

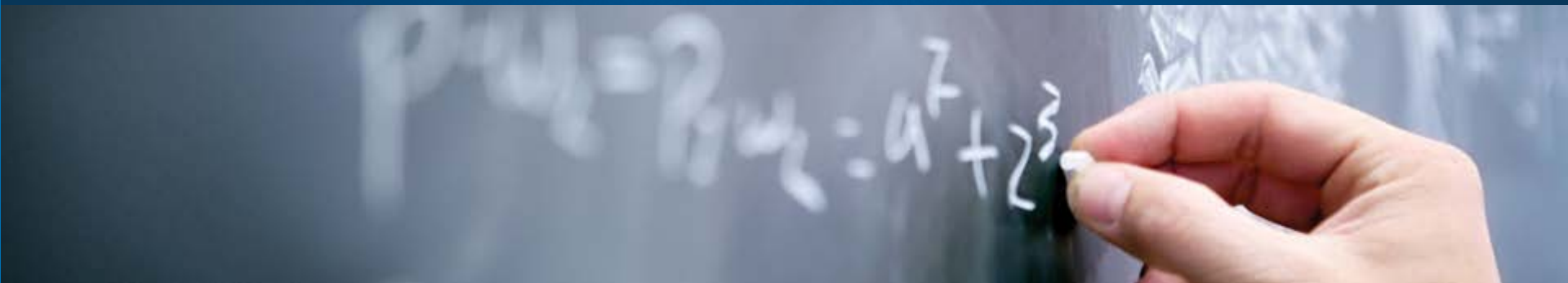
Анализ изображений и видео

Лекция 7: Классификация изображений и распознавание объектов.

Наталья Васильева

nvassilieva@hp.com

HP Labs Russia



Вопросы (время на ответы 15 минут)

1. Преобразуйте цепной код 11076765543322 так, чтобы он стал инвариантным по отношению к выбору начальной точки и к повороту.
2. Дано изображение шахматного поля с клетками размером $n \times n$ пикселей. Какие параметры сдвига будут порождать матрицу смежности диагонального вида?
3. К каким трансформациям изображения не инвариантен детектор Харриса?
4. Какая индексная структура требует меньшего объема памяти для организации индекса: kD-tree или Vocabulary tree?
5. Какова вероятность коллизии для двух объектов x и y при использовании b -битных хеш-кодов и t независимых хеш-таблиц, если известно что $\text{sim}(x, y) = p$?
 - a) p^t
 - b) $1 - (1 - p^b)^t$
 - c) $b(1 - p^t)$



Зачем сравнивать изображения?

Классификация, кластеризация

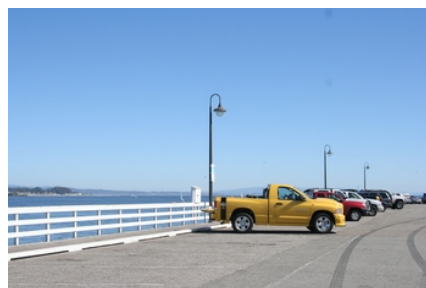


Обучающее множество:
изображения и метки класса

Обучение классификатора:
сопоставление особенностей
изображений меткам классов

Обучение

Модель
классификатора



Тестовое изображение

Сопоставление тестового
изображения модели
классификатора

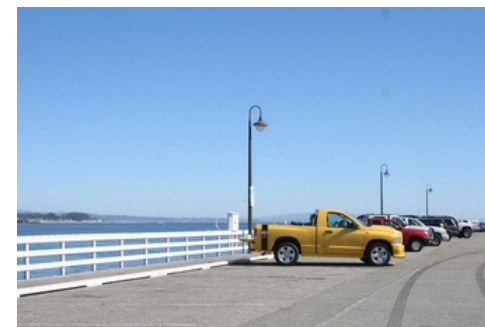
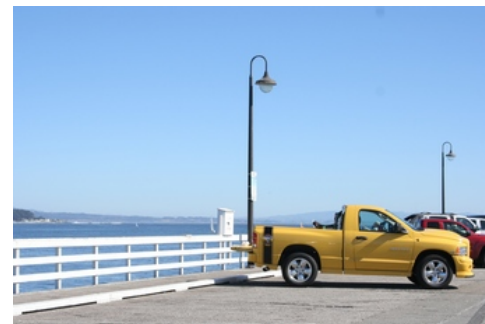
Тестирование

Предсказанные
классификатором
классы:
– open outdoor



Зачем сравнивать изображения?

Обнаружение объектов



Зачем сравнивать изображения?

Мы бы хотели...

- чтобы компьютер «понимал» семантику сцены на изображении
- автоматически распознавать **что** и **где** изображено



- категоризировать и идентифицировать объекты, определять их свойства и отношения

Что считать похожими объектами?

Разные уровни обобщения

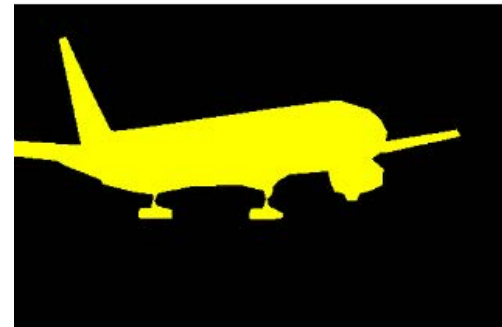
- Один и тот же объект в разные моменты времени (слежение/tracking)
- Один и тот же объект в разных позах, в разных условиях, с разным фоном (распознавание лиц)
- Разные объекты одной категории (обнаружение объектов/object detection)

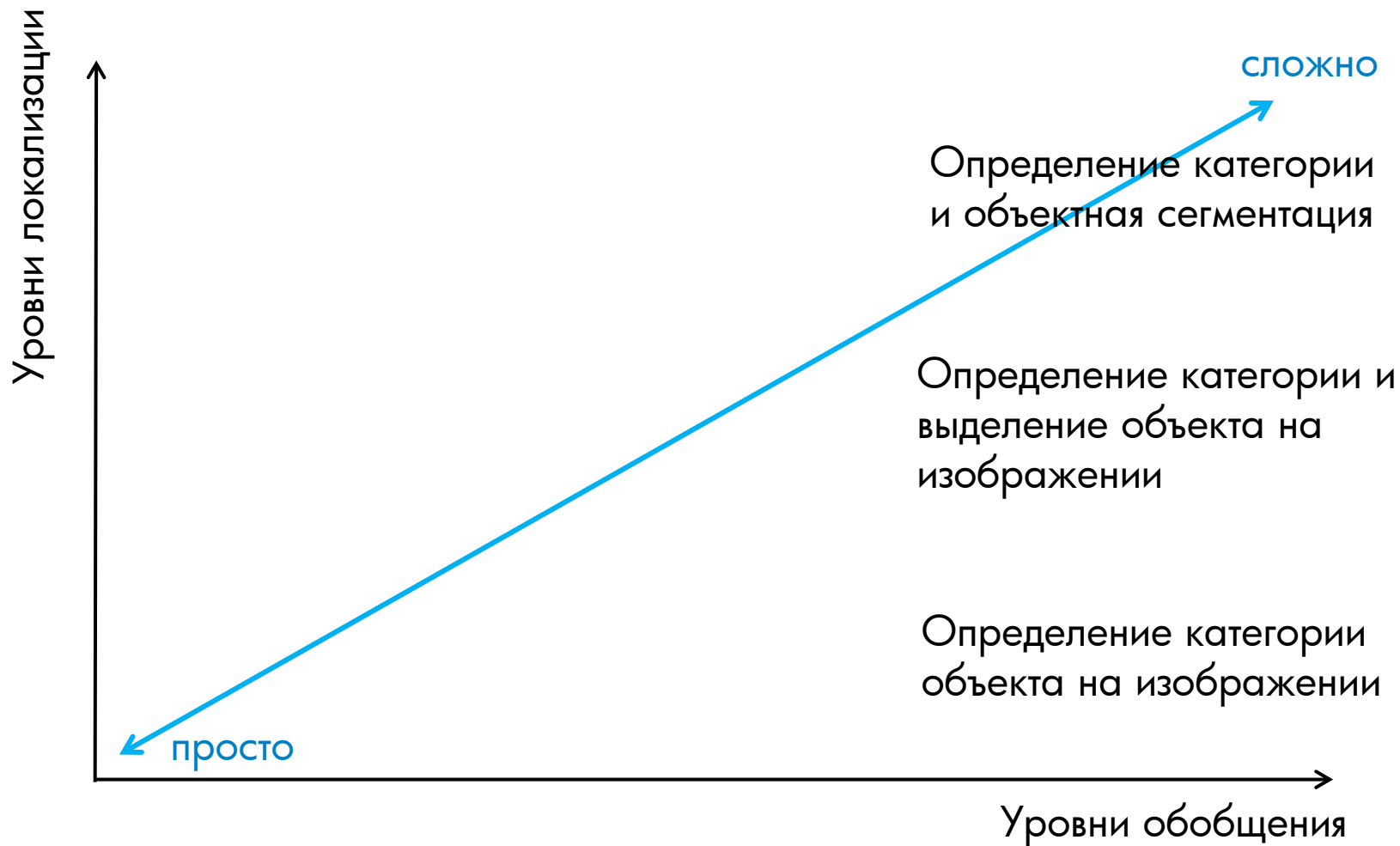


Что есть распознанный объект?

Разные уровни локализации

- Объект (сцена) присутствует на изображении
классификация, категоризация изображений (image classification)
- Известно местоположение объекта, объект выделен на изображении
обнаружение, выделение, локализация объекта (object detection, localization)
- Известны пиксели, принадлежащие объекту
объектная сегментация (object segmentation)





Насколько сложна задача выделения объекта?

