6.891
Computer Vision and Applications

Prof. Trevor Darrell

Lecture 6: Local Features
- Interest operators
- Correspondence
- Invariances
- Descriptions

Readings: Shi and Tomasi; Lowe.

Local Features

Matching points across images important for recognition and pose estimation

Tracking vs. Indexing

Today

Interesting points, correspondence, affine patch tracking

Scale and rotation invariant descriptors [Lowe]

Correspondence using window matching

Points are highly individually ambiguous…
More unique matches are possible with small regions of image.

Correspondence using window matching

Sum of Squared (Pixel) Differences

$w_x$ and $w_y$ are corresponding $m$ by $n$ windows of pixels.

We define the window function:

$W(x, y) = [a, v] | x - \frac{D}{2} \leq x \leq x + \frac{D}{2}, y - \frac{D}{2} \leq y \leq y + \frac{D}{2}$

The SSD cost measures the intensity difference as a function of disparity:

$C(x, y, d) = \sum_{(a, v) \in W(x, y)} (I_{a}(x, y) - I_{a-d}(x, y))^{2}$
Image Normalization

- Even when the cameras are identical models, there can be differences in gain and sensitivity.
- The cameras do not see exactly the same surfaces, so their overall light levels can differ.
- For these reasons and more, it is a good idea to normalize the pixels in each window:

\[
I = \frac{1}{\sum_{(u,v)} I(u,v)} \sum_{(u,v)} I(u,v) \quad \text{Average pixel}
\]

\[
\|W_{(u,v)}\| = \left(\sum_{(u,v)} W_{(u,v)}^2\right)^{1/2} \quad \text{Window magnitude}
\]

\[
\hat{l}(x,y) = \frac{l(x,y) - I}{\sqrt{\sum_{(u,v)} W_{(u,v)}^2}} \quad \text{Normalized pixel}
\]

Images as Vectors

Possible metrics

\[w_R(d)\]

\[w_L\]

Home vector of each window

Distance?

Angle?

Image windows as vectors

Local Features

Not all points are equally good for matching…
(Review) Differential approach: Optical flow constraint equation

Brightness should stay constant as you track motion

\[ I(x + u \partial_t, y + v \partial_t, t + \partial_t) = I(x, y, t) \]

1st order Taylor series, valid for small \( \partial \)

\[ I(x, y, t) + u \partial_t I_x + v \partial_t I_y + \partial I = I(x, y, t) \]

Constraint equation

\[ u I_x + v I_y + I_t = 0 \]

“BCCE” - Brightness Change Constraint Equation

Aperture Problem and Normal Flow

The gradient constraint:

\[ I_u + I_v + I = 0 \]

\[ \nabla I \cdot \vec{U} = 0 \]

Defines a line in the \((u, v)\) space

Normal Flow:

\[ u = \frac{I_x}{\nabla I \cdot \nabla I} \]

Combining Local Constraints

\[ \nabla I \cdot \vec{U} = -I \]

\[ \nabla I_\theta \cdot \vec{U} = -I \]

\[ \nabla I \cdot \vec{U} = -I \]

etc.

Lucas-Kanade: Integrate gradients over a Patch

Assume a single velocity for all pixels within an image patch

\[ E(u, v) = \sum_{x \in \text{patch}} I(x, y) u + I(x, y) v + I \]

Solve with:

\[ \begin{bmatrix} I_x \sum I_x & I_x \sum I_x \\ I_y \sum I_y & I_y \sum I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_x \sum I_x \\ I_y \sum I_y \end{bmatrix} \]

On the LHS: sum of the 2x2 outer product tensor of the gradient vector

\[ \sum \nabla I^2 \cdot \vec{U} = - \nabla I^2 \]

