Hand Gesture Recognition in Mobile Devices: Enhancing the Music Experience

Oriol Nieto
Music and Audio Research Lab (MARL)
oriol@nyu.edu

Dennis Shasha
Department of Computer Science
shasha@courant.nyu.edu

MARL
NYU Music and Audio Research Laboratory

NEW YORK UNIVERSITY

Summary

‣**What**: Efficient real-time hand gesture recognition for mobile devices.

‣**How**: Using the built-in front camera, applying a set of computer vision methods and heuristics.

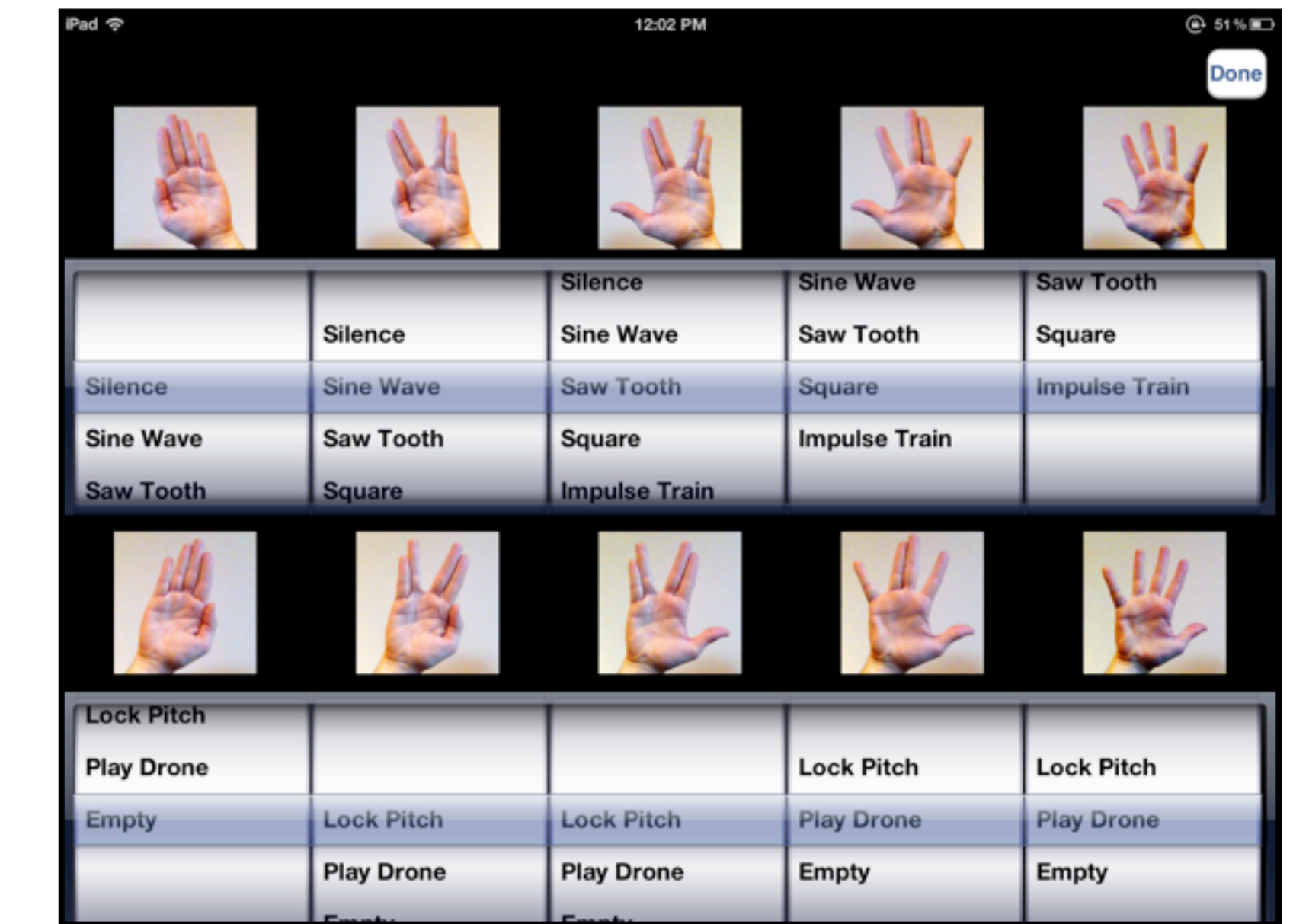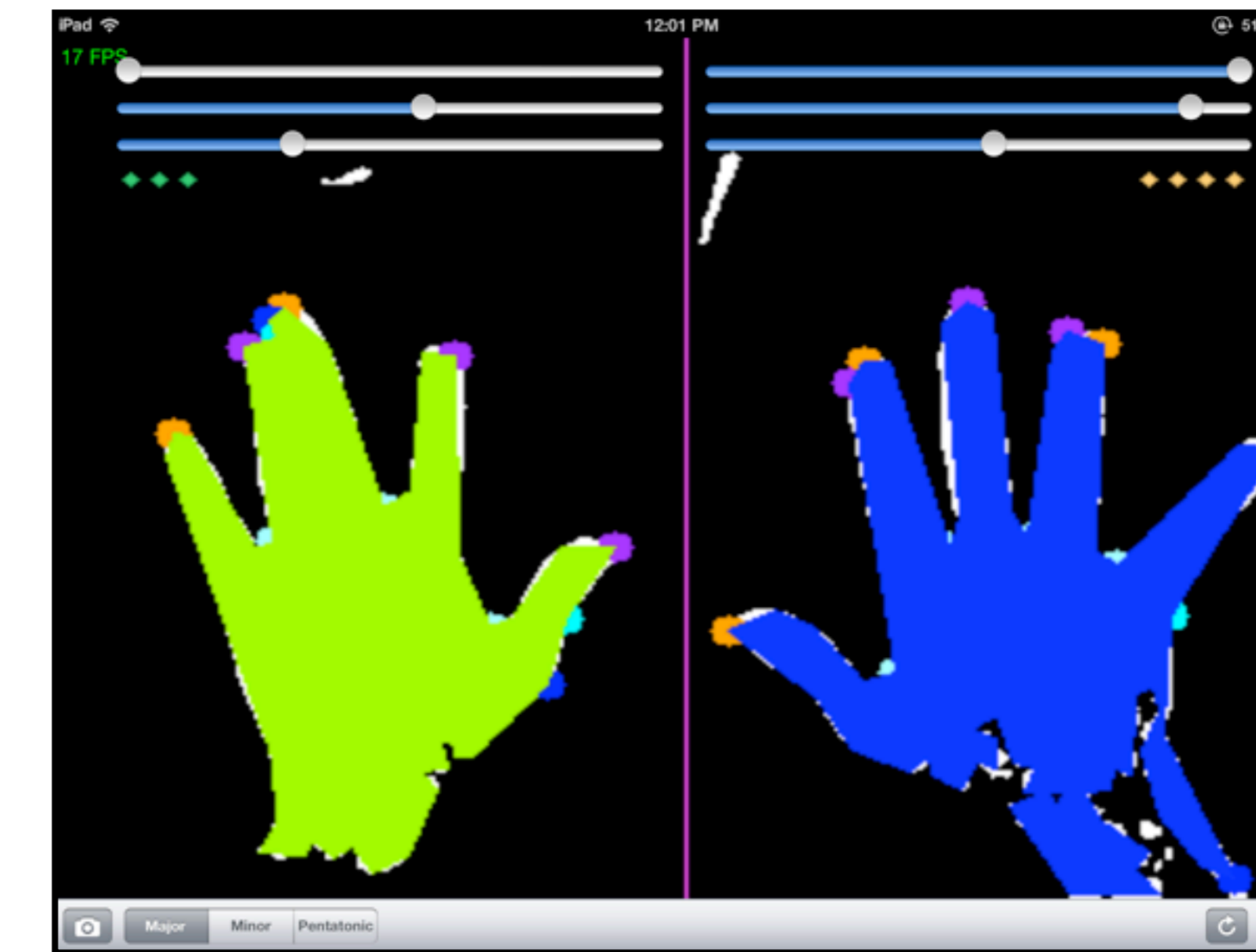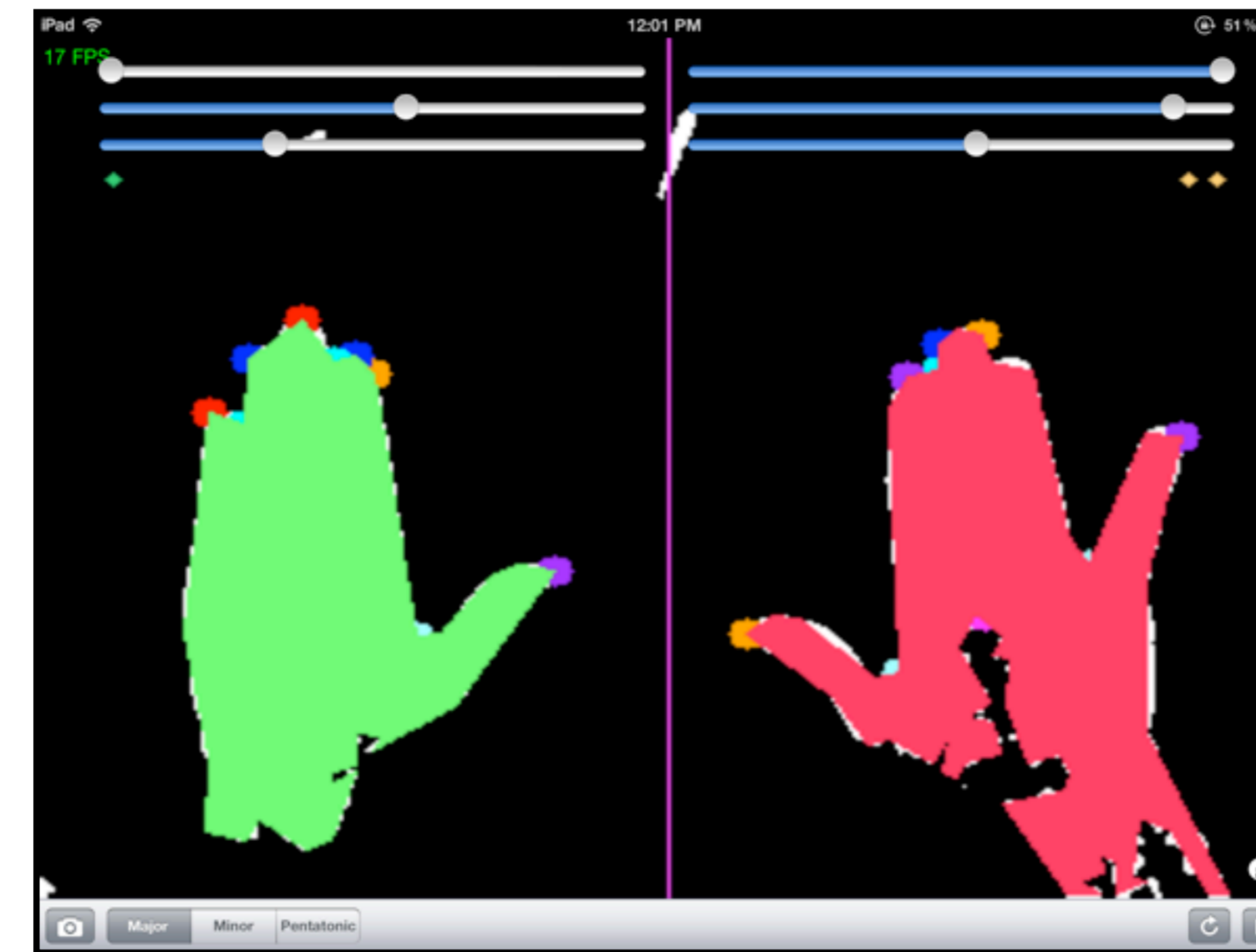‣**Why**: Control your mobile devices from the distance to, e.g. generate music, play video games.

