K-018

TOF カメラによる連続手話認識に関する検討

A Study on Continuous Sign Language Recognition Using Time of Flight Camera

森口 拓哉 † 酒向 慎司 † 北村 正 †
Takuya Moriguchi Shinji Sako Tadashi Kitamura

1. はじめに

近年，障害者が健常者と変わらない生活を営めるような環境づくりが進められている．その一つとして，聴覚障害者と健聴者の対話支援を目的としたコンピュータによる手話認識の研究がある．手話認識には手指動作の取得が必要であり，従来研究では手話動作の動画像を用いる方法やセンサを装着する方法などが用いられてきた．センサを用いた手法では，装置の特殊性や装着する煩わしさが問題となり，身体的に拘束のないカメラを用いたものが好ましいといえる．しかし，手話の動きは 3 次元的であることから，単眼では限界であるといえる．複数台のカメラで 3 次元計測を行う方法も提案されているが，装置の煩雑化もまた問題となる．

奥行き情報を容易に計測でき，かつ身体的な拘束のない方法として，TOF（Time-of-Flight）カメラを用いることが考えられる．TOF カメラによる手話認識の先行研究 [1] では，実験に使用された手話単語は数種類しかなく，小規模な認識実験であったといえる．よって本研究では，先行研究より実験に使用する語彙を増やし，TOF カメラを用いて手話のデータベースを作成する．そしてこのデータベースを使用し，これまでに提案してきた HMM（隠れマルコフモデル）に基づいた手話認識の枠組み [2] を用いて，連続手話認識を試み，その効果の検討を行った．

2. データベースの構築

TOF カメラを用いて手話のデータベースを構築する．大規模な手話単語認識実験を行うため，データベースの構築には語彙設計，収録条件等を設定する．詳細を以下の節で説明する．

2.1 TOF カメラ

手話動作の計測装置として MESA 社の赤外線 3 次元距離測定カメラ SR4000 を使用する．このカメラは Time-of-Flight（飛行時間計測）の原理に基づき，距離を測定するカメラである．この原理はカメラから放射された赤外線が，対象物に反射してからカメラに帰還するまでの時間を計測し，それを距離データとして算出する．これにより奥行き情報の取得が可能となり，対象物の 3 次元形状を得ることが出来る．TOF カメラにより取得された距離データを可視化したものを図 1 に示す．

2.2 収録文章の作成

手話は様々な音韻により構成されており，種類も多く存在する．本研究では，多様な音韻を一定の語彙でカバーできるよう，音韻のバランスを考慮した 500 文章を選定し，TOF カメラによるデータ収集を行った．

手話単語の選定基準としては，基本会話を網羅する実用的な単語として，手話技能検定低級で出題される単語を用いた．音韻の種類・出現頻度の調整方法については，先行研究で発表されている音韻表記法 [3] に基づき音韻の分類を行い，出現頻度を調整した．図 2 に音韻表記法における手話の構造を示す．

†名古屋工業大学，Nagoya Institute of Technology

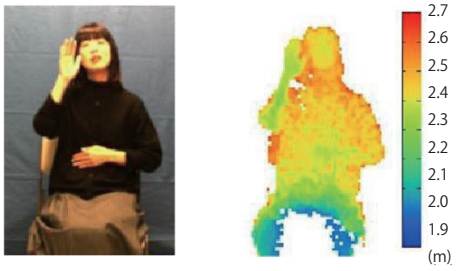

図 1: 通常のカメラ画像と距離データ

図 2: 日本手話音韻表記法における手話の構造

文章の作成については，選定された単語を用いて簡単な文構造のものを 500 文章作成する．文章を作成する上での条件として，各文章での単語数は 2〜7 語，各単語は全体で 4 回以上出現することとした．また，100 文章ごとの音韻の出現頻度がそれぞれ等しくなるよう調整した．

2.3 収録条件

先述の文章セットを用いて，TOF カメラを使用し手話動作を収録する．収録には手話通訳士の女性 1 名を対象にし，背景は一様とした．収録により得られるデータは，手話通訳士が行った手話動作のフレーム毎の距離座標である．収録後に，単語の順序の入れ替えや表現の違いなどを修正した．また，目視により手話単語の時間境界を付与した．この収録により構築されたデータベースの概要，また収録条件を表 1 に示す．

3. 連続手話認識実験

得られた 3 次元手話データを用いて，連続文章の手話認識実験を行う．フレームごとに 3 次元形状の特徴抽出を行い，単語単位で HMM を学習する．

表 1: データベース概要・収録条件

人数	手話通訳士 1 名（女性）
収録内容	手話 500 文章（単語：337 種類） 1 文章の平均単語数：3.5
取得データ	距離座標（m）
距離精度	±1cm
解像度	176（h）× 144（v）
フレームレート	30 frame/sec
フレーム数	1 文章あたり 120〜250 枚

