Introduction to Computer Vision – Chapter 1

© Yang Cao

© Chang Wen Chen

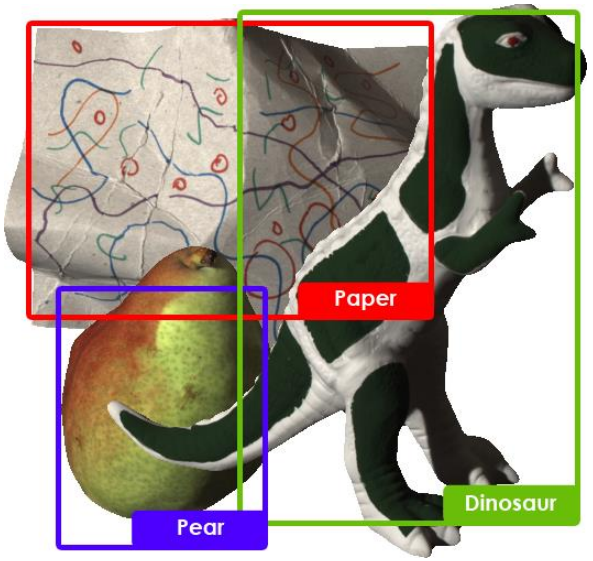
University of Science and Technology of China

What is Computer Vision?

- **Trucco and Verri**
  - computing properties of the 3D world from one or more digital images
- **Stockman and Shapiro**
  - To make useful decisions about real physical objects and scenes based on sensed images
- **Ballard and Brown**
  - The construction of explicit, meaningful description of physical objects from images
- **Forsyth and Ponce**
  - To extract descriptions of the world from pictures or sequences of pictures

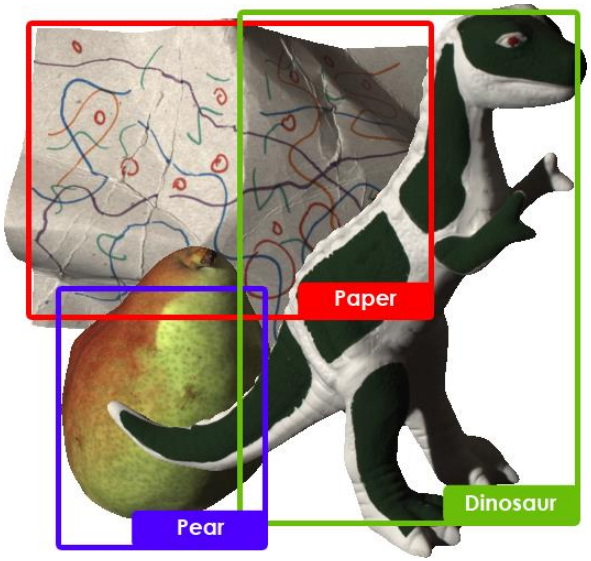
Computer Vision: A Taxonomy

Jonathan Barron and Jitendra Malik, Shape, Illumination, and Reflectance from Shading, TPAMI, 2015

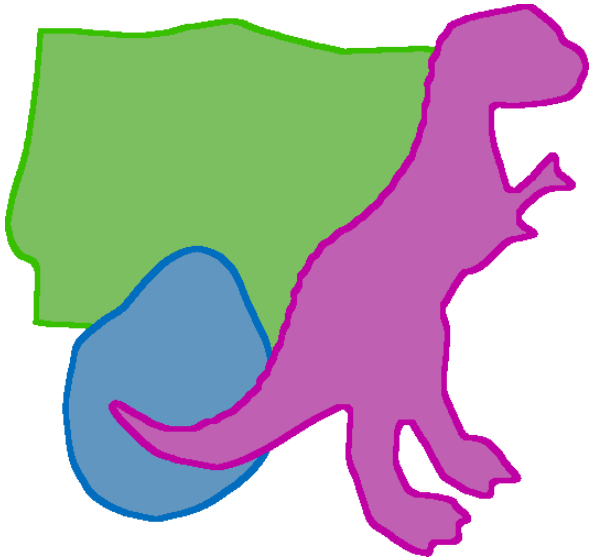


Recognition

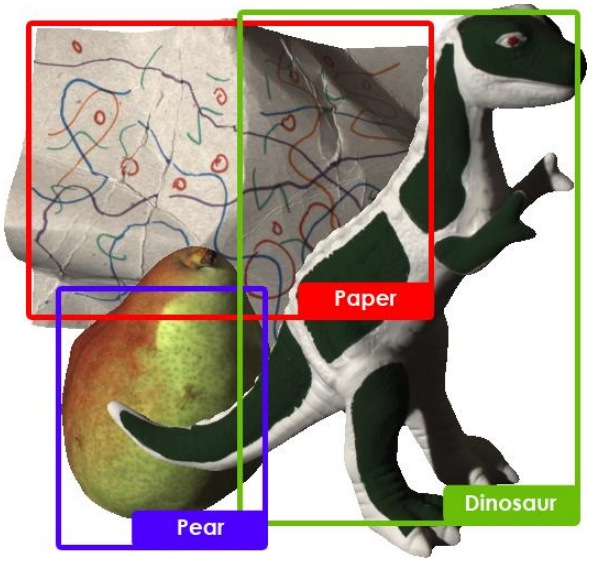
Jonathan Barron and Jitendra Malik, Shape, Illumination, and Reflectance from Shading, TPAMI, 2015



