

疎固有テンプレート追跡による実時間顔認識の検討

岡田 直城 武村 和宙 尺長 健

岡山大学大学院 自然科学研究科
〒700-8530 岡山市津島中 3-1-1

E-mail: {takemura,shaku}@chino.cs.okayama-u.ac.jp

あらまし: 我々の研究室では、従来から固有空間を用いた顔認識法の研究と、疎テンプレートコンデンセーション法 [1] による顔の実時間追跡の研究を並行して進めてきた。ここで、顔追跡と顔認識の両者で同じ固有空間を用いることができれば、両者の特長を活かした融合系を実現できると考えられるが、両者の融合には、2つの問題を解決する必要がある。まず、顔認識で要求される顔切出し精度と顔追跡で達成可能な精度のギャップを解消する必要がある。また、自然環境で顔追跡を行う際に生じる照明変動・姿勢変動・表情変化・隠れなどへの対策が必要である。本稿では、この2つの問題に対する解決するための指針を示すとともに、この指針に沿った具体的に系を構成し、その有効性を検証する。

Face Recognition by Means of Sparse Eigentemplate Tracker

Naoki Okada, Kazumichi Takemura, Takeshi Shakunaga

Department of Computer Science, Okayama University
3-1- 1 Tsushima-naka, Okayama, 700-8530, JAPAN

Abstract: This paper discusses a face recognition method implemented in the real-time framework of sparse template tracker. When an eigenface that is used in face recognition is also available for template tracker, both the tracking and the recognition problems are formally discussed in the unified framework. The unification, however, requires a pragmatic design of interface between the tracker and the recognition modules. This paper provides both a generic and specific solutions for the problem.

1 はじめに

顔画像認識では従来から固有空間が広く利用されている [2, 3, 4] が、入力画像中上での顔の切出し精度が認識性能に大きく影響することが知られている。このため、従来の顔認識系では、照明条件や顔の姿勢を制約することにより、顔認識に十分な切出し精度を実現している。一方、ヒューマンインタフェース改善を目指す立場からは、このような制約をなくして、より自然な状況で顔認識を行う必要がある。

一方、実時間で安定的な追跡を行う方法として、松原-尺長 [1] は疎テンプレートコンデンセーション法を提案している。この方法は、1枚の画像から作成した疎テンプレートを用いた追跡法であるが、テンプレートの代りに、画像集合から作成した固有空間を用いることもできる。松原-野口-尺長 [5] は、特定人物を、姿勢固定で異なる照明条件で撮影した画像集合から固有空間を構成し、疎テン

プレート（疎固有テンプレートと呼ぶ）を作成することにより、照明変動に対してロバストな顔追跡系を実現できることを報告している。

ここで、固有空間を特定個人の顔画像から構成するのではなく、何人かの登録人物の幾つかの姿勢からなる顔集合から事前に構成し、疎固有テンプレートを作成できれば、登録人物の変化に応じて、有効な顔追跡系を構成できる。一方、この固有空間を人物識別においても使用することができれば、顔追跡系と顔認識系の融合を原理的に実現できると考えられる。

本稿では、二つの系が原理的に融合可能であることを示した後、融合を妨げる要因を二つの問題に集約し、そのそれぞれについて解決策を提案する。最後に、これを実装した系により、顔追跡系と顔認識系の融合の有効性を検証する。

2 顔追跡と顔認識の融合

2.1 固有顔を用いた顔画像認識

複数人物の登録画像を学習画像として固有空間（固有顔とも呼ばれる）を作成し、識別処理をこの固有空間の中で行うことが広く行われている。本稿では、固有空間中での識別を、単純な最近傍法で取扱い、入力画像を固有空間に射影した後、固有空間内で最も距離に近い登録人物 h を選択する。

自然環境下で撮影された画像は撮影環境によって様々に変化する。ここで、輝度の変化に対応するため、正規化固有空間 [3] を用いる。（本稿では、追跡系と認識系の融合を課題とする目的から、それ以外の大きな照明変動はないものとする。）

