キーフレームを用いた手話映像要約における要約方式および再生速度の評価

髙濵彩香[1]　米村俊一[2]　筒口拳[1]

概要：我々は手話映像の時間的要約を目的として，手話実写映像から手話の特徴を強く表しているフレーム（キーフレーム）を抽出し，キーフレームのみから要約映像を再構成する手法を提案している．本研究では，聴覚障がい者の方を対象として，(1) オプティカルフロー解析に基いて自動抽出したキーフレームから再構成した要約映像と，手動により抽出されたキーフレームからの要約映像との内容理解に関する比較実験，および，(2) 複数の要約方式や再生速度による内容理解の差異を評価する被験者実験を行ったのでその結果を報告する．

キーワード：手話，キーフレーム，映像要約

Evaluation of Sign Language Keyframe Videos with Summarization Method and Playback Speed

AYAKA TAKAHAMA[†1]　SHUNICHI YONEMURA[†2]　KEN TSUTSUGUCHI[†3]

Abstract: For the purpose of temporal summarization of sign language videos, we are proposing a method that extracts frames that strongly represent sign language features (we call these frames as "keyframes") from live sign language videos and reconstructs summarized videos from only the keyframes. In this paper, we report the subject experiment results, conducted with hearing-impaired people that evaluate of the difference in content understanding; (1) summarized videos reconstructed from automatically extracted keyframes based on optical flow analysis vs. summarized videos from manually extracted keyframes, and (2) the differences due to multiple summarization methods and playback speeds.

Keywords: Sign language, Keyframe, Video summarization

1. はじめに

聴覚障がい者は日常的なコミュニケーションに手話を用いており，情報収集を行う際にもニュース映像など手話映像を用いることが多い．しかしながら，災害時のように通信環境が悪い場合には映像の閲覧は困難となり，手話映像における重要な箇所が欠落してしまった際には正確な内容の把握が難しくなることが考えられる．また，手話映像を高速で再生する際にも，単なる高倍速化では手話における重要な箇所がスキップされてしまうことにより元映像における手話の内容を正確に把握できなくなる可能性がある．

これらの課題への対策として，手話の特徴を強く表す重要な箇所（以下，キーフレームと呼ぶ）を正確に抽出し再構成することで，手話映像の重要箇所が保持された要約映像を生成することが考えられる[1]．コンピュータグラフィックスによる手話アニメーションのように，アーカイブや再生が容易な手法[2]も存在するが，制作コストや手話の地方性の面からも，本研究では実写手話映像を元に要約映像（以下，手話キーフレーム映像または要約映像と呼ぶ）を作成する手法を検討するものである．

手話キーフレーム映像の作成にあたり，以下の課題が挙げられる：

1. 原映像からキーフレームを自動抽出すること，
2. 手指の動きのみならず，非手指動作からもキーフレームを抽出すること，
3. キーフレーム映像の再生方式を検討すること．

先行研究においてキーフレームを自動抽出する手法[3]，非手指動作のうち「頷き」を抽出する手法[4]，キーフレーム映像の再生方式[5]が提案されているが，本研究では，これらの手法をもとに生成された手話キーフレーム映像を用いて，実際の聴覚障がい者の方を対象とする内容理解に関する評価実験を 2 度にわたり行ったものである．

実験の結果，自動抽出されたキーフレームによる手話キーフレーム映像と手動抽出による手話キーフレーム映像は理解度に差がなく，また，高倍速の映像では，手話キーフレーム映像の優位さが示唆される結果となった．以下，第 2 章でキーフレームおよび関連研究について述べ，第 3 章で実験について説明する．第 4 章でまとめと今後の課題について述べる．

1. 崇城大学大学院 工学研究科
Graduate School of Engineering, Sojo University
2. 芝浦工業大学
Shibaura Institute of Technology
3. 崇城大学 情報学部
Faculty of Computer & Information Sciences, Sojo University

