

視聴覚情報の統合に基づくアクティブインタラクションに関する研究 (031403009)

Active Interaction Based on Integration of Audio-visual Information

金子 正秀 電気通信大学 大学院電気通信学研究科
Masahide Kaneko, The University of Electro-Communications

長井 隆行 電気通信大学 大学院電気通信学研究科
Takayuki Nagai The University of Electro-Communications

研究期間 平成 15 年度～平成 17 年度

概要

本研究開発課題では、コンピュータやロボットと人間との間のインタフェースに関し、人間にとってより自然で負担が少なく、かつ、新たな機能を提供する方法、すなわち、視聴覚情報の統合に基づくアクティブインタラクション技術に関する研究開発を行うことを目的としている。具体的には、距離情報、肌色情報（顔・手の抽出に利用）、動き情報、多チャンネルの音響情報等を有機的に結合させ、ユーザを含めた環境内の情報を的確に取得できるようにした。また、人間どうしのコミュニケーションに見られる曖昧さを含んだ時空間指示に対する解釈、ユーザとロボットとの間での注意の共有、ユーザの行動の観測とそれに応じたロボットの動作など、コンピュータやロボットと人間との間のインタラクション形態に関する新たなコンセプトの提案と有用性の確認を行った。

Abstract

This research aims to develop the active interaction technology based on the integration of audio-visual information which provides more natural and human-friendly interface between humans and computer/robot. Firstly we have studied the methods to capture necessary information from the surroundings by combining multimodal information such as range, skin-color, motion and audio information. Then new human-computer/robot interaction styles have been proposed and their usefulness has been confirmed through experiments using real robots. Examined interaction styles are interpretation of spatiotemporal instructions including ambiguousness, sharing of attention between users and robot, control of robot's action based on observed user's behavior, and so on.

1. まえがき

コンピュータに加えて、人間との知的なインタラクションが可能な人間共存型のロボット(知能ロボットと称する)が、家庭や職場に入りつつある。コンピュータは最早単なる「計算する機械」ではなく、ユーザとの様々な形での係り合いの可能性を持っている。知能ロボットでは更に、移動、動作を伴うという点で、3次元の実環境の中でのユーザとの係り合いが重要になる。この様な状況下において、コンピュータや知能ロボットとユーザとの間のインタラクションは、人間が日頃職場や家庭で他の人間を相手にして行っているのと同様なインタラクションであることが望まれる。しかし、従来のマンマシーンインタフェースにおいては、人間同士のコミュニケーション形態に比べると著しく柔軟性に欠け、また、ユーザとのインタフェースのチャンネルも限られたものである。

本研究開発課題では、コンピュータやロボットと人間との間のインタフェースに関し、人間にとってより自然で負担が少なく、かつ、新たな機能を提供するインタラクション方法の開発を目的とした。この様なインタラクションを本研究では「アクティブインタラクション」と名付けた。具体的には、(1) 視聴覚情報の統合的利用による、ユーザを含めた環境内の情報の的確な取得、及び、(2) 人間同士のコミュニケーションに見られる曖昧さを含んだ時空間指示、注意の共有、マルチモーダル情報の利用、優先度の算出に基づく複数ユーザの中からのコミュニケーション相手の選択、といった形態を反映した新たなインタラクションの枠組みの構築、の2つの観点から研究開発を行った。

2. 研究内容及び成果

アクティブインタラクションのための基礎技術として、以下の研究成果を得た。

(1) 視聴覚情報の統合による周囲環境のセンシング技術

ロボット外周の円筒面上に配置した複数のマイクロフォンで収集した音データに基づいて、全周方向に対する音源位置推定を行える様にした。同時に複数の音源の位置を推定でき、移動音源にも対応可能である。更に、画像情報に基づいて作成した肌色尤度マップを音源位置推定結果と統合することにより、音声・非音声の識別、音源がユーザの口元である場合の位置推定精度の向上を実現した。

