Announcements

- P2 was released yesterday
- It is now due on May 8\textsuperscript{th} (instead of May 6\textsuperscript{th})
- Please form your team by Wednesday this week;
Recognition

- Classification
- Detection
- Single instance detection and localization

Forsyth, Ponce “Computer vision: a modern approach”:
- Chapter 16, Sec. 16.1
- Chapters 6 (sec. 6.2)

Szeliski, “Computer Vision: algorithms and applications"
- Chapter 14, Sec. 14.1, Sec. 14.3, sec. 14.4

Several slides in this lecture are credit from L. Fei-Fei, R. Fergus, and A. Torralba. "Recognizing and Learning Object Categories, CVPR 2007 short course".
From low level to high level vision

- **Feature Detection**
  - e.g. DoG

- **Feature Description**
  - e.g. SIFT

- Estimation
- Matching
- Classification
- Detection
Classification or indexing

Is this an image of a bridge?
Image search engines
Detection
Does this image contain a bridge? [where?]
Face detection
Human body detection and gesture recognition
Single instance detection
Does this image contain the golden gate bridge? [where?]
Or which landmark does this image contain?
Visual search and landmarks recognition

Google Goggles
Fingerprint identification
Face identification
Visual search and landmarks recognition
Challenges: illumination
Challenges: scale
Challenges: deformation
Challenges: occlusion

Magritte, 1957
Challenges: background clutter

Kilmeny Niland. 1995
Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba
~10,000 to 30,000
Challenges: intra-class variation
Recognition

- Classification
- Detection
- Single instance detection and localization
Classification or indexing

Is this an image of a bridge?
Object

Bag of ‘words’
“Bag of Words” models

• 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
definition of “BoW”

– Independent features
– histogram representation
Representation

- feature detection & representation
- image representation

Codewords dictionary

Category models (and/or) classifiers

Learning

Recognition

Category decision
1. Feature detection and description

- Detect patches
  - [Mikojaczyk and Schmid '02]
  - [Mata, Chum, Urban & Pajdla, '02]
  - [Sivic & Zisserman, '03]

- Normalize patch
- Compute SIFT descriptor
  - [Lowe'99]
2. Codewords dictionary formation
2. Codewords dictionary formation
2. Codewords dictionary formation

Cluster center = code word

E.g., Kmeans clustering
2. Codewords dictionary formation

- Image patch examples of codewords
3. Bag of word representation

- Nearest neighbors assignment
- K-D tree search strategy

Codewords dictionary
3. Bag of word representation

- Codewords
- Codewords dictionary
- Frequency
Representation

1. feature detection & representation

2. codewords dictionary

3. category models
Category models

Class 1

Class N
Recognition

codewords dictionary

category models (and/or) classifiers

category decision
Major drawback of BOW models

Don’t capture spatial information!
Spatial Pyramid Matching
Spatial Pyramid Matching

- K. Grauman and T. Darrell 2005
- S. Lazebnik et al., 2006
- D. Nister et al. 2006,

Class “street”
Caltech 101

Fei-Fei et al. (2004)

Caltech 101

Pyramid matching

BOW \sim 15\%
Major drawback of BOW models

- Don’t capture spatial information!
- As the number of images/classes to model increases, the dictionary size also increases
  - Computational cost of increasing the size of the vocabulary becomes very high!
Vocabulary tree

*Scalable Recognition with a Vocabulary Tree*. David Nistér and Henrik Stewénius. 2006

- Feature vectors are hierarchically clustered in into a k-way tree – also called vocabulary tree
  - Suppose I want to have a dictionary of 81 words
First, an initial k-means process is run on the training data, defining k cluster centers. The training data is then partitioned into k groups, where each group consists of the descriptor vectors closest to a particular cluster center. The same process is then recursively applied to each group of descriptor vectors, recursively defining quantization cells by splitting each quantization cell into k new parts.
Vocabulary tree

Scalable Recognition with a Vocabulary Tree. David Nistério and Henrik Stewénius. 2006

- Computational cost is logarithmic in the number of leaf nodes.
- Vocabularies of millions (e.g., 10^6) of codewords can be supported
  - Only 10 x 6 comparisons for quantizing each descriptor
  - Individual words can be made more discriminative
Vocabulary tree

With 40,000 images in the database, the retrieval is still real-time… (in 2006 !)
Classification methods

Nearest neighbor

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

Support Vector Machines

Guyon, Vapnik, Heisele, Serre, Poggio...

