Bag of Words
or
Bag of Features

Lecture-17

(Slides Credit: Cordelia Schmid

LEAR – INRIA Grenoble)

Contents

- Interest Point Detector
- Interest Point Descriptor
- K-means clustering
- Support Vector Machine (SVM) classifier
- Evaluation Metrics: Precision & Recall

Copyright Mubarak Shah 2003

Image classification

- Image classification: assigning a class label to the image

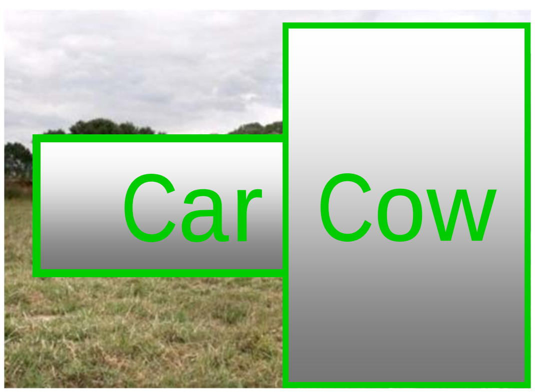
Car: present
Cow: present
Bike: not present
Horse: not present
...

Image classification

- Image classification: assigning a class label to the image

Car: present
Cow: present
Bike: not present
Horse: not present
...

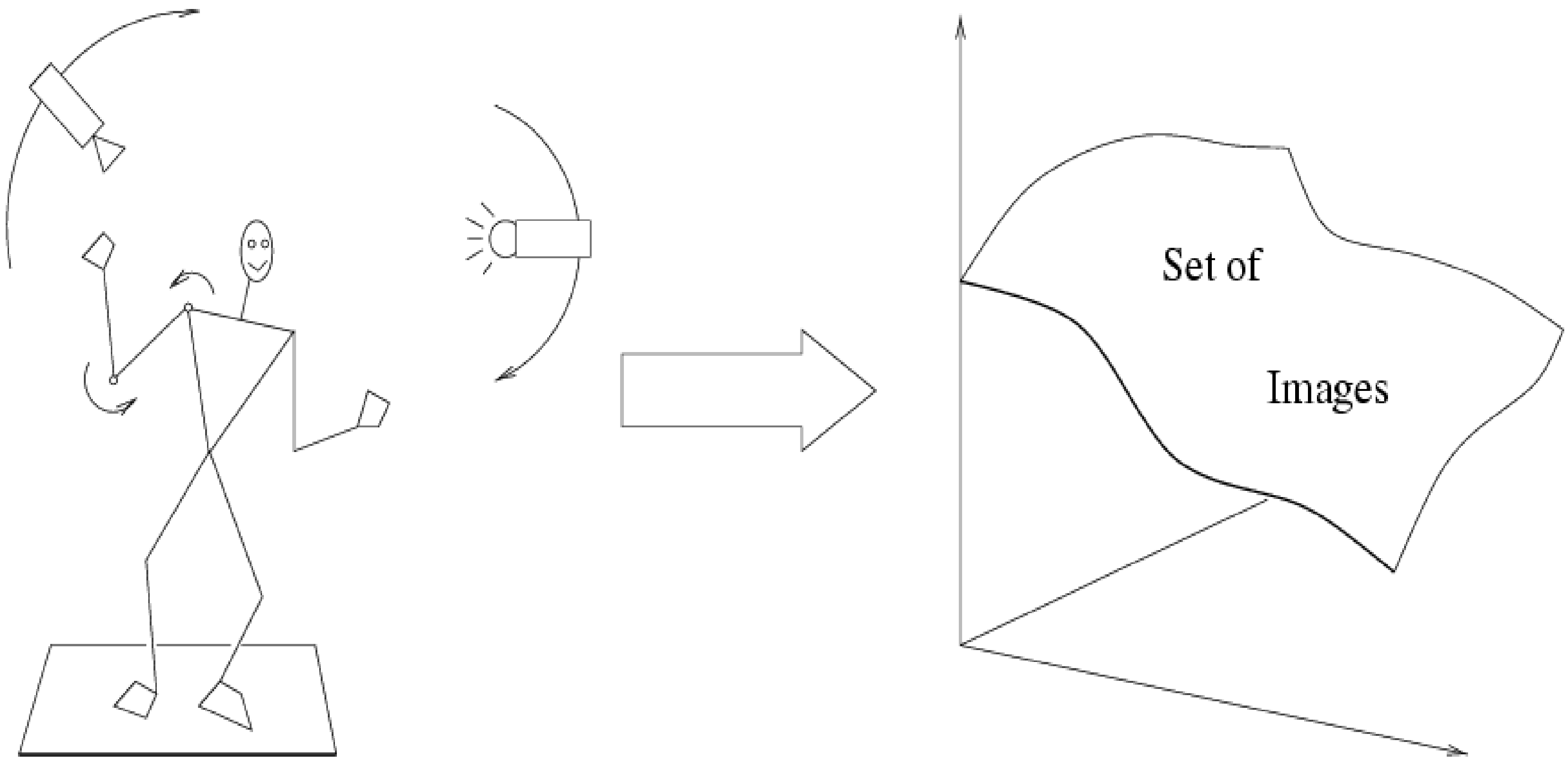
- Object localization: define the location and the category



Location

Category

Difficulties: within object variations



Variability: Camera position, Illumination,Internal parameters

⇨ Within-object variations

Difficulties: within class variations



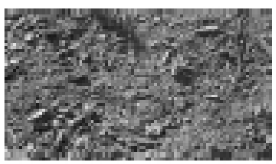



Image classification

- Given

  Positive training images containing an object class

  Negative training images that don't



- Classify

  A test image as to whether it contains the object class or not

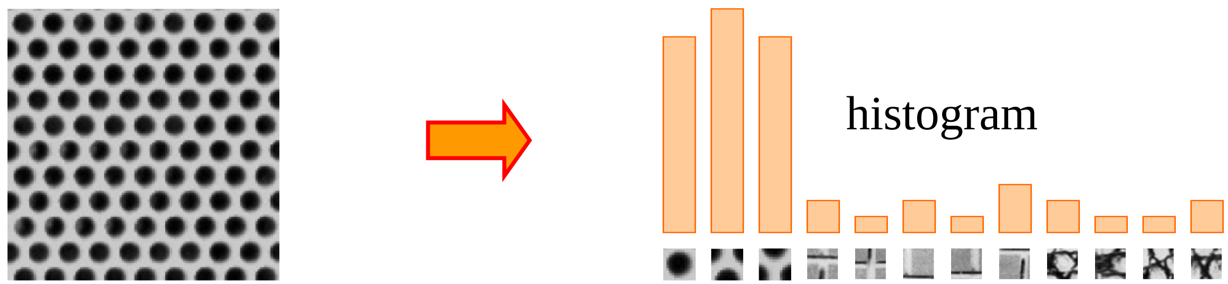
?

Bag-of-features – Origin: texture recognition

- Texture is characterized by the repetition of basic elements or *textons*

Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001 Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Bag-of-features – Origin: texture recognition



Bag of Words Model

The last duel
After quarrelling over a bank loan, two men took part in the last fatal duel staged on Scottish soil. BBC News's James Landale retraces the steps of his ancestor, who made that final challenge.

Doc-1

Bank : 1
Loan : 1
Water : 0
Farmer : 0

West Bank water row
Palestinians have accused Israel of diverting water away from their towns in order to keep Jewish settlements in the occupied territories fully supplied. Israel denies the charge saying Palestinian farmers are to blame for using illegal connections to irrigate their fields.

