手話キーフレーム映像の構成に関する検討

板井 祐太[1]　西村 洋輝[1]　米村 俊一[2]　筒口 拳[1]

概要：手話映像の効率的な伝達のため，手話映像の中から最も手話の特徴を強く表している画像（キーフレーム）を抽出し，キーフレームだけで映像を再構成する手法を検討している．本研究では抽出されたキーフレームから要約映像を生成する 2 つの手法を提案する．2 つの手法をもとに生成された映像長の異なる複数の手話要約映像に対し，手話内容の解釈および要約映像に対する注視領域について検討を行い，エキスパートによるレビューを行った結果を報告する．

キーワード：手話，映像要約，キーフレーム，構成，視線計測

A Study of the Structure of Sign Language Keyframe Video

Yuta Itai[1]　Hiroki Nishimura[1]　Shunichi Yonemura[2]　Ken Tsutsuguchi[1]

Abstract: For efficient transmission of sign language videos, we are studying a method for reconstructing video using only key frames, that strongly represent the characteristics of sign language. In this research, we propose two methods to generate abstracted video from the extracted key frames. For multiple sign language abstracted videos with different video lengths generated based on the two methods, the interpretation of the sign language content and the gaze area for the video are examined, and the results of a review by an expert are reported.

Keywords: Sign language, video abstraction, keyframe, composition, eye tracking.

1. はじめに

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，手話映像を確認する際には時間がかかってしまう．短時間で確認するには映像を一定間隔でスキップする方法（早送り）が考えられるが，重要なシーンが欠落し，内容が伝達できない可能性がある．

この問題を解決するための手段として，手話の特徴が強く表れている画像（以下，キーフレームとよぶ）のみを用いて映像を再構成すれば，手話内容を保持したまま映像を要約することが可能となる．キーフレームのみで構成された要約映像（キーフレーム映像）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先行研究により示唆されており [1][2][3]，確認時間の短縮やデータ削減に大きな効果があることが期待できる．**図 1** にキーフレームの概念を示す．図において手話映像（左側）から重要なフレームを抜き出し（右側），この重要なフレームのみからキーフレーム映像を構成することになる．

キーフレーム映像による伝達は，データ量以外の面でも，実写映像に基づくため顔の表情などが保持されている点や，手話の種類（地域性や方言など）や言語（日本語，英語など）に依存しない可能性がある点が大きな特徴である．

キーフレーム映像の作成にあたり，(1) 手話映像からど

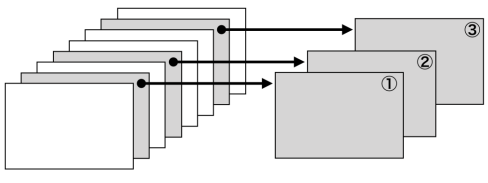

図 1　キーフレームの概念．
Fig. 1　The concept of keyframe.

[1] 崇城大学　Sojo University
[2] 芝浦工業大学　Shibaura Institute of Technology

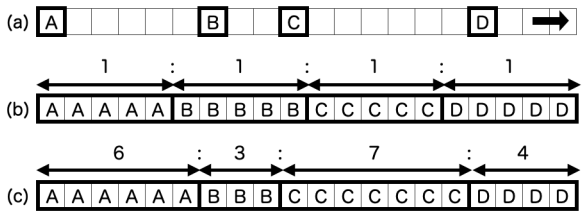

図 2 キーフレーム映像；(a) 元映像：A/B/C/D がキーフレーム；(b) Constant rate；(c) Proportional rate.

Fig. 2 The keyframe video; (a)original video (A/B/C/D are keyframes), (b) Constant rate, (c) Proportional rate.

のようにキーフレームを自動抽出するか，(2) キーフレームから映像をどのように再構成するか，という課題がある．

(1) については，キーフレームでは手指の運動が停留するという品田らの仮説 [4][5] に基づき，手指の関節位置が停留する時点をキーフレーム候補とする手法 [6] や，オプティカルフローを用いて運動の変化を検出し，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手法が提案されている [7].

