2008－CVIM－161 (17)
2008／1／17

# ステレオカメラを用いた顔検出の高速化

鈴木 一正† 呉 海元‡ 和田 俊和‡

† 和歌山大学システム工学部 〒640-8510 和歌山市栄谷 930

E-mail: † {suzuki,wuhy,twada}@vrl.sys.wakayama-u.ac.jp

**あらまし** 単眼画像から顔検出する際には，画像内の顔の位置や大きさがわからないため，位置とスケールを変えた数万回もの識別を行う必要がある．ステレオカメラでは左右2枚の画像が得られるため，それらで全探索を行った場合は識別回数が非常に多くなる．本論文では，ステレオ処理により得られた距離情報を活かし，限られた位置とスケールのみをサーチすることによって識別回数を減らし，ビデオレートでの顔検出を実現する．また，ステレオ画像をサーチすることで，検出精度が向上すると考えられる．

**キーワード** 顔検出，ビデオレート，ステレオカメラ

# Video-rate face detection by using stereo-camera

Kazumasa SUZUKI† Haiyuan WU‡ and Toshikazu WADA‡

† Faculty of System Engineering, Wakayama University

930 Sakaedani, Wakayama-city, Wakayama, 640-8510 Japan

E-mail: † {suzuki,wuhy,twada}@vrl.sys.wakayama-u.ac.jp

**Abstract** In order to detect human face without knowing the position and the size in the image, tens of thousands of classifications are required. The stereo camera provides two images and the number of classification becomes numerous when full search is performed. In this paper, the number of classification is decreased by searching for the limited position and scale by using distance information, and the video-rate face detection is achieved. Moreover, it is considered that the accuracy of the detection system improves by searching the stereo image.

**Keyword** Face detection, Video-rate, Stereo-camera

## 1. はじめに

コンピュータビジョンの分野では，一般的に物体認識を行う場合には，まず，入力画像の中からどこにその物体があるかを検出する必要がある．本論文では，ひとつの例として，その物体を人間の顔とする．

画像中に含まれる人間の顔を検出する技術は，コンピュータビジョンの中でも重要な分野であり，監視カメラや，マシンインタフェース，ロボットとの対話など様々な応用がある．また，実際に様々な応用システムを開発するために，動画から顔をビデオレートで検出することを要求されている．

顔検出を行う場合，画像内の顔の位置や大きさがわからないため，位置とスケールを変えた数万回もの識別を行う必要がある．ステレオカメラでは左右2枚の画像が得られるため，それらで全探索を行った場合は識別回数が非常に多くなる．そのため，これまでは，高い識別率を持ちながら高速に顔検出できる識別器の構築法に関する研究開発がいくつなされた[1][2][3][4]．中には，識別回数を減らすために色情報を用いる方法[5]などはあるが，距離情報を使った研究はほとんど見当たらない．

提案手法では，距離情報によって識別回数を減らすことで顔検出の高速化を実現する．ステレオカメラでは，カメラに映る物体までの距離情報が得られる．また，人間の顔の大きさにはそれほど大きな違いがないため，距離情報から画像上での顔の大きさを推定することができる．このように，距離に応じたスケールのみを探索することで識別回数を減らし，探索を高速に行うことができる．さらに，ステレオ画像を探索に用いることで，一方で検出できない顔をもう一方で検出できる場合

など，検出率の向上も期待できる．

最後に，ステレオの実画像による実験より，提案手法は高速性かつ高検出率の両面に有効であることを確認する．

## 2. 関連手法と基本アイデア

本章では，本研究に関連する手法やその問題点について述べながら,提案手法の基本アイデアを説明する．

### 2.1. 顔検出

本節ではまず，本研究で探索時に用いる識別器について述べる．次いで，従来の全探索について説明する．

#### 2.1.1. Viola の識別器

顔検出の手法は数多く提案されているが,近年，Viola らが提案している AdaBoost 学習法とカスケード構造を組み合わせた手法[1]は，実行速度の速さと検出率の高さから世間で大変注目を浴びており，顔検出の研究と応用において広く利用されている.よって本研究では AdaBoost ベースの顔検出器を用いることにした[6].

