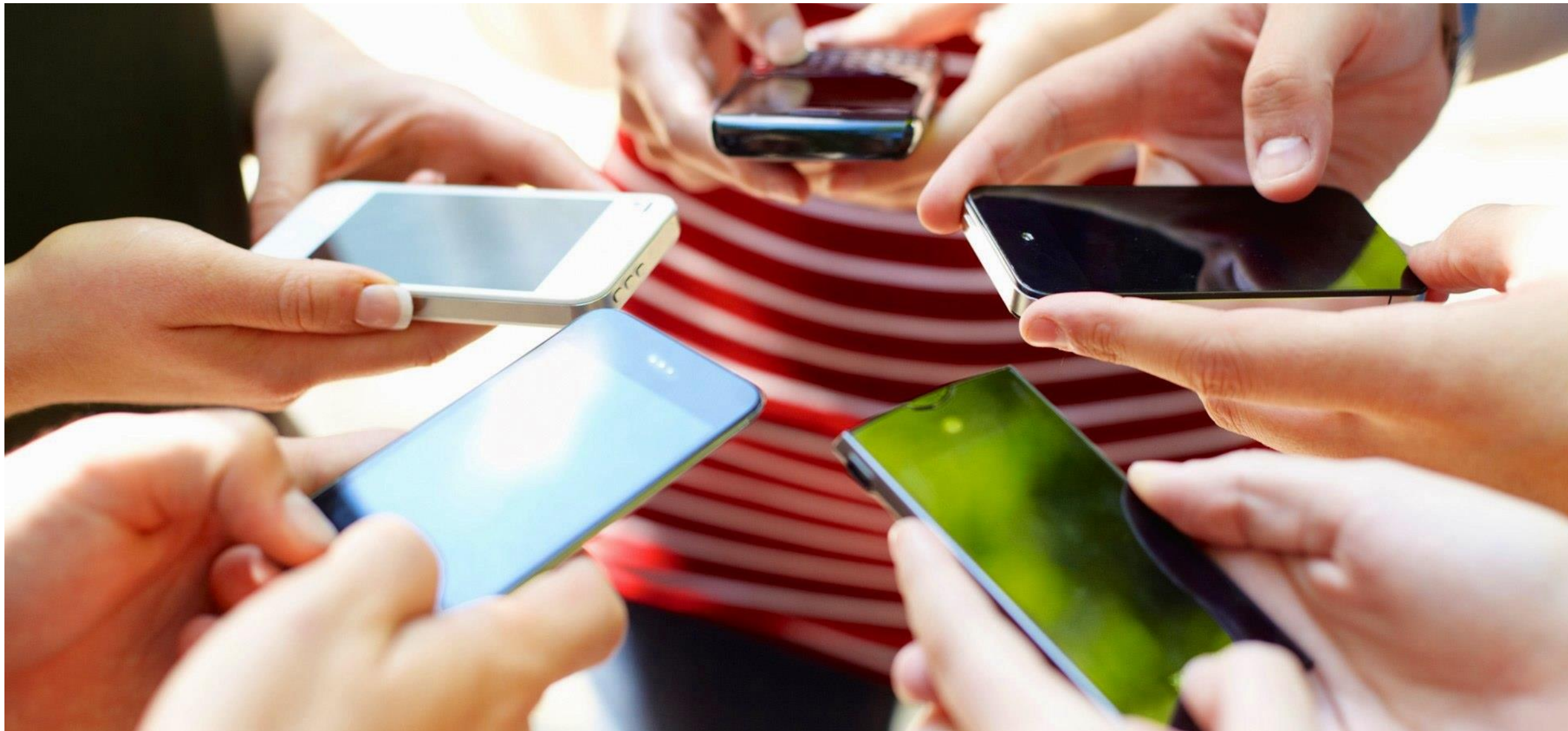


In-air Gestures Around Unmodified Mobile Devices

Jie Song¹, Gabor Sörös¹, **Fabrizio Pece**¹, Sean Ryan Fanello²,
Shahram Izadi², Cem Keskin², Otmar Hilliges¹

¹ETH Zurich

²Microsoft Research



Issues with Touch Input



Complement Touch with Gestures



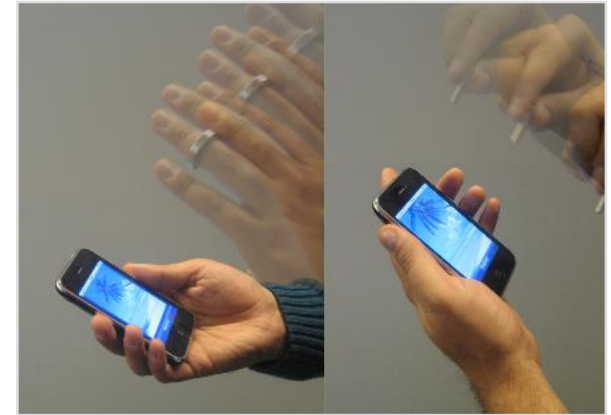
[Le Goc et al. *CHI 2014*]



[Kim et al. *ISWC 07*]



[Ketabdard et al. *MobileHCI 10*]



