No.1

コンピュータビジョン基礎

Basis of Computer Vision

担当教員：向川康博・田中賢一郎

What is Computer Vision?

- Computer vision is a method that extracts various visual information of real world from images.
- It gives computers and robots the ability for visually recognizing real world.

Real World

Observation target

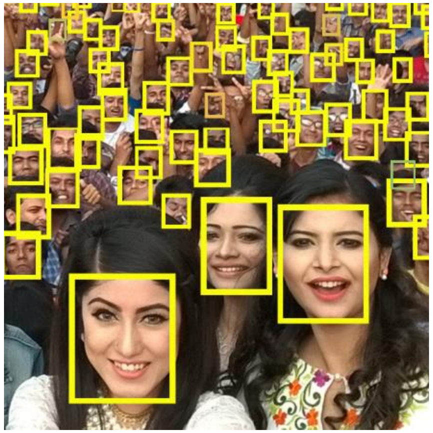
Observation

Computer

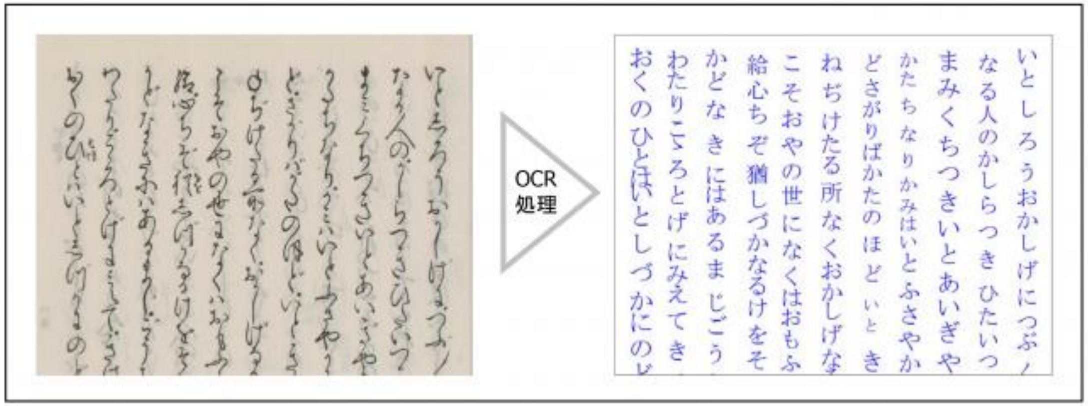
A red sphere on the white table.
It is soft.
It seems that I took a picture indoors.

Applications of computer vision

- Face recognition



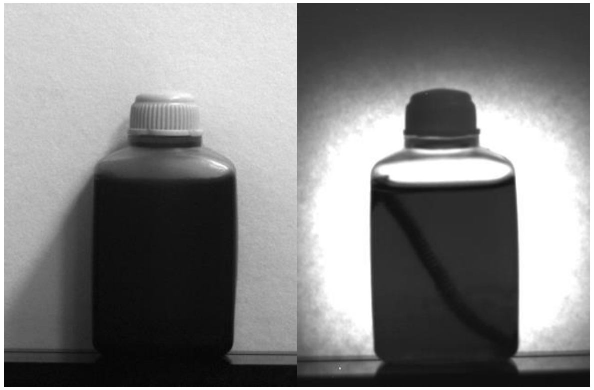
- Optical character recognition (OCR)



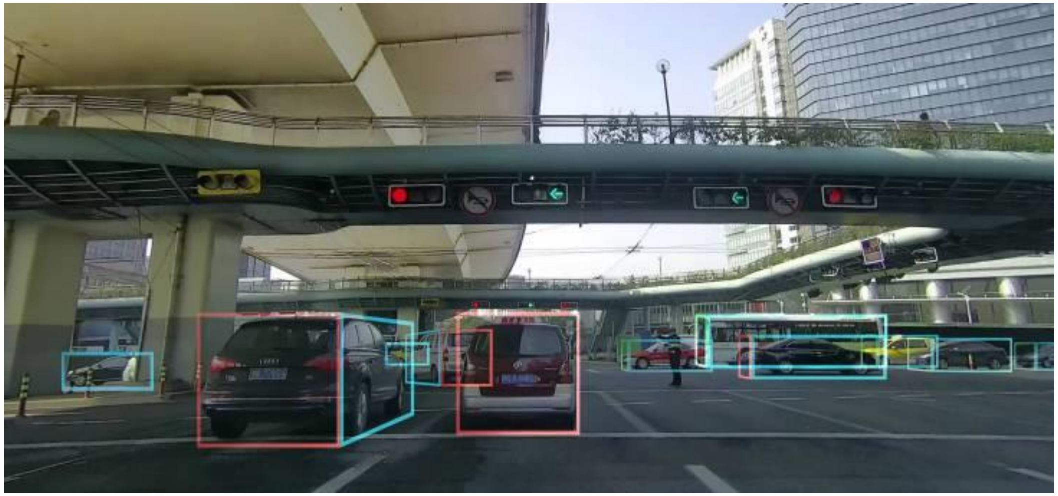
凸版印刷

- Machine inspection

Monochrome camera　Near infrared camera

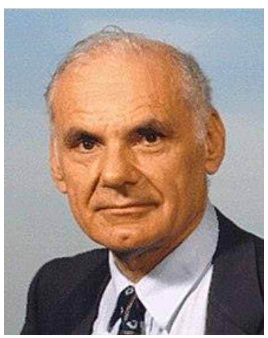


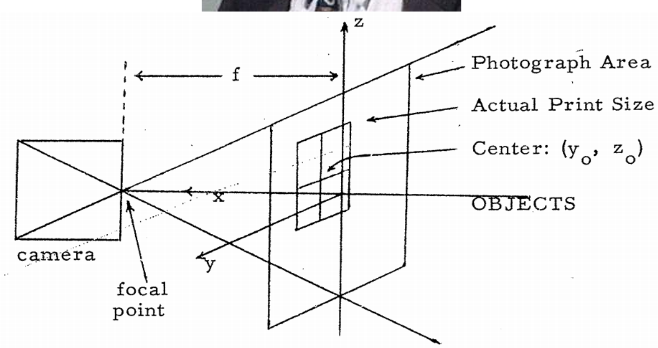
- Autonomous driving car



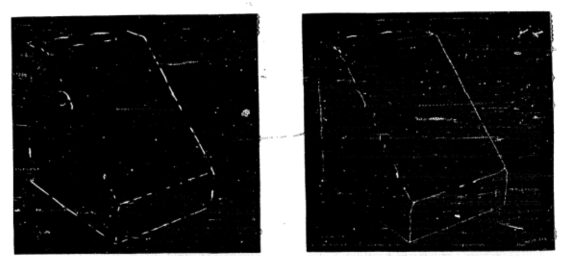
1960s : Lawrence Roberts-The Father of Computer Vision

- In 1963 he published *“Machine Perception Of Three-Dimensional Solids”*
- he discusses **extracting 3D information about solid objects from 2D photographs of line drawings**.

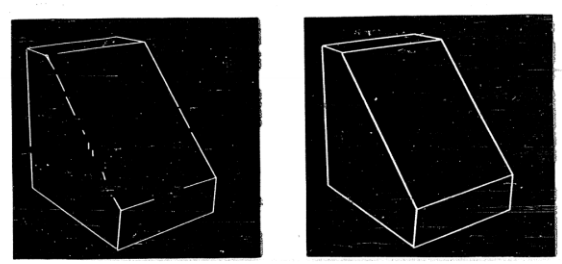




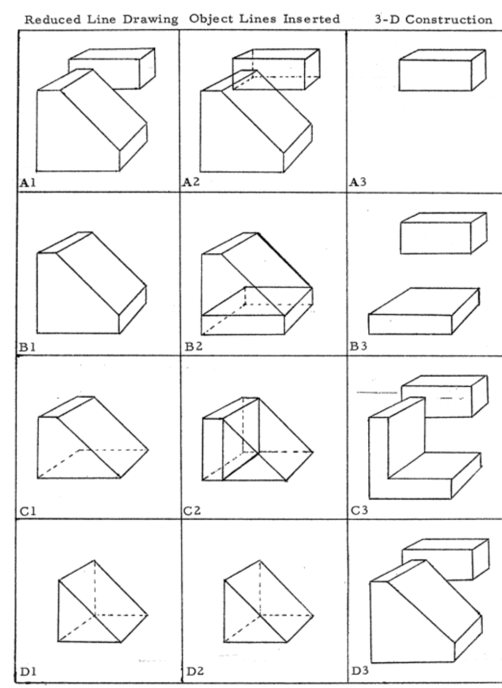
1: Camera Transformation



E. Connected Feature Points   F. After Complexity Reduction



G. After Initial Line Fitting   H. Final Line Drawing



1966 : The Summer Vision Project at the MIT

- Marvin Minsky asked "spend the summer linking a camera to a computer and getting the computer to describe what it saw"
- He gave students segmentation and recognition tasks for objects and backgrounds.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual
The particular task was chosen partly because it can be segmented
sub-problems which will allow individuals to work independently a
participate in the construction of a system complex enough to be
landmark in the development of "pattern recognition".

Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

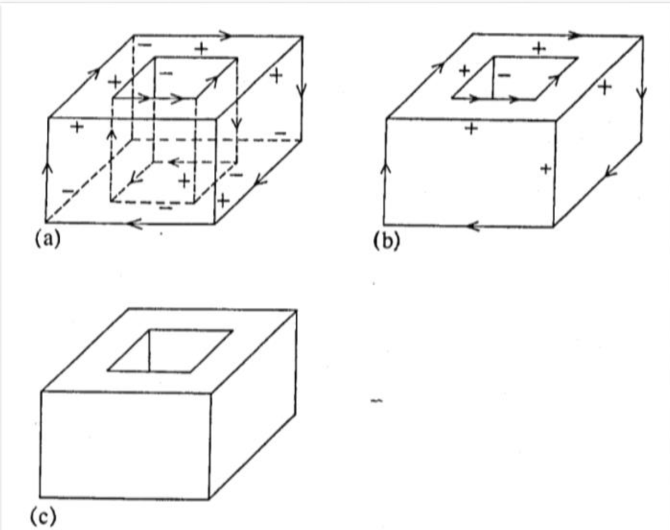
likely background areas

chaos.

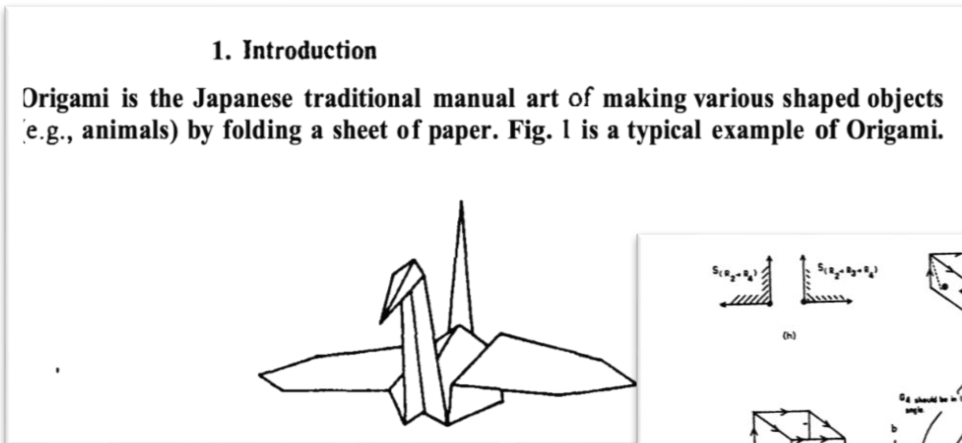
https://dspace.mit.edu/bitstream/handle/1721.1/6125/AIM-100.pdf?sequence=2

1970's : Line labelling

- Research for line drawings because real-world images were just too hard to handle at the time.
- Early attempts at scene understanding involved extracting edges and then inferring the 3D structure of an object or a "blocks world" from the topological structure of the 2D lines.

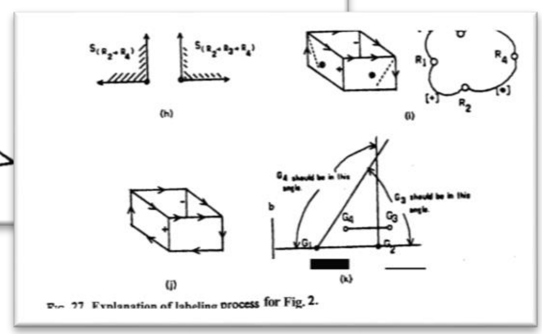


'igure 7. Illustrating three kinds of pictures: (a) labelled 'X-ray' picture-graph; b) labelled picture-graph; (c) unlabelled picture-graph



1. Introduction

Origami is the Japanese traditional manual art of making various shaped objects (e.g., animals) by folding a sheet of paper. Fig. 1 is a typical example of Origami.

