

映像から三次元顔モデルへのテクスチャマッピング におけるテクスチャ画像のアラインメント手法 Image Alignment for 3D Face Model Texture Mapping

クリッピングデル サイモン† 藤井 真人†
Simon Clippingdale Mahito Fujii

1. まえがき

映像クリップから三次元顔モデルを構築するために、複数のフレームよりモデルのテクスチャ ("UV") 画像へのテクスチャマッピングが必要であるが、各フレームでの顔の位置、サイズ、回転角度などの推定エラーを原因に、各フレームからの貢献部分がアラインメントせずにボケが出る。そのため、各フレームからの貢献画像のアラインメント手法を提案し、実験結果を示す。

2. 顔画像認識システムへの新規登録の自動化

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。実用化を目指し、新規人物の登録の自動化を図っている。

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。

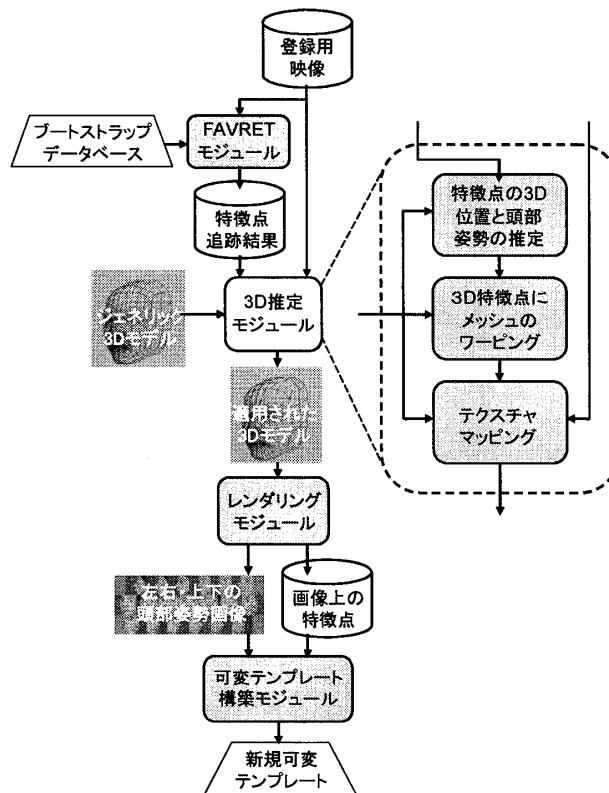


図1 新規自動登録システムの構成概要

3. テクスチャマッピング

上述の処理の流れの(B)(iii)のところ、テクスチャマッピングを行なう。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テクスチャ ("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。

テクスチャマッピングのアルゴリズム [5][6] は、登録用映像クリップの各フレームのどこから、三次元モデルの UV テクスチャ画像のどこへ写せば良いかを、各フレームで推定された頭部姿勢に基づいて推定して、写されるテクスチャの和をとる。写されるテクスチャのボケを減少させる工夫（重み付き）があるが、登録用映像クリップの各フレームで推定した頭部姿勢の誤差により、各フレームから写したテクスチャ情報がずれてしまい、その和としての UV テクスチャ画像にボケが生じることがある。

それで各フレームから写したテクスチャ情報の和をとる前に、アラインメントすることにした。しかし複数のフレームから、それぞれ異なる非線形マッピングで写された画像のア

† NHK 放送技術研究所

ラインメントの問題となるので、非線形で局所的なワーピングが必要である。

4. アラインメント手法

提案手法の処理の流れは下記のとおりである。提案手法は、多重解像度での局所的ワーピングを行なう。

- (A) テクスチャマッピングアルゴリズム[5][6]により、登録用映像クリップの各フレーム k から写されるテクスチャ $c^k(\mathbf{x})$ とする；
- (B) $c^k(\mathbf{x})$ のグレースケール情報から、多重解像度のピラミッドを構築する：

$$\begin{aligned} C_0^k(\mathbf{x}) &= G(c^k(\mathbf{x})) \\ C_l^k(\mathbf{x}) &= D(C_{l-1}^k(\mathbf{x}) \otimes g(\mathbf{x})) \\ 1 \leq l \leq L, 0 \leq k < N_{frames} \end{aligned} \quad (1)$$

