氏　　名	LE QUANG CHIEN
学位(専攻分野)	博士(情報学)
学 位 記 番 号	総研大甲第 1884 号
学位授与の日付	平成28年9月28日
学位授与の要件	複合科学研究科　情報学専攻 学位規則第6条第1項該当
学 位 論 文 題 目	Human Action Recognition from 3D Videos Using Multi-Projection-Based Approach
論 文 審 査 委 員	主　査　教授　杉本 晃宏 准教授　LE DUY DINH 教授　佐藤 いまり 准教授　CHEUNG GENE 教授　佐藤 真一 国立情報学研究所

(別紙様式 2)
(Separate Form 2)

論文内容の要旨
Summary of thesis contents

Human Action Recognition from 3D Videos Using Multi-Projection-Based Approach
LE QUANG CHIEN

Depth-based action recognition has been attracting the attention of researchers because of the advantages of depth cameras over standard RGB cameras. One of these advantages is that depth data can provide richer information from multiple projections. In particular, multiple projections can be used to extract discriminative motion patterns that would not be discernible from one fixed projection. However, high computational costs have meant that recent studies have exploited only a small number of projections, such as front, side and top. Thus, a large number of projections, which may be useful for discriminating actions, are discarded. In this thesis, we aim to enhance human action recognition by exploiting a pool of multiple projections. We propose solutions for three distinct scenarios, i.e., recognizing actions from similar, different and arbitrary view angles.

In the first scenario, some different actions can look similar in the same view. To address this challenge, the key idea is that such actions can be distinguished from another views. We propose to use a large number of projections to expand much meaningful information. These projections are performed via different viewpoints sampled on a geodesic dome to project 3D data onto multiple 2D-planes. Then we train and test action classifiers independently for each projection. To reduce the computational cost, we propose a greedy method to select a small yet robust combination of projections. The idea is that best complementary projections will be considered first for searching the optimal combination. This method provides us a reasonable choice which is more effective than using heuristic projections.

In the second scenario, we handle a challenging task due to viewpoint variation which is known as the cross-view action recognition. In this case, some same actions can look different from different views. To handle this challenging issue, existing research generally deals with this problem by transferring knowledge. The knowledge transfer is based on finding a viewpoint independent latent space in which action descriptors from different viewpoints are directly comparable. However, the approach required actions which share the target view information. In this thesis, we deal with this challenge in another direction. We select views in which trained classifiers are robust to viewpoint variations. To do that, we propose a discrimination-based selection method. The key idea is that classifiers trained on each viewpoint are evaluated on samples from all possible views. We select viewpoints in which the classifiers achieve the highest performance. This method is more flexible and widely applicable to a lot of scenarios of different viewpoints.

Finally, we argue that a robust action recognition system should be effective to the scenarios of arbitrary views, i.e., the views are either same or different. In this case, how to guarantee the final performance of action recognition system. To do that, we propose a hybrid fusion method in which all the benefits herein aforementioned are embedded. In our developed framework, we simply combine both previous methods. For each method, we obtain corresponding classification scores to predict action labels. However, for the first method, we have no information related to a corresponding combination of viewpoints from target data. This issue can compromise the final performance when target viewpoints are different from source viewpoints. Therefore, we propose a correlation-based selection method to collect possible combinations of viewpoints. Then, each test sample can be described by various representations corresponding to the selected combinations. And, we obtain classification scores corresponding to the representations from the trained classifiers. Finally, we decide the action category by fusing two sets of classification scores.

We verify the effectiveness of our approaches on the benchmark datasets, MSR Action 3D, MSR Gesture 3D, 3D Action Pairs and Northwestern-UCLA Multiview Action 3D. Our approaches can not only recognize actions from seen viewpoints, but also can apply for cross-view action recognition. Compared to other state-of-the-art approaches, our solutions achieve outstanding results.

(別紙様式 3)
(Separate Form 3)

博士論文の審査結果の要旨

Summary of the results of the doctoral thesis screening

Human Action Recognition from 3D Videos Using Multi-Projection-Based Approach
LE QUANG CHIEN

本論文は、Human Action Recognition from 3D Videos Using Multi-Projection-Based Approach (複数射影を用いた三次元映像による人物動作認識に関する研究)と題し、奥行き情報も含む三次元映像から人物の動作を認識するための技術について述べている。Kinect 等三次元センサーは一般的になってきており、これを用いた動作認識は、ロボットインタラクション、ゲーム、VR 等でも重要な技術である。本論文では、典型的な動作の映像を学習データとして動作種別を学習し、未知の動作映像の動作種別を認識する技術について、特に、与えられた三次元映像に対し、仮想的に設定した複数の観測方向に射影した映像を作成し、Dense Trajectory 特徴に基づく Bag of Visual Words 法により算出した特徴量に対して SVM により識別する方法を基本構成要素としている。その上で、観測方法の選別法、複数の特徴量の統合法、複数の識別結果の融合法等について、学習映像と対象映像との観測方向が同じ場合、異なる場合、同じか異なるか不明である場合それぞれについて工夫し、、より高精度の認識が可能な方法について検討し、英文にてまとめている。

第一章 Introduction(序論)では、本研究の動機、動作認識の概要、課題、本論文の貢献についてまとめている。

第二章 Related Work and Datasets (関連研究とデータセット)では、本研究に関連する研究ならびにベンチマークデータセットについて調査している。

第三章 Exploiting Complementary Projections for Action Recognition from Similar Views (相補的射影を用いた同様の観測方向のための動作認識)では、学習映像と対象映像の観測方向がほぼ同じ場合を対象とし、観測方向によって識別しにくい動作があることを考慮し、三次元映像から複数方向への射影を人工的に作り出し、それらのうちからもっとも識別精度が高くなると期待される複数の射影方向を相補的射影として自動抽出する方法を提案している。学習映像と対象映像の観測方向がほぼ同じであることから、射影間の対応関係は自明である。本手法により、従来法より顕著に高い識別性能を達成している。

第四章 Exploting Discriminative Projections for Action Recognition from Different Views (識別性の高い射影を用いた異なる観測方向のための動作認識)では、学習映像と対象映像との観測方向が異なる場合を対象とし、観測方向が異なっていても識別性能が高い射影をあらかじめ選んでおくことにより、高性能の識別を狙った方法を提案している。実際に観測方向が異なる場合でも精度良く識別できることが示されている。

第五章 Hybrid Fusion for Action Recognition from Arbitrary Views (ハイブリッド融合法による任意観測方向での動作認識)では、先の二章の融合方法であり、相補的射影と識別性の高い射影の双方を用い、かつ観測方向間の対応関係を相関法により求め、各射影ごとの識別結果を融合する方法であり、任意観測方向の状況で最も高い識別性能を達成している。

第六章 Conclusions (結論)にて本論文の成果をまとめている。

本論文で提案している三次元映像による動作認識手法は、ベンチマークデータに基づく評価によると、既存手法を上回る高い検索性能を達成していることが示されている。加えて、様々な観測方向の条件下でも頑健に高い識別が可能な方法を新たに考案しており、その新

(別紙様式 3)
(Separate Form 3)
規性ならびに有効性も高い。人物の動作認識は社会的要請も高く、本論文で得られた知見は学術的な意義も高い。本論文の根幹部分は、査読付き論文誌である電子情報通信学会論文誌にてすでに公表済みであり、学術的にも評価されている。このように、本論文の人物動作認識技術に関連する学術的・社会的貢献は少なくないと考えられる。
以上に基づき審査した結果、本論文は学位を授与するのに十分なレベルであるものと判定した。