Temporally optimized inverse kinematics for 6DOF human pose estimation

Gildea, Kevin; Mercadal-Baudart, Clara; Blythman, Richard; Simms, Ciaran

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Gildea, K., Mercadal-Baudart, C., Blythman, R., & Simms, C. (2022). *Temporally optimized inverse kinematics for 6DOF human pose estimation*. Paper presented at ESB 2022, Porto, Portugal.

Total number of authors:
4

TEMPORALLY OPTIMIZED INVERSE KINEMATICS FOR 6DOF HUMAN POSE ESTIMATION

Kevin Gildea (1), Clara Mercadal-Baudart (1), Richard Blythman (1) , Ciaran Simms (1)

1. Department of Mechanical, Manufacturing & Biomedical Engineering, Trinity College Dublin, Ireland

Introduction

The recent emergence of deep learning and computer vision-based 3D human pose estimation presents opportunities for a form of markerless motion-capture. State-of-the-art approaches have achieved remarkable accuracy in predicting global and relative joint positions [1], however, many potential applications require information on joint orientations, e.g., in the fields of biomechanics. Furthermore, methods that do include joint orientations are incompatible for applications with predefined *incongruent kinematic chains,* i.e., chains with differing limb lengths and proportions. Therefore, we propose a temporal inverse kinematics (IK) optimization technique to infer joint orientations in a user customizable kinematic chain from a position-based 3D pose input. This technique may be particularly useful for sports/injury biomechanics, and telehealth applications.

Methods

Due to the ambiguous joint 'twist' angle around links, the problem of mapping a kinematic chain to a hierarchical 3D position set ($[P]$) is indeterminate. Therefore, optimization is needed. A sequential least squares programming procedure [2] is developed to solve the minimization problem for a 16-joint kinematic chain (G) expressed using the Denavit-Hartenberg convention [3] as a hierarchical set of 4×4 transformation matrices ($\{T_j\}^{\underline{e}_{j_{parent}}}$) with parent-relative positions ($\{\vec{r}_{a_j}\}^{\underline{e}_{j_{parent}}}$) and orientations ($[A^{\underline{e}_j,\underline{e}_{j_{parent}}}]$) for each joint ($j$) (Eqn.s 1&2), with sequential Cardan angle rotation parameterization bounded joint ranges of motion (ROMs).

$$\{T_j\}^{\underline{e}_{j_{parent}}} = \begin{bmatrix} [A^{\underline{e}_j,\underline{e}_{j_{parent}}}] & \{\vec{r}_{a_j}\}^{\underline{e}_{j_{parent}}} \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \tag{1}$$

$$G = \left\{\{T_1\}^{\underline{e}_{j_{parent}}}, \cdots, \{T_{16}\}^{\underline{e}_{j_{parent}}}\right\} \tag{2}$$

Two IK algorithms are developed, 1) a frame-by-frame approach with local and global losses for pose (Eqn. 3), and 2) a 5-frame batch temporal approach including weighted losses for joint angular difference and positional difference (for links with missing joints) across frames ($2_{\text{5-frame}}$) (Eqn. 4).

$$L_{frame} = E_{\angle(\hat{a},\hat{b})} + \lambda_1 E_{\angle(\hat{p}^a,\hat{p}^b)} \tag{3}$$

$$L_{temporal} = E_{\angle(\hat{a},\hat{b})} + \lambda_1 E_{\angle(\hat{p}^a,\hat{p}^b)} + \lambda_2 E_{\Delta\vec{\theta}} + \lambda_3 E_{\Delta\vec{s}} \tag{4}$$

For assessment, we generate a series of motion sequences for the kinematic chain with 1) congruent, and 2) incongruent configurations, and applied both IK algorithms to the resulting joint pose sequences, i.e., with joint orientations hidden (see Fig. 1). Agreement was assessed using a 10-joint Mean Per Joint Angular Separation ($MPJAS_{11}$) between inferred and ground truth joint orientations. Randomly drawn rotations within joint ROMs average approximately 1.18 radians $MPJAS_{11}$ error (68 degrees/joint).

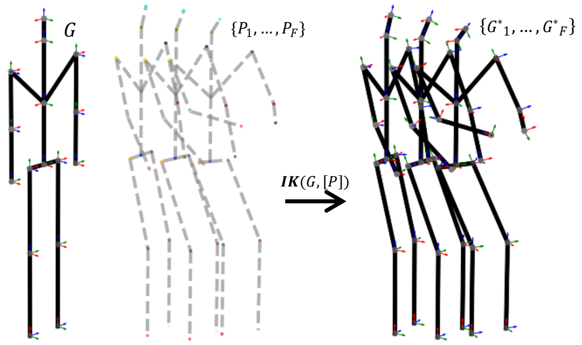

Figure 1: IK approach for 6-Degree of Freedom (DOF) pose from 3DOF pose.

Results

Table 1 shows a comparison of average angular errors. Where algorithm 1 is initialized with random weights for each frame (1_{rand_w}) or fed weights from previous frames (1_{prev_w}).

Algorithm	**Skeleton**					
	congruent			incongruent		
	e	b	tot	e	b	tot
1_{rand_w}	1.2×10^{-1}	0	4.0×10^{-2}	2.5×10^{-1}	3.5×10^{-2}	9.8×10^{-2}
1_{prev_w}	6.0×10^{-2}	0	1.8×10^{-2}	1.9×10^{-2}	3.4×10^{-2}	8.2×10^{-2}
$2_{\text{5-frame}}$	1.7×10^{-2}	0	5.0×10^{-3}	1.7×10^{-2}	3.5×10^{-2}	7.5×10^{-2}

Table 1: Accuracy of IK algorithms 1 and 2. $MPJAS_{11}$ (radians) per frame for e: frames with extended/straight limbs, and b: frames with bent limbs, for both congruent and incongruent skeleton types.

Discussion

With frame-by-frame IK and congruent skeletons (1_{rand_w}, 1_{prev_w}) we obtain uniquely optimal solutions in the case of bent elbows and knees, however, for extended/straight limbs the solution space is non-unique. With our temporal approach ($2_{\text{5-frame}}$) we reduce ambiguity for these poses, resulting in lower overall errors for both congruent and incongruent skeletons, with negligible errors associated with the former (0.3 degrees/joint), and low errors for the latter (4.3 degrees/joint). This technique allows for accurate prediction of joint orientations, for convenient post-processing of 3D pose estimates (e.g., Fig.2).

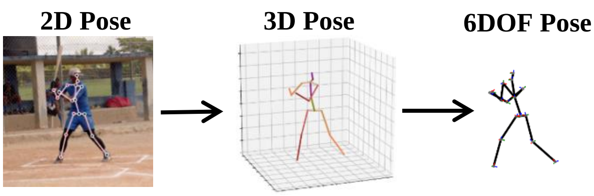

Figure 2: Example application using 3D predictions from [4].

References

1. Zheng et al, J. ACM, 37, 2022.
2. Kraft, 1998.
3. Denavit et al, J. Appl. Mech, 22: 215–221, 1955
4. Liu et al, ICCV, 2272–2281, 2019

Acknowledgements

The authors thank the Irish RSA, which funded this research as part of the RSA-Helena Winters Scholarship.