(2) 複数ユーザとロボットとの間での共同注意の形成

複数ユーザによる会話シーンを対象として、話者交替の検出、話者の顔姿勢の推定と誘目度マップの作成、誘目度マップに基づく話者の注意対象の検出を可能とした。

(3) 物体認識・学習ロボットにおける視聴覚アルゴリズム

画像による話者情報を用いた音声分離の精度向上、話者モデルの自律的学習、ロボットの身体性(腕)を利用した物体の学習、等の手法の開発を行った。

(4) ユーザの行動の理解

動画像及び距離情報として観測したユーザの日常動作に対して意味付けを行う方法を研究した。この際、ユーザと環境内のオブジェクトとの関係による拘束を利用して、動きの曖昧さの解消に役立てる様にした。

(5) Mind Reading システム (人間の心を読むメカニズム)の基礎技術

被験者への刺激と生体信号の測定に基づいて両者の対応付けをニューラルネットワークにより学習させ、ロボットに相手の心的状況を推定させる、等の基礎検討を行った。

次に、アクティブインタラクションの具体的な形態として、以下の形態について実証実験を行い、有用性を確認した。

(6) 曖昧さを伴った時空間指示に基づく対象物の探索に関わるインタラクション

ユーザからの曖昧さを含んだ時空間指示の取扱い方法を検討した。実際に、実世界中の本棚を対象に、ユーザが曖昧さを含んだ言葉で位置を指示した本をロボットに探

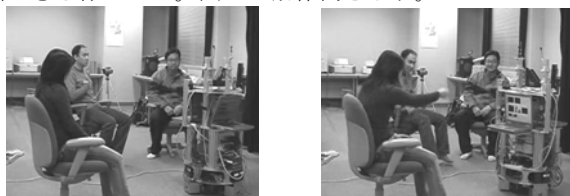
素させるという設定の実験を行った。視界画像からの棚領域の抽出、本の空間的配置に基づいた領域の分割、ロボットからの自発的な行動などを実際にシステムに組み込むことにより、あたかも人間と接しているような感覚でのインタラクションが実現できることを明らかにした。

(7) 相対的位置関係を考慮したユーザとロボットとのインタラクション

ユーザ・ロボット・物体の3者の空間的な相対的位置関係の把握に基づくインタラクション動作の実証のため、ユーザが取得したいオブジェクトの傍にユーザを近づけるように、ロボットが相対的位置関係を随時把握しながら適切な指示を出すというインタラクションを実現させた。

(8) マルチモーダル情報に基づくユーザ毎の優先度を考慮したロボットとユーザとのコミュニケーション

ロボットが複数のユーザとコミュニケーションを行う場合に、どのユーザがロボットに最も注意を向けているかを判断できる様にした。すなわち、一定の時間間隔毎に、ユーザの顔がロボットの方を向いている割合、ジェスチャーを行っている割合、発話の出現頻度、ユーザとロボットとの間の距離から得られる親密さの度合いの情報を統合し、更に各々に重み付けを行って、ユーザ毎の「優先度」を算出できる様にした。図1に動作例を示す。



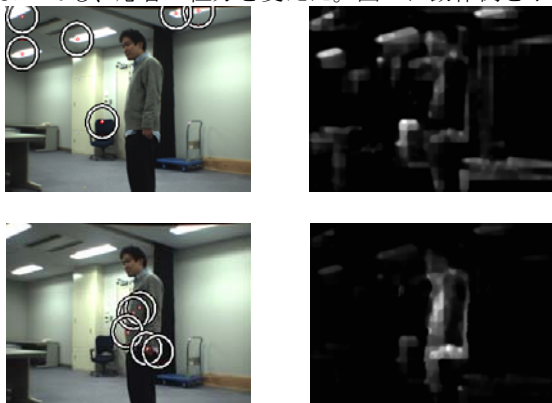
(a) フレーム #47

(b) フレーム #195

図1 複数ユーザの優先度計算に基づくコミュニケーション相手の自動選択(中央のユーザから左側のユーザへロボット(写真右端)のコミュニケーション相手が変わる)