Doc-2

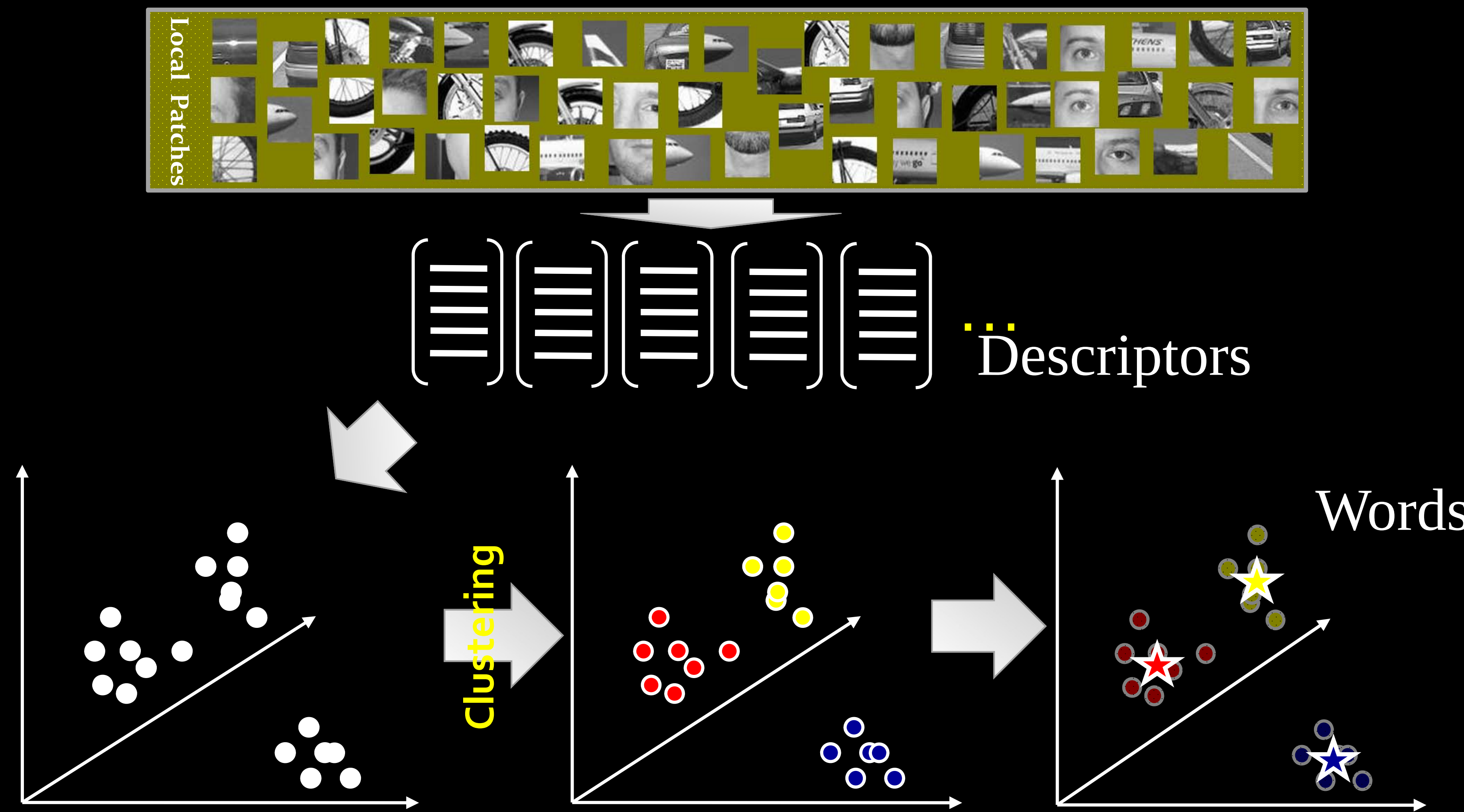
Bank : 1
Loan : 0
Water : 2
Farmer : 0

Words \ Documents	Doc-1	Doc-2	Doc-3	Doc-4
Bank				
Loan				
Water				
Farmer				

Words \ Documents	Doc-1	Doc-2	Doc-3	Doc-4
Bank	0.5	0.33	0	0
Loan	0.5	0	0	0.6
Water	0	0.66	1	0.2
Farmer	0	0	0	0.2

