Computer Vision – Lecture 1

Introduction

19.10.2016

Bastian Leibe

Visual Computing Institute
RWTH Aachen University
http://www.vision.rwth-aachen.de/

leibe@vision.rwth-aachen.de

RWTH AACHEN UNIVERSITY

Organization

- Lecturer
  - Prof. Bastian Leibe (leibe@vision.rwth-aachen.de)
- Teaching Assistant
  - Stefan Breuers (breuers@vision.rwth-aachen.de)
- Course webpage
  - http://www.vision.rwth-aachen.de/courses/
    → Computer Vision
  - Slides will be made available on the webpage
  - There is also an L2P electronic repository
- Please subscribe to the lecture on the Campus system!
  - Important to get email announcements and L2P access!

Language

- Official course language will be English
  - If at least one English-speaking student is present.
  - If not… you can choose.

- However…
  - Please tell me when I'm talking too fast or when I should repeat something in German for better understanding!
  - You may at any time ask questions in German!
  - You may turn in your exercises in German.
  - You may answer exam questions in German.

Organization

- Structure: 3V (lecture) + 1Ü (exercises)
  - 6 EECS credits
  - Part of the area “Applied Computer Science”

- Place & Time

Lecture:	Mon	10:15 - 11:45	UMIC 025
Lecture/Exercises:	Wed	10:15 - 11:45	UMIC 025

- Exam
  - Written exam
  - Dates will be communicated soon

Exercises and Demos

- Exercises
  - Typically 1 exercise sheet every 2 weeks (Matlab based)
  - Hands-on experience with the algorithms from the lecture.
  - Send in your solutions the night before the exercise class.
  - No admission requirement to qualify for the exam this year!

- Teams are encouraged!
  - You can form teams of up to 3 people for the exercises.
  - Each team should only turn in one solution.
  - But list the names of all team members in the submission.

Course Webpage

Course Schedule

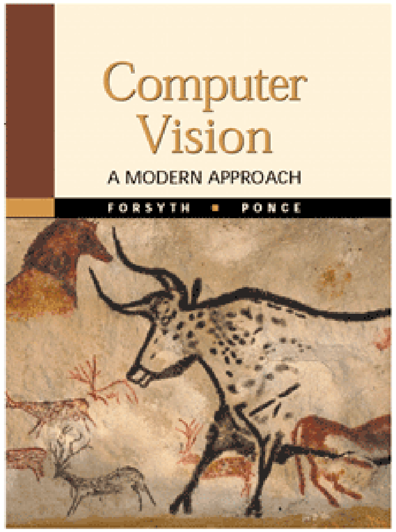
Date	Title	Content	Material
Wed, 2016-10-19	Introduction	Why vision? Applications, Challenges, Image Formation	
Mon, 2016-10-24	Exercise 1	Intro Matlab	
Wed, 2016-10-26	Image Processing I	Binary Images, Thresholding, Morphology, Connected Components, Region Descriptors	
Mon, 2016-10-31	Image Processing II	Linear Filters, Gaussian Smoothing, Median Filter	
Wed, 2016-11-02	Edge Detection	Multi-scale Representations, Image Derivatives, Edge Detection	
Mon, 2016-11-07	Structure Extraction	Chamfer Matching, Line Fitting, Hough Transform, Gen. Hough Transform	
Wed, 2016-11-09	Segmentation I	Segmentation as Clustering, k-means, EM, Mean-Shift	
Mon, 2016-11-14	Exercise 2	Thresholding, Morphology, Derivatives, Edges	
Wed, 2016-11-16	Segmentation II	Segmentation as Energy Minimization, (Markov Random Fields, Graph Cuts)	
Mon, 2016-11-21	Recognition I	Global Descriptors, Histograms, Histogram Comparison, Multidim. Histograms	
Wed, 2016-11-23	Categorization I	Sliding Window-based Object Detection, SVM, HOG	

Monday: Matlab tutorial

http://www.vision.rwth-aachen.de/courses/

Textbooks

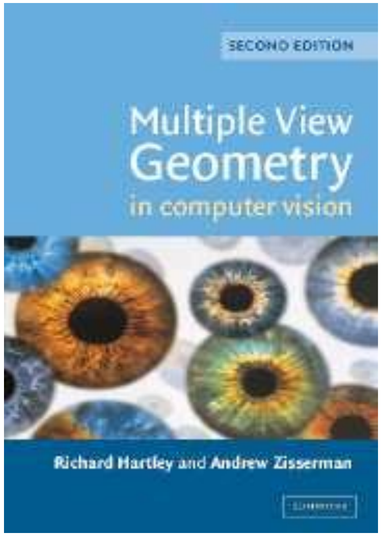
- No single textbook for the class.
- Basic material is covered in the following two books.



D. Forsyth, J. Ponce
Computer Vision - A Modern Approach
Prentice Hall, 2002

(available in the library's "Handapparat")

R. Hartley, A. Zisserman
Multiple View Geometry in Computer Vision
2nd Ed., Cambridge Univ. Press, 2004



- Additional material will be given out for some topics.
  - Tutorials and deeper introductions.
  - Application papers

How to Find Us

- **Office:**
  - UMIC Research Centre
  - Mies-van-der-Rohe-Strasse 15, room 124

- **Office hours**
  - If you have questions to the lecture, come to us.
  - My regular office hours will be announced (additional slots are available upon request)
  - Send us an email before to confirm a time slot.

Questions are welcome!

Topics of Today's Lecture

- **What is computer vision?**
- **What does it mean to see and how do we do it?**
- **How can we make this computational?**

- **First Topic: Image Formation**
  - Details in Forsyth & Ponce, chapter 1.

Why Computer Vision?






Cameras are all around us…





Images and video are everywhere…

Personal photo albums

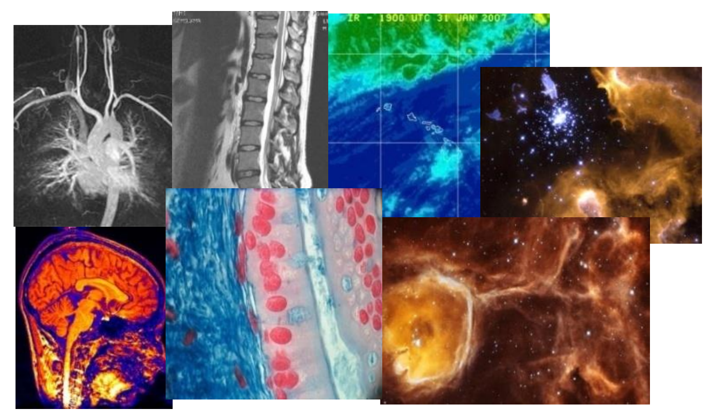
Movies, news, sports

Google Image Search · Picasa · flickr · webshots · picsearch · YouTube Broadcast Yourself · kooaba

Internet services

Surveillance and security

Mobile and consumer applications



Medical and scientific images

What is Computer Vision?

- Goal of Computer Vision
  - *Enable a machine to "understand" images and videos*

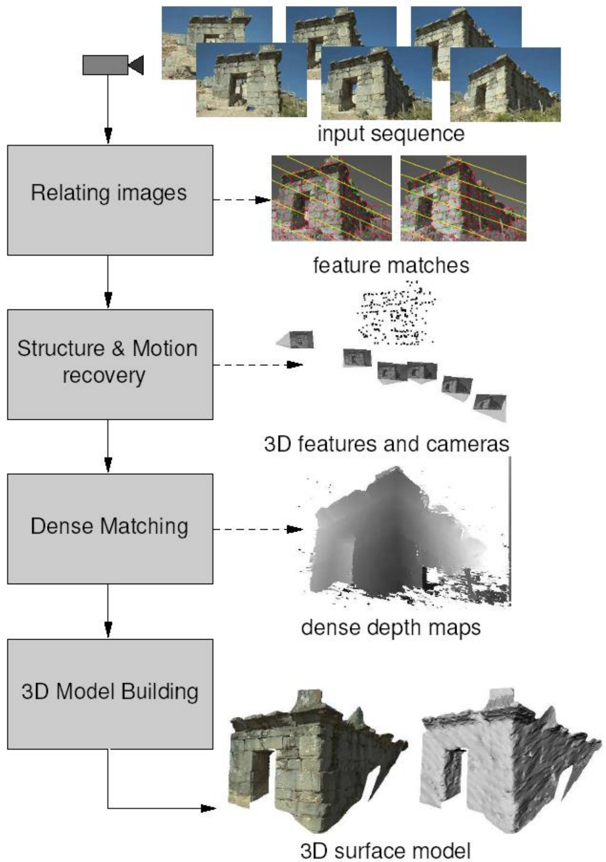
- Automatic understanding
  - Computing properties of the 3D world from visual data *(measurement)*
  - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*

Vision for Measurement

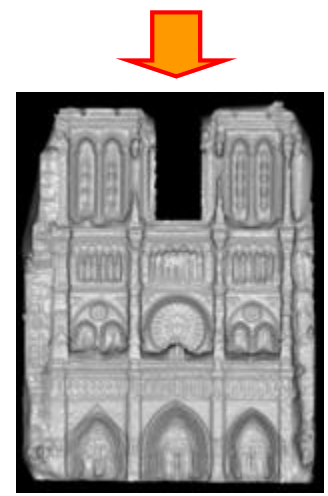
Real-time stereo

Pollefeys et al.

