

CS 6643

Histograms of Oriented Gradients (HOG)

HOG Feature

- HOG = Histograms of Oriented Gradients
- Proposed by Dalal & Triggs in CVPR 2005 for detecting human in 2D images
- One of the most widely used feature descriptor for detecting human in 2D images
 - descriptor = a feature representation computed from the images



Human Detection using HOG



HOG Feature

- Slide a detection window across different locations of an image.
- Divide detection window into small non-overlapping regions called cells.
- Compute *intensity gradient orientation histograms* in each cell
- Human shapes can be described by local gradient orientations; local histograms insensitive to small variations in the location and pose of the human shape.



HOG Feature

- Normalize local gradient orientation histograms over larger local regions called blocks
 - For handling variations in illuminations and image contrast.
 - Blocks are made up of cells and overlapping blocks are typically used in HOG.
- The normalized histograms from the individual blocks are concatenated together to form the final HoG descriptor (or feature vector.)

HoG: Implementation Details

- Assume detection window size = 128×64 pixels (rows x columns), cell size = 8×8 pixels, block size = 16×16 pixels (or 2×2 cells), block overlap step size = 8 pixels (or 1 cell).
- Window divided into 16×8 non-overlapping cells of 8×8 pixels each.
- Group of 4 cells form a block.
- There are $15 \times 7 = 105$ overlapping blocks in window.
- Can use other values for window, cell, block and step sizes.

HoG: Implementation Details

- Local gradient orientation histograms for cells
 - Quantize gradient angles into bins: 9 bins (unsigned) or 18 bins (signed)
 - Histogram is formed by weighted vote of individual edge pixels
 - Gradient magnitude used as weight and vote is split between two closest bins based on distance to bin center.
 - e.g., for an edge with gradient angle = 85 degrees, add $\frac{1}{4}$ of its magnitude to the bin centered at 70 degrees, and add $\frac{3}{4}$ of its magnitude to the bin centered at 90 degrees.

HoG: Implementation Details

- Normalization over blocks
 - Group of 4 cells form a block.
 - Concatenate histograms from the 4 cells to form a long vector.
 - For 9 bin (unsigned) histograms, dimension of vector = $9 \times 4 = 36$ for each block.
 - Normalize vector by its L2-norm, L2-hys, L1-norm, or L1-sqrt.
 - L2-norm of a vector is defined as

$$|V| = \sqrt{v_1^2 + v_2^2 + v_3^2 + \dots + v_n^2}$$

- Concatenate vectors from all blocks to form final descriptor
 - Dimension of descriptor = $36 \times 105 \text{ blocks} = 3,780$.
- A SVM (support vector machine) or other types of classifier can be trained to classify detection windows into *contain human* or *does not contain human*.