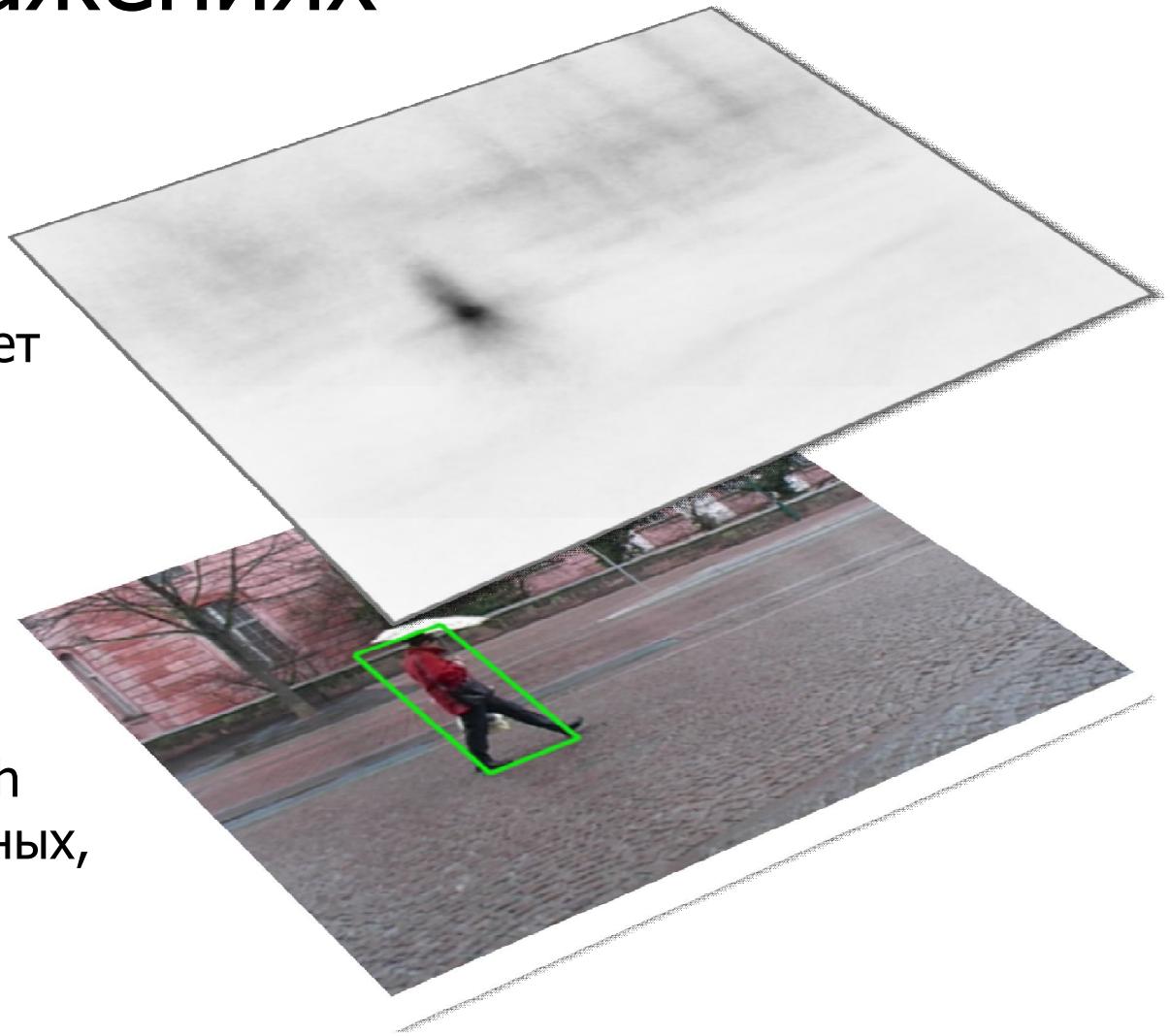


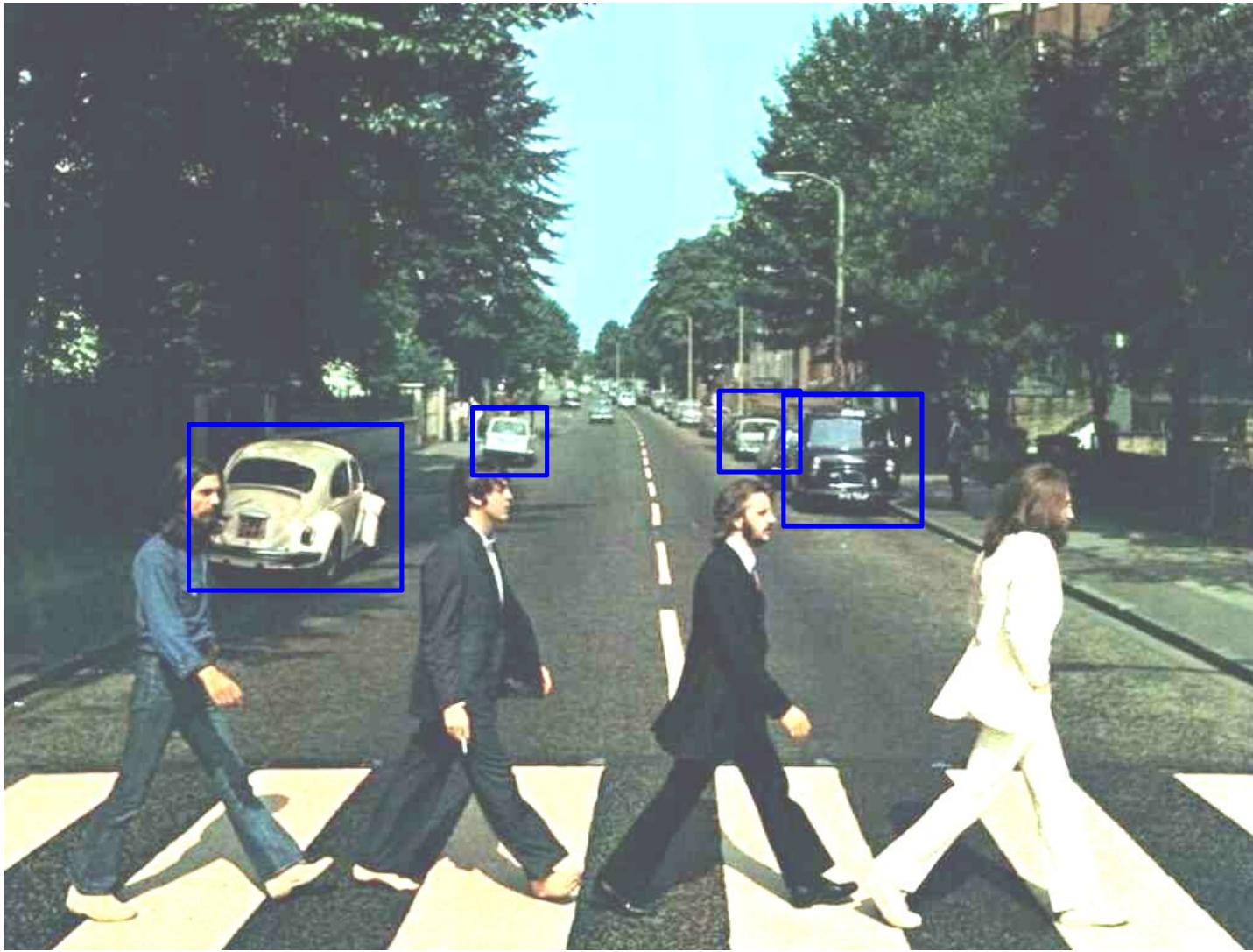
Локализация объектов на фотоизображениях

Виктор Лемпицкий,
Оксфордский Университет

Яндекс-Microsoft Research
семинар по анализу данных,
Москва, 2010

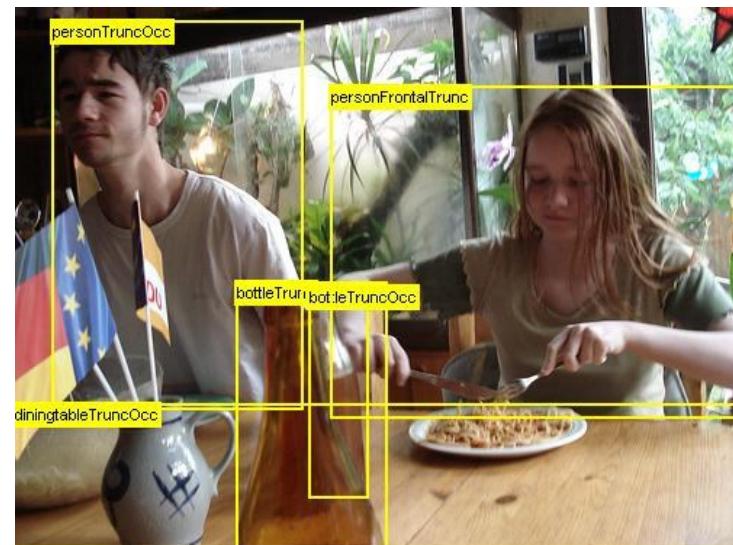
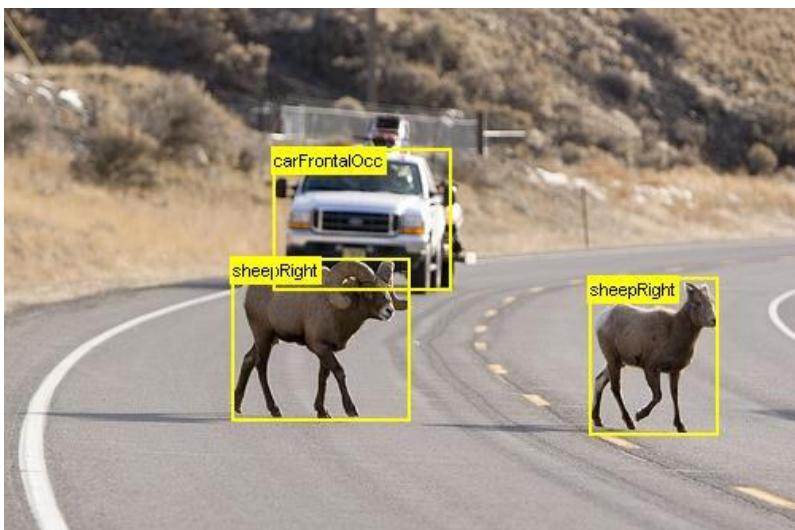


Локализация категорий объектов



Зачем это нужно?

«Что изображено на картинке?»

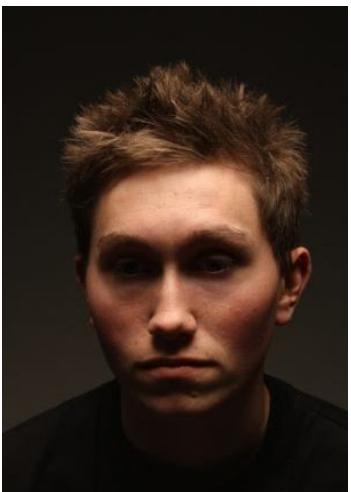


*PASCAL Visual Object Cognition Challenge
(Everingham et al. 2009)*

Сложность задачи



изменчивость угла зрения



изменчивость масштаба
/размеров

изменчивость освещения, теней, текстур

Сложность задачи



Закрытия

изменчивость формы

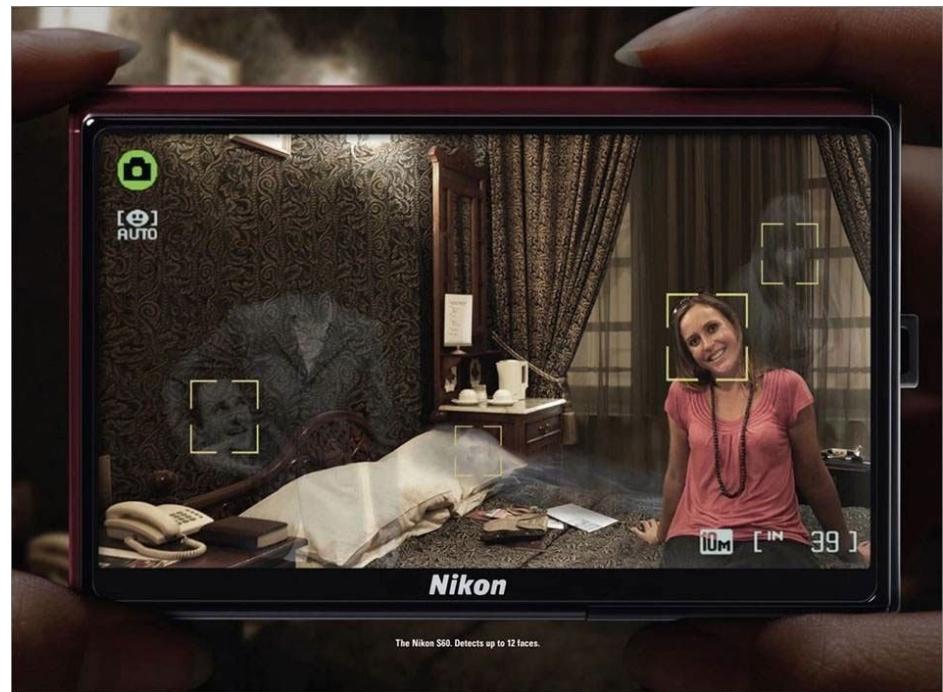


изменчивость позы

Нахождение лиц



Viola & Jones system, CVPR 2001



«Обычный» фотоаппарат сегодня

Локализация пешеходов

Пример «нерешенной» задачи

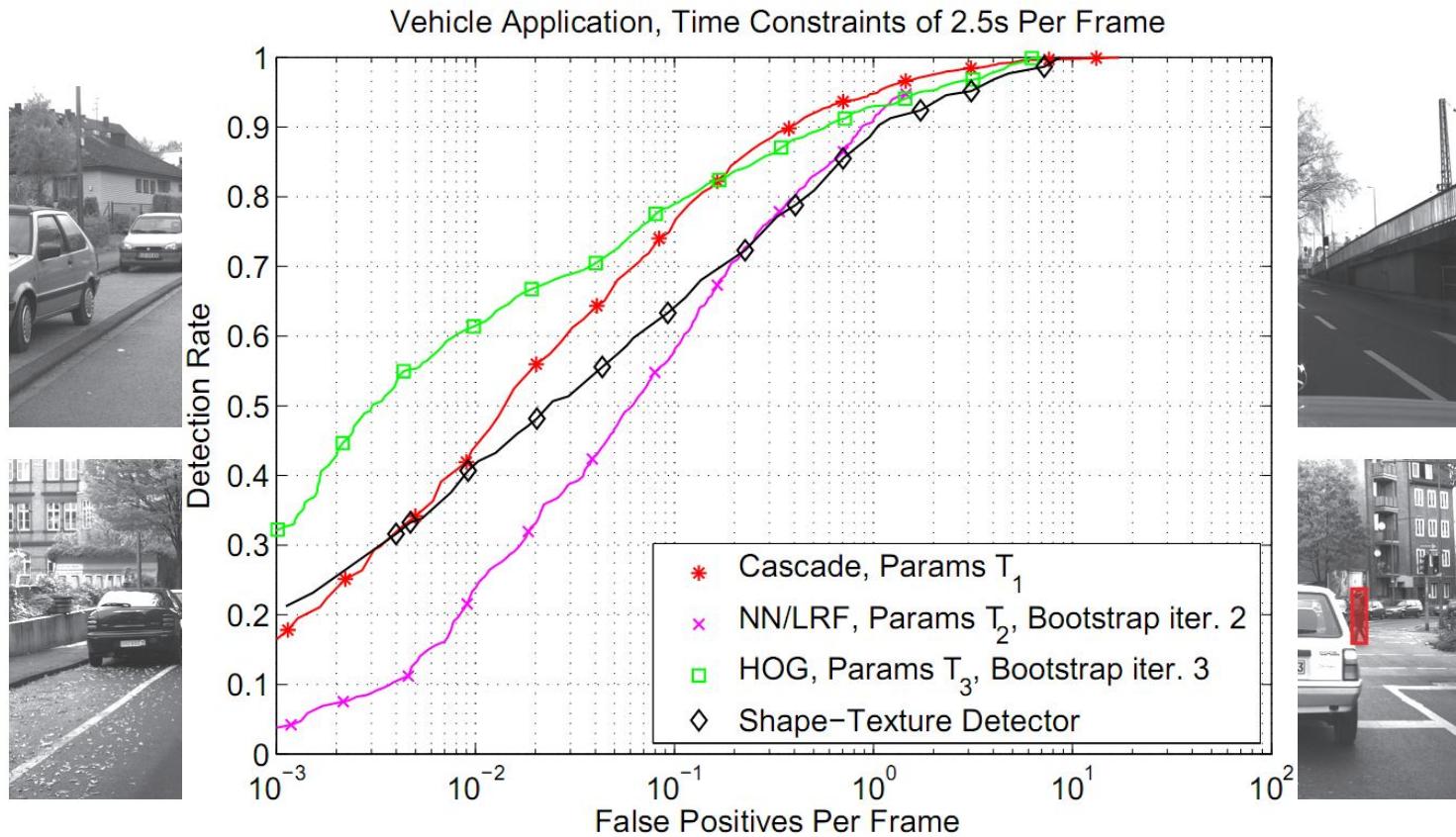


Enzweiler & Gavrila. Monocular Pedestrian Detection: Survey and Experiments. TPAMI'09

- Недостаточная скорость/аккуратность
- можно улучшить засчёт стерео камер и анализа видео

Локализация пешеходов

Пример «нерешенной» задачи



Enzweiler & Gavrila. Monocular Pedestrian Detection: Survey and Experiments. TPAMI'09

- Недостаточная скорость/аккуратность
- можно улучшить засчёт стерео камер и анализа видео

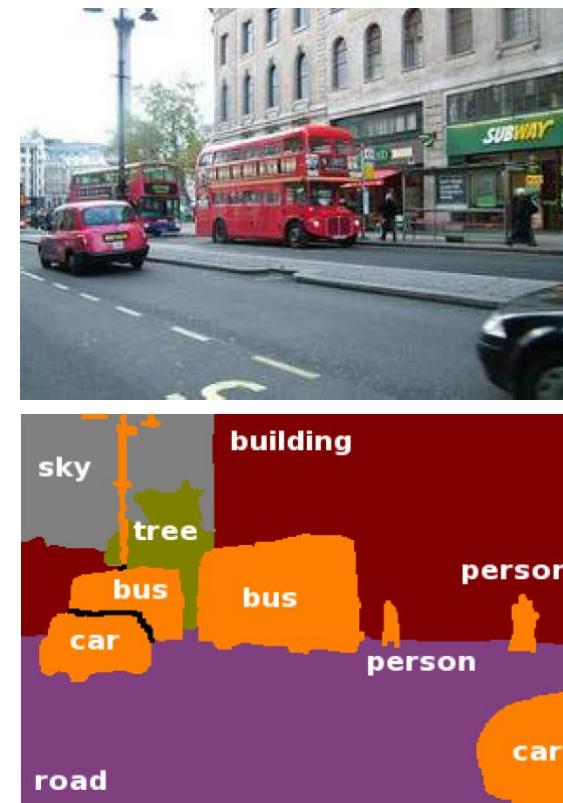
Нахождение пешеходов



Video from Dariu Gavrilă/Daimler

Схожие задачи

1. Нахождение конкретных объектов
2. Локализация для видео
3. Другие модальности:
 - Спутниковые снимки
 - Медицинские изображения
 - Стерео/range maps
4. Семантическая сегментация
5. Оценка конфигурации
(выражение лица, поза человека...)



