Computer Vision I

Introduction - 17.04.2013

TECHNISCHE UNIVERSITÄT DARMSTADT

vi visual inference

Computer Vision I

TECHNISCHE UNIVERSITÄT DARMSTADT

- Lecturer:
  - Stefan Roth <sroth AT cs.tu-darmstadt...>
  - Office hours: Wednesdays, 13:45 – 14:45
- Teaching Assistants:
  - Stephan Richter <stephan.richter AT gris.tu-darmstadt...>
  - Thorsten Franzel <thorsten.franzel AT gris.tu-darmstadt...>
  - Office hours: TBA
- Course staff email: <cv1staff AT gris.informatik.tu-d...>
  - Please direct any questions here!

Course Material

TECHNISCHE UNIVERSITÄT DARMSTADT

- Course web page: http://goo.gl/zG8GY
  - Contains slides and homework assignments
  - Contains pointers to readings
- Moodle ("HRZ Moodle"):
  - Please check if you are signed up!
  - Automatic through TUCaN!?
  - We will be using it for homework assignments etc.
- Mailing list (see webpage):
  - Please make sure you sign up!

Forum

TECHNISCHE UNIVERSITÄT DARMSTADT

- There is a forum set up at the Fachschaft's website: [http://www.fachschaft.informatik.tu-darmstadt.de/forum/viewforum.php?f=290]
  - Please use it to ask questions of public interest.
  - You are encouraged to discuss with each other.
  - However: Please do not share solutions or give strong hints about the solutions to the homework problems.

Course language

TECHNISCHE UNIVERSITÄT DARMSTADT

- will be English.
  - This applies to lectures, exercises, announcements, etc.
- Why?
  - Essentially all computer vision publications and books are written in English.
  - Knowing the original terms is crucial.
- If strongly preferred, you may contact the course staff in German.
  - English is encouraged though, because we may use your (anonymized) question to clarify points to the entire class.

Organization

TECHNISCHE UNIVERSITÄT DARMSTADT

- Class type: IV4 (new)
- Lecture:
  - ~2 hours a week
  - Wednesdays, 09:50, S3|05, room 074
  - We will cover the foundational aspects of each topic.
- Exercise:
  - ~2 hours a week (irregularly, remaining times: homework)
  - Wednesdays, after lecture, S3|05, room 074
  - We will cover some practical aspects, and discuss the homework assignments.

Exam & Exercises

TECHNISCHE UNIVERSITÄT DARMSTADT

- Exam:
  - (Most likely) written exam at the end of the semester.
  - Can be taken in English or German.
  - Exam date will be set and announced soon.
- Exercises:
  - Regular homework assignments; exercises will be graded.
  - Excercise points are part of your final grade (vorlesungsbleitende Prüfung)! (new)
- Grading: 2/3 exam, 1/3 exercise

Homework assignments

TECHNISCHE UNIVERSITÄT DARMSTADT

- Mostly programming assignments.
  - MATLAB, standard environment for scientific computing.
- Sometimes smaller pen and paper exercises
- The homework exercises are crucial for actually "digesting" the material from the lecture.
  - Hence your points count... (new)
- Time commitment:
  - Solving the homework problems will require a substantial time commitment.
  - Keep in mind that 6CP equal a workload of approximately 180h / semester.

Collaboration policy

TECHNISCHE UNIVERSITÄT DARMSTADT

- I do not tolerate plagiarism in any way!
- You may (and are encouraged to) discuss general class topics.
- But each handed in homework solution must be your **own**!
  - Any sources you used (other than provided by us) must be cited.
  - Details: TBA on the first homework assignment sheet.

- Questions? Problem cases?
  - Talk to us.

Readings

TECHNISCHE UNIVERSITÄT DARMSTADT

- Good news: New book!
  - Computer Vision: Algorithms and Applications by Richard Szeliski (Springer, 2011)
  - Much more accessible than our previous book.
  - Even better news: PDF available online - [http://szeliski.org/Book]
- Additional readings:
  - Papers and tutorials.
  - Will be available on the web or course page.
- I will often assign weekly readings:
  - Please read them and come to class prepared!

How does it fit into your course plan?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Elective:
  - Part of Human Computer Systems (HCS) track.
- Related classes:
  - Human Computer Systems: prerequisite, with exceptions
  - Computer Vision II (Roth, WS, **not in 2013**)
  - Maschinelles Lernen - Statistische Verfahren I (Peters, SS)
  - Maschinelles Lernen - Statistische Verfahren II (Roth, WS, **not in 2013**)
  - Bildverarbeitung (Sakas, SS)
- Theses and projects:
  - Topics in computer vision and machine learning.

Preliminary Syllabus

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Subject to change!**
- 17.04.2011: Introduction and Overview
- 24.04.2011: Image formation
- 01.05.2011: **No class** (public holiday)
- 08.05.2011: Appearance-based matching
- 15.05.2011: Global and local features
- 22.05.2011: **No class** (project review meeting)
- 29.05.2011: Bags of visual words

Preliminary Syllabus

TECHNISCHE UNIVERSITÄT DARMSTADT

- 05.06.2011: Object detection
- 12.06.2011: Motion estimation
- 19.06.2011: Single & two-view geometry
- 26.06.2011: **No class (CVPR)**
- 03.07.2011: Two-view geometry
- 11.07.2011: Stereo & segmentation

Goal of today's lecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Get intuitions about:
  - What is computer vision?
  - Why is it useful?
  - Why is it hard / interesting?
  - How might we go about it?

What is computer vision?

TECHNISCHE UNIVERSITÄT DARMSTADT

[from Steve Seitz]

What does it mean to "see"?

TECHNISCHE UNIVERSITÄT DARMSTADT

see [1] |sē|

verb (**sees** |sēz|, **seeing** |sē-i ng |; past **saw** |sô|; past part. **seen** |sēn|)

[trans.]
1 perceive with the eyes; discern visually *: in the distance she could see the blue sea* |
[intrans.] *Andrew couldn't see out of his left eye* figurative *I can't* ***see into*** *the future.*
• [with clause] be or become aware of something from observation or from a written or other visual source *: I see from your appraisal report that you have asked for training.*

...

[Oxford English dictionary, slide from Michael Black]

What does it mean to "perceive"?

TECHNISCHE UNIVERSITÄT DARMSTADT

perceive |pərˈsēv|

verb [trans.]
1 become aware or conscious of (something); come to realize or understand : *his mouth fell open as he perceived the truth* | [with clause] *he was quick to perceive that there was little future in such arguments.*
• become aware of (something) by the use of one of the senses, esp. that of sight : *he perceived the faintest of flushes creeping up her neck.*
2 interpret or look on (someone or something) in a particular way; regard as : *if Guy does not* ***perceive*** *himself* ***as*** *disabled, nobody else should* | [trans.] *some geographers perceive hydrology to be a separate field of scientific inquiry.*

[Oxford English dictionary, slide from Michael Black]

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- What is computer vision (or machine vision)?
  - Developing computational models and algorithms to interpret digital images / understand the visual world we live in.

[from Steve Seitz]

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- What is computer vision (or machine vision)?
  - Developing computational models and algorithms to interpret digital images / understand the visual world we live in.
- Is that important?
- What can we (already) do with it?

Face detection

e.g. Canon [powershot.com], etc.

Human pose estimation

Microsoft Kinect

[www.xbox.com/kinect]

Earth viewers

TECHNISCHE UNIVERSITÄT DARMSTADT

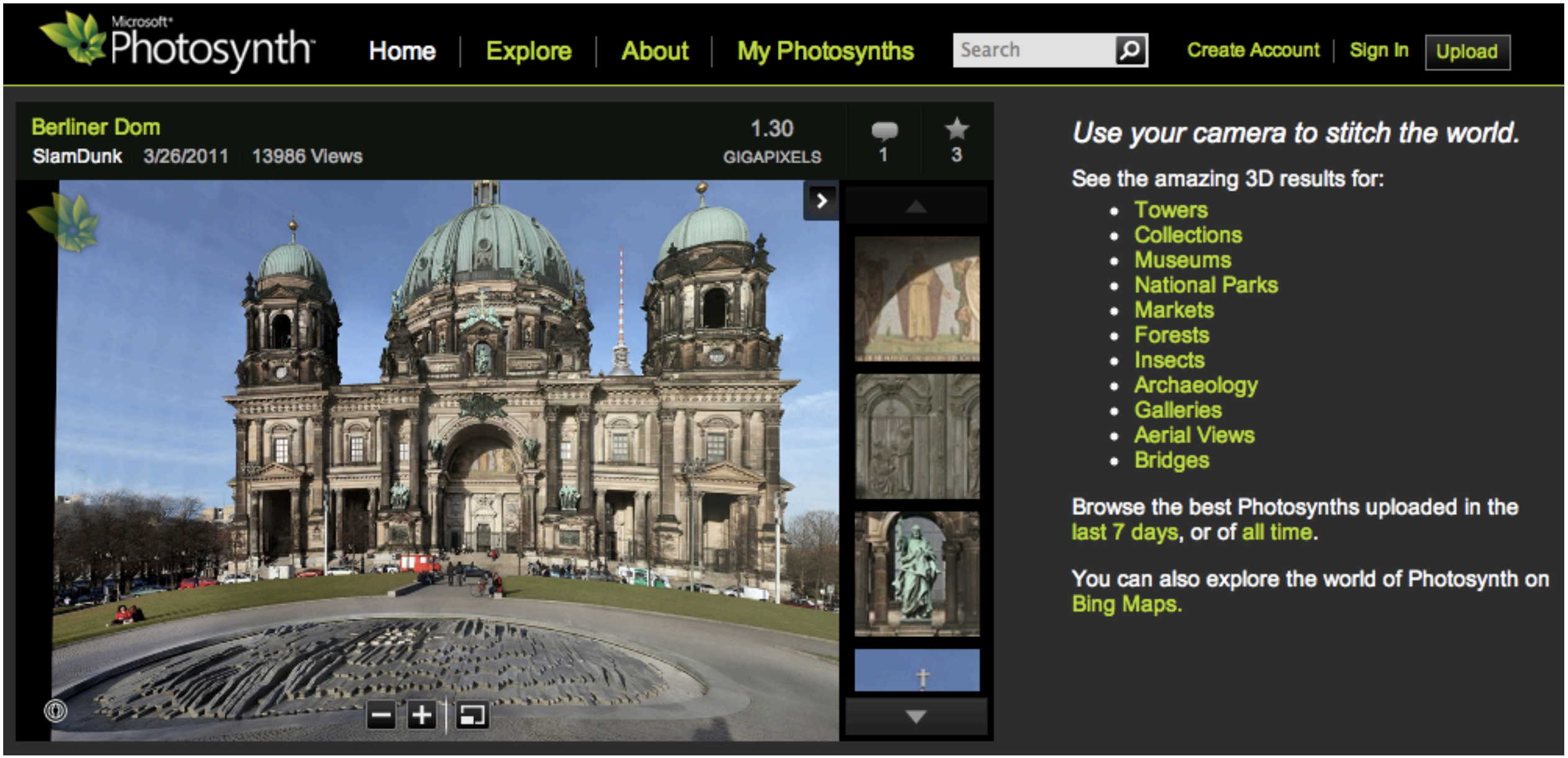
Google Street View

[www.google.com]

