

# 矩形特徴を用いた顔検出器から得られる 識別スコアの最大化による顔追跡

日高 章理<sup>†</sup> 西田 健次<sup>††</sup> 栗田多喜夫<sup>††</sup>

<sup>†</sup> 筑波大学システム情報工学研究科, つくば市

<sup>††</sup> 産業技術総合研究所脳神経情報研究部門

E-mail: <sup>†</sup>hidaka.akinori@aist.go.jp

あらまし 顔追跡は今なおコンピュータ・ビジョンの重要な課題であり続けている。本論文では静的な顔検出器をベースとした追跡手法を提案する。提案手法では入力画像の“顔らしさ”のスコアを計算するために、矩形特徴とブースティングを用いた顔検出器を利用する。顔追跡は動画像の各フレーム毎に画像平面の中で最高のスコアを持つ部分領域を追跡することで実現される。ここでは、勾配法による数値最適化と局所領域の全探索を組み合わせたスコアの最大化による追跡手法を開発し、評価実験を行った。

キーワード 顔追跡, 矩形特徴, ブースティング

## Face Tracking by Maximizing Classification Score of Face Detector Based on Rectangle Features

Akinori HIDAKA<sup>†</sup>, Kenji NISHIDA<sup>††</sup>, and Takio KURITA<sup>††</sup>

<sup>†</sup> Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai,  
Tsukuba-shi, Ibaraki-ken, 305-8577 Japan

<sup>††</sup> Neuroscience Research Institute, AIST, Central 2, 1-1-1 Umezono, Tsukuba-shi, Ibaraki-ken, 305-8568  
Japan

E-mail: <sup>†</sup>hidaka.akinori@aist.go.jp

**Abstract** Face tracking continues to be an important topic in computer vision. We describe a tracking algorithm based on a static face detector. We use rectangle feature and boosting algorithm to calculate the face score of an input image. In our method, face tracking is performed by tracking the local region where the face score is maximum. We propose and evaluate the tracking algorithm that the combination of jumping to the gradient direction and precise search at the local region.

**Key words** face tracking, rectangle feature, boosting

### 1. ま え が き

複雑な環境における実時間での対象追跡という課題は、今なおコンピュータ・ビジョンの重要な課題であり続けている。実時間での対象追跡は監視、スマートルーム、通信会議などのアプリケーションに必須の要素である。対象追跡問題に対する典型的なアプローチには次のようなものがある。特徴ベースの追跡手法では、画像内での曲線 [1], [2] や対象の見え [3] などの特徴の固有性を頼りに追跡が行われる。同様に、モデルベースの追跡手法においても曲線 [4], [5] や対象の見え [3] が用いられる。パーティクルフィルタ [6], [7] やミーンシフト [8] では、しばしば対象のダイナミクスが取り扱われる。しかし群衆の中での顔

追跡問題などでは、対象のダイナミクスよりも画像としての表れ方に注意を払うべきである。

一方で、サポートベクターマシン (SVM) [9], [10] やブースティング [11], [12] を用いた対象検出アルゴリズムは、既にビデオレートでの動作が可能なレベルに達している。特に Viola ら [11], [12] の顔検出器は、高い識別性能と高速な処理を両立した手法である。彼らは対象検出のために矩形特徴 (Rectangle Features) を提案し、また少数の矩形特徴をブースティングにより組み合わせることで高い識別率が実現されることを示した。しかしながら、実際に対象追跡を実用化する場合、追跡器の計算量は非常に重要な問題となる。ほとんどのアプリケーションにおいて、システムは対象追跡が完了した後により高次の処理

(対象の認証, 移動経路の解析, 推論など)を行うため, 追跡器には割り当てられる計算リソースは全体の中のごく一部となる. ゆえに, それらの高速な対象検出器をベースとした追跡器の開発が重要となる. 近年では S. Avidan [13] によりサポートベクタートラッカー (SVT) と呼ばれる追跡器が提唱された. このアルゴリズムでは, SVM による対象検出器から計算される識別スコアを画像平面内で最大化することで対象追跡が実行される.

