



Школа Анализа Данных
Яндекса

Курс «Анализ изображений и видео»

Лекция №1
«Введение в предмет. Цифровое
изображение. Свет и цвет»

Антон Конушин

Заведующий лабораторией компьютерной графики и мультимедиа
ВМК МГУ

9 сентября 2016 года

План лекции



- Что такое «компьютерное зрение» и почему это сложно
- История компьютерного зрения
- Цифровое изображение
- Глаз человека и цифровая камера
- Цвет и модели цвета

Что такое «компьютерное зрение» и
почему это сложно?

Что такое компьютерное зрение?



Задача зрения: понять, что находится на изображении

Компьютерное зрение:
построение компьютерной модели системы зрения

Компьютерное зрение –
часть области искусственного интеллекта (AI)

Тест Тьюринга для компьютерного зрения:

Ответить на любой вопрос про изображении, на который может ответить человек.

Что и где находится на изображении?



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

Что за сцена и откуда?



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

**Классификация
изображений**



Общая характеристика изображения

Какой?



Голубое

Ветер
слабый

Наклонная

Мао

Анфас

Профиль

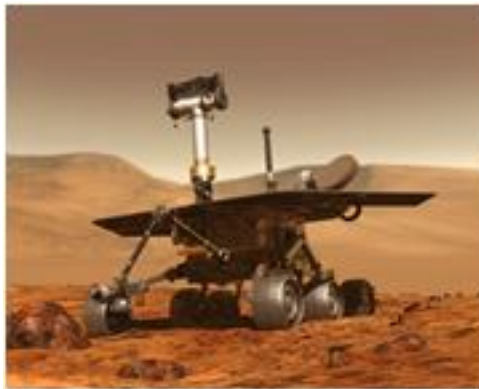


Характеристики (атрибуты) отдельных объектов

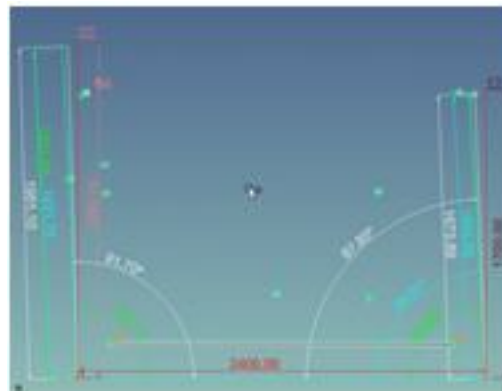
Какой формы?



Стерео-зрение



Измерения
расстояний по
снимкам



3D моделирование по
снимкам



Измерение расстояний и формы объектов

Почему зрение – это сложно?



Попробуйте придумать причины,
затрудняющие анализ изображений!

Точка наблюдения (ракурс)



Michelangelo 1475-1564



slide credit: Fei-Fei, Fergus & Torralba

Освещение



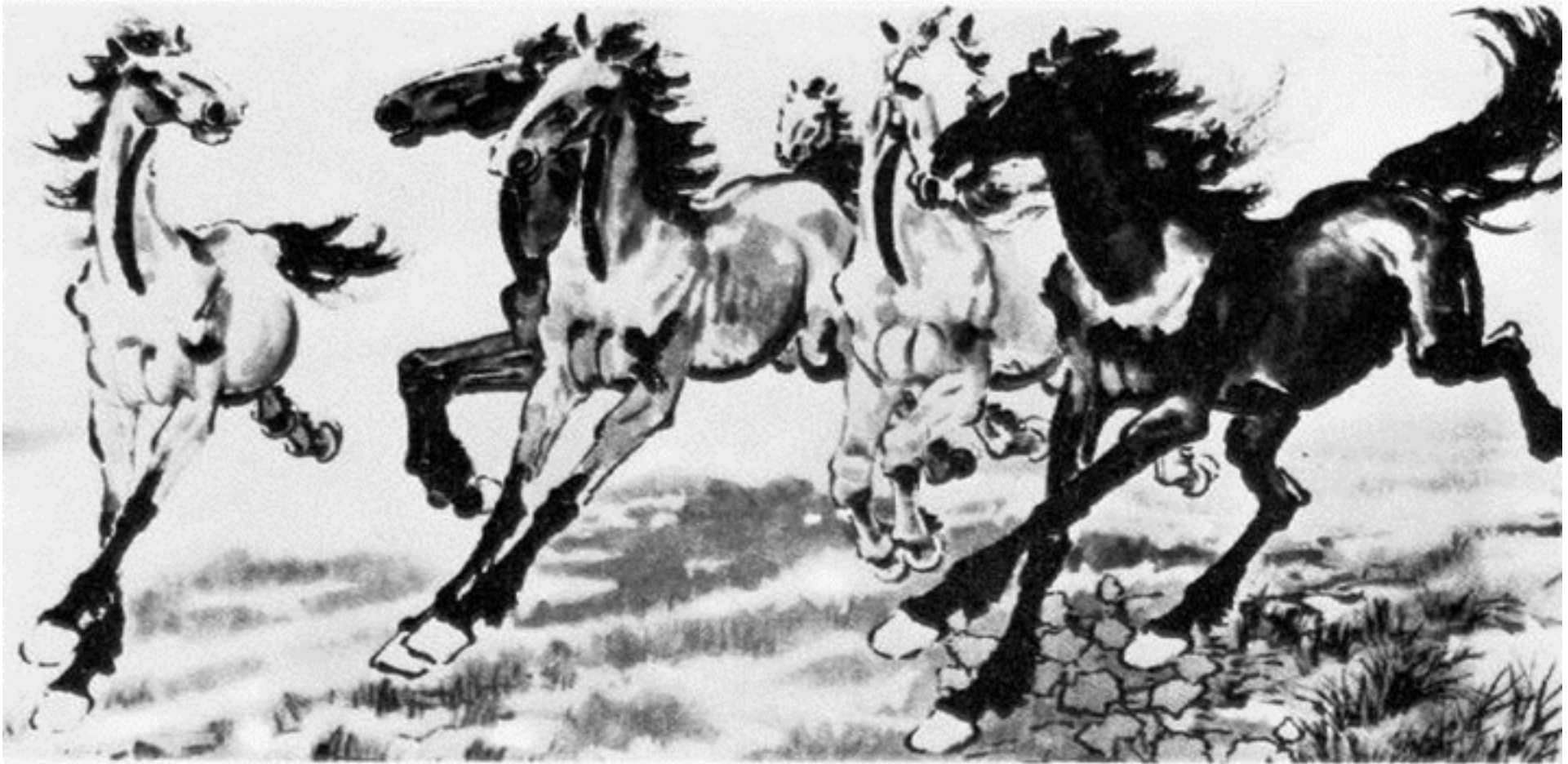
image credit: J. Koenderink

Размер



Slide credit: Fei-Fei, Fergus & Torralba

Деформация

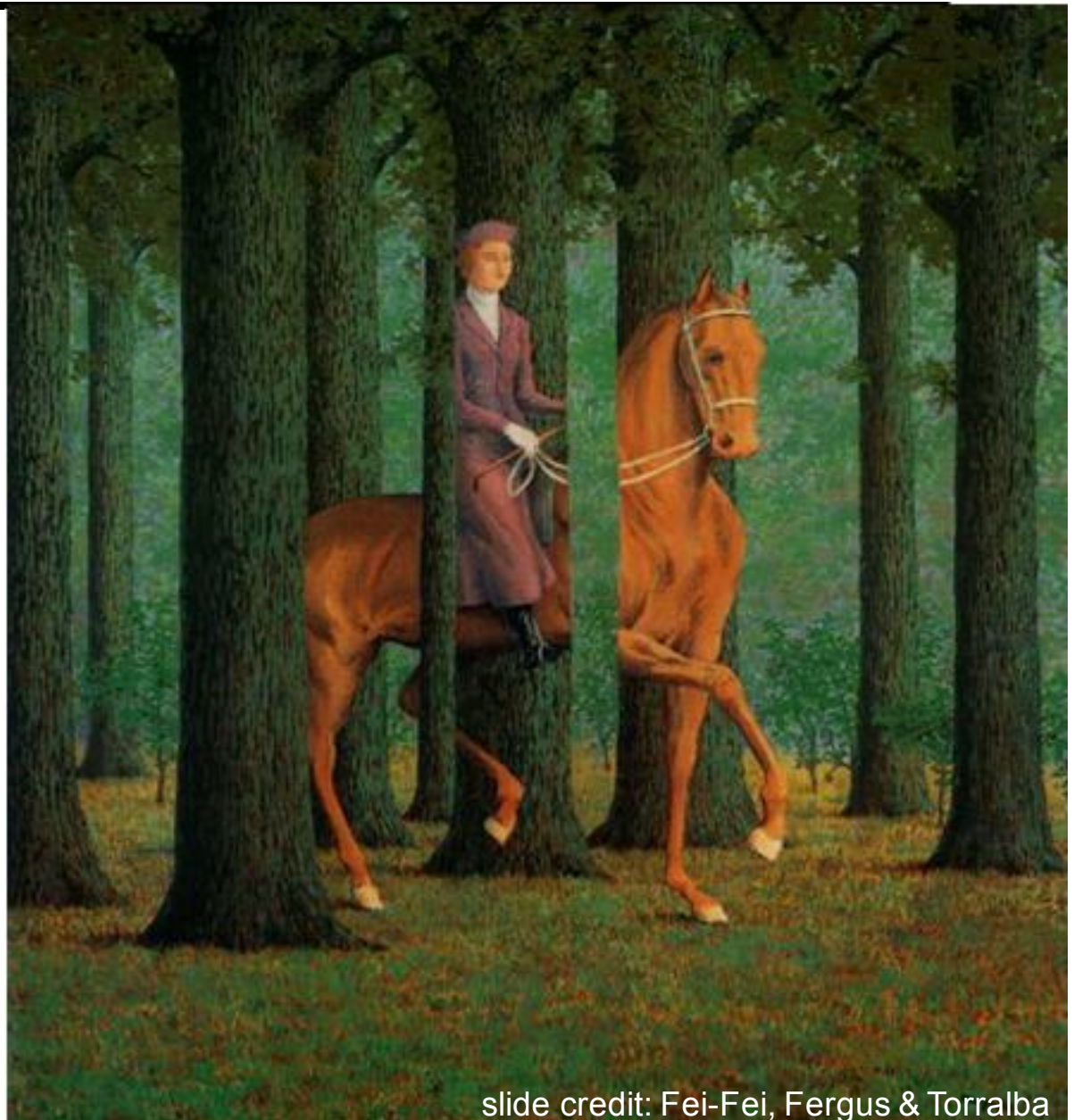


Xu, Beihong 1943

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

Несколько интерпретаций



Одно 2D изображение допускает разные 3D интерпретации

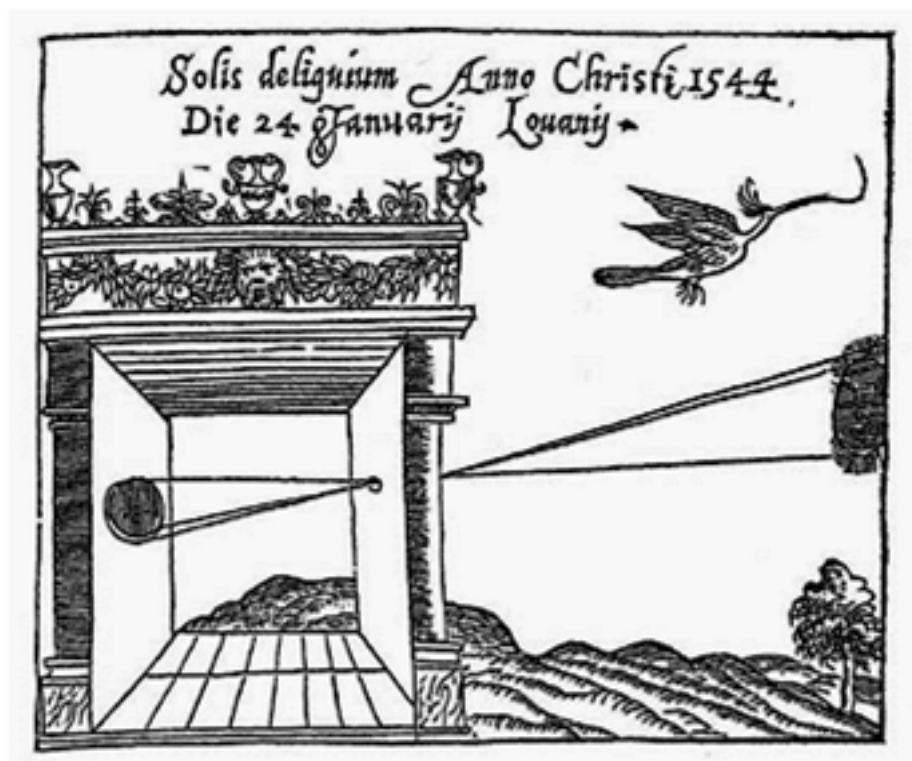
Зрение человека



- Мы сопоставляем наблюдения (подсказки) и априорные знания для интерпретации изображения
- По оценкам, 25% мозга занято решением задачи зрения
- Какие подсказки и априорные знания вы можете придумать/вспомнить?

История компьютерного зрения

Камера-обскура



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

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.

“Магический фонарь” (1492)



Figure 1. Leonardo da Vinci.

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.

Стереокамера и
теодолит

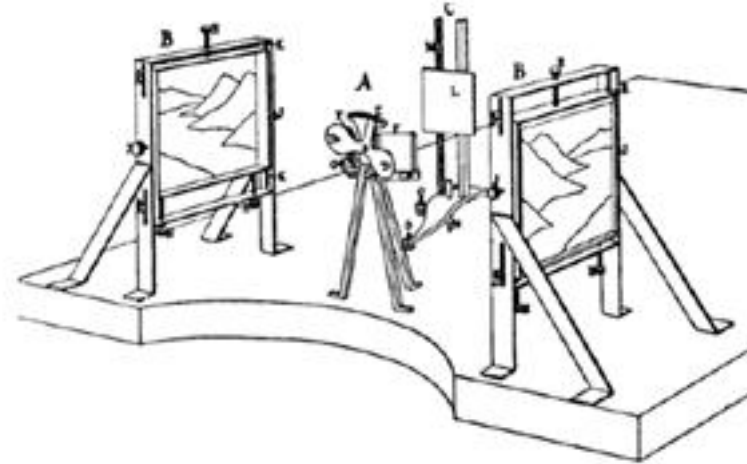


Figure 18. Deville's Stereo Planigraph.

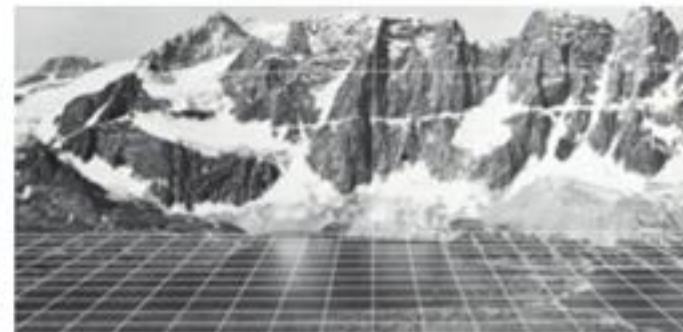


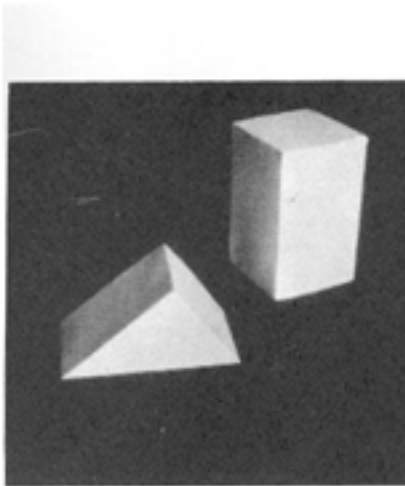
Figure 19. Illustration of the Canadian Grid Method used by Deville.

Whirlwind, MIT (1951)

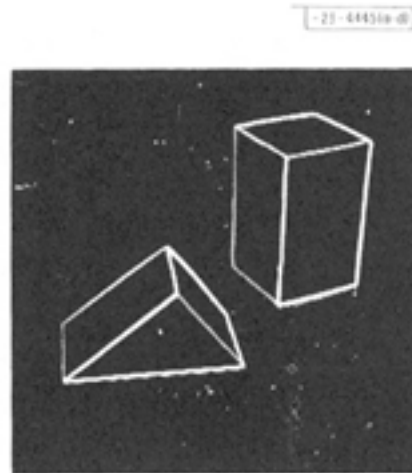


- Первый компьютер, отображающий текст и графику в реальном времени на мониторе
- Точками карту, значком самолёт.
- «Световое перо» для взаимодействия с экраном (запрос информации об объекте)

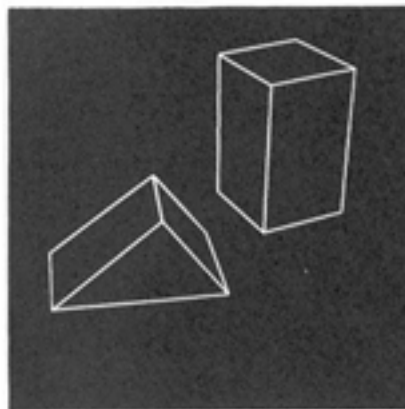
Зарождение компьютерного зрения (1960)



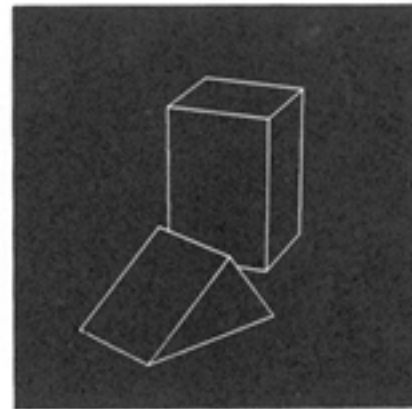
(a) Original picture.



