Interest Operators

All lectures are from posted research papers.

• **Harris Corner Detector: the first and most basic interest operator**

• **Kadir Entropy Detector and its use in object recognition**

• **SIFT interest point detector and region descriptor**

• **MSER region detector and Harris Affine in region matching**
0. Introduction to Interest Operators

• Find “interesting” pieces of the image
  – e.g. corners, salient regions
  – Focus attention of algorithms
  – Speed up computation

• Many possible uses in matching/recognition
  – Search
  – Object recognition
  – Image alignment & stitching
  – Stereo
  – Tracking
  – …
Interest Points have 2D structure.
Local invariant photometric descriptors

- Local: robust to occlusion/clutter + no segmentation
- Photometric: (use pixel values) distinctive descriptions
- Invariant: to image transformations + illumination changes
History - Matching

1. Matching based on correlation alone
2. Matching based on geometric primitives
e.g. line segments

⇒ Not very discriminating  (why?)

⇒ Solution : matching with interest points & correlation

[ A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry,
Z. Zhang, R. Deriche, O. Faugeras and Q. Luong,
Artificial Intelligence 1995 ]
Zhang Approach

- Extraction of interest points with the Harris detector
- Comparison of points with cross-correlation
- Verification with the fundamental matrix

The fundamental matrix maps points from the first image to corresponding points in the second matrix using a homography that is determined through the solution of a set of equations that usually minimizes a least square error. CH 12-13
Preview: Harris detector

Interest points extracted with Harris (~ 500 points)
Harris detector
Harris detector
Cross-correlation matching

Match two points based on how similar their neighborhoods are.

Initial matches – motion vectors (188 pairs)
Global constraints

Robust estimation of the fundamental matrix (RANSAC)

99 inliers

89 outliers
General Interest Detector/Descriptor Approach

1) Extraction of interest points (characteristic locations)
2) Computation of local descriptors (rotational invariants)
3) Determining correspondences
4) Selection of similar images
1. Harris detector

Based on the idea of auto-correlation

Important difference in all directions => interest point
Background: Moravec Corner Detector

- take a window \( w \) in the image
- shift it in four directions \((1,0), (0,1), (1,1), (-1,1)\)
- compute a difference for each
- compute the min difference at each pixel
- local maxima in the min image are the corners

\[
E(x,y) = \sum_{u,v \text{ in } w} w(u,v) |I(x+u,y+v) - I(u,v)|^2
\]
Shortcomings of Moravec Operator

• Only tries 4 shifts. We’d like to consider “all” shifts.

• Uses a discrete rectangular window. We’d like to use a smooth circular (or later elliptical) window.

• Uses a simple min function. We’d like to characterize variation with respect to direction.

Result: Harris Operator
Harris detector

Auto-correlation fn (SSD) for a point \((x, y)\) and a shift \((\Delta x, \Delta y)\)

\[
f(x, y) = \sum_{(x_k, y_k) \in W} \left[ I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y) \right]^2
\]

SSD means summed square difference

Discrete shifts can be avoided with the auto-correlation matrix

with

\[
I(x_k + \Delta x, y_k + \Delta y) = I(x_k, y_k) + (I_x(x_k, y_k) \Delta x + I_y(x_k, y_k) \Delta y)
\]

what is this?

\[
f(x, y) = \sum_{(x_k, y_k) \in W} \left( \begin{pmatrix} I_x(x_k, y_k) & I_y(x_k, y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2
\]
Harris detector

Rewrite as inner (dot) product

\[
f(x, y) = \sum_{(x_k, y_k) \in W} \left( \begin{bmatrix} I_x(x_k, y_k) & I_y(x_k, y_k) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right)^2
\]

\[
= \sum_{(x_k, y_k) \in W} \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} I_x(x_k, y_k) \\ I_y(x_k, y_k) \end{bmatrix} \begin{bmatrix} I_x(x_k, y_k) & I_y(x_k, y_k) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}
\]

The center portion is a 2x2 matrix

\[
= \sum_W \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}
\]

\[
= \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \sum_W \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}
\]

