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

呉 夢竹 [†1]　米村 俊一 [†2]　筒口 拳 [†1]

概要：我々は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研究に取り込んでいる．しかし現階段では，キーフレームを自動で抽出する手法が確立されていない．本研究ではキーフレームでは手の動きが停留するであろうという仮定のもと，手話映像のオプティカルフローを用いてキーフレームを自動抽出することを試みた．実験の結果，単語ではほぼ全てのキーフレームを抽出でき，文章では 77%の抽出率となった．

キーワード：手話，映像要約，キーフレーム，オプティカルフロー

Extraction of Keyframe Candidates from Sign Language Video using Optical-flow

Mengzhu WU [†1]　Shunichi YONEMURA [†2]　Ken TSUTSUGUCHI [†1]

Abstract: In order to efficiently transmit sign language videos, we are studying a method of extracting images called keyframes that strongly represent the characteristics of sign language from the videos, and reconstructing the videos from only the keyframes. However, extracting keyframes automatically has not been established. In this study, we attempted to automatically extract keyframes using the optical flow analysis of sign language video, assuming that hand movements would stop (or be slow) at keyframes. As a result of the experiment, almost all keyframes could be extracted for words, and the extraction rate for sentences was 77%.

Keywords: Sign language, video abstraction, keyframe, optical flow.

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，映像を再度確認する際に時間がかかってしまう．手話映像を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．重要なシーン（画像）だけで手話映像を構成することができれば，内容を損なうことなく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．

秋山らは手話映像を「キーフレーム」とよぶ「手話の特徴が強く表れている画像フレーム」のみで再構成する手法を提案している[1-6]．これらの報告によれば，キーフレームだけで構成された要約映像（「キーフレーム映像」）でも内容を伝達することが可能であることが示唆され，キーフレーム映像の有効性が示されている．しかし，キーフレームを自動で抽出する手法が確立されていない．

これに対し，品田らは，キーフレームでは手指の運動が停留するという仮定のもと，手話演者にカラー手袋を装着させることで手の領域を色相や彩度によって分離し，手領域の重心位置の時間変化を求め，その軌跡グラフが極値を取るところ（時間的フレーム位置）をキーフレーム候補として抽出をこころみているが，ノイズの影響によりキーフレームではない候補が多数検出されるという課題があった[7,8]．また，入江らは肘部の骨格の動作を OpenPose で解析し，位置座標の時間変化をプロットして曲線で近似して極値を自動検出する手法を提案している[9]．いずれの手法も，身体部位の位置の時間変化に伴うノイズの問題がある．

本研究では，キーフレームでは手指の運動が停留するであろうという品田らの仮定を前提として，手指の位置を推定するのではなくオプティカルフローを用いて大局的な運動の変化を検出し，その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ことで自動抽出することを試みた．少数ではあるが単語を表す映像と文章を表す映像に対して処理を行い，単語ではほぼ全てのキーフレームを抽出でき，文章では 77%の抽出を行うことができた．

以下，2 章でキーフレームについて説明し，3 章で提案手法の詳細について述べる．4 章で実験結果について述べ，5 章でまとめる．

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

2. キーフレーム

手話映像において手話の特徴を強く表しているフレーム画像を本研究ではキーフレームと呼ぶ．キーフレームは手話において単語の開始点や終了点，手指の移動方向が変化する点などのように，運動が停留したり，ある方向の速度が 0 になるような時点が考えられる．現実においては動きが完全に停止するわけではないこともあるが，そのようなフレームの近傍では動きが遅くなると考えられる．

このようなキーフレームのみから再構成された映像でも内容の伝達が可能である[1-6]．図 1 にキーフレームの概念を示す．(a)で示された動作を理解するには，(b)の映像すべてを再生する必要がなく，(c)のキーフレームだけでよい，というものである．反面，キーフレームが欠落すると内容伝達が困難になるため，キーフレームを抽出する際には不足よりも冗長の方が望ましいということになる．

キーフレームを自動抽出することができれば，既存映像も有効利用することが可能となる．キーフレーム映像は実写映像をもとに生成するため，表情といった非手指情報を表すことができると考えられる．また，キーフレームは手指の動きから抽出するため，外国語の手話やいわゆる方言・個人差にも対応できる可能性がある．

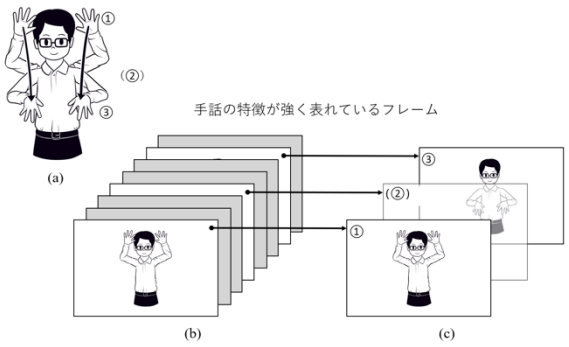

図 1　キーフレーム概念図；(a)手話，(b)実写映像，(c)キーフレーム

Figure 1. Concept of "keyframe"; (a) sign language, (b) video, (c) keyframes.

