Hand Gesture Recognition by using Logical Heuristics

Pulkit Kathuria Atsuo Yoshitaka

Japan Advanced Institute of Science and Technology
School of Information Science

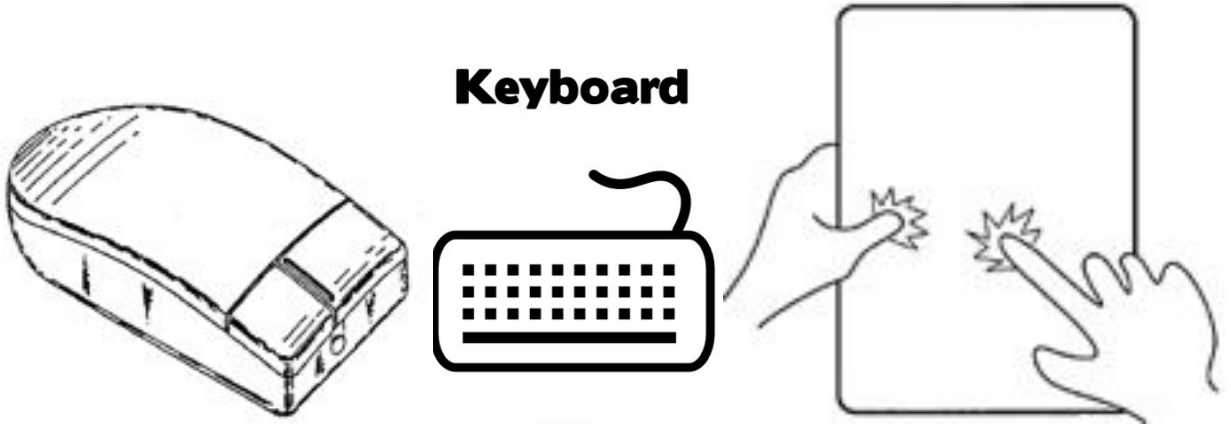
SIG, Human Computer Interaction-147, 2012

Outline

Outline

Motivation

Existing Computer input Interfaces

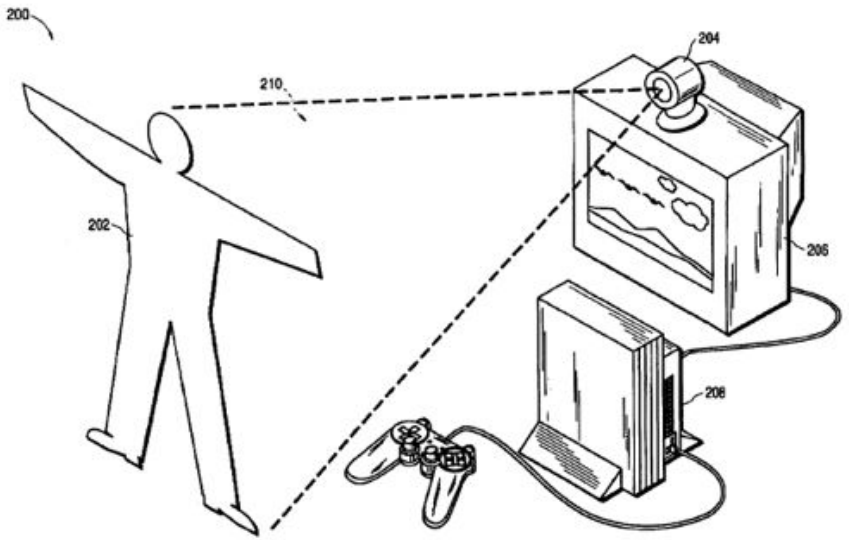


Efficient **BUT** require

- Additional equipped hardware, hence extra space
- A Physical Touch

Motivation

Emerging Category: Input by Gesture Recognition



Natural HCI **BUT** not efficient because of

- Different types of gestures in complex backgrounds
- Complicated computational procedures

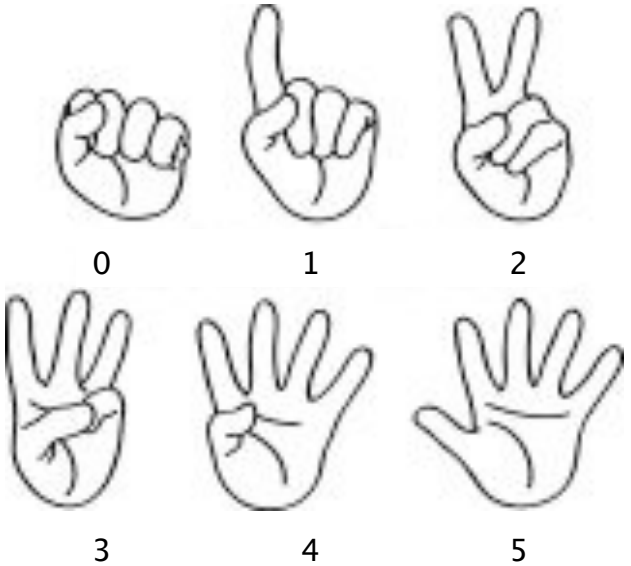
In This Paper

Aim: Natural HCI by focussing on

- Recognizing gestures in complex backgrounds
- Fast processing on standard PC and cameras

