CE6902 Computer Vision

Image Recognition

Contents and Learning Objectives

1. Introduction
2. Bag of Features
3. Support Vector Machines
4. Image Recognition
5. Applications

1. Introduction

- Computer vision → machine perception → a machine can see and understand
- Image classification and recognition
  - Image recognition: what is object in the image?
  - It can be at categorization level or instance level

1. Introduction

1 Introduction – Challenges

1. View point variation

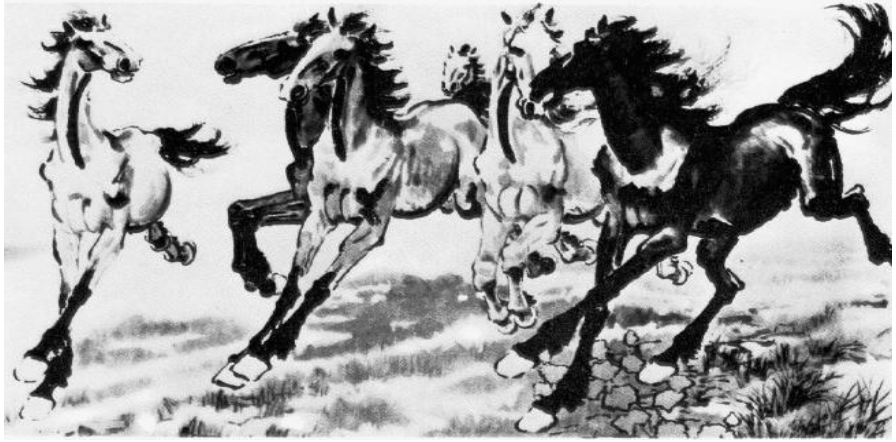
2. Illumination

3. Occlusion

1 Introduction – Challenges

4. Scales

5. Deformation



6. Background clutters

- Xu Beihong, 1943
- Klimt, 1913

1 Introduction – Challenges

7. Intra-class variations






4.1 A Gap on Features

Given feature vectors, we do classification. But what kind of feature vectors and how to organize and use them?

2 Bag of Features

The concept of bag of features comes from bag of words.

Data collection:

- **D1:** *It was the best of times,*
- **D2:** *it was the worst of times,*
- **D3:** *it was the age of wisdom,*
- **D4:** *it was the age of foolishness*

Design Vocabulary:

- 10 unique words here (ignoring case and punctuation) are: ‘it’, ‘was’, ‘the’, ‘best’, ‘of’, ‘times’, ‘worst’, ‘age’, ‘wisdom’, ‘foolishness’.

Create Document Vectors:

- D1: *It was the best of times*: [1 1 1 1 1 1 0 0 0 0]
- D2: *It was the worst of times*: [1 1 1 0 1 1 1 0 0 0]
- D: *The best of the worst*: [0 0 2 1 1 0 1 0 0 0]

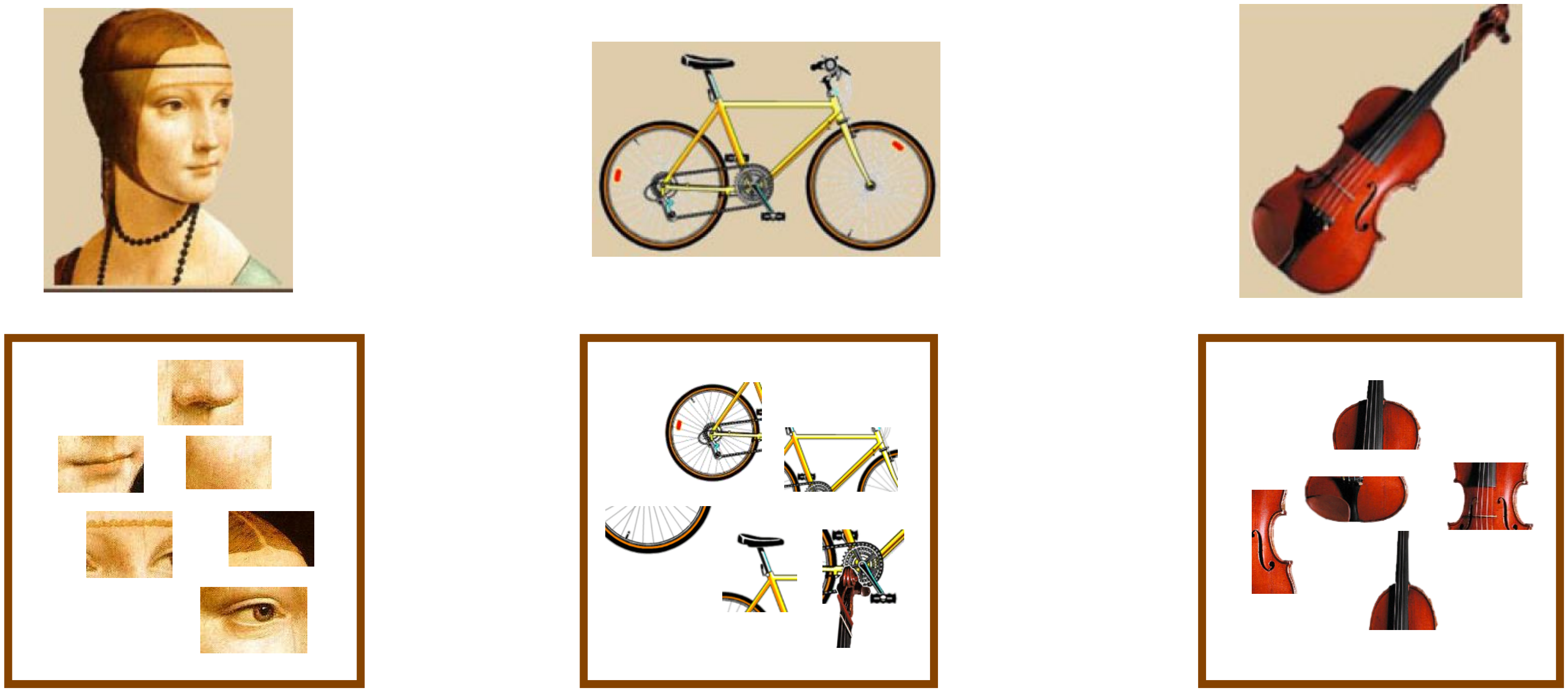
2 Bag of Features

Works pretty well for image-level classification

class	bag of features Zhang et al. (2005)	bag of features Willamowski et al. (2004)	Parts-and-shape model Fergus et al. (2003)
airplanes	**98.8**	97.1	90.2
cars (rear)	98.3	**98.6**	90.3
cars (side)	**95.0**	87.3	88.5
faces	**100**	99.3	96.4
motorbikes	**98.5**	98.0	92.5
spotted cats	**97.0**	—	90.0

