

Indian Institute of Information Technology, Allahabad



Visual Recognition

By

Dr. Shiv Ram Dubey

Computer Vision and Biometrics Lab

Department of Information Technology

Indian Institute of Information Technology, Allahabad

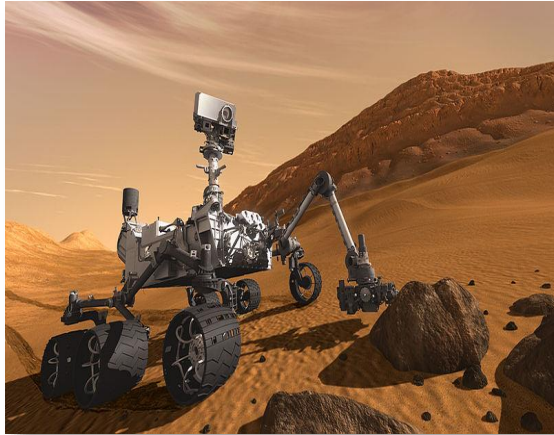


DISCLAIMER

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?

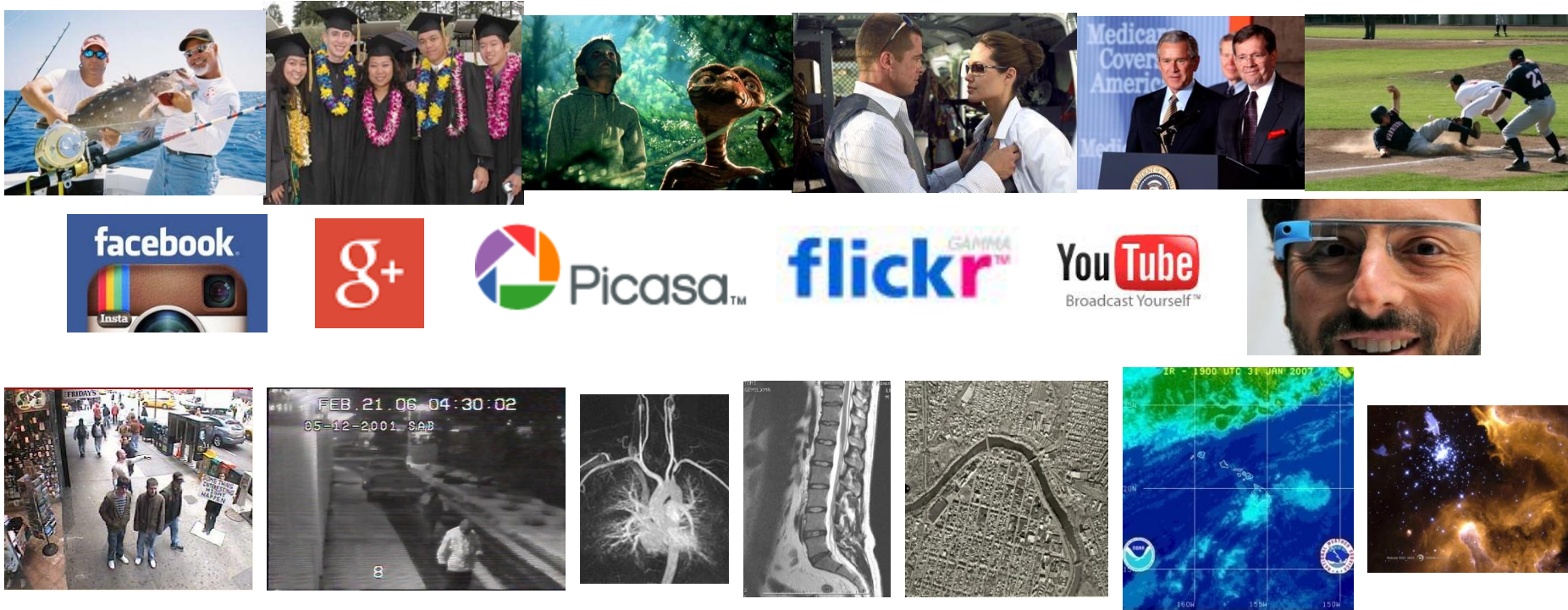


- Yes and no (mainly no)
 - computers can be better at “easy” things
 - humans are much better at “hard” things
- But huge progress has been made
 - Especially in the last 15 years
 - What is considered “hard” keeps changing



