2D Vision and Deep Learning

Introduction

Prof. Dr. Ulrich Schwanecke

RheinMain University of Applied Sciences

CVMR

Computer Vision & Mixed Reality Group

How to use the HTML slides

- All materials can be found here
  - usr: 2DCV
  - pwd: ws2324

- Use the keys **left**/**right** for navigating through the slides.
- Click icon ☰ (top left) to open the navigation menu.
- Press **f**/**ESC** to enter/leave fullscreen mode.
- **Double-click** an item (e.g. an image) to zoom in/out.
- If the bottom boundary flashes on slide change, something was written on the virtual whiteboard.
  - **Scroll down** to see it.

About myself

Who am I?

- Born in Darmstadt
- Grown up in Wiesbaden
- JoGu Mainz
- TU Darmstadt
- MPI Informatik, Saarbrücken
- Daimler Chrysler Research, Ulm
- RheinMain University of Applied Sciences, Wiesbaden

About this course

Course Goal and Content

- **Goal**
  - Gain an understanding of the theoretical and practical concepts of computer vision
    - Focus on 2D vision
  - After this course, you should be able to
    - develop and train computer vision models
    - repoduce results and
    - conduct original research

- **(Planned) Content**
  1. Introduction, Organization
  2. Primitives, Transformations, Geometric Image Formation
  3. Photometric Image Formation, Image Sensing Pipeline
  4. Image Filtering
  5. Orthogonal Basis Transformation (Fourier)
  6. Features
  7. Motion
  8. Introduction to Machine Learning, Neural Networks
  9. Transfer Learning for Image Classification
  10. Object Detection
  11. Image Segmentation
  12. Image Manipulation

Organization

- **SWS 2V + 2Ü, 6 ECTS, Total Workload: 180h**
- **Lecture**
  - Monday, 14:15-15:45, 3-428
  - 2023: Oct. 24/30, Nov. 06/13/20/27, Dec. 04/11/18, 2023: Jan. 08/15/22/29, Feb. 05
- **Exercise Sessions**
  - Monday, 8:15-9:45, 4-432. *Submission each Wednesday until 16:00 via Moodle*
  - Exercises are **mandatory**
- **Exam**
  - Content: lectures and exercises
  - Very likely written (date and time will be announced)
  - To qualify for the exam you have to
    - have $\geq 50\%$ of all achievable points ($\geq 25\%$ for each problem set)
    - present at least one assignment

Course Materials

- **Books**
  - R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer 2011
    https://szeliski.org/Book
  - I. Gooldfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press 2016
    https://www.deeplearningbook.org
  - J. E. Solem, *Programming Computer Vision with Python*, O'Reilly 2012
  - V. K. Ayyadevara, Y. Reddy, *Modern Computer Vision with PyTorch*, Packt 2020
  - M. P. Deisenroth, A. A. Faisal, C. S. Ong, *Mathematics for Machine Learning*
    https://mml-book.github.io
  - K. B. Petersen, M. S. Pedersen, *The Matrix Cookbook*
    http://www.cs.toronto.edu/~bonner/courses/2012s/csc338/matrix_cookbook.pdf

Course Materials

- **Tutorials**
  - The Python Tutorial: https://docs.python.org/3/tutorial
  - Numpy Quickstart: https://numpy.org/devdocs/user/quickstart.html
  - PyTorch Tutorial: https://pytorch.org/tutorials
- **Frameworks, IDEs**
  - Visual Studio Code: https://code.visualstudio.com/
  - Google Colab: https://colab.research.google.com
- **Courses**
  - Slide deck covering *Szeliski's book* https://szeliski.org/Book
  - I. Gkioulekas, *Computer Vision* https://www.cs.cmu.edu/~16385/
  - A. Owens, *Foundations of Computer Vision* https://web.eecs.umich.edu/~ahowens/eecs504/w20/

Prerequisites

- Basic math skills
    - Linear Algebra, Calculus, Probability
- Basis computer science skills
    - Variables, functions, loops, classes, algoritms
- Basic Python coding skills
    - https://docs.python.org/3/turorial/
- Basic PyTorch coding skills
    - https://pytorch.org/turorials

Prerequisites

- **Linear Algebra**
  - Vectors: $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
  - Matrices: $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$
  - Operations:
    - $\mathbf{x}^\top \mathbf{y}, \mathbf{x} \times \mathbf{y}$
    - $\mathbf{A}\mathbf{x}$
    - $\mathbf{A}^\top, \mathbf{A}^{-1}, \text{trace}(\mathbf{A}), \det(\mathbf{A}), \mathbf{A} + \mathbf{B}, \mathbf{AB}$
  - Norms: $||\mathbf{x}||_1, ||\mathbf{x}||_2, ||\mathbf{x}||_\infty, ||\mathbf{A}||_F$
  - Eigenvalues, Eigenvectors, SVD: $\mathbf{A} = \mathbf{UDV}^\top$
- **Calculus**
  - Multivariate functions: $f : \mathbb{R}^n \to \mathbb{R}$
  - Partial derivatives: $\frac{\partial f}{\partial x_i}, i = 1, \ldots, n$, Gradient
  - Integrals: $\int f(x)dx$
- **Probability**
  - Probability distributions: $P(X = x)$
  - Expectation: $\mathbb{E}_{x \sim p}[f(x)] = \int_x p(x) f(x) dx$
  - Variance: $\text{Var}(f(x)) = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$
  - Marginal: $p(x) = \int p(x, y) dy$
  - Conditional: $p(x, y) = p(x|y)p(y)$
  - Bayes rule: $p(x|y) = p(y|x)/p(y)$
  - Distributions: Uniform, Gaussian

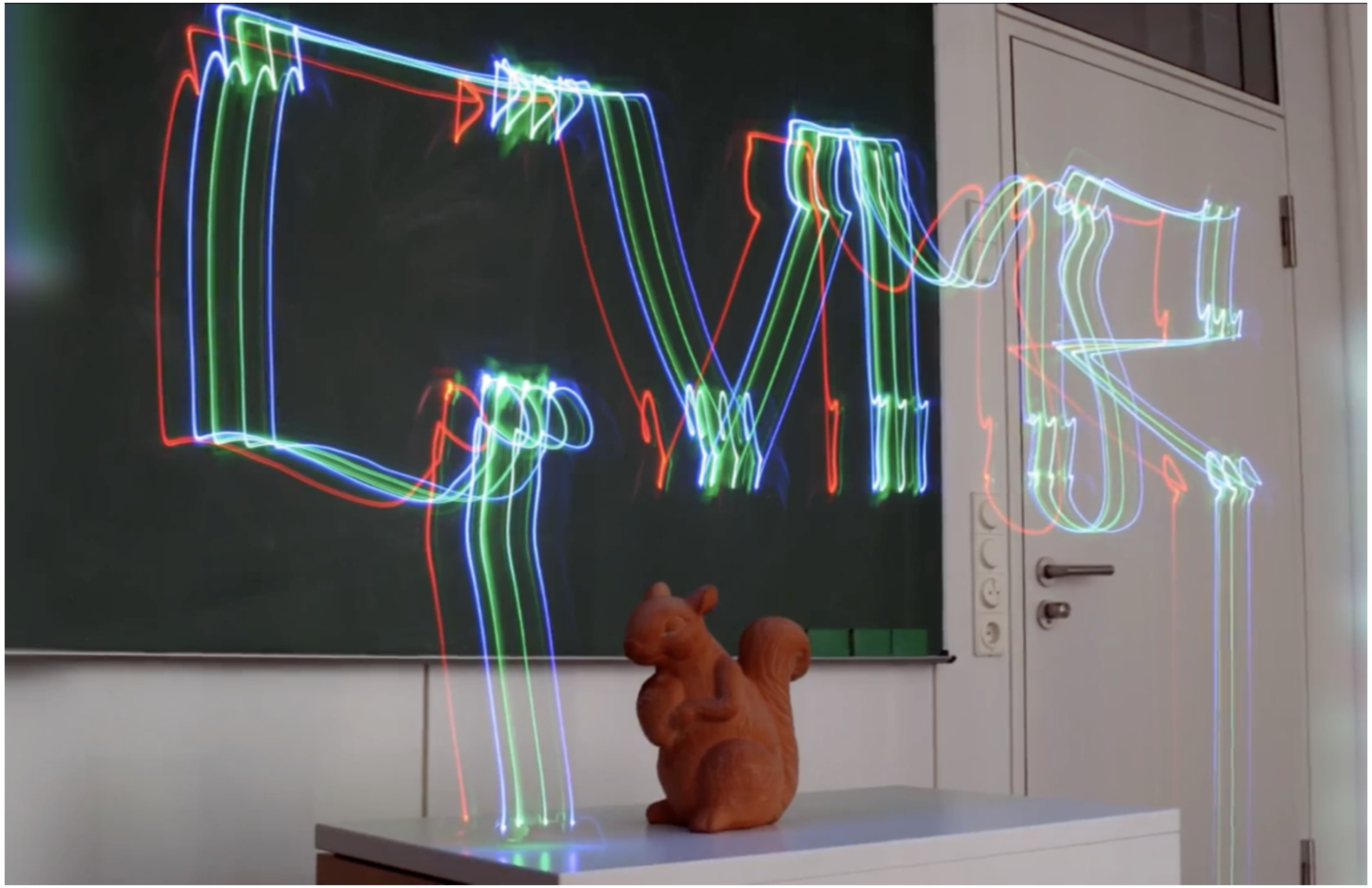
Time Management

Activity	Times	Total
Attending (watching) the lecture	2h / week	24h
Self-study of lecture materials	2h / week	24h
Participation in exercise	2h / week	24h
Solving the assignments	6h / week	72h
Preparation for the final exam	36h	36h
Total workload		**180h**