2 Bag of Features

1. Extract features

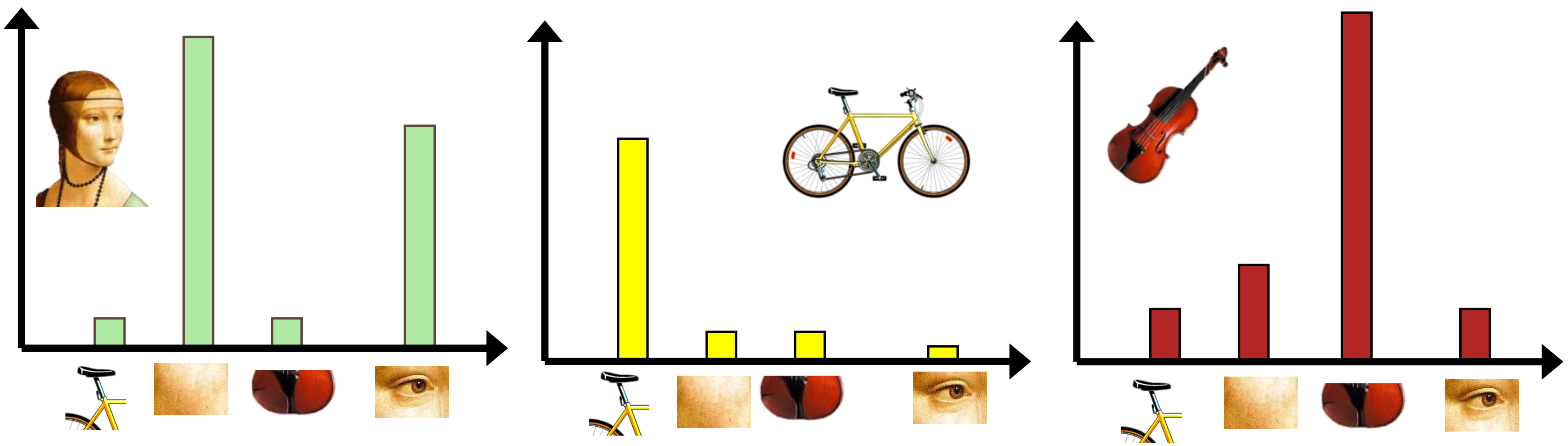


2. Learn “visual vocabulary”

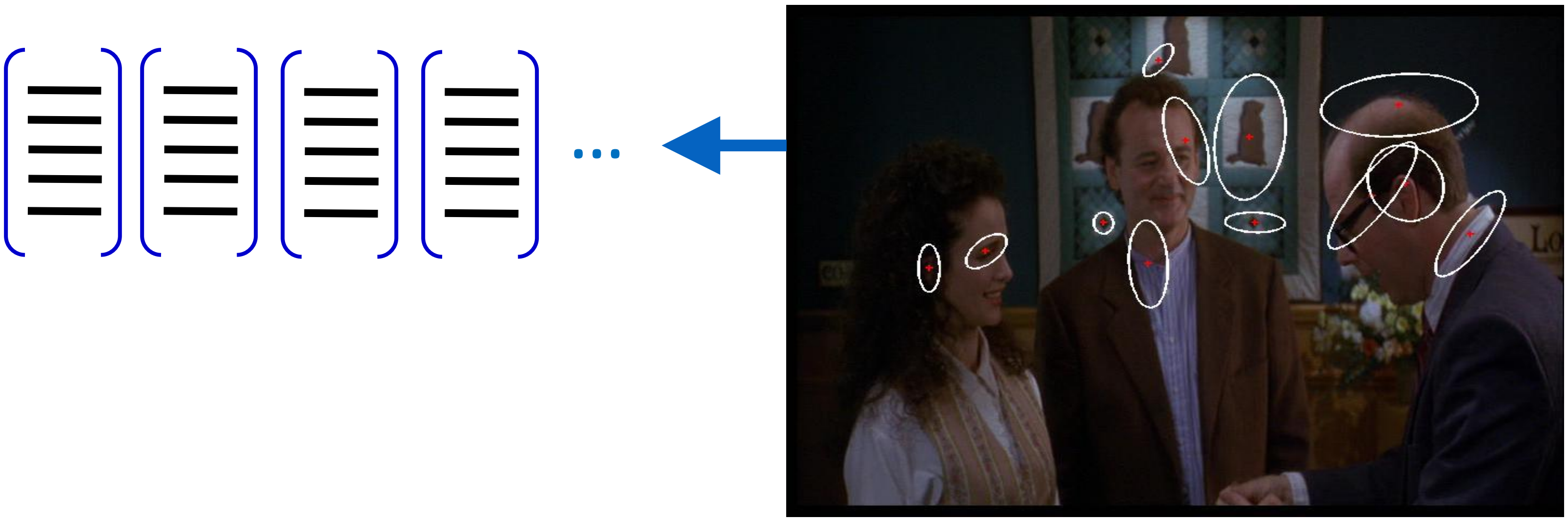
2 Bag of Features

3. Quantize features using visual vocabulary

4. Represent images by frequencies of "visual words"

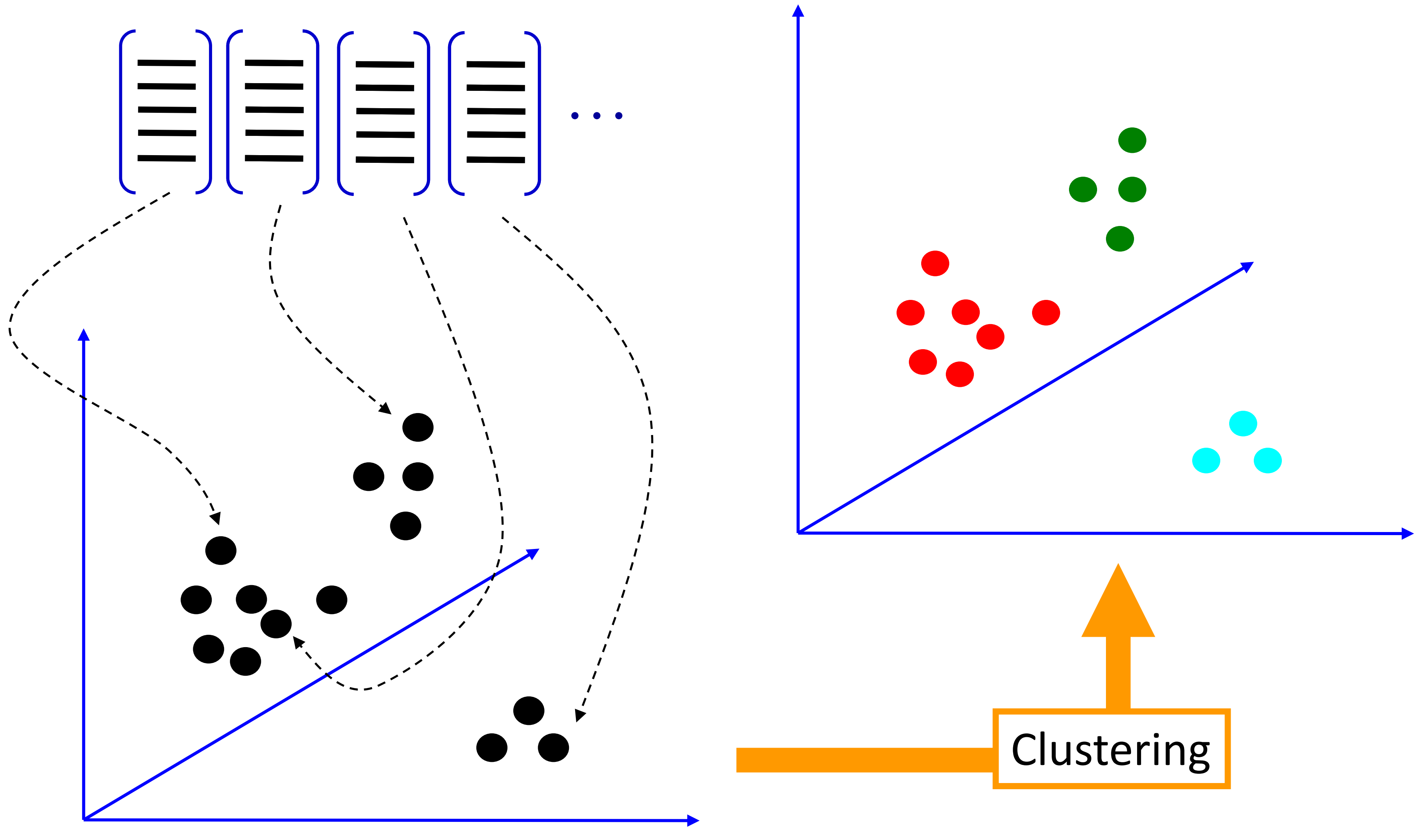


