

EEE422 (EEE6082) Computational Vision

Histograms of Oriented Gradients

Ling Shao

1

Global vs. Part-Based

- We distinguish global people detectors and part-based detectors
- Global approaches:
 - A single feature description for the complete person
- Part-Based Approaches:
 - Individual feature descriptors for body parts / local parts

2

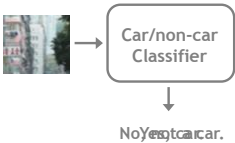
Advantages and Disadvantages

- Part-Based
 - May be better able to deal with moving body parts
 - May be able to handle occlusion, overlaps
 - Requires more complex reasoning
- Global approaches
 - Typically simple, i.e. we train a discriminative classifier on top of the feature descriptions
 - Work well for small resolutions
 - Typically does detection via classification, i.e. uses a binary classifier

3

Detection via classification: Main idea

Basic component: a binary classifier

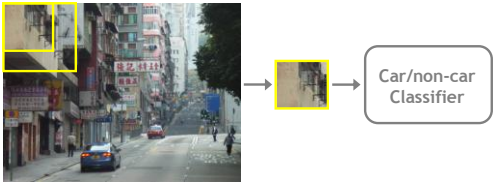


Slide credit: K. Grauman, B. Leibe

4

Detection via classification: Main idea

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



Slide credit: K. Grauman, B. Leibe

5

Gradient Histograms

6

Gradient Histograms

- Have become extremely popular and successful in the vision community
- Avoid hard decisions compared to edge based features
- Examples:
 - SIFT (Scale-Invariant Image Transform)
 - GLOH (Gradient Location and Orientation Histogram)
 - HOG (Histogram of Oriented Gradients)

7

Computing gradients

- One sided: $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$
- Two sided: $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x-h)}{2h}$
- Filter masks in x-direction
 - One sided:

-1	1
----	---
 - Two sided:

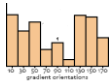
-1	0	1
----	---	---
- Gradient:
 - Magnitude: $s = \sqrt{s_x^2 + s_y^2}$
 - Orientation: $\theta = \arctan(\frac{s_y}{s_x})$



8

Histograms

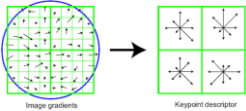
- Gradient histograms measure the orientations and strengths of image gradients within an image region



9

Example: SIFT descriptor

- The most popular gradient-based descriptor
- Typically used in combination with an interest point detector



- Region rescaled to a grid of 16x16 pixels
- 4x4 regions = 16 histograms (concatenated)
- Histograms: 8 orientation bins, gradients weighted by gradient magnitude
- Final descriptor has 128 dimensions and is normalized to compensate for illumination differences

10

Application: AutoPano-Sift

Sift matches



Blended image



Other applications:
- Recognition of previously seen objects (e.g. in robotics)

11

Histograms of Oriented Gradients

- Gradient-based feature descriptor developed for people detection
 - Authors: Dalal&Triggs (INRIA Grenoble, F)
- Global descriptor for the complete body
- Very high-dimensional
 - Typically ~4000 dimensions

12

HOG

Very promising results on challenging data sets



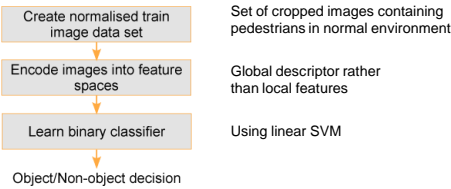
Phases

- 1. Learning Phase
- 2. Detection Phase

13

Detector: Learning Phase

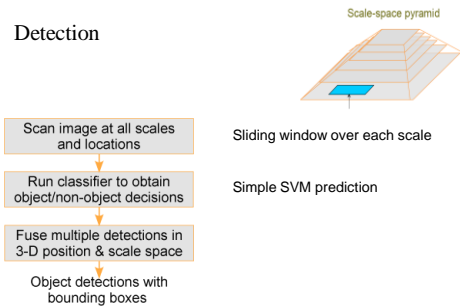
1. Learning



14

Detector: Detection Phase

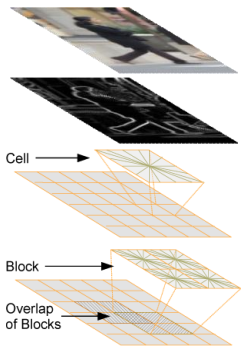
2. Detection



15

Descriptor

- 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



16

Gradients

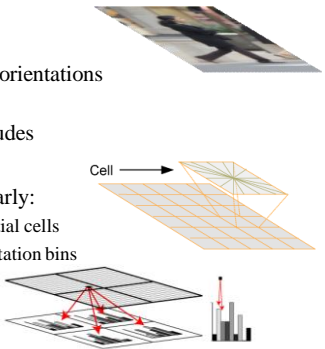
- Convolution with $[-1 \ 0 \ 1]$ filters
- No smoothing
- Compute gradient magnitude+direction
- Per pixel: color channel with greatest magnitude - > final gradient



