



Введение в компьютерное зрение

Антон Конушин

2 марта 2015 года

ВМК МГУ имени М.В. Ломоносова



Лектор



Антон
Конушин

- Доцент, к.ф.-м.н., зав. лаборатории компьютерной графики и мультимедиа, каф. АСВК, ВМК МГУ имени М.В. Ломоносова
- Доцент, академический руководитель программы бакалавриата «Прикладная математика и информатика» ФКН НИУ ВШЭ
- Лектор Школы Анализа Данных Яндекс
- Научный консультант стартапа «Технологии видеоанализа»
- Руководитель семинара «Компьютерное зрение» на АСВК

E-mail: ktosh@graphics.cs.msu.ru



Ассистент



Влад
Шахуро

- Студент лаборатории компьютерной графики и мультимедиа ВМК МГУ имени М.В. Ломоносова
- Научные интересы: выделение и распознавание объектов на изображениях

E-mail: shahurik@ya.ru



План лекции

- Введение
 - Что такое «компьютерное зрение» и почему это сложно
 - История и достижения компьютерного зрения
 - Информация о курсе
- Компьютерное зрение и зрение человека
 - Цифровое изображение
 - Глаз человека и цифровая камера
 - Цвет и модели цвета

1. Что такое «компьютерное зрение» 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



Задача компьютерного зрения

- «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
- Что это в действительности обозначает?
 - Зрение - источник семантической информации о мире
 - Зрение - источник информации о расстояниях и размерах объектов



Выделение объектов



Необходимо определить, есть ли на изображении объекты заданного типа и если да, то определить их положение



Классификация изображений

- вне помещения
- город
- Пекин, Китай
- Пл. Тяньаньмэнь



Общая характеристика изображения



Характеристики объектов



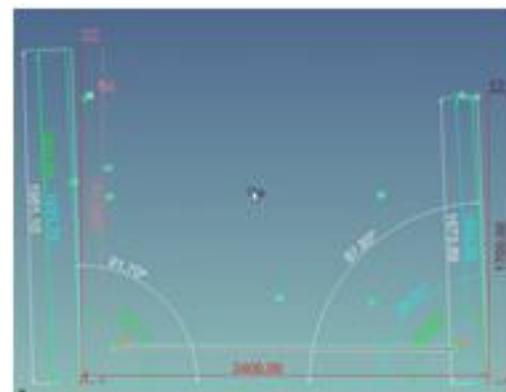


Измерения по изображению

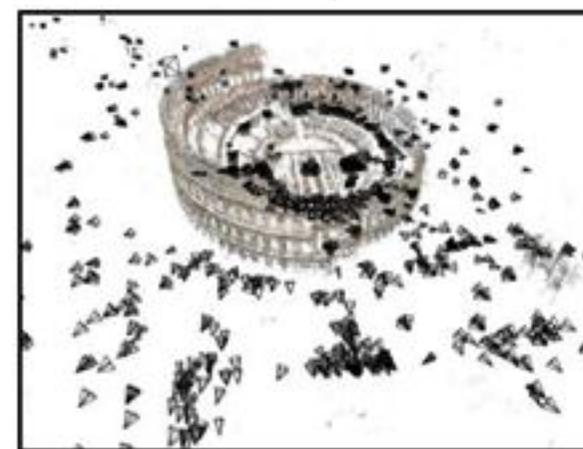
Сtereo-зрение



Измерения
расстояний по
снимкам



3D моделирование по
снимкам



Почему зрение – это сложно?



Точка наблюдения (ракурс)



Michelangelo 1475-1564



Освещение

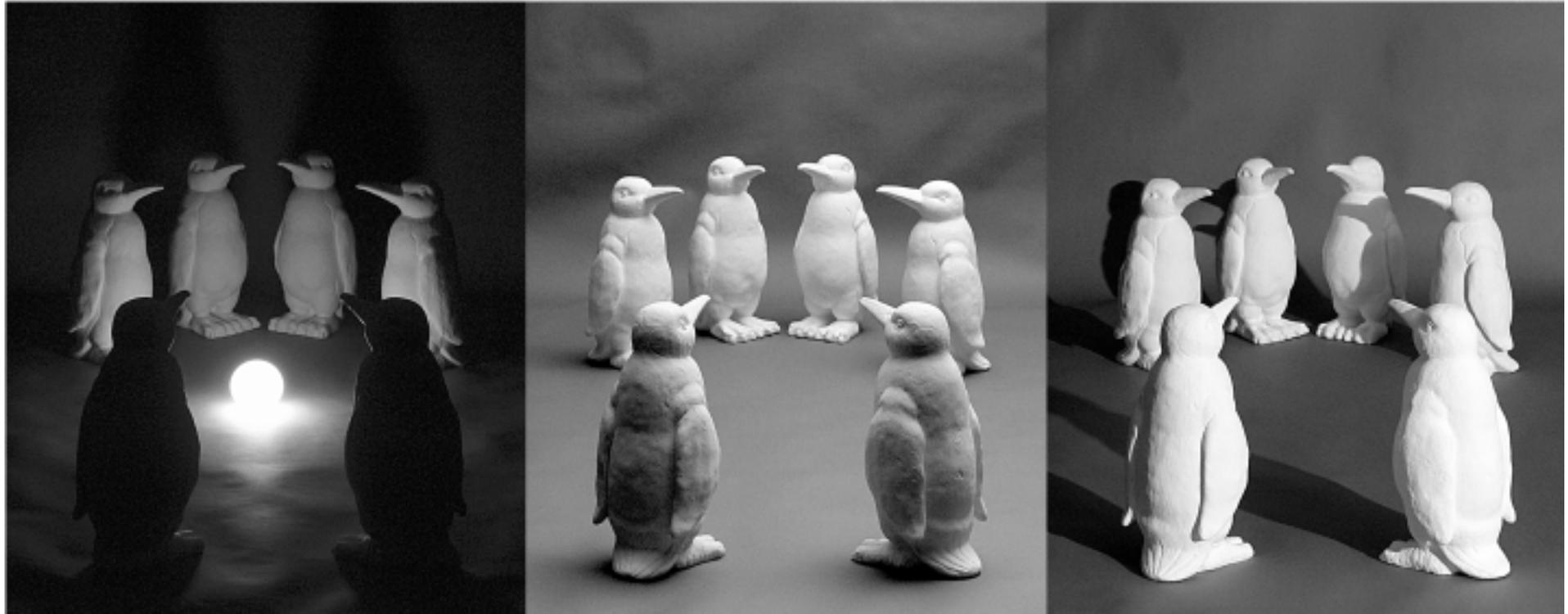


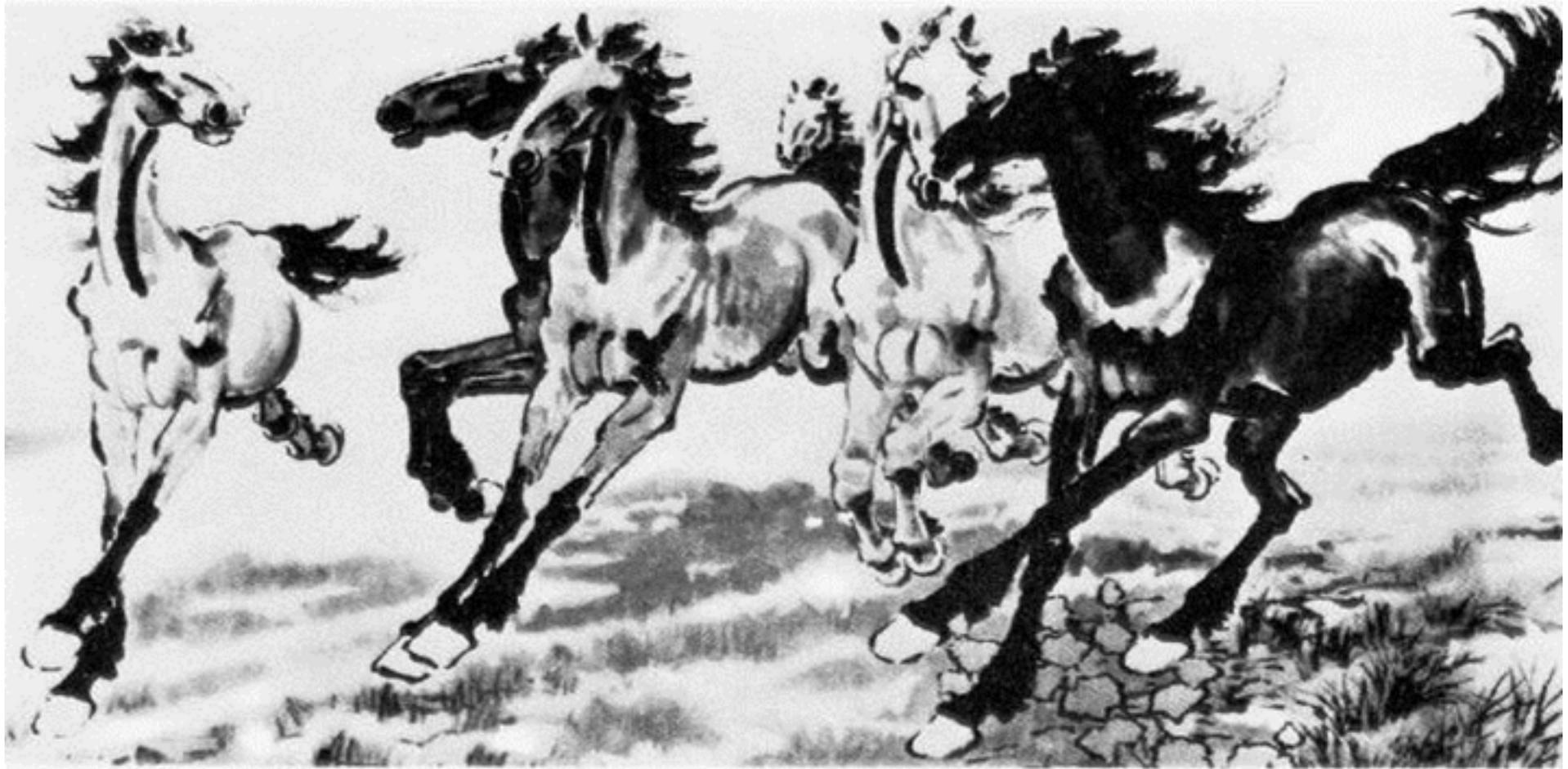
image credit: J. Koenderink

Размер



Slide credit: Fei-Fei, Fergus & Torralba

Деформация

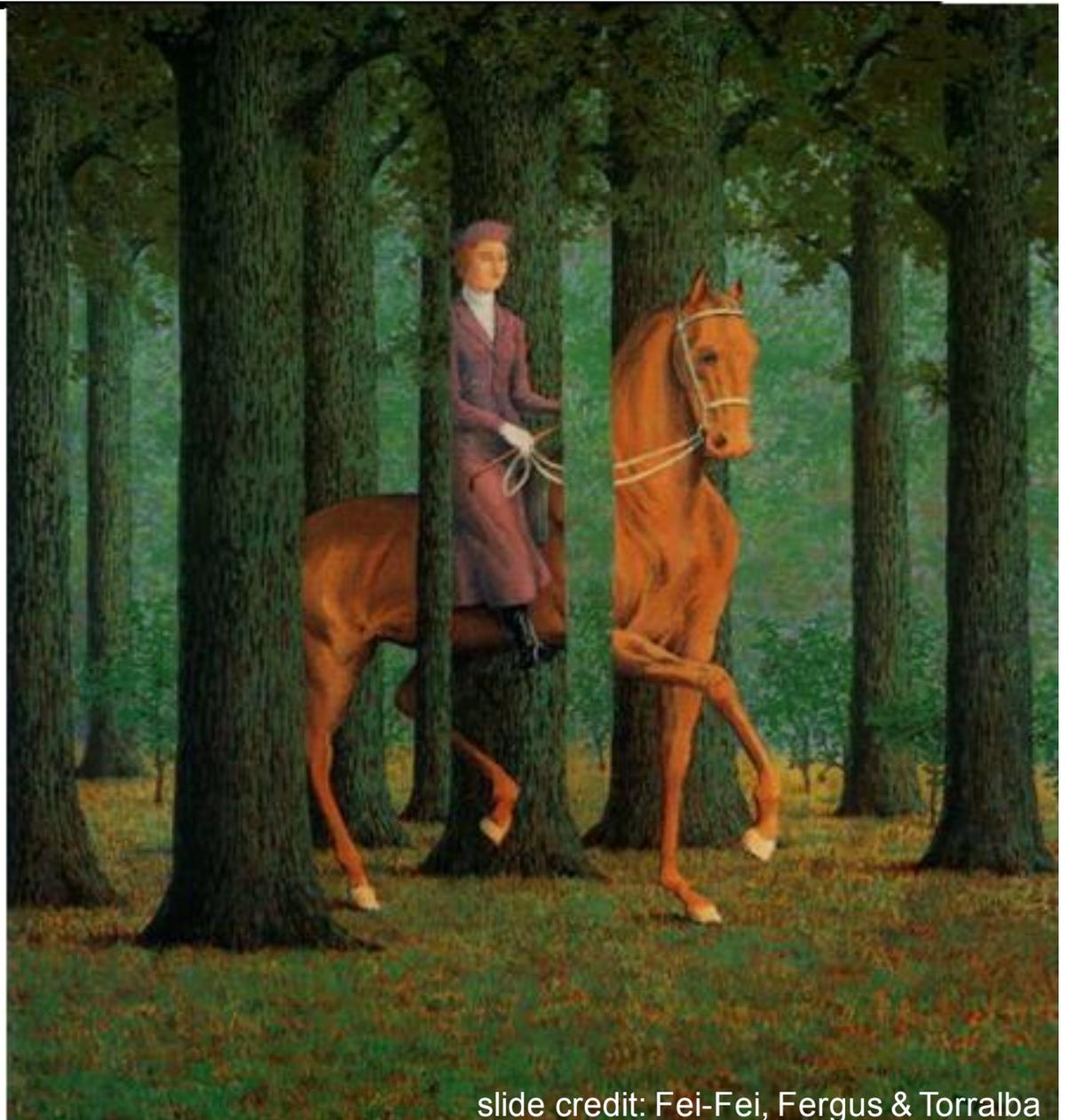


Xu, Beihong 1943

Перекрытие



Magritte, 1957



slide credit: Fei-Fei, Fergus & Torralba

Маскировка



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

Движение





Внутриклассовая изменчивость



Slide credit: Fei-Fei, Fergus & Torralba



Сложности или возможности?

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

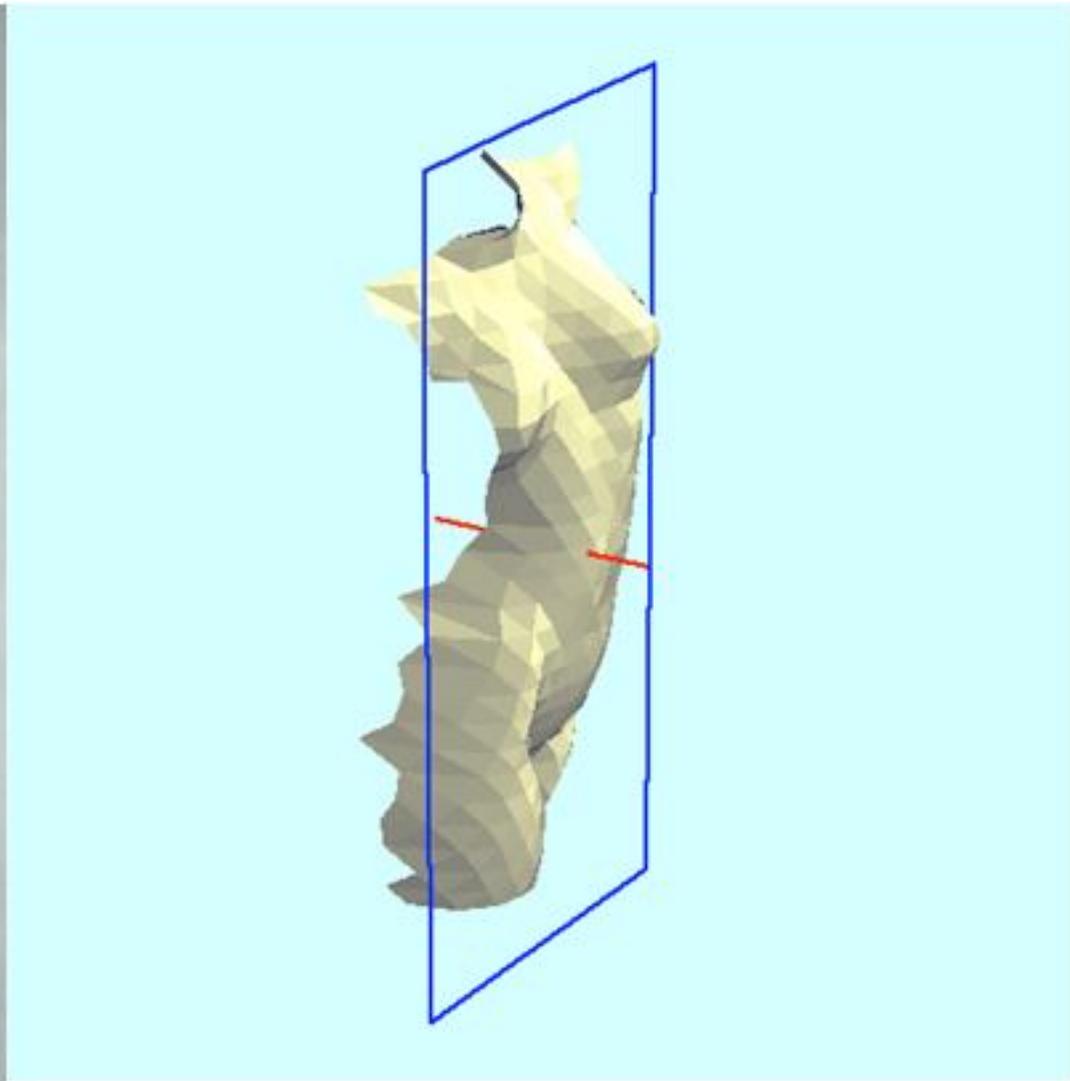


Цвет





Освещение



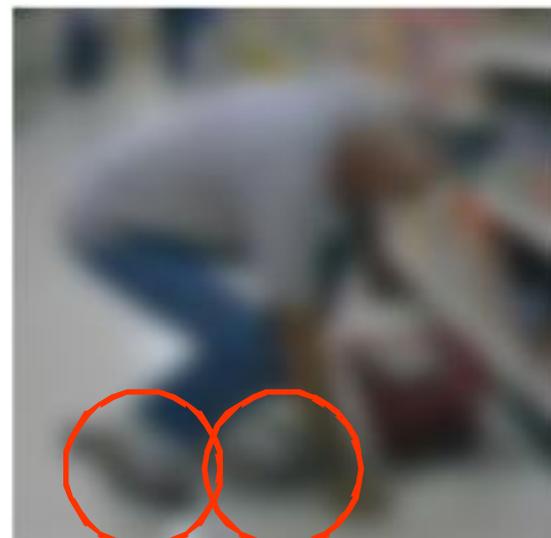
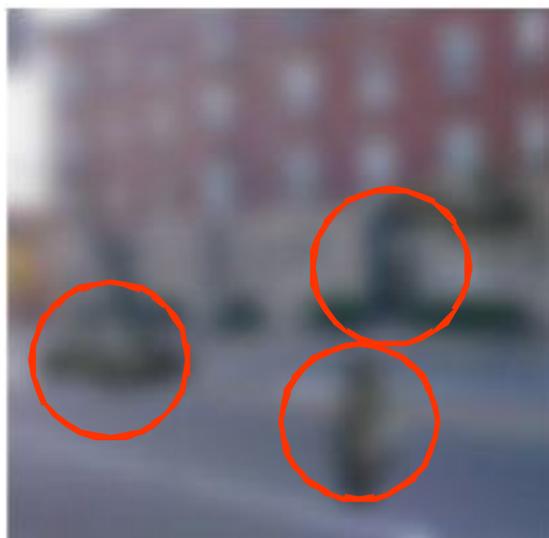
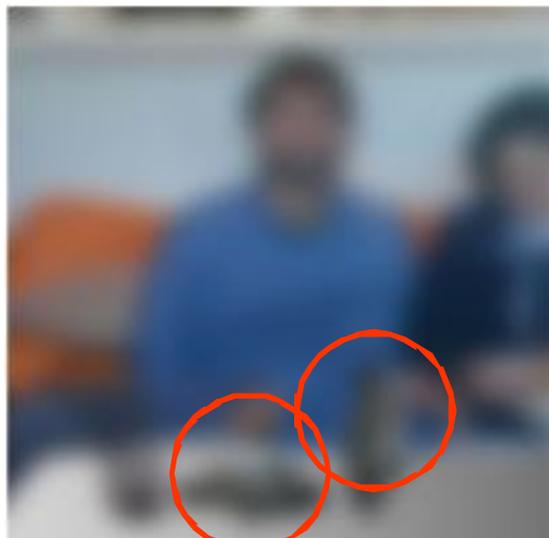
Текстура



Контекст



Что это?



