Histograms of Oriented Gradients for Human Detection by Navneet Dalal, Bill Triggs

Marzieh Sharbat Maleki
Spring 2014
Challenges of pedestrian detection

• Wide variety of articulated poses
• Variable appearance/clothing
• Complex backgrounds
• Unconstrained illumination
• Occlusions
• Different Scales
Histogram of Oriented Gradients (HOG) Steps:

- Extract fixed-sized (64x128 pixel) window at each position and scale.

- HOG feature extraction:
  Compute centered horizontal and vertical gradients orientation and magnitudes with no smoothing and create histograms over cells.
  - The combination of these histograms then represents the descriptor.
  - For color image, pick the color channel with the highest gradient magnitude for each pixel.
  - HOG descriptor assumes that the local object appearance and shape within an image is described by the distribution of intensity gradients or edge directions.
  - Score the window with a linear SVM classifier
  - Perform non-maxima suppression to remove overlapping detections with lower scores.
Main Advantages:

• Since it operates on localized cells, it shows invariance to geometric and photometric transformations.
• The HOG descriptor is particularly suited for human detection in images. Essential in contextually critical environments: surveillance of pedestrians, vehicles and groups of unknown objects.

Performance limited by

- the occlusion problem often occurring in surveillance applications.
- noise occurring in e.g. large illumination variations, persistent shadows.
• Tested with
  – RGB  Slightly better performance vs. grayscale
  – LAB
  – Grayscale
• Gamma Normalization and Compression
  – Square root of image intensity  Very slightly better performance vs. no adjustment

• This step can be omitted in HOG descriptor computation, as the descriptor normalization essentially achieves the same result.
• They used Gaussian smoothing followed by one of the several discrete derivative masks for computing gradients.
• Although, performing Gaussian smoothing before applying the derivative mask, reduces the performance.
• Centered filter outperforms the rest.
Comparison of different Sigma for calculating Gaussian:
Blocks, Cells:

- For a 64x128 image, divide the image into 16x16 blocks of 50% overlap.
- 7x15=105 blocks in total.
- Each block should consist of 2x2 cells with size 8x8.
- Quantize the gradient orientation into 9 bins.
- The vote is the gradient magnitude.

NOTE: HOG blocks typically overlap: each cell contributes more than once to the final descriptor.

- 9 Bins:
Comparison of number of Bins:

DET – effect of number of orientation bins $\beta$

- $\text{bin}= 9$ (0–180)
- $\text{bin}= 6$ (0–180)
- $\text{bin}= 4$ (0–180)
- $\text{bin}= 3$ (0–180)
- $\text{bin}= 18$ (0–360)
- $\text{bin}= 12$ (0–360)
- $\text{bin}= 8$ (0–360)
- $\text{bin}= 6$ (0–360)

miss rate vs.
false positives per window (FPPW)
Blocks:

Two main block geometries exist:

- **R-HOG blocks**: Rectangular or square grids represented by three parameters:
  - the number of cells per block.
  - the number of pixels per cell.
  - the number of channels per cell histogram.

- **C-HOG blocks**: Circular blocks
  - a) With one single, central cell.
  - b) With an angularly-divided central cell.

C-HOG blocks can be represented by these parameters:
- the number of angular and radial bins.
- the radius of the center bin.
Effect of Block and Cell Size:
• Contrast normalization is essential and results in better invariance to changes in illumination, shadowing or foreground-background contrast.

• Different methods for block normalization:
  - L1-norm: \[ L1\text{-norm} : v \mapsto v / (\| v \|_1 + \epsilon) \]
  - L2-norm: \[ L2\text{-norm} : v \mapsto v / \sqrt{\| v \|_2^2 + \epsilon^2} \]
  - L1-sqrt: \[ L1\text{-sqrt} : v \mapsto \sqrt{v / (\| v \|_1 + \epsilon)} \]

• All methods showed very significant improvement over the non-normalized data. The best methods are L2-norm and L1-sqrt.
Comparison of different Normalization methods:

- L2–Hys
- L2–norm
- L1–Sqrt
- L1–norm
- No norm
- Window norm

DET – effect of normalization methods

miss rate

false positives per window (FPPW)
Concatenate histograms:

- Make it a 1D matrix of length 3780.

\[ \text{# features} = 15 \times 7 \times 9 \times 4 = 3780 \]

- Visualization:
HOG descriptors are fed into a recognition system based on SVM supervised learning which looks for an optimal hyper plane as a decision function.
## Data Sets Evaluation:

<table>
<thead>
<tr>
<th>MIT pedestrian database</th>
<th>INRIA person database</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Database Images]</td>
<td>![Database Images]</td>
</tr>
<tr>
<td><strong>Train</strong></td>
<td><strong>Train</strong></td>
</tr>
<tr>
<td>507 positive windows</td>
<td>1208 positive windows</td>
</tr>
<tr>
<td>Negative data unavailable</td>
<td>1218 negative images</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td><strong>Test</strong></td>
</tr>
<tr>
<td>200 positive windows</td>
<td>566 positive windows</td>
</tr>
<tr>
<td>Negative data unavailable</td>
<td>453 negative images</td>
</tr>
<tr>
<td>Overall 709 annotations+ reflections</td>
<td>Overall 1774 annotations+ reflections</td>
</tr>
</tbody>
</table>
Overall Performance:

MIT pedestrian database

INRIA person database
Movie Example vs. Image Example
Thank You!
References

- Histograms of Oriented Gradients for Human Detection by Navneet Dalal, Bill Triggs – CVPR 2005
- Pedestrian Detection, Pete Barnum presentation