Lecture 6: Finding Features (part 1/2)

Professor Fei-Fei Li
Stanford Vision Lab
What we will learn today?

• Local invariant features
  – Motivation
  – Requirements, invariances

• Keypoint localization
  – Harris corner detector

• Scale invariant region selection
  – Automatic scale selection
  – Difference-of-Gaussian (DoG) detector

• SIFT: an image region descriptor
What we will learn today?

• Local invariant features
  – Motivation
  – Requirements, invariances
• Keypoint localization
  – Harris corner detector

Some background reading:
Rick Szeliski, Chapter 14.1.1; David Lowe, IJCV 2004
Image matching: a challenging problem
Image matching: a challenging problem

by Diva Sian

by swashford
Harder Case

by Diva Sian

by scgbt

Slide credit: Steve Seitz
Harder Still?

NASA Mars Rover images

Slide credit: Steve Seitz
Answer Below  (Look for tiny colored squares)

NASA Mars Rover images with SIFT feature matches
(Figure by Noah Snavely)

Fei-Fei Li  Lecture 6 -  8  2-Oct-14
Motivation for using local features

- Global representations have major limitations
- Instead, describe and match only local regions
- Increased robustness to
  - Occlusions
  - Articulation
  - Intra-category variations
General Approach

1. Find a set of distinctive keypoints

2. Define a region around each keypoint

3. Extract and normalize the region content

4. Compute a local descriptor from the normalized region

5. Match local descriptors

\[ d(f_A, f_B) < T \]
Common Requirements

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

No chance to match!

We need a repeatable detector!
Common Requirements

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

• Problem 2:
  – For each point correctly recognize the corresponding one

We need a reliable and distinctive descriptor!
Invariance: Geometric Transformations

Slide credit: Steve Seitz
Levels of Geometric Invariance
Invariance: Photometric Transformations

- Often modeled as a linear transformation:
  - Scaling + Offset
Requirements

- Region extraction needs to be repeatable and accurate
  - Invariant to translation, rotation, scale changes
  - Robust or covariant to out-of-plane (≈affine) transformations
  - Robust to lighting variations, noise, blur, quantization

- Locality: Features are local, therefore robust to occlusion and clutter.

- Quantity: We need a sufficient number of regions to cover the object.

- Distinctiveness: The regions should contain “interesting” structure.

- Efficiency: Close to real-time performance.
Many Existing Detectors Available

- Hessian & Harris [Beaudet ‘78], [Harris ‘88]
- Laplacian, DoG [Lindeberg ‘98], [Lowe ‘99]
- Harris-/Hessian-Laplace [Mikolajczyk & Schmid ‘01]
- Harris-/Hessian-Affine [Mikolajczyk & Schmid ‘04]
- EBR and IBR [Tuytelaars & Van Gool ‘04]
- MSER [Matas ‘02]
- Salient Regions [Kadir & Brady ‘01]
- Others...

- Those detectors have become a basic building block for many recent applications in Computer Vision.
Keypoint Localization

- Goals:
  - Repeatable detection
  - Precise localization
  - Interesting content

⇒ Look for two-dimensional signal changes
Finding Corners

• Key property:
  – In the region around a corner, image gradient has two or more dominant directions
• Corners are *repeatable* and *distinctive*

Corners as Distinctive Interest Points

- Design criteria
  - We should easily recognize the point by looking through a small window (locality)
  - Shifting the window in any direction should give a large change in intensity (good localization)

- "flat" region: no change in all directions
- "edge": no change along the edge direction
- "corner": significant change in all directions
Harris Detector Formulation

• Change of intensity for the shift \([u,v]\):

\[
E(u, v) = \sum_{x, y} w(x, y) \left[ I(x+u, y+v) - I(x, y) \right]^2
\]

Window function
Shifted intensity
Intensity

Window function \(w(x,y) = \)

1 in window, 0 outside

or

Gaussian

Slide credit: Rick Szeliski
Harris Detector Formulation

- This measure of change can be approximated by:

\[ E(u, v) \approx [u \ v] \ M \begin{bmatrix} u \\ v \end{bmatrix} \]

where \( M \) is a 2x2 matrix computed from image derivatives:

\[ M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \]

- Sum over image region – the area we are checking for corner

\[ M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \ I_y] \]
Harris Detector Formulation

where $M$ is a 2x2 matrix computed from image derivatives:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Gradient with respect to $x$, times gradient with respect to $y$

Sum over image region – the area we are checking for corner

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \ I_y]$$
What Does This Matrix Reveal?