2 BoF – Extract Features (Step 1)

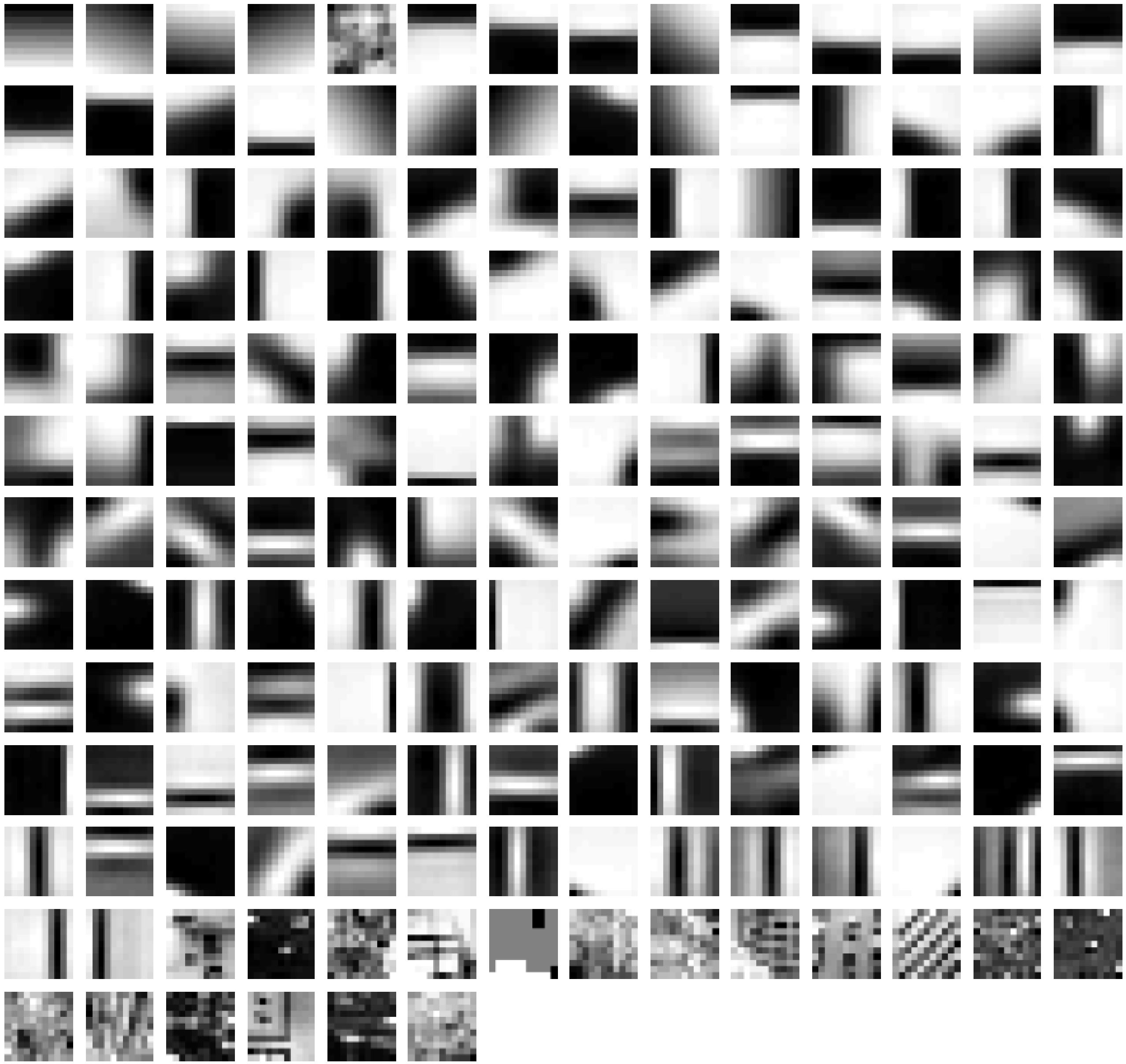


- Feature detection is of big interest for long time.
- There are many feature detectors.
- As shared, SIFT has been widely adopted. The online code is available so you can run it before you know it.