Slide credit: Fei-Fei, Fergus & Torralba



Интерпретация изображений



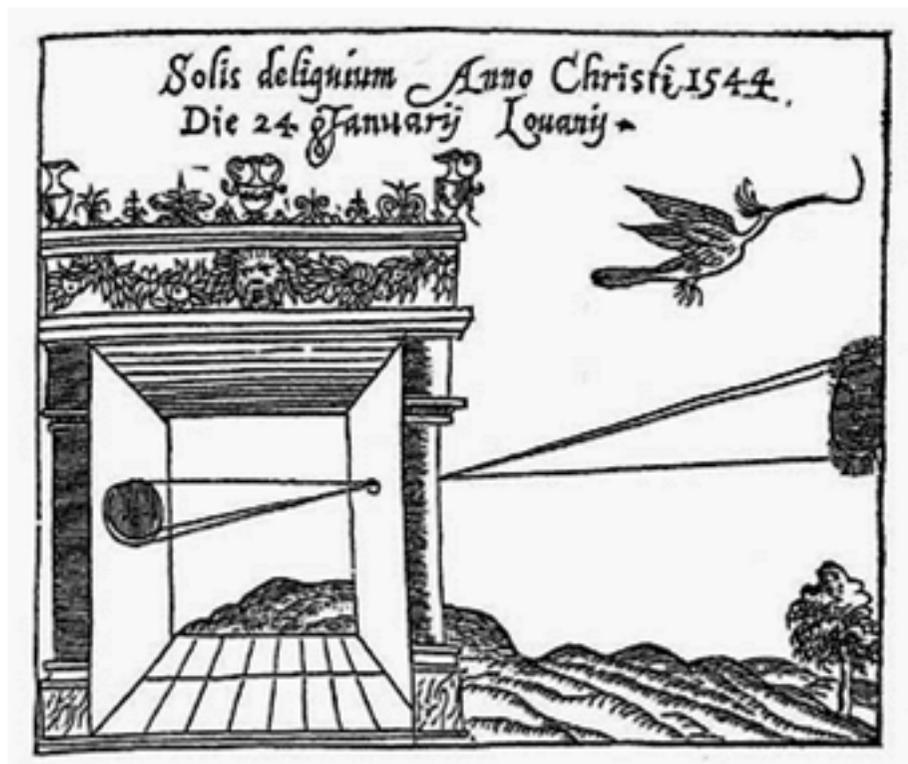
- Одно 2D изображение допускает разные 3D интерпретации
- Нужны априорные знания о структуре и свойствах мира

- Мы сопоставляем наблюдения (подсказки) и априорные знания для интерпретации изображения
- По оценкам, 25% мозга занято решением задачи зрения
- Это типичная задача искусственного интеллекта
- Это можно делать с помощью машинного обучения!
- Поэтому машинное обучение играет такую большую роль

2. История и достижения компьютерного зрения



Камера-обскура



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

“Магический фонарь” (1492)



Figure 1. Leonardo da Vinci.



Живопись Ренессанса



Яна Ван Эйк «Портрет Четы Арнольфини» 1434г



Секреты мастеров



BBC David Hockney's «Secret Knowledge»

Камера-обскура с линзой (1500-1600е)



Возможность проецировать изображение на большой холст. Побочный результат – много левшей на картинах



<http://www.adme.ru/hudozhniki-i-art-proekty/sekretnoe-znanie-543505/>



Камера-люцида (1807)



Camera Lucida — устройство, которое представляет собой конструкцию с призмой, которая крепится, например, на стойке к планшету.

Художник, глядя на свой рисунок одним глазом, видит реальное изображение, а другим — собственно рисунок и свою руку. Получается оптическая иллюзия, позволяющая точно переносить пропорции реальные на бумагу.



Первая фотография



Самая первая фотография
1825 год



Figure 5. J. N. Niepce.

Требовала 8 часов проявки



Фотограмметрия



Figure 6. Jacques Daguerre.

1837 – первые практически применимые фотографии

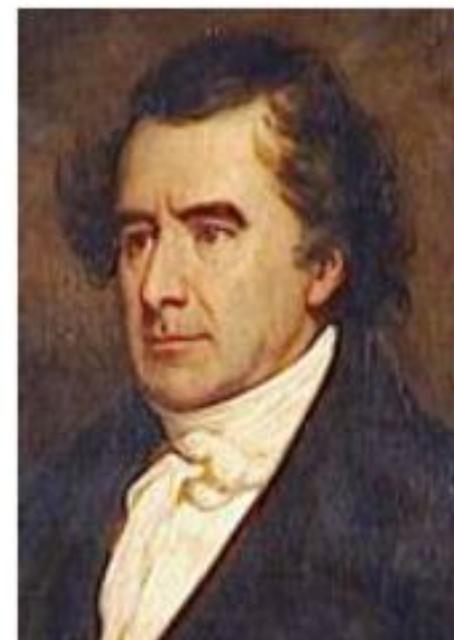


Figure 8. Dominique François Jean Arago.

1840 – «Фотограмметрия – будущее геодезии»

Фотограмметрия (Photogrammetry) - измерение расстояний между объектами по 2D изображениям

Видео



1878 – первая скоростная съемка, Eadweard Muybridge



1888 – первое кино на плёнке, Louis Le Prince



Стереофотограмметрия (1896)



Figure 17. Edouard Deville.

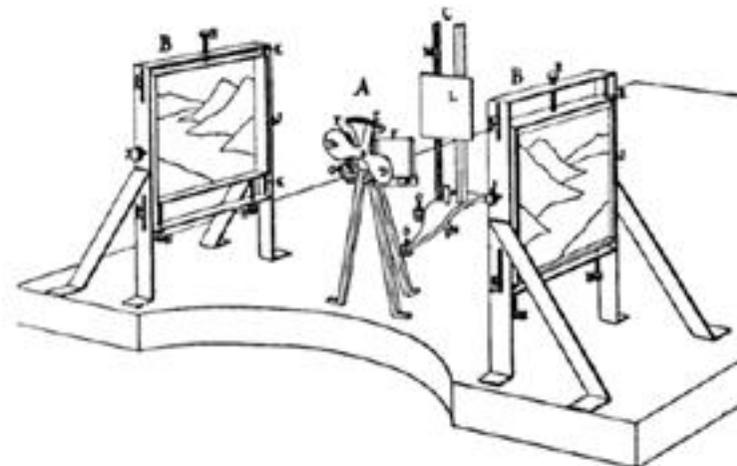


Figure 18. Deville's Stereo Planigraph.

Стереокамера и
теодолит

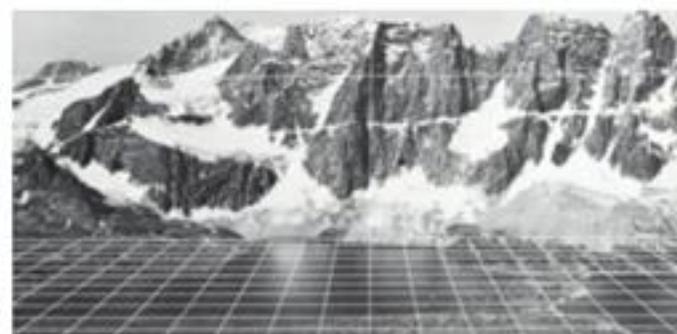
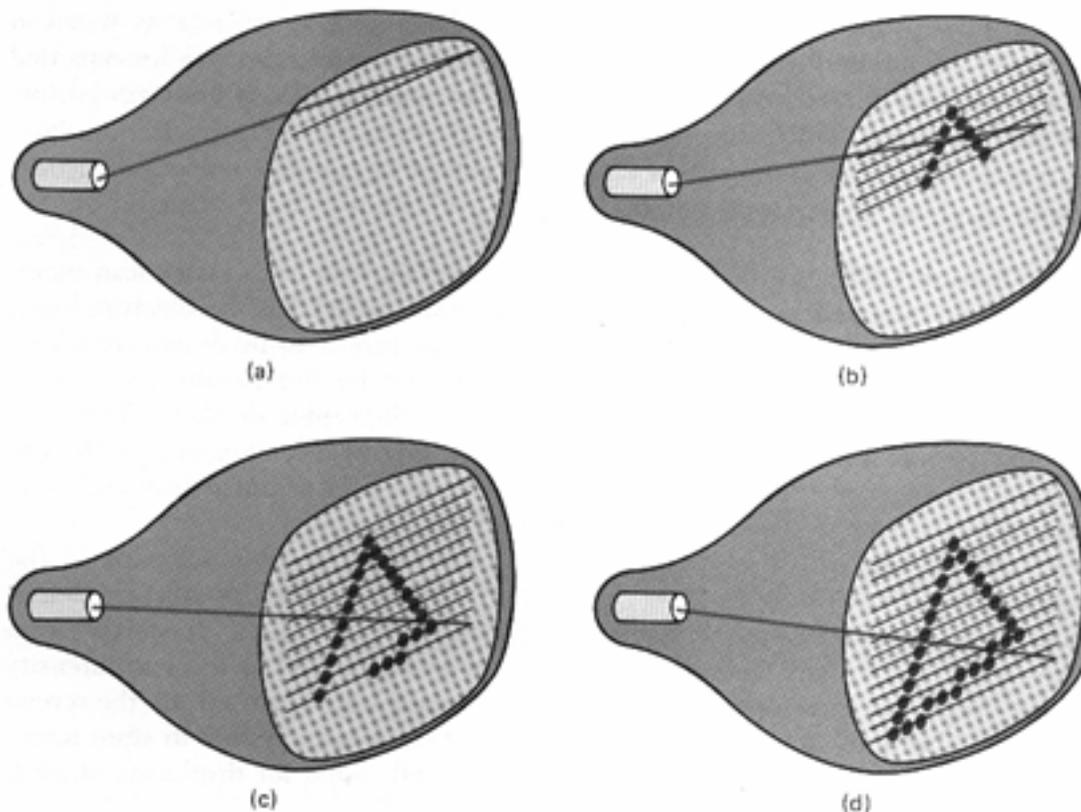


Figure 19. Illustration of the Canadian Grid Method used by Deville.



Растровый дисплей (1927)



Philo Farnsworth – 60-строчный растровый дисплей

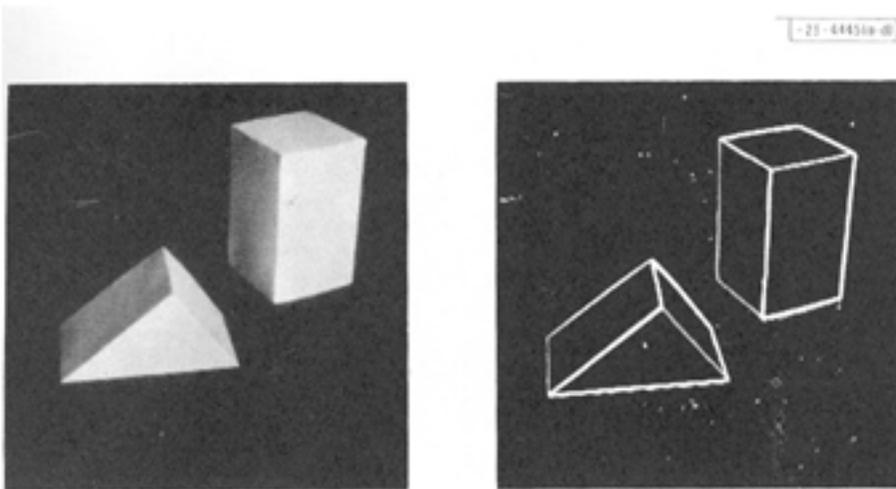


Whirlwind, MIT (1951)



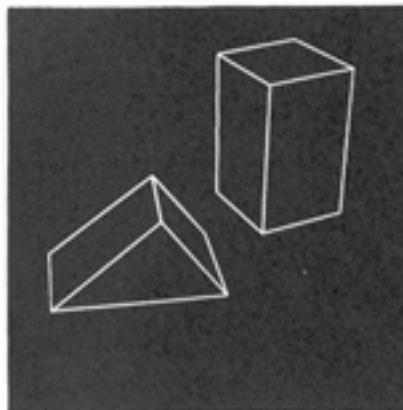
- Первый компьютер, отображающий текст и графику в реальном времени на мониторе
- Точками карту, значком самолёт.
- «Световое перо» для взаимодействия с экраном (запрос информации об объекте)

Зарождение компьютерного зрения (1960)

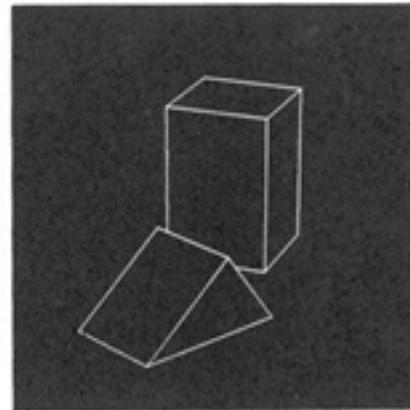


(a) Original picture.

(b) Differentiated picture.



(c) Line drawing.



(d) Rotated view.

L. G. Roberts, *Machine Perception of Three Dimensional Solids*,
Ph.D. thesis, MIT Department of
Electrical Engineering, 1960



SketchPad, MIT (1963)

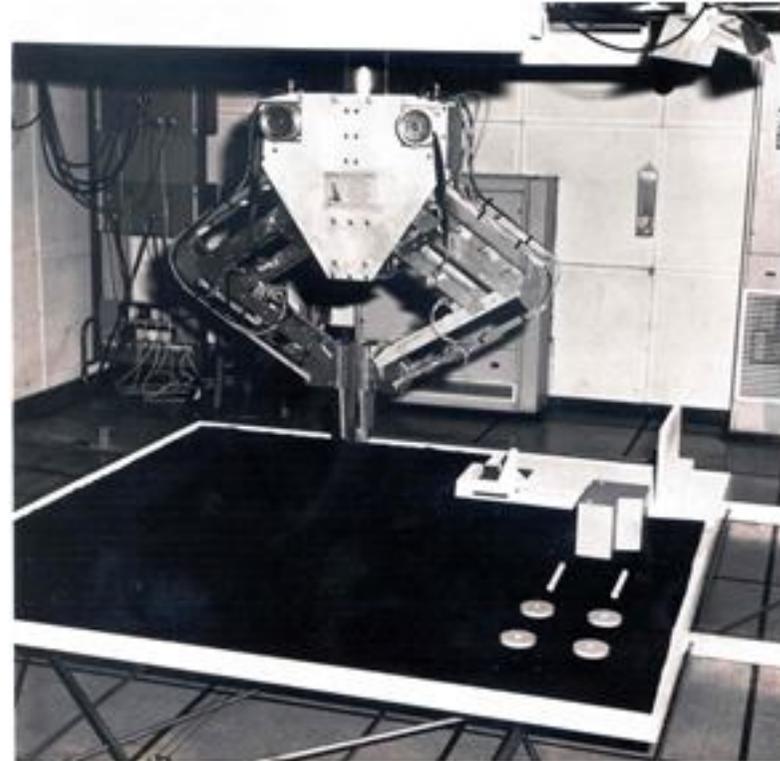


Ivan Sutherland демонстрирует интерактивный графический редактор SketchPad



Freddy II (1973)

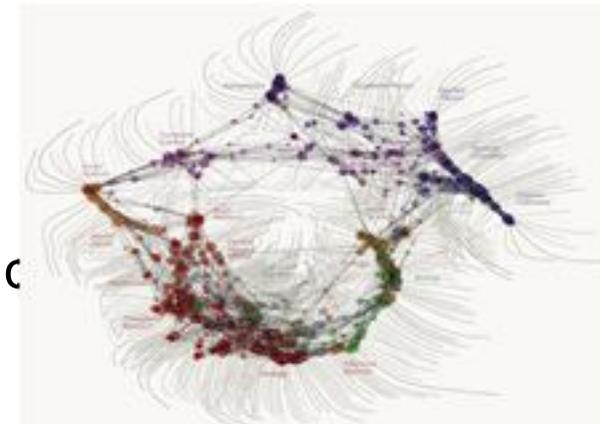
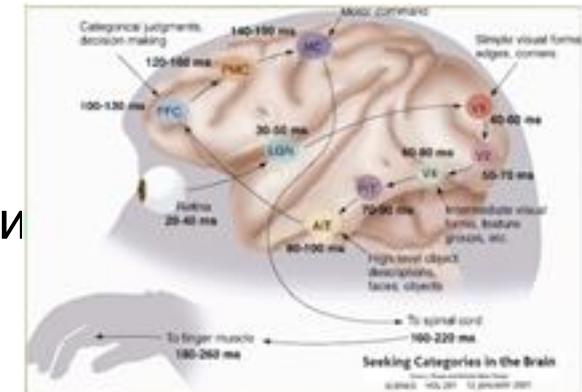
- Университет Эдинбурга
- Один из первых роботов с системой машинного зрения
- 5 степеней свободы
- Умеет собирать машинки из кубиков, разбросанных по столу
- 384Кб RAM в управляющем компьютере





Давид Марр (1970e)

- «Primal sketch»
 - Низкоуровневые («low-level») свойства изображения: направленные края, отрезки и т.д.
- «2.5D sketch»
 - Упорядочивание по глубине (бинокулярное стерео), учёт текстуры и т.д.
- «3D model»
 - Распознавание объектов и представление с 3х мерном мире



“The Marr Prize” – главная премия в области компьютерного зрения



Детектор лиц Viola-Jones (2001)



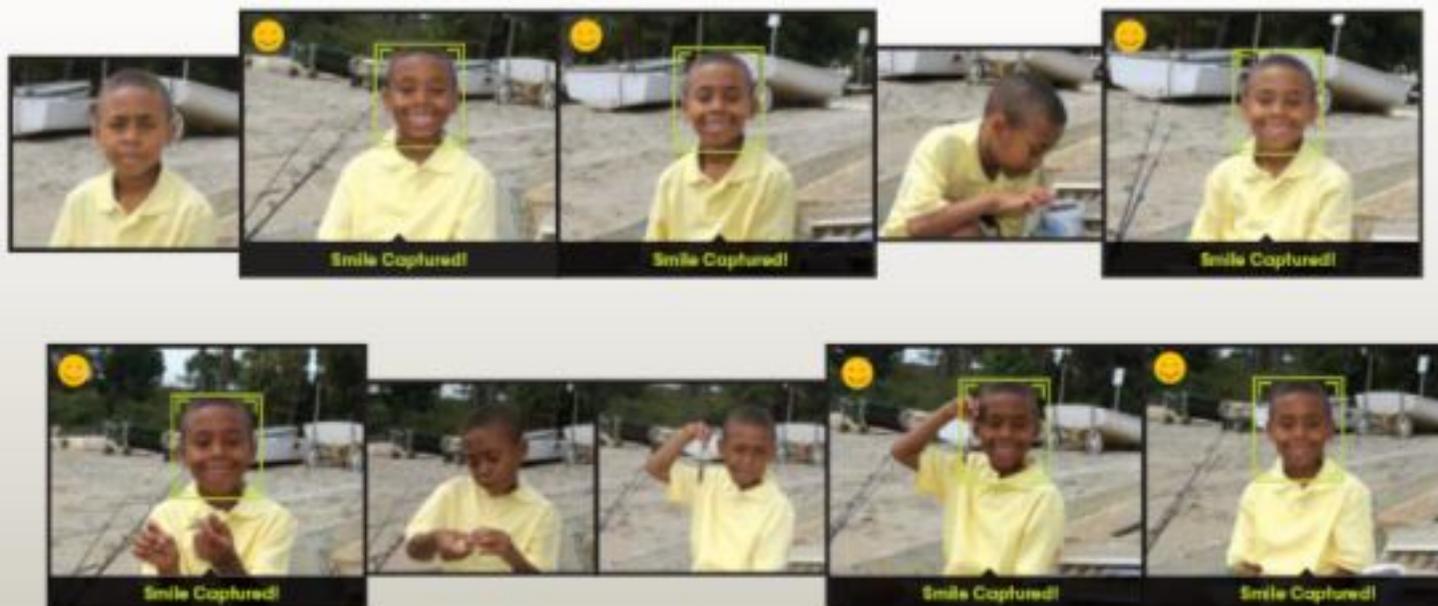
Алгоритм Viola-Jones – первый быстрый и надежный алгоритм поиска лиц. Демонстрация силы машинного обучения.



