

人物不特定可変テンプレートを用いた顔画像認識の高速化 Person-independent deformable templates for face recognition

サイモン クリッピングデル†
Simon Clippingdale

藤井 真人†
Mahito Fujii

1. まえがき

映像への自動インデックス付与やユーザーインタフェースの個人化などの応用を目指して著者らが映像中の顔画像領域を追跡し認識するシステム[1][2]を開発した。顔の表情や発話に伴う顔変形に対応するために、中心技術として可変テンプレートマッチングを採用した。基本的には、入力映像中の顔を認識するために、認識対象人物の顔画像から予め構築しデータベースに登録された可変テンプレートのすべてを一つずつ、入力フレームの顔領域とを照合（可変テンプレートマッチング）する必要がある。

本報告では、この可変テンプレートマッチングの処理に伴う計算量を軽減し高速化を実現する手法について述べる。計算量の軽減と認識性能の程度を計測するための実験の結果を紹介し、手法の効果を示す。

2. 顔画像認識システムの構成

著者等が開発した顔画像の追跡と認識システム[1][2]の構成概要を図 1 に示す。入力映像のフレームに肌色領域抽出と顔検出の結果から、顔の追跡と認識に使用される可変テンプレートマッチングを初期化する（本報告では肌色領域抽出と顔検出の詳細を省略する）。

2.1. 可変テンプレートマッチング

可変テンプレートは

- 9点の特徴点の位置 (x-y座標) ;
- 各特徴点で計測されたガボールウェーブレット特徴 (「テンプレート特徴」)

からなる。可変テンプレートマッチングのアルゴリズムは下記のとおりである：

1. 前フレームから残った追跡結果または現フレームでの顔検出結果から、マッチング対象の可変テンプレートの特徴点の位置を入力画面上に初期化する；
2. 入力画面上に、各特徴点でガボールウェーブレット特徴「入力特徴」を計測する；
3. 入力特徴とテンプレート特徴の位相の差分から、入力特徴がテンプレート特徴に最も類似するようになる各特徴点の位置ずれを推定する；
4. 各特徴点をずらして、ずらされた位置で入力特徴を再度計測し、収束するまでステップ 3 に戻る。
5. テンプレート全体のマッチスコアを各特徴点の類似度と特徴点配置のゆがみエネルギーから算出する；

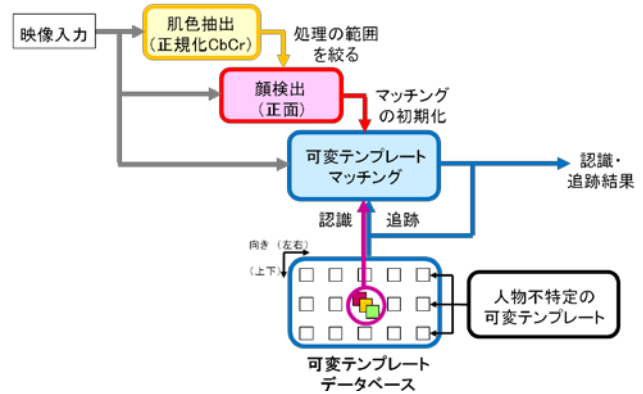


図 1. 顔追跡・認識システムの構成概要

$$S_{A,B} = 1 - \alpha_f \left(1 - \frac{\langle \mathbf{c}^A, \mathbf{c}^B \rangle}{\|\mathbf{c}^A\| \|\mathbf{c}^B\|} \right) - \alpha_s \frac{\sqrt{E_{A,B}}}{\lambda_r} \quad (1)$$

ここでは A と B は入力とテンプレート、 $S_{A,B}$ はマッチスコア、 \mathbf{c}^A と \mathbf{c}^B は特徴ベクトル、 $E_{A,B}$ はゆがみエネルギー、 α_f 、 α_s 、 λ_r は定数である。

2.2. 人物不特定可変テンプレート

図 1 の下部に現れる可変テンプレートデータベースには、認識対象人物の正面顔画像から構築された可変テンプレートが人数分ある。それに加え、正面以外の各向きに一つの「人物不特定可変テンプレート」がある（正面の向きにも一つある）。後者は各向きで、多数の学習用顔画像から構築されたものである。構築の詳細は[1][2]に説明されているのでここでは省略するが、各特徴点毎にすべての学習用画像を自動的に位置合わせて ([3]のアルゴリズムにより)、平均画像を取って、可変テンプレートを造ったものである。従来のシステム[1][2]では、人物不特定可変テンプレートを正面以外の向きでの顔追跡に使用したが、本報告では顔認識に流用することにより、認識処理の高速化を実現したことを報告する。

3. 顔画像認識の高速化

入力顔が正面近くの向きで映ったら、認識対象人物の登録された可変テンプレートのすべてとの可変テンプレートマッチングを行うことにより認識する。概念を図 2 に示す。可変テンプレートの個数は認識対象人物の人数分あるので、この数を N とし、可変テンプレートマッチングに伴う計算量を K としたら、全体的な計算量が NK となる。

