

# Компьютерное зрение. Введение

Лекция 1

Что это такое и почему это сложно?



# Задача компьютерного зрения

- Понять, что запечатлено на изображении



Мы видим

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Компьютер видит

Source: S. Narasimhan

# Тест Тьюринга

- «To see means to know what is where by looking»
  - David Marr, Vision, 1982
- «Тест Тьюринга» — компьютер должен ответить на любой вопрос об изображении, на который может ответить человек

# Вопросы про изображение

- Что и где находится на изображении? — задача **детекции**
- Вопрос о свойстве объектов и их атрибутов — задача **классификации** в зависимости от ответа
- Какой формы и какого размера объект? — задача **метрического зрения**, сложной версией которой является **построение 3-D модели** объекта

# Зрение — источник

- семантической информации о мире
- метрической информации о трехмерном мире

# Семантическая информация



# Классификация сцены



# Признаки объектов



# Выделение границ объектов

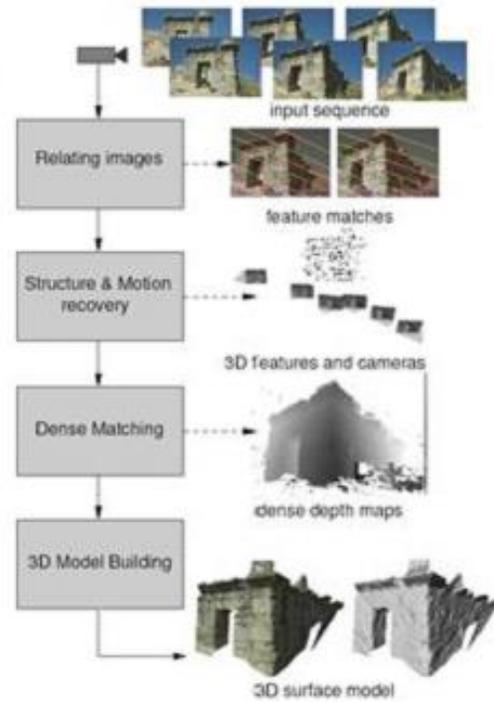


# Измерения

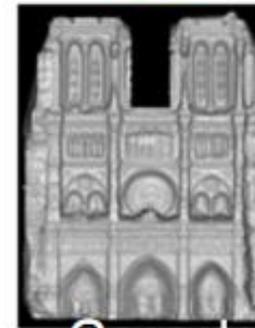
Стерео  
реконструкция



Структура из  
движения



Моделирование по  
пользовательским  
снимкам



Slide: Svetlana Lazebnik

# Зрение... принятые названия

- Фотограмметрия (Photogrammetry)
  - Ист: измерение расстояний между объектами по 2D изображениям
- Компьютерное зрение (Computer vision)
  - Восстановление 3D структуры по 2D изображения.
  - Шире: принятие решений о физических объектах, основываясь на их изображениях
- Машинное зрение (Machine vision)
  - Решение промышленных, производственных задач (исторически)
- Обработка изображений (Image processing)
  - На входе и выходе изображение
- Анализ изображений (Image analysis)
  - Фокусируется на работе с 2D изображениями
- Распознавание образов (Pattern recognition)
  - Распознавание, обучение на абстрактных величинах, полученных в том числе и из изображений

# Зачем?

- Наглядное применение многих математических методов анализа данных
- Много практических применений
- Много нерешенных задач
  - Стимул для развития методов анализа данных и высокопроизводительных вычислений
- Сложно
  - 25% мозга человека отвечает за зрение
  - «ИИ-полная» задача – решение задачи зрения на уровне человека равносильно решению задачи искусственного интеллекта

Почему зрение — это сложно?

# Точка наблюдения

---



Michelangelo 1475-1564



slide credit: Fei-Fei, Fergus & Torralba

# Освещение



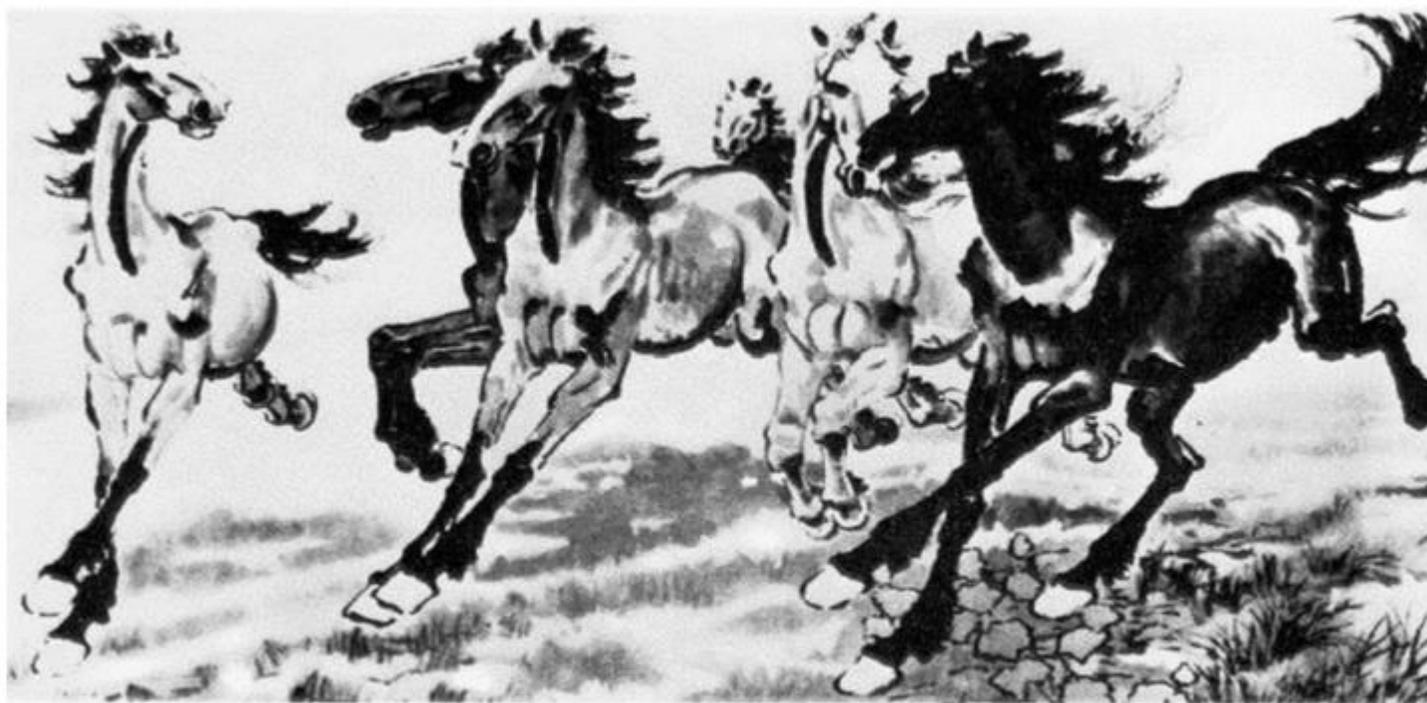
image credit: J. Koenderink

# Масштаб



Slide credit: Fei-Fei, Fergus & Torralba

# Деформация

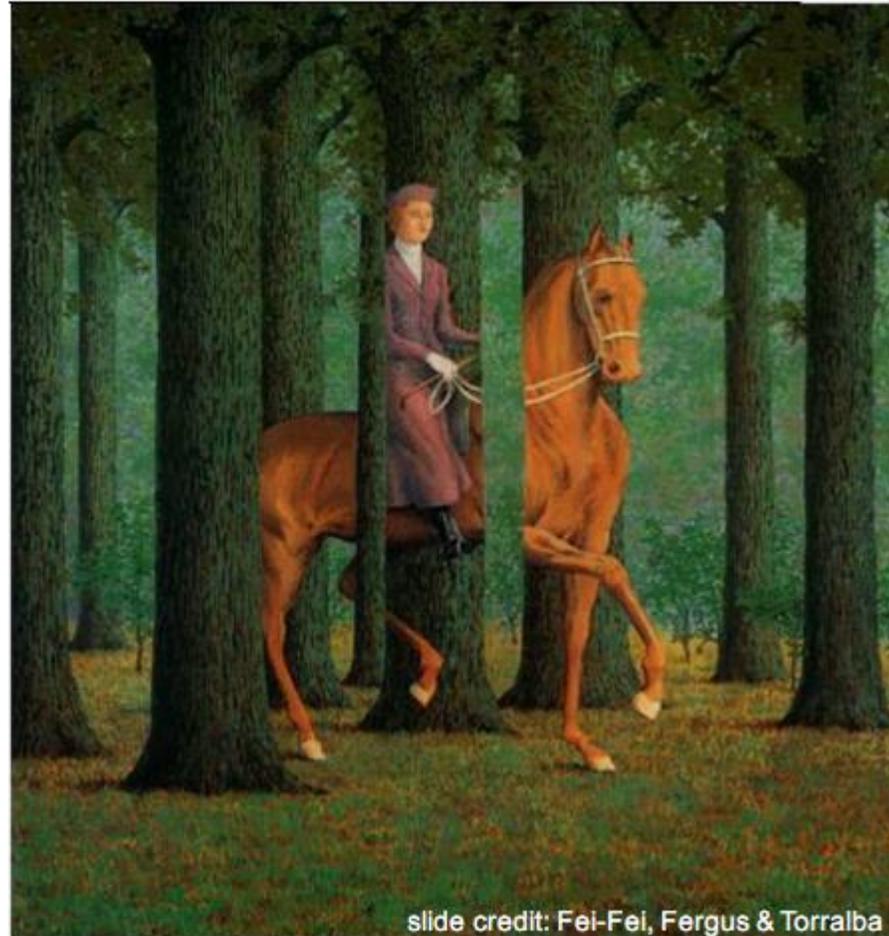


Ху, Вейхонг 1943

Slide credit: Fei-Fei, Fergus & Torralba

# Перекрытие

Magritte, 1957

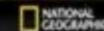


slide credit: Fei-Fei, Fergus & Torralba

# Маскировка



Spectral shrimp and commensal crab on a sea cucumber in Fiji  
Photograph by Tim Laman



© 2017 National Geographic Society. All rights reserved.