Обнаружение лица и улыбки

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



[Sony Cyber-shot® T70 Digital Still Camera](#)

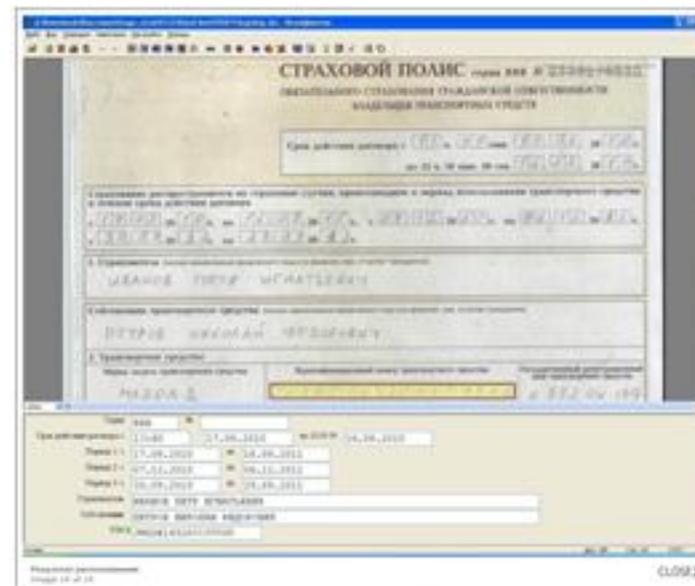
Source: S. Seitz



Распознавание текста



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



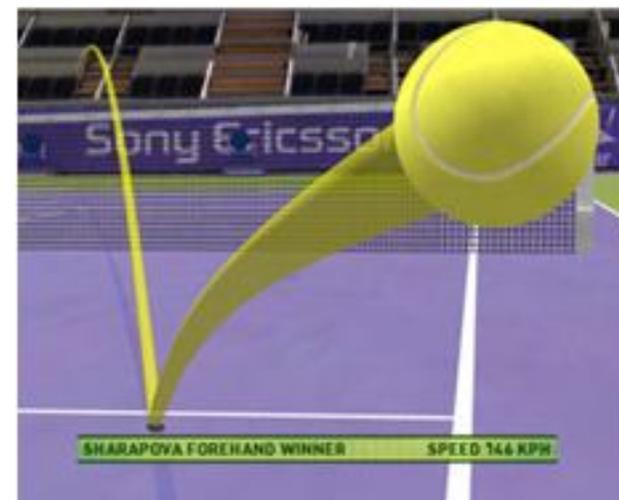
Cognitive Technologies, Россия
<http://cognitiveforms.ru/products/cognitive-forms/>



Распознавание текста в произвольных изображениях пока работает плохо!



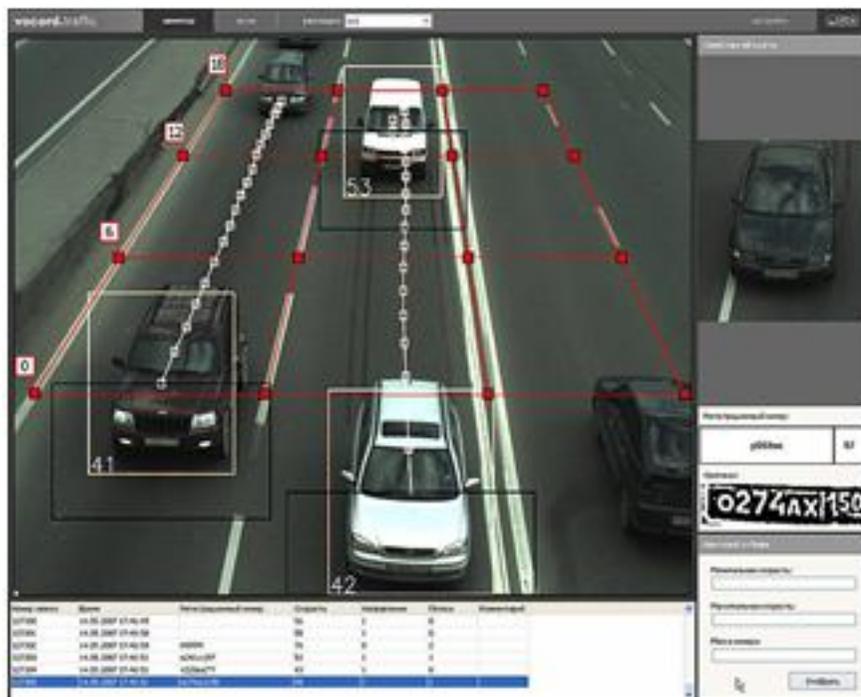
Спортивные соревнования



<http://www.hawkeyeinnovations.co.uk>



Видеонаблюдение

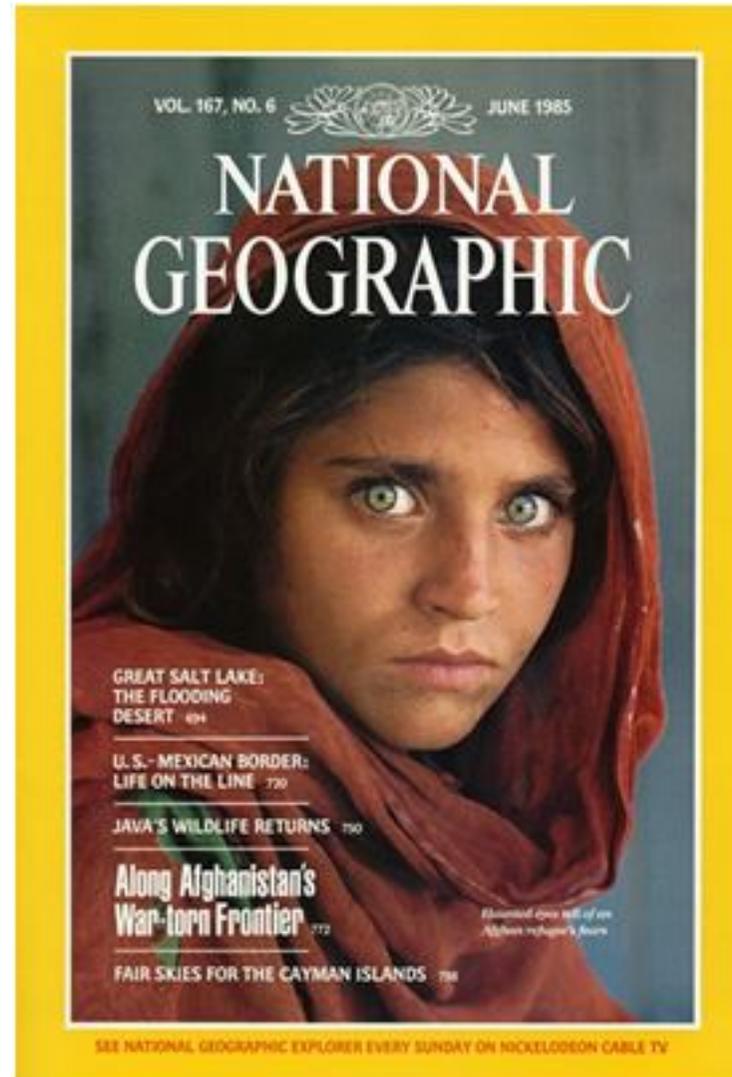


VOCORD Traffic, Vocord, Россия
<http://www.vocord.ru/218/>



Автоматизация видеонаблюдения
за счёт видеоаналитики
(в т.ч. в нашей лаборатории)

Биометрия



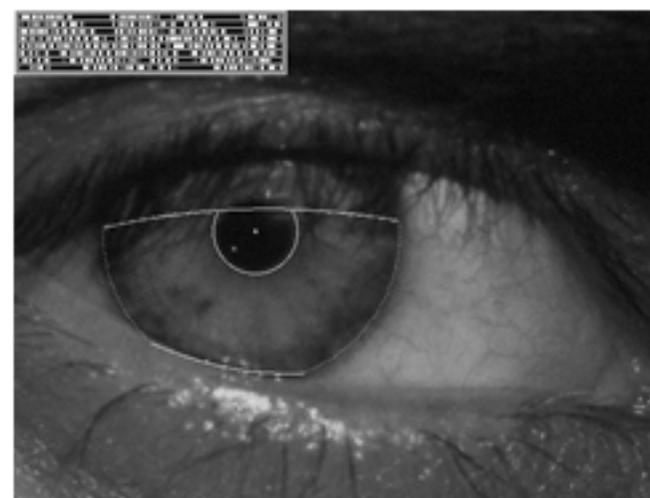
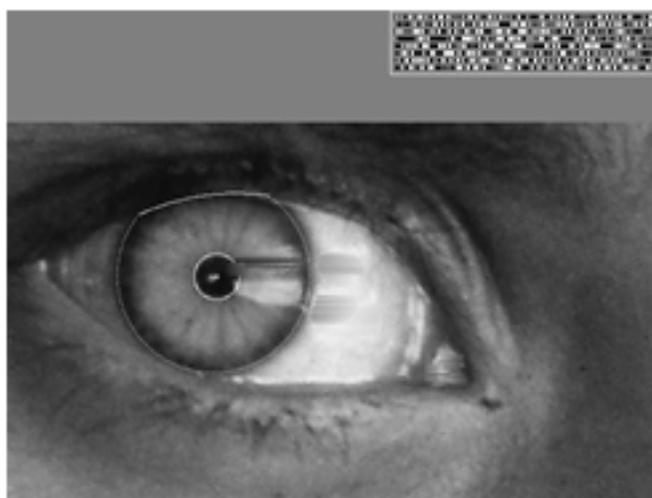
Кто она?



Биометрия



“Как девушку из Афганистана идентифицировали по радужке глаза” Читайте [ИСТОРИЮ](#)



Source: S. Seitz



3D модели и захват движения



L.A. Noire, Team Bondi/Rockstar, 2011





Microsoft Kinect (2010)



Первая потребительская система взаимодействия с компьютером с помощью жестов. При создании использовались суперкомпьютеры



Поиск изображений

Яндекс **картинки**

Поиск Почта Карты Маркет Новости Словари Блоги Видео Картинки ещё a.konushin Помощь Выход

Загруженная картинка Умеренный поиск

Размер
Большой
Средний
Маленький
Свой размер
Обои

Файл
JPEG
PNG
GIF

 Нашлось 4 таких же картинки

 Гамбург 2008 * Форум Винского
Гамбург 2008... 800×488
http://forum.awd.ru/gallery/image_page.php?album_id=17164&im...
JPG, 278 КБ

 Гамбург 2008 * Форум Винского
Гамбург 2008... 800×488
http://forum.awd.ru/gallery/image_page.php?album_id=17164&im...
JPG, 135 КБ

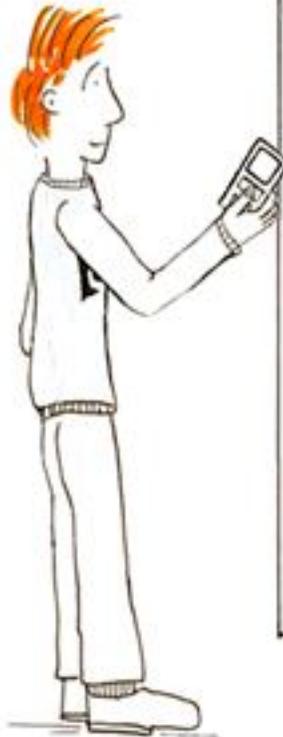
 Гамбург / Путешествия / Германия / Pinme.ru
Войдите, для того, чтобы комментировать Пин... 600×400
<http://pinme.ru/pin/50c31047c00470a8740000bb/>
JPG, 61 КБ



Мобильные приложения

kooba

MOBILE IMAGE RECOGNITION
TRY IT OUT NOW!!!



SHOW.MOVIEPOSTER

Movie data provided by:



1. POINT
YOUR MOBILE
PHONE CAMERA TO
THE MOVIE
POSTER.

2. SNAP A
PICTURE AND SEND
IT:

IN SWITZERLAND:
MMS TO 5555 (OR
079 394 57 00
FOR ORANGE
CUSTOMERS)

IN GERMANY:
MMS TO 84000

EVERYWHERE:
EMAIL TO
M@KOOABA.COM

3. FIND ALL
RELEVANT INFOR-
MATION ABOUT THE
MOVIE ON YOUR
MOBILE PHONE

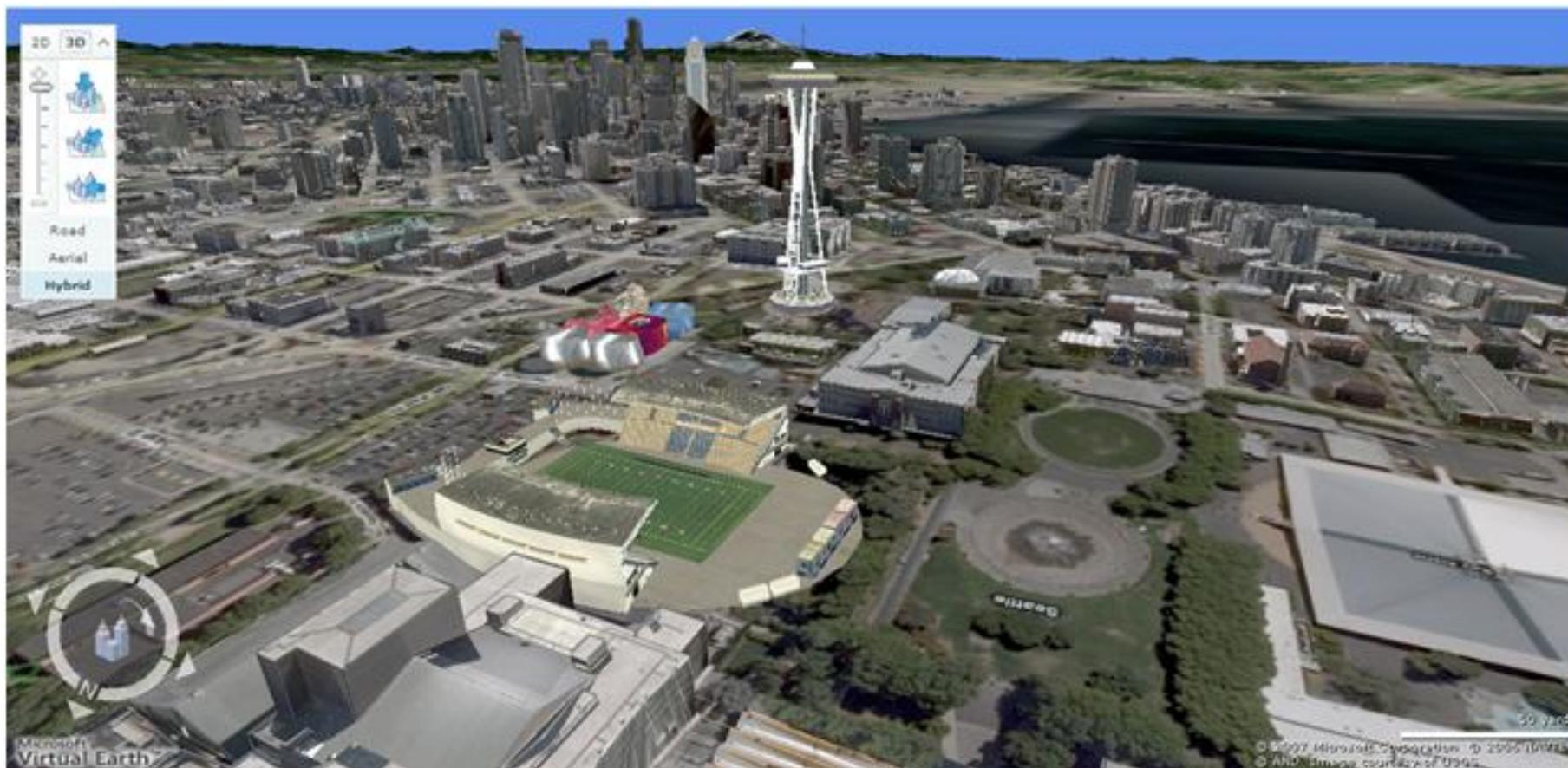


Дополненная
реальность

www.kooba.com



Трехмерные карты



Изображение из Microsoft's [Virtual Earth](#)
(аналогичные [Google Earth](#))



Робототехника



Зрение автомобилей от [Mobileye](#)



Перспектива!

3. О курсе



Зачем этот курс?

