

顔らしさ分布を利用した顔検出手法

高塚 皓正 田中 正行 奥富 正敏

東京工業大学 大学院理工学研究科

概要 顔検出はコンピュータビジョンにおいて、近年注目されている技術の一つであり、様々な研究が行なわれている。従来の顔検出に関する研究の多くは、切り取られたサブウィンドウが顔かどうかを判別する識別器を改良することを主な目的としている。この識別器は顔検出を行なう際の一プロセスとして重要である。しかし、これらの識別器は各サブウィンドウに対して、独立に顔または非顔を判定するため、識別器の出力(顔らしさ)の高い非顔画像を誤検出してしまうことがよくある。本報告では、顔と非顔における顔らしさ分布の違いに着目し、この違いを隔に利用した新しい顔検出の枠組みについて提案する。提案手法では、顔らしさ分布を生成し、統合処理により顔と非顔の違いを強調することで、従来手法で誤検出していた非顔を正しく分類することが可能になる。実験では、テストデータセットと実画像を用いて、提案手法の有効性を確認した。その結果、それぞれ20%と10%の検出率の向上が見られた。

Face Detection Algorithm using Face Likelihood Distribution

Hiromasa Takatsuka Masayuki Tanaka Masatoshi Okutomi

Graduate School of Science and Engineering,
Tokyo Institute of Technology

Face detection is a useful technique in computer vision. Many face detectors have been developed in the literature. Almost all approaches for face detection focus on the face detectors which classify a given sub-window into face or non-face. However, in face detection process, since the detectors also evaluate the scanned sub-windows independently, non-faces with high face likelihood are often misdetected. In this paper, we propose a novel face detection algorithm which explicitly uses difference of face likelihood distribution between faces and non-faces. The proposed algorithm can correctly classify the non-faces misdetected by the existing algorithm. The face likelihood distribution is generated and integrated to emphasize the difference between faces and non-faces. Experiments with pre-scanned data set and real-world images show that the proposed algorithm improves the detection rate approximately by 20% and 10%, respectively.

1 はじめに

画像中から顔を検出する“顔検出技術”は画像処理における基盤技術であり、監視カメラや顔認証システムといった高度な画像システムにおいてなくてはならない重要な技術である [3, 6]。近年、顔検出の分野はめざましい進歩を遂げており、様々な顔識別器、および学習アルゴリズムが提案されている [5, 4, 6]。たとえば、Rowley らはニューラルネットワークを用いた識別器を提案している [4]。また、Viola と Jones はカスケード構造を成す複数の AdaBoost ベースの識別器を提案している [5]。Lienhart らは異なるパラメータを持つ様々なタイプの AdaBoost ベースの識別器の比較を行なっ

ている [2]。また、識別器に用いられる特徴量の改良を行なっている研究もいくつか存在する。Lienhart らは、AdaBoost ベースの学習器に用いられる様々なタイプの Haar-Like 特徴に関しても検討を行なっている [2]。また、三田らは複数の Haar-Like 特徴量を組み合わせることで、新たな特徴量を作り出すアプローチを提案している [7]。

しかし、これらの研究は識別器、およびその特徴に着目したものであり、シーン中から顔を検出することを直接的に目的とはしていない。識別器は切り取られたサブウィンドウが顔かどうかを識別する機能だけしか持っていないため、従来の研究においては、入力画像をサブウィンドウに分割する前処理と、識別器の処

理結果をマージし、顔検出の処理結果として出力する後処理とを組み合わせ、入力画像からの顔検出の機能を実現している。

従来の顔検出の枠組みの中で顔識別器は各サブウィンドウごとに独立して評価を行なっているため、識別器が高い評価値(顔らしさ)を示す非顔画像のサブウィンドウを、誤って顔として検出してしまうことがある。

本報告では、まず顔画像と非顔画像のサブウィンドウが同様に高い顔らしさを持っていたとしても、その近傍における顔らしさ分布はそれぞれ異なることを明らかにする。次いでサブウィンドウを独立で評価するのではなく、近傍の“顔らしさ分布”まで考慮できる新しい顔検出手法の枠組みを提案する。この手法を用いることで、従来手法では誤検出していた非顔画像を正確に分類できる。

