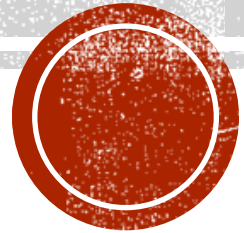
Indian Institute of Information Technology, Allahabad

Visual Recognition



By

Dr. Satish Kumar Singh & **Dr. Shiv Ram Dubey**
Computer Vision and Biometrics Lab
Department of Information Technology
Indian Institute of Information Technology, Allahabad

TEAM

Computer Vision and Biometrics Lab (CVBL)

Department of Information Technology

Indian Institute of Information Technology Allahabad

Course Instructors

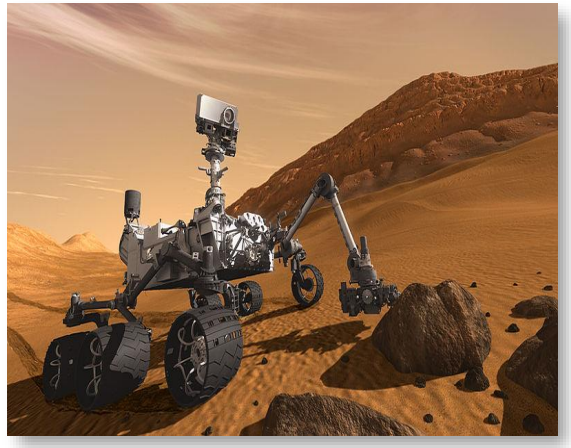
Dr. Satish Kumar Singh, Associate Professor, IIIT Allahabad (Email: sk.singh@iiita.ac.in)

Dr. Shiv Ram Dubey, Assistant Professor, IIIT Allahabad (Email: srdubey@iiita.ac.in)

DISCLAIMER

The content (text, image, and graphics) used in this slide are adopted from many sources for Academic purposes. Broadly, the sources have been given due credit appropriately. However, there is a chance of missing out some original primary sources. The authors of this material do not claim any copyright of such material.

VISUAL RECOGNITION



THIS CLASS

1. Introduction
2. Visual Recognition
3. Course Logistics

WHAT IS VISUAL RECOGNITION?

- Make computers understand images and videos.

- What kind of scene?
- Where are the cars?
- How far is the building?

EVERY IMAGE TELLS A STORY

EVERY IMAGE TELLS A STORY

- Goal of computer vision: perceive the "story" behind the picture

EVERY IMAGE TELLS A STORY

- Goal of computer vision: perceive the "story" behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

HOW COMPUTER SEE IMAGES?

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

CAN THE COMPUTER MATCH HUMAN PERCEPTION?

CAN THE COMPUTER MATCH HUMAN PERCEPTION?

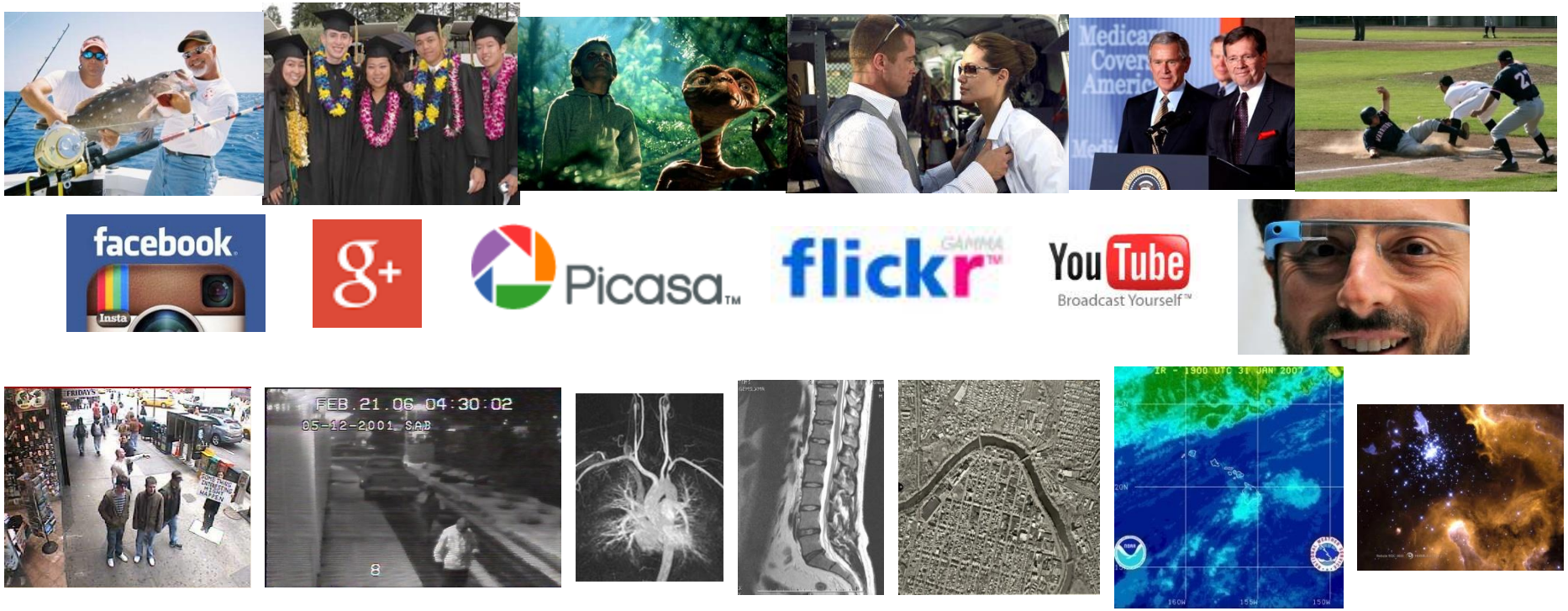
- Yes and no (mainly no)
  - computers can be better at "easy" things
  - humans are much better at "hard" things

CAN THE COMPUTER MATCH HUMAN PERCEPTION?

- Yes and no (mainly no)
  - computers can be better at "easy" things
  - humans are much better at "hard" things
- But huge progress has been made
  - Especially in the last 15 years
  - What is considered "hard" keeps changing

WHY STUDY VISUAL RECOGNITION?

- Millions of images being captured all the time



WHY IS VISUAL RECOGNITION HARD?

WHAT DID YOU SEE?

- Where this picture was taken?
- How many people are there?
- What are they doing?
- What object the person on the left standing on?
- Why this is a funny picture?

WHY IS VISUAL RECOGNITION HARD?

WHY IS VISUAL RECOGNITION HARD?

WHY IS VISUAL RECOGNITION HARD?

WHY IS VISUAL RECOGNITION HARD?

WHY IS VISUAL RECOGNITION DIFFICULT?

Viewpoint variation

Illumination

Scale

WHY IS VISUAL RECOGNITION DIFFICULT?

Intra-class variation

Motion (Source: S. Lazebnik)