人間が物体認識を行うときには, 自分の目を連続的に注視点に向ける. 人間の視覚では, 移動物体の追跡はサッケード運動と追従眼球運動を組み合わせで行われる. サッケード運動では対象のラフな位置に目が向けられ, 正確な位置は追従眼球運動によって保たれる. 我々はこのサッケード運動と追従眼球運動の組み合わせこそ, 恐らく高速かつ高精度な追跡を実現するために重要であると考える.

この論文では Viola らの矩形特徴とブースティングを用いた検出器を SVT のフレームワークに当てはめることで, 新たな対象追跡器を構築する. この追跡器はサッケード運動のような粗い探索と局所領域の精密な探索を組み合わせで対象を追跡する.

## 2. 矩形特徴とブースティングを用いた検出器

本研究では顔追跡のベースに用いる顔検出手法として, P. Viola ら [11] が提案した矩形特徴とブースティングから成る検出器を利用する. 我々は文献 [11] で用いられた手法に対し, 2つの変更を加えた. 第一に, 識別関数を 0/1 出力のしきい値関数からシグモイド関数に改め, 微分可能であるようにした. 第二に, ブースティングにおける二段階の最適化を一段階に減らすことで, 最終的な識別器の汎化性能を向上させた. 以下にそれらの詳細を述べる.

### 2.1 矩形特徴

矩形特徴は互いに隣接する  $n$  個の小矩形から構成される特徴 (図 1 左端の 4 パターン, それぞれ  $n=2,2,3,4$ ) で, 入力画像における各小矩形領域の明るさの違いを指標として対象の識別を行う. 例として, 図 1 のようにある一般の画像  $F$  内の領域  $q$  に存在する部分画像  $I_q$  が顔であるかどうかを識別するという問題を考える. ある矩形特徴  $r$  の  $k$  個目 ( $k = 1, \dots, n$ ) の小矩形が占める領域を  $s_k$  とおくと, 領域  $s_k$  の部分画像は  $I_{s_k} = \{i(x, y) | (x, y) \in s_k\}$  と定義される ( $i(x, y)$  は点  $(x, y)$  の輝度値). 各小矩形  $s_k$  には黒色/白色を示すラベル  $c_k \in \{-1, 1\}$  が割りふられる. ここで領域  $s_k$  の平均輝度値を,

$$m(I_{s_k}) = \frac{1}{S} \iint_{s_k} i(x, y) dx dy \quad (1)$$

とおく ( $S$  は  $s_k$  の面積). 今, 一つの矩形特徴を  $r = \{s_k, c_k\}_{k=1}^n$  と表すことにすると, 入力画像  $I_q$  に対する  $r$  の特徴量  $f$  は,

$$f(r; I_q) = \sum_{k=1}^n c_k m(I_{s_k}) \quad (2)$$

と定義される. すなわち,  $f$  は白色小矩形領域の平均輝度値から黒色小矩形領域の平均輝度値を引いた値である.

Viola らはこの  $f$  による対象識別関数として, 0/1 出力のしきい値関数,

$$h(r; I_q) = \begin{cases} 1 & (I_q \text{ は顔}) \text{ if } pf(r; I_q) > p\theta \\ 0 & (I_q \text{ は背景}) \text{ otherwise} \end{cases}$$

を用いた. ここで  $p \in \{-1, 1\}$  と  $\theta \in \mathbb{R}$  は機械学習によって固定される値である. すなわち, 各小矩形の明るさの差  $f$  が一定値  $\theta$  より大きい ( $p = 1$  のとき) あるいは小さい ( $p = -1$  のとき) ならば, 入力画像は顔であると判断される.

我々はこの識別関数をシグモイド関数,

$$h(r; I_q) = \frac{1}{1 + e^{p(f(r; I_q) - \theta)}} \quad (3)$$

に置き換え, 対象の位置  $x, y$  に関して微分可能であるようにした.

矩形特徴の平均輝度は, 積分画像 (integral image) 法 [11] あるいは summed area table [14] と呼ばれる手法により, 高速に計算される.

