Lecture 12

Visual recognition

• An introduction to recognition
• Image classification – the bag of words model

[FP] – Chapters 6 (sec. 6.2)
[FP] – Chapters 16 (sec. 16.1)
What we have seen so far

- Feature Detection
  - e.g. DoG

- Feature Description
  - e.g. SIFT
  - Estimation
  - Matching
  - Indexing
  - Recognition
What’s visual recognition?
Classification:
Does this image contain a building? [yes/no]

Yes!
Classification:
Is this an beach?

No!
Image Search or Indexing

Organizing photo collections
Detection:
Does this image contain a car? [where?]
Detection:
Which object does this image contain? [where?]
Detection:
Accurate localization (segmentation)
Object detection is useful...

- Computational photography
- Assistive technologies
- Surveillance
- Security
- Assistive driving
Categorization vs Single instance recognition

Which building is this? *Marshall Field* building in Chicago
Categorization vs Single instance recognition

Where is the crunchy nut?
Recognizing landmarks in mobile platforms
Detection: Estimating object semantic & geometric attributes

Object: Building, 45º pose, 8-10 meters away
   It has bricks

Object: Person, back; 1-2 meters away

Object: Police car, side view, 4-5 m away
Activity or Event recognition
What are these people doing?
Visual Recognition

• Design algorithms that are capable to
  – Classify images or videos
  – Detect and localize objects
  – Estimate semantic and geometrical attributes
  – Classify human activities and events

Why is this challenging?
How many object categories are there?

\( \approx 10,000 \text{ to } 30,000 \)
Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba
Challenges: illumination

image credit: J. Koenderink
Challenges: scale
Challenges: deformation
Challenges: occlusion

Magritte, 1957
Challenges: background clutter

Kilmeny Niland. 1995
Challenges: intra-class variation
Basic properties

• Representation
  – How to represent an object category

• Learning
  – How to learn the classifier, given training data

• Recognition
  – How the classifier is to be used on novel data
Representation

- Building blocks: Sampling strategies

Interest operators

Dense, uniformly

Multiple interest operators

Randomly

Image credits: F.-F. Li, E. Nowak, J. Sivic
Representation

- Building blocks: Choice of descriptors [SIFT, HOG, codewords....]
Representation

- Appearance only
- 2D location and appearance
- 3D location and appearance
Representation

– Invariances
  • View point
  • Illumination
  • Occlusion
  • Scale
  • Deformation
  • Clutter
  • etc.
Representation

– How to handle intra-class variability?
– It is convenient to describe object categories using probabilistic models
Object categorization: the statistical viewpoint

\[ p(\text{zebra} \mid \text{image}) \]

vs.

\[ p(\text{no zebra} \mid \text{image}) \]

- Bayes rule:

\[ p(A \mid B) = \frac{p(B \mid A) \cdot p(A)}{p(B)} \]

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})}
\]
Object categorization:
the statistical viewpoint

\[ p(\text{zebra} \mid \text{image}) \]
vs.

\[ p(\text{no zebra} \mid \text{image}) \]

- Bayes rule:

\[
p(A \mid B) = \frac{p(B \mid A) \cdot p(A)}{p(B)}
\]

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})}
\]

- posterior ratio
- likelihood ratio
- prior ratio
Object categorization: the statistical viewpoint

- Discriminative methods model posterior
- Generative methods model likelihood and prior

Bayes rule:

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})}
\]

- posterior ratio
- likelihood ratio
- prior ratio
Discriminative models

**Nearest neighbor**
- Shakhnarovich, Viola, Darrell 2003
- Berg, Berg, Malik 2005...
- 10^6 examples

**Support Vector Machines**
- Guyon, Vapnik, Heisele, Serre, Poggio...

**Neural networks**
- LeCun, Bottou, Bengio, Haffner 1998
- Rowley, Baluja, Kanade 1998

**Latent SVM**
- Felzenszwalb 00
- Ramanan 03...

**Structural SVM**
- Felzenszwalb 00
- Ramanan 03...

**Boosting**
- Viola, Jones 2001
- Torralba et al. 2004
- Opelt et al. 2006,...
Generative models

• Naïve Bayes classifier
  – Csurka Bray, Dance & Fan, 2004

• Hierarchical Bayesian topic models (e.g. pLSA and LDA)
  – Natural scene categorization: Fei-Fei et al. 2005

• 2D Part based models
  - Constellation models: Weber et al 2000; Fergus et al 2000
  - Star models: ISM (Leibe et al 05)

• 3D part based models:
Basic properties

• Representation
  – How to represent an object category; which classification scheme?

