

ステレオカメラを用いた3次元特徴距離からの顔表情認識

Facial Expression Recognition from Three-Dimensional Feature Distance Using Stereo Camera

○吉田寛和^{*}, 工藤憲昌^{*}, 釜谷博行^{*}

○Hirokazu Yoshida^{*}, Norimasa Kudoh^{*}, Hiroyuki Kamaya^{*}

^{*}八戸工業高等専門学校

^{*}National Institute of Technology,
Hachinohe College

キーワード: ステレオカメラ (Stereo Camera), OpenCV (Open Source Computer Vision),
ニューラルネットワーク (Neural Network), 表情認識 (Facial Expression Recognition)

連絡先: 〒039-1192 八戸市田面木字上野平 16-1 八戸工業高等専門学校 産業システム工学科
釜谷博行, Tel.: 0178-27-7283, E-mail: kamaya-e@hachinohe-ct.ac.jp

1. はじめに

近年、高齢化の進展につれて、日常生活における基本動作が困難で、介護を要する介護認定者の数が年々増加している。中でも要介護度が軽度とされている要支援者の増加が問題となっている。高齢による衰弱、関節疾患、骨折などが原因の要支援者は体を動かす機会が減少し、引きこもりがちになることからうつ病や認知症になりやすいといわれている^[1]。しかし、介護人材の需要に対し、供給が追いついていないという問題がある。

その対策として注目されているのがコミュニケーションロボットである。ロボットと会話をすることでうつ病や認知症の予防につながるのではないかと期待されている。高齢者を対象としたアンケート調査によると、コミュニケーションロボットは各年代に需要があり、その需要は高齢になるとともに増加している^[2]。

このようなコミュニケーションロボットにおいて、人間の感情を正しく理解して動作することが重要となる。人間の顔は感情の重要な情

報源の一つで、その認識にはカメラが用いられる。従来の表情認識システムの多くは、一般に単眼カメラを利用している。しかし、単眼カメラでは3次元情報を2次元情報として扱うので、顔の向きが変化した場合、認識性能に劣化が生じる^[3]。そこで、本研究では表情をより正確に認識できるよう、ステレオカメラを用いて3次元での顔表情の認識を行う。

2. 開発環境

2-1. システム構成

本研究では安価なUSBカメラを2台平行に並べることでステレオカメラを構成した。使用したカメラはBUFFALO製のBSW32KM03 (Table 1)である。カメラ間距離を50 [mm]として、3Dプリンタで固定用治具を製作し、市販の三脚に設置した (Fig. 1)。

ソフトウェアの開発にはMicrosoft Visual Studio 2013を使用する。カメラキャリブレーションやステレオマッチングはコンピュータビジョンライブラリのOpenCV2.4.11を使用す

る^[4]。生成した3Dモデルの解析には3D点群処理ライブラリのPoint Cloud Library^[5]を使用する。2次元画像からの顔の特徴点の抽出にはDlib 19.0^[6]を使用する。

Table 1 カメラの仕様

BUFFALO 製 BSW32KM03	
映像素子	CMOS センサ
有効画素数	320 万画素
最大解像度	2016×1512 ピクセル
最大フレームレート	30fps



Fig.1 製作したステレオカメラ

2-2. ステレオカメラの原理

人間は物を見るとき2つの目で得られた視覚情報のずれから奥行きを判断する。これと同様の原理で、ステレオカメラは2台のカメラで得られたカメラ画像の視差から三角測量の原理で奥行きを計算する(Fig.2)。視差とは、左右のカメラで見える同じ特徴の画像平面でのx座標の差である。左の画像に投影された点Pのx座標を x_l 、右の画像におけるx座標を x_r とすると、視差は $x_l - x_r$ と表される^[4]。3次元空間上の点P(P_x , P_y , P_z)は、この視差 $x_l - x_r$ を使い、式(1)で計算できる。

$$\begin{cases} P_x = \frac{x_l d}{x_l - x_r} \\ P_y = \frac{y d}{x_l - x_r} \\ P_z = \frac{f d}{x_l - x_r} \end{cases} \quad (1)$$

