

HRIにおける表情認識を用いた

ユーザの感情推定手法に関する基礎的検討

A Basic Study of Emotion Estimation Method of User using Facial Expression Recognition in HRI

朱 曜南^{*1*2}
Yaonan Zhuジメネス フェリックス^{*2}
Felix Jimenez吉川 大弘^{*2}
Tomohiro Yoshikawa古橋 武^{*2}
Takeshi Furuhashi^{*1} 中国科学技術大学ソフトウェア学院
School of Software Engineering, USTC ^{*1}^{*2} 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University ^{*2}

Human Robot Interaction has been an active research area in recent years. For better and effective human robot interaction, it is important for robots to estimate human emotions in real time and make a proper reaction. This research proposes an image recognition based method to estimate human emotions in real time. The method uses data of human facial expressions and postures, focusing on the educational-support scenario. It is able to estimate students' emotions such as happy, angry and sad from students' facial expressions, those emotions are expressed when students are learning. In addition, when learning, students can be confused about difficult problems or concentrating on solving problems, the method is capable of estimating those learning states by recognizing students' postures.

1. はじめに

近年、人とロボットの相互作用 (Human-Robot Interaction) の研究が盛んに行われている。本研究では、ユーザと社会的相互作用をするパートナー型ロボットに注目する。このようなロボットは、教育支援、介護支援、コミュニケーションなど多様な場面において応用されている[Breazeal 03]。その中でも、本研究では教育支援の場面で活躍する教育支援ロボットに注目する。

教育支援ロボットでは、ロボットが学習者の気持ちに共感することでロボットに対する親近感を向上させることが可能であると報告されている[Jimenez 16]。しかしながら、従来研究では、ロボットと学習者が問題を交互に解き合う協調学習において、ロボットが学習者の正誤判定に応じて学習者の気持ちに共感するような感情を表出する共感表出手法が提案されているが、問題を解く際の学習者の困惑や驚きなどの、感情に共感するような手法は提案されていない。ロボットが問題を解いている学習者の感情に対して共感するためには、ロボットが問題を解いている学習者の感情をリアルタイムに認識して、それに応じた感情を表出することが重要だと考える。

そこで本稿では、画像認識技術を用いて表情と身体動作から問題を解いている学習者の感情をリアルタイムに推定する感情推定手法について検討する。本手法は、学習者の表情から喜怒哀楽の感情を認識し、学習者の身体動作から学習者の学習に対する集中度合や困惑程度などを認識する手法である。実験では、画像認識技術の従来研究におけるテストデータを用いて、本手法における表情認識の性能について検証した。

2. 感情推定手法

本手法は、表情認識システムと身体動作認識システムから構成される。表情認識システムでは、学習者の表情から喜怒哀楽の感情を認識する。身体動作認識システムでは、学習者の身体動作から集中度合や困惑程度を認識する。

2.1 表情認識システム

本稿では、学習者の表情から感情を推定するために、Convolutional Neural Network (CNN)を用いてリアルタイムに表情を認識する表情認識システムを構築した。CNNは、表情認識の従来研究において最も使用されている手法であり、人の喜怒哀楽を分類する性能が高いと報告されている[Gaspar 16]。

