APPLIED COMPUTER VISION IN UBICOMP

CSE 590 Ubiquitous Computing | Lecture 7 | May 10

Jon Froehlich • Liang He (TA)

DESIGN USE BUILD

MAKEABILITY LAB

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

UNIVERSITY of WASHINGTON

SCHEDULE TODAY: 6:30-9:20

06:30-06:50: Discussion of required reading led by Pankaj Parag

06:50-07:10: A3 Nite Lite “Show-and-Tell”

07:10-07:15: Overview of A4 “Smart Spaces” assignment

07:15-08:10: Applied Computer Vision in UbiComp/HCI

08:10-08:15: Break

08:15-09:20: FaceTrackerBLE Exercises

SCHEDULE TODAY: 6:30-9:20

06:30-06:50: Discussion of required reading led by Pankaj Parag

06:50-07:10: A3 Nite Lite “Show-and-Tell”

07:10-07:15: Overview of A4 “Smart Spaces” assignment

07:15-08:10: Applied Computer Vision in UbiComp/HCI

08:10-08:15: Break

08:15-09:20: FaceTrackerBLE Exercises

SCHEDULE TODAY: 6:30-9:20

06:30-06:50: Discussion of required reading led by Pankaj Parag

06:50-07:10: A3 Nite Lite “Show-and-Tell”

07:10-07:15: Overview of A4 “Smart Spaces” assignment

07:15-08:10: Applied Computer Vision in UbiComp/HCI

08:10-08:15: Break

08:15-09:20: FaceTrackerBLE Exercises

CSE590 ASSIGNMENT

A4 "SMART SPACES" OVERVIEW

Overview

Having rolled out a very successful line of IoT night lights :), your company has tasked you with prototyping a new line of products related to smart spaces. In particular, they would like to enhance kitchen spaces with a smart tracking system that, for example, can help occupants find ingredients, track product usage for auto-reordering, and even warn parents when children get too close to a hot stove or oven. As a two-week course assignment, we are going to simplify things and only focus on the last scenario.

Using some of the computer vision (CV) concepts learned in class along with posted sample code, you are going to create a CV-based motion tracking system that:

- Detects the face of and tracks the movement of at least one user in the Android camera view
- Measures the distance to that user using the ultrasonic sensor, which is mounted on a servo motor. Using the motion tracking information from Android, the servo motor automatically rotates the ultrasonic sensor so that it is always facing the detected user. We have posted example code for the servo here and the ultrasonic sensor here. Note that the servo motor only rotates from 0-180 degrees.
- Flashes an LED and plays a tone using the piezoelectric buzzer when the detected user is within 0.5 meters
- Has a functional lo-fi stand for the servo that allows the servo to sit robustly on a table or other surface even when the motor is sweeping back and forth and a lo-fi attachment mechanism that affixes the ultrasonic sensor to the servo motor. For A5, you will design and fabricate 3D-printed versions.
- Includes one creative feature related to the "smart spaces" theme (*e.g.*, face recognition, speech interaction, hand tracking + gesture recognition)

The **midpoint checkin** includes properly tracking a face and automatically adjusting the servo motor accordingly.

SCHEDULE TODAY: 6:30-9:20

06:30-06:50: Discussion of required reading led by Pankaj Parag

06:50-07:10: A3 Nite Lite “Show-and-Tell”

07:10-07:15: Overview of A4 “Smart Spaces” assignment

07:15-08:10: Applied Computer Vision in UbiComp/HCI

08:10-08:15: Break

08:15-09:20: FaceTrackerBLE Exercises

The goal of this lecture is to provide some **inspiration** and a **broad overview of CV**. Next week, we'll dive deeper.

CV OVERVIEW

INTERPRET THIS IMAGE...

Source: Andrej Karpathy, The state of Computer Vision and AI: we are really, really far away, http://karpathy.github.io/2012/10/22/state-of-computer-vision/

CV OVERVIEW

INTERPRET THIS IMAGE...

Andrej Karpathy blog

About Hacker's guide to Neural Networks

The state of Computer Vision and AI: we are really, really far away.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

I could go on, but the point here is that you've used a HUGE amount of information in that half second when you look at the picture and laugh. Information about the 3D structure of the scene, confounding visual elements like mirrors, identities of people, affordances and how people interact with objects, physics (how a particular instrument works, leaning and what that does), people, their tendency to be insecure about weight, you've reasoned about the situation from the point of view of the person on the scale, what he is aware of, what his intents are and what information is available to him, and you've reasoned about people reasoning about people.

standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very

Source: Andrej Karpathy, The state of Computer Vision and AI: we are really, really far away, http://karpathy.github.io/2012/10/22/state-of-computer-vision

GOAL OF COMPUTER VISION

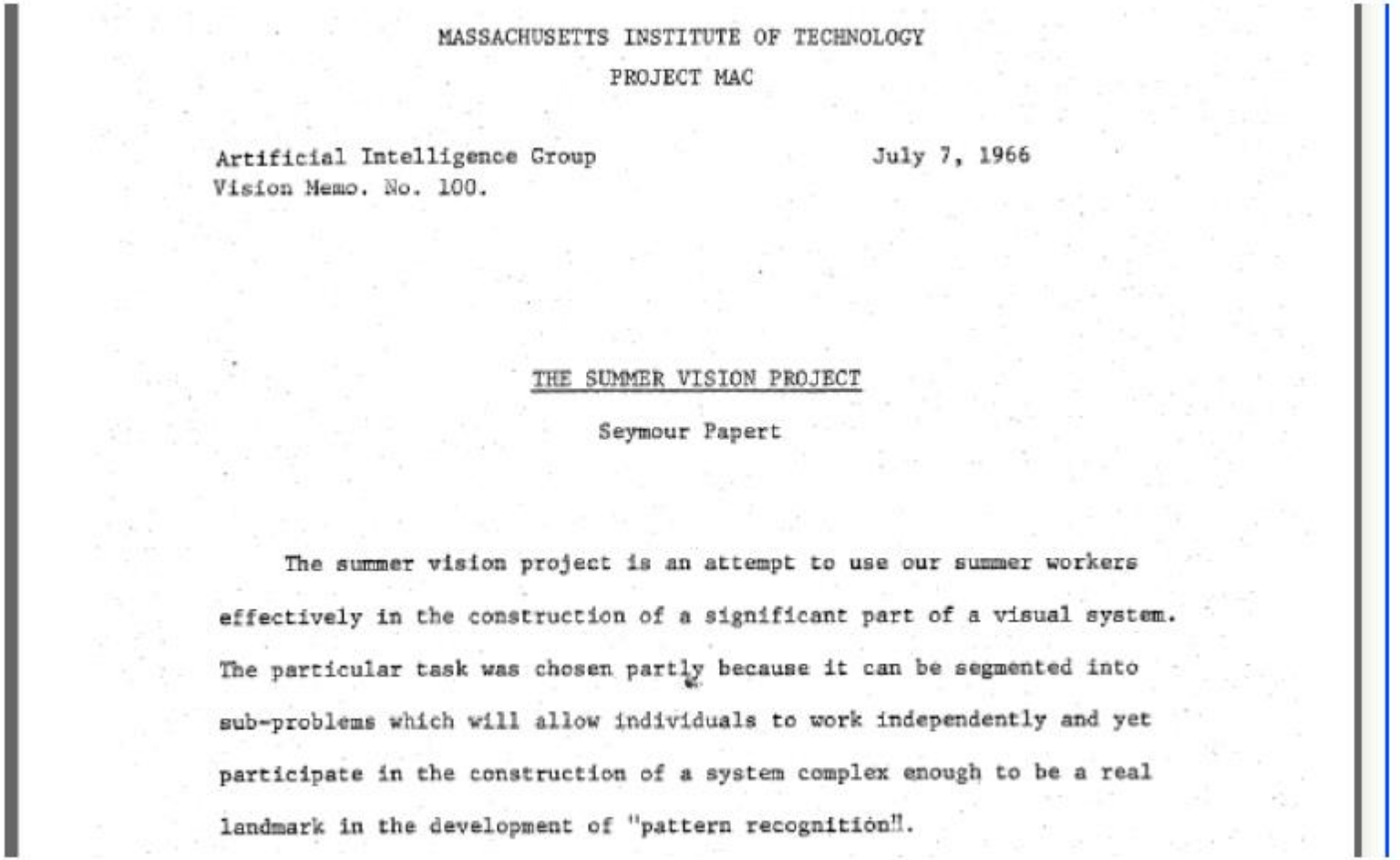
Bridge the gap between pixels and meaning (and "meaning" is deeply complex—social, cultural, *etc.*)

What we see

What a computer sees

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

ORIGINS OF CV: AN MIT UGRAD SUMMER PROJECT!



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

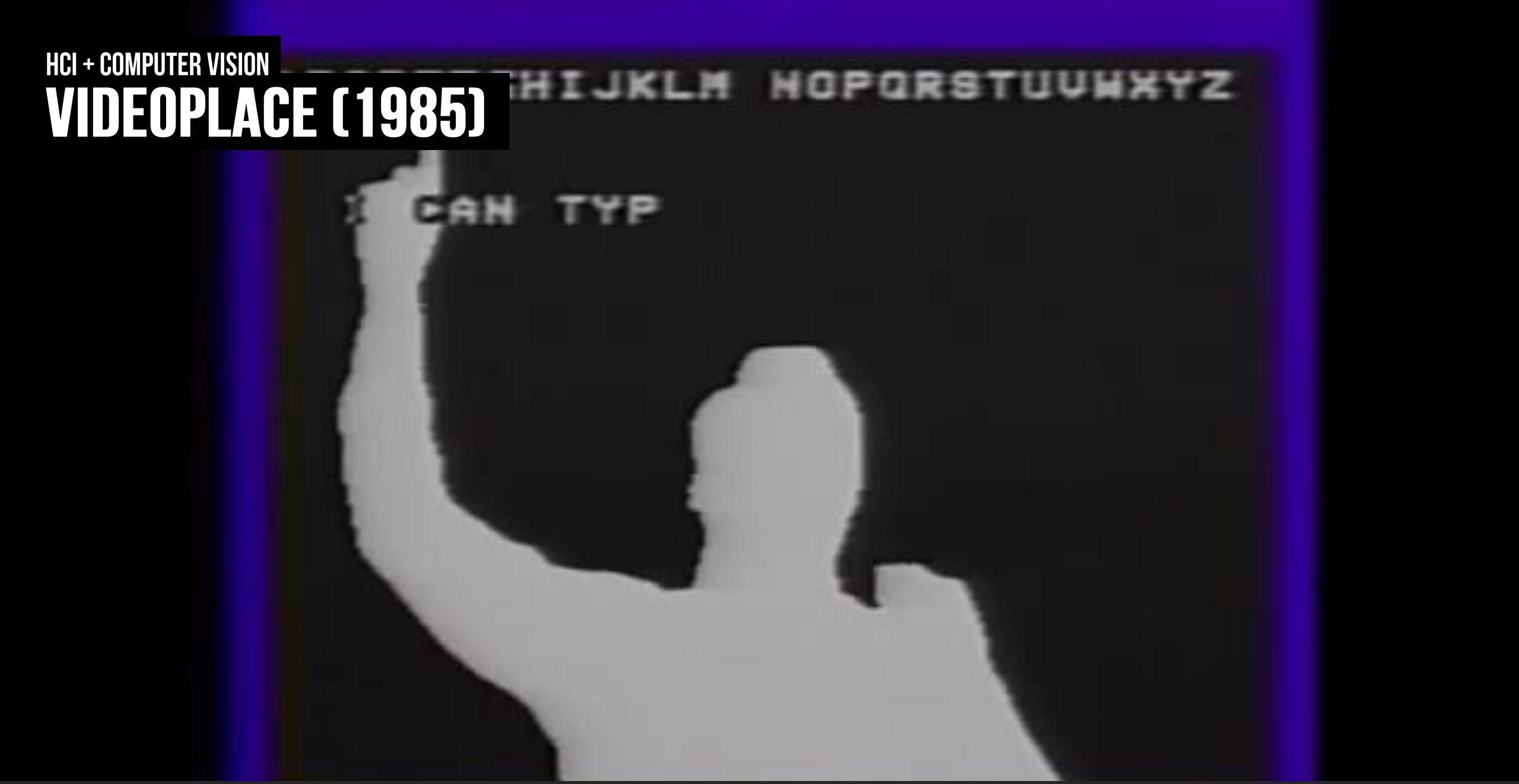
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Based on slides by Fei-Fei Li for Stanford Computer Vision course; see also Golan Levin's essay: http://www.flong.com/texts/essays/essay_cvad

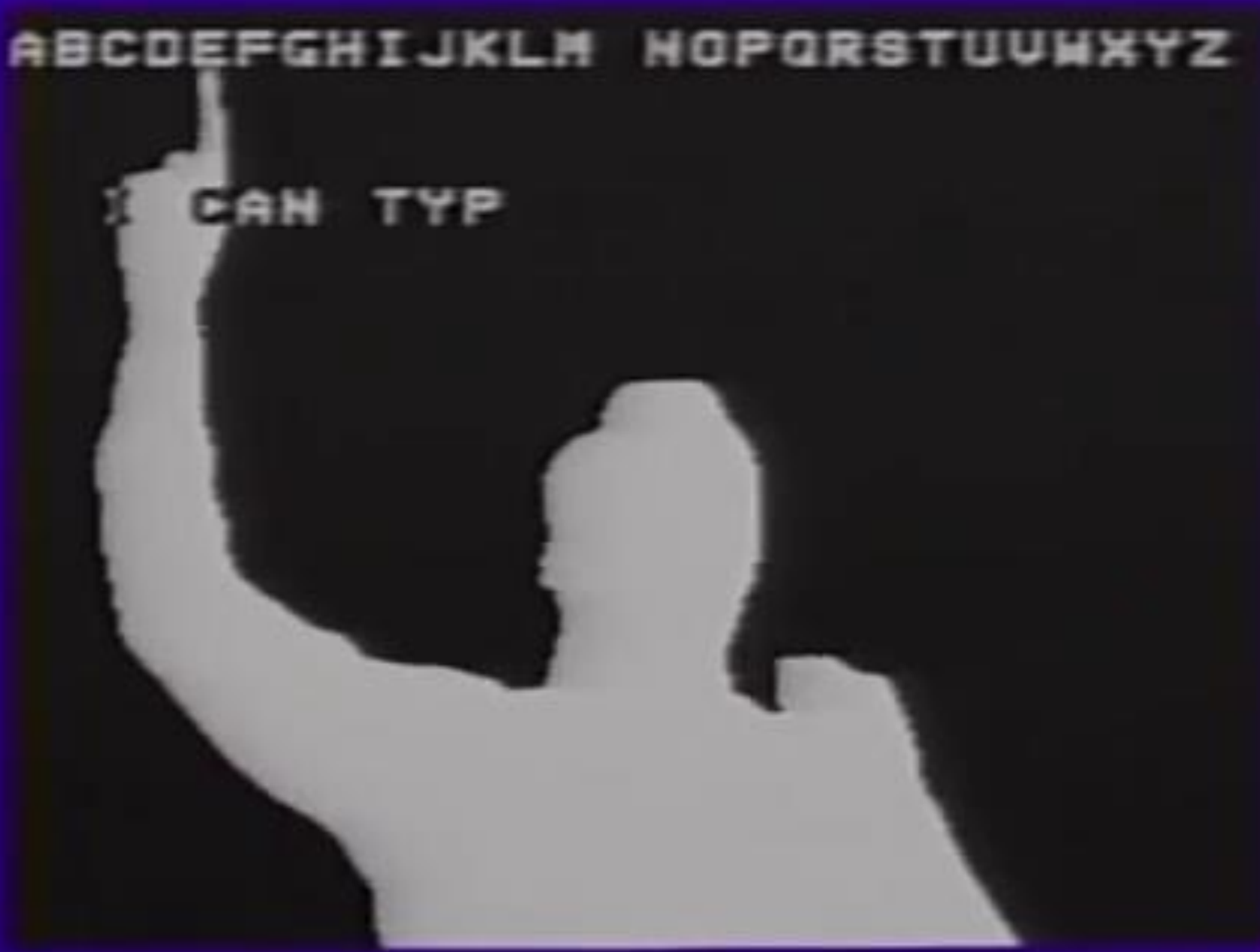
There is a **long history** of trying to use vision for HCI/UbiComp

HCI + COMPUTER VISION

VIDEOPLACE (1985)



Creator: Myron Krueger, https://youtu.be/d4DUIeXSEpk


ABCDEFGHIJKLM NOPQRSTUVWXYZ
I CAN TYP

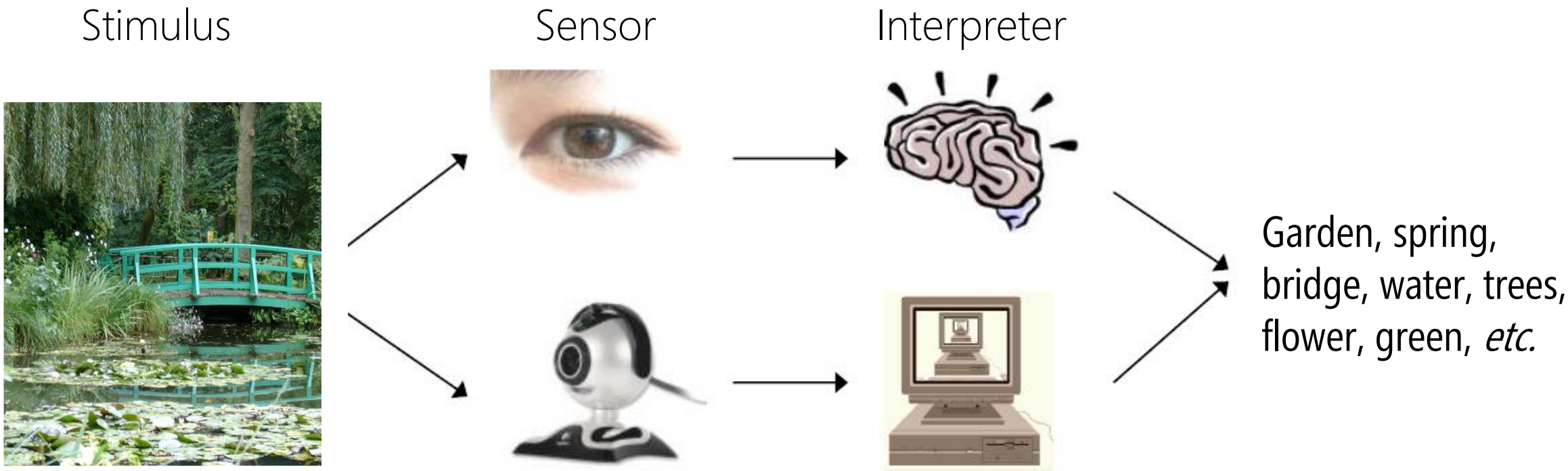
HCI + COMPUTER VISION

VIDEOPLACE (1990)

Creator: Myron Krueger, https://youtu.be/5sGeEnGos0Y

CV OVERVIEW

WHAT IS VISION?



