

パーティクルフィルタとLevenberg-Marquardt法による 顔・人物姿勢追跡の検討

筒井 健斗^{1,a)} 溝部 諒¹ 右田 剛史¹ 尺長 健¹

概要：従来より、我々は顔追跡・認識融合系、人物姿勢追跡、およびジェスチャ認識を提案している。これらの追跡系では、パーティクルフィルタによる追跡を行い、粗密探索法によって精度の高い追跡を実現している。本稿では、追跡系の性能を更に向上させるためにパーティクルフィルタとLevenberg-Marquardt法による顔・人物姿勢追跡法を提案する。シミュレーション実験では、従来の粗密探索法の場合と比較して追跡精度が向上していることが確認できた。また、1フレーム当たりの処理時間も短くなっていることが確認できた。実画像実験においても追跡精度向上が確認でき、本手法の有効性を示すことができた。

1. はじめに

我々の研究室では、従来より顔追跡・認識融合系、人物姿勢追跡、およびジェスチャ認識に関する研究を行っている。本稿では、これらの性能を高めるためにパーティクルフィルタ [1] とLevenberg-Marquardt法 [2] (以下、LM法) を利用した追跡を検討する。

岡-尺長 [3] は3次元顔追跡を行う手法として、照明変動にロバストに対応できる3次元疎固有テンプレート追跡を提案し、中岸-尺長 [4] は登録人数100人の顔追跡・認識融合系を実現している。図1にその概要を示す。追跡系では、複数の解候補を保持し評価の高いものを選択するパーティクルフィルタによる追跡を行い、粗密探索法 [5] により精度の高い追跡を実現している。認識系では、追跡系で得られた姿勢によって正面向きに変形した入力画像と登録人物の画像を比較して認識を行う。つまり、追跡精度が向上すれば認識精度も向上することになる。

また、人物姿勢のような高次元の姿勢追跡に関する研究 [6] も現在行われている。近年では、安価で高性能な距離・画像センサを利用できるようになったため、RGB画像と距離画像を併用することにより、容易に人物の領域を抽出でき、人物の姿勢追跡を行うことができる。四宮ら [7] は、距離・画像センサを用いた追跡系を構成し、人物の上半身の姿勢追跡をパーティクルフィルタと粗密探索を用いて行っている。その人物姿勢追跡系の概要を図2に示す。加藤ら [8] はこの姿勢追跡に改良を加え、さらに追跡によ

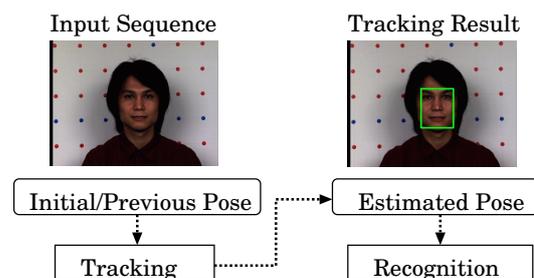


図1 追跡・認識融合系の概要

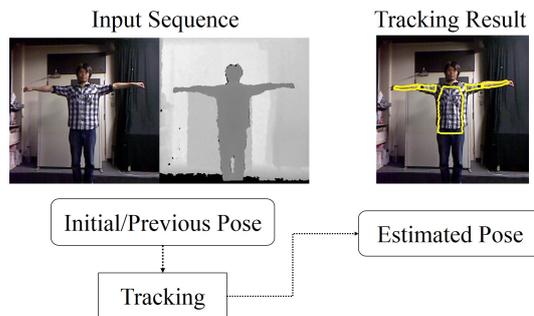


図2 人物姿勢追跡系の概要

て得られた姿勢情報を用いたジェスチャの認識法を提案している。このジェスチャの認識も顔認識と同様に追跡の精度が重要である。

従来の粗密探索法による追跡は、パーティクルフィルタを繰り返し行うことで精度を向上させている。しかし、この手法では多量の計算量が必要となる。そこで、本稿ではパーティクルフィルタにより選ばれた解候補をLM法によって最適化する方法を提案する。本手法により、計算量の削減、精度の向上を見込むことができると考えられる。本稿の実験では、従来法と本手法の追跡精度、処理時間を

¹ 岡山大学

Okayama University

^{a)} {tsutsui, mizobe, migita, shaku}@chino.cs.okayama-u.ac.jp

比較することで本手法の有効性を示す。

2. 顔・人物姿勢追跡

2.1 3次元形状モデル

本稿では、3次元形状モデルを使って追跡する。形状モデルを構成する点は物体座標系を用いて $\widetilde{\mathbf{M}}_O = [X_O \ Y_O \ Z_O \ 1]^T$ と表される（人体の関節モデルでは4章で述べるように拡張される）。任意の姿勢に対して形状モデルの各点の画像座標を計算するためには、物体座標系からカメラ座標系 $\widetilde{\mathbf{M}}_C = [X_C \ Y_C \ Z_C \ 1]^T$ 及び画像座標系 $\widetilde{\mathbf{M}}_I = [X_I \ Y_I \ 1]^T$ に変換する必要がある。この変換は姿勢行列 \mathbf{D} と内部パラメータ行列 \mathbf{K} を用いて次式で表される。

$$\widetilde{\mathbf{M}}_C = \mathbf{D} \widetilde{\mathbf{M}}_O \quad (1)$$

$$\widetilde{\mathbf{M}}_I = \widehat{\mathbf{K}} \mathbf{D} \widetilde{\mathbf{M}}_O, \text{ ただし } \widehat{\mathbf{K}} = [\mathbf{K} \mid \mathbf{0}] \quad (2)$$