- Познакомиться с основами компьютерного зрения
- Узнать и попробовать на практике некоторые методы распознавания и машинного обучения
- Научиться работать на Python
- Получить оценку за спецкурс

- Результаты курса играют большую роль при собеседовании в лабораторию компьютерной графики



Структура курса

- 8 лекций
- 4 домашних задания на Python
- Страница курса на сайте:

<http://машграф.рф/>

<http://courses.graphics.cs.msu.ru>



План курса

- (2 марта)Введение + цвет
 - *Тьюториал по питону (Прокудин-Горский)*
- (9 марта) Лекции нет, выходной
- (16 марта) Простой анализ изображений
 - *Распознавание автомобильных номеров/клад (конкурс)*
- (23 марта) Основы обработки изображений
- (30 марта) Классификация изображений и введение в машинное обучение
 - *Классификация дорожных знаков через HOG + SVM (конкурс)*
- (6 апреля) Выделение объектов на изображении
- (13 апреля) Нейросетевые алгоритмы
 - *Детектор объектов - развитие HOG + SVM (конкурс)*
- (20 апреля) Поиск и большие коллекции
- (27 апреля) Распознавание лица человека



Дополнительный практикум

- Детектор объектов с библиотекой P.Dollar
- Поиск полудубликатов через GIST и ANN с библиотекой vlfeat
- Перенос выражений лиц с библиотекой intraface + vlfeat
- Задание по нейросетям с библиотекой Caffe



Про группу «компьютерное зрение»



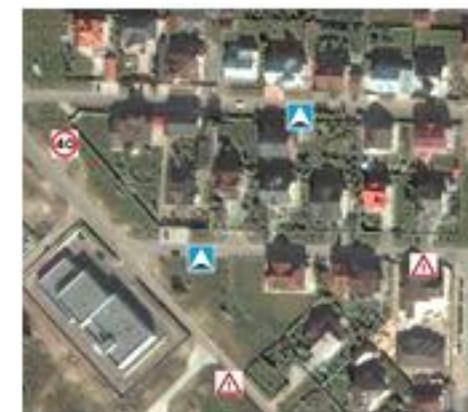
Видеоаналитика



«Person of interest»



Дорожные знаки

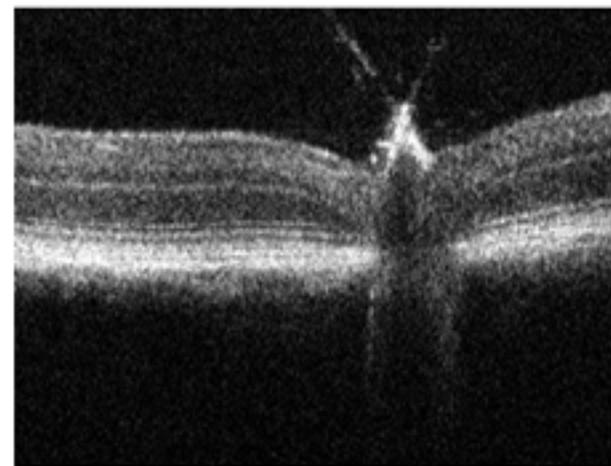


Построение карт

4. Цифровое изображение



Изображение



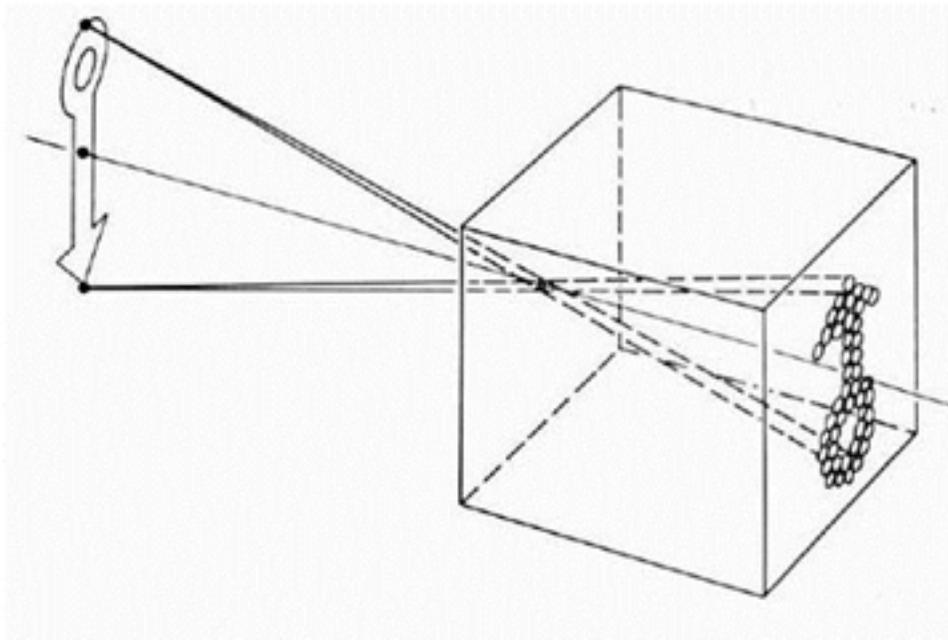
Изображение оптическое – картина, получаемая в результате прохождения через оптическую систему лучей, распространяющихся от объекта, и воспроизводящая его контуры и детали.

(Физический энциклопедический словарь.)

Оптические системы могут быть очень разными!



Камера-обскура



- Простейшее устройство для получения изображений
- Основная модель оптической системы получения изображений

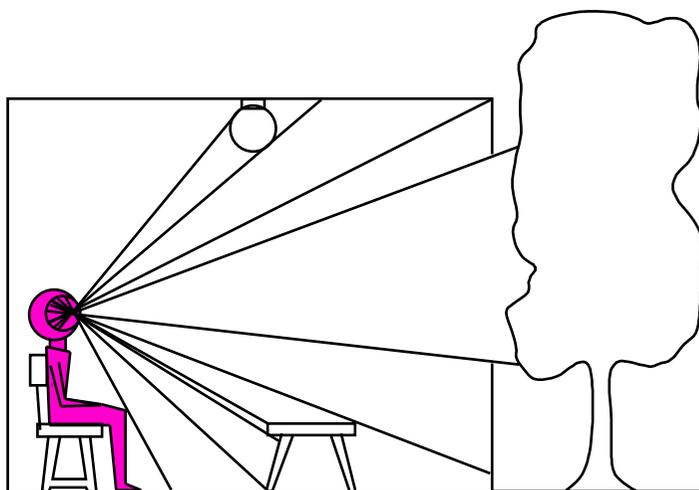
Математическая модель – перспективная проекция:

- Пучок лучей проходит через одну точку (точечное отверстие)
 - Эта точка называется «центр проекции» (фокальная точка / **focal point**)
- Изображение формируется на картинной плоскости (**image plane**)

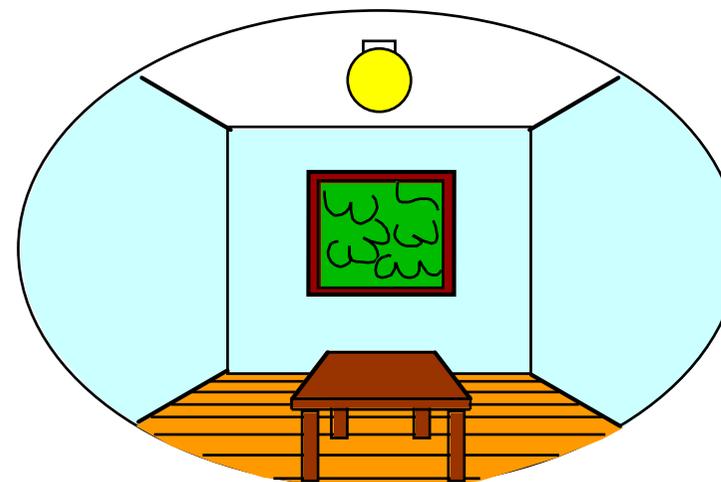


Машина Понижения Размерности

3D мир



2D картина



Что мы теряем?

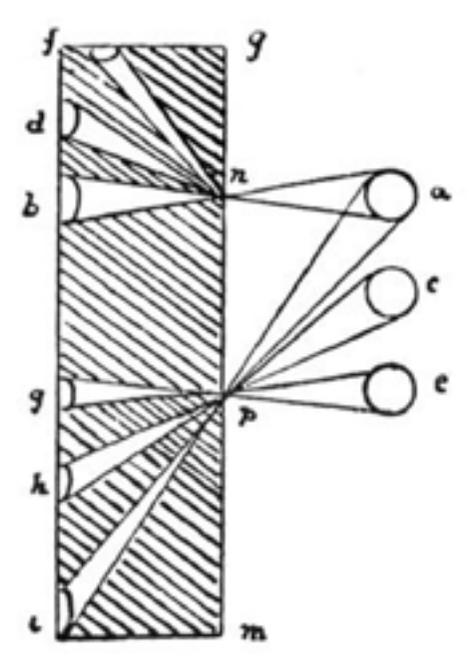
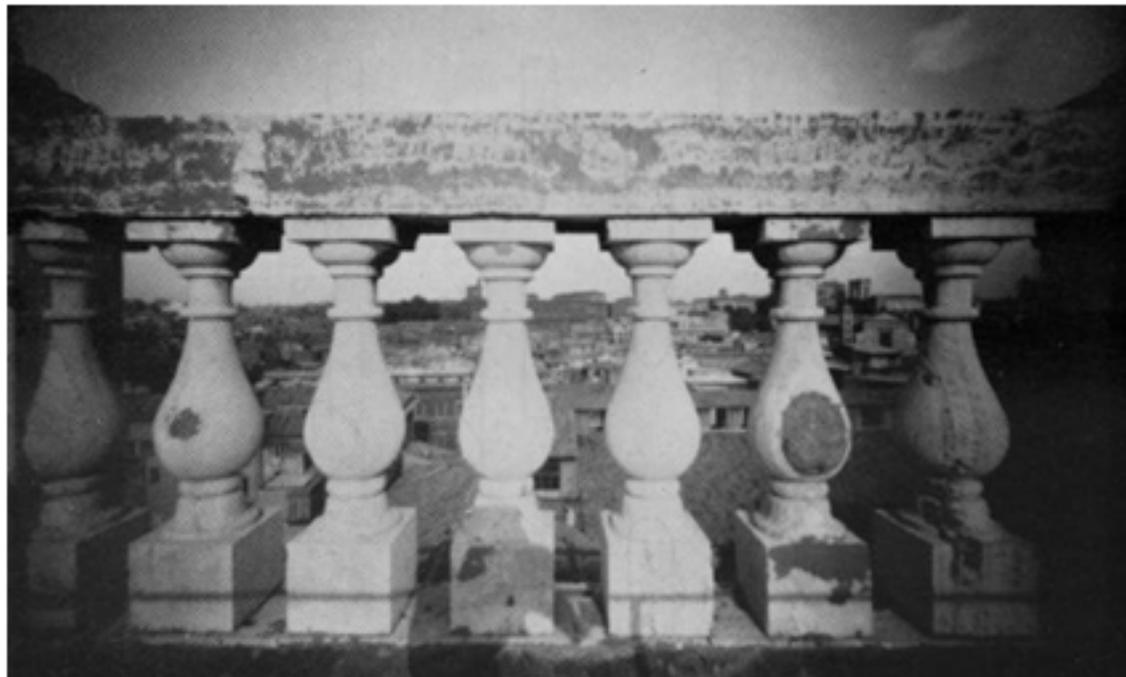
- Углы
- Расстояния и длины

Подробнее геометрическая
модель рассматривается
позже



Перспективные искажения

- Крайние колонны кажутся толще
- Эти искажения вызваны не погрешностью линз!
- Проблема была отмечена еще да Винчи

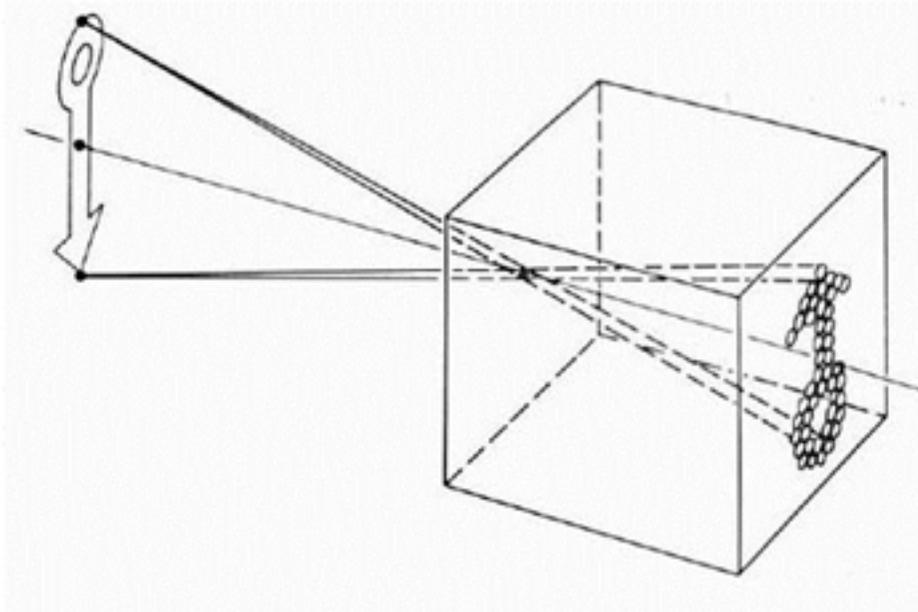




Перспективные искажения: Люди



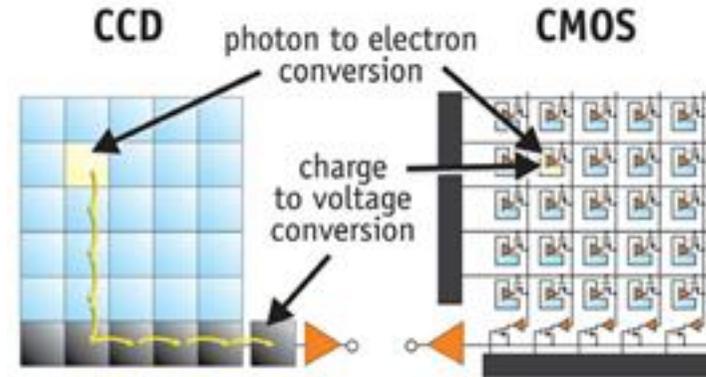
Современная камера



Тот же самый принцип, но с объективом и цифровой матрицей



Цифровая камера - дискретизация



CCDs move photogenerated charge from pixel to pixel and convert it to voltage at an output node, CMOS imagers convert charge to voltage inside each pixel.

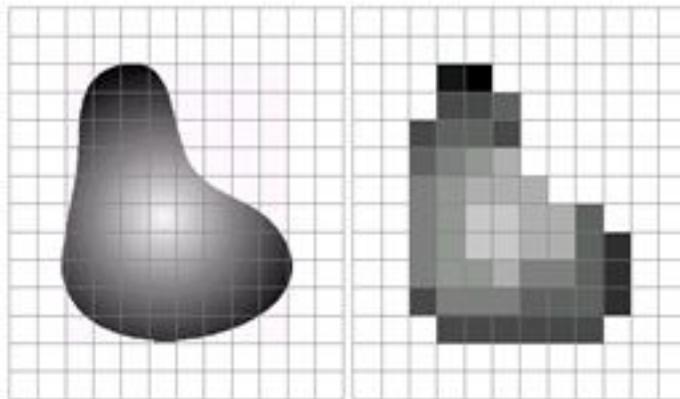
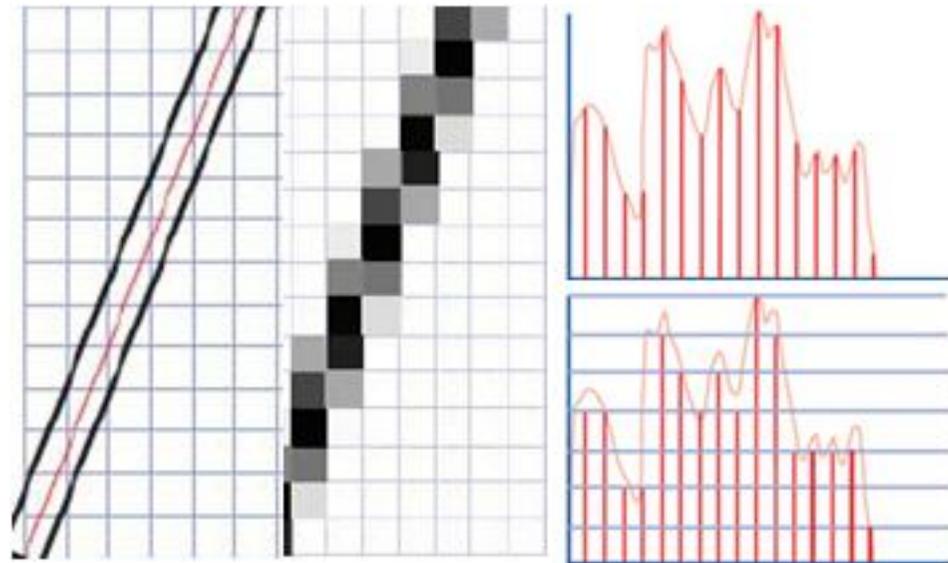


FIGURE 2.17 (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

По пространству



По яркости



Цифровое изображение

Функция интенсивности (яркости) канала, заданная на 2х мерной сетке (матрице)

$$I = g(x, y), \{x \in [x_0, x_1], y \in [y_0, y_1]\}$$

Используется дискретное представление

$$I = g(i, j), \{i = \overline{1, n}, j \in \overline{1, m}\}$$



Первое цифровое фото (1957)



Разрешение
176*176 пикселей

<http://listverse.com/history/top-10-incredible-early-firsts-in-photography/>

5. ЦВЕТ



Цветные фотографии??

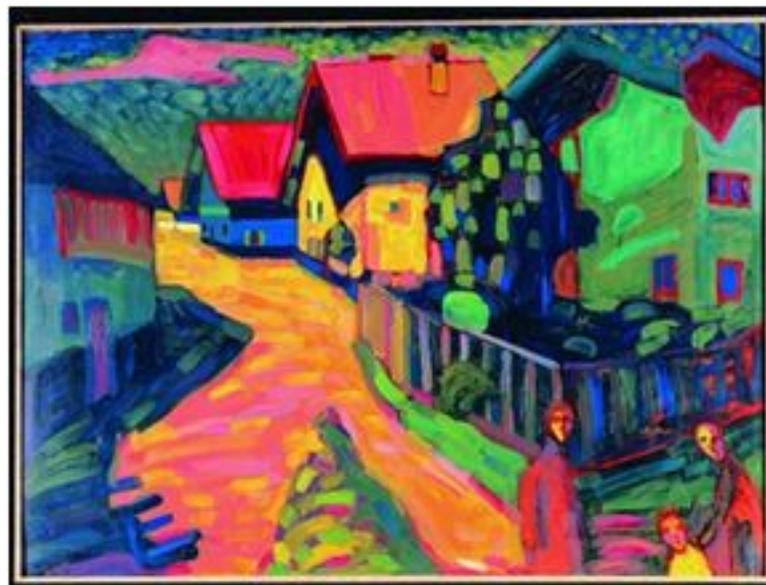




Что такое цвет?

Цвет – это психологическое свойство нашего зрения, возникающее при наблюдении объектов и света, а не физические свойства объектов и света (S. Palmer, *Vision Science: Photons to Phenomenology*)

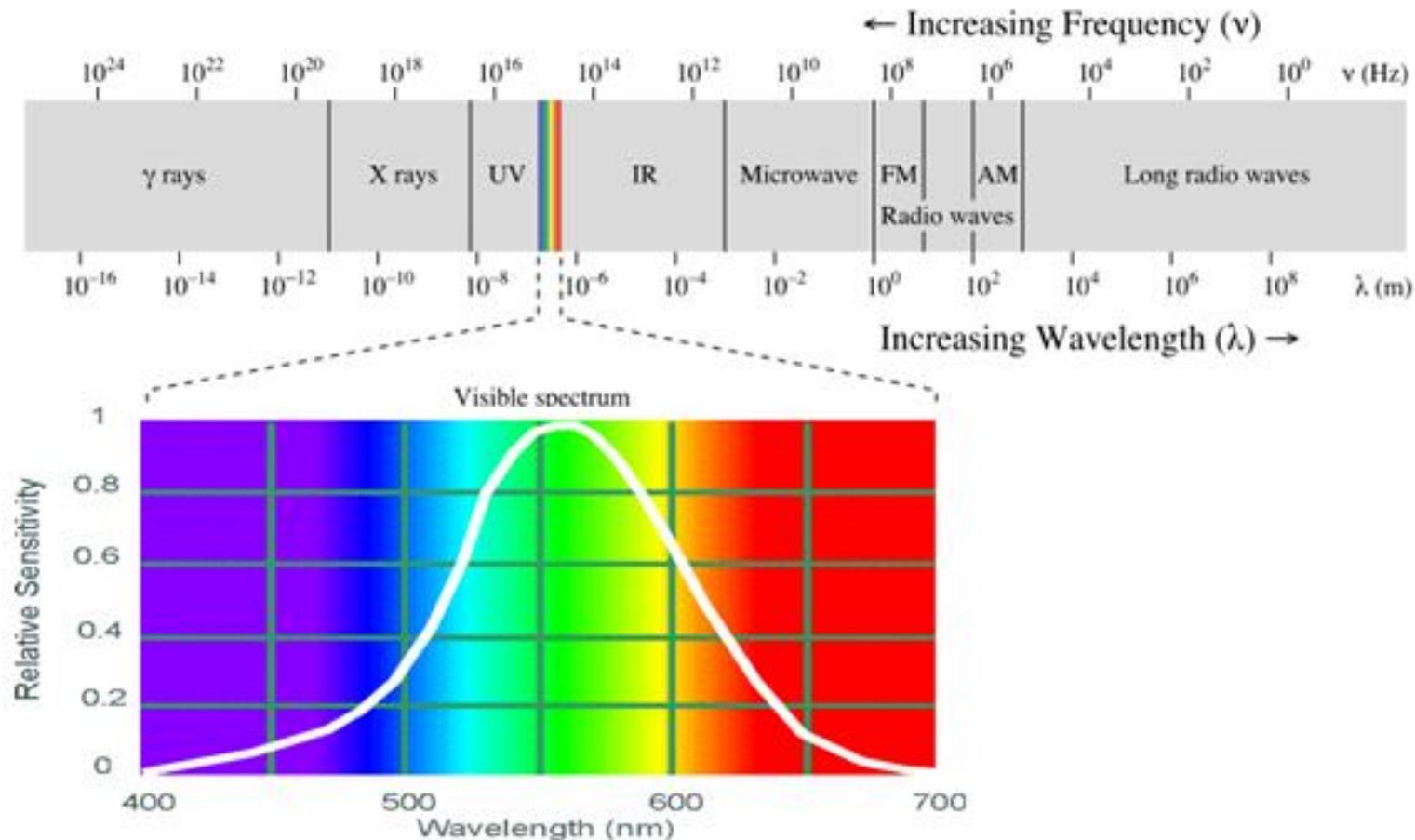
Цвет – это результат взаимодействия света, сцены и нашей зрительной системы



