Embedded Vision Summit
Santa Clara, May 2017

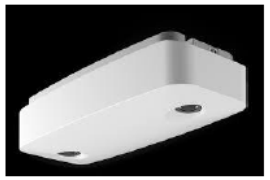
2017

Embedded Vision Summit 2017

- Yearly event since 2012
- Focus: practical computer vision and visual intelligence
- Past talks are free to access!

What is Embedded Vision anyway?

- A camera with a processor inside?
- An embedded system with an integrated camera?
- An embedded system with an external camera?
- A camera-equipped device with remote intelligence in your phone?
- A surveillance camera with intelligence in the cloud



Images: EXPO210XX.com, RetailNext, Camio, Anki,, Amazon

Overall

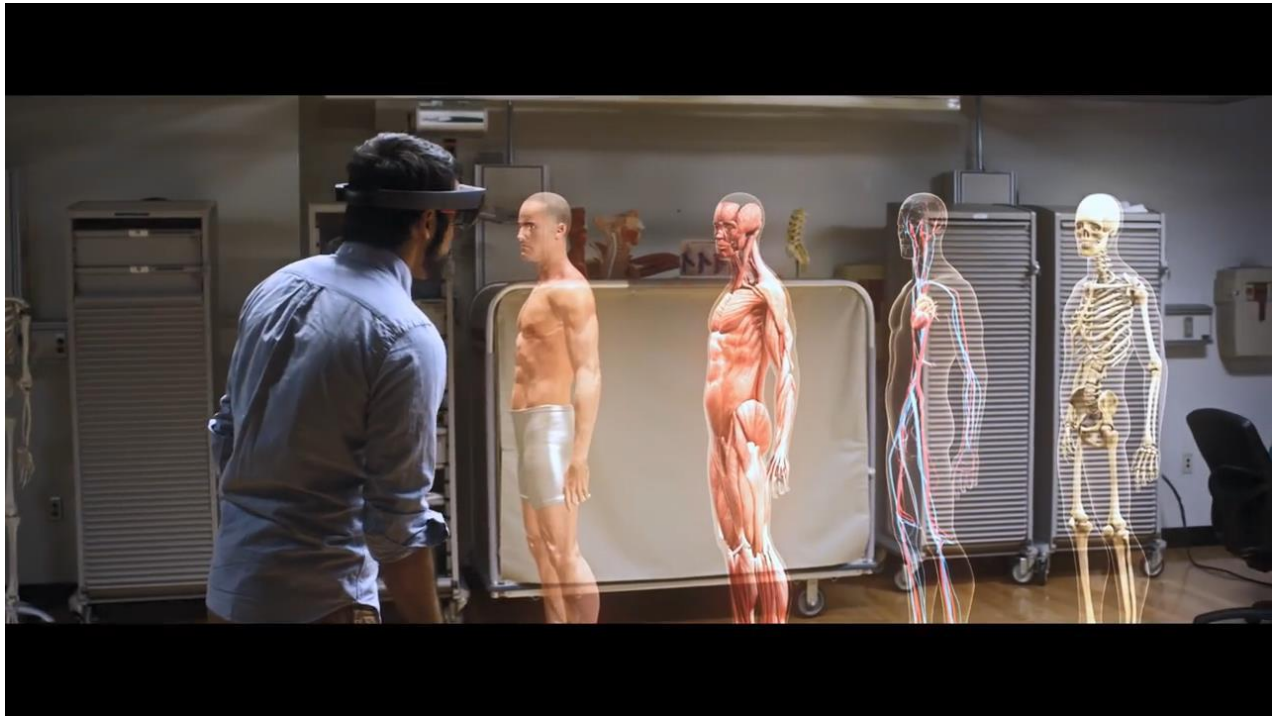
- Multiple tracks:
  - technical insights track
  - enabling technologies
  - Business
  - Fundamentals
  - Showcase
  - Not surprisingly: AI focused

Microsoft: 3D Computer Vision and Mixed Reality

- Speaker: Marc Pollefeys
- Microsoft HoloLens
- 3D computer vision
- Architecture

Applications

- 3D learning/ medical
- https://www.youtube.com/watch?v=h4M6BTYRlKQ



Applications: Medgadget

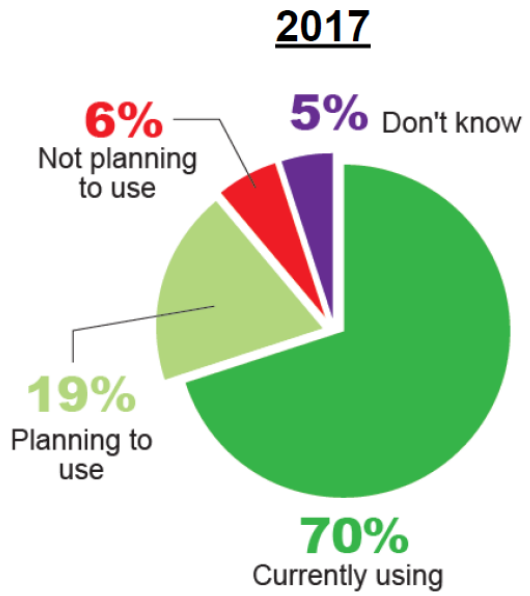
- Scopis Introduces Mixed Reality to Simplify Surgical Navigation
- https://www.youtube.com/watch?v=xvbWE4OsKxY

Other Applications

- Tele presence, design

What will happen by 2020?

- Time to reflect:
  - 2014: keynote on AI
  - 2017: AI explosion
  - 2017: Keynote on Augmented/Mixed Reality
  - 2020: ?



