

# Advanced Image Processing

## An Introduction

Have you ever used computer vision?  
How? Where?

**Laptop:** Biometrics auto-login (face recognition, 3D), OCR

**Smartphones:** QR codes, computational photography (Android Lens Blur, iPhone Portrait Mode), panorama construction (Google Photo Spheres), face detection, expression detection (smile), Snapchat filters (face tracking), Google Tango (3D reconstruction), Night Sight (Pixel)

**Web:** Image search, Google photos (face recognition, object recognition, scene recognition, geolocalization from vision), Facebook (image captioning), Google maps aerial imaging (image stitching), YouTube (content categorization)

**VR/AR:** Outside-in tracking (HTC VIVE), inside out tracking (simultaneous localization and mapping, HoloLens), object occlusion (dense depth estimation)

**Motion:** Kinect, full body tracking of skeleton, gesture recognition, virtual try-on

**Medical imaging:** CAT / MRI reconstruction, assisted diagnosis, automatic pathology, connectomics, endoscopic surgery

**Industry:** Vision-based robotics (marker-based), machine-assisted router (jig), automated post, ANPR (number plates), surveillance, drones, shopping

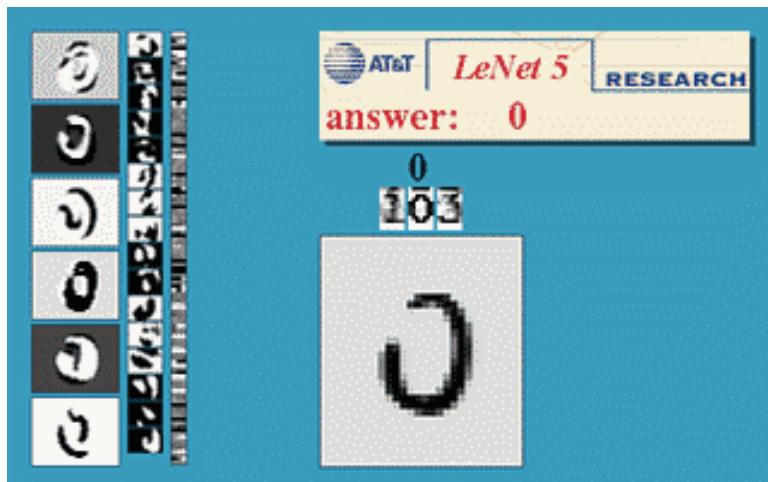
**Transportation:** Assisted driving (everything), face tracking/iris dilation for drunkenness, drowsiness, automated distribution (all modes)

**Media:** Visual effects for film, TV (reconstruction), virtual sports replay (reconstruction), semantics-based auto edits (reconstruction, recognition)

# Optical character recognition (OCR)

Technology to convert images of text into text

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



Mail digit recognition, AT&T labs  
<http://www.research.att.com/~yann/>



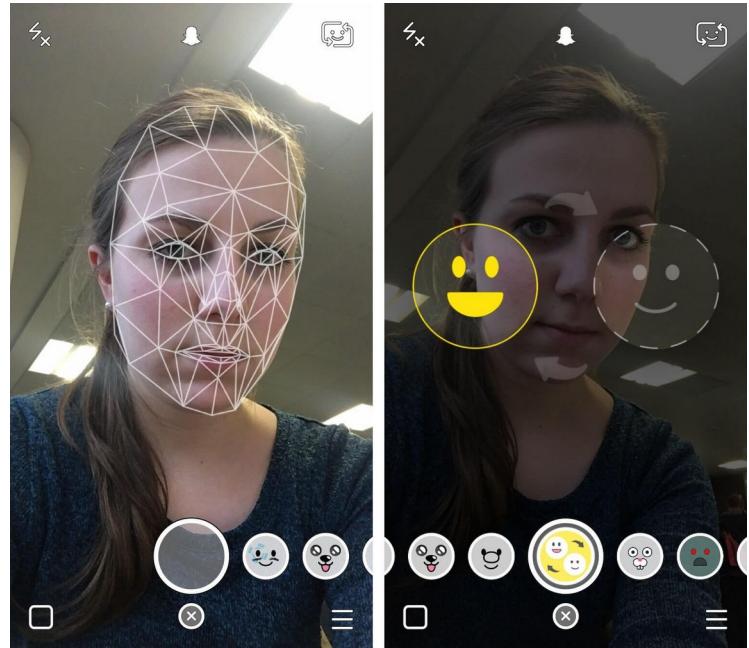
License plate readers  
[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)



Live  
Camera  
Translatio  
n



# Face detection



- Almost all digital cameras detect faces
- Snapchat face filters

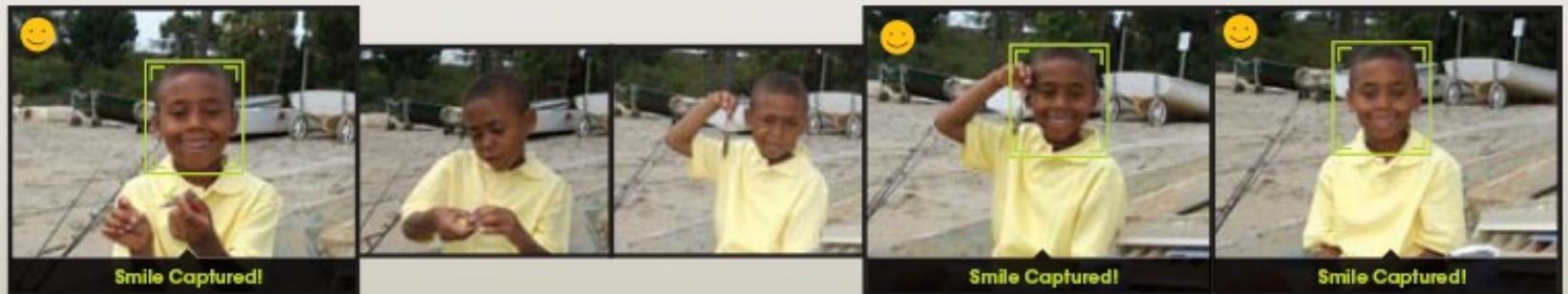




# Smile detection

## 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.

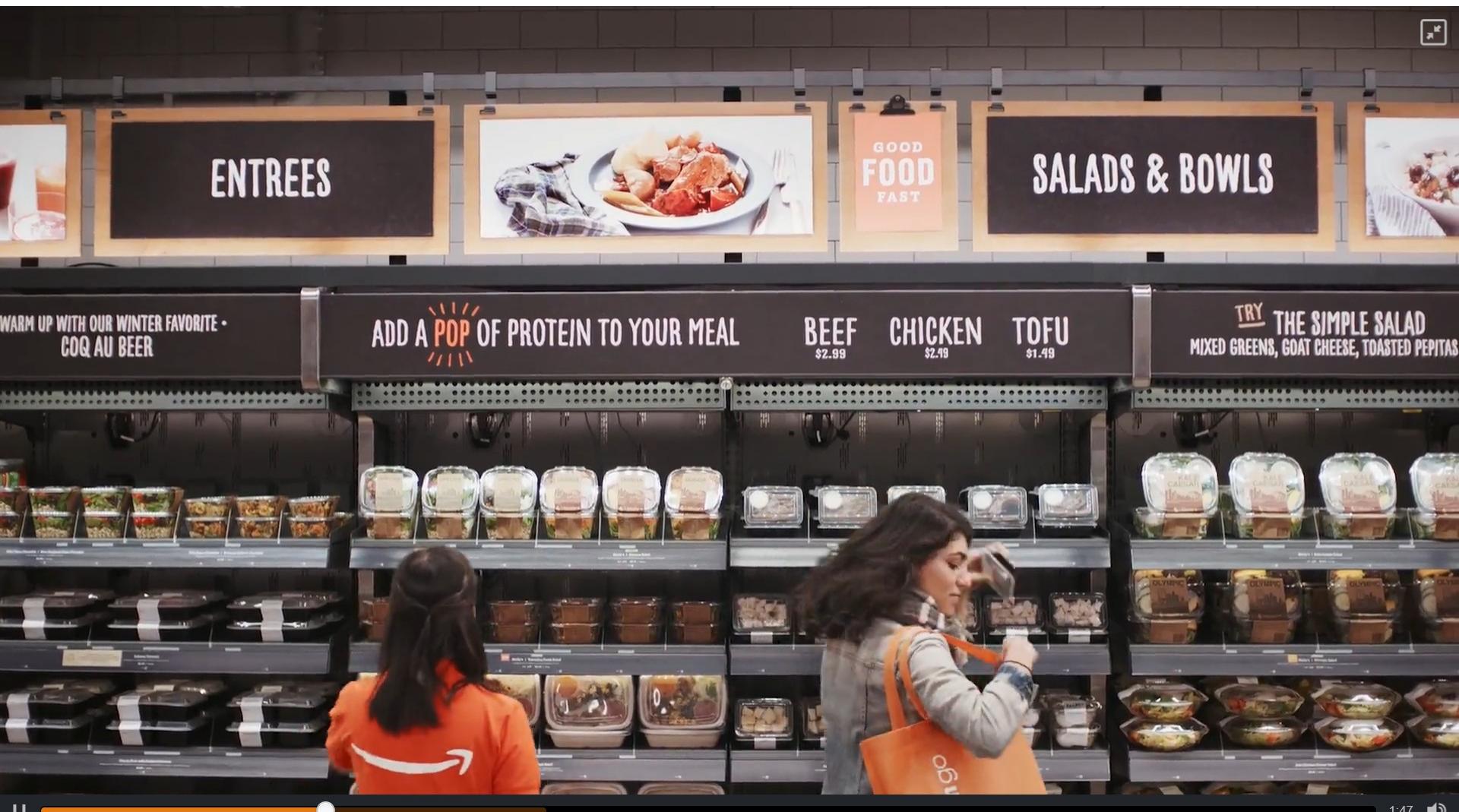


[Sony Cyber-shot® T70 Digital Still Camera](#)

