

## 色情報と領域追跡情報を用いた人物の顔と手の領域の抽出

非会員 浅沼克紀 (大阪府立大学)  
学生員 大西正輝 (大阪府立大学)  
非会員 小島篤博 (大阪府立大学)  
正員 福永邦雄 (大阪府立大学)

Extracting Regions of Human Face and Hands  
considering Information of Color and Region Tracking

Katsunori Asanuma, Non-member, Masaki Onishi, Student Member, Atsuhiko Kojima, Non-member,  
Kunio Fukunaga, Member (Osaka Prefecture University)

In this paper, we propose a method of extracting and tracking regions of human face and hands appeared on sequential images considering fusion of color information and tracking region. Conventionally, it has been hard to extract the skin regions of human in the case when a background reflects skin color objects like corrugated cardboards, or including moving picture such as TV screen. Coping with such kind of difficult situation to extract skin regions, we firstly obtain three kinds of information based on background color, skin color and its tracking skin regions, then apply Dempster-Shafer's combination rule to the three kinds of information. This approach using the fused information becomes a robust extraction method, because other information can extract the regions even if one of the information lacks. The effectiveness of our approach is examined by the experiments.

キーワード：動画像, Dempster-Shafer の結合規則, 背景, 肌, 追跡

## 1. まえがき

近年, VR やマルチメディア・エンタテインメント分野のインタフェース, さらには人間とロボットの間のコミュニケーション支援を目指して, 行動やジェスチャを認識する研究が盛んに行なわれている<sup>(1)-(3)</sup>。その中でも, 人物の手の動きや顔に対する手の相対的な動きに注目したのも数多く報告されており, 動画像から人物の手や顔の領域を抽出する重要性が高まってきている。

これまでも, 動画像において人物の顔や手の領域を抽出する研究は数多くなされており, その成果が報告されている。まず顔や手の色情報の分布に着目したものの<sup>(4)</sup>として, あらかじめ人物の肌色領域を画像から抽出し, その色情報の分布を調べておき, 画像中から肌色である領域を抽出する方法がある。しかし, この方法では背景領域に壁やダンボールなどの肌色領域が存在する場合にはその領域も抽出されてしまうという問題点があった。

そこで, これらの問題点を解決するために, 人物領域全体を抽出しておく方法が考えられる。まず背景情報を用いる方法では, あらかじめ背景画像を撮影しておき, 動物体

が現れた画像から背景画像の差分を求めることで動物体領域を調べる方法<sup>(5)-(7)</sup>が考えられる。しかし, 光の強度の変化や局所的な輝度の変化が起こると誤った領域検出を行なうことになる。そこで, この問題に対処する方法として背景画像を一定時間撮影しておき, 各画素の光の強度や色情報の揺らぎを統計的にとらえることにより, 各画素ごとに背景か否かを調べていく方法が提案されている<sup>(8)-(10)</sup>。

また, 動物体の領域を追跡していくことにより領域を抽出する方法<sup>(11)</sup>, さらには, 赤外線センサーなど人体の熱線に感応する特殊な入力装置を用いることにより, 人物領域を抽出する方法<sup>(12)</sup>なども提案されている。しかし現在のところ, どのような環境に対しても動物体領域, さらには肌領域や髪領域など人物の特定の領域を確実に抽出するロバストな方法を構成することは難しいのが現状である。しかし, 最近の動作認識や動物体認識など動画像処理の研究が活発になるに従い, その前処理としての動物体領域の抽出の精度が, 後に続く動画像処理結果に大きく影響を与えることから, 改めて多くの検討がなされている。

そこで本論文では, 色情報と領域追跡情報を用いた動画像上の人物の顔や手などの肌色領域を抽出する新たな方法

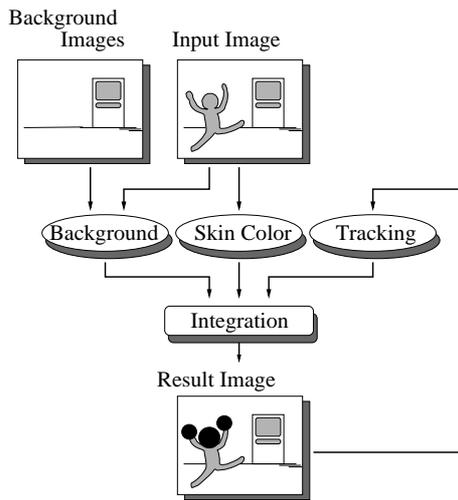


図1 処理の概要

Fig. 1. A flow of process.