Recognize 6 different hand gestures

- No Training phase
- Simple and Fast Algorithm

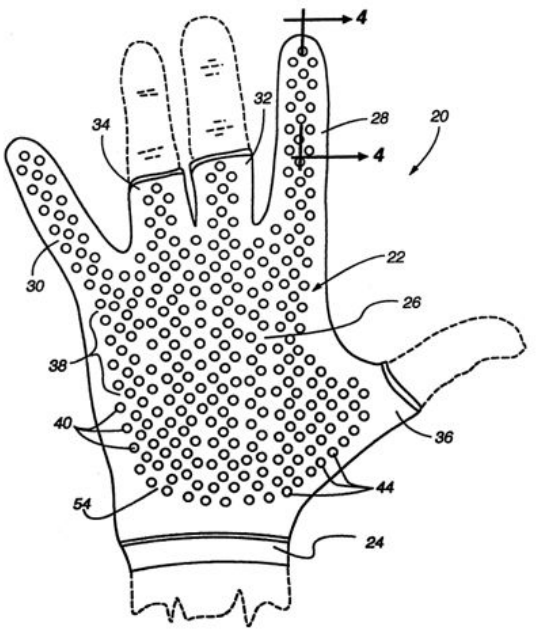


Outline

Existing Approaches

Wearable Data Gloves (Pandit et al., 2009)

- Requires data gloves with markers to be worn to extract hand posture
  - Efficient BUT not a suitable interface
  - Complex calculations



Existing Approaches

Template Matching (Alon et al., 2005)(Stefan et al., 2008)

- Requires sets of huge training database of desired gestures to be recognized
  - Lack in response time
  - Prepare new training data for new gestures

