

カウンセリングに必要な顔表情データの抽出

Extraction of Facial Expression Data for Counseling

小湊 勇弥 清瀬 藍子 西村 有紗 武藤 良 植木 一也*
Yuya Kominato Aiko Kiyose Nishimura Arisa Ryo Mutou Kazuya Ueki

明星大学
Meisei University

Abstract: We are currently working on an artificial intelligence system that aims to provide psychologically distressed people with an opportunity to receive counseling. In this paper, we describe one of the functions of the system, the face recognition technology. As a face recognition technology, we implemented two functions: one for facial expression recognition and the other for blink measurement. We conducted a unit test of the face recognition technology, and confirmed that both facial expression recognition and blink measurement were working properly.

1 はじめに

近年の自殺者数は、警視庁の web サイトによると、減少傾向にある。しかし、自殺動機は厚生労働省のデータ [1] によると、年々自殺動機が多様化しつつある。最近のニュースだと、有名男性俳優の自殺により、多くファンが彼を追いかけるように自殺をしたというものがあつた。これは、ファンの生きる活力だったものが無くなり、喪失感に襲われ、自殺という選択肢を選んできました結果と考えられる。そこで私たちは、精神的に苦痛を感じた人たちにカウンセリングを受けてもらうためのきっかけを作る人工知能システムを開発することを目指している。本論では、提案する人工知能システムの機能の 1 つである、顔画像認識について述べていく。

2 全体のシステム

本論文では、顔画像認識について述べていくが、まず初めにシステム全体 [2] について説明する。システムの理想として、画面に表示させたキャラクターと被験者が対話を通じて緊張感を削ぎつつ、悩みに関する内容に触れ、何に悩み苦しんで知るかなどを記録する。記録したデータをカウンセラーに提出し、カウンセラーの仕事の手助けになることを目標としている。このシステムを大まかに分けると、画像生成、表情認識、自然言語処理の 3 つが存在する。

画像生成は、カウンセリング中に被験者に対して表示するキャラクターを作成する。この機能は、カウンセリング中被験者の緊張を解き、悩みの核心に迫る際、心理的ハードルを下げることを目的としている。そのため、セラピー効果の高い犬や猫といった動物を表示させようと考えている。

表情認識は、被験者の表情の変化を読み取る機能を作成する。被験者が人工知能と対話をする際、どの質問に対して表情が変化したか、瞬きの回数が増えたか等の情報を収集し、対話結果のフィードバックに対する付随情報として、精度を向上させることを目的としている。詳細は次章から説明していく。

自然言語処理では、被験者と人工知能が対話するための対話機能と、文字ベースの悩み文に対するタグ付け機能がある。対話機能は、カウンセリングを通して被験者の情報を集めることが目的で、文字ベースの悩み文に対するタグ付け機能では、FastText を使って、カウンセリングで被験者が入力した文章を解析し、何に対して悩んでいるのかといったデータを蓄積することを目的としている。悩み文に対するタグ付け機能の詳



図 1: 実験に使用した RAF-DB の顔画像の例

*連絡先: 明星大学 情報学部 情報学科
〒191-8506 東京都日野市程久保 2-1-1
E-mail: kazuya.ueki@meisei-u.ac.jp

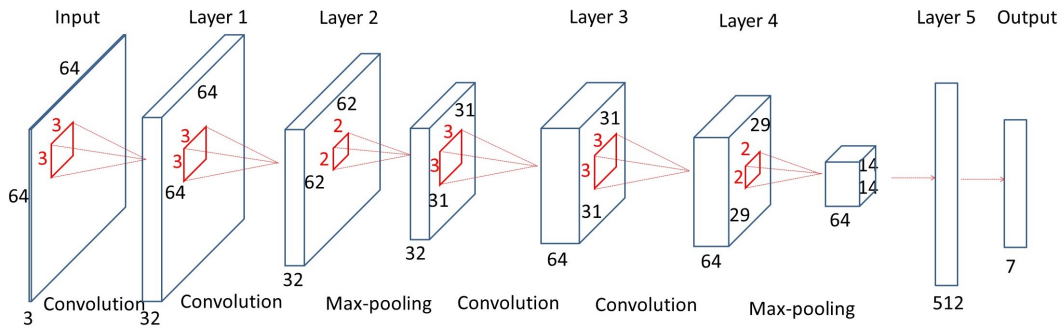


図 2: 使用した畳み込みニューラルネットワークの構造.