Microsoft: Emotion Recognition

- Applications
  - Wide array of applications

Affect-aware personal assistant/companion devices

Autism intervention

Honest signal

Affect-aware game development

Basic Emotion Types

Neutral

Sadness

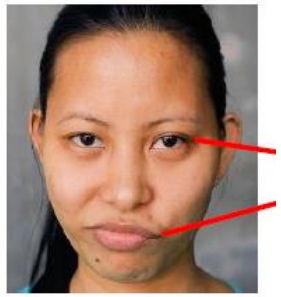
Disgust

Happiness

Anger

Surprise

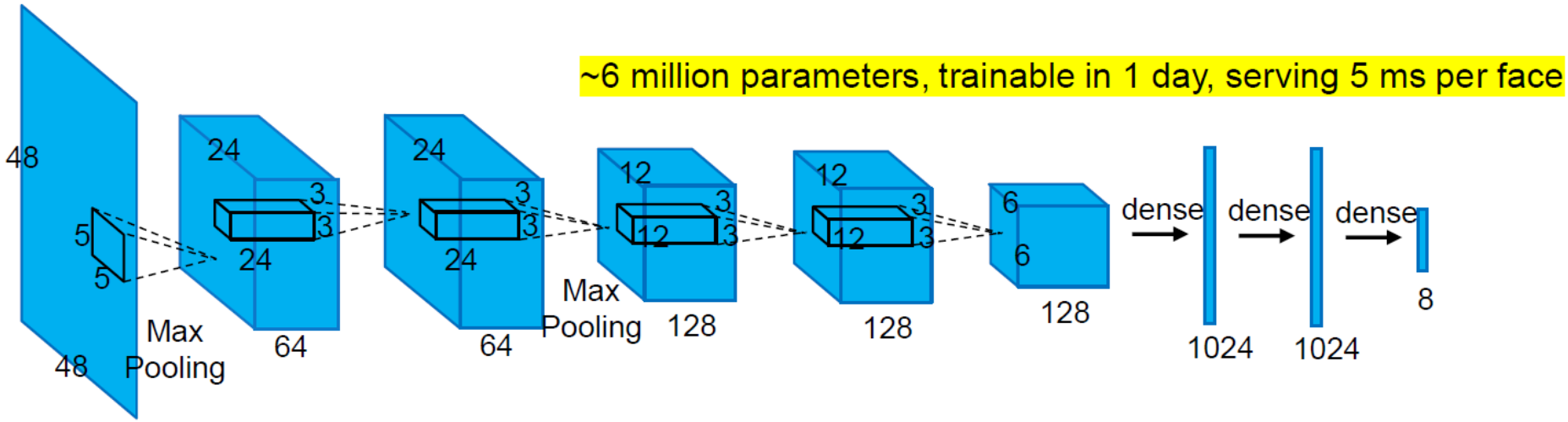
Contempt



Fear

AI Architecture

- A convolutional neural network (CNN) based approach



- Experimented with ResNet and others
  - Achieved similar performance
  - Limited by label accuracy

Data Collection

- Start with FER 2013
  - Web crawled + human labeling
  - 48x48 image resolution
  - 28709 training examples
  - 3589 validation examples
  - 3590 test examples
  - Very noisy data
- Train DCNN
  - Without data augmentation
    - 65.07%
  - With data augmentation
    - 71.73%

Data Collection (cont.)

- Crawled ~4.5 m images with emotional keywords
  - 166 emotional adjectives
  - 230 celebrity names, 100 popular first names, 166 people related words
- Face detection
- Active learning
  - Use DCNN to select confusing facial images for tagging
- Self-paced learning
  - Use DCNN to expand training data based on classification results
  - Randomly sampled
  - Biased towards rare emotion types

Crowd Sourcing Labeling

- Taggers are forced to choose one emotion out of 8, or tag the face image as “unknown”
- We started with at least 2 taggers agree and up to 5 taggers
  - Quality was very bad specially with subtle emotions
- We retagged all our data with 10 taggers
  - Quality improved drastically (detailed next)

How many taggers needed?

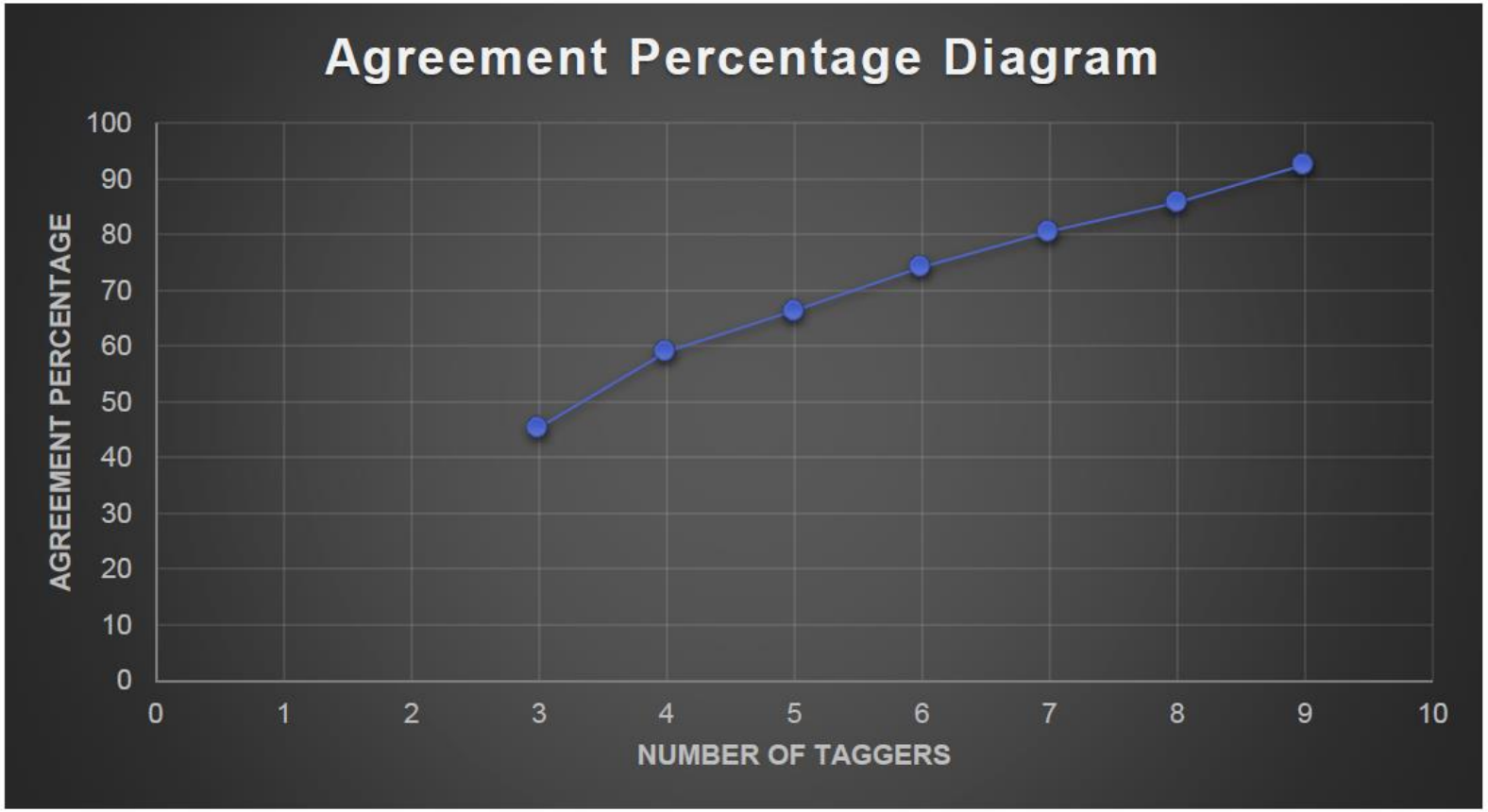
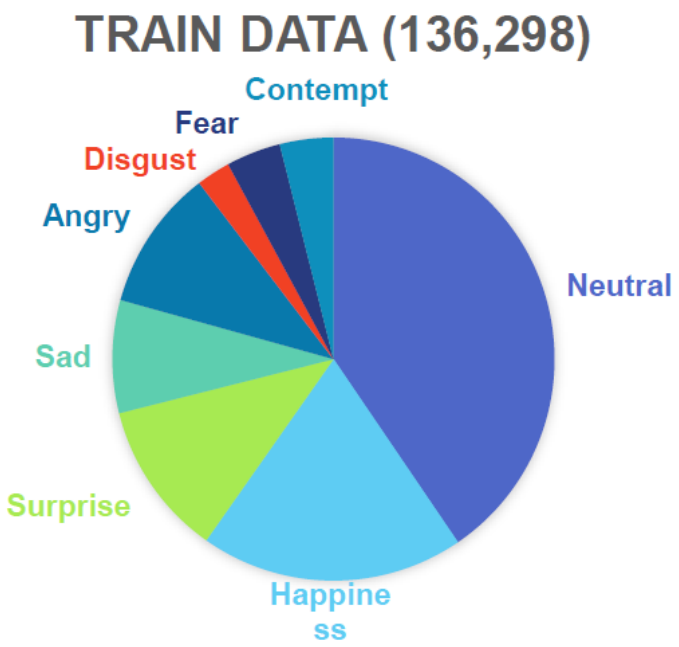
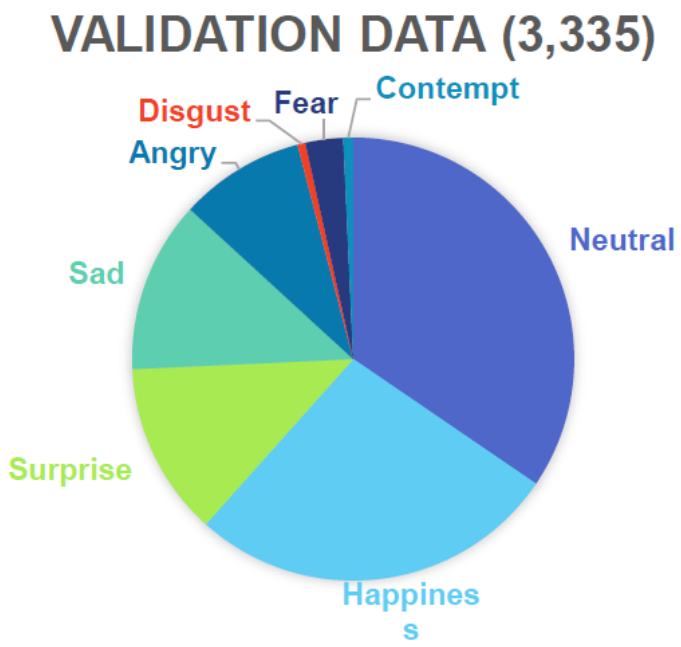
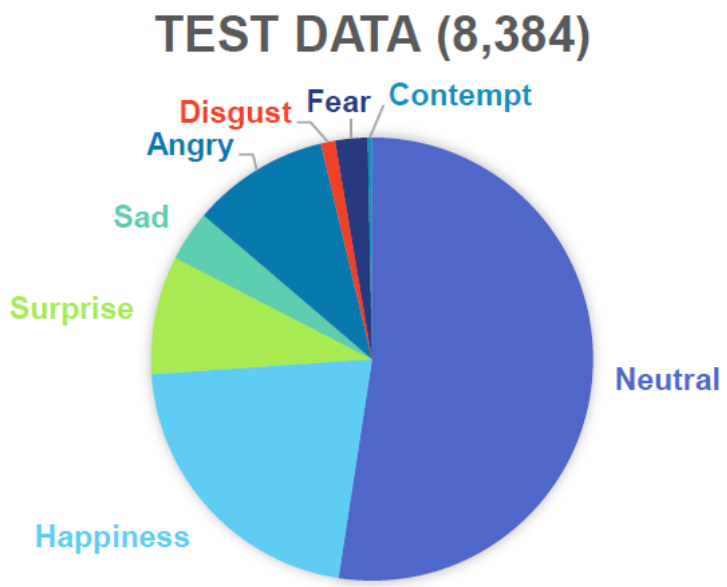


