

## 画像からの複数の情報の統合による顔の検出

### Face Detection by Integration of Multiple Cues from Images

○徐 明、赤塚 孝雄  
○ Ming XU, Takao Akatsuka

山形大学工学部電子情報工学科  
Yamagata University, Information Science and Electrical Engineering Department

キーワード：顔検出(face detection), 複数情報の統合(integration of multiple cues),  
色・運動・幾何・顔特徴 (color/motion/geometry/facial feature)

連絡先：〒9928510 山形県米沢市城南4-3-16 山形大学工学部電子情報工学科 赤塚研究室  
徐 明、Tel: (0238) 26-3323, Fax: (0238) 26-2082, Email: xum@eat72.yz.yamagata-u.ac.jp

#### 1. はじめに

画像から人間の顔を自動的に検出することは顔認識、データベース検索、very low bit-rate video coding、ヒューマンコンピュータインタラクションなど多数の応用分野において重要な課題である。画像から顔を検出するために、顔のパターンを統計モデルで表現し、パターン認識の問題として取り上げる方法が多用されている。しかし、人間の顔は変動の多いパターンである。例えば、顔の表情、特徴、色は人によって異なる。顔の化粧、眼鏡、ひげなどが顔の様子を変える。頭の姿勢や環境の照明状況も見た顔の様子を変えてしまう。人がいる環境や背景が複雑である場合、顔を検出することは更に難しくなる。

画像から顔に関わる情報をより多く抽出し、複数の情報を統合することは顔パターンの多様性多変性に対処するための有効な方法と考えられる。我々は顔の色と幾何情報を取り込んだマルチモジュールの手法を試みた[1]が、ここでは、顔と髪の色、頭の運動、顔の幾何性質、顔の特徴(瞳孔)に着目する。画像からこれらの情報を抽出する方法を議論した上

で、情報の統合による顔の検出について議論する。

#### 2. 色情報の利用

##### 2.1 色モデルの構築

色情報は画像から顔と他の物体を区別するための重要な手がかりである。色による検出手法は頭の姿勢によらず、高速な処理が可能である。顔の色分布が使用する色空間に係っているため、色空間の選択は色セグメンテーションの効果に影響してくる。RGB色空間はデジタル画像の表現に適しているため、RGB色空間を利用することにした。顔の色は様々に見えるが、その違いは主に照明強度によるため、正規化処理によって取り除くことができる。

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \quad (1)$$

$r + g + b = 1$ であるから、正規化されたRGB色空間は  $(r, g)$ 空間で表わすことができる。顔画像を  $(r, g)$ に変換し、二次元の色分布ヒストグラムを計算する。対象を東南アジア人として多数の異なる顔で実験した結果、顔色

の分布が (r, g) 空間の狭い領域に集中し、ガウシアンのような分布があることが分かった。従って、顔色の分布を二次元ガウシアンモデル ( $m_f, \Sigma_f$ ) で表示する。

$$m_f = (\bar{r}, \bar{g}) \quad (2)$$

$$\Sigma_f = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix} \quad (3)$$

$$\bar{r} = \frac{1}{N} \sum_{i=1}^N r_i, \quad \bar{g} = \frac{1}{N} \sum_{i=1}^N g_i \quad (4)$$