まず、提案手法では従来提案されている顔識別器を用いて顔らしさ分布を生成する。次に、この顔らしさ分布の位置および大きさに関して統合する。この統合処理により、顔と非顔の違いをより明確にすることができる。最後に、統合後の顔らしさ分布から顔の検出を行なう。

以下では、まず AdaBoost 識別器に関して説明をする。次に、顔画像と非顔画像における“顔らしさ分布”の違いについて具体的に示す。そして、この顔らしさ分布を用いた新しい顔検出手法について説明し、提案手法を用いた実験を行なう。実験では、画像から顔を検出する際の提案手法の有効性について確認する。

2 AdaBoost 識別器

本研究では、顔らしさ分布を作成するために AdaBoost 識別器を用いた [1, 5]。AdaBoost 識別器は従来手法の中でも実用性に優れた方法として知られている。詳細なアルゴリズムは参考文献 [1, 5] に示されている。

ここでは、簡単に説明をする。まず、AdaBoost 識別器の出力は、式 (1) のように複数の弱仮説の重み付きの足し算で与えられる。

$$H(x) = \sum_{i=1}^T \alpha_i h_i(x) \quad (1)$$

ここで T は弱仮説の総数、 $H(x)$ は識別器の出力、 α_i は各弱仮説の重み、 $h_i(x)$ は弱仮説の出力である。ただし、通常の AdaBoost 識別器は $H(x)$ に対して閾値処理したものを最終出力としている。 $H(x)$ は -1 から 1 の

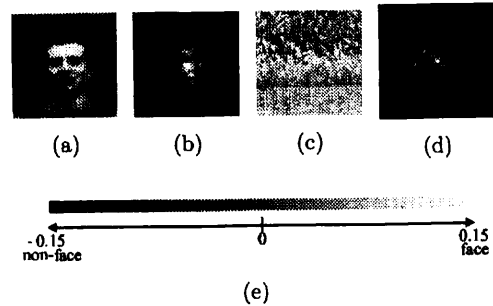


図 1: 顔らしさ分布: (a) 顔画像例, (b) 顔画像 (a) の顔らしさ分布, (c) 非顔画像例, (d) 非顔画像 (c) における顔らしさ分布, (e) 検出器の出力と表示色の対応。

値をとり、値が高いときには対応するサブウィンドウ x は顔である可能性が大きく、低いときには顔である可能性が小さい。我々は、この識別器の出力 $H(x)$ をサブウィンドウ x の“顔らしさ”と呼んでいる。

本研究では、web から集めた 5,000 の顔画像と 5,000 の非顔画像の学習データを用いて、学習を行なった。5,000 の顔画像は人の手で切り取り、非顔画像は顔の含まれていない画像よりランダムに生成した。全ての画像の大きさは 20×20 ピクセルにそろえられ、そこから 20×20 の識別器を学習した。また、この AdaBoost 識別器に使用した弱仮説の総数は 500 である。

3 顔らしさ分布

本章では、顔らしさ分布の特徴を示す。我々は本当の顔 (true positive) と、あるサブウィンドウで高い顔らしさを示す非顔 (false positive) における顔らしさ分布の違いを実験的に調べた。これら二つのサブウィンドウ近傍の顔らしさ分布には違いがある。この違いを利用することで高い顔らしさを示す非顔を正しく分類できる可能性がある。

3.1 位置方向における顔らしさ分布

図 1 は、顔画像と非顔画像において、サブウィンドウの位置を変化させた際の顔らしさ分布を示している。サブウィンドウの大きさを固定し、各サブウィンドウの顔らしさをサブウィンドウの中心位置にプロットした。また、各画像は顔らしさが最大になる位置が中心になるように調整されている。図 1 の (a), (c) には顔らしさの最大値がほぼ同じになる顔画像と非顔画像を選んだ。顔画像の顔らしさ分布 (b) においては、中心