About Computer Vision

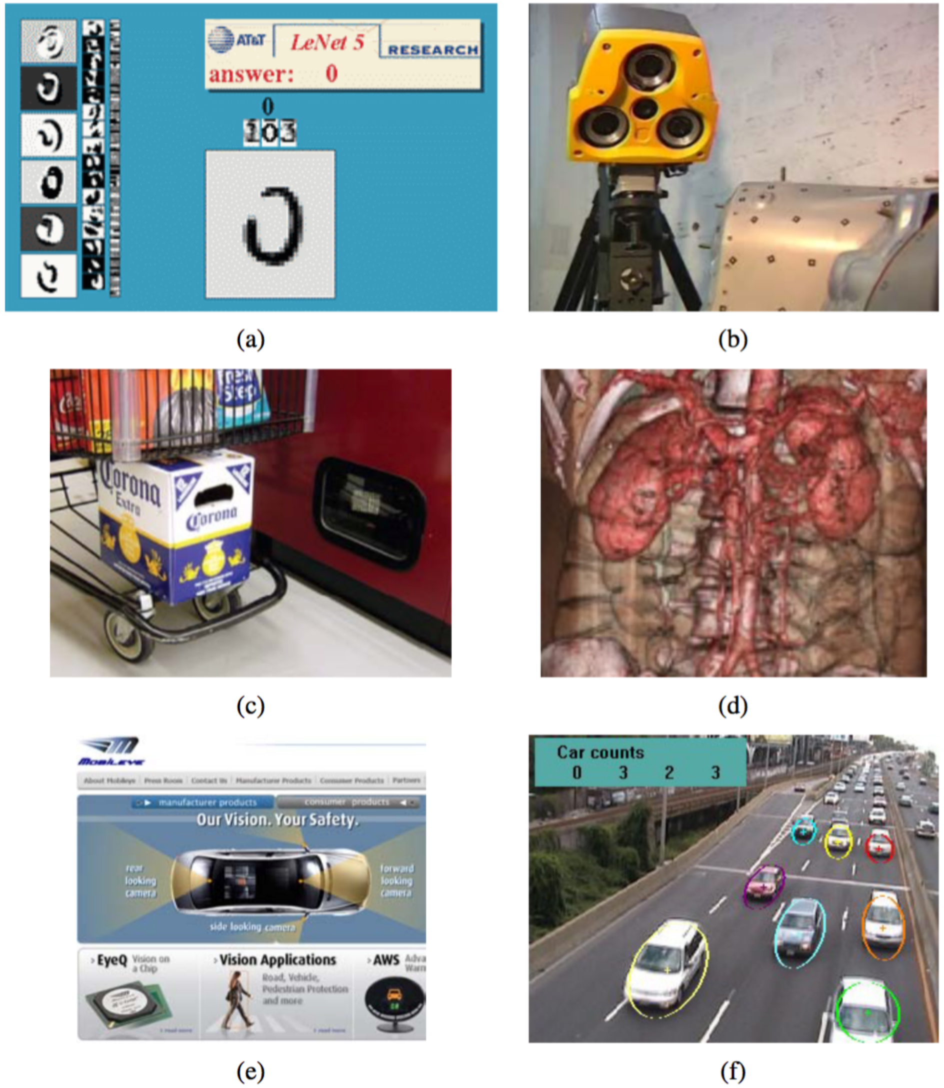
Computer Vision

- Goal of Computer Vision is to **convert light into meaning** (geometric, semantic, ...)



Computer Vision Applications

- Optical Character Recognition (a)
- Mechanical Inspection / 3D Modelling (b)
- Retail (c)
- Medical Applications (d)
- Automotive (Savety and Driving) (e)
- Surveillance (f)



(a) (b) (c) (d) (e) (f)

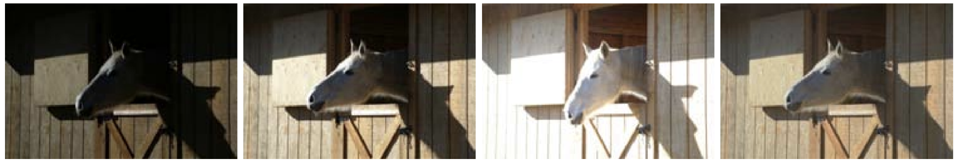
[R. Szelisky ©]

Computer Vision Applications

- Image Stitching / Video Stabilization
- Exposure Bracketing
- Robotics
- Mobile Devices
- Accessibility (e.g. Image Captioning), …

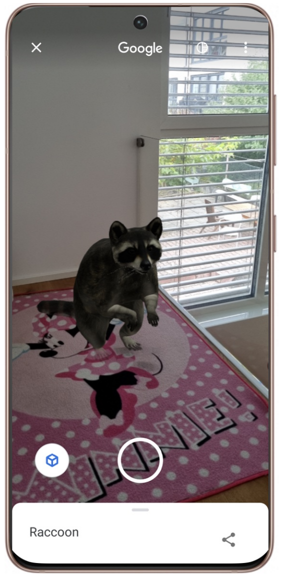
“A bird that is sitting on a branch”

[R. Szelisky ©]



[R. Szelisky ©]

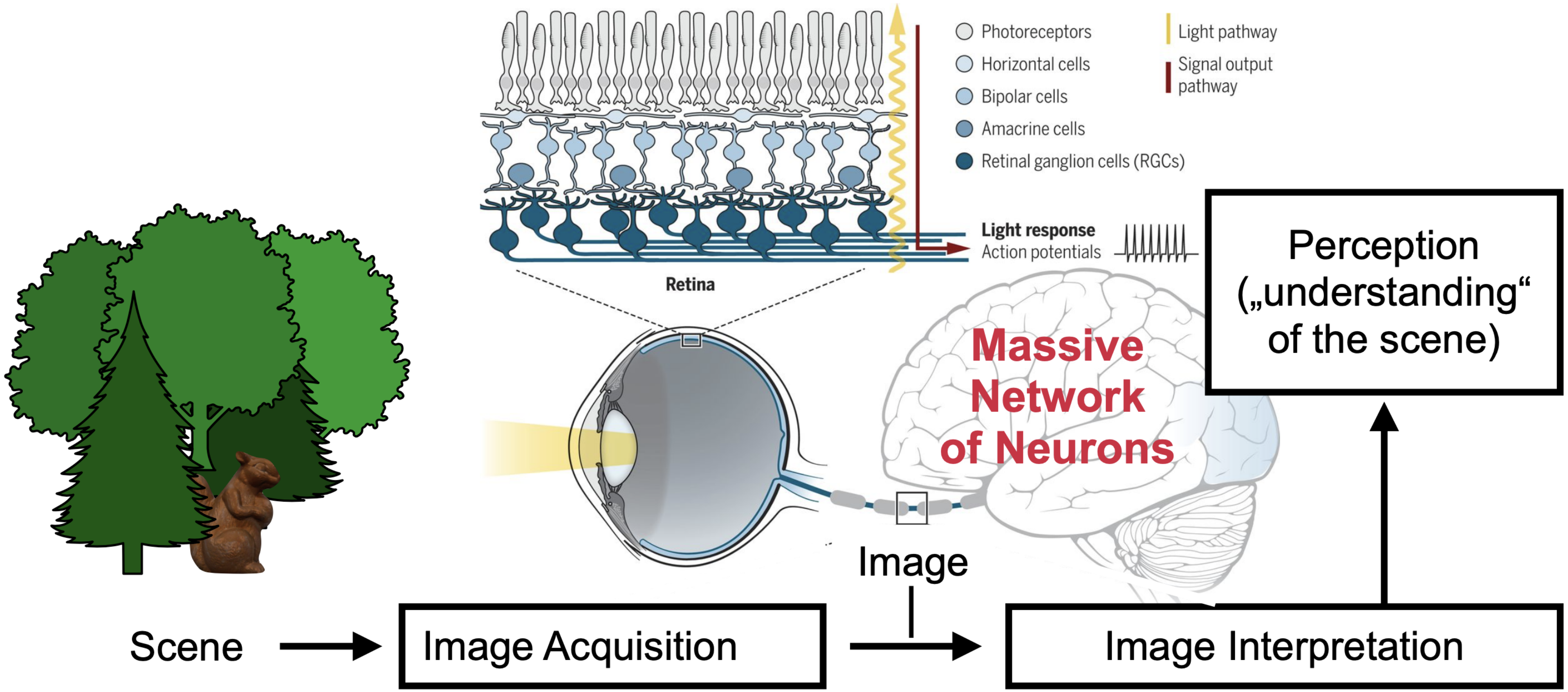
[quadruped.de ©]



