T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	デュアル間接特徴を用いた人間姿勢予測と 技能習得への応用
Title(English)	Human Pose Prediction using Dual Indirect Features and its Application in Skill Acquisition
著者(和文)	Wu Erwin
Author(English)	Erwin Wu
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12248号, 授与年月日:2022年9月22日, 学位の種別:課程博士, 審査員:小池 英樹,德永 健伸,三宅 美博,篠田 浩一,岡崎 直観
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12248号, Conferred date:2022/9/22, 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

Powered by T2R2 (Science Tokyo Research Repository)

（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　系 コース	申請学位（専攻分野）： Academic Degree Requested	博士　（ 学術 ） Doctor of
学生氏名： Student's Name	WU ERWIN	指導教員（主）： Academic Supervisor(main)	小池　英樹
		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,「Human Pose Prediction using Dual Indirect Features and its Application in Skill Acquisition」と題し，カメラ画像などの入力において二種類の異なる間接特徴(Indirect Features)を用いて人の三次元姿勢を予測する手法と，それらの技能習得への応用について述べている．本論文は英文10 章からなる．

第1 章「Introduction」では，研究の背景として，技能の習得において姿勢解析や推定の重要性を述べている．そして，現在の姿勢推定の多くは入力から直接観察できる特徴に基づく手法で，間接特徴を使用する姿勢の予測が少ないと指摘している．一方で，人間の脳には直接特徴の推定と間接特徴からの予測の能力があり，特に予測はスポーツなどの技能において重要視されていることを述べている．また，間接特徴には様々な種類があり，本研究ではそれを時間特徴と空間特徴に分類した．その上で，本論文では，(1)これらの2種類の異なる間接特徴を抽出できる深層学習のフレームワーク，(2)そのフレームワークを使用して時間と空間の両方の間接特徴を抽出するネットワーク，(3)それらのネットワークを使用する複数個の技能習得アプリケーションを提案することが述べている．

第2 章「Related Work」において，本論文の関連研究や先行研究を紹介している．人間の姿勢の推定に関しては，既存のビジョンに基づく実時間推定方法，時間間接特徴に基づく姿勢予測研究，そして空間間接特徴に基づく姿勢予測研究があることを述べている．また，技能習得の関連研究として，直接特徴を使用する既存の支援システムの問題点や，VR/AR を使用する可視化手法について紹介している．

第3 章と第4 章は本論文の「Research Proposal」と具体的に実現を紹介する「Network Architecture」の章であり，研究の主なアプローチについて述べている．先述の間接特徴抽出のため，既存の手法の優劣を分析し，新たに直接特徴と間接特徴を同時に学習する2 ストリームネットワークのフレームワークを提案した．既存のネットワークが直接特徴と間接特徴を別々に学習するのと異なり，提案手法は両方のストリームで学習済みの情報を互いに共有し，間接特徴を用いて直接推定を増強し（またはその逆も），それにより高精度なネットワークを実現できる．また，本フレームワークはモジュール化されているため，今後個々のモジュールをより洗練されたネットワークに入れ替えることも容易に実現できると述べている．

第5 章「FuturePoseNet (Temporal Prediction)」では，提案フレームワークを時系列間接特徴の抽出に応用し，過去の時系列情報から未来の三次元姿勢を予測するFuturePoseNet を提案している．ネットワークの実装において，従来の研究を参考し，間接特徴抽出モジュールを新しく提案し，Keypoint Lattice Optical Flow (KLOF)という人の関節の動きに着目したオプティカルフローのアルゴリズムを開発した．同アルゴリズムを使用したFuturePoseNet は，従来のオプティカルフローを使用したアルゴリズムより計算速度が3 倍以上(10ms から3ms まで)速くなり，予測精度も向上することが後の実験で判明した．さらに，FuturePoseNet ではLSTM とGNN を間接特徴の学習に用い，最終的に0.5秒先の動きを最大71.3%の精度(6.6 ㎝の平均誤差)で実現し，従来のオフライン手法と比べ5%以上の向上を実現し，しかも実時間の予測を可能にしている．最後に，同ネットワークを卓球の軌道予測にも応用し，ネットワークの汎用性と拡張性を示した．