Structure from motion



Multi-view stereo for community photo collections



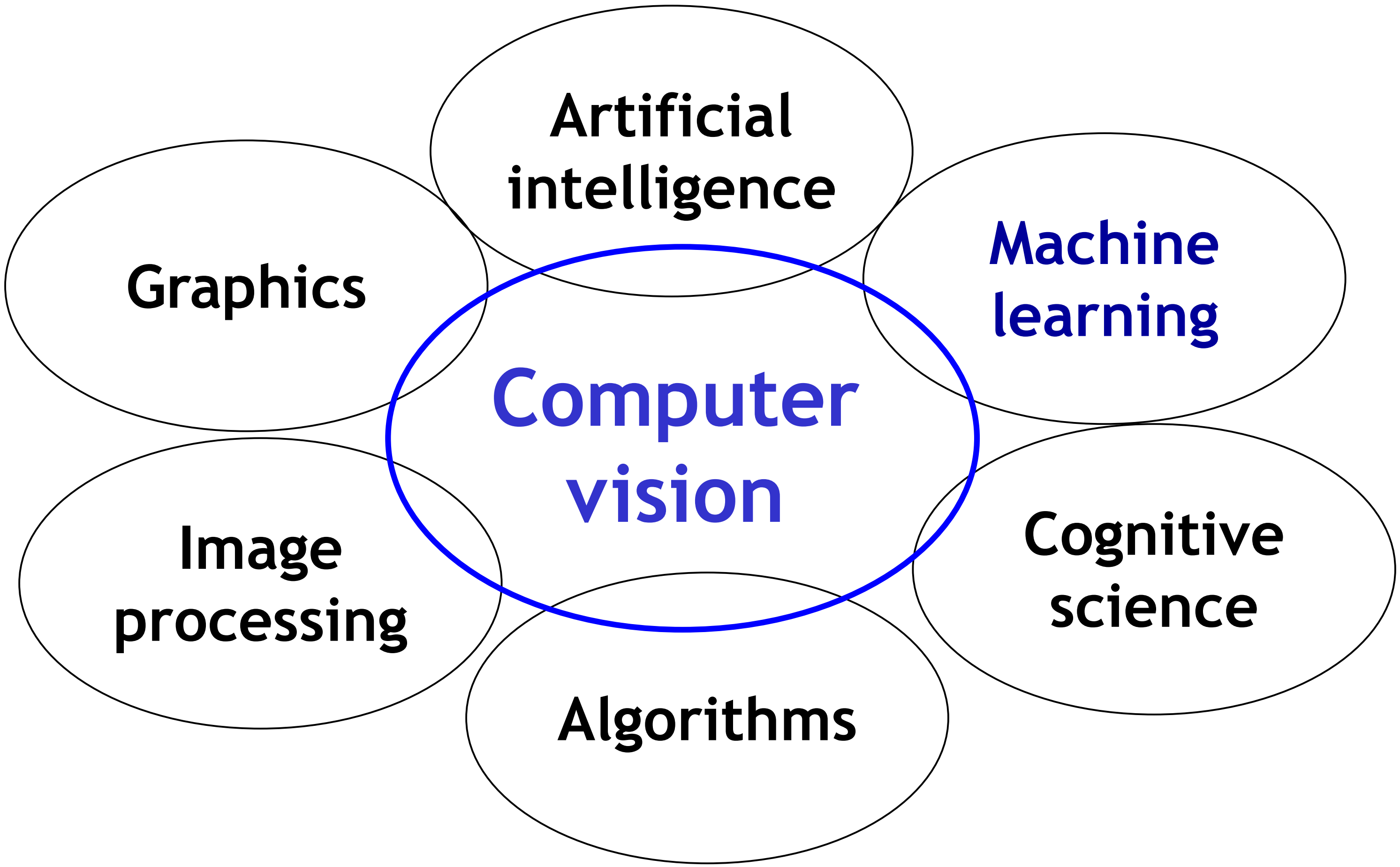
Goesele et al.

Vision for Perception, Interpretation

Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions…

Slide credit: Kristen Grauman

Related Disciplines



Directions to Computer Vision

- Science
  - Foundations of perception. How do WE see?
- Engineering
  - How do we build systems that perceive the world?
- Many applications
  - Medical imaging, surveillance, entertainment, graphics, …

Applications: Faces and Digital Cameras

Setting camera focus via face detection

Camera waits for everyone to smile to take a photo [Canon]

Automatic lighting correction based on face detection

Segmentation

- Automatic background removal from images
  - Functionality is included in Microsoft Office 2010…

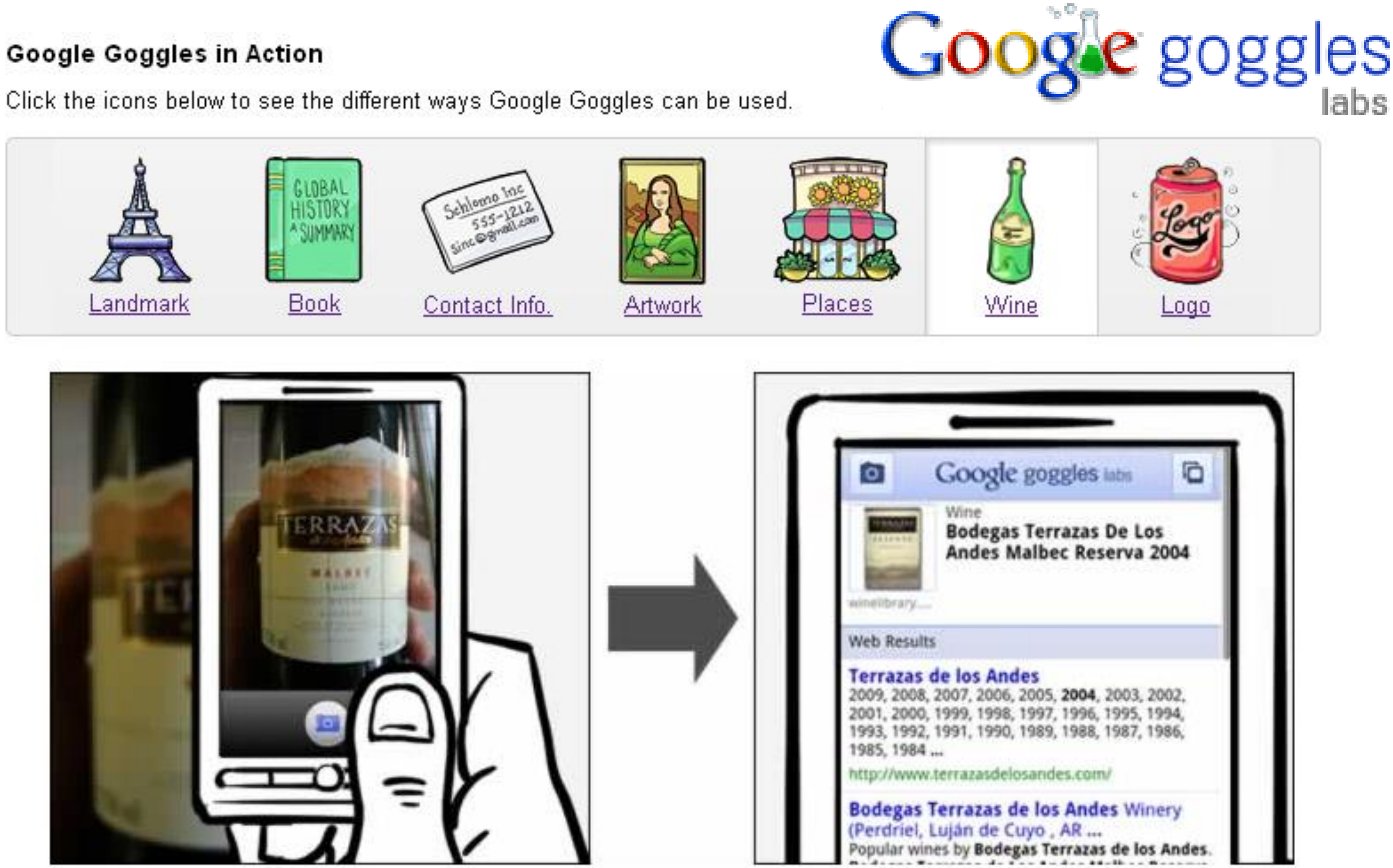
Matching

- **Stitch your photos together to create panoramas**

AUTOSTITCH

Available on the iPhone App Store

Applications: Vision for Mobile Phones



- **Take photos of objects as queries for visual search**

Slide credit: Svetlana Lazebnik

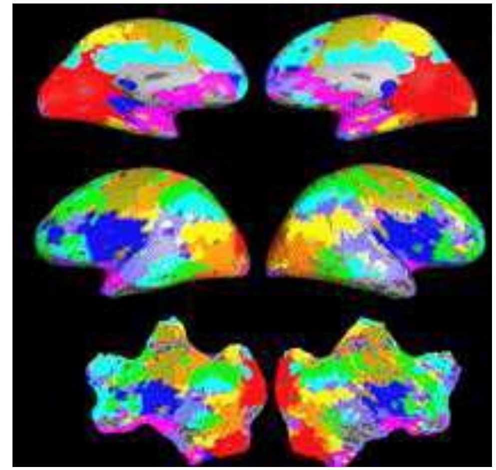
Applications: Vision-based Interfaces

Games
(Microsoft Kinect)

Assistive technology systems
Camera Mouse
Boston College

Applications: Medical & Neuroimaging

Image guided surgery
MIT AI Vision Group



fMRI data
Golland et al.