2.2 評価関数

入力画像 \widehat{I} と追跡対象のテンプレート画像 I を姿勢行列 \mathbf{D} によって同じ向きにすることで比較する。各点 i の残差 d_i にロバスト関数（Geman-McClure 関数）を適用し、合計したものを評価値とする。

$$\epsilon = \sum_{i=1}^N \rho_c(d_i), \text{ ただし } \rho_c(x) = \frac{x^2}{c^2 + x^2} \quad (3)$$

$$d_i = \widehat{I}(\widetilde{\mathbf{M}}_{Ii}) - I(\widetilde{\mathbf{M}}_{Oi}) \quad (4)$$

ただし、 $\widehat{I}(\widetilde{\mathbf{M}}_{Ii})$ は画像座標 $\widetilde{\mathbf{M}}_{Ii}$ における画素値、 $I(\widetilde{\mathbf{M}}_{Oi})$ は点 $\widetilde{\mathbf{M}}_{Oi}$ における画素値、 c はノイズレベルに対応する定数である。計算量を減らすため、疎テンプレート追跡 [5] において式 (3) は顔領域の全画素にわたる総和ではなく、 N （最大 1024）点の疎な特徴点での総和である。また、式 (4) は照明変動にロバストに対応できる顔追跡や距離情報を用いた人物姿勢追跡では適宜拡張される。

2.3 パーティクルフィルタ

パーティクルフィルタ [1] は、時刻 $t-1$ の少数の姿勢候補（例えば 10 個）からそれぞれに対して乱数を加算することで時刻 t の多数の姿勢候補（例えば 1000 個）を作成し、評価値（式 (3)）の低い少数の姿勢候補を選択する。これを毎フレーム行うことで対象物体を追跡する。

2.4 粗密探索法

パーティクルフィルタで安定した姿勢追跡を行うためには多数の姿勢候補が必要となる。粗密探索法 [5] は、パーティクルフィルタで加算される乱数の生成範囲を段階的に狭めながら繰り返し行うものである。 s 段目の標準偏差 σ_s は $s-1$ 段目の標準偏差を半分にした $\sigma_{s-1}/2$ とする。このようにパーティクルフィルタを反復することにより少ない計算量で姿勢追跡の精度を高めることができる。



図 3 3次元顔モデル

2.5 Levenberg-Marquardt 法

式 (3) を LM 法により最小化することで、更に少ない計算量で精度を高めることができると考えられる。このために式 (3) を $\mathbf{r}(\mathbf{p})^T \mathbf{r}(\mathbf{p})$ に変形する。

$$\mathbf{r}(\mathbf{p}) = \begin{bmatrix} r_1 & r_2 & \dots & r_N \end{bmatrix}^T, \quad r_i = \frac{d_i}{\sqrt{c^2 + d_i^2}} \quad (5)$$

ここで、 \mathbf{p} は M 次元の姿勢パラメータのベクトル $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_M]$ である。LM 法では次式により姿勢候補 $\mathbf{p}^{(\tau)}$ を $\mathbf{p}^{(\tau+1)}$ に修正する。

$$\mathbf{p}^{(\tau+1)} = \mathbf{p}^{(\tau)} - (\mathbf{J}^T \mathbf{J} + C\mathbf{I})^{-1} \mathbf{J}^T \mathbf{r} \quad (6)$$

ここで、

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{r}}{\partial p_1} & \frac{\partial \mathbf{r}}{\partial p_2} & \frac{\partial \mathbf{r}}{\partial p_3} & \dots & \frac{\partial \mathbf{r}}{\partial p_M} \end{bmatrix}, \quad (7)$$

C はある正数（例えば 10^{-5} ）、 \mathbf{I} は単位行列である。姿勢候補 $\mathbf{p}^{(\tau)}$ が最適解から離れている場合は、 C を大きくすることで発散しにくくなり、最適解に近づくにつれて C を小さくすることで高速に収束させることができる。

3. 顔追跡

顔追跡では、3次元顔モデル（図 3）を使って姿勢パラメータを推定する。以下にその手法について述べる。なお本稿では、顔画像における表情変化など、物体の形状が変化する場合を考慮していない。

3.1 3次元疎固有テンプレート追跡

3次元疎固有テンプレート追跡 [3] では、照明変動にロバストに対応できる追跡を実現するために固有空間を用いる。固有空間とは、多様な照明条件で撮影された複数人の画像集合を効率的に表す低次元空間のことである。これを $\widetilde{\Phi} = [\Phi \ \bar{\mathbf{v}}]$ と表現する。ただし、 Φ は固有ベクトルの集合、 $\bar{\mathbf{v}}$ は平均画像である。これを使う場合、式 (4) の代わりに $\mathbf{d} = (\widetilde{\Phi} \widetilde{\Phi}^+ - \mathbf{I}) \mathbf{V}$ を用いる。ただし、 $\mathbf{d} = [d_1 \ \dots \ d_N]$ 、ベクトル \mathbf{V} は姿勢 \mathbf{D} に従って入力画像の顔領域を切り取ったものを正面向きに变形した画像である。固有空間で顔画像は正面向きに登録してあるため、この操作が必要となる。

3.2 姿勢空間

対象物体（顔）を剛体と仮定すると、対象物体の3次元

$$D = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) & t_x \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) & t_y \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

図 4 姿勢行列

$$\mathbf{n}_i^T = \frac{1}{z_i} (I_x, I_y) \begin{bmatrix} 1 & -x_i/z_i \\ & 1 & -y_i/z_i \end{bmatrix} \mathbf{K} \begin{bmatrix} \mathbf{I} \\ \frac{[[q]_o \widetilde{\mathbf{M}}_{O_i}]^T}{q_0} \begin{pmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & & \\ & q_0 & \\ & & q_0 \end{pmatrix} \end{bmatrix}, \text{ただし } [q]_o = \begin{bmatrix} q_1 & q_2 & q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{bmatrix},$$

$$(x_i \ y_i \ z_i)^T = \widehat{\mathbf{K}} \mathbf{T} \mathbf{R} \widetilde{\mathbf{M}}_{O_i}, \quad (I_x, I_y) \text{ は画像の勾配.}$$

