计算机视觉
——目标识别

2022年春季

桑农，王岳环

目的

- 了解什么是目标识别？
- 熟悉两种目标识别方法
- 可在编程作业中应用

目标识别

- 什么是目标识别？
  - What is it?

    Object classification

  - Where is it?

    Object detection

  - Who is it?
    - Object Identification

作业

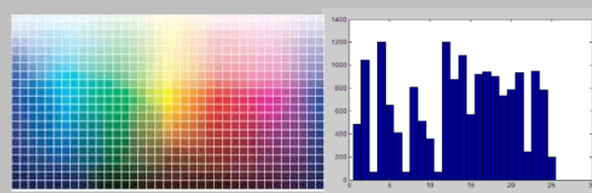
- 分析HOG、颜色直方图、PCA、shapecontext、特征点+BOW这几种表达的平移、旋转（平面内）、尺度、遮挡不变性，并分析原因

	平移	旋转	尺度	遮挡
HOG	✕	✕	✕	✕
颜色直方图	✔	✔	✔	✕
PCA	✔	✔	✔？	✕
shapecontext	✔	✔	✔	✔？
特征点+BOW	✔	✔？	✔？	✕

图像特征

Input image

<u>Color:</u> Quantize RGB values



<u>Invariance?</u>

? Translation

? Scale

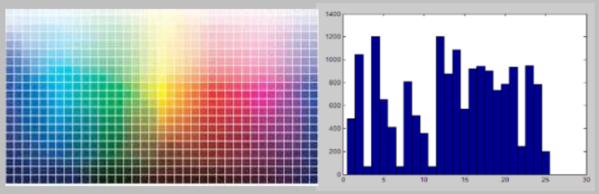
? Rotation

? Occlusion

图像特征

Input image

Color: Quantize RGB values



Invariance?

Translation

? Scale

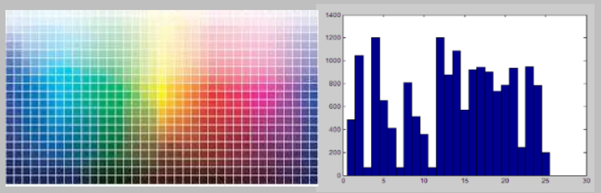
Rotation (in-planar)

Occlusion

图像特征

Input image

<u>Color:</u> Quantize RGB values



<u>Invariance?</u>
- 🙂 Translation
- 🙂 Scale
- 🙂 Rotation (in-planar)
- 🙁 Occlusion

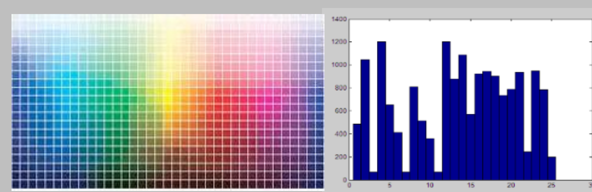
<u>Global shape:</u> PCA space

<u>Invariance?</u>
- ? Translation
- ? Scale
- ? Rotation (in-planar)
- ? Occlusion

图像特征

Input image

Color: Quantize RGB values



Invariance?
- Translation
- Scale
- Rotation (in-planar)
- Occlusion

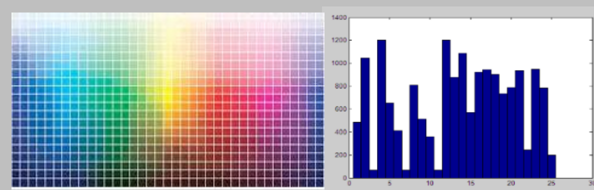
Global shape: PCA space

Invariance?
- Translation
- ? Scale
- Rotation (in-planar)
- Occlusion

图像特征

Input image

Color: Quantize RGB values

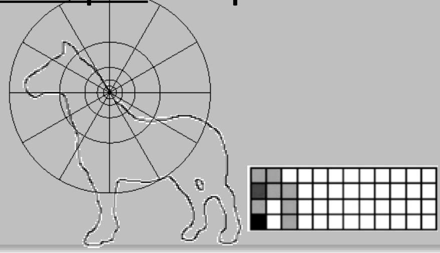


Invariance?
- 🙂 Translation
- 🙂 Scale
- 🙂 Rotation
- 🙁 Occlusion

Global shape: PCA space

Invariance?
- 🙂 Translation
- ? Scale
- 🙂 Rotation
- 🙁 Occlusion

Local shape: shape context

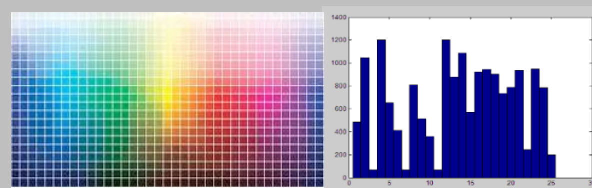


Invariance?
- ? Translation
- ? Scale
- ? Rotation (in-planar)
- ? Occlusion

图像特征

Input image

<u>Color:</u> Quantize RGB values

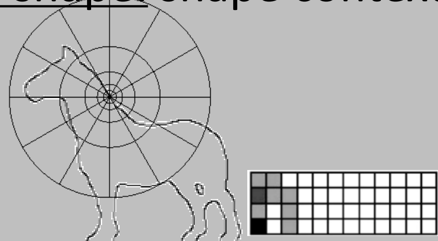


<u>Invariance?</u>
- Translation
- Scale
- Rotation
- Occlusion

<u>Global shape:</u> PCA space

<u>Invariance?</u>
- Translation
- ? Scale
- Rotation
- Occlusion

<u>Local shape:</u> shape context

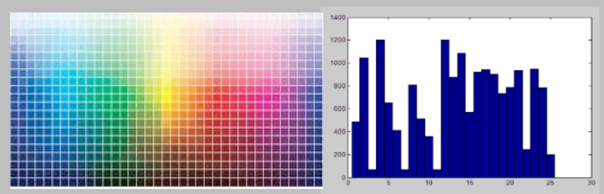


<u>Invariance?</u>
- Translation
- Scale
- ? Rotation (in-planar)
- Occlusion

图像特征

Input image

Color: Quantize RGB values

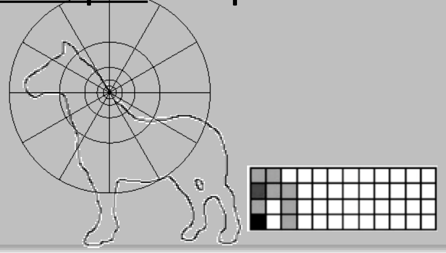


Invariance?
- Translation
- Scale
- Rotation
- Occlusion

Global shape: PCA space

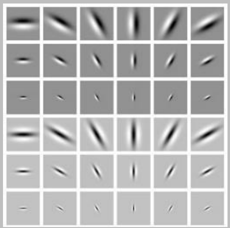
Invariance?
- Translation
- ? Scale
- Rotation
- Occlusion

Local shape: shape context



Invariance?
- Translation
- Scale
- ? Rotation (in-planar)
- Occlusion

Texture: Filter banks

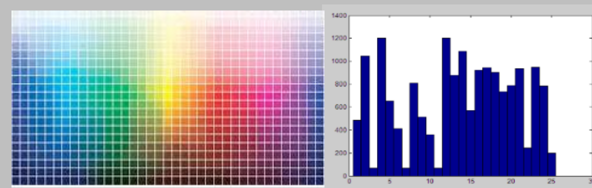


Invariance?
- ? Translation
- ? Scale
- ? Rotation (in-planar)
- ? Occlusion

图像特征

Input image

Color: Quantize RGB values

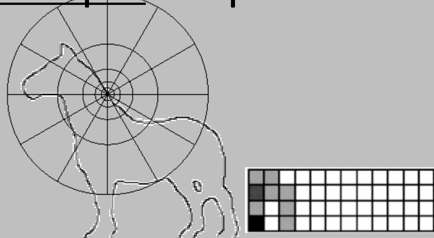


Invariance?
- Translation
- Scale
- Rotation
- Occlusion

Global shape: PCA space

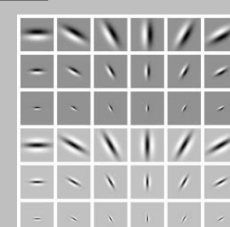
Invariance?
- Translation
- ? Scale
- Rotation
- Occlusion

Local shape: shape context



Invariance?
- Translation
- Scale
- ? Rotation (in-planar)
- Occlusion

Texture: Filter banks



Invariance?
- Translation
- ? Scale
- ? Rotation (in-planar)
- Occlusion

目标识别（分类）

- 任务：通用目标识别

目标识别（分类）

- 任务：通用目标识别

horse

person

目标识别（分类）

- 基于特征点的目标识别

- 局部特征 + BOW

Bag-of-words目标识别框架

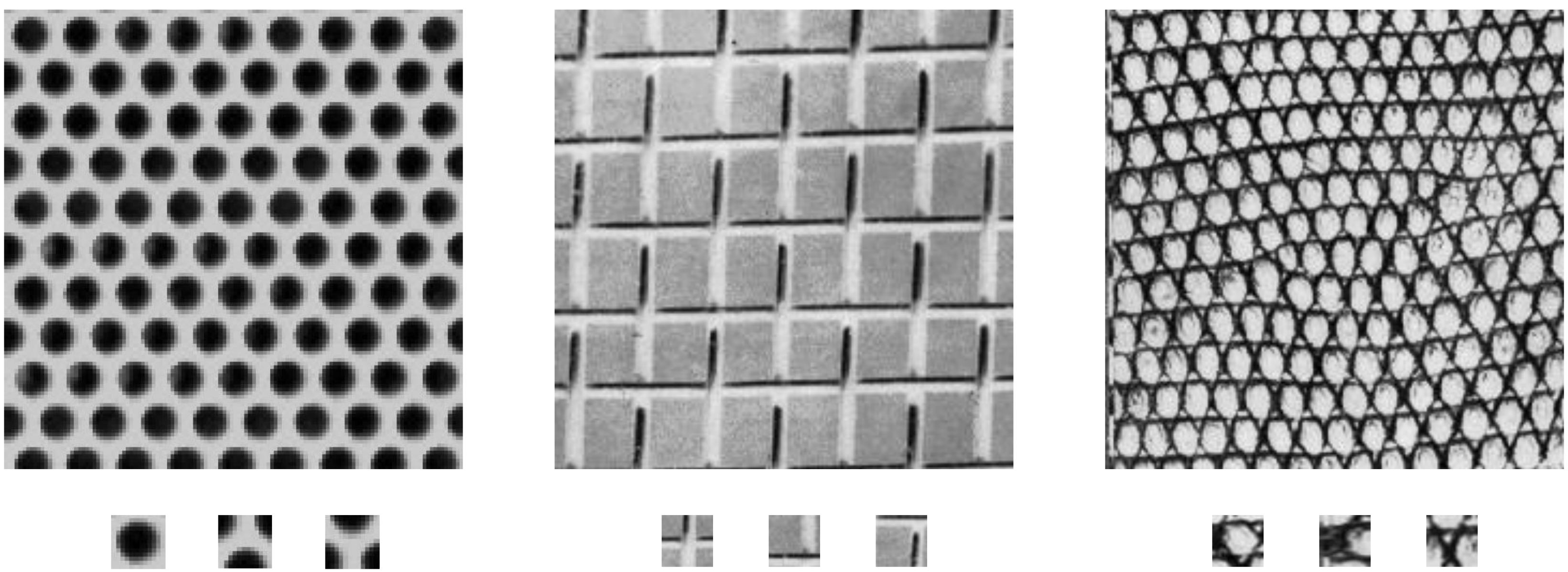
Object → **Bag of ‘words’**

(Li Fei-Fei, ICCV 2005)

