

H-011

## 人物の顔画像情報に基づくコンテンツの解析 Content Analysis based on Human Face Images

岡田 朋子†  
Tomoko Okada

滝口 哲也‡  
Tetsuya Takiguchi

有木 康雄‡  
Yasuo Ariki

### 1. はじめに

近年、ビデオやDVD等の映像記録媒体を扱うデジタル映像機器が急速に普及し、家庭においてもメディア情報が莫大となっている。そこで大量のコンテンツを種類別に分類する研究や、視聴者に番組を推薦する研究 [1] といった映像検索技術の研究が注目されている。本研究はその中でも、映像中の顔画像情報に基づいてコンテンツの解析を行うことで、ユーザの再視聴の際に見たい場面を効率よく探し出すニーズに応えることを目的とする。具体的には、あるテレビドラマの動画像を研究資料として使用し、その映像から主役の俳優の顔を検出し、顔表情認識をすることで、ユーザが希望する特定の人物の、特定の表情を表出した場面を検出する。

### 2. 提案システムの概要

#### 2.1 提案システムに用いる手法

テレビドラマの動画像中の俳優は顔を自由に動かして演技をし、また撮影するビデオカメラも場合によって視点をかえるため、人物の顔の方位や表情は様々である。そのため、従来の正面に限定するといった顔照合、顔表情認識を用いることは困難である。そこで本研究では方位、表情の変化に頑健性のある相互部分空間法 (MSM) [2] を用いて顔照合を行い、同じく顔の方位、表情が変化しても顔特徴点の追跡が可能な Active Appearance Model (AAM) [3] を用いる。AAMを使用することで、顔の方位を正面に戻すこと [4] ができるため、任意の方向を向いた顔表情認識が可能である。

#### 2.2 提案システムの流れ

図1に本研究の流れを示す。まず映像中から Haar-like 特徴を特徴量として AdaBoost によって顔検出を行い、検出された顔画像から MSM を用いて主役の俳優か否かを判定する。そして、俳優の顔画像から作成した AAM を入力画像に張り付けることで、顔特徴点を追跡し、顔の方位を正面に戻した後、Support Vector Machines (SVMs) によって顔表情認識を行う。

### 3. 提案システムの手法

#### 3.1 MSM

MSM はあらかじめ得られる複数の画像を学習データとし、未知の複数の画像を入力データとした時、その類似度をはかることができる手法で、文字認識や顔認識に

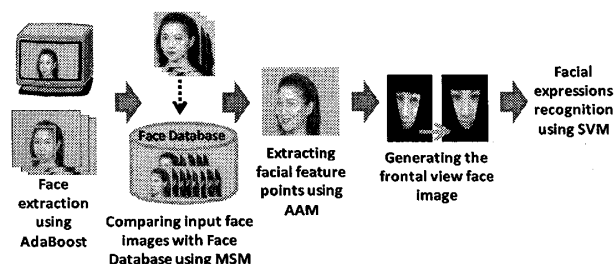


図1: 提案手法の流れ

よく用いられる。学習時、入力時の双方で PCA を用いて部分空間を作成するため、変動が少ない部分を比較することができ、顔認識においては顔の方位や表情にロバストだと言われている。比較する部分空間を  $P$ ,  $Q$  とすると、部分空間同士の類似度 (正準角)  $\theta$  は次式によって定義される。 $u, v$  は  $P$ ,  $Q$  に射影された入力ベクトル集合と学習ベクトル集合である。

$$\cos^2 \theta = \max_{u \in P, v \in Q, \|u\| \neq 0, \|v\| \neq 0} \frac{|(u, v)|^2}{\|u\|^2 \|v\|^2} \quad (1)$$

#### 3.2 AAM

AAM は shape (特徴点の座標値) と texture (shape 内の輝度値) の結合によって得られる combined parameter を持つことで構築された統計学的なモデルで、顔や臓器等の特徴点探索に広く用いられている。それぞれの特徴量を PCA によって次元削減を行うため、高速かつ安定して追跡することが可能となっており、また、顔の方位等といったモデルの変動が多い部分を取り除くことによって、顔表情はそのまま顔を正面に戻すことが可能である。以下に AAM の構成と探索、そして正面顔を生成する方法について述べる。

##### 3.2.1 AAM の構成

ここでは AAM の構成について述べる。図2のように特徴点の座標が与えられた複数の画像を学習データとし、座標値を並べた shape ベクトルを  $s$  とおく。また、これの内部のテクスチャを平均形状に正規化し、その輝度値を並べた texture ベクトルを  $g$  とする。これらをそれぞれ PCA すると次式で表現される。

$$s = \bar{s} + \Phi_s b_s, g = \bar{g} + \Phi_g b_g \quad (2)$$

ここで、 $\bar{s}$ ,  $\bar{g}$  は shape, texture の平均ベクトルであり、 $\Phi_s$ ,  $\Phi_g$  はそれぞれの空間の基底ベクトルで構成される

