Школа Анализа
Данных Яндекса

Курс «Анализ изображений и видео»

Лекция №4
«Классификация изображений»

Антон Конушин
Заведующий лабораторией компьютерной графики и мультимедиа
ВМК МГУ

7 октября 2016 года

Классификация изображений

Бинарная классификация

- Пешеход ли это?
- Бинарный ответ

$$y \in [0,1], 1 - \text{да}, 0 - \text{нет}$$

Car

Многоклассовая классификация

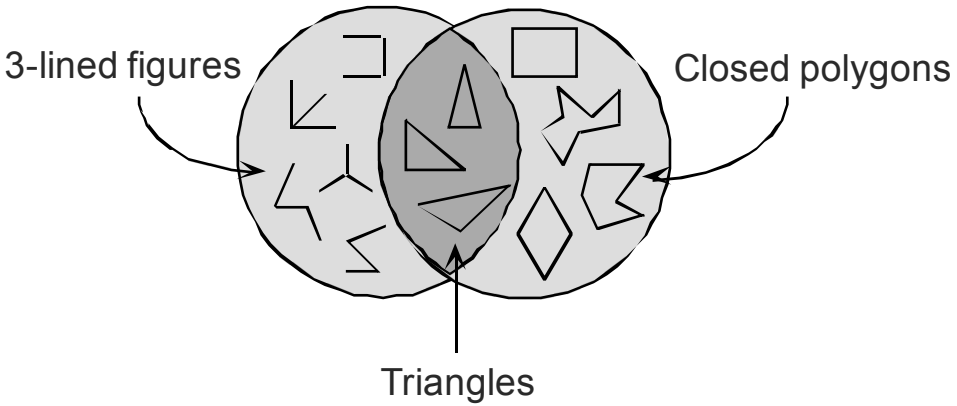
- К какому из заданных s классов относится данное изображение?
- Ответ – метка класса $c \in [1, S]$

- Другое название – категоризация изображений
- Определить, есть ли на изображении объект (сцена) заданной категории

Что такое «класс» / «категория»?

- Классическая идея прослеживается до Платона и Аристотеля
  1. Категория определяется набором свойств, общих для всех элементов из категории
  2. Принадлежность к категории бинарная
  3. Все элементы из категории одинаковы

Пример: Треугольники – это трехсторонние замкнутые ломанные



Проблема совмещения признаков разных категорий

Естественные категории

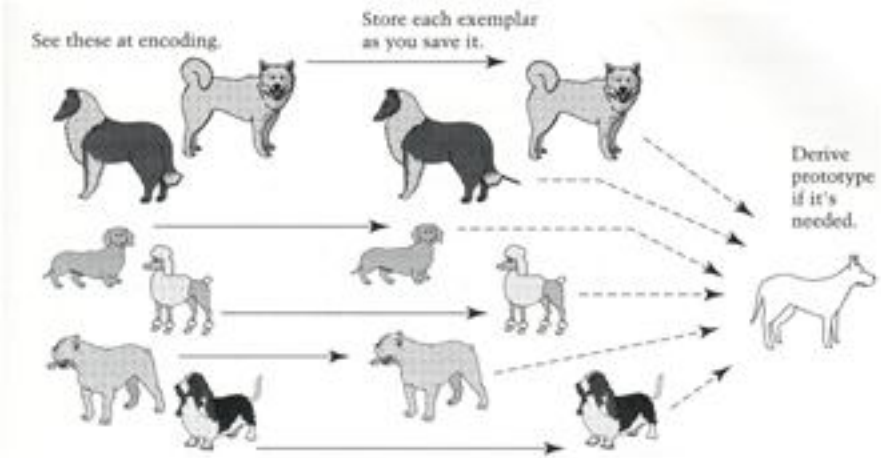


Figure 7.4. Schematic of the exemplar model. As each exemplar is seen, it is encoded into memory. A prototype is abstracted only when it is needed, for example, when a new exemplar must be categorized.

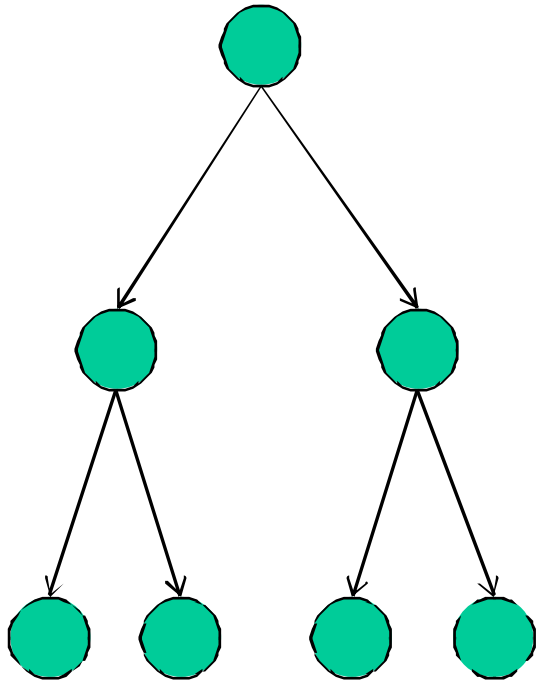
- Категория определяется «типичными» примерами (прототипами)
- Решение о соответствии категории принимается путем сравнения объекта со множеством примеров из категории
- Соответствие категории может быть «нечётким», т.е. можем оценить степень соответствия категории

Субъективность категории

Категория объекта зависит от наблюдателя.

«Женщины, Огонь и Опасные Вещи» – это отдельная категория в языке Австралийских аборигенов (Lakoff 1987)

Сколько всего категорий?

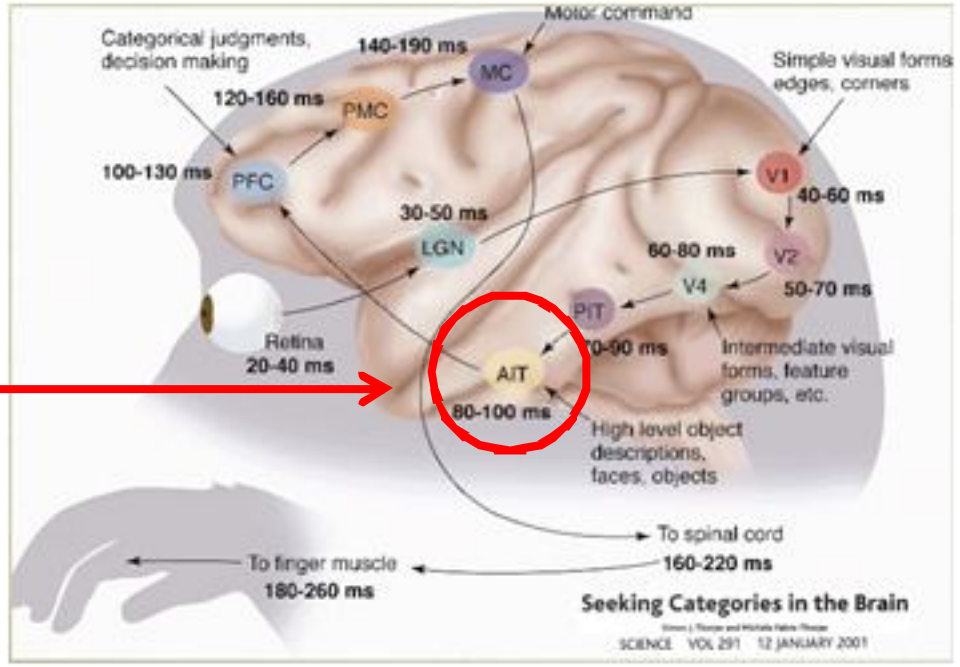


- Категории естественным образом иерархически организуются
- Группируя английские слова, получим до 100 тыс. категорий
  - 171000+ слов всего, 47000 устаревших слов, половина существительные

Biederman 1987

Распознавание человеком

Зона распознавания объектов



Методы исследования:

- Психологические эксперименты
- Изучение людей с заболеваниями и повреждениями мозга
- Измерения активности участков мозга

Крошки-картинки (Tiny images)

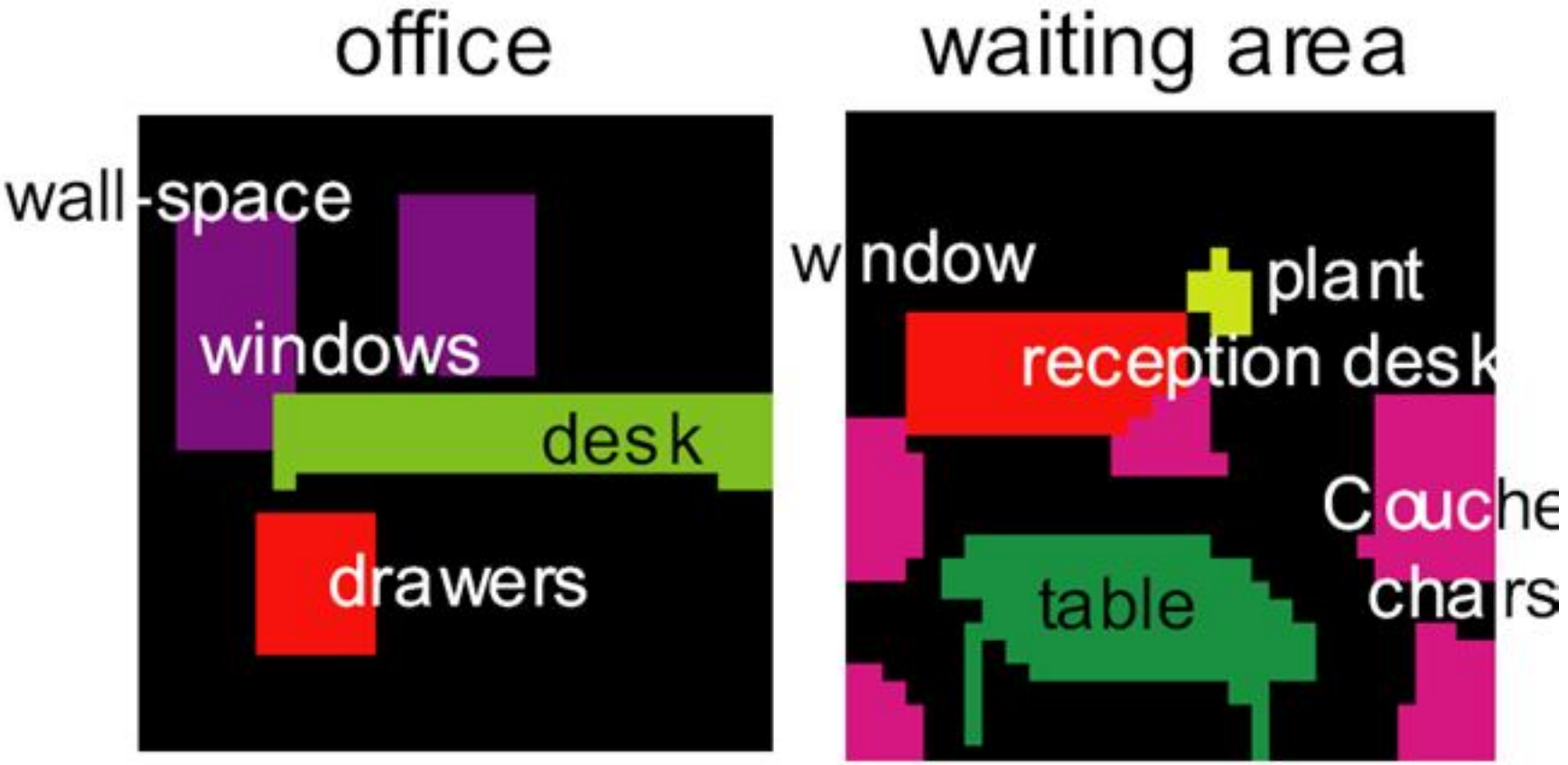
- Первый эксперимент по сбору и использованию 100 млн. коллекции изображений
- Какого размера нам нужны изображения? (Место ограничено)
- Попробуем на человеке

A. Torralba, R. Fergus, W. T. Freeman 80 million tiny images: a large dataset for non-parametric object and scene recognition IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30(11), pp. 1958-1970, 2008.

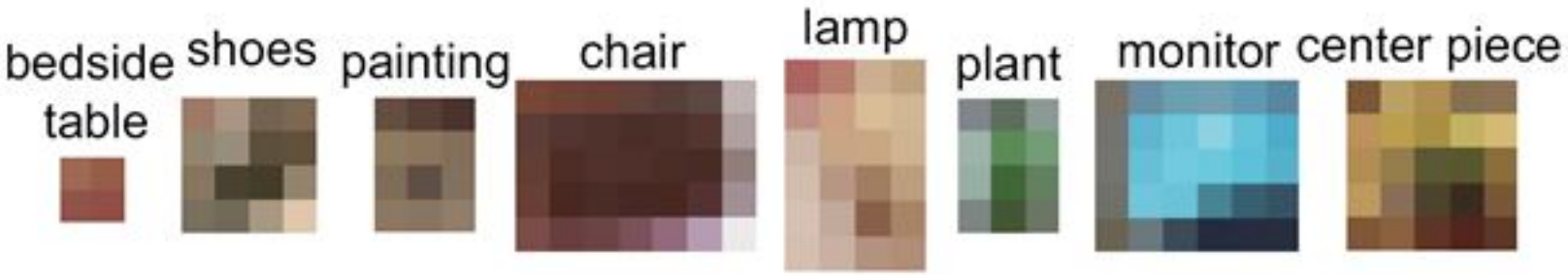
Примеры изображений

Примеры изображений

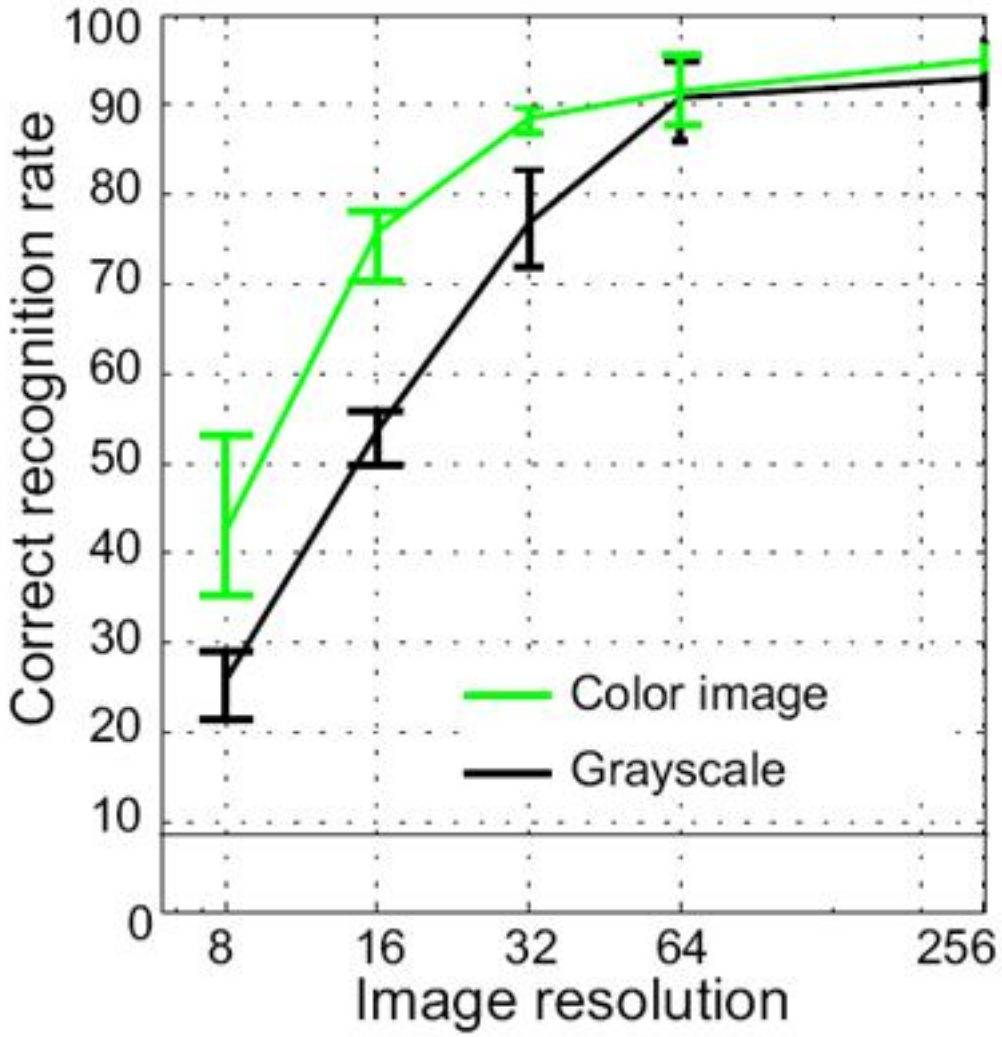
Сегментация



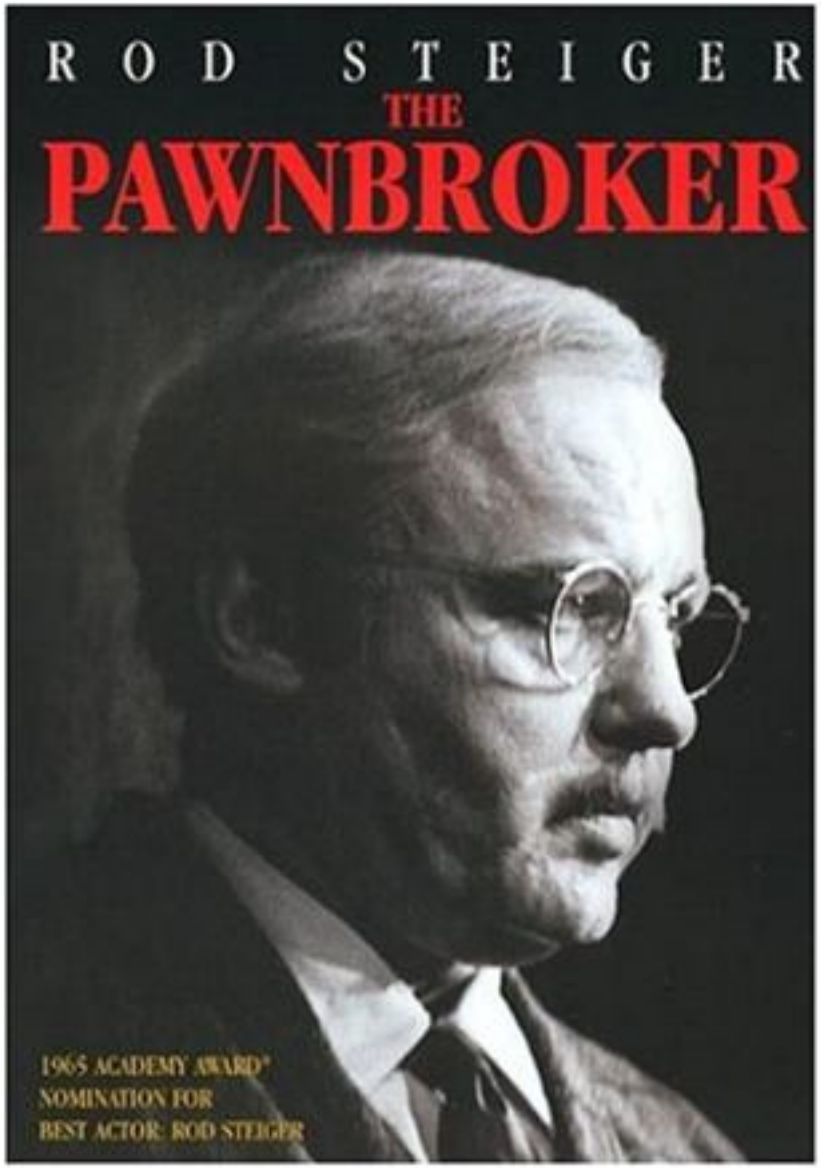
Отдельные объекты



Распознавание человеком



В кино



S. Lumet, 1965

Mary Potter (1976)

