Computer Vision 2024

Week – 1 lecture One

Miaomiao Liu

- What is Computer Vision?
- Why to study Computer Vision?
- What can Computer Vision do?
- What do you expect to learn from this course?

Synonym

Computer Vision

- Image Understanding
- Machine Vision
- Robotic Vision

What is Computer Vision?

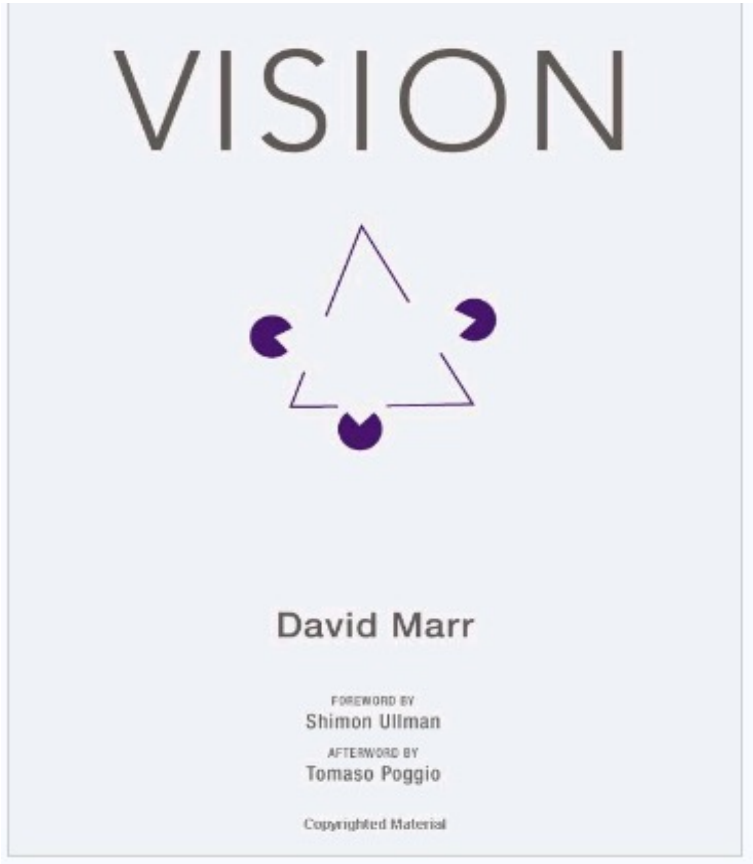
- (aka machine vision, robot vision, image understanding)
- and **applications of Computer Vision**

What is Computer Vision?

- Humans see things very easily.
- Can a computer do the same task?
  - Relieve human from tedious work
  - Perform more accurate measurement than what the human can do
- **How to teach a computer (or robot) to see (and understand what it sees)?**

What is Vision?

- Is [the art] ``*to know what is where, by looking*.''

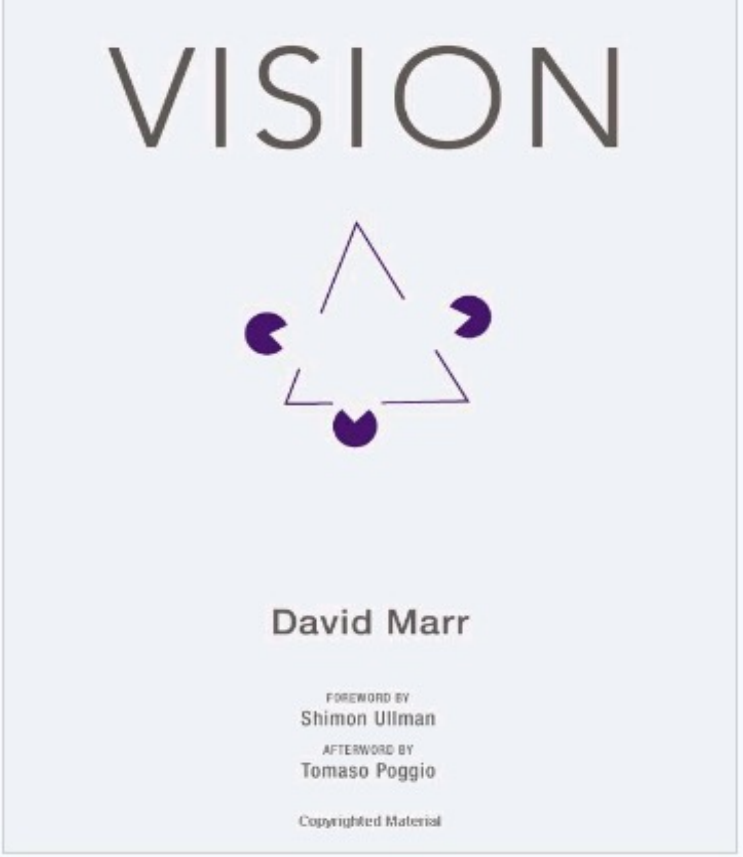


Published in 1982

David Marr (1945 --1980)

What is Vision?

- The process of discovering from images what is present in the world and where it is.

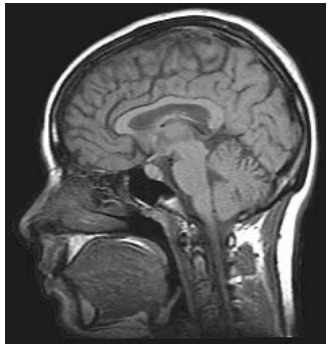


David Marr (1945 --1980)

What is Computer Vision?

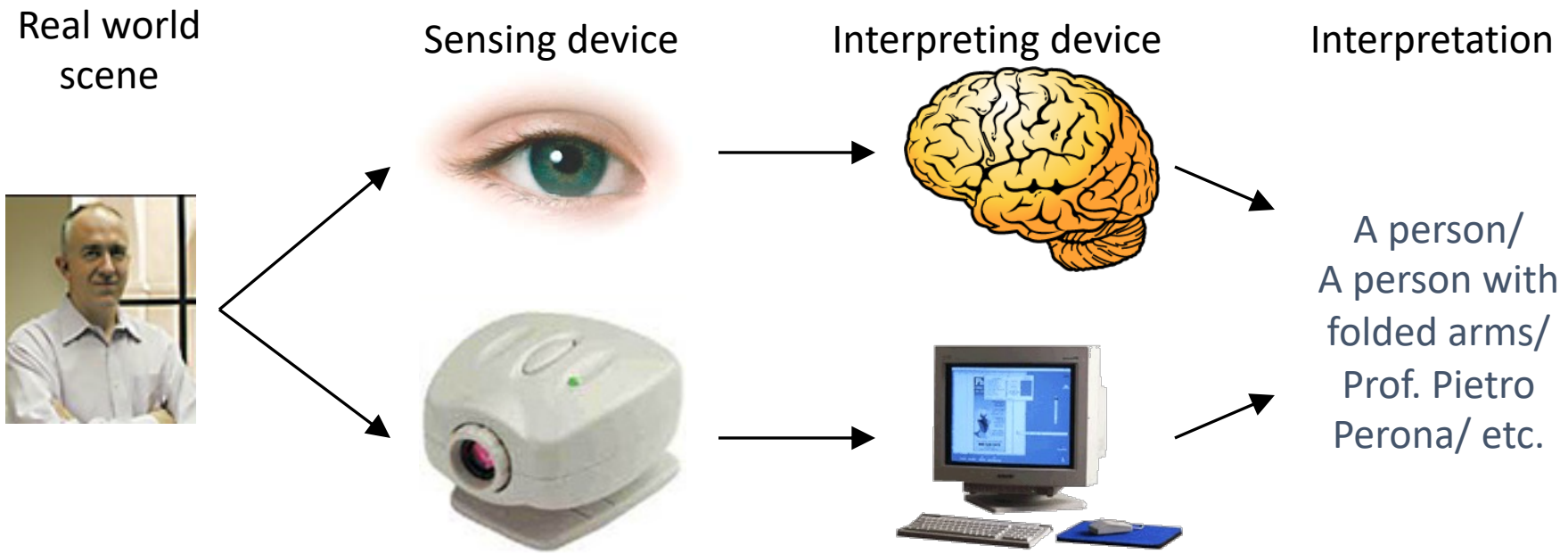
- Computer Vision is concerned with ***the theory*** of building artificial vision systems that obtain useful information from images.

- Image data taken by many forms, e.g. *video sequence, depth images, multi-dimensional data from a medical scanner etc.*



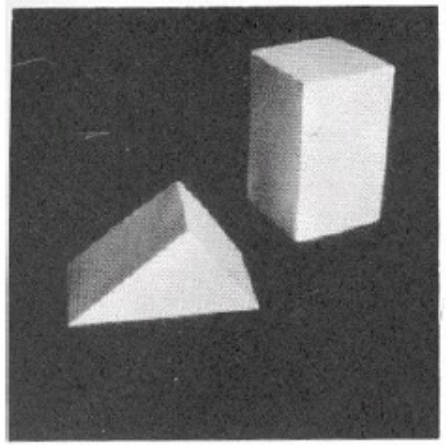
Computer Vision Problems

- Make a computer see and understand images.
- We know it is physically possible – we do it every day and effortlessly!

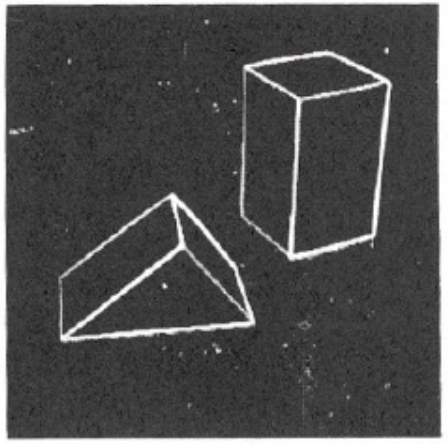


Brief history of computer vision research

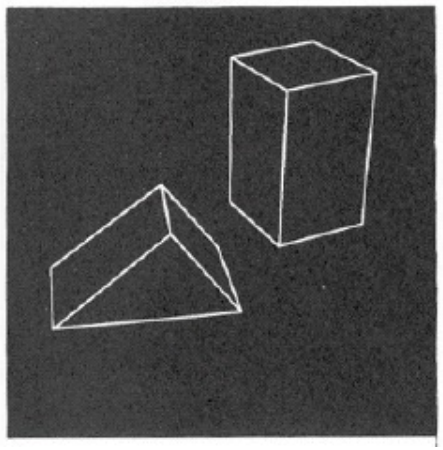
Origins of Computer Vision

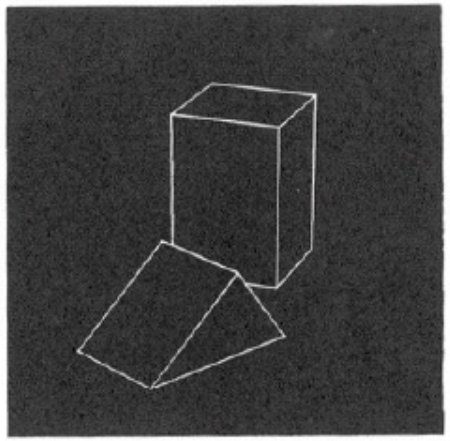


(a) Original picture.



(b) Differentiated picture.





L. G. Roberts, *Machine Perception of Three Dimensional Solids,* Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

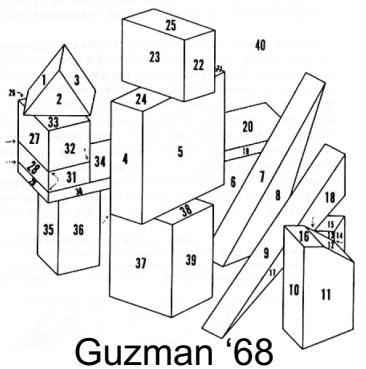
An (overly) optimistic start

In 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to "*spend 3 months in this summer linking a camera to a computer and getting the computer to describe what it saw*".

Now, more than fifty years, we know the problem is significantly more difficult than a 3-month student project.

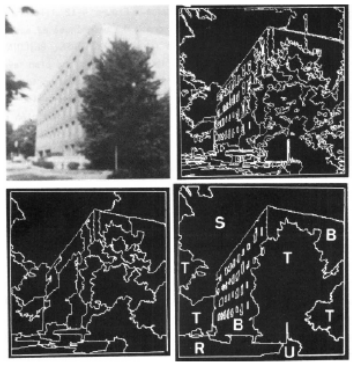
Brief history of computer vision

- 1963: Robert's thesis
- 1966: Minsky assigns computer vision as an undergrad summer project



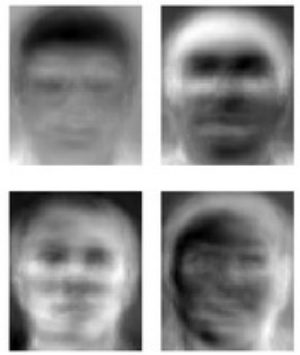
Guzman '68

- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images



Ohta Kanade '78

- 1980's: ANNs come and go; shift towards geometry and increased mathematical rigor, Marr's book published (after he died).
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts; vision & graphics; vision for HCI; internet vision, etc.
- 2012: Kinect, big-data, Google Car, … Deep Neural network, Deep learning ….


Turk and Pentland '91

Current State of the art of CV applications

- Apple Vision Pro
- https://www.youtube.com/watch?v=IY4x85zqoJM&list=PPSV&ab_channel=Apple

- Google Car
- https://www.youtube.com/watch?v=B8R148hFxPw

- Amazon Go
- https://www.youtube.com/watch?v=NrmMk1Myrxc

- HoloLens- Microsoft
- https://www.youtube.com/watch?v=aYdB2xBNFek

State of the art google self-driving car (GOOGLe)

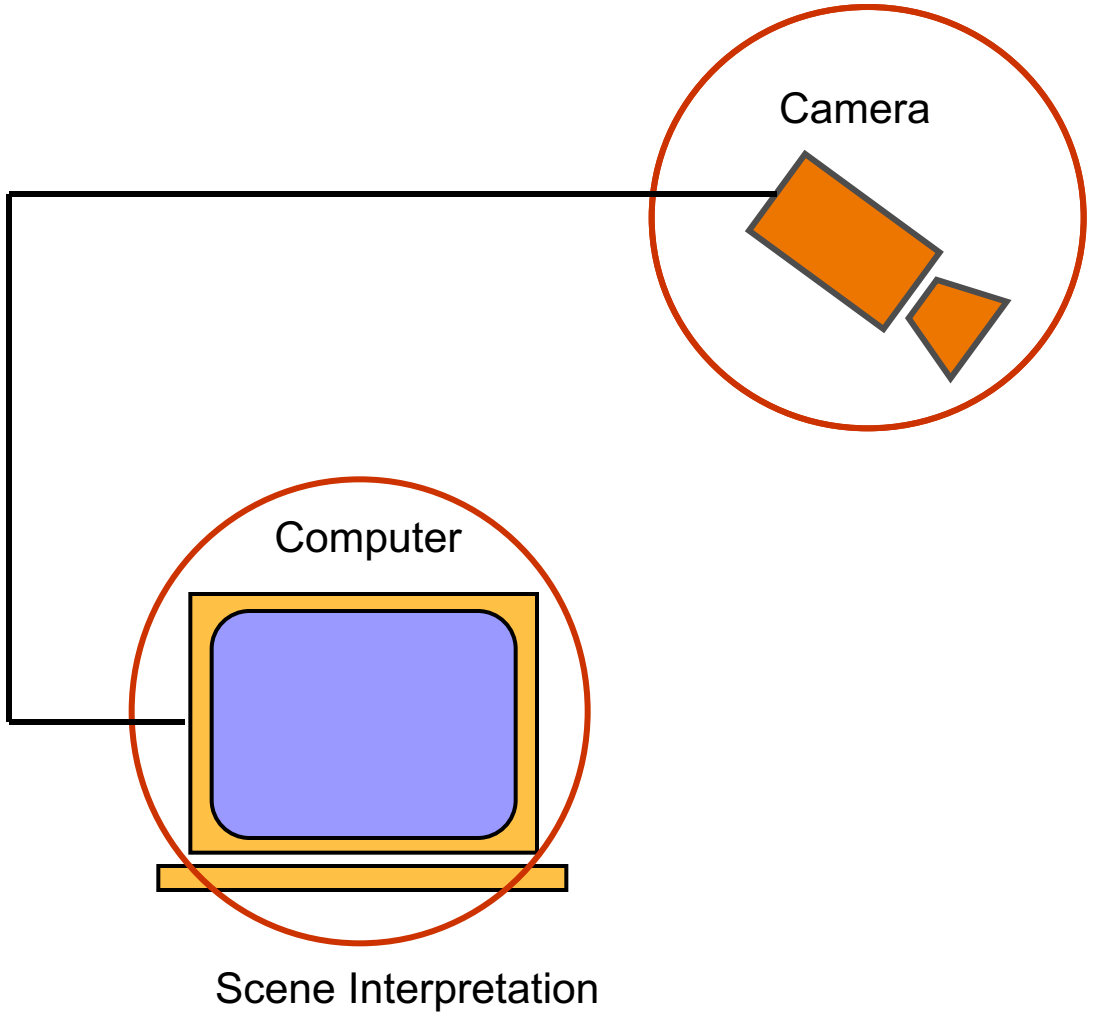
https://www.youtube.com/watch?v=B8R148hFxPw

What makes a **Computer Vision system** ?

- In a **Computer Vision System**, a camera (or several cameras) is linked to a computer.

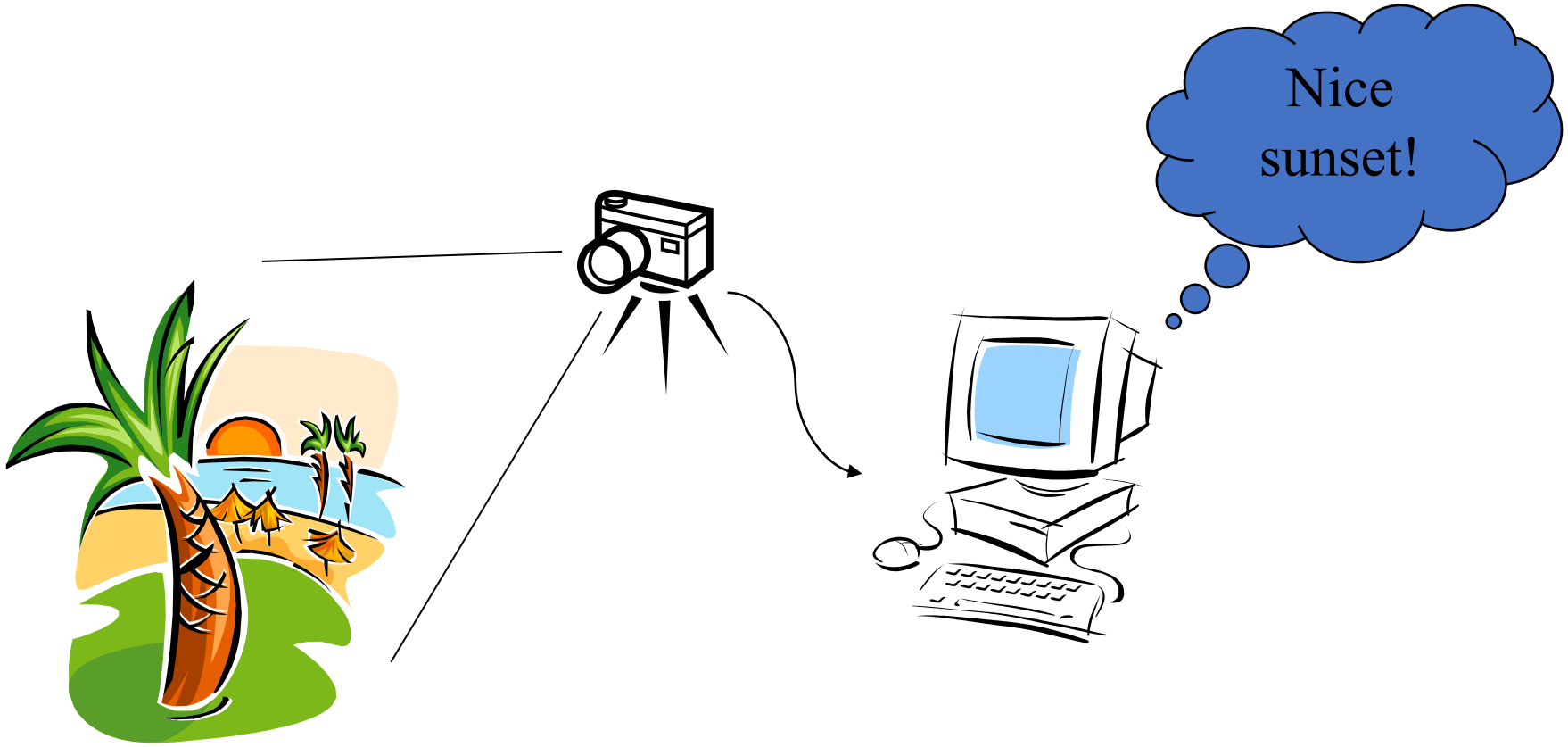
- The computer interprets images of a real world to obtain information useful for tasks such as navigation, manipulation and recognition.

Computer Vision System Example



Srinivasa Narasimhan's slide

– *“Making computers see”*



– *To ‘see’ is not only to take images, but also ‘to understand’ it.*

Understand the contents of images and videos.

What kind of scene?

Where are the buildings?

How far are the buildings from the camera?