[www.mobile3dcity.com
]

Photosynth

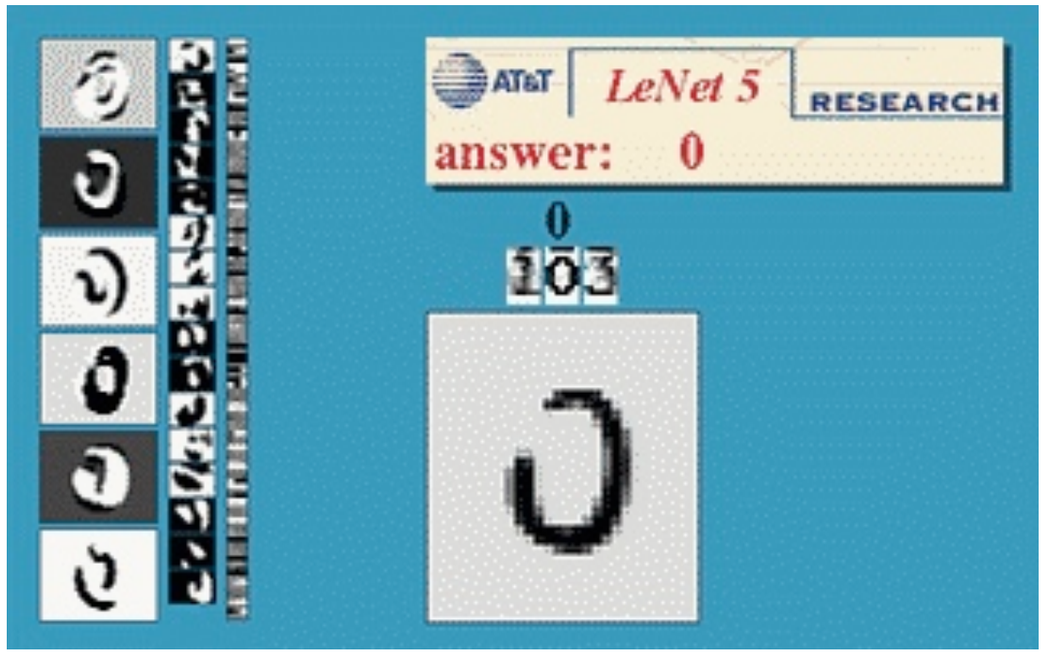
TECHNISCHE UNIVERSITÄT DARMSTADT



[photosynth.net]

Optical character recognition

- Convert scanned documents to text



Digit recognition, AT&T labs
[www.research.att.com/~yann/]

License plate readers
[en.wikipedia.org/wiki/Automatic_number_plate_recognition]

[from Steve Seitz]

Traffic sign recognition

TECHNISCHE UNIVERSITÄT DARMSTADT

Opel

VDO

Special effects: Shape capture

TECHNISCHE UNIVERSITÄT DARMSTADT

The Matrix movies, ...

[from Steve Seitz]

Special effects: Motion capture

TECHNISCHE UNIVERSITÄT DARMSTADT

Pirates of the Carribean, Industrial Light and Magic [from Steve Seitz]

Sports

3D soccer analysis

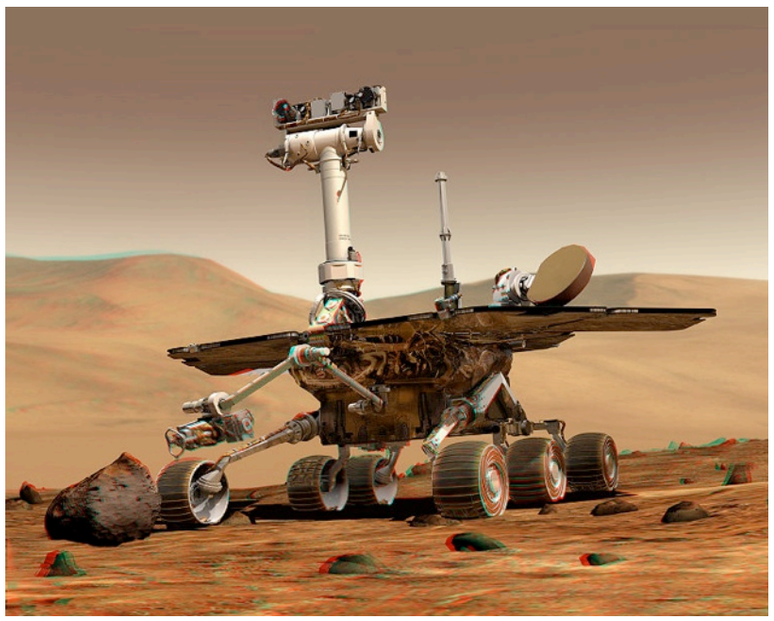
[www.kicker.de]

Vision in space

TECHNISCHE UNIVERSITÄT DARMSTADT

NASA's mars exploration rover Spirit

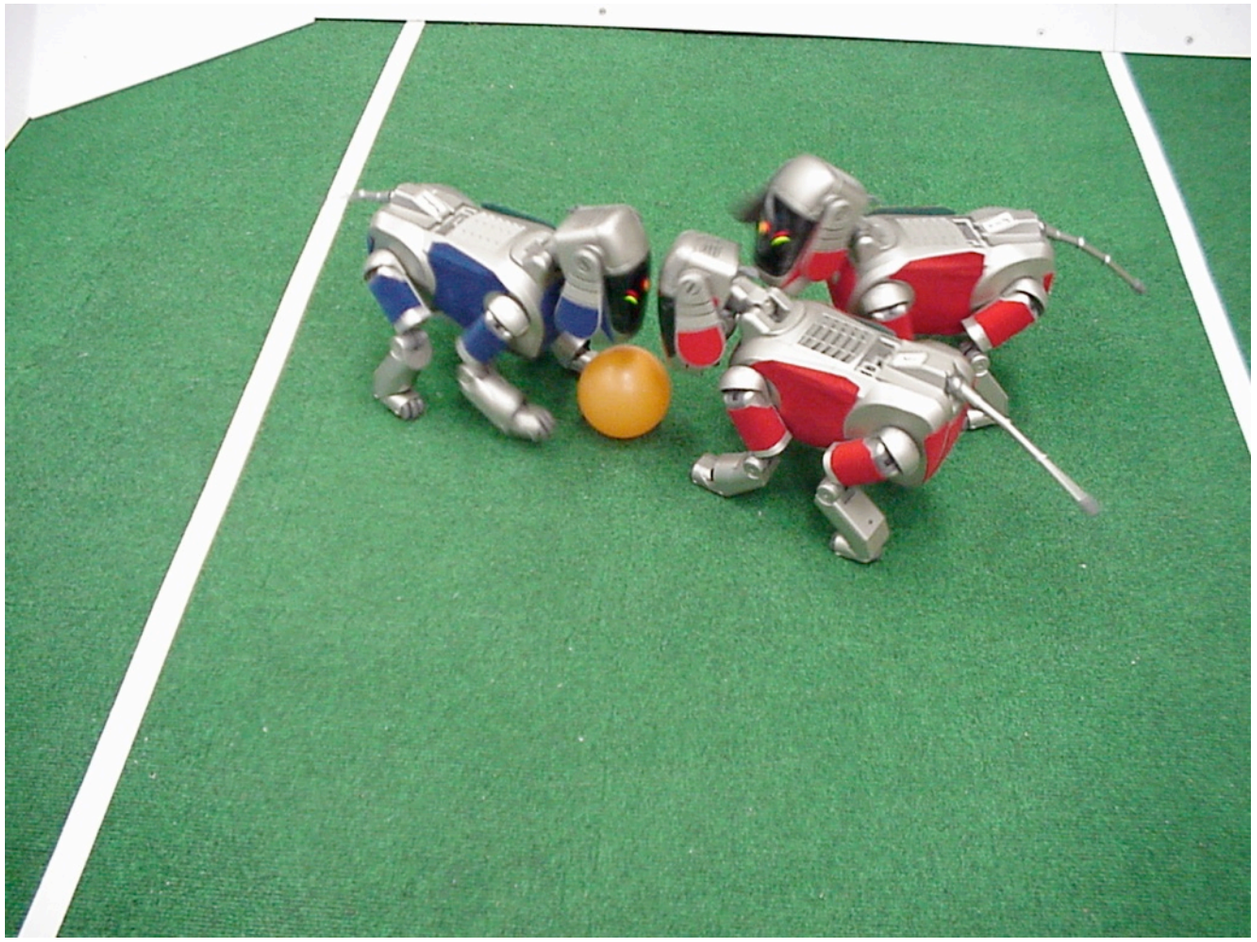
[en.wikipedia.org/wiki/Spirit_rover]



[from Steve Seitz]

Robotics

TECHNISCHE UNIVERSITÄT DARMSTADT



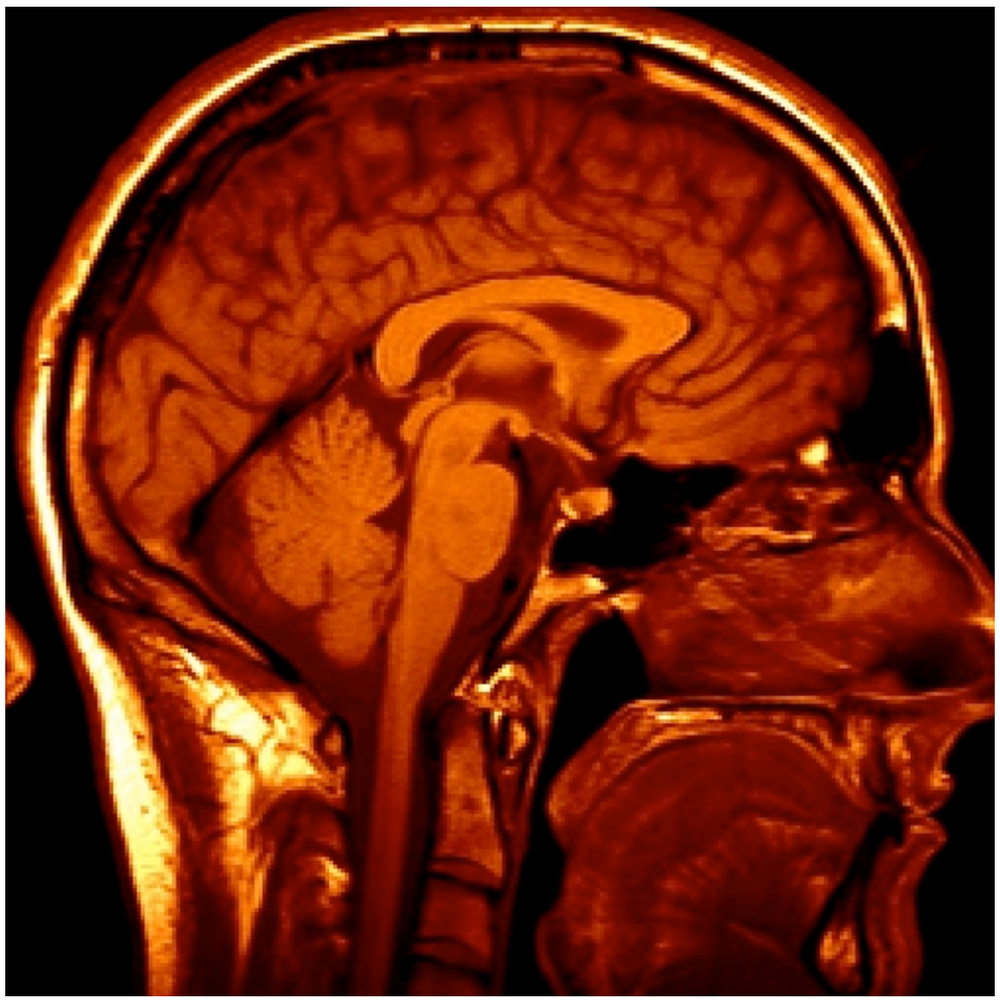
Robocup

[www.robocup.org]

Darmstadt Dribblers [FG SIM]

Medical imaging

TECHNISCHE UNIVERSITÄT DARMSTADT



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

[from Steve Seitz]

Why is computer vision interesting (for you)?

