

連続特徴空間における決定木構築法と顔検出への応用

林 拓[†] 和田 俊和[†] 加藤 丈和[†]

[†]和歌山大学 システム工学部 システム工学研究科

〒 640-8510 和歌山県 和歌山市 栄谷 930

E-mail : hayashi@vrl.sys.wakayama-u.ac.jp, twada@ieee.org, t.kato@ieee.org

あ ら ま し 近年、顔認識技術を用いた実時間顔検出の研究が盛んである。中でも Viola らが提案したカスケード型識別器を用いた顔検出手法は、複数のスケールの異なる顔画像が含まれる顔画像からビデオレートで顔検出を行うことができ、注目されている。この手法は、画像中の大半を占める非顔画像を早い段階で判断し、除去することによって高速な検出を可能としている。しかし、顔の向きの変動に対応するなど学習データの多様性が増えると、カスケードの段数が増え、速度と精度の両立が難しくなるという問題や、学習に非常に時間がかかり、学習データが追加されると識別器全体の再学習が必要になるという問題がある。本報告では、カスケード型の識別器の代わりに、木構造を持つ決定木を提案し、これらの問題の解決を目指す。

キーワード 顔検出, 決定木

Decision Tree Construction in the Continuous Feature Space for Face Detection

Hiromu Hayashi[†], Toshikazu Wada[†] Takekazu Kato[†]

[†]Graduate School of System Engineering, Wakayama University,

930 Sakaedani, Wakayama, 640-8510, Japan

E-mail : hayashi@vrl.sys.wakayama-u.ac.jp, twada@ieee.org, t.kato@ieee.org

Abstract Recently, recognition based face detection is getting a major research field in Computer Vision. Especially, Paul Viola 's cascaded classifier based face detection method is most highlighted, because of its fast detection speed while changing the scale, position, and orientation of the template on the image. The advantage of this method is that most of the inputs, i.e., non-face images are rejected in early stage of the cascade. However, this method has two limitations 1) the length of cascade becomes longer while increasing the training images including multi aspect images of human faces, 2) re-training of whole classifier is required when we add a single training sample. In this paper, we propose a decision tree based classifier for solving these problems.

Keywords Face Detection, Decision Tree

1 まえがき

顔認識技術には、監視システムやマシンインターフェースなど様々な応用が考えられる。顔の認識を行うには、まず、画像中のどこに顔があるかを検出する必要があり、様々な応用システムを開発するためには、リアルタイムに顔を検出することが要求される。しかし、画像中には、大きさや向きが異なる

顔が存在し、これを高速かつ高精度に検出するのは困難である。

画像中の大半は非顔であり、顔の部分は少ない。この特徴に注目すると、非顔に関しては、識別に時間をかけずに非顔であると識別でき、顔の部分に関しては、識別に時間がかかっても正しく顔であると識別できる識別器ができれば、高速かつ高精度な顔検出が行えると考えられる。これを実現

した研究に Viola らが提案した AdaBoost[4] とカスケード型識別器を組み合わせた顔検出手法がある [3]。Viola らの顔検出手法を改良し、顔の向きまで対応した研究として、Bo Wu らの顔検出手法 [1] がある。

しかし、これらの手法は学習に時間がかかる問題や新たな学習データが追加された場合に再学習を行わなければならないなどの問題がある。本研究では、カスケード型の識別器を用いる代わりに、木構造を持つ決定木を用いることでこれらの問題の解決を目指す。

本稿では、第 2 章で顔検出手法の従来手法を述べる。第 3 章で提案手法を述べる。第 4 章で実験結果を、第 5 章で本研究のまとめを述べる。

2 従来研究

顔検出の従来研究として近年注目されている Viola の顔検出手法について説明する。また、Viola らの手法を改良した手法である清華大学の Bo Wu らの顔検出手法についても説明する。

2.1 Viola らの手法

Viola らの提案した顔検出手法が高速に行えるのは用いた識別器と特徴量に工夫があるからである。以下では、用いた識別器と特徴量について説明する。

2.1.1 カスケード型識別器

Viola らの手法で提案されている識別器は AdaBoost により学習した計算量の異なる顔検出用の強学習器をカスケード状につないだものである (図 1)。顔検出時は、画像から切り出した様々なサイズの矩形領域が顔であるか非顔であるかをカスケード型識別器を用いて識別する。この時に、カスケード型識別器を構成するある強学習器で非顔と識別されると、その矩形領域に対しては、それ以降の強学習器で識別を行わない。また、カスケード型識別器を構成する強学習器の計算量を少ないものから徐々に多いものにする事で厳密に顔を識別できる構造になっている。このような構造にすることで、画像中の大半を占める非顔を識別の早い段階で除去でき、顔の可能性のある矩形に関しては厳密に識別が行われるので、高速で精度の良い識別が可能となっている。