Gould, Fulton, Koller, ICCV'09

План на сегодня

Обзор популярных методов:

- Общее в подходах к локализации
- Метод Далала-Тригса
- Метод Виолы-Джонса
- Визуальные словари и методы «наборов слов»
- Хаф-преобразования и «Неявная модель формы»

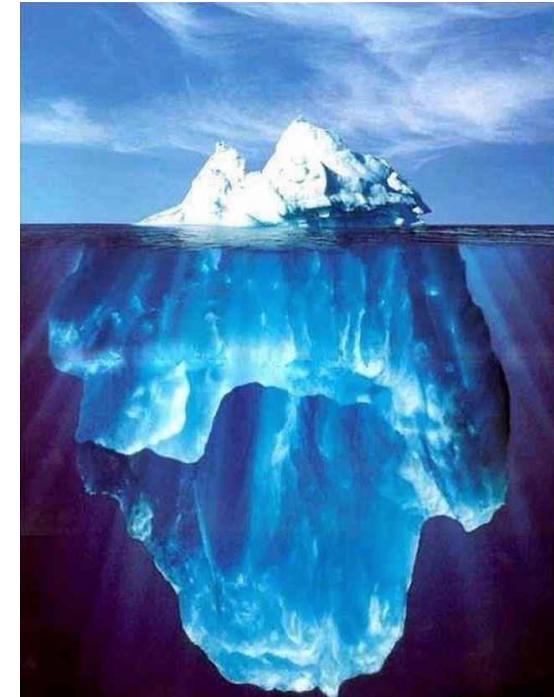
Новый метод:

- Обнаружение с помощью Хаф-лесов

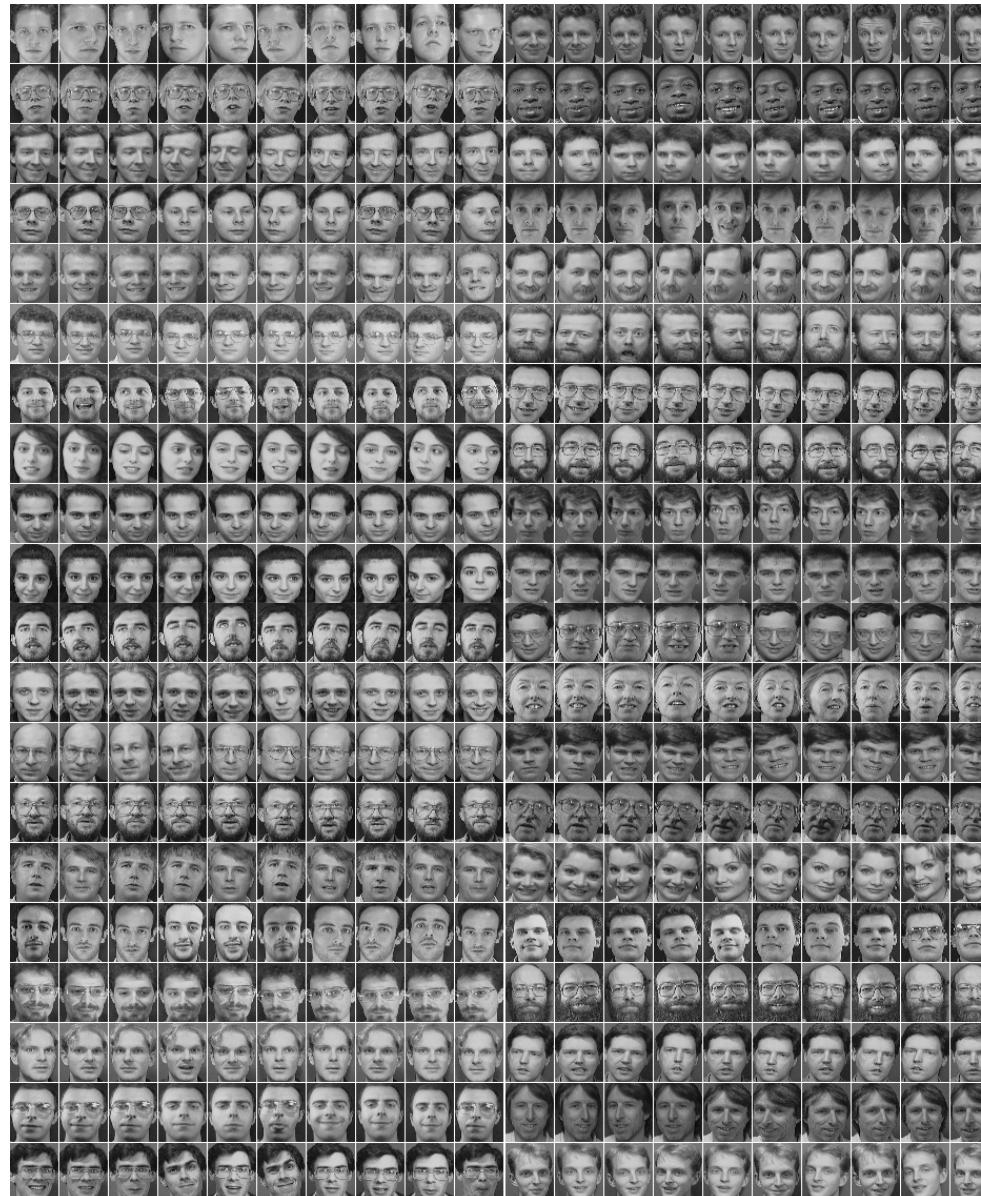
Открытые направления

Disclaimers

- Сегодня “верхушка айсберга” (сотни статей в год!)
- Пропущены важные направления:
 - Pictorial structures
 - Generative models
- Отличные курсы (слайды online):
 - *Li Fei-Fei, R. Fergus, A. Torralba, "Recognizing and Learning Object Categories", ICCV'05, CVPR'07, ICCV'09*
 - *B. Leibe, K. Grauman, "Visual Object Recognition", AAAI'08*
- Сегодня объяснение «на пальцах», математика в статьях

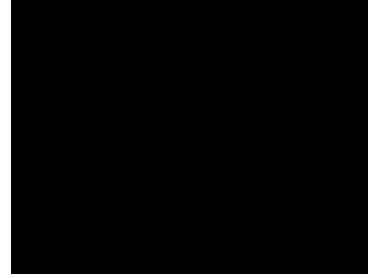
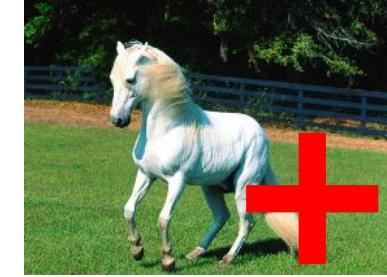
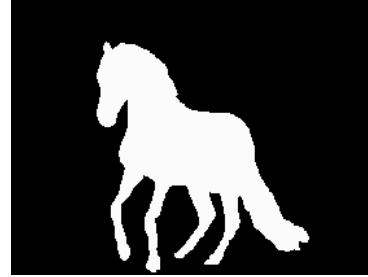
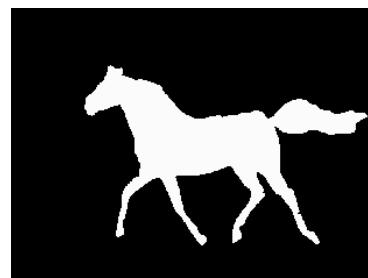
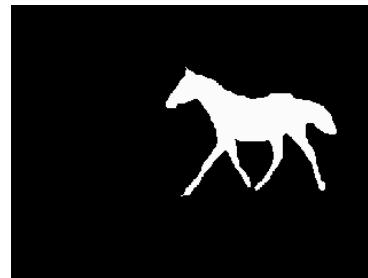


Обучение на примерах



Olivetti faces dataset

Уровни аннотации



Dataset
("Weizmann Horses")

пиксели

«окна»

изображения

“Скользящее окно”

Обучение: тренировка бинарного классификатора на окнах



“Класс 1” = объект

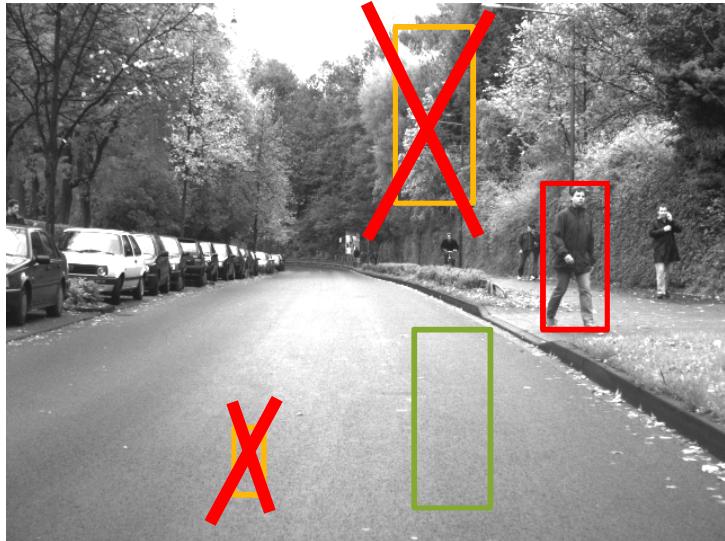
“Класс 0” = фон

Применение:

1. применить классификатор к каждому окну
2. Найти локальные максимумы выше порога



Ускорение “Скользящего окна”



1. Геометрия сцены
2. Каскад
3. Branch-and-Bound



Представление изображений

Шаг 1: тренировка бинарного классификатора $f(\text{окно})$

Машинное обучение

компьютерное зрение

«Наивное» представление:



Вектор интенсивностей

Тренировочный пример

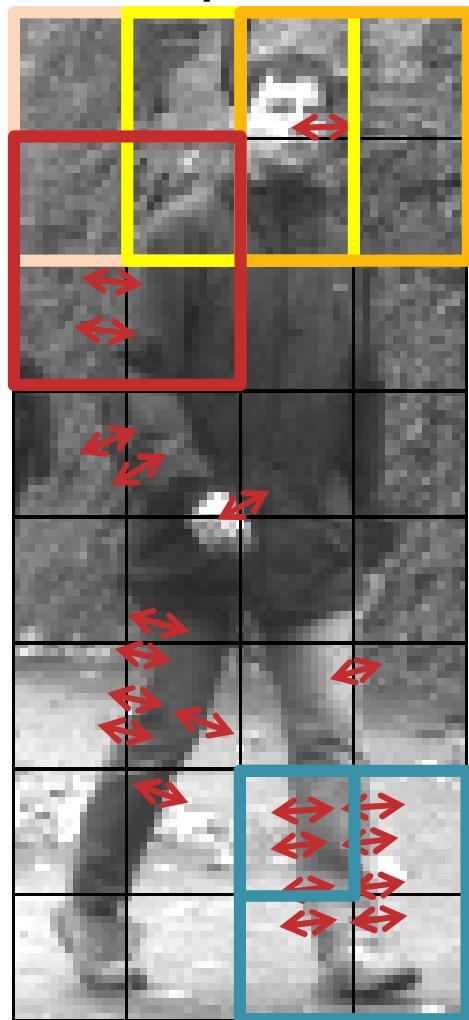


«Схожие» примеры

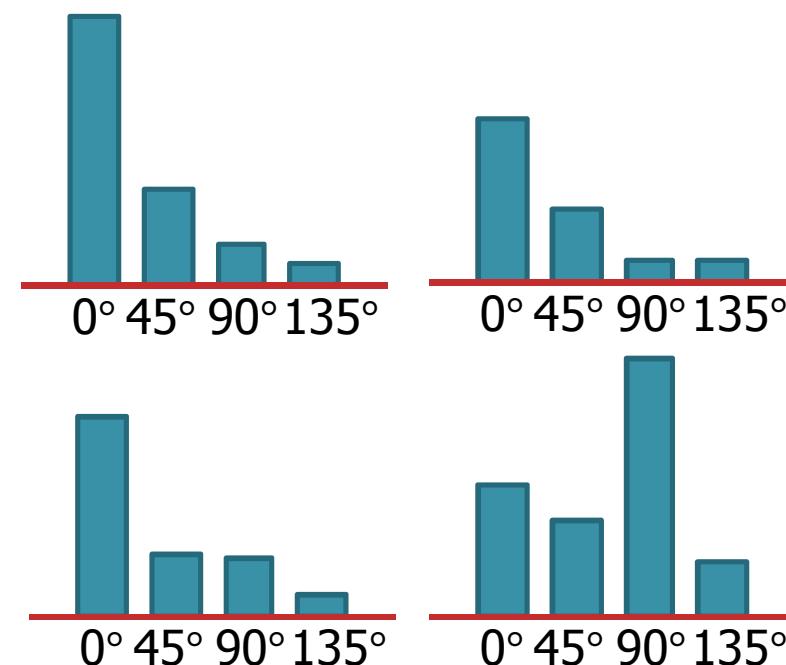
Далеки в «наивном» представлении!

Гистограммы Ориентированных Градиентов (HOG)

Dalal & Triggs, CVPR 2005

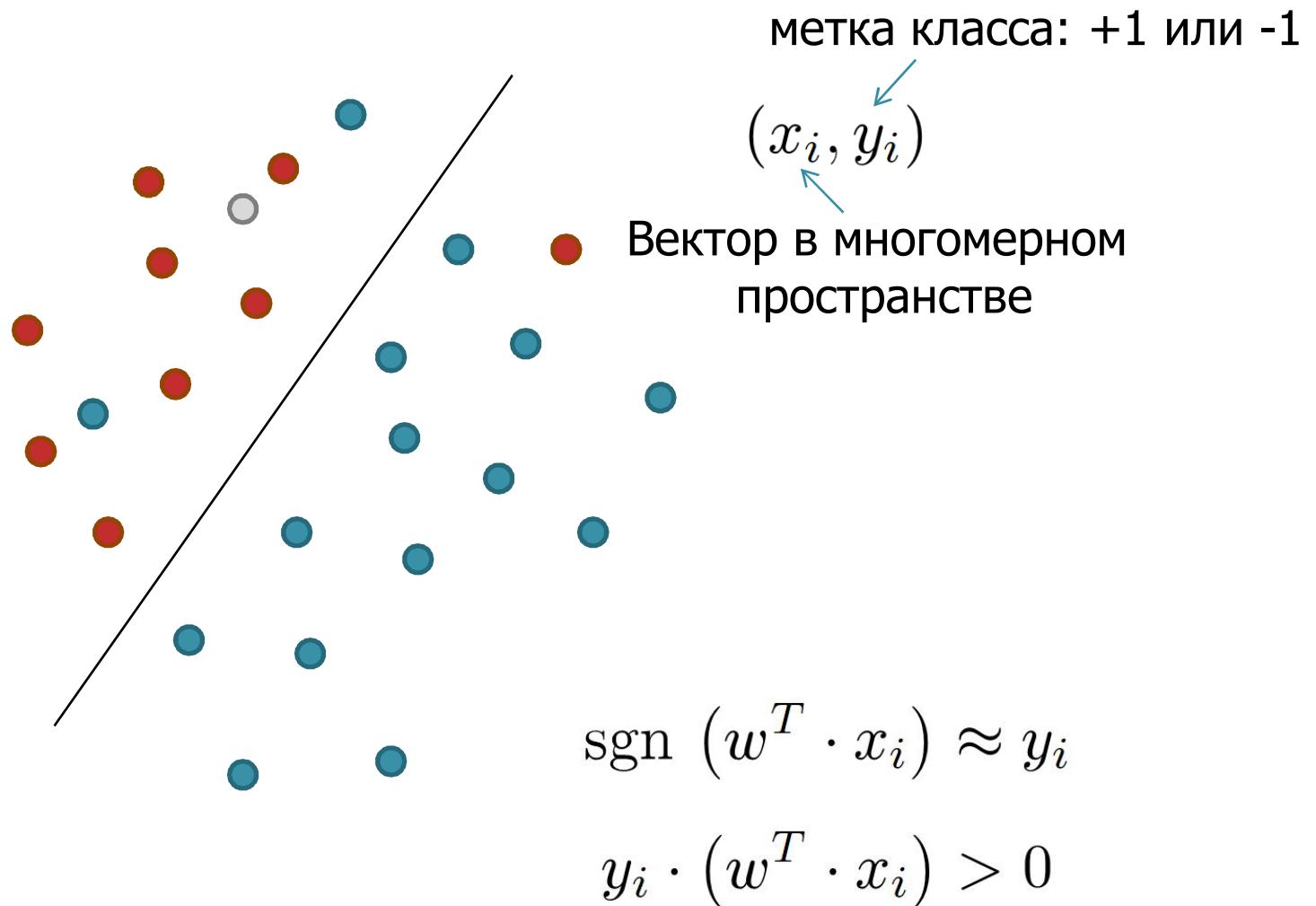


1. Подсчет ориентированных градиентов
2. Описание клеток гистограммами
3. Нормализация по блокам
4. Стыковка

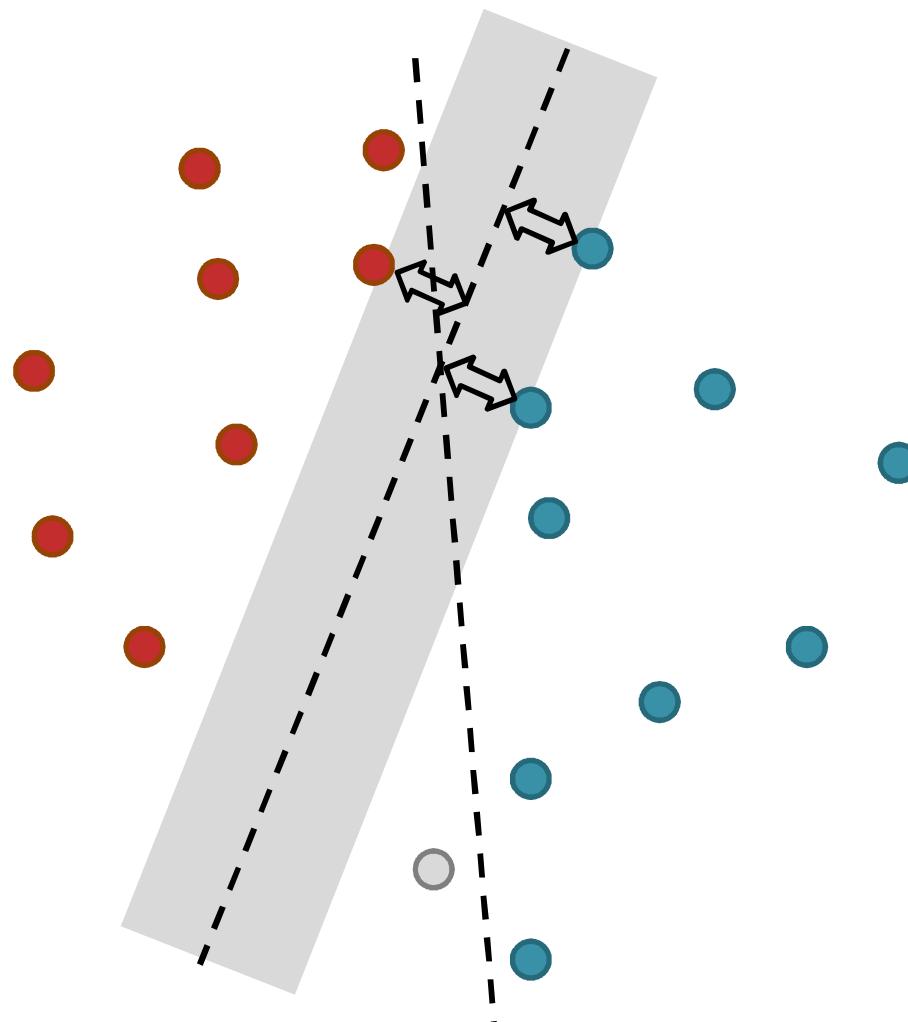


- Информация о форме сохранена
- Инвариантность к малым деформациям
- Инвариантность к существенным фотометрическим изменениям
- Похожие дескрипторы: SIFT, SURF, DAISY, etc.