CAN COMPUTERS MATCH HUMAN VISUAL SYSTEM?

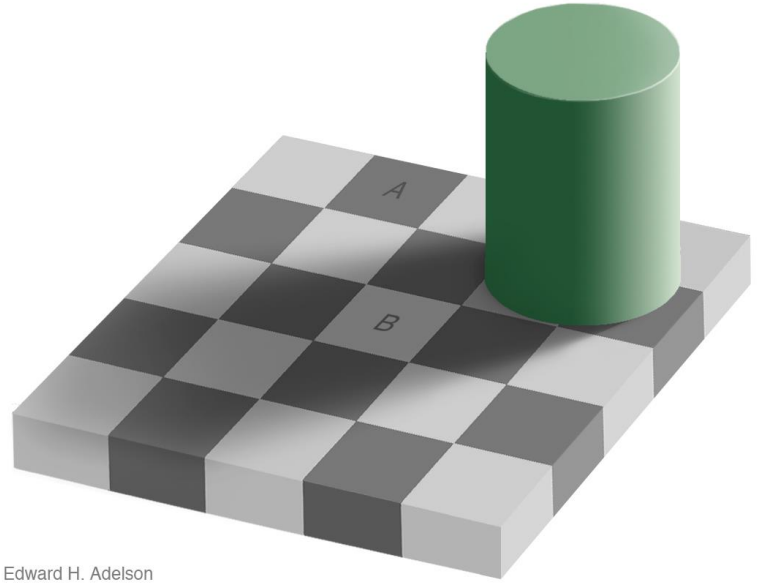
Sometimes yes but most often no

However, fascinating time to work in this area due to cloud-based AI services and deep learning

What is considered hard keeps changing…

EVEN HUMAN PERCEPTION HAS ITS SHORTCOMINGS

Sinha and Poggio, *Nature*, 1996



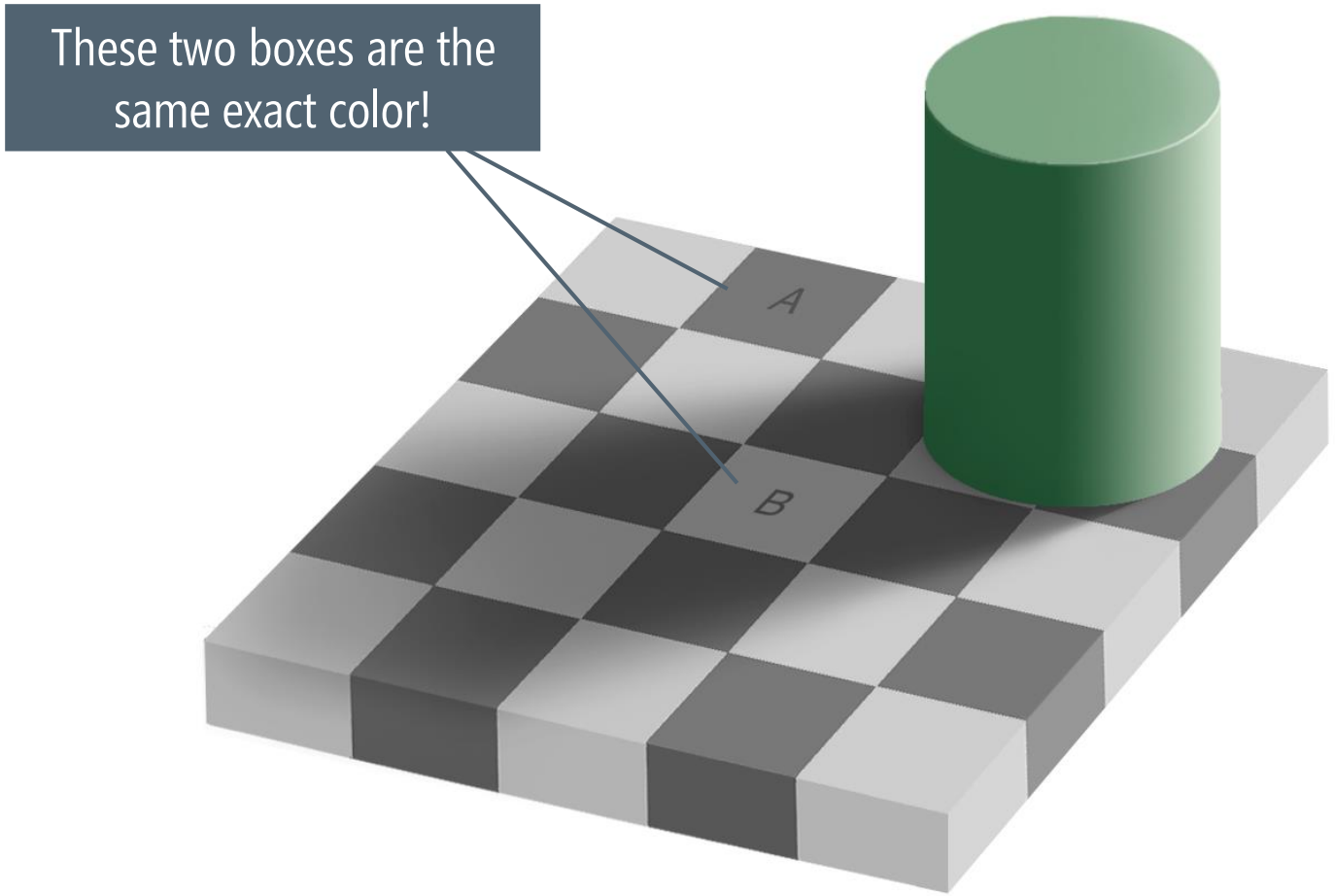
Edward H. Adelson

Adelson's Checkershadow

Change Blindness

HUMAN VISUAL SYSTEM

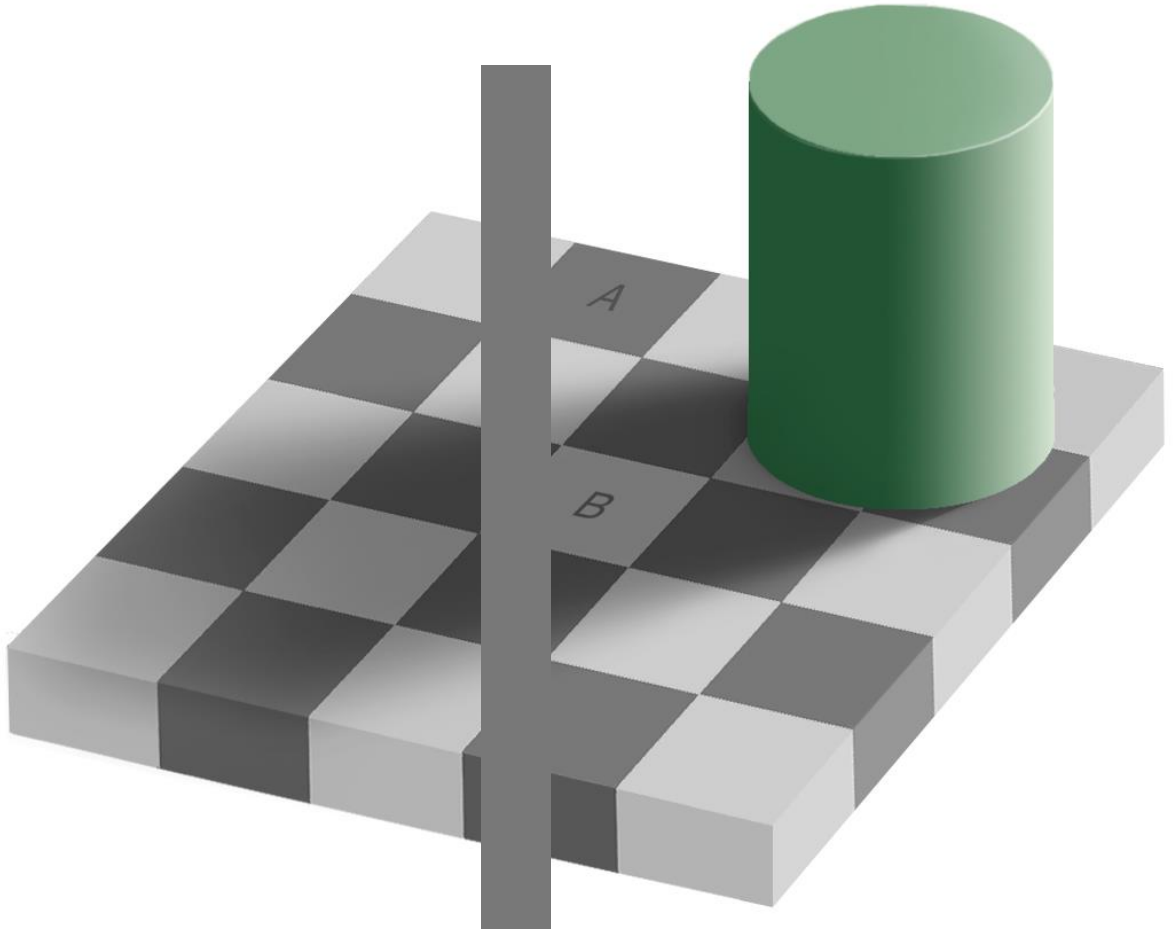
ADELSON'S CHECKERSHADOW



Source: Adelson, http://persci.mit.edu/gallery/checkershadow

HUMAN VISUAL SYSTEM

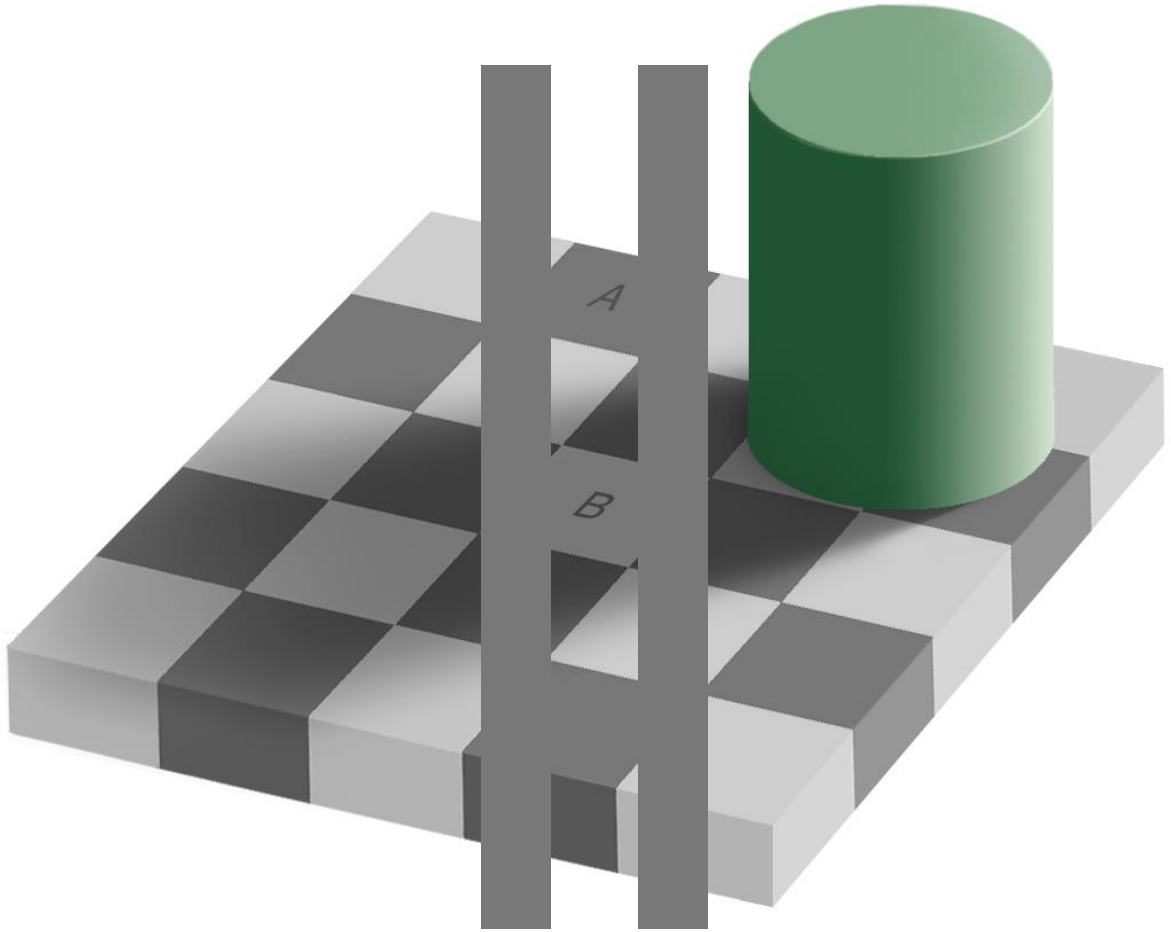
ADELSON'S CHECKERSHADOW



Source: Adelson, http://persci.mit.edu/gallery/checkershadow

HUMAN VISUAL SYSTEM

ADELSON'S CHECKERSHADOW



Source: Adelson, http://persci.mit.edu/gallery/checkershadow

HUMAN VISUAL SYSTEM

CHANGE BLINDNESS

Visual disruption: flicker

Visual disruption: noise

Visual disruption: none

Source: O'Regan, Encyclopedia of Cognitive Science, http://nivea.psycho.univ-paris5.fr/ECS/ECS-CB.html

BUT MOSTLY THE HUMAN VISUAL SYSTEM IS REMARKABLY EFFICIENT

VISUAL COGNITION, 2005, 12 (6), 852–877

Psychology Press
Taylor & Francis Group

How long to get to the ''gist'' of real-world natural scenes?

Guillaume A. Rousselet, Olivier R. Joubert, and Michèle Fabre-Thorpe

Centre de Recherche Cerveau et Cognition, UMR 5549 (CNRS-UPS), Toulouse, France

This study aimed at assessing the processing time of a natural scene in a fast categorization task of its context or ''gist''. In Experiment 1, human subjects performed 4 go/no-go categorization tasks in succession with colour pictures of real-world scenes belonging to 2 natural categories: ''Sea'' and ''mountain'', and 2 artificial categories: ''Indoor'' and ''urban''. Experiment 2 used colour and grey-level scenes in the same tasks to assess the role of colour cues on performance. Pictures were flashed for 26 ms. Both experiments showed that the gist of real-world scenes can be extracted with high accuracy (>90%), short median RT (400–460 ms) and early responses triggered with latencies as short as 260–300 ms. Natural scenes were processed faster than artificial scenes. Categories for which colour could have a diagnostic value were processed faster in colour than in grey. Finally, processing speed is compared for scene and object categorization tasks.