†神戸大学大学院工学研究科

‡神戸大学自然科学系先端融合研究環

行列,  $b_s, b_g$  はそれに対応したパラメータベクトルである。そして, shape と texture の単位を正規化する行列  $W_s$  を用いて結合させ, 再び PCA を実行することにより,  $Q$  の基底ベクトルに対応した combined パラメータベクトル  $c$  を得ることができる。

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s \Phi_s^T (s - \bar{s}) \\ \Phi_g^T (g - \bar{g}) \end{pmatrix} = Qc \quad (3)$$

この  $c$  を制御することによって shape, texture の両方を同時に扱うことが可能となる。

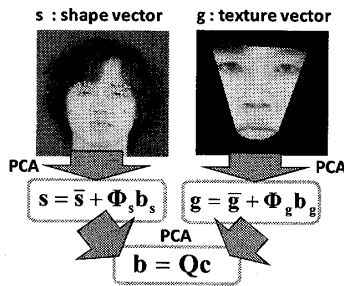


図 2: AAM の構成

### 3.2.2 AAM の探索

入力画像が与えられた時, 先に述べた方法で構築した AAM に平行移動, 回転を施し,  $c$  を用いて変形させる。それによって合成された顔画像の texture と入力画像の texture を比較し, その誤差が最小となる最適な  $c$  を再急降下法によって求め, 入力画像に最も近い AAM を探索する。これによって入力画像の特徴点を探索することができる。

### 3.2.3 顔方位を算出し正面顔を生成する方法

AAM で取得できる combined パラメータベクトル  $c$  の低次元には顔方位の成分が現れる。従って,  $c$  を低次元のベクトル,  $C_0, C_1$  を定数,  $\theta$  を顔方位角度とすると,

$$c = C_0 + C_1 * \theta \quad (4)$$

のようにおける。これに基づいて最小自乗法によって学習データから  $C_0, C_1$  を求めておく。次に入力顔画像から,

$$\theta' = (c' - C_0) / C_1 \quad (C_1 \neq 0) \quad (5)$$

のようにして取得した  $c'$  から, 顔方位角度  $\theta'$  を算出する。また顔方位を正面に戻すには, 式 (4) に  $\theta = 0$  を代入し,  $c_{res}$  を残差ベクトルとした

$$c_{front} = C_0 + c_{res} \quad (6)$$

によって算出される  $c_{front}$  を用いる。

## 4. 実験

### 4.1 実験条件

実験は, 20 分間のテレビドラマ 1 話分の中で, 主役の俳優が映っているフレームにタグ付けし, テストデータとした。AAM の作成や MSM に用いる学習データは, テストデータとは異なるドラマ 2 話分から収集した。顔表情認識は, 主役の俳優が笑顔か否かをフレーム単位で認識した。

### 4.2 実験結果

顔照合の実験結果を表 1 に示し, 顔表情認識の実験結果を表 2 に示す。表 2 の  $\alpha, \beta$  は全フレームを対象にしており,  $\alpha$  は顔照合と顔表情認識の両方に成功した結果で,  $\beta$  は顔表情認識のみが成功した結果である。また  $\gamma$  は, 顔照合ができたフレームのみを対象にした顔表情認識の結果である。

表 1: 顔照合の実験結果

適合率	再現率	F 値
0.79	0.75	0.77

表 2: 顔表情認識の実験結果

実験の種類	適合率	再現率	F 値
$\alpha$	0.56	0.57	0.56
$\beta$	0.63	0.64	0.63
$\gamma$	0.77	0.94	0.85

### 4.3 考察

実験結果から顔照合が 8 割程度で実現しており, 顔照合ができていないフレームでは 8 割 5 分程度で顔表情認識が実現できていることがわかる。顔表情の認識率の低下は, 照明や顔の動きの変化が激しい場面で, AAM による顔特徴点追跡の精度が落ちたためだと考えられる。

## 5. おわりに

本研究では, 再視聴者の場面検索のニーズに応えるために, 映像コンテンツから特定の人物, 表情を認識する方法を提案した。今後は, 顔照合, 顔表情認識の精度を向上させ, 表情の種類を増やす等, よりユーザの希望に沿う研究を行う予定である。

## 参考文献

- [1] Masanori Miyahara, Masaki Aoki, Tetsuya Takiguchi, Yasuo Ariki, "Tagging Video Contents with Positive/Negative Interest Based on User's Facial Expression," The 14th International Multimedia Modeling Conference (MMM2008), pp. 210-219, 2008.
- [2] K. Maeda and S. Watanabe, "A pattern matching method with local structure," IEICE Trans. Inf. & Syst. (Japanese Edition), vol. J68-D, no. 3, pp. 345-352, March, 1985.
- [3] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," The 15th European Conference on Computer Vision (ECCV1998), Vol. II, Freiburg, Germany, 1998.
- [4] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "View-based active appearance models," Image and Vision Computing 20, pp. 227-232, 2002.