Lecture 6: SURF and HOG

Bohyung Han
CSE, POSTECH
bhhan@postech.ac.kr

SURF

• Speed-Up Robust Features (SURF)
  ▪ Simplified version of SIFT
  ▪ Faster computation but comparable performance

• Characteristics
  ▪ Fast interest point detection
  ▪ Distinctive interest point description
  ▪ Speeded-up descriptor matching
  ▪ Invariant to common image transformations:
    ▪ Image rotation
    ▪ Scale changes
    ▪ Illumination change
    ▪ Small change in viewpoint

Interest Point Detection

• Gaussian second order derivatives
  \[
  \frac{\partial^2}{\partial x^2} G(\sigma) \\
  \frac{\partial^2}{\partial y^2} G(\sigma) \\
  \frac{\partial^2}{\partial x\partial y} G(\sigma)
  \]

• Hessian-based interest point localization

\[
H(x, y, \sigma) = \begin{bmatrix}
L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\
L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma)
\end{bmatrix}
\]

\[
L_{xx}(x, y, \sigma) = \frac{\partial^2}{\partial x^2} G(\sigma) \ast I(x, y)
\]

\[
L_{yy}(x, y, \sigma) = \frac{\partial^2}{\partial y^2} G(\sigma) \ast I(x, y)
\]

\[
L_{xy}(x, y, \sigma) = \frac{\partial^2}{\partial x\partial y} G(\sigma) \ast I(x, y)
\]

Integral Image

• Advantage of integral images
  ▪ Speed up the computation of the second order derivatives of Gaussian and Haar-wavelet responses
  ▪ Can be efficiently computed by row sum followed by column sum or vice versa.

\[
ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')
\]

\[
s(x, y) = s(x, y - 1) + i(x, y)
\]

\[
ii(x, y) = ii(x - 1, y) + s(x, y)
\]

Image

\[
\begin{array}{ccc}
0 & 1 & 1 \\
1 & 2 & 2 \\
1 & 3 & 1
\end{array}
\]

Integral image

\[
\begin{array}{ccc}
0 & 1 & 2 \\
1 & 4 & 7 \\
2 & 7 & 11
\end{array}
\]

\[
\begin{array}{ccc}
0 & 1 & 0 \\
1 & 1 & 3 \\
3 & 11 & 16
\end{array}
\]
Feature Evaluation with Integral Image

- How to compute the sum of the pixel intensities efficiently using integral image?

Pre-computed integral image

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>ii(x, y)</td>
<td>ii(x1, y1)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>ii(x3, y3)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>ii(x4, y4)</td>
<td></td>
</tr>
</tbody>
</table>

\[ \sum_{(x,y) \in D} i(x,y) = ii(x_4,y_4) - ii(x_3,y_3) - ii(x_2,y_2) + ii(x_1,y_1) \]

The evaluation of two-, three- and four-rectangle features require only 6, 8 and 9 table look-ups, respectively.

Approximated Blob Response

- Approximation with integral images

\[ \frac{\partial^2}{\partial y^2} G(\sigma) \quad F_{yy}(\sigma) \quad \frac{\partial^2}{\partial x \partial y} G(\sigma) \quad F_{xy}(\sigma) \]

- Filter responses can be computed using integral images.
- Very fast: Computation time is independent of filter size.
- Performance is comparable or even better than cropped Gaussians.

Approximated blob response: by determinant of Hessian as

\[ \det(H) = D_{xx}D_{yy} - (wD_{xy})^2 \quad w \approx 0.9 \]

Scale Space Pyramid

- Scale analysis with constant image size
  - Computation time is independent of filter size.

- Filter sizes
  - 1st octave: 9x9, 15x15, 21x21, 27x27
  - 2nd octave: 15x15, 27x27, 39x39, 51x51
  - 3rd octave: 27x27, 51x51, 75x75, 99x99
Scale Selection

- Non-maximum suppression and interpolation
- Blob-like feature detector

Orientation Assignment

- Methodology
  - The Haar wavelet responses are represented as vectors.
  - Sum all responses within a sliding orientation window covering 60 degrees.
  - The two summed response yield a new vector.
  - The longest vector is the dominant orientation.

Circular neighborhood of radius $6s$ around interest point $(s =$ the scale at the point: $9 \times 9$ filter is equivalent $s = 1.2.)$

Building the Descriptor

- Descriptor specification
  - Split the interest region up into $4 \times 4$ square sub-regions
  - Compute gradients by applying Haar-like features
  - Compute $\sum dx$, $\sum |dx|$, $\sum dy$, and $\sum |dy|$: $64D$ altogether
  - Normalize the vector into unit length

Examples of Descriptors
**Robustness of SURF**

- Image sub-region
  - Clean
  - Noisy

**SIFT gradients**
- $\sum dx |dx|$,
- $\sum dy |dy|$,
- $\sum d\theta |d\theta|$

**SURF sums**
- $\sum dx |dx|$, $\sum dy |dy|$, $\sum d\theta |d\theta|$

**SIFT vs. SURF**

- **SIFT**
  - Apply DoG or LoG
  - Find local optima
  - Remove edge responses using Hessian
  - Affine transformation
  - Orientation normalization
  - Extract descriptor

- **Surf**
  - Compute Hessian at each position
  - Identify interest points
  - Orientation normalization
  - Extract descriptor

  Missing in SURF

  Can be performed efficiently using integral image

**HOG**

- Histogram of Oriented Gradients (HOG)
  - Normalize gamma and color
  - Compute gradients in the region to be described
  - Divide the region into cells
  - Construct histogram of gradient orientations for each cell
  - Group the cells into large blocks
  - Normalize each block

**Variations**

- **Parameters**
  - Gradient scale
  - Size of blocks and cells
  - Orientation bins
  - Percentage of block overlap

- **Schemes**
  - RGB or Lab, color/gray-space
  - Block normalization
  - Block shapes: R-HOG, C-HOG

Implementation Details

- Gradients
  - $[-1\ 0\ 1]$ and $[-1\ 0\ 1]^T$ were good enough.
- Cell histograms
  - Each pixel within the cell casts a weighted vote (by gradient magnitude) for an orientation histogram.
  - 9 channels based on the values found in the unsigned gradient
- Blocks
  - Group the cells together into larger blocks, by either R-HOG or C-HOG.
  - Normalization

\[
L_1: v \leftarrow \frac{v}{\|v\|_1 + \epsilon} \quad L_1\ sqrt: v \leftarrow \sqrt{\frac{v}{\|v\|_1 + \epsilon}} \\
L_2: v \leftarrow \frac{v}{\|v\|_2 + \epsilon^2} \quad L_2\ Hys: L_2\ norm + clipping\ at\ 0.2\ and\ renormalization
\]

An Example of HOG Descriptor

- For pedestrian detection
- Specification
  - Cell size: 8x8
  - Block size: 16x16
  - Each window has 8x16 cells.
  - Each block is composed of 2x2 cells, which means that there are 7x15 blocks.
  - No orientation normalization
  - High dimensionality

Other Feature Descriptors

- Local Binary Patterns (LBP)
- Bias and gain normalization (MOPS)
- PCA-SIFT
- Gradient location-orientation histogram (GLOH)