

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：12102

研究種目：基盤研究(B) (一般)

研究期間：2013～2015

課題番号：25282173

研究課題名(和文) 手話からの特定連続指文字のスポッティングシステム

研究課題名(英文) Spotting Finger Spelled Words from Sign Language Video

研究代表者

福井 和広 (Fukui, Kazuhiro)

筑波大学・システム情報系・教授

研究者番号：40375423

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：スポッティングを実現するために、時間正則化付き正準相関分析(TRCCA)に、複数スケール対応機能と非線形直交化による特徴抽出を付加したKOTRCCAを提案した。8種類の指文字単語を混在させた手話発話動画からのスポッティング実験において、TRCCAの64.1%に対してKOTRCCAは70.2%と大きく性能向上することを確認した。また別アプローチとして、フレーム順を保存しつつ、動画像列からランダム選択した複数画像特徴を連結したTime Elastic Random 特徴を考案し、これをベースにしたRandom Time wrapping法の有効性を評価実験で確認した。

研究成果の概要(英文)：In this research, we developed a new method for spotting specific finger alphabet words from an input sign language video. We addressed the spotting task by employing the basic idea of temporal regularized canonical correlation analysis (TRCCA). The classification accuracy of TRCCA is enhanced by incorporating two functions: 1) parallel processing with multiple time scales, 2) implicit feature mapping by nonlinear orthogonalization. The enhanced TRCCA is called "kernel orthogonal TRCCA (KOTRCCA)". We demonstrated its effectiveness through the experiments in the task of spotting 8 kinds of finger alphabet words from a sign language vide. In addition, as another approach to spotting, we developed a randomized extension of the DTW, termed randomized time warping (RTW), which generates time elastic (TE) features by randomly sampling the sequential data while retaining the temporal information. We also demonstrated the applicability of RTW through experiments on three public datasets.

研究分野：コンピュータビジョン, パターン認識

キーワード：指文字 手話 スポッティング 正準相関分析 部分空間表現

1. 研究開始当初の背景

手話の見出し語に含まれない、人名、地名、外来語、専門用語などは、日本語50音やアルファベットに対応した指文字の組み合わせで表現される。この連続指文字は任意の用語を表現できる反面、指文字形状が連続的に変化していくために、手話熟達者の手話通訳者や聴覚障害者でも、安定かつ正確に把握することは難しい。特に講演などの一方のコミュニケーションにおいて、呈示された連続指文字が受け手にとって初出の場合にはさらにその困難さが増す。この状況を改善する様々な研究が行われていた。例えば、研究分担者らは、手話通訳映像とキーワード(専門用語や数値等の字幕)を合成して聴覚障害の学生に提示するシステムを開発した。しかし、この方法では、通訳者の手話に合わせて補助者がタイミング良くキーワードを提示する必要があり、キーワードリストが膨大になってくると運用困難となる。また音声認識字幕の可能性も検討済みであるが、安定な音声認識の実現において課題が多かった。このような状況を改善する一つの方策が、手話から事前登録した特定の連続指文字をリアルタイムでスポッティング(検知・認識)である(図1参照)。

このようなスポッティングに際して、留意すべき点は、複数の指文字から構成されている連続指文字は、単にそれらを語順に呈示したのではなく、会話テンポに合わせて各指形状は滑らかに音韻変化していくことであった。そのために、指文字間の境界は曖昧で、各指文字の基本形状が明確に現れることは希で、連続指文字では指形状が滑らかに変化していく。このために、連続指文字からの指文字のセグメントが困難となり、一文字毎に認識するというアプローチは取り難く、連続指文字を"一つのまとまり"として捉えるアプローチの方が有利となることが知られていた。代表者らが先行研究で開発した指文字間の3次元形状の類似度を測る"形状類似度"はこの条件を満足する方法であった。その最

