

H-009

顔画像認識に使用される顔データの映像からの新規登録の自動化について On Automatic Registration of Face Recognition Data from Video

クリッピングデル サイモン†
Simon Clippingdale

藤井 真人† 柴田 正啓†
Mahito Fujii Masahiro Shibata

1. まえがき

顔における特徴点の位置とその近傍から計測された特徴を可変テンプレートとして、各対象人物の複数の頭部姿勢（顔の向き）で可変テンプレートを構築してデータベースに登録したプロトタイプ顔画像認識システムを著者等が開発してきた。実用化のために、対象人物の新規登録の手間を軽減させる必要がある。ここで登録用映像からの新規登録を自動化するための手法について述べ、実験結果を示す。最後に、今後の課題について考察する。

2. 顔画像認識システムへの新規登録

著者等が FAVRET 顔画像認識システムを開発してきた [1][2]。本システムのデータベースに、対象人物毎に、複数の頭部姿勢（顔の向き）を示す顔画像における特徴点と、特徴点の近傍で計測したガボールウェーブレット特徴を、可変テンプレートとして登録する。

2.1 新規登録の作業

新規に人物を登録するには、要求されるすべての頭部姿勢を示す画像または映像クリップ中のフレームを集めて、候補が複数ある場合にどれを使用するか選択して、特徴点の位置を指定しなくてはならない。システムの実用化に向けて、この作業の手間を軽減させるために、なるべく自動的に登録する手法について検討している。

3. 新規登録の自動化

自動登録手法の概要を図1で示す。登録用映像クリップからの登録の処理の流れが下記のとおりである：

- (A) 約 25 人分のリファレンス顔画像を手動で登録された「ブートストラップデータベース」を備えた FAVRET モジュールを用いて、登録用の映像クリップ中の新規登録対象人物の顔の 9 特徴点を追跡する [1][2]；
- (B) 3D 推定モジュールにより、ジェネリック（人物不特定）な顔の三次元 CG メッシュモデルを登録用映像に、下記のとおり fitting する：
 - (i) 特徴点の追跡結果から、特徴点の三次元位置と映像クリップの各フレームでの頭部姿勢を、逐次バンドル調整 [3][4] のようなアルゴリズム [5] により推定する；
 - (ii) ジェネリックなメッシュモデルを、推定した特徴点の三次元位置に合わせてワーピングする；
 - (iii) 登録用映像クリップの全フレームまたは選択されたフレームセットから、メッシュモデルへのテクスチャマッピングを行なう；
- (C) FAVRET システムが必要とする各頭部姿勢にメッシュモデルを回転させてレンダリングすることにより、登録用映像クリップに含まれなかった頭部

