• lecture topics

– Interest points 1 (Linda)
  • interest points, descriptors, Harris corners, correlation matching

– Interest points 2 (Linda)
  • Kadir operator, Fergus object recognition, Sivic video indexing

– Interest points 3 (Matt Brown from Microsoft Research: April 18)
  • improving feature matching and indexing, SIFT

– Image Stitching (Matt: April 23)
  • automatically combine multiple images to give seamless, “stitched” result
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

0D structure
→ not useful for matching

1D structure
→ edge, can be localised in 1D, subject to the aperture problem

2D structure
→ corner, or interest point, can be localised in 2D, good for matching

Interest Points have 2D structure. Edge Detectors e.g. Canny [Canny86] exist, but descriptors are more difficult.
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 ]
Approach

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

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

Initial matches – motion vectors (188 pairs)
Global constraints

Robust estimation of the fundamental matrix (RANSAC)

99 inliers

89 outliers
Summary of the approach

• Very good results in the presence of occlusion and clutter
  – local information
  – discriminant greyvalue information
  – robust estimation of the global relation between images
  – works well for \textit{limited view point changes}

• Solution for more general view point changes
  – wide baseline matching (different viewpoint, scale and rotation)
  – local \textit{invariant descriptors} based on greyvalue information
Invariant Features

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

Based on the idea of auto-correlation

Important difference in all directions => interest point
Autocorrelation

Autocorrelation function (ACF) measures the self similarity of a signal

$$ACF = r(a) = \int_{-\infty}^{\infty} f(x)f(x-a)\,dx$$

$$SSD = e(a) = \int_{-\infty}^{\infty} (f(x) - f(x-a))^2\,dx$$

Autocorrelation function (ACF) measures the **self similarity** of a signal.
Autocorrelation

- Autocorrelation related to sum-square difference:

\[ \text{SSD} = 2(1 - \text{ACF}) \]

(if \( \int f(x)^2 \, dx = 1 \))
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 \((\Delta x, \Delta y)\)

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

Discrete shifts can be avoided with the auto-correlation matrix

with \(I(x_k + \Delta x, y_k + \Delta y) = I(x_k, y_k) + (I_x(x_k, y_k) \quad I_y(x_k, y_k))(\Delta x \quad \Delta y)\)

\[
f(x, y) = \sum_{(x_k, y_k) \in W} \left( I_x(x_k, y_k) \quad I_y(x_k, y_k) \right)^2
\]

what is this?
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

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

\[
= \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \sum_{W} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}
\]
Harris detector

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

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
Some Details from the Harris Paper

- 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, - for edges, and small for flat regions
- Select corner pixels that are 8-way local maxima

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

$$\text{tr}(A) = a_{11} + a_{22}$$
Determining correspondences

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.

\[
SSD = \sum\sum (W1(i,j) - (W2(i,j)))^2
\]
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 viewpoint 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

<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>?</td>
<td>?</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>?</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>Image Type</th>
<th>Translation</th>
<th>Rotation</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>YES</td>
<td>YES</td>
<td>?</td>
</tr>
<tr>
<td>Translated</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Rotated</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Scaled</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

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

Is Harris invariant?
- YES
- YES
- NO

Is correlation invariant?
- ?
- ?
- ?

<table>
<thead>
<tr>
<th>Original</th>
<th>Translated</th>
<th>Rotated</th>
<th>Scaled</th>
</tr>
</thead>
<tbody>
<tr>
<td></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>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</td>
<td>YES</td>
<td>NO</td>
<td>?</td>
</tr>
<tr>
<td>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>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>
Invariant Features

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

\[ H \]

\[ H_{ref} \]

\[ u' = Hu \]

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

\[ u' = Hu \]
\[ H'_{ref} = H_{ref}H^{-1} \]
Multi-Scale Oriented Patches

- Extract oriented patches at multiple scales
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
  – \( I' = (I - \mu)/\sigma \)
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