Diagram on percentage of agreement with final majority label

Final Dataset

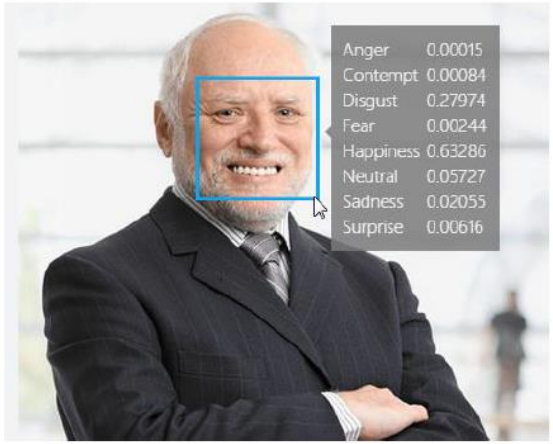






- 85.26% on test set

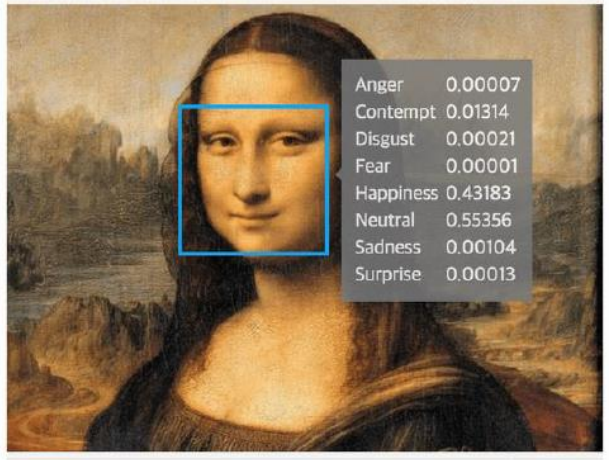
Examples



Happiness: 0.63
Disgust: 0.27

"Pretty awesome that it detected the underlying emotion."

Happiness: 0.43
Neutral: 0.55



"That oddness in her smile is contempt apparently, makes sense"

"The face of pure happiness."

Try Emotion Recognition

- https://azure.microsoft.com/en-us/services/cognitive-services/emotion/
- Search for Microsoft emotion recognition API

Jeff Bier: Conference Founder

- The power and power consumption of vision computing will decrease by **1000** over the next **three years**.

Image: BusinessInsider.com

Why does this matter?

1948:
First mobile phone

Image: Wb6nvh.com

2017 (70 years later):
More than 7 **billion** phones

Image: finder.com.au

- Mobile phones and wireless data went from “inconceivable” to “ubiquitous”
- Enabled thousands of new businesses and business models along the way
- Example: Facebook has 1.7 billion mobile users, 56% are mobile-only
- Vision is poised to be the “new wireless”

How will we get there?