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

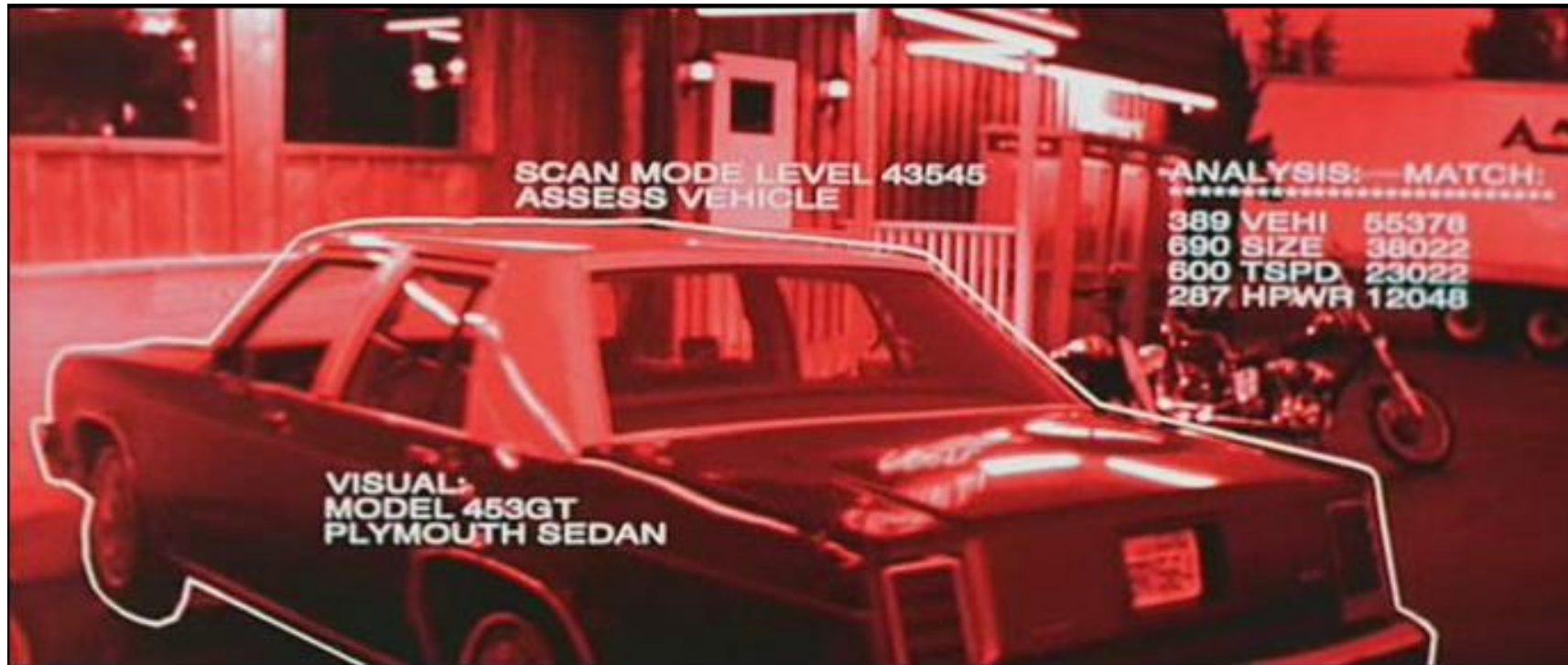


THE GOAL OF VISUAL RECOGNITION



THE GOAL OF VISUAL RECOGNITION

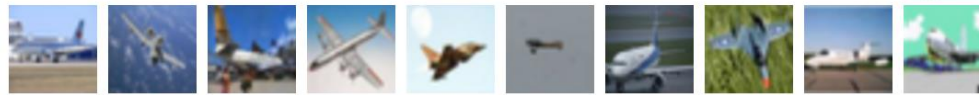
- Recognizing objects and people



THE GOAL OF VISUAL RECOGNITION

- Classification of visual images/videos

airplane



automobile



bird



cat



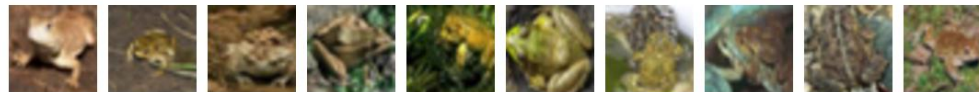
deer



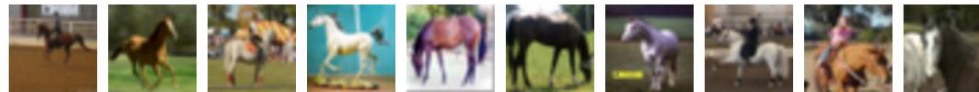
dog



frog



horse



ship



truck



Images From CIFAR dataset







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, ...



FACE RECOGNITION

Tag Your Friends

This will quickly label your photos and notify the friends you tag. [Learn more](#)



Who is this?



Who is this?



Who is this?



Who is this?
Facebook



Who is this?



Who is this?

