Local features: detection and description

Monday March 7
Prof. Kristen Grauman
UT-Austin

Midterm Wed.

- Covers material up until 3/1
- Solutions to practice exam handed out today
- Bring a 8.5"x11" sheet of notes if you want
- Review the outlines and notes on course website, accompanying reading in textbook

Last time

• Image warping based on homography
• Detecting corner-like points in an image

Today

• Local invariant features
  - Detection of interest points
    • (Harris corner detection)
    • Scale invariant blob detection: LoG
  - Description of local patches
    • SIFT: Histograms of oriented gradients

Local features: main components

1) Detection: Identify the interest points

\[ x_1 = [x_1^{(1)}, \ldots, x_1^{(n)}] \]

2) Description: Extract vector feature descriptor surrounding each interest point.

\[ x_2 = [x_2^{(1)}, \ldots, x_2^{(m)}] \]

3) Matching: Determine correspondence between descriptors in two views

Goal: interest operator repeatability

• We want to detect (at least some of) the same points in both images.

No chance to find true matches!

• Yet we have to be able to run the detection procedure independently per image.
Goal: descriptor distinctiveness

- We want to be able to reliably determine which point goes with which.
- Must provide some invariance to geometric and photometric differences between the two views.

Local features: main components

1) Detection: Identify the interest points
2) Description: Extract vector feature descriptor surrounding each interest point.
3) Matching: Determine correspondence between descriptors in two views

Recall: Corners as distinctive interest points

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

2 x 2 matrix of image derivatives (averaged in neighborhood of a point).

Notation:
\[ I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y} \quad I_x I_y = \frac{\partial I_x}{\partial x} \frac{\partial I_y}{\partial y} \]

Recall: Corners as distinctive interest points

Since \( M \) is symmetric, we have
\[ M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T \]

\[ Mx = \lambda_x x \]

The eigenvalues of \( M \) reveal the amount of intensity change in the two principal orthogonal gradient directions in the window.

Recall: Corners as distinctive interest points

"edge": \( \lambda_1 \gg \lambda_2 \quad \lambda_2 \gg \lambda_1 \)

"corner": \( \lambda_1 \) and \( \lambda_2 \) are large, \( \lambda_1 \approx \lambda_2 \)

"flat" region \( \lambda_1 \) and \( \lambda_2 \) are small.

One way to score the cornerness:
\[ f = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \]

Harris corner detector

1) Compute \( M \) matrix for image window surrounding each pixel to get its cornerness score.
2) Find points with large corner response \( (f > \text{threshold}) \)
3) Take the points of local maxima, i.e., perform non-maximum suppression
Harris Detector: Steps

Compute corner response $f$

Properties of the Harris corner detector
Rotation invariant? Yes

$$M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$$

Scale invariant?
Properties of the Harris corner detector

- Rotation invariant? Yes
- Scale invariant? No

All points will be classified as edges

Corner!

Scale invariant interest points

How can we independently select interest points in each image, such that the detections are repeatable across different scales?

Automatic scale selection

Intuition:
- Find scale that gives local maxima of some function $f$ in both position and scale.

What can be the “signature” function?

Recall: Edge detection

- Edge = maximum of derivative

Recall: Edge detection

- Edge = zero crossing of second derivative
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.

Blob detection in 2D

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

\[ \nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \]

Blob detection in 2D: scale selection

We define the characteristic scale as the scale that produces peak of Laplacian response.

Example

Original image at ¾ the size
Scale invariant interest points

Interest points are local maxima in both position and scale.

Squared filter response maps

→ List of \((x, y, \sigma)\)
Scale-space blob detector: Example

We can approximate the Laplacian with a difference of Gaussians; more efficient to implement.

$$L = \sigma^2 \left( G_\sigma(x, y, \sigma) + G_{-\sigma}(x, y, \sigma) \right)$$

(Difference of Gaussians)

Technical detail

Local features: main components

1) Detection: Identify the interest points

2) Description: Extract vector feature descriptor surrounding each interest point.

$$x_1 = [x_1^{(1)}, \ldots, x_d^{(1)}]$$

$$x_2 = [x_1^{(2)}, \ldots, x_d^{(2)}]$$

3) Matching: Determine correspondence between descriptors in two views

Geometric transformations

e.g. scale, translation, rotation

Photometric transformations

The simplest way to describe the neighborhood around an interest point is to write down the list of intensities to form a feature vector.

But this is very sensitive to even small shifts, rotations.

Raw patches as local descriptors

Figure from T. Tuytelaars ECCV 2006 tutorial
SIFT descriptor [Lowe 2004]
- Use histograms to bin pixels within sub-patches according to their orientation.

Why subpatches?
Why does SIFT have some illumination invariance?

Making descriptor rotation invariant
- Rotate patch according to its dominant gradient orientation
- This puts the patches into a canonical orientation.

SIFT descriptor [Lowe 2004]
- Extraordinarily robust matching 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 (below)
  - Fast and efficient—can run in real time
    - Lots of code available

Example

SIFT properties
- Invariant to
  - Scale
  - Rotation
- Partially invariant to
  - Illumination changes
  - Camera viewpoint
  - Occlusion, clutter

Example

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

Image from Matthew Brown

Example

NASA Mars Rover images
Local features: main components

1) Detection: Identify the interest points

2) Description: Extract vector feature descriptor surrounding each interest point.

3) Matching: Determine correspondence between descriptors in two views

Matching local features

To generate candidate matches, find patches that have the most similar appearance (e.g., lowest SSD)
Simplest approach: compare them all, take the closest (or closest k, or within a thresholded distance)

Ambiguous matches

At what SSD value do we have a good match?
To add robustness to matching, can consider ratio:
distance to best match / distance to second best match
If low, first match looks good.
If high, could be ambiguous match.

Matching SIFT Descriptors

- Nearest neighbor (Euclidean distance)
- Threshold ratio of nearest to 2nd nearest descriptor

Recap: robust feature-based alignment

Source: L. Lazebnik
Recap: robust feature-based alignment

- Extract features
- Compute **putative matches**

Applications of local invariant features

- Wide baseline stereo
- Motion tracking
- Panoramas
- Mobile robot navigation
- 3D reconstruction
- Recognition
- …
Automatic mosaicing

http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html

Wide baseline stereo

[Image from T. Tuytelaars ECCV 2006 tutorial]

Recognition of specific objects, scenes

Schmid and Mohr 1997
Sivic and Zisserman, 2003
Rothganger et al. 2003
Lowe 2002

Summary

• Interest point detection
  – Harris corner detector
  – Laplacian of Gaussian, automatic scale selection

• Invariant descriptors
  – Rotation according to dominant gradient direction
  – Histograms for robustness to small shifts and translations (SIFT descriptor)