# Object recognition (in supermarkets)

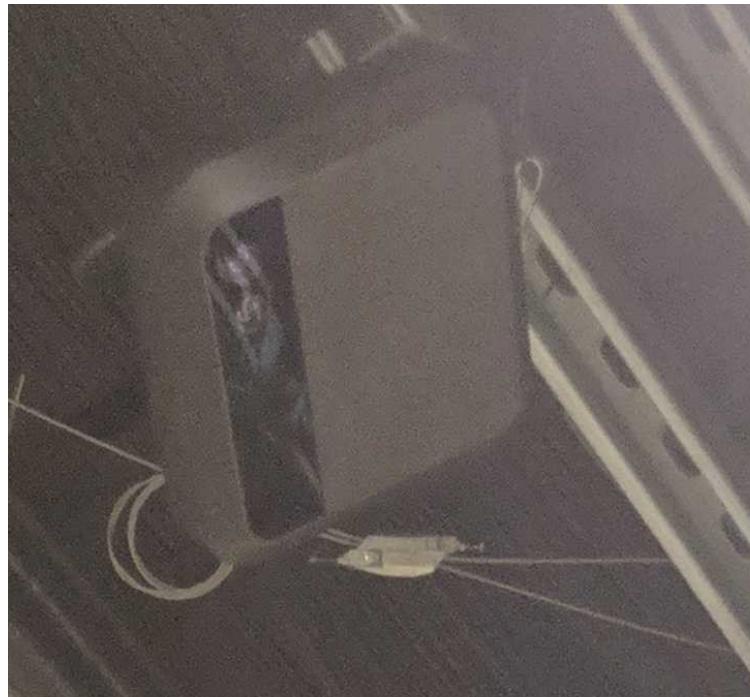
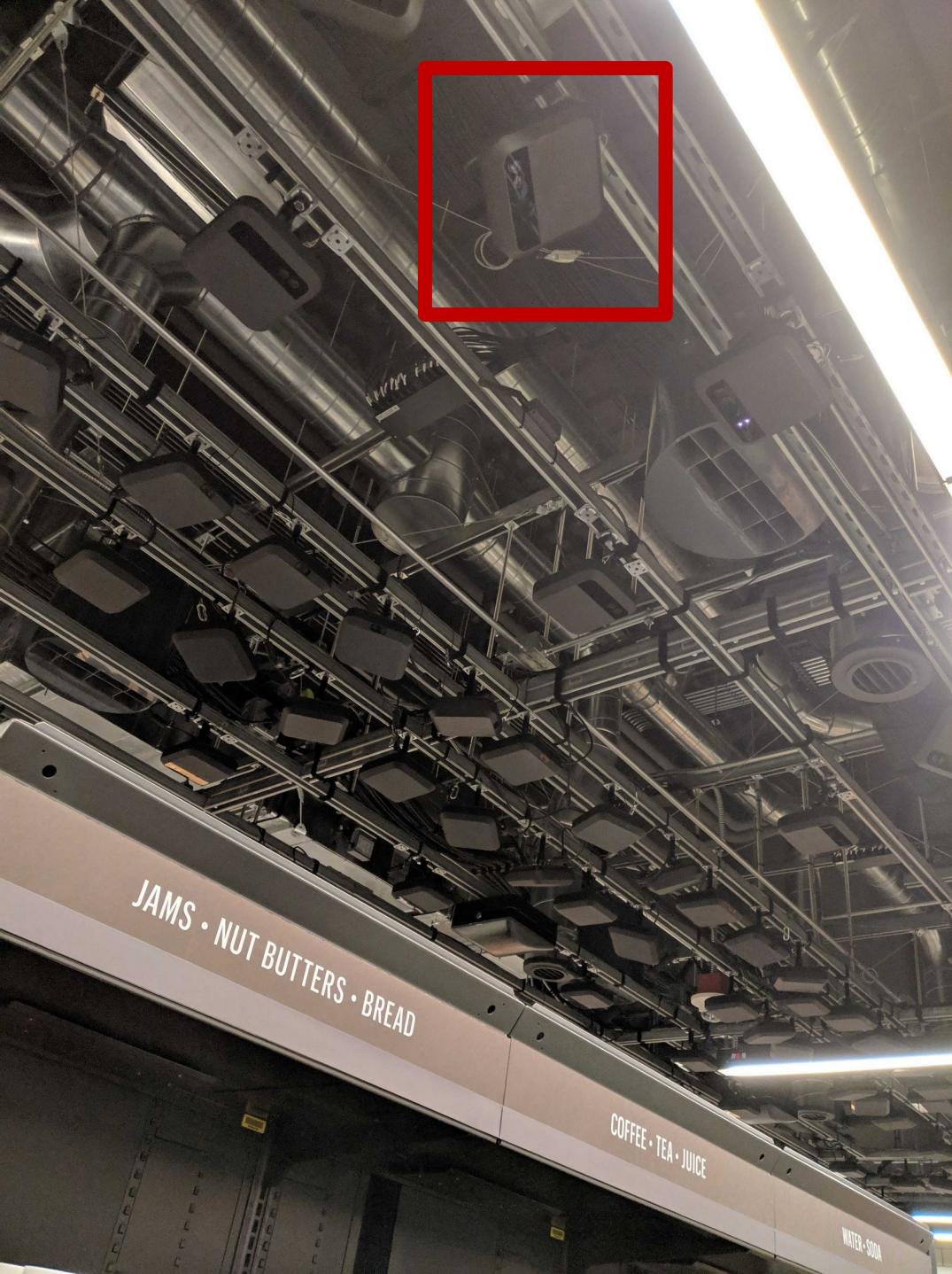


# How does it work?



# How does it work?





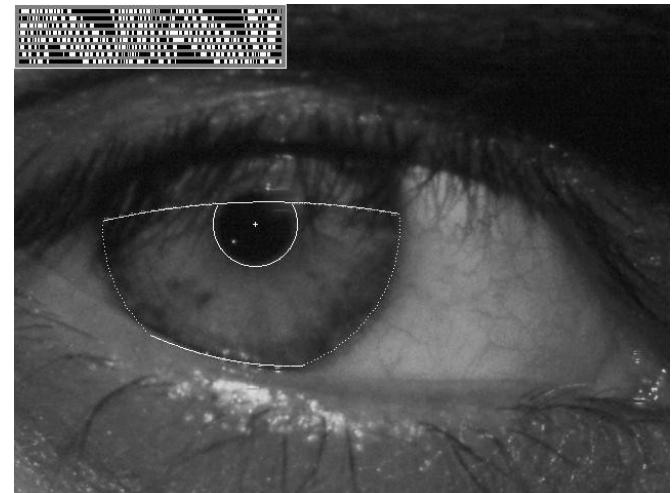
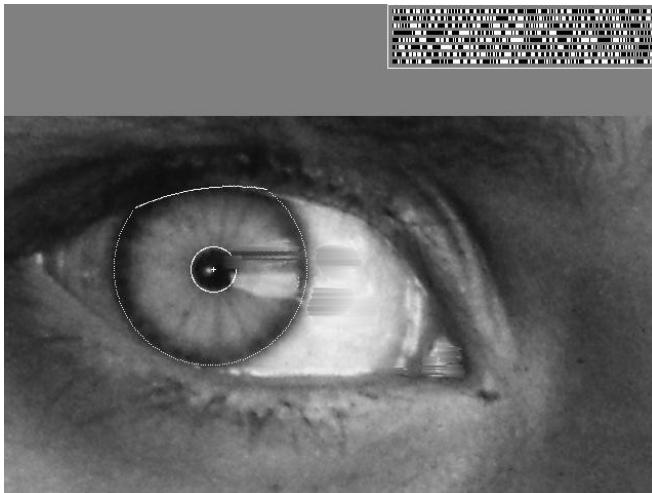
Thanks to Vivek  
Ramanujan

# Vision-based biometrics



*“How the Afghan Girl was Identified by Her Iris Patterns”*

Read the [story \(Wikipedia\)](#)



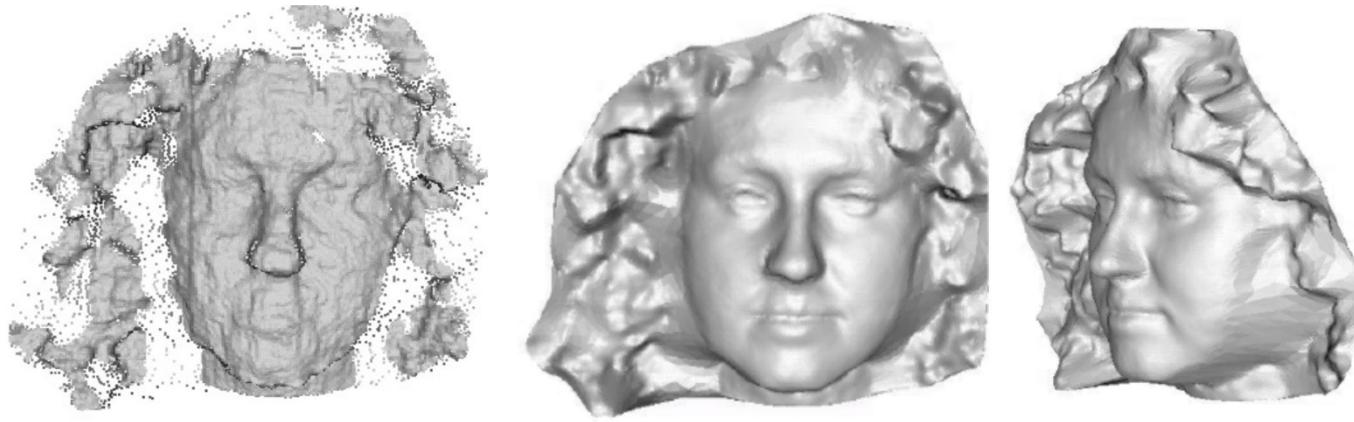
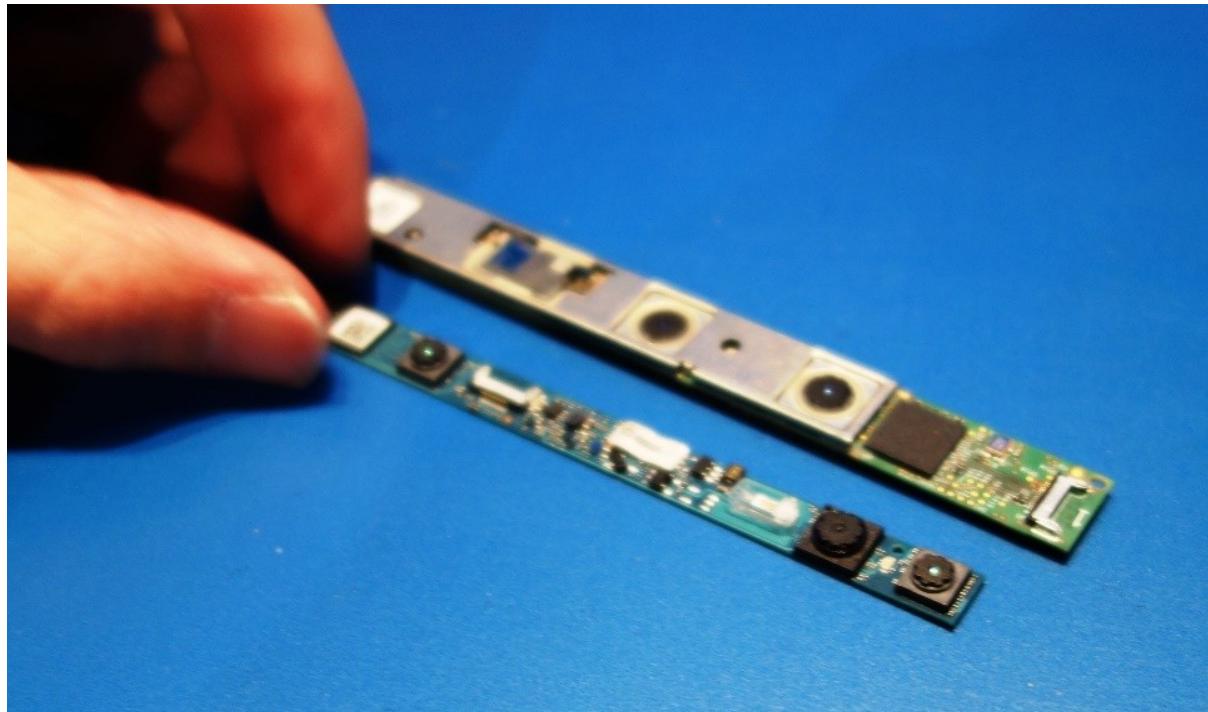
# Facial login without a password...