ただし、 y は画像平面でのy座標、 f は焦点距離[mm]、 d はカメラ間距離[mm]である。

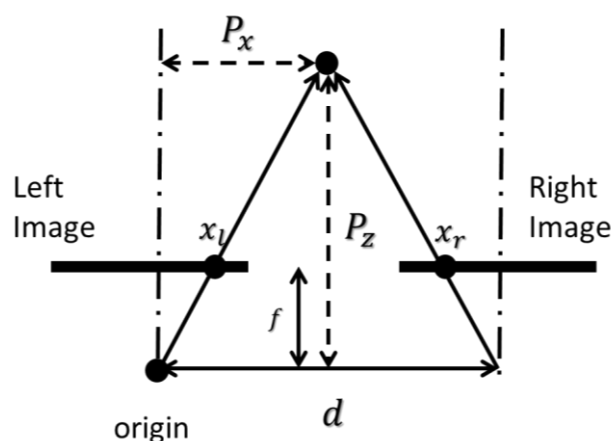


Fig.2 ステレオカメラの原理

3. 3次元特徴距離計測

3-1. 前処理

ステレオカメラにおいて左右の画像の対応点を探るステレオマッチングを行う際、y座標にずれやカメラの個体差によるレンズの歪みがあると、マッチング時間の増加、マッチングの失敗につながる。そこで、事前にカメラキャリブレーションによってカメラパラメータを導出し、得られたパラメータを元に平行化と歪み補正を行う。キャリブレーションにはチェスボードのコーナー位置から内部パラメータ・外部パラメータを推定するZhangの手法^[4]を用いた。

3-2. ステレオマッチング

ステレオマッチングの手法は大きく分けて「Localな手法」と「Globalな手法」の2つがある。「Localな手法」は画像の類似度を用いて視差を求めるが、類似したテクスチャが並んでいとうまくマッチングできないといった欠点

がある。「Globalな手法」は画像の類似度に加えて視差に対する空間的な制約を考慮するため、類似したテクスチャでもマッチングが可能である。一方で、処理時間が長くなるのでリアルタイム処理が困難という欠点がある。本研究では、精度と計算量のコストパフォーマンスを考慮した上で、高い分解能で3次元復元を行うために、サブピクセルの導出も可能なセミグローバルステレオマッチング法を使用する。

3-3. 顔特徴点の取得

3次元顔データにおける特徴点の取得手法^[7]はいくつか提案されているが、本研究では簡易的に行うために2次元画像における特徴点を3次元空間に投影することで特徴点の取得を行う。

2次元の顔特徴点の取得手法としてOpenCVに用意されているHaar-Like Cascade分類器が挙げられるが、この分類器は顔の向きがあらかじめ決められており、顔の向きが変化した場合には分類器を変更する必要がある。また、眉毛の分類器が用意されていないため、新たに眉毛の分類器を生成しなければならない。そこで、本研究ではDlibに用意されている顔特徴点データ (shape_predictor_68_face_landmarks.dat) を使用する。参照画像である左カメラの画像におけるデプス画像の奥行きをもとに特徴点を3次元空間に投影する。

3-4. 特徴距離

3-3. で得られた特徴点をもとにユークリッド距離を計算し、特徴距離を求める。特徴距離には以下の6つの距離を使用する (Fig. 3)^[8]。

- A) 目の開き： 右目の左端と右端
- B) 瞼の高さ： 眉毛の最上点と左目の最上点
- C) 口の開き： 唇の左端と右端
- D) 口の高さ： 唇の最上点と最下点
- E) 唇の伸び： 唇の右端と右目の右端
- F) 目頭の間： 右目の左端と左目の右端

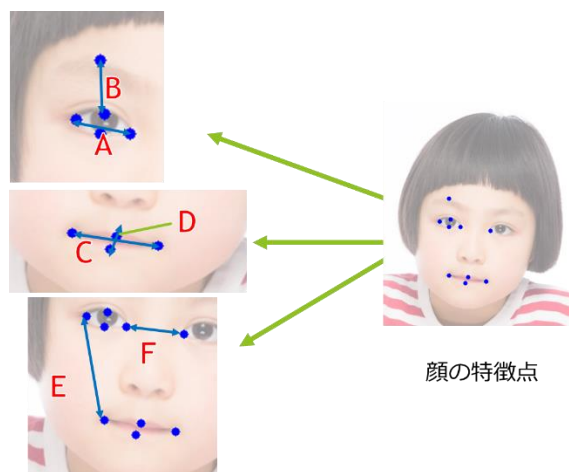


Fig. 3 使用した特徴距離

カメラまでの距離が変化すると特徴距離も変わってしまう。そこで、表情によって変化しにくいとされている目頭の間を距離を用いて、正規化を行う。

3-5. 表情認識

表情認識には、3層の階層型ニューラルネットワークを使用する。6つの特徴距離を入力とし、7種の基本表情「無表情」、「喜び」、「怒り」、「驚き」、「悲しみ」、「恐怖」、「嫌悪」に分類する。隠れ層は20である。

