Computer Vision (7CCSMCVI / 6CCS3COV)

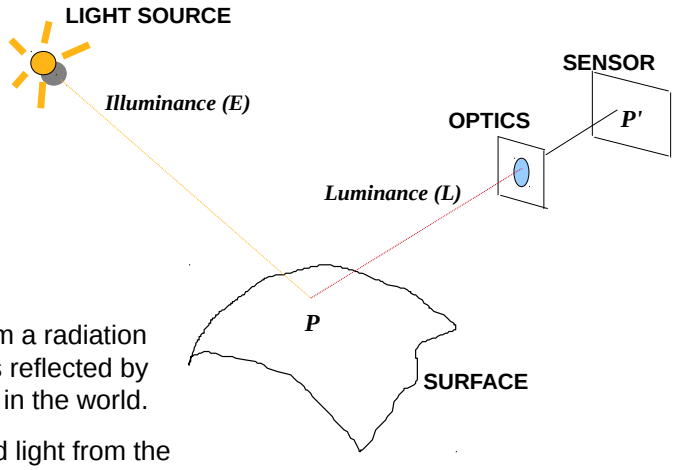
Recap

- **Introduction**
  - vision concerned with determining properties of the world from images.
  - difficult due to the problem being ill-posed (one image can have many interpretations) and being exponentially large (one object can generate many images).
  - overcoming these problems requires combining evidence obtained from the image with prior knowledge in order to make inferences about image content.
- **Image formation** ←Today
- **Low-level vision**
- **Mid-level vision**
- **High-level vision**

Today

- Physics of image formation (light, reflectance, optics)
- Geometry of image formation (camera models, projective geometry)
- Digital images (digitisation, representation)
- The Eye

Overview of image formation



Light from a radiation source is reflected by surfaces in the world.

Reflected light from the world hits the sensor.

Images are formed when a SENSOR registers RADIATION that has interacted with PHYSICAL OBJECTS

Ingredients of image formation

The resulting image is affected by two sets of parameters:

Radiometric parameters

Determine the intensity/colour of a given image pixel

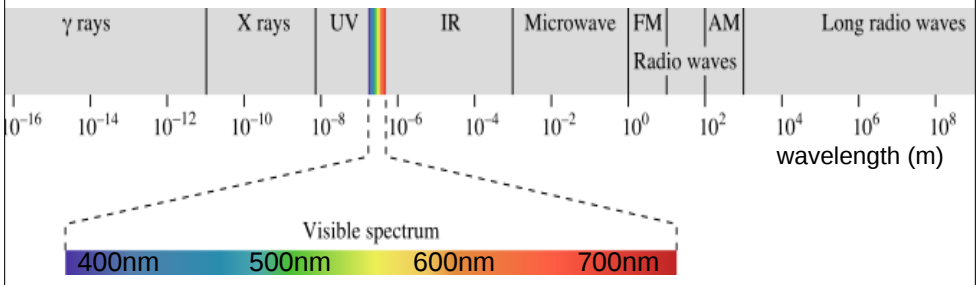
- illumination (type, number, location, intensity, colour-spectrum)
- surface reflectance properties (material, orientation)
- sensor properties (sensitivity to different electromagnetic frequencies)

Geometric parameters

Determine where on the image a scene point appears

- camera position and orientation in space
- camera optics (e.g. focal length)
- projection geometry (mapping from 3D to 2D)

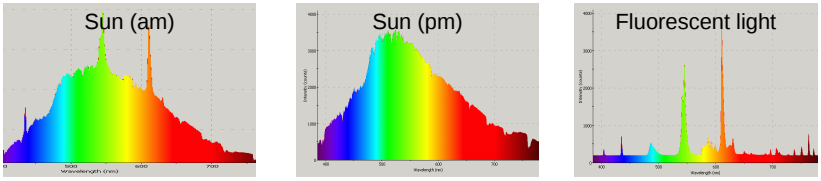
Light and Colour



- At the earth's surface the intensity of the electromagnetic radiation emanating from the sun has a peak within the 400-700nm range.
- The human eye has evolved a specific sensitivity to this part of the electromagnetic spectrum.
- Hence, visible light is that part of the electromagnetic spectrum with a wavelength (λ) between 400 and 700nm.
- Cameras and Computer Vision systems also concentrate on this part of the spectrum (but not exclusively, e.g. infra-red cameras for detecting body heat).

Colour perception

Light is produced in different amounts at different wavelengths by each light source.



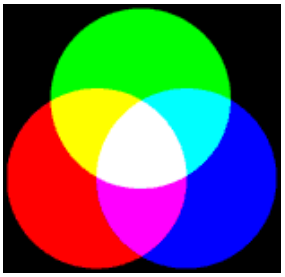
Light is differentially reflected at each wavelength, which gives objects their natural colours (albedo = fraction of light reflected at a particular wavelength).

Colour mixing

Mixing light: additive

e.g. a green light plus a blue light plus a red light gives light containing a broad spectrum of light, i.e. white.

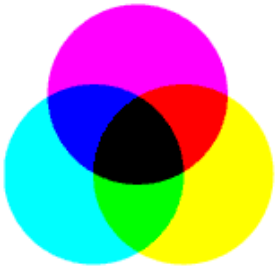
The illumination from different light sources adds.



Mixing pigments: subtractive

e.g. a green pigment plus a blue pigment plus a red pigment gives a pigment that **absorbs** light over a broad spectrum, leaving black.

The reflection from different surfaces subtracts.



Measuring surface properties

The radiation that drives the human construct of colour, is fundamentally colour less.

The sensation of colour is determined by the human visual system, based on the product of light and reflectance.

Luminance (L) amount of light striking the sensor depends on **Illuminance** (E) amount of light striking the surface as well as **Reflectance** (R) which depends on material properties

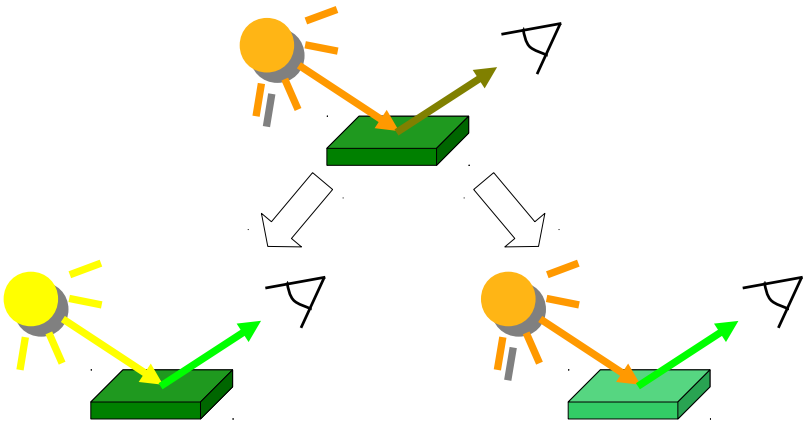
$$L(x,y,\lambda) = fn\,(E(x,y,\lambda), R(x,y,\lambda))$$

Intensity at a particular location and wavelength

값을 모르면 ill-posed

To determine properties of objects in the world (e.g. their colours), we need to recover R, but we don't know E so this problem is **ill-posed**.

Measuring surface properties



intensity = illumination

If colour/intensity of light source changes, colour/intensity of reflected light also changes (i.e. L varies with E)

If colour/reflectance of surface changes, colour/intensity of reflected light also changes (i.e. L varies with R)

Colour constancy: artificial

To recover the surface colour of a particular location, R(x,y,λ), we need to know the colour of the illumination at that point, E(x,y,λ).

Many ways of approximating E have been suggested:

- Average reflectance across scene is known (often fails)
- Fixing brightest image patch to be white
- Gamut (collection of all colours) falls within known range
- Known reference colour (colour chart, skin colour…)
- Specular reflections have the colour of the illumination

None of these are work particularly well.

