«Интернет-зрение»

Антон Конушин

Many slides adapted from Alexei Efros and Jean-François Lalonde

Сколько изображений нужно?

- Всего около ~100000 разных категорий объектов
- Сколько нам нужно изображений, чтобы научиться их всех распознавать?
- Сколько нам доступно изображений?
- Что мы про них будем знать?

Сколько всего картинок?

Число картинок на диске: **10^4**

Число картинок, виденных за 10 лет: **10^8**

(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Число картинок,
виденных всем человечеством: **10^{20}**

106,456,367,669 humans[1] * 60 years * 3 images/second * 60 * 60 * 16 * 365 =

1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Число картинок во вселенной: **10^{243}**

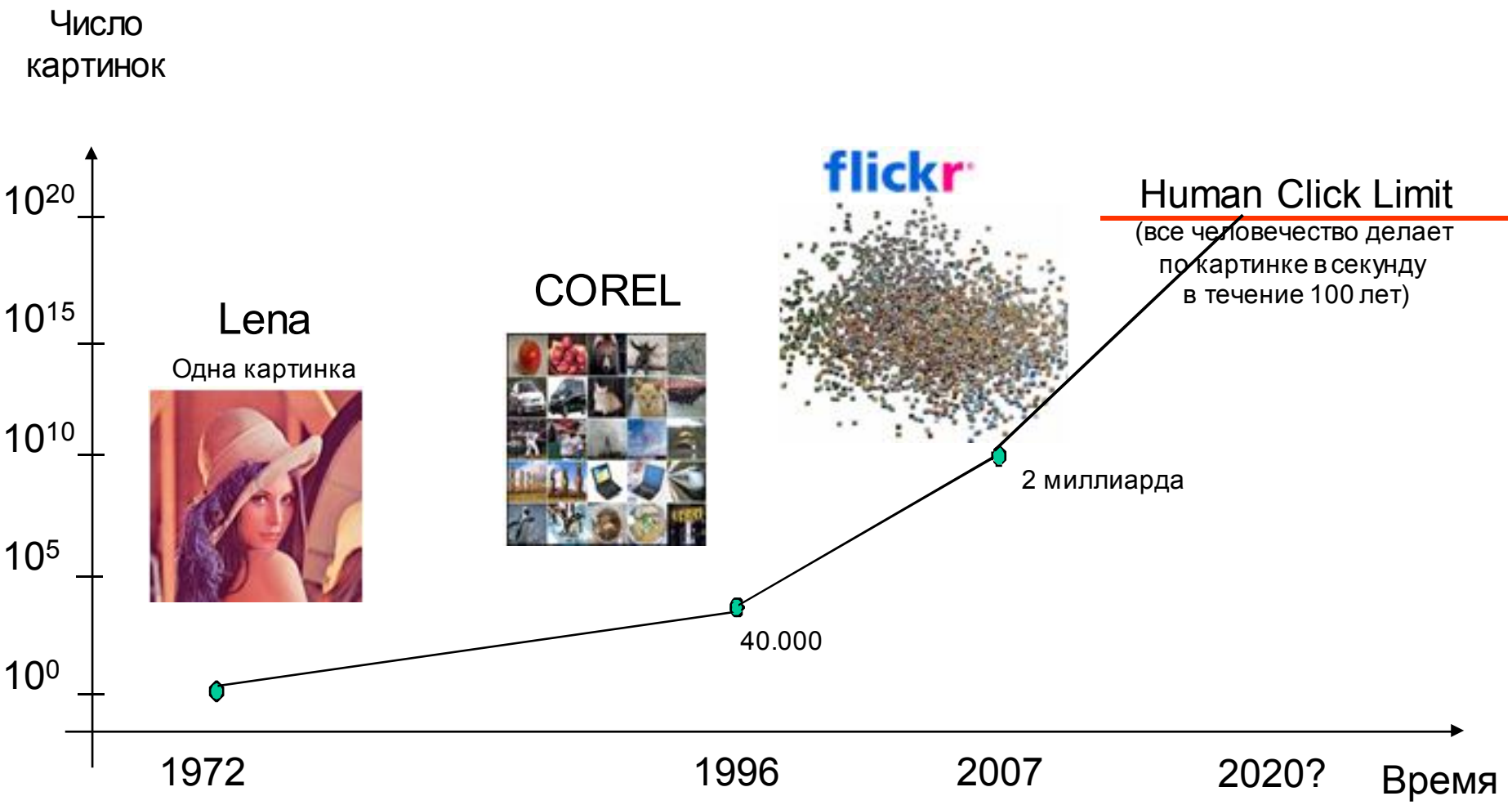
10^{81} atoms * 10^{81} * 10^{81} =

Число всех картинок 32x32 : **10^{7373}**

$256^{32*32*3} \sim 10^{7373}$

Slide by Antonio Torralba

Доступные изображения



Обычно, про само изображение мы ничего не знаем, кроме параметров в EXIF из JPEG

Slide by Antonio Torralba

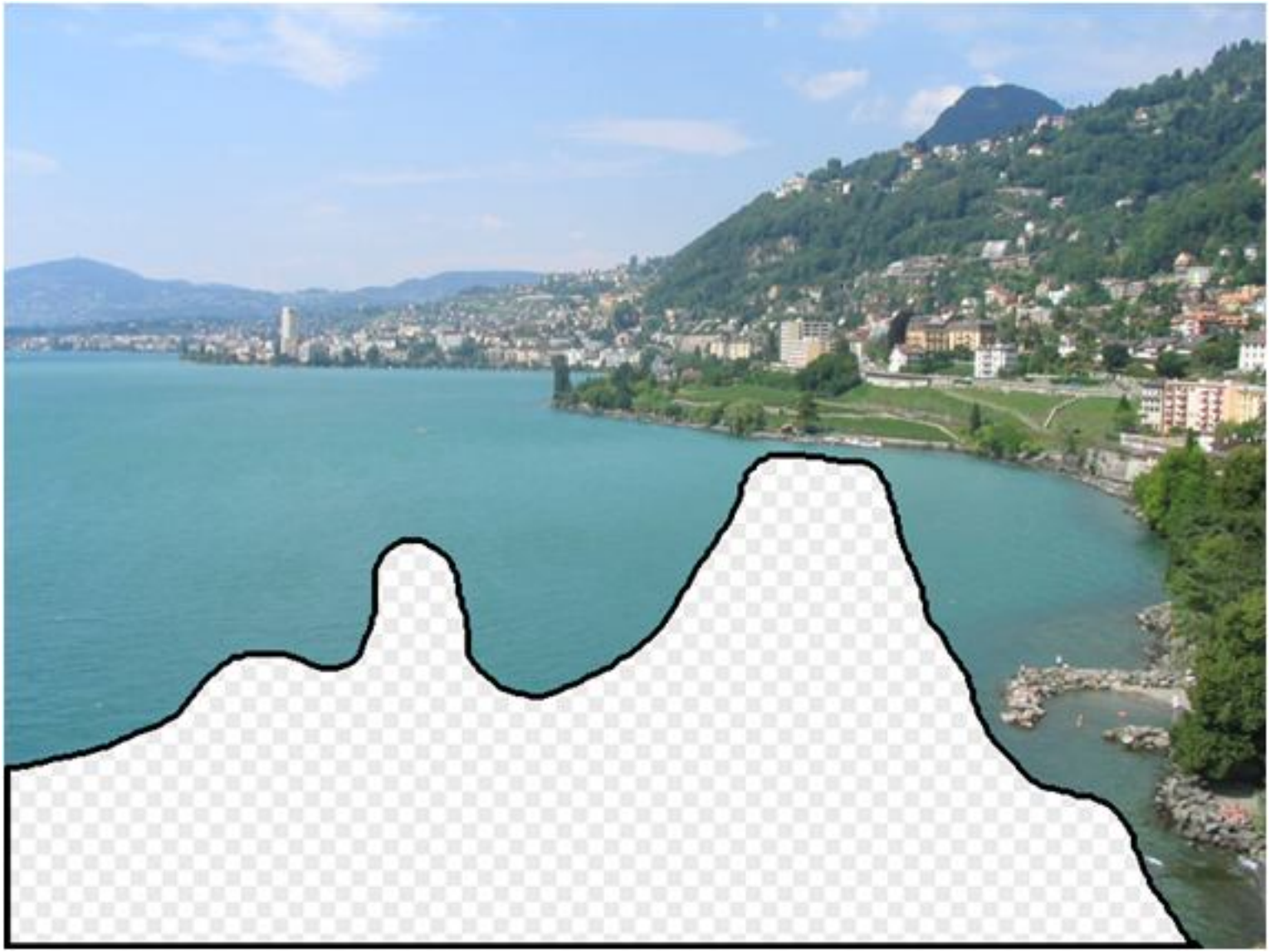
Что будем делать?

- Благодаря интернету, мы получили доступ к огромным коллекциям изображений
- Что мы с ними можем сделать, как они нам помогут?

Попробуем решить такую задачку:

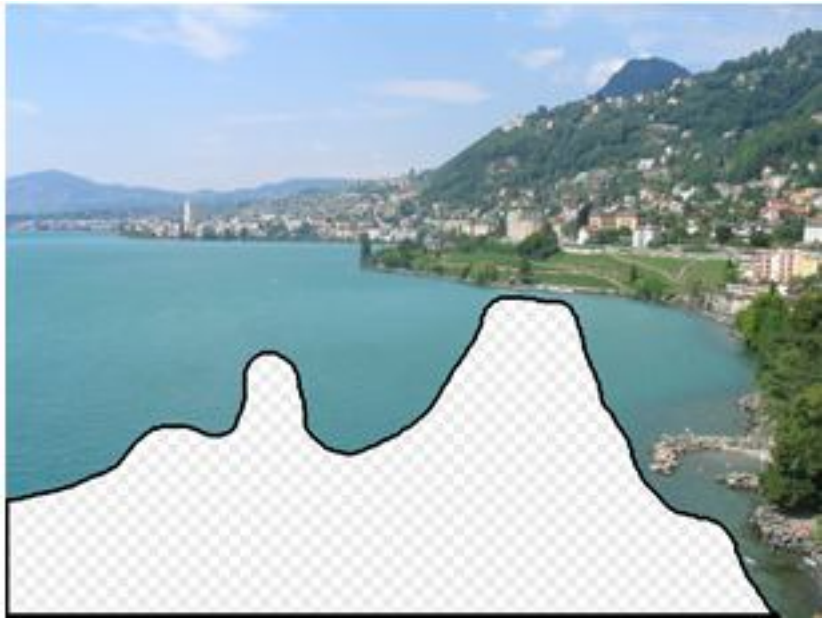
James Hays, Alexei A. Efros Scene Completion using Millions of Photographs, *SIGGRAPH 2007*

«Inpainting»



Реконструкция изображения в невидимой области

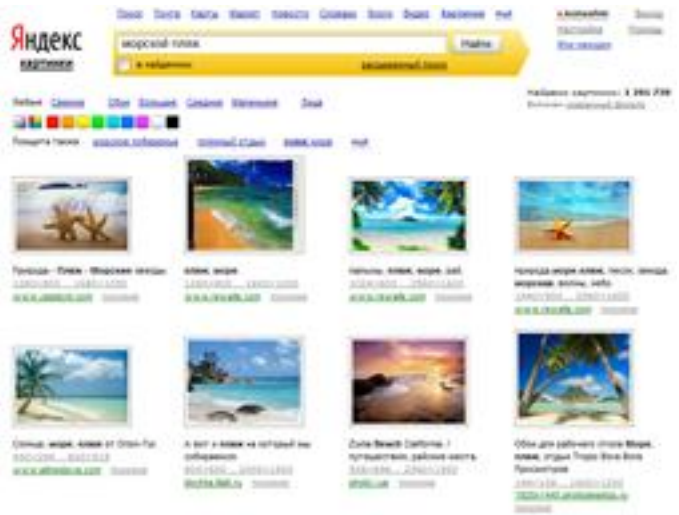
Как будем решать?



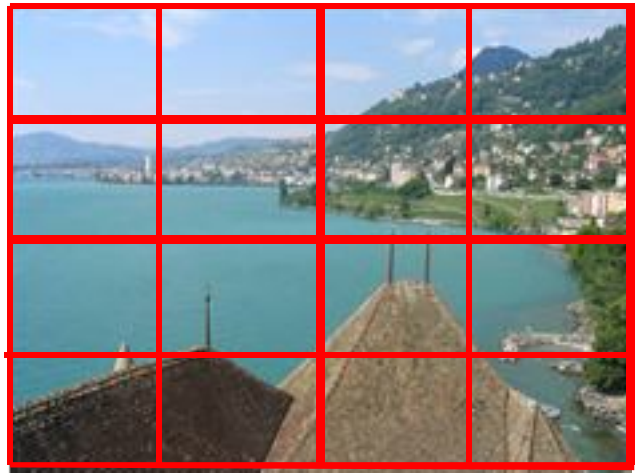
- Если бы мы нашли в интернете подходящую картинку, мы бы могли просто ей просто «закрыть дырку»

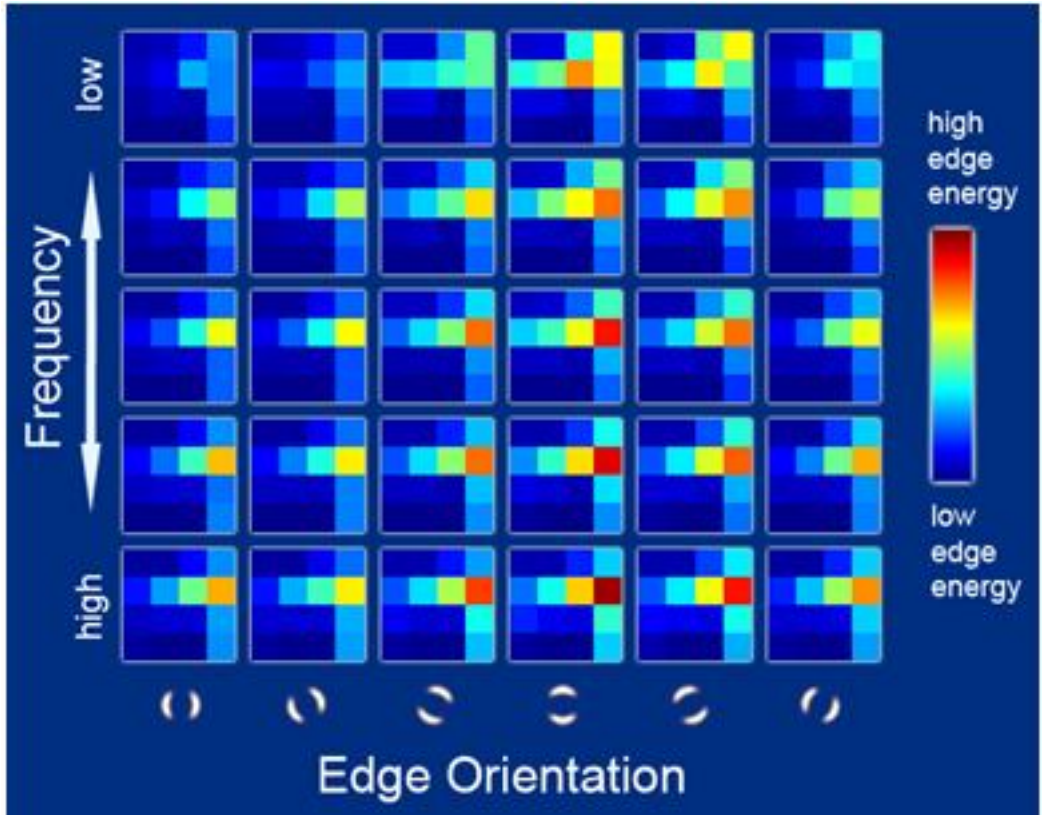
Как будем решать?

- Скачаем из интернета множество пейзажей по разным ключевым словам
- Очень много – миллион и больше картинок
- Будем искать похожие картинки в этой базе
- Воспользуемся одним из методов поиска изображений по содержанию



Дескриптор изображения GIST

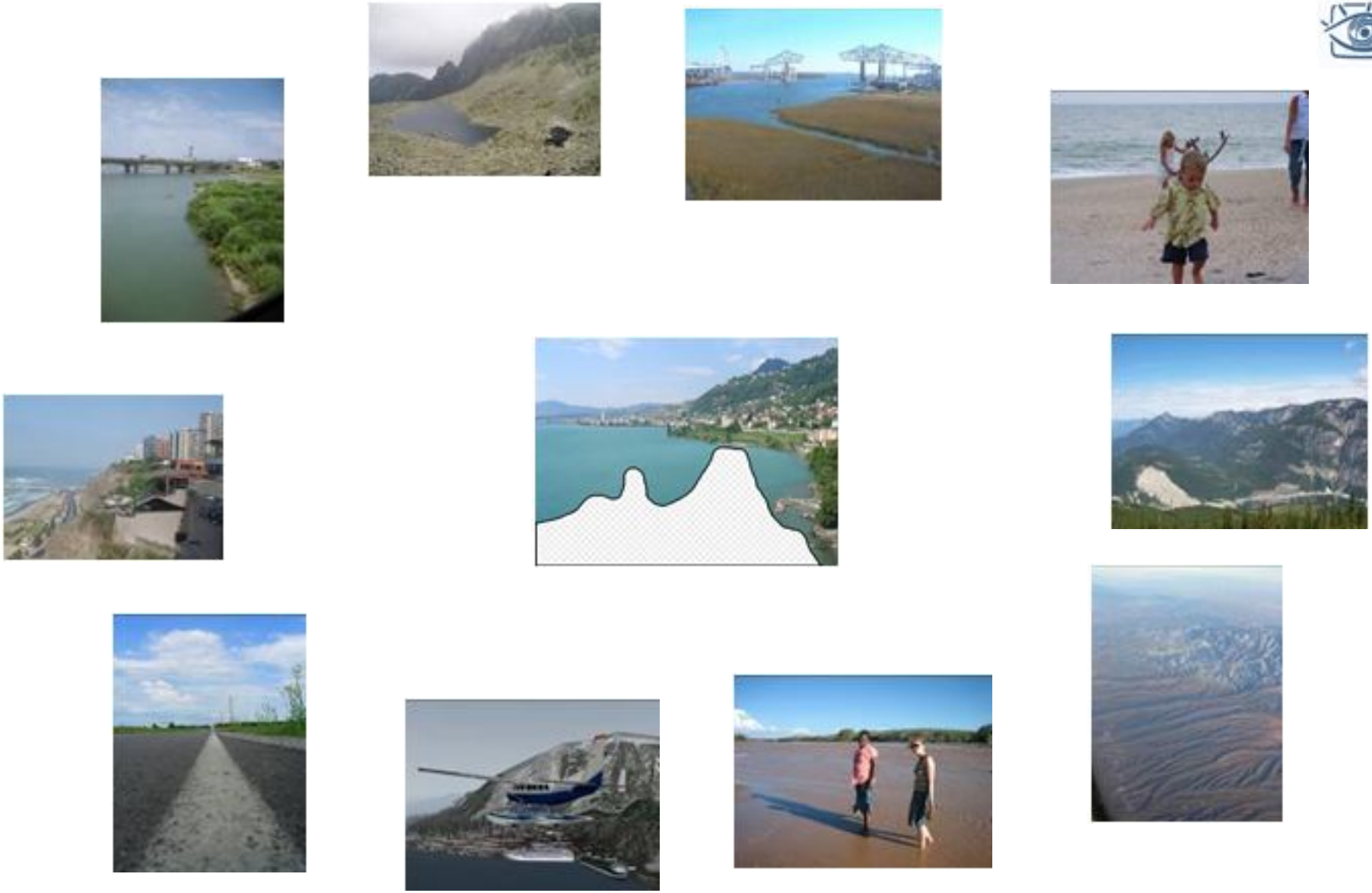




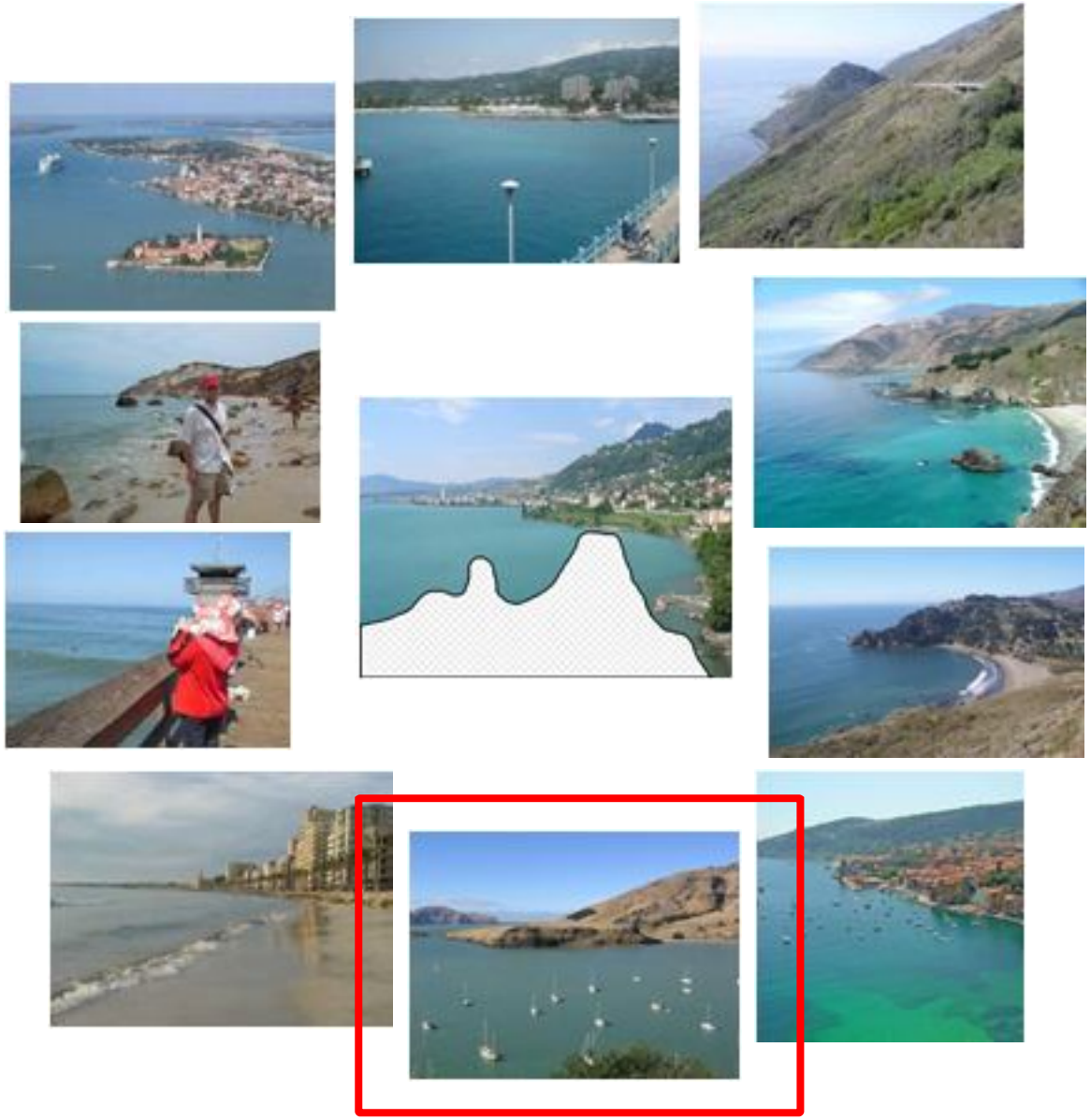
- Посчитаем отклики детекторов краёв на 5 разных масштабах и 6 ориентациях края
- Получим 33 «канала» - RGB-цвет и 30 откликов фильтров края
- Разобьём изображение сеткой 4x4 на 16 ячеек
- В каждой ячейке усредним значения всех каналов

