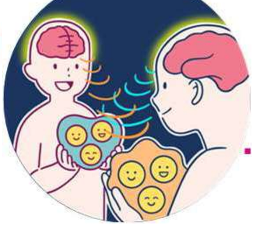
IoB Middleware

Moonshot Research and Development Program Goal 1 Kanai Project



Estimation of Human State Using Indirect Biometric Information and Its Applications

Prof. Hideki Koike Institute of Science Tokyo

Overview

Estimating human posture and motor states is a crucial technology in fields such as rehabilitation and skill acquisition support. Although advanced state estimation using electroencephalography (EEG) holds promise for future applications, current EEG-based methods still face limitations in noise robustness and real-time performance, preventing their practical deployment. To address this challenge, the Koike Laboratory focuses on more stable and readily obtainable indirect physiological signals—such as electromyography (EMG), gaze information, and plantar pressure—and aims to develop posture estimation techniques based on these modalities (Fig. 1).

In this study, we concentrate specifically on plantar pressure sensing and seek to develop a high-precision method for three-dimensional posture estimation using only the information obtained from compact, wearable pressure sensors. By eliminating the need for costly and spatially restrictive optical motion-capture systems, our goal is to establish a real-time and versatile inference system suitable for practical use in real-world environments.

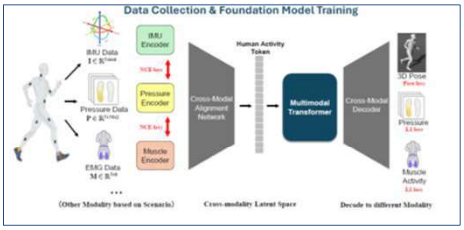
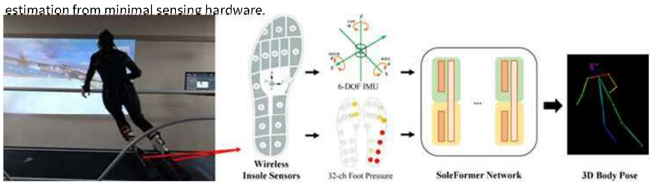


Fig 1 Overview of this entire study

SolePoser: From Foot Pressure to Posture

In this study, we propose SolePoser, a lightweight and real-time method for estimating full-body three-dimensional posture using only insole-type plantar pressure sensors. Conventional motion-capture systems require expensive camera setups or the attachment of numerous IMUs, which limits their applicability in daily-life settings and sports environments. In contrast, our method takes the plantar pressure distribution obtained from insole sensors as input and estimates the 3D positions of 17 body joints using a Dual-Stream Transformer architecture, SoleFormer. To ensure physically consistent and highly accurate estimation, we introduce two key components: a Dual Cycle Loss, which reconstructs IMU signals and plantar pressure from the predicted poses, and a Cross-Attention mechanism that emphasizes the relationships between IMU dynamics and pressure patterns. Together, these components enable stable and precise full-body pose estimation from minimal sensing hardware.



Furthermore, to evaluate posture estimation from plantar pressure, we constructed a new large-scale multimodal dataset. The SP-S dataset, specialized for sports movements, includes high-difficulty actions such as skiing, snowboarding, golf swings, and table tennis, comprising a total of 606K frames collected from 28 participants. The SP-E dataset, which focuses on daily activities such as walking and squatting, contains 302K frames obtained from 10 participants (mean age 25.9 years). Combined, the two datasets amount to 908K frames, achieving significantly greater diversity and scale than existing datasets such as TMM. Experimental results demonstrated that our method achieved an average joint position error of 65.3 mm for sports-motion data and 51.0 mm for everyday movements, with angular errors of 29.7° and 22.3°, respectively. The inference speed is 11 ms per frame (on an RTX 3090 Ti), indicating that the system is fully capable of real-time deployment.

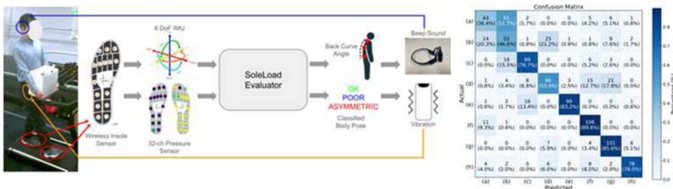
Network	Inference Time (ms)	Results w/o Elbow and Wrist (13 Joints)						Results w/ Elbow and Wrist (17 Joints)					
		SP-S		SP-E		TMM100 [40]		SP-S		SP-E		TMM100 [40]	
		MPJPE	MPJAE	MPJPE	MPJAE	MPJPE	MPJAE	MPJPE	MPJAE	MPJPE	MPJAE	MPJPE	MPJAE
PoseFormer (RGB) [70]	22.5	63.8	29.9	49.2	22.2	42.7	19.0	65.2	31.3	48.0	21.4	46.9	20.4
IMUPoser (2 IMUs) [31]	10.2	85.0	39.1	80.2	37.4	✗	✗	103.7	47.9	99.9	46.1	✗	✗
AvatarPoser (3 IMUs) [18]	6.1	74.6	37.6	69.7	31.3	✗	✗	70.7	34.9	66.7	30.0	✗	✗
Simple Foot IK [8]	5.0	119.3	49.3	100.7	46.7	✗	✗	135.3	56.6	118.5	53.3	✗	✗
Direct Transformer	8.7	72.2	35.3	69.5	31.0	✗	✗	87.6	40.3	77.4	36.6	✗	✗
PressNet (Inversed) [40]	38.9	87.3	40.9	82.8	38.2	54.3	25.8	94.4	44.6	89.9	41.9	59.7	27.5
SoleFormer (Pressure Only)	10.2	81.7	39.0	79.9	37.6	40.0	18.3	90.4	42.4	86.3	40.1	48.5	22.0
SoleFormer w/o CA	10.9	65.4	30.2	51.0	22.1	✗	✗	71.5	35.2	70.2	31.1	✗	✗
SoleFormer (Ours)	11.0	65.3	29.7	51.0	22.3	✗	✗	70.9	34.6	67.1	30.3	✗	✗

