



Лаборатория компьютерной
графики и мультимедиа
ВМК МГУ имени М.В. Ломоносова

Курс «Компьютерное зрение»

Лекция №1

«Введение в предмет. Цифровое изображение. Свет
и цвет.»

Антон Конушин

2024 год

План лекции



- Что такое «компьютерное зрение»?
- Цифровое изображение
- Свет и цвет
- Модели цвета
- Особенности человеческого зрения



Что такое «компьютерное зрение»?

Что такое компьютерное зрение?



Задача зрения: понять, что находится на изображении

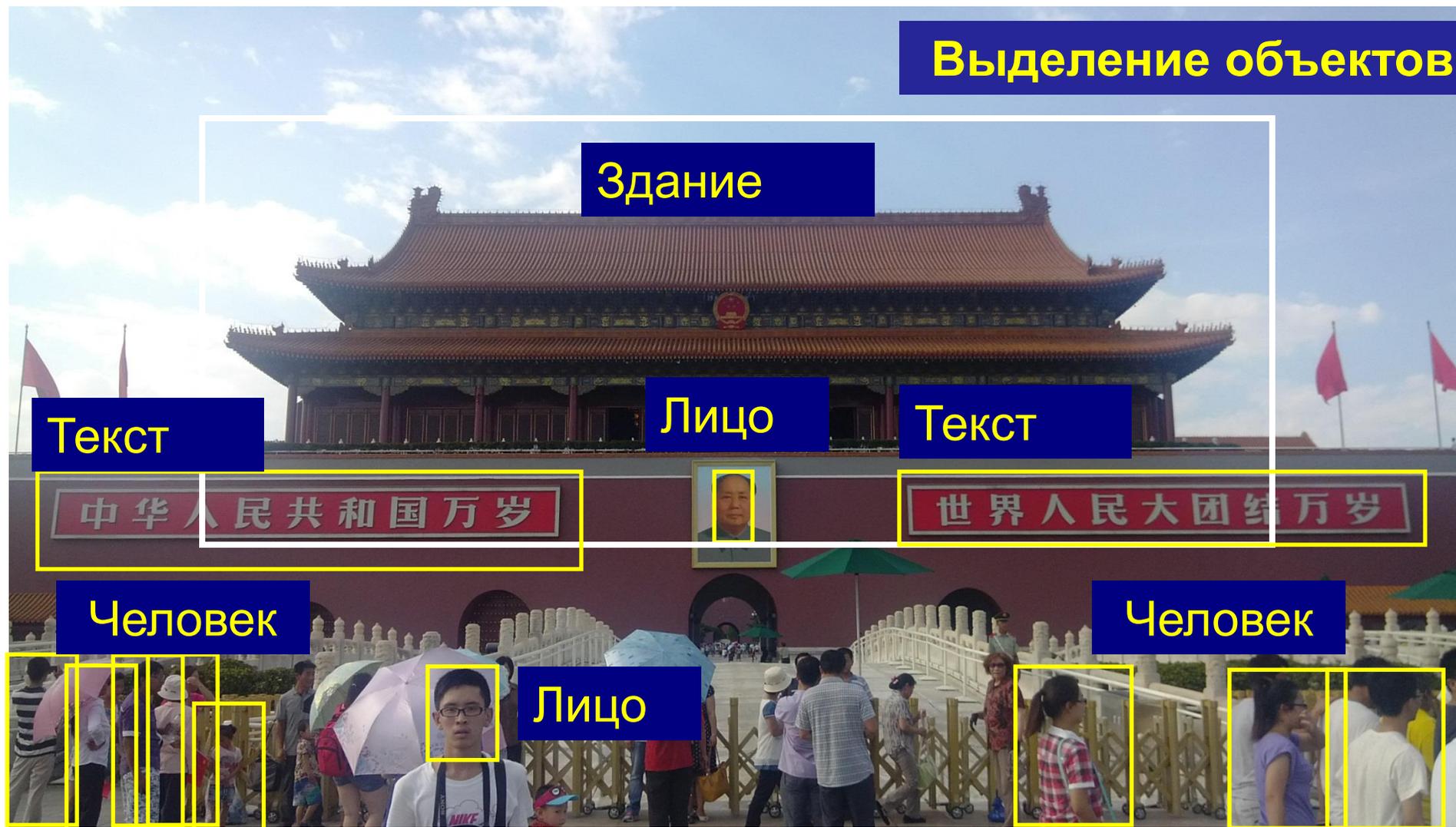
Компьютерное зрение: построение компьютерной модели системы зрения

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

Свойства объектов и атрибуты



- outdoor
- city
- Beijing, China
- Tiananmen Square

Классификация и поиск похожих

Sky is blue

Wind is slow

Male, Mao Zedong



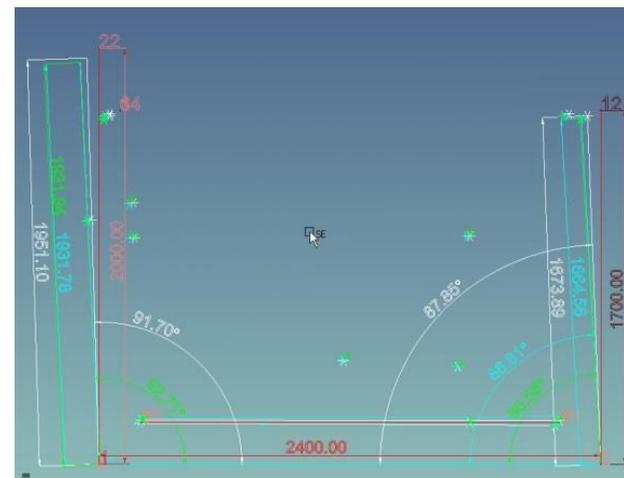
Profile, Female,
Unknown

Frontal, Male,
Unknown

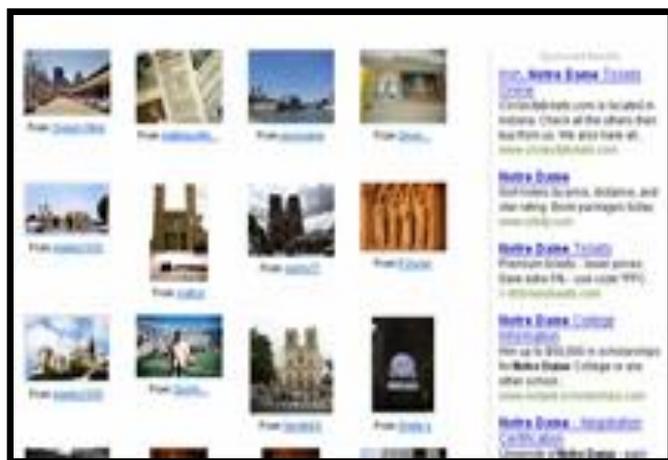




Какой формы? (Метрическое зрение)



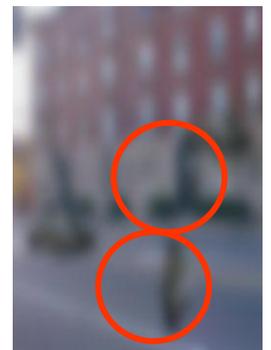
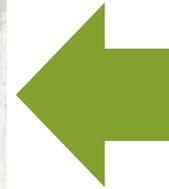
Фотограмметрия



Source: Pollefeys et al.

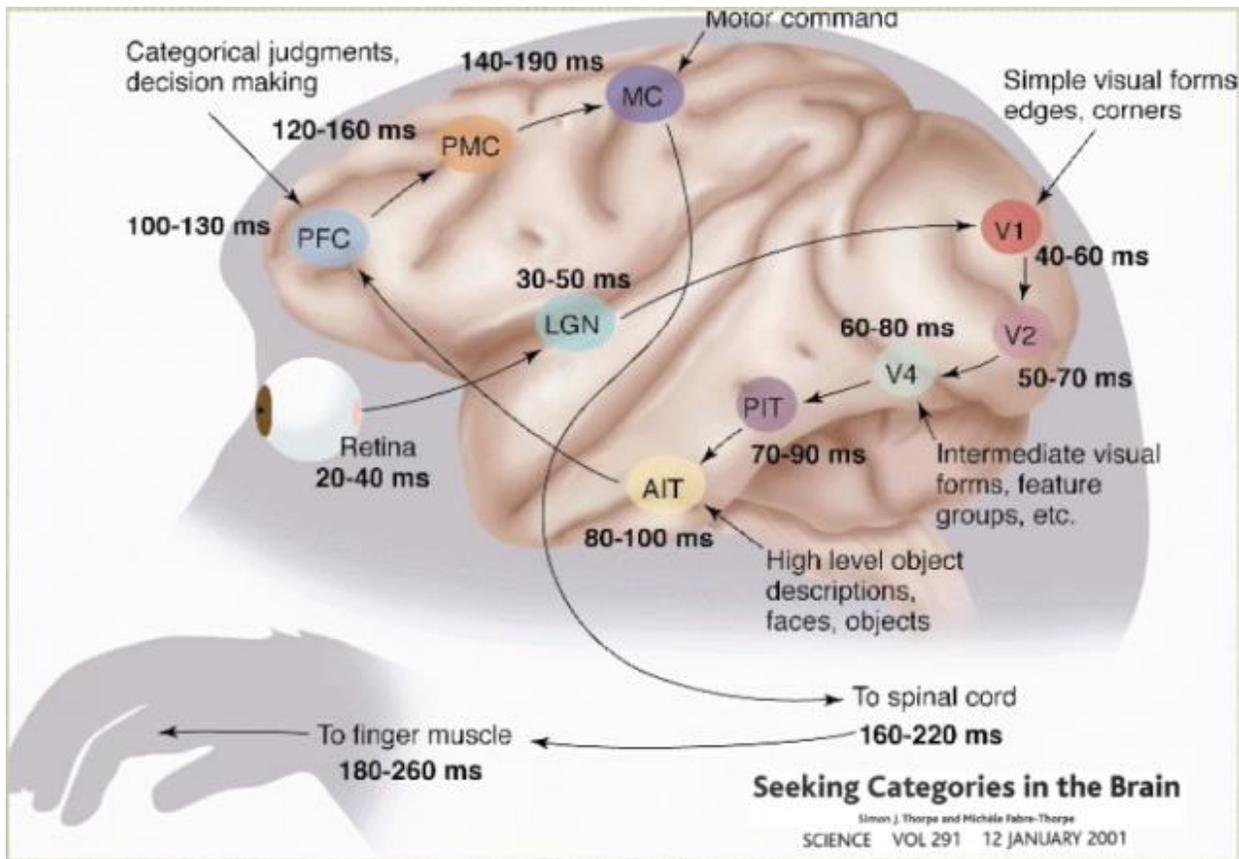
3-х мерная реконструкция по изображениям

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?

Работы Дэвида Марра (1970е)



«Primal sketch»

- Низкоуровневые («low-level») свойства изображения: направленные края, отрезки и т.д.

«2.5D sketch»

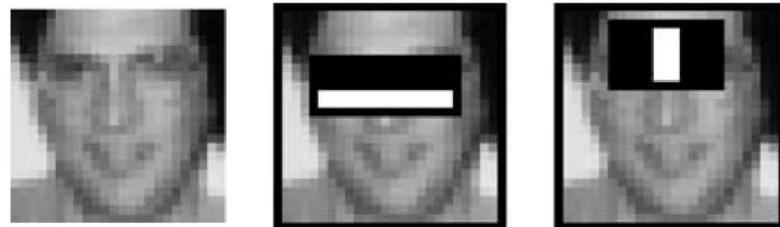
- Упорядочивание по глубине (бинокулярное стерео), учёт текстуры и т.д.

«3D model»

- Распознавание объектов и представление о 3х мерном мире

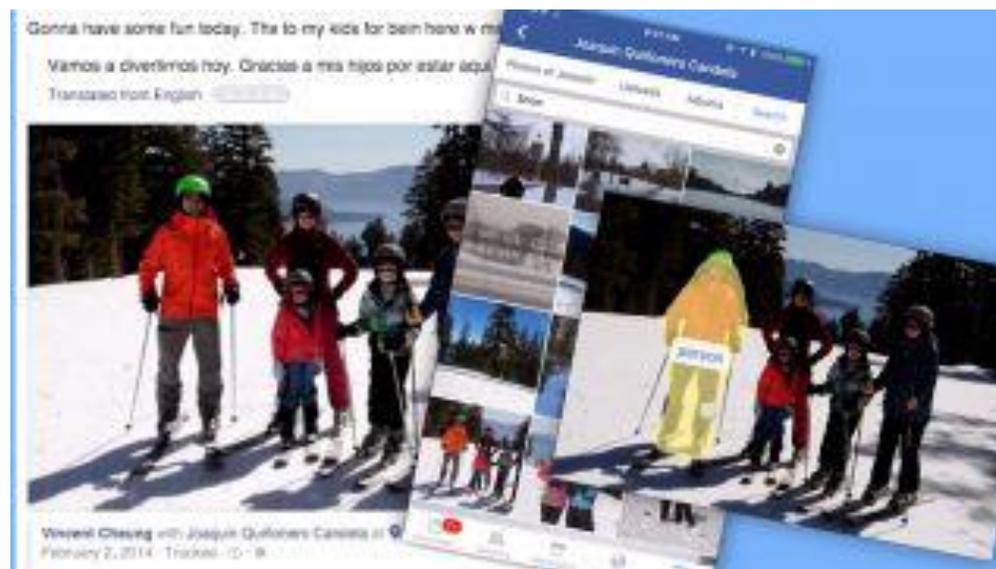
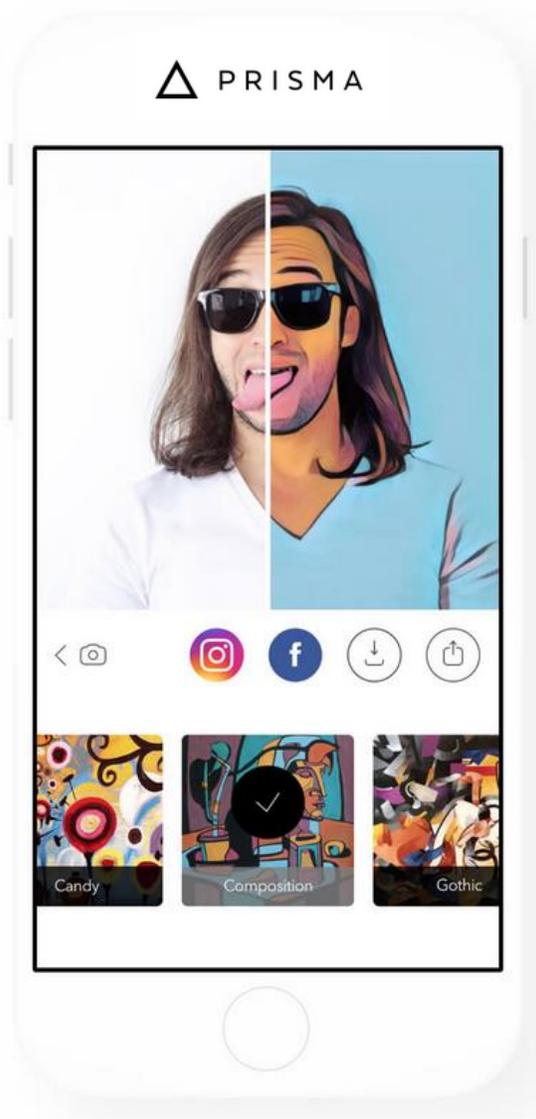
«The Marr Prize» – главная премия в области компьютерного зрения

Детектор лиц Viola-Jones (2001)



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

Глубокое обучение (2012 - н.в.)





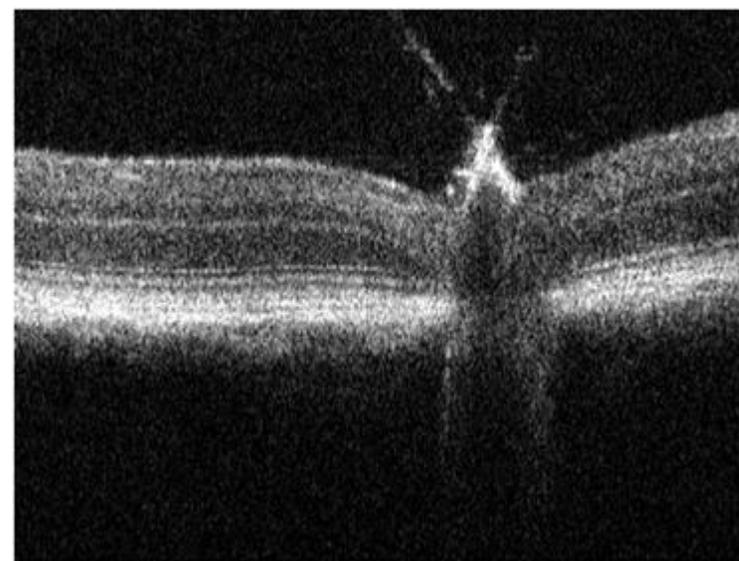
Цифровое изображение

Изображение



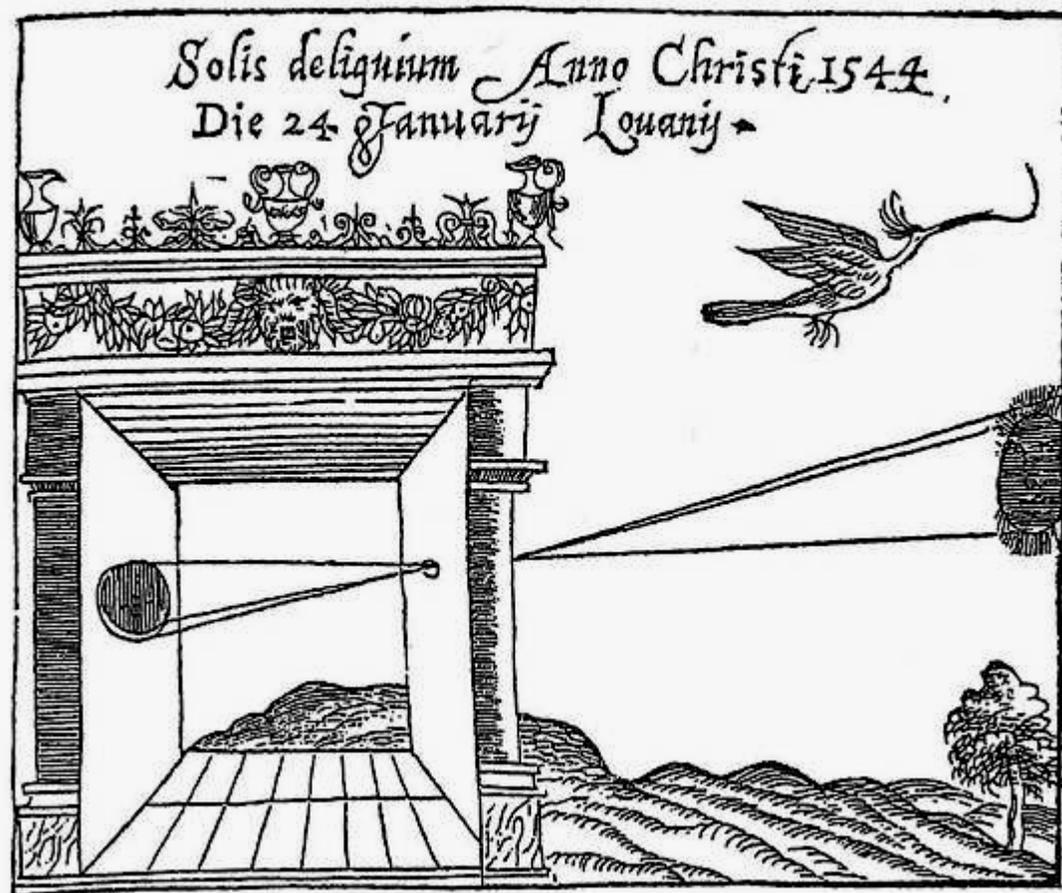
Изображение оптическое – картина, получаемая в результате прохождения через оптическую систему лучей, распространяющихся от объекта, и воспроизводящая его контуры и детали.

(Физический энциклопедический словарь.)



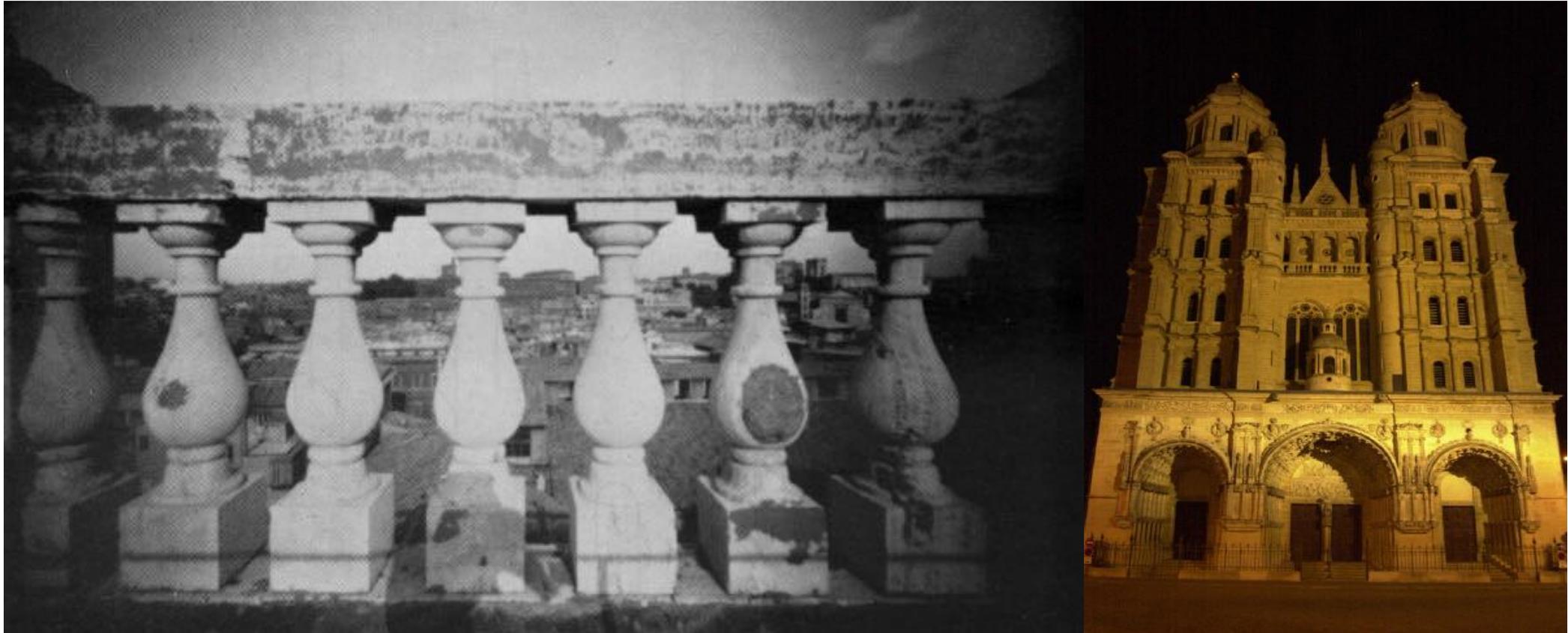
Оптические системы могут быть очень разными!

