OrCam MyMe

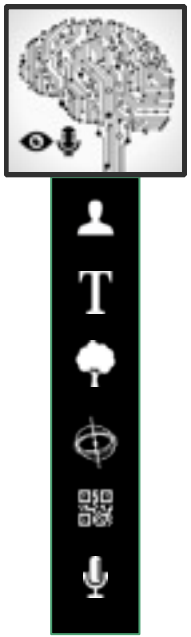
Digital Personal Assistance For Real Life

Amnon Shashua

ORCAM

What If Siri/Cortana/Now Had Eyes and Ears?

Understanding The World Around Us



Eating and fitness habits

How you spend your time

Habits: Books, TV, Phone etc'

people you spend time with

Emotion analysis of people you encounter

Conversation topics

Different Worlds

Wearable

- Computational Lean
- Restricted Battery
- Constantly Working

AI / Deep Learning

- CPU Intensive
- Power intensive
- Process On Demand

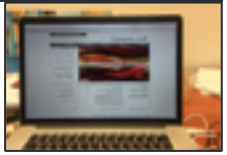
Mature Applications in Computer Vision and NLP

Image Understanding

Car, Travel, Traffic, Road

People, Crowd, Party

Computer, Screen, Desk



Surf, water, ocean, wave

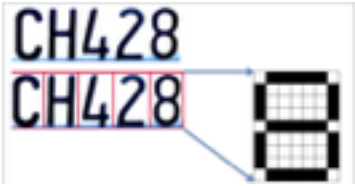
Face Recognition

Speech to Text

Emotion Recognition

Product Recognition

OCR



Introducing OrCam MyMe

The First Wearable AI

Key advantages of using wearable AI as a platform for Digital Personal Assistant

Augmented Attention

Context based operations: one trigger- different actions

Always On, Constantly Attentive

Technology

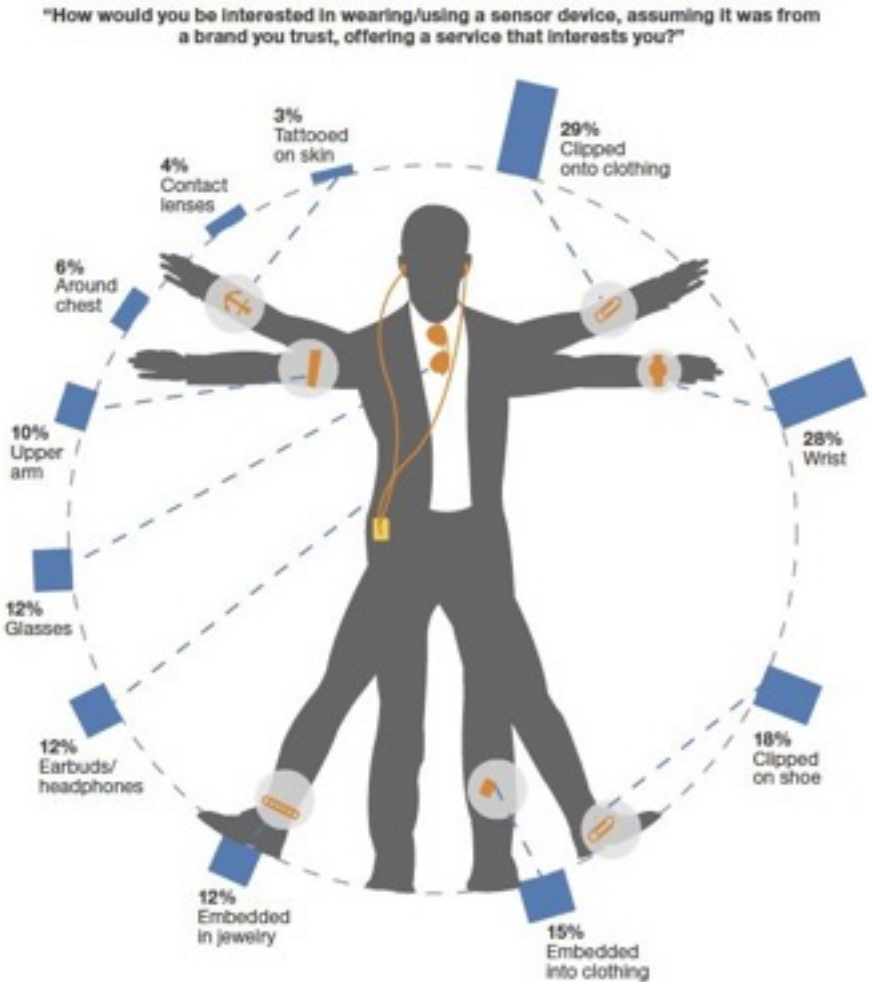
Challenges In Building Wearable AI

Unique Form Factor

Requirements

- Low key
- Clear line of view for visual sensor
- Unobstructed audio signal
- Facing what the user is interacting with

