特定人物登場シーン抽出システム

市川 雅人　坂東 忠秋　中屋敷 かほる
関東学院大学 工学部 情報ネット・メディア工学科

1. はじめに

近年ＤＶＤ・ＨＤＤレコーダの高機能化が進むと共に、大量の動画ファイルから、利用者が関心を持つシーンを取り出すことが注目されている[1]。本研究の目的はテレビ映像から、特定人物が表示されている映像のみを抽出することである。本システムをテレビ約４０分と、ビデオ撮影した約１０分の映像を使って有効性を検証した。

2. システム概要

本システムは大きく分けて、
①顔の有無を検出する顔検出
②顔が誰であるか認識する顔認識
③顔の領域を追う顔追跡
以上の３つの部分から構成される。まず、映像１フレーム毎に顔検出を行う。検出できた顔領域を、蓄積してある顔画像特徴量データとの比較により人物判別を行う。特定人物が認識されたら、そのフレームから顔が連続で表示されているフレームに対して、顔追跡する。顔が追跡できたフレームを、特定人物の登場しているシーンとして抽出する。

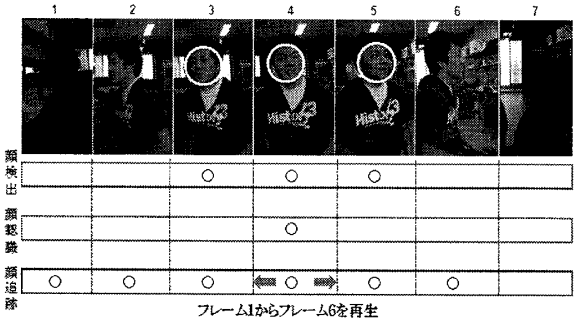

図１　システム概要

3. 顔検出

3-1 HAAR変換

顔検出には、OpenCV[2]の顔検出機能を用いる。OpenCVは色情報を使わず、明度の変化量を見て顔を検出するHAAR変換という方法を用いている。この方法は、照明変動の影響を受けない方式である。しかし、実際の映像での実験では、撮影条件にとらわれず検出できるが、明度の分布のみで判断しているため、顔でないところを顔と誤検出するケース(図2-2)も多くある。検出範囲は、顔の正面からやや横を向いている場合までである。眼鏡をかけている場合や髪の毛で顔が隠れている場合も、ある程度までなら検出できる。

3-2 肌色検出

OpenCVの誤検出を少なくする為、HAAR変換に肌色情報のチェック機能を追加した。この機能は、検出された顔の重心から±５pixelの範囲で肌色のチェックをするものである。これにより、誤検出を大幅に減らすことができる。肌色情報は色合いを表すHと、彩度を表すSを用いる。値は日本人の肌色分布から 120＜H＜160・20＜S＜120 とする。

図 2-1　成功例

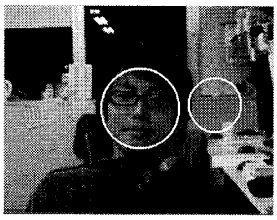

図 2-2　誤検出

4. 顔認識

4-1 主成分分析

顔認識には、主成分分析の手法を用いる。主成分分析とは“元のパターン全体の分布を最もよく近似”するように、特徴ベクトル空間の次元数を削減した部分空間を求める方法である。

4-2 顔領域と顔特徴ベクトル

顔認識に必要な顔領域（図3-2）は、個人の特徴が含まれる目・鼻・口・眉毛が含まれる領域のみである。その他の領域（背景や髪の毛）が含まれると誤認識の大きな原因になる。そこで、試行錯誤で適した数値を下記計算式によって設定した。

重心座標 =（ｘ，ｙ），半径 = ｒ

```
minY = y - r / 2
maxY = y + r * 4 / 5
width = r * 6 / 10 * 2
minX = x - width / 2
maxX = x + width * 2
```

本システムでは、顔領域を２０×２０の正規化画像に変換し、主成分分析により顔の特徴ベクトルを求める。あらかじめデータベースに、一人につき１００枚の顔画像から主成分分析で得られた特徴ベクトルを登録しておく。データベースに登録する画像の顔の向きや表情にばらつきがあると、その向きや表情に偏ってしまう。本実験では正面向き、無表情の顔画像のみを登録する。

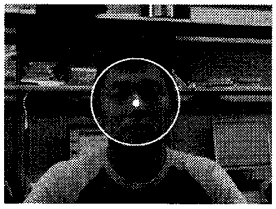

図 3-1 OpenCV 顔領域

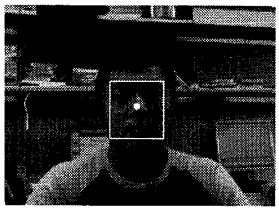

図 3-2 必要な顔領域

4-3　パターンマッチング

顔認識は、入力した顔との距離値が近い上位 10 個のデータに、距離値が最も近い人物と同じ人物が 6 人以上存在し、その距離値が 750 以下の時、特定人物と認識する。