TORRALBA, A., MURPHY, K. P., FREEMAN, W. T., AND RUBIN. Context-based vision system for place and object recognition. In ICCV 2003

Изображение-запрос

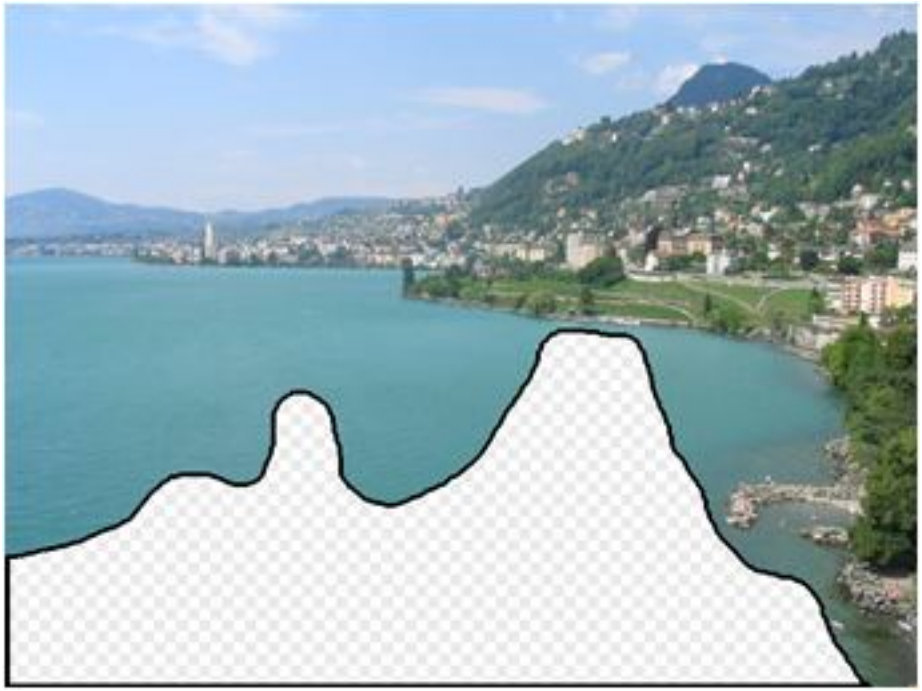


10 ближайших из 20,000 изображений



10 ближайших из 2х миллионов изображений

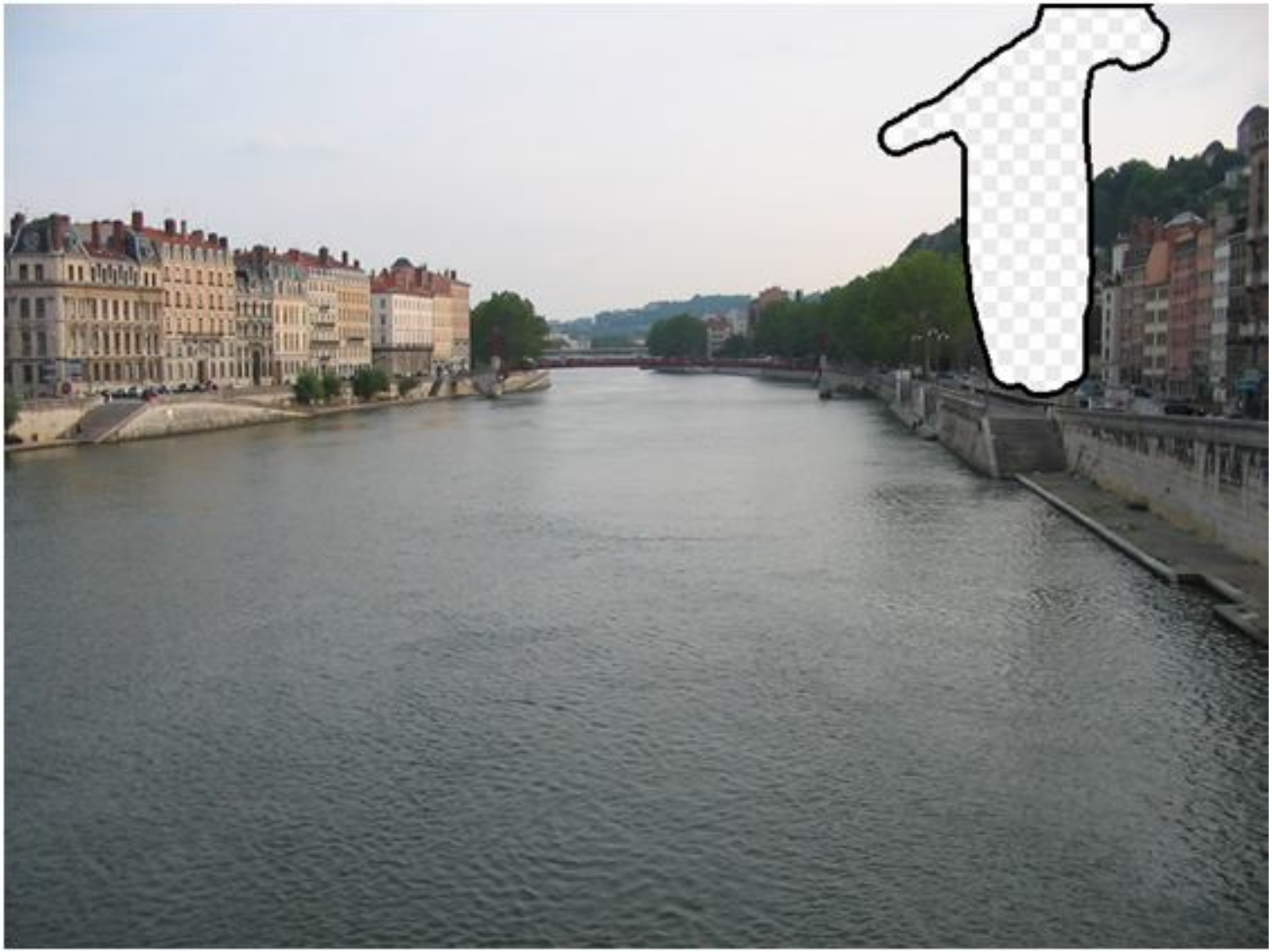
Закрываем дырку



Graph cut + Poisson blending

Пример

Пример



Пример

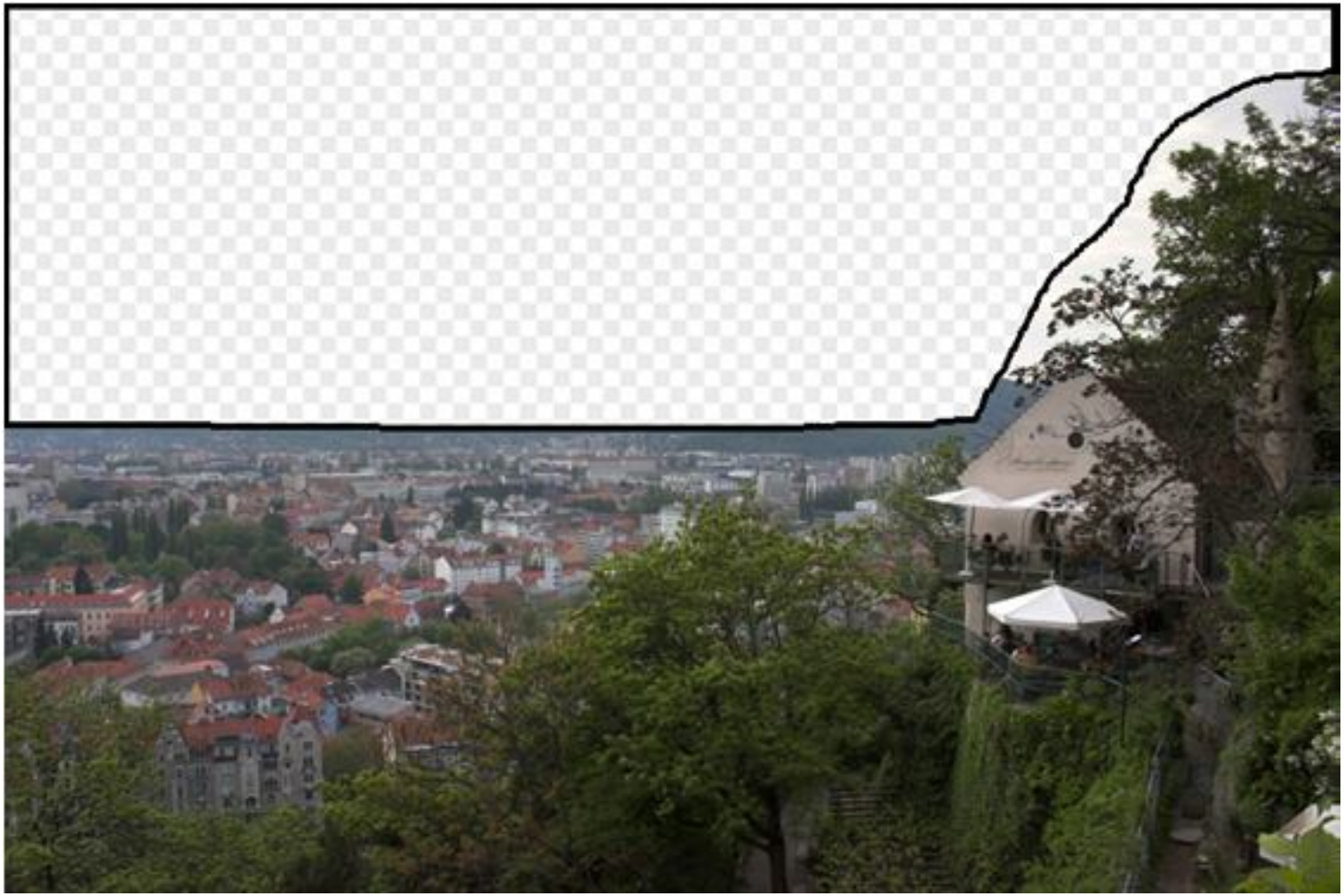
Пример

… 200 ближайших

Пример

Пример

Пример

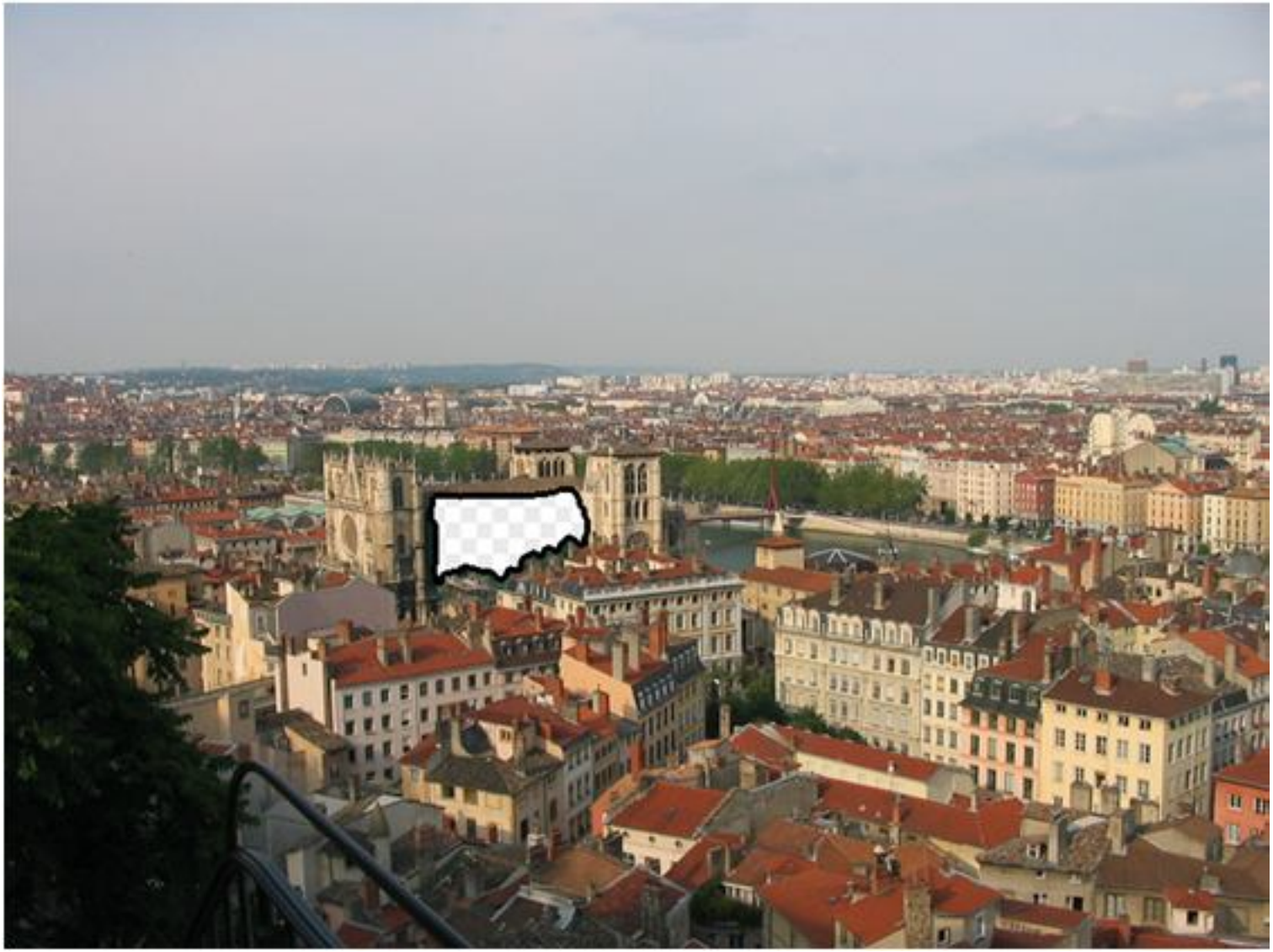


Пример

Пример

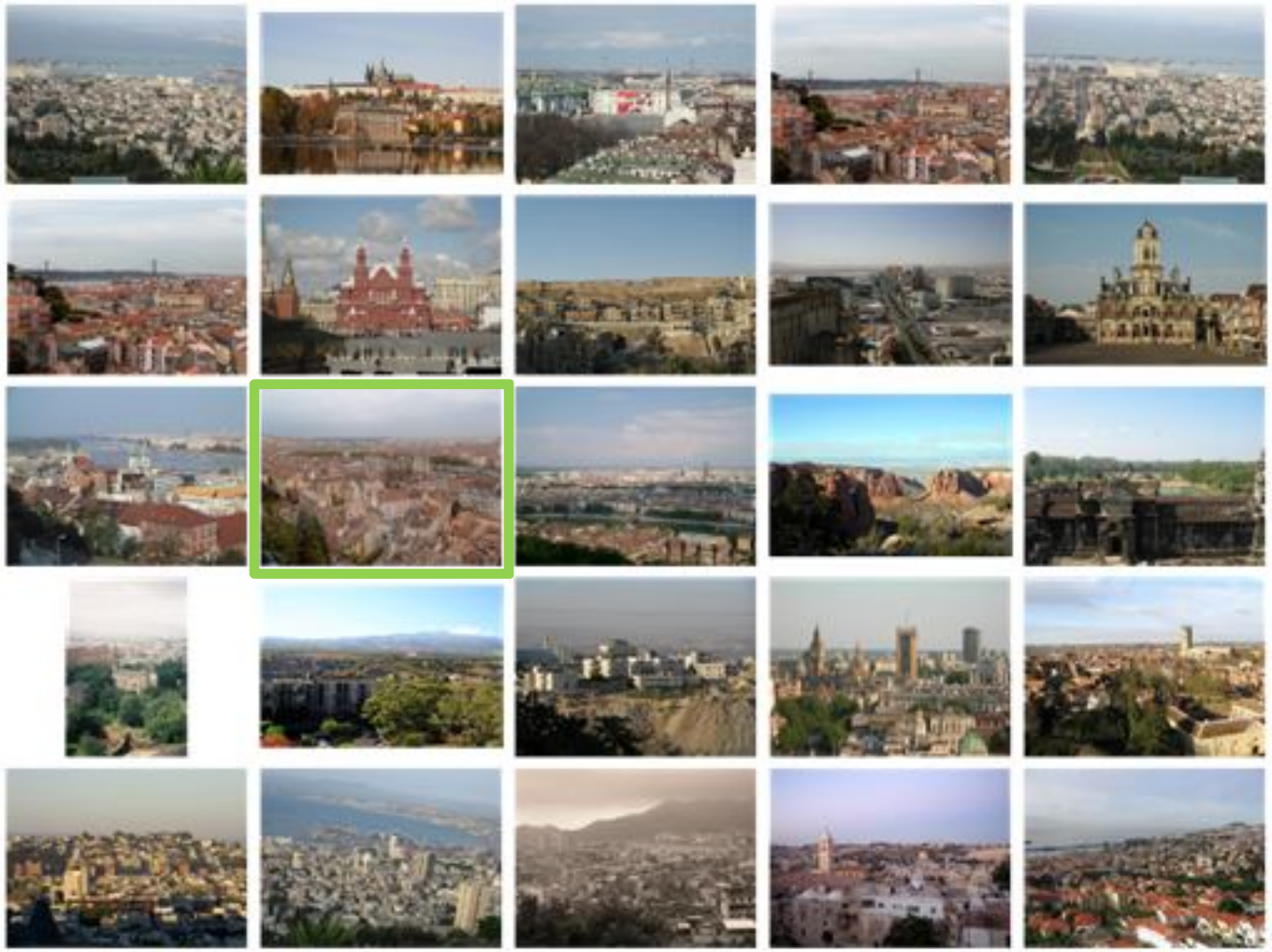
Пример

Пример



Пример

Пример



… 200 ближайших

Изображение с того же города!

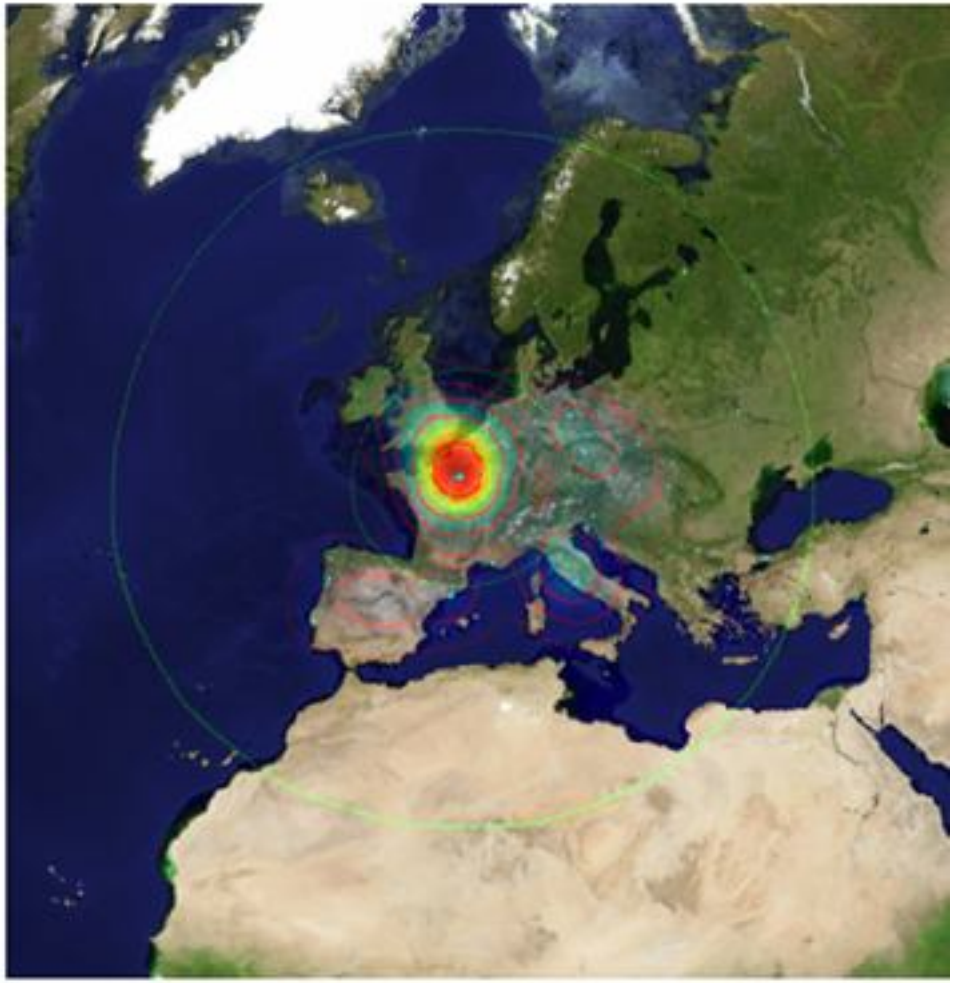
im2gps

- Собрали 6M картинок из Flikr с проставленными GPS-метками
- К дескриптору GIST добавили кучу дополнительной информации, требующей долгого обсчёта
- Кластер из 400 машин для аннотации всех 6M изображений
- Теперь по комбинированному дескриптору можем искать
- Найдем, вот эту картинку:

James Hays, Alexei A. Efros **im2gps: estimating geographic information from a single image**, *CVPR 2008*.