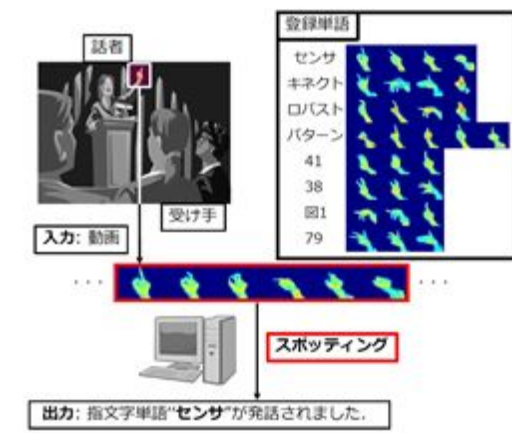


図 1

大の特徴は、従来法では標準的に行う手の形状復元を陽に行わず、比較する指文字AとBの多視点画像セットSaとSbから両者の類似度を求める事にある。各画像セットは高次元ベクトル空間中の非線形部分空間PとQで表され、最終的にはPとQの成す正準角が指文字AとBの形状類似度となる。形状類似度は、従来法のように3次元モデルを使わないために汎用性が極めて高い。

しかし、この類似度では複雑な動きを伴う連続指文字を安定かつ高精度にスポッティングすることは難しかった。主な要因は、各指文字の呈示順の情報(時間変化の情報)が反映されていないことであった。さらに、同一人物が呈示する語句を対象とするとは言え、発話状況により各指文字の呈示タイミングは微妙に変化する。この時間方向の局所伸縮への対応が不十分なことも要因であった。

2. 研究の目的

本研究の目的は、手話の中から、発話者自身が予め登録した特定の連続指文字を、多視点距離動画像を用いたパターン認識により検知・抽出(スポッティング)する方法を開発し、その有効性を検証することである。このために、以下の3つの課題に取り組んだ。1)先行研究で開発した手形状類似度の有効性を生かしつつ、指文字の呈示順も考慮した新しい類似度(動き形状類似度)や識別アルゴリズムを開発する、2)動き形状類似度に基づいてスポッティングシステムを試作する、3)本システムの有効性・問題点を評価実験により明らかにする。

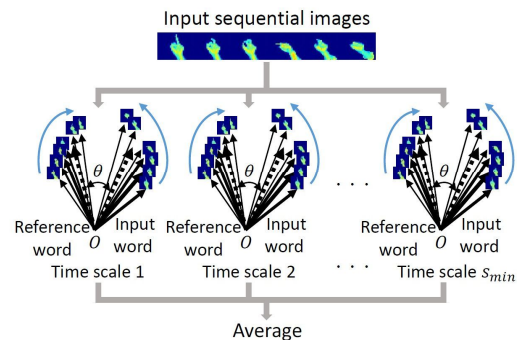


図 2

3. 研究の方法

本研究では、上記課題に対応するために、スポッティングを実現するための2つの異なる方法を開発し、それぞれの有効性を検証する。一つが時間正規化付き正準相関分析(TRCCA)の理論拡張である Kernel orthogonal TRCCA (KOTRCCA) (図2)である。もう一つが、入力画像の時間順序を考慮した Time Elastic Random 特徴に基づく Random time wrapping (RTW)法である。

(1)KOTRCCA によるスポッティング: スポッティング対象となる指文字単語の順序情報を

考慮するために、時間情報と形状情報を同時に扱うことが可能な時間正則化付き正準相関分析(TRCCA)に着目した。TRCCAによるスポッティング性能をさらに強化するために、TRCCAに以下の2つの機能を付加した。(a)マルチ時間スケールへの対応機能：複数の時間スケールでTRCCAを並列処理し、それらから得られる複数の類似度を重み付き平均して、統合類似度を求める。TRCCAでは、時間方向の連続性を考慮するために、CCAを計算する際に、隣接する重み係数が滑らかに変化するという制約を付加している。この滑らかさの局所性を制御することで、異なる時間スケールに対応し、様々なスケールの時間伸縮の影響を吸収する。(b)クラス部分空間の直交化：非線形直交化により、各クラス非線形部分空間の関係を直交状態に近づけておくことで、各単語同士の分離性能を向上させて、スポッティング性能の向上を目指す。各アルゴリズムの詳細については論文を参照のこと。

