

ライフログ映像活用のための表情強度を取り入れた 表情表出シーン検索

Emotional Scene Retrieval Using Expression Strength for Effective Use of Lifelog Video

森國 淳司† 野宮 浩揮† 宝珍 輝尚†
Atushi Morikuni Hiroki Nomiya Teruhisa Hochin

1. はじめに

ライフログとは文字通りに解釈すると生活の記録である。しかし、ライフログと一口に言っても、記録媒体、情報の種類などは多岐にわたる。例として、ブログ、GPS 記録、個人の健康状態などが挙げられる。新保はライフログの定義を「特定自然人の、特定活動に関する、特定の記録媒体に、自動的に、デジタルデータとして、包括的または連続的に記録（蓄積）し、それによって取得された、特定の個人に関する個人情報（個人識別情報）および個人に関連する個人情報に該当しない情報（非個人識別情報）の総称」としている[1]。

日常生活上の様々なデータを取得・蓄積して、後に有効利用しようというライフログが提案されてきている[2]。本研究では、ホームビデオのような映像データは手軽に作成することができ、またライフログとして有用な様々な情報を内包することができる点に着目し、ライフログ映像を対象としている。

従来はライフログ映像の取得、保存には高価な記録機器や大容量のストレージが必要であったためライフログの実用化が難しかったが、近年のマルチメディアデータ記録機器の高性能化やストレージの大容量化、価格の低下に伴い、個人でも簡単に大量のマルチメディアデータ（例えば子供の成長記録など）を作成できるようになった。しかしながら、多量の映像が蓄積されるので、映像の量が増える一方で、記録した映像を利用したくても、必要な映像を即座に引き出すことが出来ず、有効活用できていないというのが現状である。このような背景から、ライフログ映像の中から有用な映像を簡単に検索したいという要求が高まってきている。

映像に関する検索システムとして、YouTube[3]等の動画共有サイトが挙げられる。多くの動画共有サイトでは、各動画に対してタグを登録し、検索に活用している。これに対し、ライフログ映像は一般に個人が所有しているため所有者が映像一つ一つにタグを登録しなければならず、タグの登録に非常に手間がかかるので、タグを用いた検索は、ライフログ映像検索には適していないと考えられる。また、ホームビデオ等のライフログ映像では、ハンディビデオカメラなどで単純に撮影を行うだけでは、映像にタイトルをつけることや、フォルダごとに整理する事は可能でも、一つの映像の中でより重要なシーンを抽出することは難しい。1 日中ずっと撮影を行なって大量のデータを得たとしても、実際にもう一度動画として視聴したいと感じるシーンはその中のごく一部である[4]。このことから長時間の映像の中からユーザが視聴したいと感じるシーンの検索方法が必要である。

そこで本研究では、ライフログ映像から効率的に重要なシーンを検索、提示することを目的とする。

本研究では、視聴者が視聴したいと感じるシーンの検索方法として、映像中の人物の表情を利用する。ユーザが検索したくなるようなシーンでは基本的に何らかの事象が発生し、シーン中の人物はその事象に対して何らかの感情を抱き表情を表出している可能性が高い。このことから映像中の人物の表情を予め取得し、その表情を検索キーとしてシーン検索を行う。

著者らはこれまでに、映像中の人物の表情を用いた、ライフログ映像からの印象的なシーンの検索に関する研究を行ってきた[5]。この先行研究の中では、動画内の人物の表情の検索はできるのだが、表情の程度を推定することができない。例えば、「笑顔」という表情は検索可能であるが、「笑顔」の中にも「微笑み」から「高笑い」まで感情の程度が異なるものが存在し、これらは考慮されていない。また、実際に動画のシーン検索を行うにあたって表情の程度を考慮することで、より感情が大きく表出しているような事象を優先的に検索することができると考えられる。

そこで本研究では、このような表情の程度を数値化し、検索に活用することを目的として、顔の特徴点を用いて表情強度という値を導入し、表情の程度の推定を行う。

本論文の構成は次の通りである。まず 2 章で関連研究について述べ、3 章で本研究で扱う顔の特徴点について説明する。その後 4 章で本研究の提案手法を述べ、5 章でその提案手法を用いた実験の内容と結果を示し、6 章でその考察を行う。最後に 7 章でまとめを行う。