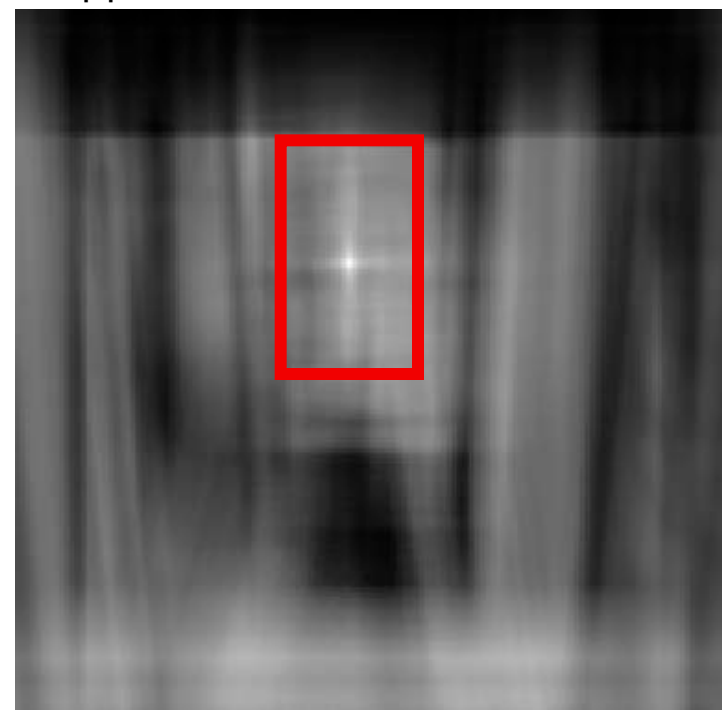
Стул



Найдем стул на этом изображении



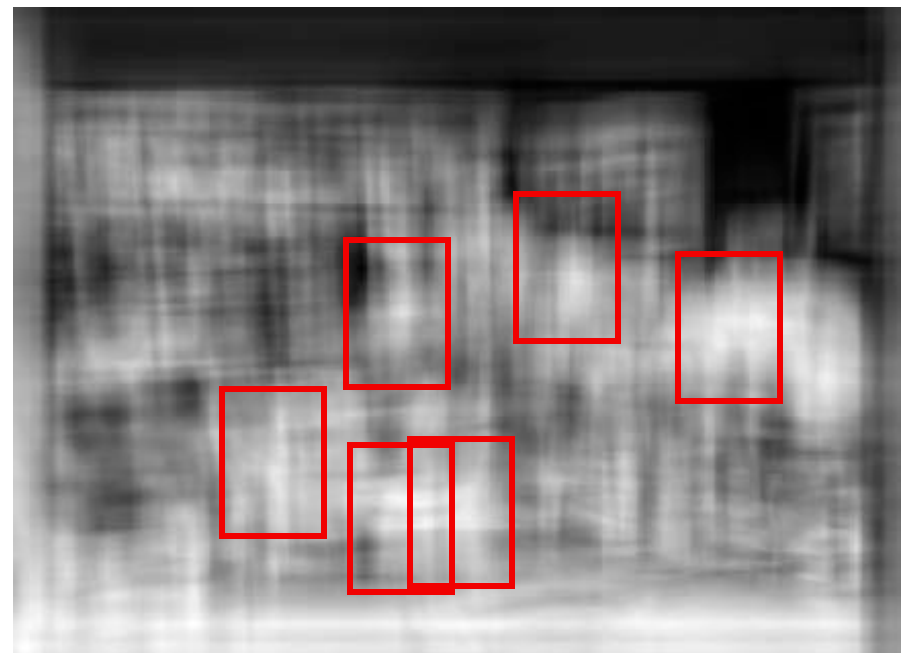
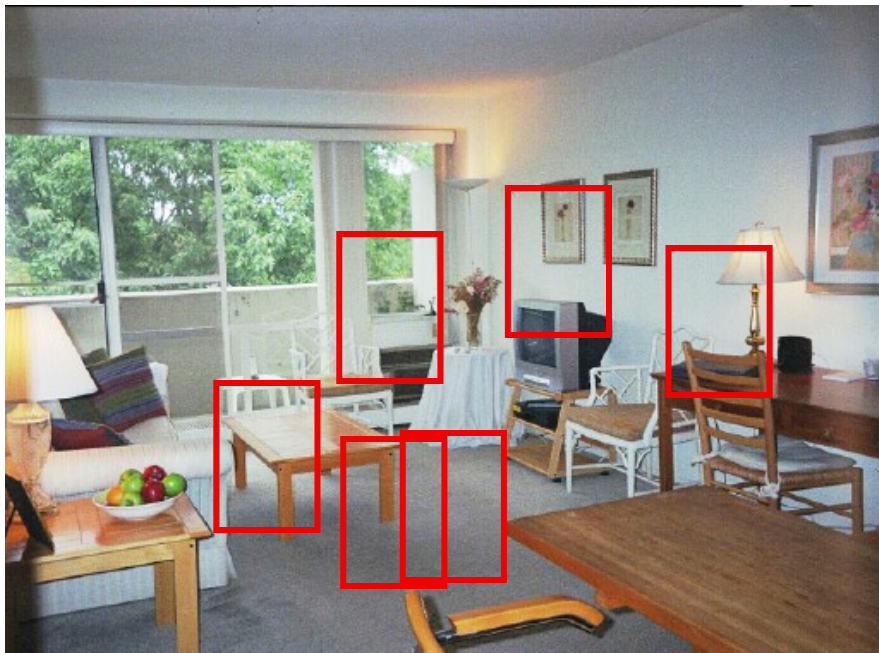
Корреляция с шаблоном



Насколько сложна задача выделения объекта?



Найдем стул на этом изображении

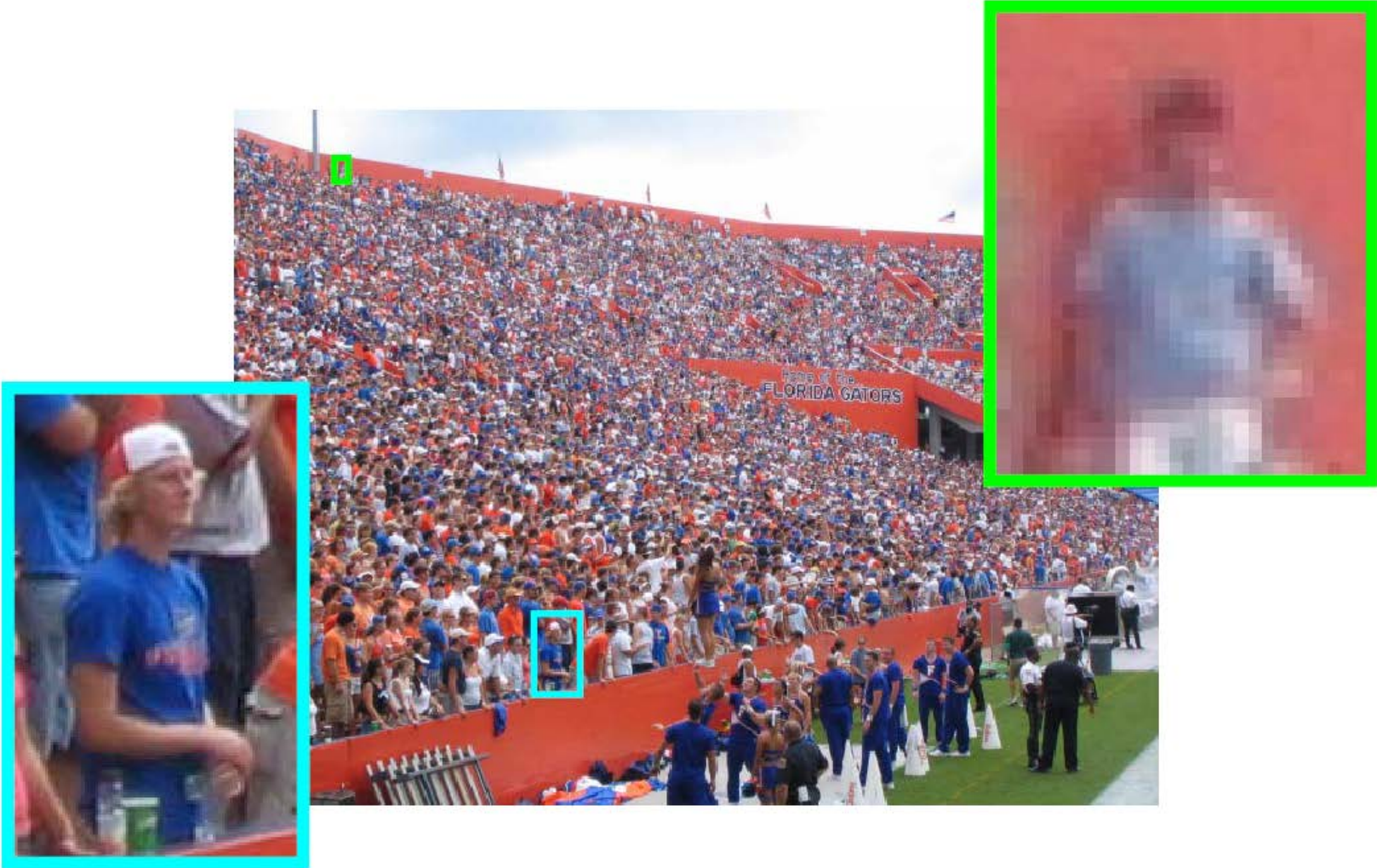


Простое сопоставление шаблонов
не решает задачу

Сложности: разные ракурсы/позы объекта



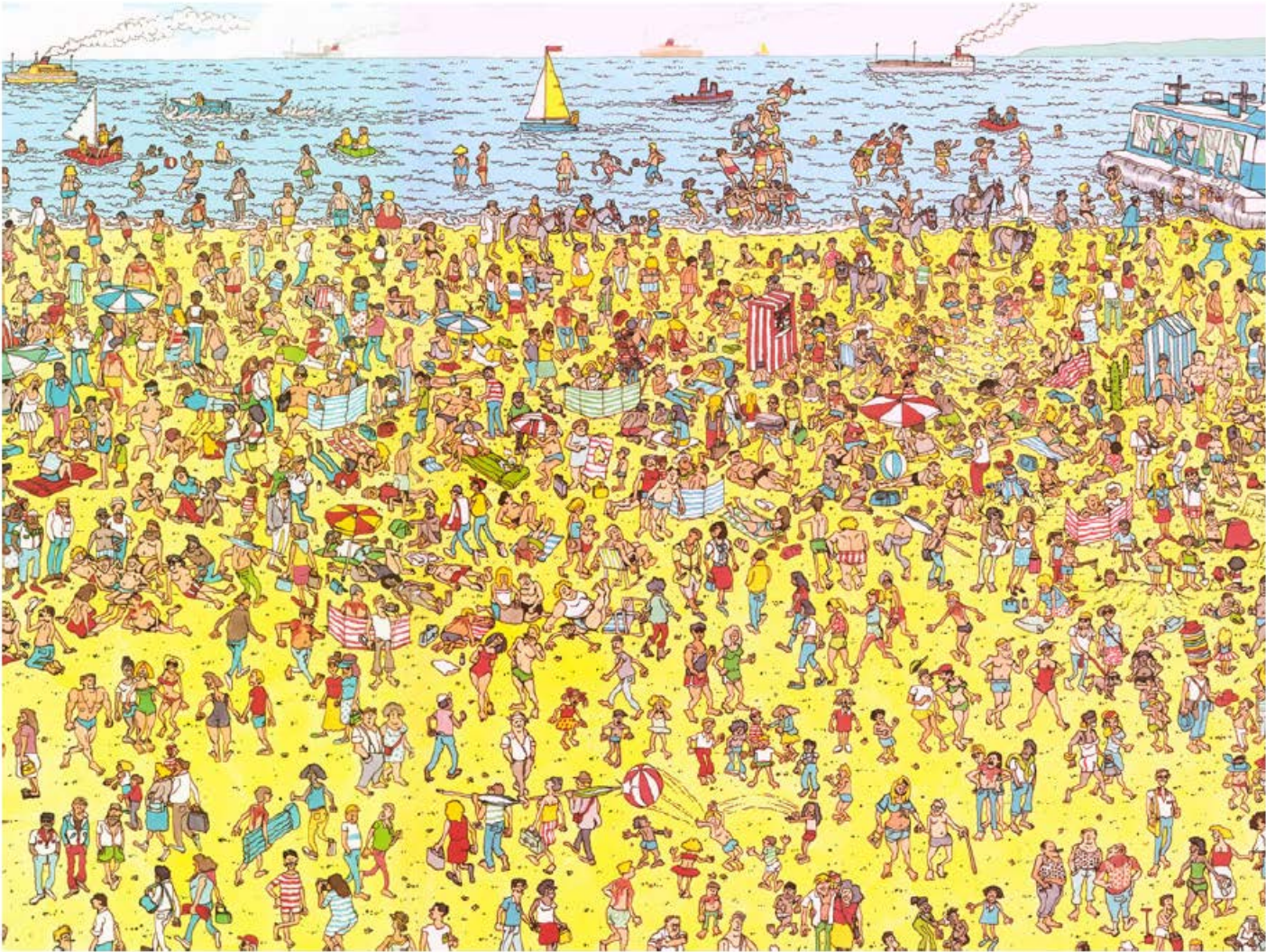
Сложности: разный масштаб



Сложности: изменение освещения



Сложности: фон



Сложности: перекрытия



Сложности: деформируемые объекты



Сложности: внутривидовые различия



Как выделить общие свойства свойства объектов из одной категории?