5.　顔追跡

5-1　肌色追跡

OpenCV の顔検出機能は横を向いてしまった場合、検出不可能なので追跡できない。そこで、正面以外を向いている顔を追跡するため、肌色情報を用いる。

特定人物が認識されたフレームの次のフレームでは、認識した顔の重心から、縦と横に半径の距離の領域内で肌色抽出（図 4 右上）する。抽出された肌色領域の雑音除去をし、顔の肌色領域の重心（図 4 右下）を求め、次のフレームからは肌色領域から求めた重心を用いて同じ処理を繰り返す。抽出できた肌色領域の面積が 100 画素を下回る場合、特定人物がフレームから消えたと判断して顔追跡を終了する。

実際の映像にはシーンの不連続点（シーンチェンジやカットなど）が含まれる。本システムの顔追跡は不連続点があると追跡不可能になってしまう。不連続点を検出するには、各種の研究があるが、今回は目視で不連続点を検出する。これにより、それぞれ連続しているシーン間で追跡を行う。

5-2　逆フレーム追跡

5-1 の肌色追跡は、特定人物を認識した後のフレームを追跡する。しかし、認識する前からその人物は映像に登場していたとも考えられる。そこで、追跡を開始したフレームの番号から逆向きにフレームをチェックし、同じように肌色追跡を実行する。

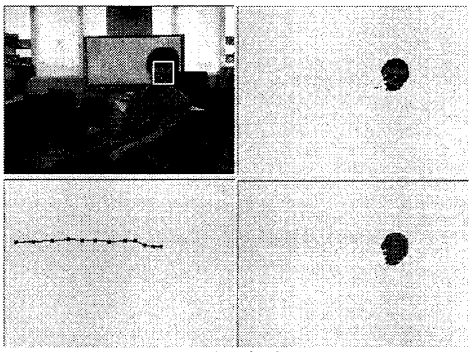

図 4　顔追跡

（左上：原画像　　右上：肌色抽出結果
左下：重心の軌跡　右下：顔肌色領域（雑音除去））

6.　評価実験

6-1　実験方法

本実験にはテレビ番組約４０分と、ビデオカメラで撮影した映像約１０分の映像を使用する。テレビ映像のコマーシャルはカットしておく。ビデオカメラで撮影した映像を使う場合、出演者数は８人・顔データベースは一人につき１００枚の登録をする。ＴＶ映像を使う場合、出演者数は１０人・顔データベースは一人につき１０枚の登録で実験を行う。

以下のデータ①～⑥を求める。

③　追跡率＝ ①特定人物を追跡できたフレーム数 / ②特定人物登場シーンのフレーム数

⑥誤追跡率＝ ④特定人物以外を追跡したフレーム数 / ⑤追跡したすべてのフレーム数

6-2　実験結果

特定人物を変えて２回ずつ実験を行った。

表１　実験結果データ

	ＴＶ映像		撮影した映像	
	人物１	人物２	人物３	人物４
①[frame]	8998	3412	1052	3884
②[frame]	12236	3525	2260	5291
③[%]	73.5	96.8	46.5	73.4
④[frame]	240	122	4471	4300
⑤[frame]	9238	3534	5523	8184
⑥[%]	2.6	3.5	81.0	52.5

追跡できたフレームを特定人物の登場しているシーンとして抽出するので、追跡率＝抽出率となる。

ＴＶ映像では OpenCV で特定人物の顔が検出できない場合があり、（顔の表示が小さい・色付きの照明で肌色検出不可能など）顔認識が行えず追跡できないケースが発生した。ビデオ撮影した映像の追跡率はあまり良いとは言えないが、顔認識で認識できたシーンからの顔追跡は精度良く追跡できた。

誤追跡の主な原因は、主成分分析での誤認識による誤追跡と、顔が重なった後に違う人物の肌色を追ってしまう誤追跡の２つである。

7.　まとめ

今回開発したシステムは、映像から特定人物が表示されている映像のみを抽出するもので、大きく分けて顔検出・顔認識・顔追跡の３つの部分から構成されている。

本システムは一度認識できれば、連続しているシーン間なら全て追跡することができる。よって、本研究の「特定人物登場シーン抽出」という目的から、認識率の高さよりも誤認識率の低さが重要になる。

残された課題として、追跡中に顔が重なる場合や後ろを向いた時、追跡できなくなる場合があるが、他の特徴量（服の色など）を取り入れた追跡手法に改良すれば、この問題は改善される。また、不連続点を毎回手動で入力するという動作は、本研究の目的から考えても大きな欠点である。本システムは、それら不連続点を自動的に検出するシステム[3]が、確実に必要になる。

参考文献

[1]山崎智弘・筒井秀樹・浦田耕二・福井美佳（東芝），“コミュテンツ記事からのシーンメタデータ抽出実験と評価”
[2]Intel Corporation, ” openCV”
[3]松井淳・サイモンクリピングデル・松本隆，“ベイズ的動画像顔検出における顔候補領域の逐次予測”