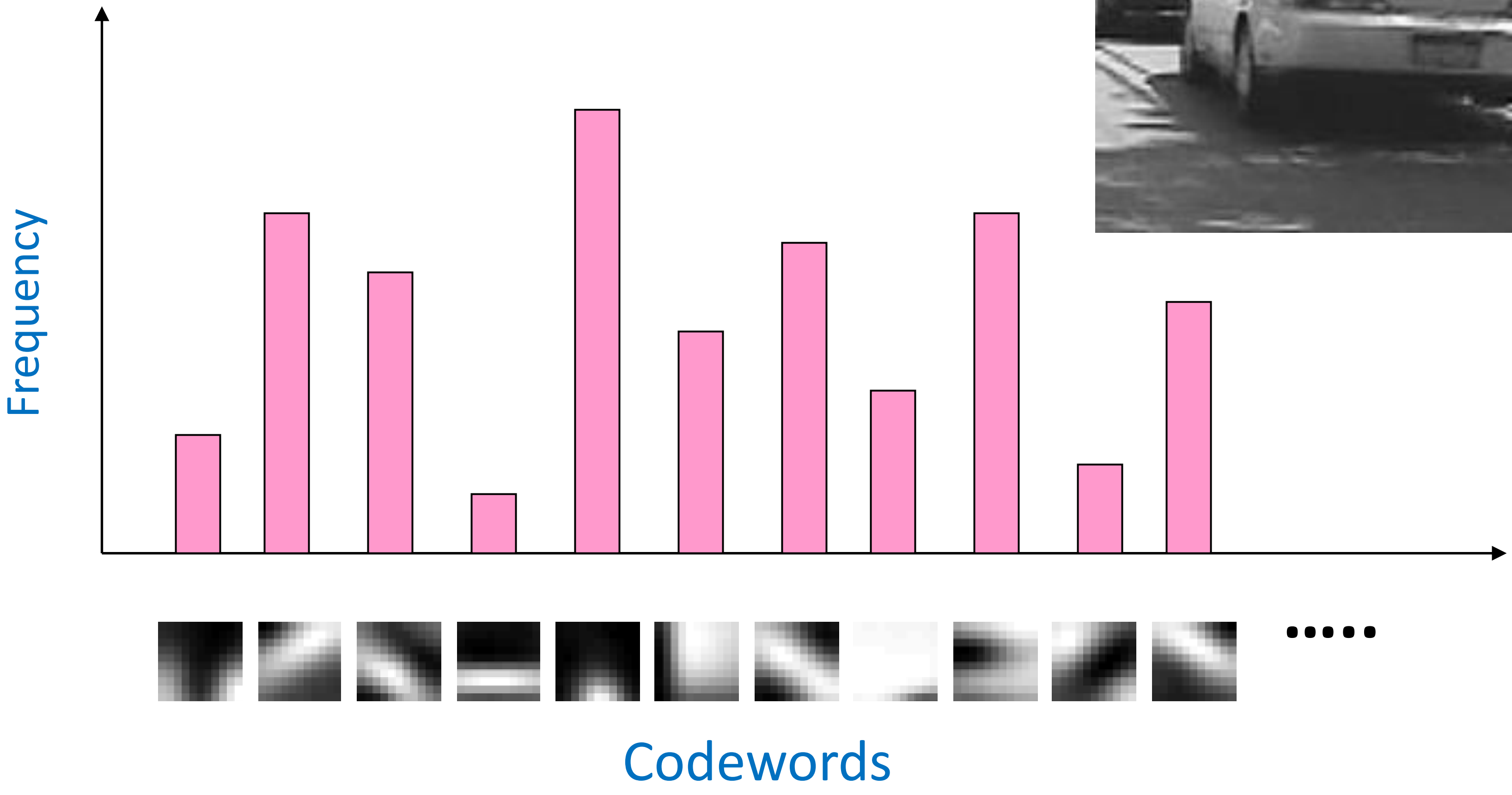
2 BoF – Learn Visual Dictionary (Step 2)



2 BoF – Learn Visual Dictionary (Step 2)



2 BoF – Quantization and Representation (3 – 4)

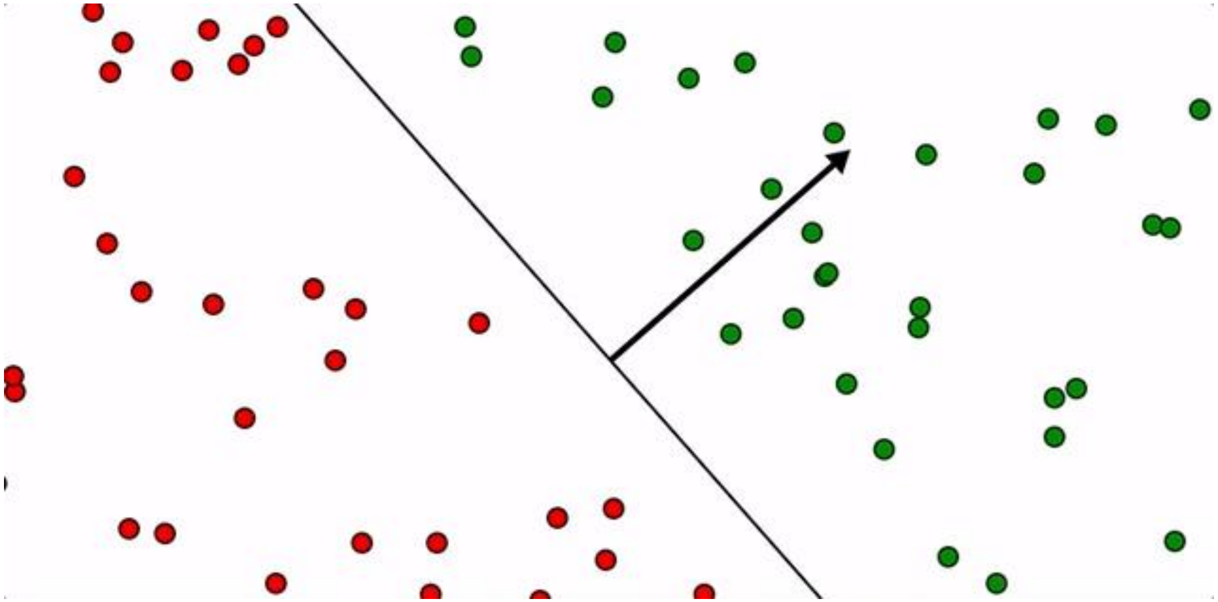


3. Support Vector Machines

Given a linearly separable dataset $\{(\mathbf{x}_1,y_1), (\mathbf{x}_2,y_2), \ldots, (\mathbf{x}_N,y_N)\}$, $\mathbf{x}\in\mathbf{R}^D$, $y\in\{-1,+1\}$, finding a separating hyperplane that satisfies:

$$\begin{cases} \mathbf{w}^T\mathbf{x}_i + b \geq 1 & for\ y_i = +1 \\ \mathbf{w}^T\mathbf{x}_i + b \leq -1 & for\ y_i = -1 \end{cases} \Rightarrow y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad \forall i$$

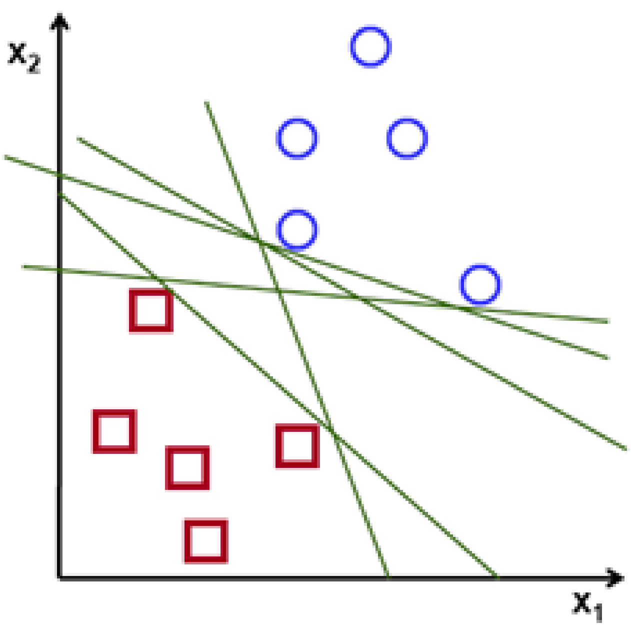
There are an infinitely number of hyperplanes