一つの矩形特徴は, それを構成する小矩形の個数・サイズ・並び方, および小矩形全体の入力画像内での位置, という 4 つの情報から構成される. 我々は小矩形の個数および並び方を図 1 (左) の 4 パターンに限定し, また一つの矩形特徴において各小矩形のサイズは同じであるという制限を加えた. この場合,  $24 \times 24$  ピクセルの訓練画像平面において, およそ 13 万通りの異なる矩形特徴が存在する. これらのバリエーションの中から, 人の顔の識別という課題に特化した矩形特徴を複数選び出して統合する学習法が, 次節のブースティングである.

### 2.2 ブースティング

ブースティングとは比較的性能が低い識別器を複数組み合わせでより強力な識別器を構築する学習アルゴリズムである. ベースとなる識別器は弱学習器と呼ばれ, ここでは個々の矩形特徴が一つの弱学習器となる.

学習アルゴリズムは下記の通りである.

- 訓練画像セット  $\{x_i, t_i\}_{i=1}^N$  を与える. ( $x_i$ : 画像ベクトル,  $t_i = 0$   $x_i$ : 背景画像,  $t_i = 1$   $x_i$ : 顔画像)
- 画像の重みの初期化:  $t_i = 0$  ならば  $w_{1,i} = \frac{1}{2m}$ ,  $t_i = 1$  ならば  $w_{1,i} = \frac{1}{2n}$ . ( $m$ : 背景の数,  $n$ : 顔の数)
- for  $t = 1, \dots, T$ 
  - 画像の重みの正規化:  $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{i=1}^N w_{t,i}}$ .
  - ( $\forall j$  について) 矩形特徴  $r^j$  のパラメータ  $p^j, \theta^j$  を最適化し, 重み付き誤差  $\epsilon^j = \sum_i w_i |h(r^j; x_i) - y_i|$  を最小にする.
  - 全矩形特徴  $r^j$  の中で重み付き誤差  $\epsilon^j$  が最小のものを  $t$  番目の識別器  $h_t$  に採用.
  - 重みの更新:  $w_{t+1,i} = w_{t,i} (\frac{\epsilon_t}{1-\epsilon_t})^{1-|h_t(x_i)-t_i|}$
- 最終的な識別関数の構築:

$$H(I) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq \Theta \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

ここで  $\alpha_t = \log(\frac{1-\epsilon_t}{\epsilon_t})$ ,  $\Theta$  は手動で決める閾値.

各矩形特徴の訓練において, 重み  $w_{t,i}$  が大きい画像は, なるべく誤識別してはいけない重要な画像と認識される.  $w_{t,i}$  の更

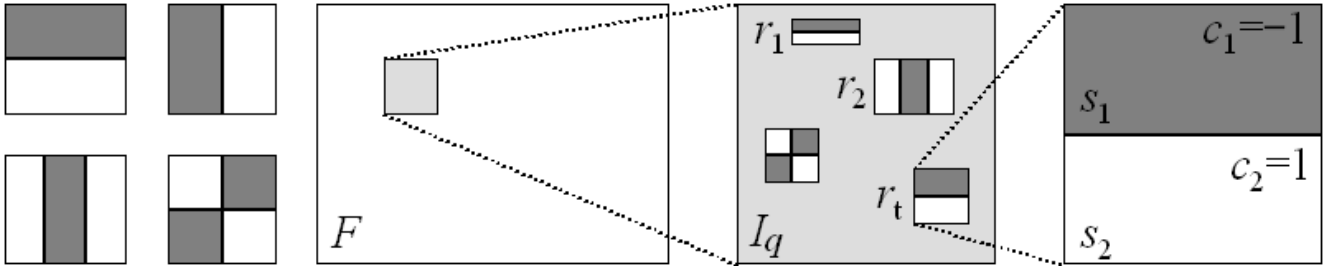


図 1 (左) 矩形特徴の例 (4 パターン) (右) 矩形特徴による識別の模式図 . 与えられた一般的な画像  $F$  に対して, その中のある領域  $q$  における部分画像  $I_q$  が人の正面顔であるかどうかを識別する際の図 . 一つの矩形特徴  $r$  は,  $I_q$  内の特定領域  $\{s_1, \dots, s_n\}$  の明るさの差を手がかりとして対象の識別を行う . すなわち白い矩形部分の平均輝度から黒い矩形部分の平均輝度を引いた値が一定値より大きい小さいかによって入力が顔か背景かの判断を決定する .

