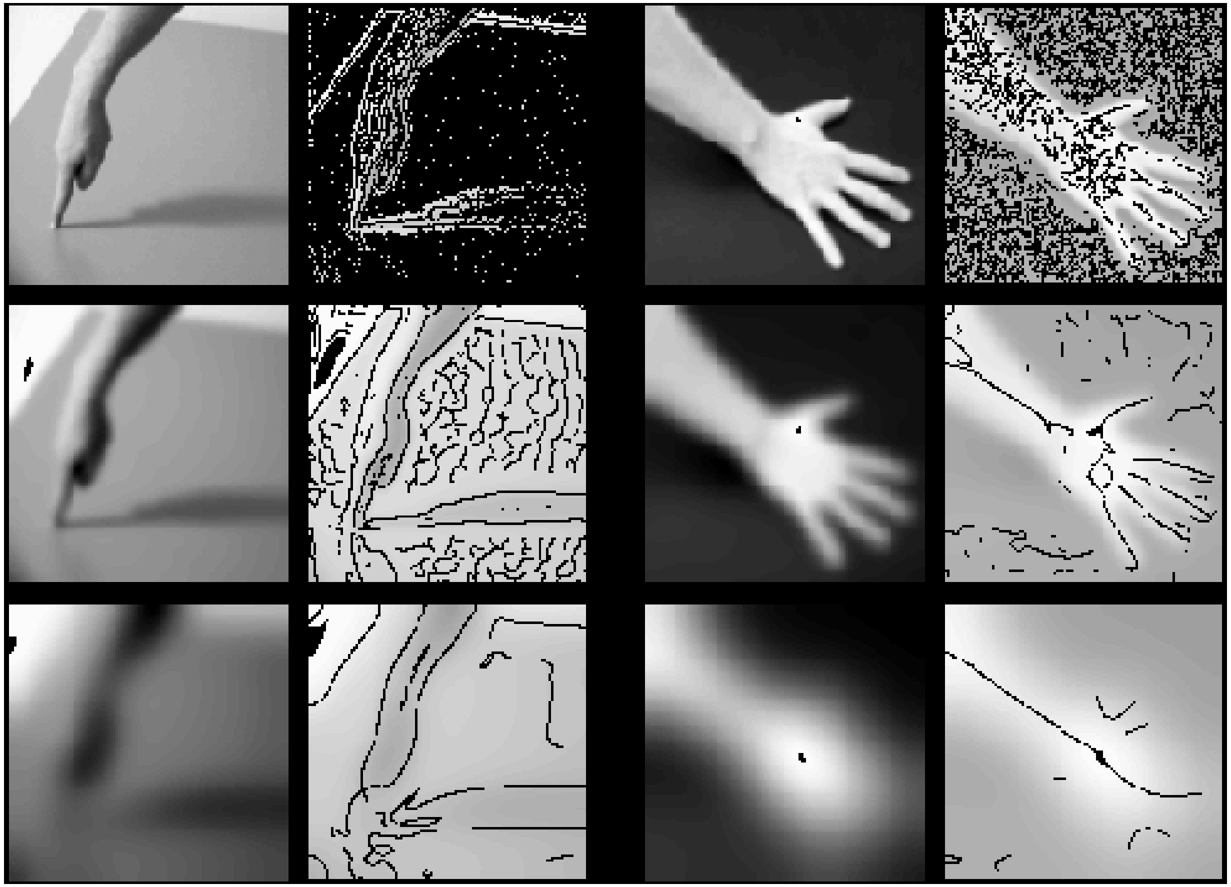
Problems…

- Moravec and Harris have problems with rotations…





The scale problem



Other features

$$L(x, y;\ t) = \int_{(\xi,\eta)\in\mathbb{R}^2} f(x-\xi, y-\eta)\, g(\xi,\eta;\ t)\, d\xi\, d\eta$$

$$g(x, y;\ t) = \frac{1}{2\pi t}\, e^{-(x^2+y^2)/2t}$$

$$L_{x^\alpha y^\beta}(\cdot,\cdot;\ t) = \partial_{x^\alpha y^\beta} L(\cdot,\cdot;\ t) = (\partial_{x^\alpha y^\beta} g(\cdot,\cdot;\ t)) * f(\cdot,\cdot).$$

