Window-based models for generic object detection

May 21st, 2019

Yong Jae Lee
UC Davis

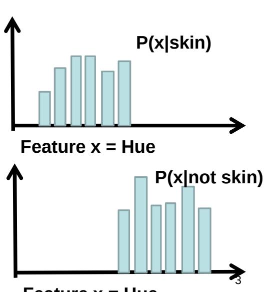
Previously

- Intro to generic object recognition
- Supervised classification
  - Main idea
  - Skin color detection example

2

Last time:
Example: skin color classification

- We can represent a class-conditional density using a histogram (a "non-parametric" distribution)

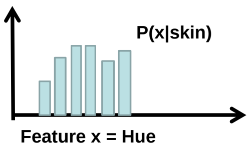


3

Kristen Grauman

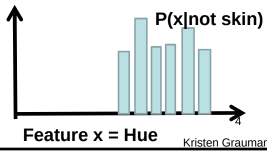
Last time: Example: skin color classification

- We can represent a class-conditional density using a histogram (a "non-parametric" distribution)



Now we get a new image, and want to label each pixel as skin or non-skin.

$P(skin \mid x) \,\alpha\, P(x \mid skin) P(skin)$



4

Kristen Grauman

Last time: Example: skin color classification

Now for every pixel in a new image, we can estimate probability that it is generated by skin.

Brighter pixels → higher probability of being skin

Classify pixels based on these probabilities

- if $p(\text{skin}|\boldsymbol{x}) > \theta$, classify as skin
- if $p(\text{skin}|\boldsymbol{x}) < \theta$, classify as not skin

5

Kristen Grauman

Today

- Window-based generic object detection
  – basic pipeline
  – boosting classifiers
  – face detection as case study

6

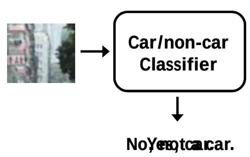
Generic category recognition: basic framework

- Build/train object model
  - Choose a representation
  - Learn or fit parameters of model / classifier
- Generate candidates in new image
- Score the candidates

7

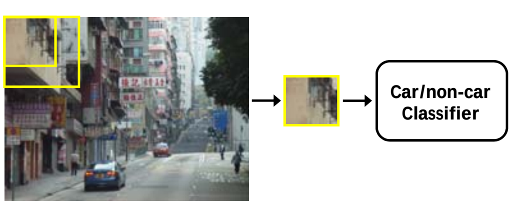
Window-based models
Building an object model

Given the representation, train a binary classifier



8
Kristen Grauman

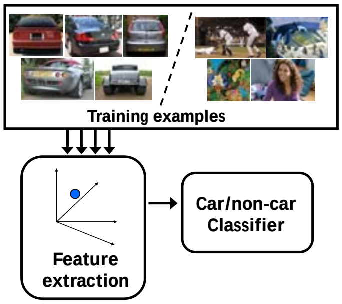
Window-based models
Generating and scoring candidates



9
Kristen Grauman

Window-based object detection: recap

Training:
1. Obtain training data
2. Define features
3. Define classifier

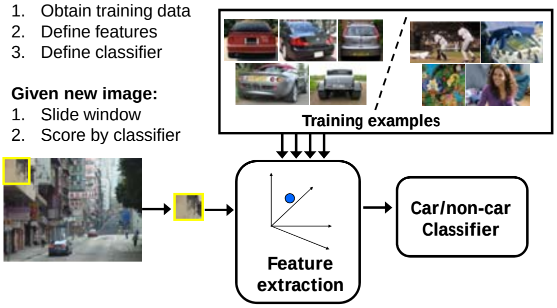


10
Kristen Grauman

Window-based object detection: recap

Training:
1. Obtain training data
2. Define features
3. Define classifier

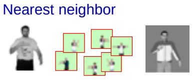
Given new image:
1. Slide window
2. Score by classifier



11
Kristen Grauman

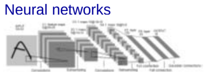
Discriminative classifier construction

Nearest neighbor



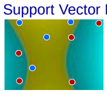
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005...