Определить «с потолка» абстрактную модель непросто
Проще обучить модель на примерах



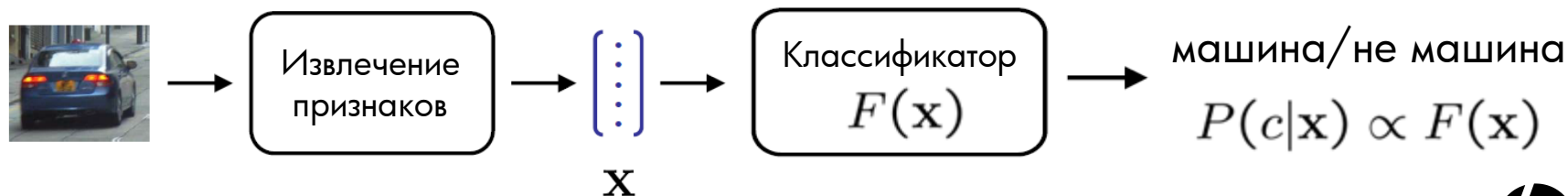
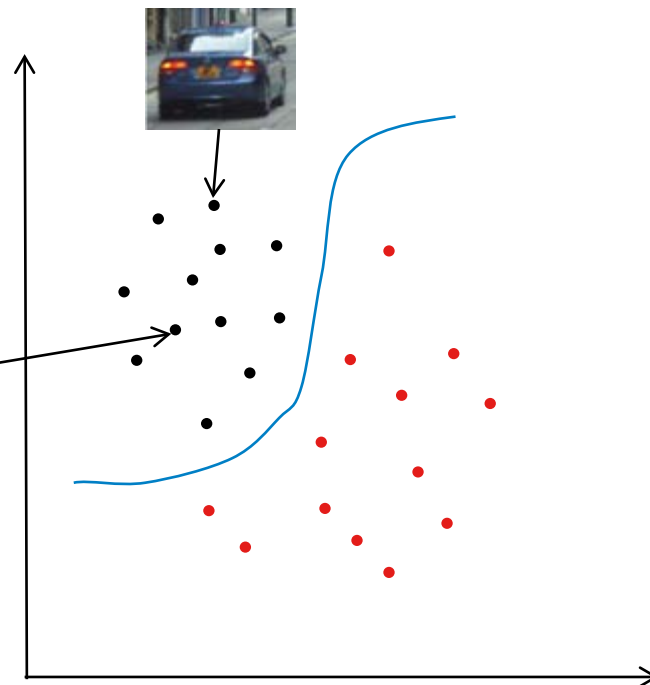
Общая схема решения

Использование классификатора:

- Представление изображения или его фрагмента в виде вектора признаков:

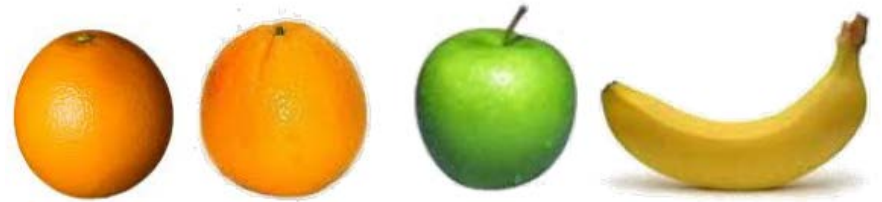


- Обучение классификатора
- Классификация векторов признаков



Какие признаки использовать?

Какие свойства объектов являются ключевыми?



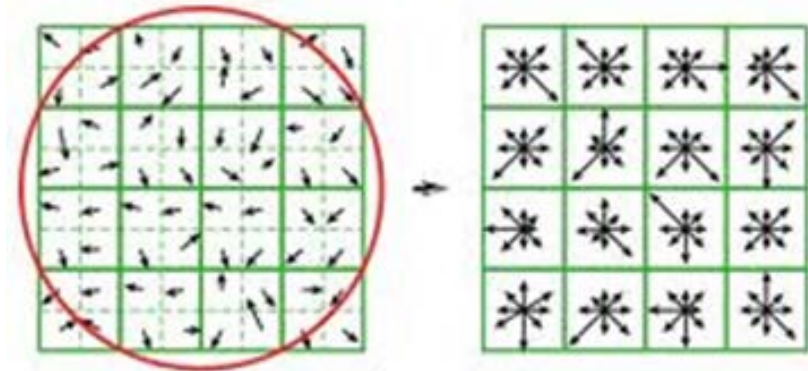
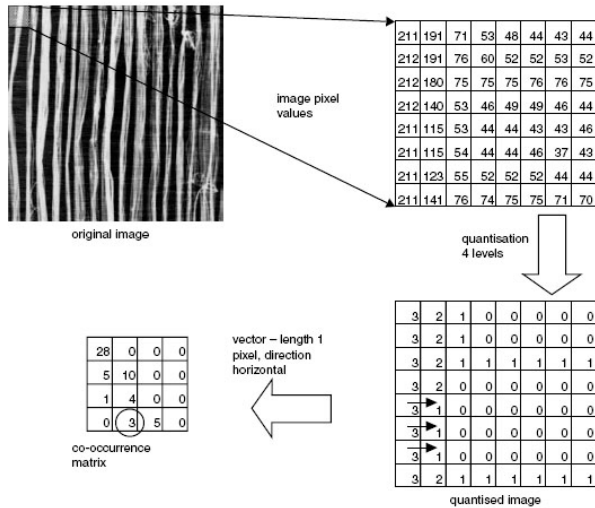
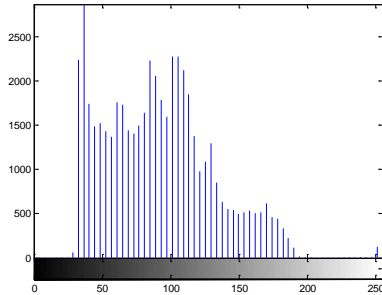
Зависят ли ключевые свойства и признаки от категории?



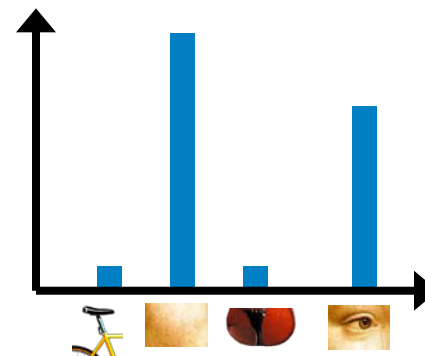
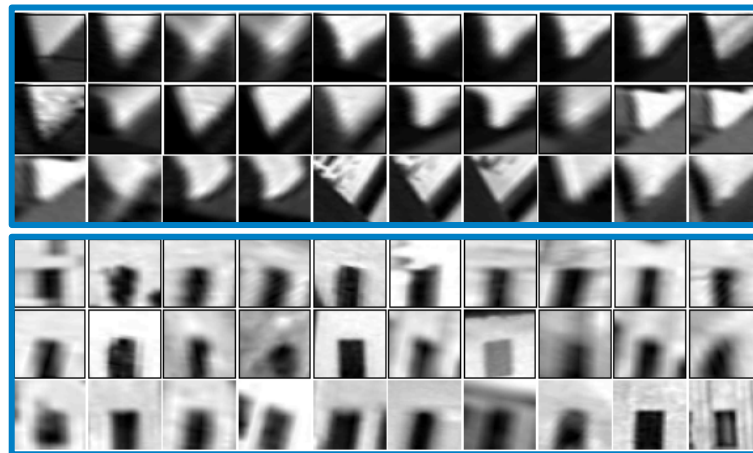
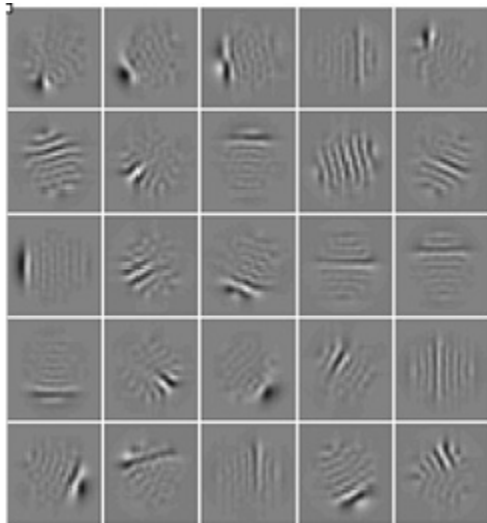
Как достичь необходимой инвариантности?



Какие признаки использовать?

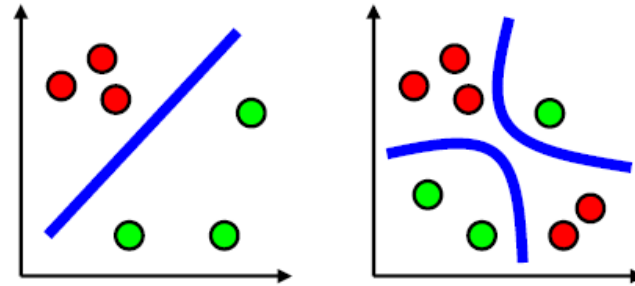


Какие признаки использовать?

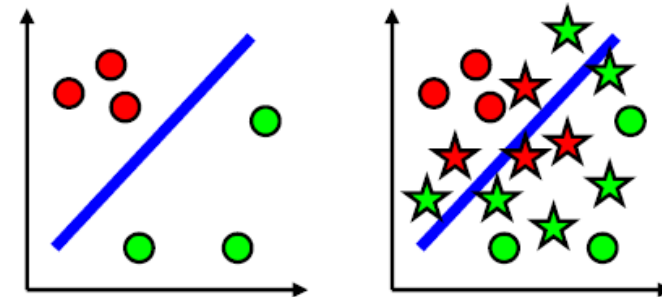


Как обучить классификатор?

Как построить разделяющую поверхность?



Как добиться необходимого уровня обобщения модели классификатора?



Как учесть дополнительные априорные знания?



Различные варианты разметки обучающего множества

Image-level label



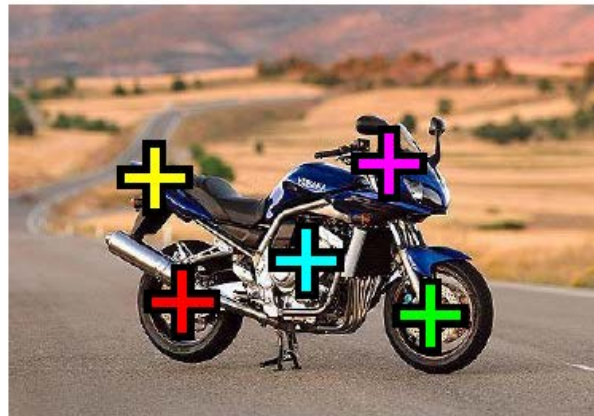
Bounding box



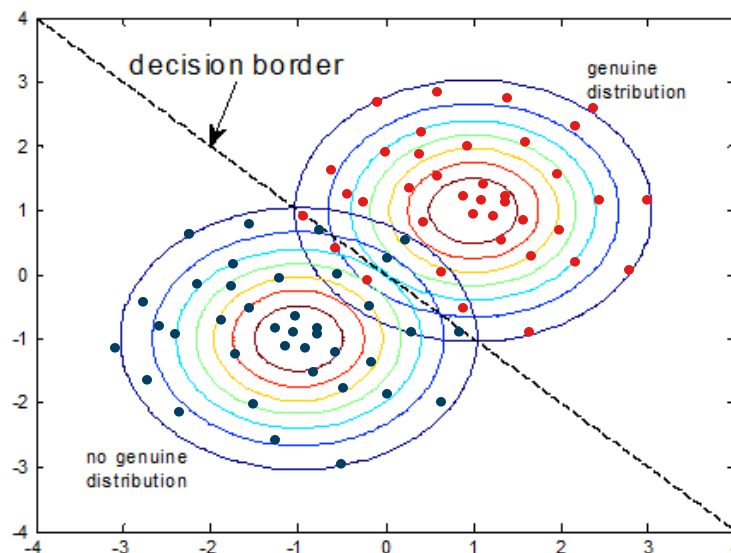
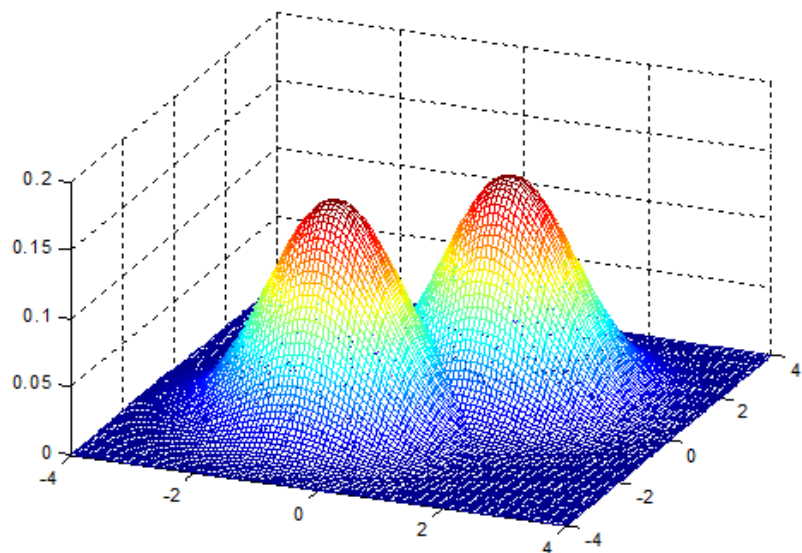
Pixel-level segmentation



“Parts”