以下に本稿で用いる記号の定義をまとめておく。正規化固有空間は、 N 次元画像 $\mathbf{X} (\neq 0)$ の正規化画像 $\mathbf{x} = \mathbf{X}/1^T \mathbf{X}$ を用いて構成され、 $(\bar{\mathbf{x}}, \Phi)$ により明示的に示す。ここで、 $\bar{\mathbf{x}}$ は固有空間の中心、 Φ は固有軸を示す。固有空間は、登録人数 H 、姿勢数 K の画像 HK 枚から構成したものを用いる。このとき、入力画像 \mathbf{X} から得られる正規化画像 \mathbf{x} の $(\bar{\mathbf{x}}, \Phi)$ への射影 \mathbf{x}^* は次式で計算される。

$$\mathbf{x}^* = \Phi^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (1)$$

このとき、固有顔法の識別結果 h は次式で表される。

$$h = \arg \min_{1 \leq h \leq H} \min_k d(\mathbf{x}^*, \mathbf{x}_{hk}^*) \quad (2)$$

$$\text{where } d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})$$

ここで、 \mathbf{x}_{hk}^* は人物 h の姿勢 k におけるプロトタイプである。

2.2 疎固有テンプレートによる顔追跡

実時間でテンプレート追跡を行う方法として、松原-尺長 [1] は疎テンプレートコンデンセーション法を提案している。また、松原ら [5] は、テンプレートを固有空間に拡張できることを示すとともに、照明変動に対してロバストな追跡が実現できることを示している。この方法では、確率的枠組としてコンデンセーション法 [6] を用いることにより、ロバストな追跡を実現している。ここで、コンデンセーション法では、多数のサンプル（パーティクル）を姿勢空間内で伝播させることを基本としている。疎固有テンプレートコンデンセーション法では、以下に示すように **drift**、**diffuse**、**measure** の各ステップを構成することにより、疎テンプレートマッチングに基づく確率的追跡を実現している。なお、以下の記述においては、時刻 t における L 個のサンプル姿勢からなる集合を $\{\mathbf{s}_t^{(l)}\}$ とし、各

サンプルの尤度を $\pi_t^{(l)}$ とする。また、固有空間は顔認識で用いる固有空間 $(\bar{\mathbf{x}}, \Phi)$ を用いる。

drift ステップでは、前時刻 $t-1$ のサンプルの中から尤度の高い $L/20$ 個のサンプルから尤度の比に従って L 個を選択する。このとき、選択されたサンプルの半分については、等速度運動を仮定して並進パラメータを予測したものを用いる。

diffuse ステップでは、選択された L 個のサンプルの各姿勢パラメータに対して、標準偏差を σ とするランダムなガウスノイズを付加する。

measure ステップでは、**diffuse** ステップで作成されたサンプル姿勢 l に対応する画像の位置を計算することにより、入力画像 \mathbf{x}_t を得ることができ、疎固有テンプレートマッチングを行う。具体的には、各 \mathbf{x}_t 中の疎な画素集合（全画素の 1% 程度）を固有空間 $(\bar{\mathbf{x}}, \Phi)$ に射影し、得られる射影残差中の、予め選択された 2% 程度の画素に対応するものに、ロバスト関数（Geman-McClure 関数）を適用して積算したものを評価値 $\epsilon^{(l)}$ として計算することができる。このとき、 $\epsilon^{(l)}$ の上位 $L/20$ 個以外のサンプルの $\epsilon^{(l)}$ を 0 とする。次に、各サンプルの尤度 $\pi_t^{(l)}$ を $\pi_t^{(l)} = \frac{1/\epsilon^{(l)}}{\sum_{j=1}^L 1/\epsilon^{(j)}}$ により求める。また、各サンプルの姿勢パラメータから尤度を重みとする重み付き平均 $\bar{\mathbf{s}}_t = \sum_{l=1}^L \mathbf{s}_t^{(l)} \pi_t^{(l)}$ を求め、時刻 t の推定姿勢とする。

2.3 融合における問題点と解決策

前の 2 節で述べた顔認識系と顔追跡系において、固有空間を共通にできる場合、両者の融合は原理的には実現可能である。しかし、これは原理上の話であり、顔追跡系と顔認識系を融合するには、解決しなければならない問題点が幾つか存在する。本稿では、これらの問題点を次の 2 項目に集約して、それぞれについて解決策を検討する。

1. 顔認識に要求される位置決め（切出し）精度と顔追跡において実現可能な精度にギャップが存在する。
2. 自然環境下で生じる照明変動・姿勢変動・表情変化・隠れなどの問題に対応する必要がある。

第一の問題は、追跡系では実時間性が要求されるために個々のフレームにおける顔の切出し精度を高くできないことによる。この問題に対する 1 つの解決策として相互部分空間法 [7] の利用が考えられる。この方法では、登録時・認識時のいずれにおいても入力画像系列から部分空間を作成し、両者の正準相関を求めることから、各顔画像の切出し精度の問題を軽減できる。この方法は、時間軸方向の統計処理により系を安定化でき