Камера-обскура



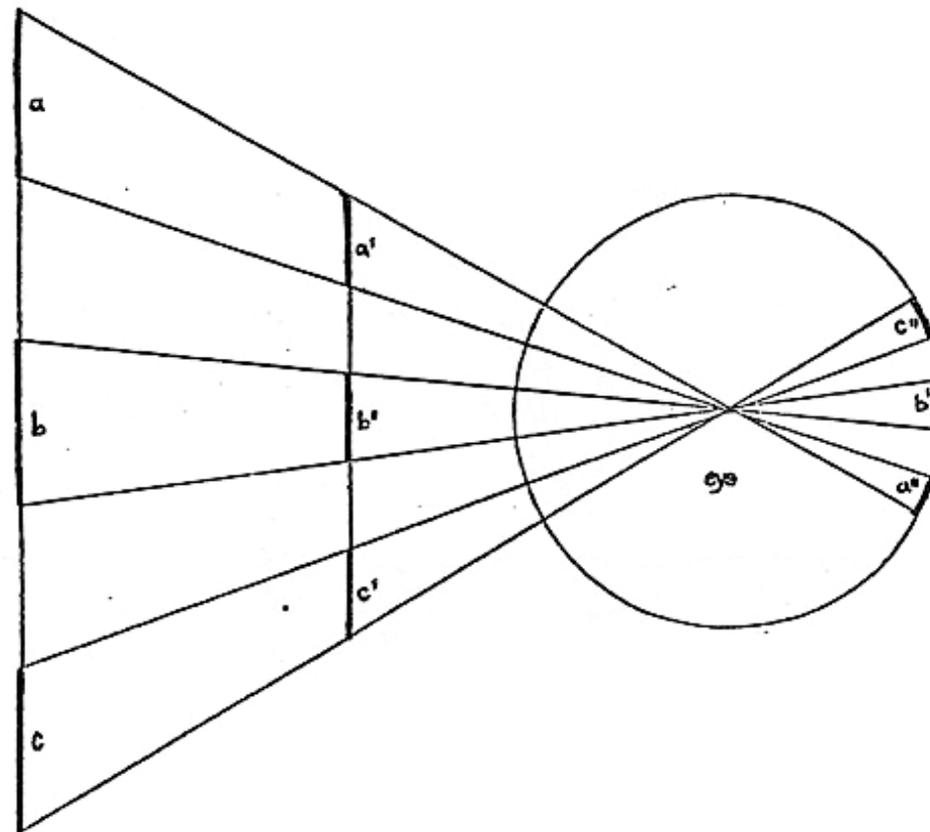
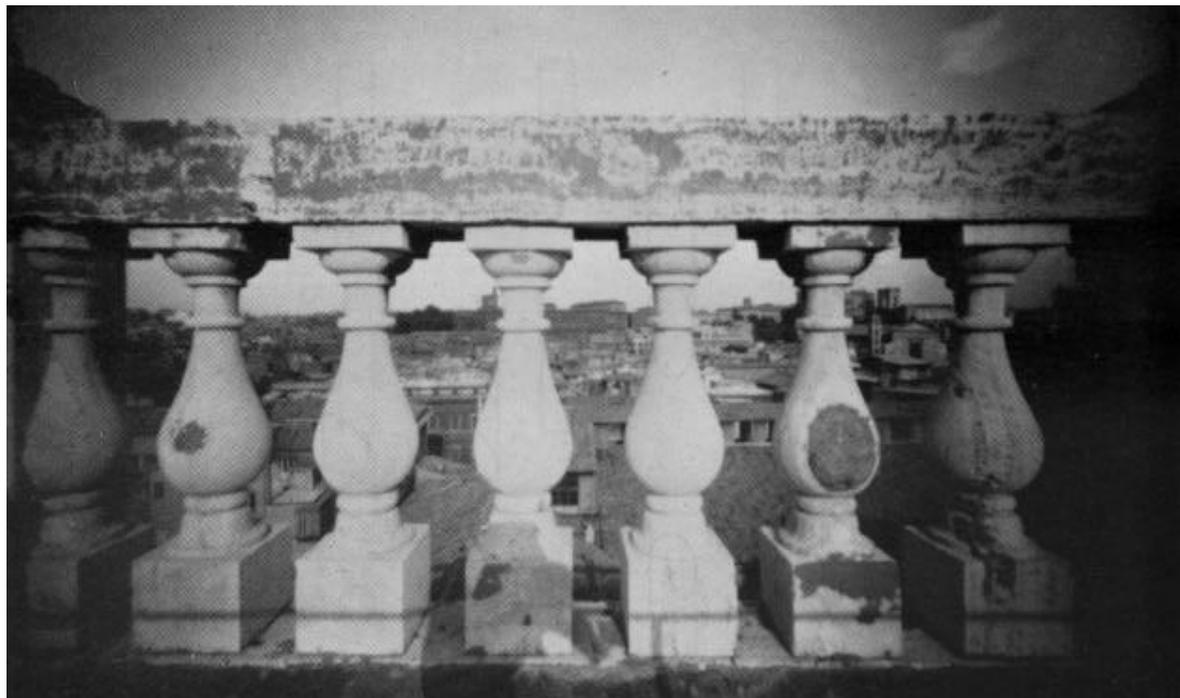
- Простейшее устройство для получения изображений
- Принцип был известен еще [Моцзы](#) (470 – 391 до Н.Э.) и Аристотелю (384-322 до Н.Э.)
- Помогала художникам: описана Леонардо да Винчи (1452-1519)
- Математическая модель - перспективная проекция.
- Лучи проходят через 1 точку (центр проекции) и формируют изображение на картинной плоскости

Перспективные искажения



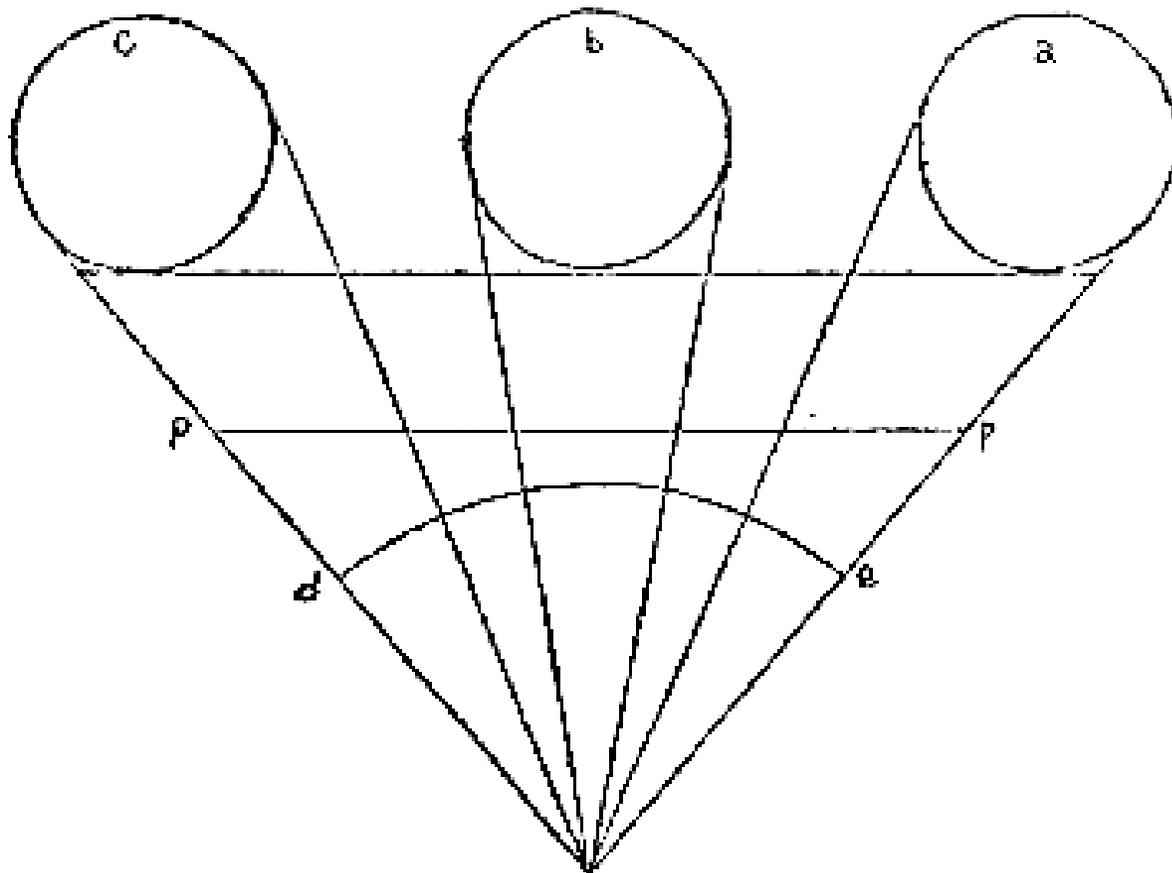
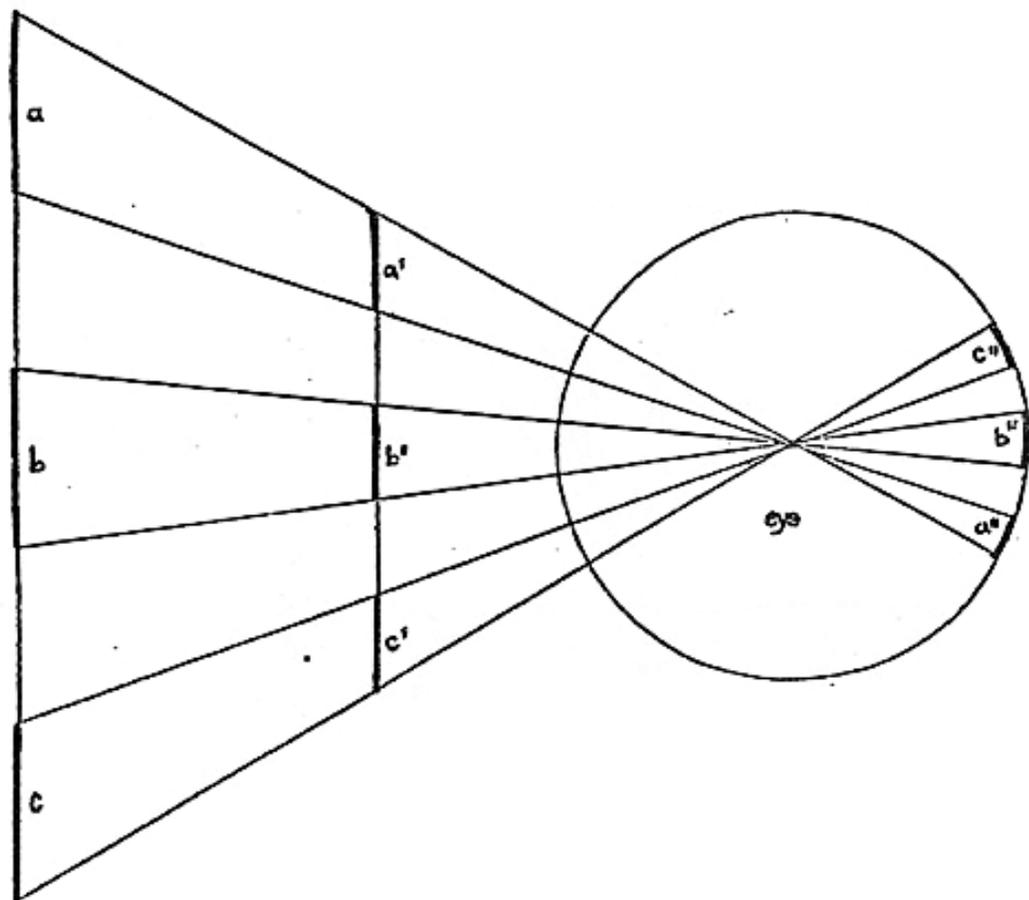
- Перспективная проекция приводит к искажениям свойств объектов
 - Сходящиеся вертикальные линии
 - Крайние колонны кажутся толще
- Эти искажения вызваны не погрешностью оптической системы!

Перспективные искажения



Почему так происходит?

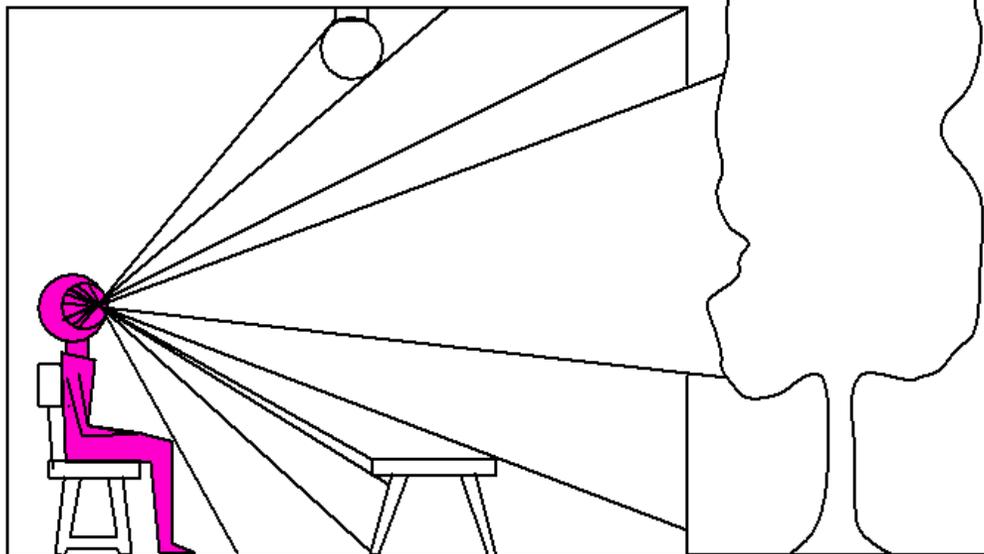
Перспективные искажения



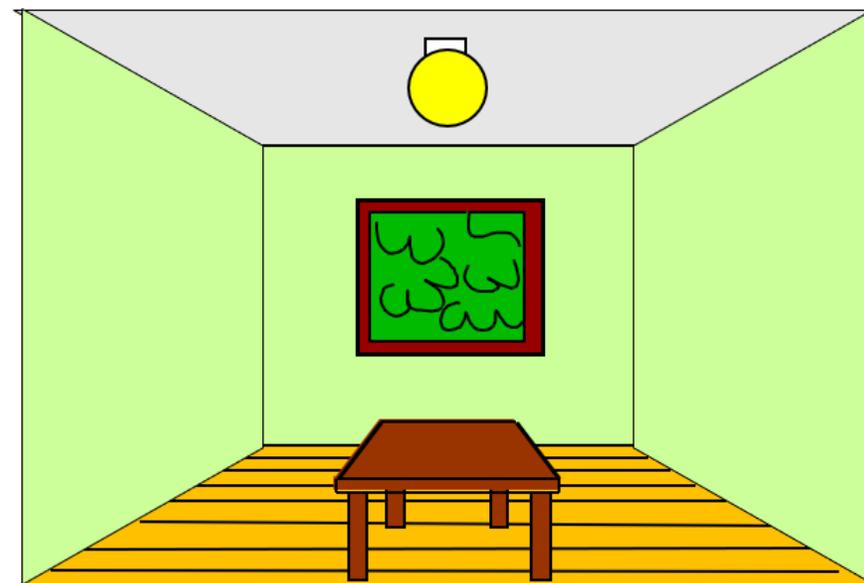
Машина Понижения Размерности



3D мир



2D картина

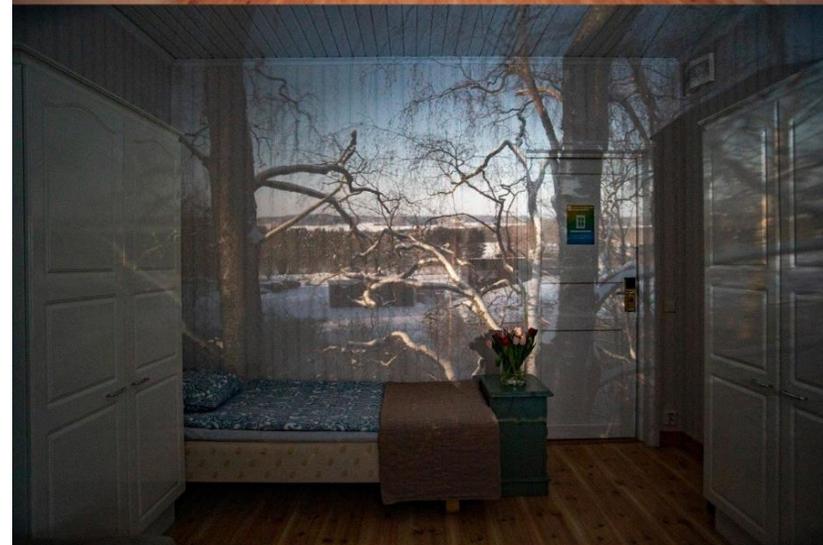
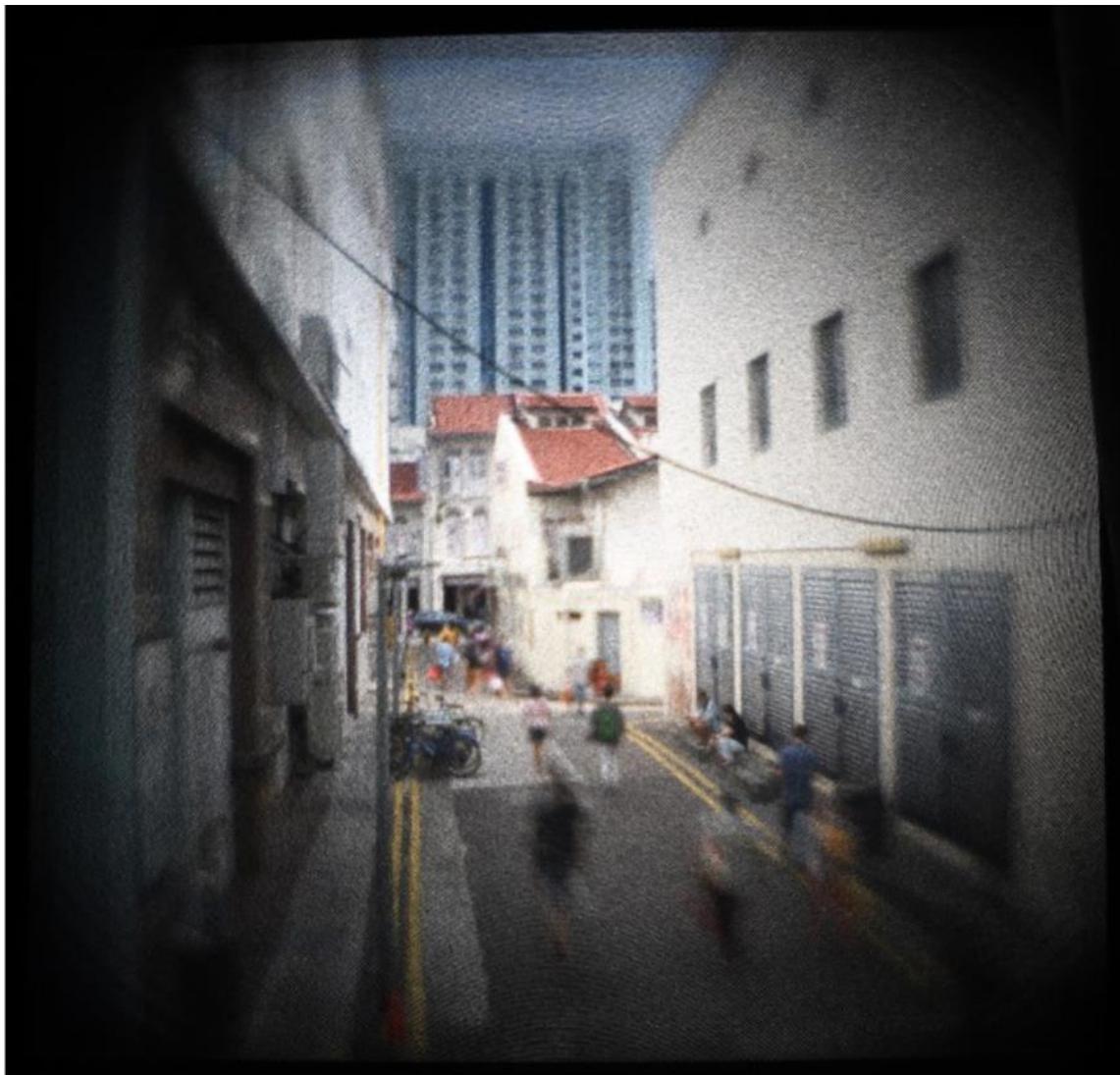


Что мы теряем?

