UNIVERSITY of ROCHESTER

Machine Vision

CSC 249/449 Spring 2019

http://www.cs.rochester.edu/~cxu22/t/249S19/

Instructor: Chenliang Xu
chenliang.xu@rochester.edu

Introductions

- Instructor:
  - Prof. Chenliang Xu

- Teaching Assistants:
  - Yapeng Tian, Graduate TA
  - Jing Shi, Graduate TA
  - Zhiheng Li, Graduate TA
  - Yutong (Kelly) He, Undergraduate TA
  - Chaoying Xue, Undergraduate TA

Today

- Course overview
- Requirements, logistics

What is computer vision?

- Automatic understanding of images and videos

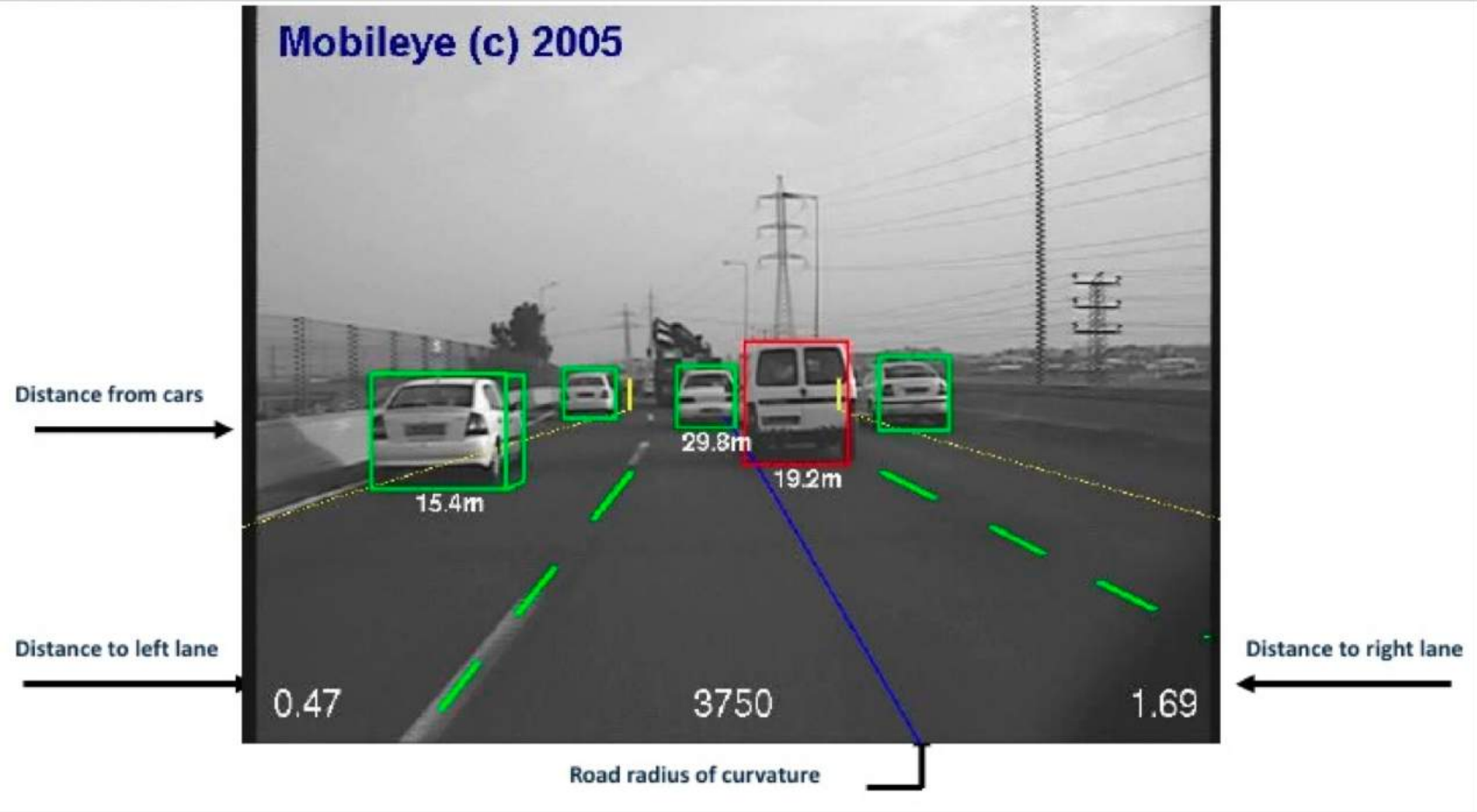
Adapted from: K. Grauman

What is computer vision?

- Automatic understanding of images and videos
  - 1. Computing properties of the 3D world from visual data (measurement)

Adapted from: K. Grauman

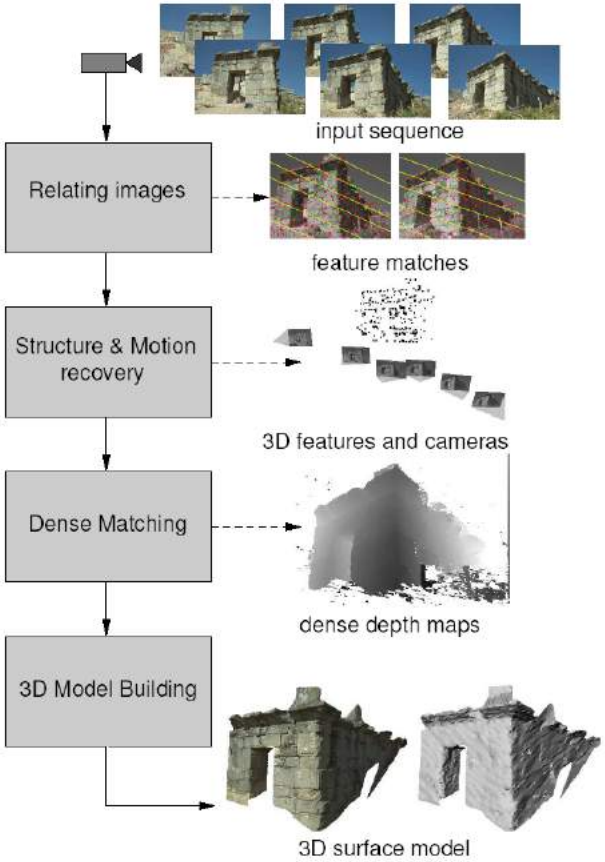
Vision as measurement device (geometric information)



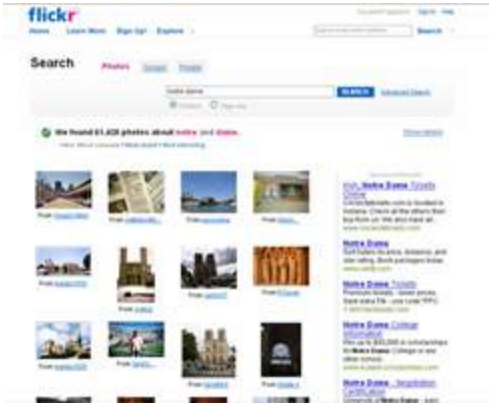
Source: Mobileye

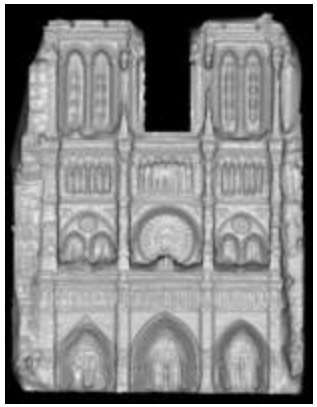
Vision as measurement device (geometric information)

NASA Mars Rover



Pollefeys et al.





Goesele et al.

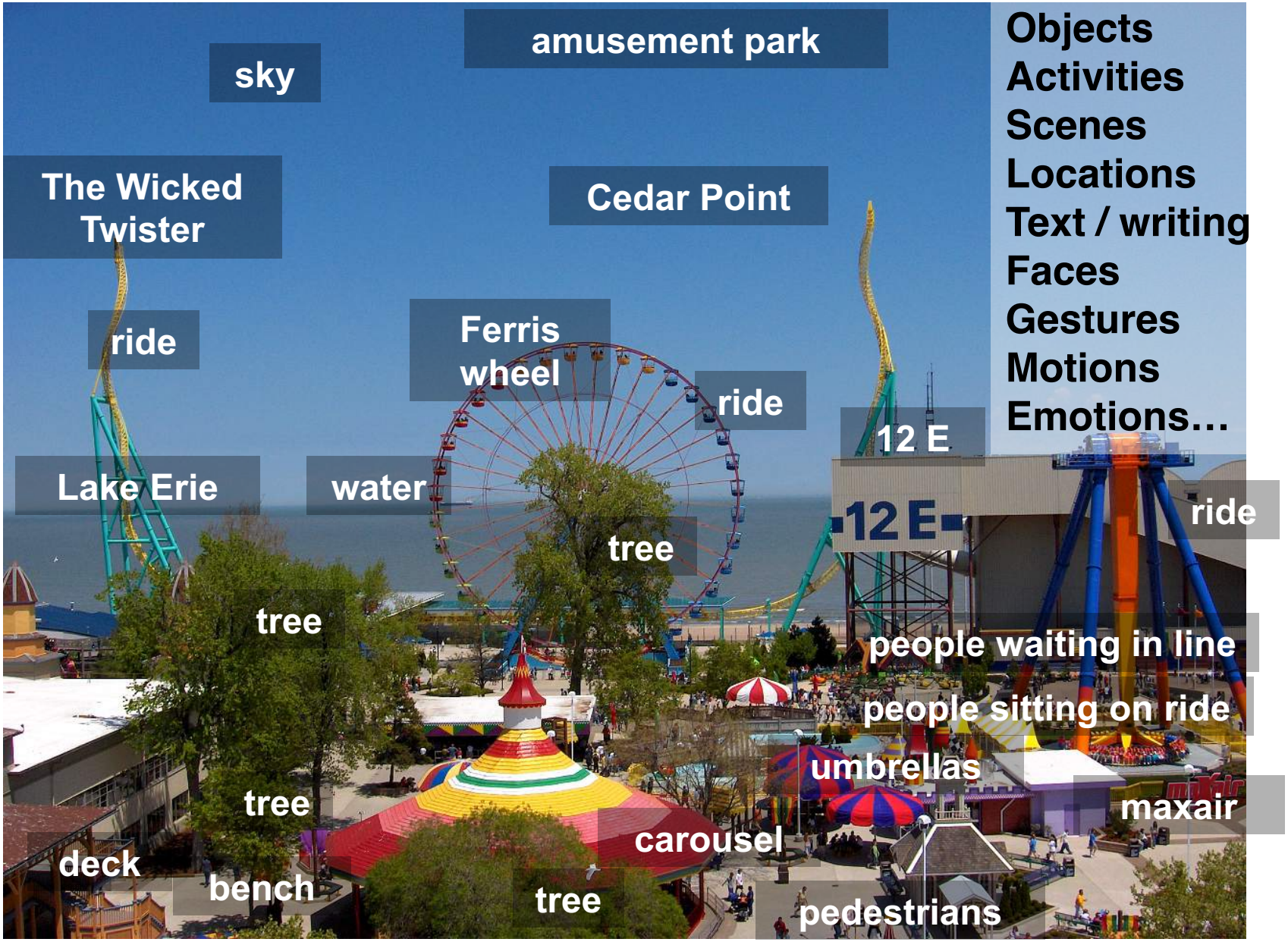
Slide Credit: Svetlana Lazebnik

What is computer vision?