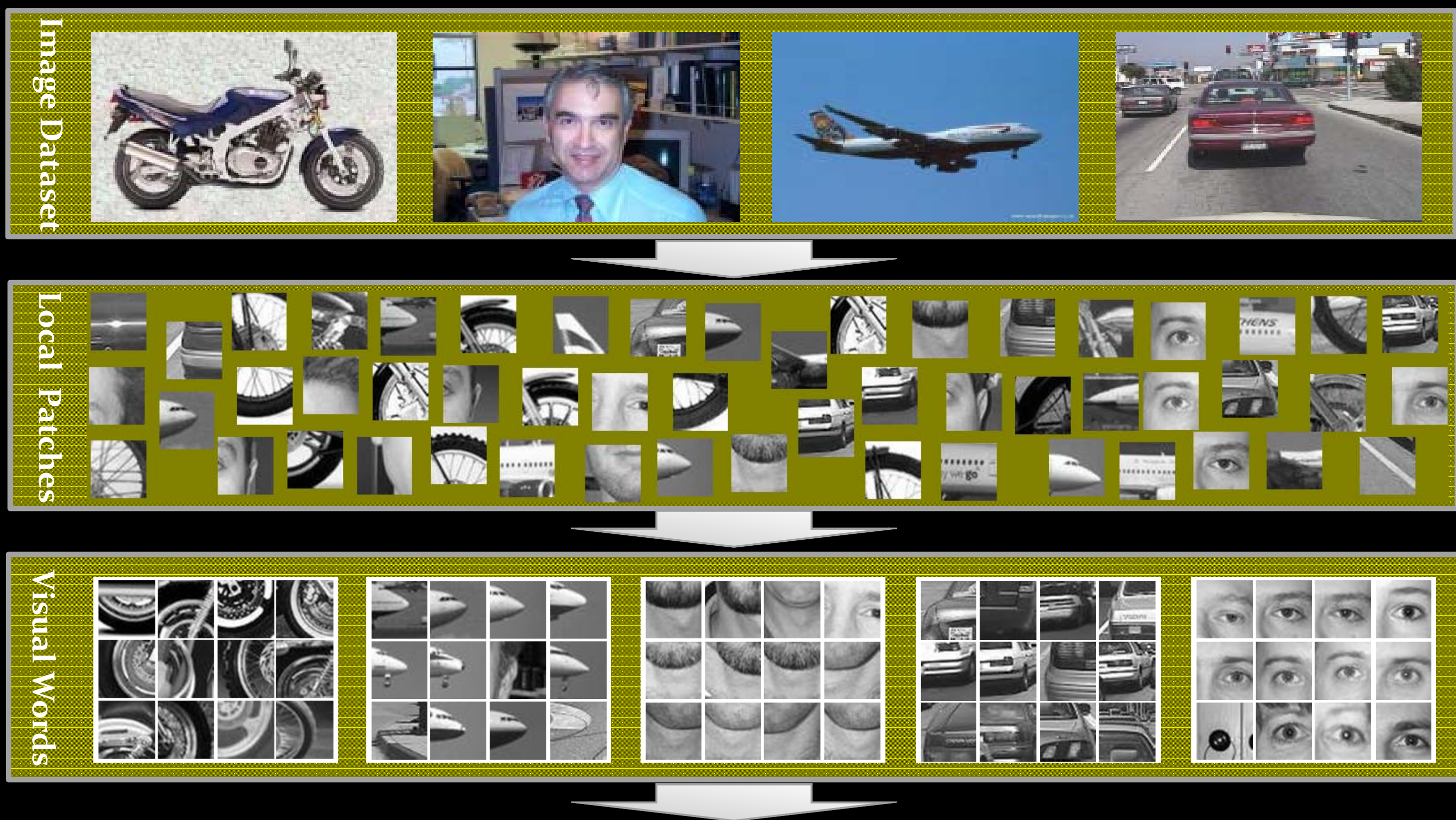
Bag of Visual Words model

Bag of Visual Words model

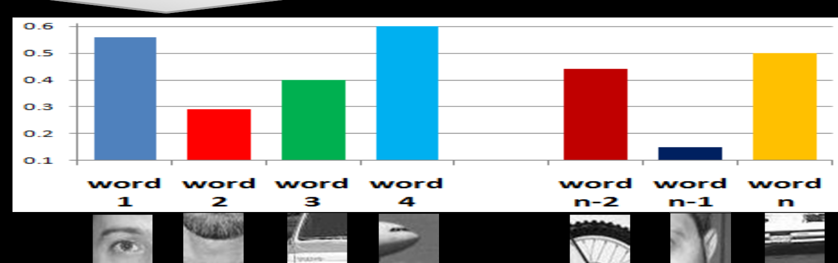


Generate the visual vocabulary

Bag of Visual Words model

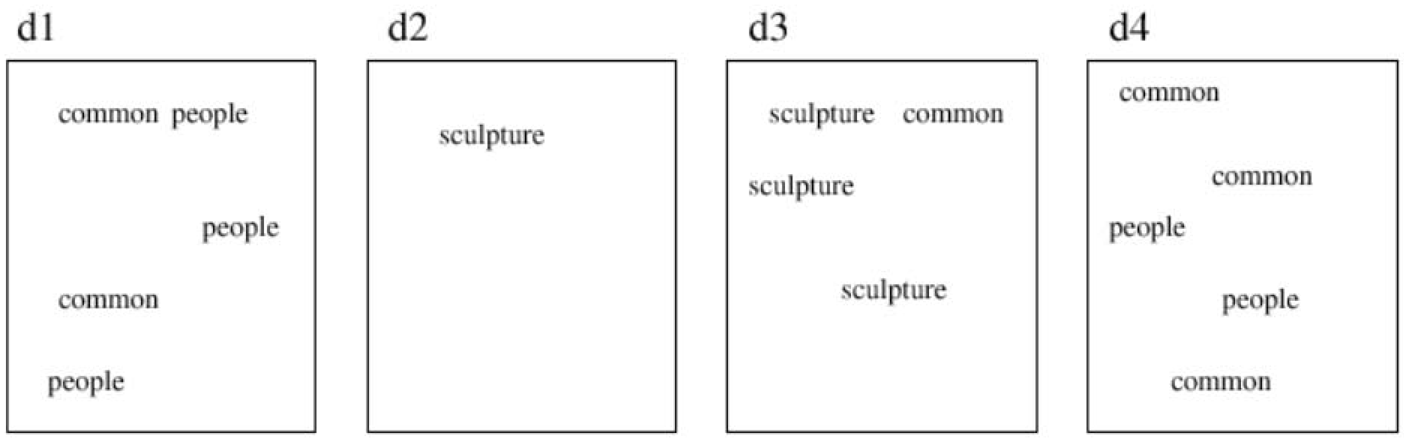


Represent an image as a histogram or bag of words



Bag-of-features – Origin: bag-of-words (text)

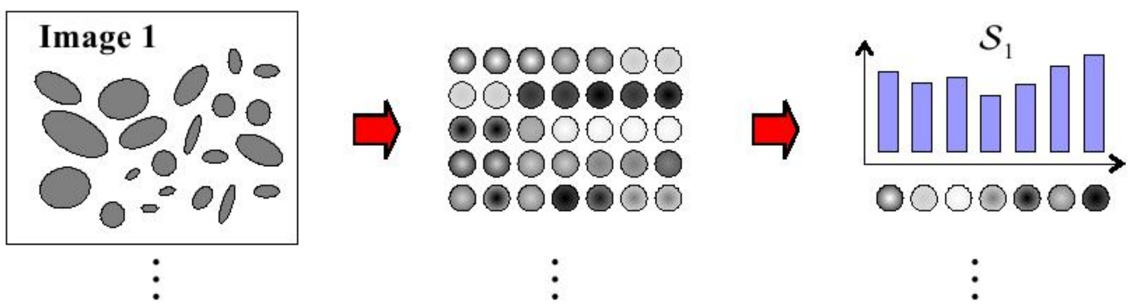
- Orderless document representation: frequencies of words from a dictionary
- Classification to determine document categories