本稿で構築した表情認識システムの構造を図1に示す。

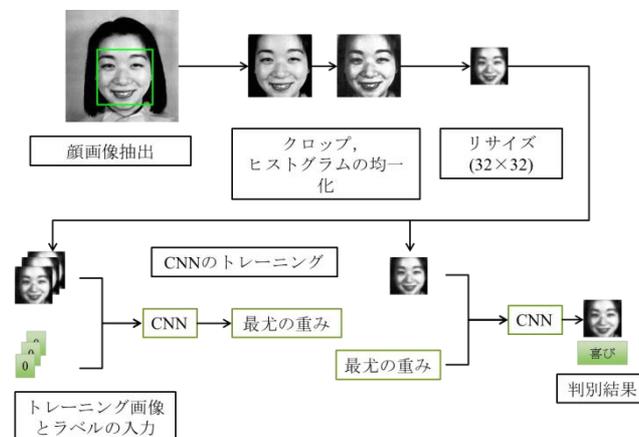


図1 リアルタイム表情認識システム

表情認識手法では、まず、カメラから学習者の顔画像を抽出する。次に抽出された顔画像をグレースケールに変換し、クロップした静止画像を生成する。ここでクロップする理由は、画像中で表情と関係のない部分を取り除き、分類器によく分類させるためである。クロップした静止画像は、撮影した際の環境に左右されてコントラストや明るさが微妙に違っているため、これにヒストグラムの均一化を施し、画像全体のコントラストと明るさのバランスを改善する。そして、改善した静止画像を32x32ピクセルにリサイズし、CNNに判別させる。32x32ピクセルの画像を使用することで、処理速度を向上させることが可能となり学習者の感情をリアルタイムに分類することができる[André 17]。本稿で使用したCNNの構造を図2に示す。

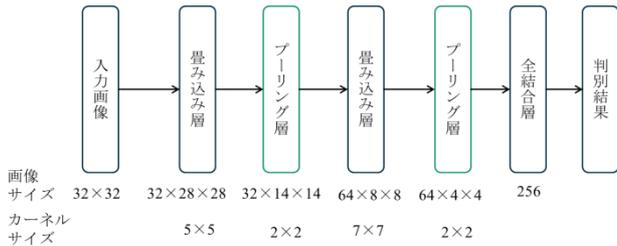


図2 本システムで使用するCNNの構造

図2に示されているように、構築したCNNは畳み込み層を2層持ち、それぞれの畳み込み層の後にプーリング層を配置している。CNNのトレーニングにはJAFFEデータセット[Lyons 98]を用いた。JAFFEは、日本人女性10人の表情データで構成されている。表情データは、喜び、怒り、悲しみ、嫌悪、驚き、恐怖、真顔で構成されており、それぞれの表情は3~4枚ずつある。表情データは、合計219枚ある。

本表情認識システムでは、一枚の画像を認識して、感情を分類するのに40msの時間しかかからないため、1秒間に25回感情を認識できる。学習中に変化していく学習者の表情からより的確に感情を推定するために、1秒間に25回感情を推定して、その結果の多数決から感情を分類する。これにより、学習中に変化していく学習者の感情を認識する性能を向上させることができると考える。例えば、25回中、15回が喜び、5回が驚き、5回が怒りと分類された場合は、その1秒間は最も回数が多かった「喜び」という感情に認識される(図3)。

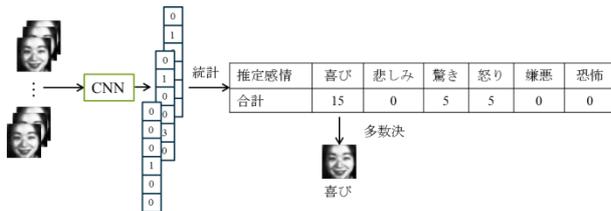


図3 多数決法

2.2 身体動作認識システム

学習中における、学習者の集中度合や困惑程度は、表情から推定しにくい。一方、手を動かしているや首を傾げているなど学習者の身体動作からなら、学習者の集中度合や困惑程度は推定しやすい。そこで学習者の学習状態を把握するために、Kinectを用いて学習者の身体動作から学習状態を推定する身体動作認識システムを構築した。身体動作認識システムは、Kinect Gesture Builderを用いて学習者の学習中のポスチャーを分類器に学習させた特定のポスチャーを識別できるようにした。ポスチャーとは、学習中の学習者の姿勢や身体動作のことを指す。身体動作認識システムの構造を図4に示す。