# Движение

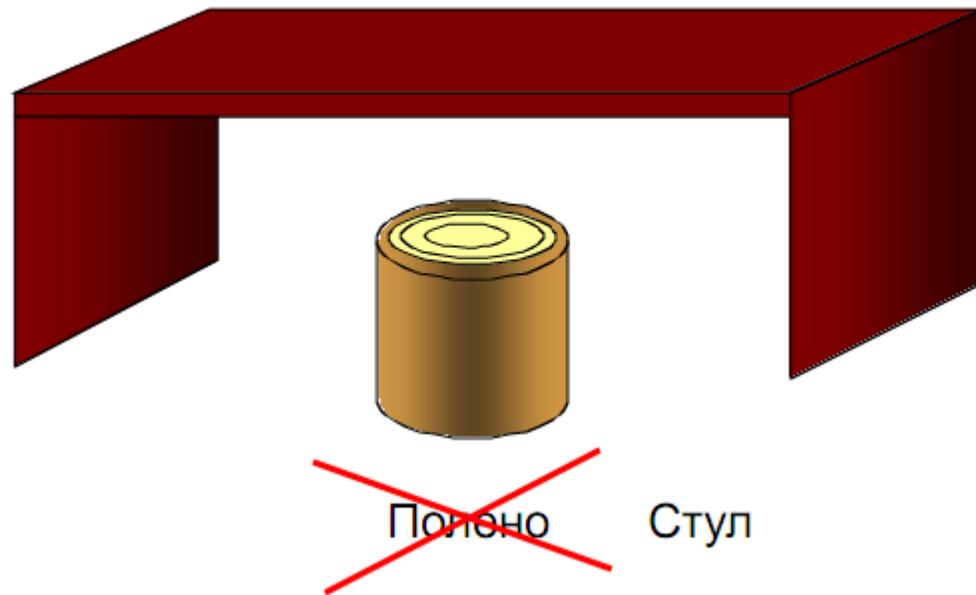


# Внутриклассовая изменчивость

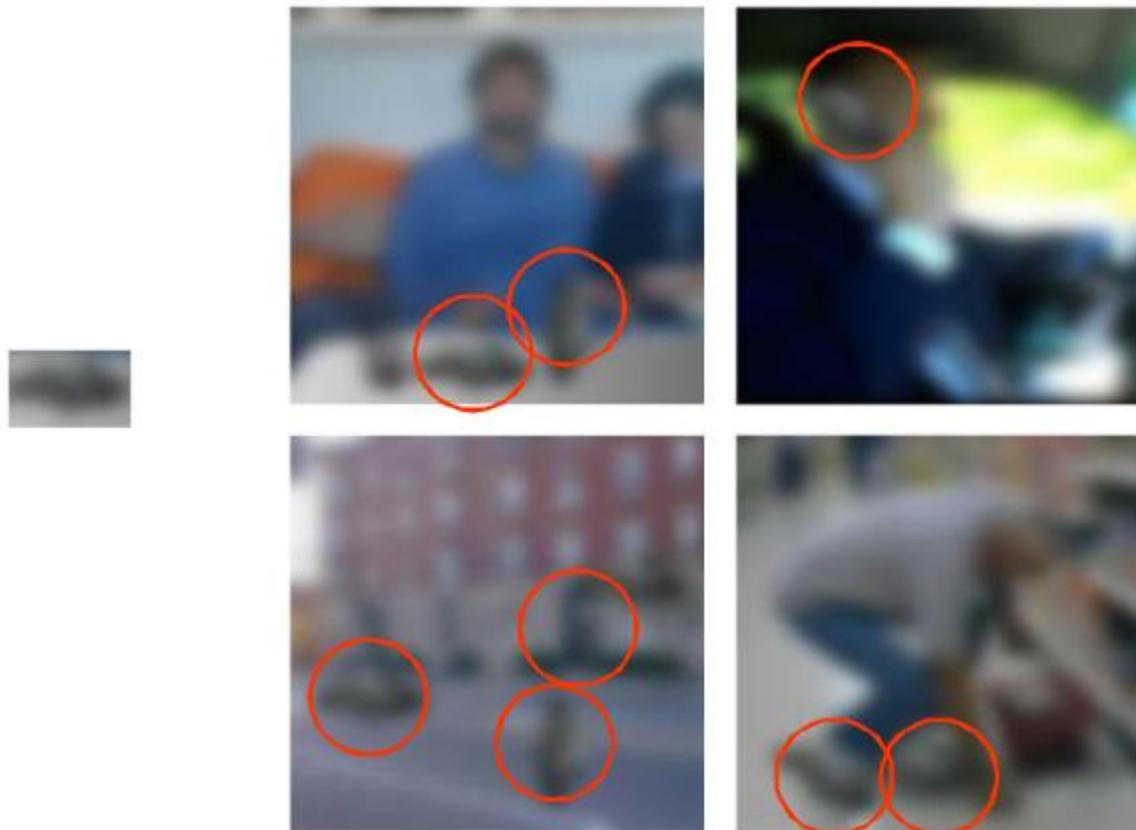


Slide credit: Fei-Fei, Fergus & Torralba

# Контекст



# Контекст и локальная неоднозначность



Slide credit: Fei-Fei, Fergus & Torralba

# Сложности или возможности?

- Изображение запутывает, но дает много подсказок
- Наша задача – интерпретировать подсказки



Image source: J. Koenderink

# Подсказки

- Цвет
- Контур
- Текстура
- Тени и освещение
- Контекст (окружение объекта)

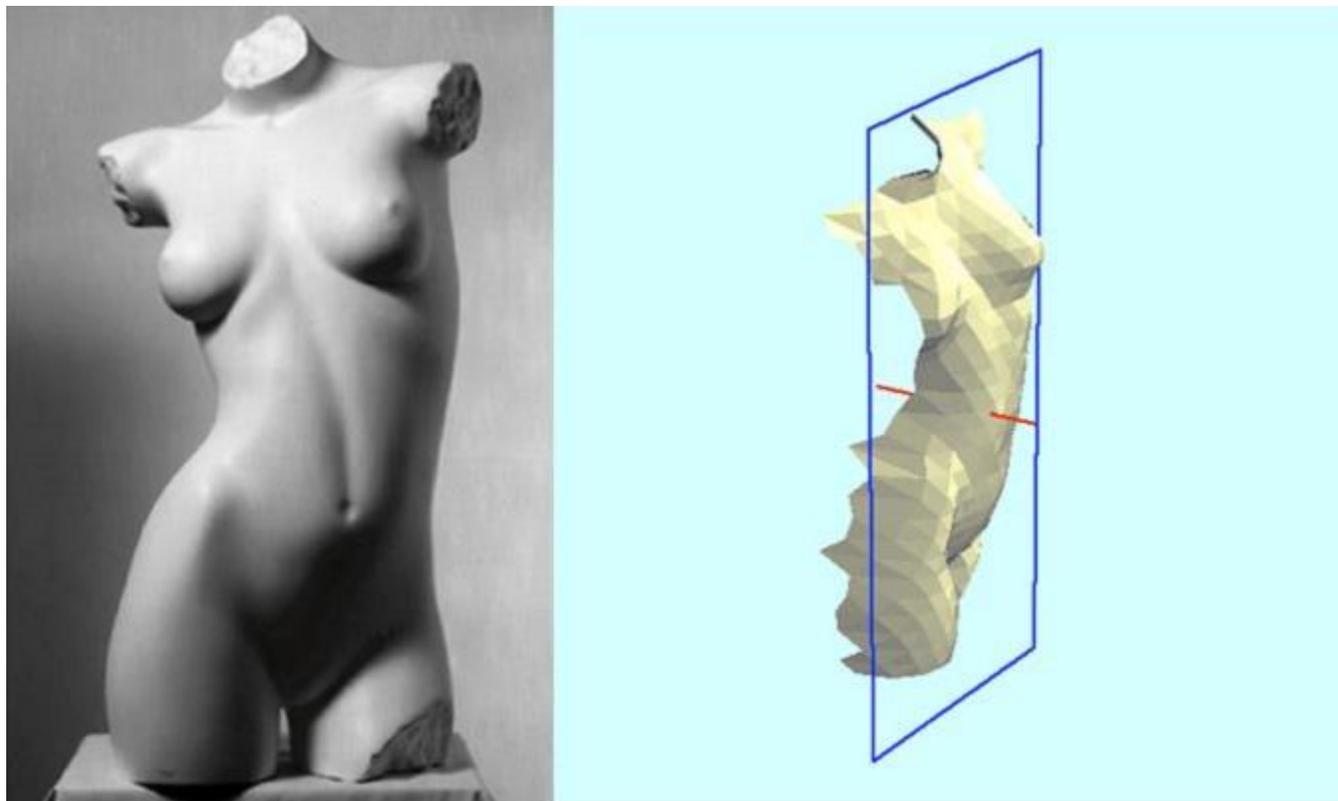
# Априорные знания об объекте

- Размеры
- Контекст
- Типичный диапазон распределения конкретной характеристики
- Эталонные примеры

# Цвет



# Тени и освещение



Source: J. Koenderink

# Отбрасываемые тени



Source: J. Koenderink

# Текстура



# Группировка

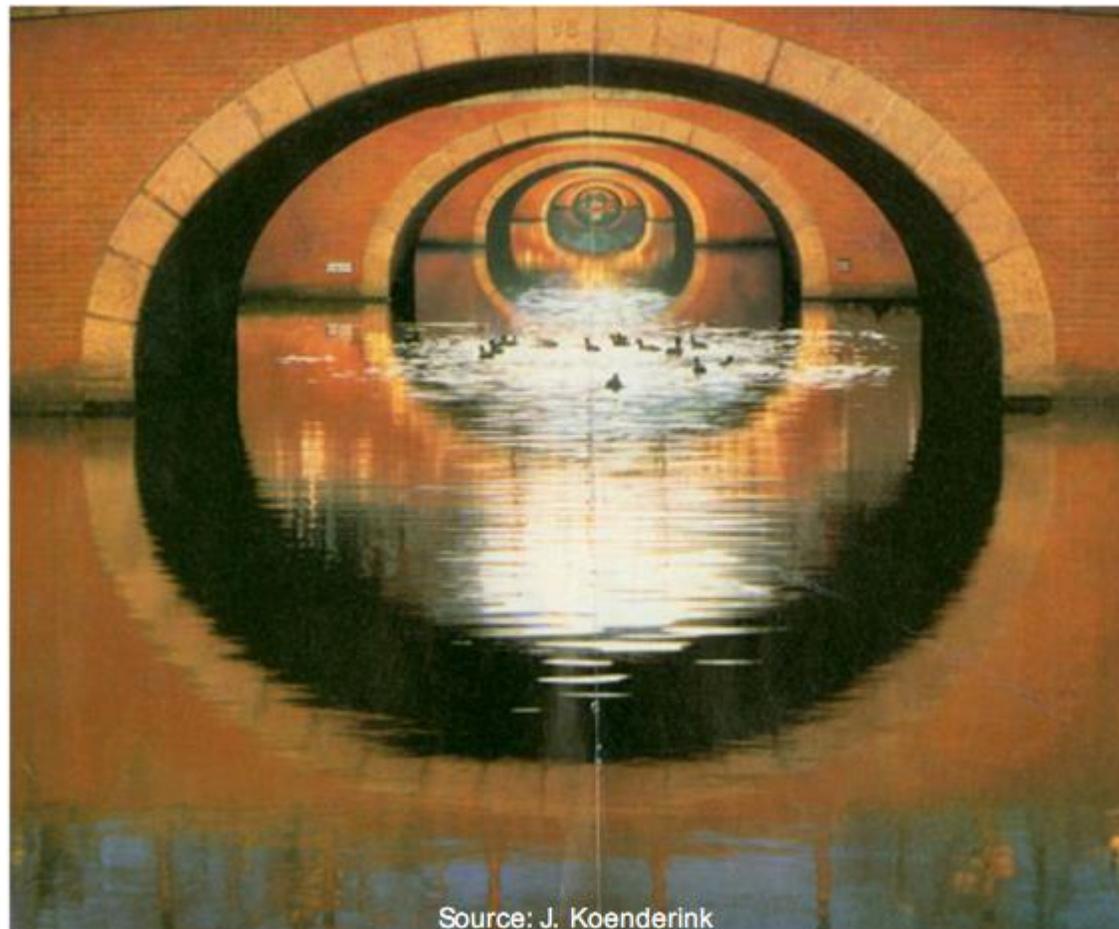


Image credit: Arthus-Bertrand (via F. Durand)

# Перспектива



# Упорядочивание по глубине



# Туман и фокусировка



# Резюме

- Зрение изначально нечеткая задача
- Разные 3D сцены могут давать одинаковое 2D изображение
- Необходимы априорные знания о структуре и свойствах мира
- Нужно сопоставлять наблюдения и априорные знания

# Обработки визуальной информации в мозге человека

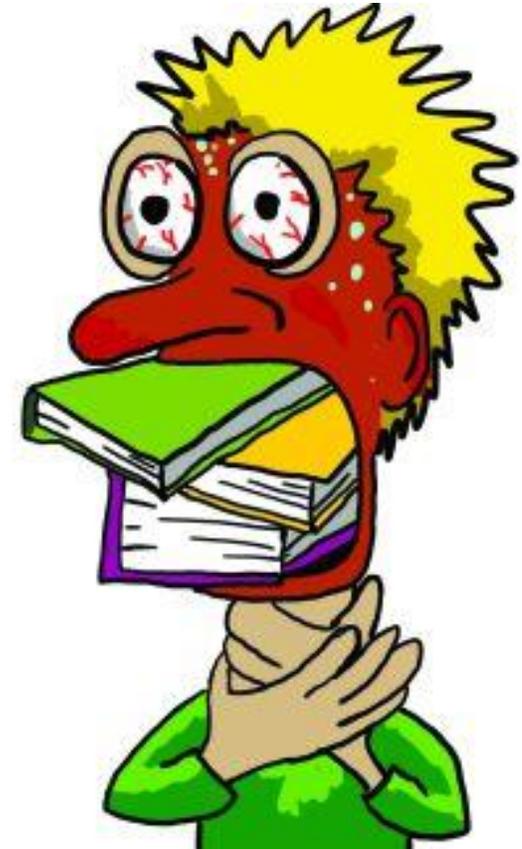
- Одними из важнейших и этапных работ в области нейрофизиологии являются работы британского нейрофизиолога Давида Марра в 1970-х, который сформулировал принципиальную структуру обработки визуальной информации в мозге человека.
- Он разделил процесс обработки на 3 этапа:
  - основной набросок
  - набросок 2.5D
  - 3D модель
- Именем Давида Марра была названа **главная премия** в области компьютерного зрения — **The Marr Prize**

Этап	Процесс
Основной набросок	Извлечение низкоуровневых (low-level) свойств изображения: направленных краев, отрезков и т.д.
Набросок 2.5D	Добавление бинокулярной информации (упорядочивание по глубине), учет текстуры и т.д.
3D модель	Распознавание объектов с учетом априорной информации, представление в 3D модели

# Сплав различных областей знаний

Что необходимо знать чтобы реализовывать алгоритмы компьютерного зрения для конкретных практических задач?