表 2: 実験条件

学習単語数	298
状態数	10，15，20
混合数	1
特徴量	主成分得点 50 次元，Δ，Δ^2
実験データ	学習用：1 名× 343 文章（298 単語） 認識用：1 名× 100 文章（228〜236 単語） leave-one-out 法（3 セット）

3.1 実験条件

得られた 3 次元データのうち，動作者の頭部から腰までの領域に限定し，その区間（105 × 86）の距離データについて主成分分析（PCA）による次元圧縮を行った．各フレームごとに得られる主成分得点を HMM の学習データとして用いる．予備実験により主成分得点は 50 次元のものを使用する．

HMM は 10〜20 状態の Left-to-Right 型を用い，単語単位で学習する．認識時の単語間における文法等の制約は施していない．このため，認識結果には尤度の高いものが順に出力される形となっている．また，データベース構築において手話単語の修正を行った際，出現頻度が極端に少ない手話単語が存在した．このような単語では信頼性の高いモデルが学習されないため，一部の語彙を除外した．なお，HMM の学習と認識には HTK[4] を用いる．実験条件を表 2 に示す．

3.2 実験結果

認識結果を図 3 に示す．また，本実験における認識率は次式で表される．

$$Acc(\%) = \frac{H - S - I}{N}$$

N は認識単語の総数，H は正解単語数，S は置換誤り単語数，I は挿入誤り単語数を表している．この認識結果より，オープンデータはクローズデータの約半分の認識率となり，あまり対応できていないことが確認できる．この原因として，一つに学習データ不足が挙げられる．語彙設計によって出現頻度は調整したが，手話独特の文法表現等により，表現が変化してしまった手話単語がいくつか存在する．これより手話単語によって，作成される学習モデルの精度にばらつきが生じ，認識率の低下につながったと考えられる．また，前後の単語による影響も考えられる．文章の流れによって手話単語の表現方法に違いが生じ，手話を始める位置や向きなどが変化してしまい，学習モデルの精度を低下させてしまった恐れがある．また，今回は手話動作の画像全体に主成分分析を施して認識実験を行った．これより，手話動作に関連のない部分の情報を認識実験に使用し，認識結果に影響を与えた可能性がある．この改善方法として，手指動作，口の動き等，手話動作と関連が密な部分のみの情報を用いることで，認識率の向上につながると考えられる．

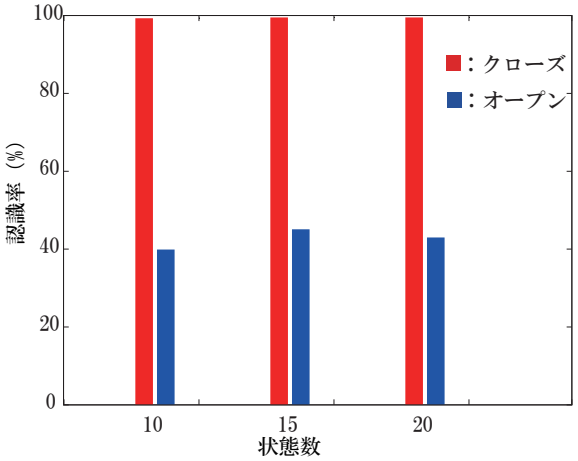

図 3: 単語認識率

4. むすび

本研究では，TOF カメラを用いて奥行き情報を取得し，得られた情報を基に手話認識実験を行った．また，先行研究より多くの語彙で認識実験を行うため，手話のデータベースを構築した．データベースには，手話文 500 文章，手話単語 337 種類が収録されている．今回，このデータベースを用いて連続手話認識実験を行った．認識結果として，オープンデータでは最高で 46.3% の認識率が得られた．認識率の向上としては，学習データ拡充，文法規則の導入，特徴抽出方法の改善等が考えられる．今後の課題として，データベースの拡充（対象者数，学習データ数等）のほか，通常のカメラ等と比較して，3 次元データを用いることの有効性を検証することが挙げられる．

謝辞 本研究の一部は文部科学省科学研究費補助金 (課題番号:22500506) によって行われた．

参考文献

[1] 佐藤 他，“TOF カメラによる 3D 手話認識”，MIRU2010，IS3-44，pp.1861–1868，2010.

[2] 柳生 他，“主成分分析を用いた HMM による手話認識”，FIT 講論集，K-3，pp.373–374，2002.

[3] 神田 他，“日本手話の音韻表記法”，日本手話学会，手和学研究，第 12 巻，pp.31–39，1991.

[4] HMM Tool Kit(HTK)，http://htk.eng.cam.ac.uk/