

氏 名 Md.Hasanuzzaman

学位（専攻分野） 博士（情報学）

学位記番号 総研大甲第 952 号

学位授与の日付 平成 18 年 3 月 24 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 Vision and Knowledge-Based Gesture Recognition for
Human-Robot Interaction

論文審査委員 主 査 教授 上野 晴樹
助教授 古山 宣洋
助教授 後藤田 洋伸
教授 佐藤 真一
教授 白井 良明（立命館大学）

Recently, human-robot symbiotic systems have been studied extensively due to the increasing demand of welfare service for the aged and handicapped under the situation of decreasing of the younger generation. In the future, it will be difficult to provide help to the aged and disable persons such as taking care, nursing, informing important information, recreation, etc., by trained human. To build human-robot symbiotic society, where robots are able to support elderly or disable people, the robot should be capable of recognizing user, gesture, gaze, speech and text commands.

To use gestures is a natural way for human to interact with robot. However, gestures are varying among individuals or varying from instance to instance for a given individual. The hand shape and human skin-color are different for different persons. The gesture meanings are also different in different cultures. A significant issue in building visual gesture-based human-robot interactive system is to utilize user-specific knowledge for gesture interpretation and adapt to new gestures and users over time.

In this work, we propose a vision and knowledge-based gesture recognition system for human-robot interaction. In the proposed method, a frame-based knowledge model is defined for the person-centric gesture interpretation and human-robot interaction. In this knowledge model, necessary frames are defined for the known users, robots, poses, gestures and robot behaviors. The system first detects a human face using a combination of template-based and feature-invariant pattern matching approaches and identifies the user using the eigenface method. Then, using the skin-color information of the identified user three larger skin-like regions are segmented from the YIQ color spaces, and face and hand poses are classified by the subspace method. The system is capable of recognizing static gestures comprised of face and hand poses, and dynamic gestures of face or hand in motion. It is implemented using the frame-based Software Platform for Agent and Knowledge Management (SPAK). Known gestures are defined as frames in SPAK knowledge base using the combination of face and hand pose frames. If the required combination of the pose components is found then corresponding gesture frame will be activated. Dynamic gestures are recognized using the state transition diagram of face poses in a sequence of time steps. The person-centric interpretation of gesture is achieved using the frame-based approach and the interaction between humans and robots is determined using the hierarchical frame-based knowledge model.

In this work adaptation method for new users, hand poses, gestures and robot

behaviors by combining computer vision and knowledge-based approaches is proposed. In this method one or more clusters are built to learn new user or new pose. New robot behavior can be learned according to the generalization of multiple occurrence of same gesture with minimum user interaction. An experimental human-robot interaction system comprised of an entertainment robot 'Aibo' and a humanoid robot 'Robovie' has demonstrated the effectiveness of the proposed method. In this research, we also compare three pattern-matching approaches for face and hand poses classification: general PCA, person-specific subspace and pose-specific subspace methods. In the pose-specific subspace method, training images are grouped based on pose and eigenvectors for each pose are generated separately. In the person-specific subspace method, hand poses are grouped based on each person and for each person one PCA is used.

This dissertation makes three major contributions in the field of human-robot symbiotic systems. First, the proposed system utilizes user specific knowledge for person-centric gesture interpretation that supports cultural variation of gesture meanings. The knowledge-based approach allows user to define or edit gestures and robot behaviors frames easily using SPAK knowledge Editor. The segmented skin regions are more noise free for known person because of using person-centric threshold values for the YIQ components. The performance of person-specific subspace method is better than the general PCA method in the same environment for face and hand pose classification.

Second, this thesis proposes a new user and hand pose adaptation method using a multi-cluster based interactive learning method, which supports different orientations of the images of a person's face and different orientations of same hand poses.

Third, this work combines several system components as an integrated system by means of a knowledge-based software platform SPAK. The system components include vision based face and gesture recognizer, knowledge manager, and various robot software and hardware components. We have demonstrated that it is simple and intuitive to develop a natural human-robot interactive system by integrating computer vision with a knowledge-based platform in the internet-based distributed environment.