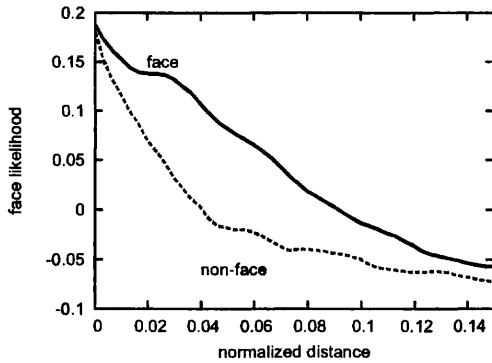


図 2: 位置方向における顔らしさ分布: 本当の顔 (true positive) と誤検出された非顔 (false positive) 近傍の顔らしさ分布の比較

から離れた領域においても顔らしさが高い値を持っていることが分かる。一方で、非顔画像の顔らしさ分布 (d) では、高い顔らしさを持つ領域が、(b) と比較して小さいことが確認できる。

図 2 は顔画像、非顔画像の顔らしさ分布を比較したものである。横軸はサブウィンドウの大きさで正規化された中心からの距離を示している。縦軸は顔らしさの値を表す。図 2 では、図 1 で使用したのと同じ顔画像と非顔画像を使用している。本当の顔 (true positive) 近傍の顔らしさ分布は誤検出された非顔 (false positive) の顔らしさ分布と明らかに異なる。中心付近では双方の顔らしさの値にあまり差はないが、距離 0.04 付近においては、顔画像と非顔画像の顔らしさの値に大きな差がみられる。もし、中心のサブウィンドウの顔らしさのみを使用して (c) を評価した場合、(c) は顔として検出されてしまう。しかし、近傍の顔らしさ分布を含めて判断した場合には (c) の画像を正しく分類することが可能である。

3.2 スケール方向における顔らしさ分布

スケール方向の顔らしさ分布に関しても位置方向の場合と同様の傾向を示す。図 3 は、サブウィンドウのスケールを変化させた際の本当の顔 (true positive) と誤検出された非顔 (false positive) の顔らしさ分布を示している。横軸は最大値をとるサブウィンドウのサイズを基準としたスケールで、縦軸は顔らしさである。これらは 3.1 章で使用したのと同じ画像を用いてプロットした。

スケール方向で比較した場合にも、二つの顔らしさ

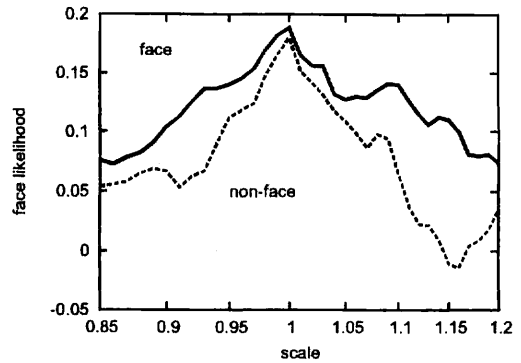


図 3: スケール方向における顔らしさ分布: 本当の顔 (true positive) と誤検出された非顔 (false positive) 近傍の顔らしさ分布の比較

分布の間には明確な差が確認できる。スケールの値が 1 のときには二つの顔らしさ値の差は小さい。一方で、スケールが変化した際には、双方の値の差は大きくなっている。たとえば、スケールが 1.1 のときは、顔らしさには約 0.1 の差が生じている。本当の顔の顔らしさ分布の方がゆるやかに変化するのに対し、誤検出された非顔の分布は急速に変化していることが分かる。

4 近傍の顔らしさ分布を利用した顔検出手法

本章では、新しい顔検出手法を提案する。前章でしめた顔らしさ分布は、既存手法においては陽には利用することができない。一方、我々の提案する手法は前章で示した顔らしさ分布を陽に利用し、従来手法では誤検出を引き起こす高い顔らしさを持つ非顔を正しく分類することを可能とする。