新は“今選ばれた矩形特徴が誤識別してしまう画像の重みを次の反復で増加させる”というルールに則って行われる . それゆえ, 次の反復では今まで識別が難しかった画像をより上手く識別する矩形特徴が選ばれやすくなる . この結果, 各弱識別器の間で一種の役割分担が生じ, 最終的に全ての訓練サンプルを正しく識別する識別器が得られる .

$\alpha_t$  は選ばれた矩形特徴に与えられる重みである . 最終的な識別関数  $H(x)$  は採用された各識別関数  $h_t(x)$  の“多数決”の形を取るが,  $\alpha_t$  はその投票における  $h_t$  の“投票力”を意味する .

### 2.3 汎化性能

前節のアルゴリズムにおけるブースティングでは, 下記のような二種類の最適化が行われていた .

- 各矩形特徴のパラメータを最適化し, 誤差を最小化 .
- 最適化された各矩形特徴の中で最良のものを選択 .

しかし我々は前者のパラメータの最適化を行わずに後者の最良な矩形特徴の選択を行う学習法, すなわち二段階の最適化を一段階に減らした学習法によって, 最終的な識別器の汎化性能が向上することを発見した . 図 2 は従来通り二段階の最適化を行った場合および最適化を一段階に減らした場合の学習曲線および ROC カーブである . 訓練セットには顔画像 300 枚 + 非顔画像 700 枚, テストセットには顔画像 425 枚 + 非顔画像 1500 枚を用いた . いずれの画像も  $24 \times 24$  ピクセルである . 訓練セットに対する学習曲線では, 二段階の場合 (破線) の方が一段階の場合 (実線) より収束が速い . しかしテストセットに関する学習曲線ではむしろ一段階の場合の方がより誤差が低くなっている . 同様に ROC カーブにおいても最適化を一段階に減らした場合の方が優れた結果になっていることがわかる . これは弱学習器のパラメータにランダムネスが加わることで学習データへの過適応が防がれ, その結果として汎化性能が向上したのではないかと考えられる .

## 3. スコア関数の最大化による顔追跡

S. Avidan の SVT [13] では, 追跡は SVM から得られる対象のスコアを最大化することで実行された . 顔の検出・追跡という問題においては, スコア関数の値には対象画像と“人の顔”との類似度という意味がある . したがって人の正面顔が含まれ

る画像一般において, ちょうど顔のみを含むような部分領域のスコア関数の値は, それ以外のあらゆる領域のスコアより高くあることが期待される . ここでもし顔領域が画像平面内のどこかに移動した場合, この顔に付随するスコア関数の頂点も顔の動きに同期して平面内を移動するはずである . ゆえに, もし信頼に足る顔のスコア関数が得られたならば, 顔追跡という問題は画像平面内でスコアが最大の領域を追跡するという問題に置き換えられる . このときもしスコア関数が微分可能ならば, その導関数は探索の助けになる . 対象が領域  $q_0$  から未知の領域  $q$  に移動した場合, 領域  $q$  は  $q_0$  から見てスコア関数の上り勾配方向にいるはずである .

### 3.1 スコア関数と最大化の対象関数

我々のスコア関数  $E(I_q)$  は式 (3) および式 (4) から次のように得られる .

$$E(I_q) \equiv \sum_{t=1}^T \alpha_t h(r_t; I_q) = \sum_{t=1}^T \frac{\alpha_t}{1 + e^{p_t(f(r_t; I_q) - \theta_t)}} \quad (5)$$

これは各矩形特徴  $r_t$  の多数決において, 入力画像を顔と判断した矩形特徴の投票力の総和である .