Computer Vision Based

- 3D hand modeling
  - Uses many approximation processes
  - Complicated

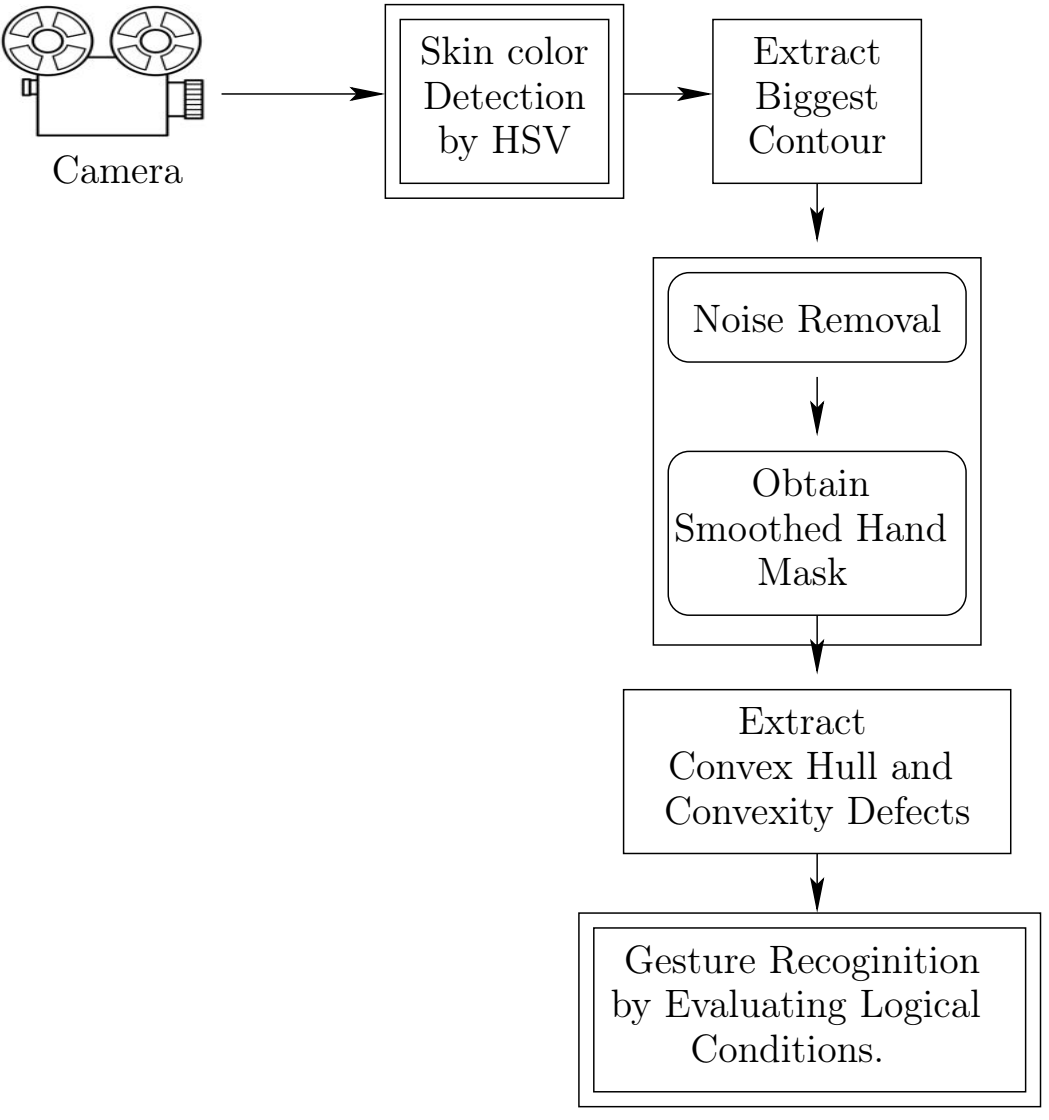
Our Approach

Computer Vision Based

- Suitable Interface
  - No additional wearable equipment
  - *Uses a standard digital camera*
- Computationally simple & fast
  - No training phase
  - *Uses Logical Conditions*
- Recognizes Gestures in
  - Static Conditions
    - *Background remains same (images)*
  - Dynamic
    - *Background lightening, colors etc change (videos)*

Outline

Overview: Block Diagram



- 2 Major Steps
  1. Skin Color Detection
  2. Gesture Recognition

Outline

Skin Color Detection

Used *HSV* (**H**ue, **S**aturation, **V**alue) color scheme

- More related to human perception (Albiol et al., 2001)
- Used classical method to detect skin pixels
  - By setting Upper & Lower bound values

$$H_{min} \leq H \leq H_{max} \begin{cases} H_{min} & 0^\circ \\ H_{max} & 20^\circ \end{cases} \tag{1}$$

$$S_{min} \leq S \leq S_{max} \begin{cases} S_{min} & 45 \\ S_{max} & 255 \end{cases} \tag{2}$$

Skin Color Detection

- Left Side (Colored)
  - Original Image
  - Houses color is similar to skin color
- Right Side (Black & White)
  - **Black**: non skin color
  - **White** dots: skin color

Skin Color Detection

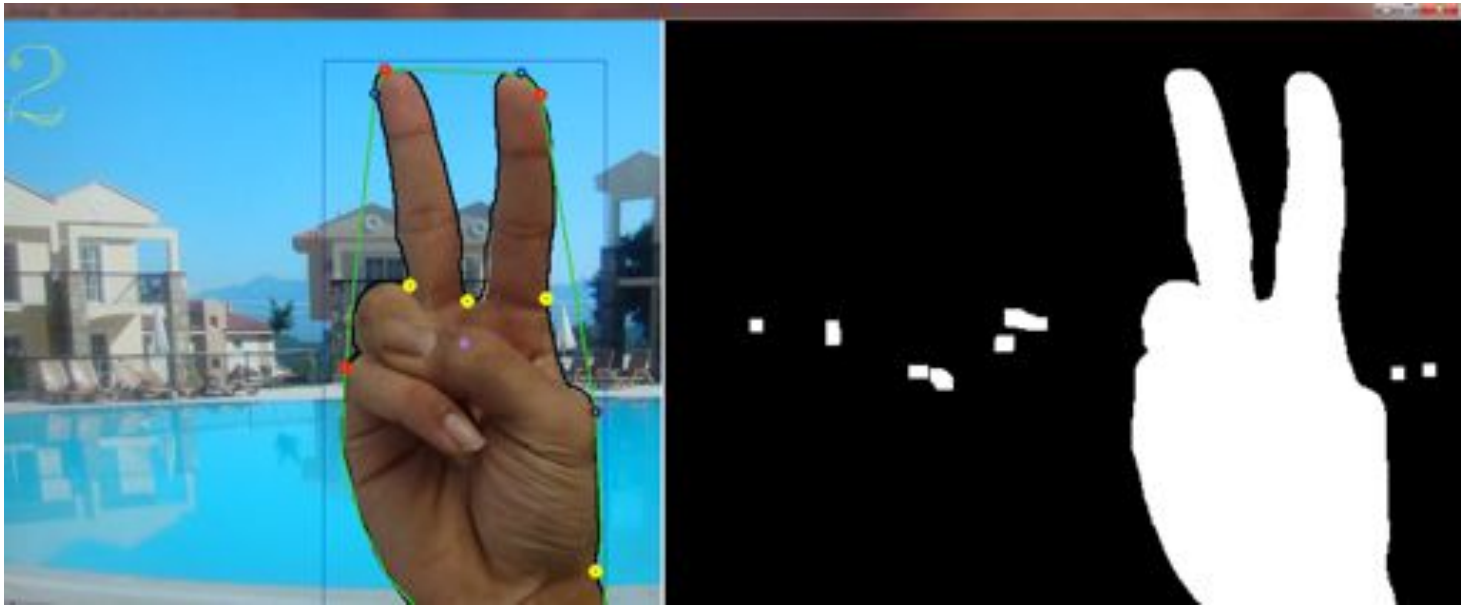
Naive & Robust BUT

- Classifies noisy objects as skin

Reduce Noise

- Assumption:
  - Biggest connected white dots area is of Human Hand

Noise Reduction & Hand Detection



Reduce Noise

1. Take median of **H** & **S** values for each pixel of biggest contour (the hand)
2. Compare color values of extracted skin objects whose:
   - *area* $< 20\%$ *of the area of biggest contour*
3. Smoothed Skin Pixel Binary Mask of a hand is obtained

