

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 7 日現在

機関番号：17104

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25730138

研究課題名(和文)一人称視点魚眼カメラを用いたモバイル手話認識システムの開発

研究課題名(英文)Development of Mobile Sign Language Recognition System using First Person View Fish-eye Camera

研究代表者

河野 英昭(Kawano, Hideaki)

九州工業大学・大学院工学研究院・准教授

研究者番号：00404096

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：これまで、手話認識システムの研究は数多く行われている。しかし、日常的な使用を目的とした手話認識システムの研究は少ない。本研究では日常的な使用を目的とした、ウェアラブルかつカメラベースの手話認識手法を提案し、その認識精度を検証した。認識処理においては、SIFTマッチングを用いて毎フレームの手形状を判別し、その結果をHMMによって学習・認識する。また、本システムの有効性を検証するために手話単語 22 単語の認識実験を行った。結果として77.3%の認識率を示した。

研究成果の概要(英文)：Indeed a lot of studies have been reported on the sign language recognition system, but there are few studies suitable for daily use. In this study, we proposed a sign language recognition system available as daily use. The system is wearable but no restraints in gesturing because sign languages are sensed by a camera mounted on the user's chest. In discriminating the form of hand, a state-of-art feature descriptor, i.e. SIFT, is employed. It is considered that the SIFT is suitable for wearable systems because of the low computational costs. The effectiveness of the proposed method is verified by an experiment using 22 sign language words. As a result, we obtained good results for 19 words among the 22 words.

研究分野：ソフトコンピューティング

キーワード：手話認識 一人称視点 SIFT HMM

1. 研究開始当初の背景

手話は元来、対話者に提示して理解を得るものである。これまでの手話認識の研究の多くは、手話話者と対面する位置から見た正面映像から認識する方法に焦点が当てられてきた [Ong et al., IEEE Trans. PAMI (2005)]。この撮影条件の場合、手話話者がいつでも・どこでも手軽に利用し、手話提示を介した発話に利用するにはカメラの設置が問題となる。そこで、手話話者自身がカメラを装着し、手話を装着カメラ（以後、一人称視点カメラと呼ぶ）から捉えて認識する方法が研究されてきている。

先駆的な試みに [Starner et al., IEEE Trans. PAMI, (1998)] の Hat-Mount Camera があり、[Brashear et al., Proc. ISWC, (2003)] でその改良として手首への加速度バンド付加が行われた。精度は向上したものの身体的拘束が増している。[Mayol et al., Proc. ISWC, (2005)] では、手動作のイベントタグ付けを行う課題に取り組んでいるが動作条件は極めて限定的である。国内では、[Nagasue et al., Proc. AROB, (2012)] が指文字認識を行っているが、手話のような動的変化が起こらない点、実験環境が制限されている点でまだ実用的ではない。

以上のように、動的複雑背景下で、装着ストレスなく、実用的に手話を一人称視点から認識可能なシステムの開発には至っていない。また、取り組み自体が非常に少ない。

2. 研究の目的

ろう者と聴者の円滑なコミュニケーションを実現するために、手話話者がいつでも・どこでも身につけて利用可能なモバイル手話認識システムを開発することを目的とする。開発するシステムは手話話者が身につける魚眼カメラから、手話を読み取り、その動作を認識し、認識結果を音声で再生するシステムを想定している。

具体的には、以下の3つが達成目標となる。

- (1) 動的で雑多な背景から手領域を安定的に検出する。
- (2) 一人称手話認識に有効な画像特徴量を明らかにする。
- (3) 一人称視点固有の画像特徴量の時系列パターンを認識する方法を確立し、ウェアラブルシステムに実装した一人称視点魚眼カメラによるモバイル手話認識を実現する。