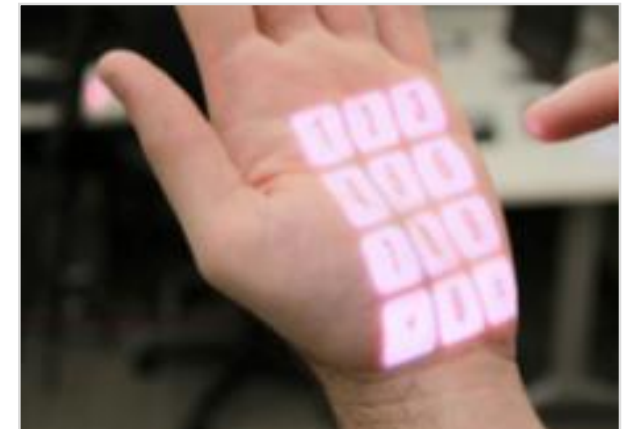
[Harrison et al. *UIST 2009*]



[Wigdor et al. *UIST 2007*]

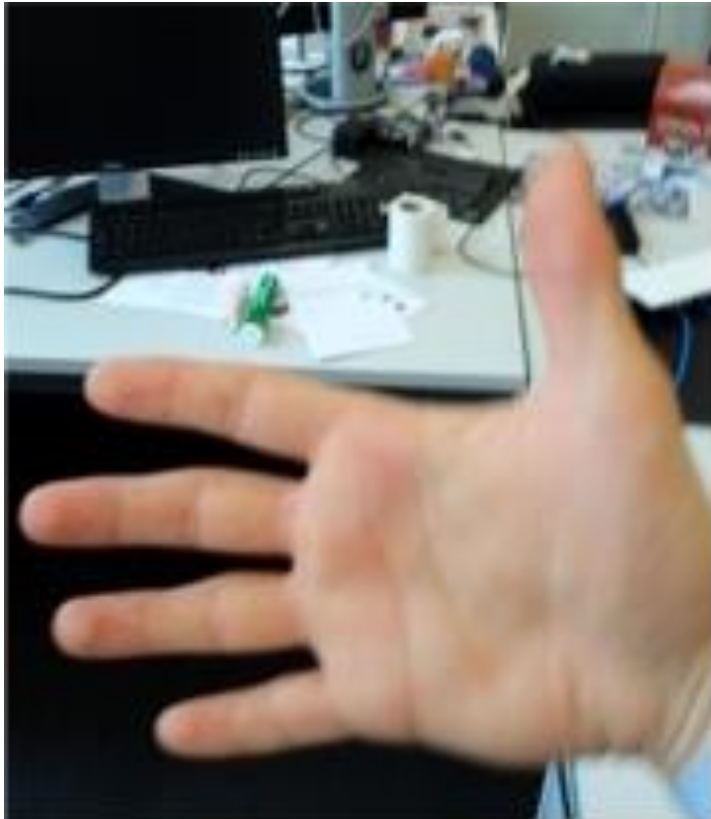


[Harrison et al. *CHI 2010*]