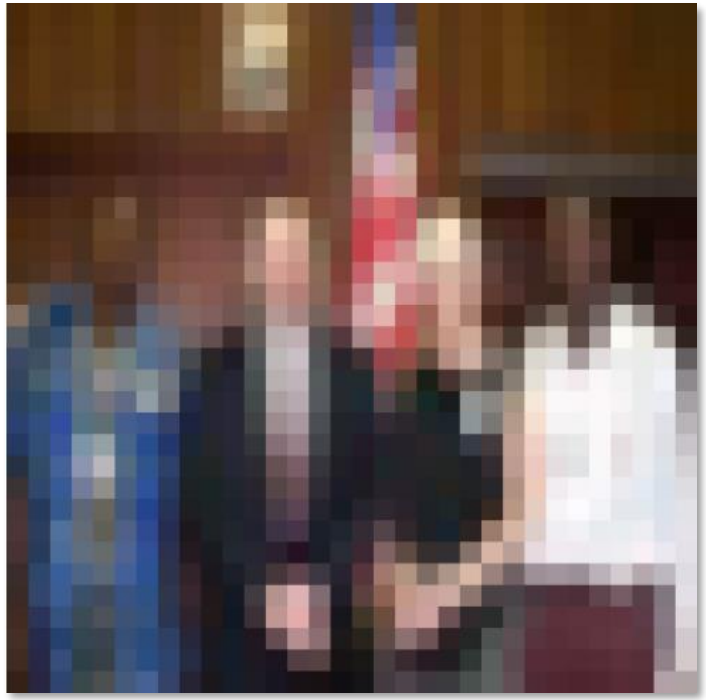
Natural scenes are more than a simple collection of objects. However, much of the research on scene processing has been devoted to the understanding of object processing in scenes, leaving aside the question of how we process the whole scene itself.

This issue is important given that we do not only process objects but we also analyse the context in which they appear. Global coarse information about a scene (mainly its category, or *gist*, and its spatial structure, or *layout*) is relatively crucial in memory-free models of scene perception in which little information is integrated across saccades. According to this idea, perception is constructed by integrating abstract scene representations with volatile object

Please address all correspondence to: Guillaume A. Rousselet, McMaster University, Department of Psychology, Hamilton, L8S4K1, ON, Canada. Email: rousseg@mcmaster.ca

This work was supported by the CNRS, the ACI ''Integrative and Computational Neurosciences'', and the Cognitique grant no. IC2. Financial support was provided to G. A. Rousselet by a PhD grant from the French government. We thank Anne-Sophie Paroissien for her very valuable help running the experimental sessions in Experiment 1.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280444000553



Source: “80 million tiny images” by Torralba, et al.

What can **CV** do?

THREE OPERATIONAL LEVELS

Low-level operations

Image enhancement, feature detection, region segmentation

Mid-level operations

Structure-from-motion, optical flow tracking

High-level operations

Recognition of people, places, events

IMPROVE PHOTOS ("COMPUTATIONAL PHOTOGRAPHY")

Super-resolution (source: 2d3)

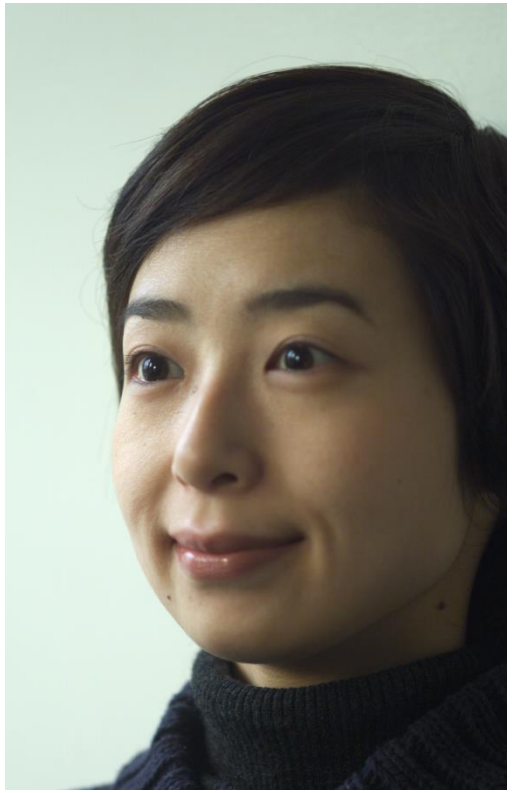
Low-light photography
(credit: Hasinoff et al., SIGGRAPH ASIA 2016)

Depth of field on cell phone camera
(source: Google Research Blog)

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

WHAT CAN CV DO?

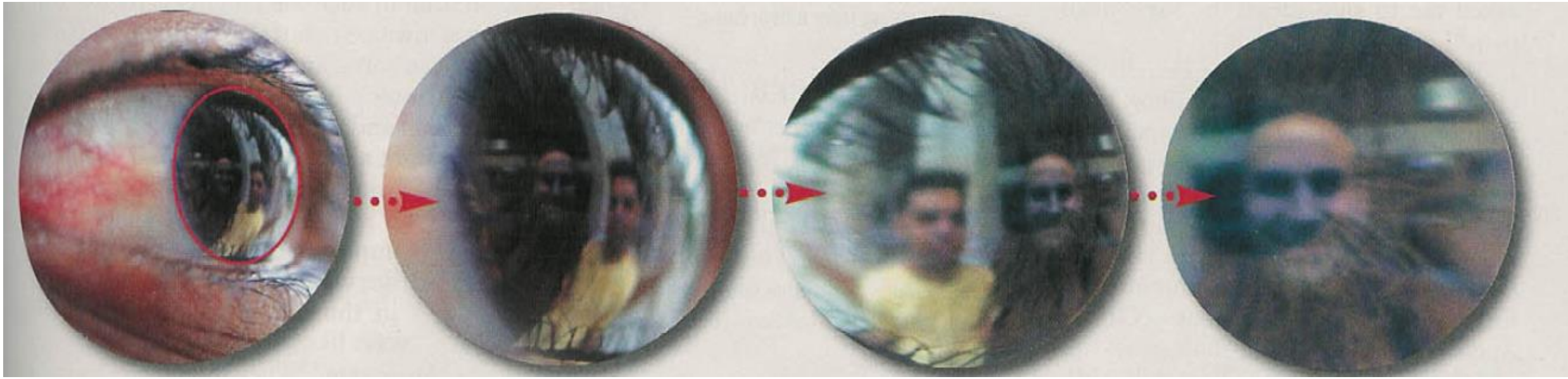
FORENSICS



Original source by Nayar and Nishino *Eyes for Relighting*, Based on slides by Noah Snavely at Cornell: http://www.cs.cornell.edu/courses/cs5670/2018sp/lectures/lectures.html

WHAT CAN CV DO?

FORENSICS



Original source by Nayar and Nishino *Eyes for Relighting*, Based on slides by Noah Snavely at Cornell: http://www.cs.cornell.edu/courses/cs5670/2018sp/lectures/lectures.html

WHAT CAN CV DO?

ACCESSIBILITY

ISRAELI VISUAL AID COMPANY ORCAM VALUED AT $1 BILLION

OrCam's device reads texts, supermarket barcodes and recognizes faces while speaking the information into the user's ear.

BY REUTERS / FEBRUARY 22, 2018 12:26

HCI + COMPUTER VISION

NANT MOBILE MONEY READER

Video source: Jon Froehlich (shot by Liang He)


FEDERAL RESERVE NOTE
THE UNITED STATES OF AMERICA
MK 07219676 B
MK07219676B
20
TWENTY DOLLARS
SERIES 2013
JACKSON

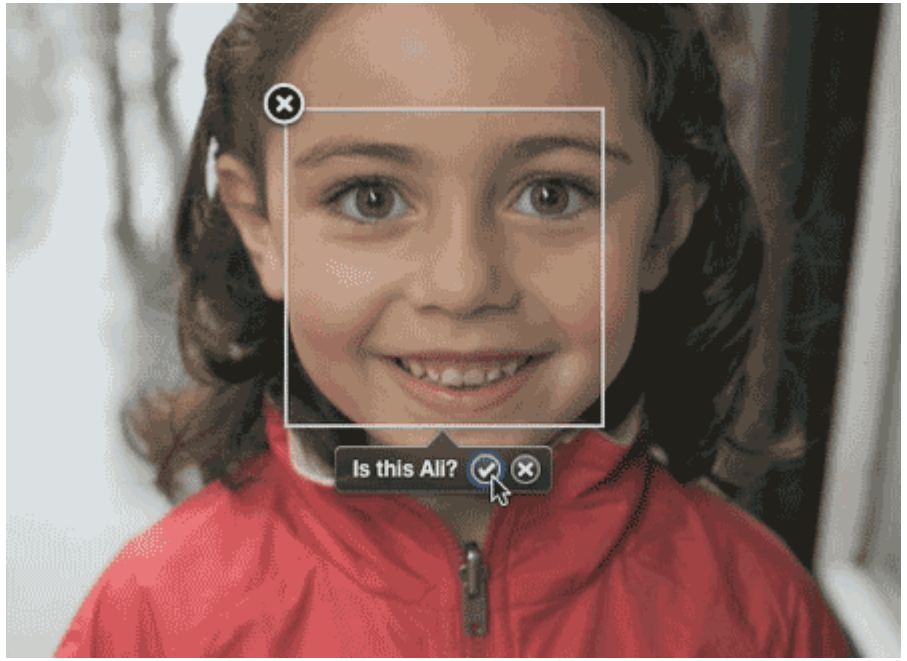
WHAT CAN CV DO?

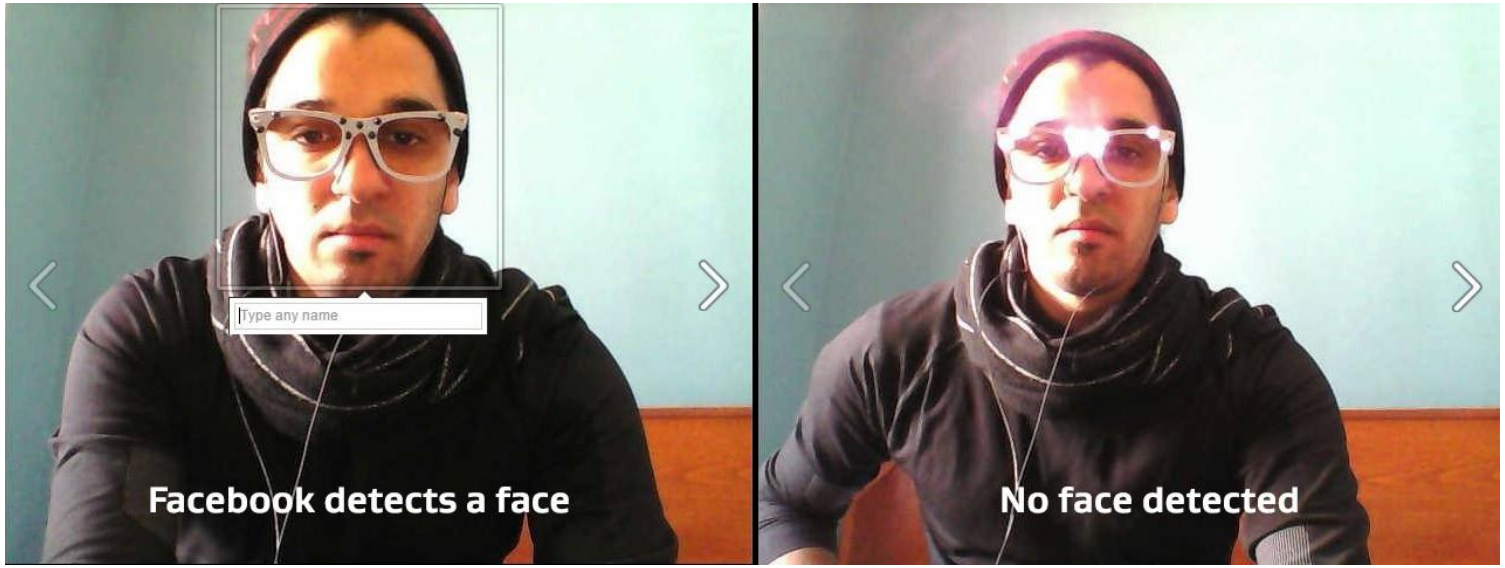
FACE RECOGNITION

Why is face detection on cameras useful?

WHAT CAN CV DO?

FACE RECOGNITION





Based on slides by Noah Snavely at Cornell: http://www.cs.cornell.edu/courses/cs5670/2018sp/lectures/lectures.html

WHAT CAN CV DO?

FACE RECOGNITION

Sighthound PRODUCTS TECHNOLOGY PARTNERS BLOG SUPPORT MY ACCOUNT | DOWNLOADS

Face Recognition Performance

Sighthound is a leader in deep learning, specializing in building intelligent convolutional neural networks that can run real time in resource-constrained environments, including on embedded systems. Sighthound's facial recognition system leads the world, with measured accuracy of 99.2% against PubFig200, some ten percentage points clear of other commercial or academic methods.

Our model uses 3.8 million images, or about 1.5% of the number of images Google lists for training FaceNet. Our pipeline uses only a single crop of the face. Crops are partial areas of training images used to enhance results. However each additional crop slows down performance in recognition tasks, meaning the models are optimized for test-taking, not for real world use. Sighthound is not specifically tailored to verification tasks, and the same model can be used with industry-leading performance on identification tasks.

The result is that Sighthound's software is highly accurate, robust to a variety of real world use cases and runs in real time.

To try Sighthound Facial Recognition, try the API on the Sighthound Cloud page or contact sales.

PubFig200 Face Recognition Accuracy

Method	Accuracy (%)
Sighthound [1]	99.2
DRM [2]	89.9
SSDML [3]	88.8
RNP [4]	88.6
MSSRC [5]	85.6
Face++ [6]	84.1
SANP [7]	80.4
OpenCV [8]	49.6

[1] Sighthound Inc. All PubFig200 images removed from model training

[2] Deep Reconstruction Models for Image Set Classification, Munawar Hayat, Mohammed Bennamoun and Senjian An, PAMI15

[3] P. Zhu, L. Zhang, W. Zuo, and D. Zhang, "From Point to Set: Extend the Learning of Distance Metrics", ICCV13

[4] M. Yang, P. Zhu, L. V. Gool, and L. Zhang, "Face Recognition Based on Regularized Nearest Points Between Image Sets", FG 2013

[5] E. Ortiz, A. Wright, and M. Shah, "Face Recognition in Movie Trailers via Mean Sequence Sparse Representation-Based Classification", CVPR13

[6] www.faceplusplus.com

[7] Y. Hu, A. S. Mian, and R. Owens, "Face Recognition Using Sparse Approximated Nearest Points Between Image Sets", PAMI12

[8] www.opencv.org

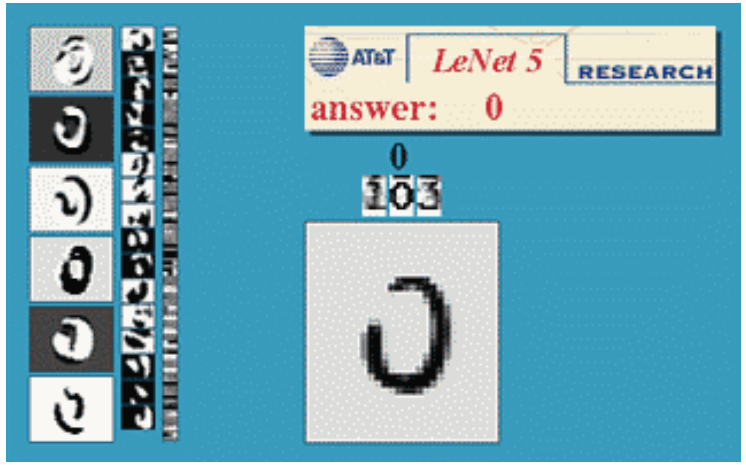
PubFig 200 is a data set of 58,797 images of 200 public figures. All academic papers cited with the presentation dates. All commercial systems tested in March 2016. Microsoft declined to make its system available for testing.

PubFig200 Face Recognition Accuracy

Method	Accuracy (%)
Sighthound [1]	99.2
DRM [2]	89.9
SSDML [3]	88.8
RNP [4]	88.6
MSSRC [5]	85.6
Face++ [6]	84.1
SANP [7]	80.4
OpenCV [8]	49.6

WHAT CAN CV DO?