- основы обработки изображений
- методы машинного обучения
- принципы распараллеливания вычислений
- немного математики – линейную алгебру, геометрию, дифуры и теорвер
- для верности можно добавить к списку оптику, физику и теорию обработки сигналов

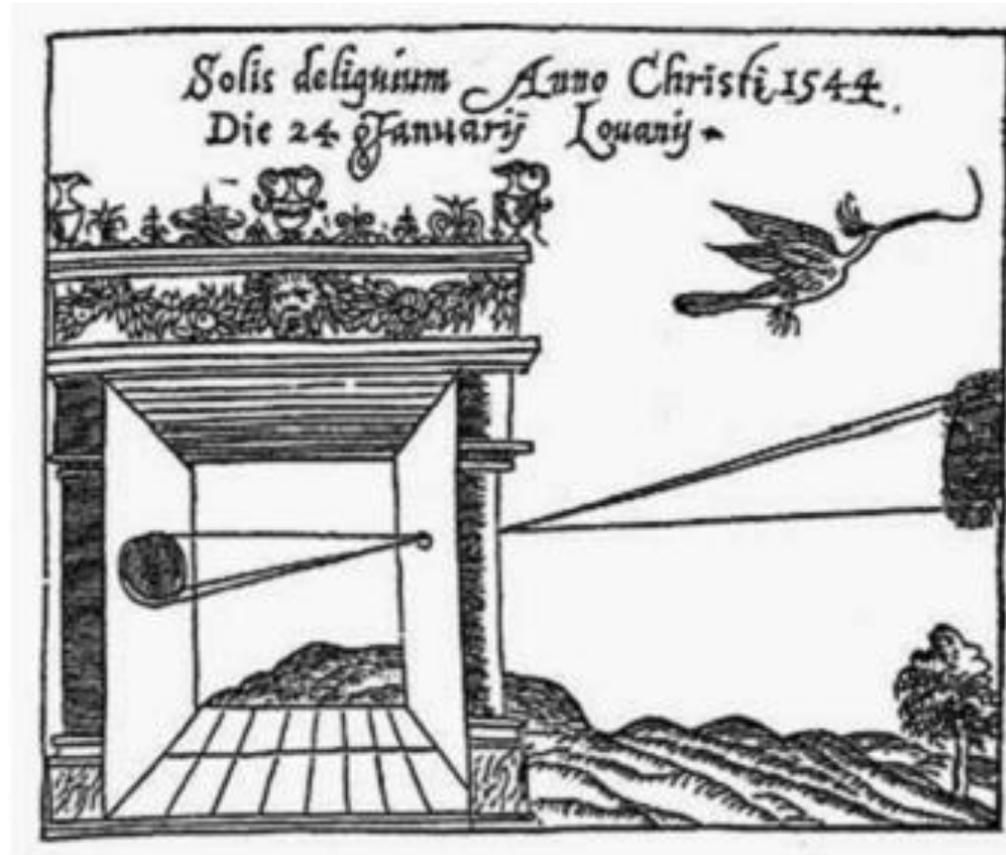


# Цифровое изображение

Изображение — объект, образ, явление, в той или иной степени подобное (но не идентичное) изображаемому или сам процесс их создания. Подобие достигается вследствие физических законов получения изображения...

Изображение оптическое – картина, получаемая в результате прохождения через оптическую систему лучей, распространяющихся от объекта, и воспроизводящая его контуры и детали.

# Камера-обскура

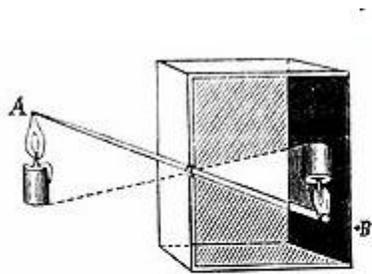


Принцип был известен еще Аристотелю (384-322 до Н.Э.)

# Принцип действия камеры-обскуры

В X веке арабский ученый Ибн ал-Хайсам (Альхазен) из Басры пользовался специальными палатками для наблюдений за затмениями Солнца.

Альхазен был первым, кто объяснил принцип действия камеры-обскуры



Схема

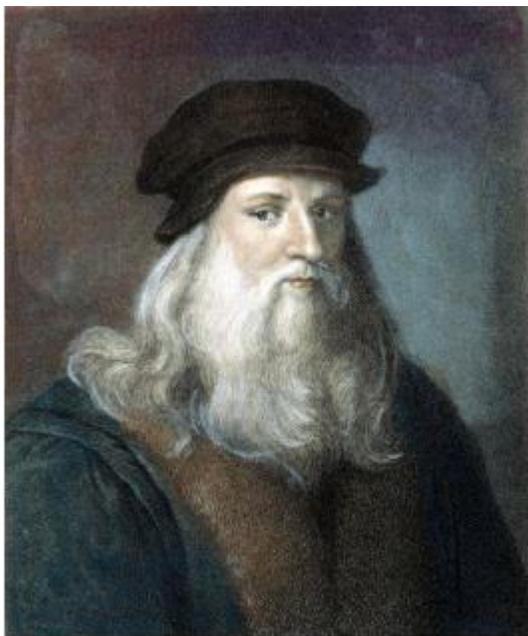


Общий вид



Её математическая модель — перспективная проекция: пучок лучей проходит через одну точку (точечное отверстие) — «центр проекции» (focal point) и на картинной плоскости (image plane) позади фокуса формируется изображение

# Леонардо Да Винчи



**Леонардо да Винчи**

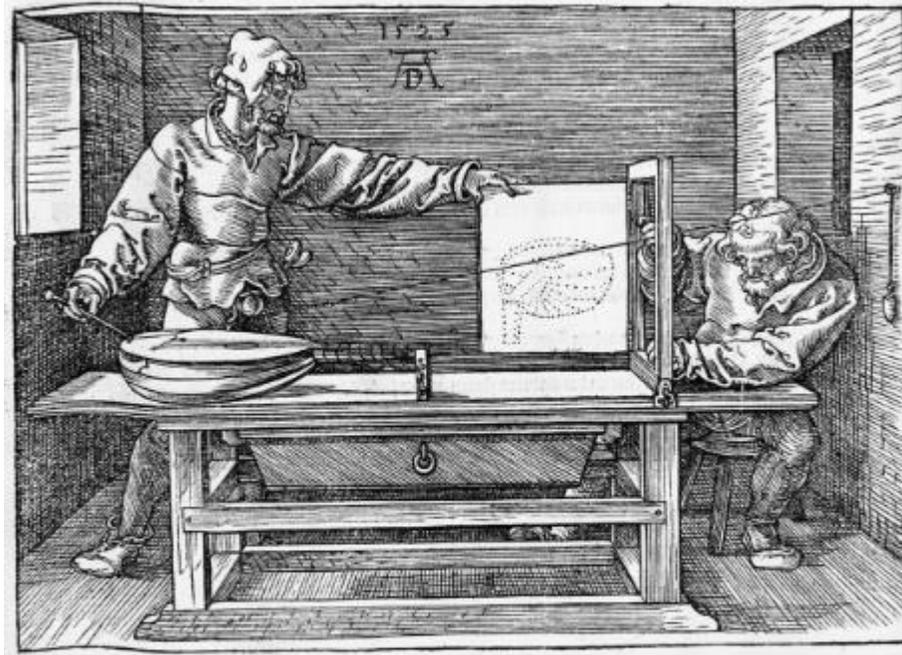


Первым использовал камеру-обскуру для зарисовок с натуры Леонардо да Винчи. Он также подробно описал её в своём «Трактате о живописи».

# Альбрехт Дюрер

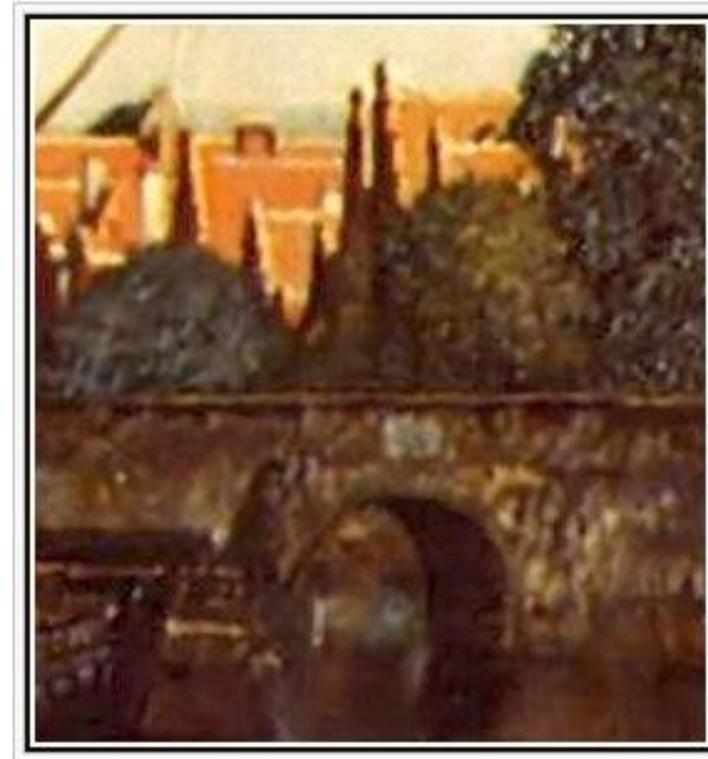


**Альбрехт Дюрер**



Художник, рисующий лютню с помощью перекрещивающихся нитей. Из «Руководства к измерению при помощи циркуля и линейки плоскостей и объёмов». Альбрехт Дюрер. 1525

## Ян Вермеер — Vermeer Jan (1632–1675)



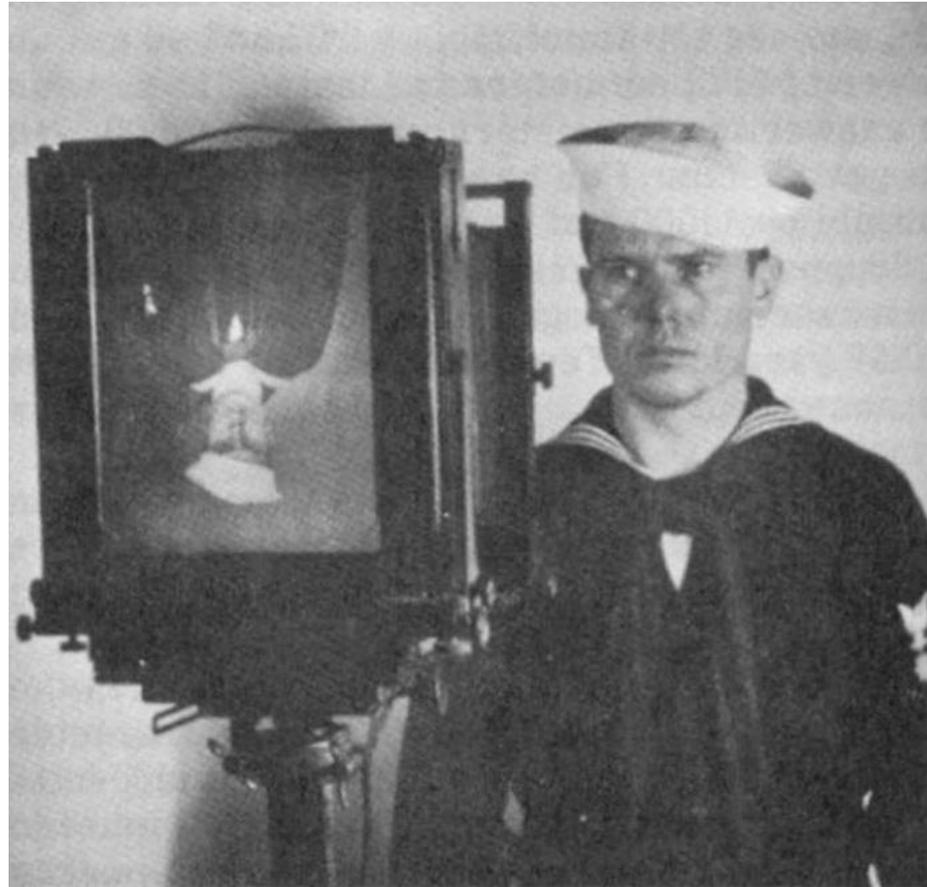
Камеры обскуры тех времён представляли собой большие ящики с системой зеркал для отклонения света, а вместо простого отверстия использовался объектив, что позволяло увеличить яркость и резкость изображения.