Problem Statement

INPUT



SEGMENTATION



LABELED OUTPUT

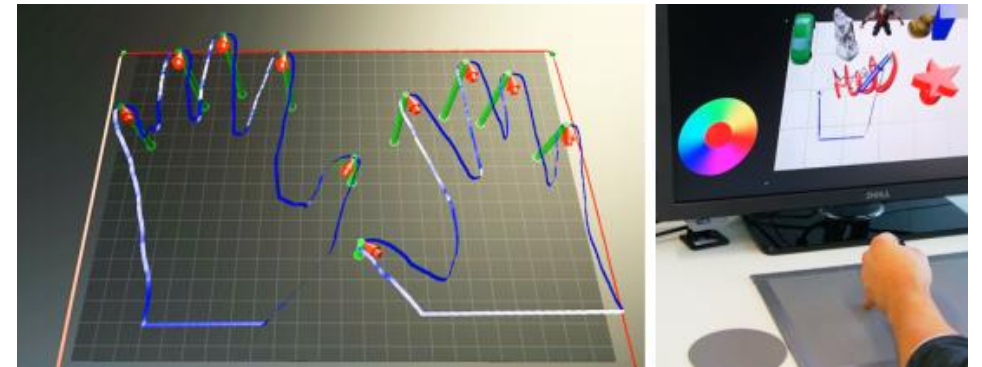


Randomized Decision Forest

Body Pose Estimation - Shotton et al. *CVPR '11*



Finger Part Classification - Kim et al. *CHI '14*



Hand Pose Estimation - Keskin et al. *ECCV '12*

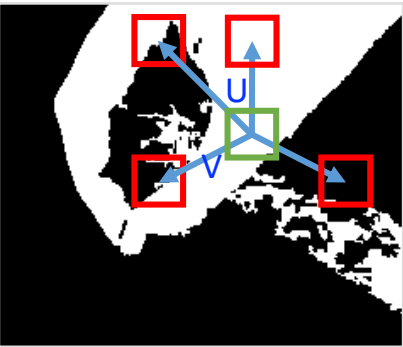


Motion Gesture Recognition - Taylor et al. *CHI '14*

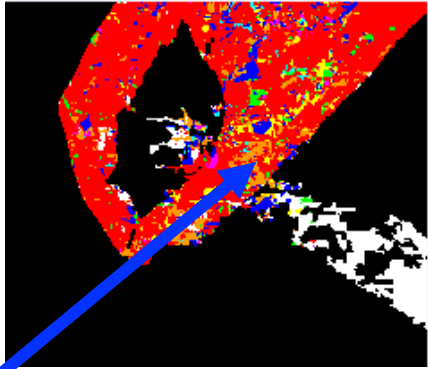


Randomized Decision Tree

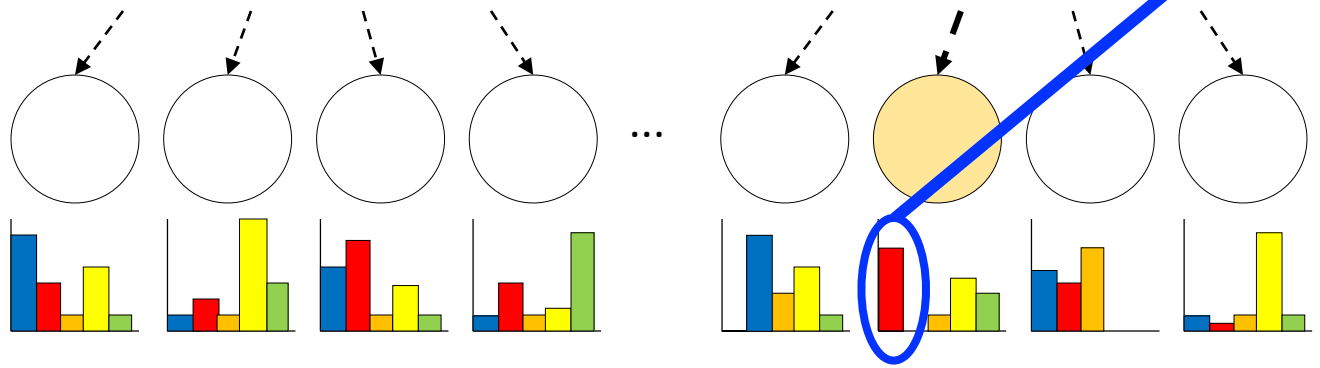
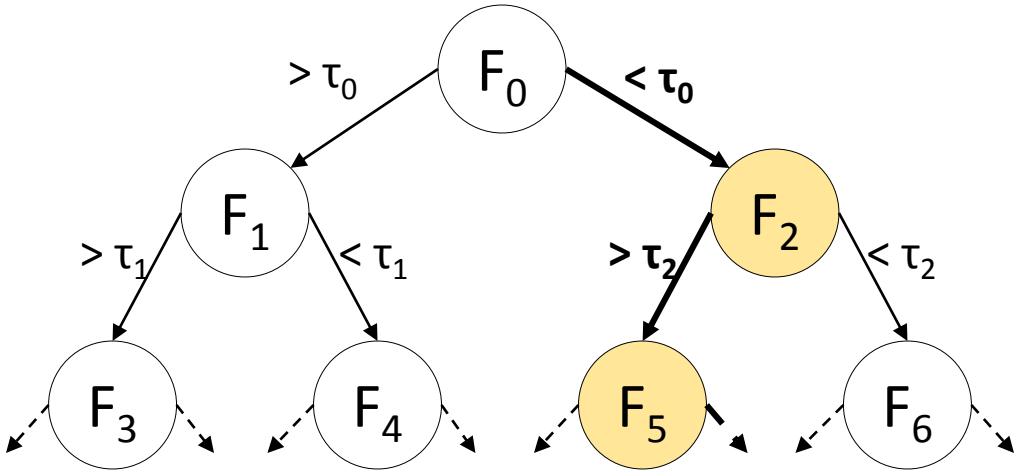
Input



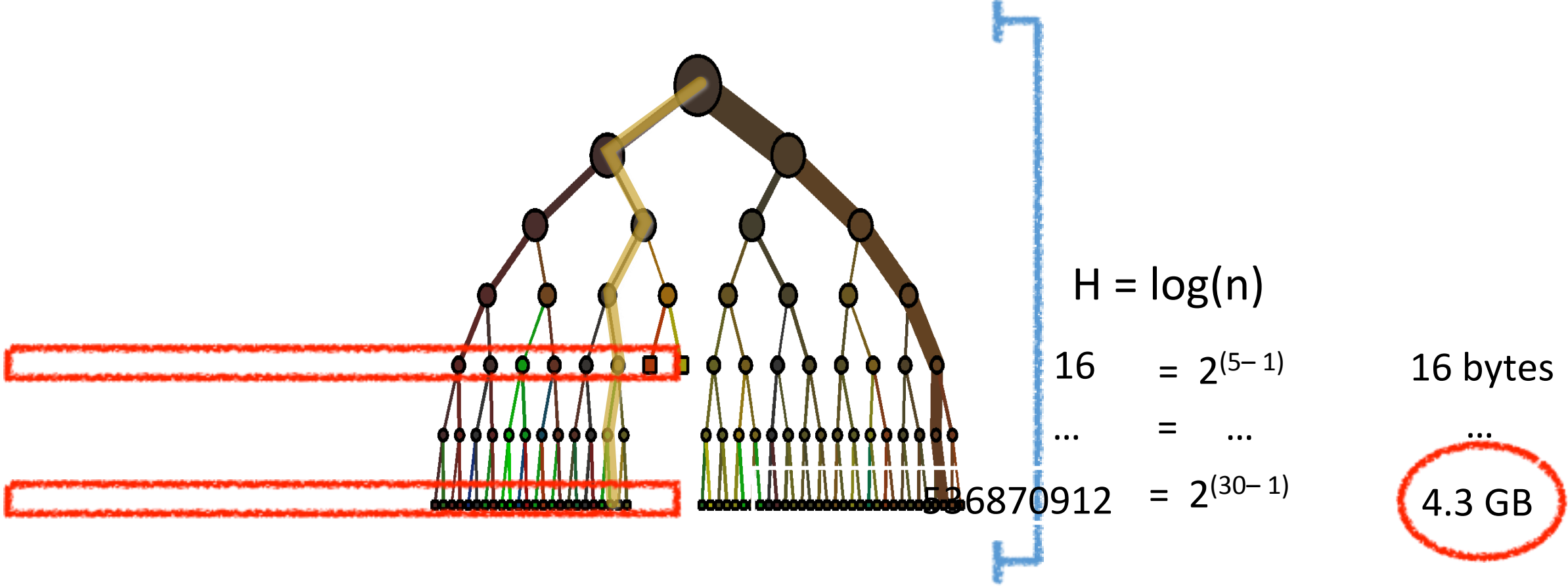
Output



■ "Class A"



Runtime Efficiency vs Memory Footprint



$H = \log(n)$

16 = $2^{(5-1)}$

... = ...

