

# 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

崔 元杰<sup>†</sup>

今井 順一<sup>†</sup>

金子 正秀<sup>†</sup>

<sup>†</sup>電気通信大学

## 1. はじめに

人間とロボットとのインタラクションに関する研究が増えてきている。それらの研究の多くでは、一台のロボットと一人のユーザとのコミュニケーションを対象としている。これに対し、複数のユーザとのコミュニケーションを扱った研究はまだ少ない。

本論文では、ユーザのロボットへの関心の度合を評価する基準になる優先度を導入することにより、ロボットと複数のユーザとの円滑なコミュニケーションを可能にすることを試みる。具体的には、複数のユーザが同時に話しかけた時、優先度が高い（ロボットに、より興味を持っており、最も親しい）ユーザを自動的に選択するシステムを提案し、その有効性を示す。

## 2. ユーザ毎の優先度

本論文では、ユーザの顔の向き、ジェスチャ、音声などのマルチモーダル情報を統合し、また、ユーザとの距離を考慮することにより、ロボットとの親しさを表す優先度を次式で算出する。

$$U = \sum_i w_i p_i \quad (1)$$

この式で、 $U$  は優先度を、 $p_i$  は各モーダルの出現頻度及び距離から判定される親密さを、 $w_i$  は  $p_i$  に対する重みを表す。

ユーザごとの優先度を計算することにより、複数ユーザの中でロボットにもっとも関心を示したユーザを優先してコミュニケーションを実現することが可能となる。

## 3. ユーザに関する情報の抽出

### 3.1 顔抽出

肌色領域を抽出し、マスク画像を作成することによって、処理範囲を限定する。図1の画像(b)は肌色抽出結果を利用したマスク画像である。このような処理を行うことにより、顔抽出の精度が高くなり、処理速度を向上させることができる。

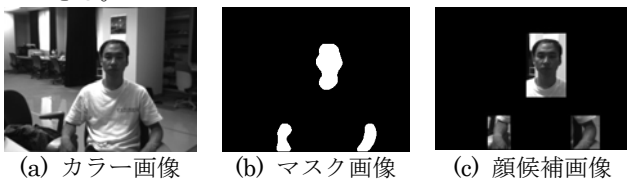


図1. マスク画像の作成

肌色抽出に関しては、文献[1]で提案されているガウス混合モデルを用いて肌色分布のモデルを作り、このモデルから入力画像における肌色の尤度を求める。その後、閾値処理により2値画像を作り、雑音除去処理を行ってマスク画像を作成する。

“Communications among Robot and Users Based on Judgment of Priority of Each User by Multimodal Information”  
<sup>†</sup> Yuanjie Cui, Jun-ichi Imai, and Masahide Kaneko  
 The University of Electro-Communications

得られた顔候補画像（図1の画像(c)）をもとにオムロン社の OKAO Vision を用いて顔の抽出を行う。図2は顔抽出結果の例である。

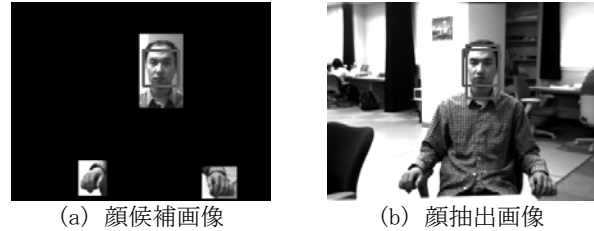


図2. 顔抽出結果

### 3.2 手の抽出

一枚の画像中に複数のユーザがいる場合、どの手がどのユーザの手であるかを確認する必要がある。本論文では、各ユーザの上半身を抽出することを通して、各ユーザの手の判定を行う。

#### 3.2.1 上半身の抽出

3.1 で得られた顔領域から、人の身体構造を利用して、図3(a)に示すように顔領域の下の部分を切出す。この部分のHSV値を求めることにより、ユーザの上半身のカラー情報が得られる。このカラー情報を用いて画像全体で抽出を行うことで、ユーザの上半身を抽出する。図3(c)にカラー画像(b)からの上半身領域抽出結果を示す。