TECHNISCHE UNIVERSITÄT DARMSTADT

- It is a challenging problem that is far from being solved.
- It combines insights and tools from many fields and disciplines:
  - Mathematics and statistics
  - Cognition and perception
  - Engineering (signal processing)
  - And of course, computer science

Why is computer vision interesting (for you)?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Allows you to apply theoretical skills
  - ... that you may otherwise only use rarely.
- Quite rewarding:
  - Often visually intuitive and encouraging results.

- It is a growing field:
  - Cameras are becoming increasingly commonplace.
  - There are a number of vision companies both here and abroad.
  - Even the big players are hiring computer vision experts.
  - Conferences are growing rapidly.

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- What is computer vision (or machine vision)?
  - Developing computational models and algorithms to interpret digital images / understand the visual world we live in.
- How could we go about it?

[from Steve Seitz]

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- We need a formal model that describes our problem *as well as* an algorithm that realizes (i.e. implements) it.
  - Neither the model alone, nor the algorithm alone suffices (in the long run).
  - Both mathematical and computational.

- What properties / cues of the visual world can we exploit or measure?
- What general (prior) knowledge of the world (not necessarily visual) should we exploit?

[adapted from Michael Black]

Case Study: Art & Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

[from Michael Black]

Case Study: Art & Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

(C) Linda Carson 2002

[from Michael Black]

Case Study: Art & Vision

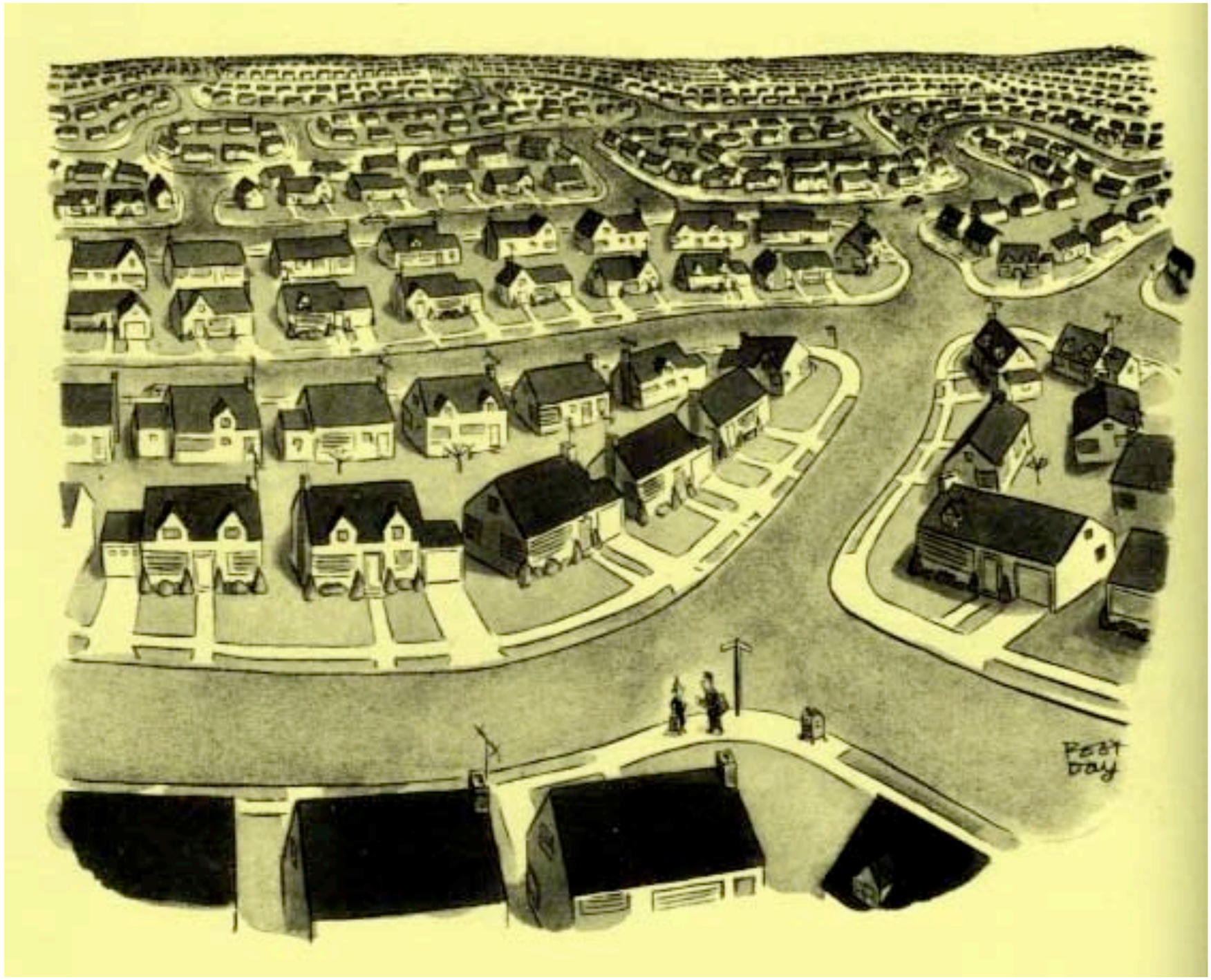
TECHNISCHE UNIVERSITÄT DARMSTADT

(C) Linda Carson 2002

[from Michael Black]

Case Study: Art & Vision

TECHNISCHE UNIVERSITÄT DARMSTADT



[from Michael Black]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

- Stereo parallax:

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

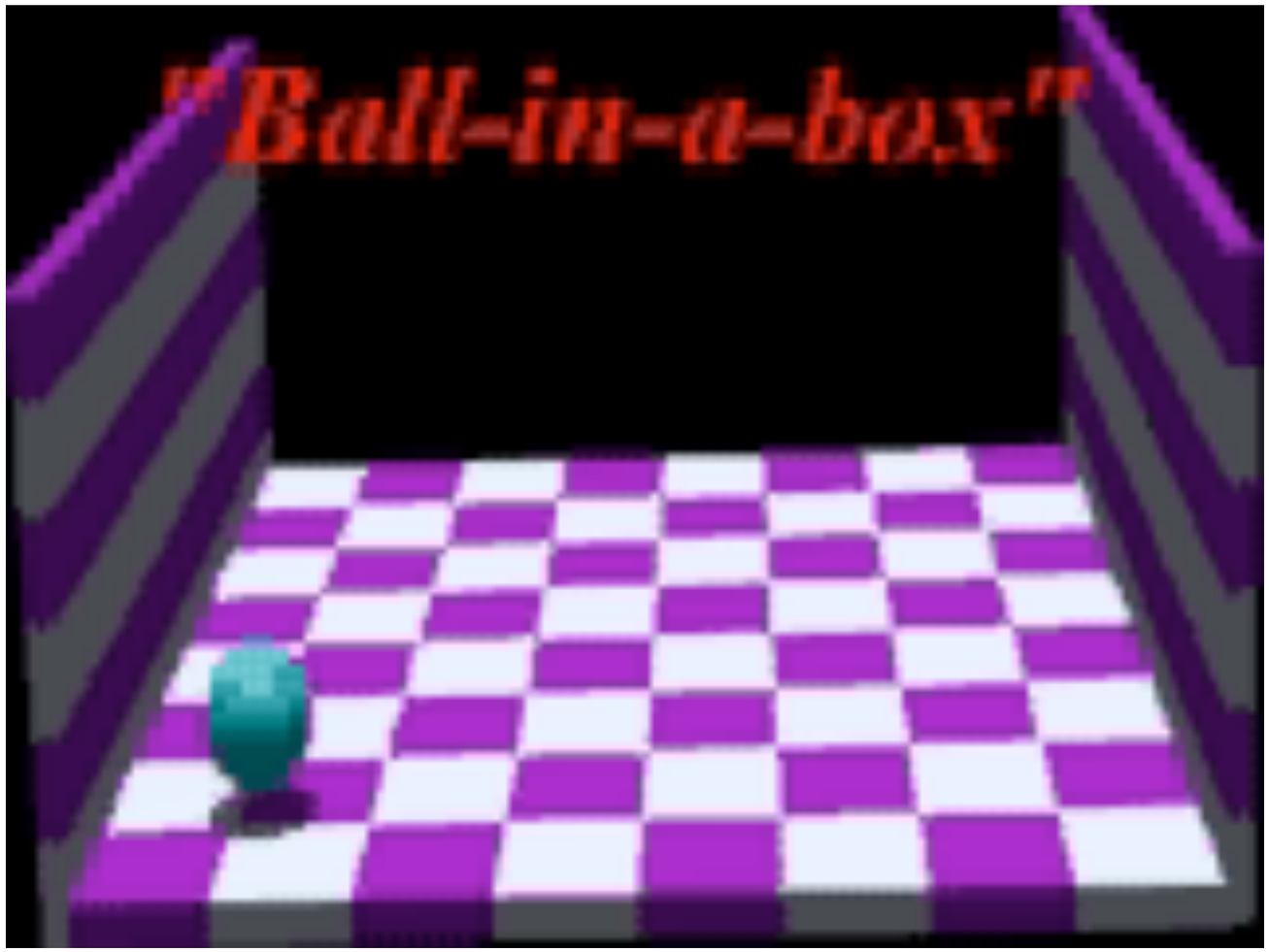
- Motion parallax:

[from Michael Black]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

- Shadows:



[from Michael Black]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

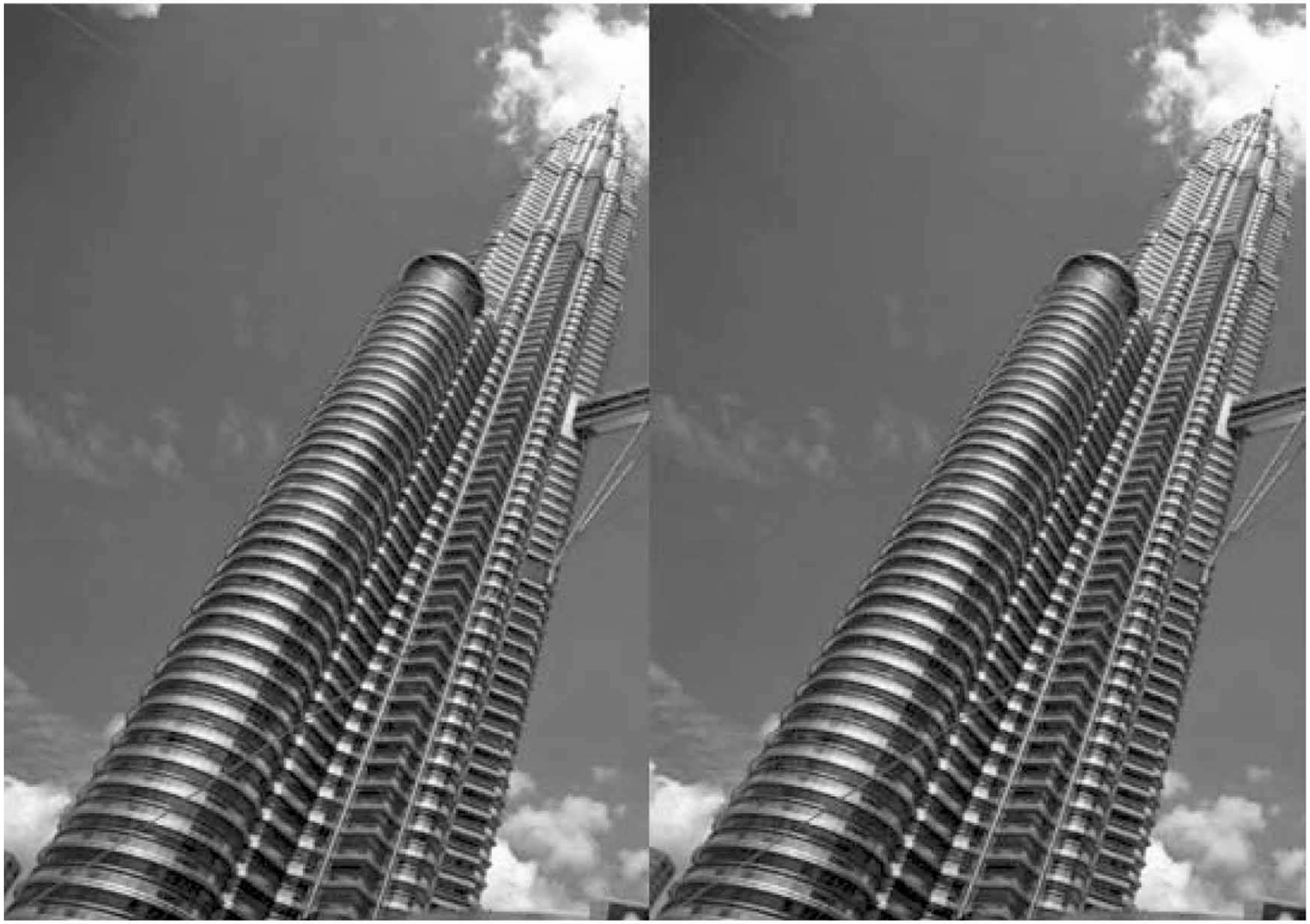
- Convergence:

[from Kingdom et al. 2007]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

- Convergence:



[from Kingdom et al. 2007]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

- Context:

[from Antonio Torralba]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

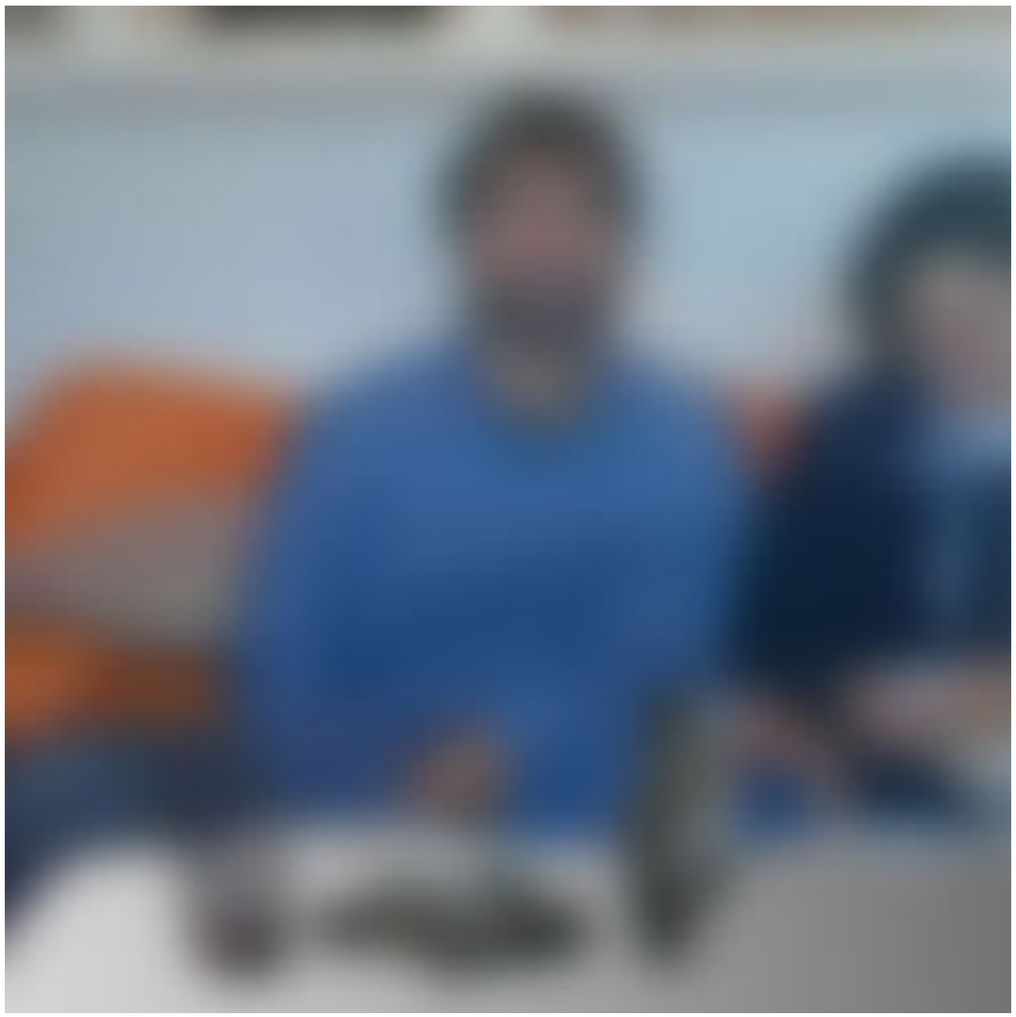
- Context:

[from Antonio Torralba]

Case Study: (Human) Visual Cues

TECHNISCHE UNIVERSITÄT DARMSTADT

- Context:



[from Antonio Torralba]

Is human vision the reference?

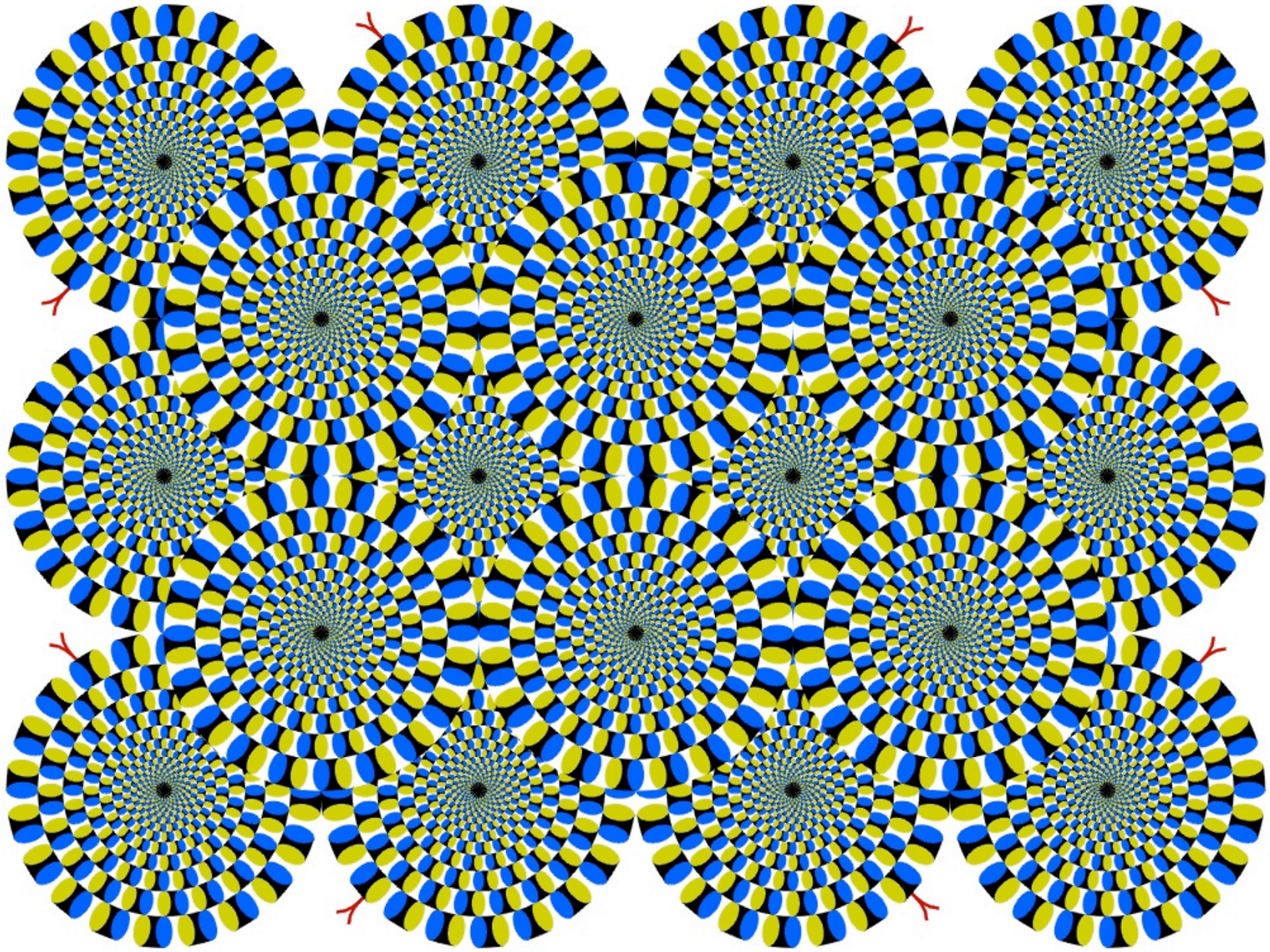
TECHNISCHE UNIVERSITÄT DARMSTADT

Sinha and Poggio, Nature, 1996

[from Steve Seitz]

Is human vision the reference?