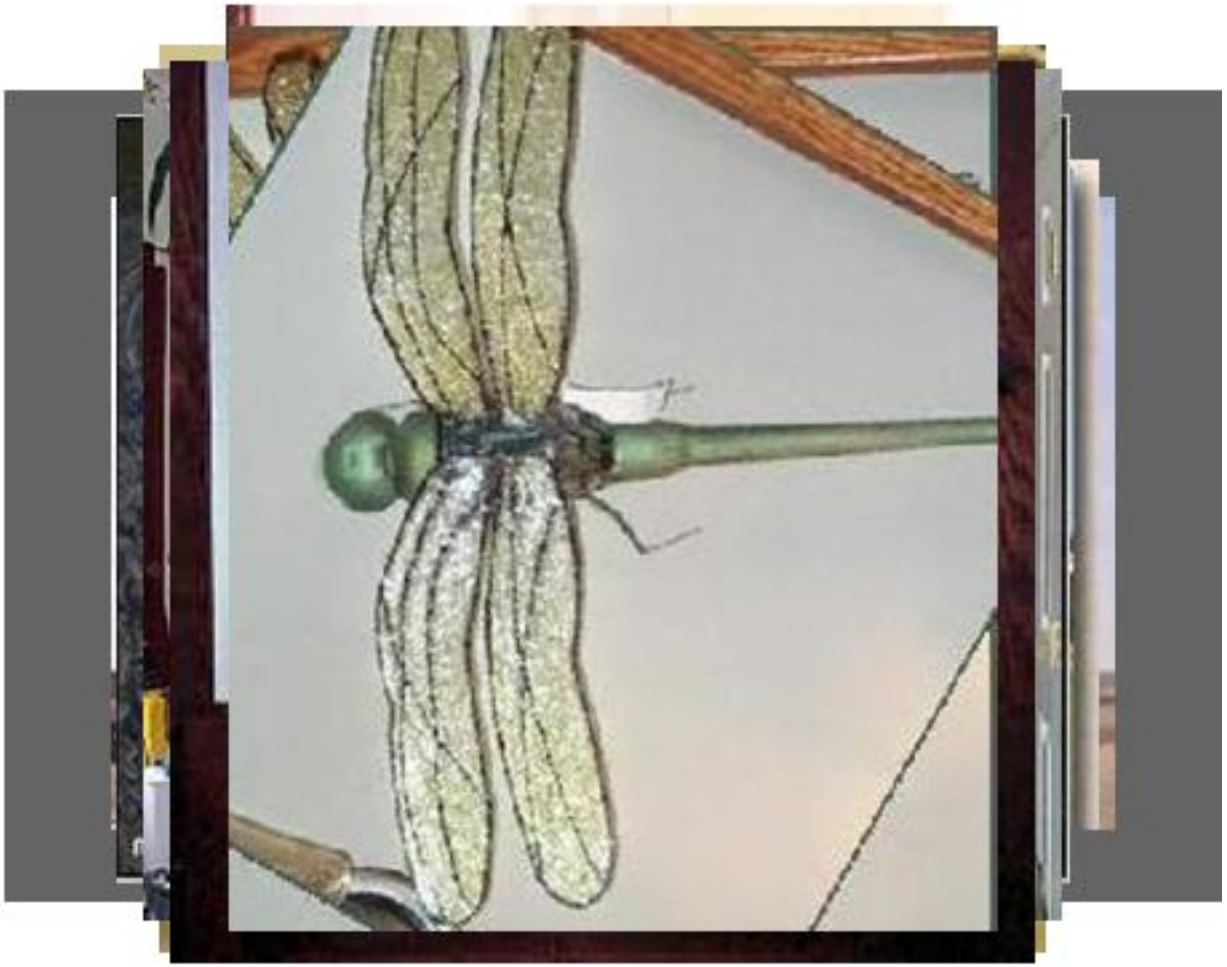
- Мэри Поттер (1975, 1976) продемонстрировала, что при скоростном показе серии изображений (100 мсек на изображение), человек успеваем извлечь и запомнить много визуальной информации из каждого изображения

Схема экспериментов

Potter, Biederman, etc. 1970s

Демонстрация

- Инструкция: 10 изображений будут показаны по полсекунды на каждую. Задача **запомнить все изображения**!



Тест на память

- Какие из фотографий вы запомнили?
- **Если вы видели эту фотографию, нужно хлопнуть в ладоши (один раз)**
- Если вы не видели эту фотографию, тогда ничего не нужно делать

ДА

HET

ДА

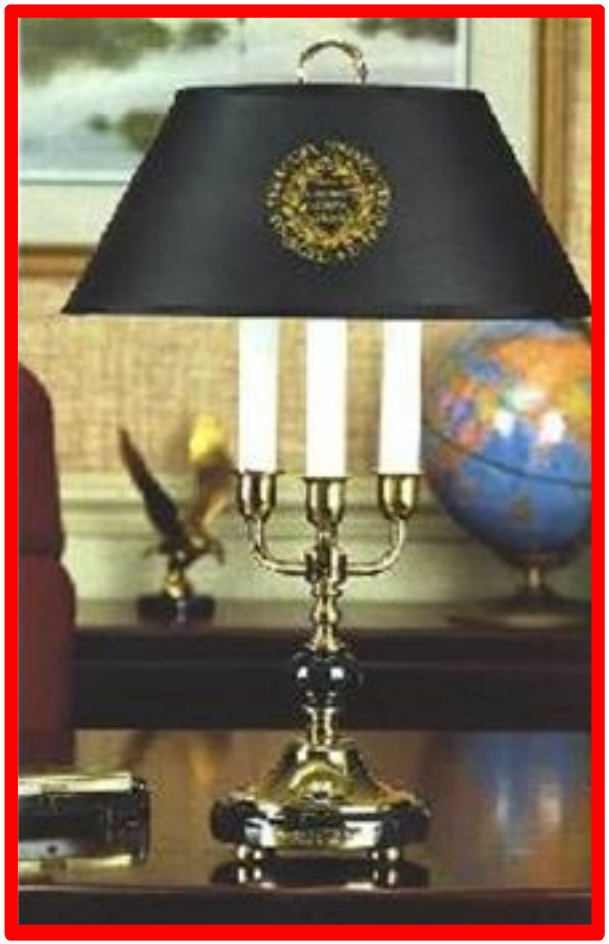
ДА

HET

HET

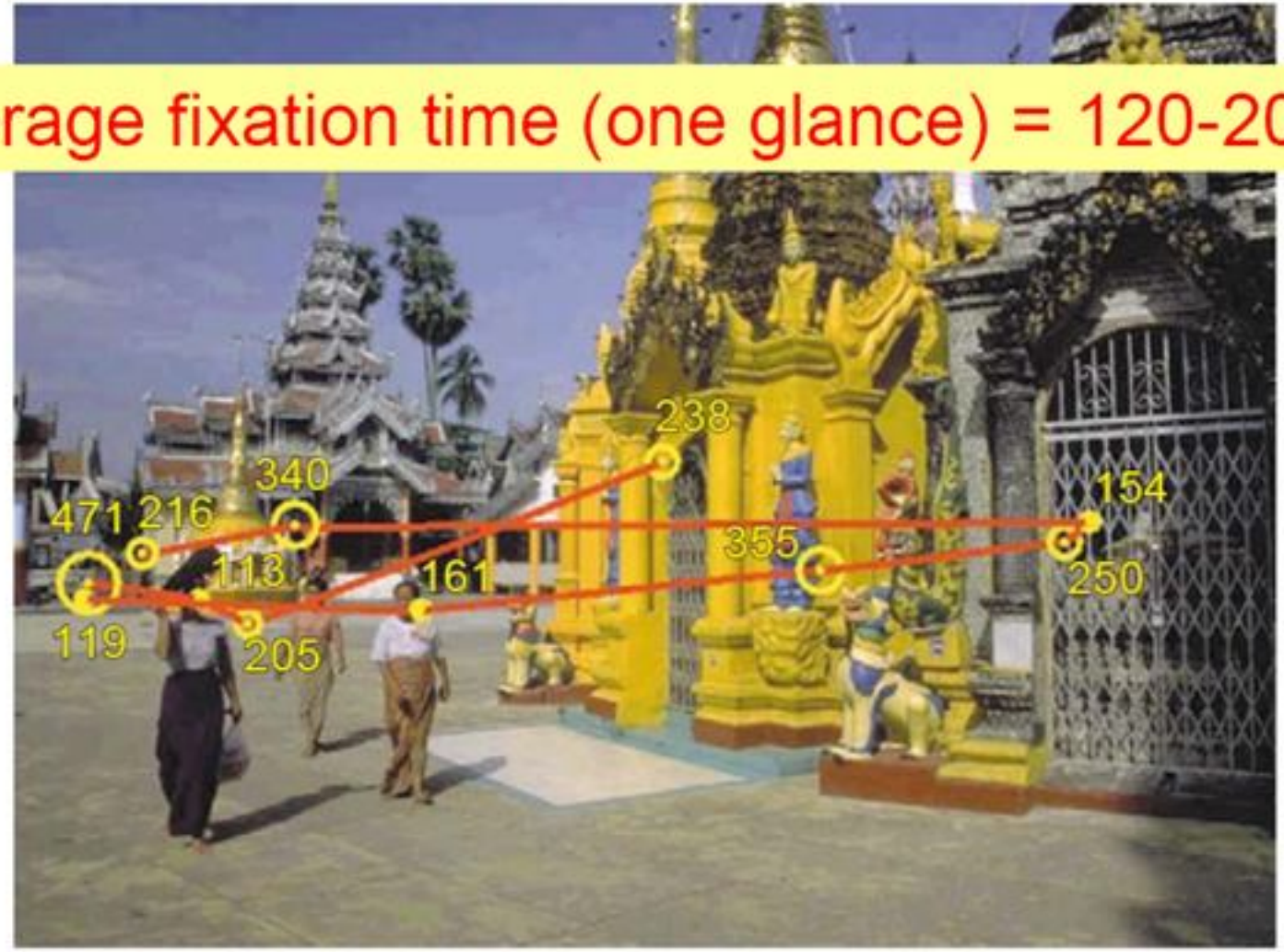
(c) Worldprints.com

ДА



HET

Один взгляд (2007)

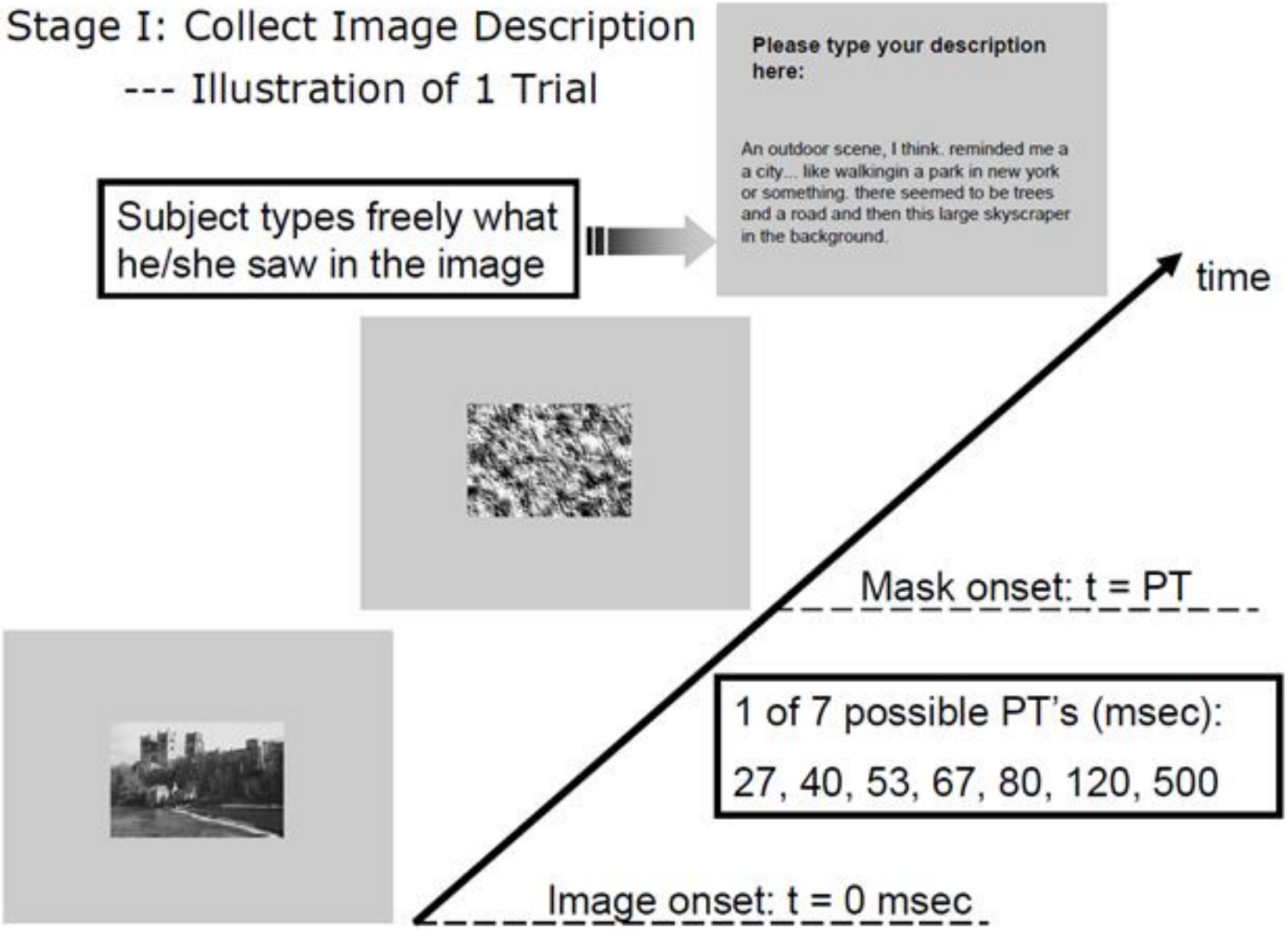


Что мы узнаём с одного взгляда?

Fei-Fei et al. JoV 2007

Примеры изображений

Схема эксперимента



Пример ответов

PT = 27ms

Couldn't see much; it was mostly dark w/ some square things, maybe furniture. (Subject: AM)

PT = 40ms

This looked like an indoor shot. Saw what looked like a large framed object (a painting?) on a white background (i.e., the wall). (Subject: RW)

PT = 67ms

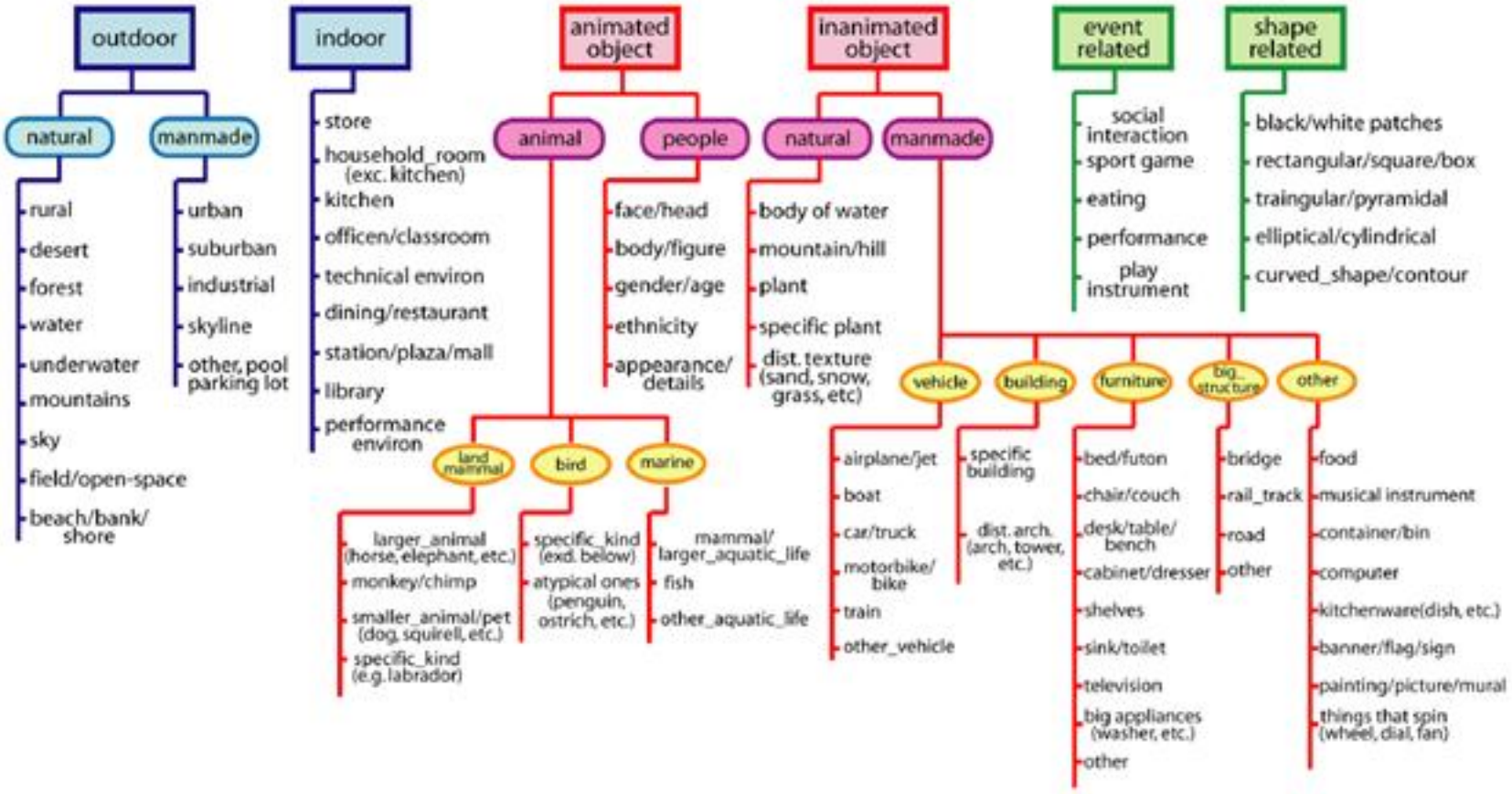
I saw the interior of a room in a house. There was a picture to the right, that was black, and possibly a table in the center. It seemed like a formal dining room. (Subject: JB)

PT = 500ms

This is indoors. It's must be a rich person's house. There are many paintings on the wall. The largest painting might have a fireplace beneath it. I think the largest painting was that of a man standing erect. The room is richly decorated and it looks like one of the rooms in Mr. Darcy's house in the A&E movie Pride and Prejudice. Or maybe it more closely resembles one of the rooms where the one of the rooms in Hungtington's house (at the Huntington).

