Interest Points and Scale Invariance
Example: Finding Correspondences Between Images

- First step toward 3-D reconstruction: Find correspondences between *feature points* in two images of a scene
- Object recognition: Find correspondences between *feature points* in “training” and “test” image
2,106 images, 819,242 points

http://grail.cs.washington.edu/projects/rome/

Building Rome in a Day
Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski
International Conference on Computer Vision, 2009, Kyoto, Japan.
• Intuitively, junctions of contours.
• Generally more stable features over changes of viewpoint
• Intuitively, large variations in the neighborhood of the point in all directions
The distribution of the $x$ and $y$ derivatives is very different for all three types of patches.
The distribution of $x$ and $y$ derivatives can be characterized by the shape and size of the principal component ellipse.

- **Corner**: $R = 28.07$
- **Flat**: $R = 0.25$
- **Linear Edge**: $R = 0.3328$
How to evaluate “interestness”?

• The distribution of gradients in a neighborhood $W$ is represented by the inertia (shape, Harris) matrix:

$$H = \begin{bmatrix}
\sum_{w} I_x^2 & \sum_{w} I_x I_y \\
\sum_{w} I_x I_y & \sum_{w} I_y^2
\end{bmatrix}$$

• Elongations of the distribution = Eigenvalues of $H$: $\lambda_{\text{min}}, \lambda_{\text{max}}$
  – We want $\lambda_{\text{min}}, \lambda_{\text{max}}$ to be approx. equal
  – We want $\lambda_{\text{min}}, \lambda_{\text{max}}$ to be large
Harris detector and its variants

- We want $\lambda_{\text{min}}, \lambda_{\text{max}}$ to be approx. equal
- We want $\lambda_{\text{min}}, \lambda_{\text{max}}$ to be large

$$R = 4 \frac{\lambda_{\text{min}} \lambda_{\text{max}}}{\lambda_{\text{min}} + \lambda_{\text{max}}}$$

• ($R = 1$ if $\lambda_{\text{min}} = \lambda_{\text{max}}$ but keep only the ones with large $\lambda_{\text{max}}$)

$$R = \frac{\lambda_{\text{min}} \lambda_{\text{max}}}{\lambda_{\text{min}} + \lambda_{\text{max}}}$$

$$R = \lambda_{\text{min}} \lambda_{\text{max}} - k \lambda_{\text{min}} + \lambda_{\text{max}}$$
Comp. efficient definition

- For any symmetric matrix $H$:
  - $\text{Det}(H) = \lambda_{\min}\lambda_{\max}$
  - $\text{Trace}(H) = \lambda_{\min} + \lambda_{\max}$ (The trace is the sum of the diagonal of $H$)

$$R = 4 \frac{\text{Det}(H)}{\text{Trace}(H)^2}$$

$$R = \frac{\text{Det}(H)}{\text{Trace}(H)}$$

$$R = \text{Det}(H) - k\text{Trace}(H)^2$$
Computation of $H$ and gradients

- $I_x$ means convolution with Gaussian of $\sigma$
- $H$ should be computed with different weights $\rightarrow$ Convolution with Gaussian

\[
H = \begin{bmatrix}
\sum_w I_x^2 & \sum_w I_x I_y \\
\sum_w I_x I_y & \sum_w I_y^2
\end{bmatrix}
\]

Constant weights

- $G_\sigma * I^2$
- $G_\sigma * I_x I_y$

Gaussian weights

\[
H = \begin{bmatrix}
G_\sigma * I_x^2 & G_\sigma * I_x I_y \\
G_\sigma * I_x I_y & G_\sigma * I_y^2
\end{bmatrix}
= G_\sigma * \begin{bmatrix}
I_x^2 & I_x I_y \\
I_x I_y & I_y^2
\end{bmatrix}
\]
Input

Response of interest point operator
Input

Response of interest point operator

Interest point detection example
\[ \det M - k \left( \text{trace } M \right)^2 \]
Interest points example
Example (from Montessinos et al.)
1. Compute $x$ and $y$ derivatives of image

$$I_x = G^x_\sigma * I \quad I_y = G^y_\sigma * I$$

2. Compute products of derivatives at every pixel

$$I_{x2} = I_x . I_x \quad I_{y2} = I_y . I_y \quad I_{xy} = I_x . I_y$$

3. Compute the sums of the products of derivatives at each pixel

$$S_{x2} = G_{\sigma I} * I_{x2} \quad S_{y2} = G_{\sigma I} * I_{y2} \quad S_{xy} = G_{\sigma I} * I_{xy}$$

4. Define at each pixel $(x, y)$ the matrix

$$H(x, y) = \begin{bmatrix} S_{x2}(x, y) & S_{xy}(x, y) \\ S_{xy}(x, y) & S_{y2}(x, y) \end{bmatrix}$$

5. Compute the response of the detector at each pixel

$$R = Det(H) - k(Trace(H))^2$$
Interpretation

• Geometry:
  – Interest point = places where the image surface has high curvature (sharp elliptic points).