2. 関連研究

人間の感情を表すいくつかの表情について、その認識、分類のみならず、無表情から最大まで表情の程度をおおまかに推定することを目的として、木村ら[6]は、顔の部位それぞれの特徴点を抽出し特徴量を作り出すのではなく、顔全体を一つのパターンとして表情認識を行う手法を提案している。処理に用いる顔画像を切り出しサイズの正規化を行った後に、ポテンシャルネット[7]と呼ばれる処理モデルを用いて解析を行っている。しかし、この手法では顔の画像を 1 枚ずつ人の手で適切に切り抜く必要があり、動画の解析で各フレームに同様の処理を行うのは非現実的である。また、取り扱っている表情についても、俳優に作為的に演技してもらった表情ではなく、素人の曖昧な表情は認識結果にノイズとして影響を与えることを言及している。そのため本研究で扱うようなライフログ映像で撮影が想定される曖昧な表情については不適であると考えられる。

3. 顔特徴点[5]

本研究で扱う顔特徴点は、顔特徴点抽出ソフトウェア Luxand FaceSDK 4.0[8]を用いて抽出しており、以下の顔の特徴点 59 点を用いている。図 1 にこれらの顔特徴点の抽出例を示す。

- 両目付近 22 点
- 鼻 9 点
- 両眉 10 点
- 口 14 点
- ほうれい線 4 点

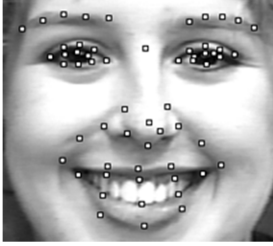


図 1 顔特徴点抽出例（四角い点が顔特徴点を表している）

4. 表情強度

表情の程度を数値化することを目的として本研究では表情強度を定める。本節ではその概要を示す。

4.1 表情強度の推定に用いる特徴量

前述の顔特徴点を組み合わせて、以下に示す 10 種類の特徴量を定め表情強度の推定に用いる。

- ・左眉と右眉の傾き f_1

左眉（右眉）の傾きは、左眉（右眉）上の 5 つの顔特徴点から最小 2 乗法（線形）により算出し、左右の眉がほぼ対象であると考え、 $f_1 = (a_l - a_r)/2$ とする。ただし a_l は左眉、 a_r は右眉の傾きである。

- ・眉と目の間の距離 f_2

両眉の 10 点と上瞼の 10 点を両方共左から順に 1 対 1 対応で結んだ 10 本の線分の長さの平均値である。

- ・眉間の面積 f_3

両眉と両目からそれぞれ最も顔の中央に近い点を結んでできる四角形の面積である。ただし、左目と右目の距離の 2 乗で割って正規化を行う。

- ・目の面積 f_4

左目（右目）の 8 つの特徴点をつないでできる八角形の面積 b_l, b_r の平均値 $(b_l + b_r)/2$ である、ただし左目と右目の距離の 2 乗で割って正規化を行う。 b_l は左目、 b_r は右目の特徴量である。

- ・目の縦横比 f_5

左目（右目）の一番上にある点と一番下にある点を結んだ線分を目の縦の長さ l_v 、最も左と右にある 2 点を結んだ線分を目の横の長さ l_h とし、 $\tan^{-1}(l_v/l_h)$ を左右の目で平均値をとって縦横比とする。

- ・口の内側の面積 f_6

口の内側の 8 つの特徴点を結んでできる八角形の面積である。ただし左目と右目の距離の 2 乗で割って正規化を行う。

- ・口の外側の面積 f_7

口の外側の 8 つの特徴点を結んでできる八角形の面積である。ただし左目と右目の距離の 2 乗で割って正規化を行う。

- ・口の内側の縦横比 f_8

口の内側の点の中で一番上にある点と一番下にある点を結んだ線分を目の縦の長さ m_v 、最も左と右にある 2 点を結んだ線分を目の横の長さ m_h とし、 $\tan^{-1}(m_v/m_h)$ を縦横比とする。

- ・口の外側の縦横比 f_9

口の外側の点の中で一番上にある点と一番下にある点を結んだ線分を目の縦の長さ m_v 、最も左と右にある 2 点を結んだ線分を目の横の長さ m_h とし、 $\tan^{-1}(m_v/m_h)$ を縦横比とする。