Paris
Paris
Paris
Paris
Paris
Paris
Paris
Madrid
Rome
Paris
Cuba
Paris
Paris
Poland
Paris
Paris

Отображение результатов на карте



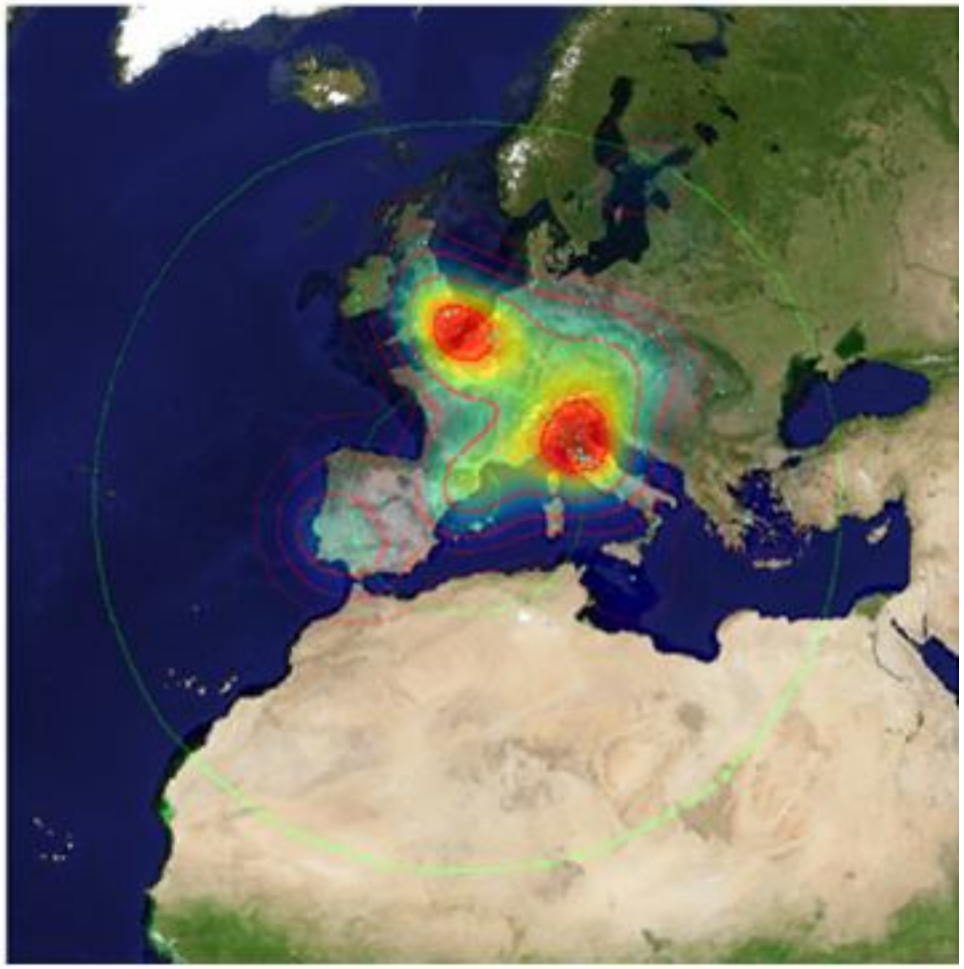
Найдем теперь такую:

Что мы нашли:

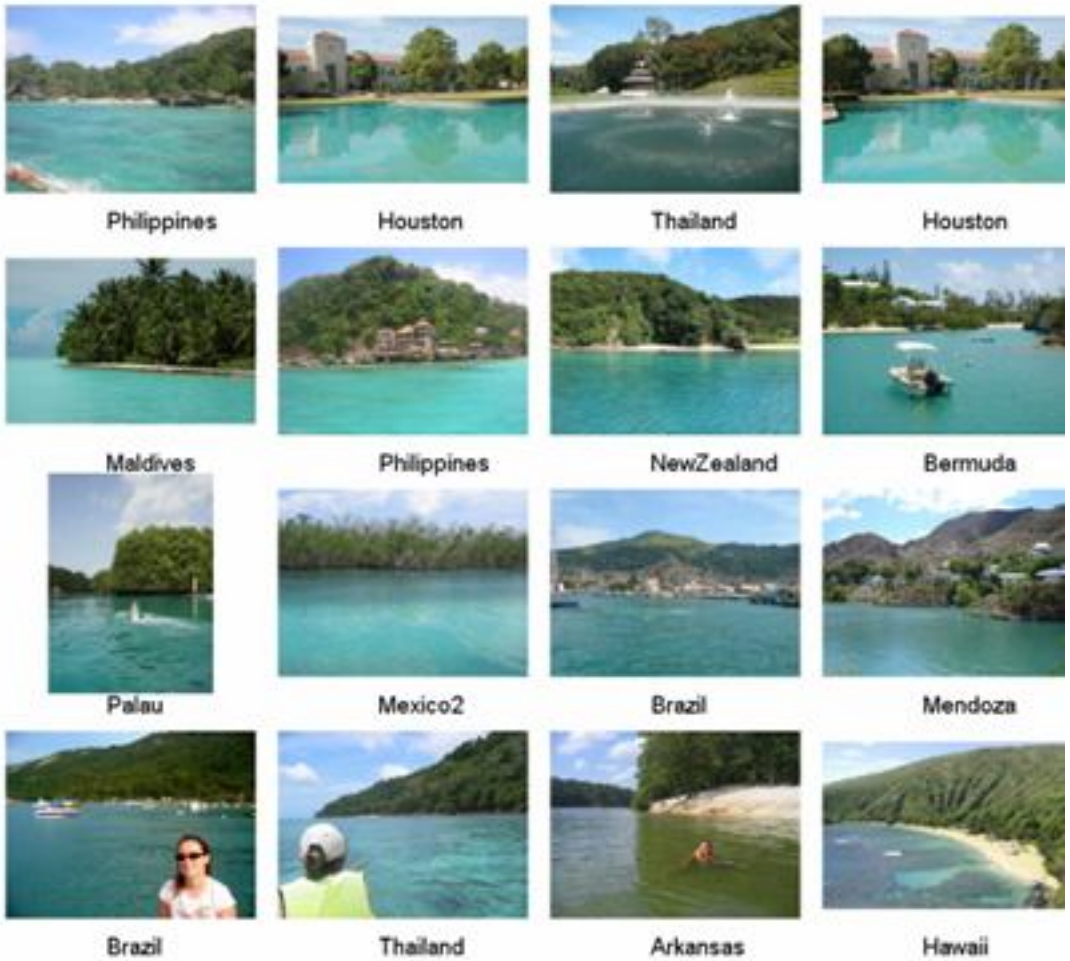
Madrid
england
France
Paris
Croatia
heidelberg
Macau
Malta
Cairo
Italy
Italy
Italy
Latvia
europe
Barcelona
Austria

Отображение результатов на карте

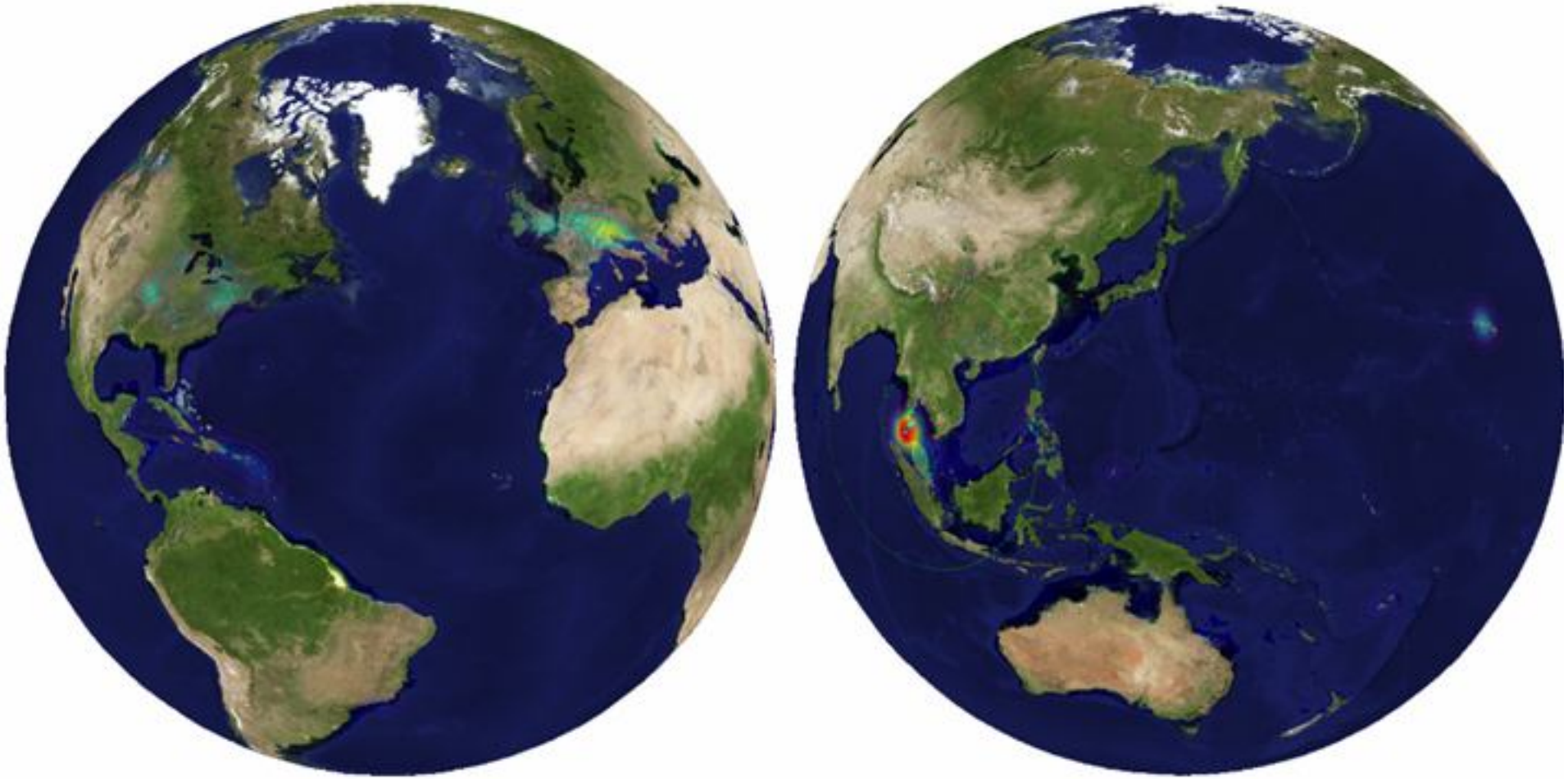
200 результатов, кластеризуем и покажем центры
и распределение картинок



Пример


Philippines
Houston
Thailand
Houston
Maldives
Philippines
NewZealand
Bermuda
Palau
Mexico2
Brazil
Mendoza
Brazil
Thailand
Arkansas
Hawaii

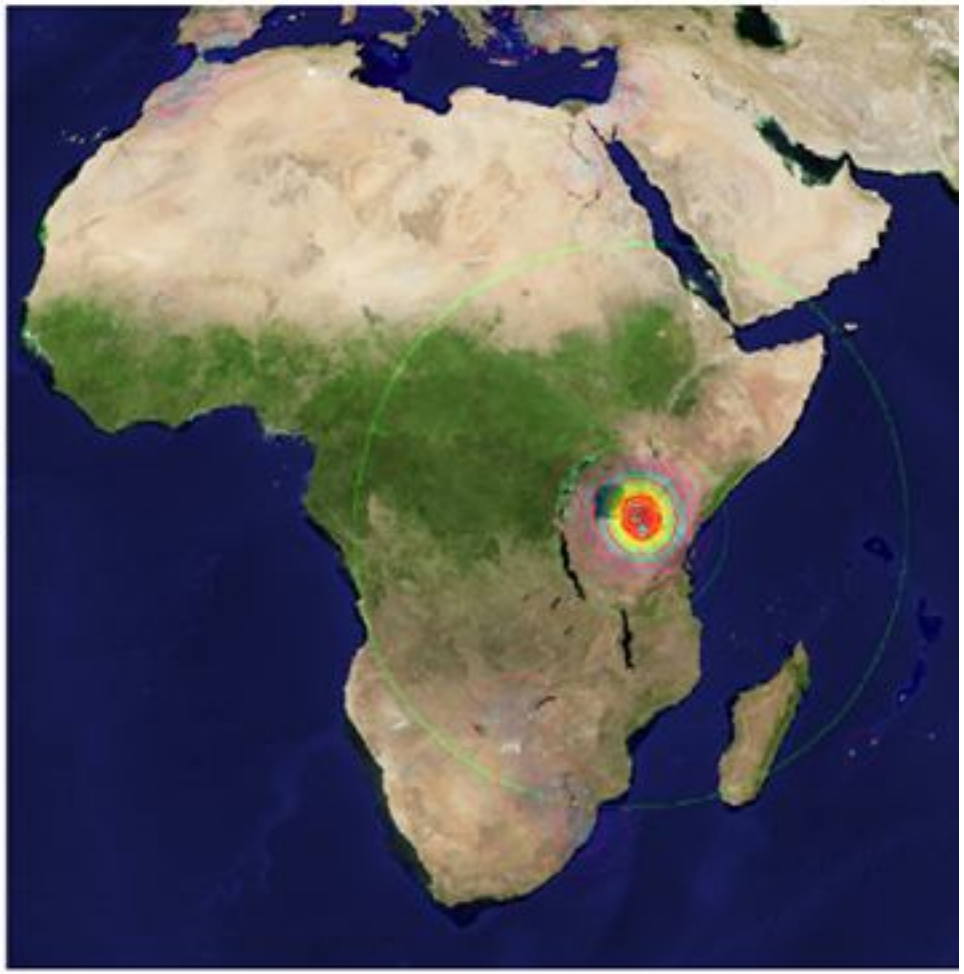
Пример



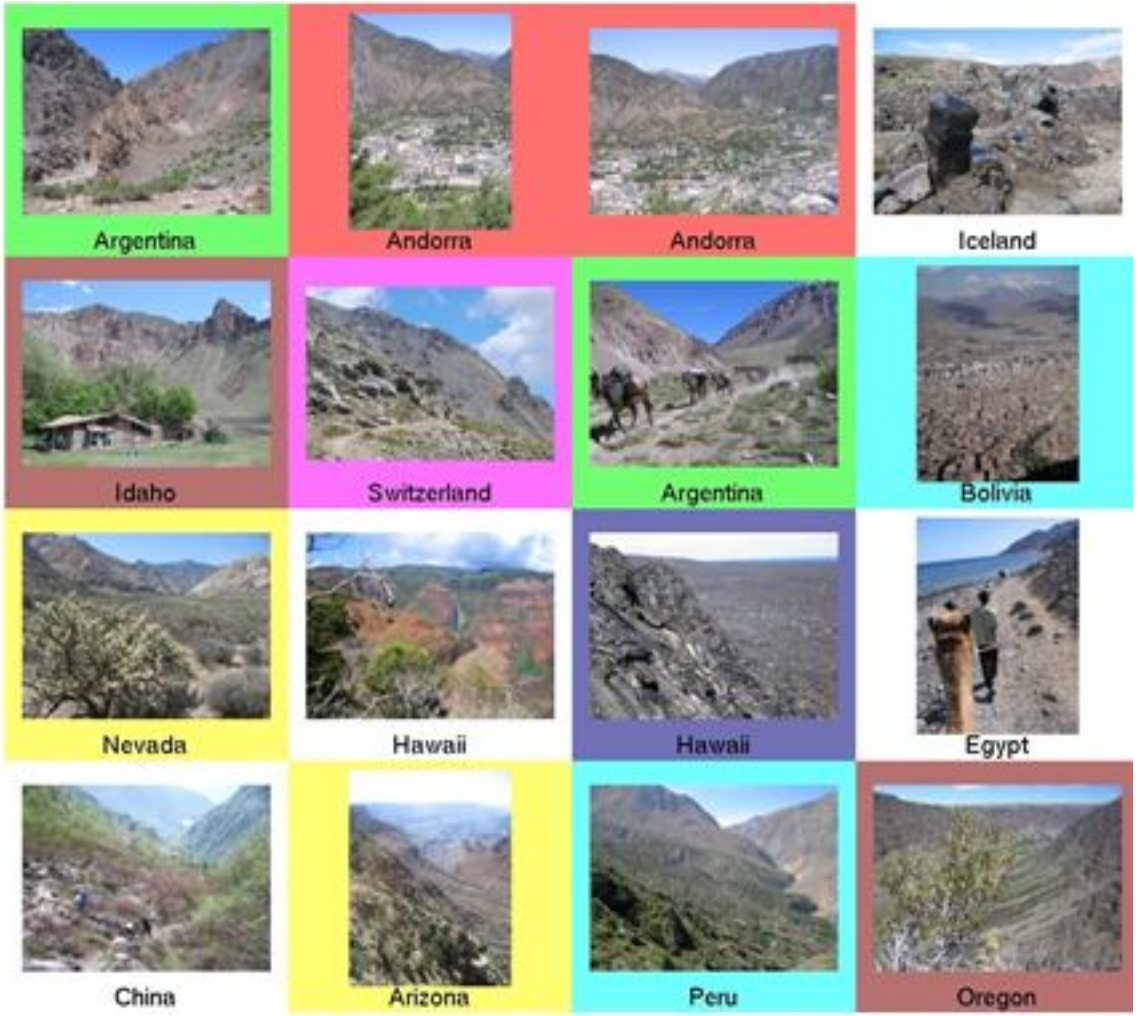
Пример

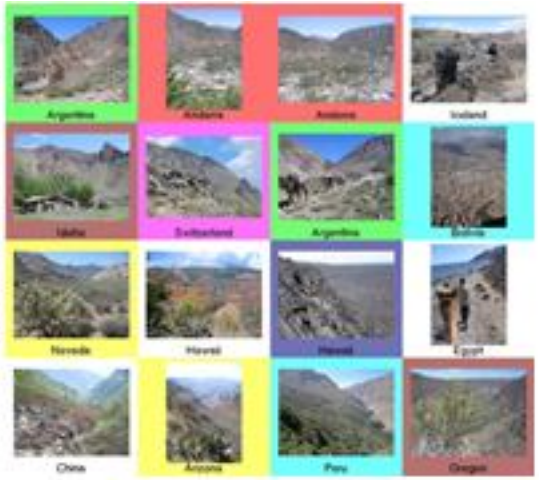
California
Oklahoma
SouthAfrica
Zambia
Kenya
Hyderabad
Mongolia
SouthAfrica
Kenya
Kenya
Zambia
Ethiopia
Nevada
africa
Morocco
Tennessee

Пример

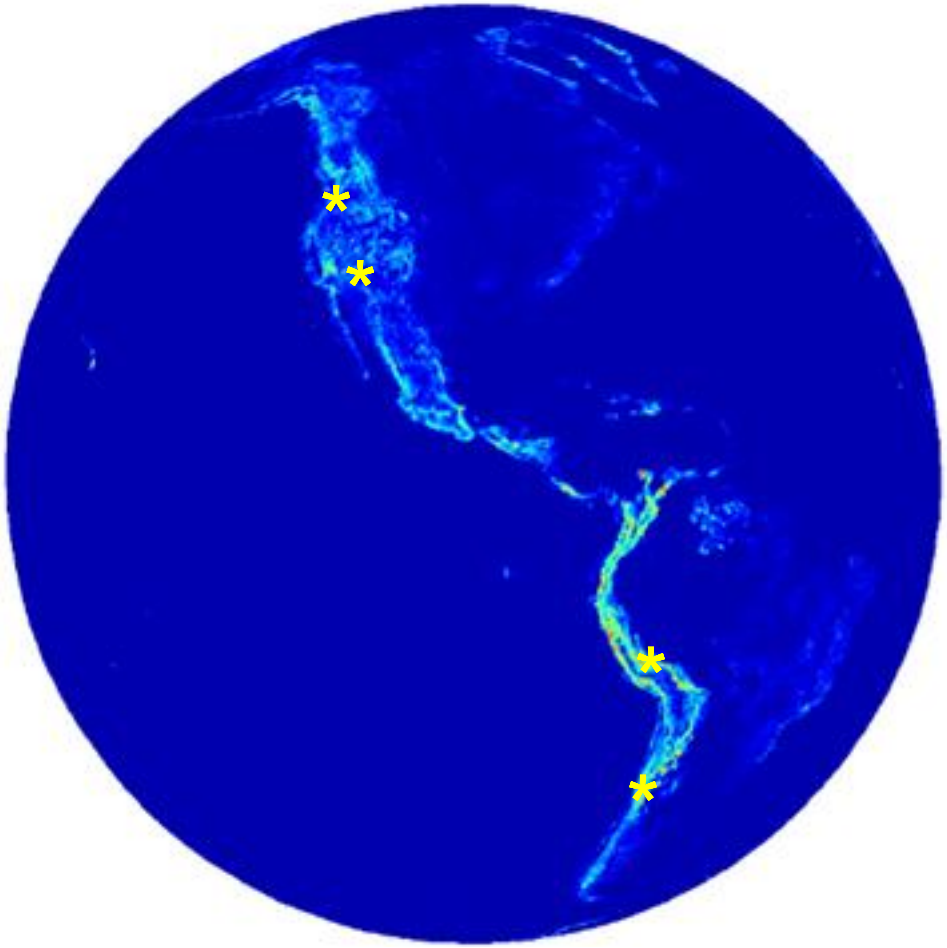


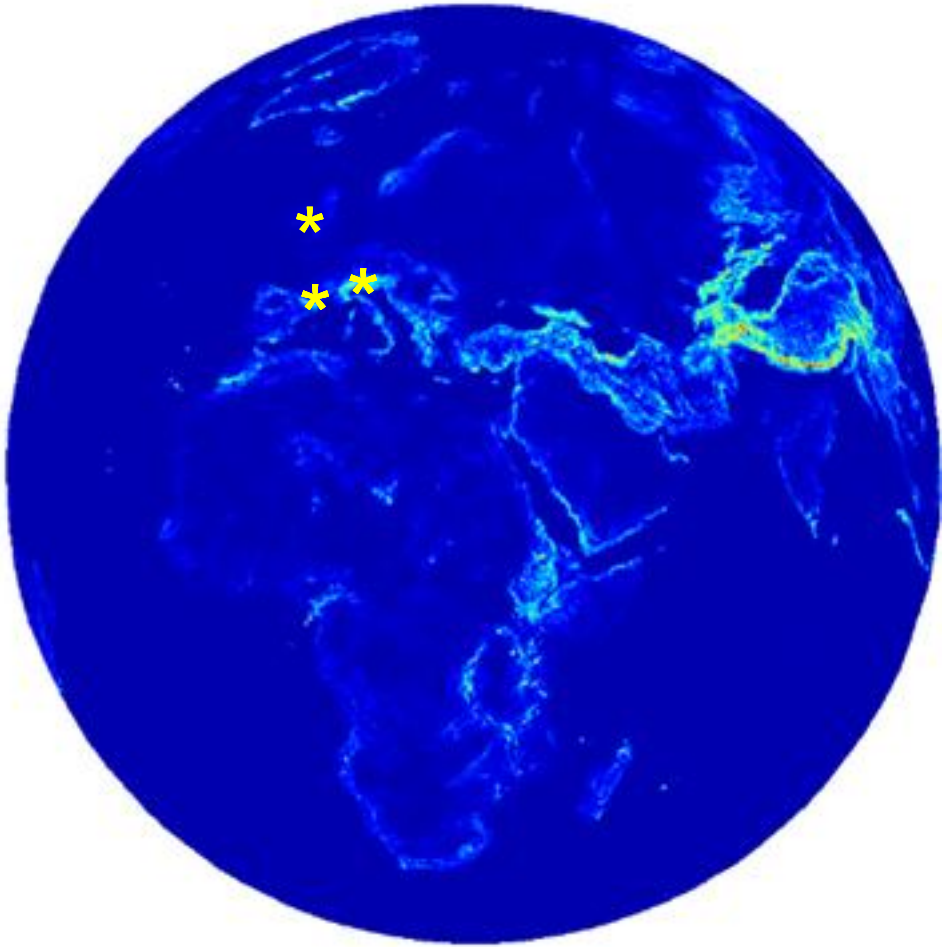
Категории данных




Argentina
Andorra
Antarctica
Iceland
Idaho
Switzerland
Argentina
Bolivia
Nevada
Hawaii
Hawaii
Egypt
China
Arizona
Peru
Oregon

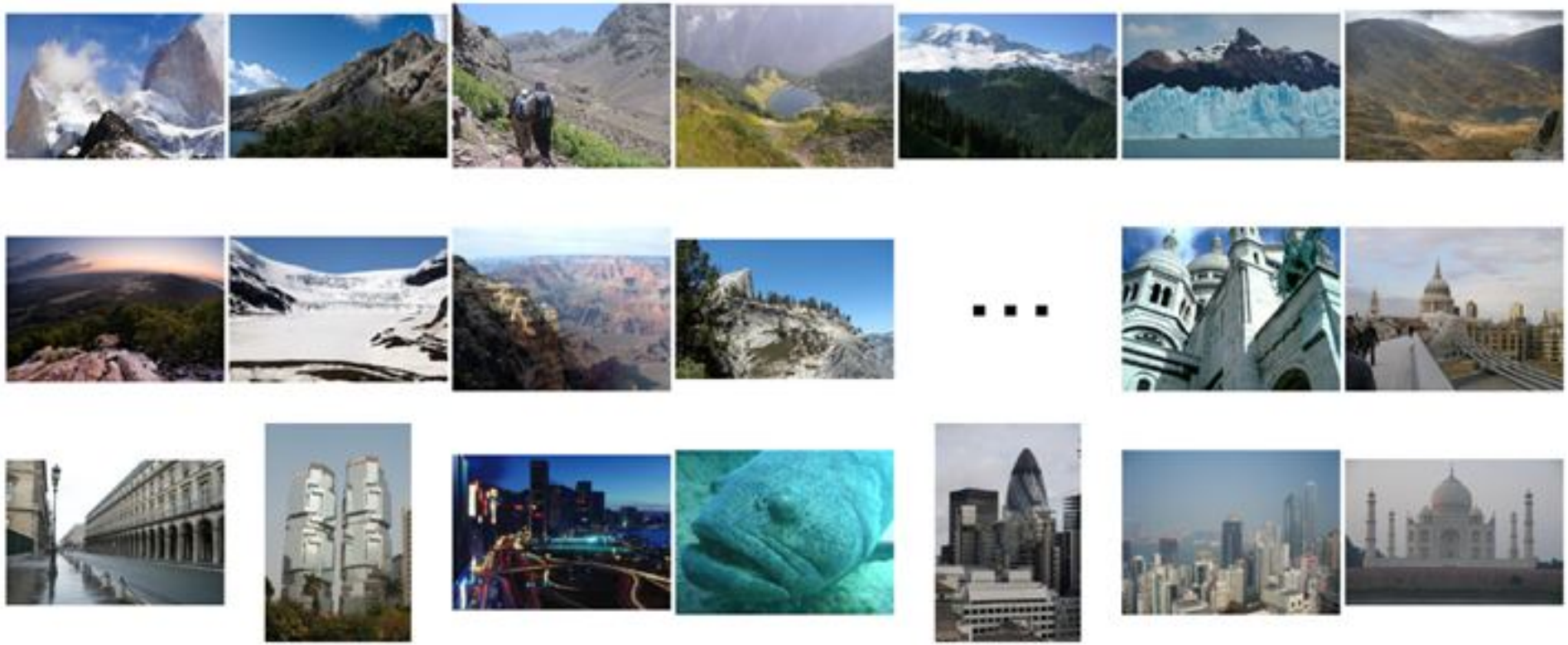
Скорость = 112 м / км





Уклон

Ранжирование изображений по уклону (от макс к мин)