TECHNISCHE UNIVERSITÄT DARMSTADT



Copyright A.Kitaoka 2003

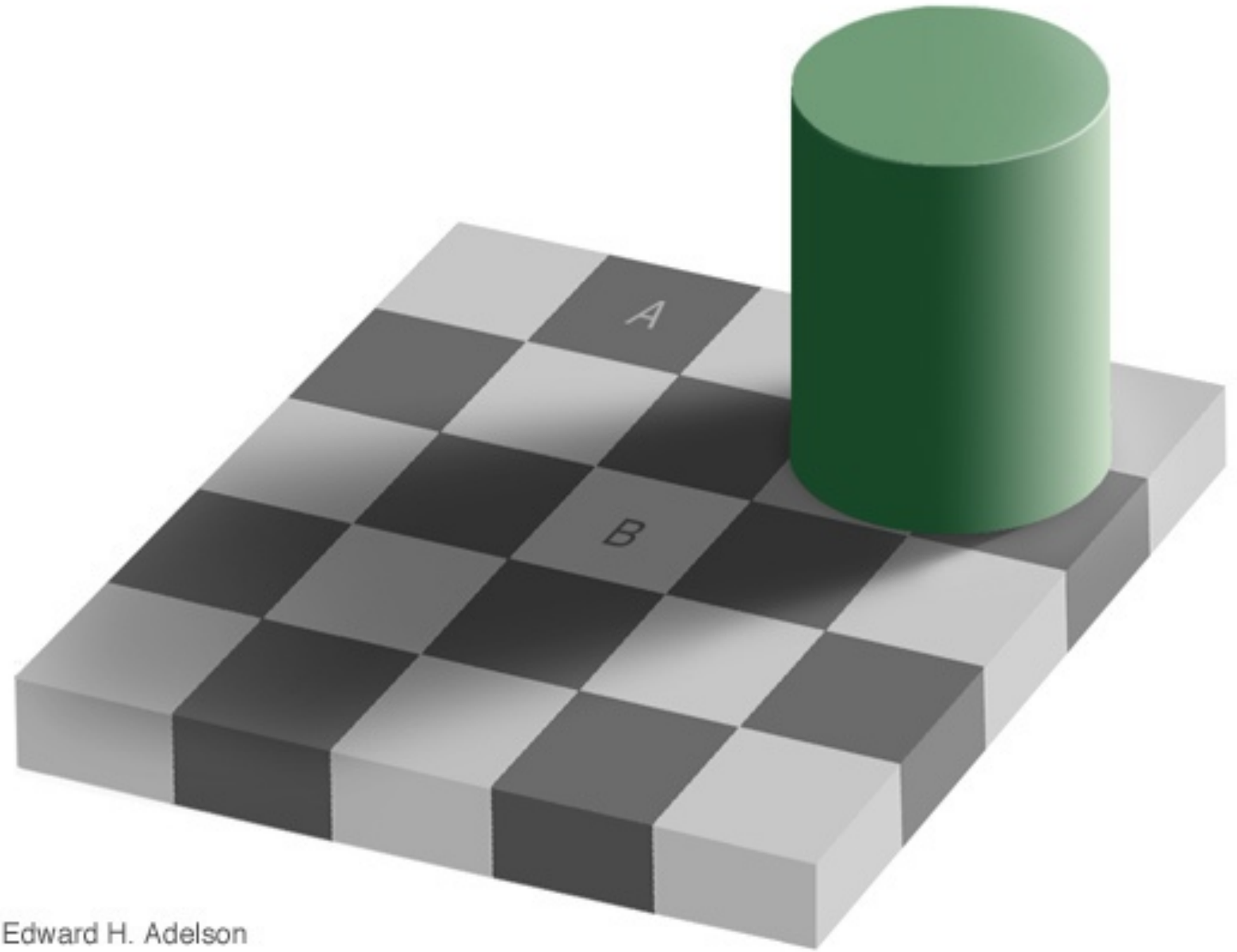
[from Steve Seitz]

Is human vision the reference?

[from Michael Black]

Measurement vs. Interpretation

TECHNISCHE UNIVERSITÄT DARMSTADT



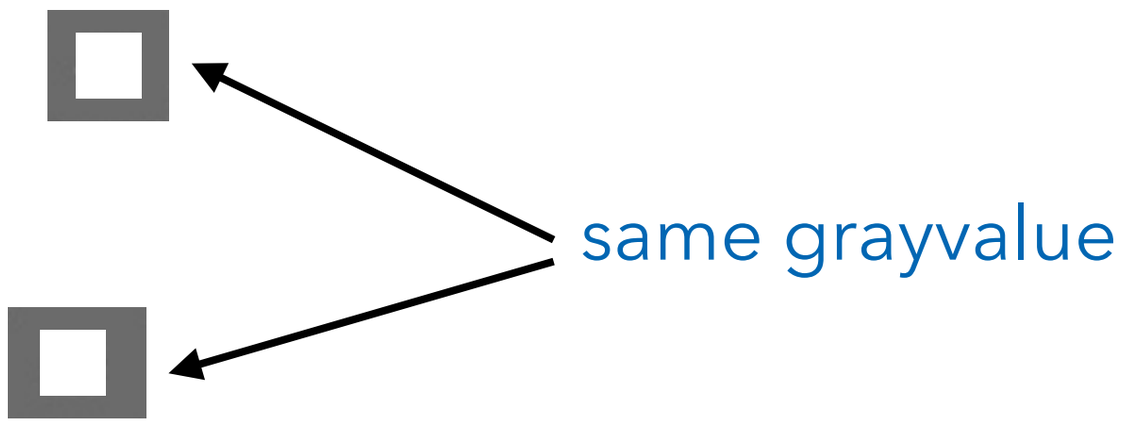
Edward H. Adelson

Measurement vs. Interpretation

TECHNISCHE UNIVERSITÄT DARMSTADT

Measurement vs. Interpretation

TECHNISCHE UNIVERSITÄT DARMSTADT



Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

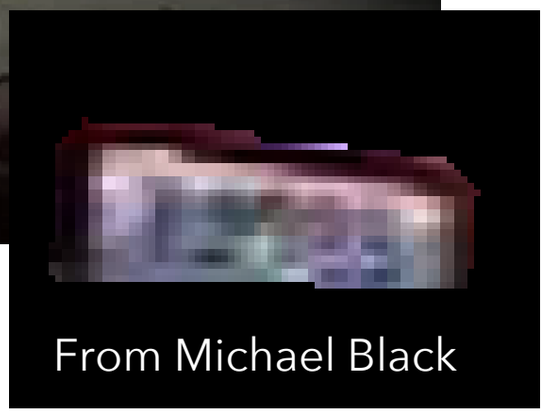
- We as humans use many different cues to interpret what is in an image.
  - They are not always right.
- We use information on what we regard as being a plausible and meaningful interpretation.
  - The measurement alone does not suffice.
- This is necessary, because an image is a 2D projection of a complex 3D world.
  - A lot of information about the world is lost when we take a picture.

Ambiguity of Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- Our image data is not only too little to fully recover and understand the “state of the visible world”.
- It may even be of poor quality:
  - Low resolution
  - (Sensor) noise
  - Etc.


Courtesy Criterion Collection


From Michael Black

- Our image data is always ambiguous.

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- We want to devise computer algorithms to understand the visual world much like we do.
- This means:
  - We have to use a lot of different cues...
  - We have to deal with ambiguity...
  - We have to exploit what is a plausible and meaningful interpretation...
  - ... in order to extract information about the 3D visual world from a small amount of data.
- CV is an inverse problem.

Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

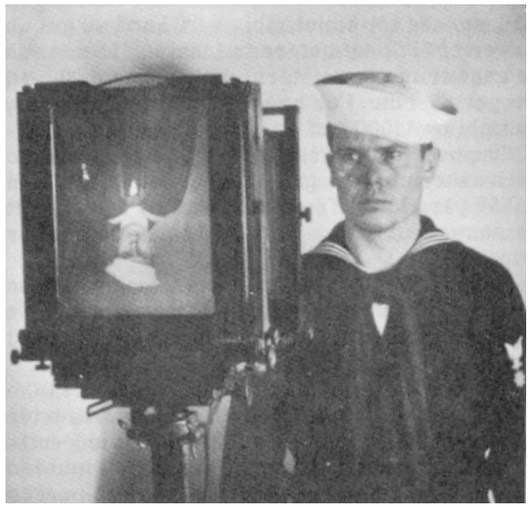
- How do we turn these cues around?
  - "inverse graphics"?

- We need:
  - To understand the geometry and physics of the world (to some extent).
  - A mathematical model of the cues that we want to exploit.
  - A mathematical model of our prior knowledge of the world.
  - An computational model and an algorithm to infer the state of the world from cues and prior knowledge.

Cameras: Introduction

TECHNISCHE UNIVERSITÄT DARMSTADT

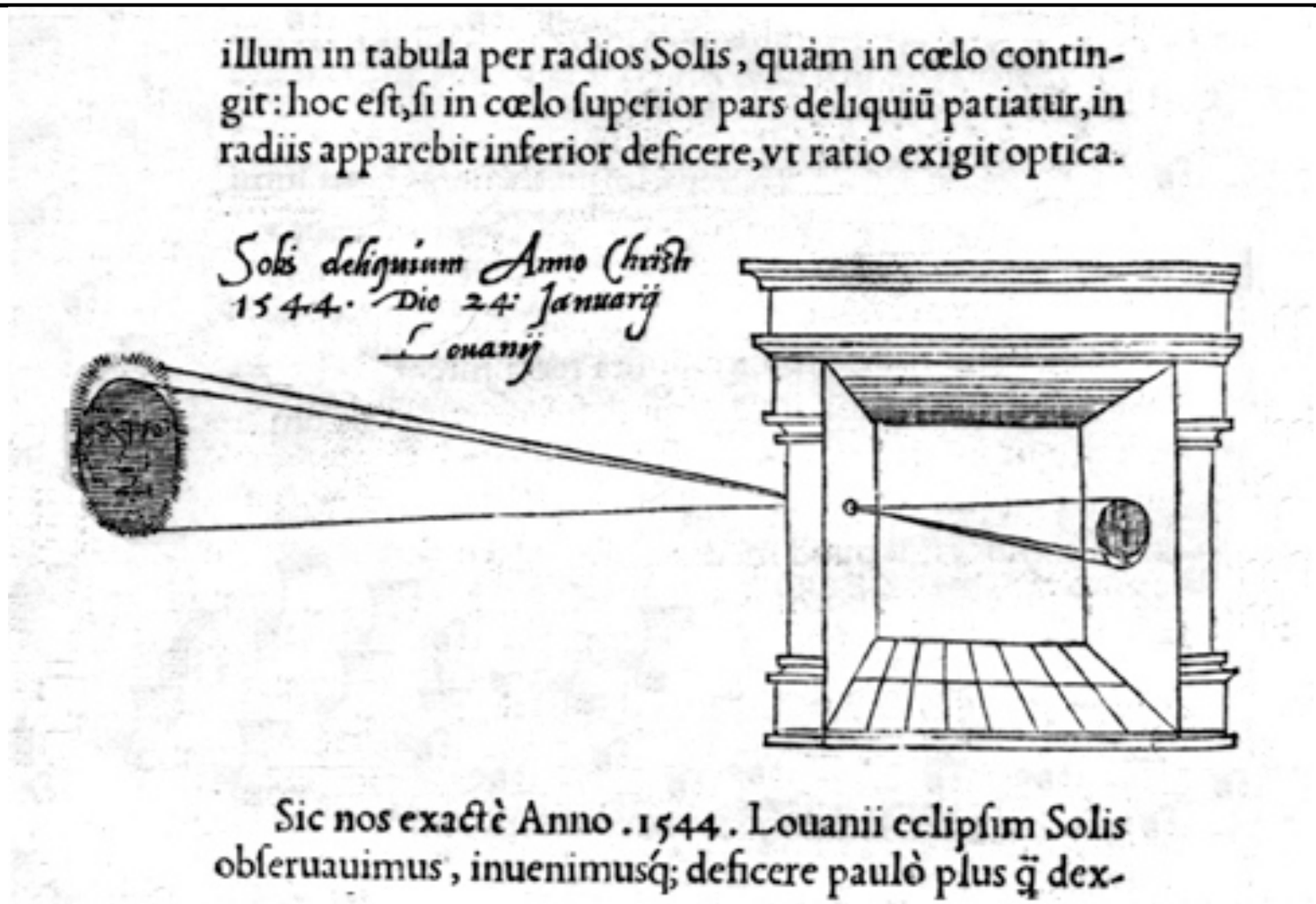
- Cameras are everywhere these days.
  - Not just your digital still and video cameras.
  - Essentially every new cell phone has a camera built in.
    - More cameras than people?
  - Many new computers have a camera built in.
- History:
  - From "camera obscura" (dark room)
  - Basic principle known to Mozi (470-391 BC) and Aristoteles (384-322 BC)
  - Described by Ibn al-Haytham (Alhazen) in his "Book of optics" (around 1000AD).