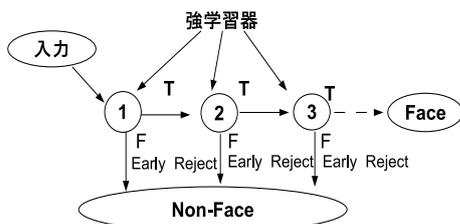


図 1: カスケード型識別器

2.1.2 Haar 型特徴量

Viola らは顔の特徴量に Haar 型特徴量を用いている。この Haar 型特徴量は矩形内の領域の平均輝度値の差であり、Integral Image を用いて解像度に関係なく高速に計算することができる。Viola らは図 2 の 4 種類の Haar 型特徴量を提案している。

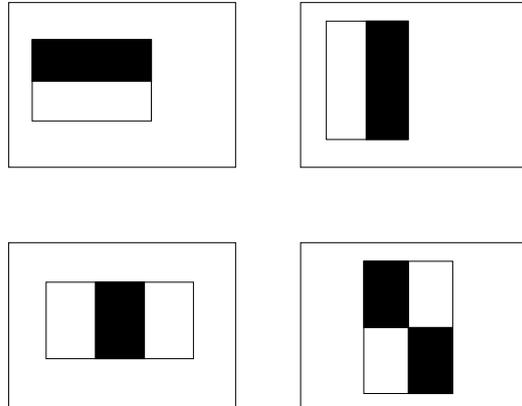


図 2: Viola の提案した Haar 型特徴量の矩形

2.2 Bo Wu らの手法

Bo Wu らの行った改良は主に次の 2 点である。まず、1 つ目は Schapire と Singer の Real AdaBoost[2] を用いた点である。Real AdaBoost は、サンプル空間 X から 2 値の予測でなく、確信度に関する実数の空間 R へ射影する弱学習器の学習アルゴリズムである。Viola らが用いた離散の AdaBoost では、1 か -1 のような離散的な出力しか得ることができないが、Real AdaBoost を用いることで連続的な値 (確信度) を出力できるようになる。これを用いることにより、より複雑な確率分布に対応できるようになっている。

2 つ目の改良はカスケード型の識別器の構成である。Viola らのカスケード型識別器では、カスケード型識別器を構成する各強学習器は独立しており、ある強学習器で識別を行う時には、その強学習器よりも前の段階の識別の結果が利用されていない。Bo Wu らは Real AdaBoost を用いることで識別結果を確信度とし、ある段階の強学習器で出力される確信度を次の段の強学習器を構成する弱学習器の 1 つとして利用する構造にした (図 3)。これにより、ある強学習器より前の識別結果を利用できるようにしている。

3 提案手法

3.1 概要

提案手法について述べる前に、簡単に決定木について説明する。決定木は与えられた学習データに対する分類規則を木構造で表現するものである。決定木の構築は次のようにして行う。まず、特徴空間を構成する各特徴に関して学習データを分割す

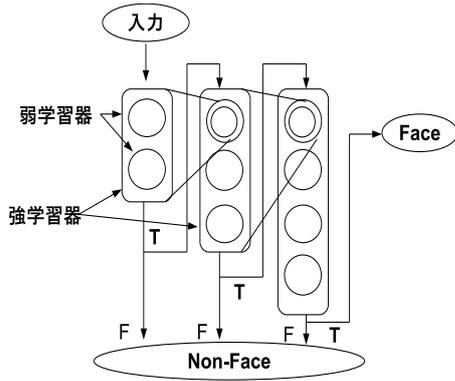


図 3: 入れ子型カスケード 識別器

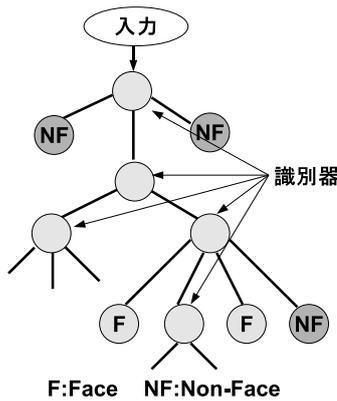


図 4: 提案手法

る．次に，定められた評価基準の下で分割を行うべき最適な特徴を選択する．評価基準には，Entropy Gain などがある．評価基準に Entropy Gain を用いた場合には，正しく分類できるデータ数が最大となる特徴が選択される．最後に分割された学習データが単一のクラスのデータになるまで，上記の処理を繰り返す．決定木の構築において問題となるのは，特徴選択である．