OPTICAL CHARACTER RECOGNITION



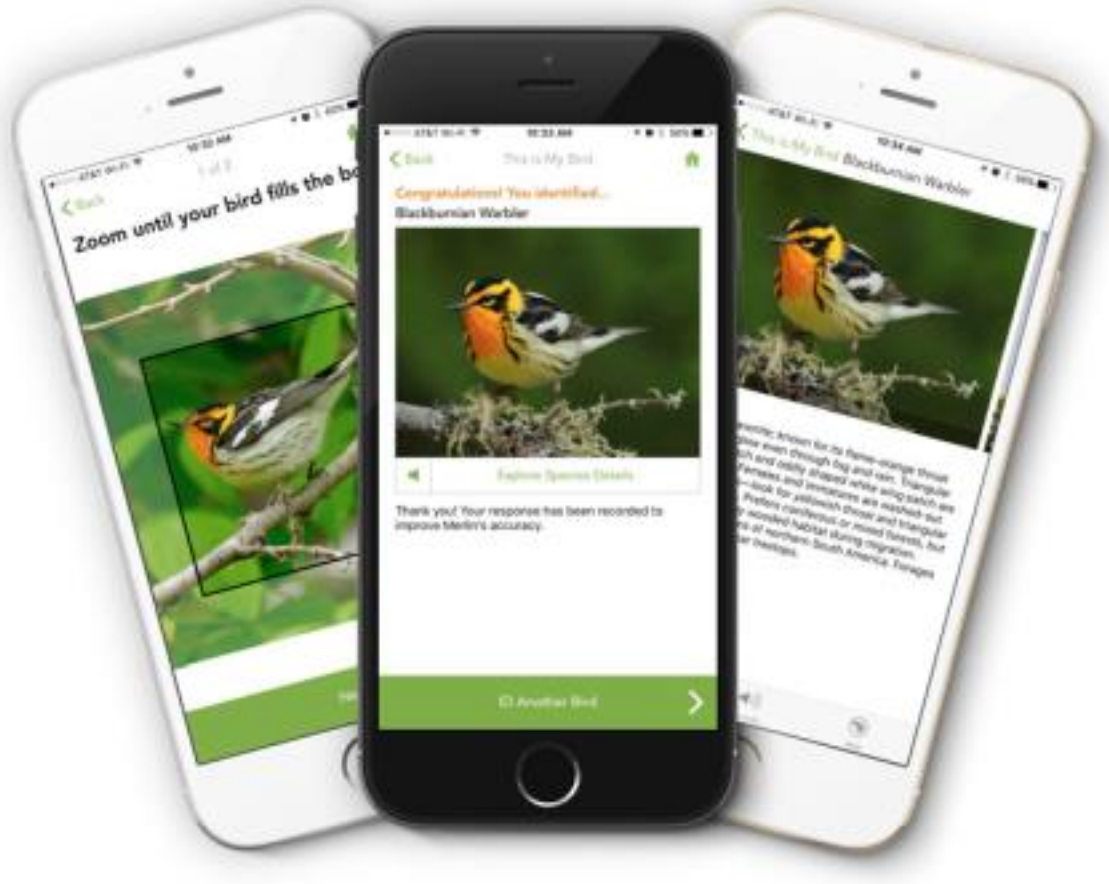
Yann LeCun's CNN applied to handwritten digit recognition
http://yann.lecun.com/exdb/lenet/

License plate readers

Automatic check processing

WHAT CAN CV DO?

NATURE IDENTIFICATION



Merlin Bird ID (based on Cornell Tech research)

WHAT CAN CV DO?

SPORTS

Sportvision first down line

Nice explanation on www.howstuffworks.com

WHAT CAN CV DO?

SMART VEHICLES & SELF-DRIVING CARS

Based on slides by Noah Snavely at Cornell: http://www.cs.cornell.edu/courses/cs5670/2018sp/lectures/lectures.html

WHAT CAN CV DO?

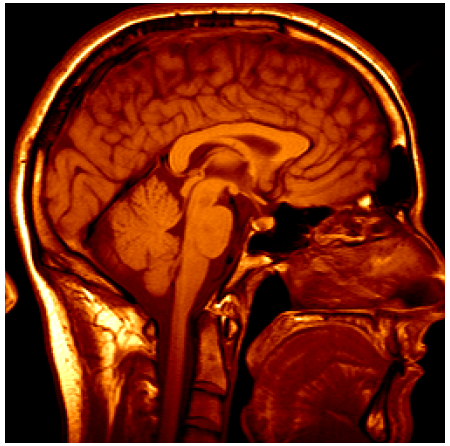
ROBOTICS

Amazon Picking Challenge
http://www.robocup2016.org/en/events/amazon-picking-challenge/

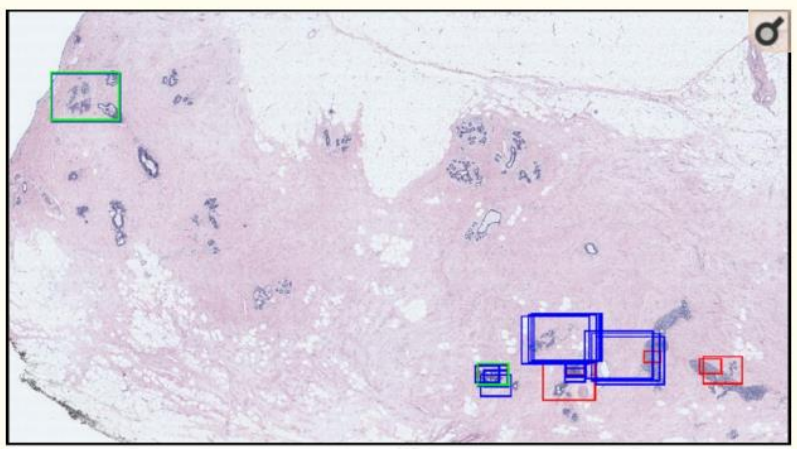
NASA's Mars Curiosity Rover
https://en.wikipedia.org/wiki/Curiosity_(rover)

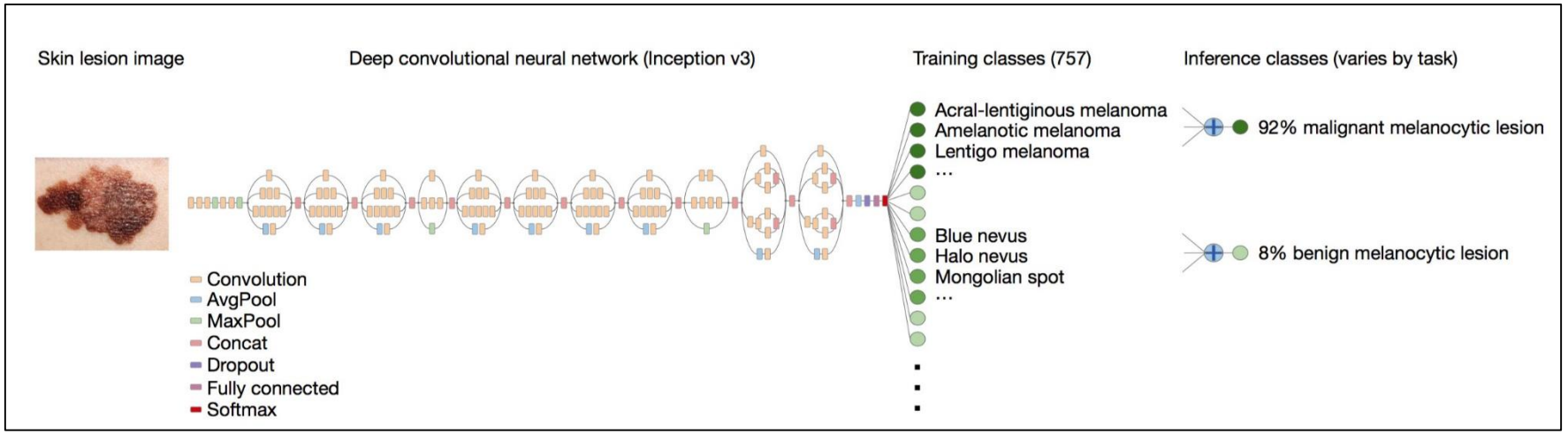
WHAT CAN CV DO?

MEDICAL IMAGING



3D imaging
(MRI, CT)

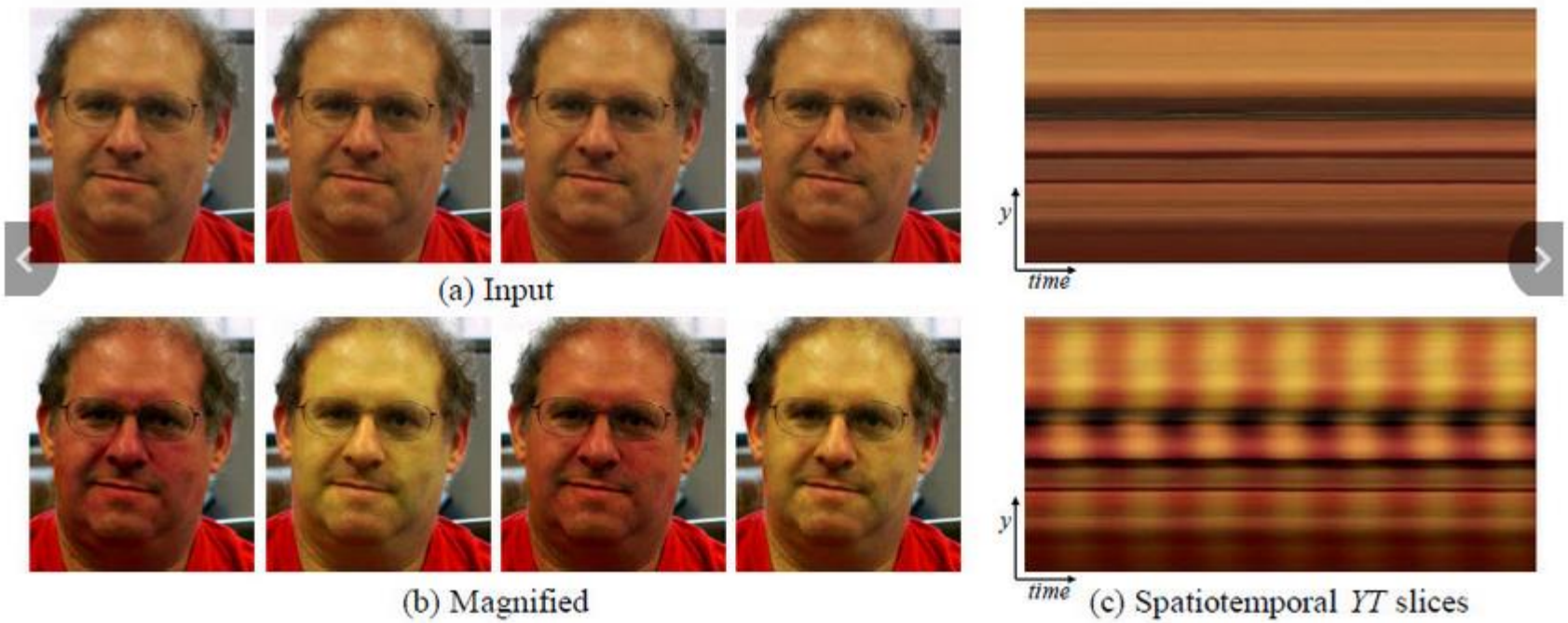




Skin cancer classification with deep learning
https://cs.stanford.edu/people/esteva/nature/

Video Magnification

Videos Software Publications Applications People Related Work Talks



An example of using our Eulerian Video Magnification framework for visualizing the human pulse. (a) Four frames from the original video sequence. (b) The same four frames with the subject's pulse signal amplified. (c) A vertical scan line from the input (top) and output (bottom) videos plotted over time shows how our method amplifies the periodic color variation. In the input sequence the signal is imperceptible, but in the magnified sequence the variation is clear.

[source: http://people.csail.mit.edu/mrub/vidmag/]

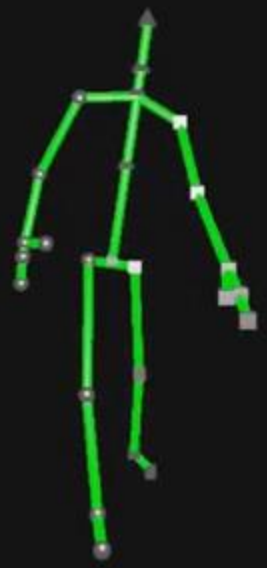
And, of course, HCI/Ubicomp

HCI + COMPUTER VISION

MICROSOFT KINECT

Source: Microsoft, https://youtu.be/ziXflemQr3A

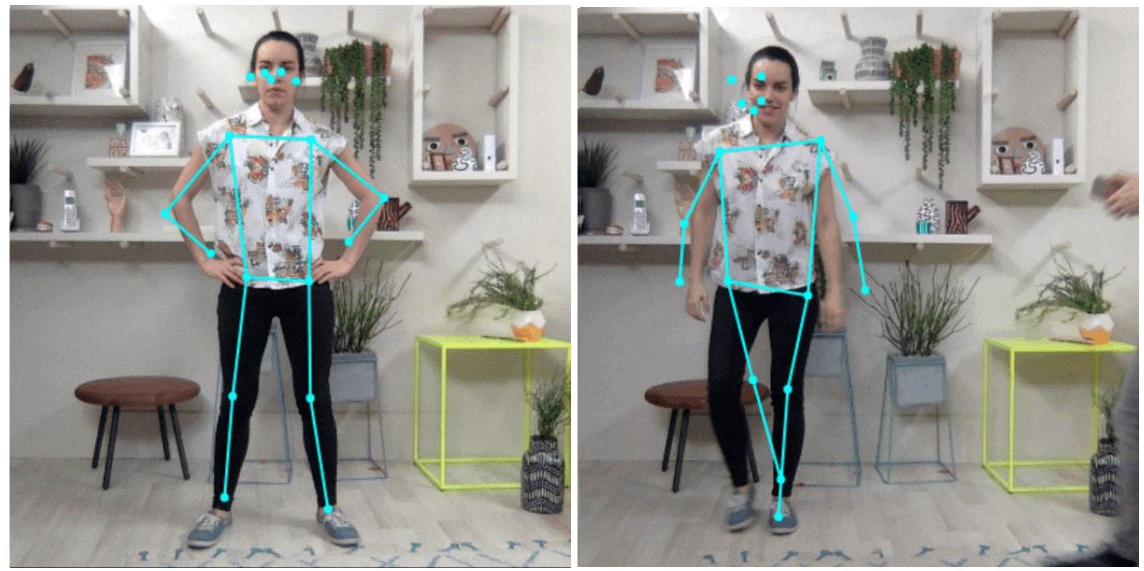
Skeleton



Real-time Human Pose Estimation in the Browser with TensorFlow.js

Posted by: Dan Oved, freelance creative technologist at Google Creative Lab, graduate student at ITP, NYU.
Editing and illustrations: Irene Alvarado, creative technologist and Alexis Gallo, freelance graphic designer, at Google Creative Lab

In collaboration with Google Creative Lab, I'm excited to announce the release of a TensorFlow.js version of PoseNet[1],[2] a machine learning model which allows for **real-time human pose estimation in the browser**. Try a live demo here.

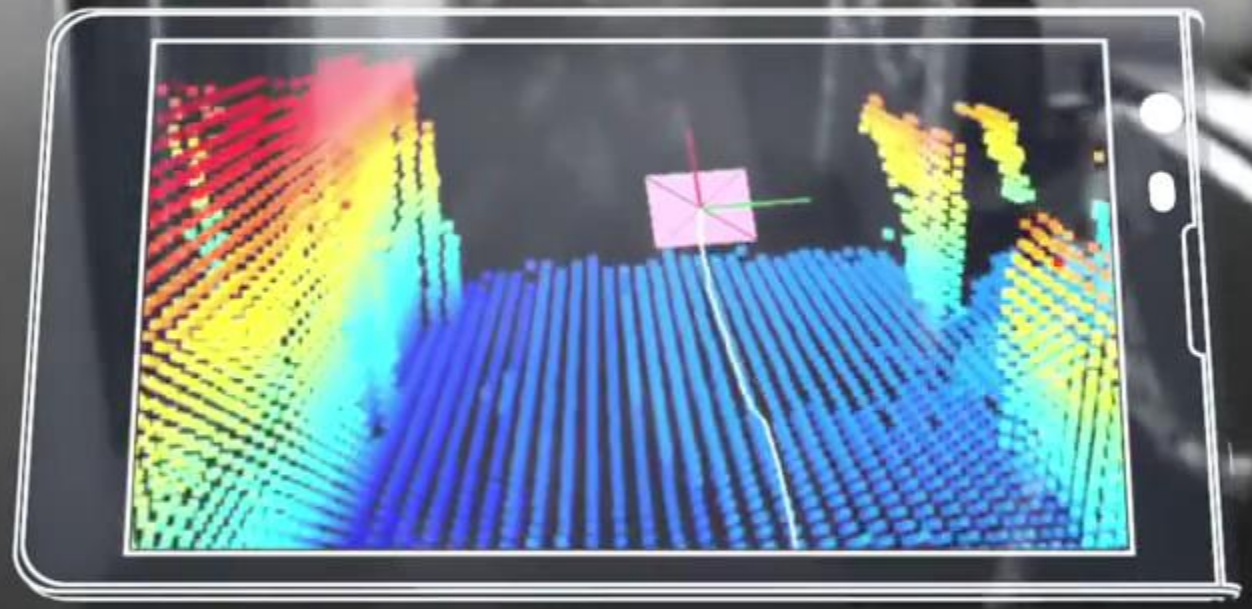


PoseNet can detect human figures in images and videos using either a single-pose or multi-pose algorithm—all from within the browser.