4.1 従来の顔検出手法

図 4 は従来手法と提案手法の比較を示している。従来の顔識別処理は顔検出処理の枠組みにおける一つのプロセスでしかない。そのため、実シーン中から顔検出を行なうためには、サブウィンドウに分割するための前処理と、識別器の出力から顔検出の出力を得る後処理が必要となる。

まず、前処理では入力画像を走査し、サブウィンドウを生成する。その後、従来の顔識別器が各サブウィンドウに対して独立に顔らしさを評価する。このとき、識

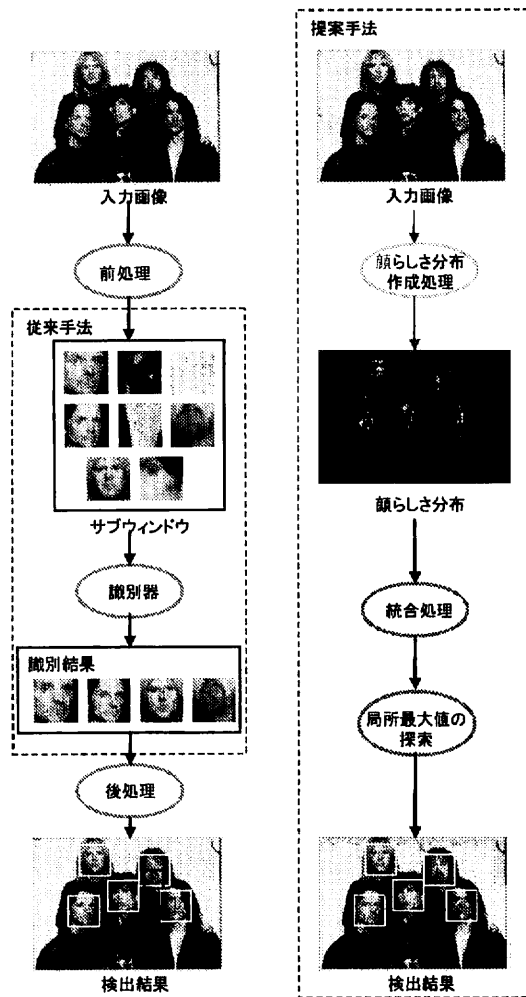


図 4: 従来手法と提案手法の比較

別器の出力が閾値以上であれば、顔として分類し、閾値以下であれば、非顔として分類する。識別器が顔として識別したサブウィンドウを直接表示すると、一つの顔の周りに複数個の検出候補が出現することがある。そのため、最後に後処理として、これらをグループングし、最終的な検出結果としている。

4.2 提案する顔検出手法

従来の顔検出においては、顔識別器の入出力に合わせ経験的に前処理と後処理を加えている。そのため、顔識別処理においては、サブウィンドウ毎に独立に評価・閾値処理をして顔を検出している。このとき、ある非

顔のサブウィンドウが高い「顔らしさ」を示すと、顔として誤検出してしまうことがよくある。

一方で、提案手法はシーン中から顔を探すことを積極的に考慮している。そのため、顔らしさ分布の違いを陽に利用することができ、高い「顔らしさ」をもつ非顔画像を正しく分類することができる。さらに提案手法を使用することで、従来手法では検出できなかった顔も正しく分類することが可能となる。図 4 の右側のプロセスが提案手法の流れを示している。まず、既存の顔識別器を用いて、顔らしさ分布を作成する。この顔らしさ分布は縦位置、横位置、スケールの三次元の分布を持っている。顔らしさ分布を作成した後、この分布に対して統合処理を行なう。統合処理では、近傍の分布に対して重みつき足し算を計算する。この処理を行なうことで、ある非顔画像のサブウィンドウが高い顔らしさを示したとしても、近傍の分布を含めて評価するため、統合後の顔らしさの評価は低くなる。これは、3章で示したように、非顔画像の近傍の分布は顔画像に比べ低い顔らしさをとるためである。統合処理後、画像全体で最大値を探索し、顔とする。次に二番目、三番目…と順に閾値を超える最大値がなくなるまで探索を繰り返す。

