Instance-level recognition:
Local invariant features

Cordelia Schmid
INRIA, Grenoble
Overview

• Introduction to local features

• Harris interest points + SSD, ZNCC, SIFT

• Scale & affine invariant interest point detectors

• Evaluation and comparison of different detectors

• Region descriptors and their performance
Local features

Several / many local descriptors per image
Robust to occlusion/clutter + no object segmentation required

Photometric : distinctive
Invariant : to image transformations + illumination changes
Local features: interest points
Local features: Contours/segments
Local features: segmentation
Application: Matching

Find corresponding locations in the image
Illustration – Matching

Interest points extracted with Harris detector (~ 500 points)
Interest points matched based on cross-correlation (188 pairs)
Global constraint - Robust estimation of the fundamental matrix

99 inliers

89 outliers
Application: Panorama stitching

Images courtesy of A. Zisserman.
Application: Instance-level recognition

Search for particular objects and scenes in large databases
Finding the object despite possibly large changes in scale, viewpoint, lighting and partial occlusion

→ requires invariant description
Difficulties

- Very large images collection → need for efficient indexing
  - Flickr has 2 billion photographs, more than 1 million added daily
  - Facebook has 15 billion images (~27 million added daily)
  - Large personal collections
  - Video collections, i.e., YouTube
Instance-level recognition: Approach

- Image content is transformed into local features invariant to geometric and photometric transformations
- Matching local invariant descriptors
Applications

Search photos on the web for particular places

Find these landmarks ...in these images and 1M more
Applications

• Take a picture of a product or advertisement
  → find relevant information on the web

PRENEZ EN PHOTO L'AFFICHE!

Accédez à la bande annonce, à tous les horaires et à la réservation.

Avec la participation de

UGC

TOUTLECINE.COM

[Pixee – Milpix]
Applications

• Finding stolen/missing objects in a large collection
Applications

- Copy detection for images and videos

Query video

Search in 200h of video
Applications

- Sony Aibo – Robotics
  - Recognize docking station
  - Communicate with visual cards
  - Place recognition
  - Loop closure in SLAM
Local features

1) Extraction of local features
   - Contours/segments
   - Interest points & regions
   - Regions by segmentation
   - Dense features, points on a regular grid

2) Description of local features
   - Dependant on the feature type
   - Contours/segments $\rightarrow$ angles, length ratios
   - Interest points $\rightarrow$ greylevels, gradient histograms
   - Regions (segmentation) $\rightarrow$ texture + color distributions
Line matching

- Extraction de contours
  - Zero crossing of Laplacian
  - Local maxima of gradients

- Chain contour points (hysteresis)

- Extraction of line segments

- Description of segments
  - Mi-point, length, orientation, angle between pairs etc.
Experimental results – line segments

images 600 x 600
Experimental results – line segments

248 / 212 line segments extracted
Experimental results – line segments

89 matched line segments - 100% correct
Experimental results – line segments

3D reconstruction
Problems of line segments

• Often only partial extraction
  – Line segments broken into parts
  – Missing parts

• Information not very discriminative
  – 1D information
  – Similar for many segments

• Potential solutions
  – Pairs and triplets of segments
  – Interest points
Overview

• Introduction to local features

• **Harris interest points + SSD, ZNCC, SIFT**

• Scale & affine invariant interest point detectors

• Evaluation and comparison of different detectors

• Region descriptors and their performance
Harris detector [Harris & Stephens’88]

Based on the idea of auto-correlation

Important difference in all directions => interest point
Harris detector

Auto-correlation function for a point \((x, y)\) and a shift \((\Delta x, \Delta y)\)

\[
A(x, y) = \sum_{(x_k, y_k) \in W(x, y)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2
\]
Harris detector

Auto-correlation function for a point \((x, y)\) and a shift \((\Delta x, \Delta y)\)

\[
A(x, y) = \sum_{(x_k, y_k) \in W(x, y)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2
\]

\(W\)

\(A(x, y)\) \{
small in all directions \quad \rightarrow \quad uniform region
large in one directions \quad \rightarrow \quad contour
large in all directions \quad \rightarrow \quad interest point
\}
Harris detector