…

What we could see

What computer sees

```
01 00 05 00 03 00 02 00 00 03 01 01 01 01 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 02 00 01
03 30 3A 38 39 2D 1D 15 10 0E 0C 0A 0A 0A 09 06 08 07 06 06 05 05 07 07 04 05 04 04 06 02 01 02 02 02 02 07 01 02 02 03
03 22 1B 16 14 0A 08 0B 0A 0D 0B 0B 0C 06 07 05 05 06 06 06 03 07 04 06 05 09 05 04 05 01 04 04 02 03 03 04 02 04 03 02
00 0F 0B 04 10 07 09 07 08 09 09 08 05 08 08 05 09 03 08 05 02 08 08 06 06 04 02 05 03 02 05 05 00 02 02 04 04 00 00 03
00 07 09 0E 0C 07 08 0A 0A 0B 0F 0A 0C 07 06 0B 07 0B 05 0B 08 09 07 03 08 04 04 02 00 04 02 04 00 04 03 08 00 06 09 04
00 0E 0C 09 09 08 08 07 08 09 09 0A 05 08 07 07 07 09 08 0A 08 09 06 0A 03 09 07 06 06 03 05 03 01 06 02 03 07 01 04 04
02 0C 0B 0A 05 08 09 0A 0C 0A 0A 08 0A 0A 06 08 06 06 04 06 02 06 07 04 04 04 06 09 05 05 08 06 04 05 04 06 01 0A 03 02
02 0B 14 0F 0F 0D 0A 0E 0A 0C 0C 0E 0A 0C 0B 09 0A 09 0A 0A 09 0B 0B 05 0C 0C 0A 04 07 06 03 05 07 04 05 03 02 01 06 03
02 10 12 0B 10 0A 0D 0D 0B 0D 0C 0B 0B 0C 0D 0B 0B 0A 0A 0A 0B 0C 17 15 1C 15 0D 08 09 08 05 05 05 04 02 05 04 04 00 04
01 15 0E 10 12 0C 0D 0C 0C 0A 0B 0B 09 0C 0F 09 09 0D 07 0B 08 15 60 5D 61 59 33 0D 0A 07 08 08 05 03 06 07 01 03 05 02
02 12 10 0F 0E 10 10 0B 0C 0F 0F 0E 0C 10 0D 15 10 09 12 11 12 50 68 66 89 71 5E 3F 08 09 0A 09 0A 03 03 02 05 05 04 02
01 11 12 0C 11 13 10 10 0B 10 0F 0C 11 11 13 0D 0F 0D 0D 0B 25 7A 7F 79 6D 80 6E 54 0C 0D 09 0A 06 04 02 05 00 05 04 03
01 10 0F 0D 12 0E 10 0E 0F 13 13 11 13 17 11 0F 14 11 11 14 39 84 88 7E 8C 73 7A 5C 1E 05 0A 0F 0E 0C 05 02 04 03 06 05
02 0F 15 0D 18 11 0D 11 14 10 12 12 14 19 13 17 13 16 16 20 73 68 87 89 93 8B 83 69 43 07 0A 12 0A 0B 06 06 03 04 05 03
02 13 14 14 16 11 13 13 17 12 17 17 28 1E 1A 17 19 14 12 4F 7D 74 85 91 93 8C 7F 6F 5F 0B 09 12 0D 0C 02 04 07 04 05 04
00 0F 16 0F 13 12 10 1D 12 21 15 1E 21 1F 1C 1D 2D 1A 2D 7C 7A 95 6B 30 48 62 87 71 5C 0A 08 11 0C 09 04 04 02 06 04 03
00 10 1C 10 11 1A 0D 1A 1A 25 28 33 30 26 2B 3E 29 35 6C 83 5E 7B 94 8A 5A 3D 42 76 5C 13 08 13 0F 0C 04 04 01 05 05 03
01 12 17 1A 19 18 15 20 29 20 3F 1F 37 29 39 49 24 33 8F 93 B4 AE 79 42 39 73 7D 89 46 12 06 12 12 0F 08 03 03 03 04 03
01 13 20 0F 14 26 1B 18 20 2F 3D 3E 42 3B 45 2E 48 70 96 9F 96 6B 24 0F 22 4B C3 A4 3F 4F 0C 18 16 0F 05 05 08 05 05 04
00 19 1C 13 13 21 1D 12 18 47 3D 47 45 3A 27 3B 33 A8 A6 91 81 4B A1 75 4B AC A1 B5 79 0C 0B 13 0F 0B 02 03 06 07 07 04
00 1B 1D 1C 1C 1C 1B 1B 1E 55 49 49 36 28 2A 24 9F AD AC AA B1 9C 8D 5F 3E 98 B7 B7 A3 31 11 14 0A 0D 04 08 07 07 07 06
02 21 18 15 16 1D 15 18 1E 36 5B 29 2C 19 29 4F AF BC AF AB 9E A1 97 82 70 9F AE AD A5 92 16 10 07 0E 0A 0C 08 05 0B 05
01 17 1B 1A 1A 2B 1B 2A 32 34 46 2C 1B 26 4C 40 BA BB B5 AE 95 94 84 7A 8A 9A B9 BB AD 9C 8A 15 09 09 05 0B 0D 0F 0B 07
00 1A 18 1C 1E 27 21 1D 3F 4E 32 25 1B 1B 93 46 AF AB B1 AC A4 93 89 91 86 90 AA 9F 91 97 AD 7F 0C 0B 0E 0B 0C 0C 09 05
00 15 1A 21 1E 2E 1B 23 47 4E 23 21 19 49 99 5B AA AC B7 AF A6 9A 93 8F 85 7F A0 A4 C2 9F 99 4E 09 08 0A 0D 0C 0A 0C 07
00 13 18 21 26 31 28 25 34 4C 1F 2B 1C 8B 9B 42 9B A7 A1 B4 B0 AA A0 9D 92 72 8E 97 71 A7 32 04 0A 0A 0D 0D 09 0D 0C 07
00 1A 1C 21 28 3A 30 26 40 4C 26 18 2C 90 A1 39 A0 97 B8 AA B2 A5 A6 A3 98 76 92 96 98 6D 08 0D 07 08 0C 0B 0E 0D 0D 0A
04 1E 29 1F 27 32 26 2E 41 4A 2C 34 46 8A A5 89 9E A3 B0 B7 AF AB AB 99 97 90 A4 94 85 7C 08 07 07 08 09 09 08 0C 0D 0B
01 1F 29 27 27 2A 2C 36 4D 50 34 42 45 95 9B AA 7E AD B3 AA B2 A8 B2 92 98 8E 9E 8E 44 34 18 05 06 0A 0D 0D 0D 0F 0C 08
00 21 2E 23 29 2C 2A 34 44 5A 39 4F 29 90 9B A5 86 AA B2 B3 AE A0 A3 9C 94 79 43 2B 25 2D 07 0E 05 06 0C 0A 0F 0D 09 0C
00 21 27 20 28 29 2F 2A 44 57 42 31 28 8C 93 A3 AC 60 BA BD B4 AE A8 A2 62 91 5F 52 4F 3F 09 0D 0D 09 0E 0E 0B 12 0B 0B
03 30 2E 2C 29 2A 3B 30 4E 3C 40 40 49 5E AE 9F A4 B1 4E AA AA A0 A4 9C 94 A2 AB A8 93 52 0E 0E 09 0B 0D 10 0C 0C 10 09
00 30 32 2E 36 39 36 24 2D 5A 46 46 68 30 8B 8C A3 AC A5 3E A1 AF A8 82 A4 AC A2 96 71 73 08 10 0B 0B 0B 0E 0F 10 11 0A
00 54 34 1E 3C 3F 3E 29 27 56 38 4C 5C 44 26 94 9A A2 A2 A6 8E 4E 70 99 AC A6 A2 89 7E 5B 11 0E 10 10 17 12 0D 0C 0D 0C
00 4B 30 23 36 44 48 3C 2E 2D 34 35 29 58 5B 0D 36 50 34 52 9C A8 B5 AA B3 AE A0 9C 8C 62 0A 12 14 0D 16 14 11 10 0E 0D
01 38 2C 24 2E 51 59 4B 30 27 39 2B 2B 24 29 69 37 25 29 82 97 A1 AB AC B2 A6 A6 A0 89 69 0F 10 1C 18 14 10 10 0F 0C 0F
03 21 2A 27 22 5C 44 31 3F 33 1F 37 24 23 36 27 24 2B 4D 50 85 90 96 86 A3 A5 99 8D 7A 4E 0E 1B 15 20 0F 0F 16 12 13 0B
01 1D 1F 2B 20 21 48 2F 40 2F 2D 2A 25 2B 2C 20 25 25 26 3E 55 5E 62 6D 6D 6E 68 5E 43 0D 10 21 18 32 1A 13 10 13 15 10
04 27 2F 2A 28 21 3B 45 2E 3A 40 33 2D 2F 1F 1E 1B 20 37 3C 3F 3C 34 30 24 17 0D 0B 0E 11 1E 23 1B 25 14 0D 10 0F 12 0F
04 22 27 37 33 1A 1B 35 4A 1D 20 2C 2F 1F 1F 3B 34 1A 2A 38 44 1E 0C 0C 06 0C 10 12 1B 21 21 34 32 20 0B 0E 10 0D 0D 0F
02 32 22 33 29 20 22 19 30 35 1D 1E 16 19 18 1C 16 18 23 39 10 13 0E 0E 1A 15 15 13 1A 18 2C 2E 19 0F 0D 10 0E 0E 14 0D
01 33 36 23 31 29 20 19 1B 1E 17 1C 1F 1F 1F 1C 31 23 1C 2F 13 11 16 10 12 16 13 19 1B 17 19 1D 13 14 10 10 12 11 12 0D
01 28 31 34 24 30 23 19 18 28 2A 1D 1F 1D 1B 1E 1B 26 31 39 16 14 13 14 13 15 1B 22 1A 1E 1B 15 13 16 0C 0D 11 0E 12 0D
00 29 20 1C 2E 25 28 28 22 1E 20 1F 1F 1D 1B 1C 29 22 43 37 17 10 15 15 12 10 14 15 1B 1E 15 1A 11 10 14 13 14 17 12 11
01 25 28 2A 23 23 29 26 1E 1D 34 38 1B 1B 22 26 18 1A 4C 33 1C 11 14 14 14 10 10 18 17 1E 29 20 1A 15 12 17 0E 14 12 12
```


01 00 05 00 03 00 02 00 00 03 01 01 01 01 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 02 00 01
03 30 3A 38 39 2D 1D 15 10 0E 0C 0A 0A 0A 09 06 08 07 06 06 05 05 07 07 04 05 04 04 06 02 01 02 02 02 02 07
01 02 02 03
03 22 1B 16 14 0A 08 0B 0A 0D 0B 0B 0C 06 07 05 05 06 06 06 03 07 04 06 05 09 05 04 05 01 04 04 02 03 03 04
02 04 03 02
00 0F 0B 04 10 07 09 07 08 09 09 08 05 08 08 05 09 03 08 05 02 08 08 06 06 04 02 05 03 02 05 05 00 02 02
0404 00 00 03
00 07 09 0E 0C 07 08 0A 0A 0B 0F 0A 0C 07 06 0B 07 0B 05 0B 08 09 07 03 08 04 04 02 00 04 02 04 00 04 03 08
00 06 09 04
00 0E 0C 09 09 08 08 07 08 09 09 0A 05 08 07 07 07 09 08 0A 08 09 06 0A 03 09 07 06 06 03 05 03 01 06 02 03
07 01 04 04
02 0C 0B 0A 05 08 09 0A 0C 0A 0A 08 0A 0A 06 08 06 06 04 06 02 06 07 04 04 04 06 09 05 05 08 06 04 05 04 06
01 0A 03 02
02 0B 14 0F 0F 0D 0A 0E 0A 0C 0C 0E 0A 0C 0B 09 0A 09 0A 0A 09 0B 0B 05 0C 0C 0A 04 07 06 03 05 07 04 05 03
02 01 06 03
02 10 12 0B 10 0A 0D 0D 0B 0D 0C 0B 0B 0C 0D 0B 0B 0A 0A 0A 0B 0C 17 15 1C 15 0D 08 09 08 05 05 05 04 02 05
04 04 00 04
01 15 0E 10 12 0C 0D 0C 0C 0A 0B 0B 09 0C 0F 09 09 0D 07 0B 08 15 60 5D 61 59 33 0D 0A 07 08 08 05 03 06 07
01 03 05 02
02 12 10 0F 0E 10 10 0B 0C 0F 0E 0E 0C 10 0D 15 10 09 12 11 12 50 68 66 89 71 5E 3F 08 06 0A 08 0A 03 03 02
05 05 04 02
01 11 12 0C 11 13 10 10 0B 10 0F 0C 11 11 13 0D 0F 0D 0D 0B 25 7A 7F 78 6D 80 6E 54 0C 0B 09 0A 06 04 02 05
00 05 04 03
01 10 0F 0D 12 0E 10 0E 0F 13 13 11 13 17 11 0F 14 11 11 14 39 84 88 7E 8C 73 7A 5C 1E 05 0A 0F 0E 0C 05 02
04 03 06 05
02 0F 15 0D 18 11 0D 11 14 10 12 12 14 19 13 17 13 16 16 20 73 68 87 89 93 8B 83 69 43 07 0A 12 0A 0B 06 06
03 04 05 03
02 13 14 14 16 11 13 13 17 12 17 17 28 1E 1A 17 19 14 12 4F 7D 74 85 91 93 8C 7F 6F 5F 0B 09 12 0D 0C 02 04
07 04 05 04
00 0F 16 0F 13 12 10 1D 12 21 15 1E 21 1F 1C 1D 2D 1A 2D 7C 7A 95 6B 30 48 62 87 71 5C 0A 08 11 0C 09 04 04
02 06 04 03
00 10 1C 10 11 1A 0D 1A 1A 25 28 33 30 26 2B 3E 29 35 6C 83 5E 7B 94 8A 5A 3D 42 76 5C 13 08 13 0F 0C 04 04
01 05 05 03
01 12 17 1A 19 18 15 20 29 20 3F 1F 37 29 39 49 24 33 8F 93 B4 AE 79 42 39 73 7D 89 46 12 06 12 12 0F 08 03
03 03 04 03
01 13 20 0F 14 26 1B 18 20 2F 3D 3E 42 3B 45 2E 48 70 96 9F 96 6B 24 0F 22 4B C3 A4 3F 4F 0C 18 16 0F 05 05
08 05 05 04

…and the output is

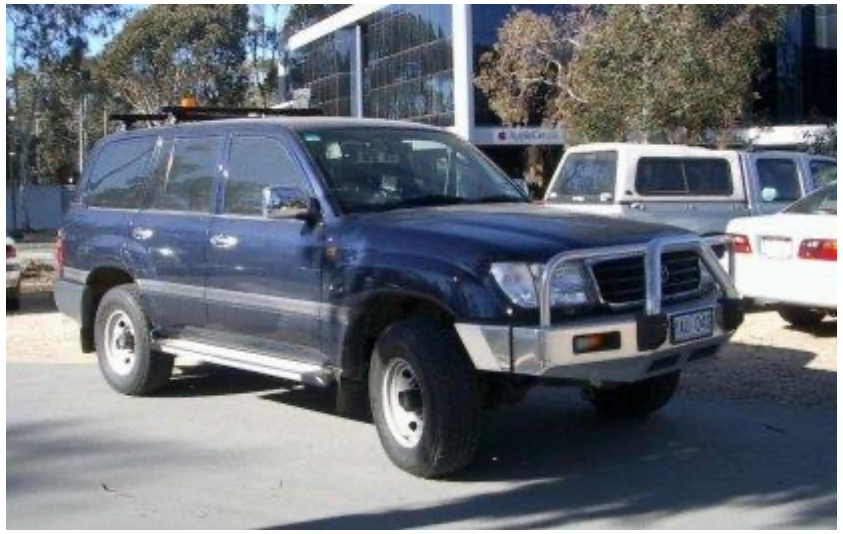
- *A harbor…*
- *… with many dozens of boats;*
- *… water is calm and glassy;*
- *… vertical masts;*
- *… mountains in background,*
- *… blue sky with a touch of clouds…*

Interpret images.

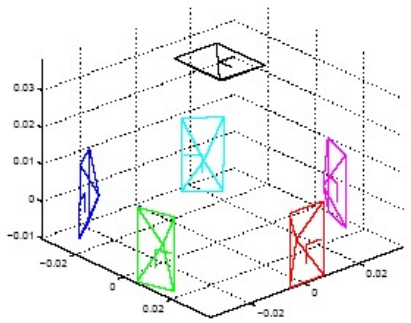
Image → Symbols, Semantics, Meanings,..

- Previous Computer Vision Projects @ ANU

Autonomous vehicles / self-driving cars

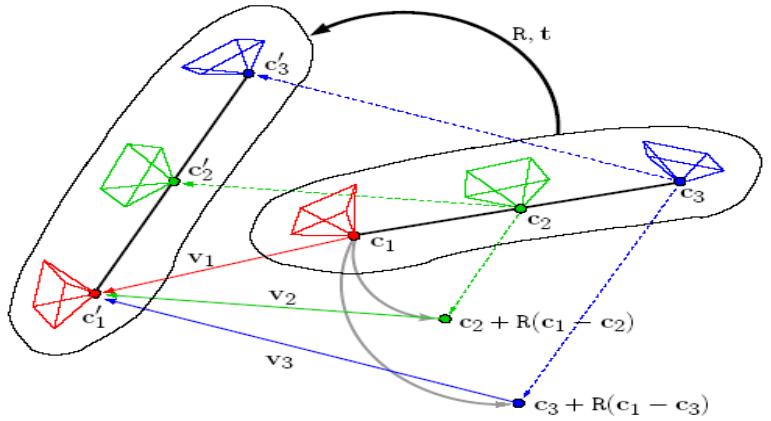


Multiple Camera City Modelling

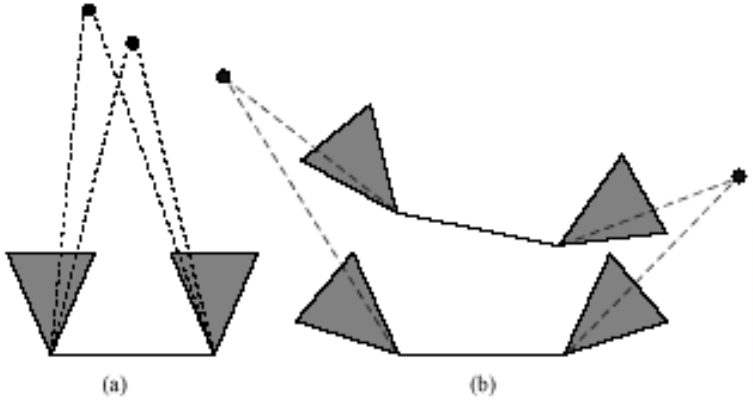


(a)

(b)



Multi-camera system



(a) Overlapping vs. (b) non-overlapping Multi-camera systems

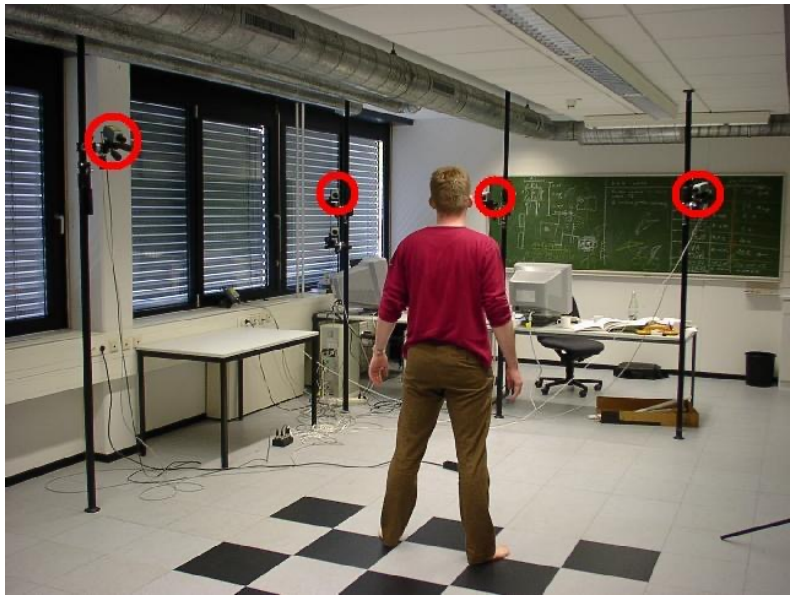
Images from one camera cluster

Kim and Li, Hartley, CVPR 08

Human Body Shape and Motion 3D Capture

- Marker-free
- No need to wear a special suit or gloves
- No dedicated hardware
- Standard video cameras
- Cost-effective

KINECT
for XBOX 360

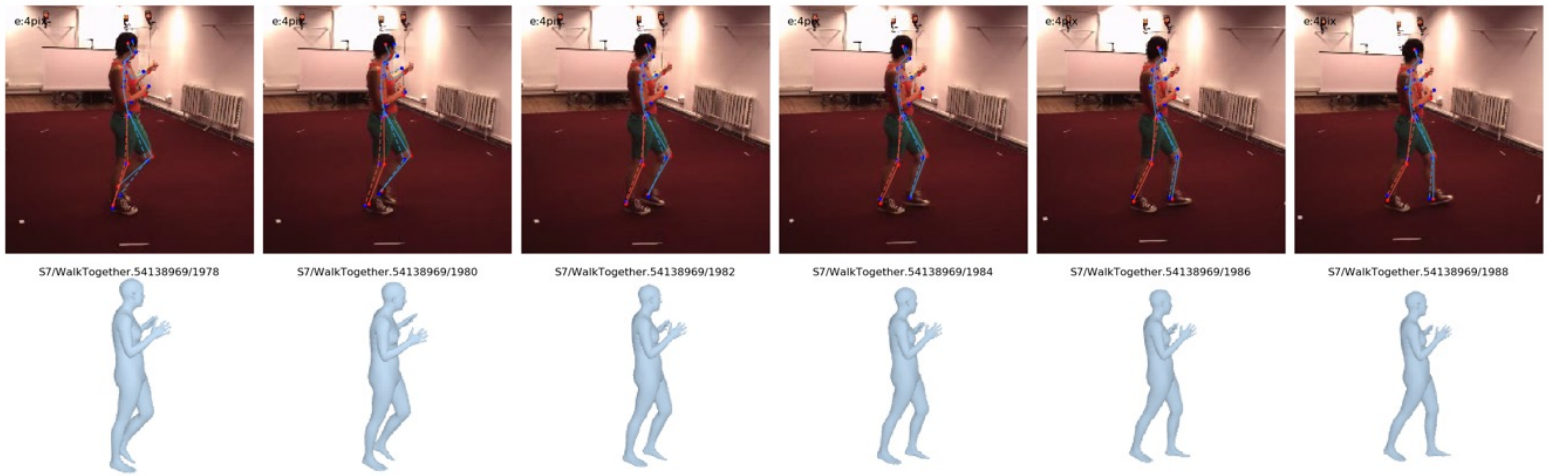


Motion Capture Environment (Image courtesy of Christian Theobalt)

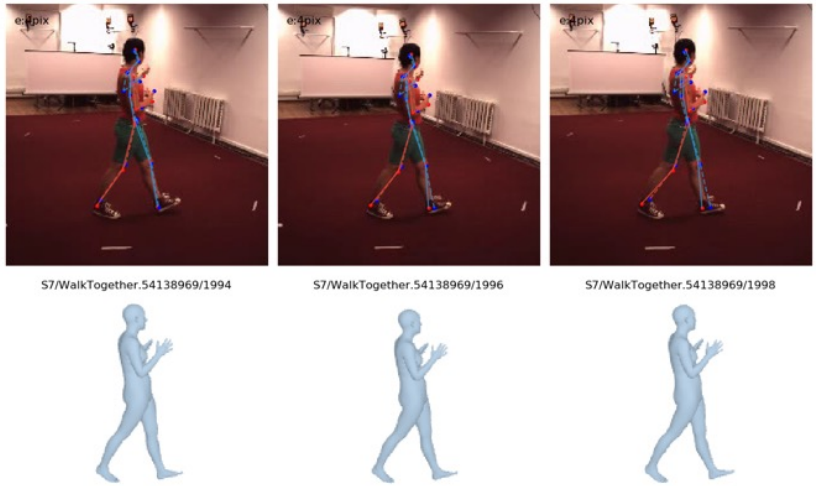
Human Body Shape and Motion Prediction

- Predict Human Motion Dynamics from a Video sequence

Historical Data:

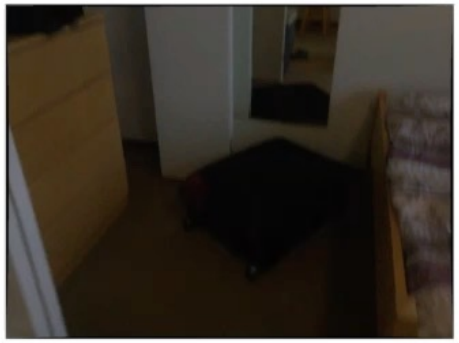


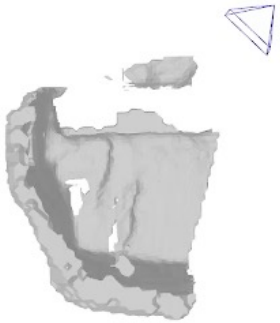
Predict Future Human Motion:



Mao, Liu, Salzmann, Li, ICCV19

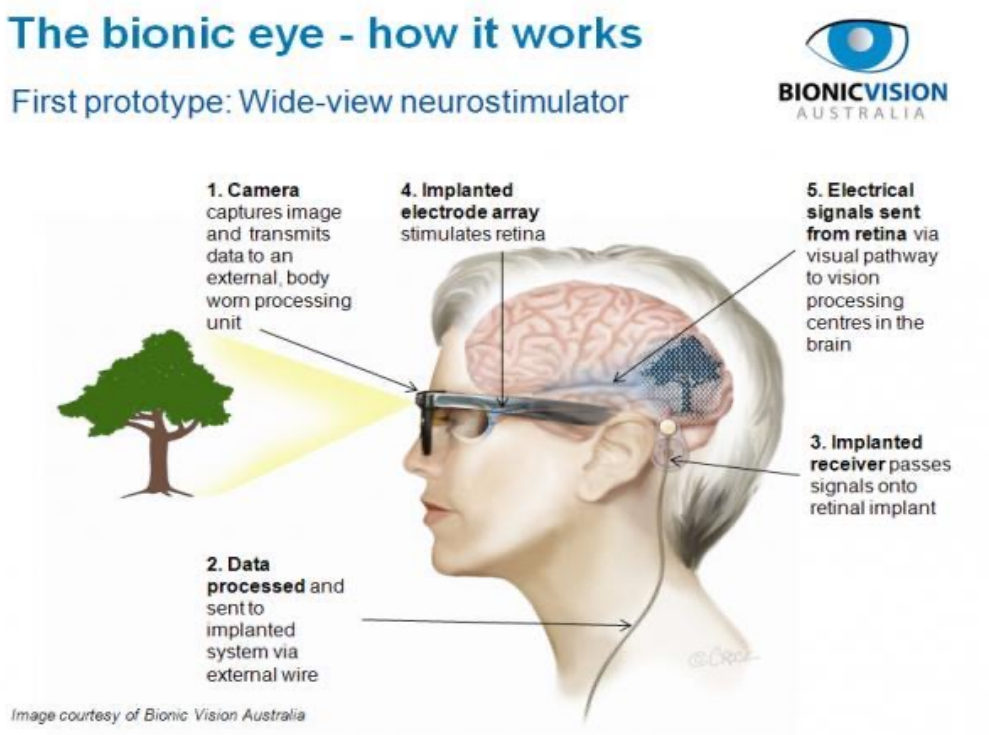
3D Reconstruction





Australian Bionic Eye

- 2020 Summit Project
- Budget $50M for first 4 years
- 5 BVA Members and 2 contributing universities
- Officially started July 2010