図 5 \mathbf{n}_i の定義

姿勢は並進 3 自由度，回転 3 自由度の計 6 自由度となる．ここで， 4×4 の姿勢行列 D を次のように表す．

$$D = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (8)$$

ただし， \mathbf{R} は 3×3 の回転行列， \mathbf{t} は 3 次元の並進ベクトルである．この姿勢行列 D は 6 自由度の姿勢パラメータ $\mathbf{p} = [t_x \ t_y \ t_z \ q_1 \ q_2 \ q_3]^T$ により表される (図 4)．ここで， \mathbf{R} は四元数で表現し，その要素の一つは次式により表される．

$$q_0 = \sqrt{1 - q_1^2 - q_2^2 - q_3^2} \quad (9)$$

3.3 Levenberg-Marquardt 法による顔追跡

2.4 節の粗密探索法で得られる複数の姿勢パラメータ \mathbf{p} を初期値として，LM 法の反復公式 (式 (6)) により最適化する．式 (7) の \mathbf{J} は次式のように表される．

$$\mathbf{J} = \mathbf{W}(\widetilde{\Phi}\widetilde{\Phi}^+ - \mathbf{I})\mathbf{N} \quad (10)$$

ただし， $\mathbf{W} = \text{diag}(w_1, w_2, \dots, w_N)$ ， $w_i = c^2(d_i^2 + c^2)^{-3/2}$ ， $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_N]^T$ であり， \mathbf{n}_i は図 5 のように表される．

各姿勢候補の評価値 (式 (3)) は，図 6 のように LM 法で最適化され，一つの値に収束することが望ましい．しかし，図 7 のように複数の点に収束する場合もある．これは姿勢パラメータが局所解に陥ってしまったためと考えられる．

4. 人物姿勢追跡

人物の姿勢は関節が多く，オクルージョンの多い姿勢があり，RGB 画像のみでの追跡は困難である．そこで，距離画像を用いることで，RGB 画像のみを用いた追跡よりも高精度な追跡を行う．実際には RGB 画像をグレースケールに変換して計算を行っている．

本稿で追跡対象とするのは，人間の上半身である．上半身は胴体，左上腕，左下腕，右上腕，右下腕の 5 つの部位が左肩，左肘，右肩，右肘の 4 つの関節によって連結されているものとする [8]．

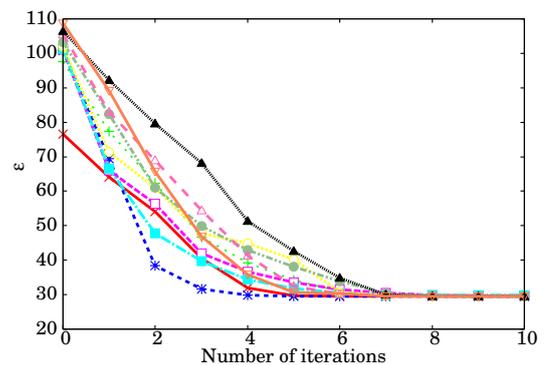


図 6 評価値の推移

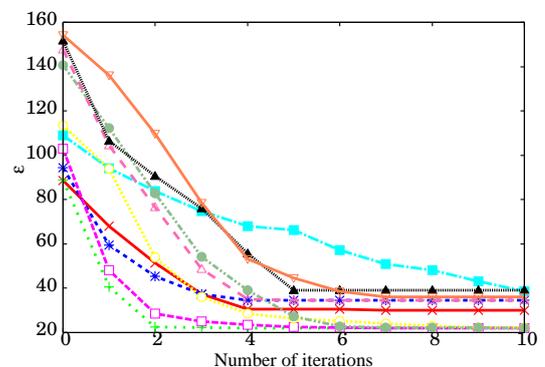


図 7 評価値の推移 (局所解)

4.1 座標系の定義

一般に，関節物体の姿勢は，主となる物体の姿勢と，それぞれの関節位置，関節角度により定義される．本稿では，上半身モデルの姿勢推定問題を考えるため，3 つの (同次) 座標系を考える．ここで，上腕と下腕については左右別々に座標系を考える必要があるが，左右の相互作用は考慮しないため，以下では 3 つの座標系について述べる．物体モデルを構成する点を胴体座標系，上腕座標系，下腕座標系の 3 つの座標系を用いて次のように記述する．

(1) $\widetilde{\mathbf{M}}_O$: 胴体座標系で表される座標

(2) $\widetilde{\mathbf{M}}_U$: 上腕座標系で表される座標

(3) $\widetilde{\mathbf{M}}_L$: 下腕座標系で表される座標

上腕座標系の点 $\widetilde{\mathbf{M}}_U$ と下腕座標系の点 $\widetilde{\mathbf{M}}_L$ は以下の式で

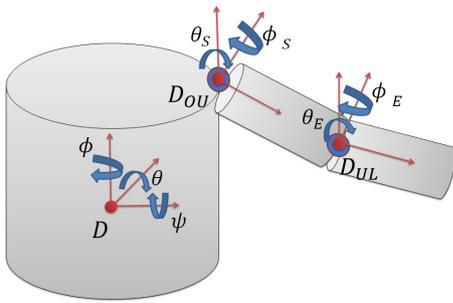


図 8 各座標系の関係図

胴体座標系の点に変換できる．

$$\widetilde{\mathbf{M}}_O = \mathbf{D}_{OU} \widetilde{\mathbf{M}}_U \quad (11)$$

$$\widetilde{\mathbf{M}}_O = \mathbf{D}_{OU} \mathbf{D}_{UL} \widetilde{\mathbf{M}}_L \quad (12)$$