Wassily Kandinsky (1866-1944), Murnau Street with Women, 1908



Электромагнитный спектр



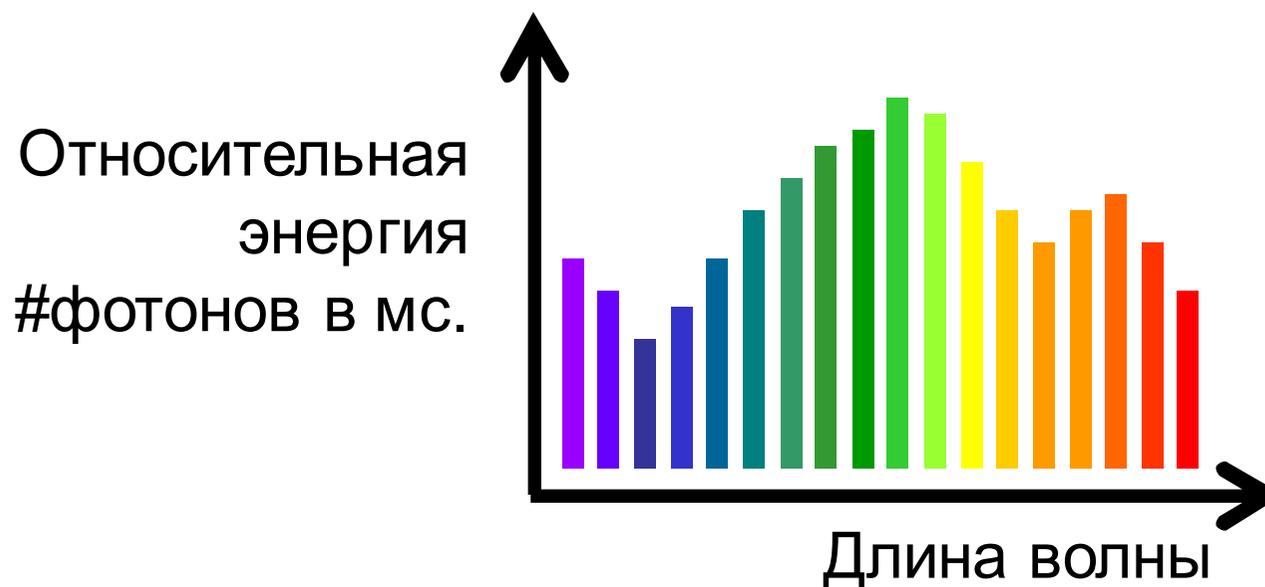
Human Luminance Sensitivity Function

Мы видим свет именно в таком диапазоне, потому что на него приходится значительная доля энергии солнца.



Физика света

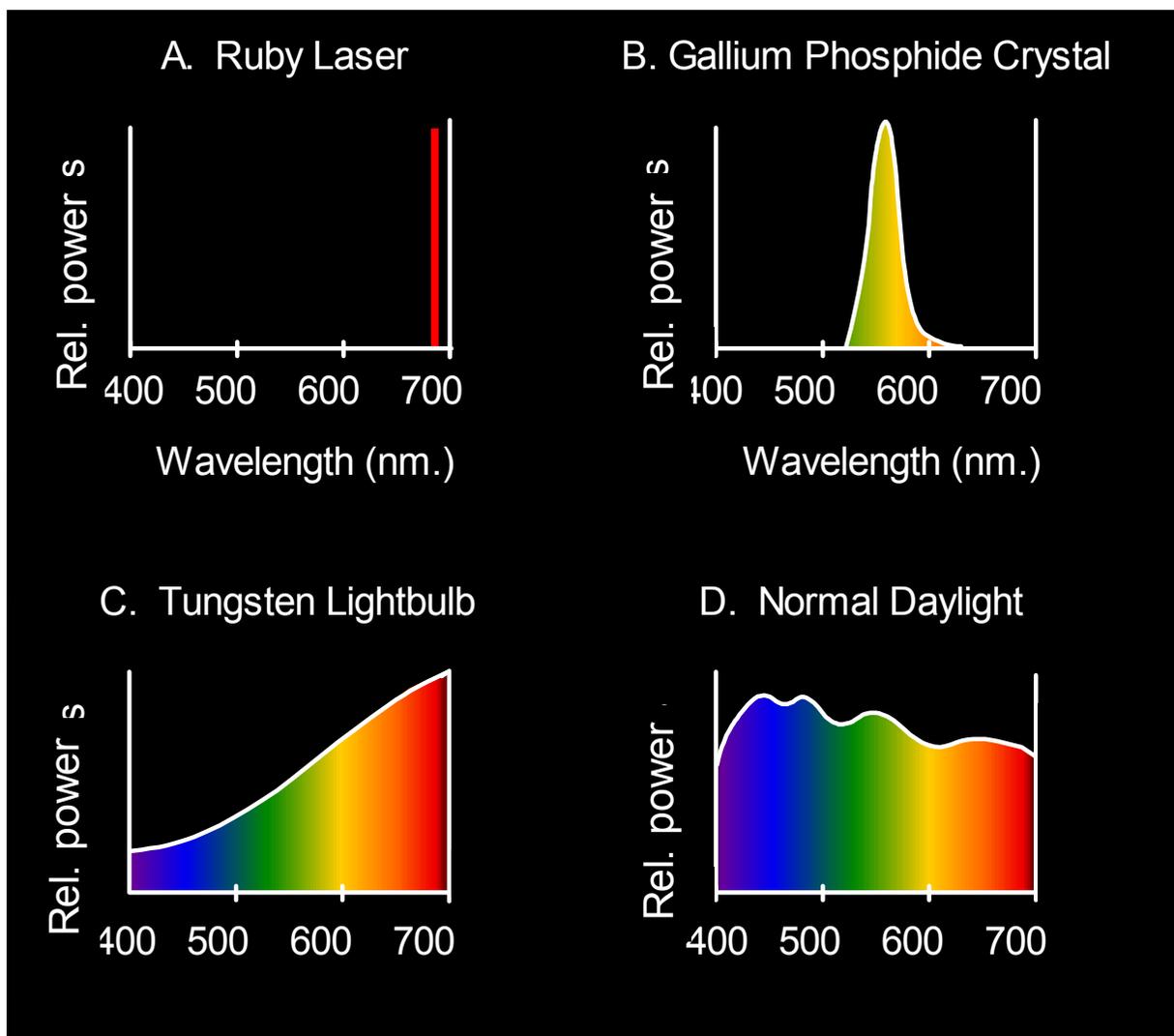
Любой источник света можно полностью описать спектром: количество излученной энергии в единицу времени для каждой длины волны в интервале 400 - 700 nm.





Физика света

Примеры спектров разных источников света

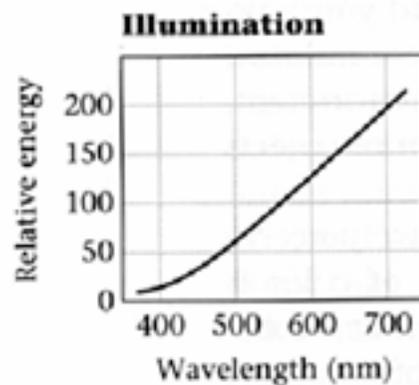




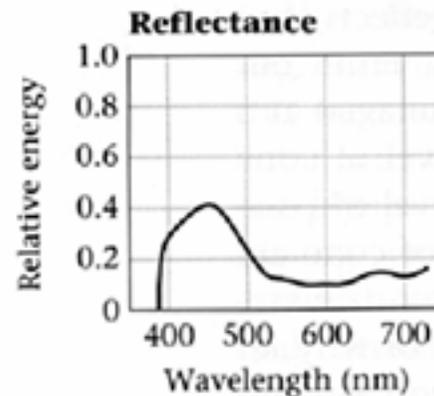
Взаимодействие света и объектов



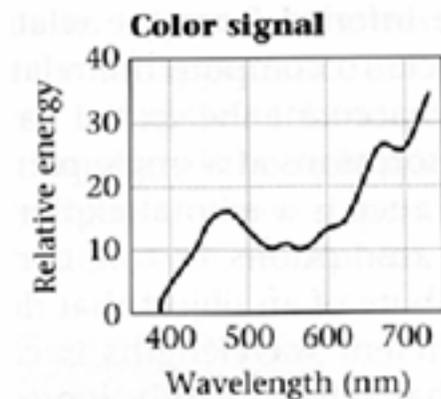
Отраженный свет это результат взаимодействия излучаемого света и поверхности



• *



=



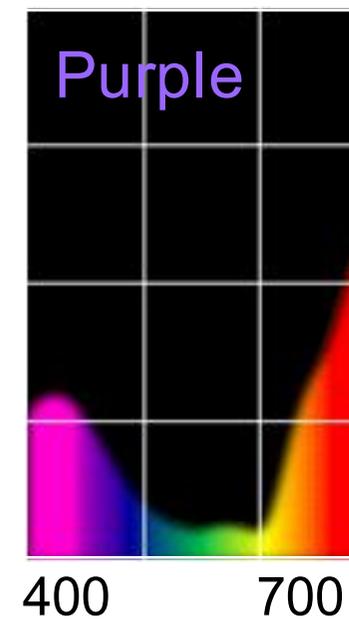
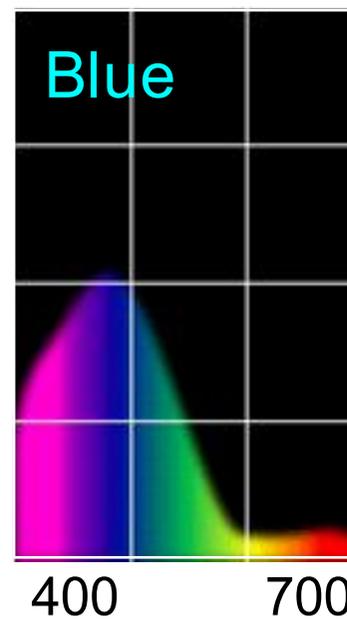
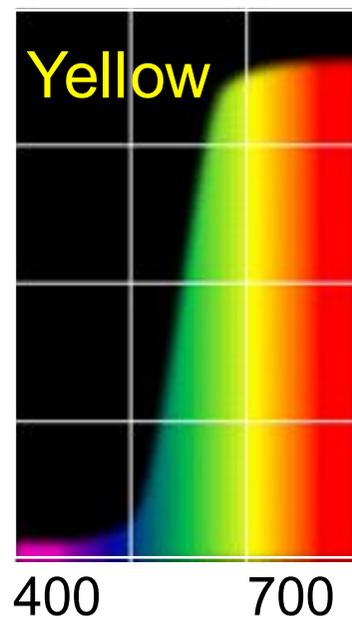
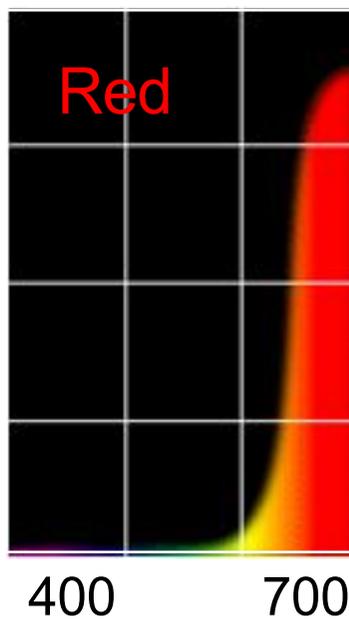


Физика света

Примеры спектров отраженного света от предметов

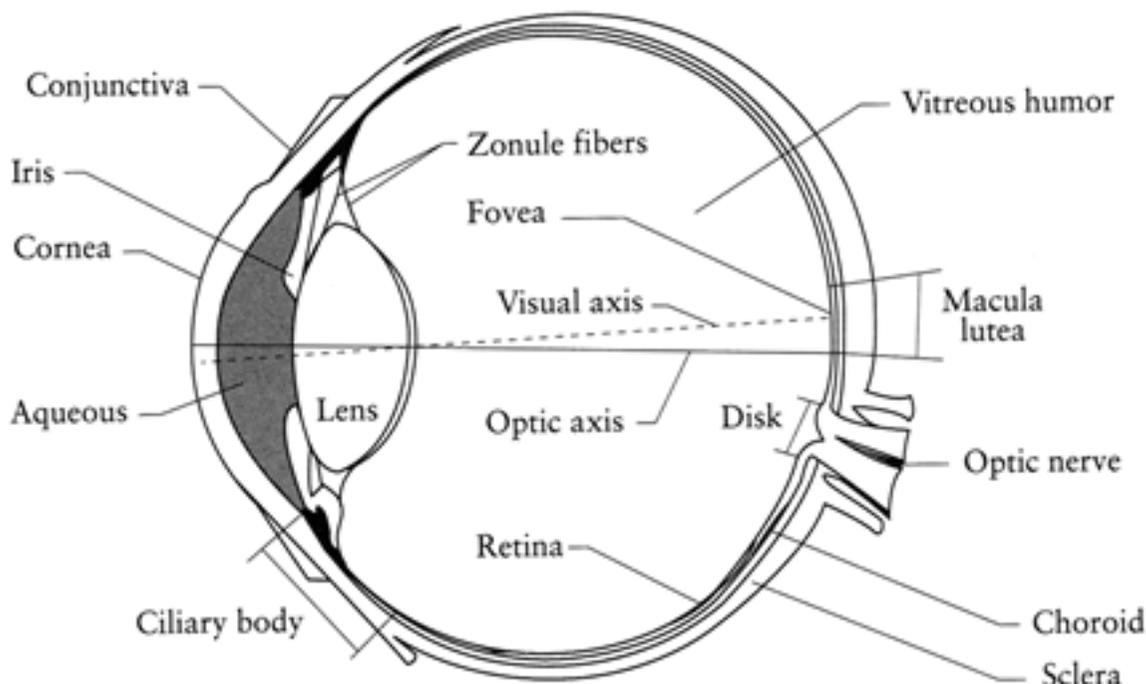


% Отраженного света





Человеческий глаз



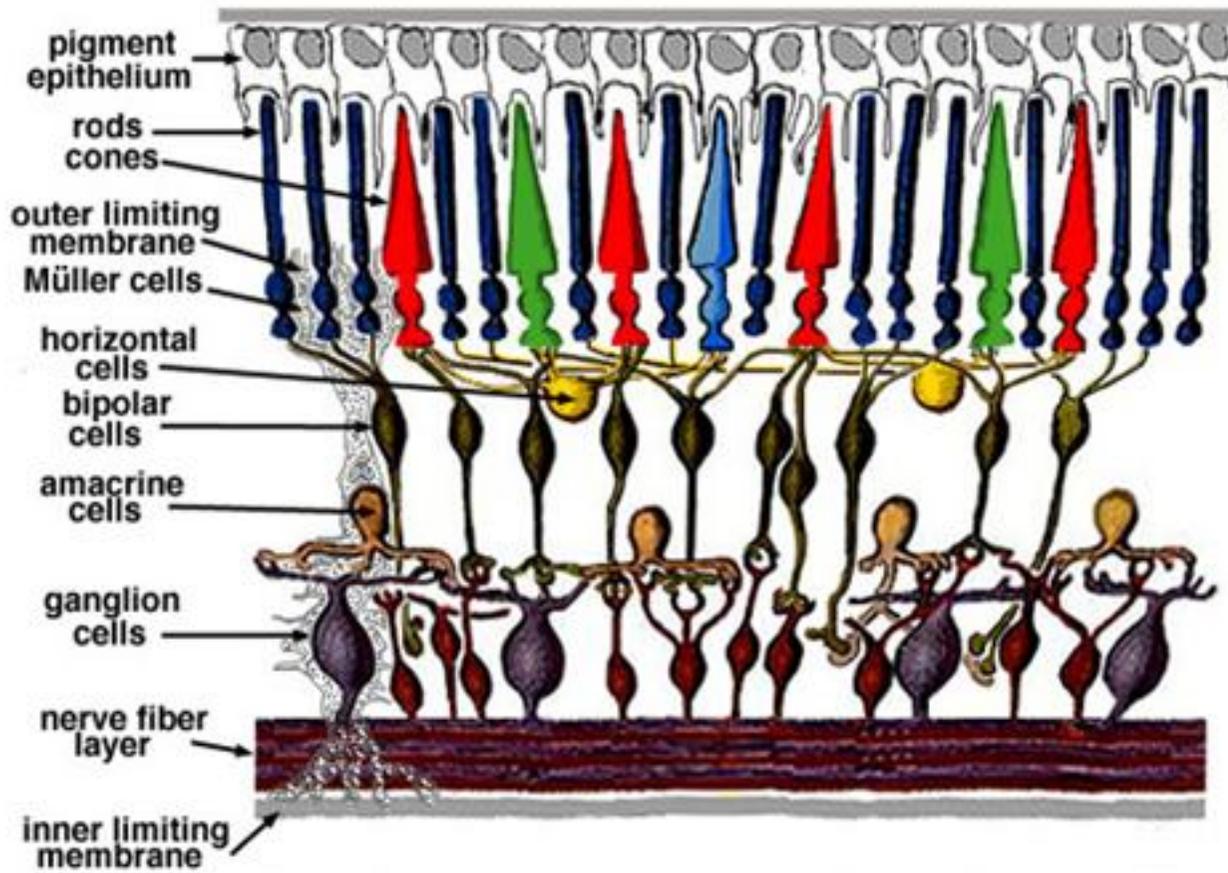
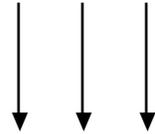
Глаз как камера!