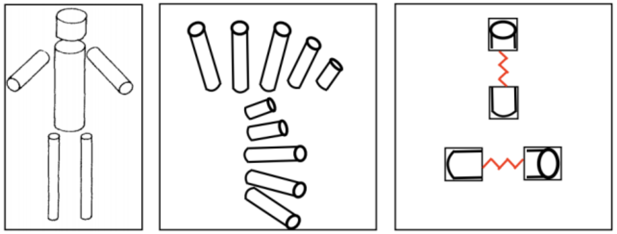


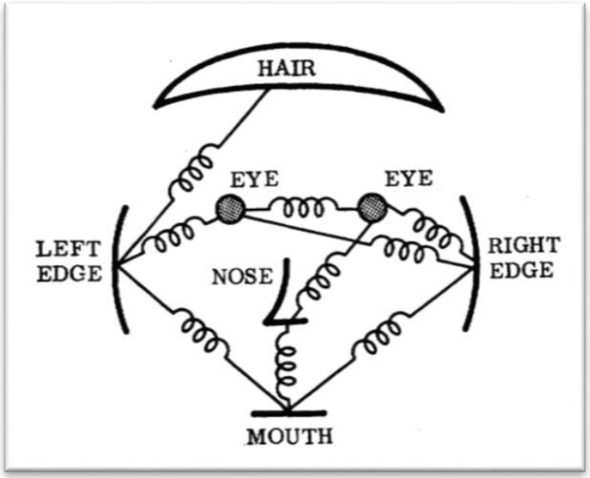
[Huffman 1971](#)

[Kanade 1980](#)

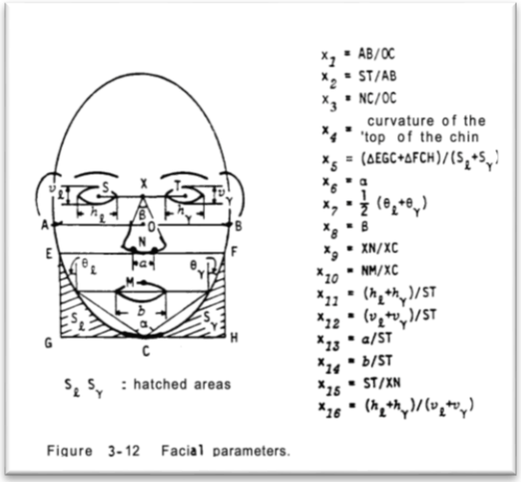
1970's : Pictorial structures

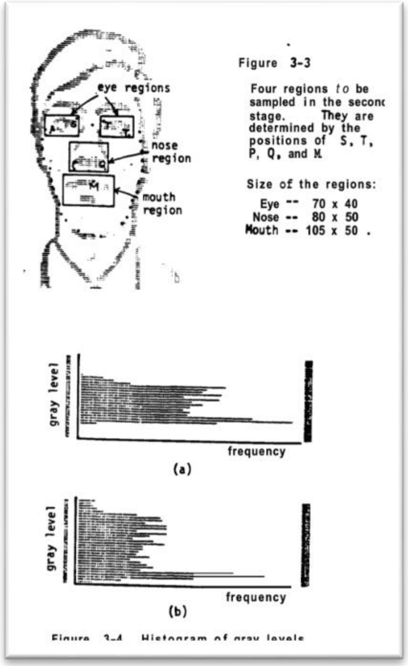
- Recognizing an object by finding its constituent parts and measuring their geometric relationships is one of the oldest approaches to object recognition





[MA Fischler 1973](#)

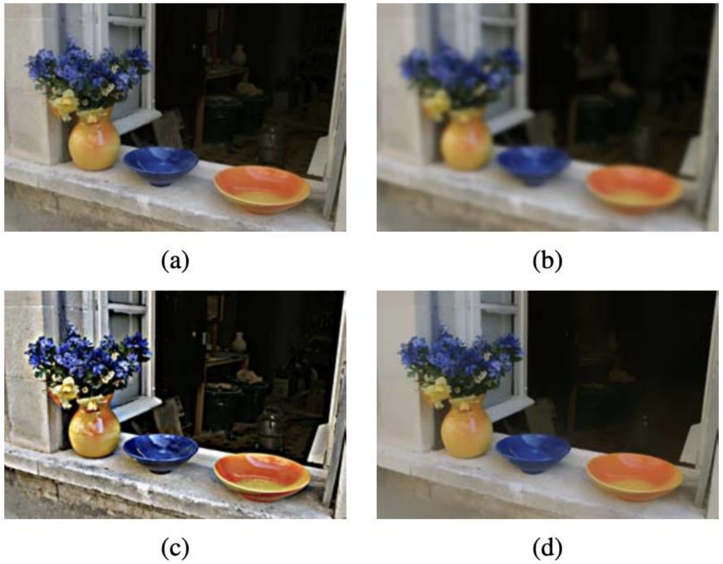




[Kanade 1977](#)

Image Processing

- Input an image and output another image that serves a purpose
  - Linear filter
    - Soft blue
    - Sharpen details
    - Accentuate edges
    - Remove noise
  - Geometric transformations



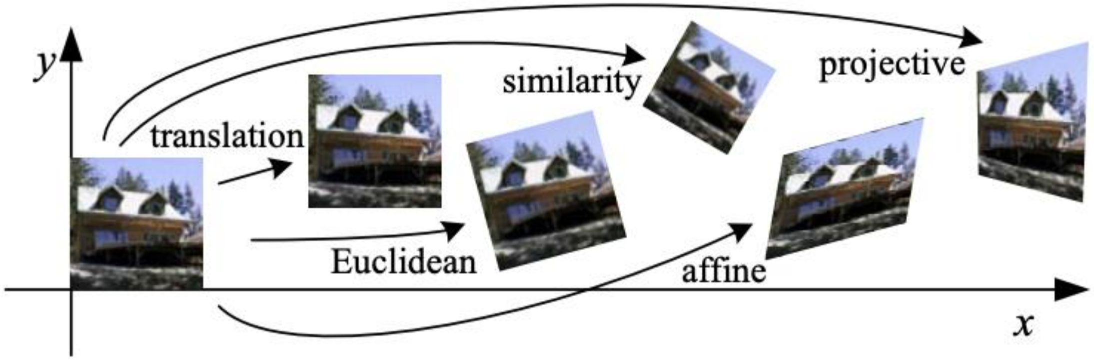
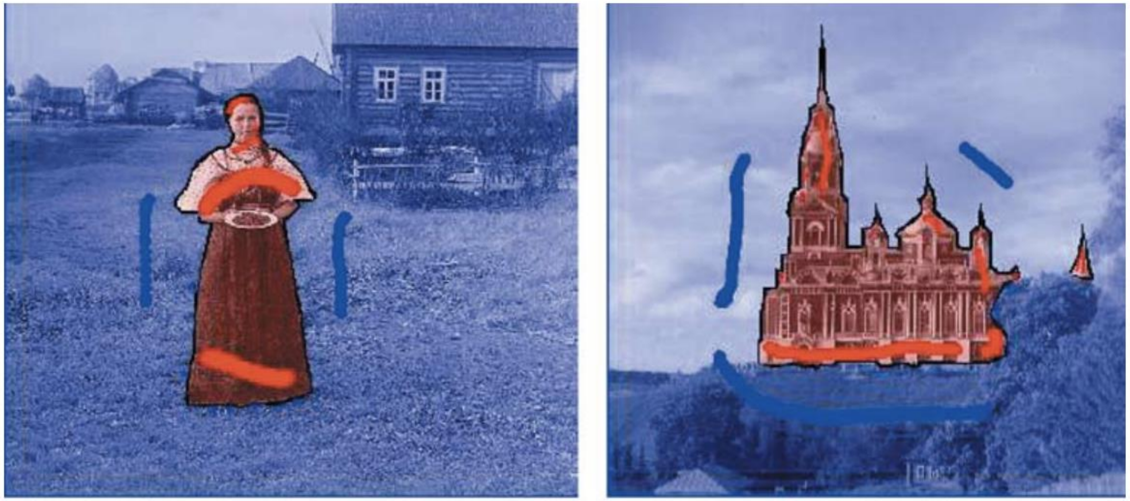


Image Segmentation（領域分割）

- The image is divided into parts or regions based on the characteristics of the pixels in the image.

Graph cut

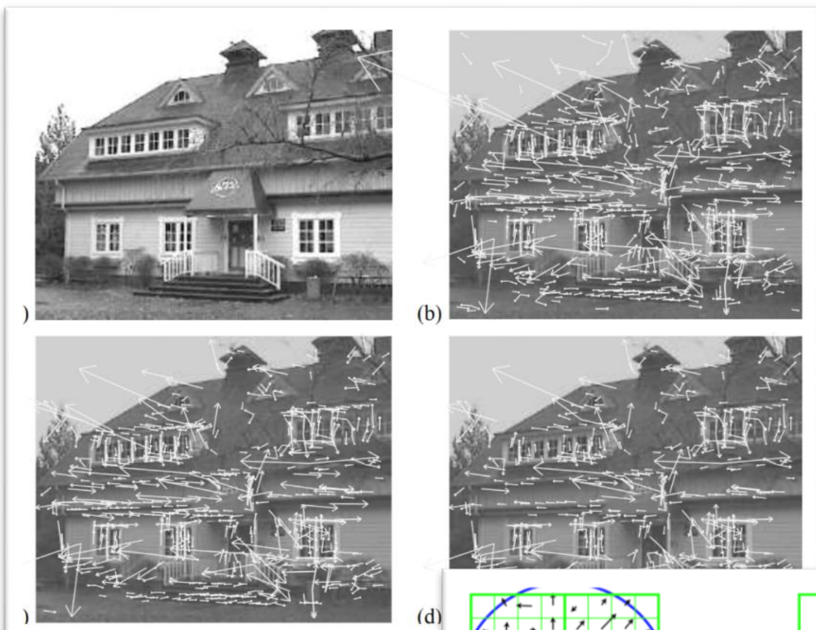


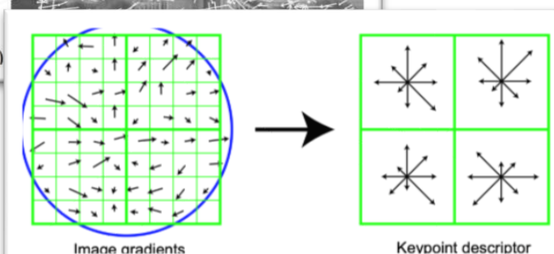
(a) A woman from a village (b) A church in Mozhaisk (near Moscow)

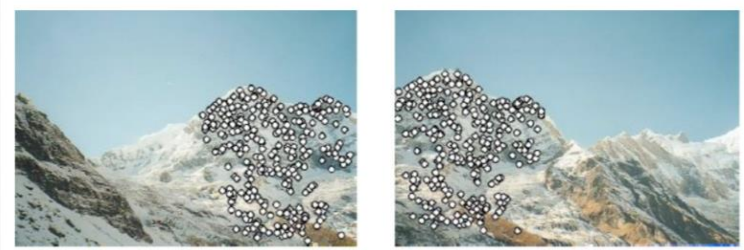
Boykov and Funka-Lea 2006

Image features

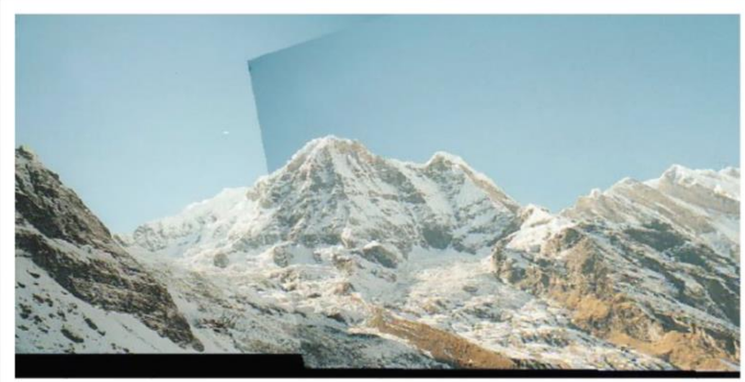
- A variety of operators
  - SIFT
  - HOG







(c) SIFT matches 1 (d) SIFT matches 2



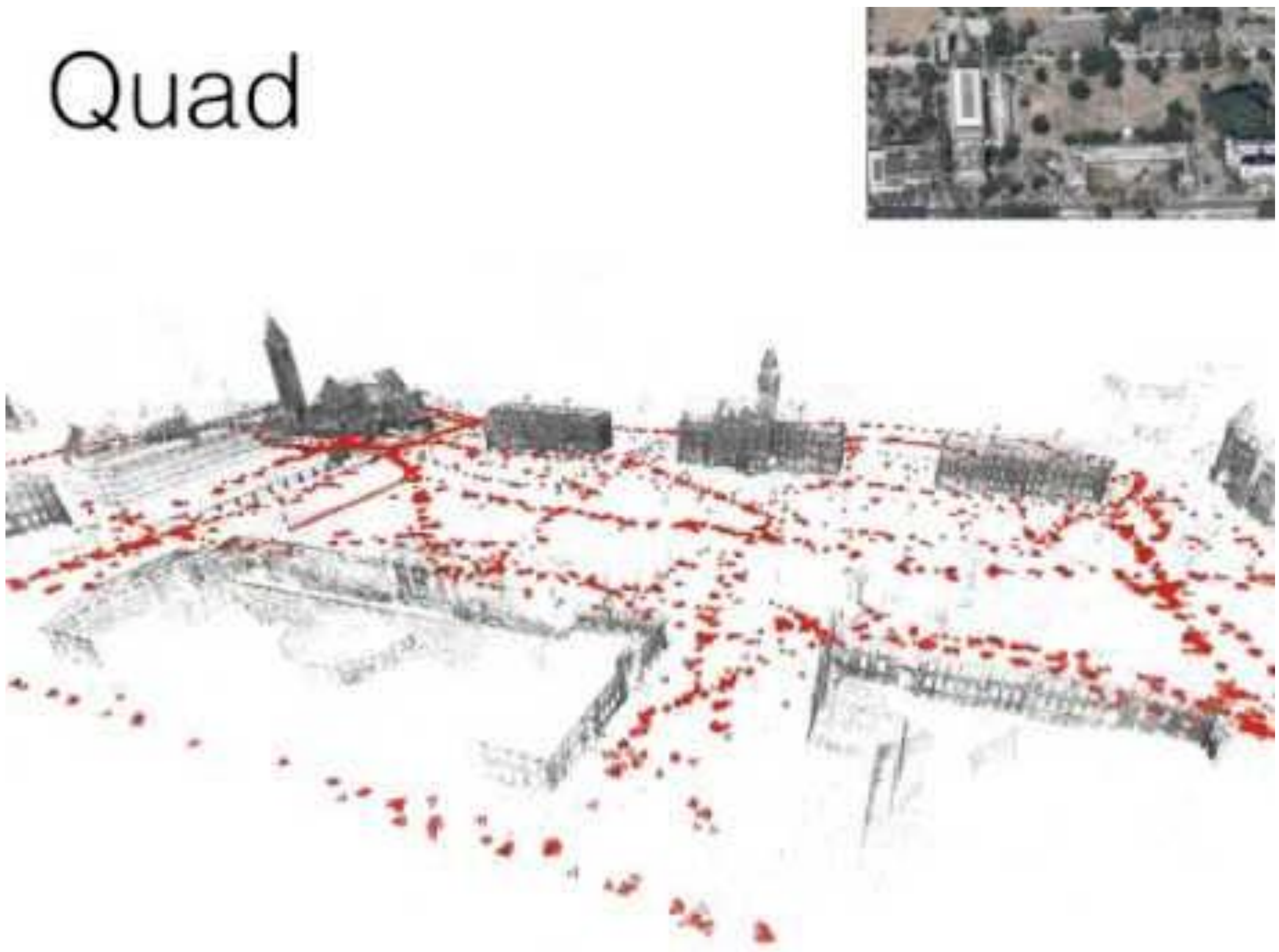
(e) Images aligned according to a homography

3D reconstruction

- Structure from motion
  - Restore the 3D shape of the subject from multiple images with different viewpoints