From [Ponce & Forsyth]

[www.grand-illusions.com]

Camera obscura


illum in tabula per radios Solis, quàm in cœlo contin-
git: hoc eſt, ſi in cœlo ſuperior pars deliquiũ patiatur, in
radiis apparebit inferior deficere, vt ratio exigit optica.

Solis deliquium Anno Christi
1544. Die 24. Januarij
Louanij

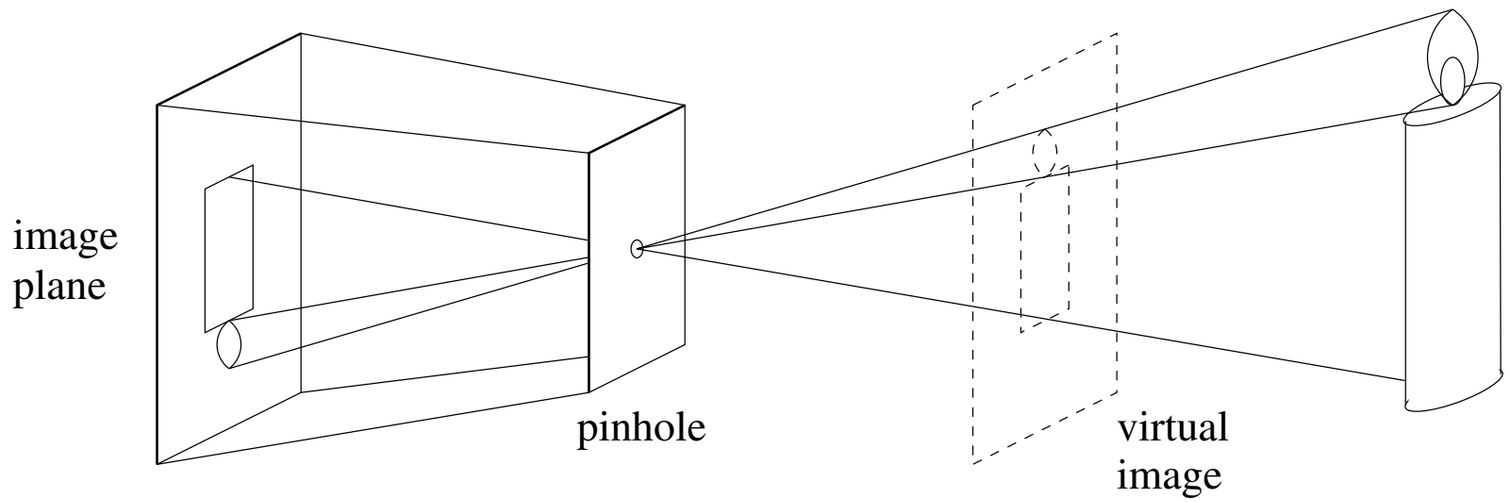
Sic nos exactè Anno .1544. Louanii eclipſim Solis
obſeruauimus, inuenimusq; deficere paulò plus q̃ dex-

"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". Da Vinci [from Michael Black]

Cameras: Introduction

TECHNISCHE UNIVERSITÄT DARMSTADT

- Idealized standard camera model:
  - Pinhole camera



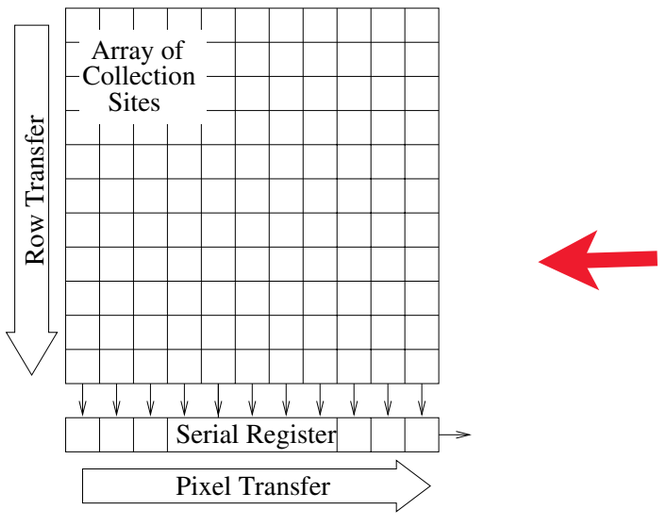
From [Ponce & Forsyth]

- Very simple, but for many purposes sufficient.
- Usually, we work with the upright virtual image.

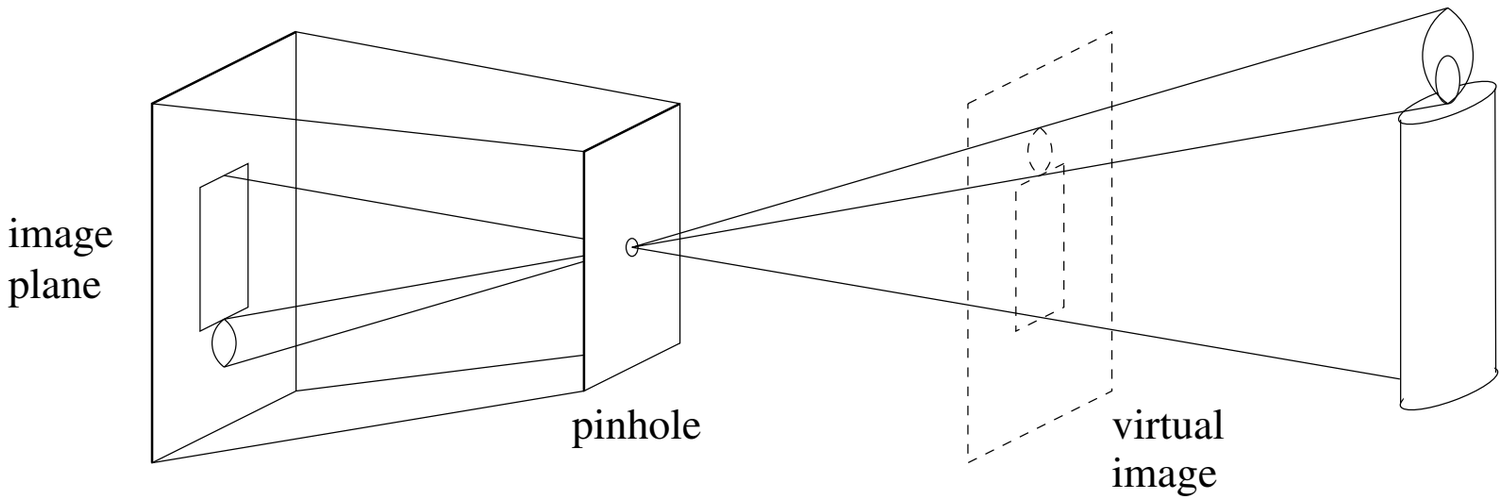
Cameras: Introduction

TECHNISCHE UNIVERSITÄT DARMSTADT

- Idealized standard camera model:
  - Pinhole camera



CCD or CMOS



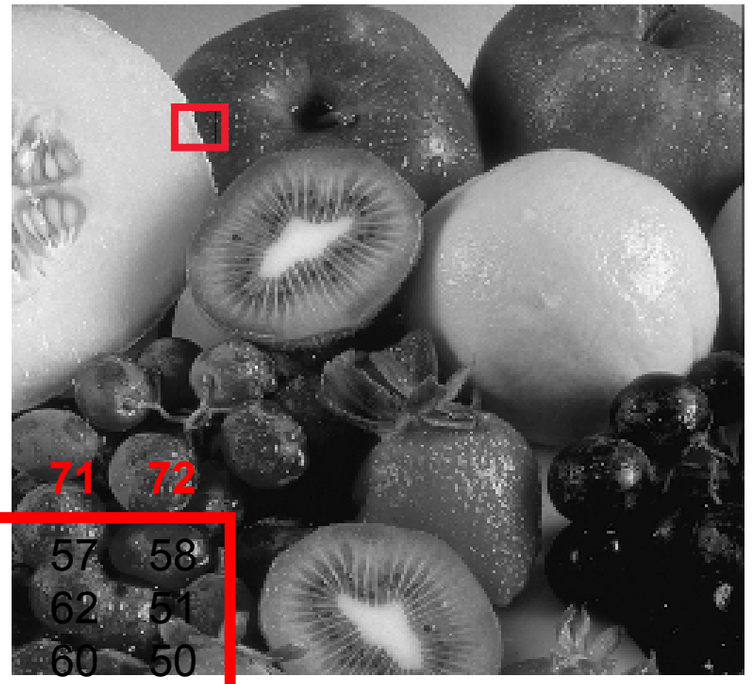
From [Ponce & Forsyth]

- Very simple, but for many purposes sufficient.
- Usually, we work with the upright virtual image.

Aside & Reminder: What is vision about?

TECHNISCHE UNIVERSITÄT DARMSTADT

- How do we make sense of this array of numbers?