Fei-Fei et al. JoV 2007

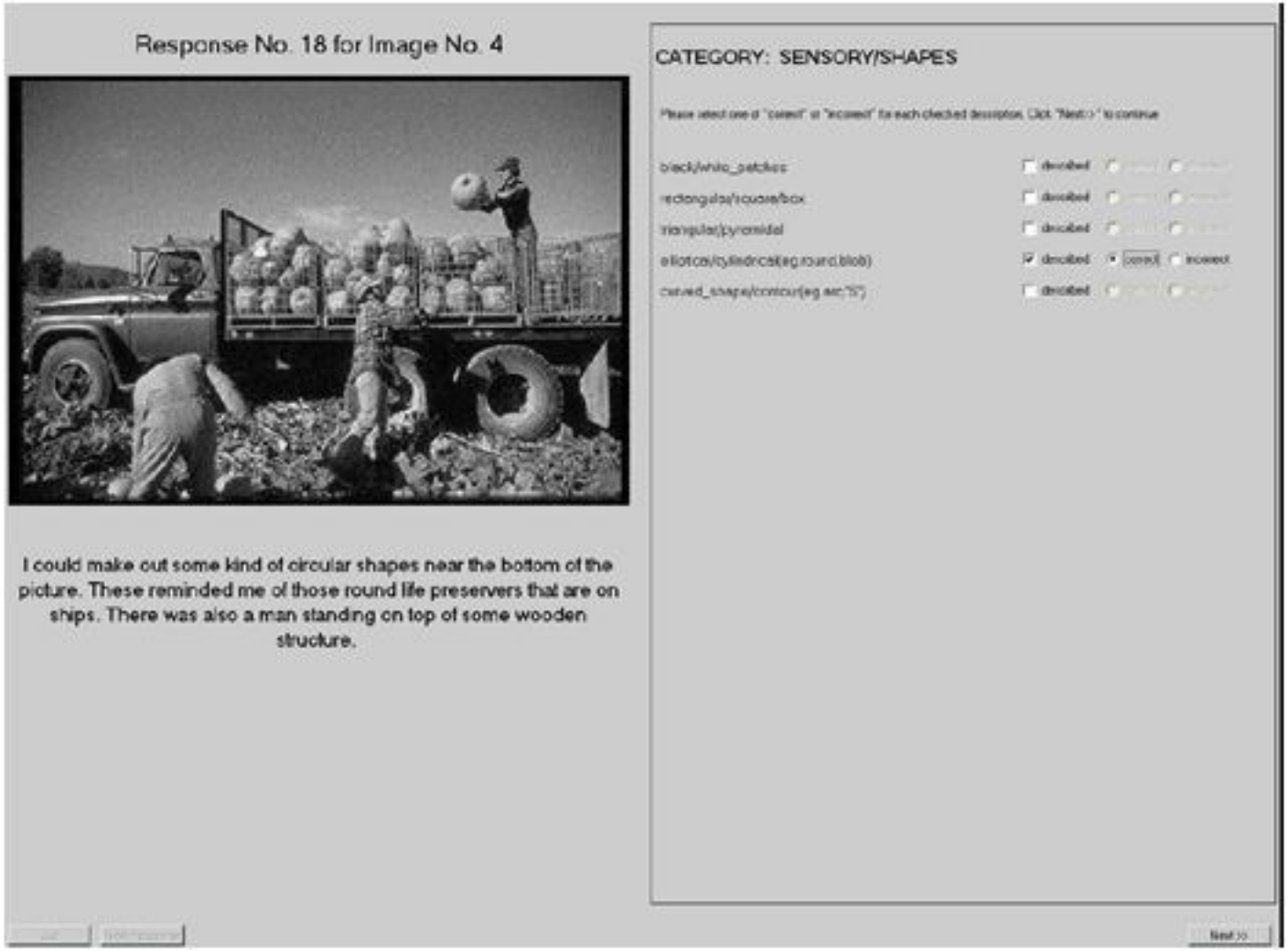
Атрибуты



Fei-Fei et al. *JoV, 2007*

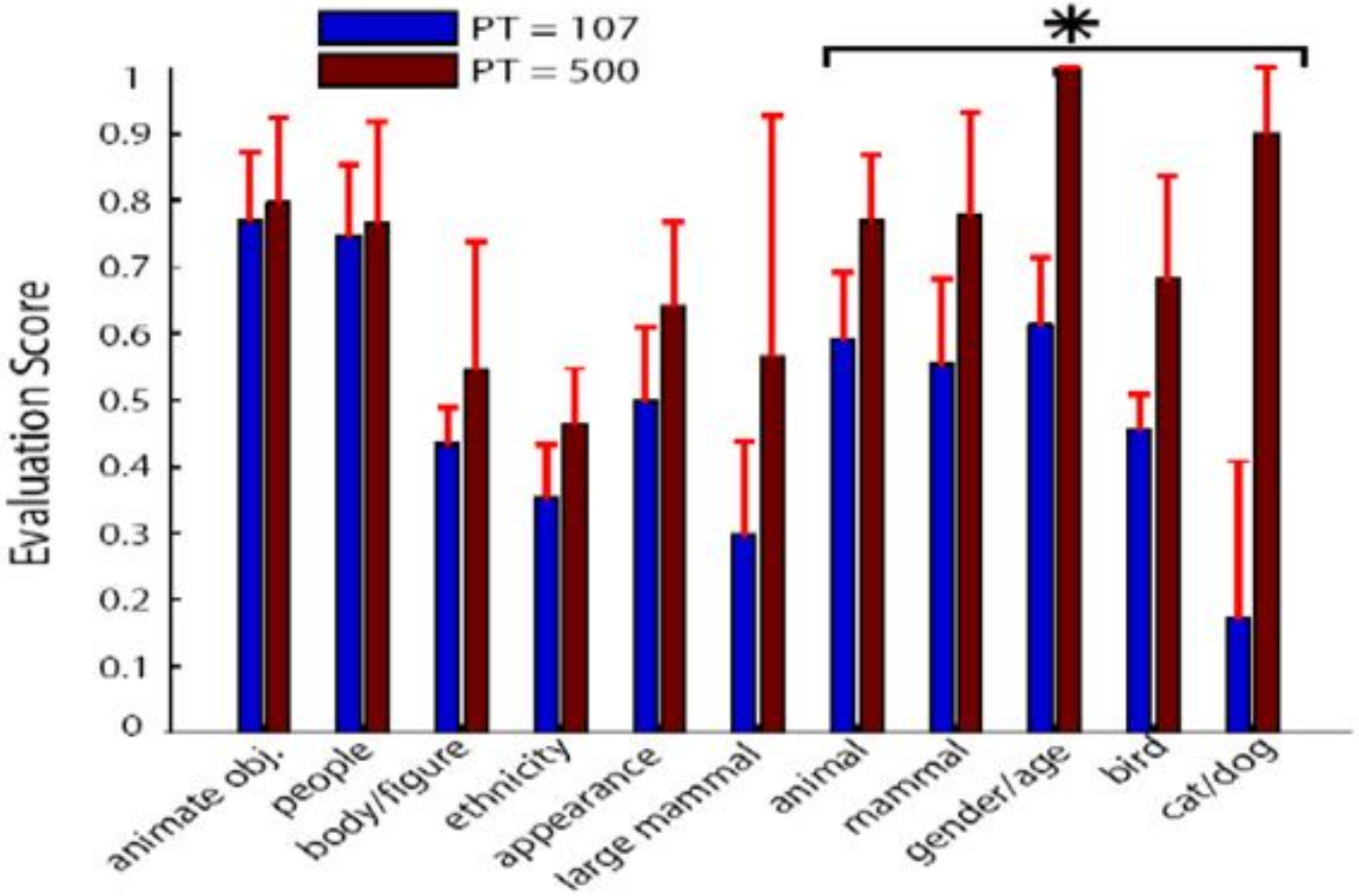
Выделили «категории» и «атрибуты», которые могли встречаться в описаниях

Обработка результатов



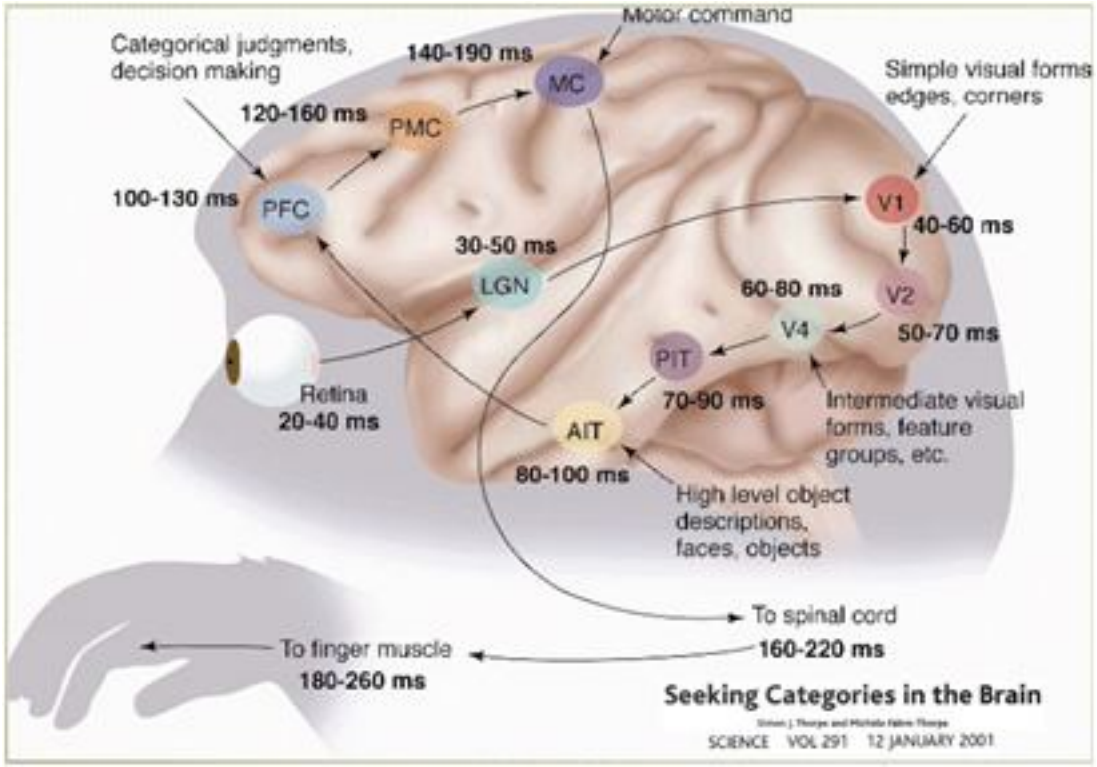
Все описания вручную обработали, определив, встречается ли в них категории из списка

Результаты



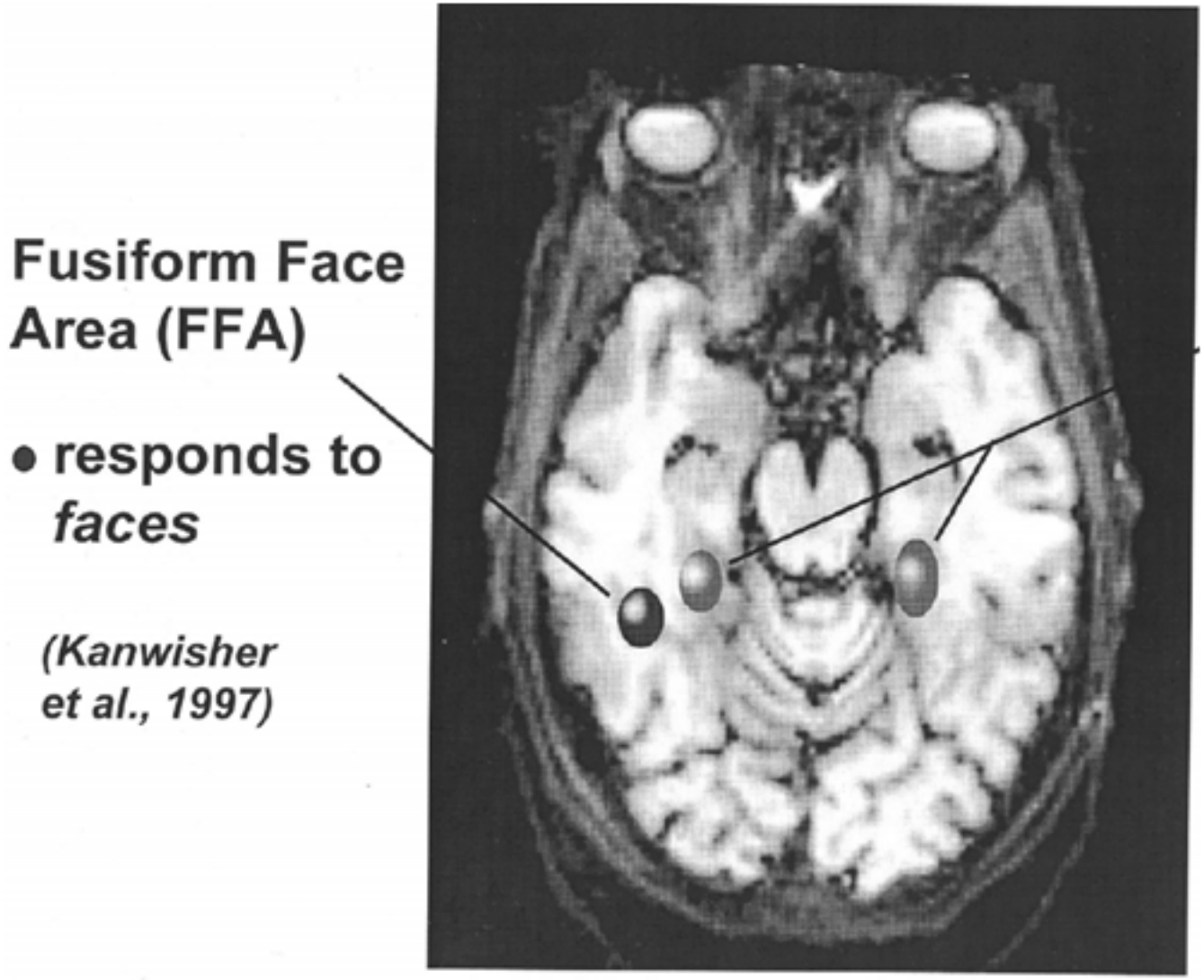
Ряд вещей человека за 100мс надежно определяет, для каких-то требуется больше времени и больше «взглядов»

Некоторые выводы



- В этих экспериментах человек не знал, изображения какого типа он увидит (нет априорной информации)
- Скорость распознавания заставлять предполагать, что в таком режиме распознавание не требует обратных связей
- Это некоторое обоснование «feed-forward architectures»

FFA (зона распознавания лиц)



FFA (зона распознавания лиц)

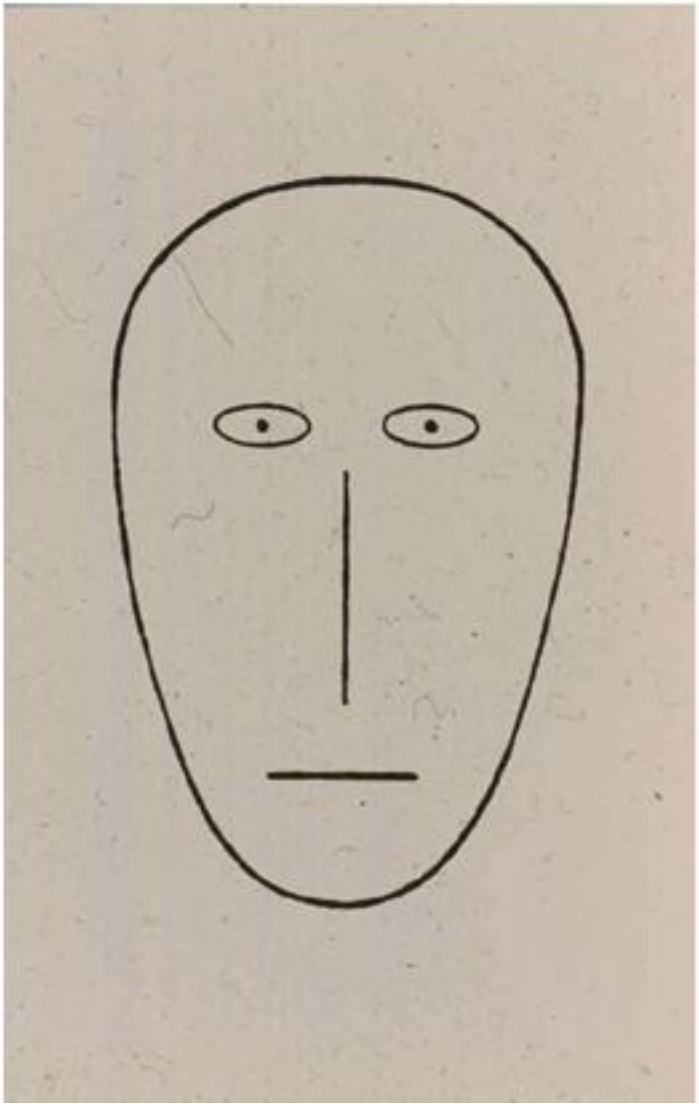
FFA responds more strongly to faces:

1.8 1.8 1.8 1.7 1.6 %MR Signal

than non-face control stimuli:

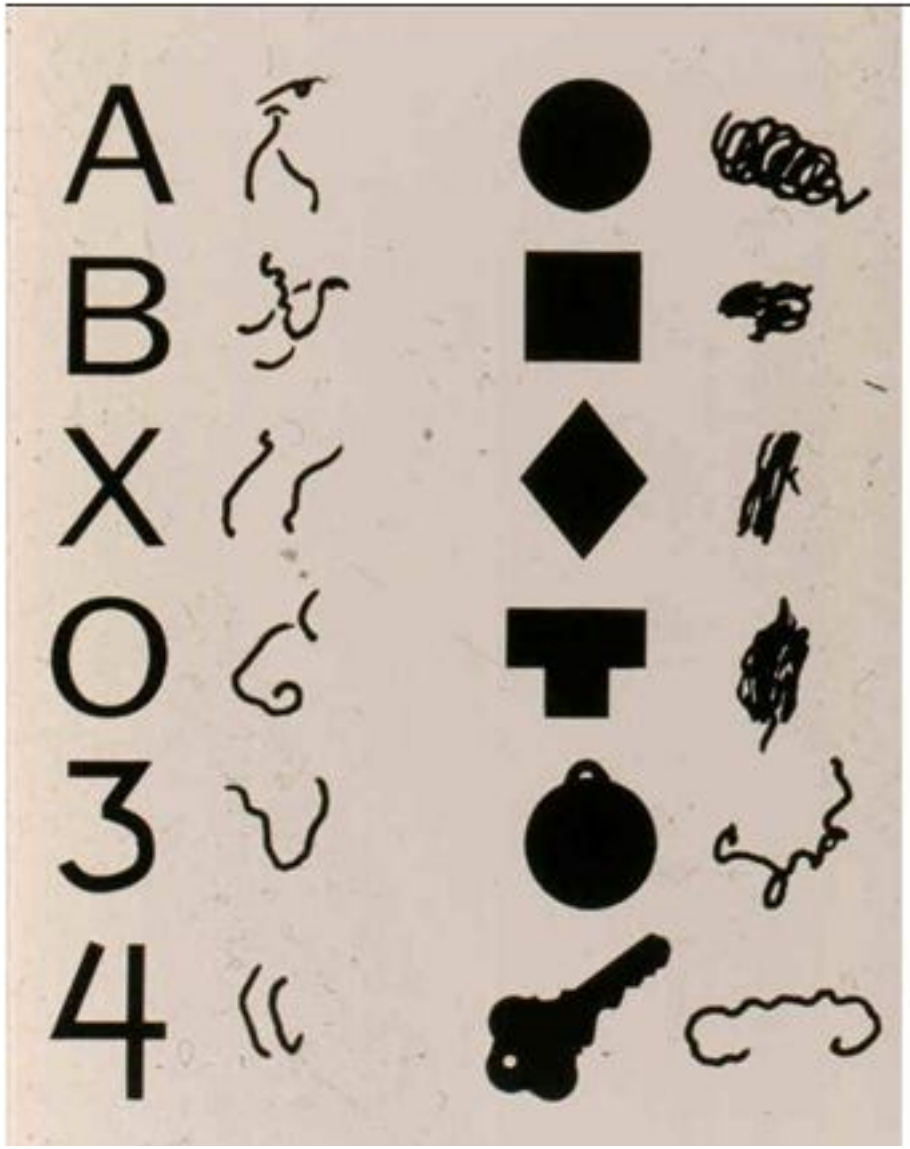
1.3 1.3 0.9 0.9 0.7 %MR Signa

Восприятие лиц



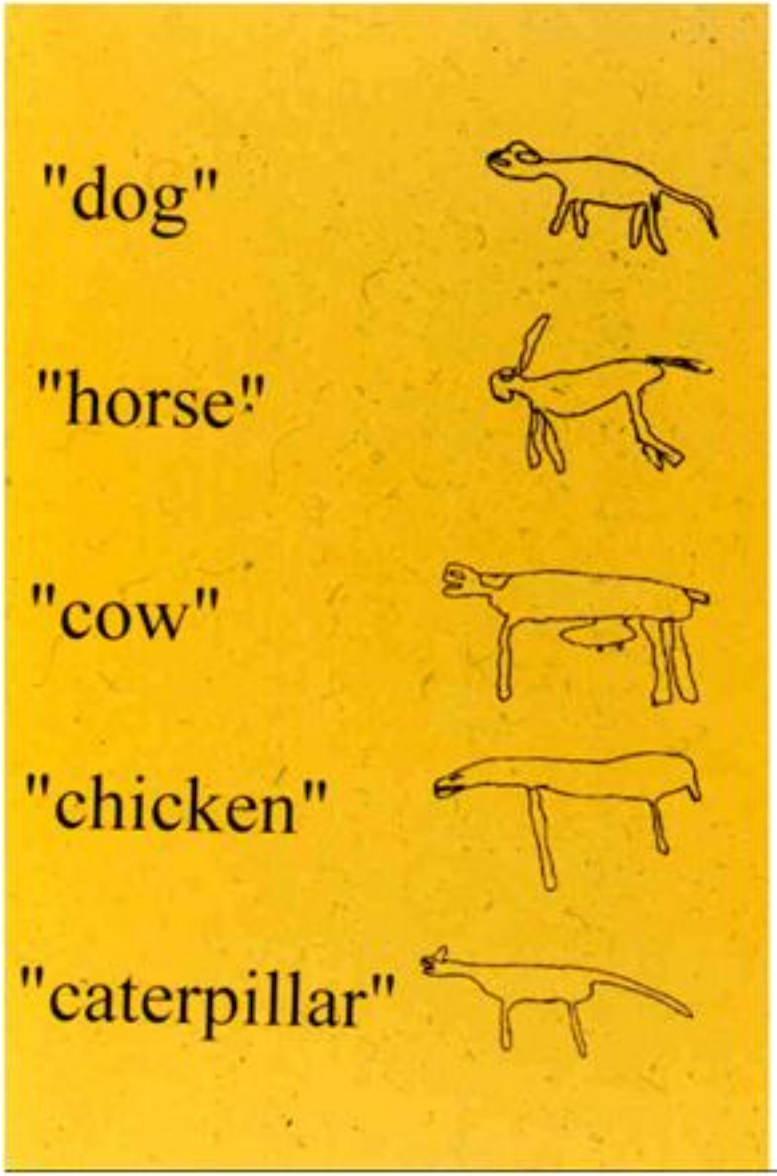
Described by a patient with prosopagnosia as "an apple with two worm holes, a folded over stem and a crease".

Восприятие объектов



Copying by a
Patient suffering
From object agnosia

Восприятие живых объектов

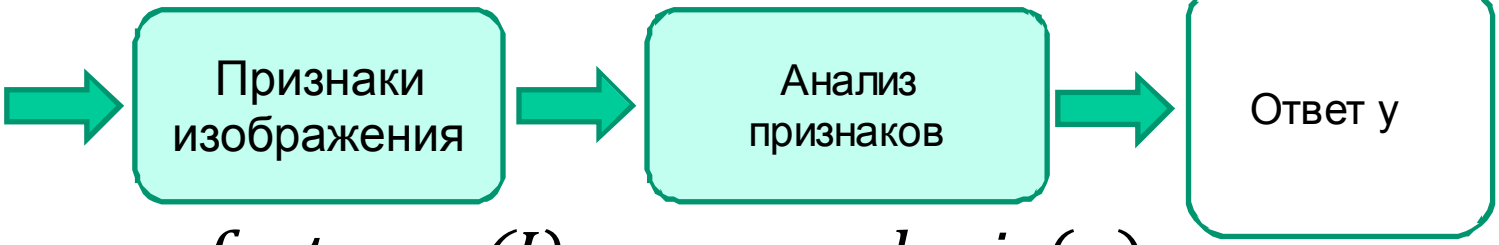


Drawing by a patient with agnosia for animate objects.

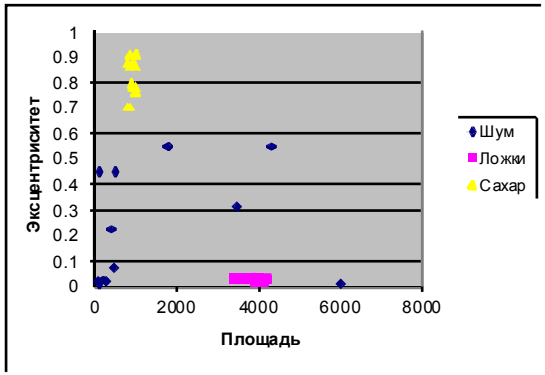
Декомпозиция задачи

Классификация:

Изображение I



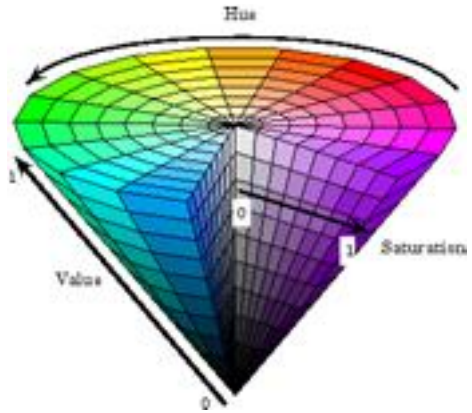
$$x = features\,(I) \quad y = analysis(x)$$



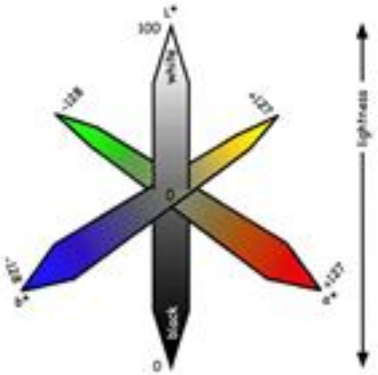
- Мы можем сконструировать так много признаков, что эвристические правила не помогут
- Придётся обучать классификаторы машинным обучением

Признаки пикселей

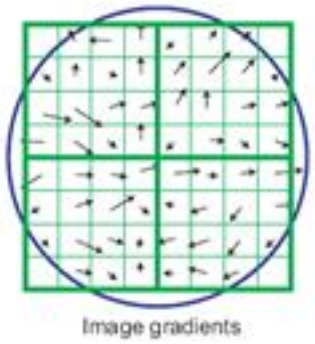
Пространство цветов RGB



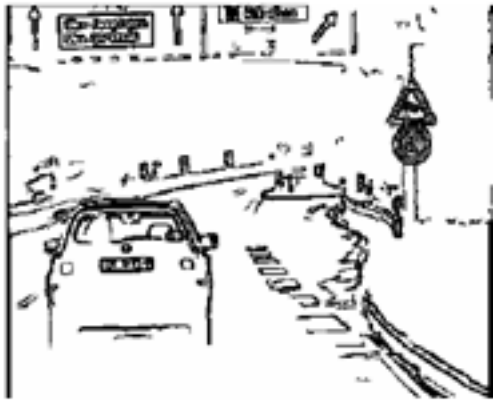
Пространство цветов HSV



Пространство цветов L*a*b*



Градиенты в каждом пикселе



Наличие и ориентация края в каждом пикселе

Использование напрямую

- Можно ли использовать признаки всех пикселей напрямую?
- Можно, если все изображения для классификации будут одинакового размера
- Нормализуем изображения, т.е. приведем их к одному размеру
- Вытянем изображение в вектор, признаки пикселей будут элементами вектора

Для распознавания, изображение должно описываться вектор-признаком фиксированной длины!

Распознавание NN

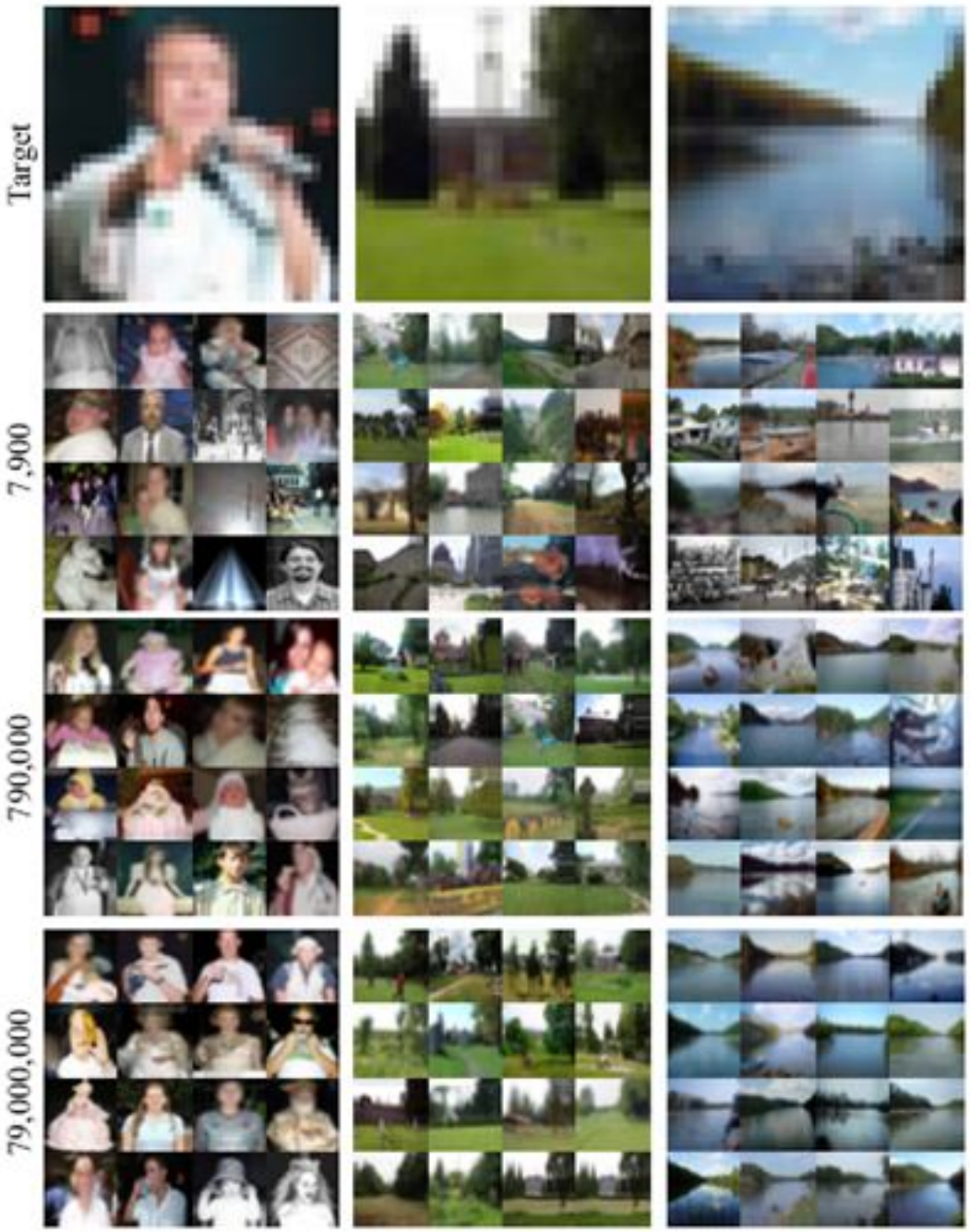
- Возьмём большую коллекцию tiny images
- Воспользуемся методом классификации NN

$$D^2_{SSD} = \sum_{x,y,c} (I_1(x,y,c) - I_2(x,y,c))^2$$

SSD-метрика

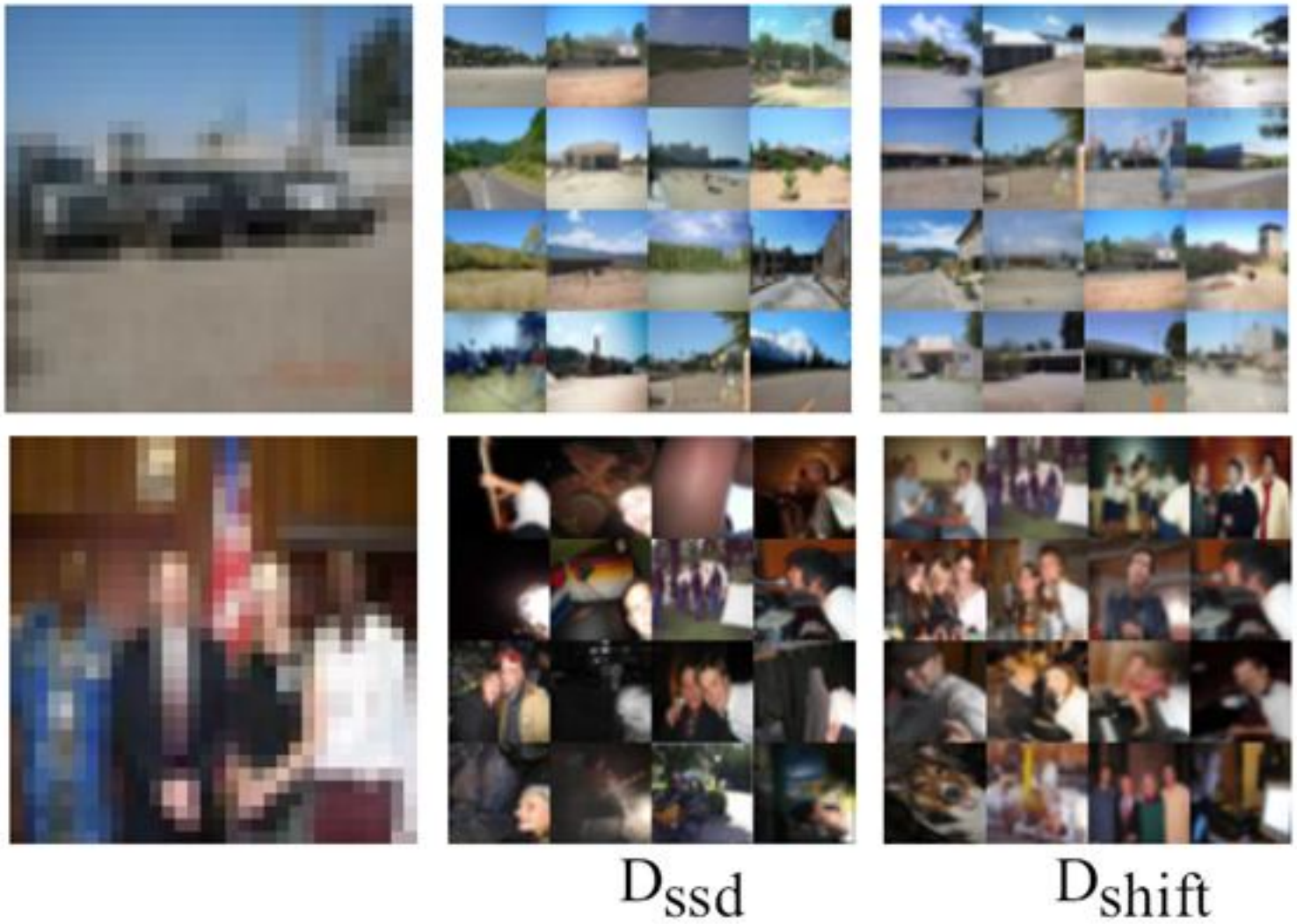
$$D^2_{warp} = \min_{\theta} \sum_{x,y,c} (I_1(x,y,c) - T_\theta I_2(x,y,c))^2$$

Сравнение с искажением



- Наиболее похожие по SSD в зависимости от размеров коллекции
- Похожие изображения содержат те же самые объекты
- Если есть метки классов, то NN или k-NN даст ответ