\mathbf{D}_{OU} 、 \mathbf{D}_{UL} はそれぞれ上腕座標系から胴体座標系、下腕座標系から上腕座標系への座標変換行列である*1．それぞれの座標系の関係を図 8 に載せる．詳細については次節 4.2 にて述べる．胴体座標系から画像生成するために、2.1 節と同様に胴体座標系をカメラ座標系に変換する．

4.2 各剛体の表現と座標変換行列

胴体については、6 自由度 (3 次元並進 + 3 次元回転) の姿勢行列で取り扱う．胴体座標系からカメラ座標系への変換 \mathbf{D} を胴体の姿勢行列と考えることができる．

$$\mathbf{D} = \begin{bmatrix} \mathbf{R}_z^{\theta} \mathbf{R}_y^{\phi} \mathbf{R}_x^{\psi} & \mathbf{t}_O \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (13)$$

ここで式 (13) における $\mathbf{R}_z^{\theta} \mathbf{R}_y^{\phi} \mathbf{R}_x^{\psi}$ は胴体の回転を表す行列であり、それぞれの項については図 9 のように表される．また、 \mathbf{t}_O は胴体の並進を表すベクトルである．

上腕の姿勢は、胴体の姿勢 \mathbf{D} と肩を中心とした回転の合成により得られる．胴体座標系における肩関節の位置を \mathbf{t}_S とする．これは追跡する人物によって異なる既知パラメータである．上腕座標系の原点は肩関節にあり、上腕の主軸を x 軸とし、肩を中心に y 軸、 z 軸回転を考える．肩での回転行列を \mathbf{R}_S とすると、胴体座標系への座標変換行列 \mathbf{D}_{OU} は以下の式 (14) で表せる．

$$\mathbf{D}_{OU} = \begin{bmatrix} \mathbf{R}_z^{\theta_s} \mathbf{R}_y^{\phi_s} & \mathbf{t}_S \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (14)$$

下腕座標系から胴体座標系への変換行列は、上腕から胴体への座標変換行列 \mathbf{D}_{OU} と、肘を中心とした回転の合成により得られる．上腕座標系における肘の位置を \mathbf{t}_E とする．これは追跡する人物によって異なる既知パラメータである．下腕座標系の原点は肘関節にあり、下腕の主軸を x 軸とし、肘を中心に y 軸、 z 軸回転を考える．肘での回転

*1 本稿では \mathbf{D}_{OU} 等の添字は、行列の右に掛るベクトルの座標系を右に、積の結果の座標系を左としており、順序が [8] とは逆になっている．

行列を \mathbf{R}_E とすると、上腕座標系への座標変換行列 \mathbf{D}_{UL} は以下の式 (15) で表せる．

$$\mathbf{D}_{UL} = \begin{bmatrix} \mathbf{R}_z^{\theta_E} \mathbf{R}_y^{\phi_E} & \mathbf{t}_E \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (15)$$

4.3 3次元疎テンプレートの作成

3次元疎テンプレートによる姿勢追跡 [7] では、現時刻 t の姿勢を推定する時、まず、時刻 $t-1$ の姿勢において各部位に対応したテクスチャ付き人体モデルを投影する．このとき隠面処理を行うことにより、モデルの表側の画素値と距離値を取得することができる．2.2 節と同様に、式 (3) の総和は、局所的に輝度値が最大・最小となる点 (疎な特徴点) のみで行う．このため、各点の 3 次元座標 $(X_i Y_i Z_i)$ と画素値 I_i の組 $\{(X_i Y_i Z_i; I_i)\}$ を作成する．同様に隠面処理に用いた Z バッファにおいて、局所的に距離値が最大・最小となる点を選び、 $\{(X'_i Y'_i Z'_i)\}$ を得る．

4.4 3次元疎テンプレートマッチング

作成した疎テンプレートを用いて、2.3、2.4 節で述べたパーティクルフィルタと粗密探索を用いて追跡を行っている．ただし、RGB 画像、距離画像を併用して 3 次元疎テンプレートマッチングを行う．

従来のカスケード姿勢推定 [8] では、胴体 (6 自由度)、右上腕・下腕 (2 + 2 自由度)、左上腕・下腕 (2 + 2 自由度) の順に姿勢の推定を行っている．本稿では、この従来法に LM 法を追加して姿勢の最適化を行う．

4.5 距離・画像センサを用いた追跡

本節では、RGB 画像と距離画像のそれぞれの評価値の定義とその併用について述べる．

4.5.1 RGB 画像に対する評価値

第 k 姿勢候補における入力画像の評価値 e_k^C は、テンプレートの i 番目の点の画素値 I_i と、その点の入力 RGB 画像の画素値 $\hat{I}(X_i Y_i Z_i; \mathbf{D}_k)$ を用いて式 (16) で求める．ここで、 \mathbf{D}_k は第 k 姿勢候補で、胴体、上腕、下腕の姿勢行列 \mathbf{D}_k 、 $\mathbf{D}\mathbf{D}_{OUk}$ 、 $\mathbf{D}\mathbf{D}_{OUk}\mathbf{D}_{ULk}$ を表す．

$$e_k^C = \sum_{i=1}^N \rho_c(d_i), \quad d_i = \beta \hat{I}(X_i Y_i Z_i; \mathbf{D}_k) - \alpha I_i \quad (16)$$