以下でスコア関数の最大化問題を定式化する . 画像  $I'$  および  $I$  を, 人間とそれ以外の対象が映った動画像系列の中の連続する 2 フレームとする . 今, フレーム  $I'$  において領域  $q$  に存在していた顔が, 次のフレーム  $I$  において未知の領域  $q_f$  に移動したとする . 古い位置情報  $q$  を手がかりに, 移動先の位置  $q_f$  を特定したい .  $I_q$  を現フレーム  $I$  における領域  $q$  の部分画像とする . ここで領域  $q$  の近傍でスコア  $E(I_q)$  がどのように変化するかを考える . 領域  $q$  を距離  $(u, v)$  だけ平行移動させた領域を  $q_{(u,v)}$  とおく . 計算量削減のため, 部分画像  $I_{q_{(u,v)}}$  をテイラー展開で次のように近似する .

$$I_{q_{(u,v)}} \simeq I_q + uI_{qx} + vI_{qy} \quad (6)$$

ここで  $I_{qx} = \frac{\partial I_q}{\partial x}$ ,  $I_{qy} = \frac{\partial I_q}{\partial y}$  である . これから式 (1) と式 (2) は次のように近似される .

$$m(I_{s_k}) \simeq m(I_{s_k}) + um_x(I_{s_k}) + vm_y(I_{s_k})$$

$$f(r_t; I_q) \simeq \sum_{k=1}^n c_k \{m(I_{s_k}) + um_x(I_{s_k}) + vm_y(I_{s_k})\} \quad (7)$$

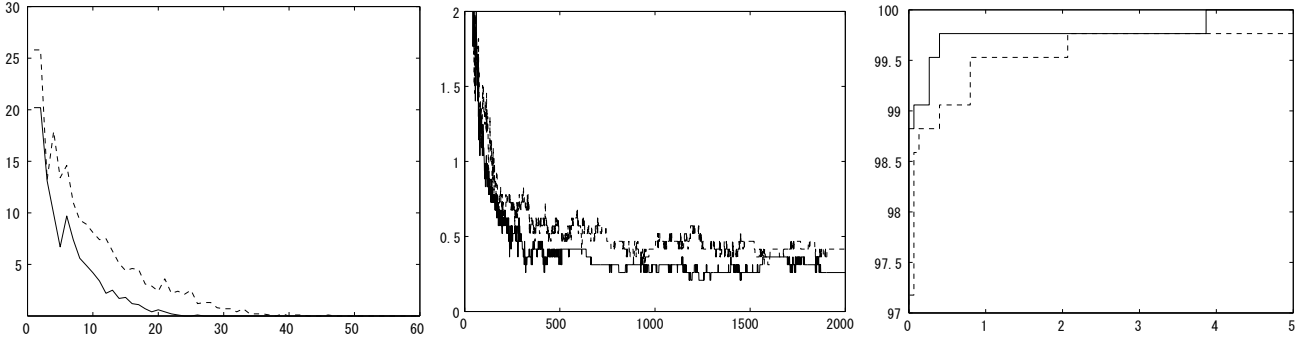


図2 (左) 訓練セットに対する学習曲線 (中央) テストセットに対する学習曲線 (右) テストセットに対する ROC カーブ. 左図および中央図では横軸がブースティングの回数, 縦軸が誤識率 (単位は%). 右図では横軸が非顔画像を顔と誤検出した率, 縦軸が顔画像を正しく顔と検出した率 (いずれも単位は%). 実線が一段階の最適化によるもの, 破線が従来の二段階の最適化によるもの.