1. Algorithms — 10x
2. Processors — 10x
3. Frameworks, tools, middleware — 10x

= 1000x

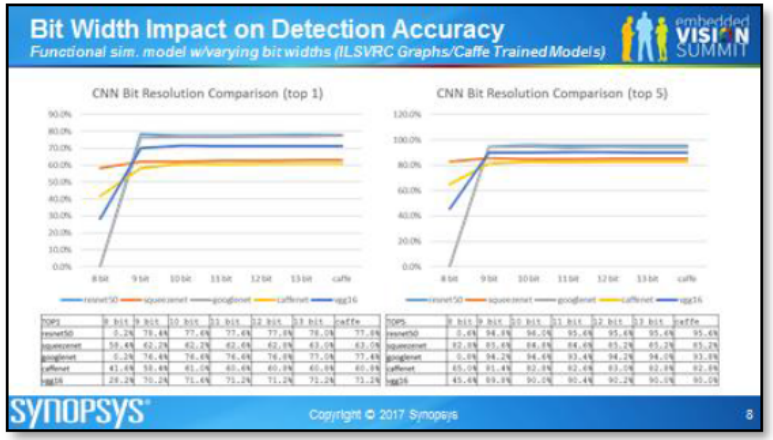
Algorithms

- Many techniques have been proven effective for slimming down existing algorithms
  - E.g., reducing data types:
    - 32-bit floating-point → 16-bit float → 16-bit fixed → 8-bit fixed →?
  - Pruning
  - Compression

Network	Top-1 Error	Top-5 Error	Parameters	Compress Rate
Baseline Caffemodel (BVLC)	42.78%	19.73%	240MB	1×
Fastfood-32-AD (Yang et al., 2014)	41.93%	-	131MB	2×
Fastfood-16-AD (Yang et al., 2014)	42.90%	-	64MB	3.7×
Collins & Kohli (Collins & Kohli, 2014)	44.40%	-	61MB	4×
SVD (Denton et al., 2014)	44.02%	20.56%	47.6MB	5×
Pruning (Han et al., 2015)	42.77%	19.67%	27MB	9×
Pruning+Quantization	42.78%	19.70%	8.9MB	27×
Pruning+Quantization+Huffman	**42.78%**	**19.70%**	**6.9MB**	**35×**