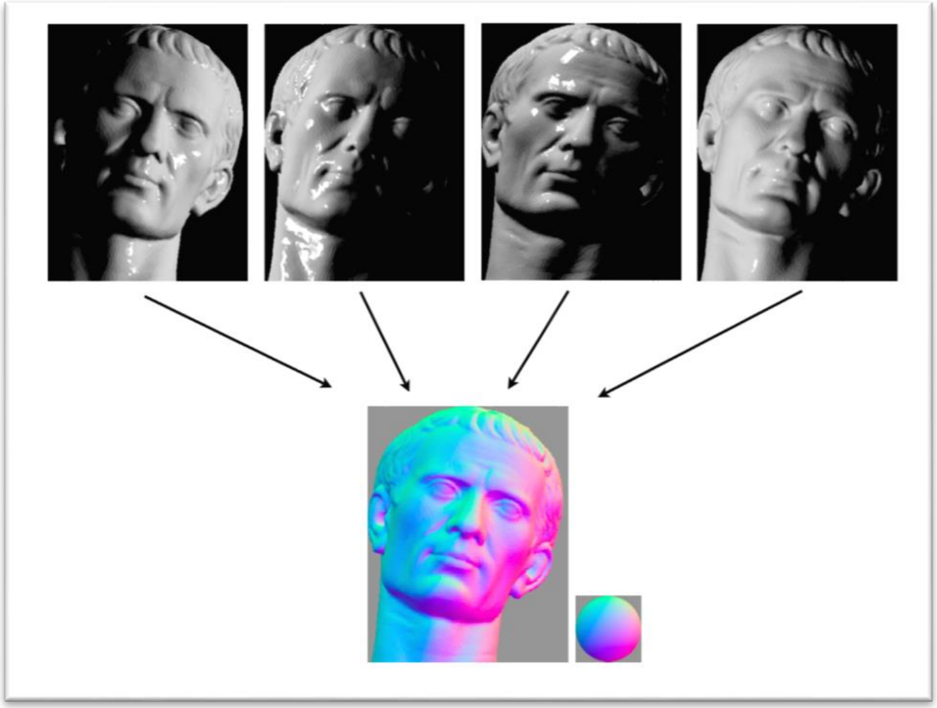
3D reconstruction

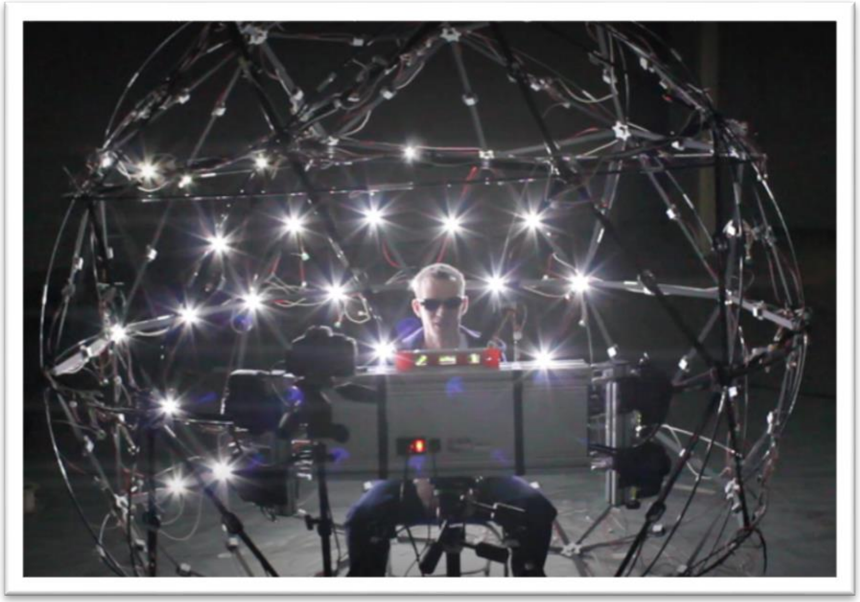
- Structure from motion



3D reconstruction

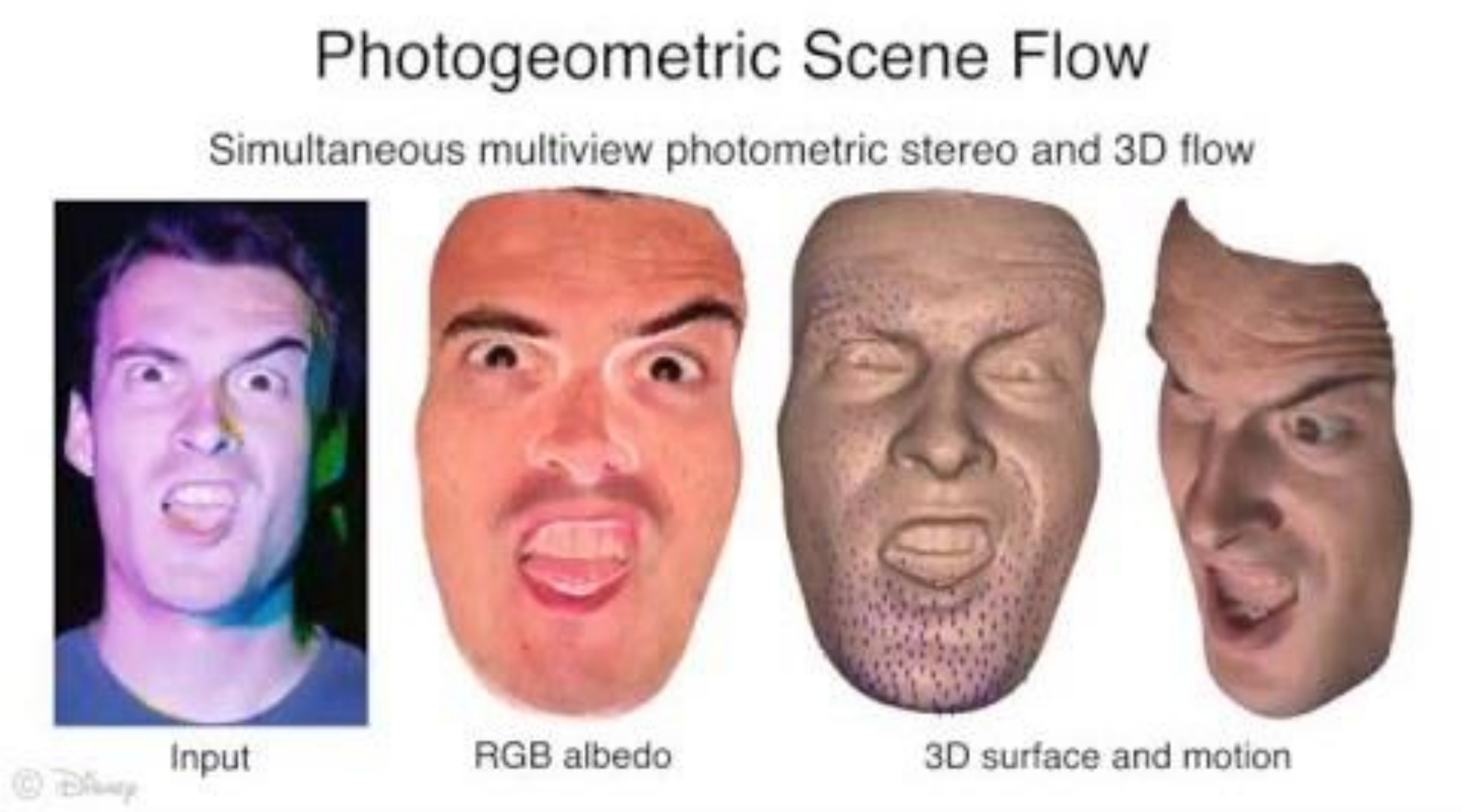
- Photometric stereo
  - Use multiple images taken with different lighting directions





3D reconstruction

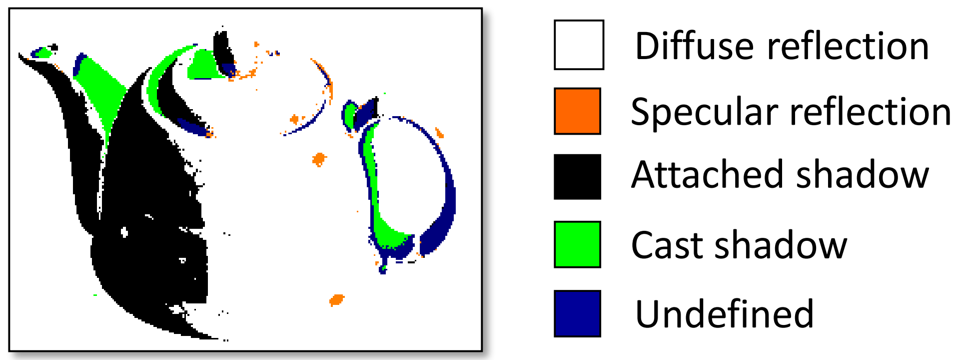
- Photometric stereo

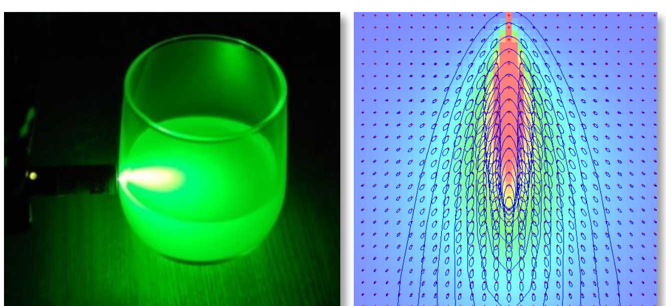


http://openaccess.thecvf.com/content_iccv_2015/papers/Gotardo_Photogeometric_Scene_Flow_ICCV_2015_paper.pdf

Photometric analysis

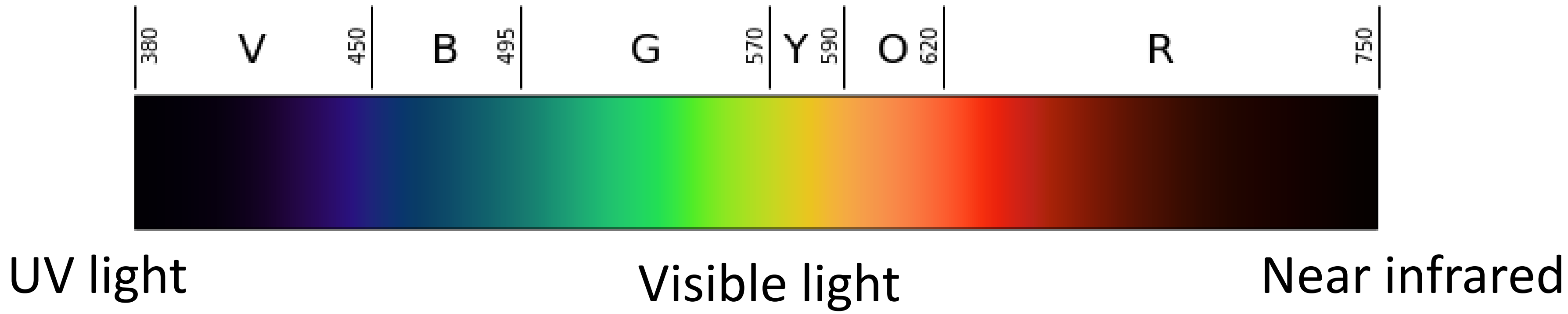
- Reflection, shade, shadow, scattering
- Analysis by physics-based model





Spectral imaging

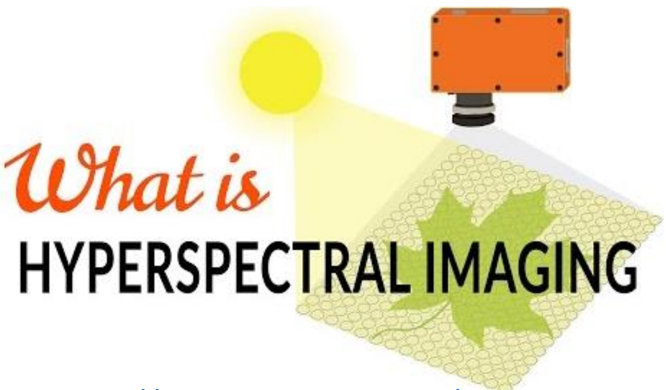
- Light is wave(electromagnetic wave)
- Light contains various wavelengths.
- Color is not a physical quantity but human sense quantity
- Generally use hyperspectral camera



Hyperspectral imaging

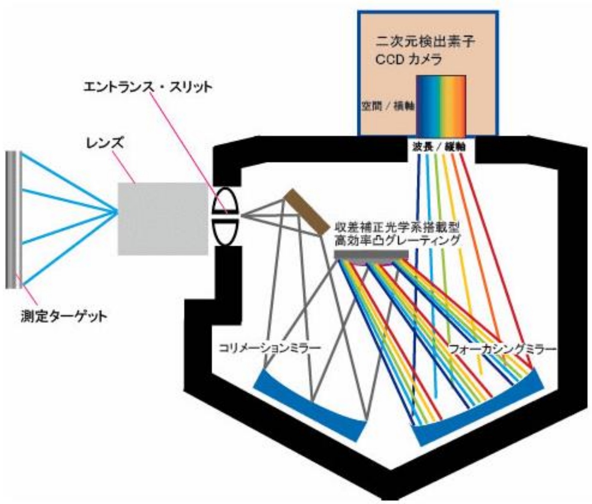
- Hyperspectral camera

http://www.specim.fi/iq/

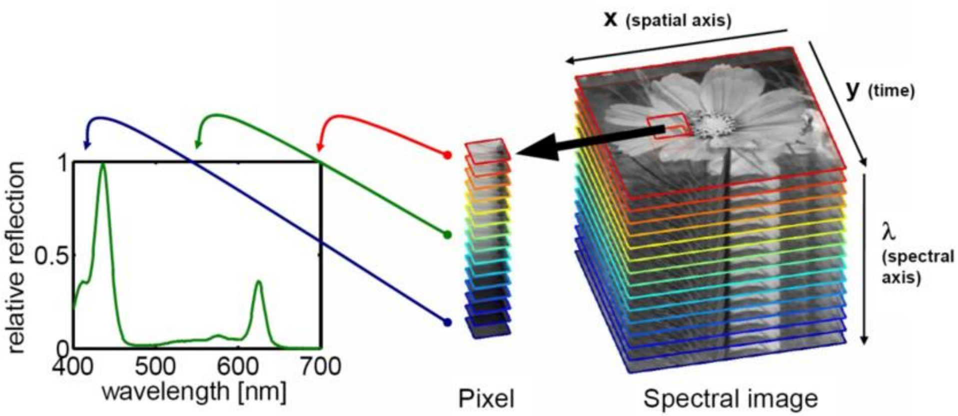


https://www.youtube.com/watch?time_continue=151&v=e6hbmSfPnMQ

Construction



Spectral Cube

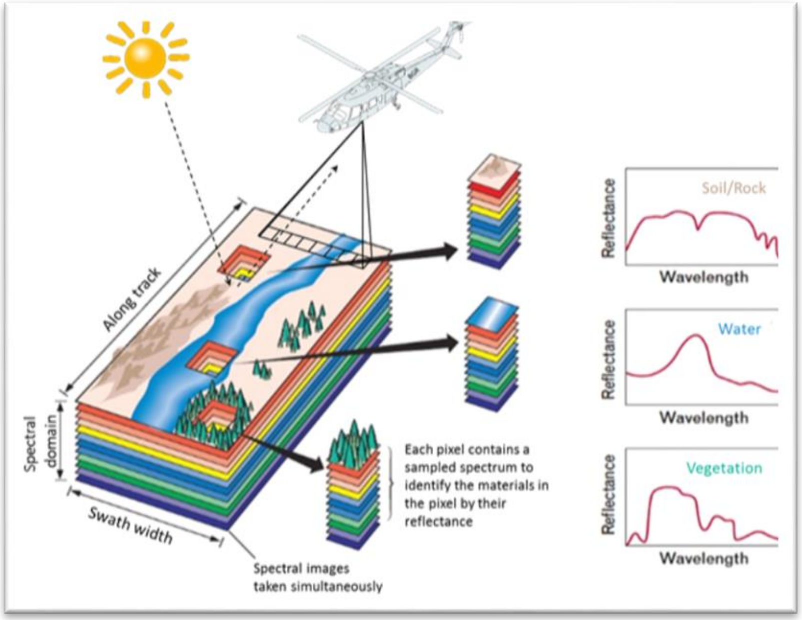


https://www.argocorp.com/cam/special/HeadWall/how_it_works.html

Spectral imaging

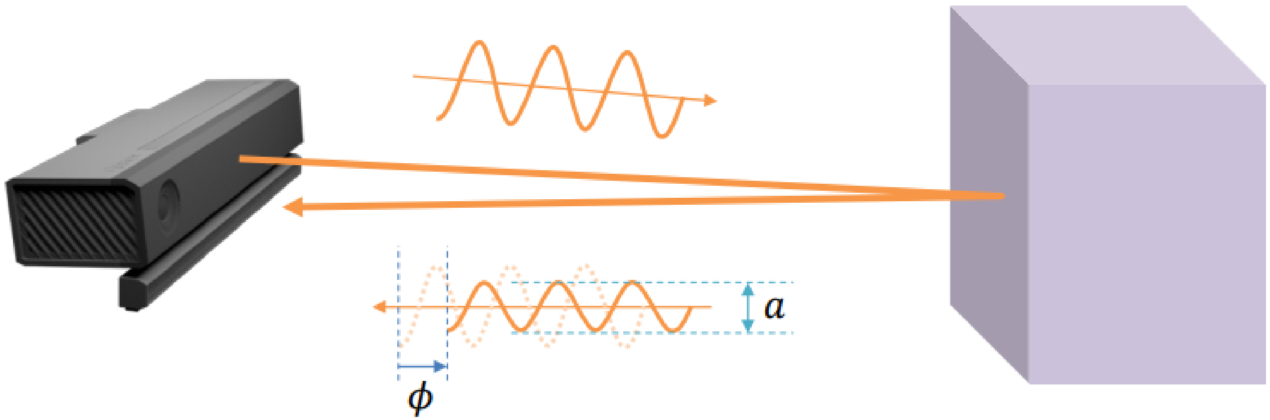
- What can Hyperspectral imaging do ?