Facebook face auto-tagging

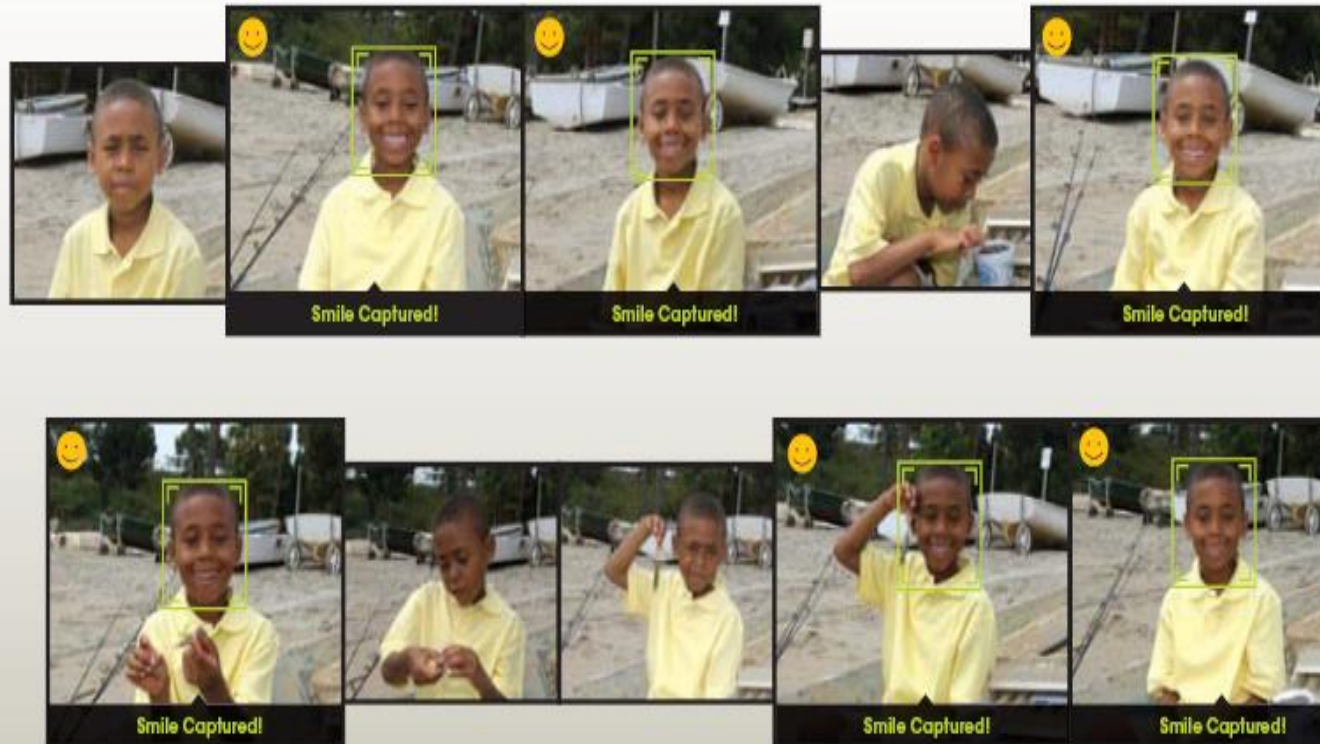
Google photo



SMILE RECOGNITION

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.

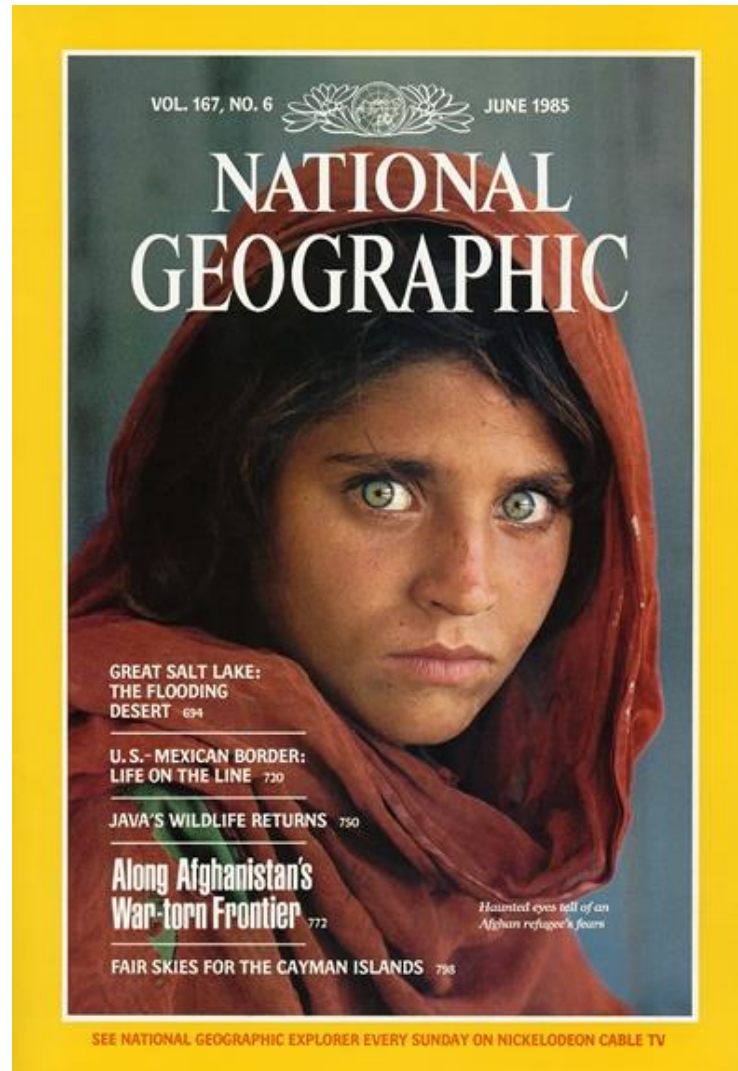


Sony Cyber-shot® T70
Digital Still Camera

Slide credit: Steve Seitz



BIOMETRIC RECOGNITION



Who is she?