- **Радужка** – цветная пленка с радиальными мышцами
- **Зрачок** - отверстие (апертура), диаметр управляется радужкой
- **Хрусталик** – «линза», меняющая форму под действием мышц
- Где матрица?
 - Клетки-фоторецепторы на сетчатке



Сетчатка глаза

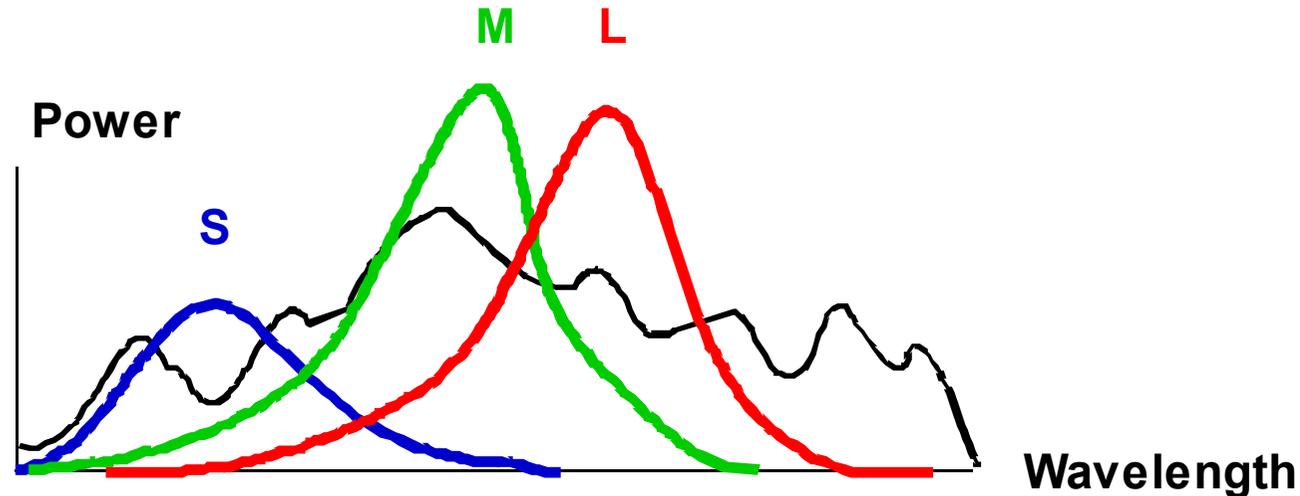
Свет



- Палочки (Rods) измеряют яркость
- Колбочки (Cones) измеряют цвет



Восприятие цвета

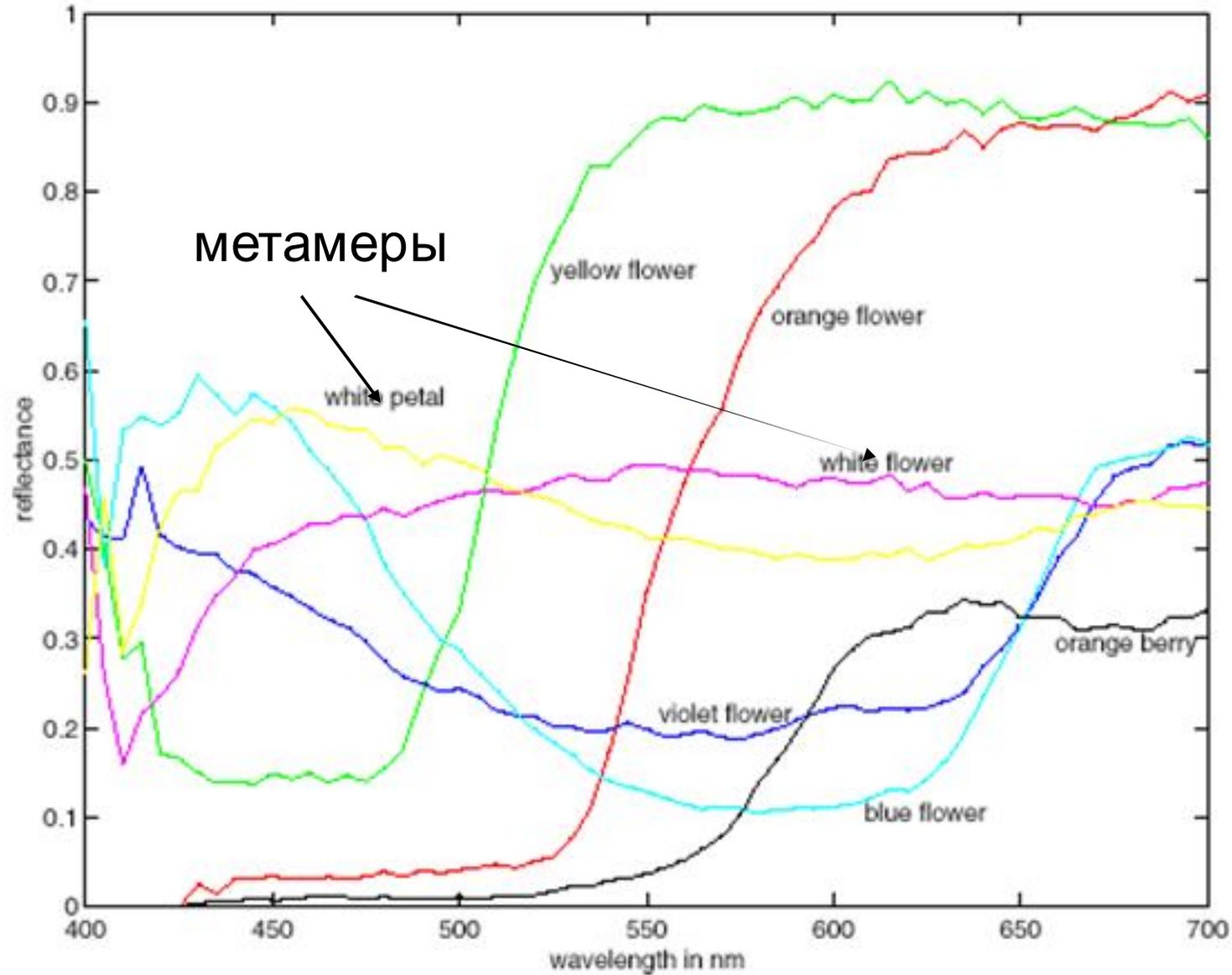


Палочки и колбочки – фильтры спектра

- Спектр умножается на кривую отклика, производится интегрирование по всем длинам волн
 - Каждый тип колбочек даёт 1 число
- В: Как же мы можем описать весь спектр 3мя числами?
- О: Мы и не можем! Большая часть информации теряется.
 - Два разных спектра могут быть неотличимы
 - » Такие спектры называются **метамеры**



Спектры некоторых объектов



6. Цветовые модели



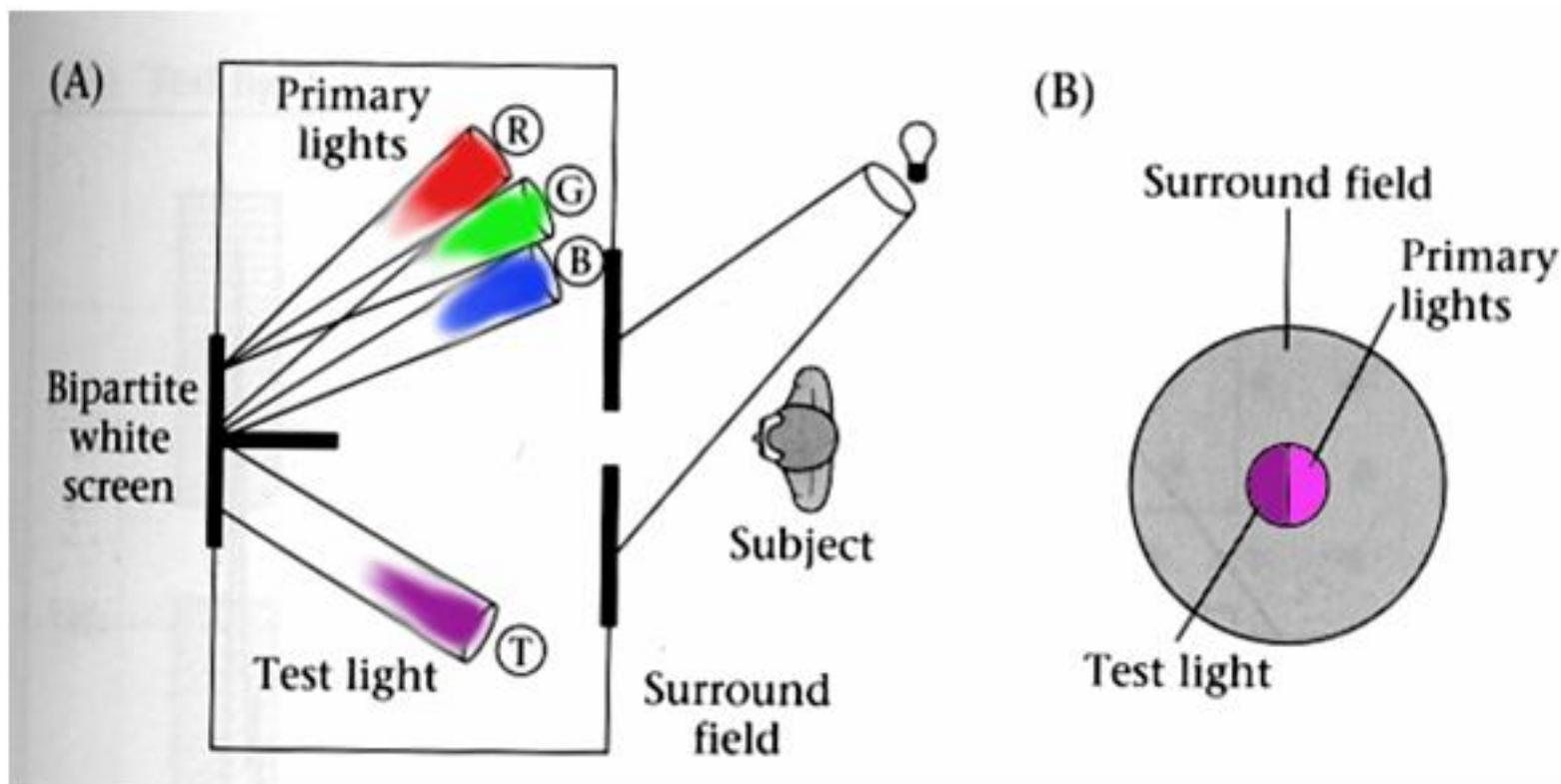
Трихроматическая теория

- Любой видимый «свет» можно описать композицией 3х основных цветов
- Для одного и того же «света» люди выберут одинаковые композиции основных цветов
 - Исключения: цветовая слепота
- История восходит к 18^у веку (Томас Юнг)

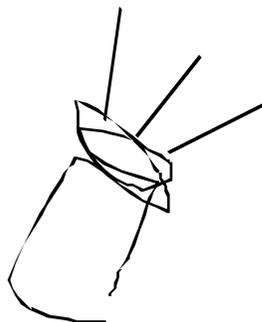
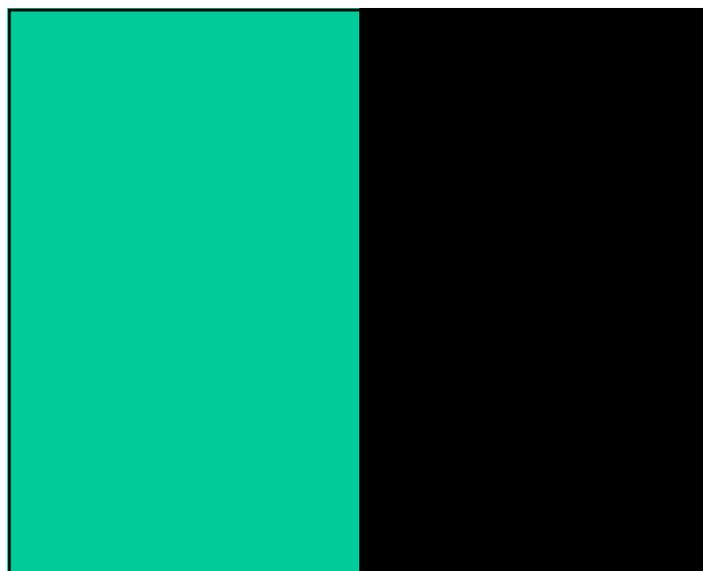


Стандартизация восприятия цвета

Экспериментальная проверка трихроматической теории – сопоставление цветов

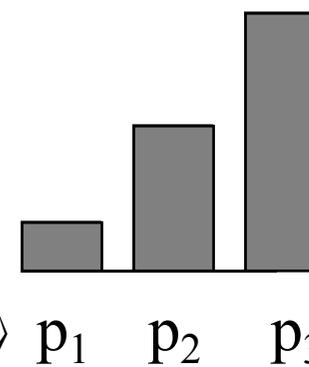
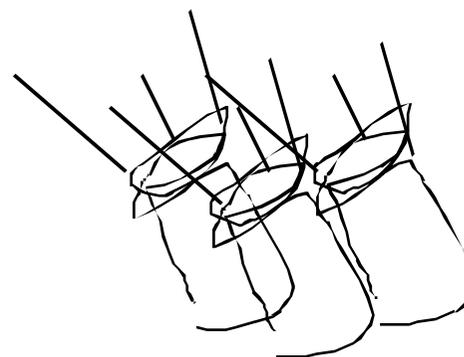
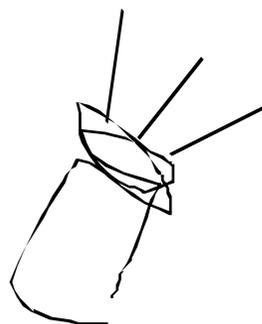
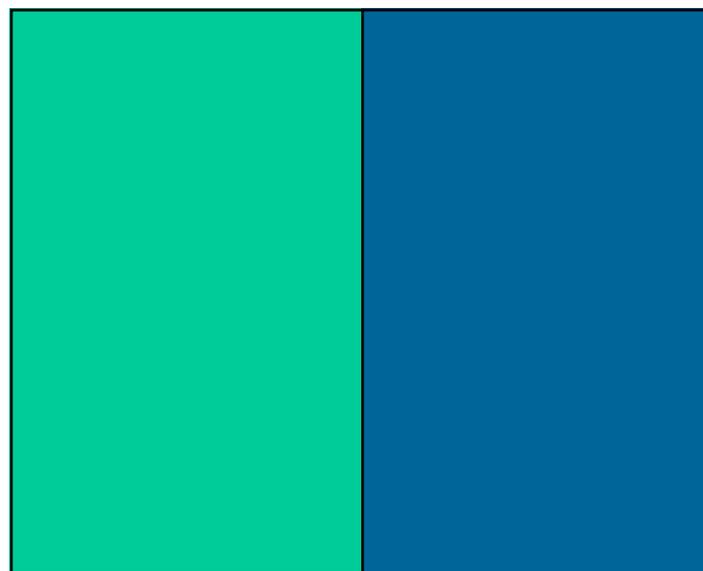


Эксперимент №1



Source: W. Freeman

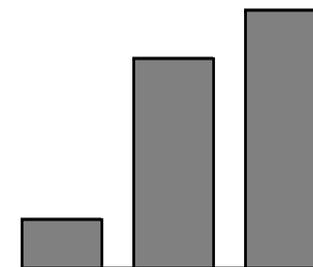
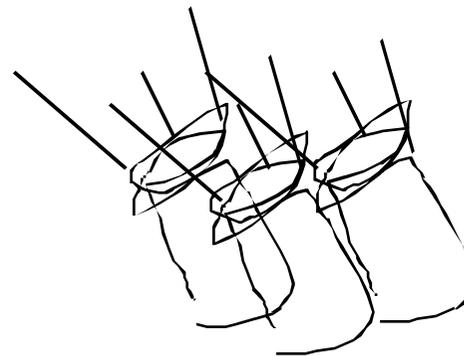
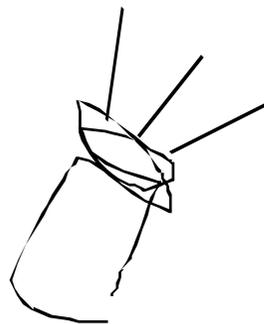
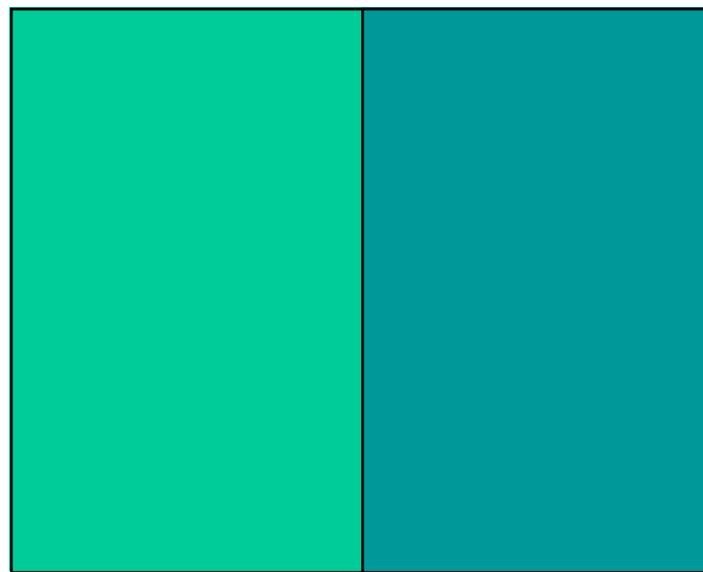
Эксперимент №1



Source: W. Freeman



Эксперимент №1

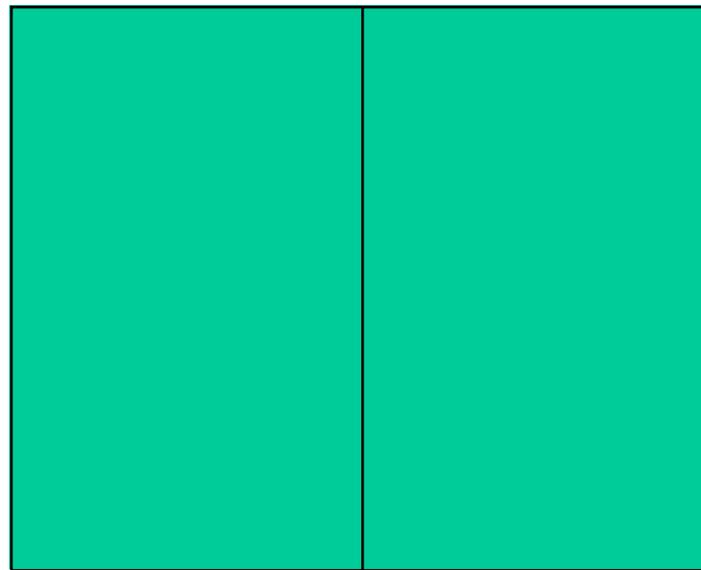


p_1 p_2 p_3

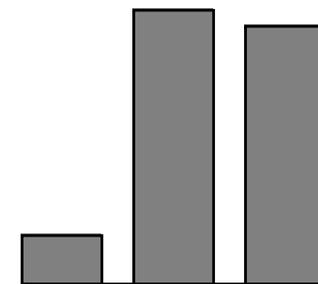
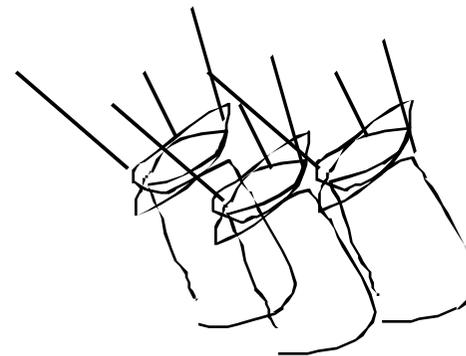
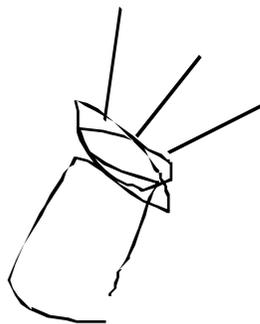
Source: W. Freeman



Эксперимент №1



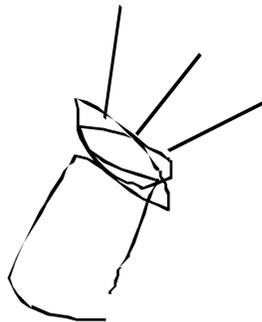
Основные цвета,
необходимые для
сопоставления



p_1 p_2 p_3

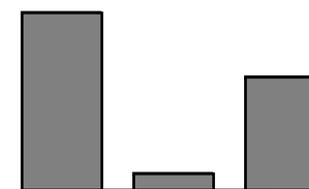
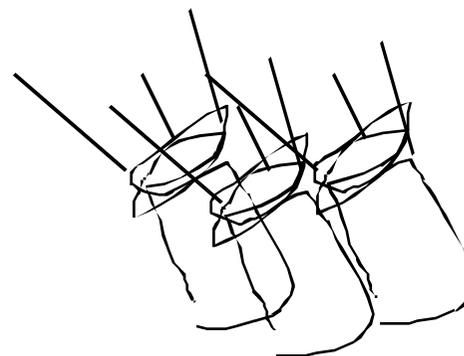
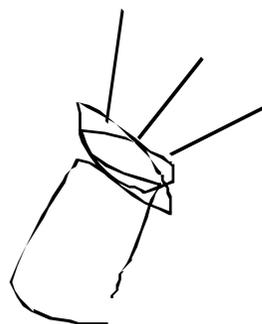
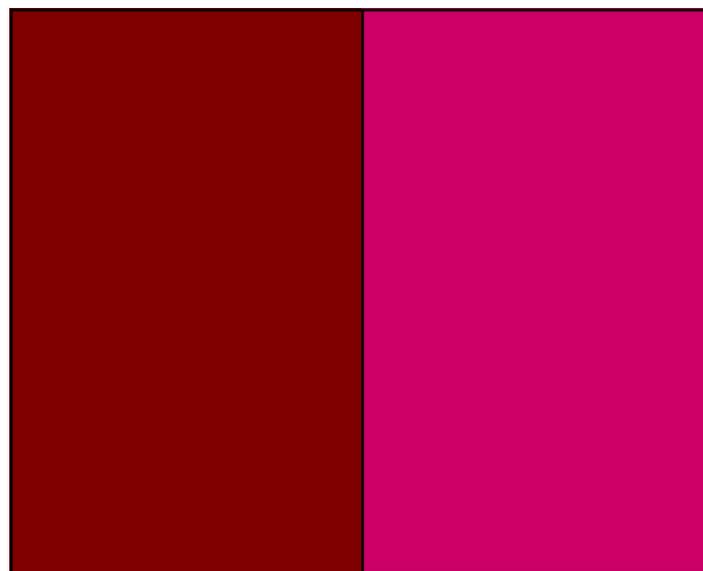
Source: W. Freeman