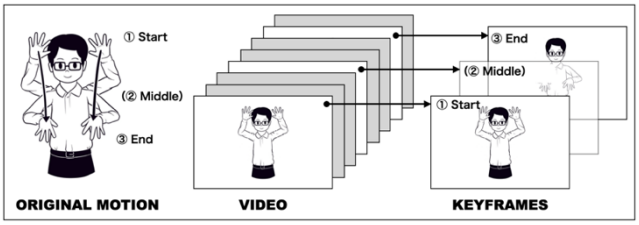

図 1　キーフレームの例

2. 手話キーフレーム映像および関連手法

本研究では手話の特徴を強く表す重要な箇所のことをキーフレームと呼ぶ．図 1 にキーフレームの例を示す．

図 1 左に示す手話の開始時点，中間時点および終了時点（START／(MIDDLE)／END）をビデオからキーフレームとして図 1 右のように抽出する．キーフレームから要約映像を生成する際には，キーフレームから次のキーフレームまで同じ画像を表示し続けるように再構成する．このようにして生成した手話キーフレーム映像は，(1) 高倍速映像と比較した際に，同じ内容理解度であれば手話キーフレーム映像の方がデータ量が少なくてすむ[1]，(2) 原映像と同等の内容理解度を保持することが示唆されている[2]，という特徴を有している．

しかし，キーフレームの抽出は手動であったため，キーフレームでは動作の強調や方向転換のために見かけ上の動きが停留するという仮説に基づき，呉らや髙濱らによってオプティカルフローを用いたキーフレーム候補の自動抽出を行う手法が提案された[3][4]. 図 2 にオプティカルフロー解析の様子を示す．

一方，表情・顔部など手・腕以外の動き（非手指動作）も手話においては重要な意味を持つ．そのため，非手指動作からもキーフレームを抽出できれば，手話キーフレーム映像の内容理解がさらに深まるものと考えられる．髙濱らの研究では，重要な非手指動作の一つである「頷き」に注目し，MediaPipe[7,8]を用いて目の座標の変化や顔領域の縦横比を算出して「頷き」のキーフレーム候補の検出を試みている[4]．図 3 に「頷き」抽出の様子を示す．

