

受取人の状態に応じて割込みのタイミングを変化させる配達ロボット

荒井 孝普 高橋 桂太 金子 正秀
 電気通信大学 大学院情報理工学研究科

1. はじめに

お手伝いロボットのように、人間に対して働きかけを行うロボットが世の中に登場し始めている。人間への働きかけの場面では、人間の作業を邪魔して不快感を与えないよう、割込むタイミングを計る必要がある。従来研究として、パソコンでの作業における忙しさをパソコンの操作状態から調べる試み[1]や、不快感の低下を狙って動作と発話のタイミングをずらす方法[2]などが報告されている。しかし、これらの研究ではロボットと人間が1対1の関係にあり、実際の生活場面で見られるような人間同士の会話などの関わり合いが考慮されていない。

そこで本稿では、受取人に荷物を届ける配達ロボットを例として、受取人が他者と会話している可能性を考慮して、割込みのタイミングを計る手法を提案する。会話状態の判断には、受取人と会話相手の関係を考慮したマルコフモデルを用い、会話の「間」を判断し、割込みを行う。

2. ロボットのセンシング・認識機能の実装

2.1. ロボットの構成

ロボットには MobileRobots 社製の Pioneer3DX を用い、受取人とその周囲の環境をセンシングするために、距離画像の取得が可能な TOF (Time of Flight) カメラ、カラー単眼カメラ、及び 5 本マイクから構成されるマイクロホンアレイを搭載した。TOF カメラとカラーカメラでは光軸がほぼ一致していると見なせる。

このロボットは図 1 のような状態を持つ。まず、「視覚探索」状態では TOF カメラによる距離情報とカラーカメラによる色情報を用いて、ロボットの視野内に人間がいるか、いるならばそれが誰かを認識する。次に、「聴覚探索」状態ではマイクロホンアレイを用いて音源位置推定を行い、誰が発話しているかを認識する。

2.2 人物検出

TOF カメラによる距離画像から人物領域を検出する手法を述べる。まず、距離情報の各点を、横軸が水平 (X) 方向に、縦軸が奥行き (Z) 方向に対応した XZ 平面画像上に投票していく。次に、投票画像を 2 値化処理し、物体の存在候補領域を得た。この時に、遠方に行くほど距離情報のサンプルが疎になることを考慮し、奥行き z [m] に応じて、式 (1) のように閾値 T_z を変化させた。

$$T_z = 50/z^2 \quad (1)$$

最後に、物体の存在候補領域にラベリングを行い、人間らしい大きさを持つ領域を人間と解釈した (図 2 の赤色領域)。この領域に対応する領域を、カラーカメラから抽出して人物抽出画像とした (図 3)。

Interruption Strategy for Delivery Robots Aware of Receiver's Condition

Takayuki ARAI, Keita TAKAHASHI and Masahide KANEKO
 The University of Electro-Communications
 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan
 {t_arai, takahashi, kaneko}@radish.ee.uec.ac.jp

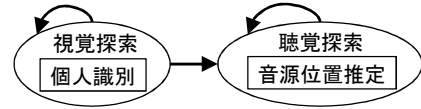


図 1. ロボットのセンシング・認識の流れ

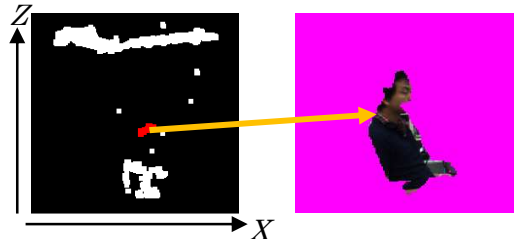


図 2. XZ 平面 図 3. 抽出された人物

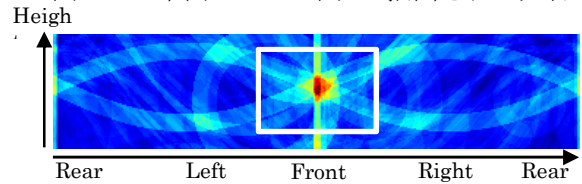


図 5. 尤度マップと TOF カメラの視野

2.3 個人識別

2.2 で得た人物抽出画像について、色情報 (HS ヒストグラム) と勾配情報 (SURF 特徴[3]) を用いて個人識別を行った。

HS ヒストグラムについては、事前に用意した 1 人につき 5 枚のデータベース画像 (以下 DB 画像) から得られたヒストグラムと比較した。最も距離が近い個人の ID を R_{hs} とした。

