

制約充足問題に基づく、顔の向きによらない登場人物の認識

清水 美穂†

†神戸大学工学部情報知能工学科

上原 邦昭‡

‡神戸大学大学院工学研究科

1 はじめに

近年、マルチメディア技術の進歩に伴って、放送映像や個人撮影映像など、大量の映像がウェブ上や個人の HDD に蓄積されるようになってきた。このような環境の中で、大量の映像の中から、ユーザが視聴したいイベントを効率的に検索するための映像検索技術に対する期待が高まっている。

最近、従来の色や動きなど“信号レベル”の特徴量を用いた映像検索よりも、登場人物やオブジェクトなどの“意味レベル”の特徴量を用いた映像検索が注目を集めている。白浜らの研究 [1] では、映像中に出現する登場人物の出現パターンに基づいて、映像をイベントに分割すれば、映像を約 80% の精度でイベントに分割可能であることが示されている。さらに、登場人物の例外的な出現パターンを検出することによって、登場人物が興味深い行動を行っているイベント（例えば、恋愛、殺人、逃走など）をトピックとして抽出できることが示されている。本論文では、白浜らの研究を自動化することを目的として、映像中に様々な向きや大きさで出現する登場人物を認識するためのアプローチを提案する。

一般に、人物認識は、顔領域内の特徴量を比較することにより行われる。しかしながら、カメラの位置や距離によって人物の顔の大きさや向きは様々な変わる。そのため、同一人物であっても、異なるショットでは顔領域内の特徴量が大きく異なってしまう。つまり、検出された顔領域の特徴量の比較だけでは、人物を認識することは困難である。

本稿では、映像においては時間的に近いショットは意味的にも類似していると仮定する。例えば、2 人の人物が会話をしているシーンの間は、ショットが切り替わっても 2 人はそれぞれ同じ場所で、同じ服を着て出現する。その仮定の下で、服の色などショット内の別の特徴量を用いて人物認識を行う方法を提案する。

本手法の流れは以下の通りである。まず、認識を行う前に、どこに顔があるのかを知るため、映像から様々な向きの顔を検出する。そして、検出された顔のうち、正面顔についてのみ認識を行う。正面顔については、2 次元 HMM を用いて認識している。最後に、服の色の照合とショットの意味の連続性に基づく時間的制約を導入し、認識済みの正面顔と検出された横・斜め顔を関連付け、横顔の認識を行う。

Multi-View Face Character Recognition in Videos Using Constraint Satisfaction Problem

†Miho Shimizu ‡Kuniaki Uehara

†Department of Computer Science and Systems Engineering, Kobe University

‡Graduate School of Engineering, Kobe University

2 様々な向きの顔の検出

本研究では、様々な方向を向いた顔を検出するために、松山らが提案した手法を用いている [2]。この手法では、Haar 型の特徴量を学習によって組み合わせる分類器を生成し、画像の分類を行っている。検出器の生成方法を図 1 に示す。(a) のように、顔の対称性から、まず半顔の検出器を作成し、それを反転させたものと組み合わせる 2 次元の全顔検出器を生成する。次に、(b) のように、顔の奥行き方向の回転を表現するために、生成した全顔検出器を 3 次元モデルにマッピングする。そして、3 次元モデルを鉛直軸方向に回転させてから、(c) のように、再び 2 次元に写像して (d) のような新たな検出器を作成する。これを予め様々な角度について実行し、各角度に対応した検出器を作成する。

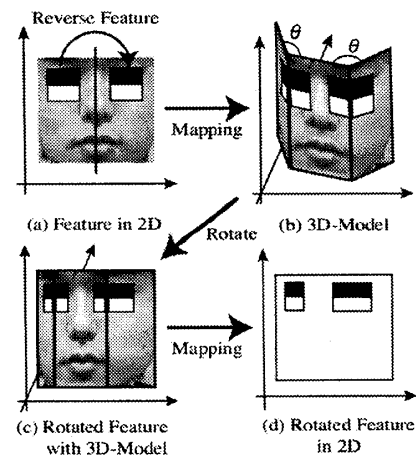


図 1: 特徴のマッピング

先述した方法により、 -60° 、 -30° 、 0° 、 30° 、 60° の向きに対応した検出器を作成し、様々な角度の顔に適用した結果を表 1 に記す。各角度ごとに 30 枚ずつの画像を用意しその全てに 5 つの検出器を適用した。error は、顔でない領域を検出した数である。

表 1: 各検出器により検出された顔の個数

		顔の向き ($^\circ$)													
		90	75	60	45	30	15	0	-15	-30	-45	-60	-75	-90	error
検出器の角度 ($^\circ$)	60	12	25	23	16	8	13	5	0	1	8	0	0	0	1
	30	0	0	0	0	6	11	11	1	0	0	0	0	0	0
	0	0	0	0	1	2	6	17	24	14	6	0	0	0	1
	-30	0	0	0	0	0	2	13	13	6	3	2	0	0	1
	-60	0	0	0	0	0	0	2	2	3	3	17	15	5	0

