

手形状3次元CGを用いた指文字認識

細江 花
Hana Hosoe

酒向 慎司
Shinji Sako

北村 正
Tadashi Kitamura

名古屋工業大学
Nagoya Institute of Technology

1 はじめに

近年、ろう者と健聴者間の対話支援を目的として、画像情報を用いた手話認識の研究が行われている。手話の単語の意味を形成するとされる手指動作は、主に手の形状と位置と動きの3要素が重要とされ画像処理によって認識を行う手法がいくつか提案されているが、手の形状の認識は対象の自由度が高く困難とされている。そこで本研究では、CGによる精密な手形状画像を作成し、手話の手形状認識に活用する手法を検討する。実験では、学習データとして静止指文字41種類からなる実画像データセットを作成し、畳み込みニューラルネットワーク(CNN)による指文字認識により、これらに手形状CG画像を加えることの効果について検証する。

2 3次元手形状モデルの作成と手形状CG画像の作成

解剖学的には人間の手は27本の骨で構成され、5方向の関節運動を基本として複雑な動きを行うが、過度な複雑さを避けるため、ここでは21本のアーマチュア(モデル内の骨に相当)で骨格を構成する。また、手形状の自由度は、各関節の運動である23DOF(Degree of freedom)と、各軸方向の移動である3DOFの合計26DOFとすることで、指文字で表出される形状には十分に対応できると考えられる。

本研究では3次元CG作成ツール(Blender)を使用して右手手首から上の3次元手形状モデルを作成する。3次元手形状モデルの作成には、手のひらを広げ正面から撮影した実画像を使用した。アーマチュアは実物の手の骨格や関節の可動域と差異がないように作成する。また、テクスチャは、メッシュの作成に使用した人物と同一である。

3 実画像データの収録と手形状CG画像の生成

実画像のデータは、手話を母語としない10人に協力してもらい、動きを伴わない41種類の指文字を、RGBカメラを使用して収録した。各被験者は、同一の指文字を手の向きや指の角度を少しずつ変化させながら最低10サンプルずつ連続で撮影し、合計5000サンプルの画像データを収録した。収録した画像は手の周辺のみを自動で切り出し、手のひらを基準に大きさの正規化を施した。このうち、指文字ごとに6~10サンプルを手作業で選出しテストデータとした。

3次元手形状モデルによる各指文字の作成は、上記の実画像データを参考に、若干の形状変化を伴う3次元手形状モデルを3または4種類作成し、滑らかに手形状が変形するアニメーションの各コマを静止画として出力したのち、その中から参照パターンとして使用する画像を手作業で選択した。

表1 CNNによる認識の実験条件

使用サンプル数	実験1:実画像5000サンプル +回転やシフトを行った実画像 1000サンプル 実験2:実験1の画像6000サンプル +手形状CG画像2000サンプル
画像サイズ	32×32 pixel
学習データ	実験1:4800サンプル 実験2:実画像4800サンプル及び 手形状CG画像2000サンプル
テストデータ数	1200サンプル
学習率	0.01
バッチサイズ	100
学習のエポック数	60

4 CNNによる指文字認識

CNNによる指文字認識では、作成した手形状CG画像が学習データとして使用した場合に効果があるか調べるために、2つの実験を行う。実験条件は表1の通りである。また、実験で使用したCNNを図1に示す。

実験1では、実画像のみを使用してモデルを学習し認識実験を行った結果、90%となった。次に、実験1の学習データに手形状CG画像を追加して実験を行った結果、認識率は91%となり、実験1と比較して1%改善した。手形状CG画像を学習データに追加することで、認識率が1%改善することが確認でき、作成した手形状CG画像は学習データとして有効であると考えられる。

5 むすび

本研究では、手話認識に必要な学習データの生成に対応できる3次元手形状モデルの作成と、CNNによる指文字認識実験を通して生成した手形状CG画像によって学習データの拡張が可能か検証した。生成された手形状CG画像を学習データの一部として使用した場合、実画像のみを学習データとした場合と比較して1%改善する結果が得られた。

今後の課題として、3次元手形状モデルを用いて連続した手形状変化のアニメーションを作成し、連続動作を伴う指文字認識について検討する。

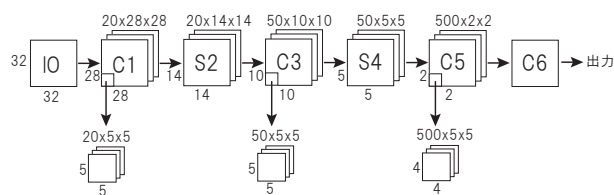


図1 CNNの構造