• Learning
  – How to learn the classifier, given training data

• Recognition
  – How the classifier is to be used on novel data
Learning

• Learning parameters

• What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
Learning

• Learning parameters
• What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)

• Level of supervision
  • Noisy labels; image labels; bounding box; manual segmentation; part annotations

• Batch/incremental

• Priors
Learning

• Learning parameters
• What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)

• Level of supervision
  • Noisy labels; image labels; bounding box; manual segmentation; part annotations

• Batch/incremental

• Priors

• Training images:
  • Issue of overfitting
  • Negative images for discriminative methods
Basic properties

• Representation
  – How to represent an object category; which classification scheme?

• Learning
  – How to learn the classifier, given training data

• Recognition
  – How the classifier is to be used on novel data
Recognition

– Recognition task: classification, detection, etc.
Recognition

– Recognition task

– Search strategy: Sliding Windows
  • Simple
  • Computational complexity \((x, y, S, \theta, N \text{ of classes})\)
    - BSW by Lampert et al 08
    - Also, Alexe, et al 10

Viola, Jones 2001,
Recognition

– Recognition task

– Search strategy: Sliding Windows

• Simple

• Computational complexity (x, y, S, θ, N of classes)
  - BSW by Lampert et al 08
  - Also, Alexe, et al 10

• Localization

  • Prone to false positive

  Non max suppression:
  Canny ’86
  …
  Desai et al , 2009

Viola, Jones 2001,
Recognition

– Recognition task

– Search strategy: Sliding Windows
  • Simple
  • Computational complexity (x, y, S, θ, N of classes)
    - BSW by Lampert et al 08
    - Also, Alexe, et al 10
  • Localization
    • Prone to false positive
      Non max suppression: Canny ’86
      ....
      Desai et al , 2009
  • Objects are not boxes

Viola, Jones 2001,
Successful methods using sliding windows

- Subdivide scanning window
- In each cell compute histogram of gradients orientation.


[Dalal & Triggs, CVPR 2005]

- Subdivide scanning window
- In each cell compute histogram of codewords of adjacent segments

Code available: [http://www.vision.ee.ethz.ch/~calvin](http://www.vision.ee.ethz.ch/~calvin)

[Ferrari & al, PAMI 2008]
Recognition

- Recognition task

- Search strategy: Segmentation

  - Bottom up segmentation
    
    Malik et al. 01
    Maire et al. 08

  - Semantic segmentation
    
    Felzenszwalb and Huttenlocher, 2004

    Duygulu et al. 02
Recognition

- Recognition task
- Search strategy
- Attributes

• Savarese, 2007
• Sun et al 2009
• Liebelt et al., '08, 10
• Farhadi et al 09

Category: car
Azimuth = 225°
Zenith = 30°

- It has metal
- it is glossy
- has wheels

• Farhadi et al 09
• Lampert et al 09
• Wang & Forsyth 09
Recognition

– Recognition task
– Search strategy
– Attributes
– Context

Semantic:
• Torralba et al 03
• Rabinovich et al 07
• Gupta & Davis 08
• Heitz & Koller 08
• L-J Li et al 08
• Bang & Fei-Fei 10

Geometric
• Hoiem, et al 06
• Gould et al 09
• Bao, Sun, Savarese 10
Agenda on recognition

• Image classification
  • Bag of words representations

• Object detection
  • 2D object detection
  • 3D object detection

• Scene understanding
Lecture 11
Visual recognition

• An introduction to recognition
• Image classification – the bag of words model
Bag of words models

• Used for image and object classification

• Designed to handle variability due to:
  • View point
  • Illumination
  • Occlusions
  • Intra-class
Basic properties

• Representation
  – How to represent an object category; which classification scheme?