るという長所を持っているが、認識時に部分空間を構成する必要がある、また、登録時と認識時の入力条件が大きく異なる場合には手法の限界があると考えられる。

一方、我々の追跡系では確率的追跡を行っているため、各フレームにおいて複数の切出しを同時に考えることができる。これは、認識系で必要とする切出し精度を確率的に解決できる可能性を示唆している。即ち、認識時に部分空間を作成することなく、通常の部分空間法で十分に対処可能である可能性がある。

しかし、確率的追跡では、切出し精度の問題も確率的にしか解決できない。これは、場合によっては誤識別が生じるフレームが残ることを意味する。従って、各フレームの識別結果に信頼性評価を付与し、その積分値で総合的に判断することで、系全体として高い信頼性を得る必要がある。

さらに、追跡系と認識系の融合において重要なのが顔テンプレートの作成法である。追跡系と認識系の両方において都合がよい位置に選択することが望ましい。

次に、第二の問題であるが、自然な生活環境を対象とする場合、様々な原因により、顔追跡が不安定になる。照明変動・姿勢変動・表情変化・隠れなどの問題は、追跡方法の改良により、ある程度は対処可能であり、また、今後も追跡系自体の性能は向上していくと考えられる。しかし、全ての状況に対応できる追跡系を作成することは難しく、むしろ、失敗することを前提として、システムを設計することが重要であると思われる。このため、各フレームにおいて追跡結果を評価することにより、追跡の状態を判定し、顔追跡の安定化と顔認識に有効なフレームの選択に利用するのが妥当と思われる。即ち、品質の低い画像を固有空間と入力画像の相関によって評価し、評価値が低い場合には、そのフレームを棄却する（認識に利用しない）ことが、本質的に重要である。ここで、棄却判定を厳しくし過ぎると、識別に使用できる画像がなかなか出現しないことが予想されるため、固有空間を作成するための画像集合の選択が重要になる。

3 融合系の構成

3.1 システムの概要

前節で論じた方針に基づき、顔追跡系と顔認識系の融合を行った。具体的には、疎固有テンプレートコンデンセーション法による顔追跡と、部分空間法による顔認識で、共通の固有顔を用いることにより、追跡系と認識系の融合を行う。図1に、2

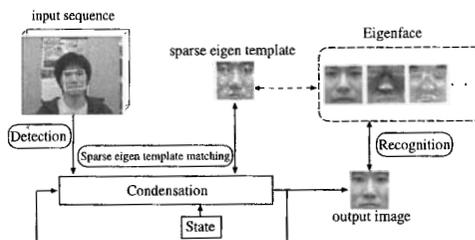


図1: 疎固有テンプレート追跡系と顔認識系の融合

つの系の概念的な接続関係を示す。ここで、固有顔は、登録者全員の幾つかの姿勢で撮影した顔画像から事前に構成したものを用い、主成分分析により構成した。

この追跡・認識融合系は、顔検出系と組み合わせることにより、自動的に動作する。実装系においては、従来から研究室で用いられている顔検出法 [8] を使用した。

3.1.1 テンプレートの作成

2.3においても指摘したように、テンプレートの切出し方は追跡系および認識系の双方にとって重要である。本稿では、顔の特徴点を用いてテンプレートの位置を正規化する。特徴点を用いた位置の正規化は、切出した後の特徴点の位置が全ての画像で等しくなるように行う。ここで、顔部品の位置情報は個人の大きな特徴であるため、特徴点の数が多くなるにつれて認識性能が低下すると考えられる。したがって、本稿では x, y 軸方向のスケールの自由度を拘束できる、3点を用いた位置の正規化を行う。

具体的には3点を次のように決定した。まず、顔の対称性を仮定すると、瞳2点を利用することで、適切に x 軸方向のスケールの自由度を拘束できる。次に、 y 軸方向のスケールの自由度を拘束するための特徴点を選択する。このとき、顔中心から遠い位置のものが望ましいと思われるため、本稿では、第3の特徴点として口の中央を用いる。

3.1.2 姿勢の取り扱い

疎テンプレートコンデンセーション法 [1] では、物体の姿勢を並進 $2(u, v)$, x, y, z 軸周りの回転 $3(\psi, \theta, \phi)$, スケール $1(s)$ の計6パラメータとして表している。上述のように位置を正規化したテンプレートに対応する部分は、ほぼ平面で近似できる。したがって、 x 軸、 y 軸周りの回転は、各方向のスケール変換で表せる。そこで、 z 軸周りの回転の前に x 軸、 y 軸方向のスケール変換を行う次

