

表情強度を用いたライフログ映像からの表情表出シーン検索

Emotional Scene Retrieval from a Lifelog Video Using Facial Expression Strength

森國 淳司, 野宮 浩揮, 宝珍 輝尚

Atsushi Morikuni, Hiroki Nomiya, Teruhisa Hochin

京都工芸繊維大学 情報工学専攻, 京都市左京区御所海道町

Kyoto Institute of Technology, Gosyokaido-cho, Sakyou-ku, Kyoto

あらまし:ライフログ映像データベースから重要なシーンを検索・提示することを目的として、表情認識に基づくシーン検索法についての検討を行う。従来では人物の表情の種類を識別することで表情表出シーンの検索が可能であったが、表情がどれだけ強く表出されているかは考慮されておらず、シーンの重要度の判断が困難であった。そこで、表情強度という表情表出の程度を表す尺度を導入し、よりの確に重要なシーンを検索できる手法を提案する。

Summary: For the purpose of retrieving the emotional scenes from a lifelog video database, and showing them, we perform the examination about the scene retrieval method based on the recognition of facial expressions. Conventionally, the retrieval of the emotional scenes has been performed by distinguishing the kind of the emotion of the person. However, the judgment of the importance of the scene is difficult without the consideration of the strength of the emotion. Therefore, we propose expression strength and the technique that can search the important scenes more precisely.

キーワード:ライフログ, 映像検索, 表情認識

Keywords: Lifelog, video retrieval, facial expression recognition

1. はじめに

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[1]。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[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に基づい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ている[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」の中の「微笑み」から「高笑い」までというような感情の程度が異なるものが考慮されていない。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。

そこで本論文では、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て、被験者の表情の映像を取得し、表情強度の測定を行い、表情強度を用いることで表情の程度を推定する。

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章では本研究で扱う顔の特徴点について説明する。その後4章で表情強度について述べ、5章で表情強度を用いたライフログ映像検索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、俳優に作為的に演じてもらった表情と比べて、素人の曖昧な表情は認識結果にノイズとしてより大きい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

本研究では、顔特徴点を用いて表情強度を推定する。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

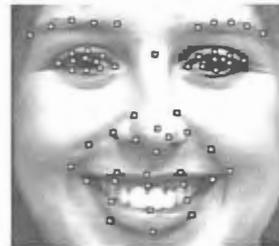


図1 顔特徴点抽出例（四角い点が顔特徴点を表している）

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。ただし、左目と右目の距離で割って正規化を行う。

・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口の縦の長さ m_v 、最も左と右にある2点を結んだ線分を口の横の長さ 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口の縦の長さ m_v 、最も左と右にある2点を結んだ線分を口の横の長さ 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。

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

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言える。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。ここで、 i はフレーム番号を表す。

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、 $s_i = \sum_{j=1}^{\lfloor N/2 \rfloor} \sum_{k=1}^{10} f_{r_{ij}k} - \sum_{j=\lfloor N/2 \rfloor + 1}^N \sum_{k=1}^{10} f_{r_{ij}k}$ (2) である。 N は映像に含まれるフレームの総数、 r_{ij} は $p_{ij} \sim p_{Nj}$ 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、

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{otherwise} \end{cases} \quad (3)$$

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 $(\text{sgn}(s_j) = 0)$ 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 実験方法

被験者は著者の一人を含む男子学生 8 名 (21 才～24 才) である。8 名 (以降、被験者 A, B, C, D, E, F, G, H と表記する。) を 4 名 × 2 グループに分け、すべての被験者の前に顔が撮影できるように 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ハーツ、大富豪をそれぞれ 2 度ずつ行なってもらい、計 4 度の撮影を行い 8 × 4 = 32 個の表情映像データを得た。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。

・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。目視による表情判断基準を表 1 に示す。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。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 実験の結果

表 1 目視時の表情判断基準表

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。
無表情		基本 6 感情(喜び、驚き、悲しみ、怒り、恐れ、嫌悪)以外の表情

表 2 表情強度平均例 (ハーツ 3E)