3. 提案手法

オプティカルフローとは物体やカメラの移動によって生じる隣接フレーム間の物体の動きの見え方のパターンである[10]．あるフレームの特徴点（または画素）と，次のフレームの対応する特徴点（または画素）の移動を表す量がフローベクトルである．図 2 にフローベクトルの例を示す．動きが多いほどフローベクトルの数も多いため，手話映像において動きが停留する時間位置ではフローベクトルの数も減少する．

本研究では，フローベクトル数の変化により手の動きの停留を検出しキーフレーム候補を抽出する手法を提案する．手の重心位置や腕の姿勢推定を用いる手法が局所的な位置変化に基づくのに対し，オプティカルフローを用いる手法は大域的な解析に基づくと言える．

フローベクトルは大きさを持つため，カウントするベクトルの大きさの下限値をどのように設定するかによって時系列データの形状が異なってくる．カウントするフローベクトルの大きさの下限値を d と書くことにすると，d 以上のフローベクトル数により動きの量を判断することができる．フローベクトルの数が少ないときは動きが穏やかな時であり，時間軸に対し極小値をとるフレームをキーフレーム候補として見なすことができる．

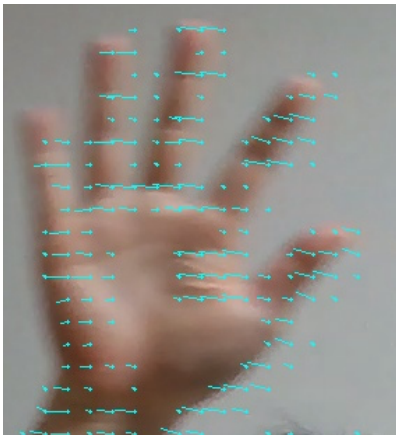

図 2　フローベクトルの例

Figure 2.　Example of flow-vector

4. 実験

4.1 対象とする手話

実験で用いた手話映像は 3 つの単語：「徹夜する」，「熱い」，「相手」と 3 つの文章：「妻の料理はとてもおいしいです」，「好きな映画は何ですか？」，「集会は毎週の金曜日に開かれます」の合計 6 種類であり，事前に正解となるキーフレーム位置を手話者の目視にて抽出しておく．映像のサイズは Full-HD（1920x1080）である．実験対象のフレーム数および正解キーフレーム数を表 1 に示す．これらの映像に対し，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抽出を行い，正解と比較し評価を行う．

表 1　実験対象の手話映像

Table 1. Sign language videos for experiment

	総数	正解 KF
単語 1:徹夜する	277	3
単語 2:熱い	247	2
単語 3:相手	277	2
文 1:妻の料理はとてもおいしいです	337	10
文 2:集会は毎週金曜日に開かれます	412	22
文 3:好きな映画は何ですか？	307	13

※総数：総フレーム数，正解 KF:正解キーフレーム数

表 2　オプティカルフローのサイズ

Table 2.　The size of optical-flow

	キーフレーム数	検出数	誤検出数	正解率	誤検出率	正解率点数	誤検出率点数	検出数点数	合計点数
d＝1	3	34	53	81%	60.92%	13	3	13	29
d＝2	3	33	61	79%	64.89%	10	1	10	21
d＝3	3	33	60	79%	64.52%	10	2	10	22
d＝4	3	33	46	79%	58.23%	10	5	10	25
d＝5	3	28	26	67%	48.15%	9	12	9	30
d＝6	3	22	24	52%	52.17%	3	11	3	17
d＝7	3	24	21	57%	46.67%	8	13	8	29
d＝8	3	23	22	55%	58.18%	4	6	4	16
d＝9	3	17	19	40%	52.78%	1	10	1	12
d＝10	3	23	27	55%	54.00%	4	9	4	17
d＝11	3	20	25	48%	55.56%	2	8	2	12
d＝12	3	23	29	55%	55.77%	4	7	4	15
d＝13	3	23	33	55%	58.93%	4	4	4	12