Bag-of-words

Common	2	0	1	3
People	3	0	0	2
Sculpture	0	1	3	0
…	…	…	…	…

Bag-of-features for image classification

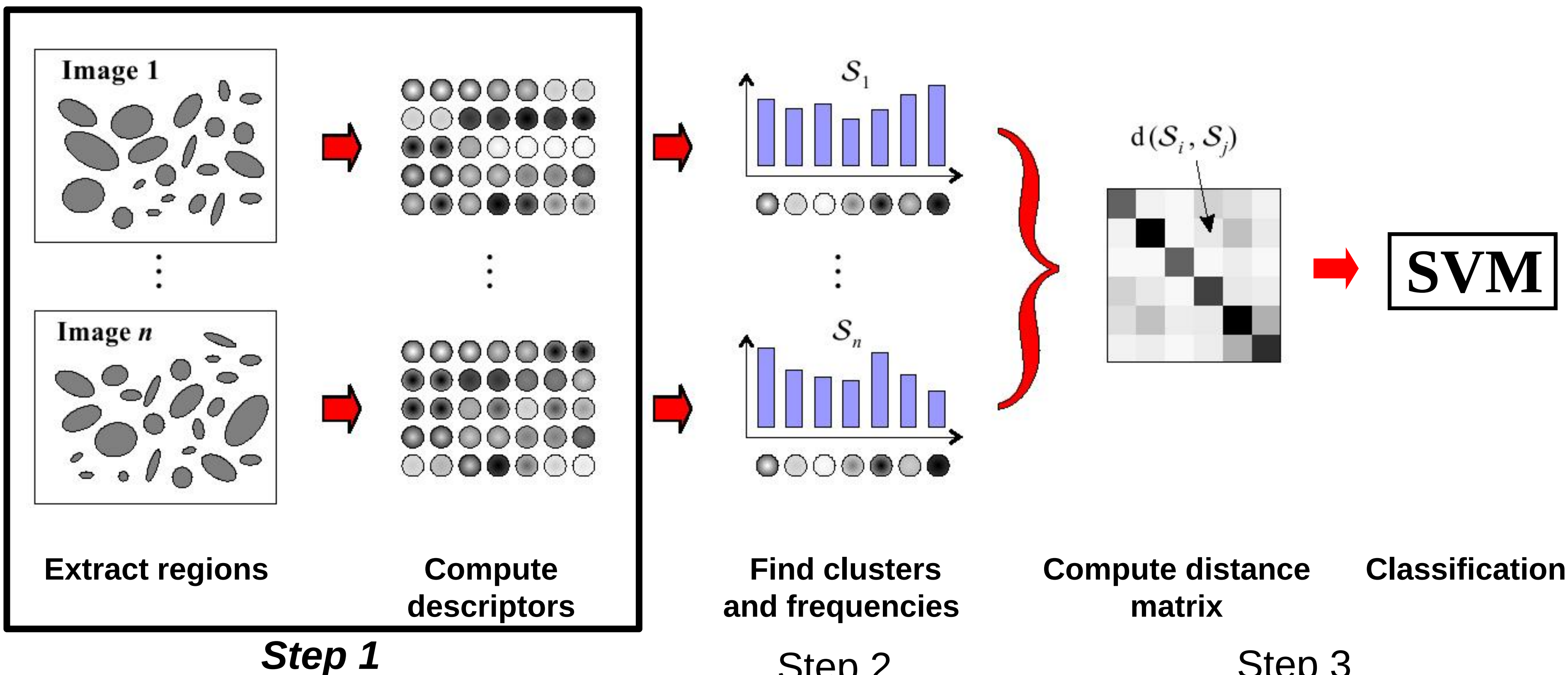


Extract regions or Interest Points

Compute descriptors

Find clusters and frequencies

Bag-of-features for image classification

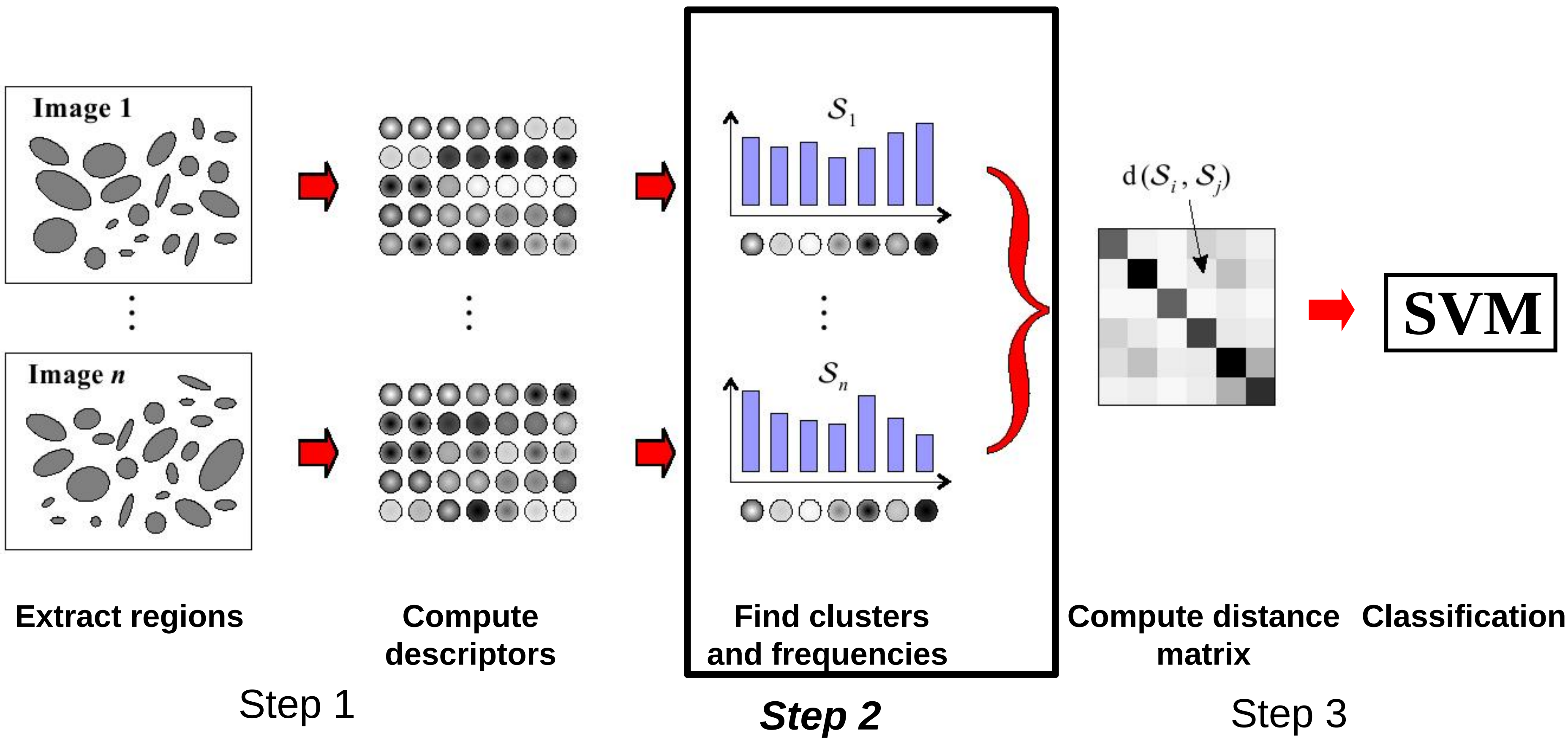


Step 1: feature extraction

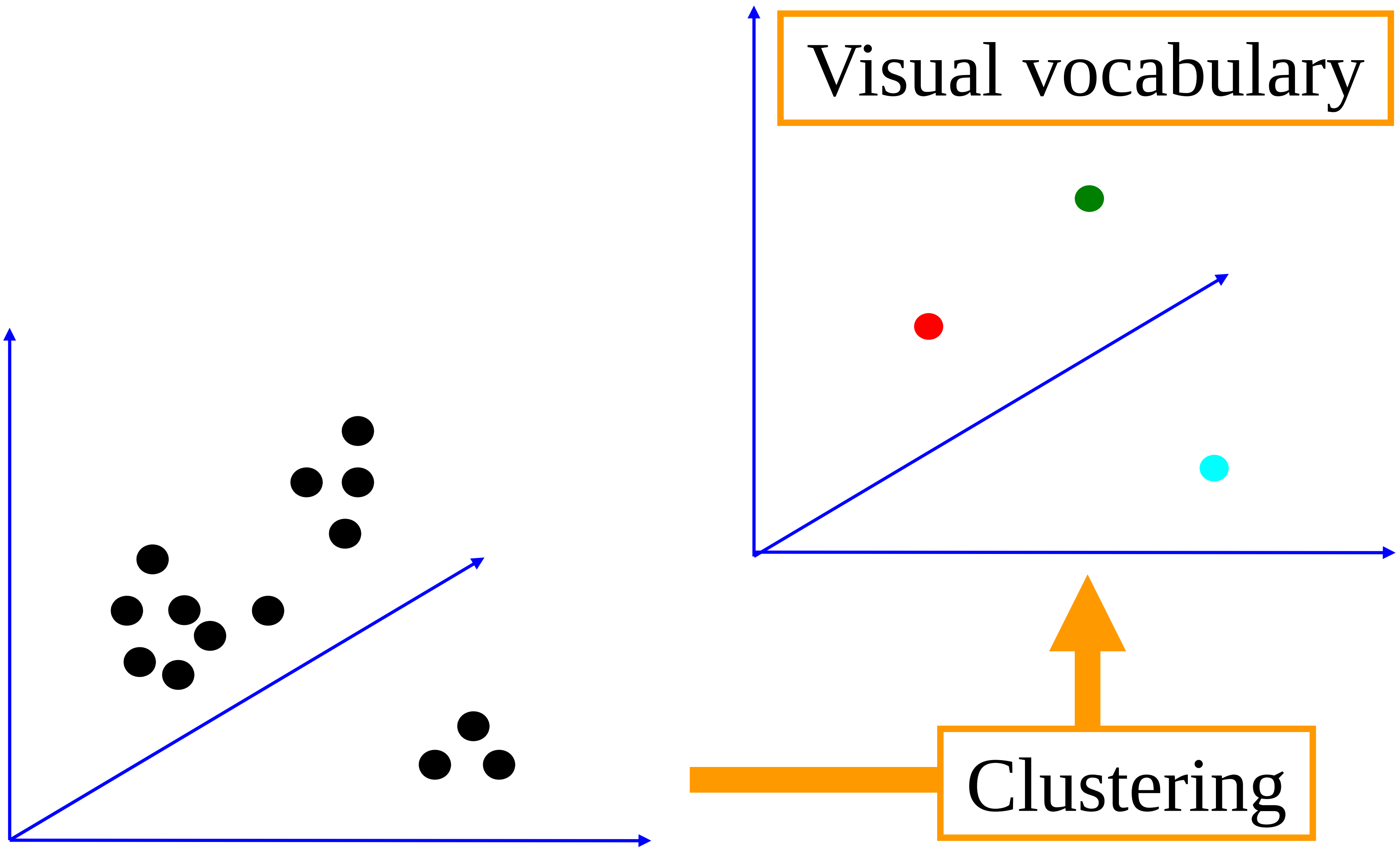
- Detect Interest Points
  - SIFT
  - Harris
  - Dense (take every nth pixel as interest point)
- Compute Descriptor around each interest point
  - SIFT
  - HOG