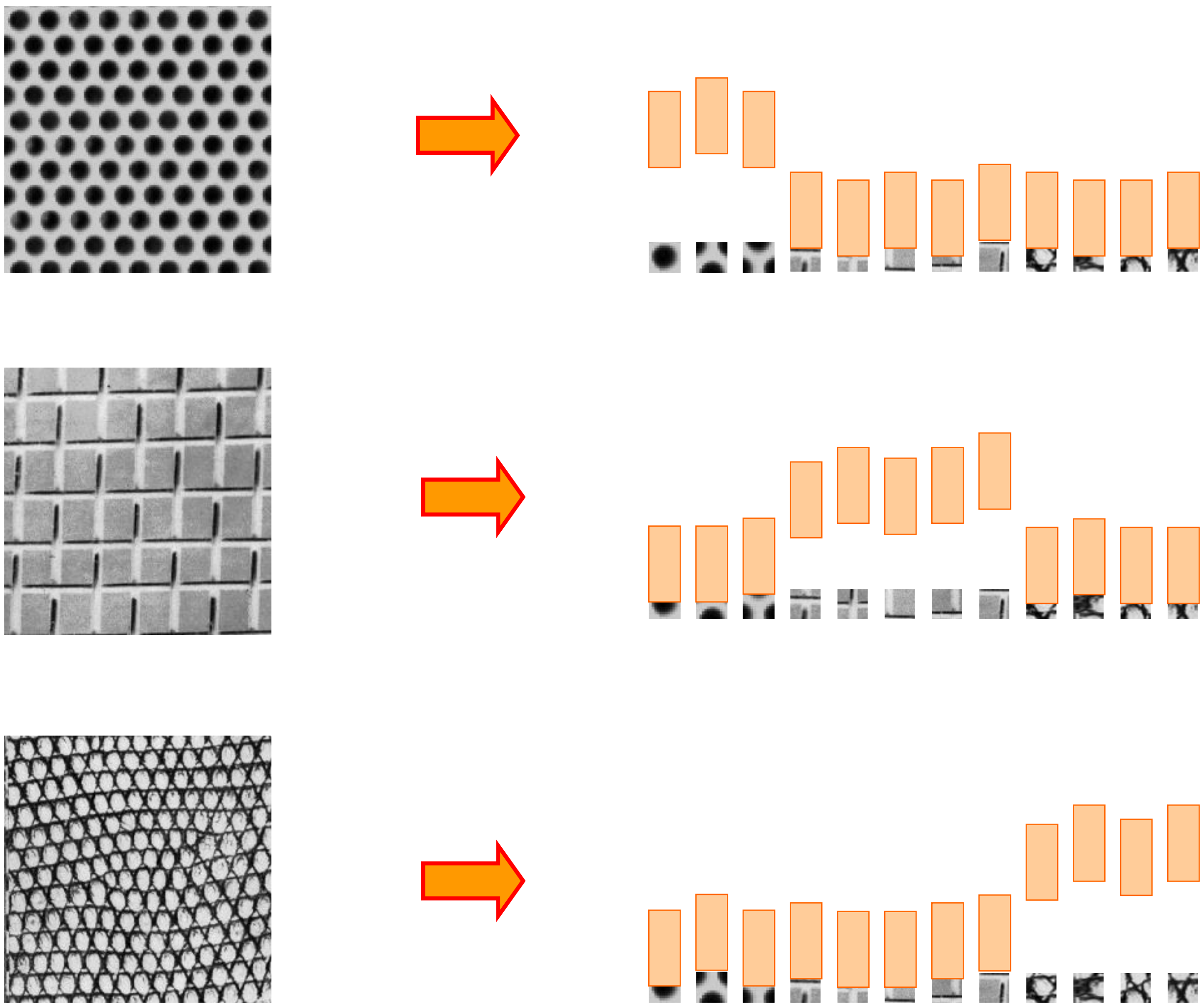
Bag-of-words目标识别框架

Bag-of-words目标识别框架

– 纹理是由一些重复的基元（texton）组成的
– 对于有明显统计特征的纹理来讲，我们只需要确认其基元，而不需要确定空间排列

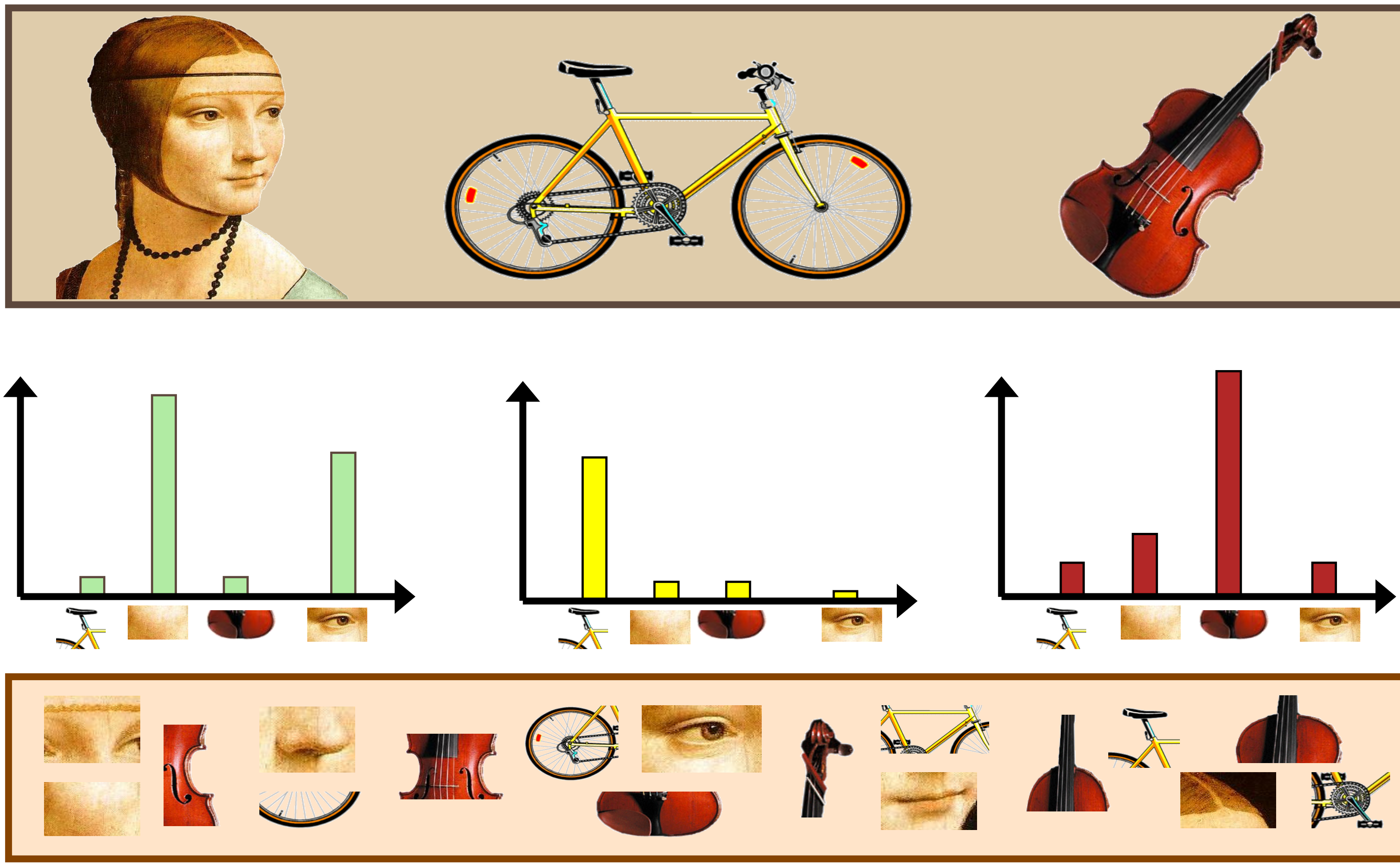


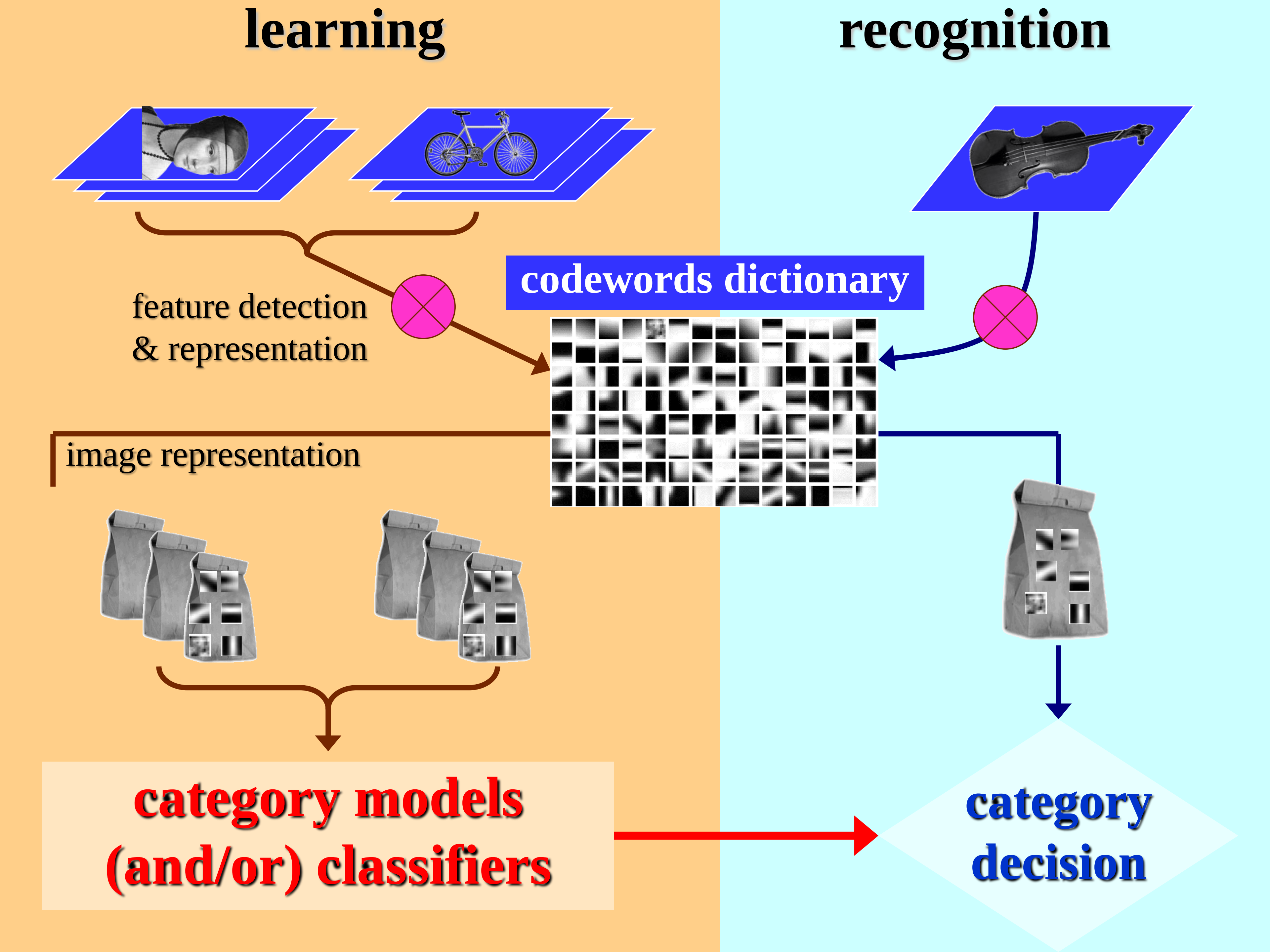
Bag-of-words目标识别框架



Bag-of-words目标识别框架

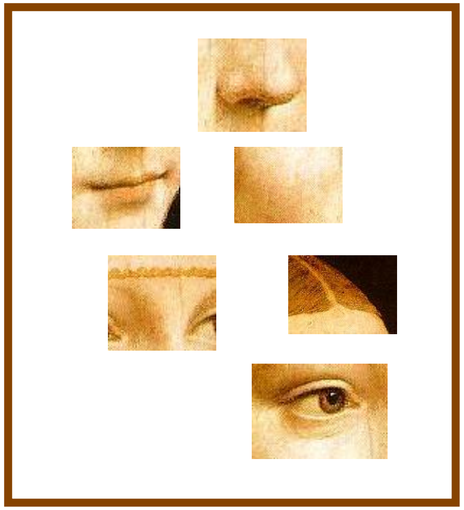
- Bag-of-words框架




learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

基于Bags of features的目标识别

1. 特征提取



基于Bags of features的目标识别

1. 特征提取
2. 学习视觉字典

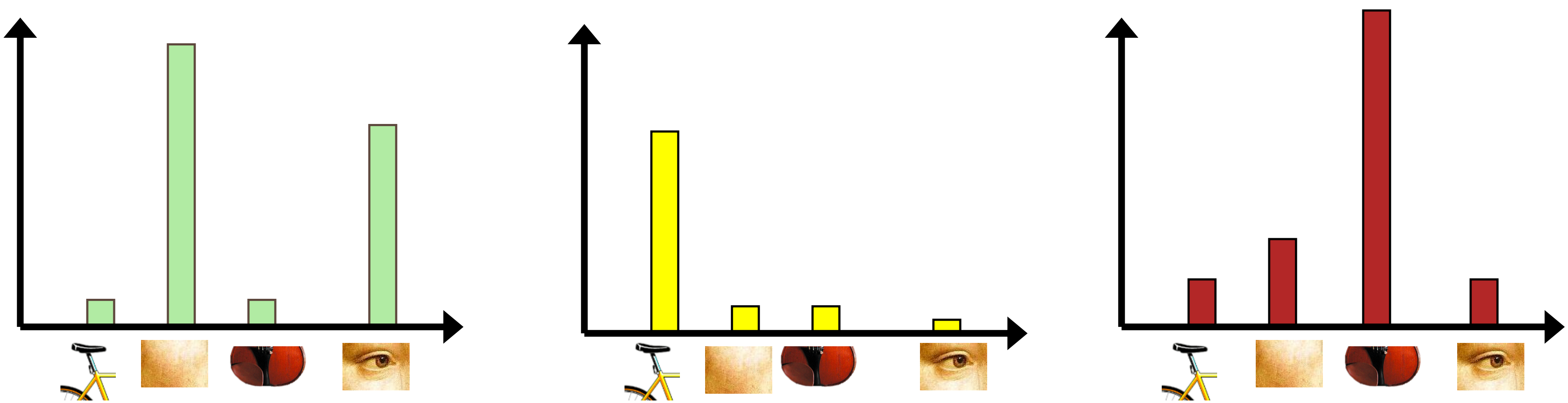
基于Bags of features的目标识别

1. 特征提取
2. 学习视觉字典
3. 利用视觉字典对特征进行量化

基于Bags of features的目标识别

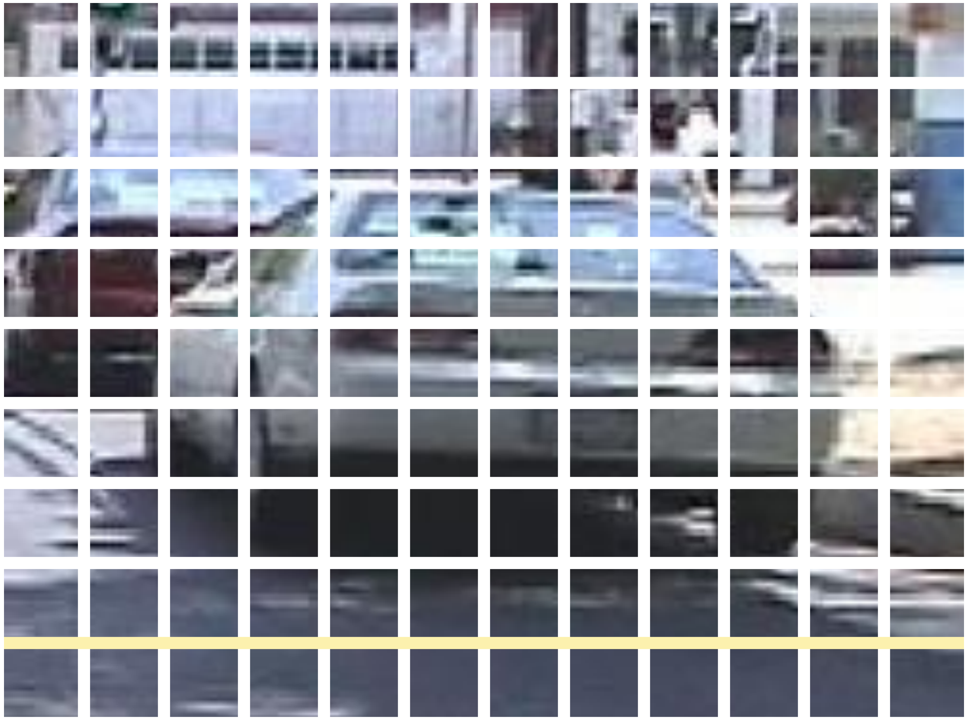
1. 特征提取
2. 学习视觉字典
3. 利用视觉字典对特征进行量化
4. 将图像表达为视觉字典出现的频率