を提案する。図1に示すように、まず背景を一定時間撮影した画像から、各画素についてその色の分布を調べ、これをもとに人間など動物体を含む画像(入力画像)の各画素が背景領域である度合を求める。次に、一般的な人間の肌色分布から、入力画像の各画素が人物の肌色領域である度合を求める。さらに顔や手など一まとまりの肌領域の位置の変化とその画像上での分布を追跡するとともに、現在の入力画像上での肌領域分布を予測し、これをもとに各画素が顔や手の領域と考えられる度合を求める。

これら全ての度合を基本確率の形で表し、Dempster-Shaferの結合則で統合することにより、一部の情報が欠落した場合においても人物の顔や手の領域をロバストに抽出し、追跡できる方法を明らかにしている。

## 2. 背景情報・肌情報と領域情報

人物の顔と手の領域を抽出し、追跡を行なうための情報として、本論文では色による背景情報と肌情報および追跡による肌領域情報の3つを考え、これら一つ一つの情報をBayesの定理をもとに求める。

ここでは色に基づく情報を抽出するが、一般に画像処理でよく用いられるRGB表色系を線形変換したCIE1976UCS表色系を用いる。一般にRGB表色系は明るさの変化に大きく影響を受けること、また色相の変化を人間の感覚に近い形で取り扱うことに難点があるとされている。そこで、本論文では明るさの変化を吸収でき、色相の変化を人間の感覚に近い形で扱うことができるCIE1976UCS表色系の $u, v$ を用いる。

**2.1 色による背景情報の抽出** ここでは、人物領域を確定するための情報の一つとして背景の色の分布に基づき背景情報を確率として求める方法について述べる。一般に、背景の色は光源などの影響により時間とともに変動している。そこで、本手法では背景画像の各画素について時

間変動による色の变化を確率分布としてとらえ、これをもとに各々の画素について背景の度合を定義する。

本手法では、背景での色の変動が色空間上では多変数正規分布になると仮定し、背景の色の分布を調べる。そこでまず、背景画像を一定時間( $T_B$ 枚)撮影する。ここで画素 $x = (x, y)$ に注目し、この画素の各時点 $t_i (i = 1, 2, \dots, T_B)$ での色 $C_{xi} = (u_{xi}, v_{xi})$ を調べ、画素 $x$ の背景色の平均値を $\mu_x$ とし、その共分散行列を $\Sigma_x$ とする。つまり、

$$\mu_x = \frac{1}{T_B} \sum_{i=1}^{T_B} C_{xi} \dots\dots\dots (1)$$

$$\Sigma_x = \frac{1}{T_B - 1} \sum_{i=1}^{T_B} (C_{xi} - \mu_x)^T (C_{xi} - \mu_x) \quad (2)$$

とする。これは、ある画素 $x$ に注目すると、その色の時間的な変動の分布の中心は $\mu_x$ であり、その共分散行列は $\Sigma_x$ の形で与えられることを示している。そこで、確率密度関数 $p_B(C_x|\omega_B)$ は背景クラス $\omega_B$ に属する画素 $x$ の色 $C_x$ が生起する確率を表し、次の2次元正規分布で与える。

$$p_B(C_x|\omega_B) = \frac{1}{2\pi|\Sigma_x|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(C_x - \mu_x) \Sigma_x^{-1} (C_x - \mu_x)^T\right\} \dots\dots\dots (3)$$