Latent SVM

Structural SVM

Felzenszwalb 00
Ramanan 03...

Neural networks

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

Boosting

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

- Classification
- Detection
- Single instance detection and localization
Detection

Does this image contain a face? [where?]
Detection

Does this image contain a face? [where?]

1. Slide a window in image
   - E.g., choose position, scale orientation

2. Compare it with a model/template
   - Compute similarity to an example object or to a summary representation

3. Compute a score for each comparison

4. Retain position with max score

[Viola, Jones 2001]
Model template:
HoG (Histogram of Oriented Gradients)

• Like SIFT, but...
  – Sampled on a dense, regular grid around the object
  – Gradients are contrast normalized in overlapping blocks

In OPEN CV: `struct CV_EXPORTS HOGDescriptor`

Navneet Dalal and Bill Triggs, Histograms of Oriented Gradients for Human Detection, CVPR05
Model template: DPM (Deformable Part Model)

• Like HOG template, but...
  – Use a star-structured part-based model made of:
    • Root filter (similar to Dalal-Triggs)
    • Set of parts and an associated deformation model

More on this on Wed!
Guest lecture by Hyun Oh Song

Felzenszwalb, et al., Discriminatively Trained Deformable Part Models,
http://people.cs.uchicago.edu/~pff/latent/
Detection

– Issue with Sliding Windows approaches:

  • Computational complexity ($x,y,S,\theta,N$ of classes)

    - Beyond sliding windows (integral images) [Lampert et al 08, Alexe et al 10]
Detection

– Issue with Sliding Windows approaches:

  • Computational complexity \((x,y, S, \theta, N \text{ of classes})\)
    
    - Beyond sliding windows (integral images) [Lampert et al 08, Alexe et al 10]

  • Localization
    • Objects are not boxes
Detection

- Issue with Sliding Windows approaches:
  
  • Computational complexity \((x, y, S, \theta, N \text{ of classes})\)

    - Beyond sliding windows (integral images) [Lampert et al 08, Alexe et al 10]

- Localization
  
  • Objects are not boxes
  • Prone to false positive

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

• Selective Search:

Selective Search: Sande et al 2011
segDPM: Fidler, Mottaghi, Yuille, Urtasun 2013
Object Detection

**Deformable Part Models (DPM)**

- **DPM**: Felzenszwalb, Girshick, McAllester, Ramanan 2010
- **Sparselet**: Song et al. 2012
- **Multi-Component model**: Gu et al. 2012

**Convolutional Neural Network (CNN)**

- **CNN**: LeCun, Bottou, Bengio, Haffner 1998
- **Deep CNN**: Krizhevsky, Sutskever, Hinton 2012

**Boosting**

- **Vila-Jones Detection**: 2001
- **Regionlet**: Wang et al 2013

**3D Object Detection**

- **ALM**: Yu & Savarese, 2012
- **3D²PM**: Pepik et al 2012
- **RGBD-CPMC**: Lin et al 2013
Recognition

- Classification
- Detection
- Single instance detection and localization
Single instance detection

• Does this image contain the golden gate bridge? [where?]
• Or which landmark does this image contain?
Recognizing single instances

- Representation
  - Detectors and descriptors

- Model learning & Recognition
  - Hypothesis generation
  - Model verification

Feature descriptor
SIFT, ORB, etc...
Representation

Feature descriptor
SIFT, ORB, etc…
Recognition

**Goal:** given a query image $I$, match objects in the image against a collection of learnt object models
**Recognition**