SURF 特徴については、事前に DB 画像から得た SURF 特徴により学習済の SVM (Support Vector Machine) を用いて、どの人間に所属するか分類した。分類された SURF 特徴の数が最も多い人間の ID を R_{surf} とした。

最終的な識別結果は、SURF 特徴の信頼度が高い時には R_{surf} 、そうでなければ R_{hs} とした。式 (2) を満たせば、SURF 特徴は信頼できるとした。

$$v_1 \geq 0.5 \sum_{i=1}^n v_i \quad (2)$$

ここで $v_1 \sim v_n$ は各人間への SURF 特徴の投票数が多い順に並べたものである。

2.4 話者推定

マイクロホンアレイにより音源位置推定を行うことで、2.3 で識別された人間のうちの誰が発話しているかを推定した。尤度マップ法[4]により、図 5 のようにロボットの全周 360°の音源尤度の画像を得た。この画像中のロボットの正面に当たる部分 (図 5 中の白枠) を切り抜き、閾値処理を行うことで TOF カメラの視野と対応した音源位置画像を得た。音源位置画像と 2.2 および 2.3 の結果を対応させることで、誰が発話したのかを推定した。

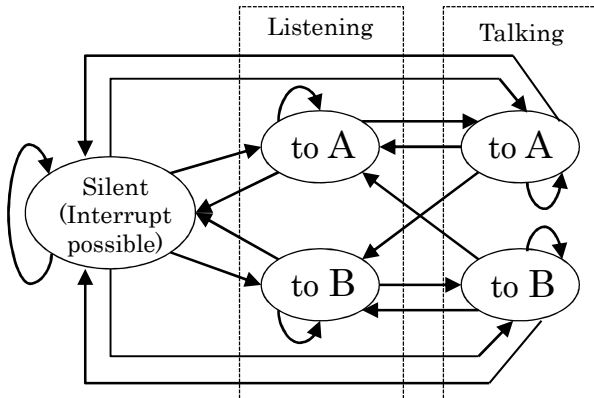


図 6. 会話状態の状態遷移図

表 1. 会話状態モデルの遷移確率例(無音時)

		To				
		S	L-A	L-B	T-A	T-B
From	S	1.0	0	0	0	0
	L-A	*1	*2	N/A	0.1	N/A
	L-B	*3	N/A	*4	N/A	0.1
	T-A	0.1	0.1	0	0.8	N/A
	T-B	0.1	0	0.1	N/A	0.8

$0.4 - 0.4C_A \dots *1$ $0.5 + 0.4C_A \dots *2$
 $0.4 - 0.4C_B \dots *3$ $0.5 + 0.4C_B \dots *4$

3. 割り込み判断のための会話状態推定

ロボットが荷物配達割り込みタイミングを決めるため、受取人が他の人間と会話している状態を、図6のようにマルコフモデルを用いてモデル化した。ただしここでは、受取人に加え人物A、Bの計3人がいる場合を考える。「Silent」(以下「S」と表記)は会話が行われていない状態、「Listening」(以下「L」と表記)は、受取人本人が発話していないものの、AかBが発話を聞いている状態、「Talking」(以下「T」と表記)は受取人がAまたはBに向けて発話している状態を示す。

また、A、Bと受取人との社会的人間関係を C_A , C_B として表す。 C_A , C_B は社会的立場の差や親しさを反映し、会話の重要性が高いほど大きな値になるものとした。これにより、会話相手によって割り込みまでのタイミングを変化させることが出来る。

