

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 5日現在

機関番号：14303

研究種目：若手研究（B）

研究期間：2010～2011

課題番号：22700098

研究課題名（和文）感情推定に基づくライフログからの印象的なシーンの検索に関する研究

研究課題名（英文）Impressive Scene Retrieval for Lifelog Videos based on Emotion Estimation

研究代表者

野宮 浩揮（NOMIYA HIROKI）

京都工芸繊維大学・工芸科学研究科・助教

研究者番号：80533116

研究成果の概要（和文）：

ライフログの主な構成要素である映像データから、映像中の人物の表情を認識し、印象的であると考えられる、感情（表情）が表出しているシーンを検索するシステムを構築した。さらに、ライフログを想定して、実際に日常生活の様子を収めた映像と表情表出シーンの情報を収めた検索用データベースを作成して、特に印象的なシーンになりやすいと考えられる、笑顔が表出しているシーン検索の評価実験を行った。本評価実験では、9割以上の適合率での表情表出シーン検索を行うことができた。

研究成果の概要（英文）：

In order to retrieve impressive scenes from lifelog videos, a retrieval system of emotional scenes, which is considered to be impressive, has been developed in this research. The retrieval system detects emotional scenes based on facial expression recognition. In addition, a lifelog video database including video sequences of daily life and the information about emotional scenes contained in the video sequences has been constructed. By using the database, emotional scene retrieval experiments were conducted to evaluate the retrieval performance. Since the video scenes including smile and/or laughter will be impressive, the retrieval experiments are focused on retrieving such video scenes. The precision of the scene retrieval in this experiment was more than 90%.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,100,000	330,000	1,430,000
2011年度	800,000	240,000	1,040,000
年度			
年度			
年度			
総計	1,900,000	570,000	2,470,000

研究分野：マルチメディアデータベース・機械学習

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：ライフログ、マルチメディア検索、表情認識、機械学習

## 1. 研究開始当初の背景

近年における映像や音声の記録装置の高性能化や低価格化に伴い、個人の行動や体験の履歴を映像や音声などのマルチメディアデータとして記録し、記憶の補助や行動パターン分析などに利用することを目的としたライフログが注目されている。しかし、ライフログでは、生活の様子の大部分を記録に残すため、特に映像のようにデータが大きくなりやすい記録のデータ量は膨大になり、その中から重要な出来事を検索することは非常に難しい。また、ライフログは個人の生活の様子を収めたデータであるという性質上、個人で所有することが多いため、大量のライフログからの検索を行う際には、利用者（ライフログデータの所有者）の記憶に依存せざるを得ない状況にあった。このことは、貴重な情報を含むライフログデータが活用されることなく退蔵されるという問題を引き起こしており、このような背景から、効果的なライフログデータの利用法の確立が急務となっていた。

## 2. 研究の目的

本研究では、ライフログの中でも特によく用いられ、またデータ量が大きくなりやすいライフログ映像を対象として、その中から利用者の記憶に頼ることなく、自動的に有用と考えられる印象的な出来事が起こったシーンを検索・提示することを目的としている。印象的な出来事が起きているシーンには、多くの場合人物が含まれており、またその人物に何らかの感情が表出していると考えられる。また、感情は表情に現れることが多いため、本研究では、以下に示す手順により、ライフログ映像の中から人物に何らかの表情が表出しているシーンを検索するシステムを構築することを目的とする。

### (1) 表情認識による感情推定手法の構築

ライフログ映像に含まれる人物の表情の認識を通じて、感情の推定を行う。映像は連続した静止画像により表現されるため、映像から各フレームの画像（静止画像）を抽出し、それらの各画像から、表情表出に関連する顔面上の点の動きを解析することにより、表出している表情を認識する手法を確立する。

### (2) ライフログ映像からの印象的なシーン検索システムの構築

(1)での各フレーム画像の表情認識結果を用いて、ライフログ映像中から感情（表情）が表出しているシーンを検索する手法を確立する。また、ライフログの規模が大きくなっても高速に検索が行えるよう、効率的な検索システムの構築を図る。

### (3) ライフログ映像データベースの作成

ライフログ映像検索の性能評価のため、日常生活の様子を記録したライフログ映像データベースを作成する。なお、ライフログ映像は、自然に表出した感情（表情）が含まれるように作成する。

### (4) ライフログ映像を用いたシーン検索性能の評価

(3)で作成した映像データベースを用いて、表情表出シーン検索性能の評価を検索精度および検索速度の点から行う。

## 3. 研究の方法

### (1) 静止画像からの表情認識

#### ① 顔特徴点に基づく特徴量の生成

ライフログ映像に含まれる各フレーム画像から、図1に示す、表情表出に深く関与すると考えられる顔面上の点（以後「顔特徴点」と呼ぶ）の集合を抽出し、表情の認識に用いる。なお、これらは、眉の周辺6点、目の周辺6点、鼻の周辺4点および口の周辺8点の計24点の顔特徴点から構成される。また、これらの顔特徴点は既存の顔特徴点抽出ソフトウェア(Luxand FaceSDK 2.0)を用いて抽出する。