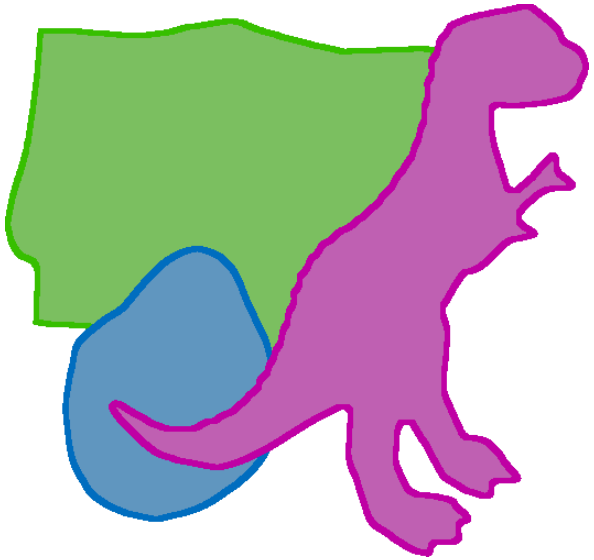
Recognition



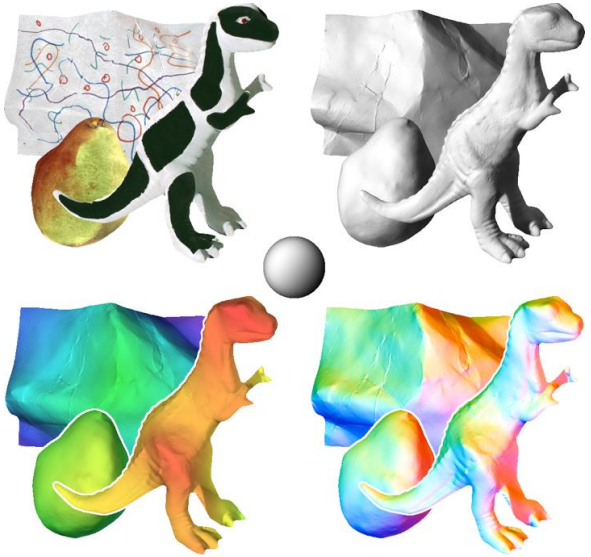
Reorganization



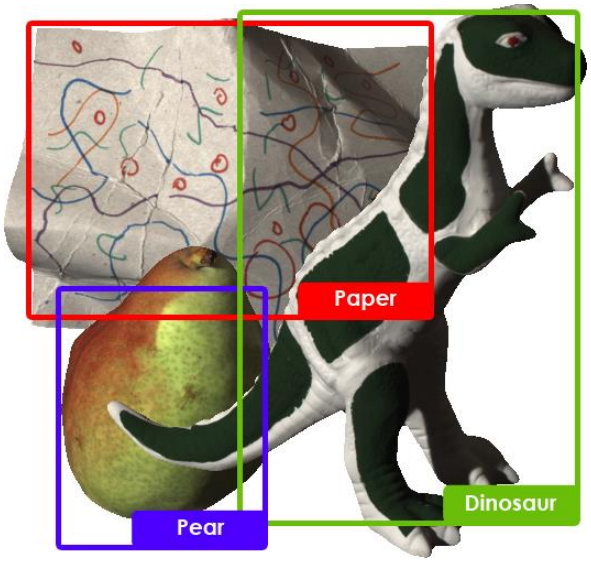
Recognition



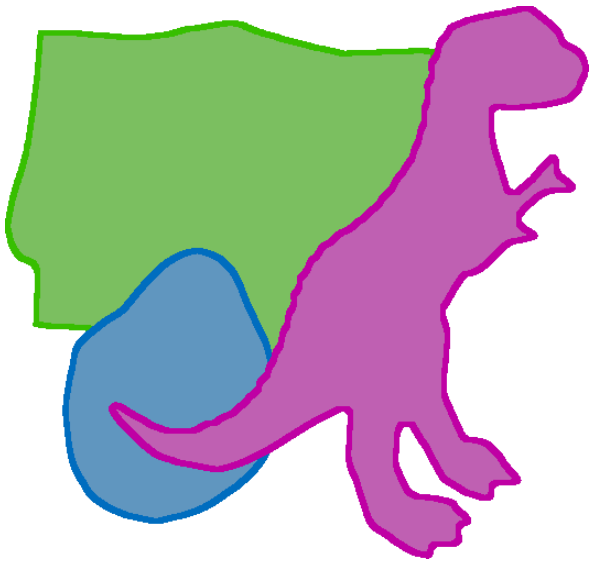
Reorganization



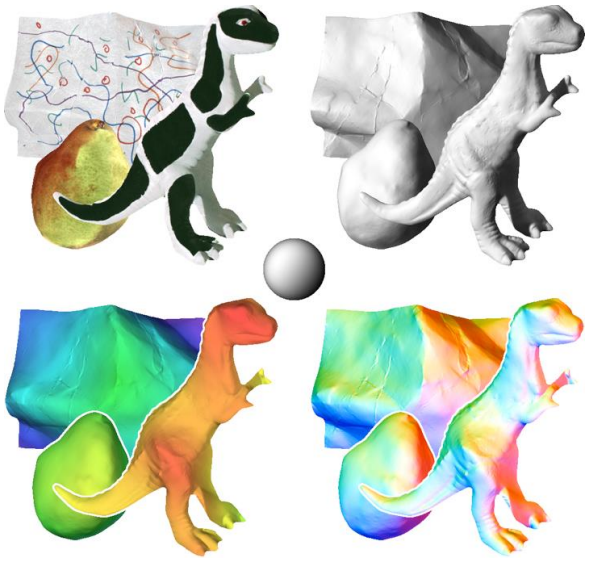
Reconstruction



Recognition

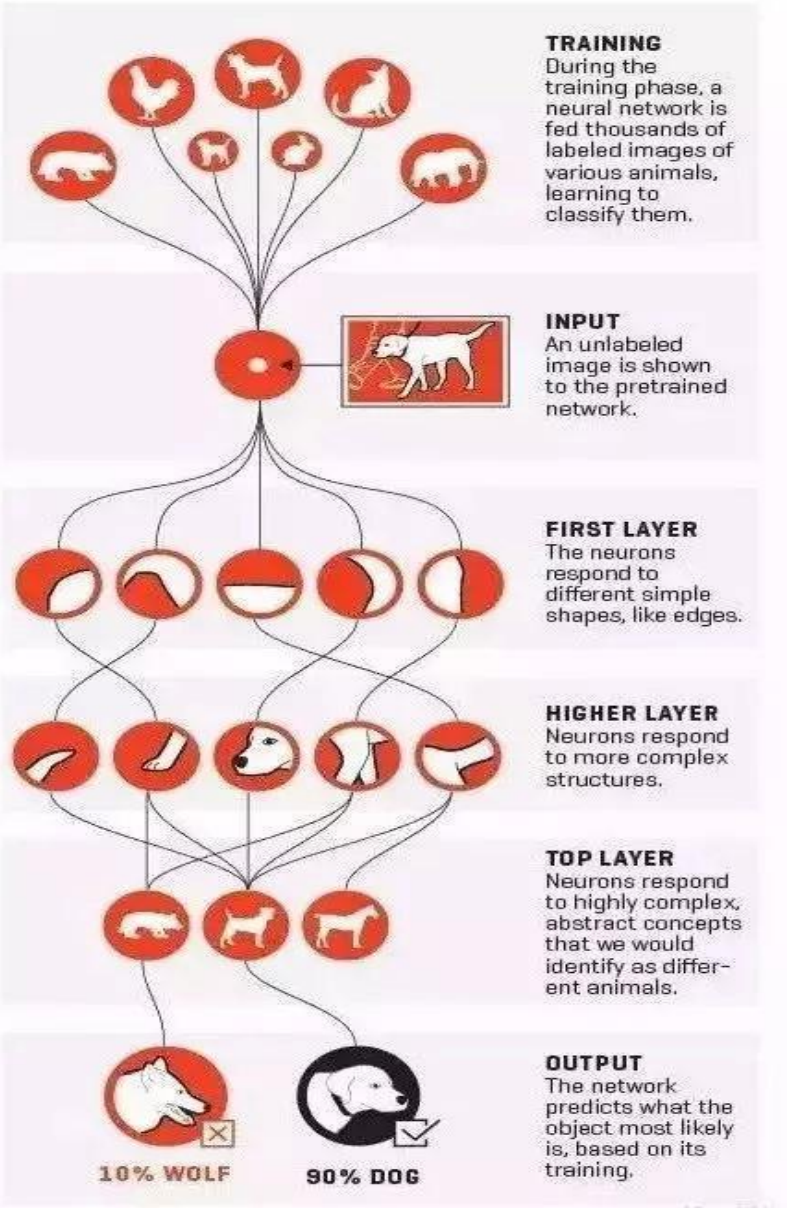


Reorganization

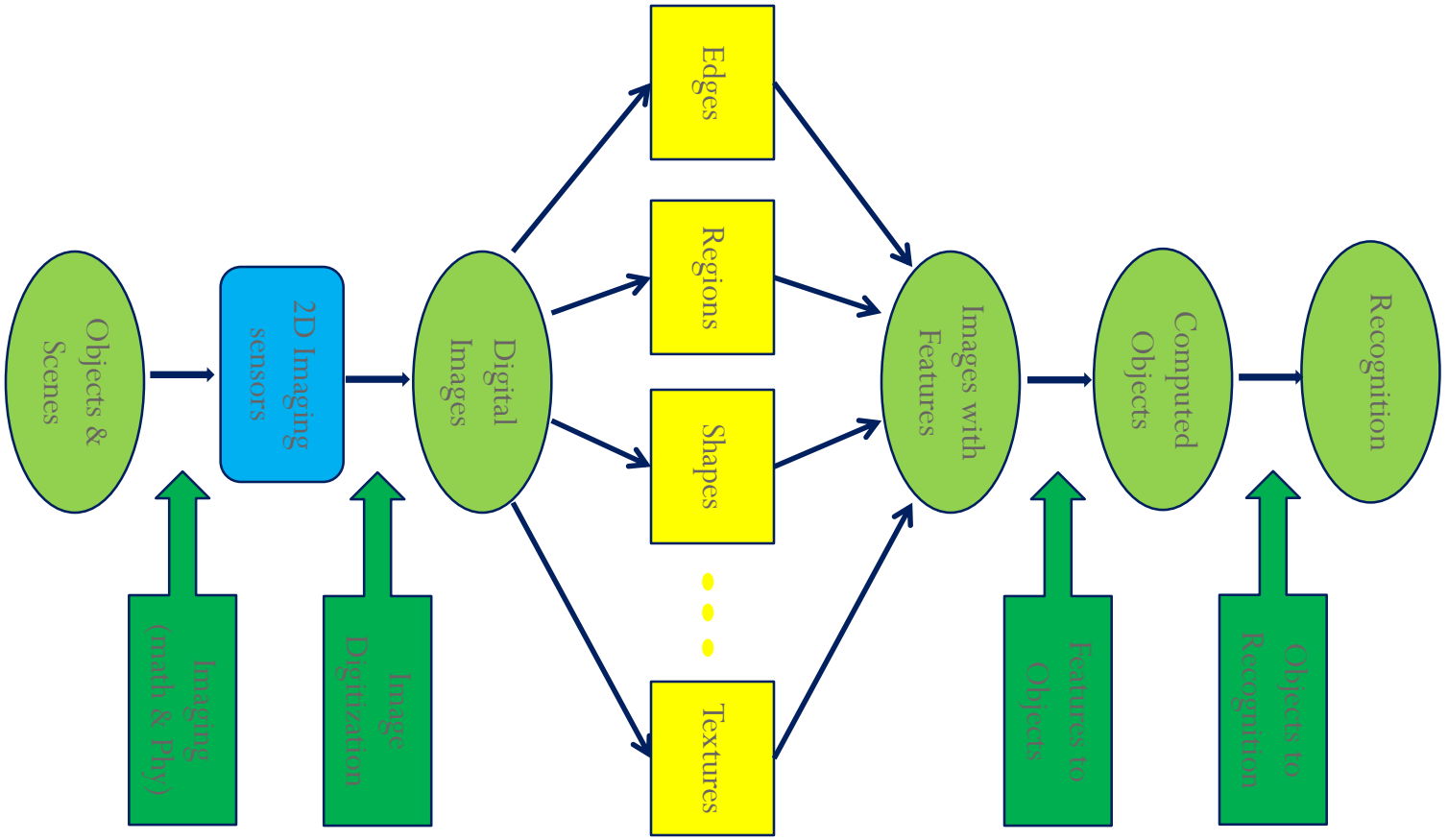


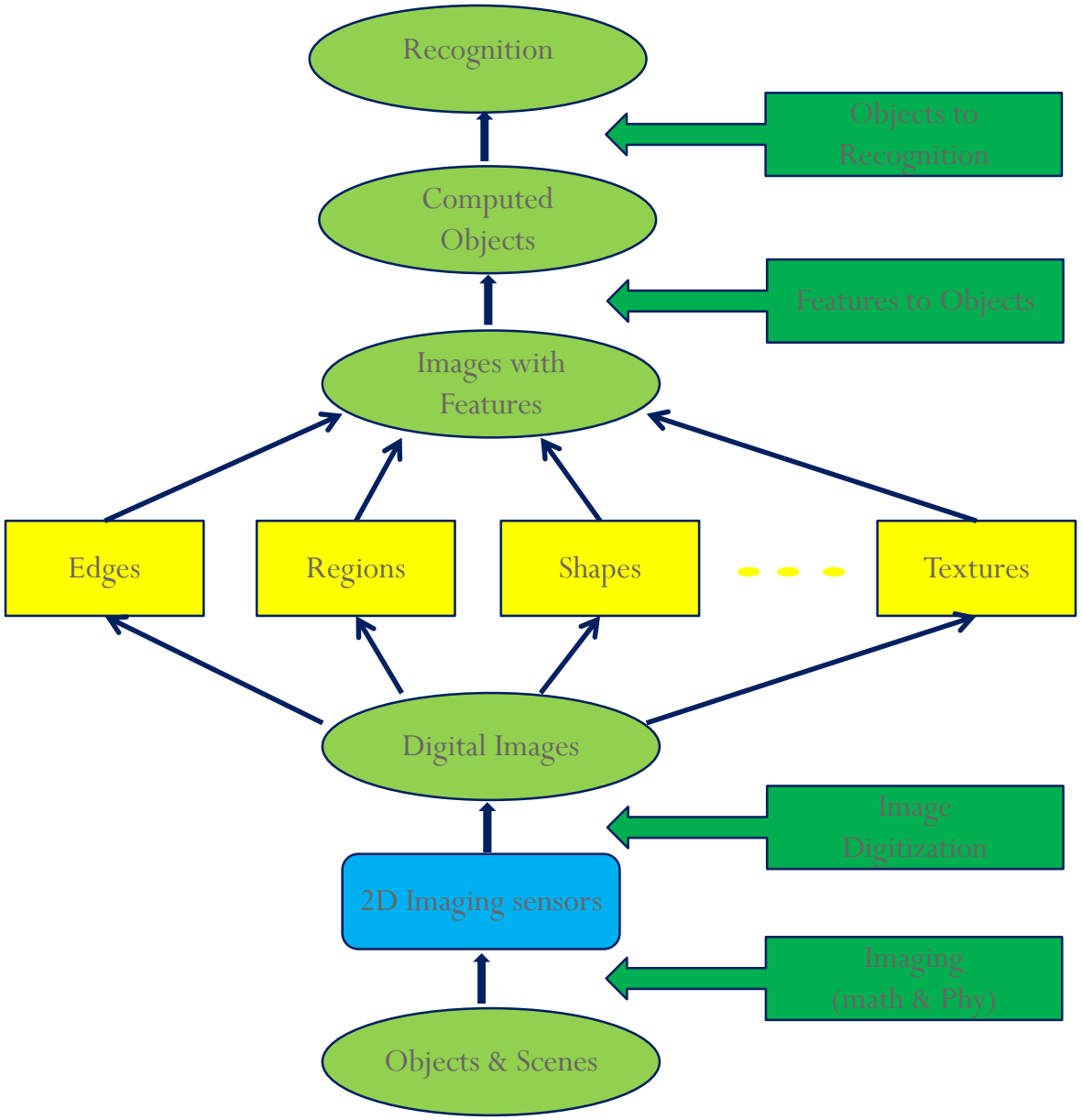
Reconstruction

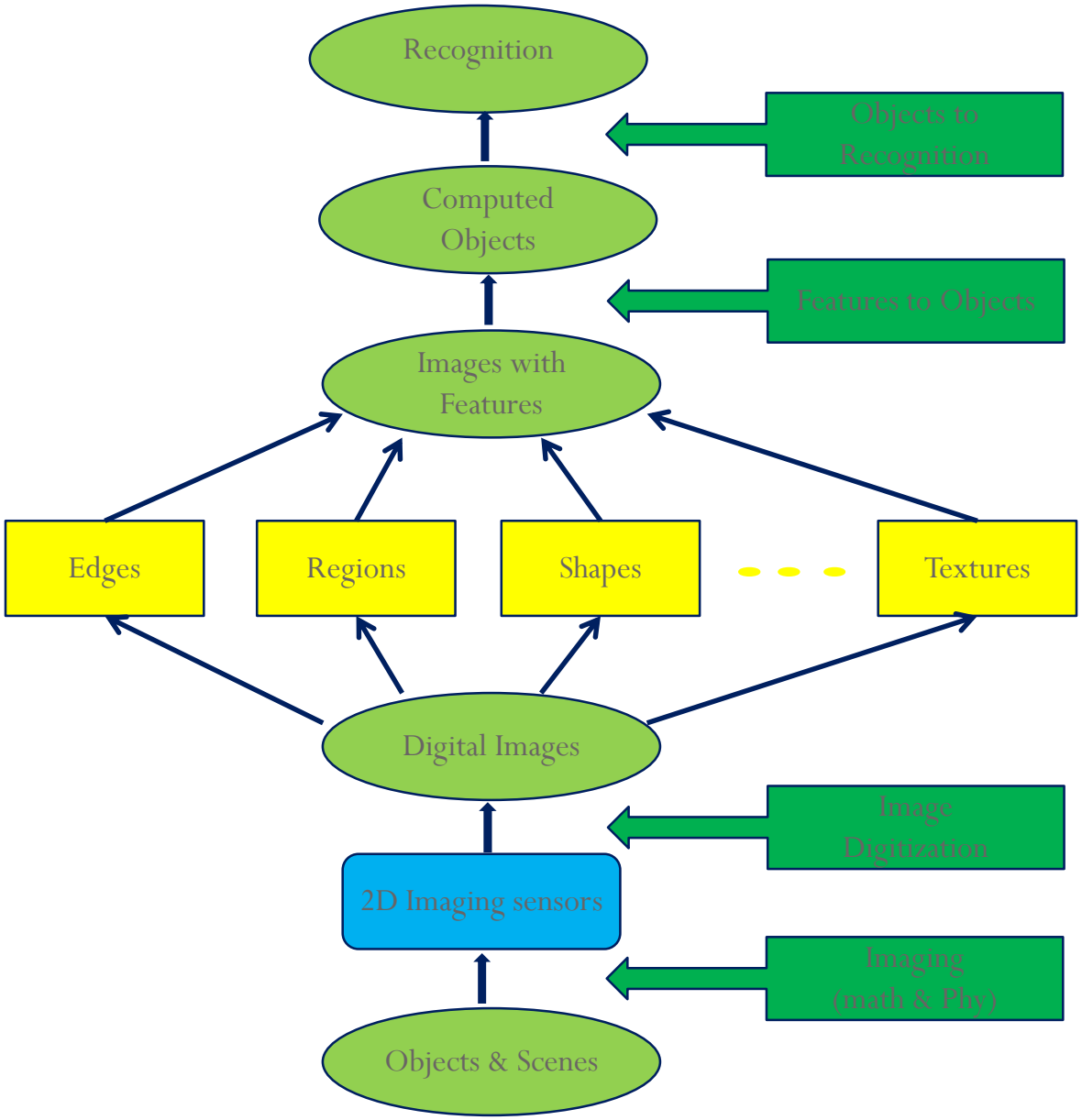
Recognition in Deep-Learning Era

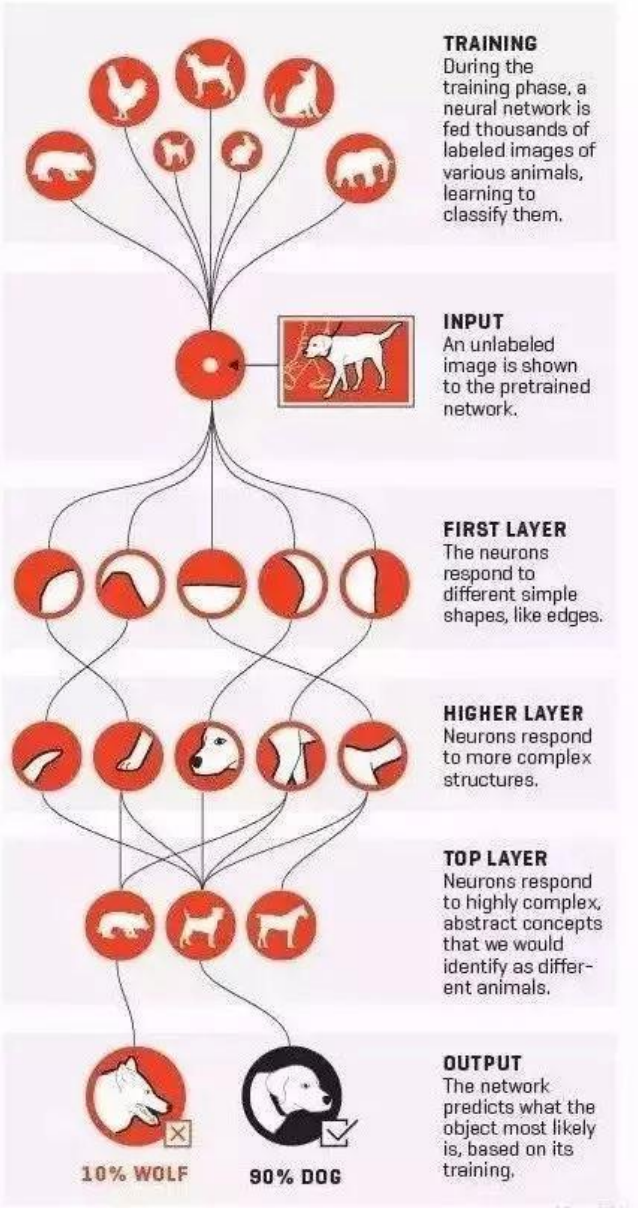


From Images to Recognition



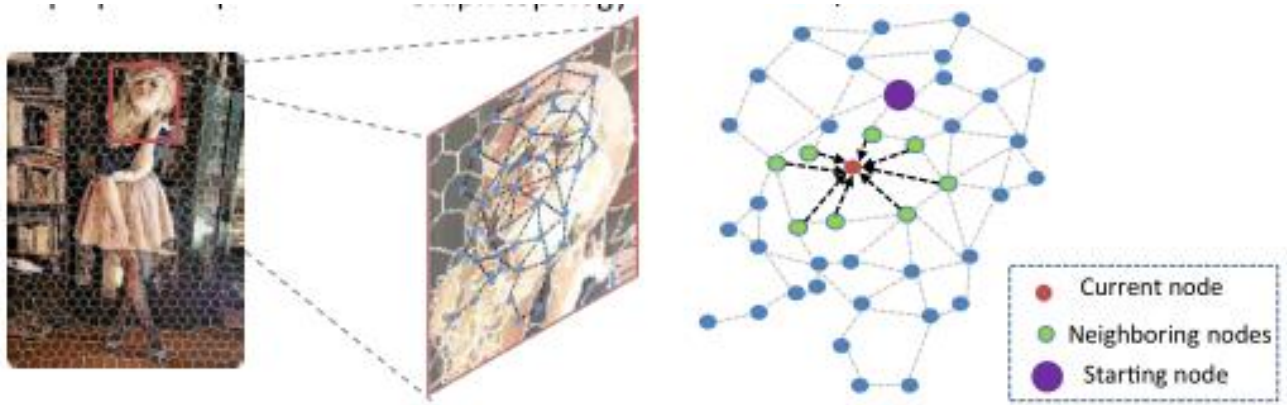

Recognition
Objects to Recognition
Computed Objects
Features to Objects
Images with Features
Edges
Regions
Shapes
Textures
Digital Images
Image Digitization
2D Imaging sensors
Imaging (math & Phy)
Objects & Scenes


Recognition
Objects to Recognition
Computed Objects
Features to Objects
Images with Features
Edges
Regions
Shapes
Textures
Digital Images
Image Digitization
2D Imaging sensors
Imaging (math & Phy)
Objects & Scenes


TRAINING
During the training phase, a neural network is fed thousands of labeled images of various animals, learning to classify them.
INPUT
An unlabeled image is shown to the pretrained network.
FIRST LAYER
The neurons respond to different simple shapes, like edges.
HIGHER LAYER
Neurons respond to more complex structures.
TOP LAYER
Neurons respond to highly complex, abstract concepts that we would identify as different animals.
OUTPUT
The network predicts what the object most likely is, based on its training.
10% WOLF
90% DOG