Обучение бинарного классификатора



Линейный метод опорных векторов



Перцептрон Розенблatta:

$$y_i \cdot (w^T \cdot x_i) > 0$$

SVM Вапника :

$$y_i \cdot (w^T \cdot x_i) > 1$$

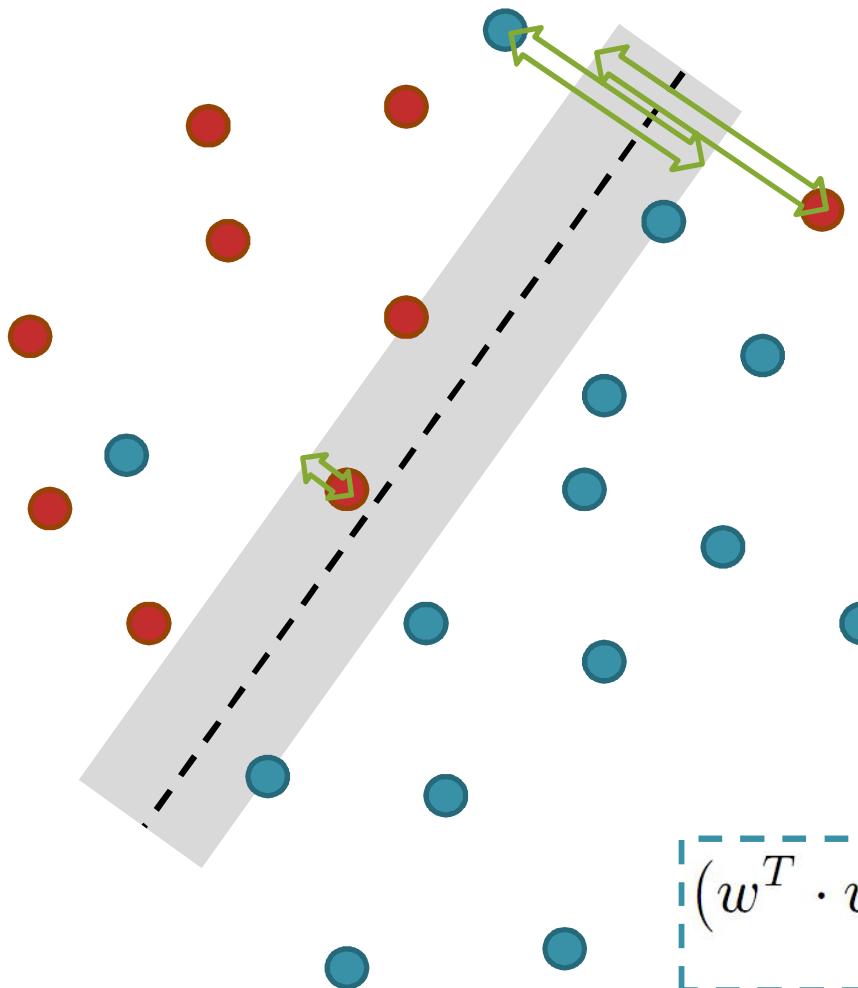
Максимизация отступа:

$$(w^T \cdot w) \rightarrow \min$$

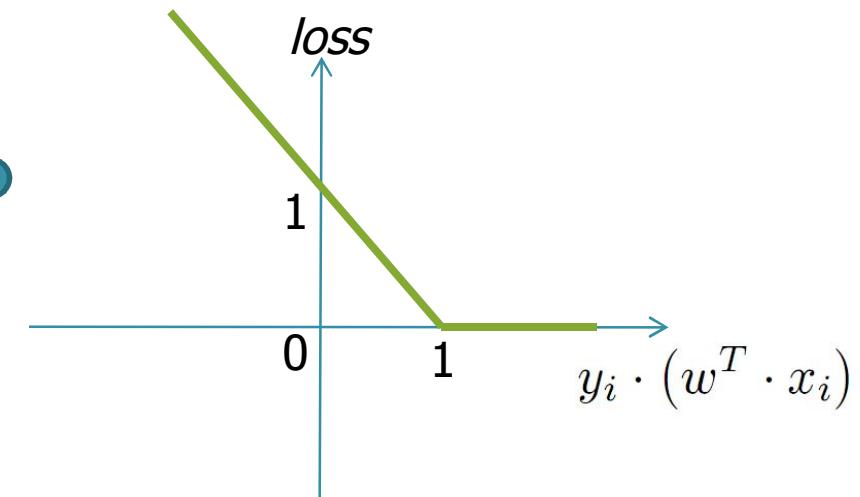
Выпуклое квадратичное программирование – эффективные алгоритмы!

Линейный метод опорных векторов

«Смягчение» граничных условий:
[Cortes&Vapnik, ML'93]



$$y_i \cdot (w^T \cdot x_i) > 1$$



$$(w^T \cdot w) + C \sum_i L(y_i \cdot (w^T \cdot x_i)) \rightarrow \min$$

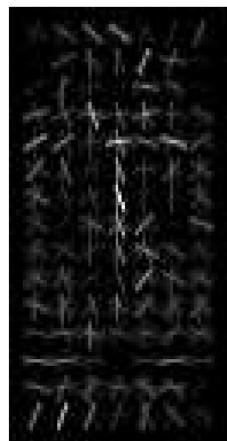
Метод Далала-Тригса

Dalal&Triggs, CVPR 2005: HOG+линейная SVM

From Navneet Dala's presentation:



$W+$



$W-$



Метод Виолы-Джонса

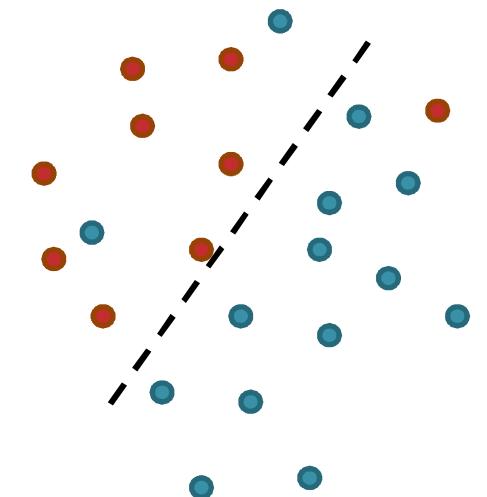
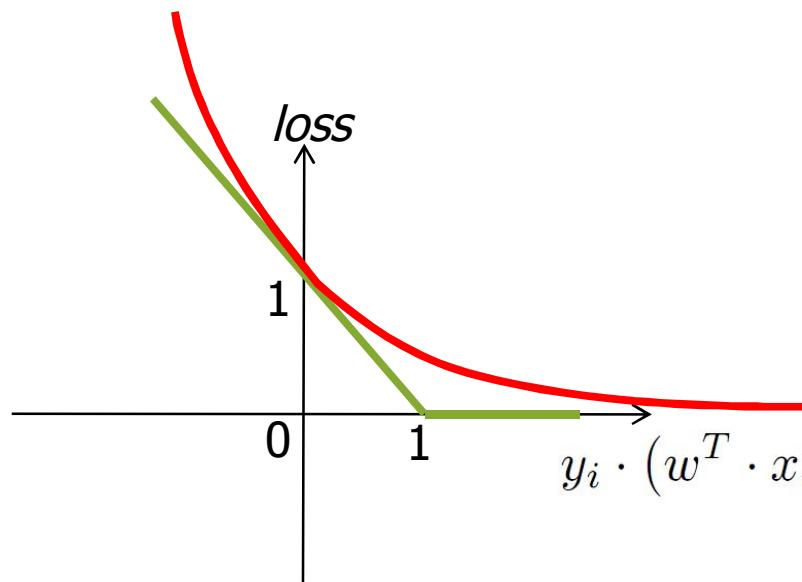
Представление (“признаки Хаара”):



Изображение представляется комбинаторным числом
бинарных признаков

Бустинг классификаторов

- Вместо глобальной оптимизации ограниченное число шагов по координатному спуска



Метод опорных векторов: $(w^T \cdot w) + C \sum_i L(y_i \cdot (w^T \cdot x_i)) \rightarrow \min$

АдаБуст: $\sum_i \underline{\exp}(-y_i \cdot (w^T \cdot x_i)) \rightarrow \min$

Бустинг классификаторов

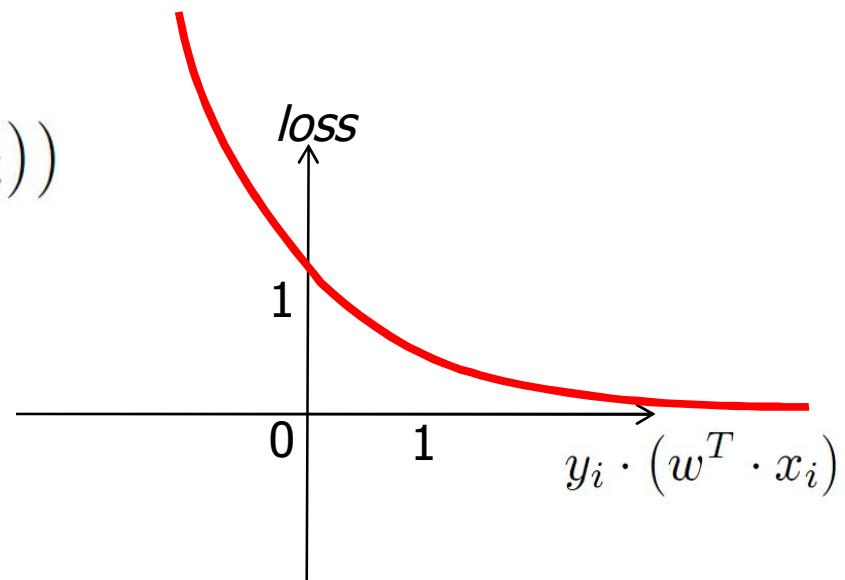
Функционал в АдаБусте:

$$F(w) = \sum_i \exp(-y_i \cdot (w^T \cdot x_i))$$

$$w \leftarrow 0$$

Повторять:

1. Выбор большого подмножества координат (признаков)
2. Выбор координаты j с наибольшим $-\frac{\partial F}{\partial w^j}$
3. Спуск вдоль координаты j до минимума $F(w)$
4. Остановка через несколько сотен/тысяч шагов

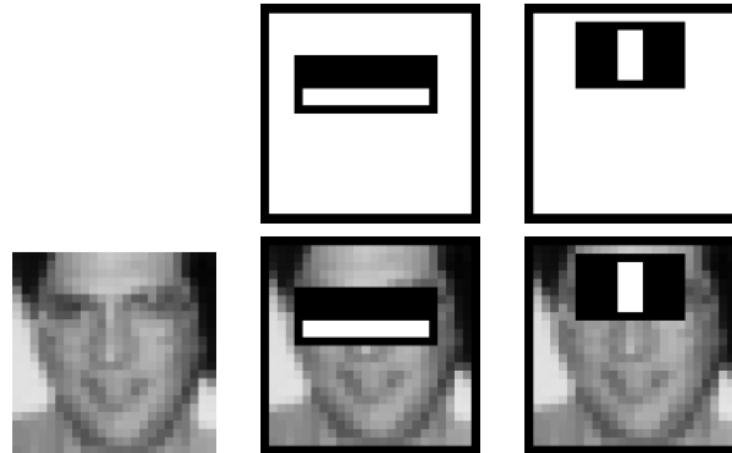


Поиск лиц методом Виолы-Джонса

[Viola & Jones, CVPR'01]

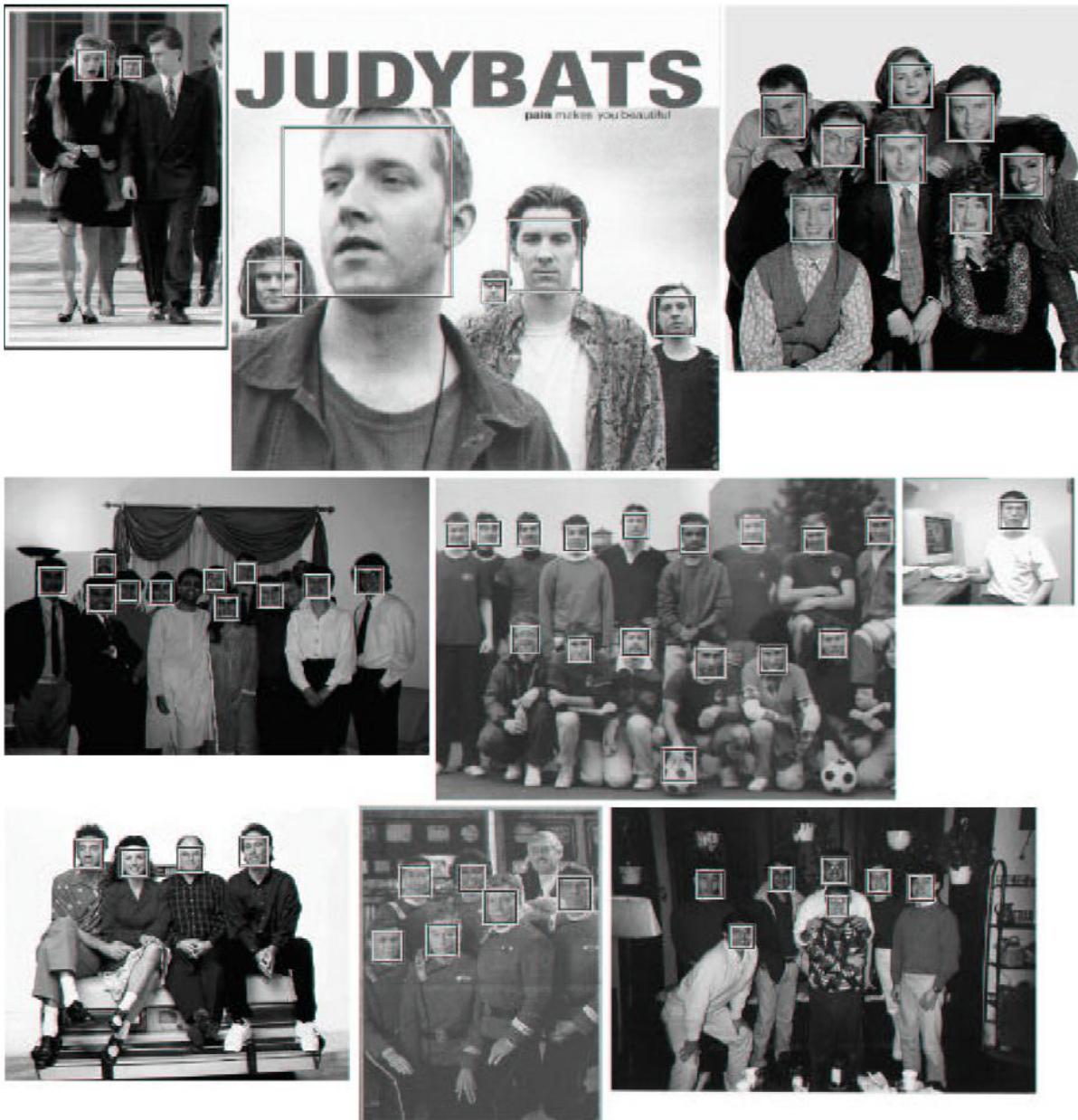


«Положительные» примеры



1-ый и 2-ой признаки,
выбранные АдаБустом

Поиск лиц методом Виолы-Джонса



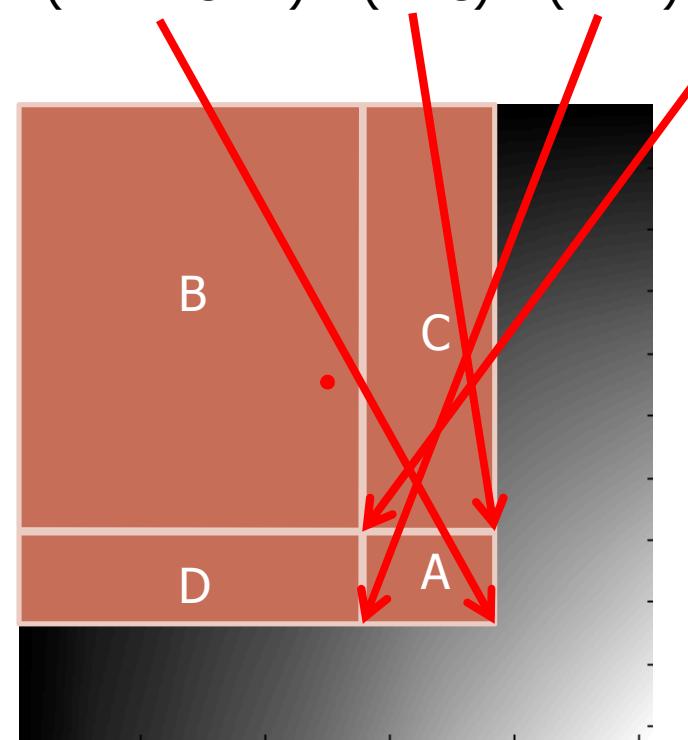
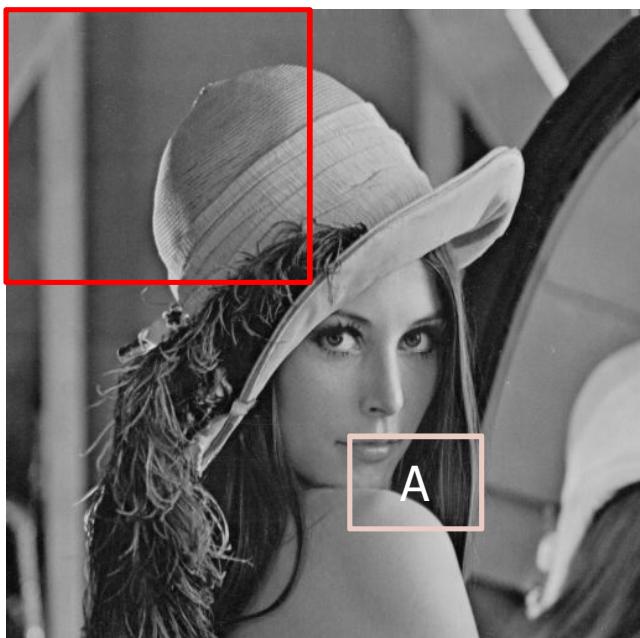
Ускорение:

- Интегральные изображения
- Каскад

Интегральные изображения

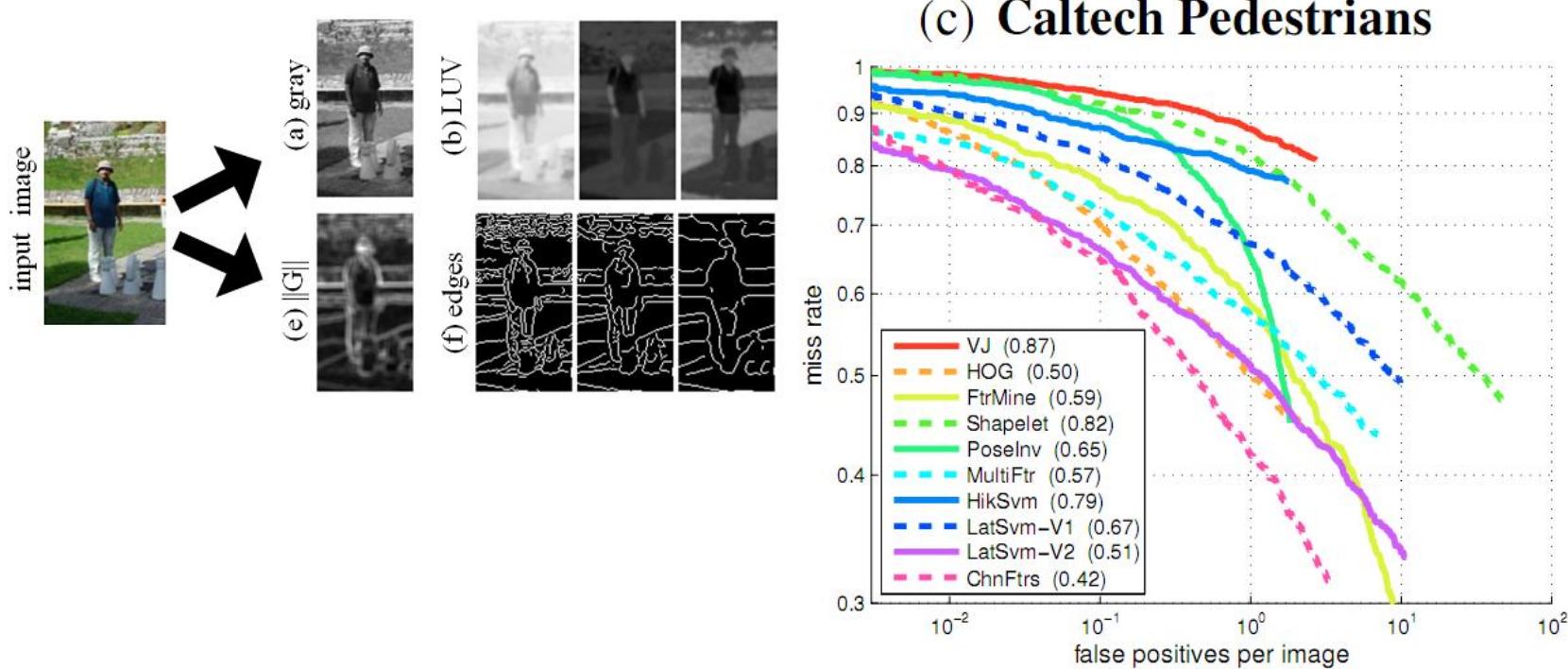
Как быстро считать прямоугольные суммы?

$$A = (A+B+C+D) - (B+C) - (B+D) + B$$



Далал-Тригс vs. Виола-Джонс

- Для лиц Виола-Джонс работает почти идеально
- Для пешеходов Далал-Тригс аккуратнее (но медленнее)
- [*Dollar et al. BMVC'09*]: Виола-Джонс на правильных каналах не хуже Далала-Тригса

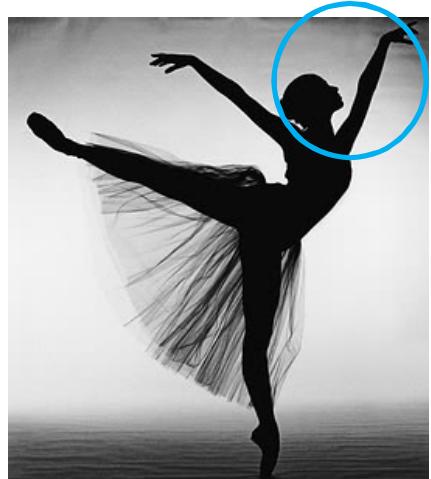


Сложности



Изменчивость
частей меньше
изменчивости
целого!

