

氏名	AYE SU PHYO
博士の専攻分野の名称	博士（学術）
学位記号番号	博理工甲第 1141 号
学位授与年月日	令和元年 9 月 20 日
学位授与の条件	学位規則第 4 条第 1 項該当
学位論文題目	Human-Robot Interaction System Based on Natural Calling Gestures (自然な手招きジェスチャによる人間とロボットのインタラクションシステム)
論文審査委員	委員長 教授 久野 義徳 委員 教授 小室 孝 委員 准教授 小林 貴訓 委員 教授 島村 徹也

論文の内容の要旨

With improved living conditions due to the continuous development of society, people are living longer. Nowadays, many countries in the world are faced with this aging of their populations, which needs help and care. Thus, there is a growing necessity for new technologies that can assist the elderly in their daily living. As one solution, service robots, may even replace home care staff (e.g. a caregiver) to take care of the elderly.

Several robotic research projects are motivated by the challenges of this ageing population with the aim of covering both care and social needs by designing service and / or companion robots. Assistive service robots may be an interesting option to support elderly people's independent living at home especially for those frail elderly people who are chronically affected by several pathologies, with instable health conditions and declining capacity to cope with stressors. Service robots can help people to do some homework, and reduce the workload of caregivers, in some cases, they are useful to improve the quality of life of the old and the disabled. To make service robots be widely used in practice, it is a very necessary job to achieve natural communication between humans and robots. Unlike in industrial robots' field where only expert operators can communicate and operate the robots, the operators are ordinary people in service robots' field, such as the old, the disabled and autistic patients, etc., and many of them have some cognitive impairment or a certain weakening ability. Thus, it is valuable to study the interaction mechanism between humans and robots.

Human Robot Interaction (HRI) is an active research field from many different points of view: from making humans understand the robot states through verbal and non-verbal communication to doing it the other way around, making the robot understand humans. One of the main aims of HRI is to design interfaces for better operation and interaction with robots. Although verbal communication tends to be primary in human-human interactions, nonverbal behaviors such as eye gaze and gestures can convey mental state, augment verbal communication, and reinforce what is being said. Gesture interfaces may make interaction with robots more attractive and friendly to older users because they are natural and intuitive, they require minimal learning time and they lead to a high degree of user satisfaction. This dissertation

proposes a human-robot interaction system based on hand gesture recognition.

At present, although the study of hand gesture recognition has made great progress and achieved high recognition rate in different areas, it is still facing many challenges: such as illumination changes and the background-foreground problem, where objects in the scene might even contain skin-like colors. Another problem related to the background occurs in the situations where a crowd of people exist and move around with many hand motions. Another important problem is that most existing works require that the patterns of gestures are predefined for specific instructions. Based on these challenges, we propose an algorithm that works in crowded environments with randomly moving objects. In real situations, the user may perform gestures in various positions and the environment may also have many people with hand motions. We propose natural hand calling gesture recognition using skeleton features in crowded environments for human-robot interaction. We make the observation that if the person does not have any intention to call the robot, he/she may not move his/her arm against the gravity. When a person calls someone, it is natural to direct his/her hand with an open towards the target person. Based on these findings from observations of people's daily activities, we detect calling gestures based on the gaze, skeleton key-points, coordinate information of the hand-wrist, and fingertips.

In conventional studies, they often assume that the user is in front of the robot. We cannot assume this. Thus the robot should find a person who may want to call the robot. Firstly, we use OpenPose developed at Carnegie Mellon University to extract the body and hand key-points of individuals in the scene. Our work detects gaze and finds hand-wrist positions to reduce the need to track all the people in the scene. And our approach only uses the key-points of people who gaze towards the camera and have clearly defined wrist positions: the position of the wrist is higher than the position of the elbow position and the position of the elbow is higher than the position of the shoulder. By detecting gaze and hand-wrist positions, we can reduce tracking of the other hand motions. There are times when we may miss the hand part in a crowded scene because the hand area is too small. To solve this, we zoom in the hand part based on the locations of body key-points. After zooming in the hand-wrist part and extracting detailed key-points of the fingertips, the system calculates positions of the fingertips. Finally, depending on the positions of the fingertips, our approach recognizes calling gestures.

The proposed method does not need for the users to remember what gestures they should make. It can recognize any natural hand gestures that may appear when humans would like to call others. In addition, it can work in cluttered environments even with many other humans.

We tested the proposed method in video under different conditions such as in case with one person to over four people who are sitting and walking around. The recognition results confirmed the effectiveness of our approach.

Then, we developed a mobile robot system by putting a small humanoid robot on a mobile robot base. We performed experiments of the robot in an actual care facility. Experimental results verified that old people can call the robot by hand gestures while they were only told that they can call the robot by hand gestures. We did not instruct them how to move their hands in advance.