The ability to understand hand gestures and person-centric meanings will improve the naturalness of human interaction with robot, and allow the user to communicate in complex tasks without using tedious sets of detailed instructions. We believe that integration of vision and knowledge-based approaches is a major step towards achieving intelligent and natural human-robot interaction.

論文の審査結果の要旨

ロボット関連技術の発達や社会の高福祉への要望により、人とロボットの共生システムの研究が盛んになってきた。また、若年層の減少は、人による福祉サービスを困難にしつつある。したがって、福祉サービスロボットへの期待は高まっているが、障害者や高齢者は機械の操作が不得意であったり、不可能である場合が少なくない。人の対話においても、身振りや手話等のジェスチャーを用いることが多く、人とロボットとの対話にジェスチャーを用いることは、極めて自然な要求であると考えられる。このようなことが本研究の背景である。一方、ジェスチャーには個人差があり、皮膚の色にも個人差や人種差がある。また、ジェスチャーには文化の違いによって意味が異なるという問題もある。人間はこれらの問題を知識や学習によってカバーしている。

本研究の目的は、コンピュータビジョンと知識ベースを融合して、ジェスチャーによる人とロボットとの自然な対話システムの概念と手法を提案し、実証実験によって有効性を示すことであるとしている。提案された方法では、知識モデルはフレームモデルを基盤にしており、人に関する情報や知識を用いた画像処理やジェスチャーの解釈、人とロボットの対話およびロボット操作コマンドの生成やセンサーベースの制御のために必要な知識が体系的に定義され、利用される。

本論文で提案されている方法は概略以下のようなものである。ジェスチャー認識は、テンプレートマッチング法に基づいており、先ずユーザの顔が eigenface 法によって識別されると、知識ベースからそのユーザの皮膚色情報等を用いて顔領域や両手領域の切り出しを行う。この結果が、部分空間法とテンプレートマッチング法の組み合わせによって分類され、認識される。認識されたテンプレートクラスにリンクされている知識情報を使って、意味と意図が理解され、対象とするロボットの発話が生成されたり、行動のためのコマンドが生成され、ロボットが制御される。

また、本研究では、各種の処理モジュールを統合して共生ロボットシステムを実現するために、分散知識処理プラットフォーム SPARK を活用するための概念と手法が提案されている。既知のジェスチャーは、顔と手の姿勢フレームのコンビネーションでフレームとして定義され、参照され、必要なとき活性化される。動的なジェスチャーは、時間ステップのシーケンスの中で顔姿勢の状態遷移図を使用して認識される。更に、アダプテーション法と呼ぶ学習法が提案されている。この方法により、システム自身がトレーニングを行ってパターンクラスターや知識を獲得することが出来る。本方法の有効性は、実験システムにより実証されている。

本研究は次世代共生ロボットに重要な貢献を行っているが、それらは次の3点に集約される。第1に、知識を活用することによって計算コストを抑え、信頼性の高いコンピュータビジョンアルゴリズムを実現し、更に背景にある文化的意味に基づいた理解を可能とするフレームワークである。これは様々な人々との共存を想定したとき極めて有用なモデルであるといえる。第2に、新しいユーザやハンドポーズのパターンクラスターや知識を自律的に獲得する対話型学習モデルとしてアダプテーション法を提案している。これによりオリエンテーションが容易となり、新しいユーザやポーズへの対応性が格段に向上している。第3に、分散知識処理プラットフォームによって色々

なソフトウェアモジュールを統合システムとして実現する方法を提案している。共生ロボットは多くのモジュールの融合体であるが、ここで提案されたフレームワークは、この課題の解決に有用な手段となろう。

審査委員会は、本研究が次世代ロボット分野に重要な学術的かつ技術的提案を与えていると判断し、博士論文として十分な内容であると認めた。