Applications: Visual Special Effects


The Matrix

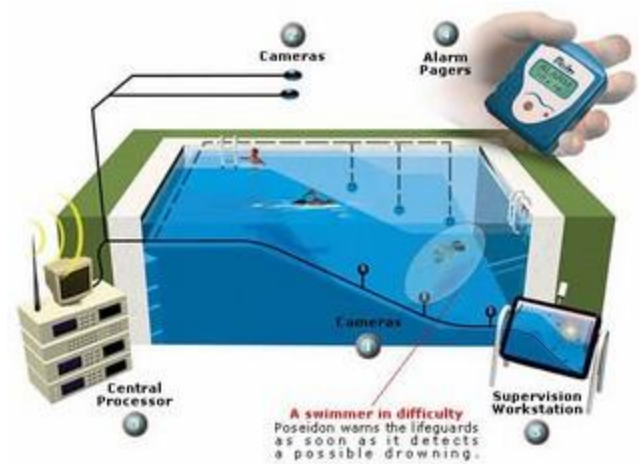

MoCap for *Pirates of the Carribean*, Industrial Light and Magic
(Source: S. Seitz)

Slide adapted from Svetlana Lazebnik, Kristen Grauman

Applications: Safety & Security

Autonomous robots

Driver assistance



Monitoring pools
(Poseidon)

Pedestrian detection
[MERL, Viola et al.]

Surveillance

Ok, Let’s Do It – Any Obstacles?

- **1966: Seymour Papert directs an undergraduate student to solve "the problem of computer vision" as a summer project.**

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

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

- **Obviously, computer vision was too difficult for that…**

Challenges: Many Nuisance Parameters

Illumination

Object pose

Clutter

Occlusions

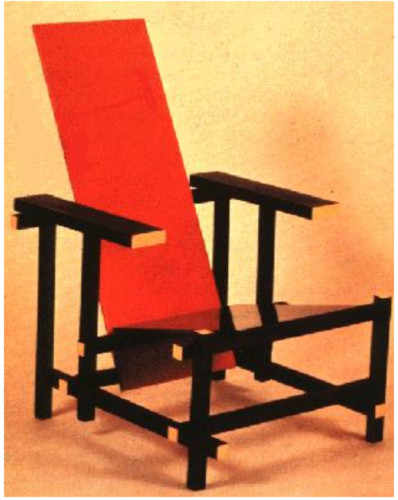
Intra-class appearance

Viewpoint

RWTH AACHEN UNIVERSITY

Challenges: Intra-Category Variation



Slide credit: Fergus, FeiFei, Torralba

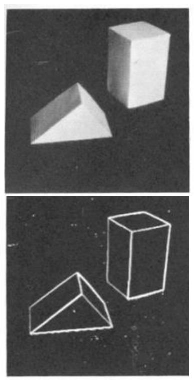
Challenges: Complexity

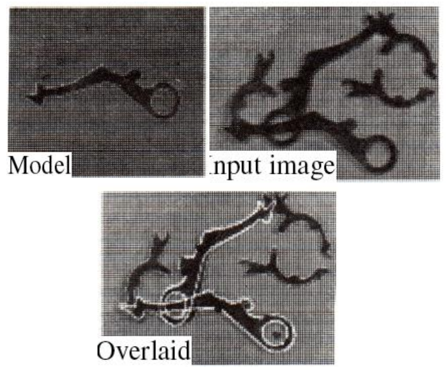
- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991].

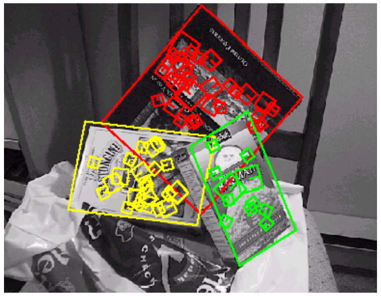
Slide credit: Kristen Grauman

So, Should We Give Up?

- NO! Very active research area with exciting progress!





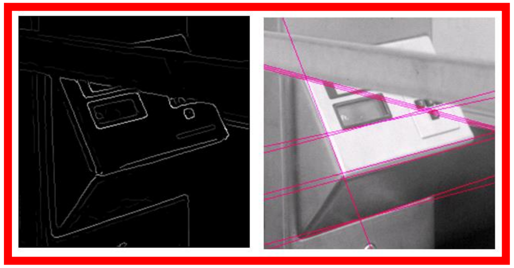


Things Are Starting to Work…

Computer Vision in realistic scenarios is becoming feasible!

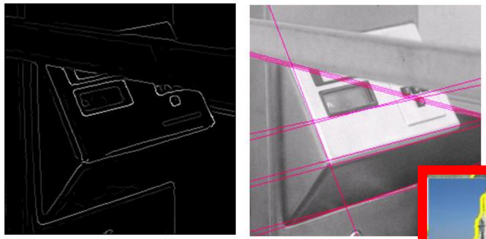
RWTH AACHEN UNIVERSITY

Course Outline



- **Image Processing Basics**
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Optical Flow

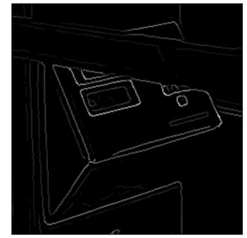
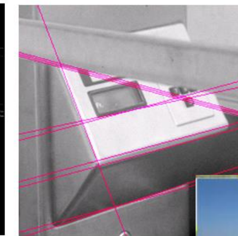
Course Outline

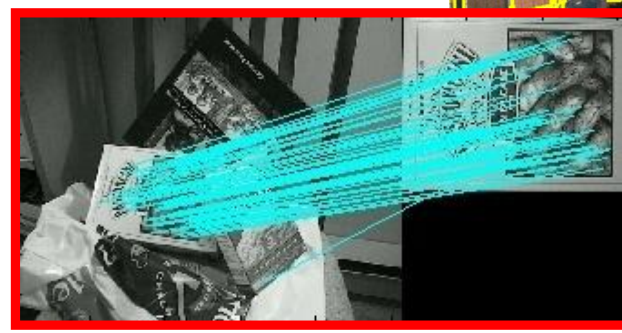


- Image Processing Basics
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Optical Flow

Course Outline

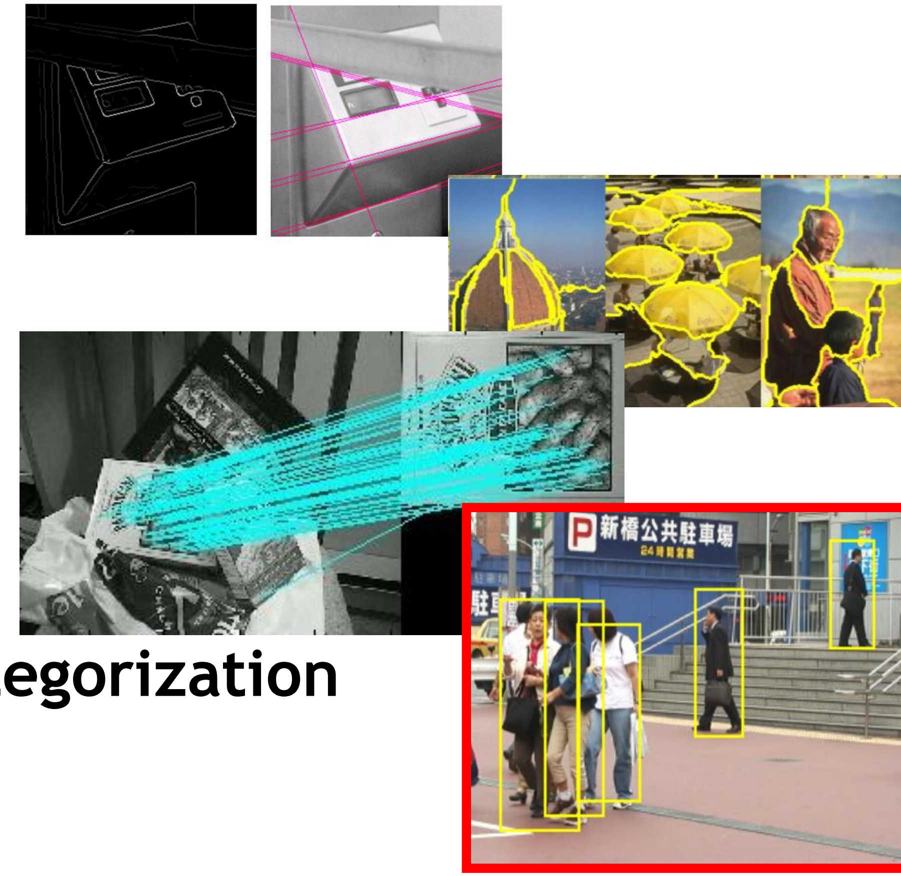
- Image Processing Basics
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Optical Flow



