

## インタラクティブ発表会場における

## 発表者支援に向けた参加者の状態認識の検討

## Recognition of Participants' Status for the Support of the Interactive Presentation

松儀 良広<sup>\*1</sup>  
Yoshihiro Matsugi

小川 環<sup>\*1</sup>  
Kan Ogawa

岡田 将吾<sup>\*1</sup>  
Shogo Okada

新田 克己<sup>\*1</sup>  
Katsumi Nitta

<sup>\*1</sup> 東京工業大学大学院総合理工学研究科知能システム科学専攻  
Department of Computational Intelligence and Systems Science,  
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

**Abstract:** Interactive presentations such as the poster presentations are complex communications where audience move around freely. In order to support a presenter to communicate with audience, we have been developing a audience recognition system. By using sensors and TV cameras, this system observes the interactive presentation, and recognizes their positions and their facing directions. We introduce the overview of this method, and show its performance by experiments.

## 1. はじめに

本研究はポスター発表会・企業の展示会のように、流動的に人が動き回るようなインタラクティブな状況で、説明者はなるべく多くの聞き手に十分な説明を行うことが求められる。説明者は、現時点での聴講者だけでなく、ポスターの後方にいる人を考慮して発表する必要がある。しかし、聴講者に説明しながら、後方の参加者の注意を払うのは困難である。

このような状況において、聴講者・説明者の位置を特定し、ある特定の聴講者への説明時間をモニタリングし、後方の参加者の存在を検出し、ユーザにフィードバックすることで、説明者の支援を行うシステムを構築することを目標とする。この目標のために、本研究では、ポスター発表のような移動が自由に出来る会話を自由移動会話と呼称し、本研究では、実際にじゆう移動会話が行われるポスター発表会における会話場をセンシングするためにセンシング環境を構築し、実際のポスター発表会における説明者・聴講者間のインタラクションコーパスを収集する。また支援システム構築の初期検討として、深度情報から、説明者・聴講者の位置情報と向きを推定するためのアルゴリズムを構築し、評価した。

## 2. 関連研究

[河原 12] では、ポスター発表における参加者の相槌や頷き等の非言語行動に着目し、聴講者の興味度や発話内容の認識に取り組んでいる。

この研究では、発表者と聴講者の位置は固定されており、実際の自由移動会話のように、人が流動的に会話に参加する設定では行われていない。

本研究では、実際に自由移動会話が行われるポスター発表会における会話場をセンシング出来る環境を構築し、自由移動会話コーパスを収集する。

[角 11]らはマイクやモーションキャプチャ等の装着型デバイスをつけて、多数会話を記録している。この研究では、発話

だけでなく視線の移動や、指差しといった非言語行動を記録し、会話構造を分析している。

本研究でも、発表者には装着型デバイスをつけて、データを取得する一方、より自然なデータを取得するため、聴講者に対してはデバイスをつけずにデータを取得するため、非接触センサを用いることにした。

## 3. ポスター発表実験のデータ収集

自由移動会話における行動データを実際に収集するために、実際にポスター発表をし、その環境にセンサやカメラを設置し、データを収集する。

参加者の状態認識のための特徴量としては、センサからの深度データやカメラからの映像データを用いる。

ポスター発表会話場では、ポスター付近に人が集まるため、ポスターの上部に深度センサを取り付け、人が存在する領域の上部から俯瞰からデータを取得することで、どこに人がいるか、人とポスターの距離、人同士の距離を容易に認識することができる。さらにポスター前だけでなく、センサを後方部に向けてことによって、後方にいる人の映像や、深度情報を取得することができる。

ポスター発表では、発表者とその他の聴講者に分かれて、コミュニケーションを行う。発表者はポスターの前に立ち、聴講者に対して、説明を行い、必要であれば質疑応答に答えるという形である。

### 3.1 実験参加者と実験設定

実験の被験者として発表者を3人用意した、年齢は24~25、全員男性であり、発表をする内容は各々の修士論文の研究内容である。また聴講者として大学院生と教員である。被験者は、そのフロアを自由に動き回り、興味のあるポスターに近づいて発表者に対して、話を聞く。

連絡先: 松儀良広, 東京工業大学大学院総合理工学研究科  
知能システム科学専攻, 神奈川県横浜市緑区長津田町  
4259 J2-53, TEL & FAX:0298-54-5204,

