

論文 / 著書情報  
Article / Book Information

題目(和文)	複数視点の時系列奥行き情報の統合による隠蔽に頑健な人物追跡と領域分割
Title(English)	
著者(和文)	福司謙一郎
Author(English)	Kenichiro Fukushi
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9695号, 授与年月日:2014年12月31日, 学位の種別:課程博士, 審査員:熊澤 逸夫,長橋 宏,佐藤 誠,山口 雅浩,小尾 高史,長尾 智晴
Citation(English)	Degree:., Conferring organization: Tokyo Institute of Technology, Report number:甲第9695号, Conferred date:2014/12/31, 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

専攻： Department of	物理情報システム	専攻	申請学位 (専攻分野)： Academic Degree Requested	博士 (工学)	Doctor of
学生氏名： Student's Name	福司 謙一郎		指導教員 (主)： Academic Advisor(main)	熊澤 逸夫	教授
			指導教員 (副)： Academic Advisor(sub)		

### 要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

スマートスペースや Natural User Interface が急速に私達の生活に身近なものになりつつある中、人物の位置やジェスチャを取得する技術の重要性がますます高まっている。ジェスチャ認識、指紋認証、顔認証のように、センサに比較的近い限られた空間を対象とした認識技術は既に製品に組み込まれるほど普及している。また、建物内や交差点のような、より広い空間内を自由に人物が動く条件での認識技術も、多くの研究者が意欲的に取り組んでおり進歩がめざましい。このような条件下で人物追跡を困難にする主な要因の一つとして、隠蔽が知られている。これは画像上で、ある物体が別の物体の背後に隠れ、一部あるいは全体が見えなくなってしまう現象である。カメラ視点画像上では奥行き情報が失われるために、画像上で重なっている領域は容易には分割できない。

近年の研究では、実時間で奥行き情報を計算できる Depth Camera が普及してきており、隠蔽を解消する手段として積極的に利用されている。通常のカメラでは失われてしまう奥行き情報は隠蔽に対して本質的な手掛かりとなる上、計測値が人物の見かけや照明変化の影響を受けにくく、実用上有用だからである。人物が鉛直方向に移動しないという仮定の下に奥行き情報を平面図で簡略化した表現は、人物追跡に利用しやすい特徴量として知られており、多くの研究で採用されている。しかし、姿勢・ジェスチャを知るのに必要な人物領域の形状は平面図表現ではほとんど失われてしまう。そのため既存研究では、隠蔽を扱いやすい、複数視点の情報を統合しやすい、といった平面図表現の特徴を活かして、頑健な人物動線検出の研究が発展している一方で、ジェスチャ認識のように人物形状を必要とする応用には適用されてこなかった。

本論文ではまず、奥行き情報を平面図に投影して得られる Occupancy Map を複数の視点で作成・統合し、複数のフレームに渡って最適化する事で、各視点の画像では領域分割を行わないまま、隠蔽に頑健な人物追跡を行う手法を提案する。次にこの人物追跡の結果をカメラ視点画像に投影することで領域分割を行う手法を提案する。一般に領域分割は物体追跡の前処理として行われるが、提案手法はその逆で、人物追跡の結果をカメラ視点画像に投影して領域分割を行う。従来手法では隠蔽によって分割が出来なかった領域でも正しく分割できる利点がある。

評価実験においては、異なる混雑度のデータセットに提案手法を適用して、隠蔽への頑健性を確認するとともに、人物追跡および領域分割のエラーを定量的に評価して手法の限界を検証する。また、従来手法では複数のフレームに渡る最適化のために動的計画法や反復法等が用いられてきたが、計算時間が大きくなる問題があった。提案手法ではヒューリスティクスなアプローチでの最適化手法を考案し、効率的な計算が可能である事を示す。さらに、本論文で構築したシステムはカメラのネットワーク化において、画像そのものは伝送せずに、処理に必要な情報のみをネットワークに送信する。通信量の評価によって、一般的なネットワーク機器でもシステムを構築出来る事も示した。

最後に、提案手法の応用としてジェスチャ認識を例にとり、混雑環境下で挙手を検知するプロトタイプシステムを構築した。提案手法は隠蔽が発生していても全ての視点で正しく領域分割が出来るため、それらの情報を統合することで各視点での隠蔽の影響を互いに補う事が出来る。本システムでは、全ての視点の領域を使って特定の人物のシルエットを作成する。実際に混雑環境下で撮影したデータセットを用いて実験を行い、安定してジェスチャ認識が可能である結果を得、これにより提案手法の実用性も確認した。

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

専攻： 物理情報システム 専攻  
Department of  
学生氏名： 福司 謙一郎  
Student's Name

申請学位 (専攻分野)： 博士 (工学)  
Academic Degree Requested Doctor of  
指導教員 (主)： 熊澤 逸夫 教授  
Academic Advisor(main)  
指導教員 (副)：  
Academic Advisor(sub)

## 要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

In this thesis, we present a computer vision-based human tracking and video segmentation system with multiple stereo cameras. Our method has two main features, as follows. First, we propose an occlusion-robust human tracking method that utilizes multi-view depth imagery acquired by the stereo cameras. Next, we propose a new segmentation technique, which can use the human tracking results to accurately extract human regions from video sequences that contain highly cluttered scenes.

Many of the methods that are used widely for human tracking, such as KLT-tracker, update the trackers “frame-to-frame,” where the features extracted from one frame are utilized to update the current state. By contrast, we propose a novel optimization technique for the “multi-frame” approach that computes the resultant trajectories directly from the video sequences. Thus, our method achieves a high level of robustness against severe occlusion, which is known to be a challenging problem in computer vision. We developed a heuristic optimization technique to estimate human trajectories, instead of using dynamic programming or an iterative approach, which means that our method is sufficiently computationally efficient to operate in realtime.

In our method, video segmentation is achieved by projecting the tracking results, which are represented as planar regions that are fitted to human shapes observed in the vertical direction, onto camera-view images. In general, video segmentation is performed before tracking in order to detect human regions. However, occlusion makes segmentation considerably difficult, where it may lead to the unfavorable fragmentation of a single human region and the confusion of multiple human regions, which both affect conventional methods adversely. Our strategy overcomes these issues successfully.

In our experimental evaluation, we processed six video sequences using our system, in which one to six people walked in a narrow laboratory space. The results demonstrate that our system is capable of handling cluttered scenes that contain severe occlusion and where people are frequently located in close proximity to each other. Moreover, minimal information is required for tracking, instead of using full camera images, which is communicated over the network in our system. Hence, the proposed method also has the advantage that commonly used network devices are sufficient to construct our tracking system.

Finally, we applied our method to a prototype demo system for detecting arm raising gestures. The results confirmed that our system is suitable for stable gesture recognition even in a highly cluttered scene.

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).