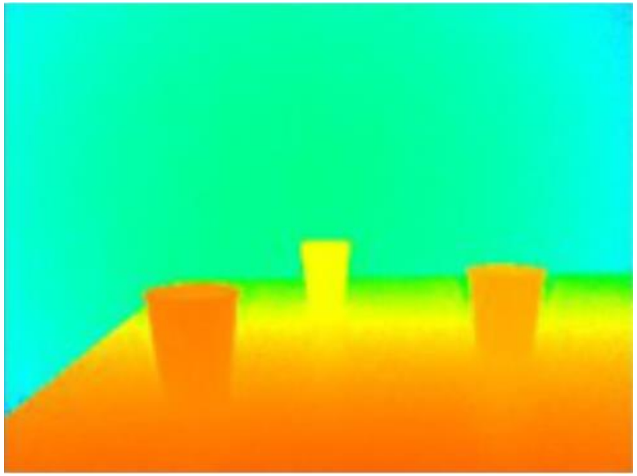
Remote sensing



Visualization

Time-of-Flight camera





depth image

$$d = \frac{c\phi}{4\pi f}$$

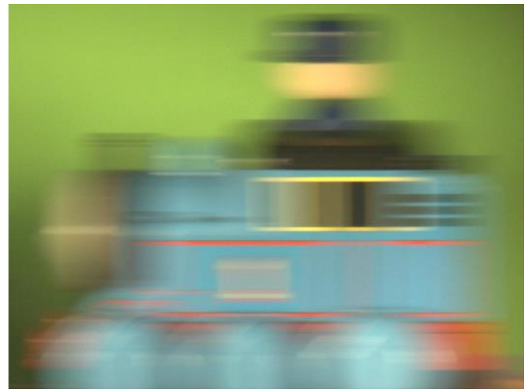
c: speed of light

f: frequency of modulation

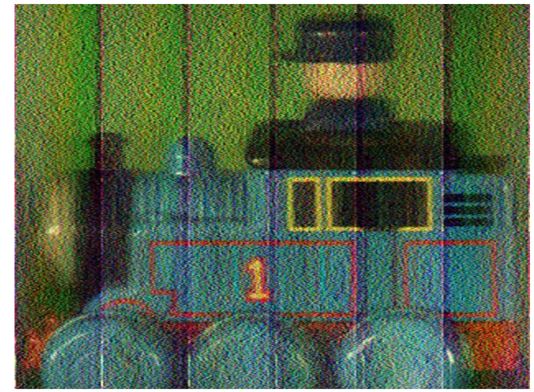
Computational photography

- Coded exposure（符号化露光）

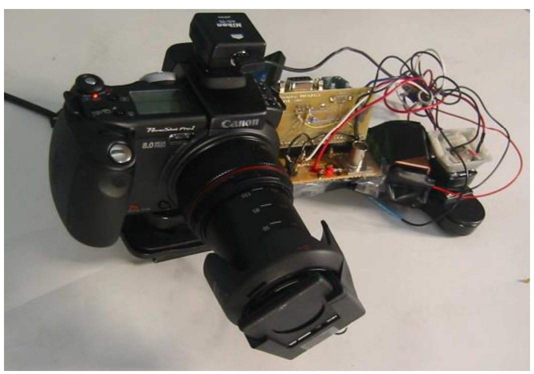
Short exposure

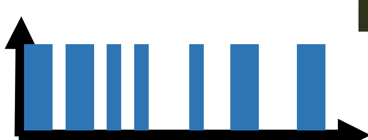


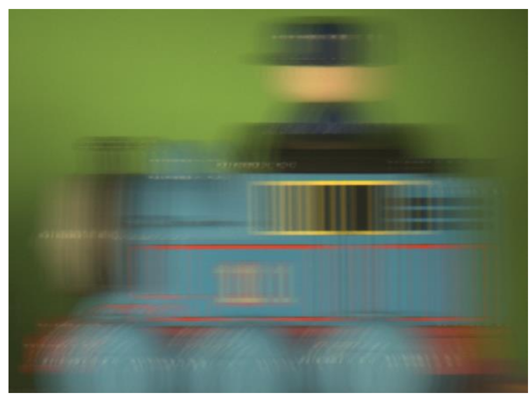
Long exposure



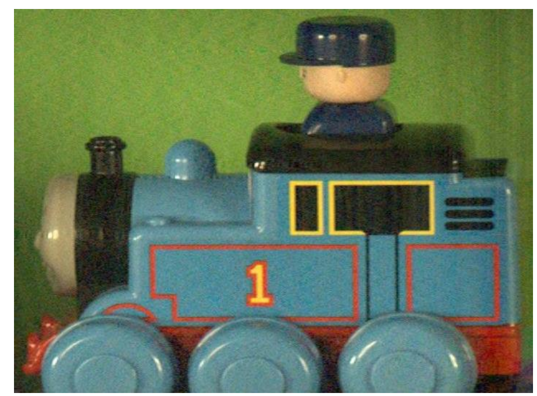
Deblurred







Coded exposure

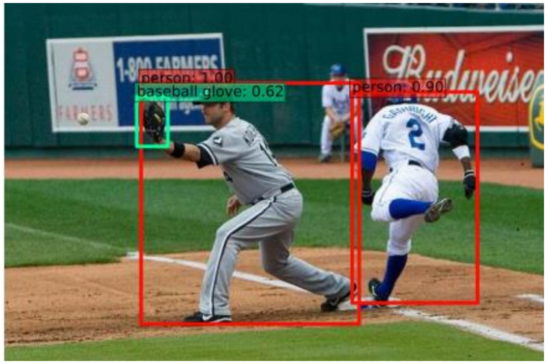


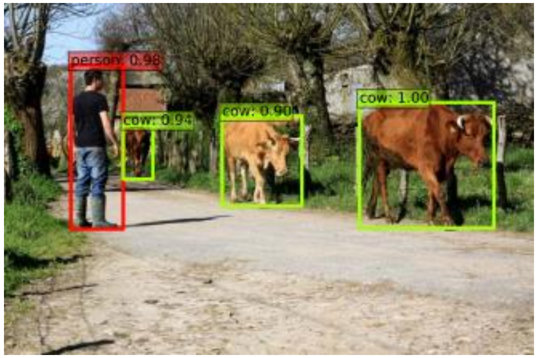
Deblurred

Raskar et al., SIGGRAPH2006

Application using deep learning

- Object Recognition

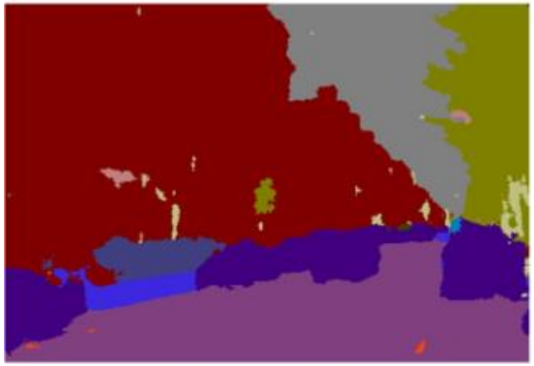




SSD: Single Shot MultiBox Detector

https://arxiv.org/abs/1512.02325

- Semantic segmentation

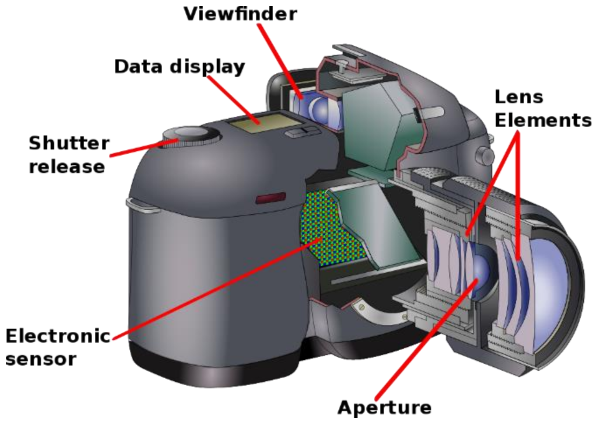
SegNet

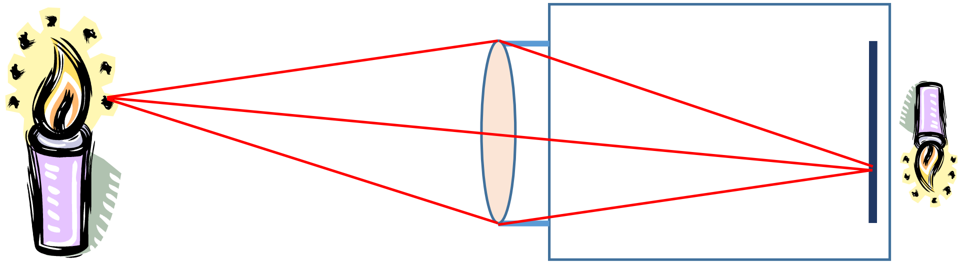
https://arxiv.org/abs/1511.00561

Input device: camera

What is Camera?

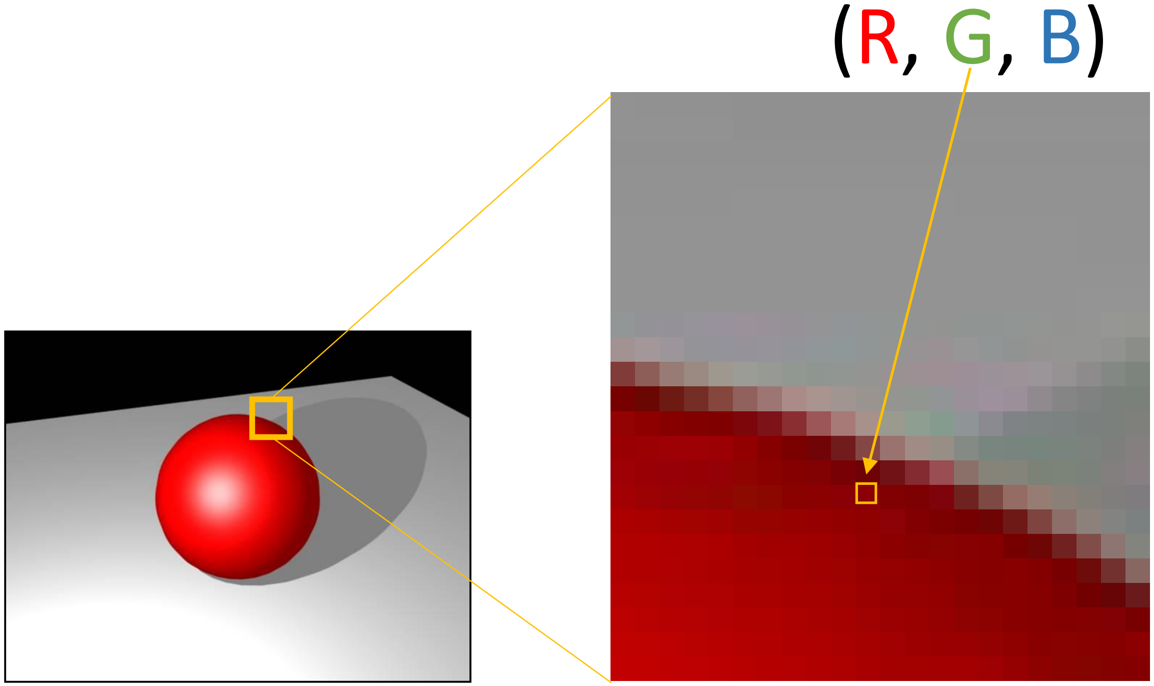
- A camera is an optical instrument for recording or capturing images, which may be stored locally, transmitted to another location, or both. The images may be individual still photographs or sequences of images constituting videos or movies. The camera is a remote sensing device as it senses subjects without any contact . (Wikipedia.org)





Digital image

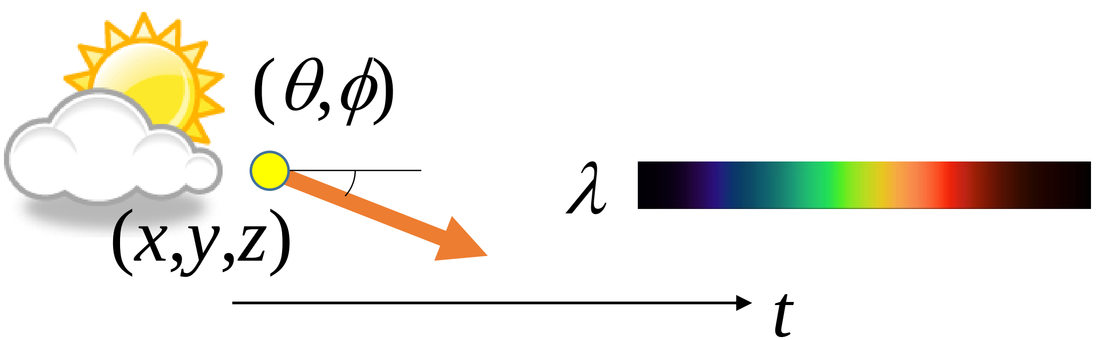
- Discrete 2D array of RGB values.
  - uncompressed or compressed
    - raw, jpeg, png, bmp, tiff,...
  - 1,920 x 1,080 = 2M pixel (FullHD)
  - 8bit / color channel (higher bit for professional)