AdaBoost 学習アルゴリズムでは低性能な特徴(弱識別器)を組み合わせて，全体として高性能な識別器(強識別器)が生成される．カスケード構造とは,結果として得られる識別器が AdaBoost で学習された幾つかの強識別器(層)から構成されているものである．強識別器は，いずれかの層でオブジェクト候補が却下されるか，あるいは全ての層をパスするまで次々に適用される．この構造により大半の識別は構造が単純な前の層，すなわち計算量の少ない層で識別してしまうため，計算効率が非常によく，高速な識別が可能になる．

#### 2.1.2. 大きさと位置が未知の顔探索方法

一般的に顔検出を行う際には，画像からサブウィンドウを切り出し，それを識別器に通すことで顔あるいは非顔と判断する．このとき，画像内の顔の位置や大きさは未知である．そのため，まず識別できる最小のサブウィンドウで画像中を走査し，続いてサブウィンドウ少し大きくして画像を走査する．これを繰り返し，画像内のあらゆる位置やスケールに対応したサブウィンドウについて識別を行う必要がある．

以上のように全探索を行うと識別回数は数万回にもなり，ビデオレートでの検出は難しい．ステレオカメラでは左右 2 枚の画像が得られるので，これらに全探索を行えば識別回数はさらに膨大なものになってしまう．

本研究ではステレオ処理によって得られた距離情報を活かすことで，識別回数を大幅に減らすことを実現する.次章でこの方法を詳しく説明する．

図 1　ウィンドウ走査

### 2.2. ステレオマッチング

ステレオマッチングとは，カメラ校正済みの左右に配置された 2 台のカメラで撮影された 2 枚 1 組のステレオ画像を用いて，右画像について距離を求めるときは，右カメラの画像内のある点が左カメラの画像のどの点と対応するかを相関の計算により求め，その対応関係を使った三角測量により，各点の 3 次元的位置を推測する方法である．

次の式(1)を用いて右画像と左画像の点の対応付けを求め，

$$\min_{d=d_{\min}}^{d_{\max}} \sum_{i=-\frac{m}{2}}^{\frac{m}{2}} \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} \left| I_{right}[x+i][y+j] - I_{left}[x+i+d][y+j] \right| \tag{1}$$

対応付けられた点の視差 $d$ から式(2)により奥行き $z$ を求める．

$$z = \frac{fB}{d} \tag{2}$$

ここで, $d_{\max}$ , $d_{\min}$ はそれぞれ最大の視差と最小の視差，$m$ はマスクサイズ,$I_{right}$ , $I_{left}$ はそれぞれ右画像と左画像，$f$ は焦点距離，$B$ はカメラ間距離である．

これを画像内の全ての点に対して行うと多くの処理時間を要してしまい，ビデオレートでの処理が不可能になる．そこで，本研究では，入力画像の全画素ではなくスパースサンプリングされた点のみで距離計算を行うことによって，検出精度を保証した上での処理時間を抑えるようにした．次章でこの方法を詳しく説明する．

## 3. ステレオカメラを用いた顔検出法

本研究ではステレオ処理によって得られた距離情報を用いることで，探索する領域とスケールを特定し，識別回数を減少させることで，ビデオレートでの顔検出を実現する．基本アイデアは，

(1)サンプリングによるステレオ処理の高速化

(2)識別回数の削減による検出の高速化

(3)ステレオ画像を用いた検出精度の向上

という3点である．以下それぞれの点について詳しく述べる．

### 3.1. サンプリングによるステレオ処理の高速化

前章で述べたように，入力画像の全画素に対してステレオマッチングを行うと，多大な計算時間となり，ビデオレートでの処理は難しい．そこで，一定間隔でサンプリングした画素に対してのみステレオマッチングを行うことで，ステレオ処理にかかる時間を抑えることが可能となる．図2(a)は320x240ピクセルの入力画像(右カメラ)で，図2(b)は入力画像についてサンプリングによるステレオ処理で得られた距離画像である．明るい画素ほど距離が近いことを意味している．

(a) (b)

図2　サンプリングによるステレオ処理の結果例

ここで，サンプル数は検出できる最小の顔サイズや，ビデオレートで処理可能な速度を考慮して80点としている．サンプルが少なすぎると，検出対象が遠くに映っているなど，画像上で顔が小さくなった場合に，検出対象がサンプル点同士の間に位置すると対象までの距離が計算できなくなるという問題がある．これはサンプル数が十分多ければ問題ないがいが，サンプルを増やすほど処理時間も多くなる．

また，各サンプルはノイズを減らすために5x5画素の領域において距離計算を行う．算出された距離値にはたいていの場合ノイズが含まれるので，これを低減するためにサーフェイスバリデイションフィルタを用いている。サーフェイスバリデイションフィルタとは，ある点において，そこで計測された距離値がその近傍の点の距離値と大きく異なっている場合に，その点を除去するものである．すなわち，計測された距離を一連のサーフェイス(物体表面)として認識し，認識できなかった点を除去するフィルタである．