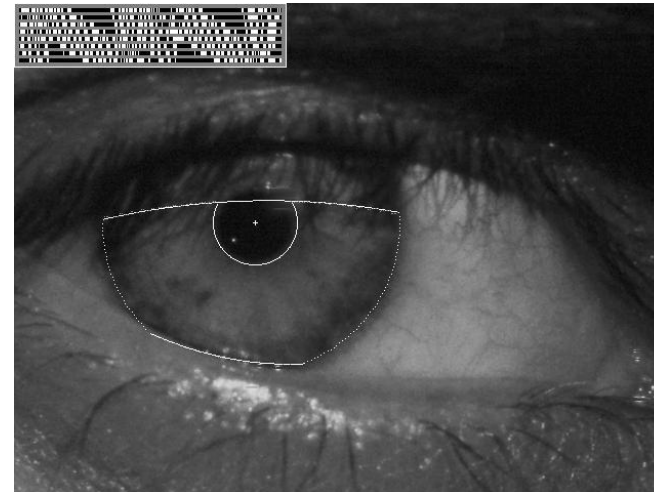
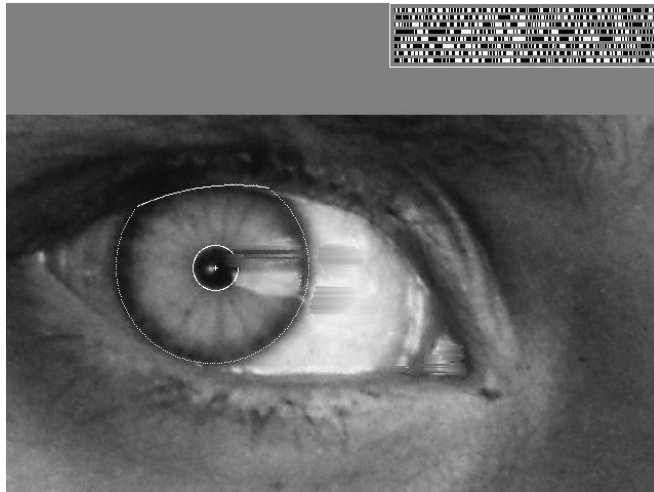
Source: S. Seitz



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



LOGIN WITHOUT A PASSWORD...



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



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

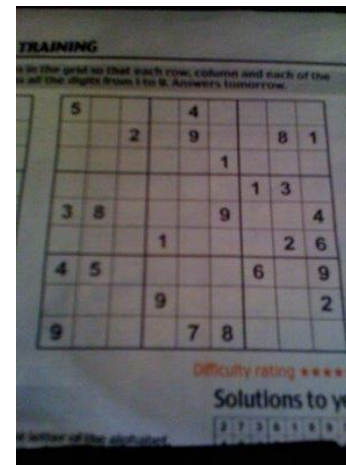


License plate readers

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)



GOOGLE SEARCH BY IMAGE



[Home](#)[Species](#)[Collectors](#)[About](#)

Leaf of the Bottlebrush Buckeye

Leafsnap: An Electronic Field Guide

Leafsnap is the first in a series of electronic field guides being developed by researchers from [Columbia University](#), the [University of Maryland](#), and the [Smithsonian Institution](#). This free mobile app uses visual recognition software to help identify tree species from photographs of their leaves.

Leafsnap contains beautiful high-resolution images of leaves, flowers, fruit, petiole, seeds, and bark. Leafsnap currently includes the trees of the Northeast and will soon grow to include the trees of the entire continental United States.

This website shows the tree species included in Leafsnap, the collections of its users, and the team of research volunteers working to produce it.

Free for iPhone:



and iPad:

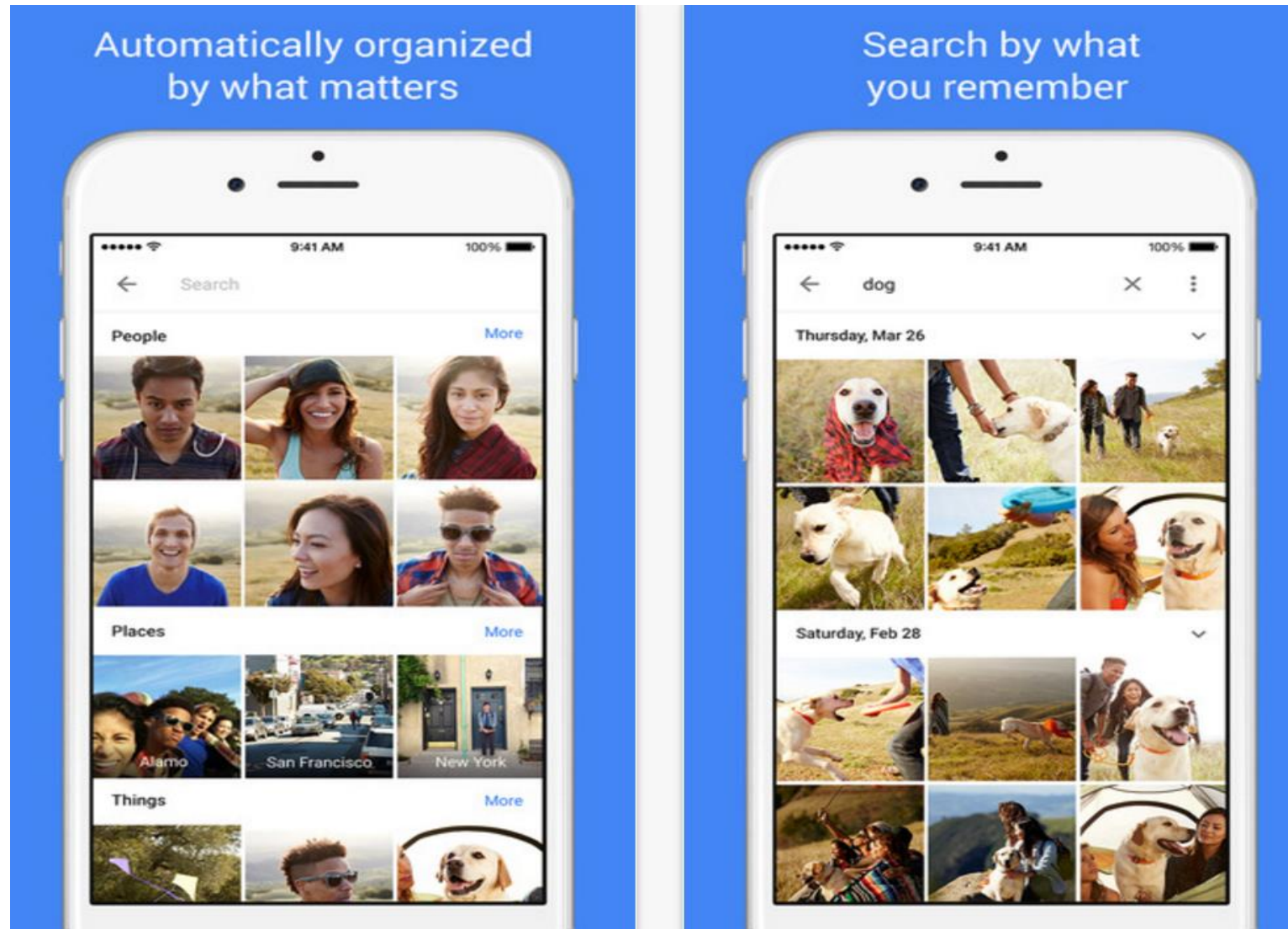








VISUAL RECOGNITION FOR PHOTO ORGANIZATION



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

[▶▶ manufacturer products](#) [consumer products ◀◀](#)

Our Vision. Your Safety.



rear looking camera

forward looking camera

side looking camera

EyeQ Vision on a Chip

[> read more](#)

Vision Applications



Road, Vehicle, Pedestrian Protection and more

[> read more](#)

AWS Advance Warning System

[> read more](#)

News

- > [Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System](#)
- > [Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end](#)

[> all news](#)

Events

- > [Mobileye at Equip Auto, Paris, France](#)
- > [Mobileye at SEMA, Las Vegas, NV](#)

[> read more](#)

- Mobileye

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



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



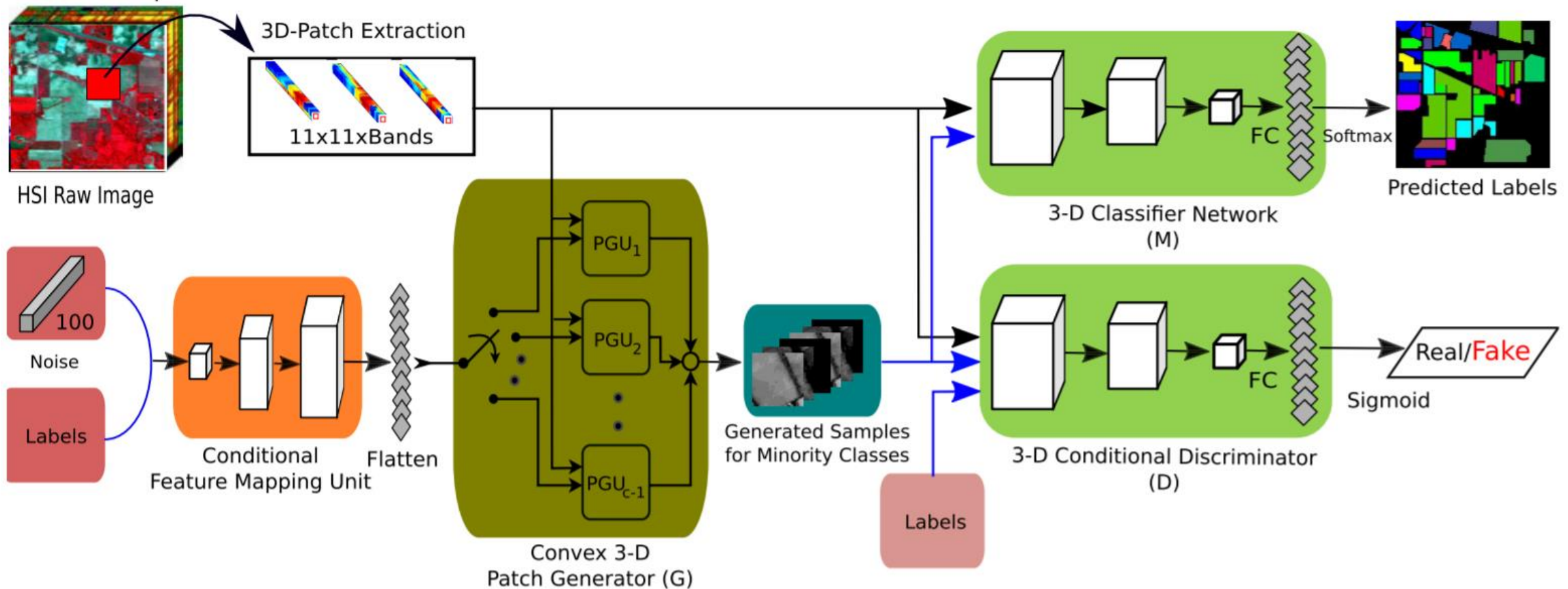
NASA's Mars Curiosity Rover (Mars Science Laboratory) Autonomous RC (Radio-controlled) Car

