

手話のキーフレームに基づく映像圧縮を用いた 災害情報伝達システム

秋山 滉太[†] 筒口 拳[‡] 米村 俊一[†]

芝浦工業大学 工学部 情報工学科[†] NTTメディアインテリジェンス研究所[‡]

E-mail: [†]al12004@shibaura-it.ac.jp

1. 背景と目的

大規模地震などの災害が発生した場合、高齢者や障がい者など弱者の被害が大きいと言われている。特に、聴覚に障がいを持つろう者は、手話によるコミュニケーションが必要となるため、映像による情報伝達を必要とするが、災害時には、通信網の混乱等が発生するため、広帯域映像の送受信が困難となることが想定される。したがって、災害時にろう者をサポートする上では、可能な限りデータ容量を削減しつつも、手話内容が読み取れる映像通信方法を提供する必要がある。

本研究では、ろう者の母語である手話をデータ容量の小さい映像に変換して伝達するシステムを検討する。映像のデータ容量を削減する方法として、手話を読み取るために最低限必要な画像、つまりそれが欠落すると手話を読み取れなくなってしまう画像をキーフレームと定義し、このキーフレームのみから構成された映像(以下、キーフレーム映像)で意味を読み取れる映像を伝送する情報伝達システムの実現をめざす。キーフレーム映像と、通常の映像のフレーム数を均等に削減した映像とを比較した実験を行ったところ、同様の内容理解度を示す場合、キーフレーム映像のデータ容量が少なく済むことを確認した。

2. キーフレームを用いた手話映像

先行研究^[1]では、ろう者が手話を読み取れる限界の時間/空間解像度の調査を行っており、空間解像度では 268x152、時間解像度では 8fps までならば、ろう者は手話を読み取れるということを示している。我々は、手話映像からキーフレーム画像を抽出し、そのみを伝送することによって、可読性の低下を抑えつつ、よりデータ容量を削減する情報伝達システムを提案する。

手話は話者の意図を手指によって空間表現する言語であり、空間表現の変化点に言語の特徴が出やすいと考える。この特徴を利用して、本研究では次の2種類のキーフレーム画像を抽出する。

(1) 静止フレーム：例えば、手話表現の開始点と

終了点のように、手指の動きが静止する瞬間を静止フレームとして抽出する。

(2) 中間フレーム：二つの静止フレームを補完するため、手話表現の中間地点の画像フレームを抽出し、これを中間フレームとする。

筆者らが実施したキーフレーム推定実験^[2]では、ろう者の男性 2 名の協力を得て最適なキーフレーム画像を実験的に抽出した。このキーフレーム映像は、次のような特徴を備えていた：

(1) 中間フレームは必ずしも必要でない

中間フレームが挿入されていない静止フレームのみで構成した映像でも、手話の意味が十分に読み取ることができる場合があった。

(2) 映像全体のコマ割りを一定に保つことが必要

一定時間内のコマ数にムラがある映像は、違和感を生じさせ、意味の読み取りも困難になる。

本研究では^[2]の結果に基づきキーフレーム映像を作成し、情報伝達の評価を行った。

3. キーフレーム映像の評価実験

3.1 刺激映像

事前に撮影した 15 種類の緊急メッセージ映像からキーフレームを抽出した。第 2 章で説明した特徴に基づき、次の 3 種類のキーフレーム映像を作成した。

(1) 中間フレームなし方式

中間フレームを使用せず、すべて静止フレームで構成したキーフレーム映像。

(2) 中間フレーム数 1 枚方式

中間フレームを 1 枚だけ、静止フレームの間に挿入したキーフレーム映像。

(3) 中間フレーム数調節方式

静止フレーム間の時間間隔に合わせ、複数の中間フレームを挿入し、映像全体のテンポがほぼ一定になるように調節したキーフレーム映像。

これら 3 種類の映像に加え、^[1]で提案されている低時間解像度(一定間隔でフレームを間引いた)映像(以降従来手法)を比較対象として使用した。

この際、ろう者のアドバイスをもとに、3 種類のキーフレーム映像中で最も使用フレーム枚数が多い「中間フレーム数 1 枚方式」のフレーム枚数から平均的な fps である 5~6fps となるように調節した。映像はいずれも空間解像度が 640x360 であり、ファイル形式は汎用性やデータ容量の圧縮率や編集効率を考慮して.gif (GIF アニメーション)

Disaster information transmission system using the video compression based on the key frame of sign language.

Kota AKIYAMA[†](al12004@shibaura-it.ac.jp)

Ken TSUTSUGUCHI[‡]

Shunichi YONEMURA[†]

Shibaura Institute of Technology[†]

NTT Media Intelligence Laboratories[‡]

とした。

3. 2 実験方法と評価手法

実験では、26～64歳のろう者20名(平均38.55歳, SD=8.96)に対し、1～15の文章それぞれにおいて4種類の映像パターンのどれか一つをディスプレイにランダムで表示し、一つ映像を見るたびに被験者の主観で映像の読み取りやすさを1:非常に悪い(わからない)から6:非常に良い(とてもよく分かる)の6段階で評価してもらった。どの被験者も、2級、もしくは1級の聴覚障害を有し、両耳全ろうであり、3歳以前に失聴している。