iPad App Prototype: AirSynth

‣**Audio Synthesizer** controlled by the proposed technology.

‣Similar to a **Theremin**.

‣**Implementation:** iPad application compatible with iOS 6.0 and higher

‣Each gesture can be associated to a user-defined sound file or action.

Detecting Hands and Gestures

‣Defect Points Acquisition

‣Apply a skin filter to a YCbCr image, apply a blur & high contrast non-linearity transforms and gray out.

‣Detect hands by finding the two largest areas of the detected color in the screen (given a minimum threshold θ).

‣Approximate the areas with polygons, and obtain the convex hull for each hand. The defect points will be the ones that define the triangle areas between the convex-hull and the approximated polygon (see Figure (f)).

‣Gestures Recognition

‣Goal: recognize **alphabet of gestures** defined in the bottom Figure.

‣Compute the average euclidean distance D between the starting and ending points to the depth point for each defect area.
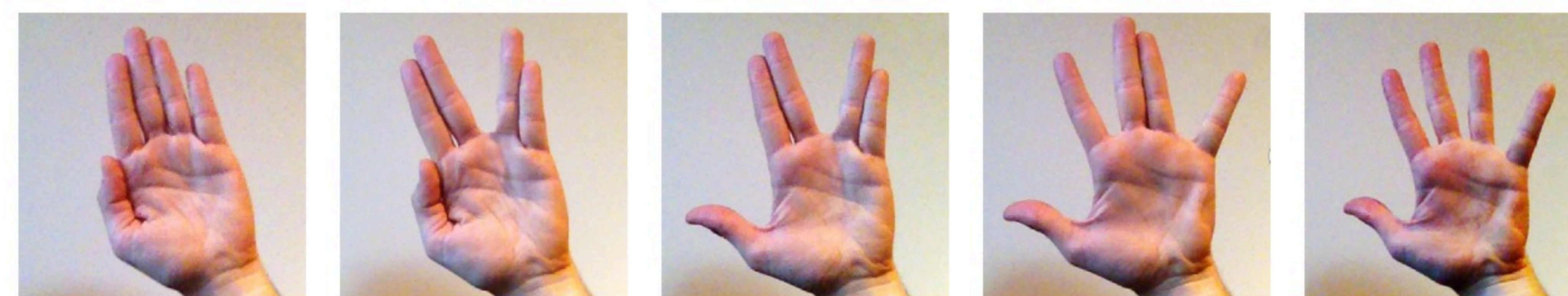
$$D_i = \frac{||P_s^i - P_d^i||_2 + ||P_e^i - P_d^i||_2}{2}$$

‣To detect the separation between two stretched fingers (i.e. a possible finger *web*) we check whether D is greater than a certain threshold η:

$$\eta_j = \frac{\sqrt{H_j}}{5}$$

‣Finally, to consider this area as a finger *web*, we have to also check that the vertical dimensions of these points are greater than its depth point:

$$D_i > \eta_j \wedge y_{P_{d_j}^i} < y_{P_{s_j}^i} \wedge y_{P_{d_j}^i} < y_{P_{e_j}^i}$$

X

YCbCr

Cb_max
Cb_min
Cr_max
Cr_min

Skin Filter

Y_max
Y_min

Blur & High Contrast

Black & White

Find Biggest Areas

θ

for each hand

Approx. Polygon

v

Convex-Hull

Defect Areas

(a) No finger webs (b) One finger web (c) Two finger webs (d) Three finger webs (e) Four finger webs