(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



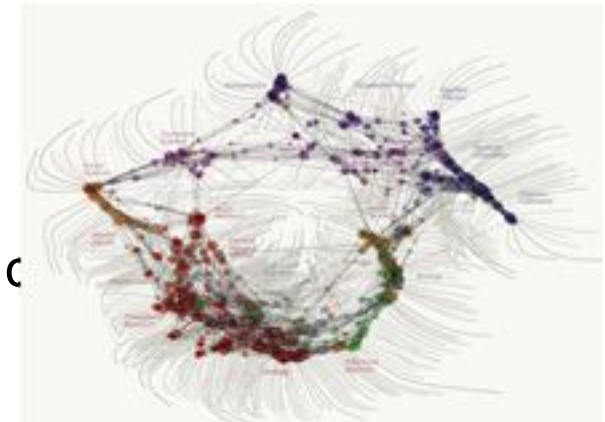
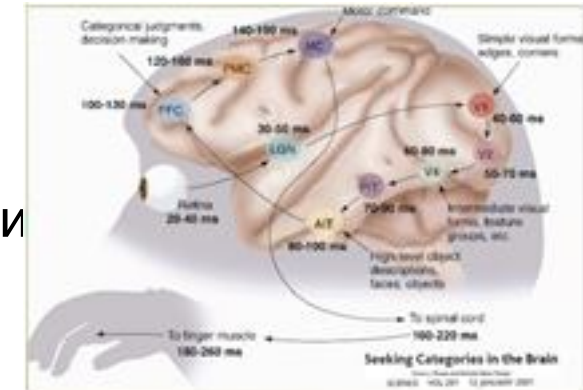
GIRL TRACED FROM PHOTOGRAPH

Ivan Sutherland демонстрирует интерактивный графический редактор SketchPad

Давид Мэпп (1970е)



- «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



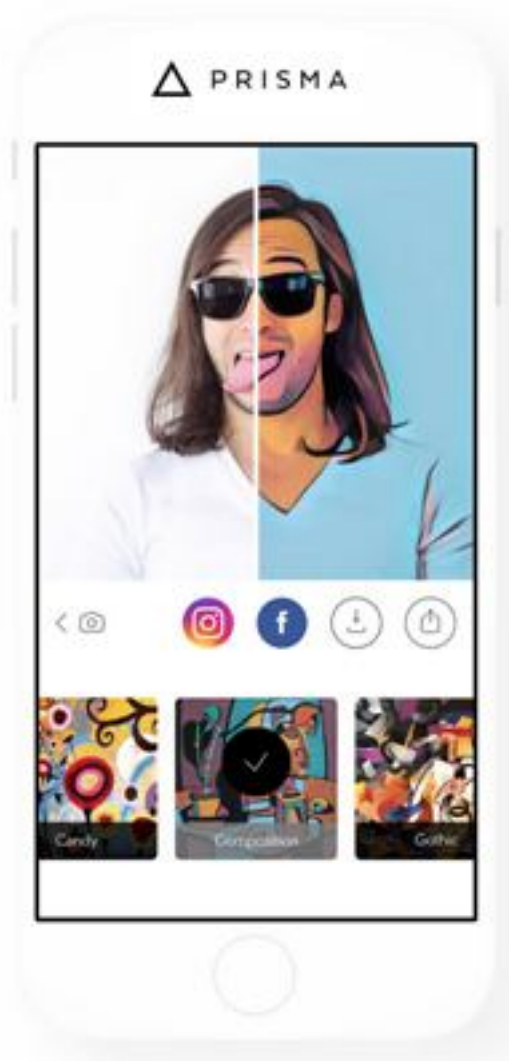
“The Marr Prize” – главная премия в области компьютерного зрения

Детектор лиц Viola-Jones (2001)



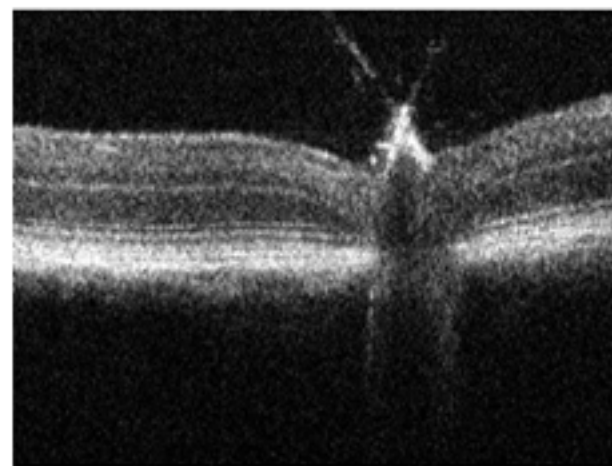
Алгоритм Viola-Jones – первый быстрый и надежный алгоритм поиска лиц. Демонстрация силы машинного обучения.

Глубинное обучение (Deep learning)



Цифровое изображение

Изображение

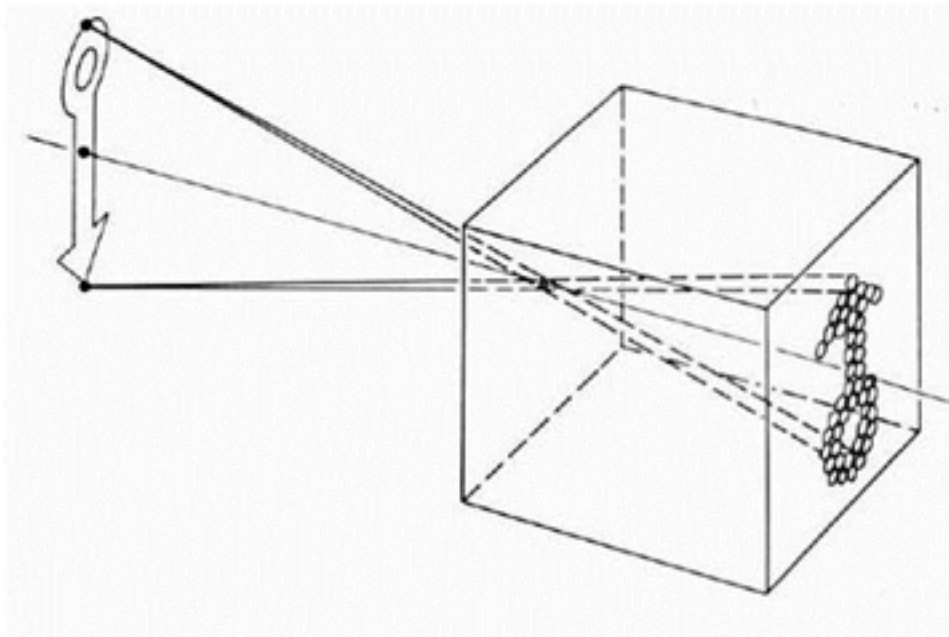


Изображение оптическое – картина, получаемая в результате прохождения через оптическую систему лучей, распространяющихся от объекта, и воспроизводящая его контуры и детали.

(Физический энциклопедический словарь.)

Оптические системы могут быть очень разными!

Камера-обскура



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

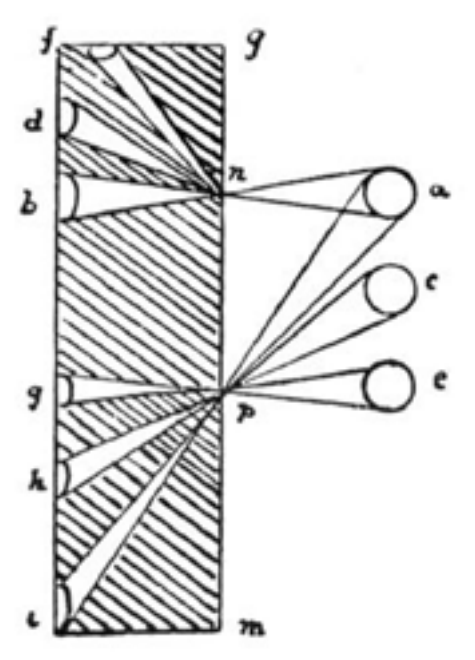
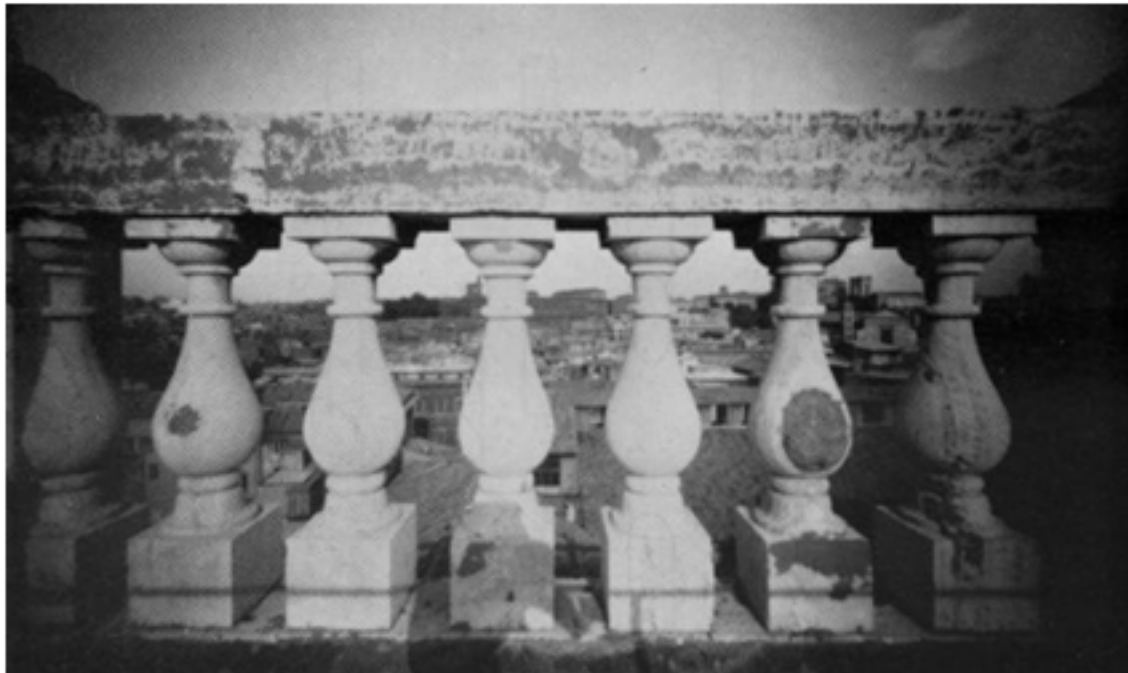
Математическая модель – перспективная проекция:

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

Перспективные искажения



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



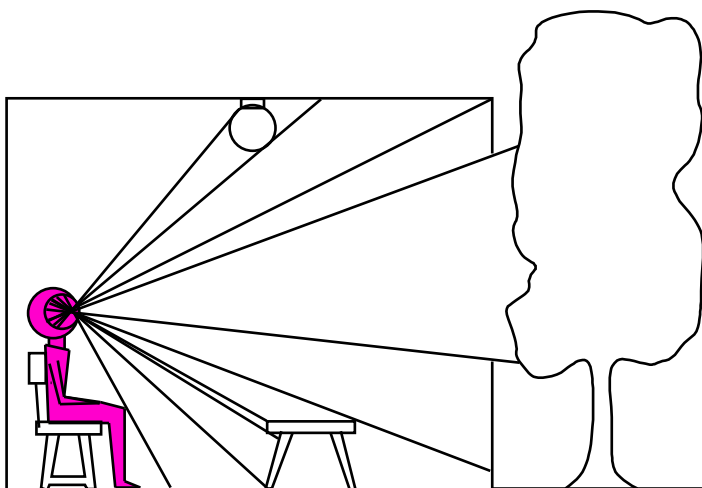
Ещё один пример



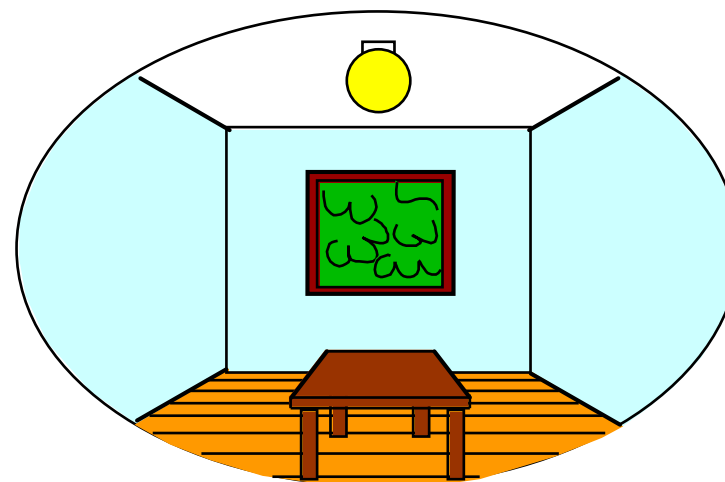
Машина Понижения Размерности



3D мир



2D картина

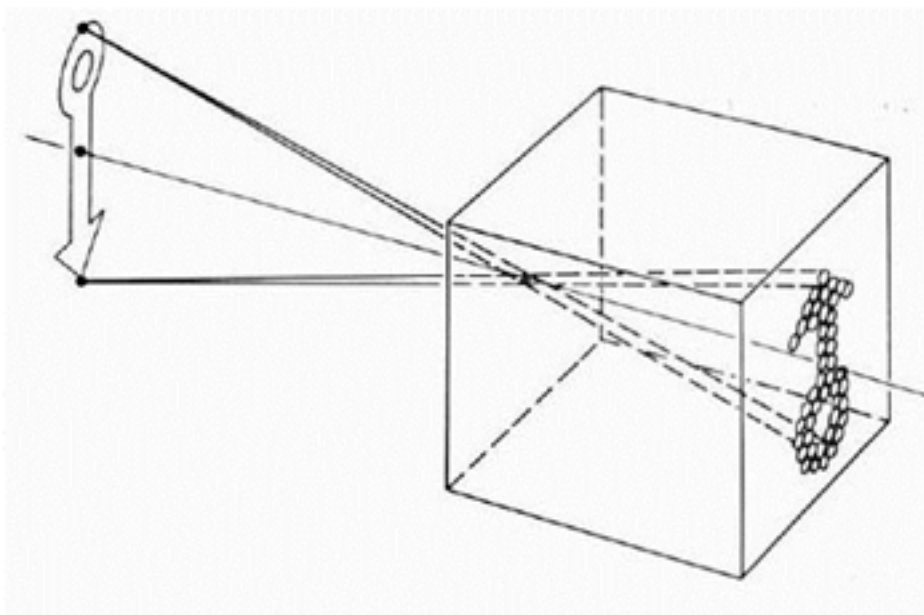


Что мы теряем?

