Supplementary Material

Nikolas Klug\textsuperscript{1*}  Moritz Einfalt\textsuperscript{2*} Stephan Brehm\textsuperscript{2} Rainer Lienhart\textsuperscript{2}
University of Augsburg
\textsuperscript{1}klug.nikolas@gmail.com
\textsuperscript{2}\{firstname.lastname\}@uni-a.de

1. Detailed Derivations

For the sake of completeness, we provide more detailed derivations for the best-case 3D estimates and the resulting minimal MPJPE under the normalized and weak perspective projection model.

1.1. Best-Case Normalized Perspective 3D Estimate

Recall that under the normalized perspective projection, the 3D pose estimate only depends on the per-point depth estimate \( \tilde{Z}_i \). Under normalization by translation as well as scaling, the squared JPE has the same form, with

\[
\Delta d^2_i(\tilde{Z}_i) = (\tilde{Z}_i - X_i)^2 + (\tilde{Z}_i - Y_i)^2 + (\tilde{Z}_i - Z_i)^2
\]

\[
= \tilde{Z}_i^2 (a^2 + b^2 + 1) - 2\tilde{Z}_i (aX_i + bY_i - Z_i) + X_i^2 + Y_i^2 + Z_i^2,
\]

(1)

where \( a \) and \( b \) are independent of \( \tilde{Z}_i \). With \( \Delta d^2_i \) being a square function in \( \tilde{Z}_i \), we obtain the best-case depth estimate by differentiating \( \Delta d^2_i \) w.r.t. \( \tilde{Z}_i \):

\[
\Delta d^2_i(\tilde{Z}_i) = 2 \left( \tilde{Z}_i (a^2 + b^2 + 1) - aX_i - bY_i - Z_i \right).
\]

(2)

Setting this expression equal to zero yields the optimal depth estimate \( \tilde{Z}_i^* \), with

\[
\tilde{Z}_i^* = \frac{aX_i + bY_i + Z_i}{1 + a^2 + b^2}.
\]

(3)

1.1.1 Minimal JPE under Translation Normalization

Under normalization by translation, we have \( a = \left( \sqrt{\frac{s}{X_i^2}} + dx \left( \frac{1}{Z_i} - \frac{1}{\tilde{Z}_i} \right) \right) \) and \( b = \frac{Y_i}{\tilde{Z}_i} \). With the optimal 3D estimate \( \tilde{P}_i = (\tilde{X}_i^*, \tilde{Y}_i^*, \tilde{Z}_i^*) \), the resulting minimal JPE for joint \( P_i \) is

\[
\Delta d_i = \sqrt{(\tilde{X}_i^* - X_i)^2 + (\tilde{Y}_i^* - Y_i)^2 + (\tilde{Z}_i^* - Z_i)^2}
\]

\[
= \sqrt{\left( aY_i - bX_i \right)^2 + (aZ_i - X_i)^2 + (bZ_i - Y_i)^2 \over 1 + a^2 + b^2}
\]

\[
= \sqrt{dx^2 \left( \frac{1}{Z_i} - \frac{1}{\tilde{Z}_i} \right)^2 \cdot (Y_i^2 + Z_i^2) \over 1 + a^2 + b^2}
\]

\[
= \left| dx \left( \frac{1}{Z_i} - \frac{1}{\tilde{Z}_i} \right) \right| \sqrt{Y_i^2 + Z_i^2 \over 1 + a^2 + b^2}.
\]

(4)

1.1.2 Minimal JPE under Scale Normalization

Under normalization by scaling, we have \( a = \rho \cdot \frac{X_i}{Z_i + dz} \) and \( b = \rho \cdot \frac{Y_i}{Z_i + dz} \). The resulting minimal JPE for joint \( P_i \) is calculated as

\[
\Delta d_i = \sqrt{(\tilde{X}_i^* - X_i)^2 + (\tilde{Y}_i^* - Y_i)^2 + (\tilde{Z}_i^* - Z_i)^2}
\]

\[
= \sqrt{(aY_i - bX_i)^2 + (aZ_i - X_i)^2 + (bZ_i - Y_i)^2 \over 1 + a^2 + b^2}
\]

\[
= \sqrt{\left( \rho \cdot \frac{X_i}{Z_i + dz} \cdot Z_i - X_i \right)^2 + \left( \rho \cdot \frac{Y_i}{Z_i + dz} \cdot Z_i - Y_i \right)^2 \over 1 + a^2 + b^2}
\]

\[
= \left| \frac{X_i^2 \left( \rho \cdot \frac{Z_i}{Z_i + dz} - 1 \right)^2 + Y_i^2 \left( \rho \cdot \frac{Z_i}{Z_i + dz} - 1 \right)^2 \over 1 + a^2 + b^2} \right|
\]

\[
= \left| 1 - \rho \cdot \frac{Z_i}{Z_i + dz} \right| \sqrt{X_i^2 + Y_i^2 \over 1 + a^2 + b^2}.
\]

(5)

1.2. Best-Case Weak Perspective 3D Estimate

Recall that the weak perspective projection model allows a perfect depth estimate \( \tilde{Z}_i = Z_i^* \). The mean squared per-joint error for the complete pose estimate is then reduced to

\[
\Delta d_2(s) = \frac{1}{n} \sum_{i} (X_i' - sx_i')^2 + (Y_i' - sy_i')^2
\]
2. Qualitative Examples

We provide additional examples of the best-case 3D human pose estimates under the normalized and weak perspective projection model. Figure 1 depicts examples from the Human3.6m [1], MPI-INF-3DHP [3] and CMU Panoptic [2] datasets at different MPJPE percentiles. The inherent simplifications in the projection models lead to guaranteed and clearly visible discrepancies compared to the 3D ground truth pose. Note that the largest deviations can be observed for persons that are either clearly off-center (top right, bottom) or very close to the camera (mid), which matches our quantitative results.

References