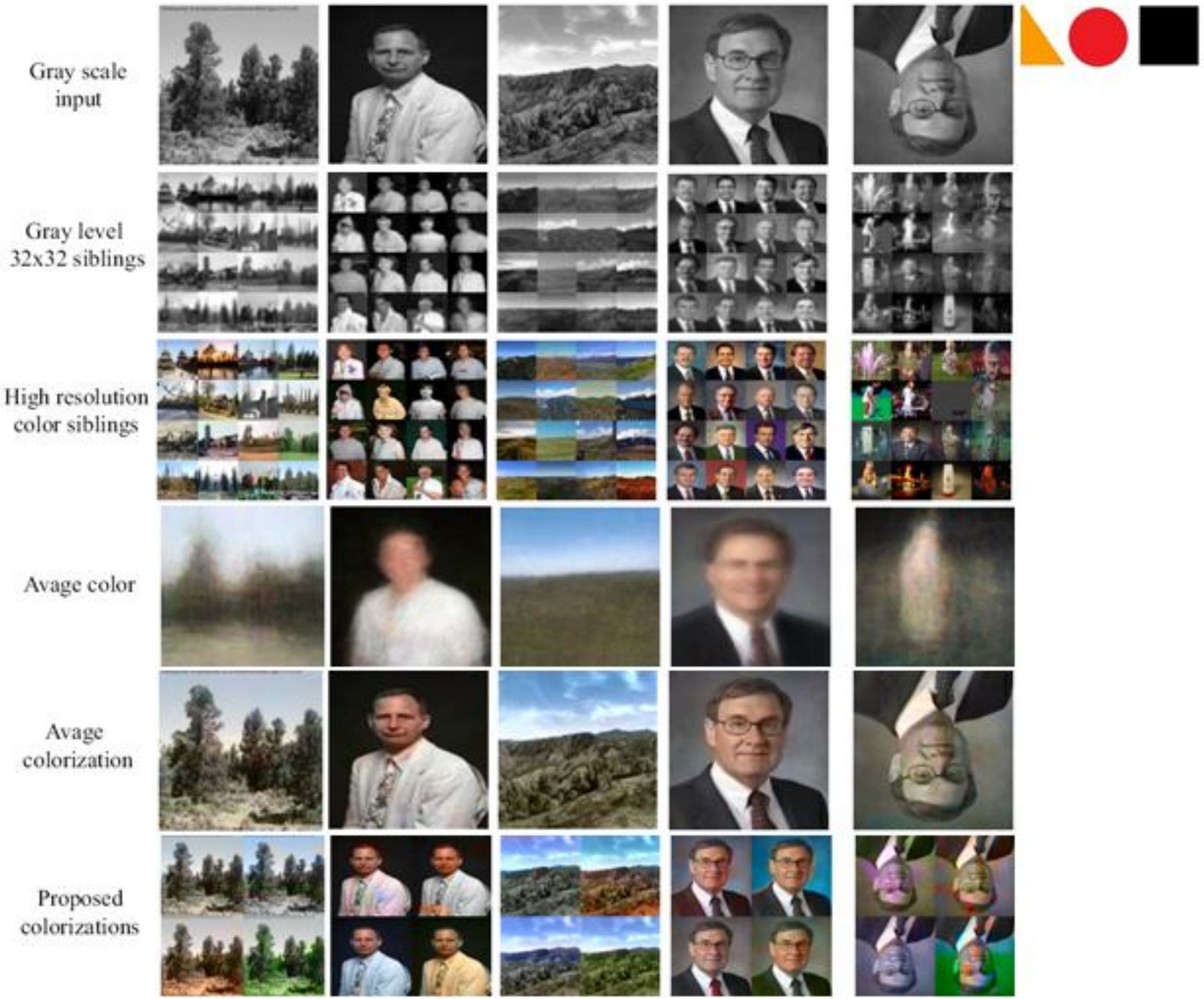
Результат улучшенной метрики



Результат улучшенной метрики

Выделение лиц через NN

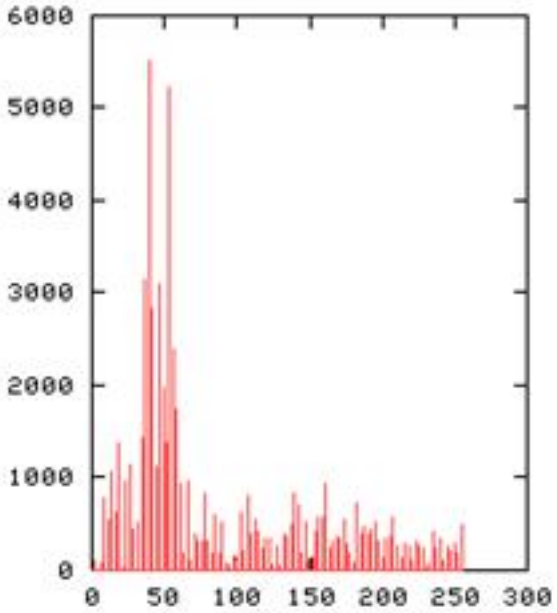

25
27
20


Gray scale input
Gray level 32x32 siblings
High resolution color siblings
Avage color
Avage colorization
Proposed colorizations

Выводы из крошек-картинок

- Человек может распознать сцену с 90% правильностью по цветному изображению 32x32
- Если есть достаточно большая выборка, то мы можем для практически любого изображения найти очень близкое и по содержанию, и по ракурсу, и т.д..
- Метод «ближайшего соседа» при больших выборках может неплохо работать

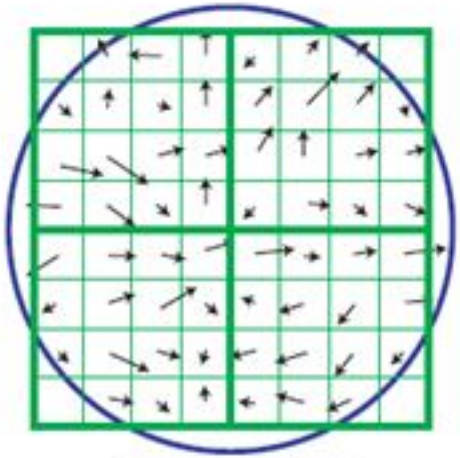
Гистограммы

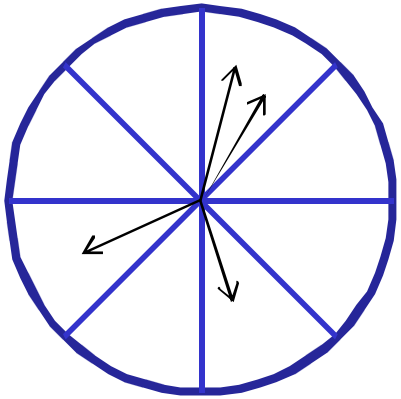


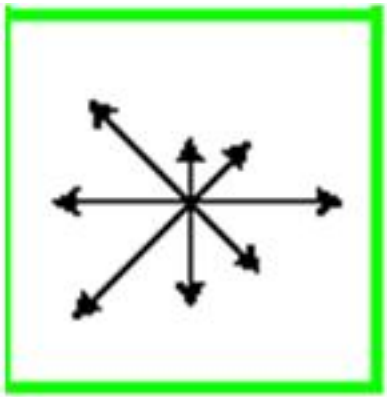
- Вместо использования пикселов напрямую, можем считать разные статистики распределения характеристик
  - Цвет, края, градиенты, текстуру и т.д.
- Гистограммы – стандартный способ непараметрического описания распределения признаков

Images from Dave Kauchak

Квантование признаков

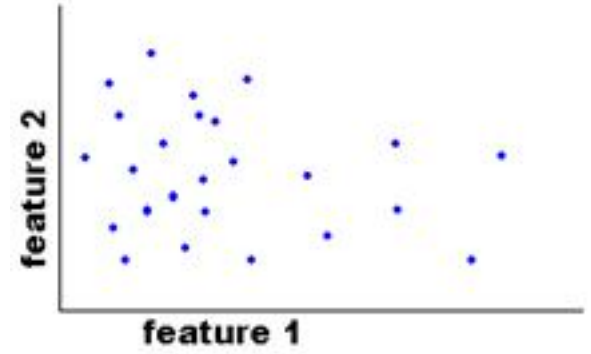


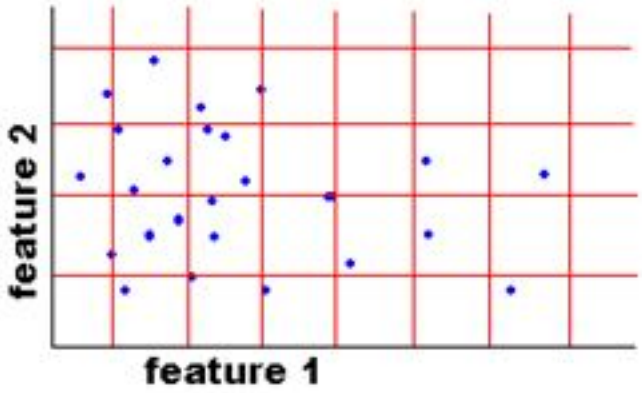


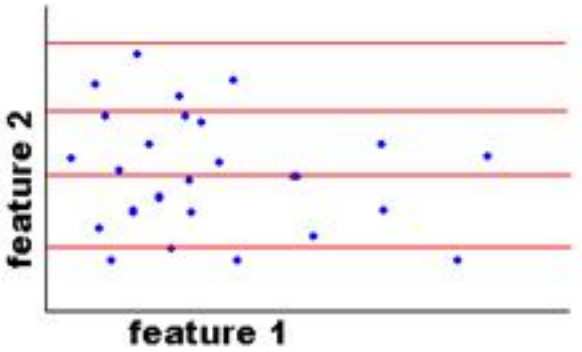


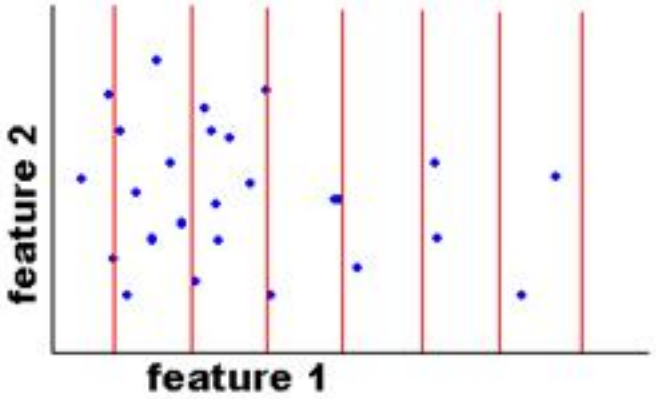
- Для некоторых признаков потребуется дискретизация (квантование)
  - Если признаки вещественные
  - Если слишком большое разрешение
- Пример для направления градиента
  - Разбиваем 360 градусов на 8 секторов
  - Каждый сектор нумеруем от 1 до 8
  - Считаем число пикселов, градиенты которых попадают в соответствующие сектора

Многомерные признаки









- Совместная (Joint) гистограмма
  - Требуется много данных
  - Теряем в разрешении, чтобы уменьшить число пустых ячеек

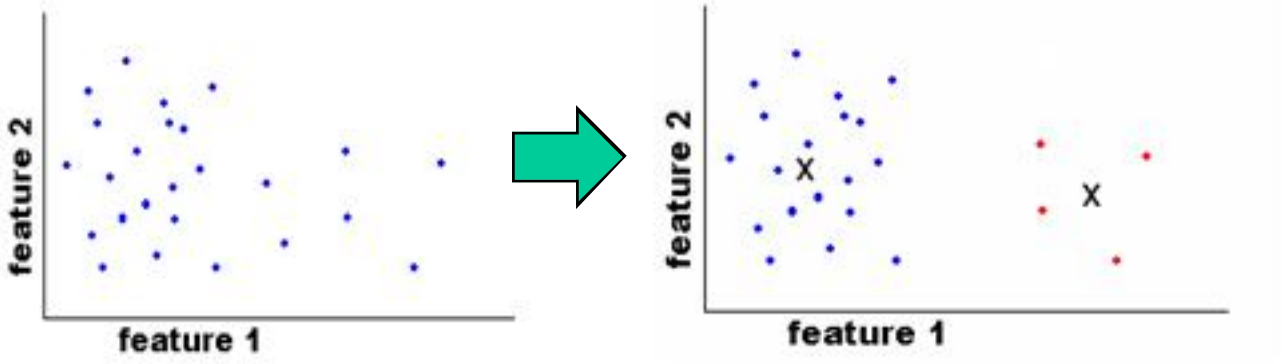
Маржинальные гистограммы

- Требуются независимо распределенные признаки
- Больше данных на одну ячейку, чем в совместных гистограммах

Images from Dave Kauchak

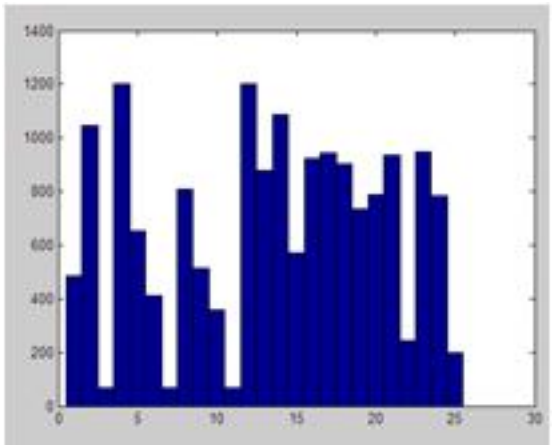
Квантование признаков

Не все цвета встречаются!



- Во многих случаях равномерное квантование по сетке не подходит
  - Высокая размерность
  - Неравномерные распределения
- Кластеризация данных
  - K-средних и производные чаще всего используются
- Для каждого нового изображения мы квантуем признаки и строим гистограмму частот

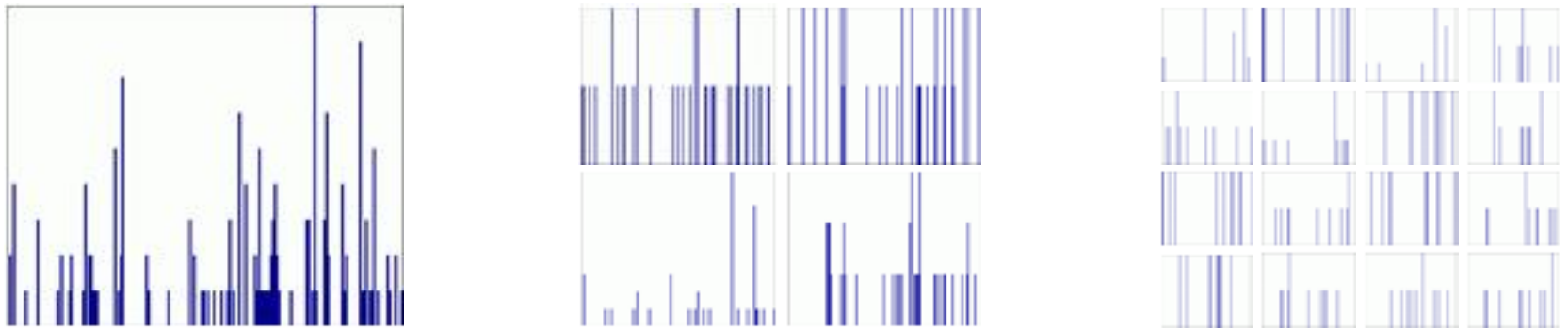
Пространственное распределение



У всех этих трех изображений похожие гистограммы цветов

Пространственная пирамида

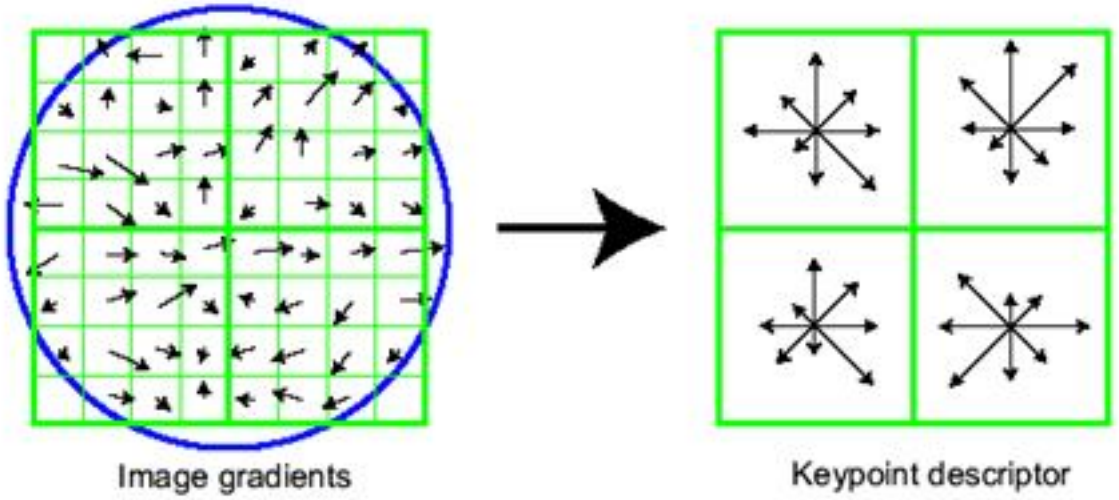
Вычислим гистограму в каждом блоке и объединим все гистограммы в один вектор-признак



Уровень 0 Уровень 1 Уровень 2

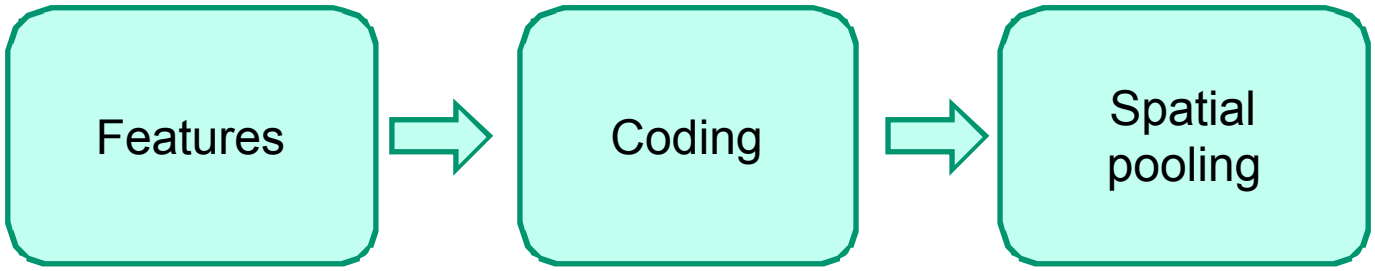
S. Lazebnik et. Al. **Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories** *CVPR 2006*

Гистограммы градиентов

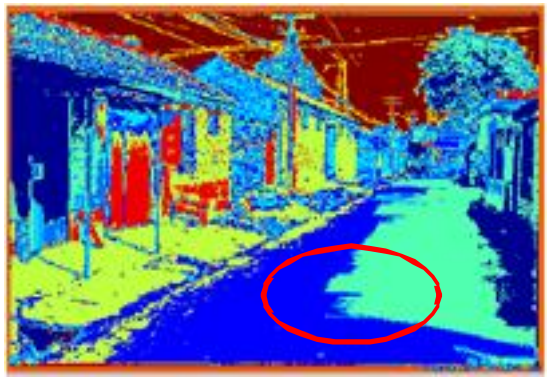


- Много названий – HOG (Histogram of Oriented Gradients), дескриптор SIFT
- Построение
  - Разбиваем изображение на области сеткой
  - В каждой сетке считаем гистограмму распределения пикселов по ориентации градиентов
  - Обычно 6-8 корзин в каждой гистограмме
  - Вычисляем градиент в каждой точке
- Какие достоинства у таких признаков?