このように，画像全体ではなくサンプルされた画素点のみ部分的に距離を計算することによって，ステレオ処理に要する時間をQVGAサイズの画像において10msec程度に抑えることができる．

### 3.2. 探索領域とスケールの決定

前節で述べた各サンプル点の距離値をもとに探索する領域とスケールの決定を行う．基本的な流れは次のようになっている．

(1)距離値から画像上での顔サイズを推定

(2)顔サイズから探索領域とスケールを決定

(3)探索領域を統合

以下それぞれについて詳しく述べる．

#### 3.2.1. 画像上の顔サイズの推定

まず，サンプル点の距離値から次式によって画像上での顔のサイズを推定する．

$$FaceSize = \frac{f}{z} RealFace \qquad (3)$$

ここで，*FaceSize*は画像上の顔サイズ，*RealFace*はワールド座標における顔サイズ，$f$はピクセル単位の焦点距離，$z$はサンプル点の距離値でワールド座標におけるカメラから顔と思われるところまでの距離である．

#### 3.2.2. 探索領域とスケールの決定

次に，式(3)で求めた顔サイズを用いて各サンプル点における探索領域とスケールを決定する．

探索領域は，図3に示すようにサンプル点から上下左右に顔サイズ分だけ離れた矩形領域とする．あるサンプル点において顔までの距離が得られたとき，そのサンプル点が顔のどこに位置しているかわからない．そのため，上述のような探索領域とすることで，サンプル点が顔上のどこに位置していても，探索領域内に顔が含まれるようになる．

探索するスケールは，識別器に通す際のサブウィンドウの大きさが式(3)で求めた顔サイズと同じになるようにする．また，顔の大きさには多少個人差があるので少し大きくしたスケールや，小さくしたスケールでも探索する必要がある．

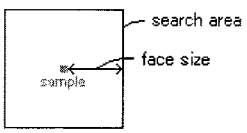


図3　各サンプル点における探索領域の決定

### 3.2.3. 探索領域の統合

前小節のようにして各サンプル点で探索領域を決めると領域の重なった部分が出てくる.例えば,図 2(b)を見ると画像中にある物体付近では,距離値のほとんど変わらないサンプル点が集まっていることがわかる.そのため,同じ位置を同じようなスケールで何度も探索するという事態が起こる.この事態を防ぐために探索領域が重なるサンプル点を距離値が近いときに限りひとまとまりにし,一つの探索領域とする.また,この領域内を探索する際のスケールは,まとめたサンプル点の距離値の平均から決める.

図 2(b)に対して,上述の処理を行った最終的な探索領域は図 4 のようになる.図中のそれぞれの矩形領域内を求められた特定のスケールのみで探索する.

さらに,ノイズによる無駄な探索領域を増やさないために,各領域において求められた顔サイズとサンプル数から,その領域を探索するか否かを決めるようにした.例えば,ある領域で,顔サイズが大きいほどサンプル点数も多くなるはずである.このとき,サンプル点数が少なすぎればノイズによる探索領域だと判断し探索しない.

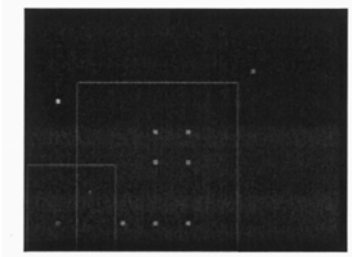

図 4　探索領域の統合例

### 3.3. 検出精度の向上

本研究で探索時に用いている識別器は,正面顔画像のみによる学習で構築されている[4].そのため,ある程度横を向いた画像などでは顔と識別することができない.

本研究ではステレオカメラを用いているので左右 2 枚の画像が得られる.このような左右 2 枚の画像での探索を行うと,一方の画像では少し横を向いていて顔と判断できない場合でも,もう一方の画像では視差があるため正面に近く顔と判断できるケースがある.このように何らかの影響で一方の画像では検出できなかった顔がもう一方の画像では検出できるということがあるため,ステレオ画像を探索することによって検出率の向上が期待できる.

また,全探索ではあらゆる位置とスケールを探索するのに対して,本手法では探索する位置とスケールが限られているため,識別回数の減少に伴って必然的に誤検出の数も減少する.

以上のことより従来の全探索に比べ検出精度が向上すると考えられる.

## 4. 実験

### 4.1. 実験環境

提案手法の処理速度と検出精度を確認する実験を行うにあたり 100 フレームほどの実画像系列(ステレオ画像 320x240 画素)を数種類用意した.