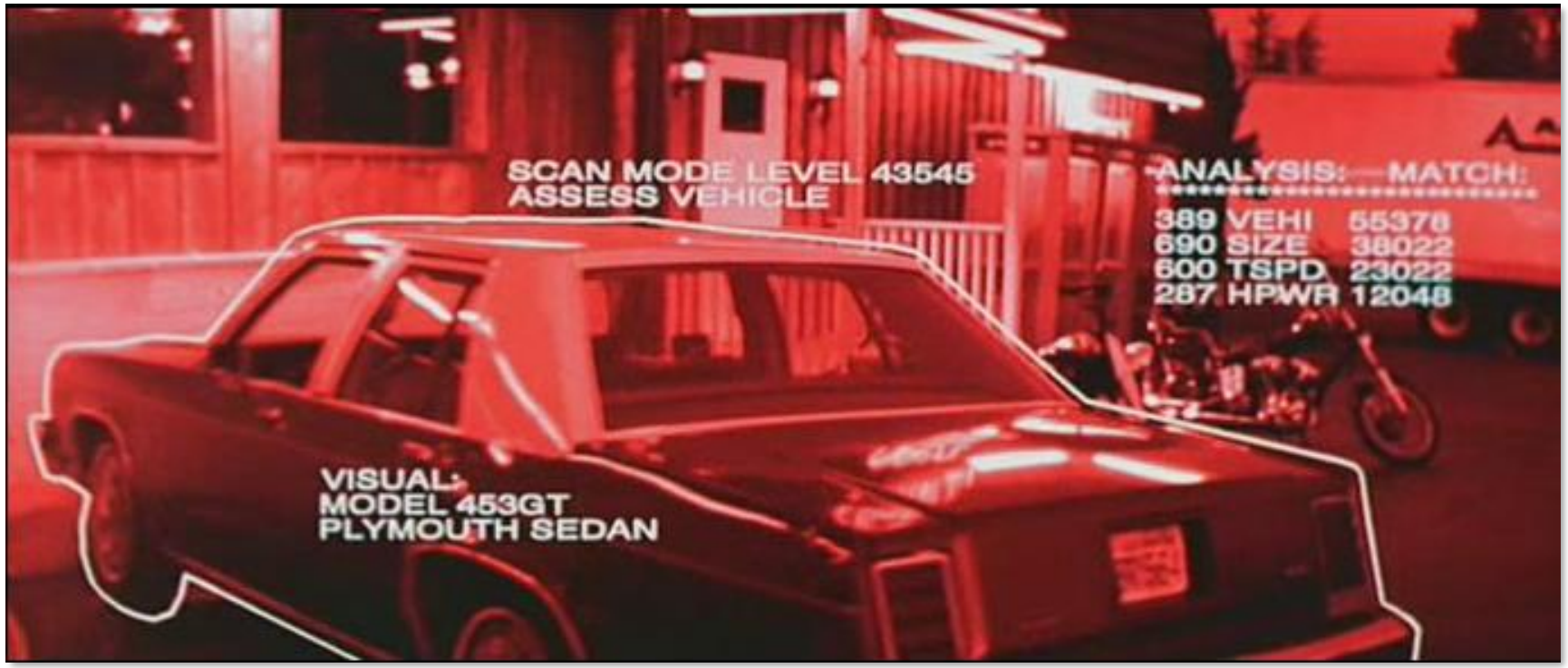
Background clutter

Occlusion

THE GOAL OF VISUAL RECOGNITION

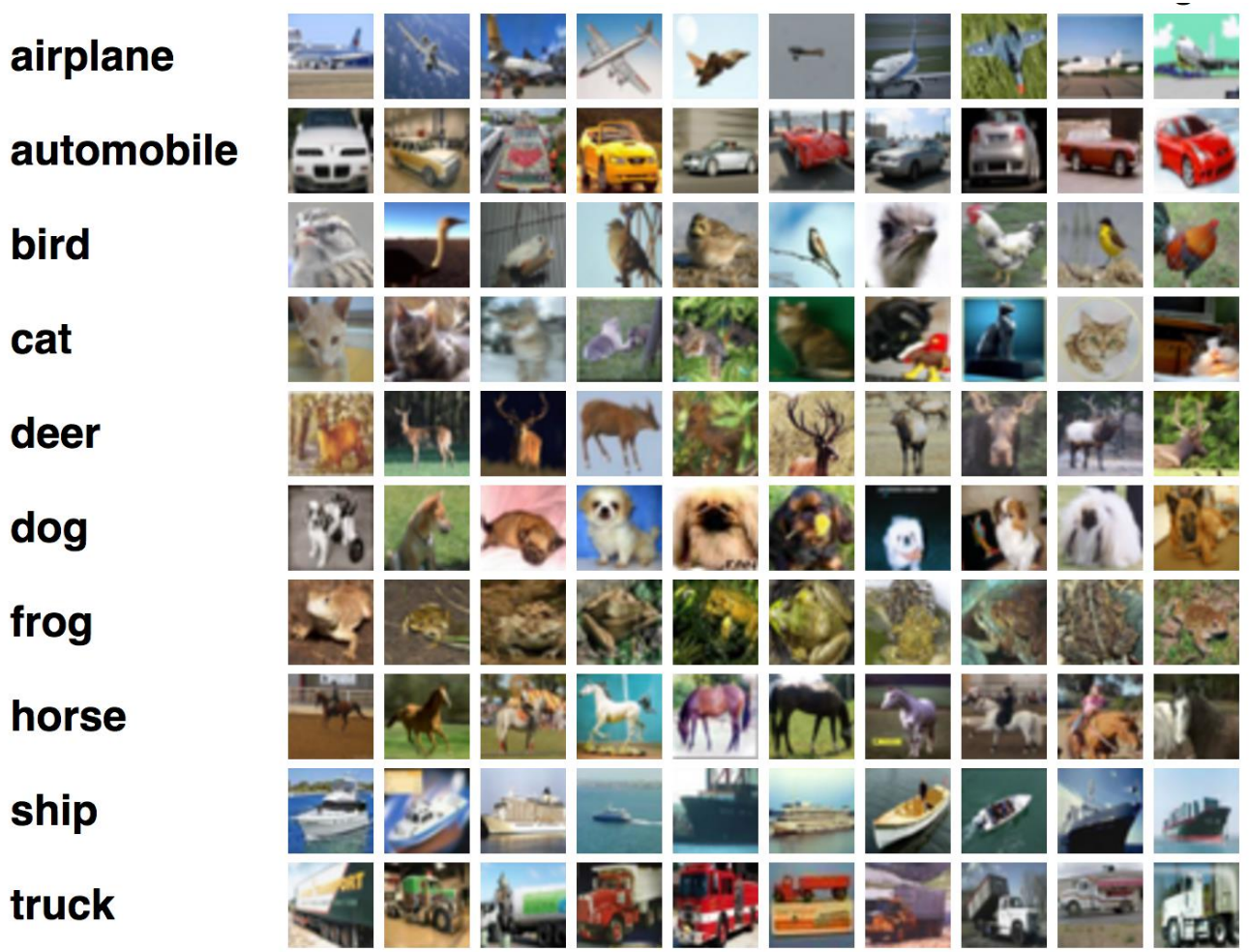
THE GOAL OF VISUAL RECOGNITION

- Recognizing objects and people



THE GOAL OF VISUAL RECOGNITION

- Classification of visual images/videos



Images From CIFAR dataset

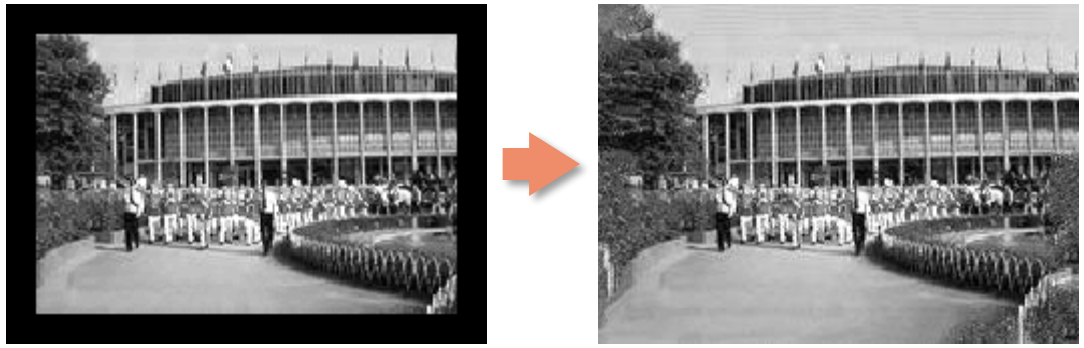

中华人民
和国万岁
世界
民大团结万岁


sky
building
flag
face
banner
wall
street lamp
bus
bus
cars
slide credit: Fei-Fei, Fergus & Torralba

THE GOAL OF VISUAL RECOGNITION

- "Enhancing" images (Computational Photography)

Super-resolution / denoising
(source: 2d3)



Texture synthesis / increased field of view
(uncropping) (image credit: Efros and Leung)