ここで、 $\bar{r}$ 、 $\bar{g}$  は色分布の r と g の平均値、 $\Sigma_f$  は r と g の分散値である。

顔画像の集合から、ペイントツールによって顔の領域を切出す。式(2)～(4)によって顔色の分布モデルを計算した。結果を Figure 1 に示している。

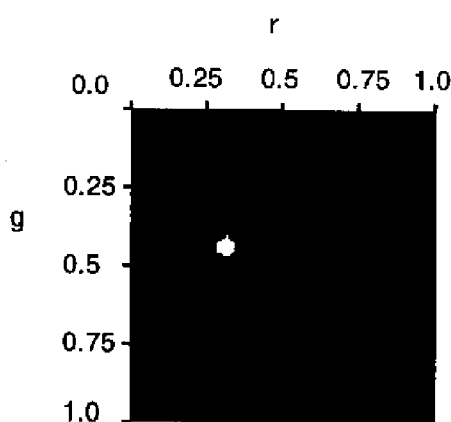


Figure 1. 顔色の分布モデル

人間の髪と顔の位置関係は、顔の位置を決めるための有用な手がかりである。髪を検出するため、髪の色分布モデルを構築する。アジア系の人の髪は殆ど黒いので、画像で髪の輝度値が低いため、色の反映が不安定になることがある。そのため、髪の色モデルには、色以外に、照明強度も要素として入れた。RGB空間を正規化する前に、強度  $I = (R + G + B)/3$  の平均値と分散値を計算しておく。

## 2.2 色変換

入力カラー画像は、式(5)、(6) に示しているように、各々の正規化された画素が顔色モデルに基づいて、算出する顔色との相違度によって濃淡画像に変換する。

$$D_{f(x,y)} = \sqrt{(r_{(x,y)} - \bar{r})^2 + (g_{(x,y)} - \bar{g})^2} \quad (5)$$

$$v_{f(x,y)} = 255 * (1 - \frac{D_{f(x,y)}}{Max}) \quad (6)$$

Max は距離値  $D_{f(x,y)}$  の最大値である。

同様に、髪色モデルに基づいて入力画像を変換する。

$$v_{(x,y)} = 255 * coeff * (1 - \frac{D_{(x,y)}}{Max}) \quad (7)$$

$$d(x,y) = |I(x,y) - m_l| \quad (8)$$

$$coeff = \begin{cases} 1 & \text{if } d_{(x,y)} \leq \sigma_i \\ \sigma_i / d_{(x,y)} & \text{if } d_{(x,y)} > \sigma_i \end{cases} \quad (9)$$

$D_{(x,y)}$  は髪色モデルとの距離で、 $d_{(x,y)}$  は髪の照明強度モデルとの距離で、 $\sigma_i$  は髪の照明モデルの分散値である

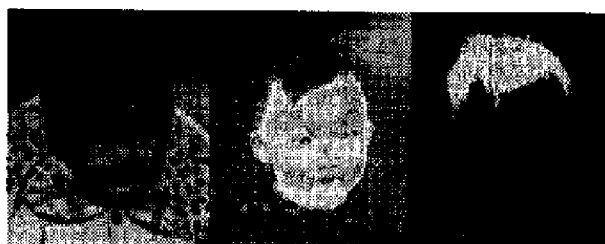


Figure 2. 色変換の例：(左) 入力画像、(中) 顔色変換画像、(右) 髪色変換画像

## 3. 運動情報の利用

### 3.1 運動情報の抽出

人間の動き情報は回りの静止物体と区別する有力な手がかりである。入力が時系列画像の場合、運動情報が顔の検出に利用することが出来る。式(10)に示すように、連続画像の差分によって画像にある運動物体を検出することが出来る。

$$M(x,y,t) = T(|I(x,y,t) - I(x,y,t-1)|, L) \quad (10)$$

$M(x,y,t)$  は差分画像の絶対値を閾値 L で二値化処理した結果であり、値がゼロでない画

素は運動画素と見なす。

顔色モデルと髪色モデルに基づいて変換した二つの画像の各画素に対して、運動情報による重みを付ける。変換画像の画素  $(x, y)$  に対して、運動画像の  $(x, y)$  画素と隣接の八画素を調べる。この九画素の中に運動画素があれば、変換画像の画素  $(x, y)$  に重み  $W_m$  を付ける。

$$W(x, y) = \begin{cases} W_m, & \sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} M(x+m, y+n, t) \neq 0 \\ 0, & \text{else} \end{cases} \quad (11)$$