4.3 位置方向の顔らしさ分布の統合

顔画像と非顔画像における顔らしさ分布を比較する方法について提案する。本研究では、近傍位置の顔らしさの重みつき足し算で表現した統合処理により顔らしさ分布の比較を行なった。この処理はあらかじめ算出した、本当の顔の顔らしさ分布との相関を求めることに相当する。まず、位置方向における顔らしさ分布の統合について考える。

$$f(x) = \sum_{u \in L(x)} w_l \left(\frac{|u-x|}{sub} \right) d(u) \quad (2)$$

ここで、 $d(u)$ は位置 u における顔らしさ、 $f(x)$ は統合後の顔らしさ、 $w_l(l)$ は重み関数、 $L(x)$ は統合する領域、 sub はサブウィンドウの大きさをそれぞれ表している。

式 (2) において、重み関数の設計は検出結果を向上させるのに大きな影響を持つ。本研究では、この重み関数を統計的に設計した。まず、50 個の顔画像を集め、それぞれの顔らしさ分布を求める。そして、顔らしさ分布の平均に一次関数で近似したものを重み関数とした。近似に利用したデータは顔らしさが 0.1 以上のもの

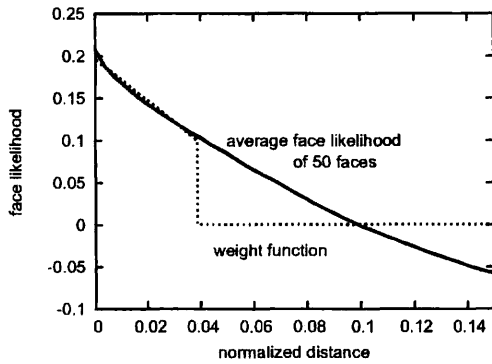


図 5: 位置方向の重み関数: 実線は顔らしさ分布の平均。破線は重み関数。

に限定した。その結果、最終的に得られた重み関数は式 (3) である。

$$w_l(l) = \begin{cases} -2.45l + 0.195, & \text{if } l < 0.039 \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

図 5 における実線が平均した顔らしさ分布、破線がこの重み関数を表している。

4.4 スケール方向の顔らしさ分布の統合

位置方向の顔らしさ分布の統合と同様に、スケール方向の統合も重みつき足し算で表される。

$$f(y) = \sum_{v \in S(y)} w_s \left(\frac{|v - y|}{sub} \right) d(v), \quad (4)$$

ここで、 $d(v)$ はスケール v における顔らしさ、 $f(y)$ は統合後の顔らしさ、 $w_s(s)$ は重み関数、 $S(y)$ は統合する領域、 sub はサブウィンドウの大きさをそれぞれ表す。重み関数は位置方向の顔らしさ分布の統合と同じく、平均値を近似し、設計した。図 6 の実線が平均した顔らしさを表している。

まず、スケールの重み関数として、 s を定義した。

$$w_s(s) = a|\log(s)| + b \quad (5)$$

顔らしさが 0.1 を超えているデータを用いて、フィッティングを行なった。その結果、得られた重み関数は、

$$w_s(s) = \begin{cases} -0.633|\log(s)| + 0.191, & \text{if } 0.866 < s < 1.155 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

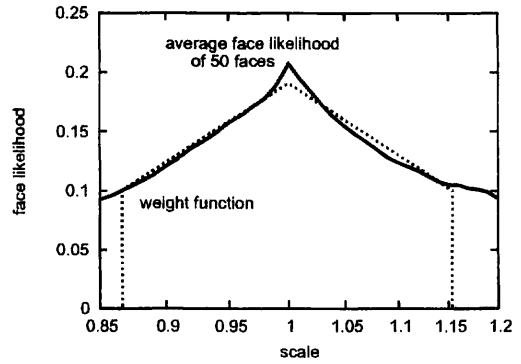


図 6: スケール方向の重み関数: 実線は顔らしさ分布の平均。破線は重み関数。

である。

4.5 位置・スケール方向の顔らしさ分布の統合

次に、位置方向とスケール方向の顔らしさ分布を共に統合する手法について説明する。このとき、顔らしさ分布は三次元になるので、統合の式は式 (7) となる。

$$f(x, y) = \sum_{u \in L(x)} \sum_{v \in S(y)} w_l \left(\frac{|u - x|}{sub} \right) w_s \left(\frac{|v - y|}{sub} \right) d(u, v) \quad (7)$$