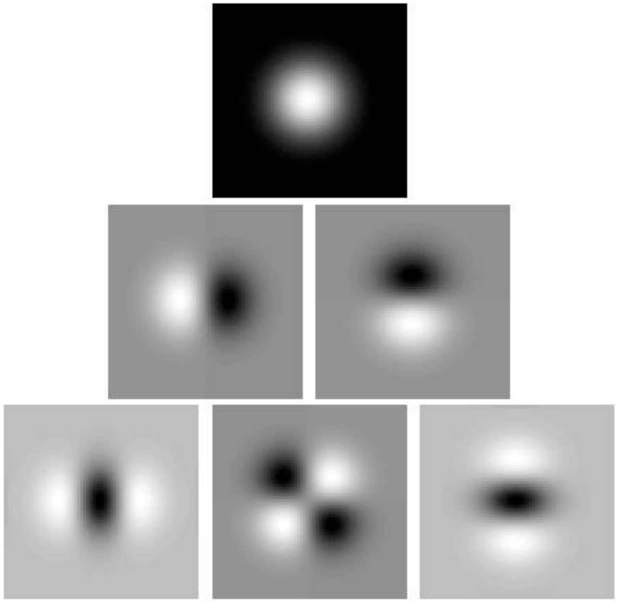


Figure 2: The Gaussian kernel and its derivatives up to order two in the 2-D case.

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix}$$

$$\nabla^2 = tr(H) = L_{xx} + L_{yy}$$

$$\det(H) = L_{xx} \, L_{yy} - L_{xy}{}^2$$

Responds to bolb-like structures, but also to edges

Responds to bolb-like structures, and saddles, insensitive to edges

Figure 4: *Differential descriptors for blob detection/interest point detection:* (left) A grey-level image of size 210×280 pixels. (middle) The Laplacian $\nabla^2 L$ computed at $t = 16$. (right) The determinant of the Hessian $\det \mathcal{H} L$ computed at $t = 16$.

Search across scale

- A common approach is to search for maxima across scales
- Since the amplitude of spatial derivatives decreases with scale, the amplitude of features is expected to decrease
- Define “normalized” features:

$$\nabla^2_{norm} = tr(H_{norm}) = t\left(L_{xx} + L_{yy}\right)$$

$$\det(H_{norm}) = t^2\left(L_{xx}L_{yy} - {L_{xy}}^2\right)$$

Lindeberg, “Feature Detection With Automatic Scale Selection”, Int. Journal of Computer Vision, 1998