Плотность населения

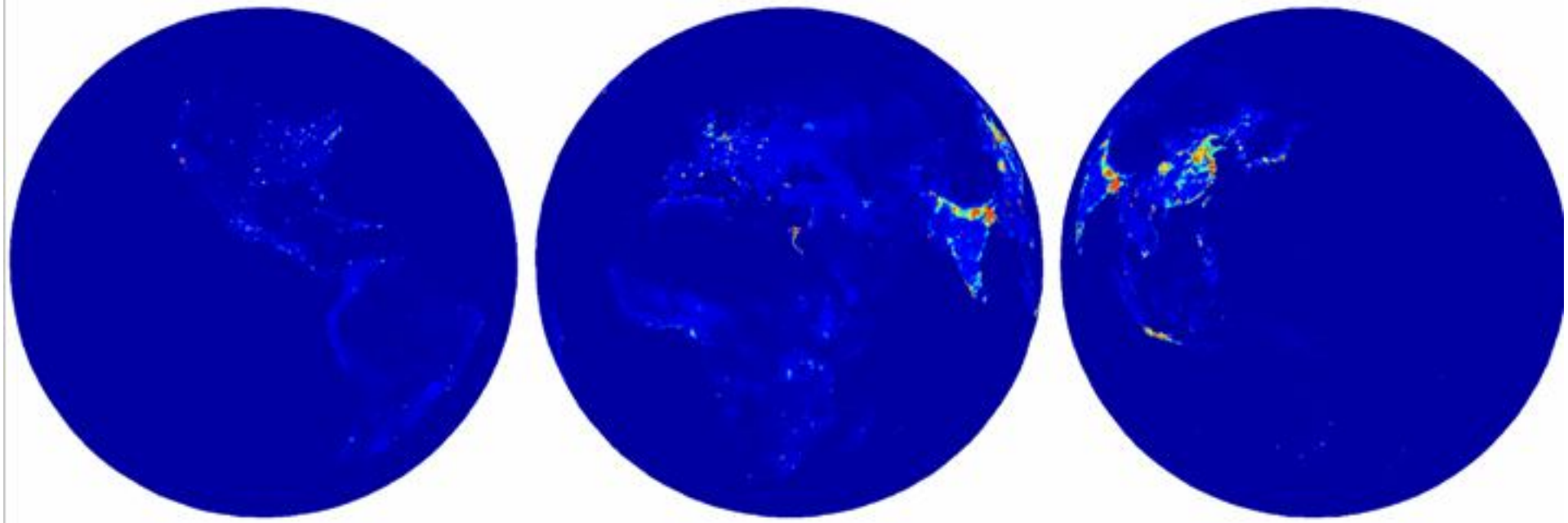


Figure 2. Global population density map.

Ранжирование по населению

Тип территории

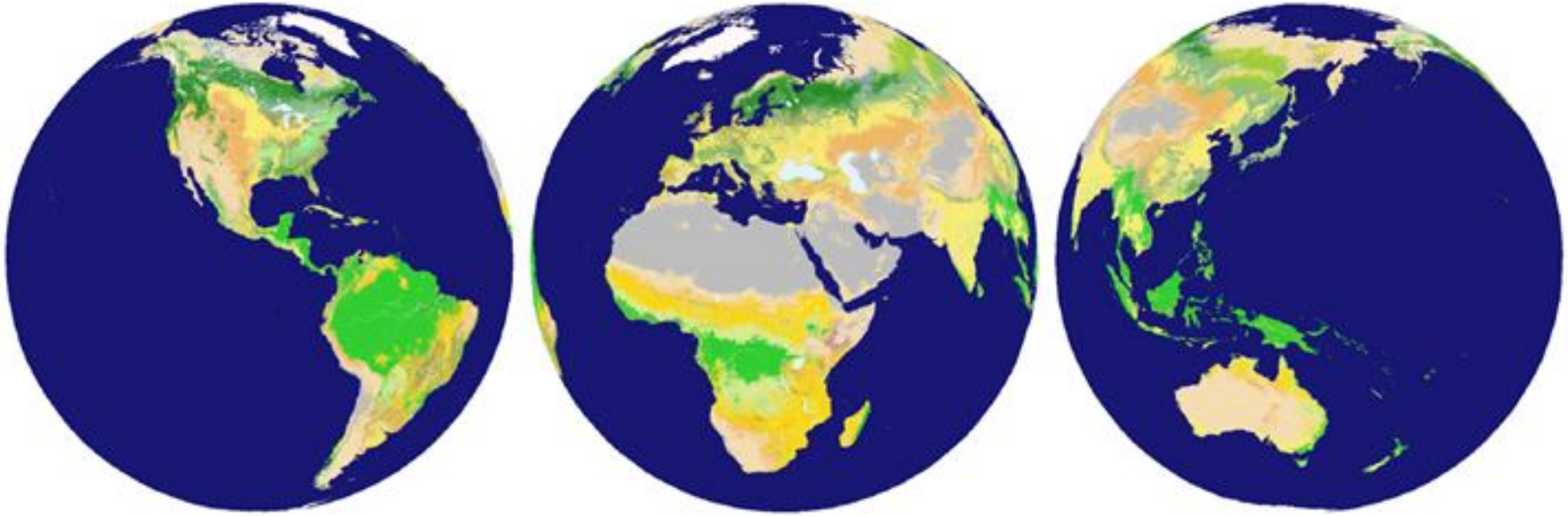
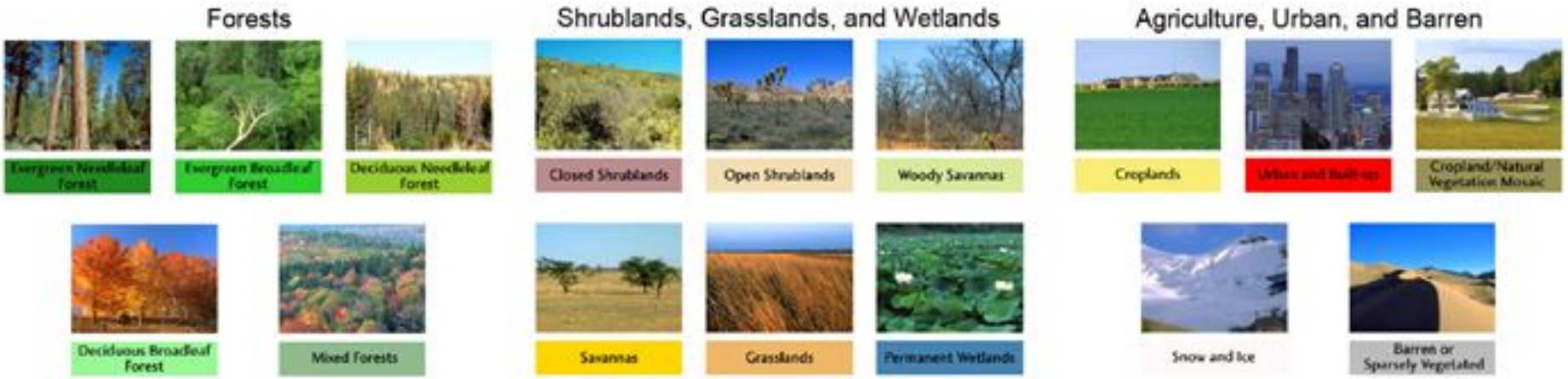
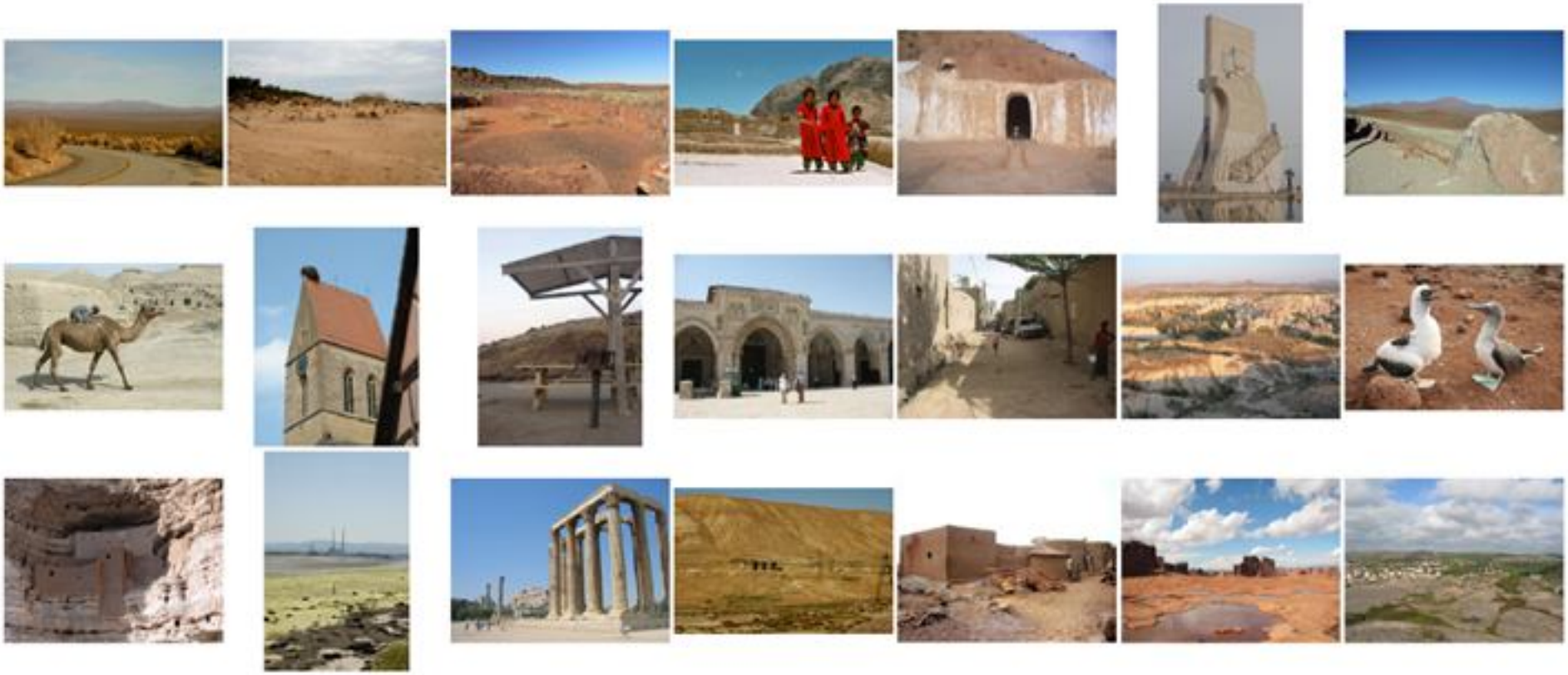


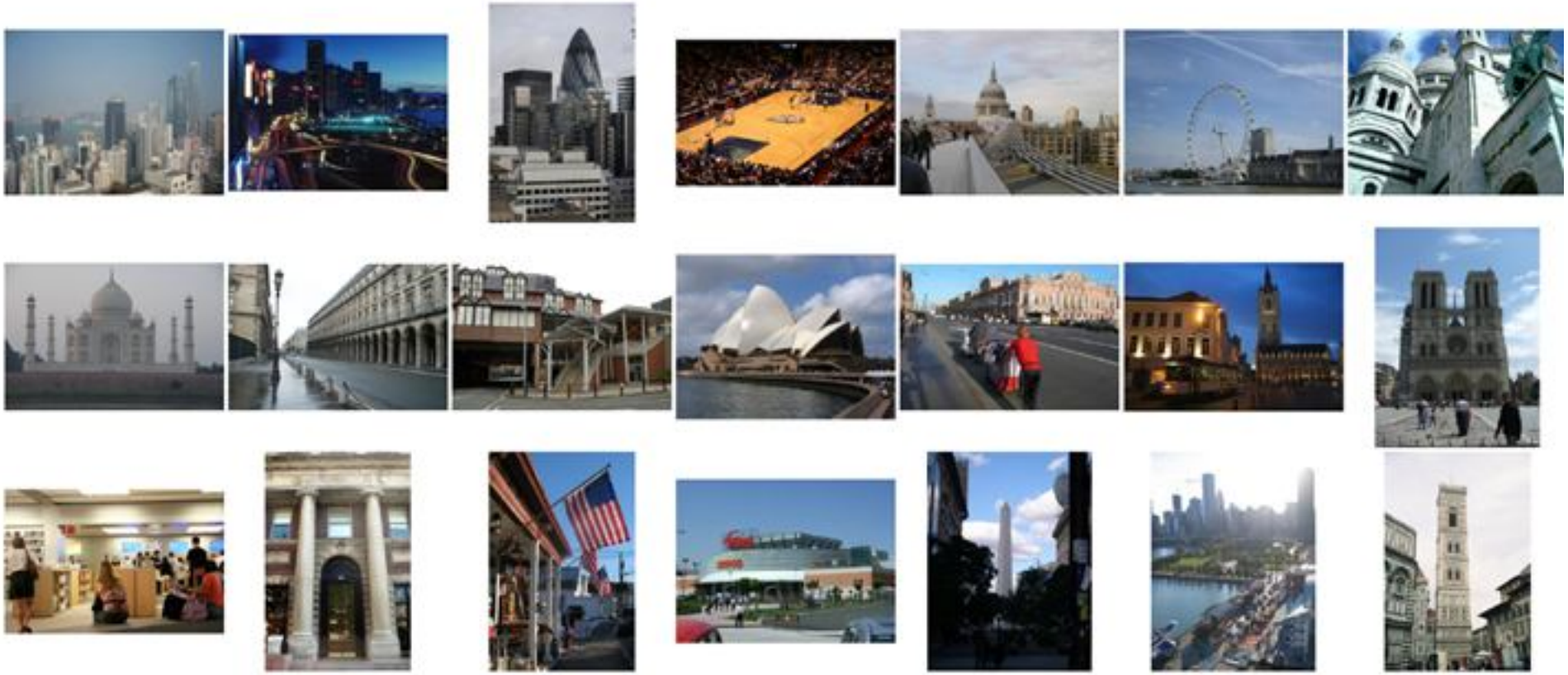
Figure 4. Global land cover classification map.



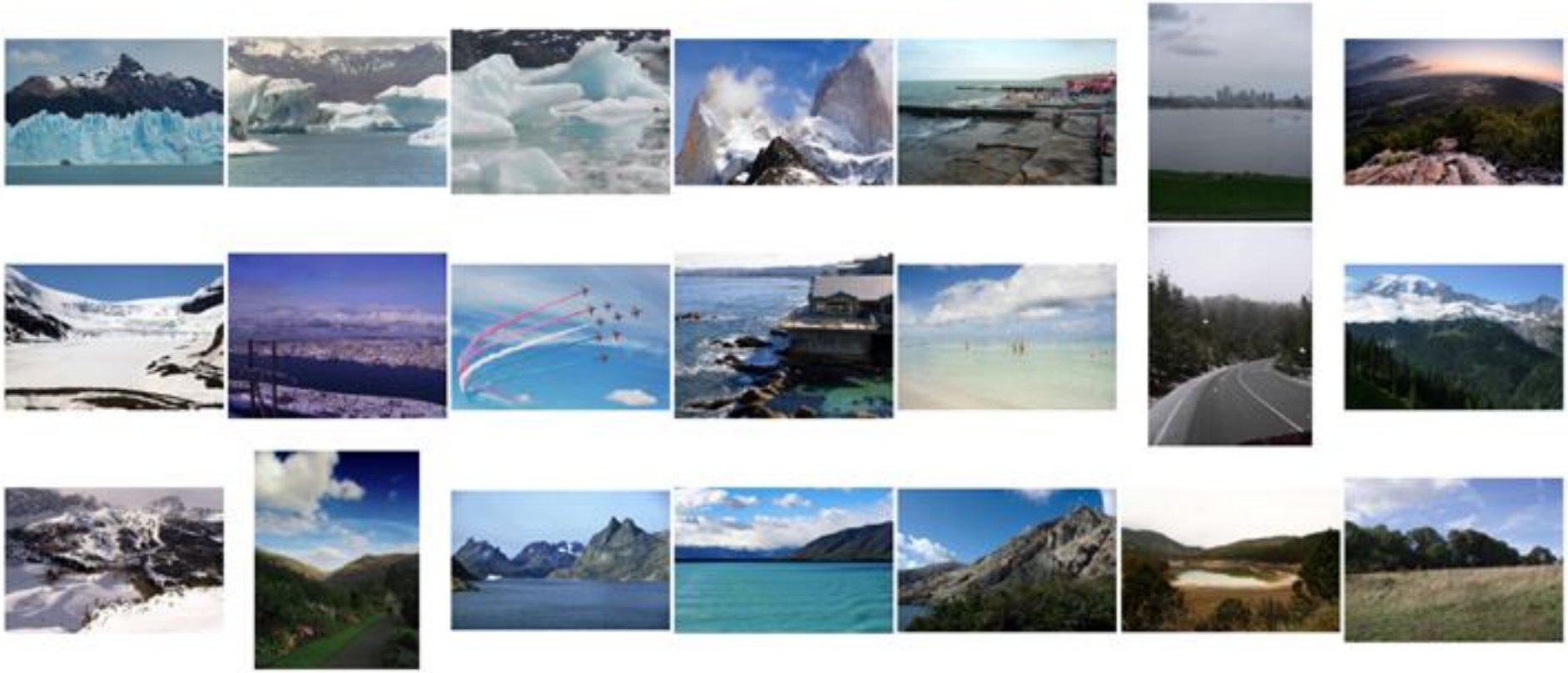
Пустыни



Города и здания

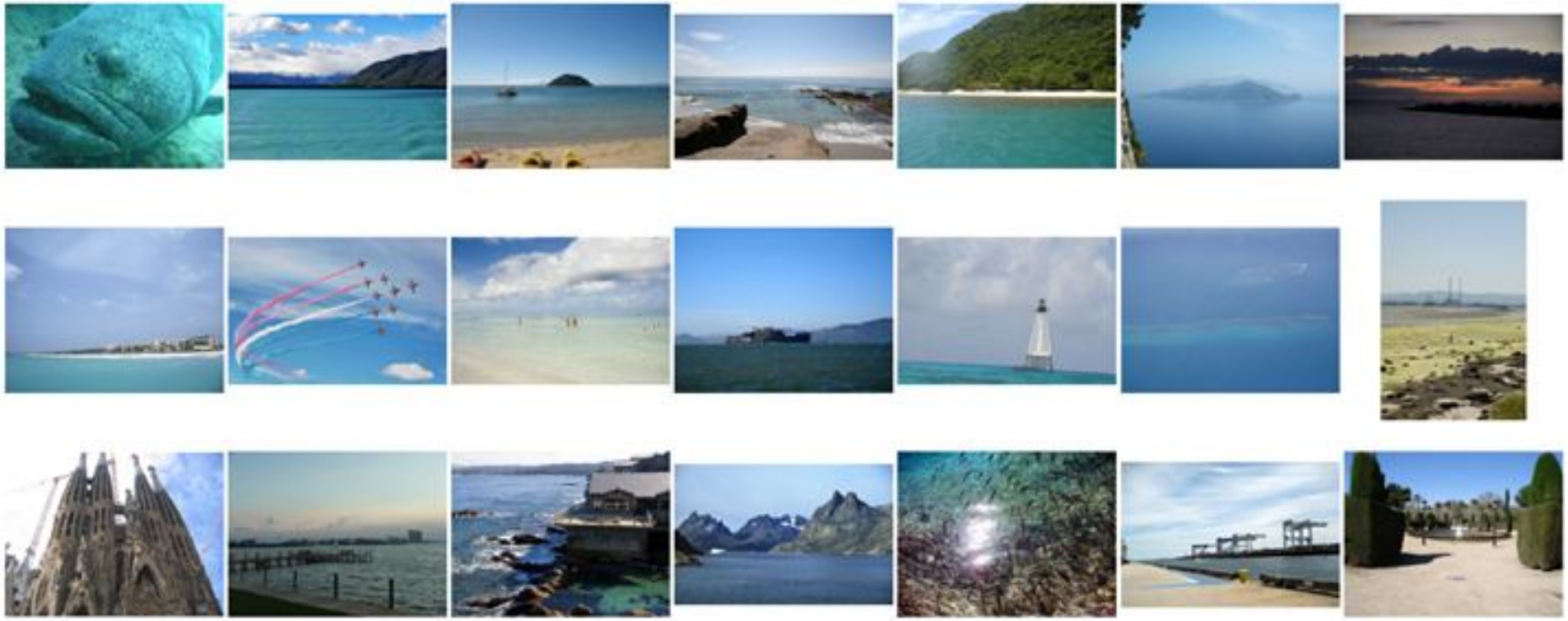


Снег и лед



Саванна

Вода



Некоторое резюме

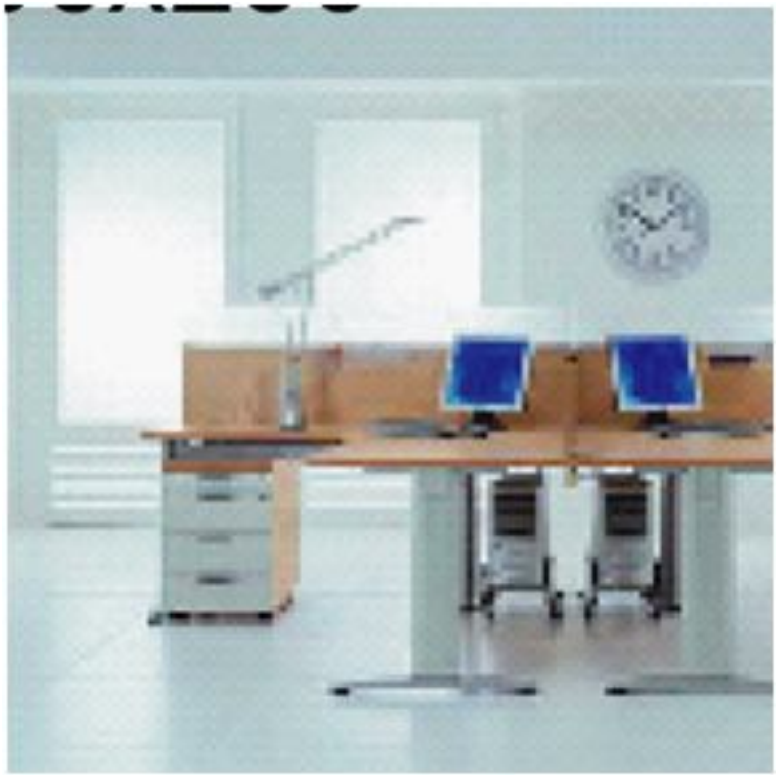
- Если нам важна общая конфигурация сцена, то достаточно дескриптора GIST для поиска похожих изображений
- При превышении порога в миллион изображений можно найти очень похожие по конфигурации
- Если есть дополнительная аннотация изображений базы, то KNN поиск по большой базе позволяет неплохо решать задачи регрессии по изображениям

Крошки-картинки (Tiny images)

А что мы можем сделать, если у нас есть сто миллионов картинок?

