Histograms of Oriented Gradient and its Application

Presenter: Seung Hwan, Bae
- Histograms of oriented gradient
- Author: Navneet Dalal and Triggs (INRIA Grenoble)
- CVPR, 2005
- Total citation: 8653
Generative vs. Discriminative Models

- **Generative:**
  - (+) possibly interpretable
  - (+) models the object class/ can draw samples
  - (-) model variability unimportant to classification task
  - (-) often hard to build good model with few variability

- **Discriminative**
  - (+) appealing when infeasible to model data itself
  - (+) currently often excel in practice
  - (-) often can’t provide uncertainty in predictions
  - (-) non=interpretable
Global vs. Part-based

- We distinguish global people detectors and part-based detector

- Global approaches:
  - A single feature description for the complete object
Global vs. Part-based

- Part-based approaches:
  - Individual feature descriptors for body parts/local parts
Advantages and Disadvantages

- **Global:**
  - Typically simple, i.e., we train a discriminative classifier on the top of the feature descriptions
  - Work well for small resolutions
  - Typically does detection via classification, i.e. uses a binary classification

- **Part-based**
  - Be better able to deal with moving body parts
  - Be able to handle occlusions, overlaps
  - Requires more complex reasoning
Detection via classification: Main idea

- Basic components: a binary classifier

Car/non-car Classifier

Yes, car.

No, not a car.
Detection via classification: Main idea

- If object may be in a cluttered scene, slide a window around looking for it.

Car/non-car Classifier
Gradient Histograms
Gradient Histograms

- Have become extremely popular and successful in the vision community

- Avoid hard decisions compared to edge based features

- Invariant to small shifts

Examples:
- SIFT (Scale-Invariant Image Transform)
- GLOH (Gradient Location and Orientation Histogram)
- HOG (Histogram of Oriented Gradients)
Computing gradients

- **One sided:**
  \[ f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \]

- **Two sided:**
  \[ f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x - h)}{2h} \]

- **Filter masks in x-direction**
  - One sided: 
    \[ \begin{bmatrix} -1 & 1 \end{bmatrix} \]
  - Two sided: 
    \[ \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \]

- **Gradient:**
  - Magnitude: 
    \[ s = \sqrt{s_x^2 + s_y^2} \]
  - Orientation: 
    \[ \theta = \arctan\left(\frac{s_y}{s_x}\right) \]
Gradient histogram measure the orientation and strengths of image gradient within an image region
Histograms of Oriented Gradients

- Gradient-based feature descriptor developed for people detection

- Global descriptor for the complete body

- Very high-dimensional
  - Typically ~ 4000 dimensions
Overall Process of HOG

1. Compute gradients on an image region of 64x128 pixels

2. Compute histograms on ‘cells’ of typically 8x8 pixels (i.e. 8x16 cells)

3. Normalize histograms within overlapping blocks of cells (typically 2x2 cells, i.e. 7x15 blocks)

4. Concatenate histograms
Gradient Computation

- Convolution with [-1 0 1] filters and its transpose
- No smoothing
- Compute gradient magnitude and direction
- Per pixel: color channel with greatest magnitude -> final gradient
Cell Histograms

- A cell: KxK (pixels), usually K = 8
- 9 bins for gradient orientations
  - (0 – 180 degrees/ or 0 – 360 degrees)
- Filed with magnitude

- Interpolated trilinearly
  - Linearly into orientation bins
  - Bilnearily into spatial cells
Linear and Bilinear Interpolation for Subsampling

- **Linear**

\[
f(x) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} (x - x_1)
\]

\[
f(x) = f(x_1) + \frac{f(x_2) - f(x_1)}{d_1 + d_2} d_i
\]

\[
f(x) = \frac{d_2}{d_1 + d_2} f(x_1) + \frac{d_1}{d_1 + d_2} f(x_2)
\]

- **Bilinear**

\[
f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}), \quad R=(x, y_1)
\]

\[
f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}), \quad R=(x, y_2)
\]

\[
f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2)
\]
Histogram Interpolation (Soft Binnig)

- $\theta = 85$ degrees
- Distance to bin centers
  - Bin 70 $\rightarrow$ 15 degree
  - Bin 90 $\rightarrow$ 5 degree
- Ratios: $5/20 = 1/4$, $15/20 = 3/4$

- Distance to bin centers
  - Left: 2, Right: 6
  - Top: 2, Bottom: 6
- Ratio Left-Right: $6/8, 2/8$
- Ratio Top-Bottom: $6/8, 2/8$
- Ratios:
  - $6/8 * 6/8 = 36/64 = 9/16$
  - $6/8 * 2/8 = 12/64 = 3/16$
  - $2/8 * 6/8 = 12/64 = 3/16$
  - $2/8 * 2/8 = 4/64 = 1/16$
Blocks

- A block: 2x2 cells
- Histograms of overlapped cells between blocks are normalized
- This normalization improves invariance to the bias
- Normalization
  - Different norms possible
    - (L1-norm, L1-sqrt, L2-norm, L2-Hys)
Dalal and Trigss use 4 normalization factors for normalizing a cell histogram $C(i, j)$:

$$N_{(\delta, \gamma)} = \left( \|C(i, j)\| \right)^2 + \left( \|C(i + \delta, j)\| \right)^2$$

$$+ \left( \|C(i, j + \gamma)\| \right)^2 + \left( \|C(i + \delta, j + \gamma)\| \right)^2 , \text{where} \quad \delta, \gamma \in \{-1, +1\}$$

HOG feature map is obtained by concatenating normalization results:

$$H(i, j) = \left[ T_\alpha \left( \frac{C(i, j)}{N_{-1,+1}(i, j)} \right), T_\alpha \left( \frac{C(i, j)}{N_{+1,-1}(i, j)} \right), T_\alpha \left( \frac{C(i, j)}{N_{+1,+1}(i, j)} \right), T_\alpha \left( \frac{C(i, j)}{N_{-1,-1}(i, j)} \right) \right], \text{where} \quad T_\alpha(v) : \text{Trunction Function with} \ \alpha$$
Normalization

- Given 5x7 cells
Final Descriptors

- Concatenation of Blocks

- Visualization
Pedestrian Detection using HOG+SVM

- Very promising results on challenging data sets

- Phases
  - Learning phase
  - Detection phase
Pedestrian Detection using HOG+SVM

Training Images

HOG Feature Extraction

SVM Classifier Learning

A Training Phase

Query Image

HOG pyramid

Confidence Map Building & Non-Maxima Suppression

Detection Results

A Testing Phase
Pedestrian Detection using HOG+SVM
Conclusion

- **Contributions**
  - Robust feature encoding for object detection
  - Gives good performance for variety of object classes
  - Real time detection is possible

- **Good works based on HOG**
  - Part based detector for handling partial occlusion
    - Object detection with discriminatively trained part based models, F. Felzenszwalb et. al., TPAMI, 2010, (total citation 2001)
  - Incorporating HOG and texture into the detection framework
    - An HOG-LBP human detector with partial occlusions handling, X. Wang et. al., CVPR, 2009, (citation: 580)
Thanks You

Q & A