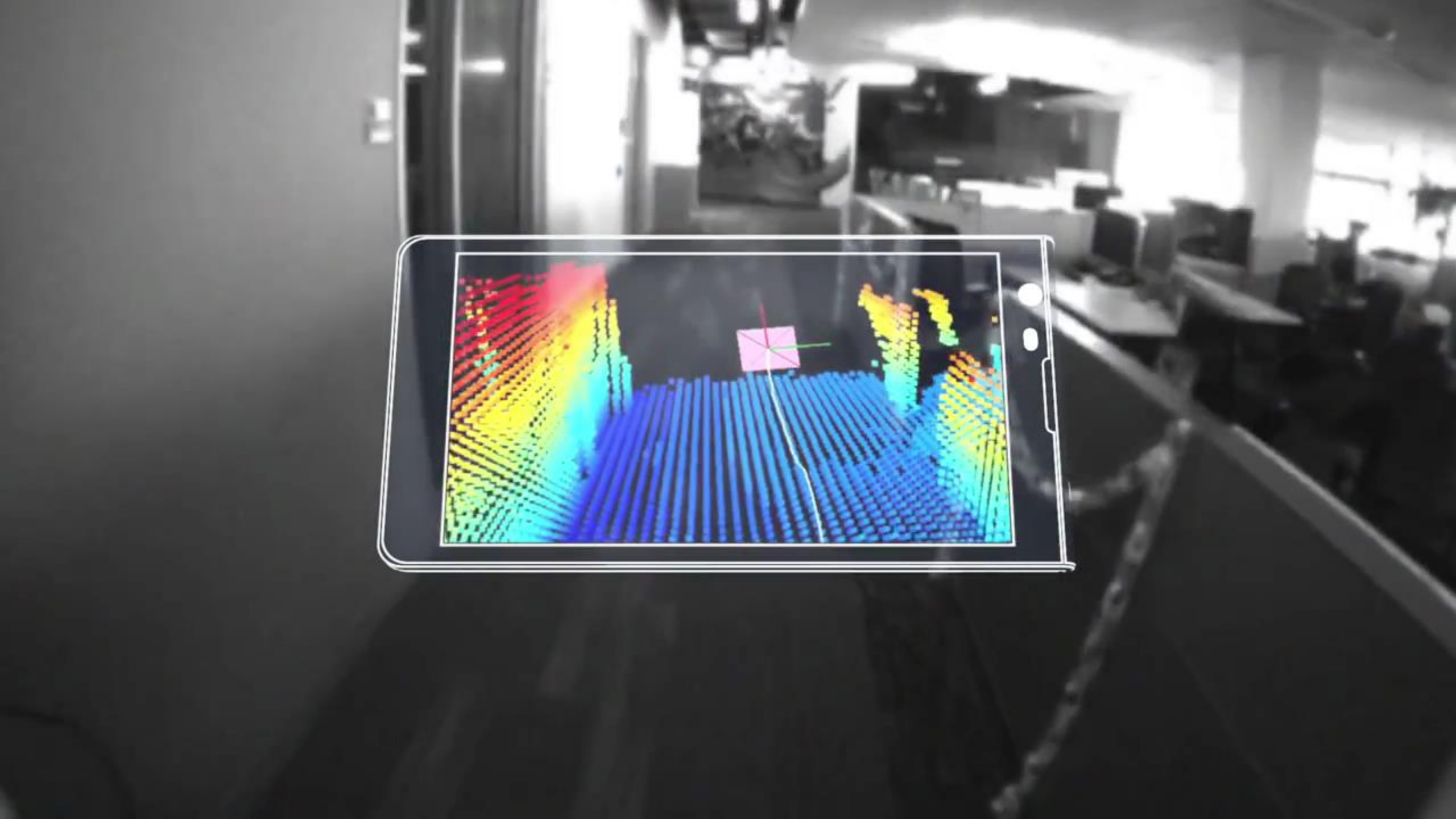
So what is pose estimation anyway? Pose estimation refers to computer vision techniques that detect human figures in images and video, so that one

HCI + COMPUTER VISION

GOOGLE'S PROJECT TANGO



Source: Google



HCI + COMPUTER VISION

INTERACTIVE PUPPET USING CV (2010)

Source: Theo Watson, https://vimeo.com/16985224

WHY IS CV HARD?

Viewpoint variation

Illumination

Intra-class variation

Occlusion

Scale

Motion (*e.g.,* Blur)

Background clutter

CV OVERVIEW

DETECTING CURB RAMPS IN GOOGLE STREET VIEW IMAGES

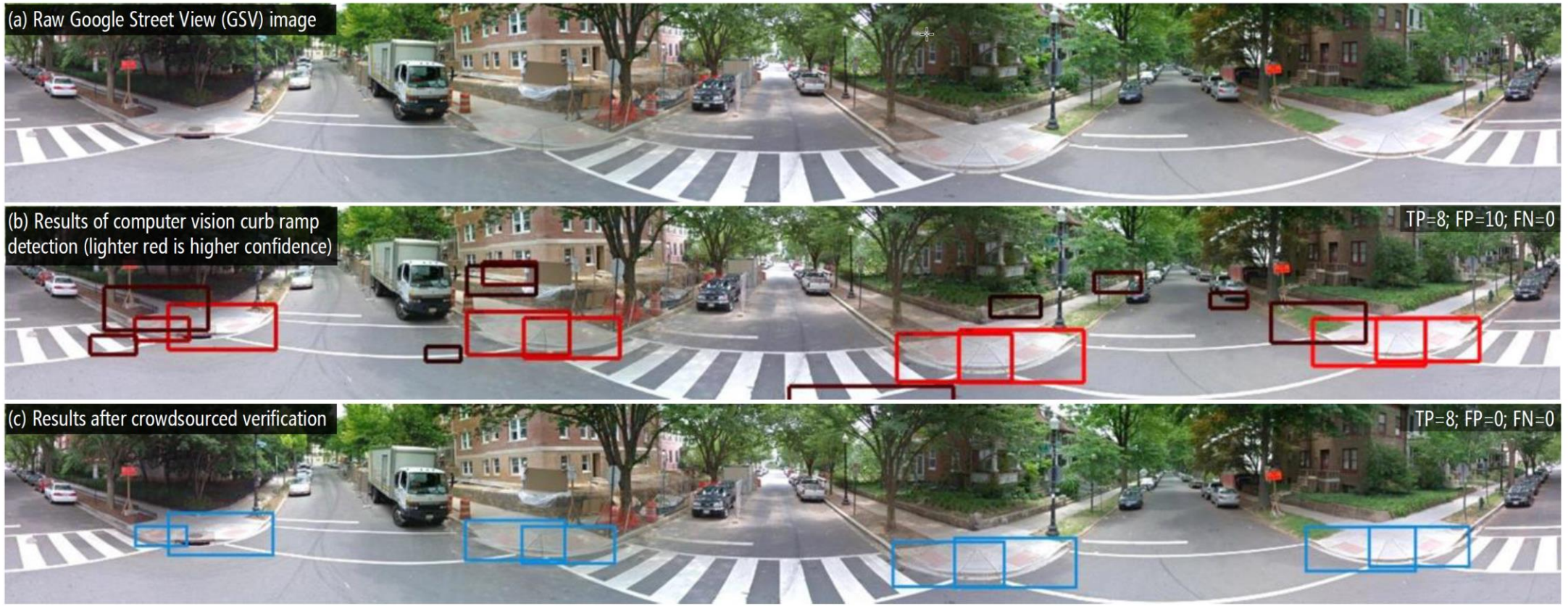


Figure 1: In this paper, we present *Tohme*, a scalable system for semi-automatically finding curb ramps in Google Streetview (GSV) panoramic imagery using computer vision, machine learning, and crowdsourcing. The images above show an actual result from our evaluation.

Source: Hara, Sun, Moore, Jacobs, & Froehlich, UIST2014

CV OVERVIEW

BUT THIS IS HARD BECAUSE...

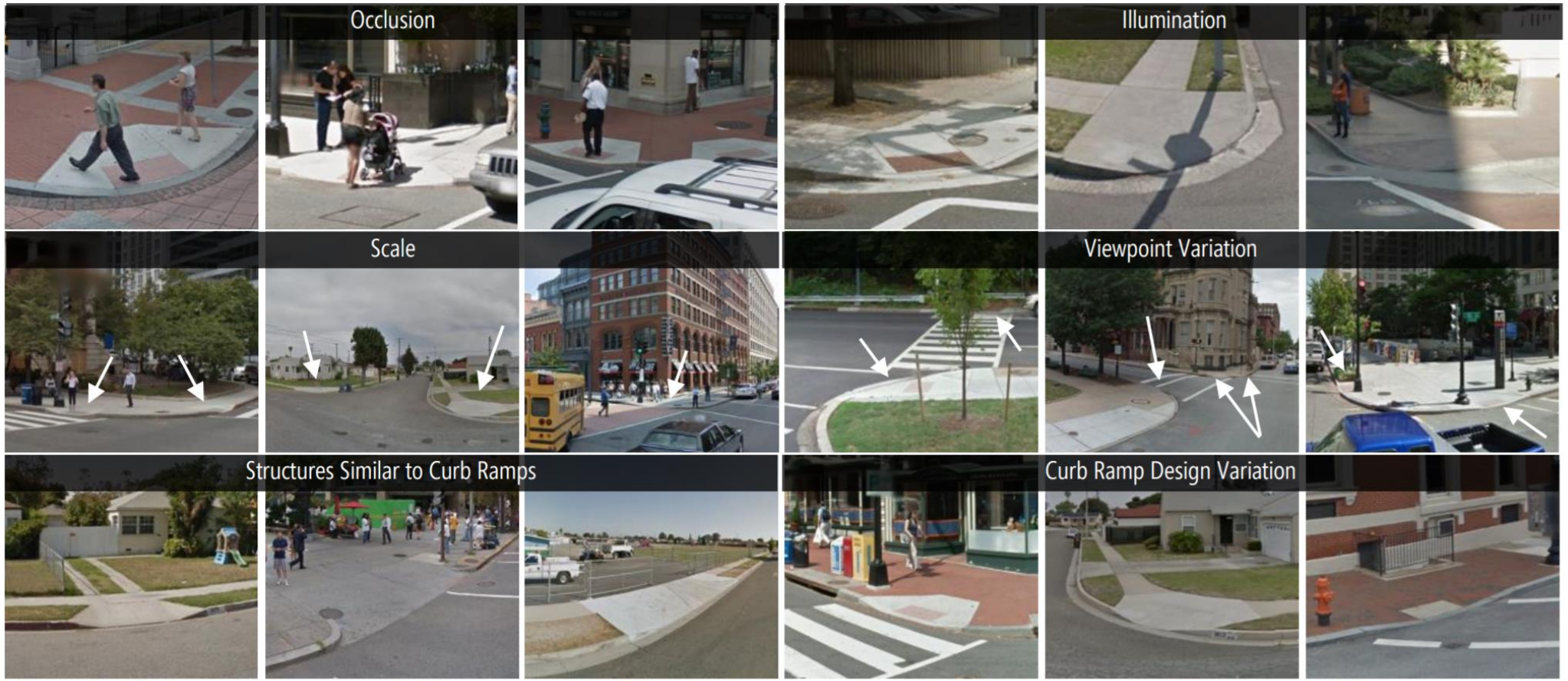


Figure 16: Finding curb ramps in GSV imagery can be difficult. Common problems include occlusion, illumination, scale differences because of distance, viewpoint variation (side, front, back), between class similarity, and within class variation. For between class similarity, many structures exist in the physical world that appear similar to curb ramps but are not. For within class variation, there are a wide variety of curb ramp designs that vary in appearance. White arrows are used in some images to draw attention to curb ramps. Some images contain multiple problems.

Source: Hara, Sun, Moore, Jacobs, & Froehlich, UIST2014

Some motivating examples for the A4 assignment.
Combining **computer vision** + **actuation**

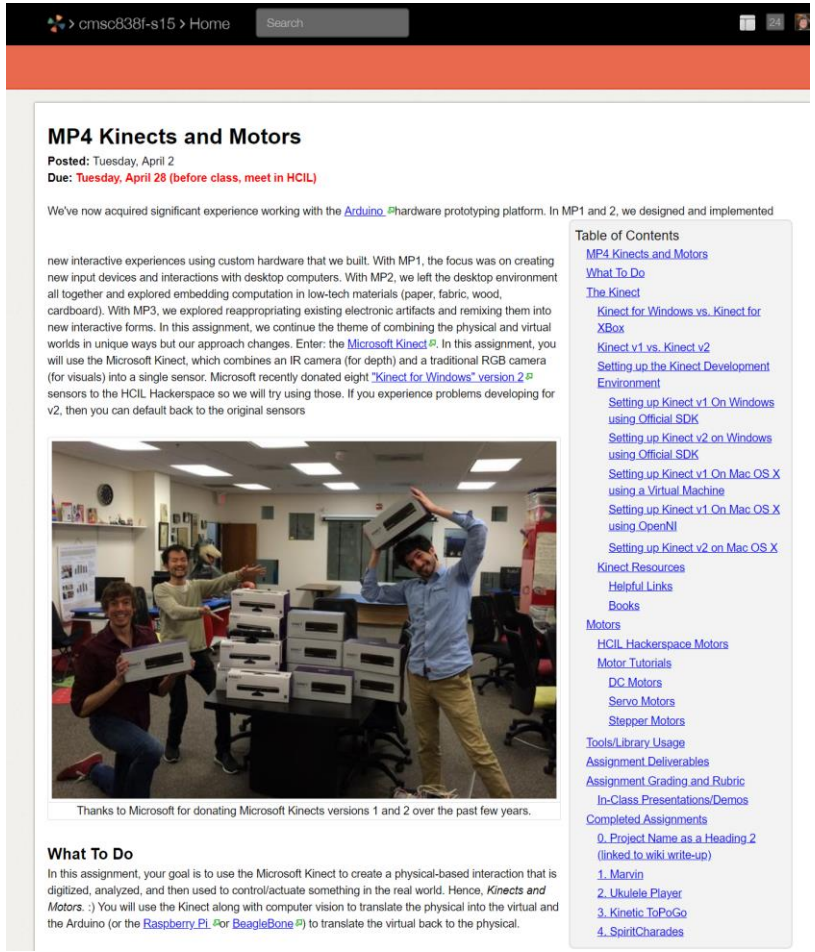
HCI + COMPUTER VISION

PINOKIO INTERACTIVE LAMP (2012)

Designers: Shanshan Zhou, Adam Ben-Dror, Joss Doggett, https://vimeo.com/52366512

HCI + COMPUTER VISION

SOME EXAMPLES FROM MY TANGIBLE INTERACTIVE COMPUTING COURSE AT UMD



> cmsc838f-s15 > Home

Search

MP4 Kinects and Motors

Posted: Tuesday, April 2
Due: Tuesday, April 28 (before class, meet in HCIL)

We've now acquired significant experience working with the Arduino hardware prototyping platform. In MP1 and 2, we designed and implemented new interactive experiences using custom hardware that we built. With MP1, the focus was on creating new input devices and interactions with desktop computers. With MP2, we left the desktop environment all together and explored embedding computation in low-tech materials (paper, fabric, wood, cardboard). With MP3, we explored reappropriating existing electronic artifacts and remixing them into new interactive forms. In this assignment, we continue the theme of combining the physical and virtual worlds in unique ways but our approach changes. Enter: the Microsoft Kinect. In this assignment, you will use the Microsoft Kinect, which combines an IR camera (for depth) and a traditional RGB camera (for visuals) into a single sensor. Microsoft recently donated eight "Kinect for Windows" version 2 sensors to the HCIL Hackerspace so we will try using those. If you experience problems developing for v2, then you can default back to the original sensors

