

遠隔会議における同一空間内の会話グループ推定手法の研究 A Method for Estimating Conversation Group in the Same Space on Teleconference

井黒 海星[†]
Kaisei Iguro

大倉 裕貴[‡]
Yuki Okura

岩本 健嗣[§]
Takeshi Iwamoto

1. はじめに

近年、遠隔地にいる相手との音声通話機能や、互いの姿を Web カメラで撮影し共有する機能をもつ遠隔会議システムが広く普及している。また、オフィスの一室や、講義室等の対面環境に複数人が集まり、そこから遠隔会議に参加する会議形式も多く見られる。この会議形式では、遠隔地にいる相手との会話と、対面している参加者の間でのみ行われる会話が並行することがある。この場合、対面している参加者側では、複数人が同時に発言していても目的の話者の発言のみを選択的に聞き取ることが可能であり、活発な会話が行われる [1]。一方で、遠隔地からの参加者は、複数人の音声が一のスピーカーから同時に再生されるため、目的の話者の音声を聞き取るのは困難となる。さらに、遠隔地からの参加者が同時に 2 人以上発言した際には、互いの発言が衝突し、どちらも会話を中断してしまうことが玉木らにより報告されている [2]。これらの問題を解決するため、対面している参加者らを並行している会話ごとにグループ分けし、遠隔地からの参加者が選択したグループに属する参加者らに対してのみ、音声の聞き取りや発言を可能にするシステムが必要であると考えた。そこで本研究では、対面で会話をを行っている参加者の集団を会話グループとして定義し、参加者らを会話グループとして分類・推定する手法を検討する。

2. 会話グループの推定方法

2.1 会話グループ推定に用いる特徴量

会話グループの推定に用いる特徴量の選定を行うにあたり、対面での会議の様子を観察した結果、会話グループには話者が必ず 1 人以上存在し、グループ内の聞き手は話者を頻繁に注視する動きが見られた。また、話者も同様に自分に向けられた聞き手の視線を確認しながら話す動作が見られた。図 1 にその様子を示す。また、会話グループが形成される際には、机を挟んで正対している参加者や、隣り合う参加者と会話する様子が多く見られた。そこで本研究では、会話グループの推定に用いる特徴量として、参加者の顔の向き、および参加者同士の位置関係を用いる。



図 1 話し手に対する聞き手の注視方向

[†] 富山県立大学 大学院 工学研究科 電子・情報工学専攻科
Graduate School of Engineering, Electrical and Computer Engineering,
Toyama Prefectural University

[‡] 富山県立大学 電気・電子工学科

[§] 富山県立大学 情報システム工学科

2.2 相互参加度

設定した特徴量から、会話グループ推定のための新たな指標を定義する。横軸・縦軸がそれぞれ水平・奥行き方向の距離を表す 2 次元座標平面を図 2 に示す。そこに 2 人の会議参加者 i, j を俯瞰するように配置し、会議参加者 i, j の位置を表す 2 点と、それらを始点とする参加者の顔の向きを方向成分にもつ大きさ 1 のベクトルを図示する。特徴量を反映した指標となるよう、相互参加度 P_{ij} の算出方法は以下のように定義する。 l_p は両会議参加者の位置を示す点同士を結ぶ線分、 l_a は両者の顔方向ベクトルの終点同士を結んだ線分である。そして相互参加度 P_{ij} は、線分 l_p および線分 l_a の長さの差をとったスカラー量として定義する。参加者 1 人につき、他の全ての参加者との間で相互参加度を計算し、得られたベクトルをその参加者の相互参加度ベクトルとして定義する。

$$\begin{cases} P_{ij} = l_p - l_a & (i \neq j) \\ P_{ij} = 2 & (i = j) \end{cases}$$

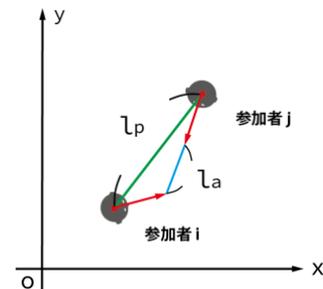


図 2 相互参加度の算出方法

相互参加度の大きさは、参加者同士の距離に関わらず、両者の顔の向きがどれだけ正対しているかにより変動する。表 1 に示すように、相互参加度の値は互いの顔の向きが完全に正対する際は最大値 2 をとり、互いに反対を向く場合は最小値 -2 をとる。

表 1 相互参加度と特徴量の関係

顔の向き	参加者間の距離	相互参加度 P_{ij}
		$P_{ij} = 2$
		$0 < P_{ij} < 2$
		$P_{ij} = 0$
		$P_{ij} = 0$
		$-2 < P_{ij} < 0$
		$P_{ij} = -2$

2.3 提案手法

本節では、設定した特徴量をセンシングする手法、および前節で定義した相互参加度を用いた会話グループの推定手法について述べる。対面での会議では、兆候もなく会話が始まる場合や、会話グループを構成する参加者の入れ替わりや増減などが頻発する可能性がある。故に、会話シーンを動的に分析し、その結果から推定を行う必要がある。そこで、深度カメラを用いて特徴量をリアルタイムに取得し、相互参加度の算出を行なう。また、相互参加度は会議参加者 1 人に対し、他の全ての参加者との間について算出する。故に、参加者の位置情報を合わせた 1 つの多次元データを分析し、会話グループの推定を行う必要がある。本研究ではこれらの要件を満たすグループ推定手法として、各参加者間の相互参加度、位置情報を入力とし、クラスタリングを行うことで会話グループの推定を行なう。また、クラスタリング手法としては、教師データを用いないクラスタリング手法である k-means 法を用いる。