式の座標変換を採用する。ここで、姿勢パラメータを並進 (u, v) 、回転 (ϕ) 、スケール (s_x, s_y) の計 5 パラメータである。

$$\begin{aligned} \mathbf{m}' &= \mathbf{R}\mathbf{S}\mathbf{m} + \mathbf{t} \\ &= \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \mathbf{m} + \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned}$$

ここで、 \mathbf{m} はテンプレート画像上の座標、 \mathbf{m}' は入力画像上の座標である。

追跡の開始時には、前節で述べたような 3 点の特徴点位置の情報が与えられるため、これを用いて初期姿勢を決定する。まず、顔の対称性を仮定して x 軸方向のスケールパラメータ s_x を、瞳 2 点間の距離を用いて計算する。そして、残りの 4 パラメータを最小二乗法により計算する。これらを初期姿勢として追跡を開始する。

3.2 状態処理を用いた追跡の安定化

2.3 では、追跡時に生じる様々な変動に対する考え方を示し、顔追跡の安定化と失敗時の判定が重要であることを述べた。この議論を踏まえて、以下に示すように、状態に応じて処理の切替えを行う追跡系を構成した。追跡時の、各状態の定義と処理内容について簡潔に説明する。

1) Initial

追跡を開始する前の初期状態である。顔検出モジュールが起動され、顔が検出されると初期姿勢を決定して、追跡状態 (Tracked) に遷移する。

2) Tracked

追跡に成功している状態である。顔認識は、この状態においてのみ起動される。追跡状態では、その時刻 t の推定姿勢 \bar{s}_t を以降の処理に利用できるように保持する。実装系では、保持姿勢数は $S = 1$ とする。

3) Temporarily Lost

一時的に追跡が不安定になっている状態である。この状態では、追跡の安定化のために以下のような処理を行う。

- 姿勢サンプル数の増加
- 保持姿勢 $\{\bar{s}_i\}$ を利用した drift
- バックグラウンドでの疎固有テンプレートコンデンセーション法を用いた周辺探索

ここで、drift ステップで、直前の姿勢の他に保持姿勢 $\{\bar{s}_i\}$ からサンプルを選択することで、追跡状態への復帰を促進する。また、保持姿勢と直前の姿勢が大きく異なる場合には、通常の追跡だけでは対応できない場合もあり、バックグラウンド処理として 2 段階の疎固有テンプレートコンデン

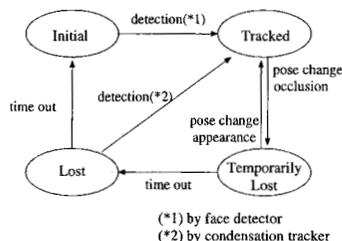


図 2: 顔追跡系の状態遷移図

セーション法による周辺探索を行う。1 段階目では diffuse で用いるガウスノイズの標準偏差を変更することで広範囲を探索し、2 段階目では狭い範囲での探索を行う。この探索により、顔が検出された場合には Tracked に遷移する。この処理を一定時間繰返しても顔が検出されない場合、状態は Lost に遷移する。

4) Lost

追跡に失敗している状態である。Temporarily Lost で述べた周辺探索のみを行う。一定時間が経過しても顔の検出ができなければ Initial に遷移する。

状態遷移図を、図 2 に示す。図中、Tracked と Temporarily Lost の間、および Lost から Tracked への遷移判定は関数 (式 (3)) によって行う。

3.3 画像系列を対象とした認識系の構成

3.3.1 相関による棄却

2.3 の議論に従って、顔追跡によって得られる画像集合 $\{x_t\}$ には、様々な品質のものが含まれている。そこで、各画像 x_t と固有空間 (\bar{x}, Φ) の相関 C_t を次式によって求めて、画像の棄却に用いる。

$$C_t = \text{Cor}(x_t, \Phi x_t^* + \bar{x}) \quad (3)$$

$$\text{where } \text{Cor}(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x} - 1/N)^T (\mathbf{y} - 1/N)}{\|\mathbf{x} - 1/N\| \|\mathbf{y} - 1/N\|}$$

C_t に関する閾値処理により、次式で定義される w_t が 1 のものについてのみ、認識を行う。

$$w_t = \begin{cases} 1 & \text{if } C_t > \theta_c \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