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

Adapted from: K. Grauman

Vision as a source of Semantic Information



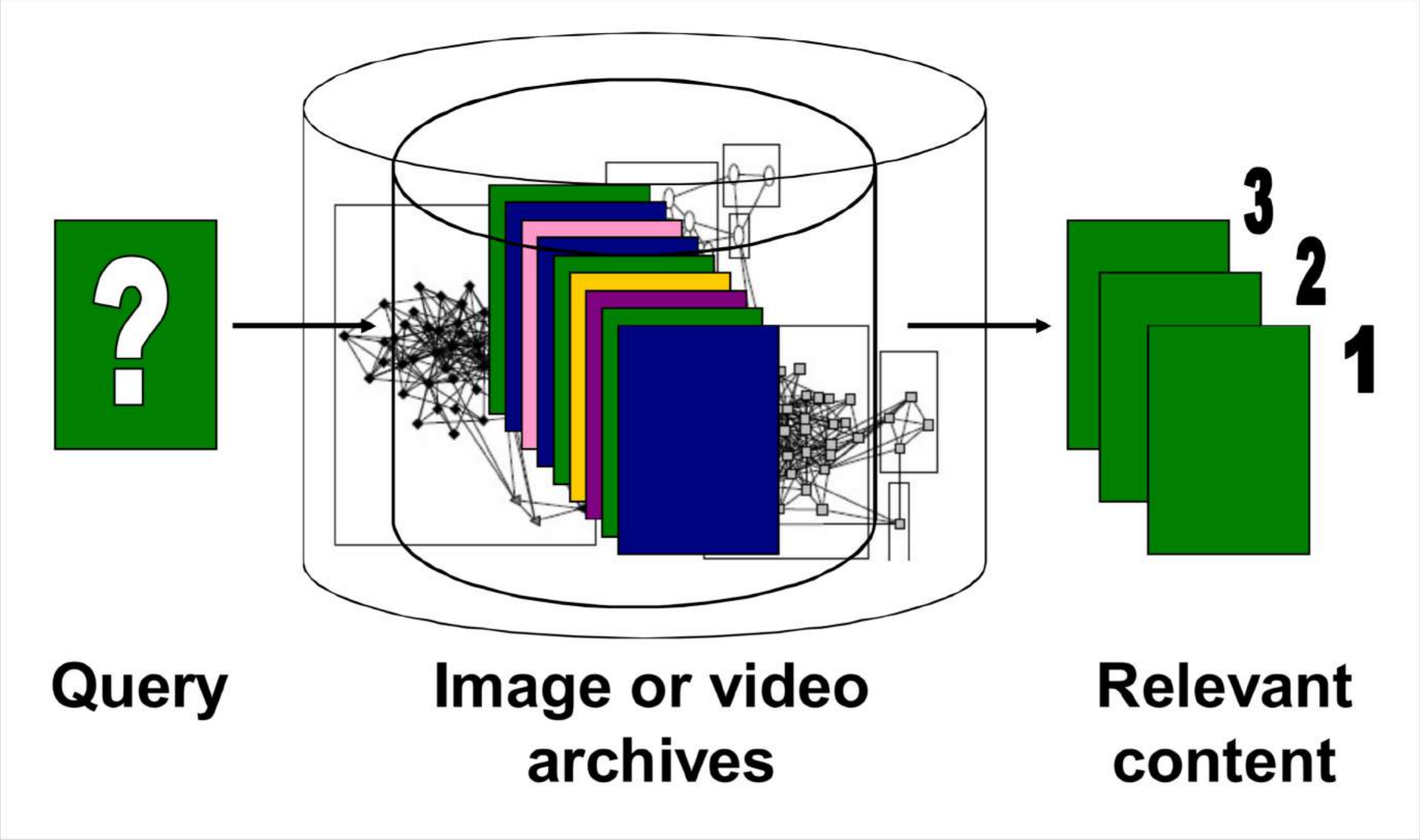
Source: Kristen Grauman

What is computer vision?

- Automatic understanding of images and videos
  - 1. Computing properties of the 3D world from visual data (measurement)
  - 2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (perception and interpretation)
  - 3. Algorithms to mine, search, and interact with visual data (search and organization)

Adapted from: K. Grauman

Visual search, organization



Slides: K. Grauman

What is computer vision?

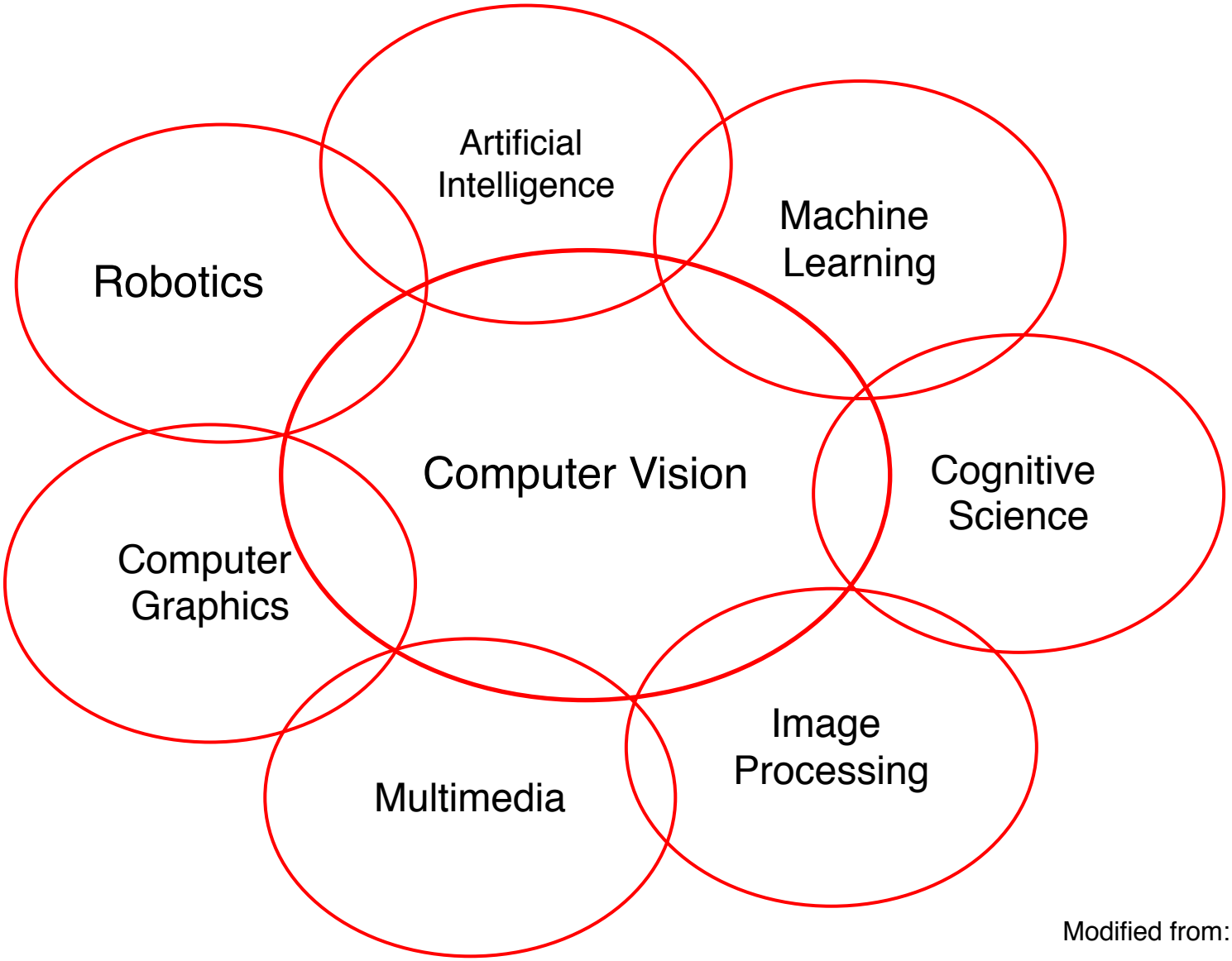
- Automatic understanding of images and videos
  - 1. Computing properties of the 3D world from visual data (measurement)
  - 2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (perception and interpretation)
  - 3. Algorithms to mine, search, and interact with visual data (search and organization)
  - 4. Algorithms to synthesize objects, actions, and physical processes (synthesis and content creation)

Adapted from: K. Grauman

Synthesis and content creation

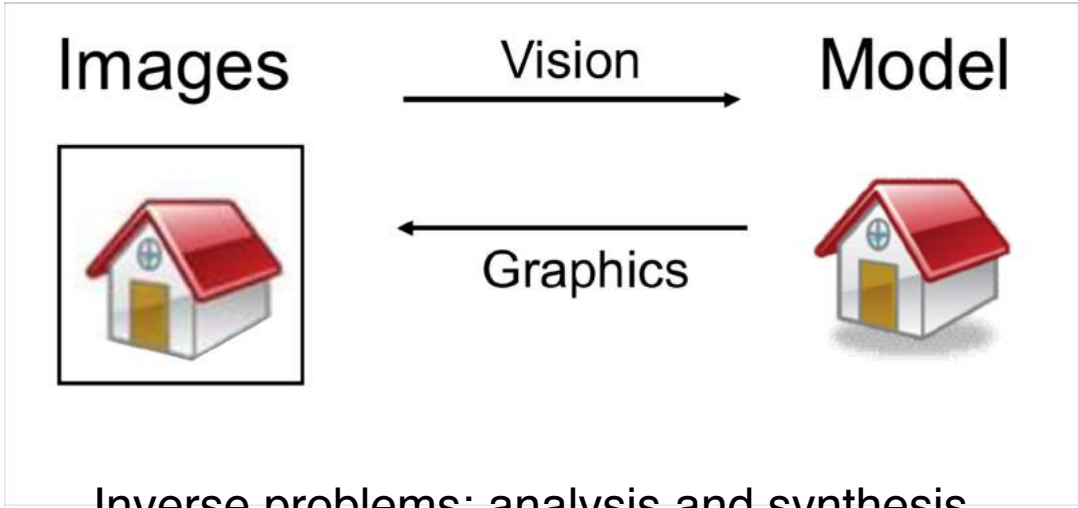
Figure: Google BigGAN

Connections to other disciplines



Modified from: K. Grauman

Computer Vision & Computer Graphics



Inverse problems: analysis and synthesis

slides: K. Grauman

Computer Vision & Cognitive Science

Image (or video)

Sensing device

Interpreting device

Interpretations

garden, spring, bridge, water, trees, flower, green, etc.

slides: Fei-Fei Li

A quick experiment
Animals or Not?