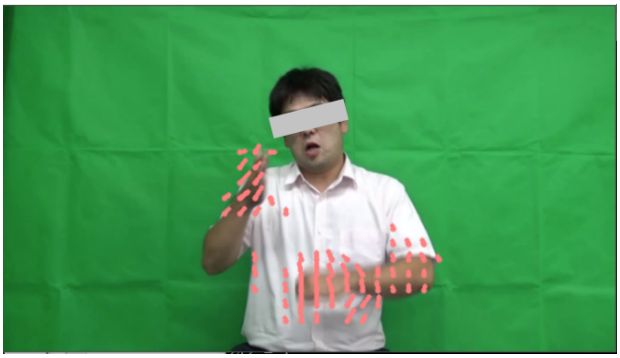

図 2　オプティカルフローによる解析の例

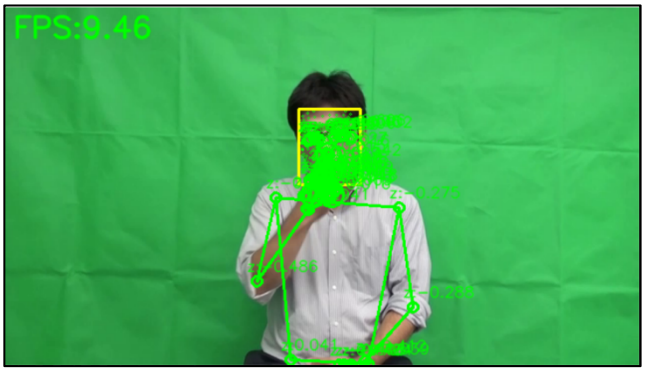

図 3　「頷き」のキーフレーム候補抽出例

手話キーフレーム映像の再構成方式については板井らによって Constant Rate と Proportional Rate が提案されている[5]．Constant Rate（以下，CR）は原映像から抽出したキーフレームを要約映像に再構成する際に，原映像におけるキーフレーム間の時間間隔にかかわらず一定の間隔でキーフレームを表示させる手法である．一方，Proportional Rate（以下，PR）は原映像におけるキーフレーム間の時間間隔の比率を，要約映像においても保持するものである．いずれの方式も，キーフレーム間の時間間隔を調整することにより 2 倍速，3 倍速といった高速度の手話キーフレーム映像を作成することができる．ただし，PR においては，もともとキーフレーム間の時間間隔が短かった箇所においては，それ以上削減できない下限値が存在することになる．また，どちらの方式においても再構成後の総フレーム数がキーフレーム数を下回るとキーフレームが欠落することになるため，そのような要約映像は作成しない（従って高速度にも物理的な意味での上限値が存在する）．

図 3 に手話キーフレーム映像における CR のキーフレーム間隔の例を，図 4 に PR のキーフレーム間隔の例をそれぞれ示す．後述するように，本研究においてはこれら 2 つの再構成方式に基いて再生時間を短くした手話キーフレーム映像と，現映像から指定した倍速に応じた等時間間隔でフレームを抽出した映像とを比較する．

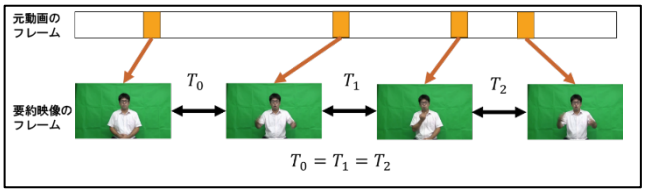

図 3　Constant Rate のキーフレーム間隔の例

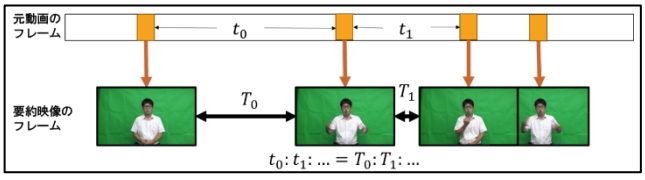

図 4　Proportional Rate のキーフレーム間隔の例

表 1 実験に用いた手話文章[4]

手話文章
1. 集会は毎週の金曜日に開かれます.
2. 小学 4 年の娘は，少女漫画を読むのが好きだそうです.
3. 妻の料理は，とてもおいしいです.
4. 娘は陸上部に入って毎日遅くまで練習しています.
5. 私の夢は小学校の先生になることです.
6. 私は運動不足で困っています.
7. 私は，家族皆が健康なことが自慢です.
8. これからやってみたいことはありますか？
9. 文化祭ではどんなことをやるのですか.
10. 父の会社は，郵便局の前にあります.
11. お兄さんには子どもがいるのですか？
12. 明日は 5 時に仕事を終えて，友達に会います.
1. 集会は毎週の金曜日に開かれます.
2. 小学 4 年の娘は，少女漫画を読むのが好きだそうです.
3. 妻の料理は，とてもおいしいです.
4. 娘は陸上部に入って毎日遅くまで練習しています.
5. 私の夢は小学校の先生になることです.
6. 私は運動不足で困っています.
7. 私は，家族皆が健康なことが自慢です.
8. これからやってみたいことはありますか？
9. 文化祭ではどんなことをやるのですか.
10. 父の会社は，郵便局の前にあります.
11. お兄さんには子どもがいるのですか？
12. 明日は 5 時に仕事を終えて，友達に会います.
13. 台風が来ていて午後は大雨らしいので，早く帰ります.
14. 大きい湖の中に小さな島があります.
15. 高いところから見るので花火がとても近くに見えました.
16. 私の趣味はテレビで野球を見ることです.
17. 私はパソコンが得意です.
18. 私の母は料理が苦手です.
19. 妹はリンゴが好きですが，ミカンは嫌いです.
20. スポーツはするのもテレビで見るのもどちらも好きです.
21. 苦手なスポーツはありますか？
22. 妹は青が好きで自転車は青です.
23. 母は花が好きで昨日黄色と赤色の花を買いました.
24. 好きな映画は何ですか.
25. 母の仕事は手話通訳です.
26. 父の仕事は日曜日は休みです.
27. 私は花を見るのが好きです.
28. 朝 8 時に家を出て自転車で店に行きます.
29. 職場はどこにありますか.
30. 仕事は何時に終わりますか.