細は, [3] で述べている.

3 作成した顔感情認識のシステム

今回作成した顔画像認識は内容は, 大きく分けて以下に示す 2 つの機能を有している.

1. カメラからユーザの表情を認識して数値化する.
2. 瞬きの回数を計測.

3.1 感情認識

今回作成するシステムでは, 被験者の心情を詳しく読み取る必要があるため, 多くの種類の表情を認識する必要がある. そのため, Surprise (驚き), Fear (恐怖), Disgust (嫌悪), Happiness (喜び), Sadness (悲しみ), Anger (怒り), Neutral (ニュートラル) の 7 つが用意されている, Real-world Affective Faces Database (RAF-DB)[4] を使用した. RAF-DB の顔画像ファイルのうち, aligned というディレクトリに置かれている, 図 1 の上部に示すような顔位置が正規化された 100x100 ピクセルの画像を用いた. システムで顔検出を行った場合の基準と合わせるため, Dlib の顔器官検出を使用し, 再度, 目の位置を基準として 64x64 ピクセルで切り取り直した. 切り取った画像の例を図 1 の下部に示す. 以上の設定で RAF-DB 内の顔画像を整理したところ, 15,339 枚あったものが 9,829 枚になった.

顔表情の学習には, 図 2 に示すような 5 層の畳み込みニューラルネットワークを用いた. 9,829 枚の画像を学習データと検証データを 8:2 の割合に分割し, 100 エポックまで学習し, 検証データの精度が最も高かった 96 エポック時のモデルを選択した. 7 種類の表情の識別精度は平均で 79.3% となった.

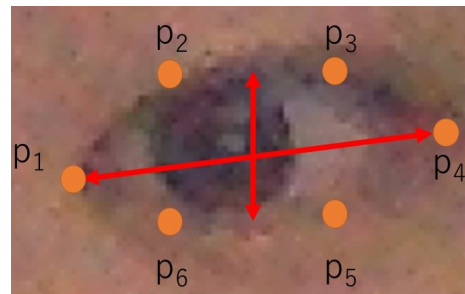


図 3: 目のランドマーク

3.2 瞬きの認識

本システムにおける瞬きの認識は, 被検者の心情を読み取る意味合いを持っている. 瞬きの回数が増えたタイミング等を記録することでカウンセリングの質問内容に対する, 心の揺れ, 動揺等を計測できると考えている.

瞬きの認識方法としては, Dlib の顔ランドマーク検出機能により検出された, 図 3 に示すような $p_1 \sim p_6$ の 6 点のランドマークを使用した. その後以下の式を使い, 目のアスペクト比 (EAR: eyes aspect ratio) を求め計測を行った [5].

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

左目, 右目のそれぞれで EAR を計算したのち, 両目の値アスペクト比の値を合計する. 目の大きさは, 人によって異なるため, 被検者ごとに目のアスペクト比は変化する. そのため, 被検者ごとに瞬きの計測が行えるよう, 計測ごとに被験者の目のアスペクト比を平均化し, 平均値が一定の基準よりも下回った場合, 瞬きとしてカウントするシステムを構築した.

```
Surprise : 0.25044
Fear      : 0.00000
Disgust   : 0.00002
Happiness : 0.00000
Sadness   : 0.01004
Anger     : 0.00000
Neutral   : 0.73950
```

図 4: 表情認識結果

4 実験

今回の実験では、システム開発に携わった筆者が被験者となり、webカメラを利用したリアルタイム方式により実験を進めた。実験中の映像を表示するウィンドウは、表情変異グラフを表示するものと、瞬きを計測する2つに分けた。

表情認識結果を表示する図4のウィンドウには、7つの表情を数値化して、パーセントで表示させた横棒グラフを作成した。7つの表情のうち、Surprise（驚き）、Happiness（喜び）、Anger（怒り）、Neutral（ニュートラル）の4つは比較的安定して認識できたが、Fear（恐怖）、Disgust（嫌悪）、Sadness（悲しみ）の3つはあまり正確に認識できていなかった。