Dense features

Bag-of-features for image classification



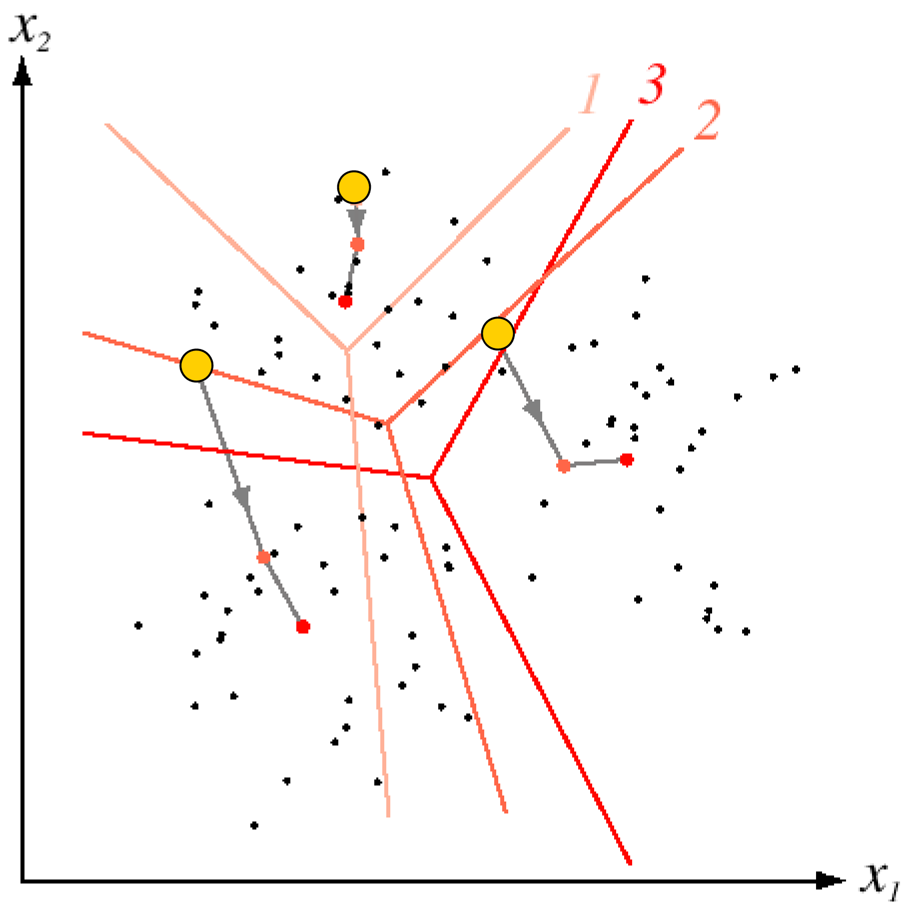
Step 2: Quantization



Step 2: Quantization

- Cluster descriptors
  - K-means
- Assign each visual word to a cluster
- Build frequency histogram

Example: 3-means Clustering



from
Duda et al.

Convergence in 3 steps

K-Means

```
Choose k data points to act as cluster centers

Until the cluster centers are unchanged

    Allocate each data point to cluster whose center is nearest





    Replace the cluster centers with the mean of the elements
    in their clusters.

end
```