Hong et al., 2016

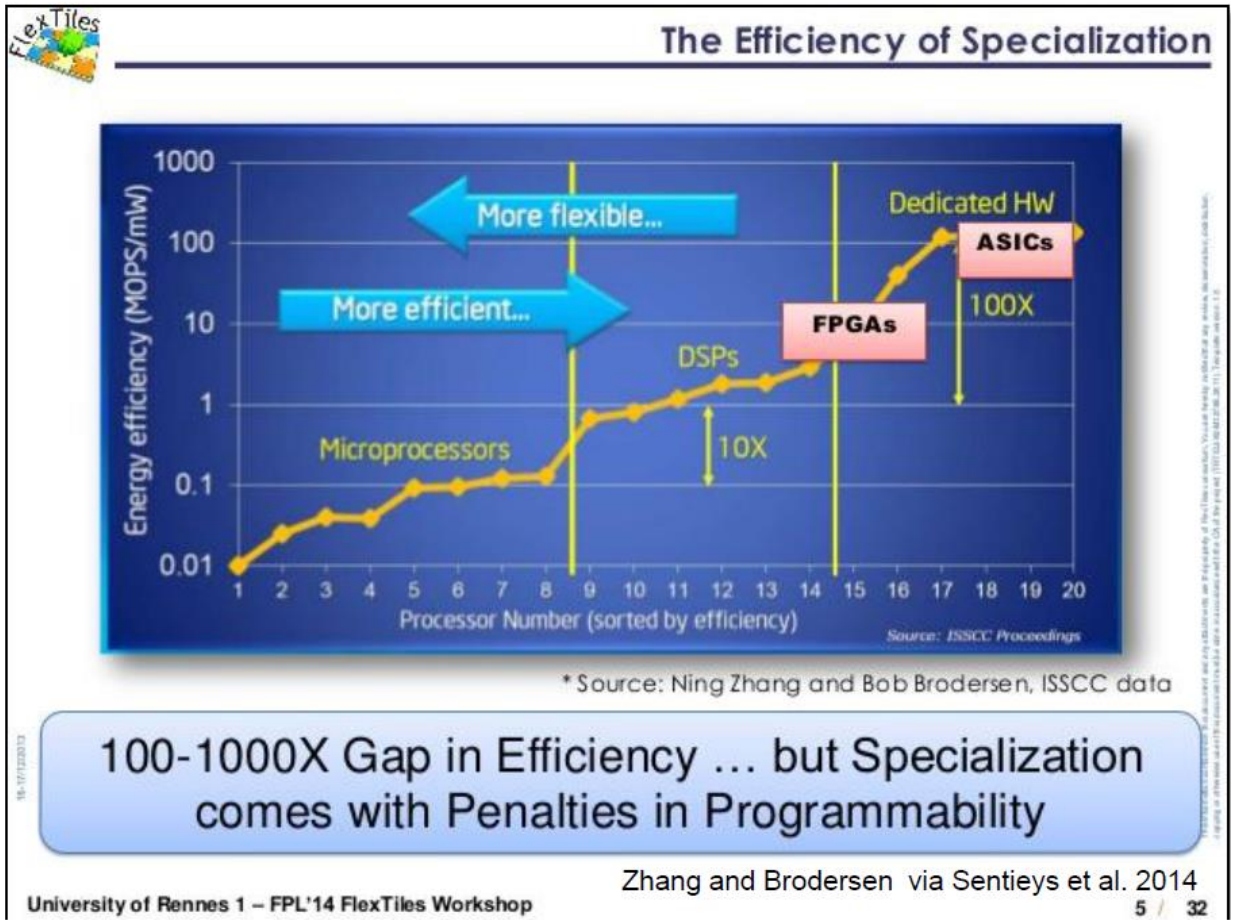
2x-
37x



Tom Michiels, Synopsys

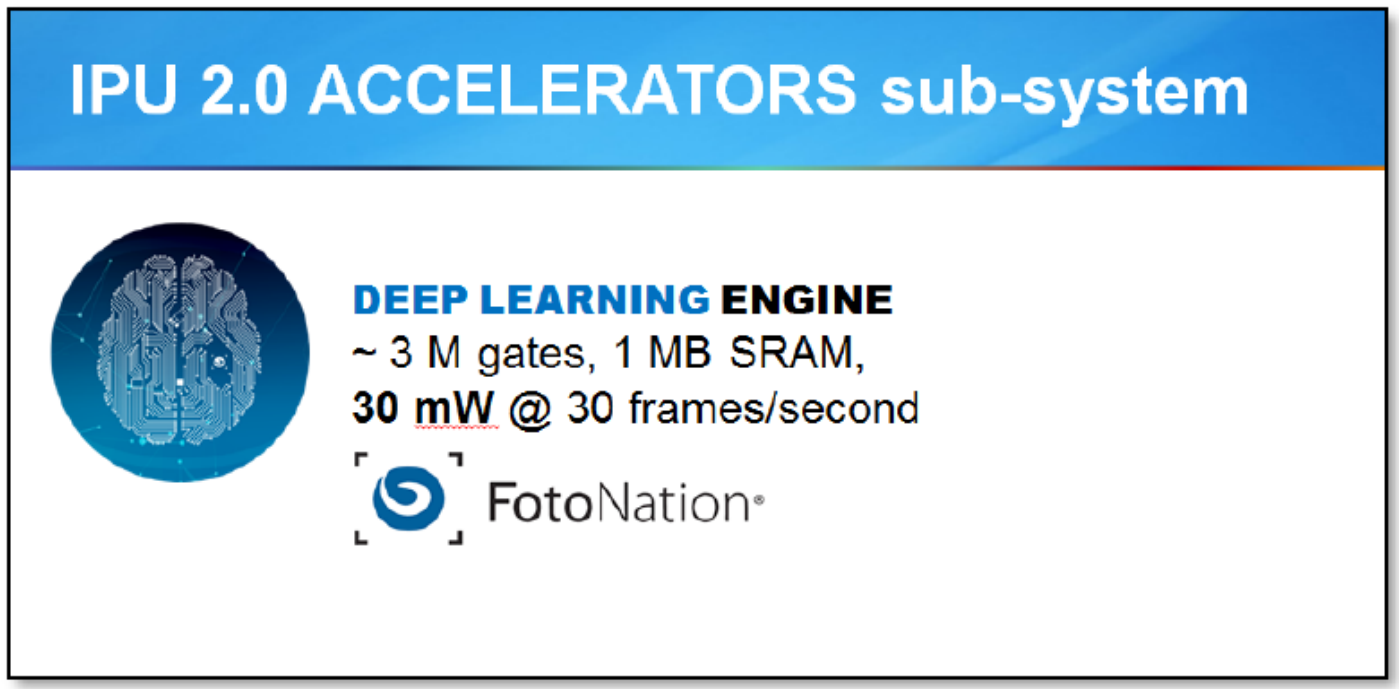
Processors

For decades, chip designers have created specialized processors to get big gains in cost/performance and energy-efficiency



Processors

Today, dozens of chip and IP core suppliers are creating processors specialized for deep neural networks

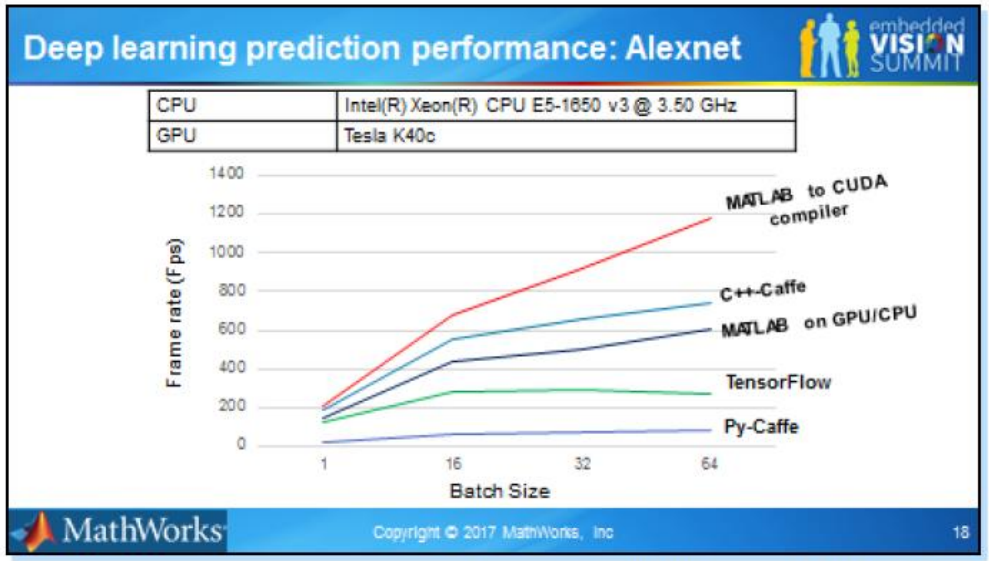


Processors

- GPU vs USB stick

Frameworks and SW tools

Today, dozens of software, chip and IP core suppliers are creating software tools specialized for deep neural networks



Venkataramani and Nehemiah, MathWorks

Highlights

- **Automate compilation** of MATLAB to CUDA
- **14x speedup** over Caffe & **3x speedup** over TensorFlow

1000 times in three years

1. Algorithms — 2x-37x ➔ ~18x
2. Processors — 7x-37x ➔ ~20x
3. Frameworks, tools, middleware — 3x-14x ➔ ~8x

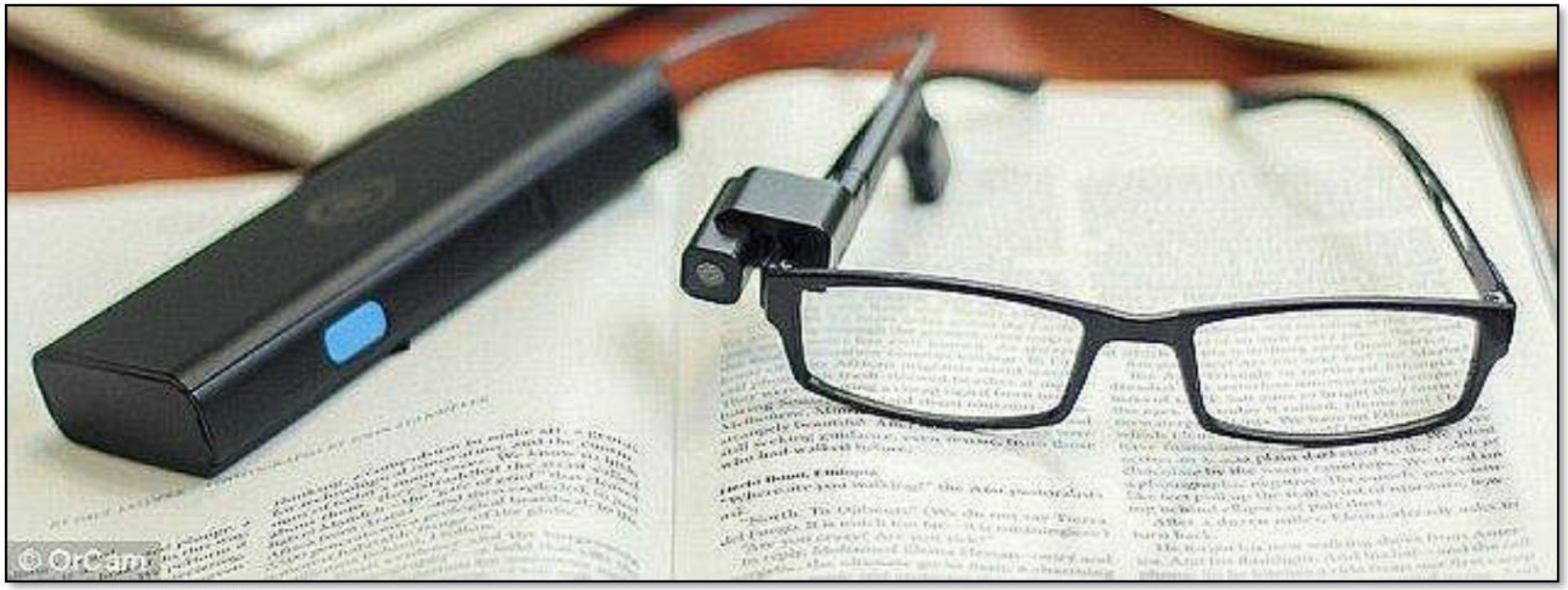
= ~2,880x

What does this mean?