本研究は (2) のキーフレームから映像をどのように再構成するか，を提案するものである．

本研究では，2 種類のキーフレーム映像の構成方法を提案する．呉らの手法 [7] を用いて手話映像から抽出したキーフレームに対し，2 つの構成方法をもとに映像長の異なる複数種類のキーフレーム映像を作成し，以下の 3 点について手話通訳士へのエキスパートレビューを行った：

(1) **映像内容の保持**（意味が伝わるか）．キーフレーム映像を閲覧しても，情報が正確に伝わるかどうか．
(2) **映像の見やすさ**（違和感のなさ）．作成された要約映像を，特に違和感なく閲覧できるかどうか．
(3) **映像の注視領域**．本研究ではキーフレームを手指動作に基づいて抽出しており，表情などの非手指動作については考慮されていない．キーフレーム映像のどのあたりを注視するかにより，キーフレーム抽出手法への指針となるのではないか．

その結果，キーフレーム映像はどの方式・種類でも「見づらい」という結果となり，手話の内容は平均すると約 2/3 が伝達されていた．また，通常映像・キーフレーム映像ともに，顔（口元）周辺に中心視線が集まることがわかった．

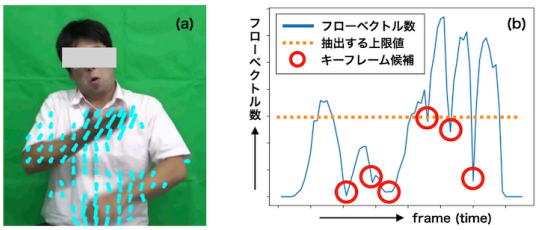

図 3 キーフレーム抽出；(a) フロー検出，(b) 候補抽出．

Fig. 3 Keyframe extraction; (a) flow vector, (b) candidates.

表 1 実験で用いた手話の内容．

Table 1 Sign language setences.

これから/やってみたい/ことは/ありますか
お兄さんには/子どもが/いるのですか
明日は/5 時に/仕事を/終えて/友達に/会います
大きい/湖の/中に/小さな/島が/あります
高いところから/見るので/花火が/とても近く/見えました
私の/趣味は/テレビで/野球を/見ることです
私は/パソコンが/得意です
私は/花を/見るのが/好きです
職場は/どこに/ありますか
仕事は/何時に/終わりますか

以下，2 章でキーフレーム映像の構成について述べ，3 章で実験（エキスパートレビュー）とその結果を報告する．4 章で考察を加え，5 章でまとめる．

2. キーフレーム映像

キーフレームから要約映像を再構成するにあたり，あるキーフレーム A から次のキーフレーム B までの間は，前のキーフレーム A を再生し続ける．その際の構成方式として，キーフレーム間を固定フレーム数だけ繰り返す方式（Constant Rate）と，キーフレーム間のフレーム数の差分の比を保持する方式（Proportional Rate）を提案する．

図 2 (a) に示す映像の各フレームのうち，A，B，C，D がキーフレームだとすると，Constant Rate（以下，CR ）はこれらキーフレームの時間的配置に関係なく，各キーフレームを同じフレーム数だけ継続して映像にする（図 2 (b)）．一方，Proportional Rate（以下，PR）は各キーフレーム間の時間間隔の比率を出し，その比率を保持したフレーム数だけキーフレームを継続するものである（図 2 (c)）．

いずれの方式も，キーフレーム間のフレーム数により，映像長の異なる様々な要約映像を作成することができる．フレーム総数が少ない方が再生時間は短くなるが，早送りのようになってしまい内容確認が難しくなる可能性がある．そこで本研究では，固定するフレーム数を 4 フレーム，5 フレーム，6 フレーム，7 フレーム，8 フレームにした CR 映像 5 種類と，キーフレーム間の比率を元映像の 0.2 倍，0.3 倍，0.4 倍，0.5 倍，0.6 倍にした PR 映像 5 種類の計 10 種類の映像を作成し，どの方式のどの映像長が聴覚障がい者にとって適切であるかを評価することとした．