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.

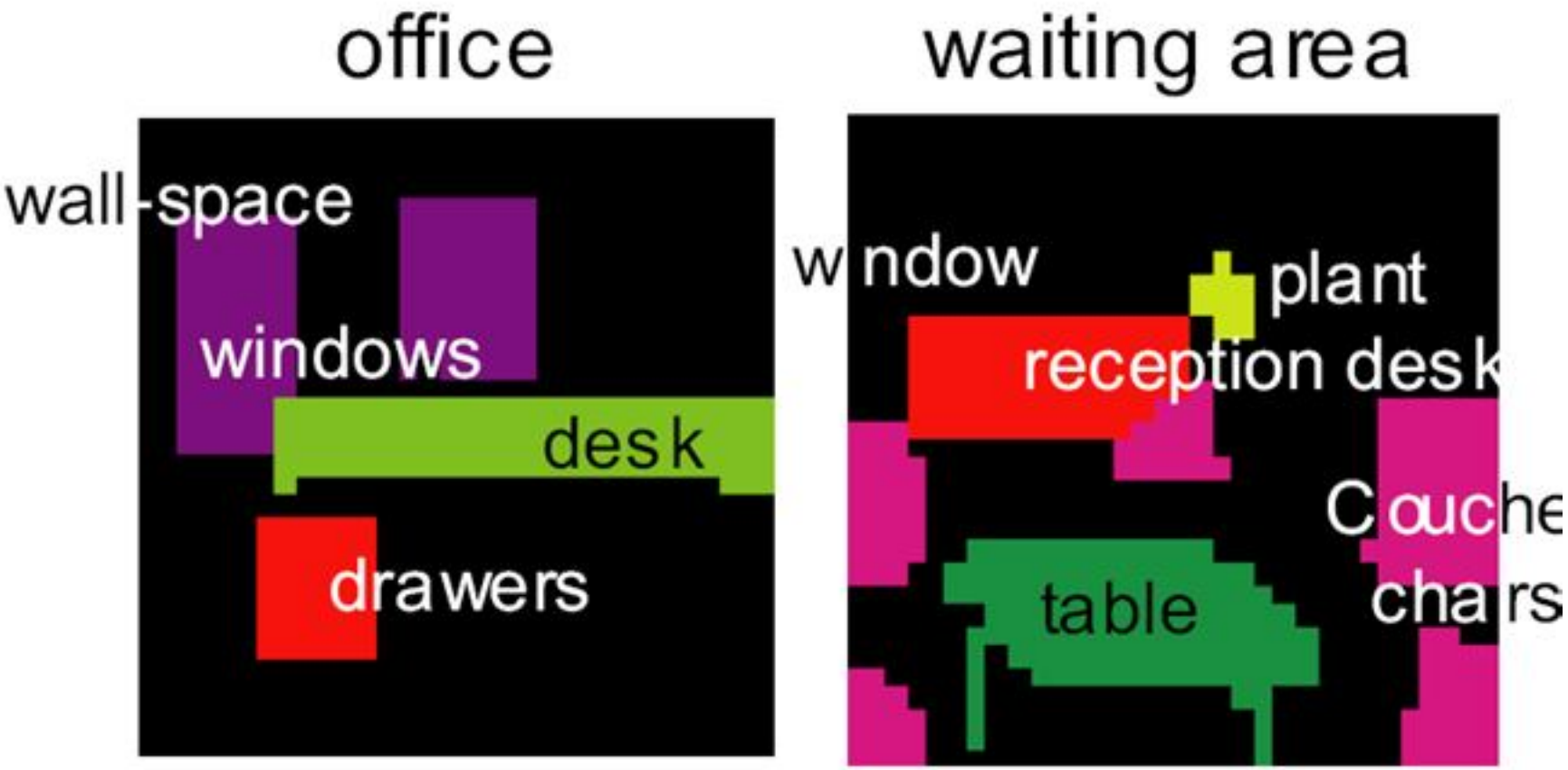
Примеры изображений



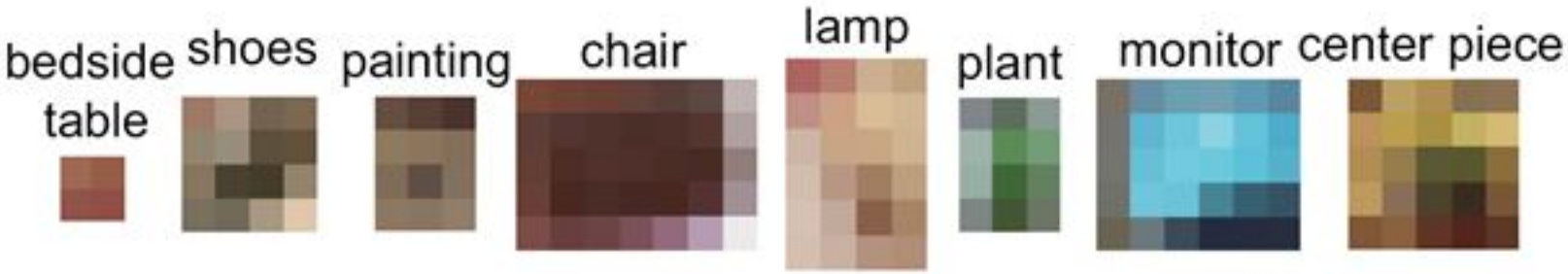
Какого размера нам потребуются картинки, чтобы их можно было использовать для распознавания? Попробуем на человеке...

Примеры изображений

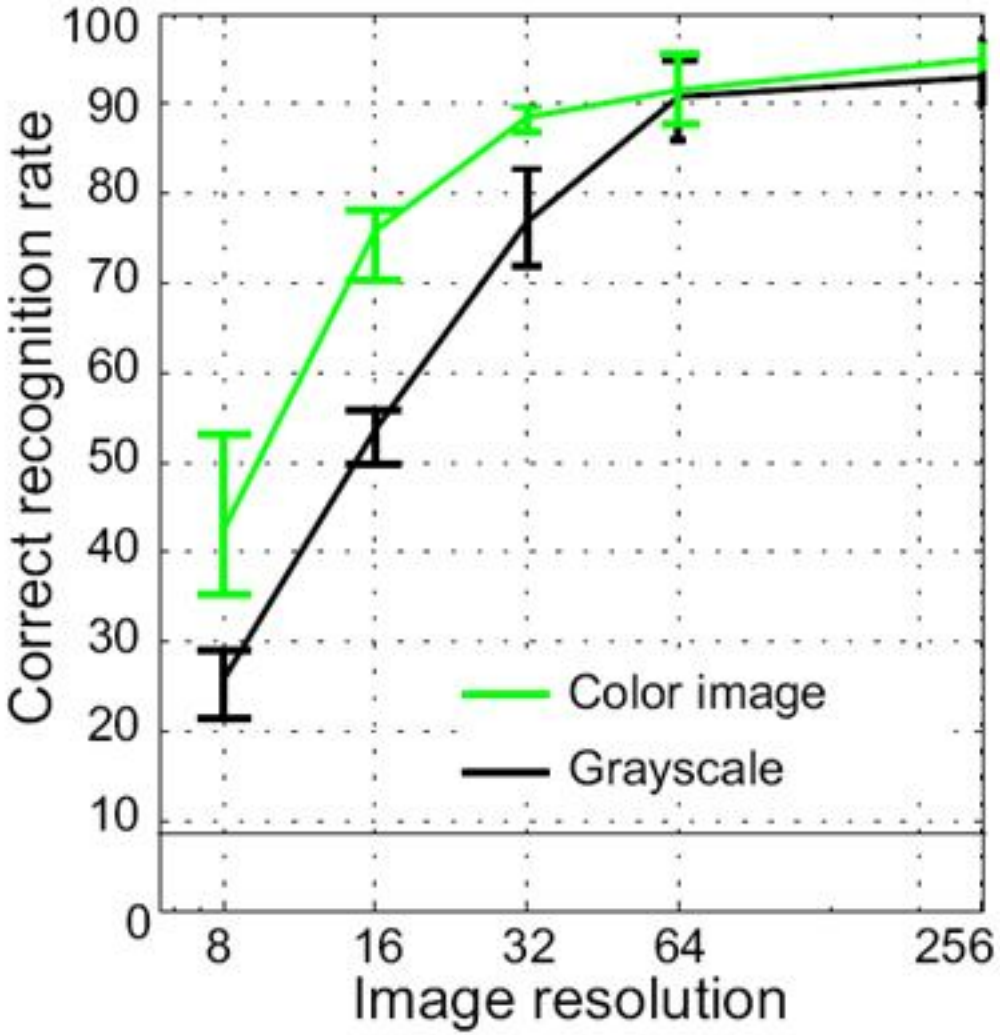
Сегментация



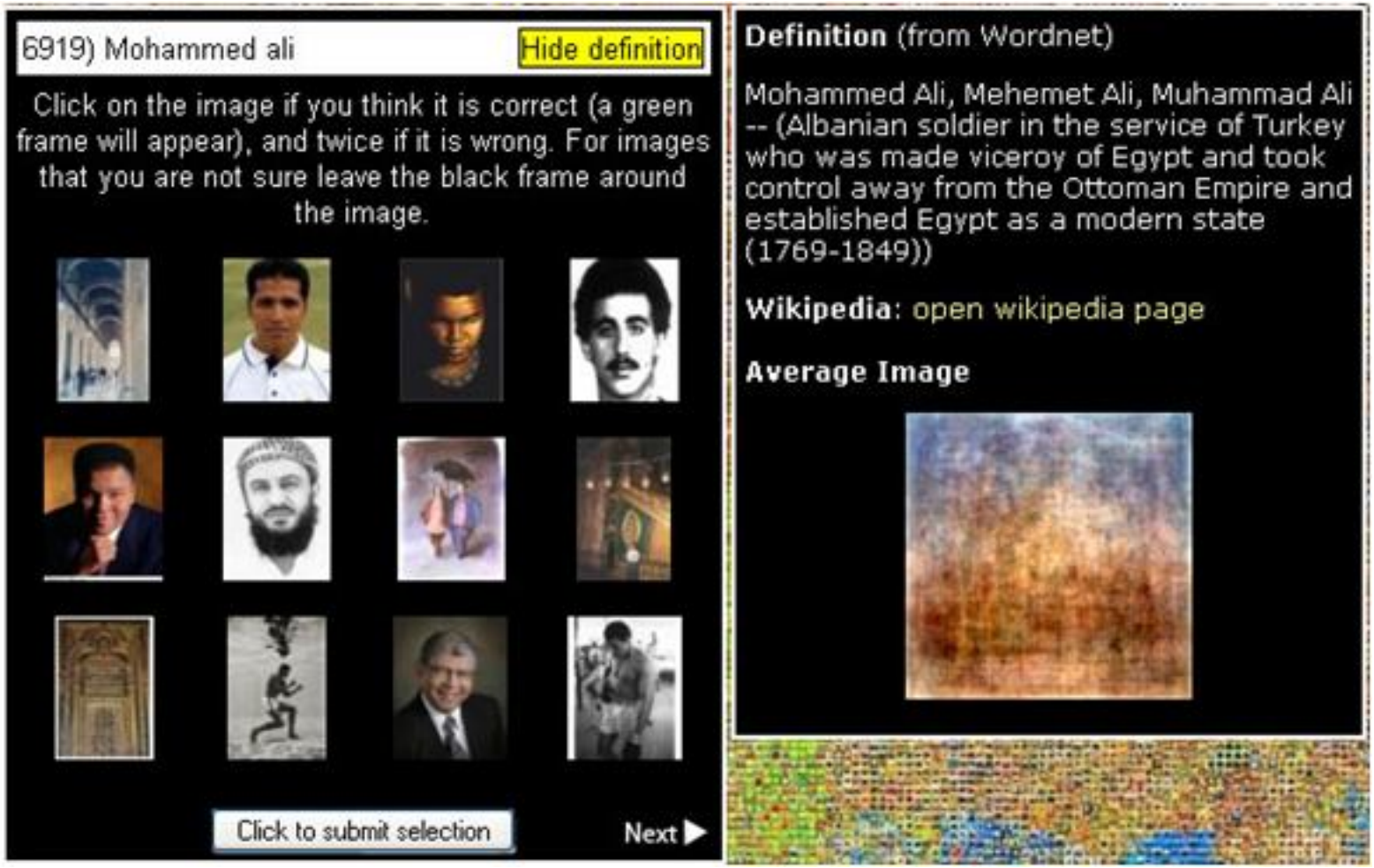
Отдельные объекты



Распознавание человеком

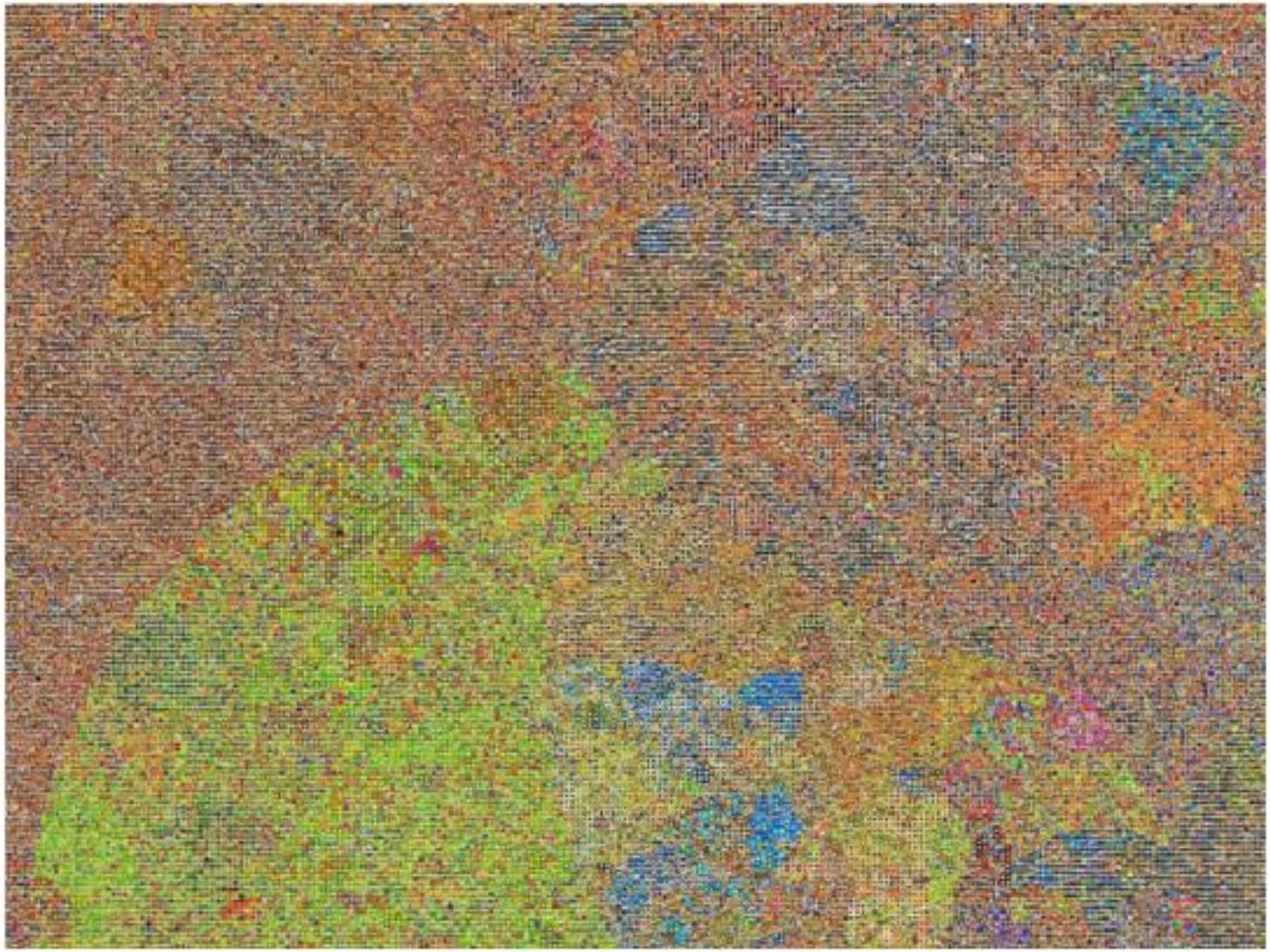


80 миллионов изображений



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

80 миллионов изображений



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

Метрики для поиска

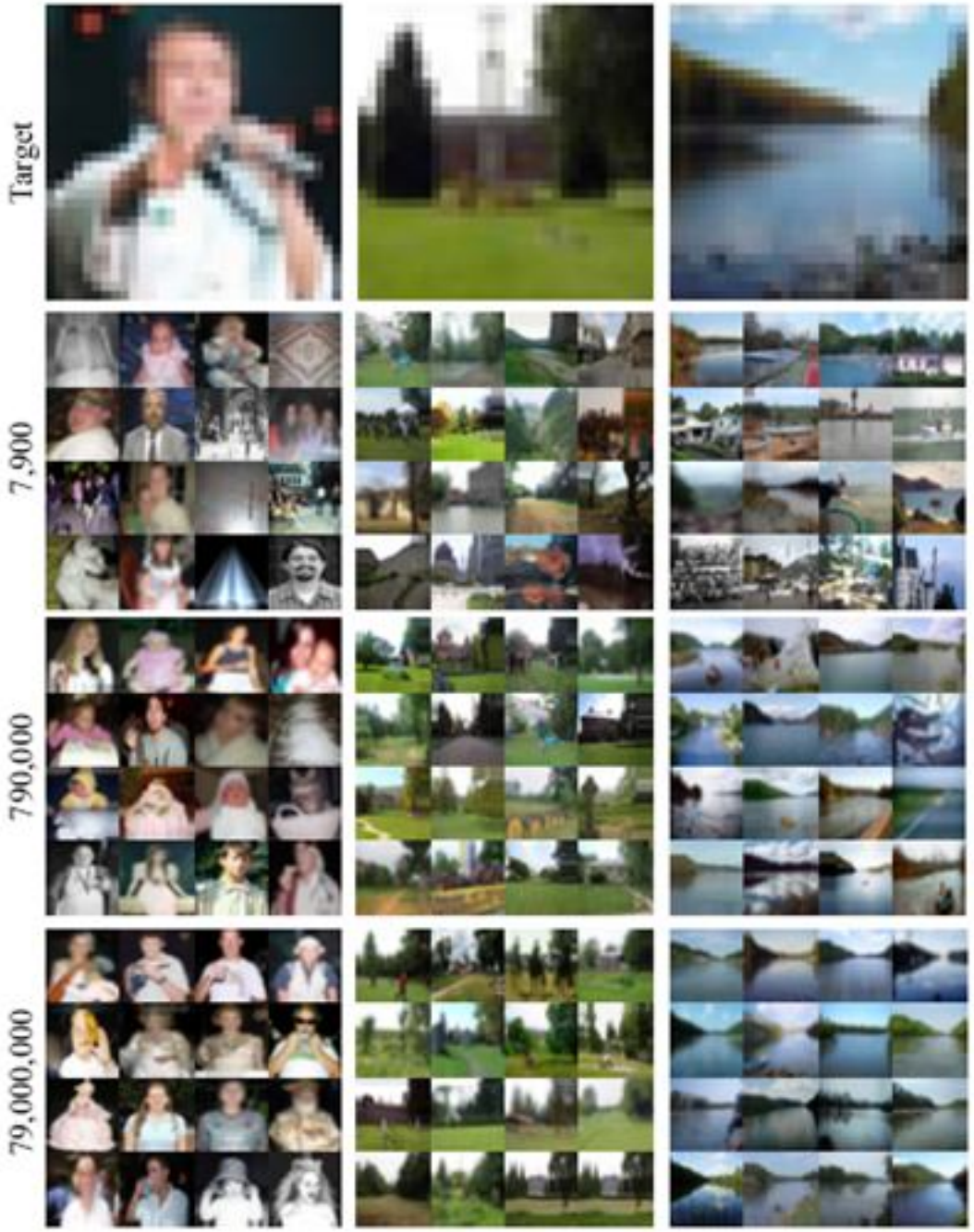
Target Neighbor Warping Pixel shifting

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

SSD-метрика

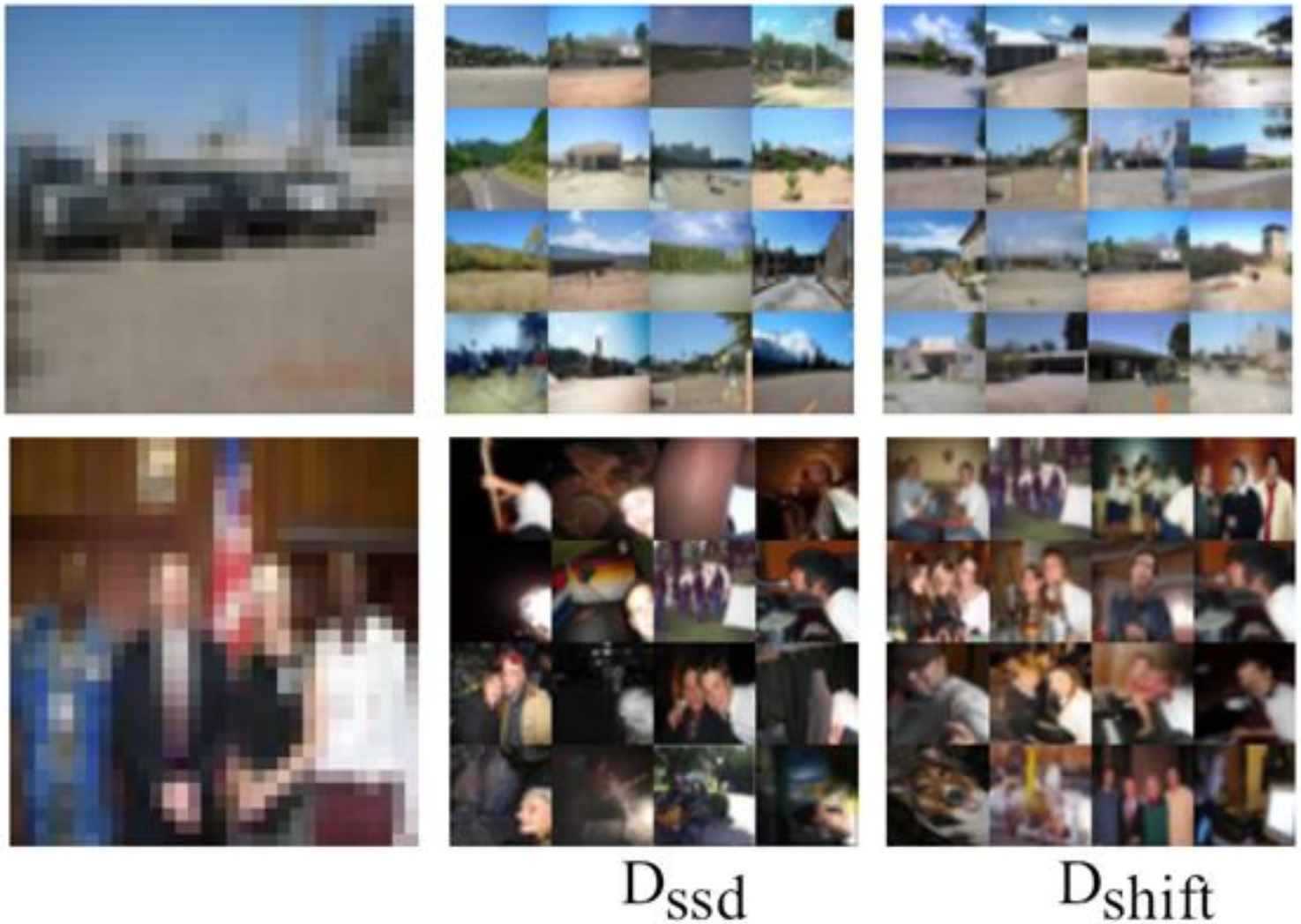
$$D_{warp}^2 = \min_{\theta} \sum_{x,y,c} (I_1(x,y,c) - T_\theta I_2(x,y,c))^2$$