表出表情	表出シーン	表情強度平均
		0.643
笑顔 1	0:54 ~ 1:04	-0.481
		-0.209
笑顔 2	2:00 ~ 2:07	0.410
笑顔 1	2:08 ~ 2:15	1.780
		-0.283
笑顔 1	2:30 ~ 2:37	0.062
		-0.027
驚き 2	2:52 ~ 2:55	0.823
		-0.115
笑顔 1	3:06 ~ 3:08	0.012
		-1.330
笑顔 1	3:44 ~ 3:46	-1.381
		-0.435
笑顔 1	4:06 ~ 4:12	-0.333
		0.210
笑顔 2	4:26 ~ 4:29	2.060
		-0.570
笑顔 1	4:38 ~ 4:45	0.795
		1.316
笑顔 1	4:50 ~ 4:52	-1.723
笑顔 2	4:58 ~ 5:06	1.786
		-0.070
笑顔 2	5:27 ~ 5:30	0.206
		0.037
笑顔 1	8:48 ~ 8:49	-0.170
		0.198
笑顔 1	9:30 ~ 9:35	0.479
		0.156
笑顔 1	10:01 ~ 10:07	0.323
		0.555
笑顔 1	10:10 ~ 10:13	0.096
		0.117
笑顔 1	10:40 ~ 10:42	0.391
		-0.001
笑顔 2	10:52 ~ 11:02	-0.033
		-0.173
笑顔 2	11:52 ~ 11:58	-0.207
		-0.414
笑顔 1	12:12 ~ 12:20	-0.228
笑顔 2	12:20 ~ 12:25	-0.033
		-0.172
笑顔 1	12:42 ~ 12:45	0.094
笑顔 2	12:52 ~ 12:58	-0.355

表 2 に実験で得られた結果の一部を示す。

表 2 は 3 回目にはハーツを行った回における被験者 E の表出表情と表情強度平均を示したものである。ただし、表出表情は筆者の一人が表 1 に示す基準を用いて、目視で判断したものである。また、「表出シーン」は時間(分:秒)を示し、表情強度平均はそのシーン内の表情強度の平均を示す。

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)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 3 に 8 人の被験者の映像の中で強度の傾向が変わった 2 フレーム、計 16 フレームを示した。32 個の映像中から被験者 8 人それぞれから特に顕著な傾向が見られた動画を 1 つずつ選択している。

左端の列「ハ 1A」はハーツ 1 回目被験者 A のフレームを指す。「大 4E」は大富豪 4 回目の被験者 E のフレームを指す。表中の薄く色を塗っている f6, f7, f9 は、主成分分析時に第一主成分の主成分負荷量の絶対値が高かった上位 3 つの特徴量であり、濃い色を塗っている f1 (被験者 G のみ f8), f2, f3 は第 2 主成分の主成分負荷量の絶対値が高かった上位 3 つの特徴量である。

表 3 各動画において特徴量の傾向が大きく変化した 2 フレームの各特徴量

	F 番号	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
ハ 1A	6060	-0.090	0.345	0.196	0.061	0.310	0.184	0.021	0.336	0.031	0.045
	6070	-0.031	0.378	0.190	0.047	0.319	0.192	0.034	0.346	0.074	-0.041
ハ 2B	5470	-0.174	0.328	0.185	0.049	0.303	0.151	0.020	0.370	0.081	0.105
	5480	-0.270	0.304	0.197	0.045	0.284	0.191	0.045	0.508	0.170	0.000
ハ 1C	4610	-0.106	0.327	0.188	0.048	0.336	0.245	0.104	0.533	0.308	0.214
	4620	-0.221	0.346	0.225	0.044	0.306	0.171	0.045	0.389	0.090	-0.056
ハ 1D	14570	-0.177	0.317	0.196	0.038	0.299	0.203	0.087	0.491	0.269	0.303
	14580	-0.055	0.320	0.180	0.038	0.292	0.250	0.110	0.472	0.219	0.313
大 4E	11570	-0.273	0.306	0.188	0.030	0.204	0.203	0.078	0.326	0.135	0.450
	11580	-0.221	0.307	0.164	0.026	0.173	0.254	0.093	0.384	0.138	0.615
ハ 4F	10470	-0.134	0.325	0.203	0.036	0.296	0.182	0.028	0.361	0.078	-0.200
	10480	-0.081	0.328	0.198	0.035	0.269	0.166	0.016	0.377	0.020	-0.150
ハ 3G	14280	-0.140	0.290	0.183	0.044	0.312	0.152	0.019	0.433	0.077	0.211
	14290	-0.136	0.310	0.182	0.039	0.273	0.156	0.012	0.432	0.026	0.111
ハ 3H	2360	-0.122	0.352	0.208	0.050	0.303	0.254	0.064	0.464	0.159	0.040
	2370	-0.143	0.355	0.212	0.047	0.333	0.262	0.094	0.576	0.254	0.133

