

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 31 日現在

機関番号：13903

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25350666

研究課題名(和文) 手指動作と非手指動作のサブユニットモデルに基づく手話認識に関する研究

研究課題名(英文) Sign Language Recognition based on Subunits of manual signals and non-manual signals

研究代表者

北村 正 (Kitamura, Tadashi)

名古屋工業大学・工学(系)研究科(研究院)・教授

研究者番号：60114865

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：手話は手指動作と非手指動作で構成される。手指動作は手の形、動き、位置によって単語の意味を表し、非手指動作は表情、頭部の傾きなどにより文法規則を付加する働きを行う。

手指動作認識では、2名が223単語を2回生成したデータに対して認識を行った。手話の3基本要素、手形状(24種類)、位置(8種類)、動き(40種類)の認識実験を行い、28.7%、78.3%、60.0%、重み付け和の単語認識率は33.8%となった。上位10位以内の認識率は79.7%であった。非手指動作認識は、出現頻度の高い傾き、顎上げ、顎下げの3種類を対象とした。66文章で学習し125文章を認識した結果、正解率は79%であった。

研究成果の概要(英文)：A sign language consists of manual signals and non-manual signals. Manual signals convey the meaning of the word by three elements, e.g. the shapes, the movements and the positions of the hands. Non-manual signals such as facial expressions, head tilting etc. work to add grammatical functions.

We carried out recognition experiments for manual signals using database recorded by 2 persons, 223 words, twice. The results have shown that the recognition accuracies for 24 hand shapes, 8 positions, 40 movements, are 28.5%, 78.3%, 60.0% respectively. The word recognition accuracy rate was 33.8% using combination of weighted scores of each element. The recognition accuracy rate for top 10 was 79.7%. The recognition experiment of non-manual signals for 3 different motions of head tilting was performed. In the experiments 66 sentences were used for training the Hidden Markov Models of the motions. The recognition rate was 79% for 125 test sentences.

研究分野：情報工学

キーワード：手話認識 手指動作 非手指動作 Kinect

1. 研究開始当初の背景

(1)視覚言語としての手話は聴覚障害者のコミュニケーションに不可欠なものである。しかし、手話を対象とする国内外の研究では、手話のデータ収集にかかる困難さや、手話の言語体系が未確立であるなど様々な問題があり、定まった知見がまだ十分に得られていない状況にある。アメリカ手話やドイツ手話をはじめとして、各国の手話を中心として手話の自動認識に関する研究が行われているが、未だ手話単語単位などの限定的なものが多く、近年の音声認識のような大規模な連続手話単語認識や、文章レベルの手話認識の研究はまだ少ない現状にある。その原因の一つとして、手話では音声言語と比べて言語資源の集約がまだ十分ではなく、同時にそれらの知見を手話認識の枠組みに活用できていない点が挙げられる。

(2)このような視点から我々の研究グループでは、実用的な手話認識を実現するための取り組みとして、サブユニット(音声言語の音素に相当する、特定の意味をなさず、語彙間で共通にみられる基本動作単位とする呼称)の考え方を提唱し、このサブユニットの組み合わせが特定の手話単語を構成するものとして、サブユニットによる手話動作のモデル化・認識を行う枠組みを提案してきた。この方法は、音声認識で主流となっている隠れマルコフモデルの枠組みを手話動作のモデル化に応用している。その中で、手話表現を基本的な動作要素に分解することで、限られた手話動画データから様々なサブユニットを自動的に定め、複数の語彙の中でそれらを共有する効率的なモデル化手法である。

2. 研究の目的

手話の動作は、手の形や動きなどによって表現される手指動作と、口の形や表情、視線などによって表現される非手指動作等から構成される。手指動作は主に手話単語の意味を表し、非手指動作は文法規則を付加する役割がある。そのため手話認識では、手指動作認識と非手指動作認識が必要となる。

手指動作は音韻学的に手形状、手の位置、動きの3要素から構成されていると言われている。また、3000単語程度の手指動作は、それらの3要素に分解すると100種類程度となることが知られており、数千程度の手指動作認識を考える場合は、それらの3要素の認識を行うことが効率的であると考えられる。そのため、基本単位としてこれらの3要素の認識を行う。

一方で、それら動作の特徴抽出を高速に行い、認識することも重要である。従来は画像を利用する方法が主であり、計算コストも高く実時間処理が困難であったが、近年大幅に技術が進んだ深度センサを活用することにより、大幅に低コスト化・高速化が可能とな

ってきた。これらを踏まえて、近年利用可能になった Kinect v2 を用いて実時間で動作する手話認識システムを実装することを考える。また、今後の研究の発展のため、手話通訳士による手話単語データベースを収録することも考える。

