

[研究助成 (C)]

深層学習による身体動作予測に基づく遠隔会議システムの開発と評価

Development of Remote Conference System with Body Motion Prediction Function
Based on Deep Learning Method

2207009



研究代表者	北陸先端科学技術大学院大学 知識科学系 (助成受領時：筑波大学 システム情報工学研究科 博士後期課程)	助 教	王 子 洋
共同研究者	筑波大学 システム情報系	助 教	川 口 一 画
	筑波大学 システム情報工学研究科	博士前期課程	野 秋 裕 己
	東京大学 情報理工学系研究科	教 授	葛 岡 英 明

[研究の目的]

遠隔会議とは、離れる場所にいる複数人がビデオ通話などのシステムを使用し、ある議題に対して意見交換・意思決定を行うことである。そのうち、本研究では一人の遠隔参加者と複数人の現地側参加者の遠隔会議に着目した。このような遠隔会議では、遠隔参加者がウェブカメラを使用し、会議に参加することが多い。しかし、現地側参加者が会議室を歩き回って、実空間の物体を他者に示す参照的動作を行っている際に（例えば、会議室のホワイトボードに近づき、ホワイトボードの上にある資料を指差ししながら話す）、遠隔参加者はウェブカメラの狭いFOVなどの原因により、参照的動作を出す人を把握できず、さらに、その人の会話内容やニュアンスが理解できないことがある。一方、遠隔参加者の存在感（社会的プレゼンス）が通常より薄いので、このような問題を生じた時、現地側参加者は気付かなくて、遠隔参加者をサポートしないことが多い（例えば、ウェブカメラの方向を調整する）。そのまま会話を続けると、意見交換の効率と会議の品質が下がる恐れがある。

これを解決するために、本研究では、参照的動作を自動的に検出または予測し遠隔参加者に提示することでより円滑かつ効率的な遠隔会議を実現することを目的とした。本研究の実施にあたって、主に二つの課題を行った。

1. 適切な参照的動作の提示方法の調査
2. 参照的動作を検出/予測するモデルの構築

[研究の内容、成果]

課題 1. 適切な参照的動作の提示方法の調査

本課題では、検出/予測された参照的動作の適切な提示方法と提示内容を検討する調査を行った。

まず、遠隔会議システムの開発について、「Kubi」という商用テレプレゼンスロボットを用いて、遠隔参加者が「Kubi」にあるタブレットのカメラの方向を変更することで、自らの視点を自由に操作できる機能を備える多人数遠隔会議システムを開発した。その後、検出/予測された参照的動作の適切な提示方法と適切な提示内容を調査するために、以下の機能を実装した。