4.2 オプティカルフローのサイズ

フローベクトル数をカウントするサイズの下限値 d を設定する．手話映像「徹夜する」を対象として，d＝1～13 としたときのキーフレーム候補検出数，正解率，誤検出数，誤検出率は表 2 に示すようになる．それらに対し，1 位（13 点）から 13 位（1 点）まで点数を付け，最も点数が高い d を採用する．表 2 の結果より，最も高得点であった d ＝ 5 を他の映像の解析においても採用することとした．

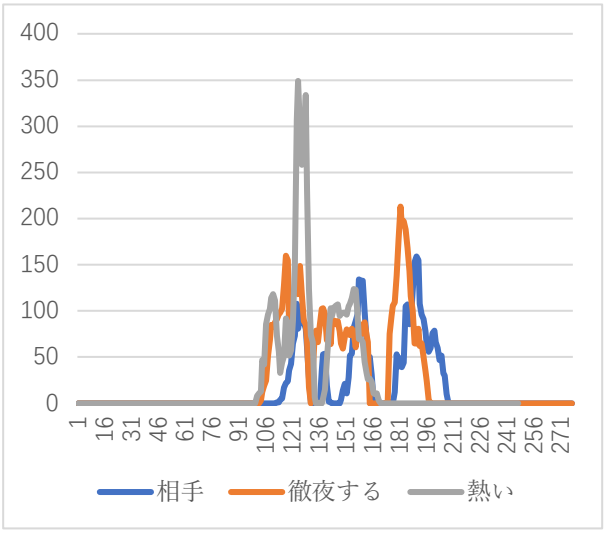

図 3　フローベクトル数（単語，d=5）

Figure 3.　Number of flow vector (word, d=5)

4.3 実験結果

（1）単語

フローベクトル数が極値をとる場合でも，ベクトル数が比較的多い場合には動きが停留しているとは考えにくい．従って，カウントしたフローベクトル数がある閾値 T よりも少ないという条件を満たした極小値のみをキーフレーム候補とみなす．

図 3 は d=5 のときのフローベクトル数の時間変化を表している．横軸はフレーム数，縦軸はベクトル数である．開始から 100 フレームあたりまでは話者が待機状態にあり，また 200 フレームあたりから最後までは手話が完了している状態であるため解析対象から外し，「徹夜する」の正解キーフレームとグラフとを比較し，閾値 T を 60 に設定した．

以上の結果，単語手話映像からのキーフレーム候補の抽出条件として以下の 3 つを全て満たすものとする：

1）d ＝ 5 である

2）T ＝ 60 である

3）フローベクトル数のグラフが極小値をとる

ただし，解析対象であるがフレーム数 0 が連続する場合はその中の 1 つを極小値とみなしている．抽出結果を表 3 に示す．正検出数は正解キーフレームのうち提案手法によって抽出されたものに含まれている数であり，検出漏れはキーフレームであったにもかかわらず候補として抽出されなかったものの数である．

表 3　単語抽出結果

Table 3.　The result of word extraction

（単位：数）	正解	検出候補	正検出	検出漏れ
徹夜する	3	3	3	0
相手	2	7	*5 (2)	0
熱い	2	6	3	0

＊1 つの正解キーフレームの許容範囲内で複数個検出されたもの
カッコ内は検出できた正解キーフレーム数

（2）文章

文章は単語に比べ手の動きが複雑になる傾向にあるため，フローベクトル数も単語に比べ多くなる．文章においても d=5 とし，「妻の料理はとてもおいしいです」のフローベクトル数を正解と比較し，T=160 と設定した．図 4 に文章のフローベクトル数の時間変化を示す．

