

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)	HWANG Dong-Hyun
Author(English)	Dong-Hyun Hwang
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第11848号, 授与年月日:2022年3月26日, 学位の種別:課程博士, 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), Conferring organization: Tokyo Institute of Technology, Report number:甲第11848号, Conferred date:2022/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

## 論文要旨

THESIS SUMMARY

系・コース： 情報工学 系  
Department of, Graduate major in 情報工学 コース  
学生氏名： HWANG Dong-Hyun  
Student's Name

申請学位 (専攻分野)： 博士 (Philosophy)  
Academic Degree Requested Doctor of  
指導教員 (主)： KOIKE Hideki  
Academic Supervisor(main)  
指導教員 (副)：  
Academic Supervisor(sub)

### 要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters )

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た人の3次元姿勢推定と擬似3次元映像の合成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コンピュータグラフィックス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の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章では、ポータブルマルチモーダルヒューマンモーションキャプチャシステム「MonoEye」を紹介する。このポータブルシステムは超広角カメラを用いて、カメラ着用者や周囲の人々の3次元モーション、ユーザーの視線領域などのマルチモーダル情報を撮影する。「MonoEye」で使用されるニューラルネットワークたちを訓練するため、様々な体型、衣服、アクション、背景、照明条件を反映した68万枚のフレームの合成データセットを提示し公開している。

第4章では、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、知識蒸留に基づいた効率的な学習法を提案する。さらに、人物の姿勢と推論時間のベンチマークに関する広範な評価を行い、その結果に基づき、提案する知識蒸留学習法が有意に精度を向上させた事とネットワークが実世界の様々なモバイルデバイスで迅速な推論できる事を確認する。

第5章では、拡張現実ヘッドマウントディスプレイ(HMD)向けに、単眼動画から擬似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」を提案する。本システムでは、動画の中の人体を自動的に検出し、テクスチャを抽出し、擬似的な3次元位置を自動で算出する。これらの情報を基にインタラクティブな拡張現実コンテンツを生成し、ユーザーは拡張現実HMDを通して自由な視点で合成されたコンテンツを楽しむことができる。ユーザスタディーにより、生成されたコンテンツは原本の単眼映像よりも没入感が高く、魅力的であることを確認した。

第6章では、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する。

備考：論文要旨は、和文2000字と英文300語を1部ずつ提出するか、もしくは英文800語を1部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報工学 情報工学	系 コース	申請学位 (専攻分野)： Academic Degree Requested	博士 Doctor of	(Philosophy)
学生氏名： Student's Name	HWANG Dong-Hyun		指導教員 (主)： Academic Supervisor(main)	KOIKE Hideki	
			指導教員 (副)： Academic Supervisor(sub)		

### 要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words )

In this thesis, we propose methods for motion capture and a method to create and visualize AR content by retargeting the captured motion from monocular videos.

Chapter 1 introduces the background of conventional 3D human pose estimation methods and computer graphics modeling. Then, the problems with current approaches, such as the requirement of multiple cameras for more accurate pose estimation and modeling, are described.

Chapter 2 briefly discuss related works about motion capture system, 2D and 3D human pose estimation, monocular video-based content synthesis, and free-viewpoint video systems.

Chapter 3 presents MonoEye, a multimodal human motion capture system using a single RGB camera with an ultra-wide fisheye lens, mounted on the user's chest. The system estimates the various human-motion information such as the 3D motion of the camera wearer and surrounding people and the user's viewport using an ultra-wide fisheye camera.

Chapter 4 proposes MoVNect, a lightweight deep neural network to capture 3D human pose using a single RGB camera. In order to improve the overall performance of the model, we apply the teacher-student learning method-based knowledge distillation to 3D human pose estimation. Extensive evaluations show the advantages of our lightweight model with the proposed training method.

Chapter 5 presents MonoMR, end-to-end system that estimates motion information, creates a 2D texture that could be retargeted, and generates pseudo-2.5D AR content based on this information. The MonoMR system can be used by casual end-users to generate MR content from a single camera only. In order to evaluate the efficiency and interactive potential of MonoMR, we conducted performance evaluations and a user study with 12 participants.

Chapter 6 summarizes the contributions and scientific findings of the systems and applications described in Chapters 3, 4, and 5 and proposes future work to implement a virtual motion capture studio using collective vision.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).