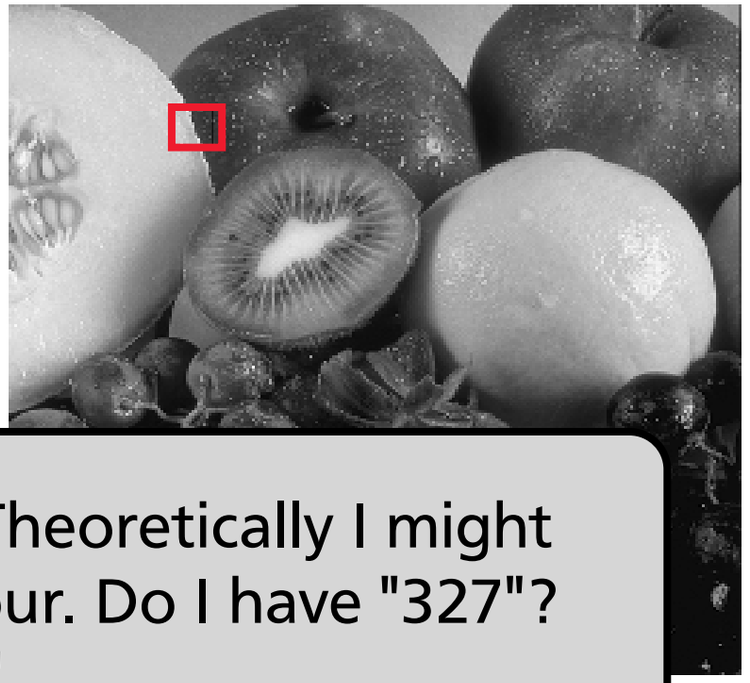
	x = 58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
y = 41	210	209	204	202	197	247	143	71	64	80	84	54	54	57	58
42	206	196	203	197	195	210	207	56	63	58	53	53	61	62	51
43	201	207	192	201	198	213	156	69	65	57	55	52	53	60	50
44	216	206	211	193	202	207	208	57	69	60	55	77	49	62	61
45	221	206	211	194	196	197	220	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

[from Michael Black]

Aside & Reminder: What is vision about?

TECHNISCHE UNIVERSITÄT DARMSTADT

- How do we make sense of this array of numbers?



> "I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of colour. Do I have "327"? No. I have sky, house, and trees."
>
> Max Wertheimer, 1923

x =

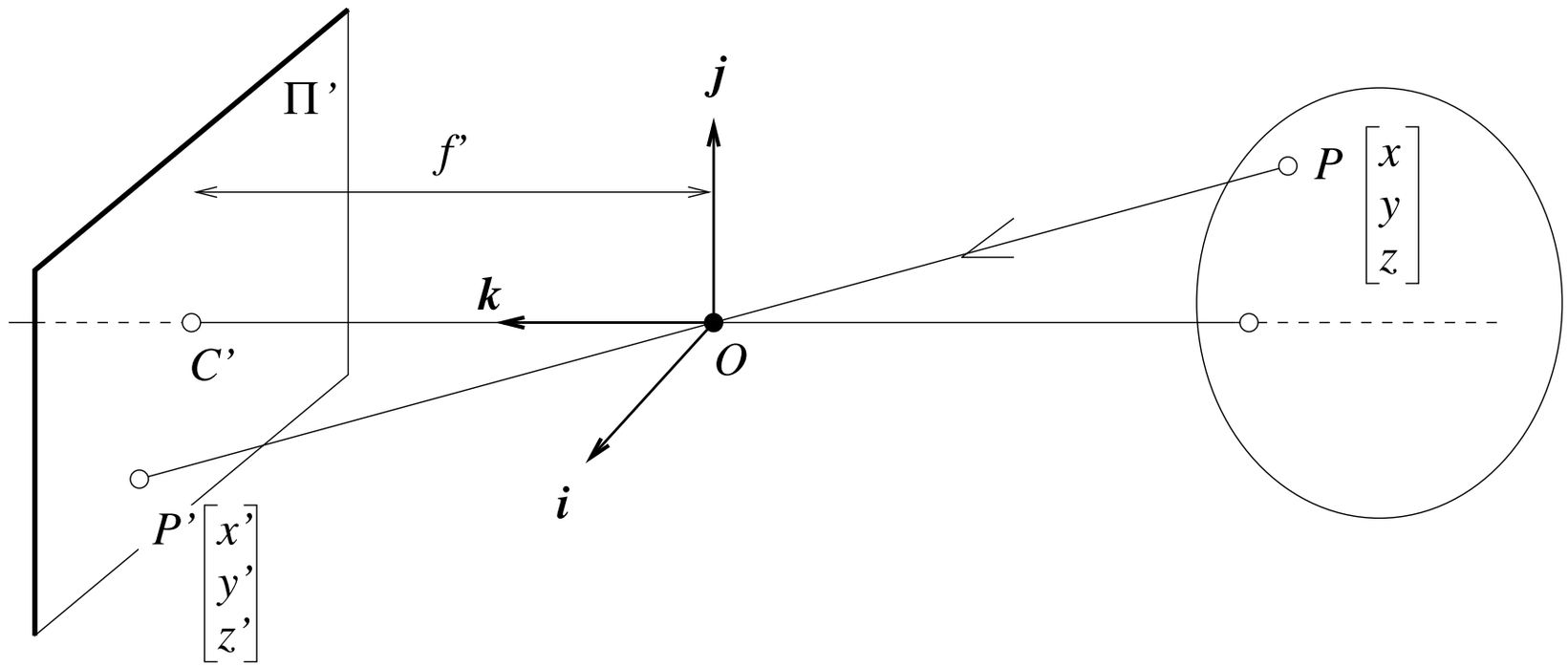
y

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

Perspective cameras

TECHNISCHE UNIVERSITÄT DARMSTADT

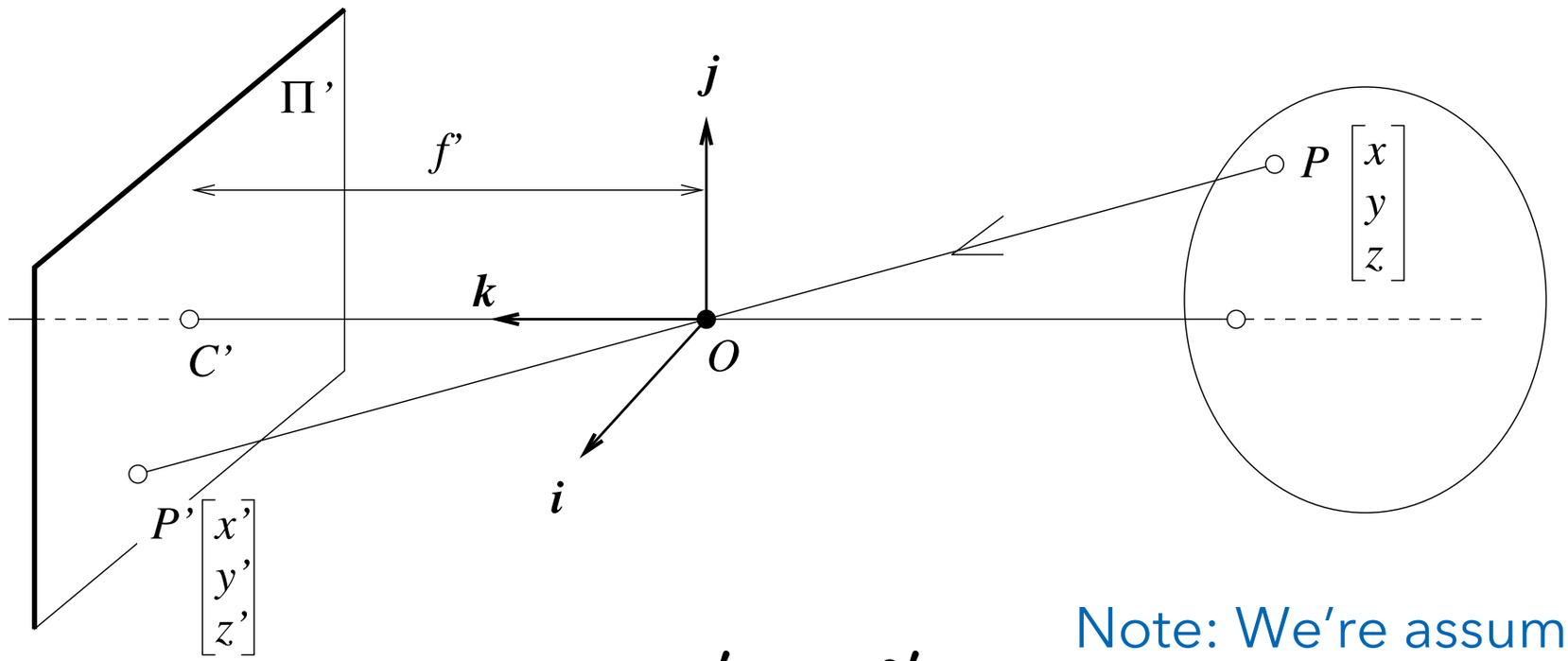
- Perspective projection with a pinhole camera:



- World point: $P = (x, y, z)^{\mathrm{T}}$
- Image point: $P' = (x', y', z')^{\mathrm{T}}$
- Focal length: f'

Perspective cameras

- Perspective projection with a pinhole camera:



- Projected depth:
- Projective projection:

$$z' = f'$$

$$\frac{x}{x'} = \frac{z}{f'} \quad \Rightarrow \quad x' = f' \cdot \frac{x}{z}$$

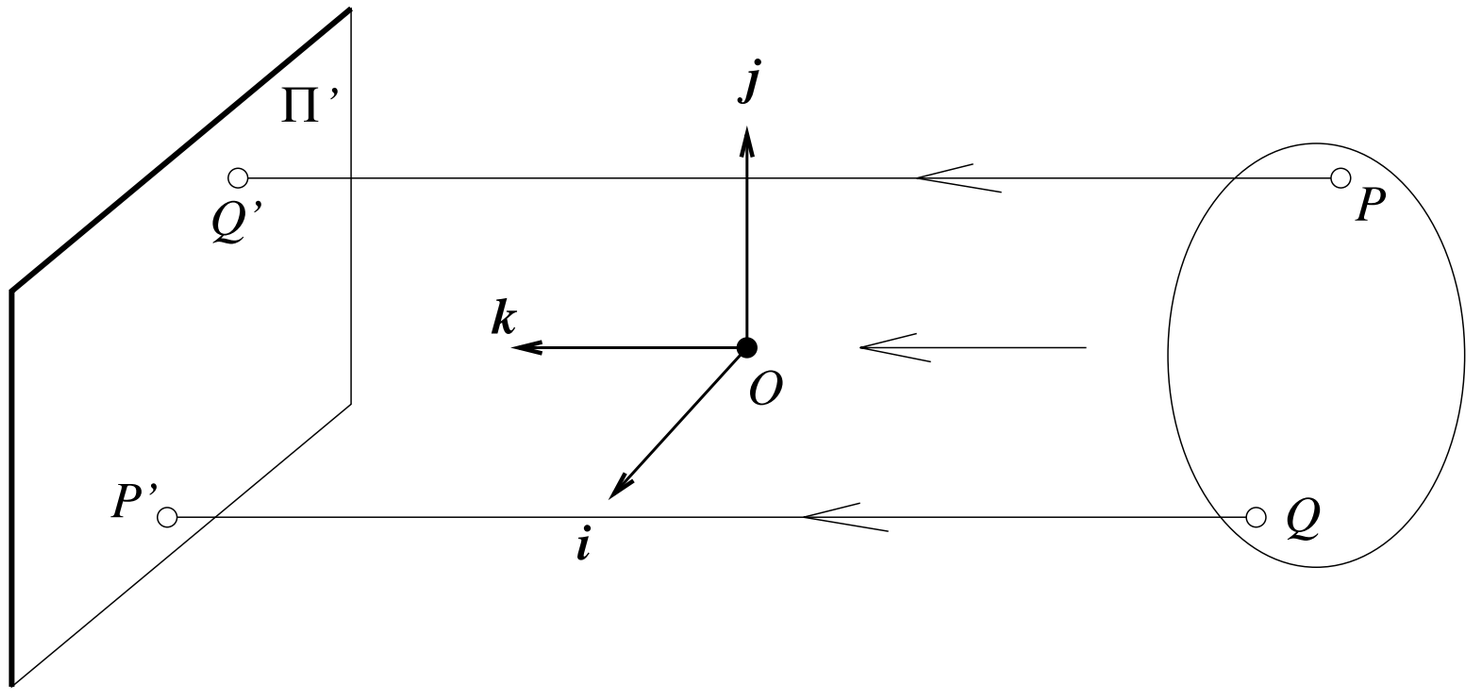
$$\frac{y}{y'} = \frac{z}{f'} \quad \Rightarrow \quad y' = f' \cdot \frac{y}{z}$$

Note: We're assuming a virtual image!