3. 研究の方法

(1) 最終的な使用形態であるウェアラブル魚眼カメラ環境で取得される一人称手話映像を得る実験環境を整備し、手話データを収集し



図1 一人称視点手話認識システムで撮影される映像の例

た。データの収集は文献を参照し、研究室で独自に作成した。

(2) 手話映像に対して、画像特徴量の検討を行い、手話の認識に有効な画像特徴量を選定した。

(3) 種々の時系列パターン認識系の検討と評価を行い、本研究課題に相応しい認識系を構築した。

(4) 構築したシステムを日常手話単語の認識に適用し、評価を行った。

(2), (3) および (4) のサイクルを段階的に繰り返しながら、アルゴリズムの精緻化を図った。

(5) ウェアラブルシステムとして、実装した。

(6) 研究の進行に合わせて、手話話者に着用してもらい、ユーザビリティの評価も行った。

初年度は (1)～(2) までの一人称視点魚眼カメラからの手話データ取得を実施し、次年度以降に (3)～(5) の認識系の構成およびウェアラブルシステムへの実装を行い、モバイルシステムとして完成した。

4. 研究成果

本研究では、カメラを手話話者の胸の位置に装着し、手話動作を撮影する。撮影される映像は、手話話者側の視点から見一人称視点映像となる。撮影される映像の例を図1に示す。

本研究では、この一人称視点映像を用いた場合の手話認識精度について検証した。認識手法は、Auephanwiriyaikul らの手法 [Patten Recognition Letters, 34(11), 2013] を用いた。この手法は SIFT を用いて手話動作のマッチングを行っている。そのため計算コストが低く、ウェアラブルシステムに適している。

本研究で開発した手話認識における処理の流れは以下の通りである。

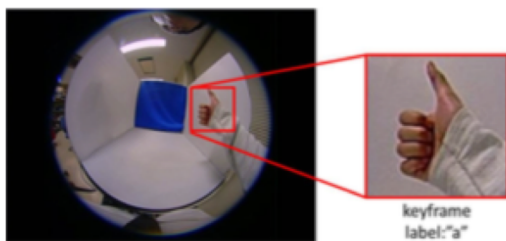


図2 keyframe の抽出例

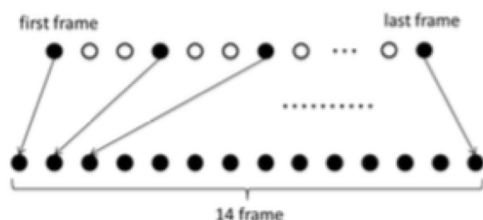


図3 手話動画の間引き処理

前処理:

(1) keyframe の登録

手話動画から特徴的な手の形の画像 (keyframe) を切り出し、テンプレートとして登録する。手話動画中の特徴的な手の形の画像をテンプレートとして登録する。このテンプレートは keyframe と呼ぶ。keyframe は図2 に示すように手領域を含む四角形で切り出す。一人称視点映像では手の大きさの変化が大きいので、keyframe の画像サイズは固定しない。その後、手の形の種類で分類し、各クラスにラベルを割り当てる。

本処理:

(2) 手話動作を撮影

カメラは手話話者の胸の位置に装着する。また、手話が行われる領域を視野に収めるために魚眼レンズを用いる。

(3) 動画を間引く

撮影した動画について、1つの動画の長さが14フレームになるようにフレームを間引く。選択する14フレームは図3に示すように、動画の最初の1フレームと最後の1フレームに加えて、等間隔に12フレームを選ぶ。

(4) SIFT マッチング

動画の各フレームについて、SIFT 特徴量を用いて全ての keyframe とのマッチングを行う。SIFT マッチングの例を図4に示す。これにより、フレーム画像と keyframe 画像間において対応するキーポイントの位置、スケール、オリエンテーションが得られる。

(5) 重なりの度合を求める

各キーポイントについて、keyframe との手のシルエットの重なりの度合を求める。重なりの度合は以下の式で求める。

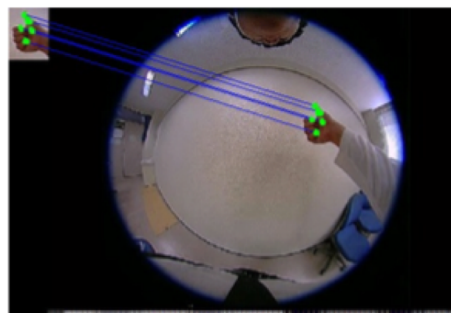


図4 SIFT マッチング処理の結果例

重なり度合 = 重なっている画素数 ÷ シルエットに該当する画素数 (1)