$$v_{(x,y)} = v_{(x,y)} * (1 + W(x, y)) \quad (12)$$

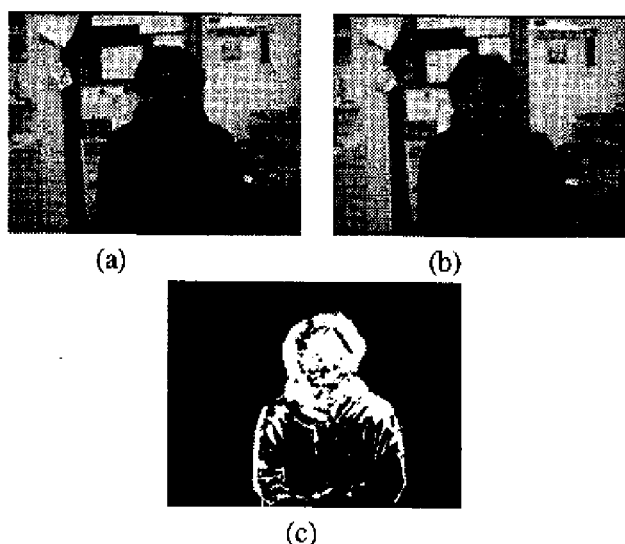


Figure 3. 運動画素の抽出

(a)と(b)は二つの連続画像、(c)は運動画像

### 3.2 顔画素の抽出

運動情報からの重みを付けた顔色変換画像に対して、閾値処理によって顔画素を抽出する。

$$R = \sqrt{\sigma_{rr}^2 + \sigma_{gg}^2} \quad (13)$$

$R$  は顔色の分布半径であり、式(6)に代入して閾値  $Thresh$  を算出する。

$$Thresh = 255 * (1 - \frac{R}{Max}) \quad (14)$$

運動体は頭が主であると仮定できれば、運動重み  $W_m$  を付けた  $Thresh * (1 + W_m)$  を新たな閾値にする。

閾値処理の結果を更に  $5*5$  又は  $7*7$  のモルフォロジーの開放演算を行い、顔画素をつなげたクラスタにする。

## 4. 幾何情報の利用

### 4.1 顔の形状情報

形状情報は顔を判定する要素の一つである。エッジ検出して楕円を当てはめる方法やモーメント分析などは顔の姿勢や顔型に影響されやすいので、安定的な形状判定法にはならない。ここでは、 $s_1$  と  $s_2$  という二つのパラメータを定義し、クラスタの形状を記述する。クラスタの外接四方形を求める (Figure 4(a))。その両辺長が  $W$  と  $H$  であり、クラスタの面積を  $A$  とすれば、

$$s_1 = A/(W * H), \quad (15)$$

$$s_2 = H/W, \quad (16)$$

顔を楕円で近似する時、楕円の辺長は  $a$  と  $b$  となる。顔が正面・垂直の姿勢をとる時、 $s_1$  は最

大値  $\pi b/4a$  になる。顔はこの姿勢を離れると、

$s_1$  の値が減少する。 $s_2$  は  $[b/a, a/b]$  の間の値で

ある。姿勢や髪型の変動を考慮し、 $\frac{a}{b} \leq 2a$  と

いう比を設定した。従って、クラス形状パラ

メータ  $s_1$  と  $s_2$  は以下の数値であれば、顔候補と

判断することにした。

$$\pi/8 < s_1 \leq 1 \quad (17)$$

$$0.5 < s_2 < 2 \quad (18)$$

### 4.2 顔と髪的位置関係

顔と髪の空間関係はクラスタが顔であるか否かを定めるもう一つの手がかりである。Figure 4(b)に示している様に、クラスタの外接四方形を  $ABCD$  とし、 $ABCD$  と重なる四方形  $EMNF$  に着目する。もしこのクラスタが顔領域であれば、四方形  $EMNF$  に髪画素要素が多くある。ここでは、髪色変換した画像を閾値処理せず、その値を直接利用する。四方形  $EMNF$  中のクラスタ以外の画素の値を集計し、平均値を計算する。計算した結果をサンプル画像集合から得られた数値と比較し、既知の値の範囲内であれば、四方