- Углы
- Расстояния и длины

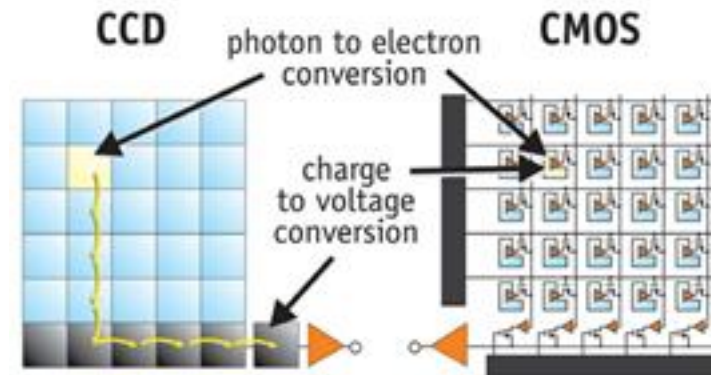
Подробнее геометрическая
модель рассматривается
позже

Современная камера

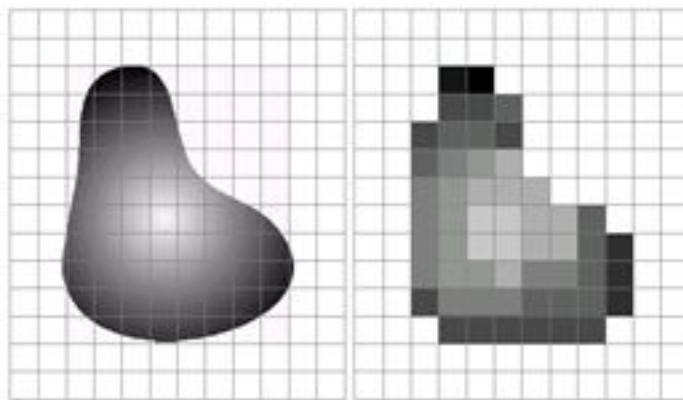


Тот же самый принцип, но с объективом и цифровой матрицей

Цифровая камера - дискретизация



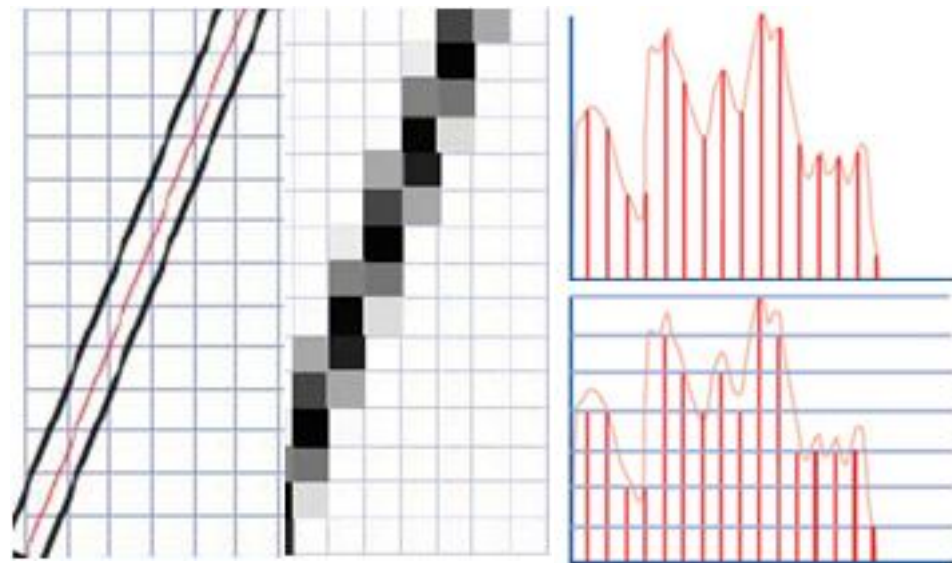
CCDs move photogenerated charge from pixel to pixel and convert it to voltage at an output node. CMOS imagers convert charge to voltage inside each pixel.



a b

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

По пространству



По яркости

Первое цифровое фото (1957)



Разрешение
176*176 пикселей

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

Цифровое изображение



- Цифровое полутоновое изображение – матрица $I \in (b_{ij})^{n,m}$, элементами которой b_{ij} являются значения яркости света, измеренного на 2х мерной прямоугольной сетке
 - Про то, что такое «яркость» мы ещё отдельно поговорим
 - b (яркость) можно описать вещественным числом. Обычно ограничиваются интервалом $b_{ij} \in [0,1]$, где 0 (нет света), 1 (максимальная яркость)
 - Обычно b кодируем 1 байтом, т.е. $b \in [0,255]$
 - Можем использовать большую точность (10-16бит)

Свет и 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*)

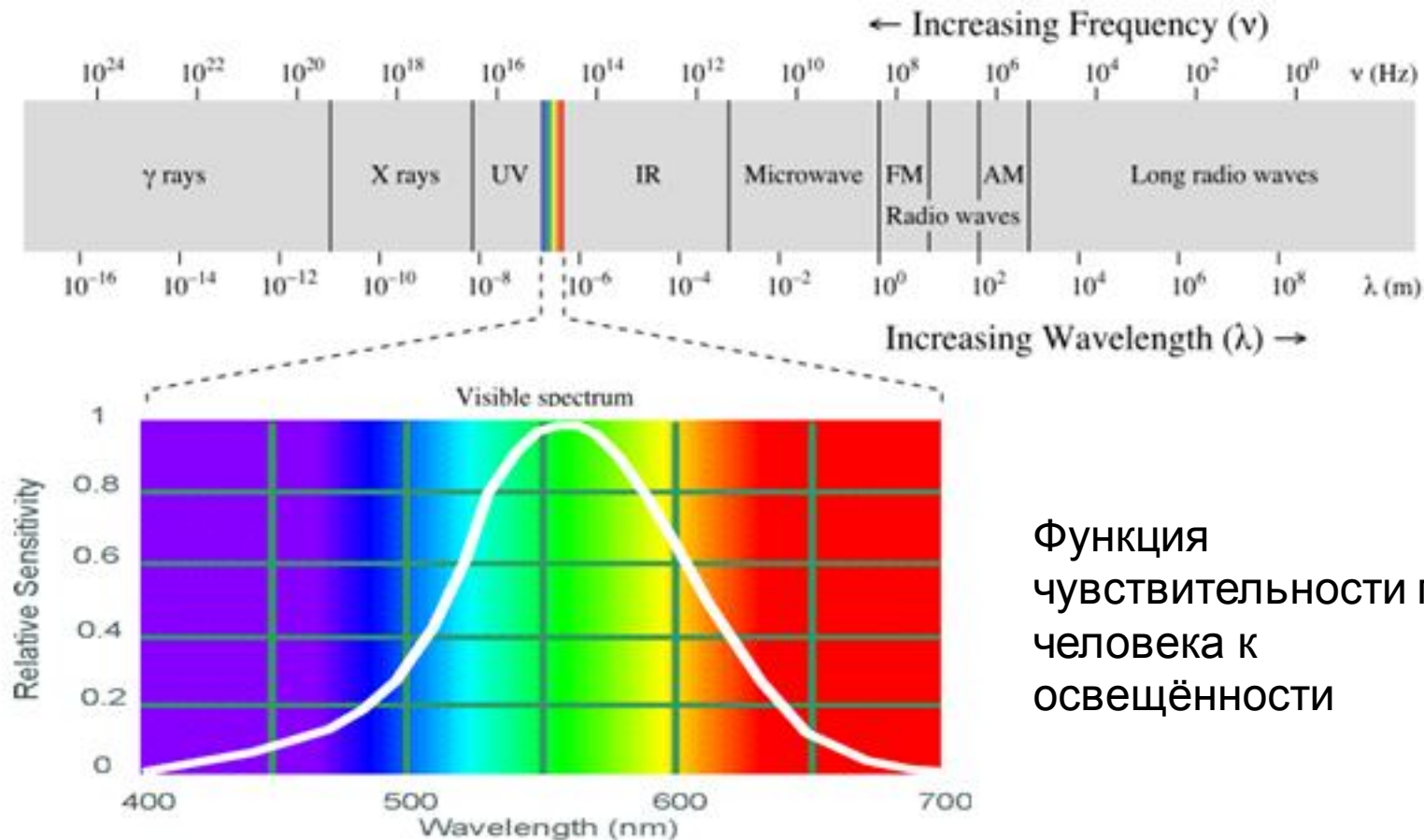
Цвет – это результат взаимодействия света, сцены и нашей зрительной системы



Восприятие света человеком изучают науки фотометрия и колориметрия

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

Видимый свет



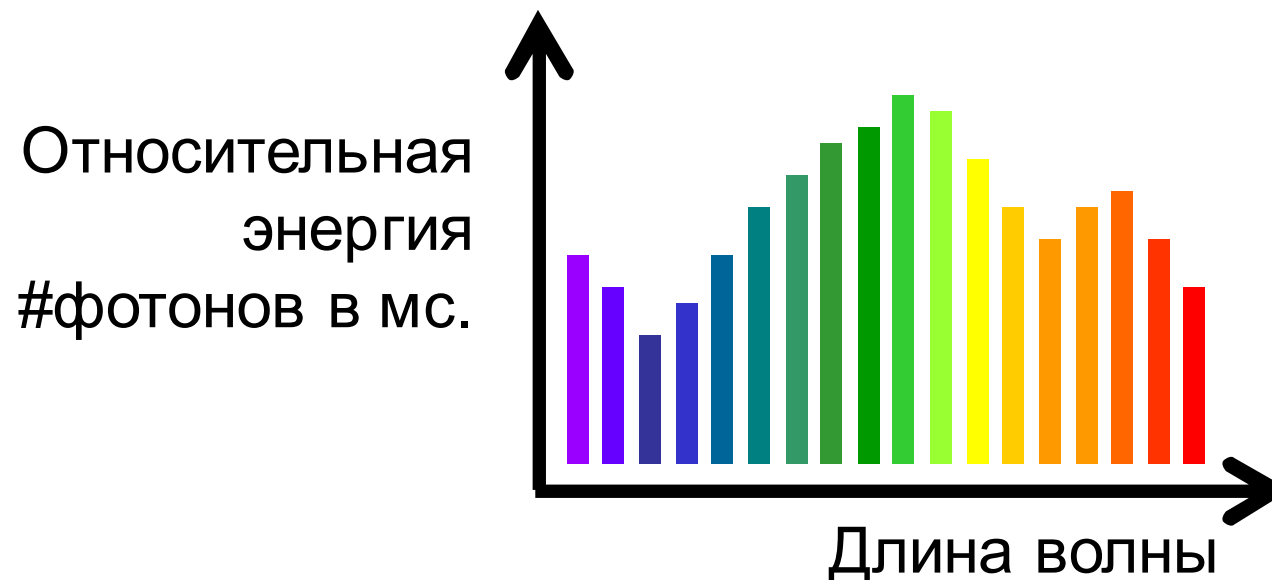
Функция
чувствительности глаза
человека к
освещённости

- Электромагнитное излучение в диапазоне [380nm,780nm]
- Видимый свет попадает в основное «оптическое окно» земной атмосферы (~46% энергии)

Физика света



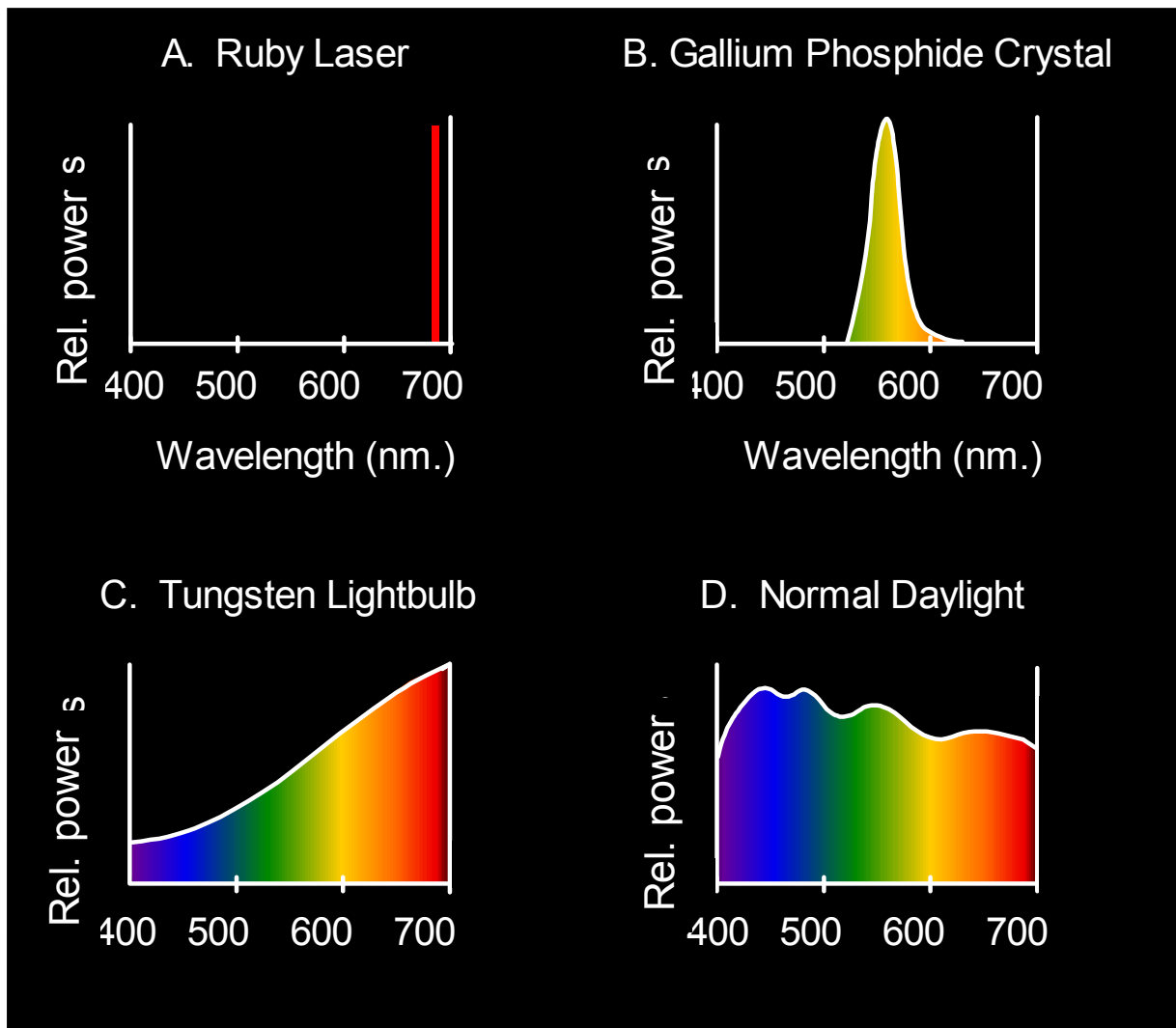
Любой источник света можно полностью описать спектром: количество излученной энергии в единицу времени для каждой длины волны в интервале 380 - 780 nm.



Физика света



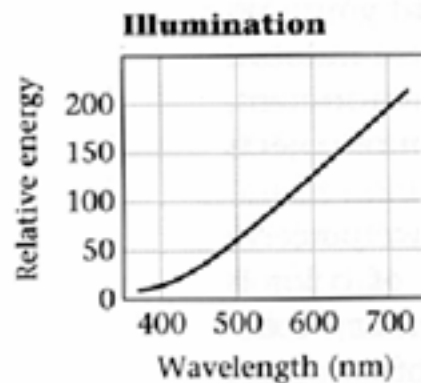
Примеры спектров разных источников света



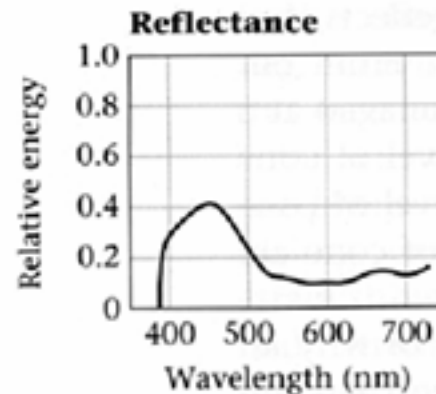
Взаимодействие света и объектов



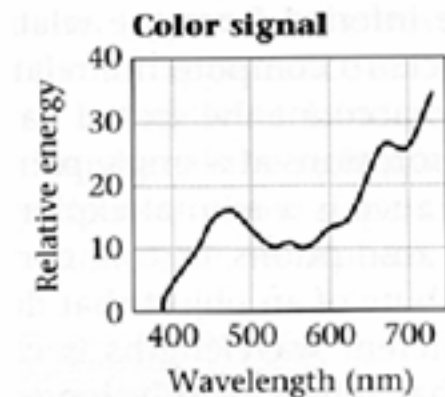
Отраженный свет это
результат взаимодействия
излучаемого света и
поверхности