“flat” region: no change in all directions

“edge”: no change along the edge direction

“corner”: significant change in all directions
Harris detector

Discret shifts are avoided based on the auto-correlation matrix

with first order approximation

\[ I(x_k + \Delta x, y_k + \Delta y) = I(x_k, y_k) + (I_x(x_k, y_k) \quad I_y(x_k, y_k)) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \]

\[ A(x, y) = \sum_{(x_k, y_k) \in W(x, y)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2 \]

\[ = \sum_{(x_k, y_k) \in W} \left( \begin{pmatrix} I_x(x_k, y_k) \\ I_y(x_k, y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2 \]
Harris detector

\[
\begin{align*}
\Delta x & \Delta y \\
= & \begin{pmatrix}
\sum_{(x_k, y_k) \in W} (I_x(x_k, y_k))^2 & \sum_{(x_k, y_k) \in W} I_x(x_k, y_k)I_y(x_k, y_k) \\
\sum_{(x_k, y_k) \in W} I_x(x_k, y_k)I_y(x_k, y_k) & \sum_{(x_k, y_k) \in W} (I_y(x_k, y_k))^2
\end{pmatrix}
\begin{pmatrix}
\Delta x \\
\Delta y
\end{pmatrix}
\end{align*}
\]

Auto-correlation matrix

the sum can be smoothed with a Gaussian

\[
\begin{align*}
\Delta x & \Delta y \\
= & \begin{pmatrix}
(I_x^2 & I_xI_y \\
I_xI_y & I_y^2
\end{pmatrix}
\begin{pmatrix}
\Delta x \\
\Delta y
\end{pmatrix}
\]
Harris detector

- Auto-correlation matrix

\[ A(x, y) = G \bigotimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \]

- captures the structure of the local neighborhood
- measure based on eigenvalues of this matrix
  - 2 strong eigenvalues => interest point
  - 1 strong eigenvalue => contour
  - 0 eigenvalue => uniform region
Interpreting the eigenvalues

Classification of image points using eigenvalues of autocorrelation matrix:

- \( \lambda_1 \) and \( \lambda_2 \) are small;
- \( \lambda_1 \) and \( \lambda_2 \) are large, \( \lambda_1 \sim \lambda_2 \);
- \( \lambda_1 \gg \lambda_2 \);
Corner response function

\[ R = \text{det}(A) - \alpha \text{trace}(A)^2 = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2 \]

\( \alpha: \) constant (0.04 to 0.06)
Harris detector

• Cornerness function

\[ f = \det(A) - k(\text{trace}(A))^2 = \lambda_1 \lambda_2 - k (\lambda_1 + \lambda_2)^2 \]

Reduces the effect of a strong contour

• Interest point detection
  – Threshold (absolute, relative, number of corners)
  – Local maxima

\[ f > \text{thresh} \land \forall x, y \in 8-neighbourhood \quad f(x, y) \geq f(x', y') \]
Harris Detector: Steps
Harris Detector: Steps

Compute corner response $R$
Harris Detector: Steps

Find points with large corner response: $R > \text{threshold}$
Harris Detector: Steps

Take only the points of local maxima of $R$
Harris Detector: Steps
Harris detector: Summary of steps

1. Compute Gaussian derivatives at each pixel
2. Compute second moment matrix $A$ in a Gaussian window around each pixel
3. Compute corner response function $R$
4. Threshold $R$
5. Find local maxima of response function (non-maximum suppression)
Harris - invariance to transformations

• Geometric transformations
  – translation
  – rotation
  – similitude (rotation + scale change)
  – affine (valid for local planar objects)

• Photometric transformations
  – Affine intensity changes ($l \rightarrow a \cdot l + b$)
Harris Detector: Invariance Properties

- Rotation

Ellipse rotates but its shape (i.e. eigenvalues) remains the same

Corner response $R$ is invariant to image rotation
Harris Detector: Invariance Properties

- Affine intensity change
  - Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$
  - Intensity scale: $I \rightarrow a \cdot I$

*Partially invariant* to affine intensity change, dependent on type of threshold
Harris Detector: Invariance Properties

- Scaling

Corner

All points will be classified as edges

Not invariant to scaling
Comparison of patches - SSD

Comparison of the intensities in the neighborhood of two interest points

SSD : sum of square difference

$$\frac{1}{(2N+1)^2} \sum_{i=-N}^{N} \sum_{j=-N}^{N} (I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j))^2$$

Small difference values $\Rightarrow$ similar patches
Comparison of patches

SSD : \( \frac{1}{(2N+1)^2} \sum_{i=-N}^{N} \sum_{j=-N}^{N} (I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j))^2 \)

Invariance to photometric transformations?