Mobile AR

Biological Vision vs. Computer Vision

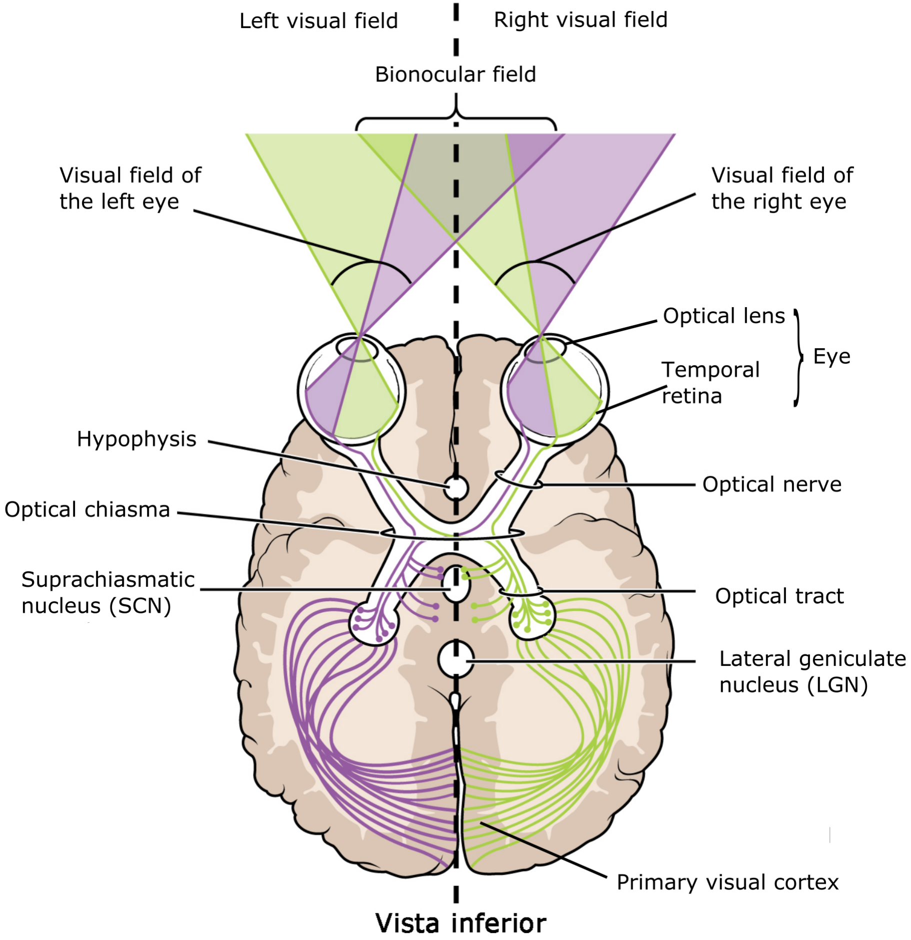
- Human Vision is the process of discovering what is present in the world and where it is by looking



[Adapted from K. Sutliff/Science ©]

Biological Vision vs. Computer Vision

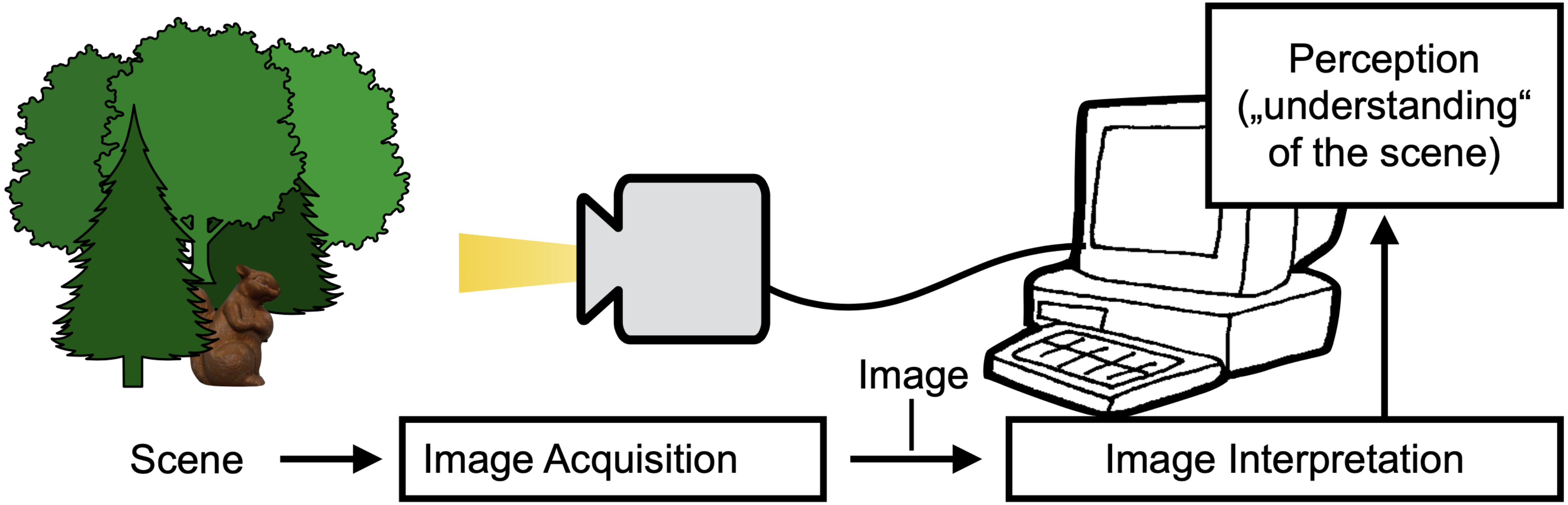
- Over **50%** of the processing in the **human brain** is dedicated to **visual information**



[OpenStax College ©]

Biological Vision vs. Computer Vision

- Computer Vision is the study of analyzing images to achieve results similar to those as by humans



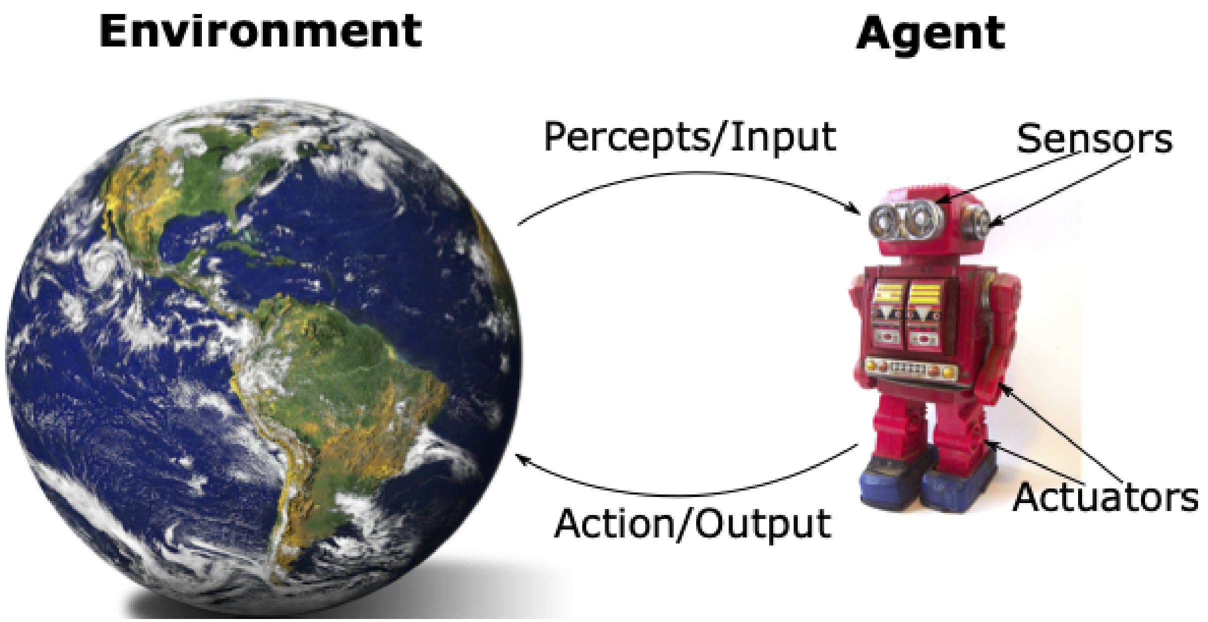
[Adapted from K. Sutliff/Science ©]