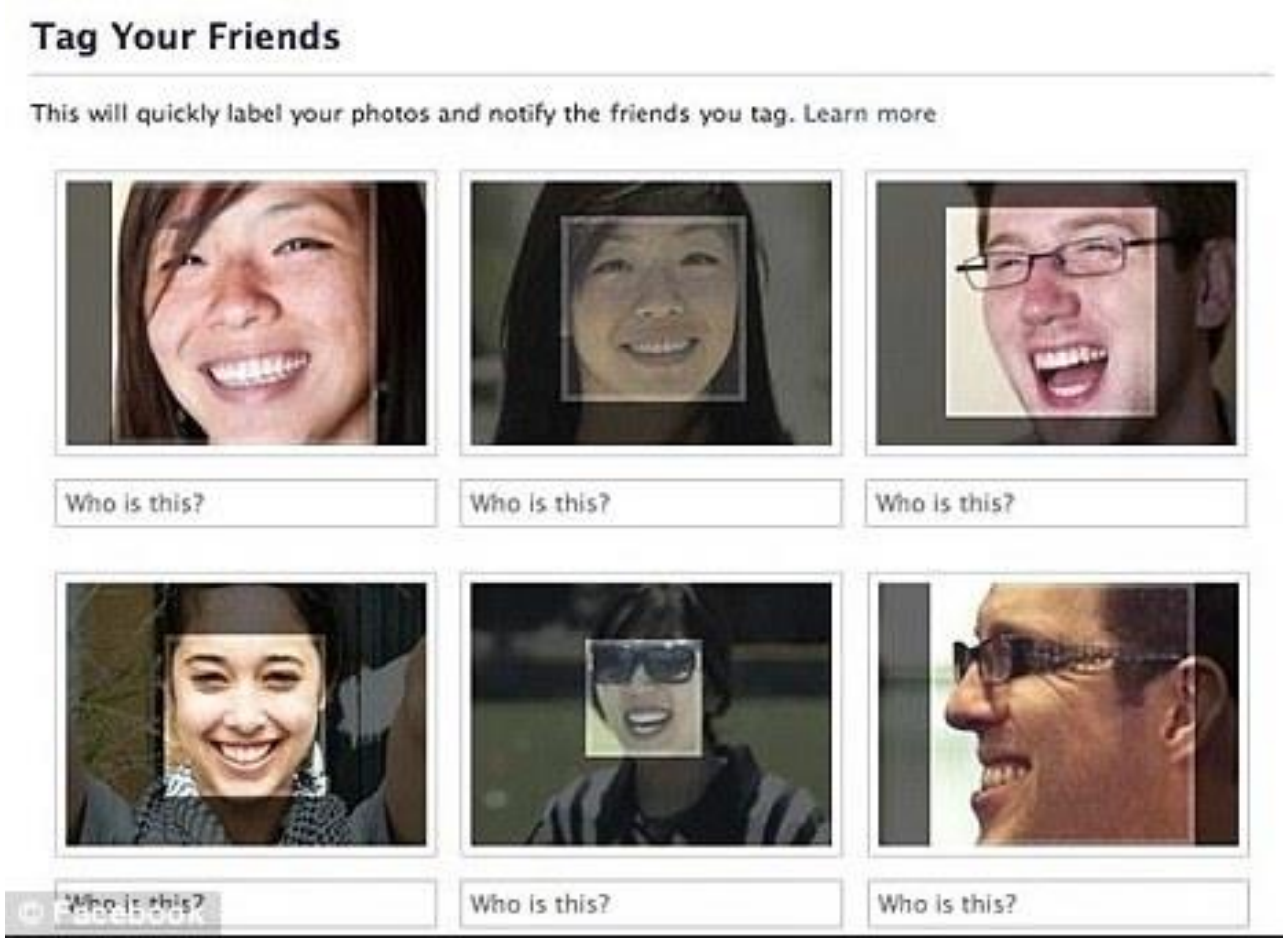
Inpainting / image completion
(image credit: Hays and Efros)

FACE DETECTION

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …
- Very challenging under severe imaging conditions
  - Pose, Light, Scale, …