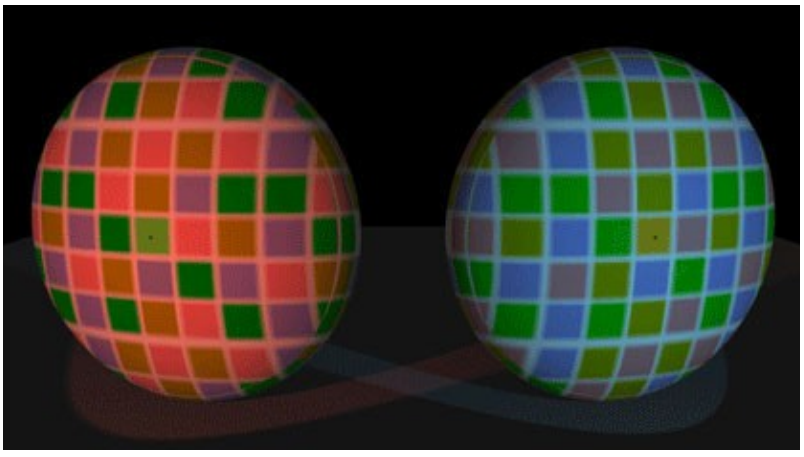
Colour constancy: biological

However, the human visual system does seem able to recover surface colour, since despite large changes in illumination (and consequently the intensity spectrum that enters our eyes), we usually experience the colour of an object as being constant.

artificial light hazy daylight clear blue sky

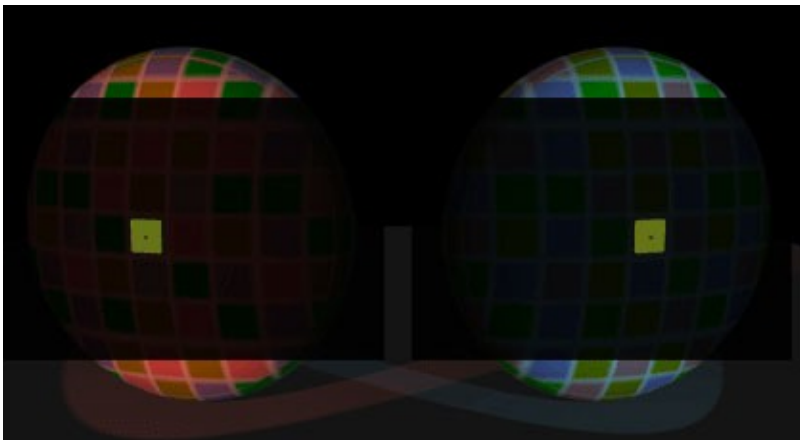
We are not normally aware of this variation because colour constancy mechanisms discount the effects of illumination.

Effects of inference (illumination)



Are the central patches the same colour?

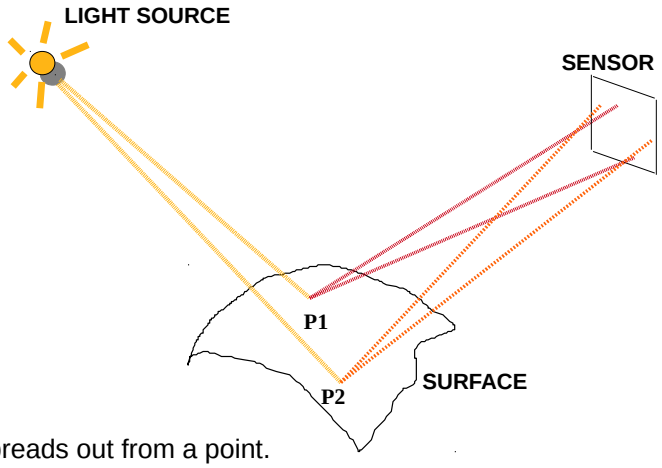
Effects of inference (illumination)



Are the central patches the same colour?

Visual system sees them as different due to inference about different lighting conditions

Focusing Light

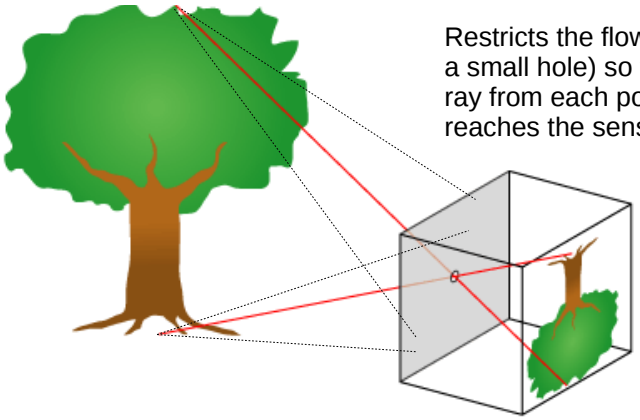


Light spreads out from a point.

Without some kind of optics each location on the sensor will register light coming from many different points in the world.

No image will be formed.

Pinhole camera



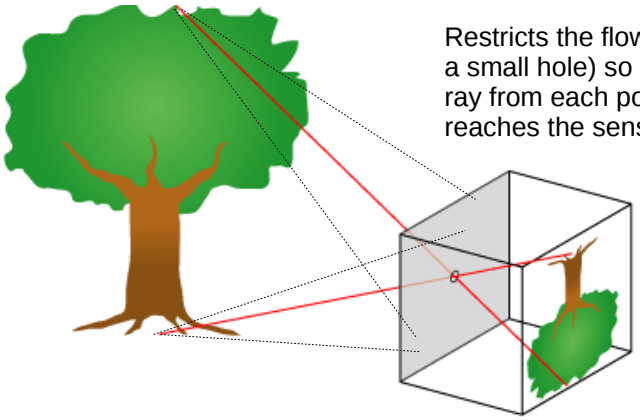
Restricts the flow of light (using a small hole) so that only one ray from each point in the world reaches the sensor.

Note: image inverted

"Focus" means that all rays coming from a scene point converge into a single image point.

"Exposure" is the time needed to allow enough light through to form an image (the smaller the aperture, the longer the exposure time). The longer the exposure the more blurred an image is likely to be.

Pinhole camera



Restricts the flow of light (using a small hole) so that only one ray from each point in the world reaches the sensor.

Note: image inverted

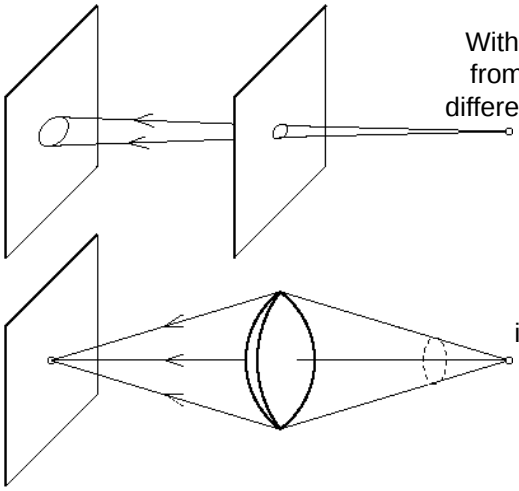
Small pinhole: sharp focus but dim image (long exposure time)

Large pinhole: brighter image (shorter exposure) but blurred

To produce an image that is both bright and in focus requires a lens

Pinhole 로 맞추기 힘드니깐 렌즈가 필요하다는 소리

Lensed camera



With a large pinhole, light rays from the same point project to different locations on the image. The image is blurred.

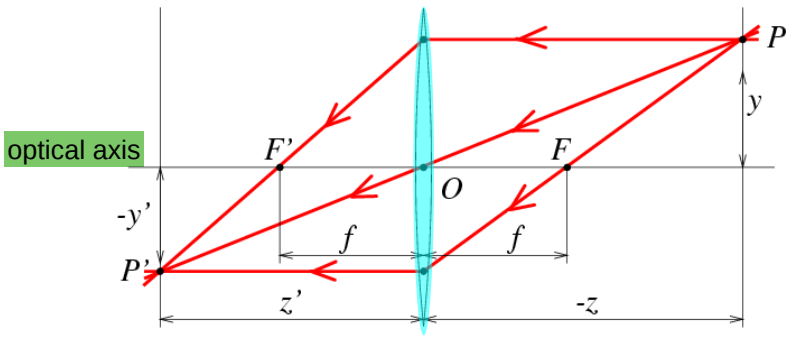
By focusing rays from the same point onto a single image location, a lens can keep the image sharp while gathering more light.

Cost: image focused for only a restricted range of object positions

Thin lenses

A Lens works by refracting light.

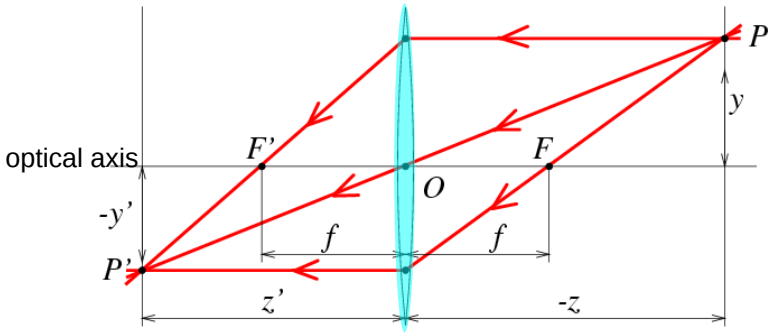
렌즈들은 빛을 굴절시키는 작용을 한다.



F' = focal point, f = focal length

- Rays passing through the centre point O are not refracted
- Rays parallel to the optical axis are refracted to pass through the focal point.
- Rays passing through point F are refracted to be parallel to the optic axis.