Artificial Intelligence

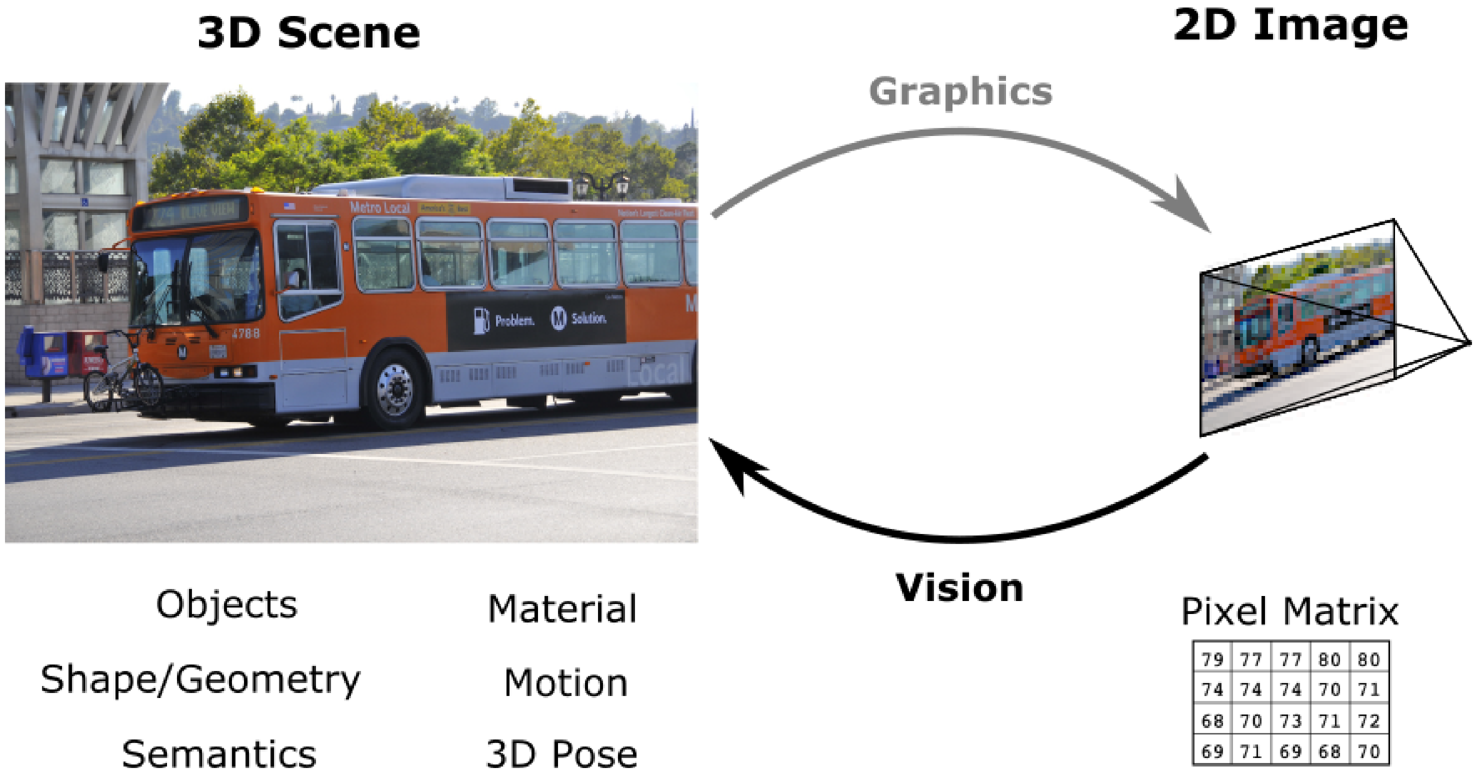
"An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves"

[John McCarthy at *Dartmouth Summer Research Project on Artificial Intelligence*, 1956]

- Machine Learning
- Computer Vision
- Computer Graphics
- Natural Language Processing
- Robotics & Control
- Art, Industry 4.0, Education, …



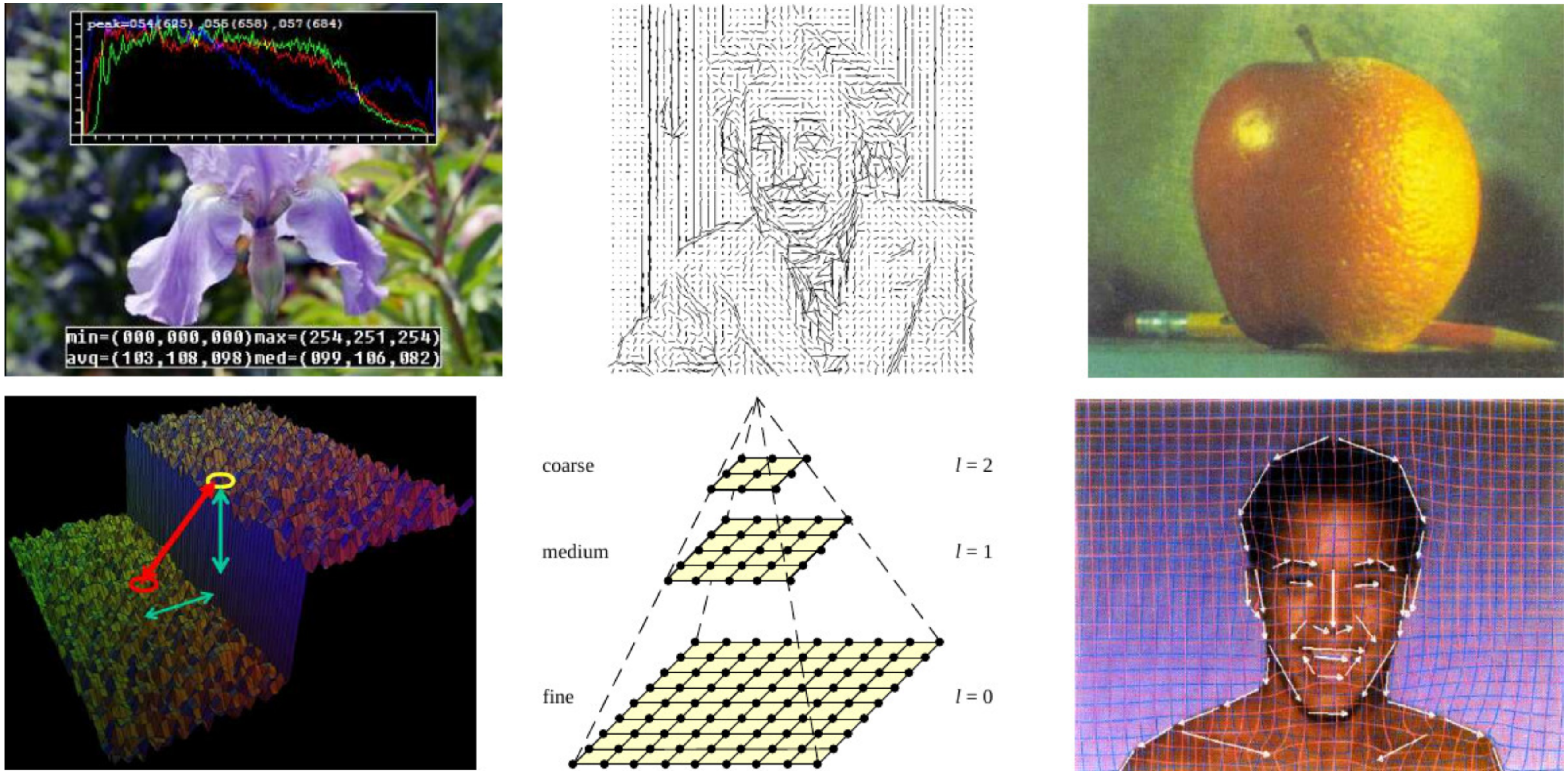
Computer Vision vs. Computer Graphics



- **Computer Vision is an ill-posed inverse problem**
  - Many 3D scenes yield the same 2D image
  - Additional constraints (knowledge about world) are required

