

令和 4 年 6 月 9 日現在

機関番号：37401

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K12032

研究課題名（和文）手話映像の時間的要約方式に関する研究

研究課題名（英文）Temporal Summarization of Sign Language Videos

研究代表者

筒口 拳（Tsutsuguchi, Ken）

崇城大学・情報学部・教授

研究者番号：70828227

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究は手話の効率的な映像コミュニケーション実現をめざし、手話の実写映像を、映像から抽出した少数の特徴的な画像（キーフレーム）のみを用いて要約するものである。本研究では、キーフレームから作成した要約動画の効果を確認したうえで、キーフレームでは手指動作が（動画映像中において）停留する、という仮説のもと、動画内の手話者の動きを解析することにより、課題であったキーフレームの自動抽出を可能とする新しい手法を提案し、実現した。この手法は手話者に依存せず、実時間に近い時間でキーフレーム候補を抽出できるという大きな特徴がある。

研究成果の学術的意義や社会的意義

手話映像を特徴的な少数の画像により要約するという手法は著者が提案した手法であり、学術的新規性が高い。本研究により、手話映像から自動的にキーフレーム候補を抽出することが可能となり、当初計画で想定していた特定手話者のみならず、不特定の手話者に対しても抽出することが可能となり、またリアルタイムでの抽出の可能性も示唆されているため、その学術的な意義は極めて高い。本成果を応用することで、災害時に情報を素早く伝達することが可能となり、また、手話映像を再確認する際にも、より短い時間で内容を確認することが可能となるなど、ろう者の利便性をはかることができ、その社会的な意義は非常に大きい。

研究成果の概要（英文）：This study aims to summarize sign language videos using only a small number of key images (we call as "keyframes") extracted from the video, to realize efficient video communication of sign language.

In this study, we first confirmed the effect of the summarized video created from the keyframes. Then, based on the hypothesis that the movement of hand or arm is stopped or moved slowly (in the video) at the key frame, we analyzed the movement of sign language video, and proposed and realized a new method that enables automatic extraction of keyframes. This method has two major features that it does not depend on the sign language person, and can extract keyframe candidates in near real time.

研究分野：ビジュアル・インタフェース

キーワード：手話映像 要約 キーフレーム 自動抽出 オプティカルフロー

## 1. 研究開始当初の背景

自然災害等が発生した場合、被害を最小限に抑えるためには避難を促すような緊急情報を素早く、かつ、内容を損なわないような手段で伝達する必要がある。手話を母語とする聴覚障がい者の方々の中には、文字を中心とした情報伝達やコミュニケーションを得意としない人たちが一定数存在する。情報通信技術の進歩により、手話映像を用いたコミュニケーションが可能な環境が整っているが、災害時には通信帯域が圧迫されることが多いため、これらの方々を支援するには、できるだけ手話映像のデータ容量を小さくして伝達することが重要である。

また、手話は一連の動きを見なければ手話の内容を理解することが難しいため、例えば手話によるニュース映像を再確認する際にも、実時間を要してしまう。内容を損なうことなく映像を要約することができれば、手話映像を短時間で確認することができ、聴覚障がい者の方々にとって有用となり得る。

我々は、手話映像を構成する時間的に連続する画像のうち、手話の特徴を強く反映している画像を「手話キーフレーム」と定義し、手話キーフレームのみを伝達することによる効率的な手話映像伝達の実現をめざす映像伝達方式を提案してきた。図1は手話映像とキーフレームの概念を示す。図1において、もとの手話映像の単語の開始点及び終了点をキーフレームとして抽出し、この2枚のみを提示することで単語の意味を損ねず、送信する画像数を劇的に削減するというものである。我々のこれまでの検討で、手話キーフレームのみで構成された要約映像であっても、通常の手話映像と同等の内容を伝達できることが示唆されており、映像コミュニケーションにおけるデータ容量の削減に大きな効果があることが期待できる。