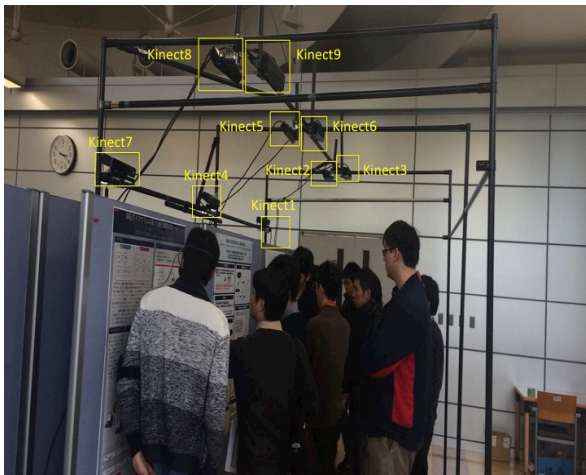


図1:ポスター発表の実験

### 3.2 行動データ

使用したセンサは Kinect センサ\*2 とウェアラブルカメラである。

Kinect センサはそれぞれのポスターにつき、真下、正面、後方の方向に向け、3 台設置しデータを取得した(図 2)。この論文では、便宜的にそれぞれ名前をつけ、その位置を下図に示す(図 3)。

しかし、ポスター2の真下のデータを取得するために設置した Kinect5 はデータ欠損により、データを取得することができなかった。そのため合計8台の Kinect センサのデータを取得した。センサはあらかじめポスター会場に設置された骨組みを組んで作成した槽に設置された。

ウェアラブルカメラは発表者に装着し、映像データと音声データを取得した。

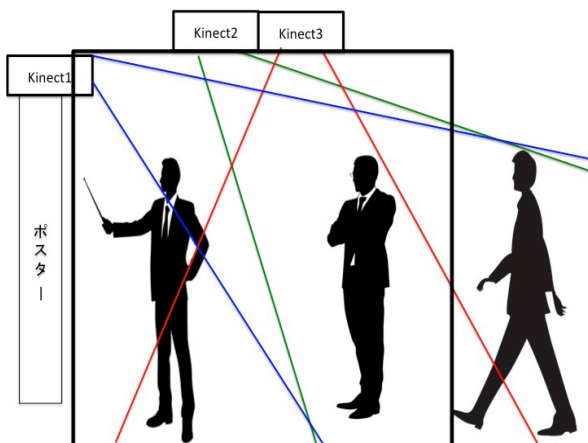


図2:実験環境を横から見た図

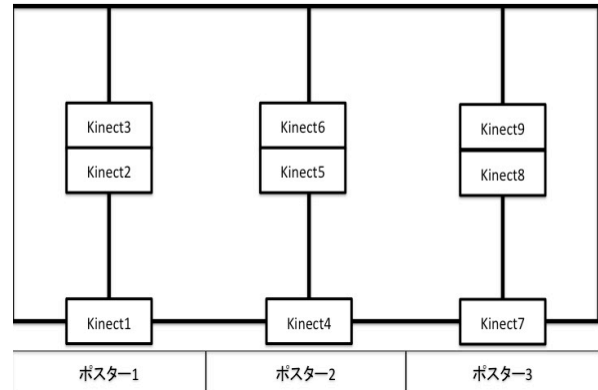


図3:実験環境の俯瞰図

### 3.3 Kinect のデータ

設置した 8 台の Kinect センサから RGB データ、深度データ、骨格データを取得した。RGB データは 15fps で avi 形式で出力し取得した。各 Kinect センサから約 80 分のデータを取得した。今後、画像処理による顔認識や教師データを作る際のデータとして取得された。

RGB データと同じように、8 台全ての Kinect センサから深度データを取得した。深度データは Kinect センサの位置からの距離を示したデータであり、15fps でデータを取得し、png の形式で出力し取得した。深度データは 16bit の画像データで出力され、500mm から 8000mm まで距離を取得することができる。

RGB データと同様に約 80 分のデータを取得可能である。

深度データにより、人の位置情報を取得することが容易になると考えられる。

骨格データは Kinect1, Kinect4, Kinect7 のポスターの正面に設置した Kinect センサから、ポスターの前にいる人間の骨格情報を 15fps で取得した(図 3)。骨格データは体の骨格の 25 箇所の座標を取得した。Kinect の仕様上同時に 6 人のデータを取得するのが限界である。