以後、このような2次元正規分布を次式で表す。

$$p_B(C_x|\omega_B) = N_2(C_x, \mu_x, \Sigma_x) \dots\dots\dots (4)$$

次に、色の生起確率 $p(C_x)$ を求める。画像中に現れる色は、全ての色についてその生起確率が均等ではないと考えられる。そこで本手法では、一般的と考えられる複数の画像から全画素に対する色をCIE1976UCS表色系の $u-v$ 色度図上に投影し、そのヒストグラムを調べ、そのヒストグラムを調べた全画素数で割ることにより色の生起確率 $p(C_x)$ を求める。

そして、画素 $x$ における色情報 $C_x$ の背景への帰属度として、Bayesの定理により次式を用いて確率を求める。

$$P_B(\omega_B|C_x) = \frac{p_B(C_x|\omega_B) P(\omega_B)}{p(C_x)} \dots\dots\dots (5)$$

ただし、背景クラス $\omega_B$ の事前確率 $P(\omega_B)$ は既知とし、あらかじめ与えるものとする。

**2.2 色による肌情報の抽出** ここでは顔と手の領域の抽出が目的であるため、その大きな特徴である肌色に注目する。人物の肌色は同じ人種であればその色に大きな差はなく、 $u-v$ 色度平面上でも閉じた狭い領域に分布している。そこで、本手法では同一人種の肌色サンプルからその色の分布を調べ、これをもとに色による肌確率を定義する。

まず、同一人種の肌色の分布を多変数正規分布と仮定し、その色の分布を調べる。一般的な画像から人物の顔や手などの肌色領域の画素を $S_n$ 個選ぶ。次に、 $S_n$ 個の色 $C_i = (u_i, v_i) (i = 1, 2, \dots, S_n)$ に対し、平均値 $\mu_S$ 、その共分散行列 $\Sigma_S$ を求める。つまり、

$$\mu_S = \frac{1}{S_n} \sum_{i=1}^{S_n} C_i \dots\dots\dots (6)$$

$$\Sigma_S = \frac{1}{S_n - 1} \sum_{i=1}^{S_n} (C_i - \mu_S)^T (C_i - \mu_S) \dots\dots (7)$$

とする。これは、人物の肌色は  $u-v$  色度平面上において中心が  $\mu_S$ 、共分散行列が  $\Sigma_S$  で分布していることを示しており、中心の色が最も肌らしいことを表している。

したがって、画素  $x$  の肌クラス  $\omega_S$  における確率密度関数  $p_S(C_x|\omega_S)$  は、先と同様に 2 次元正規分布と仮定する。

$$p_S(C_x|\omega_S) = N_2(C_x, \mu_S, \Sigma_S) \dots\dots\dots (8)$$

事前確率  $P(\omega_S)$  を既知として 2.1 と同様、肌色への帰属度として確率  $P_S(\omega_S|C_x)$  を次式で求める。

$$P_S(\omega_S|C_x) = \frac{p_S(C_x|\omega_S)}{p(C_x)} P(\omega_S) \dots\dots\dots (9)$$

これは、画素位置  $x$  における色  $C_x$  が肌クラス  $\omega_S$  に属する確率を表している。

2.3 領域の追跡による肌情報の抽出 次に領域追跡情報として、過去のフレームにおける肌領域をもとに現在のフレームの肌領域の中心と広がり予測することで追跡による肌確率を求める。ここでは、まず前フレームまでに予測された肌領域の中心と広がりから、予測追跡により肌領域に属する割合を求める方法について述べ、次にその中心を予測する方法について説明する。