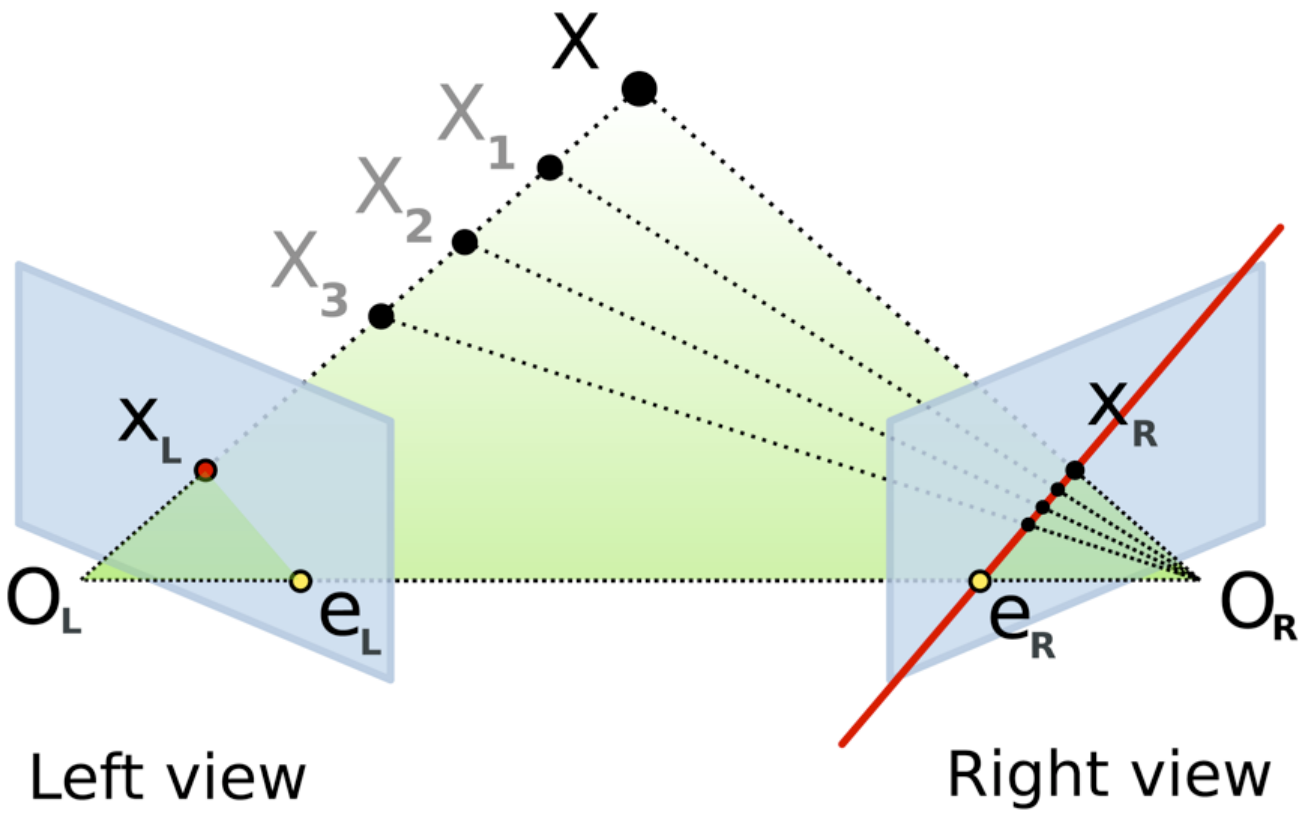
Reorganization

Reorganization



Parsing Graph

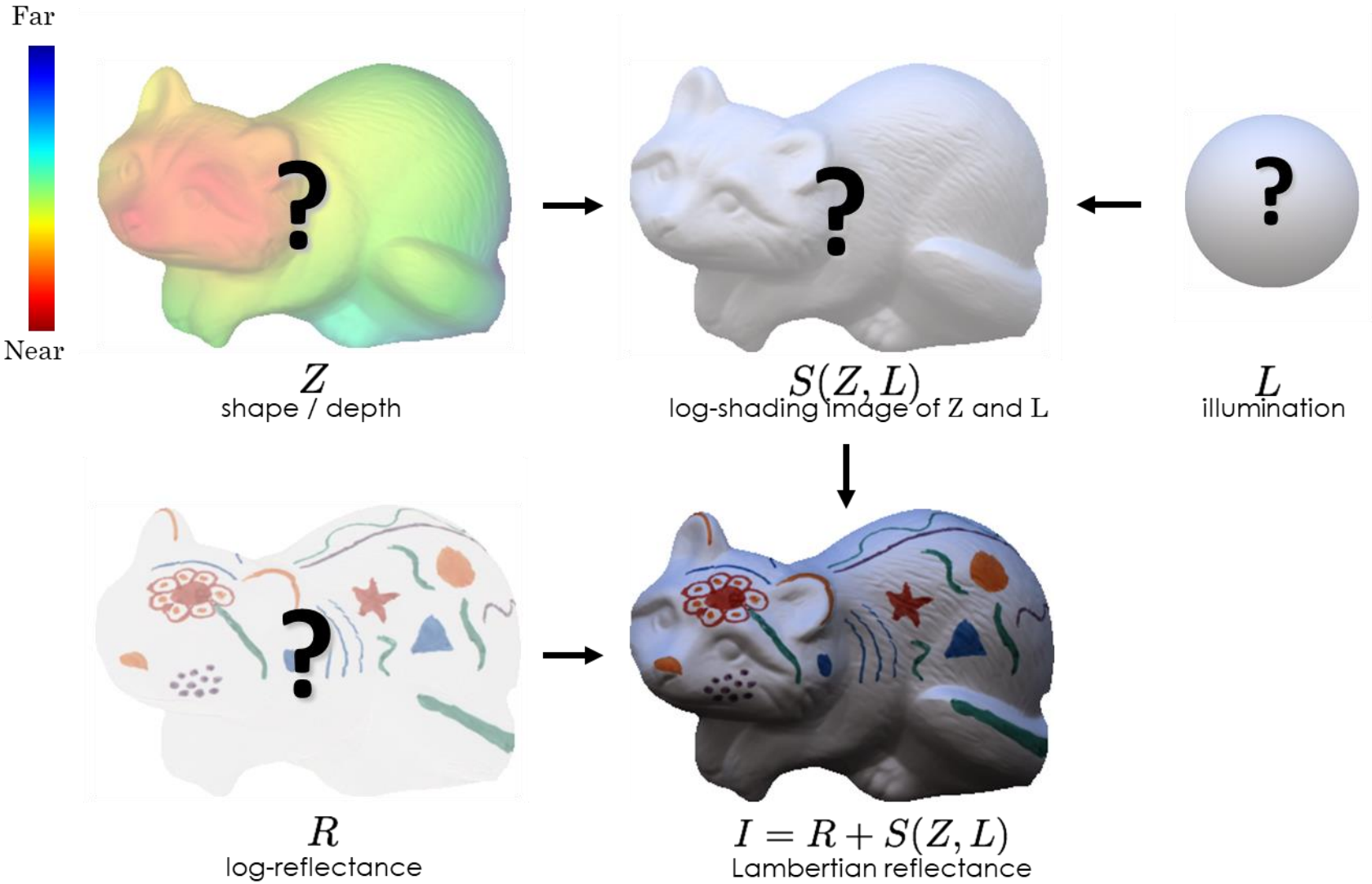
Reconstruction



Reconstruction

Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski.
Building Rome in a Day, ICCV'09

Reconstruction



Jonathan T. Barron, Jitendra Malik. Shape, Illumination, and Reflectance from Shading. TPAMI, 2015

Schedule

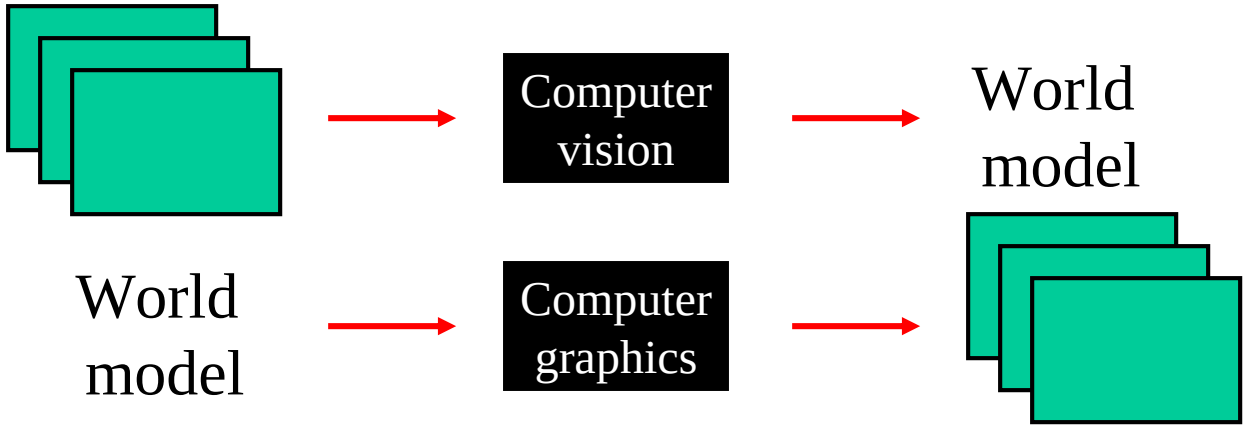
09.02 (3 lectures)	Introduction	10.21 (3 lectures)	Reorganization II
09.09 (3 lectures)	Recall: Image Processing I	10.28 (3 lectures)	Reconstruction I
09.16 (3 lectures)	Recall: Image Processing II	11.04 (3 lectures)	Reconstruction II
09.23 (3 lectures)	Representation I	11.11 (3 lectures)	Recognition I
09.30 (3 lectures)	Representation II	11.18 (3 lectures)	Recognition II
10.07 (3 lectures)	Representation III	11.25 (3 lectures)	Recognition III
10.14 (3 lectures)	Reorganization I	12.02 (1 lecture)	Final Report

Computer Vision vs. Image Processing

- Image Processing – Mostly concerned with image-to-image transformations
  - Filtering, enhancement, restoration, compression, …
- Computer Vision – Almost always concerned with how the given images/videos reflect the real world (3D scenes/objects)
  - Filtering (for feature extraction)
  - Enhancement (for scene/object detection)
  - Segmentation (for region processing and identification)
  - Object Recognition (for image understanding)

Computer Vision vs. Computer Graphics

- Computer vision is also considered by many as the inverse of computer graphics (Richard Szeliski, MSR)