Thin lenses

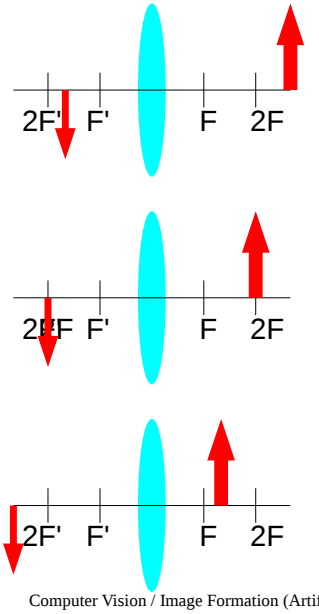


An object located at distance z from the lens has an image formed at depth at z' from the lens according to the "thin lens equation":

$$\frac{1}{f}=\frac{1}{|z|}+\frac{1}{|z'|}$$ location

Hence, depth of the image depends on the depth of the object as well as the focal length of the lens.

Depth of focus



$$\frac{1}{f}=\frac{1}{|z|}+\frac{1}{|z'|}$$

For a lens with a fixed focal length, the depth of the image plane required to bring an object into focus varies inversely with the depth of the object.

At the extremes: if the object is at infinity the image is formed at F', if the object is at F (or closer) no image can be formed.

For a short focal length camera, almost all objects will be located more than 2 focal lengths from the lens (top figure)

Placing the receptor plane at a depth in the range z' ∈ [F', 2F'] will provide an acceptable image for objects in a wide range of depths greater than 2F.

Focal range

Focal range is defined by the range of object locations such that blurring due to the difference between the receptor plane and the focal plane is less than the resolution of the receptor device.

Decreasing the aperture size increases the focal range, but it decreases the amount of light available to the receptor.

As the aperture decreases to a pinhole we recover the infinite focal range of the pinhole camera!

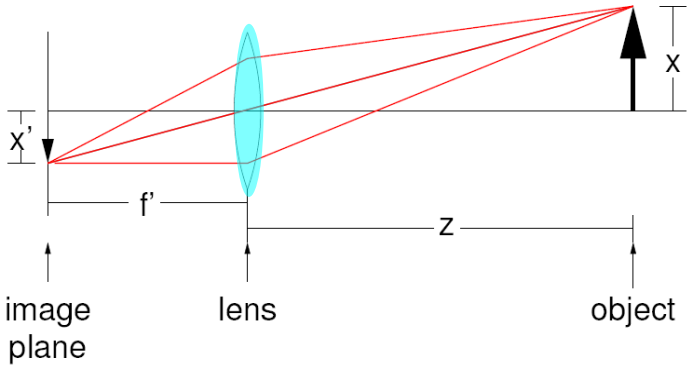
Geometric camera models

Given the coordinates of a point in the scene, what are the co-ordinates of this point in the image?

i.e. given P = (x,y,z) how do we calculate P'=(x',y',z')? How

To answer, we need a mathematical model of the geometric projection implemented by the camera, often called simply a camera model.

Pinhole (perspective) camera model



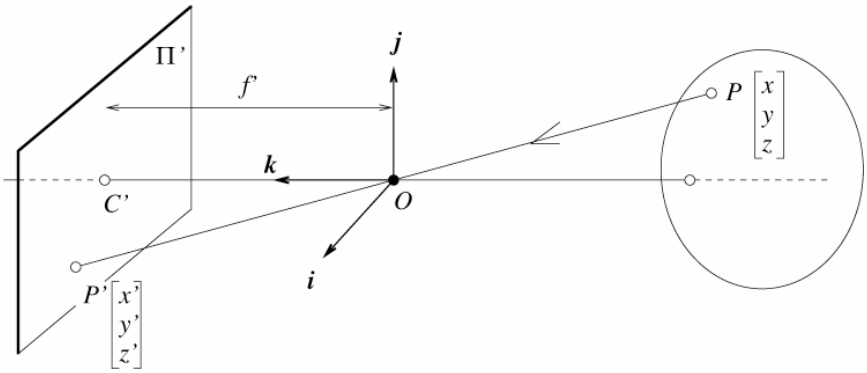
A lens follows the pinhole model for objects that are in focus.

The pinhole camera is therefore an acceptable mathematical approximation (i.e. a model) of image formation in a real camera.

For a pinhole camera everything is in focus regardless of image plane depth

Pinhole model arbitrarily assumes that image plane is at distance f '

Perspective camera model



P: a scene point with coordinates (x,y,z), P': its image with coordinates (x', y',z')

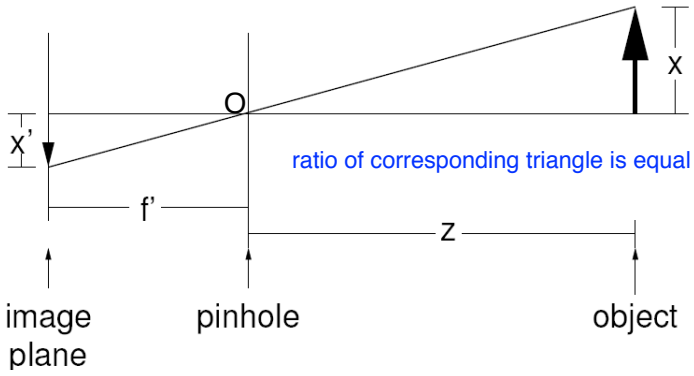
O: origin (pin hole / centre of lens)

Image plane Π' is located at a distance f' from the pinhole along the vector k

optical axis: line perpendicular to image plane and passing through O

C': image centre (intersection of optical axis and image plane)

Equation of projection (2D)



From similar triangles:

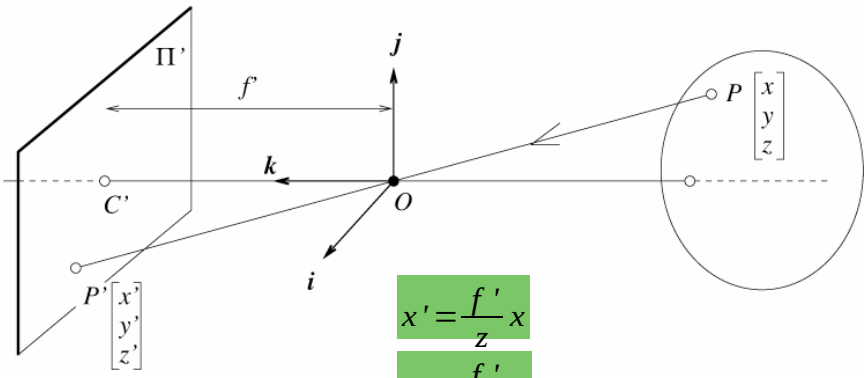
$$x'/x = f'/z$$

Hence:

$$x' = (f'/z)\, x$$

z : distance

Equation of projection (3D)



$$x' = \frac{f'}{z} x$$

Similarly, for y coordinates:

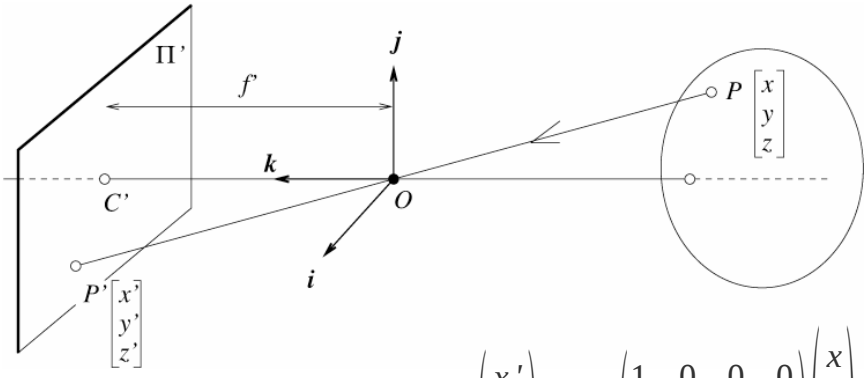
$$y' = \frac{f'}{z} y$$

Since P' lies on image plane:

$$z' = f'$$

All coordinates are relative to camera reference frame [mm] – i.e. axes rigidly attached to camera with origin at pinhole

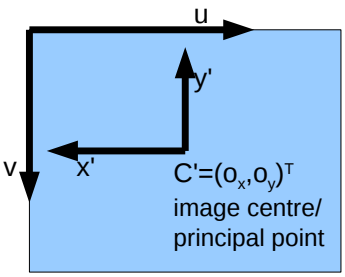
Equation of projection (3D)



We can re-write these equations in matrix form using homogeneous coordinates:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \frac{f'}{z} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection operator}} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Relating pixels to camera coordinates



To convert from camera reference frame [mm] to image reference frame [pixel], we need to take account of:

- Origin of image (in corner, not centre)
- Pixel size (s_x, s_y [mm/pixel])

$$u = \frac{-x'}{s_x} + o_x = \frac{-f'x}{s_x z} + o_x = \alpha \frac{x}{z} + o_x$$

$$v = \frac{-y'}{s_y} + o_y = \frac{-f'y}{s_y z} + o_y = \beta \frac{y}{z} + o_y$$

s : scale factors

Where:

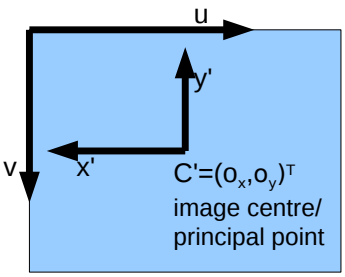
- u, v are integer image coordinates [pixel]
- α, β magnification factors in x' and y' directions ($\alpha = -f'/s_x$, $\beta = -f'/s_y$)

o : offset of center of image

o_x,o_y, α, β are the 4 intrinsic camera parameters

α/β = aspect ratio of camera

Relating pixels to camera coordinates



We can re-write these equations in matrix form:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \underbrace{\begin{pmatrix} \alpha & 0 & o_x \\ 0 & \beta & o_y \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic camera parameters}} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection operator}} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

u, v 를 구할 수 있다.