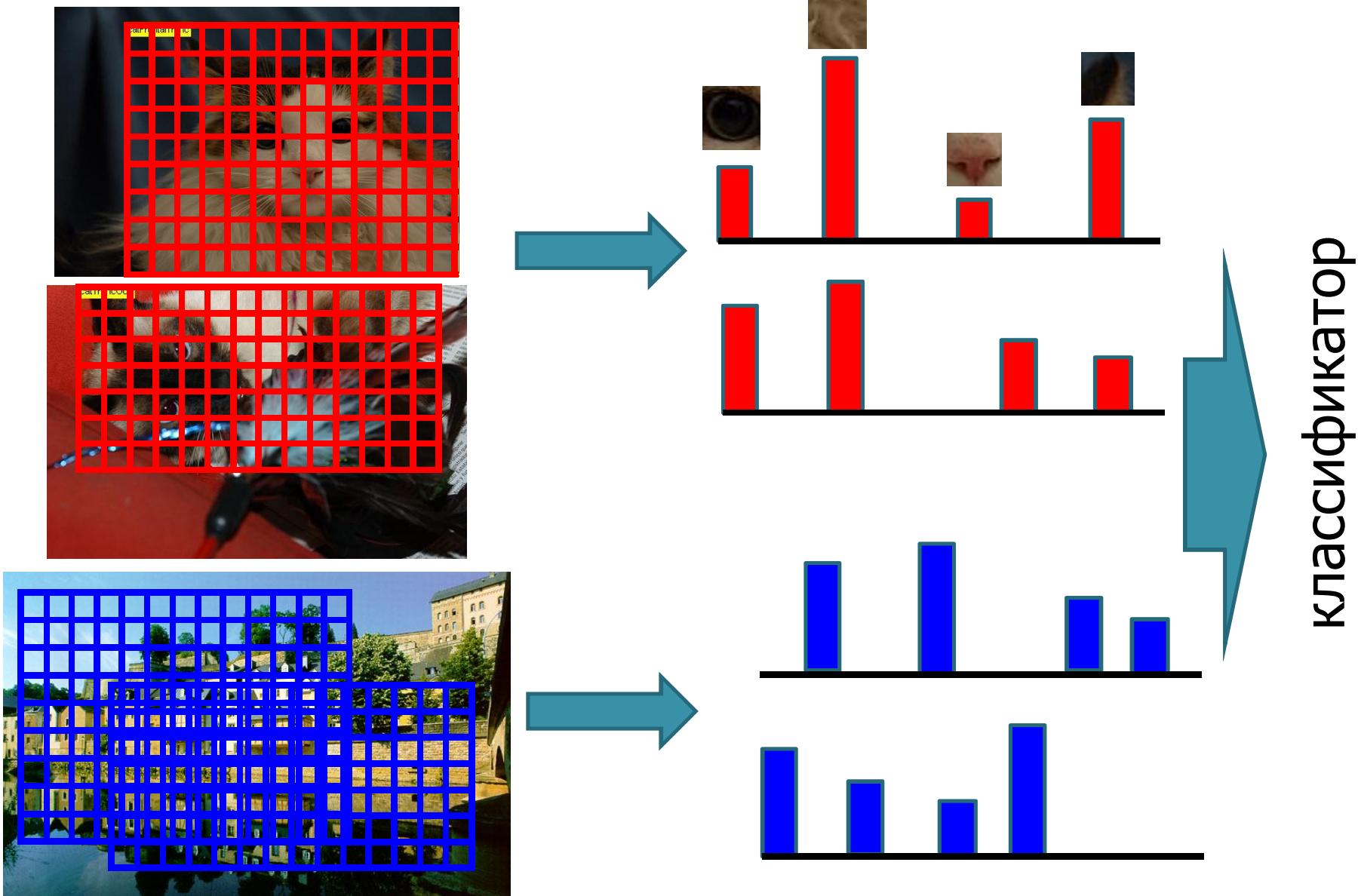
Изменчивость формы



закрытия

изменчивость позы

Изображение как набор «визуальных слов»



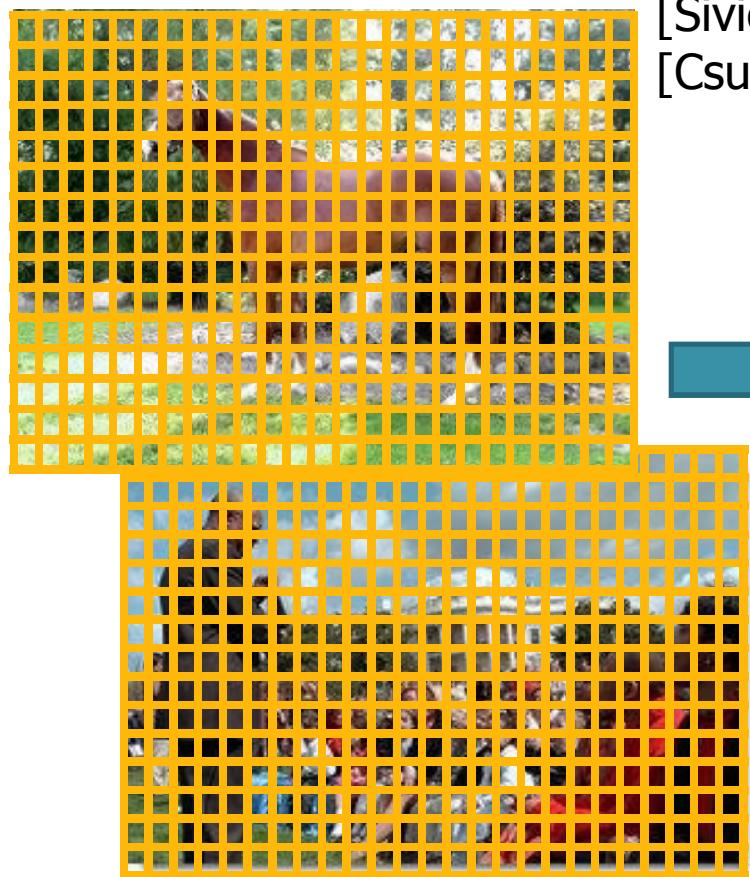
Построение словаря для «наборов слов»

[Sivic & Zisserman, ICCV'03]

[Csurka et al. ECCVw' 04]

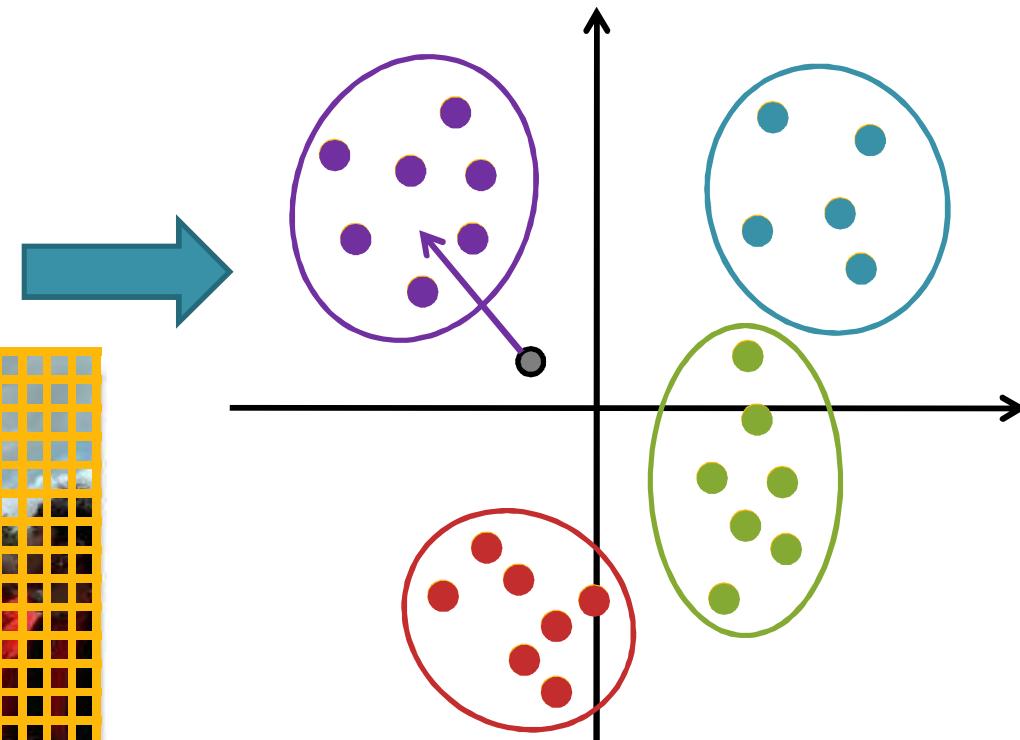


Построение словаря для «наборов слов»



[Sivic & Zisserman, ICCV'03]

[Csurka et al. ECCVw' 04]

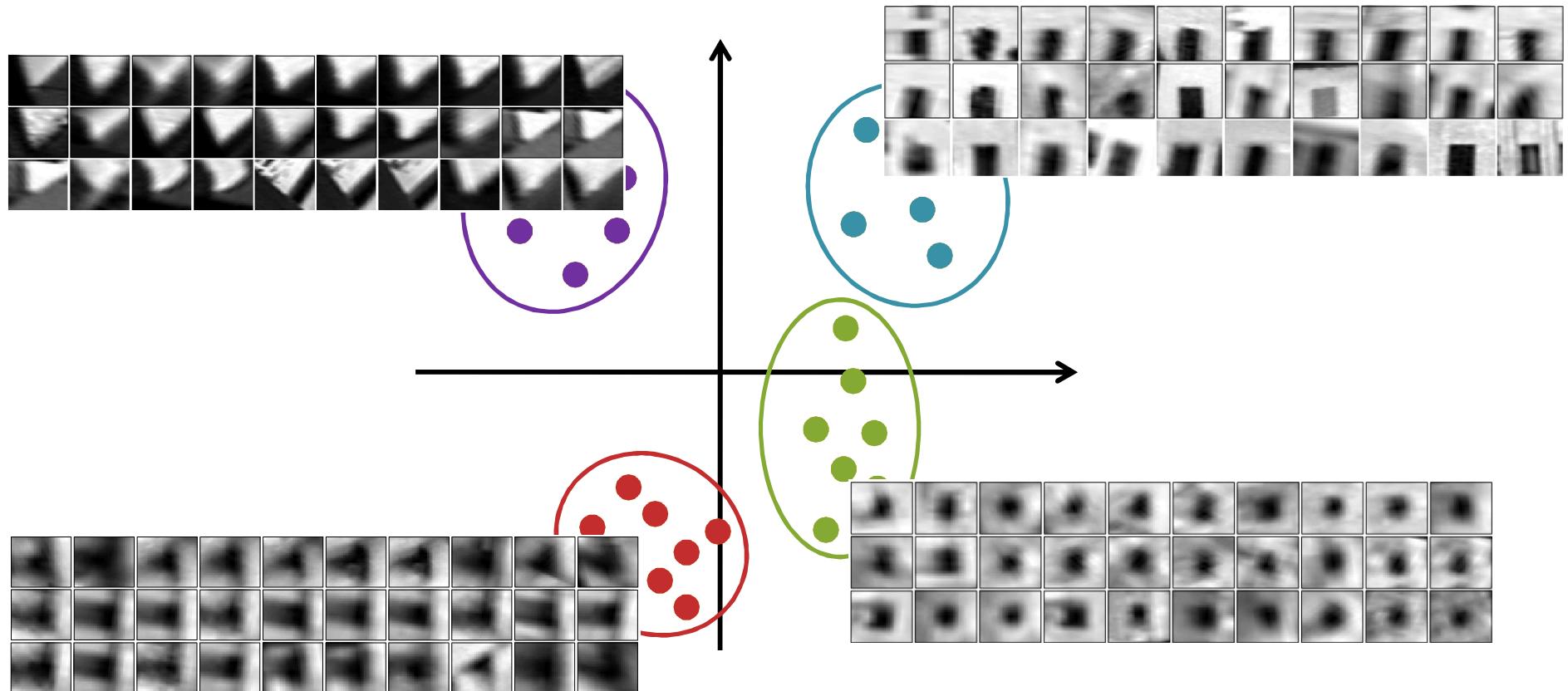


Пространство дескрипторов

- Вместо патчей можно взять контуры или суперпиксели
- Возможный сэмплинг: “интересные точки”, рёбра, случайный, плотный

«Визуальные слова»

Sivic & Zisserman, ICCV 2003



Поиск изображений

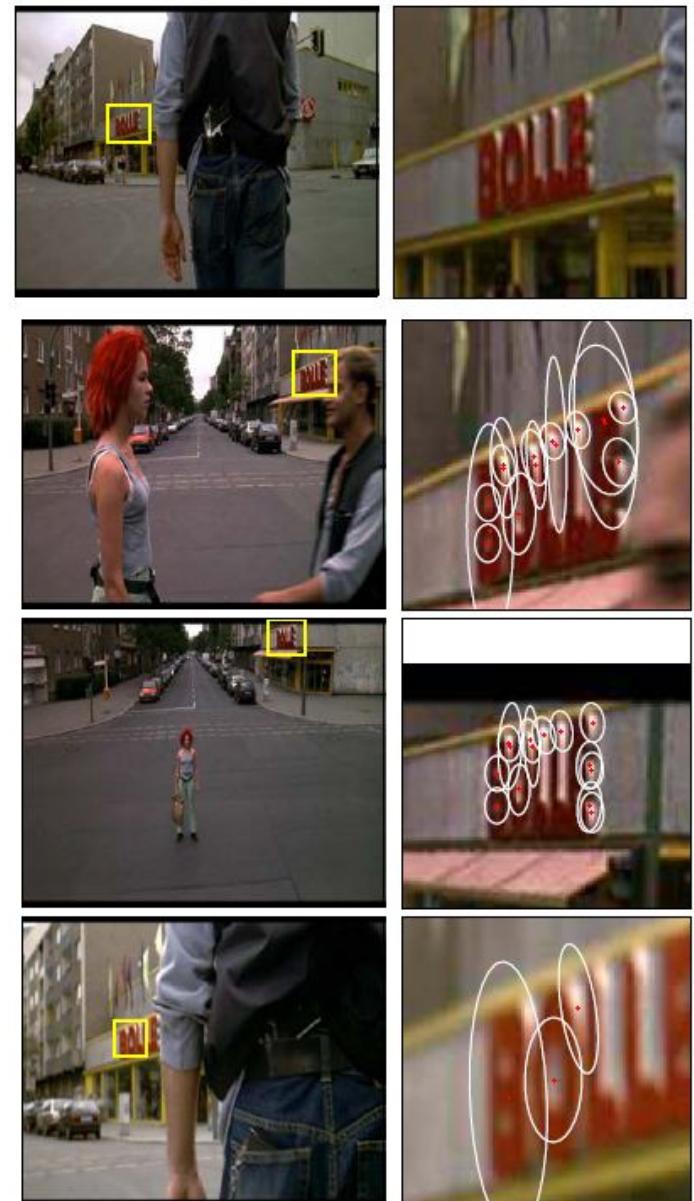
Sivic & Zisserman, "Video Google", ICCV 2003

Демо онлайн:

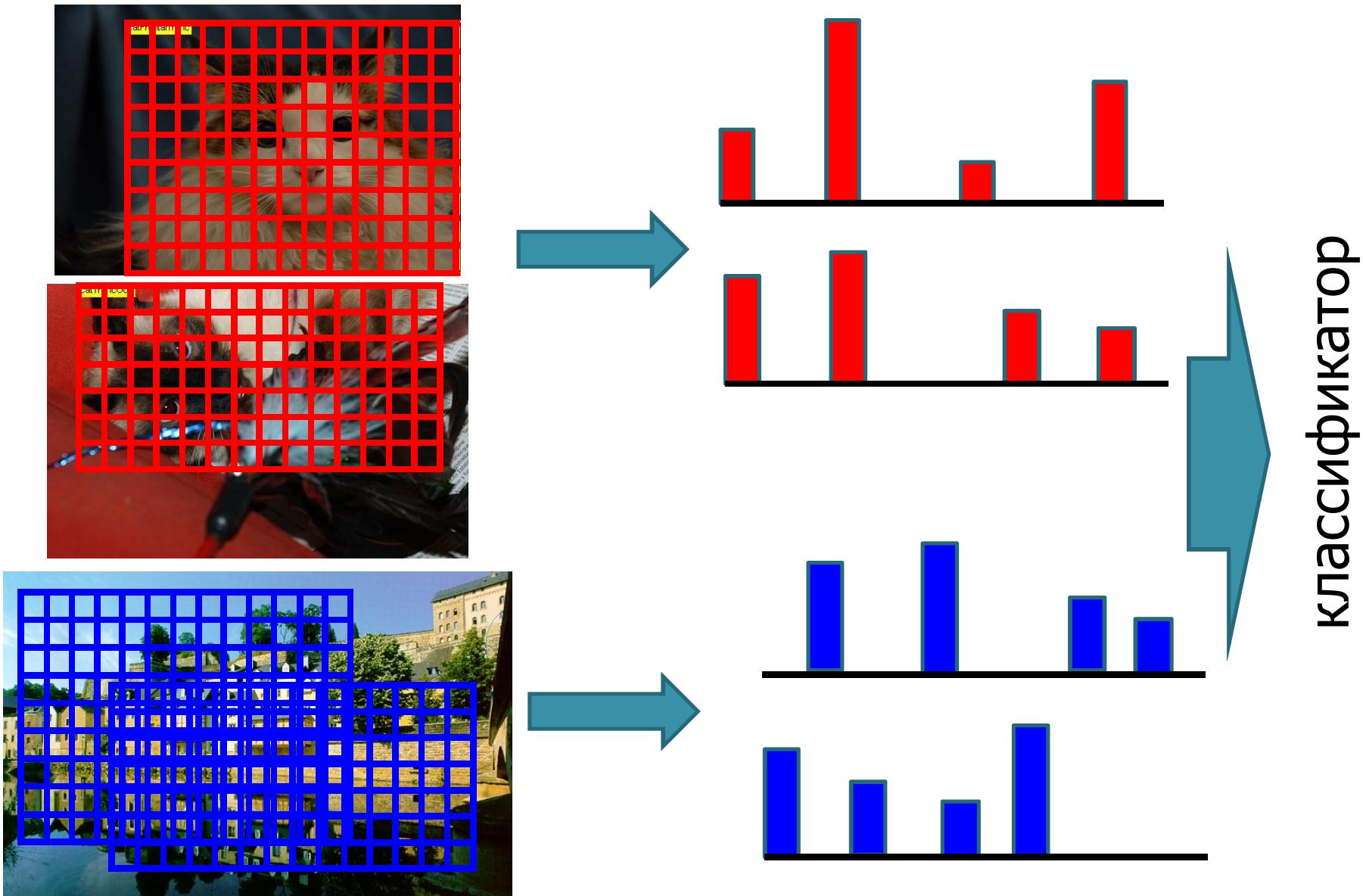
<http://www.robots.ox.ac.uk/~vgg/research/vgoogle/>