# Современный вид



Камера-обскура в г. Мюльхайм-на-Руре (Северный Рейн-Вестфалия)

Этот же принцип в фотографии



Формирование изображения на задней панели фотокамеры

# Одна из первых из сохранившихся фотографий



la table servie (Nicéphore Niépce, 1822) Коллекция Harlinge-Viollet

# Видео



1878 – первая скоростная  
съемка, Eadweard Muybridge



1888 – первое кино на плёнке,  
Louis Le Prince

# Начало цифровой обработки изображений

Передача иллюстраций по подводному кабелю между Лондоном и Нью-Йорком

1920 гг — Система «Бартлейн» для передачи изображений по кабелю

Задержка снизилась с недели до трех часов



Цифровое изображение, полученное с кодовой ленты на телеграфном буквопечатающем аппарате с особым шрифтом — 1921

Ранние системы «Бартлейн» кодировали 5-ю градациями яркости

В 1929г. – уже 15 градаций



Зарождение компьютерного зрения  
Начало пути  
История до наших дней

# Virtual Reality, Harvard, 1968

Айвен Сазерленд с помощью своих студентов Боба Спроула, Квинтина Фостера и Дэнни Коэна в 1968 году создал то, что считается **первым шлемом виртуальной реальности и дополненной реальности — Дамоклов Меч.**

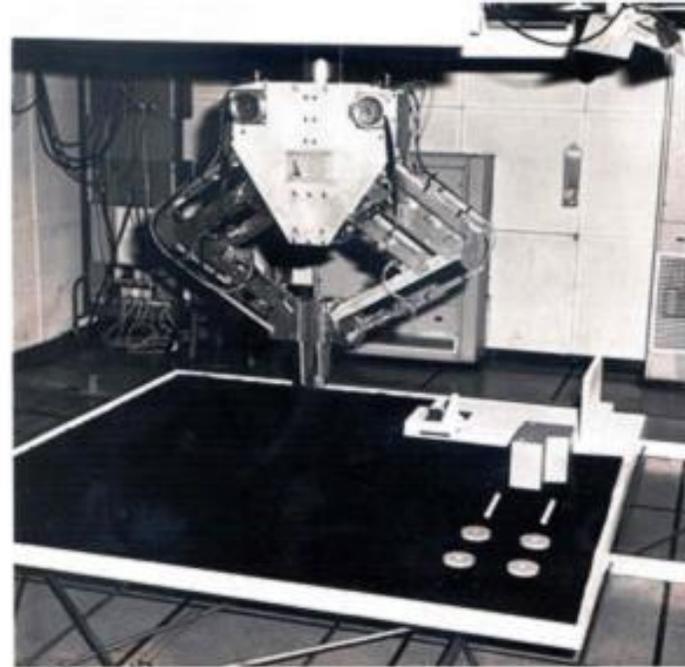
Шлем был примитивным как с точки зрения интерфейса, так и по реализму изображения, а его вес был таким большим, что он подвешивался к потолку.

Виртуальная среда состояла из простых каркасных моделей комнат



# Freddy II, 1973

- Университет Эдинбурга
- Один из первых роботов с системой машинного зрения
- 5 степеней свободы
- Умеет собирать машинки из кубиков, разбросанных по столу
- 384Кб RAM в управляющем компьютере



# Детектор лиц Viola-Jones (2001)



Первым известным алгоритмом, демонстрирующий силу ML в CV, стал алгоритм Viola-Jones (2001 г.) — надежный и быстрый (работа в режиме реального времени) детектор лиц.

## Аватар (2009) – 3D кино



Демонстрация достижений в цифровой композиции, захвате движения и т.д.

# Kinect (2010)



Первая потребительская система взаимодействия с компьютером с помощью жестов

# Практическое применение

Автоматизация обработки текстовых данных привела к революционным изменениям в организации бизнеса и жизни

Изображения дают 90% информации, но пока обрабатываются вручную

В перспективе, компьютерное зрение — один из главных компонентов робототехники

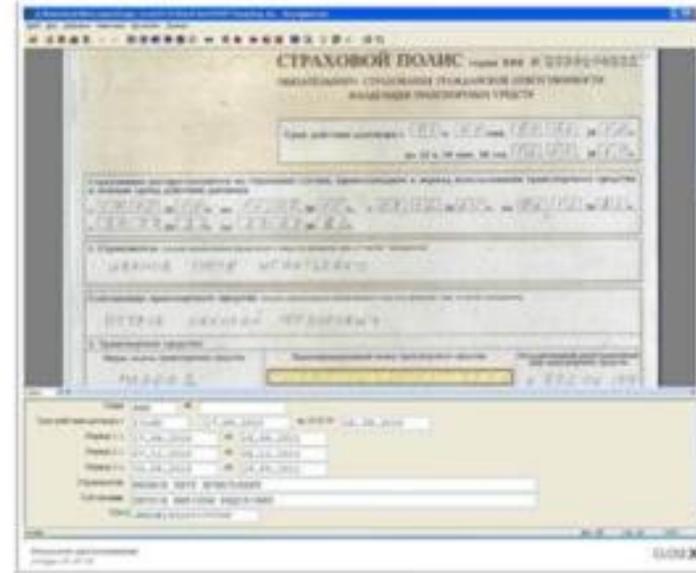


# Области применения

# Распознавание текста



FineReader, ABBYY, Россия  
<http://www.abbyy.ru/finereader/>



Cognitive Technologies, Россия  
<http://cognitiveforms.ru/products/cognitive-forms/>

# Автоматическое чтение номерных знаков и видеонаблюдение



The screenshot displays the VOCORD Traffic software interface. It features a central traffic camera feed showing a multi-lane road with several vehicles. Red lines and boxes track the movement of these vehicles. On the right side, there is a sidebar with a smaller camera view and a data entry form. At the bottom, there is a table with columns for 'Номер знака', 'Дата', 'Регистрционный номер', 'Скорость', 'Направление', 'Полоса', and 'Комментарий'. The table contains several rows of data.

Номер знака	Дата	Регистрционный номер	Скорость	Направление	Полоса	Комментарий
027206	14.08.2007 17:40:49		56	2	0	
027206	14.08.2007 17:40:58		58	2	0	
027206	14.08.2007 17:41:08	MMPK41965	79	2	2	
027206	14.08.2007 17:41:17	MMPK41965	52	2	1	
027206	14.08.2007 17:41:25	MMPK41965	43	2	0	
027206	14.08.2007 17:41:33	MMPK41965	56	2	0	

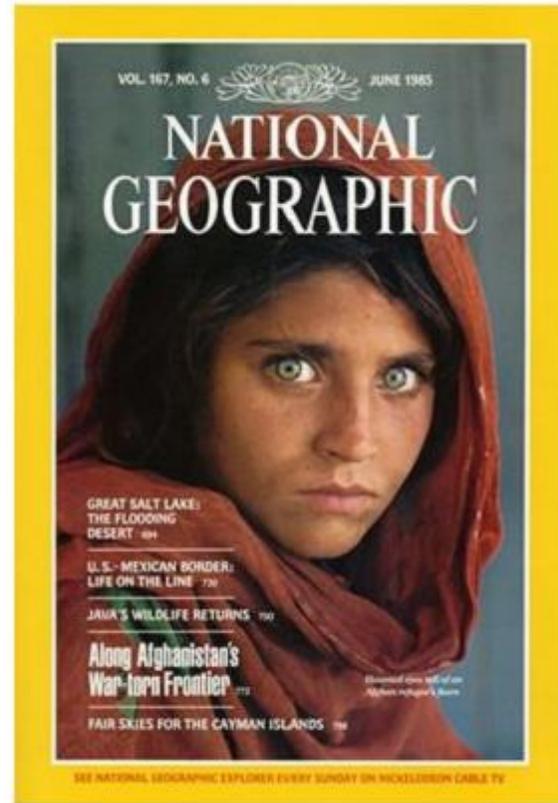
VOCORD Traffic, Vocord, Россия  
<http://www.vocord.ru/218/>

# Анализ движущихся объектов на видео



Обнаружение изменяющихся областей видео, анализ их формы и динамики изменения (обычно для систем безопасности)

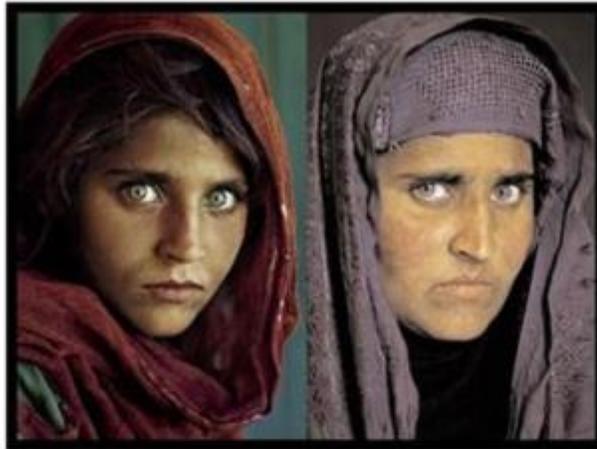
# Незнакомка



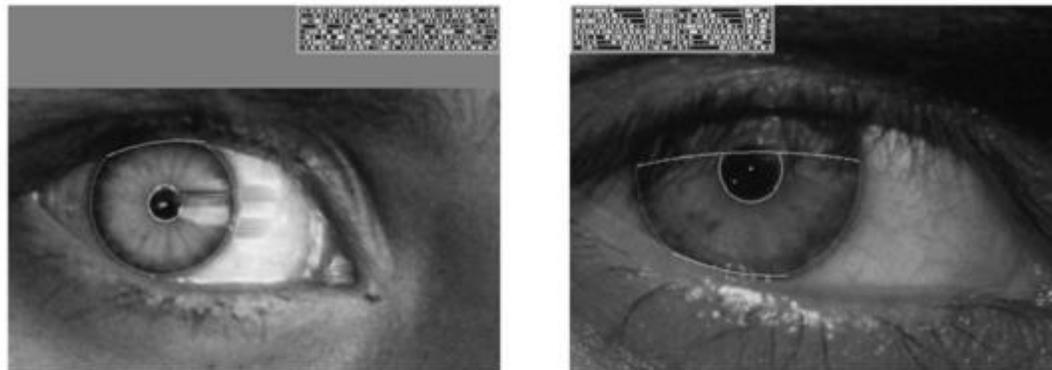
Кто она?

Source: S. Seitz

# Биометрия



“Как девушку из Афганистана идентифицировали по радужке глаза”



Source: S. Seitz

# Умные машины

The screenshot displays a website interface for mobileye products. At the top, there are navigation tabs for "manufacturer products" and "consumer products". The main header reads "Our Vision. Your Safety." Below this, a central image shows a car from a top-down perspective, with four camera views highlighted: "rear looking camera", "forward looking camera", and "side looking camera".

Below the main image, there are three product highlights:

- EyeQ Vision on a Chip**: Accompanied by an image of a microchip and a "read more" link.
- Vision Applications**: Described as "Road, Vehicle, Pedestrian Protection and more", accompanied by an image of a pedestrian and a "read more" link.
- AWS Advance Warning System**: Accompanied by a circular graphic showing a car and a "read more" link.

On the right side of the page, there is a "News" section with two articles:

- "mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System"
- "Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end"

Below the news is an "Events" section with two items:

- "Mobileye at Equip Auto, Paris, France"
- "Mobileye at SEMA, Las Vegas, NV"

A "read more" link is located at the bottom of the events section.

