Robust Localization of Ears by Feature Level Fusion and Context Information

Anika Pflug, Adrian Winterstein, Christoph Busch
Madrid, 5th of June 2013
Motivation

Outer Ear as a biometric characteristic

Face, finger, ear, eye, iris, retina, vein pattern, palm, gait, foot, exotic
Related Work

Ear Detection in 3D images

Detection rate

<table>
<thead>
<tr>
<th>Method</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prakash and Gupta</td>
<td>1604</td>
</tr>
<tr>
<td>Zhou et al.</td>
<td>142</td>
</tr>
<tr>
<td>Ansari and Gupta</td>
<td>700</td>
</tr>
<tr>
<td>Chen and Bhanu</td>
<td>312</td>
</tr>
<tr>
<td>Our approach</td>
<td>1777</td>
</tr>
</tbody>
</table>

88% 90% 92% 94% 96% 98% 100% 102%
Solved:
Ear detection using either texture or depth from full profile images.

Unsolved:
• Use 2D and 3D information for more accurate detection
• Include facial structures into the detection process
• Increase robustness against in-plane rotations and pose variation
Proposed System in a nutshell

- Texture image
- Edge image
- Selected edges
- Ear components
- Depth image
- Mean curvature after threshold
- Shapes
- Best ear candidate
- Ear with components
Core Ideas

Curvature, edges and proximity

- Calculate mean curvature and apply threshold
- Extract longest shapes
- Superimpose curvature image and edge image
- Select edges in proximity to shapes
Core Ideas

Modular scoring scheme

\[ S = \omega_1 \times I + \omega_2 \times R + \omega_3 \times C \]

- Does it look like an ear?
- How ear-like does it look in relation to other candidates?
- Is it located at a feasible position?

I = Individual Score
R = Relative Score
C = Context Score
Simple features for individual score

- Calculate convex hull of ear candidate
- Determine key points

→ Total slope from all pixels
→ Slope ratio from pixels on convex hull
→ Ratio between major and minor axis of enclosing ellipse
Core Ideas

Relative scoring System

Initial shape

- Length of ear candidate reduced by the number of moved and interpolated pixels
- Excludes candidates in regions with low number of shapes

→ The less adaptations, the better the candidate
• Calculate largest distance between two points on convex hull
• Calculate distance between line segment $\overline{p_1p_2}$

→ Ratio between h and d should not exceed an anatomically determined threshold
Overlap ratio calculation

\[ O = \frac{2 |G \cap E|}{|G| + |E|} \]

- **O** = overlap
- **G** = ground truth
- **E** = detected region

→ A detection is counted as successful, if \( O > 0.5 \)
Evaluation

Datasets

UND Collection J2
- Left profiles (-90 degrees)
- Minor occlusions though earrings

UND Collection NDOff-2007
- Yaw poses between -90 and -35, as well as 35 and 90 degrees
- Occusions through earrings and hair
- Not all images contain an ear
Results

Rotation invariance

-90 degrees | 99,5 | 96,5 | 97 | 98,5 | 99 | 99,5 | 100
90 degrees | 99,5 | 96,5 | 97 | 98,5 | 99 | 99,5 | 100
Flipped vertically | 99,5 | 96,5 | 97 | 98,5 | 99 | 99,5 | 100
Flipped horizontally | 99,5 | 96,5 | 97 | 98,5 | 99 | 99,5 | 100
No rotation | 99,5 | 96,5 | 97 | 98,5 | 99 | 99,5 | 100

<table>
<thead>
<tr>
<th>No rotation</th>
<th>Flipped horizontally</th>
<th>Flipped vertically</th>
<th>90 degrees</th>
<th>-90 degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>detection rate in %</td>
<td>99</td>
<td>98,6</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>
Results

Detection rates versus pose variation

Detected ears per yaw pose and database

<table>
<thead>
<tr>
<th>NDOff</th>
<th>-90</th>
<th>-60</th>
<th>-45</th>
<th>-30</th>
<th>30</th>
<th>45</th>
<th>60</th>
<th>90</th>
<th>-90 UND-J2</th>
</tr>
</thead>
<tbody>
<tr>
<td>depth image only</td>
<td>86,8571</td>
<td>70,9</td>
<td>50,49</td>
<td>23,7</td>
<td>19,79</td>
<td>49,36</td>
<td>86,1</td>
<td>91,76</td>
<td>92,9</td>
</tr>
<tr>
<td>texture and depth</td>
<td>96,47</td>
<td>83,47</td>
<td>76,49</td>
<td>58,91</td>
<td>42,7</td>
<td>76</td>
<td>85</td>
<td>93,53</td>
<td>99</td>
</tr>
</tbody>
</table>
Advantages and disadvantages

Pros
• Good detection accuracy
• Rotation invariance
• Some robustness to pose variation
• Returns good approximation of ear outline (normalization)
• Simple set of rules
• Uses shapes from the edge of the 3D-model
Advantages and disadvantages

Cons
• Runtime increases with number of shapes (5 secs on average)
• Needs rendered profile views
• Detects either one or no ear
• Can get trapped by long hair
Insights

• Fusion of co-registered texture and depth information overcomes disadvantages of single channel analysis

• The ear shape can be described with a set of simple features

• Context information from other face regions increases robustness against false positives
- Comments or Questions? -

Contact:
Anika Pflug, anika.pflug@cased.de

https://www.dasec.h-da.de