ShaRPy: Shape Reconstruction and Hand Pose Estimation From RGB-D with Uncertainty

Vanessa Wirth, A. Liphardt, B. Coppers, J. Bräunig, S. Heinrich, S. Leyendecker, A. Kleyer, G. Schett, M. Vossiek, B. Egger, and M. Stamminger

First markerless RGB-D approach tailored to medical applications

Focus on plausibility of motions
- Speculated hidden parts
- Speculated skeleton
- Misclassified predictions

Focus on reliability of motions with Uncertainty Estimation

Keypoint-based SOTA [2,3]

Results

Evaluation on H2O Pose Estimation

<table>
<thead>
<tr>
<th>Method</th>
<th>MEPE (mm)</th>
<th>3D PCK@15mm</th>
<th>Left</th>
<th>Right</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hassan et al.</td>
<td>30.07</td>
<td>30.07</td>
<td>41.87</td>
<td>41.87</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tekin et al.</td>
<td>41.32</td>
<td>41.32</td>
<td>38.86</td>
<td>38.86</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kwon et al.</td>
<td>41.45</td>
<td>41.45</td>
<td>37.21</td>
<td>37.21</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Aboukhandra et al.</td>
<td>36.80</td>
<td>36.80</td>
<td>36.50</td>
<td>36.50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cho et al.</td>
<td>24.44</td>
<td>24.44</td>
<td>25.80</td>
<td>25.80</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wen et al.</td>
<td>35.02</td>
<td>35.02</td>
<td>35.63</td>
<td>35.63</td>
<td>12.67</td>
<td>12.67</td>
</tr>
<tr>
<td>Cho et al.</td>
<td>14.40</td>
<td>14.40</td>
<td>15.90</td>
<td>15.90</td>
<td>70.75</td>
<td>70.75</td>
</tr>
<tr>
<td>Luo et al.</td>
<td>20.80</td>
<td>20.80</td>
<td>24.70</td>
<td>24.70</td>
<td>40.77</td>
<td>40.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27.81</td>
<td>27.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Motivation

Musculoskeletal diseases on the hand can be detected through motion. How can we objectively evaluate the hand function of a patient? Current SOTA with OMS is time-consuming and not intuitive → Markerless methods

Single RGB-D Camera (arbitrary model)

Intuitive 3D visualization of pose and personalized hand shape

Contributions

Neural Object Detector

Methods

RGB Image

Depth Map

Correspondence Matching

Pose and Shape Tracking

Uncertainty Estimation

Direct conversion

Anatomically aligned Segmentation canonical correspondence space ($H, S, V$)

Optimal Params $\Omega^* = \arg \min \Omega$

$E_{3d}(H, \theta)$

$E_{2d}(S, V)$

$E_{shape}(\theta)$

$E_{pose}(\theta)$

$E_{temp}(\theta^1, \theta^2)$

Predicted

Resolved

Predicted

Uncertainties

Cluster residuals per segment Thresholding Error-prone & unobserved segments

Error-prone

$\Rightarrow$ Cluster

Unobserved

Error-prone & unobserved segments

Thresholding

$f_{pred} = \sum_{j \in \mathbb{R}} f_{\text{seg}}(j) + \sum_{j \in \mathbb{R}} f_{\text{temp}}(j)$

$\Rightarrow$ Cluster

$\Rightarrow$ Cluster

Discarded poses

Qualitative Evaluation with OpenPose

Hand function tests on PsA patient → discard unreliable poses

THE END