3. 実験

前章で述べた先行研究においては，アルゴリズムの提案と実装は行っていたものの，実際に評価実験は行っておらず，有効性が確認できていなかった．本研究では実際に数種類の手話キーフレーム映像を作成し，手話者を対象とする被験者実験を行った．本実験での確認事項は以下のとおりである：

1. 自動抽出キーフレームからの手話キーフレーム映像と手動抽出キーフレームからの手話キーフレーム映像との内容理解に関する比較.
2. CR，PR，等間隔映像の内容理解に関する比較.
3. 何倍速まで内容を理解できるか.

以上の計画のもと，表 1 に示す文章の手話映像から各種手話キーフレーム映像を自動作成した．再生速度は 2 倍速（総フレーム数が原映像の 1/2），3 倍速（同 1/3），4 倍速（同 1/4），5 倍速（同 1/5），6 倍速（同 1/6）を用意した．従って，キーフレーム抽出方法 2 種類（手動／自動），再構成方式 3 種類（CR／PR／等間隔），速度 5 種類，原映像 30 本の，合計 900 本の刺激映像を作成し，以下の実験ではこの中からランダムに抽出して被験者に提示した．

実験に先立って，熊本県ろう者福祉協会に所属する手話者 2 名，手話通訳士 1 名を対象にエキスパートレビューを行った．抽出は手動のみとし，再構成方式を CR と PR，等間隔の 3 種類，速度を 2 倍速，4 倍速，6 倍速とした．合計 9 種類の要約パターンで 9 本の手話キーフレーム映像（1 つのパターンごとに異なる文章を用いた）を提示した．レビュー対象者はこれらの手話キーフレーム映像を 1 つの映像につき繰り返し 3 回閲覧したのちに，手話の内容をどの程度把握できたかを「とてもよく理解できた」「よく理解できた」「やや理解できた」「やや理解できなかった」「あまり理解できなかった」「ほとんど理解できなかった」の 6 件法（5 点から 0 点までを付与）で回答頂いた．その集計結果を図 5 に示す．