瞬きの計測を表示するウィンドウには、動画のフレームレート「FPS」、瞬きの回数「blink count」、両目のアスペクト比「RL_eye_ear」、両目のアスペクト比の平均値「RL_eye_ear_ave」、瞬きとして認識する基準値を「EYE_THRESHOLD」として表示させた。その他に、瞬きしている瞬間を確認しやすいように瞬き中は、「Closed eyes」と表示させた。表示させた数値と実験中の様子を図5に示す。瞬きの計測で使用する目を閉じているかどうかの閾値である「EYE_THRESHOLD」は、両目のアスペクト比の平均値を0.9で掛けた値である。「EYE_THRESHOLD」が「RL_eye_ave」の数値を2フレーム以上連続で上回った場合、瞬きをしたと判定し、「blink count」の数値を1増加させた。2フレーム以上連続という条件を付けることで、瞬きのカウントミスを減少することができた。

5 むすび

今回の実験では、システムの大まかな構造は完成したが、まだまだ改善する余地があることが判明した。

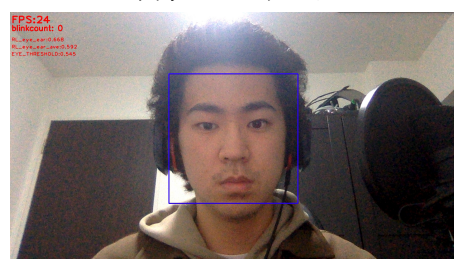
表情認識では簡単な表情は読み取れたが、Fear（恐怖）、Disgust（嫌悪）、Sadness（悲しみ）などの心の状態が深く関わる表情は現段階では読み取ることが難しかった。そのため、今後は判別が困難な表情の認識

を可能とし、深い心情を読み取るシステムの実現に近づけていこうと考えている。その他にも、システムの被験者を一人だけではなく、幅広い年代の男女を対象とし、瞬きの基準値「EYE_THRESHOLD」等のパラメータを自動で設定できるようにしていくことも検討している。

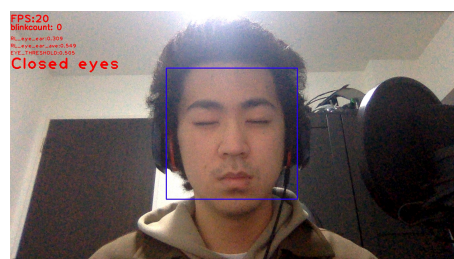
今後追加する機能として、対話中のしぐさ、声色から感情の揺れなどを認識できればと考えている。これらから感情を読み取ることにより、より正確な感情データを記録することができると考える。

```
FPS:22
blinkcount: 1
RL_eye_ear:0.658
RL_eye_ear_ave:0.519
EYE_THRESHOLD:0.477
```

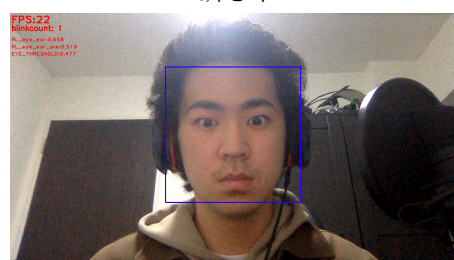
表示させたデータ



瞬き前



瞬き中



瞬き後

図 5: 瞬き前、瞬き中、瞬き後の映像の例

謝辞

本研究を行うにあたり、カウンセリングについて取材に協力してくださった明星大学心理学相談室カウンセラーの田村友人さん、心より感謝申し上げます。

参考文献

- [1] 厚生労働省: 平成 27 年版自殺対策白書 (2015)
- [2] 西村 有紗, 勝田 暁子, 小湊 勇弥, 清瀬 藍子, 鈴木 和也, 能戸 誓音, 武藤 良, 植木 一也: 人工知能でカウンセリングを身近に, HAI シンポジウム 2021 (2021)
- [3] 勝田 暁子, 西村 有紗, 植木 一也: FastText を用いた悩み文へのタグ付け, HAI シンポジウム 2021 (2021)
- [4] Shan Li, Weihong Deng, and JunPing Du: Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild, *In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2584–2593 (2017)
- [5] Tereza Soukupová and Jan Čech: Real-Time Eye Blink Detection using Facial Landmarks, *In 21st Computer Vision Winter Workshop (CVWW 2016)*, pp. 1–8 (2016)