Ray in 3D space

- Ray has rich information
  - passing point: (x,y,z)
  - passing direction: (θ,ϕ)
  - wavelength: λ
  - time: t
- Plenoptic function

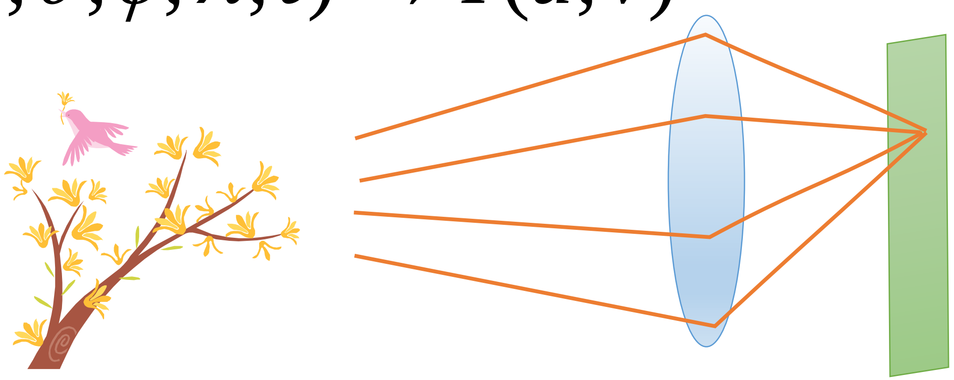
$$P(x, y, z, \theta, \phi, \lambda, t)$$

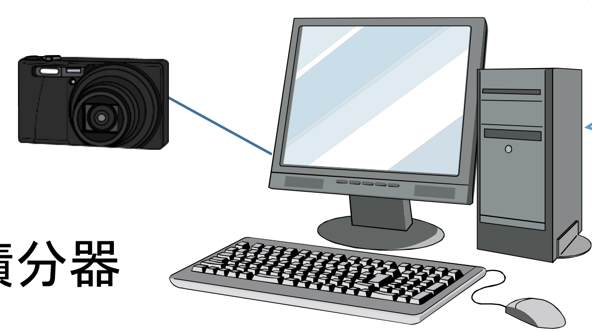


From ray to image

- Image is an integral of rays. Camera is integrator.
  - angular integration by lens
  - temporal integration by shutter
  - spectral integration by color filter

$$P(x, y, z, \theta, \phi, \lambda, t) \rightarrow I(u, v)$$





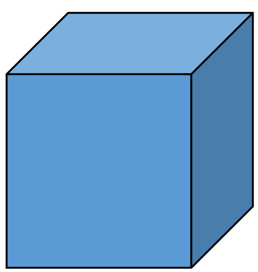
カメラは光線の積分器

109	93	75	82	83
103	98	103	95	85
129	120	106	107	102
133	117	84	68	110
106	90	61	46	74

2D array of pixels

Conversion from 3D scene to 2D image

- Camera as a converter from 3D scene to 2D image
- Geometric conversion
  - from 3D world coordinate (x,y,z)
  - to 2D image coordinate (u,v)
- Photometric conversion
  - from wavelength to RGB values



World coordinates (x,y,z)

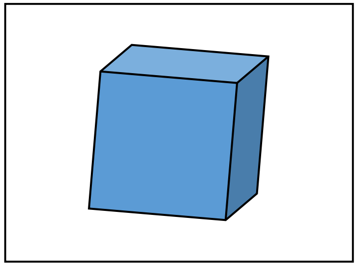


Image coordinates (u,v)

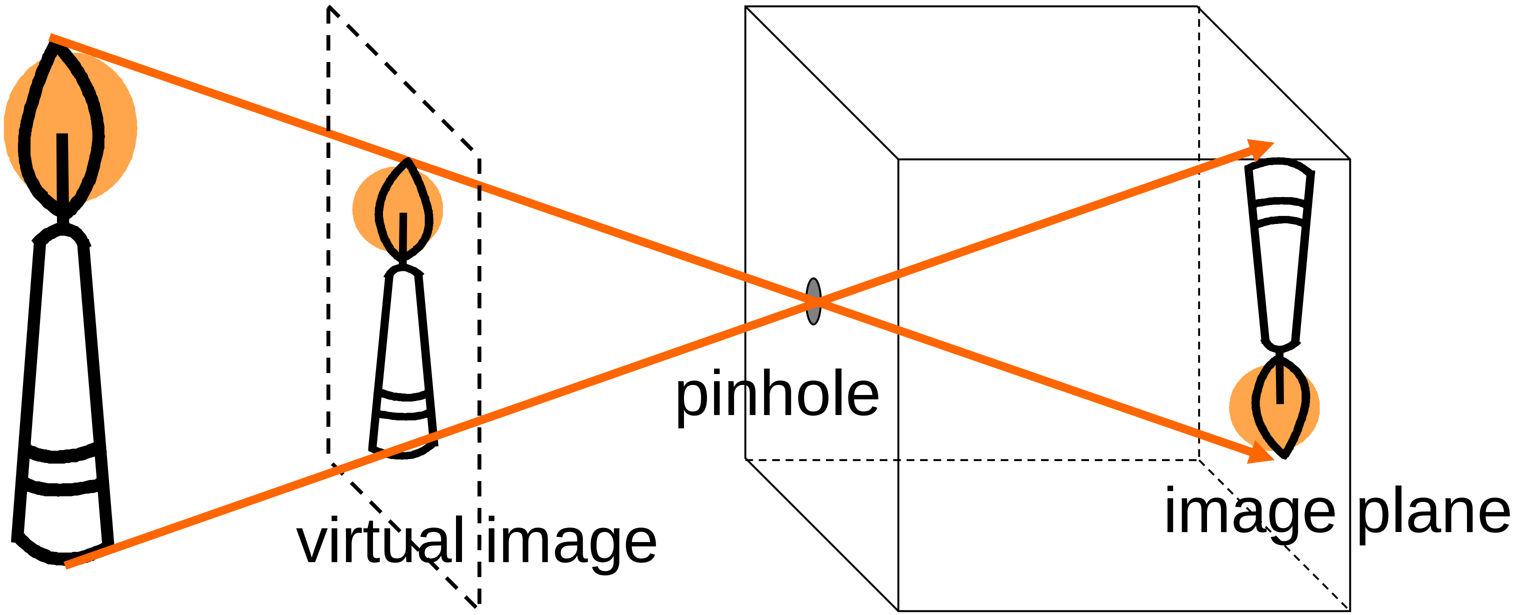
Geometric conversion
- 3D to 2D -

Different views

- Two images were captured using a same camera from same direction.
- Q1: What is different?

Pinhole perspective projection

- Extremely simple camera model.
- A box with a small hole in one of its sides with a pin.
- An inverted image on the opposite side.
- A virtual image associated with a plane lying in front of the pinhole at the same distance from it.



Perspective effects

- The apparent size depends on the distance.
  - Far objects appear smaller than close one.
- A line is observed as a line.
- Parallel lines intersect at the horizon.
  - The projections of two parallel lines lying in plane Π converge on a horizon line H formed by the intersection of the image plane with the plane parallel to Π and passing through the pinhole.

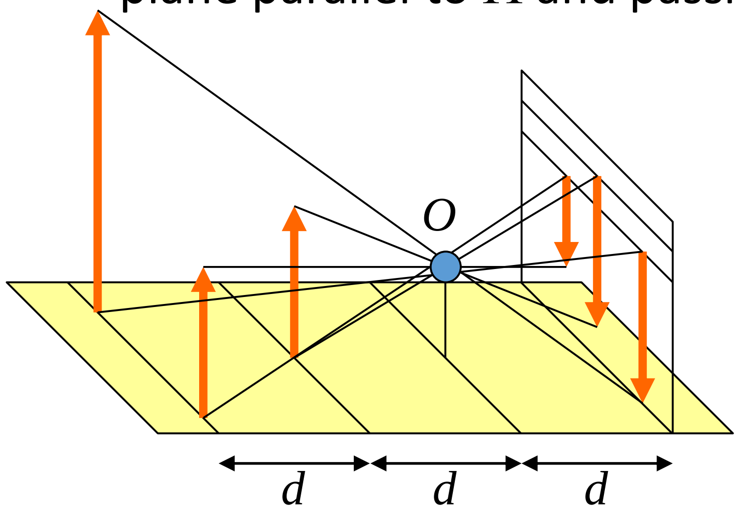


image plane behind the pinhole

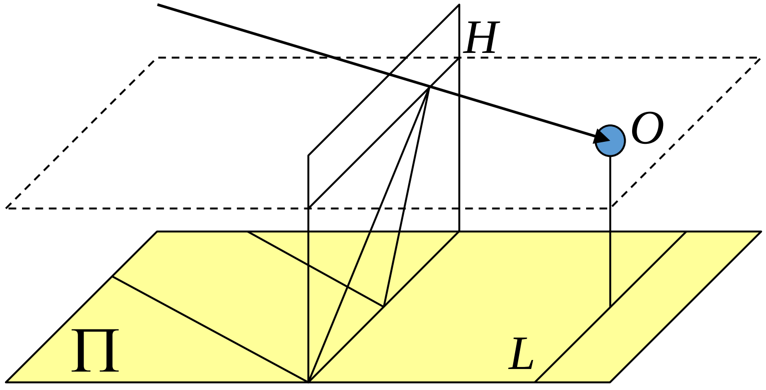


image plane in front of the pinhole

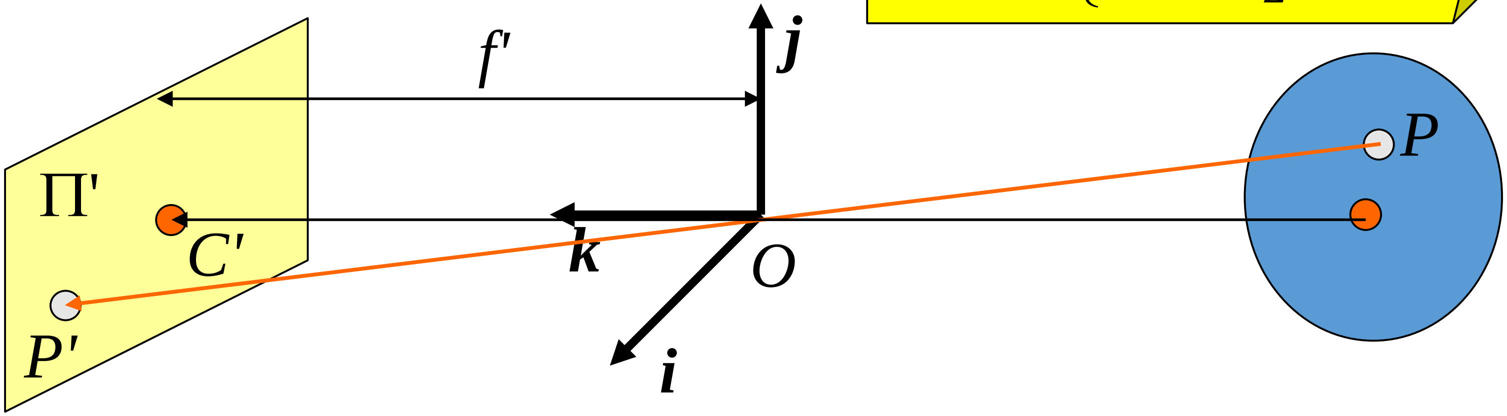
Perspective projection equations

- P: a scene point with coordinate (x,y,z)
- P': its image with coordinate (x',y',z')
- Since P' lies in the image plane, $z'=f'$.
- Since the three points P,O,P' are collinear, $OP'=\lambda OP$.

$$\begin{cases} x' = \lambda x \\ y' = \lambda y \\ z' = \lambda z \end{cases} \Leftrightarrow \lambda = \frac{x'}{x} = \frac{y'}{y} = \frac{z'}{z}$$

Therefore,

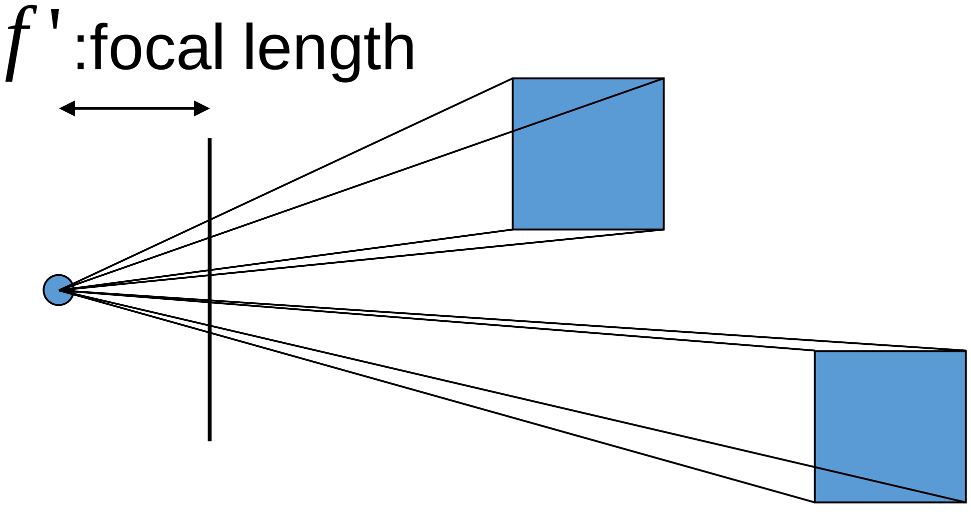
$$\begin{cases} x' = f'\dfrac{x}{z} \\ y' = f'\dfrac{y}{z} \end{cases}$$



Perspective projection

- A model of pinhole camera (no lens)
- Nonlinear equation
- division by z

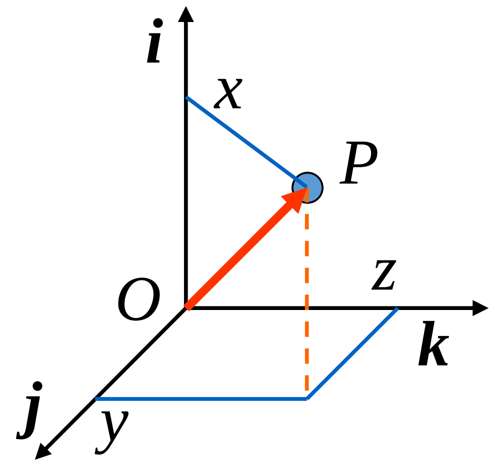
$$\begin{cases} x' = f'\dfrac{x}{z} \\ y' = f'\dfrac{y}{z} \end{cases}$$



Euclid coordinate system

- Orthonormal（正規直交） coordinate frame defined by (O, $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$)
  - O: origin of the coordinate system.
  - $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$: basis vectors.