图像表达



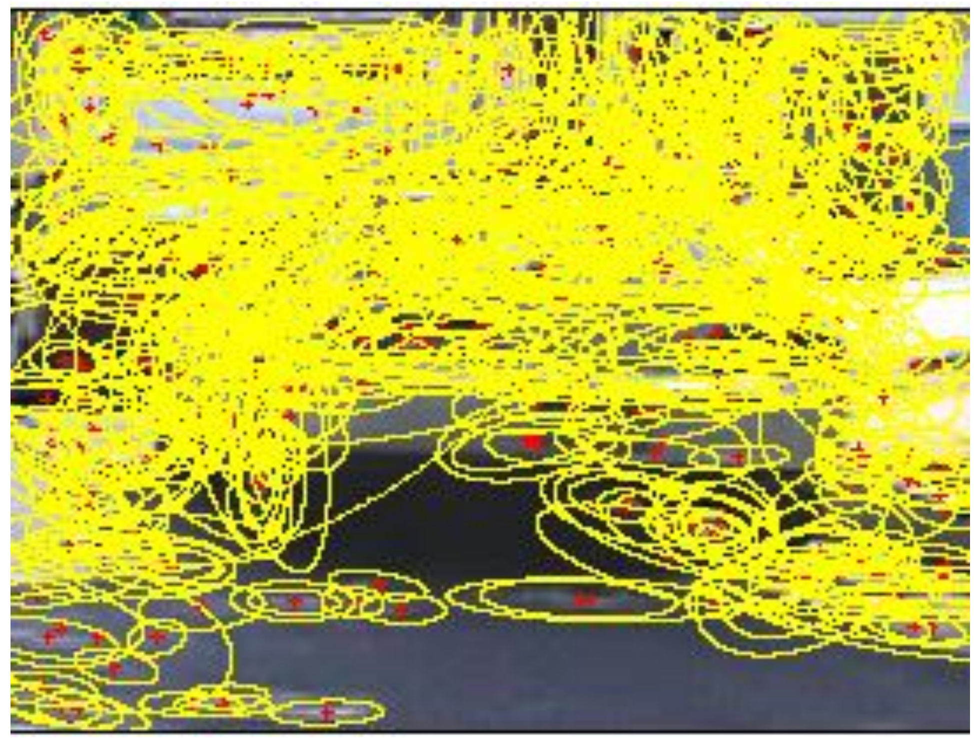
特征提取

1. 规则的网格区域



特征提取

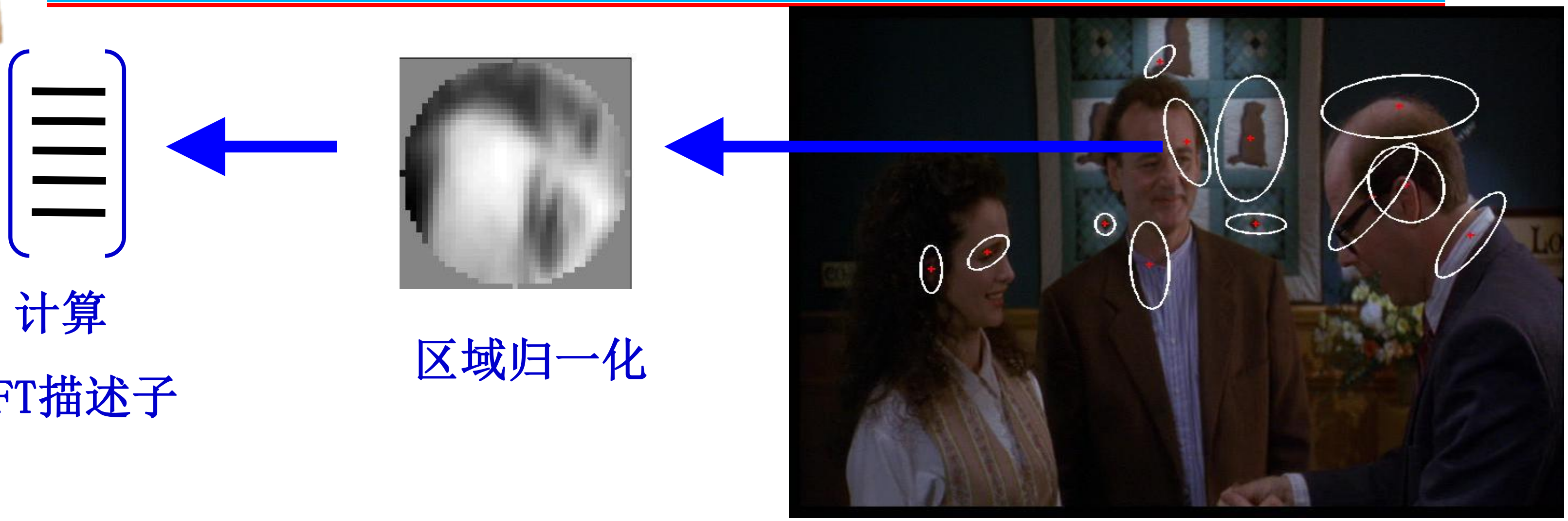
1. 规则的网格区域
2. 感兴趣点



特征提取

1. 规则的网格区域
2. 感兴趣点
3. 其他方法（随机采样、基于分割的区域）

特征提取



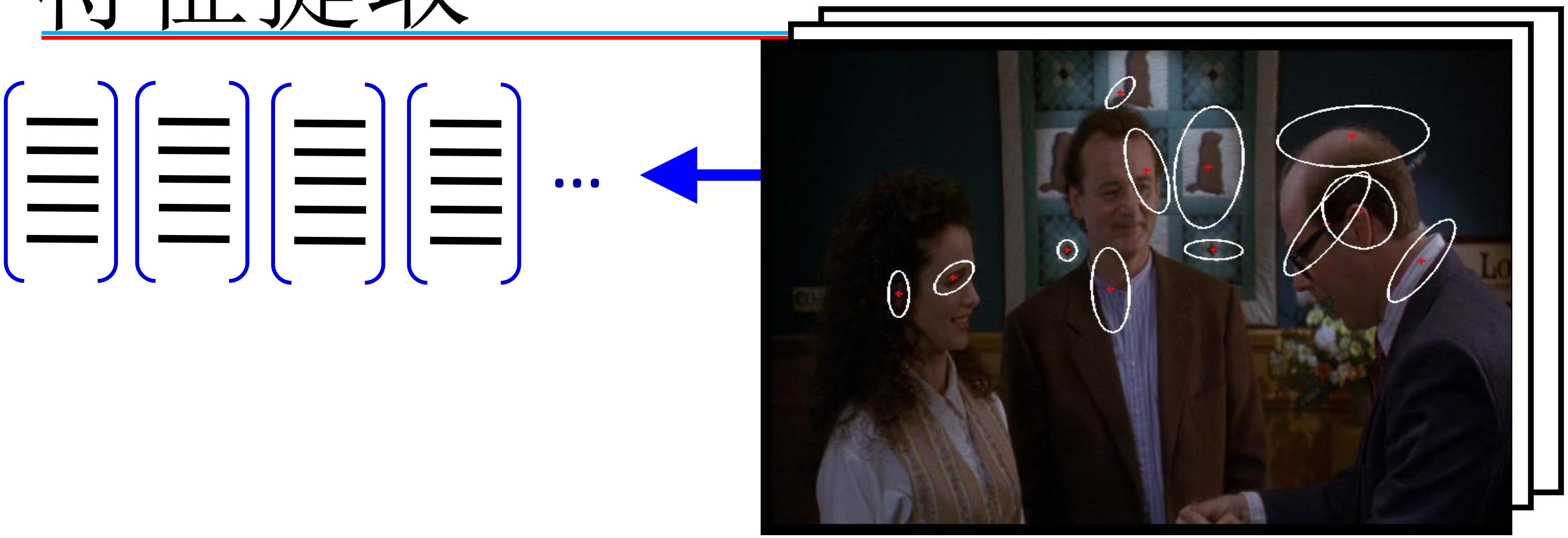
计算 SIFT描述子

区域归一化

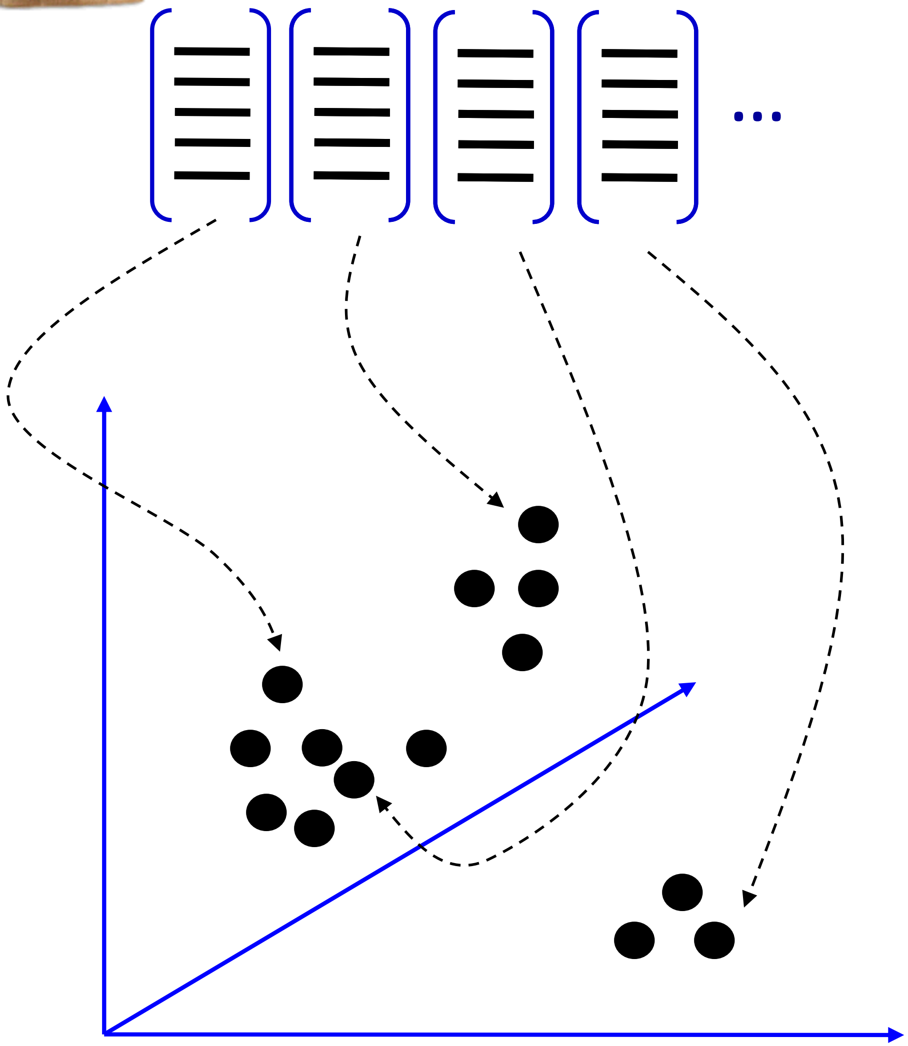
感兴趣点检测

Slide credit: Josef Sivic

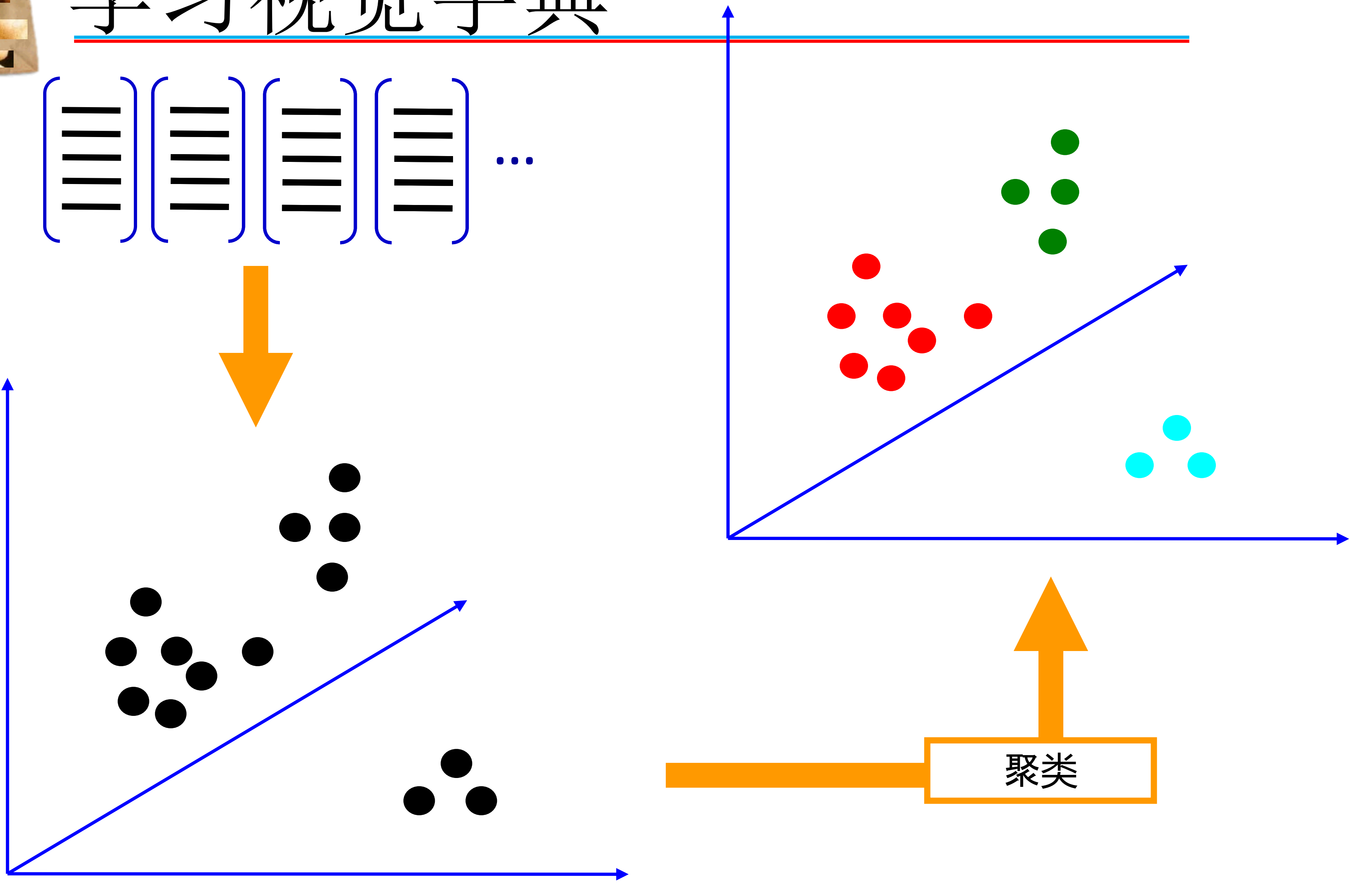
特征提取



学习视觉字典

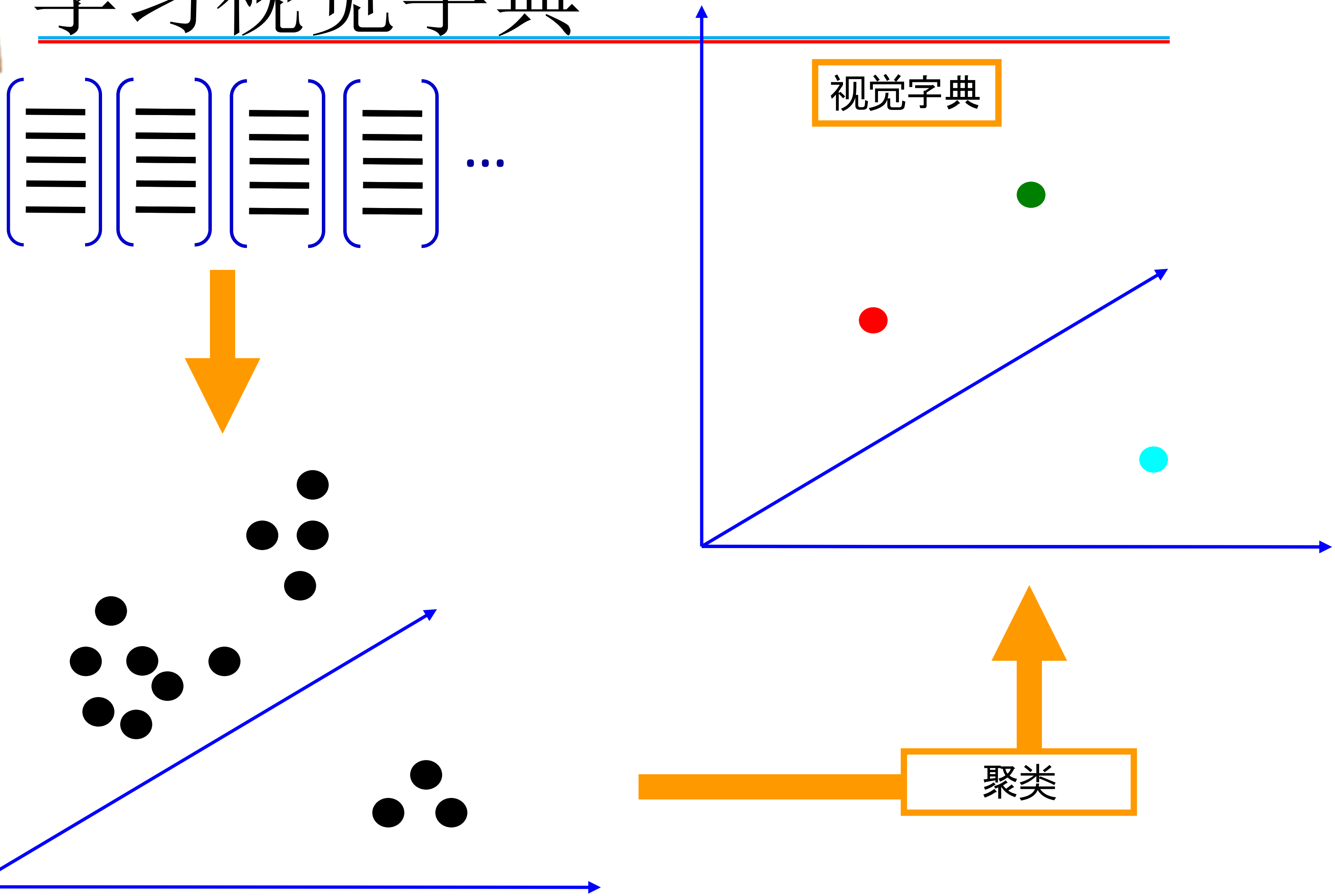


学习视觉字典



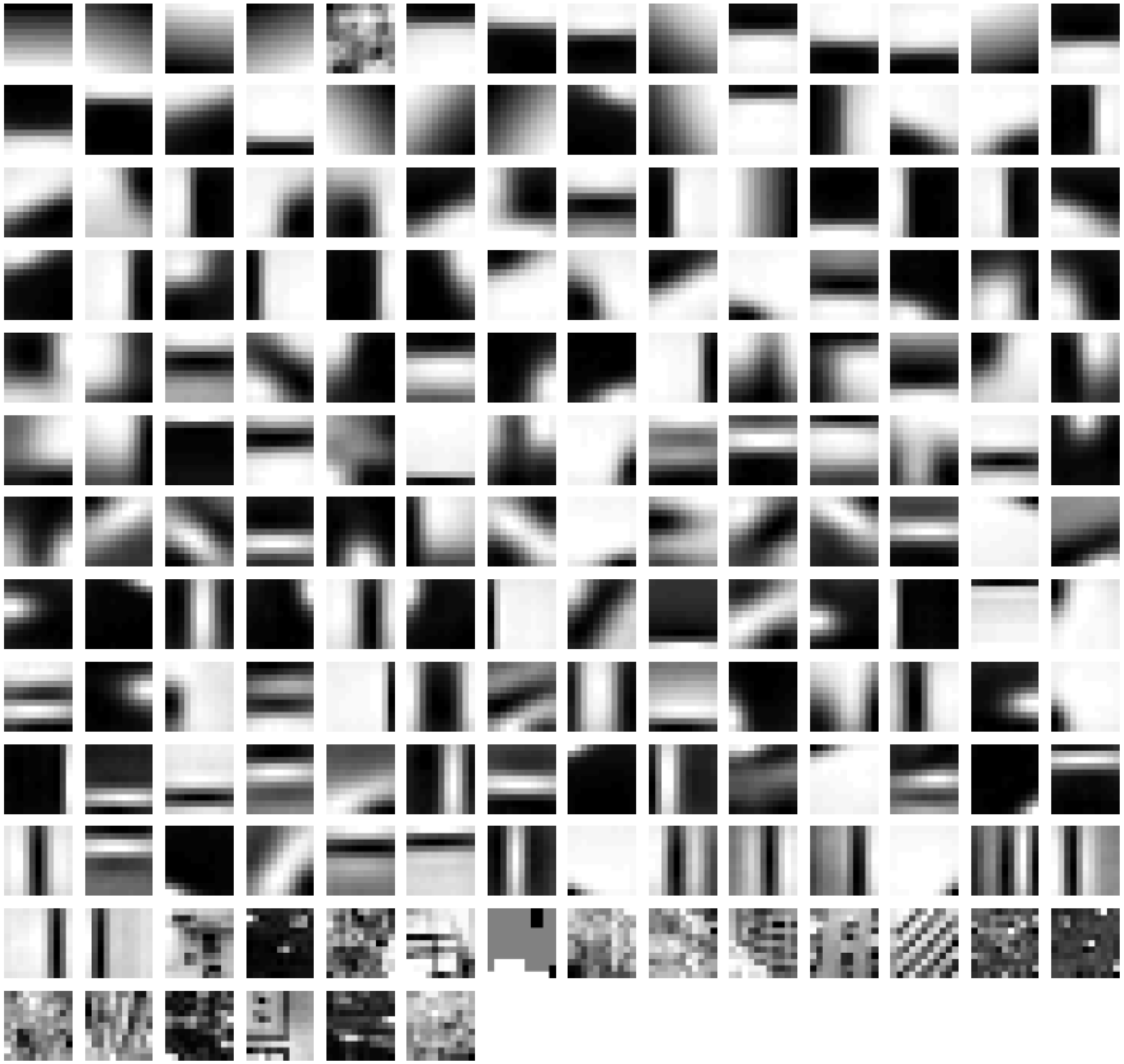
Slide credit: Josef Sivic

学习视觉字典



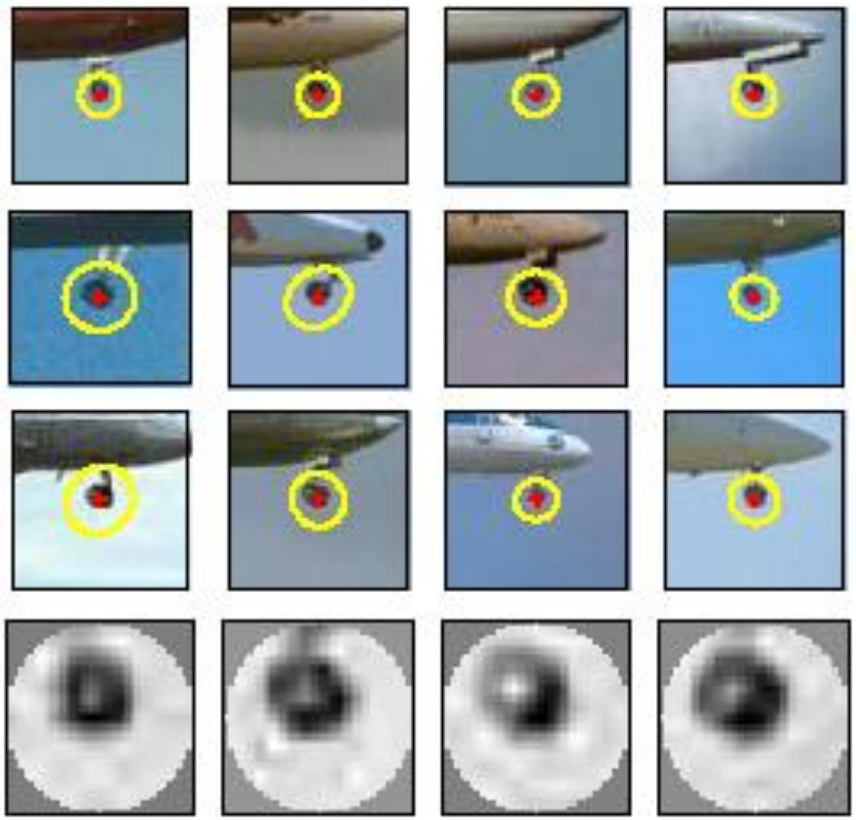