Основная идея:

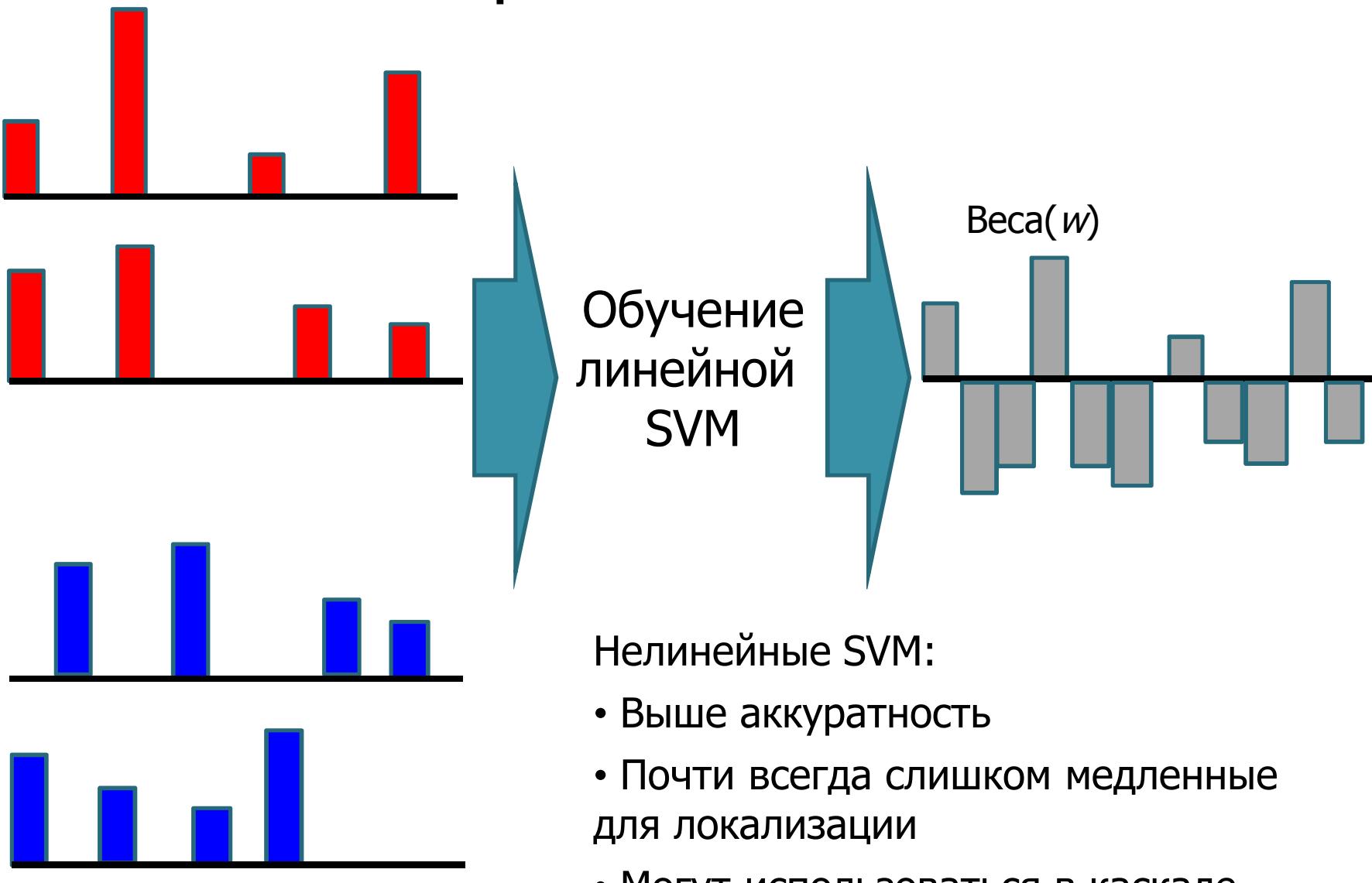
- Использование словаря
- Инвертированные индексы



Изображение как набор визуальных слов



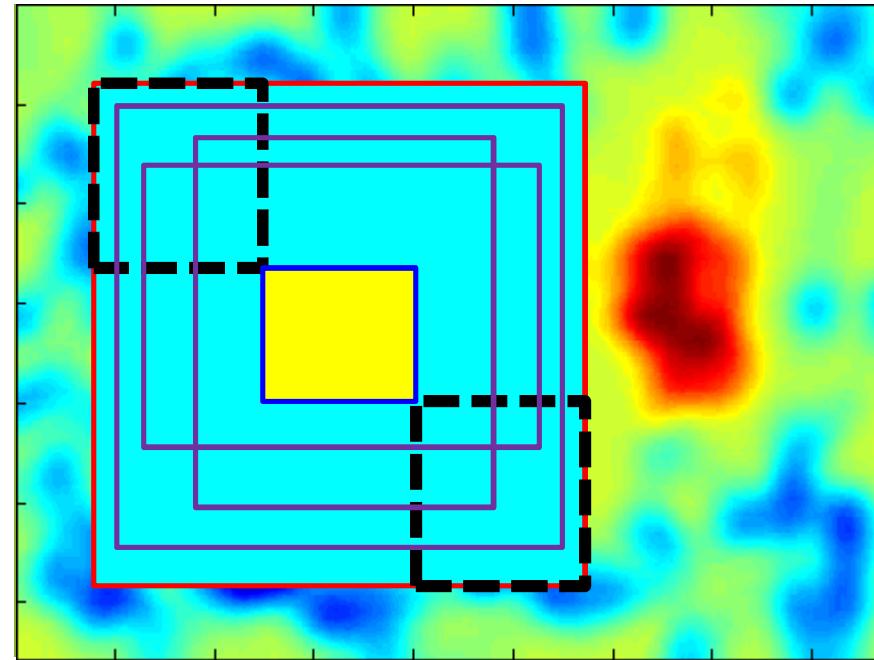
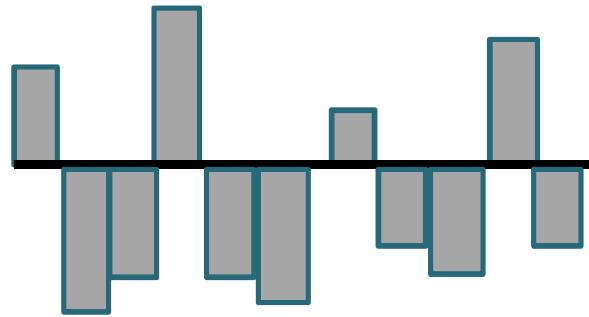
Наборы слов и SVM



Локализация, SVM и Branch-and-Bound

[Lampert , Blaschko, Hofmann CVPR'08]

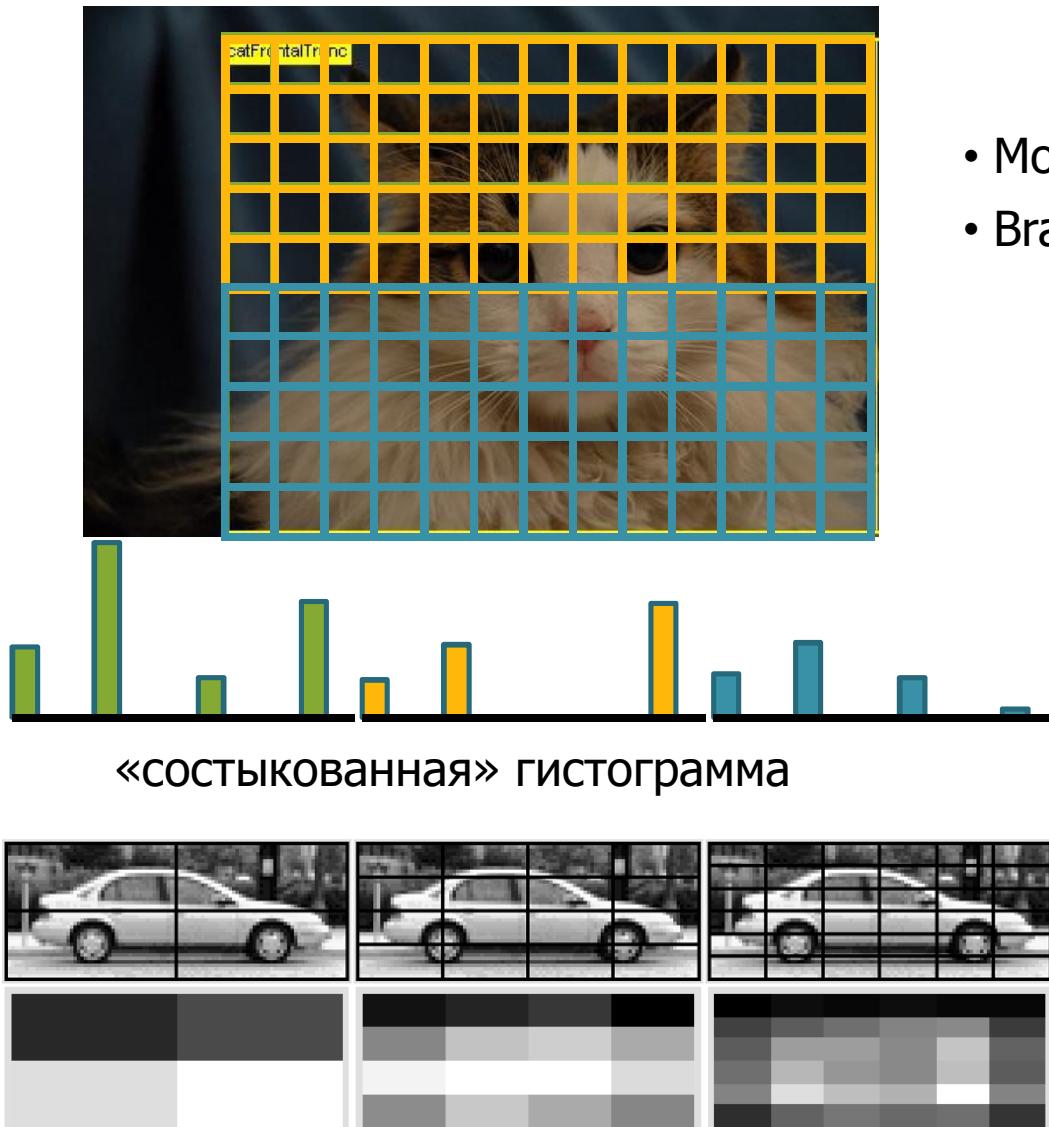
Веса SVM (w)



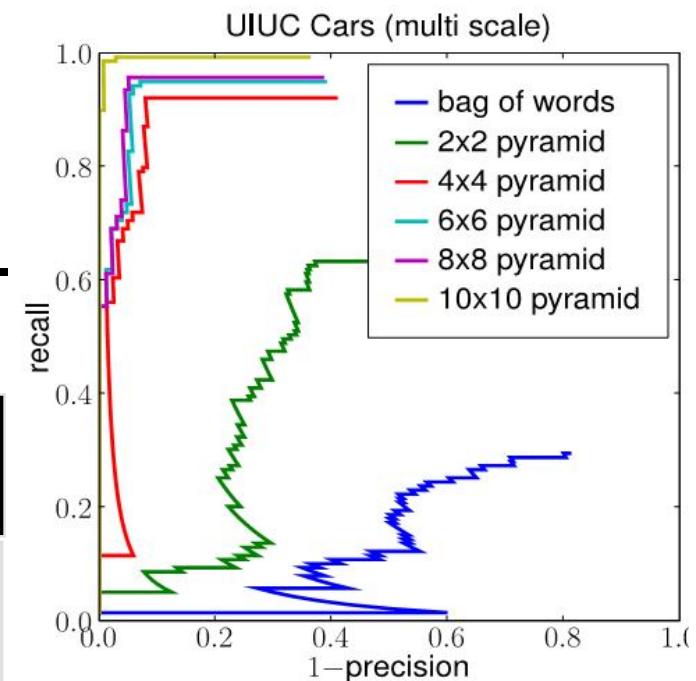
$$(w \cdot h(\square)) \leq (w^+ \cdot h(\square)) + (w^- \cdot h(\square))$$

Возвращение геометрии

[Lazebnik, Schmid & Ponce, CVPR 2006]



- Можно взять несколько разбиений
- Branch-and-Bound обобщается



Скользящее окно: итоги



Общий подход:

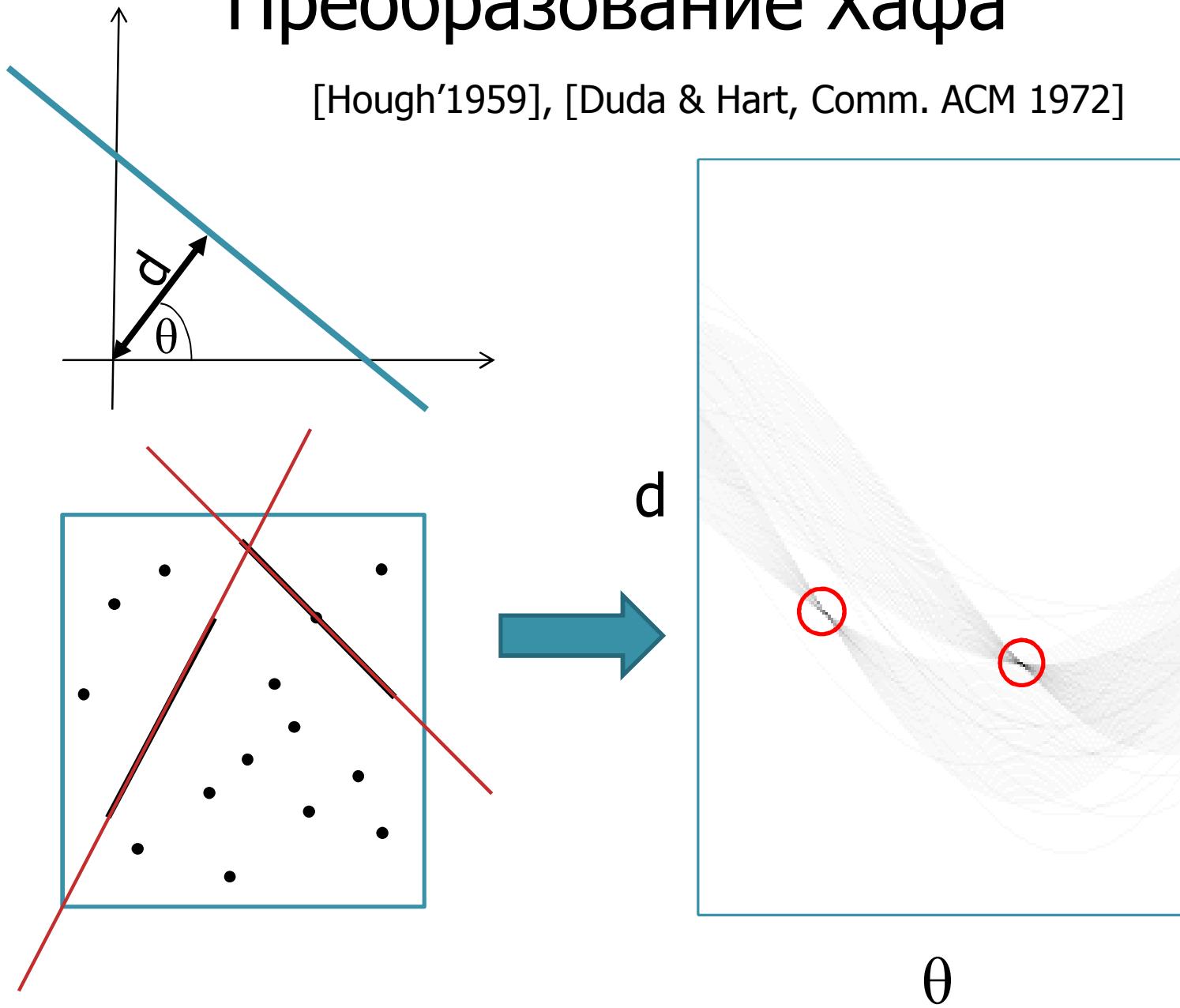
- Найти представление изображений в окне
- Натренировать бинарный классификатор

+ Бинарная классификация

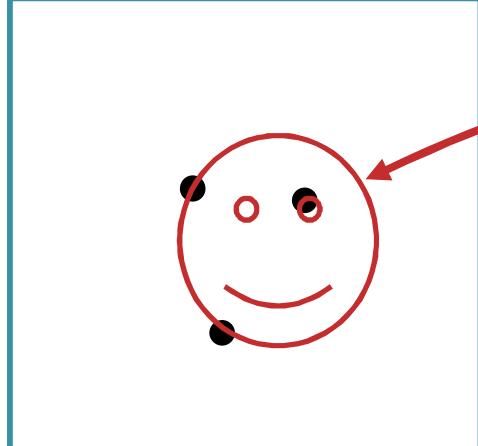
- Ограниченная обобщающая способность

Преобразование Хафа

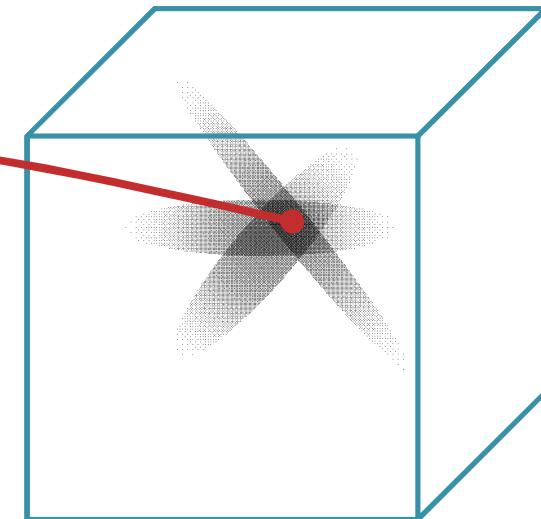
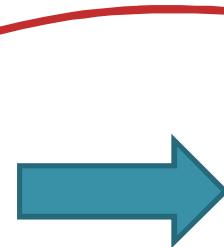
[Hough'1959], [Duda & Hart, Comm. ACM 1972]



