

L_013

番組の出演者一覧生成のための顔画像クラスタリング手法

Clustering of Face Features for Listing Performers in TV-Programs

奥村真澄† 大網亮磨† 平田恭二十
Masumi Okumura Ryoma Oami Kyoji Hirata

1. はじめに

地上デジタル放送の開始や大容量低価格ハードディスクレコーダの普及により膨大な番組映像が蓄積されつつあるのを受けて、視聴したい番組を効率的に選択できる番組一覧への需要が高まっている。我々は、出演者の代表的な顔を出演率順に一覧表示することによって、各番組の出演者情報を容易に把握できる番組一覧と、その実現を目的とした代表顔抽出技術について提案した[1]。代表顔抽出技術では、番組映像から検出された顔を顔画像クラスタリングにより出演者ごとのグループに分け、各グループから出演者一覧に表示する顔画像を抽出する。この顔画像を利用した図1に示す番組一覧により、利用者は、出演者の名前を知らない場合でも、各番組に誰がどんな頻度で登場するか表情を含めて即座に把握でき、指定人物の出演番組を容易に検索できる。

各番組の出演者一覧への要求条件として、1. 出演者が重複無く表示できること、2. 出演率の高い出演者が上位に表示できること、3. 特に上位1位から3位程度の出演者が確実に表示できることが挙げられる。

上記条件を満たす出演者一覧の実現には、顔向き・表情・照明条件などの様々な変動に対して頑強に顔画像をクラスタリングする必要がある。そこで本稿では、顔画像クラスタリング手法として重心法および最短距離法を用い、45分ドラマ映像118本に両手法を適用した際の検出精度の比較結果について報告する。

2. 出演者の代表顔抽出方式

提案した代表顔抽出方式は以下に示す、顔検出・正面顔選択・顔画像クラスタリング・代表顔選択の4ステップで構成される。

処理1. 顔検出

リングフィルタによる目検出とGLVQによる顔判定処理[2]により番組映像から人物の顔画像を抽出する。

処理2. 正面顔選択

抽出された全顔画像から、左右対称性を利用して正面顔を選択する。

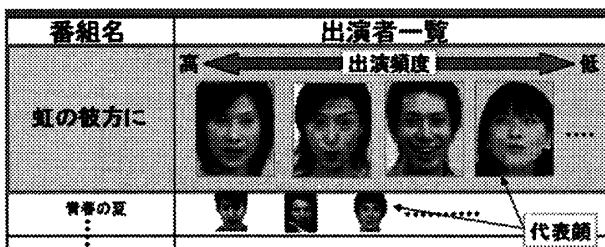


図1. 出演者の顔による番組一覧のイメージ図

処理3. 顔画像クラスタリング

顔特徴量の値が類似する顔画像をグループ化し、顔画像クラスタを生成する。顔特徴量には、振動空間法により顔の変動要因に対応したガボールウェーブレット基底に基づく特徴量[3]を用いる。

処理4. 代表顔選択

顔画像クラスタから、クラスタ重心に最も近い画像を出演者の代表的な顔(代表顔)として抽出する。また、クラスタに属する顔画像数から出演者の出演頻度を算出し、出演者一覧における代表顔の表示順を決定する。

3. 顔画像クラスタリング手法の検討

本稿では、出演者一覧において出演者の顔が重複・漏れ無く表示されるように、変動に対して頑強な顔画像クラスタリング手法について検討した。顔画像クラスタリングでは、ショット内で特徴量の時間的連続性に基づき初期クラスタを生成後、ショット間で階層的クラスタリングによりクラスタ間距離が所定の閾値以上になるまでクラスタを統合する。階層的クラスタリングとして、以下で述べる重心法および最短距離法を用いる。

(1) 重心法

クラスタ間距離としてクラスタ重心間の距離を用いる。重心法は、特徴量がクラスタの重心付近に集中的に存在し、かつ等方的に分布している場合に、高精度にクラスタリングできるという特性がある。同一人物の正面顔の特徴量は互いに値が類似するため、同一人物の顔特徴量は重心付近に集中した等方的な分布になると推測される。同一人物の正面顔の特徴量分布がこの仮定を満たす場合に、本手法が有効である。

(2) 最短距離法

クラスタ間距離として、一方のクラスタの全特徴量ともう一方の全特徴量の間で算出される距離の最小値を用いる。最短距離法は、特徴量がクラスタ重心から離れて分布する場合であっても、2クラスタ間で値の類似する特徴量対が存在すれば1つに統合できる。そのため、多様な表情の顔や正面顔選択で除去できずに混入した非顔・非正面顔によって同一人物の顔特徴量の分布に等方性がない場合に、本手法が有効である。

4. 評価実験

45分ドラマ映像24タイトルの各5話分(1タイトルだけ3話分)からなる全118話について、重心法および最短距離法を適用したときの出演者の検出精度を比較する。

検出精度は、出演者一覧に表示する顔画像の枚数(表示顔画像数)が指定されたときに、本方式および正解データをもとにした出演者一覧の比較により算出される。正解データは、目視により各フレームに手動付与した人物情報を用いる。評価尺度には以下3種を用いる。

† NECメディア情報研究所
NEC Media and Information Research Labs.