参照的動作の提示方法は、自動移動法と

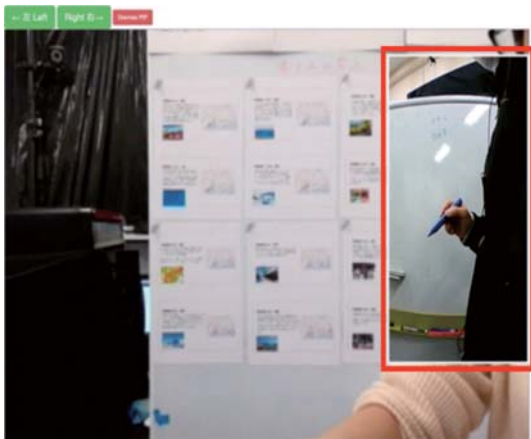


図1 PIP 提示方法の様子

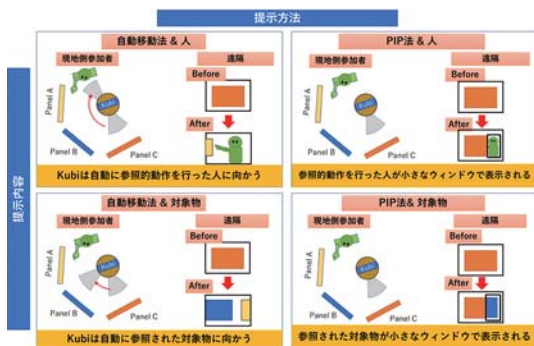


図2 四種類の条件

Picture-in-Picture 法 (PiP 法) の 2 種類とした。自動移動法は、現地側参加者がある物体を指している際に、「Kubi」が自動的にその人またはその物体に向かうように回転することにより、遠隔参加者の画面に直接反映する方法である。一方、PiP は、参照的動作を発生した時、超広角カメラで記録した画像から対象者または対象物をクロップして、遠隔参加者の本来の画面に小さなウィンドウを設け、そこで表示する方法である (図 1)

一方、参照的動作の提示内容は、参照的動作を行った人と参照的動作の対象物の 2 種類とした。そして、この四種類 (2 種類の提示方法と 2 種類の提示の組み合わせ) の機能の効果を調査するために (図 2)、実際の遠隔会議を模擬する実験が行われた。

一回の実験では、遠隔参加者 1 人と現地側参加者 2 人のグループは旅行計画のタスクを実施した。実験参加者らは交通時間を考慮しながら、



図3 実験の様子

10～15 箇所の観光名所から希望の観光名所を選択して 10 時間の旅行を計画した。また、実験参加者は個別の目標が提供されたので、自分の目標をなるべく達成するように他の参加者と交渉することにした。なお、実験参加者がほかの実験参加者の目標が分からなかった。15 分経過後、旅行の計画が完成していなくてもタスクを終了させ、直ちに質問紙の回答を行った。質問紙の内容はシステムのユーザビリティ、コミュニケーションの質、およびグループディスカッションのコンセンサスが含まれており、合計 36 問があった。

このあと、10 分間の休憩を挟み、次のタスクを実施した。同じグループはこの旅行計画のタスクを 4 回実施し、遠隔参加者は上述の 4 種類の参照的動作の提示機能を備えた遠隔会議システムを介し現地側参加者と会話した (図 3)。

結果は、遠隔参加者には自動移動法と比べ PiP 法のユーザビリティが高く感じた。自動移動法では、ビデオ分析により遠隔参加者が意図しない視点移動に戸惑う発言 (「故障している」「動かされている」「嫌がらせを受けている」等) が確認でき、視点の自動移動がユーザビリティを低下させていることが考えられる。また、参照的動作を行った人を提示すると比較し、参照的動作の対象物を提示した方は、ユーザビリティが高い傾向にあった。この要因は現地側参加者がパネルに注視誘導をした際に、遠隔参加者が視野を現地側参加者からパネルまで移動させる必要があったことが考えられる。

また、会話への参加機会について、PiP 法と参照的動作を行った人の条件は、自動移動法と

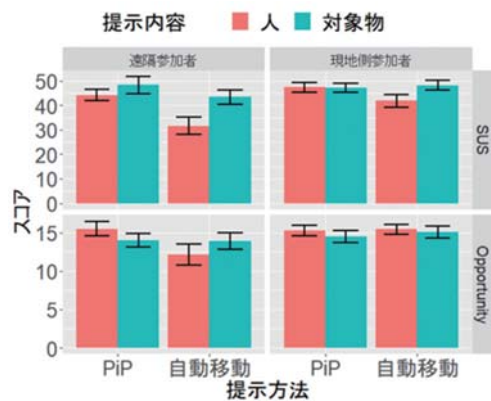


図4 課題1の結果

参照的動作を行った人の条件より遠隔参加者は高く感じた。このことから、PiPの方が部分的に会話の円滑性が高いことが示された。

総合的に考慮した場合、PiP法と対象物の条件は最も会話が円滑性であったことになる。これらの機能により、遠隔参加者は現地側参加者の参照的動作を見逃すことなく、遠隔会議にスムーズに参加することができる。

課題2. 参照的動作を検出/予測するモデルの構築

本課題では、深層学習の手法を用い、現地側参加者の参照的動作を検出するモデルを開発した。

参照的動作を検出/予測するための入力データについて、モーションキャプチャシステムで取得したスケルトンデータはビデオデータと比較してノイズが少ないので、スケルトンデータを使用した。学習データは、旅行計画の模擬実



