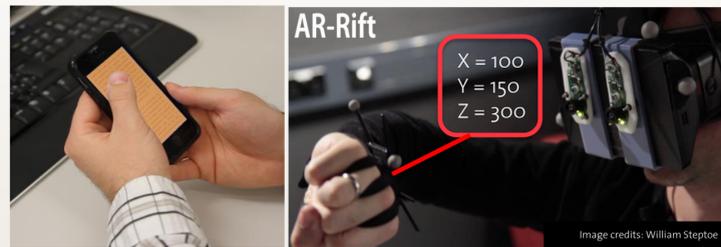


## INTRODUCTION

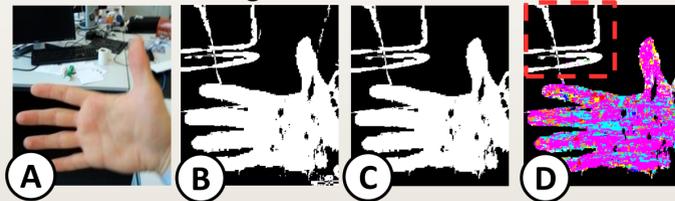
We present a novel algorithm that performs real-time hand gesture classification and hand depth regression, using *only* an RGB camera.



Our algorithm extends the interaction space from the touch screens to around the mobile devices