呉らの手法 [7] に基づき，オプティカルフローを用いてキーフレームを抽出し（**図 3**），その後，キーフレームだけを用いて映像を再構成した．要約対象として，日本手話と日本語対応手話の両者が混在する中間型の手話映像を用いた．作成した文を**表 1** に示す．

なお，現在のキーフレーム抽出の課題として，手指動作からキーフレームを抽出しているため，表情の変化や頷きなどの非手指動作がキーフレームとして抽出されない可

表 2　内容の正答率の例（75%）.
Table 2　Correct answer rate (75 %).

（本文）	私は	花を	見るのが	好きです
（翻訳）	僕は	この花が		好き
（正誤）	○	○	×	○

能性があり，内容が正確に伝わらないおそれがある．このため，生成したキーフレーム映像について，実際に表情のキーフレームを必要とするのかどうかについても，映像の注視領域を検出することによって明らかにする．

3. 実験（エキスパートレビュー）

どのようなキーフレーム映像が適切であるかを評価するため，手話通訳士（60 代女性・手話通訳歴 30 年）1 名に対し 1 章で述べた 3 つの実験を行った．刺激映像として，前章で述べた CR 映像 5 種類，PR 映像 5 種類の計 10 種類を用意した．被験者はキーフレーム映像を閲覧後に日本語に翻訳し，次にキーフレーム映像の「見やすさ」を評価する．その後，改めてキーフレーム映像を閲覧して視線計測を行った．

3.1 キーフレーム映像の翻訳

映像を 1 つ閲覧するごとに日本語への翻訳を行う．10 種類の映像はランダムな順序で閲覧することとし，各映像は 3 回ずつ閲覧するものとした．キーフレーム映像が表す内容と翻訳した内容を文節ごとに比較し，正誤をとる．**表 2** に正答率の計算例を示す．この例では 4 個の文節に対し 3 個の文節を翻訳できているため，75%の正答率としている．

図 4 に各キーフレーム映像に対する正答率を示す．グラフの “CR4” はキーフレーム間の固定フレーム数を 4 フレームにした映像，“CR5” は 5 フレームにした映像，等である．また，“PR2” はキーフレーム間のフレーム数の比率を保持しつつもとの映像長の 0.2 倍にしたもの，“PR3” は 0.3 倍にしたもの，等である．

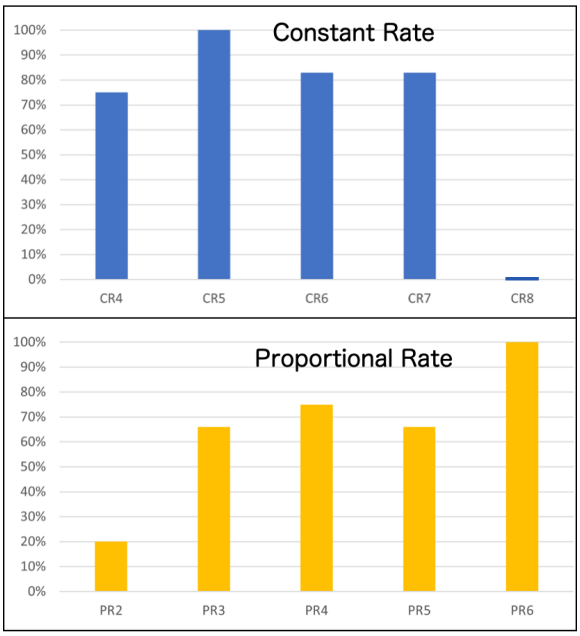

図 4　キーフレーム映像翻訳の正答率．上：CR，下：PR.
Fig. 4　The score of interpretation of keyframe videos.

