

# 実践知能アプリケーション開発プラットフォーム PRINTEPS のための RGB-D カメラによる来場者検出と属性判定

田中 康浩† 中山 祐介† 齋藤 俊太† 齋藤 英雄†  
慶應義塾大学†

## 要旨

接客業を代行するロボットの実現のためには、来場者の人数や属性を検出し、ロボットに情報を送るための手法が必要である。本稿では RGB-D カメラによる三次元情報を用い、来場者の人数を把握し、人物の属性を判別するモジュールに入力するための画像を用意する手法を検討した。これにより、ロボットに来場者の情報を送ることができた。

## 1. はじめに

近年、AI アプリケーションが次々と登場し、社会の関心が集まっている。しかし、その次のステップとして、人と機械が様々なレベルで連携し、互いに知能を進化させていくマルチモーダルな枠組みが必要とされてきている。そこで我々は、人と機械が共同するための相互作用を持つことを特徴とする知能アプリケーションの開発プラットフォーム PRINTEPS[1]の構築を行っており、現在はロボットが接客業を代行する際に必要なモジュール群の開発・検討に注力している。その中でも、接客を担当するコミュニケーションロボットが来場者を検知し、実際にコミュニケーションを始めるといった状況を想定した際、ロボットは、来場者が何人いるか、その人たちがどんな属性を持っているかを判別した上で、適切に振る舞う必要がある。これらを実現するために、本稿では、場内の出入り口付近に RGB-D カメラを設置し、その RGB-D カメラを用いて取得した画像列から、人物の検出、及びその人物の性別、年齢の把握をし、ロボットにフィードバックを与える手法の提案・検討を行う。

提案手法では、接客業において人物を検出した際、その人が来場者なのか、退場者なのかの区別をつけることは重要であるため、カメラか

らの三次元情報を活用して、それらの区別をつける手法についても検討する。また、人物の性別、年齢を検出する際には、検出した人物の画像を入力として深層畳み込みニューラルネットワークを用いた手法[2]などを提案手法内に適用することが考えられる。しかし、このような手法を本提案手法で導入する場合、入力画像において人物が正面を向いているかどうかによって出力結果の精度が異なってしまう。そこで本稿では、来場者の顔画像をある時間内で撮影し、最も入力画像として適しているときのフレームを送る前処理の手法も検討する。

## 2. 提案手法

本提案手法では、RGB-D カメラとして Microsoft 社の Kinect v2 を利用し、実装には Kinect v2 の Software development kit (SDK)の機能を用いた。撮影した RGB-D 画像列と SDK の機能を元に、そのフレームでの人数を把握する。また、検出された人については関節の情報を得ることもでき、頭部の位置検出も行える。頭部の位置座標を把握して、カラー画像での座標に変換することで、カラー画像から人の顔画像を切り取ることができる。この画像を深層畳み込みニューラルネットワークを用いた性別・年齢推定手法[2]の入力として用いる。これにより性別・年齢の情報が得られるため、来場者の ID、年齢、性別、カメラからの距離をまとめてロボットに送ることができる。さらに、フレーム毎の距離情報を保存しておくことで、カメラからの距離が離れていく人と近づいていく人とを区別し、近づいていく人を来場者として検出する。

性別・年齢推定モジュールに出力するための顔写真の選出について説明する。SDK を利用することで検出された人物の頭部の情報として、図 1 における Pitch, Yaw, Roll の各回転軸の回転量が得られる。各回転軸はすべて頭部の位置座標を通り、各軸の回転量は顔が正面に近づくほど小さくなる。最も正面を向いている画像を選出するため、(1)式に示す計算により各軸の回転量の平方和をとる。(1)式で  $\theta_p$ ,  $\theta_y$ ,  $\theta_r$  は各回

Customers Detection and Attributes Recognition Using RGB-D Camera for a Platform PRINTEPS to Develop Practical Intelligent Applications

† Yasuhiro Tanaka, Yusuke Nakayama, Shunta Saito, Hideo Saito, Keio University.