• *



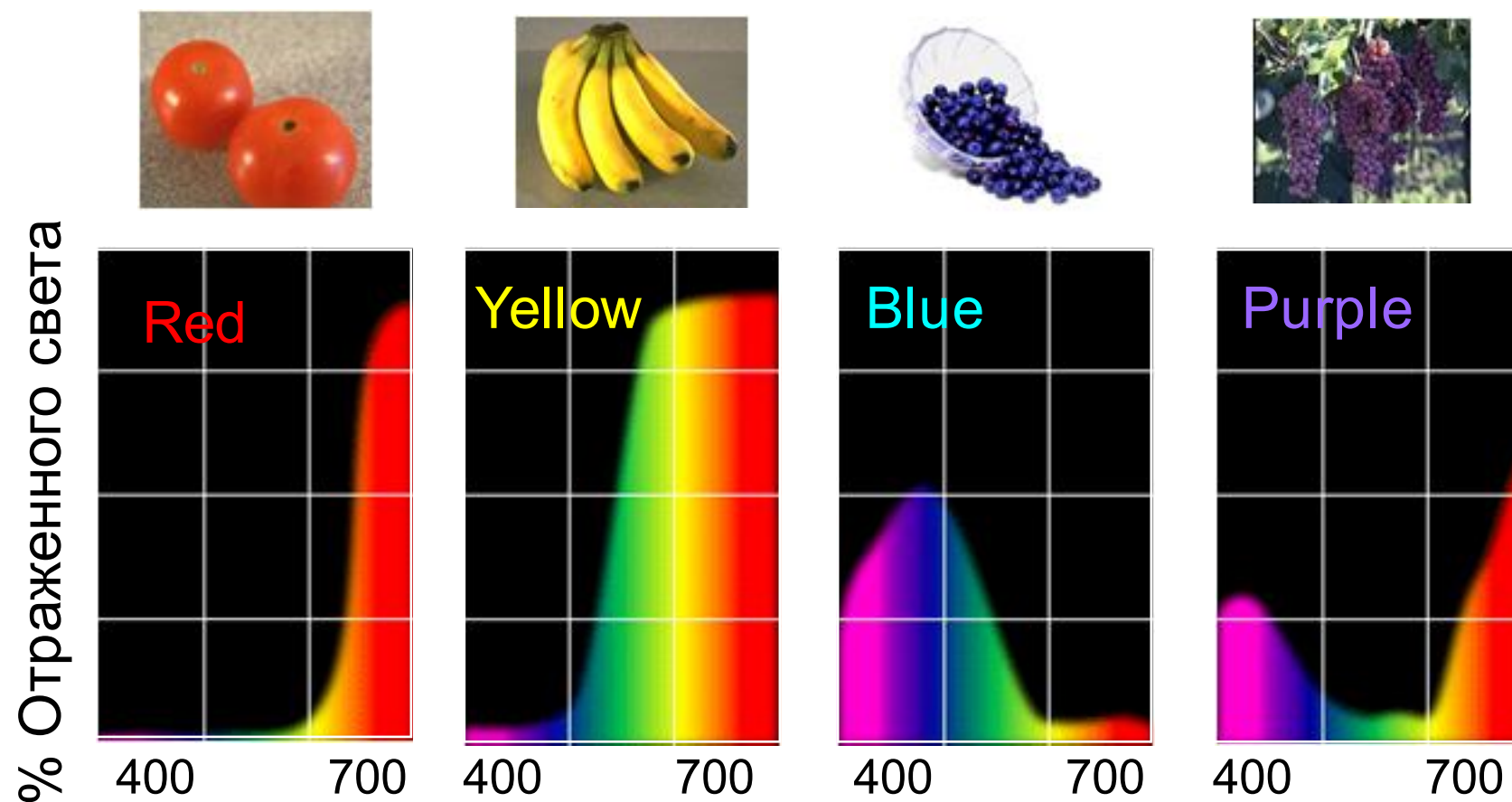
=



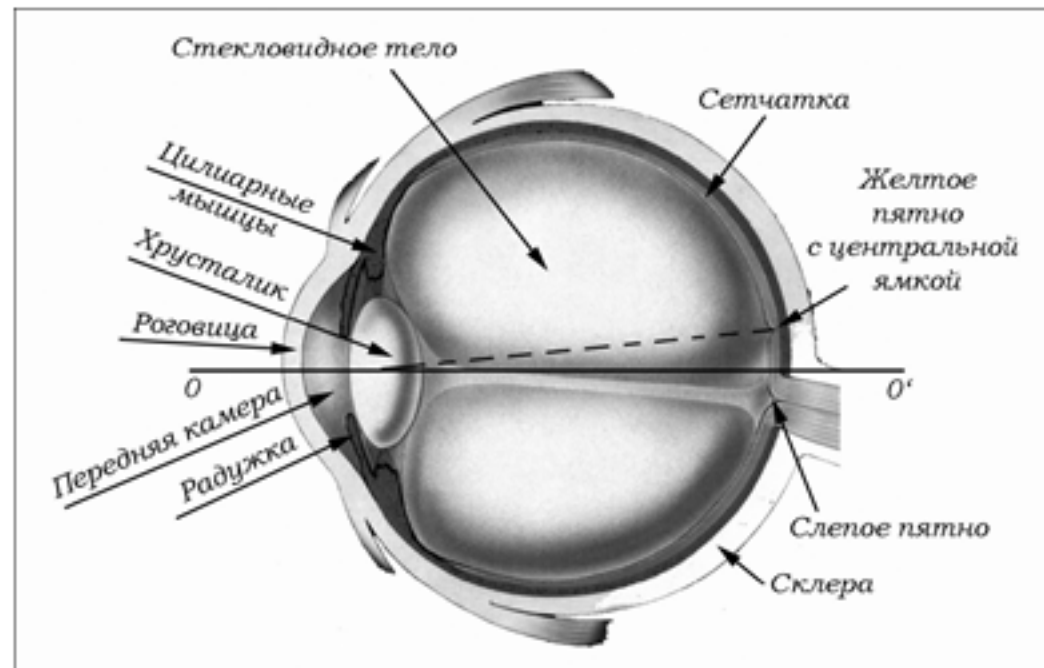
Физика света



Примеры спектров отраженного света от предметов



Человеческий глаз



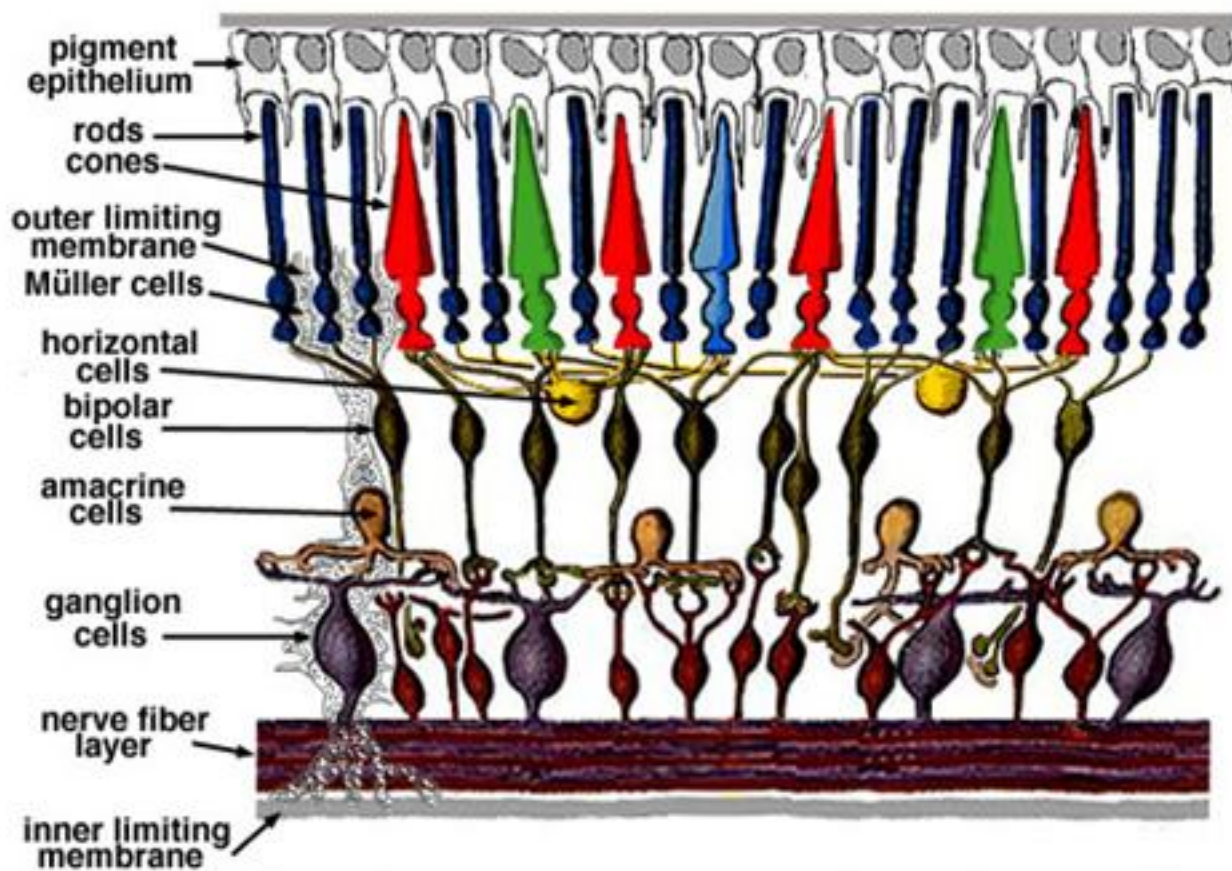
Глаз как камера!

- **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

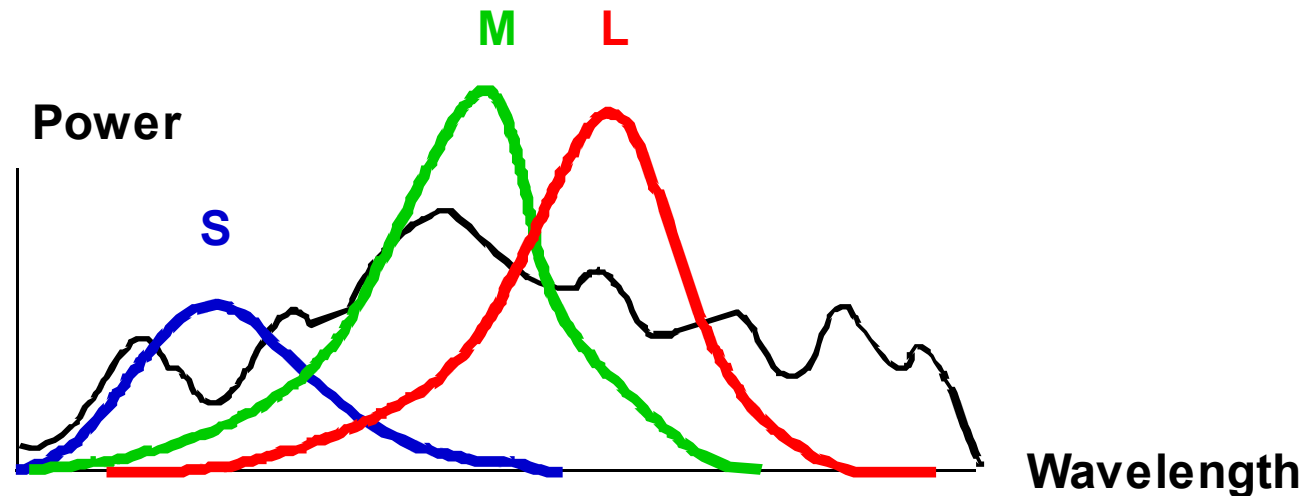


Свет ↓ ↓ ↓



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

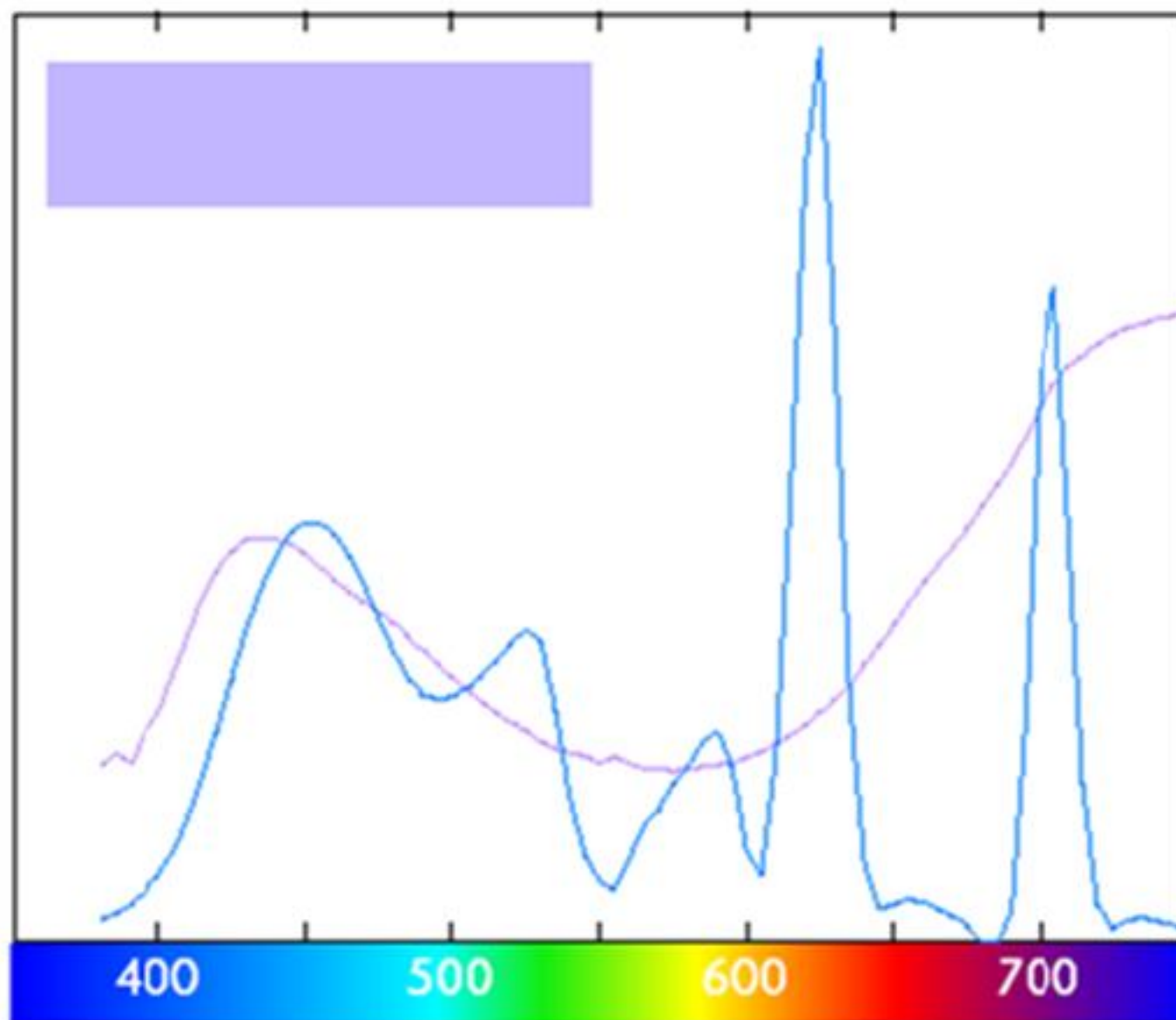
Восприятие цвета



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**

Метамеры фиолетового



Цветовые модели

Цифровое цветное изображение



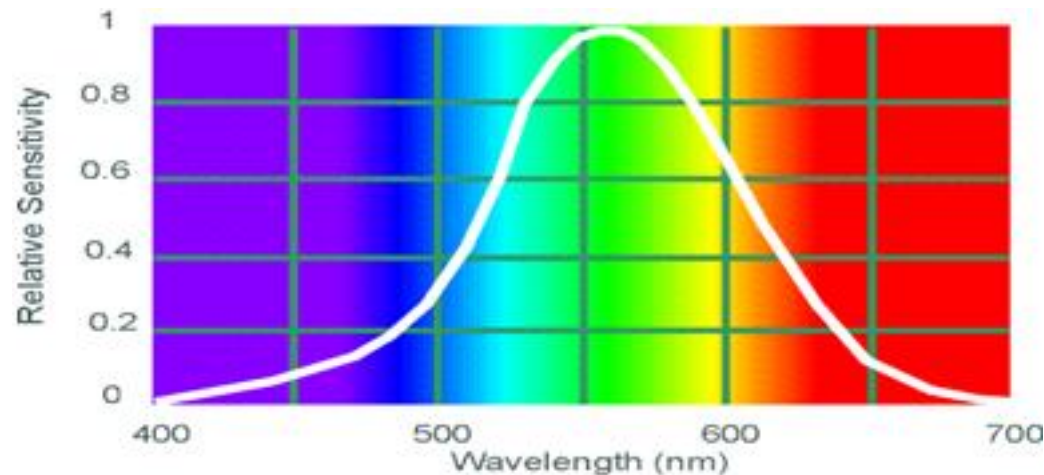
- Цифровое цветное изображение $X \in R^{n \times m \times c}$ – тензор ранга 3, в котором для каждого пиксела (x,y) хранится цвет.
- n, m – высота и ширина изображения
- c – число каналов (длина вектора, описывающего цвет пиксела)
- Обычно каждая компонента x_{ijk} дискретизируется до $[0, 255]$ в 8 бит
- Какое наиболее точное описание света, падающего на пиксел?