また、映像の内容を正しく認識できているかの確認のため、映像の意味内容を手話で表現してもらい、実験の協力者であるろう者の男性一名(33歳)にその表現内容を日本語で記録してもらった。

3. 3 結果

各映像パターンの全文章の評価の平均得点を図1に示す。縦軸は評価値、横軸は各方式である。

図1より、中間1方式の評価値が最も高いが、どの方式も評価値は大きく異ならなかった。図2は、被験者が読み取った内容が提示した文章内容と異なった回数、図3は、映像の読み取りが不可能だった回数である。縦軸は回数、横軸は各映像伝達方式である。図1に示した評価点数について分散分析を行った結果($p=0.36>0.05$)、有意差は見られなかった。図2の読み間違い回数では、従来方式と中間なし方式が最も少なかった。図3の

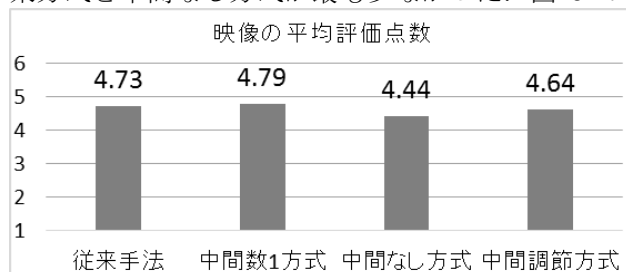


図1. 各映像伝達方式の平均評価点

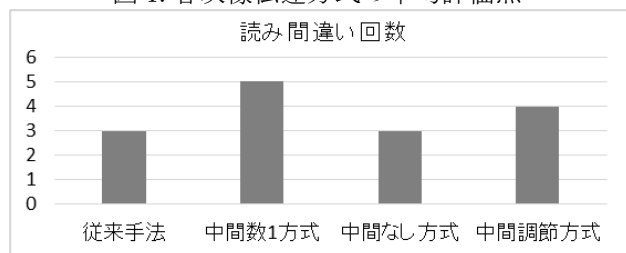


図2. 各映像伝達方式の読み間違い回数

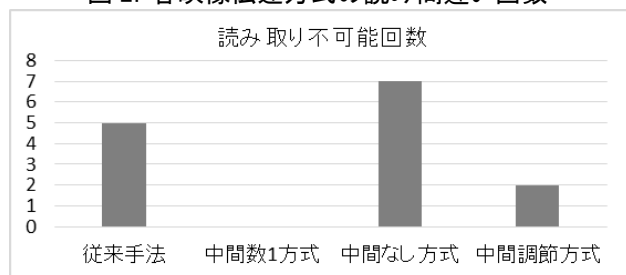


図3. 各映像伝達方式の読み取り不可能回数

表1. 各方式のフレーム数平均及び平均データ量

	従来手法	中間数1	中間なし	中間調節
枚数(枚)	69.9	47.5	61.4	31.2
データ量(MB)	4.25	2.91	3.76	1.99

読み取り不可能回数では、中間1方式において読み取り不可は発生せず、また中間なし方式で最も多い結果となった。

また、各方式の15種類のメッセージ映像にわたる総フレーム数の平均枚数を表1に示す。

3. 4 考察

[1]では、手話が読み取れる時間解像度の限界は8fpsであると報告されていたが、本実験では5～6fpsでも手話の読み取りが可能であり、かつ、映像の評価点数も他の方式と差異がないという結果が得られた。実験協力者であるろう者に確認してもらった結果、[1]の手話表現は話者の癖が強く、今回の手話表現は癖がほとんどないという見解であり、これが影響したと考えられる。

また、今回使用した4種の映像パターンでは、それぞれフレーム総数が異なっている。例を挙げると、従来手法の映像に使用されているフレーム枚数に対し、中間フレームなし方式は半分以下のフレーム枚数で構成されている。方式によってフレーム数がかなり異なるにもかかわらず、どの方式も被験者が同程度の理解度を示していることが分かる。すなわち、キーフレームを扱うことでより少ないフレーム数で同等の性能を出すことが出来たと考え、データ容量の削減が可能となることを期待できる。

実験終了後一部の被験者に対し、従来手法の映像とキーフレーム映像を同時に提示しどちらの映像がよりわかりやすいかアンケートを実施した。その結果ほぼすべての被験者がキーフレーム映像の方がより見やすいと回答していた。

4. まとめ

本稿では、手話表現の空間的な特徴に着目して映像圧縮を行い、災害時のような低容量の回線で手話映像を伝送するキーフレーム映像伝達システムを提案した。評価実験の結果、キーフレーム映像を用いることで均等にフレーム数を削減する手法より少ないフレーム数で同等の可読性を持つ映像を作成でき、データ容量を削減できることが示唆された。今後は最適なキーフレーム映像の方式を検討する予定である。

参考文献

- [1] 山崎真光, 米村俊一, 筒口拳「手話映像を用いた災害時の緊急案内の画質に関する検討」, 信学技報 WIT2014-58(2014-12), pp29-34.
- [2] 秋山滉太, 米村俊一, 筒口拳「キーフレーム通信方式を用いた災害時情報伝達システムの提案」, 信学技報 WIT2015-46(2015-08), pp53-58.