ただし、

$$\alpha = \frac{1}{\sum_{i=1}^N I_i}, \quad \beta = \frac{\alpha \sum_{i=1}^N I_i^2}{\sum_{i=1}^N I_i \hat{I}(X_i Y_i Z_i; \mathbf{D}_k)}$$

なお、 α はテンプレート画像の明度の正規化係数、 β は入力画像の明度の正規化係数、 N は疎な特徴点の数である．また c は定数であり、ここでは、 $c = 0.5/N$ とする．

4.5.2 距離画像に対する評価値

第 k 姿勢候補における入力距離画像の評価値 e_k^D は、RGB

$$\mathbf{R}_x^\psi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{bmatrix}, \quad \mathbf{R}_y^\phi = \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix}, \quad \mathbf{R}_z^\theta = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

図 9 回転行列

画像と同様にして、テンプレートの i 番目の点の距離値 Z_{ik} とその点の入力距離値 $\hat{Z}(X'_i Y'_i Z'_i; D_k)$ を用いて式 (17) で求める。

$$\epsilon_k^D = \sum_{i=1}^{N'} \rho_{c'}(d'_i), \quad d'_i = \hat{Z}(X'_i Y'_i Z'_i; D_k) - Z_{ik} \quad (17)$$

なお、距離センサの出力が mm を単位とする整数値であることを勘案し、予備実験により、 $c' = 5.0$ とした。

4.5.3 RGB 画像と距離画像を併用した評価値

RGB 画像に対して算出した評価値 ϵ_k^C と距離画像に対して算出した評価値 ϵ_k^D を併用して姿勢推定を行うことで、より適切な姿勢を推定できると考えられる。

第 k 姿勢候補における評価値 ϵ_k を両者の重み付き和によって式 (18) で定義する。

$$\epsilon_k = (1 - \lambda)\epsilon_k^C + \lambda\epsilon_k^D \quad (18)$$

ここで、 λ は定数 ($0 \leq \lambda \leq 1$) である。

4.6 Levenverg-Marquardt 法による人物姿勢追跡

14 次元の姿勢パラメータを同時に推定すると、追跡に時間がかかるため、従来法ではカスケード姿勢推定により、胴体 (6 自由度)、右上腕下腕 (4 自由度)、左上腕下腕 (4 自由度) と分割して追跡を行っていた。しかし、LM 法による最適化を用いることによって、14 パラメータの同時推定を行うことができる。また、より高次元のパラメータにおいても、高速に最適化できると考えられる。

本稿では、LM 法の式 (6) により姿勢候補 \mathbf{p} (14 次元の姿勢パラメータのベクトル) を最適化して、精度を向上させる。このために、式 (18) を $\mathbf{r}(\mathbf{p})^\top \mathbf{r}(\mathbf{p})$ の形にする。

$$\mathbf{r}(\mathbf{p}) = \begin{bmatrix} r_1 & r_2 & \dots & r_N & r'_1 & \dots & r'_{N'} \end{bmatrix}^\top \quad (19)$$

ただし、 $r_i = \sqrt{1 - \lambda}d_i / \sqrt{c'^2 + d_i^2}$ 、 $r'_i = \sqrt{\lambda}d'_i / \sqrt{c'^2 + d_i'^2}$ である。式 (19) の前半は ϵ_k^C に対応し、後半が ϵ_k^D に対応する。

また、 $\frac{\partial \mathbf{r}}{\partial p_n}$ については以下の式で近似を行う。

$$\frac{\partial \mathbf{r}}{\partial p_n} \approx \frac{\mathbf{r}(\mathbf{p} + \mathbf{e}_n \epsilon) - \mathbf{r}(\mathbf{p} - \mathbf{e}_n \epsilon)}{2\epsilon} \quad (20)$$

ただし、 \mathbf{e}_n は 14 次元のベクトルで、 n 番目の要素が 1、その他の要素が 0 となるベクトルである。 ϵ は各パラメータに応じた微小定数である。

表 1 実験条件

項目	設定値
姿勢候補数	1000
選択する姿勢候補数	10
式 (3) の N	1024

表 2 初段の乱数の生成範囲 (σ)

t_x	t_y	t_z	q_1	q_2	q_3
5.0	5.0	10.0	5.0	5.0	2.5



図 10 Simulation1 (変動量小)



図 11 Simulation2 (変動量大)

5. 実験

5.1 顔追跡実験

顔追跡の追跡精度が LM 法によって向上していることを確認する。まずシミュレーション実験を行い、実験条件 (粗密探索法の段数, LM 法の反復回数) を変更して追跡精度を比較する。次に、中岸らの顔認識系 [4] に組み込むことで実画像での実験を行い、提案手法の有効性を示す。実験に用いる各種パラメータを表 1, 2 に示す。また本稿では、各フレームで姿勢候補を複数選択している。そのため、そのフレームでの代表となる推定姿勢は、評価値 (式 (3)) に従った重みを用いて重み付き平均を求めたものとする。

5.1.1 シミュレーション実験

シミュレーション実験では、3 次元顔モデルと動きを表す並進・回転の姿勢 6 パラメータを与えて投影し追跡対象の画像を生成する。このとき、変動量の異なる 2 つの画像列 (Simulation1 (変動量小), Simulation2 (変動量大)) を生成する (図 10, 11)。各フレームでの正解姿勢 \mathbf{R}_0 と推定姿勢 \mathbf{R} との差は、回転推定誤差を用いて次式で評価する。

$$\arccos \frac{\text{tr}(\mathbf{R}_0 \mathbf{R}^\top) - 1}{2} \quad [\text{deg}] \quad (21)$$

表 3 平均回転推定誤差 [deg] と処理時間 [ms] (Simulation1)