ここでは $G(\cdot)$ はグレースケール変換、 $D(\cdot)$ はサブサンプルのオペレータ、 \otimes は畳み込みオペレータ、 $g(\mathbf{x})$ は2D ガウシアンフィルタである。

- (C) ピラミッドのトップレベル (L) で平均をとることにより、リファレンステクスチャを生成する：

$$R_L(\mathbf{x}) = \frac{1}{N_{frames}} \sum_{k=0}^{N_{frames}-1} C_L^k(\mathbf{x}) \quad (2)$$

- (D) 格子グリッド頂点にて、 $C_L^k(\mathbf{x})$ と $R_L(\mathbf{x})$ との位置ずれ (シフト) $\hat{\mathbf{s}}_L^k(\mathbf{x})$ をガボールウェーブレット係数の位相の差分から推定する[1][2]；
- (E) $\hat{\mathbf{s}}_L^k(\mathbf{x})$ を格子頂点の間に線形的に補間し、次に高い解像度 (ピラミッドレベル $L-1$) で各フレームからのテクスチャをずらす：

$$S_{L-1}^k(\mathbf{x}) = C_{L-1}^k(\mathbf{x} - I(\hat{\mathbf{s}}_L^k(\mathbf{x}))) \quad (3)$$

ここでは $I(\cdot)$ は線形補間とアップサンプルのオペレータである。

- (F) 各フレームのずらされたテクスチャから、レベル $L-1$ でのリファレンステクスチャを生成する：

$$R_{L-1}(\mathbf{x}) = \frac{1}{N_{frames}} \sum_{k=0}^{N_{frames}-1} S_{L-1}^k(\mathbf{x}) \quad (4)$$

- (G) (D)と同様に、格子グリッド頂点にて、 $S_{L-1}^k(\mathbf{x})$ と $R_{L-1}(\mathbf{x})$ との位置ずれ $\hat{\mathbf{s}}_{L-1}^k(\mathbf{x})$ をガボールウェーブレット係数の位相から推定する；
- (H) 次々のピラミッドレベルにて (E)(F)(G)を再帰的に繰り返す；
- (I) レベル0で得られた、フレーム k での累積シフト

$$\mathbf{r}_0^k(\mathbf{x}) = I(\hat{\mathbf{s}}_L^k(\mathbf{x}) + \hat{\mathbf{s}}_{L-1}^k(\mathbf{x}) + \dots + \hat{\mathbf{s}}_1^k(\mathbf{x})) \quad (5)$$

を用いて、RGB テクスチャをずらして平均をとる：

$$R_0(\mathbf{x}) = \frac{1}{N_{frames}} \sum_{k=0}^{N_{frames}-1} c^k(\mathbf{x} - \mathbf{r}_0^k(\mathbf{x})) \quad (6)$$

これで得られた平均テクスチャ $R_0(\mathbf{x})$ は、アラインメントアルゴリズムの出力テクスチャである。

4.1 実験結果

図2の左側は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[5]。図2の右側は提案手法を用いてマッピングした場合の結果を示す。



図2 テクスチャマッピングの例 (80フレームから)。左：従来のテクスチャマッピング手法の結果。右：提案手法でアラインしたときの結果。

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。

そこで提案手法により各フレームでのテクスチャの対応点情報を正確で細かく推定できれば、上記2(B)での顔形状の推定結果を修正できるか現在検討中である。

- [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, 柴田正啓, “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”, 情報技術フォーラム FIT2009 H-009.
- [6] S. Clippingdale, M. Fujii, M. Shibata, “Multimedia Databases for Video Indexing: Toward Automatic Face Image Registration”, *IEEE MIPR2009*, in *Proc. ISM2009*, pp.639-644, San Diego (2009)