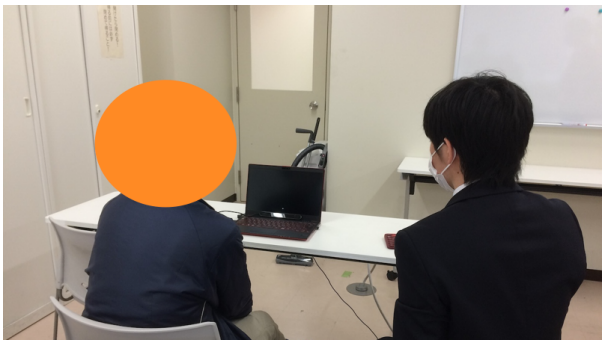

図 5　視線計測の実施状況.
Fig. 5　The situation of eyetracking experiment.

CR 映像では，キーフレーム間のフレーム数が 8 のときは翻訳が出来なかったが，それ以外では 70%以上の正答率となった．PR 映像では全体として高い数値ではなかったが，映像長が元の映像の映像長に近づくにつれて翻訳の正答率が高くなる傾向があった．

3.2 映像の見やすさ

前節の映像翻訳のあと，映像 1 種類ごとに見やすいかどうかについて点数で評価してもらった．ここで「見やすさ」は映像が表す手話が理解できるかどうか，という内容にかかわるものではなく，視覚的に違和感がないか，というものである．評価は「見づらい（1 点）」から「見やすい（6 点）」までの 6 段階で評価する 6 件法を用いた．

その結果，10 種類の映像すべて 2 点に評価されており，全体として見づらいという結果になった．

実験後に頂いた意見では，被験者はキーフレーム映像に違和感を感じるということであったが，映像によっては翻訳の正答率が高いという結果になった．

結局，今回の実験は被験者が一人ということもあるが，キーフレーム映像の適切な再生方法が定まらなかった．

3.3 視線計測

次に，通常映像やキーフレーム映像を閲覧してもらい，視線検出を行った．被験者には，映像の意味を読み取ってもらうよう指示を行った．実験の状況を**図 5** に示す．

一般に手話者どうしでコミュニケーションを行う際は，相手の口元を注視し，周辺視野で手の動きを捉えていると言われている．キーフレーム映像は，もとの手話映像を要約することから情報が少なくなっている．本研究で用いたキーフレーム抽出は手指の動作をもとに行っており [7]，顔の表情や頷きなどの非手指動作については重要なフレーム

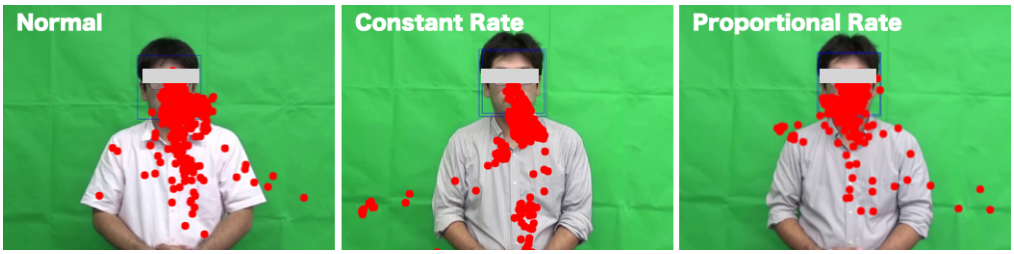

図 6 視線計測結果（手話通訳士）；左：通常映像，中：CR，右：PR.
Fig. 6 Eye tracking of interpreter; from left to right: normal video, CR, PR.