	PF1LM0	PF3LM0	PF1LM10	PF3LM10
P1	3.374	1.100	0.868	0.738
P2	3.026	1.083	0.725	0.635
P3	3.189	1.026	0.658	0.636
P4	3.273	1.022	0.740	0.669
P5	2.928	1.174	0.801	0.690
P6	3.055	0.998	0.671	0.638
P7	3.268	1.192	0.881	0.800
P8	3.681	1.150	0.742	0.706
P9	3.274	1.115	0.845	0.766
Time	30ms	75ms	45ms	85ms

表 4 平均回転推定誤差 [deg] と処理時間 [ms] (Simulation2)

	PF1LM0	PF1LM10	PF1LM30
P1	失敗	0.860	0.851
P2	5.480	0.746	0.716
P3	5.053	0.740	0.721
P4	4.317	0.683	0.673
P5	失敗	0.785	0.766
P6	失敗	0.677	0.670
P7	失敗	失敗	0.928
P8	4.506	0.749	0.755
P9	5.237	0.898	0.819
Time	30ms	45ms	60ms

まず、様々な人物の顔モデルを使って追跡精度を確認する。今回は9人の顔モデルを用いる。本手法では粗密探索法(2.4節)とLM法(3.3節)を組み合わせる追跡を行う。従来法は、粗密探索法3段、LM法反復回数0回(LM法を用いていない)なのでPF3LM0と表現する。

表3はSimulation1で実験したときの平均回転推定誤差と1フレーム当たりの処理時間を示している。粗密探索法段数とLM法反復回数が増加すると、誤差が減少していることが確認できる。また、PF3LM0とPF1LM10を比較すると、PF1LM10の方が追跡精度と処理時間のどちらも良いことが確認できる。

次にSimulation2で実験したときの追跡精度を比較する。表4がその実験結果である。Simulation2は顔の動きが大きいのでPF1LM0では追跡失敗するものが多数あるが、PF1LM10では精度の高い追跡が行えている。しかし、P7は追跡失敗している。これはLM法が10回の反復では収束していないためであると考えられる。PF1LM30で行ったところ、全て精度の高い追跡を行うことができていることが確認できる(ただし、P8は少し悪化)。

5.1.2 実画像実験

次に、実画像を用いて実験する。中岸らの認識系[4]では登録人物を正面向きに登録しているため、追跡系で得られた姿勢で入力画像を正面向きに變形させて認識を行う。実画像では追跡精度を直接確認できないため、認識の成功率によって追跡精度を評価する。そこで、異なる50種類の

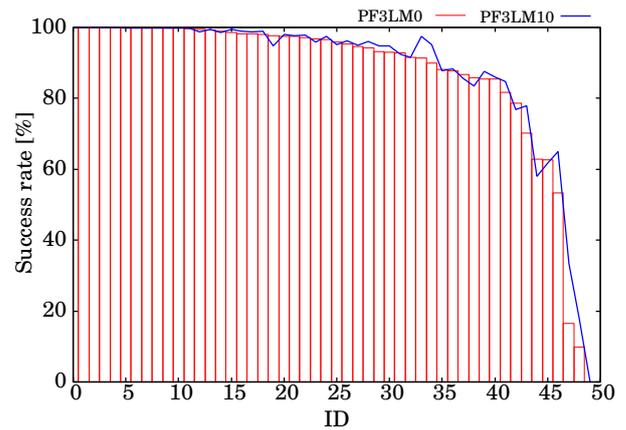


図 12 認識成功率 (全 500 フレーム)

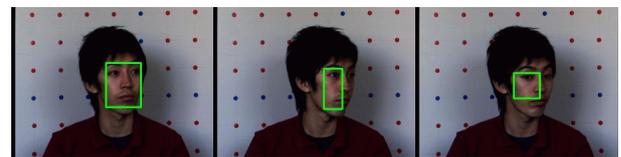


図 13 追跡失敗例

実画像(各500フレーム)で追跡を行い、フレーム毎に個人認識が成功しているかどうかを確認する。実験条件は、PF3LM0(従来法)とPF1LM10(本手法)で比較する。

図12はフレーム毎の成功率を示したものである。PF3LM0の成功率(赤い棒グラフ)が高い順にソートし、PF1LM10の成功率(青い折れ線グラフ)と比較できるようにしている。この結果から、PF3LM0に比べてPF1LM10の方が25個の実動画で成功率が向上、16個の実動画で低下、9個の実動画で同等であることを確認した。成功率が向上したのは、本手法によってより正しい姿勢が得られたためであると考えられる。低下したのは、LM法の初期値(パーティクルフィルタで選ばれた姿勢候補)が正しくなかったため、局所解に陥ったためであると考えられる。ただし、どちらの実験条件でも1個の実動画で顔の検出失敗、2個の実動画で顔が横または下に向きすぎて追跡失敗(図13)した。また、成功率の低い実動画の人物は、他の人物に認識されることが多かった。

実画像実験によって、処理速度が速いPF1LM10で従来法と同等の成功率を出すことができることが確認できた。

5.2 人物姿勢追跡実験

次に、人物姿勢追跡において、本手法の効果をj確認する。実験はシミュレーション画像と実動画を用いて、従来法と本手法での追跡性能の比較を行う。シミュレーション実験では、精度向上の確認を行い、実動画実験では提案手法が従来よりも効率的であることを確認する。

本稿では、距離・画像センサとしてKinectを用いている。距離情報はmm単位で表され、500mm以上離れた対

表 5 実験条件

項目	設定値
姿勢候補数	1500
式 (16) の N , 及び式 (17) の N' の値	128
評価値 ϵ における距離情報の比率 (λ)	0.90

表 6 初段の乱数の生成範囲 (σ)