• Learning
  – How to learn the classifier, given training data

• Recognition
  – How the classifier is to be used on novel data
Related works

• Early “bag of words” models: mostly texture recognition

• Hierarchical Bayesian models for documents (pLSA, LDA, etc.)
  – Hoffman 1999; Blei, Ng & Jordan, 2004; Teh, Jordan, Beal & Blei, 2004

• Object categorization
  – Csurka, Bray, Dance & Fan, 2004; Sivic, Russell, Efros, Freeman & Zisserman, 2005; Sudderth, Torralba, Freeman & Willsky, 2005;

• Natural scene categorization
  – Vogel & Schiele, 2004; Fei-Fei & Perona, 2005; Bosch, Zisserman & Munoz, 2006
Object → Bag of ‘words’
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based on the messages that reach our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the brain; the cerebral cortex was a movie screen, so to speak, upon which the image in the eye was projected. Through the discoveries of Hubel and Wiesel we now know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path to the various cell layers of the optical cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a step-wise analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.

China is forecasting a trade surplus of $90bn (£51bn) to $100bn this year, a threefold increase on 2004’s $32bn. The Commerce Ministry said the surplus would be created by a predicted 30% jump in exports to $750bn, compared with a 18% rise in imports to $660bn. This is likely to further annoy the US, which has long argued that China’s exports are unfairly helped by a deliberately undervalued yuan. Beijing agrees the surplus is too high, but says the yuan is only one factor. Bank of China governor Zhou Xiaochuan said the country also needed to do more to boost domestic demand so that more goods stayed within the country. China increased the value of the yuan against the dollar by 2.1% in July, but allowed the trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

Hoffman 1999; Blei, Ng & Jordan, 2004; Teh, Jordan, Beal & Blei, 2004
definition of “BoW”

– Independent features

face

bike

violin
definition of “BoW”

– Independent features
– histogram representation

codewords dictionary
Representation

- feature detection & representation
- image representation

Codewords dictionary

Category models (and/or) classifiers

Learning

Recognition

Category decision
1. Feature detection and description
1. Feature detection and description

• Regular grid
  – Vogel & Schiele, 2003
  – Fei-Fei & Perona, 2005
1. Feature detection and description

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005

- Interest point detector
  - Csurka, et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic, et al. 2005
1. Feature detection and description

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005

- Interest point detector
  - Csurka, et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic, et al. 2005

- Other methods
  - Random sampling (Vidal-Naquet & Ullman, 2002)
  - Segmentation based patches (Barnard, Duygulu, Forsyth, de Freitas, Blei, Jordan, 2003)
1. Feature detection and description

Compute descriptor
E.g. SIFT [Lowe’99]

Detected features

Slide credit: Josef Sivic
2. Codewords dictionary formation
2. Codewords dictionary formation
Example: color feature

- (R=0, G=200, B=20)
- (R=255, G=200, B=250)
- (R=245, G=220, B=248)
- (R=15, G=189, B=2)
- (R=3, G=12, B=2)
Example: color feature
2. Codewords dictionary formation

Cluster center = code word

E.g., K-Means clustering
2. Codewords dictionary formation

Image patch examples of codewords
2. Codewords dictionary formation

Fei-Fei et al. 2005
3. Bag of word representation

- Nearest neighbors assignment
- K-D tree search strategy

Codewords dictionary
3. Bag of word representation

Codewords dictionary

frequency

codewords

Codewords dictionary
Visual vocabularies: Issues

• How to choose vocabulary size?
  – Too small: visual words not representative of the object appearance distribution
  – Too large: quantization artifacts, sparse histograms, overfitting

• Computational efficiency
  – Vocabulary trees
    (Nister & Stewenius, 2006)
Invariance issues

• Scale? Rotation? View point? Occlusions?
  – Implicit
  – Depends on detectors and descriptors
Representation

1. feature detection & representation

2. codewords dictionary

3. category models
Category models
Recognition

codewords dictionary

category models (and/or) classifiers

category decision
Next Lecture

- Object classification – BoW models part 2
- 2D object detection