3. 研究の方法

(1)手指手話の基本動作

手話は音韻学的に、主に手形状、手の位置、動きの3要素から構成されていると言われ、手話認識においてもこれらの要素を認識することが重要であるとされてきた。

本研究では、手話の基本要素を利用することを考えてきたが、上記の考えに基づき、深度センサから得られた特徴量を3要素におけるパターンに変換し、それらを基本単位として学習、認識を行う手法を提案する。そのため、他研究者により提案されている日本語・日本語辞書システム(見出し数2596単語)を利用する。この方法を、我々が従来考えていたサブユニットを用いる方法に変わるものとして用いることとした。辞書に記載された2596単語は、これらの3要素で記述すると、手の形状39種類、位置17種類、動き44種類の計100種類の基本ユニットにより記述できる。そのため、これらの基本ユニットを学習、認識に利用すればいいので非常に効率的であり、今後の語彙拡張も容易に行うことができる。

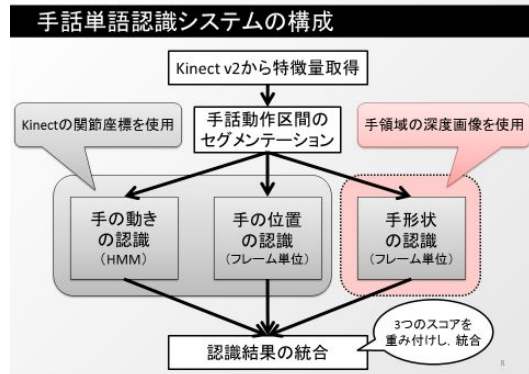


図1 手話単語認識システム(手指動作)

学習、認識についてのブロック図を図1に示す。手の位置は混合正規分布、手の動きは手座標の時系列を隠れマルコフモデル、手形状は輪郭特徴を用いてモデル化し、それらの3要素を個別認識した後、3要素のスコアの重み付け和で手話単語のスコアとする。

(2)非手指動作の認識

研究では、出現頻度の高い手指動作である顎き、顎上げ、顎下げを対象とし、動作全体を3状態の隠れマルコフモデル(1次元、単一正規分布)によりモデル化する。1次元データとしては、抽出された額と頬の座標から計算した頭の傾きを用いた。その時系列から

静止、頷き、顎上げ、顎下げの区間をラベル付けしたものを、非手指動作のモデルの学習、評価に利用する。

(3) データベース収録

今後の手話認識に必要なデータベースの構築のため2名の手話通訳士により全国手話検定5級の出題単語全400単語を1名につき2回ずつ動作してもらい、Kinect v2により1600単語を収録する。非手指動作は、1名の手話通訳士により平叙文、肯定疑問文、命令文を各々72文ずつ収録する。また、手指動作、非手指動作のための特徴量を抽出し、特徴量をもとにセグメンテーションを手動で行う。更に、各データのラベル付けも手動で行い、学習、評価データとして用いる。

(4) 実時間認識システム

実時間認識システムの構築のため、特徴抽出には、奥行き情報を含む3次元特徴量の抽出が容易に行えるという理由から、深度センサであるKinect v2と利用ソフトであるKinect for Windows SDK 2.0を使用する。これらを利用して汎用の計算機で実時間認識システムを構築する。

4. 研究成果

Kinect v2を用いた汎用計算機により実時間認識システムを構築して、表1の実験条件で実験を行った。

表1 手指動作認識実験の条件

認識対象	223単語×話者2名(手話通訳士)×2回=892データ
動き	種類 40種類
学習データ	動き1種類につき疑似的な手の動き10回
特徴量	3次元の方向ベクトル+速さ
HMMのパラメータ	状態数:5~8, GMMの混合数:1
位置	種類 8種類
特徴量	3次元の利き手座標
GMMの混合数	1~6(位置の種類によって異なる)
学習データ	1種類の位置につき: 該当する位置で円を3回連続して描く間の手の座標×6セット
手形状	種類 24種類 (手型と掌の向きの組み合わせにより決定)
深度画像の大きさ	120×120 pixel(縦30×横30×奥行35cm)
輪郭特徴系列の長さ	30(輪郭線上の点のサンプル数)
ベルトの数	手型×掌の向き×話者につき25
学習データ数	手型×掌の向き×話者につき150枚

(1) 手指動作の認識

手話の言語学的知見に基づき、手話の3要素である手の位置、動き、形状によって手話単語を定義しモデル化する方法で認識を行った。本研究では、基本動作の学習はデータベースとは異なる他の疑似的なデータを用いて行った。223単語の未知語に対して孤立手話単語の認識実験を行った。実時間手話単語認識を行うため、Kinect v2を用いてそれぞれの特徴量を抽出した。

3要素である手の形状(24種類)、位置(8種類)、動き(40種類)の個別の認識率はそ

れぞれ、28.7%、78.3%、60.0%となった。単語認識率は、形状:位置:動きに対して、0.1:0.4:0.5の重み付け和のスコアを用いることにより33.8%となった。また、上位10位以内の単語認識率は79.7%であった。