(9) 複数の異なるモダリティ情報の統合に基づく移動ロボットの行動選択

画像特徴情報、シーンの奥行き情報、音源位置推定結果の情報を組合せ、シーンに対してロボットが注意を向けるべき度合いの2次元的な分布を求める。これをロボットの行動状態の情報と組合せることにより、注意先の選択を行えるようにした。この際、相手が人間であるか物であるかによっても、応答の仕方を変えた。図2に動作例を示す。



(a) 入力画像と注意対象の選択結果 (○で表示) (b) 注意を向けるべき度合いを示す2次元マップ

図2 ロボットの行動状態を考慮しない場合(上段)と考慮した場合(下段)の注意先の選択結果

(10) ユーザの仕草に応じたロボットの応答動作

ユーザの顔アクション(ウィンク、あくび等)に応じて、ロボットがユーザに共感し、その結果に応じた動作をユー

ザに返すシステムを構築した。

3. むすび

視聴覚情報はコンピュータやロボットにおいても、外界から情報を得るためにきわめて重要である。本研究では、単なる画像処理、画像認識や音声認識という観点ではなく、ユーザや対象、ロボットが存在する空間を視聴覚情報により認識し、その環境の中から必要な情報を取り出し、ロボットの動作やコンピュータからの応答に反映させることを目指した。これにより、環境からの情報の取得と、それに基づく動作、その結果による環境の変化の取得、等々という連続的なスムーズな流れが実現される。このような技術は、人間のパートナーとしてのコンピュータやロボットの役割を高め、日常生活に活用していくためのキーとなる技術の一つであり、将来におけるより高度でユーザに優しい情報化社会の実現に向けて、重要な意義を持っている。

本研究成果を今後更に発展させることにより、普段人間を相手にしているのと同様な感覚でコンピュータやロボットと情報のやり取りや指示を行うことが可能になり、これらをストレスなく十分に活用することが可能となる。また、人間とロボットが共存する空間が、従来の生産現場だけでなく、教育・娯楽活動までも含む通常の生活環境にまで広がる。すなわち、様々な場面で知能ロボットが物理的・精神的に人間を支援・介助することが期待される。

【誌上発表リスト】

- [1] 陳彬、目黒光彦、金子正秀：“ベイジアンネットワークに基づく視聴覚情報の統合を用いた画像からの3次元音源位置推定,” 電気学会電子・情報・システム(C)部門誌, vol.124, no.3, pp.720-728, 2004.3.
- [2] 陳彬、金子正秀：“複数の異なるモダリティ情報の統合に基づく移動ロボットの行動選択,” 電気学会電子・情報・システム(C)部門誌, vol.125, no.5, pp.765-773, 2005.5.
- [3] Takayuki Nagai and Truong Nguyen：“Appearance Model-Based Face-to-Face Transform,” Proc. of IEEE Int. Conf. on Acoust. Speech & Signal Proc. 2004, vol.5, pp.749-752, 2004.5.

【申請特許リスト】

- [1] 金子正秀、今井順一、高橋哲史：“肌色領域検出装置及び肌色領域検出方法,” 日本, 特願 2006-77692, 2006.3.20.出願
- [2] 金子正秀、今井順一、崔元杰：“ロボット装置及びそのコミュニケーション方法,” 日本, 特願 2006-77693, 2006.3.20. 出願
- [3] 今井順一、金子正秀：“物体追跡装置、異常状態検知装置及び物体追跡方法,” 日本, 特願 2006-95587, 2006.3.30. 出願

【受賞リスト】

- [1] 立木翔一:情報処理学会第68回全国大会学生奨励賞、「ロボットとユーザとのインタラクションにおけるロボット・ユーザ・オブジェクトとの相対的位置関係の取扱い」の発表に対して受賞, 2006.3.10.
- [2] 崔元杰:情報処理学会第68回全国大会学生奨励賞、「複数ユーザとコミュニケーションするロボットにおけるマルチモーダル情報に基づくユーザ毎の優先度判定」の発表に対して受賞, 2006.3.10.

【報道発表リスト】

- [1] 日経産業新聞：“ロボ君「これは何だろう」 見て、つかんで、覚える 電気通信大で試作,” 2005.6.23.
- [2] テレビ東京ワールドビジネスサテライト(トレンドたまご):“興味津々ロボット,” 2005.7.13.