# Неразрушающая диагностика

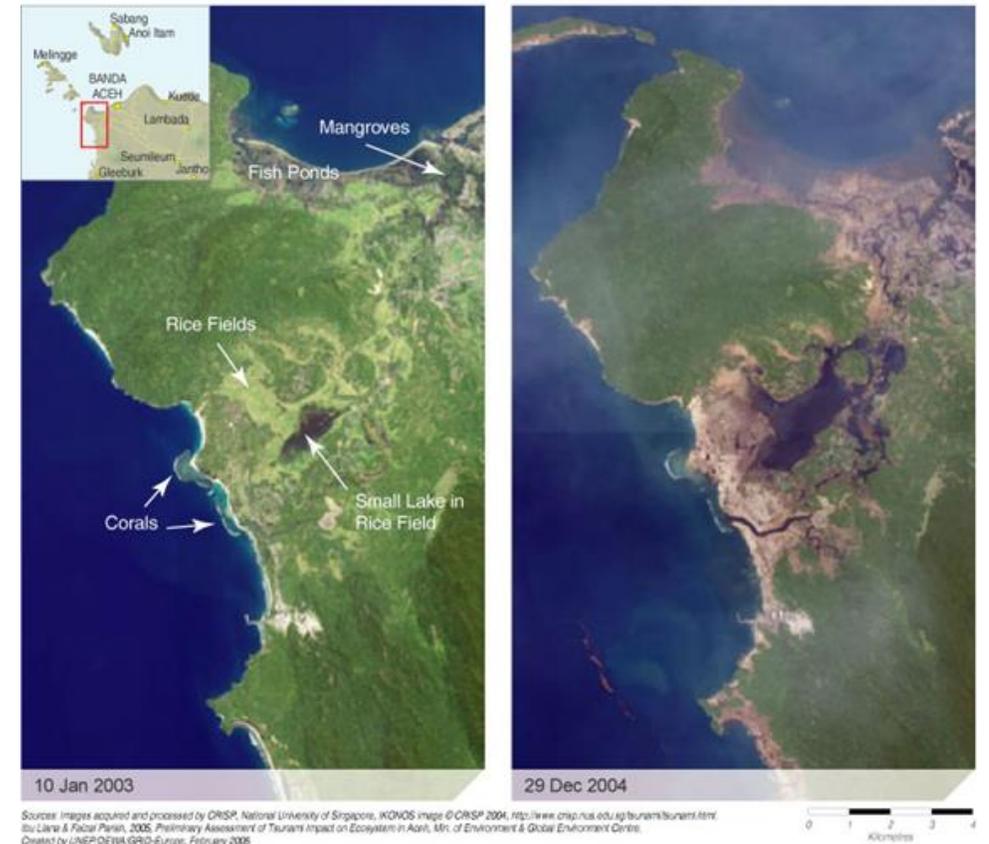


Автоматический поиск трещин в асфальте по ИК изображениям

# Анализ спутниковых снимков

- погода,
- геологические процессы (напр. таяние ледников),
- экология

## Overall Damage to Ecosystems Aceh, Sumatra, Indonesia



# Поиск конкретных объектов



# Human detection и Pose estimation

In addition to pose, 2D joints convey information about 3D shape.



Bogo, Kanazawa, Lassner, Gehler, Romero, Black, Keep it SMPL: Automatic estimation of 3D human shape and pose from a single image, ECCV'16

Слайд с презентации М. Black

# 3D модели и захват движения



*L.A. Noire*, Team Bondi/Rockstar



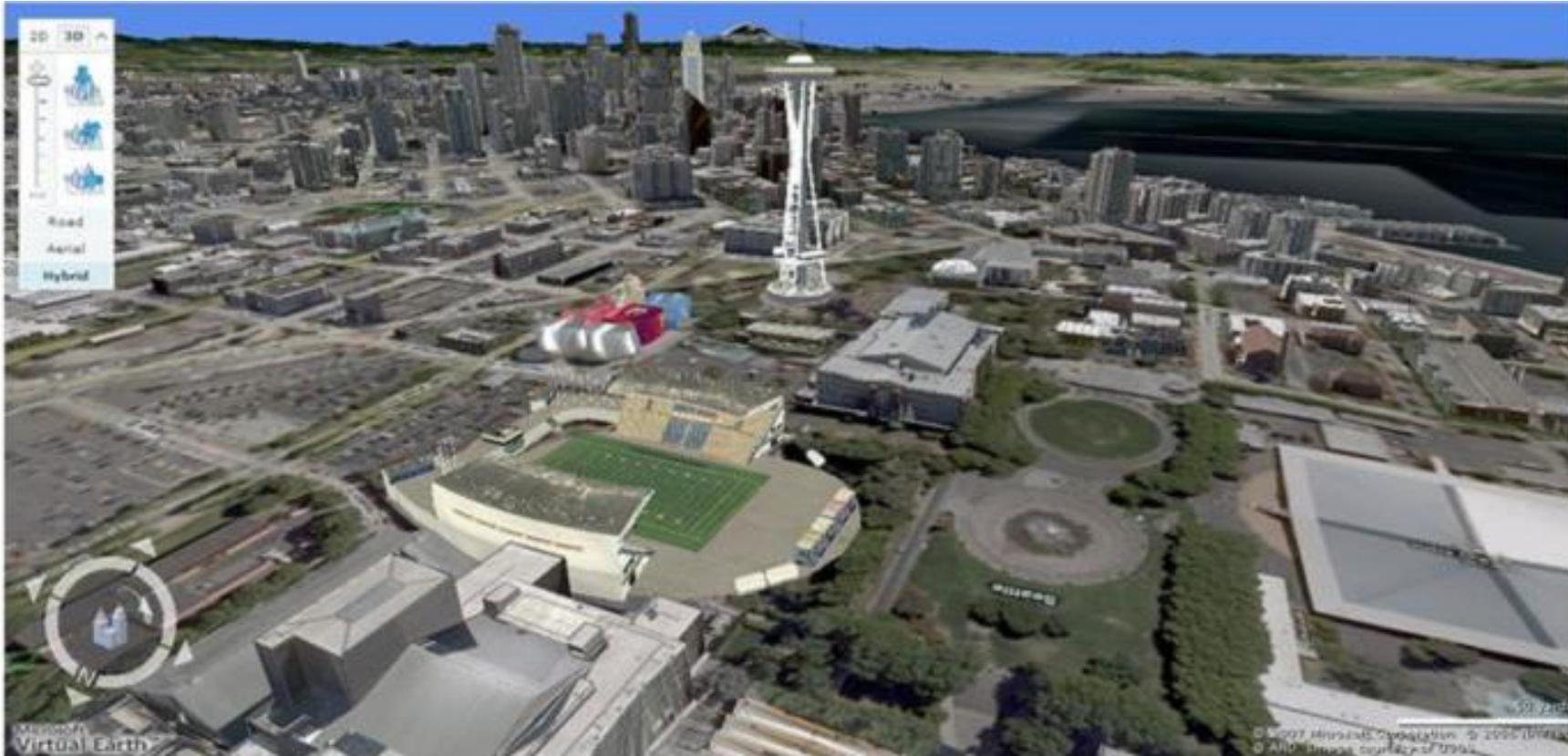
*The Matrix* movies, ESC Entertainment, XYZRGB, NRC

# Зрение в космосе



- NASA'S Mars Exploration Rover Spirit
- Системы зрения использовались для:
  - Склейка панорам
  - 3D моделирование местности
  - Поиск препятствий, определение местоположения
  - Подробнее см. "Computer Vision on Mars" by Matthies et al.

# Трёхмерные карты

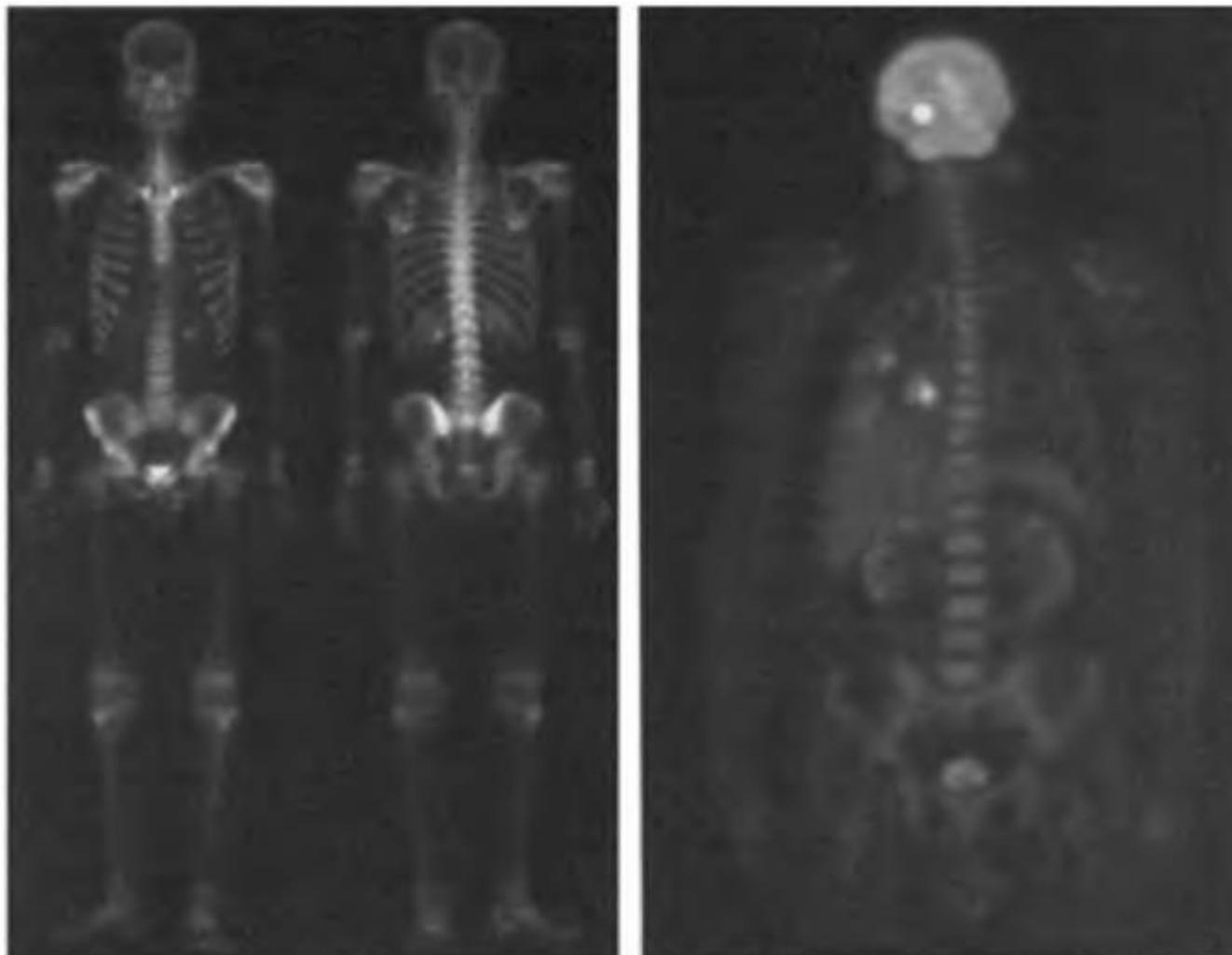


Изображение из Microsoft's Virtual Earth (аналогичные Google Earth)

# Классификация изображений по видам формирующих их ИСТОЧНИКОВ

- Видимые изображения (Оптические).
- Изображения, полученные с помощью гамма-лучей.
- Рентгеновские изображения.
- Изображения в ультрафиолетовом диапазоне.
- Акустические и ультразвуковые изображения.
- Изображения, получаемые посредством электромагнитных излучений.
- Изображения, получаемые в электронной микроскопии.
- Изображения, генерируемые компьютером.

# Изображения, полученные с помощью гамма-лучей

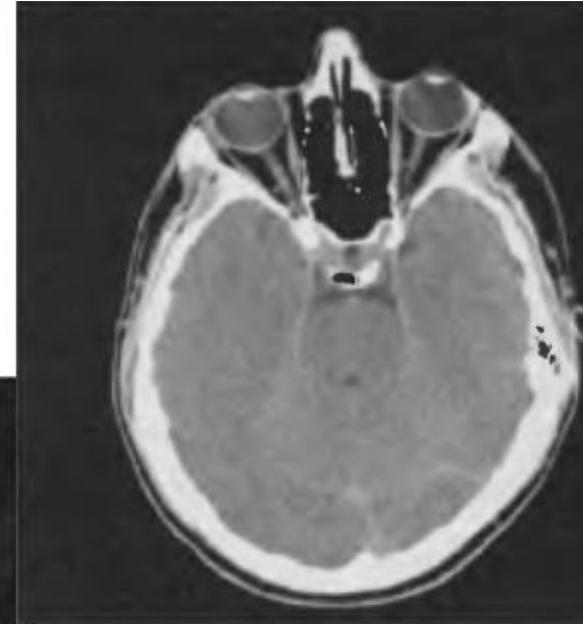


# Рентгеновские изображения



Рентгенограмма

Ангиограмма  
аорты



Томограмма  
ГОЛОВЫ

Дискретизация — перевод изображения в цифровой вид

Цифровое полутоновое изображение — матрица  $I \in (b_{ij})^{n,m}$ , элементами которой  $b_{ij}$  являются значения интенсивности света, измеренного на двумерной прямоугольной сетке.

