Lecture 13: Local invariant features

Tuesday, Oct 30
Prof. Kristen Grauman

Outline

• Types of transformations and invariance
  – Scale invariance
• Local features: detectors and descriptors
  – SIFT

What would we like our image descriptions to be invariant to?

Geometric transformations

And other nuisances…

• Noise
• Blur
• Compression artifacts
• Appearance variation for a category
Classes of transformations

- **Euclidean/rigid**: Translation + rotation
- **Similarity**: Translation + rotation + uniform scale
- **Affine**: Similarity + shear
  - Valid for orthographic camera, locally planar object
- **(Projective**: Affine + projective warps)
- **Photometric**: affine intensity change
  - $I \rightarrow aI + b$

Exhaustive search
A multi-scale approach

We want to extract the patches from each image independently.
Invariant local features

Subset of local feature types designed to be invariant to
- Scale
- Translation
- Rotation
- Affine transformations
- Illumination

1) Detect distinctive interest points
2) Extract invariant descriptors

(Good) invariant local features

- Reliably detected
- Distinctive
- Robust to noise, blur, etc.
- Description normalized properly

Interest points: From stereo to recognition

- Feature detectors previously used for stereo, motion tracking
- Now also for recognition
  - Schmid & Mohr 1997
    - Harris corners to select interest points
    - Rotationally invariant descriptor of local image regions
    - Identify consistent clusters of matched features to do recognition

Review: corner detection as an interest operator

“flat” region: no change in all directions
“edge”: no change along the edge direction
“corner”: significant change in all directions

Review: Harris Detector Workflow

Compute corner response $R$
**Review: Harris Detector Workflow**

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

**Review: Harris Detector Workflow**

Take only the points of local maxima of $R$

**Harris Detector**

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

  *Corner response $R$ is invariant to image rotation*

  But, for corner detection we must search windows at a *pre-determined scale.*

**Scale space (Witkin 83)**

- first derivative peaks
- contours of $f'' = 0$ in scale-space

**Scale space**

Scale space insights:

- edge position may shift with increasing scale ($\sigma$)
- two edges may merge with increasing scale (edges can disappear)
- an edge may *not* split into two with increasing scale (new edges do not appear)
Scale Invariant Detection

- Consider regions of different sizes around a point
- At the right scale, regions of corresponding content will look the same in both images

Solution:
- Design a function on the region (circle), which is "scale invariant" (the same for corresponding regions, even if they are at different scales)
  - Example: average intensity. For corresponding regions (even of different sizes) it will be the same.
  - For a point in one image, we can consider it as a function of region size (circle radius)

Common approach:
- Take a local maximum of this function
- Observation: region size, for which the maximum is achieved, should be invariant to image scale.
  
  Important: this scale invariant region size is found in each image independently!

Automatic scale selection
Lindeberg et al., 1996
Scale Invariant Detection

- A "good" function for scale detection: has one stable sharp peak

- For usual images: a good function would be one which responds to contrast (sharp local intensity change)

Scale space insights:
- edge position may shift with increasing scale ($\sigma$)
- two edges may merge with increasing scale (edges can disappear)
- an edge may not split into two with increasing scale (new edges do not appear)

What could be an approximation of an image’s scale space?

Scale invariant detection

Requires a method to repeatably select points in location and scale:
- Only reasonable scale-space kernel is a Gaussian (Koenderink, 1984; Lindeberg, 1994)
- An efficient choice is to detect peaks in the difference of Gaussian pyramid (Burt & Adelson, 1983; Crowley & Parker, 1984)
- Difference-of-Gaussian is a close approximation to Laplacian

Scale selection principle

- Intrinsic scale is the scale at which normalized derivative assumes a maximum -- marks a feature containing interesting structure. (T. Lindeberg ‘94)

$\Rightarrow$ Maxima/minima of Laplacian

Lowe’s DoG

Difference of Gaussians as approximation of the Laplacian of Gaussian
Scale Invariant Detection

Kernels:
\[ L = \sigma^2 \left( G_x(x,y,\sigma) + G_y(x,y,\sigma) \right) \]
(Laplacian)
\[ \text{DoG} = G(x,y,k\sigma) - G(x,y,\sigma) \]
(Difference of Gaussians)

where Gaussian
\[ G(x,y,\sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}} \]

\[ f = \text{Kernel} \ast \text{Image} \]

SIFT: Key point localization

- Detect maxima and minima of difference-of-Gaussian in scale space
- Then reject points with low contrast (threshold)
- Eliminate edge responses (use ratio of principal curvatures)

Candidate keypoints: list of \((x,y,\sigma)\)

SIFT: Example of keypoint detection

Threshold on value at DOG peak and on ratio of principle curvatures (Harris approach)

(a) 233x189 image
(b) 832 DOG extrema
(c) 729 left after peak value threshold
(d) 536 left after testing ratio of principle curvatures

Scale Invariant Detectors

- Experimental evaluation of detectors w.r.t. scale change

Scale Invariant Detection: Summary

- Given: two images of the same scene with a large scale difference between them
- Goal: find the same interest points independently in each image
- Solution: search for maxima of suitable functions in scale and in space (over the image)
Affine Invariant Detection

• Above we considered: Similarity transform (rotation + uniform scale)

• Now we go on to: Affine transform (rotation + non-uniform scale)

Affine Invariant Detection

• Intensity-based regions (IBR):
  – Start from a local intensity extrema
  – Consider intensity profile along rays
  – Select maximum of invariant function $f(t)$ along each ray
  – Connect local maxima
  – Fit an ellipse

Affine Invariant Detection

• Maximally Stable Extremal Regions (MSER)
  – Threshold image intensities: $I > I_0$
  – Extract connected components ("Extremal Regions")
  – Seek extremal regions that remain "Maximally Stable" under range of thresholds

Point Descriptors

• We know how to detect points
• Next question: How to describe them for matching?

Point descriptor should be:
1. Invariant
2. Distinctive

Rotation Invariant Descriptors

• Harris corner response measure: depends only on the eigenvalues of the matrix $M$

Rotation Invariant Descriptors

• Find local orientation
  Dominant direction of gradient

• Rotate description relative to dominant orientation

**Scale Invariant Descriptors**

- Use the scale determined by detector to compute descriptor in a normalized frame

---

**SIFT descriptors: Select canonical orientation**

- Create histogram of local gradient directions computed at selected scale
- Assign canonical orientation at peak of smoothed histogram
- Each key specifies stable 2D coordinates (x, y, scale, orientation)

---

**SIFT descriptors: vector formation**

- Thresholded image gradients are sampled over 16x16 array of locations in scale space
- Create array of orientation histograms
- 8 orientations x 4x4 histogram array = 128 dimensions

---

**SIFT properties**

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

---

**SIFT matching and recognition**

- Index descriptors
- Generalized Hough transform: vote for object poses
- Refine with geometric verification: affine fit, check for agreement between image features and model

---

**Value of local (invariant) features**

- Complexity reduction via selection of distinctive points
- Describe images, objects, parts without requiring segmentation
  - Local character means robustness to clutter, occlusion
- Robustness: similar descriptors in spite of noise, blur, etc.
Coming up

• Problem set 3 due 11/13
  – Stereo matching
  – Local invariant feature indexing

• Thursday: image indexing with bags of words
  – Read Video Google paper