# Facial login without a password...



# Facial login without a password...



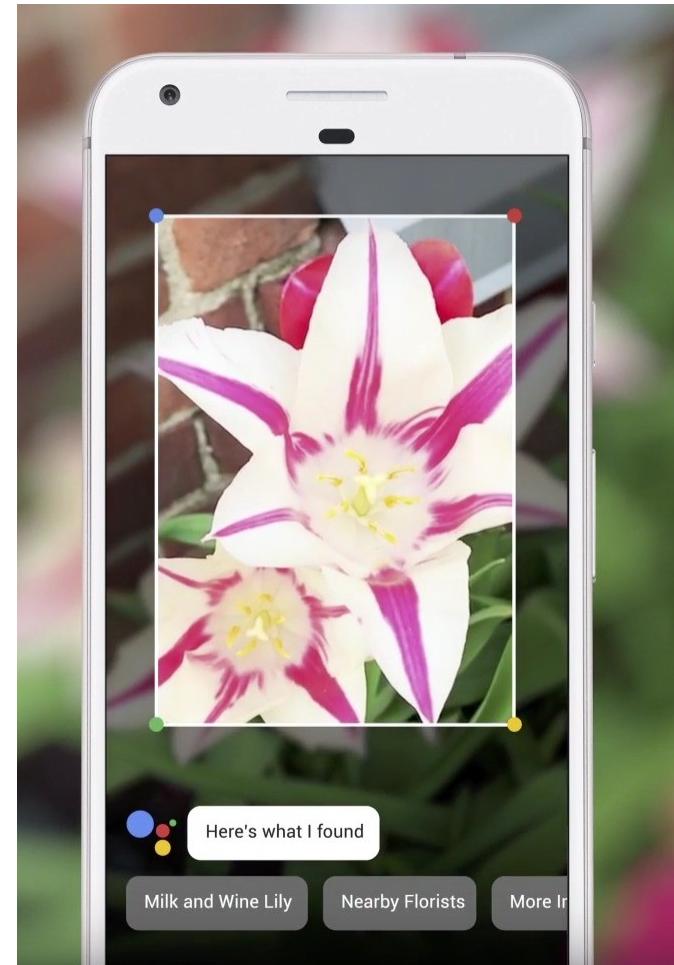
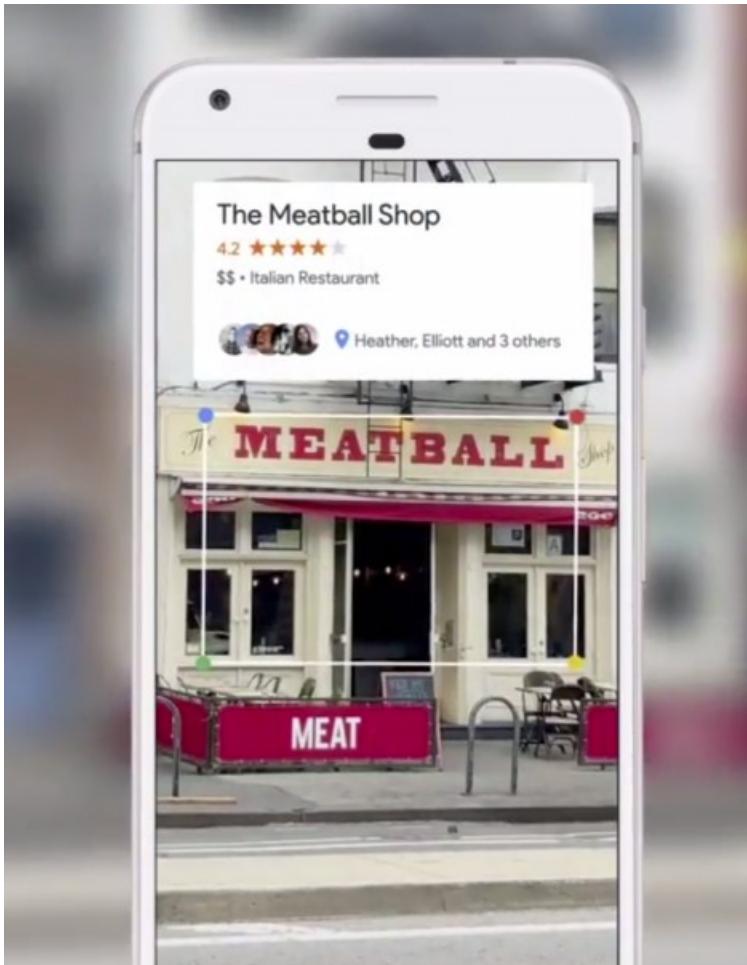
Single depth frame