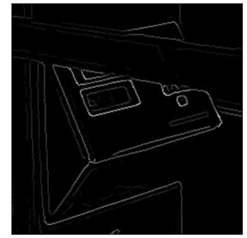
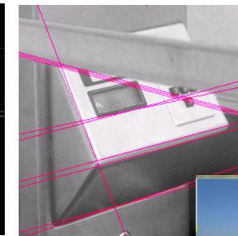
Course Outline

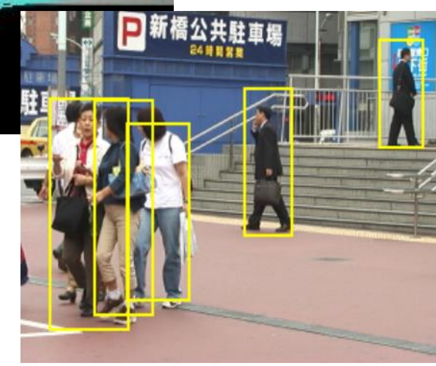
- Image Processing Basics
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Optical Flow



Course Outline

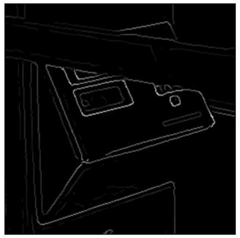
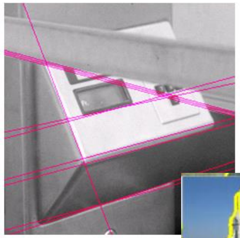
- Image Processing Basics
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Optical Flow

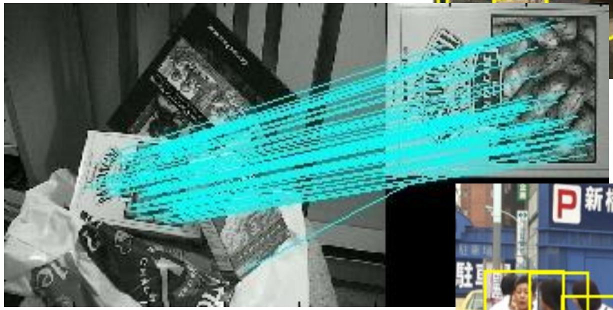
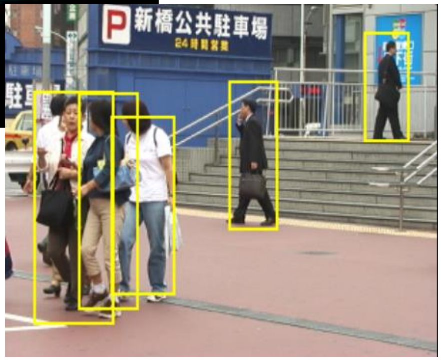



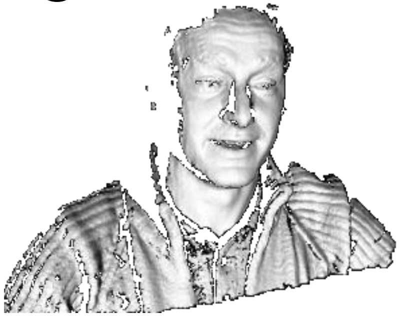



Course Outline

- Image Processing Basics
- Segmentation
- Local Features & Matching
- Object Recognition and Categorization
- 3D Reconstruction
- Motion and Optical Flow



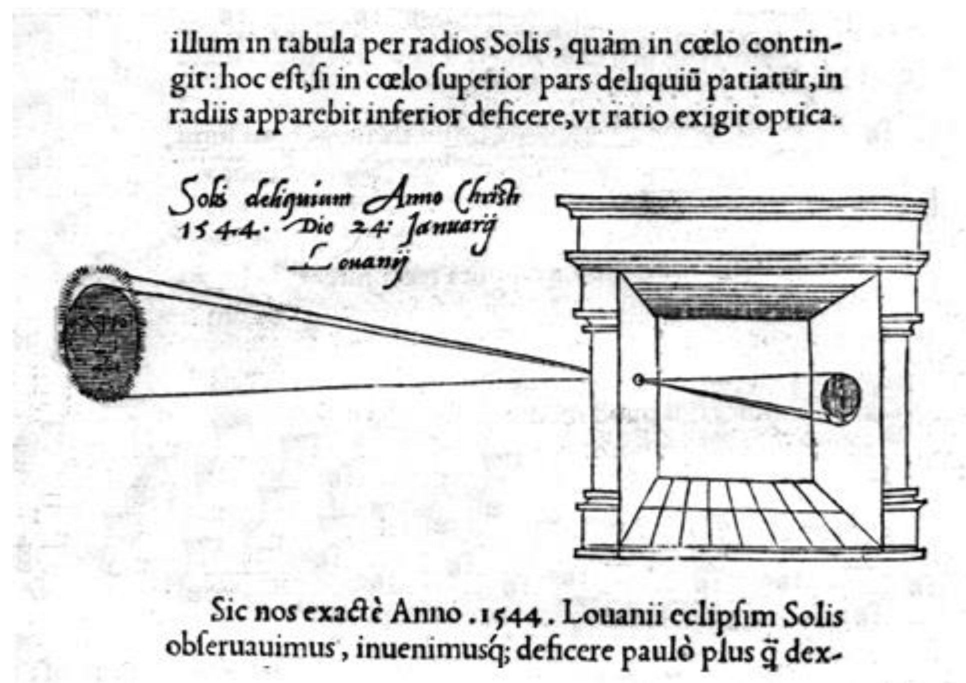

Topics of Today's Lecture

- What is computer vision?
- What does it mean to see and how do we do it?
- How can we make this computational?

- **First Topic: Image Formation**
  - Details in Forsyth & Ponce, chapter 1.

Camera Obscura

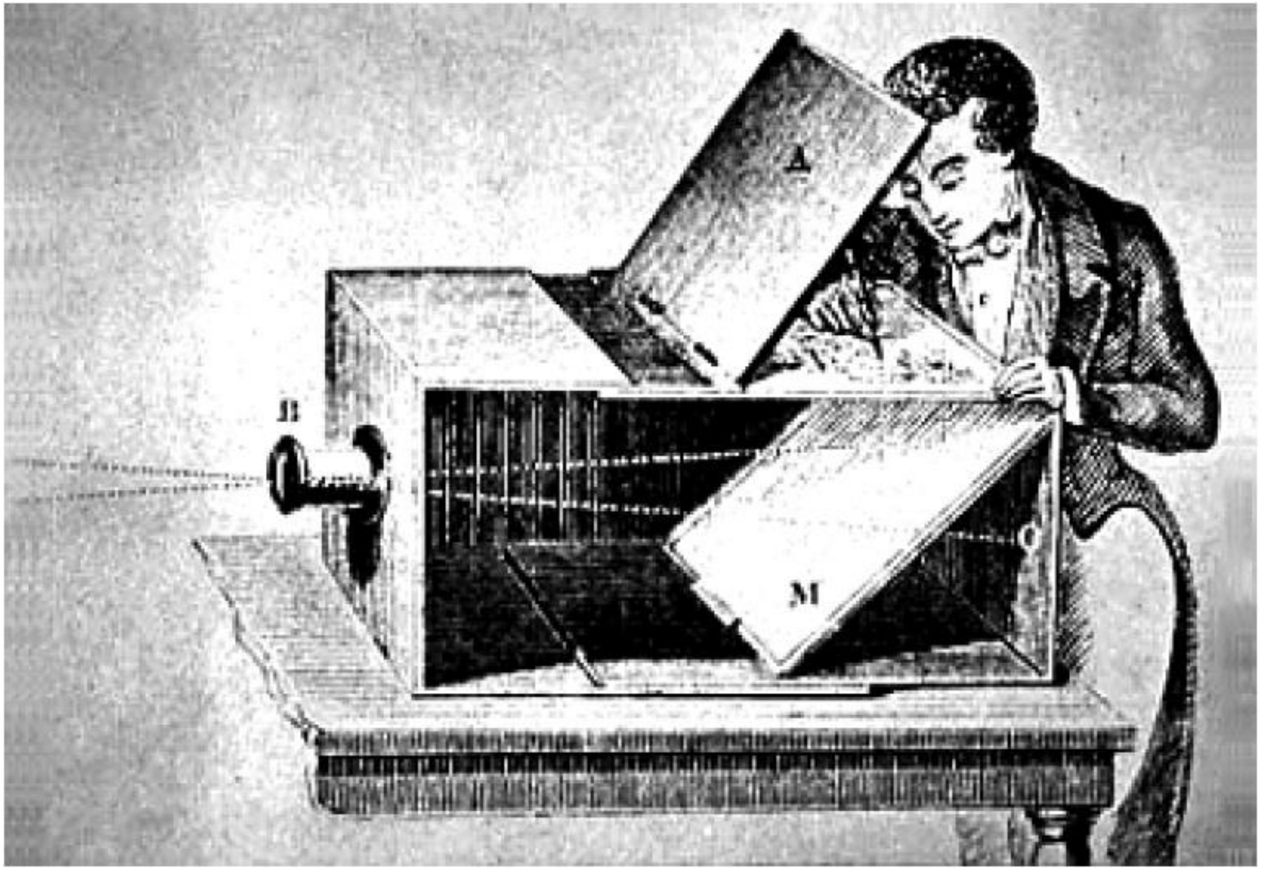
- **Around 1519, Leonardo da Vinci (1452 – 1519)**
  - **“When images of illuminated objects … penetrate through a small hole into a very dark room … you will see [on the opposite wall] these objects in their proper form and color, reduced in size … in a reversed position owing to the intersection of the rays”**



Slide credit: Bernt Schiele

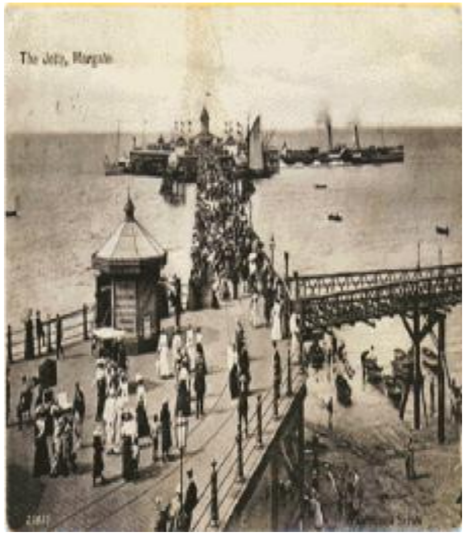
Source: http://www.acmi.net.au/AIC/CAMERA_OBSCURA.html

Camera Obscura



- **Used by artists (e.g. Vermeer 17th century) and scientists**

Camera Obscura



Jetty at Margate England, 1898.