(2)KOTRCCAに基づくスポッティングシステムの構築：

KOTRCCAに基づくスポッティングの流れは図3に示す通りである。入力としてはKinect距離センサから得られる3次元情報も含んだ距離画像を用いる。まず学習時に発話者自身が登録したい連続指文字をKinect距離センサに向けて呈示する。システムは多視点動距離画像の各画像から手領域を切り出し、辞書

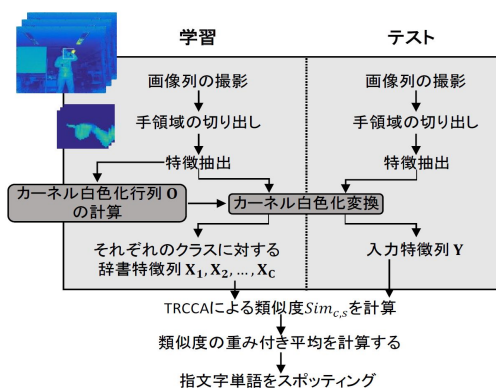


図 3

| | | |
|------|------|--|
| クラス1 | センサ | |
| クラス2 | キネクト | |
| クラス3 | ロバスト | |
| クラス4 | パターン | |
| クラス5 | 41 | |
| クラス6 | 38 | |
| クラス7 | 図1 | |
| クラス8 | 79 | |

図 4

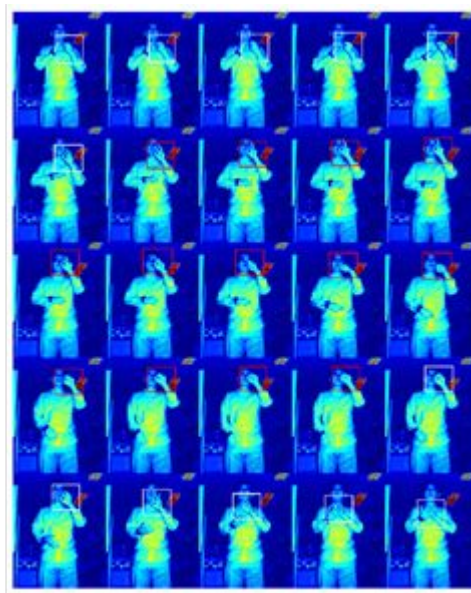


図 5

部分空間 D_s を生成する。辞書部分空間は連続指文字クラス ($s=1-C$) 毎に生成する。運用時には、入力される多視点動距離画像から手領域を抽出し、非線形入力部分空間を生成する。非線形入力部分空間と各クラス部分空間との正準角を測り、それらの最大値がしきい値以上に対応する連続指文字クラスがスポッティングされたと判断しモニタ表示する。

(3)KOTRCCAに基づくスポッティング性能の評価：

講師が手話と日本語指文字表現による単語呈示を混在させて、ある文章を発話する状況で性能評価を行った。混在する指文字単語は図4に示す8種類である。データの収集にあたっては手話常用者に依頼し、動画の撮影を行った。図5に収集した動画像列の一部を示す。

(4)Time Elastic Random (TER)特徴：動画像列からその呈示順を保存しつつ、ランダム選択した画像パターンを連結したものを新しい特徴パターン(Time Elastic Random (TER)特徴)を考案した(図6)。TER特徴は元の動画像の時間方向の情報を保持しつつ、時間方向の多様な局所伸縮を施した特徴ベクトルとなっている。

(5) Randomized Time Warping (RTW)法：上記で順序制約付きランダム選択により生成されたTER特徴ベクトルのセットを用いた識別法がRTW法である。予め、スポッティング対象となる各単語クラスのTEA特徴セットに対して、非線形主成分分析を適用して非線形部分空間を生成する。同様に連続入力される手話から特定区間の画像列を抽出し、入力非線形部分空間を生成する。入力非線形部分空間と各単語クラス非線形部分空間の成す正準角が入力の各指文字クラスに対する動き形状類似度となる。