図3. 上半身領域の抽出

#### 3.2.2 手の抽出

ユーザの手は腕とつながっているため、手の周りにユーザの上半身領域が存在するはずである。3.1 で得られた肌色領域中、顔領域以外の肌色領域を手の候補領域とする。手の候補領域を1.4倍に拡大し、この拡大領域内にユーザの上半身領域が入っているか否かで、そのユーザの手を抽出する。図4は手の抽出結果の例である。

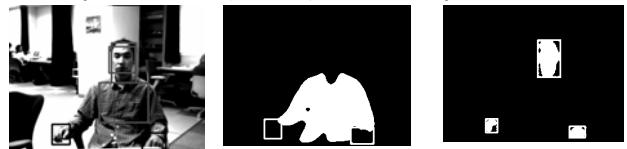


図4. 手の抽出結果

### 3.3 ジェスチャ認識

本論文では6方向（上・下・左・右・前・後）指示のジェスチャ認識を行う。6方向指示のジェスチャを手の移動ベクトルで調べてみた結果、指示する方向の手の動き（移動ベクトル）が戻りの手の動きより速いことが分かった。そこで、本論文では手の移動ベクトルを用いてジェスチャ

認識を行うことにする。まず、3.2で抽出した手の3次元座標を求める。前フレームの手の座標と現フレームの手の座標の差を移動ベクトルとして求める。ジェスチャをしている間の移動ベクトルを調べることで認識を行う。

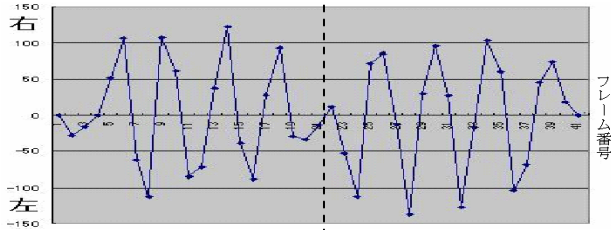


図 5. 移動ベクトル

図5は点線を境にして、二つのジェスチャ(右と左)の移動ベクトルを表わしたグラフである。図から分かるように、「右」を指示するジェスチャ(一番目のジェスチャ)の時、右方向移動ベクトルの平均値(68)が左方向移動ベクトルの平均値(53)より高い。

### 3.4 音源定位

本論文では、5つのマイクロホンから構成されるアレーを利用する。CSP法(Cross-power Spectrum Phase) [2]を用いて、2つずつのペアに対する白色化相互相関係数を計算し、その結果を同期加算することにより音源定位を行う。推定した音源の位置がユーザの顔の領域に入っている場合には、このユーザが話していると判断する。

### 3.5 距離の計算

撮影用カメラとしてはステレオカメラ(Digiclops)を使用する。画像の各画素までの距離を計算できるので、顔領域の画素の平均距離をロボットとユーザとの距離と定義する。近接学における親密な距離(~50cm程度)、個人的な距離(50~120cm程度)、社会的な距離(120~360cm程度)、公共的な距離(360cm以上)の関係を用いて、ユーザとの親密さ $p_d$ を計算する。ユーザとロボットとの距離が近いほど $p_d$ の値を大きくする。

## 4. マルチモーダル情報の統合

ある枚数のフレーム(例えば $nFrame = 50$ )を1つの観察区間とする。その中で顔を抽出したフレーム数を $nFrontFace$ 、ジェスチャをしているフレーム数を $nGesture$ 、話しているフレーム数を $nVoice$ とする。各モーダルの出現頻度を次の式で定義する。

$$p_f = nFrontFace / nFrame \quad (2)$$

$$p_g = nGesture / nFrame \quad (3)$$

$$p_v = nVoice / nFrame \quad (4)$$

観察区間ごとに $p_f, p_g, p_v$ 及び $p_d$ を用いて式(1)により優先度を計算する。コミュニケーション時における顔の向きの重要性を考慮して、 $p_f$ の重み $w_f$ を0.5、