THE UNIVERSITY OF MELBOURNE
UNSW
The Bionic Ear Institute
Centre for Eye Research Australia
NICTA

Patient trial in Canberra 2014

- Ms Ashworth had her first 'unplugged' trial in *Canberra*, 2014.
- https://www.abc.net.au/news/2014-04-30/bionic-eye-patients-start-first-navigation-tests/5422174

Australian Centre for Robotic Vision

- $25M government funding for 2014 - 2020
- 4 Centre Nodes (QUT, ANU, Adelaide, Monash)
- 13 Chief Scientists national-wide


Oxford University, Oxford, UK
Imperial College, London, UK
ETH, Zurich, Switzerland
INRIA, Rennes, France
Georgia Tech, Atlanta, Georgia, USA
QUT, Brisbane
NICTA, Sydney, Brisbane, Canberra, Melbourne
UoA, Adelaide
Monash, Melbourne
ANU, Canberra

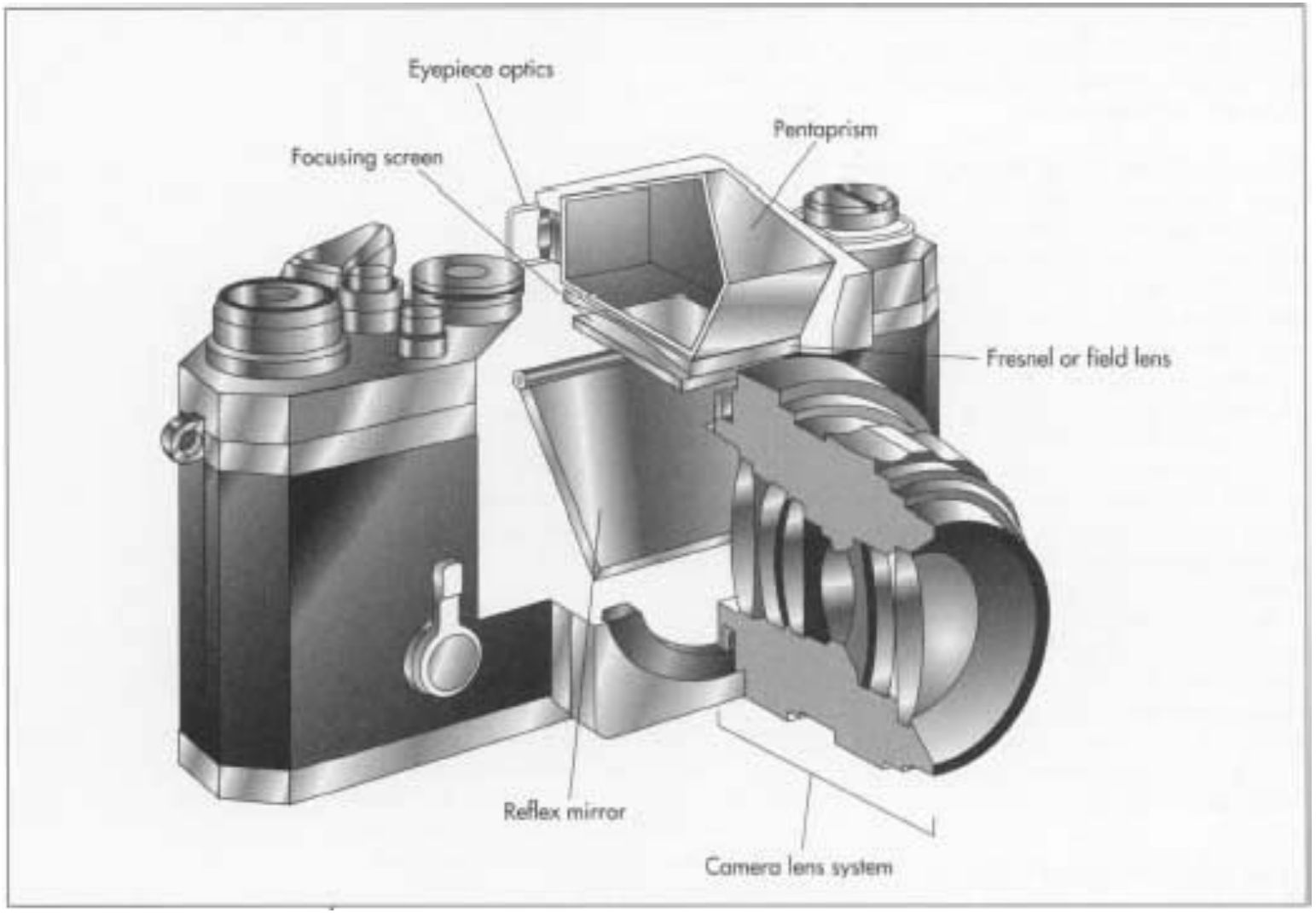
Outline

- What is a camera (sensor) ?
- Geometric image formation (details in 3D Vision Lectures)
- Photometric image formation
- Image representation
- Point Operation

Outline

- What is a camera (sensor) ?
- *Geometric image formation (details in 3D Vision Lectures)*
- Photometric image formation
- Image representation
- Point Operation

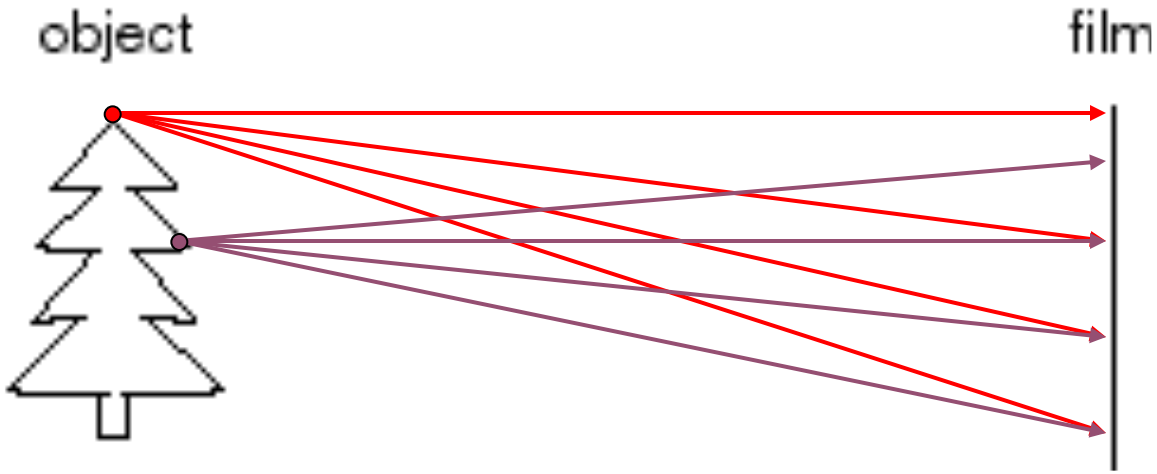
What is a camera ?



Slides from: F. Durand, S. Seitz, S. Lazebnik, S. Palmer

Let us design a camera

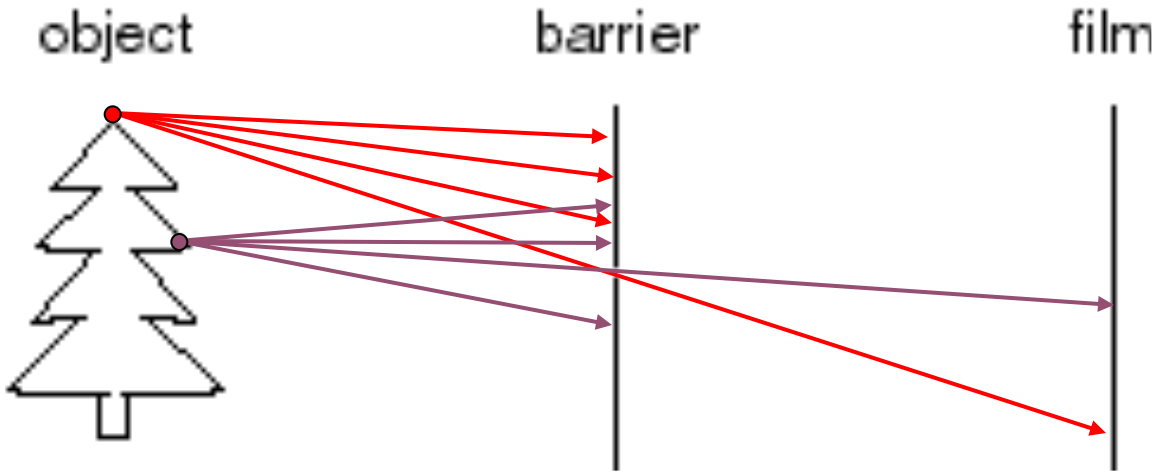
Let's design a camera



- Idea 1: put a piece of film in front of an object
- Will we get an image?

Slide by Steve S

Pinhole camera

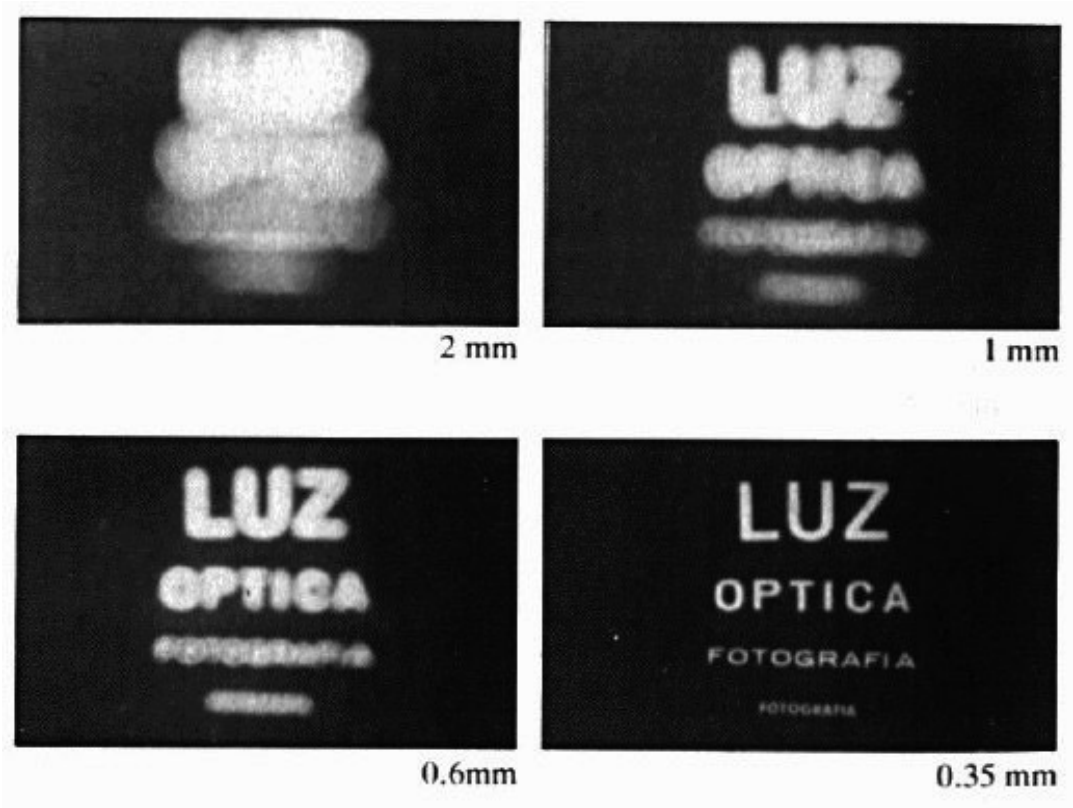


- Add a barrier to block off most of the rays
  - This reduces blurry effect
  - The opening is known as the ***aperture***

A pinhole camera is a simple camera without a lens but with a tiny aperture (the so-called pinhole)—effectively a light-proof box with a small hole in one side. Light from a scene passes through the aperture and projects an inverted image on the opposite side of the box, which is known as the camera obscura effect. The size of the images depends on the distance between the object and the pinhole.

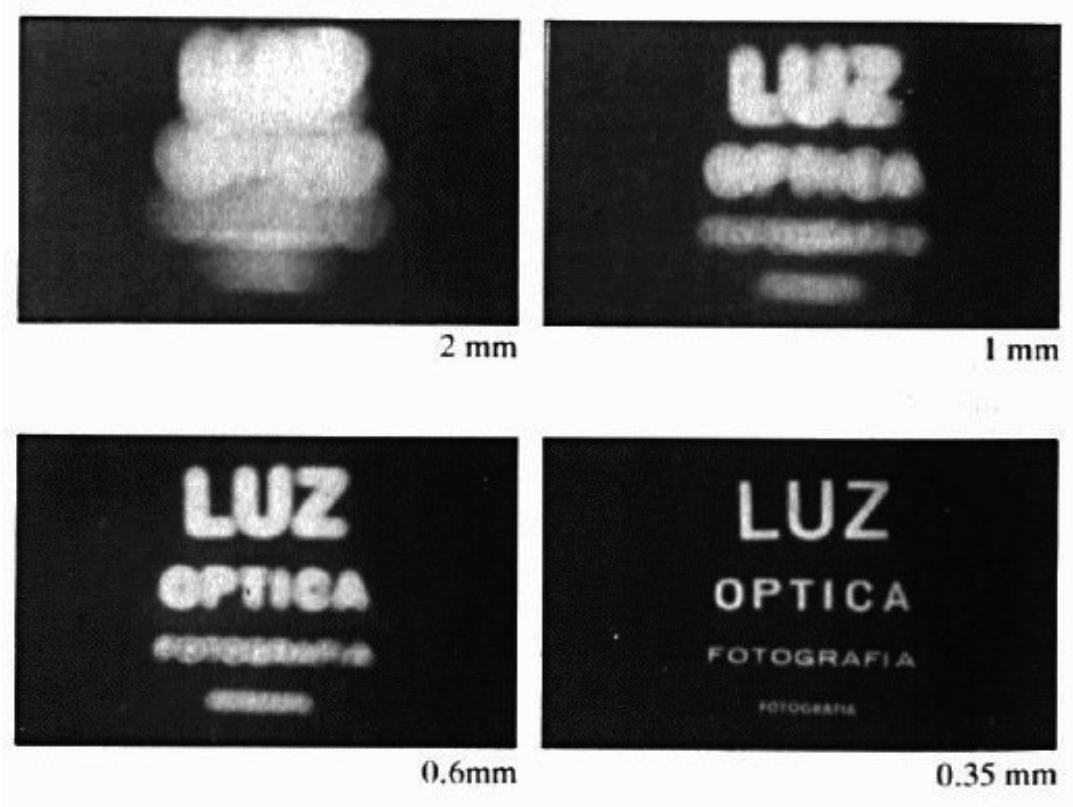
Slide by Steve Seitz

Shrinking the aperture

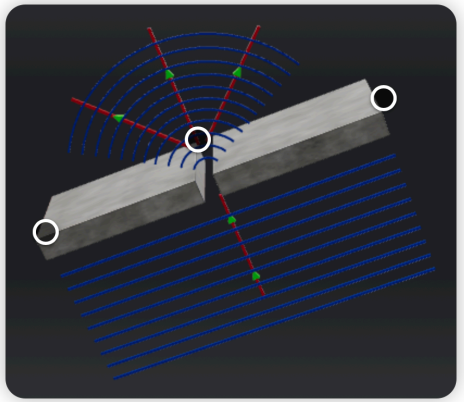


Slide by Steve Seitz

Shrinking the aperture

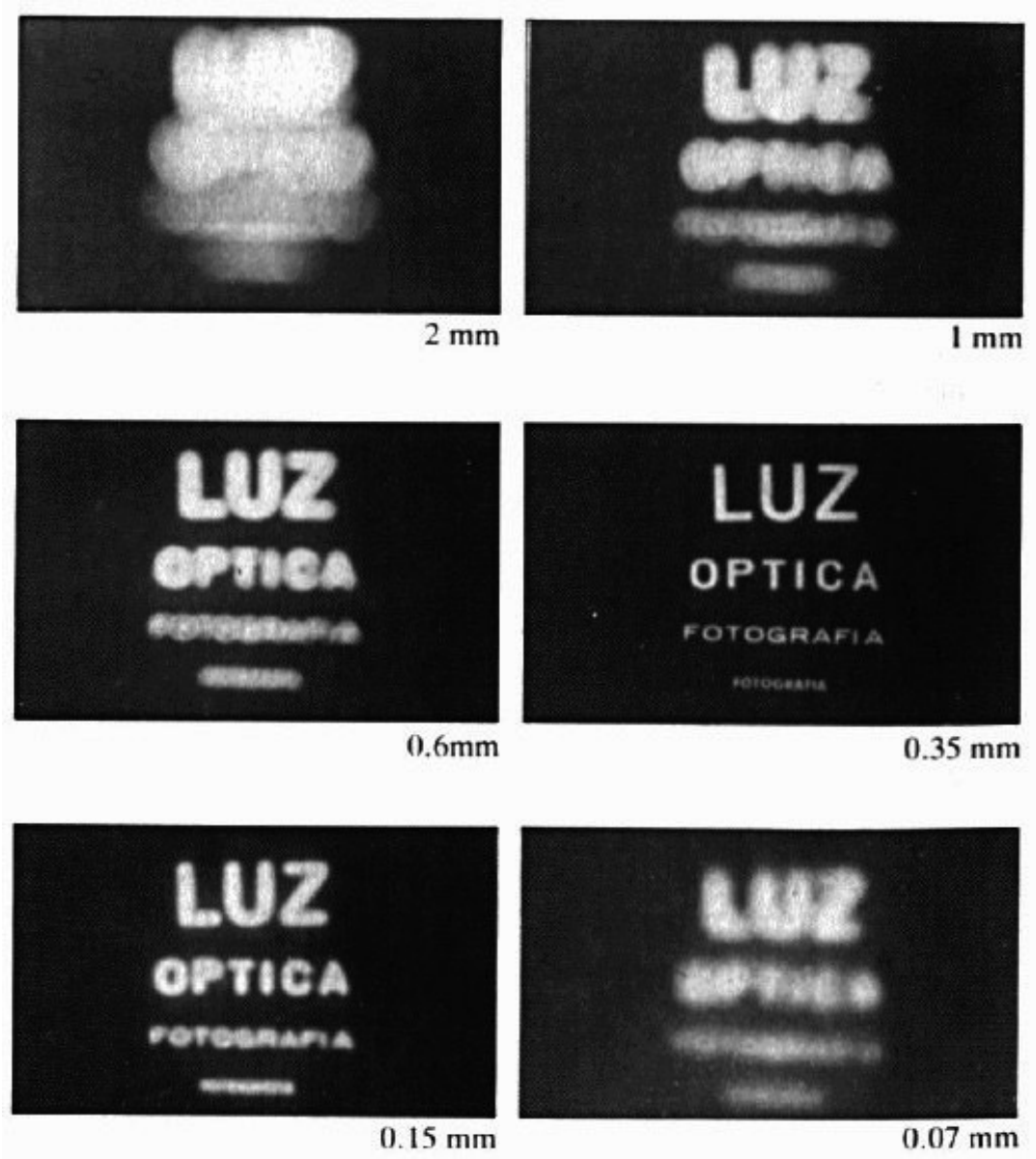


- Why not make the aperture as small as possible?
  - Less light gets through
  - Diffraction effects…



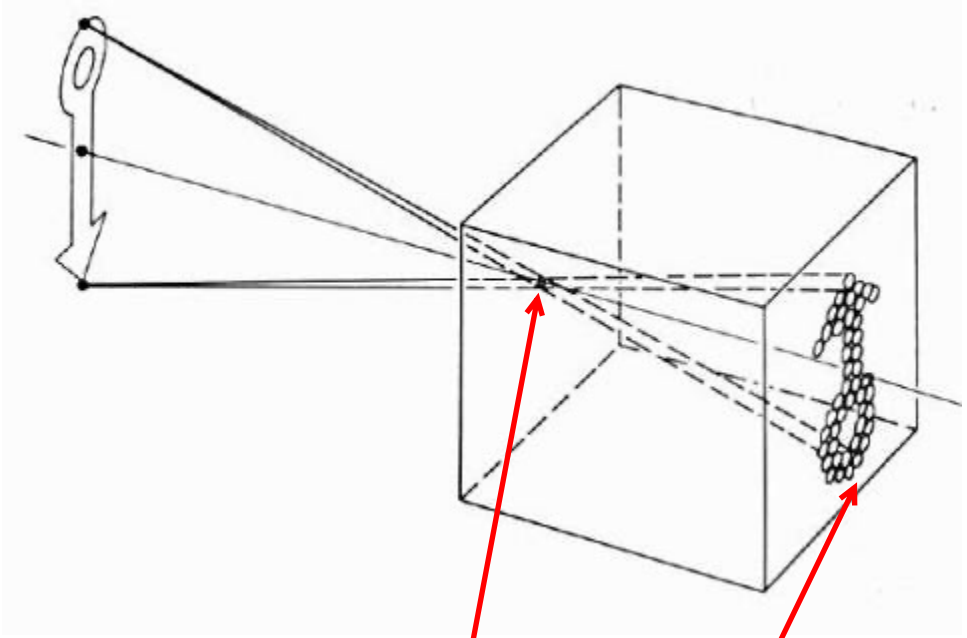
Slide by Steve Seitz

Shrinking the aperture



Extreme small pinhole:
Diffraction effect

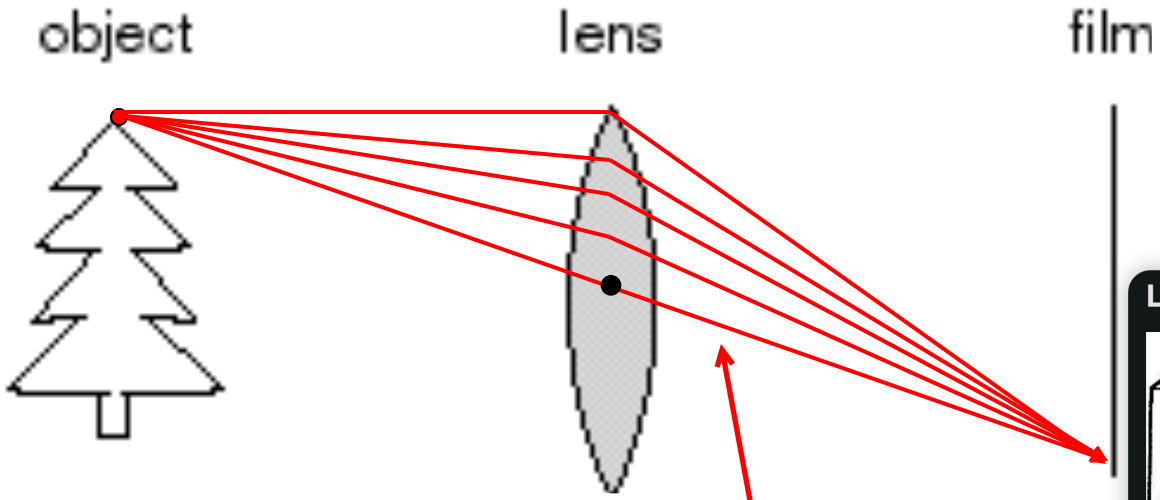
Pinhole camera model

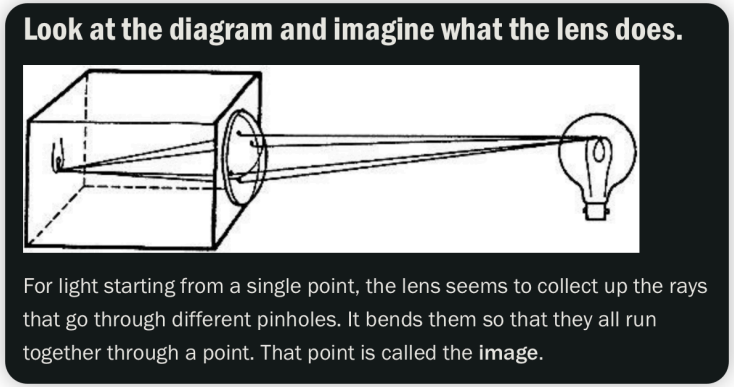


- Pinhole model:
  - Captures ***pencil of rays*** – all rays through a single point
  - This point is called ***Center of Projection (focal point)***
  - The image is formed on the **Image Plane**