• (Small) motion analysis:
  – Interest point = If we move the image by a small amount $dx \ dy$, the image value changes a lot, no matter what direction $(dx,dy)$ is used.
Curvature is defined as the rate of variation of the orientation of tangent vector along the curve:

\[ k = \frac{d\theta}{ds} \]

Geometrically, the curvature is the inverse of the radius of the locally best fitting circle:

\[ k = \frac{1}{R} \]
At any point \( p \) on a surface \( z = f(x,y) \), we define:
- The tangent plane.
- The surface normal \( N \) at \( p \): The vector normal to the tangent plane.

Consider the family of planes that contain the line \((p,N)\). This family is obtained by rotating a plane about the line \((p,N)\) by an angle \( \theta \).

- A plane at orientation \( \theta \) intersects the surface along a curve of curvature \( k(\theta) \).
- As \( \theta \) varies \( k(\theta) \) reaches two extreme values \( k_{\text{min}} \) and \( k_{\text{max}} \), the principal curvatures of the surface at \( p \).

- The tangents to the two curves, \( t_{\text{min}} \) and \( t_{\text{max}} \), are the principal directions of the surface at \( p \):
  - \( t_{\text{min}} \) and \( t_{\text{max}} \) are orthogonal to each other.
  - Intuitively \( t_{\text{min}} \) (resp. \( t_{\text{max}} \)) is the direction in which the normal to the surface varies the fastest (resp. the slowest).
  - \( t_{\text{min}} \) and \( t_{\text{max}} \) are the eigenvectors of a matrix \( M \) computed from the derivatives of \( f \).
  - \( k_{\text{min}} \) and \( k_{\text{max}} \) are the corresponding eigenvalues.
$I(x, y)$

$\nabla I(x, y)$

$k_{\text{min}}$

$k_{\text{max}}$

$N$
$k_{min}$ small (parabolic point)

$k_{min}$ and $k_{max}$ small

$k_{min}$ and $k_{max}$ large and similar magnitudes
→ Good selection of interest point (elliptic point)
Interest points are points at which both principal curvatures are large and are of similar magnitudes.

Principal curvatures approximated by the eigenvalues of the moment matrix $H$.

The shape matrix $H$ is used to approximate $M$. 
Scale Selection

- We need to decide what window size to use for computing the Harris matrix.
- Equivalently, we need to choose the value of $\sigma'$.
- The window size (or $\sigma'$) must be consistent between different magnifications of the image.
How to choose a neighborhood size? \( \rightarrow \) Intuition

Best radius: Local extrema of function that measures amount of interesting stuff
\[ \sigma_1 = 2 \times \sigma_2 \]
• Why did we use the normalized Laplacian in the previous example?
• Justification (and basis for most scale selection operations in computer vision):
  • **Scale Selection Principle (T. Lindeberg):**

In the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data.

What are normalized derivatives?  Example using 2\textsuperscript{nd} order derivatives

\[ \sigma^{n+m} \frac{\partial^{n+m} f}{\partial x^n \partial y^m} \]

\[ \sigma^2 \nabla^2 f = \sigma^2 \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) \]
The Laplacian of a Gaussian can be approximated by the difference of two Gaussian:

\[ \nabla^2_{\sigma} I = \nabla^2 G_{\sigma} * I \]

To compare the Laplacian at different scales we need to explain more carefully what the approximation is.

In practice: the scaled Laplacian can be computed by taking the difference between level in a Gaussian pyramid.

\[ G_{k\sigma} - G_{\sigma} \approx (k - 1)\sigma^2 \nabla^2 G_{\sigma} \]
(From Mikolajczyk and Schmid'02)
Laplacian
Combining Harris and Laplace: Harris-Laplace detector

\[ H = G_{\sigma_I} \ast \begin{bmatrix} I_x^2 & I_x I_y, \sigma_D \\ I_x I_y, \sigma_D & I_y^2 \end{bmatrix} \]

\[ R = \text{Det}(H) - k(\text{Trace}(H))^2 \]

- The location \( x \) of the points detected by Harris is not scale invariant \( \rightarrow \) Depends on the choice of \( \sigma_I \) and \( \sigma_D \).