- ・口角の上がっている度合 f_{10}

口の両端の点 x_1, x_2 と口の最も上の点と下の点 x_3, x_4 から次式によって求められる。

$$\frac{(y(x_3) + y(x_4)) - (y(x_1) + y(x_2))}{|y(x_3) + y(x_4)|}$$

ここで、 $y(x)$ は特徴点 x の y 座標を表す。 x_1 と x_2 の y 座標の平均値が x_3 と x_4 の y 座標の平均値より大きければ正の値をとるため、この値が大きいほど口角が上がっていると考える。分母は正規化のための項である。

以上、10 個の特徴量をライフログ映像中の各フレーム画像（10 フレームおきに抽出したもの）に対して求め、フレームごとに 10 次元特徴ベクトル (f_{i1}, \dots, f_{i10}) を構成している。

4.2 表情強度の算出

4.1 節では人間が表情を表出する際に大きく変化すると考えられる 10 個の顔の特徴量を示した。しかし、どの特徴量を重視すればよいかは、表情の種類や個人差などに依存すると考えられる。そこで、各特徴量に対して主成分分析を行い、その固有値と主成分得点にもとづいて表情強度を定める。 i 番目のフレームの表情強度 I_i は、4.1 節で示した 10 個の特徴量 (f_{i1}, \dots, f_{i10}) に対して主成分分析を行い、得られた第 1 ～第 10 主成分得点 (p_{i1}, \dots, p_{i10}) を用いて次の式(1)のように定める。なお、主成分得点は平均 0、分散 1 に正規化する。

$$I_i = \sum_{j=1}^{10} \lambda_j \text{sgn}(s_j) p_{ij} \quad (1)$$

ここで、 λ_j は第 j 主成分の固有値である。また、

$$s_i = \sum_{j=1}^{\lfloor N/2 \rfloor} \sum_{k=1}^{10} f_{r_{ij}k} - \sum_{j=\lfloor N/2 \rfloor + 1}^N \sum_{k=1}^{10} f_{r_{ij}k} \quad (2)$$

である。 N は映像に含まれるフレームの総数、 r_{ij} は $p_{1j} \sim p_{Nj}$ を降順に並べた時の p_{ij} の順位を示す。例えば、 p_{ij} が $p_{1j} \sim p_{Nj}$ の中で 30 番目に大きい値を持つ場合、 $r_{ij} = 30$ となる。また、

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{otherwise} \end{cases} \quad (4)$$

である。

一般に、顔面上の変化が大きい場合は、それぞれの特徴量 (f_{i1}, \dots, f_{i10}) の値は大きくなる傾向にある。したがって、式(2)の第 1 項 (p_{ij}) の値が大きいものから順に $\lfloor N/2 \rfloor$ フレームの特徴量の値の総和)が第 2 項 (残りの $N - \lfloor N/2 \rfloor$ フレームの特徴量の値の総和) よりも大きい場合は、 p_{ij} の値が大きいと表情強度が強いと考えられる。したがって、そのような場合 ($s_j > 0$) には、 p_{ij} に正の係数 ($\text{sgn}(s_j) = 1$) を乗じる。逆に、式(2)の第 1 項が第 2 項よりも小さい場合 ($s_j < 0$) には、 p_{ij} に負の係数 ($\text{sgn}(s_j) = -1$) を乗じる。また、式(2)の第 1 項と第 2 項が等しい場合 ($s_j = 0$) には、 p_{ij} から表情強

度を求めることが難しいと考えられるため、 p_{ij} に $(sgn(s_j) = 0)$ を乗じて、表情強度が p_{ij} の値に影響されないようにしている。

5. 実験

本節では4節で示した表情強度を用いて、表情の程度を考慮した表情表出シーンの検出実験を行った結果を示す。

5.1 実験設定

被験者は男子学生4名(21~22才)である。4名を2名×2グループに分け、一方の被験者の前にWebカメラを設置した。一回の撮影につき10分を目安にゲームの区切りの良いところまでトランプの神経衰弱を行なってもらい、4名がそれぞれ一回ずつ表情の映像が撮影できるように4度撮影を行った。その後、別途、そのトランプを行なっている様子をホームビデオ風に撮影していた映像を被験者に視聴してもらい、検索したいシーンをその映像中からアンケートで回答を行ってもらった。撮影機材は以下のとおりである。