Intersection between Vision and Graphics

- From Richard Szeliski, MSR

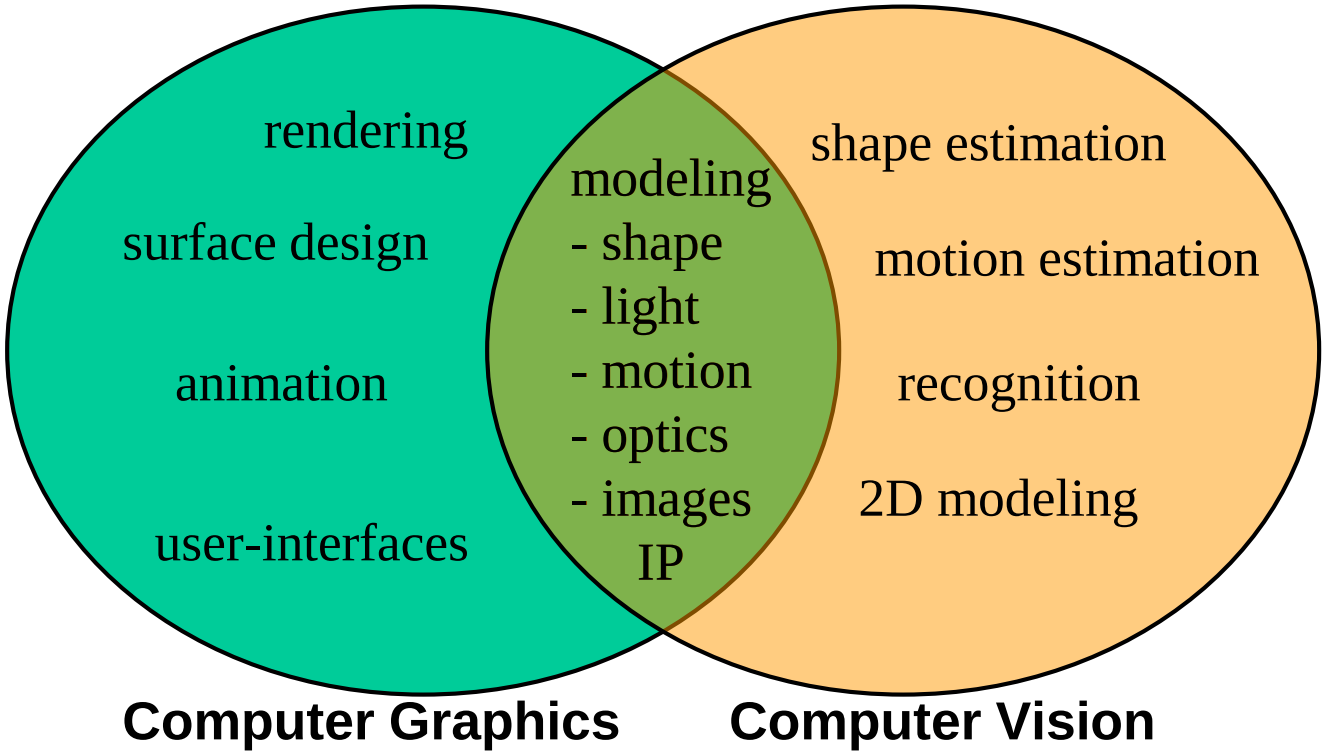


Image Understanding Basic Operations

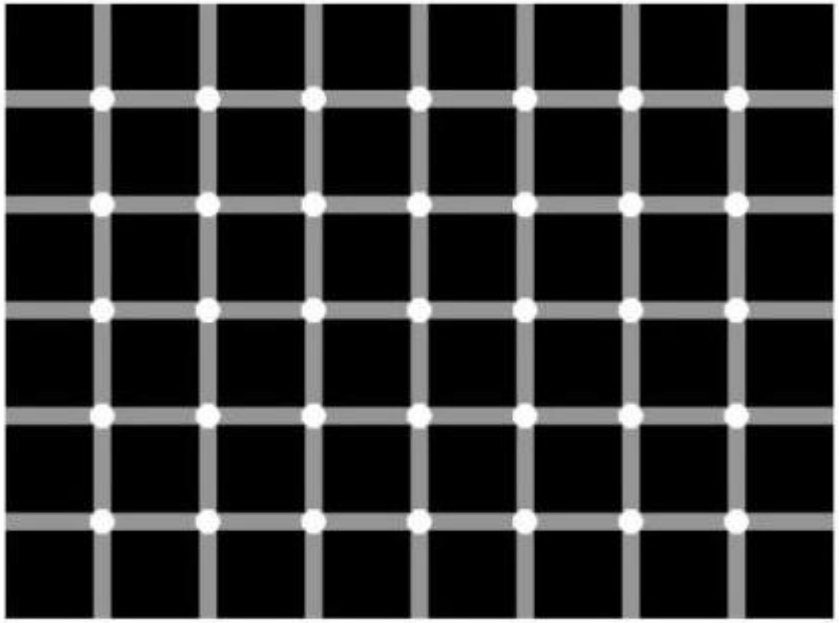
- From Objects/Scene to Digital Images
  - Imaging, Sampling and Quantization
- From Digital Images to Image Features
  - Low Level Image Processing – Edges, Regions, …
- From Image Features to Objects
  - High Level Image/Feature Processing
- From Objects to Recognition
  - Image Understanding and Computer Vision Tasks

Evaluating Computer Vision Algorithms

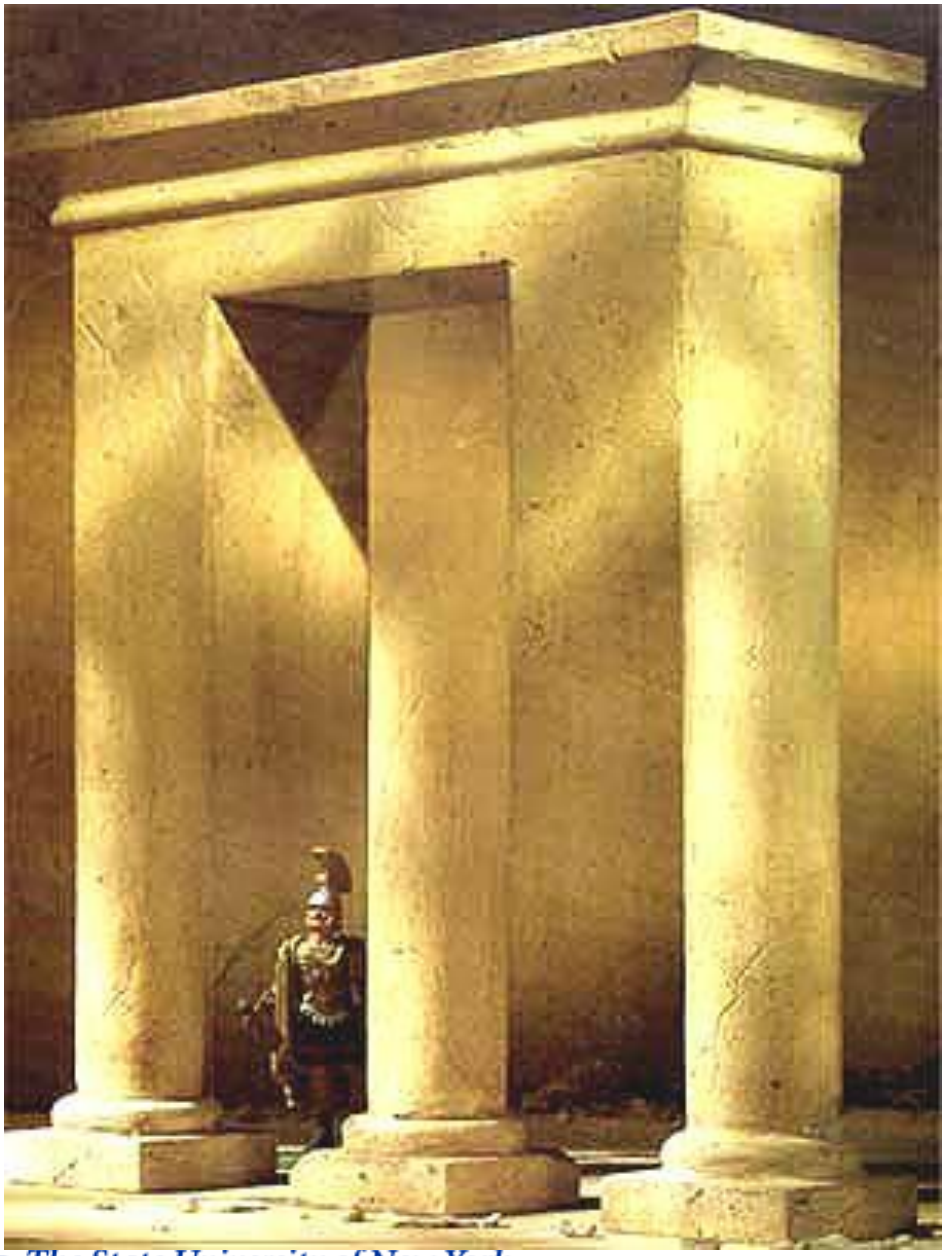
- We need to understand what are the physical process (model) that governs the generation of images from 3D scenes/objects.
- If we cannot understand the forward process (imaging process or computer graphics), it will be difficult to solve the inverse problem (computer vision)
- The best way to evaluate computer vision algorithm is almost always to build a forward model to test the inverse solutions