An attraction in the late 19th century

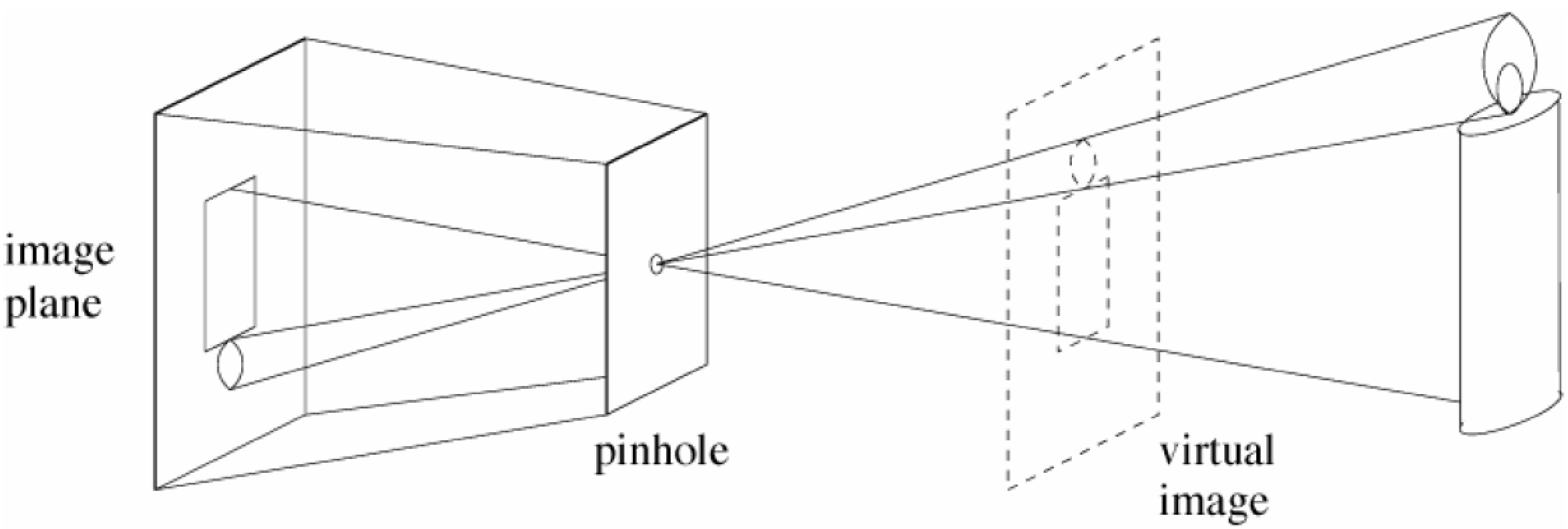
http://brightbytes.com/cosite/collection2.html

Adapted from R. Duraiswami

RWTH AACHEN UNIVERSITY

Pinhole Camera

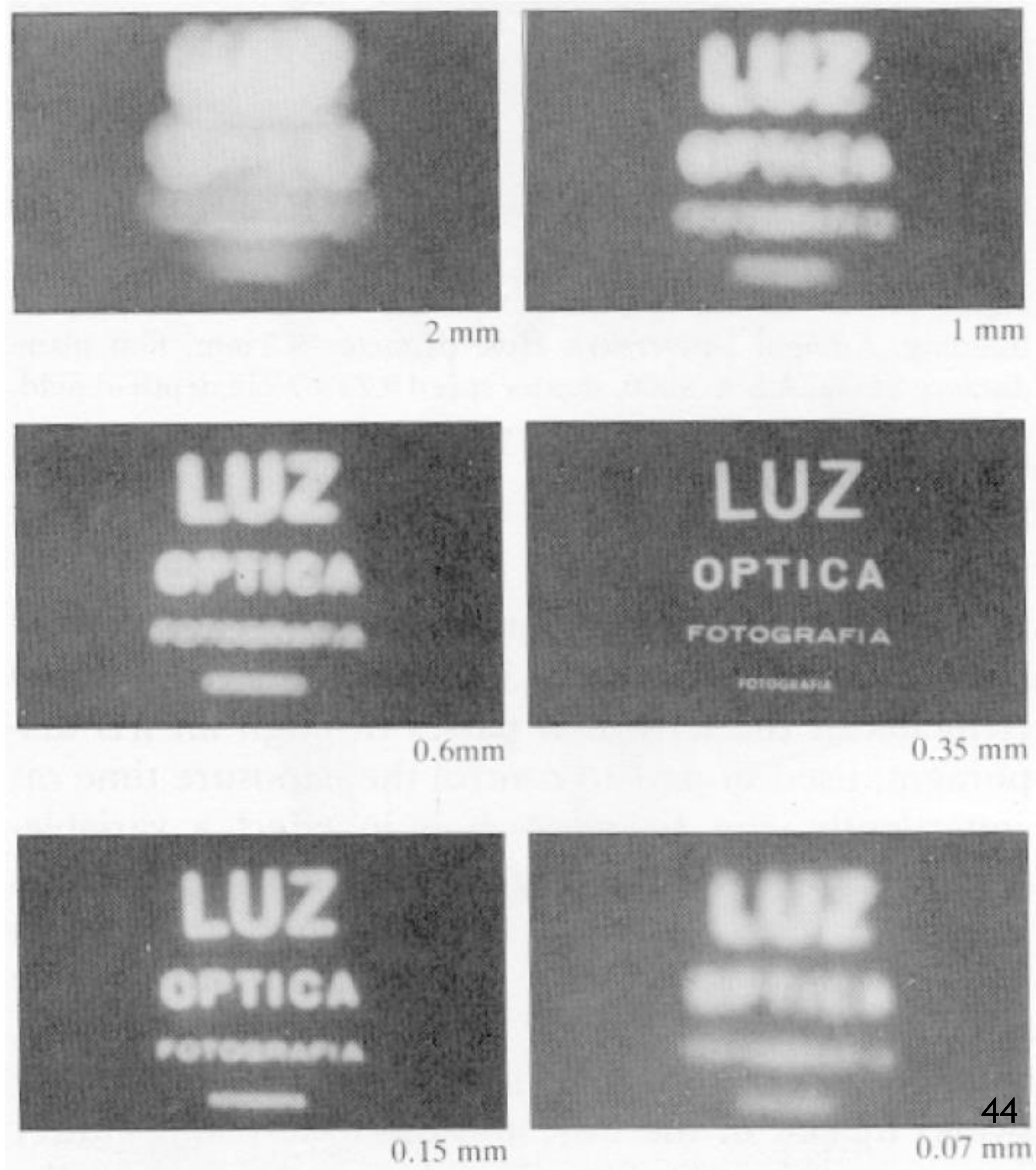
- (Simple) standard and abstract model today
  - Box with a small hole in it
  - Works in practice