Reconstructed 3D mesh

Liang et  
al. 2014

# Object recognition (in mobile phones)

e.g., Google Lens



# 3D from images

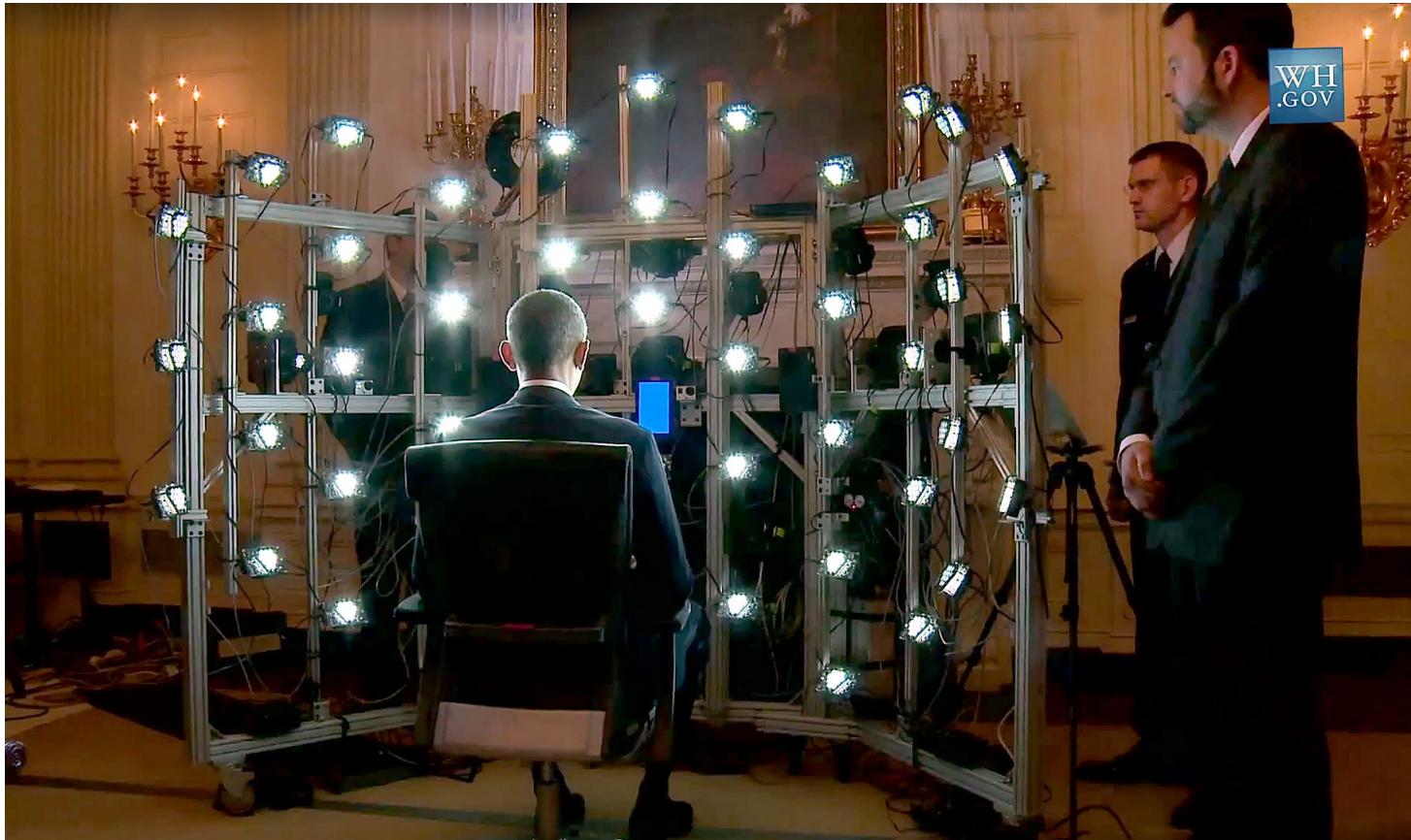


Building Rome in a Day: Agarwal et al. 2009

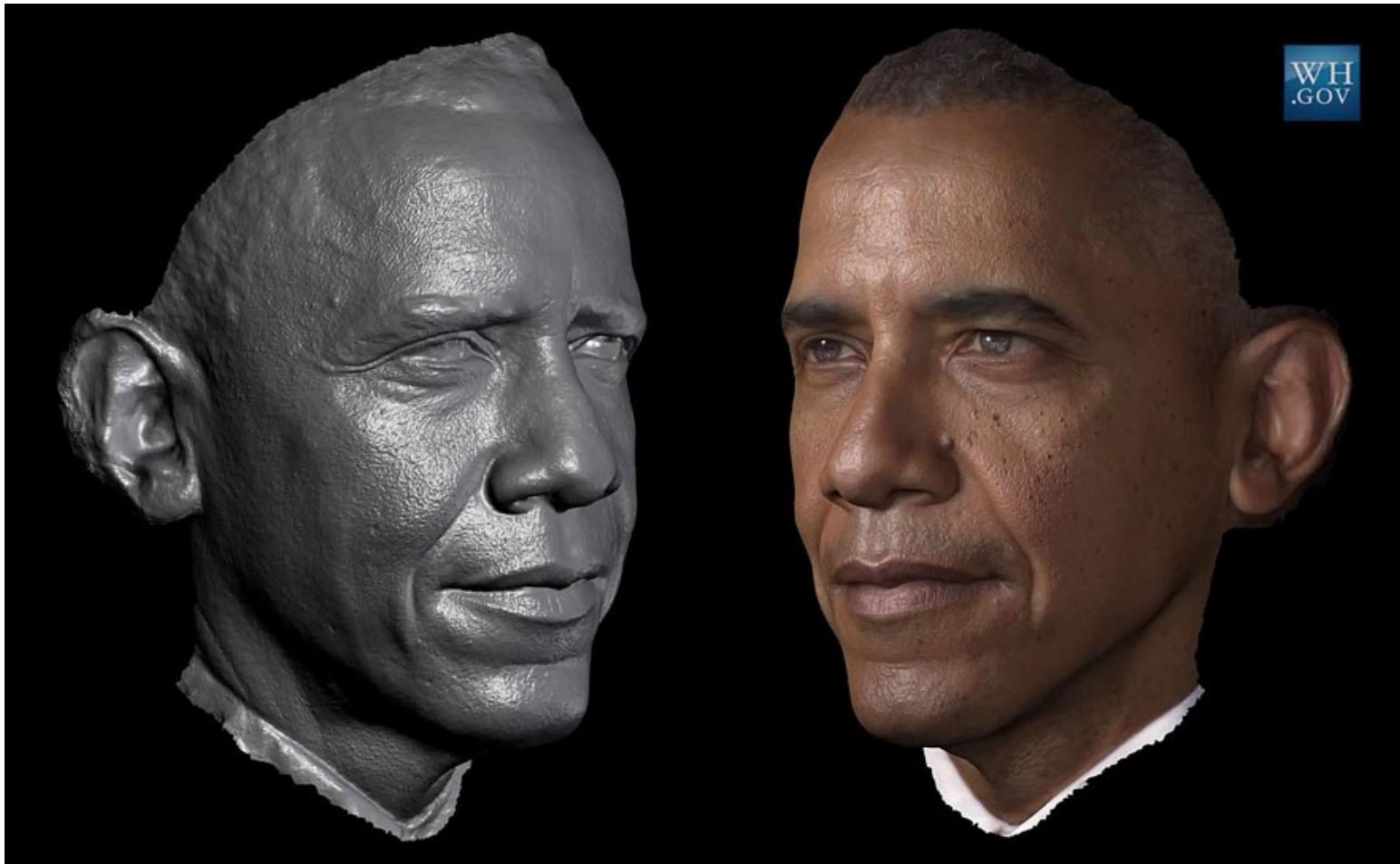
# Human shape capture



# Human shape capture



# Human shape capture



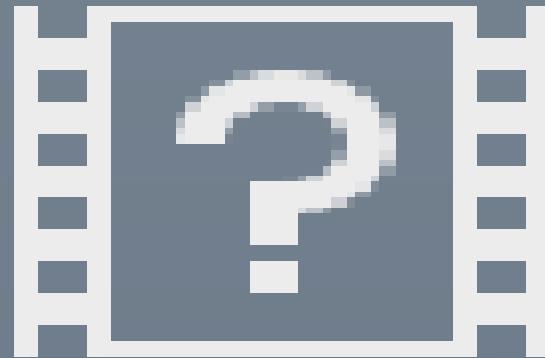
# Human shape capture



# Microsoft Holoportation



# Microsoft Holoportation



# Special effects: shape capture



*Star Wars: Rogue One – Peter Cushing / Admiral Tarkin*

# Special effects: shape capture



# Special effects: shape capture



# Special effects: Bad Examples

\$300 million budget



\$500 used computer



I taught an AI to shave Henry Cavill's mustache  
(<https://www.youtube.com/watch?v=2PZ3W1W20bk>)

# Special effects: Bad Examples?!



# Special effects: motion capture



# Interactive Games: Kinect

Object Recognition:

<http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o>

Mario: <http://www.youtube.com/watch?v=8CTJL5IUjHg>

3D: <http://www.youtube.com/watch?v=7QrnwoO1-8A>

Robot: <http://www.youtube.com/watch?v=w8BmgtMKFbY>

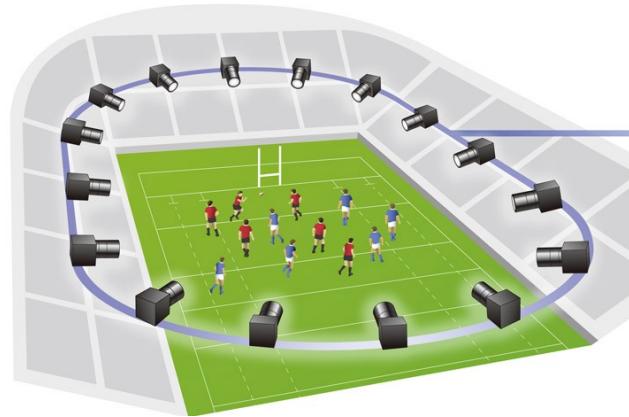


# Sports

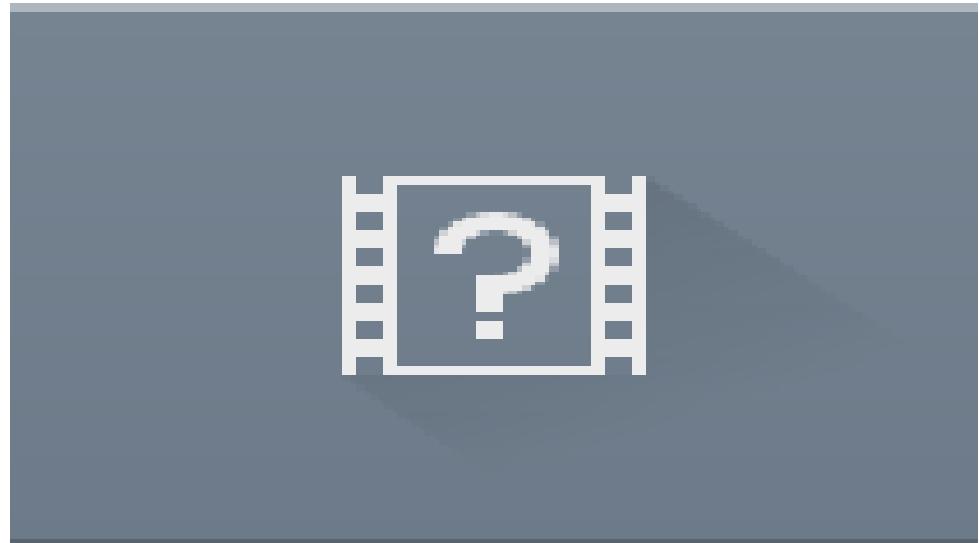


*Sportvision* first down line

Nice explanation on [www.howstuffworks.com](http://www.howstuffworks.com)

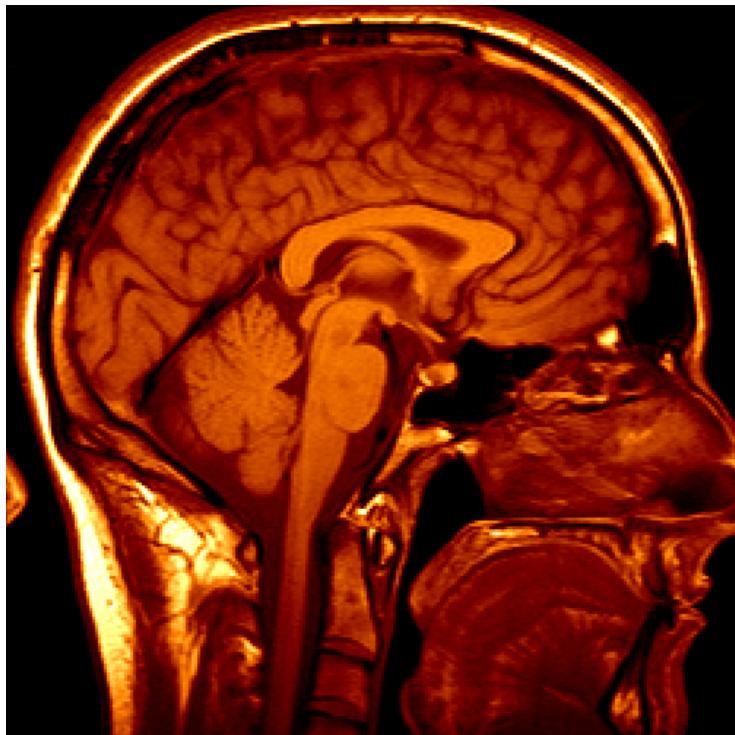


Free viewpoint video



[Canon 2017]