実験条件は異なるが、本実験では単語データから学習を行った先行研究と同等、またはそれ以上の精度が得られており、提案法の有効性が確認できた。

(2) 非手指動作の認識

1名の手話通訳士のデータを対象とし、平叙文、肯定疑問文、命令文で表出される、出現頻度の高い頷き、顎上げ、顎下げを認識対象とした。Kinect v2により得られた額と頬の座標を用いて頭の傾きを求めて特徴量とした。その時系列から静止、頷き、顎上げ、顎下げの区間をラベル付けした非手指動作のデータベースを作成した。非手指動作は、3状態の隠れマルコフモデル(1次元、単一正規分布)によりモデル化した。

実験では、平叙文、肯定疑問文、命令文を各々72文ずつ収録し、学習用に66文、認識用に125文のデータを用いた。認識対象は、平叙文(頷き36データ)、肯定疑問文(顎上げ45データ)、命令文(顎下げ44データ)である。3種類の頭部動作全体の正解率は79%であった。

(3) 実時間動作実験

実験では実時間動作を目指して、Kinect v2と下記の性能の汎用の計算機を用いた。手指動作の3要素の認識時間はそれぞれ0.1秒以内となった。今後の計算機の性能向上を考えると実時間動作はほぼ実現出来ると考える。

表2 計算機の性能

OS	Windows 8.1 64 bit
CPU	Intel Core i5-4547 クロック:3.20 GHz, コア数:4
メモリ	4GB

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 5件)

Shinji Sako, Mika Hatano, and Tadashi Kitamura, "Real-time Japanese sign language recognition based on three phonological elements of sign", HCI 2016 International Conference on Human-Computer Interaction, 査読有, 採択済み. DOI:未定, 謝辞の記載あり
酒向 慎司, 北村 正, "隠れマルコフモデルによる手話の音韻構造に基づいた

自動手話認識”, 日本福祉工学会論文誌, Vol. 17, No. 2, pp. 2-7, 2015. DOI: なし, 謝辞の記載あり

Mika Hatano, Shinji Sako, and Tadashi Kitamura, “Contour-based Hand Pose Recognition for Sign Language Recognition”, Proc. of 6th Workshop on Speech and Language Processing for Assistive Technologies, 査読有, Sep. 2015. DOI: なし, 謝辞の記載あり

Filip Malawski, Bogdan Kwolek, and Shinji Sako, “Using Kinect for Facial Expression Recognition under Varying Poses and Illumination”, Active Media Technology Lecture Notes in Computer Science, LNCS 8610, 査読有, pp. 395-406, Aug. 2014. DOI: 10.1007/978-3-319-09912-5_33, 謝辞の記載あり

Shinji Sako, and Tadashi Kitamura, “Subunit Modeling for Japanese Sign Language Recognition Based on Phonetically Depend Multi-stream Hidden Markov Models”, UAHCI/HCI 2013, Part I, LNCS 8009, 査読有, pp. 548-555, Jul. 2013. DOI: 10.1007/978-3-642-39188-0_59, 謝辞の記載なし

〔学会発表〕(計 6 件)

加藤 里奈, 酒向 慎司, 北村 正, “HMM を用いた手話の非手指信号の認識”, 電子情報通信学会 2016 年総合大会 学生ポスターセッション, ISS-SP-255, p. 254, Mar. 2016.

波多野 美歌, 酒向 慎司, 北村 正, “Kinect v2 による手話動作の 3 要素に基づく実時間手話認識”, 電子情報通信学会研究会技術報告 福祉情報工学研究会 (WIT), Vol. 115, No. 491, WIT2015-99, pp. 59-64, Mar 2016.

波多野 美歌, 酒向 慎司, 北村 正, “手話認識のための輪郭特徴を用いた手形状認識”, 電子情報通信学会研究会技術報告 福祉情報工学研究会 (WIT), Vol. 115, No. 100, WIT2015-31, pp. 175-180, Jun. 2015.

不破 大樹, 酒向 慎司, 北村 正, “Kinect を用いた手話の非手指信号の認識”, 電子情報通信学会 2015 年総合大会 ISS 学生ポスターセッション, ISS-SP-163, p. 163, Mar. 2015.

波多野 美歌, 酒向 慎司, 北村 正, “手話動作の 3 要素に基づく実時間手話認識”, 電子情報通信学会研究会技術報告 福祉情報工学研究会 (WIT), Vol. 114, No. 92, WIT2014-13, pp. 69-74, Jun. 2014.

小山 智己, 齊藤 剛史, 酒向 慎司, 北村 正, “Kinect を用いた手話認識”,

計測自動制御学会第 32 回九州支部学術講演会, 103A5, pp. 159-162, Nov. 2013.

〔図書〕(計 0 件)

〔産業財産権〕
出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕
ホームページ等
<http://www.mmsp.nitech.ac.jp/>

6. 研究組織

(1) 研究代表者

北村 正 (KITAMURA, Tadashi)
名古屋工業大学・大学院工学研究科・教授
研究者番号: 60114865

(2) 研究分担者

酒向 慎司 (SAKO, Shinji)
名古屋工業大学・大学院工学研究科・助教
研究者番号: 30396791