$$\begin{cases} x = OP \cdot \boldsymbol{i} \\ y = OP \cdot \boldsymbol{j} \\ z = OP \cdot \boldsymbol{k} \end{cases} \Leftrightarrow OP = x\boldsymbol{\imath} + y\boldsymbol{j} + z\boldsymbol{k}$$



  - coordinate vector of the point $\boldsymbol{P}$

$$\boldsymbol{P} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathrm{R}^3$$

Homogeneous coordinates

（同次座標系・斉次座標系）

- Add a component equal to 1 to the ordinary coordinate vector $\boldsymbol{P}$.
- Coordinates are expressed by the ratio of the components whose dimension is increased.
- Defined up to scale.
- Q2: What is the merit?

$$\boldsymbol{P} = \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \qquad \boldsymbol{P} \cong \lambda \boldsymbol{P}$$

equivalence relation

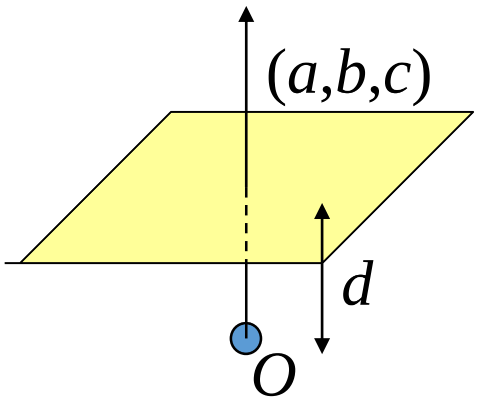
e.g. $$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix} \cong \begin{pmatrix} 2 \\ 4 \\ 6 \\ 2 \end{pmatrix} \cong \begin{pmatrix} -10 \\ -20 \\ -30 \\ -10 \end{pmatrix}$$

Homogeneous coordinate for describing geometric figures

- Plane

$$ax + by + cz + d = 0$$

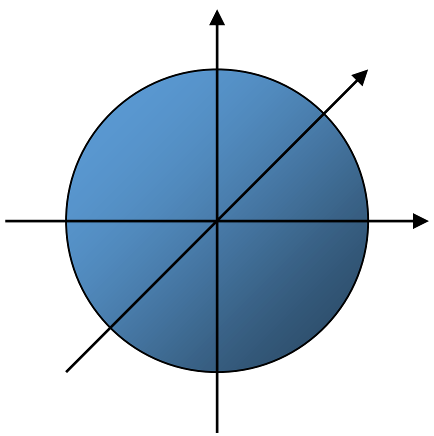
$$(a, b, c, d)(x, y, z, 1)^{\mathrm{T}} = 0$$



- Sphere

$$x^2 + y^2 + z^2 = R^2$$

$$(x, y, z, 1)\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -R^2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = 0$$



Perspective projection equation

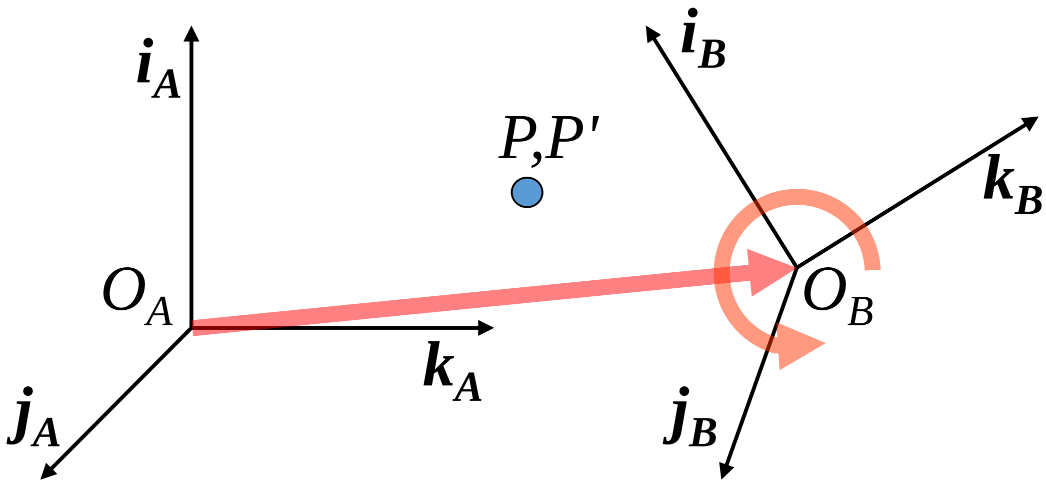
- Non-linear equation by linear algebra.
  - normalization of the fourth component corresponds to division by z.

$$
\begin{cases}
x' = f'\dfrac{x}{z} \\
y' = f'\dfrac{y}{z}
\end{cases}
\qquad
\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} f' & 0 & 0 & 0 \\ 0 & f' & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} f'x \\ f'y \\ z \end{pmatrix}
$$

Rigid transformation

rigid:剛体

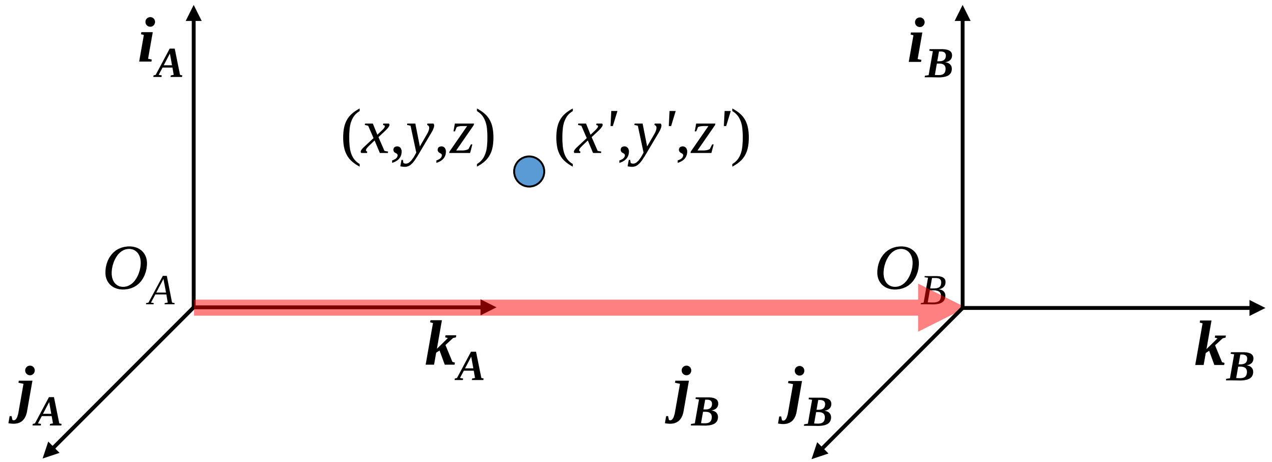
- Rigid transformation = translation + rotation
- Conversion between two different coordinate system from $\boldsymbol{P}$ to $\boldsymbol{P}'$.
  - $\boldsymbol{P}$: defined by the coordinate system $(O_A, \boldsymbol{i}_A, \boldsymbol{j}_A, \boldsymbol{k}_A)$
  - $\boldsymbol{P}'$: defined by the coordinate system $(O_B, \boldsymbol{i}_B, \boldsymbol{j}_B, \boldsymbol{k}_B)$



Translation

- Both coordinate systems are parallel to each other.
- $\boldsymbol{i_A}=\boldsymbol{i_B}, \boldsymbol{j_A}=\boldsymbol{j_B}, \boldsymbol{k_A}=\boldsymbol{k_B}$
- the origins O_A and O_B are distinct.

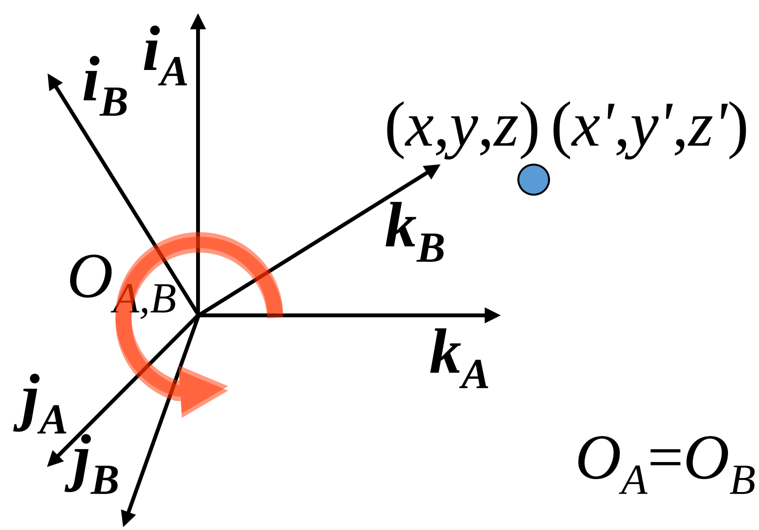
$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \qquad \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$



Rotation

- Origin of the two coordinate systems are coincide.
- Pure rotation around the origin.
- R: 3x3 rotation matrix

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} & & \\ & \boldsymbol{R} & \\ & & \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \boldsymbol{R} = \begin{pmatrix} i_A \cdot i_B & j_A \cdot i_B & k_A \cdot i_B \\ i_A \cdot j_B & j_A \cdot j_B & k_A \cdot j_B \\ i_A \cdot k_B & j_A \cdot k_B & k_A \cdot k_B \end{pmatrix} \qquad \boldsymbol{R}^{-1} = \boldsymbol{R}^{\mathrm{T}}$$



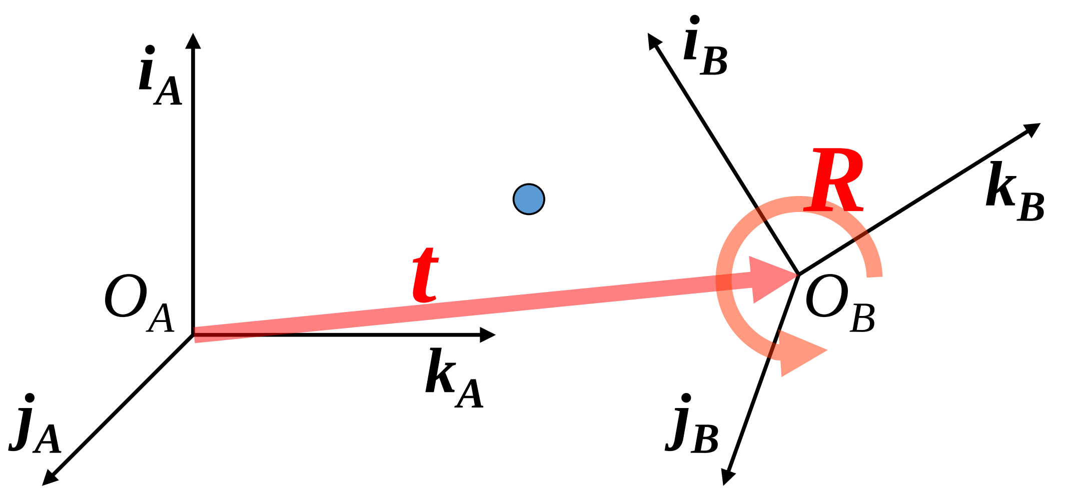
$O_A = O_B$

Translation and Rotation

- Using homogeneous coordinates allows us to write a general change of coordinates as the product of a 4x4 matrix and a 4 vector.

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} & R & \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \qquad \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

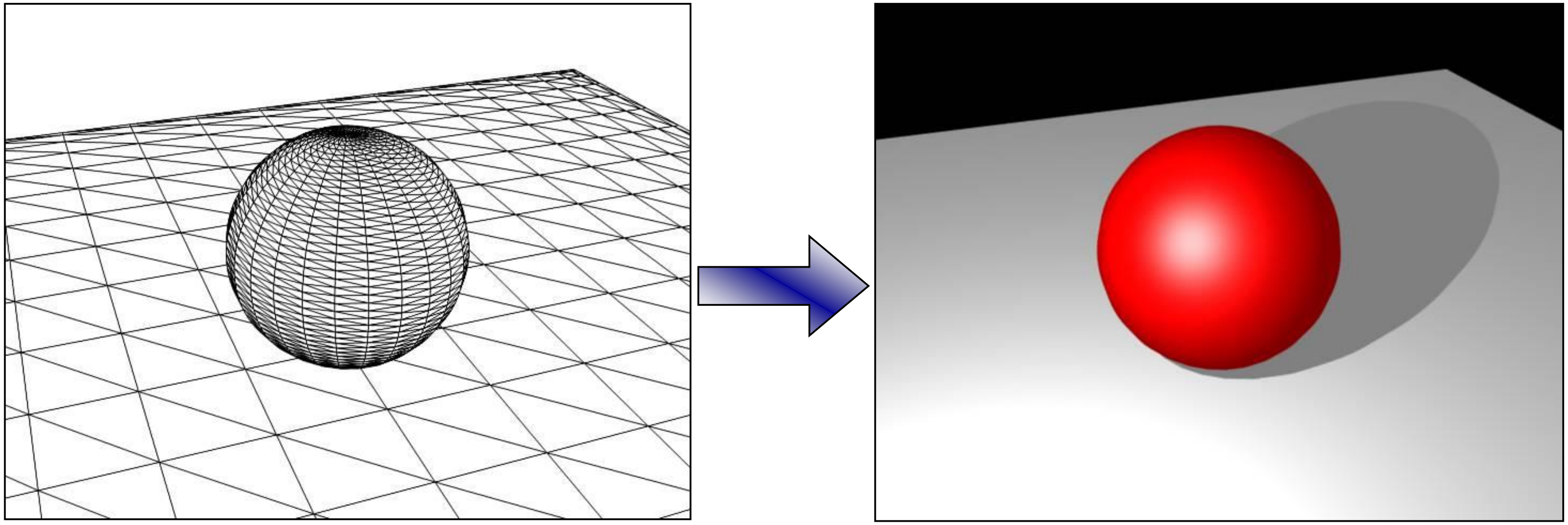
$$\boldsymbol{0} = (0,0,0)^{\mathrm{T}}$$



Photometric conversion
- wavelength to RGB -

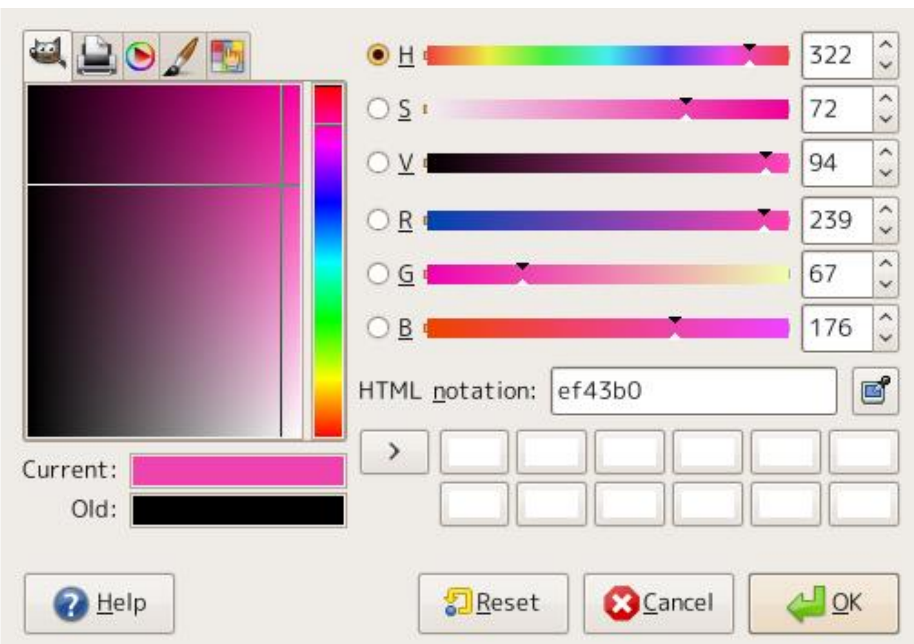
Appearance of object

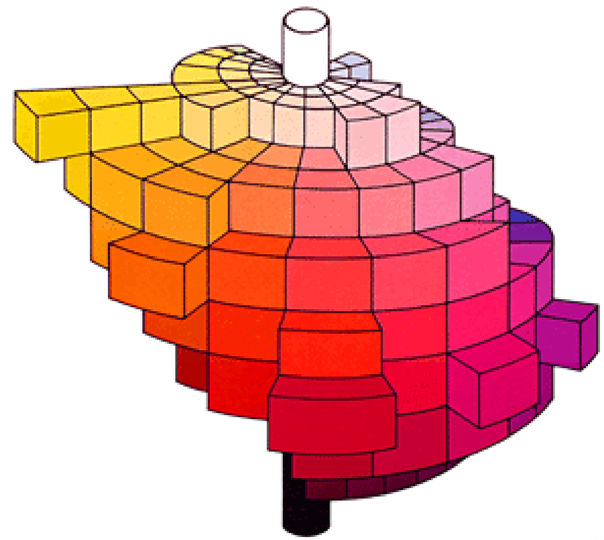
- A sphere is placed on a table.
- The camera position is fixed.
- But there are many appearances.