胴体						肩		肘	
t_x	t_y	t_z	θ_O	ϕ_O	ψ_O	ϕ_S	θ_S	ϕ_E	θ_E
5.0	0.5	5.0	1.0	1.0	1.0	8.0	8.0	15.0	15.0

象について距離情報がリアルタイムで得られる。

人物モデルは, KinectFusion[9] を用いて作成した 3 次元モデルを用いる。

5.2.1 シミュレーション追跡実験

追跡精度の確認のために, 7 種類の動画でシミュレーション実験を行った。シミュレーション実験では, Kinect の入力情報の擬似表現として距離情報の分解能を粗くし, 入力距離値を 5[mm] 単位としている。

シミュレーション実験では, 従来のパーティクルフィルタによる追跡と, LM 法を追加した追跡の比較実験を行う。実験に用いた各種パラメータを表 5, 6 に示す。精度評価の尺度として, 右上腕, 下腕のそれぞれの主軸 (x 軸) と正解姿勢の主軸が成す角度を用いる。

各動画に対して 5 回試行を行い, その各部位に対する平均誤差を表 7, 表 8 に示す。表 7 は LM 法の反復回数 0 回 (従来法), 表 8 は LM 法の反復回数 50 回のものである。

表 7 と表 8 を比較してみると, 表 8 の*を除いて誤差が小さくなっていることが確認できる。段数 1(PF1) での実験では, 従来法に比べて大きく誤差を低減することができていることが分かる。従来法では表 7 の太字のように誤差が大きく, 追跡が失敗に終わっていた。しかし, LM 法を追加した本手法では, no5, 7 については少しブレがあるものの追跡を行うことが可能となった。段数 2(PF2) では, 本手法では表 8 の太字以外の場合に平均 1[deg] 以下の高精度な追跡が行えている。

また, 各部位についての誤差の平均をとりグラフにしたものが図 14 である。図 14 は LM 法の反復回数を 0 回 (従来法), 15 回, 50 回 (それぞれ LM0, LM15, LM50 と表す) として実験を行い, 比較したものである。図 14 より, LM15 についても十分に誤差が低くなっていることが分かる。LM 法は収束が早い, LM50 でさらに精度が上がっていることから, LM15 では完全な収束はしていないことが分かる。その原因として距離画像にノイズがあることが考えられる。しかし, LM15 から LM50 への精度の差が小さく, 追跡時間のことも考慮すると, LM15 が有効であると考えられる。

5.2.2 実動画追跡実験

実動画では, 本手法が従来よりも効率的に追跡できるこ

表 7 各部位の誤差 [deg](PFxLM0)

段数 (x)	動画	右上腕	左上腕	右下腕	左下腕
1	no1	4.340	3.476	5.214	4.170
	no2	3.416	3.539	4.183	3.354
	no3	3.142	3.378	4.195	5.115
	no4	4.557	4.894	6.306	6.764
	no5	3.449	3.510	11.433	4.441
	no6	6.016	4.121	21.360	4.232
	no7	3.910	5.123	9.615	11.832
2	no1	1.692	1.475	1.609	1.475
	no2	1.582	1.553	1.517	1.537
	no3	1.330	1.315	1.239	1.457
	no4	1.737	1.693	1.662	1.560
	no5	1.190	1.496	1.711	1.613
	no6	1.242	1.729	1.464	1.702
	no7	1.376	1.185	2.149	1.456
3	no1	1.007	0.969	0.679	0.718
	no2	0.867	0.873	0.669	0.660
	no3	0.973	0.790	0.576	0.611
	no4	1.038	1.162	0.744	0.748
	no5	0.698	0.917	0.651	0.671
	no6	0.822	0.999	0.706	0.682
	no7	0.772	0.705	0.691	0.619

表 8 各部位の誤差 [deg](PFxLM50)

段数 (x)	動画	右上腕	左上腕	右下腕	左下腕
1	no1	2.627	1.459	2.313	1.188
	no2	1.501	1.223	1.254	0.938
	no3	2.192	1.335	2.430	0.977
	no4	1.757	1.507	1.648	1.133
	no5	1.549	1.559	2.055	1.578
	no6	3.381	1.349	9.960	0.929
	no7	2.279	1.474	4.389	2.188
2	no1	1.220	0.927	0.818	0.759
	no2	0.949	0.999	0.732	0.864
	no3	0.974	0.859	0.626	0.614
	no4	1.077	1.008	0.772	0.752
	no5	0.759	0.951	0.732	0.896
	no6	0.884	0.957	0.930	0.766
	no7	0.961	0.817	0.914	0.927
3	no1	0.896	0.802	0.557	0.607
	no2	0.806	0.746	0.581	0.597
	no3	0.781	0.791*	0.562	0.558
	no4	1.051*	0.902	0.659	0.611
	no5	0.604	0.817	0.540	0.545
	no6	0.762	0.877	0.648	0.532
	no7	0.750	0.557	0.671	0.471

とを確認する。実験条件はシミュレーションと同様に表 5, 6 である。粗密探索の段数を 1(PF1) として, 実動画の追跡実験を行った。LM 法の反復回数は, シミュレーションで有効と考えられた 15 回とした。7 種類の実動画について実験を行い, 従来法で追跡できる動画は本手法でも全て追跡できることを確認した。また, 従来法では失敗してい