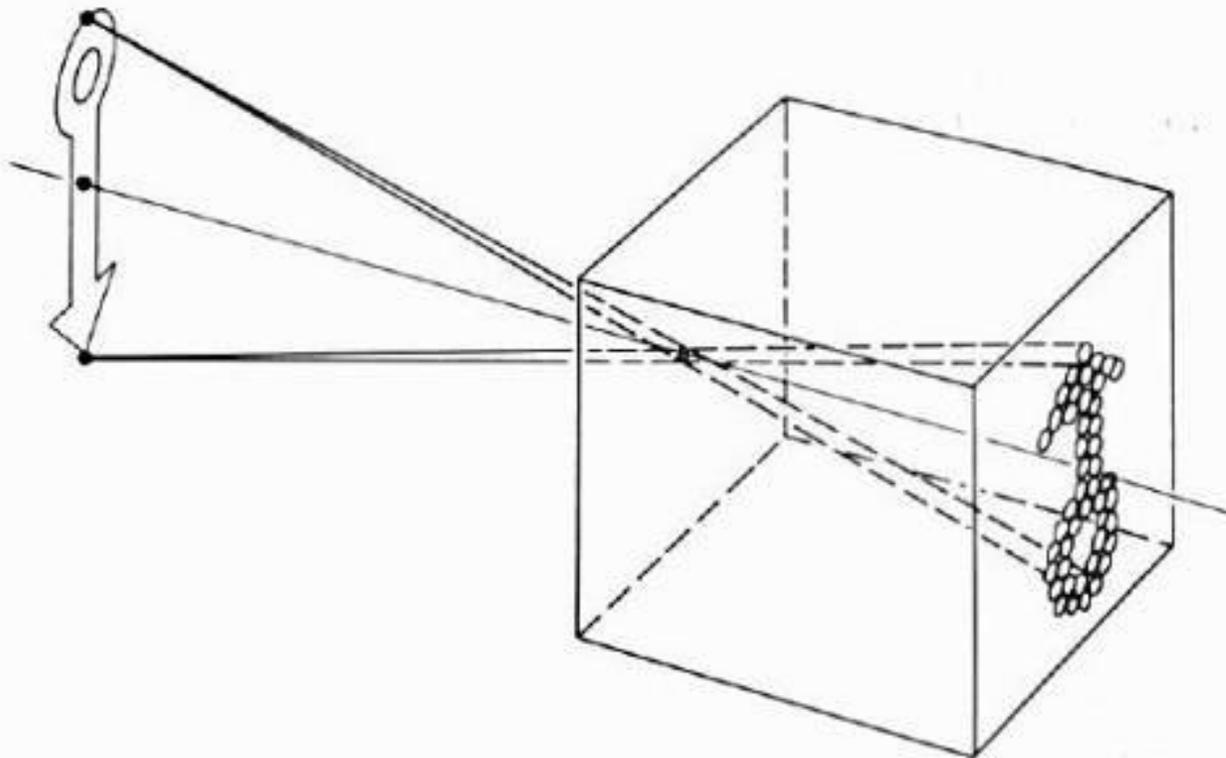
- Углы
- Расстояния и длины

Подробнее геометрическая
модель рассматривается
позже

Камера-обскура в реальной жизни

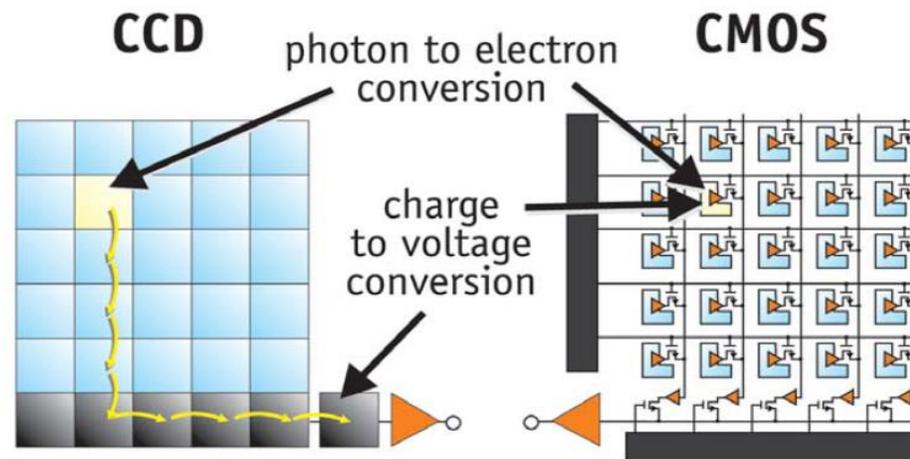


Цифровая камера

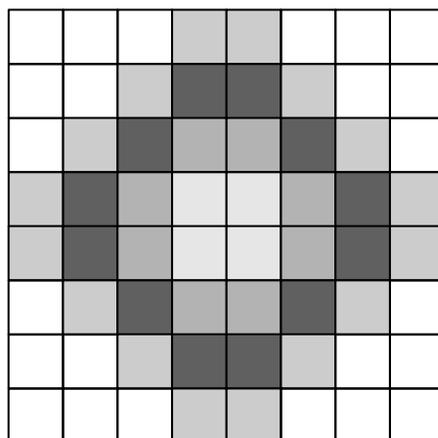
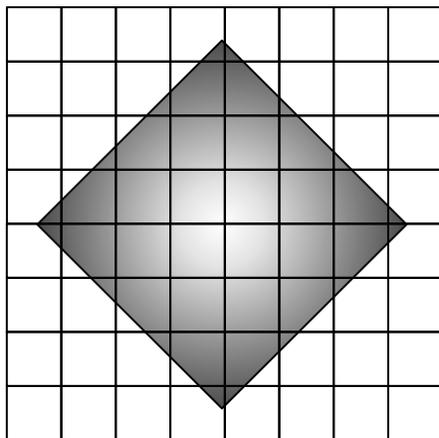


Тот же самый принцип, но с объективом и цифровой матрицей

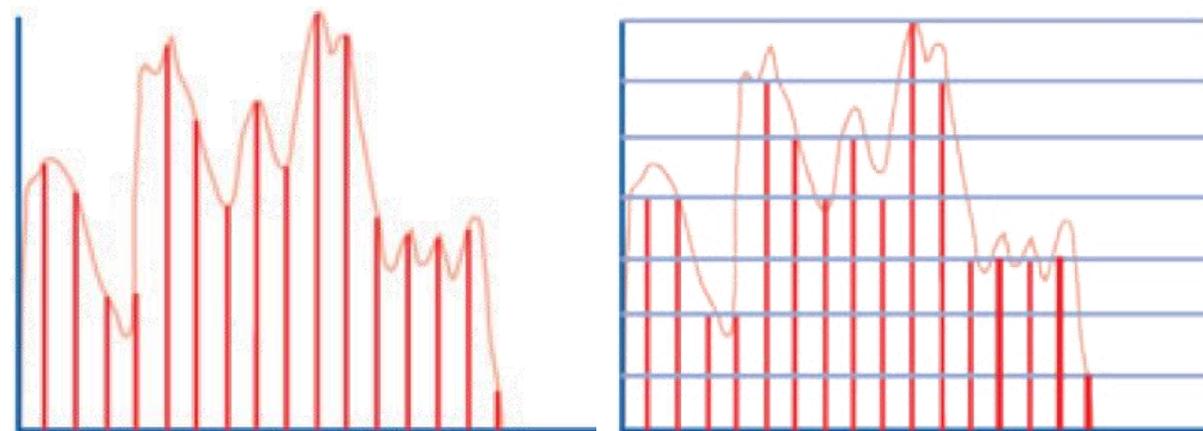
Цифровая камера - дискретизация



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.



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 интенсивности света, измеренного на 2х мерной прямоугольной сетке

- b (интенсивность) 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

Как получить цветные фотографии?





Что такое цвет?

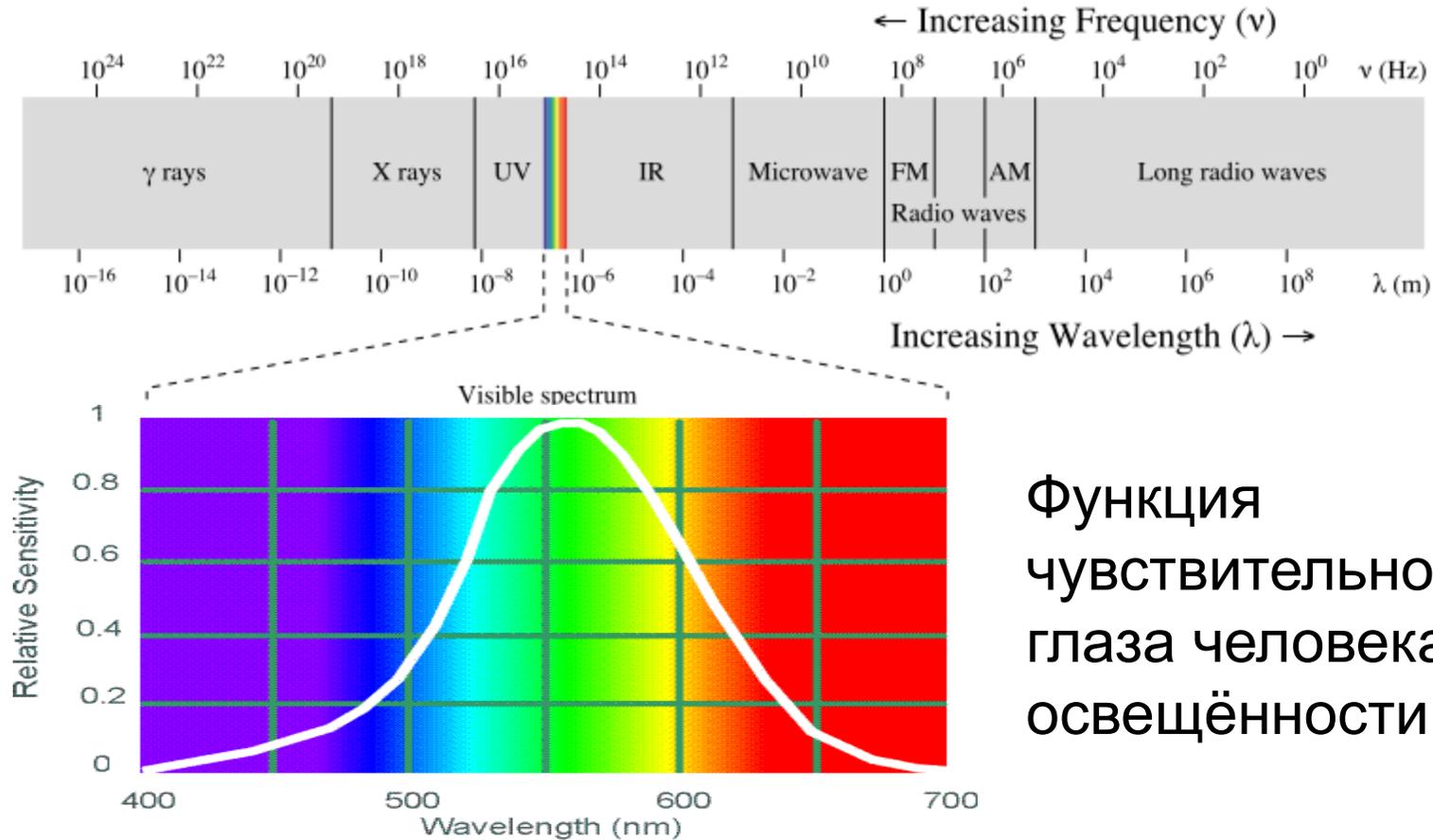
- Цвет – это психологическое свойство нашего зрения, возникающее при наблюдении объектов и света, а не физические свойства объектов и света (S. Palmer, *Vision Science: Photons to Phenomenology*)
- Цвет – это результат взаимодействия света, сцены и нашей зрительной системы



Восприятие света
человеком изучают
науки фотометрия и
колориметрия

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

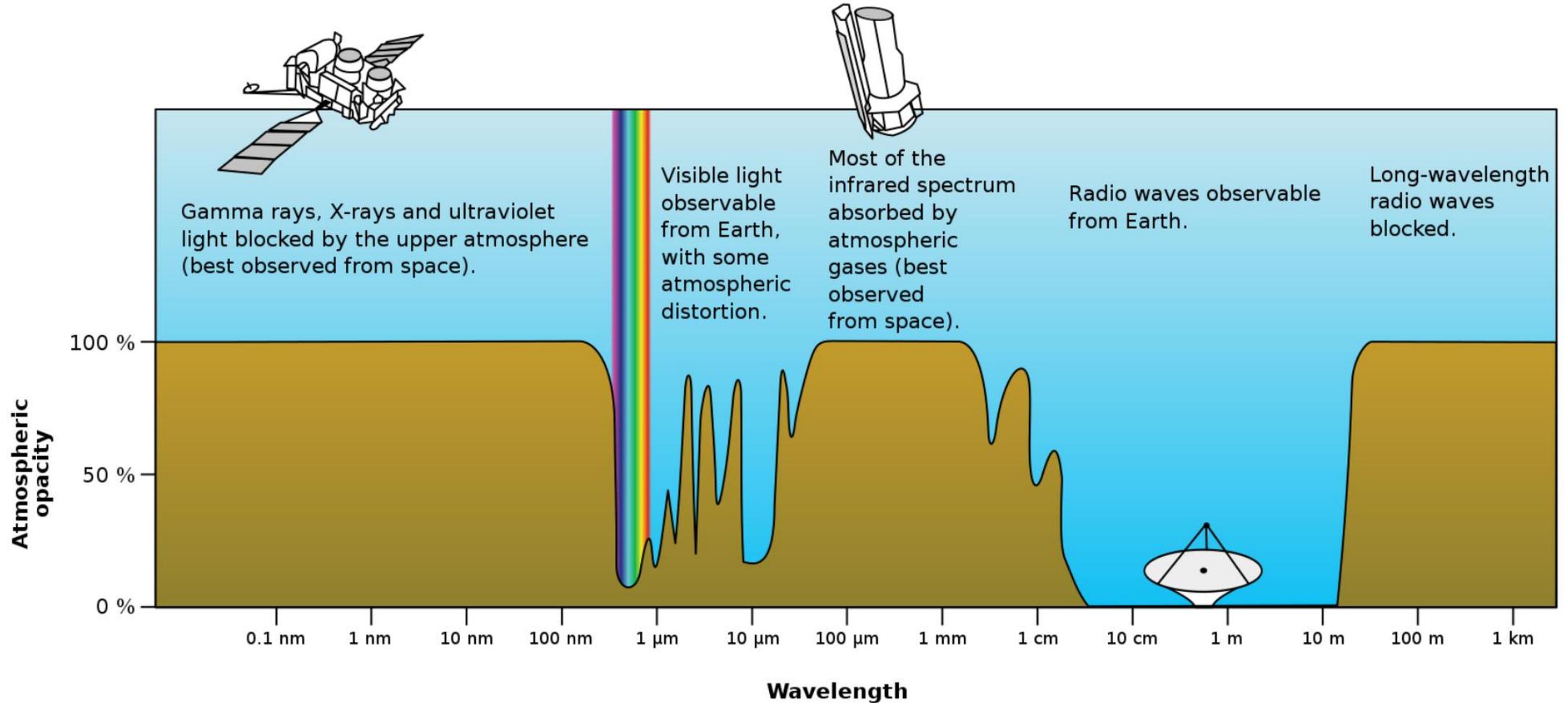
Видимый свет



Функция
чувствительности
глаза человека к
освещённости

- Электромагнитное излучение в диапазоне [380nm,780nm]
- Как так получилось?

Видимый свет

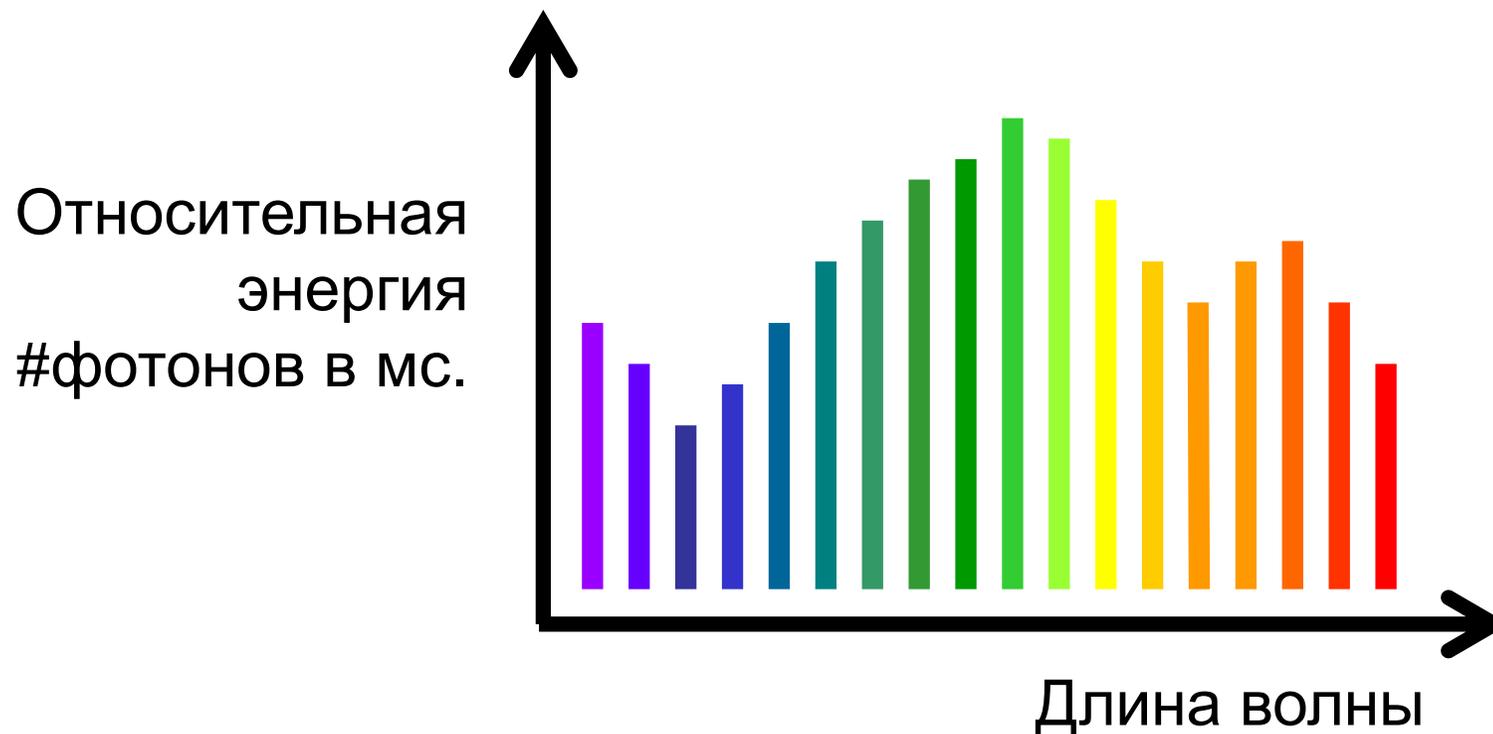


Видимый свет попадает в основное «оптическое окно» земной атмосферы (~46% энергии)

Физика света

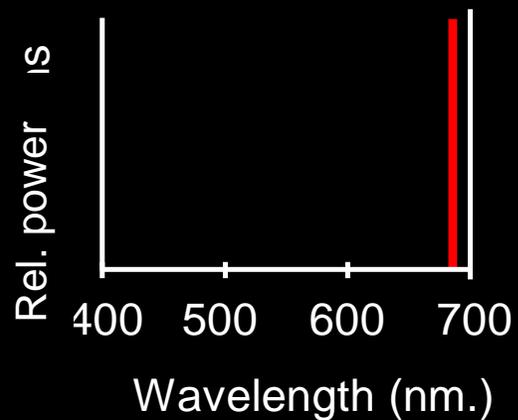


Любой источник света можно полностью описать спектром: количество излученной энергии в единицу времени для каждой длины волны в интервале 380 - 780 nm.

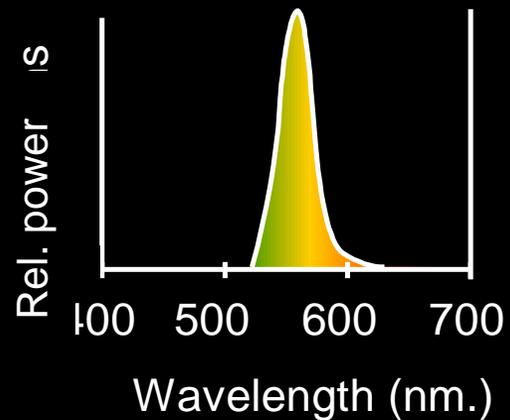




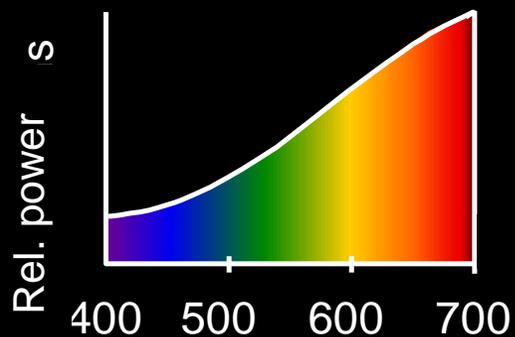
A. Ruby Laser



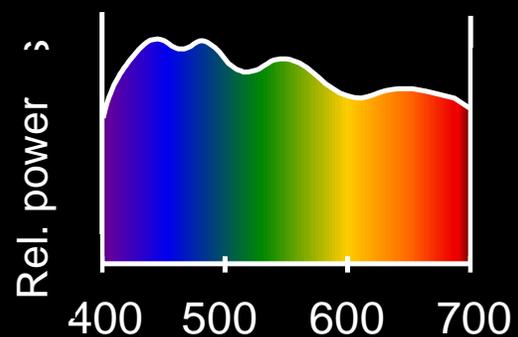
B. Gallium Phosphide Crystal



C. Tungsten Lightbulb



D. Normal Daylight

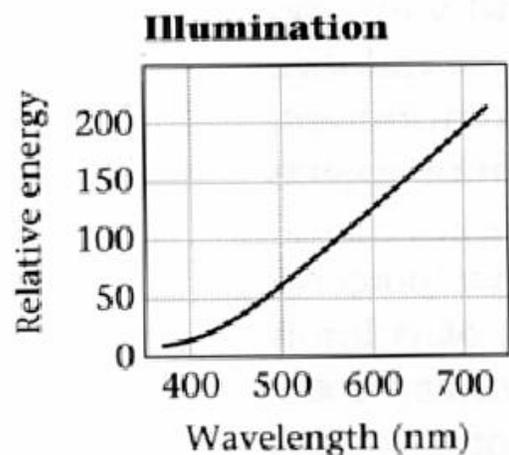


Примеры спектров
разных источников
света

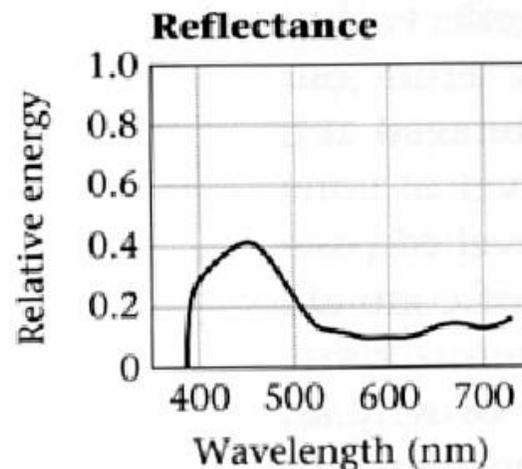
Взаимодействие света и объектов



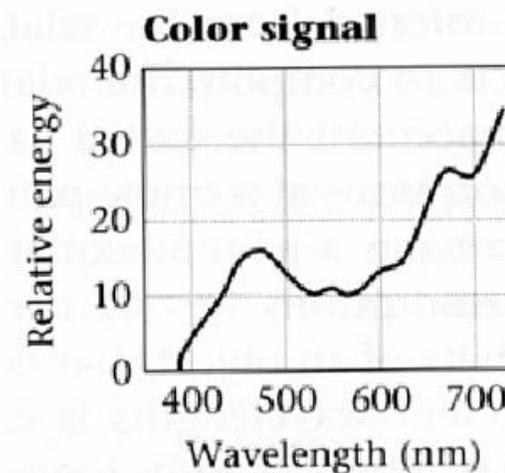
Отраженный свет это результат взаимодействия излучаемого света и поверхности