Algorithm 16.5: *Clustering by K-Means*

Examples for visual words

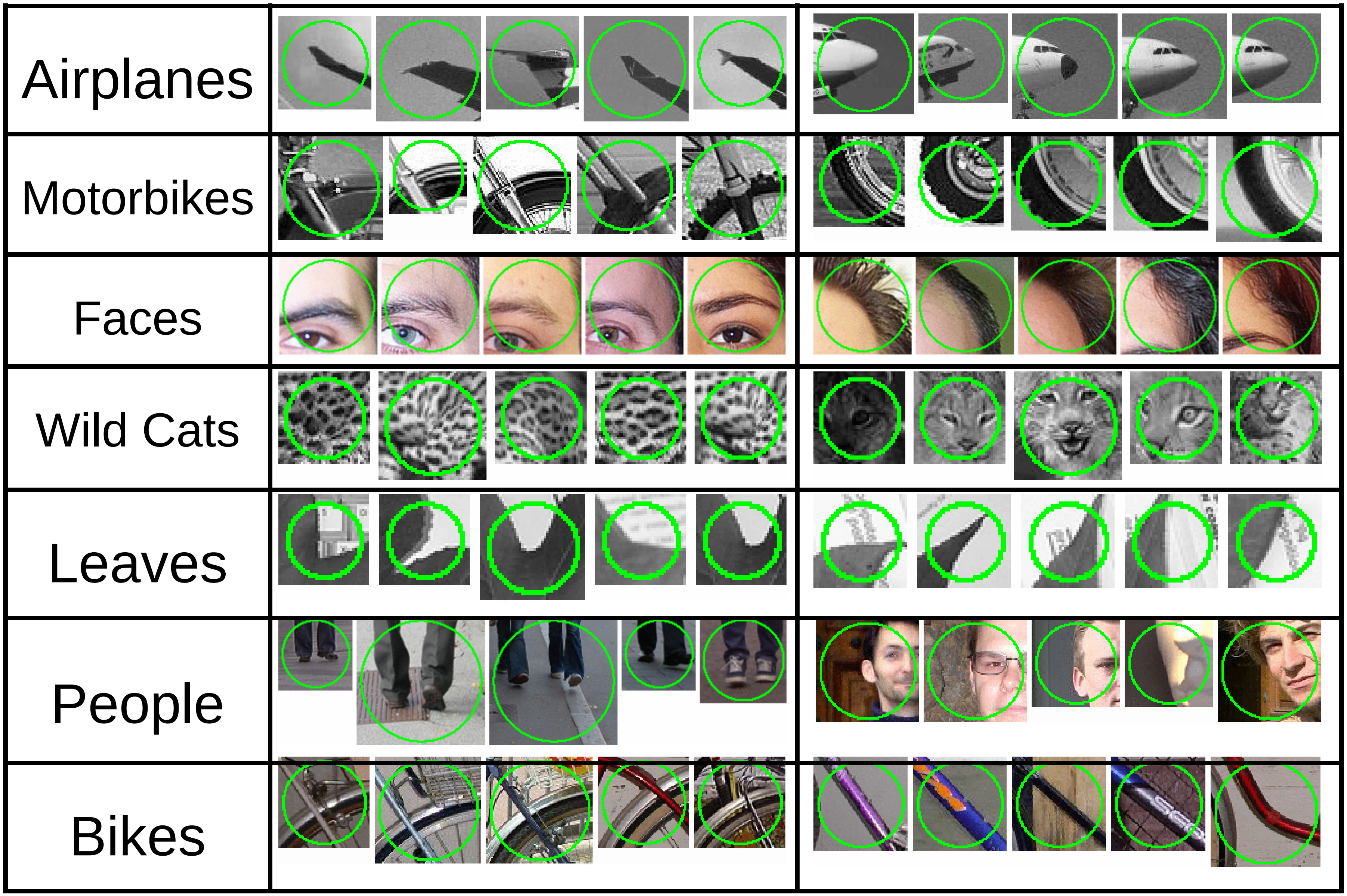
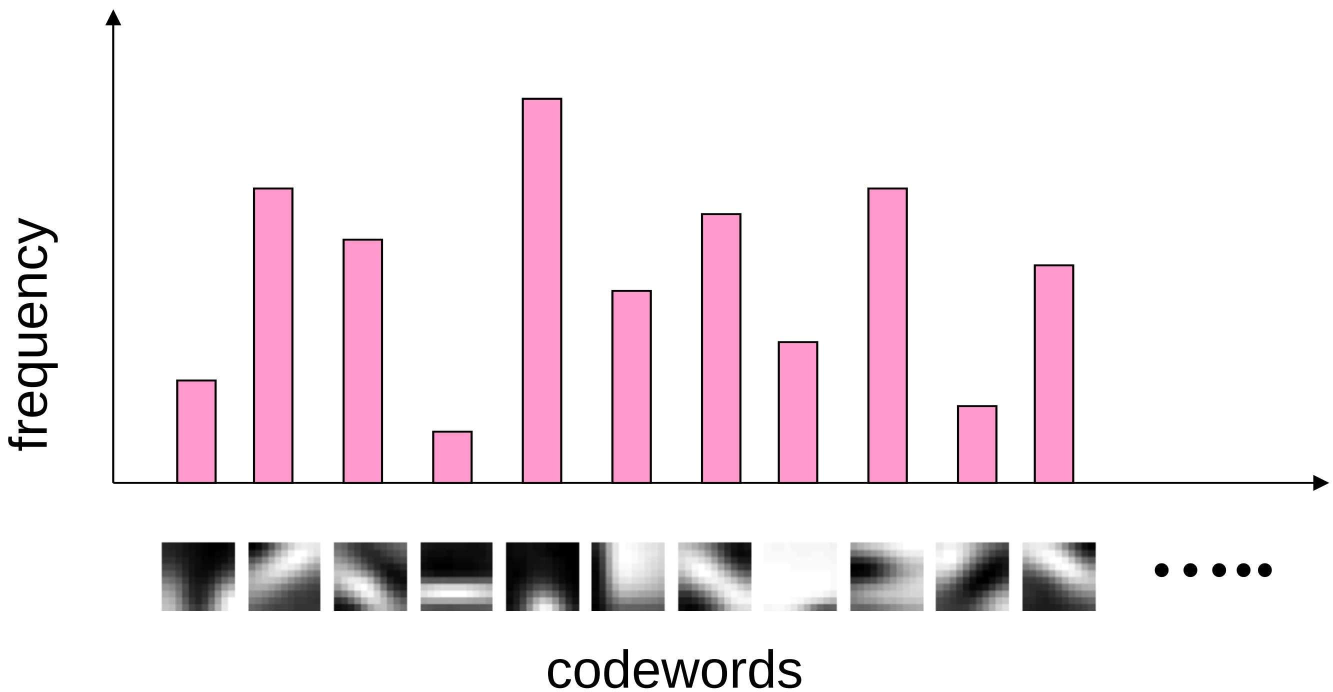
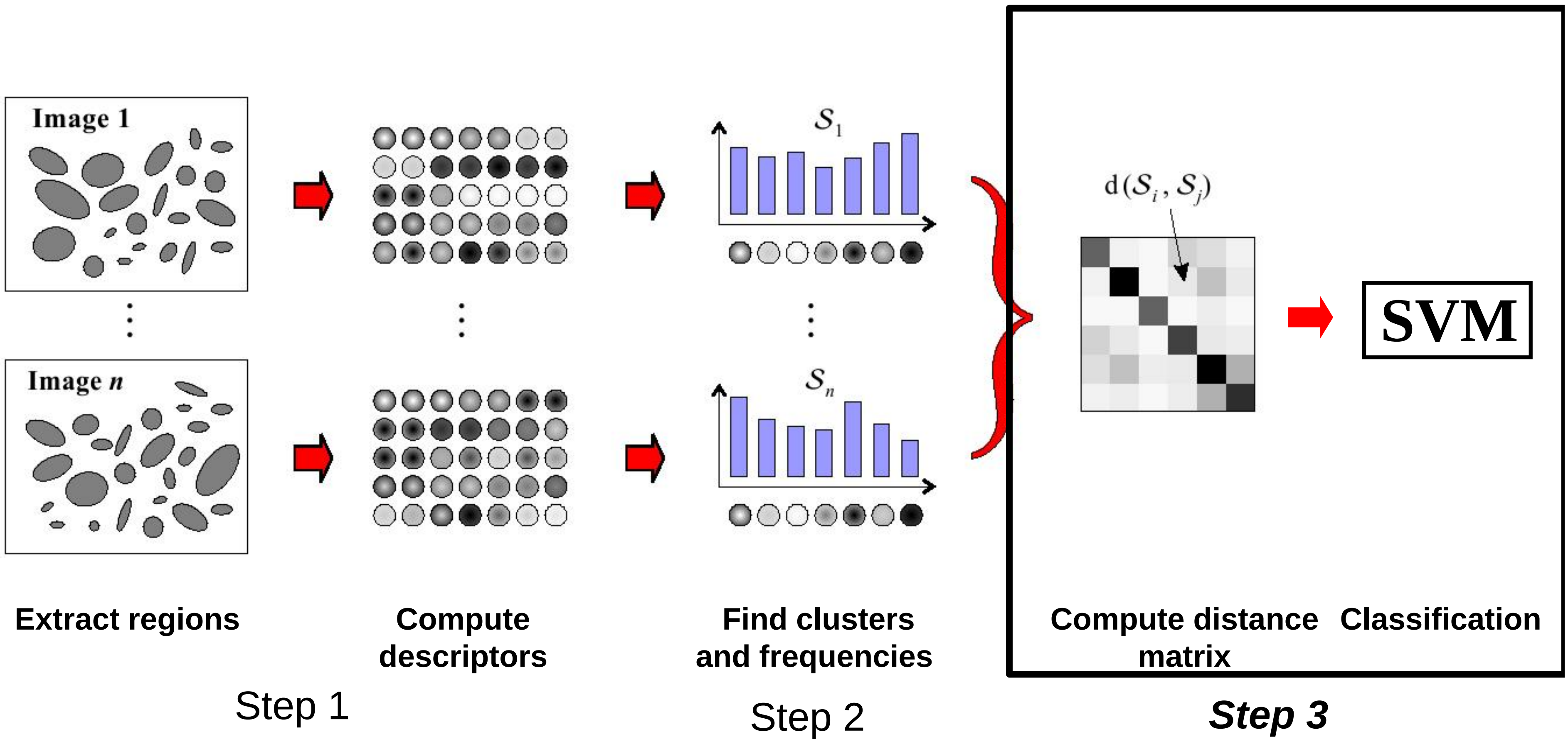


