Today’s lecture:
- Invariant Features, SIFT

Interest Points and Descriptors
- Recap: Harris, Correlation

Adding scale and rotation invariance
- Multi-Scale Oriented Patch descriptors

SIFT
- Scale Invariant Feature Transform
- Interest point detector (DOG)
- Descriptor
Harris Corners & Correlation

- Classic approach to image feature matching
- Harris corners
  - Low autocorrelation / high SSD in 2D
- Correlation matching
  - Equivalent to SSD for normalised patches
- Robust image matching
  - cf. pixel based error functions
- BUT: poor invariance properties
## Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is Harris invariant?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Is correlation invariant?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is Harris invariant?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Is correlation invariant?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
# Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is Harris invariant?</strong></td>
<td>YES</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td><strong>Is correlation invariant?</strong></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

![Image of mountain with annotations showing original, translated, rotated, and scaled versions.](image-url)
# Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is Harris invariant?</strong></td>
<td>YES</td>
<td>YES</td>
<td>?</td>
</tr>
<tr>
<td><strong>Is correlation invariant?</strong></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is Harris invariant?</strong></td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td><strong>Is correlation invariant?</strong></td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is Harris invariant?</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Is correlation invariant?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

4/19/2007
Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is Harris invariant?</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Is correlation invariant?</td>
<td>YES</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
## Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is Harris invariant?</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Is correlation invariant?</td>
<td>YES</td>
<td>NO</td>
<td>?</td>
</tr>
</tbody>
</table>
# Rotation/Scale Invariance

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is Harris invariant?</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Is correlation invariant?</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>
Invariant Features

Local image descriptors that are invariant (unchanged) under image transformations
Interest Points and Descriptors

- **Interest Points**
  - Focus attention for image understanding
  - Points with **2D structure** in the image
  - Examples: Harris corners, DOG maxima

- **Descriptors**
  - Describe region around an interest point
  - Must be **invariant** under
    - Geometric distortion
    - Photometric changes
  - Examples: Oriented patches, SIFT
Canonical Frames

\[ H'_{ref} = H_{ref} H^{-1} \]

\[ u' = Hu \]
Multi-Scale Oriented Patches

- Extract oriented patches at multiple scales

[ Brown, Szeliski, Winder CVPR 2005 ]
Multi-Scale Oriented Patches

- Interest point
  - x, y, scale = multi-scale Harris corners
  - rotation = blurred gradient

- Descriptor
  - Bias gain normalised image patch
  - Sampled at 5x scale
Multi-Scale Oriented Patches

- Sample scaled, oriented patch
Multi-Scale Oriented Patches

- Sample scaled, oriented patch
  - 8x8 patch, sampled at 5 x scale
Multi-Scale Oriented Patches

- Sample scaled, oriented patch
  - 8x8 patch, sampled at 5 x scale
- Bias/gain normalised
Application: Image Stitching
Image Sampling: Scale

downsampling ↓ 2

blur, downsampling ↓ 2
Image Sampling : Scale

This is aliasing!

downsample ↓ 2

blur, downsample ↓ 2
Image Sampling: Scale

↓ 2

↓ 4

↓ 8
Problem: sample grids are not aligned…
Bilinear Interpolation

- Use all 4 adjacent samples
The SIFT (Scale Invariant Feature Transform) Detector and Descriptor

developed by David Lowe
University of British Columbia
Initial paper ICCV 1999
Newer journal paper IJCV 2004
Motivation

- The Harris operator is not invariant to scale and correlation is not invariant to rotation\(^1\).

- For better image matching, Lowe’s goal was to develop an interest operator that is invariant to scale and rotation.

- Also, Lowe aimed to create a descriptor that was robust to the variations corresponding to typical viewing conditions.

\(^1\)But Schmid and Mohr developed a rotation invariant descriptor for it in 1997.
Idea of SIFT

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters.
Claimed Advantages of SIFT

- **Locality**: features are local, so robust to occlusion and clutter (no prior segmentation)
- **Distinctiveness**: individual features can be matched to a large database of objects
- **Quantity**: many features can be generated for even small objects
- **Efficiency**: close to real-time performance
- **Extensibility**: can easily be extended to wide range of differing feature types, with each adding robustness
Overall Procedure at a High Level

1. Scale-space extrema detection
   Search over multiple scales and image locations.

2. Keypoint localization
   Fit a model to determine location and scale.
   Select keypoints based on a measure of stability.

3. Orientation assignment
   Compute best orientation(s) for each keypoint region.

4. Keypoint description
   Use local image gradients at selected scale and rotation to describe each keypoint region.
1. Scale-space extrema detection

- **Goal**: Identify locations and scales that can be repeatably assigned under different views of the same scene or object.

- **Method**: search for stable features across multiple scales using a continuous function of scale.

- **Prior work** has shown that under a variety of assumptions, the best function is a **Gaussian function**.

- **The scale space of an image is a function** $L(x,y,\sigma)$ that is produced from the convolution of a Gaussian kernel (at different scales) with the input image.
Aside: Image Pyramids

And so on.

3\textsuperscript{rd} level is derived from the
2\textsuperscript{nd} level according to the same function

2\textsuperscript{nd} level is derived from the original image according to some function

Bottom level is the original image.
Aside: Mean Pyramid

And so on.

At 3\textsuperscript{rd} level, each pixel is the mean of 4 pixels in the 2\textsuperscript{nd} level.

At 2\textsuperscript{nd} level, each pixel is the mean of 4 pixels in the original image.

Bottom level is the original image.
Aside: Gaussian Pyramid
At each level, image is smoothed and reduced in size.

And so on.

At 2\textsuperscript{nd} level, each pixel is the result of applying a Gaussian mask to the first level and then subsampling to reduce the size.