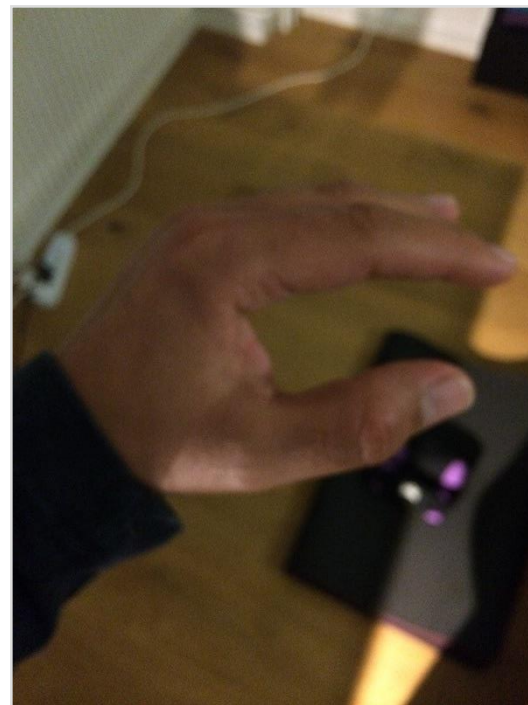
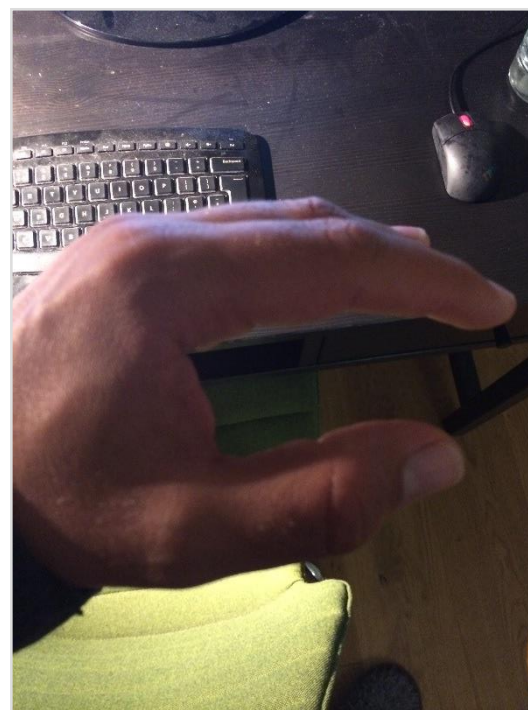
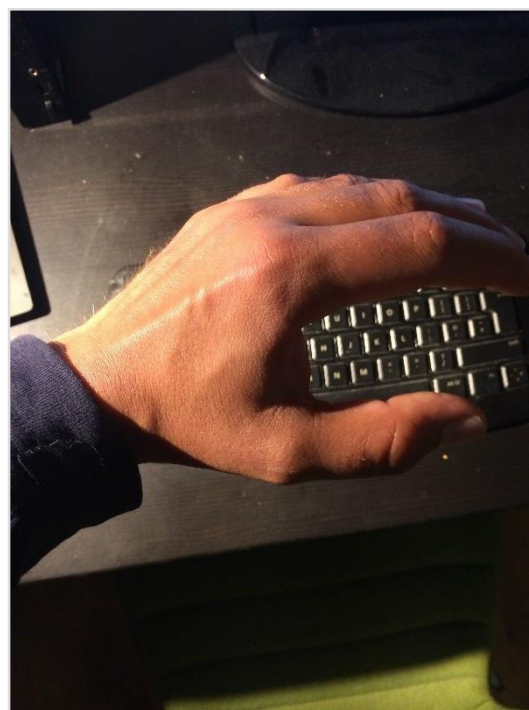
536870912 = $2^{(30-1)}$

16 bytes

...