姿勢でも、顔画像とそれにおける特徴点の位置を算出する；

- (D) 各顔画像の各特徴点で、ガボールウェーブレット特徴を計測して、特徴点の位置と特徴より可変テンプレートを構築する。

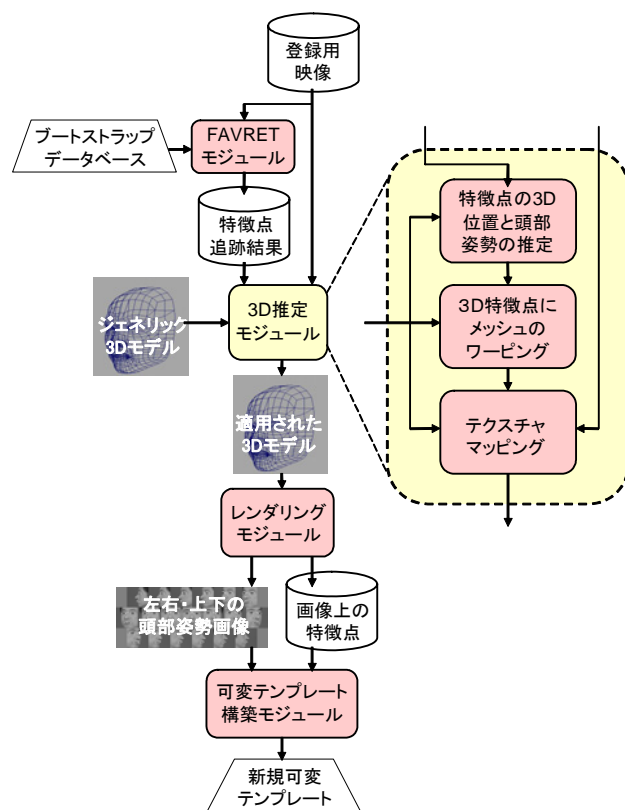


図1 新規登録システムの構成概要

4. 3D 推定モジュール

自動登録の処理の中心に「3D 推定モジュール」がある。3段階の処理を行なう（3節の(B)と図1の右側参照）。処理の最終結果として、登録用映像クリップに適用されたメッシュモデルを次のレンダリング手段に出力する。

3D 推定モジュールの最初の手段（特徴点の三次元位置と映像クリップの各フレームでの頭部姿勢の推定）は [5] で記述してあるのでここで省く。その後の2手間は下記のとおりである。

4.1 メッシュモデルのワーピング

ジェネリックなメッシュモデルを、推定した特徴点の三次元位置に合わせてワーピングする（図2）。 $N_{FP}=9$ 点の FAVRET 特徴点に相当するメッシュ頂点とメッシュの

† NHK 放送技術研究所

境にある幾つかの頂点 (固定頂点) の位置を固定し、図2で示す三角毎にメッシュをワーピングする:

- (A) ジェネリックモデルの各メッシュ頂点 $n, 0 \leq n < N_{vt}$ を三角の一つに割り当てる;
- (B) 頂点の座標 $\mathbf{x}^n \in R^3$ を三角の頂点の線形組み合わせとして表現する;
- (C) FAVRET 特徴点に相当する三角の頂点を、推定した特徴点に置き換えて、 \mathbf{x}^n を三角の頂点の同一組み合わせ $\hat{\mathbf{x}}^n$ に置き換える。

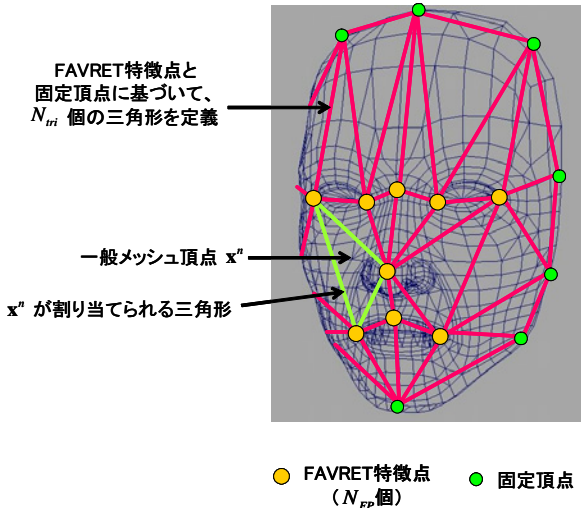


図2 メッシュのワーピング

4.2 テクスチャマッピング

メッシュ頂点毎に、テクスチャ ("UV") 画像へのポインタがあり、レンダリングの時にその頂点に映る色 (RGB 値) をコントロールする。登録用映像クリップから、対象人物の顔のテクスチャをモデルの UV 画像に写すことにより、レンダリングした時に対象人物の顔テクスチャがモデルの表面としてレンダリングされる。

しかし、登録用映像クリップの全フレームから顔のテクスチャを写すために、各フレームでの頭部姿勢によって調整する必要がある。テクスチャマッピングのアルゴリズムは下記のとおりである:

- (A) メッシュの多角形のすべてを三角形に変換する (三角の数は数千あり)。メッシュ頂点 $n, 0 \leq n < N_{vt}$ の UV ポインタを $(u^n, v^n), 0 \leq u^n, v^n \leq 1$ とする;
- (B) UV 画像の各画素 (u, v) を、4.1 節の(A)と同様に、メッシュ頂点のポインタ (u^n, v^n) を頂点とする $i = i(u, v)$ 番目の三角に割り当てる;
- (C) 映像クリップの k 枚目のフレームで推定した頭部姿勢を \hat{Q}^k 、特徴点の重心を $\bar{\mathbf{y}}^k$ として、 i 番目の三角の頂点が対応する位置を $\hat{\mathbf{r}}_k^{i,l} = \hat{Q}^k \hat{\mathbf{x}}^{i,l} + \bar{\mathbf{y}}^k, l = 0,1,2; 0 \leq k < N_{frames}$ として推定する;
- (D) (u, v) が対応する位置 $\hat{\mathbf{r}}_k(u, v)$ を、 $\hat{\mathbf{r}}_k^{i,l}, l = 0,1,2$ の 4.1 節の(A)と同様な線形組み合わせとして算出する;
- (E) そこでの RGB 値を $\mathbf{p}_k(\hat{\mathbf{r}}_k(u, v))$ とする;

- (F) ジェネリックモデルのテクスチャ画像 $\mathbf{t}_o(u, v)$ を

$$\mathbf{t}(u, v) = \frac{\gamma \mathbf{t}_o(u, v) + \delta(u, v) \sum_{k=0}^{N_{frames}-1} \alpha^k(u, v) \mathbf{p}_k(\hat{\mathbf{r}}_k(u, v))}{\gamma + \delta(u, v) \sum_{k=0}^{N_{frames}-1} \alpha^k(u, v)}$$

に置き換える。ここで γ, δ, α は重みである。

α は表面法線とカメラ軸との角度により設定する重みで、カメラ軸との直交線に近いパッチにテクスチャマッピングを制限することができる。そうすると、テクスチャの解像度を維持できるが、映像クリップに映っていない頭部姿勢を補間する性能が落ちる。

図3は 80 フレーム (正面→左横顔) の映像からマッピングしたテクスチャを示す。ここで重みの α は $\cos^2(\bullet)$ とした。鼻孔にアーティファクトが見えるが、そこは鼻の中の見えない場所に相当するなのでマッピングされない。



図3 テクスチャマッピングの例 (80 フレームから)

5. 今後の課題

形状推定には9点の FAVRET 特徴点しか使用しないので特徴点以外の場所に形状エラーが発生する。テクスチャが正しくマッピングされても、特徴点近傍の形状にエラーがあると、計測される特徴にもエラーが生じるので、より細かく形状を推定する必要がある。

全フレームから形状を推定しても、テクスチャをマッピングに少数のフレームを使用すると解像度が維持できると思われるが、フレーム選択アルゴリズムは要検討。

- [1] S. Clippingdale, 伊藤 崇之, “動画像の顔検出・追跡・認識への統一されたアプローチ,” 通学技報, PRMU98-200 (1999).
- [2] S. Clippingdale, T. Ito, “A Unified Approach to Video Face Detection, Tracking and Recognition,” Proc. ICIP'99, Kobe, Japan, 1999.
- [3] Y. Shan, Z. Liu, Z. Zhang, “Model-Based Bundle Adjustment with Application to Face Modeling”, Proc. ICCV2001, Vancouver, pp. 644-651 (2001).
- [4] 菅野裕介, 佐藤洋一, “顔変形を伴う3次元頭部姿勢の単眼推定”, 情報処理学会論文誌 CVIM, 1, 2, pp.41-49 (2008).
- [5] S. Clippingdale, 藤井真人, 八木伸行, “遮蔽とノイズのある二次元観測データから三次元顔特徴点推定の一検討”, 通学技報 PRMU2008-42, pp.133-138 (2008).