Сравнение с искажением



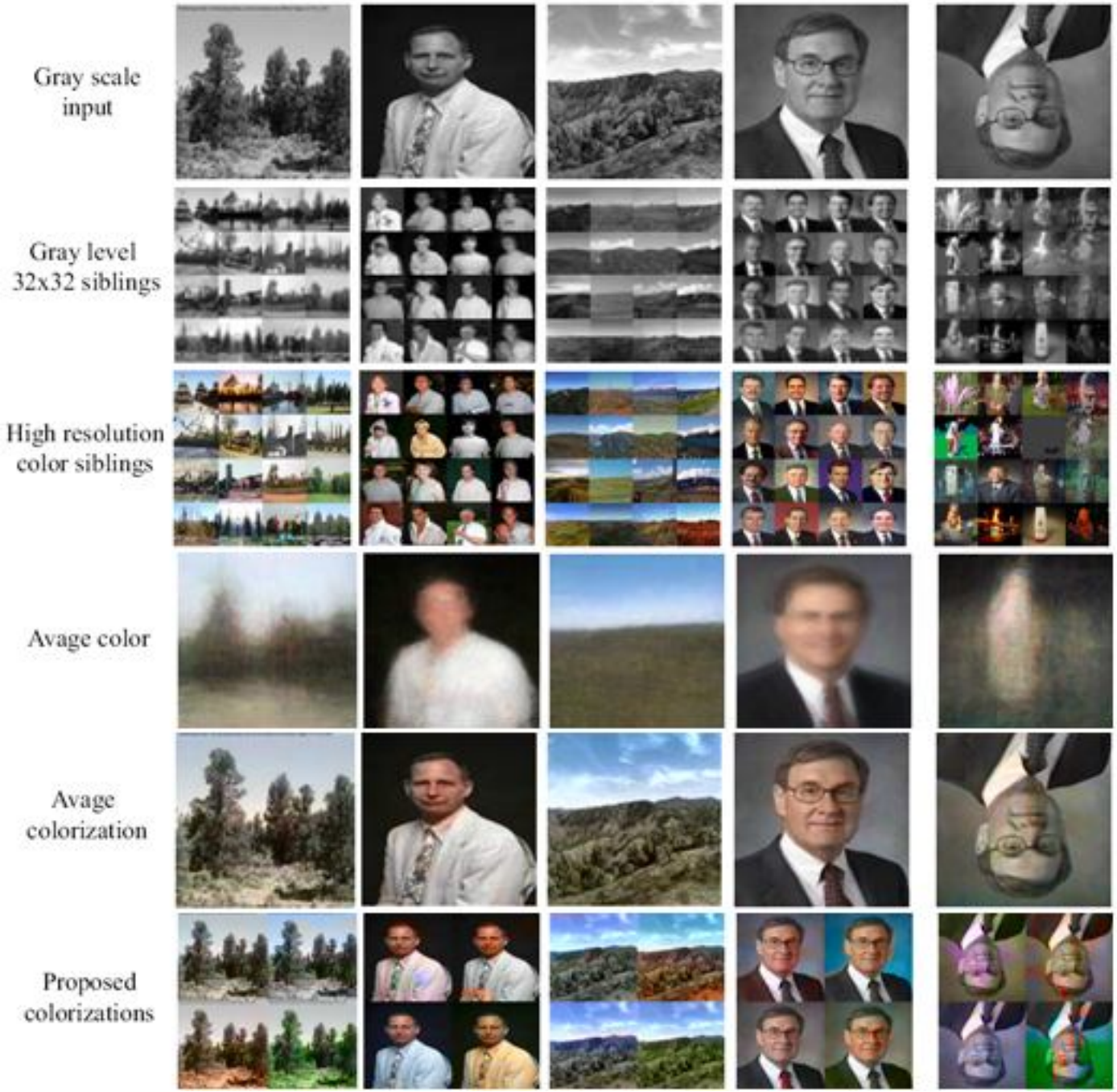
Наиболее похожие по SSD в зависимости от размеров коллекции

Результат улучшенной метрики



Поиск лиц на основе коллекции

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

Расширенная аннотация

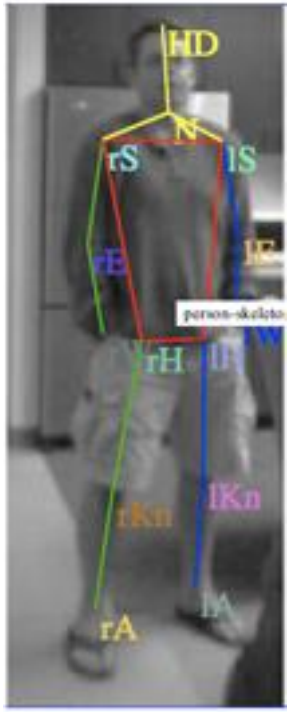
- Хотелось бы больше информации про изображения в коллекции
  - GPS-метка и атлас уже дали интересные результаты
- Нужна дополнительная аннотация
- Как бы нам её получить?

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/

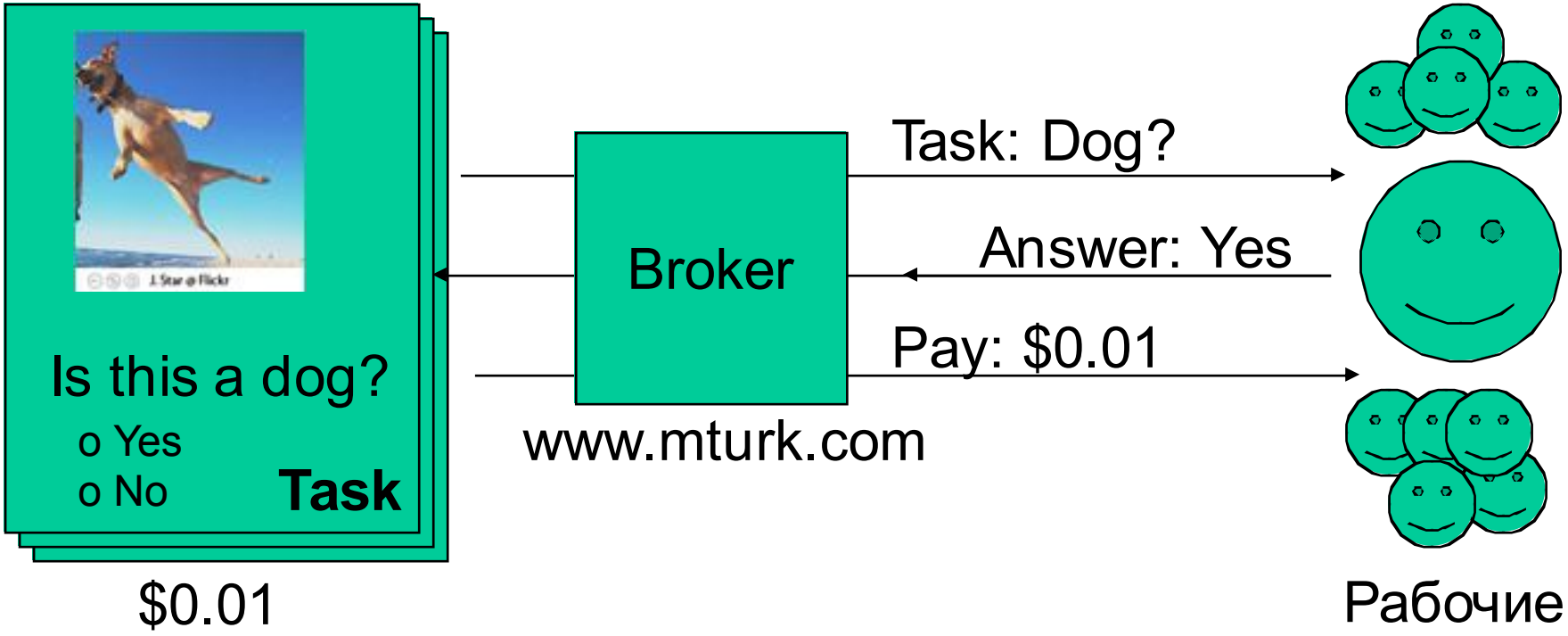
- Классификация изображений галактик
- Первый масштабный проект такого рода
- Более 150000 волонтёров за первый год бесплатно сделали более 60 млн. меток

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

Amazon Mechanical Turk



- Интернет-брокер для выполнения интеллектуальных заданий по анализу изображений
- Есть и другие брокеры
- В целом можно обеспечить наиболее дешевую из возможных разметок данных

Построение коллажей

Если у нас есть хорошая разметка, можем ли мы сделать хороший инструмент для коллажей?

Jean-François Lalonde, Derek Hoiem, Alexei Efros, Carsten Rother, John Winn, Antonio Criminisi **«Photo Clip Art»**, *SIGGRAPH 2007*.

Внедрение объектов в изображение

Высокодетализированная модель
Качественные материалы

- Фотореалистичная визуализация
- Дорого и «домохозяйке» не под силу

Paul Debevec. Rendering Synthetic Objects Into Real Scenes: Bridging Traditional and Image-Based Graphics With Global Illumination and High Dynamic Range Photography, SIGGRAPH 1998

Альтернатива: картинки

- Быстро и дешево
- Далеко от реальности….

Почему коллаж тяжело делать?

- Если возьмём произвольную машинку с другой фотографии…

Опять много-много данных…

- Возьмём подходящую картинку из коллекции…

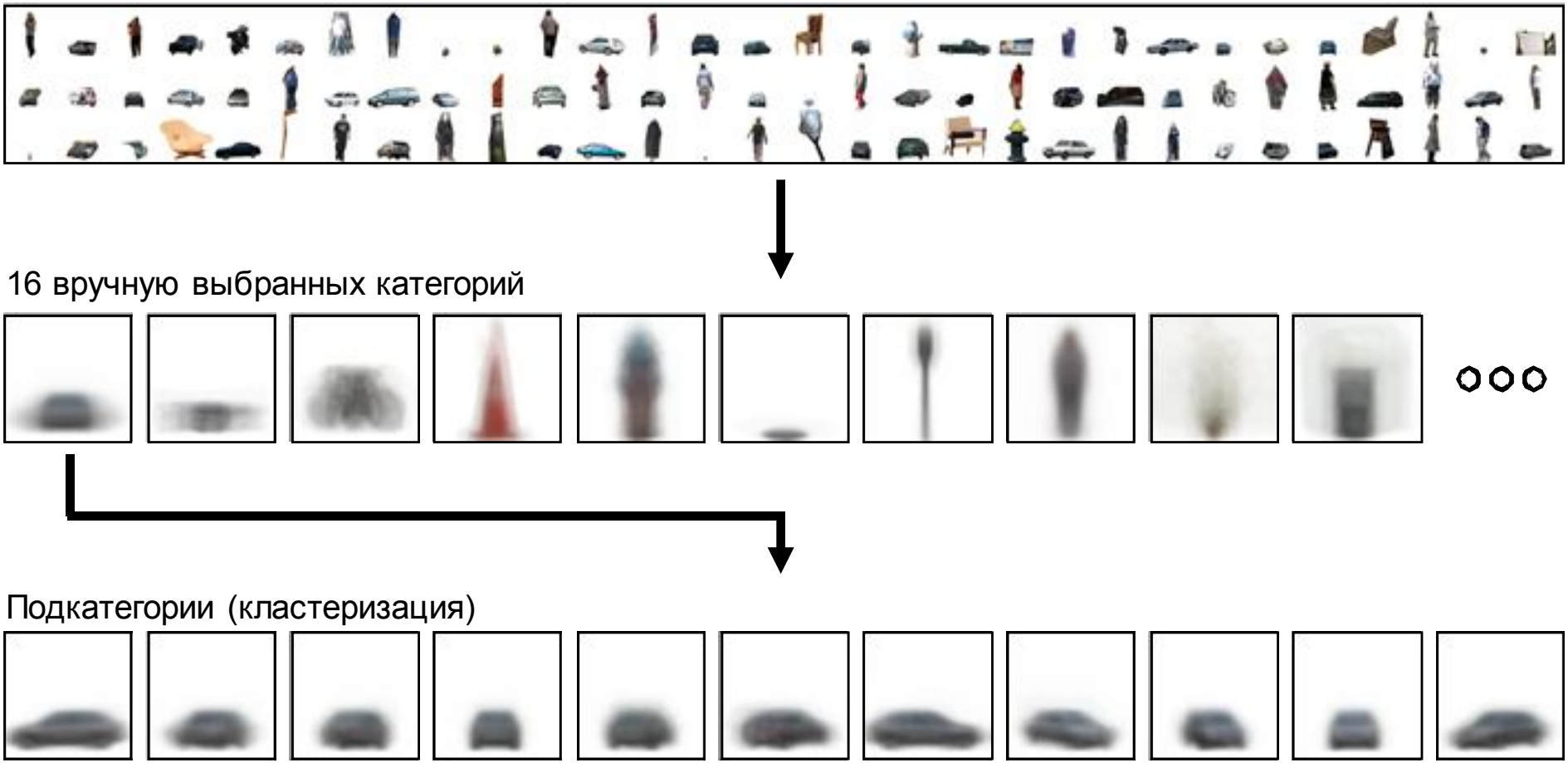
Источник данных: LabelMe

- Но просто контура и категории нам недостаточно, нужны и другие параметры
- Можно их вычислить самим!

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

Кластеризация данных

- Разобьём объекты на группы по похожим силуаэтам

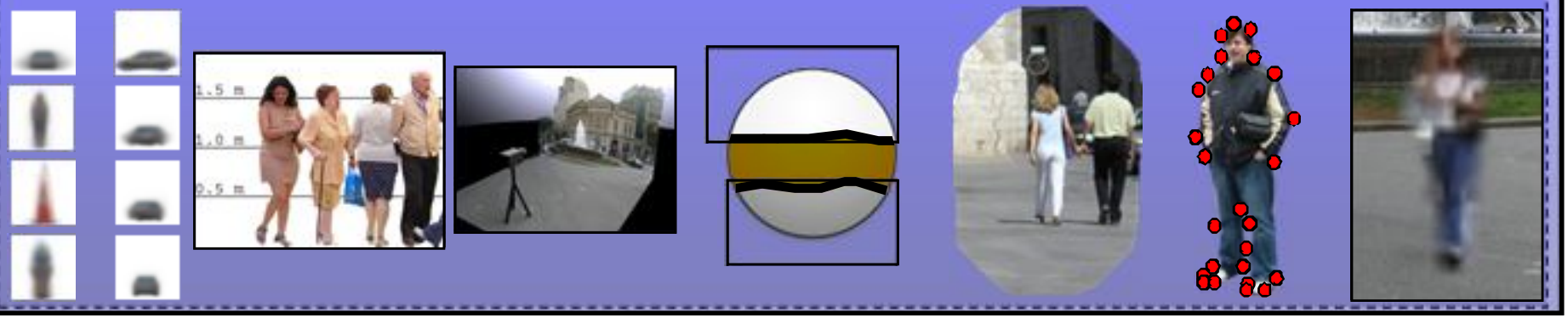


Все вместе

Этап I: Аннотация базы

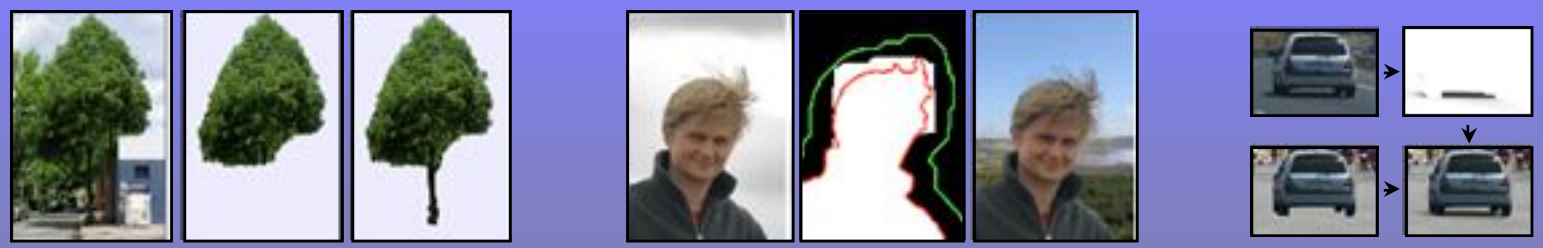
Object properties (used for sorting the database)

Label | Cluster | 3-D height | Camera | Illumination context | Local context | Segmentation | Blur



Этап II: Вставка объекта

Segmentation | Blending | Shadow transfer

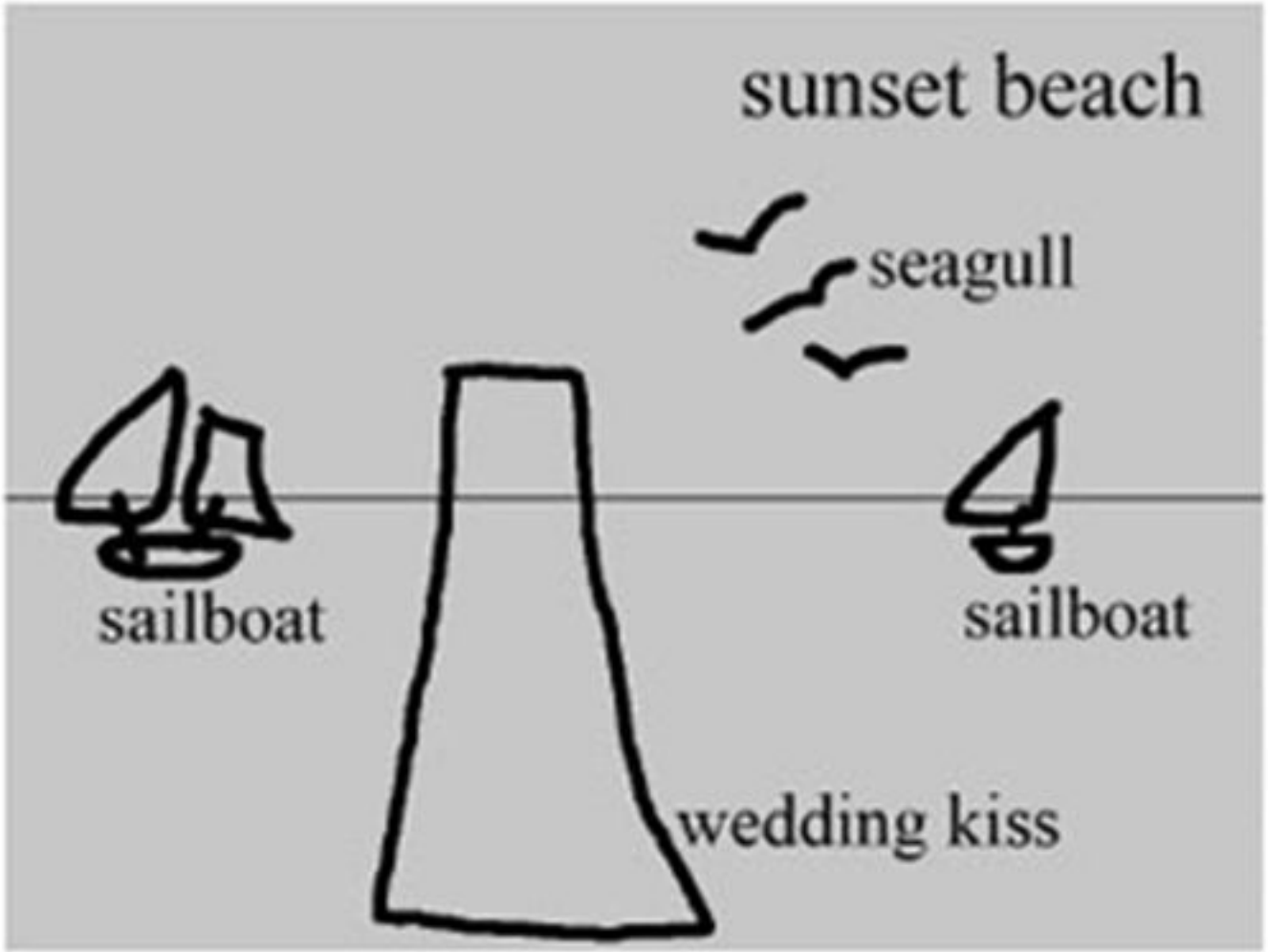


Авария

Мост

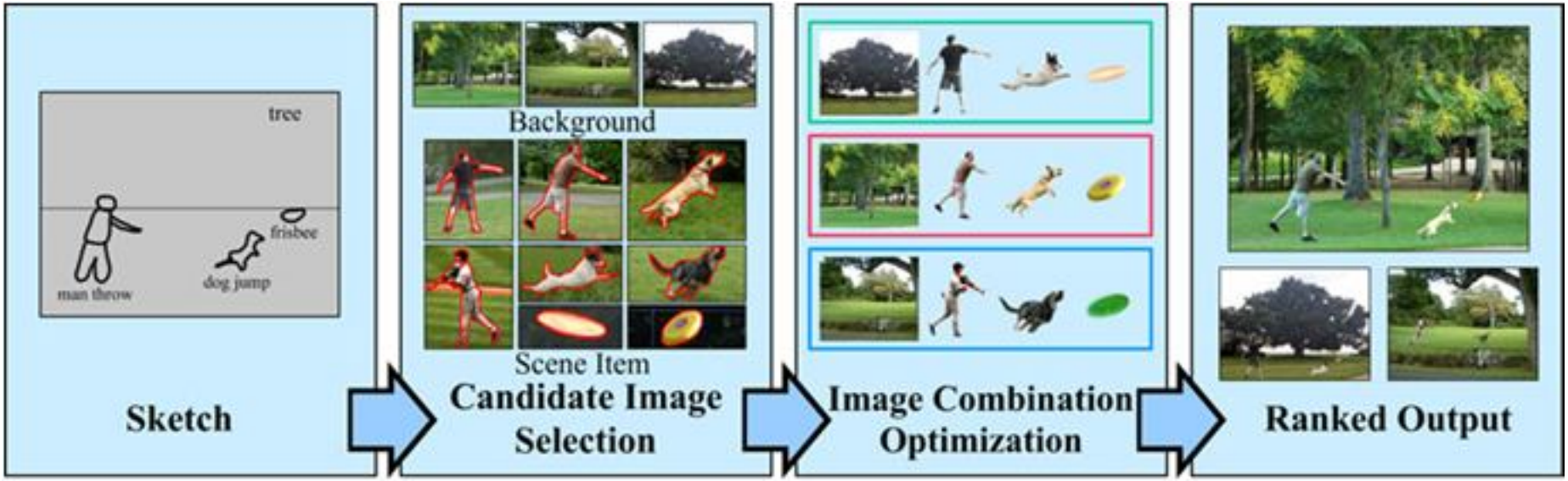
Улица

Sketch2photo



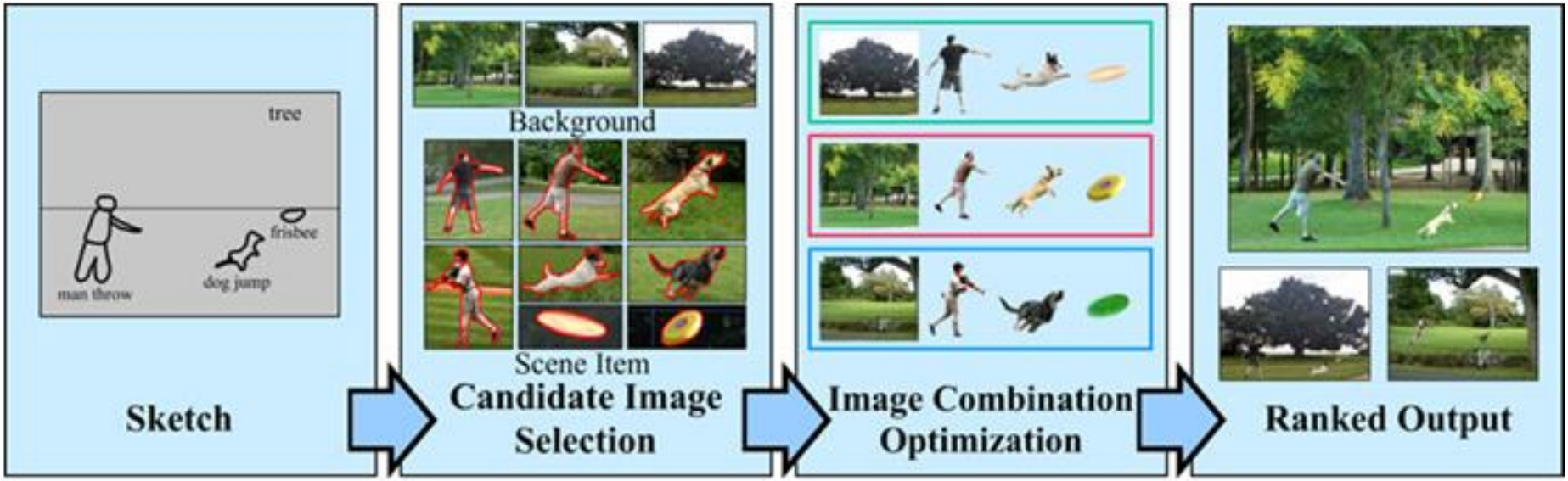
Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, Shi-Min Hu
Sketch2Photo: Internet Image Montage *ACM SIGGRAPH ASIA 2009*