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

<http://www.cs.cornell.edu/~asaxena/rccar/>



HYPERSPPECTRAL IMAGE CLASSIFICATION



Source: [HyperGAMO](#)



EARTH VIEWERS (3D MODELING)

USING LAND COVER CLASSIFICATION

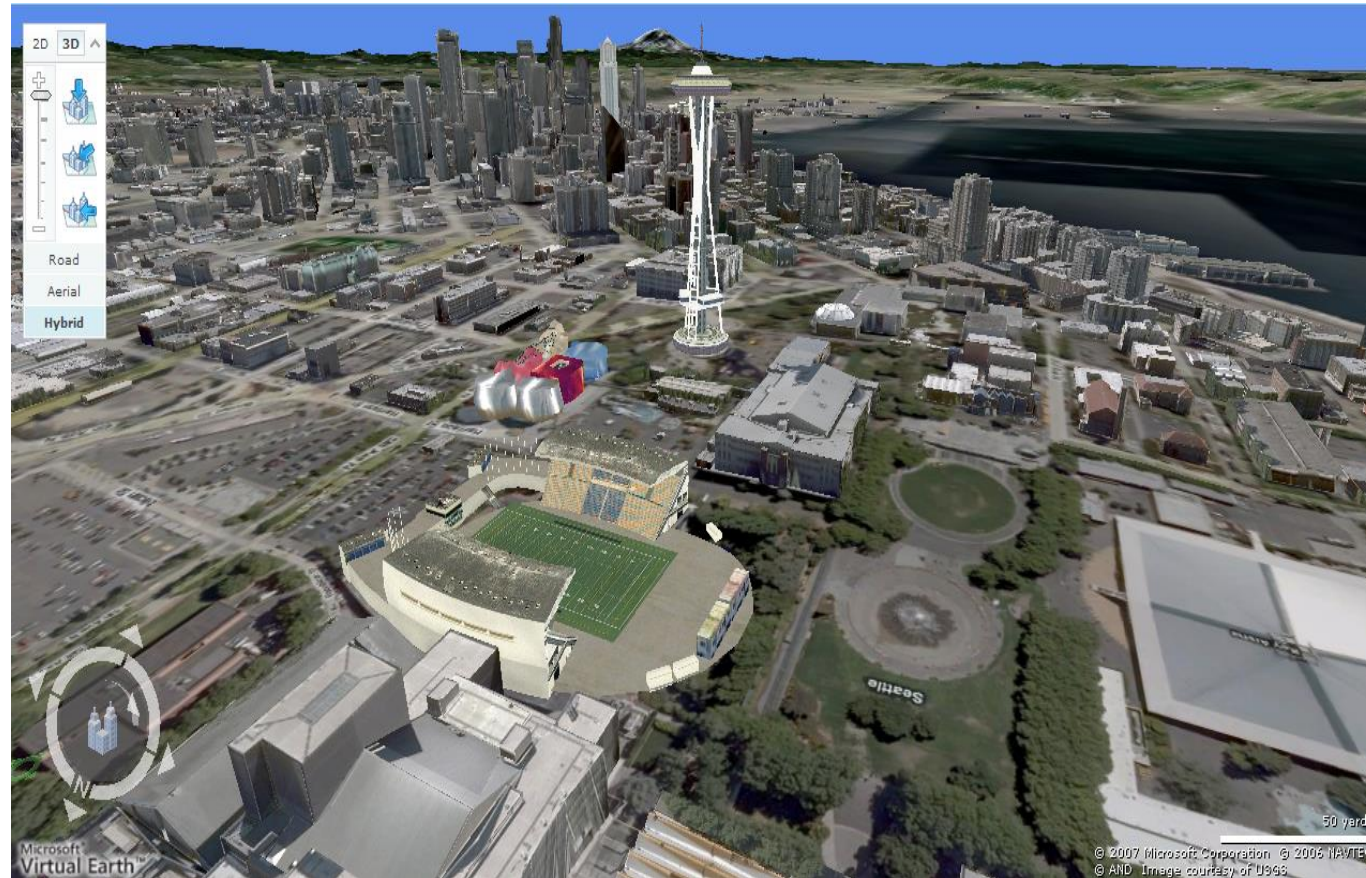
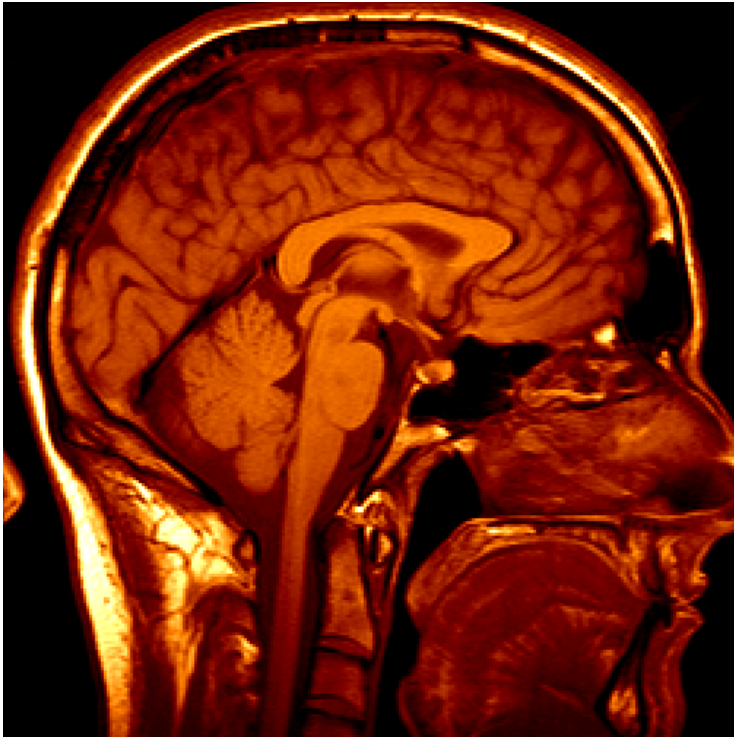