Яркость и цветность



Интуитивно можно выделить характеристики света:

- Яркость (brightness)
- Цветность (chromaticity)



Функция чувствительности
глаза человека к свету
(воспринимаемая яркость)

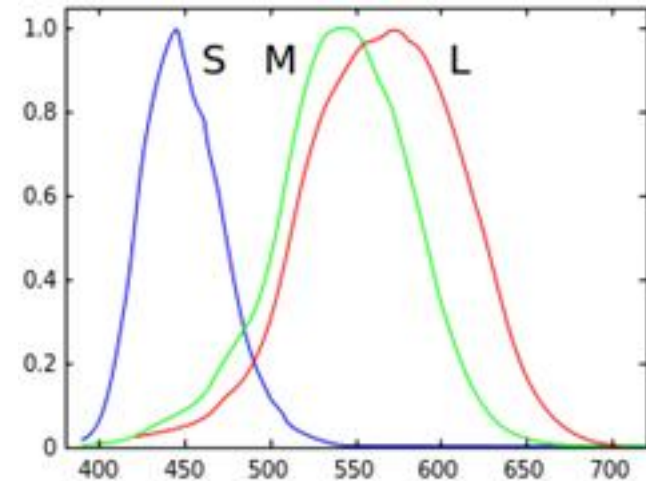
В цветности можно выделить тон (hue) и насыщенность (saturation)

Модель LMS



- Три значения, описывающие вклад (возбуждение), каждого из трёх типов фоторецепторов

- L(long)
- M(medium)
- S(short)

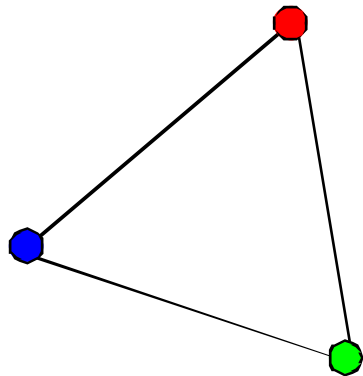


- Некоторые комбинации значений LMS невозможны для реального света
 - $M > 0, S = L = 0$
- Невозможно на практике построить дисплей, работающий в LMS

Трихроматическая теория и линейные модели



- Выбираем базовые цвета P1, P2, (P3)
- Остальные цвета задаются линейной комбинацией базовых цветов (primaries). Веса – «координаты» цвета



$$A = u_1 P_1 + u_2 P_2 + u_3 P_3$$

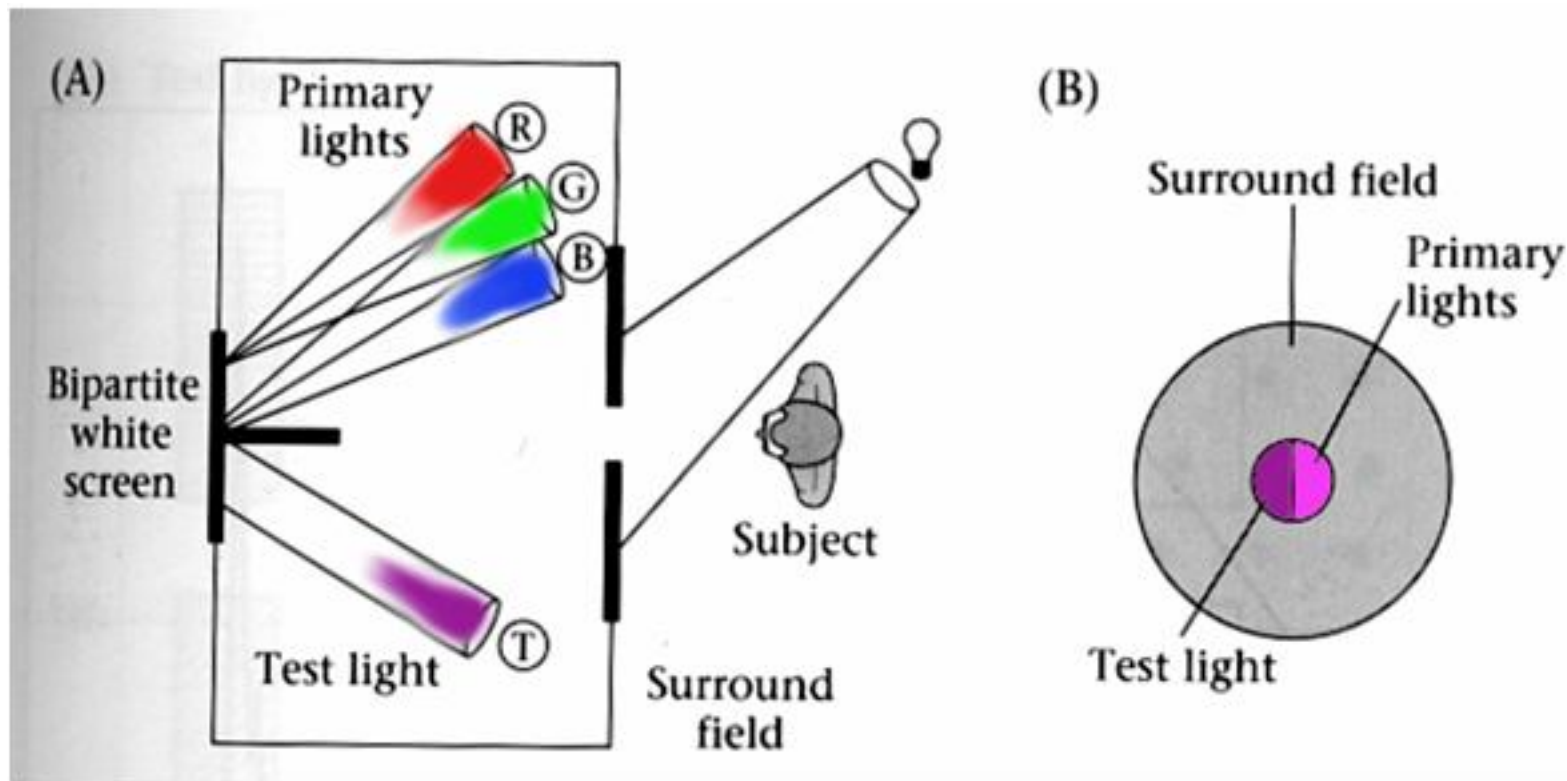
Смешение трех цветов

- Трихроматическая теория гласит, что все видимые цвета можем получить комбинацией трёх базовых цветов
- Можем делать эксперименты по сопоставлению цветов

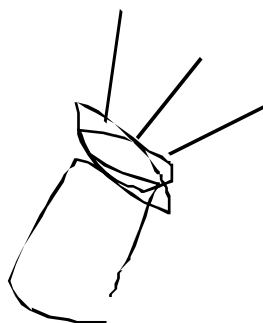
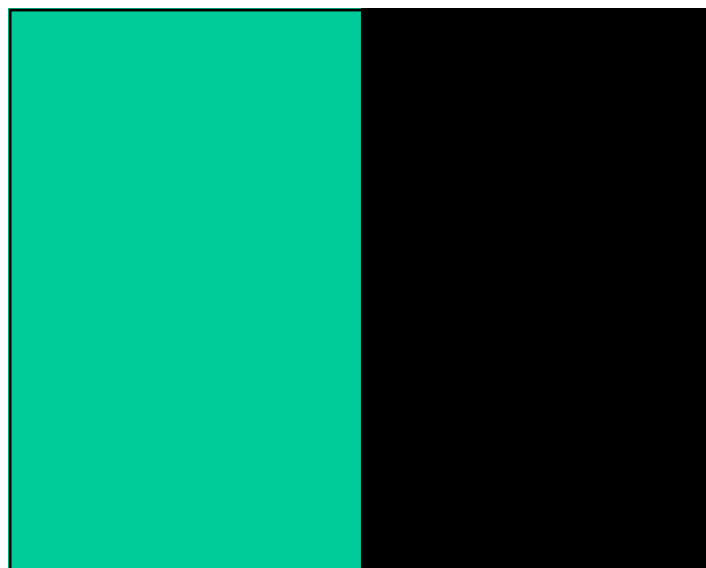
Сопоставление цветов



- Берём «тестовый» источник и пробуем сопоставить его комбинацией базовых цветов (источников света)

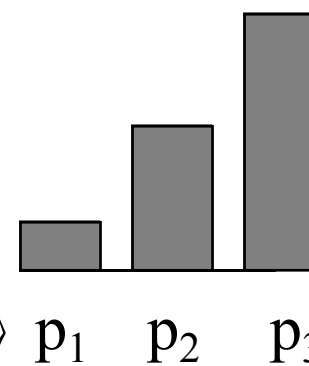
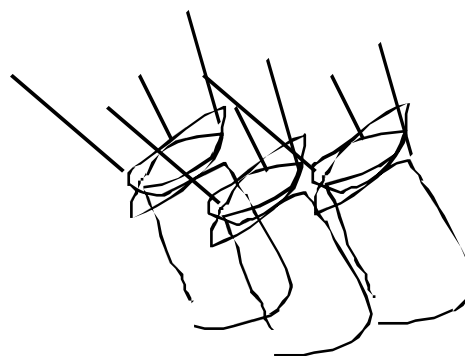
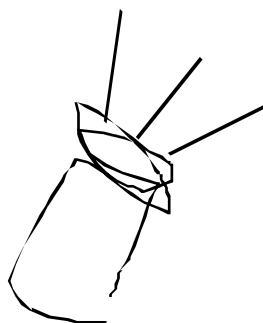
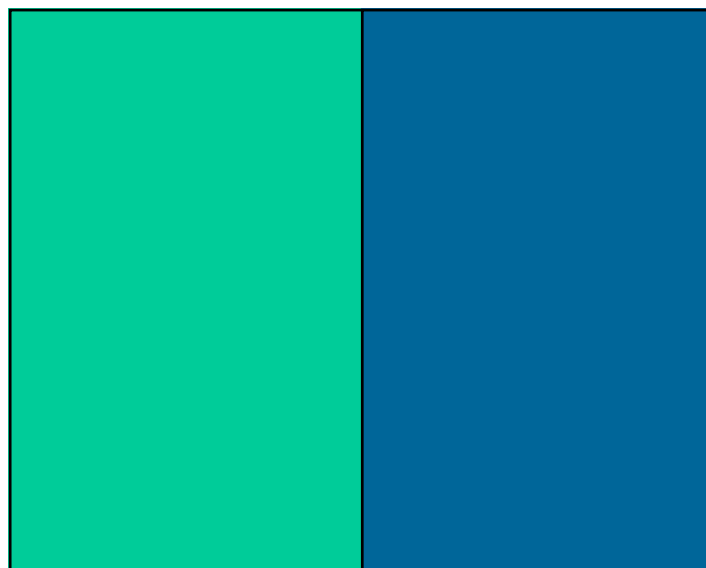


Эксперимент №1



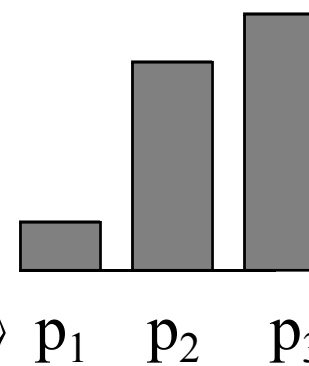
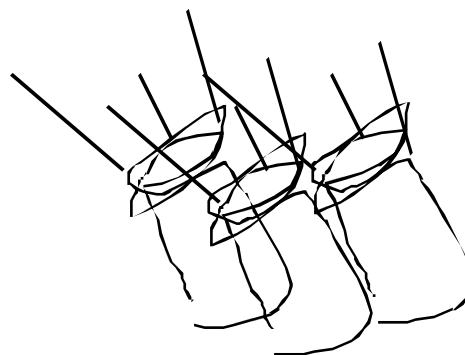
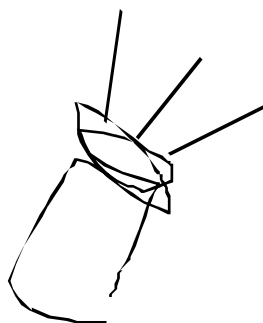
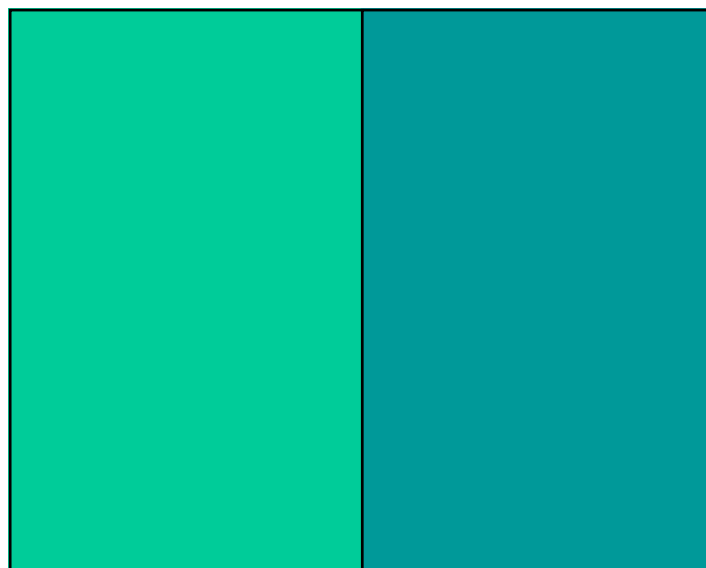
Source: W. Freeman

Эксперимент №1



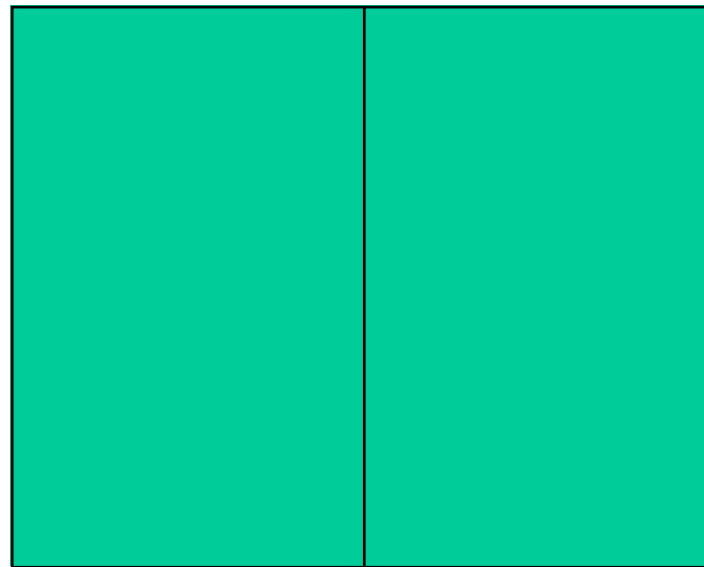
Source: W. Freeman

Эксперимент №1

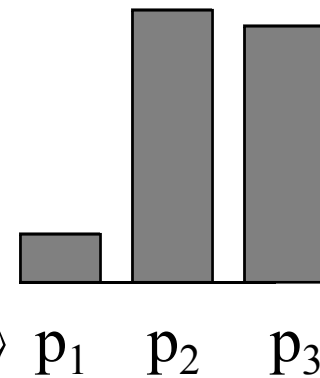
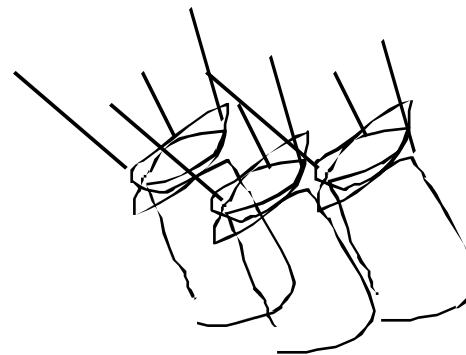
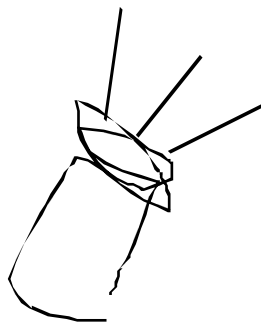


Source: W. Freeman

Эксперимент №1



«Координаты»
света в нашей
модели



Source: W. Freeman

Закон аддитивности Грассмана (1853)



Эмпирический закон о линейности человеческого зрения.

Верно:


1. Если наблюдатель задаст цвет лучей 1 и 2 как (R_1, G_1, B_1) и (R_2, G_2, B_2) , тогда если мы сложим источники 1 и 2, то мы их можем воспроизвести как $(R_1+R_2, G_1+G_2, B_1+B_2)$
2. Соответствие цветов выполняется на всех уровнях яркости. Т.е. если $C_1 = (R_1, G_1, B_1)$, то $kC_1 = (kR_1, kG_1, kB_1)$


Модели цвета, основанные на выборе базовых цветов и их линейных комбинациях – линейные цветовые модели

Модель CIE RGB 1931



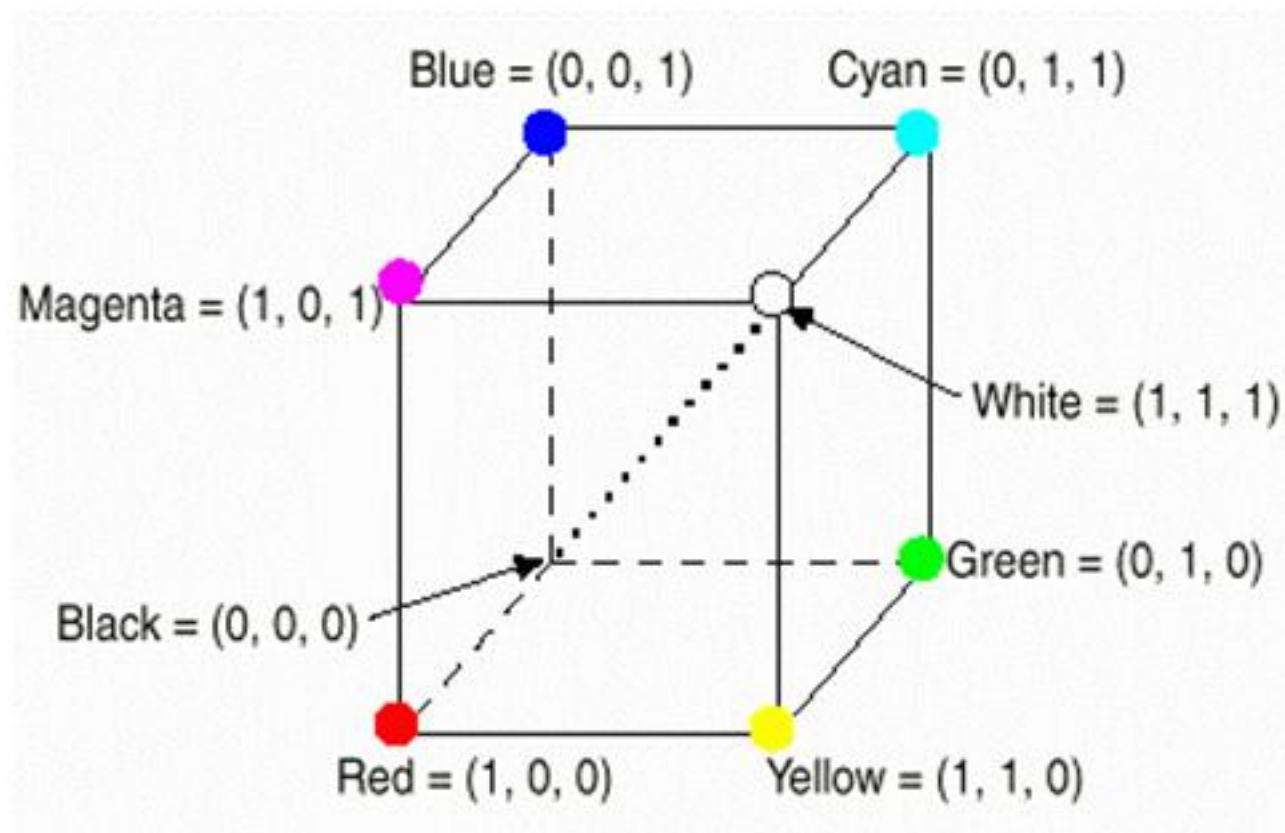
- Выберем три монохроматических цвета в качестве основных на основании информации о чувствительности колбочек