図6のマルコフモデルに対して、「無音」、「受取人が発話」、「Aが発話」、及び「Bが発話」のイベントに対応する計4個の遷移確率テーブルを作成した。表1は「無音」時のテーブルである。この会話状態モデルを用いて「S」状態確率が0.7を超えた場合、ロボットは割り込みを行うことにした。

4. 予備実験

人物認識、個人識別、及び話者推定をロボットに実装し、実験を行った。実験では、受取人Rと会話相手A、会話に参加しないBの計3人がいる会話環境を記録し、その結果に対して3.の会話状態モデルを当てはめて、 C_A を変化させつつ各状態確率をオフラインで計算した。

図7及び8に、 $(C_A, C_B) = (0.3, 0.5)$ 及び $(0.6, 0.5)$ の時の、ロボットが検出した発話と会話状

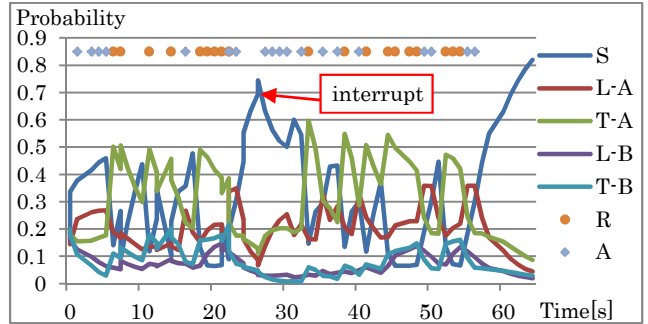


図 7. $C_A=0.3, C_B=0.5$ の時の各状態確率

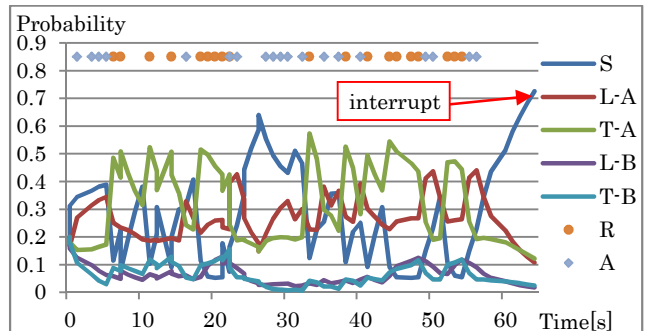


図 8. $C_A=0.6, C_B=0.5$ の時の各状態確率

態モデルの各状態の確率を示す。図7より、時刻25秒付近の発話非検知区間で「S」状態の確率が上昇し、0.7を超えていることがわかる。また図8より、 C_A が大きくなると、Aの話を書いている確率「L-A」が高くなり、「S」状態の確率の上昇が鈍くなる。この場合、割り込みが可能になったのは時刻65秒となり、会話相手の重要度が反映された結果になった。

5. むすび

本稿では、視覚情報による個人識別と音情報を組み合わせることで、受取人が誰と会話しているかを考慮した割り込みを行うための会話状態モデルを提案した。今後の課題として、事前に与えられた地図情報を用いることで、観測対象がすべてTOFカメラの視野内に収まらないケースにも対応することが挙げられる。

参考文献

- [1] 田中貴紘, 藤田欣也: “ユーザの割り込み拒否度を考慮した円滑な会話開始支援エージェント,” 信学会論, Vol. J92-A, No. 11, pp. 852-863, 2009.
- [2] M. Yamamoto and T. Watanabe: “Timing Control Effects of Utterance to Communicative Actions on Embodied Interaction with a Robot,” RO-MAN, pp.20-22, Japan, September 2004
- [3] H. Bay, T. Tuytelaars and L. V. Gool: “SURF: Speeded Up Robust Features,” European Conference on Computer Vision, pp. 404-417, 2006.
- [4] 高橋哲史, 今井順一, 金子正秀: “全周方向での音源位置推定及び肌色情報との統合による話者位置推定への適用,” 信学会論, Vol. J90-D, No.8, pp. 2232-2241, 2007-8.