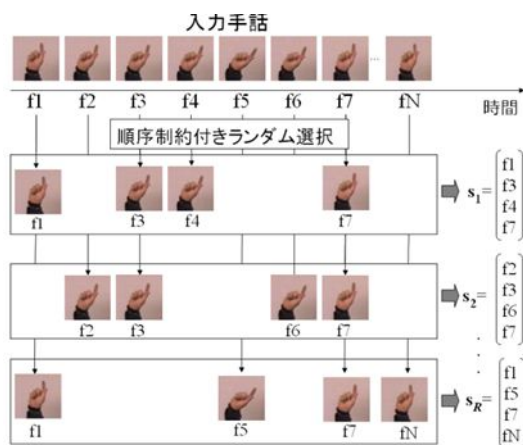


図 6

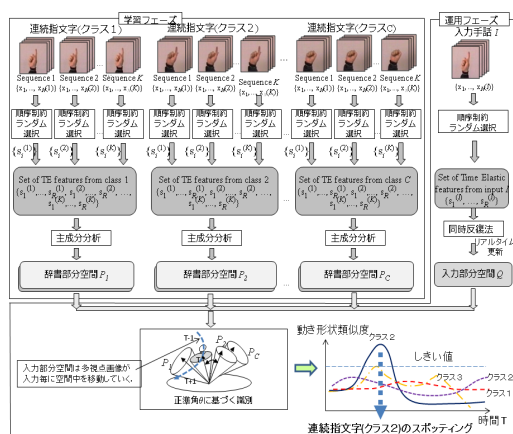


図 7

スポッティング全体の流れは図7に示すようになる。全体はオフライン処理の学習フェーズとオンライン処理の運用フェーズから構成される。(i)学習フェーズでは、カメラに向けて呈示された連続指文字の多視点動画画像から手領域を抽出し、それから抽出されたTER特徴セットに非線形主成分分析を適用して、非線形部分空間を生成する。登録したいC個の連続指文字に対してこれらの部分空間を用意する。(ii)運用フェーズでは、多視点画像から手領域が抽出される毎に、入力非線形部分空間Qを同時反復法によりリアルタイム更新する。そして入力非線形部分空間Qと各非線形辞書Piクラス部分空間の成す正準角を動き形状類似度として求める。こうして入力ステップ毎に、Cクラス分の動き形状類似度が得られることになる。これらの類似度の時間変化をモニタリングして、最大類似度がしきい値よりも大きくなった時、該当する連続指文字クラスがスポッティングされたと判断する。

4. 研究成果

本研究では以下の研究成果を得た。

(1) スポッティング実現に向けてコア技術となる部分空間表現に基づく形状類似度の算

出法の性能強化を図るために、実環境での使用を想定した様々なタイプの指文字認識システムを構築し、改良すべき点を探った。その結果を下記の各開発項目に反映させた。

(2) KOTRCCA の考案

安定なスポッティングを実現するために、時間正則化付き正準相関分析 (TRCCA) の機能強化した KOTRCCA を構築した。

(3) KOTRCCA に基づくスポッティング

手話発話に8種類の指文字単語を混在させた動画画像に対して、従来法 (TRCCA, KOMSM, KOCCA) との性能比較実験を行い、TRCCA 64.16% / KOMSM 67.90% / KOCCA 63.93% に対して、提案法は 70.21% の識別率を実現し、提案法の有効性を確認した。

(4) TER 特徴量の考案

動画像列からその呈示順を保存しつつ、ランダム選択した画像パターンを連結したものを新しい特徴パターン (Time Elastic Random (TER) 特徴) を考案した。

(5) RTW 法の考案

公開データベース Cambridge hand gesture database を用いて、標準的な方法である DTW に比べて、RTW 法の識別率は大きく向上することを確認した。さらに、最新識別法である Hankel 行列を用いた方法に比べても 92.9% から 93.6% へと識別性能を更に向上できた。また KTH action dataset を用いた評価でも、最新法と同等レベルの識別性能を実現できることを確認した。本研究成果は英文ジャーナル Image and Vision Computing に投稿し、現在、2 回目の査読が完了し、査読者からの追加コメントに対応中である。