なお、閾値 θ_c は学習セット上での相関の平均値 λ と標準偏差 σ から、 $\theta_c = \lambda - \sigma$ により求めた。

3.3.2 選択されたサンプル集合における競争的顔認識

2.2 で述べたように、各時刻において、サンプル姿勢集合から推定姿勢が計算されるが、これは

疎テンプレートを形成する少数の点における射影残差の評価 $\epsilon^{(l)}$ の大小によって行われるため、必ずしも顔認識に有効なサンプルが選択されているとは言えない。即ち、推定姿勢で切出した顔画像は、固有空間によって十分には表現できない情報を含む可能性があり、これをそのまま用いたのでは、識別性能の低下を引き起こす恐れがある。

顔認識には固有空間で十分に表現できる画像を用いるべきであると考えられるため、diffuse ステップ内で、上位に絞り込まれた各サンプル姿勢に対応する画像を評価する。具体的には、各サンプル姿勢に対応する画像と固有空間の相関 (式 (3)) を求め、相関が最も高いサンプルから得られる画像を用いて認識を行う。即ち、こうして得られた画像に対して、2.1 で述べた最近傍法による識別を行う。

ここで、最近傍法 (式 (2)) は、プロトタイプ間の決定境界 (decision boundary) を定めることに等しい。しかし、決定境界付近に射影されたものについては、その決定境界を構成する人物同士の識別が難しいと考えられる。また、プロトタイプからの距離が遠いものについても、登録者以外の人物である可能性が高いため、誤識別を引き起こす可能性がある。そのため、時刻 t の画像に対する識別結果の信頼度を決定する。入力画像 \mathbf{x} と識別結果である人物 h 、姿勢 k との固有空間内距離 D_{hk} は以下ようになる。

$$D_{hk} = d(\mathbf{x}^*, \mathbf{x}_{hk}^*) \quad (5)$$

ここで、人物 h 、姿勢 k と最も近い人物 h' 、姿勢 k' までの固有空間内距離を D_{\min} で表す。即ち、

$$D_{\min} = \min_{\substack{1 \leq k' \leq K \\ h' \neq h}} d(\mathbf{x}_{hk}^*, \mathbf{x}_{h'k'}^*) \quad (6)$$

これら D_{hk} 、 D_{\min} を用いて、識別信頼度 μ_{th} を以下のように定義する。

$$\mu_{th} = \min\left(\frac{D_{\min}}{2D_{hk}}, 1\right) \quad (7)$$

ここで、最近傍法により選択される人物以外 $h' \neq h$ については、 $\mu_{th'} = 0$ とする。また、相関による棄却処理がなされた場合、全登録人物に対して、 $\mu_{th} = 0$ とする。

このようにして、各時刻で得られる識別信頼度を累積することにより、任意の時刻 T までの画像集合に対して、累積信頼度 R_{Th} を計算でき、その最大値によって識別を行うことができる。即ち、識別結果は次式で与えられる。

$$h = \arg \max_{1 \leq h \leq H} R_{Th} = \arg \max_{1 \leq h \leq H} \sum_{t=0}^T \mu_{th} \quad (8)$$

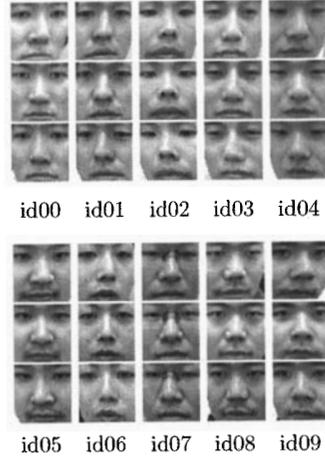


図 3: 登録画像 (10 人, 各 3 枚)

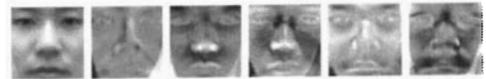


図 4: ES30: 固有空間の中心 (左端) と上位 5 軸



図 5: ES10: 固有空間の中心 (左端) と上位 5 軸

4 追跡・認識実験

4.1 実験条件

本稿で提案した方法を用いて顔の追跡・認識実験を行った。登録画像は、室内環境で撮影した 10 人の人物が正面、右、左方向に向いている画像とした。左右に向く角度は、約 15° とした。各方向とも両目と口の画像上の 2 次元位置を手動で与え、全ての画像で特徴点の位置が等しくなるように 48×48 のサイズで顔領域を切出した (図 3)。30 枚の登録画像で構成される 15 次元固有空間 (ES30: 図 4) を作成した。また、比較のため、10 枚の正面顔のみで構成される 7 次元固有空間 (ES10: 図 5) を作成し、追跡、認識の比較実験を行った。ここで、ES30 および ES10 の次元数は累積寄与率が 90% を越える最小の次元数を選択した。