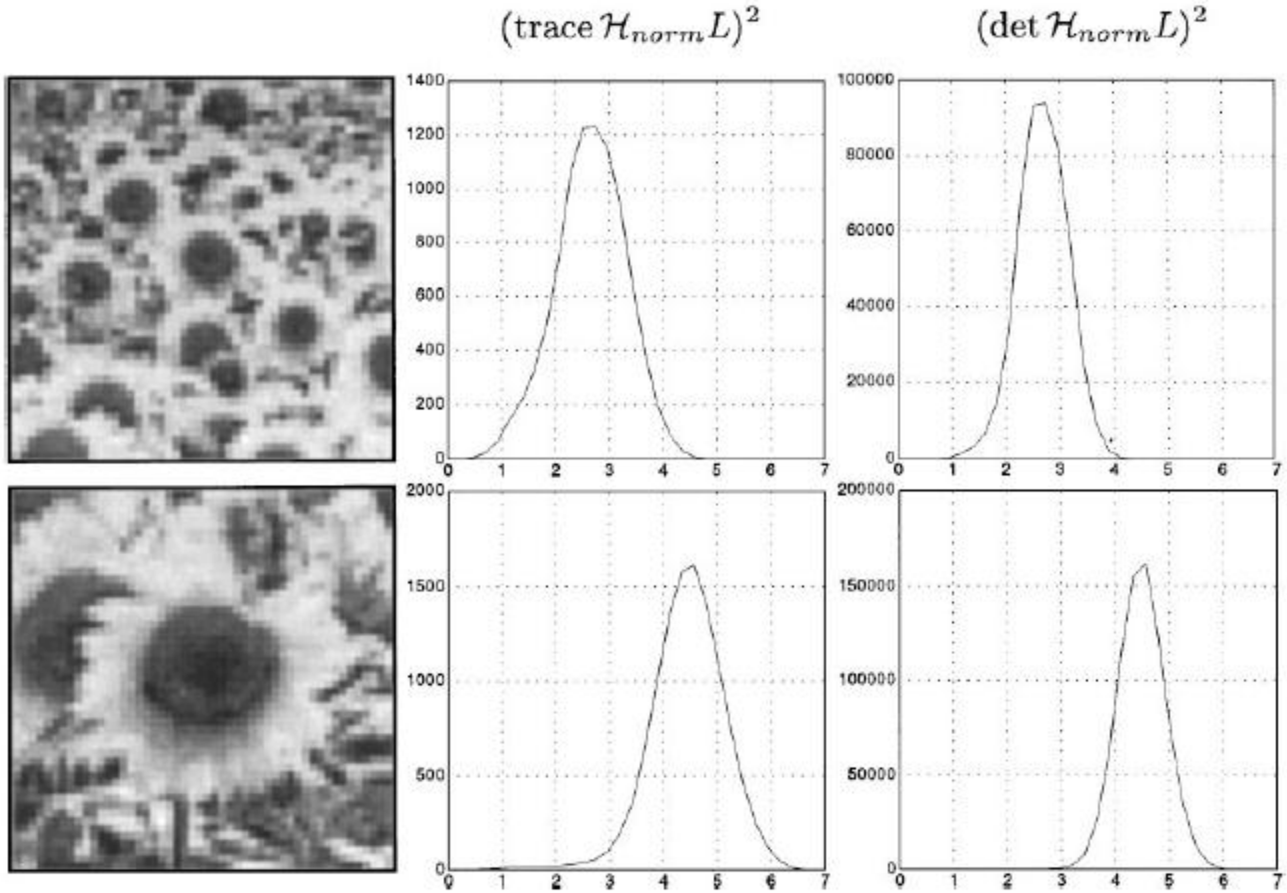


Figure 2. Scale-space signatures of the trace and the determinant of the normalized Hessian matrix computed for two details of a sunflower image; (left) grey-level image, (middle) signature of $(\text{trace}\,\mathcal{H}_{norm}L)^2$, (right) signature of $(\det\mathcal{H}_{norm}L)^2$. (The signatures have been computed at the central point in each image. The horizontal axis shows effective scale, essentially the logarithm of the scale parameter, whereas the scaling of the vertical axis is linear in the normalized operator response.)