This is only an approximation, due each individual camera having small manufacturing errors (i.e. the image plane may not be perfectly perpendicular to the optical axis, or may be rotated slightly about the optical axis ('skew'))

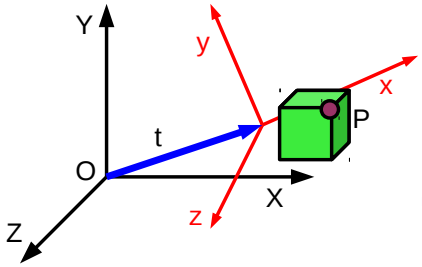
Relating pixels to world coordinates

So far it has been assumed that the location of the scene point $P=(x,y,z)^T$ is given in camera coordinates.

However, more generally scene points may be known relative to some external reference frame.

An external reference frame is especially useful when:

- the camera is moving
- multiple cameras are used (e.g. for stereopsis)



To convert coordinates from one reference frame to another requires a translation and a rotation.

빨간색 : axis

Relating pixels to world coordinates

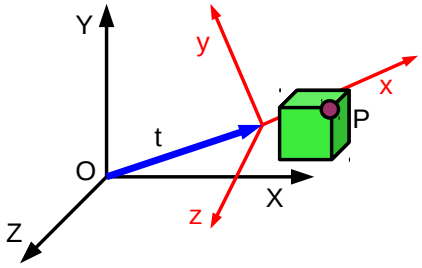
Coordinates of point in frame O_{XYZ}

Rotation matrix describing orientation of frame O_{xyz} with respect to frame O_{XYZ}

$$P = t + R\,p$$

Coordinates of origin of frame O_{xyz} (in frame O_{XYZ})

Coordinates of point in frame O_{xyz}



World reference frame O_{XYZ}

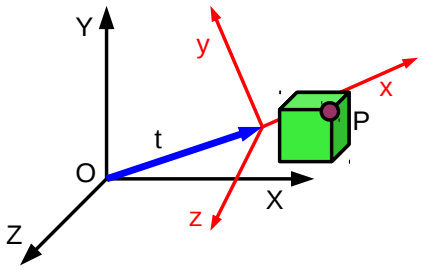
Camera reference frame O_{xyz}

R and t are the rotation matrix and translation vector going from O_{XYZ} to O_{xyz}

Relating pixels to world coordinates

We can re-write these equations in matrix form:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$



Where:

$\boldsymbol{R}$ is a 3x3 rotation matrix

$\boldsymbol{t}$ is a 3x1 translation vector

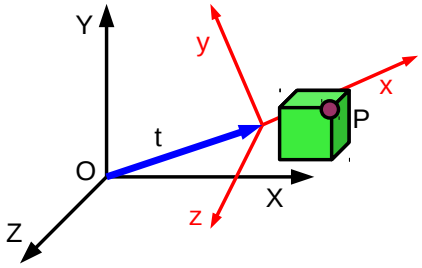
$\boldsymbol{0}$ is a 3x1 vector of zeros

Relating pixels to world coordinates

To go from the coordinates of a point in the world reference frame to the camera reference frame we require the inverse mapping:

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}^T & -\boldsymbol{R}^T\boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

reverse



Relating pixels to world coordinates

The complete transformation is thus:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \underbrace{\begin{pmatrix} \alpha & 0 & o_x \\ 0 & \beta & o_y \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic camera parameters}} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection operator}} \underbrace{\begin{pmatrix} \boldsymbol{R}^T & -\boldsymbol{R}^T \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{pmatrix}}_{\text{extrinsic camera parameters}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

intrinsic camera parameters	projection operator	extrinsic camera parameters
2D → 2D	3D → 2D	3D → 3D
Maps points on the image plane into pixel image coordinates	Projects points in the camera reference frame onto the image plane	Transforms the world coordinates to the camera reference frame

Camera Parameters

- **Intrinsic** parameters
  - Depend on properties of camera
    - » focal length, sensor dimensions/resolution
- **Extrinsic** parameters
  - Depend on the camera location
    - » Translation and Rotation of camera relative to world

Mapping between 2D and 3D

Given a point in 3D space we can model where that point will appear in a 2D image.

» A well-posed, forward problem

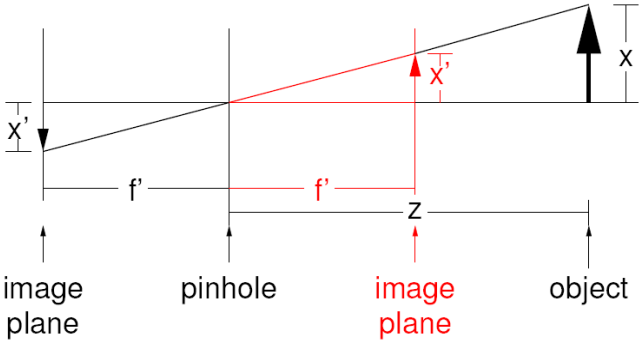
However, given a point in a 2D image we cannot determine the corresponding point in 3D space (depth information has been lost).

» An ill-posed, inverse problem

To recover depth, need extra information – e.g.

- another image (the need for stereo vision), or
- prior knowledge about the structure of the scene.

Virtual image



A pinhole camera creates inverted images.

It is traditional to draw the image plane in front of the pinhole at the same distance from it as the actual image plane.

Resulting "virtual" image is identical to the real image except it is the right way up.

This does **not** change the mathematics of the perspective camera model.

Projective geometry

Euclidean geometry describes objects "as they are".

It describes transformations within a 3D world (i.e. translations and rotations)

These mappings do not change the shape of an object (i.e. lengths, angles, and parallelism are preserved).

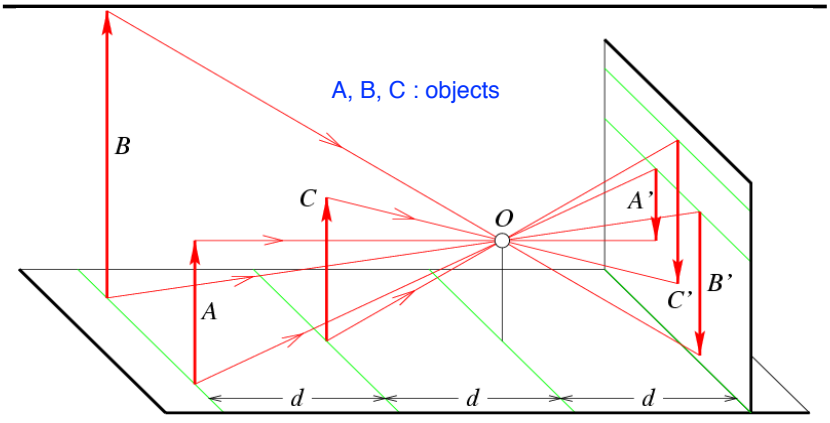
Projective geometry describes objects "as they appear".

It describes the transformation from the 3D world to a 2D image (i.e. scaling and shear in addition to translations and rotations)

This mapping does distort the shape of an object (i.e. lengths, angles, and parallelism are not preserved).

Some properties of (i.e. distortions caused by) projective geometry are described on the following slides...