しかし、これまでの検討では手話キーフレームは人手で抽出しており、キーフレームによる要約映像の生成に手間がかかっていたため、撮影した手話映像からどのようにキーフレームを自動抽出するかが大きな課題である。

本研究ではキーフレームの自動抽出方法の検討と評価実験を行い、有効性の確認と手話映像要約システムの実装をめざすものである。本研究により、特に聴覚障がい者をはじめとする情報弱者に対して、手話映像による次に挙げるような支援ができると期待できる：

- 災害時の緊急連絡、被災時の連絡手段
- 手話ニュース映像などの短時間での確認
- 通常の（スマートフォンなどによる）コミュニケーション

本研究のきっかけは上述のとおり災害時の緊急連絡の通報であるが、会話の内容を損ねることなく低容量で映像通信ができるという点は、通常のスマートフォンを用いたコミュニケーションに適用することもでき、当初の目的を超えて汎用的に活用できるものと考えている。

また、本研究の成果は手話の種類（地域性や方言など）や言語（日本語、英語ほか）に依存しないところも大きな特徴であると考えられる。

さらに、手話以外の対象、例えばスポーツ映像の中からの重要なポーズを抽出し、指導者が選手にポイントをわかりやすく説明する、といった用途にも応用できる可能性がある。



図1: 手話映像とキーフレーム。上: 手話映像, 下: キーフレーム。2枚のキーフレーム間の画像は送信しない。

## 2. 研究の目的

上述のキーフレーム自動抽出含め、我々の目標を実現するための課題として、

- (1) キーフレームだけで構成される映像は情報を十分に伝えうるか
  - (2) キーフレームを自動で抽出するにはどうすればよいか
  - (3) キーフレーム抽出をできるだけリアルタイムで行うにはどうすればよいか
- が挙げられる。これらを解決することが我々の研究の目的である。

本研究においては、(2)の自動抽出の実現性を明らかにすることを主目的としている。(2)の有効性を確認したうえで、(1)および(3)の可能性についても検討を行う。

なお、我々は研究を進める上で、「キーフレームでは、映像上で手指動作が(見かけ上)停留する」という仮説をたて、自動抽出にあたっての方針とした。キーフレームから作成された要約映像の有効性により、この仮説も検証できると考えている。

本研究期間終了後も含めた最終目標は、不特定多数の話者が演ずる不特定内容の手話蓄積映像あるいは手話通信映像から、実時間(に近い処理速度)でキーフレームを抽出するシステムの構築にある。

## 3. 研究の方法

研究対象を「特定話者の手話映像からのキーフレーム自動抽出」とし、以下の手順で進めていった。

### (1) 研究対象とする手話

日本国内で使われている手話には、日本手話、日本語対应手話、及び中間型手話の3種類が存在する。本研究では、対象とするユーザの広さという観点から、中間型手話を対象とする。

### (2) 研究環境の構築、および手話映像データの取得

映像解析処理用のパーソナルコンピュータ(GPU)を調達し、研究対象データとして、同一の手話者(ろう者)による手話映像データを取得した。

### (3) 手話者の姿勢推定および解析によるキーフレーム抽出

手話映像内の手話者の骨格モデルを抽出して姿勢推定を行い、腕部の動き(軌跡)を時系列データとして抽出する。キーフレームでは動きが見かけ上停留する、という仮説に基づき、時系列データを縦方向、横方向の2軸に分けて時系列データを曲線近似し、近似曲線が極値をとるような時間的位置をキーフレーム候補とする。

### (4) オプティカルフローを用いた動作解析によるキーフレーム抽出

手話映像に対し、オプティカルフローによる動作解析を行い、キーフレームでは動きが見かけ上停留する、という仮説に基づき、ある条件を満たすフローベクトル数の時系列変化が極小値をとる時間的位置をキーフレーム候補とする。本手法は手話者に依存せず、実時間に近い処理速度で実行可能であるという大きな特徴がある。