式(3)のワールド座標における顔の大きさ *RealFace* を 14cm と設定した.1 スケールだけの探索では個人による顔の大きさの違いや距離計算時の誤差などがあるため十分ではない.そのため,処理時間も考慮して左右 2 枚の画像で 2 スケール分を探索するようにする.実験環境は次のようになっている.

PC は Windows XP, Pentium(R) 4 CPU 3.0GHz, 1024MB RAM のものを使用し,ステレオカメラは Point Grey Research, Inc.-Bumblebee Stereo Vision Camera を使用している.プログラミング言語は Microsoft Visual C++を使用している.顔検出器は OpenCV の中のライブラリ関数を利用している.

### 4.2. 処理速度

用意した 2 種類の画像系列に対し,提案手法による顔検出実験を行った.ステレオ画像のうち一方の画像では求められたスケールで探索し,もう一方はサブウィンドウの大きさが 1.15 倍となるようにして探索している.

図 5,図 6 に 2 画像系列の処理時間のグラフをそれぞれ示す.

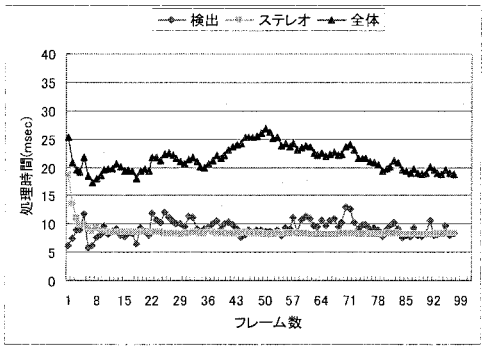


図 5　処理時間 1

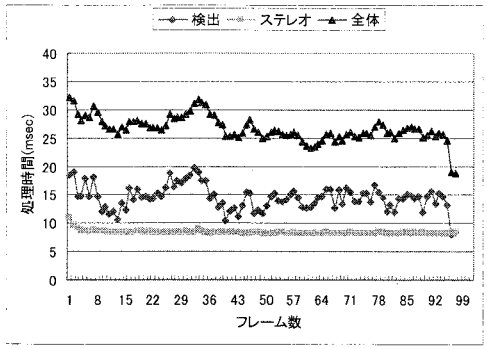


図 6　処理時間 2

横軸はフレーム番号で，縦軸は処理時間(msec)である．緑色(●)のグラフは検出にかかる時間，オレンジ色(■)のグラフはステレオ処理にかかる時間であり，青色(▲)のグラフは 1 フレーム全体の処理時間となっている．

図 6 で用いた画像系列は，カメラから少しはなれたところで人体の大部分が映っているような画像が多く，小さいスケールで探索領域が広い場合が多かったので，図 5 の結果に比べると，全体的に処理時間が多くなっている．両図から，ステレオ処理にかかる時間はほとんど一定で，10msec 弱となっており,検出にかかる時間によって全体の処理時間が変化しているが，ほとんどのフレームにおいて 33msec 以内で処理が終わっている．また，様々な種類の画像系列を用いた場合でも実験を行ったが，平均処理時間はおよそ 25msec となっている．

### 4.3. 検出精度

用意した 5 種類の画像系列に対して次の手法，

(1)単眼画像からの全探索（右画像を使用）

(2)提案手法で右画像のみの探索

(3)提案手法で左右の画像を用いた探索

の計 3 通りでの探索を行った．

単眼画像からの全探索の際には検出できる最小スケール(24x24 画素)から，スケールを 1.25 倍ずつ大きくしていった．また，提案手法で右画像のみの場合は，求めたスケールとその 1.25 倍の計 2 スケールを探索，提案手法で左右画像の場合は，右画像で求めたスケール，左画像でその 1.25 倍の計 2 スケールを探索した．一般的に，物体検出ではスケールを細かく見るほど検出率は上がるが，処理速度などを考慮してスケール変化を 1.25 倍としていることが多い．全探索では，およそ 60msec ほどかかっている．

比較実験の結果を表 1 に示す．表の数値は 101 フレーム中，顔を検出できたフレーム数を表している．

| 画像 | 系列 1 | 系列 2 | 系列 3 | 系列 4 | 系列 5 |
|---|---|---|---|---|---|
| 手法(1) | 90 | 87 | 83 | 59 | 42 |
| 手法(2) | 93 | 87 | 86 | 78 | 60 |
| 手法(3) | 94 | 87 | 86 | 87 | 65 |