Distant objects appear smaller



The apparent size of an object depends on its distance

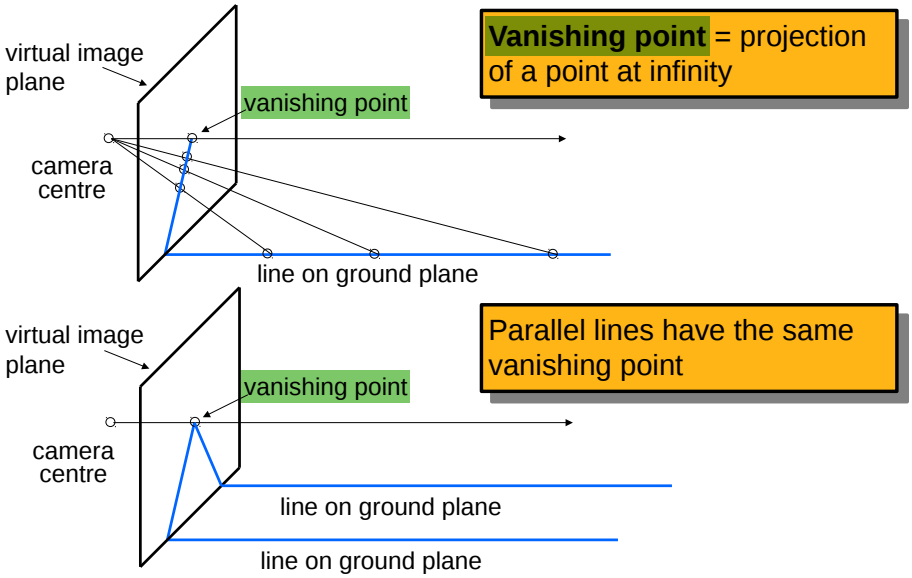
In world:

A and C have equal length, B twice this length

In image:

B' and C' have equal length, A' half this length

Vanishing points



Vanishing point = projection of a point at infinity

Parallel lines have the same vanishing point

Example of projective distortion

converging at a vanishing point

Reality is distorted by projection:
In world:

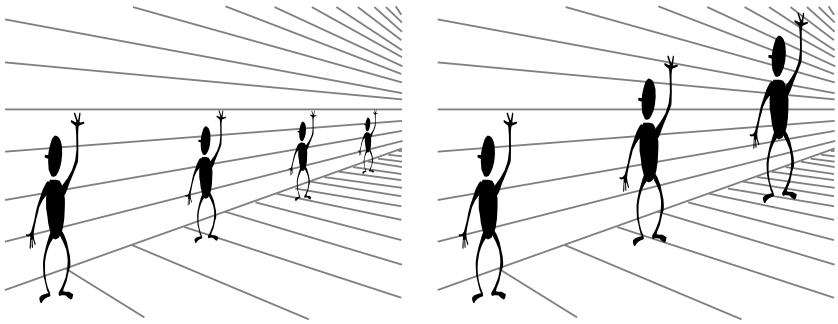
- rail tracks parallel
- ties perpendicular to tracks
- ties equal in length
- ties evenly spaced

In image:

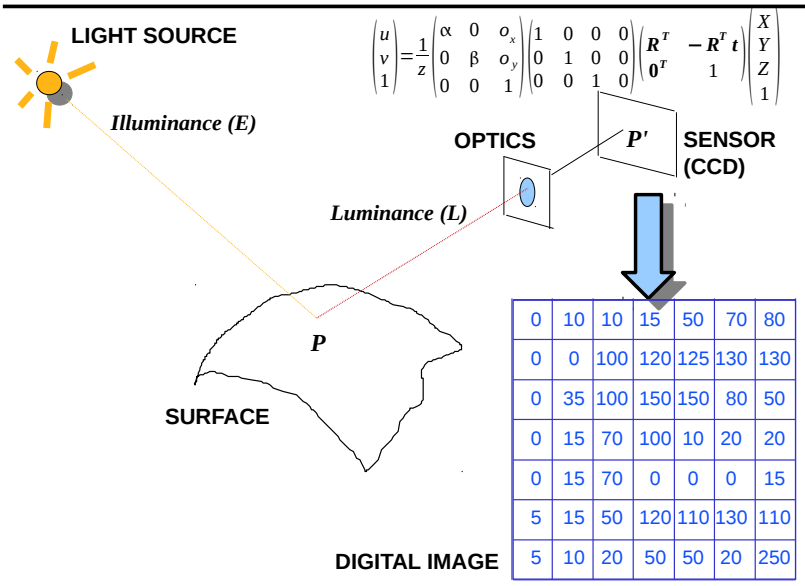
- rail tracks converge at a vanishing point
- ties not at right angles to track
- ties get shorter with distance
- ties get closer together at longer distances.

Vanishing points provide cues to size

Our visual system is good at detecting vanishing points and using this to extract information about object size.
Can be used in computer vision too.



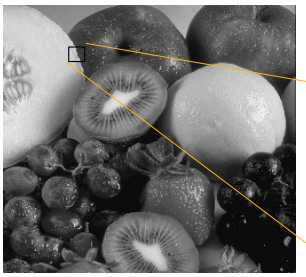
Digital image formation



$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \begin{pmatrix} \alpha & 0 & o_x \\ 0 & \beta & o_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R^T & -R^T t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

0	10	10	15	50	70	80
0	0	100	120	125	130	130
0	35	100	150	150	80	50
0	15	70	100	10	20	20
0	15	70	0	0	0	15
5	15	50	120	110	130	110
5	10	20	50	50	20	250

Digital image representation (greyscale)



A digital image is a 2D array (matrix) of numbers:

	x = 58	59	60	61	62	63	64
y = 41	7	10	10	15	50	70	80
42	1	0	67	123	25	30	130
43	2	35	100	150	150	80	50
44	8	15	70	100	10	20	20
45	12	15	76	5	17	0	15
46	5	15	50	120	110	130	110
47	5	10	20	50	50	20	250

The scene, which is a continuous function, is sampled at discrete points (called *picture elements* or *pixels* for short).

Value of each pixel = measure of light intensity at that point.

Typically:
0 = black
255 = white
integers

Or:
0 = black
1 = white
floats

Image axes and notation

A digital image is usually denoted as I.
- I is a matrix of pixel values
- Origin is top left corner

A point on the image, $p = (x,y)^T$
- p is a vector

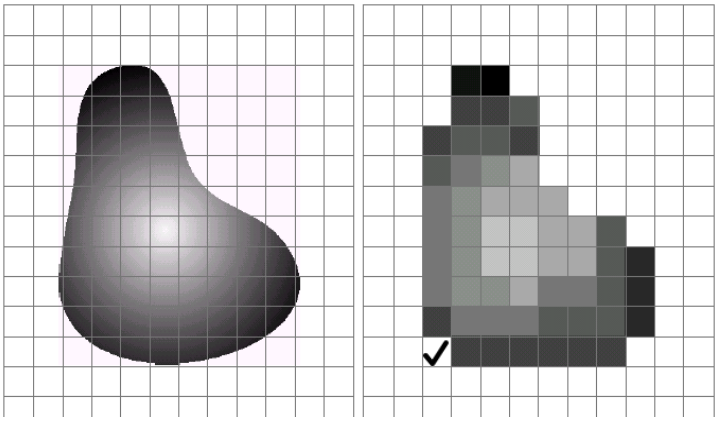
A pixel value is $I(p)$ or $I(x,y)$

Note: in mathematics (and Matlab) $I(r,c)$ would refer to the r^{th} row and the c^{th} column of I, so I is the transpose of a matrix in standard format.

	x = 58	59	60	61	62	63	64
y = 41	7	10	10	15	50	70	80
42	1	0	67	123	25	30	130
43	2	35	100	150	150	80	50
44	8	15	70	100	10	20	20
45	12	15	76	5	17	0	15
46	5	15	50	120	110	130	110
47	5	10	20	50	50	20	250

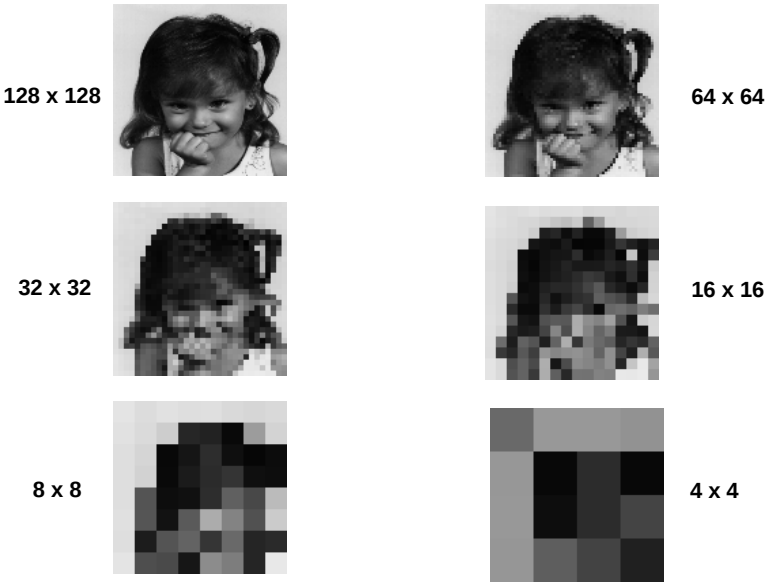
이미지를 보기 위해서는 transpose 시켜야 된다.