# Разделы компьютерного зрения

## Low-level vision

Этап	Процесс
Основной набросок	Извлечение низкоуровневых (low-level) свойств изображения: направленных краев, отрезков и т.д.
Набросок 2.5D	Добавление бинокулярной информации (упорядочивание по глубине), учет текстуры и т.д.
3D модель	Распознавание объектов с учетом априорной информации, представление в 3D модели

## Mid-level vision

## High-level vision

# Примитивные операции типа предобработки

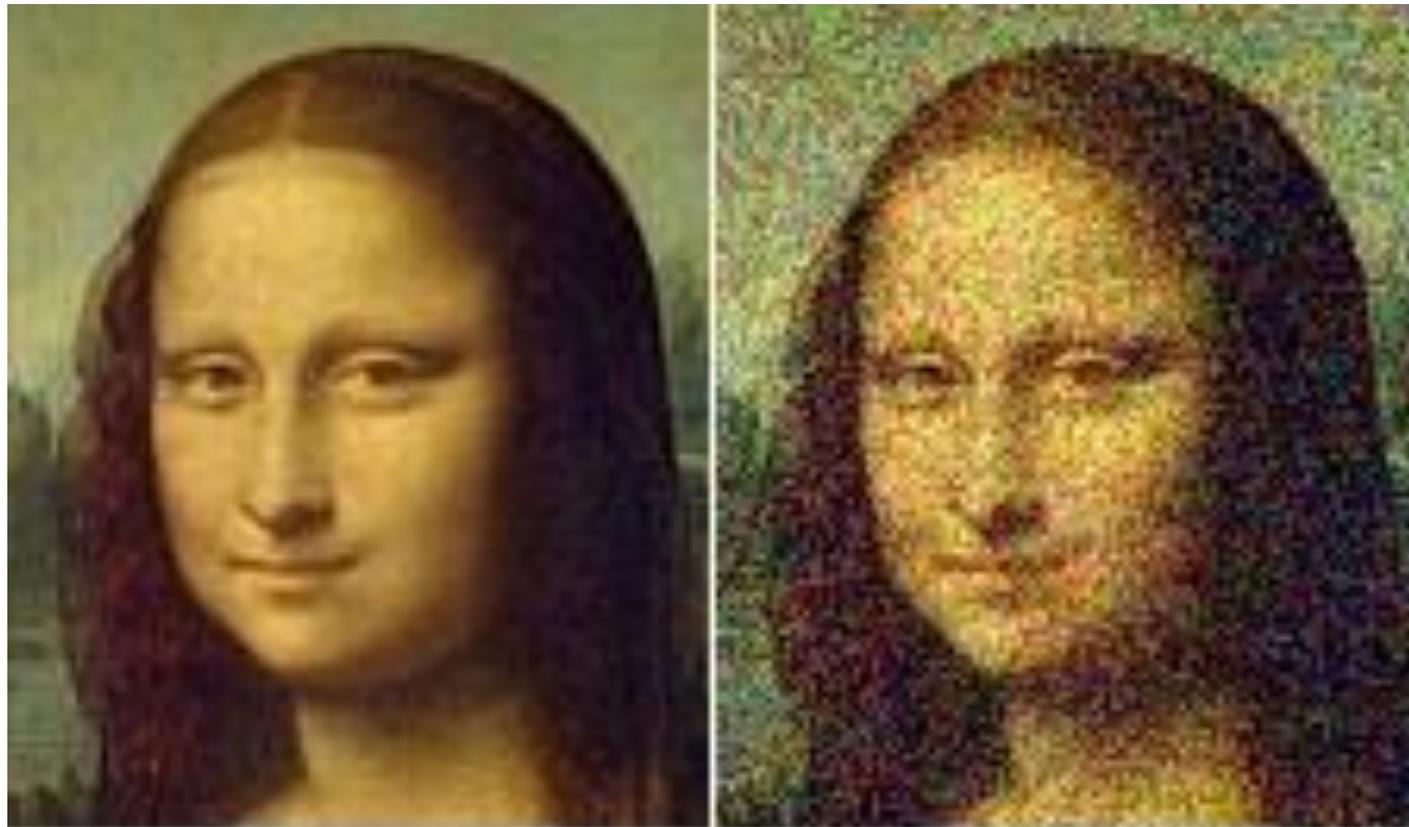
- уменьшение шума
- повышение контраста
- улучшения резкости изображений
- морфологические операции