 $p_1 = 645.2 \text{ nm}$

 $p_2 = 525.3 \text{ nm}$

 $p_3 = 444.4 \text{ nm}$




Цветовой куб RGB



Множество цветов, задаваемых цветовой моделью
RGB

Модель RGB

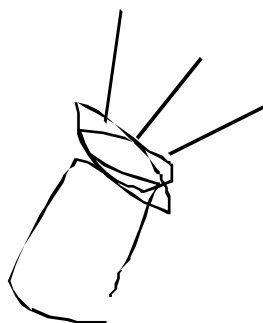


 $p_1 = 645.2 \text{ nm}$
 $p_2 = 525.3 \text{ nm}$
 $p_3 = 444.4 \text{ nm}$



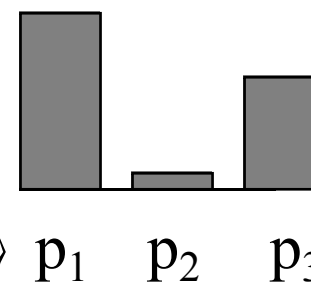
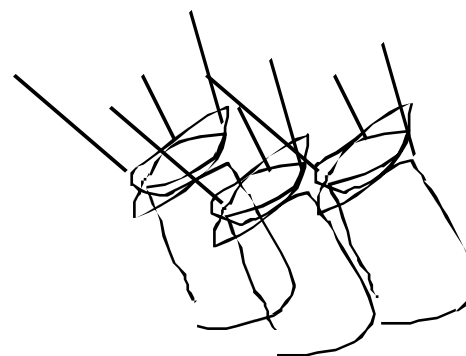
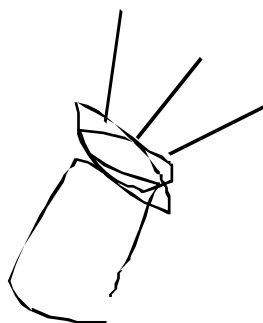
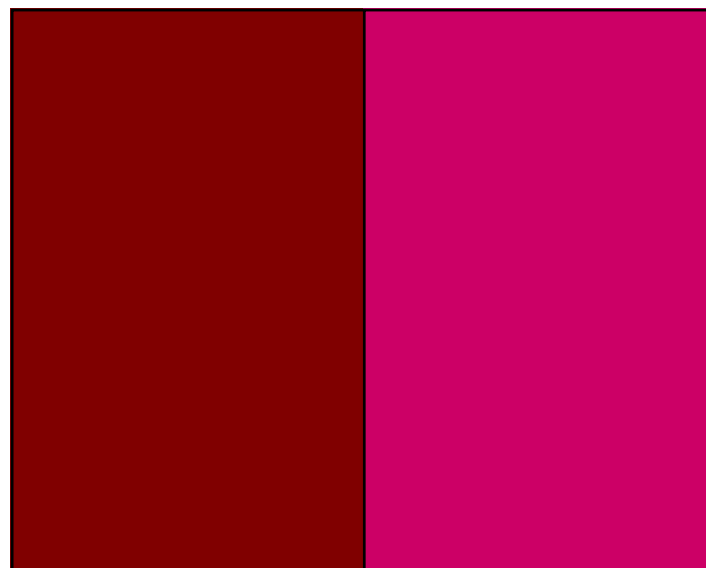
- Описывает ли модель RGB всё множество видимых человеком цветов?
- Для проверки проведём эксперимент и попробуем сопоставить каждой длине волны видимого света (когерентному источнику света) цвет в модели RGB
- *Функции сопоставления* - веса, необходимые для сопоставления с когерентными источниками света
- Будем усреднять результаты участников эксперимента

Эксперимент №2



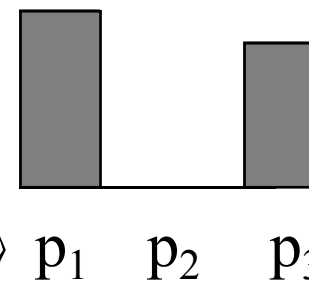
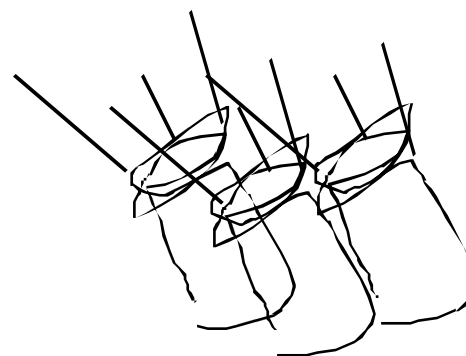
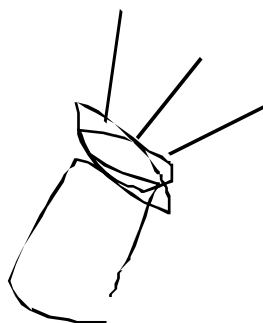
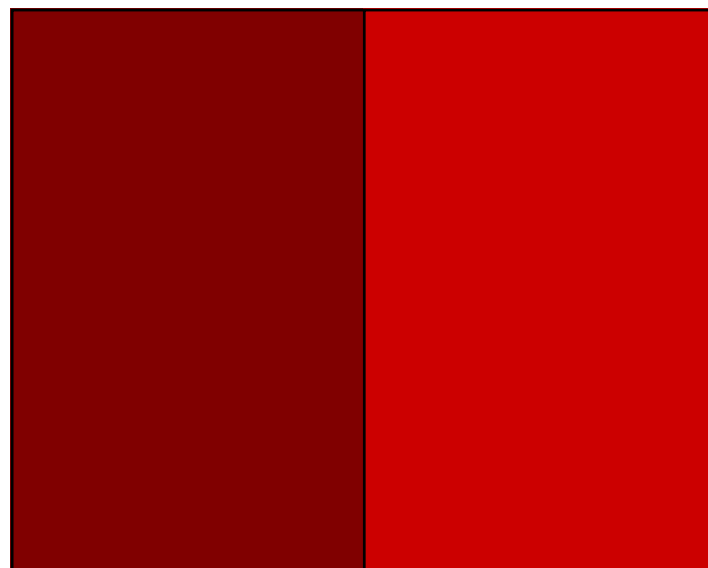
Source: W. Freeman

Эксперимент №2



Source: W. Freeman

Эксперимент №2

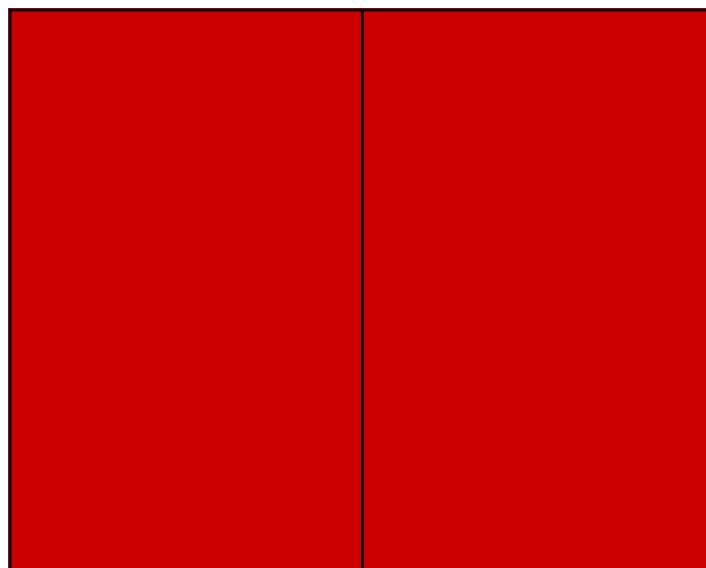


Source: W. Freeman

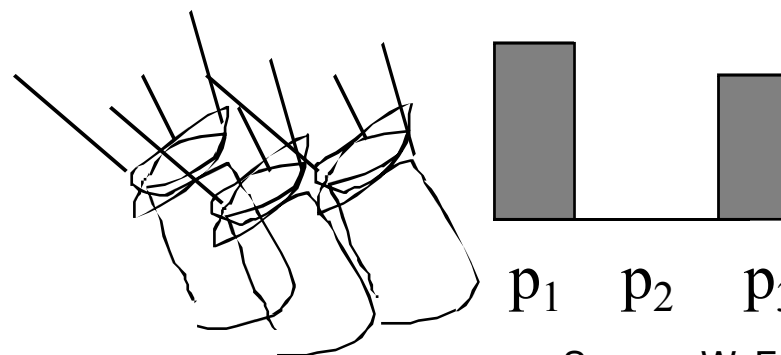
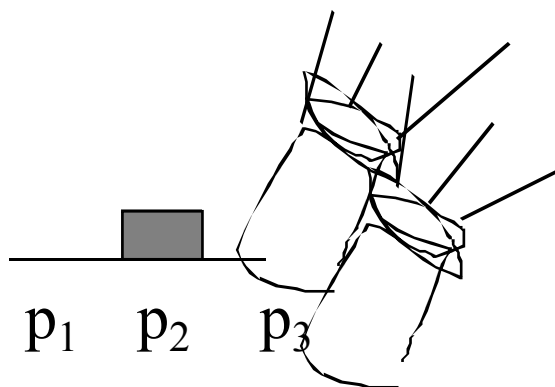
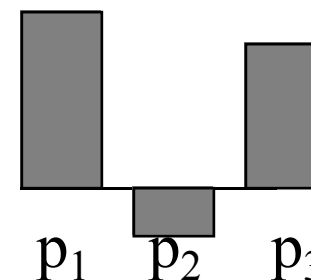
Эксперимент №2



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






Веса основных цветов, необходимых для сопоставления:



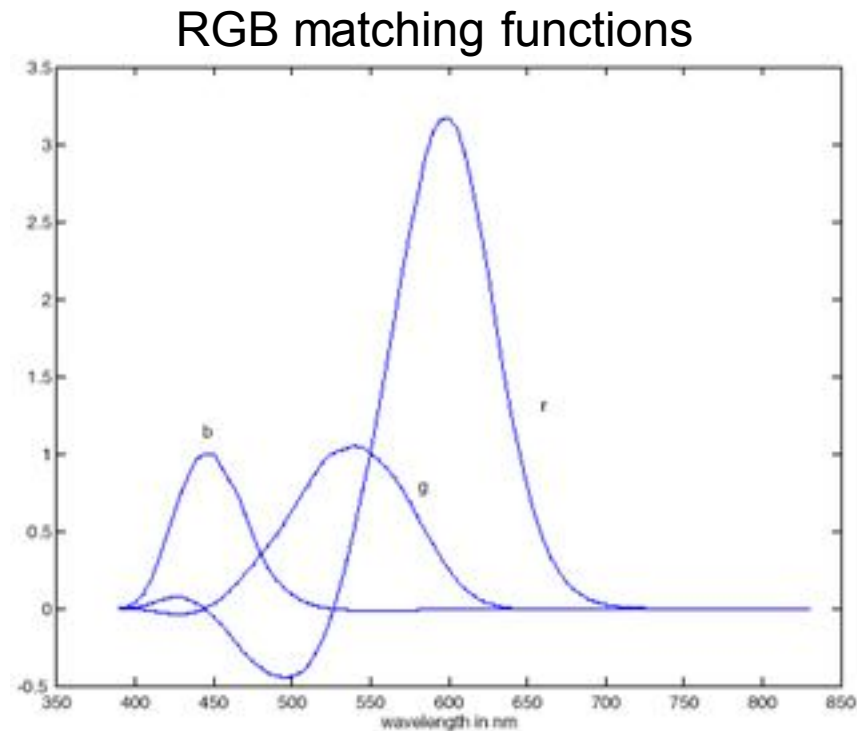
Source: W. Freeman

Линейная цветовая модель RGB



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

Модель RGB не
полностью описывает
всё множество видимых
цветов

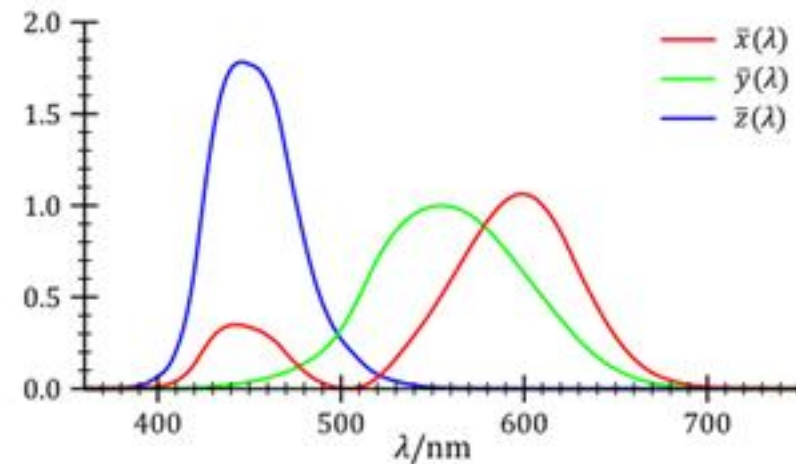


«Вычитание» необходимо
для соответствия некоторым
длинам волны

CIE 1931 (XYZ)



- Хотим получить линейную аддитивную модель XYZ:
 - покрывает всё множество видимых цветов
 - Y должен соответствовать видимой яркости света
 - Функции сопоставления должны быть неотрицательными
- X,Z тогда будут описывать «хроматическую» (цветовую) компоненту
- Точки (1,0,0), (0,1,0), (0,0,1) будут мнимыми базовыми цветами
- X,Y,Z изменяются от 0 до ∞

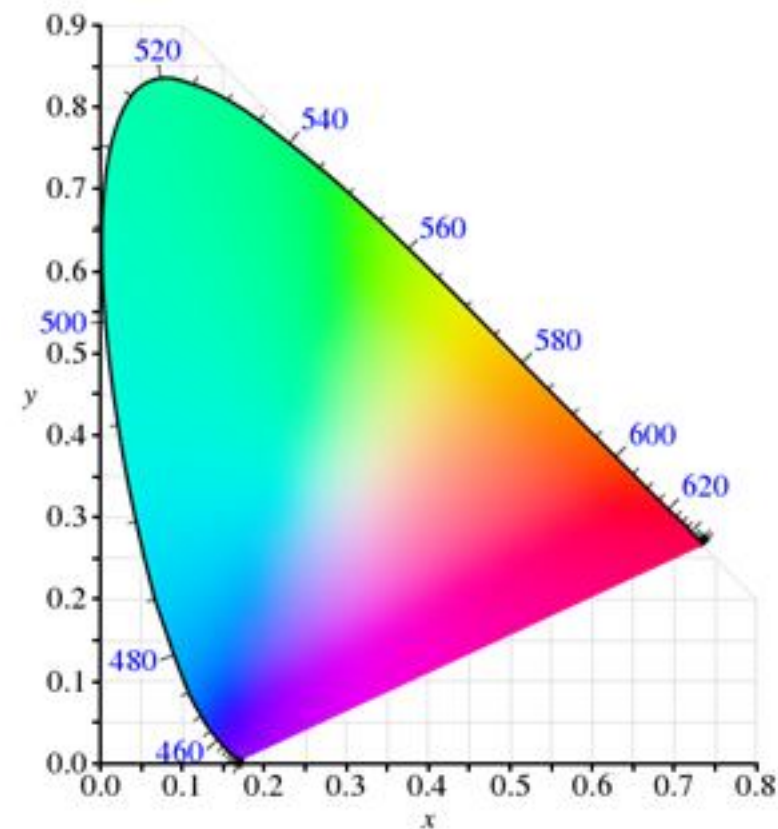


$$X = \int_{380}^{780} L_{e,\Omega,\lambda}(\lambda) \bar{x}(\lambda) d\lambda,$$
$$Y = \int_{380}^{780} L_{e,\Omega,\lambda}(\lambda) \bar{y}(\lambda) d\lambda,$$
$$Z = \int_{380}^{780} L_{e,\Omega,\lambda}(\lambda) \bar{z}(\lambda) d\lambda.$$