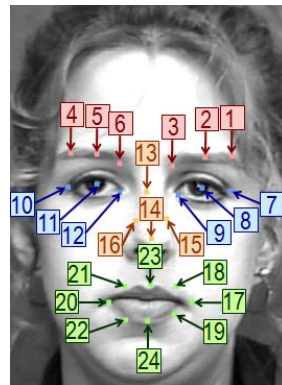


図1 表情認識に使用する顔特徴点

上記の顔特徴点を2点または3点組み合わせることにより、以下に示す6タイプの特徴量を定める。

1. 2つの顔特徴点間の距離
2. 2つの顔特徴点を結ぶ線分の傾き（角度）
3. 2つの顔特徴点を対角線とする矩形領域内に含まれる画素の輝度平均値
4. 3つの顔特徴点から構成される三角形の面積
5. 3つの顔特徴点から構成される三角形の内角
6. 3つの顔特徴点から構成される三角形内

に含まれる画素の輝度平均値

無表情の画像と表情が表出している画像から上記の各特徴量を求め、その差分を表情認識に用いる。これらは全ての可能な顔特徴点の組み合わせについて定めるため、特徴量の総数は11500種類となる。なお、これらの特徴量はいずれも顔特徴点および顔画像から高速に求めることができるため、効率的な表情認識が可能となっている。

## ② 特徴選択

特徴量は非常に数が多く、その中には表情の認識に寄与しないものも多く含まれるため、有用な特徴量のみを選択的に用いる。特徴量の選択基準として、各特徴量の有用度を定義し、有用度の高い特徴量から順に一定数の特徴量を選択する。有用度は、特徴量の分散比（級間分散と級内分散の比）と定める。なお、選択する特徴量の数は、実験的に1500としている。

## ③ 表情認識モデルの構築

選択された特徴量を並べて特徴ベクトル（1500次元）を生成し、これを用いて機械学習により表情認識モデルを構築する。表情認識問題は認識対象の表情の種類が多くなるほど難しくなると考えられることから、表情ごとに特徴ベクトルを生成し、認識対象と同数の表情認識モデルを構成する。表情認識モデルは Support Vector Machine により構成し、投票法を用いて複数の表情認識モデルの認識結果を統合することにより、最終的な認識結果を決定する。

### (2) 映像中の表情表出シーンの特定

#### ① 瞬きの検出

瞬きをしている顔画像に対しては、表情の認識を誤りやすくなるため、瞬きをしていると判断されたフレームは、表情表出シーンの検出には用いないこととする。

瞬きの検出は、まず目周辺の顔特徴点の位置から両目の領域（矩形）を定め、各領域内から円形のエッジを抽出することにより行う。目が開いている時は、瞳の部分に円形のエッジが強く検出されるため、検出されたエッジが弱い場合は、瞬きをしていると判断する。

#### ② 表情表出シーンの検出

表情が表出しているシーンを検出するため、映像中から表情の表出が開始したフレームと表情の表出が終了したフレームを特定する。

まず、各フレームの画像を順番に見ていき、次の2つの条件をみたすフレーム（ $P$ 番目のフレームとする）を表情表出開始フレームと

定める。

- ( $P-3$ )番目から( $P+3$ )番目のフレームまでの間に、表情が表出しているフレームが4フレーム以上ある。
- $P$ 番目から( $P-1$ )番目のフレームまでの間に表情表出開始フレームがない。ただし、 $P$ は最初のフレームまたは直近の表情表出終了フレームの番号であり、 $P' < P$ とする。

次に、表情表出開始フレームから順に各フレームの画像を順番に見ていき、次の2つの条件をみたすフレーム（ $Q$ 番目のフレームとする）を表情表出終了フレームと定める。

- ( $Q-3$ )番目から( $Q+3$ )番目のフレームまでの間に、表情が表出していないフレームが4フレーム以上ある。
- $Q$ 番目から( $Q-1$ )番目のフレームまでの間に表情表出終了フレームがない。ただし、 $Q$ は直近の表情表出開始フレームの番号であり、 $Q' < Q$ とする。

ここで、表情表出開始・終了フレーム検出の例を図2に示す。図2のグラフの横軸はフレーム番号、縦軸は各フレームでの表情認識結果（Angerは怒り、Neutralは無表情を表す）を示している。

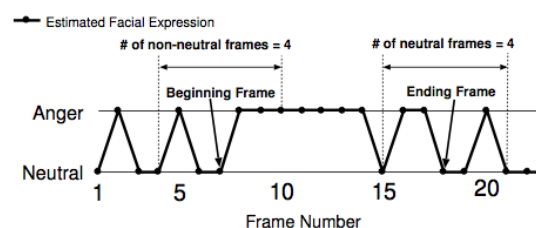


図2 表情表出開始・終了フレームの例

図2では、前述の条件を満たす $P$ が7であり、 $Q$ が18であるため、7番目のフレームが表情表出開始フレーム、18番目のフレームが表情表出終了フレームとなっている。

最終的に、表情表出開始フレームと表情表出終了フレームの間を表情表出シーンと定め、表情表出シーン内のフレームで最も多く表出している表情をそのシーンで表出している表情と定めることにより、表情表出シーンを検出する。