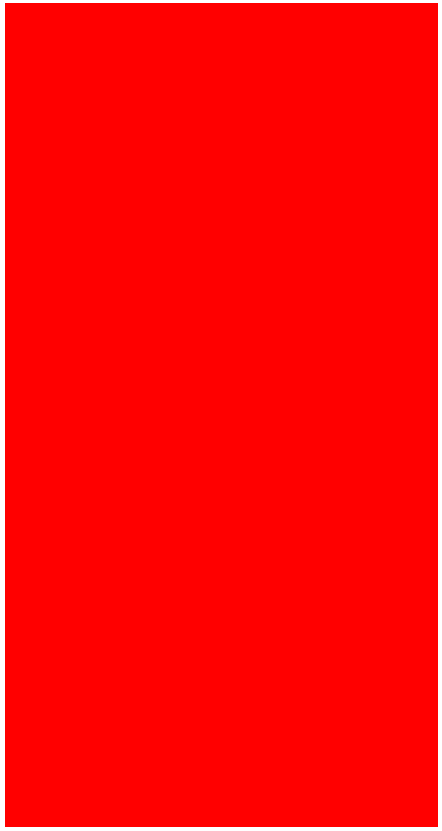
Challenges in Computer Vision

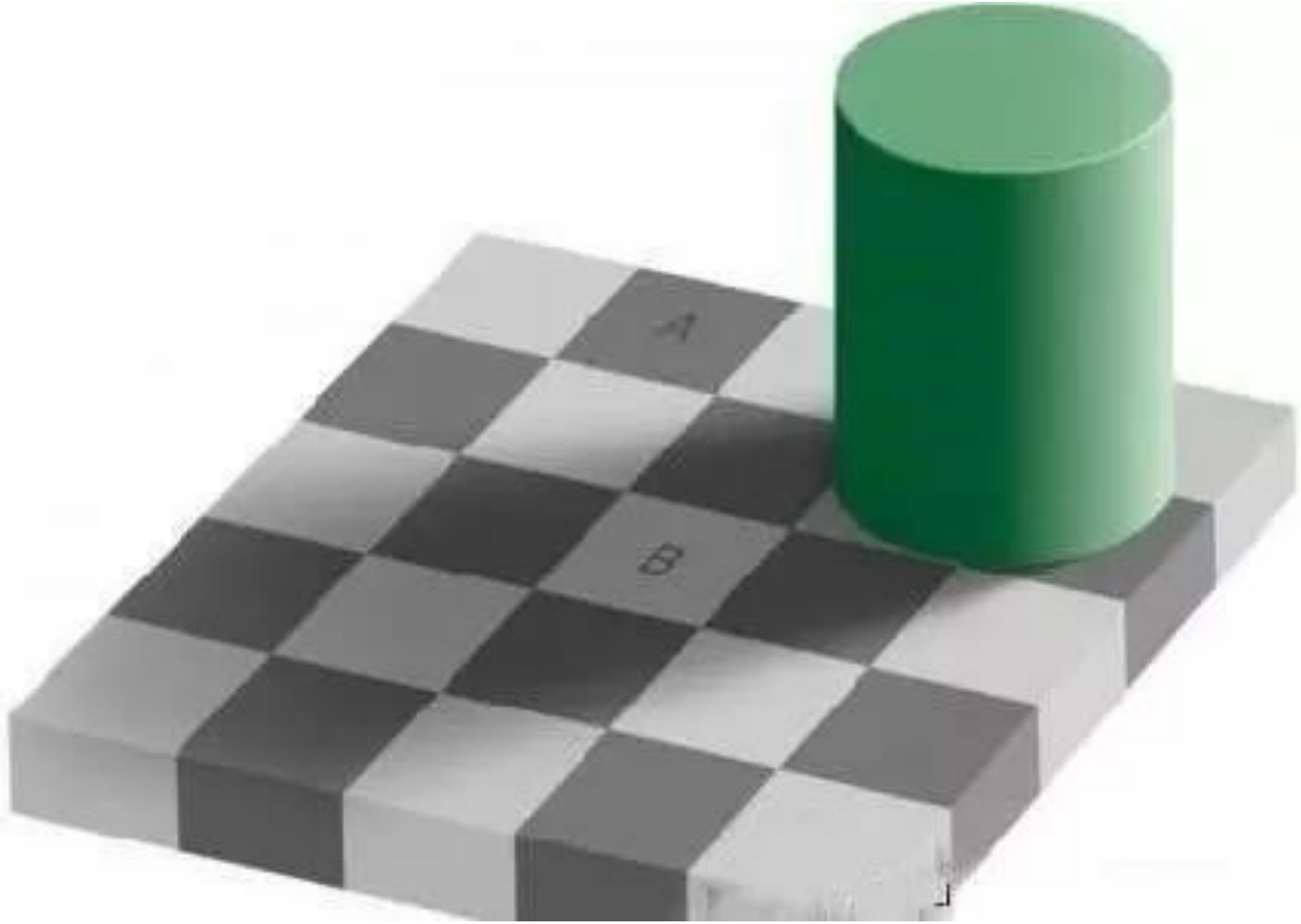
- Loss of dimension (information)
  - 3D Scene/Objects to 2D Images or Image Sequences
- Human visual perception model
  - Study of human visual perception principles is still on-going
- Noise and processing errors
  - Noise exist in both imaging and image processing system
- Data volume
  - Image and video data volume is usually enormous; Internet scale
- Semantic gap
  - Significant gaps between low level features and high level semantics

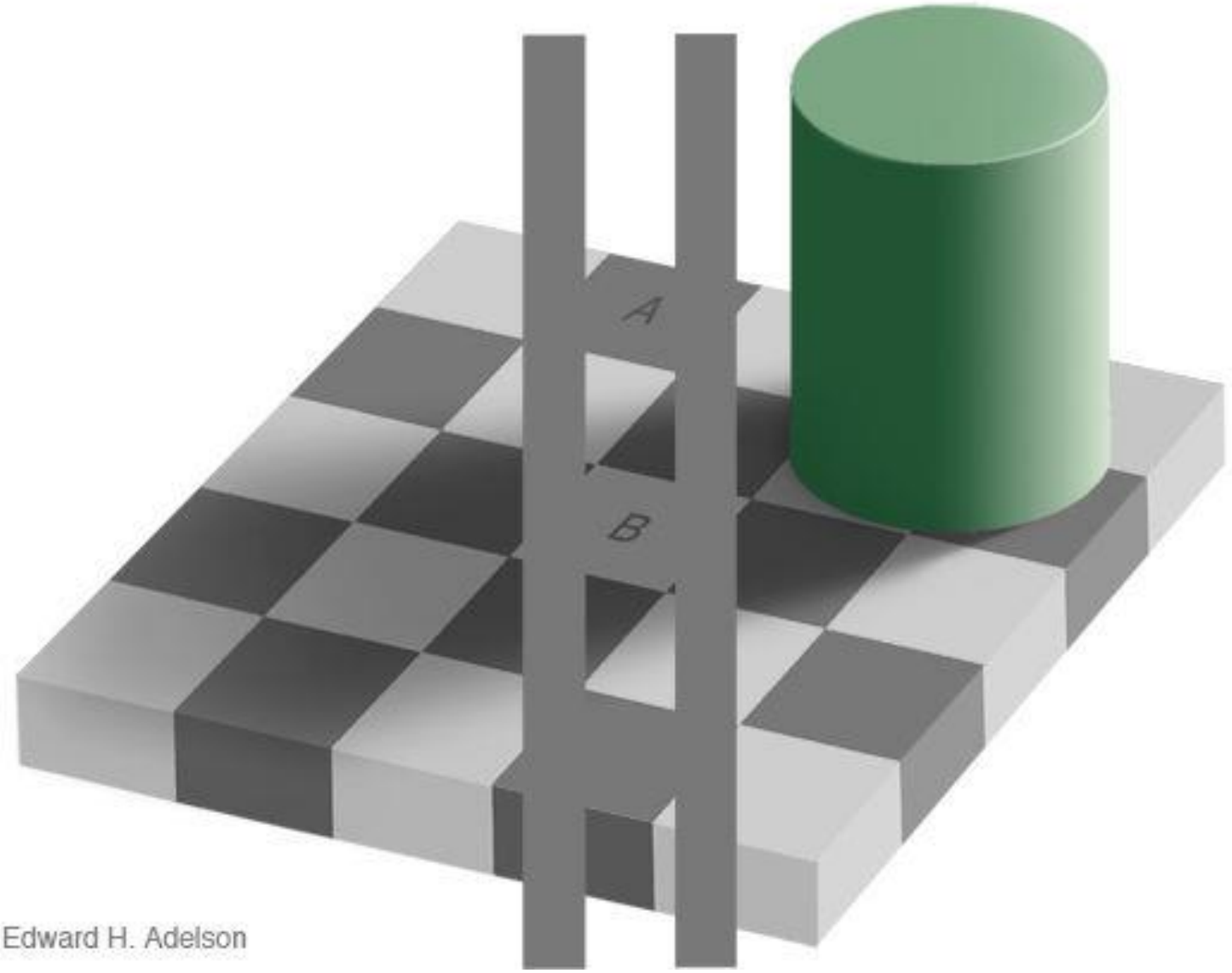


Count the black dots! :o)