形EMNFは髪領域と判断し、四方形ABCDの中のクラスタは顔候補とする。

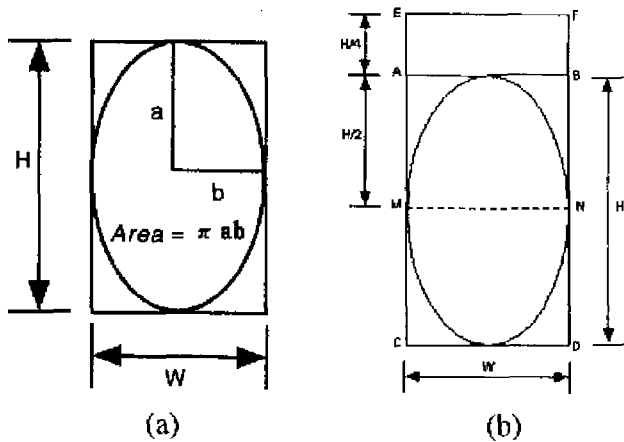


Figure 4. 幾何情報の説明図: (a) 顔形状パラメータ (b) 顔と髪の間関係

### 5. 顔特徴による検証

顔の特徴、例えば、目、鼻、口などは人間の顔を確定するには有力な手がかりである。検出された顔候補クラスタ領域から顔特徴を検出することができれば、検出結果の確信度が一層高まる。瞳孔は照明や表情などに影響されにくいので、より安定的に抽出できる。顔領域において、瞳孔が最も低い画素値を持つ。この性質を利用し、閾値処理によって瞳孔を抽出することが出来る。

まず、候補クラスタの領域で、入力画像を対象にモルフォロジーの膨張処理を行い、低い輝度値から反復閾値処理を行う。クラスタの形状にあう一定の範囲で、二つの小さい領域を検出できれば、その中心を瞳孔の位置と見なす。一定の反復処理の後でも、瞳孔を抽出することが出来なければ、このクラスタは顔でないとか、顔瞳孔が見えない位置であるとか、という結論を出すことが出来る。

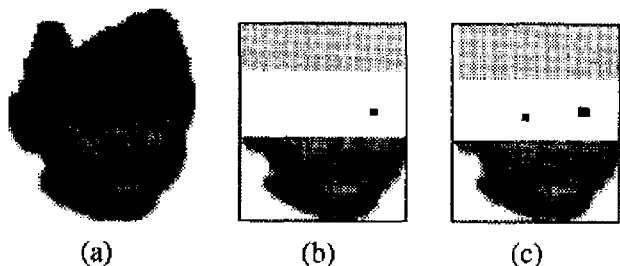


Figure 5. 反復閾値処理による瞳孔の検出 (a) Dilated input image (b) result when  $T = 25$  (c) result when  $T = 26$ .  $T$  は閾値である。

### 6. 複数情報の統合

色、運動、幾何特性、顔特徴など各々の情報で顔を確定することができるが、安定性と唯一性が保証できない。従って、これら複数の情報を統合することで真の顔領域を信頼性良く絞り出すことができると考えられる。