- Many types of devices and systems can become safer, more autonomous, easier to use and more insightful
- Unheard-of new applications of vision
- Many new, visually intelligent devices that are:
  - Inexpensive
  - Battery powered
  - Always on

Orcam: for visually impaired

- **cost today: $3500**



The people person Tie Tack

- Recognizes every person in your network
  - Face, iris
- Whispers their name to you via Bluetooth if you don’t seem to remember! ☺

Empathetic Teddy Bear

- Recognizes your child, his or her friends, your pets, other family members
- Can read emotions and interact socially
- Sends you cute photos!
- (No mobile phone required!)

Today we have

- Production-worthy augmented reality glasses costing $3,700 and weighing 1.3 lbs.

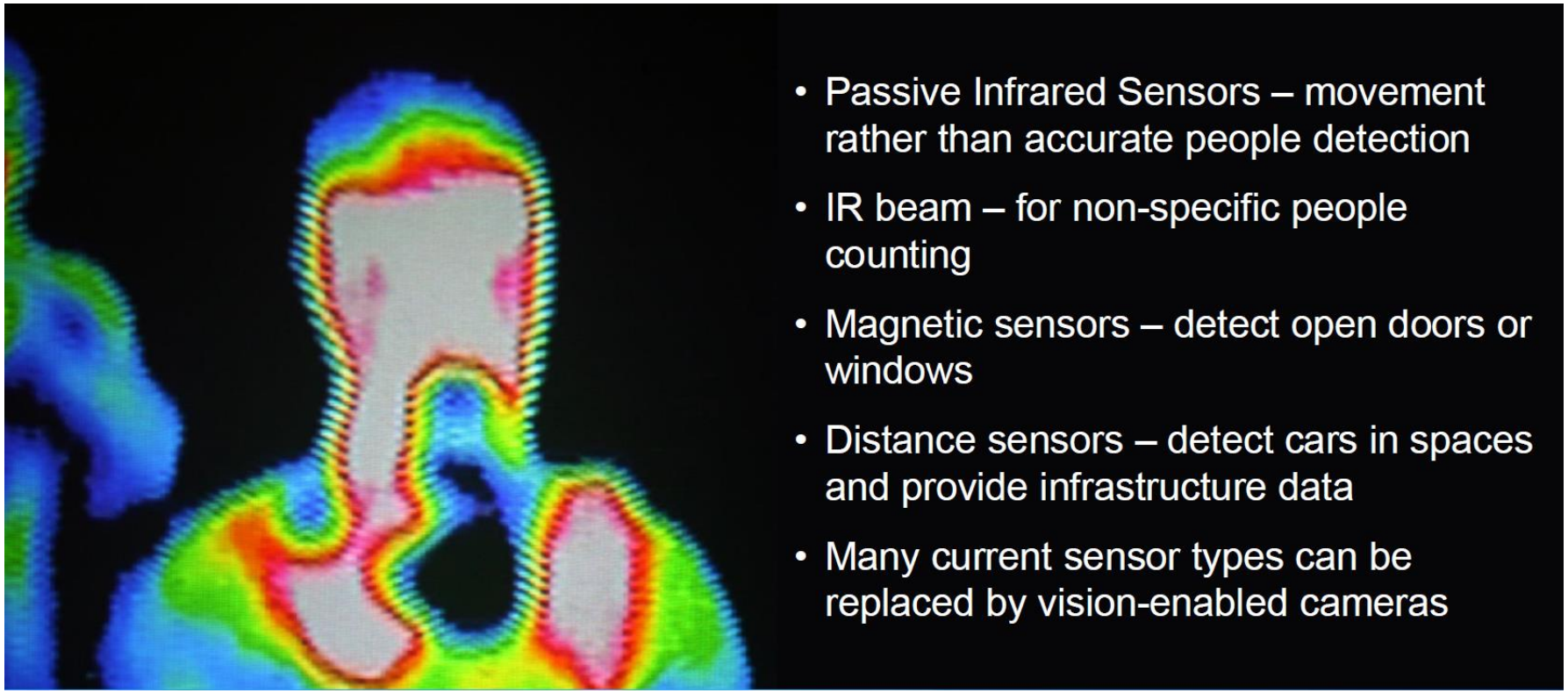
Image: Microsoft

ARM: Tim Ramsdale

- This changes everything. Why computer vision will be everywhere.