提案手法は，従来の決定木のように特徴空間を構成する各特徴で分割を行い，定められた評価基準の下で最適な特徴を選択するということを行わない．提案手法では，データを分割するための特徴を判別分析によって得られる判別軸とする．以下では，判別分析について説明した後，連続特徴空間における決定木の構築法について述べる．

3.2 判別分析

判別分析とは，フィッシャー比を最大化させる 1 次元特徴 A を求めることである．以下では，判別分析で得られる 1 次元特徴 A の求め方について説明する．まず，フィッシャー比とは，式 1 である．

$$\frac{A^T \Sigma_B A}{A^T \Sigma_W A} \quad (1)$$

式 1 の Σ_B と Σ_W は，以下の式で求める．

$$\Sigma_B = \frac{n_1 n_2 (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T}{n_1 + n_2} \quad (2)$$

$$\Sigma_W = \frac{n_1 \Sigma_1 + n_2 \Sigma_2}{n_1 + n_2} \quad (3)$$

ここで， μ_i はクラス i の平均ベクトルで， Σ_i はクラス i の共分散行列である．式 1 を最大化させる問題は， A の大きさが変わっても式 1 の値は変わらないので， $A^T \Sigma_W A = 1$ の下で $A^T \Sigma_B A$ の最大化させる問題とみなすことができる．ここで，Lagrange の未定係数法を用いると，式 4 が得られる．

$$J(A) = A^T \Sigma_B A - \lambda (A^T \Sigma_W A - 1) \quad (4)$$

J の最大値は $\Sigma_W^{-1} \Sigma_B$ の最大固有値であり，それに対応する固有ベクトルが A となる．

3.3 構築アルゴリズム

図 5, 6 に構築アルゴリズムの擬似コードを示す．この擬似コードに沿って連続特徴空間における決定木の構築方法について説明する．決定木の内部ノード τ には，以下の変数郡が格納されているものとする．

$D(\tau)$: τ に含まれるデータ集合

$A(\tau)$: τ のデータ集合に対する分割軸

$SplitPoint(\tau)$: τ のデータ集合に対する分割位置

$Label(\tau)$: τ のクラスラベル

あるノード τ に含まれるデータ集合 D に異なるクラスラベルのデータが存在するかどうかを調べる．これが *checkdata* である．1 つのクラスのデータしか無ければ，そのノードは分割を停止し，ノード τ に含まれるデータ集合のクラスラベルを出力とする終端ノードとなる．異なるクラスラベルのデータが有れば，線形な超平面で特徴空間を分割する．この分割によって，1 つのクラスラベルのデータしか存在しない領域と異なるクラスラベルのデータが交じっている領域に分ける．特徴空間の分割についてももう少し詳しく説明する．まず，データ集合 D に対して判別分析 (*Discriminant Analysis*(D)) を行い，得られる判別軸 A を分割軸とする．次に，判別分析によって得られた分割軸 A にデータ集合 D を射影する (図 7(a))．そして，判別軸上でヒストグラムを作り (*MakeHistogram*(A, D))，他クラスのデータがない領域と異なるクラスラベルのデータが交じった領域の境界を分割位置として保存する (図 7(b), (c))．保存した分割位置でデータ D を分割し，ノード τ を親ノードとする子ノードを作る (図 7(d))．

しかし，常に他クラスのデータがない領域と異なるクラスラベルのデータが交じった領域の境界が存在するとは限らない (図 8(a))．図 8(a) のよう

に境界が存在しない場合には，分割軸に直交する線形な超平面でデータ数が等しくなるように特徴空間を分割し ($Dummy(D, A)$)，2つの子ノードを作成する (図 8(b), (c), (d))．新たに作られた各子ノードに対して上記の処理を再帰的に行い決定木を構築する．