3. 実装

Intel RealSense D415 デプスカメラ[3]を用いて会議参加者の骨格情報を取得する。また、骨格検知には、RealSense用の骨格検知ライブラリ Cubemos Skeleton Tracking SDK を用いた。Skeleton Tracking SDK を用いることで、図 3 左に示すような RealSense のキャプチャ映像上に骨格情報を重畳し、各関節部の深度情報をもとに、実空間の 3 次元座標を取得する。特徴量の算出には、図 3 右のような主に顔周辺の深度情報を利用した。

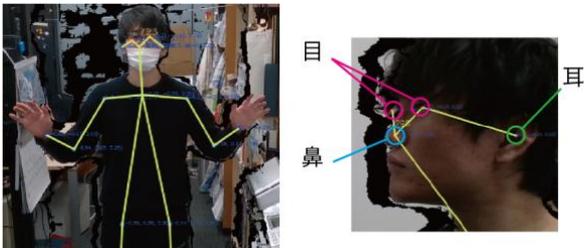


図 3 重畳された骨格情報 (左)、顔周辺の骨格情報 (右)

3.1 参加者の位置、顔の向き取得

本節では、参加者の位置と顔の向きを求める方法について述べる。本研究では、図 4 左のように、両目の中点から鼻の座標へのベクトルを規格化したものを顔の方向として用いている。また、RealSense に対して参加者が真横を向いている場合には、両目の座標のうちどちらかの深度情報が欠損することがある。その場合の導出方法を図 4 右に示す。本研究では、欠損していない方の耳の座標から目の座標への方向を顔の向きとした。

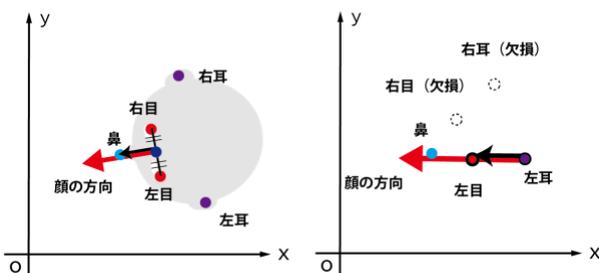


図 4 顔の向き方向の算出

RealSense に対する顔の向きが 45° 斜め方向からの角度を向いている場合、 90° 真横を向いている場合におけるそれぞれの算出方法によって生じる誤差を検証した。誤差の算出方法は、 45° 、 90° とともに 5 回計測し、2 通りの算出方法で得られる顔の向きの角度の差を平均した。結果として、 45° 斜めの場合には 2 つの算出方法で約 18° 、 90° 真横の場合には約 11° の誤差が生じた。

4. 評価実験

提案した手法を用いて、実際の会議に近い環境で、相互参加度がグループ推定に有効であるかを検証した。図 5 のように、8 人の被験者全員を RealSense の画角に入るように机を囲んで座らせ、会議内に会話グループが常に 3 つ存在するように、参加者が会話する相手を指定した。評価方法として、約 1 秒分に当たる連続した 20 フレームのキャプチャ映像を取得し、それぞれにおいて参加者の位置、相互参加度を入力とし、k-means 法によるクラスタリングを行なった。各参加者が正しいクラスターにクラスタリングされた回数を記録した。



図 5 会話グループ推定の評価実験環境

4.1 実験結果

20 回のクラスタリングの結果、参加者 8 人のうち、2 名が 70% (14/20) の精度でクラスタリングされた一方で、最も結果が悪かった参加者は 5 名おり、精度は 25% (5/20) であった。平均では、40.6% の精度となった。

5. まとめと考察

本研究では、対面環境に存在する複数の会話グループについて、参加者がどのグループに属しているかを推定する手法を検討した。評価の結果、精度が 70% でクラスタリングされた参加者がいる一方で、25% の参加者も含む結果となった。精度に差が生まれる要因としては、机を挟んで会話しているグループでは、より顔の向きが正対しやすくなる一方で、会話の相手が隣り合う位置関係にある会話グループでは、顔の向きが正対する頻度がやや低くなり、参加者間での相互参加度が高くないタイミングがあることが考えられる。今後は、相互参加度の算出方法の改善に加えて、RealSense が取得する深度情報がノイズを多く含む問題の解決のほか、マイクを使用して参加者の発話状態を利用することで、会話グループ推定の精度の改善を行う。

参考文献

- [1] 武川直樹, “コミュニケーションにおける視線の役割 – 視線が伝える意図・気持ち –”, 電子情報通信学会誌, Vol.85, No.10 (2002).
- [2] 玉木秀和, 東野豪, 小林稔, 井原雅行, “音声遅延が遠隔会議中の発話衝突と精神的ストレスに与える影響”, 電子情報通信学会論文誌 D, Vol.96 (2013).
- [3] Intel® RealSense Depth Camera D415, new Technologies Group, Intel Corporation, 2019, Revision 1.5.0.