Slide credit: Josef Sivic

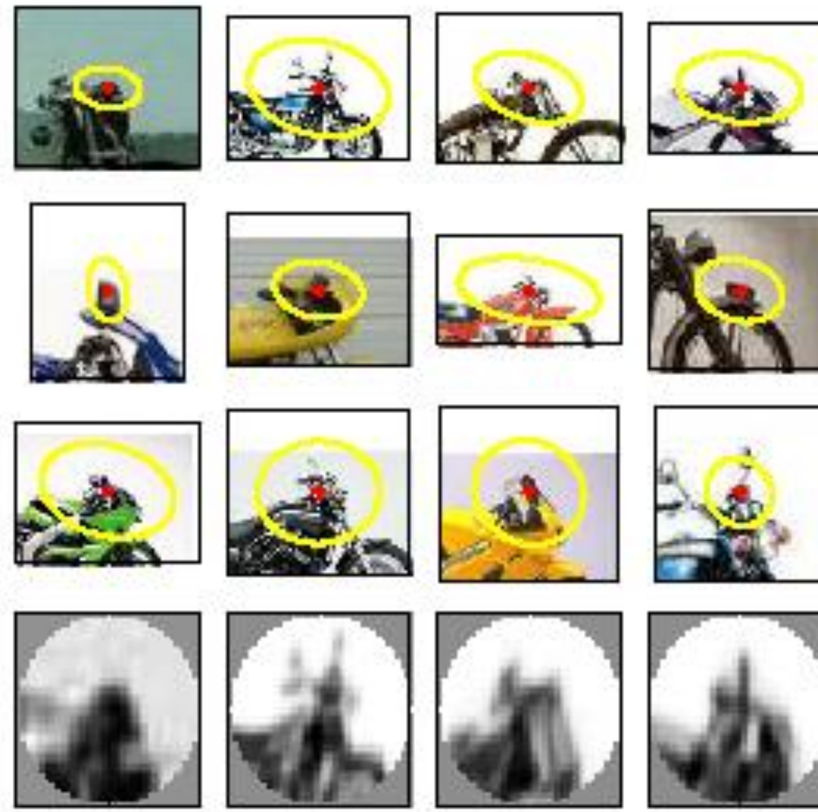
视觉字典举例



Fei-Fei et al. 2005

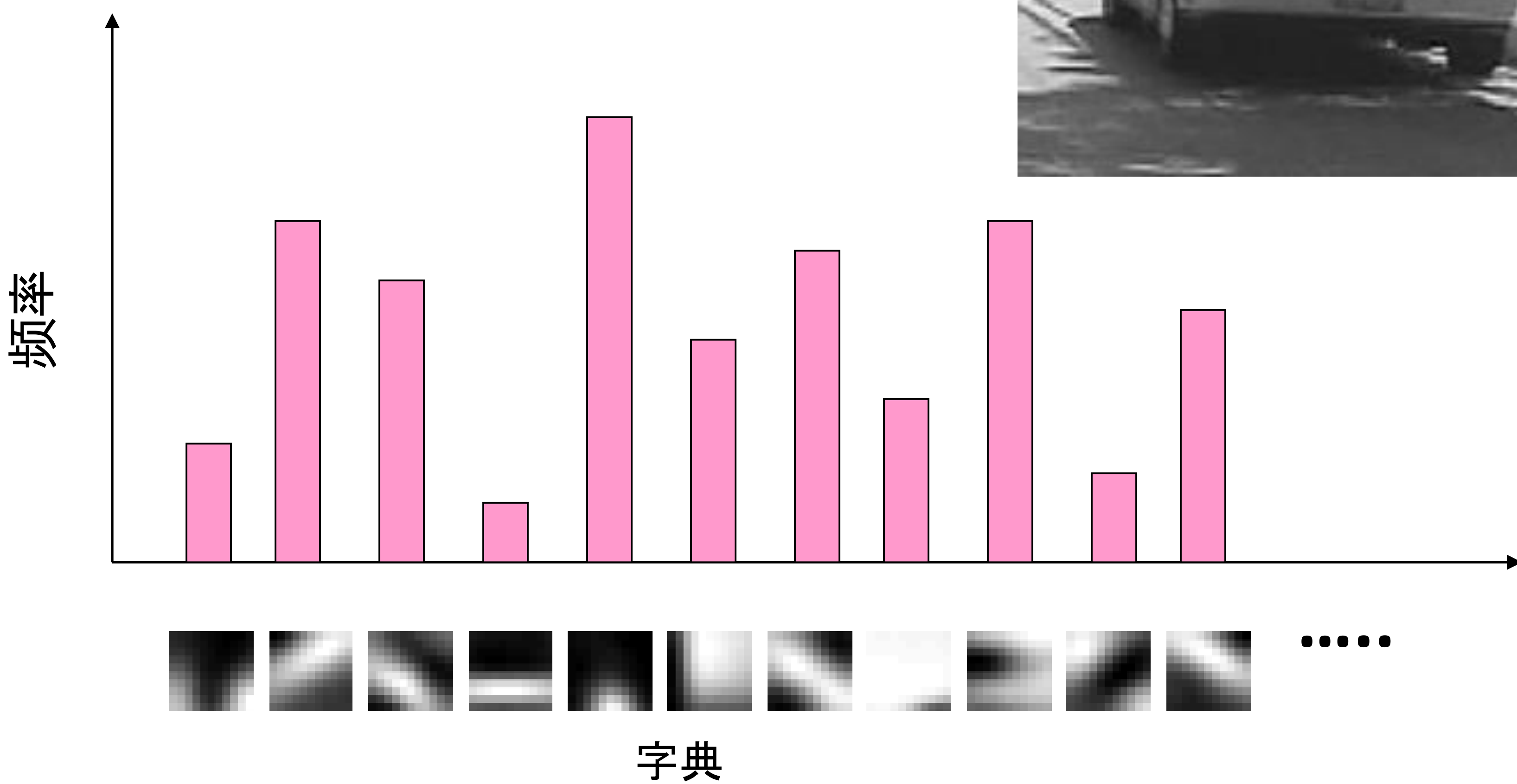
字典对应的图像区域





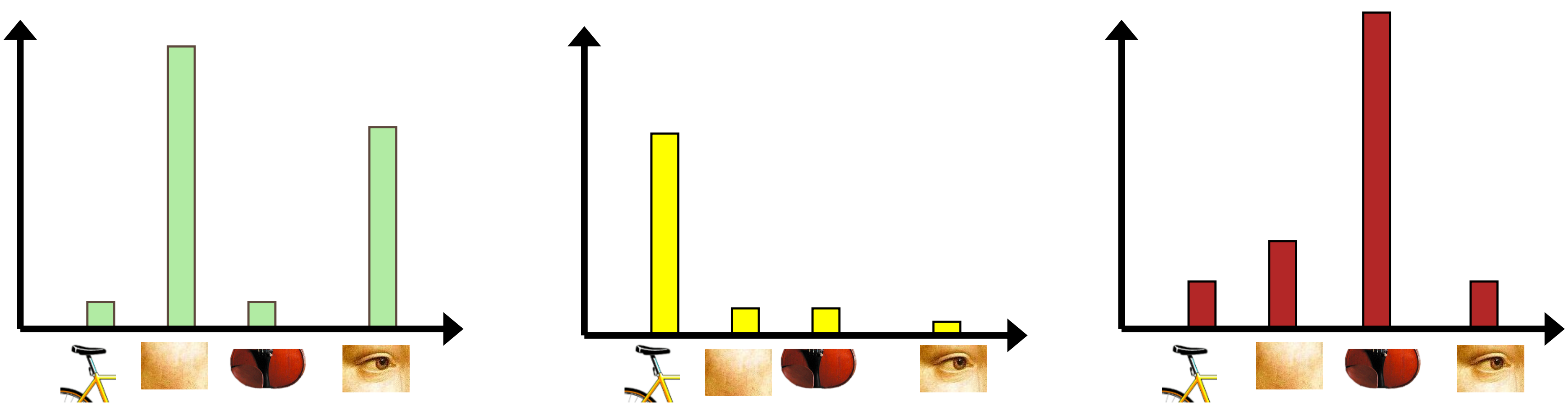
Sivic et al. 2005

图像表达



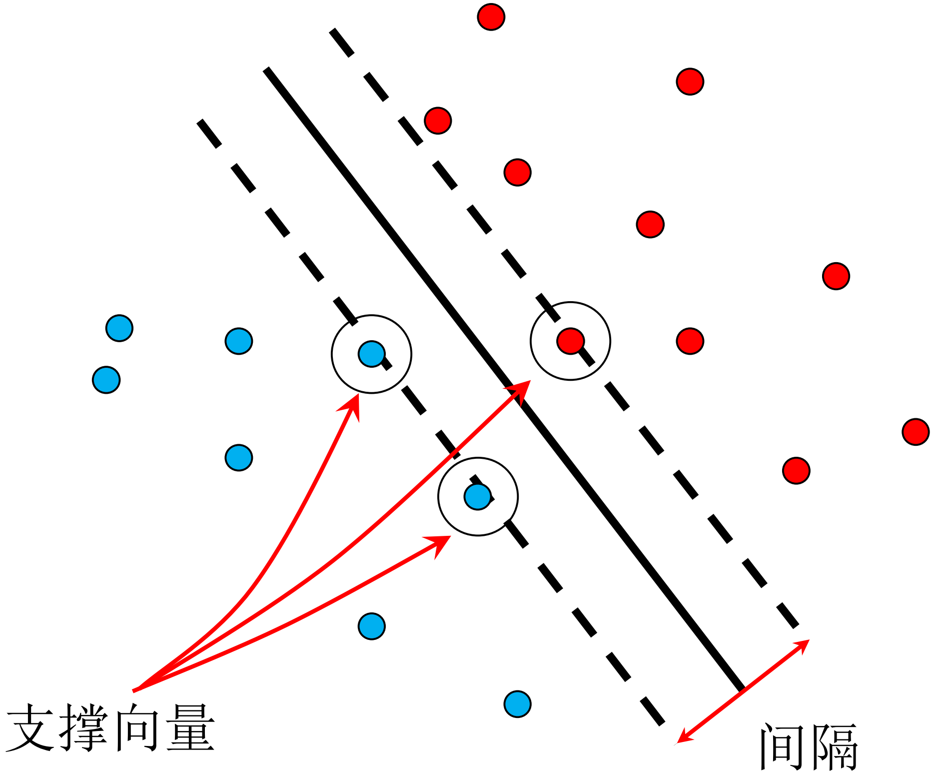
图像表达

- 得到不同类型目标的bag-of-features的表达后，我们如何对它们进行区分？



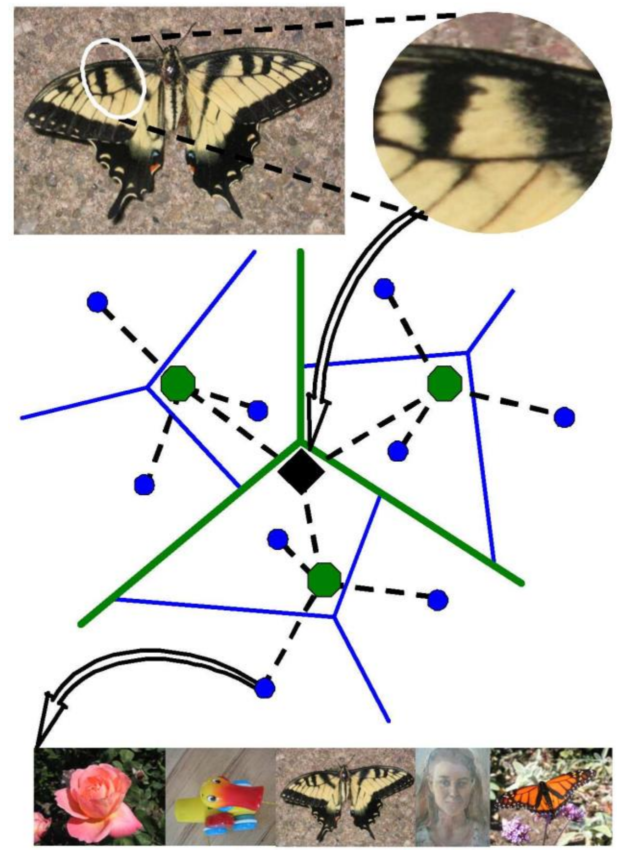
目标分类

- 产生式与判别式
  - 最近邻分类器
  - K近邻分类器
  - SVM分类器



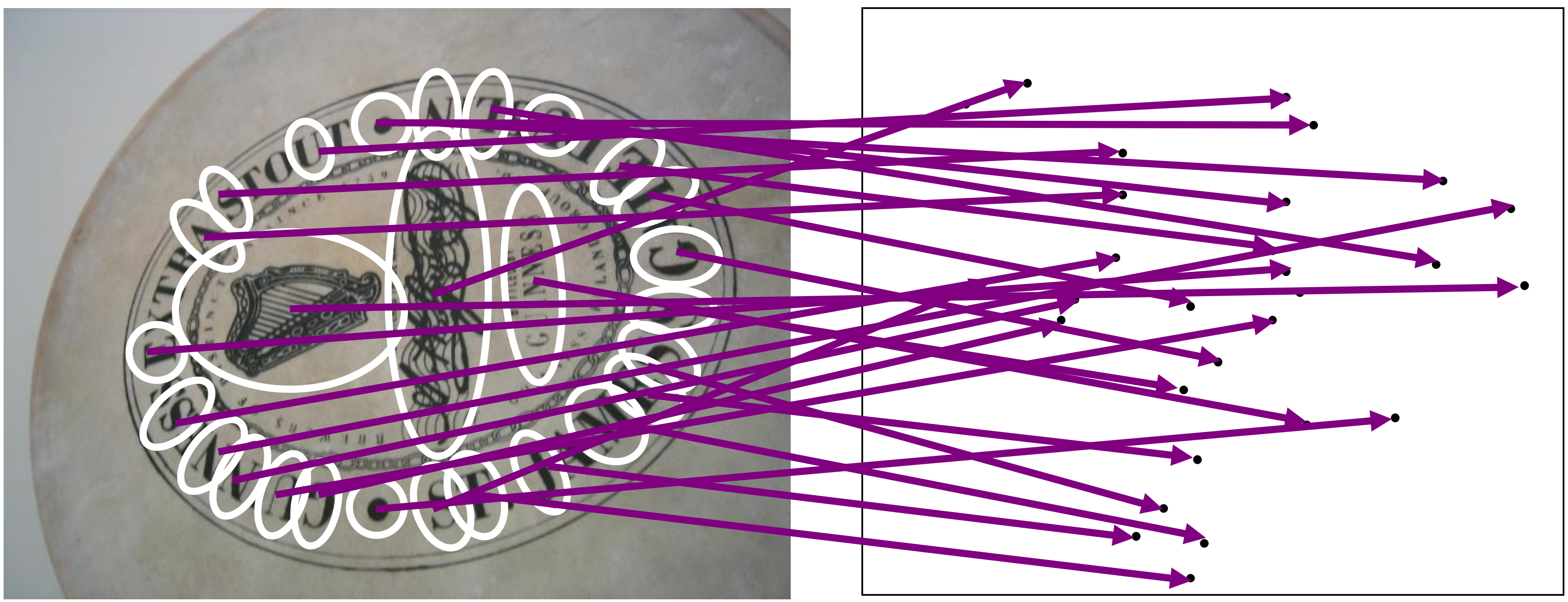
问题

- 如何选择字典大小
- 产生式还是判别式
- 计算效率
  - 字典树
    (Scalable Recognition with
    a Vocabulary Tree
    Nister & Stewenius, 2006)