3asig Al5ail

10636

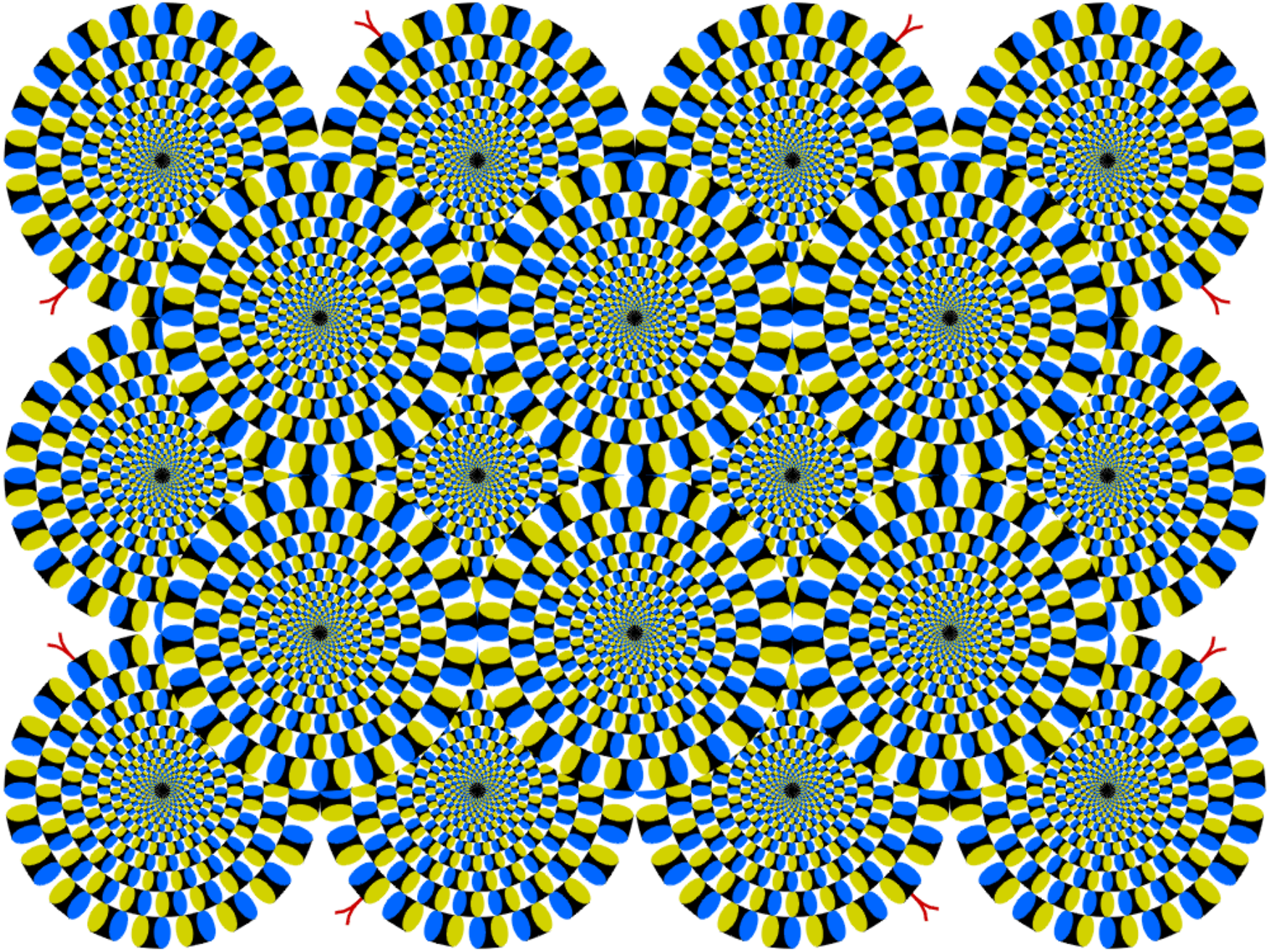
Zaehg Albait

Thorpe, et al. *Nature,* 1996

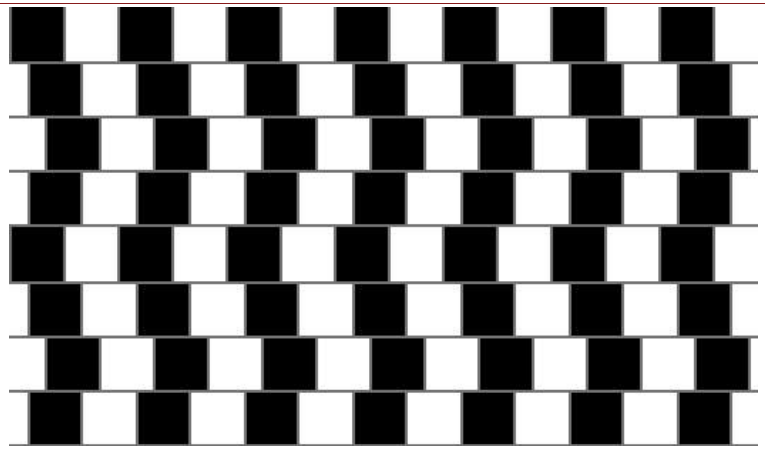
Under 150 ms from measured brain responses, ERP!!

Human vision is not perfect

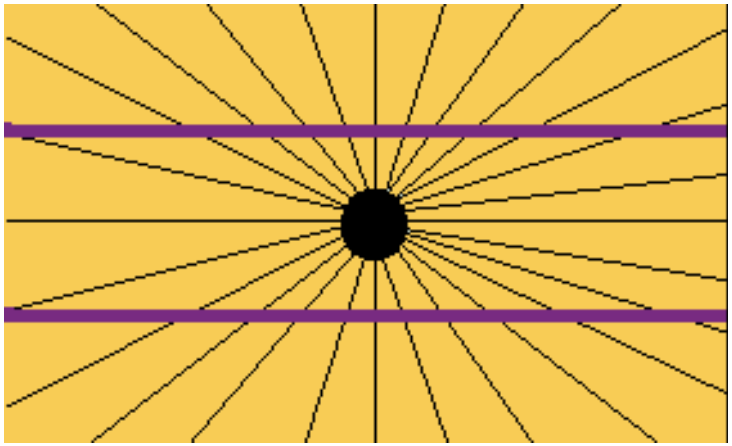
Sinha, Pawan & Poggio, *Nature*, 1996



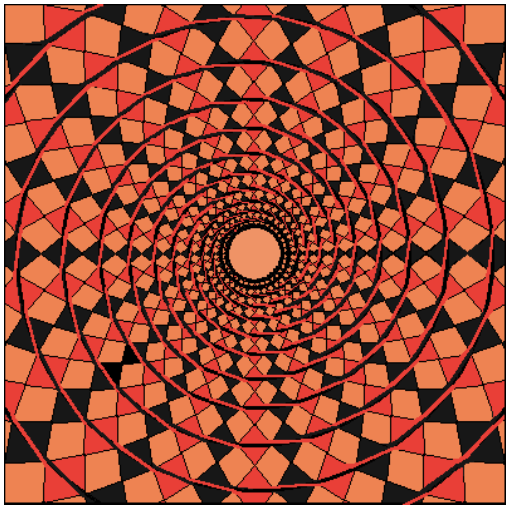
Copyright A.Kitaoka 2003



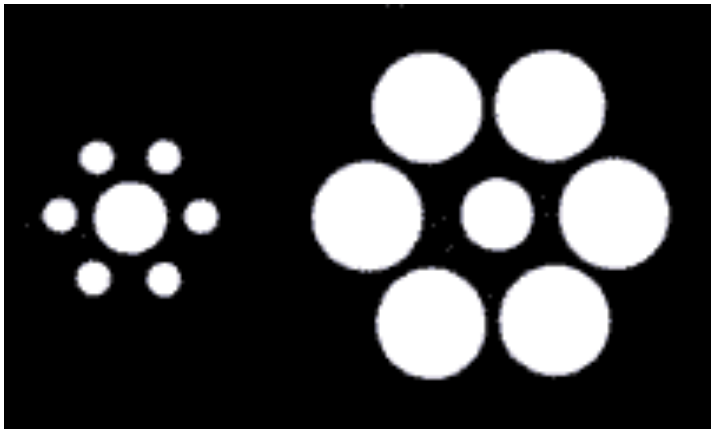
Are the horizontal lines parallel?



Are the purple lines straight?



Is this a spiral?



is the left circle (in the center) bigger?

Source: S. Maji

Why study computer vision?

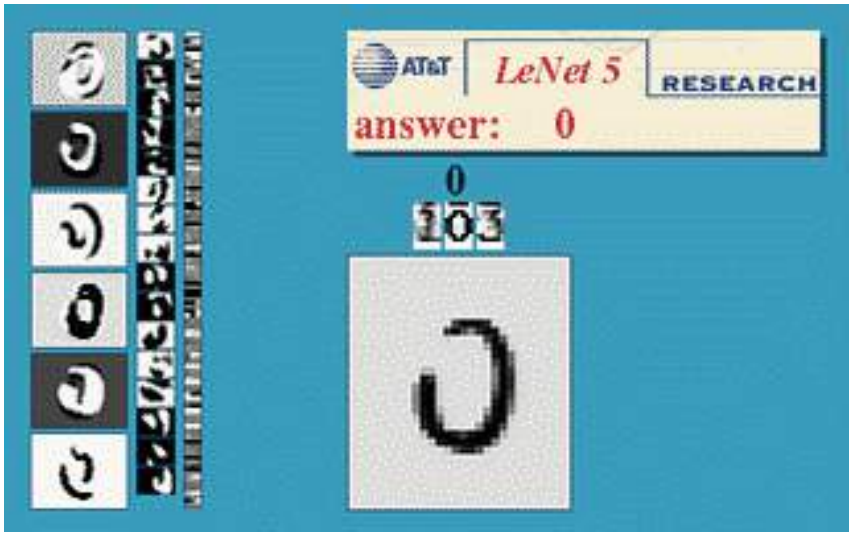
Images and videos are everywhere

Surveillance and security

Medical and scientific images

Source: Fei-Fei Li

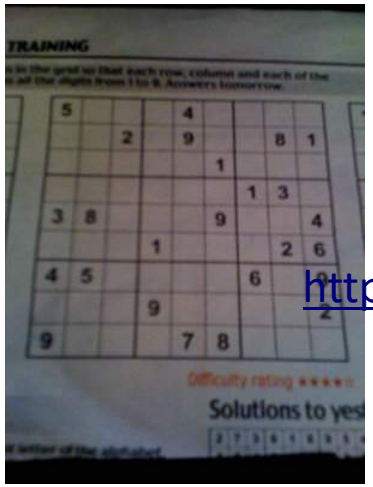
Optical character recognition (OCR)



Digit recognition
yann.lecun.com

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_rec

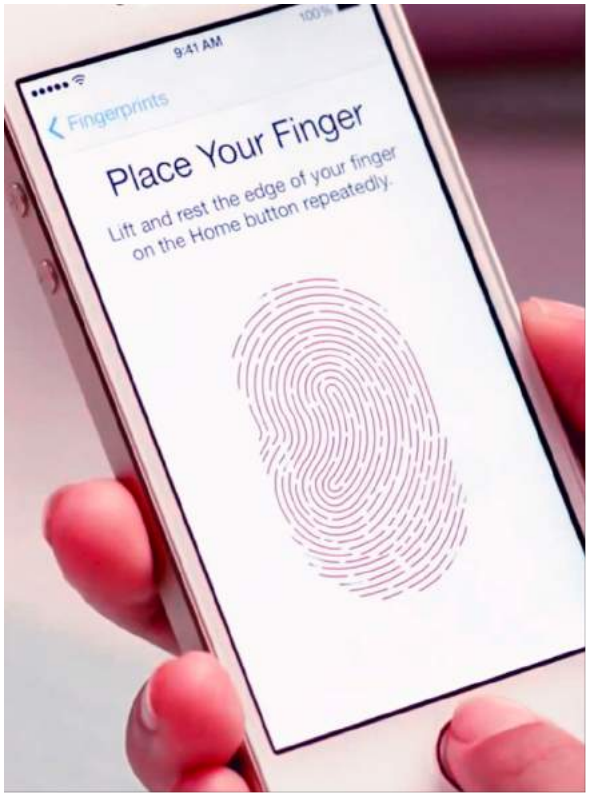
Automatic check processing



Sudoku grabber
http://sudokugrab.blogspot.co

Source: S. Seitz, N. Snavely

Biometrics



Fingerprint scanners on many devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Source: S. Seitz

Face detection

- Many consumer digital cameras now detect faces

Source: S. Seitz

Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.

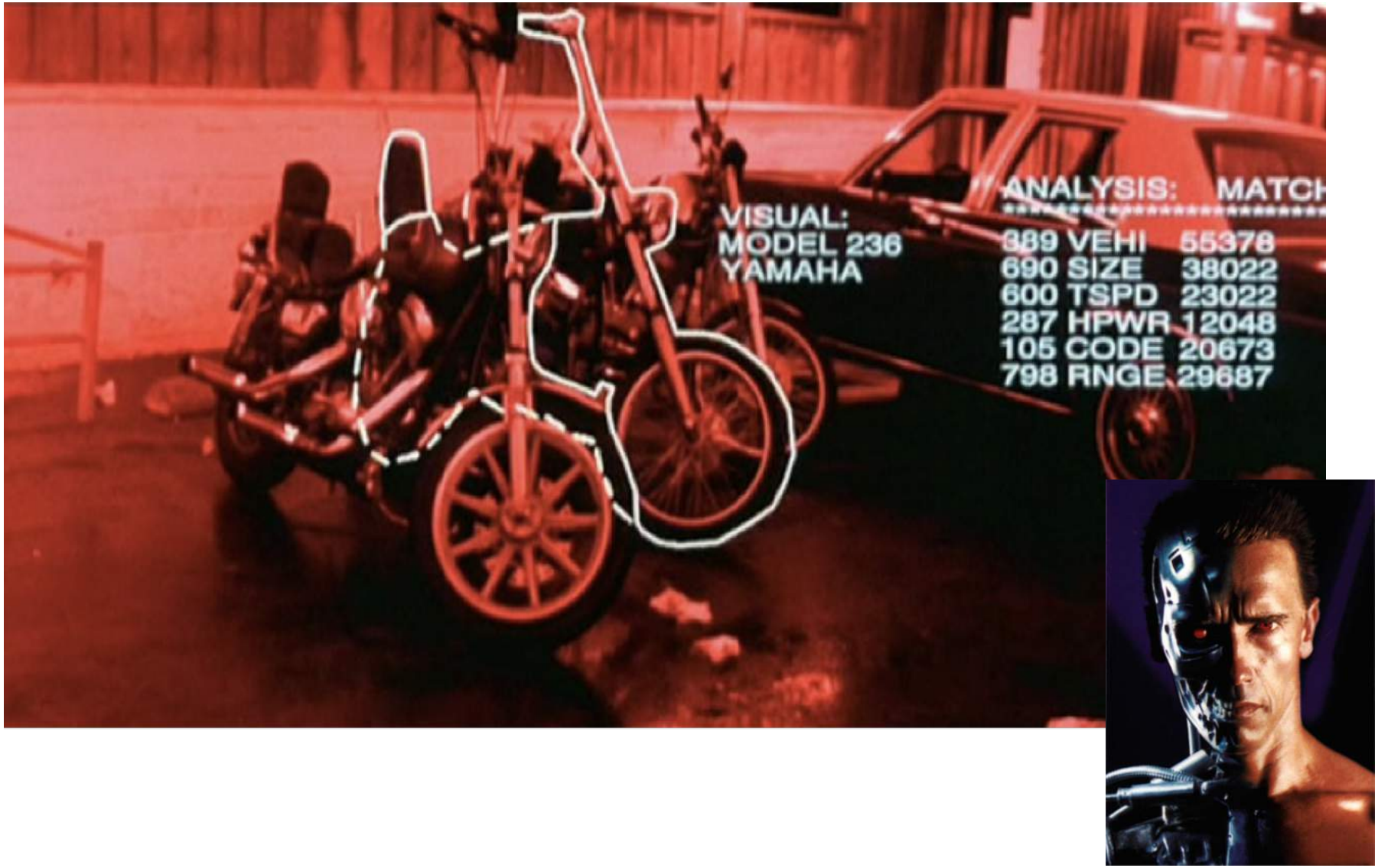
Sony Cyber-shot® T70 Digital Still Camera