Pixelisation and Quantization



Pixelisation: intensity values averaged at each sampling point (i.e. within each grid location in the sensor array).
Quantization: intensity values rounded to the nearest integer

Effects of pixelization



Effects of quantization

256 gray levels (8 bits/pixel)
16 gray levels (4 bits/pixel)
8 gray levels (3 bits/pixel)
4 gray levels (2 bits/pixel)

2 gray levels (1 bit/pixel)
BINARY IMAGE

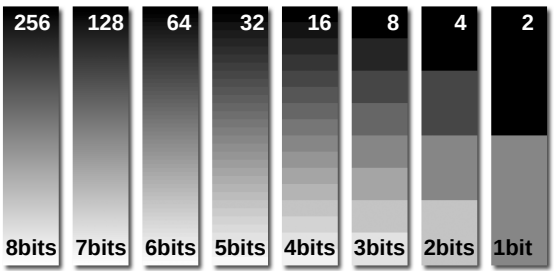


Image formats

Binary or Monochrome:
- 1 binary value per pixel, 1-bit ⇒ 2 discrete levels.
- I(p) = 0 or 1.

Greyscale or Intensity:
- 1 real value per pixel, typically 8-bit ⇒ 256 discrete levels.
- I(p) ∈ [0,1]

Colour:
- 3 real values per pixel, i.e. 3 colour channels, e.g. RGB
- $I_R(p) \in [0,1]$, $I_G(p) \in [0,1]$, $I_B(p) \in [0,1]$
- Each colour channel is a greyscale image representing the intensity of light at a particular wavelength (R=645.2nm, G=526.3nm, B=444.4nm)
- 24-bit 'True Color' ⇒ 8-bits for each colour.

Colour image representation (RGB)

I_R I_G I_B

Note: many other formats are used to represent colour images.

Switching between formats

RGB to Greyscale

Take weighted average over three colour channels:

$$I_{grey} = \frac{rI_R + gI_G + bI_B}{r+g+b}$$

where r, g, b are three weighting factors (if r=g=b, this is simply the mean).

Greyscale to binary

Apply a threshold t:

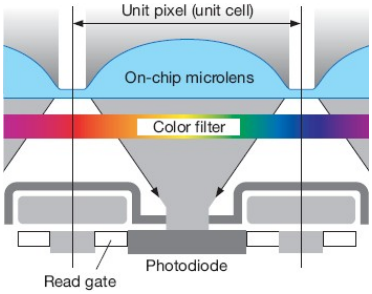
$$I_{bin}(p) = \begin{cases} 1 & \text{if } I_{grey}(p) \geq t \\ 0 & \text{otherwise} \end{cases}$$

Charge Coupled Device (CCD) Cameras

- A semiconductor carrying a two-dimensional matrix of photo-sensors (photo-diodes)
- Each photo-sensor is a small (usually square) electrically isolated capacitive region that can accumulate charge
- Photons incident on a photo-sensor are converted into electrons (via the photoelectric effect)
- The charge accumulated by each element is proportional to the incident light intensity and exposure time

To improve efficiency, microlenses are fabricated on chip to focus incoming light onto each sensor

The pattern of charge across the CCD corresponds to the pattern of incident light intensity (i.e. it is an image)



Colour from CCD devices

The photo-sensors in the CCD array are not selective to the wavelength of incoming light.

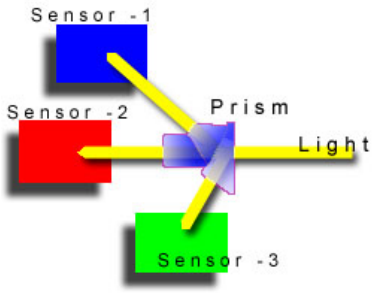
Colour sensitivity is achieved by adding a colour filter that allows through light from a small band of frequencies associated with a specific colour.

Colour from CCD devices

3 CCD solution

A prism splits light into 3 beams of different colour. 3 separate CCD devices record each colour

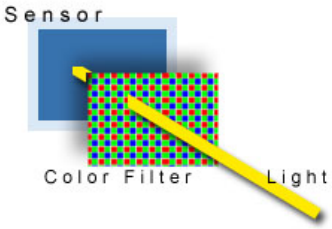
- Expensive, Bulky
- High image quality



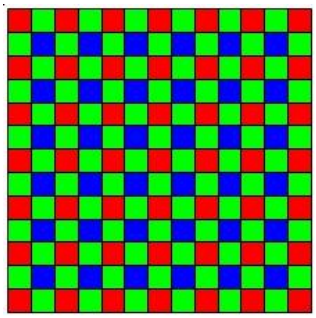
1 CCD solution

An array of coloured filters is placed over the pixels of single CCD to make different pixels selective to different colours

- Cheap, Compact
- Effectively generates coarser sampling



CCD colour masks



The Bayer mask (GRGB) is the most common colour filter used in digital cameras.

There are twice as many green filters as red and blue filters: improves sensitivity to green to which humans are most sensitive.

The raw output from a image sensor array is an array of pixels each with different colour

Demosaicing

The individual colour components provide a subsampled response from the original image, each subsampled response sensitive to a different colour.

Demoscaising is an process that computes the colour (RGB values) at every pixel based on the local red, green and blue values in the subsampled images.

Common demosaicing algorithms are nearest neighbour, bilinear interpolation, bicubic interpolation, spline interpolation, and Lanczos resampling.

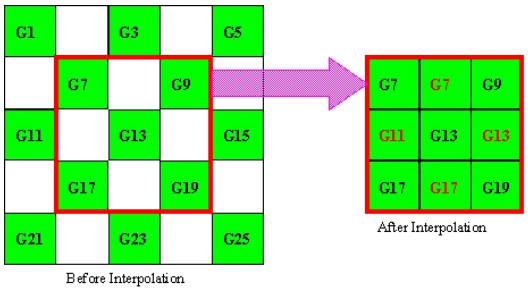
The same interpolation methods can be used when enlarging or rotating an image.

Demosaic이란 단어를 뭐져더리에선 [To convert a sample provided by the mosaic-like color filter array of a digital camera into a full-color image]라고 정의하는데, 해적하자면 "디지털 카메라의 모자이크 형태의 컬러필터배열(CFA)에 의해 만들어진 샘플을 풀컬러 이미지로 변환하는 것"이 됩니다.

Nearest Neighbour Interpolation

Copies an adjacent pixel value from the same colour channel.

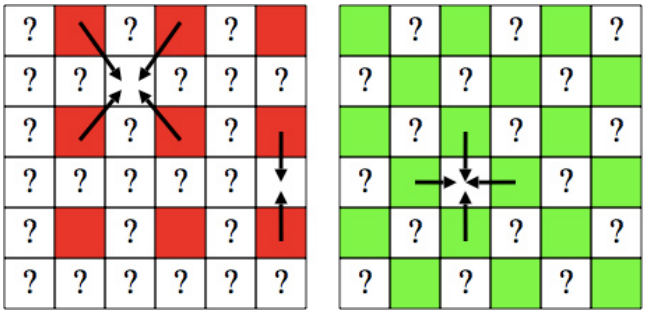
- V. Fast.
- V. Inaccurate.



Bilinear Interpolation

Takes average value of nearest two or four pixels from the same colour channel.

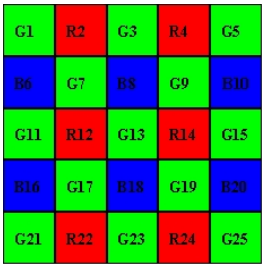
- Fast.
- Accurate in smooth regions, inaccurate at edges



Smooth Hue Transition Interpolation

Interpolation of green pixels: same as in bilinear interpolation.

Interpolation of red/blue pixels: bilinear interpolation of the ratio ("hue") between red/blue and green.



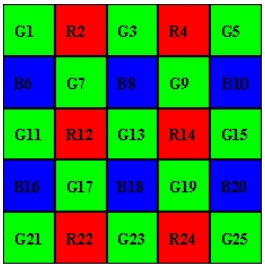
ratio of red green values

$$B12 = G12 * \frac{(B6/G6)+(B8/G8)+(B16/G16)+(B18/G18)}{4}$$

Edge-Directed Interpolation

The region around a pixel is analysed to determine if a preferred interpolation direction exists

Interpolation performed along axis where change in value is lowest (i.e. not across edges)



e.g. calculating green value at position 8

calculate horizontal and vertical gradients:

$\Delta H = |G7 - G9|$ and $\Delta V = |G3 - G13|$

perform interpolation:

If $\Delta H < \Delta V$,
$G8 = (G7 + G9)/2$;
Else if $\Delta H > \Delta V$,
$G8 = (G3 + G13)/2$;
Else
$G8 = (G3 + G7 + G9 + G13)/4$
End

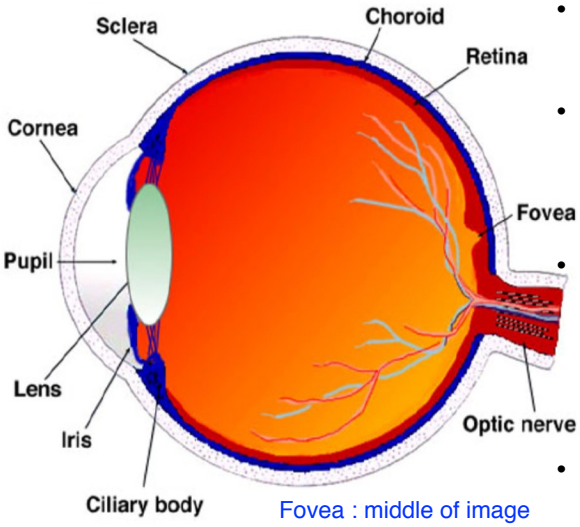
Interpolation of red/blue pixels : same as in smooth hue transition interpolation

Demosaicing

Raw output from Bayer mask image sensor

RGB output (at higher resolution) after demoscaising

The Eye



Fovea : middle of image

- Cornea performs the initial bulk of the refraction (at fixed focus)
- Lens performs further refraction and can be stretched to change its shape and hence change focal length
- Iris allows the eye to regulate the amount of light that can enter in order to both protect from over stimulation, and improve focus (as with the pinhole camera).
- Optic nerve: 1 million nerve fibers transmit information sensed by eye to the brain.