その結果，要約手法に関わらず，内容把握のしやすさは再生速度が高速化するほど減少していく結果となった．

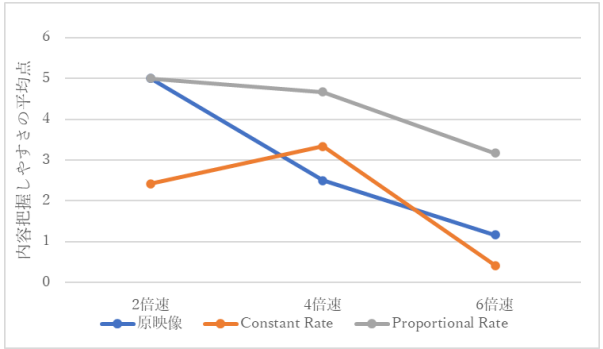

図 5 再構成方式によるエキスパートレビュー結果

図 6　提示映像（実験では顔部分のマスクはない）

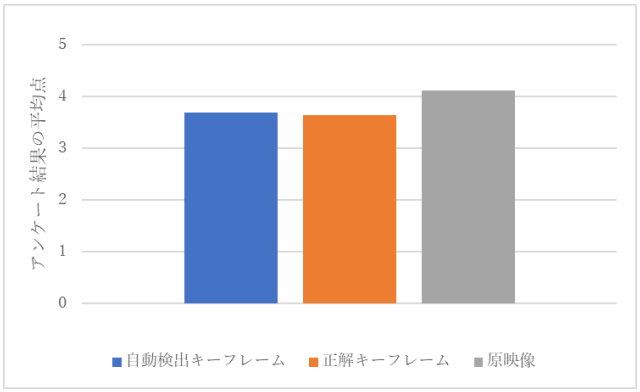

図 7 フレーム抽出方法ごとの手話内容把握しやすさ

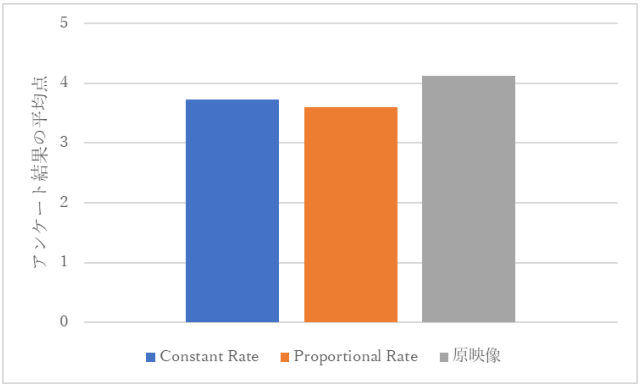

図 8 要約手法別の内容把握のしやすさ

その一方で，3 種類の要約手法を比較すると PR が他の手法と比べ，2 倍速，4 倍速，6 倍速のいずれの速度においても手話文章の内容把握しやすさが最も優れているという結果となった．その理由として，PR では原映像でのキーフレームが表示される間隔が保持されており，CR よりも原映像のリズムに近いことが考えられる．

3.1 実験 1：キーフレーム抽出方法の比較

手話映像から手話者により手動で抽出した「正解」キーフレームと，自動抽出キーフレームをそれぞれ CR，PR で要約映像として再構成し，これらと等間隔映像との比較を行なった．実験では 13 名の被験者を 5 グループに分け，

・キーフレーム抽出方式　2 種類（手動／自動）
・再構成方式　3 種類（CR／PR／等間隔）
・再生速度　1 種類（2 倍速のみ）

の 5 パターン（等間隔映像はキーフレーム抽出方式とは無関係である）で作成した要約映像を一人あたり 10 本提示した．映像は 21 インチカラーモニターで，映像の左右上下に空白を含む．図 6 に提示映像の例を示す．

被験者は 50 代から 80 代であり，手話を使い始めた年代は 10 代が最も多く，10 代から 60 代までであった．エキスパートレビュー同様に要約手話映像 1 本あたり 3 回繰り返し見せたのちに，6 件法（5 点～0 点を付与）で回答頂いた．

まず，キーフレーム抽出方法ごとの評点の平均値は図 7 に示す通りとなった．T 検定で比較したところ，いずれの組み合わせでも有意な差は認められなかった．このことから，2 倍速では等倍倍速でもキーフレーム映像でも差異がなく，また，自動抽出キーフレームのみからなる要約映像であっても手動（正解）キーフレームと同等の内容理解を得られるということが言える．したがって，文献[3][4]で報告された自動抽出方法の妥当性が示されることとなった．

次に要約手法別の手話内容の認識のしやすさの平均値は図 8 に示す通りとなった．この結果も T 検定で有意な差は認められなかった．理由として，2 倍速程度では特に重要なフレームが欠損することがないか，あるいは被験者が内容を容易に補完できる，ということが考えられる．

3.2 実験 2：再生方式及び再生速度についての比較

実験 1 の結果を踏まえ，再生速度を変更すると要約手法による手話内容の認識に差が出るかを検証する．

実験 2 では，実験 1 と異なる日程で，13 人の手話者を実験 1 同様 5 グループに分けて作成した要約映像に対する評価実験を行った．なお，13 名の属性はほぼ実験 1 と同様であり，実験 1 と重複する者もいたが，実験 1 から約 40 日経過していたため影響は少ないと判断した．実験 2 では

・キーフレーム抽出方式　1 種類（自動のみ）
・再構成方式　3 種類（CR／PR／等間隔）
・再生速度　5 種類（2，3，4，5，6 倍速）

の 15 パターン（すべて異なる手話文章）で要約映像を作成した．実験 1 同様に映像を 3 回閲覧した後に実験 1 と同じ項目にて 6 件法で評価頂いた．評点の集計結果（平均値）を図 9 に示す．