### (5) キーフレーム要約映像生成手法の検討

抽出したキーフレームから要約映像を作成する手法の検討と生成映像に対する評価実験を実施する。

### (6) その他の研究

ろう者の手話読取に関する研究や、キーフレームの手話学習システムへの適用について検討する。

## 4. 研究成果

### (1) キーフレーム映像の効果の確認

本研究の第1の成果は、当初計画には含まれてないが、キーフレーム映像の有効性を確認した評価実験の効果をとりとまとめたことである。

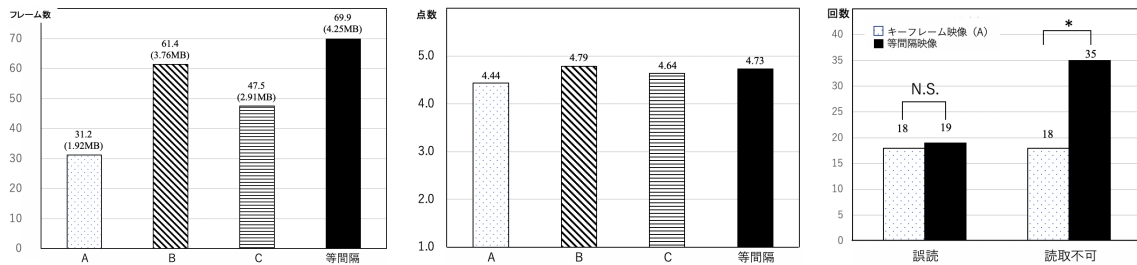


図 2: 左: 各方式映像のフレーム数平均とデータ容量, 中: 読み取りやすさの評価点, 右: 比較実験での誤答数

本研究では、手話の実写映像から手作業でキーフレームを抽出し、いくつかのパターンの要約映像を作成して、どのようなパターンが適切であると考えられるかを検討した。抽出においては、手話表現の空間的特徴が強く出ている単語の始点・終点および、その途中の点（単語内の途中点、単語間の途中点など）を抽出して作成した多くのパターンのキーフレーム映像の中から、ろう者および手話通訳者へのエキスパートレビューを介して3種類の抽出パターンと、比較のためのパターンを選定した：

- A：単語の始点・終点をキーフレームとする
- B：Aに単語内及び単語間の画像を1枚加える
- C：Aのキーフレーム間隔に応じて画像を追加する

等間隔：映像から一定間隔で画像を抽出する

要約映像は、「キーフレーム1から次のキーフレーム2まで、キーフレーム1を表示し続ける」こととし、映像長を原映像と揃えることとした。15種類の手話映像それぞれについて4パターンの刺激映像を作成し、20名のろう者に対して読み取り実験を行いキーフレーム映像の有効性を評価した。

図2左は各実験用刺激映像の総フレーム数を平均したものであり、図2中は映像の読み取りやすさを6件法で評価した結果である。この結果、映像パターンAはデータ量が最も少ないにもかかわらず、読み取りやすさにおいて他の映像と有意差がないことがわかった。

次に、最もデータ圧縮率が高いパターンAと、Aと同程度のデータ量にした等間隔映像とで、読み取り内容の誤りについて比較実験を行った（図2右）。その結果、誤読数では差がなかったが、読取不可数についてはキーフレーム映像の方が少ない（5%水準で有意）であることが判明した。

以上により、キーフレーム映像は等間隔映像よりメッセージの読み取りやすさにおいて優れていることが明らかとなった。

## (2) 姿勢推定によるキーフレーム候補抽出

本成果は、手話映像中の被写体（手話者）の姿勢を推定して腕の関節部位の軌跡（位置の時間変化）を取得し、その軌跡を曲線近似して極値をとる時間的位置をキーフレーム候補とする手法である。図(a)に示すように人物像の骨格および関節位置を抽出し、その時間変化をとって曲線近似を行った上で、図(b)に示すように軌跡の「極値」をとるフレームを求めることでキーフレーム候補抽出が可能であることを確認した。本手法は姿勢推定の段階で多くのノイズを含むため、極値が過剰に検出されてしまう。