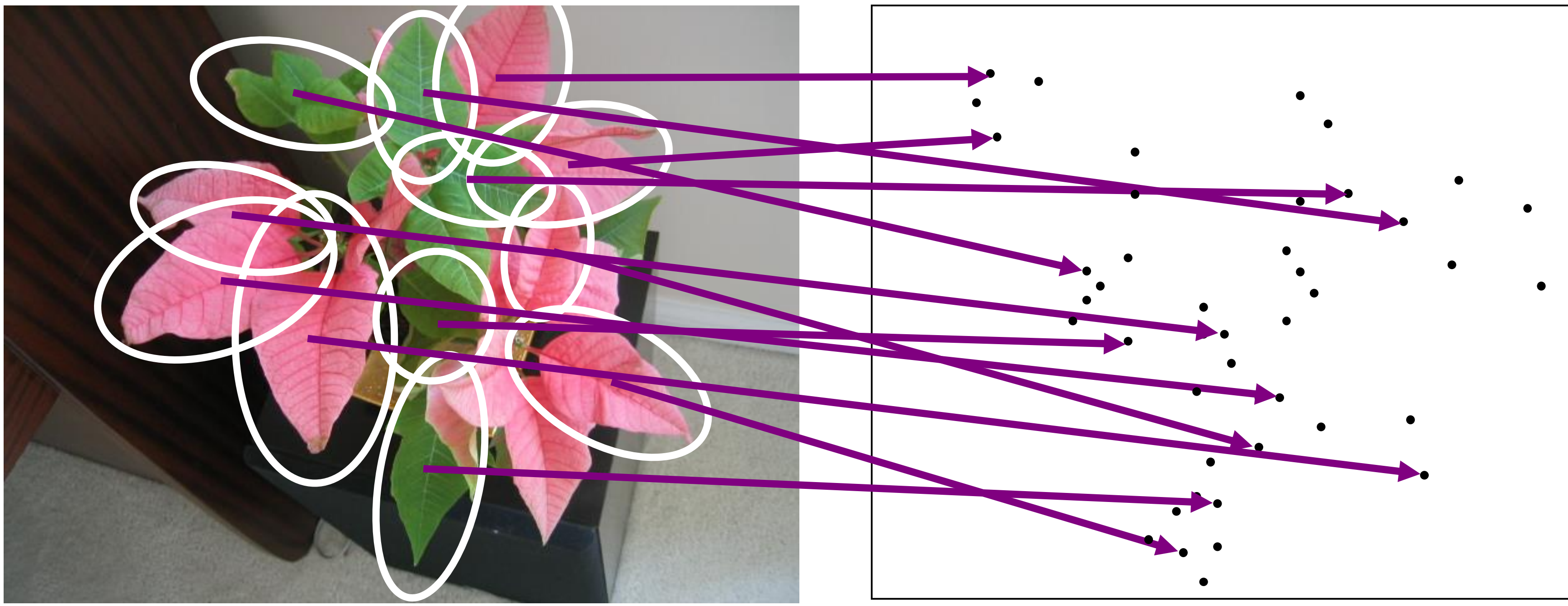


示例：基于字典树的识别

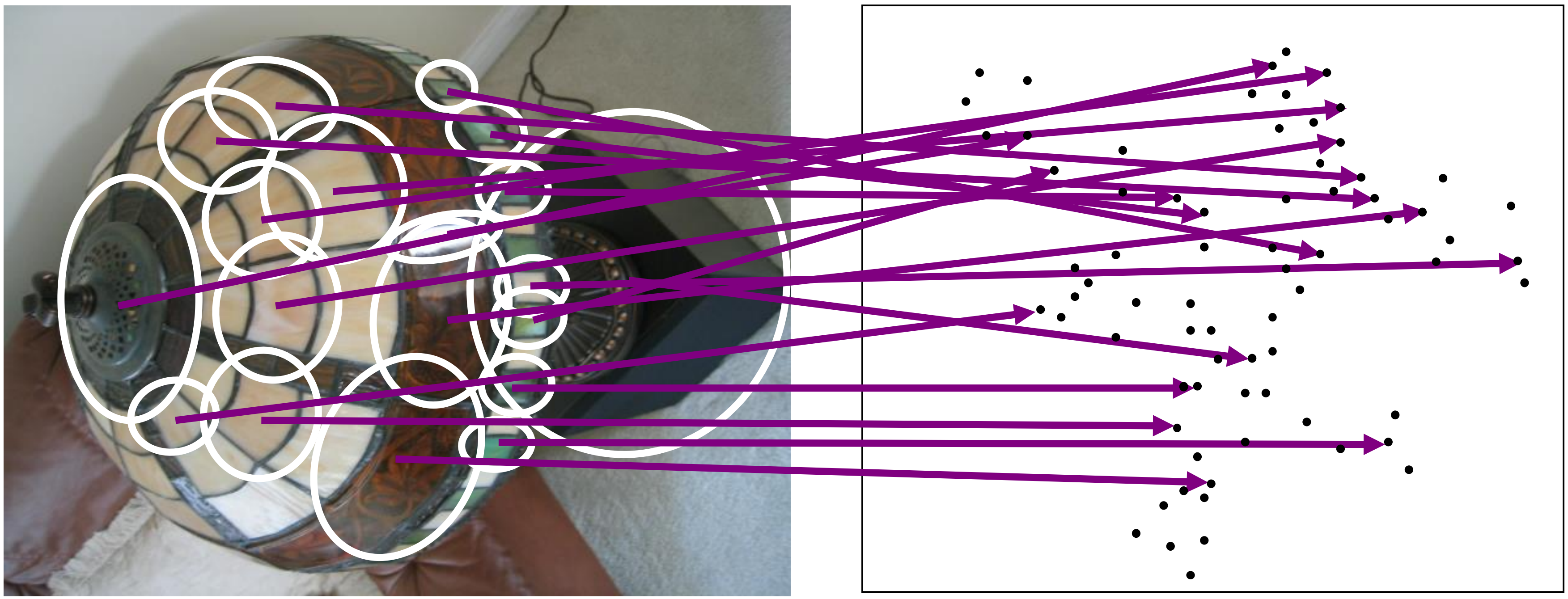
- 提取图像的局部特征



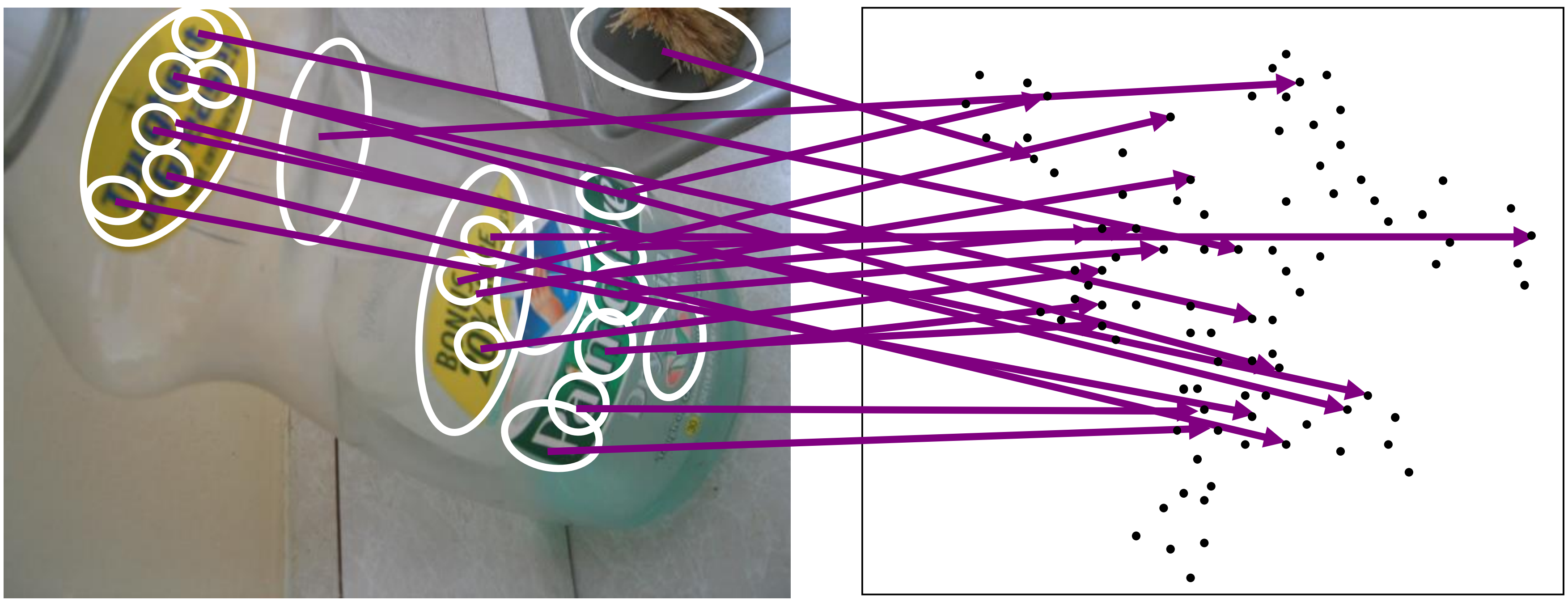
示例：基于字典树的识别

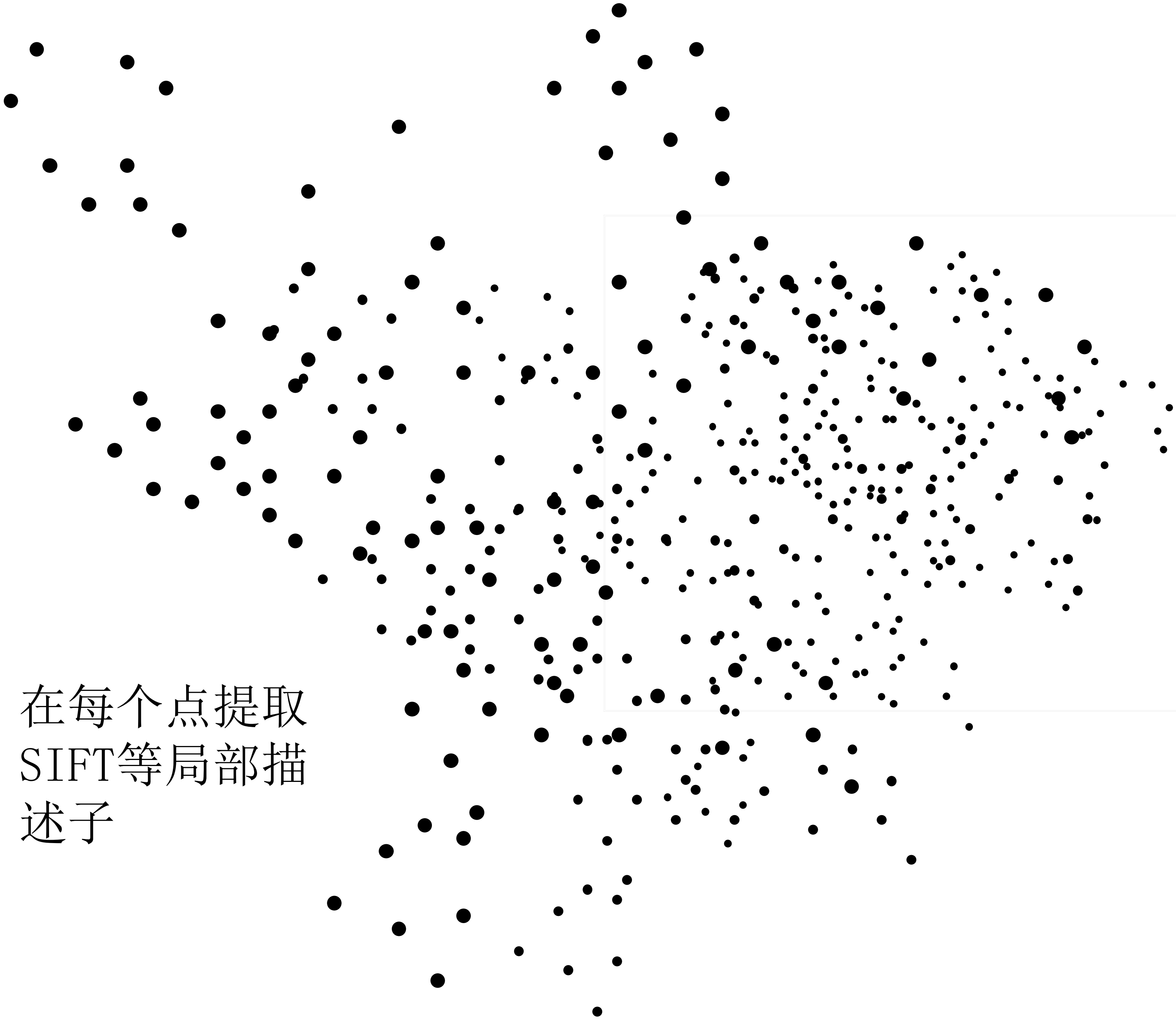


示例：基于字典树的识别

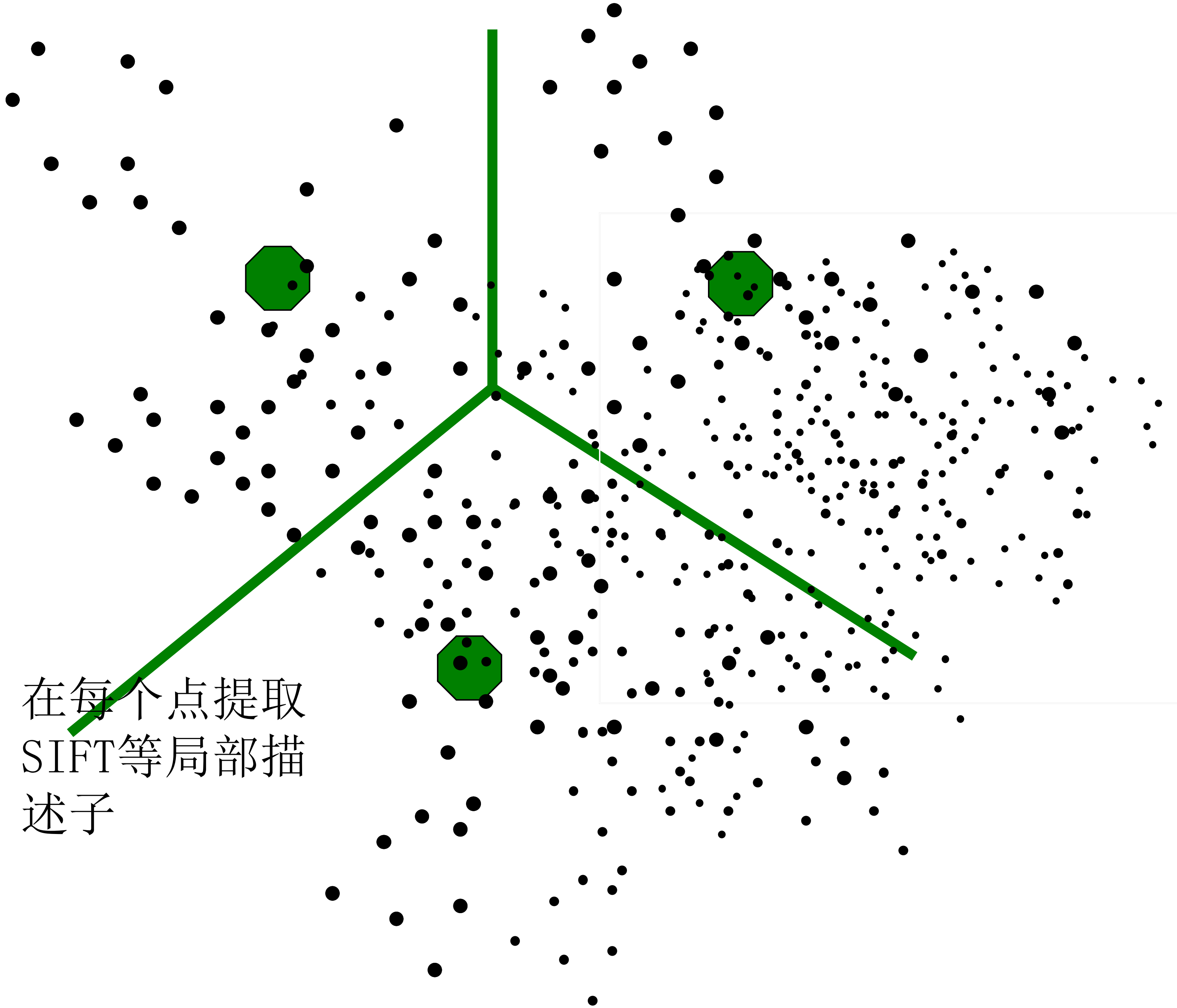


示例：基于字典树的识别



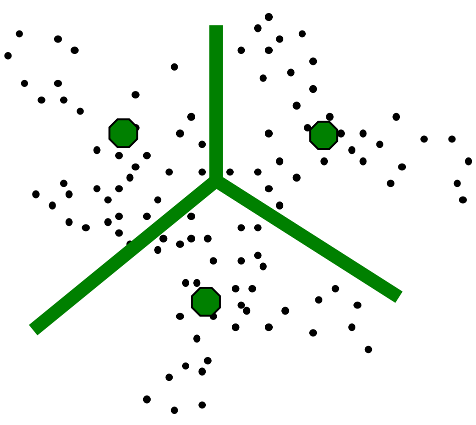


在每个点提取SIFT等局部描述子


在每个点提取SIFT等局部描述子

示例：基于字典树的识别