回転の回転量を， $S$ はそれらの平方和を表す．

$$S = \theta_p^2 + \theta_y^2 + \theta_r^2 \quad (1)$$

そして 10 秒以内で人物の顔情報をフレーム毎に取得し続け，平方和の値がより小さくなったときの画像を切り取って保存する．最終的に最も  $S$  の値が小さいときの画像を年齢・性別判定手法[2]の入力画像として用いることで，出力結果をその人の属性判定結果とする．

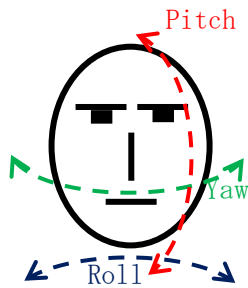


図 1: 顔の向きの情報

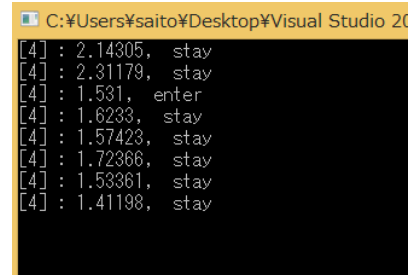


図 3: 来場者に対しての検出結果

### 3. 実験

提案した手法を用いて，入り口付近に設置した Kinect v2 により得た画像列から，来場者の検出及び来場者の性別・年齢推定を行い，コミュニケーションロボットにその情報を送信する実験を行った．その結果，図 2 のように頭部の位置を検出し，距離情報と顔情報を送ることができた．これによりロボットが来場者の人数を把握し，「いらっしゃいませ。2名様ですか？」と声をかけるといった適当な対応がとれた．



図 2: 来場者の検出とそのときの対応

また，距離情報を保存しておくことで，来場者と退場者の区別をつけ，図 3 のように「enter」「stay」といった情報をロボットに送ることができた．図 3 において，数値はカメラと人の頭部座標との距離である．

顔画像の選出についても，回転量の算出を行うことで，図 4 と図 5 のように顔の向きが異なる画像の区別を行えた．例えば，図 4 の顔の向きは「 $\theta_p: 20, \theta_y: -42, \theta_r: -17$ 」，図 5 の顔の向きは「 $\theta_p: 3, \theta_y: -1, \theta_r: -6$ 」であり，適切に値がとれている．平方和をとると図 5 のほうが小さいため，図 4 より図 5 の画像を優先して送る．



図 4: 横を向いた顔画像 図 5: 正面を向いた顔画像

### 4. まとめ

ロボットが接客業を代行するにあたって，来場者に対して適切に振る舞うために，来場者の人数や属性を得る必要があった．そのために，Kinect v2 により検出された人の情報や，頭部の位置座標を用いて，来場者の人数や属性の情報を送るための手法を検討した．距離情報を用いることで，退場者と来場者の区別をつけ，来場者のみの情報をロボットに送れるようにした．また，検出された顔情報から最も正面を向いている画像を選出して，性別・年齢判定のモジュールに最適な画像を送るための手法を検討した．今後の方針としては，正面を向いた場合とそうでない場合を比較して，性別・年齢推定の精度にどれほど影響を与えているかの比較をする．

謝辞 本研究は，科学技術振興機構 (JST) 戦略的想像研究推進事業 (CREST) 「実践知能アプリケーション構築フレームワーク PRINTEPS の開発と社会実践」の支援によって実施した．

### 参考文献

- [1] 森田武史, 西村良太, 山口高平, ストリーム推論と ROS に基づく総合知能アプリケーション開発ツール PRINTEPS, 第 19 回クラウドネットワークロボット研究会(CNR), 2015 年.
- [2] Gil Levi and Tal Hassner, Age and Gender Classification using Convolutional Neural Networks, *IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, June 2015.