(a) Input Image (b) Skin Filter

(c) Non-linear Transforms (d) Black & White
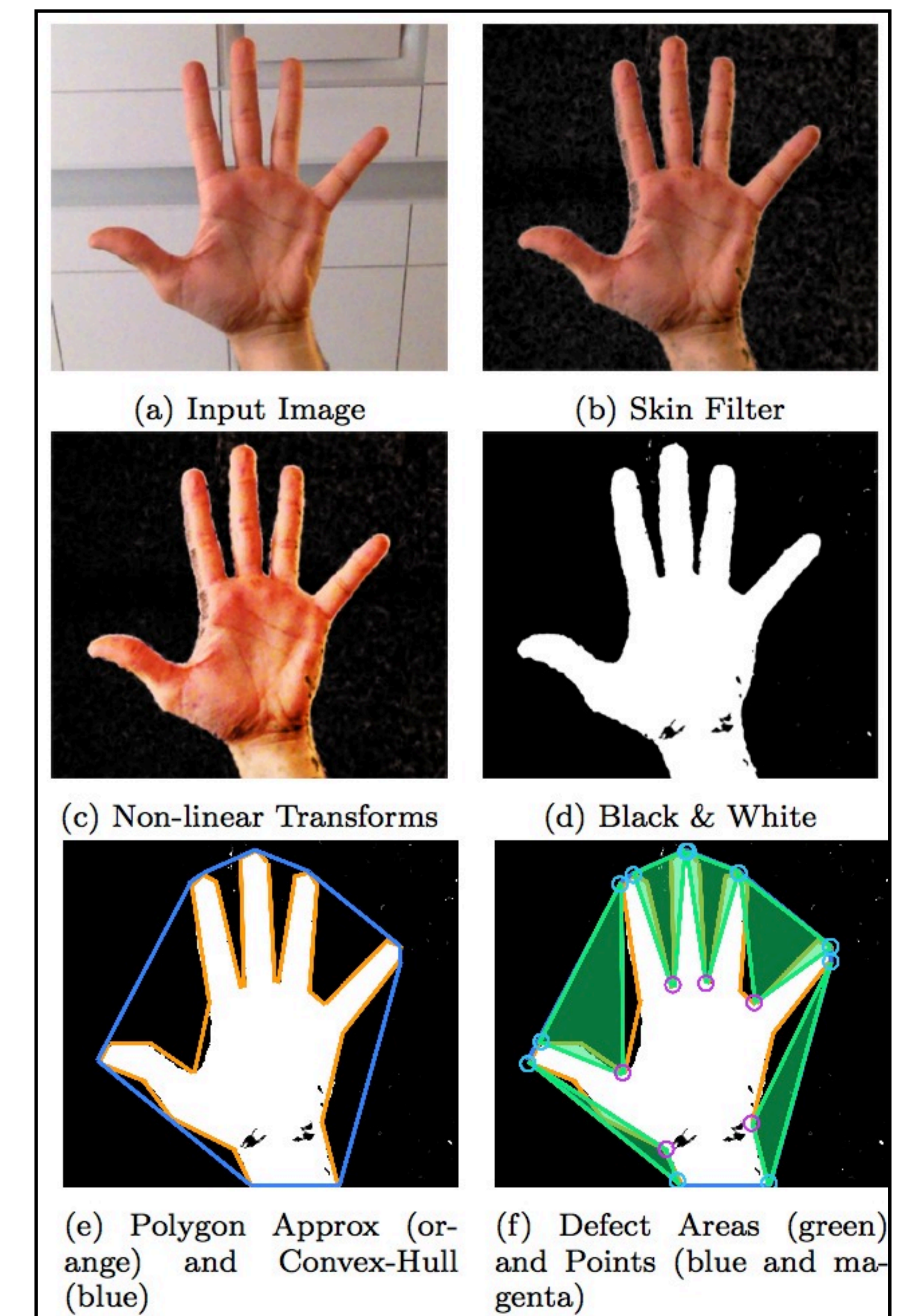