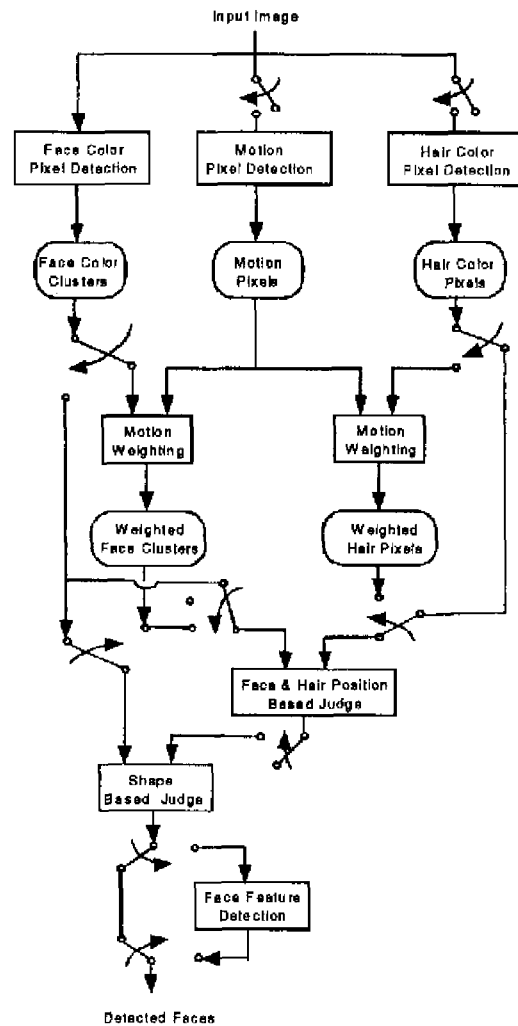


Figure 6. 複数情報の統合の流れ

Figure 6 は複数の情報を統合する一例を示している。各情報が各々のモジュールに組み込まれ、スイッチによって多種の統合方式が選択できる。顔色と顔形状情報はすべてのカラー画像に適応するため、基本モジュールとして使われる。各情報の使用可能かどうかの状況によって、Table 3.1 の五つの統合例が考えられる。

また、各情報を選択的に組み合わせることによって、どの情報が、どんな画像に対して、顔検出に一番役に立つかをしらべる事が出来る。

Table 1: 情報の統合例

Prototype	Cues in use
A	Face color, shape
B	A + hair
C	A + motion
D	B + motion
E	D + face feature verification

顔は非常に多様であるから、提案した四つの特徴情報でもカバーしきれないことがある。例えば、顔色について東南アジアの人を対象に限定していても、髪がないとか、カラーリングによる多種多様な髪色とか、というような例にも対応できない。また、頭が動かず、手だけ動くという場合、運動画素の利用も問題がある。従って、複数情報の統合による顔の検出法は一般的な状況において有効な手法であるが、以上のような例外の場合、統合の効果が落ちることがある。

ヒューマンコンピュータインタラクションなどの応用において、顔色情報と運動情報が最も基本的な役割を果たしている。例外的でないような場合、髪の情報有力な決め手である。瞳孔という顔の特徴は正面に近い顔に対して特に有効である。

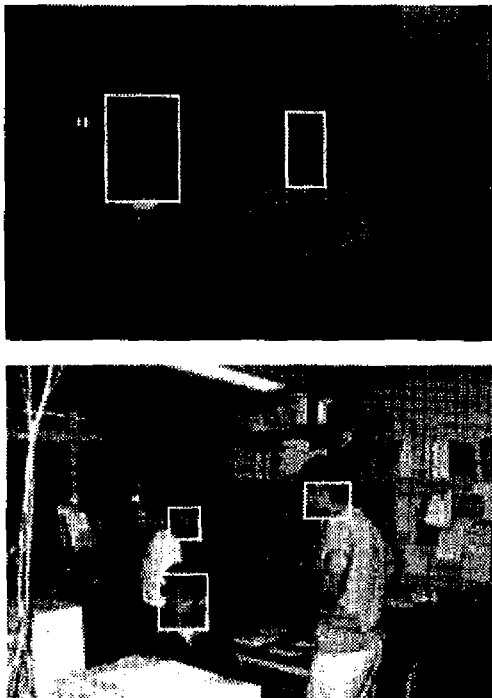


Figure 7. 顔検出の結果例

## 8. 終わりに

本文は画像から複数の情報の検出と統合による顔の検出法について検討した。単一の情

報を用いる場合よりも安定的かつ信頼性の高い検出結果ができる。

特例の場合、情報の統合をどうするか、照明状況が激しく変動する場合に、カラーモデルの自動更新はどうするか等々、今後の問題も少なくない。

## 参考文献

- 1) M. XU and T. Akatsuka: "Multi-Module Method for Detection of Human Face from Complex Background", SPIE Proceeding of Conference on Applications of Digital Image Processing, Vol. 3460, pp.793-802, 1998.
- 2) Kah-Kay Sung and Tomaso Poggio: "Example-Based Learning for View-Based Human Face Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No.1, pp.39-51, January 1998.
- 3) J. Yang and A. Wailbel, "A Real-Time Face Tracker", *Proceedings of WACV'96*, Sarasota, Florida, USA, 1996.
- 4) Haiyuan Wu, Qian Chen and Masahiko Yachida, "Face Detection from Color Images by Fuzzy Pattern Matching", *Transactions of IEICE* Vol. J80-D-II No.7 pp.1774-1785, July 1997.