$(\mathrm{trace}\,\mathcal{H}_{norm}L)^2$ $(\det\mathcal{H}_{norm}L)^2$

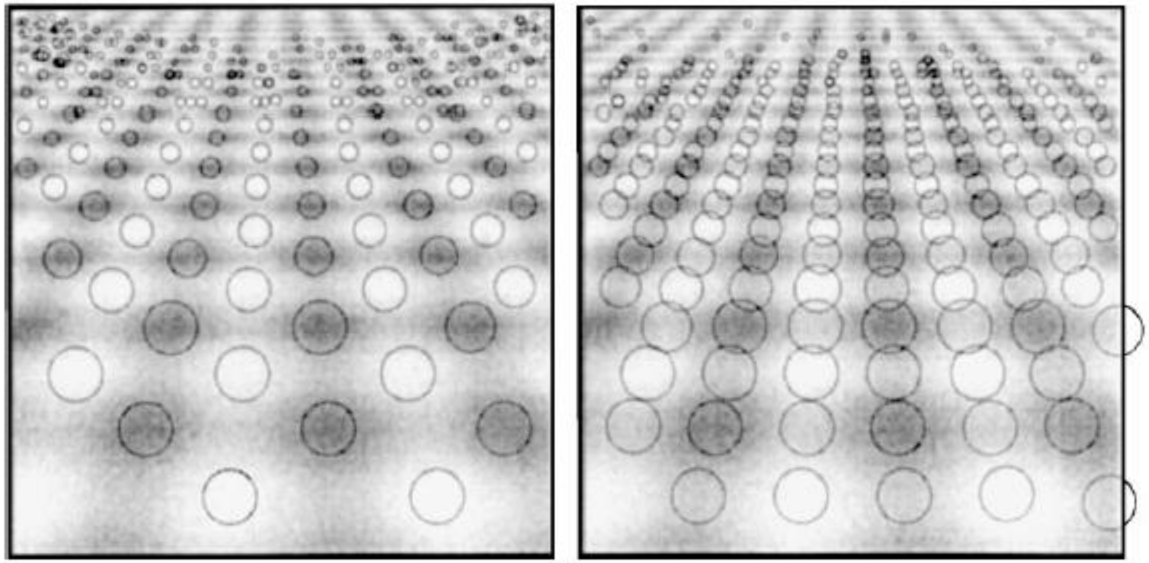
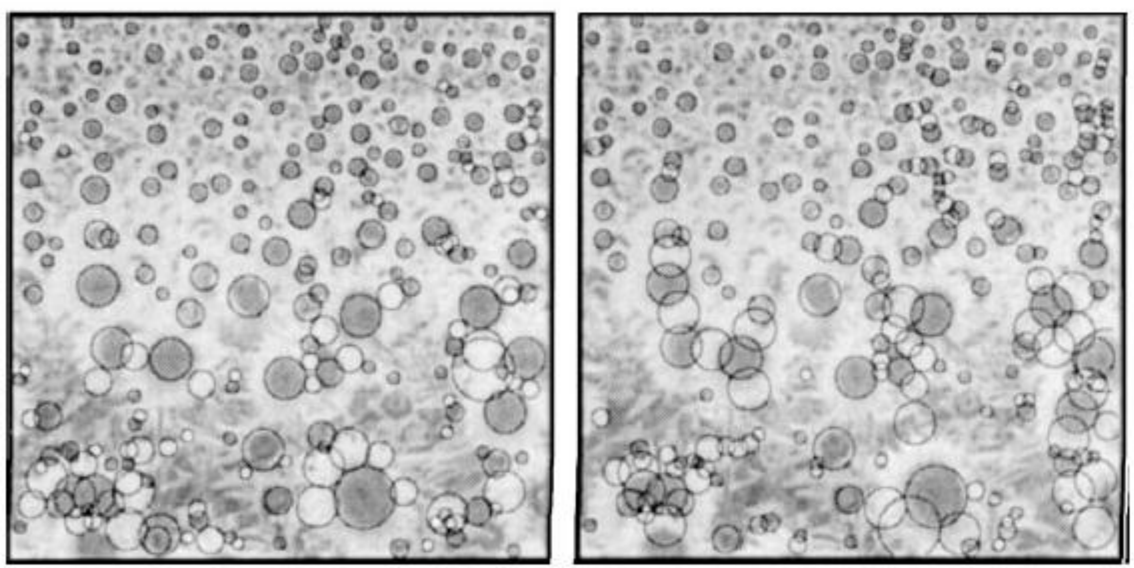


Figure 4. The 250 most significant normalized scale-space extrema detected from the perspective projection of a sine wave (with 10% added Gaussian noise).

$(\mathrm{trace}\,\mathcal{H}_{norm}L)^2$ $(\det\mathcal{H}_{norm}L)^2$



Material from:

- Lecture 10, Advances in Computer Vision, MIT (online).
- Harris, C. and Stephens, M., *A Combined Corner and Edge Detector,* Proceedings of the 4th Alvey Vision Conference, pp. 147–151, 1988
- D.Lowe and Cordelia Schmid, *Recognition and Matching based on local invariant features*, CVPR 2003
- Lowe, D., *Object Recognition from Local Scale-Invariant Features*, Proc. Of the International Conference on Computer Vision, Corfu, September 1999
- Lindeberg, T., *Edge and Ridge Detection with Automatic Scale Selection*, Intl. Journal of Computer Vision 30(2), 117-154, 1998
- *Lindeberg, T., Feature Detection With Automatic Scale Selection, Int. Journal of Computer Vision,* 30(2), 79-116, 1998
- Lindeberg, T., *Scale-space*, In Encyclopedia of Computer Science and Engineering, 2008