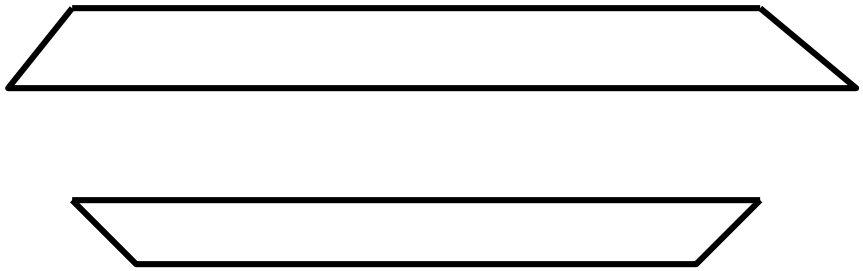


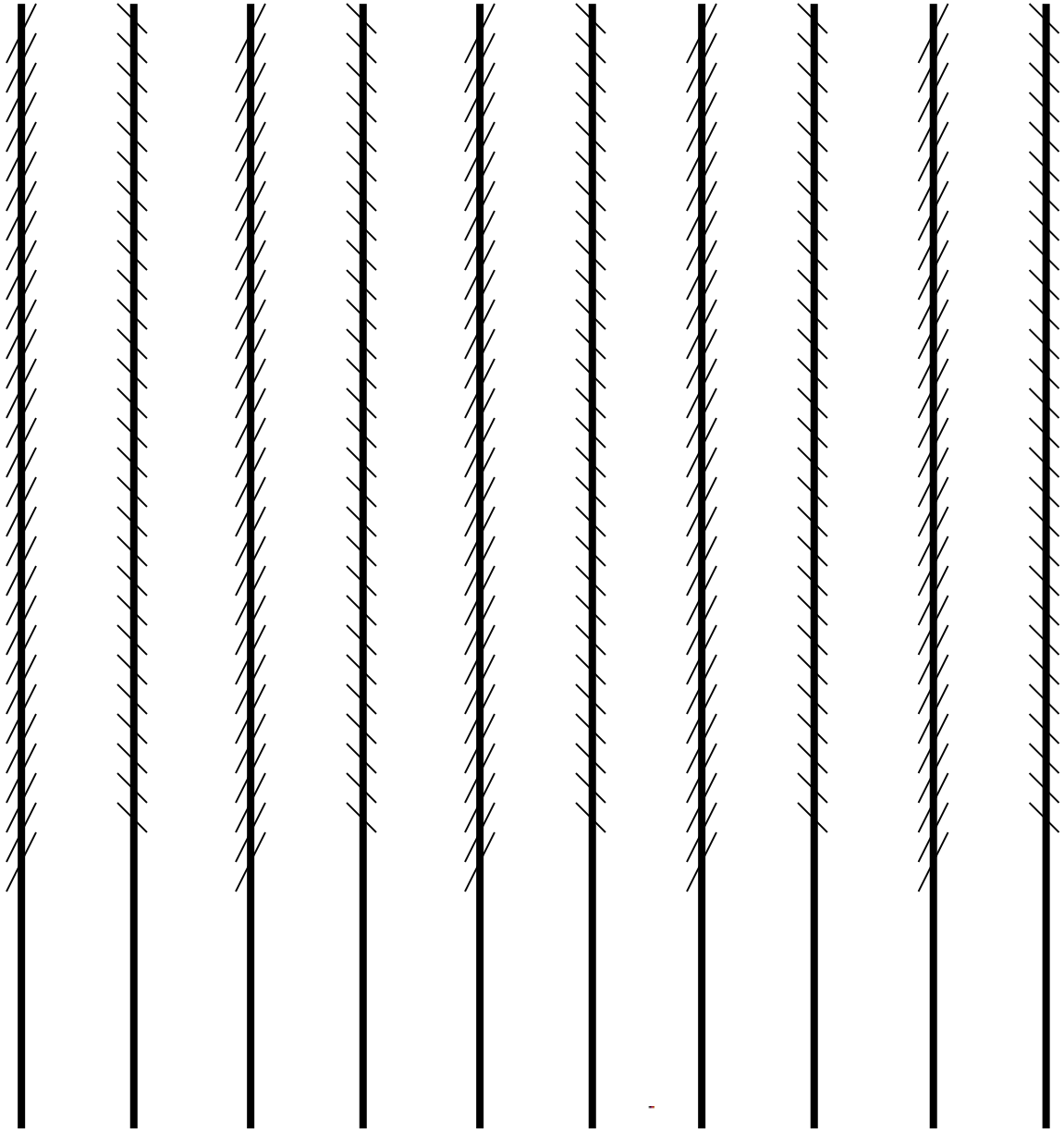


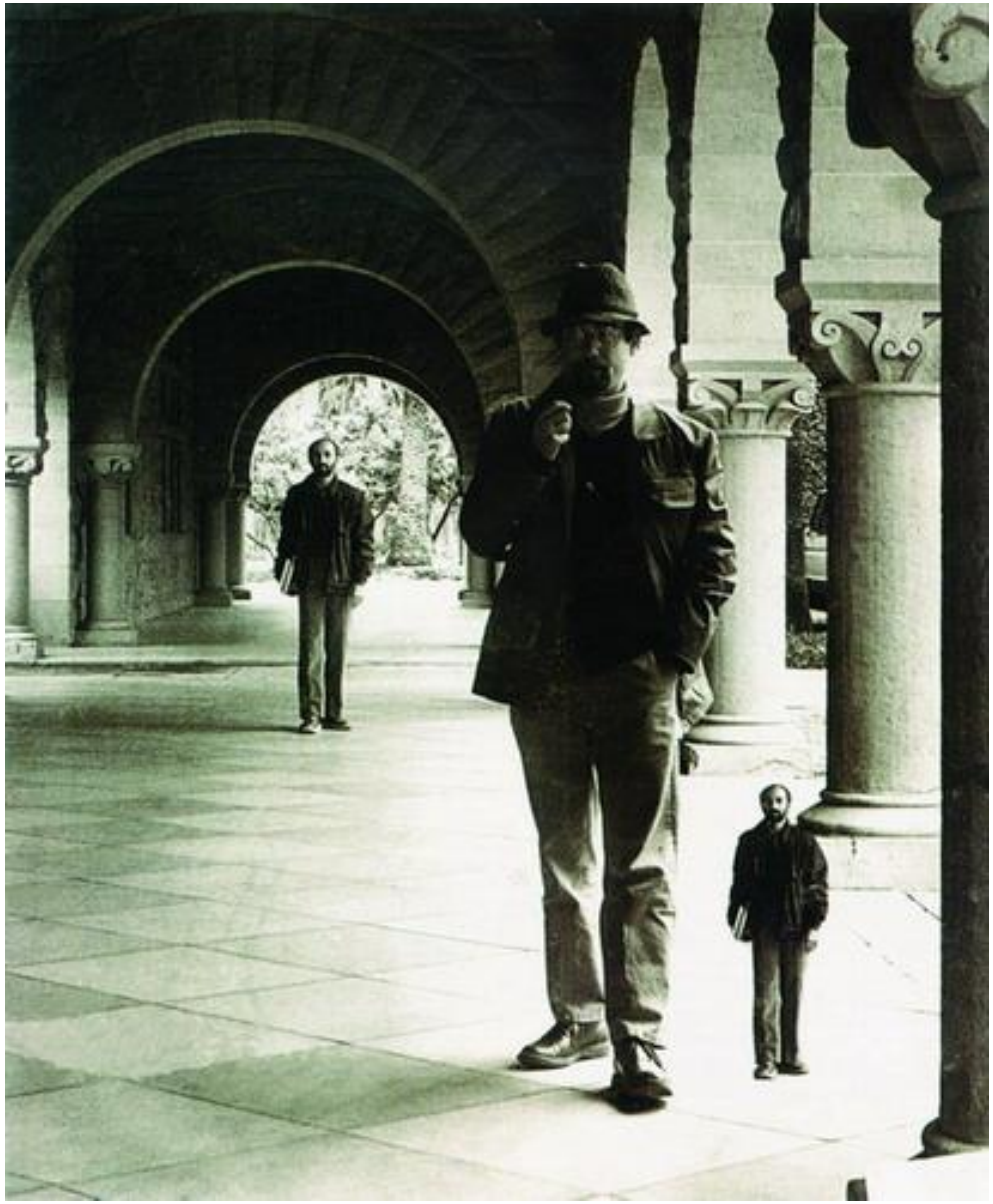

A
B



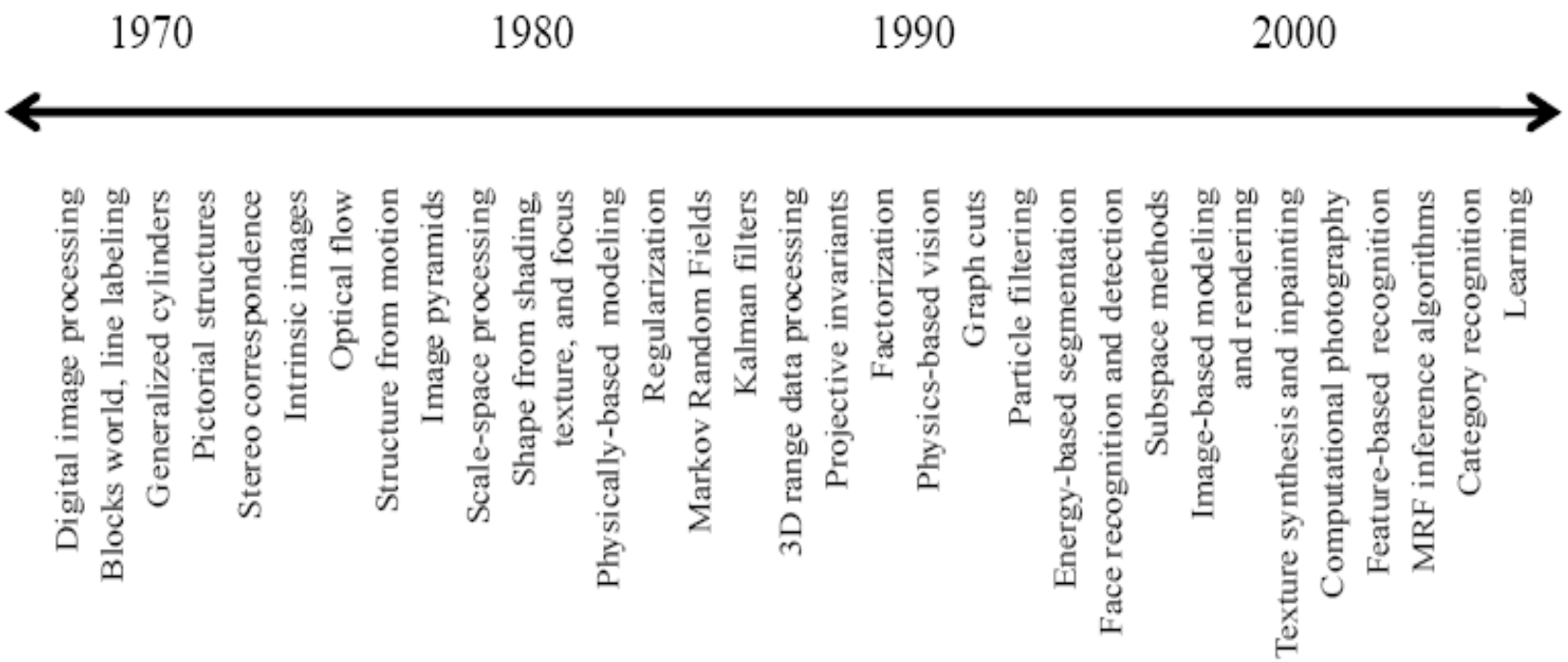
Edward H. Adelson

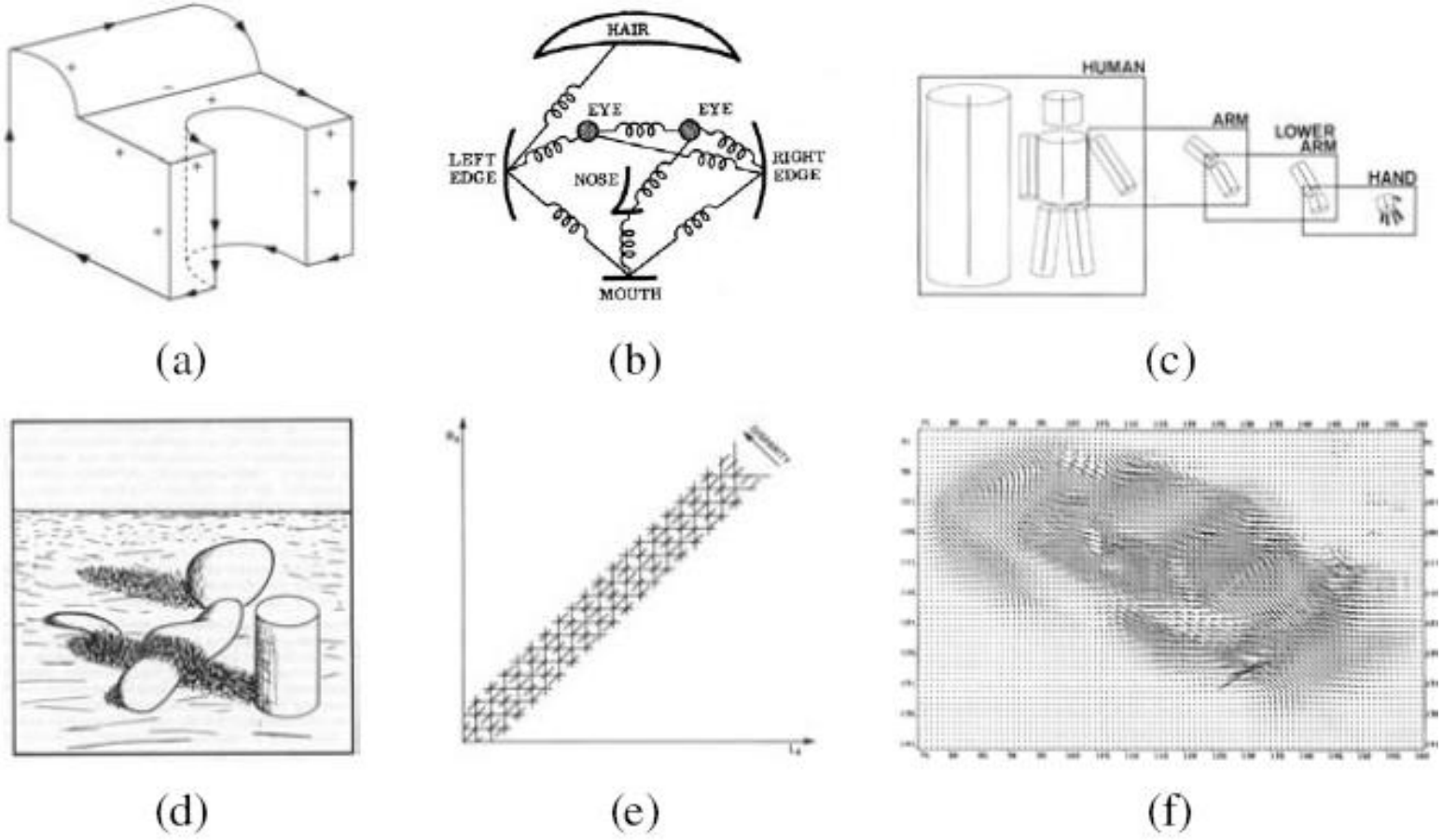






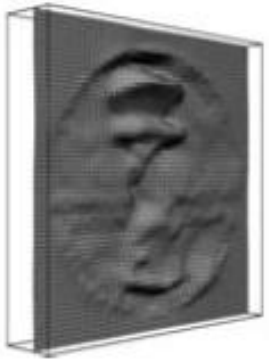
Development of Computer Vision



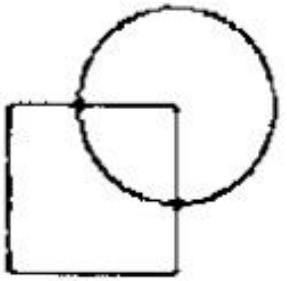


(a) line labeling(Nalwa 1993), (b) pictorial structures (Fischler and Elschlager 1973) (c) articulated body model (Marr 1982) (d) intrinsic images Barrow and Tenenbaum 1981) (e) stereo correspondence (Marr 1982 (f) optical flow (Nagel and Enkelmann 1986)