ここで  $m_x = \frac{\partial m}{\partial x}$ ,  $m_y = \frac{\partial m}{\partial y}$  である.  $m_x(I_{s_k})$  および  $m_y(I_{s_k})$  は, それぞれ矩形領域  $s_k$  における左右または上下の辺の平均輝度の差である. 矩形の各辺は, 幅あるいは高さが単位長さであるような矩形と見なせるので, その平均輝度は通常の矩形と同様に積分画像法を用いて高速に計算できることに注意する. 式(7) 右辺を  $f^*(r_t; u, v)$  とおき, 式(5) の  $f$  を  $f^*$  と置き換えれば,  $E(I_q)$  の近似が得られる. これを移動量  $(u, v)$  を用いて次のように表す.

$$E^*(u, v) = \sum_{t=1}^T \frac{\alpha_t}{1 + e^{p_t(f^*(r_t; u, v) - \theta_t)}}. \quad (8)$$

この  $E^*(u, v)$  が最大化の対象となる関数である.  $E^*$  が ( $E$  と同じく) 領域  $q_f$  で最大になるという仮定が成り立つならば, フレーム間での対象の移動量  $(u_f, v_f)$  は  $E^*(u, v)$  を最大にする  $(u, v)$  として特定できる.

以上のようにして, 領域  $q_f$  の追跡という問題は  $E^*(u, v)$  の最大化という問題に置き換えられる. 例として最急降下法を用いた場合の解法を示す.

$$f_{tu}^* = \frac{\partial f^*(r_t; u, v)}{\partial u} = \sum_{k=1}^n c_k m_x(I_{s_k})$$

$$f_{tv}^* = \frac{\partial f^*(r_t; u, v)}{\partial v} = \sum_{k=1}^n c_k m_y(I_{s_k})$$

$$\frac{\partial E^*}{\partial u} = - \sum_{t=1}^T \alpha_t \frac{e^{p_t(f^*(r_t; u, v) - \theta_t)}}{(1 + e^{p_t(f^*(r_t; u, v) - \theta_t)})^2} p_t f_{tu}^*$$

$$\frac{\partial E^*}{\partial v} = - \sum_{t=1}^T \alpha_t \frac{e^{p_t(f^*(r_t; u, v) - \theta_t)}}{(1 + e^{p_t(f^*(r_t; u, v) - \theta_t)})^2} p_t f_{tv}^*$$

$$(u, v) \leftarrow (u + \eta \frac{\partial E^*}{\partial u}, v + \eta \frac{\partial E^*}{\partial v}),$$

ここで  $\eta$  は学習係数.

### 3.2 ぼやけ画像による局所解の回避

勾配法による数値最適化において, 対象関数の多峰性は局所解への没入を引き起こす. これを避けるため, 複数レベルのぼやけ画像が用いられる. ぼやけ画像においては対象関数の多峰性はトーンダウンするが, 代わりに真の解 (最大点) の位置が

いくぶん不明瞭になる (図3). ここで, まず最もぼやけた画像から対象関数の最大化を始め, その到達点を次にぼやけた画像での最大化の出発点とする, という再帰的なアルゴリズムを用いることにより (原画像における) 局所解への没入を避けつつ, 少しずつ真の解へ近づくことができる.

このアルゴリズムを下にまとめる.

- (1)  $m \leftarrow 0$  とする.
- (Detection Mode)
- (2)  $m \leftarrow m + 1$  とする;  $I^m$  (第  $m$  フレーム) を入力.
- (3)  $I^m$  を顔検出器にかける; 検出された顔の位置を  $q^m$  とおく; 顔が検出されなければ (2) へ.
- (Tracking Mode)
- (4)  $m \leftarrow m + 1$  とする;  $I^m$  (第  $m$  フレーム) を入力.
- (5)  $I^m$  から  $L$  段階のぼやけ画像セット  $\{I_l^m\}_{l=0}^L$  を生成 ( $I_0^m$ : 原画像,  $I_L^m$ : 最も強くぼかされた画像).
- (6)  $q_L^m \leftarrow q^{m-1}$  とおく.
- (7) For  $l = L, \dots, 1$ 
  - (a) ぼやけ画像  $I_l^m$  で  $q_l^m$  を出発点として式(8)を最大化; 到達点を  $q_{l-1}^m$  とおく.
- (8) 原画像  $I_0^m$  で  $q_0^m$  を出発点として式(8)を最大化; 到達点を  $q^m$  とおく.
- (9) もし  $I_0^m$  における  $q^m$  でのスコアが予め決められたしきい値  $\theta_{track}$  より低かったならば, 追跡失敗とみなし (2) へ; そうでなければ (4) へ.