- 视觉单词
- 每一组属于一个语义



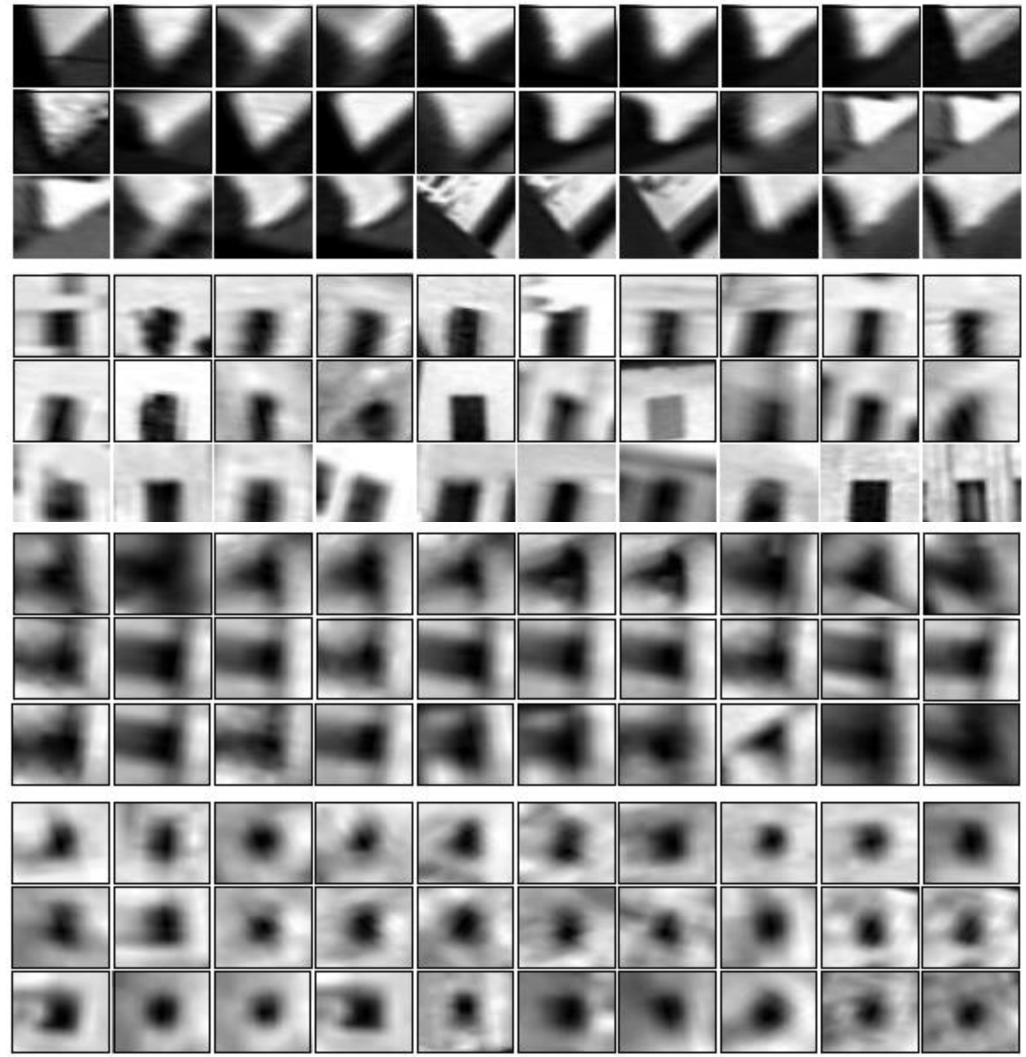
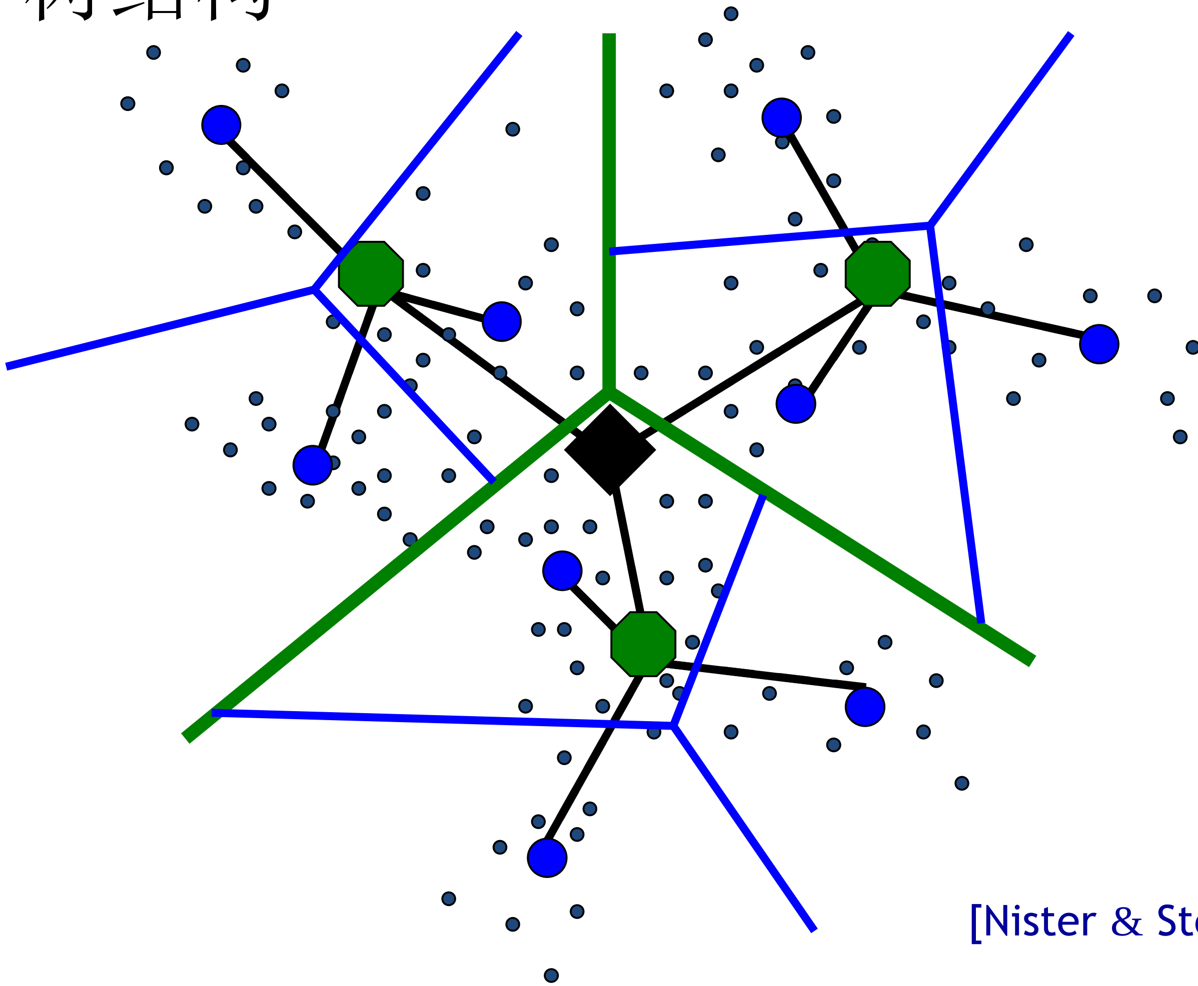


Figure from Sivic & Zisserman, ICCV 2003

示例：基于字典树的识别

- 树结构

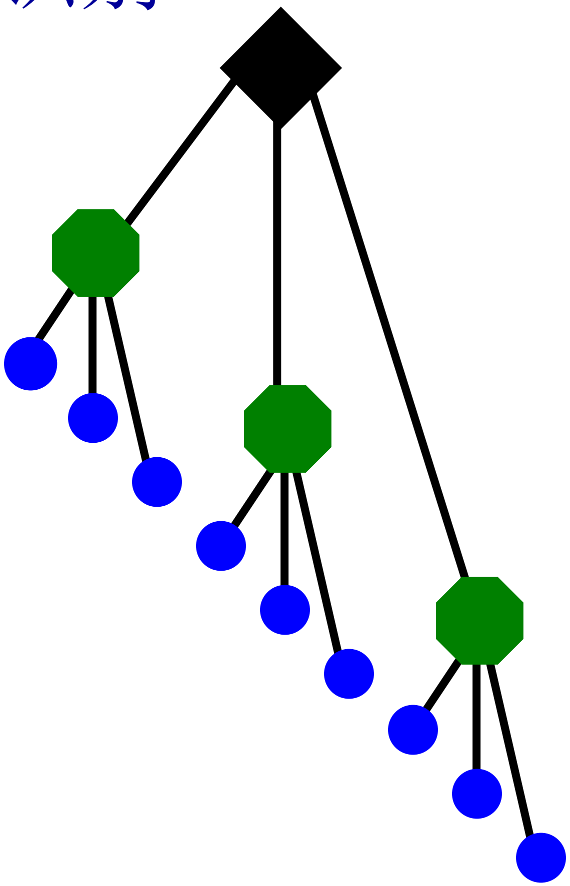


[Nister & Stewenius, CVPR'06]

Slide credit: David Nister

示例：基于字典树的识别

- 字典树
- 训练：完成树结构

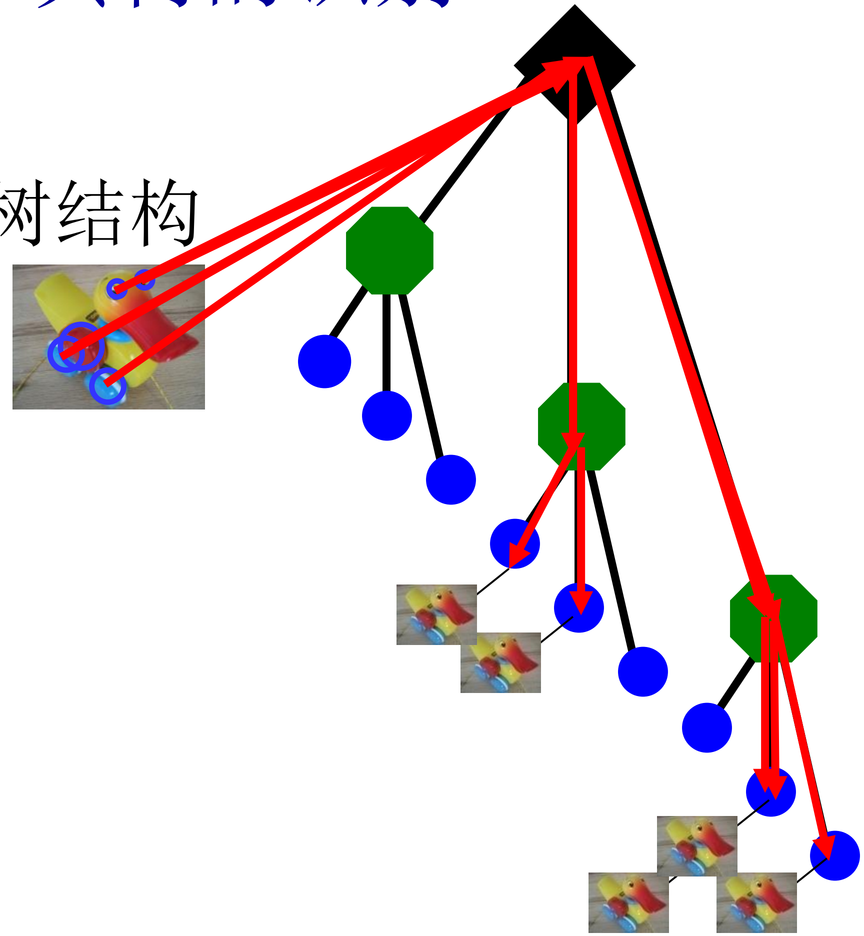


[Nister & Stewenius, CVPR'06]