Outline

Hand Gesture Recognition

6 Gesture Types

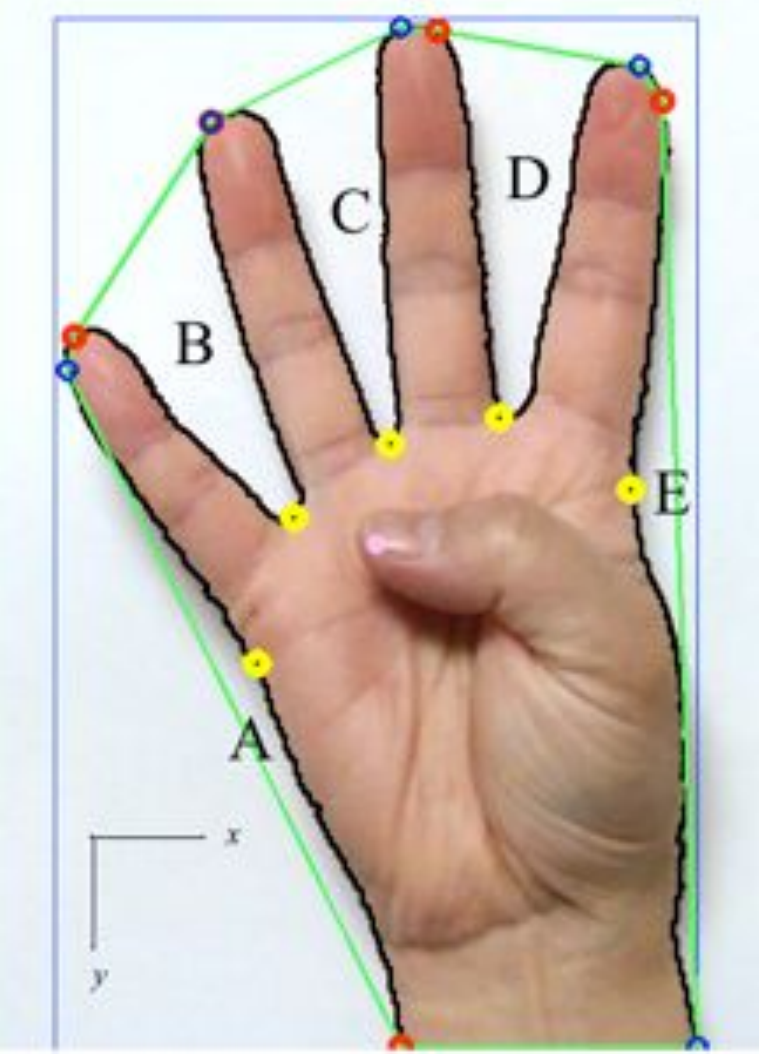
- Counting Number of fingers in a Hand *(zero - five)*
  - *zero* gesture of a closed hand

Gestures Recognition by evaluating Logical Conditions derived using

- Convex Hull
- Convexity Defects

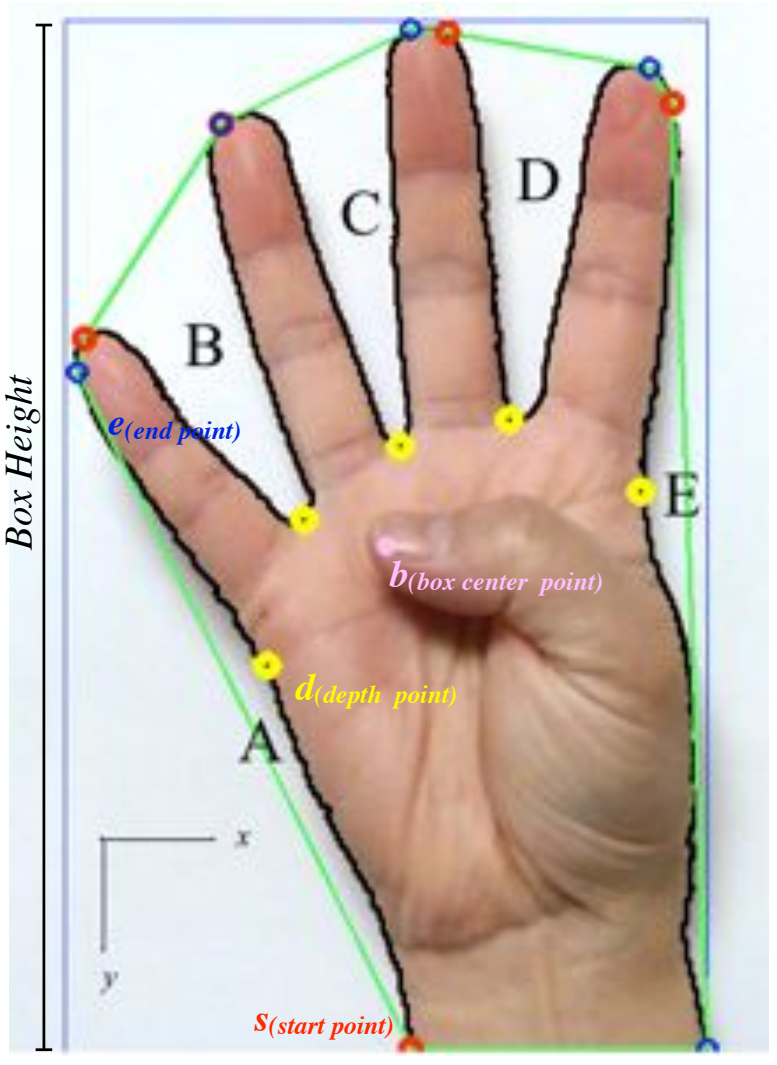
Convex Hull & Convexity Defects

- Convex Hull
  - Outer green line around the hand
- Convexity Defects
  - Holes in difference b/w convex hull & hand
  - Five convexity defects in picture
    - **A** through **E** bounded by black & green outlines