Source: S. Seitz

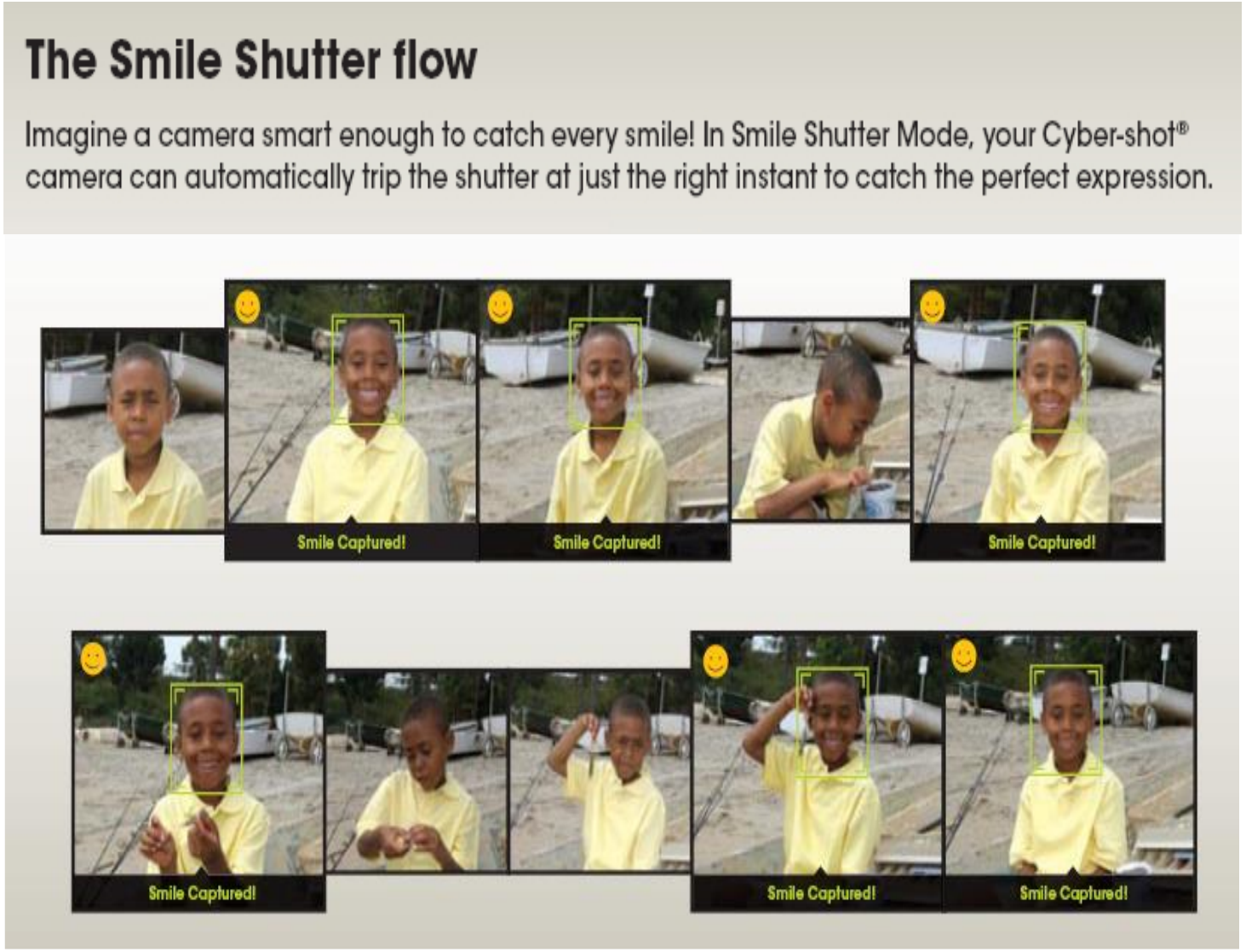
FACE RECOGNITION



Facebook face auto-tagging

Google photo

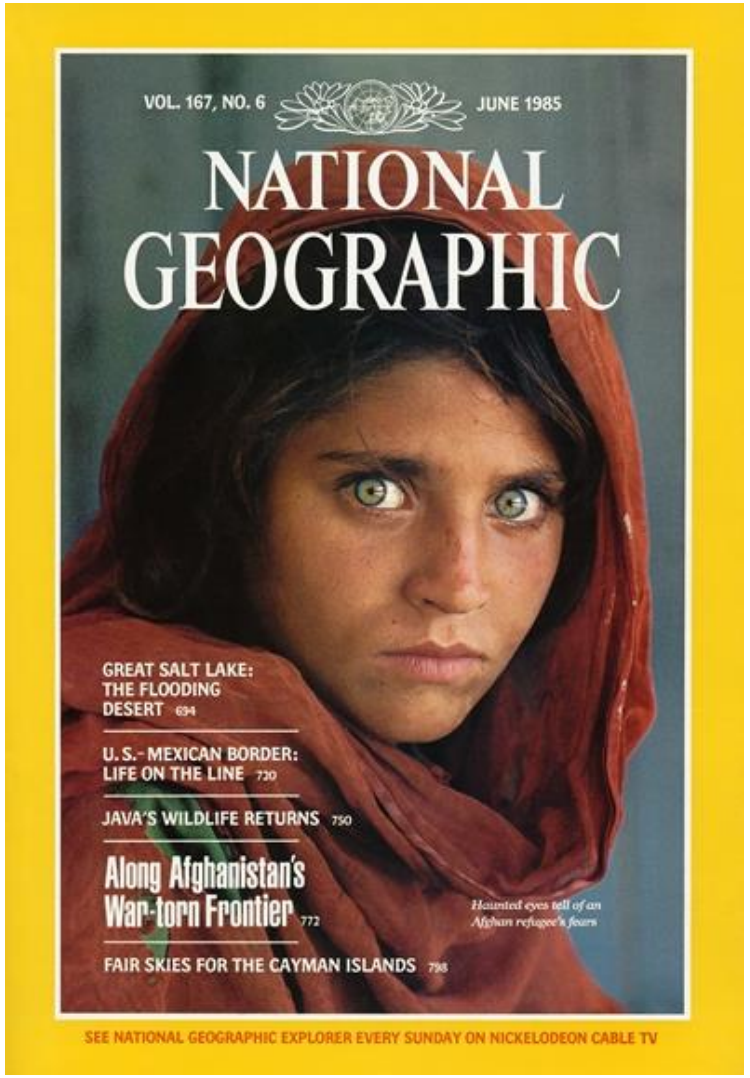
SMILE RECOGNITION



[Sony Cyber-shot® T70 Digital Still Camera]

Slide credit: Steve Seitz

BIOMETRIC RECOGNITION

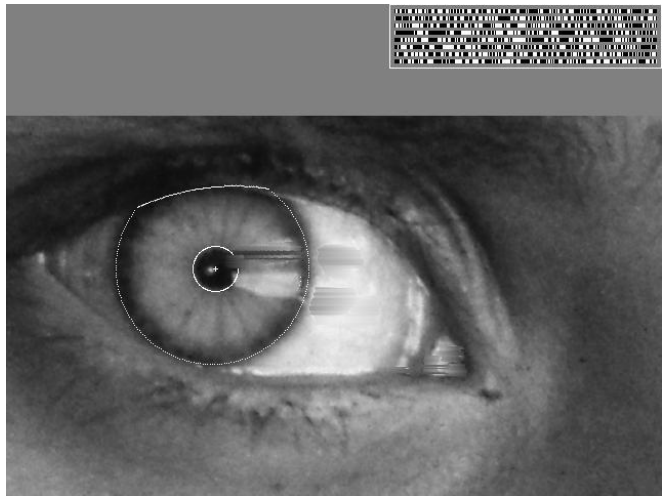


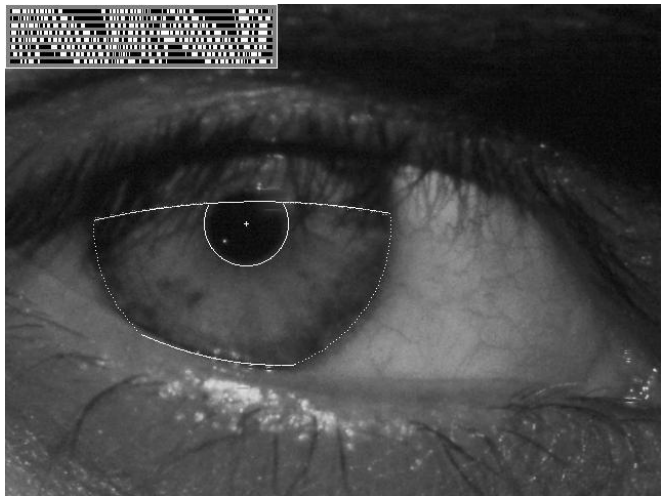
Who is she?

Source: S. Seitz

VISION-BASED BIOMETRICS

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story





Source: S. Seitz

LOGIN WITHOUT A PASSWORD…

Fingerprint scanners on many new laptops, other devices

Source: S. Seitz

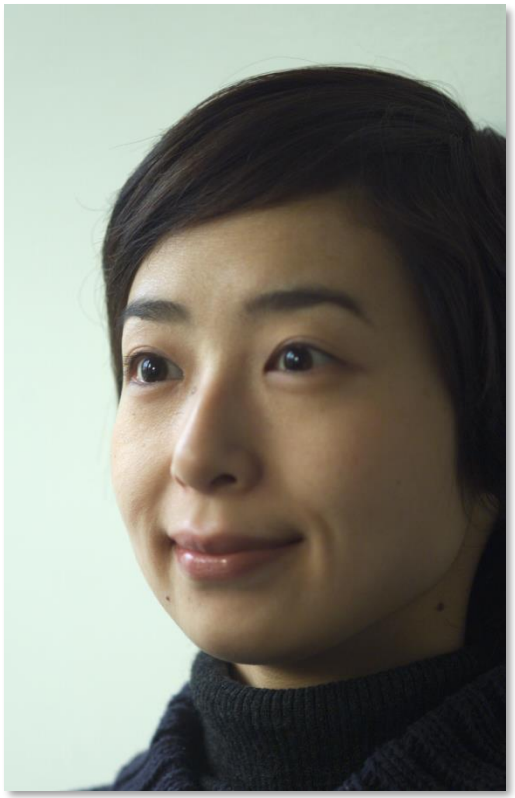
LOGIN WITHOUT A PASSWORD…

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

Source: S. Seitz

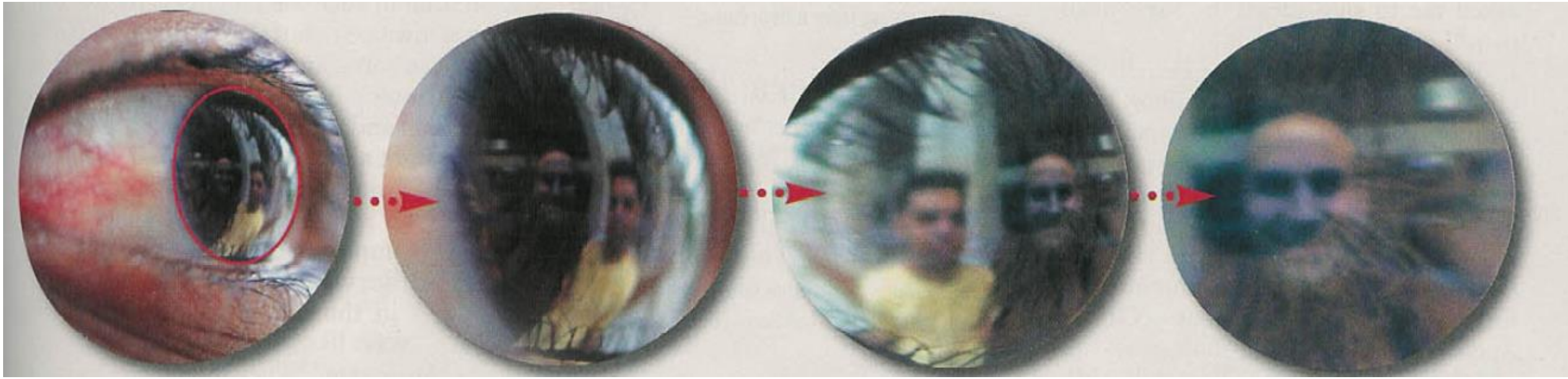
THE GOAL OF VISUAL RECOGNITION

- Forensics



Source: Nayar and Nishino, “Eyes for Relighting”

Source: Nayar and Nishino, “Eyes for Relighting”



Source: Nayar and Nishino, “Eyes for Relighting”

OPTICAL CHARACTER RECOGNITION (OCR)