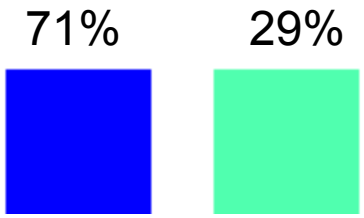
Общая схема для признаков



RGB



Квантование на 10 уровней



Посчитанная гистограмма

Spatial pooling (пространственная агрегация) = выбор областей для агрегации и метод агрегации

Мешок слов

Классификация текстов

- Представление документа без порядка: частоты слов из словаря («мешок слов») Salton & McGill (1983)

Классификация текстов

- Представление документа без порядка: частоты слов из словаря («мешок слов») Salton & McGill (1983)

Классификация текстов

- Представление документа без порядка: частоты слов из словаря («мешок слов») Salton & McGill (1983)

«Визуальные слова»

- Что в нашем случае «слово» и «словарь»?
- «Визуальное слово» – часто повторяющийся фрагмент изображения («visual word»)
- В изображении визуальное слово может встречаться только один раз, может ни разу, может много раз

Sivic, J. and Zisserman, A. Video Google: A Text Retrieval Approach to Object Matching in Videos. ICCV 2003

Словарь и мешок слов

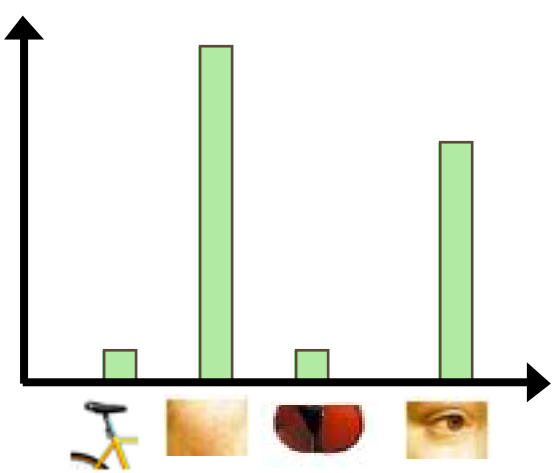
- «Словарь» – набор фрагментов, часто повторяющихся в коллекции изображений

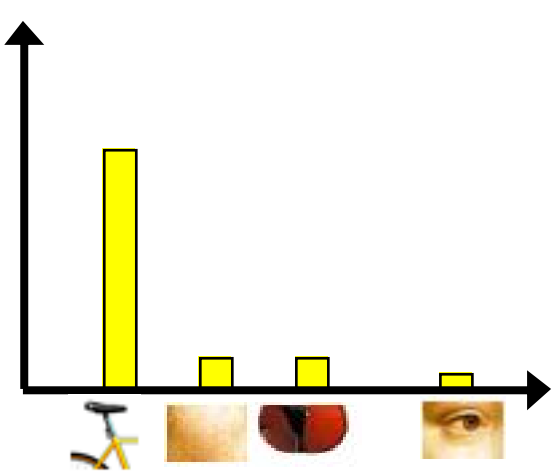
- Как составить словарь?
  - Составить большой список всех фрагментов по всей коллекции
  - Разделить весь список на похожие группы
  - Будем считать все фрагменты в одной группе – «экземплярами» одного и того же слова
  - «Средний» элемент возьмём как слово для словаря

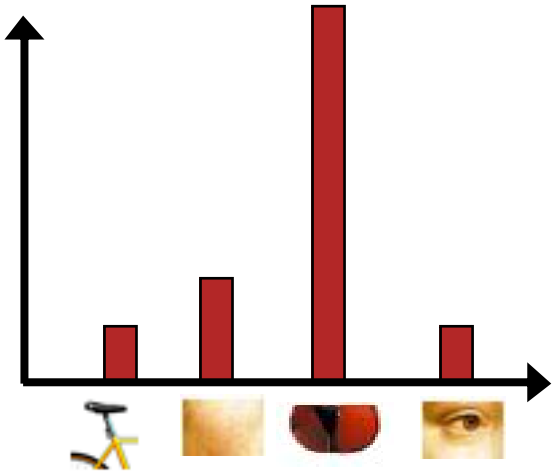
- «Мешок слов» - гистограмма частот появления визуальных слов в изображении

Схема метода «мешок слов»

1. Построение словаря
    1. Извлечение фрагментов из обучающей выборки
    2. Кластеризация и построение «визуального» словаря
2. Построение дескрипторов
    1. Извлечение фрагментов из изображения
    2. Квантование фрагментов по словарю
    3. Построение гистограммы частот визуальных слов

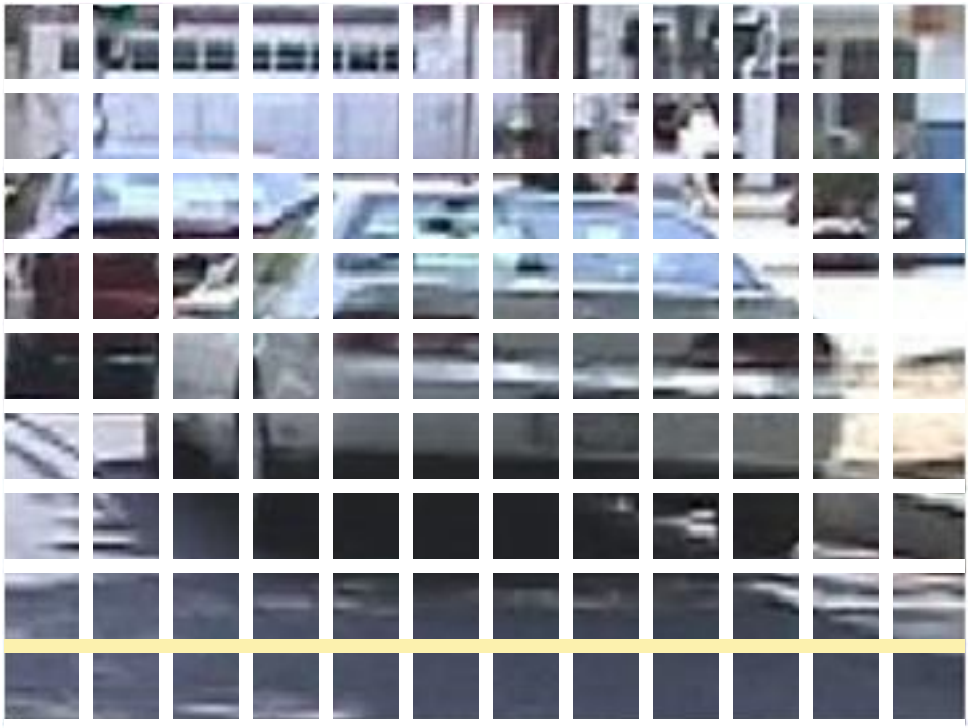


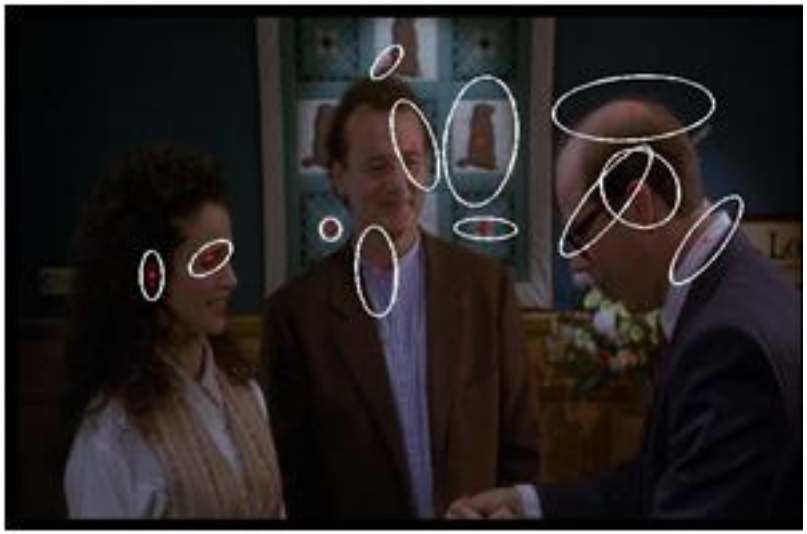




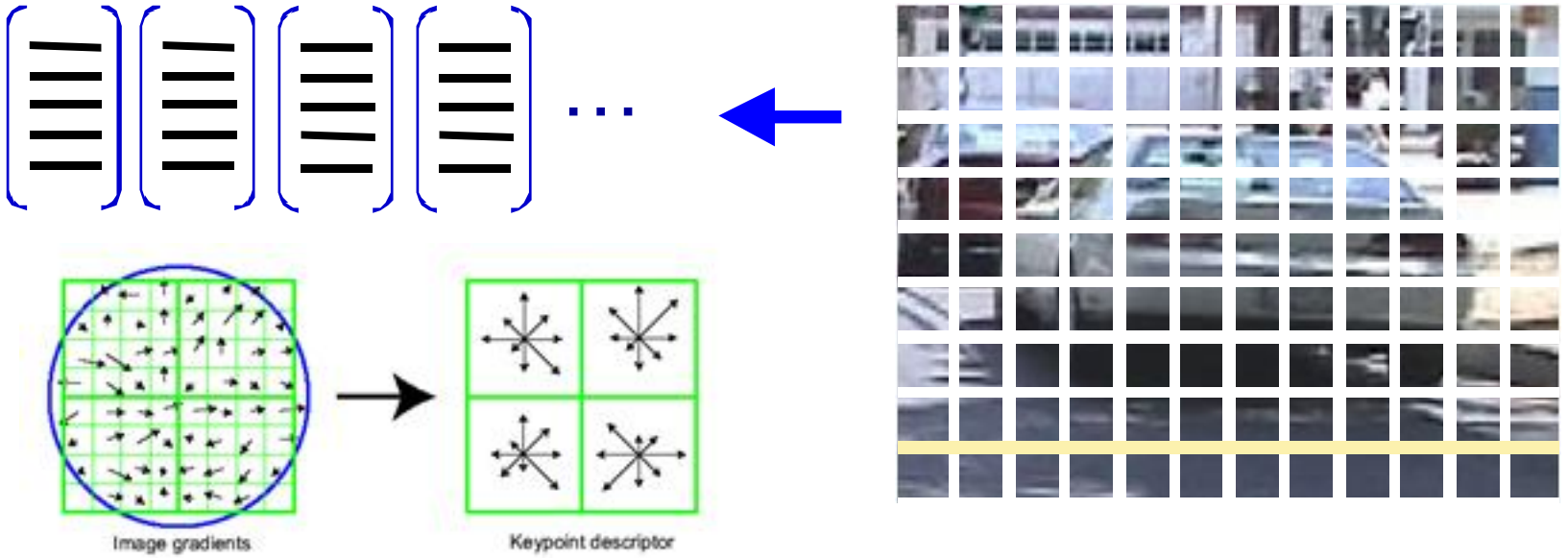
1. Извлечение фрагментов

- По регулярной сетке
  - Можем без перекрытия, можем с перекрытием
- Вокруг характерных точек
  - Точки, которые можно устойчиво идентифицировать на изображении после деформации (рассматриваем отдельно)



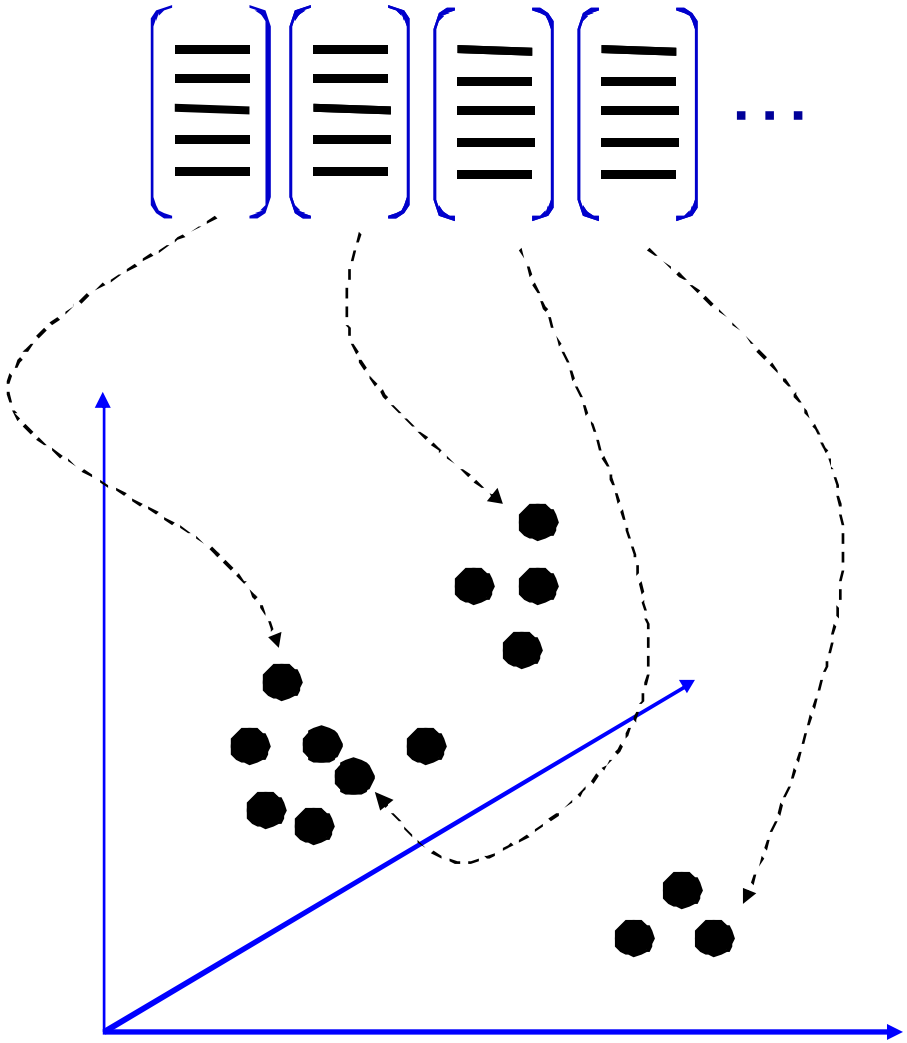


Вычисление признаков фрагментов

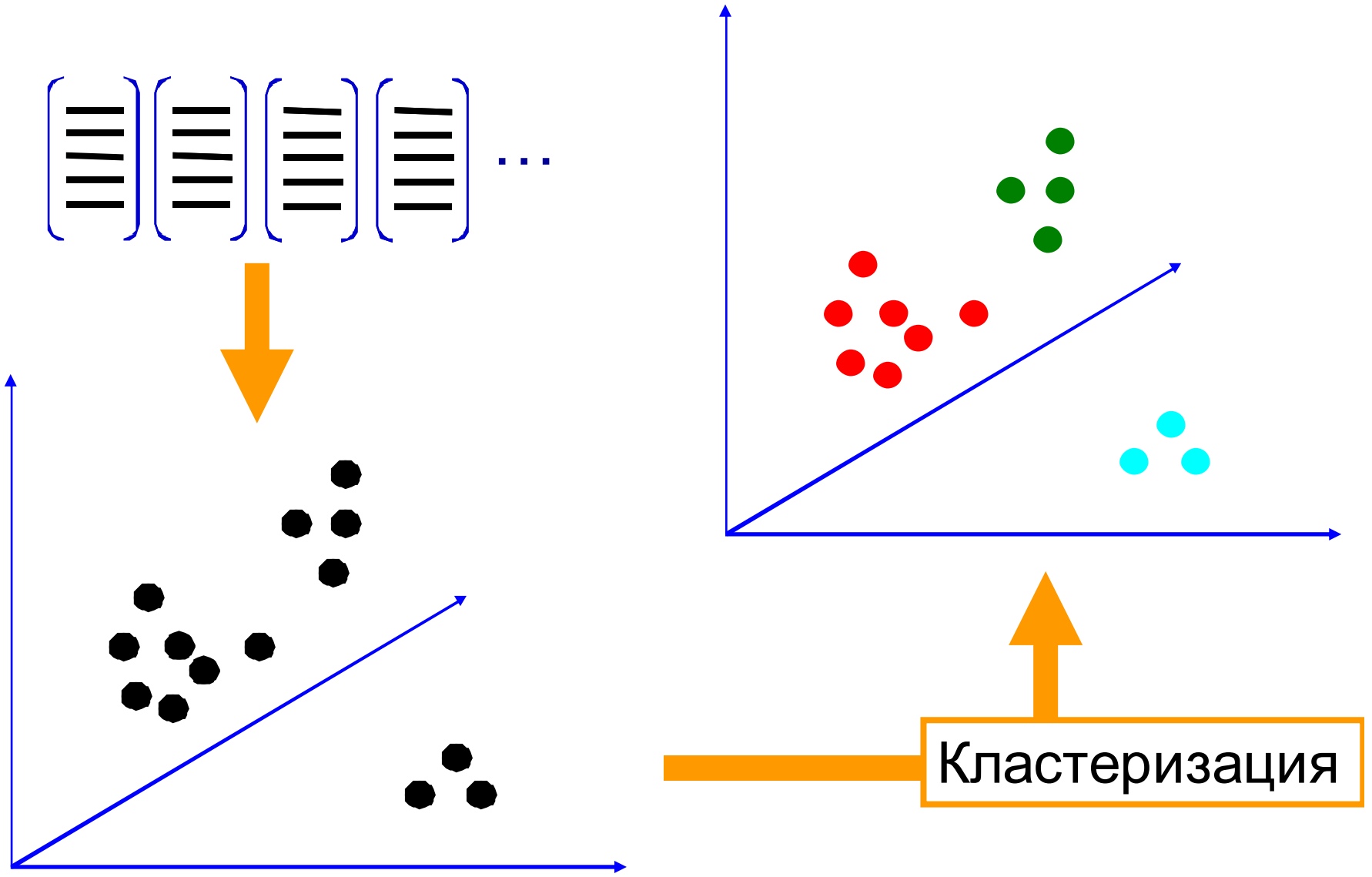


- Неупорядоченный список фрагментов из множества изображений
- Нужно описать каждый фрагмент вектор-признаком
- Чаще всего используют HOG-подобные признаки