Идея метода



- Идея:
  - Поиск в интернете подходящих для композиции изображений
  - 3000 кандидатов для каждого объекта (выбирается 100)
  - 1000 кандидатов для фона (выбирается 20)
  - Выбираем по 100 лучших примеров объектов и 20 фонов
  - Ищем оптимальную комбинацию

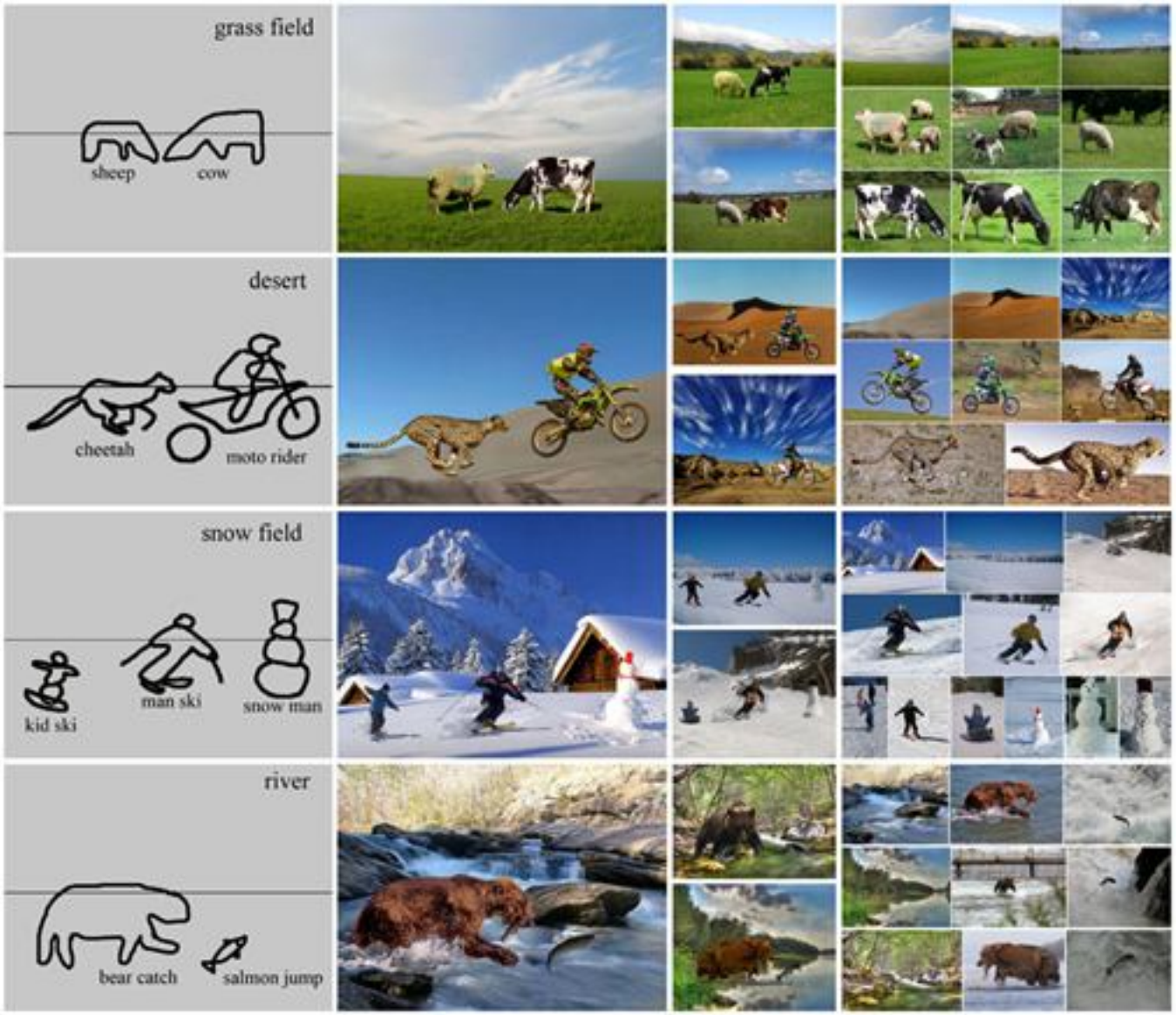
Идея метода



- Идеи:
  - Фон должен быть «простым», без мусора
  - Объект должен хорошо выделяться
  - По форме объект должен быть как можно похоже на рисунок пользователя
  - Фон объектов и фон для изображения должны быть похожи
  - И т.д. (можно много всего придумать)

Пример результата

Примеры результатов



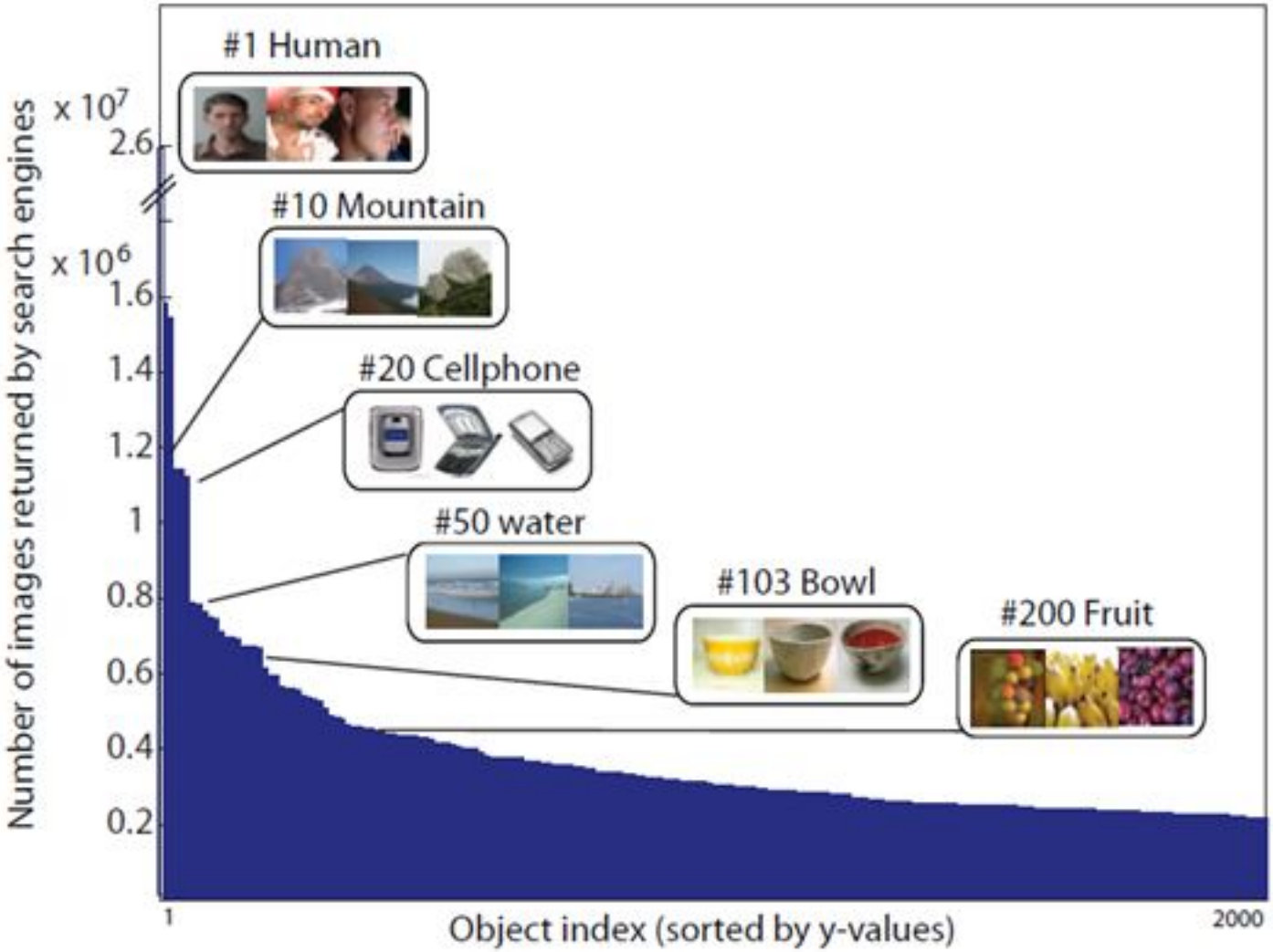
Пока всё работает медленно - 15 мин * число объектов + 3 мин на фон

Описание сцены через объекты

- Если бы мы могли хорошо находить объекты и материалы (воду, небо, и т.д.), то можно было бы классифицировать и искать изображения по этим описаниям
- Но теперь мы можем!!!
  - Большие базы для обучения
  - HOG + SVM + части = детектор объектов

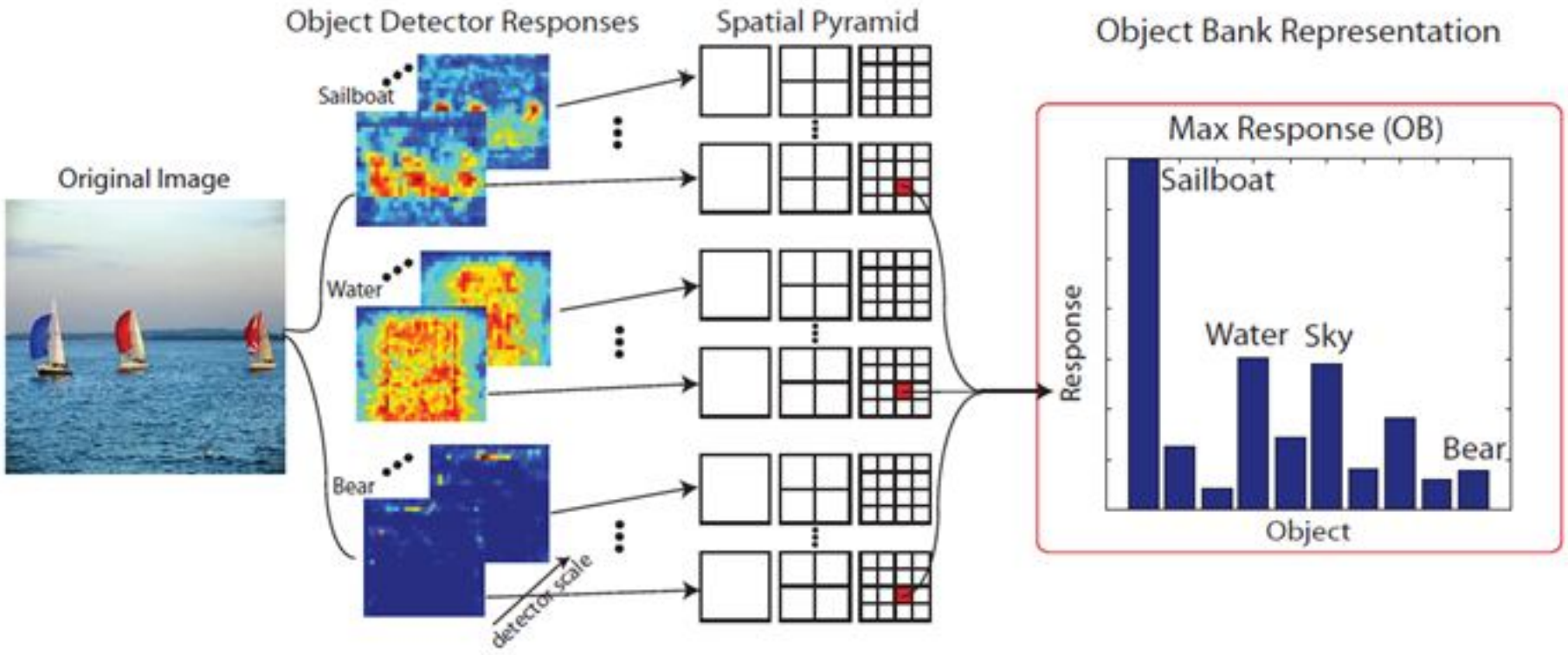
L.J. Li*, H. Su*, Y. Lim and L. Fei-Fei, Objects as Attributes for Scene Classification. ECCV 2010

Выбор объектов



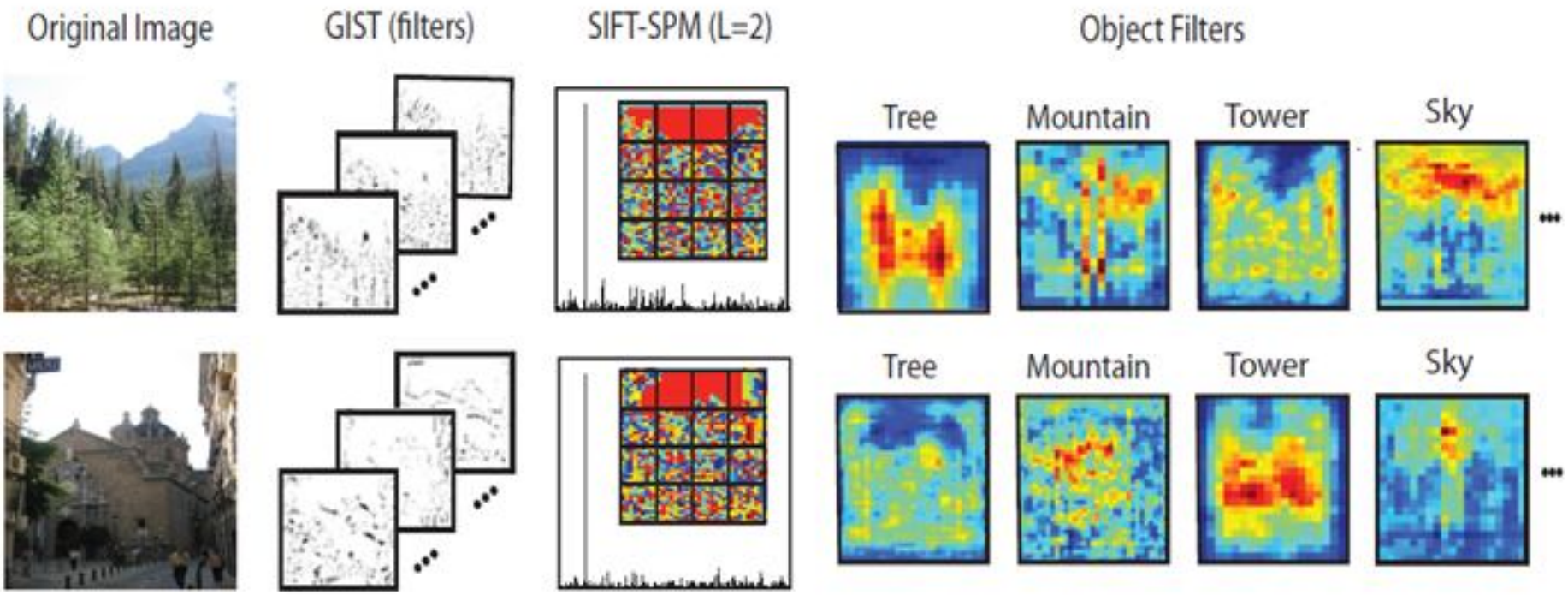
Число изображений, найденных по поиску через ключевые слова

Фильтры объектов



- «Фильтруем» изображения с помощью детекторов классов
- Суммируем области по прямоугольным областям
- Строим несколько уровней такого разрешения

Объединение признаков



- Фильтры объектов, GIST и «мешок слов» дополняют друг друга, кодируя разный тип информации
- По всем признакам можно обучить линейный SVM для классификации изображений

Резюме лекции

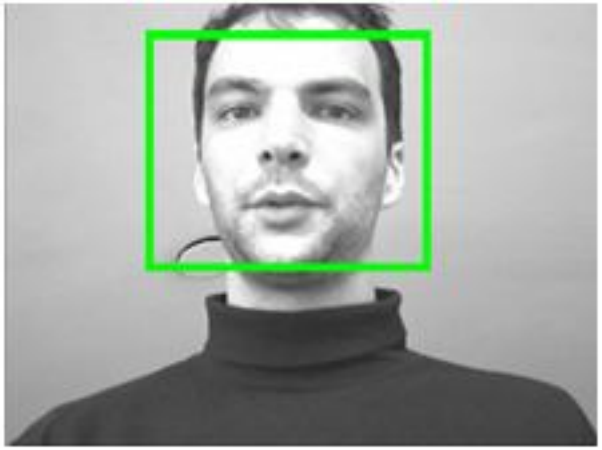
- Данные – один из краеугольных элементов для решения задачи компьютерного зрения
- Если данных очень много, то ряд задач можно решить «по аналогии»
- Нужно уметь собирать большие массивы данных и быстро искать в них
- Интернет и глобализация дают нам удобный инструмент для сбора и обработки данных

- GIST, LabelMe, ImageNet, WordNet

Распознавание и анализ лиц

Антон Конушин

Что предполагаем

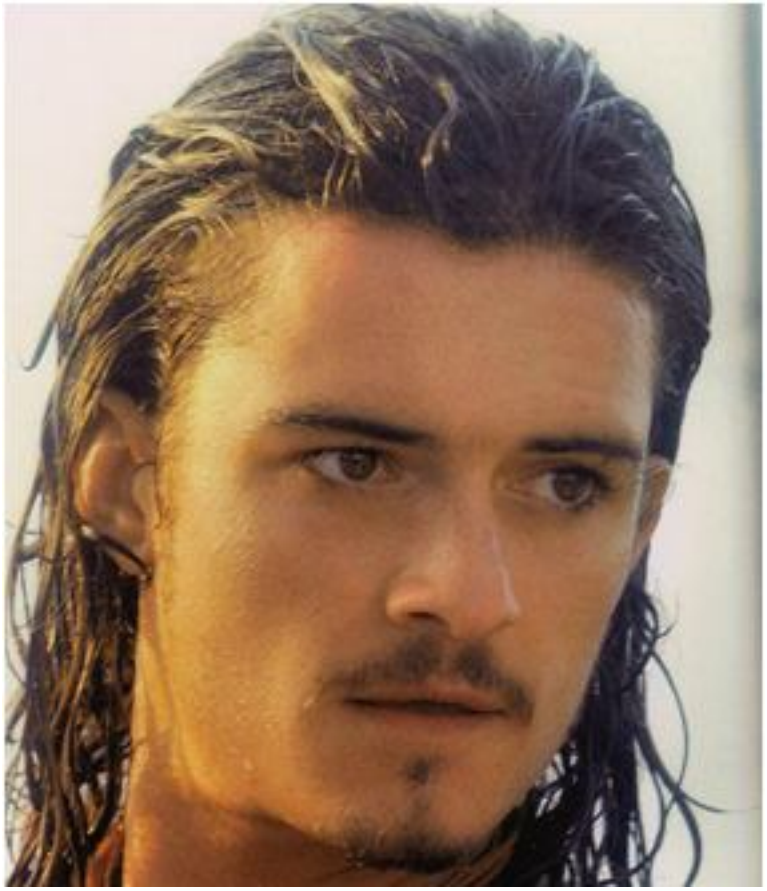


- Умеем:
  - Выделять лица в изображении
- Знаем, что можно (эти задачи рассматриваются в лекции про модели из набора частей)
  - Выделять контрольные (антропометрические) точки на изображении лица, такие как уголки глаз, и т.д.
  - Строить 2D сеточные модели лица по изображению (триангулирование сетки)

Что будем рассматривать

- Основные задачи и тестовые базы
- Распознавание лиц
- Распознавание атрибутов по лицу
- Фокусы с лицами

Верификация



На обоих изображениях один и тот же человек, или нет?

Базовая задача распознавания лиц, которую и человеку проще всего решать

Идентификация

- Есть фиксированный список людей (база) и тестовое изображении
- Нужно определить, кто из списка изображени на тестовой фотографии?

«Watch List»

- Есть список «подозрительных людей»
- Необходимо определить, входит ли человек в этот список по его фотографии
- Самая сложная постановка задачи

Аттрибуты

Мужчина

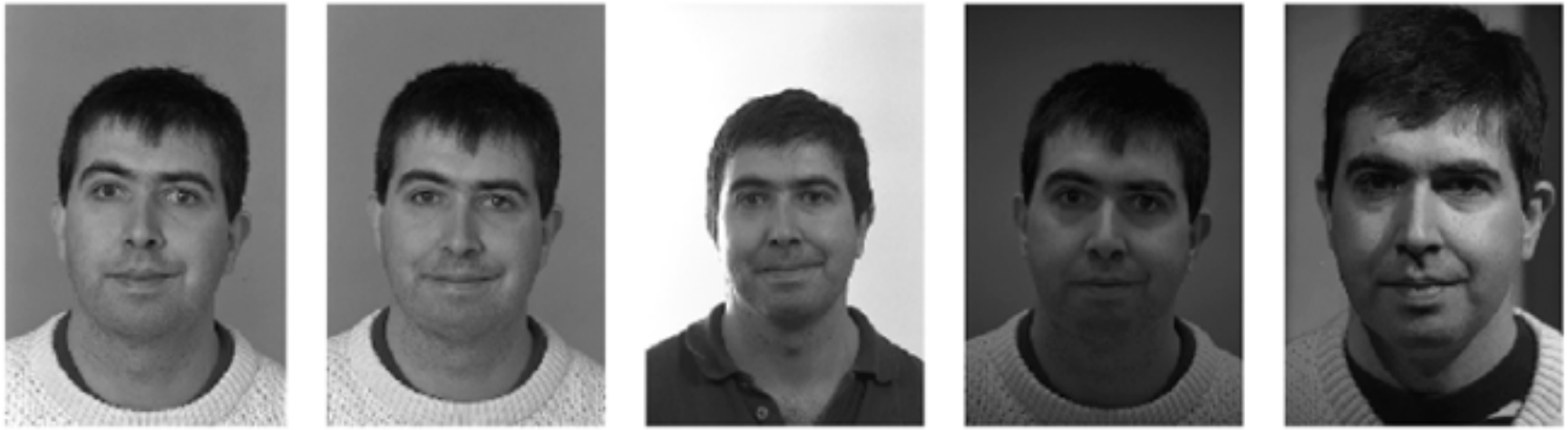
Азиат

Бородат

Улыбается

- Что мы можем сказать про этого человека?
- Атрибуты – «типичные» характеристики объекта
- Для человека - пол, возраст, раса, борода, усы, улыбка, очки и т.д.