## 4. 実験

実験で用いる顔検出器は, 725 枚の顔画像と 2200 枚の非顔画像 (いずれも  $24 \times 24$  ピクセル) を用いた反復 200 回のブースティングによって構成された. 顔追跡に用いられる動画系列は  $320 \times 240$  ピクセルのフレーム 1286 枚から成る. 各フレームには高々一人の人間が映っている. 動画中の人間は, 背景に本棚とドアがある部屋の中を動き回る (図4). 動画中の大半のフレームにおいて, 映っている顔は概ね同じ方向・サイズを維持しているが, たまに首の傾ぎ, 振り向き, カメラへの接近などの動作も発生する.

追跡実験では全 1286 フレームのうちで 15 回, 概ね正面を向

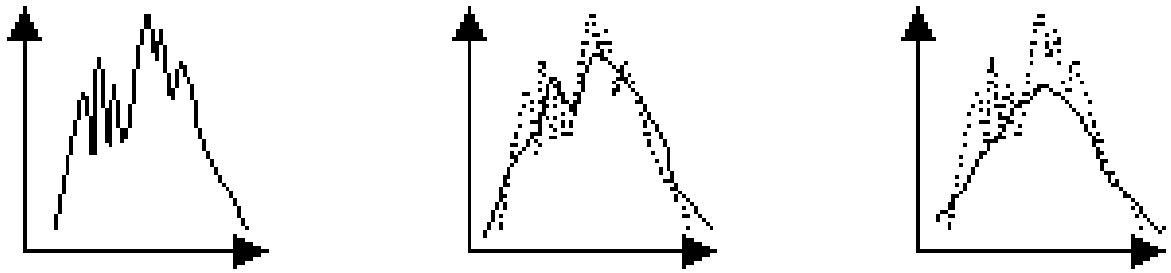


図 3 ぼやけ画像におけるスコア関数の多峰性の緩和の模式図．左から順に原画像，一段階ぼかした画像，二段階ぼかした画像におけるスコア関数の一次元断面図を表している．ぼかし強度が大きくなるにつれて関数の多峰性がトーンダウンし，代わりに最大値のポジションが不鮮明になっていく．



図 4 (左) 動画系列中の 1 フレーム (中央) 一段階ぼかした画像 (右) 二段階ぼかした画像．

いた顔において追跡失敗が生じた．それらの失敗の状況を確認したところ，いくつかのケースでは対象の移動量が大きすぎたのが原因であったと思われる．問題はそれ以外のケースであり，それらは対象がほぼ正面を向いており，なおかつほとんど移動も回転もしていない状況で追跡失敗が生じていた．これは対象関数の最大化の打ち切りから生じていると考えられる．実験では各フレーム毎の計算量を削減，また均一にするため，勾配法の反復の反復回数が一定回に達した場合，そこで最適化を打ち切るようにしていた．したがって，一見追跡がスムーズに行われているように見えるシーンでも，実は追跡された位置と真の位置の間には微妙なギャップが生じている可能性がある．このギャップは次のフレームでの追跡に悪い影響を与える．ギャップが同じ方向に蓄積されれば追跡位置は真の位置から次第に離れていくことになる．またギャップがランダムな方向に現れた場合，追跡結果にノイズが乗ることになってしまう．実際，実験ではこのノイズにより追跡結果を示すボックスが見苦しく振動させられていた．

#### 4.1 勾配方向への跳躍と局所全探索の組み合わせ

これらの問題は，勾配法が極値へ収束しないことが原因で発生する．しかしこれは局所領域を全探索することで解決できる．ここで我々は勾配法による最適化に局所全探索を組み合わせることで追跡アルゴリズムを改良する．まずぼやけ画像に関しては今まで通り勾配法により最適化を行う．勾配法による最適化の役目は，スコア関数の勾配方向へ跳ぶことで，少ない計算によってラフに位置を修正することである．そして原画像では，勾配法によって到達された位置の近傍を全探索することで（少なくとも近傍内の）最大値に到達する．近傍は予め定められた