• First, let’s consider an axis-aligned corner:

\[ M = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \]

- Dominant gradient directions align with \( x \) or \( y \) axis.
- If either \( \lambda \) is close to 0, then this is not a corner, so look for locations where both are large.

What if we have a corner that is not aligned with the image axes?
What Does This Matrix Reveal?

• First, let’s consider an axis-aligned corner:

\[
M = \begin{bmatrix}
\sum I_x^2 & \sum I_x I_y \\
\sum I_x I_y & \sum I_y^2
\end{bmatrix} = \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix}
\]

• This means:
  – Dominant gradient directions align with \( x \) or \( y \) axis
  – If either \( \lambda \) is close to 0, then this is not a corner, so look for locations where both are large.

• What if we have a corner that is not aligned with the image axes?
General Case

- Since $M$ is symmetric, we have $M = R^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R$

  (Eigenvalue decomposition)

- We can visualize $M$ as an ellipse with axis lengths determined by the eigenvalues and orientation determined by $R$

adapted from Darya Frolova, Denis Simakov
Interpreting the Eigenvalues

- Classification of image points using eigenvalues of $M$:

- $\lambda_1$ and $\lambda_2$ are small; $E$ is almost constant in all directions

- $\lambda_1$ and $\lambda_2$ are large, $\lambda_1 \sim \lambda_2$; $E$ increases in all directions

- $\lambda_1 \gg \lambda_2$ — “Edge” region

- $\lambda_2 \gg \lambda_1$ — “Corner” region

Slide credit: Kristen Grauman
Corner Response Function

\[ \theta = \det(M) - \alpha \operatorname{trace}(M)^2 = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2 \]

- Fast approximation
  - Avoid computing the eigenvalues
  - \( \alpha \): constant (0.04 to 0.06)
Window Function $w(x, y)$

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- **Option 1: uniform window**
  - Sum over square window
    $$M = \sum_{x,y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$
  - Problem: not rotation invariant

- **Option 2: Smooth with Gaussian**
  - Gaussian already performs weighted sum
    $$M = g(\sigma) * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$
  - Result is rotation invariant
Summary: Harris Detector [Harris88]

- Compute second moment matrix (autocorrelation matrix)

\[
M(\sigma_I, \sigma_D) = g(\sigma_I)^* \begin{bmatrix}
I_x^2(\sigma_D) & I_xI_y(\sigma_D) \\
I_xI_y(\sigma_D) & I_y^2(\sigma_D)
\end{bmatrix}
\]

1. Image derivatives
2. Square of derivatives
3. Gaussian filter \( g(\sigma_I) \)
4. Cornerness function - two strong eigenvalues
\[
\theta = \det[M(\sigma_I, \sigma_D)] - \alpha[\text{trace}(M(\sigma_I, \sigma_D))]^2
\]
\[
= g(I_x^2)g(I_y^2) - [g(I_xI_y)]^2 - \alpha[g(I_x^2) + g(I_y^2)]^2
\]
5. Perform non-maximum suppression
Harris Detector: Workflow

Slide adapted from Darya Frolova, Denis Simakov
Harris Detector: Workflow
- computer corner responses $\theta$
Harris Detector: Workflow
- Take only the local maxima of $\theta$, where $\theta > \text{threshold}$
Harris Detector: Workflow
- Resulting Harris points
Harris Detector – Responses [Harris88]

**Effect:** A very precise corner detector.
Harris Detector – Responses [Harris88]
• Results are well suited for finding stereo correspondences
Harris Detector: Properties

• Translation invariance?
Harris Detector: Properties

• Translation invariance
• Rotation invariance?

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

Corner response $\theta$ is invariant to image rotation
Harris Detector: Properties

- Translation invariance
- Rotation invariance
- Scale invariance?

Not invariant to image scale!

All points will be classified as edges!
What we have learned today?

• Local invariant features
  – Motivation
  – Requirements, invariances

• Keypoint localization
  – Harris corner detector

• Scale invariant region selection
  – Automatic scale selection
  – Difference-of-Gaussian (DoG) detector

• SIFT: an image region descriptor

Some background reading:
Rick Szeliski, Chapter 14.1.1; David Lowe, IJCV 2004