Neural networks



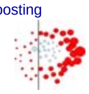
LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998
…

Support Vector Machines



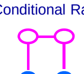
Guyon, Vapnik
Heisele, Serre, Poggio, 2001,…

Boosting



Viola, Jones 2001, Torralba et al. 2004, Opelt et al. 2006,…

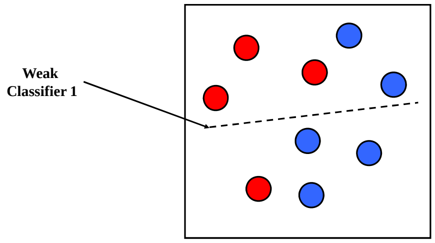
Conditional Random Fields



McCallum, Freitag, Pereira 2000; Kumar, Hebert 2003
…

12
Slide adapted from Antonio Torralba

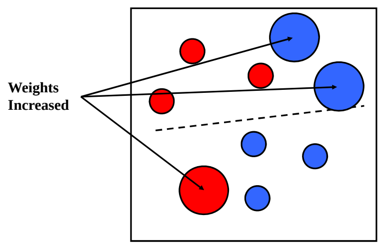
Boosting intuition



13

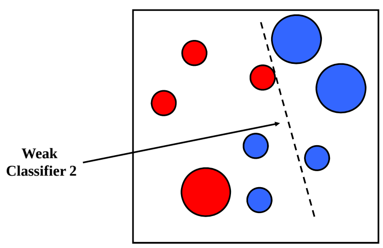
Slide credit: Paul Viola

Boosting illustration



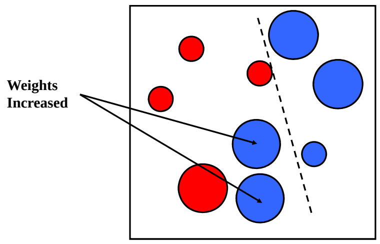
14

Boosting illustration



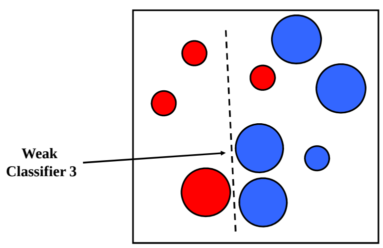
15

Boosting illustration



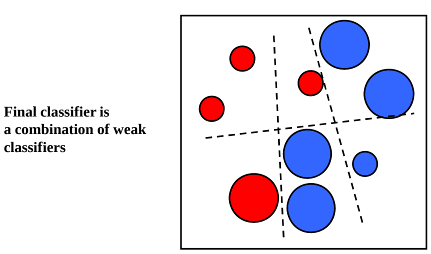
16

Boosting illustration



17

Boosting illustration



18

Boosting: training

- Initially, weight each training example equally
- In each boosting round:
  - Find the weak learner that achieves the lowest *weighted* training error
  - Raise weights of training examples misclassified by current weak learner
- Compute final classifier as linear combination of all weak learners (weight of each learner is directly proportional to its accuracy)
- Exact formulas for re-weighting and combining weak learners depend on the particular boosting scheme (e.g., AdaBoost)

19

Slide credit: Lana Lazebnik

Viola-Jones face detector

ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001

Rapid Object Detection using a Boosted Cascade of Simple Features

Paul Viola
viola@merl.com
Mitsubishi Electric Research Labs
201 Broadway, 8th FL
Cambridge, MA 02139

Michael Jones
mjones@crl.dec.com
Compaq CRL
One Cambridge Center
Cambridge, MA 02142

Abstract

This paper describes a machine learning approach for vi-

tected at 15 frames per second on a conventional 700 MHz Intel Pentium III. In other face detection systems, auxiliary information, such as image differences in video sequences,