Source: S. Seitz

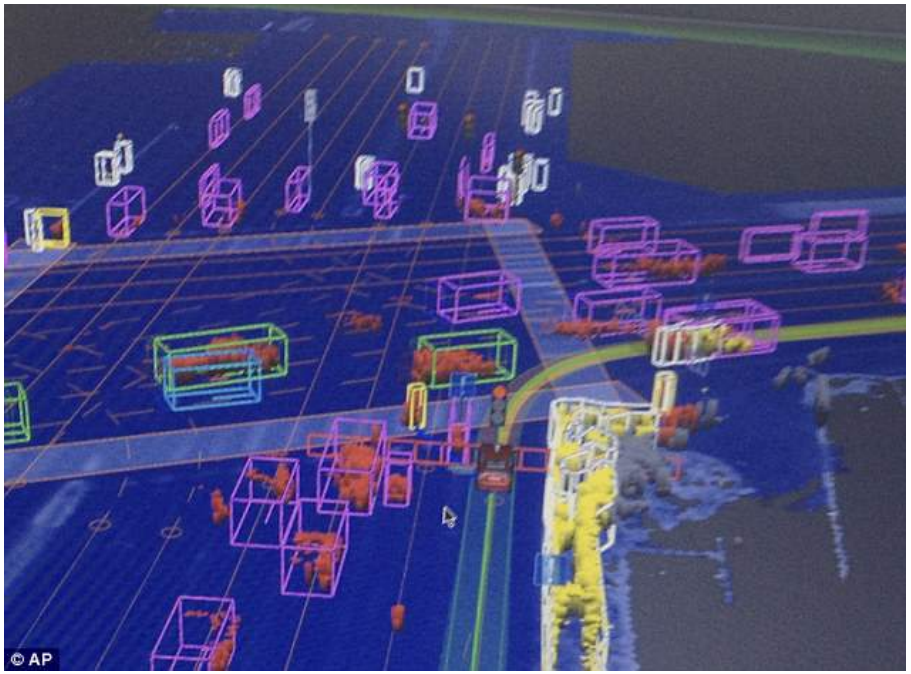
Face recognition: Apple iPhoto software

http://www.apple.com/ilife/iphoto/

Object Recognition



Self-driving cars

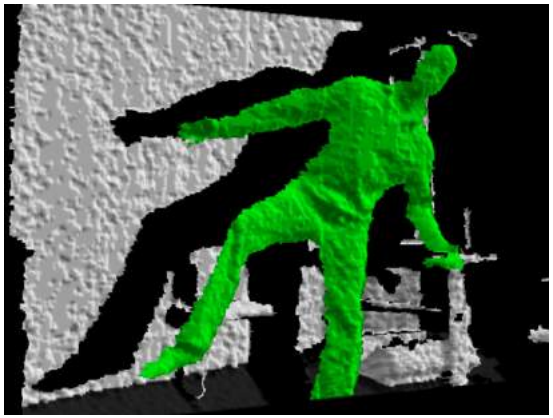


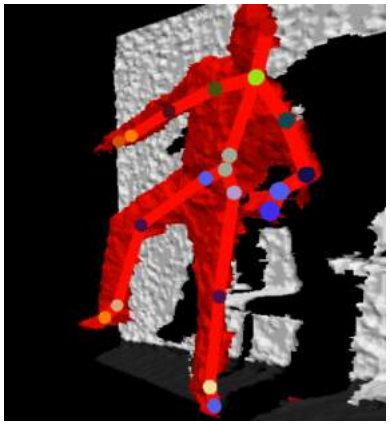
Automotive safety

- Mobileye: Vision systems in high-end BMW, GM, Volvo models
  - Pedestrian collision warning
  - Forward collision warning
  - Lane departure warning
  - Headway monitoring and warning

Source: A. Shashua, S. Seitz

Vision-based interaction: Xbox Kinect



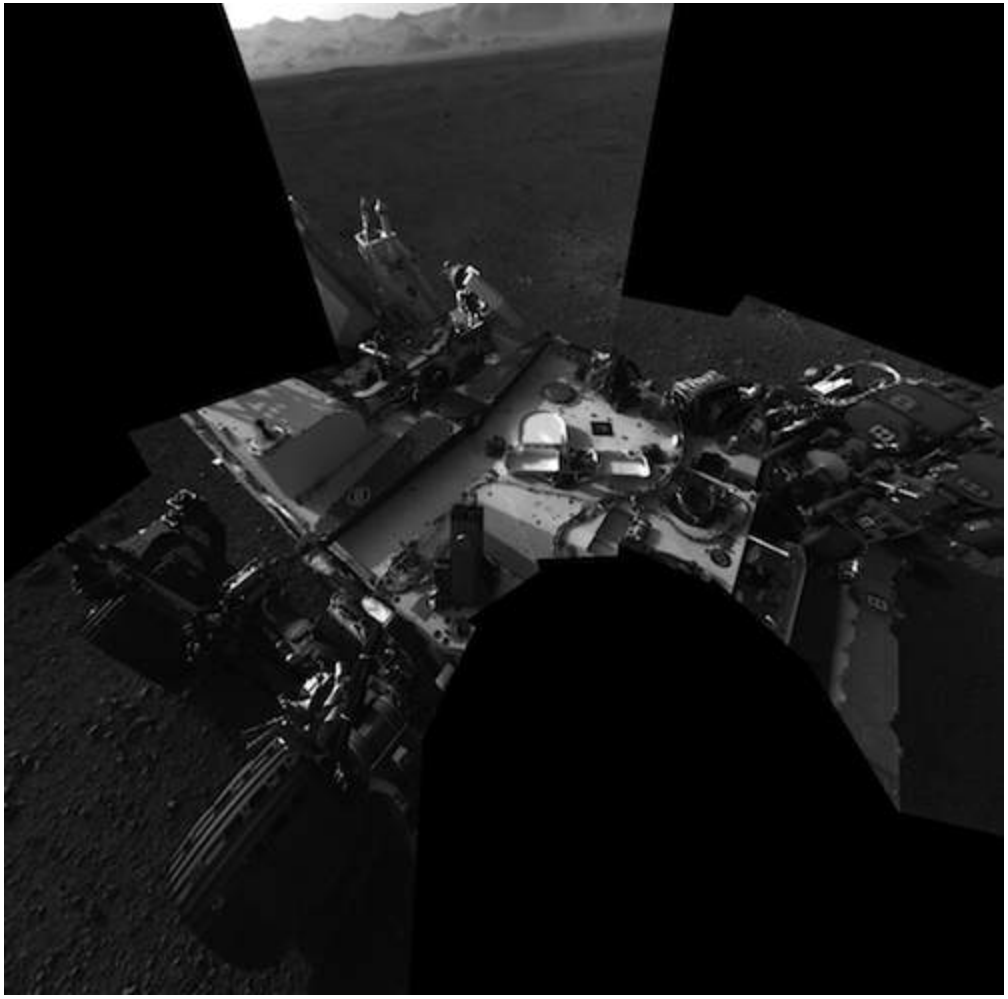


Source: S. Lazebnik

Special effects: shape and motion capture

Source: S. Seitz

Vision for robotics, space exploration



Source: S. Lazebnik

NASA'S Curiosity Rover has a system consisting of 17 cameras

Inspirational Applications

- Automatic Colorization of Black and White Images

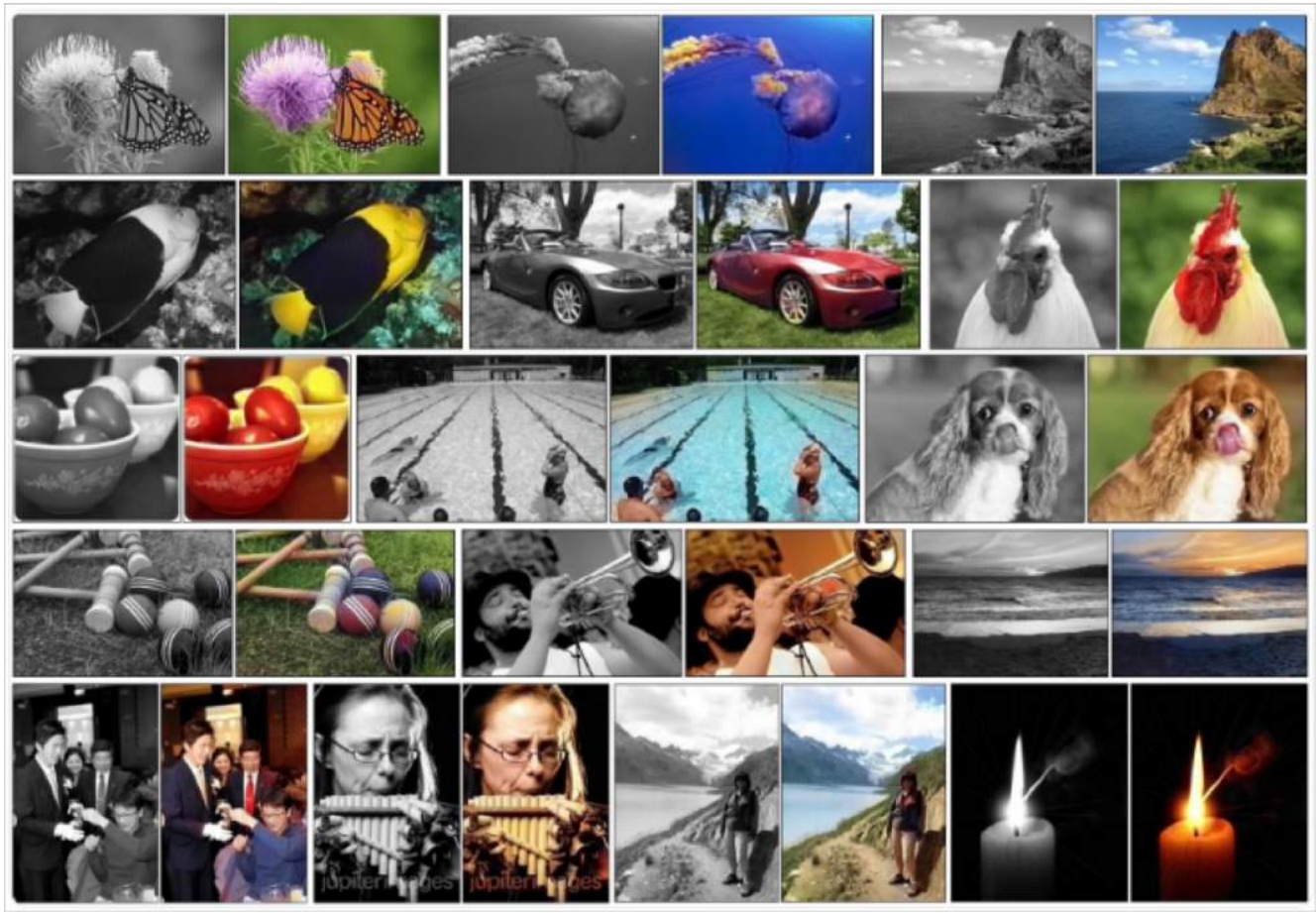
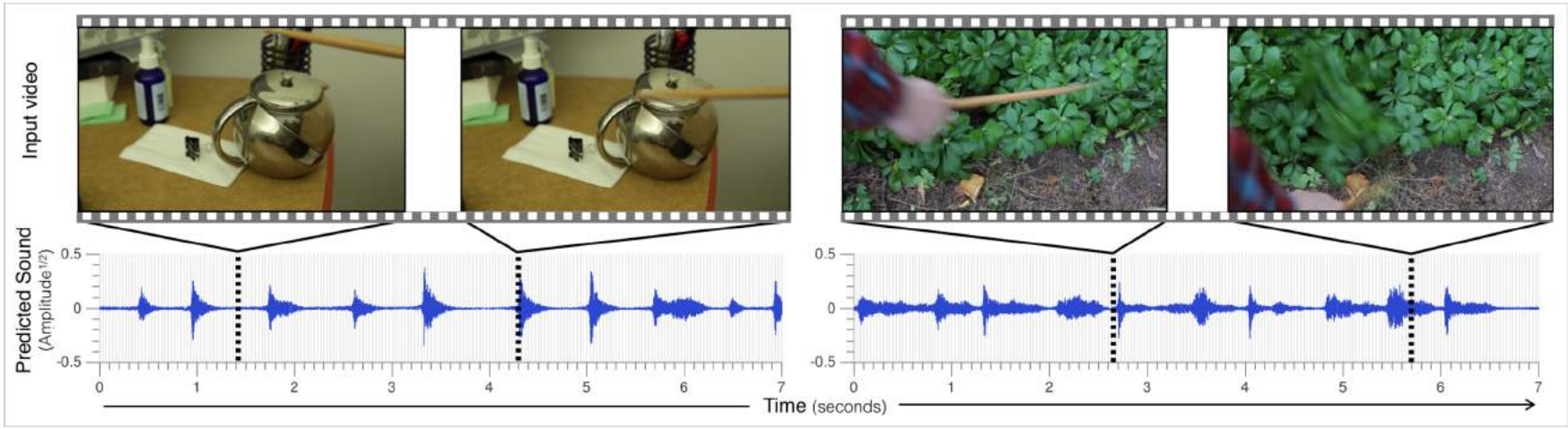


Image taken from Richard Zhang, Phillip Isola and Alexei A. Efros.