- Reduce to one parameter: \( \sigma_D = s \sigma_I \) \( (s = 0.7) \)

- The Laplacian trick gives us a good \( \sigma \) but not where the interest point is.

- Chicken and egg problem:
  - If we knew \( x \) we could estimate \( \sigma \)
  - If we knew \( \sigma \) we could find \( x \)
Harris-Laplace: Algorithm summary

1. Try Harris at different scales and report initial points and associated scales
   - Try different $\sigma_i$ of the form $k^n\sigma_o$ ($k = 1.4$)
   - Report points with large $R$

2. For each detected point $x$
   - Estimate characteristic scale $\sigma_c$ as maximum of
     $\sigma^2\nabla^2 G_\sigma$
   - Find the point $x'$ with the maximum of $R$ in a 8x8 neighborhood of $x$ by using the new scale $\sigma_i = \sigma_c$
   - Replace $x$ by $x'$

Iterate until convergence
Example from Mikolajczyk and Schmid 2004
Affine-invariant detection (overview only)

- Need to define a richer description of “neighborhood” or scale
- Use directional derivatives instead: $\sigma_I$ replaced by $\Sigma_I$, $\sigma_D$ replaced by $\Sigma_D$
- $\Sigma_I$ represent an elliptical “neighborhood” instead of a circular one
- More degrees of freedom to search through but conceptually similar algorithm:
  - Assume $\Sigma_D = s \Sigma_I$
  - Find $x$’s with initial selections of $\Sigma_I$
  - Iterate:
    - Re-estimate “scale” $\Sigma_I$
    - Adjust the location of $x$ based on new $\Sigma_I$
State of the Art: Affine Invariance

Example from Mikolajczyk and Schmid 2004
Affine Invariance
Application: Finding Correspondences
Initial detections
Scale: 4.9
Rotation: 19°

Example from Mikolajczyk and Schmid 2004
Final matches: 32 correct correspondences
Scale: 4.9
Rotation: 19°

Example from Mikolajczyk and Schmid 2004
Application: Finding Correspondences

Scale change: 1.7
Viewpoint change: 50°

Example from Mikolajczyk and Schmid 2004
Descriptors

- We have scale or affine invariant interest points with region around each of them
- Descriptor = vector describing the image content in the neighborhood around the interest point
- Classical descriptor (128-vector): SIFT (Scale Invariant Feature Transform)
Descriptors

• First transform neighborhood to canonical square window, for example: 16x16 window
SIFT Descriptor

- Image gradients are sampled over a 4x4k array of locations around interest points.
- Create an array of orientation histograms over 8 orientations in each of k² 4x4 blocks.
- 8 orientations x k x k histogram array N = 8 k² dimensions.
- In practice k = 4 N = 128.

SIFT Descriptor

\[ k = 2 \]

\[ N = 32 \]
Descriptor is not rotationally invariant $\rightarrow$ Select a dominant direction and express all the gradient orientations with respect to the dominant direction

Example from Lowe2004
60° rotation for planar objects

Example from Lowe2004
Example from Lowe2004
References


Software can be downloaded from Schmid’s and Lowe’s pages


Other descriptors/detectors: MSER

- MSER = Maximally stable extremal regions
- Objective find stable (= interest) regions

Extremal region $R_i$: All the intensity values inside $R_i$ are greater (or lower) than those at the boundary of $R_i$.
Other descriptors/detectors: MSER

- MSER = Maximally stable extremal regions
- Objective find stable (= interest) regions
From MSER to SIFT

Find extremal region
Fit ellipse
Eigenthings of:
Rotate + scale to canonical patch
Compute 128-dim SIFT vector

\[
\sum_{\text{region}} (P - \bar{P})(P - \bar{P})^T = 0
\]

\[
M = \begin{bmatrix}
\sum_{\text{region}} (x - \bar{x})^2 & \sum_{\text{region}} (x - \bar{x})(y - \bar{y}) \\
\sum_{\text{region}} (x - \bar{x})(y - \bar{y}) & \sum_{\text{region}} (y - \bar{y})^2
\end{bmatrix}
\]
Examples from image matching

Example from Matas et al. BMCV 2002
Examples from image matching

Example from Matas et al. BMCV 2002
References


- **MSER:**

- **Even more detectors/descriptors:**
  - **SURF:** Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, 2008.