そのため、両肘および両手首の4つの関節についてX方向・Y方向それぞれの近似曲線を取り、8種類の曲線から得られたキーフレーム候補から時間的に近い候補のみを残す方法（AND法）を取ることにより、キーフレーム候補の絞り込みを行った。その結果、過剰に検出された極値数を削減することができたが、それでも不要なキーフレーム候補が多く検出されてしまう結果となった。

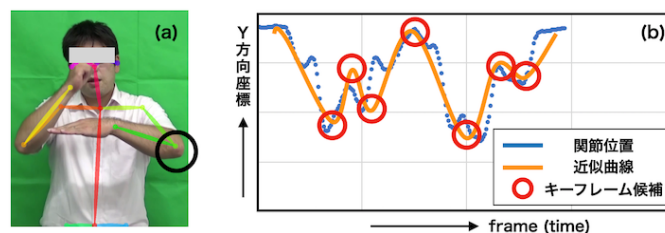


図 3: 関節位置解析によるキーフレーム抽出。(a) 骨格構造を表示した画像例, (b) 関節位置の時間変化の例。

### (3) オプティカルフロー解析を用いたキーフレーム候補抽出

本成果は、図 4(a) に示すように手話映像内の動きをオプティカルフローを用いて解析し、一定の条件を満たすフローベクトル数をフレーム毎にカウントしたうえで、図 4(b) に示すようにフローベクトル数の時間変化が極小値をとるフレームをキーフレーム候補とするものである。

検討すべきパラメータは、カウントするフローベクトルの長さの閾値、および、カウントしたフローベクトル数の閾値である。

手話単語の映像に対し複数のパラメータの組み合わせの中から最適なものを選定し、それを他の手話単語映像および文章映像に適用したところ、単語において再現率 100%・適合率 62%、文章において再現率 78%・適合率 57% となった。再現率は「正検出数 / (正検出数 + 検出漏れ)」であり、適合率は「正検出数 / (正検出数 + 誤検出数)」であるから、本手法は「検出漏れが比較的少ないが、キーフレームではないものも多く検出される」という結果になる。キーフレーム映像においてはキーフレームでないものが検出されてもデータ量が増えて表現が冗長になる程度の影響であるが、キーフレームが欠落すると手話の内容が伝達できなくなるため、今回実験対象とした単語の手話映像に関しては良好な結果であったと言える。

本手法は実装が容易であり、キーフレーム候補を実時間に近い処理時間で自動検出することができるという大きな特徴があり、本研究の大きな成果である。

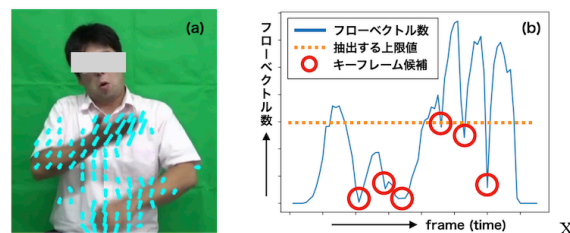


図 4: オプティカルフロー解析によるキーフレーム抽出. (a) フローベクトルを表示した画像例, (b) フローベクトル数の時間変化の例.

### (4) その他の研究成果

オプティカルフロー解析を用いて自動抽出したキーフレーム候補をもとにした 2 種類の要約映像 (キーフレーム間を一定のフレーム数で補間するもの: 図 5 (b) と、原映像におけるキーフレーム間隔に比例したフレーム数で補間するもの: 図 5 (c)) の自動生成方法を検討・実装し、手話通訳士によるエキスパートレビューまでを実施した。その他、要約映像の注視領域を視線検出装置を用いて解析したり、ろう者が手話映像を閲覧する際、手話文章をどのように理解し記憶するのか、について実験的な検証を行ったり、キーフレームの手話学習者向け学習ツールへの適用を試みるなど、本研究の応用事例などについても、前倒しで検討した。

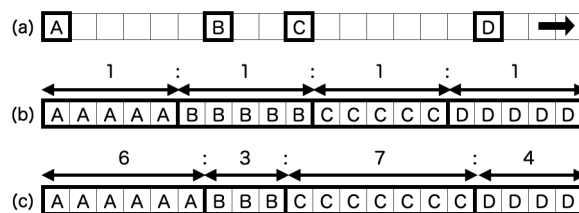


図 5: キーフレーム映像例. (a) 原映像: A/B/C/D がキーフレーム, (b) Constant rate, (c) Proportional rate.