動画像として、同様の環境で撮影した 10 人の動画像 I~III (30 秒, 30fps) を用いた。これらは並進、 z 軸回転、スケール変化、および 3 次元回転 (x, y 軸周りの回転) によって顔の姿勢が変動する。3 次元回転の範囲は、動画像 I では $\pm 15^\circ$ 、動画像 II, III では $\pm 45^\circ$ である。ここで、動画像 I, II は指示を与えることで全ての人物が同様の動きをする

表 1: 疎固有テンプレート追跡の実験条件

$\text{trace}(P^*), (\text{trace}(P^i))$	128 (64)
点選択法	領域内最大・最小
サンプル数 (Tracked)	1000
(Temporarily Lost)	2000
σ (tracking)	(2.5, 2.5, 4, 0.05, 0.05)
σ (search)	(20, 20, 15, 0.1, 0.1)

(図 6,7). 一方, 動画像 III は被撮影者が自由に動き, 表情変化とオクルージョンが発生する. また, 動画像 I,II では初期フレームにおいて正面を向いているため初期フレームから追跡を開始し, 動画像 III では人物のフレームインから始まるため, 検出された顔画像と固有空間の類似度を閾値処理することで追跡を開始する. ここで, 閾値を 0.9 とした. また, 全ての動画像では対象となる人物のみが登場する.

4.2 状態の導入による追跡安定化効果

疎固有テンプレートコンデンセーション法(従来法)に状態処理を導入することの有効性を確認するために, 隠れや大きな姿勢変動を含む動画像 III を用いて追跡実験を行った. 実験条件を表 1 に示す. ここで, diffuse ステップでのガウスノイズの標準偏差を $\sigma = [\sigma_u, \sigma_v, \sigma_\phi, \sigma_{s_x}, \sigma_{s_y}]^T$ で表す.

状態処理を含まない従来法による動画像 III (id07) の追跡例を図 8 に示す. ここで, 図中の四角の枠は顔の姿勢の推定結果を表し, 白点はサンプル位置(正確にはサンプル姿勢に対応するテンプレートの重心位置)を表す. この結果は, 手が顔の大部分を隠す場面で, 追跡に失敗していることを示している. 一方, 状態処理を導入した場合の追跡例を図 9 に示す. 上述の隠れが生じた時点で追跡結果の相関が小さくなり, $t = 85$ で Tracked 状態から Temporarily Lost 状態への移行が起こる. この例では, 保持姿勢を用いた drift が旨く働き, $t = 95$ で隠れがなくなった時点で, 保持姿勢から選択されたサンプルの評価値が大きくなり, Tracked 状態に復帰することができた.

さらに, 動画像 III (id06) の追跡例を図 10 に示す. ここで, 図中の暗い点は追跡とは別に周辺探索を行う際のサンプル位置を表し, 明るい点は評価値が上位のサンプル位置を表す. この動画像では, 隠れが発生したまま顔の姿勢が大きく変化するため, 保持姿勢を用いた drift では追跡に復帰することが難しい. しかしながら, 周辺探索を行うことで Tracked 状態に復帰できることが分かる.



図 6: 動画像 I の一部



図 7: 動画像 II の一部

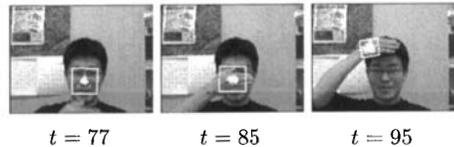


図 8: 従来法の追跡結果(動画像 III (id07))

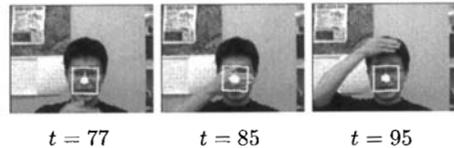


図 9: 状態処理を導入した場合の追跡結果(動画像 III (id07))

周辺探索では, まず, 図 10 の暗い点が示すように大きな範囲にサンプルを配置する. これにより大まかな位置を推定した後, 小さな範囲で再び探索を行うことで, 図 10 の明るい点が示すように顔の位置を推定することができる. この動画像では, $t = 672$ で相関が大きくなり追跡に復帰できた.