骨格データにより、ポスター前にいる人のジェスチャー等の非言語情報を取得することを目的とし、取得された。

### 3.4 ウェアラブルカメラのデータ

ウェアラブルカメラを用いて、RGB データと音声データを取得した。RGB データは発表者 3 人視点からの映像データであり、発表者の一人称視点映像データが得られる。Kinect から得られるデータと統合して、発表者の体の向きと、顔の向きの違いなどを抽出するために用いる予定である。また RGB データと音声データは実験の初めから終わりまで約 80 分のデータを取得した。音声データは、今後文字データや韻律データとして使用されることを目的として取得された。

### 4. 参加者の位置・向きの検出

収集されたデータから、参加者の位置情報、向き情報の推定を行った。使用したデータは、ポスター前にいる人を正しく認識する目的から真下に向けられた Kinect8 により取得された深度データを使用した。

\*2 Kinect v2

[http://www.microsoftstore.com/store/msjp/ja\\_JP/pdp/Xbox-One-対応-Kinect-センサー/productID.307394800](http://www.microsoftstore.com/store/msjp/ja_JP/pdp/Xbox-One-対応-Kinect-センサー/productID.307394800)

#### 4.1 参加者の位置情報取得

深度データから位置情報を取得する。位置情報はそれぞれの人が上から見て、どこにいるかを示したデータである。よって、本研究では、位置情報を頭の中心座標と定義する。

まず閾値を用いて深度データを出力することで、Kinect から一定の近さ以内のピクセルを出力する。本研究では、この閾値は 1090mm と設定した。つまり、Kinect からの距離が 850mm 以内の物を頭として認識する。上記の処理を経て、深度データから参加者の頭部領域を抽出した。次に認識された頭部領域に対して、中心座標を求める。

#### 4.2 参加者の向き情報取得

各参加者が向いている方向を取得する。体の向きを取得するために、頭部領域と肩の領域を利用する。

そこで、頭部から肩までの領域を抽出するための閾値(Kinect からの距離 1090mm)を設定し取得した。向き情報を取得する手順は[Draz'en 13]らが用いた手法を参考にした。[Draz'en 13]らはショッピングモールにセンサを設置し、深度データから人の位置情報、向き情報抽出した。

向きを取得する手順としては以下の通りである。

- 位置情報を取得する際と同じように、頭部領域を取得する。
- 閾値を使用して、頭から肩までの領域を取得する。
- 頭部領域からは中心座標を取得する。
- 頭から肩までの領域内で、肩の上で一番長い直線が引ける直線を引く。
- 中心座標から、一番長い直線に垂線を下ろし、その直線を参加者の向きとして取得する。(図4)

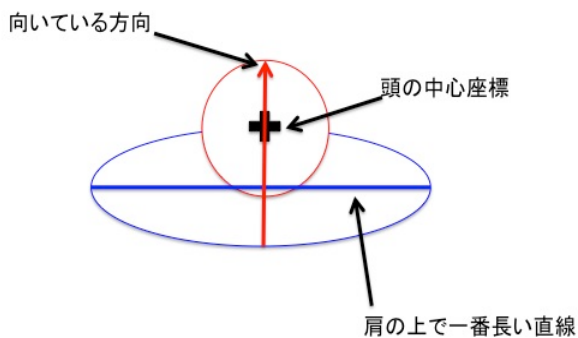


図4: 肩のデータと頭のデータによって向きを求める手法

参加者の向きの推定結果の例を以下に示す。

赤で塗られた領域が、人を示す領域である。緑の直線が人の向きであり、青の直線は肩の上で一番長い直線である。緑の直線上にある黒い点が、それぞれの人の頭の中心座標であり、位置情報とする。(図5)

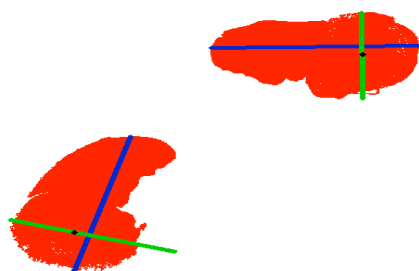


図5: 位置情報と向き情報を取得した例

以上のように、深度データを画像データとして処理することによって、容易に位置情報と向き情報を得ることができる。データフローは以下の通りである。(図6)

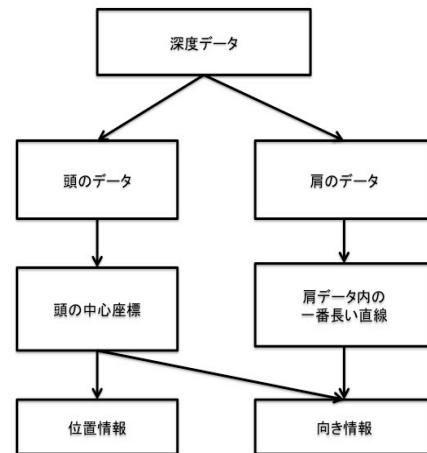


図6: データフロー

上記のアルゴリズムを実行した際、人同士の距離が近く、肩の領域が重なっている時は、頭部領域が 2 つ以上検出された場合にも同一の人として認識されてしまう問題が起きる。