4 Oct 2018

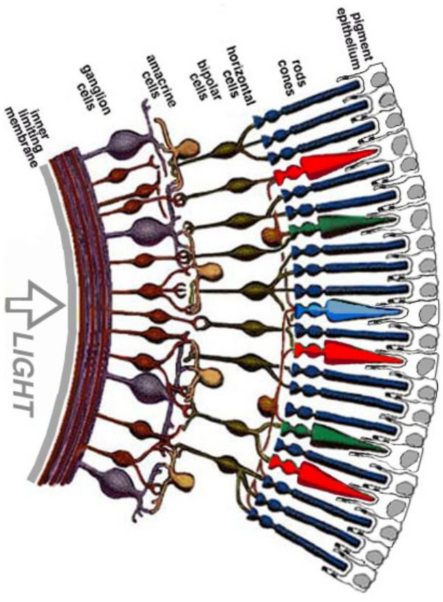
The Retina

The retina contains light sensitive **photoreceptors** (approx 130 million).

These photoreceptors are farthest from the light. But intervening cells are transparent

Photoreceptors **transduce** light to electrical signals (voltage changes).

Transduction = the transformation of one form of energy to another.



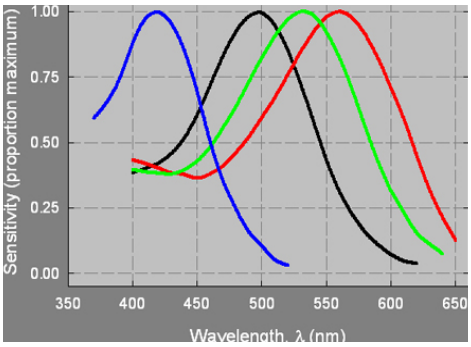
Photoreceptor types

Human eyes have 2 classes of photoreceptor:

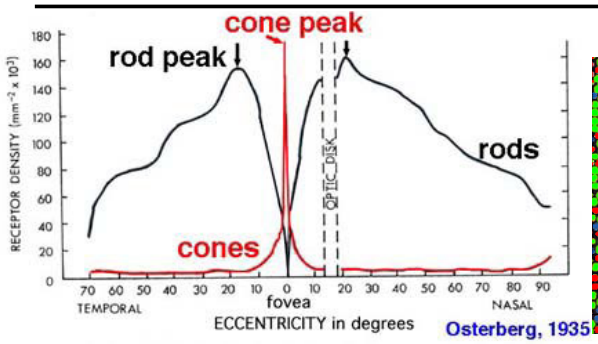
- **Rods**:
  - high sensitivity (can operate in dim light)
- **Cones**:
  - low sensitivity (require bright light)
  - 3 sub-types that are sensitive to different wavelengths

Hence cones provide colour information:

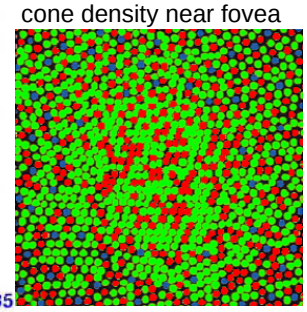
- blue: short-wavelength cones
  - peak sensitivity 440nm
- green: medium-wavelength cones
  - peak sensitivity 545nm
- red: long-wavelength cones
  - peak sensitivity 580nm



Distribution of photoreceptors



cone density near fovea



Photoreceptor types are not evenly distributed across the retina.

- **blind spot**: no photoreceptors (location where optic nerve leaves eye)
- **fovea**: no rods, high density of cones #(blue) << #(red) <= #(green)
- **periphery**: high concentration of rods, few cones

more rods overall

Foveal vs peripheral vision

The fovea is small region of high resolution containing mostly cones

Fovea:

- high resolution (acuity) – due to high density of photoreceptors
- colour – due to photoreceptors being cones
- low sensitivity – due to response characteristics of cones

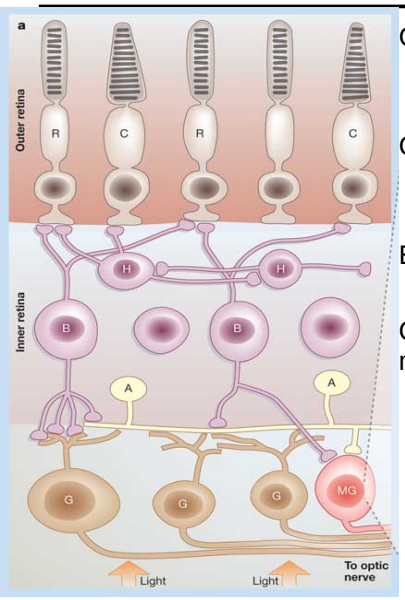
Periphery:

- low resolution (acuity)– due to low density of photoreceptors
- monochrome – due to photoreceptors being rods
- high sensitivity – due to response characteristics of rods

Far more of the brain is devoted to processing information from the fovea than from the periphery.

Retinal processing

How does the image transmit to brain?



Ganglion cells produce the output of the retina (the axons of all ganglion cells combine to form the optic nerve).

Ganglion cells produce action potentials: binary signals. This allows accurate transmission over long distance.

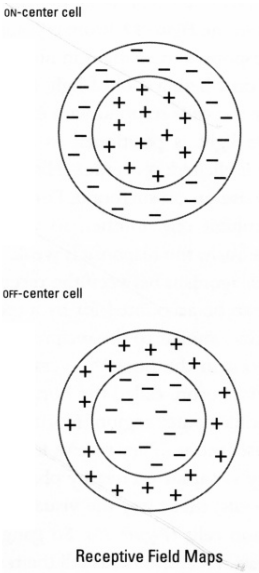
Each eye has approx. 1 million ganglion cells and 100million photoreceptors.

One ganglion cell collects input from multiple photoreceptors:

- Large convergence in periphery
- Smaller convergence in fovea

Another reason for the lower acuity in periphery compared to fovea

Ganglion cell responses



Ganglion cells have centre-surround **receptive fields**.

Receptive Field (RF) = area of visual space (i.e. area of retina) from which a neuron receives input. Or more generally, the properties of a visual stimulus that produces a response from a neuron.

Two types of ganglion cell:

On-centre, off-surround:

- active if central stimulus is brighter than background

Off-centre, on-surround:

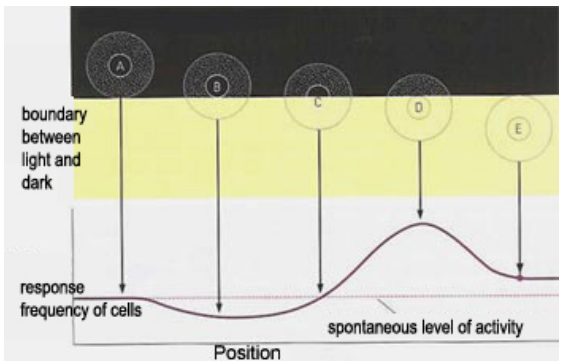
- active if central stimulus is darker than background

Behaviour is generated by ganglion cell being excited (inhibited) by photoreceptors in the centre of its receptive field and being inhibited (excited) by photoreceptors surrounding its receptive field.