論文の審査結果の要旨

当論文審査委員会は、当該論文の発表会を令和元年7月22日に公開で開催し、詳細な質問を行い論文内容の審査を行った。その論文発表を含む学位論文の審査の結果、本提出論文を博士（学術）の学位論文として合格と判定した。以下に審査結果の要約を示す。

本提出論文は、高齢者の介護ロボットなど人間の身の回りにいてサービスを提供するロボットと人間のインタラクションに関して検討したものである。高齢化社会の進行に伴い、高齢者の活動を支援するロボットの実現への期待が高まっている。介護の作業は多様であり、それを実行できるようにするためにはロボットの機構や制御の研究が必要であるが、身の回りにいるロボットに対しては、まず、ロボットを呼んで支援をして欲しいということを伝えることが必要である。ロボットを呼ぶには専用のリモコン等を使うことも考えられるが、だれもがいつでも使えるようにするという事を考えると、人間同士の場合のように音声や手招きなどのジェスチャで呼べるようにすることが望まれる。音声とジェスチャではどちらかが優れているということはなく、状況に応じて使い分ければよいと考えられる。そこで、本研究では、このジェスチャによるインタラクションについて、だれでも簡単に手招きでロボットを呼ぶことができるようにする技術を検討している。

本論文は6章からなる。まず、第1章では、上で述べたような本研究の背景、研究目的について述べている。各種のロボットを分類し、介護ロボットのように人間の身の回りにいるサービスロボットについては、人間とのインタラクションの向上が特に必要なことを議論している。

第2章では、関連研究を調査し、サービスロボットとのジェスチャによるインタラクションについて解決すべき問題点を明らかにしている。多くの介護ロボットが研究されており、人間とのインタラクションの手段としては、タッチパネルによるものや音声によるものなどが提案されている。それらは、もちろん有効であるが、人間同士の場合は、ちょっと目を合わせて手を少しあげれば相手が来てくれるというように、非言語的行動によるインタラクションも重要である。そこで、ジェスチャによるロボットへの指示の研究も行われている。しかし、従来の研究では、指示をする人間がロボットの前にいるという状況を仮定しているものが多かった。すなわち、ロボットはその前の人間に手の動きがあれば、それを指示として認識を試みればよいということになる。また、その手の動き等によるジェスチャは、どのようなジェスチャがどのような指示かを、あらかじめ取り決めてあった。すなわち、ロボットを使用する人間は指示のジェスチャを覚えておく必要があった。本研究は、この2つの問題を解決し、介護施設等を巡回するロボットに対して、事前に取り決めのない、自然に行われるジェスチャで、呼び止めることができるシステムを提案するものである。

第3章では、自然な手招きジェスチャの認識法を提案している。認識にはカーネギーメロン大学で開発された深層学習を利用した人体の姿勢認識法である OpenPose を活用している。ロボットに用があって、ロボットを呼ぼうとする人はロボットの方を見るはずである。そこで、ロボットは広視野のカメラの画像から OpenPose により人を検出し、その顔の上の特徴点から視線の方向を求め、それがロボットの方を向いている人がいないかを調べる。そのような人がいた場合は、OpenPose で得られる腕の上の特徴点を調べ、手が上に挙げられているかを調べる。これは、人を呼ぶという意識的動作をするなら、手が下がっているという通常の状態から変化するだろうという仮定に基づく。そして、もし、上に挙げられていたら、手の部分を拡大した画像を入力する。OpenPose には全身の画像から全身の姿勢を求めるモードと、手の部分の画像から、手指の姿勢を求めるモードがある。この後者を入力した画像に適用する。そして、手のひらと指の関係を調

べる。手で何らかの作業をしているような場合は指が閉じられるが、人を呼ぶときは一般に手は開いている。以上のようなヒューリスティックルールにより、事前に取り決めておかなくても、ロボットを呼ぶジェスチャを認識できる。

第4章では、移動ロボット台車の上に、小型の人型ロボットを乗せたハードウェアを準備し、実際の介護施設等で実験の行えるロボットシステムを開発している。ロボットは、室内を動きながら、前章で述べた方法でロボットを呼ぶ人を見つける。そして、その人の位置を画像から求め、その人のところへ自動的に移動するようになっている。

第5章では、ジェスチャ認識とロボットシステムを実験により評価している。ジェスチャ認識については、「ロボットを手で呼んでください」という指示だけでロボットを呼んでもらって集めたデータ、および多くの人が動いているところで、人を呼ぶのと紛らわしいと思われるような手の動きのデータを集めて、認識実験を行い、ロボットを呼ぶ動作を認識できることを確認した。ロボットシステムについては、実際の高齢者介護施設で実験を行った。高齢者には「ロボットを他の人にすると同様に手のジェスチャで呼んでください」という指示だけを行い、ロボットを呼んでもらった。定量的な評価には至らなかったが、特に整備を行わない普通の状況の介護施設内で、自然なジェスチャでロボットを呼ぶことができることを確認した。

最後に第6章で、全体を総括し、今後に残された課題を議論している。

今回の研究では、ロボットを呼ぶという部分に絞って検討している点に限定がある。他のジェスチャの認識への拡張は今後の課題だが、ジェスチャが特に有効なのはこのロボットを呼ぶところであり、ロボットが呼ばれて来たら、その後の詳細な依頼には音声に適していると思われる。この点を考えると、本研究は介護ロボットに限らず、さまざまなサービスロボットに必要な基盤技術を開発したものと評価できる。

以上のように、本論文の内容は、学術的に意義のある研究であると判断できる。よって、当学位論文審査委員会は、本論文を博士（学術）の学位論文として合格と判定した。