Computer Vision vs. Image Processing

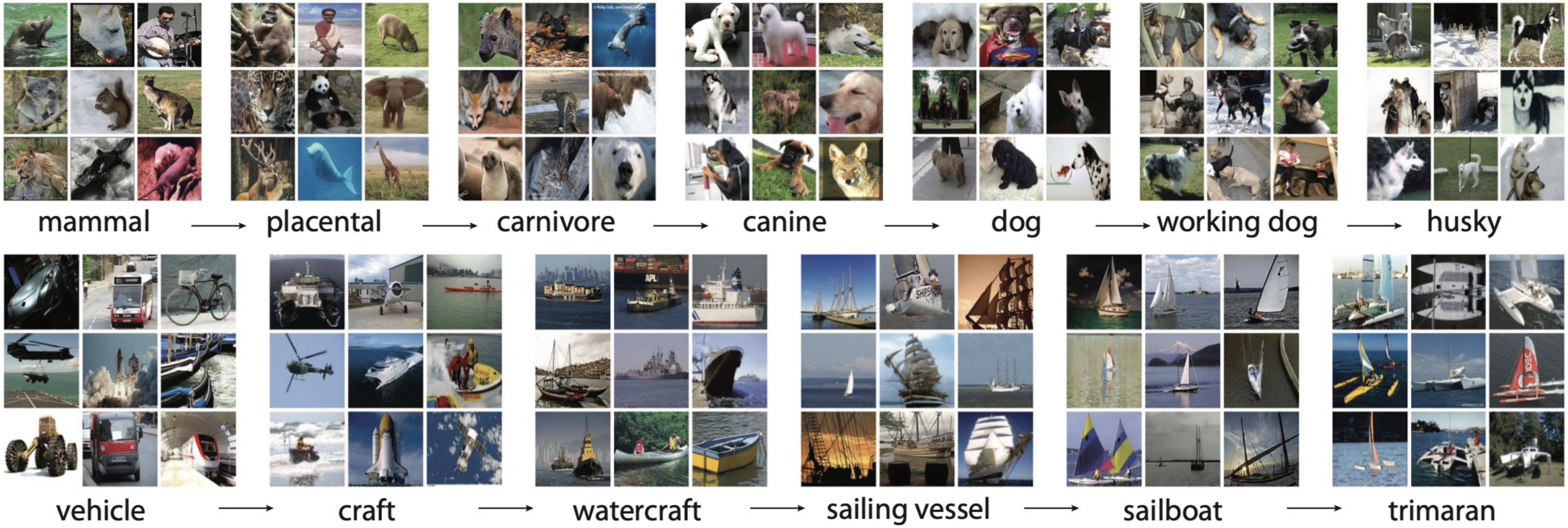
- Computer Vision seeks to achieve **full scene understanding** (in contrast to (classical) Image Processing)



[R. Szelisky ©]

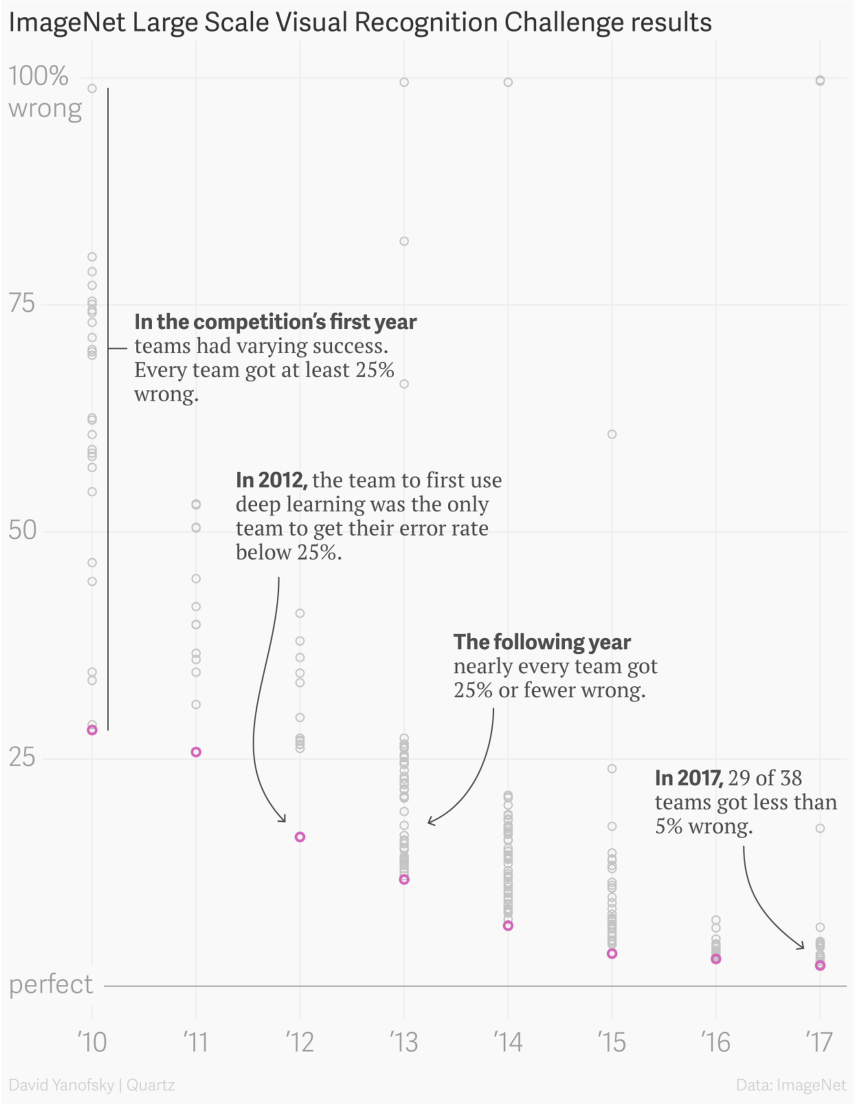
Computer Vision and Machine Learning

- **ImageNet** https://www.image-net.org/



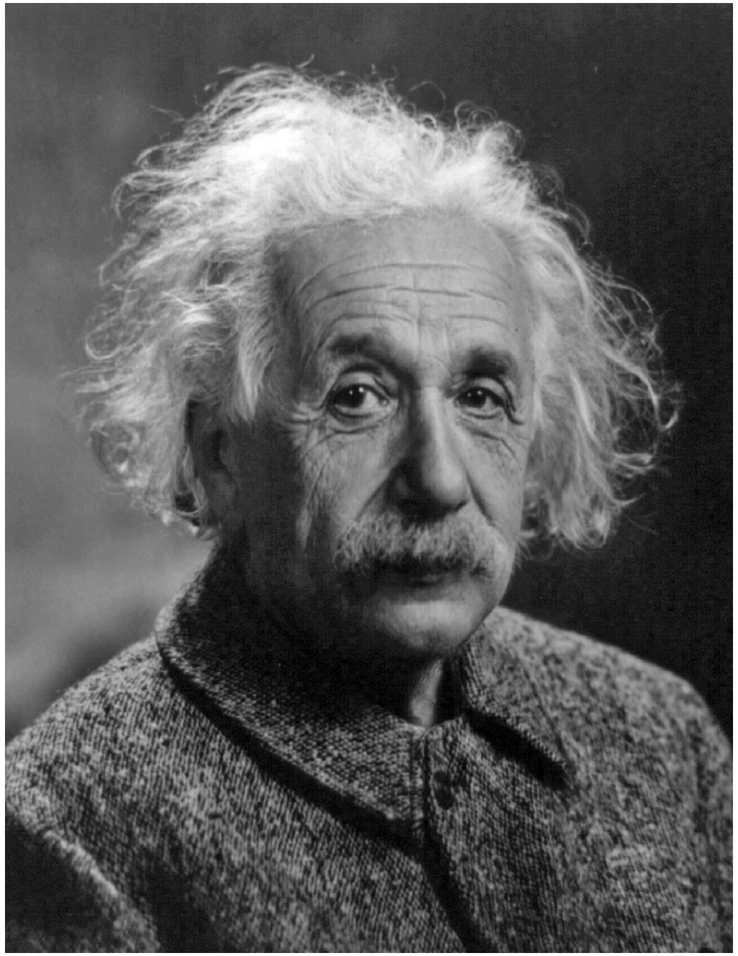
[https://image-net.org/static_files/papers/imagenet_cvpr09.pdf]

The Deep Learning Revolution



[https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/ ©]

Why is Visual Perception hard?