In comparative experiments, our method outperformed RGB camera–based approaches such as PoseFormer and IMU-based methods including IMUPoser and AvatarPoser, and substantially surpassed the existing plantar-pressure–based method PressNet. Ablation studies further demonstrated that incorporating plantar pressure input improved performance by 17 mm in average joint position error and by 8° in angular error compared to the use of IMU signals alone. We also confirmed that the introduction of Dual Cycle Loss and Cross-Attention contributed significantly to the overall accuracy improvements.

For real-world evaluation, we conducted validation experiments in alpine skiing and natural outdoor walking environments. As baselines, we compared against Xsens (using 17 IMUs) and third-person video captured using chest- and head-mounted cameras. The results showed that SolePoser achieved performance second only to TransPose and PIP, both of which rely on six IMUs, and exceeded RGB-based PoseFormer in skiing tasks. For walking motions, SolePoser matched the performance of PoseFormer, with only a 0.4 mm difference in joint position error.

These results were presented at the international conference UIST 2024[1].

SoleLoadEvaluator: Application to Posture Correction



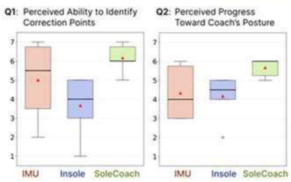
Building upon SolePoser, we propose SoleLoadEvaluator, a system that estimates forward-load walking posture using only plantar pressure sensors and provides real-time feedback. Conventional camera-based methods are sensitive to occlusions and environmental conditions, while multi-IMU configurations impose substantial wearing burden. In this study, we collected 8 types of load-bearing postures from 12 participants, constructing a dataset comprising approximately 670K frames.

In quantitative evaluations of posture classification and back flexion angle estimation, **the system achieved 72.4% accuracy for 8-class posture classification and a 7.3° error in flexion angle estimation**. Furthermore, adapting the model to a new user with one minute of personalized data increased performance to 82.4% accuracy and 4.2° error, demonstrating practical adaptability.

This system enables safe load-carrying assistance with minimal sensing hardware and low user burden, indicating strong potential for applications in industrial settings and rehabilitation. **This work was presented as an oral presentation at AHs 2025 [2].**

SoleAICoach: Multimodal Coaching

In this study, we propose SoleAICoach, a system that integrates plantar pressure sensing with a multimodal large language model (LLM) to automatically generate coaching feedback aimed at supporting skill acquisition [3]. While conventional coaching-support systems rely on cameras or motion-capture technologies, SoleAICoach directly converts plantar-pressure data into natural language feedback using an LLM, **enabling users to understand individualized points for improvement without requiring visual recording equipment**.



The dataset proposed in this study is **the first coaching dataset in which audio coaching comments, body posture, and plantar-pressure signals are fully synchronized**, comprising 3.1 million frames collected from 22 skiers.

In quantitative evaluations, the **similarity** between the generated coaching comments and those provided by expert coaches reached **85.2%**. User studies further demonstrated that participants were able to correct their own posture more effectively with feedback generated by SoleAICoach compared to conventional methods.

[1] Erwin Wu, Rawal Khirodkar, Hideki Koike, Kris Kitani, “SolePoser: Full Body Pose Estimation using a Single Pair of Insole Sensor”, ACM Symposium on User Interface Software and Technology (UIST) 2024.
[2] Shino Ito, Yichen Peng, Erwin Wu, Hideki Koike, “SoleLoadEvaluator: A Real-Time Feedback System for Walking Posture with Anterior Load Using Insole Sensors”, Augmented Humans (AHs) 2025.
[3] Toshihiro Hirano, Yuki Tabei, Yichen Peng, Chen-Chieh Liao, Erwin Wu, Hideki Koike, “SkiTechCoach: A Multimodal Alpine Skiing Dataset with 3D Body Pose, Sole Pressure, and Expert Coaching”, ACM Multimedia (MM) 2025.

Future Prospects

Traditionally, skill acquisition has focused primarily on imitating the external appearance of experts, particularly their body posture. However, in the future, understanding internal physiological signals—such as EEG, EMG, and gaze—will become equally important. Although measuring such biosignals remains costly and technically demanding at present, developing methods that enable simplified measurement or estimation of these signals through computer vision and AI technologies will be essential.

Hideki Koike, Ph.D.
Professor Institute of Science Tokyo

He is a Professor at the School of Computing, Tokyo Institute of Science (formerly Tokyo Institute of Technology). He received his Ph.D. in Engineering from the University of Tokyo in 1991. His research focuses on Human–Computer Interaction (HCI), particularly Vision-based HCI using computer vision techniques. In recent years, as a principal investigator for projects such as JST CREST and JST ASPIRE, he has conducted research on skill-acquisition support for professional athletes, expert pianists, and surgeons using computer vision, AI, and VR/AR technologies. He has received the IEEE Best Paper Award, the Japan Society for Software Science and Technology (JSSST) Fundamental Research Award, and is a Fellow of the Information Processing Society of Japan (IPSJ).

Peng Yichen
Appointed Assistant Professor Institute of Science Tokyo

Wu Erwin
Appointed Associate Professor Institute of Science Tokyo

Liao Chen-Chieh
Researcher Institute of Science Tokyo

Internet of Brains