Изображение → Изображение

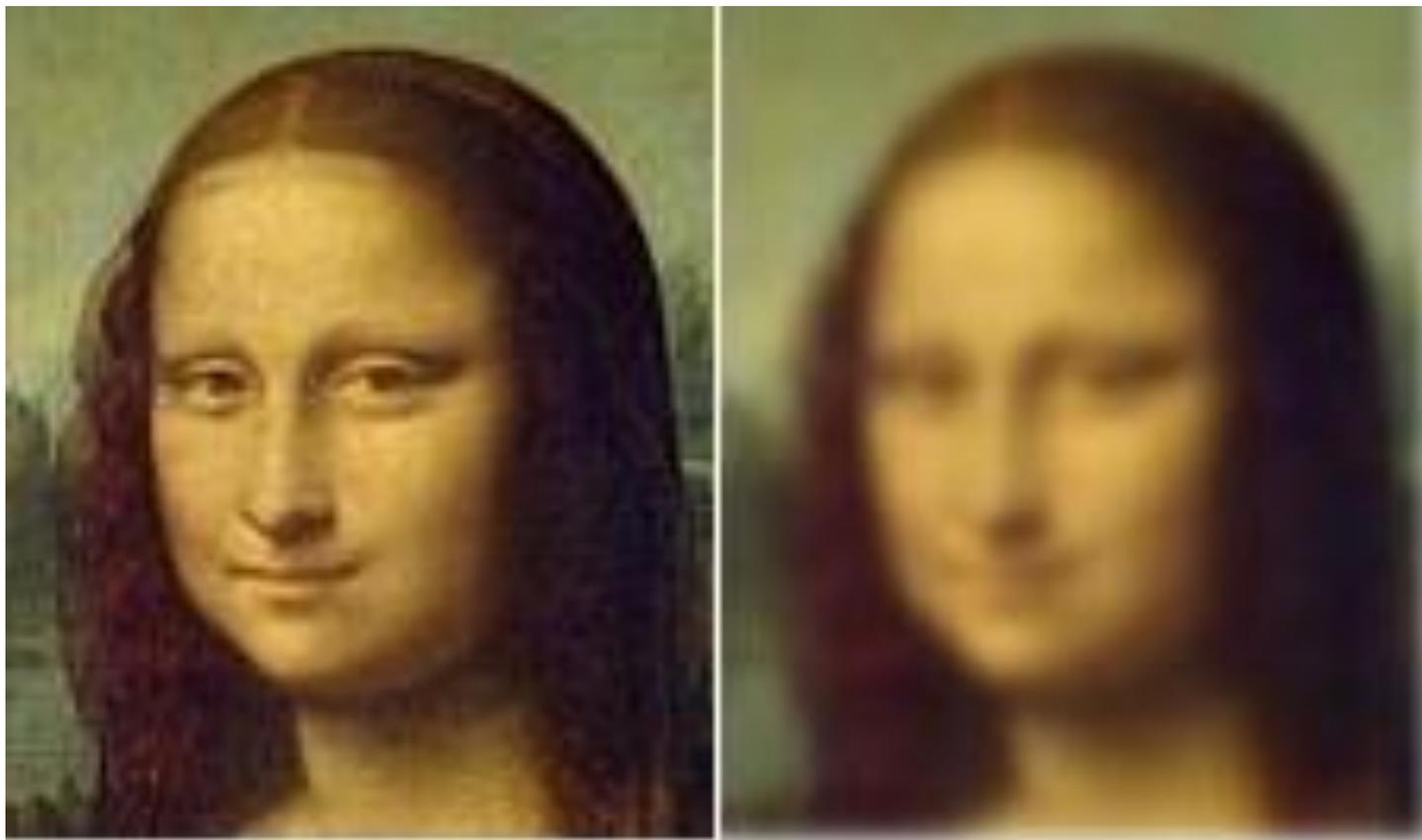
# Проблемы со светом



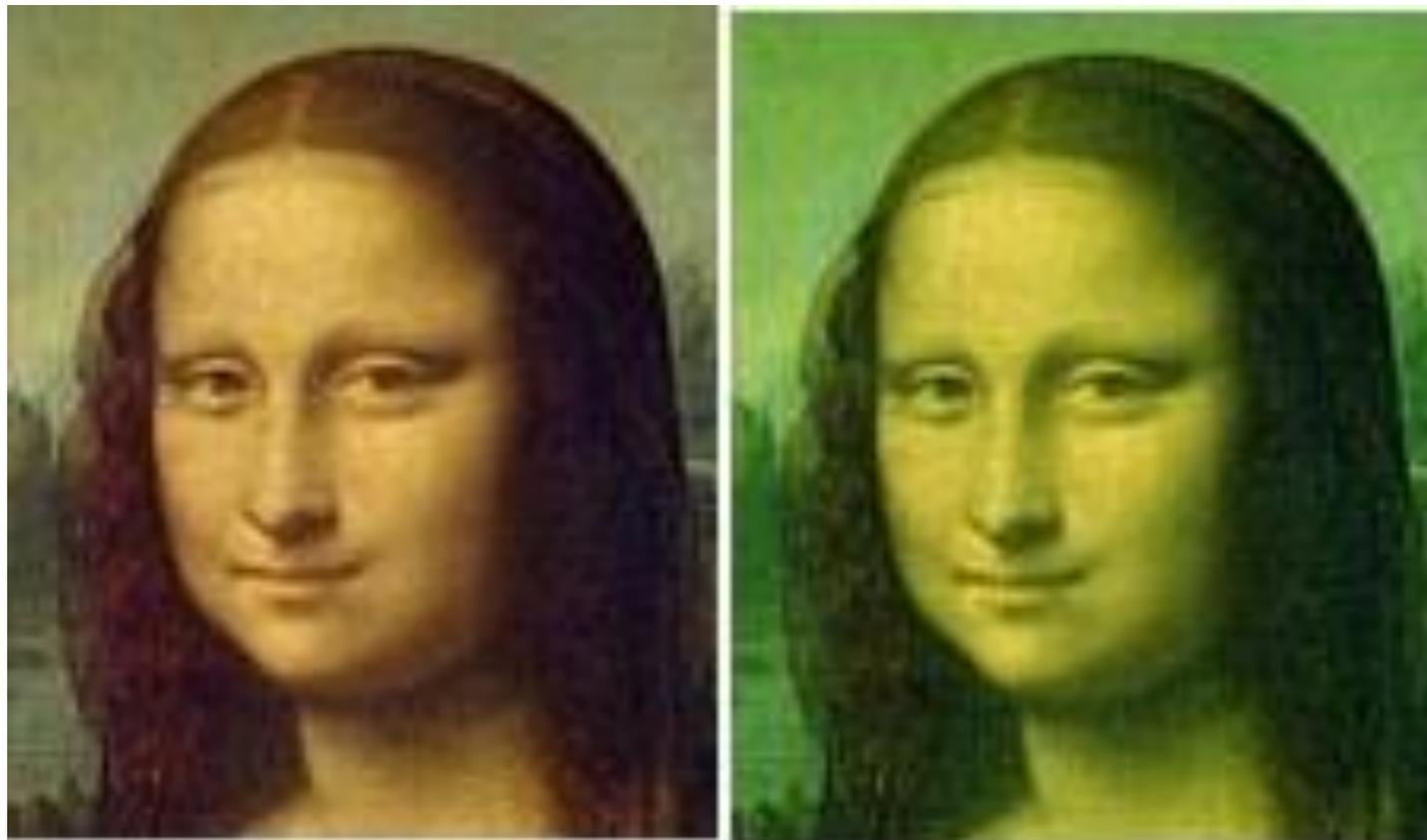
# Зашумленность



# Проблемы с резкостью



## Проблемы с балансом цвета

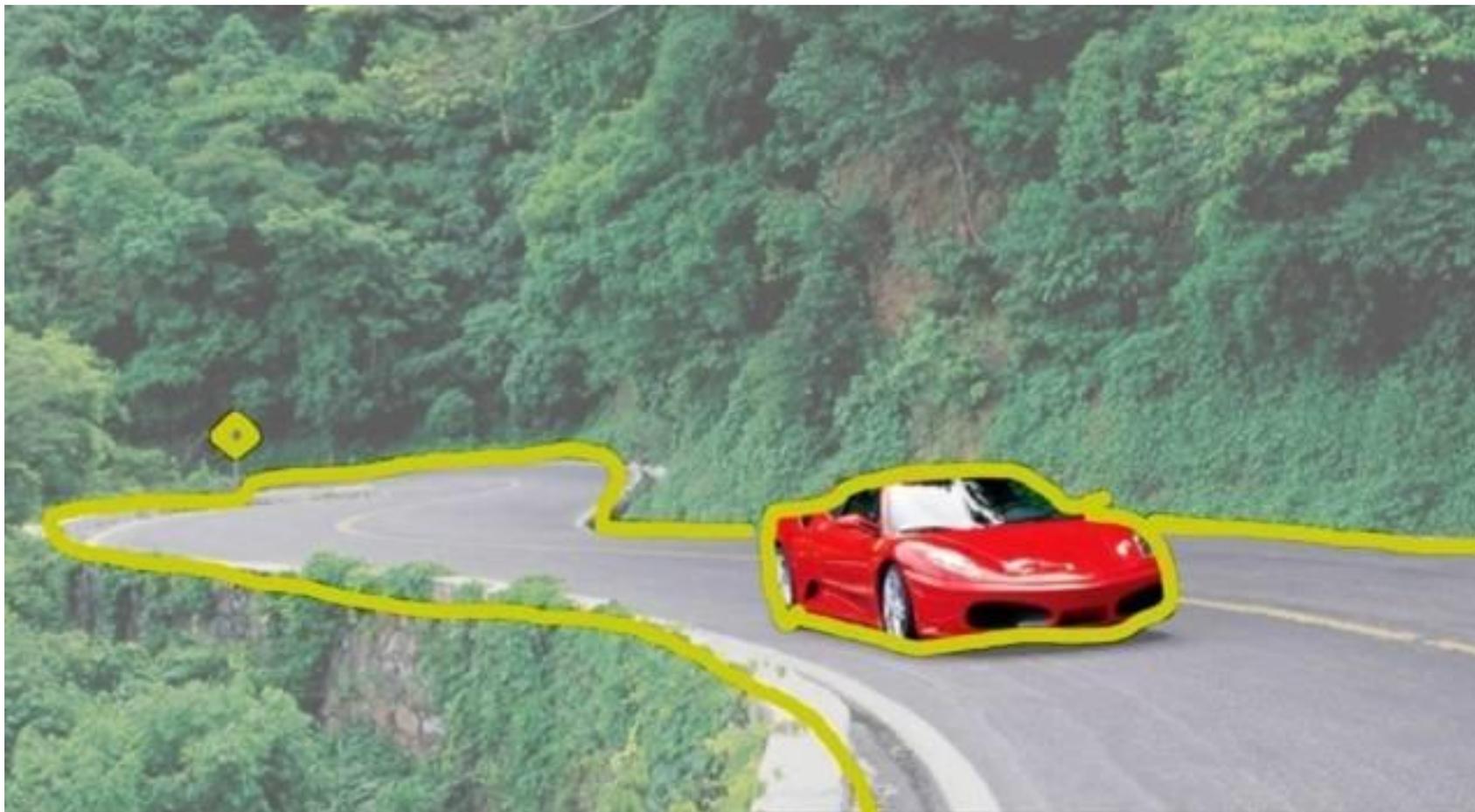


# Средний уровень обработки

- сегментация
- описание объектов
- сжатие объектов в удобную форму для компьютерной обработки

Изображение → Признаки и атрибуты

# Детектирование и сегментация

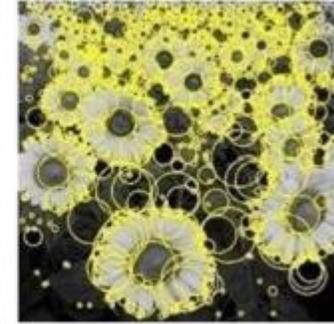
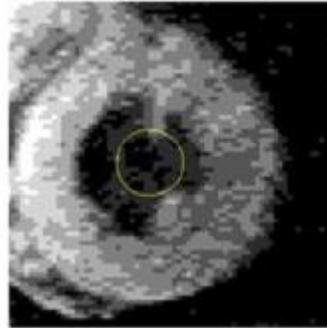
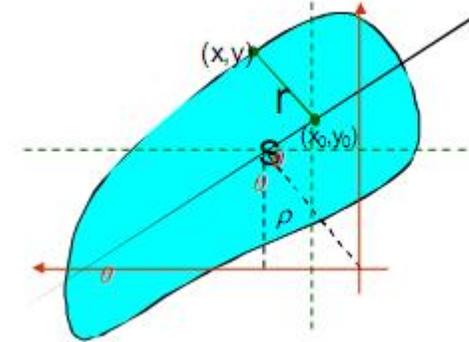
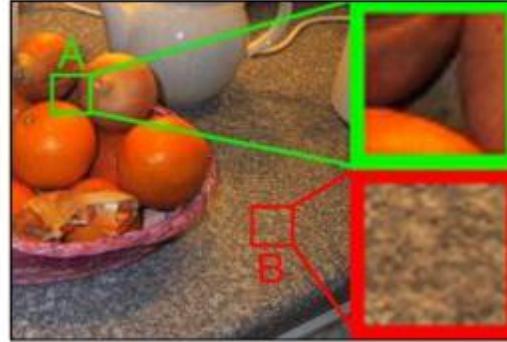
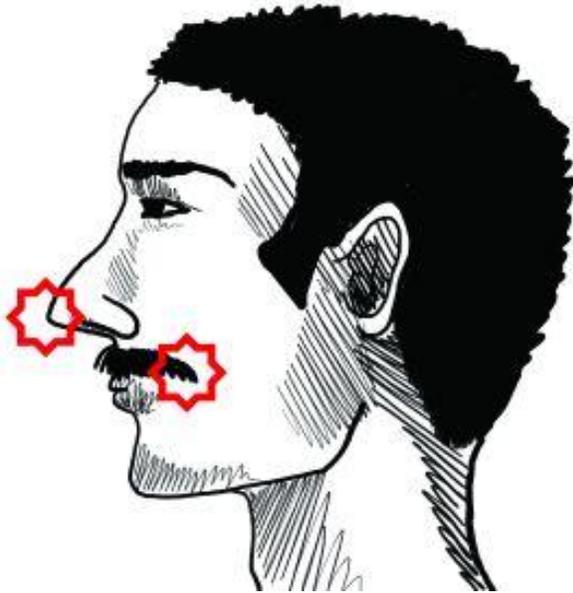


# Какую информацию можно извлечь из изображения?

- гистограмму ориентированных градиентов – histogram of oriented gradients
- оптический поток – optical flow
- disparity – невязка
- границы – линии и углы

Изображение → Признаки и атрибуты

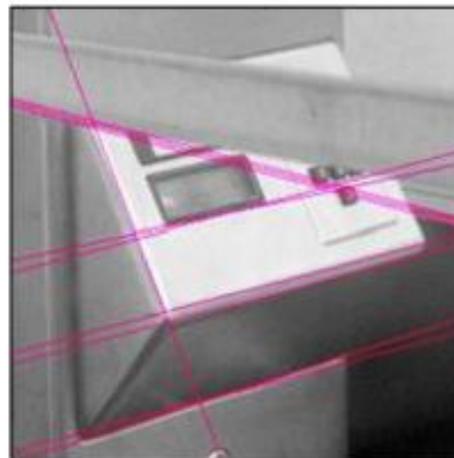
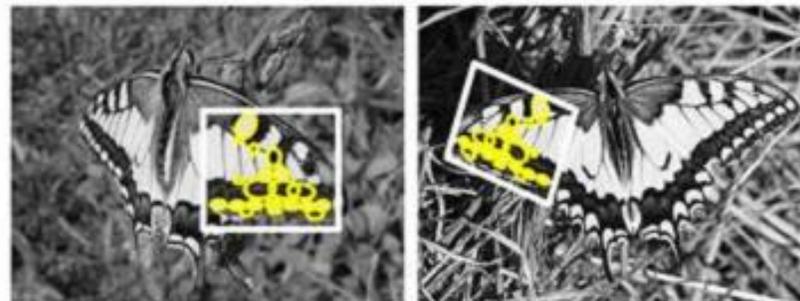
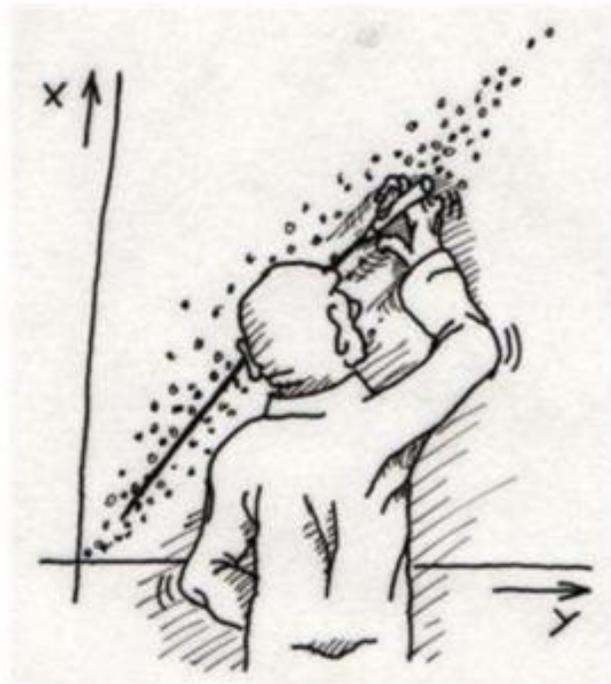
# Локальные и глобальные признаки



# Высокоуровневая обработка

- Распознавание изображений, поиск изображений
- Выделение и отслеживание объектов, распознавание событий