**Goal:** given a query image $I$, match objects in the image against a collection of learnt object models

- Match features between query image $I$ and object model
- Generate hypothesis with a few matches
- Verify hypothesis with all the remaining matches
- Select hypothesis with lowest fitting error
Recognition

• Which model to use?
• How generate hypotheses?
• How to verify these hypotheses

• Detecting planar objects
• Detecting arbitrary objects and estimate camera/object pose
Algorithm:

1. Select random sample of minimum required size to fit model \[?] = [2]
2. Compute a putative model from sample set
3. Compute the set of inliers to this model from whole data set

Repeat 1-3 until model with the most inliers over all samples is found
Sample set = set of points in 2D

Algorithm:
1. Select random sample of minimum required size to fit model [?] = [2]
2. Compute a putative model from sample set
3. Compute the set of inliers to this model from whole data set
Repeat 1-3 until model with the most inliers over all samples is found
Algorithm:
1. Select random sample of minimum required size to fit model [?]
2. Compute a putative model from sample set
3. Compute the set of inliers to this model from whole data set
Repeat 1-3 until model with the most inliers over all samples is found

$|O| = 6$
Recognition

Goal: given a query image $I$, detect object instance and estimate its pose

Equivalent to: from a collection of learnt object models, find object model that matches object in image

- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results
Recognition

**Goal:** given a query image $I$, find object model that matches with $I$
Recognition

Model: collection of points on planar surface

- Find matches between “model” points and “query” points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small

query

model
Recognition

Model: collection of points on planar surface

- Find matches between “model” points and “query” points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small
Recall that the hypothesis generates low fitting error.
Verification: The hypothesis generates low fitting error

**Model:** collection of points on planar surface
- Find matches between “model” points and “query” points
- Using N matches to fit homographic transformation
- If matches and selected model are correct, the fitting error is small

- Generate hypothesis
- Verify hypothesis
- Select hypothesis with lowest fitting error
- Generate recognition results

Verification: The hypothesis generates *low* fitting error
Large-scale visual search
Recent related work on large scale and efficient image search

- Kernelized Locality Sensitive Hashing for Scalable Image Search, by B. Kulis and K. Grauman, ICCV 2009
- Improving Image-based Localization by Active Correspondence Search. T. Sattler, B. Leibe, L. Kobbelt. ECCV 2012.
- Learning Binary Projections for Large-Scale Image Search. K. Grauman and R. Fergus. Chapter to appear in Registration, Recognition, and
- City-Scale Location Recognition, G. Schindler, M. Brown, and R. Szeliski, CVPR 2007. [pdf]
Single instance object detection on a mobile device

• G. Takacs et al. "Outdoors augmented reality on mobile phone using loxel-based visual feature organization", MIR’08


CS231M • Mobile Computer Vision

Next lecture:

- Object detection by DPM and sparselets
Shape matching

- Match shape against database
- Retrieve relevant information

- Shape context (Belongie et al 00)
- Shape Classification Using the Inner-Distance [Ling and Jacobs 07]
Shape matching

Recognizing single instances

**Goal:** given a query image $I$, identify object model in the image $I$

**Model:** collection of 3D points with descriptors
Recognition

Class: toy house #3

1. Find matches between model and test image features
Recognition

Class: toy house #3

1. Find matches between model and test image features
2. Generate hypothesis:
   - Compute transformation $M$ from $N$ matches  \( (N=2; \text{affine camera; key points with scale and rotation}) \)
   - Generate hypothesis of object location and pose w.r.t. camera

$$M = K [R \ T]$$
1. Find matches between model and test image features
2. Generate hypothesis:
   • Compute transformation $M$ from $N$ matches
   • Generate hypothesis of object location and pose w.r.t. camera
3. Model verification
   • Use $M$ to project other 3D model features into test image
   • Compute residual $= D(\text{projections, measurements})$
Recognition

Class: toy house #3

4. Repeat steps 2 and 3 until residual doesn’t decrease anymore
5. Repeat steps 1-4 for different object instances
6. $M$ and $C$ corresponding to min residual return the estimated object pose and object instance