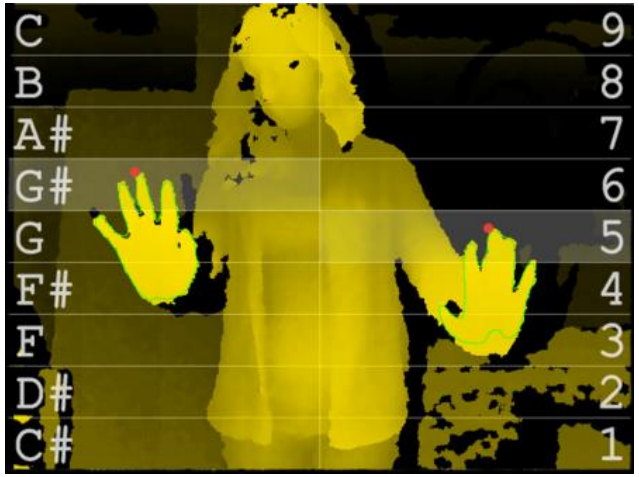
Table of Contents

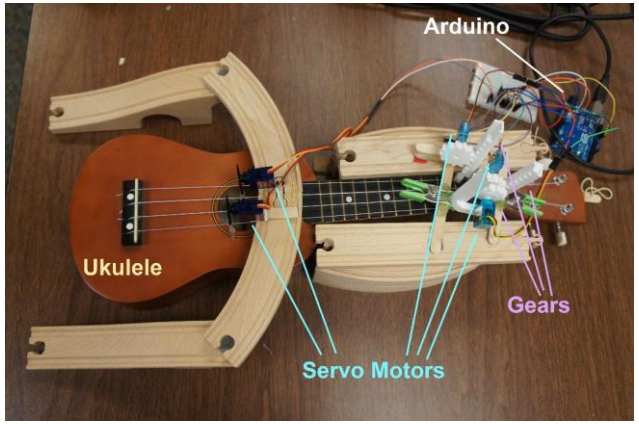
- MP4 Kinects and Motors
- What To Do
- The Kinect
  - Kinect for Windows vs. Kinect for XBox
  - Kinect v1 vs. Kinect v2
  - Setting up the Kinect Development Environment
    - Setting up Kinect v1 On Windows using Official SDK
    - Setting up Kinect v2 on Windows using Official SDK
    - Setting up Kinect v1 On Mac OS X using a Virtual Machine
    - Setting up Kinect v1 On Mac OS X using OpenNI
    - Setting up Kinect v2 on Mac OS X
  - Kinect Resources
    - Helpful Links
    - Books
- Motors
  - HCIL Hackerspace Motors
  - Motor Tutorials
    - DC Motors
    - Servo Motors
    - Stepper Motors
- Tools/Library Usage
- Assignment Deliverables
- Assignment Grading and Rubric
  - In-Class Presentations/Demos
- Completed Assignments
  - 0. Project Name as a Heading 2 (linked to wiki write-up)
  - 1. Marvin
  - 2. Ukulele Player
  - 3. Kinetic ToPoGo
  - 4. SpiritCharades

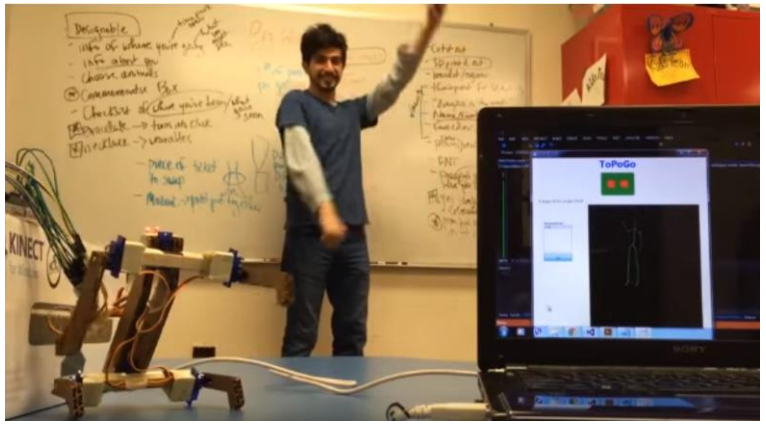
Thanks to Microsoft for donating Microsoft Kinects versions 1 and 2 over the past few years.

What To Do

In this assignment, your goal is to use the Microsoft Kinect to create a physical-based interaction that is digitized, analyzed, and then used to control/actuate something in the real world. Hence, *Kinects and Motors*. :) You will use the Kinect along with computer vision to translate the physical into the virtual and the Arduino (or the Raspberry Pi or BeagleBone) to translate the virtual back to the physical.







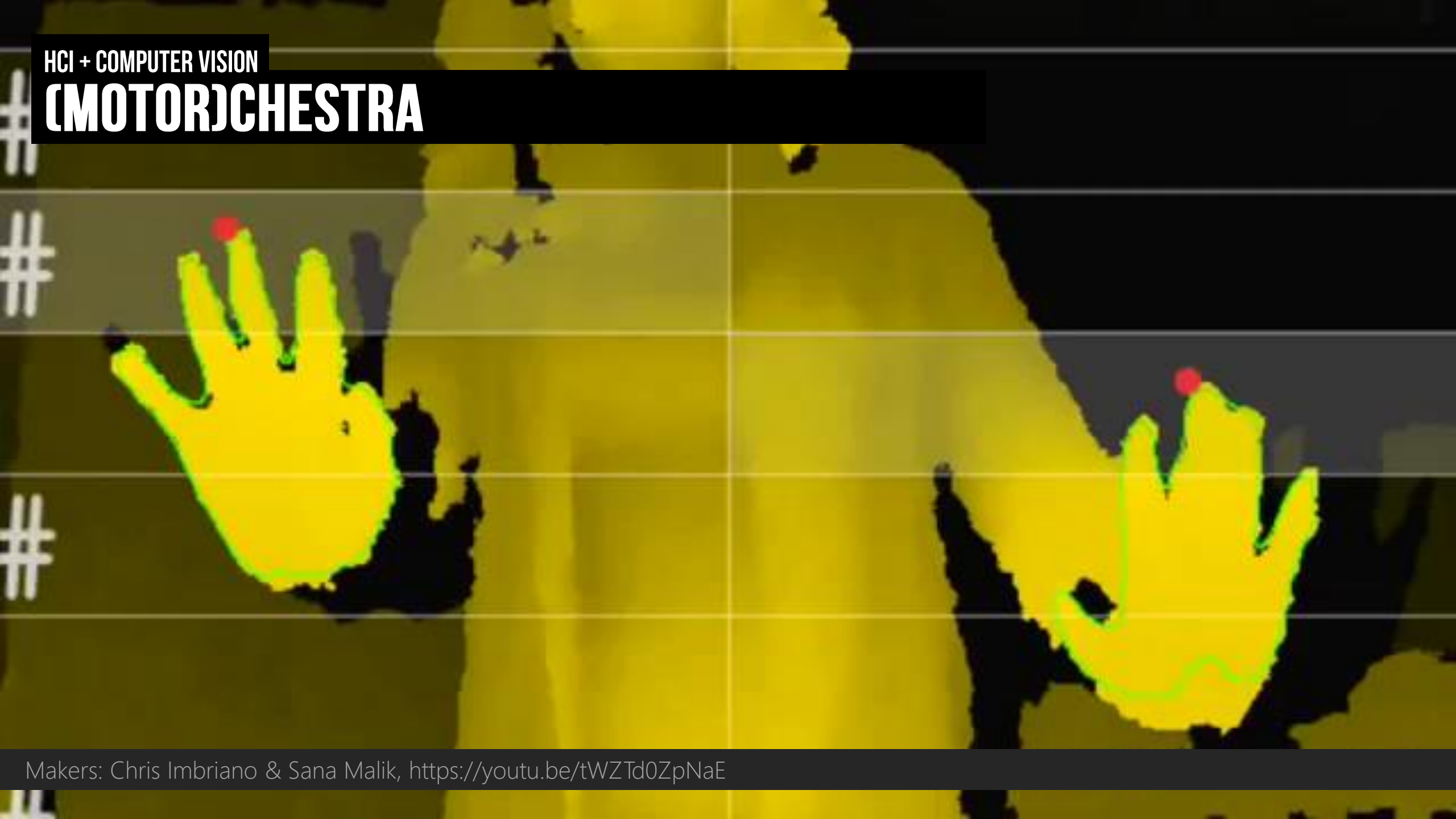
HCI + COMPUTER VISION

INTERACTIVE HELICOPTER CONTROL (2012)

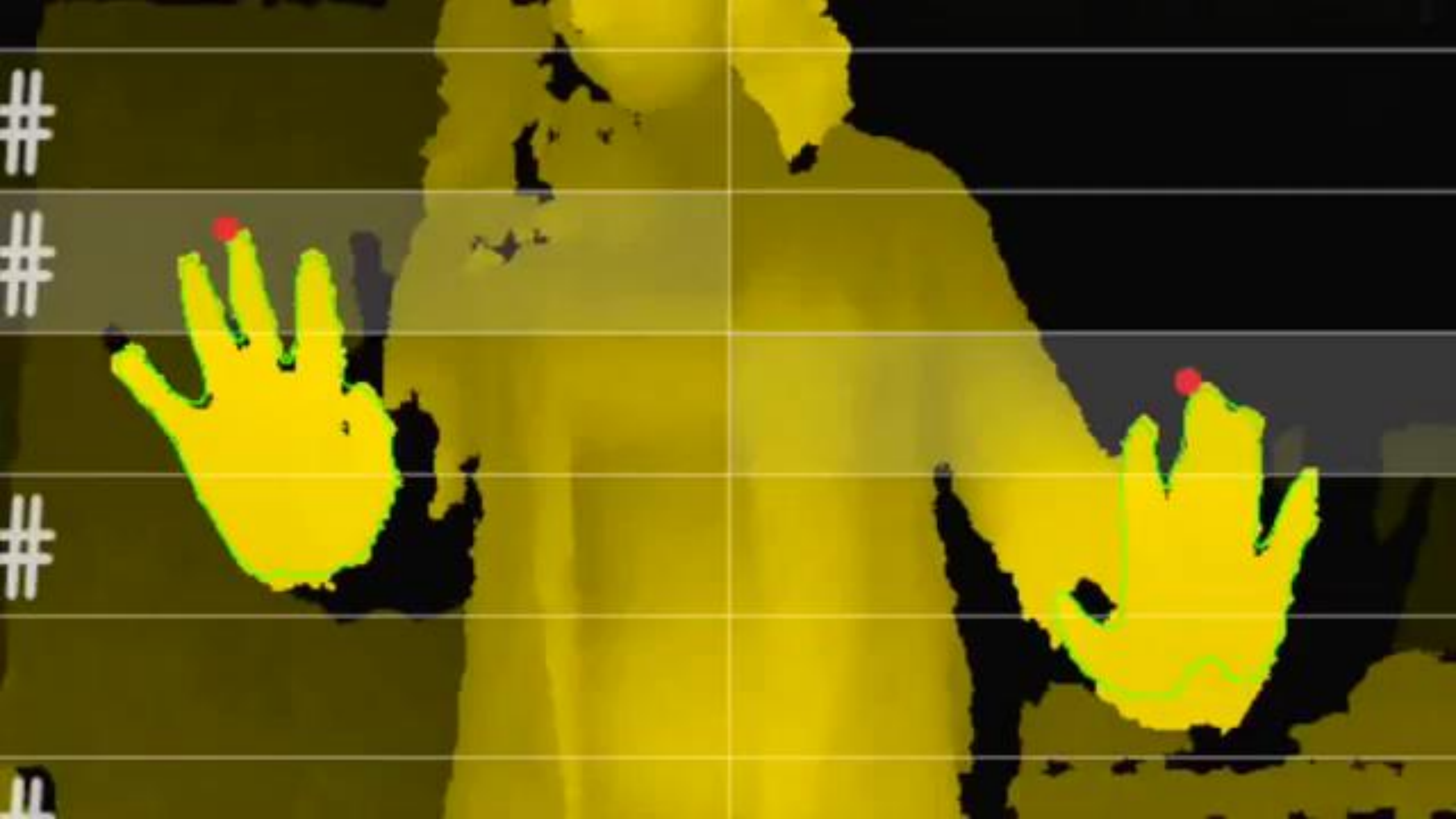
Makers: Leyla Norooz & Darren Smith, https://youtu.be/p9AV_Ws96Aw

HCI + COMPUTER VISION

(MOTOR)CHESTRA



Makers: Chris Imbriano & Sana Malik, https://youtu.be/tWZTd0ZpNaE



HCI + COMPUTER VISION

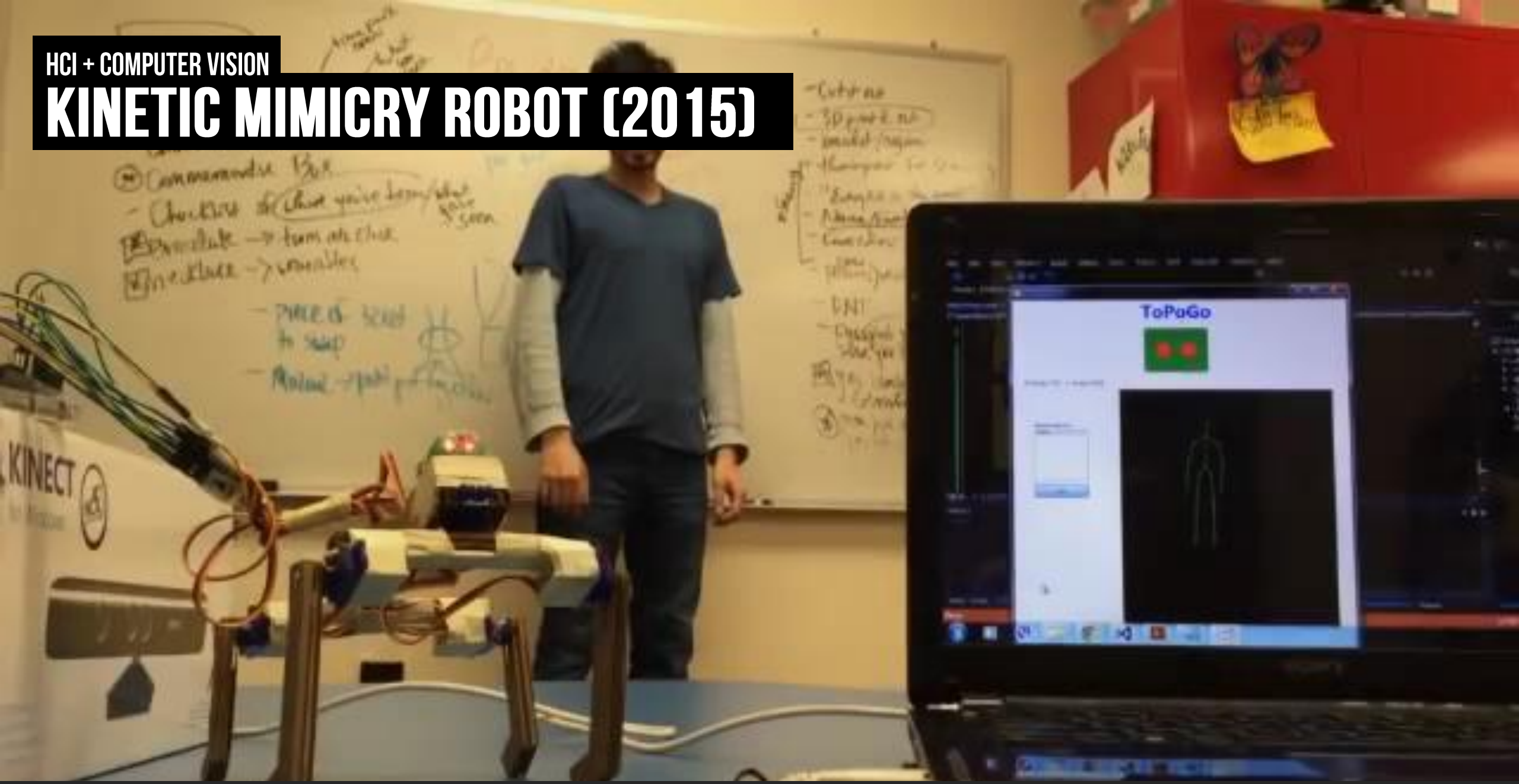
CUSTOM KINETIC ART SAND TABLE (2014)

Makers: Ruofei Du, Max Potasznik, & Kent Willis, https://youtu.be/LZZAKrGy9eM & http://www.instructables.com/id/AtmoSPHERE-XY-table-from-Knex-human-movement-visua/

3D-printed gears and tracks assist the precision-controlled XY table.