Many sensors today are proxies of vision

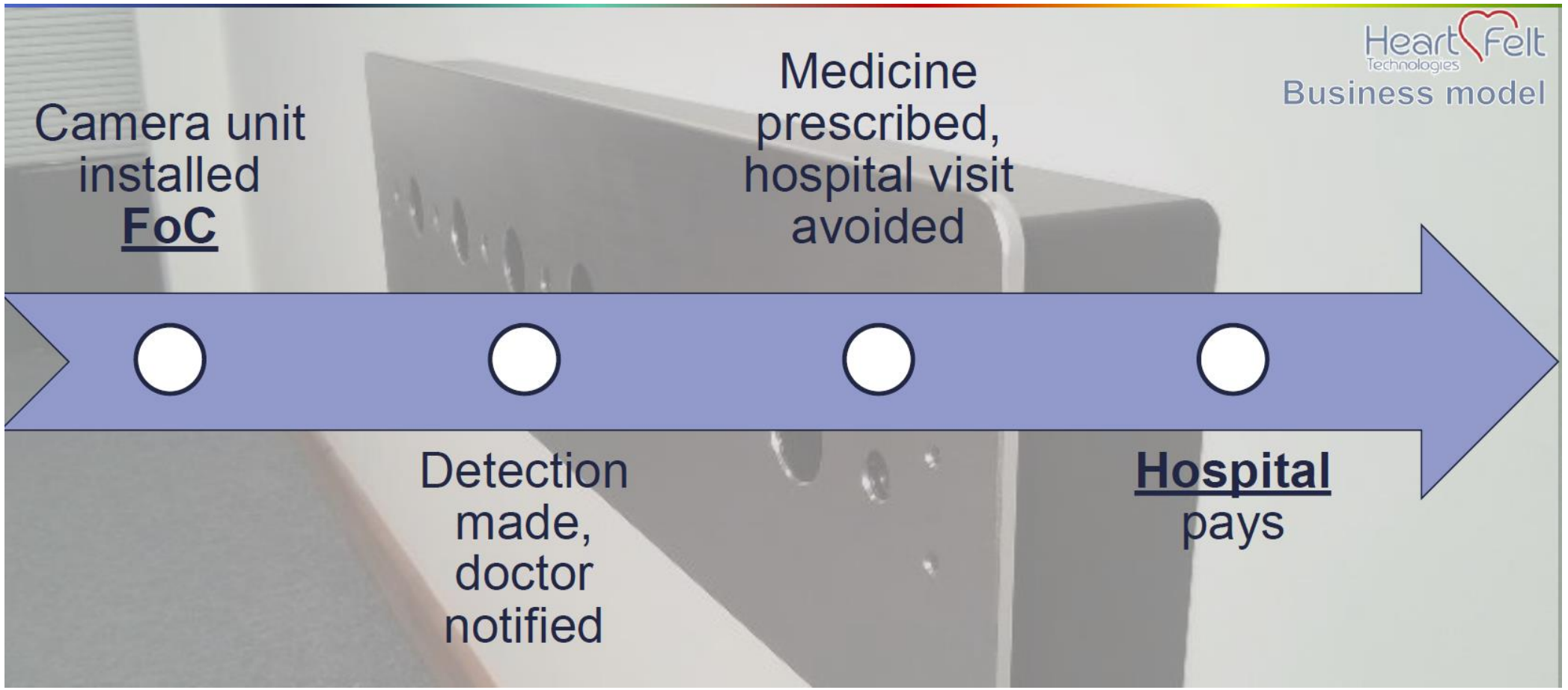
- Sensors today will be replaced by cameras



Vision technology to prevent heart failure

- Home Heart Failure Monitoring
- Swollen ankles and heart failure
- http://www.hftech.org/

Vision technology to prevent heart failure



Vision Technology in logistics

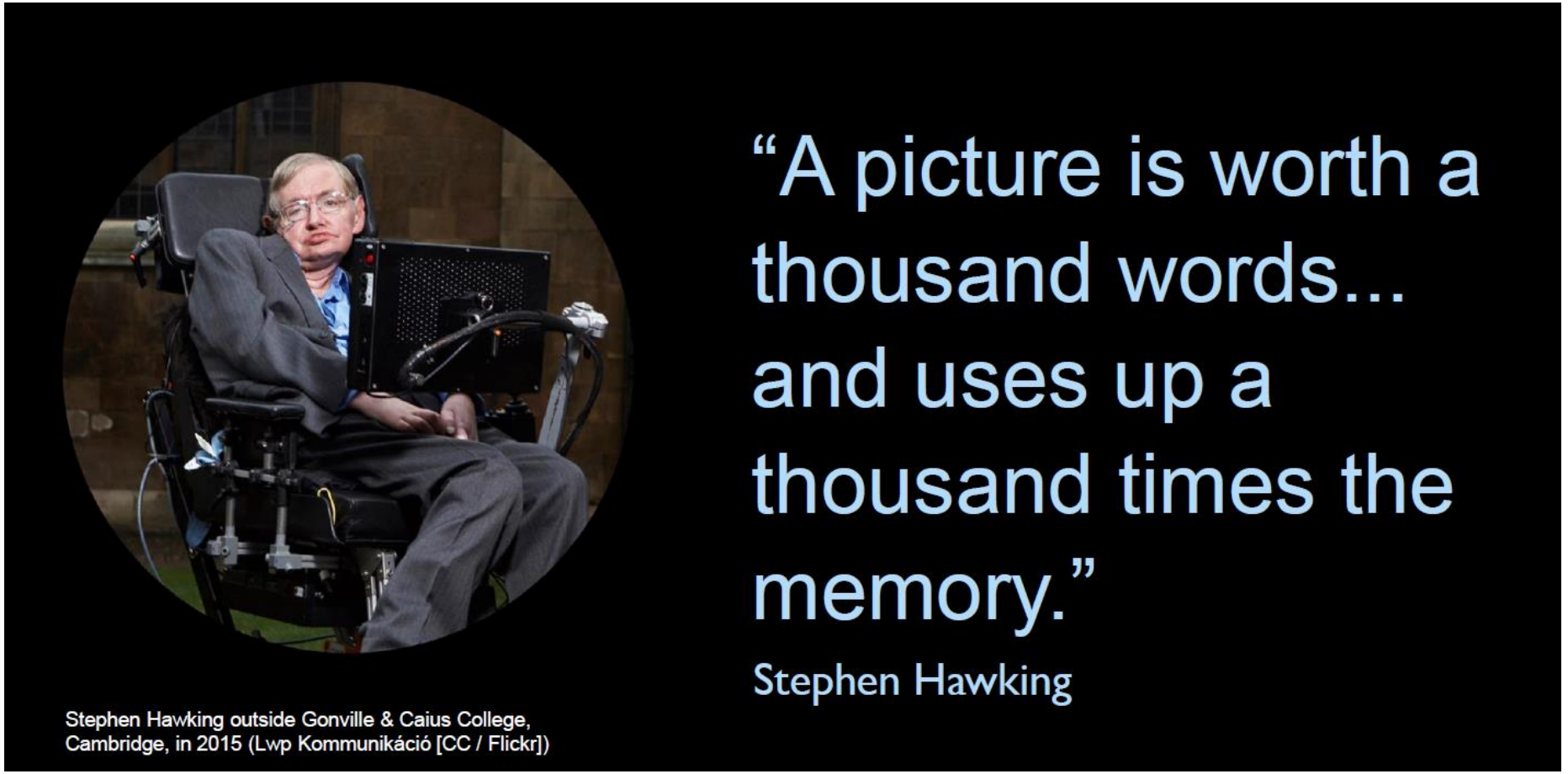
Vision Technology in Agriculture

- CV being used to selectively harvest crops:
  - Vision allows crops to be picked only when they're ready
  - Autonomous harvesting vehicles further reduce human interaction
- Business model:
  - Potential for significantly improved yields for non-uniform crops
  - Reduction in man power & human error

CROPBS "Clever Robots for Crops"

Vision in the home

The Time for imaging is now …



What does the future hold?