Have we seen this matrix before?
Harris detector

\[
W \begin{bmatrix}
(x, y)
\end{bmatrix}
\]

\[
= \begin{pmatrix}
\Delta x & \Delta y
\end{pmatrix}
\begin{bmatrix}
\sum_{(x_k, y_k) \in W} (I_x(x_k, y_k))^2 & \sum_{(x_k, y_k) \in W} I_x(x_k, y_k)I_y(x_k, y_k) \\
\sum_{(x_k, y_k) \in W} I_x(x_k, y_k)I_y(x_k, y_k) & \sum_{(x_k, y_k) \in W} (I_y(x_k, y_k))^2
\end{bmatrix}
\begin{pmatrix}
\Delta x \\
\Delta y
\end{pmatrix}
\]

Auto-correlation matrix M
Harris detection

• Auto-correlation matrix
  – captures the structure of the local neighborhood
  – measure based on eigenvalues of M
    • 2 strong eigenvalues => interest point
    • 1 strong eigenvalue  => contour
    • 0 eigenvalue       => uniform region

• Interest point detection
  – threshold on the eigenvalues
  – local maximum for localization
Harris Corner Detector

- Corner strength \( R = \text{Det}(M) - k \text{Tr}(M)^2 \)
- Let \( \alpha \) and \( \beta \) be the two eigenvalues
- \( \text{Tr}(M) = \alpha + \beta \)
- \( \text{Det}(M) = \alpha \beta \)

- \( R \) is positive for corners, negative for edges, and small for flat regions
- Selects corner pixels that are 8-way local maxima

\[
A = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix} \quad \text{det}(A) = a_{11}a_{22} - a_{12}a_{21} \quad \text{tr}(A) = a_{11} + a_{22}
\]

\( R = \text{Det}(M) - k \text{Tr}(M)^2 \) is the Harris Corner Detector.
Now we need a descriptor

Vector comparison using a distance measure

What are some suitable distance measures?
Distance Measures

• We can use the sum-square difference of the values of the pixels in a square neighborhood about the points being compared.

\[
\text{SSD} = \sum\sum (W_1(i,j) - (W_2(i,j))^2)
\]

Works when the motion is mainly a translation.
Some Matching Results from Matt Brown
Some Matching Results
Summary of the approach

• Basic feature matching = **Harris Corners & Correlation**
• Very good results in the presence of occlusion and clutter
  – local information
  – discriminant greyvalue information
  – invariance to image rotation and illumination
• Not invariance to scale and affine changes
• Solution for more general view point changes
  – local invariant descriptors to scale and rotation
  – extraction of invariant points and regions
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

Is Harris invariant?

Is correlation invariant?

<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>YES</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>YES</td>
<td>YES</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>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>
# Rotation/Scale Invariance

![ images showing original, translated, rotated, and scaled versions of a mountain scene with a circle indicating a point of interest. ]

<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>Original</th>
<th>Translated</th>
<th>Rotated</th>
<th>Scaled</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rotation/Scale Invariance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Is Harris invariant?</strong></td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><strong>Is correlation invariant?</strong></td>
<td>YES</td>
<td>NO</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>NO</td>
</tr>
</tbody>
</table>
Matt Brown’s Invariant Features

- Local image descriptors that are *invariant* (unchanged) under image transformations
Canonical Frames

$H$

$H_{ref}$

$u$

$u'_{ref}$

$H'_{ref}$
Canonical Frames
Multi-Scale Oriented Patches

- Extract oriented patches at multiple scales using dominant orientation
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 normalized (subtract the mean of a patch and divide by the variance to normalize)
  - $I' = (I - \mu)/\sigma$
Matching Interest Points: Summary

- **Harris corners / correlation**
  - Extract and match repeatable image features
  - Robust to clutter and occlusion
  - BUT not invariant to scale and rotation

- **Multi-Scale Oriented Patches**
  - Corners detected at multiple scales
  - Descriptors oriented using local gradient
    - Also, sample a blurred image patch
  - Invariant to scale and rotation

Leads to: **SIFT** – state of the art image features