

MediaPipe のブレンドシェイプ予測係数を用いた表情認識

1. はじめに

近年、オンライン会議の普及に伴い、感情を推定するための非言語的な情報の重要性が高まっている。これに関連した顔画像を用いた表情認識の研究として、顔筋肉の情報から表情を認識する手法[1]や、日本人の表情を分類する手法[2]などが提案されている。しかし、被験者ごとの精度差が大きく、実用が難しいという課題がある。そこで、本研究ではオンライン分類に向けて、MediaPipe でブレンドシェイプ予測係数を取得し、それに基づいて表情を分類する手法を提案する。

2. 提案手法

2.1 顔画像データの取得

本研究では先行研究[2]と同様に Ekman[1]の基本 6 感情（笑顔、悲しみ、怒り、嫌悪、恐怖、驚き）に真顔を追加した 7 つの表情の認識を行う。なお、表情データを撮影する際には、真顔は指示は行わずに撮影し、6 感情は画像例を提示してその表情をまねするように指示して撮影を行った。

2.2 MediaPipe を用いたブレンドシェイプ抽出

本研究では MediaPipe で抽出できる顔のブレンドシェイプ予測係数を用いて表情認識を行う。Google が提供するオープンソースのフェイシャルランドマーク検出ライブラリであり、画像から 468 点の顔の 3 次元座標を取得する。図 1 は特徴点抽出の例であり、取得した xy 座標を用いて 52 種類のブレンドシェイプ（眉やの動きや口の開き方など顔の特徴の動きを表した値）を予測することができる。

2.3 外れ値の除去

顔画像の撮影の際、まばたきや、顔の特徴点を正しく取得できない場合があるため、四分位数を用いて外れ値を除去する。各特徴の値をサイズ順に並べたものを 4 分割し、25%（第 1 四分位数）と 75%（第 3 四分位数）の間を四分位範囲(IQR)とする。そして、IQR の 1.5 倍以上離れた値を外れ値とみなし、範囲内の上限値・下限値に置き換える。

2.4 Random Forest を用いた表情分類

表情の判別には Random Forest (RF)を用いる。RF は教師あり学習に基づくパターン認識モデルの一種であり、並列に学習した複数の決定木で分類を行い、その結果を基に最終的な分類を行う手法である。外れ値を除いたデータに対し、各ブレンドシェイプごとに 0 から 1 の範囲に正規化を行い、それらを並べた 51 次元の特徴ベクトルを作成する。それを用いて RF により 7 つの表情の分類を行う。

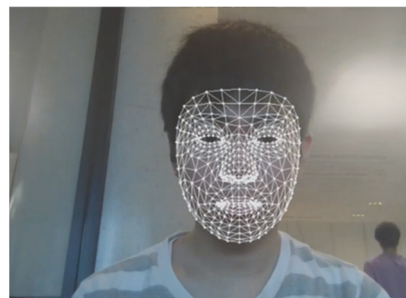


図 1 MediaPipe を用いた特徴点抽出の例

表 1 学習枚数を変更した場合の正答率の違い

学習枚数	被験者 A	被験者 B	被験者 C
各 200 枚	84.9%	90.0%	100%
各 500 枚	99.0%	95.5%	100%

3. 評価実験

表情分類の評価実験を行った。分類する表情は真顔、笑顔、悲しみ、怒り、嫌悪、恐怖、驚きの 7 クラスである。被験者は 3 名とし、学習データとして各クラス 200 枚合計 1400 枚と各クラス 500 枚合計 3500 枚の 2 通りを用意し、各クラス 200 枚のテストデータを用いて正答率を比較した。なお、学習にはそれぞれ本人のみのデータを用いた。分類結果を表 1 に示す。実験結果から各クラス 200 枚では正答率が平均 90% 程度となり、高い正答率で分類を行えた。これに対し、各クラス 500 枚では正答率が平均 98% 程度となり、より高い正答率で分類が行えた。結果が改善された理由として、一つの表情の撮影でも全く同じ顔にはならず、値にばらつきがあったため、200 枚では学習データが不足していたと考えられる。

4. おわりに

本研究では MediaPipe のブレンドシェイプ予測係数を用いて表情を認識する手法を提案した。MediaPipe で取得した顔特徴から外れ値を除去し、51 次元の特徴ベクトルを用いて Random Forest により 7 表情の分類を行った。評価実験では学習データを増やした際に正答率が向上し、被験者ごとの正答率の差も改善された。今後の課題として複数人モデルでの正答率向上や、オンライン分類の実現などが挙げられる。

【参考文献】

- [1] P. Ekman: Facial action coding system (FACS), A human face, 2002
- [2] 佐藤 恵亮：MediaPipe Face Mesh と SVM を用いた表情認識，福岡工業大学情報工学部情報工学科卒業論文，2024