ニューラルネットワークの学習アルゴリズムにはバックプロパゲーション法を使用する。学習率は0.1、慣性項は0.01とし、学習回数を100,000回とした。パラメータは予備実験により決定した。

テスト時には、出力層からの値が最も大きい表情を認識表情として出力する。なお、出力層の最大値が0.5以下の場合には、認識不可として扱う。

4. 実験

あらかじめ取得した画像を用いて認識精度を求める実験を行う。7つの表情の画像データを60セットずつ取得する。データをA、Bの2つのグループに分け、それぞれ訓練データ、テスト

データとして、学習と評価を行う。まず初めにグループAを訓練データとして学習し、グループBをテストデータとして認識率を求める。次に、データを入れ替えてグループBを訓練データとして学習し、グループAをテストデータとして認識率を求める。最後に、それぞれの認識率の平均をとる。実験結果をTable 2に示す。左がニューラルネットワークへの入力、それぞれの数値が認識率を表わす。平均認識率は91.1%となった。「喜び」の認識率が100%となり最も高かった。一方、「無表情」と「悲しみ」、「怒り」と「嫌悪」では、誤認識が生じやすかった。この原因として、誤認識が多かった表情では特徴距離の分布が重なっていることが原因と考えられる。

Table 2 認識率

		出力							
		[%]	無表情	喜び	怒り	驚き	悲しみ	恐怖	嫌悪
入力	無表情	91.1	0.0	0.0	0.0	8.9	0.0	0.0	
	喜び	0.0	100.0	0.0	0.0	0.0	0.0	0.0	
	怒り	5.6	0.0	79.6	0.0	0.0	1.9	13.0	
	驚き	1.7	0.0	0.0	96.7	0.0	1.7	0.0	
	悲しみ	15.2	0.0	0.0	0.0	84.8	0.0	0.0	
	恐怖	0.0	0.0	3.3	0.0	0.0	96.7	0.0	
	嫌悪	4.0	0.0	6.9	0.0	0.0	0.0	89.1	

平均認識率：91.1%

5. まとめ

ステレオカメラを用いて、表情認識を行った。2次元平面で求めた特徴点を3次元空間に投影し、特徴点間のユークリッド距離を特徴距離とした。認識する表情は「無表情」、「喜び」、「怒り」、「驚き」、「悲しみ」、「恐怖」、「嫌悪」の7つとし、認識にはニューラルネットワークを使用した。平均認識率は91.1%であった。「喜び」の認識率が最も高かった。しかし、特徴距離が類似した表情では誤認識が生じやすかった。

今後の課題として、まず、誤認識が多かった

表情において認識率の向上を図ることが挙げられる。また、今回は一人の人物を対象に評価実験を行ったが、他の人を対象として実験を行い、提案システムの評価・改善を行う。さらに、複数の人間と会話するコミュニケーションロボットにおいて、個人毎に学習を行ってはいない不便である。そこで、不特定人物を対象にした表情認識システムに拡張したいと考えている。

参考文献

- [1] 厚生労働省, “政策レポート (介護予防)”, <http://www.mhlw.go.jp/seisaku/2009/07/02.html> (2016年9月閲覧)
- [2] 読売新聞, “暮らし・家族「人型ロボ、高齢者ほど関心」”, 2016年8月20日付朝刊
- [3] 中島寛、中島聖、青木孝文: 2次元画像と3次元立体形状を用いたハイブリッド顔認識技術, azbil Technical Review (2012年1月)
- [4] Gray Bradski and Adrian Kaebler: 詳解 OpenCV - コンピュータビジョンライブラリを使った画像処理・認識, オーム社 pp.413-462 (2009)
- [5] Point Cloud Library, <http://pointclouds.org/> (2016年11月閲覧)
- [6] Dlib, <http://dlib.net/> (2016年11月閲覧)
- [7] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, Lijun Yin, “Static and dynamic 3D facial expression recognition: A comprehensive survey”, Image and Vision Computing (2012)
- [8] Hamit Soyel and Hasan Demirel, “Facial Expression Recognition Using 3D Facial Feature Distances”, International Conference Image Analysis and Recognition (2007)