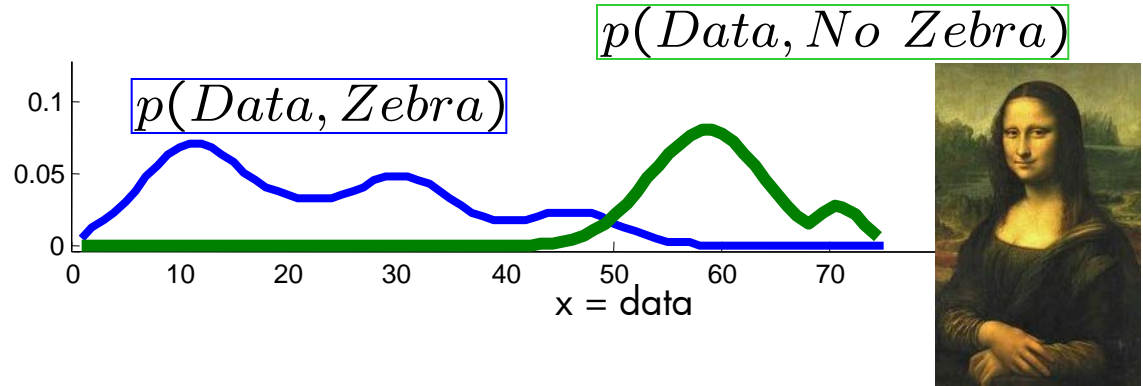


Модели: Generative vs. Discriminative

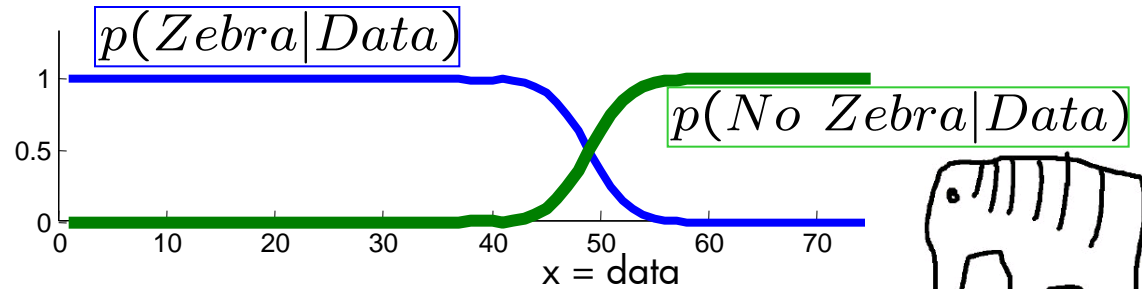
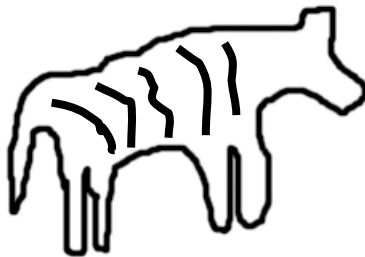


Модели: Generative vs. Discriminative

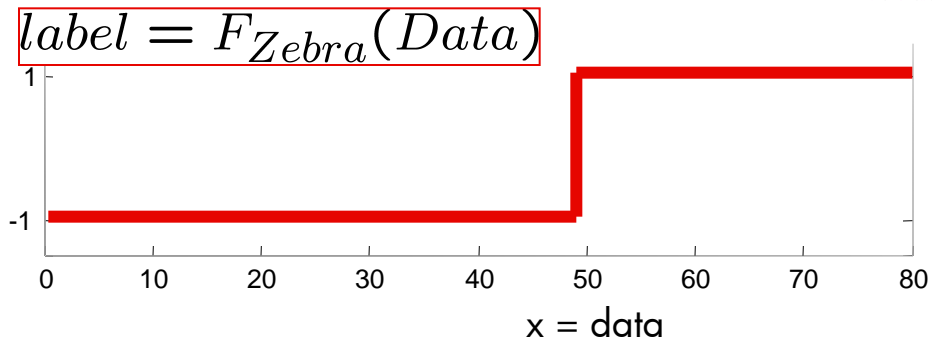
- Generative model



- Discriminative model

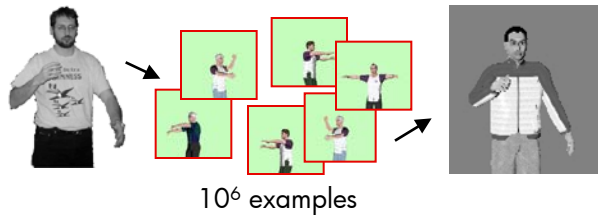


- Classification function



Discriminative methods

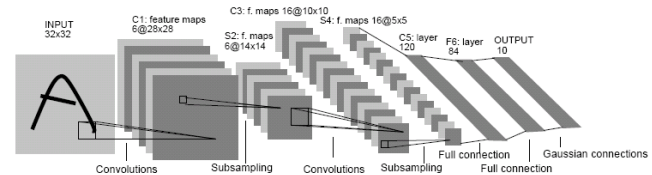
Nearest neighbor



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

...

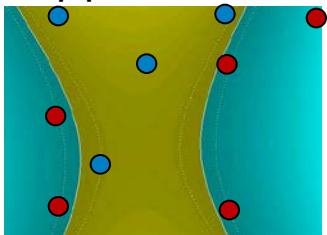
Neural networks



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

...

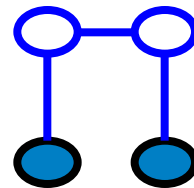
Support Vector Machines and Kernels



Guyon, Vapnik
Heisele, Serre, Poggio, 2001

...

Conditional Random Fields



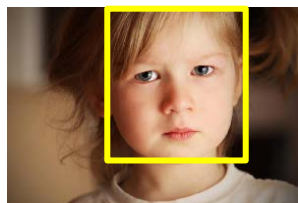
McCallum, Freitag, Pereira 2000
Kumar, Hebert 2003

...

Классификация для разных подзадач и типов объектов



- Определение категории сцены/объекта
Подход: классификация признаков изображения (глобальных или локальных) без выделения и анализа отдельных объектов/частей



- Выделение объектов
Подход: классификация локальных признаков фрагментов изображения (скользящее окно)



- Объекты как множество составляющих их частей
Подход: классификация локальных признаков составляющих частей объектов с учетом их взаимного расположения

Определение категории объекта

- Объект (сцена) присутствует на изображении
классификация, категоризация
изображений (image classification)



Определение категории: пример

“Caltech-5” Dataset

Training

- 50% of data
(100 - 200 images per class)
- Image label only

Aeroplane



Background



Testing

- 50% of data
- Object present/absent

Car
(Rear)



Face
(Frontal)



Motorbike



Learning

- Bag of words representation
- SVM classifier with χ^2 kernel

Results

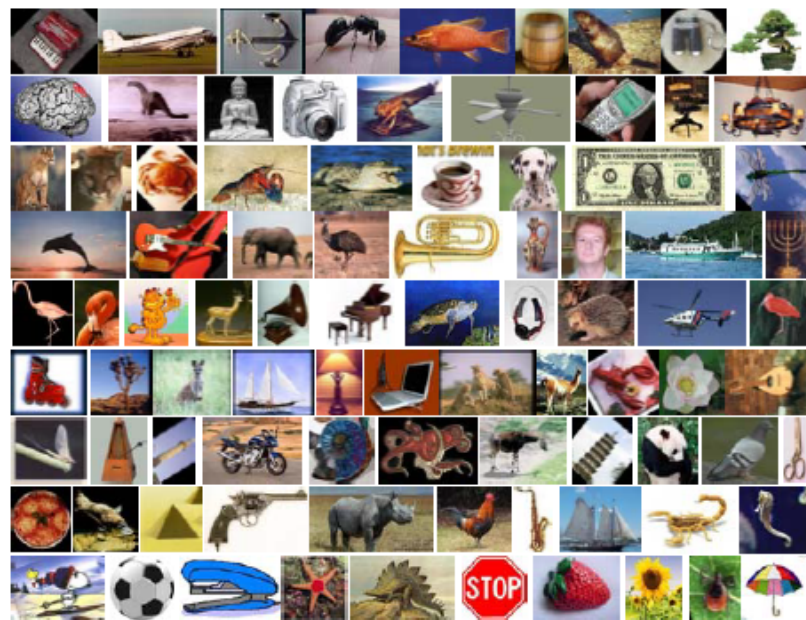
- 98 - 100% accuracy depending on class

[Csurka et al, 2004]

[Zhang et al, 2005]

Datasets & Evaluation: Caltech 101

- 101 Categories
 - accordion, airplanes, anchor, ant, barrel, bass, beaver, binocular, bonsai, brain, brontosaurus, ...
- 40-800 images per category
 - Minimum 31
- Train on **5-30** images
- Test on remainder
 - **Forced choice**: Classify test image as one of 101 classes
 - Report accuracy per class



Caltech 101

- Average images reveal low variability for some classes

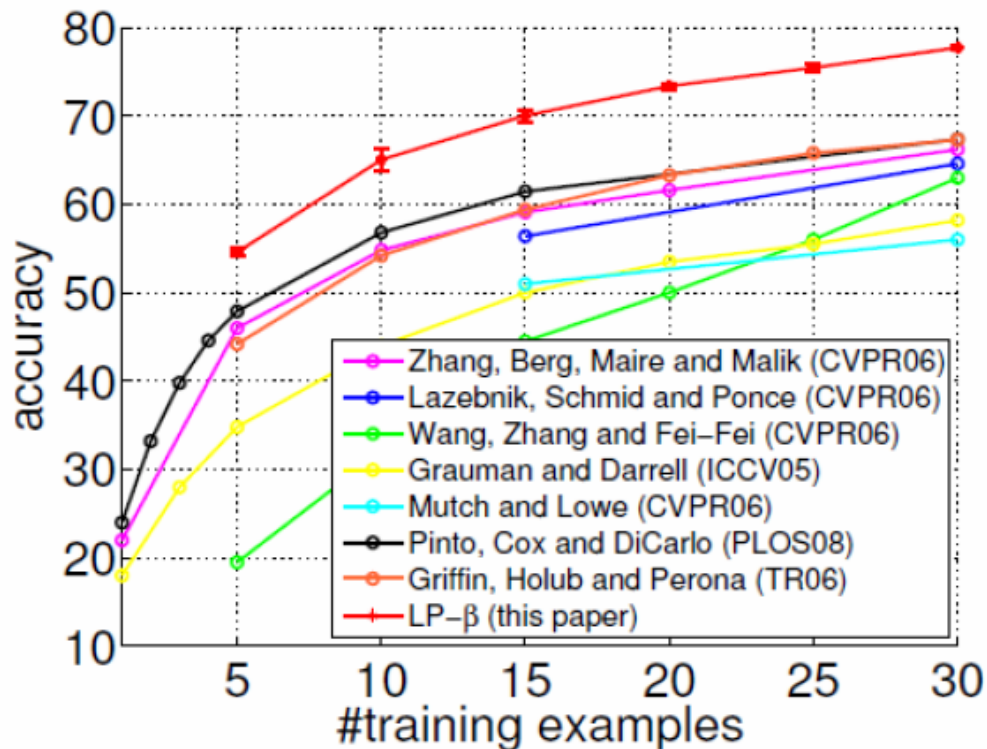


[Torralba]

Slide credit: M. Everingham

Caltech 101

- Performance is close to saturated?
- ~80% accuracy with just 30 training examples per class