Information from Convexity Defects

- For each Convexity defect
  - Start Point s_x
  - Depth Point d_y
  - Box Center Point b_y
  - Length of Defect l_d
  - Box Height



Heuristics towards Gestures Recognition

For each Convexity defect c_d

$$\text{count} = \begin{cases} 1 & \textbf{if } (s_y < b_y \textbf{ or } d_y < b_y) \textbf{ and} \\ & (s_y < d_y) \textbf{ and} \\ & l_d > \frac{\text{box height}}{n} \\ 0 & \textbf{otherwise} \end{cases} \tag{3}$$

$$l_d = \sqrt{(s_x - d_x)^2 + (s_y - d_y)^2} \tag{4}$$

$$\textit{num of fingers} = \sum_{c_d \in C_D} \text{count} \tag{5}$$

- In Equation (3)
  - **First condition:** Checks if convexity defect is of a straight finger
  - **Third condition:** To filter momentary convexity defects caused due to hand rotation

Recognized Gestures of Different Forms

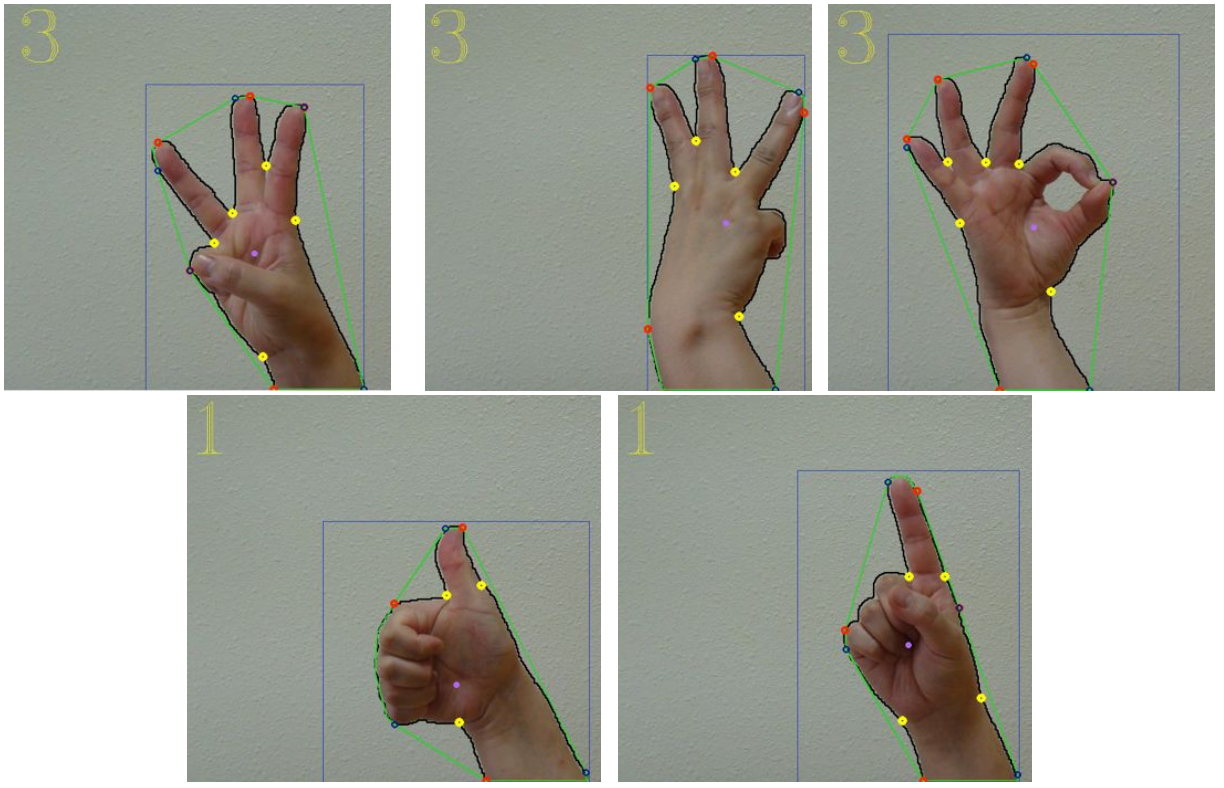


Figure: Different Forms of Three & One Counts

Recognized Gestures in Different Orientation

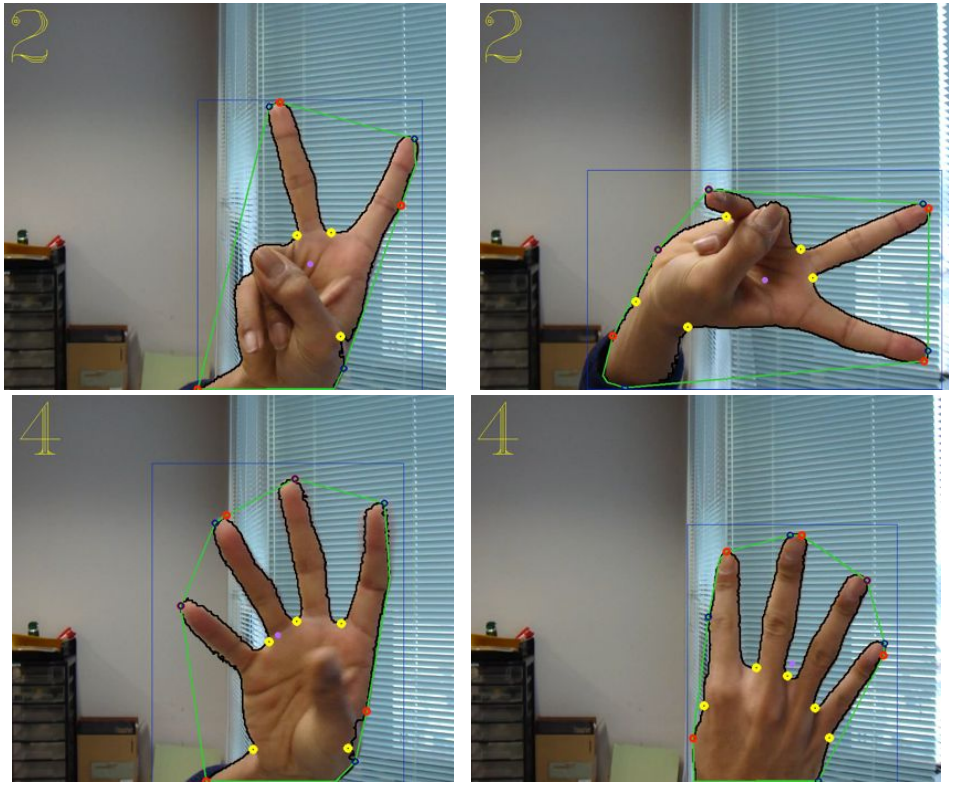


Figure: Counts of Two & Four in Different Orientation in Complex background

Outline

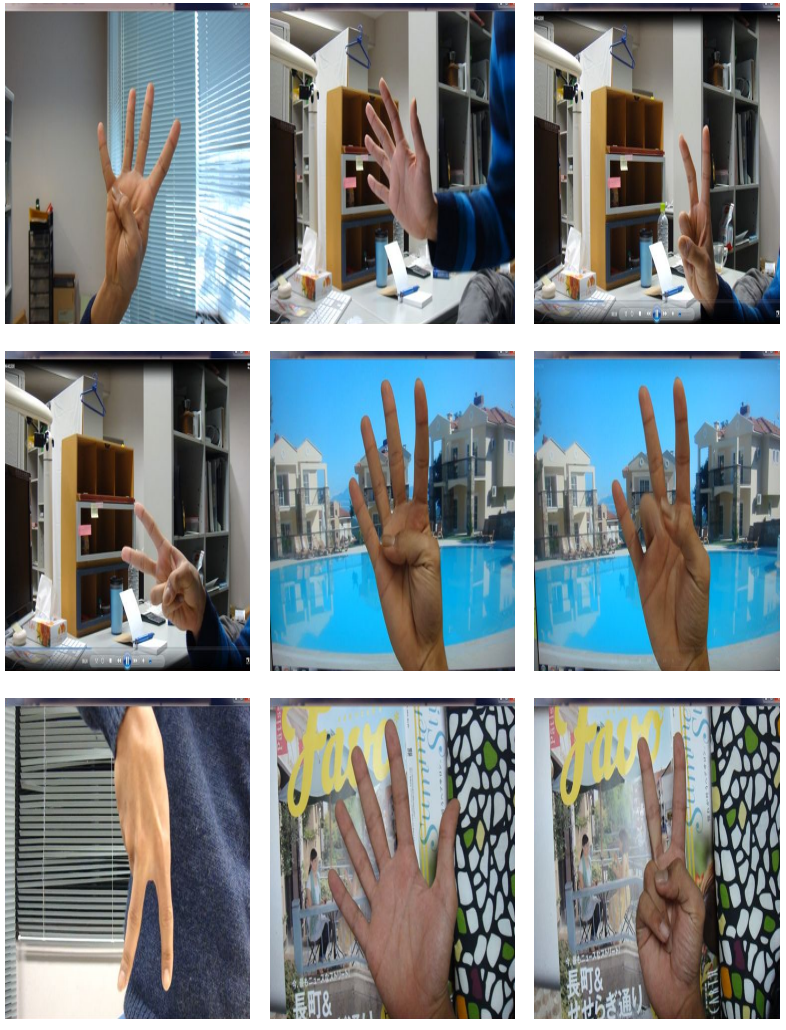
Data

- Development Data D_d
  - To Design Heuristics
- Evaluation Data D_e
  - To Evaluate Performance

