“Saliency” – detection and description of “salient” / “interesting” / “relevant” regions

Also: “Key-features”, “points and regions of interest”

• Detectors
  – Regions ✓
  – Edges/lines ✓
  – Salient “points” (corners, blobs) → always detected/described as a point coordinate and the local region surrounding it !!!
  – Affine covariance

• Descriptors
  – Their role for image/object categorization
  – Their role for measurement purposes
Detectors - Examples

Qualitative comparison of examples for detection of

• Regions
• Edges
• Corners

a. Original image (from Graz databases)
b. Graph-based segmentation
   [Felzenszwalb+Huttenlocher]
c. Canny edges [Canny]
d. Harris affine corners [Mikolajczyk+Schmid]
Regions, Edges, Corners – Examples (1)
Regions, Edges, Corners – Examples (2)
Regions, Edges, Corners – Examples (3)
Salient points (corners) based on 1\textsuperscript{st} derivatives

- **Autocorrelation of 2D image signal** [Moravec]
  - Approximation by sum of squared differences (SSD)
  - Window $W$
  - Differences between grayvalues in $W$ and a window shifted by $(\Delta x, \Delta y)$

$$f(x, y) = \sum_{r_w=-\delta}^{\delta} \sum_{c_w=-\delta}^{\delta} [I(x + r_w, y + c_w) - I(x + r_w + \Delta x, y + c_w + \Delta y)]^2$$

$$= \sum_{(x_w, y_w)\in W} [I(x_w, y_w) - I(x_w + \Delta x, y_w + \Delta y)]^2$$

- Four different shift directions $f_i(x,y)$: $f_{\text{Moravec}} = \sum_{i=1}^{4} f_i$

- A corner is detected, when $f_{\text{Moravec}} > th$
Salient points (corners) based on 1\textsuperscript{st} derivatives

• Autocorrelation (second moment) matrix:
  – Avoids various shift directions
  – Approximate \( I(x_w + \Delta x, y_w + \Delta y) \) by Taylor expansion:

\[
I(x_w + \Delta x, y_w + \Delta y) \approx I(x_w, y_w) + \left( I_x(x_w, y_w) \quad I_y(x_w, y_w) \right) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} + \cdot \times
\]

  – Rewrite \( f(x,y) \):

\[
f(x, y) = \sum_{(x_w, y_w) \in W} \left[ \begin{pmatrix} I_x(x_w, y_w) & I_y(x_w, y_w) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right]^2
\]

\[
= \sum_{(x_w, y_w) \in W} (\Delta x \quad \Delta y) \begin{pmatrix} I_x(x_w, y_w) & I_y(x_w, y_w) \end{pmatrix} \begin{pmatrix} I_x(x_w, y_w) & I_y(x_w, y_w) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}
\]

\[
= \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \mathbf{M} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}
\]

“second moment matrix \( \mathbf{M} \)”
Salient points (corners) based on 1\textsuperscript{st} derivatives

- Autocorrelation (\textit{second moment}) matrix:

\[
M = \mu = \begin{pmatrix}
\sum_{(x_w, y_w) \in W} I_x^2(x_w, y_w) & \sum_{(x_w, y_w) \in W} I_x(x_w, y_w)I_y(x_w, y_w) \\
\sum_{(x_w, y_w) \in W} I_x(x_w, y_w)I_y(x_w, y_w) & \sum_{(x_w, y_w) \in W} I_y^2(x_w, y_w)
\end{pmatrix}
\]

- M can be used to derive a measure of “cornerness”
- Independent of various displacements (\(\Delta x, \Delta y\))
- Corner: significant gradients in >1 directions \(\Rightarrow\) rank \(M = 2\)
- Edge: significant gradient in 1 direction \(\Rightarrow\) rank \(M = 1\)
- Homogeneous region \(\Rightarrow\) rank \(M = 0\)

- Several variants of this corner detector:
  - KLT corners, Förstner corners
Salient points (corners) based on 1\textsuperscript{st} derivatives

- **Harris corners**
  - Most popular variant of a detector based on $\mathbf{M}$
  - Local derivatives with “derivation scale” $\sigma_D$
  - Convolution with a Gaussian with “integration scale” $\sigma_I$
  - $\mathbf{M}_{\text{Harris}}$ for each point $\mathbf{x}$ in the image

$$
\mathbf{M}_{\text{Harris}}(\mathbf{x}, \sigma_I, \sigma_D) = \mu(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 G(\sigma_I) \begin{pmatrix}
I_x^2(\mathbf{x}, \sigma_D) & I_x(\mathbf{x}, \sigma_D)I_x(\mathbf{x}, \sigma_D) \\
I_x(\mathbf{x}, \sigma_D)I_x(\mathbf{x}, \sigma_D) & I_y^2(\mathbf{x}, \sigma_D)
\end{pmatrix}
$$