表 4 第 1 主成分 主成分負荷量

第1主成分	ハ 1A	ハ 2B	ハ 1C	ハ 1D	大 4E	ハ 4F	ハ 3G	ハ 3H
f1	0.080	-0.008	-0.355	0.047	0.051	0.357	0.036	-0.136
f2	-0.012	-0.322	0.241	0.015	0.433	-0.395	0.115	-0.053
f3	0.051	-0.293	0.439	-0.041	0.247	-0.368	0.015	-0.080
f4	-0.231	-0.425	0.325	-0.166	-0.324	0.340	-0.377	0.447
f5	0.103	-0.613	0.191	-0.342	-0.444	0.380	-0.488	0.006
f6	-0.914	0.893	0.895	0.923	0.910	-0.819	0.867	0.951
f7	-0.945	0.936	0.929	0.960	0.943	-0.828	0.934	0.969
f8	-0.556	0.187	0.375	0.859	0.516	-0.295	0.124	0.509
f9	-0.915	0.849	0.808	0.942	0.920	-0.757	0.874	0.927
f10	-0.518	0.651	0.504	-0.032	0.586	-0.049	0.412	0.664
固有値	3.217	3.598	3.229	3.548	3.727	2.697	2.970	3.632
寄与率	32.168	35.976	32.290	35.475	37.271	26.970	29.696	36.317
累積寄与率	32.168	35.976	32.290	35.475	37.271	26.970	29.696	36.317

表 5 第 2 主成分 主成分負荷量

第2主成分	ハ 1A	ハ 2B	ハ 1C	ハ 1D	大 4E	ハ 4F	ハ 3G	ハ 3H
f1	0.340	0.496	-0.099	0.585	0.585	0.567	0.672	0.155
f2	-0.841	-0.763	-0.884	-0.813	-0.690	-0.709	-0.680	-0.832
f3	-0.880	-0.891	-0.759	-0.929	-0.896	-0.762	-0.928	-0.872
f4	-0.603	0.126	0.027	-0.155	-0.454	0.464	-0.086	-0.598
f5	-0.367	0.144	0.270	-0.212	-0.438	0.209	-0.084	-0.505
f6	-0.118	-0.177	0.052	-0.046	-0.056	0.258	-0.217	-0.020
f7	0.074	-0.143	0.061	-0.015	0.046	0.472	0.008	0.044
f8	0.012	-0.155	0.597	-0.061	-0.030	0.063	-0.076	0.056
f9	0.111	-0.199	0.314	-0.031	0.085	0.468	0.053	0.120
f10	0.045	0.198	-0.258	0.465	0.162	0.513	0.335	0.020
固有値	2.129	1.813	1.969	2.158	2.059	2.438	1.958	2.109
寄与率	21.295	18.130	19.693	21.585	20.587	24.381	19.576	21.094
累積寄与率	53.463	54.106	51.983	57.060	57.858	51.352	49.272	57.411

表 6 第 3 主成分 主成分負荷量

第 3 主成分	ハ 1A	ハ 2B	ハ 1C	ハ 1D	大 4E	ハ 4F	ハ 3G	ハ 3H
f1	0.712	-0.354	-0.030	-0.130	-0.292	0.068	0.190	-0.441
f2	0.033	-0.117	0.046	0.079	0.285	0.302	0.060	0.303
f3	-0.206	-0.068	0.183	0.139	0.286	0.352	-0.032	0.358
f4	0.505	-0.676	0.817	-0.856	-0.634	0.642	0.712	-0.425
f5	0.308	-0.574	0.858	-0.768	-0.648	0.758	0.599	-0.747
f6	-0.060	-0.216	-0.139	-0.056	-0.278	0.153	0.204	0.061
f7	0.002	-0.179	-0.159	-0.078	-0.152	0.002	0.061	0.048
f8	-0.416	-0.513	-0.100	-0.168	-0.546	0.497	0.619	-0.110
f9	-0.048	-0.273	-0.157	-0.134	-0.154	0.023	0.163	0.019
f10	0.558	-0.019	-0.187	-0.095	0.252	-0.186	0.229	0.171
固有値	1.390	1.345	1.554	1.429	1.555	1.511	1.414	1.201
寄与率	13.904	13.453	15.540	14.291	15.550	15.108	14.136	12.008
累積寄与率	67.367	67.559	67.523	71.351	73.408	66.459	63.407	69.419