$p_g, p_v, p_d$ 各々の重み $w_g, w_v, w_d$ を0.2、0.2、0.1にする。

## 5. 実験と結果

### 5.1 実験

複数のユーザがいる時、各ユーザの優先度(ロボットに対する関心の高さ)を上述の方法で計算することにより、ロボットが自分に最も関心を持っているユーザを選ぶシ

ステムを作成した。

### 5.2 結果

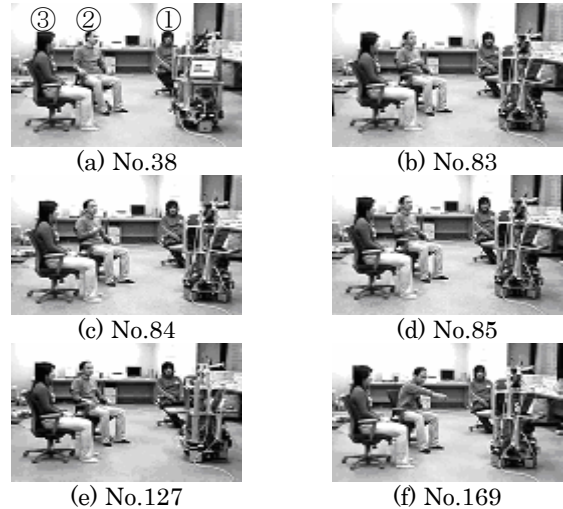


図 6. 優先度計算を行ったシーン画像

図6は優先度計算を行ったシーン画像である。図(a)で、ユーザ①(一番右)はただロボットを見ているだけで優先度は低い。画像(b)~(d)で、ユーザ②(真中)はロボットに話しながらジェスチャしているので優先度が高い。図(e)で、ユーザ③(一番左)はロボットを見ながら話しているだけなので優先度があまり高くない。3人のユーザの優先度を比べてみるとユーザ②の優先度が一番高いので、画像(f)では、ロボットがユーザ②のところに向きを変えてコミュニケーションをすることになる。

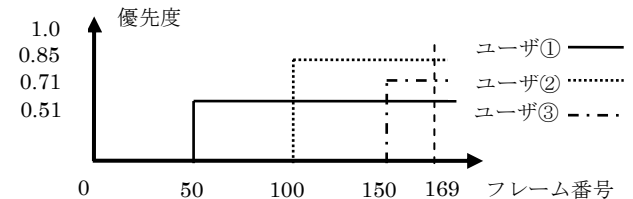


図 7. ユーザ毎の優先度

図7は、上のシーンにおける各ユーザの優先度のグラフである。フレームNo.50でユーザ①の優先度が計算され、No.100でユーザ②の優先度が計算され、No.150でユーザ③の優先度が計算される。No.169の時、3人の優先度が比べられた結果、最も優先度の高いユーザ②とコミュニケーションする。

### 6. むすび

本論文では、優先度を定義することによりロボットと複数のユーザとの円滑なコミュニケーションを可能にした。今後は、ロボットがもっと人間らしくコミュニケーションできるように、マルチモーダル情報の最適な活用方法を検討していきたい。

### [参考文献]

- [1] Quan Huynh-Thu, Mitsuhiro Meguro and Masahide Kaneko: "Skin-color extraction in images with complex background and varying illumination," IEEE Workshop on Applications of Computer Vision(WACV2002), pp.280-285, Florida, USA, 2002.12.
- [2] 西浦敬信, 山田武志, 中村哲, 鹿野清宏: "マイクロホンアレーを用いたCSP法に基づく複数音源位置推定" 信学論, vol.J83-D2, no.8, pp.1713-1721, 2000.8.