時刻  $t$  における画面上の人物の肌領域  $R_T$  での分布を考える。今、この領域の分布の中心が画像上の座標  $\mu_T = (x_T, y_T)$ 、その広がりが共分散行列  $\Sigma_T$  と予測できるとする。このとき、位置  $x = (x, y)$  の肌クラス  $\omega_S$  における確率密度関数  $p_T(x|\omega_S)$  は次式で表される。

$$p_T(x|\omega_S) = N_2(x, \mu_T, \Sigma_T) \dots\dots\dots (10)$$

このとき、画素  $x = (x, y)$  が肌領域に属する確率は Bayes の定理により、次式で与える。

$$P_T(\omega_S|x) = \frac{p_T(x|\omega_S)}{p(x)} P(\omega_S) \dots\dots\dots (11)$$

ただし、画素の生起確率  $p(x)$  およびクラスの事前確率  $P(\omega_S)$  は既知とし、 $P(\omega_S)$  の値は式 (9) の事前確率と同じ値とする。

次に、分布の中心を追跡するアルゴリズムについて述べる。時刻  $t$  における分布の中心座標を  $\mu_T^{<t>}$  とする。また、移動する領域の運動状態を表現するために、時刻  $t$  から時刻  $t+1$  までの間における分布の中心の速度を  $v^{<t>}$  とし、さらに、中心の加速度を  $a^{<t>}$  とする。

まず、時刻  $t=0$  では速度、加速度共に 0 とする。そして、時刻  $t-1$  における中心の位置  $\mu_T^{<t-1>}$  と、中心の速度  $v^{<t-1>}$  を用いて、時刻  $t$  における中心の予測値  $\hat{\mu}_T^{<t>}$  を次式により求める。

$$\hat{\mu}_T^{<t>} = \mu_T^{<t-1>} + v^{<t-1>} \cdot \Delta t \dots\dots\dots (12)$$

ここで、 $\Delta t$  はフレーム間隔を表す。この予測値  $\hat{\mu}_T^{<t>}$  を時刻  $t$  における中心の位置  $\mu_T^{<t>}$  として用いる。

ここで、時刻  $t-1$  における分布の中心  $\mu_T^{<t-1>}$  および中心の速度  $v^{<t-1>}$  から、加速度  $a^{<t-1>} = 0$  を仮定して時刻  $t$  における中心の位置を予測している。このとき、時刻  $t-1$  から  $t$  までの間、加速度が生じていなければ予測値  $\hat{\mu}_T^{<t>}$  と実際の中心  $\mu_T^{<t>}$  の差は 0 である。ところが加速度が生じた場合、これらの値は異なる値を示す。したがって、時刻  $t-1$  から  $t$  までの間、予測値からの修正量

$$\Delta\mu_T^{<t>} = \mu_T^{<t>} - \hat{\mu}_T^{<t>} \dots\dots\dots (13)$$

に相当する位置の変化を与える加速度

$$a^{<t-1>} = \frac{2}{(\Delta t)^2} \cdot \Delta\mu_T^{<t>} \dots\dots\dots (14)$$

が生じていたものと考えられる。

このとき、時刻  $t$  における中心の速度  $v^{<t>}$  は、時刻  $t-1$  における速度  $v^{<t-1>}$  および式 (14) により求めた加速度  $a^{<t-1>}$  を用いて、次式のように求められる。

$$v^{<t>} = v^{<t-1>} + a^{<t-1>} \cdot \Delta t \dots\dots\dots (15)$$

このようにして、時刻  $t-1$  からの予測値をもとに時刻  $t$  における肌領域の抽出結果から、予測値と実際の値との差により速度および加速度を修正することで追跡が可能となる。また、追跡している領域が抽出されなかった場合は、その領域が他の領域と重なっているか、あるいは人物の後ろに隠れていると考え、その間に、追跡している領域は等速直線運動していると仮定して予測値を求めている。また、前フレームとの対応は現在得られている肌領域との中心間の距離により対応付けする。一方、分布の広がりを表す共分散行列には前フレームの共分散行列を用いることにする。