手のシルエットを重ねる上では、キーポイントの位置とスケールの情報を用いる。手のシルエットを重ねる手順の例を図3に示す。図中の矢印の始点はキーポイントの位置、矢印の長さはスケール、矢印の向きはオリエンテーションを表す。図3(b): フレーム画像上のキーポイントの位置と keyframe 画像のキーポイントの位置を合わせる。図3(c): 両者のスケールの比だけ keyframe 画像を拡大縮小する。図3(d): 両者のオリエンテーションが同じ方向になるように keyframe の画像を回転させる。

(6) 各フレームにラベルを割り当てる

重なりの度合に基づき、各フレームに対してラベルを割り当てる。あるフレームにおいて求めたすべての重なりの度合を総和し、その数値が最も大きい keyframe のラベルをそのフレームに割り当てる。これにより、1つの動画は長さが14のラベル時系列データになる。

(7) HMM を用いた学習・認識

得られたラベル時系列データを用いて、学習および認識を行う。学習処理では Baum-Welch 法を用いて HMM の作成を行う。認識処理では Viterbi アルゴリズムを用いて、最も尤度の高い HMM を求める。その HMM に対応した単語を認識結果とする。

先行研究で用いられている主要手話単語22語に対して検証実験を行い、77.3%の認識率を得た。22単語中19単語は良好な認識結果を示したものの、3単語については50%にも満たなかった。原因としては、類似形状の存在、手の位置・動き情報の未考慮があげられる。これらへの対策が今後の課題として位置付けられる。

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 1件)

- ① Hideaki Kawano, Kouichirou Hayashi,

Hideaki Orii, and Hiroshi Maeda, "A proposal of image color pre-conversion for elderly employing neural networks," ICIC Express Letters, Part B: Applications, 7(4), 745-752, 2016. 査読あり

[学会発表] (計 7件)

- ① Hideaki Kawano, Kouichirou Hayashi, Hideaki Orii, and Hiroshi Maeda, "Prior Image Transformation for Presbyopia Employing Serially-Cascaded Neural Network," 22nd International Conference on Neural Information Processing, Istanbul, Turkey, November 9-12, 2015. 査読あり
- ② Yuriko Tsunoda, Kouichirou Hayashi, Hideaki Kawano, Hiroshi Maeda, "Cluster Analysis with One-Dimensional Topology Self-Organizing Map," International Conference on Intelligent Agents, Web Technologies and Internet Commerce Jointly with International Conference on Computational for Modelling, Control and Automation (CIMCA2014/IAWTIC2014), pp.7-12, Online Conference, December 10-12, 2014. 査読あり
- ③ 田島裕介, 河野英昭, 前田博, "SIFT を用いた手形状判別による一人称視点手話認識," スマートインフォメディアシステム研究会, 京都府京都市, 2014年12月18~19日. 査読なし
- ④ Yusuke Tajima, Hideaki Kawano, Hideaki Orii, Hiroshi Maeda, "First-person View Sign Language Recognition Employing SIFT and HMM," International Workshop on Smart Info-Media Systems in Asia (SISA2014), pp. 30-35, Ho Chi Minh City, Vietnam, October. 8-10, 2014. 査読あり
- ⑤ 田島裕介, 古賀経邦, 柴田裕介, 河野英昭, 前田博, "魚眼レンズを用いた一人称視点手話認識におけるレンズ歪みの影響に関する検討," 第18回日本知能情報ファジィ学会中国・四国支部大会・第15回日本知能情報ファジィ学会九州支部大会学術講演会合同大会, pp.47-50, 山口県下関市, 2013年12月21日. 査読なし
- ⑥ 柴田裕介, 折居英章, 河野英昭, 前田博, "一人称視点手話認識のための手のフレームアウトを考慮した手領域判別," ファジィシステムシンポジウム講演論文集

29, pp. 345-348, 大阪府枚方市, 2013年9月9~11日. 査読なし

- ⑦ 田島裕介, 折居英章, 河野英昭, 前田博, "手の動きに伴う面積変化を用いた一人称視点手話認識," ファジィシステムシンポジウム講演論文集 29, 501-504, 大阪府枚方市, 2013年9月9~11日. 査読なし

6. 研究組織

(1) 研究代表者

河野 英昭 (KAWANO, Hideaki)

九州工業大学・大学院工学研究院・准教授

研究者番号：00404096