Slide by Steve Seitz

Adding a lens… to capture more light

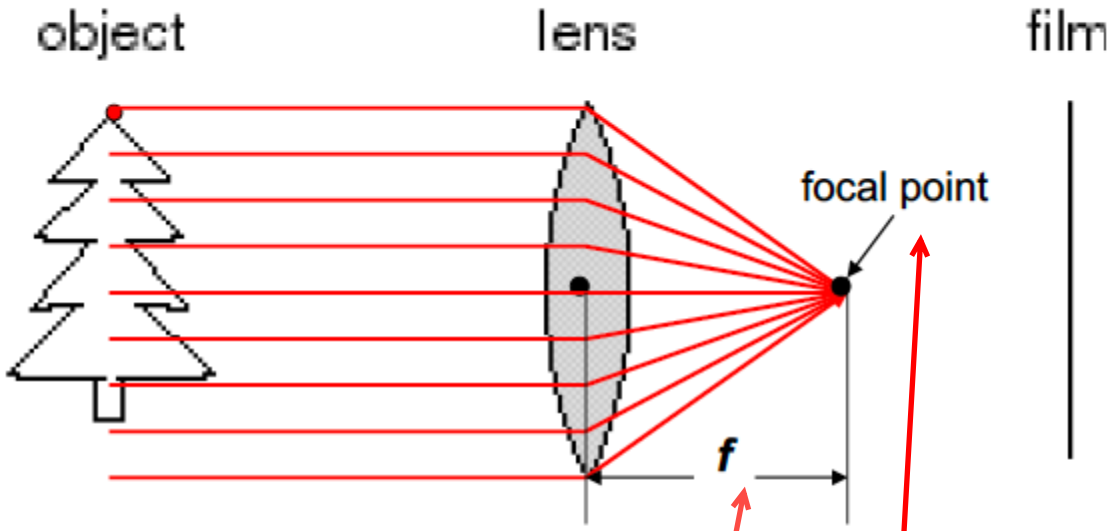




- A lens focuses light onto the film
  - Rays passing through the center are not deviated

Slide by Steve Seitz

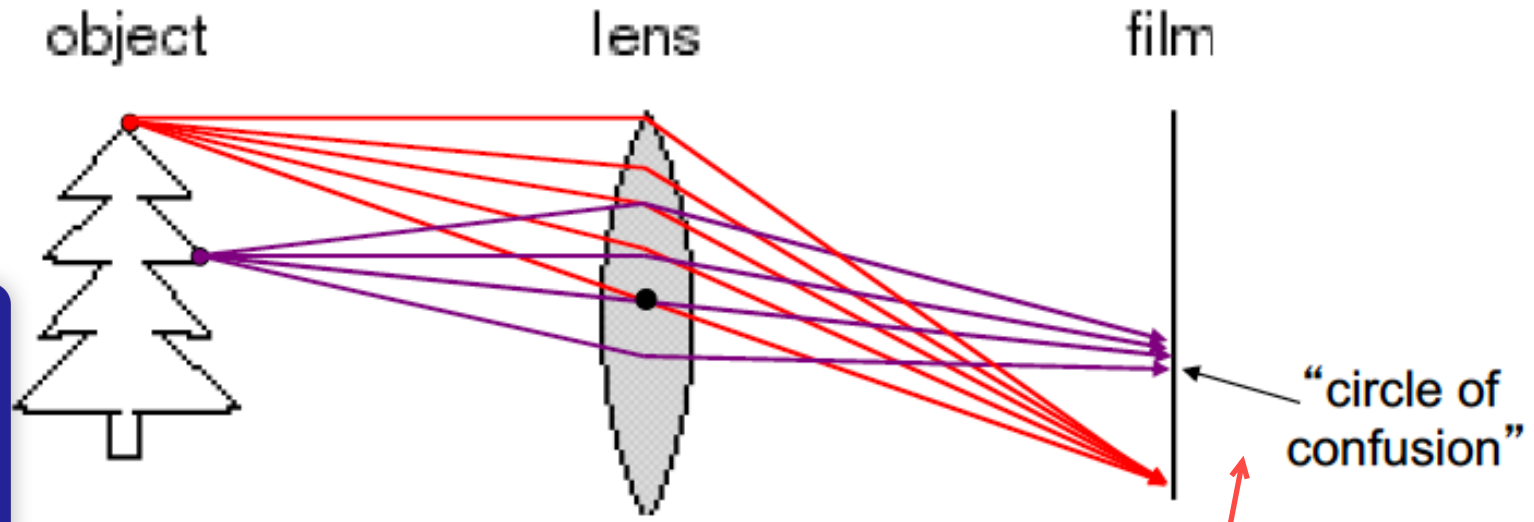
Adding a lens… to capture more light



- A lens focuses light onto the film
  - Rays passing through the center are not deviated
  - All parallel rays converge to one point on a plane located at the focal length f

Slide by Steve Seitz

Adding a Lens



What is the circle of confusion?

The **circle of confusion** is the measurement of where a point of light grows to a circle you can see in the final image. Also called the zone of confusion, it's measured in fractions of a millimeter. The circle of confusion is what defines what's in or out of focus. This number is also what calculates depth of field. The circle's size is what affects the sharpness of an image. The smaller the circle, the sharper the image. And the larger the circle, the blurrier. It is often written as CoC.

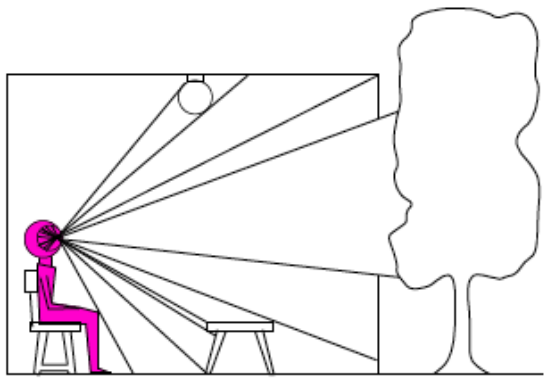
https://youtu.be/Pdq65lEYFOM
https://youtu.be/eJHIVR4_dEE

- A lens focuses light onto the film
  - There is a specific distance at which objects are "in focus"
    - other points project to a "circle of confusion" in the image

Geometric Image Formation: Pinhole Camera Model

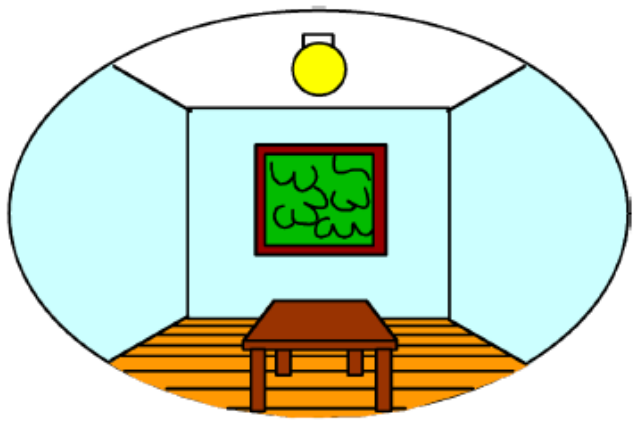
(brief introduction, details to be covered in Multiple-view geometry, Week-7)

3D world



Point of observation

2D image



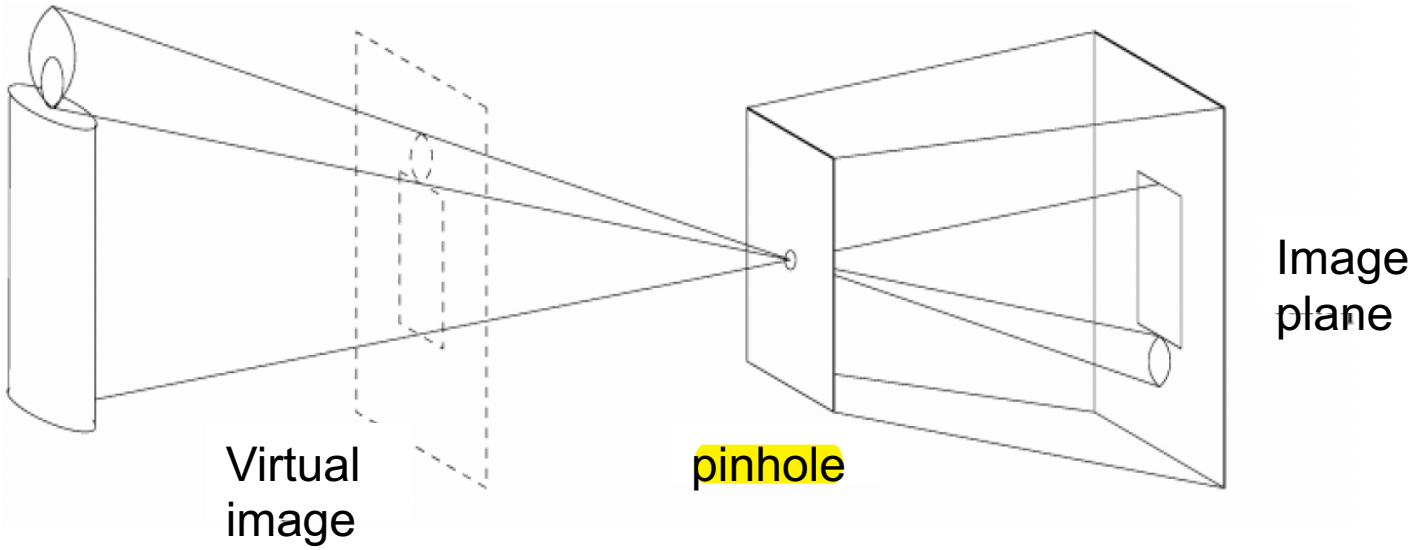
What have we lost?

- Angles
- Distances (lengths)

Slide by A. Efros

Pinhole camera

- Pinhole camera is a simple model to approximate imaging process: **perspective projection**.



- If we treat pinhole as a point, only one ray from any given point can enter the camera.

Fig from Forsyth and Ponce

Image Formation in the Eye

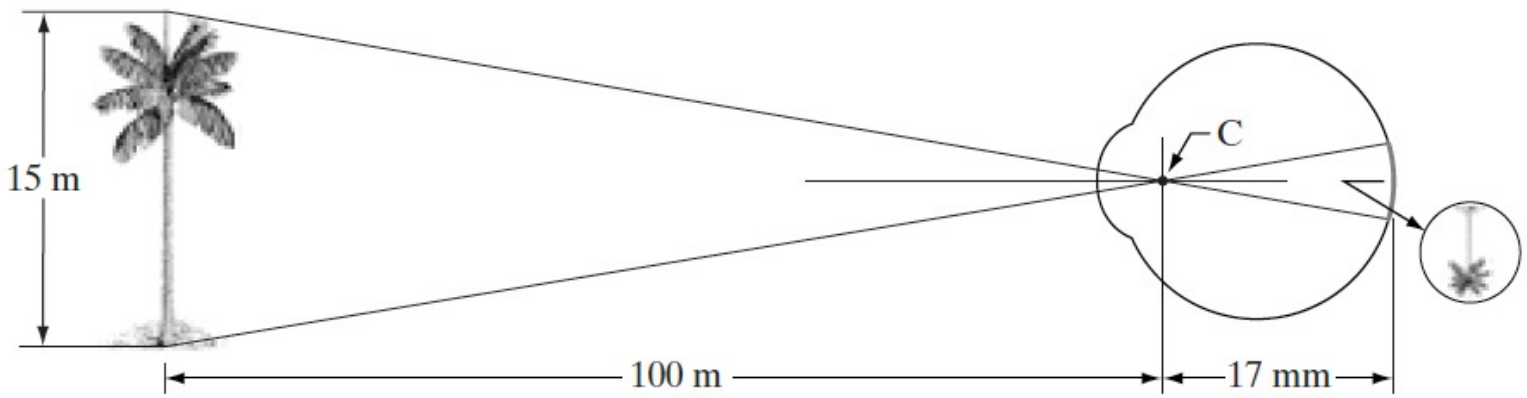


FIGURE 2.3 Graphical representation of the eye looking at a palm tree. Point *C* is the optical center of the lens.

- (reading pp37-38, Image formation in the eye)
- Book: Digital Image Processing, Second Edition

Linear Algebra

① $P = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ $P \in R^3$

② $p = \begin{pmatrix} u \\ v \end{pmatrix}$ $p \in R^2$

③ P and p are vectors.

④ I: identity matrix

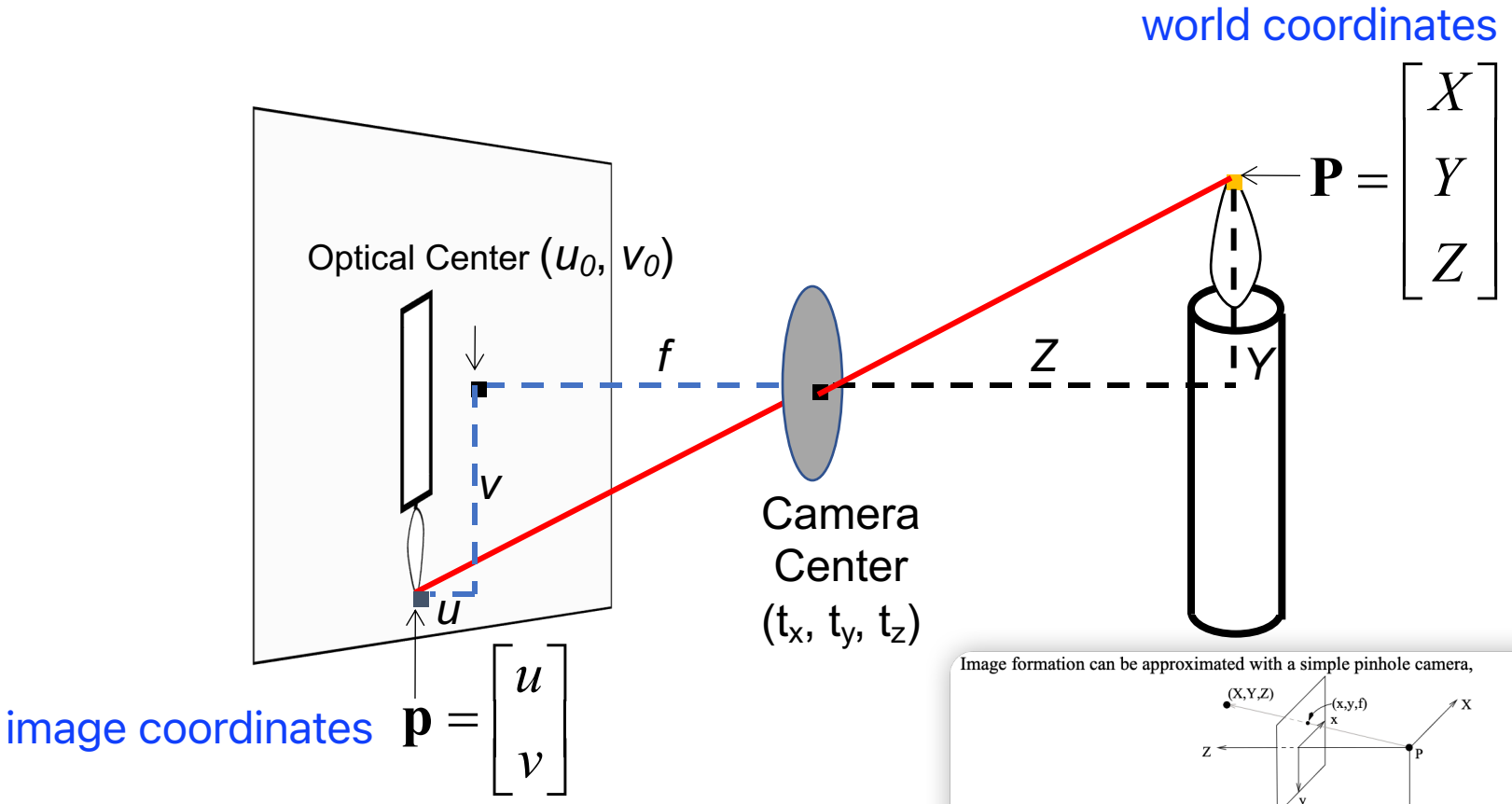
$[I \ 0]$

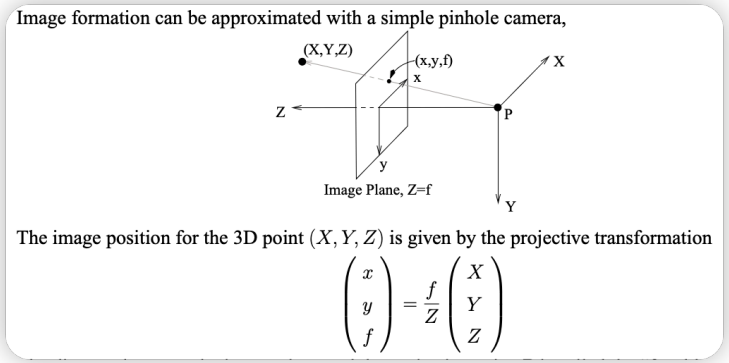
Assume $I \in R^{3\times3}$

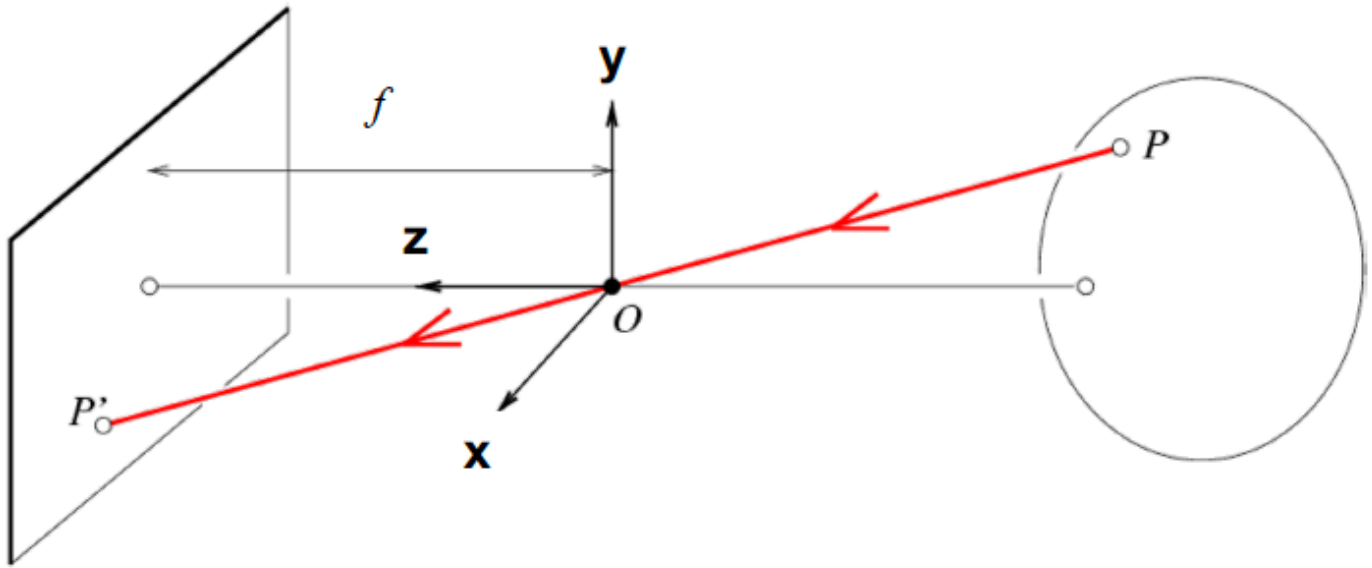
$0 \in R^{3\times1}$

then: $[I \ 0]_{3\times4}$

Projection: world coordinates → image coordinates







- Projection equations
  - Compute intersection with image plane of ray from $P = (x,y,z)$ to O
  - Derived using similar triangles

$$(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z}, f)$$

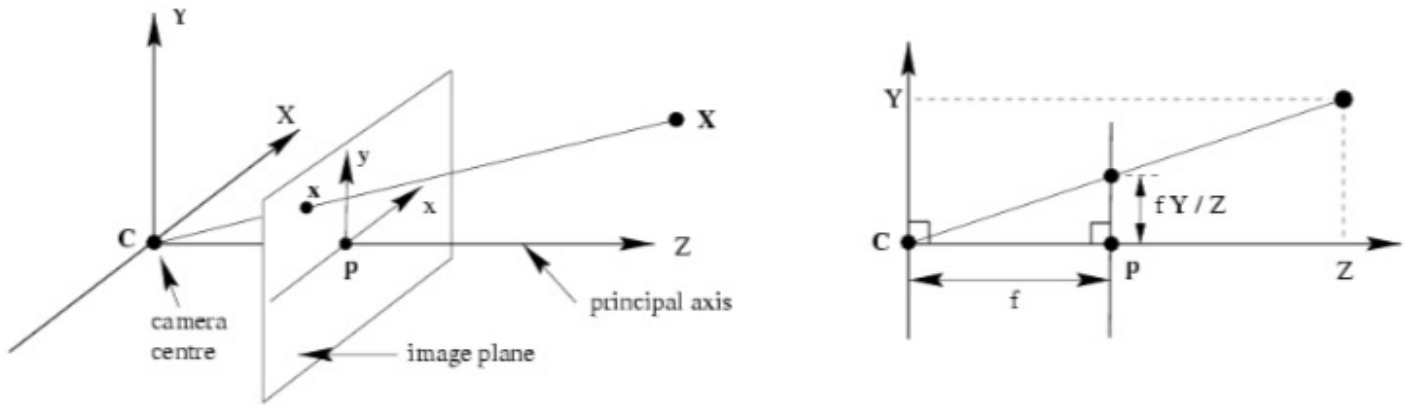
- We get the projection by throwing out the last coordinate:

$$(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z})$$

Source: J. Ponce, S. Seitz

world coor. → img coor.