20

Viola-Jones face detector

Main idea:

- Represent local texture with efficiently computable "rectangular" features within window of interest
- Select discriminative features to be weak classifiers
- Use boosted combination of them as final classifier
- Form a cascade of such classifiers, rejecting clear negatives quickly

21

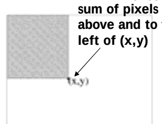
Kristen Grauman

Viola-Jones detector: features

"Rectangular" filters

Feature output is difference between adjacent regions

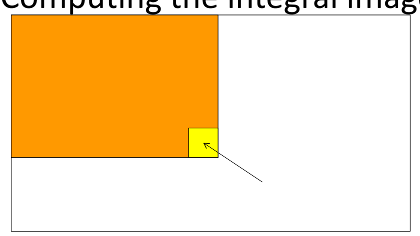
Efficiently computable with integral image: any sum can be computed in constant time.



22

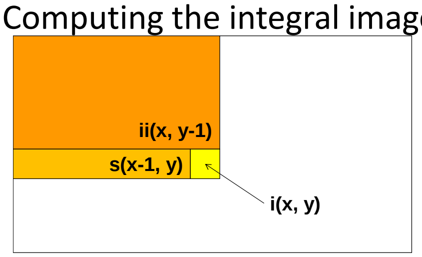
Kristen Grauman

Computing the integral image



Lana Lazebnik

Computing the integral image

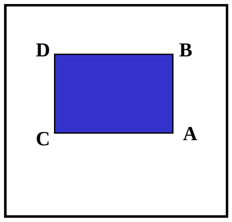


- Cumulative row sum: s(x, y) = s(x–1, y) + i(x, y)
- Integral image: ii(x, y) = ii(x, y−1) + s(x, y)

Lana Lazebnik

Computing sum within a rectangle

- Let A,B,C,D be the values of the integral image at the corners of a rectangle
- Then the sum of original image values within the rectangle can be computed as:
  sum = A – B – C + D
- Only 3 additions are required for any size of rectangle!

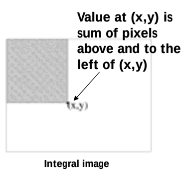


25
Lana Lazebnik

Viola-Jones detector: features

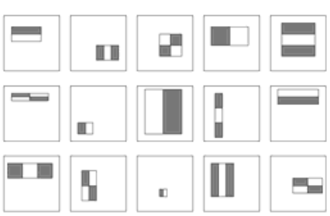
"Rectangular" filters
Feature output is difference between adjacent regions

Efficiently computable with integral image: any sum can be computed in constant time



26
Kristen Grauman

Viola-Jones detector: features



Considering all possible filter parameters: position, scale, and type:

180,000+ possible features associated with each 24 x 24 window

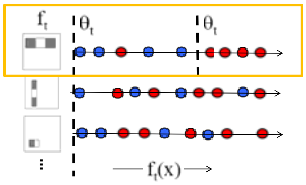
Which subset of these features should we use to determine if a window has a face?

Use AdaBoost both to select the informative features and to form the classifier

27
Kristen Grauman

Viola-Jones detector: AdaBoost

- **Want to select the single rectangle feature and threshold that best separates positive (faces) and negative (non-faces) training examples, in terms of *weighted* error.**



Outputs of a possible rectangle feature on faces and non-faces.

Resulting weak classifier:

$$h_t(x) = \begin{cases} +1 & \text{if } f_t(x) > \theta_t \\ -1 & \text{otherwise} \end{cases}$$

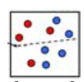
For next round, reweight the examples according to errors, choose another filter/threshold combo.

28
Kristen Grauman

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.

AdaBoost Algorithm

Start with uniform weights on training examples



For T rounds

Evaluate *weighted* error for each feature, pick best.