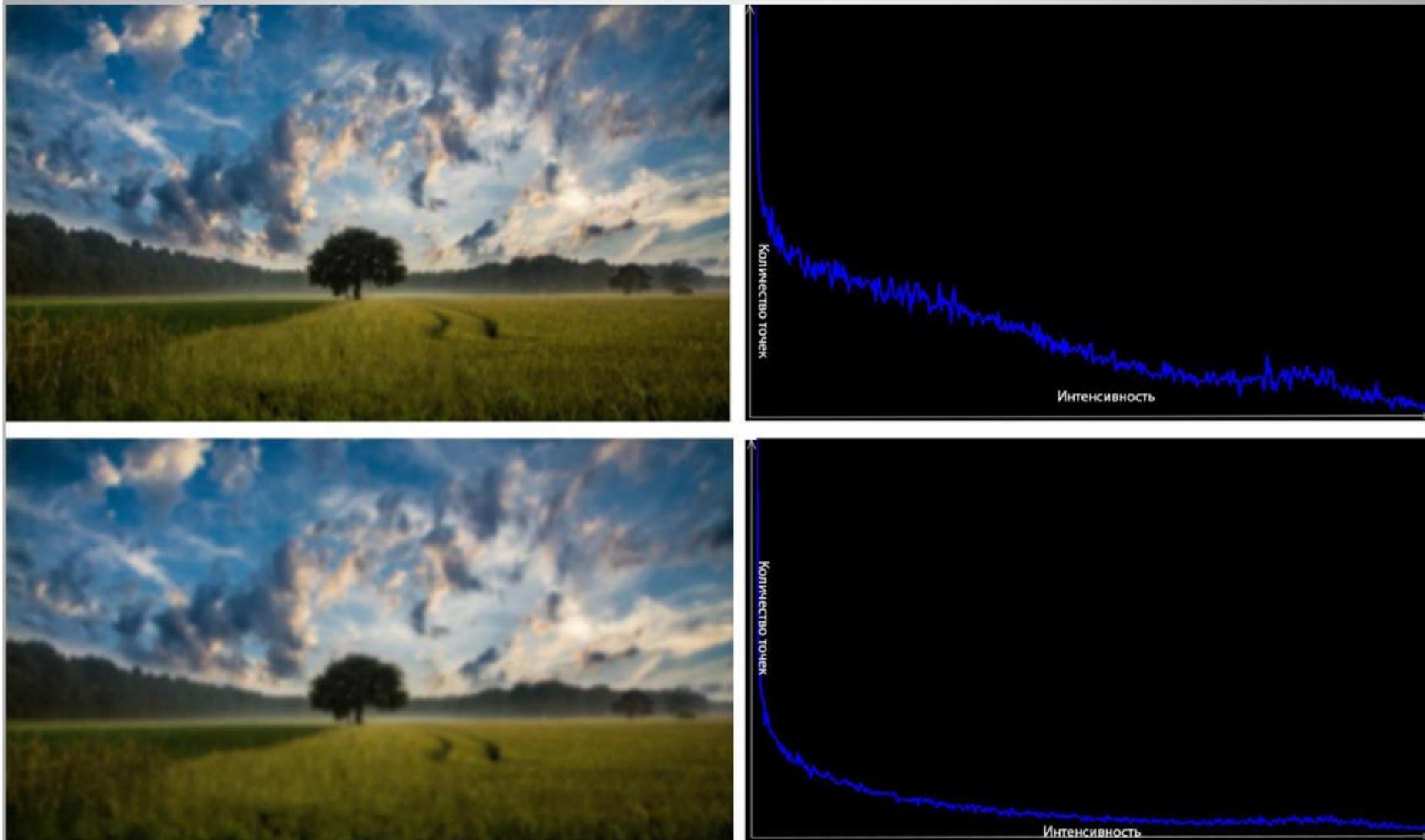
# Сопоставление изображений



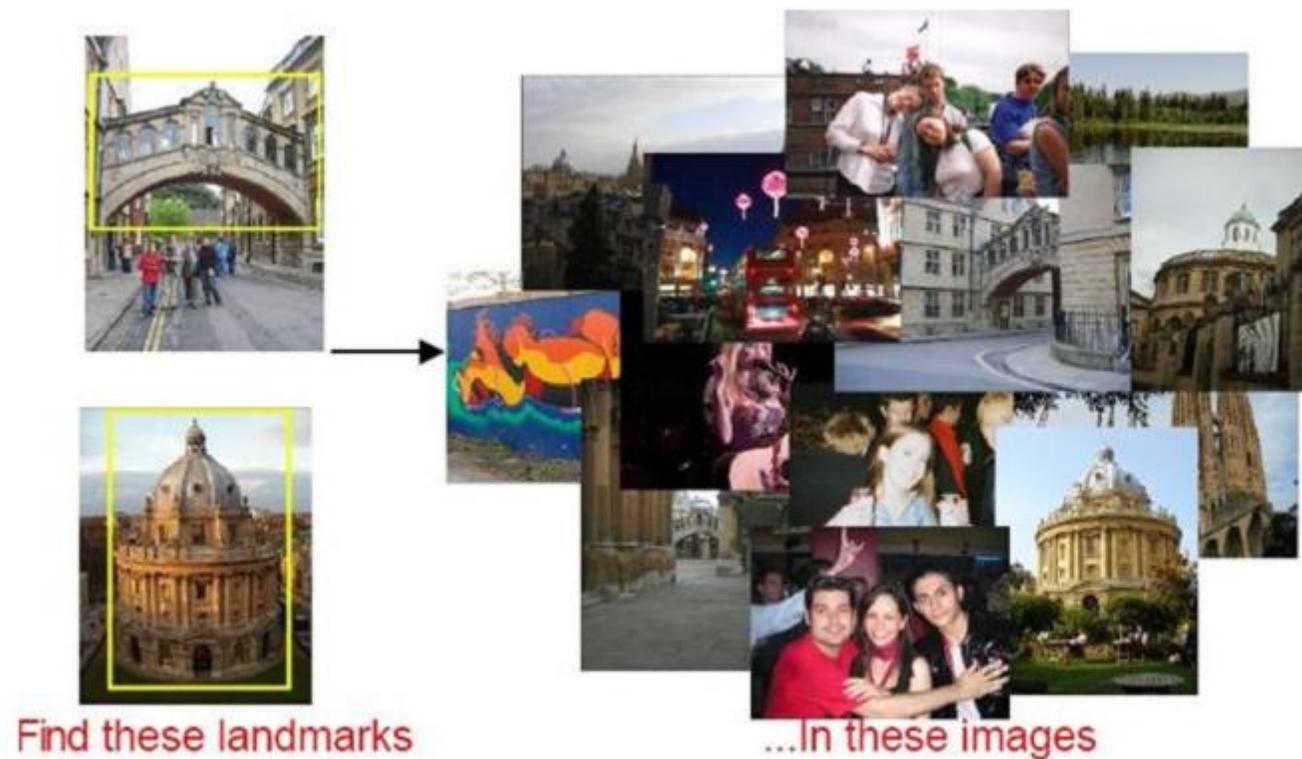
# Выделение объектов



# Автоматический отбор изображений (Раскин Антон)

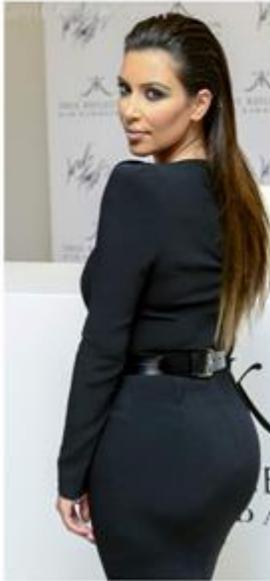


# Поиск изображений по содержанию



# Автоматическое описание фигуры человека

*curvy*  
*feminine*  
*attractive*  
*hourglass*



*big*  
*heavysset*  
*stocky*  
*short torso*



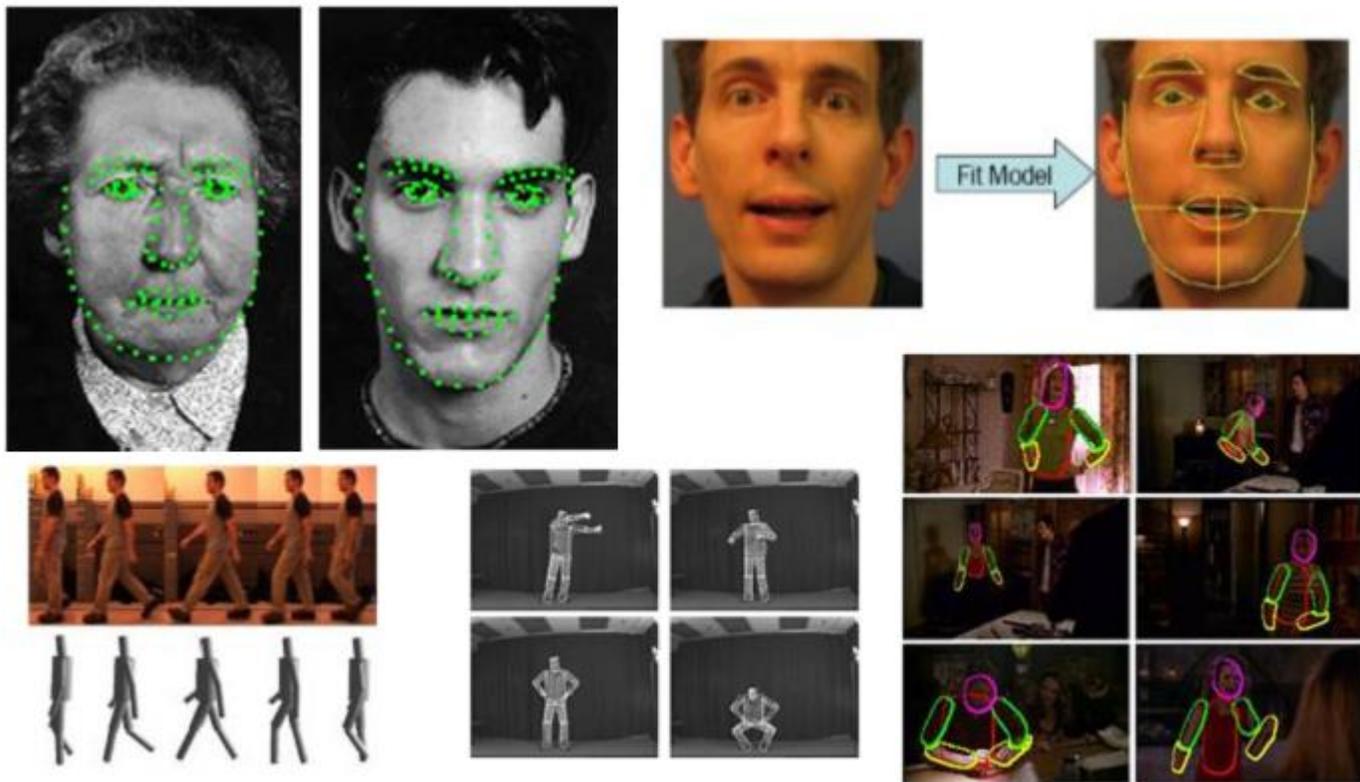
*lean*  
*petite*  
*skinny*  
*small*



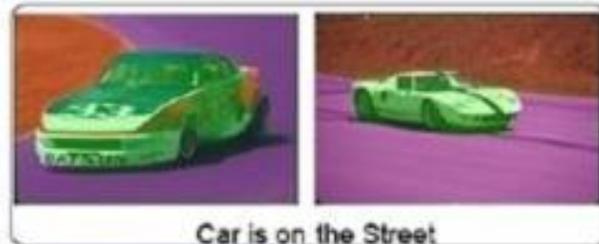
*short legs*  
*short*  
*short torso*  
*small*



# Изображения человека



# Семантическая сегментация



Car is on the Street



Bear In water

Bear is on the field



a



b

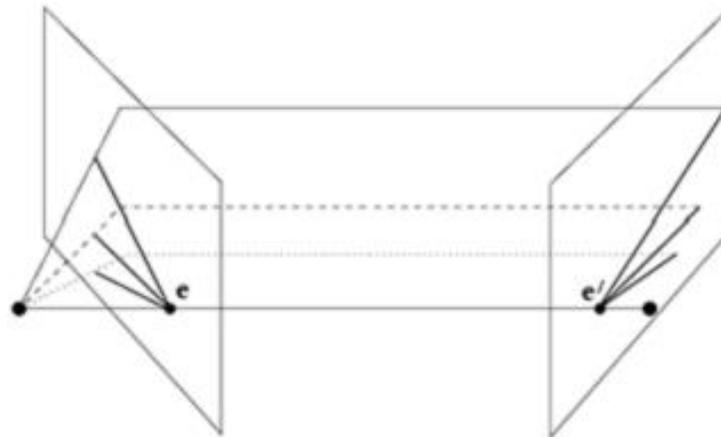
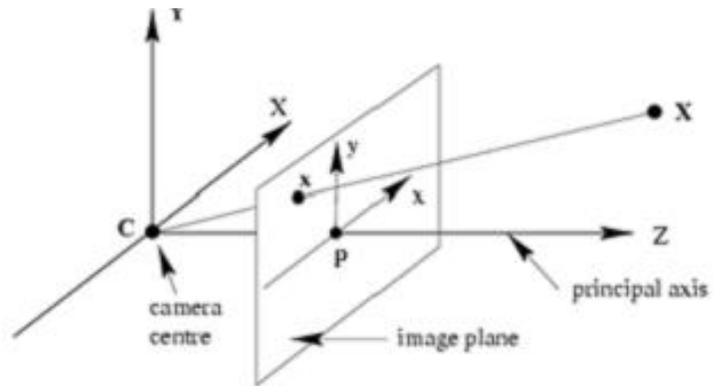


c



d

# Трёхмерная реконструкция



# Литература

1. Р. Гонсалес, Р. Вудс. Цифровая обработка изображений. R.Gonzalez, R. Woods Digital Image Processing
2. Л. Шапиро, Дж. Стокман. Компьютерное зрение Shapiro L., Stockman G. Computer Vision
3. Дэвид Форсайт, Жан Понс. Компьютерное зрение. Современный подход David A. Forsyth, Jean Ponce. Computer Vision: A Modern Approach
4. R.Szeliski «Computer vision: Algorithm and applications»