Pinhole camera model



- By similar triangles

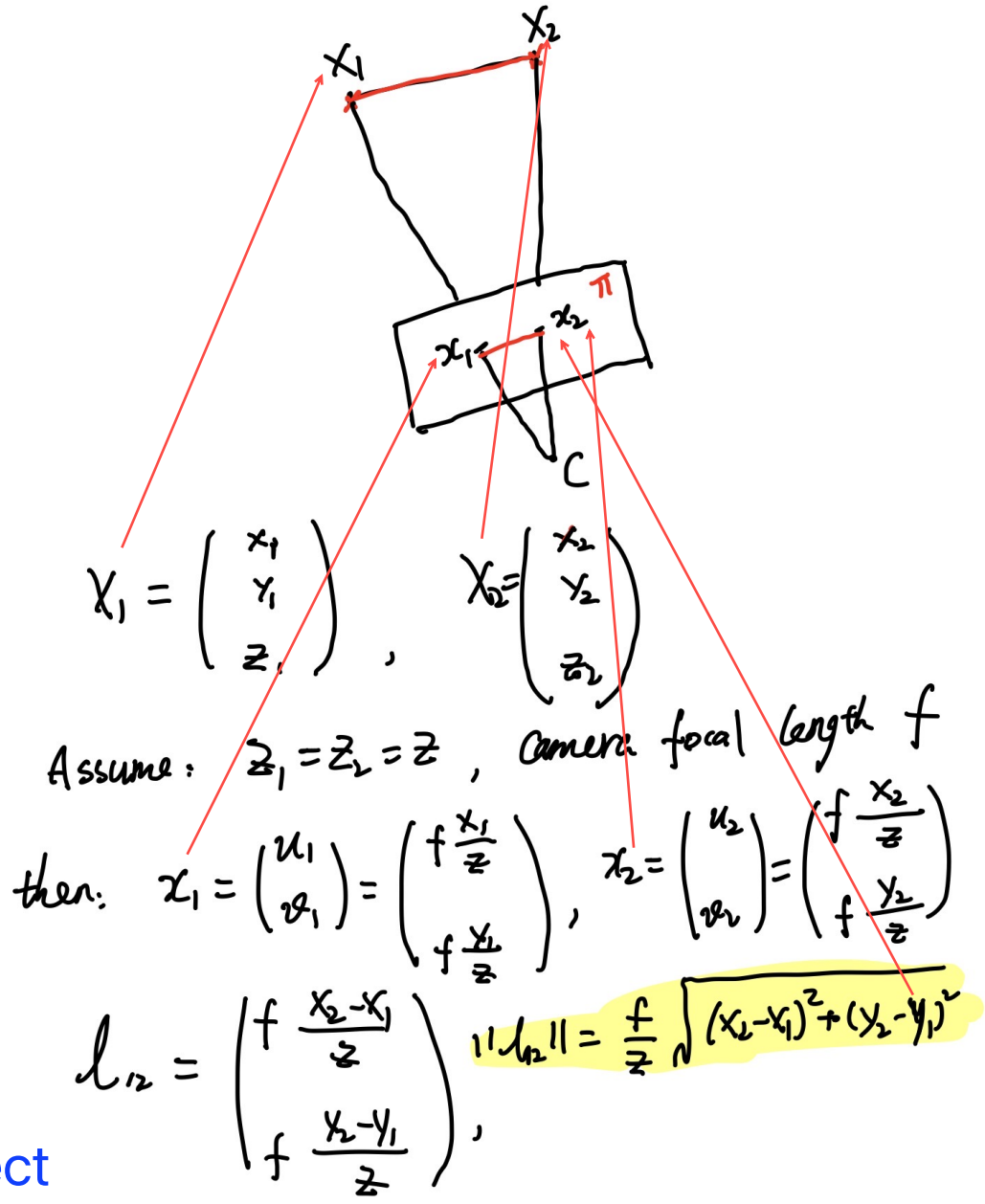
$$(X,Y,Z)^T \mapsto (fX/Z, fY/Z, f)^T$$

- Dropping third coordinate

$$(X,Y,Z)^T \mapsto (fX/Z, fY/Z)^T$$

Explanation of scale and size of the object in the image

- Object far away from the camera is projected as a smaller object in the image
- Object close to the camera is projected as larger size in the image



length scaled by f/z
smalled z, closer object, larger scaling effect

Homogeneous coordinates Representation

Conversion

- Converting from Cartesian *to homogeneous* coordinates

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

- Converting *from* homogeneous coordinates to cartesian

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

the additional w refered as "weight"

Homogeneous Coordinates

- Invariant to scaling

$$k\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ kw \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{kx}{kw} \\ \frac{ky}{kw} \end{bmatrix} = \begin{bmatrix} \frac{x}{w} \\ \frac{y}{w} \end{bmatrix}$$

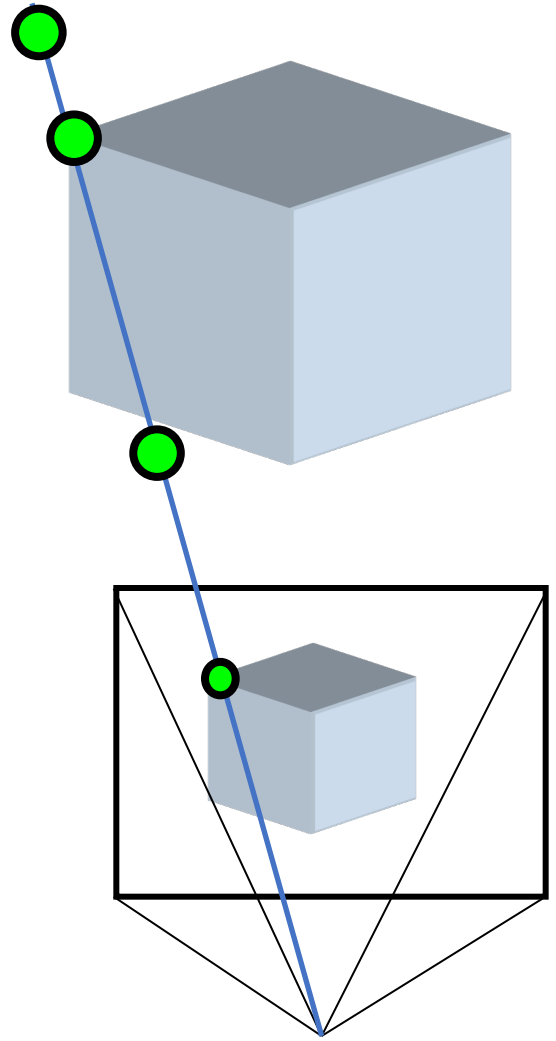
Homogeneous Coordinates

Cartesian Coordinates

- Point in cartesian is a ray in homogeneous

consider k as an arbitrary number represent the scaling effect of the coordinates on the ray of the projection

Example



$$(X, Y, Z)^T \rightarrow \quad (fX/Z, fY/Z)$$

- Cartesian coordinate represented by homogenous coordinate

Cartesian

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} fX/Z \\ fY/Z \end{pmatrix} \overset{\text{times Z}}{\rightarrow} \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix}$$ homogenous

add 1

homogenous 3D

$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

Perspective effects

3D computer vision aims to solve an inverse problem of computer graphics

Photometric (radiometric) image formation

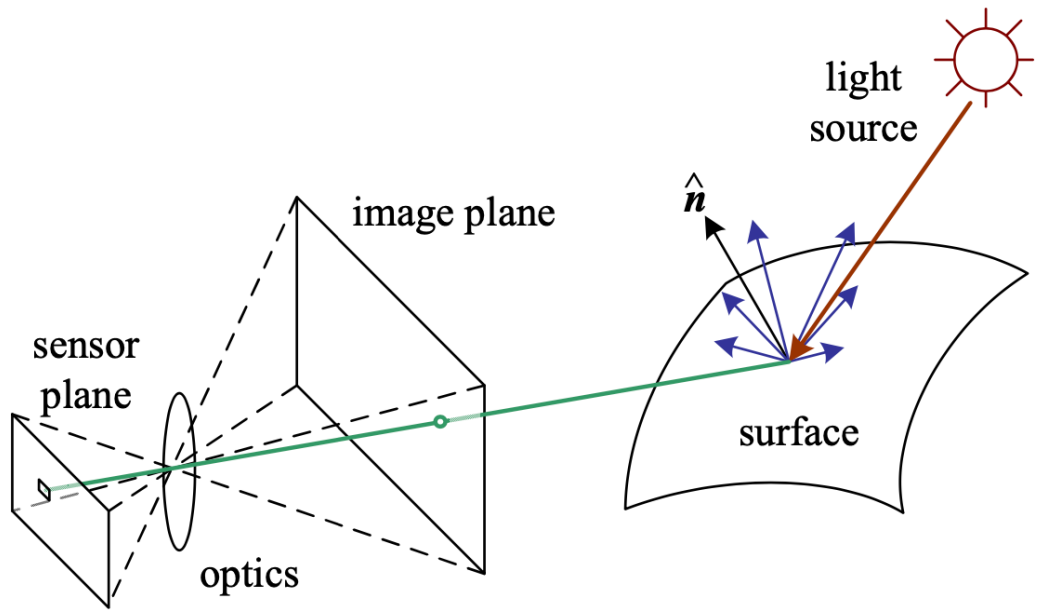
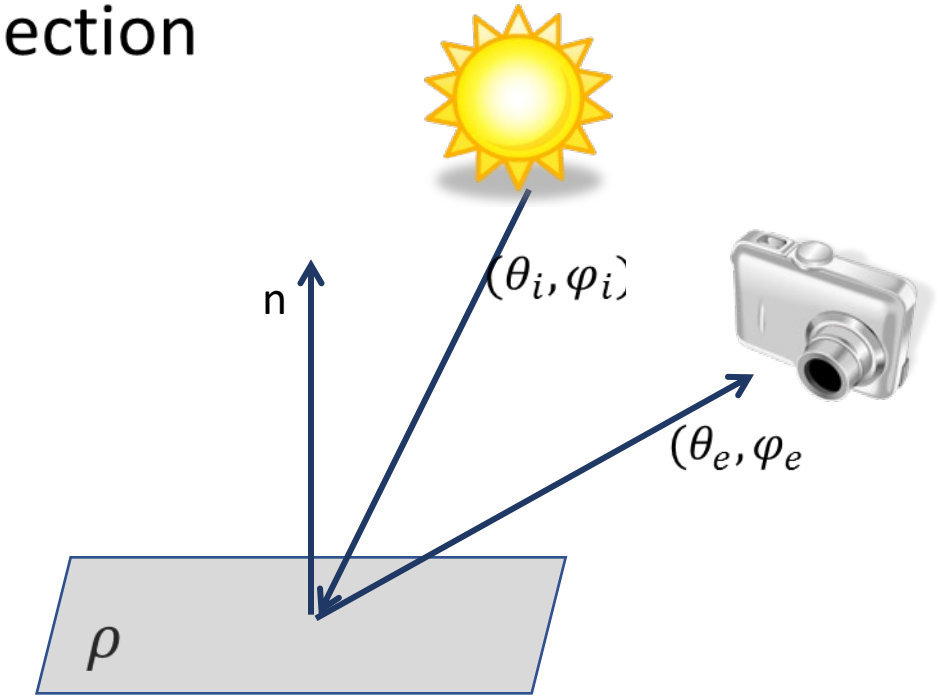


Figure 2.14 A simplified model of photometric image formation. Light is emitted by one or more light sources and is then reflected from an object's surface. A portion of this light is directed towards the camera. This simplified model ignores multiple reflections, which often occur in real-world scenes.

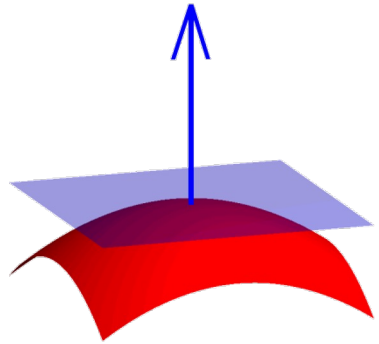
What determines the scene radiance?

- The amount of light that falls on the surface
- The fraction of light that is reflected (albedo)
- Geometry of light reflection
  - Shape of surface
  - Viewpoint

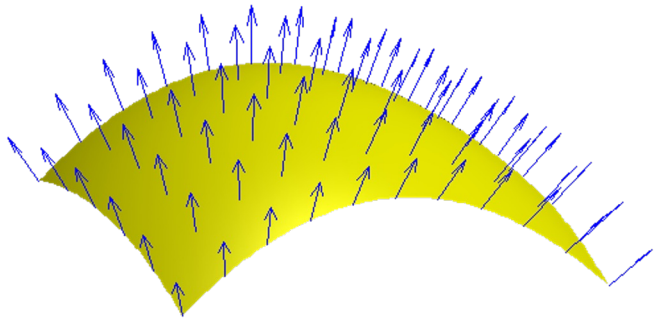


Surface Normal

- Convenient notation for surface orientation
- A smooth surface has a tangent plane at every point



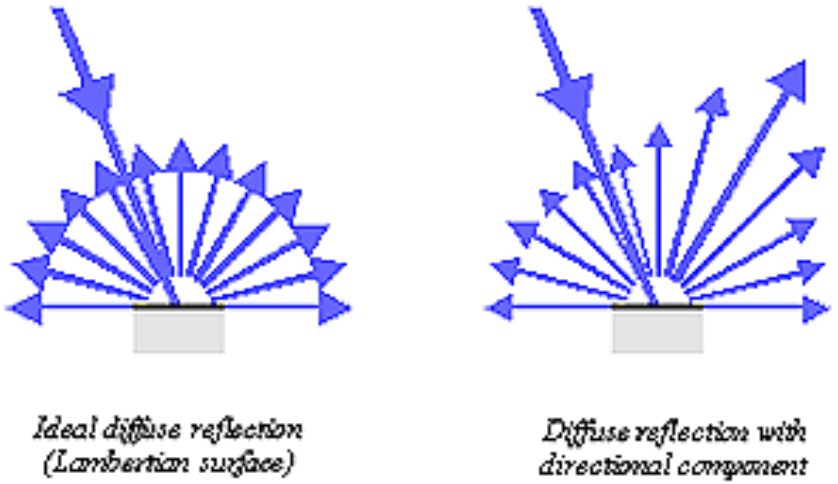
- We can model the surface using the normal at every point

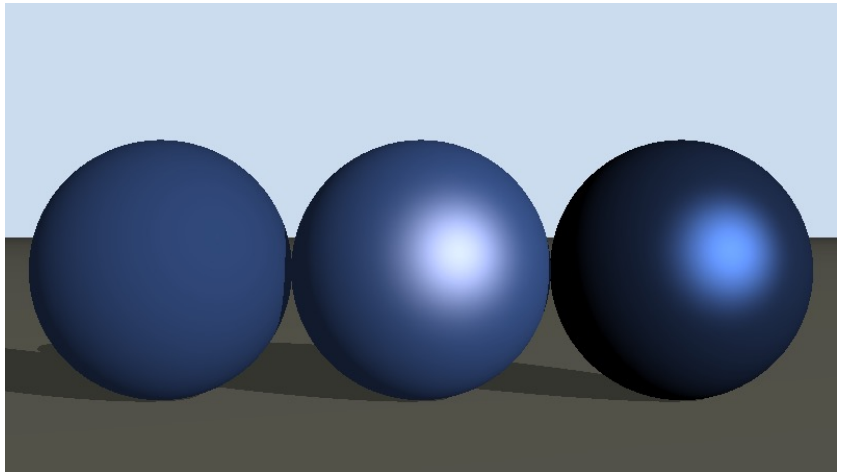


Lambertian surface

- Ideal diffuse reflectors.

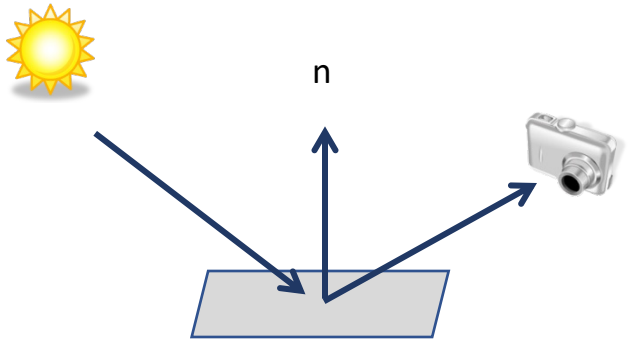
- The apparent brightness of such a surface to an observer is the same regardless of the observer's angle of view.



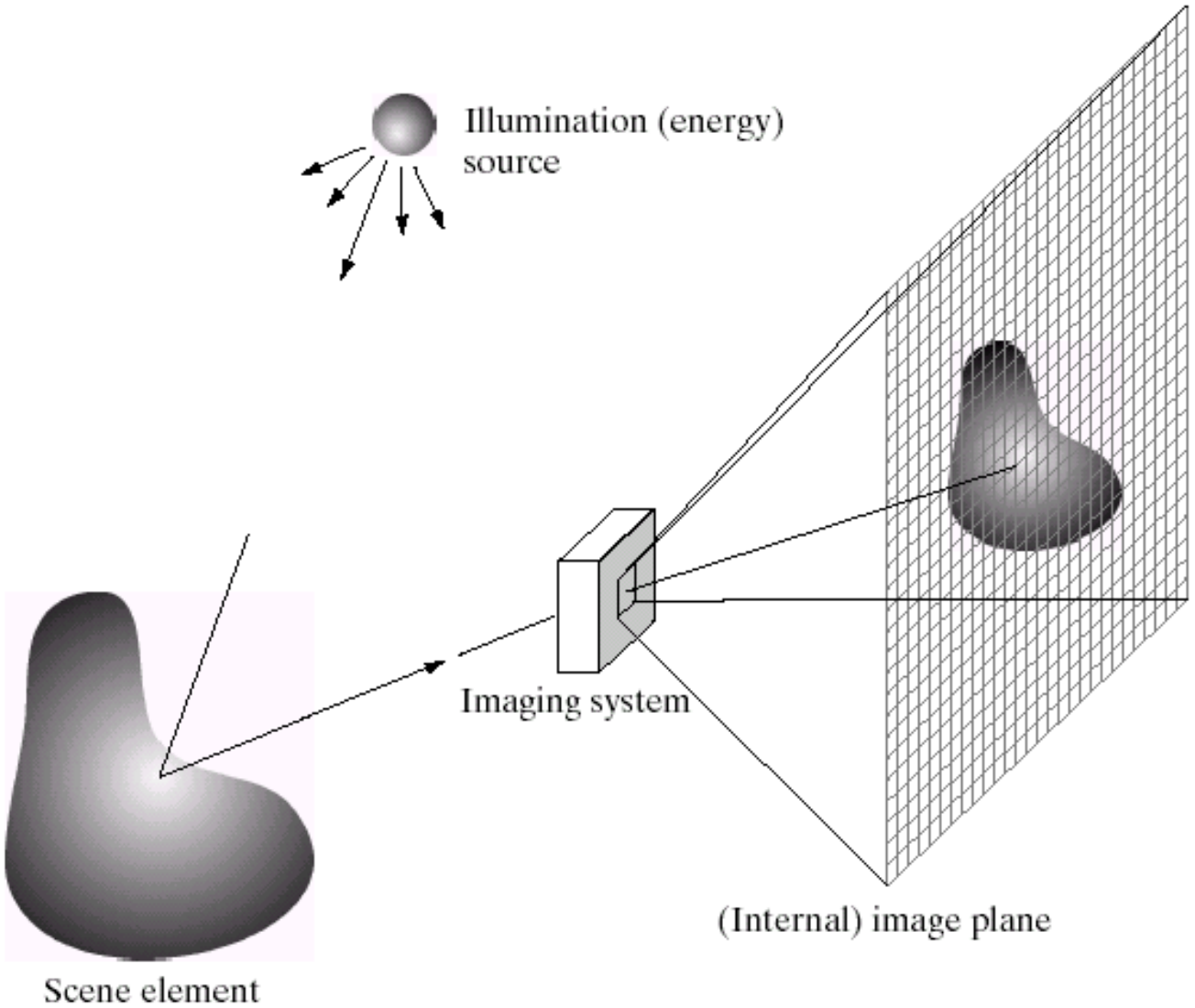


Lambertian Surface

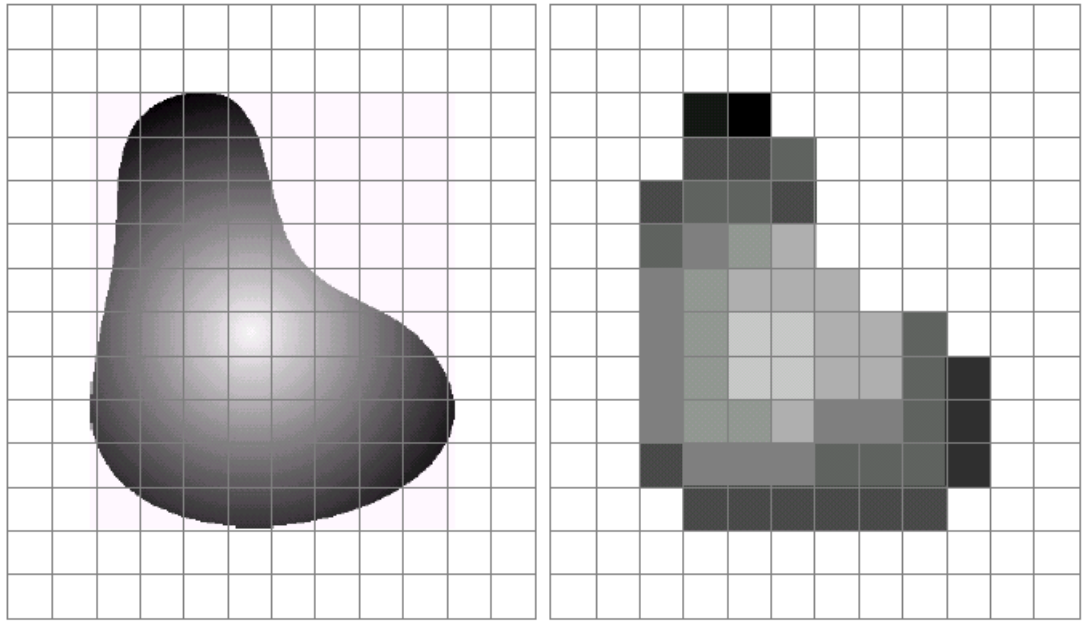
- Appears equally bright from all viewing directions
- Reflects all light without absorbing
- Matte surface, no “shiny” spots
- Brightness of the surface as seen from camera is linearly correlated to the amount of light falling on the surface



Photometric Image Formation

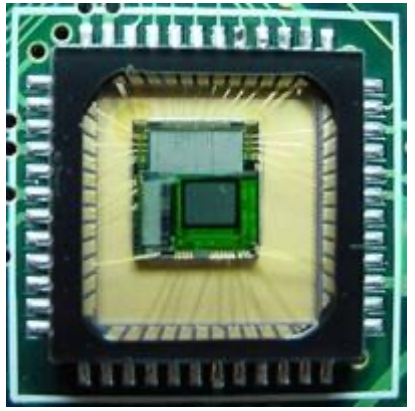


Sensor Array



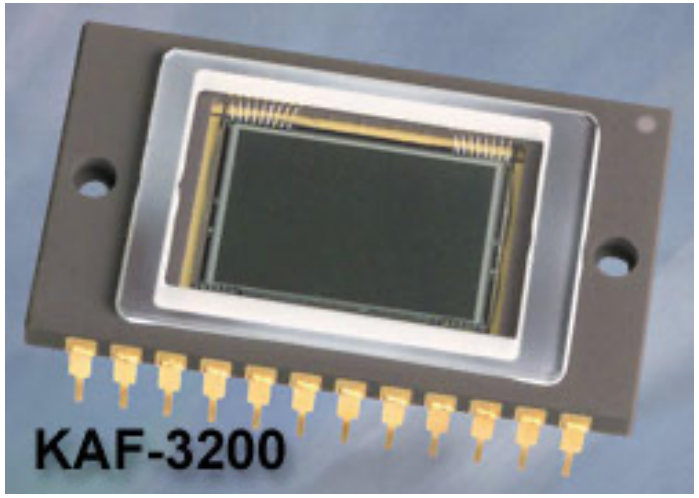
a b

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



CCD/CMOS sensor

Image Sensors : Array Sensor

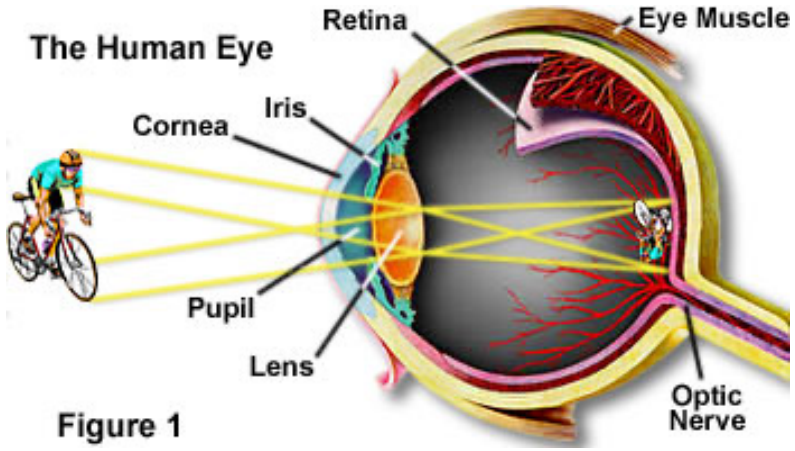


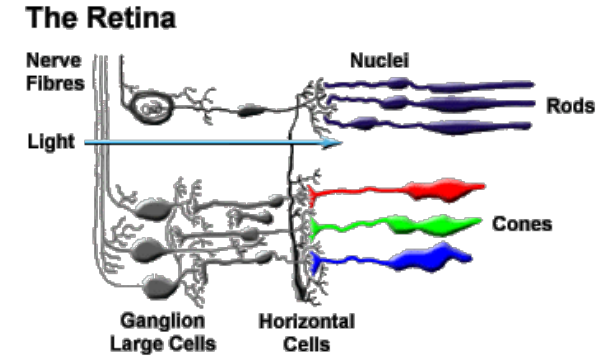
CCD KAF-3200E from Kodak.
(2184 x 1472 pixels, Pixel size 6.8 microns^2)

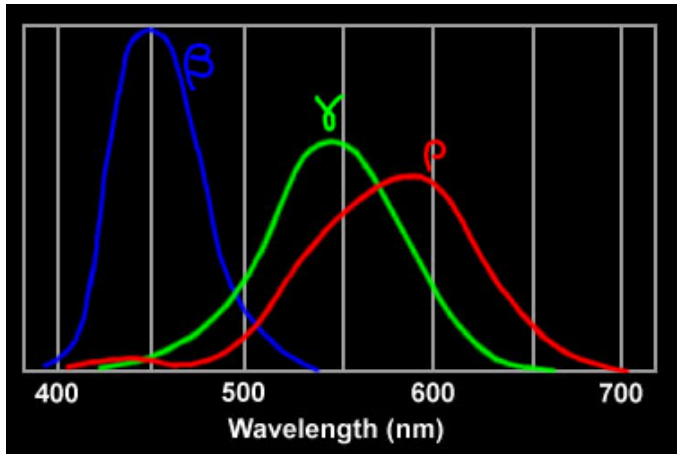
Charge-Coupled Device (CCD)