Inspirational Applications

- Automatically Adding Sounds To Silent Movies



[Owens et al. 2016]

Inspirational Applications

- Automatic Machine Translation

Image Credit: https://research.googleblog.com/2015/07/how-google-translate-squeezes-deep.html.

Inspirational Applications

- Object Classification and Detection in Photographs

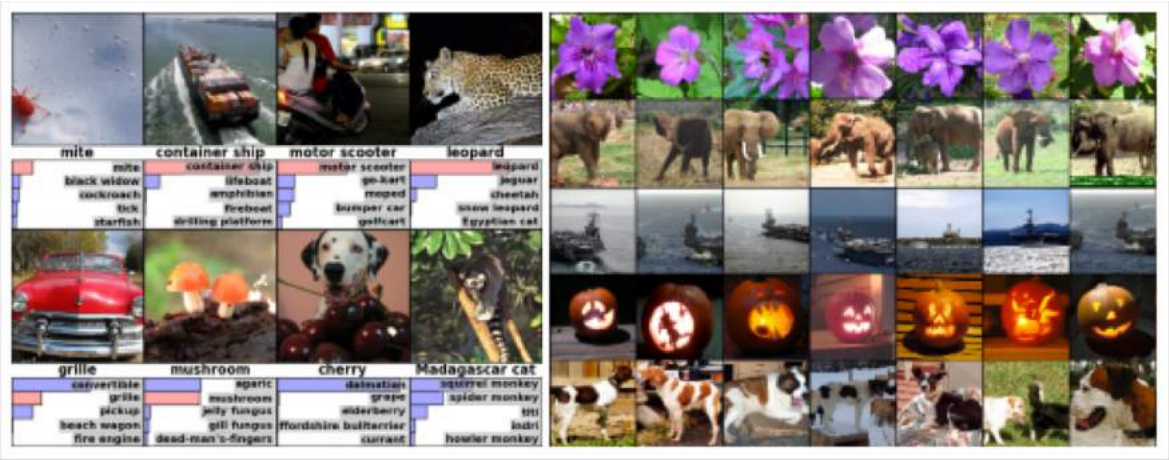


Image Credit: http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf.

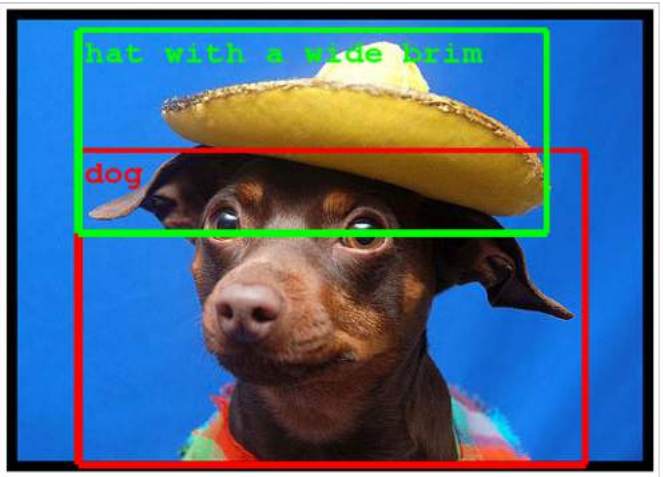


Image Credit: https://research.googleblog.com/2014/09/building-deeper-understanding-of-images.html.

Inspirational Applications

- Automatic Handwriting Generation

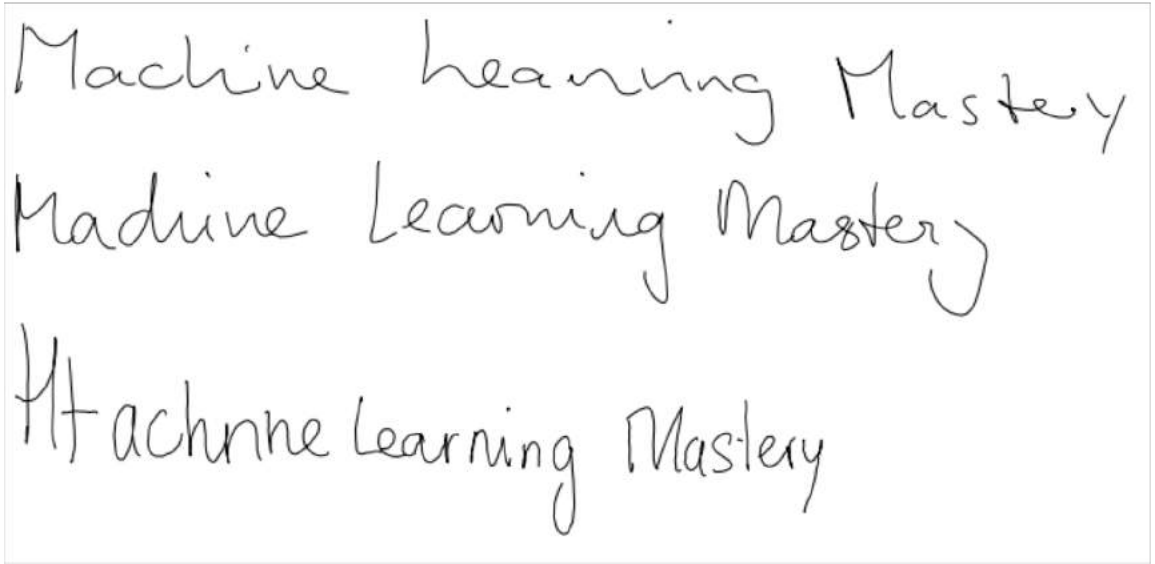


Image Credit: http://machinelearningmastery.com/inspirational-applications-deep-learning/.

Inspirational Applications

- Automatic Image Caption Generation

Image Credit: https://arxiv.org/pdf/1606.04621v3.pdf.

Inspirational Applications

- Image-To-Image Translation

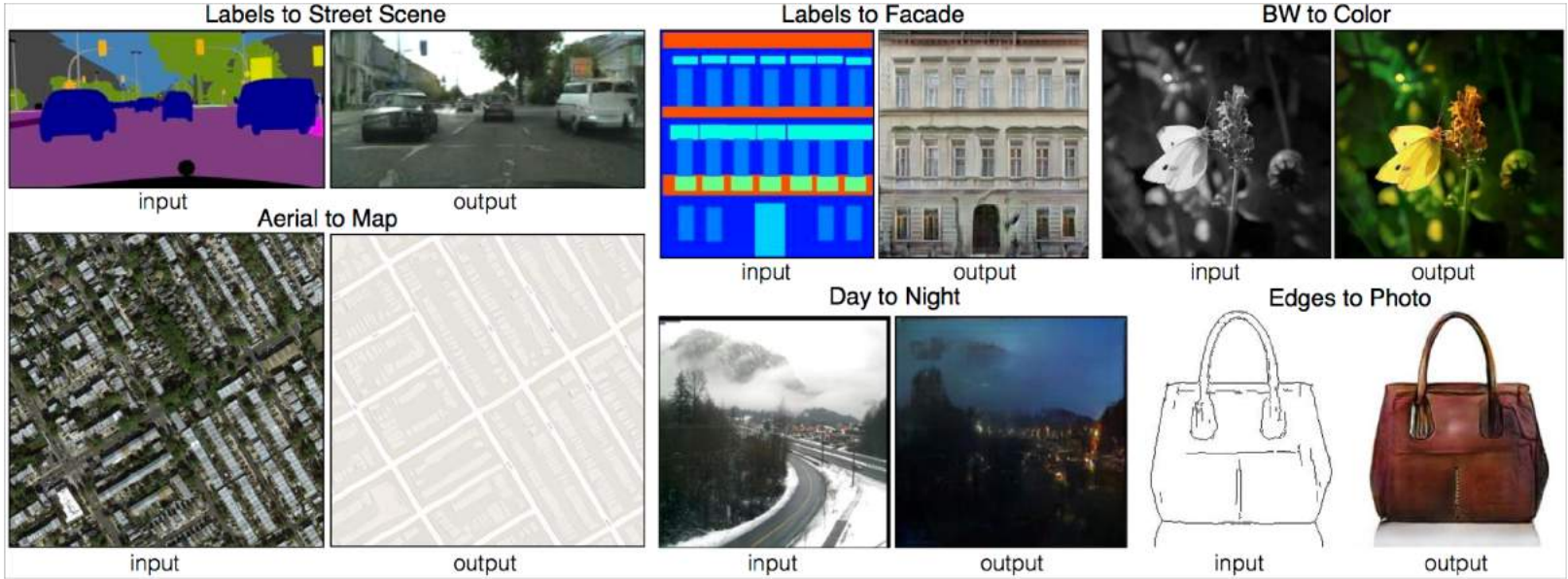


Image Credit: https://phillipi.github.io/pix2pix/.

The computer vision industry
CVPR 2017 Sponsors

SPONSORS

Platinum Sponsors

UISEE 驭势 · DEEPGLINT 格灵深瞳 · Apple · A9 · NVIDIA · Movidius · Microsoft · intel Nervana

Google · facebook · amazon · Tencent 腾讯 · NAVER · LINE

DiDi · Alibaba Group 阿里巴巴集团 · MOMENTA · NSF · SAMSUNG · Baidu 百度 · JD.COM 京东

lyft

Gold Sponsors

HISCENE 亮风台信息科技 · Mighty Ai · Adobe · CVTE · Dream-Future 视源股份 · iRobot · MALONG 码隆 · IBM

Sighthound · dji · MITSUBISHI ELECTRIC MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC · meitu · MMTLAB · Face++ 旷视 · here

SENSETIME · COGNEX · 图森 tuSimple · ARGO AI · nuTonomy · Snap Inc.

QUALCOMM · 360 www.360.cn · SAP · BOSCH Invented for life · Tencent AI Lab

The computer vision industry
CVPR 2017 Sponsors

Silver Sponsors

Kitware, DATATANG, TOYOTA RESEARCH INSTITUTE, UBER ATG, Pinterest, Disney Research

VionVision, AIMATTER, SIEMENS Healthineers, HRI Honda Research Institute USA, XILINX ALL PROGRAMMABLE.

Bronze Sponsors

Mapillary, BODY LABS, Orbital Insight, MathWorks, Lunit

Horizon Robotics, matterport, OMRON Sensing tomorrow™, Second Spectrum, SPORTLOGIQ

Zillow, ELSEVIER, FLIR, Springer, MVTec Software GmbH, King Abdullah University of Science and Technology, VCC VISUAL COMPUTING CENTER, ALTUM VIEW

LINKFACE, cruise, CloudSight™, gumgum, 唯品会 vip.com, DEEP VISION

SEGWAY ROBOTICS, 头条 TOUTIAO, Skydio, Yandex, ZOOX, Panasonic, AXON

NetPosa, NIO, Synaptics, ObjectVideo LABS, AiCure, 3dMD, playground

SIEMENS Ingenuity for life, CrowdFlower, CEPTON, GENIUS PROS 光珀智能, ef., CITADEL | Securities

The computer vision industry
CVPR 2017 Sponsors

Start-up Sponsors

wrnch
AutoX Democratizing autonomy
umbo cv
iniLabs
MORPX
PERCEPTIVE AUTOMATA
spotscale Buildings Decoded
deepcognai
TURING
FEATURE X
MLJKH TECHNOLOGIES
ANANTAK
水滴科技 WATRIX TECHNOLOGY
MARKABLE
EgoVid
eyeris
PIXM
Shopagon
OCTI
VanGogh Imaging
AUGMENTED PIXELS
SAIKOU OPTICS
SURFING TECH
ISEE
EYENUK
ARTISENSE
speechocean 海天瑞声

Non-Profit Sponsors

sama SOURCE
NEXT AI

Computer vision is hard!

Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination

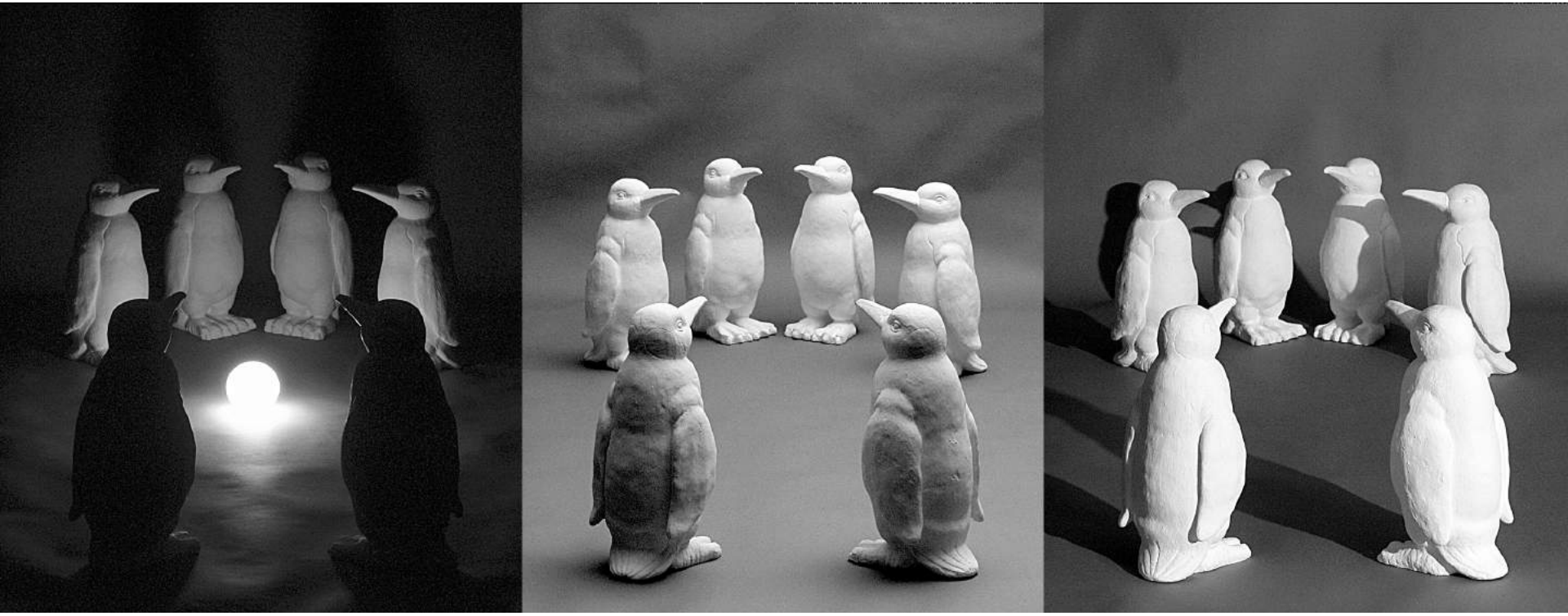
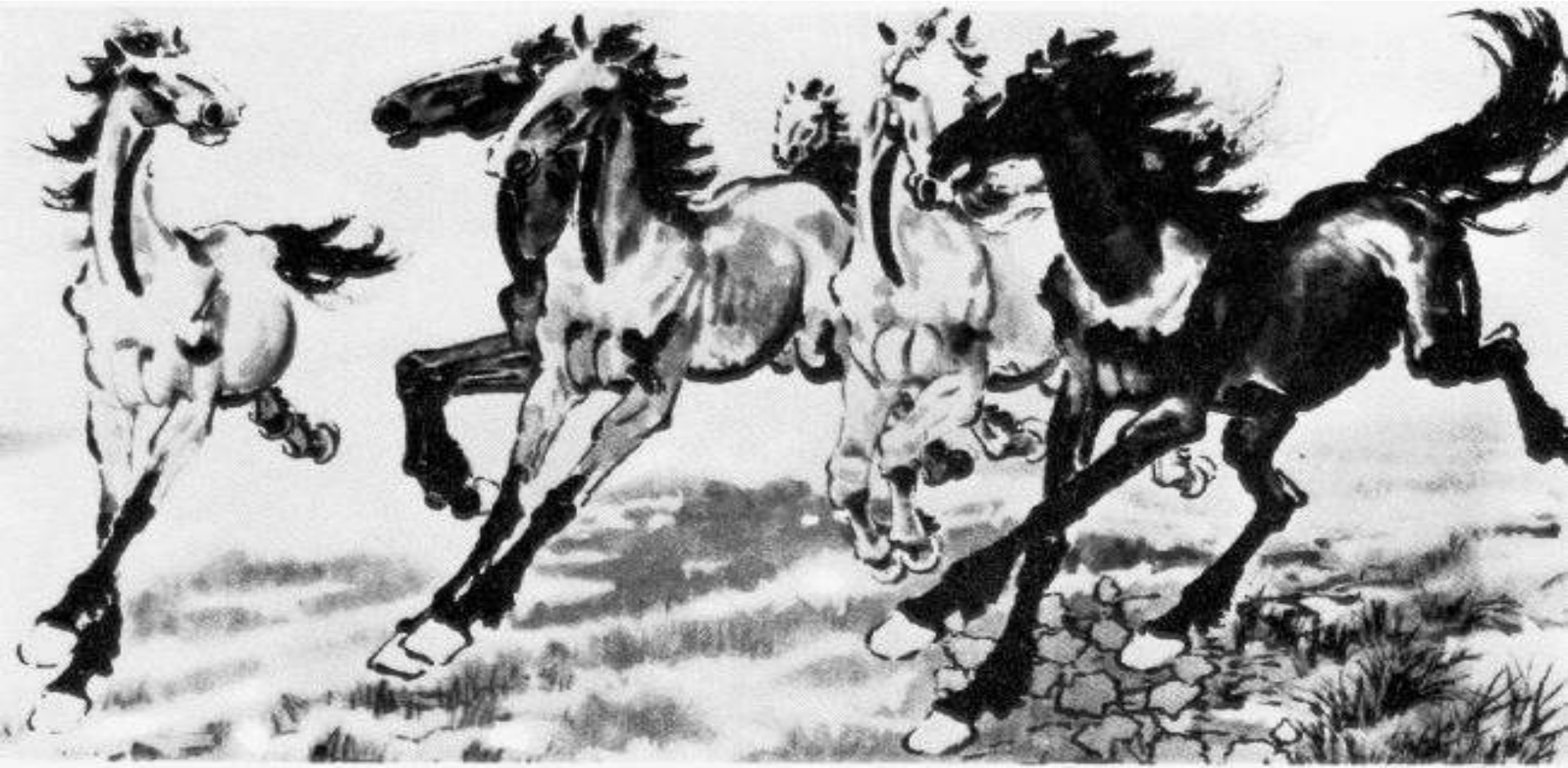


image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

Challenges: deformation



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

Challenges: object intra-class variation

slide credit: Fei-Fei, Fergus & Torralba

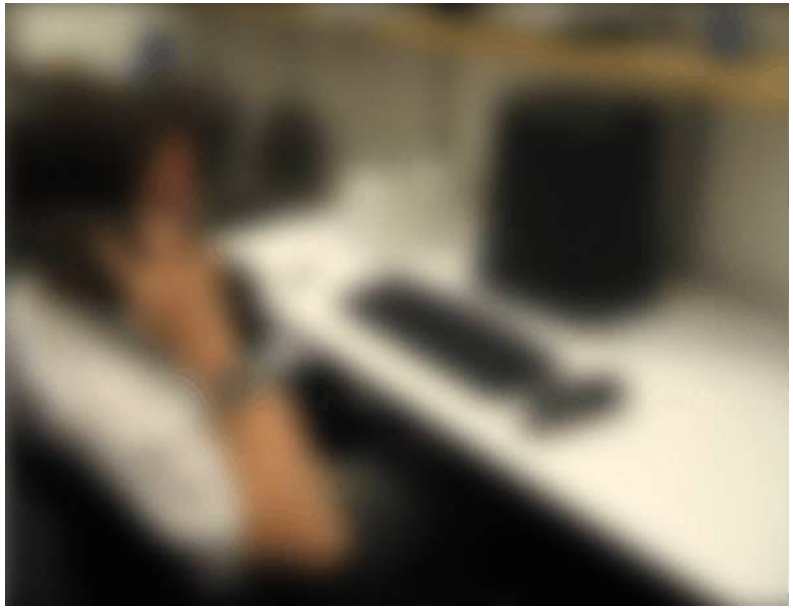
Challenges: occlusion, clutter

Image source: National Geographic
Slide : S. Lazebnik

Challenges: Motion

Source: S. Lazebnik

Challenges: ambiguity



Source: Rob Fergus and Antonio Torralba

Challenges: ambiguity

Source: Rob Fergus and Antonio Torralba

Challenges: ambiguity

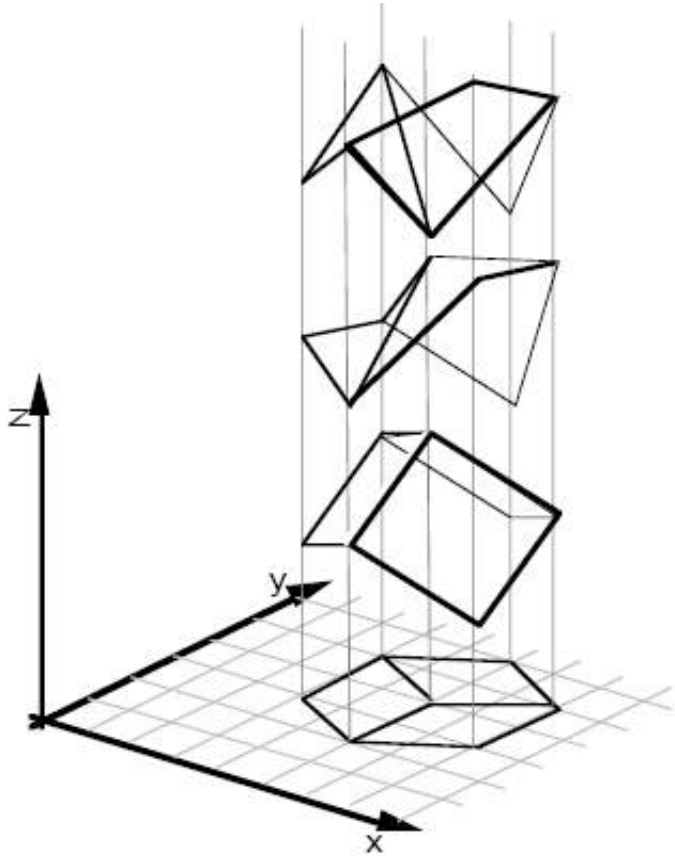
slide credit: Fei-Fei, Fergus & Torralba

Challenges: ambiguity

- Many different 3D scenes could have given rise to a particular 2D picture

Source: S. Lazebnik

Ambiguity



[Sinha and Adelson 1993]

Challenges or opportunities?

- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Source: S. Lazebnik

What will we study in this course?