- Cornerness $c_{\text{Harris}}$ does not require to compute eigenvalues

$$
c_{\text{Harris}} = \det \mathbf{M} - \alpha \text{ trace}^2 \mathbf{M}
$$

- Corner detection: $c_{\text{Harris}} > t_{\text{Harris}}$

$\sigma_I = 2, \quad \sigma_D = 0.7$

$\alpha = 0.05, \quad t_{\text{Harris}} = 25000$
Salient points (corners) based on 1st derivatives

- Harris corners
Salient points (corners) based on 2nd derivatives

- **Hessian determinant**

\[
\det \mathbf{H} = \det\begin{pmatrix}
I_{xx} & I_{xy} \\
I_{xy} & I_{yy}
\end{pmatrix} = I_{xx}I_{yy} - I_{xy}^2
\]

- Local maxima of \( \det \mathbf{H} \) [Beaudet]
- Zero crossings of \( \det \mathbf{H} \) [Dreschler+Nagel]
- Detectors are related to *curvature*
- Invariant to rotation
- Similar cornerness measure: local maxima of \( K \) [Kitchen+Rosenfeld]

\[
K = \frac{I_{xx}I_y^2 - 2I_{xy}I_xI_y + I_{yy}I_x^2}{I_x^2 + I_y^2}
\]
Salient points (corners) based on 2\textsuperscript{nd} derivatives

- **DoG / LoG [Marr+Hildreth] ✓**
  - Zero crossings
  - “Mexican hat”, “Sombrero”
  - Edge detector !
  
  \[
  \text{LoG} : \quad L = \Delta (G \ast I) = (\Delta G) \ast I = 0 \\
  \text{DoG} : \quad D = G_1 \ast I - G_2 \ast I = 0
  \]

- **Lowe’s “DoG keypoints” [Lowe]**
  - Edge $\rightarrow$ zero-crossing
  - Blob at corresponding scale: local *extremum* !
  - Low contrast corner suppression: threshold
  - Assess curvature $\rightarrow$ distinguish corners from edges

\[
H_D = \begin{pmatrix}
D_{xx} & D_{xy} \\
D_{xy} & D_{yy}
\end{pmatrix}
\]


- Keypoint detection:

\[
\frac{\text{trace}^2 H_D}{\det H_D} < th
\]
Lowe‘s DoG Keypoints
http://www.cs.ubc.ca/~lowe/keypoints/

SIFT Demo code, 400 x 300 pixel, 358 keypoints found
Salient points (corners) without derivatives

- **Morphological corner detector [Laganière]**
  - 4 structuring elements:
    - +, ◊, X, □
  - Assymetrical closing
Salient points (corners) without derivatives

- SUSAN corners [Smith+Brady]
  - Sliding window
  - Faster than Harris
Salient points (corners) without derivatives

- Kadir/Brady saliency [Kadir+Brady]
  - Histograms
  - Shannon entropy
  - Scale selection
  - Used in constellation model [Fergus et al.]
Salient points (corners) without derivatives

- **MSER** – maximally stable extremal regions [Matas et al.]
  - Successive thresholds
  - Stability: regions “survive” over many thresholds
Affine covariant corner detectors

- Locally planar patch $\rightarrow$ affine distortion

- Detect "characteristic scale"
  - see also [Lindeberg], scale-space

- Recover affine deformation that fits local image data best

Affine covariant corner detectors

elliptical support

normalization "canonical view"

correspondence

[Mikolajczyk+Schmid]
Scaled Harris Corner Detector

• “Harris Laplace” [Mikolajczyk+Schmid, Mikolajczyk et al.]
Scaled Hessian Detector

- “Hessian Laplace” [Mikolajczyk+Schmid , Mikolajczyk et al.]
Harris Affine Detector

• “Harris affine” [Mikolajczyk+Schmid , Mikolajczyk et al.]
Hessian Affine Detector

• “Hessian affine” [Mikolajczyk+Schmid , Mikolajczyk et al.]
Qualitative comparison of detectors (1)

Harris
Harris affine
Harris Laplace

Hessian affine
Hessian Laplace
Qualitative comparison of detectors (2)

Kadir/Brady

morphological

MSER

SUSAN
Corners vs. Blobs (1) [Diss. Markus Brandner, 2009]

Figure 4.15: Corner detection as an example of drawing samples from an ensemble. (a) Image out of a $N = 120$ frames sequence with changing illumination. (b) Enlarged corner with superimposed 99% regions of confidence for different corner detection algorithms. A description and comparison of the different corner detection algorithms is given in [77].
Corners vs. Blobs (2) [Diss. Markus Brandner, 2009]

Figure 4.16: Geometric bias introduced by the mapping of a planar blob feature using a perspective sensor. (a) The centre of the blob and the intersection of the diagonals of the enclosing square intersect. (b) Due to perspective distortion thecentre of gravity and the intersection of the transformed square do not intersect which gives rise to a position bias.
Descriptors (1)

• Representation of salient regions
• “descriptive” features $\rightarrow$ feature vector
• There are many possibilities!