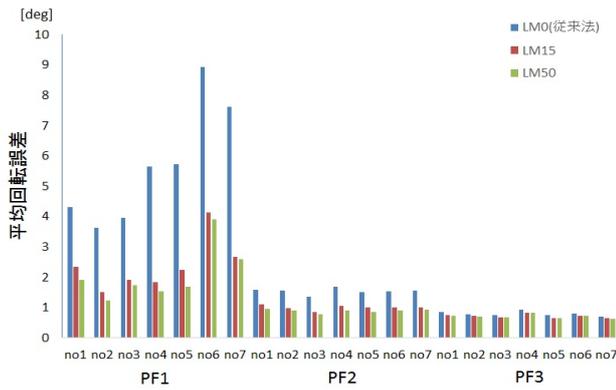


図 14 全関節の平均回転誤差

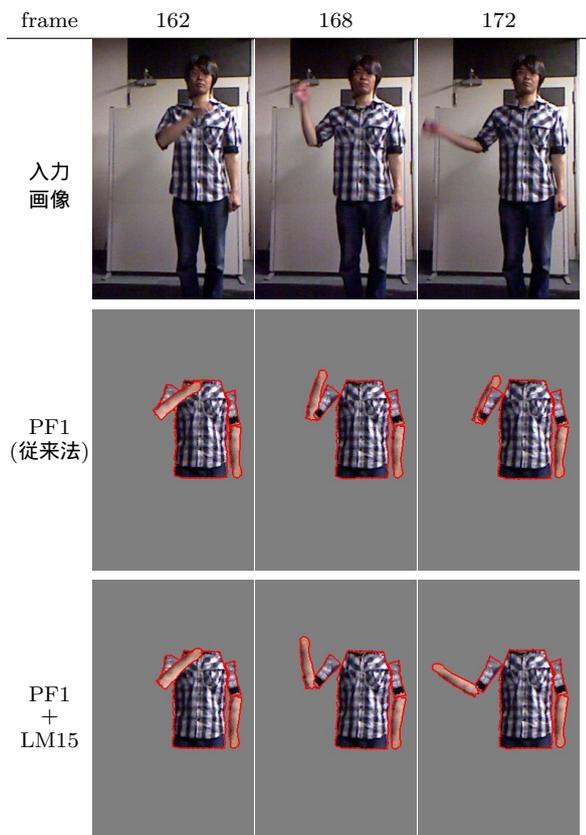


図 15 実動画実験における追跡結果

たが、本手法で成功した一例が図 15 である。162 フレーム目までは、従来法 (PF1LM0)、本手法 (PF1LM15) とともに追跡できている。しかし、168 フレーム目において、従来法は右下腕が右上腕に引き寄せられ、その後追跡に失敗している。一方、本手法では LM 法によって、右下腕の位置が修正され、正しく追跡できていることから、精度が向上していることが分かる。また、従来では図 15 の動作の追跡を安定させるために段数 3 として行っていたので、1 フレーム毎に平均 328[ms] かかっていた。PF1LM15 の追跡時間は 1 フレーム毎に平均 228[ms] で追跡が行え、従来よりも効率よく追跡が行えた。

6. まとめ

本稿では、従来から研究してきた顔追跡・認識融合系、人物姿勢追跡及びジェスチャ認識の性能を高める手法として、パーティクルフィルタと LM 法を利用した追跡を検討し、その有効性を示した。具体的には、顔追跡・人物姿勢追跡の粗密探索法で得られる複数の姿勢パラメータを初期値として LM 法の反復公式により最適化を行った。実験では、粗密探索法の段数、LM 法の反復回数をそれぞれ変更して追跡精度を確認した。シミュレーション実験の結果、従来の粗密探索法の場合と比較して追跡精度が向上していることが確認できた。また、1 フレーム当たりの処理時間も短くなることが確認できた。実画像実験においても追跡精度向上が確認でき、本手法の有効性を示すことができた。

今後の課題として、顔追跡では顔の向きが正面から大きくずれた際の追跡安定性が挙げられる。人物姿勢追跡では、オクルージョンに強い評価関数の設計や、肩関節の同時推定が考えられる。また、顔・人物姿勢追跡ともに GPU による実時間追跡の実現などが挙げられる。

参考文献

- [1] 松原康晴, 尺長 健: 疎テンプレートマッチングとその実時間物体追跡への応用, 情報処理学会論文誌, Vol. 46, No. 1, pp. 60-71 (2005).
- [2] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P.: *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press (2007).
- [3] Oka, Y. and Shakunaga, T.: Real-time face tracking and recognition by sparse eigentracker with associative mapping to 3D shape, *Image and Vision Computing*, Vol. 30, No. 3, pp. 147-158 (2012).
- [4] Chugan, H. and Shakunaga, T.: Hierarchical Approach to Weight Equations in Face Tracking and Recognition Framework, *FG2013* (2013).
- [5] Oka, Y., Kuroda, T., Migita, T. and Shakunaga, T.: Tracking 3d pose of rigid object by 3d sparse template matching, *In Proceedings of International Conference on Image and Graphics (ICIG2009)*, pp. 390-397 (2009).
- [6] Bray, M., Loller-Meier, E. and Gool, L. V.: Smart Particle for High-Dimensional Tracking, *Computer Vision and Image Understanding*, Vol. 106, pp. 116-129 (2007).
- [7] 四宮洋平, 加藤秀章, 尺長 健: Kinect を用いた 3 次元疎テンプレートによる人物姿勢追跡, 電子情報通信学会技術研究報告 PRMU2011-264, pp. 153-158 (2012).
- [8] 加藤秀章, 小林宏章, 山根 亮, 尺長 健: カスケード姿勢推定に基づく距離・画像センサを用いたジェスチャ認識, 情報処理学会研究報告, Vol. 2014-CVIM-190, No. 14 (2014).
- [9] Newcombe, R. A., Davison, A. J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D. and Fitzgibbon, A.: KinectFusion: Real-Time Dense Surface Mapping and Tracking, *Proc. International Symposium on Mixed and Augmented Reality (ISMAR2011)*, pp. 127-136 (2011).