Data	Total num. of Images	Images in Complex Backgrounds
D_d	60	20
D_e	90	60
Total	150	80

- Front Facing Camera
  - 8 MegaPixels Sony Digital
  - 1280x720 Resolution
  - 29.970 FPS
- Each image consists of a gesture from one hand

Data Samples



Outline

Results on Development & Evaluation Data

G	Development D_d			Evaluation D_e			$D_d + D_e$		
	#I	P	R	#I	P	R	#I	P	R
0	10	0.70	0.70	15	0.64	0.47	25	0.67	0.56
1	10	0.75	0.60	15	0.55	0.40	25	0.63	0.48
2	10	0.78	0.70	15	0.70	0.47	25	0.74	0.56
3	10	0.89	0.80	15	0.73	0.53	25	0.80	0.64
4	10	1.00	0.90	15	0.75	0.60	25	0.86	0.72
5	10	1.00	0.90	15	0.82	0.60	25	0.90	0.72
Total	60	0.85	0.77	90	0.70	0.51	150	0.77	0.61

G=Gesture Type, I=Num of Images

$$P = \frac{\text{Num. of correctly detected}}{\text{Total num. detected}}, \quad R = \frac{\text{Num. of correctly detected}}{\text{Total num of corrects}}$$

Discussion

G	Development D_d			Evaluation D_e			$D_d + D_e$		
	#I	P	R	#I	P	R	#I	P	R
0	10	0.70	0.70	15	0.64	0.47	25	0.67	0.56
1	10	0.75	0.60	15	0.55	0.40	25	0.63	0.48
2	10	0.78	0.70	15	0.70	0.47	25	0.74	0.56
3	10	0.89	0.80	15	0.73	0.53	25	0.80	0.64
4	10	1.00	0.90	15	0.75	0.60	25	0.86	0.72
5	10	1.00	0.90	15	0.82	0.60	25	0.90	0.72
Total	60	**0.85**	**0.77**	90	**0.70**	**0.51**	150	0.77	0.61

- P & R on D_e are worse than D_d
  - Gesture Detection in D_e is more difficult because of more complex backgrounds

Data	**Total num. of Images**	**Images in Complex Backgrounds**
D_d	60	20
D_e	90	**60**

Discussion

G	Development D_d			Evaluation D_e			$D_d + D_e$		
	#I	P	R	#I	P	R	#I	P	R
0	10	0.70	0.70	15	0.64	0.47	25	0.67	0.56
1	10	0.75	0.60	15	0.55	0.40	25	0.63	0.48
2	10	0.78	0.70	15	0.70	0.47	25	0.74	0.56
3	10	0.89	0.80	15	0.73	0.53	25	0.80	0.64
4	10	**1.00**	**0.90**	15	**0.75**	**0.60**	25	**0.86**	**0.72**
5	10	**1.00**	**0.90**	15	**0.82**	**0.60**	25	**0.90**	**0.72**
Total	60	0.85	0.77	90	0.70	0.51	150	0.77	0.61

- P & R increases towards higher finger counts
  - Hand spread is much wider
    - Therefore, logical evaluation and hand detection are relatively accurate

Overall Performance

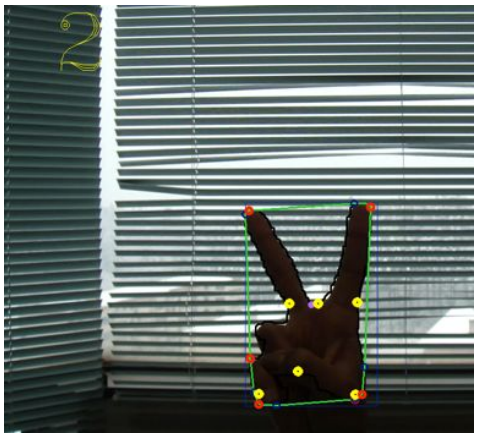
G	Development D_d			Evaluation D_e			$D_d + D_e$		
	#I	P	R	#I	P	R	#I	P	R
0	10	0.70	0.70	15	0.64	0.47	25	0.67	0.56
1	10	0.75	0.60	15	0.55	0.40	25	0.63	0.48
2	10	0.78	0.70	15	0.70	0.47	25	0.74	0.56
3	10	0.89	0.80	15	0.73	0.53	25	0.80	0.64
4	10	1.00	0.90	15	0.75	0.60	25	0.86	0.72
5	10	1.00	0.90	15	0.82	0.60	25	0.90	0.72
Total	60	0.85	0.77	90	0.70	0.51	**150**	**0.77**	**0.61**

- 77% Precision at 61% Recall from
  - 150 Gestures, 53% in complex/colorful backgrounds
- **Computational Cost**
  - OpenCV Library, Windows CPU 2 GB RAM
    - Usage $< 20\%$
  - 1 min HD video clip with 15 gestures took
    - 1 min 2 secs