(6) 今後の課題

今回の評価では、熟達者によって発話される1種類の文章の発話データで実施した。今後はさらに大規模なデータを収集し、より総合的な評価を行うことが課題として挙げられる。また、スポッティングが有効となる場面では、手話と指文字の発話比率は、前者が後者に比べて圧倒的に大きい。手話には指文字の手形状と類似したクラスが多数ある。この事を考慮すると、前処理として、手話と指文字の2分類する識別器を用意しておくことが有効であると考えられる。このように、特定の指文字以外のデータもライバルクラスとして利用することで、スポッティング性能の向上が可能であることは、予備実験により明らかになっている。

5. 主な発表論文等

〔雑誌論文〕(計12件)

Shouhei Tanaka, Akio Okazaki, Nobuko Kato, Hideitsu Hino, Kazuhiro Fukui,

“ Spotting Fingerspelled Words from Sign Language Video by Temporally Regularized Canonical Component Analysis ”, The IEEE International conference on Identity, Security and Behavior Analysis (ISBA2016), IEEE ISBA2016, pp.1-11, 2016. DOI: 10.1109/ISBA.2016.7477238. 査読有

近藤真暉, 加藤伸子, 福井和広, 岡崎彰夫, “ 動きを伴う指文字を含めた距離画像を用いた対話的指文字練習システムの開発と評価 ”, 電子情報通信学会 WIT 研究会技報, vol.114, pp.23-28, 2015. 査読無

菊池優介, 日野秀逸, 福井和広, “ 距離画像から得られる形状空間に基づく手の形状識別と姿勢推定 ”, 情報処理学会 CVIM 技報, vol.196, pp1-8, 2015. 査読無

Daisuke Takabayashi, Hideitsu Hino, Kazuhiro Fukui, “ Finger alphabets recognition with multi-depth images for developing their learning system ”, FCV2014, pp.154-159, 2014. 査読有

Bernardo Gatto, Hideitsu Hino, Kazuhiro Fukui, “ Block-based KOMSM for hand shape recognition with occlusion ”, FCV2014, pp. 375-381, 2014. 査読有

近藤真暉, 加藤伸子, 福井和広, 岡崎彰夫, “ 距離画像を用いた動きのある指文字を含めた指文字練習システムの開発 ”, ヒューマンインタフェースシンポジウム 2014 予稿集, vol.2532, pp.1-6, 2014. 査読無

田中翔平, 日野英逸, 加藤伸子, 岡崎彰夫, 福井和広, “ 距離動画像列からの特定ジェスチャのスポットティング ”, FIT2014, K-027, pp1-4, 2014. 査読無

渋谷美彩, 田中翔平, 日野英逸, 加藤伸子, 岡崎彰夫, 福井和広, “ 指文字認識システムにおける効果的な辞書生成の検討 ”, FIT2014, K-026, pp.1-4, 2014. 査読無

Yan Zhang, Chendra Hadi SURYANTO, Kazuhiro Fukui, “ Hand shape recognition using position invariant feature from the sequence of depth images ”, FIT2014, J-014, pp1-4, 2014. 査読無

Graciliano Gacia Torres Galindo Jr,

Chendra Hadi SURYANTO, Kazuhiro Fukui, “ Gesture recognition with multiple features from hands and face using depth sensor ”, FIT2014. J-013, pp.1-4, 2014. 査読無

高林大輔, 田中陽士, 加藤伸子, 岡崎彰夫, 福井和広, “ フィードバック機能を備えた指文字学習支援システムの開発 ”, ビジョン技術の実利用ワークショップ (ViEW2013), IS2-C1, 2013. 査読無

田中陽士, 高林大輔, 近藤真暉, 加藤伸子, 福井和広, 岡崎彰夫, “ Kinect を用いた指文字練習システムにおける識別処理とユーザインタフェースの改良 ”, 電子情報通信学会 HCG シンポジウム 2013 論文集, pp.291-296, 2013. 査読無

[学会発表](計13件)