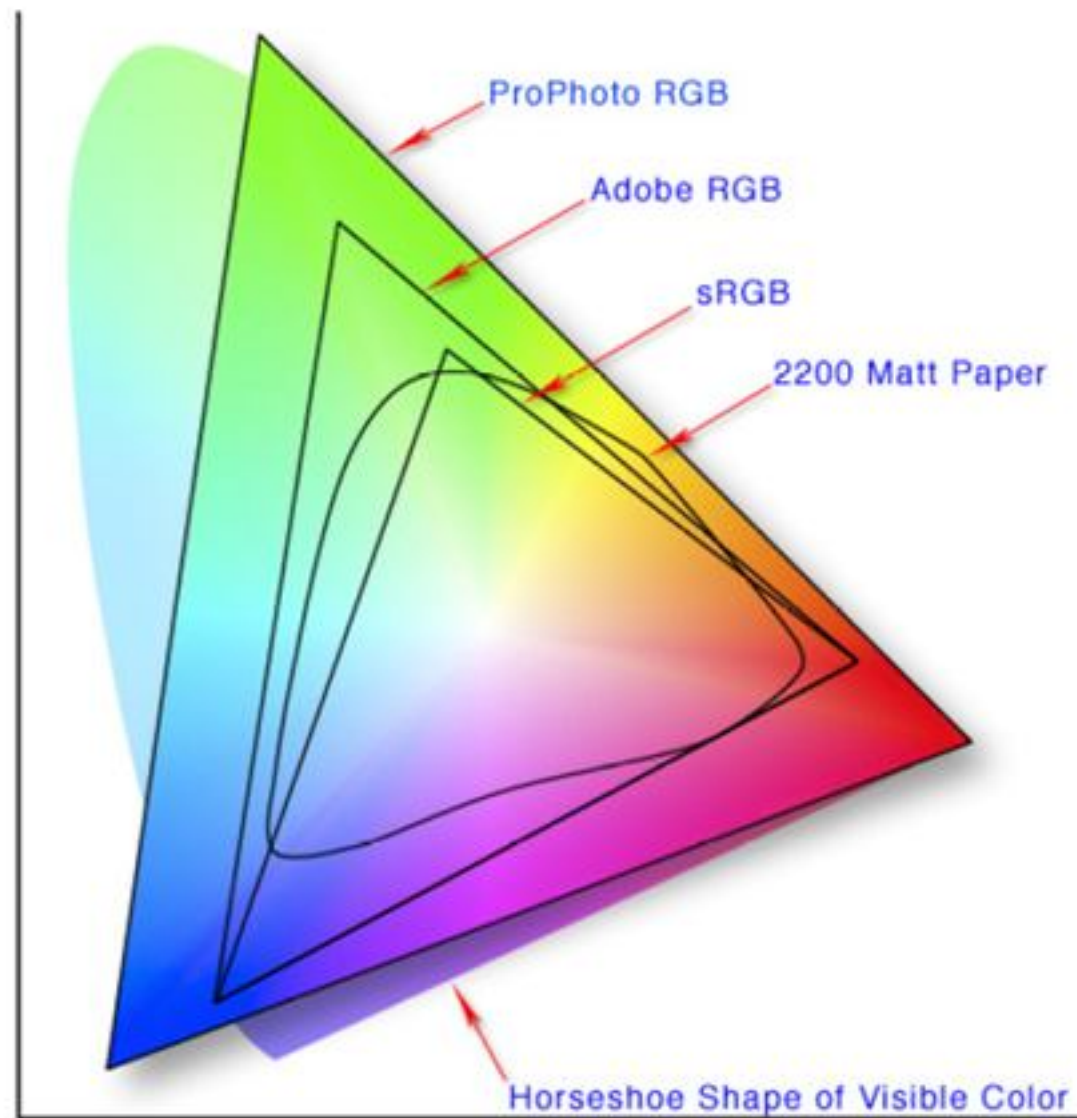
CIE xyY и gamut



- Опишем цветность двумя нормированными параметрами x и y
 - $x = X/(X+Y+Z)$
 - $y = Y/(X+Y+Z)$
- Можем построить диаграмму цветов для $x, y \in [0,1]$
- Наблюдения:
 - Когерентные источники света располагаются по дуге
 - Нижняя прямая «фиолетового» соответствует цветам, которые невозможно получить когерентным источником света
 - Никакими тремя реальными базовыми цветами невозможно покрыть видимый диапазон цветов



Различные модели



Нелинейность яркости



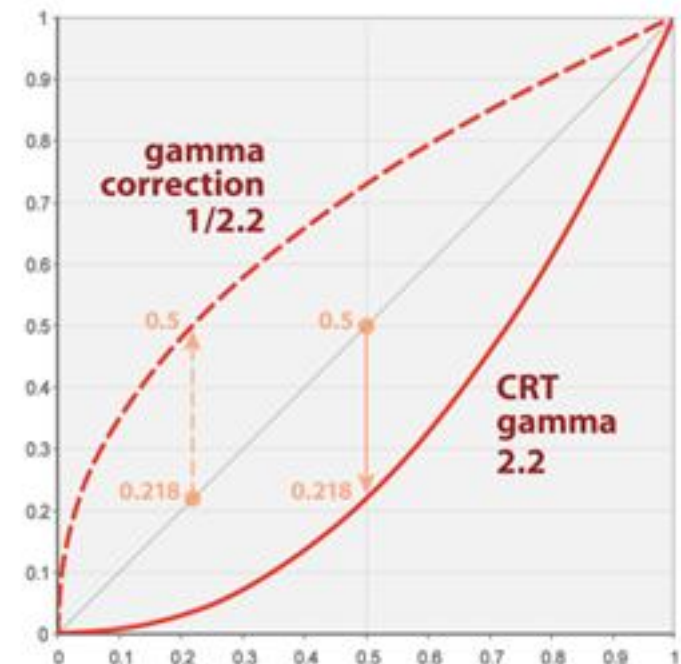
Физически равномерные яркости



Субъективно равномерные яркости

- Глаз лучше различает изменения яркости в тёмных областях, чем в светлых
- Поэтому хранить линейную яркость неэффективно
- Помогает гамма-преобразование

$$y = c \cdot x^\gamma$$



sRGB

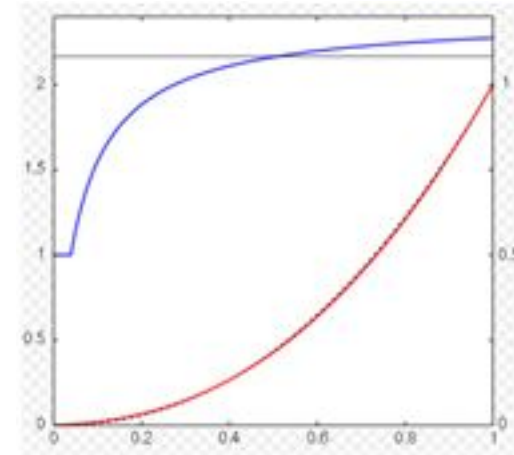


- Стандартная RGB модель для HDTV, мониторов, цифровых камер и т.д.
- Из XYZ в sRGB

$$\begin{bmatrix} R_{\text{linear}} \\ G_{\text{linear}} \\ B_{\text{linear}} \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$C_{\text{srgb}} = \begin{cases} 12.92 C_{\text{linear}}, & C_{\text{linear}} \leq 0.0031308 \\ (1 + a) C_{\text{linear}}^{1/2.4} - a, & C_{\text{linear}} > 0.0031308 \end{cases}$$

$$a=0.055$$



- Из sRGB и XYZ

$$C_{\text{linear}} = \begin{cases} \frac{C_{\text{srgb}}}{12.92}, & C_{\text{srgb}} \leq 0.04045 \\ \left(\frac{C_{\text{srgb}} + a}{1 + a} \right)^{2.4}, & C_{\text{srgb}} > 0.04045 \end{cases}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R_{\text{linear}} \\ G_{\text{linear}} \\ B_{\text{linear}} \end{bmatrix} \quad \text{- вычисление яркости}$$

Модель YIQ



В RGB нет отдельной «яркости», что бывает неудобно

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

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

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

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

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

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

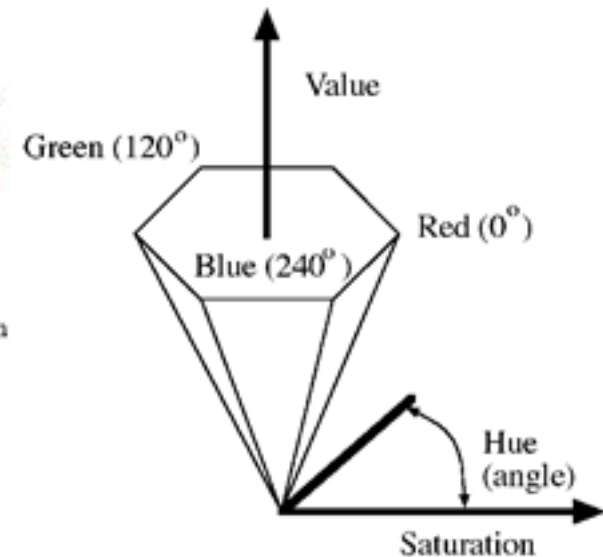
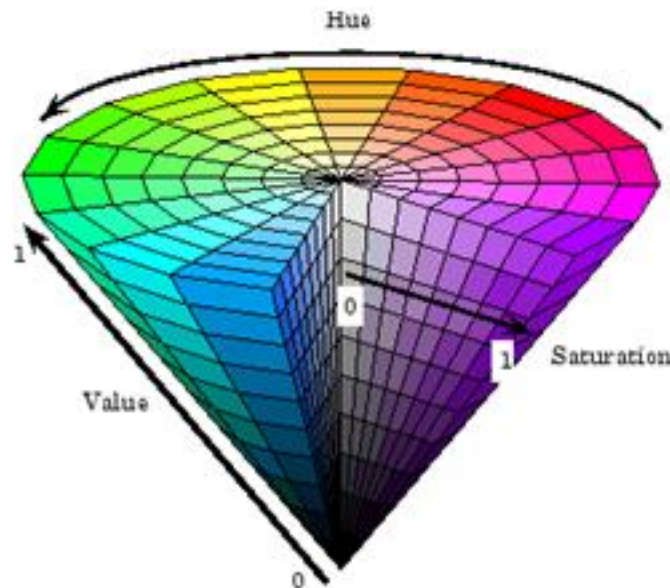
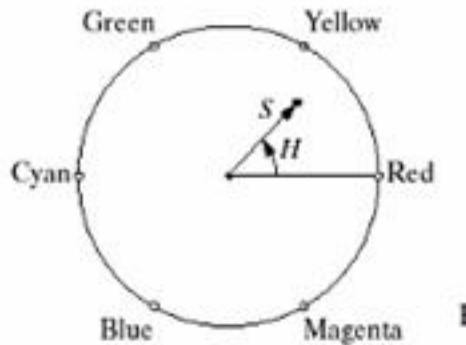
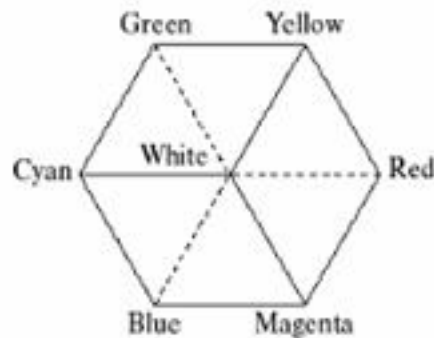
- Цветовая модель YIQ используется в коммерческом цветном телевидении США

- Модель YIQ совместима с черно-белым телевидением

- Модель YIQ используется в стандарте JPEG

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

Модель HSV (HIS)



В отличие от YIQ, координаты выбраны с учетом человеческого восприятия: Hue (Тон), Saturation(Насыщенность), Value (Intensity) (Интенсивность)

Является ли эта модель линейной?

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];

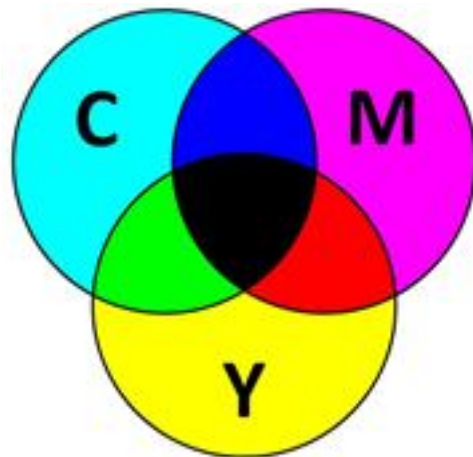
H : output value of hue in range $[0, 2\pi)$, -1 if S is 0;