Re-weight the examples:
Incorrectly classified -> more weight
Correctly classified -> less weight

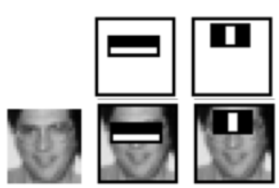
Final classifier is combination of the weak ones, weighted according to error they had.

Freund & Schapire 1995

29

Visual Object Recognition Tutorial

Viola-Jones Face Detector: Results

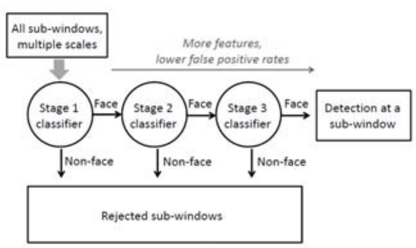


First two features selected

30

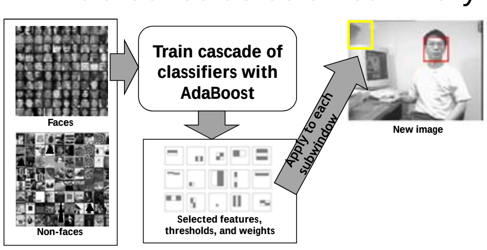
- Even if the filters are fast to compute, each new image has a lot of possible windows to search.
- How to make the detection more efficient?

Cascading classifiers for detection



- Form a *cascade* with low false negative (high recall) rates early on
- Apply less accurate but faster classifiers first to immediately discard windows that clearly appear to be negative

Viola-Jones detector: summary



Train with 5K positives, 350M negatives
Real-time detector using 38 layer cascade
6061 features in all layers

[Implementation available in OpenCV: http://www.intel.com/technology/computing/opencv/]

Kristen Grauman

Viola-Jones detector: summary

- A seminal approach to real-time object detection
- Training is slow, but detection is very fast
- Key ideas
  - *Integral images* for fast feature evaluation
  - *Boosting* for feature selection
  - *Attentional cascade* of classifiers for fast rejection of non-face windows

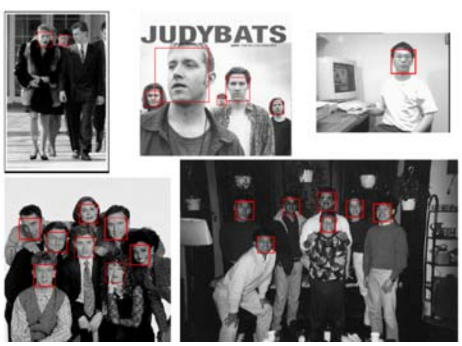
P. Viola and M. Jones. *Rapid object detection using a boosted cascade of simple features.* CVPR 2001.

P. Viola and M. Jones. *Robust real-time face detection.* IJCV 57(2), 2004.

34

Visual Object Recognition Tutorial

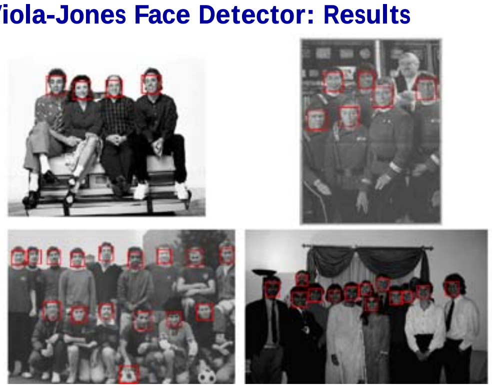
Viola-Jones Face Detector: Results



35

Visual Object Recognition Tutorial

Viola-Jones Face Detector: Results



36

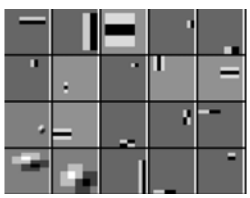
Viola-Jones Face Detector: Results

Visual Object Recognition Tutorial

37

Detecting profile faces?