また、主成分分析を行い得られた第 1～第 3 主成分の主成分負荷量を表 4～表 6 に記述する。一番上の行におけるハ 1A は、1 回目のハーツにおける被験者 A の表情映像を、大 4E は 4 回目の大富豪における被験者 E の表情映像を意味する。

6. 考察

(a)表 1 の全シーンについて

笑顔 2, 笑顔 1, 無表情のそれぞれをすべてのシーンで表情強度の平均を取った結果を、表 7 に示す。

表 7 被験者 E～H の各表情の全シーンにおける強度平均

映像	笑顔 2	笑顔 1	無表情
ハーツ 3E	0.583	0.045	-0.040
ハーツ 3F	0.008	-0.104	-0.021
ハーツ 3G	1.230	0.317	0.544
ハーツ 3H	0.944	-0.298	0.227

平均値を見ると確かに、大きく笑っている笑顔 2 は笑顔 1 や無表情に比べて高く、強く表情が表出しているシーンに関しては表情強度が高い値をとることが分かる。しかし、笑顔 1 と無表情に関しては、無表情のほうが笑顔 1 より高いことも多く、表情表出の度合いが小さい時の精度は改善の余地がある。

(b)表 4～6 について

第一主成分に関しては、どの被験者に関しても、f6, f7, f9 の主成分負荷量が高く、全被験者について第一主成分は口の概形の成分といえる。

第二主成分に関しては、被験者 C 以外は f1, f2, f3 が強く、C に関しても f2, f3 の負荷量が高いため、全被験者について眉間の概形の成分と考えられる。

第三主成分に関しては、f4, f5 が全被験者について強く、目の概形の成分であると考えられる。

以上から、どの被験者についても主成分分析を行った後の傾向が似通っていると言え、表情の変化には口、目、眉間の概形が大きな影響を及ぼしている。直感的にも、人間が表情の変化を視覚的に感知する時は口や目を特に参考にすると考えられるので、概ね妥当な結果と考えられる。

(c)被験者がアンケートで実際に検索したいと回答したシーン

表 8 に、大富豪 4 回目において、被験者が実験後にアンケート回答した「検索したいシーン」の表情強度の平均を示す。その他は被験者が選択しなかったシーンすべての強度平均を意味する。

表 8 被験者がアンケートで検索したいと回答したシーン

(a)大富豪 4E

欲しいシーン	強度平均
1:13-1:22	4.905
2:02-2:07	5.302
3:08-3:12	0.185
6:22-6:27	0.095
7:02-7:12	-5.116
8:46-8:56	0.471
9:22-9:30	-2.330
その他	-0.038

(b)大富豪 4F

欲しいシーン	強度平均
0:45-0:55	3.546
1:25-2:10	5.269
6:08-6:15	-3.951
7:50-8:00	-4.079
8:10-8:15	-1.073
その他	-0.363

(c)大富豪 4G

欲しいシーン	強度平均
0:13-0:30	1.412
0:58-1:10	-0.992
1:10-1:30	-1.067
2:01-2:13	1.481
3:00-3:17	-1.839
4:36-4:46	-1.361
6:49-6:57	0.861
その他	-0.039

(d)大富豪 4H

欲しいシーン	強度平均
2:55-3:10	1.530
9:08-9:28	0.405
9:30-9:35	0.237
その他	-0.062

この結果から、被験者が検索したいと回答した全 22 シーン 中 13 シーンはその他のシーンよりも平均が高かった。このことから検索したいシーンの 6 割程度は、強

度平均によって検索できる可能性があると考えられる。また検索できなかったシーンでは、考察(b)に示したように表情表出度合いが小さいシーンが多く、うまく強度では判断ができなかった。この点については今後改善していく予定である。

7. まとめ・今後の課題

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。その結果、表情表出シーンならびに被験者が検索したいと感じるシーンのある程度検出することが出来た。

今後の課題としては、表情強度に改善の余地が見られたので、表情強度の推定に使用する特徴量の改善を考えている。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] 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), pp.284-290, 2004.
- [2] 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.
- [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", 電気情報通信学会論文誌 D II, No.8, pp.1591-1600, 1994
- [8] Luxand, Inc., "Luxand FaceSDK4.0, "http://www.luxan.com/facesdk.