- Used for convert a continuous image into a digital image
- Contains an array of light sensors
- Converts photon into electric charges accumulated in each sensor unit

Human Color Perception

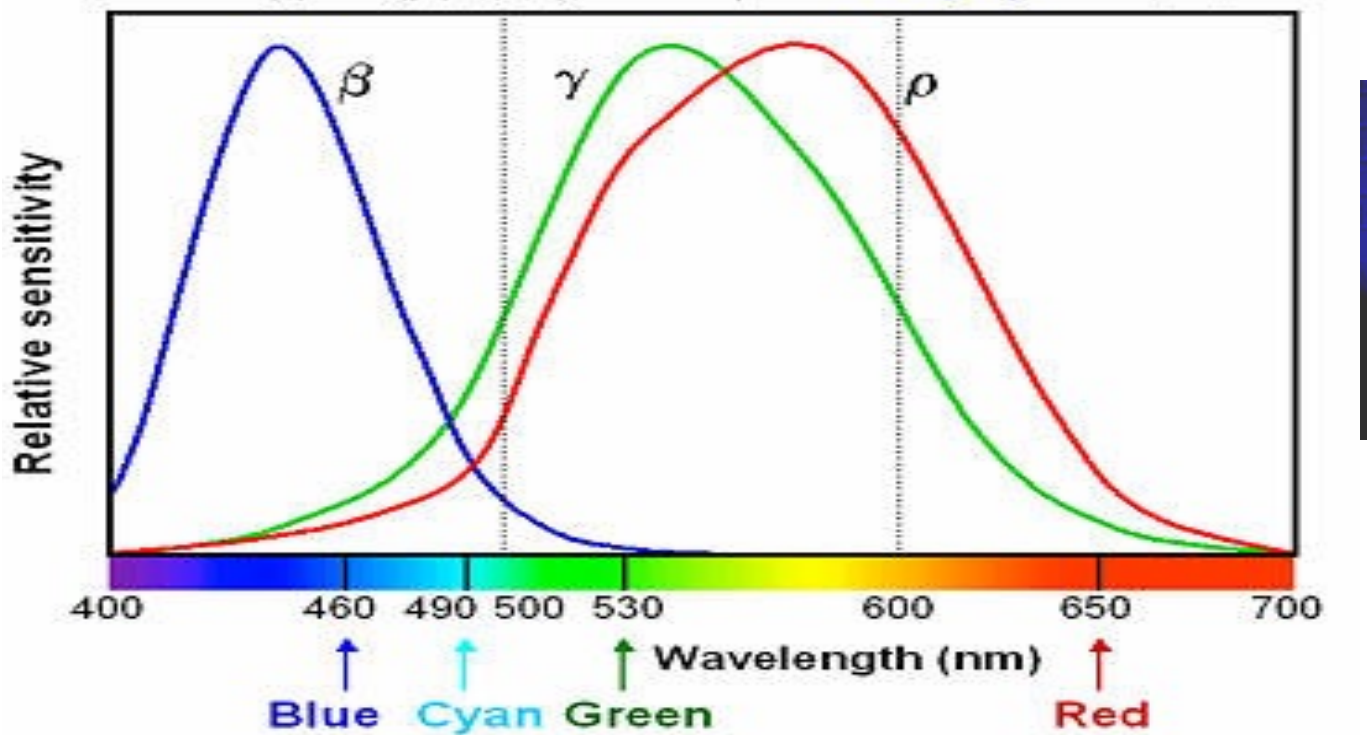


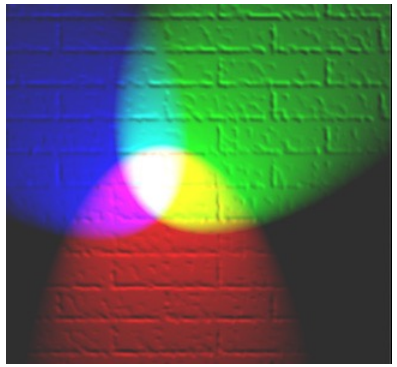




https://en.wikipedia.org/wiki/Color_vision

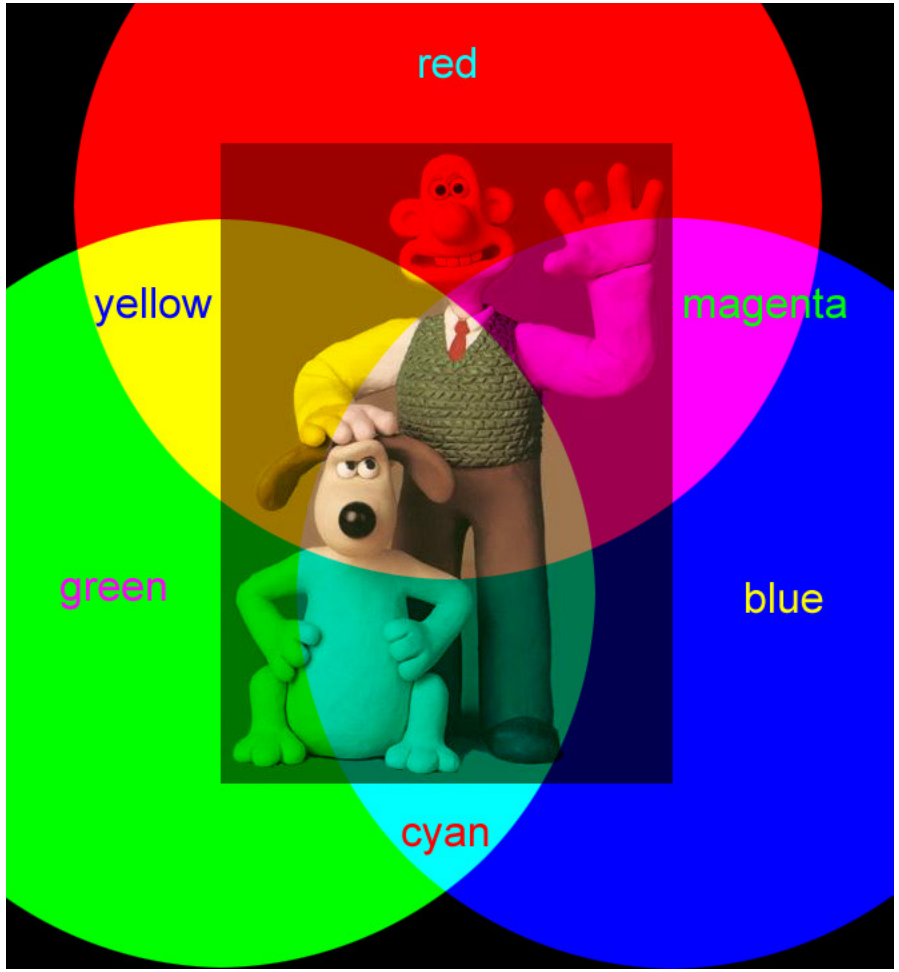
- Wavelength of the light





Colour Images

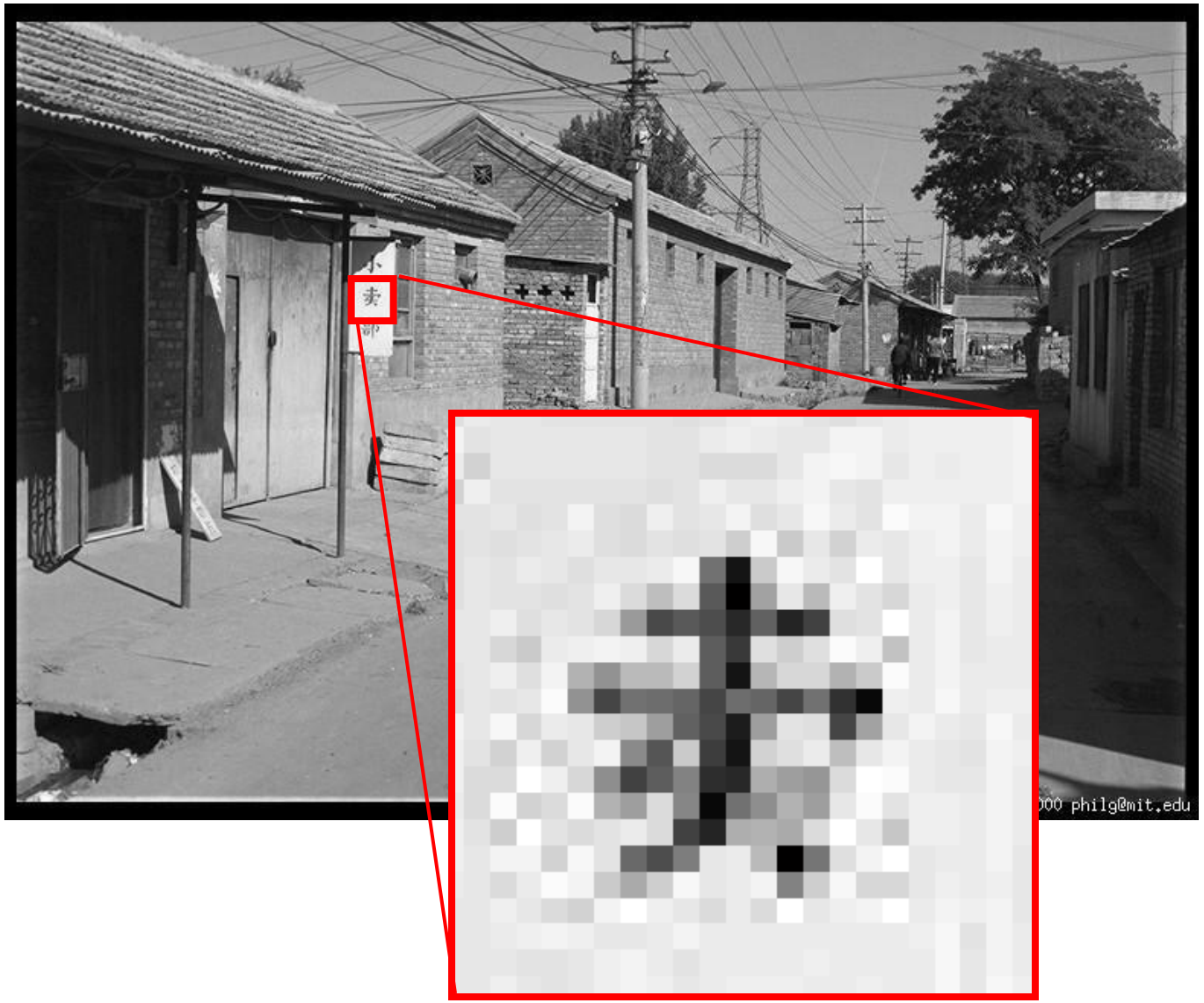
- are constructed from three intensity maps.
- Each intensity map is projected through a colour filter (*e.g.,* red, green, or blue, or cyan, magenta, or yellow) to create a monochrome image.
- The intensity maps are overlaid to create a color image.
- Each pixel in a color image is a three-dimensional vector.



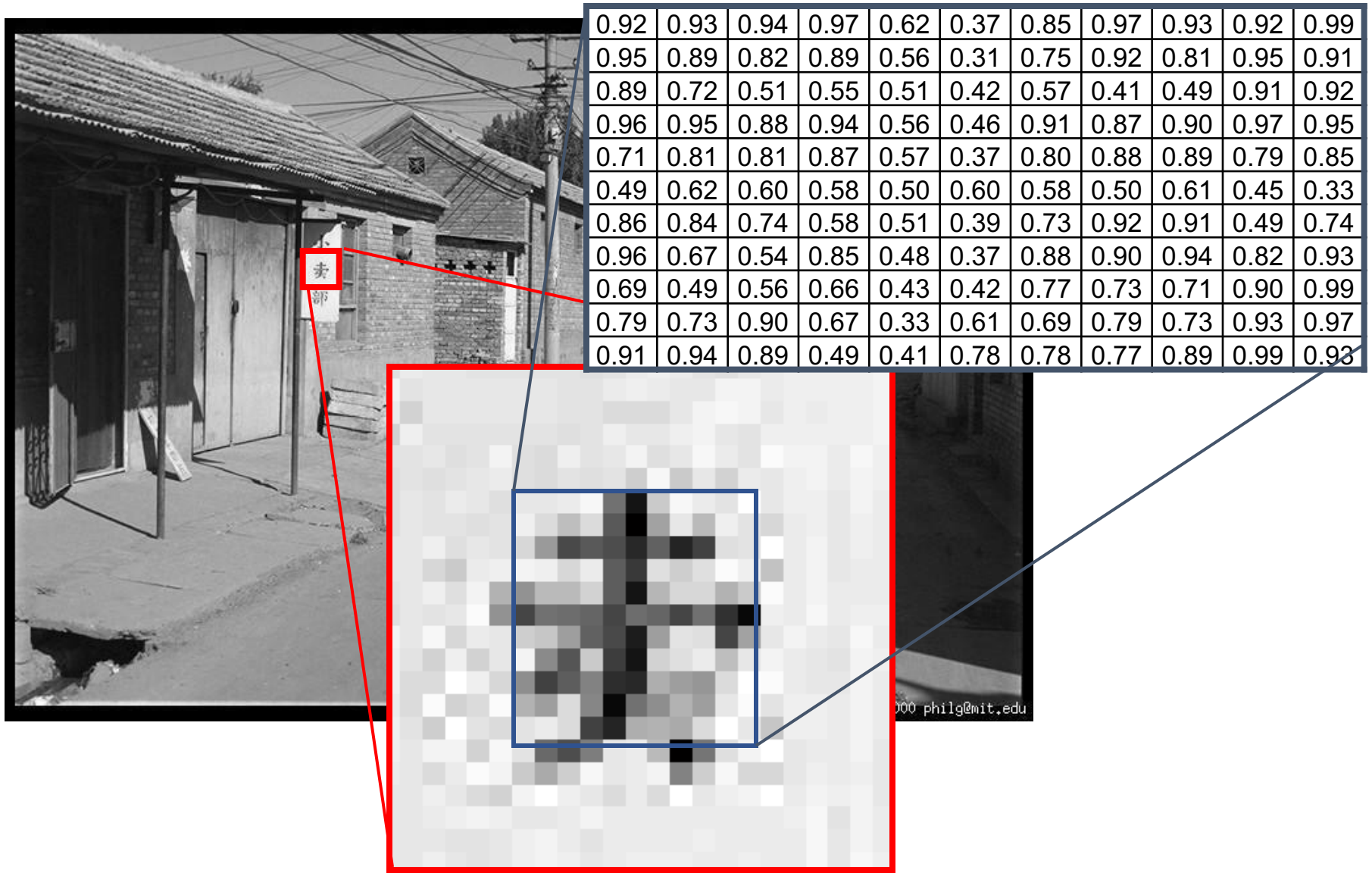
75

Digital image representation

The digital image (pixel matrix)



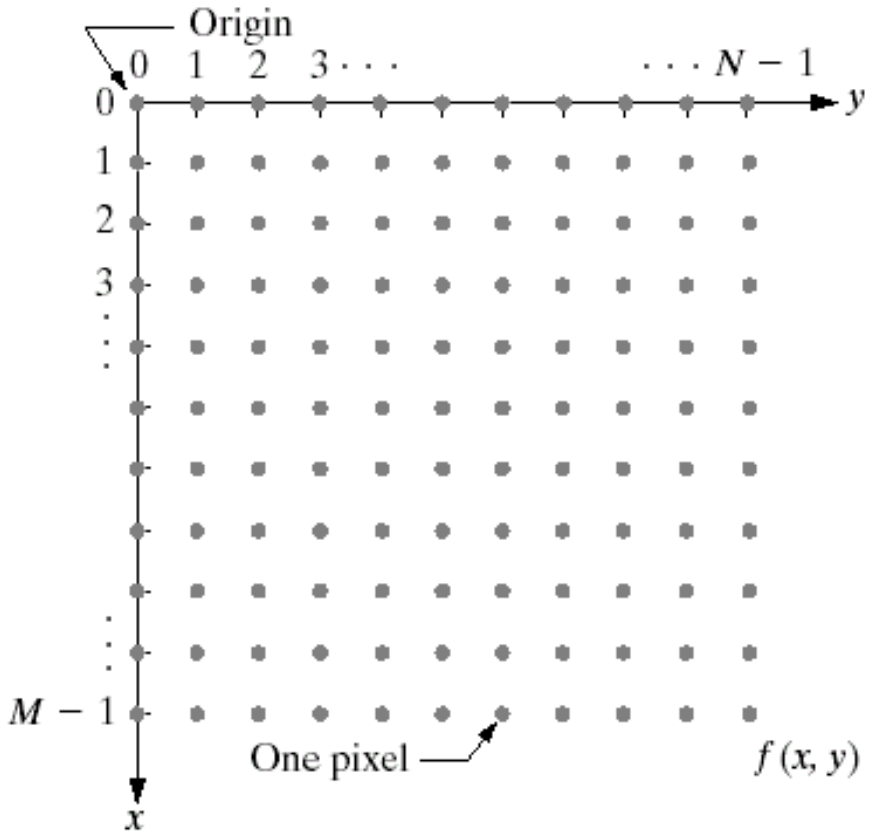
The digital image (pixel matrix)



Fundamentals of Digital Images

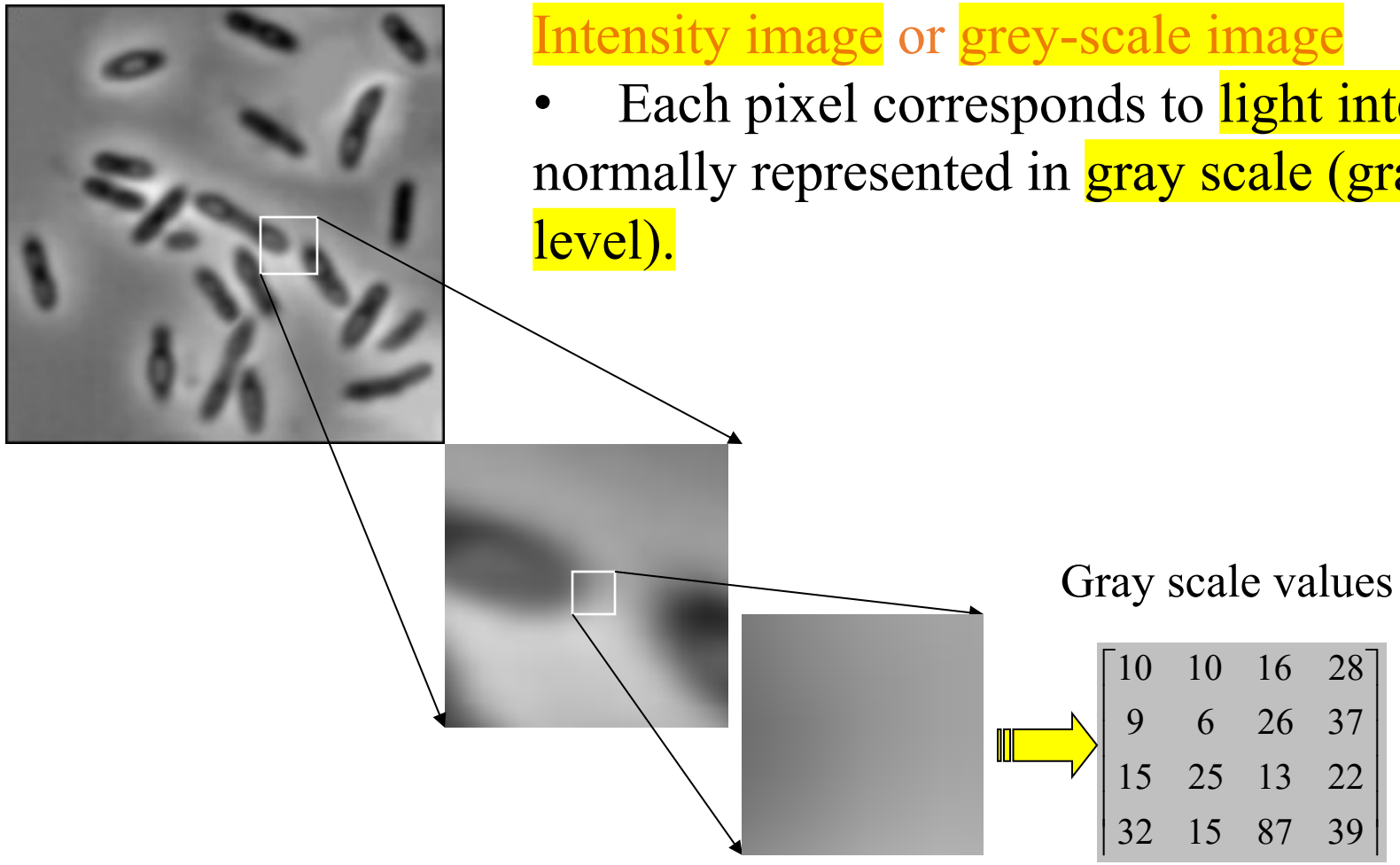
- An image = a function of spatial coordinates.
- Spatial coordinate: (x,y) for 2D case such as photograph,
  (x,y,z) for 3D case such as CT scan images
  (x,y,t) for video.
- The function f may represent the intensity (for greyscale images) or color (for color images) or other associated values.

Conventional Coordinate for Image Representation



(Images from Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, 2nd Edition.

Grey-scale Image



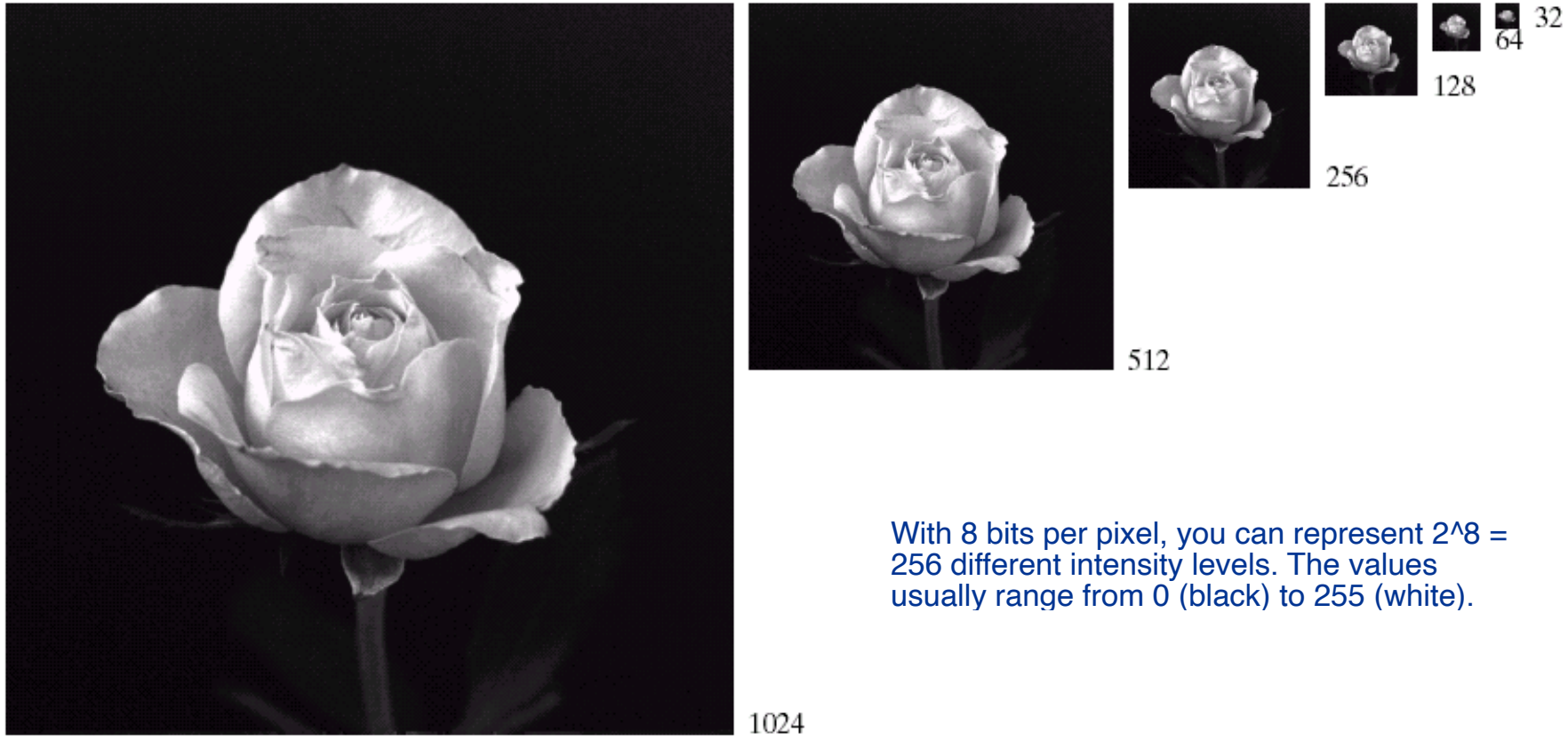
Intensity image or grey-scale image

- Each pixel corresponds to light intensity normally represented in gray scale (gray level).