Эксперимент №2



Source: W. Freeman

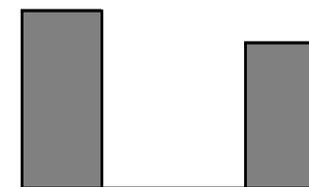
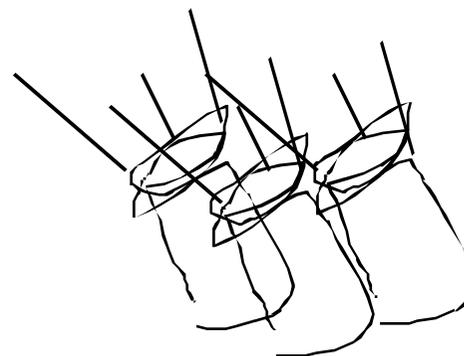
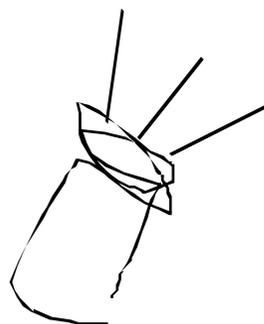
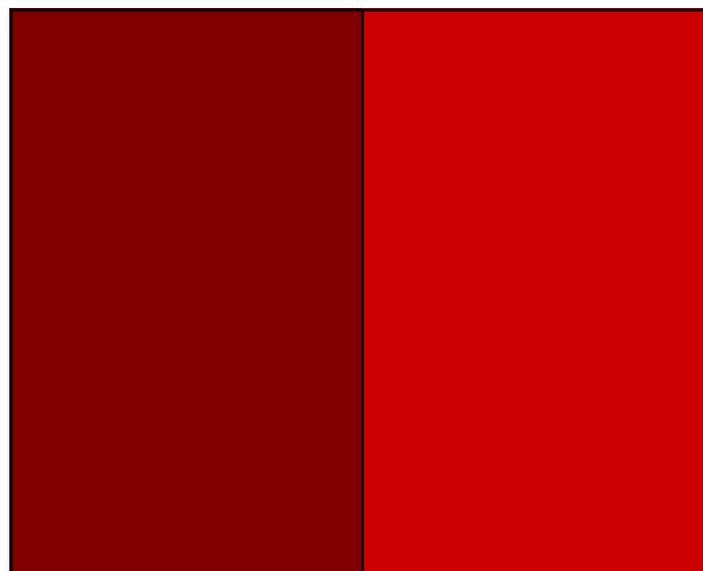
Эксперимент №2



p_1 p_2 p_3

Source: W. Freeman

Эксперимент №2



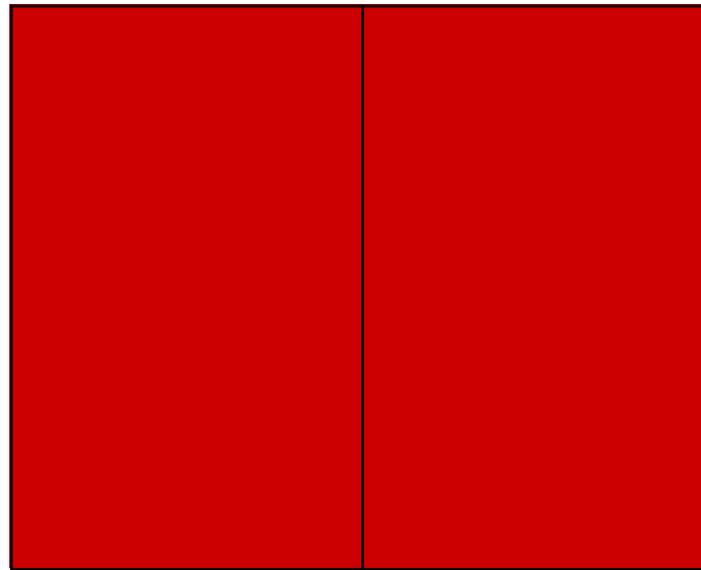
p_1 p_2 p_3

Source: W. Freeman

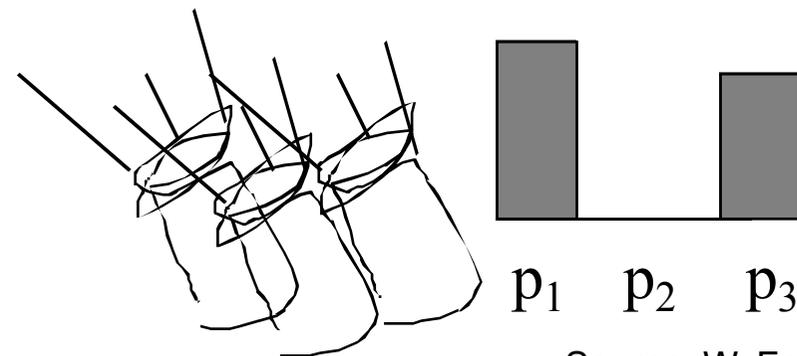
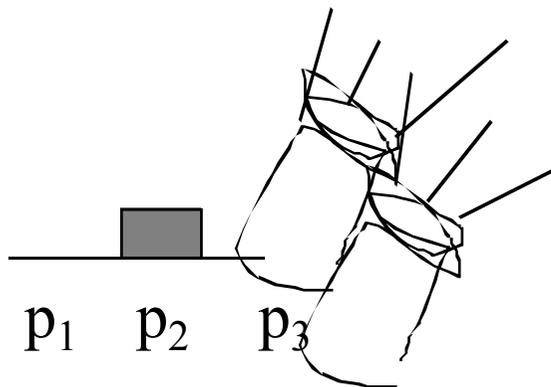
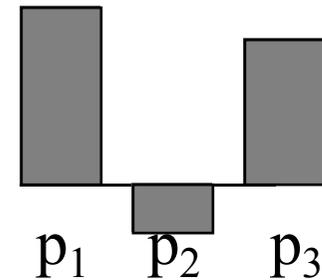


Эксперимент №2

Мы называем m «отрицательным» весом основного цвета, если цвет нужно добавлять к сопоставляемому свету.



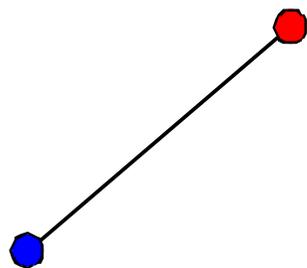
Веса основных цветов, необходимых для сопоставления:



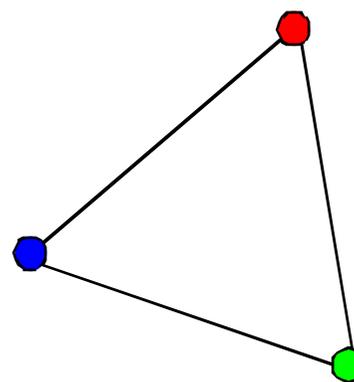


Линейные цветовые пространства

- Определяются выбором 3х основных цветов
- «Координаты цвета» задаются весами основных цветов, необходимых для сопоставления
- Каждая координата кодируется 1-2 байтами
- *Функции сопоставления: веса, необходимые для сопоставления с когерентными источниками света*



Смешение двух основных цветов

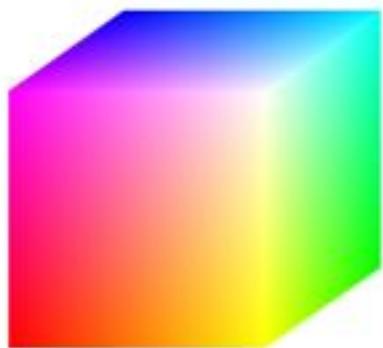


Смешение трех цветов



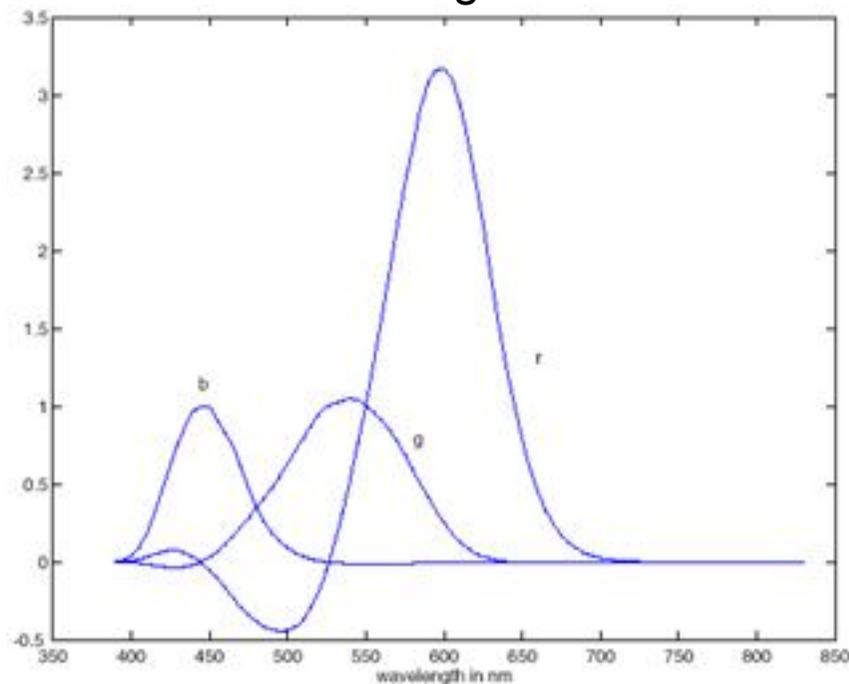
Линейные цветовые модели: RGB

- Основные цвета – монохроматические (в мониторе им соответствует три вида фосфоров)
- *Вычитание* необходимо для соответствия некоторым длинам волны



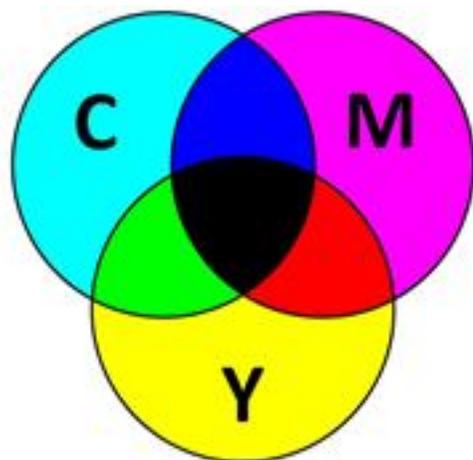
■ $p_1 = 645.2 \text{ nm}$
■ $p_2 = 525.3 \text{ nm}$
■ $p_3 = 444.4 \text{ nm}$

RGB matching functions

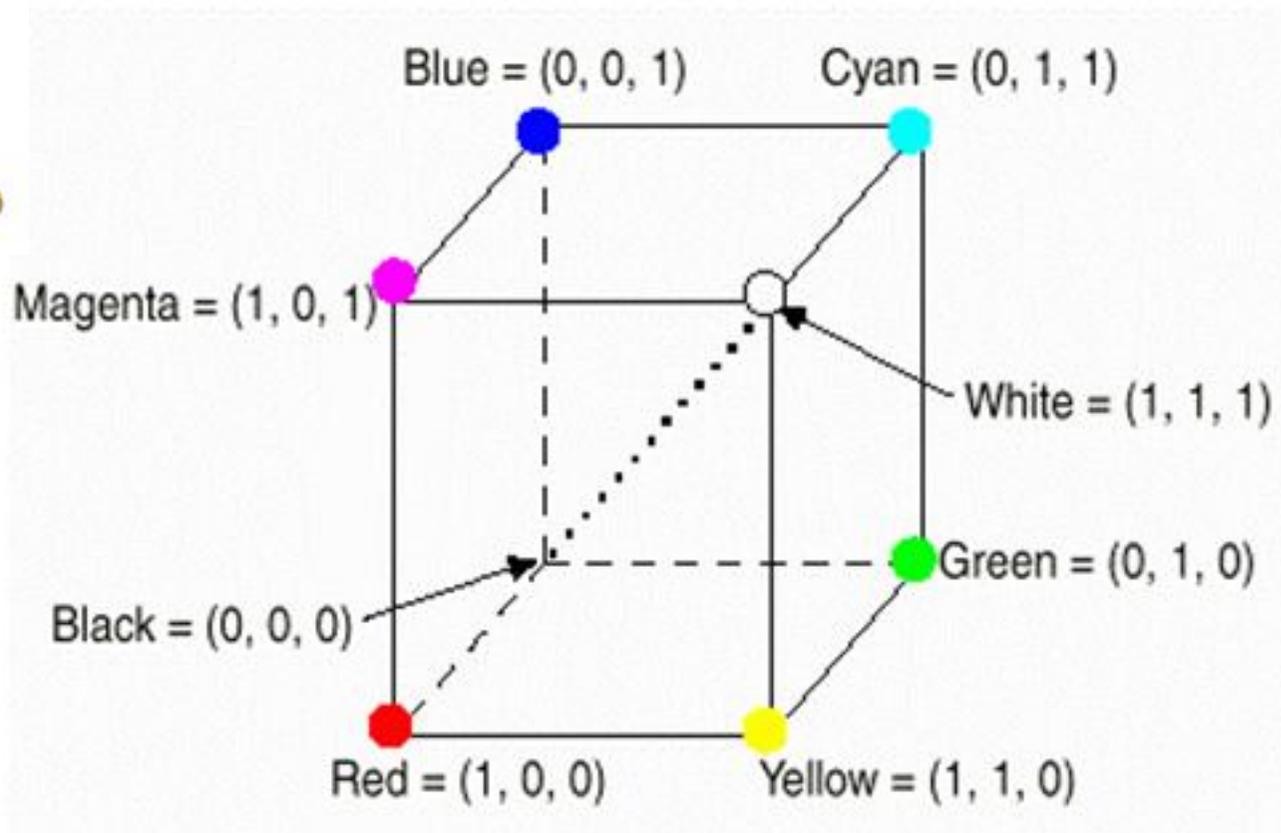




Цветовой куб и CMYK



CMY



- Аддитивная система – RGB
- Субтрактивная система – CMY
- CMYK для повышения качества

$$C = G + B = W - R$$

$$M = R + B = W - G$$

$$Y = R + G = W - B$$



Модель YIQ

В RGB нет отдельной «яркости», что бывает неудобно

- $Y = .299R + .587G + .114B$

- $I = .596R - .275G - .321B$

- $Q = .212R - .528G + .311B$

- $R = 1.000 Y + 0.956 I + 0.621 Q$

- $G = 1.000 Y - 0.272 I - 0.647 Q$

- $B = 1.000 Y - 1.106 I + 1.703 Q$

- Цветовая модель YIQ используется в коммерческом цветном телевидении США

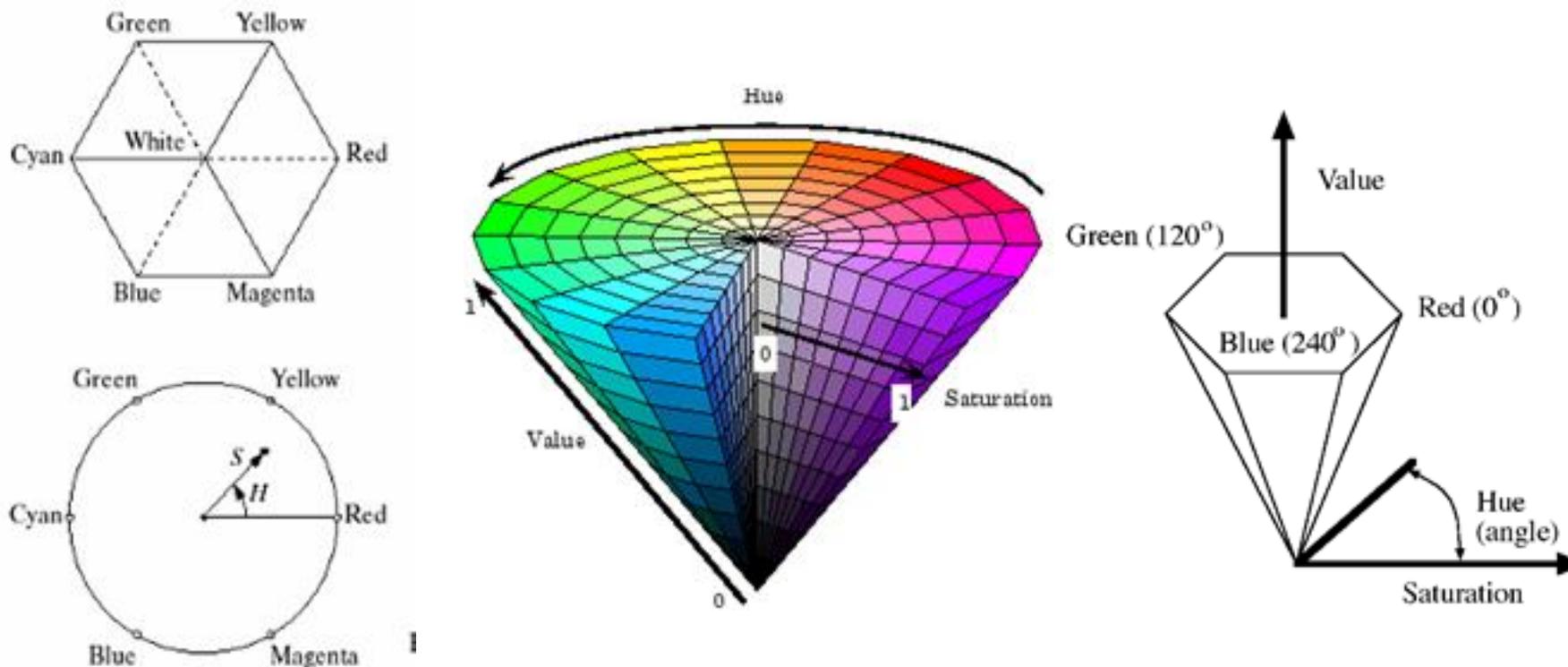
- Модель YIQ совместима с черно-белым телевидением

- Модель YIQ используется в стандарте JPEG

- $I = R - C ; \quad Q = M - G$



Модель HSV (HIS)



Координаты выбраны с учетом человеческого восприятия: Hue (Тон), Saturation(Насыщенность), Value (Intensity) (Интенсивность)



Перевод из RGB в HSV

Conversion of RGB encoding to HSI encoding.