## 4. 研究成果

### (1) 表情表出シーン検出精度

#### ① 既存のデータセットに対する性能評価

提案手法の表情表出シーン検出性能を評価するため、まず表情認識手法の評価に広く用いられている MMI Facial Expression Database を用いて表情表出シーン検出実験を行った。本データセットには、怒り、嫌悪、恐怖、幸福、悲しみ、驚きの6種類の表情が

含まれている。これらの表情のうち、幸福の表情については、およそ 68% と最も高い検出精度（適合率）が得られた。幸福の表情が表出しているシーンは印象的なシーン検索の対象になりやすいと考えられるため、本実験結果は、提案手法の有効性を示すものであるといえる。一方、恐怖の表情は、悲しみや驚きと類似した顔面上の動きが見られたため、これらに誤認識されやすく、検出精度は 22% 程度であった。そのため、今後は表情表出時の顔面上の変化をより正確に捉えられるよう、表情認識手法の改善を行うことを検討している。

## ② ライフログ映像に対する性能評価

次に、ライフログ映像を想定して、日常生活の様子を記録した映像から表情表出シーンを検出する実験を行った。本実験では、被験者（男子大学生 1 名）が被験者を含む 2 名でトランプをしている様子を撮影した、およそ 36 分間の映像を使用しており、検出する対象は笑顔が表出しているシーンとしている。

実験の結果、67 の表情表出シーンが検出され、それらのうち 63 のシーンで実際に笑顔が表出していた。したがって、検出精度（適合率）はおよそ 94% であり、非常に高い精度が得られているといえる。

### (2) 表情表出シーン検出速度

表情表出シーン検出の速度を評価するため、前述のライフログ映像から表情表出シーンを検出する際に要した時間を計測した。その結果、CPU が Xeon W3580(3.33GHz)、メモリが 8GB の計算機でのシーン検出におよそ 52 秒を要した（但し、顔特徴点抽出には既存のソフトウェアを用いているため、顔特徴点抽出に要した時間は除外している）。すなわち、映像 1 分間あたりおよそ 1.5 秒と短時間で表情表出シーンの検出ができるため、規模の大きいライフログ映像データベースにも適用可能と考えられる。

また、予めライフログ映像に対して表情表出シーン検出を行い、その結果をデータベースに保存しておくことにより、検索時に要する時間をさらに低減することも可能である。

### (3) 本研究の意義

既存の表情認識に関する研究における評価実験では、多くの場合、意図的に表出された表情のみを含む映像データを対象としているのに対し、本研究では、自然に表出された表情を含むライフログ映像を対象として、特に検索対象となりやすいと考えられる、笑顔の表出シーンを精度よく、かつ効率的に検出できている。したがって、本研究は、ライフログ映像をはじめとした、人物を主体とす

る様々な映像に対する印象的なシーンの検索に対して有用であるといえる。

## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 2 件）

- ① 野宮浩揮、宝珍輝尚、アンサンブル学習を用いた効率的な映像からの表情表出シーン検出、電子情報通信学会論文誌、査読有、Vol.J95-D、2012、pp.193-205
- ② 野宮浩揮、宝珍輝尚、顔特徴量の有用性推定に基づく特徴抽出による表情認識、日本知能情報ファジィ学会誌、査読有、Vol.23、No.2、2011、pp.46-61

〔学会発表〕（計 7 件）

- ① Hiroki Nomiya、Emotional Video Scene Detection from Lifelog Videos using Facial Feature Selection、4<sup>th</sup> International Conference on Applied Human Factors and Ergonomics、2012 年 7 月 25 日（発表確定）、Hilton San Francisco Union Square（アメリカ合衆国）
- ② 野宮浩揮、表情表出シーン検出を用いたライフログ映像からの印象的なシーンの検索、日本データベース学会、2012 年 3 月 5 日、シーサイドホテル舞子ビラ神戸
- ③ 野宮浩揮、ライフログ映像からの顔特徴点を用いた表情表出シーン検出手法の一検討、日本顔学会、2011 年 9 月 23 日、日本歯科大学
- ④ 野宮浩揮、簡潔な顔特徴量の選択と統合による映像からの効率的な表情表出シーン検出、電子情報通信学会、2011 年 6 月 6 日、慶応大学
- ⑤ 野宮浩揮、顔特徴点を用いた特徴選択と特徴抽出による表情認識に基づく映像中の表情表出シーン検出、日本データベース学会、2011 年 2 月 27 日、ラフォーレ修善寺
- ⑥ 野宮浩揮、表情表出時の顔特徴点の相互関係に基づく特徴選択による表情認識、日本顔学会、2010 年 10 月 24 日、東京医科歯科大学
- ⑦ 野宮浩揮、表情表出時の顔特徴点の変化を用いたアンサンブル学習による表情認識、電子情報通信学会、2010 年 9 月 5 日、福岡大学

6. 研究組織

(1) 研究代表者

野宮 浩揮 (NOMIYA HIROKI)

京都工芸繊維大学・工芸科学研究科・助教

研究者番号：80533116

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：