Bottom level is the original image.
Example: Subsampling with Gaussian pre-filtering

Gaussian 1/2

G 1/4

G 1/8
Lowe’s Scale-space Interest Points

- **Laplacian of Gaussian** kernel
  - Scale normalised (x by scale^2)
  - Proposed by Lindeberg

- **Scale-space detection**
  - Find local maxima across scale/space
  - A good “blob” detector

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

\[ \nabla^2 G(x, y, \sigma) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} \]

[T. Lindeberg IJCV 1998]
Lowe’s Scale-space Interest Points

- Gaussian is an ad hoc solution of heat diffusion equation

\[
\frac{\partial G'}{\partial \sigma} = \sigma \nabla^2 G.
\]

- Hence

\[
G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G.
\]

- k is not necessarily very small in practice
Lowe’s Scale-space Interest Points

- **Scale-space function** $L$
  - Gaussian convolution
    $$L(x, y, \sigma) = G(x, y, \sigma) \ast I(x, y),$$
    where $\sigma$ is the width of the Gaussian.

- **Difference of Gaussian kernel** is a close approximate to scale-normalized Laplacian of Gaussian
  $$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \ast I(x, y)$$
  $$= L(x, y, k\sigma) - L(x, y, \sigma).$$

- Can approximate the Laplacian of Gaussian kernel with a difference of **separable** convolutions

\[ G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)}. \]
Lowe’s Pyramid Scheme

• Scale space is separated into octaves:
  • Octave 1 uses scale $\sigma$
  • Octave 2 uses scale $2\sigma$
  • etc.

• In each octave, the initial image is repeatedly convolved with Gaussians to produce a set of scale space images.

• Adjacent Gaussians are subtracted to produce the DOG.

• After each octave, the Gaussian image is down-sampled by a factor of 2 to produce an image $\frac{1}{4}$ the size to start the next level.
Lowe’s Pyramid Scheme

\[ \sigma_{s+1} = 2^{(s+1)/s} \sigma_0 \]

\[ \cdots \]

\[ \sigma_i = 2^{i/s} \sigma_0 \]

\[ \sigma_2 = 2^{2/s} \sigma_0 \]

\[ \sigma_1 = 2^{1/s} \sigma_0 \]

\[ \sigma_0 \]

The parameter \( s \) determines the number of images per octave.
Key point localization

- Detect maxima and minima of difference-of-Gaussian in scale space.

- Each point is compared to its 8 neighbors in the current image and 9 neighbors each in the scales above and below.

For each max or min found, output is the location and the scale.
Scale-space extrema detection: experimental results over 32 images that were synthetically transformed and noise added.

Stability

- Sampling in scale for efficiency
  - How many scales should be used per octave? S=?
    - More scales evaluated, more keypoints found
    - S < 3, stable keypoints increased too
    - S > 3, stable keypoints decreased
    - S = 3, maximum stable keypoints found
Keypoint localization

- Once a keypoint candidate is found, perform a detailed fit to nearby data to determine
  - location, scale, and ratio of principal curvatures
- In initial work keypoints were found at location and scale of a central sample point.
- In newer work, they fit a 3D quadratic function to improve interpolation accuracy.
- The Hessian matrix was used to eliminate edge responses.
Eliminating the Edge Response

- **Reject flats:**
  - $|D(\hat{x})| < 0.03$

- **Reject edges:**

  Let $\alpha$ be the eigenvalue with larger magnitude and $\beta$ the smaller.

  $\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$

  $\operatorname{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta,$
  $\operatorname{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$

  Let $r = \alpha/\beta$. So $\alpha = r\beta$

  $r < 10$

- **What does this look like?**

  $\frac{\operatorname{Tr}(\mathbf{H})^2}{\operatorname{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r},$ (r+1)^2/r is at a min when the 2 eigenvalues are equal.
3. Orientation assignment

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

If 2 major orientations, use both.
Keypoint localization with orientation

(a) 233x189 initial keypoints
(b) 832 keypoints after gradient threshold
(c) 729 keypoints after ratio threshold
(d) 536 keypoints after ratio threshold
4. Keypoint Descriptors

- At this point, each keypoint has
  - location
  - scale
  - orientation

- Next is to compute a descriptor for the local image region about each keypoint that is
  - highly distinctive
  - invariant as possible to variations such as changes in viewpoint and illumination
Normalization

- Rotate the window to standard orientation
- Scale the window size based on the scale at which the point was found.
In experiments, 4x4 arrays of 8 bin histogram is used, a total of 128 features for one keypoint.
Statistical Motivation

- Interest points are subject to error in
  - position, scale, orientation

[ Geometric Blur Berg Malik, GLOH Mikolajczyk etc. ]
Biological Motivation

- Mimic complex cells in primary visual cortex
- Hubel & Wiesel found that cells are sensitive to orientation of edges, but insensitive to their position
- This justifies spatial pooling of edge responses

[“Eye, Brain and Vision” – Hubel and Wiesel 1988]
Lowe’s Keypoint Descriptor

- use the **normalized** region about the keypoint
- compute gradient magnitude and orientation at each point in the region
- **weight them by a Gaussian** window overlaid on the circle
- create an **orientation histogram** over the 4 X 4 subregions of the window
- 4 X 4 descriptors over 16 X 16 sample array were used in practice. 4 X 4 times 8 directions gives a vector of **128 values**.
Using SIFT for Matching “Objects”
Uses for SIFT

- Feature points are used also for:
  - Image alignment (homography, fundamental matrix)
  - 3D reconstruction (e.g. Photo Tourism)
  - Motion tracking
  - Object recognition
  - Indexing and database retrieval
  - Robot navigation
  - … many others

[ Photo Tourism: Snavely et al. SIGGRAPH 2006 ]