Image representation



• each image is represented by a vector, typically 1000-4000 dimension,

Bag-of-features for image classification



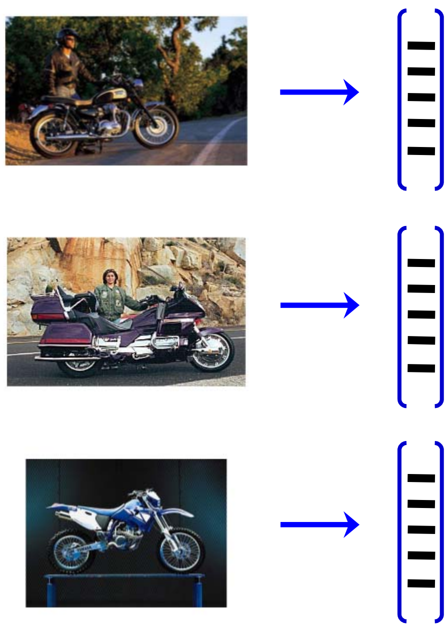
Step 3: Classification

- Learn a decision rule (classifier) assigning bag-of-features representations of images to different classes

Training data

Vectors are histograms, one from each training image

positive



Train classifier,e.g.SVM

Support Vector Machines (SVM)

Application

- Pattern recognition

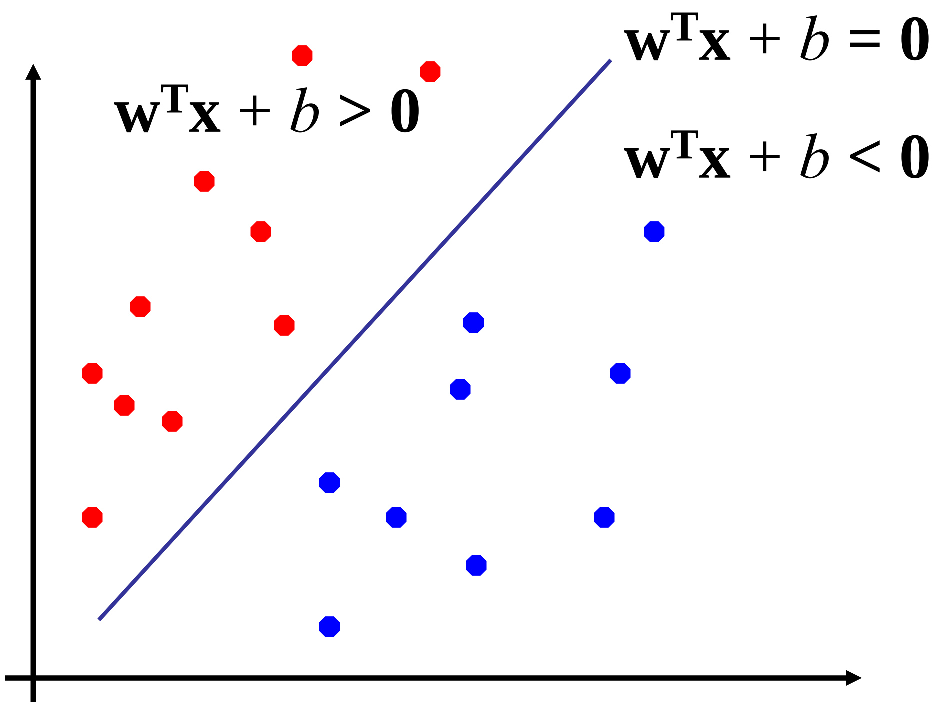
- Object classification/detection

Usage

- The classifier must be trained using a set of negative and positive examples.
- The classifier "learns" the regularities in the data
- If training was successful classifier is capable of classifying an unknown example with a high degree of accuracy.

Linear Classifier

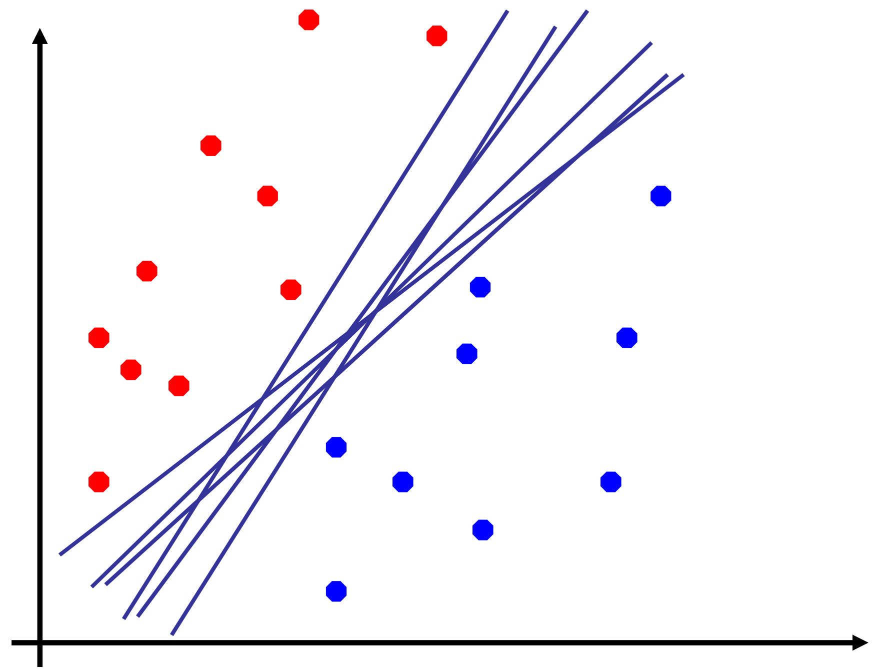
- Binary classifier → Task of separating classes in feature space



$$f(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^T\mathbf{x} + b)$$

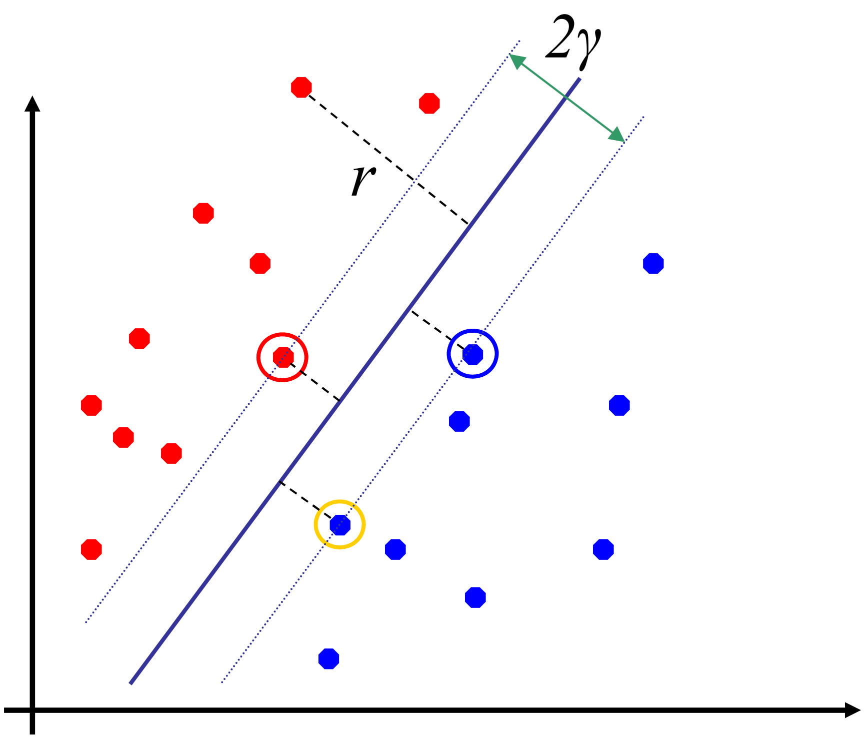
Linear Classifier cont'd

- Which of the linear separators is optimal?