Fisheye lens are used to make sure nothing is left undetected



source: North American Technograpics, Consumer Technology Survey, 2013

Place of computation

	Wearable Device	*Smart Phone*	*Cloud*	*Result*
Option A	*Conduit*	*Conduit*	*Compute*	
Option B	*Conduit*	*Compute*	*N/A*	
Option C	*Compute*	*N/A*	*N/A*	

Place of computation

	Wearable Device	*Smart Phone*	*Cloud*	*Result*
Option A	*Conduit*	*Conduit*	*Compute*	• *Battery life* • *Internet Access* • *Delay* • *Privacy*
Option B	*Conduit*	*Compute*	*N/A*	• *Battery life* • *Delay* • *Privacy*
Option C	*Compute*	*N/A*	*N/A*	*Challenge: How to make it work for a full day*

Privacy In Mind

Local

All processing are done locally on device. Captured raw data are never sent to the cloud.

No Storage

The device does not contain any storage. Photos and audio can not be saved. The goal is processing, not archiving

Lean Throughput

OrCam MyMe uses the lean low energy bluetooth (BLE) that allows for only the processing results to be transmitted

Example:

Face Recognition

Disabilitymatch Podcast

DisabilityMatch Podcast

iTunes Stitcher SoundCloud TuneIn Spreaker Blubrry YouTube Twitter Facebook Google+ Pinterest

Mentoring For Disabled Youth, Innovation For The Blind & Spring Brides

00:00 / 00:00

Play in New Window | Download

The Face Recognition Task

- Appearance
- Lighting
- Occlusion
- Aging

DeepFace (Facebook)

- Taigman et al. 2014
- 3D alignment,
- 120M parameters
- Training: 4,030 people, 1,000 photos each
- On 2.2Ghz Intel CPU: 50ms alignment, 330ms total
- LFW result: 97% with single net
  97.35% with ensemble

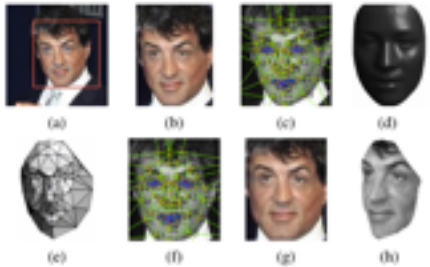
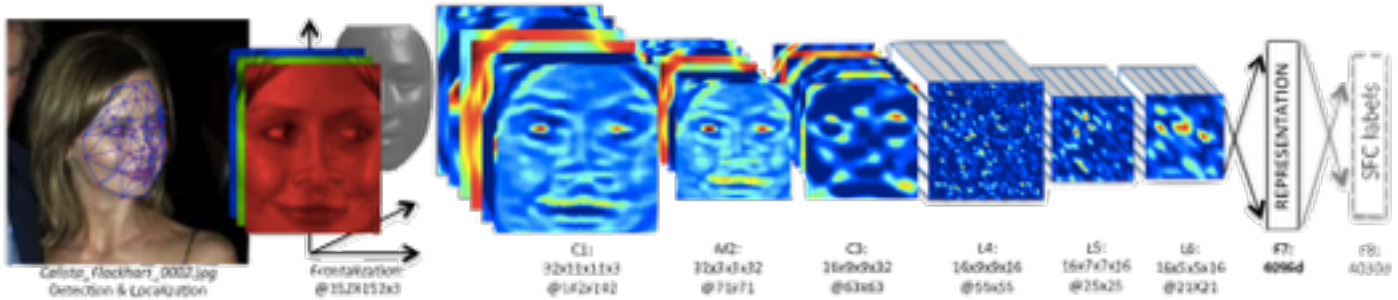


Figure 1. **Alignment pipeline.** (a) The detected face, with 6 initial fiducial points. (b) The induced 2D-aligned crop. (c) 67 fiducial points on the 2D-aligned crop with their corresponding Delaunay triangulation, we added triangles on the contour to avoid discontinuities. (d) The reference 3D shape transformed to the 2D-aligned crop image-plane. (e) Triangle visibility w.r.t. to the fitted 3D-2D camera; darker triangles are less visible. (f) The 67 fiducial points induced by the 3D model that are used to direct the *piece-wise* affine warpping. (g) The final frontalized crop. (h) A new view generated by the 3D model (not used in this paper).