Selecting Good Features

- What’s a “good feature”?
  - Satisfies brightness constancy
  - Has sufficient texture variation
  - Does not have too much texture variation
  - Corresponds to a “real” surface patch
  - Does not deform too much over time
**Good Features to Track**

\[
\begin{bmatrix}
\sum_{i,j} f_i^2 \\
\sum_{i,j} f_i f_j
\end{bmatrix}
\begin{bmatrix}
u \\
u
\end{bmatrix}
-
\begin{bmatrix}
\sum_{i,j} f_i \\
\sum_{i,j} f_i f_j
\end{bmatrix}
\begin{bmatrix}
u \\
u
\end{bmatrix}
\]

\( \mathbf{A} \mathbf{u} = \mathbf{b} \)

**When is This Solvable?**
- \( \mathbf{A} \) should be invertible
- \( \mathbf{A} \) should not be too small due to noise
- Eigenvalues \( \lambda_1 \) and \( \lambda_2 \) of \( \mathbf{A} \) should not be too small
- \( \lambda_1, \lambda_2 \) should not be too large (\( \lambda_1 = \) larger eigenvalue)

Both conditions satisfied when \( \min(\lambda_1, \lambda_2) > c \)

---

**Harris detector**

Same idea, based on the idea of auto-correlation

Important difference in all directions \( \Rightarrow \) interest point

---

**Selecting Good Features**

\( \lambda_1 \) and \( \lambda_2 \) are large

---

**Selecting Good Features**

large \( \lambda_1 \), small \( \lambda_2 \)
Selecting Good Features

Feature Distortion

- Feature may change shape over time
  - Need a distortion model to really make this work

Find displacement $(u,v)$ that minimizes SSD error over feature region

$$
\sum_{(x,y) \in F} [(W(x,y), W(x,y)) - I(x,y)]^2
$$

(minimize with respect to $W_x$ and $W_y$)

Shi and Tomasi: use affine model for verification

$$
W_x(x,y) = ax + by + c
$$

$$
W_y(x,y) = ex + fy + g
$$

Affine Motion

$$
\begin{align*}
\Phi(x,y) &= \Phi_1(x,y) + \Phi_2(x,y) \\
\Phi_1(x,y) &= a_1 + a_2x + a_3y \\
\Phi_2(x,y) &= a_4 + a_5x + a_6y
\end{align*}
$$

Substituting into the B.C.E.:

$$
I_s \cdot u + I_s \cdot v + I_s \approx 0
$$

Each pixel provides 1 linear constraint in 6 global unknowns

(least 6 pixels necessary)

Least Square Minimization (over all pixels):

$$
\text{Err}({\bar{a}}) = \sum [(I_s(a_1 + a_2x + a_3y) + I_s(a_4 + a_5x + a_6y) + I_s)^2]
$$

Dissimilarity

- Translation
- Affine
- Similarity
- Affine deformation
- Epipolar
Convergence

Translation Dissimilarity

Affine Dissimilarity

Tracking vs. Indexing

But….

What if you can’t track over time?

CVPR 2003 Tutorial

Recognition and Matching
Based on Local Invariant Features

David Lowe
Computer Science Department
University of British Columbia

Today

Interesting points, correspondence, affine patch tracking

Scale and rotation invariant descriptors [Lowe]
**Invariant Local Features**
- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters

**Scale invariance**
Requires a method to repeatably select points in location and scale:
- The 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 – but examining more scales)
- Difference-of-Gaussian with constant ratio of scales is a close approximation to Lindeberg’s scale-normalized Laplacian (can be shown from the heat diffusion equation)

**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)
Example of keypoint detection

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

(a) 213x150 image  
(b) 32 DOG extrema  
(c) 129 left after peak value threshold  
(d) 53 left after testing ratio of principle curvatures

SIFT 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

Feature stability to noise

- Match features after random change in image scale & orientation, with differing levels of image noise
- Find nearest neighbor in database of 30,000 features

Feature stability to affine change

- Match features after random change in image scale & orientation, with 2% image noise, and affine distortion
- Find nearest neighbor in database of 30,000 features

Distinctiveness of features

- Vary size of database of features, with 30 degree affine change, 2% image noise
- Measure % correct for single nearest neighbor match

Today

Interesting points, correspondence, affine patch tracking

Scale and rotation invariant descriptors [Lowe]