4.3 認識結果とその分析

4.3.1 フレーム単位での識別率の比較

動画像 I~III について, 状態処理を導入した場合の認識結果をまとめる. 顔認識法としては, 3.3.2 で述べた方法の他に, 疎固有テンプレートコンデンセーション法によって評価値が上位 10 位, 100

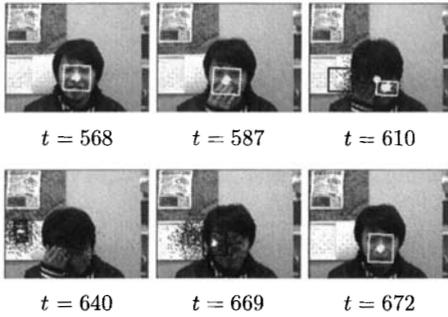


図 10: 状態処理を導入した場合の追跡結果 (動画 III (id06))

表 2: ES30 によるフレーム単位の識別率 (%)

Data set	I	II	III
相関の上位 1 位	97.4	90.5	66.2
平均姿勢 (10 位まで)	97.3	90.2	69.3
平均姿勢 (100 位まで)	97.1	90.4	70.4

表 3: ES10 によるフレーム単位の識別率 (%)

Data set	I	II	III
相関の上位 1 位	91.7	86.9	58.7
平均姿勢 (10 位まで)	93.3	87.5	60.5
平均姿勢 (100 位まで)	93.1	87.2	58.4

位のサンプルから重み付き平均姿勢を求め、これに対応する画像を射影する方法を比較した。なお、棄却判定に用いる相関の閾値 θ_c は 0.829 とした。

固有空間として、ES30 あるいは ES10 を用いた場合のフレーム単位の識別率を、それぞれ、表 2、表 3 に示す。ここで、識別率とは、棄却されなかった画像集合 (有効画像集合と呼ぶ) 内において、本人が正しく識別された率である。結果を比較すると、ES30 の方が識別率が高く、特に姿勢変動の大きい動画 III では I、II と比べて差が顕著であり、姿勢変動を含む固有空間の有効性が確認できた。

表 2 の動画 I、II に対しては、提案法 (相関の上位 1 位による識別) が平均姿勢を用いる方法よりも少し勝っているが、動画 III に対しては逆に平均姿勢の方が勝っている。これは、姿勢変動や表情変化が大きい動画 III では、個々のサンプル姿勢による切出し精度が、認識系で要求する精度より低いと考えられる。特に、表情変化の発生時間が長い id06 の人物の動画 III (図 11) については、識別率が 33% と低く、表情変化が認識に与える影響の大きさが確認できた。



図 11: 誤識別を起こした画像の例 (III (id06))

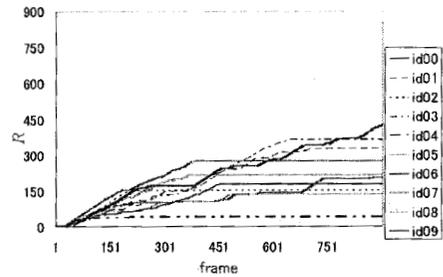


図 12: 累積信頼度の変化, 動画 III, ES30

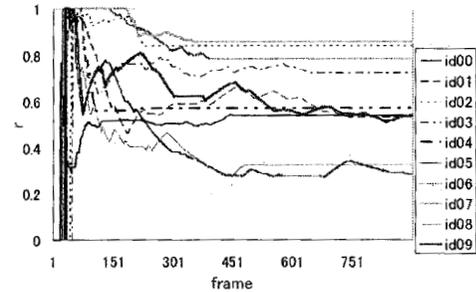


図 13: 累積信頼率の変化, 動画 III, ES30

4.3.2 累積信頼度による認識

動画 III を対象として、ES30 を用いた場合について、各登録人物が正しく識別される場合の累積信頼度の経時変化を図 12 に示す。このグラフより、累積信頼度は、画像系列中で識別に有効な画像が出現する時間に上昇し、一方、何らかの事情で画像が棄却される場合には、一定値を保っている。なお、このデータセットでは、全ての動画 III について登録人物本人の累積信頼度が高くなり、正しく識別が行われることが確認された。

次に、横軸に時刻 T を取り、累積信頼率 ($r =$ 累積信頼度 / 有効画像数) の経時変化を求めた結果を、図 13 に示す。ここで、有効画像数とは有効画像集合に含まれる画像の枚数である。この r が高いほど識別結果の信頼性が高いことを示すと考えられる。実験結果より、部分的に誤識別が行われた場合にも、その結果が蓄積されるのを抑え、誤識別を引き起こさないでいることが確認できた。

データセットを II に変えた場合の累積信頼率の変化を図 14 に示す。このデータは隠れや表情変化が少ないため図 13 に比較して、累積信頼率が高いことが確認できる。

