Histograms of Oriented Gradients for Human Detection

Praneeth A S
Anuroop K

IIT Jodhpur

November 3, 2014

Author

Navneet Dalal and Bill Triggs

Abstract

- Review existing edge and gradient based descriptors

Abstract

- Review existing edge and gradient based descriptors
- Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection

Abstract

- Review existing edge and gradient based descriptors
- Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection
- Study the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks

MIT Pedestrian Dataset

- 509 training and 200 test images

MIT Pedestrian Dataset

- 509 training and 200 test images
- Limited range of poses

INRIA Dataset

- 1805 64×128 images of humans

INRIA Dataset

- 1805 64×128 images of humans
- No bias on pose

INRIA Dataset

- 1805 64×128 images of humans
- No bias on pose
- Positive Training Samples 1239*2 : Left-Right Reflections

INRIA Dataset

- 1805 64×128 images of humans
- No bias on pose
- Positive Training Samples 1239*2 : Left-Right Reflections
- Negative(Person free Item Sets) 12180 patches sampled from 1218 person-free photos(hard examples: False Positives)

INRIA Dataset

- 1805 64×128 images of humans
- No bias on pose
- Positive Training Samples 1239*2 : Left-Right Reflections
- Negative(Person free Item Sets) 12180 patches sampled from 1218 person-free photos(hard examples: False Positives)
- Diverse environments, lighting conditions and large range of poses and backgrounds.

Method

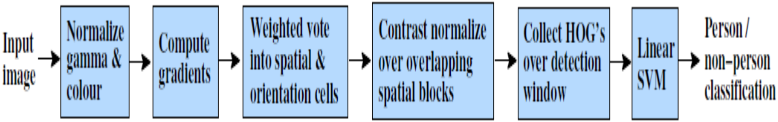

Figure 1: Method for human detection using HOG Descriptors

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients
- Divide the image into small spacial regions called "cells".

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients
- Divide the image into small spacial regions called "cells".
- Each cell accumulates a local 1-D histogram of gradient directions

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients
- Divide the image into small spacial regions called "cells".
- Each cell accumulates a local 1-D histogram of gradient directions
- Combined histograms of all cells form the representation

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients
- Divide the image into small spacial regions called "cells".
- Each cell accumulates a local 1-D histogram of gradient directions
- Combined histograms of all cells form the representation
- 50% Overlapping cells

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients
- Divide the image into small spacial regions called "cells".
- Each cell accumulates a local 1-D histogram of gradient directions
- Combined histograms of all cells form the representation
- 50% Overlapping cells
- HOG - a measure of local histogram "energy" over larger spatial regions ("blocks") and using the results to normalize all of the cells in the block.

Method

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients
- Divide the image into small spacial regions called "cells".
- Each cell accumulates a local 1-D histogram of gradient directions
- Combined histograms of all cells form the representation
- 50% Overlapping cells
- HOG - a measure of local histogram "energy" over larger spatial regions ("blocks") and using the results to normalize all of the cells in the block.
- Train the HOG using linear SVM for human/non-human classification

Advantages

- Captures edge or gradient structure that is very characteristic of local shape

Advantages

- Captures edge or gradient structure that is very characteristic of local shape
- Invariance - Translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size

Advantages

- Captures edge or gradient structure that is very characteristic of local shape
- Invariance - Translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size
- Permits limbs and body segments to change appearance and move from side to side provided they maintain an upright orientation

Error Plots

- Detection Error Tradeoff (DET) $\frac{\text{FalseNeg}}{\text{TruePos+FalseNeg}}$ vs FPPW(False Positivies per Window)

Error Plots

- Detection Error Tradeoff (DET) $\frac{\text{FalseNeg}}{\text{TruePos}+\text{FalseNeg}}$ vs FPPW(False Positivies per Window)
- Much like ROC - Receiver Operating Characteristics

Algorithm for HOG

Algorithm 1 HOG Descriptors

Input: Set of Images of size 128×64
Output: HOG Descriptors for each image

1. Divide the Image window of size 128×64 into 8×8 blocks // Total 15×7 blocks
2. Calculate Gradient Histograms for every block
3. Collect 2×2 cells and normalize the histograms obtained above. // 4×9 histograms per cell // Normalization
4. Concatenate Histograms // Total vector size $= 15 \times 7 \times 2 \times 2 \times 9 = 3780$

Algorithm for histogram calculation

Algorithm 2 Histgram Calculation

Input: Image block of size 8×8
Output: Histogram of gradients

1. Take number of bins $= 9$
2. Divide $0 - 180^{\circ}$ into 9 bins $0 - 20, 20 - 40, \ldots, 160 - 180$ // May use any voting methods for giving weight to different orientations

Gradient variation

- $[-1, 0, 1]$ - Centered

Gradient variation

- $[-1, 0, 1]$ - Centered
- $[-1, 1]$ - Uncentered

Gradient variation

- $[-1, 0, 1]$ - Centered
- $[-1, 1]$ - Uncentered
- $[1, -8, 0, 8, -1]$ - Cubic centered

Orientation Binning

- Change in number of bins: 9, 8

Orientation Binning

- Change in number of bins: 9, 8
- The orientation bins are evenly spaced over $0 - 180^{\circ}$ ("unsigned" gradient) or $0 - 360^{\circ}$ ("signed" gradient).

Normalization

- L2-Norm $v \rightarrow \frac{v}{\sqrt{\|v\|^2 + \epsilon^2}}$

Normalization

- L2-Norm $v \rightarrow \dfrac{v}{\sqrt{\|v\|^2 + \epsilon^2}}$
- L1-Norm $v \rightarrow \dfrac{v}{\sqrt{\|v\| + \epsilon}}$

Normalization

- L2-Norm $v \rightarrow \frac{v}{\sqrt{\|v\|^2 + \epsilon^2}}$
- L1-Norm $v \rightarrow \frac{v}{\sqrt{\|v\| + \epsilon}}$
- ϵ to ensure denominator is not zero

SIFT(Scale Invariant Feature Transform)

- The scale invariant feature transform (SIFT) algorithm is used to detect and describe local features in images
- Image features are transformed into local feature coordinates that are invariant to image translation, scaling, rotation and partially invariant to illumination changes
- To use the SIFT operator for object recognition purposes, it is applied on two object images, a model and a test image
- Calculation of SIFT image features is performed through the four consecutive steps

SIFT Algorithm

- **Scale-space extrema detection**
  - Extract scale and rotation invariant interest points (i.e., keypoints).
- **Keypoint localization**
  - Determine location and scale for each interest point.
  - Eliminate "weak" keypoints
- **Orientation assignment**
  - Assign one or more orientations to each keypoint.
- **Keypoint descriptor**
  - Use local image gradients at the selected scale

SIFT Features Matching

- To find corresponding features between the two images, which will lead to object recognition, different feature matching approaches can be used.
- According to the Nearest Neighbourhood procedure for each F_1^i feature in the model image feature set the corresponding feature F_2^j must be looked for in the test image feature set.
- The corresponding feature is one with the smallest Euclidean distance to the feature F_1^i
- A pair of corresponding feature $(F_1^i,\ F_2^j)$ called a match $M(F_1^i, F_2^j)$

Thank You