

マーカレスモーションキャプチャを用いた 距離画像による手話認識システム

西 悠介[†] 大川 茂樹[†]

[†]千葉工業大学 工学部 未来ロボティクス学科

1 はじめに

近年，障がい者の就労を支援する法制度の整備により，聴覚障がい者の積極的な社会進出が進んでいる．それに伴い，聴覚障がい者と健聴者との間のスムーズなコミュニケーションを必要とする機会が社会生活のさまざまな場面で増えつつある．聴覚障がい者とのコミュニケーションの手段として，手話や筆談などが挙げられる．聴覚障がい者同士のコミュニケーション手段として最も一般的な手法は手話である．聴覚障がい者と手話を知らない健聴者とのコミュニケーションは，多くの場合，筆談や手話通訳者を介して行われる．筆談を用いた場合，手話は健聴者が普段話している日本語とは文法などが異なる言語であるため，日本語の読み書きを苦手とする聴覚障がい者が存在し，スムーズな意思疎通ができない場合がある．また，健聴者も紙などに字を書くことは煩わしいことやコミュニケーションの速度が遅いなどの理由から負担を感じることもある．一方，手話通訳者を介する場合，聴覚障がい者と健聴者は負担なくコミュニケーションをとることができる．しかし，手話通訳者は慢性的に不足しており，十分な通訳サービスを受けられる状況にはなっていない．健聴者が多数を占める社会に生きる聴覚障がい者にとって，言語コミュニケーションの障壁は依然高いままである．情報技術の発展が進んだ現在，言語バリアフリーを目指すためには技術的側面からの支援が重要である．本研究では手話の習得に注目し，工学的アプローチによる手話通訳者の学習支援システムの開発を行った．

2 システム概要

手話通訳者の手話習得・手話学習支援ツールとして，手話入力システム，手話認識システムの開発を行った．ユーザが行った手話動作が正しいものかどうかを自動的に判断するシステムである．手話の入力は，ユーザの負担を考慮し非接触式センサである Microsoft 社の Kinect を使用して距離画像を画像処理し，手指の形状を得る [1]．手指の状態遷移の認識には，手にマーカなどを付着せずにモーションキャプチャを行う，マーカ

レスモーションキャプチャ[2]で行った．得られたピクセルデータから，掌，指先，指の節などの部位を特定し，手の 3D モデルを構築，この 3D モデルと実際に撮影している手を比較しながら，粒子群最適化 (PSO) で手指の状態遷移の追跡を行う．CG で表現した手モデルに沿って，ユーザはセンサの正面で手話動作を行い，手の状態遷移とあらかじめ用意した手話動作の学習データから手話の識別をし，識別結果を提示画面に出力する．

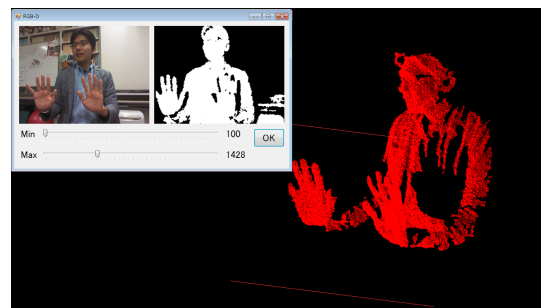


図 1: 距離画像出力の様子

2.1 手話入力システム

入力システムにおける手指の形状の抽出アルゴリズムについて述べる．ユーザには Kinect 正面で手を開いた状態で突き出す形の初期状態をとってもらい．Kinect の深度センサより得られた距離画像を用いて，最も大きな領域を手の領域とする．距離画像を用いることで，RGB 画像だけでは対応できなかった背景や照明による誤認識を防ぐことができる．手の領域内での最大内接円を求め，その中心を掌の中心とする．手の輪郭追跡をし，輪郭上の 3 点で表された角度と閾値を比較する K-curvature のアルゴリズムを用いて，指先，指の付け根にポイントを設置する．得られたポイントにより指の関節を計算し，ポイントを結んで骨組モデルを作成する．骨組モデルに沿って 3D の手モデルを成形する．手モデルと実際に Kinect の前で撮影している手の映像を合わせ，差異が最小になるように手モデルを遷移し，マーカレスモーションキャプチャを行う．遷移を求めることで動作や手の形状が似ている手話を区別することができる．対象となるデータは片手の場合，掌の重心，各指先，各指関節の合計 20

Sign language recognition system based on distance image using markerless motion capture

NISHI Yusuke[†] and OKAWA Shigeki[†]