Source: Forsyth & Ponce

Pinhole Size / Aperture

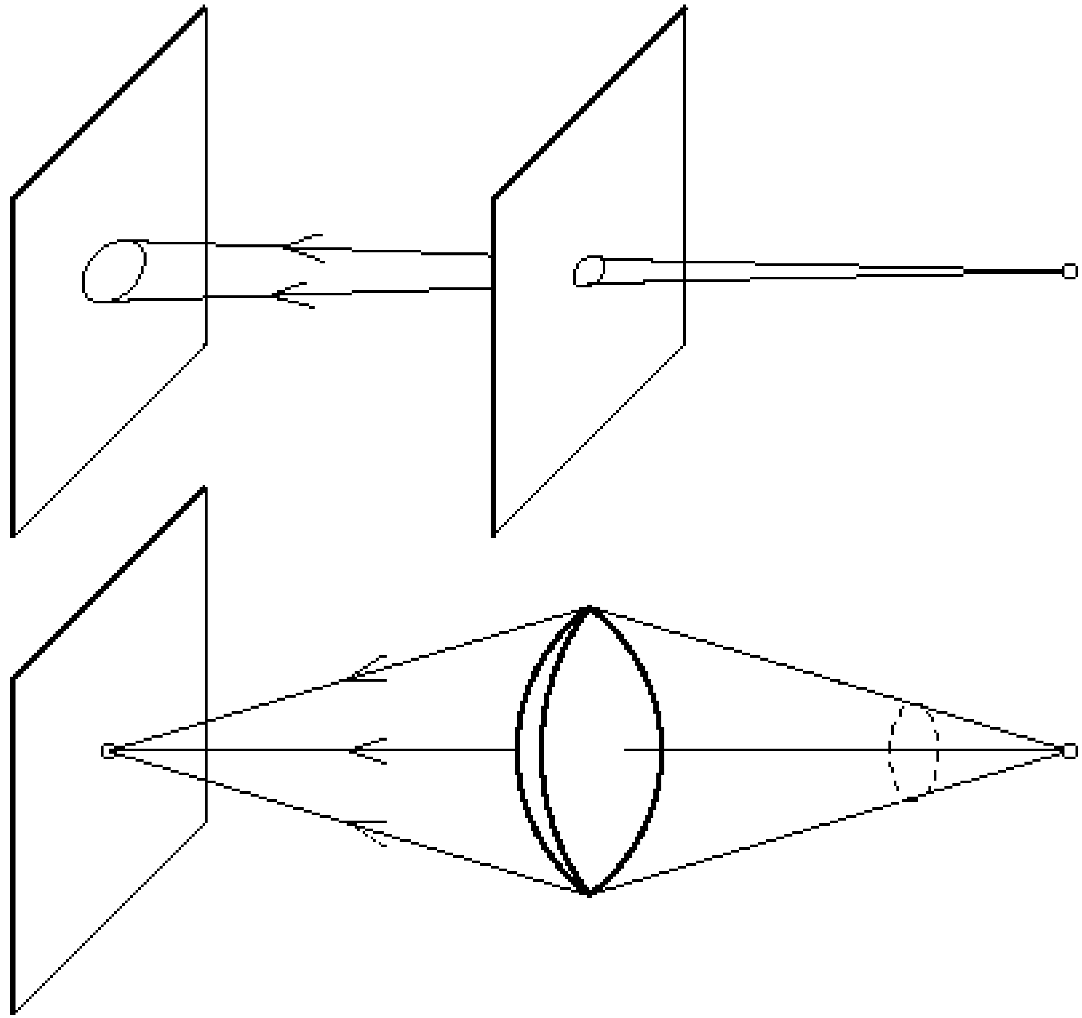
- Pinhole too big - many directions are averaged, blurring the image
- Pinhole too small - diffraction effects blur the image
- Generally, pinhole cameras are *dark*, because a very small set of rays from a particular point hits the screen.



Source: Forsyth & Ponce

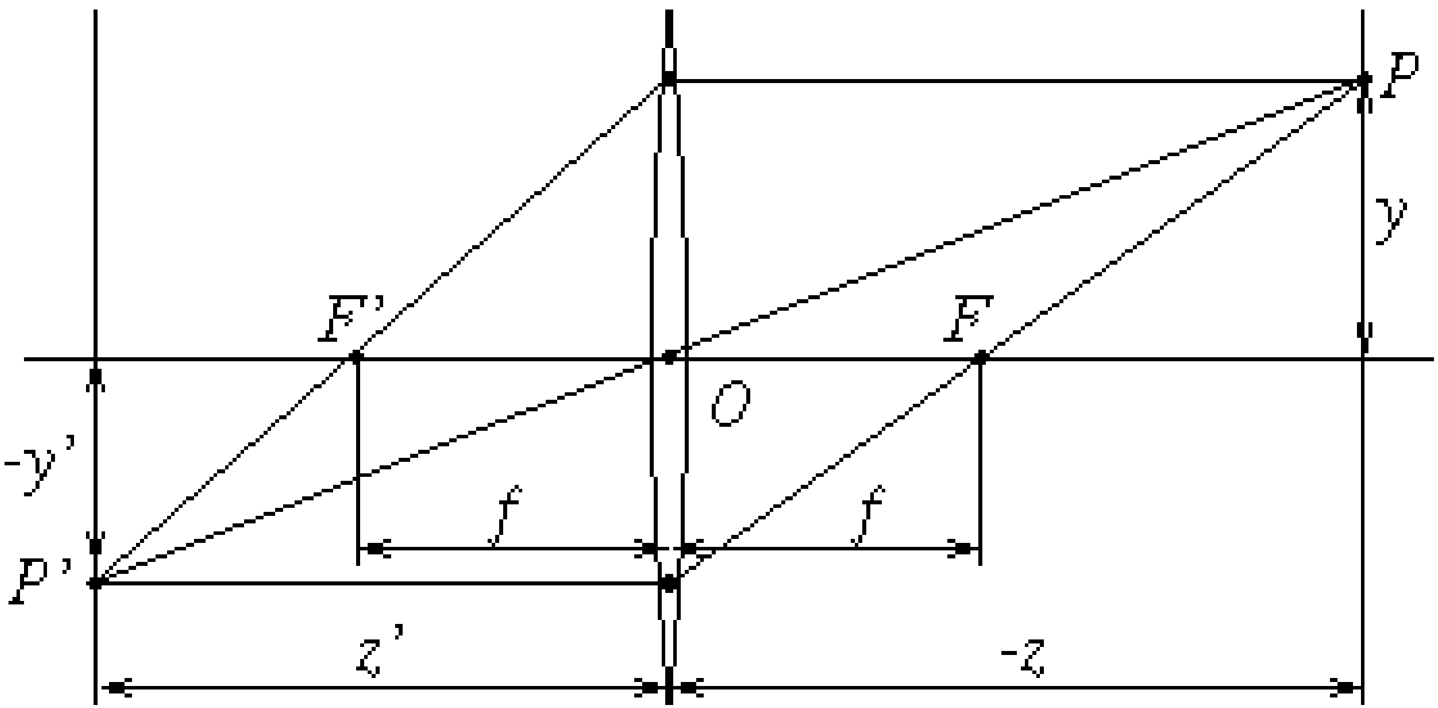
The Reason for Lenses

- **Keep the image in sharp focus while gathering light from a large area**



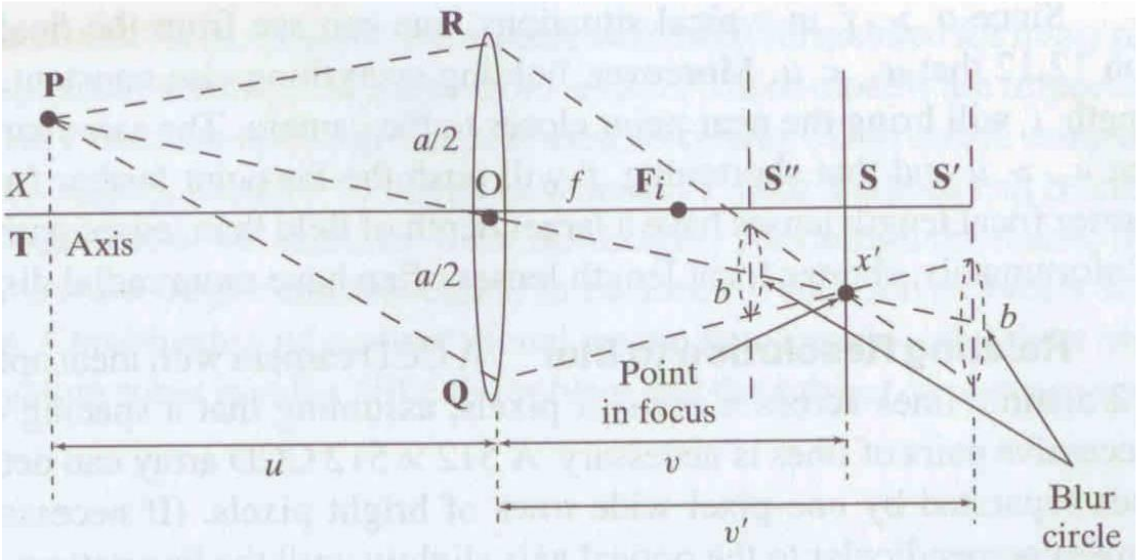
Source: Forsyth & Ponce

The Thin Lens



$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$

Focus and Depth of Field



“circles of confusion”

Thin lens: scene points at distinct depths come in focus at different image planes.

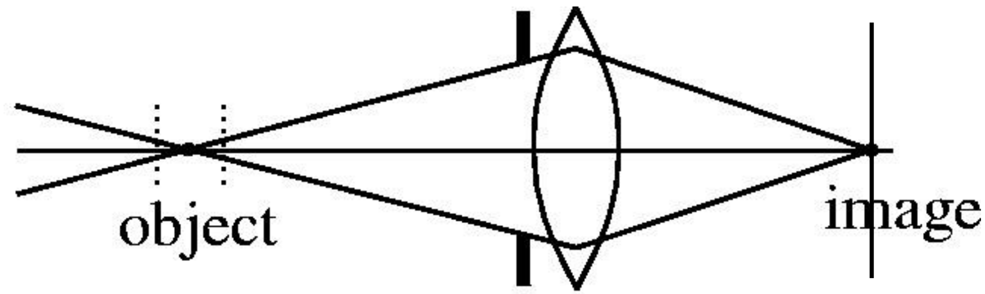
(Real camera lens systems have greater depth of field.)