- 表情撮影カメラ ELECOM UCAM-DLU130HWH
このカメラの主な仕様を以下に示す。
- サイズ：640×480
- フレームレート：30fps
- ファイル形式：WMV

なお用いた録画ソフトはWindows Movie Makerである。以上の方法で撮影した4つの動画の長さはそれぞれ以下のとおりである。

- 表情映像1：16分43秒 総フレーム数30090
- 表情映像2：12分46秒 総フレーム数22980
- 表情映像3：12分29秒 総フレーム数22470
- 表情映像4：11分53秒 総フレーム数21390

- 全体像撮影カメラ Sony HDR-CX560V
このカメラの主な仕様を以下に示す。

- サイズ：1440×1080
- フレームレート：30fps
- ファイル形式：MPEG4 AVC/H.264

次に、得られた表情映像に対して、著者の一人が目視による表情の種類と表情の程度の判定を行った。その後、4節で示した表情強度の推定を行い、目視による判定と強度の照らし合わせを行った。

5.2 表情の種類

表1に今回の実験で用いた表情の種類を示す。

今回の実験ではトランプを行なってもらったため、表出していた表情はほとんどが喜び(笑顔)と驚きの表情であった。そこで、目視で表情を確認した際の判断基準として表1に記述する表情を用いた。表1中では、「表情一覧」には表情の種類、「分類される表情」にはその表情に分類される代表的な表情、「特徴」にはその表情の判断基準を記述した。

表1 目視の表情確認の際に用いた判断基準

表情一覧	分類される表情	特徴
笑顔1	微笑み、苦笑	口はあまり開けず唇の両端が上がる
笑顔2	笑う	目尻が下がり、口を少し開けて笑い声をあげる。
笑顔3	高笑い、破顔	目尻が下がり口を大きく開け、大きな声や身振りなどを伴うような笑い
驚き1	驚嘆	目を見開き口をすぼめるような驚き
驚き2	驚嘆+半笑い	唇の両端が上がり、両まゆの間が挟まる
無表情		基本6感情(喜び、驚き、悲しみ、怒り、恐れ、嫌悪)のいずれとも言い難い表情
その他	怒り顔、泣き顔等	上記6つの表情のいずれにも属さない表情

5.3 実験の結果

以下、表2に実験で得られた結果の一部を示す。

表2 実験結果(表情映像3)

No	シーン時間	表情	平均	最大値	分散
1	0:37 ~ 0:39	笑顔1	2.66	4.98	2.89
2	0:46 ~ 0:53	笑顔2	2.23	3.86	1.19
3	1:25 ~ 1:33	笑顔1	1.95	6.35	2.54
4	1:42 ~ 1:54	笑顔1	2.95	10.09	7.68
5	2:09 ~ 2:12	笑顔2	4.31	9.78	7.44
6	2:16 ~ 2:22	笑顔2	2.59	6.1	3.28
7	2:27 ~ 2:36	笑顔2	3.49	8.45	6.4
8	2:41 ~ 2:44	笑顔2	3.95	6.26	3.12
9	2:51 ~ 2:54	笑顔2	3.13	6.3	3.12
10	3:30 ~ 3:31	笑顔2	3.66	4.77	1.24
11	3:40 ~ 3:44	笑顔1	3.62	8.93	5.72
12	3:54 ~ 3:58	笑顔2	4.21	7.58	9.85
13	4:09 ~ 4:11	笑顔2	1.41	9.45	0.9
14	4:19 ~ 4:22	笑顔1	1.23	3.22	0.96
15	4:35 ~ 4:37	笑顔1	1.39	1.39	0.87
16	5:07 ~ 5:08	笑顔1	1.45	2.81	0.83
17	5:13 ~ 5:14	その他	3.03	5.13	2.21
18	5:18 ~ 5:21	笑顔2	14.23	70.82	444.14
19	5:37 ~ 5:41	笑顔1	1.55	4.67	2.52
20	5:43 ~ 5:45	笑顔2	1.48	3.62	1.06
21	5:53 ~ 5:58	笑顔2	3.94	7.69	8.34
22	6:15 ~ 6:19	笑顔1	2.11	3.58	2.2
23	6:36 ~ 6:38	笑顔1	2.54	6.43	3.26
24	6:52 ~ 6:55	笑顔1	2.78	7.69	5.28
25	7:09 ~ 7:14	笑顔1	2.13	4.95	2
26	7:26 ~ 7:28	その他	3.27	6.42	4.06
27	7:42 ~ 7:43	驚き1	7.63	9.46	1.31
28	7:44 ~ 7:47	笑顔1	2.38	5.28	2.42
29	7:53 ~ 8:00	笑顔1	2.04	8.44	3.79
30	8:09 ~ 8:16	笑顔2	2.77	6.59	4.22