図5 深層学習用データの収集実験

験から収集された時系列のデータを用いた(図5)。

このデータの使用にあたり、まず、手動で時系列データを動作セグメントに分割し、それぞれの動作セグメントが参照的動作であるかどうかをラベリングした。

その後、持ち上げ、置き、指差し、回転の四種類の動作を検出・予測するLSTMモデルを構築した。学習した特徴は4クラスあり、動作を行った人の3次元の13個のジョイント座標、各ジョイントから隣接ジョイントへのベクターと距離(例えば右肩-右上腕のベクター)、各ジョイント隣接フレームのベクターと距離(例えばフレームNの右肩-フレームN+1の右肩のベクター)、及び各ジョイントから対象物へのベクターと距離であった。

モデルの構造は図6に示した。それぞれの特徴クラスは右半身、胴体、左半身の三つのLSTMモデルを作成し、そして三つのLSTMモデルの出力は一つのLSTMに入力した。さらに、それぞれの特徴クラスのLSTMモデルの出力はConcatenate Layerに導入して最終出力を求めた。

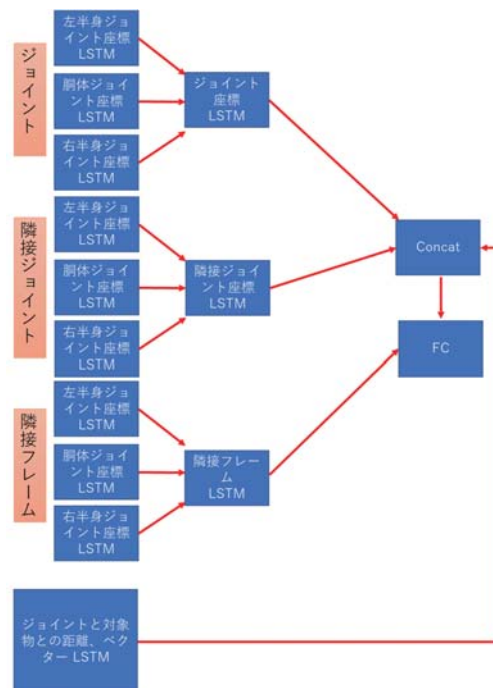


図6 深層学習モデル

本モデルの出力について、動作予測を実現するために、動作セグメントのすべてのフレームを入力した後、可能性が最も高い動作が出力されることなく、各フレームに対して、可能性が最も高い動作を出力するとした。

本モデルの検証について、特徴クラス「各ジョイントから対象物へのベクターと距離」を含めないモデル（モデル 1）とそれを含めるモデル（モデル 2）を比較した。損失関数について、指差しを有効に学習するために、指差しにより高い重みをつけた。その結果、モデル 1 の適合率は 64% で、再現率は 61.54% で F1 スコアは 62.75% であった。一方、モデル 2 の適合率は 54.81% で、再現率は 81.43% で F1 スコアは 65.52% であった。このことから、開発したモデルの有効性が確認され、モデル 2 の方が精度が高いことが分かった。

そして、モデル 2 を課題 1 の PiP 機能を備えた遠隔会議システムに導入して、参照的動作

の検出・予測機能を備えた遠隔会議システムを実現した。

[今後の研究の方向、課題]

この二年間、新型コロナウイルス感染症のため、実験実施することが困難であった。そのため、今後の課題として、システムの有用性を含め、ユーザー行為の変化を調べ、開発した参照的動作の検出・予測機能を備えた遠隔会議システムの評価を実施する。

[成果の発表、論文等]

- ・ Tzu-Yang Wang, Yuki Noaki, and Hideaki Kuzuoka. 2021. Exploring How to Display Referential Action to Support Remote Group Discussion. In Asian CHI Symposium 2021 (Asian CHI Symposium 2021), May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3429360.3468188>