検出時には，中間ノードに保存してある分割軸に入力を射影し，分割軸上での入力的位置とノードが持つ分割位置との比較を行い，たどるべきノードを決定し次のノードをたどる．終端ノードに到達するまで，同様の処理を繰り返す．終端ノードに到達すると，そのノードが持つクラスラベルが識別結果となる．

```

Procedure Built-DecisionTree( $\tau, D$ )
1:begin
2: ConstructTree( $\tau_0, D$ );
3:end

```

図 5: Pseudo code for Bulding DedisionTree

```

Procedure ConstructTree( $D$ )
1:begin
2: if checkdata then
3:    $A = DiscriminantAnalsis(D)$ ;
4:    $MakeHistogram(A, D)$ ;
5:    $SplitPoint = SearchCutSpace(A, D)$ ;
6:   if  $SplitPoint > 0$  then
7:     foreach  $SplitPoint$   $i = 1:Child$  do
8:        $ConstructTree(\tau_i, D_i)$ ;
9:     end
10:  else
11:     $SplitPoint = DummyCut(A, D)$ 
12:     $ConstructTree(\tau_1, D_1)$ 
13:     $ConstructTree(\tau_2, D_2)$ 
14:  end
15: end
16:end

```

図 6: Pseudo code in Constructing phase

4 実験

4.1 実験概要

正面顔検出の検出実験について説明する．CAS-PAL の顔データベースの正面顔画像から顔部分を切出した画像 1029 枚から顔の学習データを取った．特徴量の数は 37 である．残りの 10 枚の正面顔画像を検証用の画像とした．また，非顔のデータには，研究室の背景などから得た 6000 個程度のデータを用意した．与えた学習データを用いて決定木

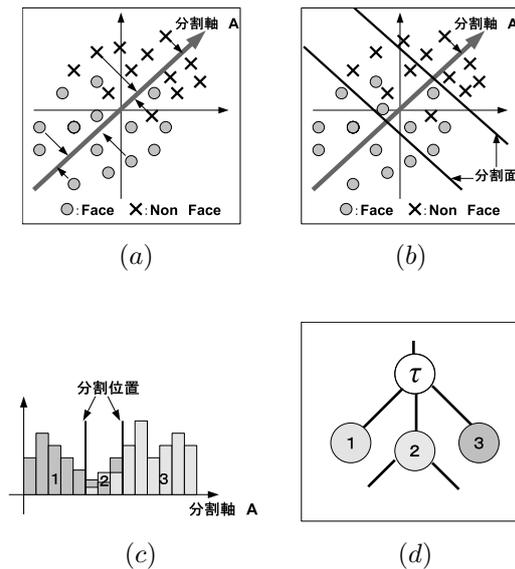


図 7: 判別分析

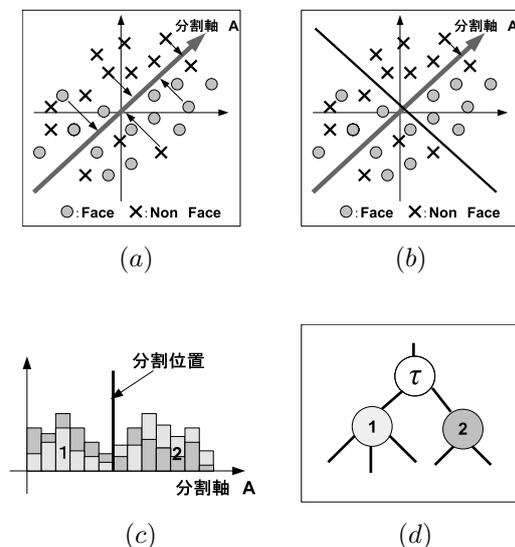


図 8: ダミーカット (a) はデータ集合 D に対して判別分析をし，分割軸にデータ集合 D を射影した図である．(b) は分割軸に直交する超平面でデータ数が等しくなるように分割した図である．(c) は判別軸上でのヒストグラムを示している．(d) はノード τ で作られる子ノードを示している．

を構築し，CASPAL の顔部分を切出す前の正面顔画像 1029 枚に対して検出を行い，誤検出された矩形の特徴量を非顔のデータとして登録する．検出に用いた 1029 枚の正面顔画像の顔の位置と大きさは既知であることから，誤検出の判定は，既知の顔の位置と検出された矩形の重なりで行っている．非顔データを追加した学習データで同様の事を繰り返した．