• *



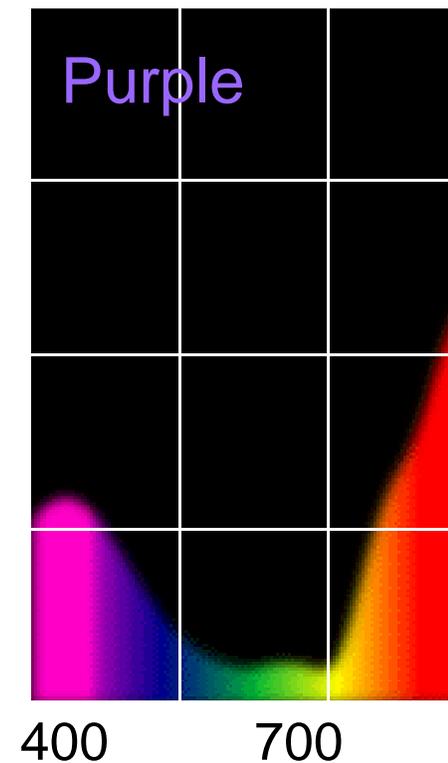
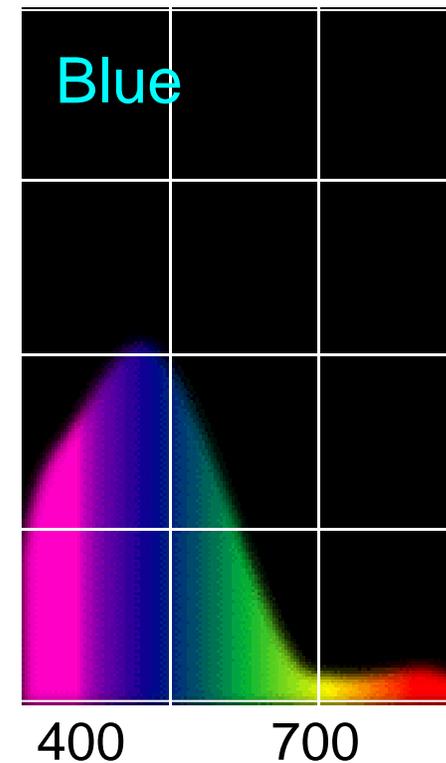
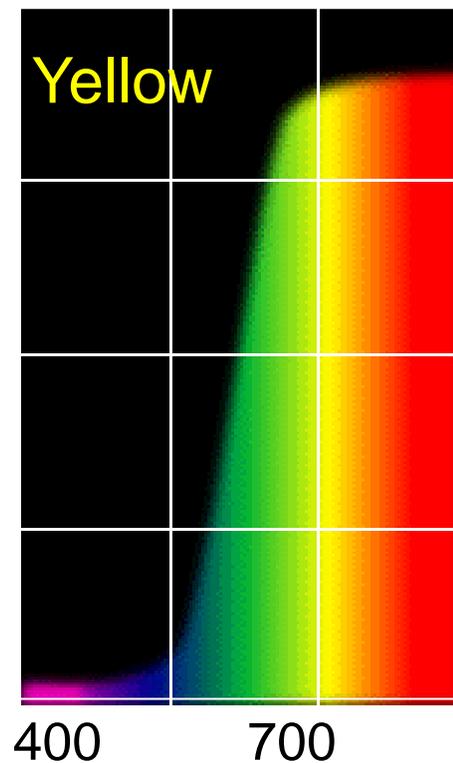
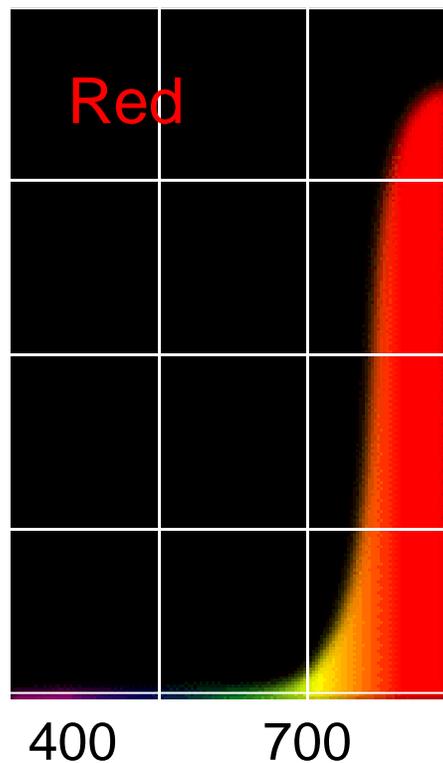
=



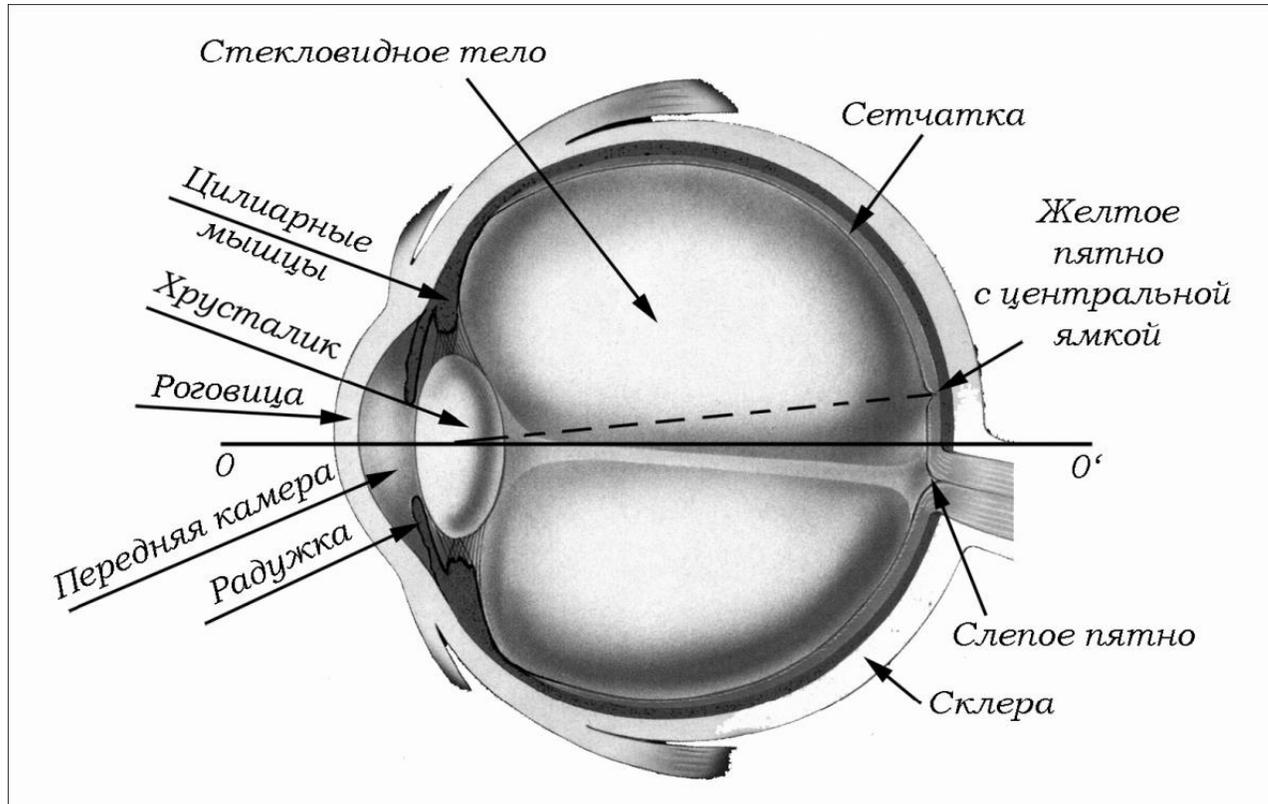
Примеры спектров отражённого света



% Отраженного света

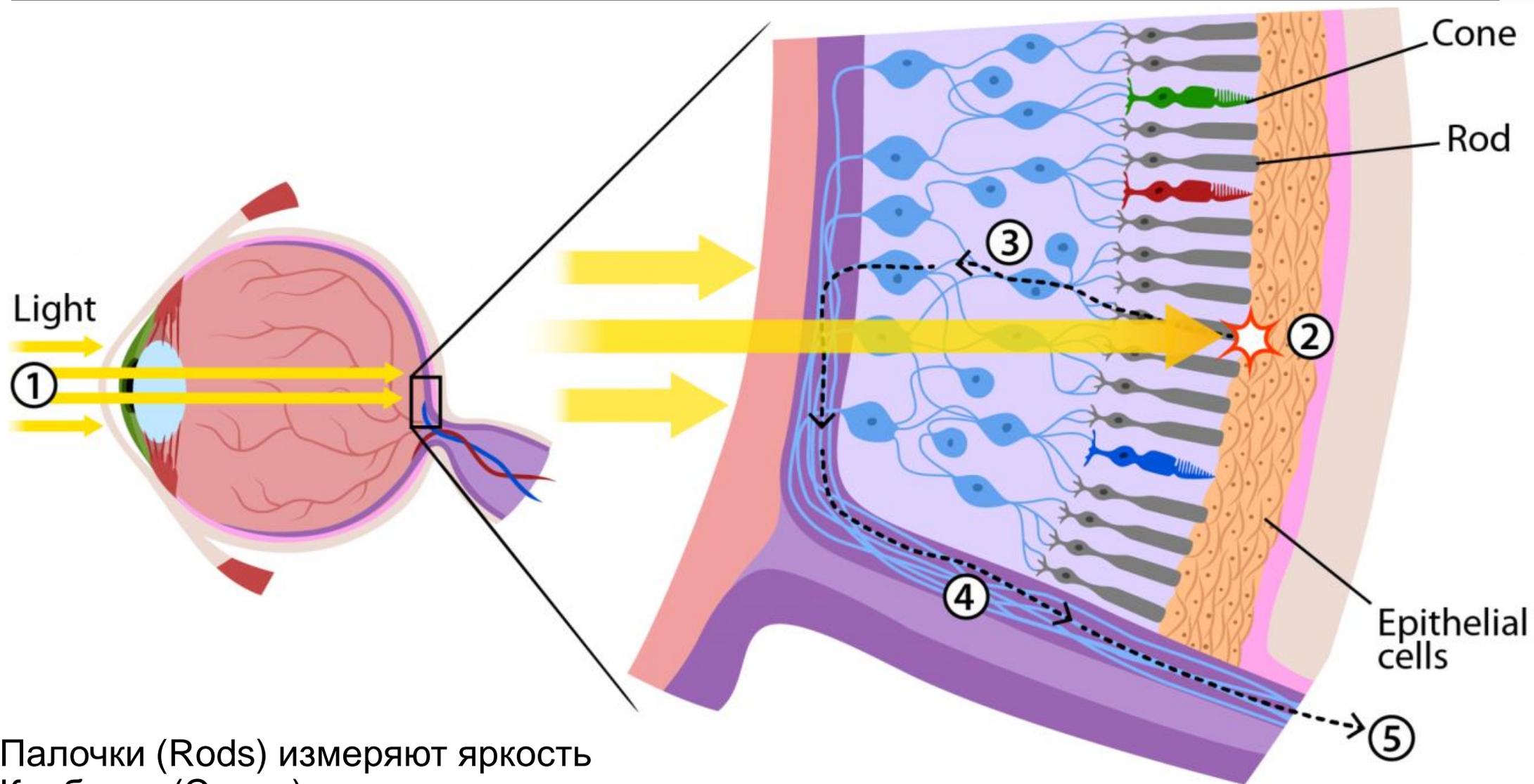


Человеческий глаз как камера



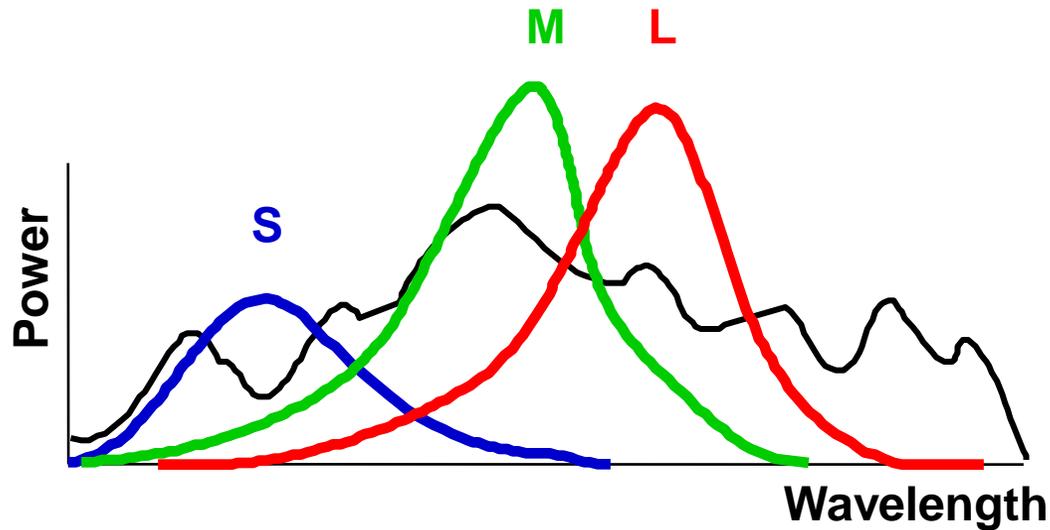
- **Радужка** – цветная пленка с радиальными мышцами
- **Зрачок** - отверстие (апертура), диаметр управляется радужкой
- **Хрусталик** – «линза», меняющая форму под действием мышц
- Где матрица?
 - Клетки-фоторецепторы на сетчатке

Сетчатка глаза



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

Восприятие цвета

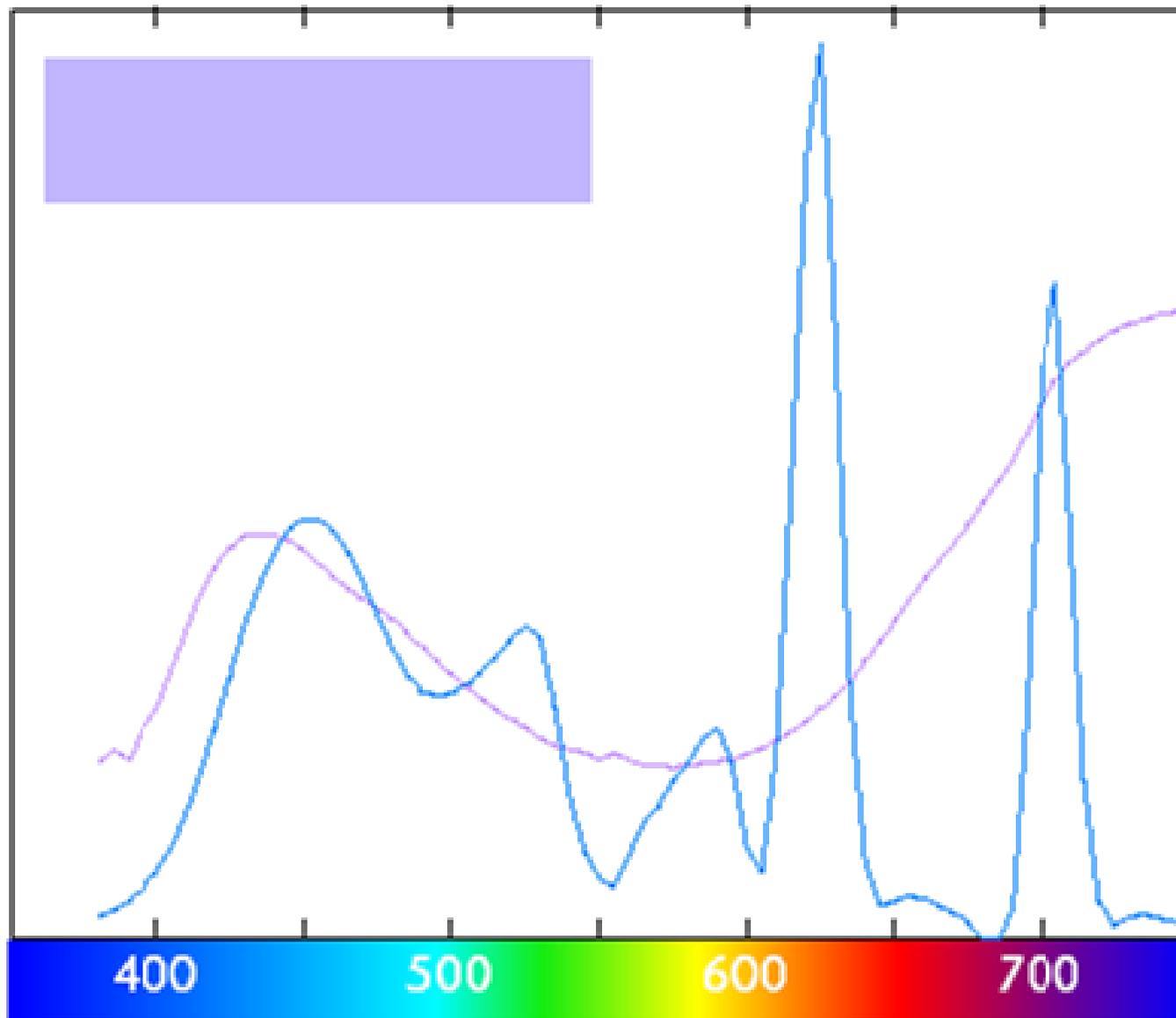


Палочки и колбочки – фильтры спектра

- 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**

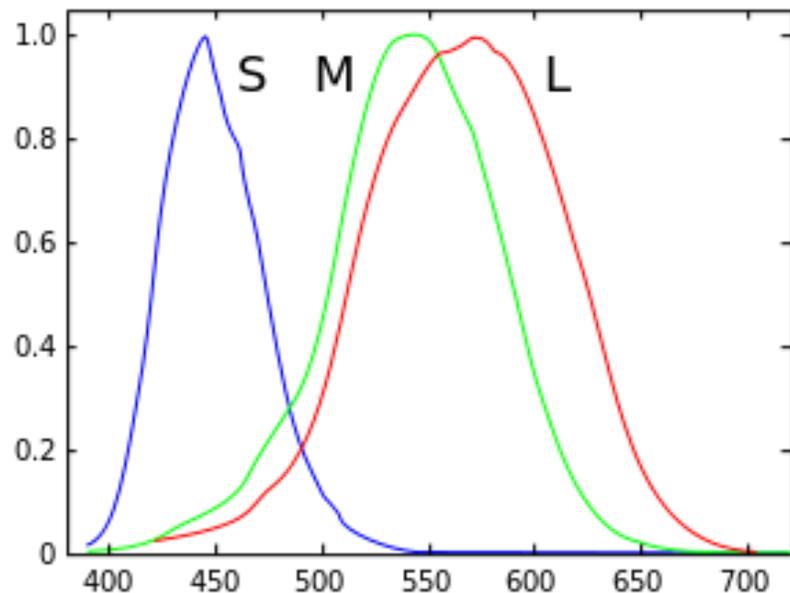
Метамеры фиолетового





Модели цвета

Модель LMS



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

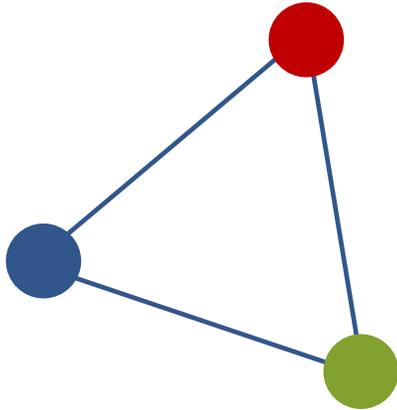
- L(long)
- M(medium)
- S(short)

- 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 P_1, P_2, P_3
- Остальные цвета задаются линейной комбинацией базовых цветов (primaries). Веса – «координаты» цвета

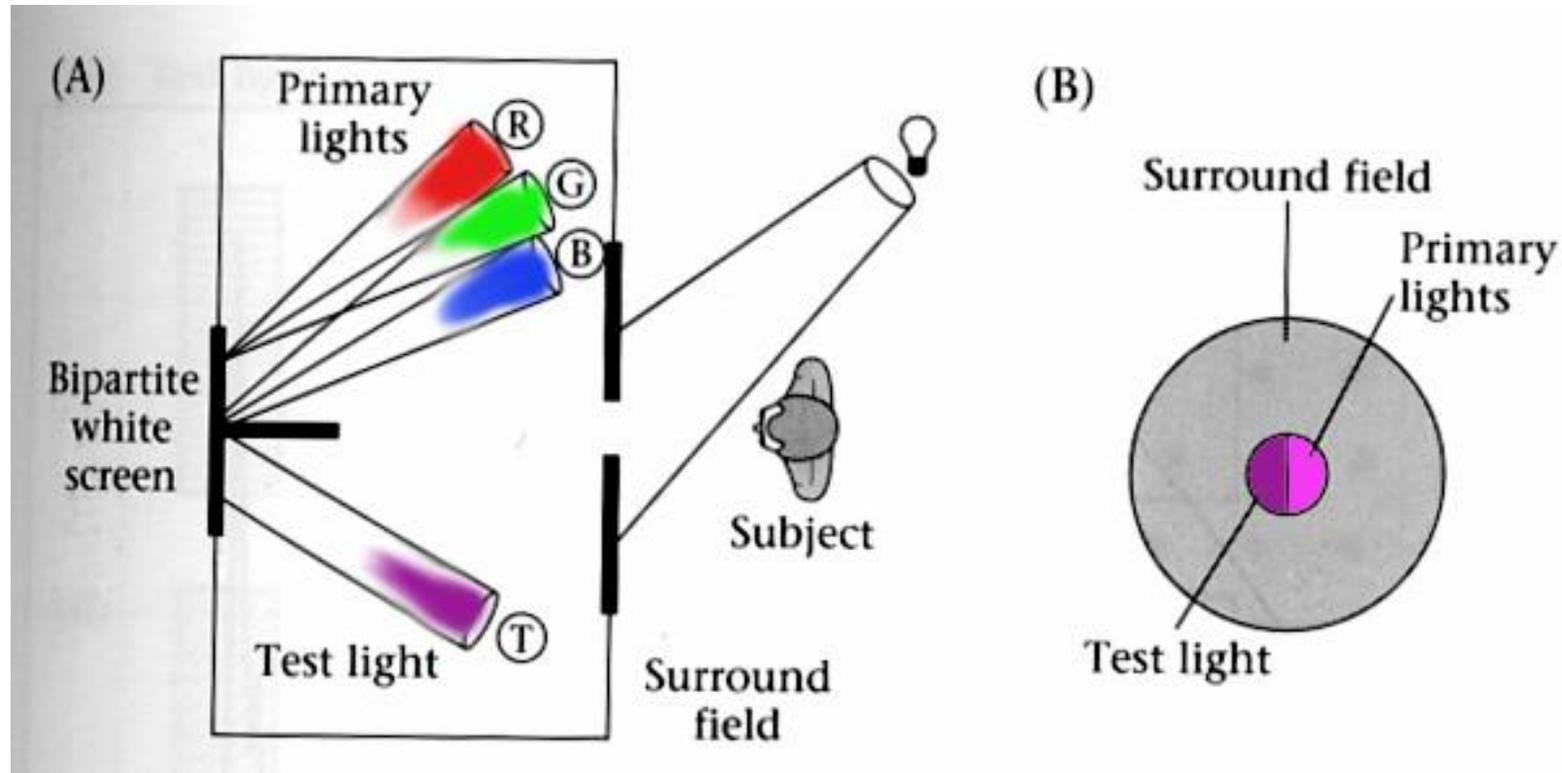


$$A = u_1 P_1 + u_2 P_2 + u_3 P_3$$

Смешение трех цветов

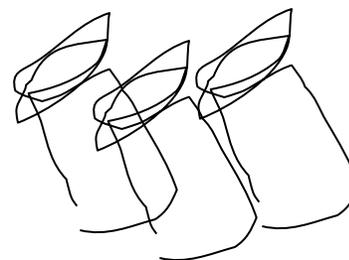
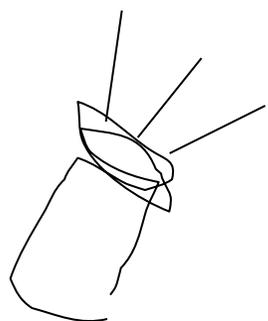
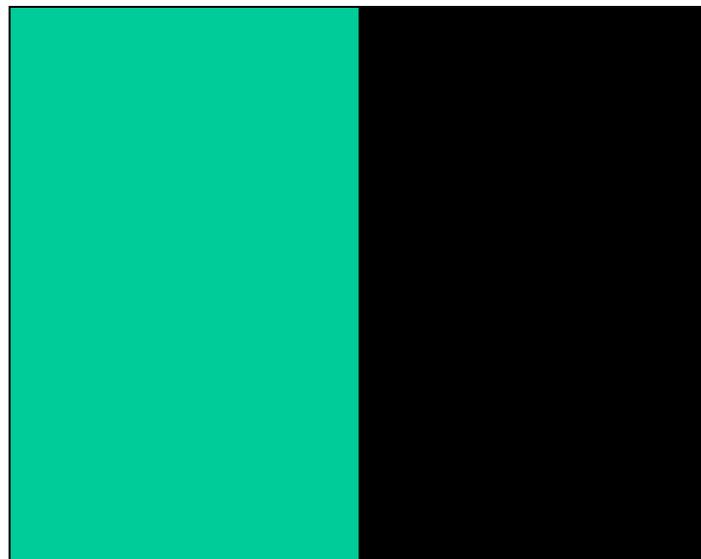
- Трихроматическая теория гласит, что все видимые цвета можем получить комбинацией трёх базовых цветов
- Можем делать эксперименты по сопоставлению цветов