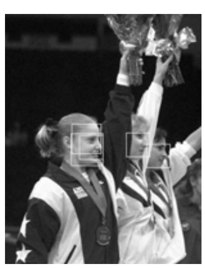
Can we use the same detector?



Visual Object Recognition Tutorial

38

Viola-Jones Face Detector: Results

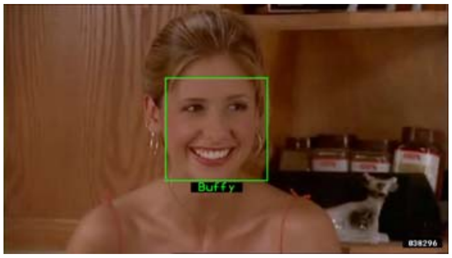


Visual Object Recognition Tutorial

Paul

39

Example using Viola-Jones detector



Frontal faces detected and then tracked, character names inferred with alignment of script and subtitles.

Everingham, M., Sivic, J. and Zisserman, A.
"Hello! My name is... Buffy" - Automatic naming of characters in TV video, BMVC 2006. http://www.robots.ox.ac.uk/~vgg/research/nface/index.html

40


ZDNet Asia

Where Technology Means Business

See how he stays with Cisco Collab Solutions

Home | News | Insight | Reviews | TechGuides | Jobs | Blogs | Videos | Community | Downloads | IT Library

News > Internet

Google now erases faces, license plates on Map Street View

Friday, August 24, 2007 01:37 PM

Google has gotten a lot of flack from privacy advocates for photographing faces and license plate numbers and displaying them on the Street View in Google Maps. Originally, the company said only people who identified themselves could ask the company to remove their image.

But Google has quietly changed that policy, partly in response to criticism, and now anyone can alert the company and have an image of a license plate or a recognizable face removed, not just the owner of the face or car, says Marissa Mayer, vice president of search products and user experience at Google.

"It's a good policy for users and also clarifies the intent of the product," she said in an interview following her keynote at the Search Engine Strategies conference in San Jose, Calif., Wednesday.

The policy change was made about 10 days after the launch of the product in late May, but was not publicly announced, according to Mayer. The company is removing images only when someone notifies them and not proactively, she said. "It was definitely a big policy change inside."

News from Countries/Region

What's Hot | Latest News

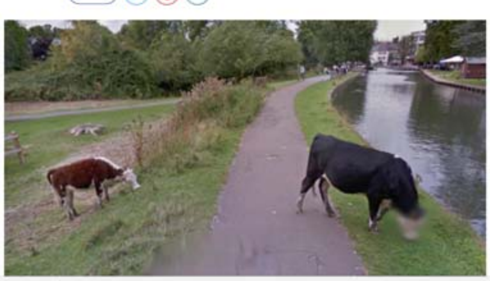
Cisco Collaboration Solut

41

Technology

Google street view blurs face of cow to protect its identity

share



Slide: Kristen Grauman

Consumer application: iPhoto 2009

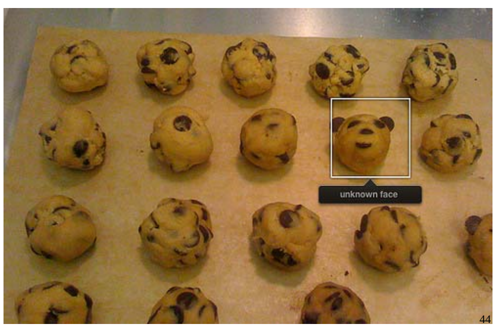
http://www.apple.com/ilife/iphoto/

43

Slide credit: Lana Lazebnik

Consumer application: iPhoto 2009

Things iPhoto thinks are faces

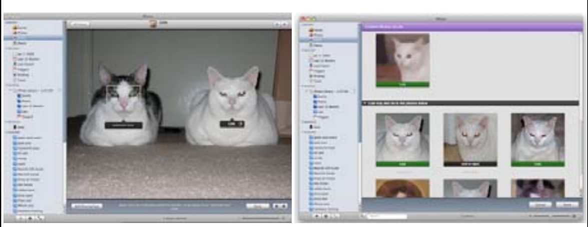


44

Slide credit: Lana Lazebnik

Consumer application: iPhoto 2009

Can be trained to recognize pets!