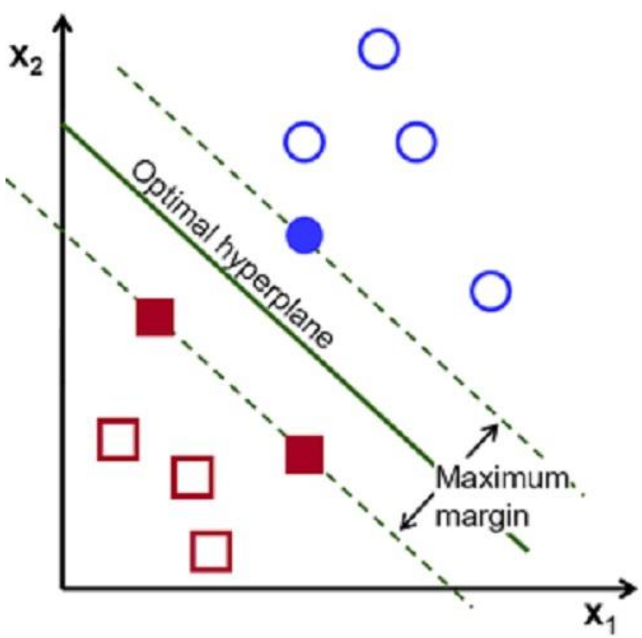


3. Support Vector Machines

Given a linearly separable dataset $\{(\mathbf{x}_1,y_1), (\mathbf{x}_2,y_2), \ldots, (\mathbf{x}_N,y_N)\}$, $\mathbf{x}\in\mathbf{R}^D$, $y\in\{-1,+1\}$, finding a separating hyperplane

- A hyperplane too close to the training examples will be sensitive to noise and less likely to generalize well for data unseen data during the training.
- The optimal separating hyperplane will be the one with the largest margin





3. Support Vector Machines

The hyperplane equation

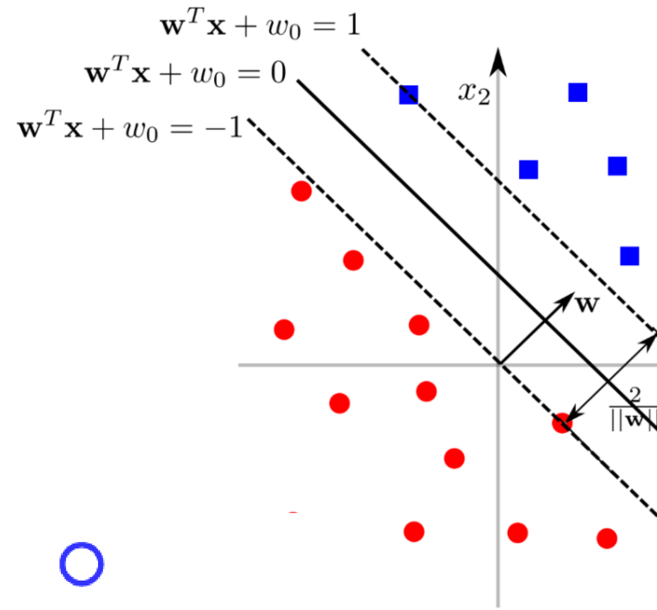
$$\mathbf{w}^T\mathbf{x} + b = 0$$

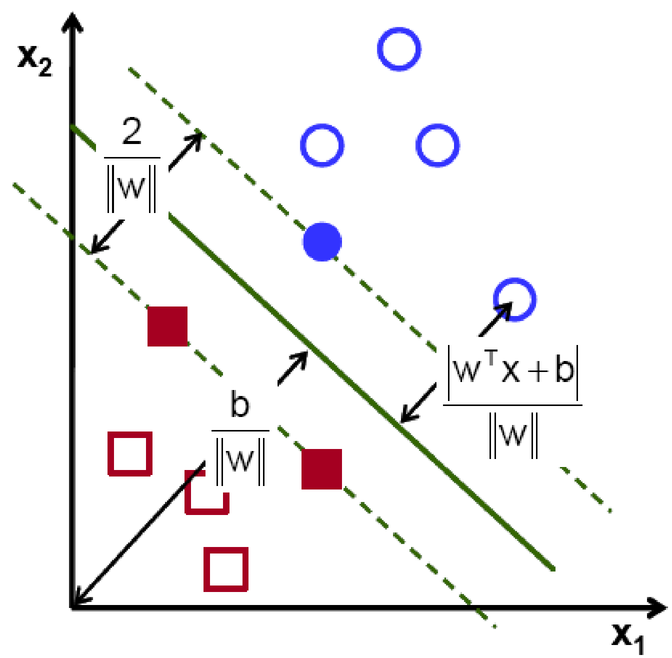
The canonical hyperplane

$$\left|\mathbf{w}^T\mathbf{x}_i + b\right| = 1$$

The margin

$$m = \frac{2}{\|\mathbf{w}\|}$$





3. Support Vector Machines

The mathematical problem

$$\text{minimize} \qquad \mathrm{J}(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2$$