† NHK 放送技術研究所

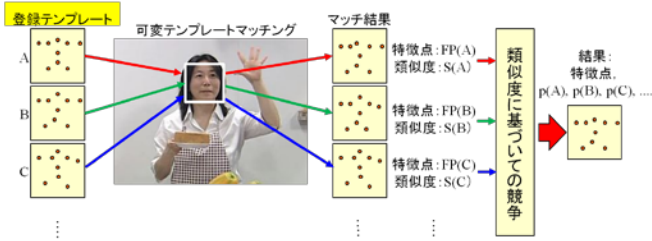


図 2 従来の顔認識の処理 (全テンプレートを順番に変形させる)

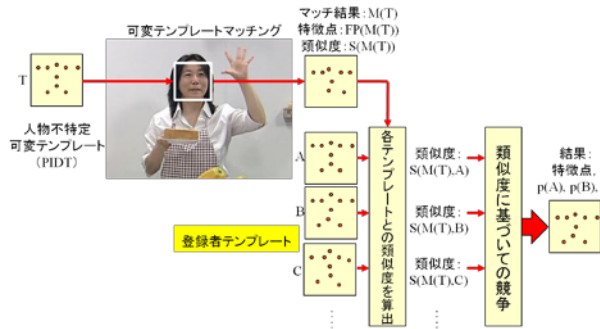
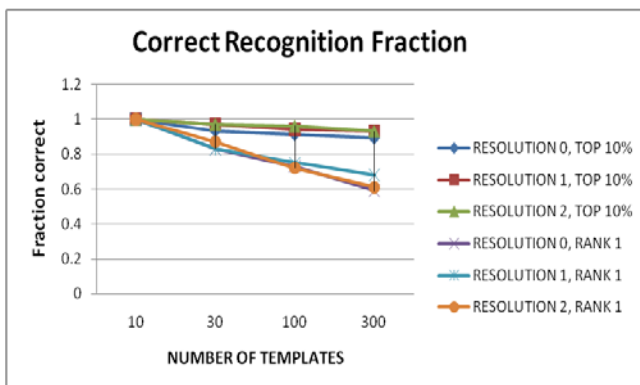


図 3 今回の手法の処理 (人物不特定テンプレートを変形させる)

高速化の手法の概念を図 3 に示す。一つだけの人物不特定可変テンプレートとの可変テンプレートマッチングを行うことにより、入力フレーム中の特徴点の位置を探索し、探索された特徴点の位置でガボール特徴を計測する (この処理は可変テンプレートによる特徴抽出となる)。

そして探索された特徴点の位置とそこで計測された特徴と、登録された人数分の可変テンプレートとの類似度を(1)式により算出し、認識結果を出力する。処理は、図 2 に示された N 回の可変テンプレートマッチングを、図 3 に示された一回だけの可変テンプレートマッチングと N 回の類似度の計算に変更した。

類似度の計算に伴う計算量が H としたら、計算量は NK (図 2) から、 $K+NH$ (図 3) に変更した。 K が H の数千倍 (現行の実装形態では $K/H = \text{約 } 3,000 \sim 8,000$ 倍) あるので、登録者の人数 N が大きくなったら軽減の倍率も大きくなる。

図 4. 図 2 のシステムに比べて図 3 のシステムの認識率 (縦) vs 登録者数 N (横)

4. 実験結果

特徴点の位置を手動で打たれた正面顔画像の 300 枚から 300 個の可変テンプレートと一個の人物不特定可変テンプレートを構築した。後者を使用して 300 枚の各画像に特徴点を探索し、300 個の各テンプレートとの類似度を算出した (図 3)。理想的には、最大類似度が入力画像と同一人物のテンプレートにあるが、人物不特定可変テンプレートによる特徴点探索が失敗した場合に、違うテンプレートにあり、誤認識が発生する。

入力画像の人物が最大類似度になった条件 ("RANK 1") と最大の 10% の類似度に含まれた条件 ("TOP 10%") で、図 2 に比べて図 3 のシステムの認識性能を図 4 に示す。計算量軽減に伴った認識性能の劣化が納得可能な範囲にあるかは応用によると思われる。

5. まとめと今後の課題

本報告では、複数の可変テンプレートとのマッチングによる顔画像認識の高速化を実現出来る手法を紹介した。手法は登録者数分の可変テンプレートとのマッチングを行う代わりに、特別に用意された「人物不特定可変テンプレート」の一つだけを可変特徴抽出器として使用し、抽出された特徴をスタティックに、各登録者の可変テンプレートとの類似度を計算する。

手法の導入により、計算量に相当な軽減が実現出来るが、特徴点の探索が失敗することがある。現在、失敗した状況を自動的に検出して対策をとるといった機能を検討中である。

参考文献

- [1] Clippingdale, S., Fujii, M., "Video Face Tracking and Recognition with Skin Region Extraction and Deformable Template Matching", International Journal of Multimedia Data Engineering and Management, 3(1), 36-48, January-March 2012
- [2] Clippingdale, S., Fujii, M., "Skin Region Extraction and Person-Independent Deformable Face Templates for Fast Video Indexing", IEEE MIPR2011, in Proc. IEEE International Symposium on Multimedia ISM2011, 416-421 (2011).
- [3] Clippingdale, S., Fujii, M., "Texture Image Alignment for Automatic Face Registration", Proc. Workshop on Picture Coding & Image Processing PCSJ2010 / IMPS2010, pp. 91-92 (2010).