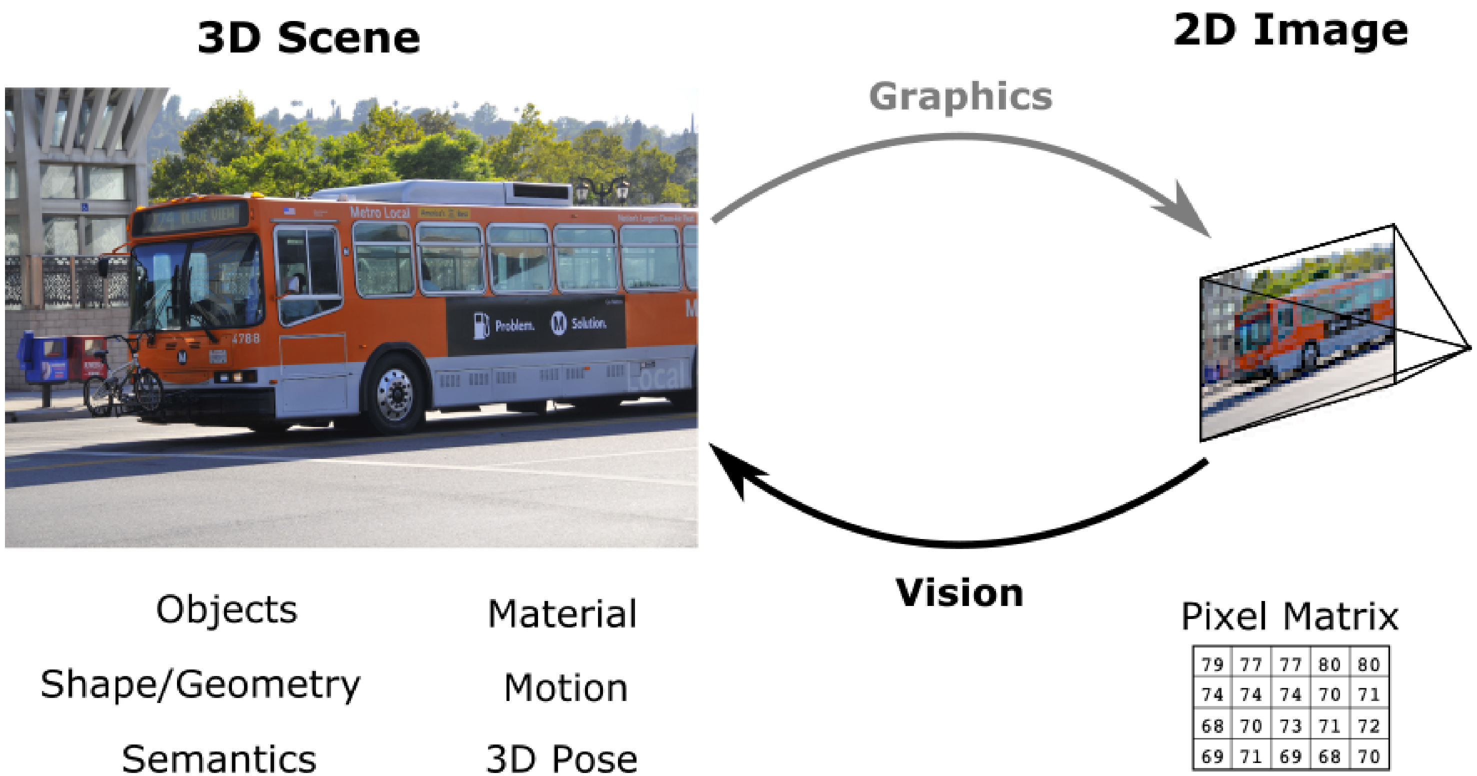


What we see

80	77	80	79	78	80	79	80	79	79	75
79	79	82	82	80	78	77	81	81	79	76
79	77	80	79	77	77	80	80	74	75	69
77	78	77	74	74	74	70	71	73	72	65
70	68	71	68	70	73	71	72	69	73	62
71	73	72	69	71	69	68	70	71	73	59
75	75	73	72	76	76	74	76	74	74	59
76	75	75	73	74	75	72	71	71	69	54
67	66	65	67	67	67	69	67	68	68	53
65	64	62	63	62	61	64	67	69	69	52
68	69	70	70	71	73	72	73	71	70	56
70	72	69	70	70	69	70	71	70	70	55

What the computer sees

Why is Visual Perception hard?

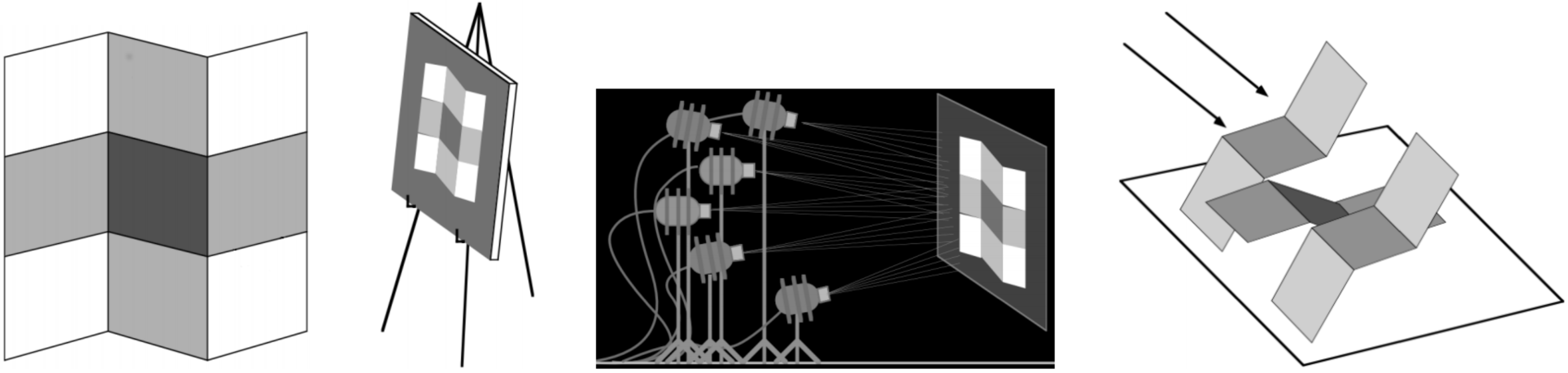


- **Image are 2D Projections of the 3D World**
  - Many 3D scenes yield the same 2D image
  - Additional constraints (knowledge about world) are required

Images are 2D Projections of the 3D World

Adelson and Pentland's workshop metaphor:

- To explain an *image* (a) in terms of reflectance, lighting and shape, a *painter* (b), a *light designer* (c) and a *sculptor* (d) will design three different, but plausible, solutions.



(a) (b) (c) (d)

E. H. Adelson, A. P. Pentland: The perception of shading and reflectance, 1996. D. C. Knill: Perception as Bayesian inference, 1996

Images are 2D Projections of the 3D World

Perspective Illusion:

Images are 2D Projections of the 3D World

Perspective Illusion (Ames Room)

Challenges: Occlusion

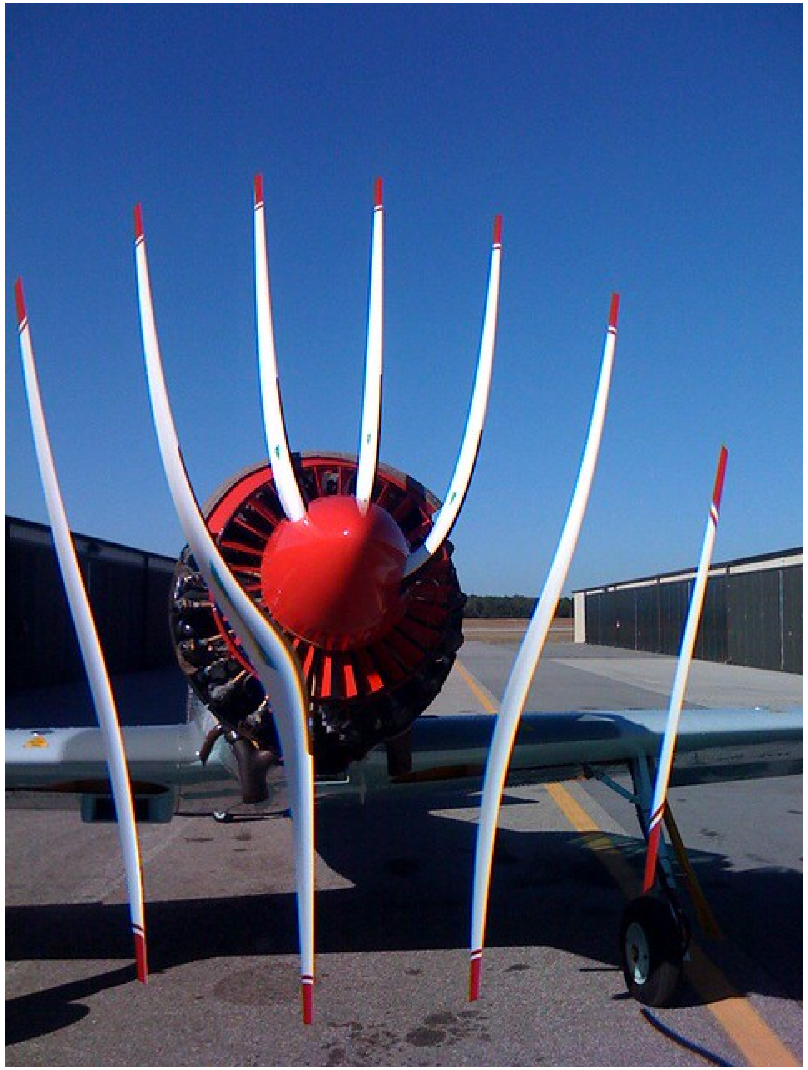
[https://imgur.com/a/nQJss ©]