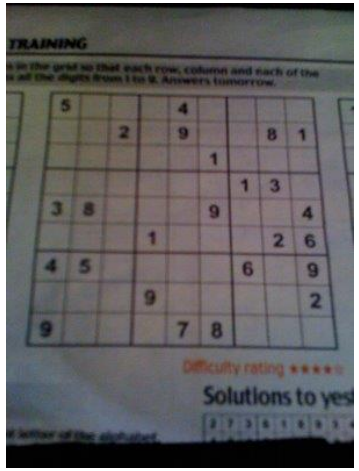
- If you have a scanner, it probably came with OCR software

Digit recognition, AT&T labs

License plate readers

http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Automatic check processing



Augmented Reality iPhone
Sudoku Grab
http://sudokugrab.blogspot.com/

Source: S. Seitz

OBJECT RECOGNITION (IN SUPERMARKETS)

SUPERMARKET CHECKOUT SCANNER USES OBJECT RECOGNITION INSTEAD OF BAR CODES

“This Toshiba scanner, demonstrated in Japan, knows what vegetables look like -- just hold up your daikon or mizuna to the camera at the cash register, and it tots up the item. No need for stickers on your food, no need to consult a human, no need to even know what kind of onions you're buying. This is the future.”

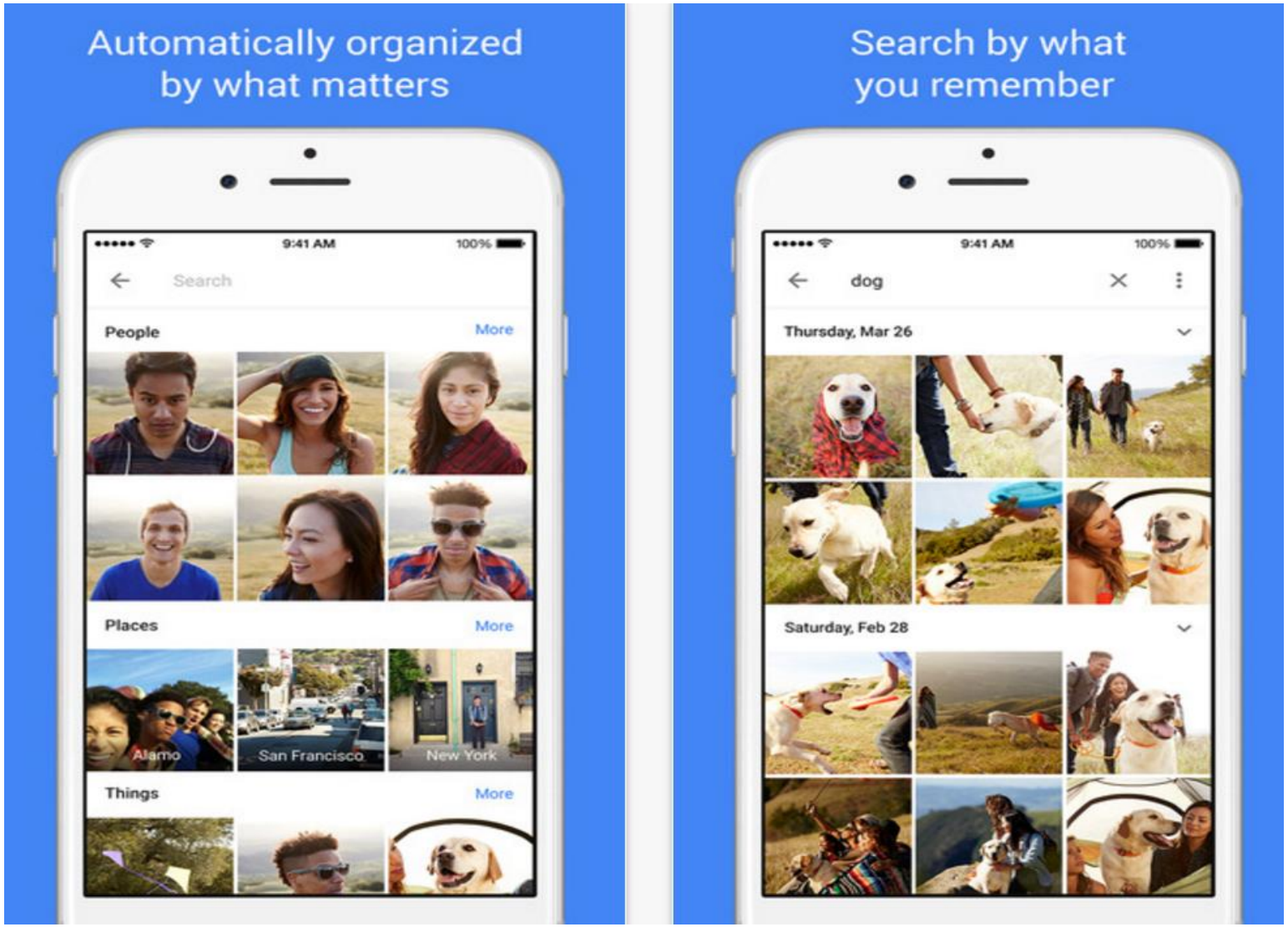
Source: S. Seitz

OBJECT RECOGNITION (IN MOBILE PHONES)

Source: S. Seitz

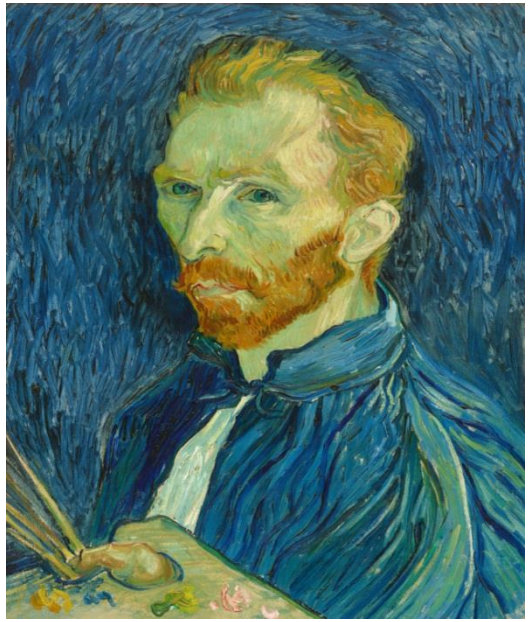
GOOGLE SEARCH BY IMAGE

VISUAL RECOGNITION FOR PHOTO ORGANIZATION



Google photo

STYLE TRANSFER



Source image (Style)

Target image (Content)

Output (deepart)