Сопоставление цветов

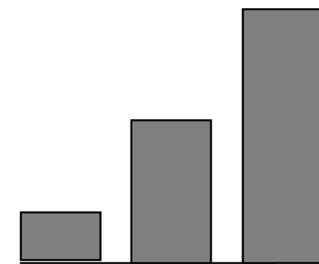
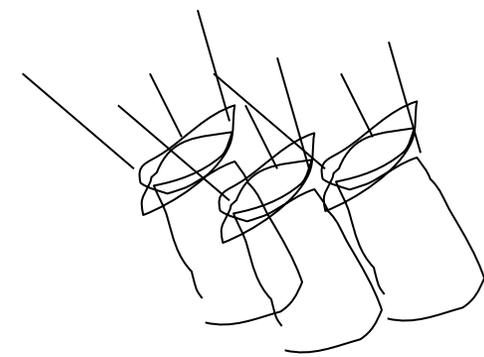
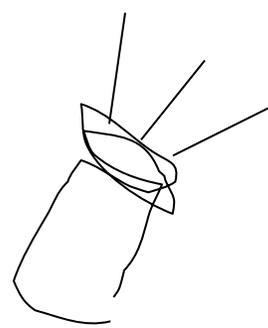
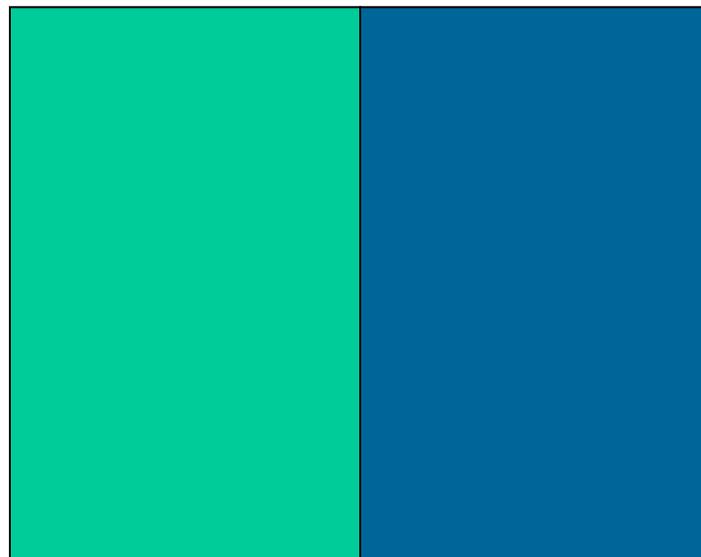


Берём «тестовый» источник и пробуем сопоставить его комбинацией базовых цветов (источников света)

Эксперимент №1

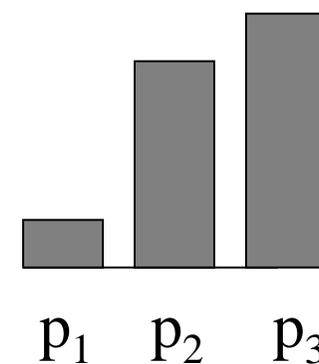
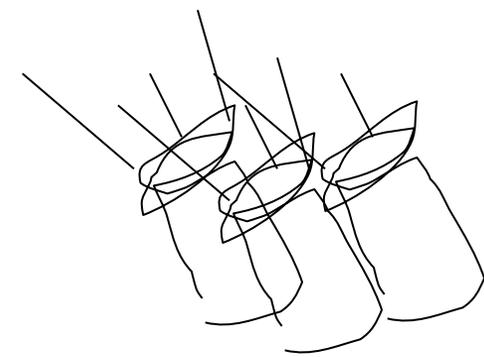
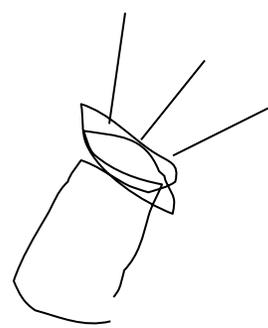
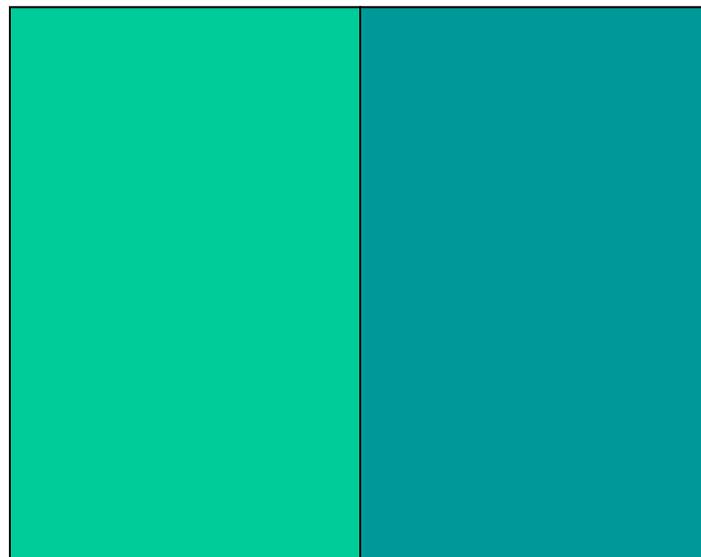


Эксперимент №1

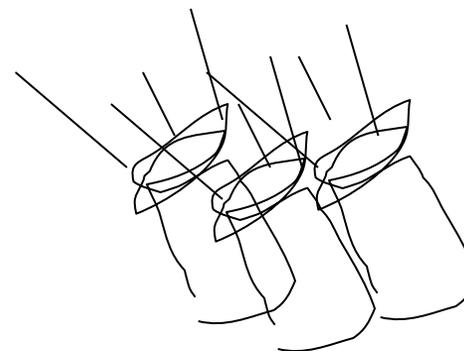
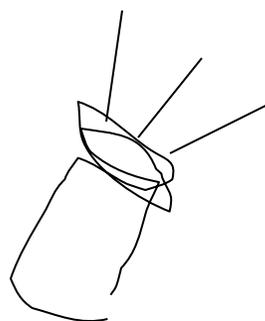
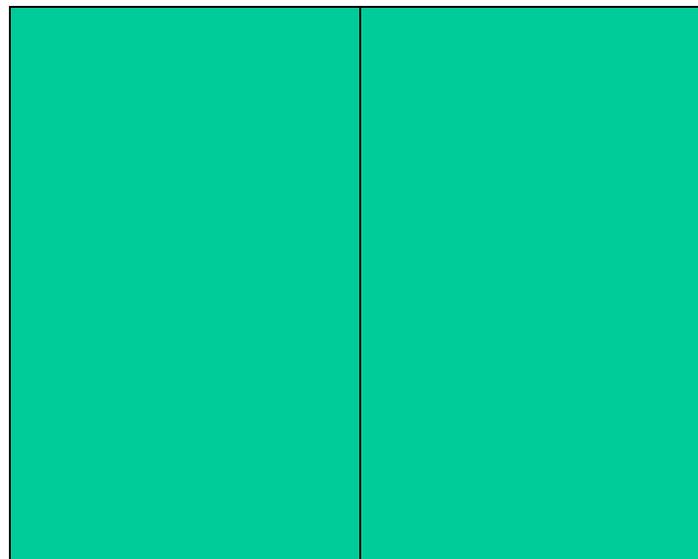


p_1 p_2 p_3

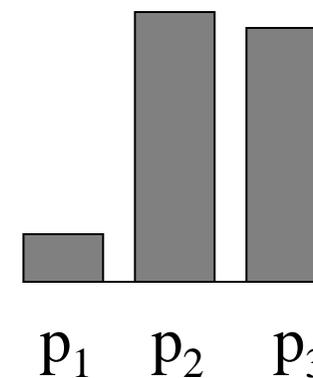
Эксперимент №1



Эксперимент №1



«Координаты»
света в нашей
модели





Закон аддитивности Грассмана (1853)

- Сопоставление цветов линейное
- Если два источника света сопоставляются одинаковыми весами базовых источников, то они воспринимаются одинаково
 - Пусть $A = u_1P_1 + u_2P_2 + u_3P_3$ и $B = u_1P_1 + u_2P_2 + u_3P_3$
 - Тогда $A = B$
- Если мы смешиваем два источника, тогда смешение соответствующих им базовым источников будет восприниматься также
 - Пусть $A = u_1P_1 + u_2P_2 + u_3P_3$ и $B = v_1P_1 + v_2P_2 + v_3P_3$
 - Тогда $A + B = (u_1 + v_1)P_1 + (u_2 + v_2)P_2 + (u_3 + v_3)P_3$
- Если мы увеличиваем яркость источника света, тогда яркость соответствующих базовых светов должна быть увеличена на такое же значение
 - Пусть $A = u_1P_1 + u_2P_2 + u_3P_3$
 - Тогда $kA = k(u_1)P_1 + (ku_2)P_2 + (ku_3)P_3$

Модель CIE RGB 1931

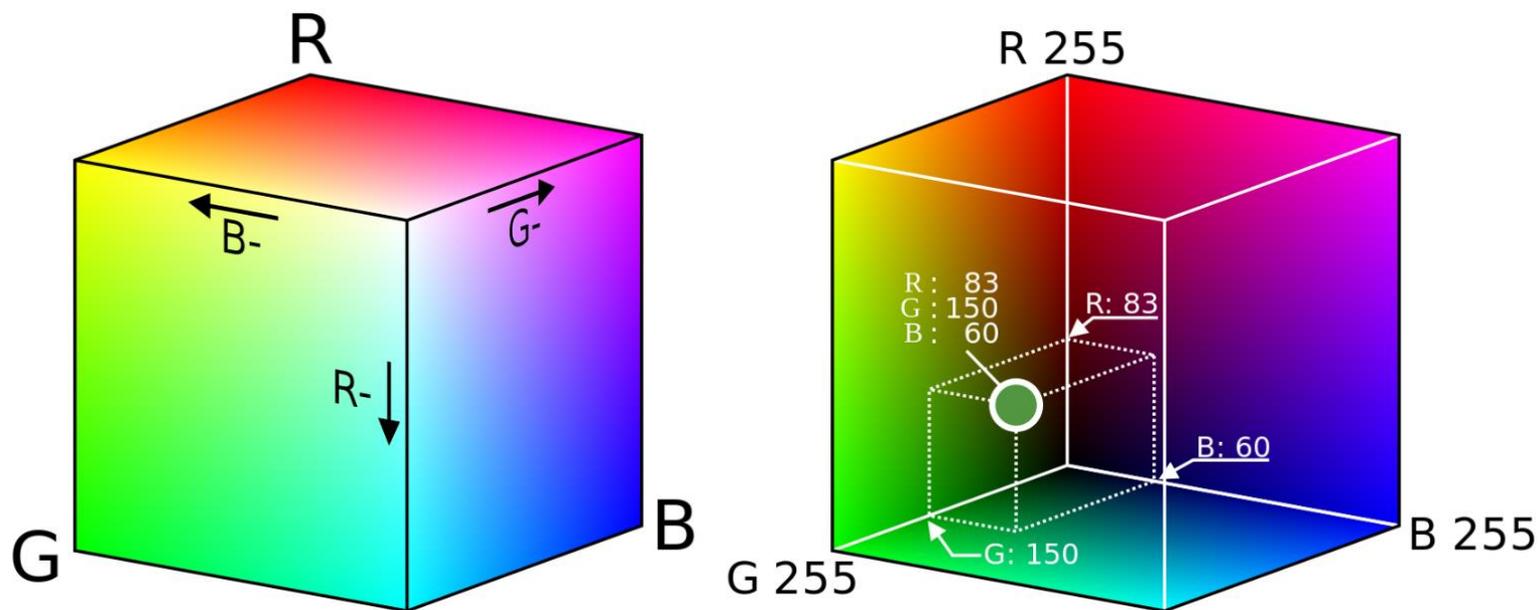


- Выберем три монохроматических цвета в качестве основных на основании информации о чувствительности колбочек
- CIE RGB 1931 от International Commission on Illumination (CIE)

 $p_1 = 645.2 \text{ nm}$

 $p_2 = 525.3 \text{ nm}$

 $p_3 = 444.4 \text{ nm}$



Цветовой куб - множество цветов, задаваемых цветовой моделью RGB

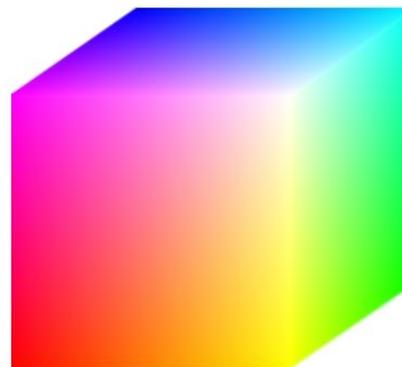
Модель RGB



 $p_1 = 645.2 \text{ nm}$

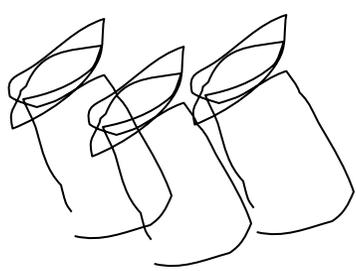
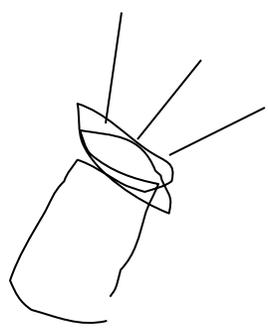
 $p_2 = 525.3 \text{ nm}$

 $p_3 = 444.4 \text{ nm}$

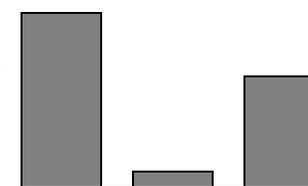
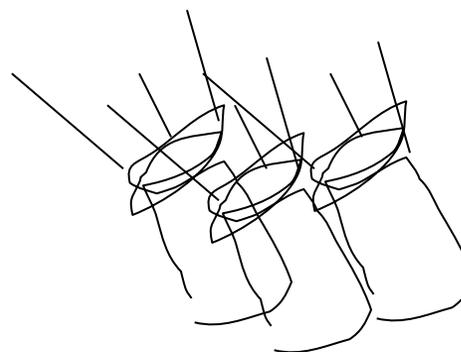
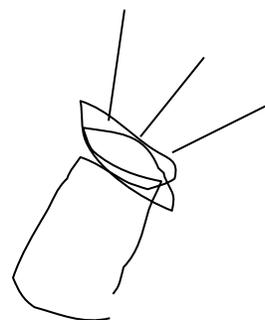
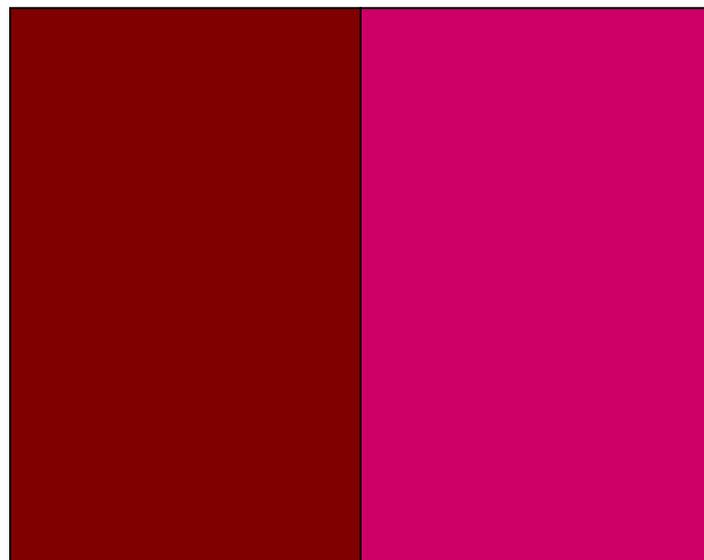


- Описывает ли модель RGB всё множество видимых человеком цветов?
- Для проверки проведём эксперимент и попробуем сопоставить каждой длине волны видимого света (когерентному источнику света) цвет в модели RGB
- *Функции сопоставления* - веса, необходимые для сопоставления с когерентными источниками света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

Эксперимент №2

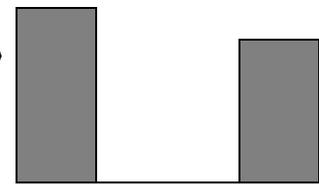
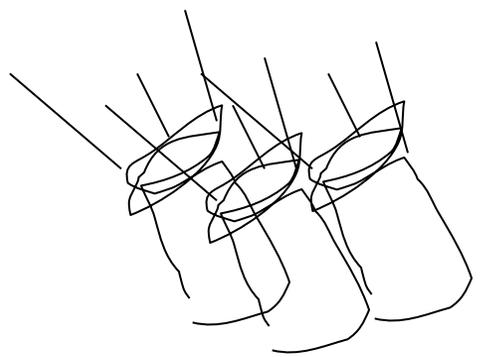
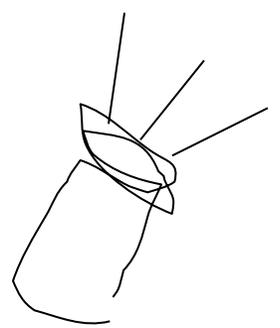
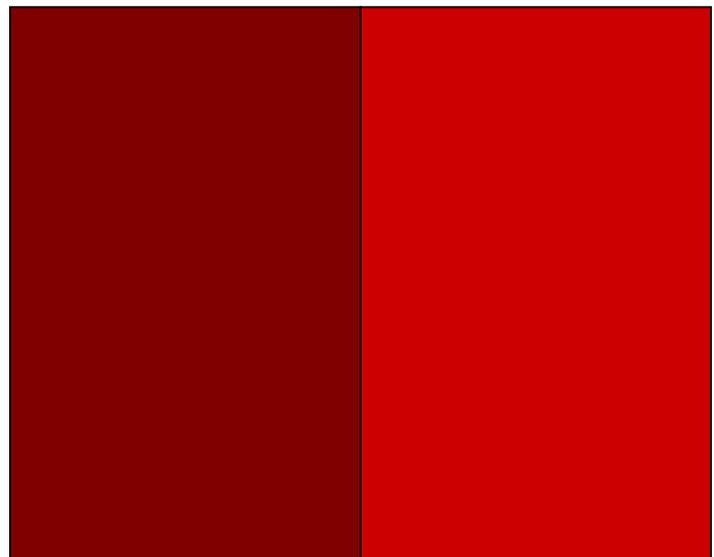


Эксперимент №2



p_1 p_2 p_3

Эксперимент №2

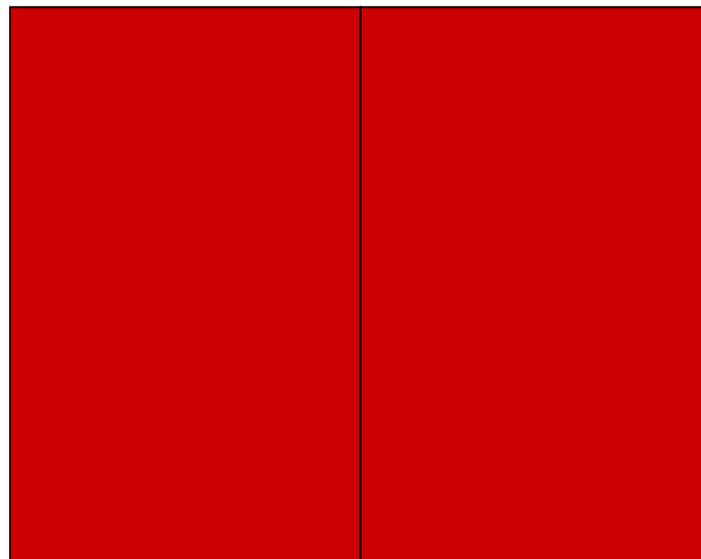


p_1 p_2 p_3

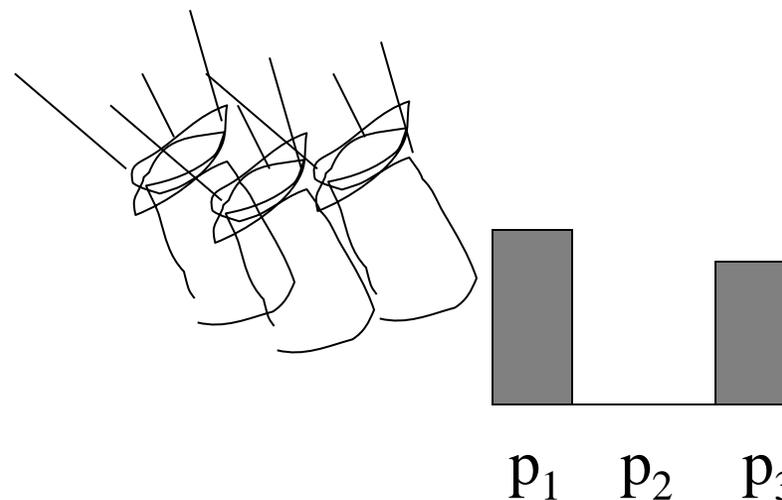
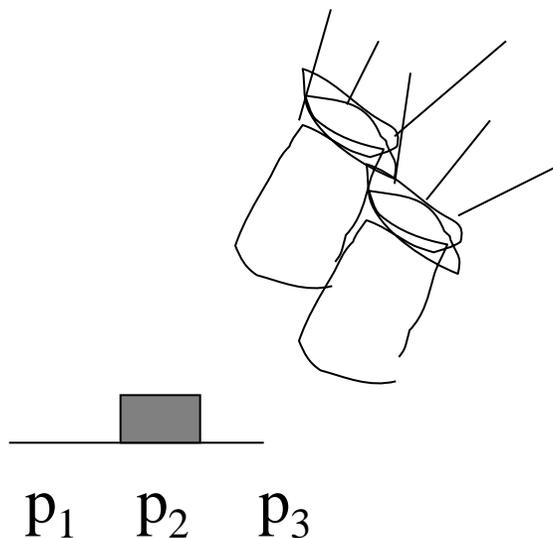
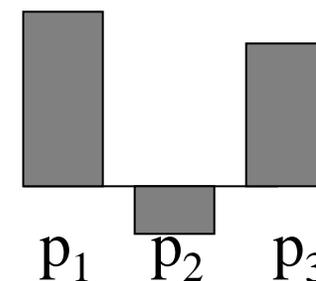
Эксперимент №2



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



Веса основных цветов, необходимых для сопоставления:





Линейная цветовая модель RGB

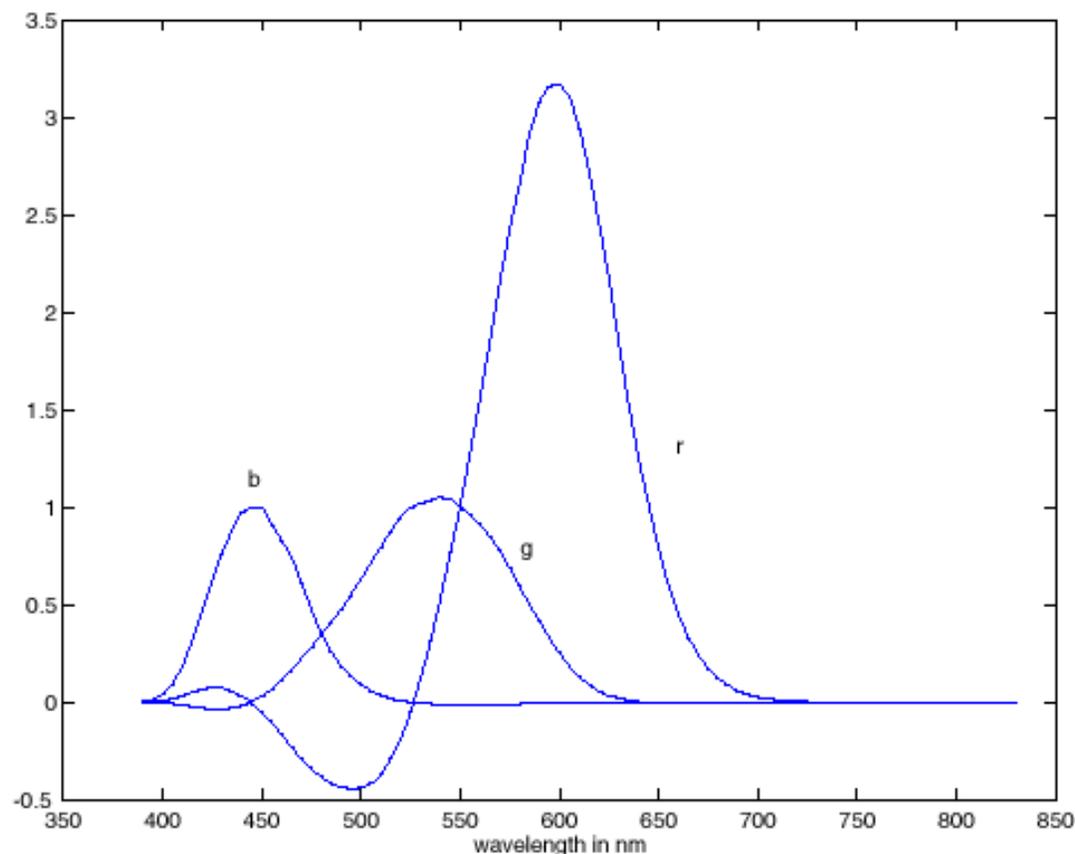
 $p_1 = 645.2 \text{ nm}$

 $p_2 = 525.3 \text{ nm}$

 $p_3 = 444.4 \text{ nm}$

Модель RGB не полностью описывает всё множество видимых цветов

«Вычитание» необходимо для соответствия некоторым длинам волны



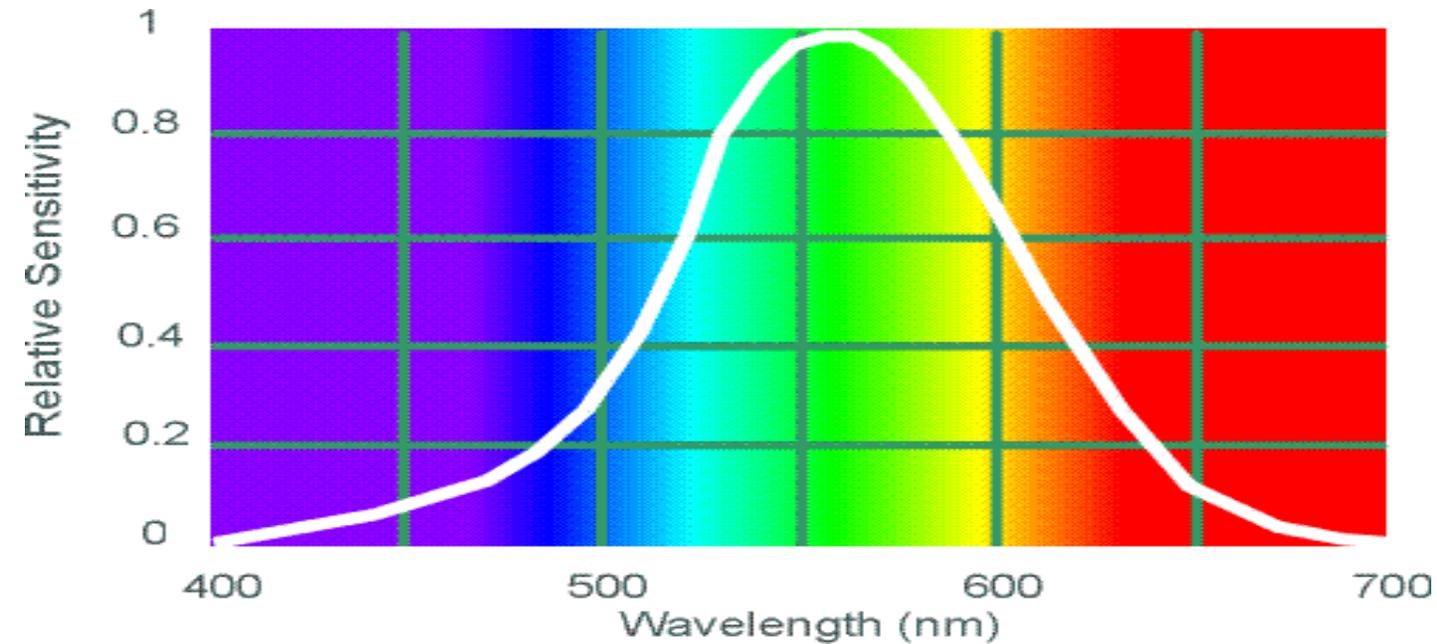
Функции сопоставления для RGB

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 YIQ



Пример модели с отдельной яркостью

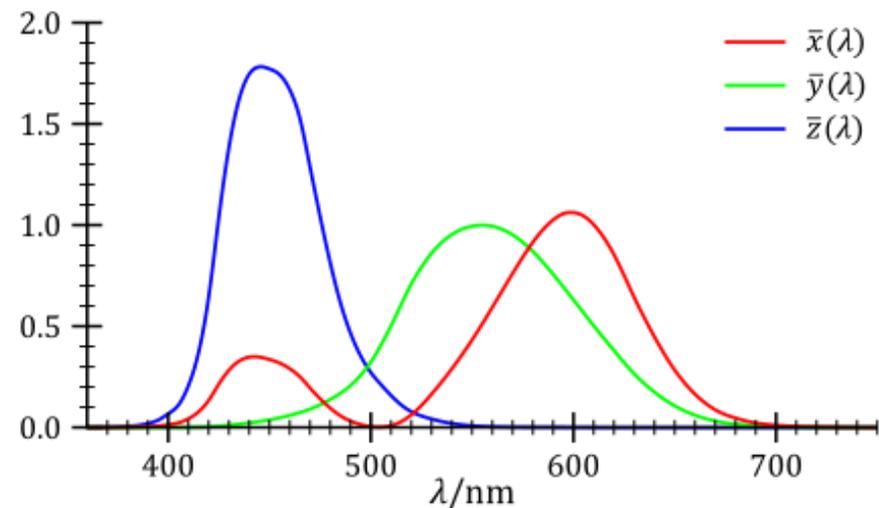
- $Y = .299R + .587G + .114B$
- $I = .596R - .275G - .321B$
- $Q = .212R - .528G + .311B$
- $R = 1.000 Y + 0.956 I + 0.621 Q$
- $G = 1.000 Y - 0.272 I - 0.647 Q$
- $B = 1.000 Y - 1.106 I + 1.703 Q$
- Цветовая модель YIQ используется в коммерческом цветном телевидении США
- Модель YIQ совместима с черно-белым телевидением
- Модель YIQ используется в стандарте JPEG
- $I = R - C$; $Q = M - G$

CIE 1931 (XYZ)



- Хотим получить линейную аддитивную модель XYZ:

- покрывает всё множество видимых цветов
- Y должен соответствовать видимой яркости света
- X, Y описываются цветность
- Функции сопоставления должны быть неотрицательными



Функции сопоставления для XYZ

- 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 ∞

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

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

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



Перевод из XYZ в RGB

- XYZ, RGB – линейные трихроматические модели
- Из одной модели в другую можем перейти с помощью линейного преобразования

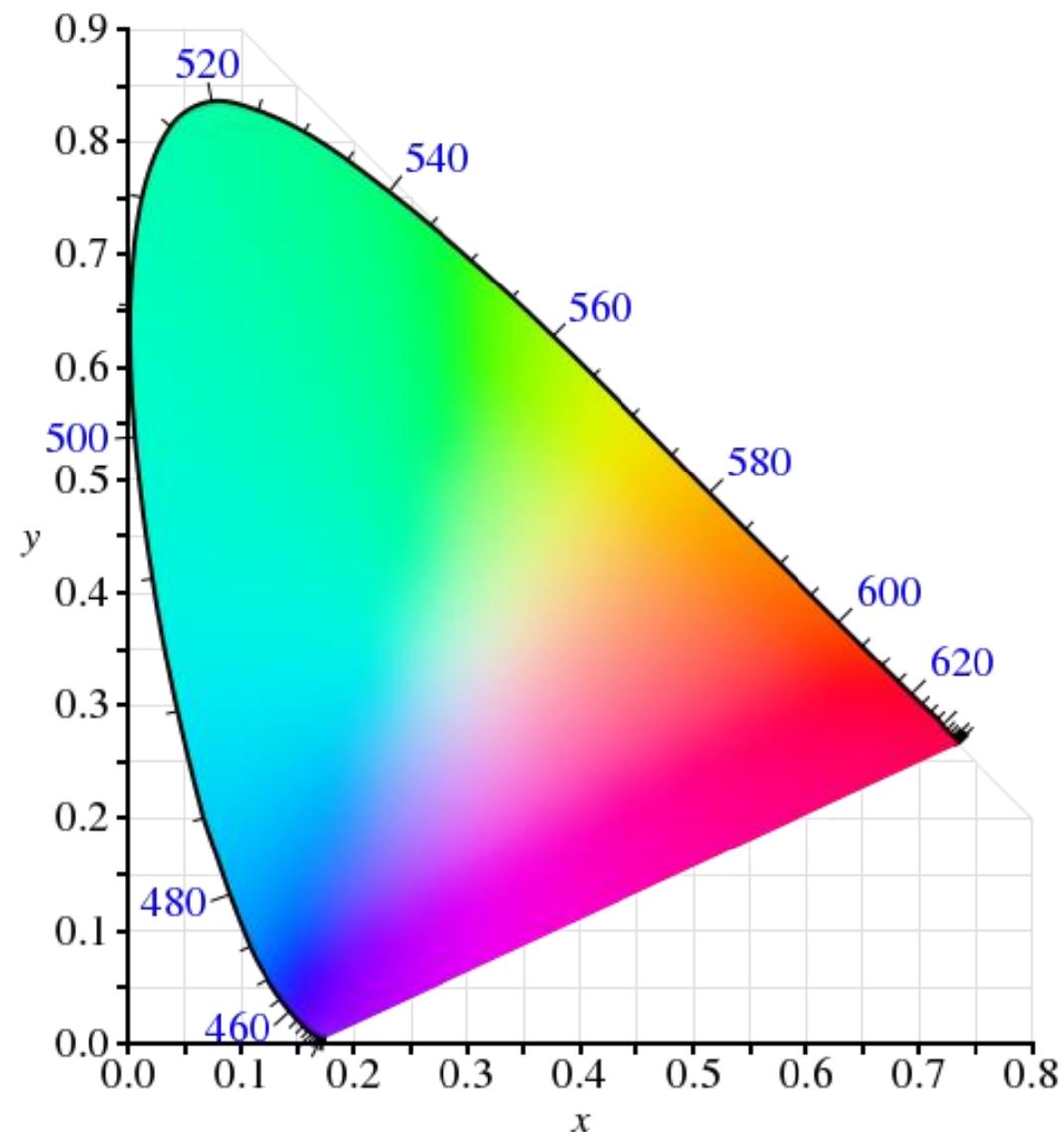
$$\begin{bmatrix} R_{\text{linear}} \\ G_{\text{linear}} \\ B_{\text{linear}} \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

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

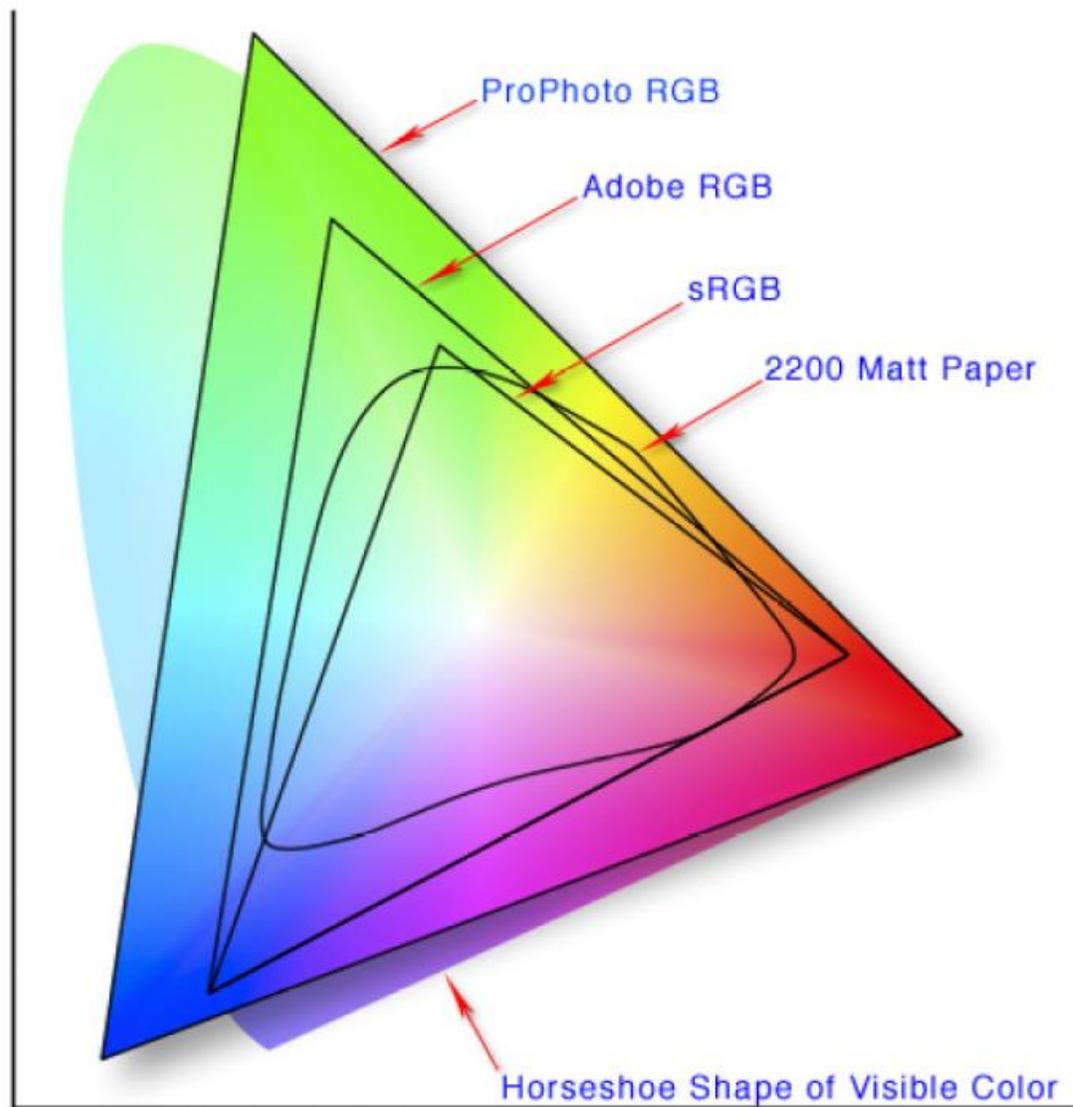
CIE xyY и диаграмма цветов (Gamut)



- 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 RGB





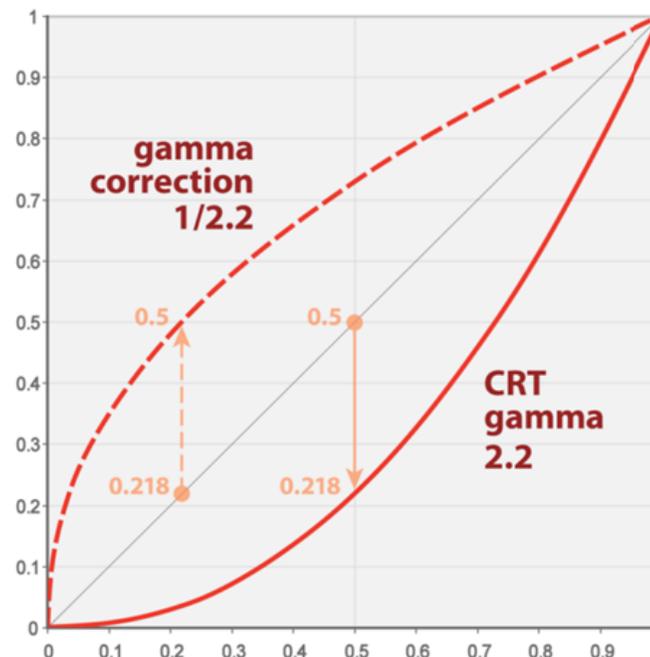
Нелинейность яркости



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 $y = c \cdot x^\gamma$



Модель sRGB



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

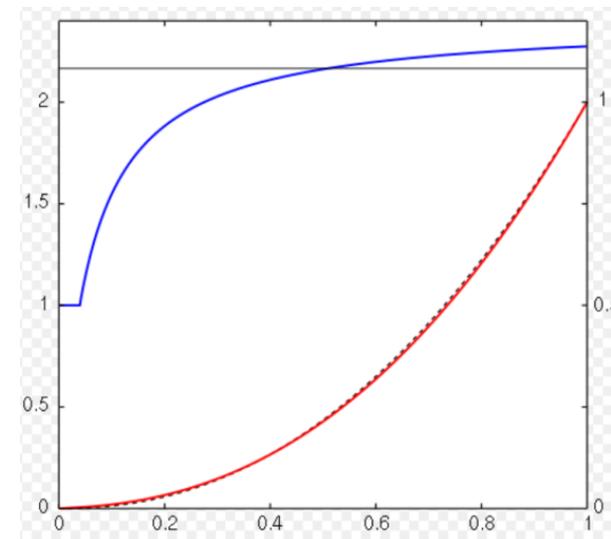
$$\begin{bmatrix} R_{\text{linear}} \\ G_{\text{linear}} \\ B_{\text{linear}} \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$C_{\text{srgb}} = \begin{cases} 12.92C_{\text{linear}}, & C_{\text{linear}} \leq 0.0031308 \\ (1 + a)C_{\text{linear}}^{1/2.4} - a, & C_{\text{linear}} > 0.0031308 \end{cases} \quad a=0.055$$

- Из sRGB и XYZ

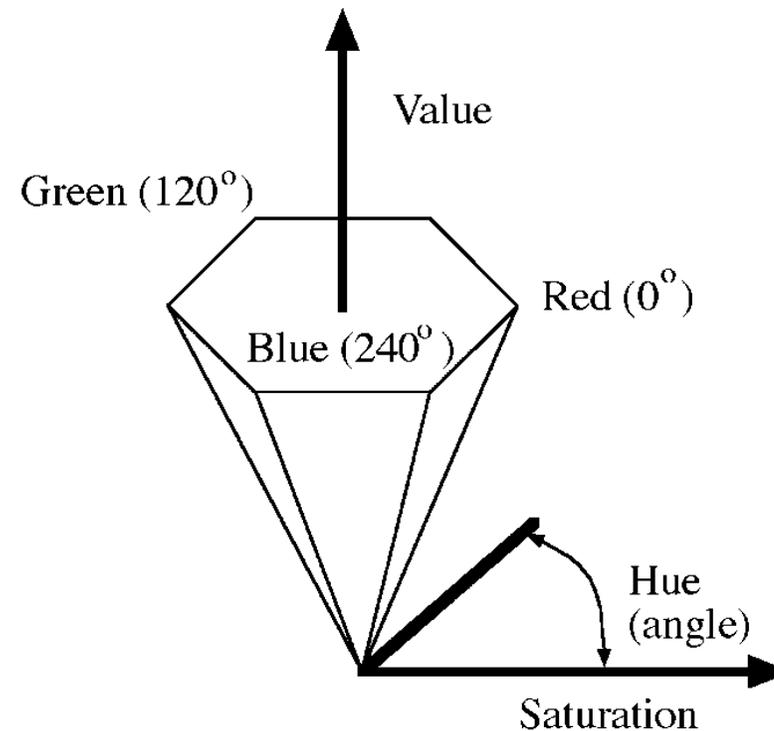
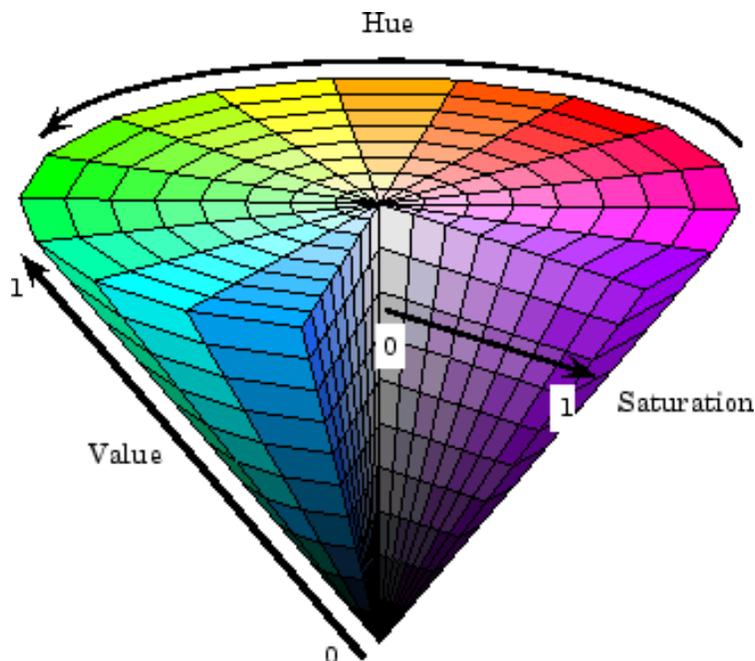
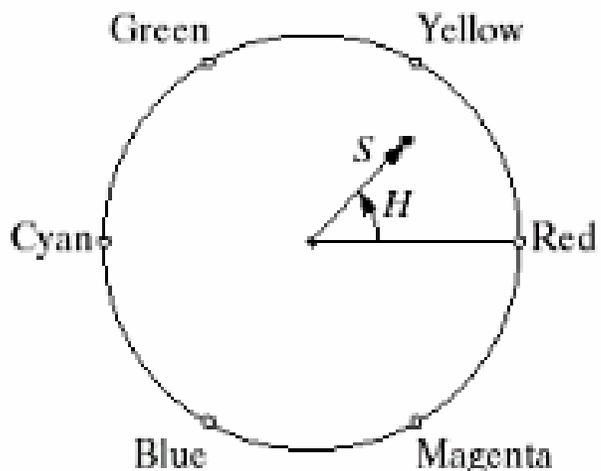
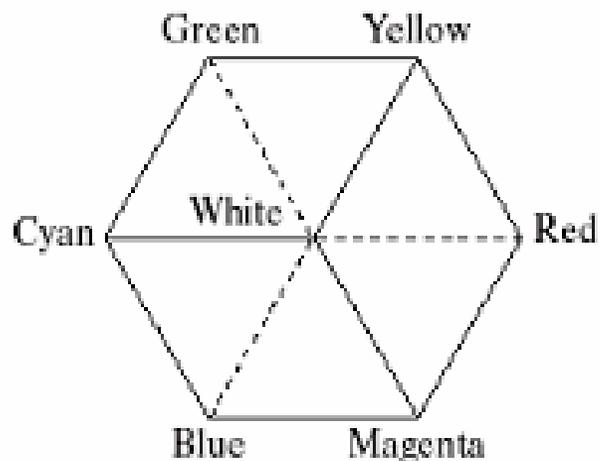
$$C_{\text{linear}} = \begin{cases} \frac{C_{\text{srgb}}}{12.92}, & C_{\text{srgb}} \leq 0.04045 \\ \left(\frac{C_{\text{srgb}} + a}{1 + a} \right)^{2.4}, & C_{\text{srgb}} > 0.04045 \end{cases}$$

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





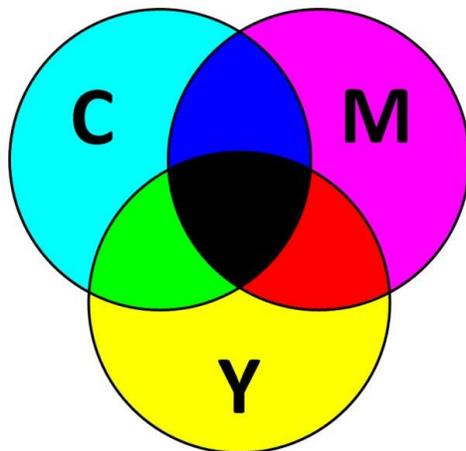
Модель HSV (HIS)



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

Является ли эта модель линейной?