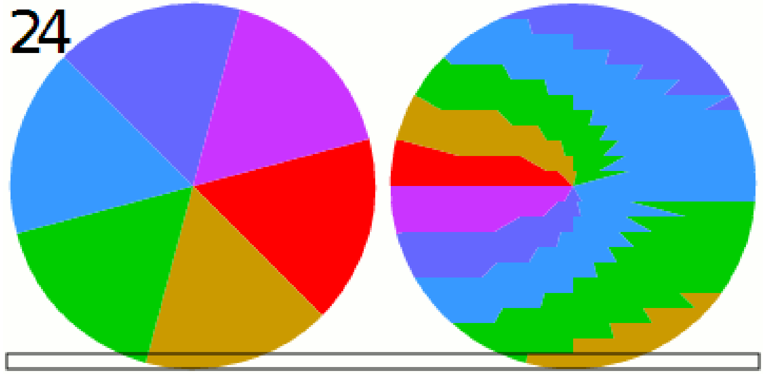
Challenges: Illumination

Challenges: Motion

[https://commons.wikimedia.org/wiki/File:Heliopsis_helianthoides_var._scabra_Summer_Sun_4zz.jpg#/media/File:Heliopsis_helianthoides_var._scabra_Summer_Sun_4zz.jpg]

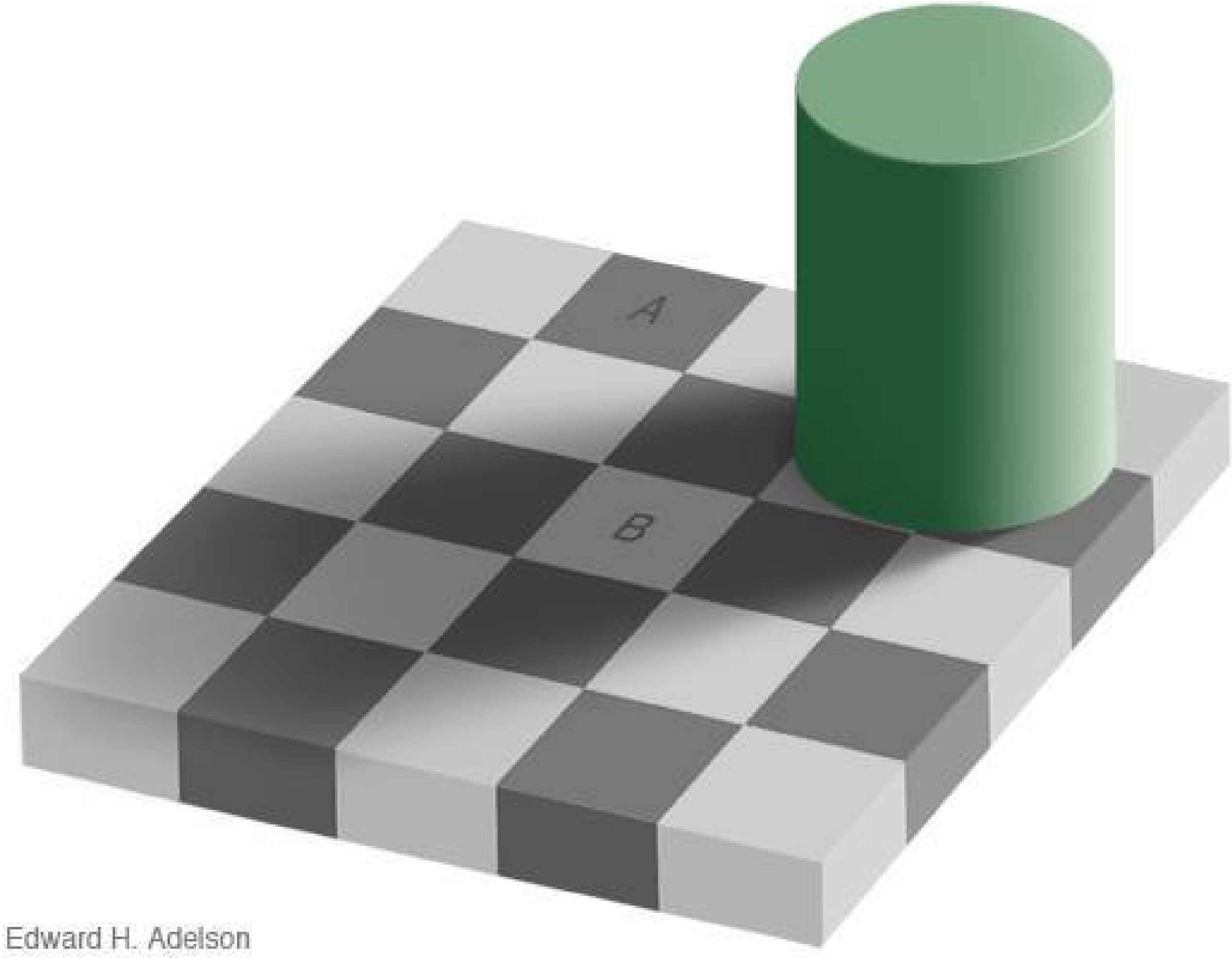
Challenges: Motion


[https://commons.wikimedia.org/wiki/File:Rolling_shutter_näidis.png]



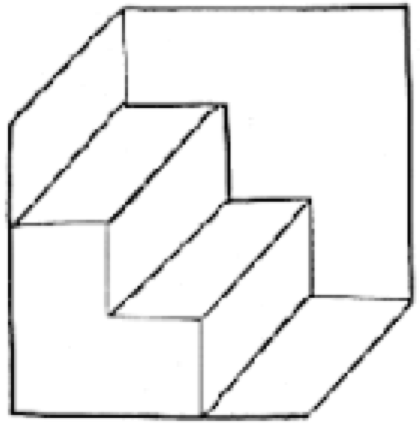
[https://commons.wikimedia.org/wiki/File:Rolling_shutter_effect.svg]

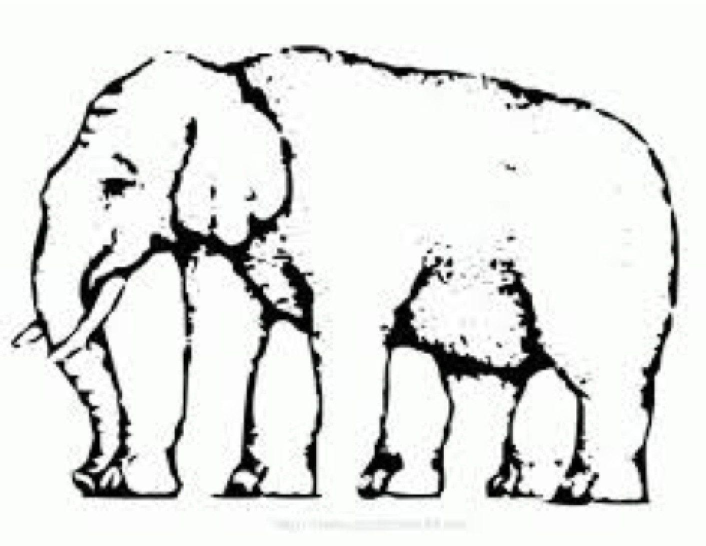
Challenges: Perception vs. Measurement

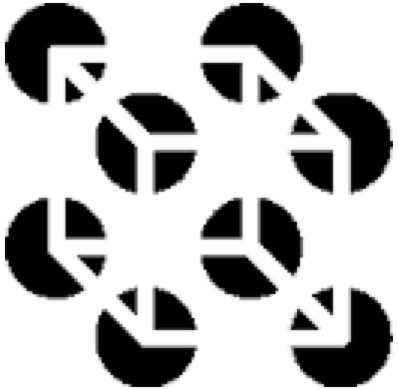


[http://persci.mit.edu/gallery/checkershadow]