Gray scale values

$$\begin{bmatrix} 10 & 10 & 16 & 28 \\ 9 & 6 & 26 & 37 \\ 15 & 25 & 13 & 22 \\ 32 & 15 & 87 & 39 \end{bmatrix}$$

Effect of Spatial Resolution

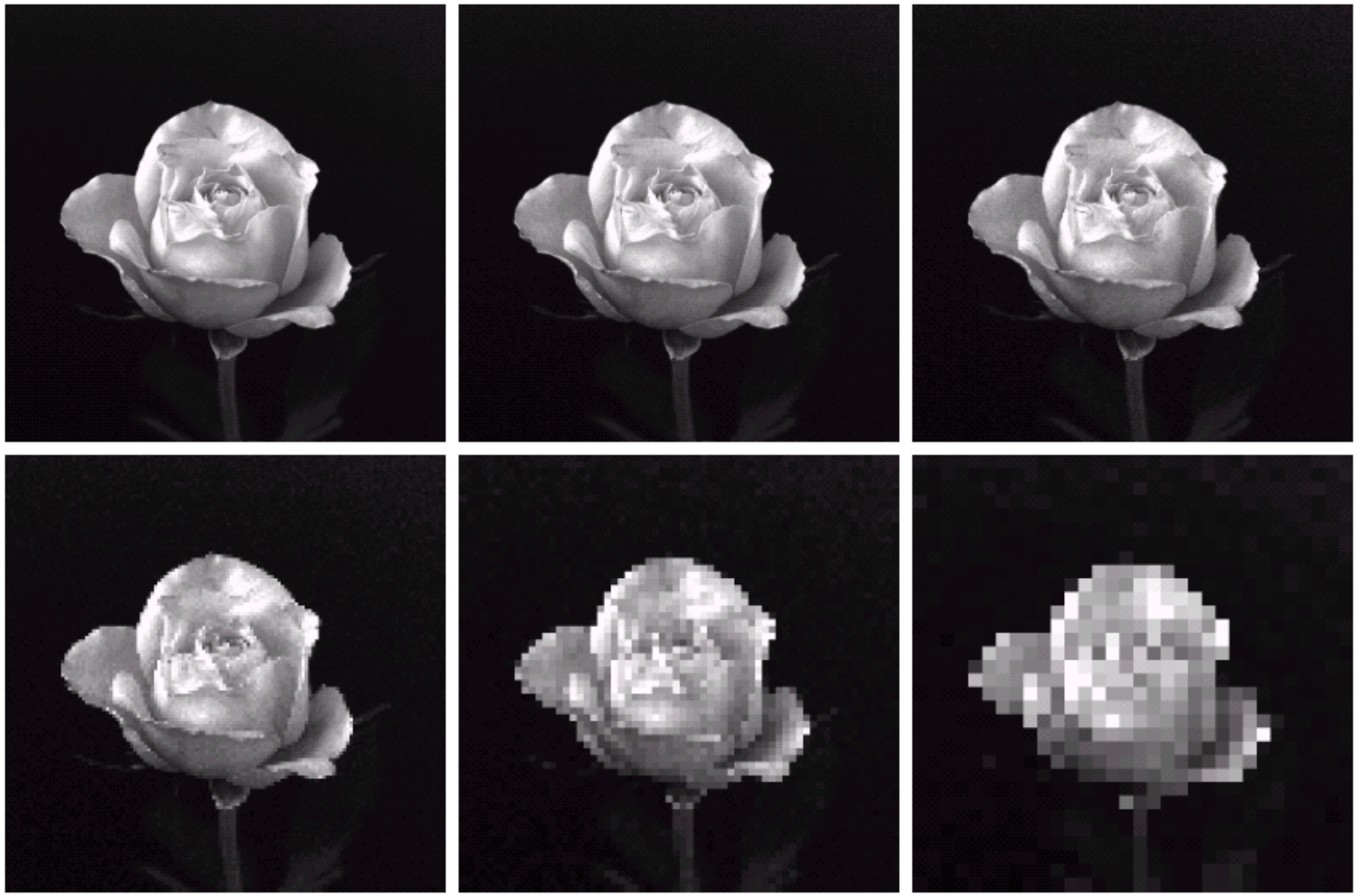


With 8 bits per pixel, you can represent 2^8 = 256 different intensity levels. The values usually range from 0 (black) to 255 (white).

FIGURE 2.19 A 1024 × 1024, 8-bit image subsampled down to size 32 × 32 pixels. The number of allowable gray levels was kept at 256.

(Images from Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, 2nd Edition.

Effect of Spatial Resolution

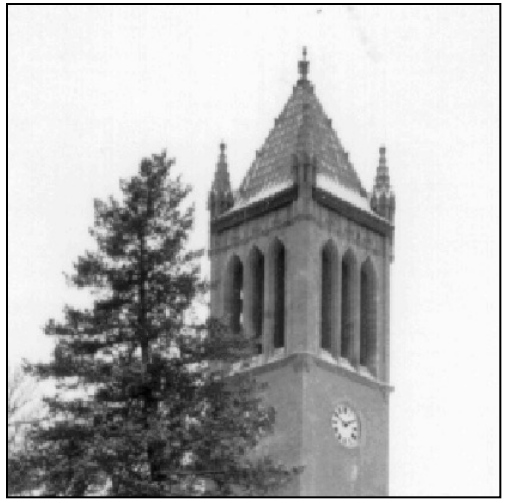


a b c
d e f

FIGURE 2.20 (a) 1024 × 1024, 8-bit image. (b) 512 × 512 image resampled into 1024 × 1024 pixels by row and column duplication. (c) through (f) 256 × 256, 128 × 128, 64 × 64, and 32 × 32 images resampled into 1024 × 1024 pixels.

(Images from Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, 2nd Edition.

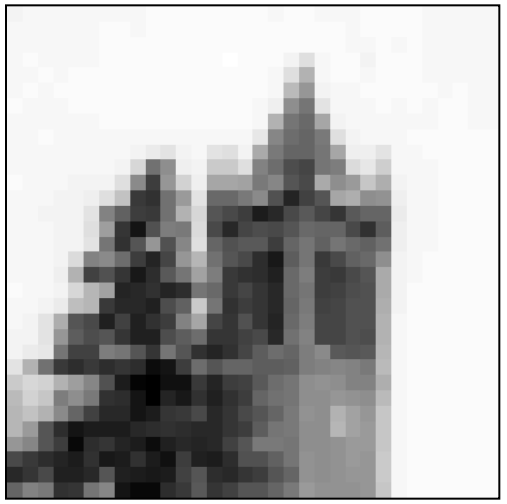
Effect of Spatial Resolution



256x256 pixels

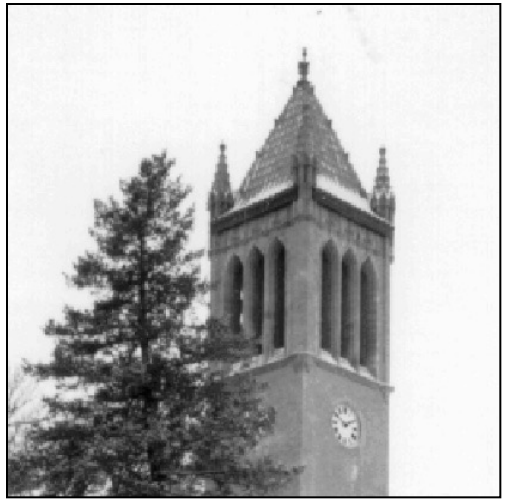
128x128 pixels

64x64 pixels



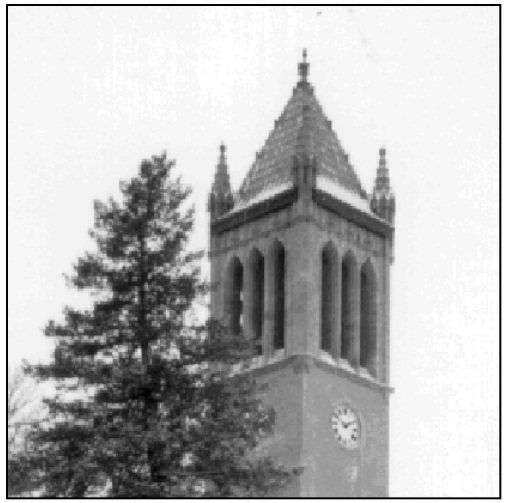
32x32 pixels

Effect of Quantization Levels



8 bit 256 levels

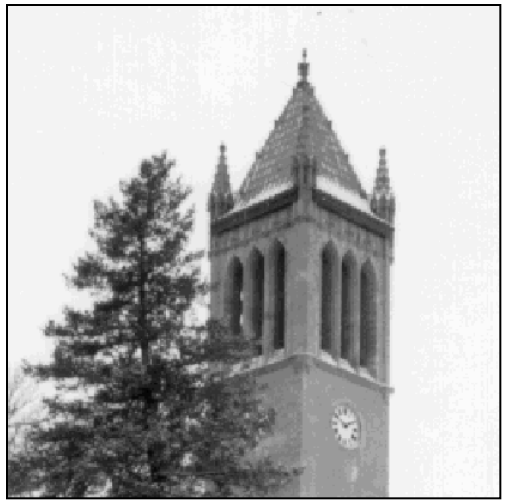
7 bit 128 levels



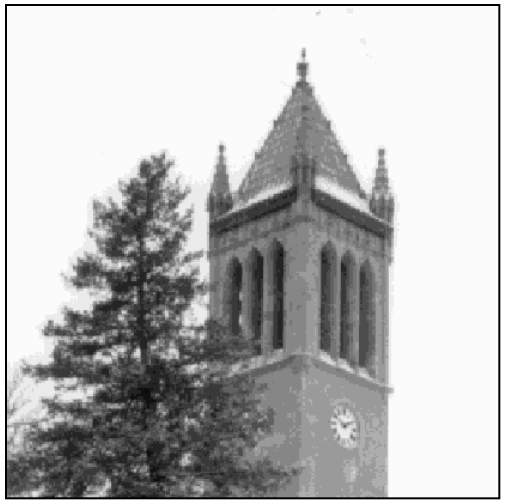
64 levels

32 levels

Effect of Quantization Levels (cont.)

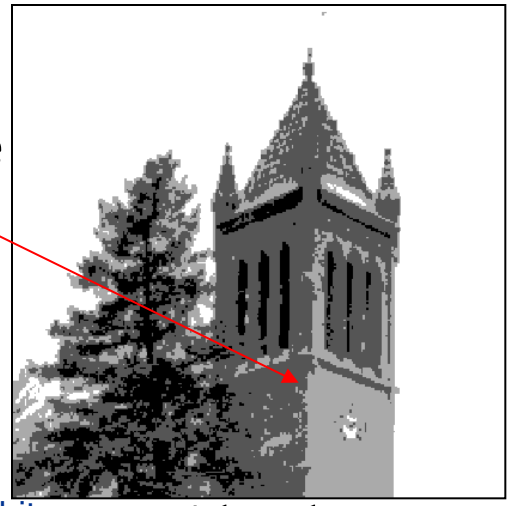


16 levels



8 levels

In this image, it is easy to see false contour.



2 bit

4 levels

1 bit
binary image

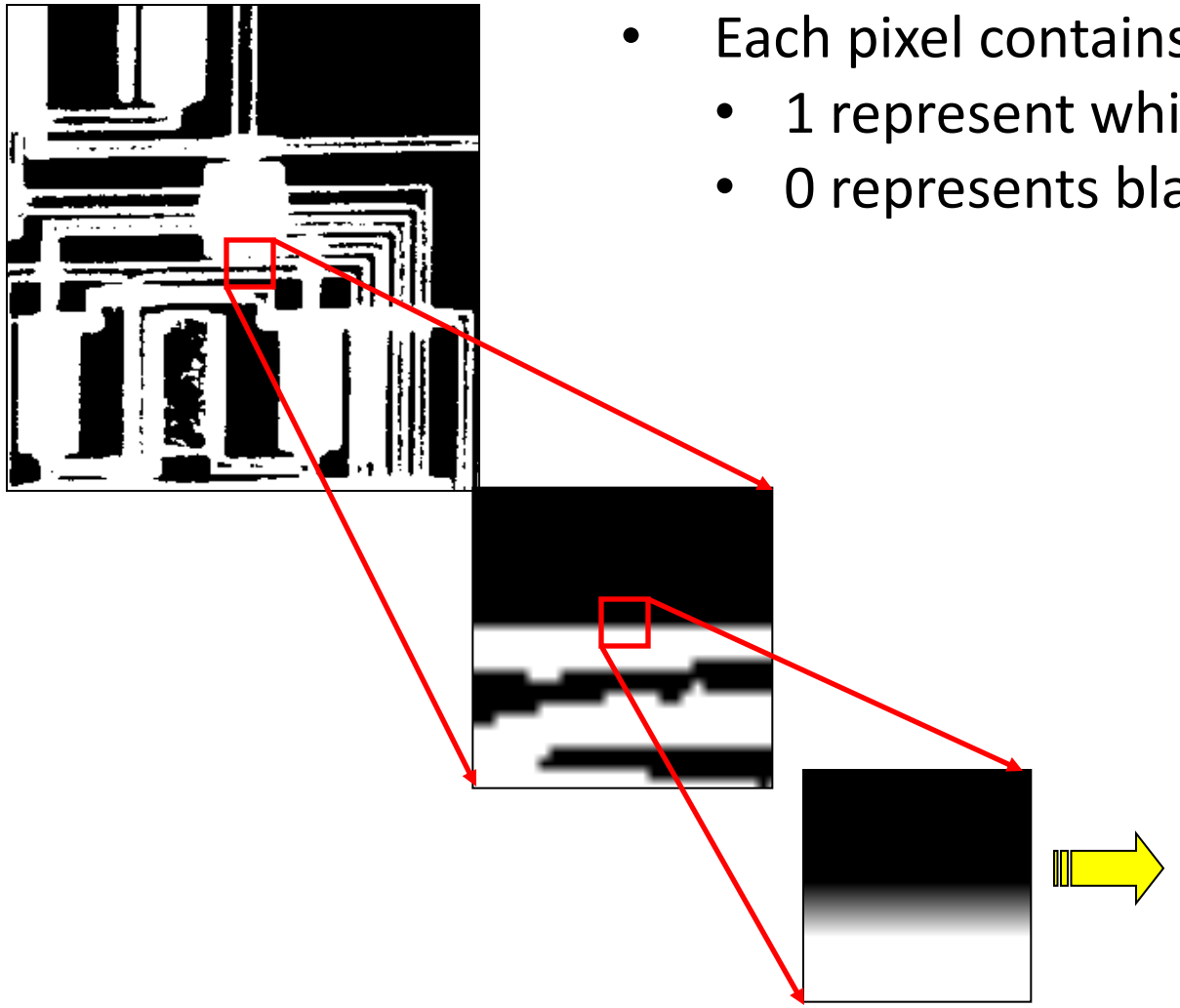
2 levels

Two-level image: binary image

2 levels, 1 bit

Binary image or black and white image

- Each pixel contains one bit :
  - 1 represent white
  - 0 represents black



Binary data

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Colour Images

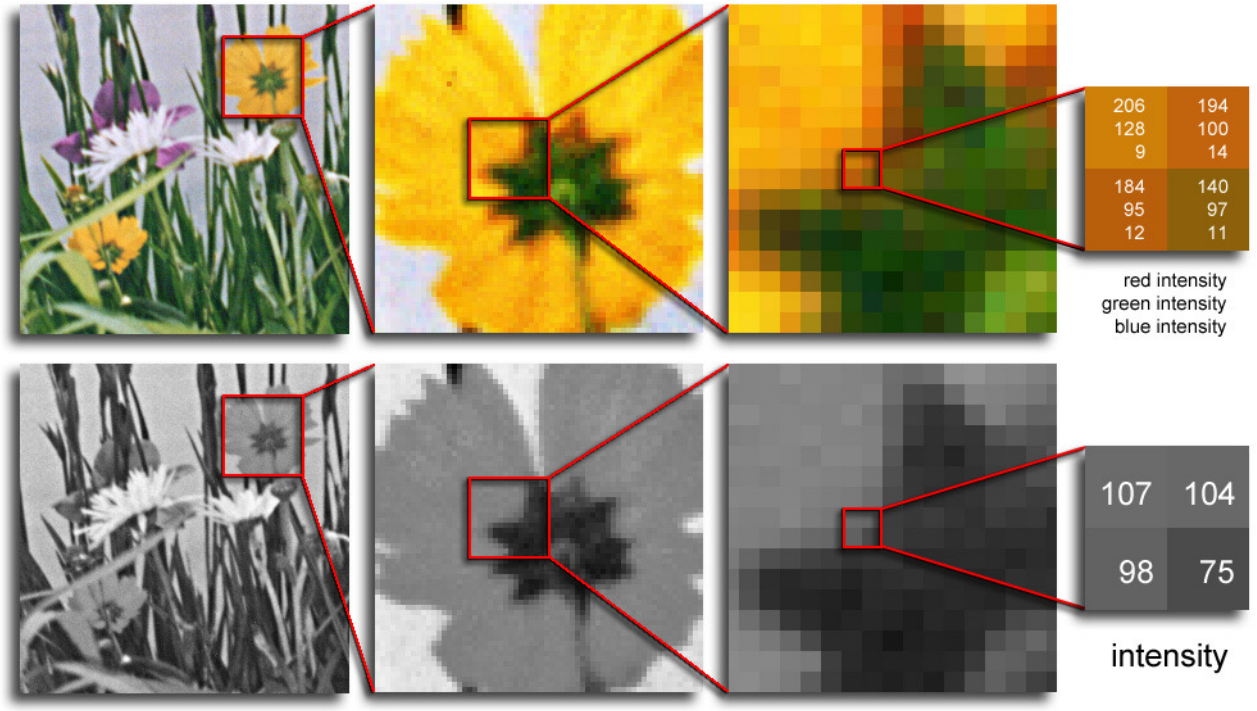
A colourful image

Colour Image

- Colour images have 3 values per pixel;
- Greyscale images have 1 value per pixel.

a grid of squares, each of which contains a single colour

each square is called a pixel (for *picture element*)

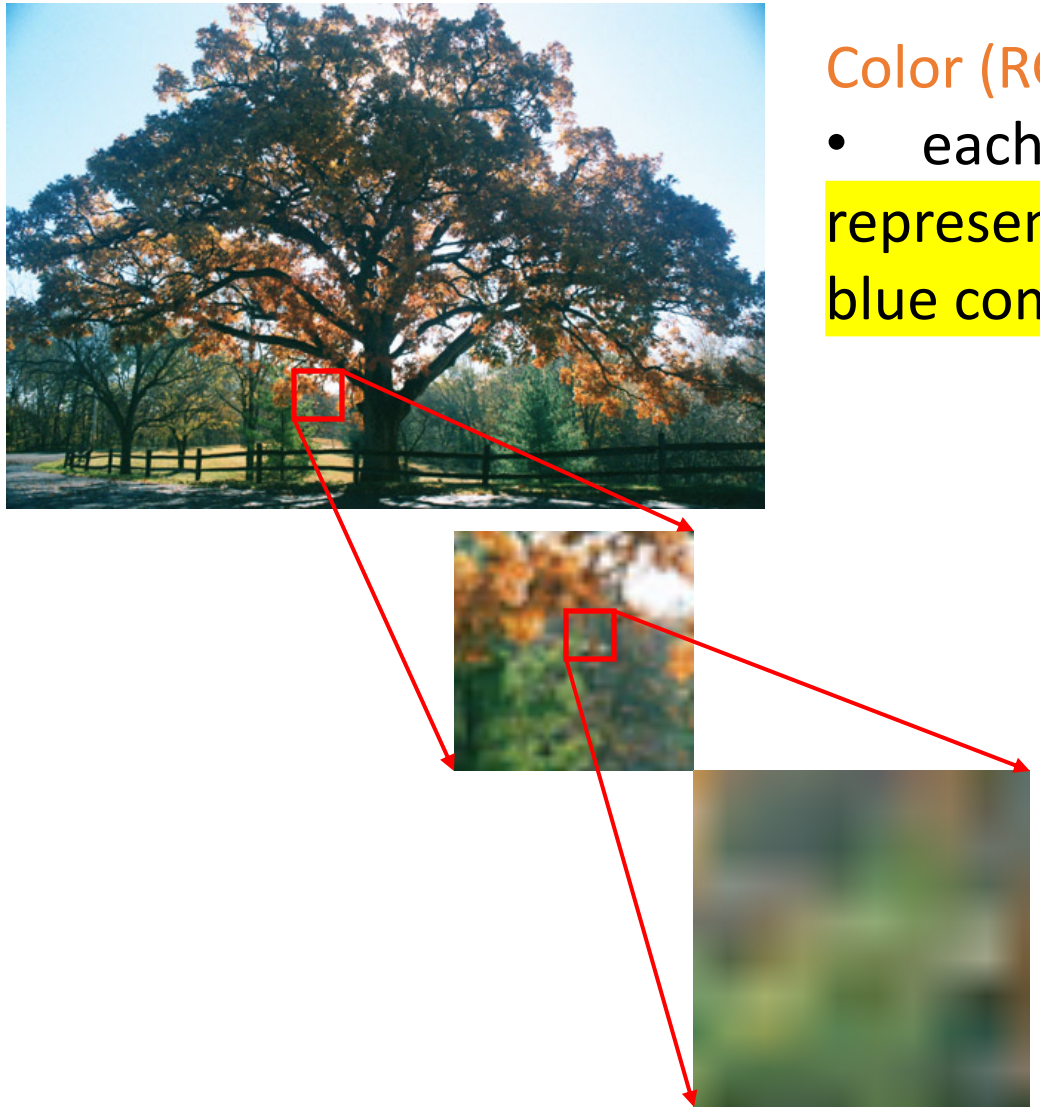


RGB

Grayscale

91

Colour Image



Color (RGB) image:

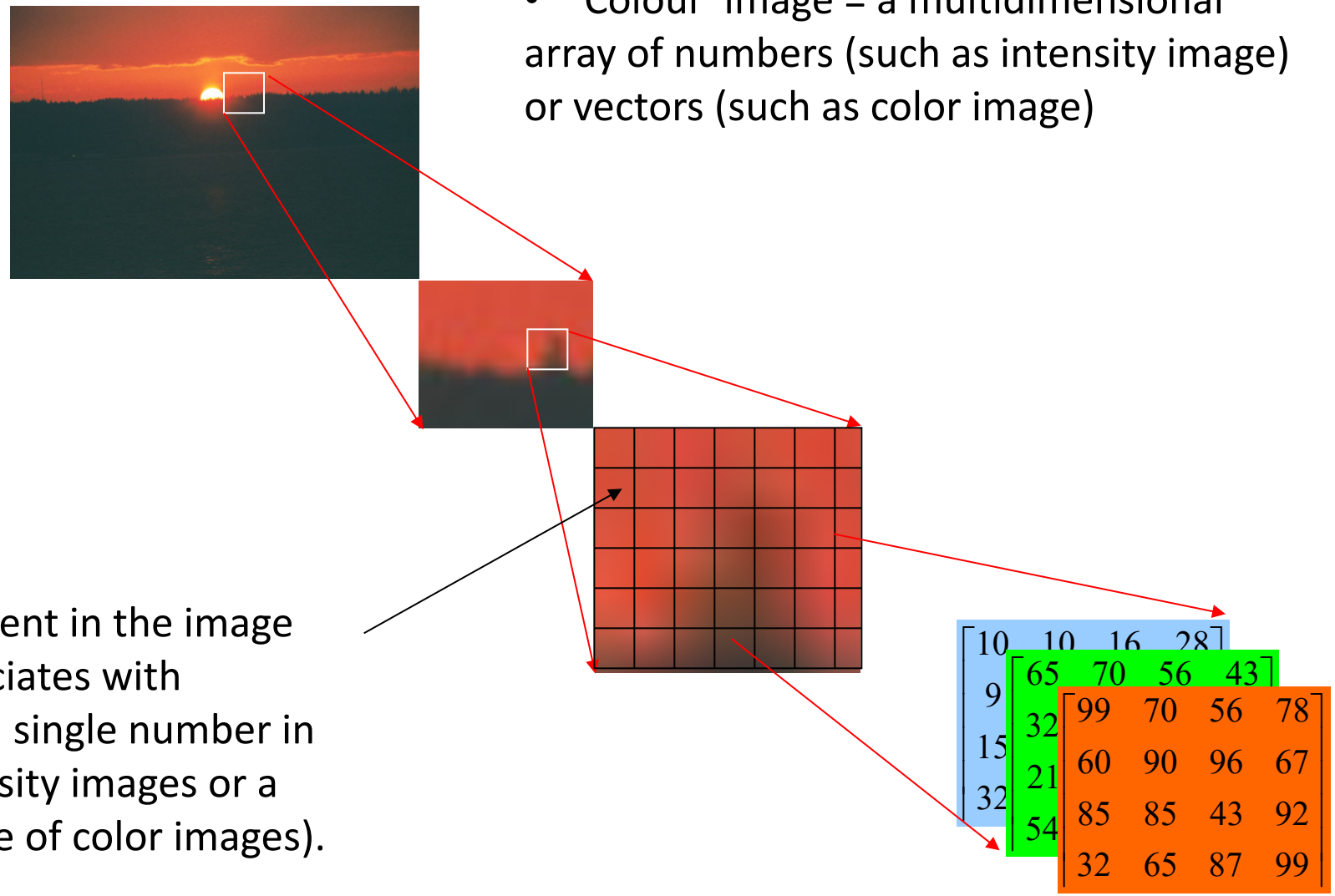
- each pixel contains a vector representing red, green and blue components.