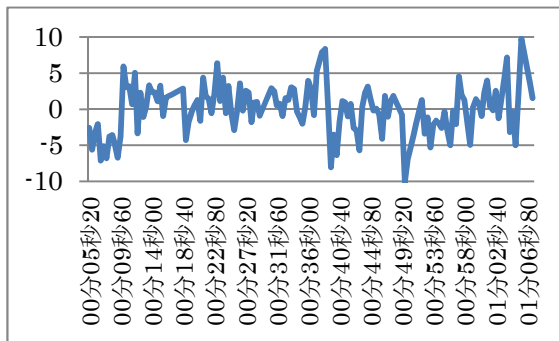
表 2 の列について、それぞれ、「No」はシーンごとに振り分けた番号、シーン時間はそのシーンの長さであり、「分:秒」の形式で示している。「表情」は表 1 にもとづいて目視で確認した表情、「平均」と「最大値」及び「分散」はそれぞれ該当シーンのフレームを 10 フレームごとに表情強度を算出しそのシーン内の平均値、最大値、分散を求めたものである。また、表 2 の中で被験者 3 が実際にアンケートで検索したいと回答したシーンは、No.7, No.9, No.18, No.27, No.28 の計 5 シーンであった。

5.3 先行研究の手法により得られた表情表出シーン

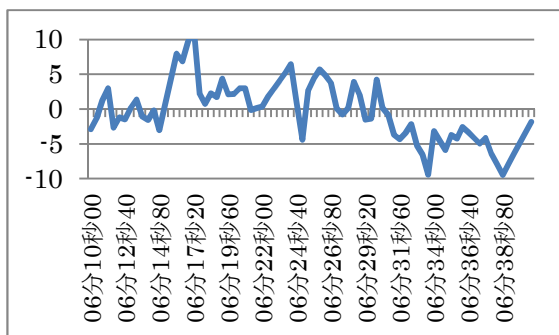
先行研究[5]の表情表出シーン検索を表情映像 3 に対して行った。表 4 に、表情表出シーン検索を行い「笑顔」と判定されたシーンの一覧を、図 2 にそれらのシーンにおける表情強度の推移を一部示す。図 2 の縦軸は表情強度、横軸は動画時間を示す。動画全体での表情強度の平均値は 0、最大値は 70.81、最小値は -11.43 であった。

表 3 表情映像 3 で笑顔のシーンを検索した結果

シーン	動画時間
A	0:05~1:07
B	1:26~1:50
C	2:54~2:54
D	3:48~4:26
E	4:28~4:46
F	4:50~5:08
G	5:15~5:35
H	6:10~6:40
I	6:55~7:34
J	8:26~9:00
K	9:13~9:52



シーン A 平均 0.1592 分散 11.48331



シーン H 平均 -0.3356 分散 18.32581

図 2 シーン A、H における表情強度の推移

5.4 先行研究との比較

著者らが目視で表情が表出していると判断したシーンの中で先行研究[5]の手法では検索ができなかったシーンとそのシーンの表情強度の平均をまとめたものを表 4 に示す。

「動画時間」は該当シーン時間、「強度平均」はそのシーンに含まれる各フレームの表情強度の平均を示している。表情強度の平均が映像の全フレームの表情強度の平均を上回ったシーンの「平均以上」の欄に丸を記述している。ちなみに、表情映像 1~4 全てにおいて、全フレームの表情強度の平均は 0 である。

表 4 先行研究の手法で取得できなかった表情表出シーンの表情強度平均

(a) 表情映像 1

動画時間	強度平均	平均以上
0:46~0:46	16.76	○
1:36~1:42	3.53	○
1:46~1:46	3.41	○
3:49~3:51	3.46	○
4:37~4:40	4.37	○
7:43~7:44	-0.82	

(b) 表情映像 2

動画時間	強度平均	平均以上
1:35~1:41	9.52	○
2:13~2:13	11.11	○
3:49~3:51	9.23	○
7:43~7:44	6.61	○
8:46~8:46	-3.51	