Challenges: Perception vs. Measurement



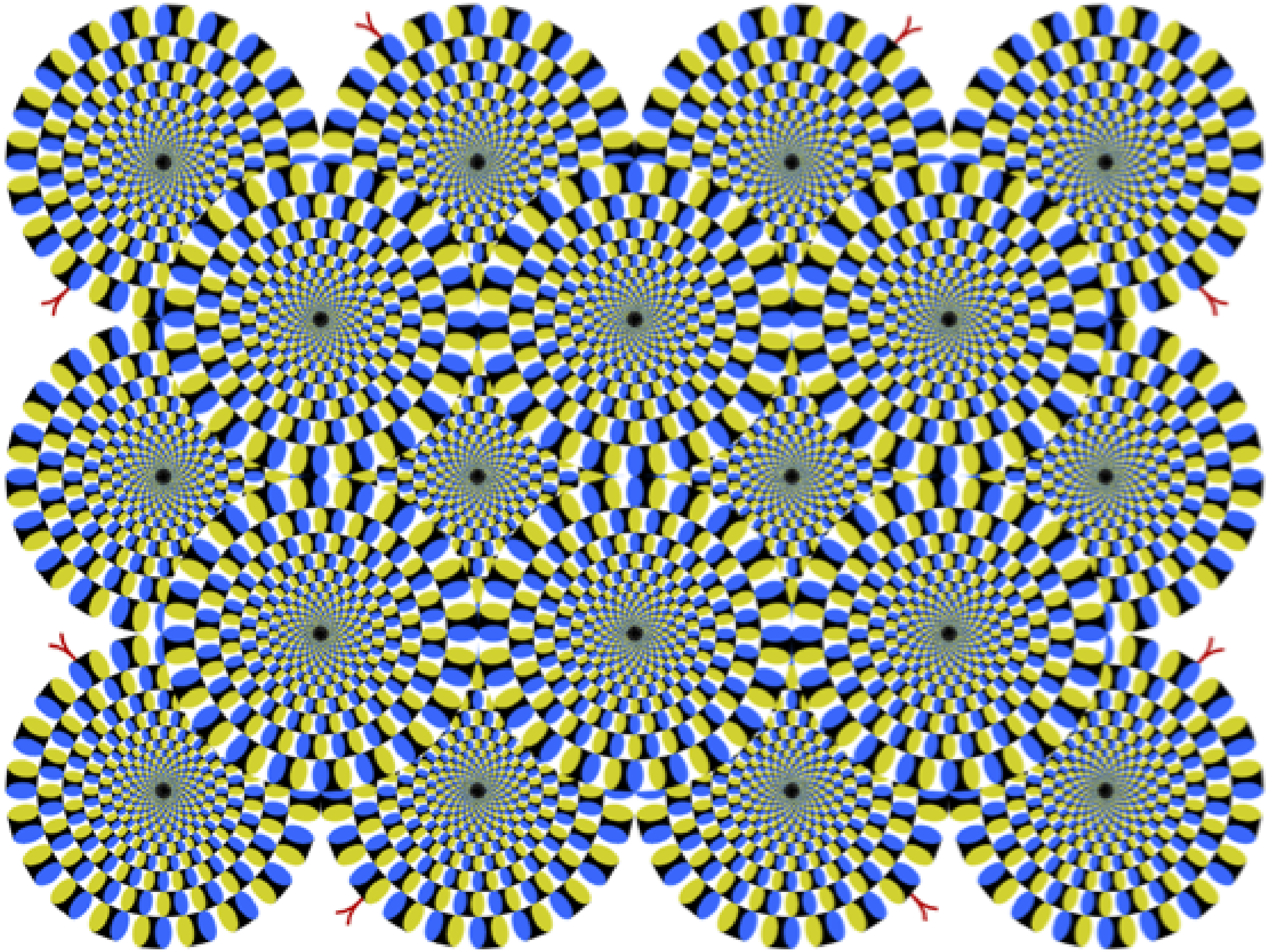




Challenges: Perception vs. Measurement

Challenges: Perception vs. Measurement

Challenges: Perception vs. Measurement

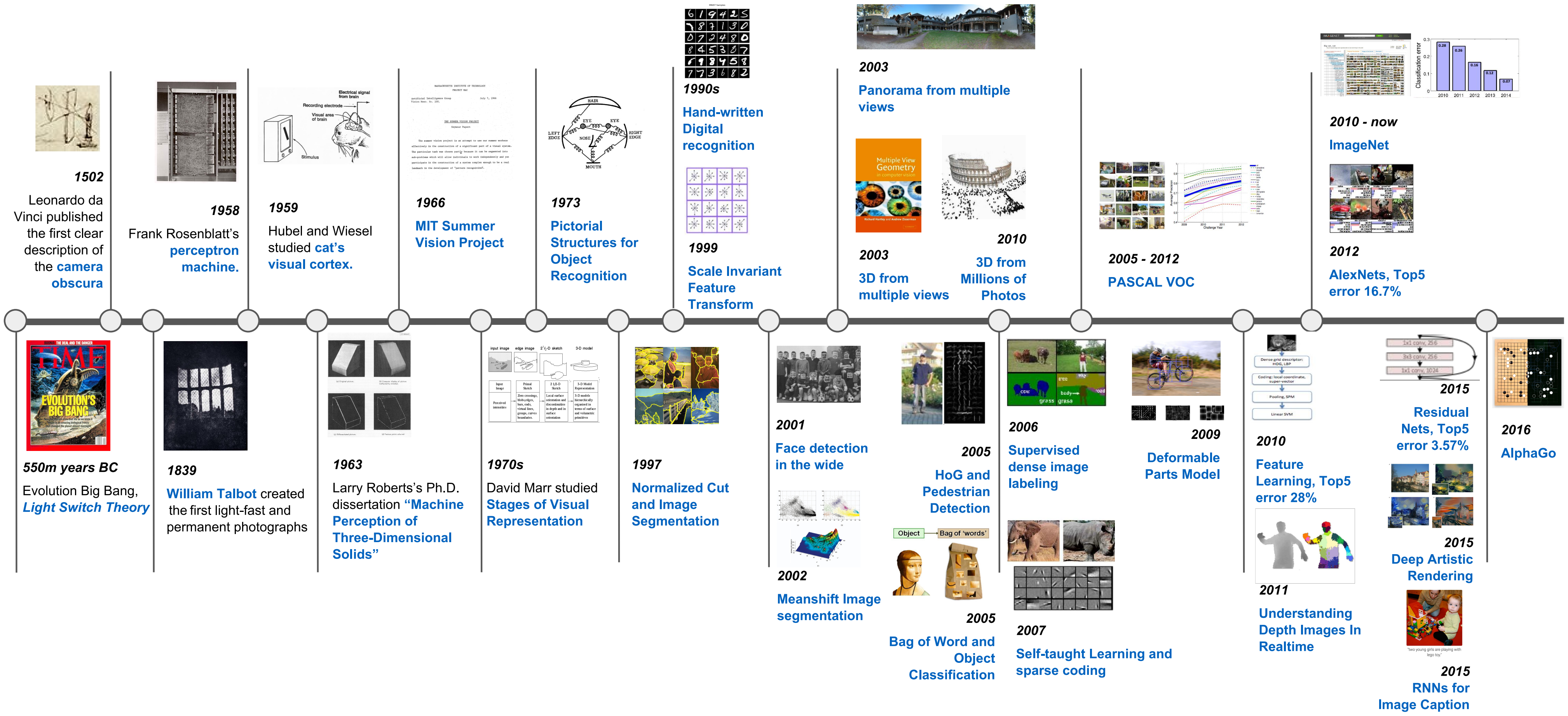


Rotation Snakes by Kitaoka Akiyoshi http://www.ritsumei.ac.jp/~akitaoka/index-e.html

Challenges: Deformation and Intra Class Variation

[M. Aubry, D. Maturana, A. Efros, B. Russel and J.Sivic, Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models]

Timeline of Computer Vision



550m years BC
Evolution Big Bang, Light Switch Theory
1502
Leonardo da Vinci published the first clear description of the camera obscura
1839
William Talbot created the first light-fast and permanent photographs
1958
Frank Rosenblatt's perceptron machine.
1959
Hubel and Wiesel studied cat's visual cortex.
Electrical signal from brain
Recording electrode
Visual area of brain
Stimulus
1963
Larry Roberts's Ph.D. dissertation "Machine Perception of Three-Dimensional Solids"
1966
MIT Summer Vision Project
THE SUMMER VISION PROJECT
1970s
David Marr studied Stages of Visual Representation
Input image
edge image
2½-D sketch
3-D model
1973
Pictorial Structures for Object Recognition
HAIR
EYE
EYE
LEFT EDGE
RIGHT EDGE
NOSE
MOUTH
1990s
Hand-written Digital recognition
1997
Normalized Cut and Image Segmentation
1999
Scale Invariant Feature Transform
2001
Face detection in the wide
2002
Meanshift Image segmentation
2003
Panorama from multiple views
2003
3D from multiple views
Multiple View Geometry in computer vision
2005
HoG and Pedestrian Detection
2005
Bag of Word and Object Classification
Object
Bag of 'words'
2010
3D from Millions of Photos
2005 - 2012
PASCAL VOC
2006
Supervised dense image labeling
cow
grass
tree
sky
body
road
2007
Self-taught Learning and sparse coding
2009
Deformable Parts Model
2010
Feature Learning, Top5 error 28%
Pooling, SPM
Linear SVM
2010 - now
ImageNet
Classification error
0.28
0.26
0.16
0.12
0.07
2010
2011
2012
2013
2014
2011
Understanding Depth Images In Realtime
2012
AlexNets, Top5 error 16.7%
2015
Residual Nets, Top5 error 3.57%
1x1 conv, 256
3x3 conv, 256
1x1 conv, 1024
2015
Deep Artistic Rendering
2015
RNNs for Image Caption
2016
AlphaGo

Next Lecture

- Primitives
  - Points, Lines and Planes
  - Homogeneous Coordinates
- Transformations
  - 2D / 3D Transformations
  - Homography Estimation
- Geometric Image Formation
  - Pinhole Camera
  - Projection Models
  - Lens Distortion