Image from Microsoft's
[Virtual Earth](#)
(see also: [Google Earth](#))

Slide credit: Steve Seitz



MEDICAL IMAGING



3D imaging
MRI, CT



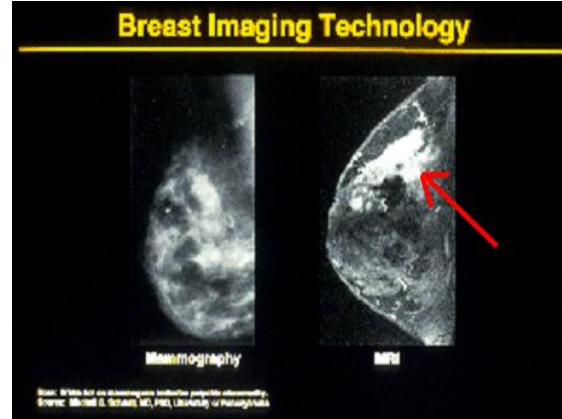
Image guided surgery
[Grimson et al., MIT](#)



THE GOAL OF VISUAL RECOGNITION



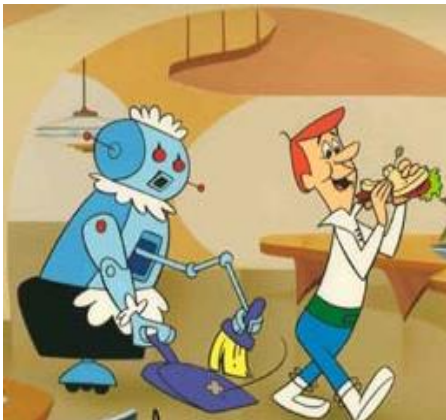
Safety



Health



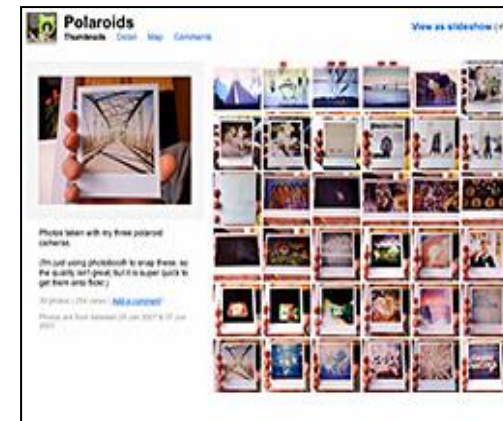
Security



Comfort



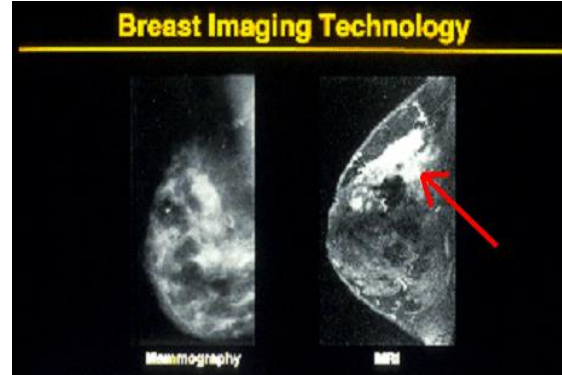
Fun



Access



THE GOAL OF VISUAL RECOGNITION



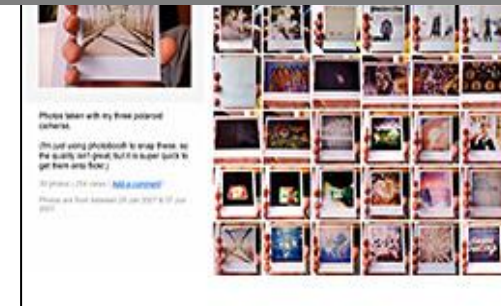
Computer Vision
Technology
Can Better Our Lives