それぞれの検出器に対応する角度付近の顔が検出されているようすが分かる。しかし、まだ精度に改善の余地が見られる。今回検出できなかった顔の特徴として、眼鏡をかけているもの、頭部が左右に傾いているもの、照明状態によって顔の濃淡が大きく変わってしまっているものなどが見られた。

また、この検出器をあるホームドラマに適用したところ、精度が低下してしまった。問題として、顔が小さいと検出ができない点、顔の特徴と似た背景の部分を顔であると誤認識してしまう点があげられる。これらの改善は今後の課題とする。

3 正面顔の認識

本章では、2次元HMMを用いた正面顔の認識手法について述べる[3]。まず、多数の2次元顔画像を参照画像として保持し、テスト画像と比較する。ここで、テスト画像の画素の輝度値と参照画像の画素の輝度値の差が最小となるようにテスト画像の画素を移動させ、参照画像と対応づける。この時、2次元HMMの状態推移の仕方に基づいて、一つ前の画素の対応関係を考慮しながら、現在の画素を対応させる。ただし、もとのテスト画像の顔の構造を極端に崩すのを避けるために、対応づけの範囲を限定している。

図2は、一番左の画像がテスト画像で、一番右の画像が参照画像を表している。上の段は垂直方向、下の段は水平方向にテスト画像と参照画像ともにSobelフィルタをかけており、エッジがよりはっきりするようにしている。そして、左から2番目の画像に示すように、テスト画像を変形させて参照画像と対応づける。また、左から3番目の画像はテスト画像をどのように変形させたかをグリッド表示したものである。最後に、テスト画像と参照画像の対応づけから算出される距離に基づいて、1-NN法を用いて、テスト画像に一番近い参照画像のラベルを割り当てる。

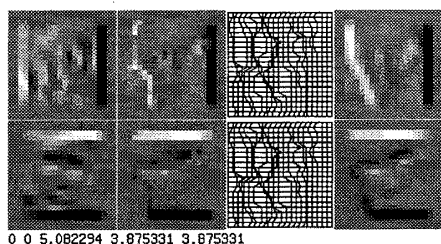


図2: テスト画像の変形

上記の2次元HMMに基づく認識手法を用いて、実際に16×16画素の画像を訓練に748枚、テストに20枚用いて分類を行ったところ、85%の精度という良好な結果が得られた。主な要因として、以下の2点が挙げられる。1点目は、上記の認識手法を用いれば、顔の構造をあまり損なわず、かつ異なる人物の顔の微妙な差異を分類に用いることができるからである。2点目は、テスト画像と参照画像との距離の算出に周りの輝

度値の差との平均を用いており、画素間の関係の情報も分類に用いることができるため、顔の位置や大きさにロバストな分類ができるからである。

4 横・斜め顔の認識

横・斜め顔の分類にあたり、正面顔についてはあらかじめ正しく認識ができているということ、さらに映像における意味の連続性から、一定時間内ではショットが変わっても同一人物は同じ色の服を着ているということ仮定する。例えば映像中において、図3のように、人物Aが時間的に近いショット中に正面向き、斜め向き、正面向きで登場しているシーンを考える。この場合、この仮定を用いれば、まず1つめと3つめのショットのキーフレーム内の正面顔については既に認識がなされている。次に、それぞれのショットに登場している人物の着ている服の色が一致することが分かれば、2つめのショット中の斜め顔についても人物Aであるとの認識ができる。この手法の検証については今後の課題とする。

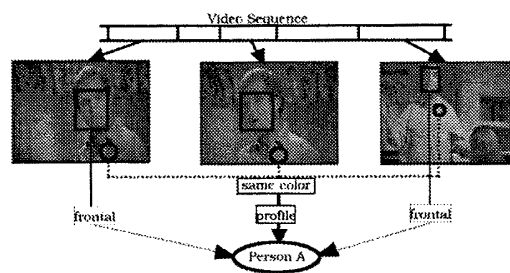


図3: 服の色などによる横顔検出手法

5 まとめと今後の課題

本稿では、映像に登場する人物を顔の向きによらずに認識する方法を提案した。さらに、その中の一部である顔検出部分について、様々な向きの顔検出器の精度の検証を行った。今後、映像の特徴(ロングショットや背景の映り込みなど)を考慮し、顔検出器を改良することが必要である。さらに、最終段階として、服の色を抽出して実際に横顔を正面顔と対応させる部分の実装と、意味が連続していると仮定する時間を実験的に求める必要がある。

参考文献

- [1] Shirahama, K. and Uehara, K.: Video Data Mining: Discovering Topics by Burst Detection in Video Streams, *Proc. of the KDM*, pp. 57-62 (2007).
- [2] Matsuyama, J. and Uehara, K.: Multidirectional Face Tracking with 3D Face Model and Learning Half-Face Template, *Computer Graphics and Vision Theory and Applications*, Springer (to appear).
- [3] Keyzers, D., Gollan, C. and Ney, H.: Local Context in Non-linear Deformation Models for Handwritten Character Recognition, *Bildverarbeitung für die Medizin*, pp. 366-370 (2004).