Centre-surround RF function

A and E: on plane surface input to centre and surround cancels (output at resting – spontaneous - state, greater than zero)

B and D: at contrast discontinuity input to centre and surround unequal (output increased or decreased)



Functional consequences:

→ efficient coding

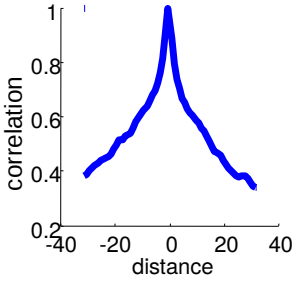
→ invariance to ambient light

→ accentuates edges

Centre-surround RF: efficient coding

Cells with RFs that fall on plane surfaces are only weakly active
Cells with RFs that fall on areas where contrast is changing are strongly active

Natural images have strong spatial correlation (i.e. little contrast change over short distances)



Hence centre-surround RF structure:

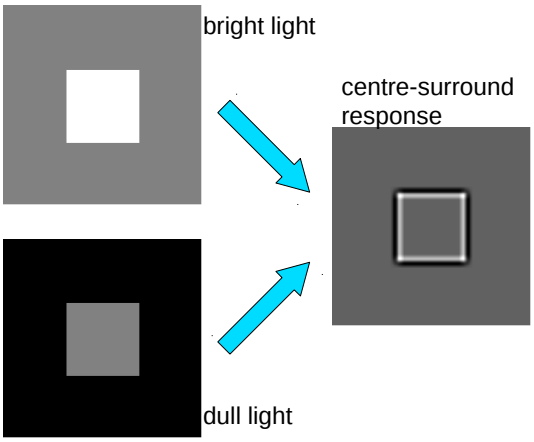
- minimises neural activity (efficient in terms of energy consumption)
- minimises bandwidth (efficient in terms of information coding) often referred to as "redundancy reduction" or "decorrelation" or "predictive coding"

Centre-surround RF: invariance to lighting

Ambient lighting conditions, i.e. the **Illuminance** (E), are generally irrelevant for most perceptual tasks.
e.g. an object should look the same in bright light and dull light.

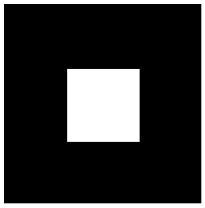
Centre-surround RFs measure the change in intensity (contrast) between adjacent locations.

This relative contrast remains constant independent of lighting conditions.

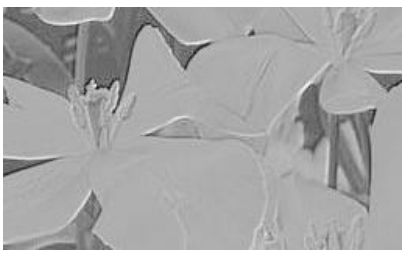


Centre-surround RF: edge enhancement

input image

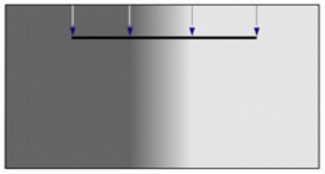


retinal response



centre-surround cells respond weakly where input is uniform, and strongly where input changes. Hence, strong response at edges.

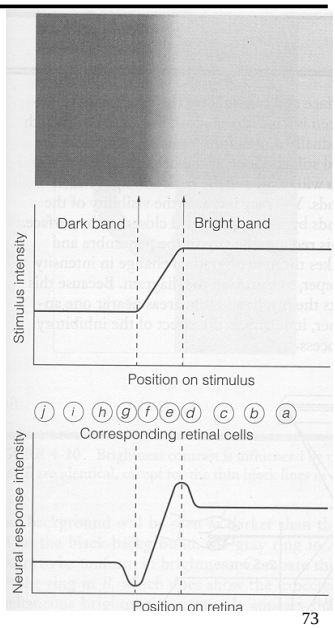
Centre-surround RF: edge enhancement



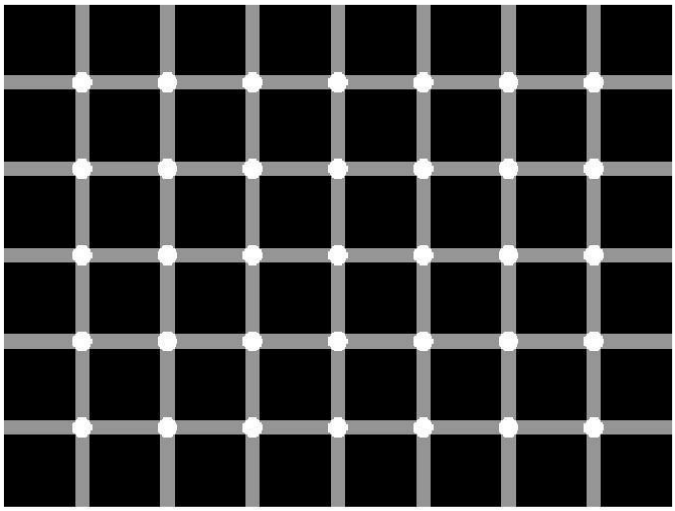
The light and dark lines (indicated by the central two arrows above) are illusory. They are called Mach bands.

Mach bands are a consequence of the edge enhancement generated by centre-surround RFs.

At a border of a light and dark surface, the light surface appears lighter, and the dark surface appears darker.



Hermann Grid



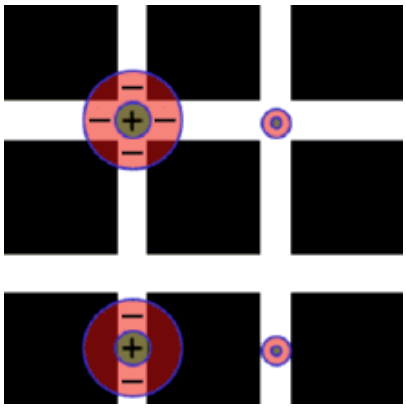
How many black dots are there?

Herman Grid Explanation

Ganglion cells in periphery have larger RFs:

Cell at intersection more inhibited than cell at edge

Therefore, response at intersection lower, and intersection appears darker

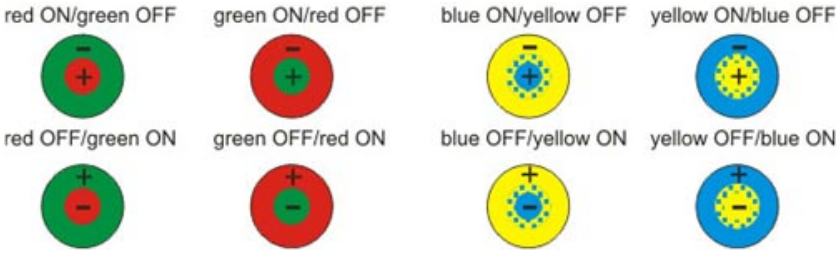


Ganglion cells in the fovea have smaller RFs:

Cell at intersection receives same input as cell at edge

Therefore, response the same, and intersections appear to have same contrast as edges

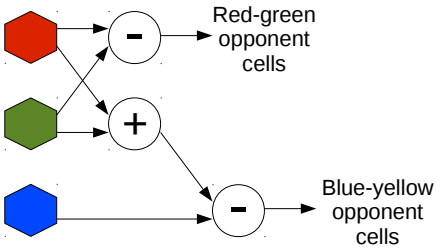
Centre-surround RFs: colour opponent cells



Ganglion cells combine inputs from both rods and cones in a centre-surround configuration.

Input from cones produces colour opponent cells.

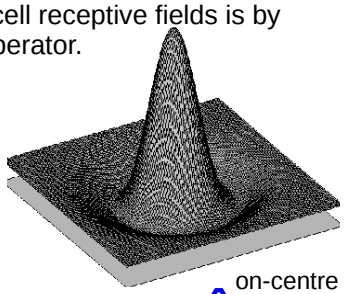
Only certain colour combinations occur in the human retina.

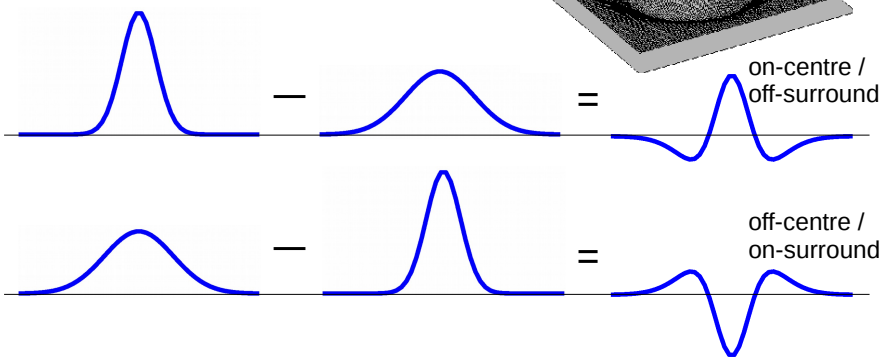


Modelling centre-surround RFs

The standard way of modelling ganglion cell receptive fields is by using a Difference of Gaussians (DoG) operator.

The DoG operator will reoccur in the next lecture.





Summary

Image formation is described by the pinhole (perspective) camera model

A point in 3D world projects to a point in 2D image dependent on extrinsic camera parameters, projection operator, intrinsic camera parameters

A lens is required to collect sufficient light to make an image that is both in focus and bright

Light reflected from an object is transduced into a electronic signal by a sensor (CCD array, retinal rods and cones)

The image is sampled at discrete locations (pixels), sampling is uniform in a camera, non-uniform in the retina (periphery vs fovea)

Following image formation the image needs to be analysed:

in biological vision system 1st stage of analysis is performed by centre-surround filters

in artificial vision systems .... (next lecture)

4 Oct 2018