位置・スケール方向の顔らしさ分布の統合を用いると、位置とスケールの二つの情報を統合することになるので、よりよい検出結果が得られることが期待される。

5 実験結果

実験をはじめると、AdaBoost 識別器の評価を行なった。AdaBoost 識別器においては、識別器の検証の際に用いるテストデータセット中の顔画像と非顔画像の比によって結果が大きく異なるため、注意が必要である。

図 7 に顔画像の数を固定し、非顔画像の数が異なる複数のテストセットを用いた ROC カーブが示されている。横軸、縦軸はそれぞれ誤検出の数と、検出率を表している。たとえば、非顔画像の数が 7,085,540 個 (比率は 1/10900) を用いた場合、誤検出数が 400 で、検出率は 90% を超えている。一方で、500,365,159 個 (比率は

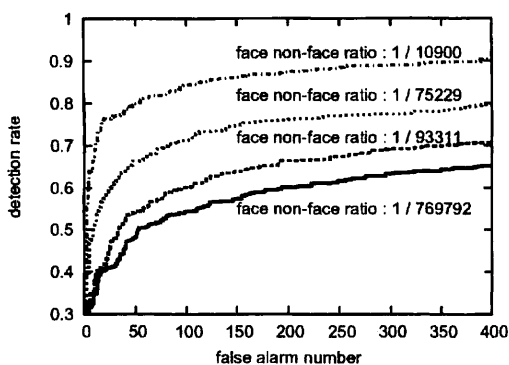


図 7: 異なる非顔画像の比率のテストセットを使用した ROC カーブ: 横軸は誤検出の数, 縦軸は検出率. 数値はそれぞれの ROC カーブに対応する顔画像と非顔画像の比率.

1/769792) の非顔画像を使用したときには, 同じ 400 の誤検出数にもかかわらず検出率は 65% になってしまう.

今回の実験では, 顔画像 650 個と非顔画像 500,365,159 個を合わせたカードタイプのテストセットを使用する. この顔画像と非顔画像の比率は, VGA サイズの画像中に顔が 10 個あるときの比率とほぼ同じであり, 妥当な人数だと考えられるためである.

5.1 カードタイプのデータセットによる検証

カードタイプのデータセットとは, 従来の識別器を評価する際に良く用いられる画像中からあらかじめ切り出されたデータセットのことを示す. まず, このカードタイプのデータセットを用いて提案手法の評価を行った. 今回使用したデータセットは 650 の顔画像と 500,365,159 (約 5 億) の非顔画像から構成されている. 顔画像のデータは全て実際の顔よりも少し大きく切り出した. これは, 提案手法が顔の近傍領域まで含めて顔かどうかを判断する為である.

ここで, 従来手法と提案手法が扱う情報量が異なることに注意する. 従来手法はサブウィンドウに対応する領域の情報しか用いていないのに対して, 提案手法においては, 統合処理においてサブウィンドウの近傍領域を間接的に用いているので従来手法よりも多くの情報を利用していると考えられる. しかし, シーン中から顔を検出する際には顔領域の周りにはこのような近傍領域が存在する. 近傍領域は従来の枠組みでは利用していないだけである.