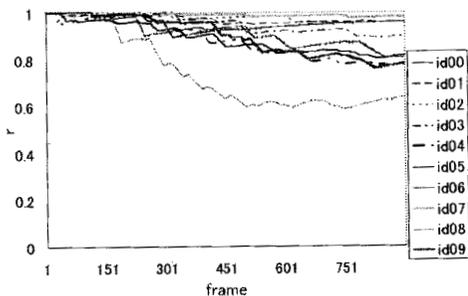


図 14: 累積信頼率の変化, 動画 II, ES30

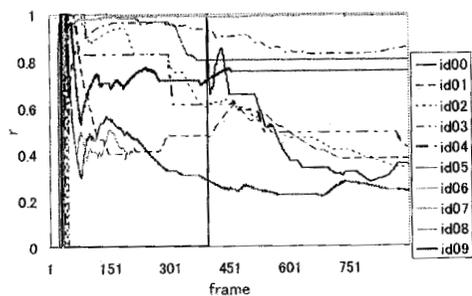


図 15: 累積信頼率の変化, 動画 III, ES10

次に、動画 III において、固有空間を ES10 に変えた場合の累積信頼率の経時変化を図 15 に示す。ES10 には姿勢変動が含まれていないため、図 13 と比較すると、全体的に累積信頼率は低くなっている。特に、人物 id00 においては、396 フレーム目まで相関が低い状態が続くことにより、認識処理が行われていないことが判り、固有空間にどのような変動を含めるかにより認識性能が変わることを顕著に示している。

5 まとめ

本稿では、疎固有テンプレートコンデンセーション法による顔追跡と、部分空間法に基づく顔認識で、共通の固有空間を使用することにより、追跡系と認識系の融合が可能であることを示した。同一の固有空間を使用することにより、両者は形式的には容易に融合できると考えられるが、認識時に必要とする位置決め精度を追跡時に安定的に実現できる保証がないため、確率的な取り組みが必須である。この問題に対して、本稿ではコンデンセーション法の確率的枠組を利用した解決策を示した。また、様々な状況下での追跡系の安定的動作を目指し、状態の概念を導入することにより、追跡が一時的に失敗した場合にも対応できる追跡系

を実現した。

本稿で示した追跡系と認識系の融合では、人物識別を目的としたため、姿勢や表情変化の取扱いが陽に取扱われていない。これは、ある程度の棄却率を許容しても人物識別が可能であることにも関連があるが、動画画像系列中には利用されない画像が多く存在する。今後は、これらの画像の取扱いを含め、照明変動、姿勢変動、表情変化などへの柔軟な対応方法を検討する必要があると思われる。

実験データの収集に協力頂いた方々に感謝する。本研究の一部は科学研究費補助金基盤研究 (B)(課題番号 15300062) により行われた。

参考文献

- [1] 松原, 尺長: “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, **46**, SIG 9(CVIM 11), pp. 60-71 (2005).
- [2] M. Turk and A. Pentland: “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience*, **3**, 1, pp. 71-86 (1991).
- [3] 重成, 坂上, 尺長: “固有顔の直交分解と仮想化による照明変動に影響されない顔画像認識”, 電子情報通信学会論文誌 (D-II), **J86-D-II**, 9, pp. 996-1004 (2003).
- [4] 坂上, 尺長: “ガウシアンフィルタによる画像分解を用いた照明変動にロバストな顔画像認識”, 電子情報通信学会技術報告 PRMU-2005-91, pp. 45-50 (2005).
- [5] 松原, 野口, 尺長: “疎固有テンプレートマッチングに基づくコンデンセーション法”, 画像の認識・理解シンポジウム (MIRU 2004) 論文集, 第 1 巻, pp. 261-266 (2004).
- [6] M. Isard and A. Blake: “Condensation-conditional density propagation for visual tracking”, *International Journal of Computer Vision*, **29**, 1, pp. 5-28 (1998).
- [7] 福井, 山口, 鈴木, 前田: “制約相互部分空間法を用いた環境変動にロバストな顔画像認識 - 照明変動の影響を抑える制約相互部分空間の学習”, 電子情報通信学会論文誌 (D-II), **J82-D-II**, 4, pp. 613-620 (1999).
- [8] F. Sakaue, M. Kobayashi, T. Migita, T. Shakunaga and J. Satake: “A real-life test of face recognition system for dialogue interface robot in ubiquitous environments”, *Proc. ICPR2006*, pp. 1155-1160 (2006).