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)/diff;
  "is the point within +/- 60 degrees of the green axis?"
  else if (g = I) then H := ( $2 * \pi/3$ ) +  $\pi/3$  *(b - r)/diff;
  "is the point within +/- 60 degrees of the blue axis?"
  else if (b = I) then H := ( $4 * \pi/3$ ) +  $\pi/3$  *(r - g)/diff;
  if (H ≤ 0.0) H := H +  $2\pi$ ;
}
```

Algorithm 15: Conversion of RGB to HSI.

Субтрактивная модель СМУК

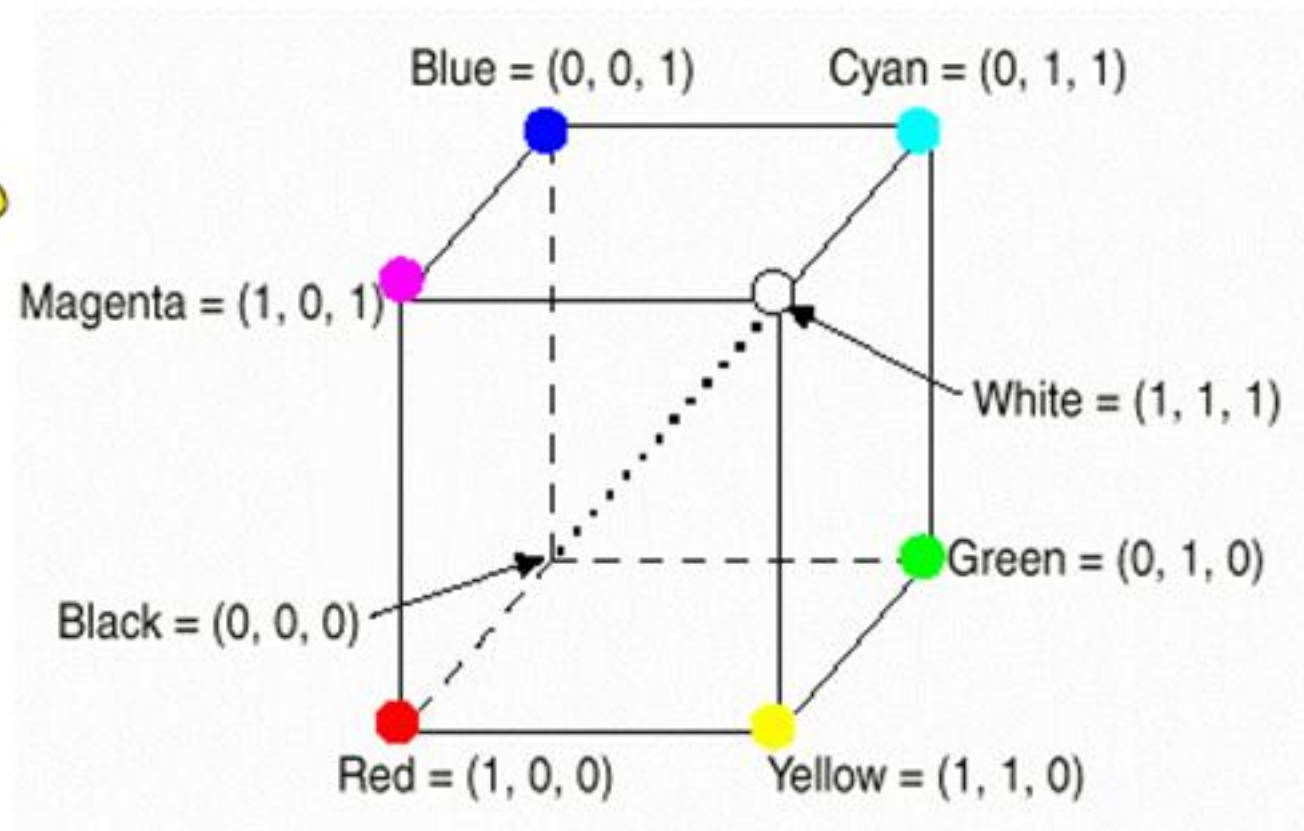


CMY

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

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

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

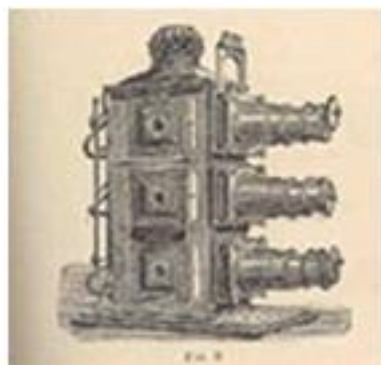


- Аддитивная система – RGB
- Субтрактивная система – CMY
- СМУК для повышения качества

Первые цветные фотографии



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



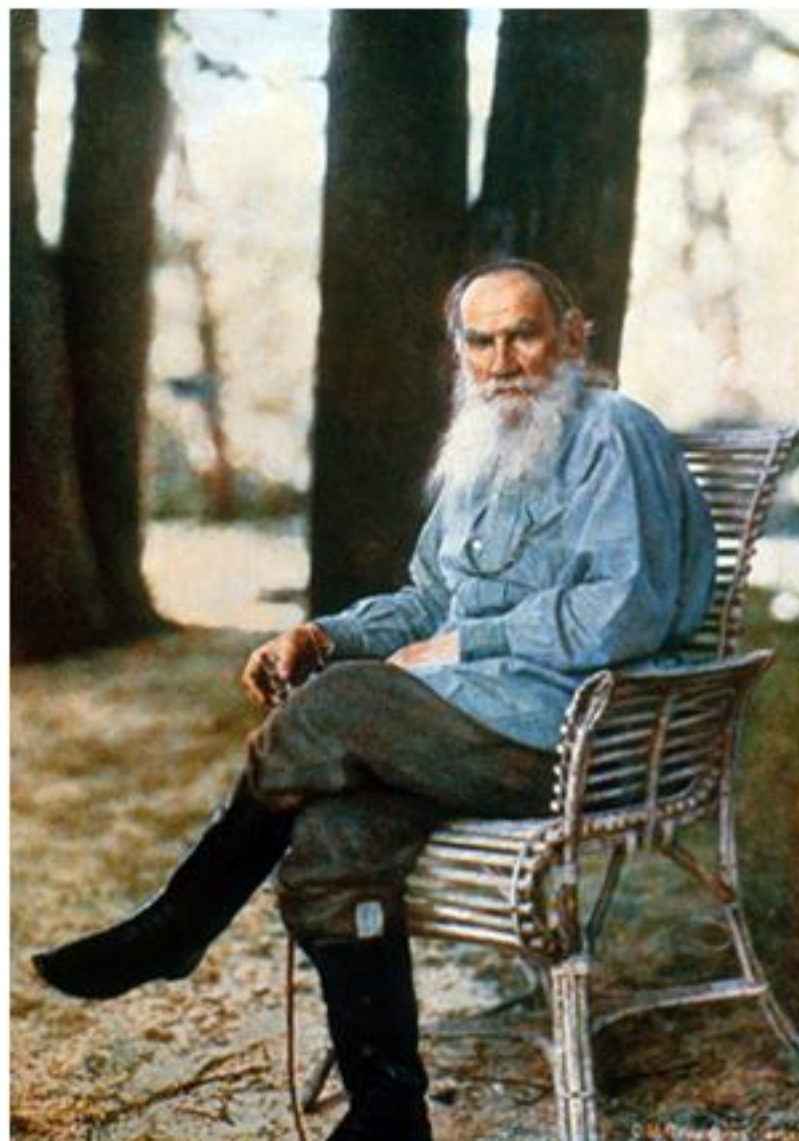
Ламповый
проектор



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

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

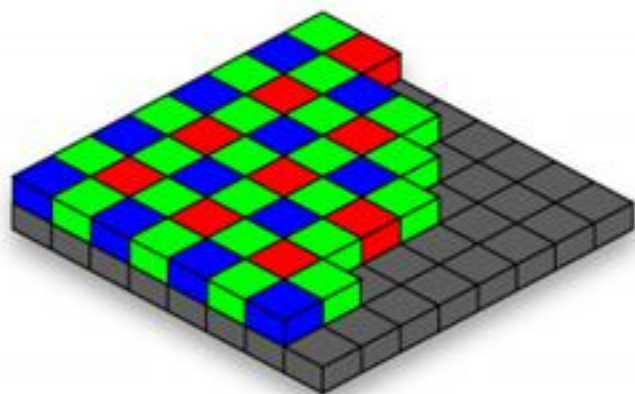
Лев Толстой



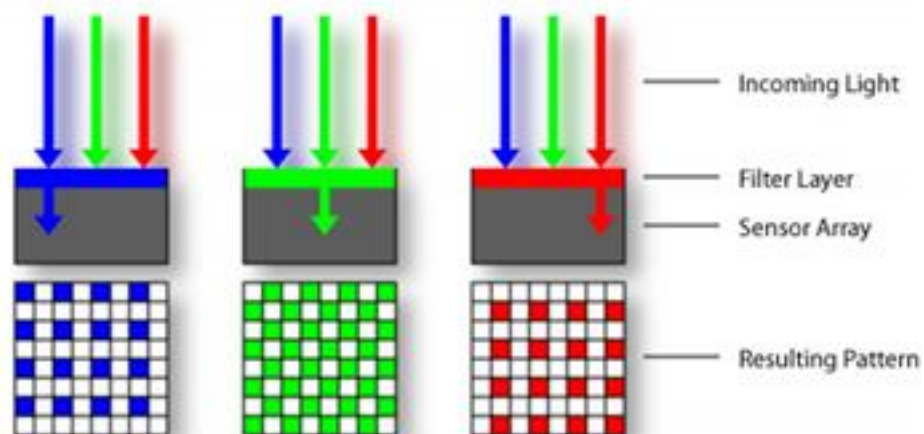
Цветное цифровое изображение



Байеровский шаблон



Демозаикинг (оценка пропущенных значений цвета)



Устранение мозаичности и ошибки



Original image



Bilinear interpolation

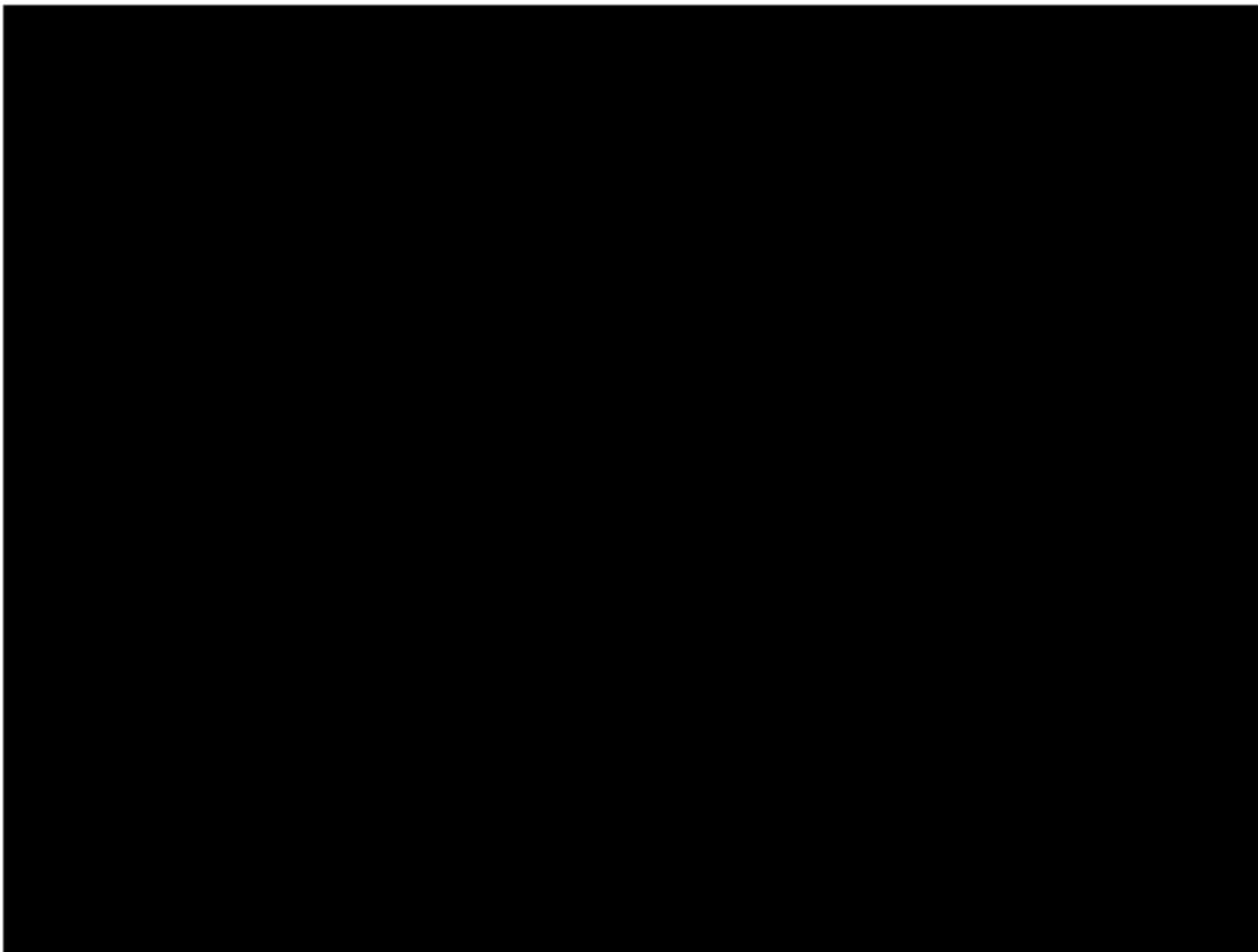


Proposed method

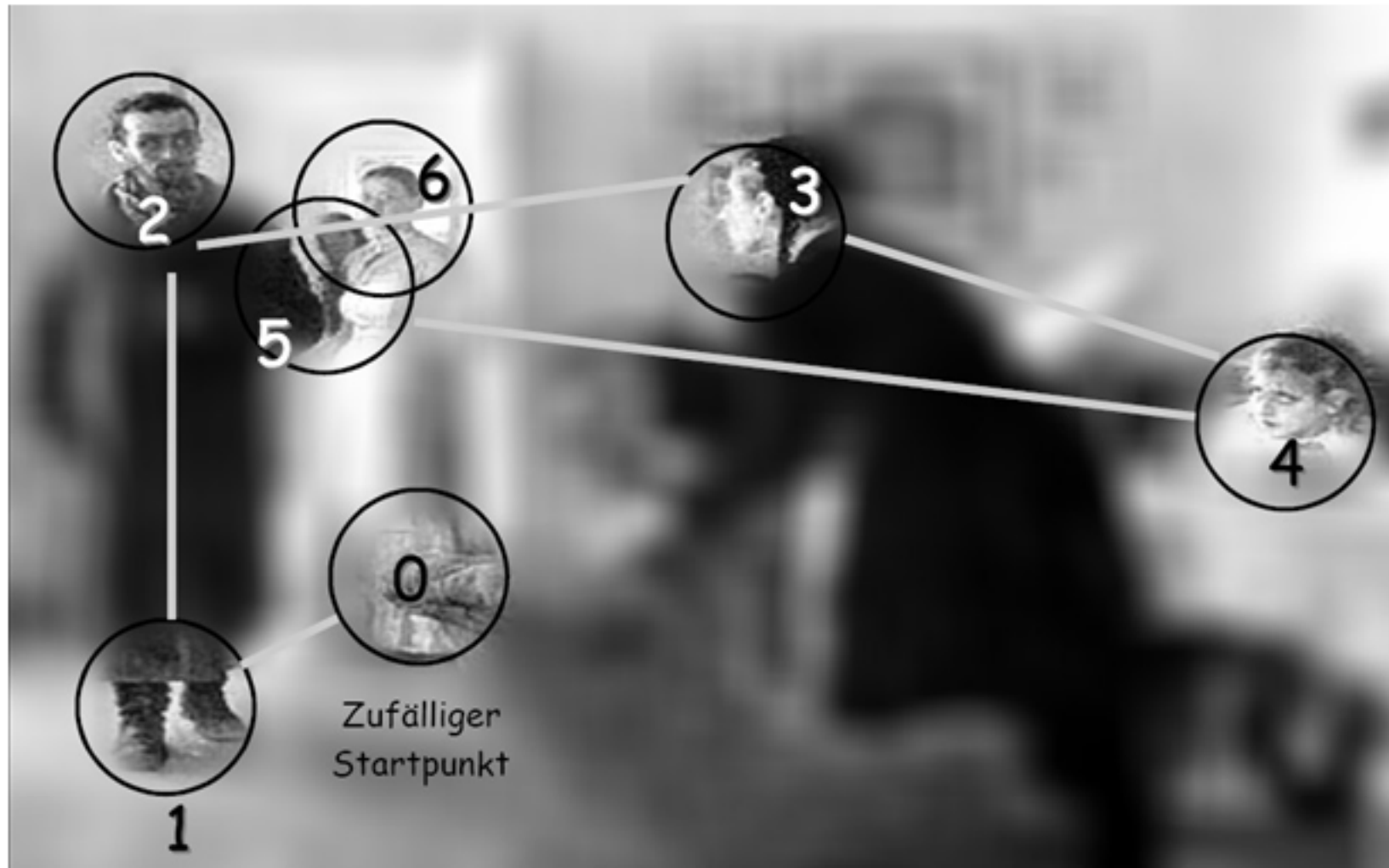
Тонкие черные и белые детали
интерпретируются как изменения цвета

Фокус внимания

Тест на внимательность

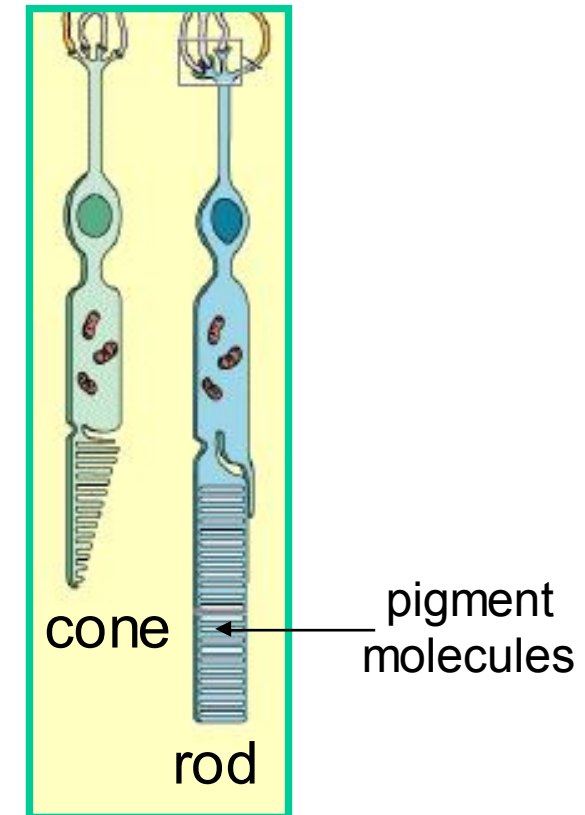
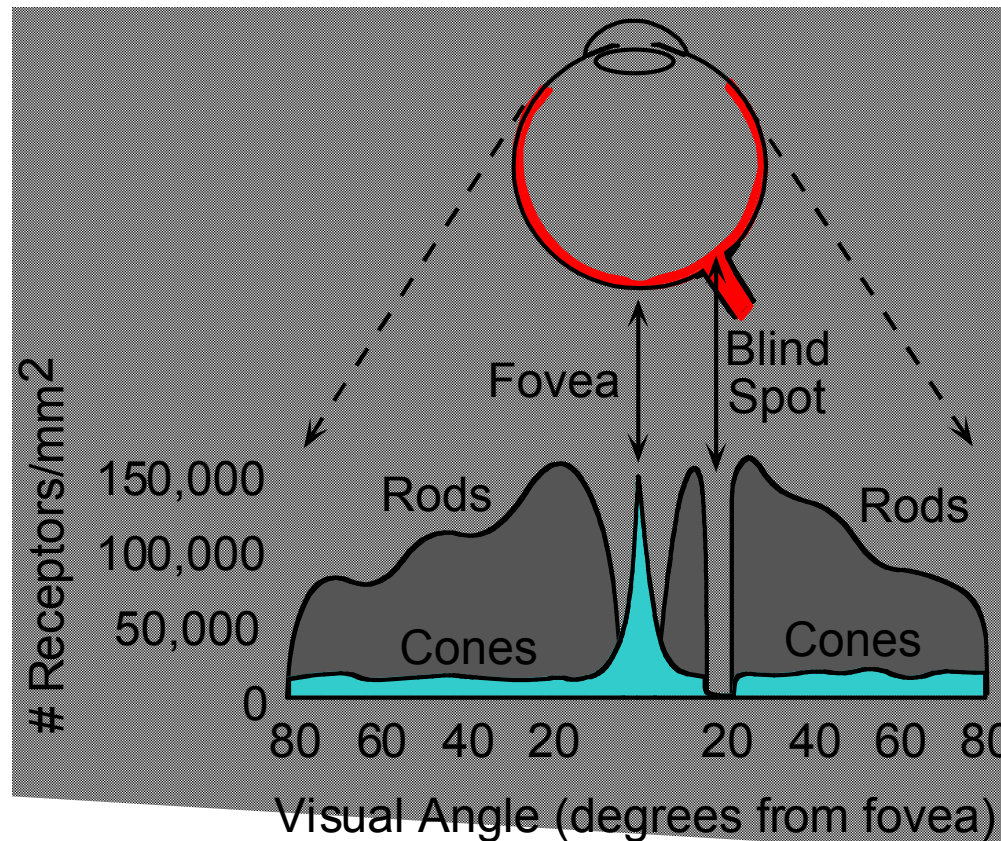


Что мы на самом деле видим



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



2

Estimate material circumstances of the family



3

Give the ages of the people.



4

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



5

Remember the clothes worn by the people.



6

Remember positions of people and objects in the room.



7

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

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