FERET (1995-2000)

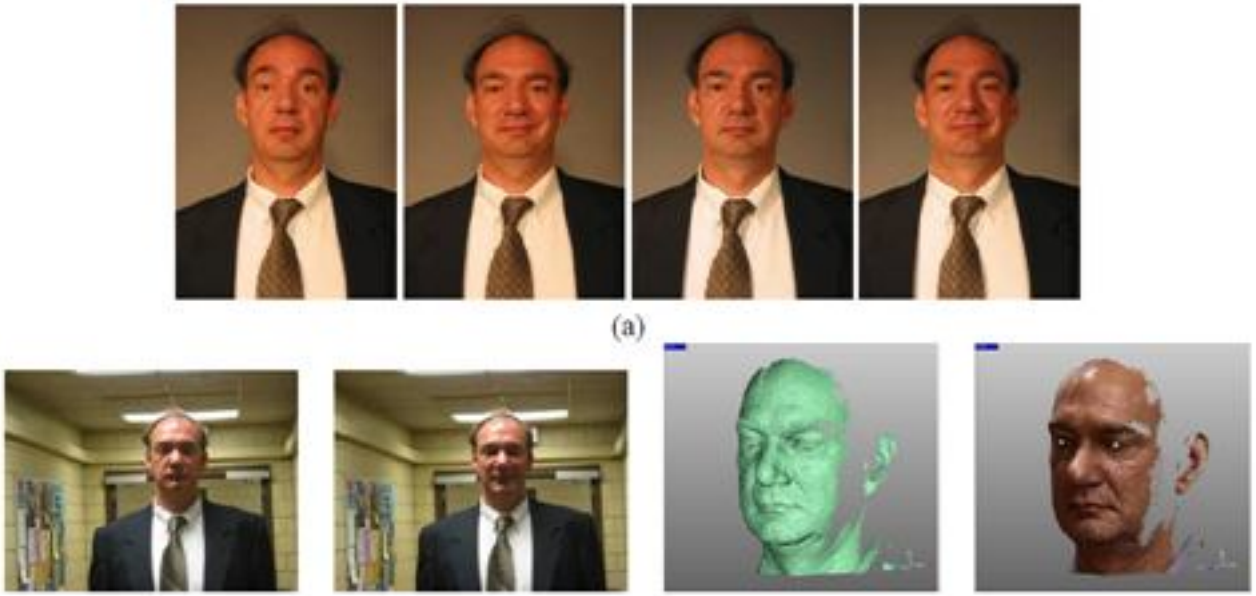


fa fb duplicate I fc duplicate II

- 14000 изображений, 1200 человек
- 2 фронтальных фотографии с разным выражением лица (fa, fb)
- Иногда одна фотография другой камерой и освещением
- Несколько дупликатов (фотографий того же человека через 1-2 года после первой съёмки)

P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms", PAMI 2000.

FRGC



- Face Recognition Grand Challenge (для Face Recognition Vendor Test)
- Идея: исследовать возможность повышения качества распознавания за счёт улучшенных данных разной природы (контролируемые, неконтролируемые, 3д), высокое разрешение
- 220 человек, 12К изображений в обучающих данных
- 465 человек в валидационной выборке (год спустя)

P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. **Overview of the Face Recognition Grand Challenge**. In CVPR, 2005.

Labeled Faces in the Wild (LFW)

Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller.
Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. *University of Massachusetts, Amherst, Technical Report 07-49*, 2007.

Labeled Faces in the Wild

- 5749 человек, 12К изображений, 1680 человек по 2 и более фотографии, остальные – по одной
- Фотографии разрешения 250*250, JPEG
- Неконтролируемые условия, очень разные фоны, позы, разное время съёмки

- “In the wild” сейчас очень популярный термин, обозначающий работу с реальными изображениями без ограничений

PubFig

Публичные люди:

- Политики
- «Звезды»

Большая коллекция

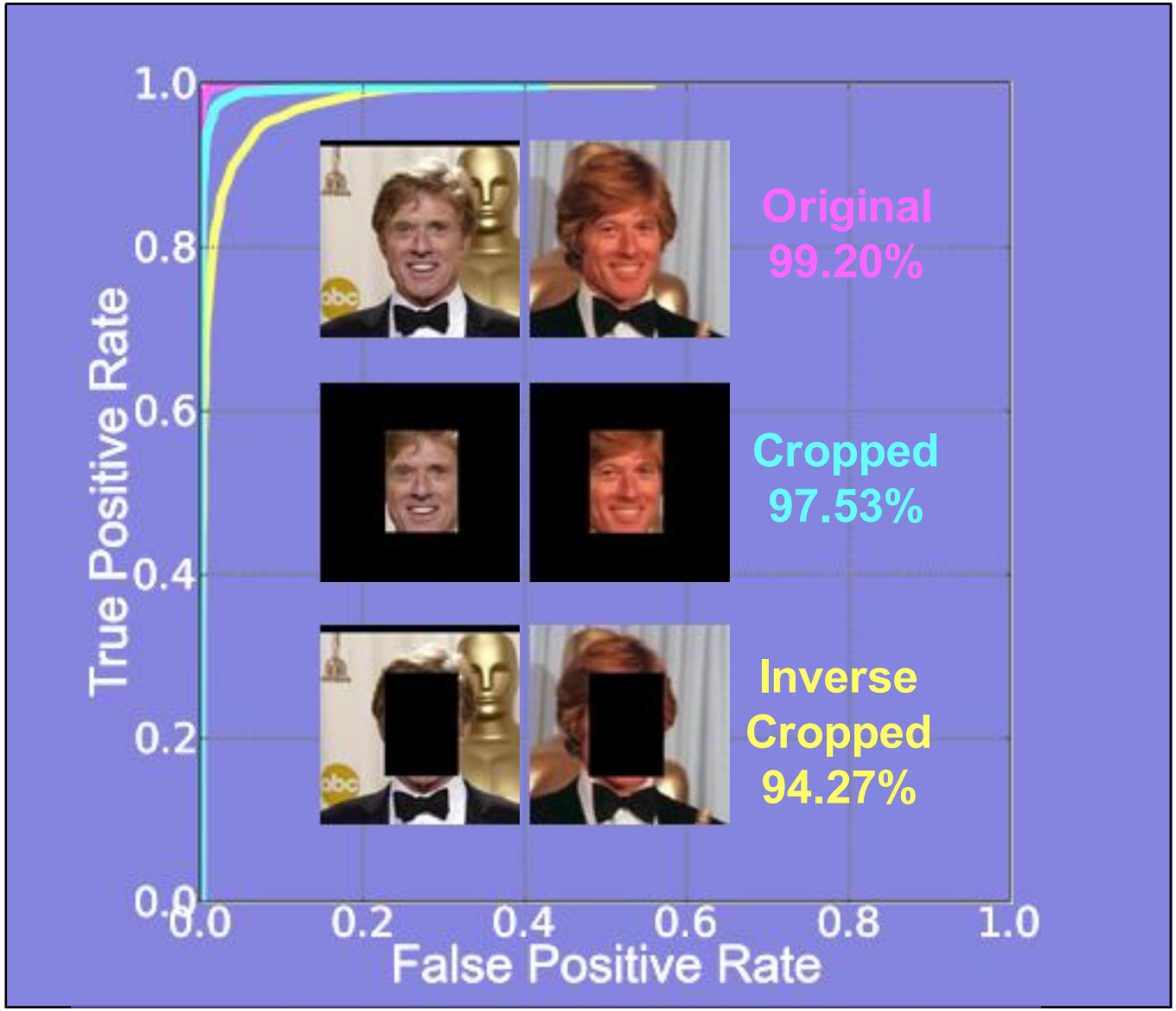
- 60,000 изображений
- 200 человек
- 300 изображений на человека

Поднаборы:

- Поза
- Освещение
- Выражение лица

http://www.cs.columbia.edu/CAVE/databases/pubfig/

Распознавание человеком



Нормализация изображения лица

- Обычно лицо нормализуется по центрам глаз
- Центры глаз должны быть на одной строке и с фиксированным расстоянием
- «Идеал» – построить 3D модель, и получить «ортографическое» фронтальное изображение

Верификация

Изображения Признаки → Верификация

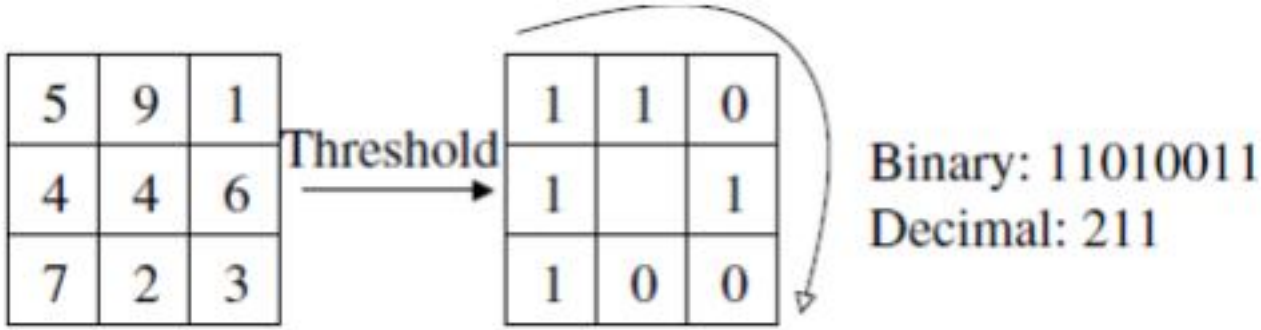
RGB
HOG
LBP
SIFT
…

RGB
HOG
LBP
SIFT
…

Разные

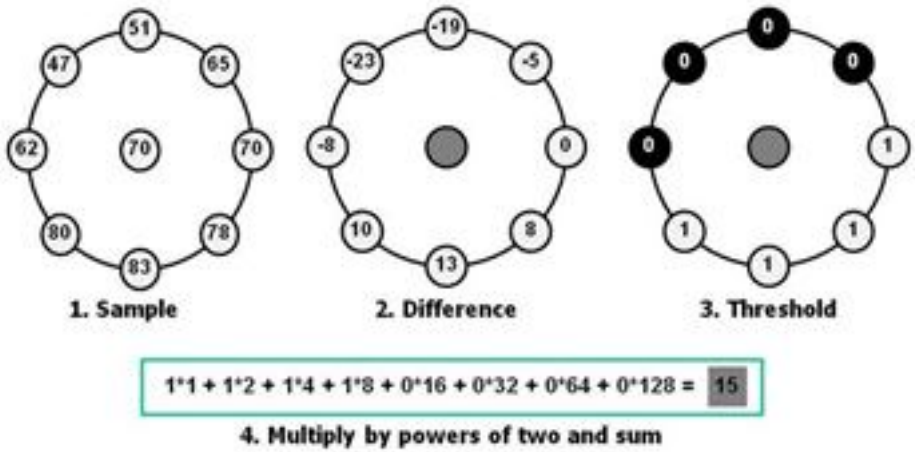
- Все виды признаков
- Все виды метрик
- Все виды классификаторов

Local Binary Patterns



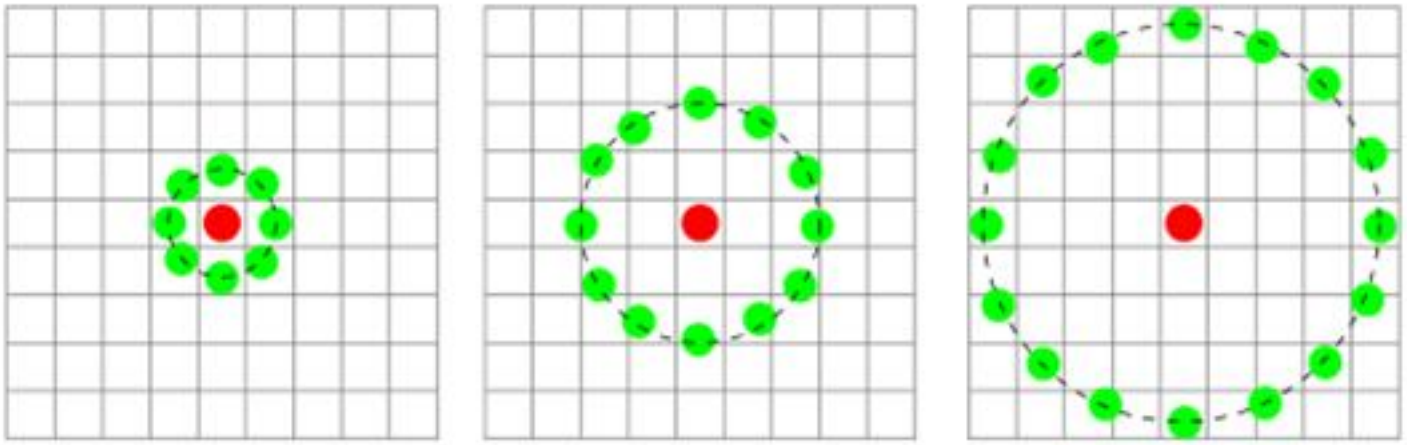
The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, if\ x \geq 0; \\ 0, otherwise. \end{cases}$$



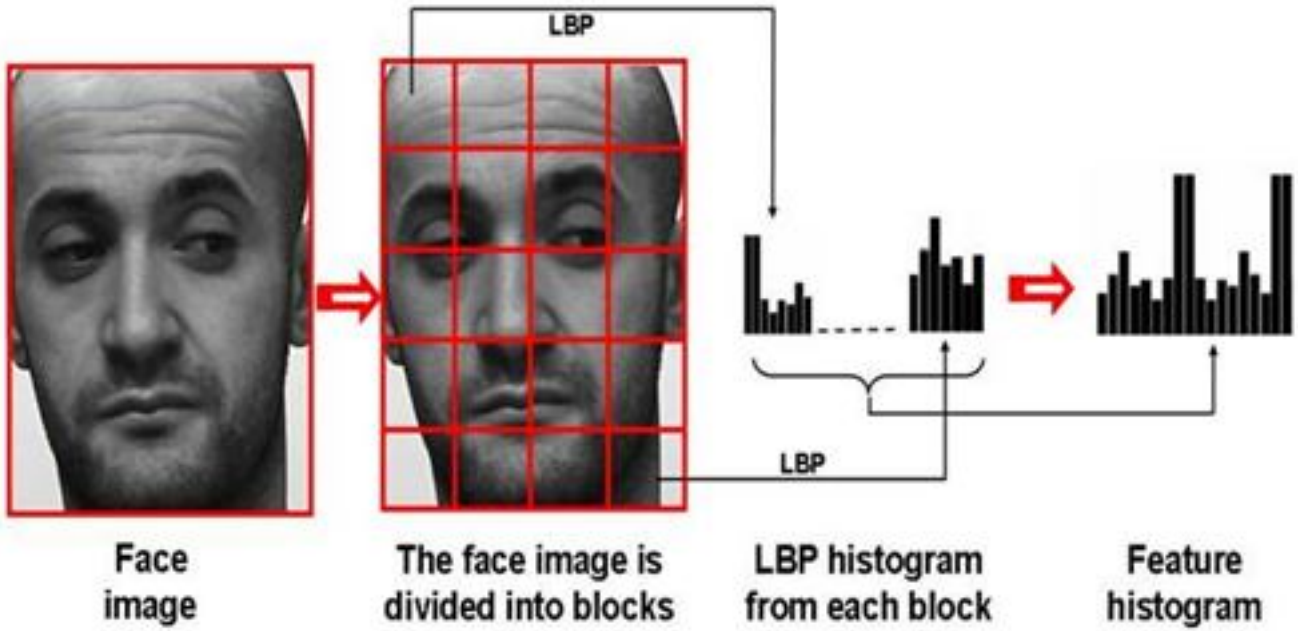
T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", ICPR 1994

Расчёт LBP



Может потребоваться интерполировать значения в точках, необходимых для расчёта LBP

Применение LBP



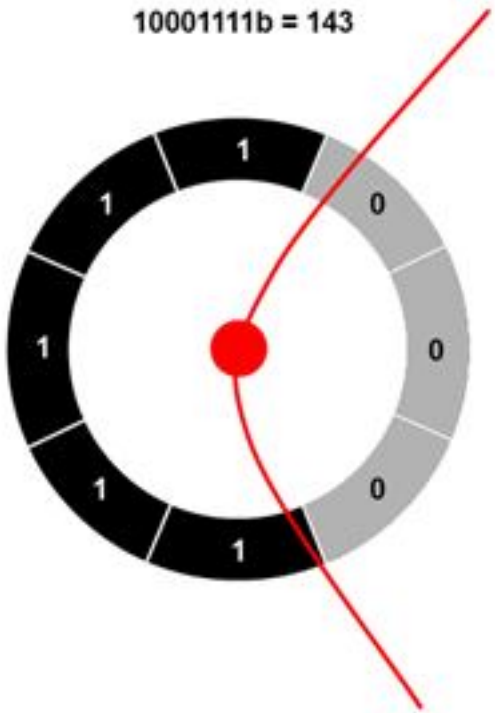
- Изображение разбивается на области. В каждой области применяются LBP операторы к каждому пикселю. Строится гистограмма.
- Объединение гистограмм – LBP дескриптор для изображения.
- Для пары изображений считается разность дескрипторов по какой-нибудь метрике (например, Хи-квадрат)

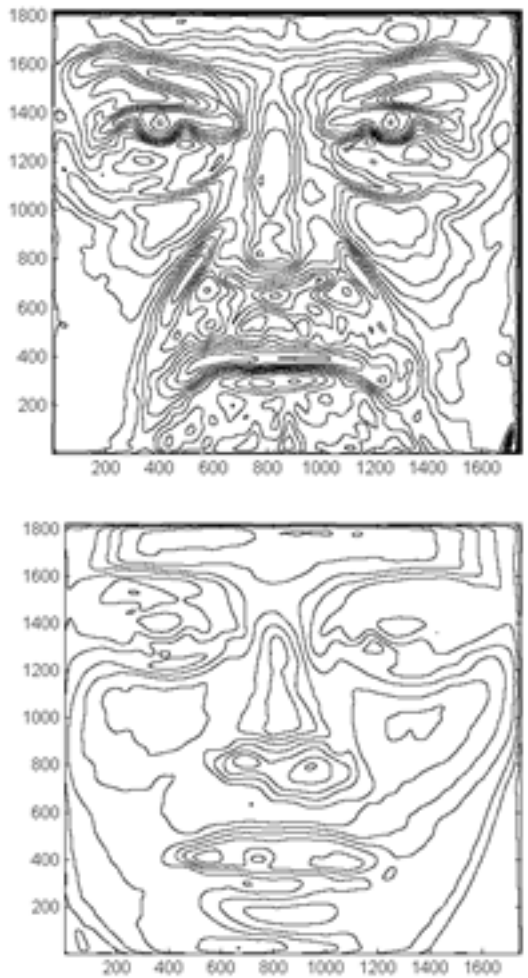
Ahonen, T., Hadid, A. and Pietikäinen, M. (2006), Face Description with Local Binary Patterns: Application to Face Recognition. IEEE PAMI 28(12):2037-2041.

Uniform LBP

- Рассмотрим количество переходов:
  - 00000000 (0 transitions)
  - 01110000 (2 transitions)
  - 11001111 (2 transitions)
  - 11001001 (4 transitions)
  - 01010010 (6 transitions)
- Однородными LBP называются такие коды, в которых переходов <= 2
- Однородные составляют от 80 до 90% всех кодов
- Неоднородные коды иногда объединяют в один
- Для кода из 8 бит
  - 256 кодов всего
  - 58 однородных (59 разных меток)

Интерпретация однородных LBP





- Коды с 2мя переходами - кодирование изолиний

Аттрибуты

Мужчина

Азиат

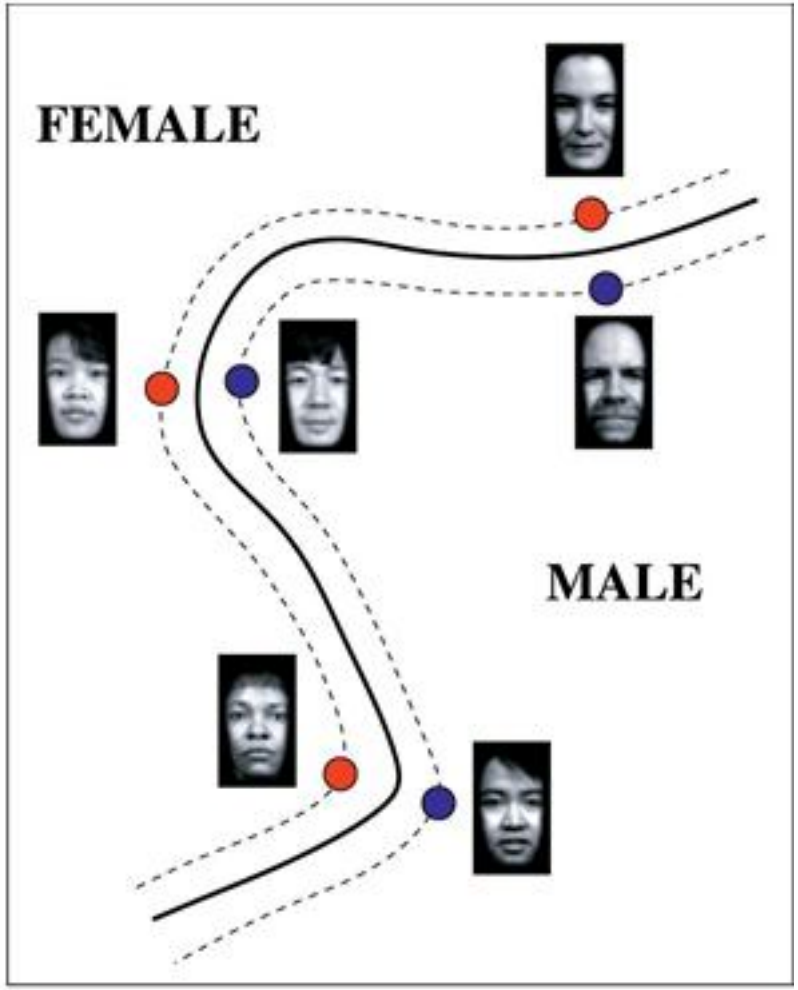
Бородат

Улыбается

- Самый сложный атрибут – возраст, ключевая проблема – данные

SVM-подход

- Уменьшенные изображения до 12*21 пикселей (thumbnails)
- SVM-RBF классификатор
- 300 опорных векторов
- FERET база
- Точность – 96.6%



B. Moghaddam and M.-H. Yang, “Learning gender with support faces,” *TPAMI*, vol. 24, no. 5, pp. 707–711, 2002

«Boosting Sex»

- Применим “boosting” подход для классификации пола лиц
- Опираемся на классификатор «позы лица», использующий сравнение интенсивности двух пикселей в изображении

S. Baluja and H. Rowley, “Boosting sex identification performance,” *IJCV*, 2007

Thank you for evaluating AnyBizSoft PDF Splitter.

A watermark is added at the end of each output PDF file.

To remove the watermark, you need to purchase the software from

http://www.anypdftools.com/buy/buy-pdf-splitter.html