(a) 出演者検出率

表示顔画像数に対して正しく抽出できた出演者の顔画像数の割合。同一人物が重複して検出されると値が低下するため、同一人物の重複検出を評価できる。

(b) 上位者検出率

表示顔画像数が N 枚のときに出演率の高い上位 N 人を上位者とし、表示顔画像数に対して正しく抽出できた上位者の顔画像数の割合。同一人物の重複検出や下位出演者の検出に起因する上位者の検出漏れを評価できる。

(c) 高出演率登場人物検出率

正解の出演率が 4%以上の高出演率出演者(通常 1 位から 3 位)の数に対し、正しく抽出できた高出演率登場人物の顔画像数の割合。高出演率出演者の検出漏れを評価する。

5. 結果と考察

5.1 結果

重心法および最短距離法による出演者の検出精度を図 2 に示す。表示顔画像数の増加に伴い、両手法で高出演率登場人物検出率は向上し、出演者検出率と上位者検出率は低下する。また、表示顔画像数の増加に伴って出演者検出率と上位者検出率の差が小さくなることから、上位者の検出漏れの要因のうち同一人物の重複検出の割合が増え、下位出演者の検出の割合が減ることがわかる。出演者検出率と上位者検出率は、表示顔画像数に限らず最短距離法の方が約 10%高い。よって重心法よりも最短距離法の方が適しているといえる。

番組選択のための一覧に表示する顔画像数として妥当と考えられる 4 枚の代表顔表示時に、最短距離法では高出演者検出率 90%を得たことから、上位 1 位から 3 位の出演者をほぼ漏れなく検出可能であることがわかる。また、出演者検出率が約 80%、上位者検出率が約 66%であった。

以上から、最短距離法を用いた代表顔検出技術は、代表顔を 4 名程度表示する出演者一覧に十分適用できると考えられる。

5.2 考察

重心法より最短距離法の方が適する理由として分布の非等方性が考えられる。実際に、顔画像クラスタリングのショット内処理において、多様な表情の顔画像や、正面性判定で除去できなかった非正面の顔画像に起因する初期クラスタの非等方性が観測された。以下では、非等方的なクラスタを模式化した図 3 を用いて説明する。

重心法では、人物 A のクラスタ同士の重心間距離よりも人物 B のクラスタとの距離が近い場合、さらにクラスタリングを進めると A と B の混合クラスタが生成される。多様な変動を含む顔画像が混合した非等方的なクラスタの重心は、例えば右向きと左向きが混合した顔のように、実際には存在しえない顔特徴量となり得る。重心法では、この重心をもとに算出したクラスタ間距離に基づいて統合判定を行うため、不正確な判定が行われる。

一方、最短距離法では、実在する顔の特徴量間距離を用いるため、変動が混合した顔による不正確な判定は行われない。また、最短距離を利用するため、図 3 の様に距離の近い顔特徴量対を 1 組以上含んでいれば、クラスタ同士を結合できる。そのため、クラスタ重心間の距離が遠くても、例えば同一人物の同じ表情・向き・照明条件の顔のように互いに特徴量間距離の近い顔を含む場合には、同一人物の

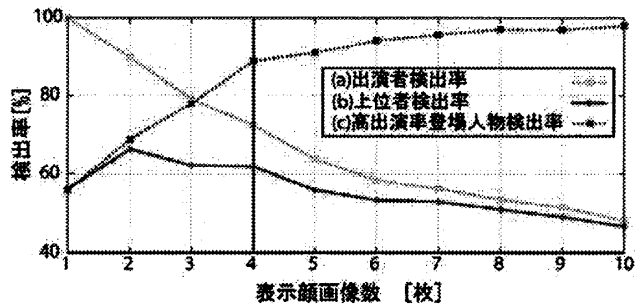
クラスタを統合できる。以上より、最短距離法は重心法よりも良好な検出率が得られたと考えられる。

6. まとめ

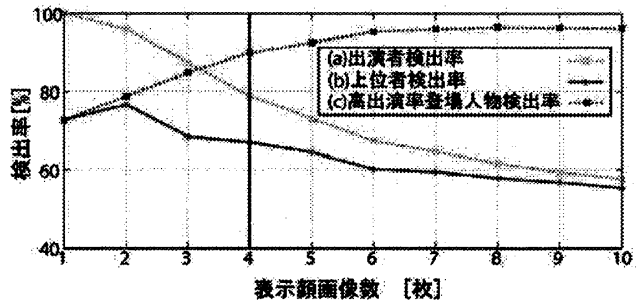
出演者一覧を実現する代表顔抽出技術において、変動に対して頑強な顔画像クラスタリングについて検討した。顔画像クラスタリング手法として、重心法および最短距離法を用いた際の出演者の検出精度を比較した。45 分ドラマ映像 118 本に対し両手法を適用した結果、出演者検出率および上位者検出率について、表示顔画像数によらず最短距離法の方が重心法よりも約 10%高い値を得た。最短距離法では、顔画像 4 枚表示時に、出演者検出率 80%、上位者検出率 66%、高出演率登場人物検出率 90%を実現し、代表顔抽出技術が出演者一覧に十分適用可能であることを示した。

参考文献

- [1]奥村真澄, 平田恭二, “出演者の代表顔に基づく番組一覧による番組選択方式,” IEICE 春季総大 2006 D-12-37.
- [2]細井利憲, 鈴木哲明, 佐藤敦, “一般化学習ベクトル量子化に基づく顔検出,” FIT2002 I-30, 2002.09.
- [3]今岡仁, 佐藤敦, “判別分析と摂動空間方を用いた顔照合アルゴリズム”, FIT2005 I-12, 2005.09.



(1) 重心法による出演者の検出精度



(2) 最短距離法による出演者の検出精度

図 2. 出演者の検出精度の比較

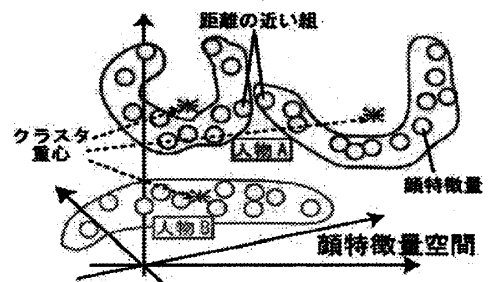


図 3. 非等方的なクラスタの一例