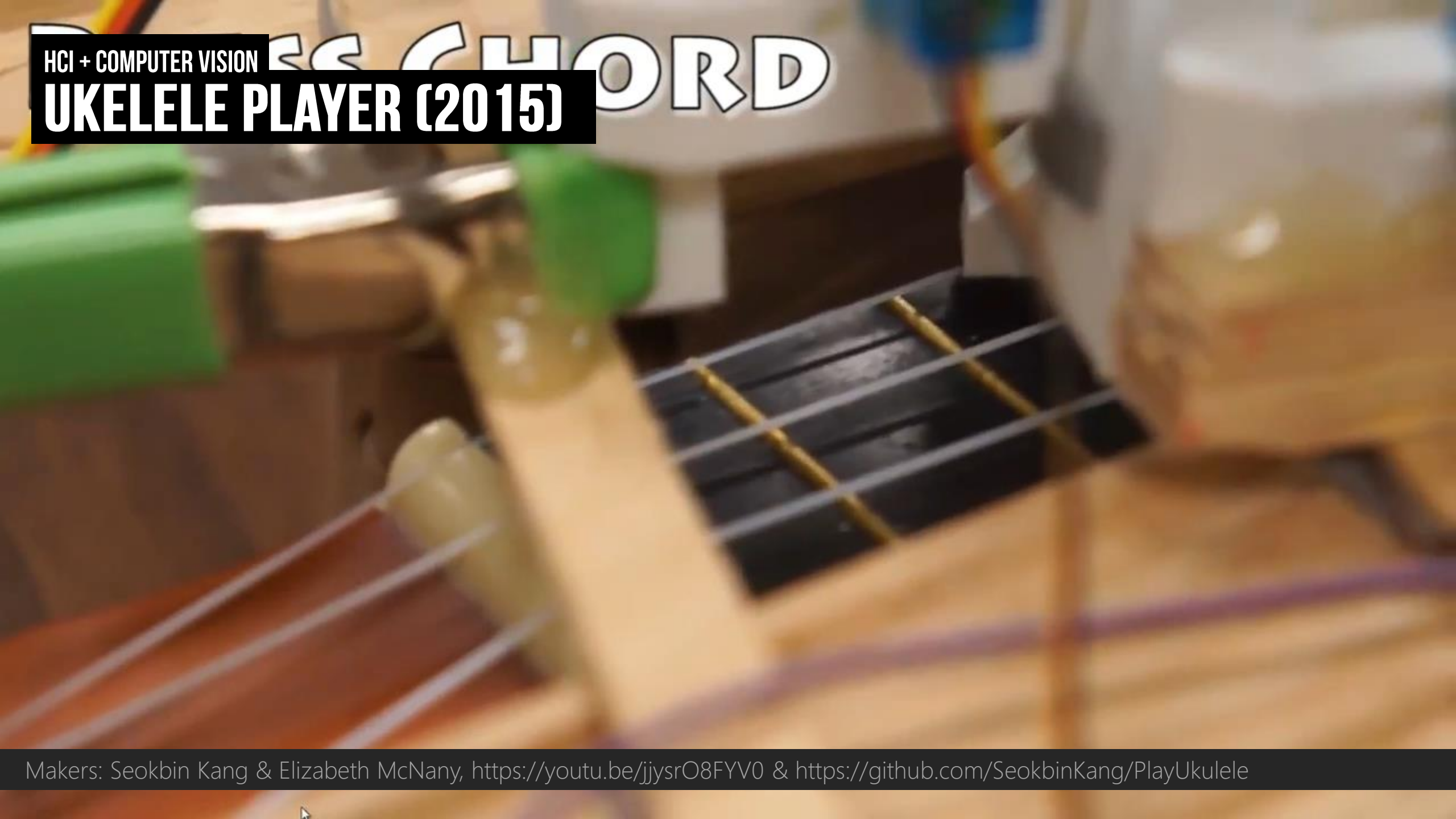
HCI + COMPUTER VISION

KINETIC MIMICRY ROBOT (2015)

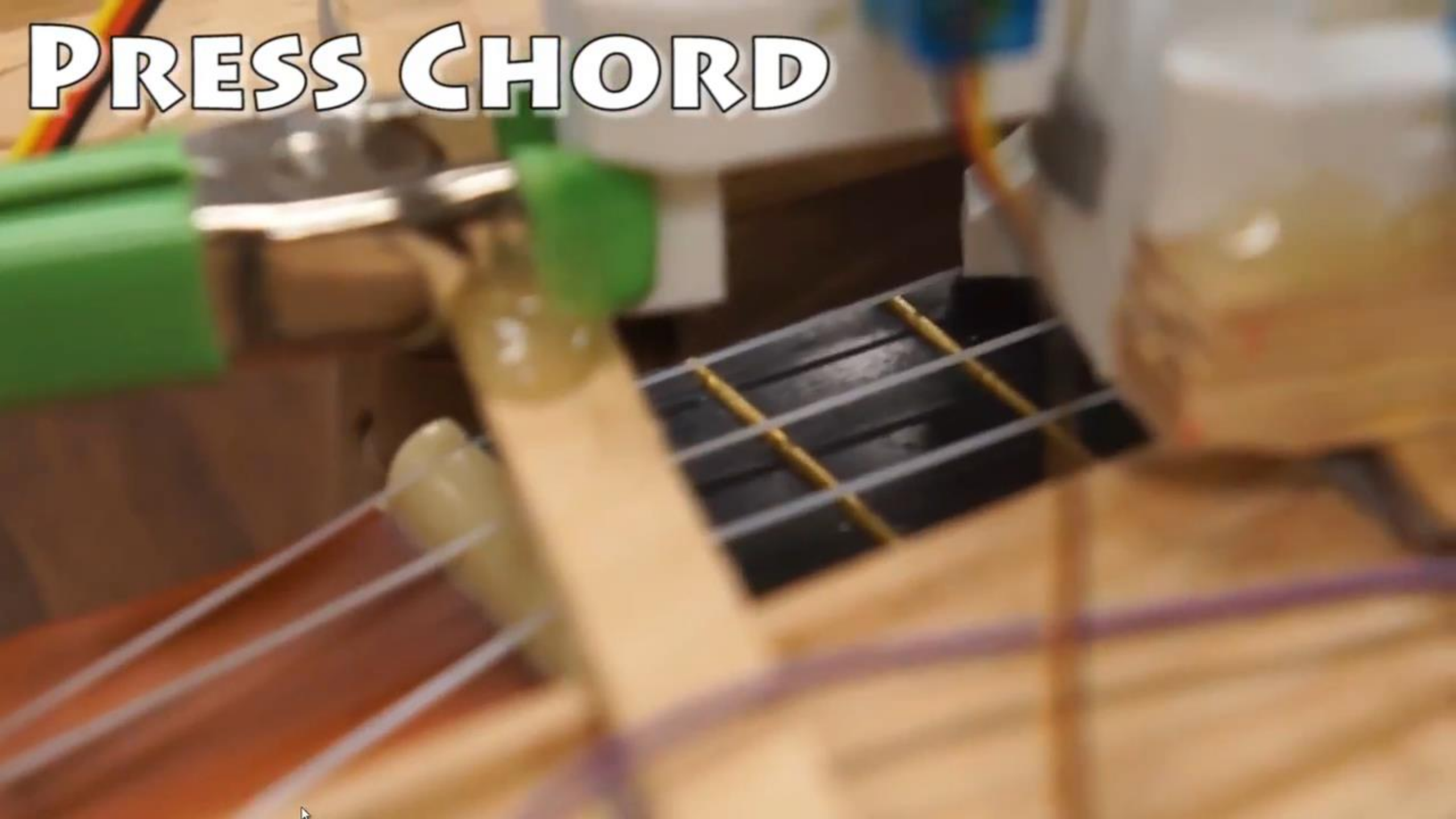


Makers: Majeed Kazemitabaar & Brian, https://youtu.be/luHxXKFdOlw & http://www.instructables.com/id/Kinect-Based-Mimicry-Robot/


ToPoGo



Makers: Seokbin Kang & Elizabeth McNany, https://youtu.be/jjysrO8FYV0 & https://github.com/SeokbinKang/PlayUkulele


PRESS CHORD

Some inspirational **digital interactive art** using vision + actuation

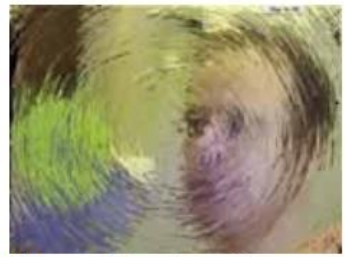
Mechanical Mirrors

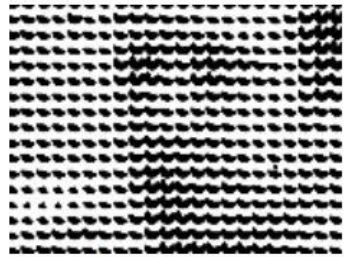

Wooden Mirror


Trash Mirror


Shiny Balls Mirror

Software Mirrors


Mirror Number 2


mirror5.html


Mirror Number 6

Video Painting


Easel

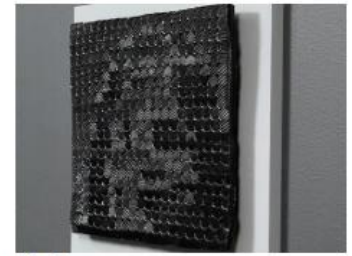

Paint-Cam


Origin Of Species

Glass Sculptures


Broken-Mirror


Self Centered Mirror


Yulie

Proxxi Prints


Yves-Merilyn/Skyline


Yonatan/Subway Map


Sharon / West Bank

HCI + COMPUTER VISION

INTERACTIVE WOODEN MIRROR (1990)

Designer: Daniel Rozin, https://youtu.be/BZysu9QcceM

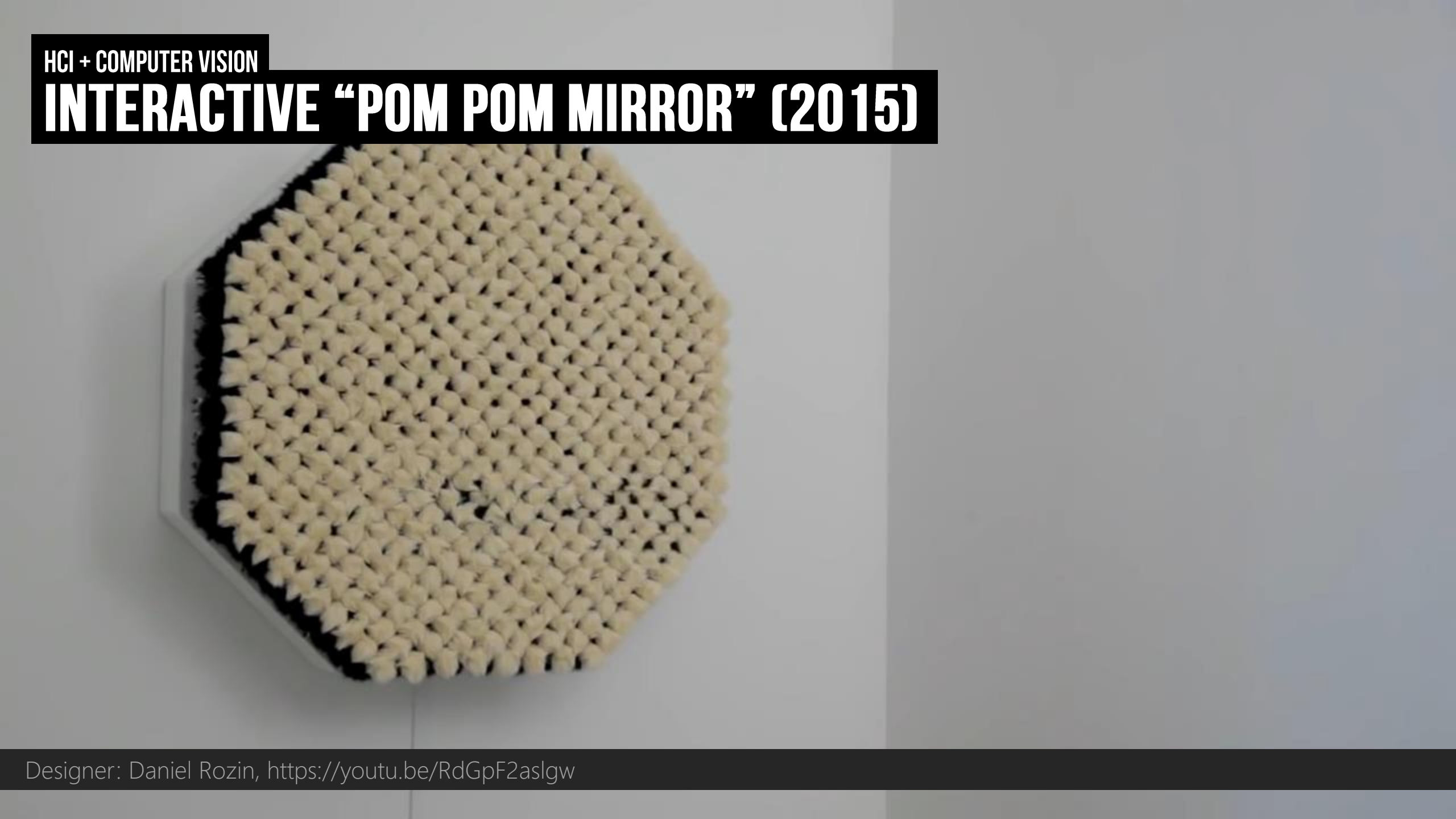
HCI + COMPUTER VISION

INTERACTIVE “WEAVE MIRROR” (2007)

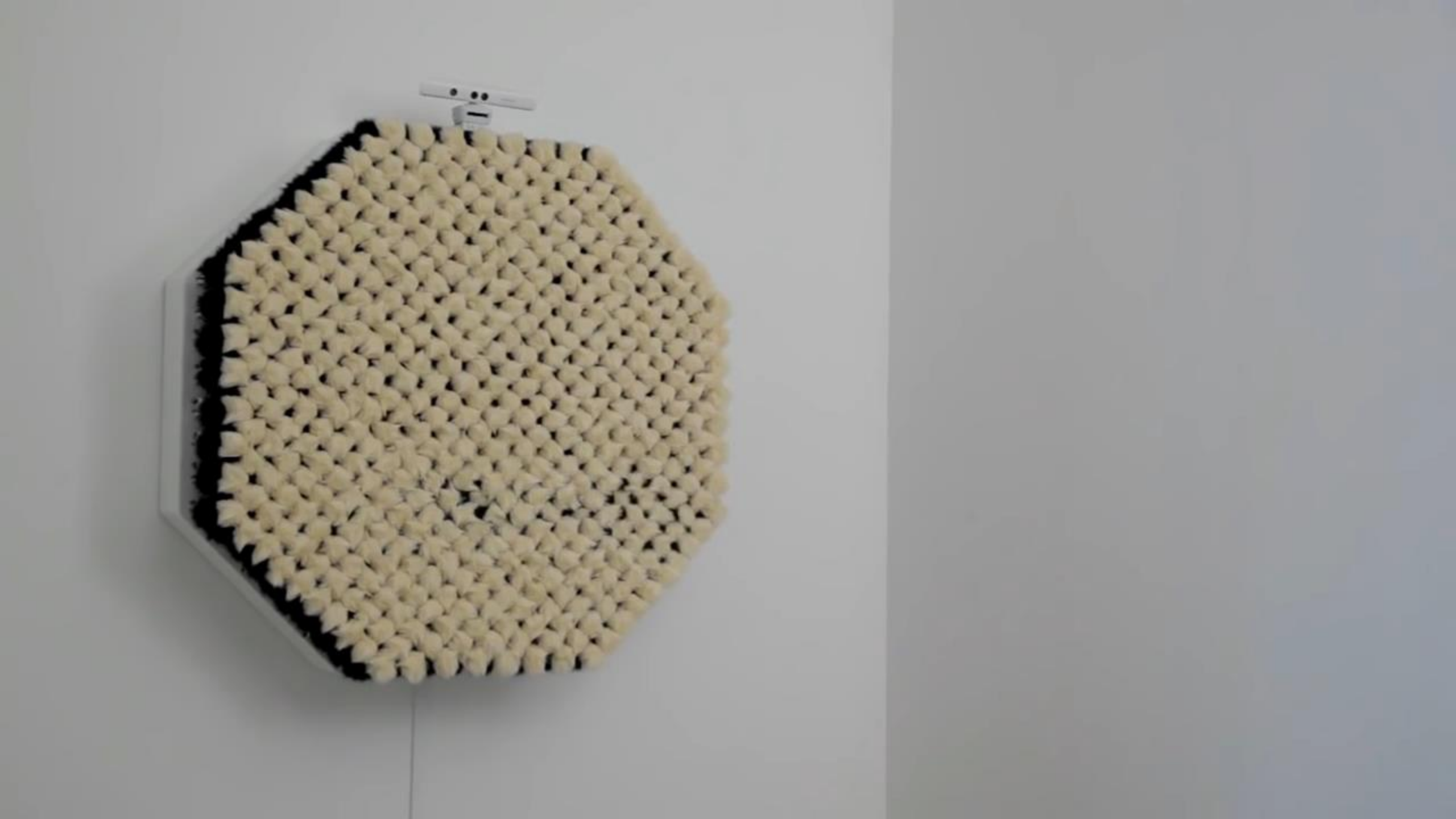
Designer: Daniel Rozin, https://vimeo.com/7067089

HCI + COMPUTER VISION

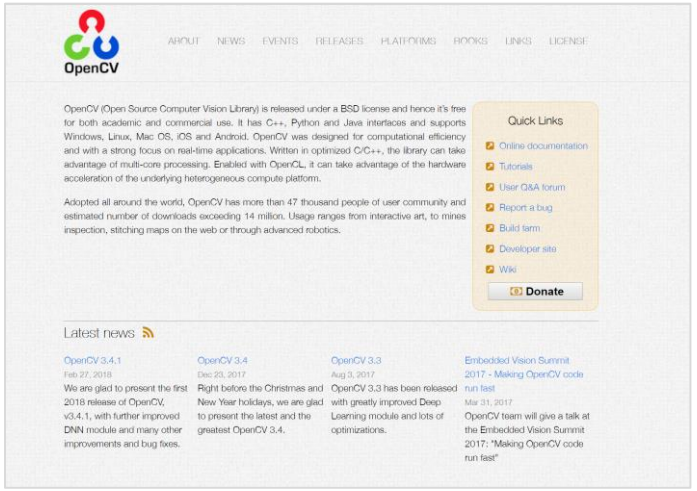
INTERACTIVE “POM POM MIRROR” (2015)



