# 論文 / 著書情報
# Article / Book Information

| | |
|---|---|
| 題目(和文) | |
| Title(English) | Markerless Human Motion Capture and Visualization from Monocular Videos |
| 著者(和文) | HWANGDong-Hyun |
| Author(English) | Dong-Hyun Hwang |
| 出典(和文) | 学位:博士(学術),<br>学位授与機関:東京工業大学,<br>報告番号:甲第11848号,<br>授与年月日:2022年3月26日,<br>学位の種別:課程博士,<br>審査員:小池 英樹,徳永 健伸,三宅 美博,岡崎 直観,齋藤 豪,佐藤 洋一 |
| Citation(English) | Degree:Doctor (Academic),<br>Conferring organization: Tokyo Institute of Technology,<br>Report number:甲第11848号,<br>Conferred date:2022/3/26,<br>Degree Type:Course doctor,<br>Examiner:,,,,, |
| 学位種別(和文) | 博士論文 |
| Category(English) | Doctoral Thesis |
| 種別(和文) | 審査の要旨 |
| Type(English) | Exam Summary |

（博士課程）

# 論文審査の要旨及び審査員

<table>
<tr><td>報告番号</td><td colspan="3">甲第　　　　　　　　　号</td><td>学位申請者氏名</td><td colspan="2">Hwang Dong-Hyun</td></tr>
<tr><td rowspan="5">論文審査<br>審 査 員</td><td></td><td>氏　名</td><td>職　名</td><td></td><td>氏　名</td><td>職　名</td></tr>
<tr><td>主査</td><td>小池英樹</td><td>教授</td><td rowspan="4">審査員</td><td>斎藤豪</td><td>准教授</td></tr>
<tr><td rowspan="3">審査員</td><td>徳永健伸</td><td>教授</td><td>佐藤洋一</td><td>教授</td></tr>
<tr><td>三宅美博</td><td>教授</td><td></td><td></td></tr>
<tr><td>岡崎直観</td><td>教授</td><td></td><td></td></tr>
</table>

論文審査の要旨（2000 字程度）

本論文では，「Markerless Human Motion Capture and Visualization from Monocular Videos」と題し，1 台のカメラだけを用いて人の 3 次元姿勢推定と擬似 3 次元映像合成を行う手法について述べている．本論文は英文 6 章からなる．

第 1 章「Introduction」では，本論文の背景として従来の人の 3 次元姿勢推定手法と 3 次元映像のモデリングについて述べている．続いて，それらの問題点として複数台のカメラを必要とすること，特により正確な姿勢推定及びモデリングには数台～数十台のカメラが必要であることを述べている．これに対して，本論文では，(1)胸に装着した 1 台の小型カメラだけを用いた 3 次元姿勢推定手法，(2)モバイルデバイスでも実行可能な軽量な 3 次元姿勢推定手法，(3)1 台のカメラ映像から疑似 2.5 次元の複合現実コンテンツを生成する手法を提案することが述べられている．

第 2 章では「Related Work」と題し，複数視点映像を用いた動作計測手法，第三者視点映像による動作計測手法，自己中心視点映像を用いた動作計測手法，頭部姿勢推定手法，単一視点映像からのコンテンツ生成技術，複数視点映像を用いたコンテンツ生成技術等が述べられている．