この結果を T 検定で比較したところ，各再生速度ごとでは要約手法の組み合わせによる比較で有意な差が認められたものは 2 倍速における原映像と PR 映像，原映像と CR 映像のみとなり，残りの組み合わせでは有意な差は認められないという結果となった．

結果から伺える大まかな傾向としては，どの再構成方式でも，再生速度が上がるにつれて内容把握のしやすさが減少する傾向にあるが，手話キーフレーム映像の方が減少する割合が少ないということが言える．

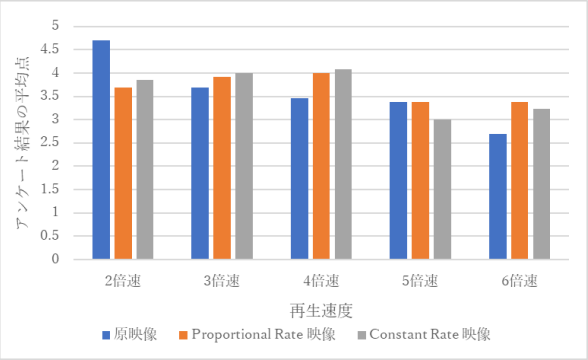

図 9　再生速度，再構成方式による評価結果

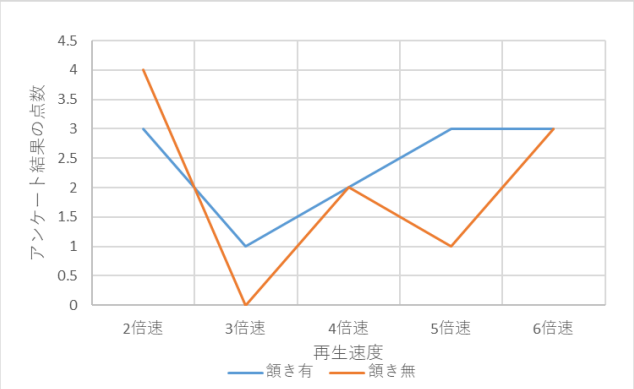

図 10　頷きの有無による認識のしやすさの変化

この結果より，手話文章の内容把握のしやすさは大まかな傾向として，再生速度が高倍速の場合では Proportional Rate 映像が優れていると考えられる．

3.3 エキスパートレビュー：頷きを含む要約映像の評価

さらに，頷きを含む手話キーフレーム映像の有効性について，今後の検証に向けてエキスパートレビューを行った．3.1 節，3.2 節の実験結果を踏まえ，自動抽出キーフレームを PR 手法で再構成したものをそれぞれ 2 倍速，3 倍速，4 倍速，5 倍速，6 倍速の 5 つの速度で作成し，頷きを含むものと含まないものの 2 種類の合計 10 パターンからなる要約映像を用いて，手話者 1 人を対象に意見を伺った．これまでの実験同様，要約映像を要約映像 1 本あたり 3 回ずつ提示し，6 件法で回答いただいた．結果を図 10 に示す．

図 10 より，全般的に頷きを含む要約映像の点数が頷きを含まない要約映像の点数を上まった．このことから，自動抽出キーフレームからなる要約映像に頷きのキーフレームを加えることで手話文章の内容把握のしやすさが向上する可能性が高い．

しかしながら一方で，頷き候補の箇所を加えたことで認識のしやすさが向上したのではなく，単に要約映像中に含まれるフレーム枚数が増えたことが認識のしやすさに影響したという可能性も考えられるため，本レビューで用いた要約映像のキーフレーム数と評価点との相関を調べたところ，相関関係はほとんどないということが分かった．このことから，原映像に対するフレーム抽出の割合が手話文章の内容把握のしやすさに与えている影響はほとんどないと言える（頷きのフレームは 1 本あたり数枚程度である）．