この問題を解決するために、頭部と肩の領域が 1 対 1 に対応するまで、肩の領域抽出のための閾値を変化させる方法を採用した。結果として、参加者のセグメンテーション精度が向上し、位置・向きの推定精度も向上した。

#### 5. 位置・向きの推定精度の評価

参加者の位置情報と向き情報を実際に取得し、50 枚のデータに対して精度の評価を行った。精度の求め方としては、本研究で抽出した位置情報と向き情報を与えた 50 枚のデータに対して、人手で行った正解データと比較して、どのくらいの精度が出るかということである。

位置情報では、頭の位置に点をつける。その点が、深度データから抽出された頭のみ領域に入っていれば正解とする。実験結果は以下のようになった

表1: 位置情報の精度

	精度
位置情報	50/50 100%

向き情報に関しては、位置情報と同じように頭の中心から向いている方向に向かって直線を引く。その直線と、本研究で得られた向き情報の直線との鋭角の角度( $0 \leq \theta \leq 90$ )を求める。その角度の平均を求めることによって、精度とする。結果は以下のようになった。

表2: 向き情報の精度

	角度の平均	最大値	最小値
向き情報	17.88	86	0

角度の平均は  $17.88^\circ$  となった。この値は小さければ正しいので、

最大値は  $86^\circ$  と大きく間違えてしまった。これは被験者が後ろを向いた際に、頭から背中にかけて長い直線が引かれてしまい真横を向いていると誤認識されたからである。

## 6. まとめ

本研究では、ポスター発表における発表者の難点を解決するために、実際に流動的に人が動き回るポスター発表の会場をデータとして採集した。また初期検討として Kinect センサから得られた深度データを用いて、参加者の位置情報、向き情報を取得した。取得された情報の精度としては、位置情報は頭部領域を見つけないという観点で見ると 100%の精度であった。また向き情報に関しては、正解との差の平均が約 18° と小さいものであった。しかし、提案した位置の推定方法では、以下の問題が生じる。

参加者同士が縦(y 軸方向)近いケースにおいては、複数の人間が同一の人として、認識されてしまうケースがあるため、今後は頭を中心座標から深度データのクラスタリングを行い、人領域の分割制度を向上させる。

また、向き情報に関する問題点として、頭のデータと肩のデータを使用して直線を引いたが、直線の端のうち、どちらが向きであるかという情報はまだ決められていない。今後向き情報を割り振るために、以下の手法を検討している。

ポスター正面からの RGB データより顔認識を行う。顔認識の結果を、深度データからの向きの推定結果と統合して、参加者の向きの推定制度を向上させる。

位置・向き情報を統合した参加者役割認識モデルの構築も今後の課題である。

役割認識では、[Clark 96]が挙げた多人数会話における役割を用いる予定である。それらの役割は、話をしている「話し手」(Speaker)、話を聞いている「聞き手」(addressee)、会話に参加はしているが現在話しかけられていない人物である「傍参与者」(side participants)、これらの3つの役割が参与者とされ、その参与者に認識され、その場にいるが会話に参加していないのが「傍観者」(bystander)参与者に認識されずに話を聞いている「盗み聞き人」(eavesdropper)という5つの役割を用いるつもりである。

また発表者支援について、具体的な手法の検討も進める。

## 参考文献

- [河原 12] 河原 達也:「招待講演」スマートポスターボード:ポスター会話のマルチモーダルなセンシングと認識, 電子情報通信学会技術研究報告,SP20120-51,(2012)
- [角 11] 角 康之, 矢野 正治, 西田 豊明:マルチモーダルデータに基づいた多人数会話の構造理解, 社会言語科学会誌, Vol.14, No.1, pp.82-96, (2011)
- [Draž' en 13]Draž' en Brs' c' ic, Takayuki Kanda ,Tetsushi Ikeda and Takahiro Miyashita : Person Tracking in Large Public Spaces Using 3-D Range Sensors , IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS , VOL. 43, NO. 6,
- [Clark 96]H. H. Clark, "Using language", Cambridge University Press,(1996)