Slide credit: David Nister

示例：基于字典树的识别

- 字典树
- 训练：完成树结构

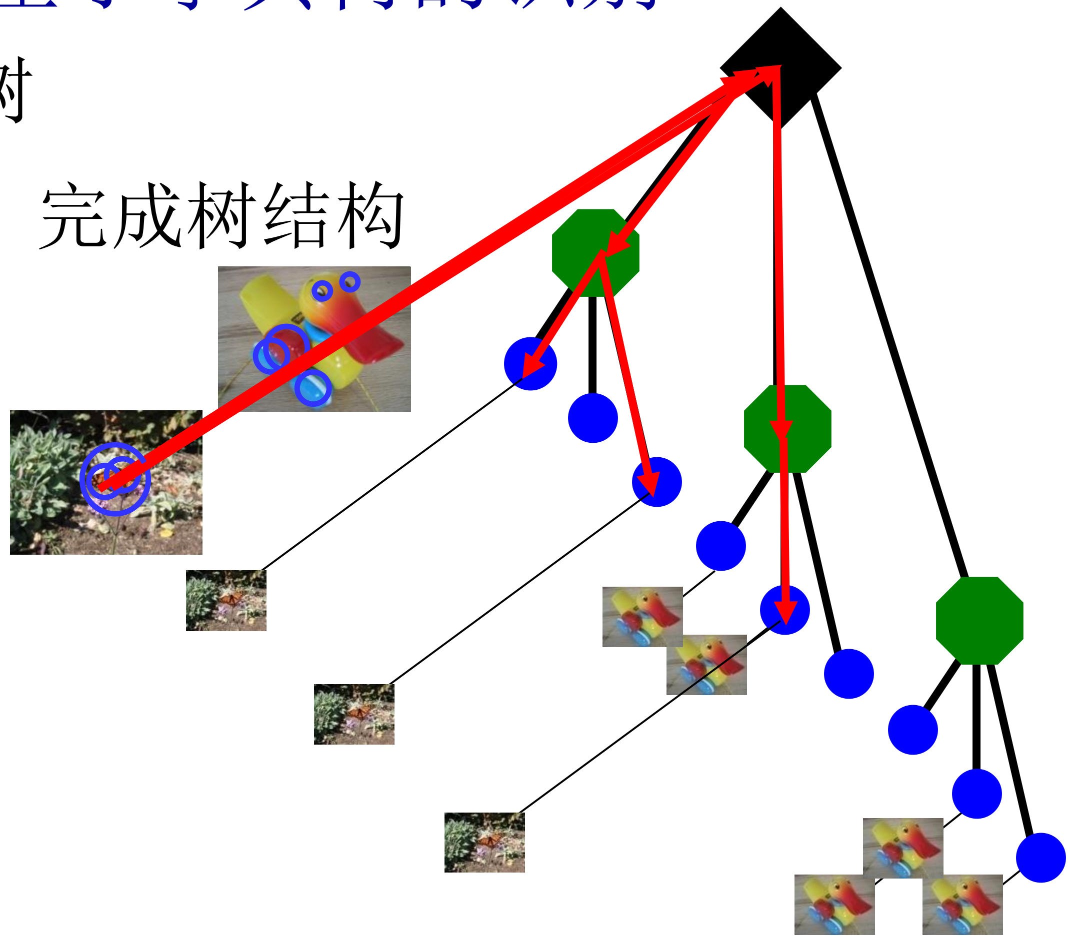


[Nister & Stewenius, CVPR'06]

Slide credit: David Nister

示例：基于字典树的识别

- 字典树
- 训练：完成树结构

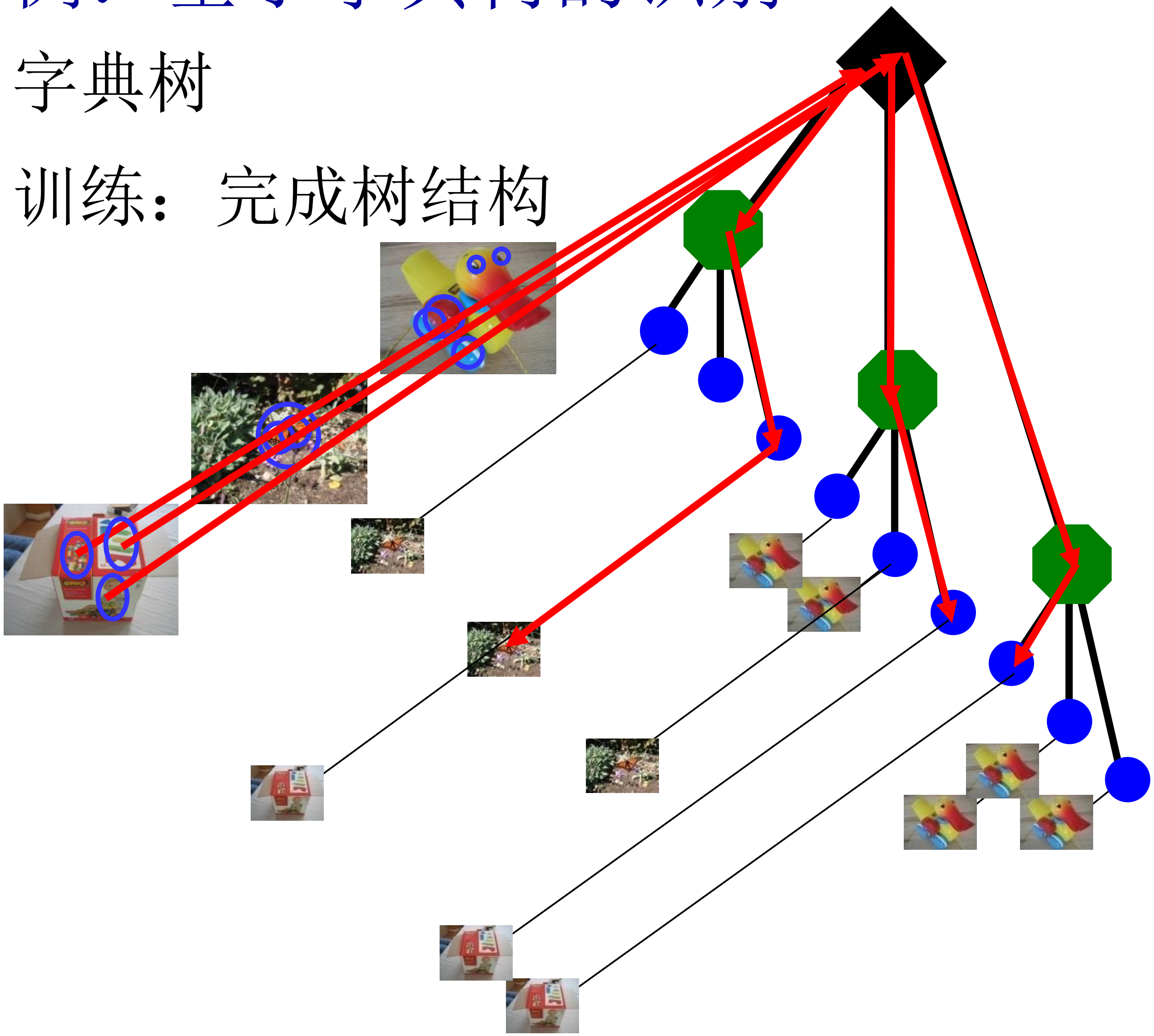


[Nister & Stewenius, CVPR’06]

Slide credit: David Nister

示例：基于字典树的识别

- 字典树
- 训练：完成树结构

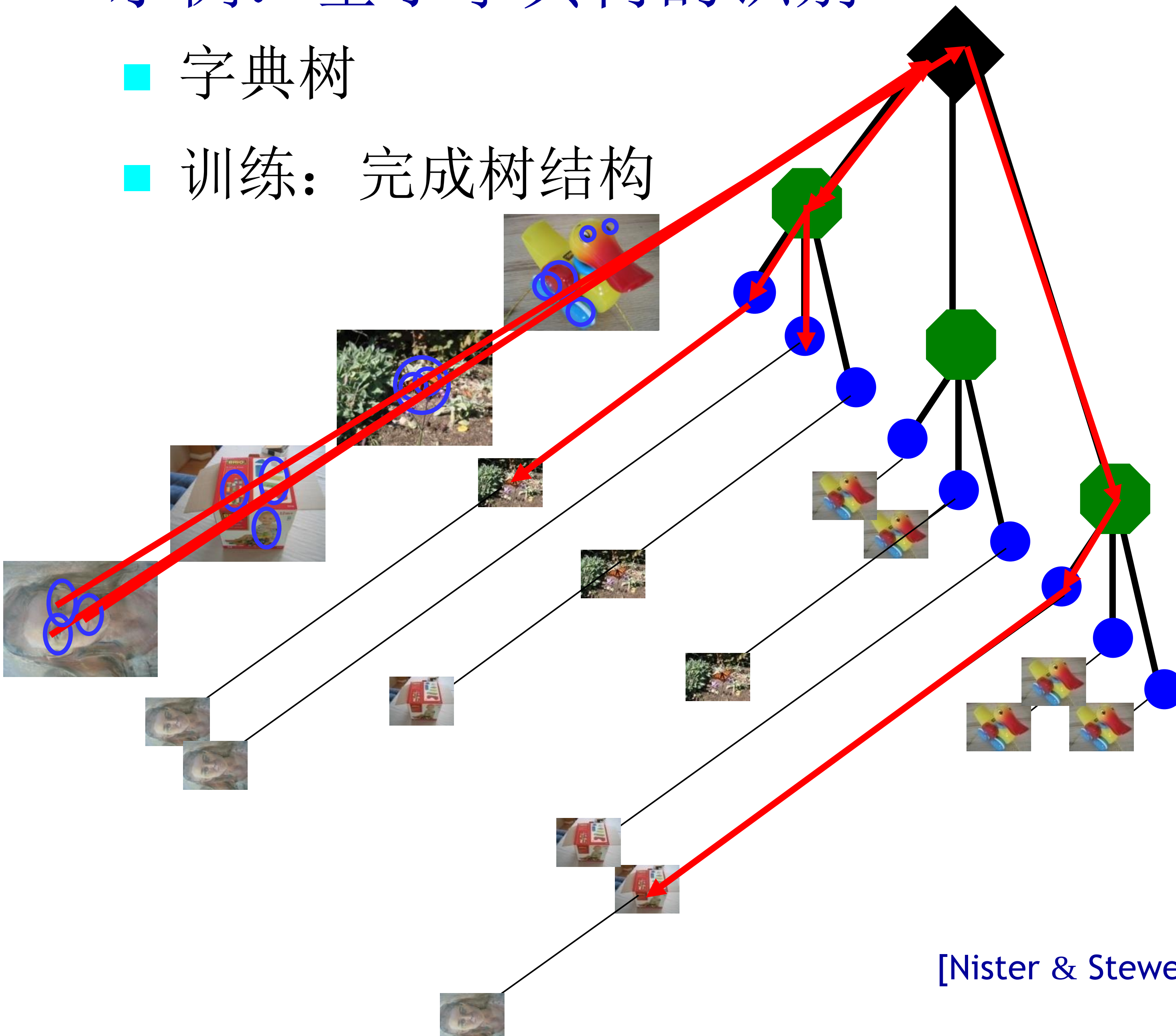


[Nister & Stewenius, CVPR’06]

Slide credit: David Nister

示例：基于字典树的识别

- 字典树
- 训练：完成树结构

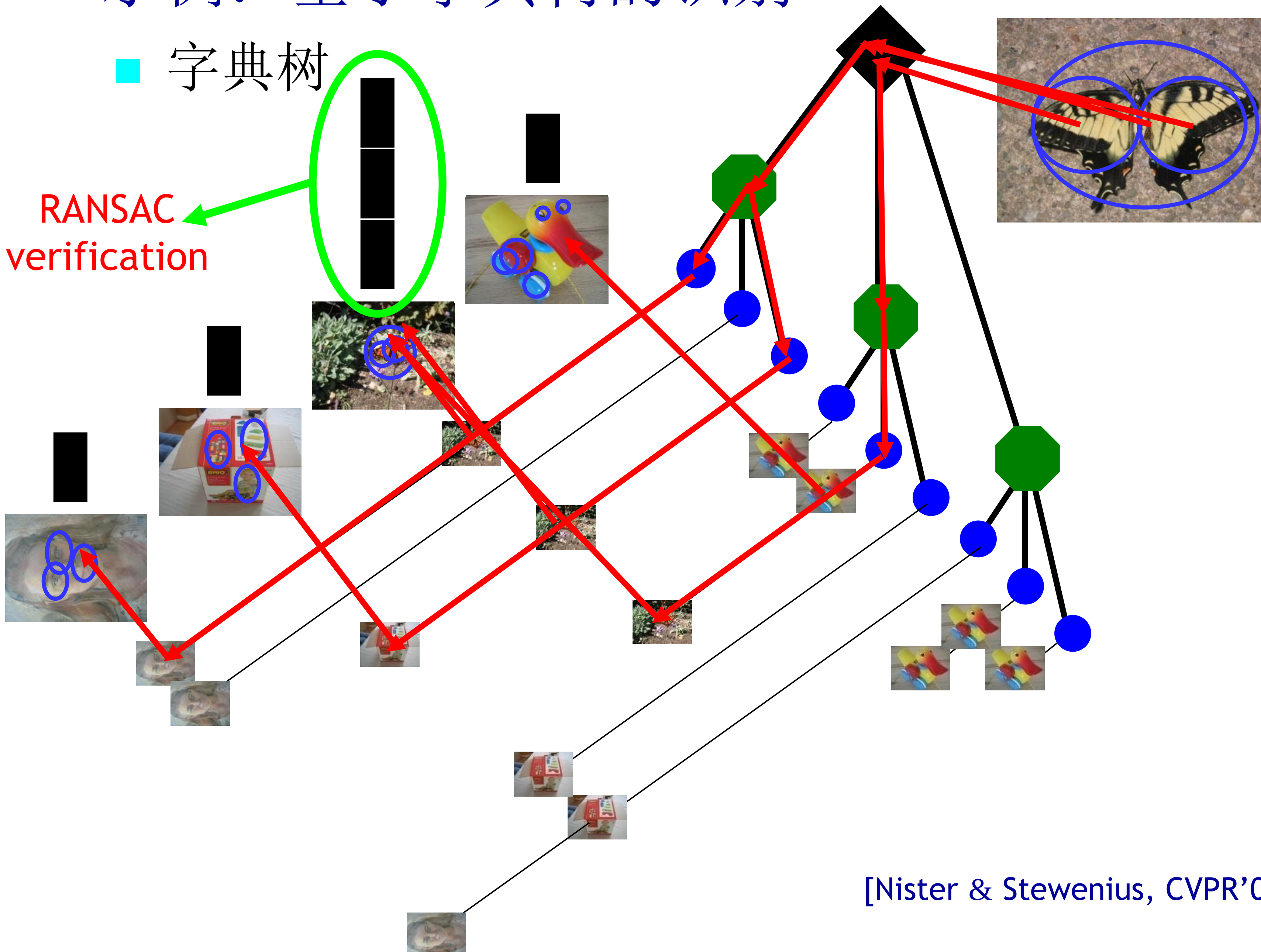


[Nister & Stewenius, CVPR'06]

Slide credit: David Nister

示例：基于字典树的识别

- 字典树



[Nister & Stewenius, CVPR’06]

Slide credit: David Nister

示例：基于字典树的识别

- 节点权重设置问题

$$w_i = \ln \frac{N}{N_i},$$

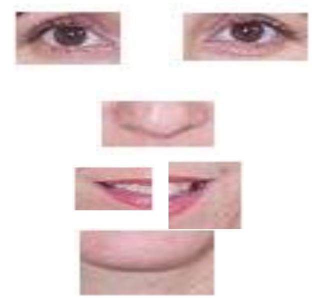
- 距离度量问题

$$\begin{aligned}
\| \mathbf{q}-\mathbf{d} \|_p^p &= \sum_i |q_i - d_i|^p \\
&= \sum_{i|d_i=0} |q_i|^p + \sum_{i|q_i=0} |d_i|^p + \sum_{i|q_i\neq 0, d_i\neq 0} |q_i - d_i|^p \\
&= \|\mathbf{q}\|_p^p + \|\mathbf{d}\|_p^p + \sum_{i|q_i\neq 0, d_i\neq 0} (|q_i - d_i|^p - |q_i|^p - |d_i|^p) \\
&= \quad 2 + \sum_{i|q_i\neq 0, d_i\neq 0} (|q_i - d_i|^p - |q_i|^p - |d_i|^p),
\end{aligned}$$

$$\| \mathbf{q}-\mathbf{d} \|_2^2 = 2 - 2 \sum_{i|q_i\neq 0, d_i\neq 0} q_i d_i,$$