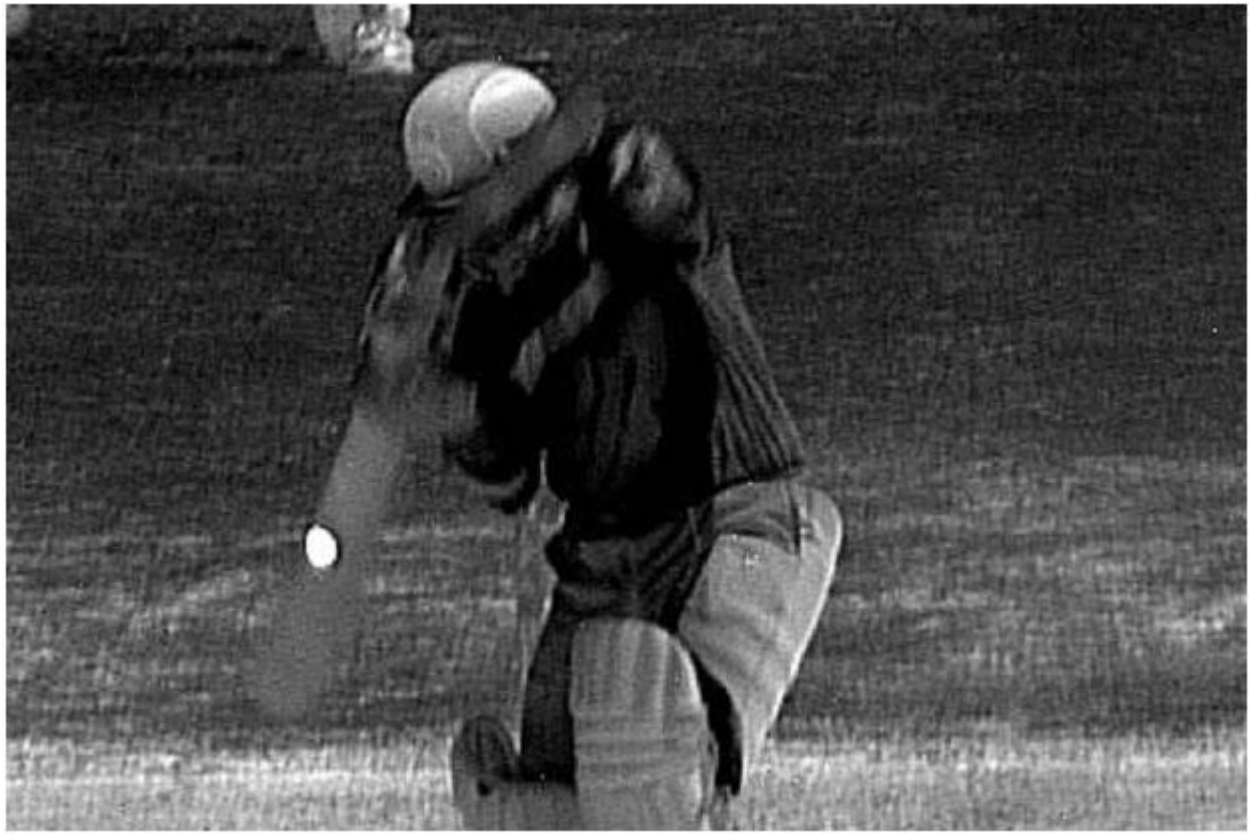
A Neural Algorithm of Artistic Style [Gatys et al. 2015]

SPECIAL EFFECTS: SHAPE CAPTURE

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

SPORTS



Infra Red Cricket

Technology In Cricket

Source: S. Seitz

SPORTS

Sportvision first down line

Nice explanation on www.howstuffworks.com

Source: S. Seitz

VISION-BASED INTERACTION (AND GAMES)

Assistive technologies

Nintendo Wii has camera-based finger tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

SMART CARS

- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models

Sources: A. Shashua, S. Seitz

GOOGLE CARS

Google in talks with Ford, Toyota and Volkswagen to realise driverless cars

http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871/

VISUAL RECOGNITION IN SPACE

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

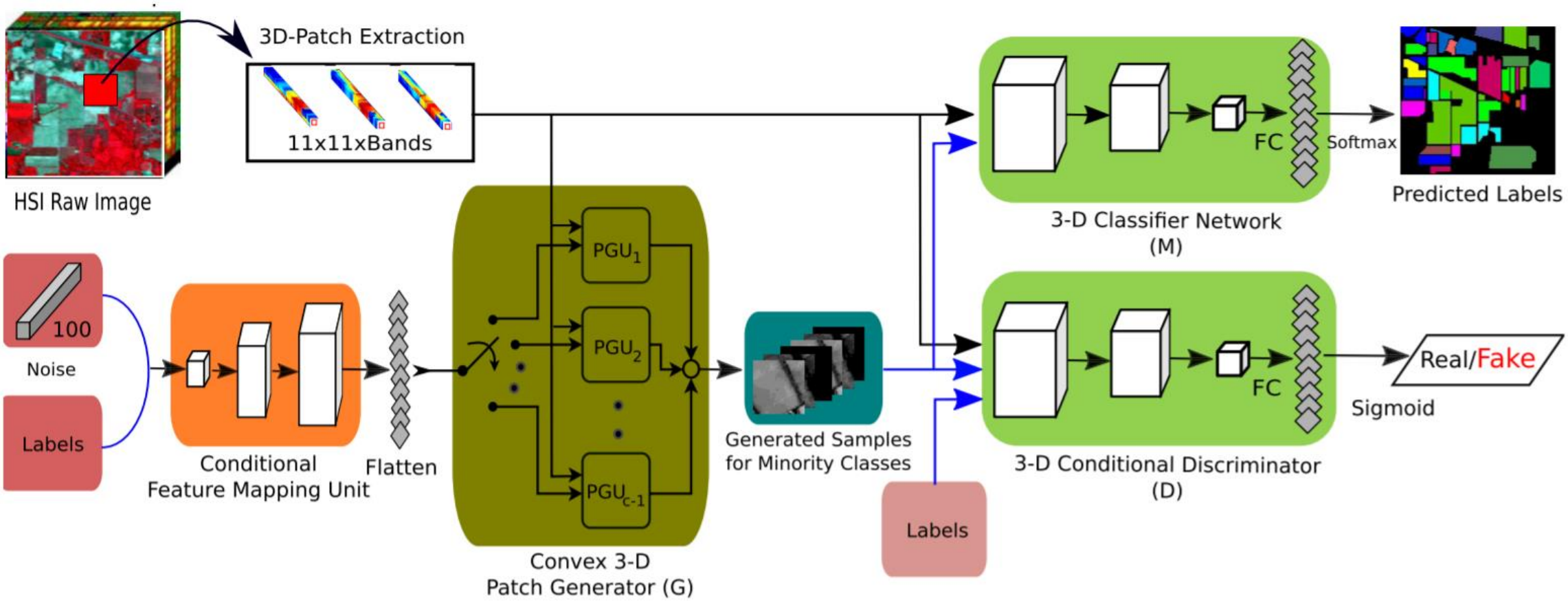
- Obstacle detection,
- Position tracking
- and many more ……

ROBOTICS

NASA's Mars Curiosity Rover (Mars Science Laboratory) Autonomous RC (Radio-controlled) Car

http://en.wikipedia.org/wiki/Spirit_rover http://www.cs.cornell.edu/~asaxena/rccar/

HYPERSPECTRAL IMAGE CLASSIFICATION



Source: HyperGAMO

EARTH VIEWERS (3D MODELING)