Обобщенное преобразование Хафа

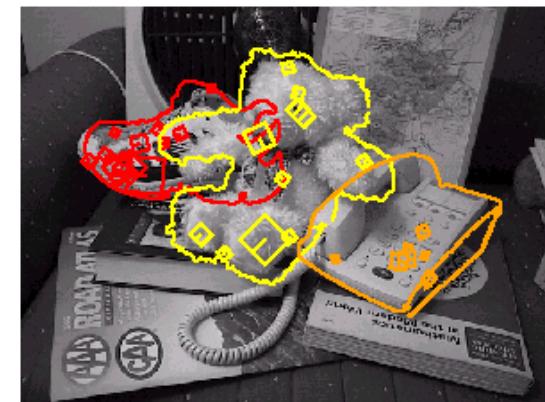


Пространство изображения



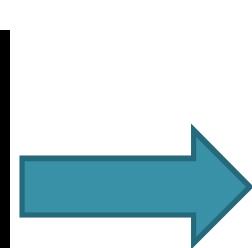
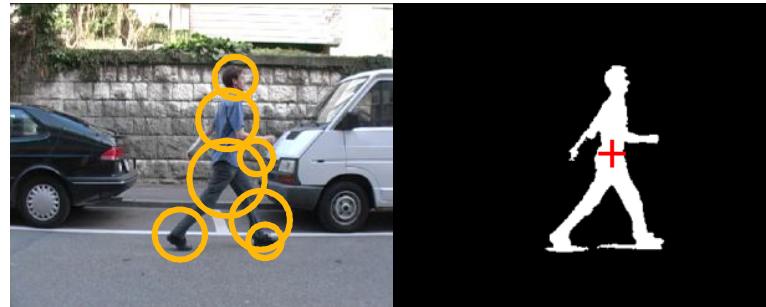
Пространство гипотез
(Хаф-пространство)

- Геометрические фигуры [Ballard PR'81]
- Конкретные объекты [Lowe ICCV'99]
- Категории объектов [Leibe & Schiele DAGM'04]

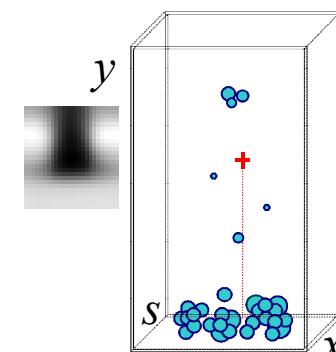
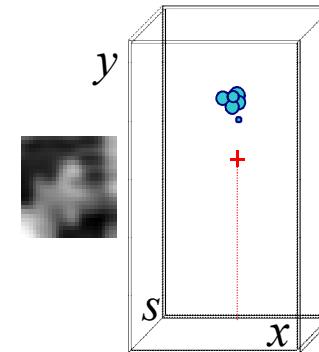
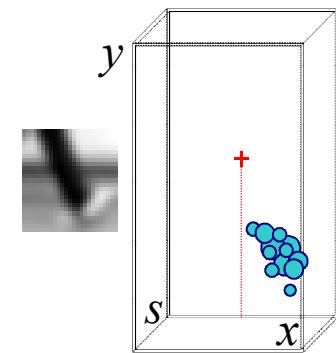
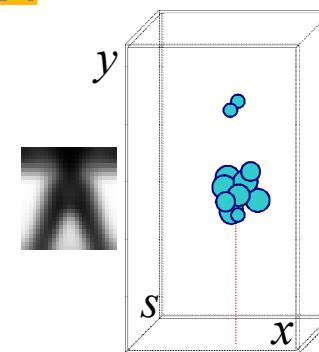


Неявная модель формы

[Leibe, Leonardis & Schiele IJCV'08]



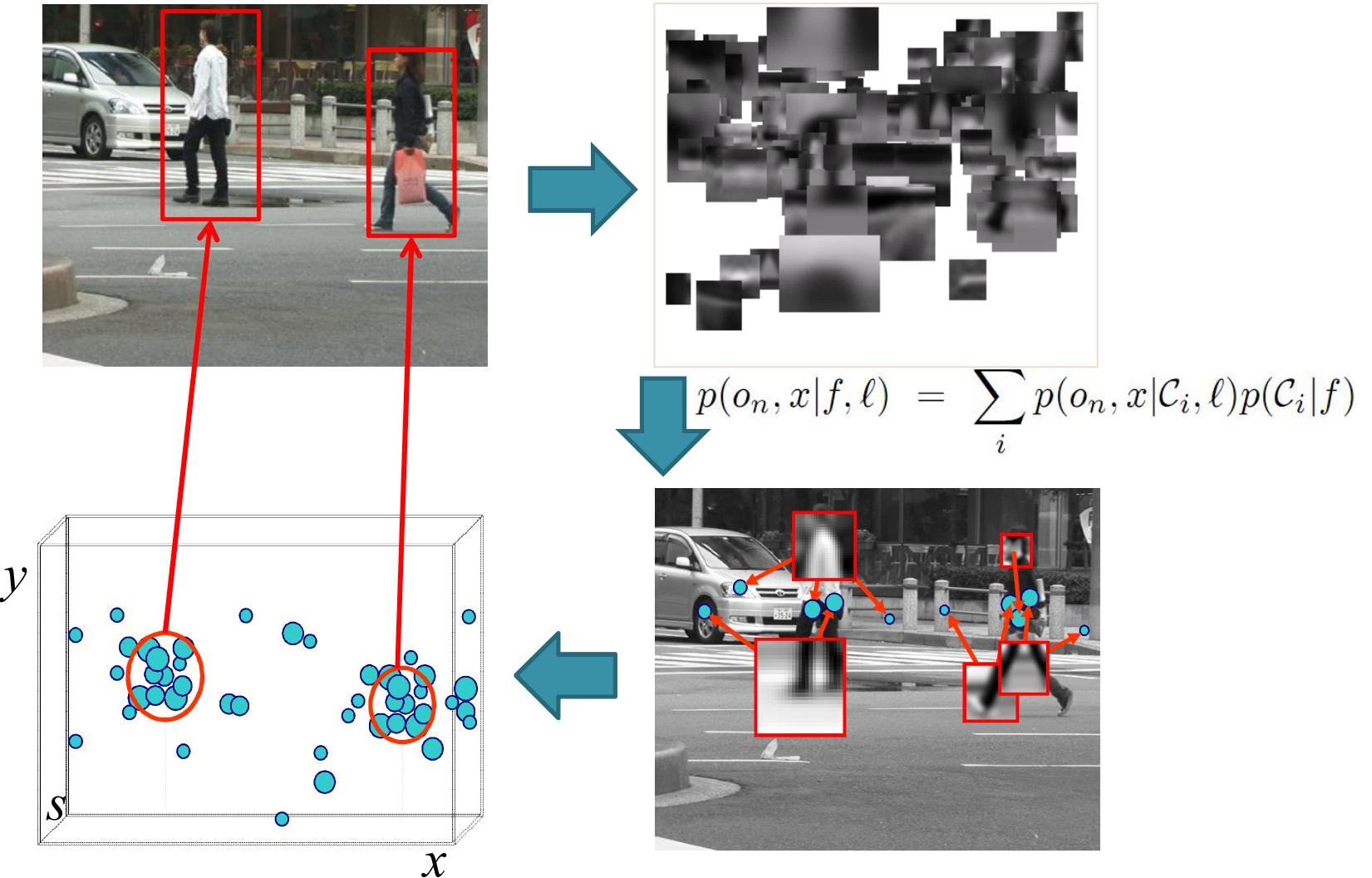
Каждое слово имеет Хаф голос



Slide based on Leibe's presentation

Неявная модель формы

[Leibe, Leonardis & Schiele IJCV'08]



Попиксельная сегментация позволяет исправить часть ошибок

Slide based on Leibe's presentation

Локализация объектов и Хаф-леса

Juergen Gall & Victor Lempitsky,
Class-Specific Hough Forests for Object Detection, CVPR 2009

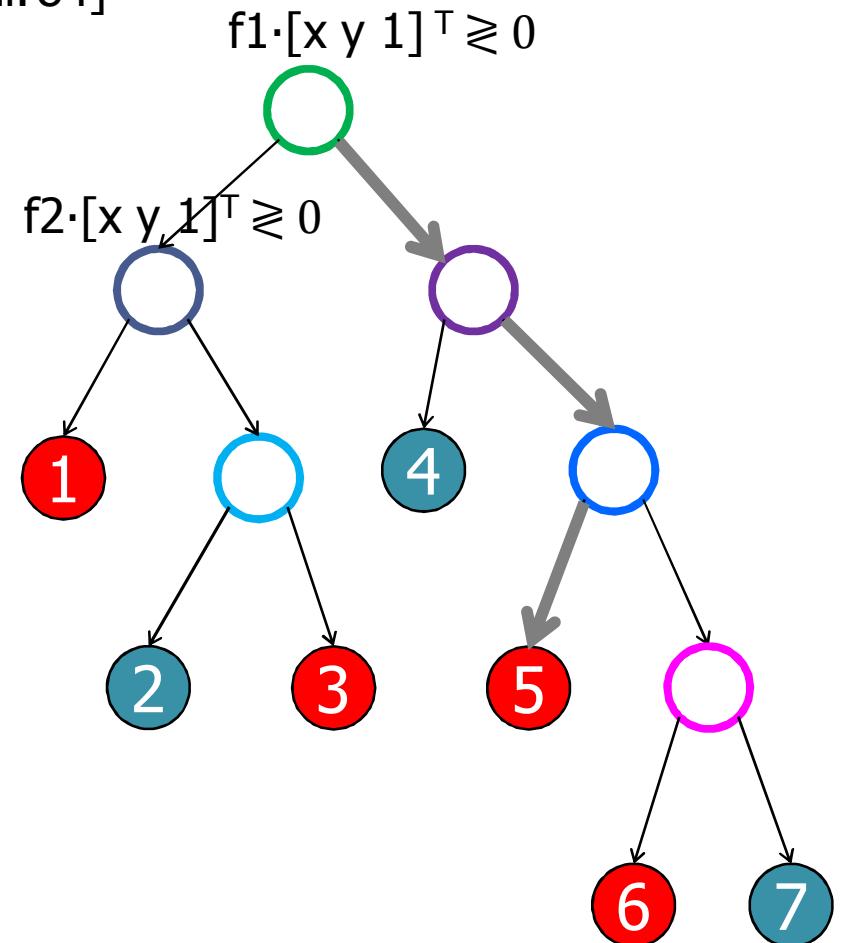
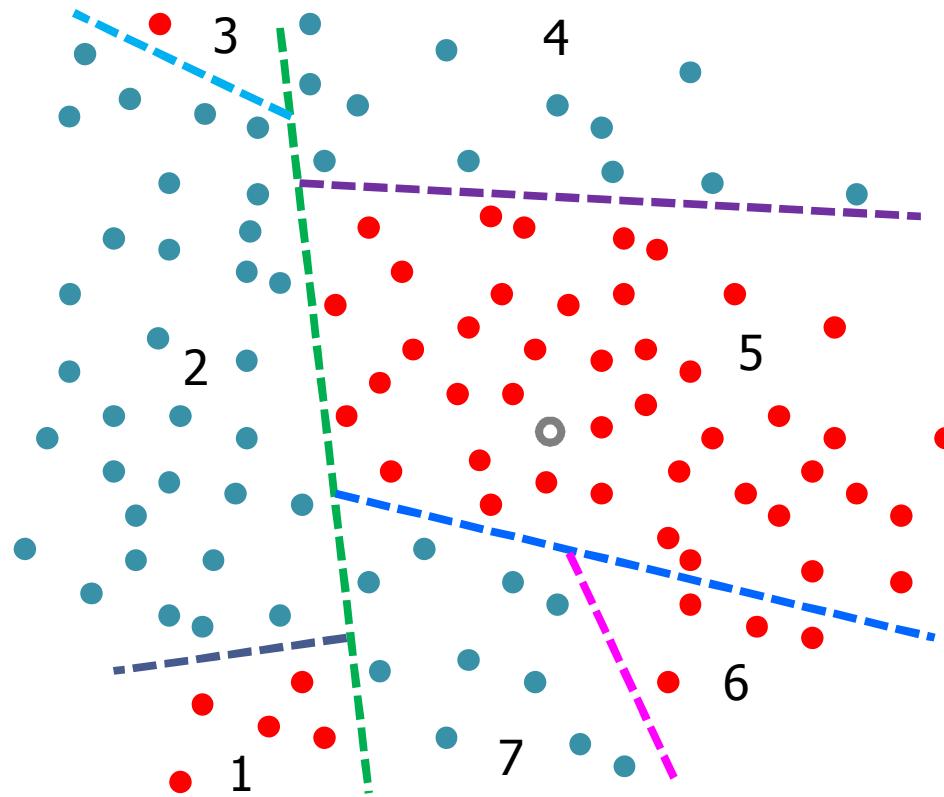
Совместная работа с Jürgen Gall



Jürgen (currently at ETHZ)

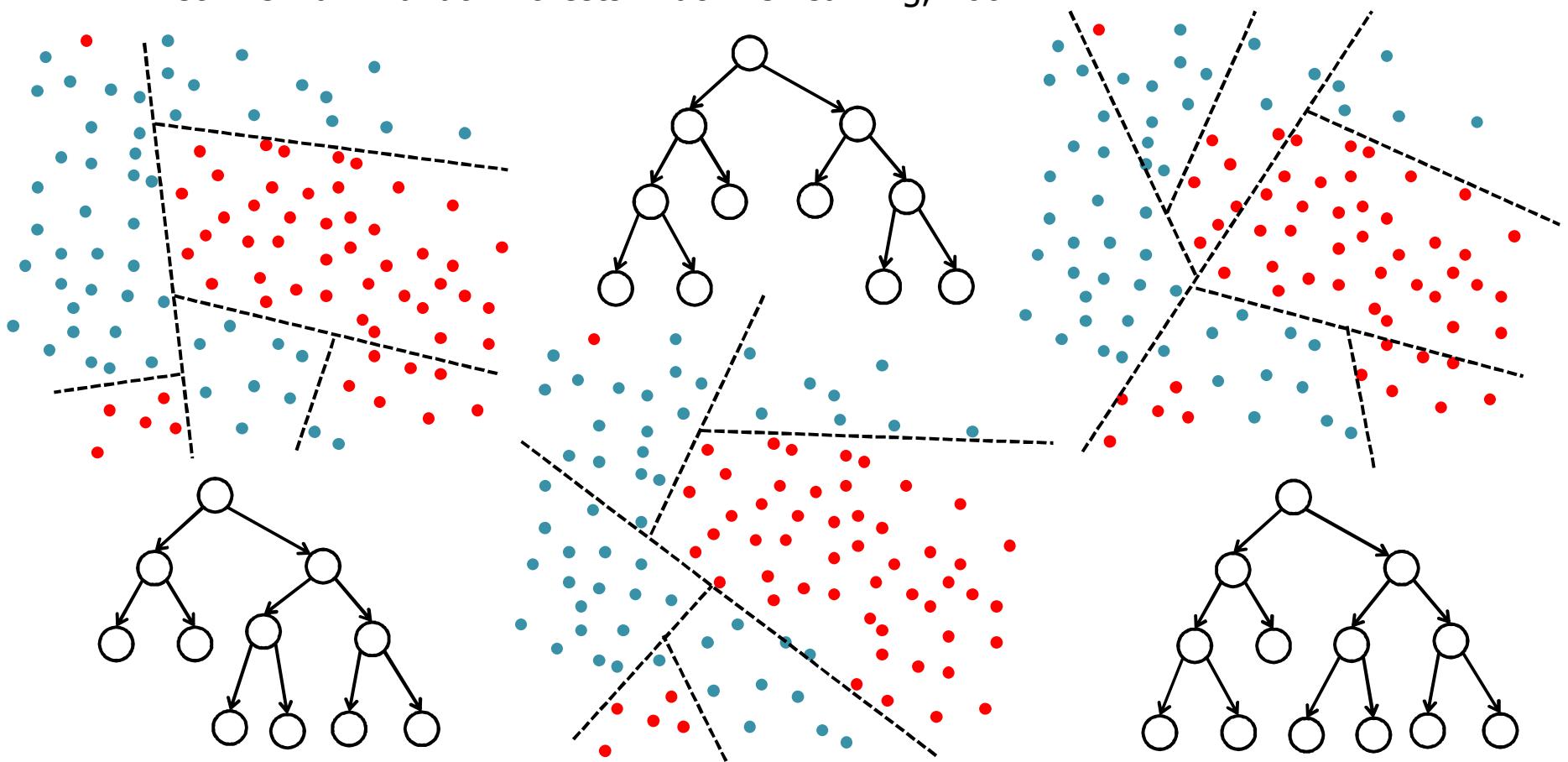
Деревья решений

[Breiman et al.'84]



Случайные леса

1. Yali Amit, Donald Geman: *Shape quantization and recognition with randomized trees.* Neural Computation, 1997.
2. Leo Breiman: *Random forests.* Machine Learning, 2001.