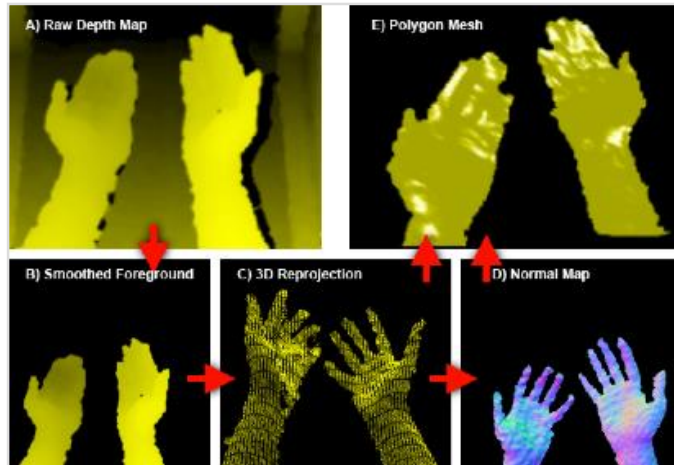
4.3 GB

$N = \exp(H)$

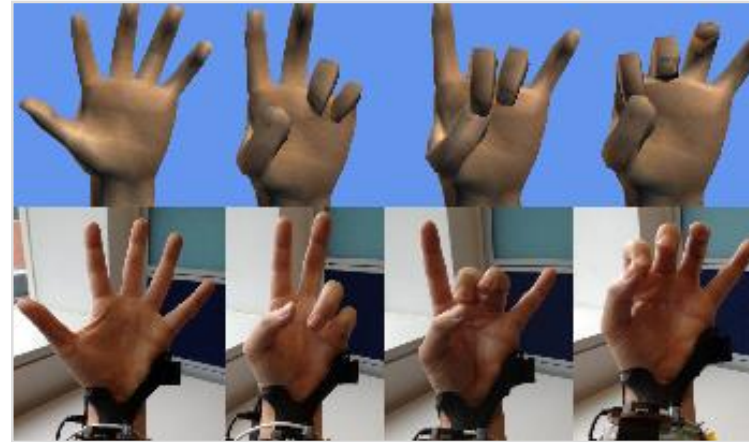


Gesture Recognition with Depth Camera

[Hilliges *et al.* CHI'12]



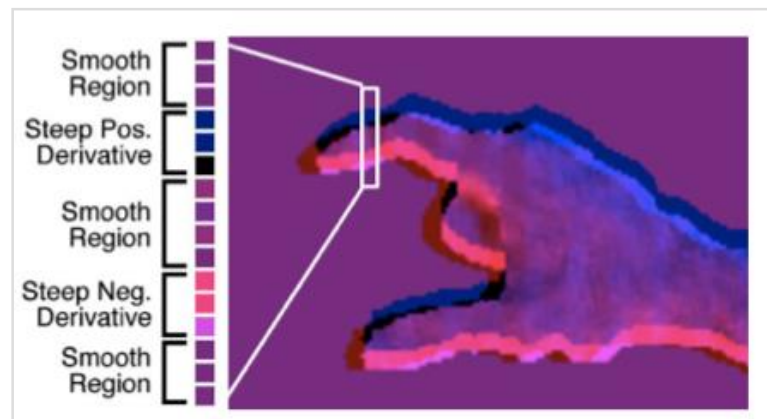
[Kim *et al.* UIST'12]



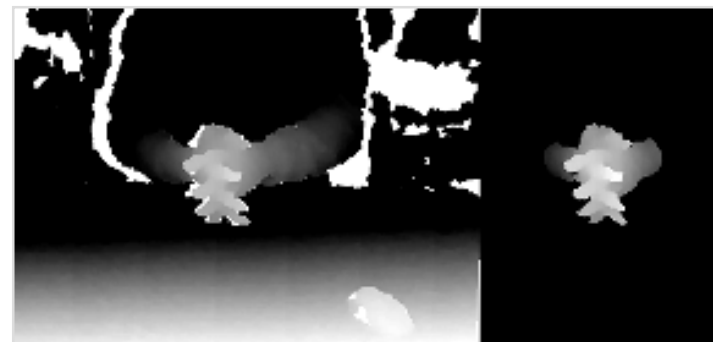
[Wang *et al.*, UIST '11]



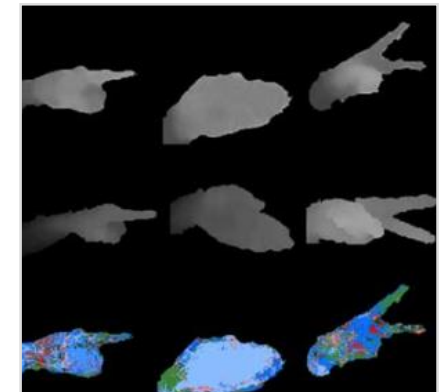
[Harrison *et al.* UIST'11]



[Oikonomidis *et al.* UIST'11]