(c) 表情映像 3

動画時間	強度平均	平均以上
1:14~1:15	-1.07	
1:17~1:18	-1.49	
1:54~1:56	-4.86	
1:58~2:02	-0.75	
2:08~2:11	-0.43	
2:18~2:21	1.52	○
2:56~2:57	-0.17	
3:08~3:11	-1.77	
3:22~3:25	2.66	○
3:35~3:36	0.92	○
3:46~3:48	1.06	○
4:02~4:03	-0.55	
5:05~5:07	-0.61	
5:09~5:12	-1.06	
6:04~6:05	-1.30	
7:19~7:20	-1.81	
7:38~7:42	-2.26	

(d)表情映像 4

動画時間	強度平均	平均以上
0:35~0:36	7.53	○
0:39~0:40	10.76	○
2:00~2:03	7.15	○
2:50~2:52	9.91	○
3:01~3:04	8.52	○
3:16~3:18	6.95	○
6:08~6:09	-0.58	
6:12~6:12	5.53	○
8:35~8:37	8.14	○
8:44~8:46	10.58	○
9:15~9:22	6.37	○

また、表 5 に先行研究で取得できたシーンの強度を示す。表情映像 3 についてはシーン数が多いため表 2 と重複していないシーンから抜粋して掲載する。

表 5 先行研究の手法で取得できたシーンの表情強度

(a) 表情映像 1

表情映像 1	表情強度平均
2:13~2:14	-0.97
2:16~2:17	-1.59
9:34~9:36	17.92

(b) 表情映像 2

表情映像 2	表情強度平均
0:49~0:52	5.64
2:13~2:13	12.28
4:37~4:41	5.41
9:34~9:36	7.83

(c) 表情映像 3

表情映像 3	表情強度平均
0:52~0:57	1.47
1:35~1:37	3.6
6:37~6:41	1.43
8:44~8:46	3.24

(d) 表情映像 4

表情映像 4	表情強度平均
0:18~0:23	9.19
1:07~1:09	12.78
1:41~1:46	15.12
3:51~3:52	11.79
5:41~5:44	9.26
6:48~6:50	14.02

6. 考察

6.1 映像全体の表情強度に対する考察

まず、感情が強く表出しているシーンを検索できるかを考える。動画全体での強度の平均値は 0 であったので、これを検索時の閾値にすると目視で選んだ笑顔 1 の（あまり表情が表出していないと思われる）19 シーン 中 14 シーン

は閾値に届いていないため切り捨てることができる。逆に笑顔 2、笑顔 3 と判定されている 14 シーンの中でこの閾値を下回っているものはわずか 1 シーンであった。このことから表情強度の平均値を用いることで感情表出が強いシーンのある程度抽出することが可能である。

ただし、強度は図 2 のように局所的に高い数値が示され、長時間にわたって強度が高い状態が維持されるわけではないので、7:42~7:43 のようなごく短いシーンでは平均が高く、分散が小さい理想的な値になるのだが、シーンが長くなればなるほど平均値は下がる傾向がある。そのためシーンの時間が長くなると、誤検出を抑えるために、表情表出区間内のピーク時のみを見るなどする必要があると考えられる。

次に、表情強度の最大値を閾値とする場合は、最大値が仮に 8 に届いていない場合を削るとすると、笑顔 1 である弱い笑顔のシーン 19 シーン 中 15 シーンを削ることができる。また笑顔 2,3 以上のそれなりに強い笑顔を示しているシーンも 14 シーン 中 1 シーンを除いて最大値が 8 を上回っている。しかし、最大値はたった 1 フレームしか見ておらず、そのフレームがたまたま顔の特徴点があまく取れずに偶然高い値を示してしまう場合があるため、シーン検索の閾値として用いるには平均値がより適していると考えられる。実際に 5:18~5:21 に異常に高い値が得られているが、顔をクシャッとさせて笑っているシーンであった。顔を大きく変えるような表情があった場合は特徴点があまく取得できず強度が跳ね上がる可能性がある。このため表情強度が他と比べて高い値を示す場合はそのフレームで正しい特徴量が取得できているかどうか吟味する必要がある。