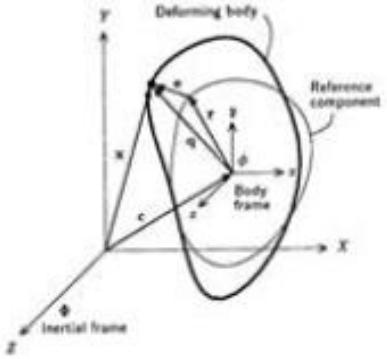
(a)



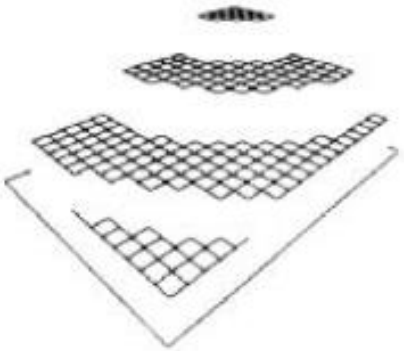
(b)



(c)



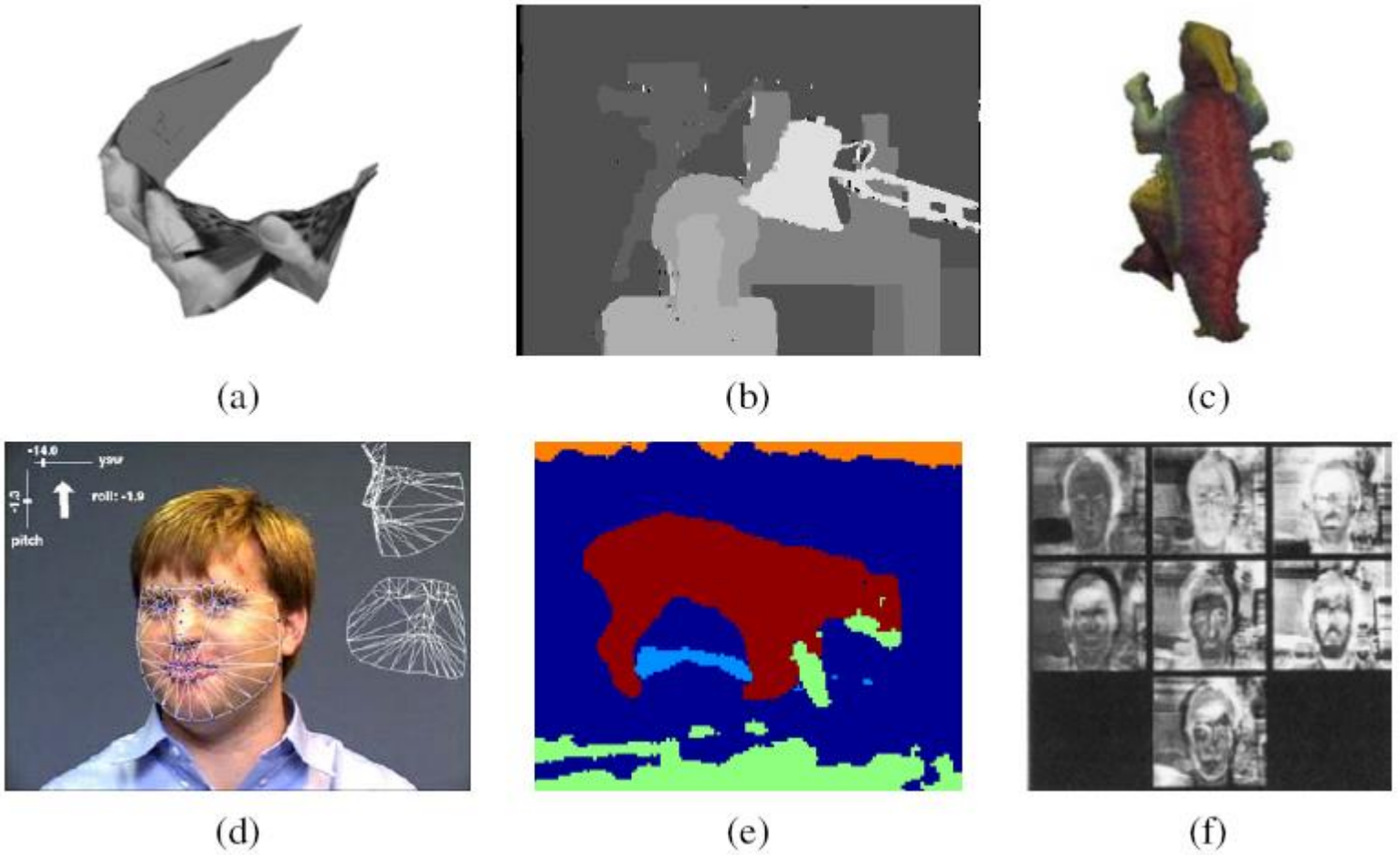
(d)



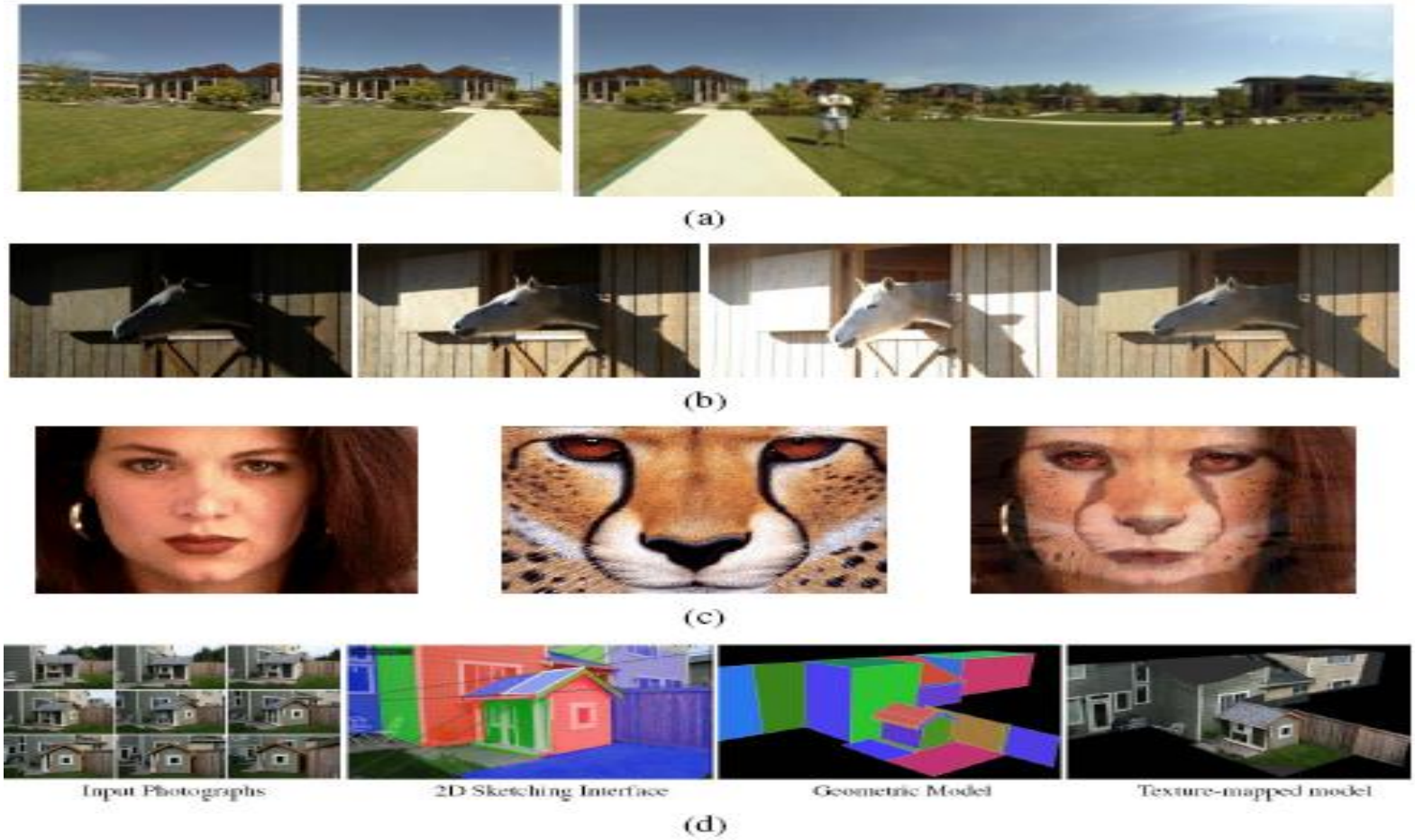
(e)

(f)

(a) pyramid blending (Burt and Adelson 1983) (b) shape from shading (Freeman and Adelson1991) (c) edge detection (Freeman and Adelson 1991) (d)physically based models (Terzopoulos and Witkin 1988) (e) regularization based surface reconstruction (Terzopoulos 1988) (f) range data acquisition and merging (Banno, Masuda, Oishi et al. 2008)



(a) factorization-based structure from motion (Tomasi and Kanade 1992), (b) dense stereo matching (Boykov, Veksler, and Zabih 2001), (c) multi-view reconstruction (Seitz and Dyer 1999) (d) face tracking (Matthews, Xiao, and Baker 2007), (e) image segmentation (Belongie, Fowlkes, Chung et al. 2002) (f) face recognition (Turk and Pentland 1991a).



(a) image stitching: merging different views (Szeliski and Shum 1997) (b) exposure bracketing: merging different exposures; (c) morphing: blending between two photographs (Gomes, Darsa, Costa et al. 1999 (d) turning a collection of photographs into a 3D model (Sinha, Steedly, Szeliski et al. 2008)

Summary

- Computer vision is still far from human vision systems, but recent advances are encouraging
- With the growth of computing power, it is possible to perform more and more complex processing on more and more images....
- More recent approaches/advances take advantage of heavily data-driven approaches – Internet Vision
- In this course, we will learn some basic image processing and computer vision techniques – starting from the bottom up – mathematics and physics of imaging process