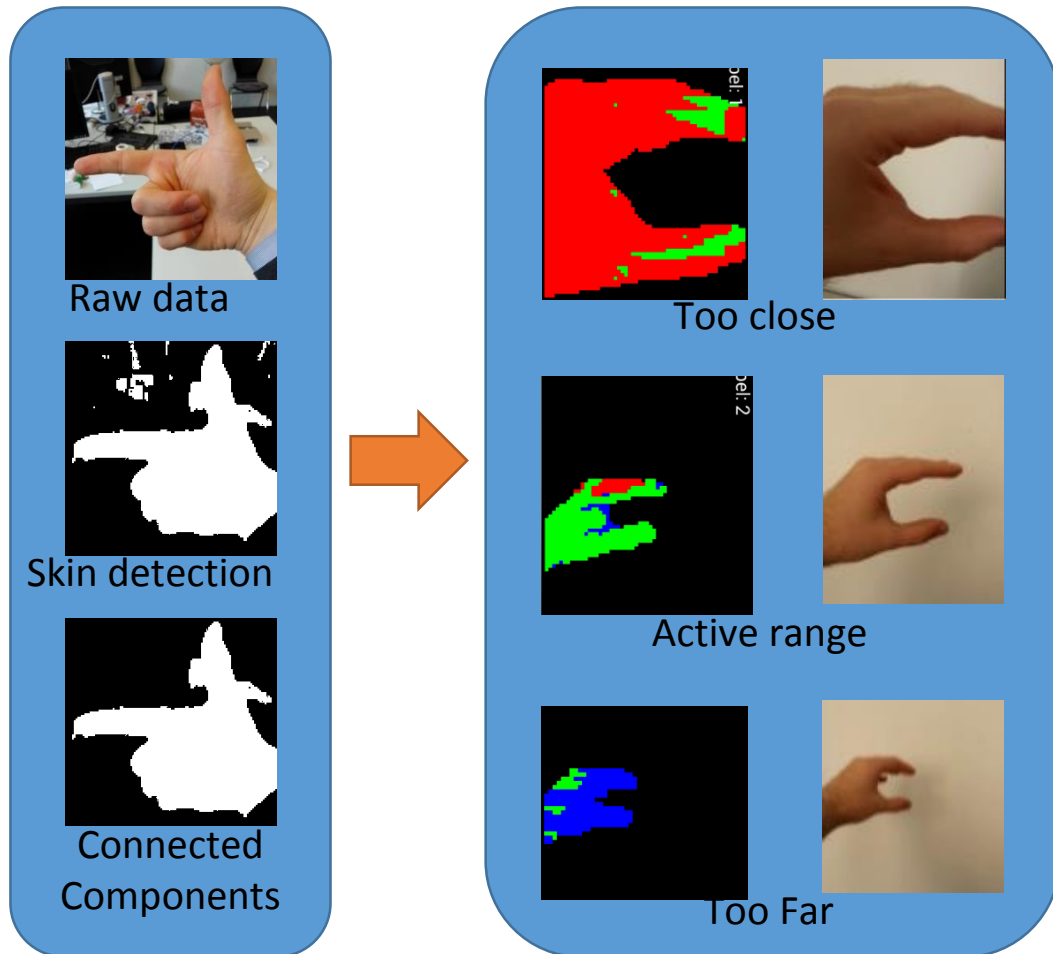
[Keskin *et al.* CPVR'12]



The Pipeline

Hand Segmentation

Coarse Depth Classification



Depth Classification Forest (DCF)

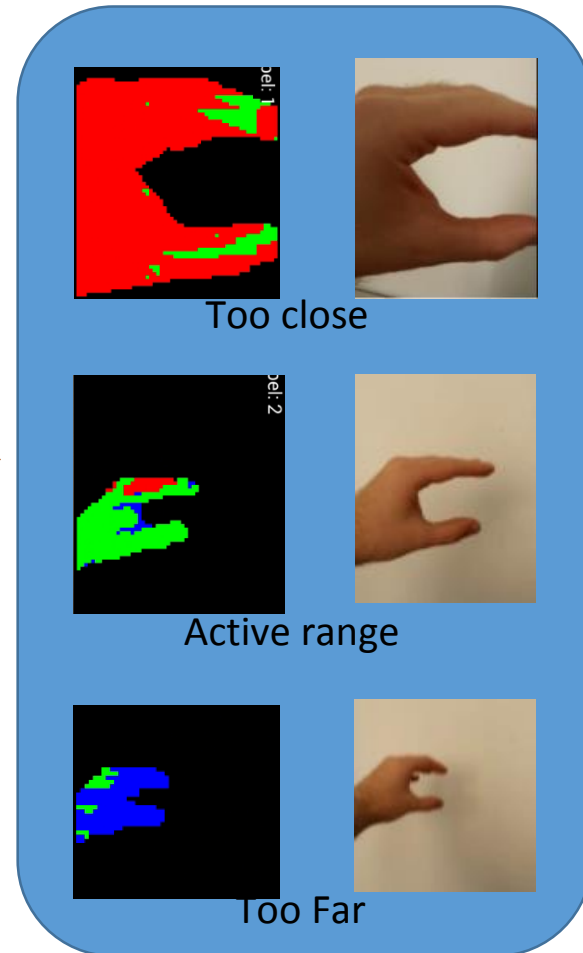
COARSE
DEPTH ESTIMATION

The Pipeline

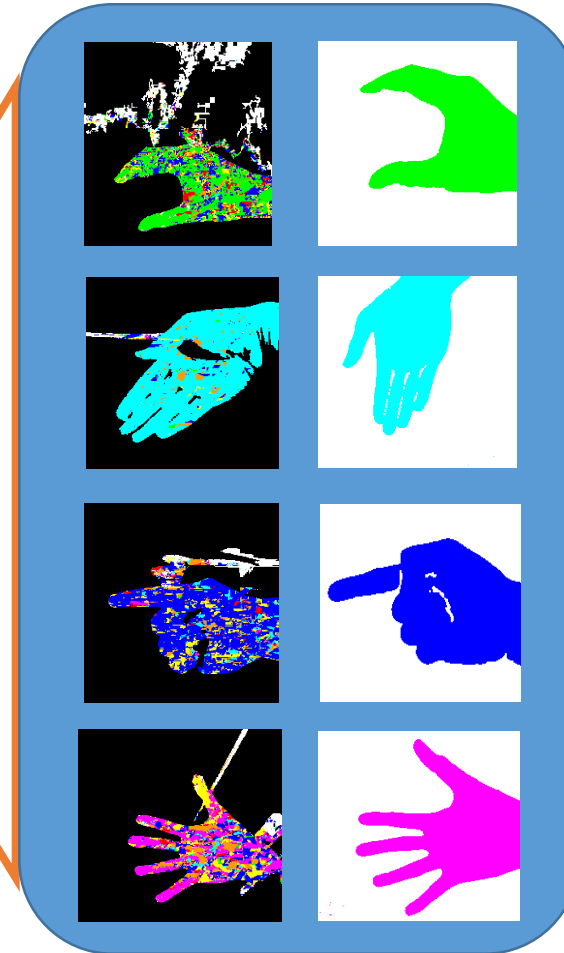
Hand Segmentation



Coarse Depth Classification



Shape Classification



Shape Classification Forest (SCF)