USING LAND COVER CLASSIFICATION

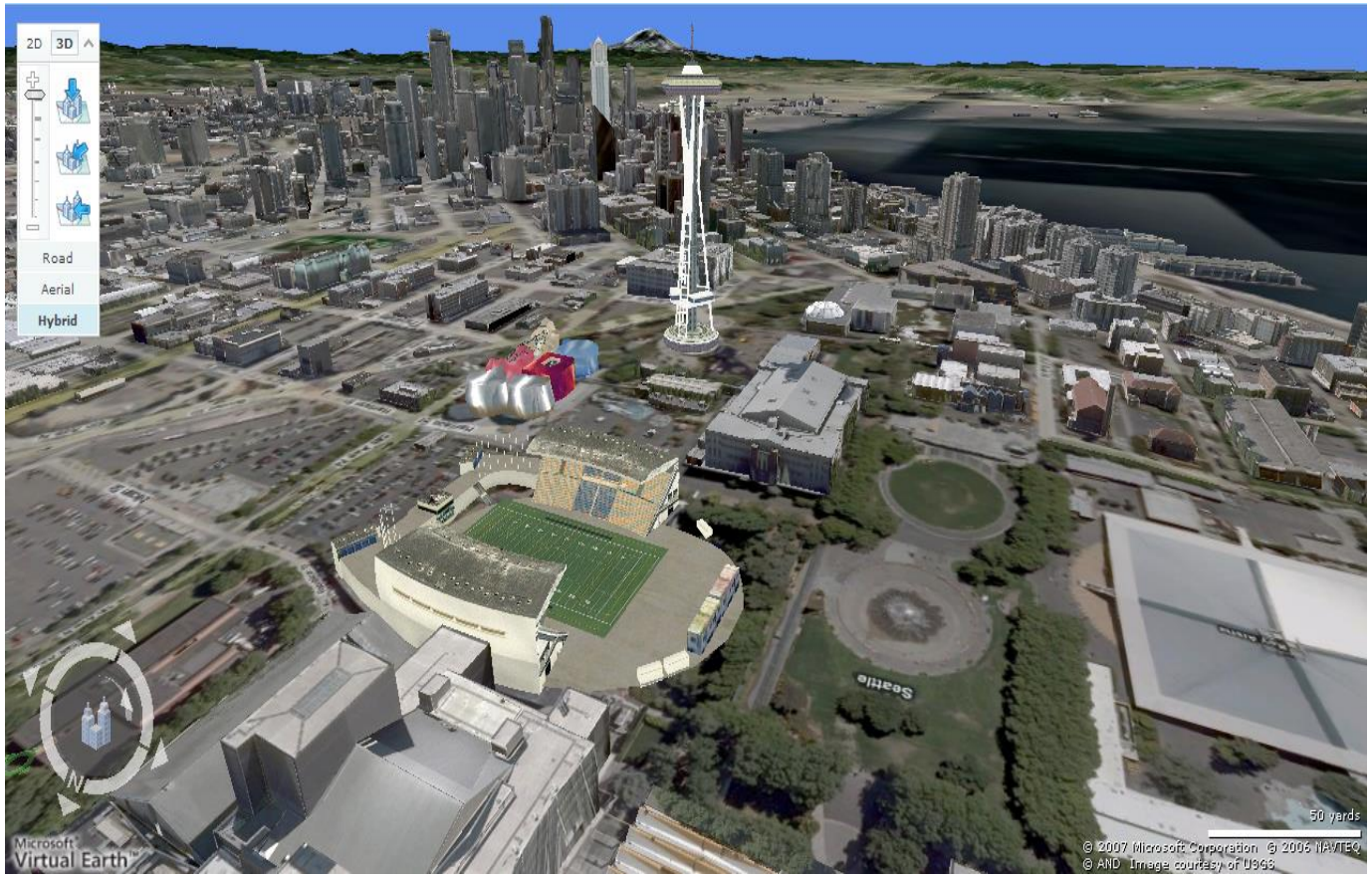
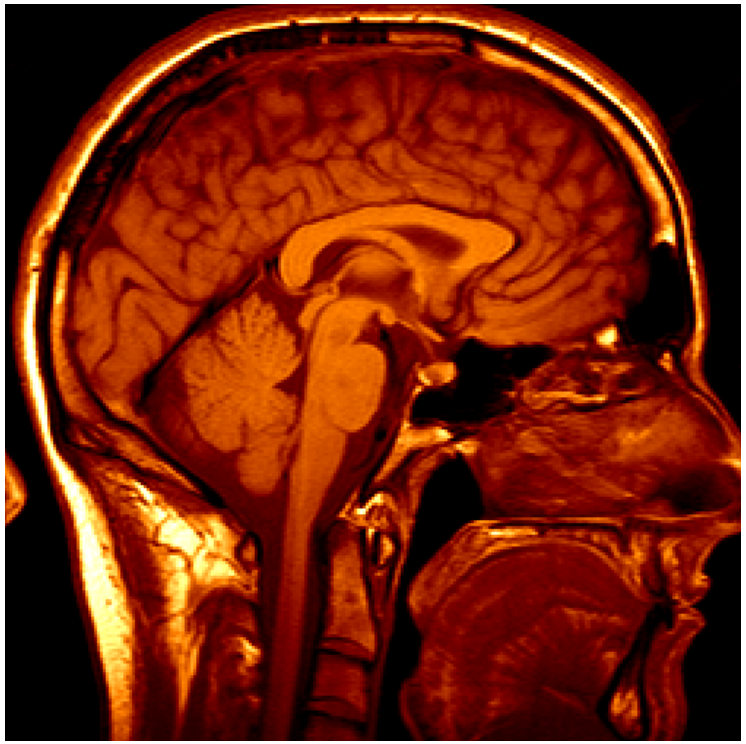


Image from Microsoft's
Virtual Earth
(see also: Google Earth)

Slide credit: Steve Seitz

MEDICAL IMAGING

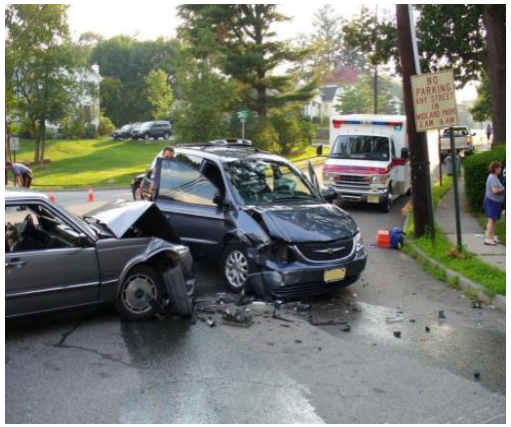


3D imaging
MRI, CT

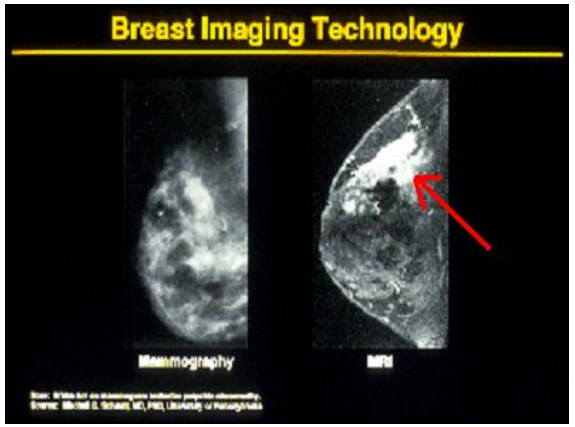
Image guided surgery
Grimson et al., MIT

Source: S. Seitz

THE GOAL OF VISUAL RECOGNITION

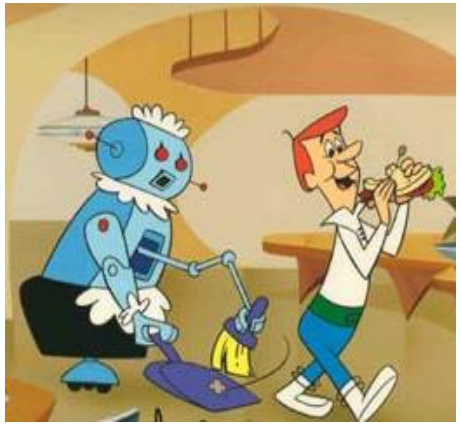


Safety



Health

Security

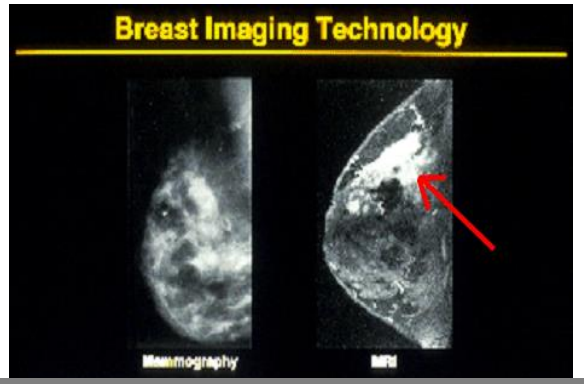


Comfort

Fun

Access

THE GOAL OF VISUAL RECOGNITION



Computer Vision
Technology
Can Better Our Lives

Comfort

Fun

Access

CURRENT STATE OF THE ART

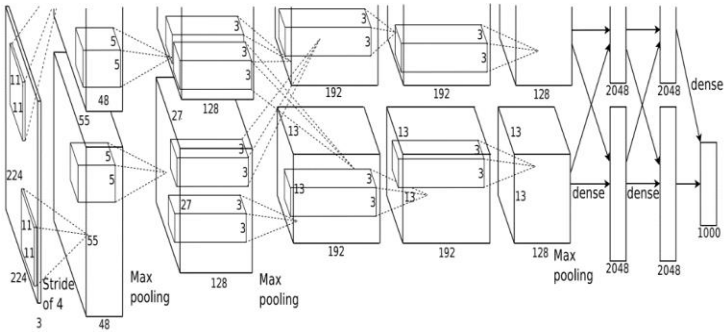
- Many of these are less than 10 years old
- Very active and exciting research area!
- To learn more about vision applications and companies
  - David Lowe maintains an excellent overview of vision companies
    - http://www.cs.ubc.ca/spider/lowe/vision.html