Субтрактивная модель СМУК

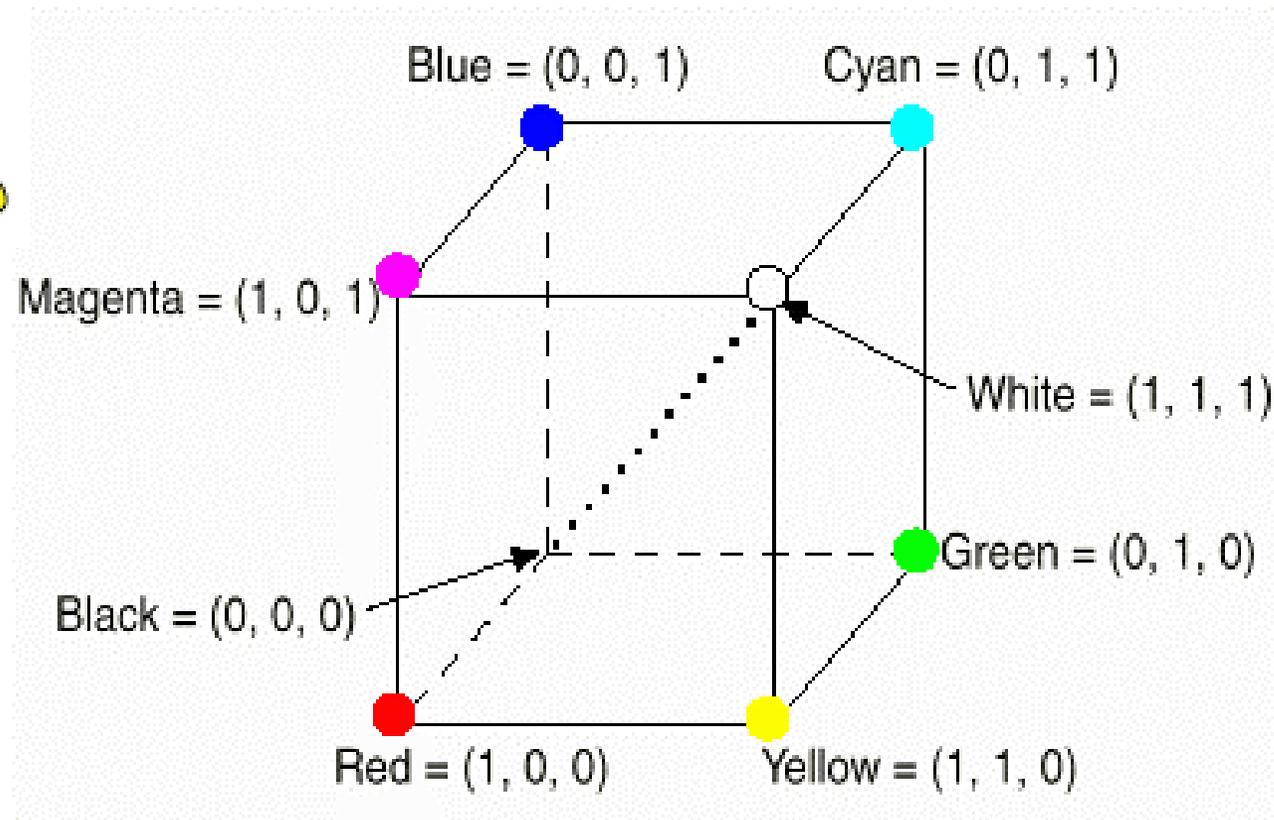


CMY

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

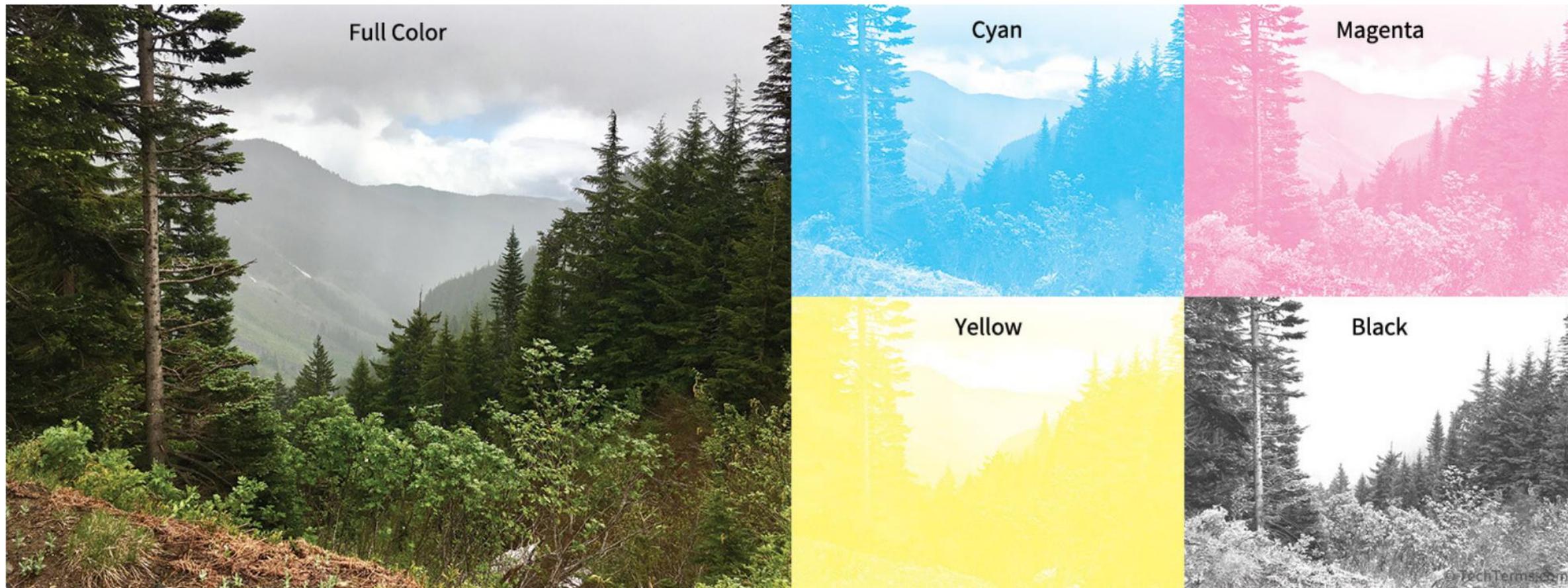
$$M = R + B = W - G$$

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



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

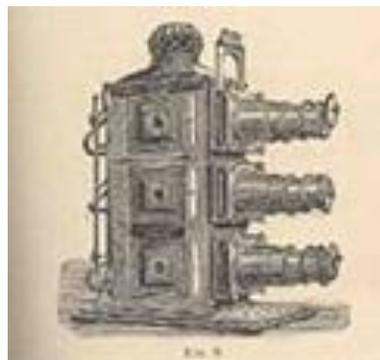
Красивый пример СМУК



Первые цветные фотографии



Отдельные каналы



Ламповый проектор



Цветное изображение после совмещения каналов

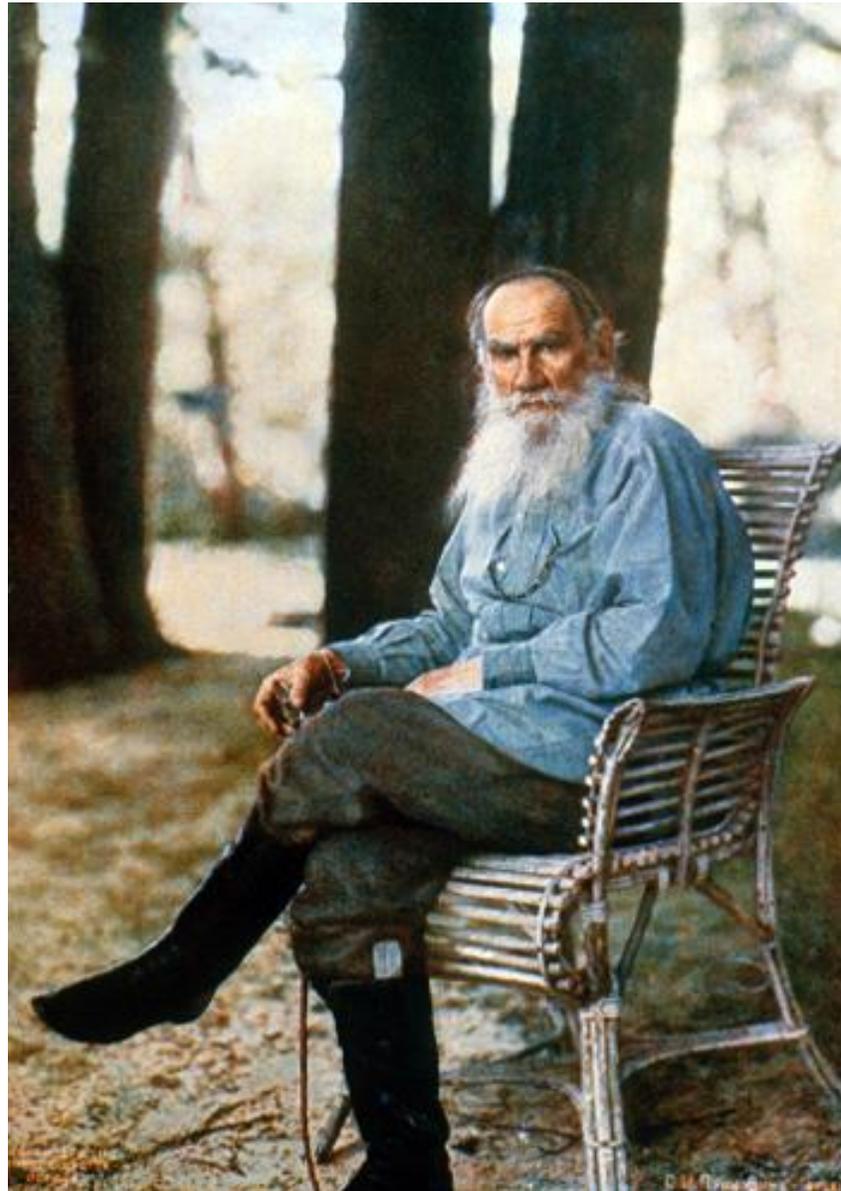


Сергей Прокудин-Горский (1863-1944)
Фотографии Российской империи (1909-1916)

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

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

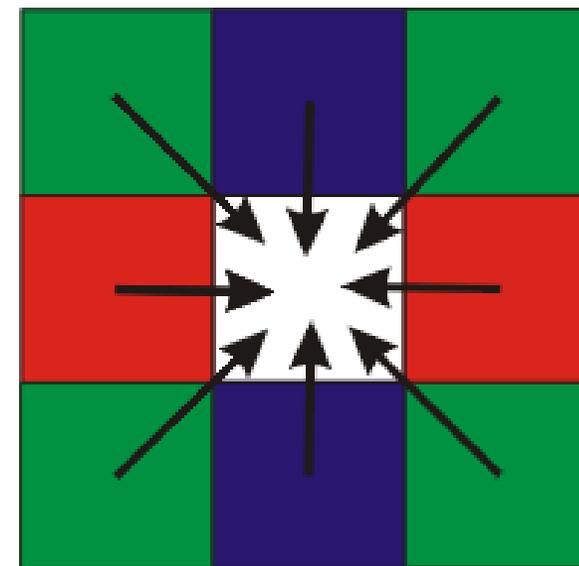
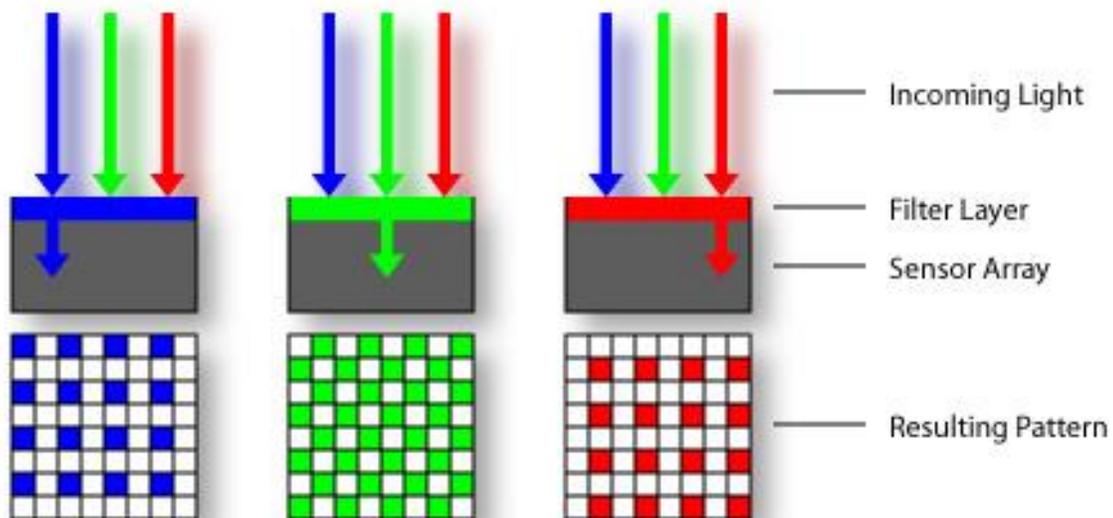
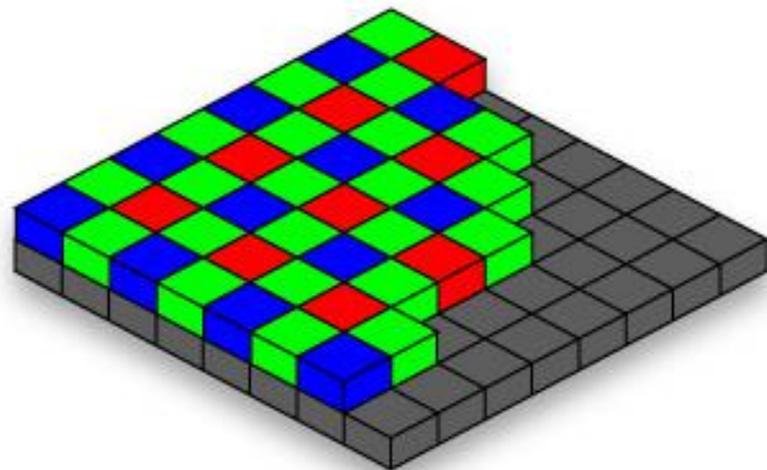
Лев Толстой



Цветное цифровое изображение

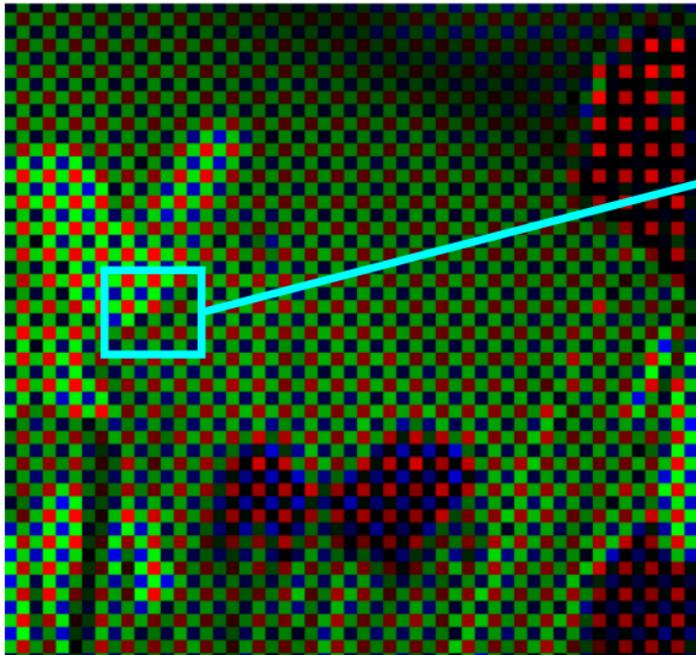


Байеровский шаблон

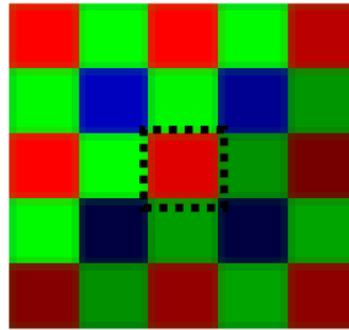


Демозаикинг (оценка пропущенных значений цвета)

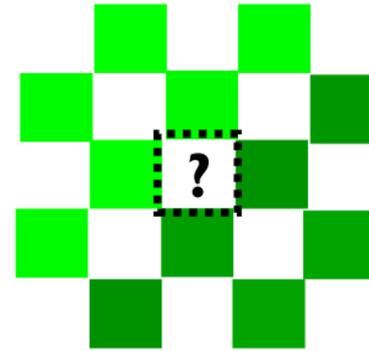
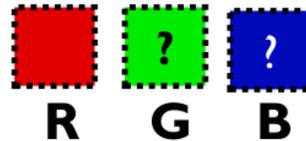
Демозаикинг на практике



Captured raw-Bayer image



Neighborhood about red pixel



Neighboring green values

| | | | | |
|-----|-----|-----|-----|-----|
| 0.8 | 0.8 | 0.8 | 0.4 | 0.2 |
| 0.8 | 0.9 | 0.9 | 0.3 | 0.2 |
| 0.7 | 0.9 | 1.0 | 0.3 | 0.2 |
| 0.5 | 0.3 | 0.3 | 0.2 | 0.2 |
| 0.1 | 0.2 | 0.2 | 0.2 | 0.2 |

Weight mask based on red pixel's similarity to neighboring red values.

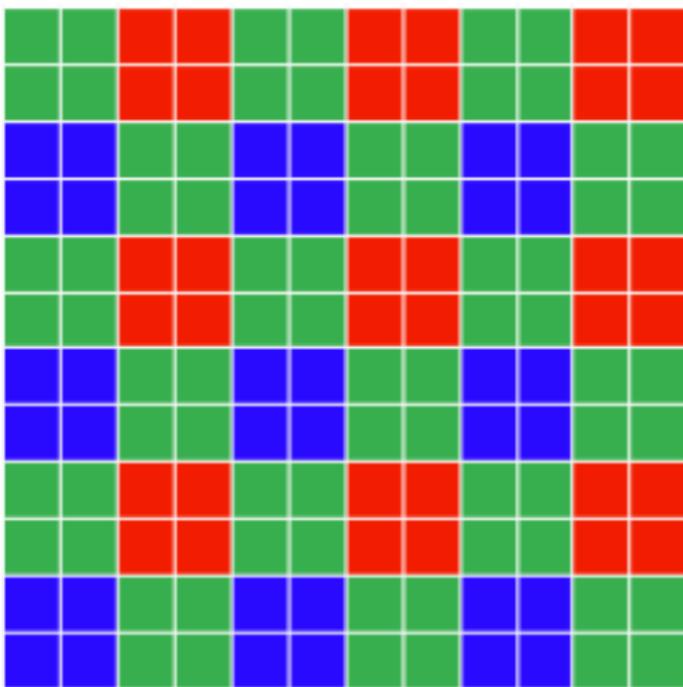
Missing green pixel value is computed as a weighted-interpolation of the neighboring green values.