Orthographic cameras

TECHNISCHE UNIVERSITÄT DARMSTADT

- Even simpler: Orthographic camera

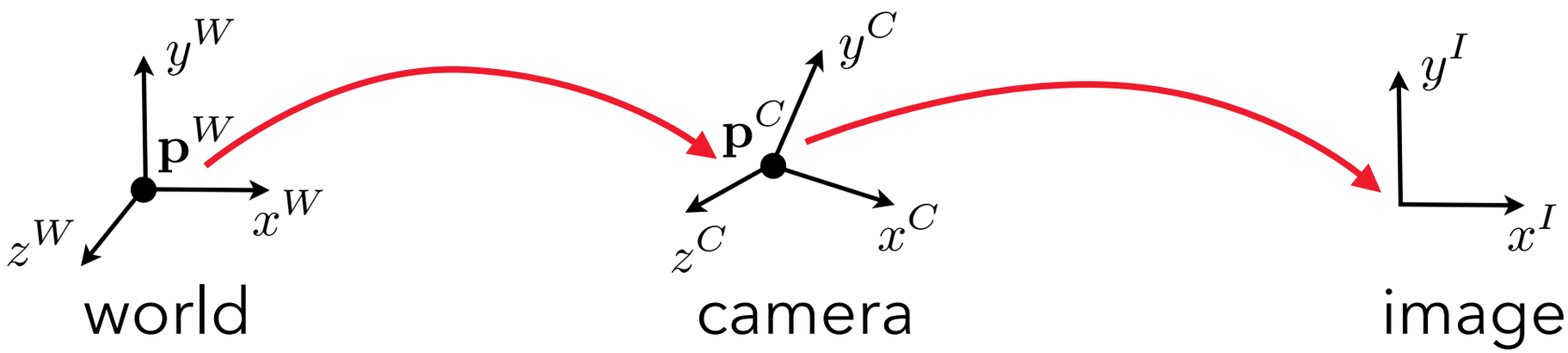


$$x' = x$$
$$y' = y$$

Coordinate Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

- Objects live in "world coordinates".
- Camera has "camera coordinate" system.
- Image is defined in "image coordinates".
- Transformations:
  - Extrinsic camera transformation takes world into camera coordinates.
  - Intrinsic camera transformation describes the image formation process.



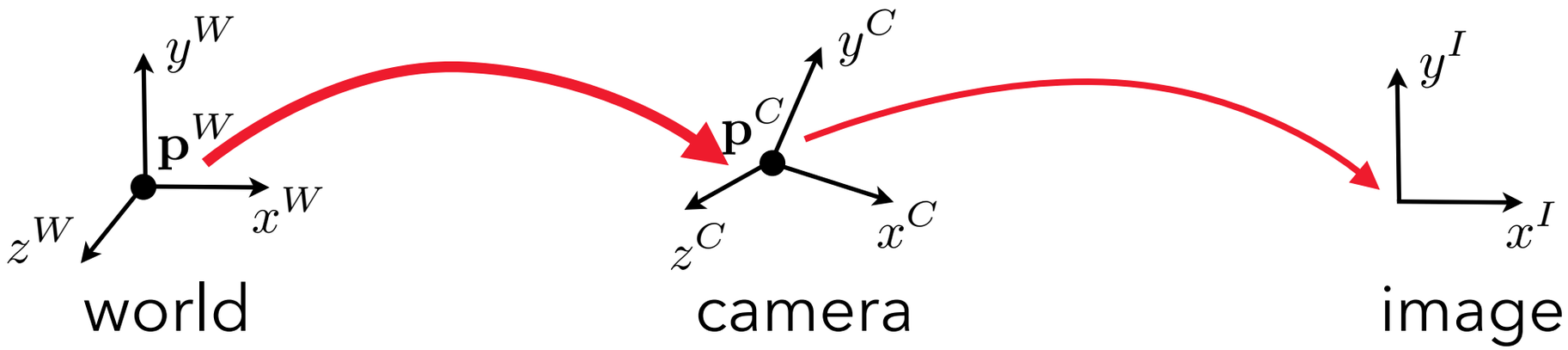
Coordinate Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

- Extrinsic camera transformation:
  - Described as a linear transformation
  - Using homogeneous coordinates
    - Augment vector with a constant 1: $\mathbf{x} = (x_1, x_2, x_3, 1)^{\mathrm{T}}$
  - Combination of rotation and translation:

$$(\mathbf{R}, \mathbf{t}) \in \mathbb{R}^{4\times4}$$

$$\mathbf{x}^C = (\mathbf{R}, \mathbf{t}) \cdot \mathbf{x}^W$$



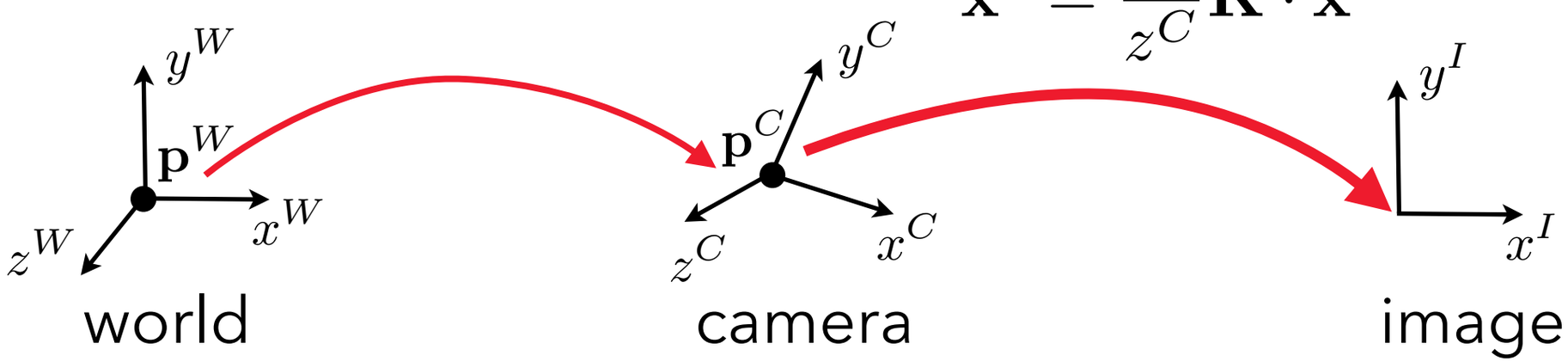
Coordinate Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

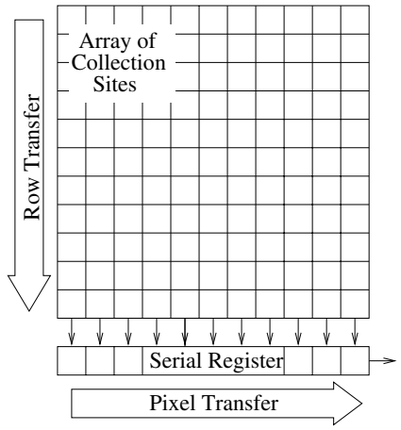
- Intrinsic camera transformation:
  - Usually described as a linear transformation + perspective division
  - Pinhole camera model is a special case
  - Can deal with more general camera models as well

$$\mathbf{K} \in \mathbb{R}^{4\times 4}$$

$$\mathbf{x}^I = \frac{1}{z^C}\mathbf{K}\cdot\mathbf{x}^C$$



Images



- Images arriving at our CCD or CMOS sensor are spatially discrete with individual pixels.
- But is the visual world spatially discrete? No.
  - Light hits the sensor everywhere.
  - Spatially continuous intensity function:

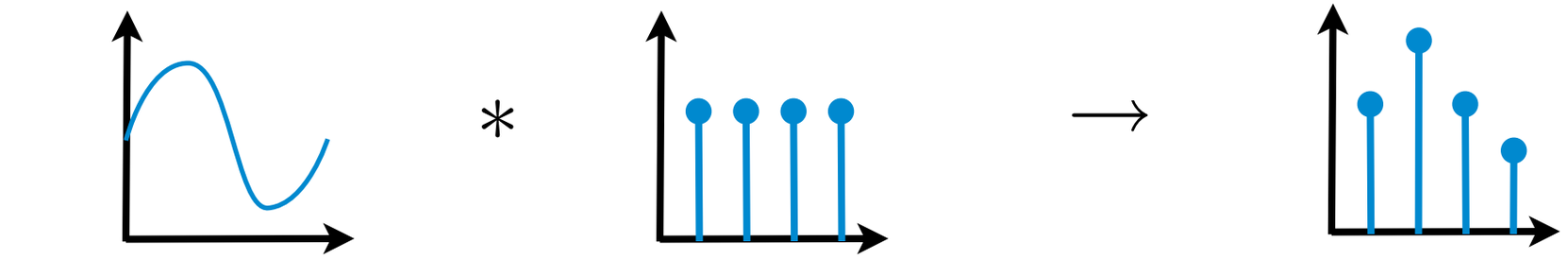
$$I(x, y), \quad I : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$$

- The image sensor performs a "sampling" of this function.
  - Turns it into a discrete array of intensity values.

Images

TECHNISCHE UNIVERSITÄT DARMSTADT

- Idealized spatial sampling (1D analogy)



continuous signal sampling operator sampled image

  - Details: Signal processing, Fourier theory

- For our purposes here, we will regard the image as spatially discrete.
  - Array of pixels

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

Readings for next week

TECHNISCHE UNIVERSITÄT DARMSTADT

- Introduction (Ch. 1)
  - What is computer vision? (Sec. 1.1)
  - For the interested: History of computer vision (Sec. 1.2)
- Image formation (Ch. 2)