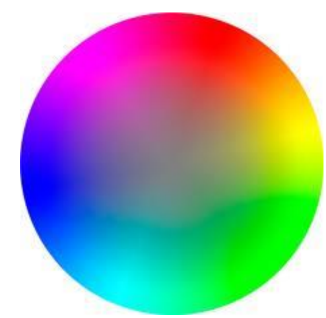


What is Color?

- Human visual perception described through color categories, with names such as red, orange, yellow, green, blue, or purple.
  - What kind of physical phenomenon?
  - How many parameters are necessary to define colors?
  - How do we make color measuring devices?



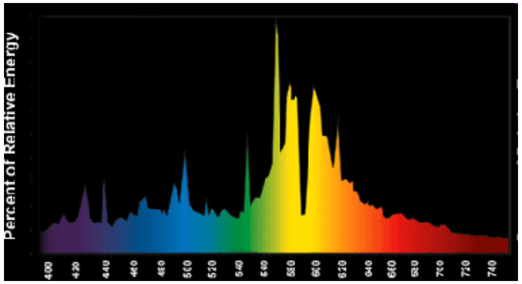




Three Aspects of Color

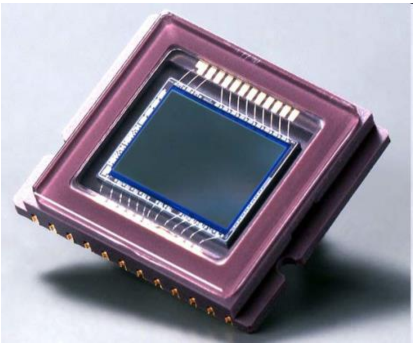
(1) Physics of color

- Spectral quantities



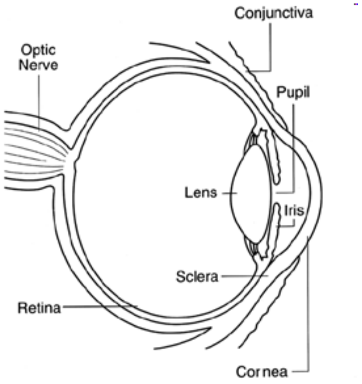
(2) Model for color camera

- Color filter



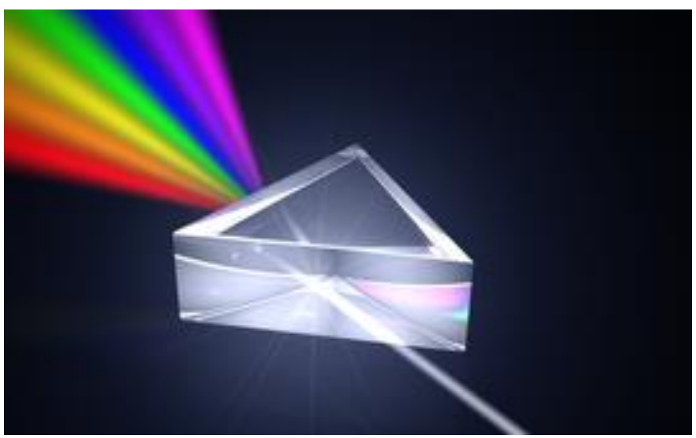
(3) Human color perception

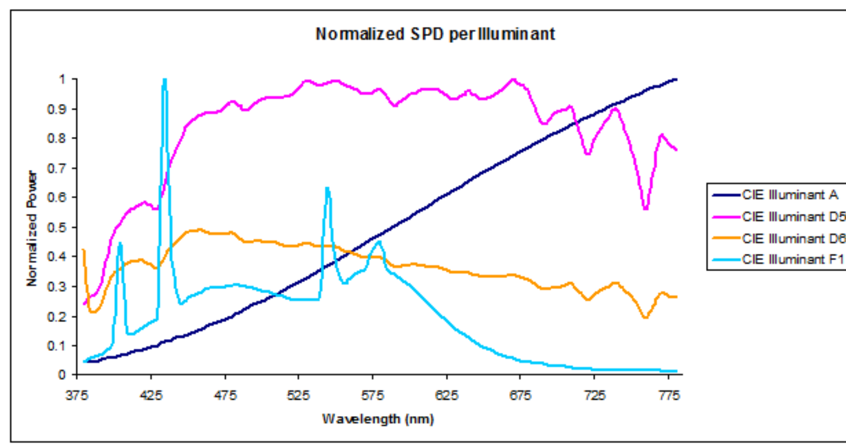
- Color receptors in the eye

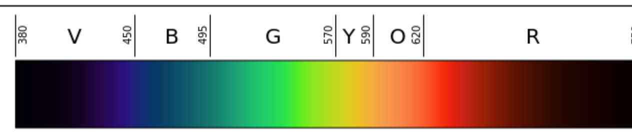


Physics of Color

- Light is electromagnetic wave.
- Human can sense lights between 380nm to 780nm.
- " The rays are not colored." by Sir Isaac Newton.
  「光線には色はついていない」
- Color is not physical quantity but psychological quantity.



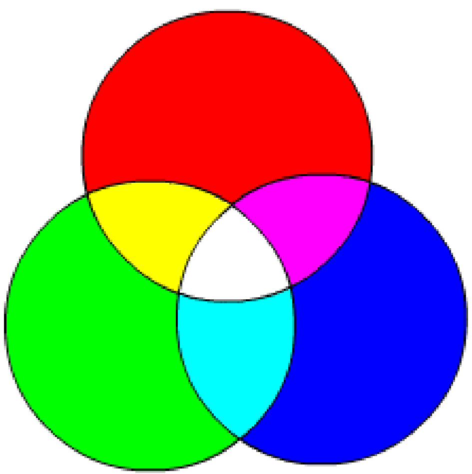


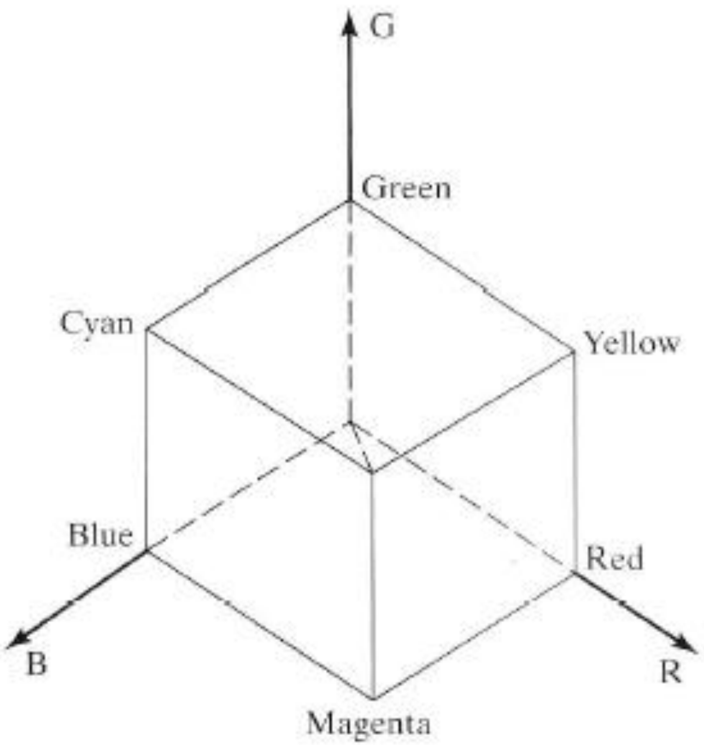


Wikipedia

Linear color spaces: RGB

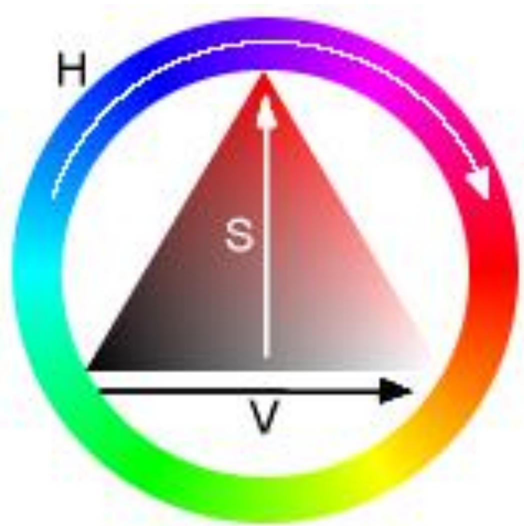
- Invented for practical reasons.
- Single wavelength primaries
  - R: 645.16nm
  - G: 526.32nm
  - B: 444.44nm
- Represented as unit cube

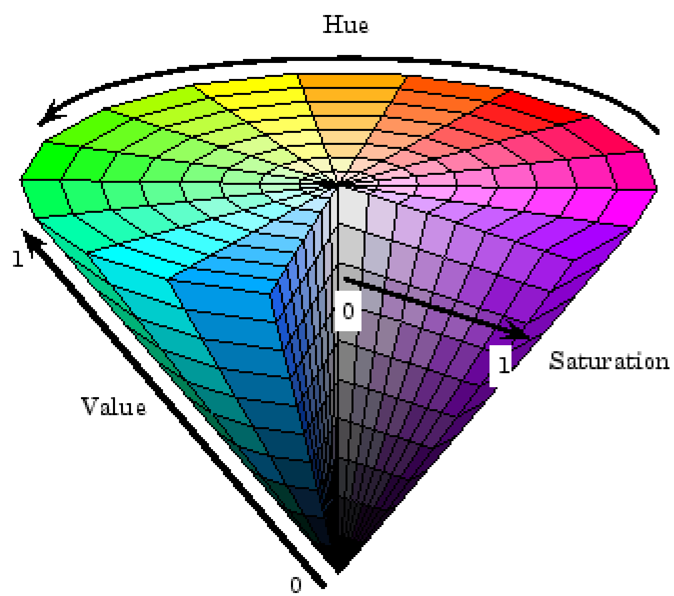




Non-Linear color spaces: Hue, Saturation, and Value

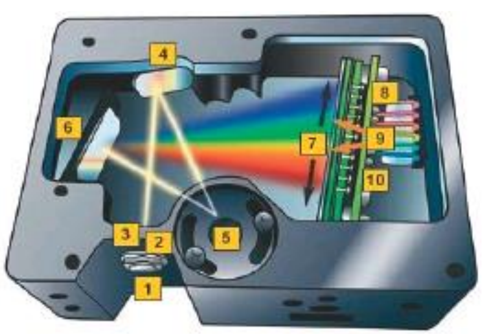
- Representing human intuitions about the topology of colors.
  - hues forms a circle from red through orange to yellow and then green, cyan, blue, purple, and then red again.
- Looking down the center axis of the RGB cube.
  - Hue（色相）, Saturation（彩度）, Value（明度）



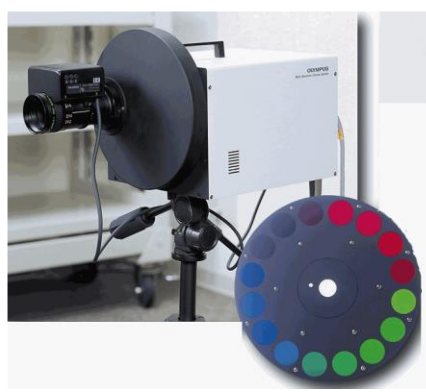


Cameras

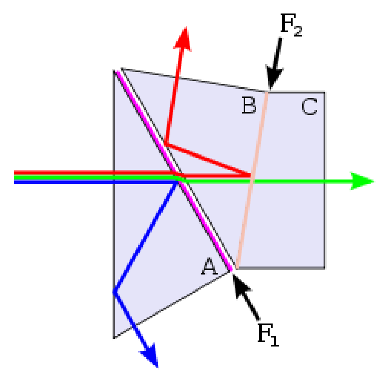
- CCD/CMOS has no function to sense color without color filters.
  - Spectral imaging camera
  - Multi band color camera
  - 3CCD camera for capturing RGB images
  - 1CCD with Bayer pattern