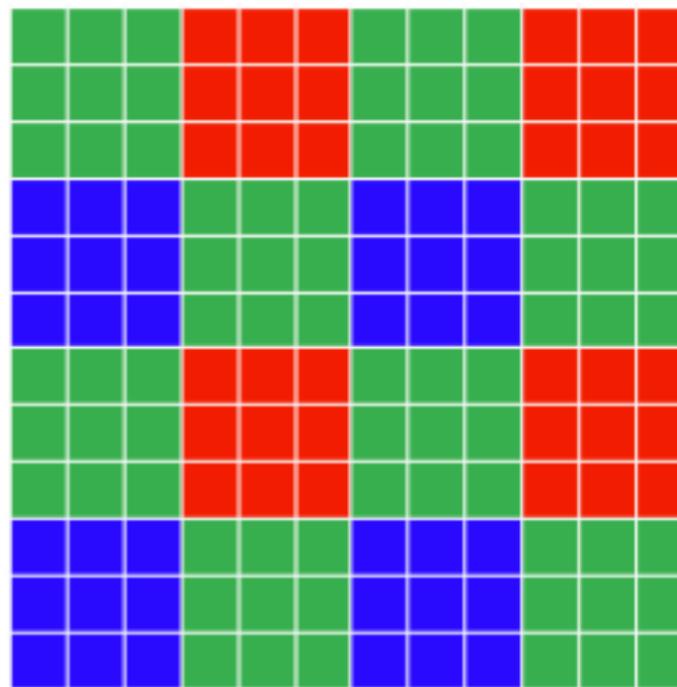
Демозаикинг



- Сейчас используют и другие паттерны, например, 2x2 и 3x3
- В условиях низкого освещения группы пикселей трактуют как один пиксель для расчёта весов интерполяции цветов



Tetra CFA

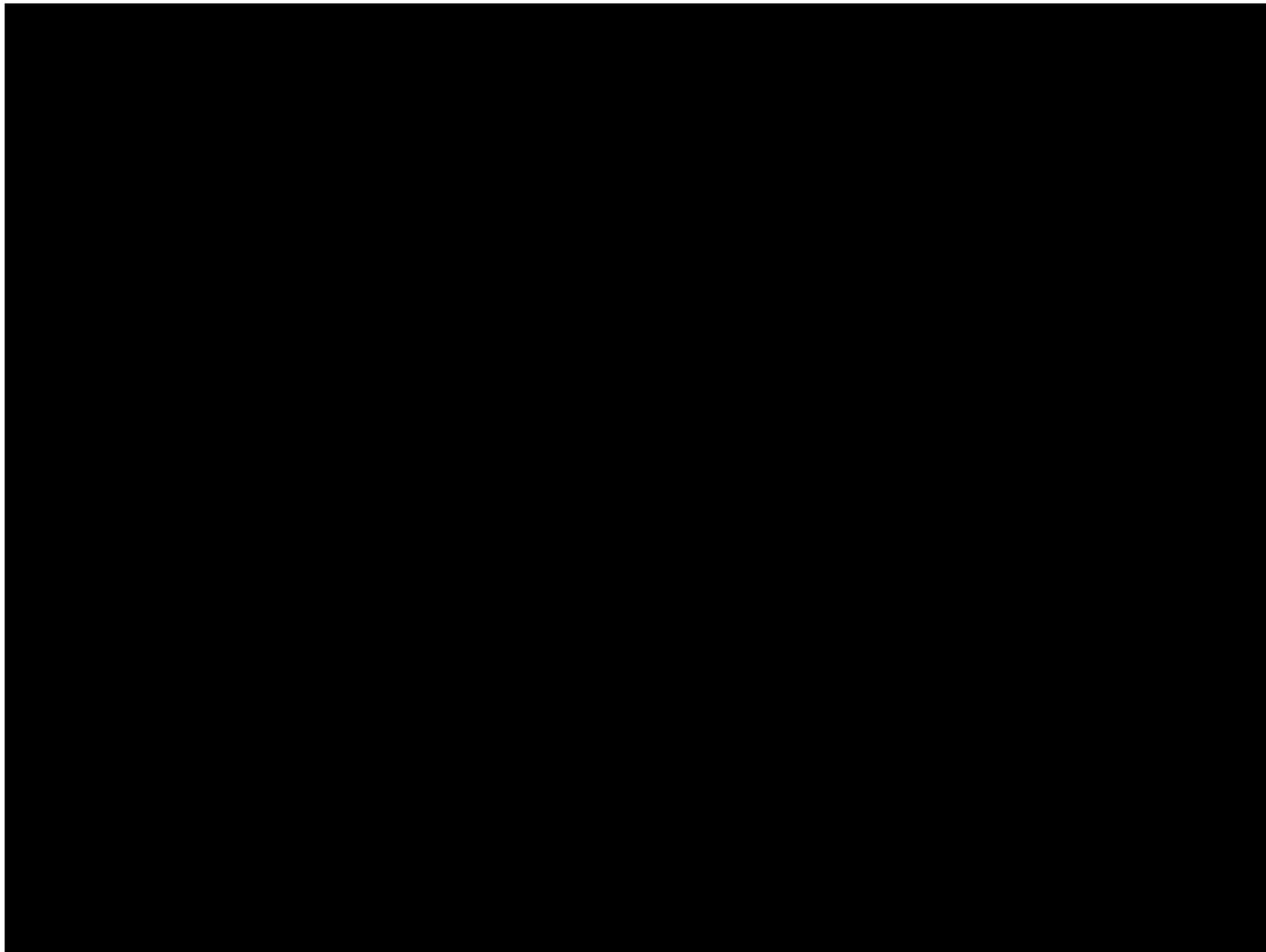


Nona CFA

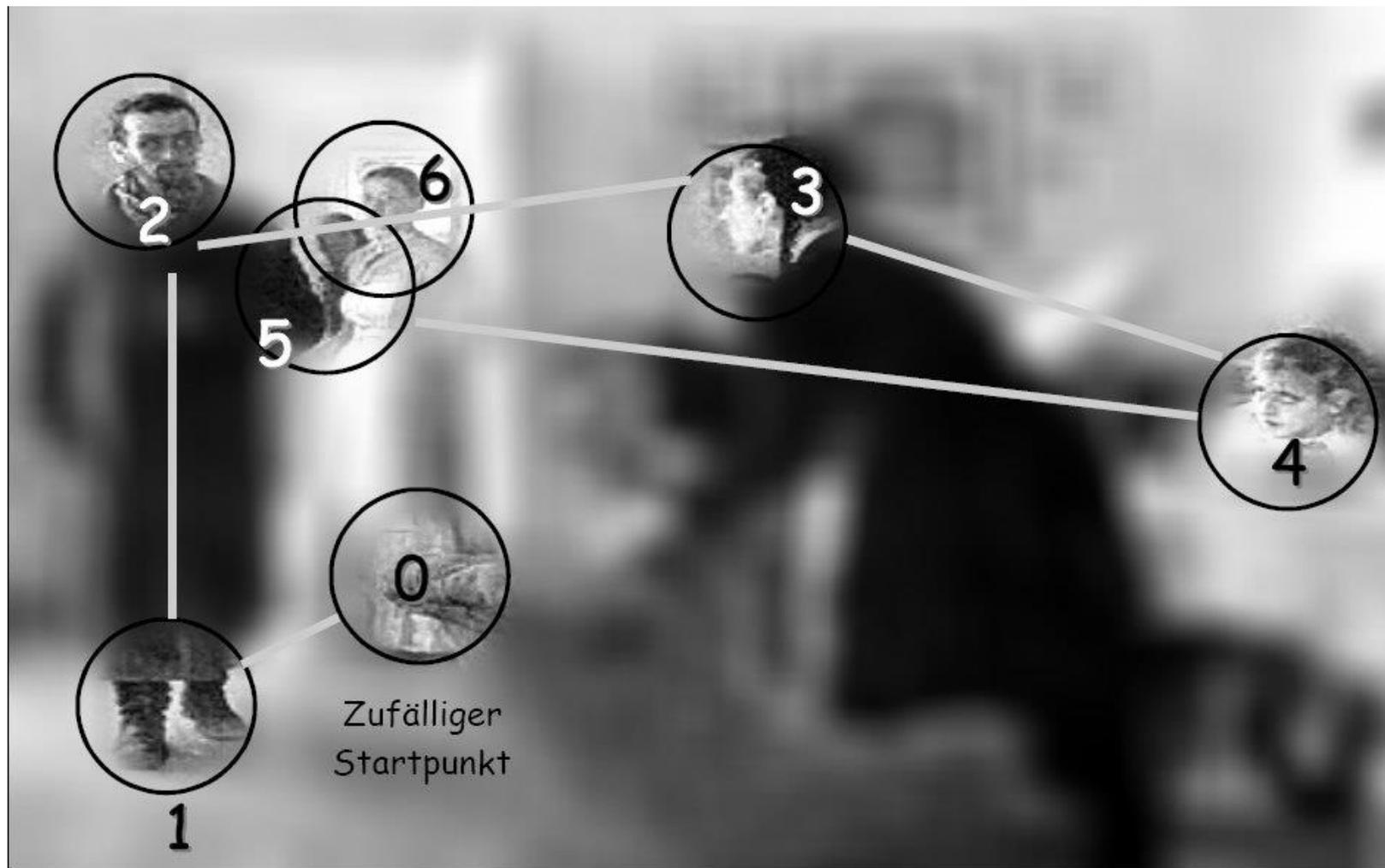


Особенности человеческого зрения

Тест на внимательность

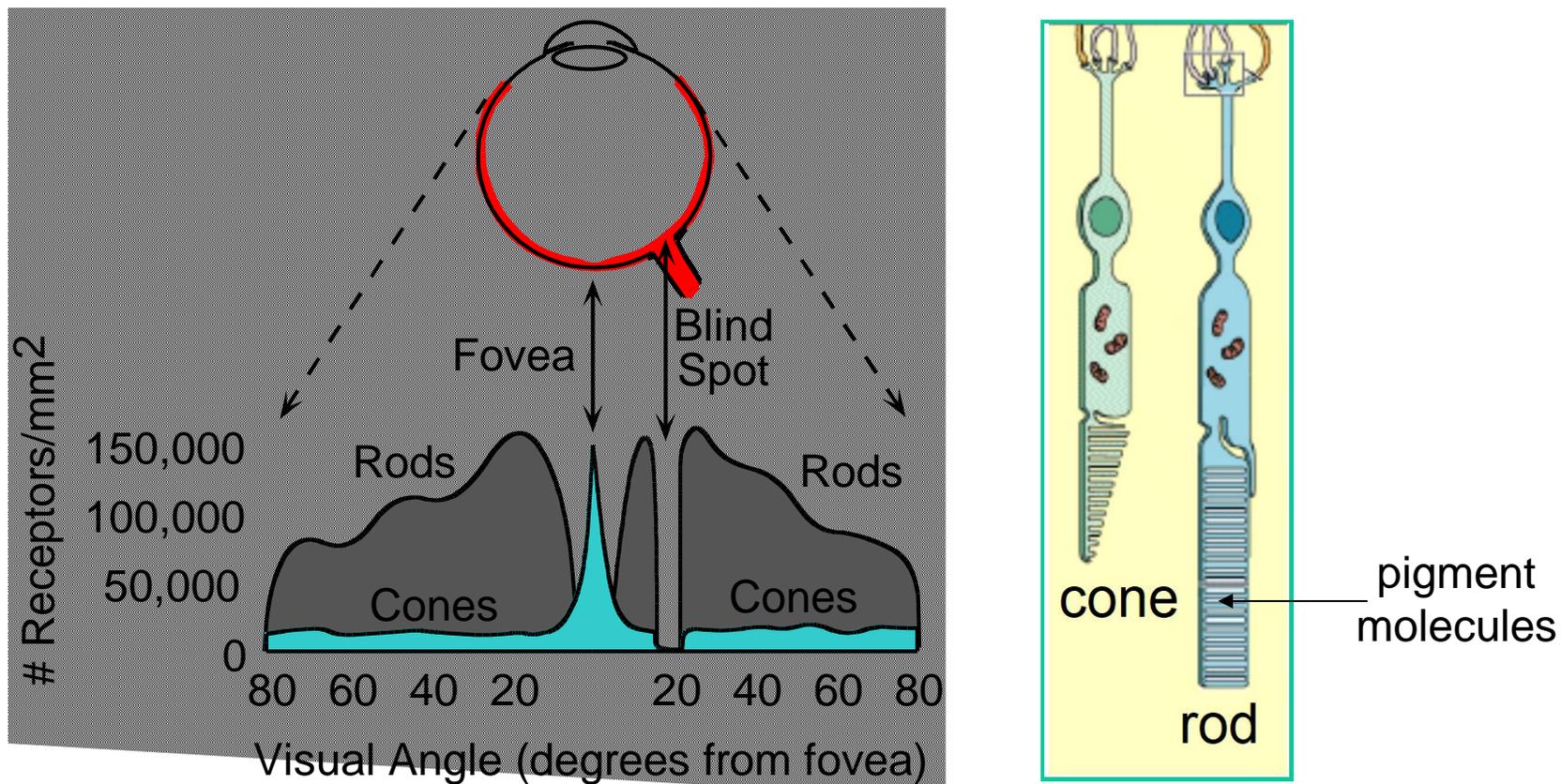


Что мы на самом деле видим



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

Неравномерность пространственного разрешения



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

Постоянство цвета и освещенности



J. S. Sargent, The Daughters of Edward D. Boit, 1882

Способность зрительной системы человека оценивать собственные отражательные свойства поверхностей в не зависимости от условий освещенности

Освещённость:

- Полнолуние – 0.27 люкс
- Пасмурный день – 100-1000 люкс
- Яркий день (в тени) – 20000 люкс
- На солнце – 100000 люкс

Следствие - проблема передачи яркости



Нужна **тональная коррекция** - осветление, затемнение или повышение контрастности всего изображения или его частей.



Следствие - проблема передачи цвета

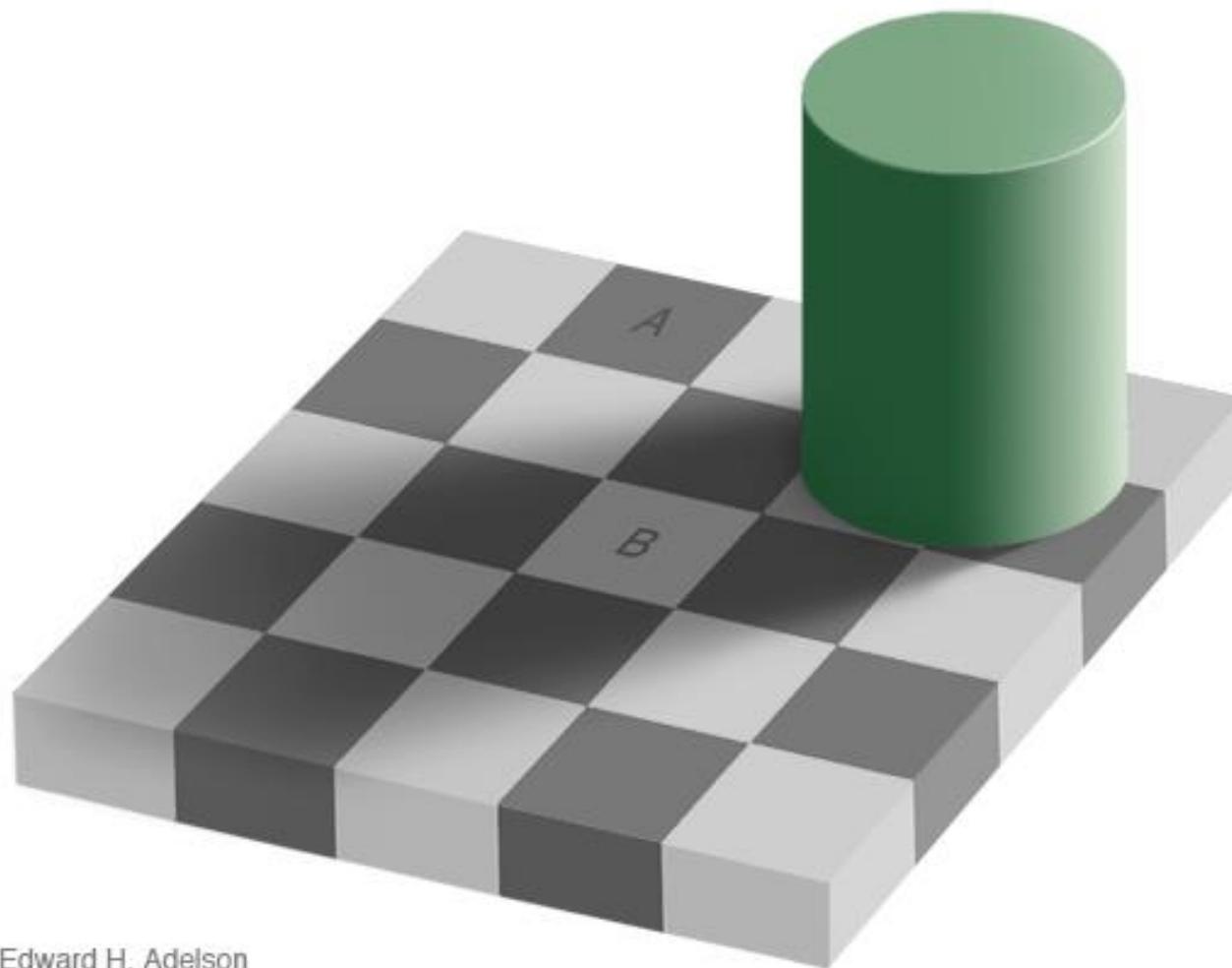
Когда мы смотрим на фотографию или монитор, глаза адаптируются к освещению в комнате, а не к освещению сцены на фотографии. Поэтому цвета фотографии могут казаться неестественными.



<http://www.cambridgeincolour.com/tutorials/white-balance.htm>

Нужна цветокоррекция для обеспечения «естественности» цветов

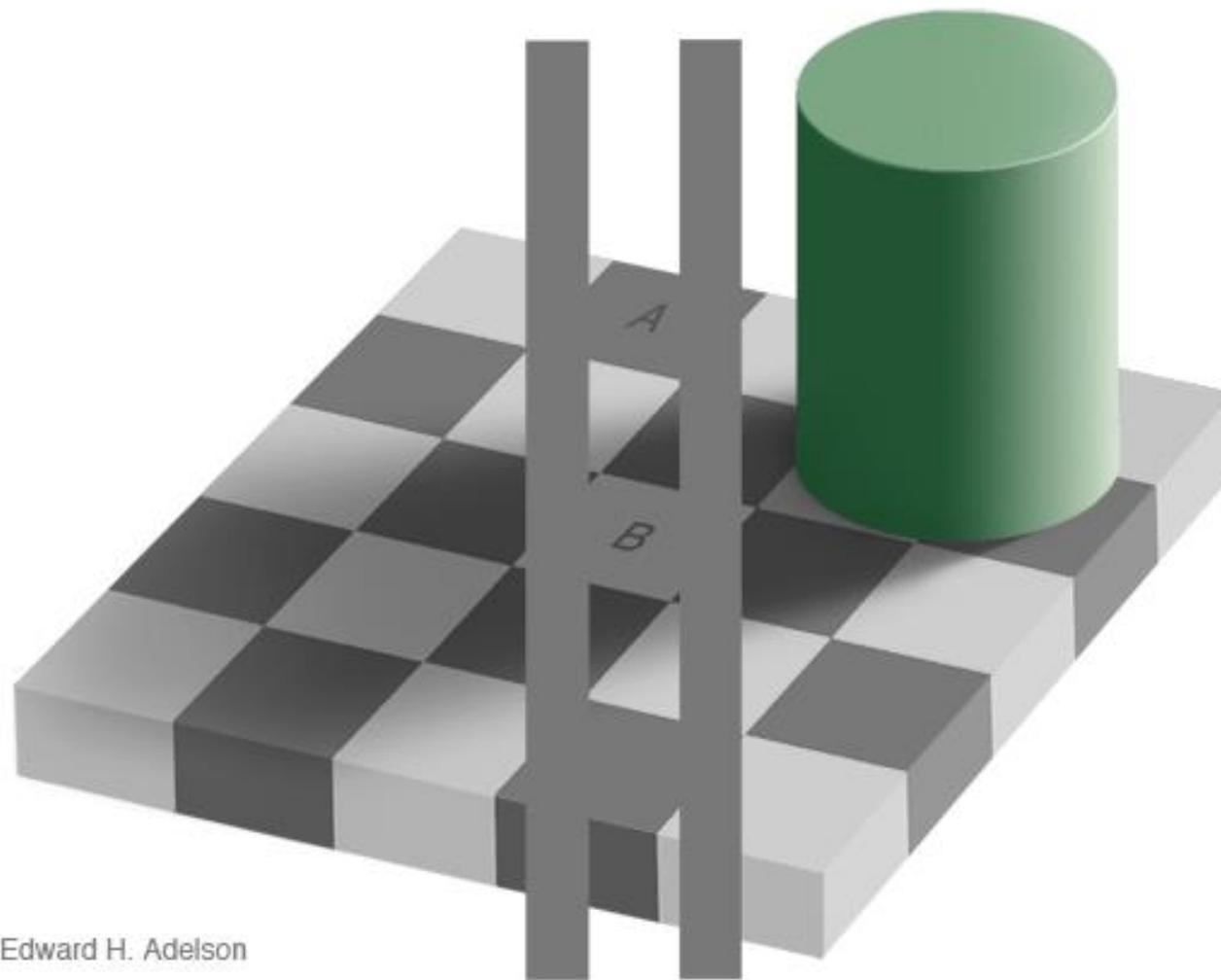
Локальность оценки яркостей



Edward H. Adelson

http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

Локальность оценки яркостей

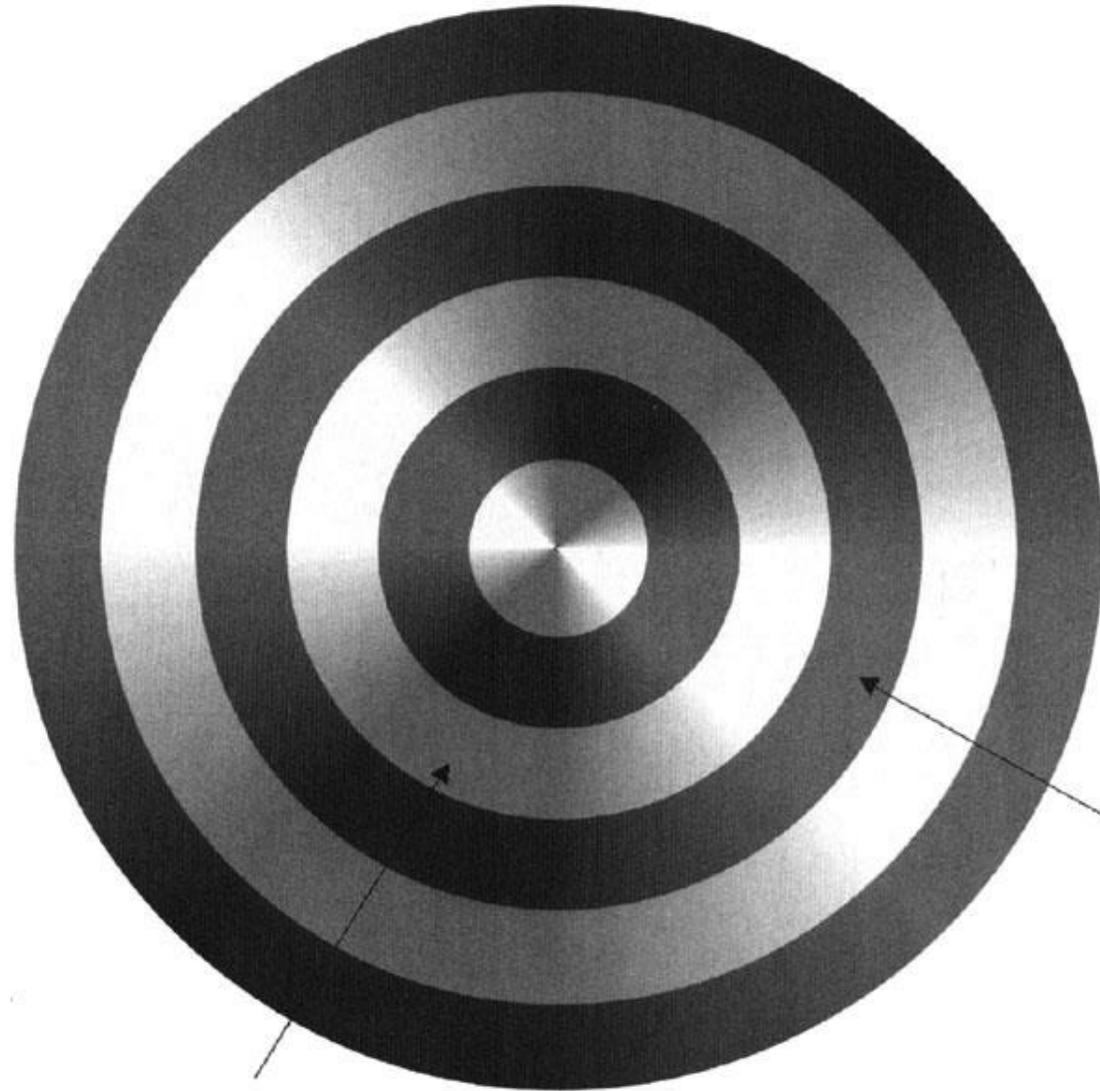


Edward H. Adelson

http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

Slide by S. Lazebnik

Локальность оценки яркостей





- Понятие о компьютерном зрении
- Устройство глаза и цифровой камеры
- Трихроматическая теория цвета
- Цветовые модели LMS, RGB, XYZ, CMYK, YIQ, HSV
- Особенности зрения человека
 - Неравномерная разрешающая способность и фокус внимания
 - Постоянство яркости и освещённость
 - Локальность оценки яркости