# Medical imaging



3D imaging  
MRI, CT



Image guided surgery

# AutoCars - Uber bought CMU's lab





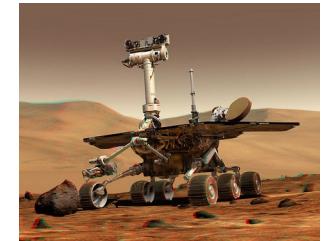


# Industrial robots



Vision-guided robots position nut runners on wheels

# Vision 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

Panorama stitching

3D terrain modeling

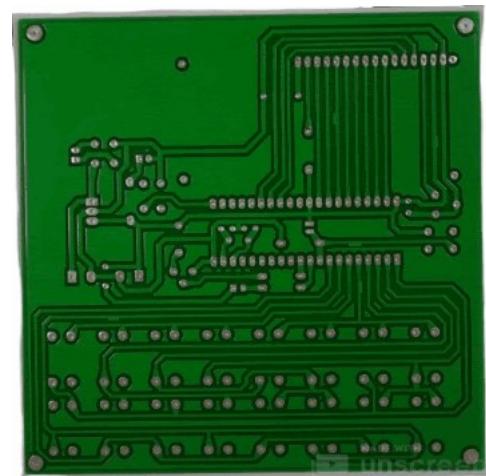
Obstacle detection, position tracking

# Industrial Settings

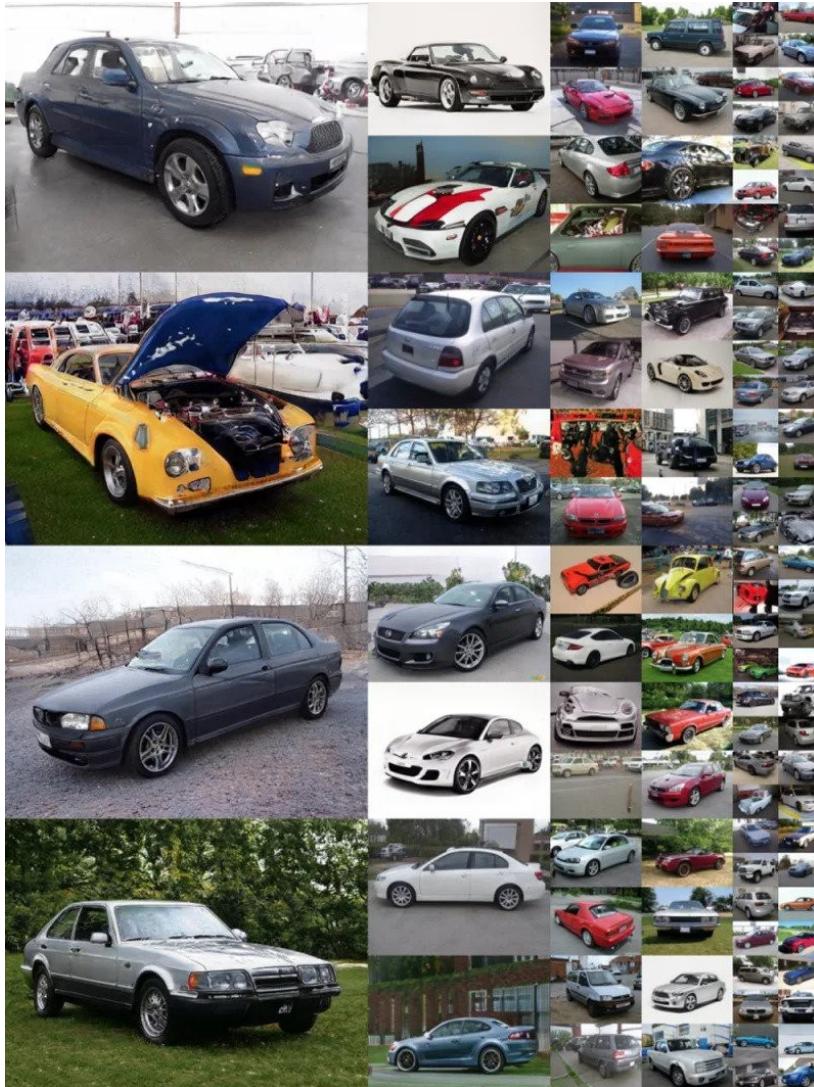
Worker Safety



Defect Inspection



# Deep Learning



StyleGAN  
(<https://arxiv.org/pdf/1812.04948.pdf>)

# Generative Networks

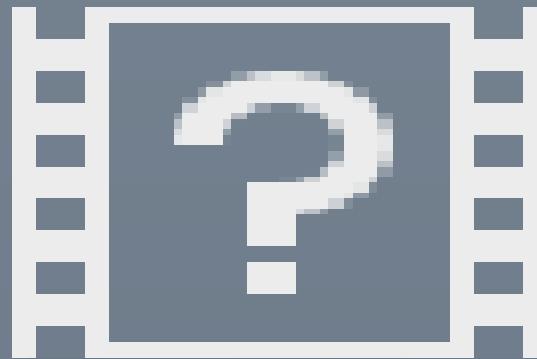


StyleGAN  
(<https://arxiv.org/pdf/1812.04948.pdf>)

# Living Portraits



# Even Further



# Scope of SYDE 671

Computer Vision  
Image Processing  
Recognition  
Deep Learning  
Geometric Reasoning

