

Learning-based Pose Estimation
with Multimodal Sensing
and Physics-based Modeling
for Practical Human Motion Capture

March 2026

Ryosuke Hori

主 論 文 要 旨

No.1

報告番号	㊦ 乙 第	号	氏 名	堀 涼介
主 論 文 題 名 : Learning-based Pose Estimation with Multimodal Sensing and Physics-based Modeling for Practical Human Motion Capture (学習ベースの姿勢推定を基盤としたマルチモーダルセンシングおよび物理ベースモデリングによる実用的な人間動作計測)				
(内容の要旨) 人間の動作をデジタル化するモーションキャプチャ (MoCap) 技術は、スポーツ科学、リハビリテーション、ロボティクスなど多様な分野における基盤技術として重要性を増している。しかし、従来の MoCap システムは多数のカメラを必要とし、設備コストや設置環境に大きな制約があるため、日常生活空間で広く活用することは依然として困難である。本研究は、日常環境において高精度かつ信頼性の高い人間動作計測を実現することを目的とし、多様なセンサモダリティを活用した4つのアプローチを体系的に示す。 第一のアプローチでは、固定カメラの撮影範囲の制約や遮蔽の問題に対応するため、手首装着型の全天球カメラを用いる MoCap 手法を構築する。人物シルエット画像を活用する機械学習モデルにより、日常環境においても頑健な全身3次元姿勢推定を可能にする。第二のアプローチでは、悪照明条件に対する頑健性向上のための、輝度変化を非同期に取得するイベントカメラを利用した手法を示す。低照度環境や高速動作といった従来の RGB カメラが不向きな状況下においても、姿勢および形状を高精度に推定でき、イベントデータの有効性を明らかにする。第三のアプローチでは、日常生活での歩行動作の簡便な取得を目的とし、靴内部に内蔵された少数の慣性センサ (IMU) 情報から全身歩行姿勢を推定する枠組みを示す。最小限の個数のセンサー構成と、長時間計測における誤差蓄積を回避する仕組みにより、日常的な歩行解析や長期的モニタリングへの実用性を高める。第四のアプローチでは、少数のウェアラブルセンサ計測を物理シミュレーションと統合し、物理的整合性を備えた MoCap を行うためのフレームワークを構築する。スマートウォッチやスマートインソールから得られる運動学的・力学的情報をもとに、関節トルク駆動の人体モデルが物理シミュレーター内で動作を再現する。 これらのアプローチは、視野制限、照明条件、装着負荷、物理的妥当性といった多面的な技術課題に対してそれぞれ独立して有効性を示す一方で、センサ構成やモダリティ、推定原理が相互補完的に機能する。将来的には、状況に応じてこれらを組み合わせて利用することで、日常環境における MoCap に求められる「精度と利便性の両立」をより高い水準で実現し得る。それにより、産業・医療・介護などの現場で求められるモニタリングや機能評価を支える基盤技術としての役割を果たすとともに、さらにその先のヒューマンデジタルツインや次世代ロボティクスの実現へと発展する可能性を持つ。				

Thesis Abstract

No. _____

Registration Number	☒ “KOU” ☐ “OTSU” No. _____ *Office use only	Name	Ryosuke Hori
Thesis Title Learning-based Pose Estimation with Multimodal Sensing and Physics-based Modeling for Practical Human Motion Capture			
Thesis Summary Motion capture (MoCap), which digitizes human movement, is becoming increasingly important as a fundamental technology in many fields such as sports science, rehabilitation, and robotics. However, conventional MoCap systems require multiple cameras and involve substantial constraints related to installation environments and equipment costs, making it difficult to deploy them widely in daily living spaces. This study aims to enable accurate and reliable human motion capture in everyday environments and systematically presents four approaches that utilize diverse sensor modalities. The first approach addresses the limited field of view and occlusion issues of fixed-camera systems by employing a wrist-mounted omnidirectional camera. A machine learning model that leverages human silhouette images enables robust full-body 3D pose estimation even in daily environments. The second approach enhances robustness under challenging lighting conditions by using an event camera that asynchronously captures brightness changes. This method achieves accurate estimation of body pose and shape in situations where conventional RGB cameras fail, such as low-light environments or rapid movements, and demonstrates the effectiveness of event-based data. The third approach focuses on convenient gait capture in daily life and proposes a framework for estimating full-body walking posture using only a small number of inertial measurement units (IMUs) embedded in shoes. A minimal sensor configuration combined with a mechanism that suppresses error accumulation during long-duration measurements enhances practicality for everyday gait analysis and long-term monitoring. The fourth approach integrates wearable sensor measurements with physics-based simulation to achieve physically consistent MoCap. Using kinematic and dynamic information from smartwatches and smart insoles, a torque-driven human model in a physics simulator reconstructs physically plausible movement. These approaches independently address issues such as field-of-view constraints, lighting variability, sensor intrusiveness, and physical consistency, while their sensor configurations and estimation principles complement one another. Combining them according to situational needs may achieve a higher level of accuracy and convenience required for MoCap in daily environments. This potential supports monitoring and functional assessment in industrial, medical, and caregiving settings and contributes to the development of advanced frameworks such as human digital twins and next-generation robotics.			