4.2 実験結果

上述の学習を 82 回繰り返した学習データを用いた場合の検証用画像の検出結果を図 9 に示す。図 9 より、多少の誤検出もみられるが、顔が検出されていることが分かる。この時に構築される決定木の深さはルートノードを 0 とすると、最大で 94 であった。次に、検出時に非顔と識別された矩形が決定木のどの段階で識別されたかを調べた。図 10 の実線で決定木の深さと非顔と識別されたデータの割合の関係を示す。図 10 の実線より、決定木の深さが 20 あたりで非顔と識別される矩形の数が急激に増えていることが分かる。これの原因について調べると、次のようなことがわかった。深さ 18 のノードの 1 つでダミーカットが起きていた。ダミーカットによって作られた深さ 19 のノードに対して分割の処理を行うと、深さ 20 のノードのみで学習データの総数の 15% ぐらいが取り除かれていた。この事から、ダミーカットを入れてデータを分割することにより、より識別に有効な特徴が選択されやすくなったと考えられる。また、図 10 から、深さが 10 あたりからダミーカットが起こった深さ 18 の間において、非顔と識別されたデータの数が非常に少ないことが分かる。さらに、学習時にこれらの深さで除去される学習データの数を調べると、非常に少ないことがわかった。これらの事から、深さ 10 から 18 までに選択されている分割軸は識別に対してあまり有効でないと考えられる。顔と識別される矩形の大半は、深さ 94 で識別されていた。また、あるノードで入力矩形を判別軸に射影し、次にたどるノードを決定するまでを 1 回の識別と数えることにすると、1 つの矩形に対する平均識別回数は、16 回であった。

ダミーカットが有効であるかを調べるために、決定木の学習段階で、除去できる学習データが少ない場合に、ダミーカットを行うようにして決定木を構築した。この決定木の深さと非顔と識別されたデータの割合の関係を図 10 の破線に示す。図 10 の破線より、より早い段階で非顔と識別できるようになっていることが分かる。また、この時の決定木の最大の深さは 20 であり、1 つの矩形に対する平均識別回数は 16 回から 4 回まで減っていた。このことより、ダミーカットを入れやすくなることで、非顔を識別の早い段階で除去できる決定木が構築できると考えられるが、誤検出は少し増加していた。

次に、検出の速度について調べた。320 × 240 の画像 1 枚に対して、検出の矩形を 24 × 24 から始め、矩形の大きさを 1.25 倍していき、240 × 240 まで変化させて検出を行った。提案手法では、検出を行うのに 0.75 秒と非常に時間がかかっている。検出に時間がかかる原因は、決定木の中間ノードから次にどのノードをたどるかを定めるために、入力矩形を分割軸に射影する必要があるからである。