COURSE OVERVIEW

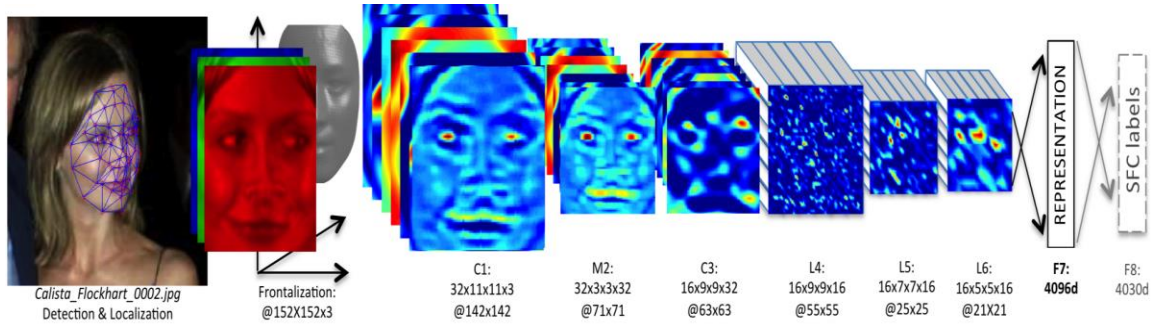
1. Feature Level based Visual Recognition
   - Feature Detection, Representation and Matching.
2. Neural Networks based Visual Recognition
   - Neural Networks, MLP, Training, Optimizers, Best Implementation Practices
3. Convolutional Neural Networks
   - Architectures for Visual Recognition: Convolution Neural Network: Convolution and pooling; AlexNet, VGG, GoogLeNet, ResNet, etc.
4. CNN for Visual Recognition Tasks
   - Object Detection, Semantic Segmentation
5. Generative Tasks
   - Generative Adversarial Network, Diffusion Models
6. Attention and Vision Transformers
   - Vision-Text Applications, Attention, Self-Attention, Transformer, ViT, etc.
7. Recent Trends
   - Explainability in Visual Recognition, Applications to Biometric Recognition, Medical Image Analysis, etc.

PROJECT IDEAS: PEDESTRIAN DETECTION

PROJECT IDEAS: CNN FOR DIFFERENT PROBLEMS

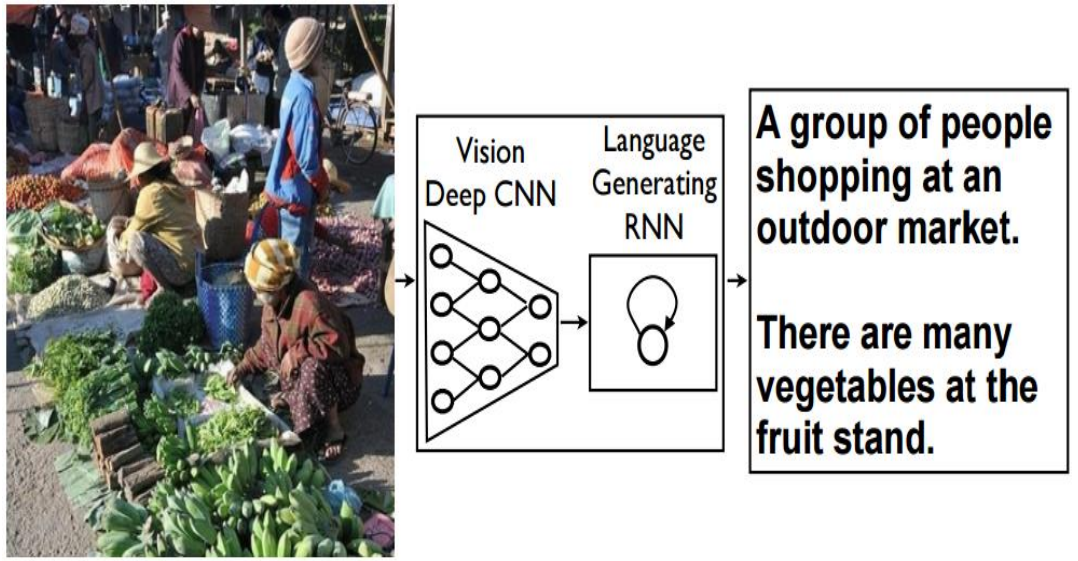


[AlexNet NIPS 2012]



[DeepFace CVPR 2014]

[DeepPose CVPR 2014]



[Show, Attend and Tell ICML 2015]

MORE PROJECT IDEAS

- Object Detection for Autonomous Vehicles
- Human Action Recognition
- Group Activity Recognition
- Identifying Car Model from Photographs
- Face Recognition under Challenging Scenario
- Person Tracking under Distributed Cameras
- Object Detection in Aerial Images
- Mask and Social Distance Recognition
- Micro-expression recognition
- Image Captioning
- Image to Image Translation
- Image Retrieval
- Multi-View Vision
- and many more ….

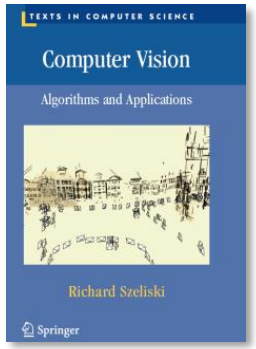
COURSE REQUIREMENTS

- Prerequisites—*these are essential*!
  - Data structures
  - A good working knowledge of python programming
  - Linear algebra
  - Vector calculus

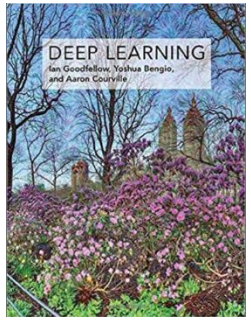
- Course does ***not*** assume prior imaging experience
  - visual recognition, computer vision, image processing, graphics, etc.

RESOURCES

- Textbook:



Rick Szeliski, *Computer Vision: Algorithms and Applications*
online at: http://szeliski.org/Book/



Ian Goodfellow and Yoshua Bengio and Aaron Courville,
Deep Learning, MIT Press, 2016
online at: http://www.deeplearningbook.org/

RECOMMENDED BOOKS (OPTIONAL)

- Simon Prince, Computer Vision: Models, Learning, and Interface, Cambridge University Press, 2012 [PDF]
- Richard Szeliski, Computer Vision: Algorithms and Applications, Springer, 2010, [PDF]
- Forsyth and Ponce, Computer Vision: A Modern Approach, Prentice Hall, 2002, [PDF]
- Mubarak Shah, Fundamentals of Computer Vision, 1997 [PDF]
- Palmer, Vision Science, MIT Press, 1999,
- Duda, Hart and Stork, Pattern Classification (2nd Edition), Wiley, 2000,
- Koller and Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009,
- Strang, Gilbert. Linear Algebra and Its Applications 2/e, Academic Press, 1980.

ONLINE MATERIALS

- CS231n: Convolutional Neural Networks for Visual Recognition, Stanford University (http://cs231n.stanford.edu/)
- Deep Learning Specialization by Prof. Andrew Ng, Coursera (https://www.coursera.org/specializations/deep-learning)
- Deep Learning for Computer Vision (NPTEL)

PROGRAMMING

- Python will be the preferred language.
  - Following book (Python programming samples for computer vision tasks) is freely available: Python for Computer Vision

ACADEMIC INTEGRITY

- Can discuss HW with peers, but don't copy and/or share code
- Carefully document any sources within HW hand-in
- Do not use code from Internet unless you have permission
  If you're not sure, ask
- Do not use your published work as your final project
  If you're not sure, ask

WHAT TO EXPECT FROM THIS COURSE

- Broad coverage
  - Feature extraction, representation and matching, recognition, learning, CNN, GAN, recent trends in visual recognition
  - Focus is on algorithms, rather than specific systems.
- Background to delve deeper into any visiual recognition - related topic
- Practical experience
- Lots of work, tough material, fast pace, but hopefully lots of learning too!

ACKNOWLEDGEMENT

Thanks to the following courses and corresponding researchers for making their teaching/research material online

- Convolutional Neural Networks for Visual Recognition, Stanford University
- Deep Learning, Stanford University
- Introduction to Deep Learning, University of Illinois at Urbana-Champaign
- Introduction to Deep Learning, Carnegie Mellon University
- Natural Language Processing with Deep Learning, Stanford University
- And Many More Publicly Available Resources ……

Questions?