Multi-view 3D human pose recovery in complex environment
Hofmann, K.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
A Main System Parameters and Their Settings

Overall

- $|C| = 3$ Number of cameras (Section 3.3)

Pre-processing

- $(S_x \times S_y \times S_z) = (10 \ m \times 10 \ m \times 2 \ m)$ Size of cuboid in which volume carving is performed (Section 3.3)

- $(D_x \times D_y \times D_z) = (200 \times 200 \times 75)$ Number of voxel grid divisions in the ground-plane oriented coordinate system (Section 3.3)

- $N_{\text{VoxelThresh}} = 5000$ Threshold on number of voxel elements below which 3D blob is removed (Section 3.3)

- $m_{\text{min}} = 0.5$ Minimum fraction of accumulated “mass” required of 3D blob in projection image (Section 3.3)

- $h_{\text{torso}} = 70 \ cm$ Threshold below which voxels are removed w.r.t. the $z$-coordinate (height) in the ground-plane oriented coordinate system (Section 3.3)

Pose exemplar representation (Sections 3.4.1 and 3.4.2)

- Below table lists upper- and lower- bounds (in degrees) and step sizes for the pose space discretization (Section 3.4.1)
APPENDIX A. SYSTEM PARAMETERS

<table>
<thead>
<tr>
<th>Angle</th>
<th>upper bound</th>
<th>lower bound</th>
<th>nr steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_t$</td>
<td>75</td>
<td>90</td>
<td>3</td>
</tr>
<tr>
<td>$\theta_t$</td>
<td>80</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>$\psi_t$</td>
<td>-180</td>
<td>157.5</td>
<td>16</td>
</tr>
<tr>
<td>$\phi_h$</td>
<td>-30</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$\psi_h$</td>
<td>-30</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>$\phi^l_s, \phi^r_s$</td>
<td>-45</td>
<td>144</td>
<td>8</td>
</tr>
<tr>
<td>$\theta^l_s, \theta^r_s$</td>
<td>15</td>
<td>148</td>
<td>8</td>
</tr>
<tr>
<td>$\psi^l_s, \psi^r_s$</td>
<td>-80</td>
<td>100</td>
<td>9</td>
</tr>
<tr>
<td>$\theta^l_e, \theta^r_e$</td>
<td>0</td>
<td>140</td>
<td>6</td>
</tr>
</tbody>
</table>

Table A.1: Pose space discretization bounds (in degrees) and step sizes.

- Non-colliding 3D poses are considered “allowable” (Section 3.4.1) if they fulfill following inequalities

$$\theta_s \leq 1.5 \theta_e + 310^\circ \quad (A.1)$$

$$\theta_s \leq -0.5 \psi_s + 110^\circ \quad \text{if } \psi_s > 40^\circ$$

$$\phi_s \geq 3 \psi_s - 160^\circ$$

where $(\phi_s, \theta_s, \psi_s)$ are the Euler angles (in degrees) of the shoulder joint and $(\theta_e)$ is the elbow joint angle, see Figure 3.2. The above bounds were determined experimentally.

- $N^* \approx 20000$ Maximum number of 2D pose exemplars that can still be practically handled with a partitional clustering algorithm of quadratic complexity, on a particular hardware (Section 3.4.2)

- $B = 10$ Branching ratio of tree of 2D pose exemplars (non-leaf level)

- $L^* = 3$ Level of tree that is constructed with a partitional clustering algorithm of quadratic complexity. Value is determined by $N^*$ and $B$.

- $|S_l|$ Number of 2D pose exemplars at tree level $l$. $|S_1| \approx 200$, $|S_2| \approx 2000$, $|S_3| \approx 20000$, and $|S_4| \approx 150000$ (Section 3.4.2)

- $\{f^1_{\text{sub}}, f^2_{\text{sub}}, f^3_{\text{sub}}, f^4_{\text{sub}}\} = \{3, 2, 1, 1\}$ Sub-sampling factors for the exemplars on the respective tree levels 1, $\ldots$, 4
Single frame pose recovery (Sections 3.4.3 – 3.5.3)

- \( \sigma_l \): Subsampling factor (in pixels) for exemplar data at tree level \( l \).
  \( \sigma_1 = 3 \) px, \( \sigma_2 = 2 \) px, \( \sigma_3 = 1 \) px, \( \sigma_4 = 1 \) px (Section 3.4.2).