[†]Dept. Advanced Robotics, Chiba Institute of Technology

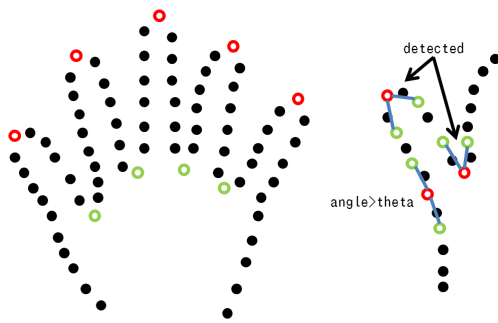


図 2: K-curvature のアルゴリズム

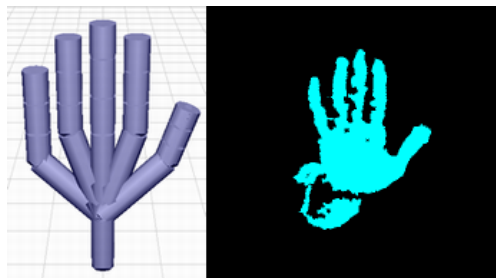


図 3: 左：手の骨組モデル，右：深度画像

パラメータである．手モデルの遷移との差異は PSO を用いることで収束させる．

2.2 手話認識システム

入力システムで得られた各手指状態 20 パラメータの 3 次元位置データを 200ms ごとに記録し，手話辞書を作成した．連続的な手話パターンの認識を行うため，識別には技術的な容易さの観点から DP マッチングを使用した．ユーザが手話動作を行うと，登録した辞書と照合し，距離が最も少ない手話単語を認識単語とし出力する．

3 検証実験

入力システムの性能を確認するため，認識実験を行った．今回は日本手話の指文字 76 文字 (清音，濁音，半濁音，長音) を遷移データとして記録し辞書とした．指文字とは手の形を文字言語に対応させた視覚言語の 1 つであり，五十音を片手で表現するものである．「あ」「い」「う」「え」「お」を各 20 パターン入力し，システムで認識させ辞書と照合した．また，手の状態遷移の認識を確認するため，動きのある指文字の認識を行った．指文字における濁音は，右手の場合，清音の形を保ったまま手を右に移動させることで表現できる．そこで「か」「き」「く」「け」「こ」と「が」「ぎ」「ぐ」「げ」「ご」の各 20 パターン入力し評価した．入力時間長は 3 秒間で固定し，その間に得られた手の遷移データを入力とした．静止する指文字につい

表 1: 指文字 5 文字の正解率

あ	い	う	え	お
90%	90%	95%	80%	100%

表 2: 動きのある指文字の正解率

か	き	く	け	こ
90%	90%	85%	90%	85%
が	ぎ	ぐ	げ	ご
60%	55%	60%	65%	60%

ては手の状態を 3 秒間維持したものを入力とした．正解数を試行回数で割った正解率を表 1 と表 2 に示す．

4 おわりに

手話教則本や，手話サークルなどで手話を習得する既存の方法に加えて，工学的アプローチによる手話学習の支援システムの一部として，マーカレスモーションキャプチャを用いた距離画像による手話認識システムを提案した．距離画像を用いた手モデル形成とモデルと実映像を比較するマーカレスモーションキャプチャを用いることで，手指の状態遷移の認識を試みた．認識実験の結果，静止した指文字の認識は高い認識率を得られた．動きのある指文字については遷移したデータを反映できず，「が」を「か」と誤認識するケースなどがあり，認識精度は高くなかった．

今後は，より精度の高い認識システムを開発し，さまざまな状況に柔軟に対応できるシステムを構築する必要がある．また，今回は結果出力時間を考慮していないため，リアルタイムで手話認識を行える効率的なシステムの考案も必要である．加えて，ユーザの熟練度を考慮し，提示される手話映像の再生速度を可変にすること，特殊な文法体系を用いる手話に対応して日本語の文法と合わせて手話の文法を提示する，次に動作する手話映像を表示する教示手話 CG の実装などの，より学習を行いやすいインタフェースを検討する．

参考文献

- [1] 西村 洋介，今村 大輔，堀内 靖雄，川本 一彦，篠崎 隆宏，黒岩 眞吾，“Kinect とパーティクルフィルタを用いた HMM 手話認識手法の検討，”電子情報通信学会技術研究報告，111(431)，pp.161-166，(2012)
- [2] I.Oikonomidis, N.Kyriazis, A.Argyros, “Efficient model-based 3D tracking of hand articulations using Kinect,” BMVC 2011, (2011)