ここでは顔らしさ分布の統合について, 統合なし, 位

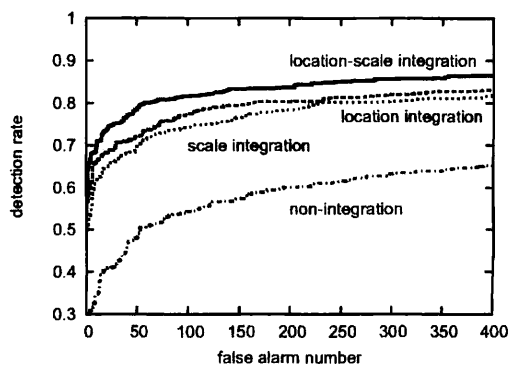


図 8: カードタイプのテストセットを使った際の ROC カーブ: 横軸は誤検出の数, 縦軸は検出率.

置方向の統合, スケール方向の統合, 位置・スケール方向の統合の四つの手法について比較検討する. 統合なしの場合には, 画像の中心のサブウィンドウにおける顔らしさをその画像の評価値とし, そのほかの手法の場合には, 中心の近傍の顔らしさ分布を統合した, 統合後の顔らしさを画像の評価値とする. 閾値を変化させ, ROC カーブを算出した. 図 8 に結果を示す.

提案する統合手法を適用した場合, 統合なしの場合と比較して誤検出数が同じときの検出率が改善していることが分かる. 特に, 位置・スケール方向の統合を共に適用した場合には, 約 20% 検出率が向上している. また, 位置方向とスケール方向の統合を単独で適用した場合には, 位置方向の統合を適用した ROC カーブの方がよりよい結果を示す. これは, 位置方向における顔と非顔の顔らしさ分布の違いの方が, スケール方向の違いに比べ, 顕著であるためだと考えられる.

5.2 実シーンによる検証

顔検出の本来の目的はシーン中から顔の位置と大きさを検出することである. よって, 顔検出手法の評価としてはカードタイプのテストセットを用いた顔識別器の評価ではなく, シーン中から顔を検出した最終的な結果に対する評価の方がより適切であると考えられる.

実験では, 395 個の顔画像を含んだ 119 枚の画像を用いた. このテスト画像は CMU+MIT テストセット [4] の一部と Web から集めた画像から成る. サブウィンドウの総数は 618,360,931 であった. 5.1 章での実験同様, 四つの統合方法について比較を行った.

この実験では, 最終的な出力結果を持って評価を行なう. よって, 図 4 のプロセスを実行した結果, 検出

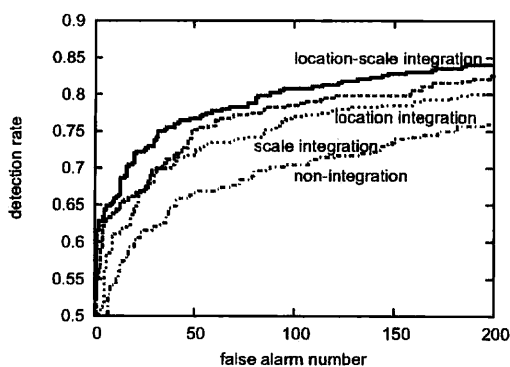


図 9: 実シーンを使った際の ROC カーブ: 横軸は誤検出の数. 縦軸は検出率. 提案手法を適用することで, 検出率が改善していることが分かる.

されたサブウィンドウがあらかじめ与えられた正解の値に近いときに, その顔が検出されたと判定し, 検出率を計算した. 正解の値はあらかじめ人間が与えた.

図 9 にこの実験における ROC カーブを示す. 実シーンをういた場合にも, 全ての範囲において提案手法による統合処理を行なった処理の方が検出率がよくなる事が分かる. また, 提案手法の中では位置・スケール方向の顔らしさ分布を統合した場合が, 最も検出率がよく, 統合をしないものと比較すると約 10% 検出率が改善している. この実験では, カードタイプでの評価より改善率が低い. これは, カードタイプでは人間が選んだ顔の中心と検出器の最大値が異なることがあるためだと考えられる.

図 10 は MIT+CMU テストセットの画像の検出結果である. 正しい顔を検出した結果は白い枠で, 誤検出した結果は黒い枠でそれぞれ示してある. 閾値は使用したテスト画像全体の検出率が 85% になるよう図 9 から算出した.