表 1　手法による検出数の比較

ほとんどの場合に，単眼画像からの全探索に比べると提案手法の検出数が多くなっていることがわかった．

画像系列 1,2,3 についてはあまり変化が見られないが，これらの画像系列は顔がほぼ正面を向きながら移動していて比較的検出しやすい画像系列であったため，あまり差が見られなかったと思われる．また，単眼画像からの全探索では検出できているのに提案手法で検出できなかった例としては，ステレオ画像から顔までの距離を正確に求められず(例えば，画像の右端の方は視差の関係で正確な距離が出せない)顔を含むような探索領域が正確に設定されていない場合などがあった．

画像系列 4,5 については，正面とは少しずれた方向を向いている顔が多かったり，被験者の動きが早く，ぶれて顔がぼやけていたりするような画像があった．そのため，全探索での検出数は少なくなっているが，提案手法では大幅に上がっていることが分かった．(2),(3)の手法を比べると，このような画像の場合のステレオ画像での探索が有効であることが分かった．

また，(1),(2)の手法では同じ右画像のみによる探索であるが，提案手法の検出数が多くなっているのは次の 2 つの理由が考えられる．1 つは，単眼画像からの全探索の場合，スケールを 1.25 倍ずつ大きくしているため，顔がその間の大きさにであったときに検出されにくくなっていが，提案手法では距離に応じたスケールを探索しているため，そのような問題は起こらない．2 つ目は，単眼画像からの全探索では誤検出が多くなってしまうので，それを除去するために，近傍に自身を含めた 2 つ以上の顔が検出されないときはノイズと判断して除去する，という処理を行っているため，本当の顔も除去してしまっている場合がある．一方提案手法では，探索する位置とスケールが限定されているので，この処理をするほどの誤検出はな

いと思われた．

図7は検出結果の2例である．これらの画像は手法(3)による探索でしか検出することができなかった．つまり，右画像では検出できなかったものが左画像では検出できたということである．図7(a)では視差によって左画像ではより正面顔に近かったためであり，図7(b)では左画像の方が少し鮮明に映っていたため検出できたものと思われる．

(a) (b)

図7　検出結果例

### 4.4. その他の検出結果

複数の顔が映っている場合，距離が異なる位置にいると，スケールの異なる探索領域が重り，識別回数が増える．そのため,重なっているところはサンプル数の多い方のスケールのみ探索することで，識別回数を抑えている．この処理で多少検出率は落ちるが，ほぼビデオレートで処理できる．図8は結果例である．

図8　複数顔の検出結果例

## 5. まとめ

本研究では，顔検出の際の膨大な識別回数を削減するため，ステレオカメラを用いることによって得られる距離情報をもとに探索範囲を限定し，ビデオレートでの顔検出を実現する方法を提案した．また，計算コストが大きいステレオ処理については，全画素ではなくスパースに配置されたサンプル画素についてのみの計算によって処理時間を抑制する方法を提案した．提案手法を用いて，ステレオ画像を探索することによって検出率を向上させることを可能にした．

実画像系列を用いた実験によって，処理時間がビデオレート内に収まっていることを確認し，また，ステレオ画像を探索することの有効性も確認できた．

今回，実験で用いたQVGAサイズの画像ではビデオレートでの検出が可能であったが，VGAサイズの画像ではビデオレートでの検出は困難である．このサイズで十分な処理時間を実現するため，今後の課題としては，的確なサンプルの配置によって，ステレオ処理時間の更なる短縮と探索領域の絞込みが必要であると考えている．

## 謝　辞

本研究の一部は，文部科学省科学研究費補助金基盤研究(c)18500131の補助を受けている．

## 文　献

[1] Paul Viola and Michael Jones, “Rapid object detection using a boosted, CVPR, Vol.1, pp.511-518, 2001.

[2] Bo Wu, Haizhou Ai, Chang Huang, and Shihong Lao, “Fast Rotation Invariant Multi-View Face Detection Based on Real AdaBoost, FGR, pp.79-84, 2004.

[3] M.Propp and A.Samal, “Artificial neural network architectures for human face detection, Intelligent Eng. Systems through Artificial Neural Networks, Vol.2, 1992.

[4] Edgar Osuna, Robert Freund and Federico Girosi, “Training support vector machines:an application to face detection, CVPR, pp.130-136, 1997.

[5] Oliva Aude, Schyns Philippe, and Akamatsu Shigeru, “The role of color for face detection in a complex background, IEICE, Vol.96, No.499, pp.55-60.

[6] OpenCV, http://www.intel.com/technology/computing/opencv