が含まれていない可能性がある．手話者がキーフレーム映像を閲覧する際，情報を補うために手指を注視するようであれば，キーフレーム映像は手指動作をもとにする手法でよいと考えられ，通常と同様に口元を注視するようであれば，キーフレーム映像に対し非手指動作をもとにするキーフレームを付加した方がよい，と考えられる．

視線検出は Tobii 社の Tobii pro ナノを使用した [8]．サンプリングレート 60Hz で視線を計測することができる．被験者はキャリブレーションを行ったあと，Tobii pro ナノを設置した PC にてキーフレーム映像の翻訳と同じ刺激映像を閲覧した．

図 6 に通常の映像，CR 映像，PR 映像に対し，手話通訳士の視線位置をプロットしたものを示す．手話通訳士は通常映像でもキーフレーム映像でも，ほとんどの映像において手話者の口元に視線が集まっていた．キーフレーム映像に対しては初見ということもあるが，要約映像においても非手指動作の重要性が示唆されていると言える．

4. 考察

今回，1 名のエキスパートレビューにとどまったものの，有用な知見を得ることができた．

手話通訳士の意見として，日本手話・日本語対応手話・それらの中間型といった種類による違いや，地域性の違いがあるとのことで，こういった要因も考慮する必要がある．また，そういった違いに加え，キーフレーム映像において手指のつながりが分かりにくくなる部分がある，という意見も頂戴した．

キーフレーム映像の内容保持については，今回の実験では平均すると約 67%の正答率となったが，今後，被験者及び実験数を増やし，精度の高いデータを得る必要がある．

また，手話熟練者は意識せずに手話者の口元を注視し，手指の動きは周辺視で把握しており，それはキーフレーム映像でも同様であることが示唆された．このことから，手指動作のキーフレームに非手指動作に基づく情報を付加することが有効になるのではないかと考えられる．

5. おわりに

本研究では，2 種類のキーフレーム映像の構成方法を提案し，2 つの方法それぞれに対し映像長の異なる複数種類のキーフレーム映像を作成し，映像内容の保持，映像の見やすさ，映像の注視領域の 3 点について手話通訳士へのエキスパートレビューを行った．

その結果，被験者が 1 名ではあるが，キーフレーム映像は「見づらい」という結果となり，手話の内容は平均すると約 2/3 が伝達されていた．また，キーフレーム映像においても，通常映像と同様に手話者の口元周辺に中心視線が集まることがわかった．

今後は被験者数を増やし，より精度の高いデータを取得したうえで解析を行うことが必要である．また，従来の手指動作に基づくキーフレーム抽出方式に加え，非手指動作に基づくキーフレーム抽出方式も検討を進めていく．

謝辞 貴重な意見を賜った手話通訳士の方に感謝する．また，本研究は科研費（基盤研究（C））19K12032 「手話映像の時間的要約方式に関する研究」の助成による．

参考文献

[1] 筒口拳，秋山滉太，品田紗弥花，米村俊一：“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”，画像電子学会誌，**50** (3)，pp.373–382 （2021）.
[2] 秋山滉太，筒口拳，米村俊一：“手話の空間的特徴に基づく映像圧縮を用いた災害情報伝達システム～キーフレーム映像による手話伝達の了解度の検証～”，ヒューマンインタフェースシンポジウム 2016 (Sep. 2016).
[3] 秋山滉太，筒口拳，米村俊一：“手話の空間的特徴に基づく映像圧縮を用いた災害情報伝達システム～無圧縮映像における手話の了解度についての考察～”，第 87 回福祉情報工学研究会 (Dec. 2016).
[4] 品田紗弥花，筒口拳，米村俊一：“カラー手袋を用いた手話の空間的特徴抽出方法に関する基礎検討”，ヒューマンコミュニケーション基礎研究会 (May 2017).
[5] 品田紗弥花，筒口拳，米村俊一：“ローパスフィルタを用いた圧縮率向上のための極値フレーム削減条件の検討”，ヒューマンインタフェース学会研究会，(Dec. 2017).
[6] 丸亀泰作，米村俊一，筒口拳：“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”，火の国情報シンポジウム 2021（Feb. 2021）.
[7] 呉夢竹，米村俊一，筒口拳：“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”，火の国情報シンポジウム 2021（Feb. 2021）.
[8] Tobii: “Tobii Pro/SDK”, `https://developer.tobiipro.com/` .