第 3 章「Multimodal Human Motion Capture using A Ultra-wide Fisheye Camera」では，胸に装着した 1 台の小型カメラによる 3 次元姿勢推定手法について述べている．まず，試作ハードウェアとして小型アクションカメラに画角 280 度の超魚眼レンズを装着したものが紹介されている．このカメラから得られる超魚眼画像には，前方画像だけでなく装着者の手足，頭部の一部が撮影されている．この超魚眼画像を申請者の開発した深層学習器で処理することで，装着者の身体および頭部の 3 次元姿勢の推定を可能とした．この深層学習器は身体姿勢を推定する BodyPoseNet，カメラの姿勢を推定する CameraPoseNet，頭部の姿勢を推定する HeadPoseNet から構成され，それぞれが ResNet を拡張して実現されている．学習にはコンピュータグラフィックスを用いて作成した 68 万枚の合成画像による MonoEye データセットと CMU Panoptic Studio で撮影した 1.6 万枚の実画像データセットを使用した．なお，MonoEye データセットは他の研究者用にインターネット上に広く公開されている．定量的評価の結果，姿勢推定では MonoEye データセットの場合，関節平均誤差(MPJPE)が 43.6mm，実画像データセットでは 84.9mm を達成した．同様に頭部姿勢推定では平均誤差(MAE)が MonoEye データセットで 4.1 度，実画像データで 13.2 度であった．

第 4 章「Lightweight 3D Human Pose Estimation with Teacher-Student Learning」では，RGB カメラ 1 台を用いた軽量の 3 次元姿勢推定ネットワーク「MoVNect」と、知識蒸留(knowledge distillation)

に基づいた効率的な学習法を提案している．MoVNect は広範な CNN 構成において，姿勢推定精度と推論時間のベンチマークを行った結果， Bilinear+Conv2D の組み合わせが最適であるとの結果を得て，さらに Teacher-Student Learning を組み合わせることで学習の効率化を図り，スマートフォンのような限りある計算資源のデバイスでの実時間動作を実現した．

第 5 章「Synthesized Pseudo-2.5D Mixed Reality Content form Monocular Videos」では、1 台のカメラで撮影した動画から擬似 2.5D コンテンツを合成するエンドツーエンドシステム「MonoMR」を提案している．本システムでは、まず固定された単眼カメラ映像から OpenPose を用いて人体を検出する．次にあらかじめ映像内に設定した床の矩形領域の情報をもとに各人の奥行き情報を推定する．背景差分を用いて人体部分のテクスチャを抽出し，各人のテクスチャ画像を推定された奥行きに配置することで擬似 2.5 次元コンテンツを生成する．利用者は Head Mounted Display(Microsoft Hololens)を用いて実空間内に重畳表示された擬似 2.5 次元コンテンツを視点を変えながら見ることができる．なお，上記人体テクスチャ画像は 3 次元空間に配置された 2 次元画像であるため，利用者が視点を変更すると人体の厚みのない不自然な映像となる．この問題に対しては Billborad レンダリングの手法を用いて利用者の視点方向に 2 次元テクスチャ平面を正対させることで不自然さを解消している．評価としては，人の奥行き情報の正確さを実験により確認し，3m 程度の奥行きでは誤差 24cm 程度，20m 程度の奥行きでは誤差 76cm 程度であることが示された．本システムは，サッカー映像やダンス映像に適用され，擬似的 2.5 次元映像が合成できることが示された．さらに，本システムを監視カメラの映像に応用し，複数台の監視カメラ映像から，人が 3 次元的な部屋のどこを歩いているかを 3 次元的に可視化するアプリケーションを実装している．

第 6 章「Conclusions」では，第 3 章，第 4 章，第 5 章で述べたシステムやアプリケーションの貢献と科学的な発見をまとめ，集合視を用いたバーチャルモーションキャプチャスタジオの実装のための今後の課題を提案している．

以上，本論文では，既存の 3 次元動作計測，3 次元合成映像作成において，同期した複数台のカメラが必要であるという問題点に対し，前者では超魚眼カメラと深層学習を用いることで胸に装着した 1 台のカメラ映像から装着者の 3 次元身体姿勢および頭部姿勢の推定の概念を提案，実装，評価し，その有効性を確認した．また，こうした動作計測がモバイル環境で実現できるように，軽量な 3 次元動作推定手法を提案した．さらに，後者では単眼映像から擬似 2.5 次元映像を合成する手法を提案，実装，評価し，その有効性を確認した．以上は学術的に高い貢献がある．よって，本論文は博士（学術）の学位として十分価値があると認められる．