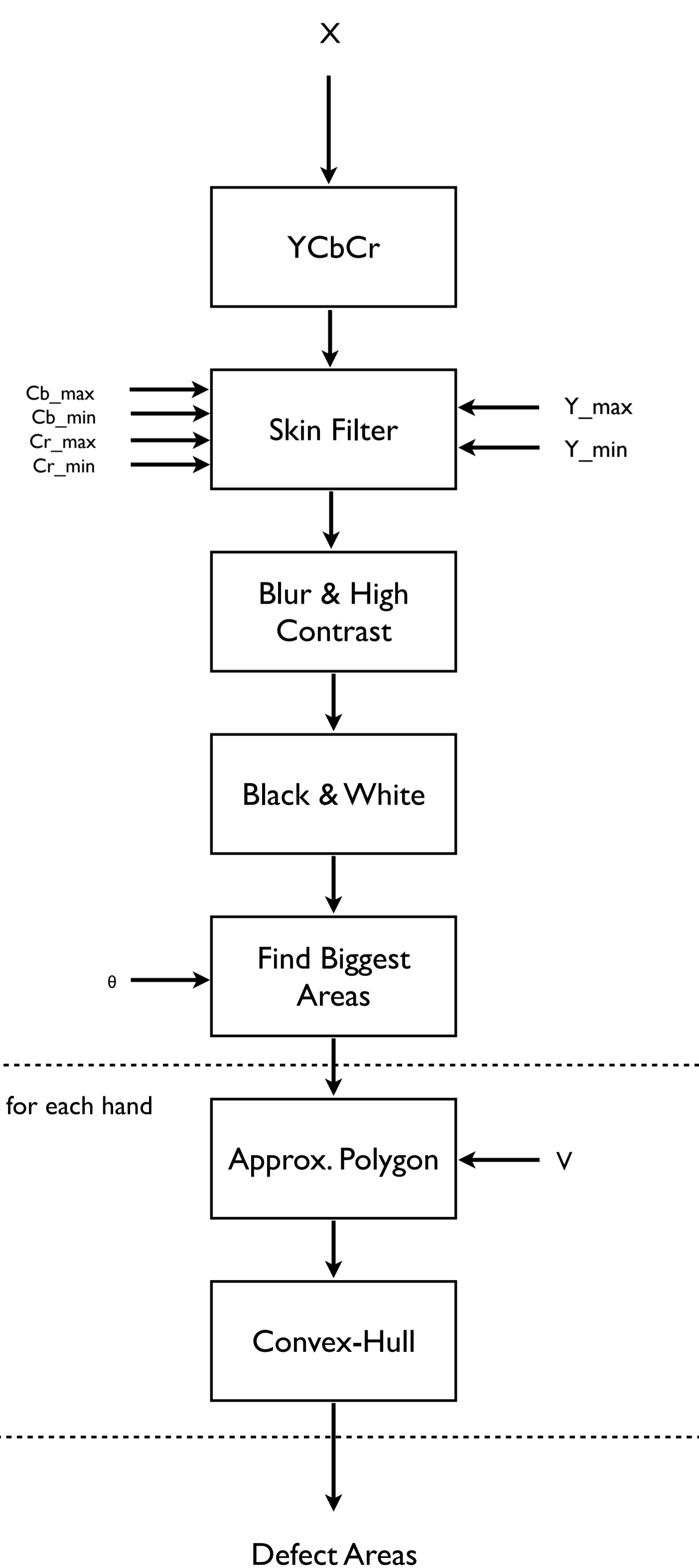
(e) Polygon Approx (orange) and Convex-Hull (blue) (f) Defect Areas (green) and Points (blue and magenta)