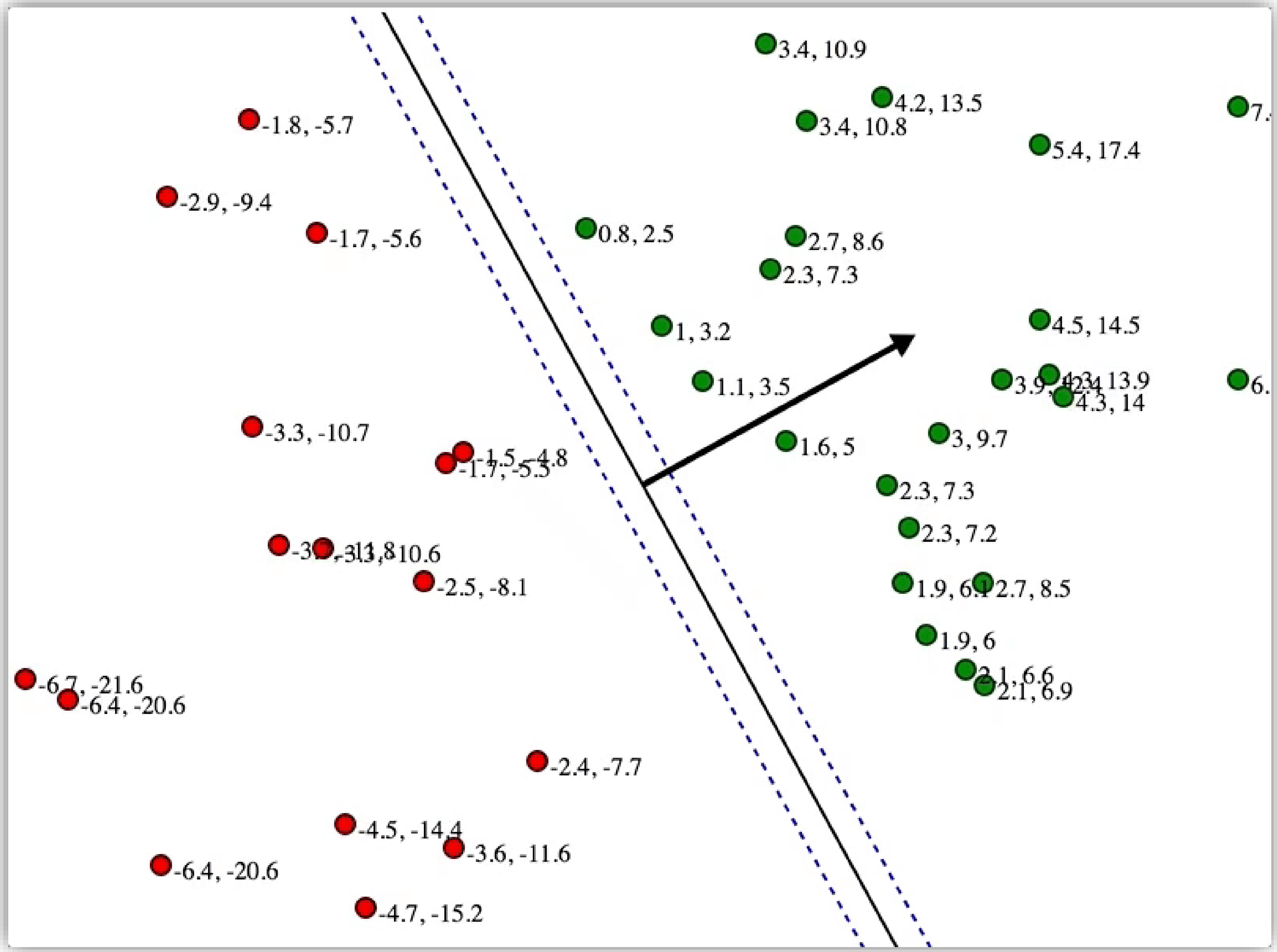
$$\text{subject to} \qquad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \forall i$$

J(**w**) is a quadratic function, which means that there exists a single global minimum and no local minima

This converts to the Lagrangian primal problem

$$\text{minimize} \qquad L_p(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i \left[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right]$$

3. Support Vector Machines



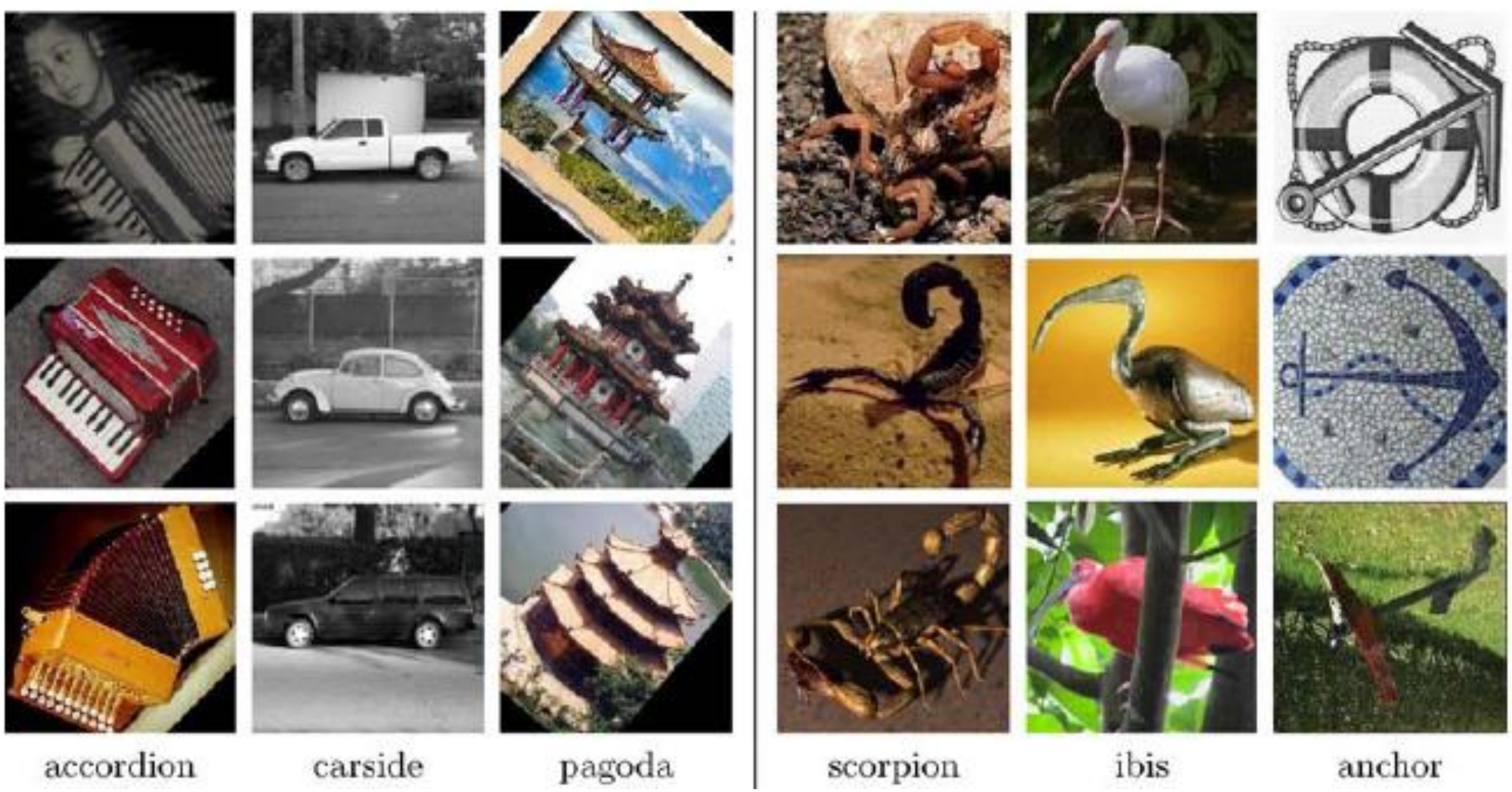
4. Image Classification – Methods

- For each image in the training set, a number of SIFT feature points can be detected and the corresponding SIFT descriptors can be determined.
- A dictionary can be constructed by clustering the SIFT descriptors that are extracted from all training images.
- Each training image can then be quantitated into a feature vector that records the feature-point frequency and its dimension is the same as the dictionary size.
- A SVM classifier can then be trained based on the determined feature vectors and their labels.
- Each image in the test set can be converted into a feature vector similarly and classified by the trained SVM classifier.