(以上)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 筒口 拳, 秋山 滉太, 品田 紗弥花, 米村 俊一	4. 巻 50
2. 論文標題 手話の空間的特徴に基づくキーフレームを用いた手話映像要約の検討	5. 発行年 2021年
3. 雑誌名 画像電子学会誌	6. 最初と最後の頁 373-382
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計11件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 猪岡 翔, 筒口 拳, 米村 俊一
2. 発表標題 手話学習支援システム -手話キーフレームによる学習ポイントの明示機能の提案-
3. 学会等名 子情報通信学会 福祉情報工学研究会
4. 発表年 2020年

1. 発表者名 丸亀 泰作, 米村 俊一, 筒口 拳
2. 発表標題 姿勢推定を用いた手話映像からのキーフレーム候補抽出
3. 学会等名 情報処理学会 火の国情報シンポジウム2021
4. 発表年 2021年

1. 発表者名 呉 夢竹, 米村 俊一, 筒口 拳
2. 発表標題 オプティカルフローを用いた手話映像からのキーフレーム候補抽出
3. 学会等名 情報処理学会 火の国情報シンポジウム2021
4. 発表年 2021年

1. 発表者名 入江 健太, 筒口 拳
2. 発表標題 手話の動作近似に関する検討
3. 学会等名 映像情報メディア学会研究会 (SIP, IE)
4. 発表年 2019年

1. 発表者名 品田 紗弥花, 筒口 拳, 米村俊一
2. 発表標題 ろう者の手話緊急メッセージ読み取りにおける文章理解の方略
3. 学会等名 映像情報メディア学会研究会 (SIP, IE)
4. 発表年 2019年

1. 発表者名 猪岡 翔, 筒口 拳, 米村 俊一
2. 発表標題 視線計測を用いた手話初学者の学習方略の分析
3. 学会等名 映像情報メディア学会研究会 (SIP, IE)
4. 発表年 2019年

1. 発表者名 品田 紗弥花, 筒口 拳, 米村 俊一
2. 発表標題 ろう者の手話映像読み取り方略を考慮した手話文章構成形式の検討
3. 学会等名 電子情報通信学会 福祉情報工学研究会
4. 発表年 2019年

1. 発表者名 入江 健太, 米村 俊一, 筒口 拳
2. 発表標題 関節軌道の多項式近似に基づく手話映像からのキーフレーム抽出
3. 学会等名 画像電子学会第292回研究会
4. 発表年 2020年

1. 発表者名 猪岡 翔, 筒口 拳, 米村 俊一
2. 発表標題 手話学習支援システム -手話キーフレームを用いた同期映像フィードバック機能の提案-
3. 学会等名 子情報通信学会 福祉情報工学研究会
4. 発表年 2021年

1. 発表者名 板井 裕太, 西村 洋輝, 米村 俊一, 筒口 拳
2. 発表標題 手話キーフレーム映像の構成に関する検討
3. 学会等名 情報処理学会 火の国情報シンポジウム2022
4. 発表年 2022年

1. 発表者名 松丸 和暉, 筒口 拳, 米村 俊一
2. 発表標題 非手指動作の学習を促進する手話学習支援システムの検討
3. 学会等名 電子情報通信学会 ヒューマンコミュニケーション基礎研究会
4. 発表年 2022年



〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	米村 俊一  (Yonemura Shunichi)  (60631033)	芝浦工業大学・工学部・教授    (32619)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------