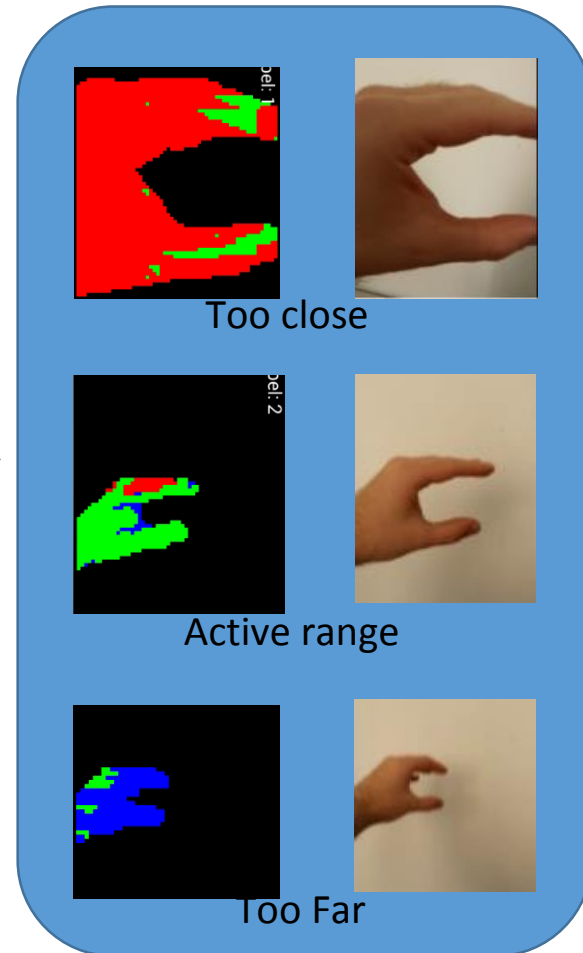
SHAPE
CLASSIFICATION

The Pipeline

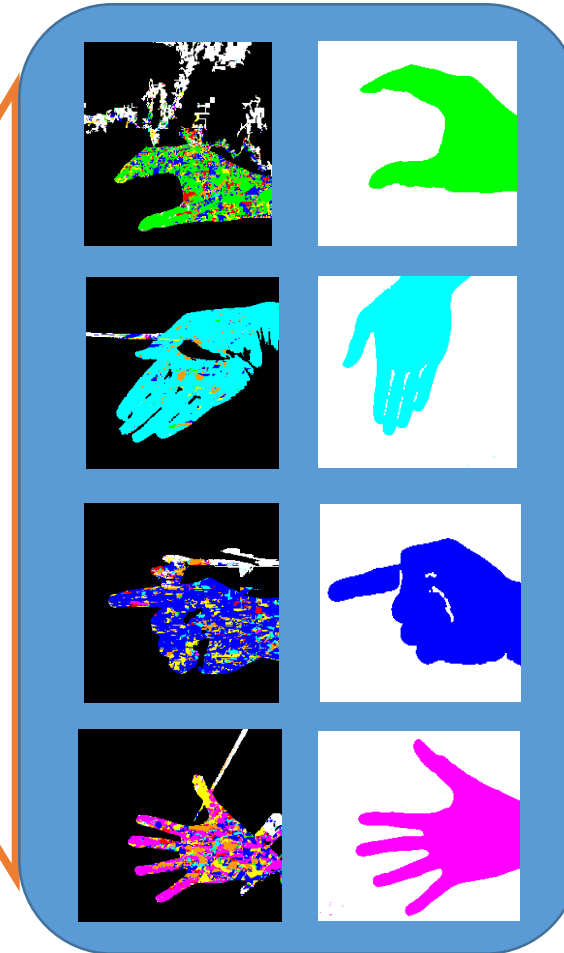
Hand Segmentation



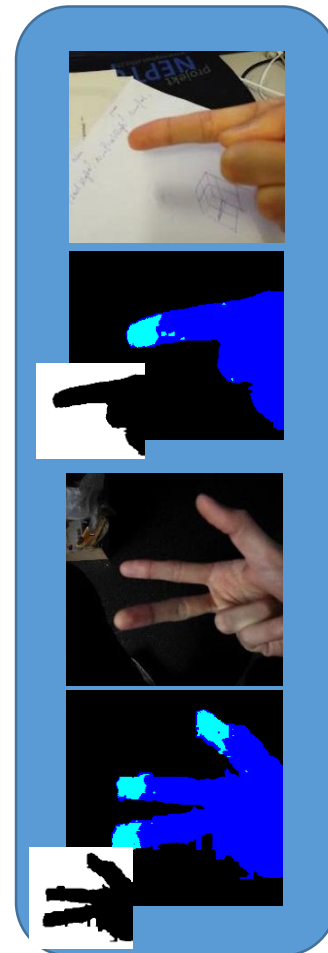
Coarse Depth Classification



Shape Classification



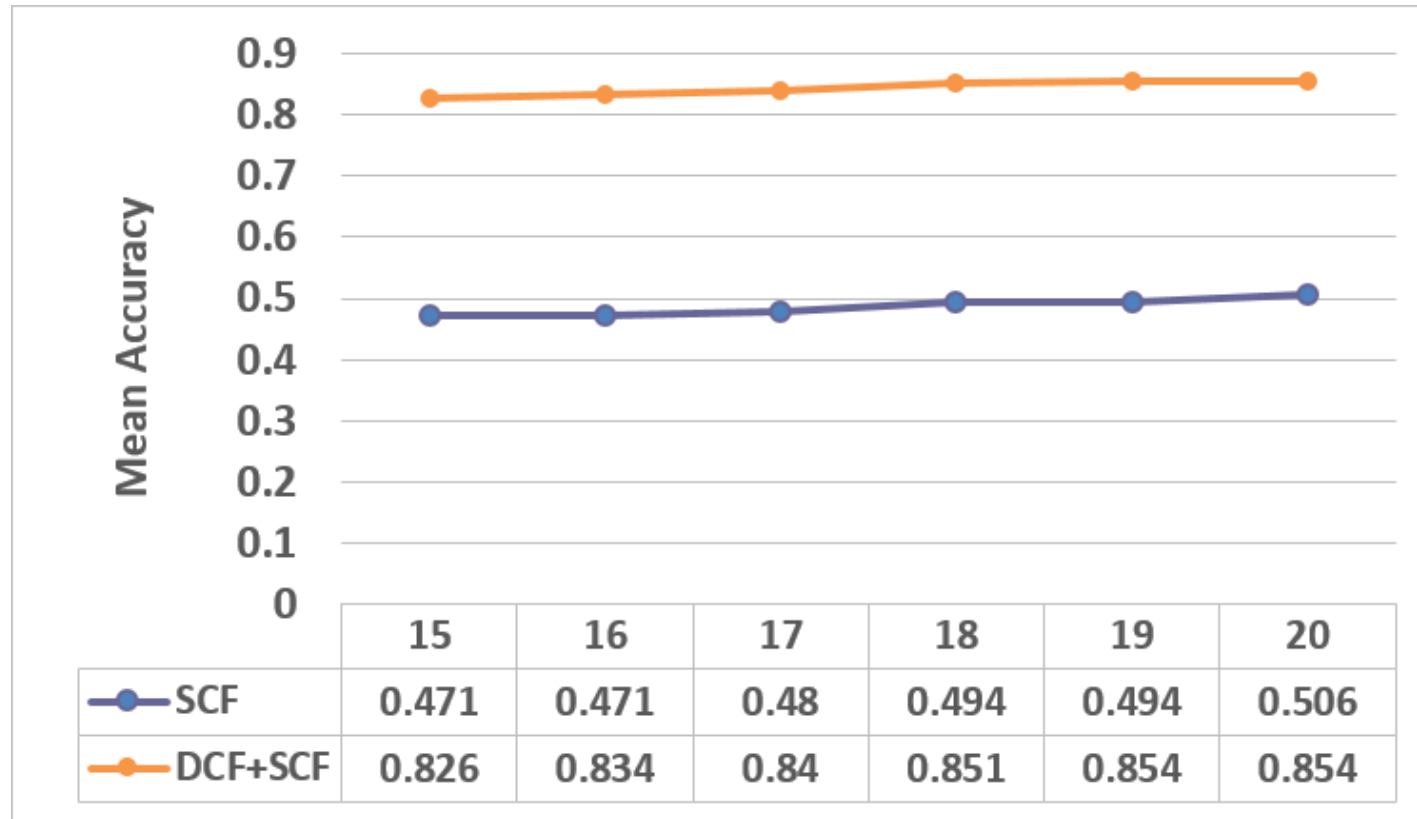
Part Classification



Part Classification Forest (PCF)

PART
CLASSIFICATION

DCF+SCF vs SCF Only



DCF+SCFs, Depth 5+10, 4.5MB, 85%
SCFs, Depth 15, 110MB, 50%

Confusion Matrix

<i>Pinch Open</i>	0.88	0.03	0.0	0.0	0.0	0.0	0.02
<i>Pinch Close</i>	0.0	0.93	0.05	0.0	0.0	0.0	0.01
<i>Pointing</i>	0.02	0.01	0.9	0.04	0.0	0.0	0.01
<i>Gun</i>	0.0	0.0	0.02	0.95	0.0	0.0	0.0
<i>Splayed Hand</i>	0.0	0.0	0.0	0.01	0.99	0.0	0.0
<i>Flat Hand</i>	0.05	0.0	0.0	0.0	0.01	0.99	0.11
<i>No-Gesture</i>	0.05	0.03	0.03	0.0	0.0	0.01	0.85

93% per-frame accuracy
Leave One Subject Out

Applications



Not Only Mobile Phones



Not Only Mobile Phones



Limitations

- Low-lighting still a problem
- Having more robust segmentation would help
- Gesture with non-unique contour are still a problem
- Mostly static gestures

Contributions

- Robust and real-time gesture recognition on **unmodified** mobile devices
- Extending multi-layered Random Forest framework for **low-memory devices**
- Variety of different mobile platforms, including **phones, tablets** and even **smart-watches**
- Variety of interaction techniques and scenarios where **in-air gestures complement touch**

Live Demo

Thank You!

In-air Gestures Around Unmodified Mobile Devices

Jie Song¹, Gabor Soros¹, Fabrizio Pece*,
Sean Fanello², Shahram Izadi²,
Cem Keskin², Otmar Hilliges¹

¹ETH Zurich

²Microsoft Research

{jsong|gabor.soros|fabrizio.pece|otmar.hilliges}@inf.ethz.ch
{seanfa|shahrami|cemke}@microsoft.com