Machine Learning

Graphics

Computational Photography

Optics

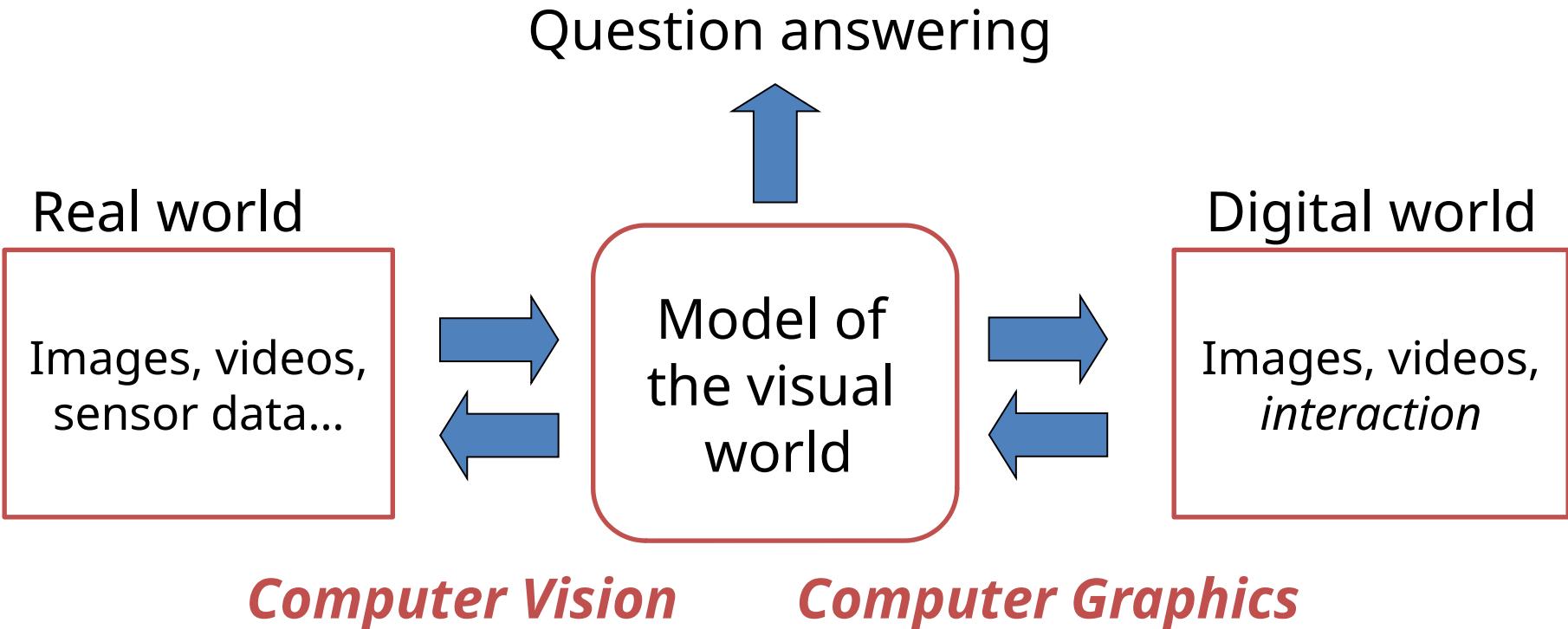
Robotics

Human Computer Interaction

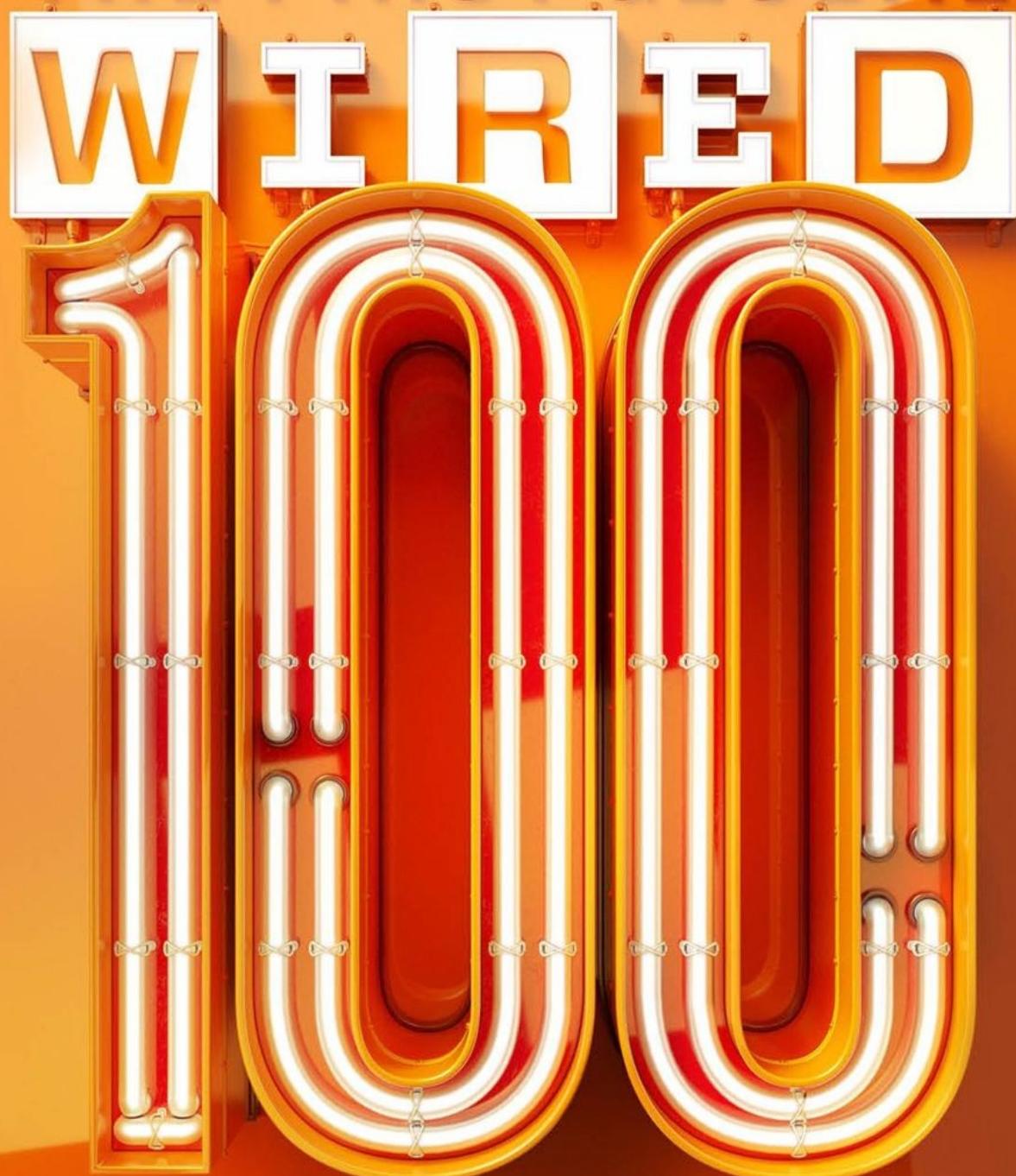
Medical Imaging

Neuroscience

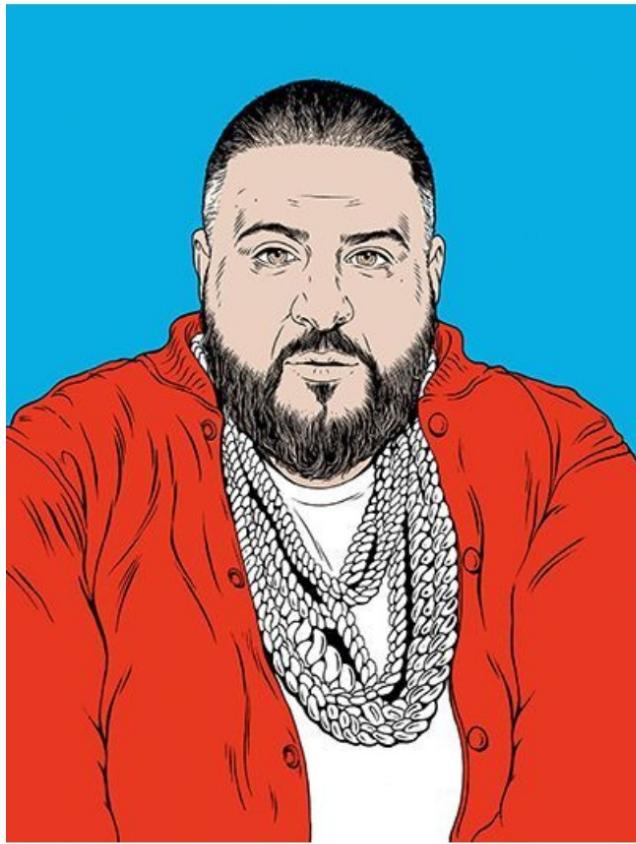
# Computer Vision and Nearby Fields



# WIRED 100



WHO'S SHAPING THE DIGITAL WORLD?



**DJ Khaled**

Credit [Louise Zergaeng Pomeroy](#)

## 73. DJ Khaled

*Snapchat icon; DJ and producer*

Louisiana-born Khaled Mohamed Khaled, aka DJ Khaled, cut his musical chops in the early 00s as a host for Miami urban music radio WEDR. He proceeded to build a solid if not dazzling career as a mixtape DJ and music producer (he founded his label We The Best Music Group in 2008, and was appointed president of Def Jam South in 2009).

# 69. Geoffrey Hinton

*Psychologist, computer scientist; researcher, Google Toronto*

British-born Hinton has been dubbed the "godfather of deep learning". The Cambridge-educated cognitive psychologist and computer scientist started being an ardent believer in the potential of neural networks and deep learning in the 80s, when those technologies enjoyed little support in the wider AI community.

But he soldiered on: in 2004, with support from the Canadian Institute for Advanced Research, he launched a University of Toronto programme in neural computation and adaptive perception, where, with a group of researchers, he carried on investigating how to create computers that could behave like brains.

Hinton's work – in particular his algorithms that train multilayered neural networks – caught the attention of tech giants in Silicon Valley, which realised how deep learning could be applied to voice recognition, predictive search and machine vision.

The spike in interest prompted him to launch a free course on neural networks on e-learning platform Coursera in 2012. Today, 68-year-old Hinton is chair of machine learning at the University of Toronto and moonlights at Google, where he has been using deep learning to help build internet tools since 2013.

## 66. PewDiePie

*YouTuber, Brighton*

In just six years, 26-year-old Swedish YouTuber Felix Kjellberg has risen from dropout who funded his videos by working at a hot-dog stand to a \$12 million global internet sensation. Kjellberg - aka [PewDiePie](#) - has built a quasi-cult fandom by uploading longish videos (Is This Game Too Sexual?!; Worst Game on the Planet!; What Happened To Resident Evil 7??) where he fools around and cracks jokes while playing video games.

In May 2016, Kjellberg's YouTube channel passed 45 million subscribers and, with 12 billion views, it's now the most-watched channel in the history of the video-sharing website. In the gaming industry, it is thought that he can single-handedly determine the success of a game by simply mentioning it in one of his videos.

Kjellberg - who lives in Brighton with his girlfriend and fellow YouTuber Marzia Bisognin, aka CutiePieMarzia - has tried to use his online clout for good, asking his fanbase (dubbed the "Bro Army") to raise money for charities on several occasions.

Recently, he teamed up with The Walking Dead creator Robert Kirkman to produce his first web series - Scare PewDiePie. The series, which features the YouTuber cavorting in horror-game-like situations, debuted on YouTube Red in February 2016.

## 63. Yann Lecun

*Director of AI research, Facebook, Menlo Park*

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

## 61. Taylor Swift

*Entertainer, Los Angeles*





Credit [Google DeepMind](#)



**Google**-backed startup DeepMind Technologies has built an [artificial intelligence](#) agent that can learn to successfully play 49 classic Atari games by itself, with minimal input.

## 8. Demis Hassabis

*Co-founder and CEO, DeepMind, London*

In March 2016, DeepMind's AI [AlphaGo](#) beat the Go world champion [Lee Se-dol](#). The Google-owned startup is moving machine learning forward at a pace that could affect every industry, from healthcare to commerce.

# Why is computer vision so difficult?

The complexity of the world is infinitely superior to the complexity of the measurements of its images.

The geometry of a scene: collection of surfaces

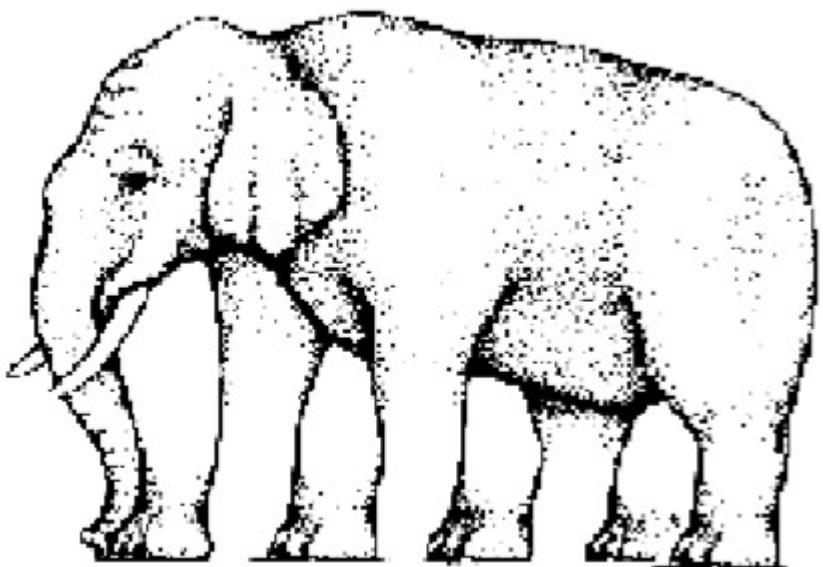
Photometry: a collection of functions that describe how light interacts with the underlying material in the scene

Dynamics of the scene: Differential equations

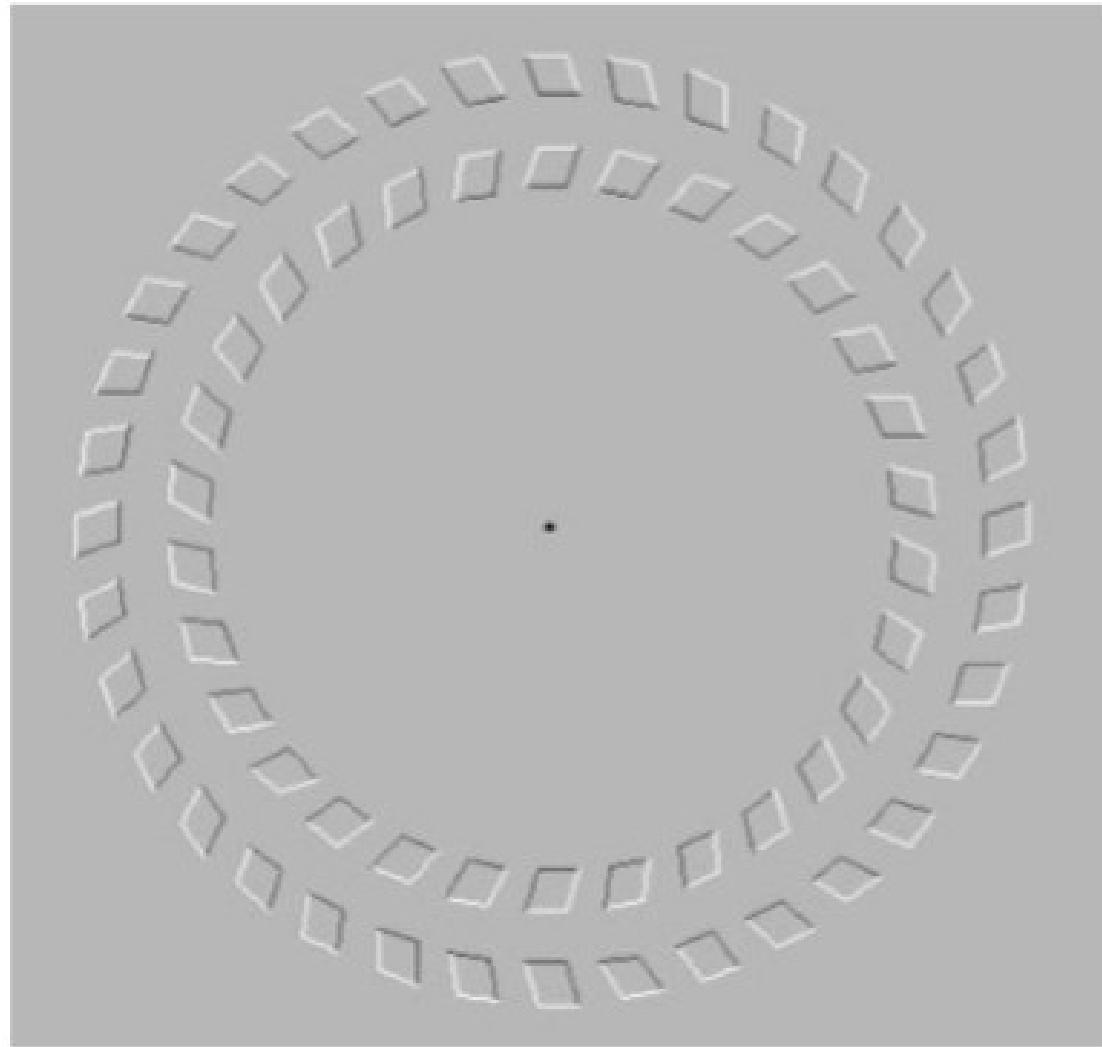
Inferring the world from finite-dimensional images is impossible without additional constraints and hypothesis.