2. Обучение словаря

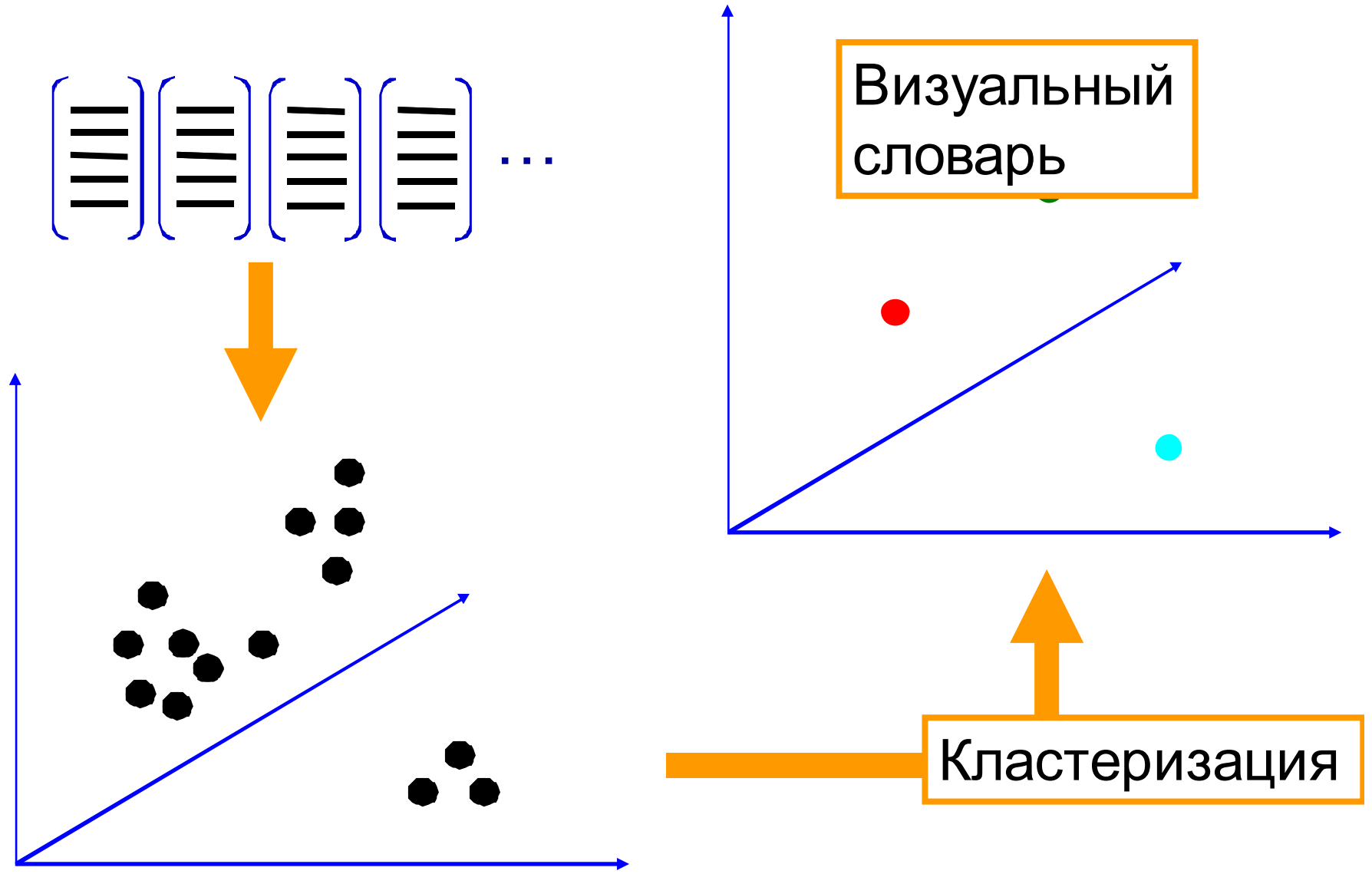


2. Обучение словаря



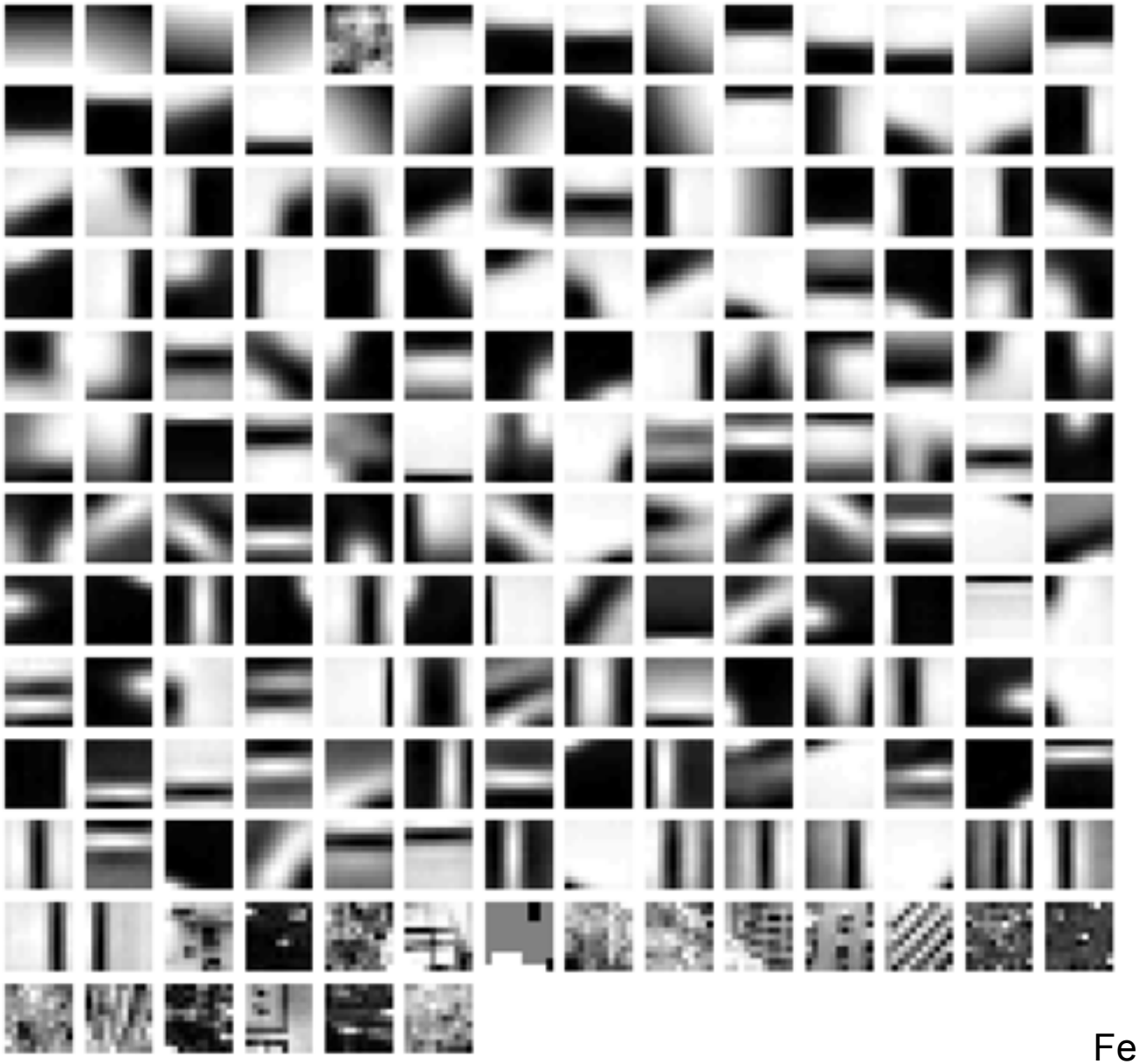
Slide credit: Josef Sivic

2. Обучение словаря



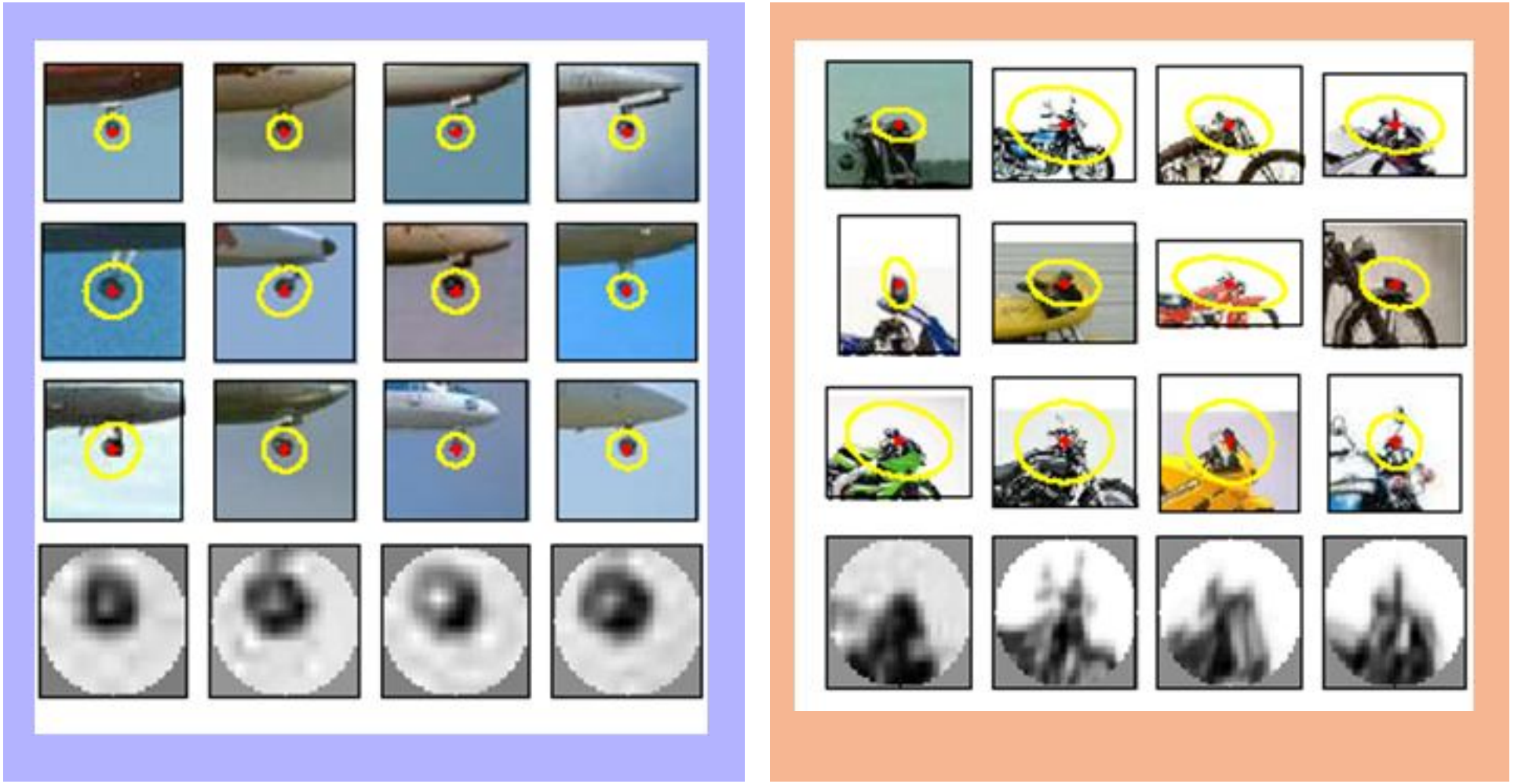
Slide credit: Josef Sivic

Пример словаря



Fei-Fei et al. 2005

Примеры визуальных слов

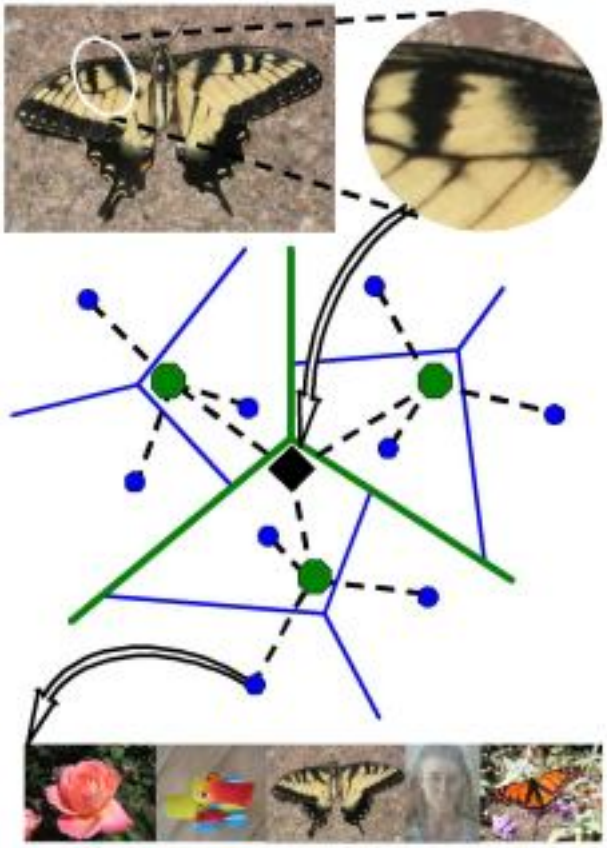


Построены по характерным точкам

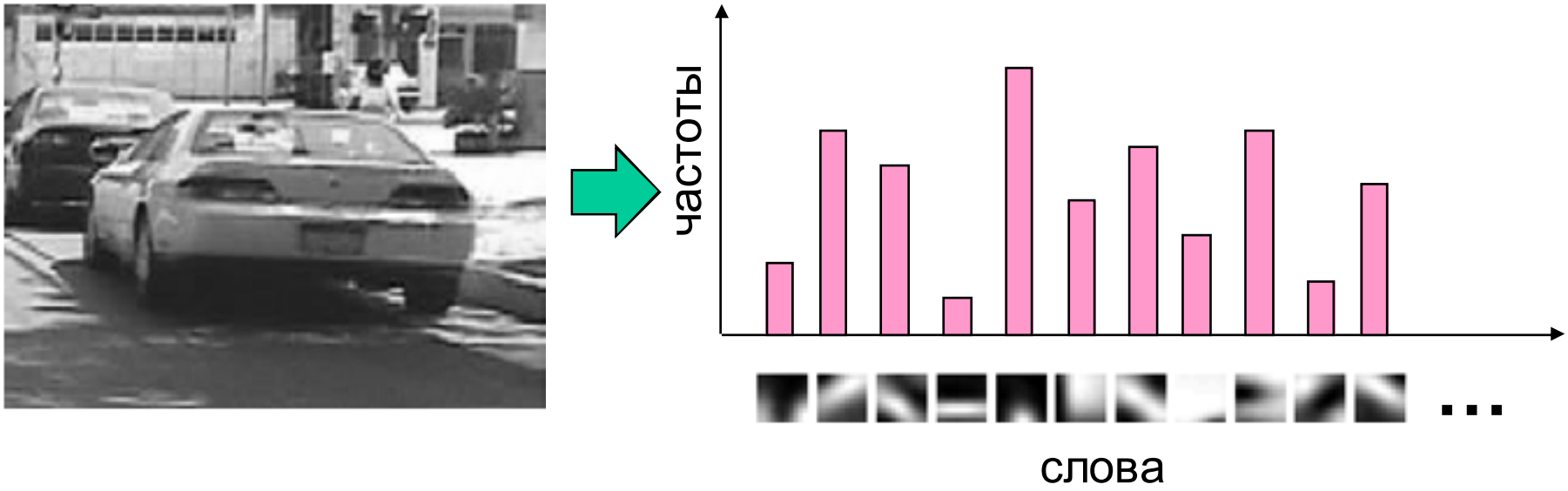
Sivic et al. 2005

Визуальные словари

- Как выбрать размер словаря?
  - Маленький: слова не могут описать все особенности
  - Большой: переобучение
- Вычислительная сложность
  - Деревья словарей (Nister & Stewenius, 2006)
  - Приближенные методы
  - Хеширование



3. Построение «мешка слов»



- Соберём тем же способом фрагменты для выбранного изображения
- «Квантуем» каждый фрагмент
  - Определим, какому слову из словаря он соответствует (фактически, сопоставим номер от 1 до N, где N – размер словаря)
- Построим гистограмму частот фрагментов для изображения

Резюме мешка слов

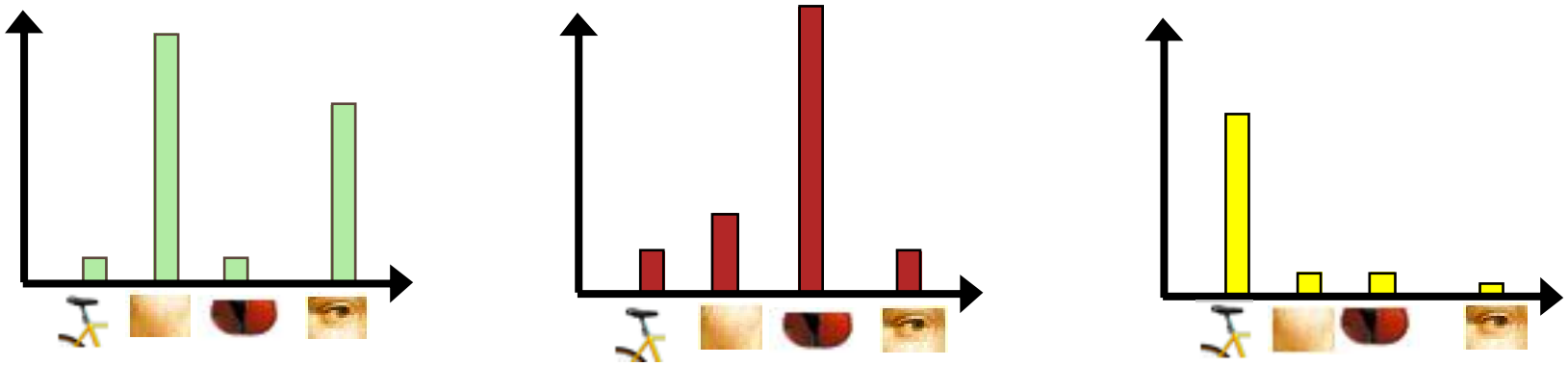
- Идея «визуальных слов» и «мешка слов» оказалась очень интересной и эффективной
- Мешок слов используется как сам по себе, так и совместно с другими признаками изображения
- Часто используются совместно и разреженная и плотная версии:
  - Мешок слов по характеристическим точкам
    - Bags of visual words
  - Плотные слова
    - Dense words (PhowGray, PhowColor)

- Реализация – библиотека VLFeat
  http://www.vlfeat.org/

Резюме признаков

- Построение вектор-признака изображение обычно разделяют на 3 этапа
  - Вычисление признаков пикселов (features)
  - Кодирование признаков (coding)
  - Пространственная агрегация (spatial pooling)
- Для кодирования и агрегации часто используют гистограммы
- Визуальные слова можно рассматривать как расширение обычной схемы, в которой мы рассматриваем окрестности пикселов (фрагменты, а не отдельные пикселы)
- Основные – гистограммы цветов, HOG, BoW признаки

Категоризация изображений



- Поскольку рассмотренные признаки основаны на гистограммах, то для классификации нам потребуется сравнивать гистограммы
  - Для нелинейных SVM строить ядра k(x1,x2) на гистограммах

Функции сравнения гистограмм

- Histogram intersection

(нормализованные гистограммы)

$$D(h_i, h_j) = 1 - \sum_{m=1}^{K} \min\left(h_i(m), h_j(m)\right)$$

- L1 distance

$$D(h_1, h_2) = \sum_{i=1}^{N} | h_1(i) - h_2(i) |$$

- χ^2 distance

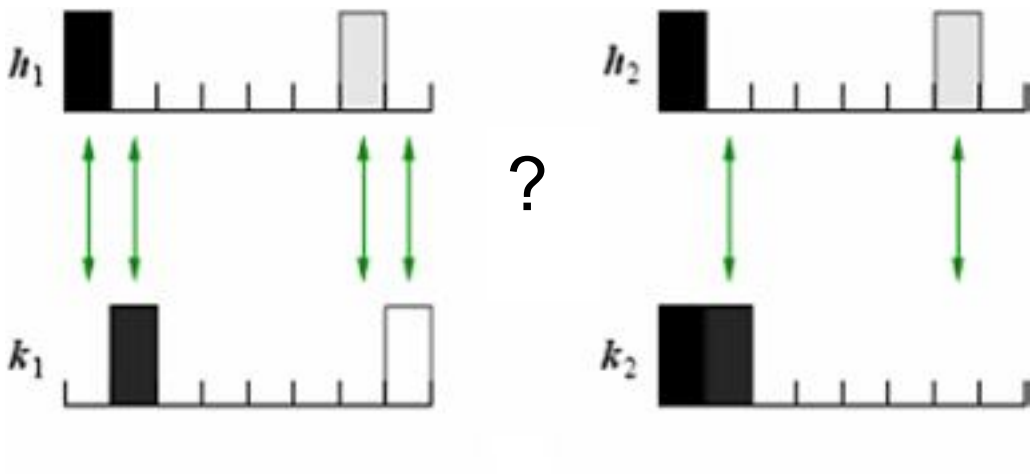
$$D(h_1, h_2) = \sum_{i=1}^{N} \frac{\left(h_1(i) - h_2(i)\right)^2}{h_1(i) + h_2(i)}$$

- Quadratic distance (*cross-bin*)

$$D(h_1, h_2) = \sum_{i,j} A_{ij} \left(h_1(i) - h_2(j)\right)^2$$

Jan Puzicha, Yossi Rubner, Carlo Tomasi, Joachim M. Buhmann: Empirical Evaluation of Dissimilarity Measures for Color and Texture. ICCV 1999

Проблемы сравнения гистограмм



- Все метрики сравнения гистограмм чувствительны к размеру ячеек
- Можно использовать более широкие ячейки, но при этом теряем «разрешающую способность»

Ядра для SVM

- Ядро пересечения гистограмм:

$$I(h_1, h_2) = \sum_{i=1}^{N} \min(h_1(i), h_2(i))$$

- Обобщенное Гауссово ядро:

$$K(h_1, h_2) = \exp\left(-\frac{1}{A} D(h_1, h_2)^2\right)$$

- *D* может быть евклидовым расстоянием, χ^2 , и т.д.