これらの結果から，要約映像に頷きの候補となる箇所を加えることで手話文章の内容把握のしやすさの向上に繋がることが考えられる．

3.4 考察

要約映像を構成するフレームの抽出方法について検証したところ，手話者によって判別された正解キーフレームからなる要約映像と，オプティカルフローを用いて抽出した自動抽出キーフレームからなる要約映像に有意な差が認められなかった．文献[3][4]では，自動抽出キーフレームと正解キーフレムとの適合率が高いとは言えなかったが，この結果により自動抽出キーフレームでも十分であるということが言え，キーフレーム映像作成のコストの面で非常に大きな成果である．

また，要約手法について，特に高倍速の場合であれば PPR 映像が有効だと考えられるが，今回の実験では有意な差が現れなかったため，今後さらに検証が必要である．また，我々が想定していたよりも高倍速において手話者が内容を把握可能であったことに関しては，文章の内容にも依存する可能性があるが，ある程度，動きを補間できるのではないかと考えられることや，手話を用いた会話に慣れているため完全には伝わらなかった箇所を手話映像の前後の文脈から推測できるのではないかと考えられる．

要約映像に頷きの候補となる箇所を加えることによる手話文章の内容把握のしやすさについては，レビューにとどまるものの，特に高倍速の状況下であれば手話文章の内容把握のしやすさに貢献する可能性が示唆された．

これらのことから，特に高倍速の状況下であれば，手話映像から抽出した自動抽出キーフレームと頷きの候補として抽出した箇所を PR を用いて要約映像として再構成したものが手話文章の内容伝達に優れていると考えられる．

4. まとめと今後の課題

本研究は，これまで技術的に提案されていた手話キーフレーム映像について，実際に手話者による被験者評価実験を行い，どのような抽出方法や再構成方式が有効であるがを検証した．

その結果，自動抽出されたキーフレームでも，正解キーフレームのみからなる要約映像と同程度の内容伝達を行うことが確認できた．このことは，災害時など通信料が制限される中で手話映像による情報伝達・情報収集を行う際に活用できるのではないかと考える．統計的な有意差は表出しなかったものの，高速な再生速度においては，PR 方式での再構成が適していることが示唆された．まさらに，非手

指動作加えた手話キーフレーム映像の有用さの可能性も確認することができた.

今後の課題としては，本文では述べなかったが，背景が動く状況でもキーフレームを安定して抽出できる手法の検討や動画像からキーフレームを抽出する際の「ブレ」の除去，などの技術的な面が挙げられる．さらに，現段階では要約映像を実際に「ビデオ映像」として生成しているため，キーフレーム画像のみから例えばブラウザ上で再生できるツールの開発も課題の一つである．また，評価実験の際には使用する映像中に登場する手話者と要約映像を評価する手話者の居住地域が近い状況で行うなど，手話の地域性による影響を最小限に抑えた形で要約映像の評価実験を行いたい.

謝辞 実験に多大なるご協力を頂きました熊本県ろう者福祉協会の皆様，手話に関する多くの貴重な助言を頂きました同協会の一條眞理子理事に深く感謝いたします.

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，Vol.50，No.3，pp.373-382（2021）.
[2] NHK：“天気・防災 手話 CG”，https://www.nhk.or.jp/handsign/（2024/2 閲覧）.
[3] 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”，情報処理学会 火の国情報シンポジウム 2021，2021 年 3 月.
[4] 髙濵 彩香，米村 俊一，筒口 拳：“手話実写映像における「頷き」シーンの抽出方法”，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，2023 年 1 月.
[5] 板井 裕太，西村 洋輝，米村 俊一，筒口 拳：“手話キーフレーム映像の構造に関する検討”，情報処理学会 火の国情報シンポジウム 2022，2022 年 3 月.
[6] 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，2016 年 12 月.
[7] MediaPipe: https://google.github.io/mediapipe/（2023/12 閲覧）
[8] KazuhitoTakahashi: "mediapipe-python-sample", (2023/12), https://github.com/Kazuhito00/mediapipe-python-sample/