What we can do is reconstruct a best 'model' of the world or internal representation

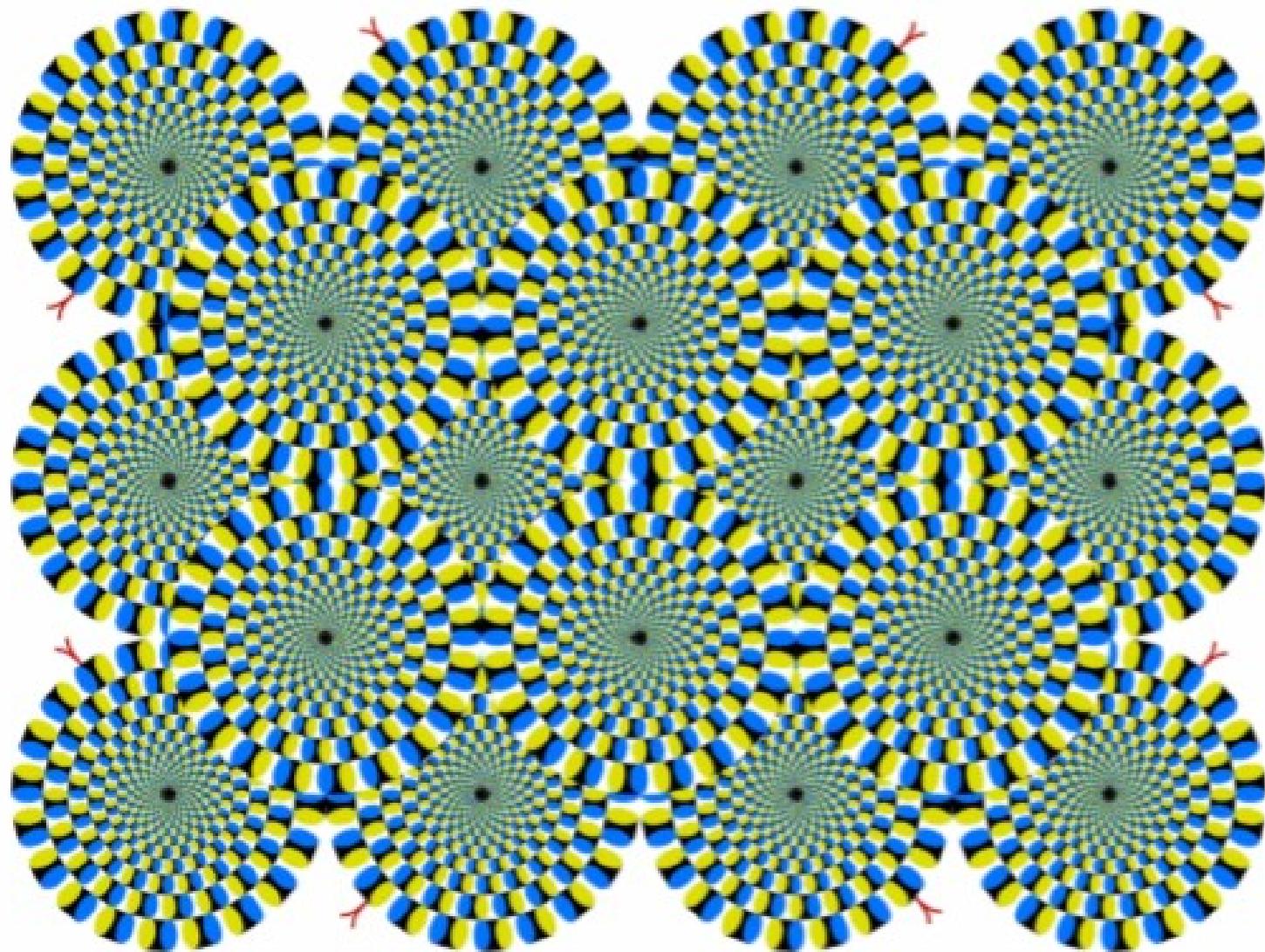
# What can go wrong



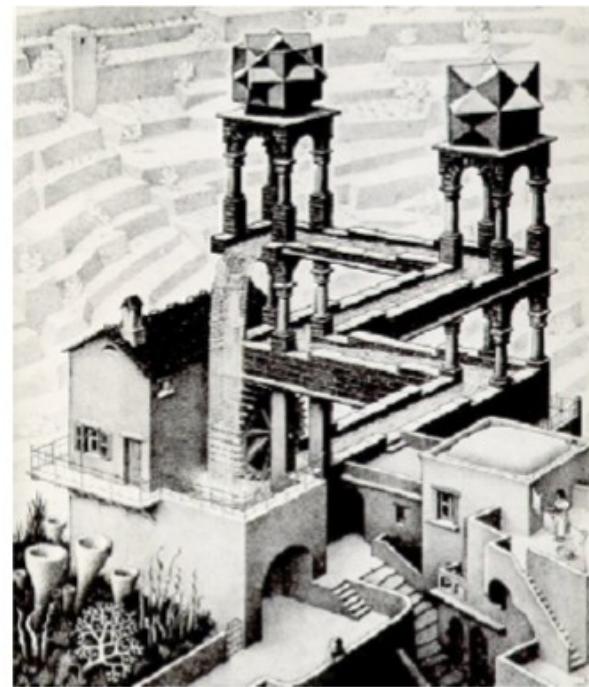
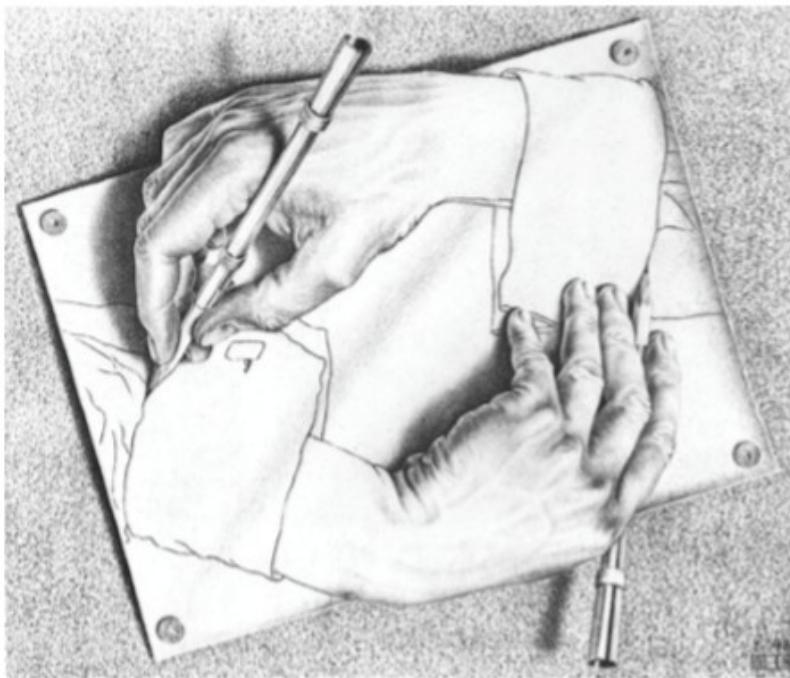
# What can go wrong