基于特征点的目标检测方法示例

- Bag-of-words框架
- 缺点？



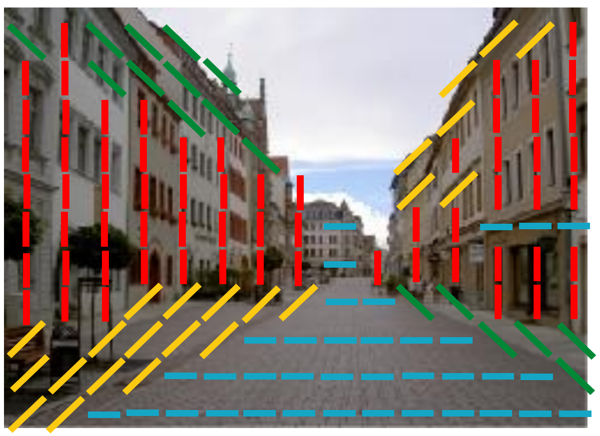
基于特征点的目标检测方法示例

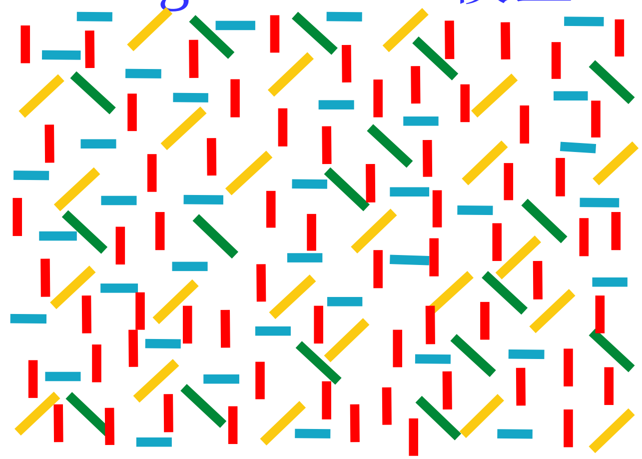
- Bag-of-words框架
- 缺点？

- 如何引入空间信息
- 特征点之间的空间关系
  - 星群模型
  - 空间金字塔模型

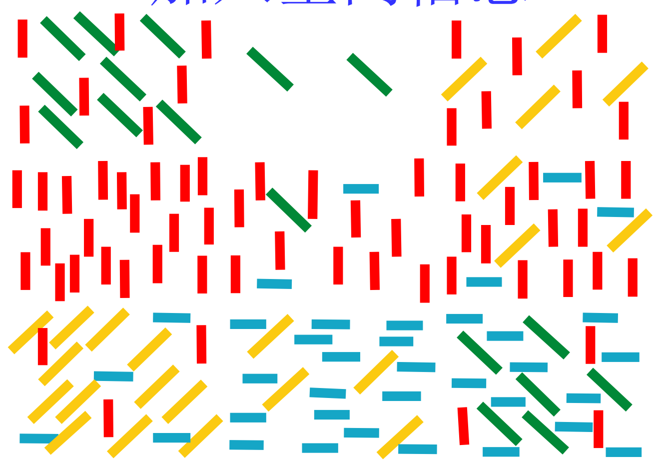
基于特征点的目标检测方法示例

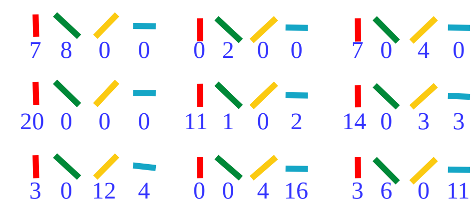
Bag of words 模型



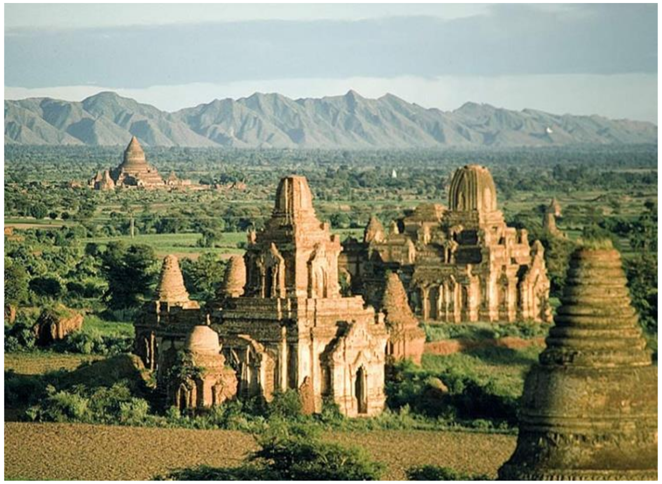


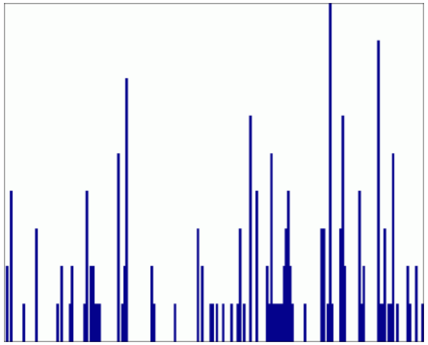
加入空间信息



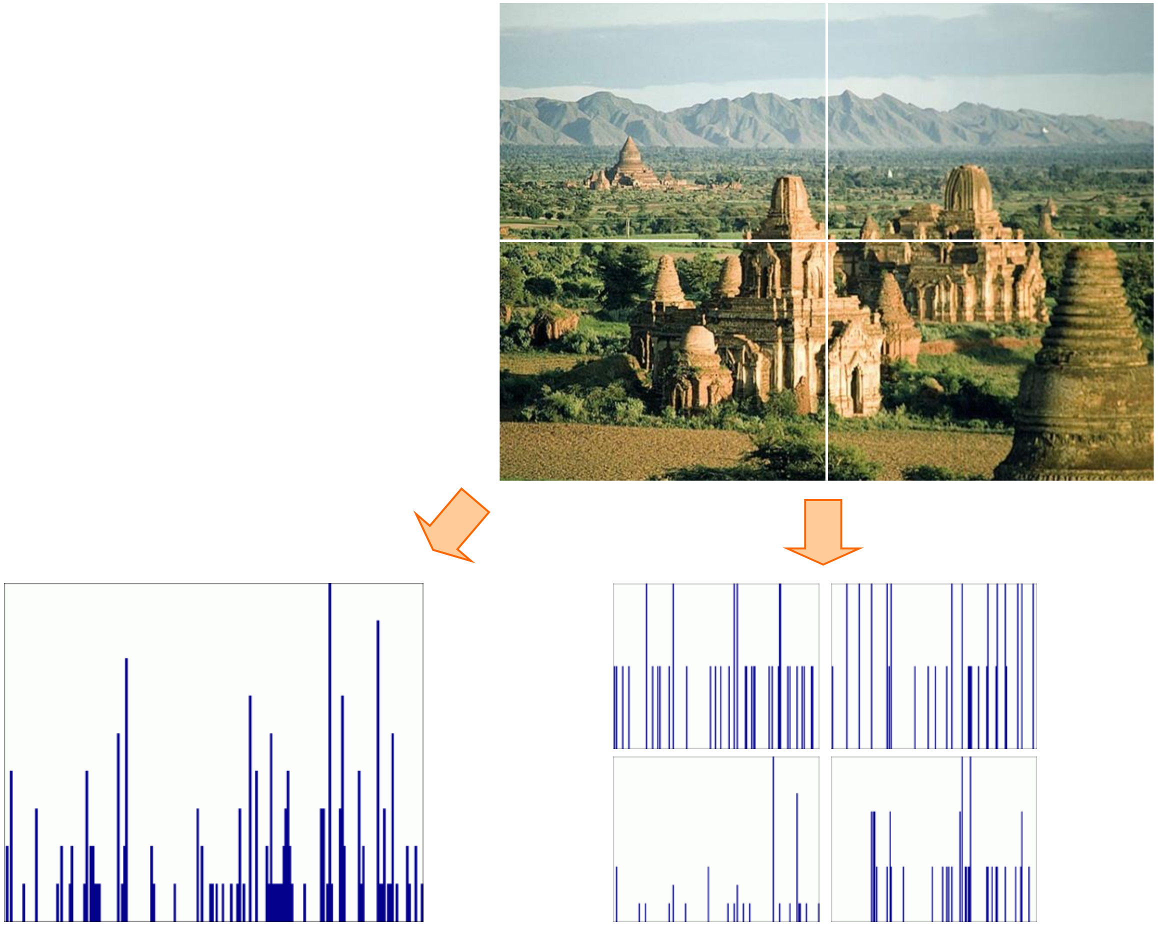


空间金字塔表达





空间金字塔表达



空间金字塔表达

The end！

How do we evaluate object detection?

- predictions
- ground truth

How do we evaluate object detection?

- predictions
- ground truth

True positive:

- The overlap of the prediction with the ground truth is MORE than 0.5

How do we evaluate object detection?

- predictions
- ground truth

True positive:

False positive:

- The overlap of the prediction with the ground truth is LESS than 0.5

How do we evaluate object detection?

- predictions
- ground truth

True positive:

False positive:

False negative:

- The objects that our model doesn't find

How do we evaluate object detection?

- predictions
- ground truth

True positive:

False positive:

False negative:

- The objects that our model doesn't find

What is a True Negative?

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

=

	Predicted 1	Predicted 0
True 1	TP	FN
True 0	FP	TN

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

=

	Predicted 1	Predicted 0
True 1	TP	FN
True 0	FP	TN

	Predicted 1	Predicted 0
True 1	hits	misses
True 0	false alarms	correct rejections

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

=

	Predicted 1	Predicted 0
True 1	TP	FN
True 0	FP	TN

	Predicted 1	Predicted 0
True 1	hits	misses
True 0	false alarms	correct rejections

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

How do we evaluate object detection?

- predictions
- ground truth

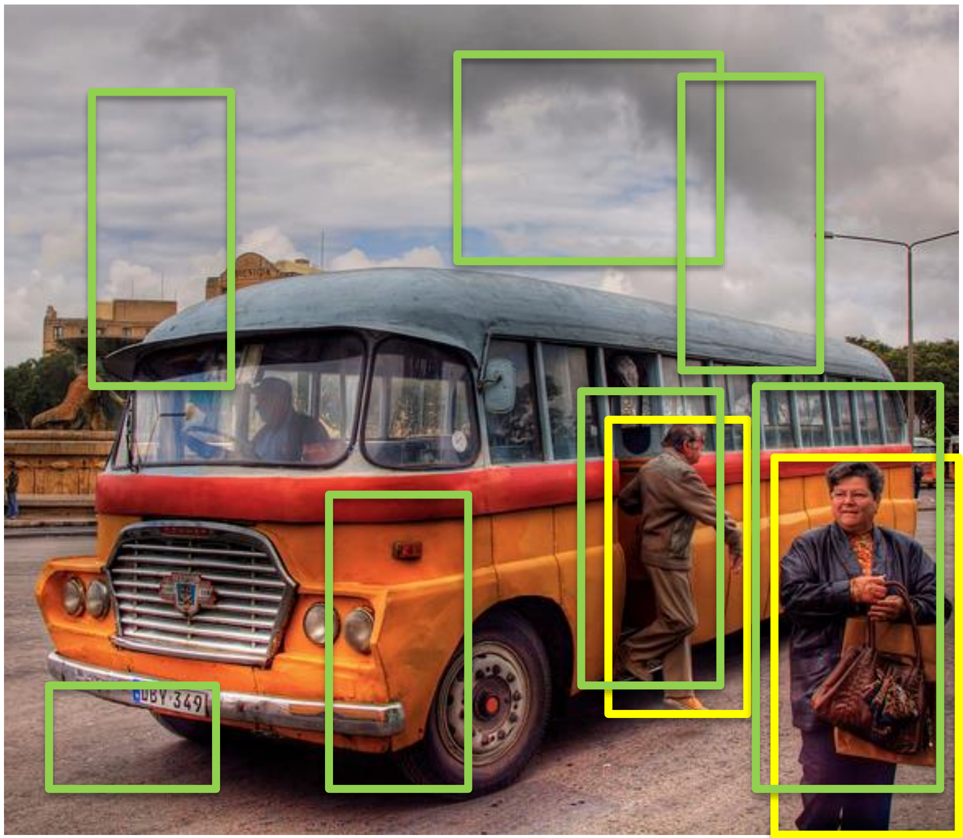
True positive: 1
False positive: 2
False negative: 1

So what is the
- precision?
- recall?

Precision versus recall

- Precision:
  - how many of the object detections are correct precision
- Recall:
  - how many of the ground truth objects can the model detect?

In reality, our model makes a lot of predictions with varying scores between 0 and 1



- predictions
- ground truth

Here are all the boxes that are predicted with score > 0.

This means that our
- Recall is perfect!
- But our precision is BAD!

In reality, our model makes a lot of predictions with varying scores between 0 and 1

- predictions
- ground truth

There are no boxes that are predicted with score = 1.

This means that our

- Precision is undefined!
- And our recall is BAD!

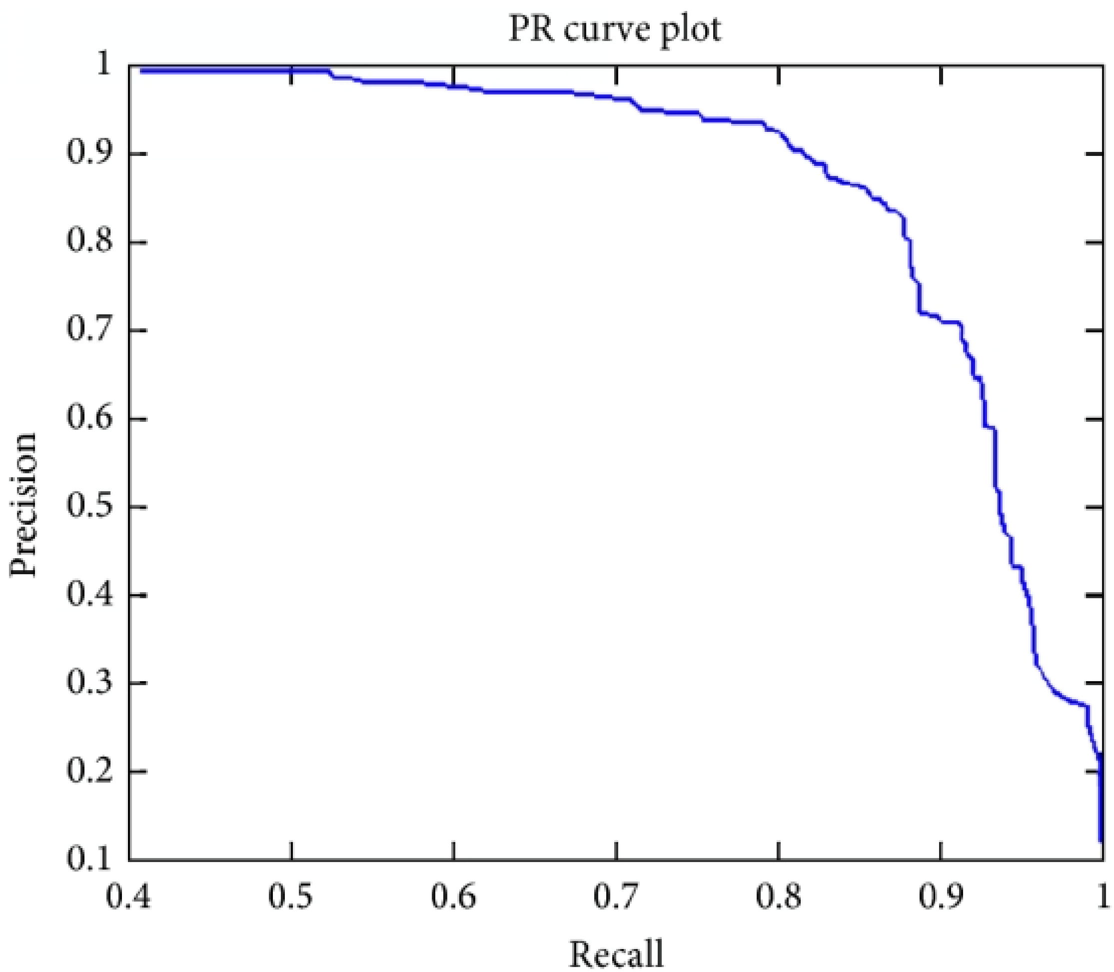
How do we evaluate object detection?

- predictions
- ground truth

Here are all the boxes that are predicted with score > 0.5

We are setting a threshold of 0.5

Precision – recall curve (PR curve)



Which model is the best?