統合処理をしていない図 10 の (a) を見ると, 15 個の誤検出が確認できる. 一方で, スケール方向の統合処理を行なった (b), 位置方向の統合処理を行なった (c), 位置・スケール方向の統合処理を行なった (d) はそれぞれ 6 個, 3 個, 1 個と誤検出数が減少していることが分かる.

その他の画像における検出結果を図 11 に示した. 閾値に関しては図 10 と同じものを使っている. 図 10 と同様, 提案手法の有効性が確認できる.

6 おわりに

本報告では, 新しい顔らしさ分布を陽に利用した顔検出手法を提案した. まず, 顔画像と非顔画像の顔らしさ分布を示し, その形状に明確な差が見られることを確認した. そして, 顔らしさ分布を統合する手法を提案し, 実際にこれらが検出率の改善に有効であることを実験結果を通して示した.

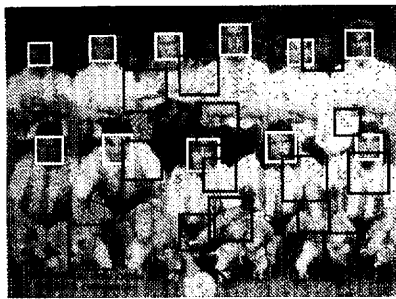
今回提案した手法は位置とスケールにおける顔らしさ分布に着目し, 近傍の分布を統合する. この統合によって, 顔画像と非顔画像の差が強調されるため, 検出率が改善すると考えられる. 従来手法ではこの顔らしさ分布を陽に利用することができなかった. この点において, 提案手法は新しいと考えられる.

実験では, AdaBoost 識別器を使用し, 提案手法の有効性を確認した. その結果, 統合処理を行なった場合, 統合処理を行わない場合と比較して, テストデータセットを使用した場合には 20%, 実画像を使用した場合には 10%, それぞれ検出率が改善した.

今後の課題としては, 提案した統合手法を他の識別器と組み合わせたとときの比較・検討が挙げられる. 特に, AdaBoost アルゴリズムを高速化した Boosted Cascade [5] の識別器に提案手法を組み合わせることが課題として考えられる.

参考文献

- [1] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, Vol. 38, No. 2, pp. 337–407 (2000).
- [2] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. *DAGM03*, pp. 297–304 (2003).
- [3] K. Okuma, A. Taleghani, N. de Freitas, J.J. Little, and D.G. Lowe. A boosted particle filter: multitarget detection and tracking. *In European Conference on Computer Vision*, pp. 28–39 (2004).
- [4] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 23–38 (1998).
- [5] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154 (2004).
- [6] M. H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34–58 (2002).
- [7] 三田雄志, 金子敏充, 堀修. 顔検出に適した joint haar-like 特徴の提案. 画像認識・理解シンポジウム (MIRU2005), pp. 104–111 (2005).



(a)



(b)



(c)



(d)

図 10: 実画像における顔検出結果: 四つの顔らしさ統合手法を用いたときのそれぞれの検出結果. (a) 統合なし, (b) スケール方向の統合, (c) 位置方向の統合, (d) 位置・スケール方向の統合.



Image 1 - (a)



Image 1 - (b)



Image 1 - (c)

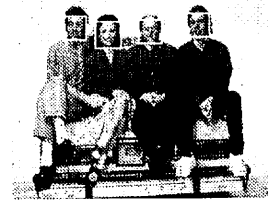


Image 1 - (d)



Image 2 - (a)



Image 2 - (b)



Image 2 - (c)

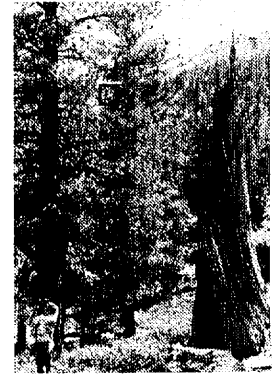


Image 2 - (d)

図 11: そのほかの顔検出結果: (a) 統合なし, (b) スケール方向の統合, (c) 位置方向の統合 (d) 位置・スケール方向の統合.