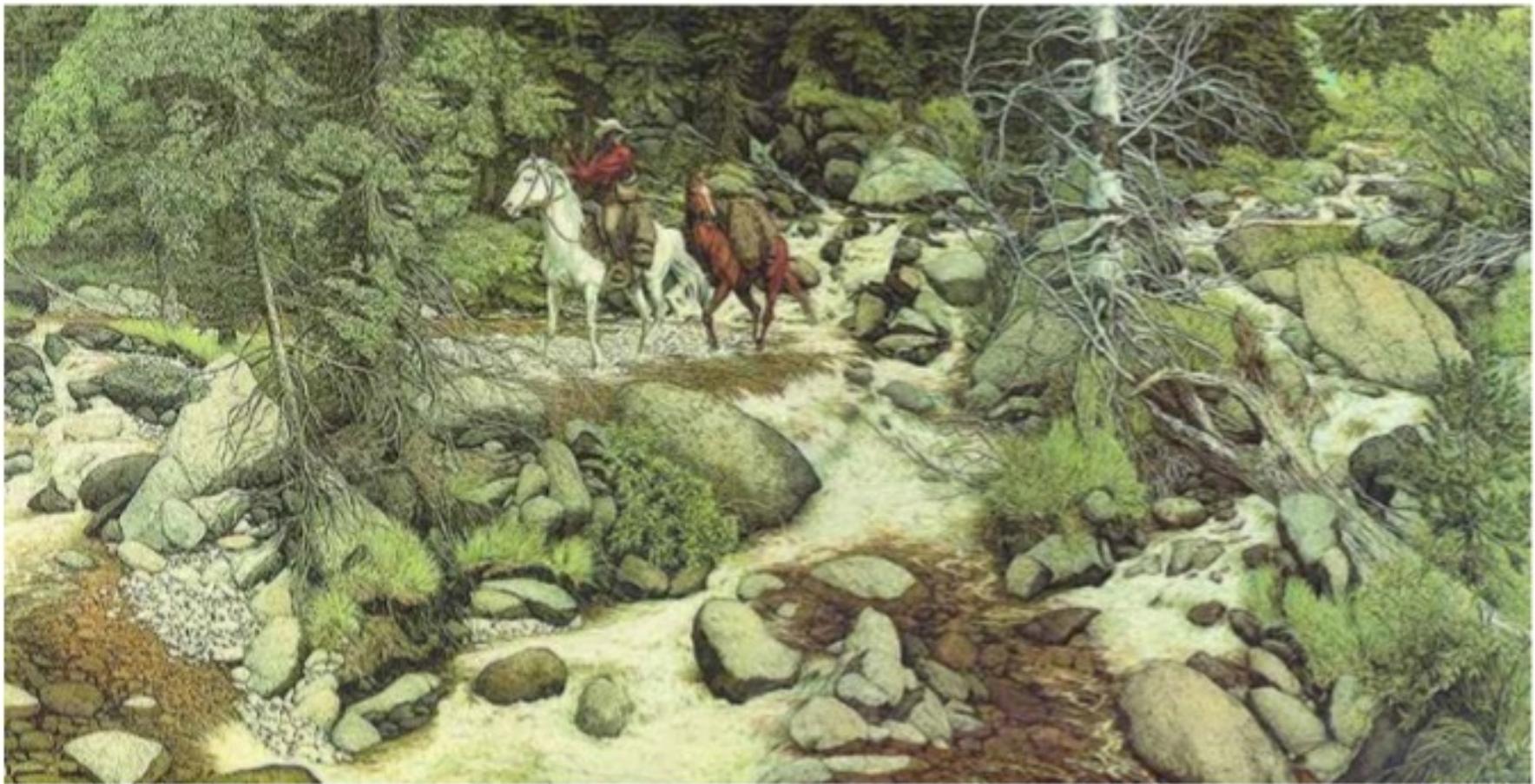
# What can go wrong



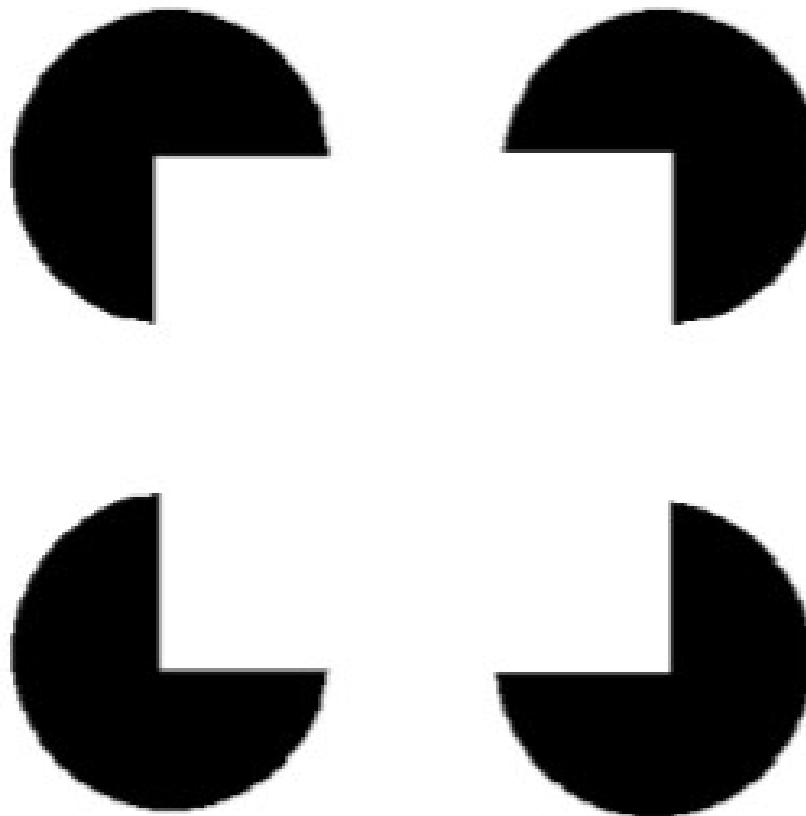
# Challenges: Impossibilities



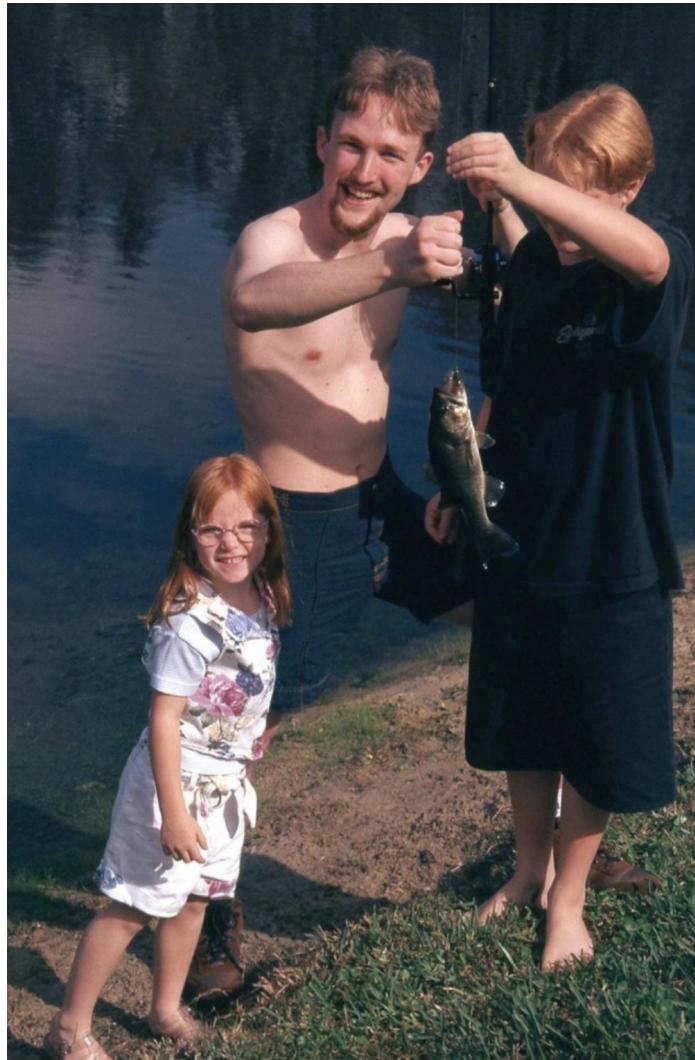
# Challenges: Subjectivity



# Challenges: Subjectivity



# Challenges : Hard Mode



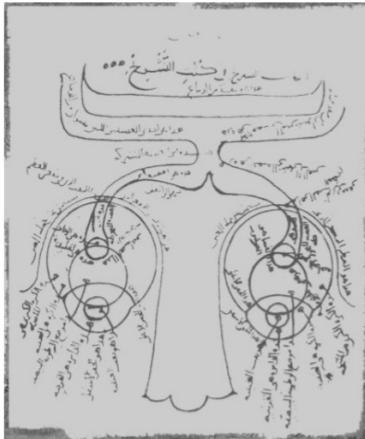
# How did we get here?!



Ibn al-Haytham

Born: 965 AD in Basra, Persia, Iran at that time (now it is in Iraq)  
Died: 1040 AD in Cairo, Egypt

Diagrammatic representation of the visual system from the oldest existing copy of the *Book of Optics*



Leonardo DaVinci

Born: 15 April 1452 in Vinci (near Empolia), Italy  
Died: 2 May 1519 in Cloux, Amboise, France

Drawing by Leonardo DaVinci of the projection of the eyes to the ventricles of the brain.

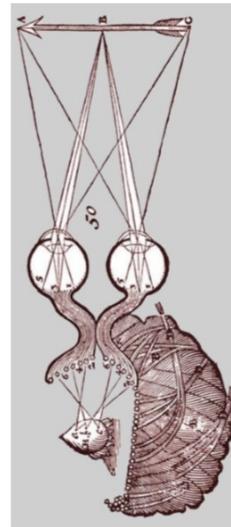


"cogito ergo sum"

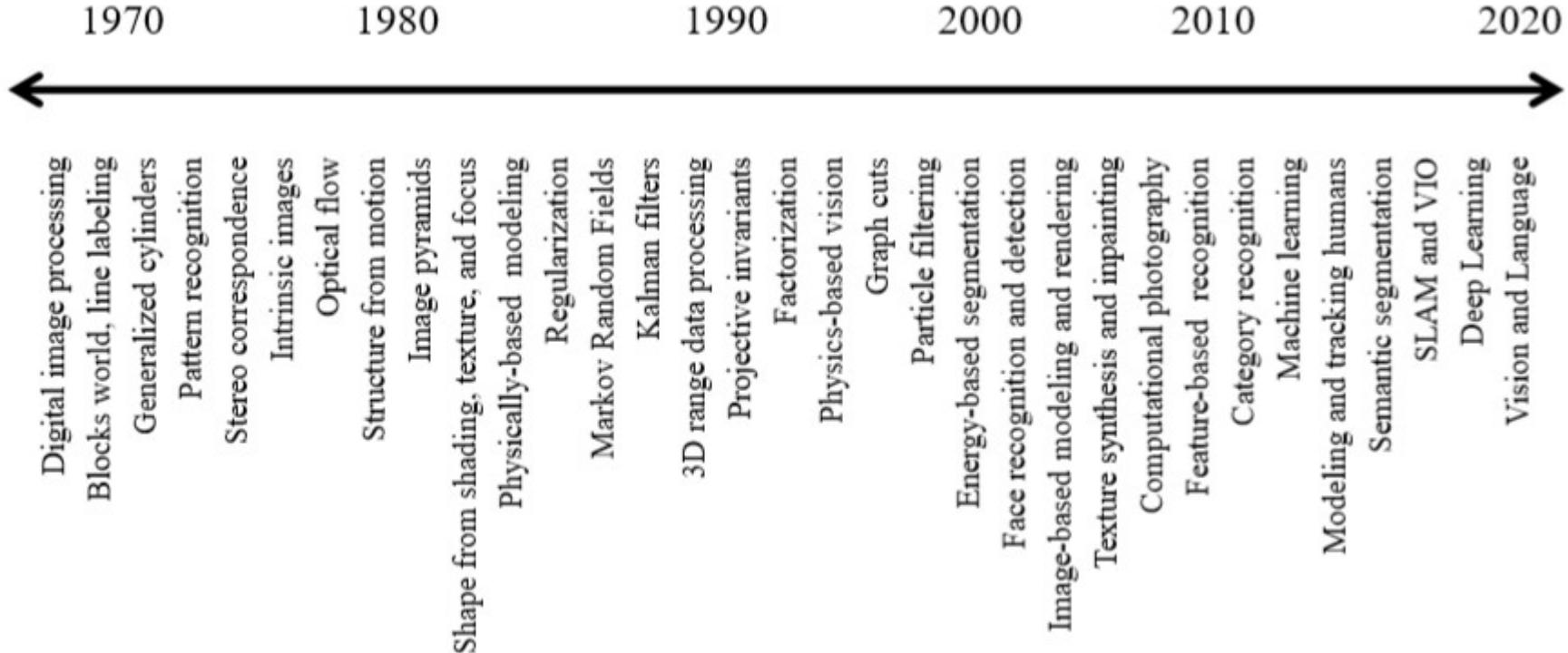
René Descartes

Born: 31<sup>st</sup> March 1596 in Touraine, France  
Died: 11<sup>th</sup> February 1650 in Stockholm, Sweden

Binocular stereoscopic visual system as imagined by Descartes. The two retinal images of the arrow are accurately, point for point, projected upon the surface of the cerebral ventricles and thence to the centrally located pineal gland, the supposed seat of imagination and common sense.



# Greatness from small beginnings



# Superhuman state of the art?

Deep learning is an enormous disruption to the field.  
Since 2012, rapid expansion and commercialization.

Why?

“With enough data, computer vision matches or even outperforms human vision at most recognition tasks.”

## WHAT?

# Vision and Society

Lots of data = lots of potential bias in the data.

Needs understanding of possible failures.

+

Responsible approach.

+

Techniques to overcome bias.