最後に、CASPAL の顔データベースに登録されていない顔画像に対して、顔の検出を行った。結果

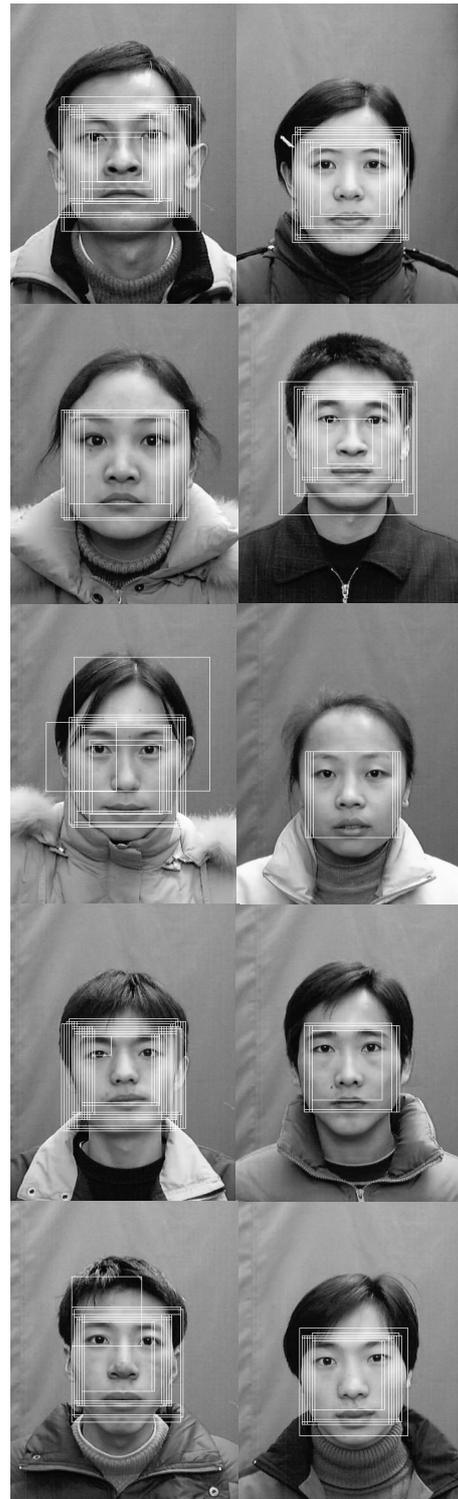


図 9: 検出成功例

を 図 11 に示す。図 11 を見ても分かるように、顔は検出されなかった。この原因は、特徴量に Haar 型特徴量を用いているため、登録された顔の照明条件と異なる照明条件の画像では、顔が検出されにくいからだと考えられる。この問題に対しては、様々な照明条件の顔データを登録することで解決できると考えている。

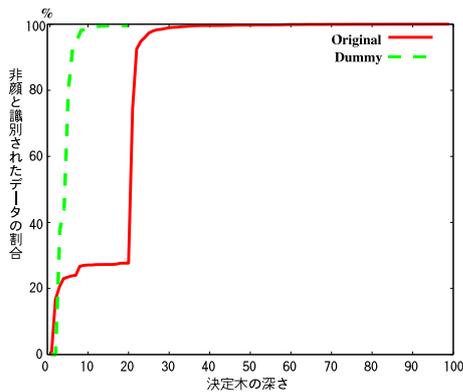


図 10: 決定木の深さと非顔と識別されるデータの関係



図 11: 検出失敗例

5 まとめ

連続特徴空間における決定木の構築方法を提案し、顔検出を行った。実験の結果より、提案手法を用いて、顔の検出が行えることを確認した。

今後の課題について述べる。課題の1つ目としては、検出時間の改善があげられる。提案手法の決定木をたどり識別を行うためには、各中間ノードで分割軸との内積を計算しなければならない。これが原因で検出に時間がかかる。これを改善する方法として次のようなことを考えている。判別分析で得られる分割軸を Walsh-Hadamard 基底などを用いてより少ない基底で判別軸を近似する。これにより内積の計算量が減り検出にかかる時間を減らすことができると考えている。

2つ目の課題は、分割に使用する軸の求め方である。本稿では、分割に使用する軸を判別分析で求めている。しかし、判別分析は、正規分布に従うようなデータに対しては有効な1次元特徴を求めることができるが、非顔のデータのようにほぼランダムに近い分布をしたデータに対して有効でない。このようなデータに対しては、Linear Support Vector Machine を用いて分割軸を求める方が識別に有効な分割軸を得ることができると考えられる。よって、決定木の構築に Linear Support Vector Machine を

導入することで、より検出の性能を上げることができると考えられる。

本稿で提案している決定木の出力は、Viola らの手法と同じように顔か非顔かのバイナリの値である。図 11 を見ても分かるように顔周辺に多数の矩形が検出されている。3つ目の課題は、決定木の出力を顔か非顔かのバイナリ値から顔らしさに変更し、Non-maximum Supression をかけることで、より顔らしい部分の検出を行うと言うものである。また、提案手法の決定木は、単一の終端ノードで識別を行っているが、これを終端ノードに到達するまでにたどった中間ノードの識別結果も利用し、より精度の高い識別を行えるようにするという課題も上げられる。

カスケード型識別器の代わりに決定木を用いる利点と次のようなことが考えられる。Viola らや Bo Wu らの手法では、新たな学習データが追加されると、再学習を行う必要がある。提案手法の決定木では、新たに追加されるデータを決定木で識別し、その結果誤識別した場合のみ、そのデータがたどったノードを変更することにより、識別器を変更することができると考えられる。この決定木の変更が可能となれば、提案手法はインタラクティブに顔検出を行えると考えられる。

参考文献

- [1] Bo Wu Haizhou Ai Chang Huang Shihong Lao. Fast rotation invariant multi-view face detection based on real adaboost. *FGR*, pp. 79–84, 2004.
- [2] Robert E. Schapire and Yoram Singer. Improved boosting using confidence-rated predictions. *Machine Learning*, Vol. 37, No. 3, pp. 297–336, 1999.
- [3] M Viola, P. and Jones. Rapid object detection using a boosted cascade of simple features. *Im Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, USA*, 2001.
- [4] Y. Freund and R. Schapire. Decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119–139, 1997.