R,G,B : input values of RGB all in range [0,1] or [0,255];

I : output value of intensity in same range as input;

S : output value of saturation in range [0,1];

H : output value of hue in range $[0, 2\pi)$, -1 if S is 0;

R,G,B,H,S,I are all floating point numbers;

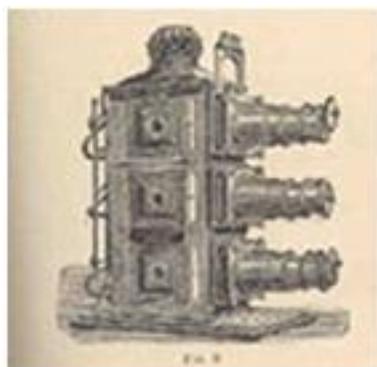
```
procedure RGB_to_HSI( in R,G,B; out H,S,I)
{
  I := max ( R, G, B );
  min := min ( R, G, B );
  if (I ≥ 0.0) then S := (I - min )/I else S := 0.0;
  if (S ≤ 0.0) then { H := -1.0; return; }
  “compute the hue based on the relative sizes of the RGB components”
  diff := I - min;
  “is the point within +/- 60 degrees of the red axis?”
  if (r = I) then H :=  $(\pi/3) * (g - b) / \text{diff}$ ;
  “is the point within +/- 60 degrees of the green axis?”
  else if (g = I) then H :=  $(2 * \pi/3) + \pi/3 * (b - r) / \text{diff}$ ;
  “is the point within +/- 60 degrees of the blue axis?”
  else if (b = I) then H :=  $(4 * \pi/3) + \pi/3 * (r - g) / \text{diff}$ ;
  if (H ≤ 0.0) H := H + 2 $\pi$ ;
}
```

Algorithm 15: Conversion of RGB to HSI.



Первые цветные фотографии

Сергей Прокудин-Горский (1863-1944)
Фотографии Российской империи (1909-1916)



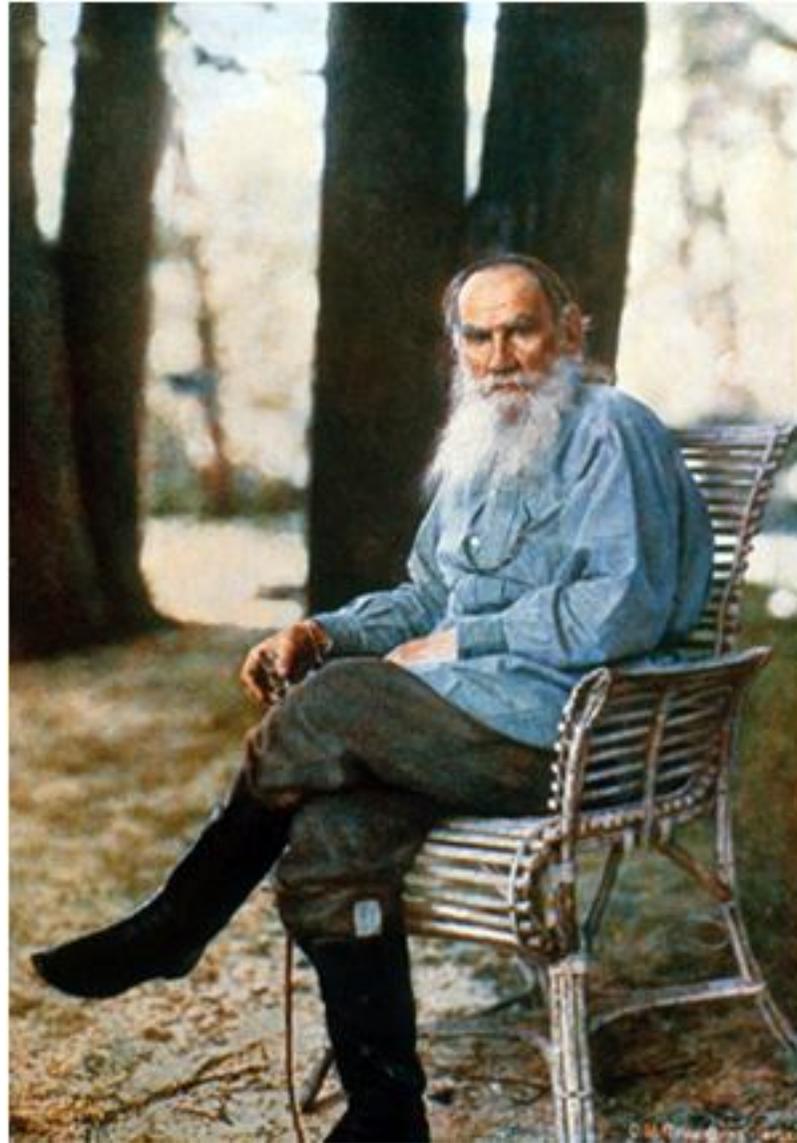
Ламповый проектор



http://en.wikipedia.org/wiki/Sergei_Mikhailovich_Prokudin-Gorskii

<http://www.loc.gov/exhibits/empire/>

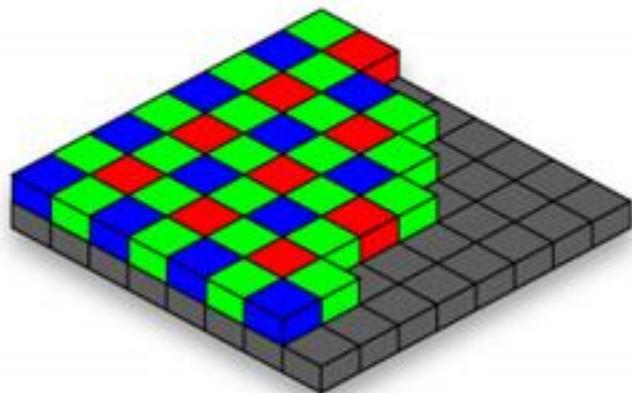
Лев Толстой



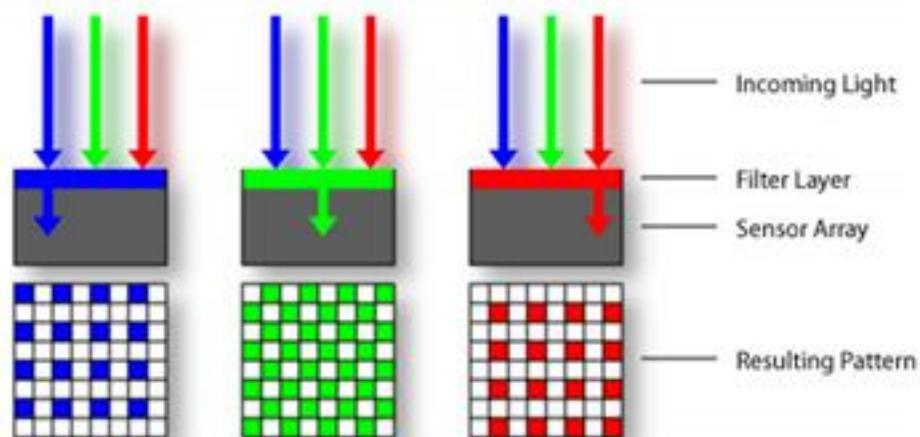
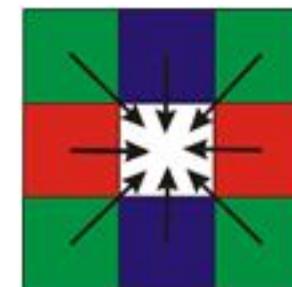


Цветное цифровое изображение

Байеровский шаблон



Демозаикинг (оценка пропущенных значений цвета)





Устранение мозаичности и ошибки



Original image



Bilinear interpolation



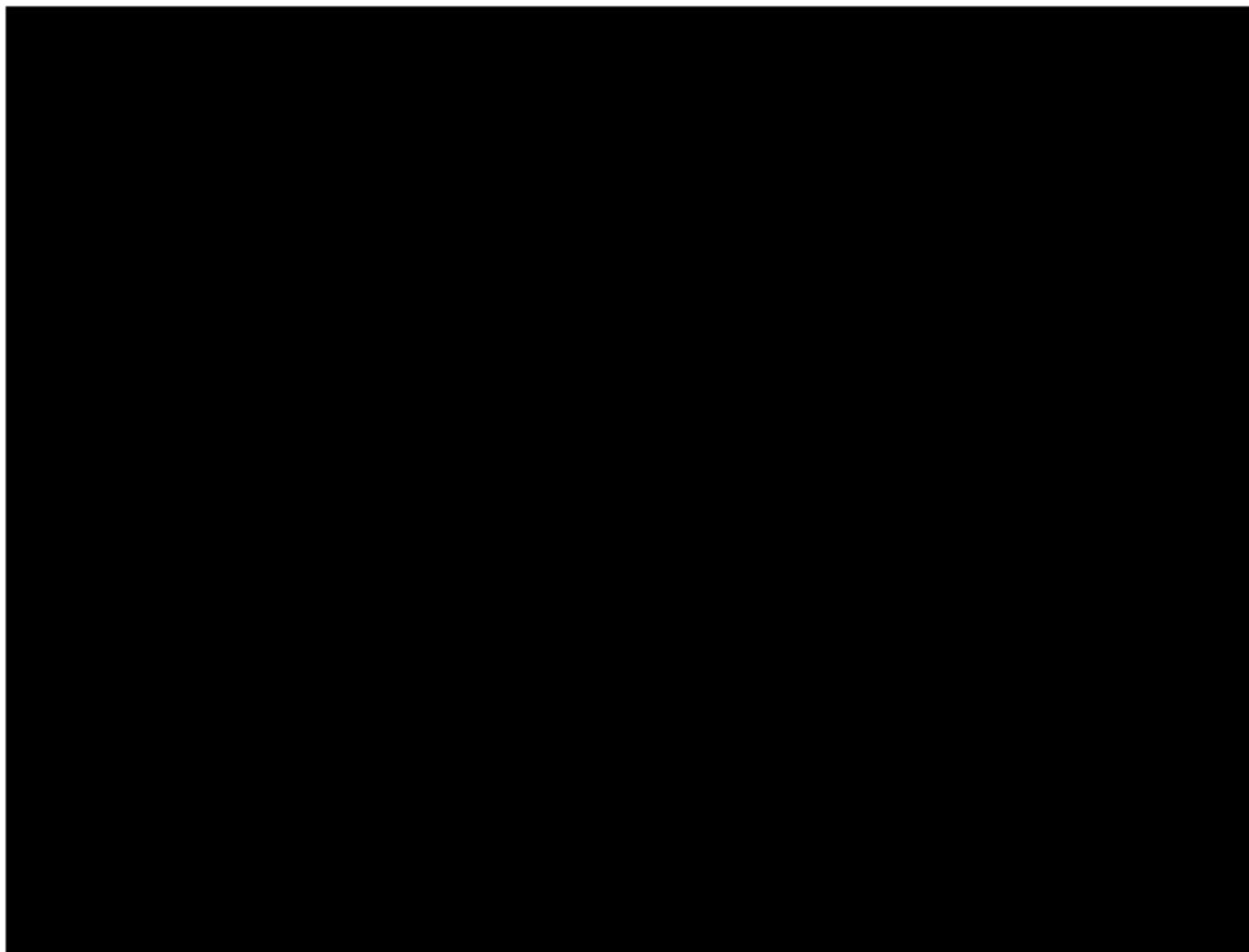
Proposed method

Тонкие черные и белые детали
интерпретируются как изменения цвета

7. Фокус внимания

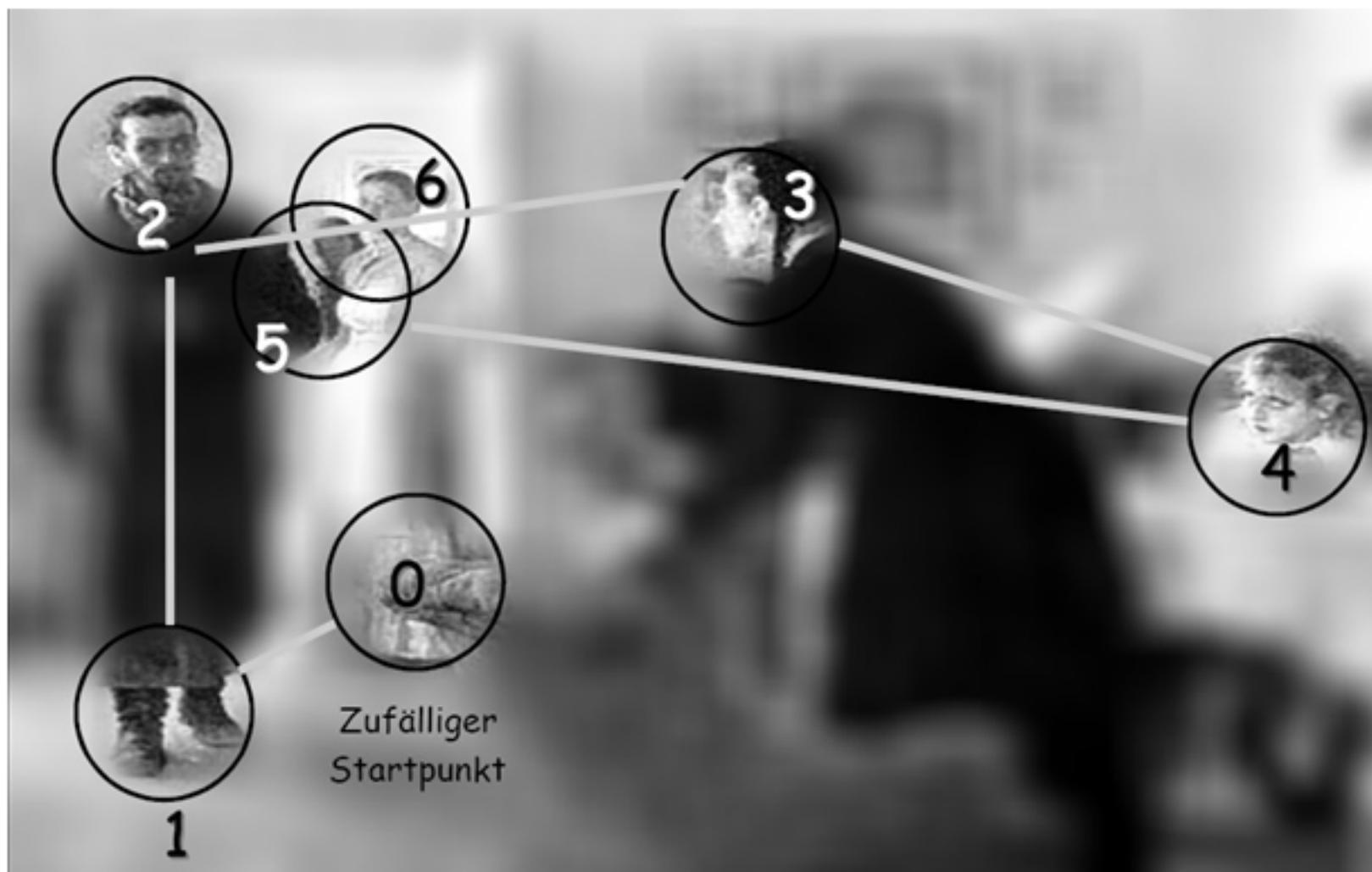


Тест на внимательность





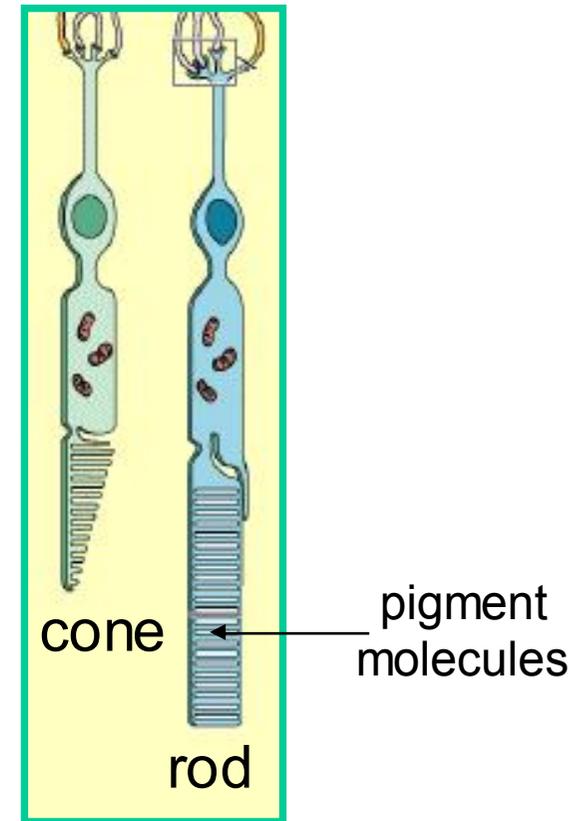
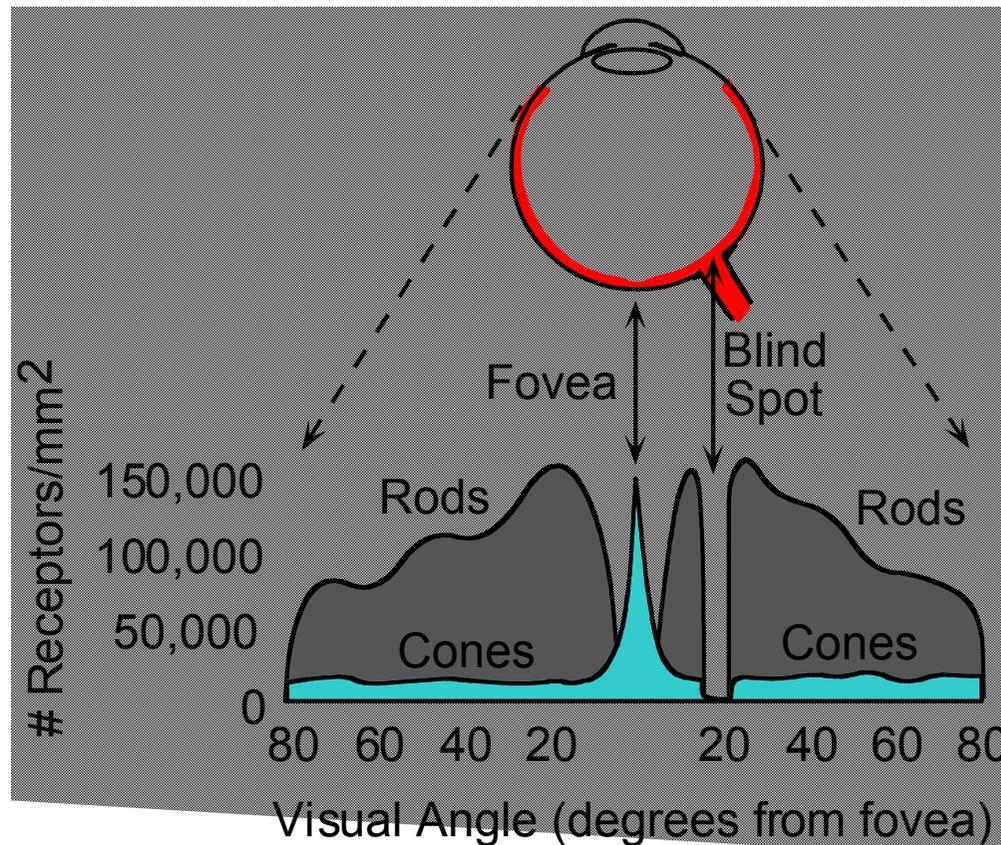
Что мы на самом деле видим



Yarbus, A. L. (1967), *Eye Movements and Vision*, New York: Plenum.



Плотность палочек и колбочек



Палочки и колбочки распределены неравномерно

- Палочки измеряют яркость, колбочки цвет
- **Fovea** («желтое пятно») – маленькая область (1 or 2°) в центре визуального поля с наибольшей плотностью колбочек и без палочек
- На периферии все больше палочек подсоединены к одному нейрону



Движения глаз



Free examination.

1



Estimate material circumstances of the family

2



Give the ages of the people.

3



Surmise what the family had been doing before the arrival of the unexpected visitor.

4



Remember the clothes worn by the people.

5



Remember positions of people and objects in the room.

6



Estimate how long the visitor had been away from the family.

7

3 min. recordings of the same subject



Резюме

- Понятие о компьютерном зрении
- История и задачи компьютерного зрения
- Устройство глаза и фотокамеры
- Трихроматическая теория цвета
- Цветовые модели RGB, CMYK, YIQ, HSV