Vision will be everywhere and will change everything

Sensors will be replaced by cameras

Real-time edge processing and Security are key

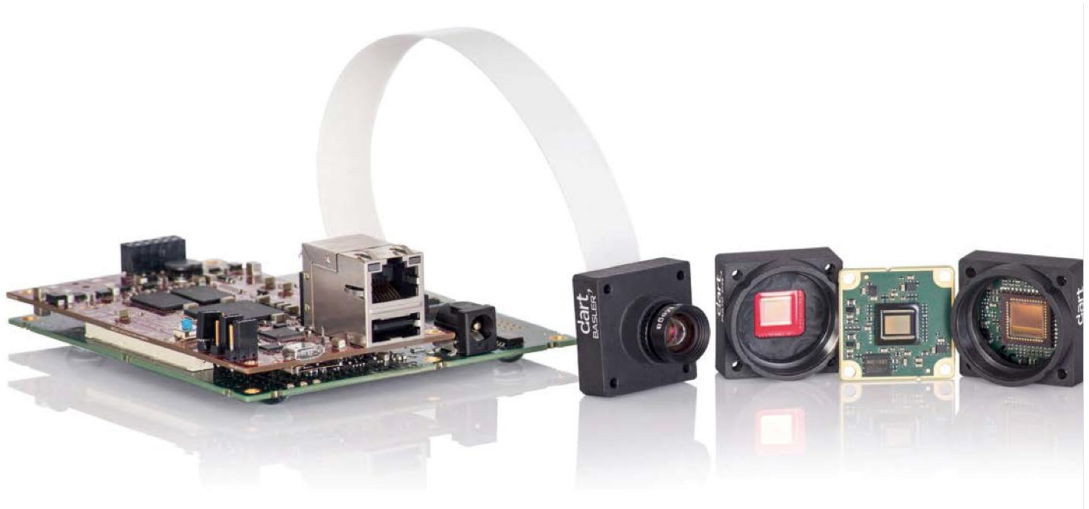
Interesting new business models based on early detection and optimizing processes

We need to optimize the complete system

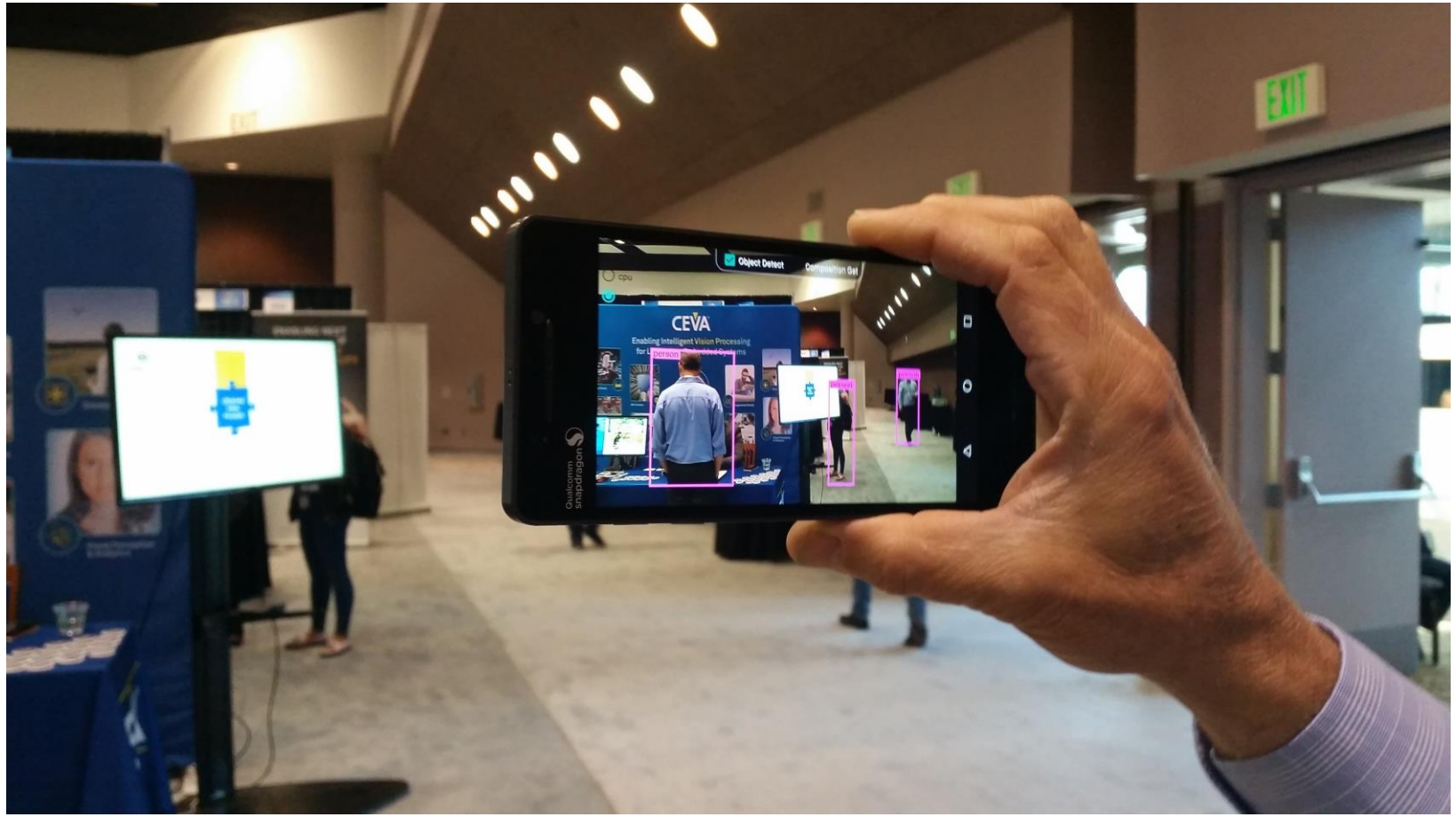
Showcases


BDTi
cadence
CEVA
CrowdFlower
D3 Engineering
embedded VISION ALLIANCE
EUREKA
eWBM
FotoNation
GDEL
HSA FOUNDATION
Imagination
iMerit
human-empowered computing
intel
IntelliVision
INUITIVE
LATTICE
LUCID
LUXOFT
machines with vision
MathWorks
Mentor
A Siemens Business
mipi alliance
Morpho
MVTec
MVTec Software GmbH
NEW ELECTRONIC TECHNOLOGY
nextchip
NXP
PATHPARTNER
PEARL
PERCEPTIN
REM
RENESAS
RTA
sama SOURCE
SHAPER
STRADVISION
SYNOPSYS
tend
ThunderSoft
VectorBlox
embedded supercomputing
VeriSilicon
videantis
passion for video
vision components
Wave Computing
XILINX
ALL PROGRAMMABLE
ximea

Semiconductors



Object Detection on Smartphone



Object Detection on USB Stick

Traffic Sign Classification

Pain Management

- Entertainment for pain management!
- Cozmo robot
- Autonomous Robot
- Interactive, aware
- https://www.youtube.com/watch?v=ldi1NC

LUCID

- https://www.youtube.com/watch?v=9SZldrmCg10
- 3D VR Camera
- Cost: $499.00

Summary

- Vision is everywhere
- Machine vision is getting smarter using AI
- Specialized processors for smart machine vision are coming
- Lots of opportunities for healthcare industry