Course Overview

- **Features and Filters:** image transforms, linear filtering, edges and corners, scale space, feature extraction, feature matching, optical flow

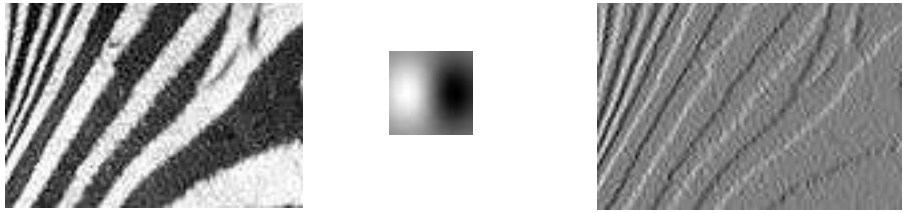
- **Segmentation and Grouping:** Gestalt principles, segmentation by clustering, image as graphs, interactive segmentation, robust fitting and RANSAC, Hough transform

- **Recognition:** feature learning, convolutional and recurrent neural networks, advanced CNN architectures, attention mechanisms, feature visualization, image classification, object detection, instance segmentation, scene parsing, image captioning, visual and sound, video analysis

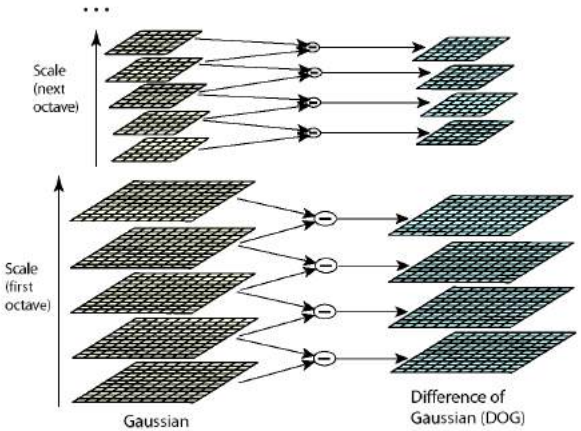
- **Image Formation and Multiple View Geometry**: camera models, light and color, camera calibration, alignment, stereo vision

Features and Filters

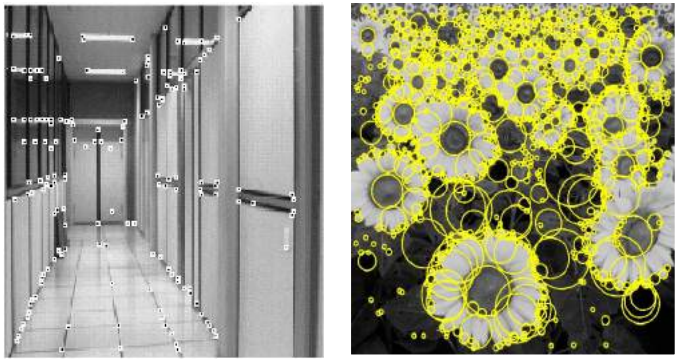
- Image transforms, linear filtering, edges and corners, scale space, feature extraction, feature matching, optical flow



Linear filtering, Edge detection



Scale space & Scale invariant features

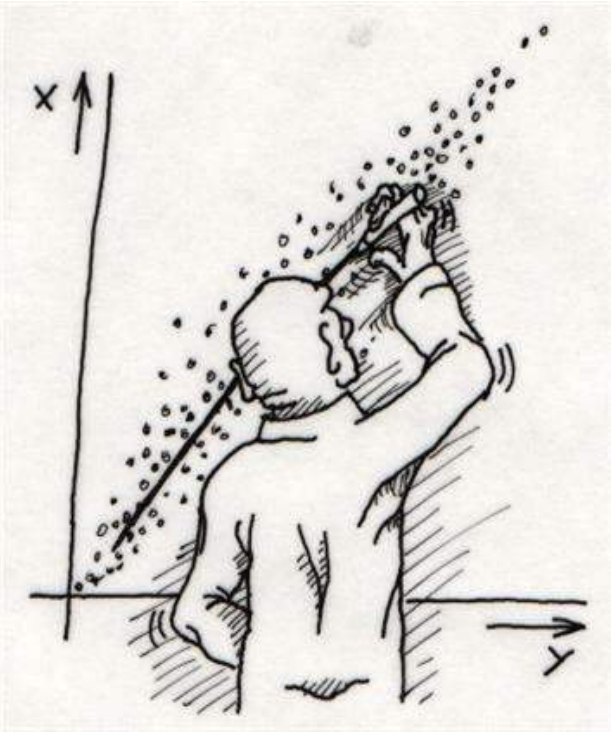


Feature extraction: corner and blob detection

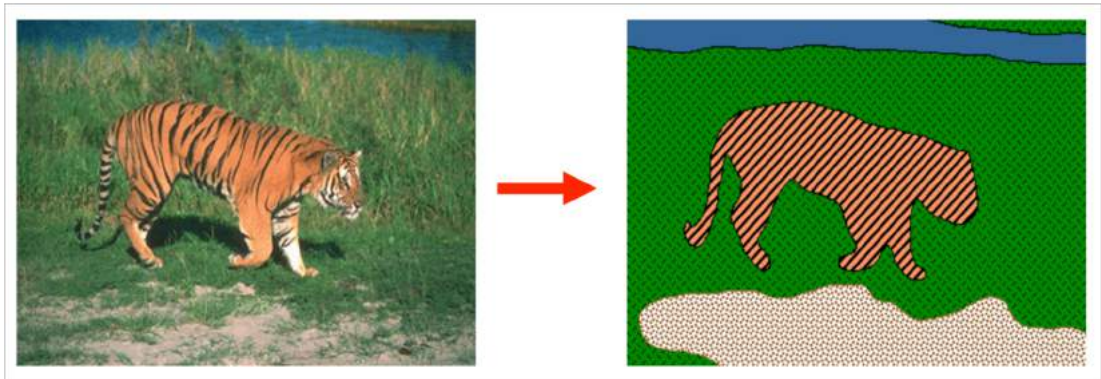
Motion & Optical flow

Segmentation and Grouping

- Gestalt principles, segmentation by clustering, image as graphs, interactive segmentation, robust fitting and RANSAC, Hough transform



Fitting: Least squares
Hough transform
RANSAC



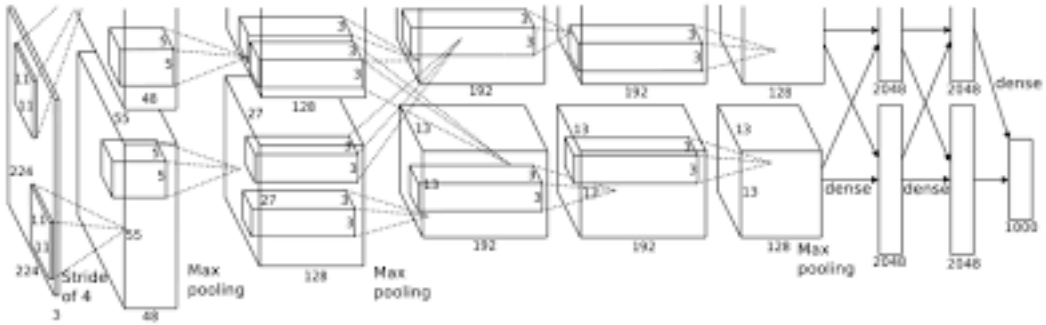
Segmentation by clustering, graph partitioning

Gestalt principles

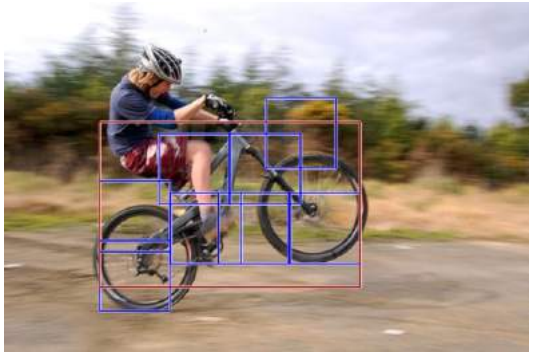
Recognition

- Feature learning, convolutional and recurrent neural networks, advanced CNN architectures, attention mechanisms, feature visualization, image classification, object detection, instance segmentation, scene parsing, image captioning, visual and sound, video analysis

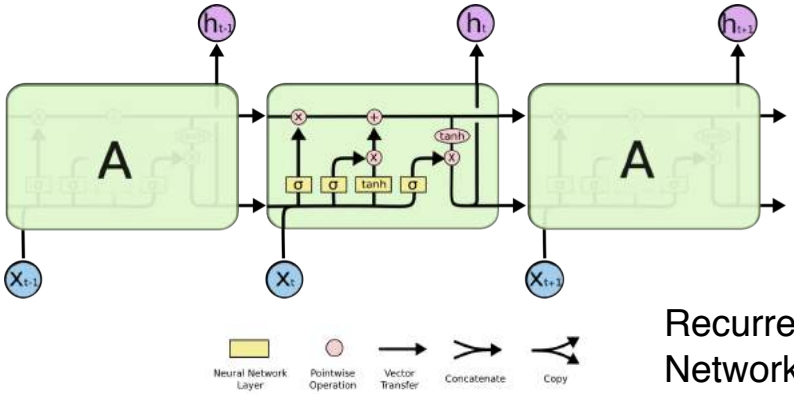
Image Classification



Convolutional Neural Networks



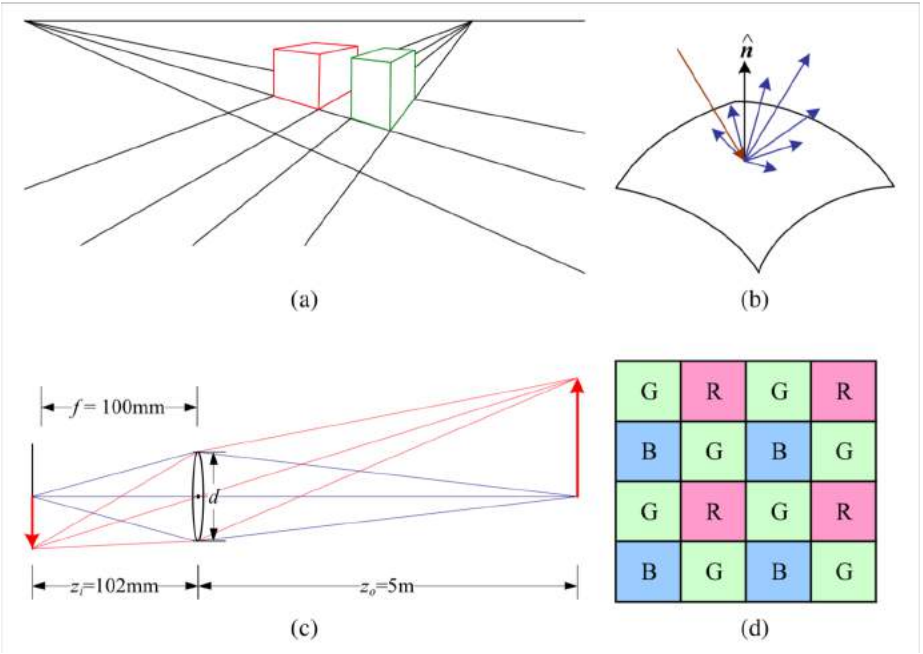
Object detection



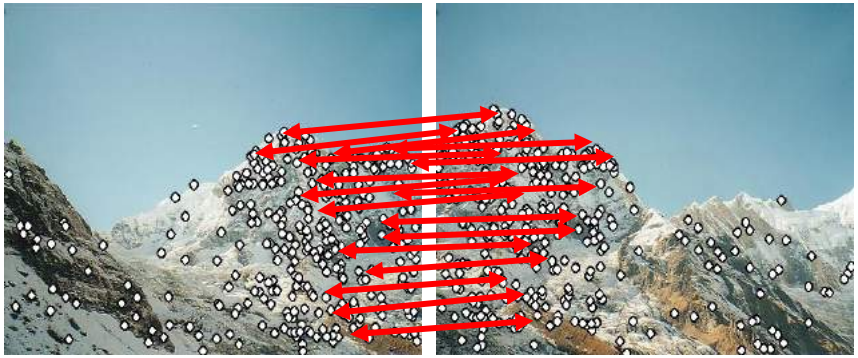
Recurrent Neural Networks

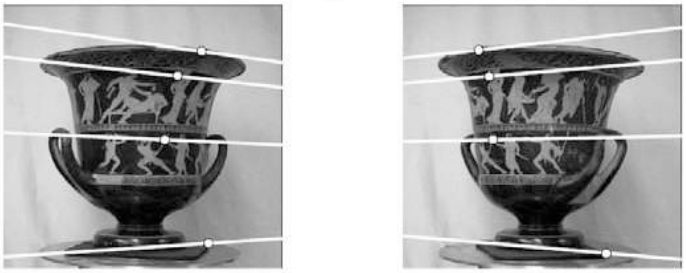
Image Formation and Multiple View Geometry

- Camera models, light and color, camera calibration, alignment, stereo vision



A few components of the image formation process: (a) perspective projection; (b) light scattering when hitting a surface; (c) lens optics; (d) Bayer color filter array.





The Course

- Course website:
  - http://www.cs.rochester.edu/~cxu22/t/249S18/
  - Schedule, office hours, syllabus, policies, etc.