Designer: Daniel Rozin, https://youtu.be/RdGpF2aslgw

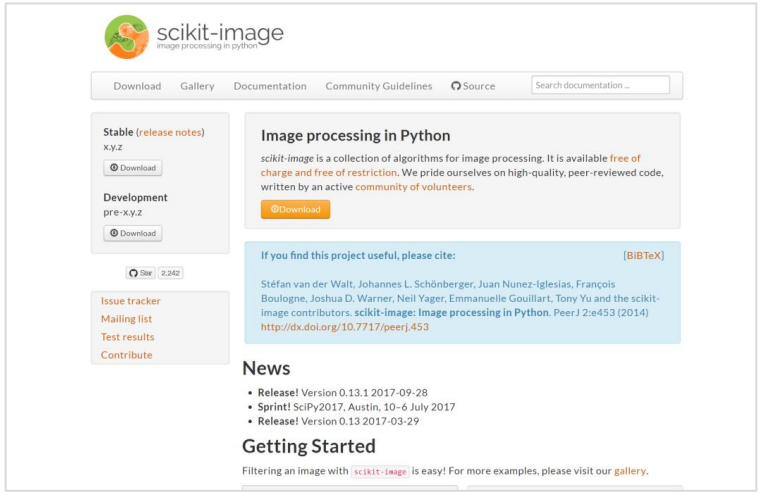


COMMON CV TOOLKITS



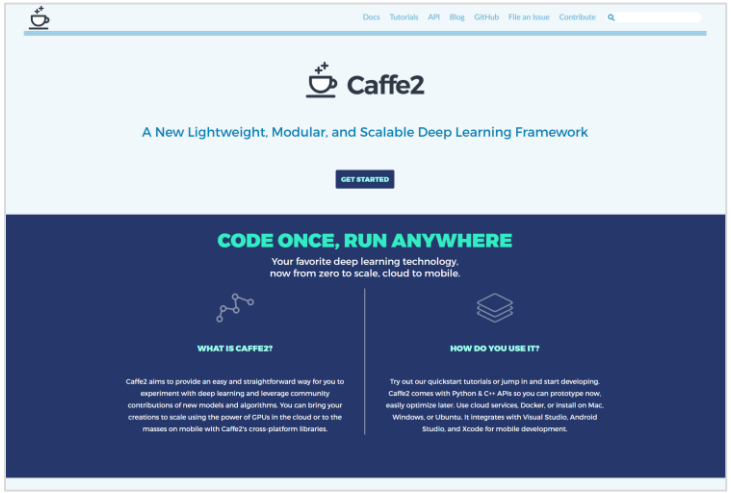
OpenCV

The most popular CV toolkit in history with an estimated 14 million downloads. It is the quintessential CV library; however, it may be losing its reign to cloud-based APIs.



Scikit-Image

Part of the scikit library. Python-based library for image processing including feature and object detection, filtering and restoration, object segmentation, etc.

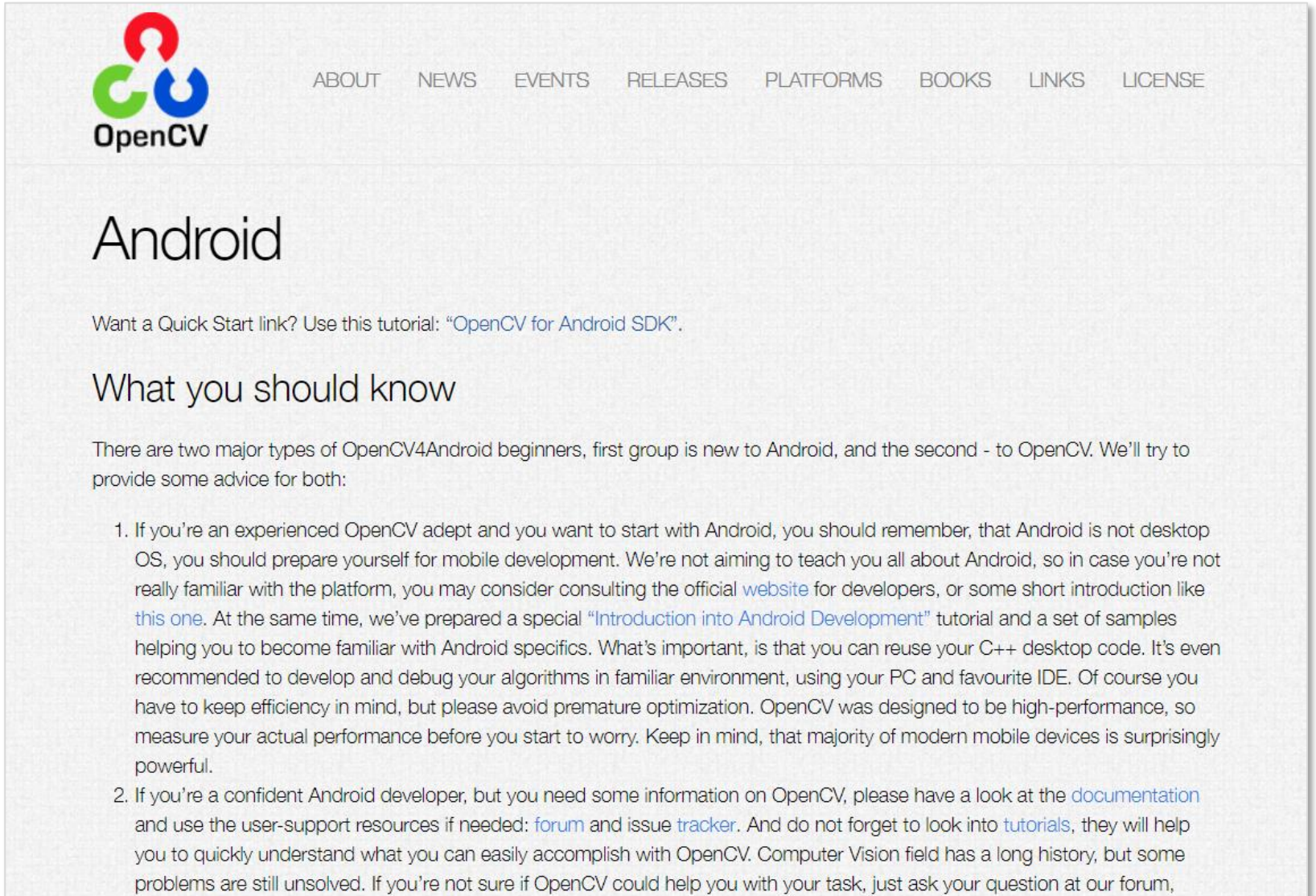


Caffe2

Scalable deep learning framework from Facebook

CV TOOLKITS

WE WERE ORIGINALLY GOING TO USE OPENCV!



OpenCV

ABOUT NEWS EVENTS RELEASES PLATFORMS BOOKS LINKS LICENSE

Android

Want a Quick Start link? Use this tutorial: "OpenCV for Android SDK".

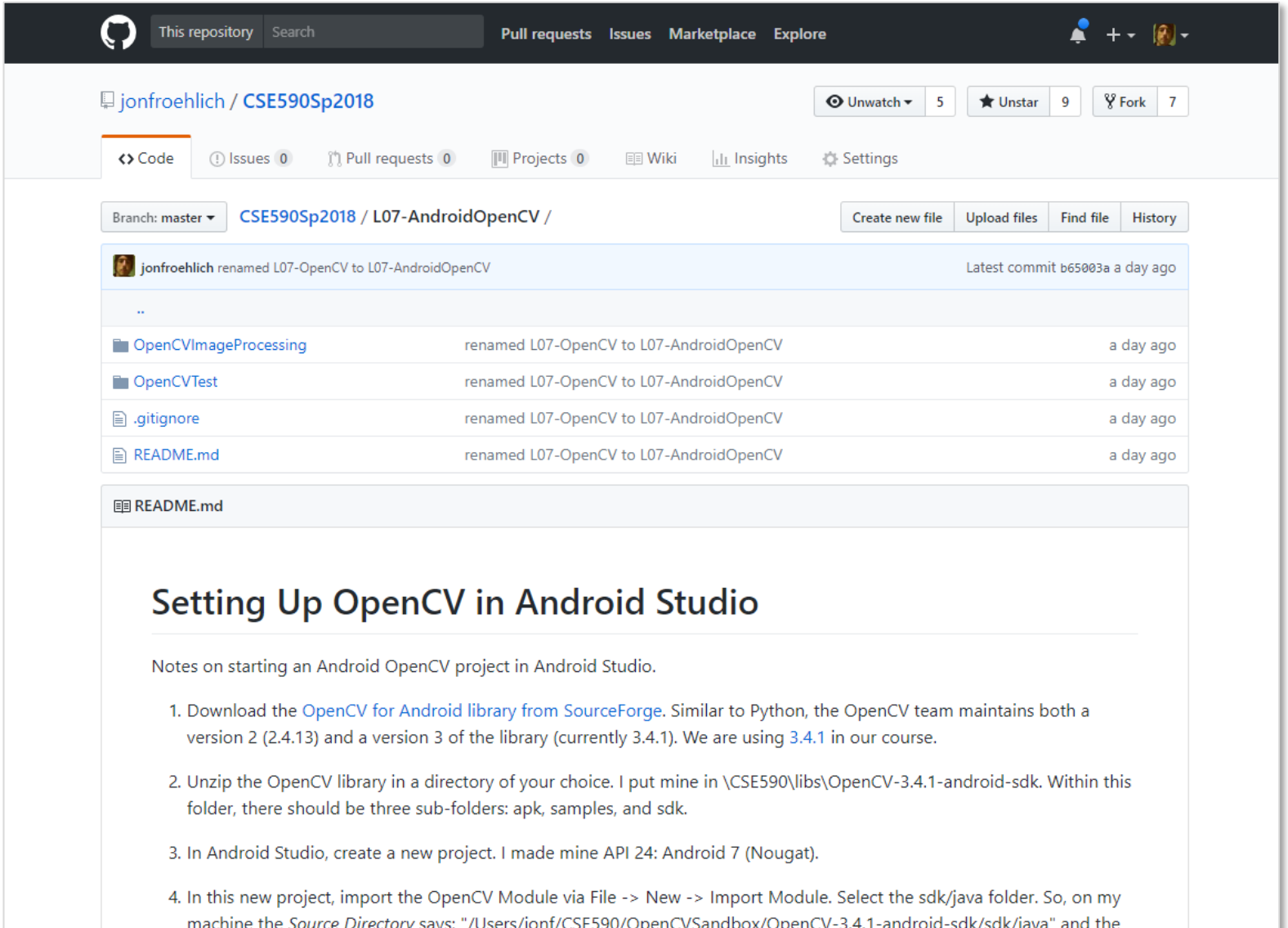
What you should know

There are two major types of OpenCV4Android beginners, first group is new to Android, and the second - to OpenCV. We'll try to provide some advice for both:

1. If you're an experienced OpenCV adept and you want to start with Android, you should remember, that Android is not desktop OS, you should prepare yourself for mobile development. We're not aiming to teach you all about Android, so in case you're not really familiar with the platform, you may consider consulting the official website for developers, or some short introduction like this one. At the same time, we've prepared a special "Introduction into Android Development" tutorial and a set of samples helping you to become familiar with Android specifics. What's important, is that you can reuse your C++ desktop code. It's even recommended to develop and debug your algorithms in familiar environment, using your PC and favourite IDE. Of course you have to keep efficiency in mind, but please avoid premature optimization. OpenCV was designed to be high-performance, so measure your actual performance before you start to worry. Keep in mind, that majority of modern mobile devices is surprisingly powerful.
2. If you're a confident Android developer, but you need some information on OpenCV, please have a look at the documentation and use the user-support resources if needed: forum and issue tracker. And do not forget to look into tutorials, they will help you to quickly understand what you can easily accomplish with OpenCV. Computer Vision field has a long history, but some problems are still unsolved. If you're not sure if OpenCV could help you with your task, just ask your question at our forum,

CV TOOLKITS

INDEED, I BUILT A FEW EXAMPLES FOR YOU!



BUT IT'S NOT A GREAT LEARNING PLATFORM

At least not on Android

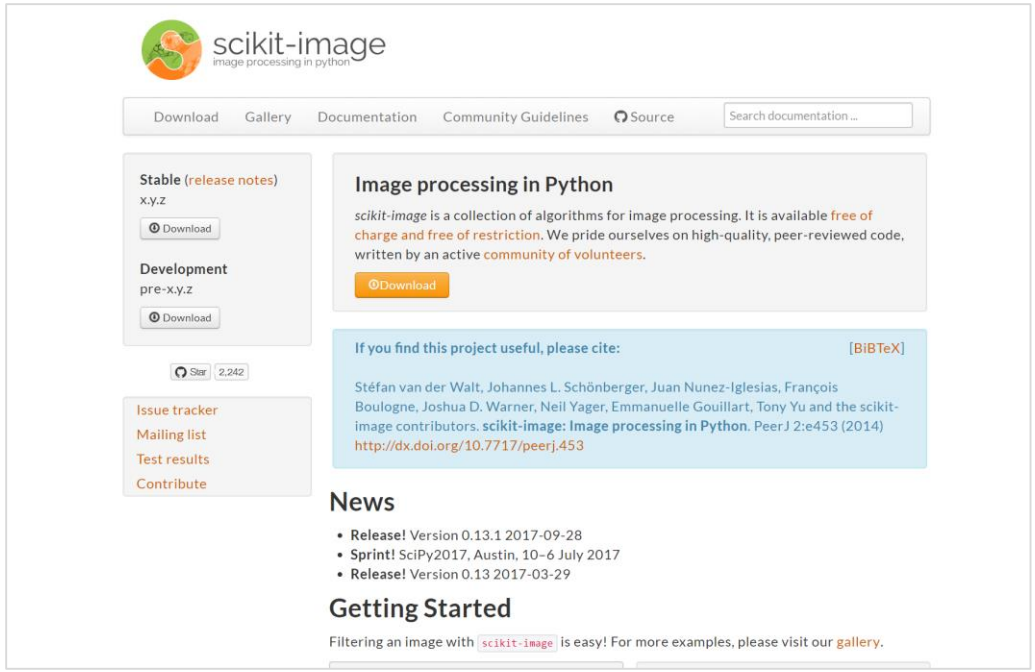
Setup is difficult

Compile + try cycle is elongated and arduous

Android is not a high priority platform for open source team

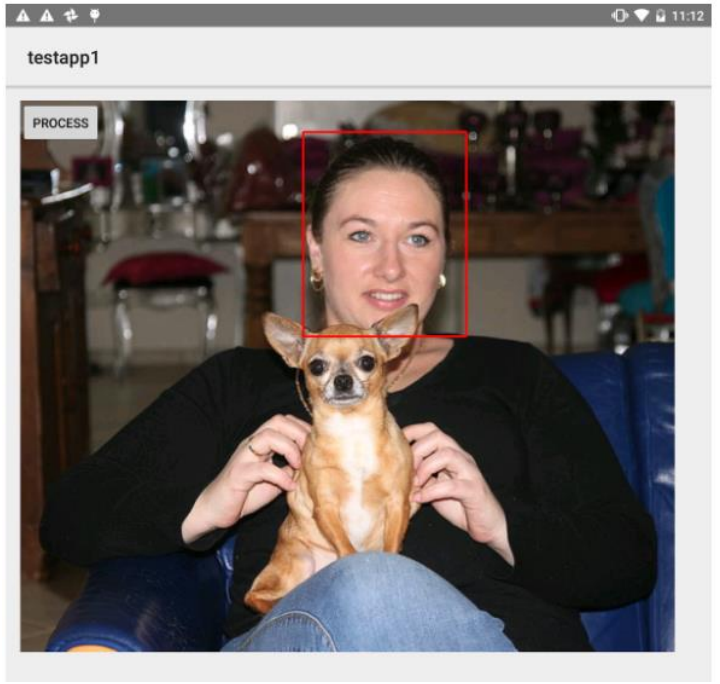
CV TOOLKITS

SO WE ARE GOING TO USE...



For our in-class exercises next week: Scikit-Image

Part of the scikit library. Python-based library for image processing including feature and object detection, filtering and restoration, object segmentation, etc.



For A4 and today's exercises: Google's Mobile Vision API

An incredibly fast and well-designed API for Android-based computer vision.

SCHEDULE TODAY: 6:30-9:20

06:30-06:50: Discussion of required reading led by Pankaj Parag

06:50-07:10: A3 Nite Lite “Show-and-Tell”

07:10-07:15: Overview of A4 “Smart Spaces” assignment

07:15-08:10: Applied Computer Vision in UbiComp/HCI

08:10-08:15: Break

08:15-09:20: FaceTrackerBLE Exercises