• Categorization vs. specific OR, matching, measurement
  – Sufficient descriptive power vs. extremely high descriptive power
  – Not too much emphasis on specific individuals vs. high specificity
  – Performance is often category-specific vs. noise, chromatic aberration, corners/blobs, etc…

$$f_n = (f_{n,1}, \ldots, f_{n,j})^T$$

feature vector extracted from patch $P_n$
Descriptors (2)

- **Grayvalues**
  - Raw pixel values of a patch $P$
  - “local appearance-based description”
  - “local affine frame” LAF [Obdržálek+Matas] for MSER

- **General moments of order $p+q$:** $m_{pq} = \sum_{x\in P} \sum_{y\in P} x^p y^q I(x, y)$

- **Moment invariants:**
  - Central moments $\mu_{pq}$: invariant to translation

\[
\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}: \quad \mu_{pq} = \sum_{x\in P} \sum_{y\in P} (x - \bar{x})^p (y - \bar{y})^q I(x, y)
\]
Descriptors (3)

- **Moment invariants:**
  - Normalized central moments
    \[ \eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \text{with} \quad \gamma = 1 + \frac{p + q}{2} \]
  - Translation, rotation, scale invariant moments \( \Phi_1 \) ... \( \Phi_7 \) [Hu]
    \[
    \Phi_1 = \eta_{20} + \eta_{02} \\
    \Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2
    \]
  - Geometric/photometric, color invariants [vanGool et al.]

- **Filters**
  - “local jets” [Koenderink+VanDoorn]
  - Gabor banks, steerable filters, discrete cosine transform DCT
Descriptors (4)

- **SIFT descriptors [Lowe]**
  - Scale invariant feature transform
  - Calculated for local patch $P$: 8x8 or 16x16 pixels
  - Subdivision into 4x4 sample regions
  - Weighted histogram of 8 gradient directions: $0^\circ$, $45^\circ$, …
  - SIFT vector dimension: 128 for a 16x16 patch

- **Similar recent descriptors:**
  - E.g. Histogram of Oriented gradients HOG [Dalal+Triggs]
The Role of Detectors / Descriptors for Categorization

• **Descriptors are required**
  – Building blocks for object representation
  – They should be sufficiently descriptive
  – They should not be too specific (intra-class variability!)

• **Detectors may have benefits, but also deficiencies!**
  – Good to learn category models under strong supervision (e.g. constellation)
  – Poor under weak supervision (e.g. caught in background clutter)

• **There are good results without detectors**
  – Random selection of points
  – Fixed / dense grid
Bag of Keypoints

- No model – “model-free”, “geometry-free”
- Keypoints / random points / dense grid
- Descriptors
- Discriminative model (category vs. counterexamples)
- Example: Graz02 bikes → bag of 100 bike keypoints

Graz02: bike #229

\[ f_n = (f_{n,1}, \cdots, f_{n,j})^T \]

\[ \text{e.g. moment invariants} \]
Bag of 100 Keypoints Example

- many training images
- many descriptors
- 100 weak classifiers
- learned by Boosting

#243
The Role of Detectors / Descriptors for Measurement

• Camera Calibration, Stereo Reconstruction, Tracking, …

• Detectors are *required*
  – As *precise* as possible (center coordinate, subpixel accuracy, …)
  – As *reproducible* as possible (varying illumination, exposure, camera pose, noise, …)

• Descriptors are *required*
  – Establish *point correspondences*
  – They should be *extremely* descriptive
  – But they should be robust against noise, illumination, viewpoint, exposure, …
Enabling methods – S+M ("SLAM")

[Schweighofer 2008]
Point Correspondences - Example S+M

Structure and motion from “natural” landmarks [Schweighofer]

→ Stereo reconstruction of Harris corners
Noteworthy Issues with Detectors/Descriptors

- **Evaluation of detectors/descriptors**
  - Repeatability
  - Descriptivity (extremely high for tracking vs. some generalization for object recognition)
  - Computational efficiency (e.g. comp.complexity, avoiding floating point arithmetic, etc.)
  - Storage requirements

- **Recently published new algorithms**
  - Detectors: FAST – Features from Accelerated Segment Test [ECCV’06]
    SURF – Speeded Up Robust Features [CVIU’08]
  - Descriptors: ORB – Oriented FAST and Rotated BRIEF [ICCV’11]
    BRISK – Binary Robust Invariant Scalable Keypoint [ICCV’11]
    FREAK – Fast Retina Keypoint [CVPR’12]