SIFT keypoint detection

D. Lowe, Distinctive image features from scale-invariant keypoints, IJCV 60 (2), pp. 91-110, 2004
Keypoint detection with scale selection

- We want to extract keypoints with characteristic scales that are covariant w.r.t. the image transformation
Basic idea

• Convolve the image with a “blob filter” at multiple scales and look for extrema of filter response in the resulting scale space

T. Lindeberg, Feature detection with automatic scale selection, IJCV 30(2), pp 77-116, 1998
Blob detection

Find maxima and minima of blob filter response in space and scale

Source: N. Snavely
Blob filter

Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D

$$\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$$
Recall: Edge detection

Edge = maximum of derivative

Source: S. Seitz
Edge detection, Take 2

Edge = zero crossing of second derivative

Source: S. Seitz
From edges to blobs

- Edge = ripple
- Blob = superposition of two ripples

**Spatial selection**: the magnitude of the Laplacian response will achieve a maximum at the center of the blob, provided the scale of the Laplacian is “matched” to the scale of the blob.
Scale selection

- We want to find the characteristic scale of the blob by convolving it with Laplacians at several scales and looking for the maximum response.
- However, Laplacian response decays as scale increases:

![Graph showing the unnormalized Laplacian response for different values of σ.](image)
Scale normalization

• The response of a derivative of Gaussian filter to a perfect step edge decreases as $\sigma$ increases:

\[ \frac{1}{\sigma \sqrt{2\pi}} \]

• To keep response the same (scale-invariant), must multiply Gaussian derivative by $\sigma$

• Laplacian is the second Gaussian derivative, so it must be multiplied by $\sigma^2$
Effect of scale normalization

Original signal

Unnormalized Laplacian response

Scale-normalized Laplacian response

maximum
Blob detection in 2D

- **Scale-normalized** Laplacian of Gaussian:

\[
\nabla^2_{\text{norm}} g = \sigma^2 \left( \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right)
\]
Blob detection in 2D

- At what scale does the Laplacian achieve a maximum response to a binary circle of radius $r$?
Blob detection in 2D

- At what scale does the Laplacian achieve a maximum response to a binary circle of radius $r$?
- To get maximum response, the zeros of the Laplacian have to be aligned with the circle.
- The Laplacian is given by (up to scale):
  \[ \left( x^2 + y^2 - 2\sigma^2 \right) e^{-\left( x^2 + y^2 \right)/2\sigma^2} \]
- Therefore, the maximum response occurs at $\sigma = r / \sqrt{2}$. 

![Diagram of a binary circle and the Laplacian response](image)
Scale-space blob detector

1. Convolve image with scale-normalized Laplacian at several scales
Scale-space blob detector: Example
Scale-space blob detector: Example
Scale-space blob detector

1. Convolve image with scale-normalized Laplacian at several scales
2. Find maxima of squared Laplacian response in scale-space
Scale-space blob detector: Example
Efficient implementation

- Approximating the Laplacian with a difference of Gaussians:

\[ L = \sigma^2 \left( G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma) \right) \]

\((\text{Laplacian})\)

\[ \text{DoG} = G(x, y, k\sigma) - G(x, y, \sigma) \]

\((\text{Difference of Gaussians})\)
Efficient implementation

Eliminating edge responses

- Laplacian has strong response along edges
Eliminating edge responses

- Laplacian has strong response along edges

- Solution: filter based on Harris response function over neighborhoods containing the “blobs”
From feature detection to feature description

- To recognize the same pattern in multiple images, we need to match appearance "signatures" in the neighborhoods of extracted keypoints
  - But corresponding neighborhoods can be related by a scale change or rotation
  - We want to normalize neighborhoods to make signatures invariant to these transformations
Finding a reference orientation

• Create histogram of local gradient directions in the patch
• Assign reference orientation at peak of smoothed histogram
SIFT features

- Detected features with characteristic scales and orientations:

From keypoint detection to feature description

Detection is \textit{covariant}:

\[ \text{features} (\text{transform}(\text{image})) = \text{transform} (\text{features}(\text{image})) \]

Description is \textit{invariant}:

\[ \text{features} (\text{transform}(\text{image})) = \text{features}(\text{image}) \]
SIFT descriptors

• Inspiration: complex neurons in the primary visual cortex

D. Lowe, Distinctive image features from scale-invariant keypoints, *IJCV* 60 (2), pp. 91-110, 2004
Properties of SIFT

Extraordinarily robust detection and description technique

- Can handle changes in viewpoint
  - Up to about 60 degree out-of-plane rotation
- Can handle significant changes in illumination
  - Sometimes even day vs. night
- Fast and efficient—can run in real time
- Lots of code available

Source: N. Snavely
A hard keypoint matching problem

NASA Mars Rover images
Answer below (look for tiny colored squares…)
What about 3D rotations?
What about 3D rotations?

- Affine transformation approximates viewpoint changes for roughly planar objects and roughly orthographic cameras
Affine adaptation

Consider the second moment matrix of the window containing the blob:

\[
M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = R^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R
\]

Recall:

\[
[u \ v] \ M \begin{bmatrix} u \\ v \end{bmatrix} = \text{const}
\]

This ellipse visualizes the “characteristic shape” of the window
Affine adaptation

K. Mikolajczyk and C. Schmid, Scale and affine invariant interest point detectors, IJCV 60(1):63-86, 2004
Keypoint detectors/descriptors for recognition: A retrospective

Figure 5: Samples of the ten texture classes used in the experiments of Section 3.2.

Keypoint detectors/descriptors for recognition: A retrospective

Keypoint detectors/descriptors for recognition: A retrospective

Keypoint detectors/descriptors for recognition: A retrospective

Keypoint detectors/descriptors for recognition: A retrospective

Amusing how some computer vision researchers jokingly refer to work done before 2012 as "prehistoric".