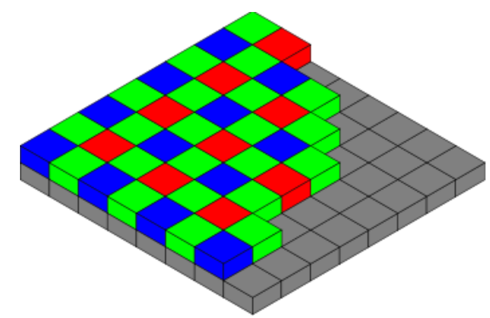
Spectral camera



16-band color camera



3CCD

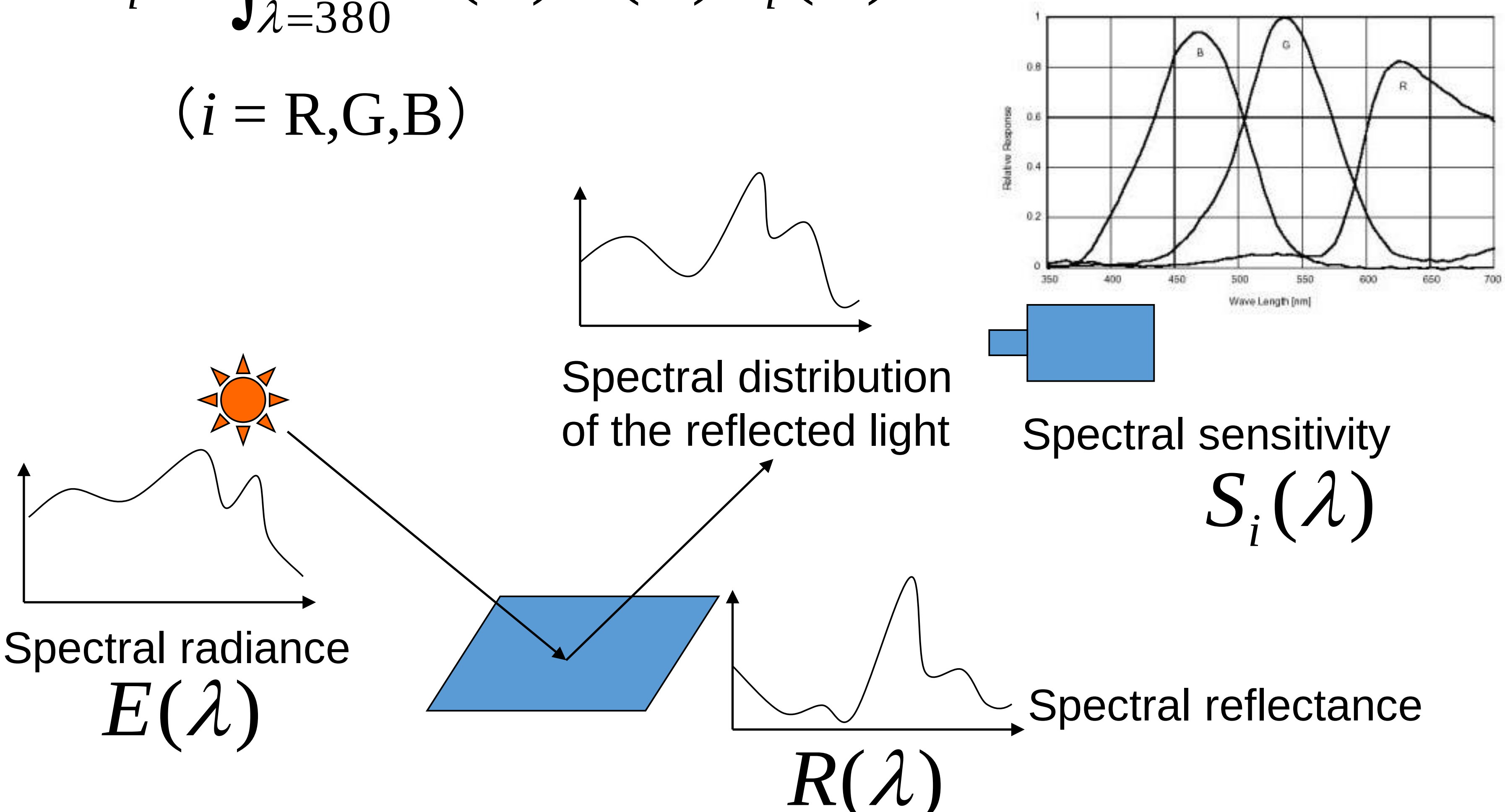


Bayer pattern

Model for Image Color

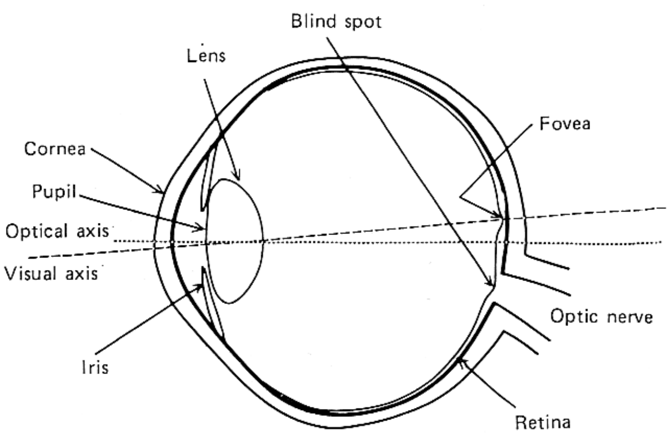
$$V_i = \int_{\lambda=380}^{780} E(\lambda)R(\lambda)S_i(\lambda)d\lambda$$

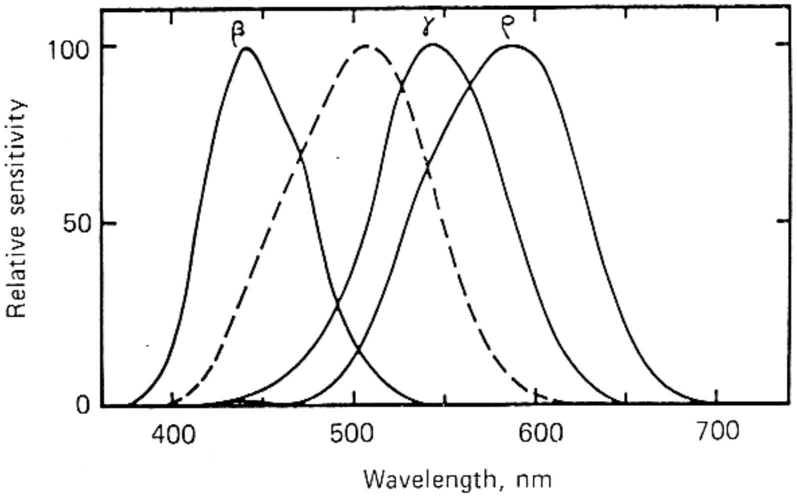
$$(i = \text{R,G,B})$$



Color Receptors

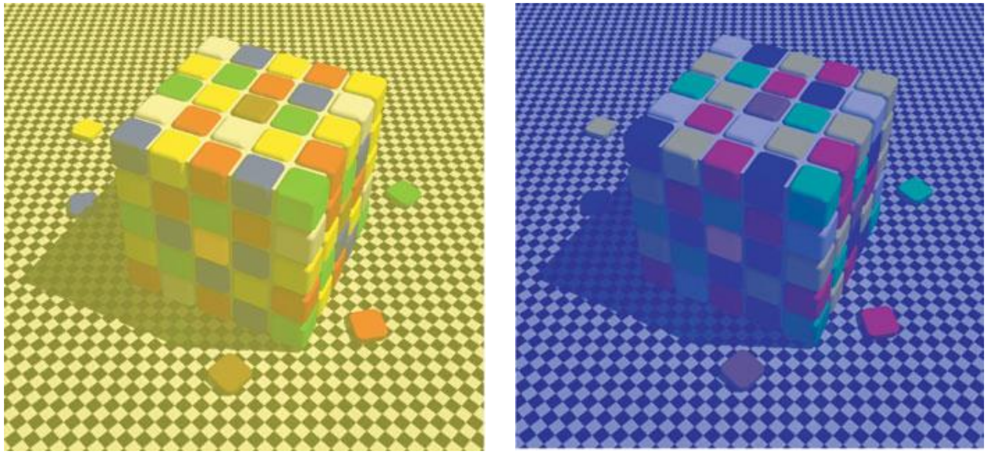
- Trichromacy suggests that there are three distinct types of receptor in the eye.
- Anatomical investigation of the retina shows two types of cell
  - Rods (杆体; かんたい):
    - sensitive in low light, color vision is poor
  - Cones (錐体):
    - three types of cones for color vision





Human color perception

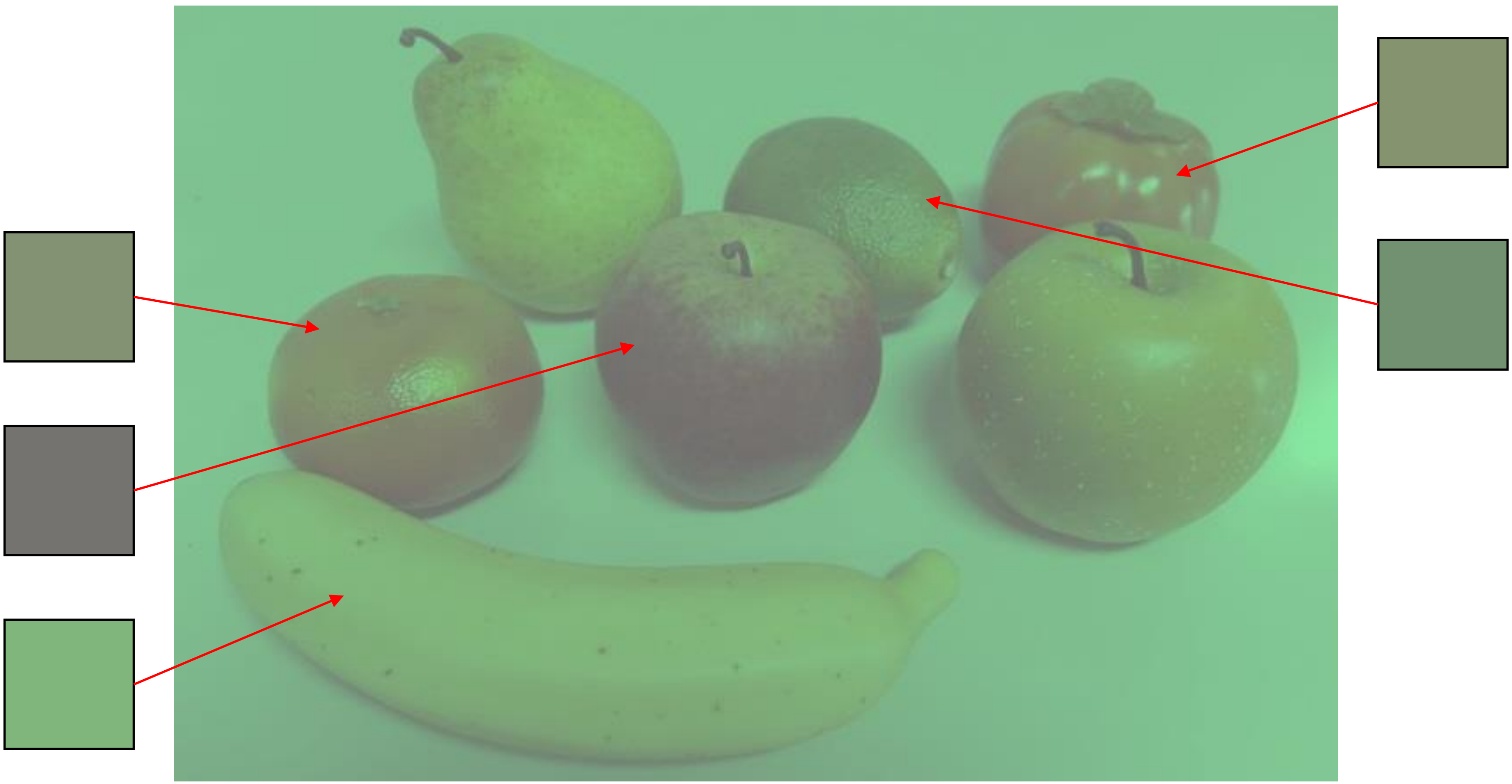
- How people respond to colors?
- Human perception of colors is a complex function of context; illumination, memory, object identity, and emotion.



https://engineering.purdue.edu/~bouman/ece637/notes/ColorConstancy/color/

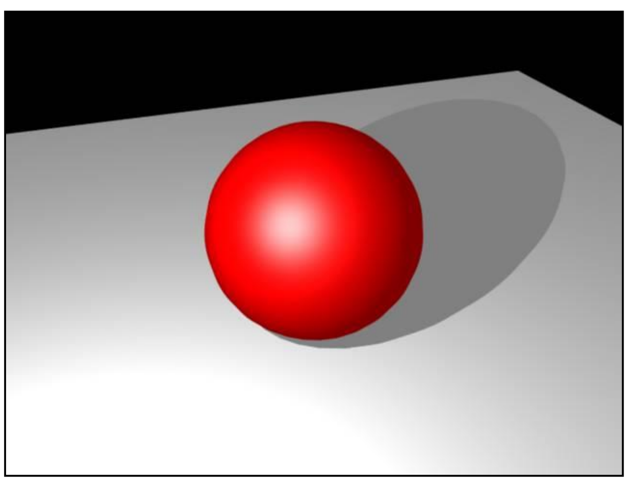
Color constancy（色恒常性）

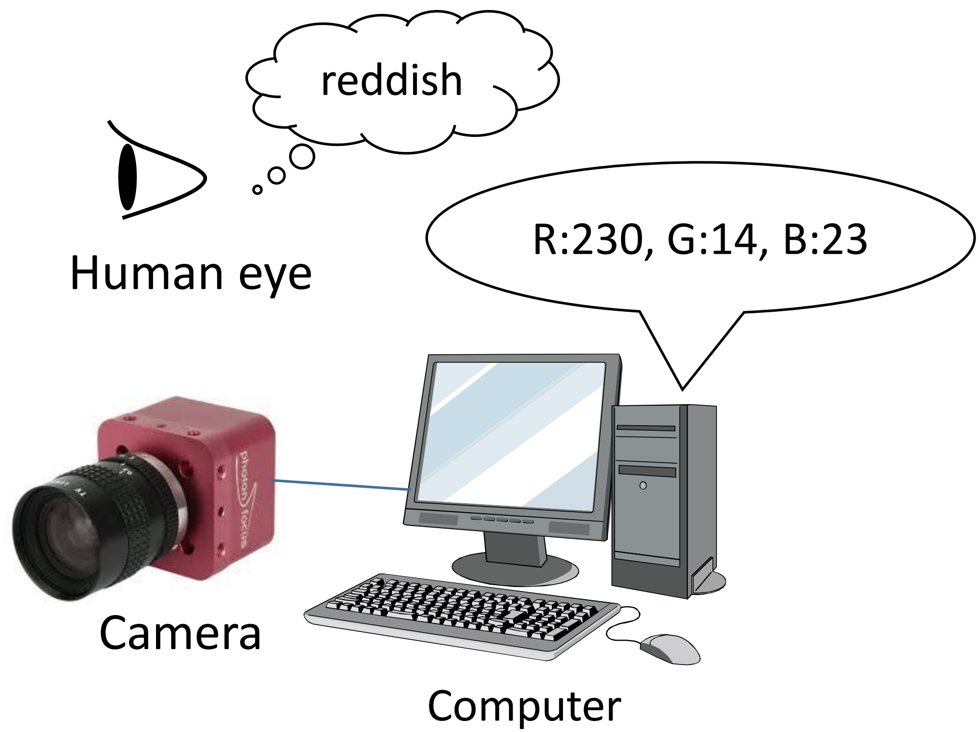
- Advanced visual ability of human that perceive true color under varying illumination conditions.



Human eye vs camera

- Different color perception
  - difficult to be same
  - no need to be same





Q3: Color perception

- Why did people see differently?