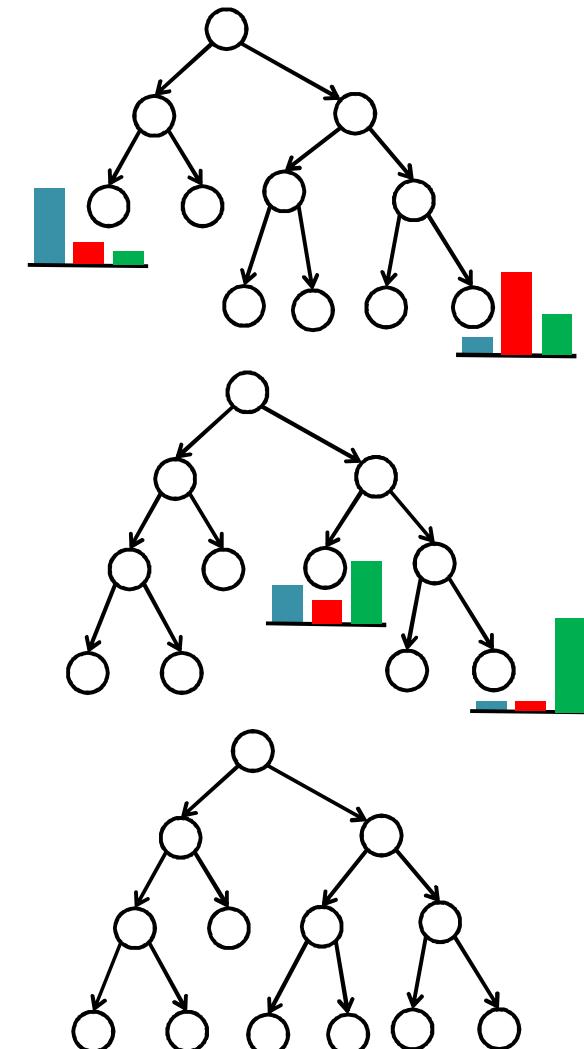


Случайные леса - обучение

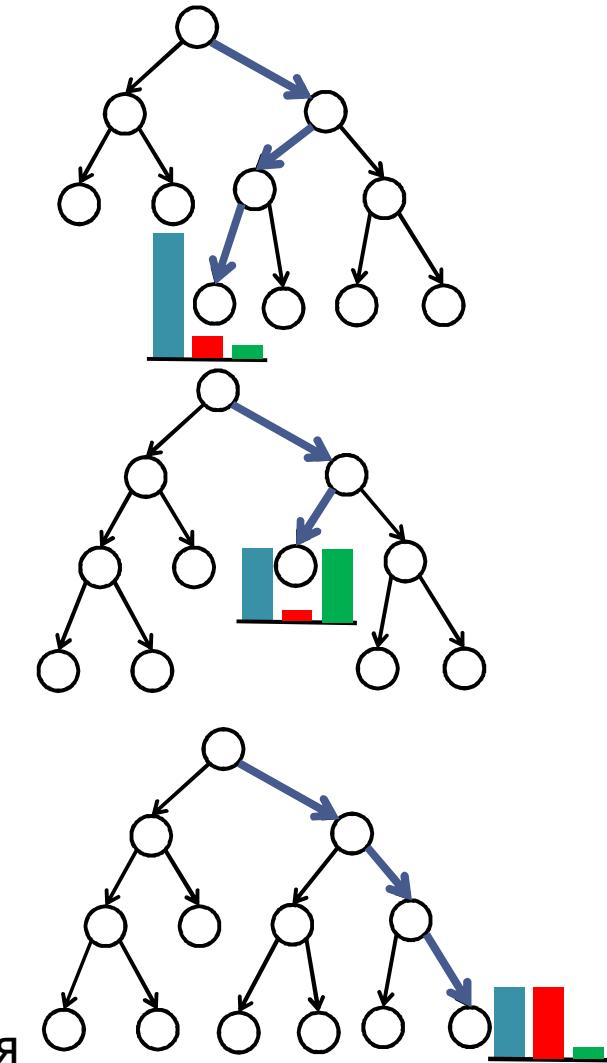
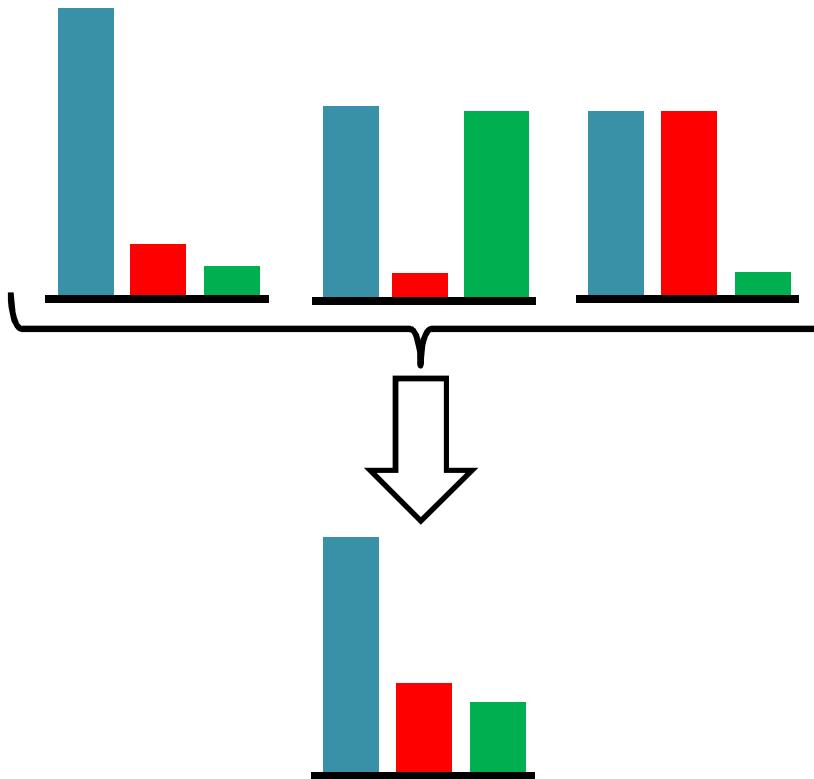
[Breiman JMLR'01]

```
function Node = Обучение_Вершины( {(x,y)}, Level) {
    if {y} одинаковые или Level == maxLevel
        return Создать_Лист(Распределение y);
    {tests} = Создать_N_Случайных_Тестов({(x,y)},N);
    test = Выбрать_лучшее_разбиение_из({tests});
    {(x0,y0)} = {(x,y) | test(x) = 0};
    {(x1,y1)} = {(x,y) | test(x) = 1};
    LeftChild = Обучение_Вершины( {(x0,y0)}, Level+1);
    RightChild = Обучение_Вершины( {(x1,y1)}, Level+1);
    return Создать_Вершину(test, LeftChild, RightChild);
}

//Обучение леса
function main() {
    {X,Y} = Прочитать_Обучающие_Данные();
    for i = 1 to N
        {Xi,Yi} = Случайное_Подмножество({X,Y});
        TreeRoot_i = Обучение_Вершины({Xi,Yi});
    end
}
```

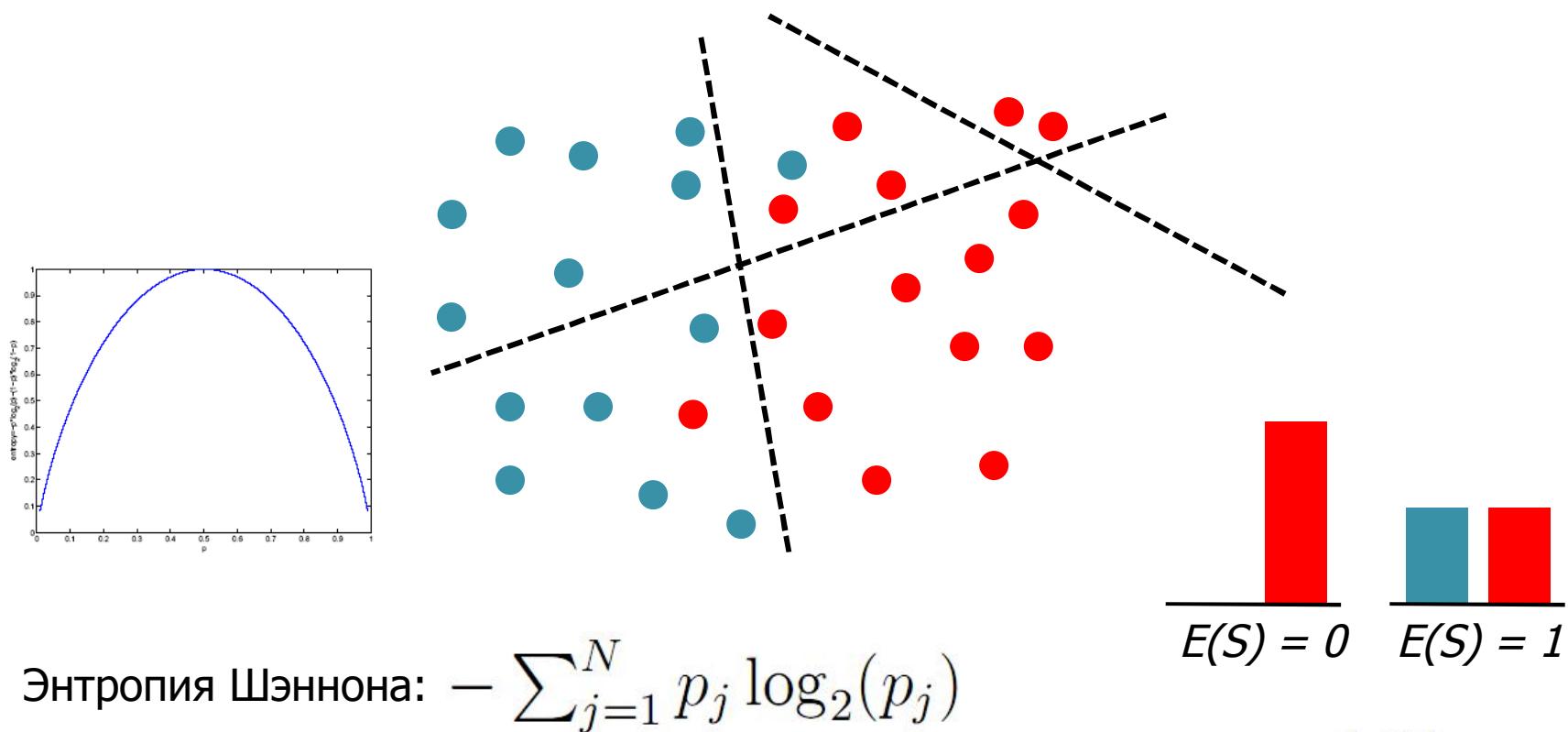


Случайные леса - применение



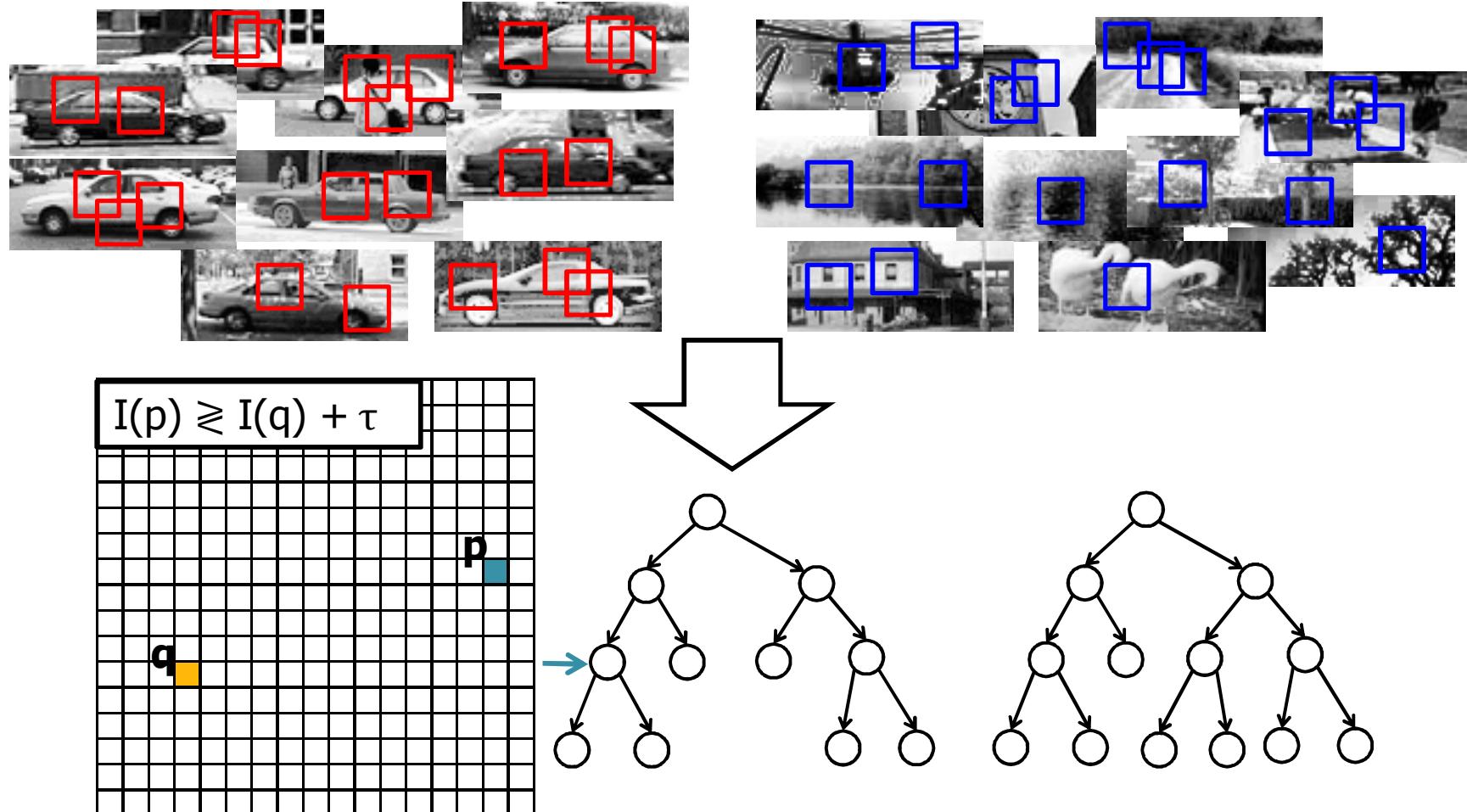
- Один из самых точных методов классификации
- Один из самых гибких методов машинного обучения
- Один из самых быстрых методов обучения

Качество разбиений: энтропия



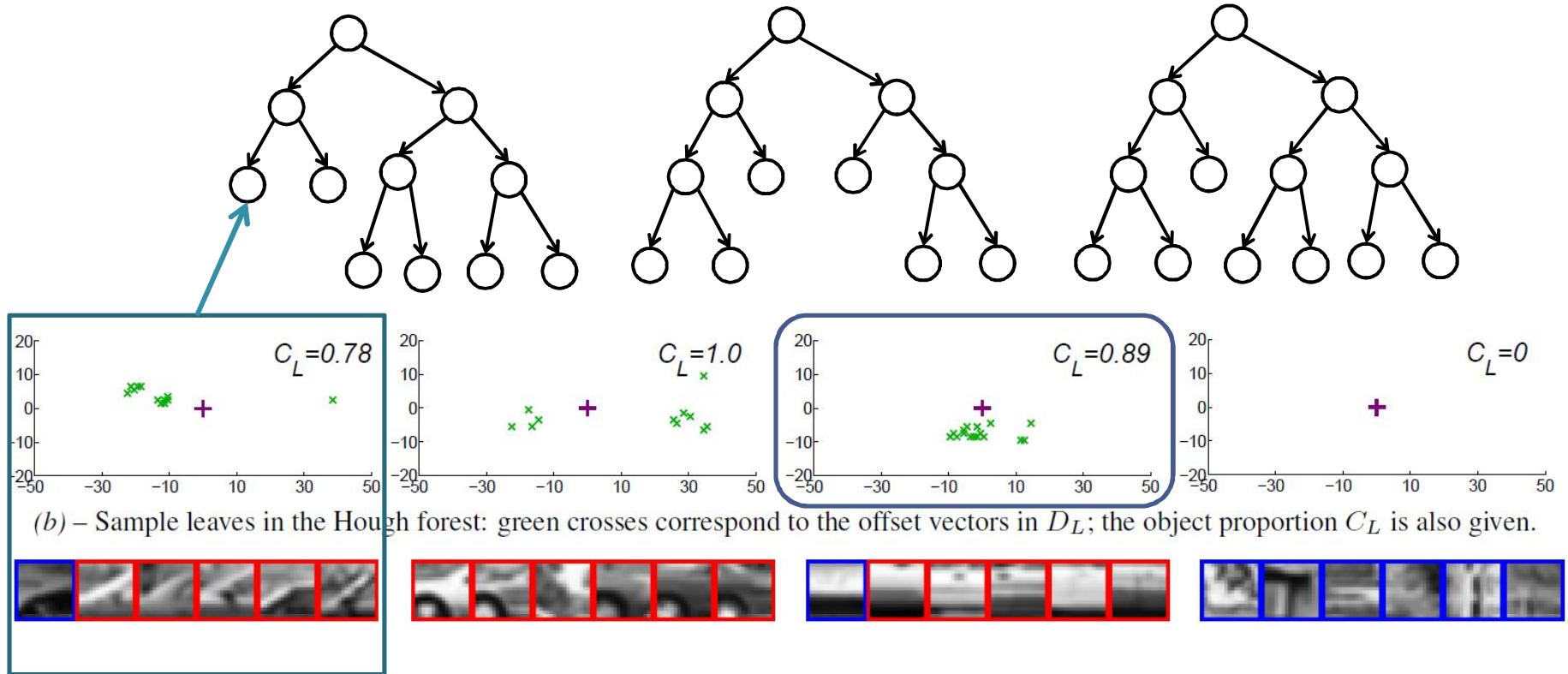
Информационный выигрыш (information gain) $\Delta E = -\sum_i \frac{|S_i|}{|S|} E(S_i)$

Хаф-леса: построение словаря



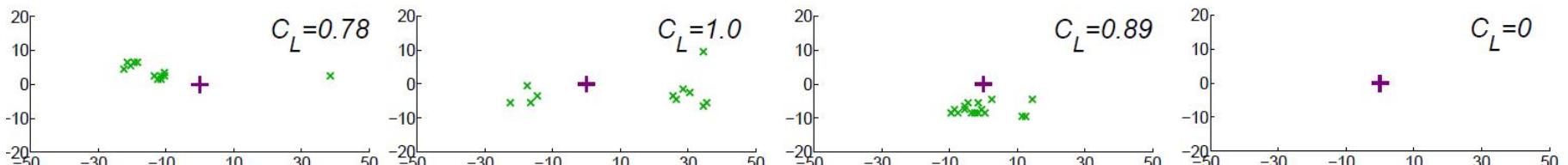
- 30 различных каналов (интенсивность, производные, HOG-like)...
- Аналогичные словари в [Moosmann et al. NIPS' 06], [Winn & Shotton CVPR' 06]

Хаф-леса: построение словаря

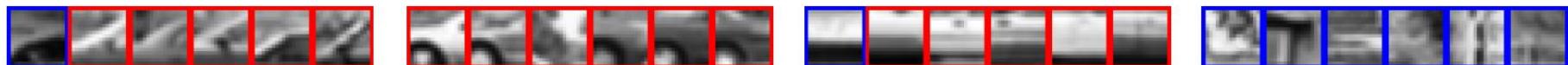


На выходе: вероятность принадлежности объекту
+ распределение положений относительно центра

Хаф-леса: построение словаря



(b) – Sample leaves in the Hough forest: green crosses correspond to the offset vectors in D_L ; the object proportion C_L is also given.



На выходе: вероятность принадлежности объекту
+ распределение положений относительно центра

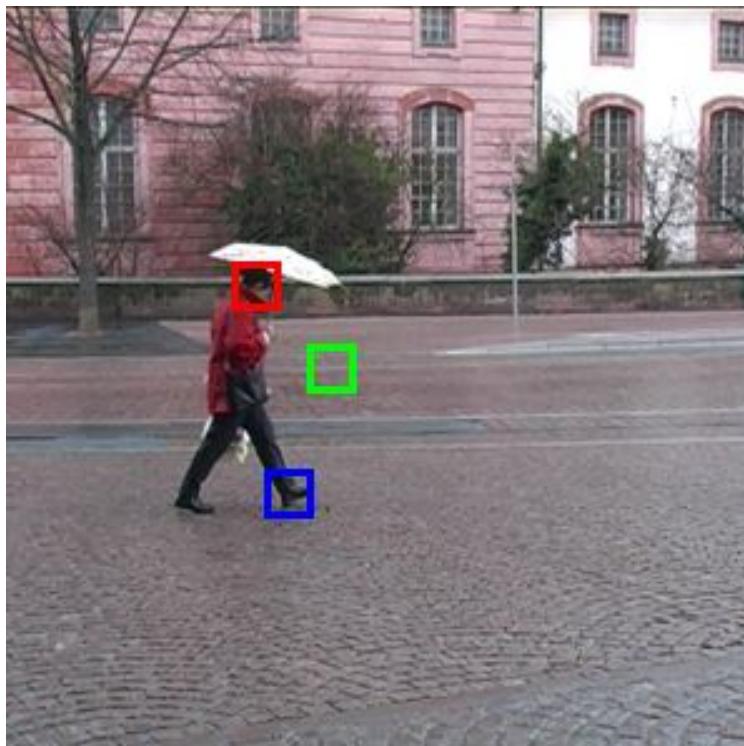
Две меры упорядоченности набора фрагментов:

$$U_1(A) = |A| \cdot \text{Entropy}(\{c_i\}) \quad U_2(A) = \sum_{i:c_i=1} (\mathbf{d}_i - \mathbf{d}_A)^2$$