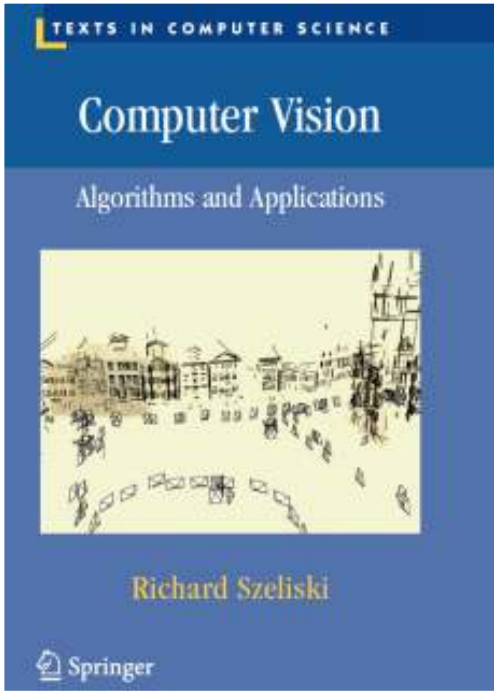
- Blackboard:
  - Announcements, posting lecture notes and assignments, and submitting work.

- Piazza:
  - https://piazza.com/rochester/spring2019/csc2492019spring52339/home
  - All course related discussions.

Textbooks

- Recommended book:
  - Computer Vision: Algorithms and Applications, Szeliski, Springer, 2011.
  - Free online: http://szeliski.org/Book/

- Additional books:
  - Computer Vision: A Modern Approach, 2nd Edition, Forsyth and Ponce, Pearson, 2012.
  - Deep Learning, Goodfellow, Bengio and Courville, The MIT Press, 2016.
  - Multiple View Geometry in Computer Vision, 2nd Edition, Hartley and Zisserman, Cambridge University Press, 2004.



Prerequisites

- Basic knowledge of:
  - probability and linear algebra;
  - data structures, algorithms;
  - programming experience.

- For assignments and projects, the most common programming language is Python.

Schedule

This is a **tentative** schedule of the course and will be modified over the semester.
Course materials will be posted on Blackboard.

Week	Tuesday (T)	Thursday (R)	HW Due	Project
1/14	No Class: Before Term	Introduction		
1/21	Linear Filtering and Image Transforms	Edges and Corners	T: 1 Out	
1/28	SIFT and Feature Matching	Neural Network and Backpropagation	R: 1 Due R: 2 Out	
2/4	CNN and Feature Visualization	Fitting		
2/11	Fitting	Alignment	R: 2 Due R: 3 Out	
2/18	Motion and Tracking	Motion and Tracking		
2/25	Segmentation	Segmentation		
3/4	RNN and Image Captioning	RNN and Image Captioning, Midterm Review	R: 3 Due R: 4 Out	
3/11	No Class: Spring Break	No Class: Spring Break		Train/Val Out
3/18	Advanced CNNs	**Midterm Exam**		
3/25	Object Recognition	Object Recognition	R: 4 Due R: 5 Out	
4/1	Generative Adversarial Networks and Image Synthesizing	Visual and Sound		
4/8	Video Analysis	Video Analysis		
4/15	Camera Models, Calibration	Light and Color		T:Test Out, R:Submit Due
4/22	Stereo Vision	**Project Presentation**		
4/29	**Project Presentation**	No Class: Term Finished	T: 5 Due	R: Report Due

Course Work and Evaluation

- **Homework Assignments (50%):**
  There will be five homework assignments over the term. Each will be a mix of analytical problems and programming problems to test whether students are understanding the foundational topics. They will be done independently by each student.

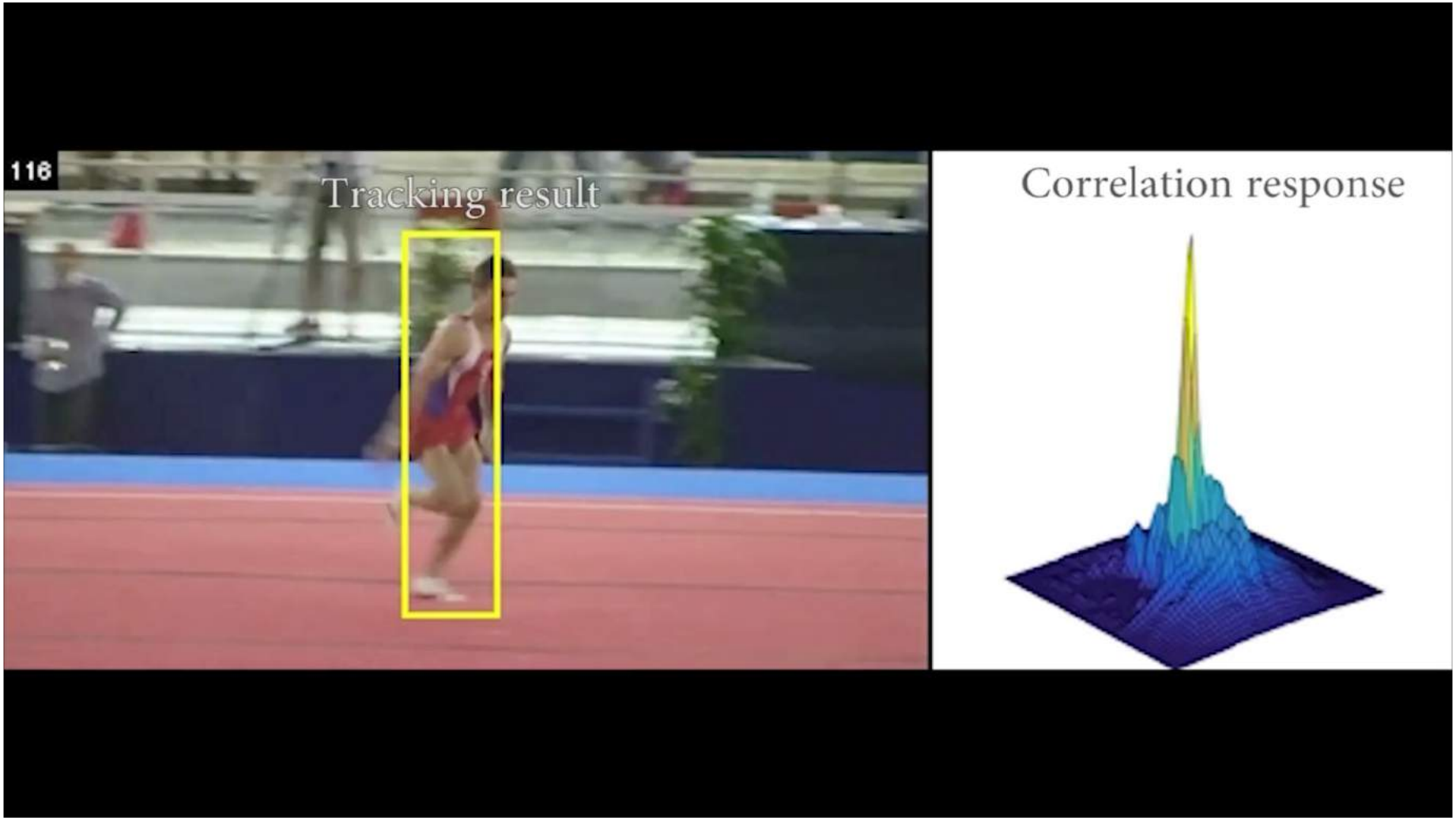
- **Midterm Exam (20%):**
  There will be one in-class midterm exam in March.

- **Project & Presentation (30%):**
  Students will form groups with up to two members to implement a course project. Full-scale project developments will be made through the term culminating in an oral/poster presentation and a report at the end of the term.

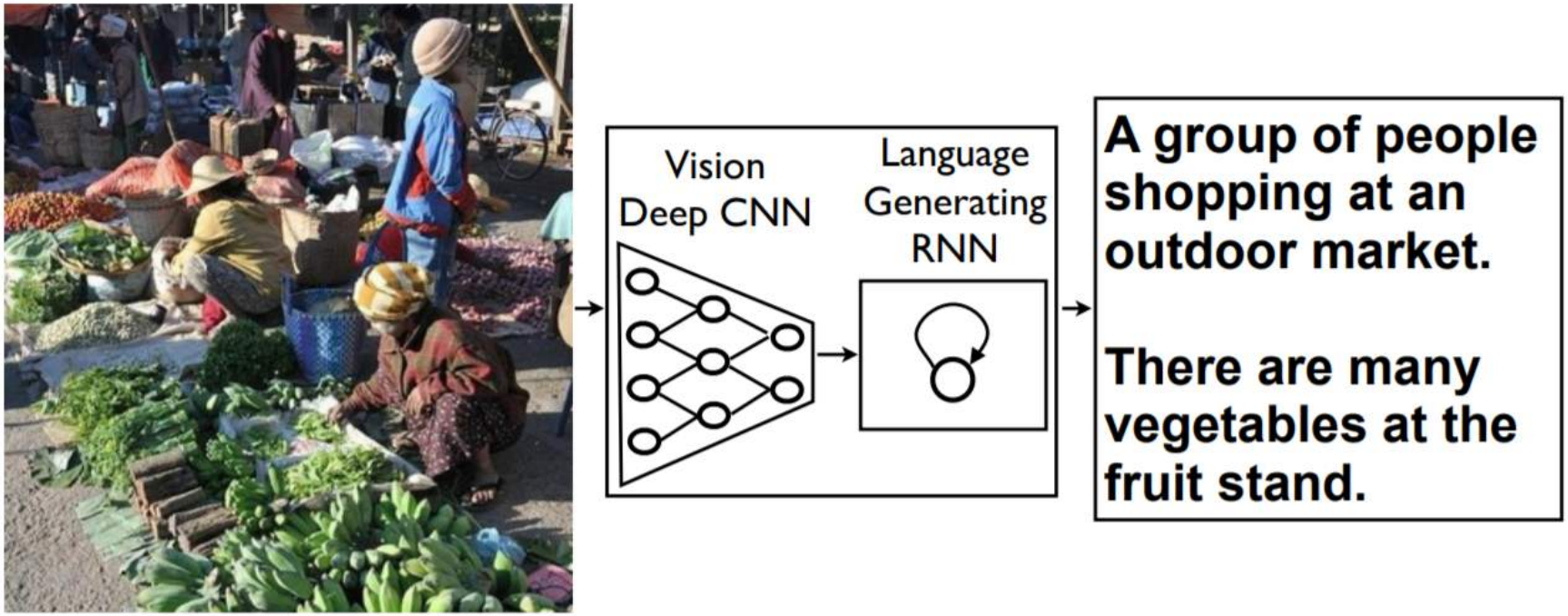
Preview of Assignments

- Programming components of planned assignments.
- Homework 1 and 2
  - Basic operations: convolution, pooling, backpropagation etc.
- Homework 3
  - Implement an object tracking system based on correlation of CNN feature maps.
- Homework 4
  - Combine CNN and RNN to build a deep learning model that produces captions given an input image.
- Final Project
  - An open challenge to detect, segment, and recognize objects and actions in images.

Preview of Assignments: HW3



Preview of Assignments: HW4



Build an image captioning network with a CNN visual encoder and a RNN language decoder

Figure credit: Vinyals et. al 2015

Preview of Assignments: Final Project

Course Work and Evaluation

- **Late Work Policy:**
  - Lose 1% per hour passing the deadline for each assignment.
  - We will not accept your submission after 72 hours over the deadline.
  - All assignments should be submitted on Blackboard.

- **Regrading:**
  - Any questions about the grading of a piece of work must be raised within one week of the date that the work was returned.

Others

- **Code of Conduct:**
  Check here: http://www.cs.rochester.edu/~cxu22/t/coc.txt for classroom "etiquette" expectations.

- **Academic honesty:**
  All assignments and activities associated with this course must be performed in accordance with the University of Rochester's Academic Honesty Policy. More information is available at here: http://www.rochester.edu/college/honesty/. You are expected to read, understand and follow the policy.

- **Disabilities:**
  If you need an accommodation based on the impact of a disability, please contact me privately during the first two weeks to discuss your specific needs. Alternatively, contact Center for Excellence in Teaching and Learning at 585-275-9049 in room 1-154 Dewey Hall to coordinate reasonable accommodations.

What's next?

- Familiar yourself with Python programming.
- Read textbook Sec. 1.1-1.3

- Next lecture:
  - Linear filters and image transforms
  - Read textbook Sec. 3.1.1-2, 3.2, 3.3.1, 3.4-3.4.2