Recognition in Different Conditions

- Colorful Background

- Poor Lightening

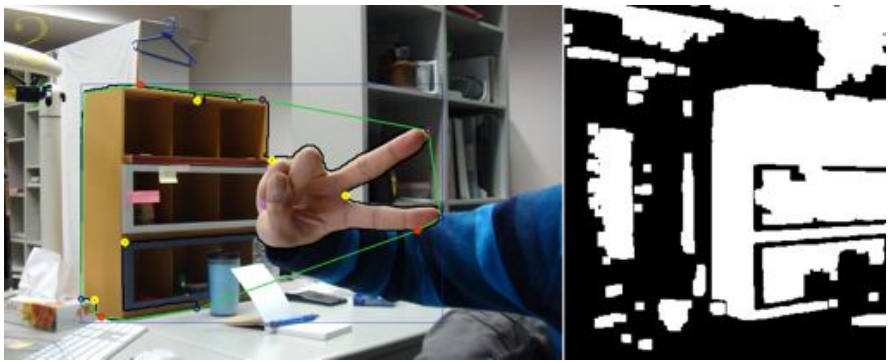


Sample Outputs

Difficulties with Skin Color & Hand Detection

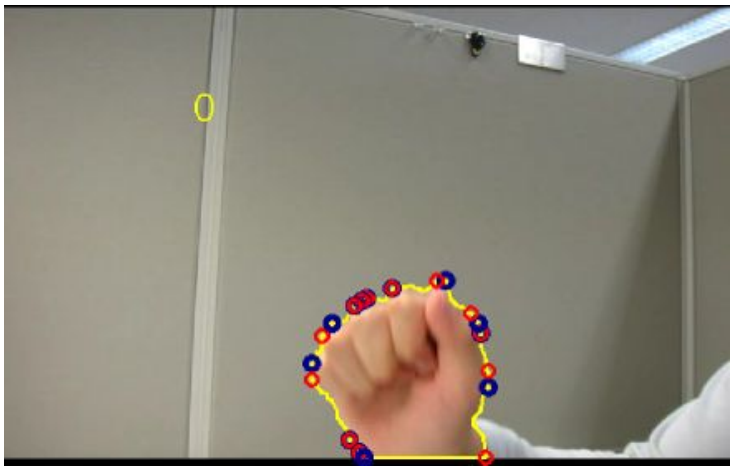
- Large Skin Color Objects

- Incorrect hand detection
  - Combined contour of wooden box and hand is extracted

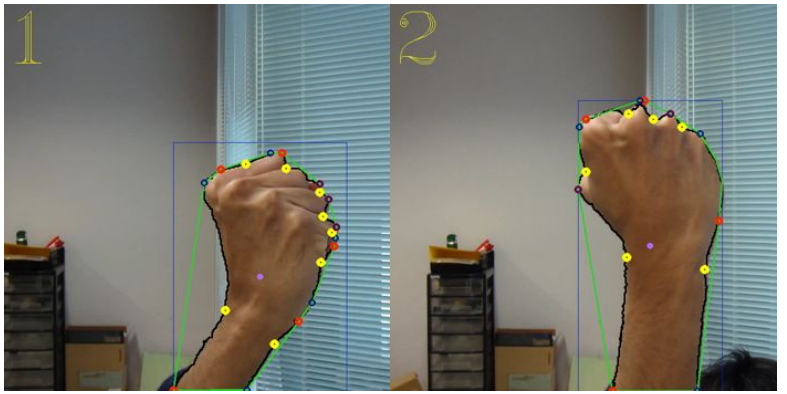


Difficulties with Logical Conditions

- Gesture *zero (closed hand)* count produces no convexity defects



- Incorrect gesture recognition of count *zero*
  - Inclusion of arm causing formation of convexity defects



Outline

Conclusion

- Proposed heuristics derived from convex hull and convexity defects to recognize hand gesture.
  - Computationally fast and simple
  - Recognizes 6 hand gestures in complex backgrounds
  - Orientation & Rotation Free
- **Future**
  - Better Hand Detection Technique to bring robustness to system
  - Derive robust logical conditions towards more complex gestures
    - Full Body Posters (Running, Walking)

Finish

Thank you for your attention.

References

- Alberto Albiol, Luis Torres and Delp, E., J.: Optimum color spaces for skin detection, 2001. In Proceedings of 2001 Image Processing International Conference, vol.1, pp.122-124 (2001).
- Jonathan Alon, Vassilis Athitsos, Quan Yuan, and Stan Sclaroff: Simultaneous localization and recognition of dynamic hand gestures. In Proceedings of IEEE Workshop on Motion and Video Computing, vol.2, pp.254-260 (Jan, 2005).
- Alexandra Stefan, Vassilis Athitsos, Jonathan Alon, and Stan Sclaroff: Translation and scale-invariant gesture recognition in complex scenes. In Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments, PETRA '08, vol. 7, pp. 1-8, New York, NY, USA (2008).
- Pandit, A. Dand, D. Mehta, Sabesan, S. Daftery, A.: "A Simple Wearable Hand Gesture Recognition Device Using iMEMS," Soft Computing and Pattern Recognition, 2009. SOCPAR '09. International Conference of , vol., no., pp.592-597, 4-7 (Dec. 2009)

Backup

- Focus on gestures from standard camera
  - Laptop
  - Smartphone etc
- Focus on robust heuristically define the gestures
- Object of interest
  - Hand usually takes up the largest area
- Delivers high precision in plain backgrounds
  - Needs work on complex backgrounds
- Manually annotated the gestures