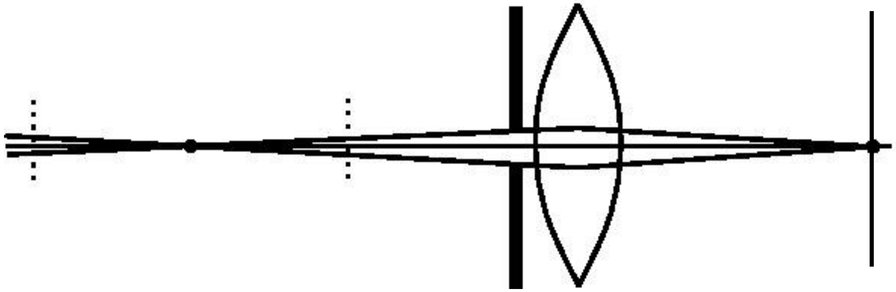
- Depth of field: distance between image planes where blur is tolerable

Source: Shapiro & Stockman

Focus and Depth of Field

- How does the aperture affect the depth of field?



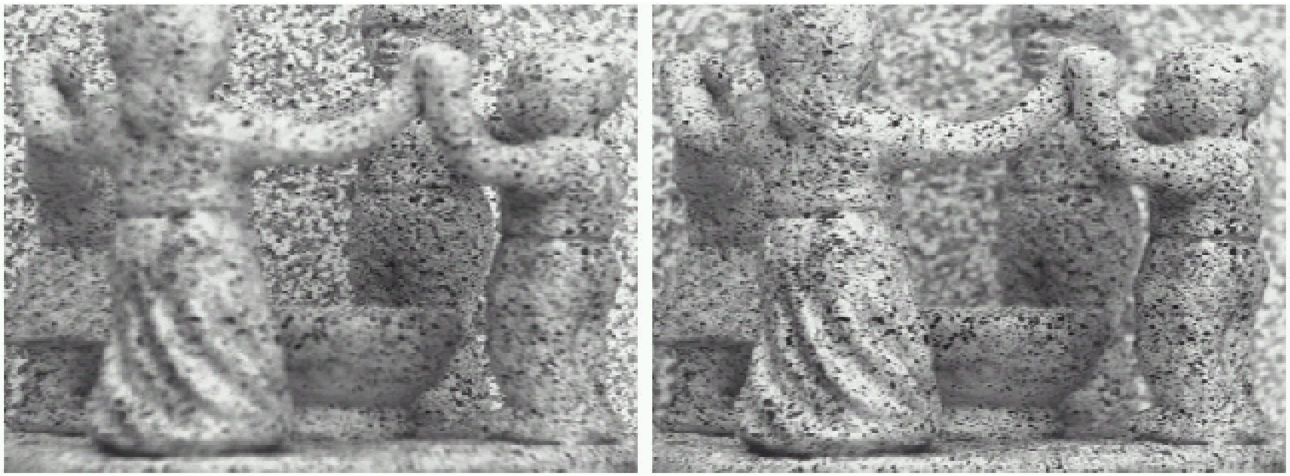


- A smaller aperture increases the range in which the object is approximately in focus

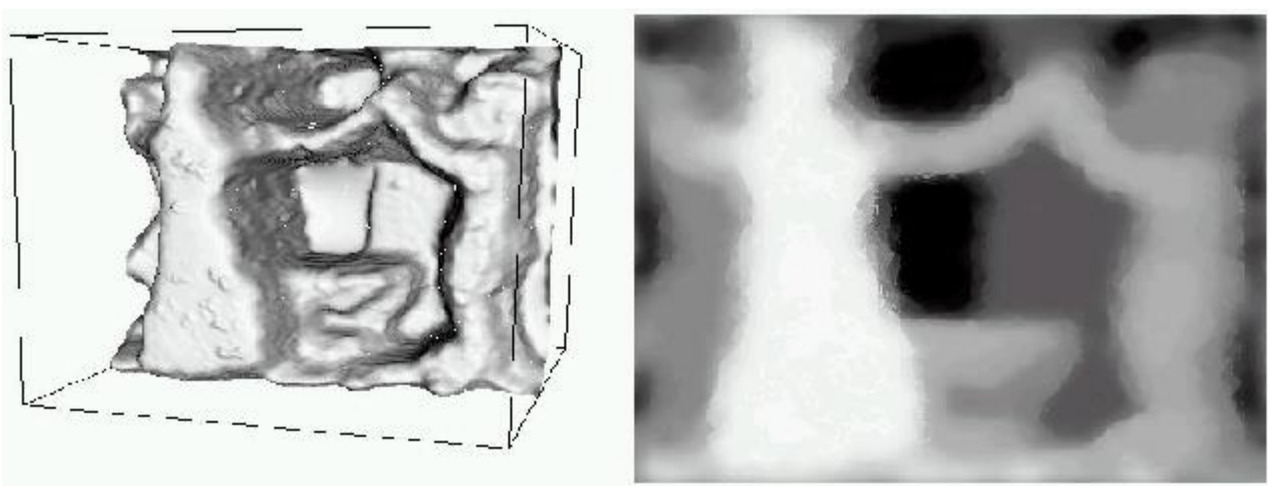
Flower images from Wikipedia http://en.wikipedia.org/wiki/Depth_of_field

Slide from S. Seitz

Application: Depth from (De-)Focus



Images from same point of view, different camera parameters



3D Shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]

Field of View

28 mm lens, 65.5° × 46.4°

50 mm lens, 39.6° × 27.0°

70 mm lens, 28.9° × 19.5°

210 mm lens, 9.8° × 6.5°

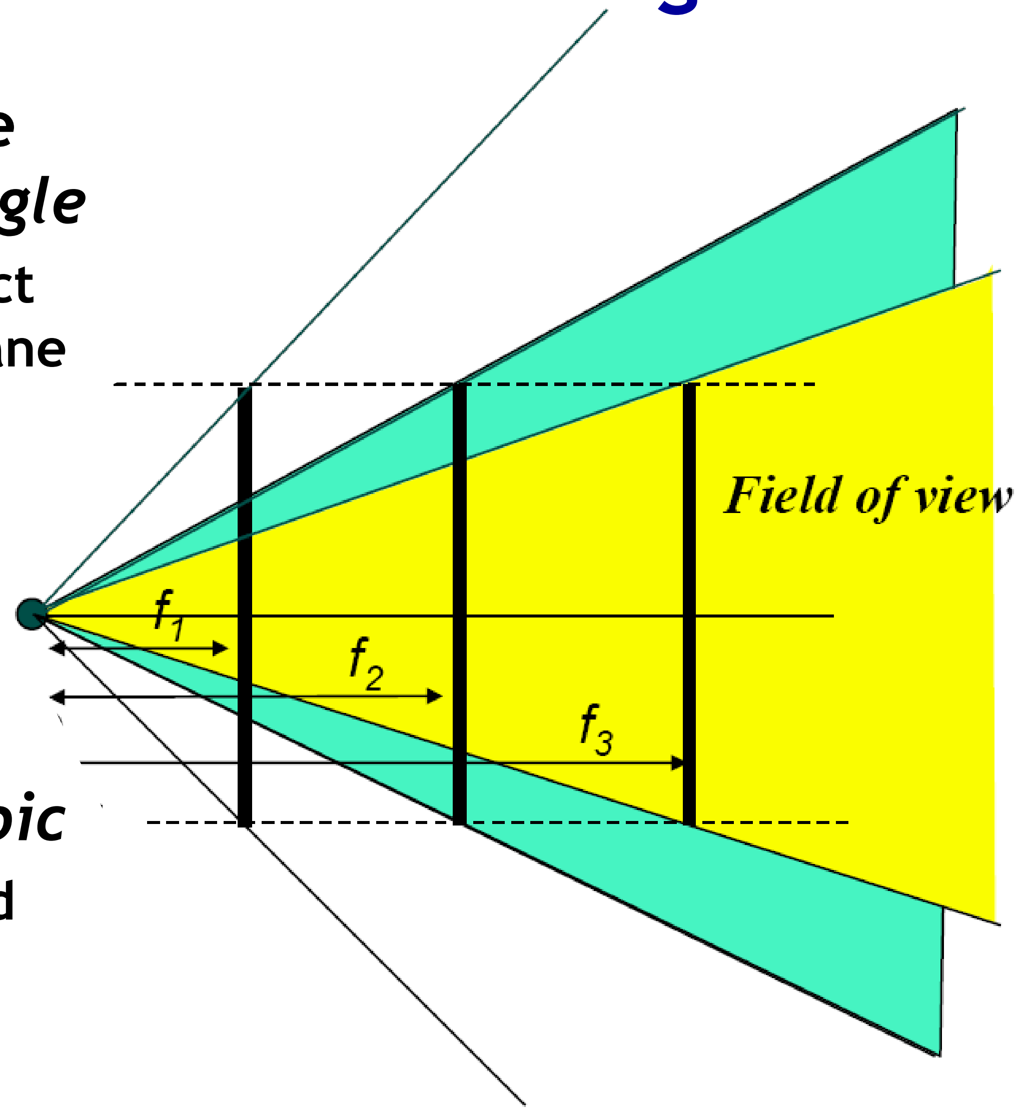
- **Angular measure of the portion of 3D space seen by the camera**