- \( \gamma_l \): Image grid size (in pixels) for interest locations for matching at tree level \( l \).
  \( \gamma_1 = 6 \) px, \( \gamma_2 = 3 \) px, \( \gamma_3 = 1 \) px, \( \gamma_4 = 1 \) px (Section 3.4.3).

- \( \tau_l \): Threshold on posterior for nodes at tree level \( l \).
  \( \tau_1 = 0.01 \), \( \tau_2 = 0.016 \), \( \tau_3 = 0.02 \), \( \tau_4 = 0.024 \) (Section 3.4.3).

- \( u_x \): Area of non-maximum suppression at process stage \( x \), before truncation. In particular,
  \( u_{PosHyp} = 2 \) pixels (Section 3.4.3),
  \( u_{PosSel} = 1.25 \) cm (Section 3.4.4),
  \( u_{PosVerEx} = 1.75 \) cm (Section 3.5.1),
  \( u_{PosVerRdr} = 0 \) cm (no non-maximum suppression) (Section 3.5.2).

- \( K_x \): Number of hypotheses that are generated by process stage \( x \), after truncation. In particular,
  \( K_{PosHyp} \approx 300000 \) Value is determined by \( \tau_1...\tau_4 \) (Section 3.4.3),
  \( K_{PosSel} = 30000 \) (Section 3.4.4),
  \( K_{PosVerEx} = 2000 \) (Section 3.5.1),
  \( K_{PosVerRdr} = 800 \) (Section 3.5.2), and
  \( K_{PosVerRdrOpt} = 50 \) (Section 3.5.3).

Temporal Integration (Sections 4.2 – 4.3)

- \( T = 50 \): Number of image frames of time interval for which the best trajectories are computed.

- \( K_{traj} = 500 \): Number of best trajectories obtained by List-Viterbi algorithm.

- \( \lambda = 0.28 \): Weighting parameter for influence of temporal model.

Appearance Model Adaptation (Section 5.1)

- \( \alpha = 0.1 \): Learning rate pixel-based texture adaptation.
APPENDIX A. SYSTEM PARAMETERS

- \( R_b \) Texture model resolution (in pixels) for different body parts. 
  \( R_{\text{torso}} = 32 \text{ px} \times 32 \text{ px}, \ R_{\text{head}} = R_{\text{u.arm}} = R_{\text{l.arm}} = 16 \text{ px} \times 16 \text{ px}, \)
  \( R_{\text{neck}} = R_{\text{hand}} = 8 \text{ px} \times 8 \text{ px} \)

Shape model adaptation (Section 5.2)

- \( F = 8 \) Number of frames selected
- \( \tau = 0.6 \) Weighting parameter in frame selection cost function (Section 5.2.4)
- \( N_{F_{\text{SelIter}}} = 25000 \) Number of iterations of cost function evaluation in frame selection optimization process (Section 5.2.4)
- \( \lambda_{\text{opt}} = 0.003 \) Texture term weighting parameter in objective function (Section 5.2.5)
- \( \mu_{\text{opt}} = 0.5 \) Penalty term weighting parameter in objective function (Section 5.2.5)
- \( \mu_g = 0.05 \) Meta-parameter for stochastic gradient descent (Section 5.2.5)
- \( \lambda_g = 0.6 \) Meta-parameter for stochastic gradient descent (Section 5.2.5)
- \( n_{\text{iterLocal}} = 20 \) Number of iterations for one step of pose or shape parameter optimization
- \( n_{\text{iterGlobal}} = 250 \) Maximum number of iterations for optimization process
- \( D^{\text{init}}_{\text{vertex}} = 6 \text{ cm} \) Initial (coarsest) vertex sampling distance (Section 5.2.5)
- \( D^{\text{end}}_{\text{vertex}} = 2 \text{ cm} \) Finest vertex sampling distance (Section 5.2.5)