Выбор разбиения принимает во внимание одну из двух мер упорядоченности

Хаф-леса: пример работы

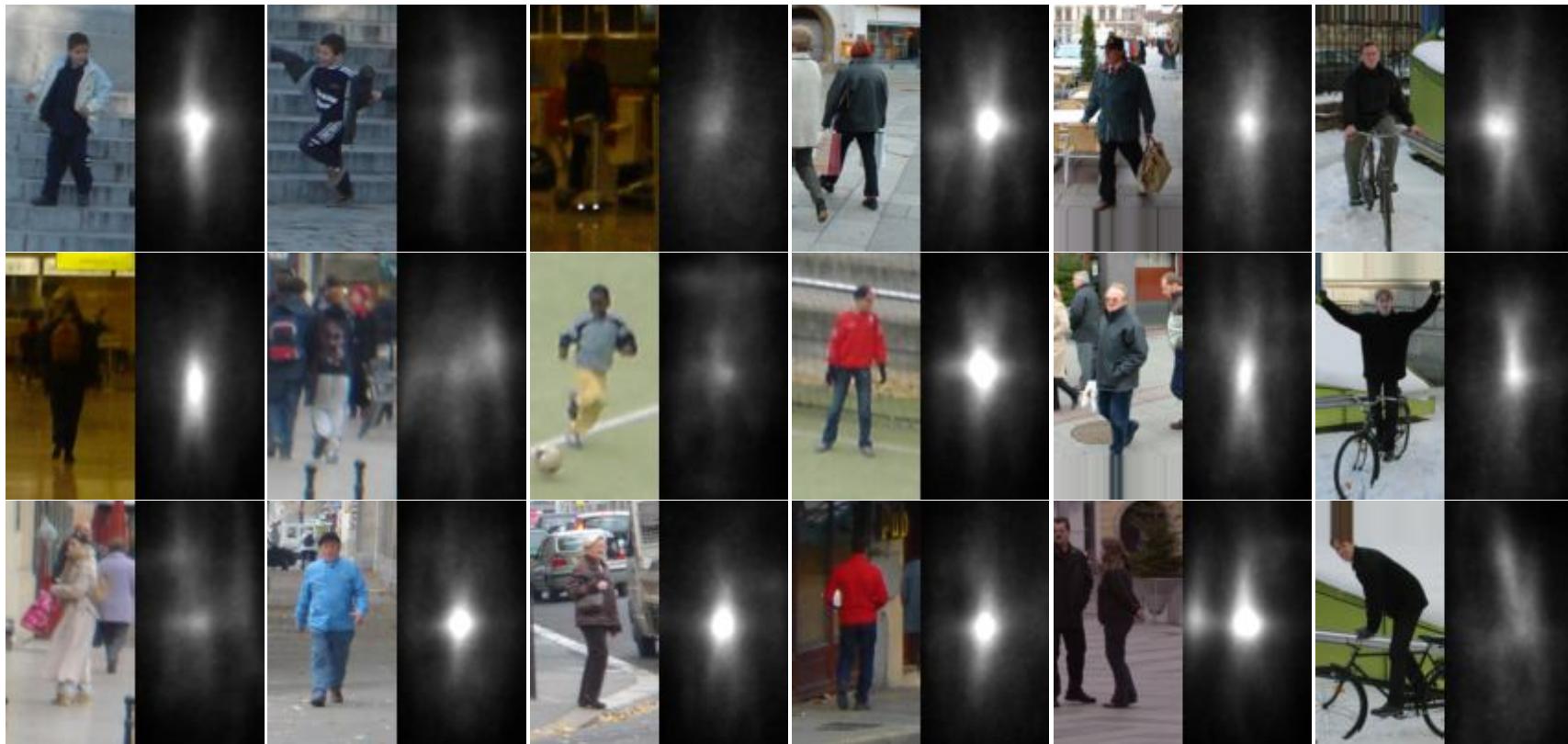
$$p(E(\mathbf{x}) | \mathcal{I}(\mathbf{y}); \mathcal{T}) = \\ \left[\frac{1}{|D_L|} \sum_{d \in D_L} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|(\mathbf{y} - \mathbf{x}) - d\|^2}{2\sigma^2}\right) \right] \cdot C_L$$



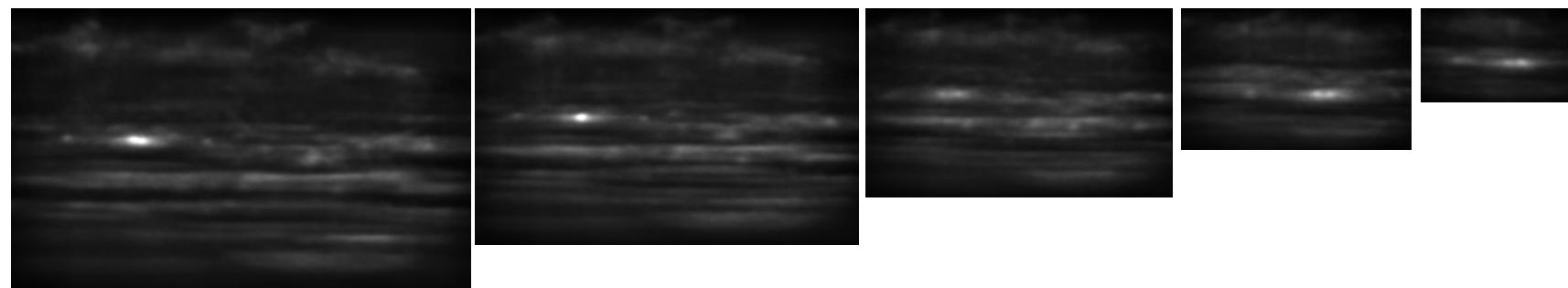
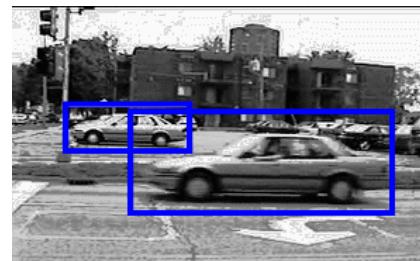
Хаф-леса: результаты



Хаф-леса: результаты



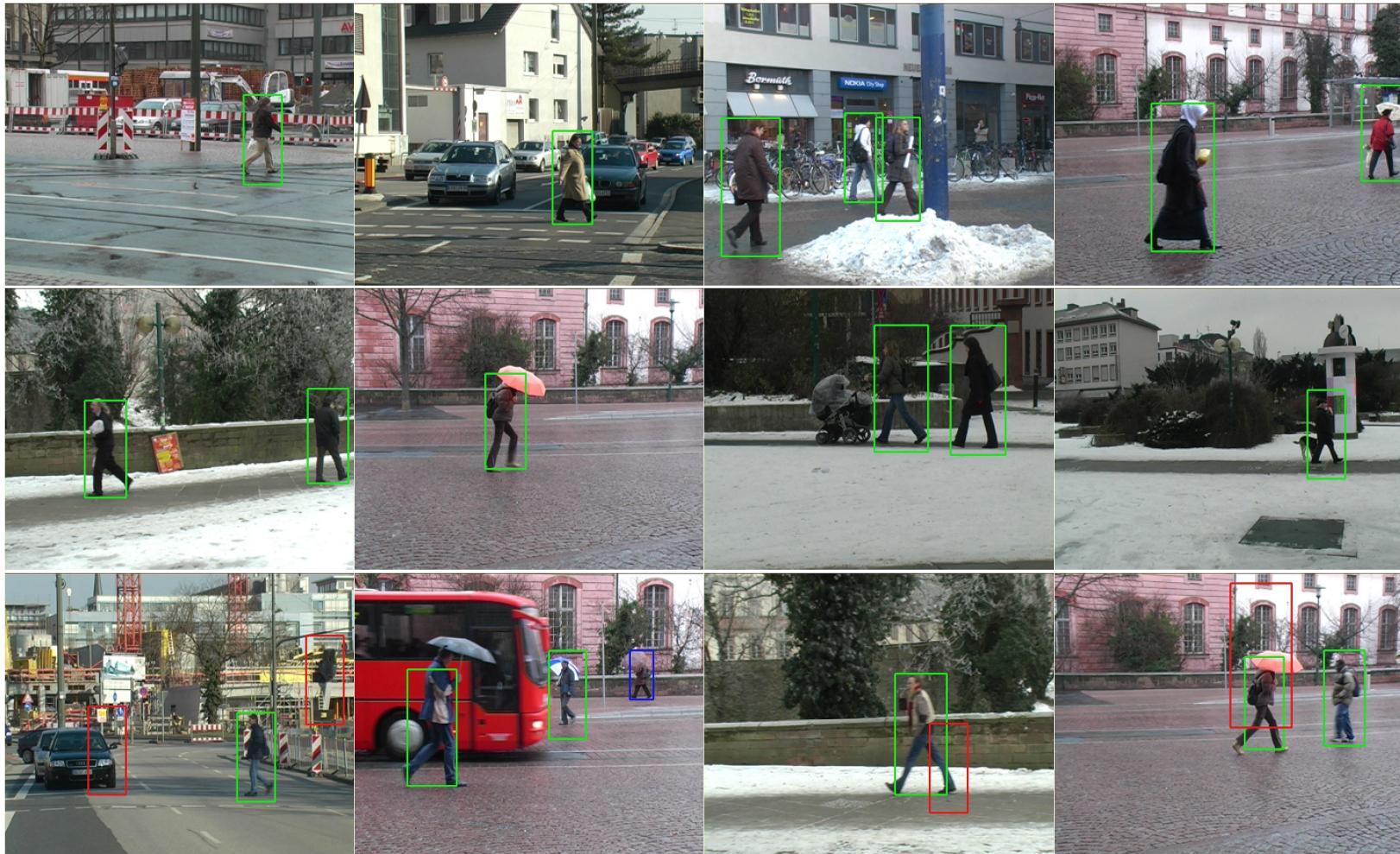
Хаф-леса: разные размеры



Хаф-леса: результаты

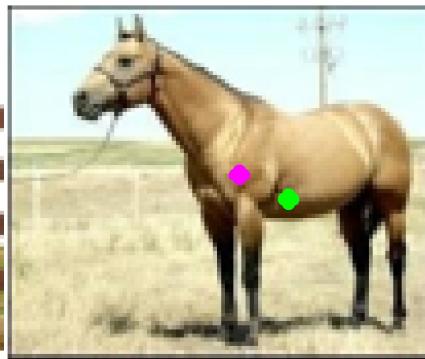


Хаф-леса: результаты

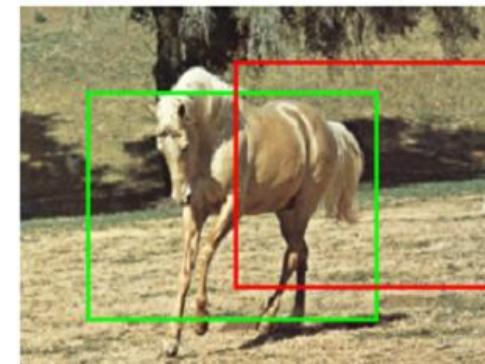
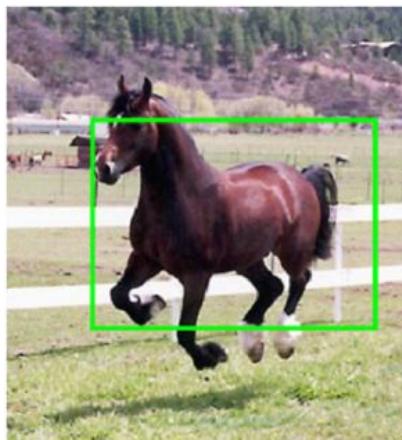


Хаф-леса: результаты

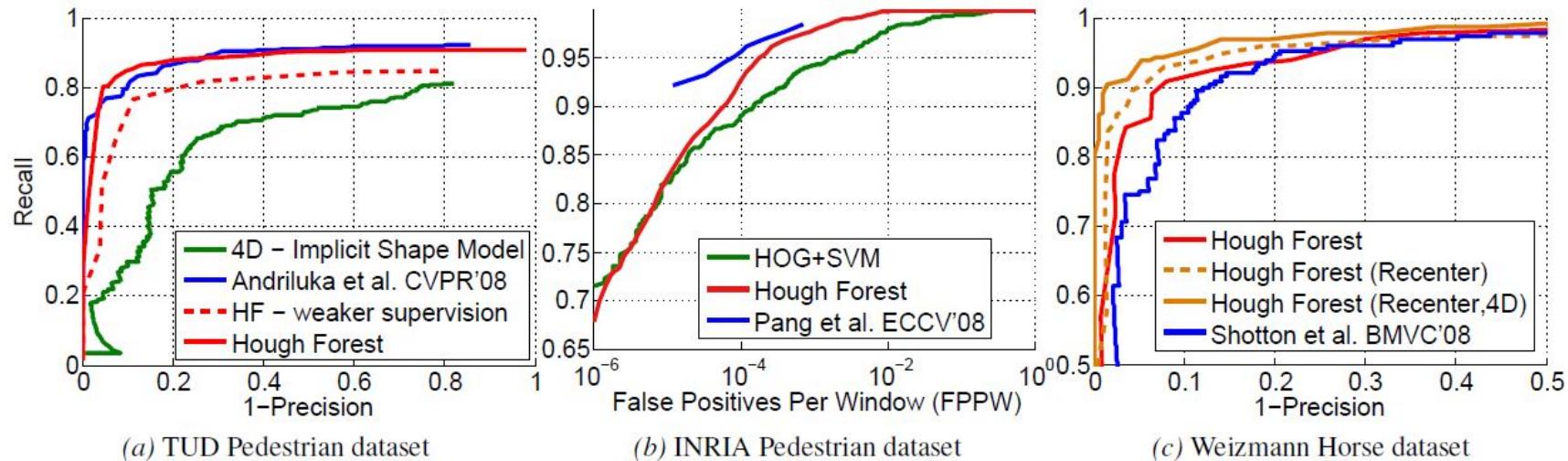
Обучение:



Обнаружение:



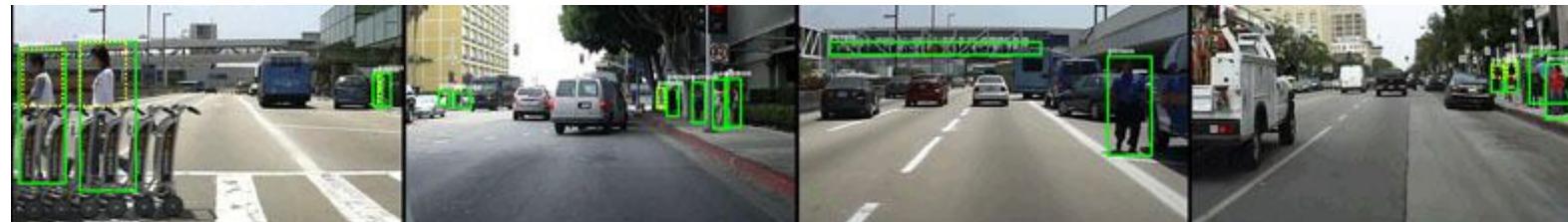
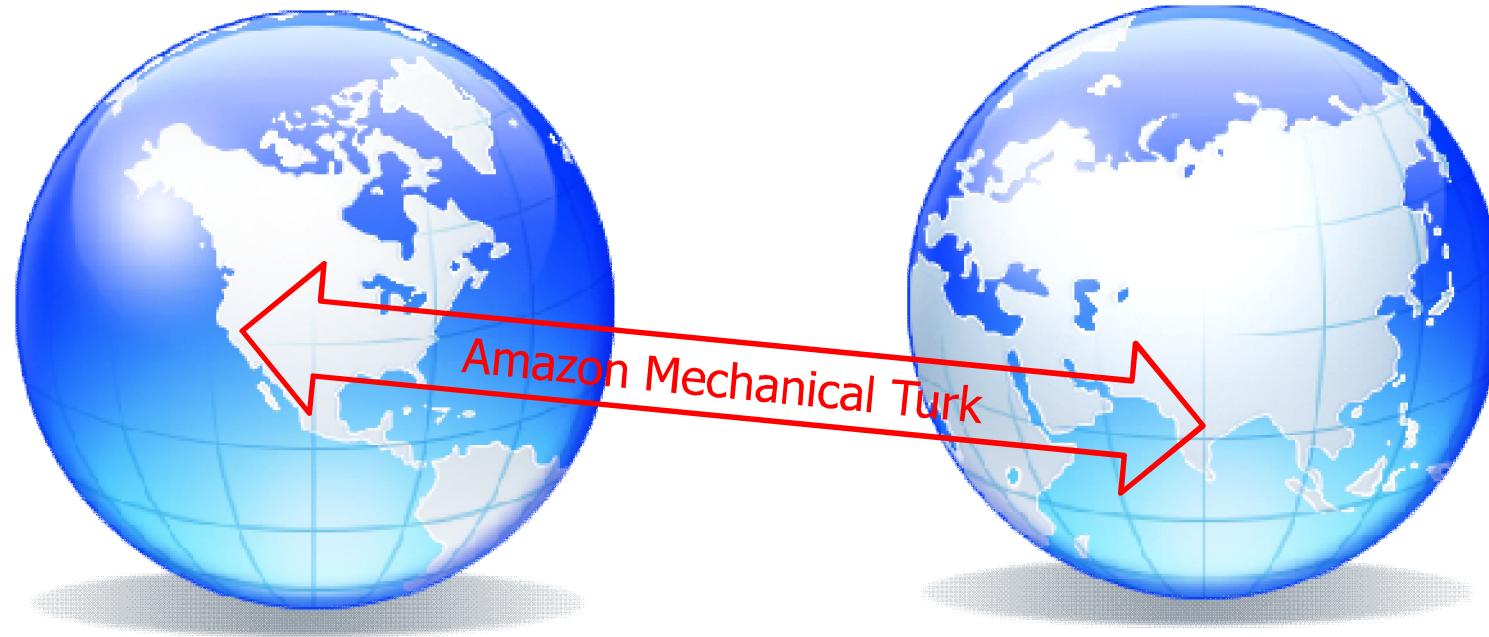
Хаф-леса: итоги



- + Аккуратный метод локализации
- + Достаточно быстрый метод локализации (секунды)
- + Не требует разметки на пикельном уровне
- Тестирование проводилось на наборах данных ограниченной сложности

Локализация объектов: открытые проблемы

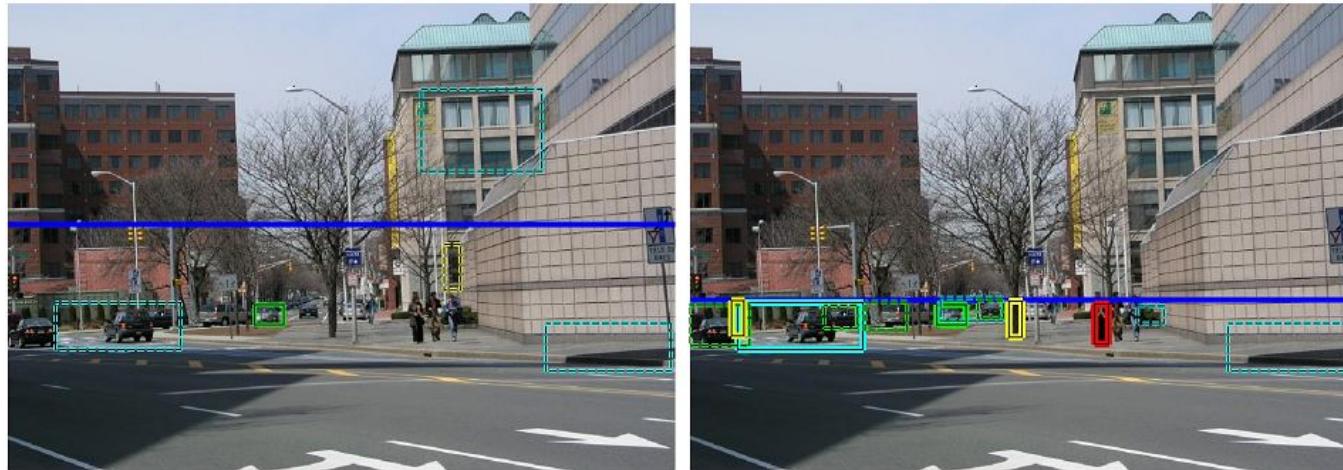
Обучение на больших выборках (100.000+ картинок)



- Caltech pedestrians, [Dollar et al. CVPR'09]: 10 часов (350.000 прямоугольников)
- ImageNet project at Princeton, 10.000.000+ изображений

Локализация объектов: открытые проблемы

Локализация через понимание сцены



[Hoiem, Efros & Hebert, "Putting objects in perspective", CVPR'06 best paper]



[Desai, Ramanan & Fowlkes. "Discriminative models for multi-class object layout", ICCV'09 best paper]