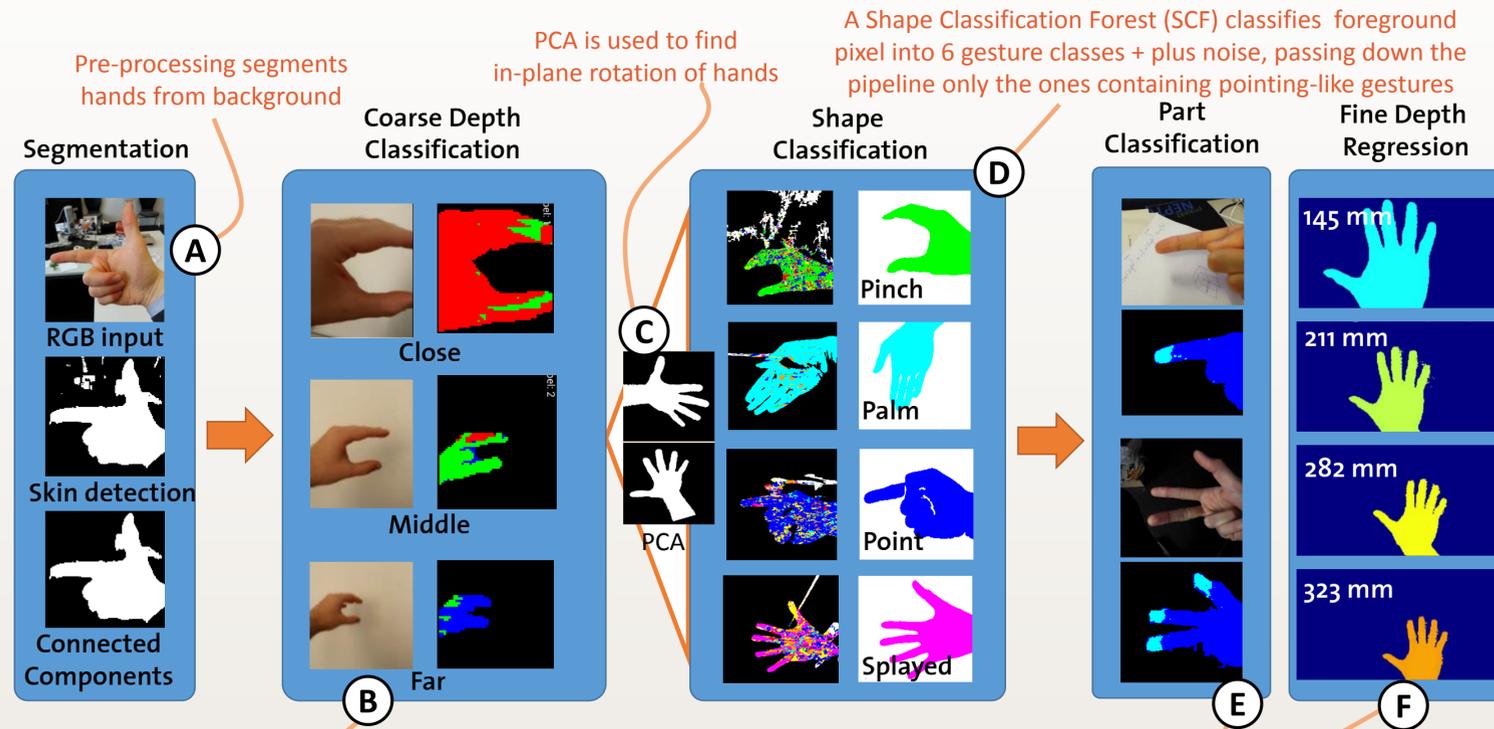
## ADVANTAGES

- runs in real time on unmodified smartphones, smartwatches, smartglasses
- has significantly lower memory footprint than traditional random forest approaches
- offers great flexibility in the gesture set
- is robust to segmentation errors



## CLASSIFICATION-REGRESSION PIPELINE

Our technique implements a multi-layer random forest (RF) architecture, which consists of established image processing steps interwoven with a new, staged classification-regression pipeline



A Depth Classification Forest (DCF) coarsely estimates the depth of the hand, greatly reducing the variation in depth of frames which are forwarded

A Part Classification Forest (PCF) classifies location of fingertips for pointing-like gestures. Alternatively, a Depth Regression Forest (DRF) estimates the distance of the hand from the camera.

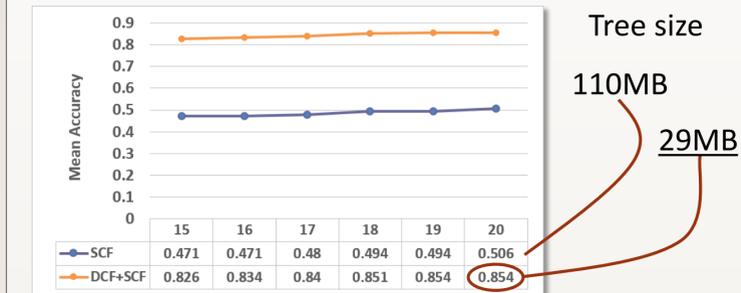
Each stage of the pipeline is trained to perform a different task. As a result, the input to each pipeline stage is modified by the previous stage, reducing the variation of the data passed to the subsequent stages

## RESULTS

PinchOpen	0.88	0.03	0	0	0	0	0.02	0.88	0.03	0	0	0	0	0.02
PinchClose	0	0.93	0.05	0	0	0	0.01	0	0.93	0.05	0	0	0	0.01
Pointing	0.02	0.01	0.9	0.04	0	0	0.01	0.02	0.01	0.9	0.04	0	0	0.01
Gun	0	0	0.02	0.95	0	0	0	0	0	0.02	0.95	0	0	0
SplayedHand	0	0	0	0.01	0.99	0	0	0	0	0	0.01	0.99	0	0
FlatHand	0.05	0	0	0	0.01	0.99	0.11	0.05	0	0	0	0.01	0.99	0.11
No-Gesture	0.05	0.03	0.03	0	0	0.01	0.85	0.05	0.03	0.03	0	0	0.01	0.85

Confusion matrix(20 users).

Left: half-test/half-train cross-validation; avg. accuracy 98% Right: leave-one-subject-out; avg. per-frame accuracy 93%.



Classification accuracy as function of architecture when extreme gesture depth variations occur. DCF+SCF (orange) outperforms single SCF (blue) in terms of accuracy averaged over all classes.



Depth regression accuracy. Green: ground truth, Orange: our method, Blue: naive pixel counting

## IN-AIR GESTURES ON THE MOVE



For more information, please visit our homepage  
<http://ait.inf.ethz.ch/projects/2014/InAirGesture/>  
<http://ait.inf.ethz.ch/projects/2015/3Dfrom2D/>