RGB components

$$\begin{bmatrix} 10 & 10 & 16 & 28 \\ 9 & 65 & 70 & 56 & 43 \\ 15 & 32 & 99 & 70 & 56 & 78 \\ 32 & 21 & 60 & 90 & 96 & 67 \\ & 54 & 85 & 85 & 43 & 92 \\ & & 32 & 65 & 87 & 99 \end{bmatrix}$$

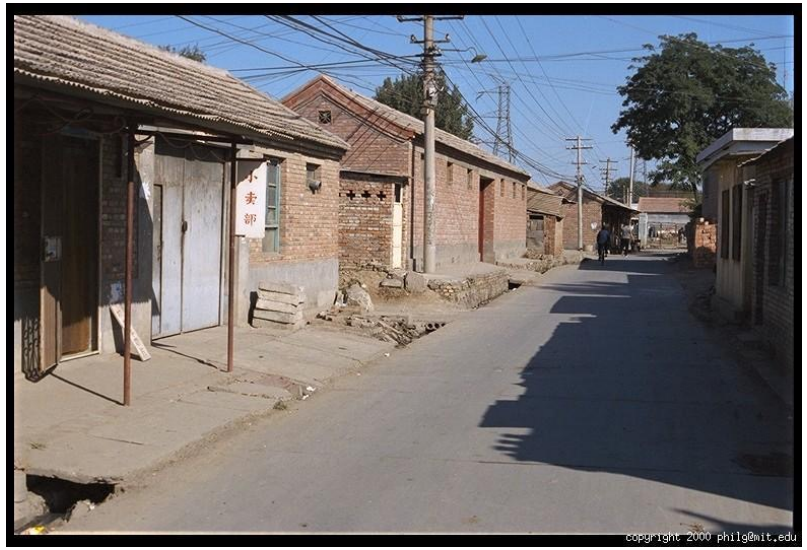
Colour Image

- Colour image = a multidimensional array of numbers (such as intensity image) or vectors (such as color image)



- Each component in the image called pixel associates with the pixel value (a single number in the case of intensity images or a vector in the case of color images).

Colour Image



Grayscale images, where each R, G, B value is converted to a grayscale value

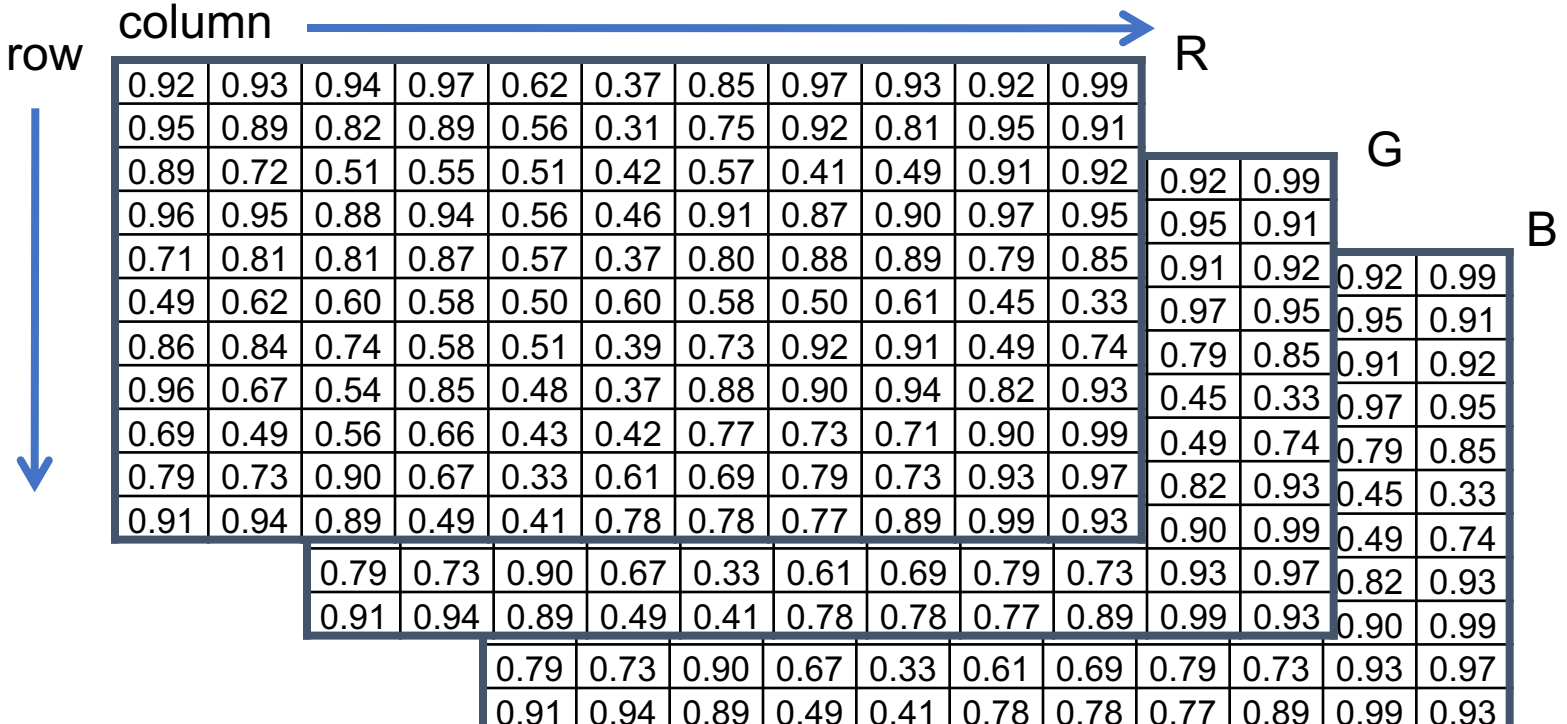
R

G

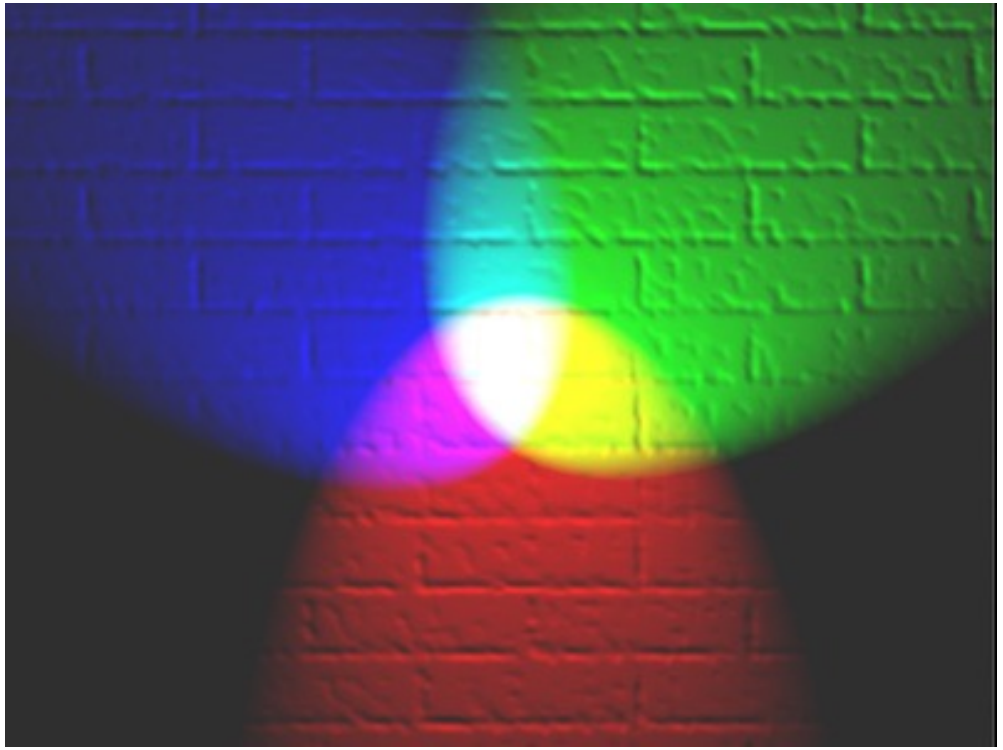
B

Colour Images in Matlab

- Images represented as a matrix
- Suppose we have a NxM RGB image called “im”
  - im(0,0,0) = top-left pixel value in R-channel
  - im(y, x, b) = y pixels down, x pixels to right in the bth channel
- cv2.imread(filename) returns a uint8 image (values 0 to 255)
  - **Convert to double format (values 0 to 1) (important !)**



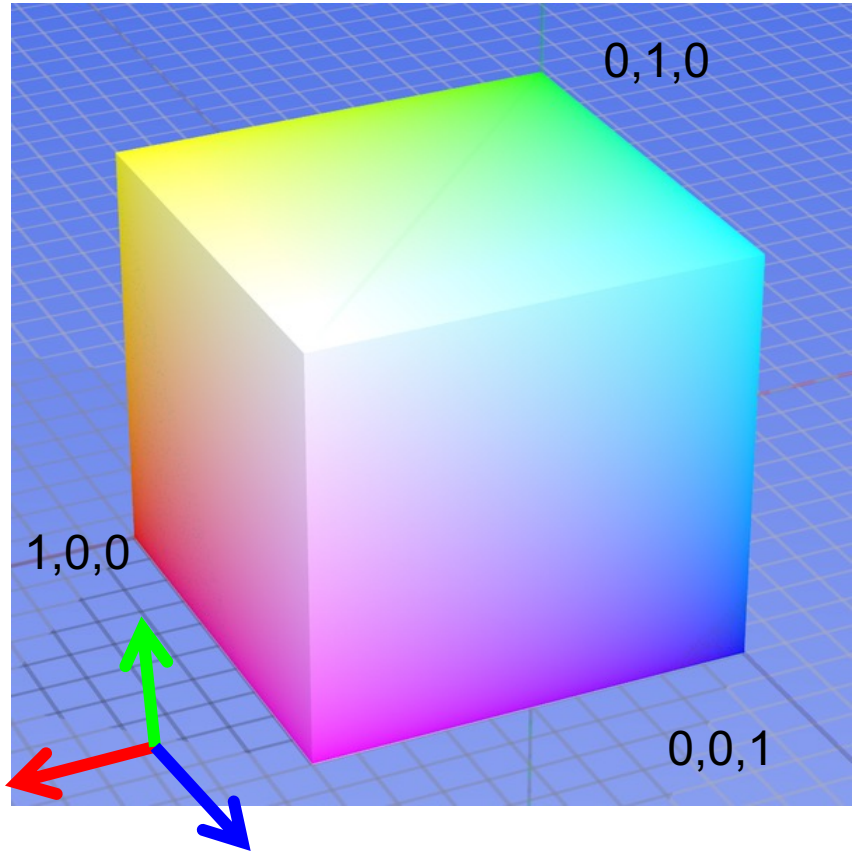
Color spaces



http://en.wikipedia.org/wiki/File:RGB_illumination.jpg

RGB space

Default color space



R
(G=0,B=0)

G
(R=0,B=0)

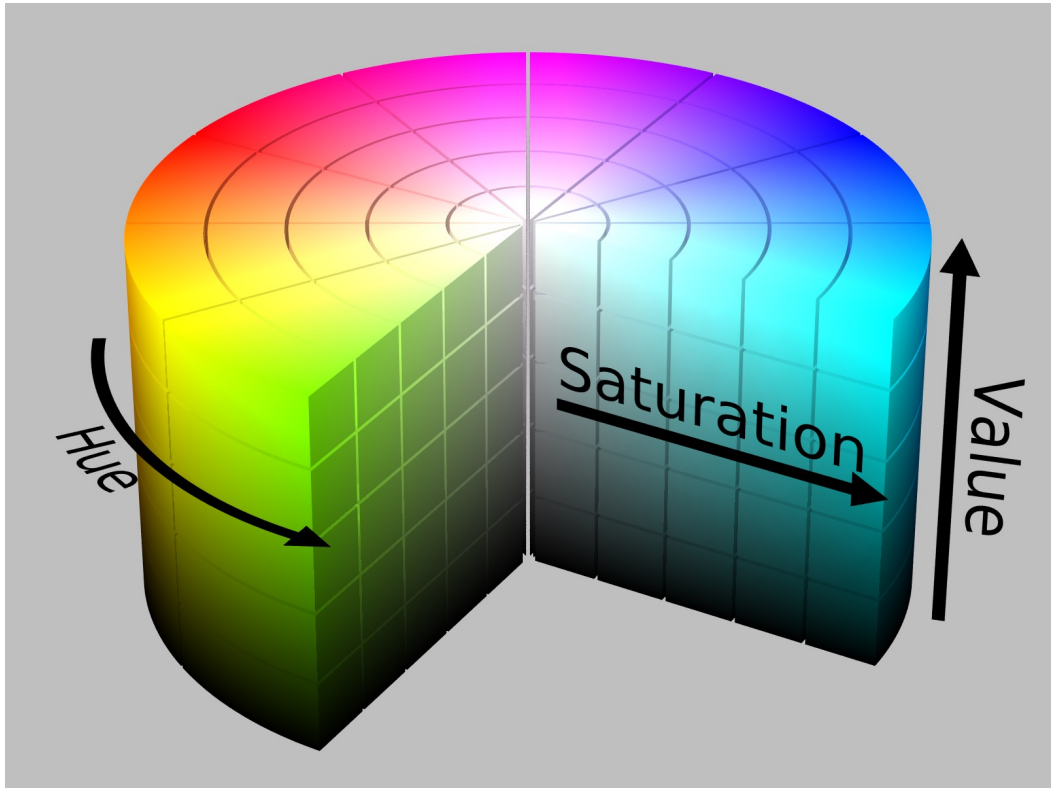
B
(R=0,G=0)

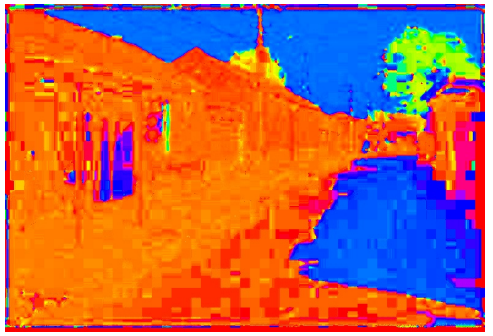
Some drawbacks

- Strongly correlated channels
- Not perceptually meaningful.

Image from: http://en.wikipedia.org/wiki/File:RGB_color_solid_cu

HSV space





H
(S=1,V=1)

S
(H=1,V=1)

V
(H=1,S=0)

Intuitive color space

Subjective terms to describe color

Hue
Name of the color
(yellow, red, blue, green, . . .)

Value/Lightness/Brightness
How light or dark a color is.

Saturation/Chroma/Color Purity
How "strong" or "pure" a color is.

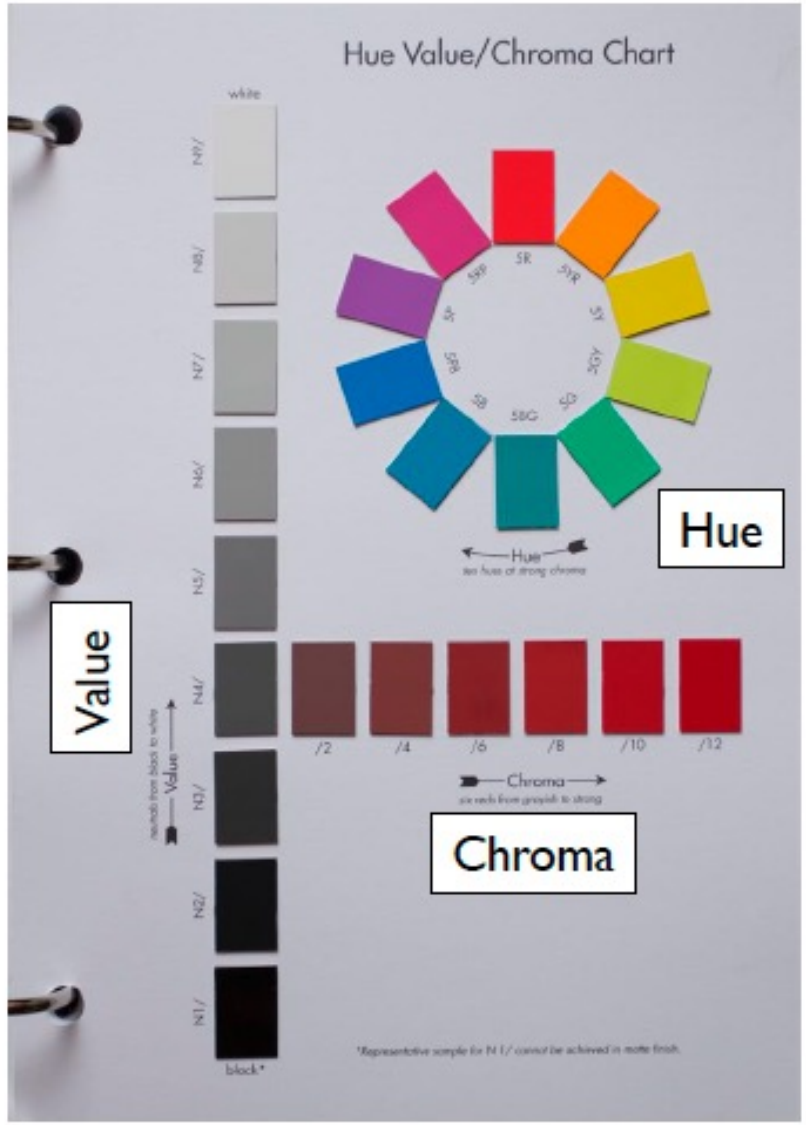


Image from Benjamin Salley
A page from a Munsell Student Color Set

HSV space

$$X_{max} := \max(R, G, B) =: V$$

$$X_{min} := \min(R, G, B)$$

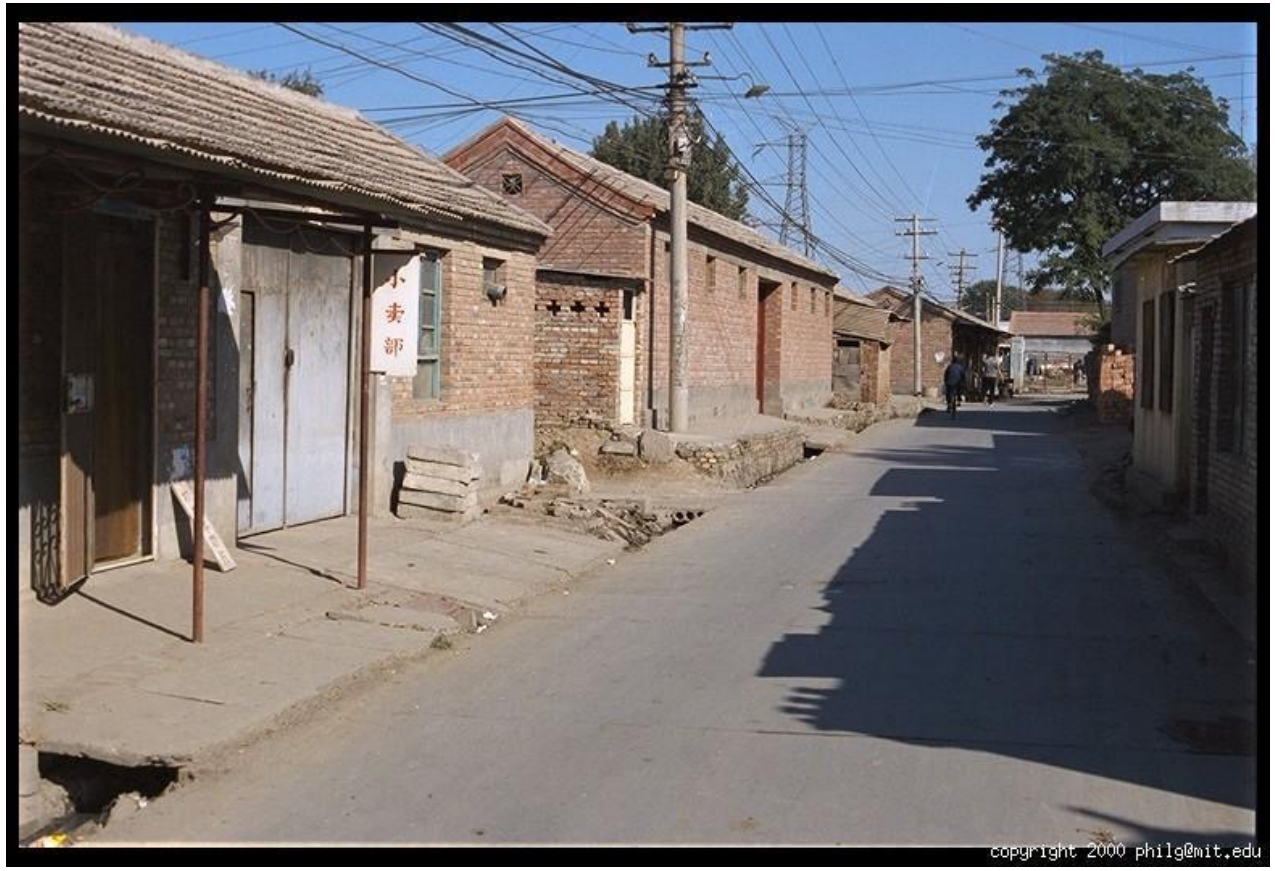
$$C := X_{max} - X_{min}$$

$$H := \begin{cases} 0, & \text{if } C = 0 \\ 60^\circ \cdot \left(0 + \frac{G-B}{C}\right), & \text{if } V = R \\ 60^\circ \cdot \left(2 + \frac{B-R}{C}\right), & \text{if } V = G \\ 60^\circ \cdot \left(4 + \frac{R-G}{C}\right), & \text{if } V = B \end{cases}$$

$$S_V := \begin{cases} 0, & \text{if } V = 0 \\ \frac{C}{V}, & \text{otherwise} \end{cases}$$

From wikipedia. : https://en.wikipedia.org/wiki/HSL_and_HSV

Most semantic information is contained in the intensity band



Original image

Most semantic information is contained in intensity channel

Summary

- Geometric image formation describes where the 3D objects are projected in the image. (location)

- Photometric image formation describes the appearance of the objects. (intensity, color, appearance.)

Readings

- Computer Vision: Algorithms and Applications
- 2nd Edition, Chapter-2 (2.1.4- perspective, camera intrinsics, camera matrix, 2.3.2 color)
- Digital image processing, Chapter 2.4.2, 2.4.3

•Take home message:

- Have a general idea of geometric and photometric image formation process.
- Understand the color image representation.