6.2 先行研究の手法で得られた表情表出シーンの表情強度に対する考察

前節でも述べたように表情強度自体が高い値で止まることはほとんどなく、一瞬高い値を出しても大半は数秒以内に下落するケースがほとんどである。今回の実験ではシーン H,K は前半にピークが来て若干右肩下がりになるパターンを示した。逆に、シーン A のように全体的に平均値前後で推移する場合、もしくは、やや右肩上がりの傾向が見られたのは、シーン A,B,C,D,E,F,G,I,J の 9 シーンであった。中央にピーク値が来るパターンは見られなかった。前者のパターンは最初に大きな笑いや驚きを示してその後普通に笑うようなシーン、後者は終始ニコニコ笑いながらしゃべるようなシーンが多く見られた。このことからどちらのパターンを好むかをユーザに選択してもらえれば検索に活用できると考えられる。

6.3 表情強度と既存手法の比較の考察

5.4 節で示した表 4 より、表情映像 1 と表情映像 2 と表情映像 4 では、既存手法では検索できなかった 22 シーンのうち 19 シーンで表情強度平均が 0 を超えており、表情強度を利用することで既存手法よりも大幅に検索漏れを減少させることができる事がわかる。また表 5 で示した通り先行研究の手法で取得できたシーンについても表情強度平均が 0 を超えているシーンが大半であり問題なく取得できると考えられる。

また、表情映像 3 については 6.1 節や 6.2 節で示した通り、既存手法でも取得できるシーンに関しては十分検索可能ではあるが、表 4 で示した既存手法の検索漏れシーンについては 18 シーン 中 4 シーンしか正の値を取っておらず

大きな改善は見られなかった。残りの 14 シーンの特徴として、顔のパーツで笑いを示すよりも、肩を震わせるような仕草や顔の動きなどで視ている人間が笑顔と感じるようなシーンであったので、こういった仕草で喜びを示すシーンに対しても対応できるように特徴量を改善する必要がある。例えば隣接するフレーム同士で特徴点の移動を追い、どれだけ特徴点が動いたかなどを新たな特徴量として扱うなどが考えられる。

7. まとめ

本論文では、ライフログ映像から効率的に重要なシーンを検索、提示することを目的に視聴者が視聴したいと感じるシーンの検索に用いることが出来る表情強度の提案を行った。実際に被験者の表情の映像を取得し、表情強度の測定を行った。その結果として、表情強度を用いることで表情の程度を推定することが出来た。

今後の課題としては、今回表情強度に用いた特徴量の幾つかは、顔が向いている方向が変わることに対して脆弱であると考えられるため、よりそういった外乱に対して頑健な特徴量に改良することがあげられる。また、今回はあくまで 1 人のユーザが自分一人の表情を使うことしか想定していないが、ホームビデオのような映像では他の人物が映っている場合も十分考えられるため、そういった複数人の表情に対して表情強度を算出し、それら全員の強度を考慮した上で検索精度を向上させたいと考えている。

8. 参考文献

- [1] 新保史生, "ライフログの定義と法的責任 個人の行動履歴を営利目的で利用することの妥当性", 情報管理, vol.53, No.6, pp.295-310, 2010.
- [2] 相澤清晴, "体験の情報処理 -ライフログの取得と処理-", 電子情報通信学会技術研究報告, PRMU, パターン認識・メディア理解 103(738), 1-9, pp.284-290, 2004.
- [3] YouTube, <http://www.youtube.com/>.
- [4] Shogo Shimura, Yasushi Hirano, Shoji Kajita and Kenji Mase, "Experiment of Recalling Emotions in Wearable Experience Recordings", Advances in Pervasive Computing: Adjunct Proceedings of the Third International Conference on Pervasive Computing, pp.19-22, 2005
- [5] 野宮浩揮, 森國淳司, 宝珍輝尚, "ライフログ映像検索に向けた顔特徴点の位置関係に基づく効率的な表情表出シーン検出", 電子情報通信学会論文誌, Vol.J96-D, No.5, pp.1279-1289, 2013.
- [6] 木村聡, "表情認識とその程度推定", 情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア] 97(10), pp.99-106, 1997.
- [7] 松野 勝弘, 李 七雨, 辻 三郎, "ポテンシャルネットと KL 展開を用いた顔表情の認識", 進学会論文誌 D II, Nc8, pp.1591-1600, 1994
- [8] Luxand, Inc., "Luxand FaceSDK4.0," <http://www.luxan.com/facesdk>.