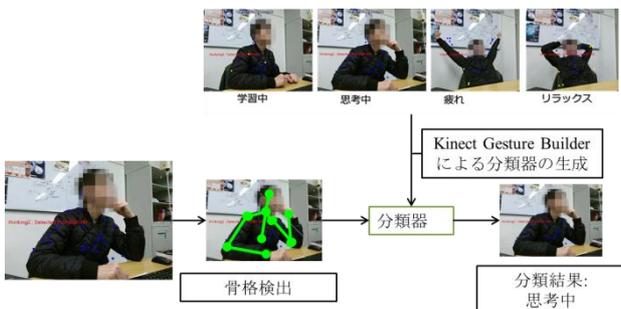


図4 身体動作認識システムの構造

身体動作認識システムでは、まず、Kinectを用いて学習者の身体動画を撮る。そして学習者の身体動画から骨格を検出する。検出された骨格の情報を用いて、あらかじめ学習させておいた分類器にどのポスチャーであるかを分類させる。今回、身体動作認識システムでは、学習者が学習中に行う可能性が高い、学習中、思考中、疲れ、リラックスの4つのポスチャーを分類器に学習させた。

3. 実験

表情認識システムに対して、クロスバリデーションを用いて判別精度(accuracy)の確認を行った。データセットはJAFFEを使用し、サンプル画像を10セットに分け、1セットをテストデータに用い、残りの9セットをトレーニングデータとし、テストデータとトレーニングデータを変更して10回繰り返した。また、用いた表情は6クラス(喜び、怒り、悲しみ、嫌悪、驚き、恐怖)である。クロスバリデーションの結果判別精度は70.0%であった。10回の内、判別精度が最も良かった分類器を用いてデータセットの各クラスのサンプル画像を判別させその結果を表1の混同行列にまとめた。

表1 混同行列

	喜び	悲しみ	驚き	怒り	嫌悪	恐怖
喜び	28	0	1	0	0	2
悲しみ	1	22	2	1	5	0
驚き	0	0	30	0	0	0
怒り	0	2	0	28	0	0
嫌悪	0	0	0	4	25	0
恐怖	0	1	3	3	2	23

身体動作認識システムに対して、学習者のポスチャーを認識できるかどうかを確認するために、被験者1名の学習中の感情を推定する実験を実施した。その結果、図5のように、被験者が学習中に現したポスチャーをそれぞれ正しく分類することが可能であることが確認できた。分類の結果は画面上に赤色の文字で表示されるようにしてある。



図5 身体動作認識システムによる感情推定

4. おわりに

本稿では、画像認識技術を用いて表情と身体動作から問題を解いている学習者の感情をリアルタイムに推定する感情推定手法について基礎的な検討をした。本手法は、表情認識システムと身体動作認識システムから構成される。表情認識システムでは、学習者の表情から喜怒哀楽の感情を認識する。身体動作認識システムでは、学習者の身体動作から集中度合や困惑程度を認識する。基礎的な実験を通して、表情認識システムは、JAFFEによるテストデータにおいて、感情の判別精度は70.0%の性能をもつことを示した。身体動作認識システムでは、Kinect Gesture Builderを用いて特定のポスチャーを学習させ分類器を生成した。そして被験者1名を対象とした基礎的な実験において、学習中の感情を推定することが可能であることを確認した。

今後は、両システムを一つにまとめ、教育支援ロボットに実装する。そして、被験者実験を通して本感情推定手法が学習者の感情を的確に認識できるかどうか検証していく。

参考文献

- [André 17] André Teixeira Lopes, Edilson de Aguiar, Alberto F.DeSouza, Thiago Oliveira-Santos: Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order, Pattern Recognition, Elsevier, 2017.
- [Breazeal 03] Cynthia Breazeal: Emotion and Sociable Humanoid Robots, International Journal of Human-Computer Studies, Elsevier, 2003.
- [Gaspar 16] Luis-Alberto Perez-Gaspar, Santiago-Omar Caballero-Morales, Felipe Trujillo-Romero: Multimodal emotion recognition with evolutionary computation for human-robot interaction, Expert Systems With Applications, Elsevier, 2016.
- [Jimenez 16] Felix Jimenez, Tomohiro Yoshikawa, Takeshi Furuhashi, Masayoshi Kanoh: Effects of Collaborative Learning with Robots using Model of Emotional Expressions, journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 2016.
- [Lyons 98] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba: Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic and Gesture Recognition, 1998