Images from http://en.wikipedia.org/wiki/Angle_of_view

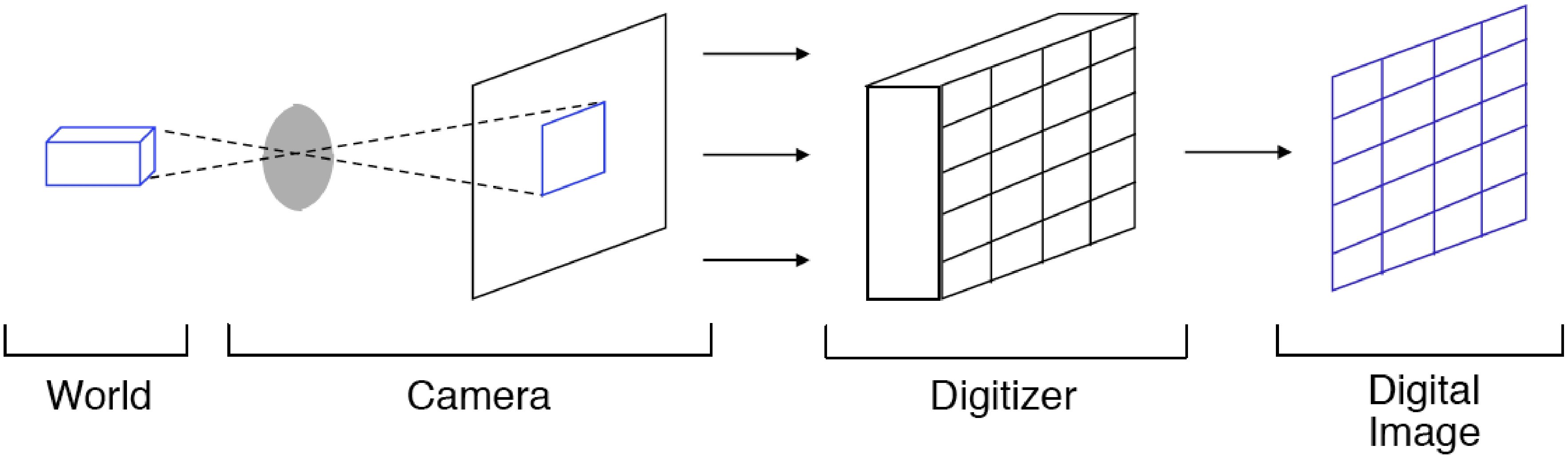
RWTH AACHEN UNIVERSITY

Field of View Depends on Focal Length

- As f gets smaller, image becomes more *wide angle*
  - More world points project onto the finite image plane
- As f gets larger, image becomes more *telescopic*
  - Smaller part of the world projects onto the finite image plane



Digital Images



- Film is replaced by a sensor array
- Current technology: arrays of *charge coupled devices* (CCD)
- *Discretize* the image into pixels
- *Quantize* light intensities into pixel values.

Resolution

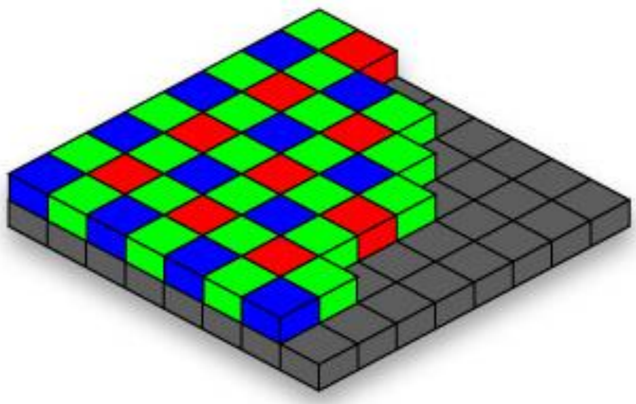
- Sensor: size of real world scene element that images to a single pixel
- Image: number of pixels
- Influences what analysis is feasible, affects best representation choice

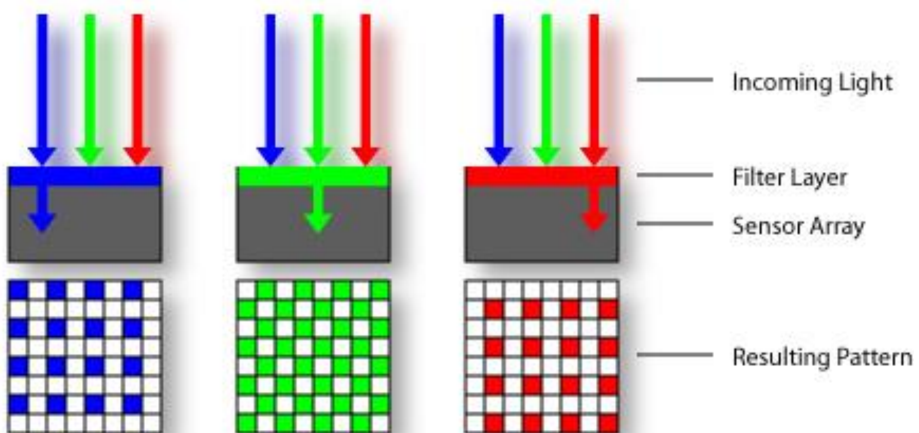
Slide credit: Kristen Grauman

[figs from Efros et al., Mori et al.]

Color Sensing in Digital Cameras

Bayer grid



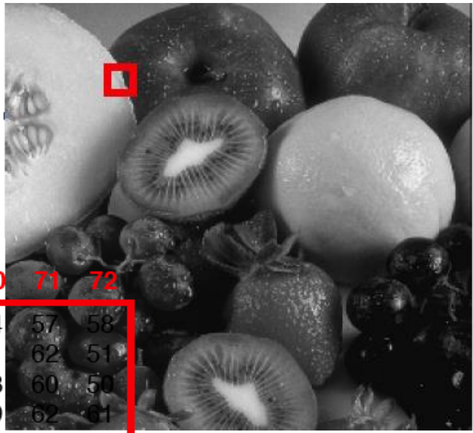


Estimate missing components from neighboring values (demosaicing)

Source: Steve Seitz

Grayscale Image

- **Problem of Computer Vision**
  - How can we recognize fruits from an array of (gray-scale) numbers?
  - How can we perceive depth from an array of (gray-scale) numbers?
  - …

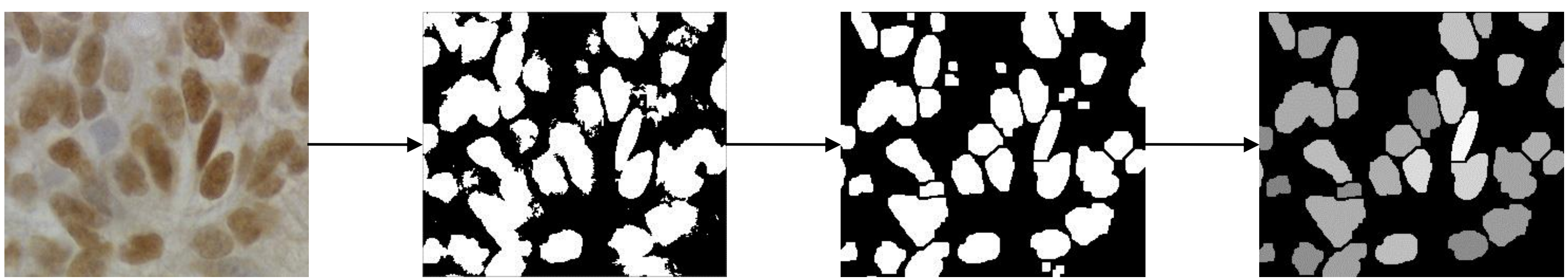


y = \ x =	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
41	210	209	204	202	197	247	143	71	64	80	84	54	54	57	58
42	206	196	203	197	195	210	207	56	63	58	53	53	61	62	51
43	201	207	192	201	198	213	156	69	65	57	55	52	53	60	50
44	216	206	211	193	202	207	208	57	69	60	55	77	49	62	61
45	221	206	211	194	196	197	220	56	63	60	55	46	97	58	106
46	209	214	224	199	194	193	204	173	64	60	59	51	62	56	48
47	204	212	213	208	191	190	191	214	60	62	66	76	51	49	55
48	214	215	215	207	208	180	172	188	69	72	55	49	56	52	56
49	209	205	214	205	204	196	187	196	86	62	66	87	57	60	48
50	208	209	205	203	202	186	174	185	149	71	63	55	55	45	56
51	207	210	211	199	217	194	183	177	209	90	62	64	52	93	52
52	208	205	209	209	197	194	183	187	187	239	58	68	61	51	56
53	204	206	203	209	195	203	188	185	183	221	75	61	58	60	60
54	200	203	199	236	188	197	183	190	183	196	122	63	58	64	66
55	205	210	202	203	199	197	196	181	173	186	105	62	57	64	63

  - How do we humans do it? How can we make a computer do it?

Next Lectures

- First few lectures: low-level vision
  - Binary image processing
  - Filtering operations
  - Edge and structure extraction
  - Color
  - Segmentation and grouping
- Next week: Binary image processing



- Monday 24.10.: *Exercise 1*
  - Intro Matlab, basic image operations

Questions ?