17

Cell histograms

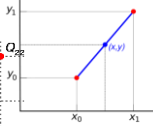
- 9 bins for gradient orientations (0-180 degrees)
- Filled with magnitudes
- Interpolated trilinearly:
 - Bilinearly into spatial cells
 - Linearly into orientation bins



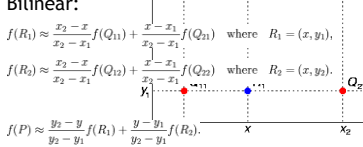
18

Linear and Bilinear interpolation for subsampling

Linear:

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0}$$


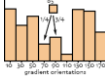
Bilinear:

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \text{ where } R_1 = (x, y_1)$$
$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \text{ where } R_2 = (x, y_2)$$
$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2)$$


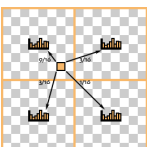
19

Histogram interpolation example

- 0=85 degrees
- Distance to bin centers
 - Bin 70 -> 15 degrees
 - Bin 90 -> 5 degrees
- Ratios: 5/20=1/4, 15/20=3/4



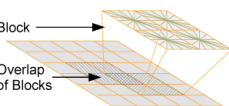

- Distance to bin centers
 - Left: 2, Right: 6
 - Top: 2, Bottom: 6
- Ratio Left-Right: 6/8, 2/8
- Ratio Top-Bottom: 6/8, 2/8
- Ratios:
 - 6/8*6/8 = 36/64 = 9/16
 - 6/8*2/8 = 12/64 = 3/16
 - 2/8*6/8 = 12/64 = 3/16
 - 2/8*2/8 = 4/64 = 1/16



20

Blocks

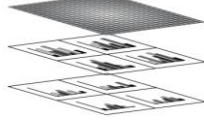
- Overlapping blocks of 2x2 cells
- Cell histograms are concatenated and then normalized
 - Note that each cell several occurrences with different normalization in final descriptor
- Normalization
 - Different norms possible (L2, L2hys etc.)
 - We add a normalization epsilon to avoid division by zero



21

Blocks

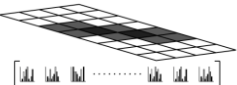
- Gradient magnitudes are weighted according to a Gaussian spatial window
- Distant gradients contribute less to the histogram




22

Final Descriptor

- Concatenation of Blocks



- Visualization:



23

Engineering

- Developing a feature descriptor requires a lot of engineering
 - Testing of parameters (e.g. size of cells, blocks, number of cells in a block, size of overlap)
 - Normalization schemes (e.g. L1, L2-Norms etc., gamma correction, pixel intensity normalization)
- An extensive evaluation of different choices was performed, when the descriptor was proposed
- It's not only the idea, but also the engineering effort

24

Training Set

- More than 2000 positive & 2000 negative training images (96x160px)
- Carefully aligned and resized
- Wide variety of backgrounds



25

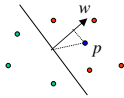
Model learning

- Simple linear SVM on top of the HOG Features
 - Fast (one inner product per evaluation window)
 - Hyper plane normal vector:

$$w = \sum \alpha_i y_i x_i \text{ with } y_i \text{ in } \{0,1\} \text{ and } x_i \text{ the support vectors}$$

$$f(p) = \sum \alpha_i y_i \langle p, x_i \rangle = p^T w$$

- Decision: $sign(p^T w)$

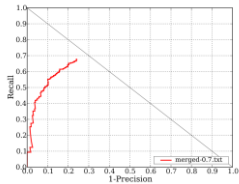


- Slightly better results can be achieved by using a SVM with a Gaussian kernel
 - But considerable increase in computation time

26

Result on INRIA database

- Test Set contains 287 images
- Resolution ~640x480
- 589 persons
- Avg. size: 288 pixels



Demo



28