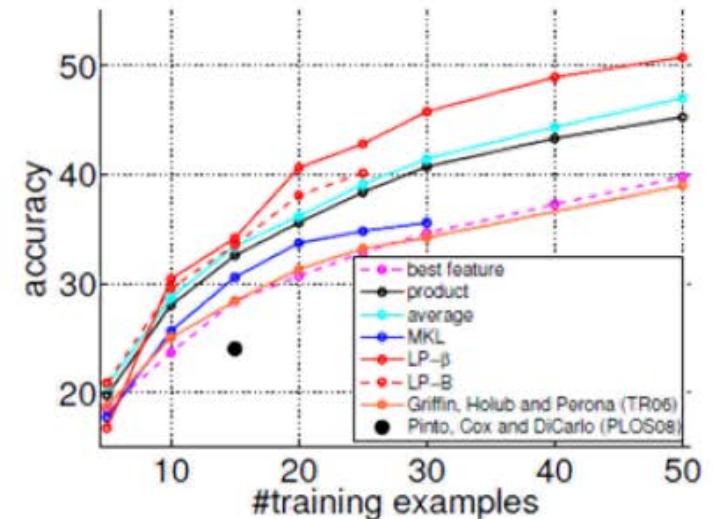
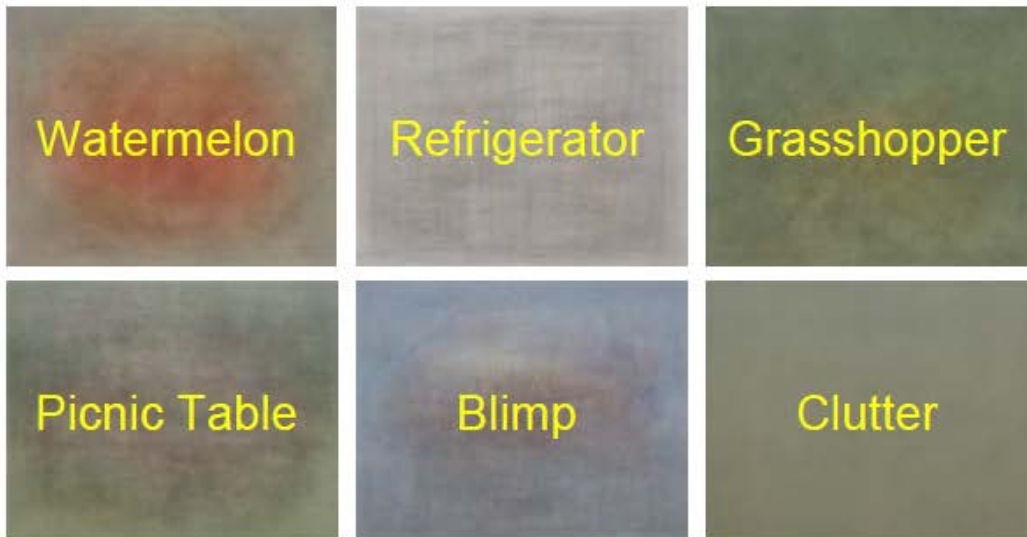


[Gehler & Nowozin, 2009]

Slide credit: M. Everingham

Caltech 256

- 256 Categories
- Minimum 80 images per category
- Unaligned and no artefacts



~50% accuracy

Slide credit: M. Everingham

PASCAL VOC (Visual Object Classes) Challenge

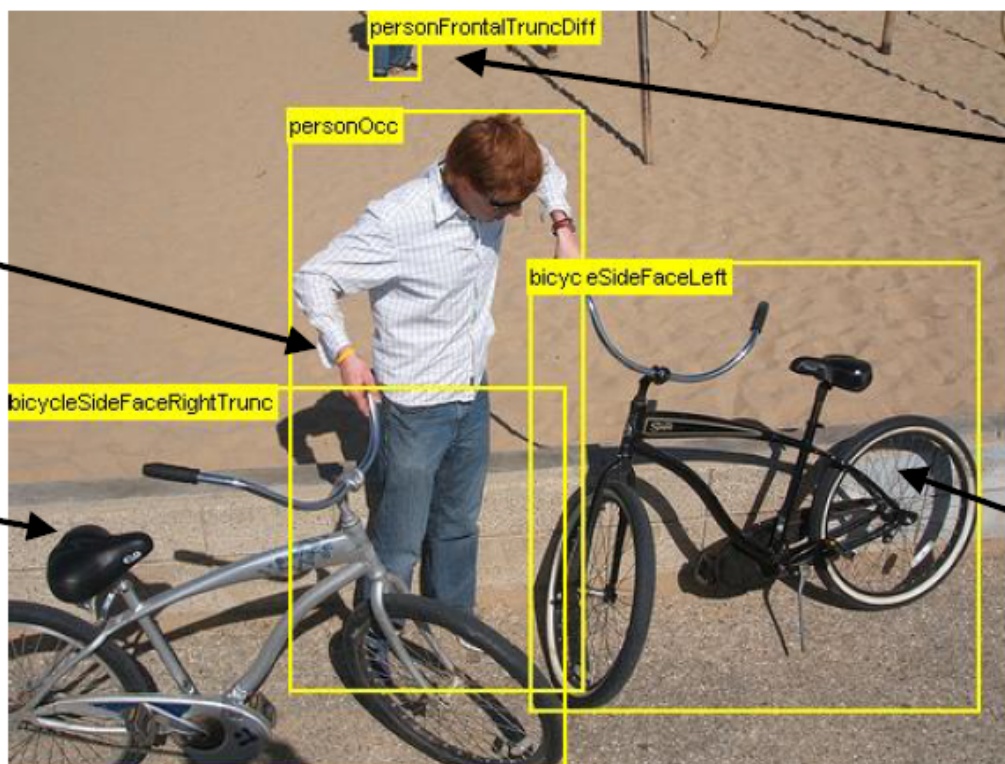
- 20 object categories: aeroplane, bicycle, bird, boat, ...
- Images from **flickr** unfiltered e.g. for difficulty/quality
- Complete bounding box annotation of all objects
- ~20,000 images, 47,000 objects (VOC2010)

Occluded

Object is significantly occluded within BB

Truncated

Object extends beyond BB



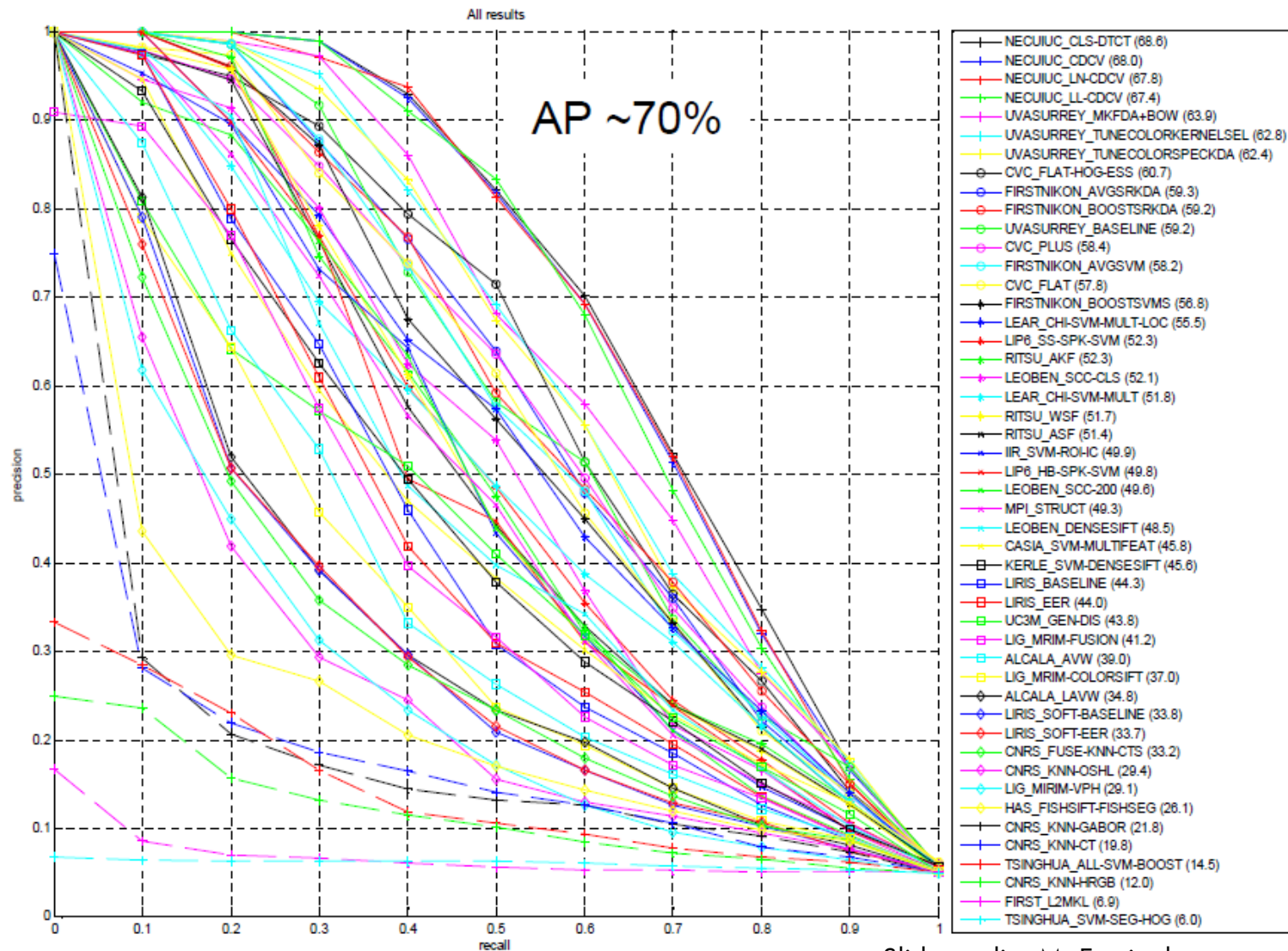
Difficult

Not scored in evaluation

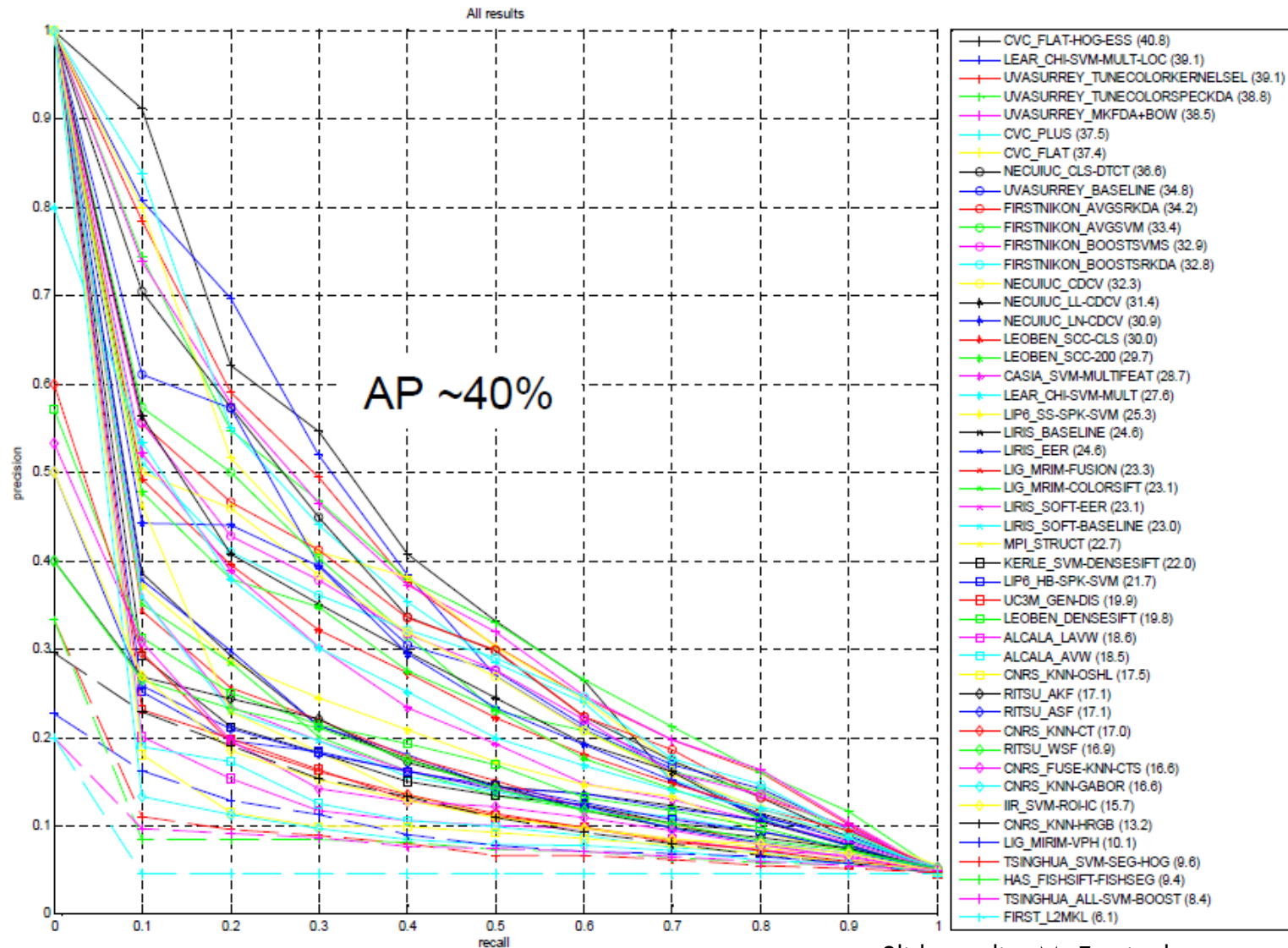
Pose

Facing left

Precision/Recall: Bicycle (VOC 2009)



Precision/Recall: Potted Plant (VOC 2009)

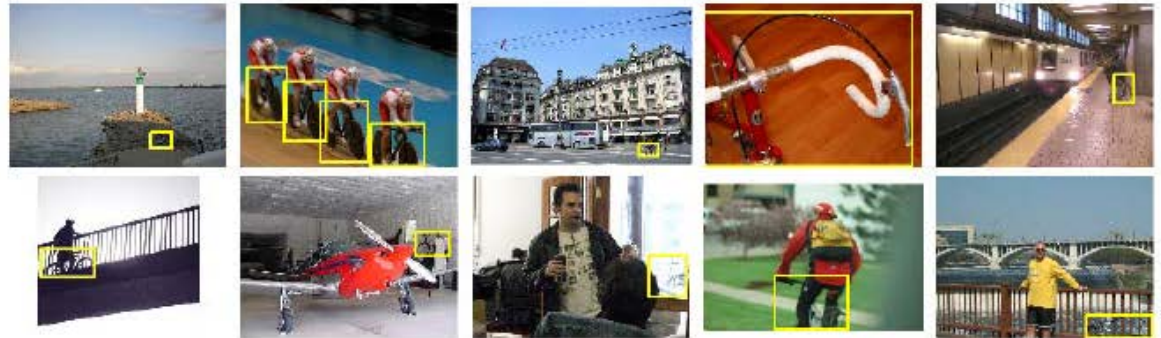


Examples: Bicycle

- Class images:
Highest ranked



- Class images:
Lowest ranked



- Non-class images:
Highest ranked



- “Texture”?