FaceNet (Google)

- Schroff-et-al, 2015
- No alignment
- Training: 8M people, 260M faces total
- 140M & 7.5M parameters. 1.6B FLOPs
- "…and trained on a CPU cluster for 1,000 to 2,000 hours."
- LFW result: 98.87% no alignment
  99.63% with 2D alignment

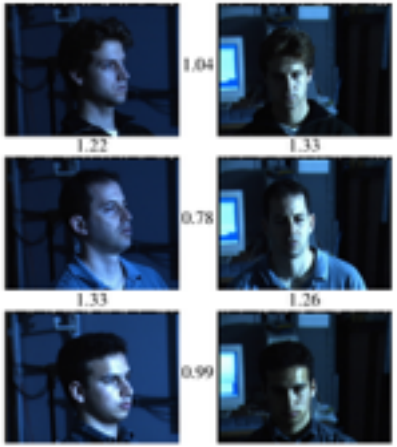


Figure 1. **Illumination and Pose invariance.** Pose and illumination have been a long standing problem in face recognition. This figure shows the output distances of FaceNet between pairs of faces of the same and a different person in different pose and illumination combinations. A distance of 0.0 means the faces are identical, 4.0 corresponds to the opposite spectrum, two different identities. You can see that a threshold of 1.1 would classify every pair correctly.

Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$

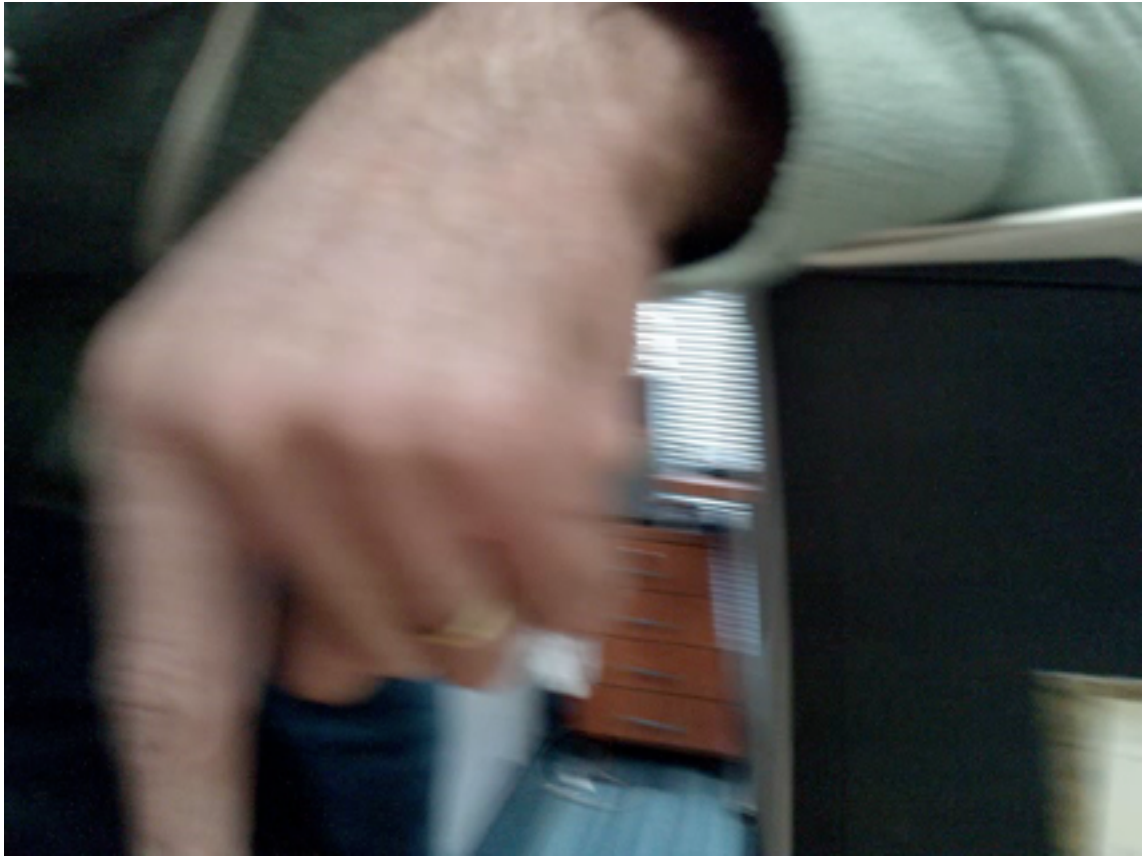
OrCam

- alignment: DNN with 4.8M FLOPs, 69K param
- Training: 2.5M faces total
- 1M parameters, 82M FLOPs
- Training time: 10 hours
- LFW result: **97.8%**
- Runtime: **30ms** on Cortex A9 core

1/100 of training data
1/100 of training time
1/20 runtime

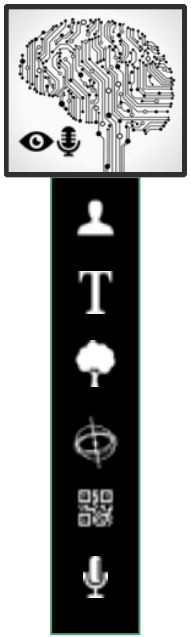
Embedding evolution in test data


6.32
8.02
8.79
5.73
4.44
6.82



Summary

Understanding The World Around Us



Eating and fitness habits

How you spend your time

Habits: Books, TV, Phone etc'

people you spend time with

Emotion analysis of people you encounter

Conversation topics