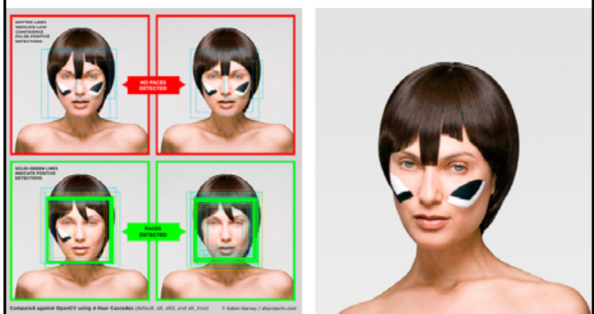


http://www.maclife.com/article/news/iphotos_faces_recognizes_cats

45

Slide credit: Lana Lazebnik

Privacy Gift Shop – CV Dazzle



http://www.wired.com/2015/06/facebook-can-recognize-even-dont-show-face/
Wired, June 15, 2015

Slide: Kristen Grauman

Privacy Visor

http://www.3ders.org/articles/20150812-japan-3d-printed-privacy-visors-will-block-facial-recognition-software.html

Slide: Kristen Grauman

Boosting: pros and cons

- Advantages of boosting
  - Integrates classification with feature selection
  - Flexibility in the choice of weak learners, boosting scheme
  - Testing is fast
  - Easy to implement
- Disadvantages
  - Needs many training examples
  - Often found not to work as well as an alternative discriminative classifier, support vector machine (SVM)
    - especially for many-class problems

48

Slide credit: Lana Lazebnik

What other categories are amenable to *window-based representation*?

49

Visual Object Recognition Tutorial

Pedestrian detection

- Detecting upright, walking humans also possible using sliding window's appearance/texture; e.g.,

SVM with Haar wavelets [Papageorgiou & Poggio, IJCV 2000]

Space-time rectangle features [Viola, Jones & Snow, ICCV 2003]

SVM with HoGs [Dalal & Triggs, CVPR 2005]

50
Kristen Grauman

Visual Object Recognition Tutorial

Window-based detection: strengths

- Sliding window detection and global appearance descriptors:
  - Simple detection protocol to implement
  - Good feature choices critical
  - Past successes for certain classes

51
Kristen Grauman

Window-based detection: Limitations

- High computational complexity
  - For example: 250,000 locations x 30 orientations x 4 scales = 30,000,000 evaluations!
  - If training binary detectors independently, means cost increases linearly with number of classes
- With so many windows, false positive rate better be low

Visual Object Recognition Tutorial

52
Kristen Grauman

Limitations (continued)

- Not all objects are “box” shaped

Visual Object Recognition Tutorial

53
Kristen Grauman

Limitations (continued)

- If considering windows in isolation, context is lost

Sliding window

Detector’s view

Visual Object Recognition Tutorial

Figure credit: Derek Hoiem

54

Visual Object Recognition Tutorial

Limitations (continued)

- In practice, often entails large, cropped training set (expensive)
- Requiring good match to a global appearance description can lead to sensitivity to partial occlusions

Image credit: Adam, Rivlin, & Shimshoni

55

Kristen Grauman

Summary

- Basic pipeline for window-based detection
  - Model/representation/classifier choice
  - Sliding window and classifier scoring
- Boosting classifiers: general idea
- Viola-Jones face detector
  - Exemplar of basic paradigm
  - Plus key ideas: rectangular features, Adaboost for feature selection, cascade
- Pros and cons of window-based detection

56

Questions?

See you Thursday!

57