Examples: Cat

- Class images:
Highest ranked



- Class images:
Lowest ranked



- Non-class images:
Highest ranked



- “Composition”?

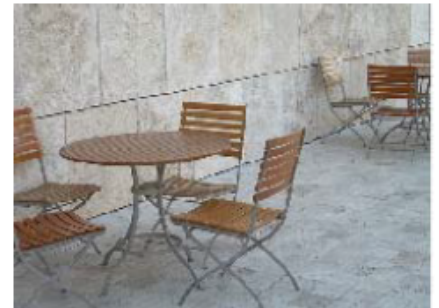
Summary: Bag of Visual Words

- Advantages

- High level of invariance to position and orientation in image
- Fixed length vector \Rightarrow standard machine learning methods applicable
- Can perform well despite clutter
- Very successful for image classification i.e. does image contain X?

- Disadvantages

- Cannot naturally localize objects within the image
- No explicit modelling of spatial information



Выделение объекта

- Известно местоположение объекта, объект выделен на изображении
обнаружение, выделение, локализация объекта (object detection, localization)



Использование классификатора

Изображение разбивается на фрагменты (возможно пересекающиеся)
... и принимается решение относительно категории каждого фрагмента («монитор» или «не монитор»).

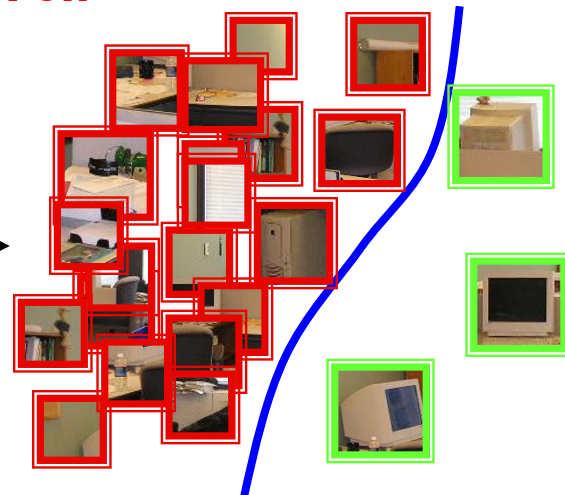
Где мониторы?



Набор фрагментов



Фон

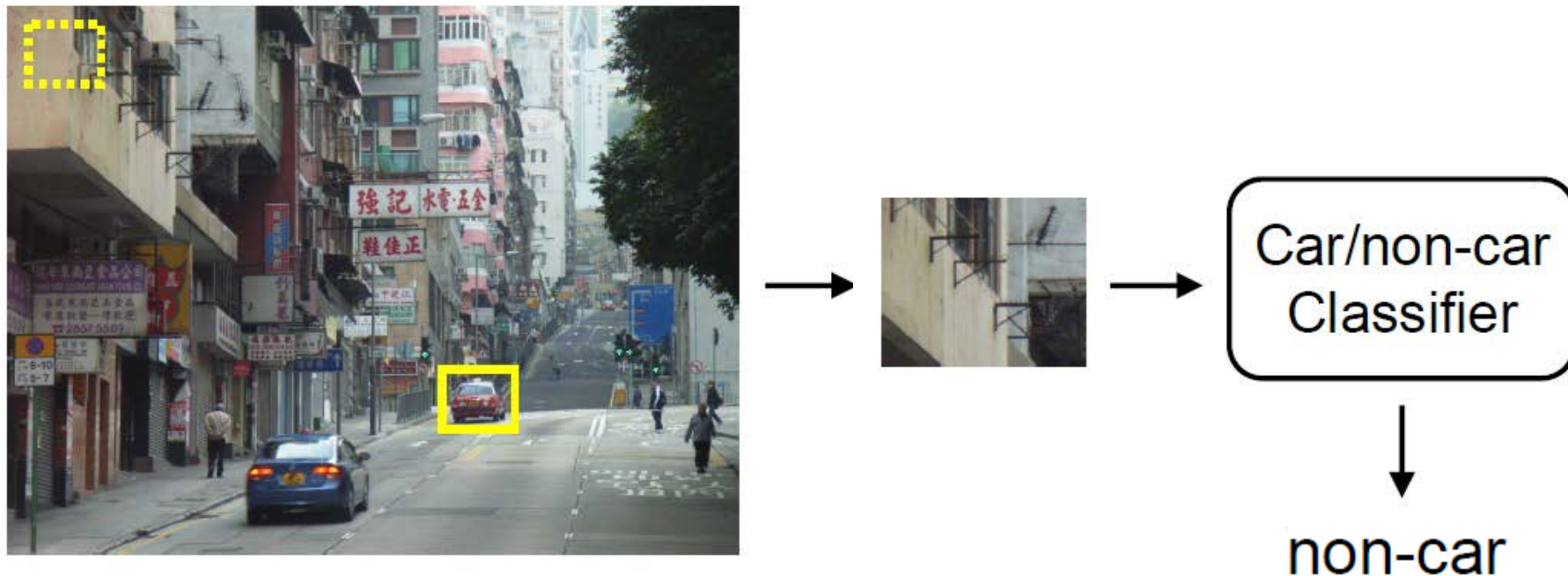


Монитор

В пространстве признаков



Использование скользящего окна



Полный перебор всех позиций и размеров окна!

Использование скользящего окна

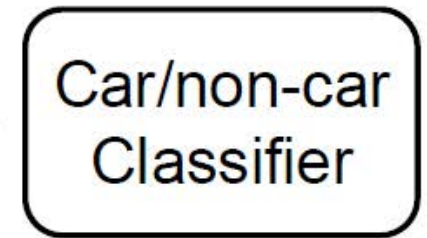


Car/non-car
Classifier

Car

Полный перебор всех позиций и размеров окна!

Использование скользящего окна



Car.

Полный перебор всех позиций и размеров окна!

Feature Extraction

Classification



Does the image contain a car?

Detection



Does the image contain a car?

- Classification: Unknown location + clutter \Rightarrow lots of invariance
- Detection: Uncluttered, normalized image \Rightarrow more "detail"

Добавим информацию о пространственном расположении в модель «мешка слов»



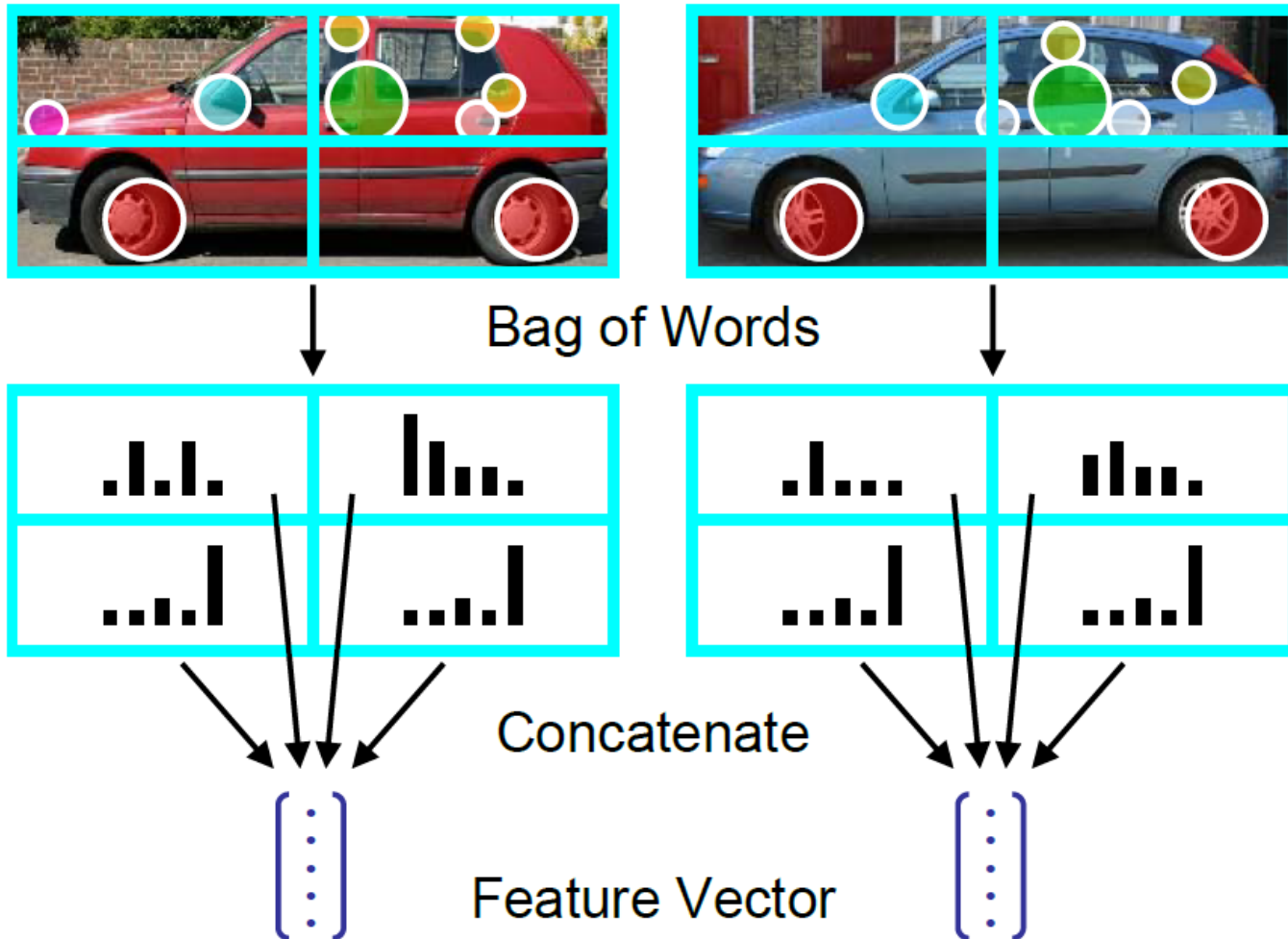
Bag of Words



Feature Vector



Добавим информацию о пространственном расположении в модель «мешка слов»