サイズを持つ． $N_s(q_0^m)$  を  $q_0^m$  を中心としたサイズ  $s$  の近傍とする．前節のトラッキングアルゴリズムは (8) を (8') に置き換えればよい．

(8') 原画像  $I_0^m$  において，全探索により  $N_s(q_0^m)$  でスコアが最大の領域を見つける；見つかった領域を  $q^m$  とおく．

この結果，追跡精度が大幅に向上した (表 1)．勾配法と局所全探索を組み合わせた方法では，対象がほとんど動いていないなどの簡単な状況での追跡失敗は発生しなかった．また，追跡結果の振動もほぼ完全に防がれた．

## 5. あとがき

本研究では矩形特徴とブースティングを用いた顔検出器をベースとした顔追跡手法を提案し，実験により評価した．提案手法では検出器から得られるスコア関数を最大化することで追跡が行われる．テスト用の動画を用いた実験から，最急降下法によるスコアの最大化に加えて局所領域の全探索を行うことで，追跡精度が大きく向上することを確認した．今度の課題として，現在のフレームワークを拡張し，追跡対象の拡大・回転にも対応したアルゴリズムの構築を目指している．

## 文 献

- [1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," In Proc. Int. Conf. on Computer Vision, pp.259-268, 1987.
- [2] A. Blake and M. Isard, *Active contours* Springer 1998.
- [3] M. Bkackand A. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," In Proc. European Conf. on Computer Vision, Vol.1, pp.329-342, 1996.
- [4] D. Lowe, "Robust model-based motion tracking through the

表 1 最急降下法 (SDM), 近傍全探索 (NS), そしてその両方のコンビネーションにより追跡を行ったときの追跡結果. コンビネーションアルゴリズムにより, 追跡失敗回数が大幅に減少した.

	SDM only (ぼやけなし)	SDM only (ぼやけ: 2 レベル)	NS only (近傍サイズ: 5×5)	SDM+NS (ぼやけ: 2 レベル) (近傍サイズ: 5×5)
追跡失敗	50 回以上	15 回	40 回以上	3 回

integration of search and estimation,” Int. J. Computer Vision, Vol.8, No.2, pp.113-122, 1992.

- [5] A. Yuille and P. Hallianan, “Deformable templates,” In A.Blake and A.Uille, editors, *Active Vision*, pp.20-38, MIT, 1992.
- [6] M. Isard and A. Blake, “CONDENSATION – conditional density propagation for visual tracking,” Int. J. Computer Vision, Vol.29, No.1, pp.5-28, 1998.
- [7] J.-M. Odobez, S. Ba, and D. Gatica-Perez, “An implicit motion likelihood for tracking with particle filters,” In Proc. of Brithish Machine Vision Conf., 2003.
- [8] D. Comaniciu, V. Ramesh, P. Meer, “Kernel-based object tracking,” IEEE Trans. on Pattern Analysis Machine Intell., Vol. 25, No. 5, 564-575, 2003.
- [9] E. Osuna, R. Freund, and F. Girosi, “Training support vector machines: An application to face detection,” In Proc. of Conf. Computer Vision and Pattern Recognition, pp.130-136, 1997.
- [10] S. Romdhani, P. Torr, B. Scholkopf, and A. Blake, “Computationally efficient face detection,” In Prof. Int. Conf. on Computer Vision, Vol.2, pp.524-531, 2001.
- [11] P. Viola and M. Jones, “Robust real time object detection,” In IEEE ICCV Workshop on Statistical and Computational Theories of Vision, July 2001.
- [12] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” In Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, Dec. 2001.
- [13] S. Avidan, “Support Vector Tracking”, In IEEE CS Conf. Computer Vision and Pattern Recognition, Dec. 2001.
- [14] F. C. Crow, “Summed-Area Tables for Texture Mapping”, In Proc. of the 11th annual conference on Computer graphics and interactive techniques, 18(3):207-212, 1984.