J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, Local Features and Kernels for Classifcation of Texture and Object Categories: A Comprehensive Study, IJCV 2007

Эталонные коллекции

PASCAL
Pattern Analysis, Statistical Modelling and Computational Learning

http://www.vision.caltech.edu/Image_Datasets/Caltech101

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

Как сейчас получают эталонные коллекции?

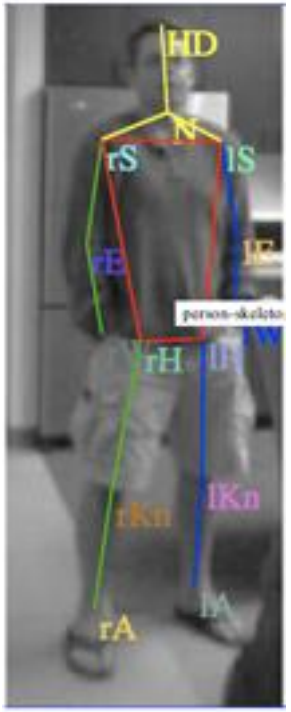
1. Берут готовые
2. Размечают вручную
3. Сопоставление имеющихся коллекций для получения доп. разметки

Mechanical Turk (1770)

- Automaton Chess Player – робот, игравший в шахматы
  - Автоматон двигает фигуры, говорит «Чек» и обыгрывает всех!
- С 1770 по 1854 развлекал публику, только в 1820 году раскрыли обман

Human Intelligence Task

- Есть много задач, простых для человека, но крайне сложных для компьютера
- Пример: классификация и разметка изображений



Galaxy Zoo

GALAXY ZOO http://www.galaxyzoo.org/

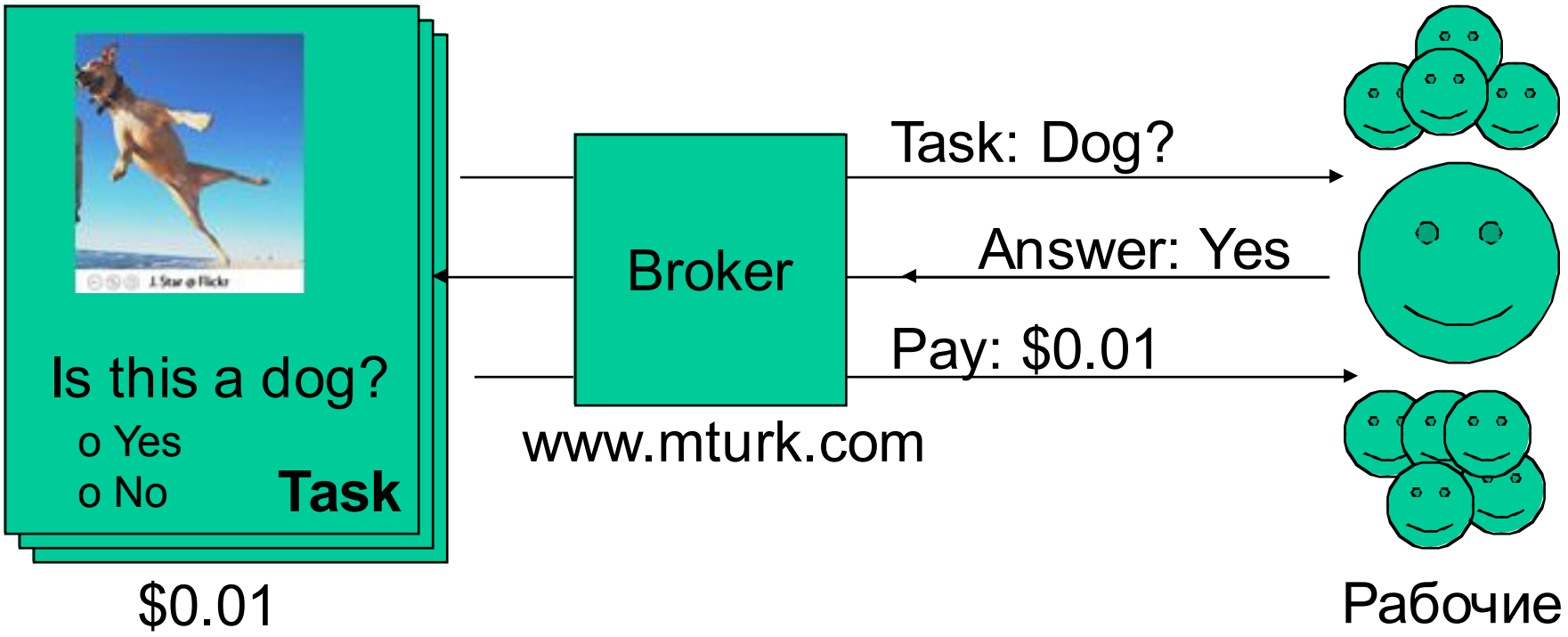
- Классификация изображений галактик
- Первый масштабный проект такого рода
- Более 150000 волонтёров за первый год бесплатно сделали более 60 млн. меток

Система LabelMe

- Просим пользователей выделять любые объекты замкнутыми ломаными и давать им название

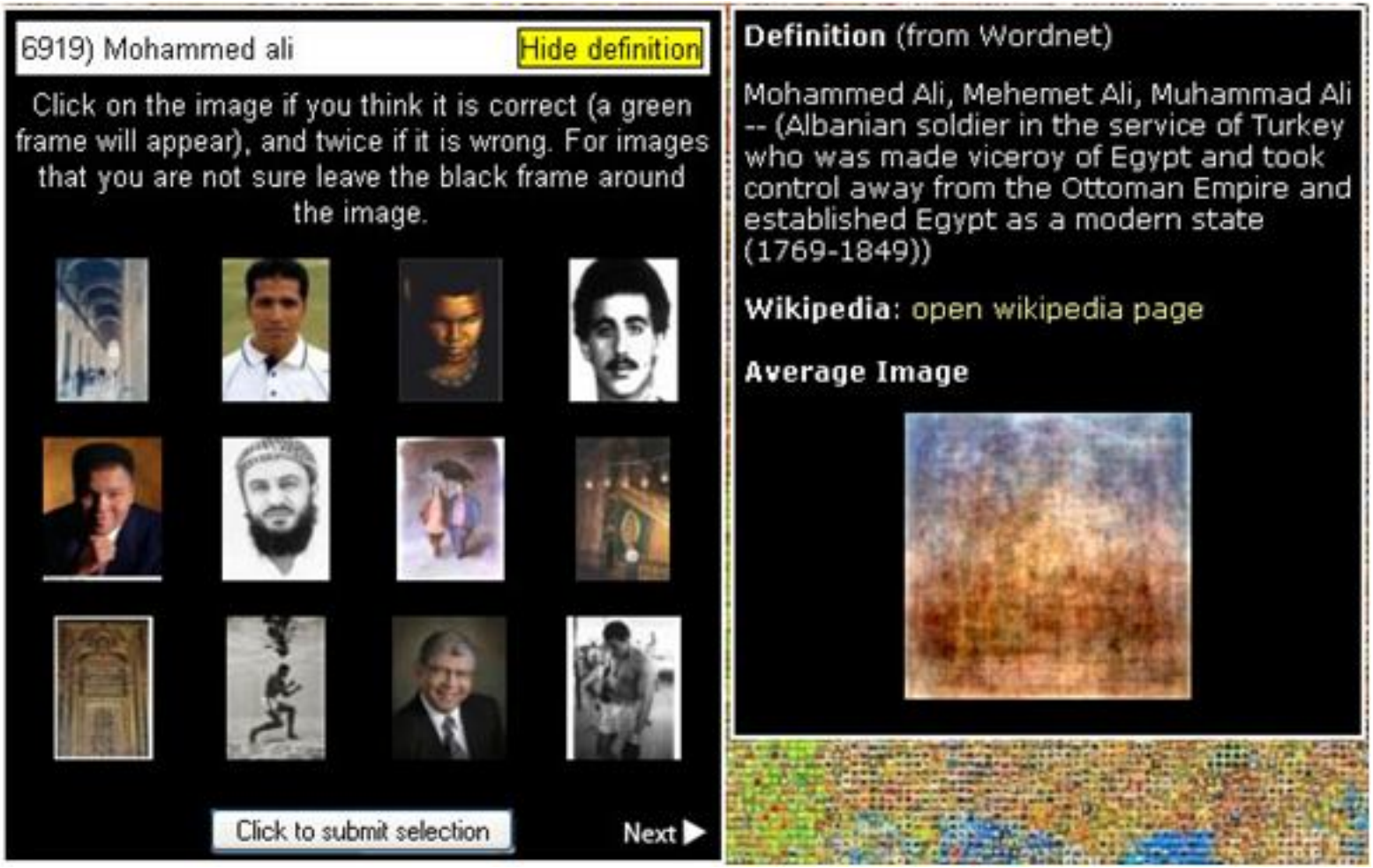
B. Russell, A. Torralba, K. Murphy, W. T. Freeman LabelMe: a database and web-based tool for image annotation, IJCV, 2008

Фрилансеры



• Возникли интернет-брокеры для сведения заказчиков и работников
• Пример: Amazon Mechanical Turk, Яндекс Толока
• Разметчики – просто подработка (студенты, домохозяйки), люди из более бедных стран

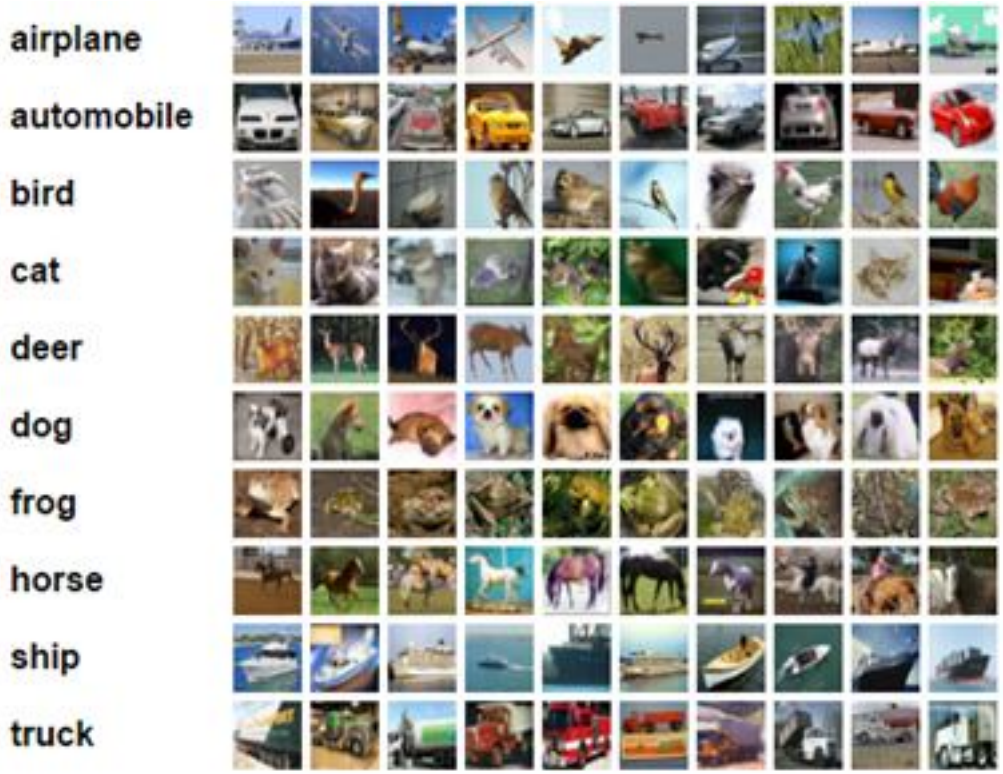
Tiny images (80 млн.)



http://people.csail.mit.edu/torralba/tinyimages/

CIFAR-10 и CIFAR-100

Выборки из TinyPictures



- CIFAR-10
  - 10 классов
  - 60000 изображений
  - 5000 обучающих и 1000 тестовых на класс
- CIFAR-100
  - 100 классов
  - 60000 изображений
  - 500 обучающих и 100 тестовых на класс

http://www.cs.toronto.edu/~kriz/cifar.htm

Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009.

IMAGENET

Цель: собрать коллекцию с 1000 изображений на каждый из 117 тыс. synsets.

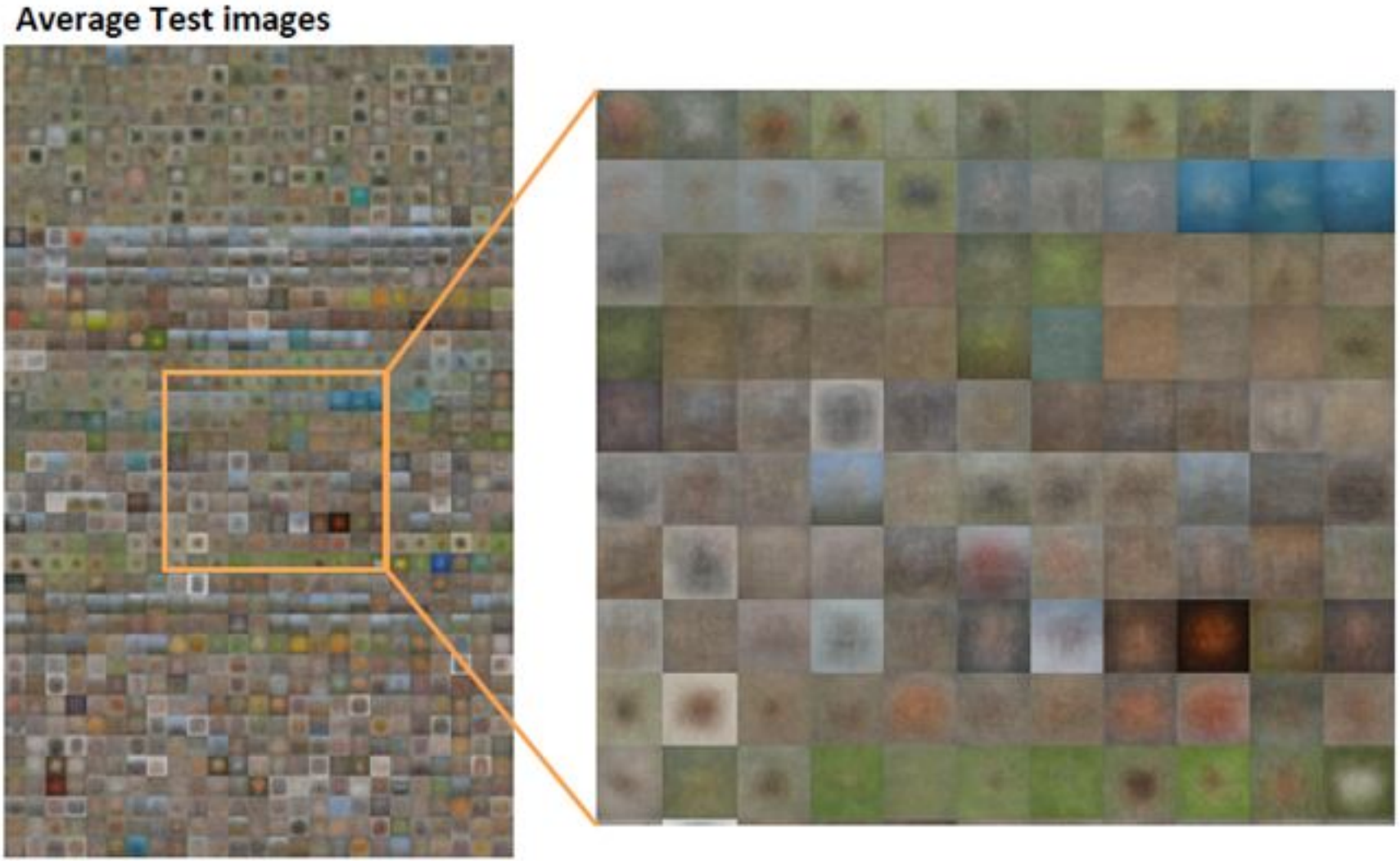
Текущая статистика:

- Total number of non-empty synsets: 21841
- Total number of images: 14,197,122
- Number of images with bounding box annotations: 1,034,908
- Number of synsets with SIFT features: 1000
- Number of images with SIFT features: 1.2 million

Конкурс large-scale image recognition

http://www.image-net.org

Изображения в среднем



Резюме лекции

- Категории – упрощение задачи понимания мира
- Рассмотрели некоторые факты о распознавании человек
- Стандартная схема построения признаков из features, coding, spatial pooling
- В качестве признаков чаще всего используют гистограммы распределений цвета, градиентов, краёв, визуальных слов
- Любой метод классификации на основе машинного обучения можно использовать
- Есть ряд хороших эталонных коллекциях, но часто приходится размечать вручную