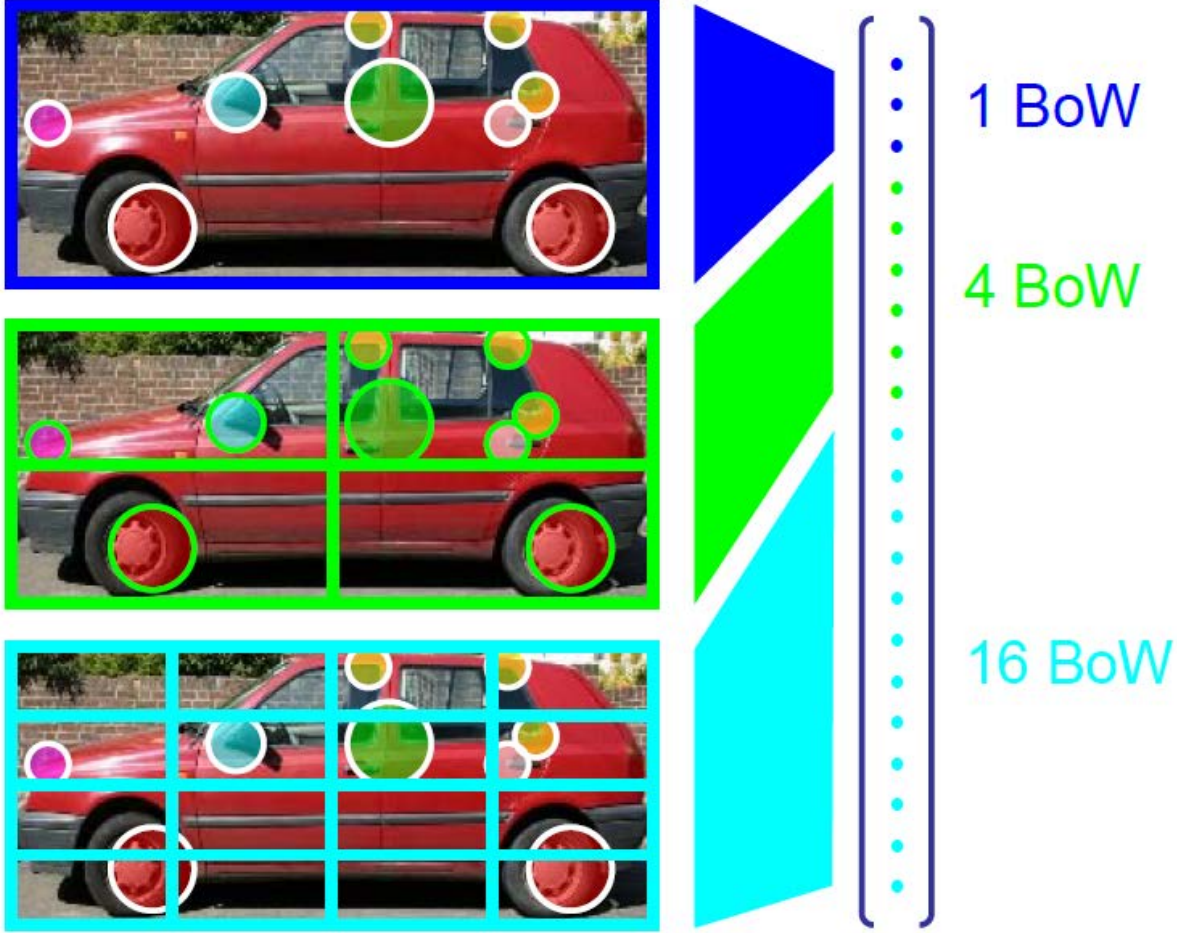


[Fergus et al, 2005]

Slide credit: M. Everingham



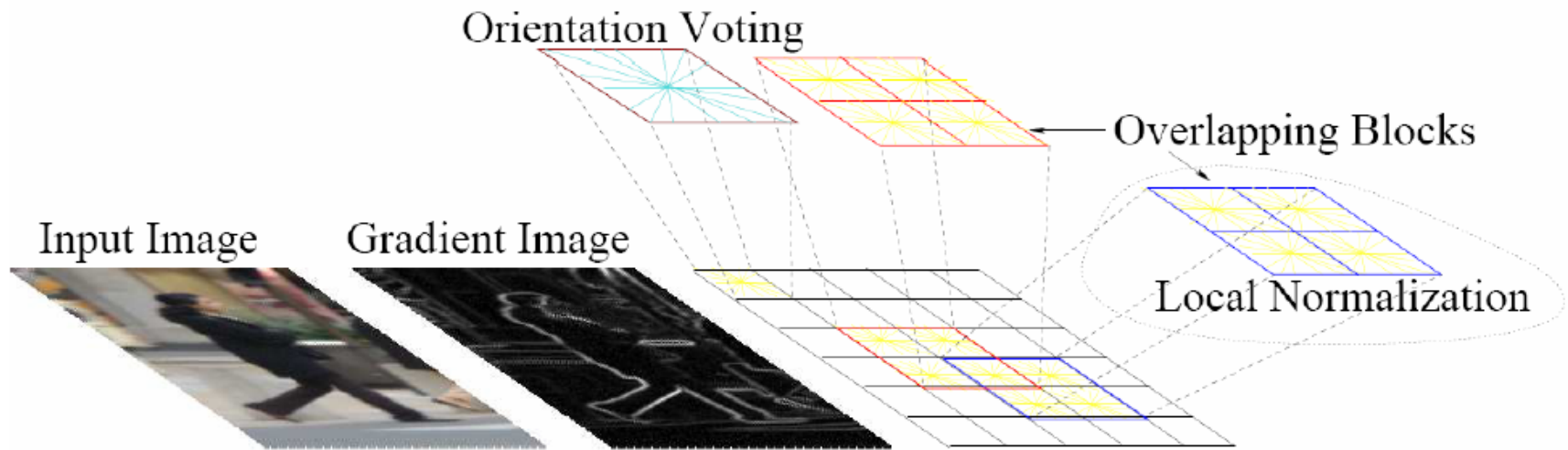
Можно использовать пирамиду



[Lazebnik et al, 2006]



Histogram of Oriented Gradients (HOG)



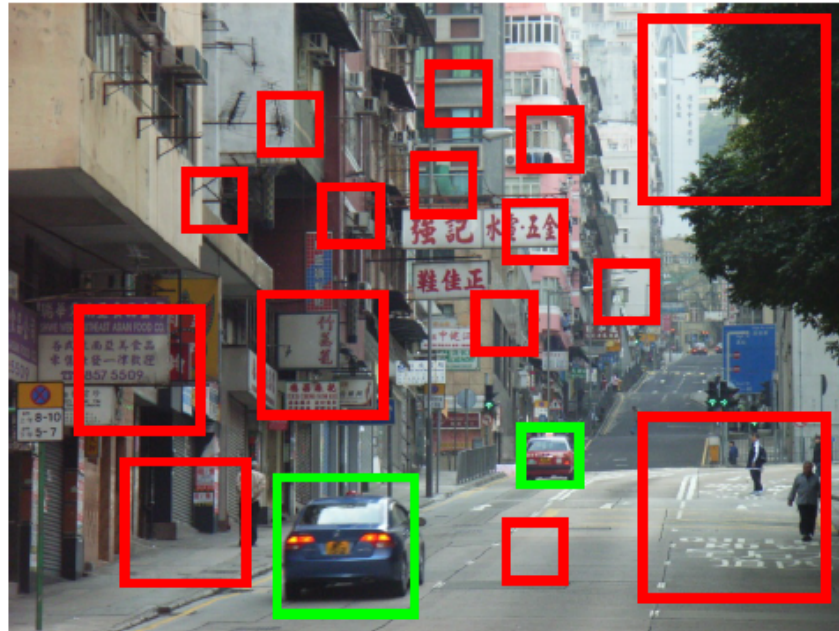
- Adds a second level of overlapping spatial bins re-normalizing orientation histograms over a larger spatial area
- Successfully applied to object category detection

[Dalal & Triggs, 2005]

Slide credit: M. Everingham

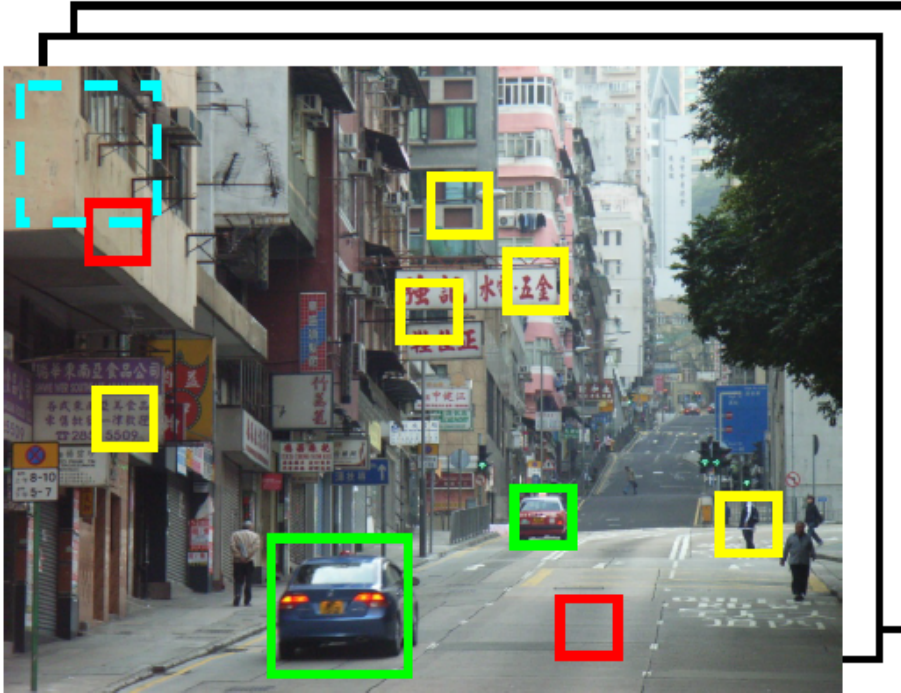
Practical Issues: Training

- Object **detection** is inherently asymmetric: much more “non-object” than “object” data



- Classifier needs to have very low false positive rate
- Non-object category is very complex – need lots of data
- Current learning methods scale poorly with amount of data

Bootstrapping



1. Pick negative training set at random
2. Train classifier
3. Run on training data
4. Add false positives to training set
5. Repeat from 2

- Collect a finite but diverse set of non-object windows
- Force classifier to concentrate on **hard** examples
- For some classifiers can ensure equivalence to training on entire data set

Jittering

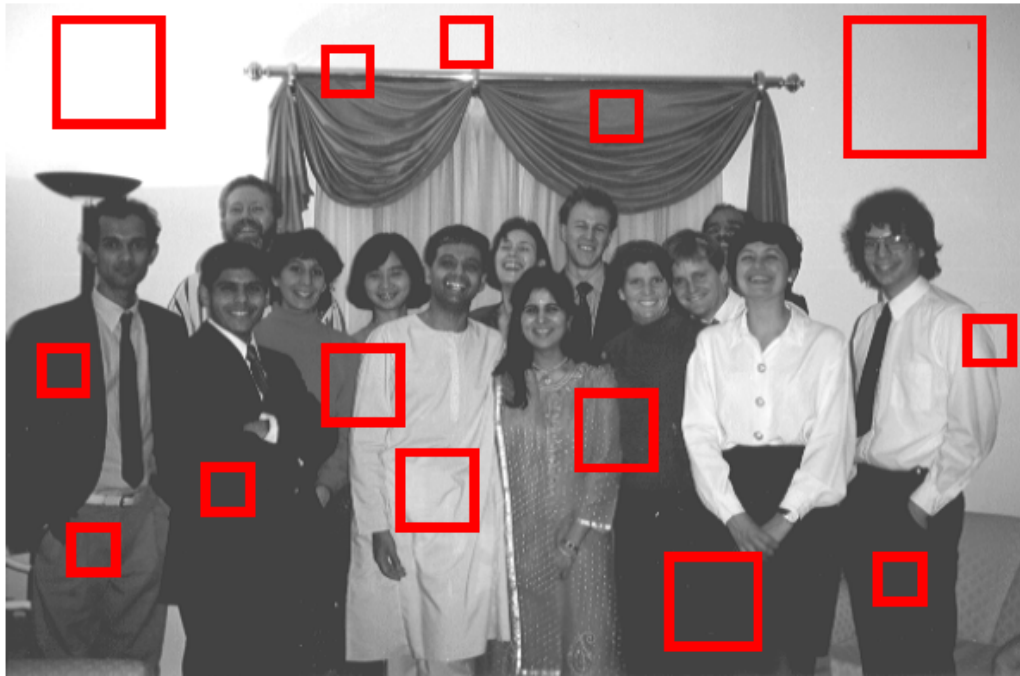
Добавление искусственных положительных примеров в выборку



Отражения, повороты, сдвиги, изменение масштаба, добавление шума

Practical Issues: Speed

- Sliding window search is slow because so many windows are needed e.g. $x \times y \times \text{scale} \approx 100,000$ for a 320×240 image



- Most windows are clearly not the object class of interest
- Can we speed up the search?