Intensity changes (I \( \rightarrow \) I + b)

\( \Rightarrow \) Normalizing with the mean of each patch

\( \frac{1}{(2N+1)^2} \sum_{i=-N}^{N} \sum_{j=-N}^{N} ((I_1(x_1 + i, y_1 + j) - m_1) - (I_2(x_2 + i, y_2 + j) - m_2))^2 \)

Intensity changes (I \( \rightarrow \) aI + b)

\( \Rightarrow \) Normalizing with the mean and standard deviation of each patch

\( \frac{1}{(2N+1)^2} \sum_{i=-N}^{N} \sum_{j=-N}^{N} \left( \frac{I_1(x_1 + i, y_1 + j) - m_1}{\sigma_1} - \frac{I_2(x_2 + i, y_2 + j) - m_2}{\sigma_2} \right)^2 \)
Cross-correlation ZNCC

zero normalized SSD

\[
\frac{1}{(2N+1)^2} \sum_{i=-N}^{N} \sum_{j=-N}^{N} \left( \frac{I_1(x_1+i, y_1+j)-m_1}{\sigma_1} - \frac{I_2(x_2+i, y_2+j)-m_2}{\sigma_2} \right)^2
\]

ZNCC: zero normalized cross correlation

\[
\frac{1}{(2N+1)^2} \sum_{i=-N}^{N} \sum_{j=-N}^{N} \left( \frac{I_1(x_1+i, y_1+j)-m_1}{\sigma_1} \right) \left( \frac{I_2(x_2+i, y_2+j)-m_2}{\sigma_2} \right)
\]

ZNCC values between -1 and 1, 1 when identical patches in practice threshold around 0.5
Introduction to local descriptors

• Greyvalue derivatives

• Differential invariants [Koenderink’87]

• SIFT descriptor [Lowe’99]
Greyvalue derivatives: Image gradient

- The gradient of an image:  
  \[ \nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \]

- The gradient points in the direction of most rapid increase in intensity

- The gradient direction is given by  
  \[ \theta = \tan^{-1} \left( \frac{\partial f}{\partial y} / \frac{\partial f}{\partial x} \right) \]
  – how does this relate to the direction of the edge?

- The edge strength is given by the gradient magnitude  
  \[ \| \nabla f \| = \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2} \]
Differentiation and convolution

- Recall, for 2D function, f(x,y):
  \[
  \frac{\partial f}{\partial x} = \lim_{\varepsilon \to 0} \left( \frac{f(x + \varepsilon, y) - f(x, y)}{\varepsilon} \right)
  \]

- We could approximate this as
  \[
  \frac{\partial f}{\partial x} \approx \frac{f(x_{n+1}, y) - f(x_n, y)}{\Delta x}
  \]

- Convolution with the filter
  \[
  \begin{bmatrix}
  -1 & 1
  \end{bmatrix}
  \]
Finite difference filters

- Other approximations of derivative filters exist:

- **Prewitt:**
  \[
  M_x = \begin{bmatrix}
  -1 & 0 & 1 \\
  -1 & 0 & 1 \\
  -1 & 0 & 1 \\
  \end{bmatrix} ;
  M_y = \begin{bmatrix}
  1 & 1 & 1 \\
  0 & 0 & 0 \\
  -1 & -1 & -1 \\
  \end{bmatrix}
  \]

- **Sobel:**
  \[
  M_x = \begin{bmatrix}
  -1 & 0 & 1 \\
  -2 & 0 & 2 \\
  -1 & 0 & 1 \\
  \end{bmatrix} ;
  M_y = \begin{bmatrix}
  1 & 2 & 1 \\
  0 & 0 & 0 \\
  -1 & -2 & -1 \\
  \end{bmatrix}
  \]

- **Roberts:**
  \[
  M_x = \begin{bmatrix}
  0 & 1 \\
  -1 & 0 \\
  \end{bmatrix} ;
  M_y = \begin{bmatrix}
  1 & 0 \\
  0 & -1 \\
  \end{bmatrix}
  \]
Effects of noise