4. Image Classification – Datasets

The CalTech101 dataset

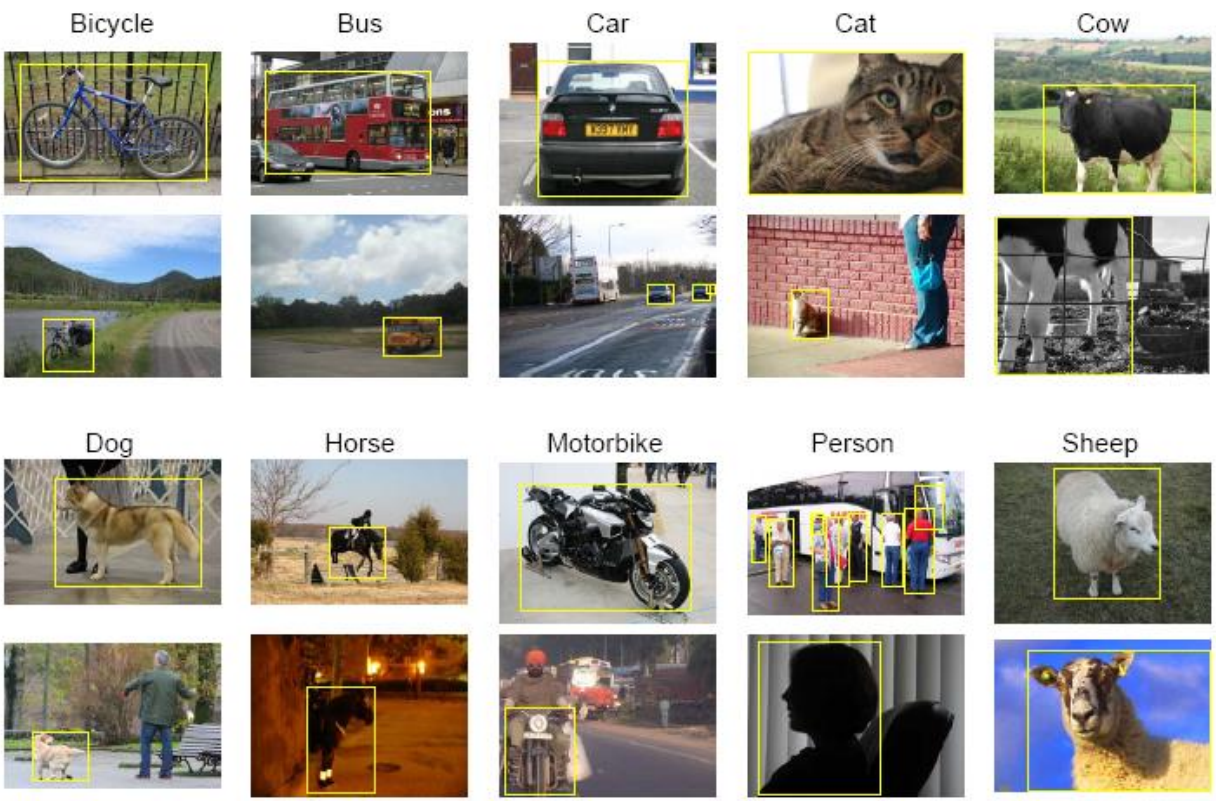
- 101 object categories with 40 to 800 images per category.
- www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html
- CalTech256 dataset



4. Image Classification – Datasets

PASCAL 2006 Dataset

- Ten object classes: bicycle, bus, car, cat, cow, dog, horse, motorbike, person, and sheep
- http://host.robots.ox.ac.uk/pascal/VOC/voc2006/index.html



4. Image Classification – Datasets

- ImageNet dataset:
  - http://www.image-net.org/
  - Total number of non-empty synsets: 21,841
  - Total number of images: 14,197,122

- SUN dataset:
  - http://groups.csail.mit.edu/vision/SUN/
  - Total number of scene categories: 908
  - Total number of images: 131,067
  - Total number of object categories: 4,479
  - Number of segmented objects: 313,884

4. Image Classification – Evaluations

Image classification can be evaluated by using precision, recall, and F-score.

Suppose the cutoff threshold is chosen to be 0.8. In other words, any instance with posterior probability greater than 0.8 is classified as positive.

Instance	P(+|A)	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

4. Image Classification – Evaluations

	PREDICTED CLASS		
		Class= Yes	Class= No
ACTUAL CLASS	Class= Yes	(TP) 3	(FN) 2
	Class= No	(FP) 3	(TN) 2

Instance	P(+|A)	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

4. Image Classification – Evaluations

	PREDICTED CLASS	
ACTUAL CLASS		
	Class=Yes	Class=No
Class=Yes	(TP) 3	(FN) 2
Class=No	(FP) 3	(TN) 2

P = TP/(TP+FP)= 3/(3+3) = 1/2

R = TP/(TP+FN) =3/(3+2) = 3/5

F-measure = 2pr/(p+r) = 6/11

4. Image Classification – Evaluations

Given 10K fruit pictures including 50 apple pictures, a retrieval system searches for apple pictures, the search returns:

1. 500 pictures: 50 apple pictures (TP) and 450 non-apple pictures (FP):

   R = TP/(TP+FN) = 50/50 = 100%; (FN = 0)

   P = TP/(TP+FP) = 50/500 = 10% (FP = 450)

   F = 2*100%*10%/(100%+10%) = 2/11

2. 5 pictures: all are apple pictures:

   R = TP/(TP+FN) = 5/50 = 10%; (FN = 45)

   P = TP/(TP+FP) = 5/5 = 100% (FP = 0)

   F = 2*10%*100%/(10%+100%) = 20/110 = 2/11

3. 100 pictures: 40 apple pictures and 60 non-apple pictures:

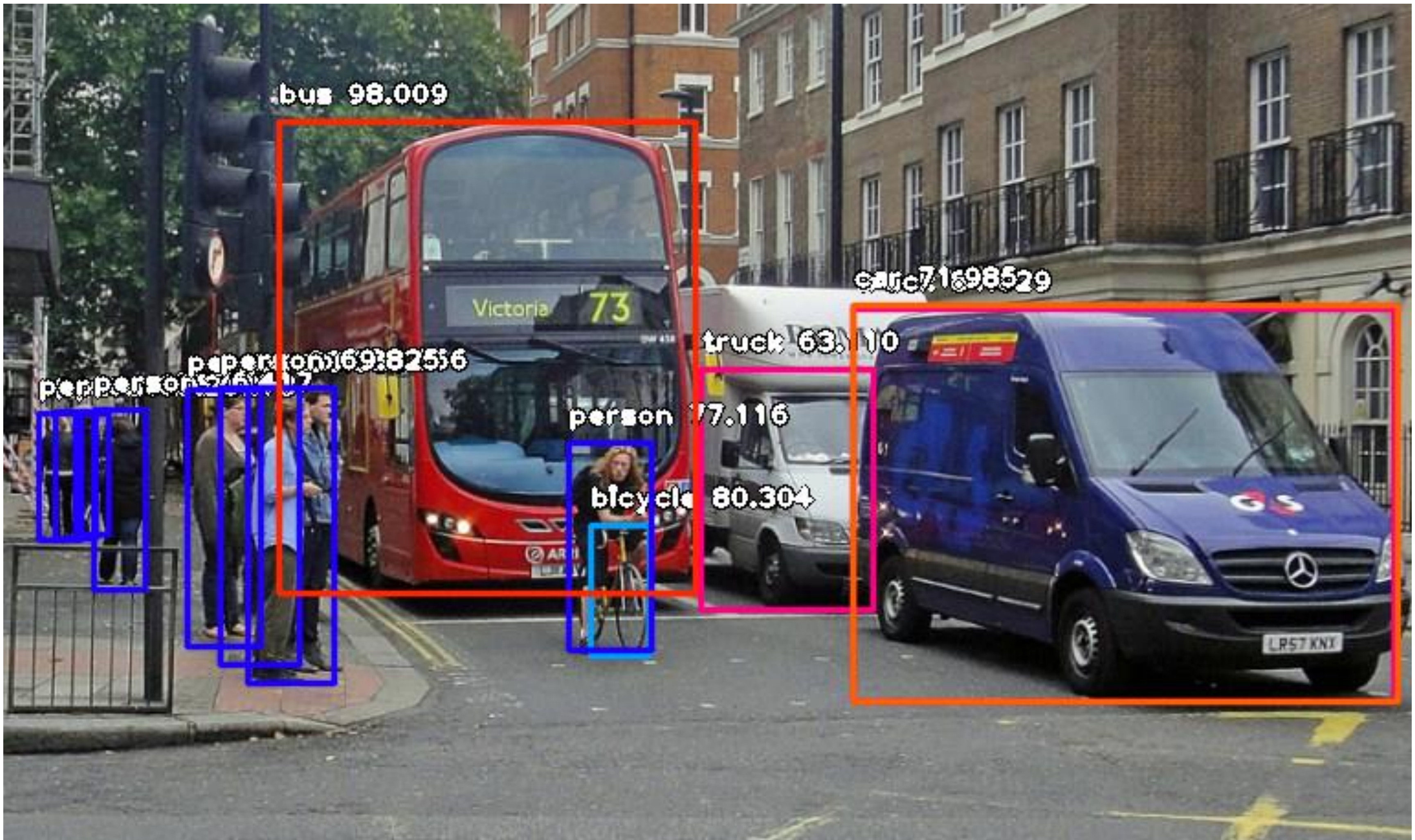
   R = 40/50 = 80%; (FN = 10)

   P = 40/100 = 40% (FP = 60)

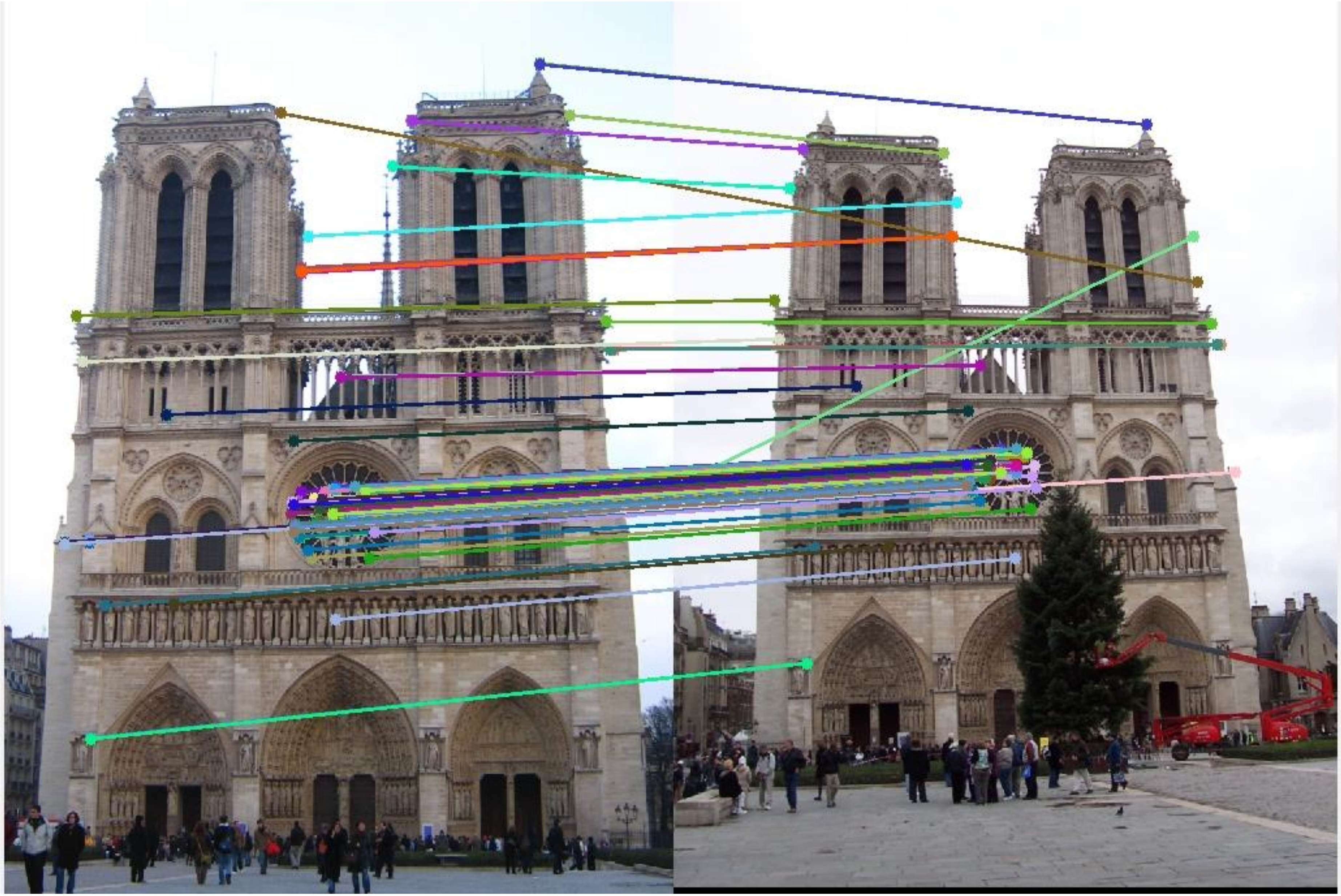
   F = 2*80%*40%/(80%+40%) = 64/120 = 8/15

5. Applications

Object detection and object recognition



5. Applications



5. Applications

Productivity and efficiency issue

Safety issues

5. Applications

Implementation issue

5. Applications

Recording and manpower issues

5. Applications

We developed a mobile vision system that detects and recognize illegal parking by just driving a enforcement vehicle around.

Summary

1. Basics about features
2. Feature detection
3. Feature description
4. Feature matching
5. Applications