Детектор Viola-Jones

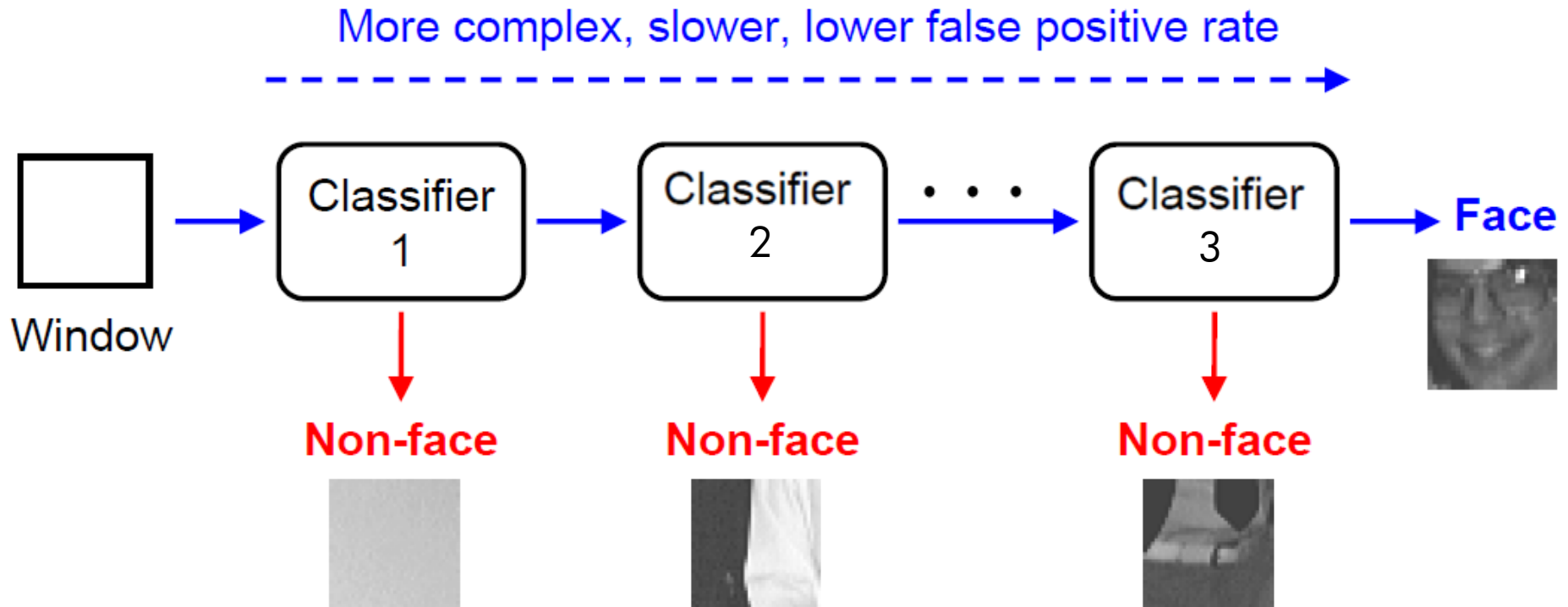
Основные идеи:

- «Быстрые», «простые» признаки объектов
 - Интегральные изображения, свертка с приближенным базисом Хаара
- Использование адаптивного бустинга (AdaBoost) для выбора наиболее информативных признаков
- Использование каскада классификаторов для быстрой отбраковки объектов



Cascaded Classification

- Build a sequence of classifiers with increasing complexity



- Reject easy non-objects using simpler and faster classifiers

Cascaded Classification



- Slow expensive classifiers only applied to a few windows \Rightarrow significant speed-up
- Controlling classifier complexity/speed:
 - Number of support vectors [Romdhani et al, 2001]
 - Number of features [Viola & Jones, 2001]
 - Type of SVM kernel [Vedaldi et al, 2009]

Part arrangement models

- Объекты, как набор составных частей
- Вероятностная модель присутствия объекта на изображении: $P(\text{Object} | \text{Image})$
- Вероятность наличия объекта зависит от:
 - Вероятности наличия необходимых частей объекта «в нужных местах»:
 $P(\text{part } k \text{ is at } [x_k, y_k] | \text{Image})$
 - Вероятности «правильного» расположения частей относительно друг друга
 $P(\text{part } 1 \text{ is at } [x_1, y_1] | \text{Image}) \ \&\& \ P(\text{part } 2 \text{ is at } [x_2, y_2] | \text{Image}) \ \&\& \ \dots$



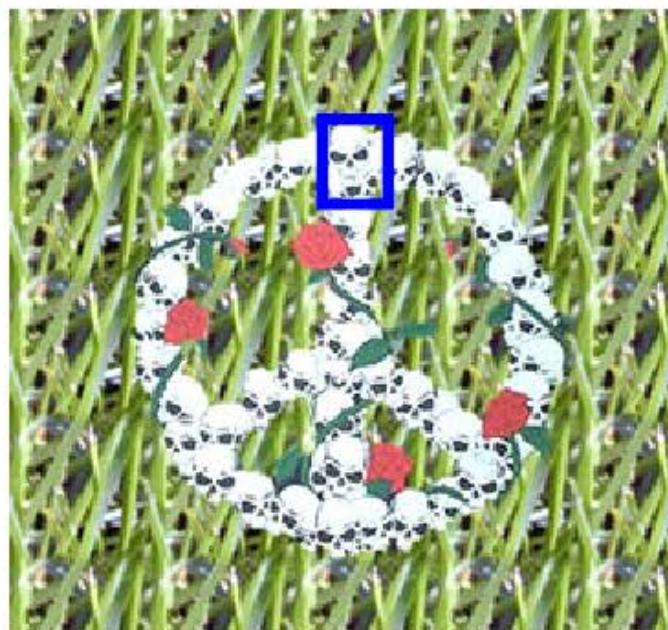
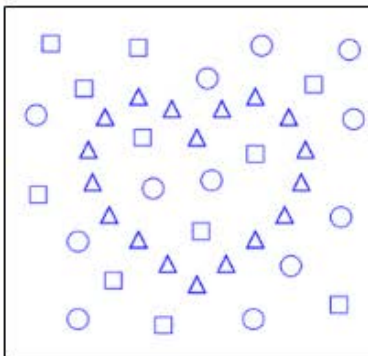
Part arrangement models

- Голосование
- Ограничение на углы
- Использование Марковских случайных полей



Detecting Similar Objects & Actions by Matching Local Self-Similarities

[CVPR'07 -- E. Shechtman & M. Irani]

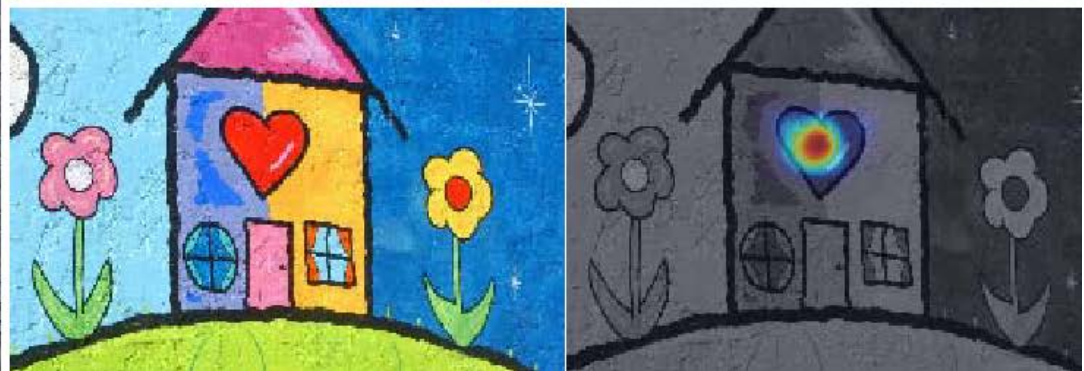
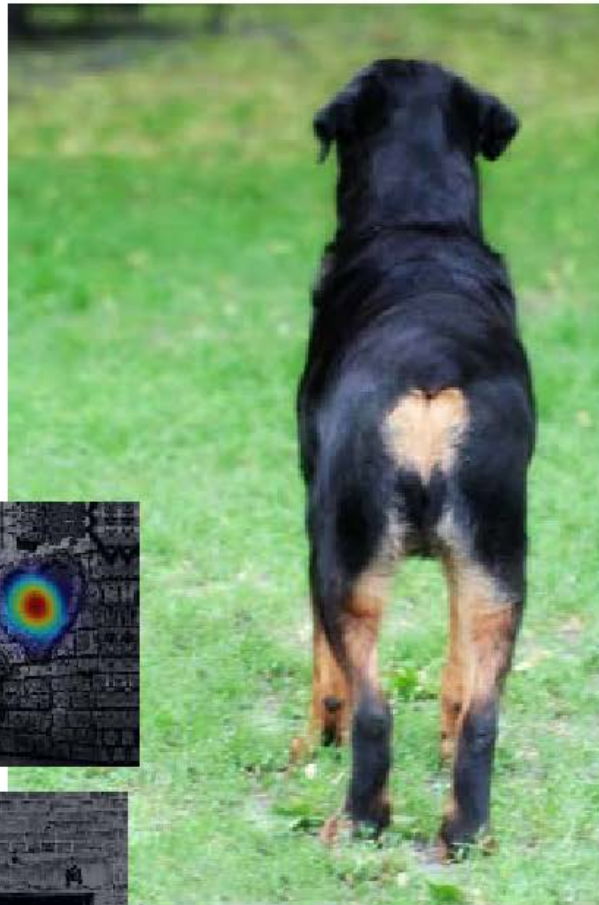


Slide credit: M. Irani

Object detection

[CVPR'07 -- E. Shechtman & M. Irani]

Query image:

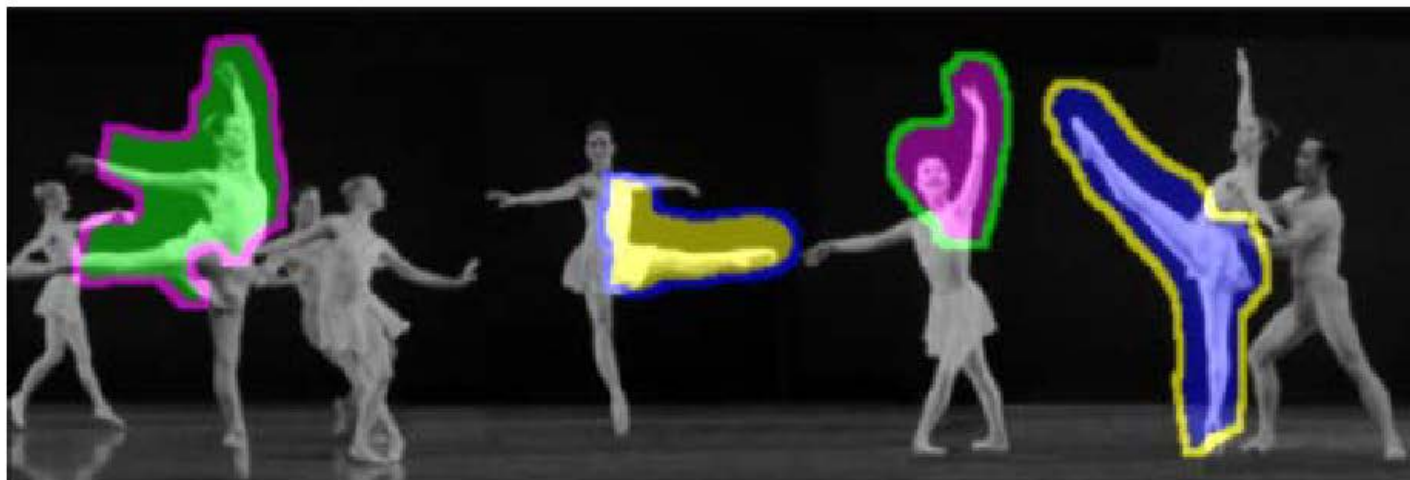


Slide credit: M. Irani

Similarity by Composition

[NIPS'06 -- *O. Boiman & M. Irani*]

LESS similar



similar



Заключение

- Определение категории и выделение объектов решается при помощи классификации
- Определение категории: классификация локальных (глобальных) признаков с использованием машинного обучения
- Выделение объекта – использование скользящего окна
- Проблемы
 - Репрезентативность обучающей выборки
 - Bootstrapping and jittering
 - Скорость работы
 - Каскады классификаторов
 - Зависимость от контекста