第6 章「IndirectPoseNet (Spatial Prediction)」において，提案フレームワークの別の可能性を探るため，空間上の間接特徴の抽出に応用したIndirectPoseNet を提案し，体の一部から別の一部もしくは全身の予測に成功している．前述のネットワークと同じく，細かい空間的変化を抽出できるOptical Flow-based Motion History Image (OF-MHI)を新たに開発し，画像の空間上の変化の間接特徴のより正確な抽出に成功した．それを基づくIndirectPoseNet を用いて，手の甲の皮膚や筋肉の変化から手の指の三次元姿勢の予測に成功し，より自然な腕装着型の3 次元手指モーションキャプチャを実現している．さらに，ネットワークに微分可能なカルマンフィルタを組込むことにより，平均9 度前後の指角度誤差を実現し，従来の手法とくらべ20%以上の向上が見られる．最後に，提案する手法の異なる応用例として，足圧からスキーの全身姿勢の推定への応用に成功した．

第7 章「Application on Skill Acquisition」では，第5 章と第6 章で提案したネットワークを種類が大きく異なる三種類の技能：アルペンスキー、ピアノの演奏、卓球の返球に応用し，それぞれ被験者実験を行った．三つの実験において，自然な練習システムの重要性、間接特徴による高精度推定の有用性、そして予測を用いる練習の効果を検証したことを述べている．

第8 章「Discussion」において，先ず第5 章と第6 章のネットワークの実験結果について議論し，提案手法が既存研究と比べて優位であることに触れ，また評価の問題点と今後の改善点を述べている．続いて，第7 章の三つのアプリケーションの被験者実験についても議論を行い，それぞれなぜ提案システムが有効なのかを分析し，被験者実験の改善点をあげ，今後の研究の手がかりを与えている．

第9 章「Future Vision」と第10 章「Conclusions」では，第3 章から第7 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，今後の課題としてもっと汎用的に統合されたシステムの開発と技能習得以外への応用を述べている．

以上，本論文では，多くの技能において動作の予測が重要であるにもかかわらず，既存の多くの技能習得方法は間接特徴に基づく人間の姿勢予測が用いられていない問題点に対して，新しい深層学習のフレームワークを提案し，時間と空間両方の間接特徴を抽出できるネットワークを実現した．それらネットワークを用いることで，人の過去の動きから未来の三次元姿勢の予測や体の一部の動きから全体の姿勢の予測などが可能となる．さらに，提案手法を三種類の異なる技能に応用し，実環境の練習アプリを開発し，それらの効果を複数の被験者実験によって検証、評価して，その有効性を確認した．

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　系 コース	申請学位（専攻分野）： Academic Degree Requested	博士　（ 学術 ） Doctor of
学生氏名： Student's Name	WU　ERWIN	指導教員（主）： Academic Supervisor(main)	小池　英樹
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

In this paper, we proposed an indirect feature-based real time pose prediction system and introduced several applications for skill acquisition using the proposed system.

Different from conventional direct feature-based pose estimation, the proposed system try to make use of those features which are indirect related with human posture (e.g. estimate full body pose from feet pressure).

The proposed network consists of two parts: a FuturePoseNet which aims to extract temporal indirect features from the input video sequences and a InvisiblePoseNet which finds out the spatial indirect relationship within each image. For each network, a special indirect feature extraction module is developed to enhance the learning of an indirect feature. The performance of both networks is quantitatively and qualitatively evaluated in the experiment, and the results suggest that the proposed indirect feature-based prediction can achieve similar accuracy as the conventional methods, without observing the direct features.

For applications, three types of different skill acquisition are introduced: Skiing, Piano, and Table Tennis, which aims to study the results from three different perspectives. The Skiing is mainly focus on spatial indirect features while the piano requires temporal one. Table Tennis is the most well-studied application which includes both temporal and spatial indirect features.

Finally, the contribution and limitation of the current work is discussed. The proposed framework is compared with other existing methods to provide clues for future research. To the best of our knowledge, this work is the first real-time 3D pose prediction using a dual-module indirect feature-based network, which is proved to be useful in different types of skill training and might open a new way for using indirect features in training.

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).