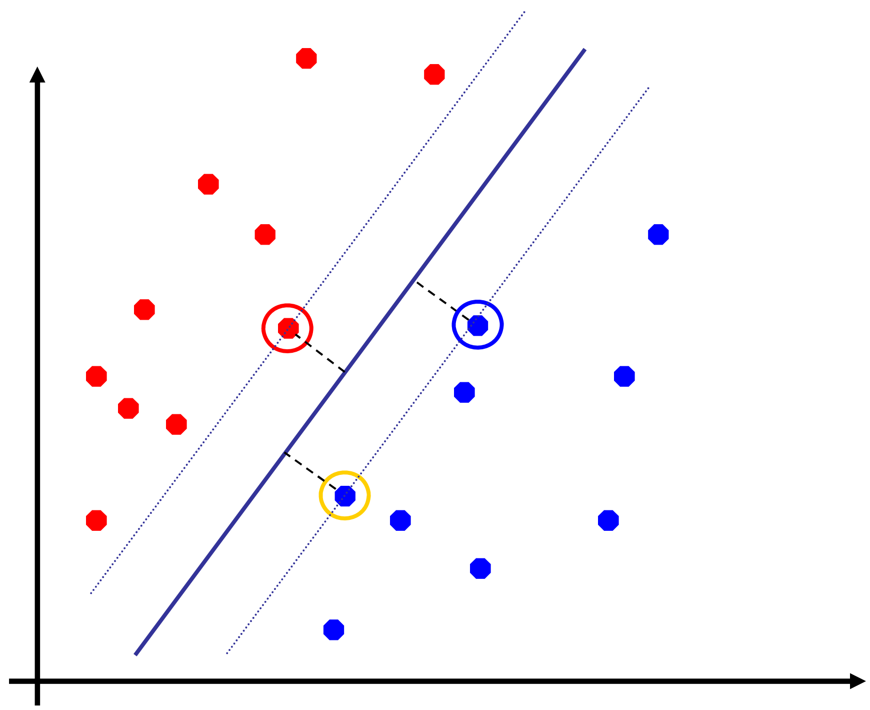
Margin

- Distance from example to the separator is (Point to Plane Distance Equation) $r = \dfrac{\mathbf{w}^T\mathbf{x} + b}{\|\mathbf{w}\|}$
- Examples closest to the hyperplane are ***support vectors***.
- ***Margin*** 2γ of the separator is the width of separation between classes.



Maximum Margin Classification

- Maximizing the margin is good according to intuition.
- Implies that only support vectors are important; other training examples are ignorable.



LibSVM

SVM implementation

- http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- http://www.cs.wisc.edu/dmi/svm/

Bag-of-features for image classification

- Excellent results in the presence of background clutter

bikes books building cars people phones trees

Examples for misclassified images

Books- misclassified into faces, faces, buildings

Buildings- misclassified into faces, trees, trees

Cars- misclassified into buildings, phones, phones

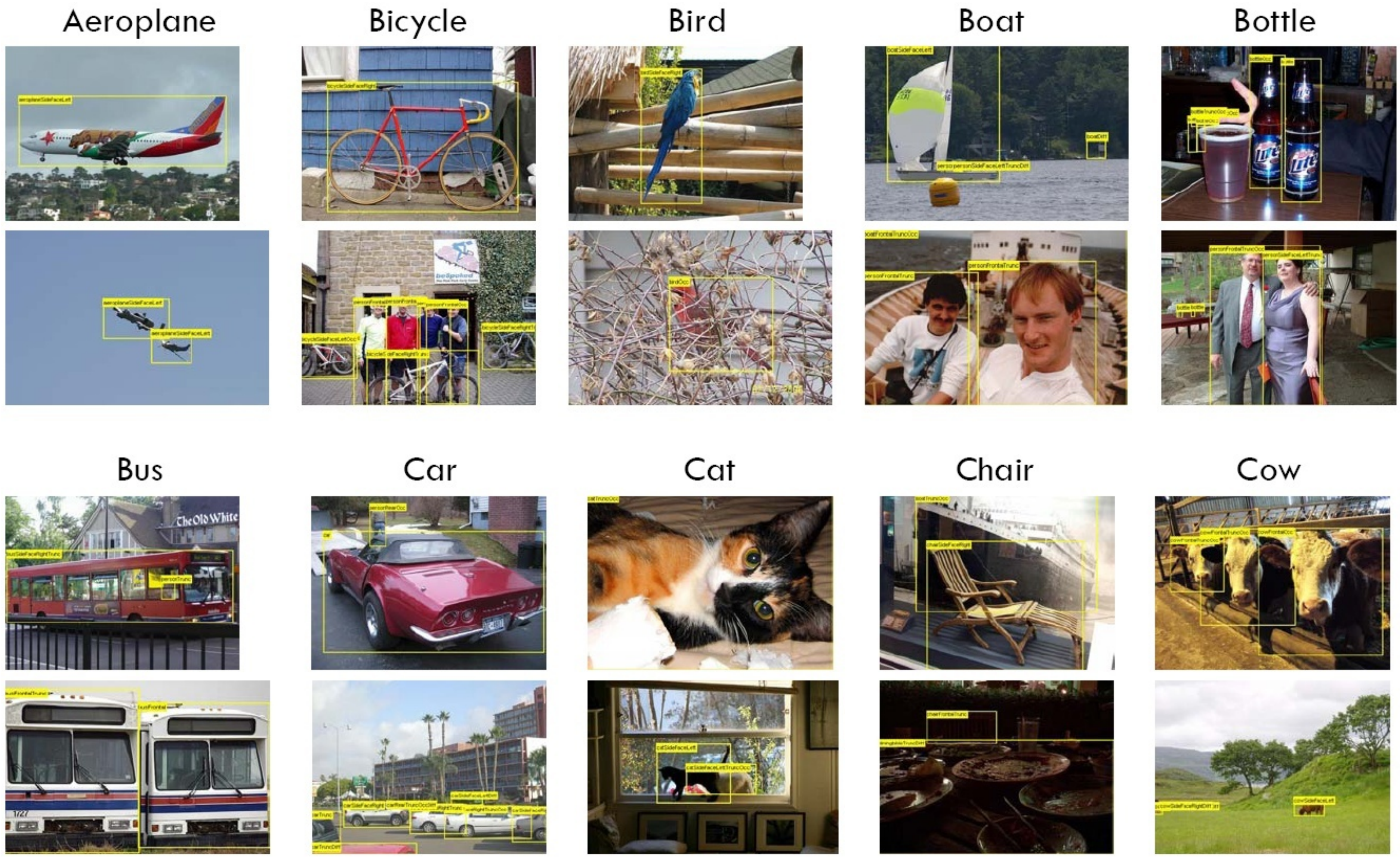
Evaluation of image classification

- PASCAL VOC [05-10] datasets

- PASCAL VOC 2007
  - Training *and* test dataset available
  - Used to report state-of-the-art results
  - Collected January 2007 from Flickr
  - 500 000 images downloaded and random subset selected
  - 20 classes
  - Class labels per image + bounding boxes
  - 5011 training images, 4952 test images

- Evaluation measure: average precision

PASCAL 2007 dataset



PASCAL 2007 dataset

Evaluation Metrics

$$\text{precision} = \frac{GT \cap RM}{RM} = \frac{TP}{RM}$$

$$\text{recall} = \frac{GT \cap RM}{GT} = \frac{TP}{GT}$$