Shouhei Tanaka, Akio Okazaki, Nobuko Kato, Hideitsu Hino, Kazuhiro Fukui, “ Spotting Fingerspelled Words from Sign Language Video by Temporally Regularized Canonical Component Analysis ”, The IEEE International conference on Identity, Security and Behavior Analysis (ISBA2016), 2016年2月29日~3月2日. 東北大学(宮城県 仙台市)

Chendra, Jing-Hao Xue, Kazuhiro Fukui, “ Randomized Time Warping (RTW) for Motion Recognition ”, The IEEE International conference on Identity, Security and Behavior Analysis (ISBA2016), 2016年2月29日~3月2日. 東北大学(宮城県 仙台市)

近藤真暉, 加藤伸子, 福井和広, 岡崎彰夫, “ 動きを伴う指文字を含めた距離画像を用いた対話的指文字練習システムの開発と評価 ”, 信学技法 WIT 研究会, 2015年3月13日. 筑波技術大学(茨城県 つくば市)

菊池優介, 日野秀逸, 福井和広, “ 距離画像から得られる形状空間に基づく手の形状識別と姿勢推定 ”, 情報処理学会 CVIM 研究会, 2015年3月6日. 東北大学(宮城県 仙台市)

近藤真暉, 加藤伸子, 福井和広, 岡崎彰夫, “ 距離画像を用いた動きのある指文字を含めた指文字練習システムの開発 ”, ヒューマンインタフェースシンポジウム 2014 予稿集, 2014年9月11日. はこだて未来大学(北海道 函館市)

田中翔平, 日野英逸, 加藤伸子, 岡崎彰夫, 福井和広, “距離動画像列からの特定ジェスチャのスポッティング”, FIT2014, 2014年9月4日. 筑波大学(茨城県 つくば市)

渋谷美彩, 田中翔平, 日野英逸, 加藤伸子, 岡崎彰夫, 福井和広, “指文字認識システムにおける効果的な辞書生成の検討”, FIT2014, 2014年9月4日. 筑波大学(茨城県 つくば市)

Yang Zhang, Chendra Hadi SURYANTO, Kazuhiro Fukui, “Hand shape recognition using position invariant feature from the sequence of depth images”, FIT2014, 2014年9月3日. 筑波大学(茨城県 つくば市)

Graciliano Gacchia Torres Galindo Jr, Chendra Hadi SURYANTO, Kazuhiro Fukui, “Gesture recognition with multiple features from hands and face using depth sensor”, FIT2014. 2014年9月3日. 筑波大学(茨城県 つくば市)

Daisuke Takabayashi, Hideitsu Hino, Kazuhiro Fukui, “Finger alphabets recognition with multi-depth images for developing their learning system”, FCV2014, 2014年2月5日. 沖縄工業高等専門学校(沖縄県 名護市)

Bernardo Gatto, Hideitsu Hino, Kazuhiro Fukui, “Block-based KOMSM for hand shape recognition with occlusion”, FCV2014, 2014年2月6日. 沖縄工業高等専門学校(沖縄県 名護市)

田中陽士, 高林大輔, 近藤真暉, 加藤伸子, 福井和広, 岡崎彰夫, “Kinectを用いた指文字練習システムにおける識別処理とユーザインタフェースの改良”, 電子情報通信学会 HCG シンポジウム, 2013年12月19日. 愛媛大学(愛媛県 松山市)

高林大輔, 田中陽士, 加藤伸子, 岡崎彰夫, 福井和広, “フィードバック機能を備えた指文字学習支援システムの開発”, ビジョン技術の実利用ワークショップ (ViEW2013), 2013年12月6日. パシフィコ横浜(神奈川県 横浜市)

6. 研究組織

(1) 研究代表者

福井 和広 (Fukui, Kazuhiro)
筑波大学・システム情報系・教授
研究者番号: 40375423

(2) 研究分担者

岡崎 彰夫 (Okazaki, Akio)
筑波技術大学・産業技術学部・教授
研究者番号: 20516679

加藤 伸子 (Kato, Nobuko)
筑波技術大学・産業技術学部・教授
研究者番号: 90279555