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

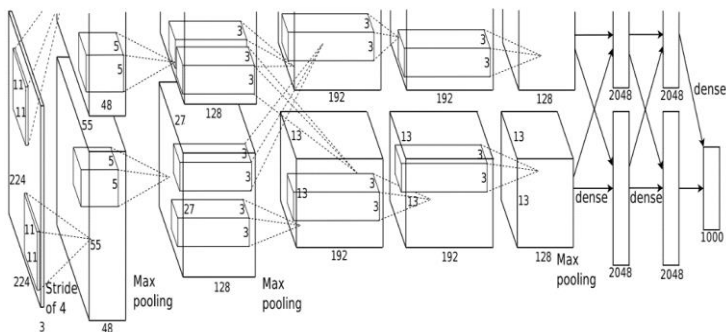
- **Foundations of Visual Recognition:** Introduction, Fundamentals of Machine Learning and Deep Learning for Vision. Convolutional Neural Networks (CNNs), Training Aspects, CNN Architectures for Image Classification, Object Detection, and Dense Prediction.
- **Generative Models for Visual Recognition:** Adversarial Attack and Defense, Image Generation, Generative Adversarial Networks, DCGAN, Pix2Pix, CycleGAN, Diffusion Models, Denoising Diffusion Models, Latent Diffusion Models, Stable Diffusion.
- **Vision Foundation Models:** Attention Mechanism, Vision Transformers (ViT), Hierarchical Transformers, Foundation Models for Vision. Vision Language Models - Image Captioning, Visual Question Answering (VQA).
- **Emerging Trends and Applications:** Explainability in Models for Visual Recognition, Video Recognition, Applications of Visual Recognition in Biomedical Health Informatics, Biometrics, Underwater Scenarios, Surveillance, etc.



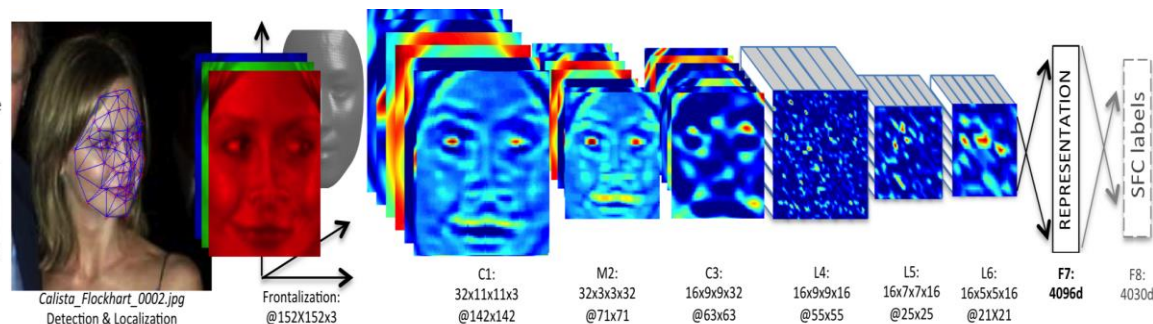
PROJECT IDEAS: PEDESTRIAN DETECTION



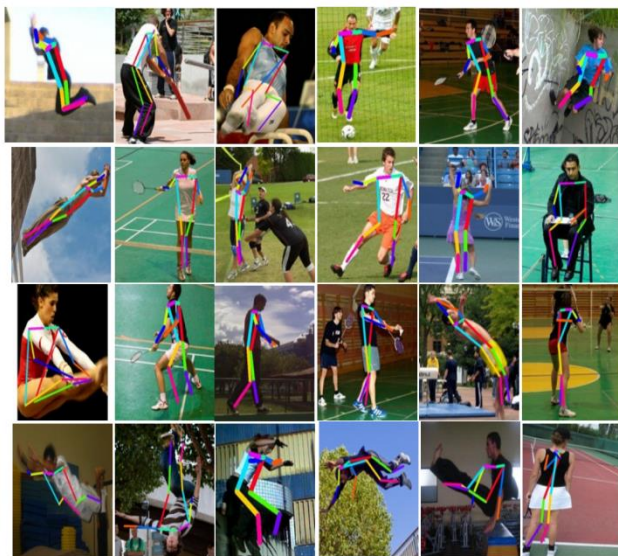
PROJECT IDEAS: CNN FOR DIFFERENT PROBLEMS



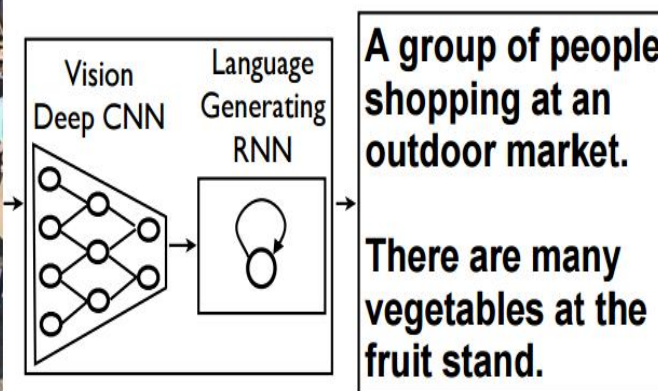
[[AlexNet NIPS 2012](#)]



[[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



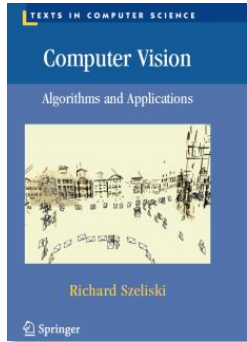
COURSE REQUIREMENTS

- 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.

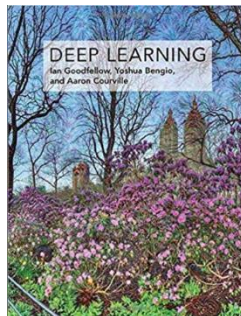


RESOURCES

- Textbook:



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



Ian Goodfellow and Yoshua Bengio and Aaron Courville,
[Deep Learning](http://www.deeplearningbook.org/), 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 visual 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?