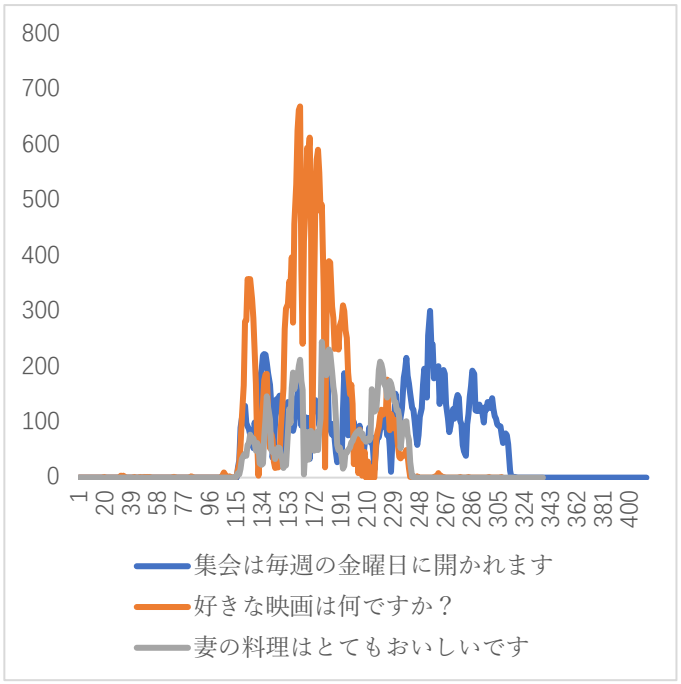

図 4　フローベクトル数（文章，d=5）

Figure 4.　Number of flow vector (sentence, d=5)

表 4　文章抽出結果
Table 4.　The result of sentence extraction

	正解数	候補数	正検出数	検出漏れ
文 1	10	22	*14(8)	2
文 2	22	31	*17(16)	6
文 3	13	16	*12(11)	2

＊1 つの正解キーフレームの許容範囲内で複数個検出されたもの
カッコ内は検出できた正解キーフレーム数

この文章（文 1）を含め，文 2「集会は毎週金曜日に開かれます」，文 3「好きな映画は何ですか？」を d=5，T=160 でキーフレーム候補を抽出した結果を表 4 に示す．

4.4 考察

前節で行った実験より単語と文章両方の実験結果をまとめ，それぞれの再現率と適合率を表 5 に示す．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．

実験結果により単語・文章いずれにおいても検出漏れよりも誤検出が多い傾向にあることがわかる．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．

一方，文章については実験の設定値（d=5，T=160）において検出漏れが発生した．現段階では単語 1 例，文章 1 例からこれらの値を設定しているため，より多くの例で最適な値を設定できるようにする必要がある．また，文章の中には短い時間内で繰り返し動作を含むものがあり，フローベクトル数の極小値により検出することが難しい場合があるという課題がある．

しかしながら，先行研究でキーフレーム候補の数が正解数の 8〜9 倍程度だったのに対し，提案手法では 2 倍程度になっており，提案手法の有効性が伺える．

表 5　再現率と適合率
Table 5.　Recall and precision.

	正検出	誤検出	検出漏	再現率	適合率
単語	8	5	0	100%	62%
文章	35	26	10	78%	57%

5.　まとめ

手話映像からのキーフレーム自動抽出を目的として，密なオプティカルフローのフローベクトルの数値により手話映像からキーフレーム自動抽出する仮説を立て，手話映像の単語と文章を対象として実験を行った．単語 1 例，文章 1 例からカウントするフローベクトルのサイズ d と数の上限 T を設定し，他の単語 2 例，文章 2 例に適用し，キーフレーム候補の検出を試みた．

その結果，誤検出はあるものの単語では比較的良好にキーフレーム候補の抽出が行え，文章でもキーフレーム候補数を先行研究に対し減らすことができた．一方で文章においては検出漏れも見られたため，より適切な d や T を設定する必要がある．

今後はこれらの課題を解決するとともに，他のキーフレーム候補抽出手法とも組み合わせ，より精度の高い抽出をめざす．

謝辞

本研究は科研費（基盤研究（C））19K2032「手話映像の時間的要約方式に関する研究」に基づくものである．また，実験を行うにあたってご協力いただいた NTT クラルティ株式会社の関係者のみなさま，芝浦工業大学 猪岡翔氏に感謝する．

文　献

[1] 秋山滉太，筒口拳，米村俊一：“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（福祉工学研究会）(Aug.　2015).
[2] 秋山滉太，筒口拳，米村俊一：“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”，情報処理学会全国大会(Mar.　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：20 名のろう者によるキーフレーム映像の有効性評”，信学技報（ヒューマンコミュニケーション基礎研究会）116(31)，201-206(MaY 2016).
[4] 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).
[5] 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).
[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 〜手話映像の繰り返し再生におけるろう者の情報取得ストラテジー 〜”，第 91 回福祉情報工学研究会(Aug.　2017).
[7] 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).
[8] 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，2017 年 12 月
[9] 入江健太，米村俊一，筒口拳：“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”，画像電子学会 292 回研究会，2020 年 2 月．
[10] OpenCV-Python Tutorials: "Dense Optical Flow in OpenCV″, https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_video/py_lucas_kanade/py_lucas_kanade.html .