- Consider a single row or column of the image
  - Plotting intensity as a function of position gives a signal

\[ f(x) \]

\[ \frac{d}{dx} f(x)_0 \]

- Where is the edge?
Solution: smooth first

- To find edges, look for peaks in \( \frac{d}{dx}(f \ast g) \)
Differentiation is convolution, and convolution is associative:

\[
\frac{d}{dx}(f * g) = f * \frac{d}{dx}g
\]

This saves us one operation:
Local descriptors

- Greyvalue derivatives
  - Convolution with Gaussian derivatives

\[
v(x, y) = \begin{pmatrix}
I(x, y) * G(\sigma) \\
I(x, y) * G_x(\sigma) \\
I(x, y) * G_y(\sigma) \\
I(x, y) * G_{xx}(\sigma) \\
I(x, y) * G_{xy}(\sigma) \\
I(x, y) * G_{yy}(\sigma) \\
\vdots
\end{pmatrix}
\]

\[
I(x, y) * G(\sigma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(x', y', \sigma) I(x - x', y - y') dx' dy'
\]

\[
G(x, y, \sigma) = \frac{1}{2\pi \sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)
\]
Local descriptors

Notation for greyvalue derivatives [Koenderink’87]

\[
\mathbf{v}(x, y) = \begin{pmatrix}
I(x, y) * G(\sigma) \\
I(x, y) * G_x(\sigma) \\
I(x, y) * G_y(\sigma) \\
I(x, y) * G_{xx}(\sigma) \\
I(x, y) * G_{xy}(\sigma) \\
I(x, y) * G_{yy}(\sigma)
\end{pmatrix}
= \begin{pmatrix}
L(x, y) \\
L_x(x, y) \\
L_y(x, y) \\
L_{xx}(x, y) \\
L_{xy}(x, y) \\
L_{yy}(x, y)
\end{pmatrix}
\]

Invariance?
Local descriptors – rotation invariance

Invariance to image rotation: differential invariants [Koen87]

**gradient magnitude**

**Laplacian**

\[
\begin{bmatrix}
L \\
L_x L_x + L_y L_y \\
L_{xx} L_x L_x + 2L_{xy} L_x L_y + L_{yy} L_y L_y \\
L_{xx} L_{xx} + 2L_{xy} L_{xy} + L_{yy} L_{yy}
\end{bmatrix}
\]

\[\ldots\]

\[\ldots\]
Laplacian of Gaussian (LOG)

\[ LOG = G_{xx}(\sigma) + G_{yy}(\sigma) \]
SIFT descriptor [Lowe’99]

• Approach
  – 8 orientations of the gradient
  – 4x4 spatial grid
  – Dimension 128
  – soft-assignment to spatial bins
  – normalization of the descriptor to norm one
  – comparison with Euclidean distance
Local descriptors - rotation invariance

- Estimation of the dominant orientation
  - extract gradient orientation
  - histogram over gradient orientation
  - peak in this histogram

- Rotate patch in dominant direction
Local descriptors – illumination change

• Robustness to illumination changes

  in case of an affine transformation \( I_1(x) = aI_2(x) + b \)
Local descriptors – illumination change

• Robustness to illumination changes

in case of an affine transformation \( I_1(x) = aI_2(x) + b \)

• Normalization of derivatives with gradient magnitude

\[
\frac{(L_{xx} + L_{yy})}{\sqrt{L_xL_x + L_yL_y}}
\]
Local descriptors – illumination change

- Robustness to illumination changes in case of an affine transformation \( I_1(\mathbf{x}) = aI_2(\mathbf{x}) + b \)

- Normalization of derivatives with gradient magnitude

\[
(L_{xx} + L_{yy}) \sqrt{L_x L_x + L_y L_y}
\]

- Normalization of the image patch with mean and variance
Invariance to scale changes

• Scale change between two images

• Scale factor $s$ can be eliminated

• Support region for calculation!!
  – In case of a convolution with Gaussian derivatives defined by $\sigma$

\[
I(x, y) \ast G(\sigma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(x', y', \sigma)I(x - x', y - y')dx'dy'
\]

\[
G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)
\]