### 3. 抽出された情報の統合

本手法では、人物における肌色領域の抽出が目的である。しかし、一般に背景にも人物以外の肌色に類似した領域が存在し、人物のみの肌色領域を抽出するのは困難である。これは、人物の肌は色による肌情報のみで抽出することが困難であり、他の判定条件を導入する必要があることを示している。このような場合に対し、Dempster-Shafer 理論 (以後、DS 理論) では複数の情報源から得られた情報を統合することで、一部の情報では判定できないとき他の情報により補完することができる。そこで、本手法では 2. で得られた 3 つの異なる情報を統合し、動画像から人物の肌領域を抽出するために必要な情報を抽出する方法について述べる。

ここでは、まず Bayes の統計的決定法と Dempster-Shafer の結合規則を用いた基本確率の統合方法について述べ、統合による肌領域の抽出方法について明らかにする。

3.1 情報の統合 クラス  $\omega_1, \omega_2$  の確率密度関数  $p(x|\omega_1), p(x|\omega_2)$  と事前確率  $P(\omega_1), P(\omega_2)$  および、クラスによらない  $x$  の生起確率  $p(x)$  が与えられているとする。Bayes の統計的決定理論に基づくパターン分類では観測された未知パターン  $x_0$  を次式で示す事後確率の大小によって識別する。

$$P(\omega_1|x_0) = \frac{p(x_0|\omega_1)}{p(x_0)} P(\omega_1) \dots\dots\dots (16)$$

$$P(\omega_2|x_0) = \frac{p(x_0|\omega_2)}{p(x_0)} P(\omega_2) \dots\dots\dots (17)$$

しかし、この分類法は誤分類を認めているにも関わらず、あいまいな分類ができない。これは、Bayes 理論では無知量を表すことができないためである。つまり、 $\bar{\omega}_i$  を  $\omega_i$  の補集合として

$$P(\omega_i|x) + P(\bar{\omega}_i|x) = 1 \dots\dots\dots (18)$$

の関係が要求されるため、信用の欠如と不信用とを区別することができない。このため、信用の欠如の多い環境下での識別は困難である。

一方、DS 理論では基本確率をクラスの集合に割り当てることができるため信用の欠如を取り扱うことができる<sup>(13)</sup>。つまり、“ $\omega_1$  か  $\omega_2$  か判断できない”といった部分集合  $\{\omega_1, \omega_2\}$  に基本確率を割り当てることが可能である。

$\Theta$  個の有限な全集合を  $\omega_0$  , その部分集合を  $\omega_i$  ( $i = 1, 2, \dots, 2^\Theta - 1$ ) とすると、基本確率  $m(\omega_i)$  ( $i = 0, 1, \dots, 2^\Theta - 1$ ) は  $[0, 1]$  の値をとり、次の条件を満たす。

$$\begin{cases} m(\phi) & = 0 \\ \sum_{\omega_i \in \omega_0} m(\omega_i) & = 1 \end{cases} \dots\dots\dots (19)$$

ただし、 $\phi$  は空集合を表し、 $m(\omega_i) > 0$  のとき  $\omega_i$  を焦点要素と呼ぶ。

このように DS 理論では、信用の欠如をあいまいな確率として扱うことができるため、これらを用いたパターン分類、画像認識がいくつか報告されている<sup>(14)(15)</sup>。

また、DS 理論における重要な法則に Dempster-Shafer の結合規則がある。これは、独立した情報源から得られた基本確率を統合する手法を与えており、独立した情報源  $\alpha, \beta$  から得られた基本確率  $m^\alpha, m^\beta$  から次式を用いて新しい基本確率  $m$  を求める。

$$m(\omega_k) = \frac{\sum_{\omega_i \cap \omega_j = \omega_k} m^\alpha(\omega_i) \cdot m^\beta(\omega_j)}{1 - \sum_{\omega_i \cap \omega_j = \phi} m^\alpha(\omega_i) \cdot m^\beta(\omega_j)} \dots\dots\dots (20)$$

ここでは、積集合が空集合となる場合が起こり得るため、空集合となる場合を除外し分母で正規化を行なっているが、空集合に割り当てられる確率が大きくなる場合などには、

不都合が生じる場合があることも指摘されている<sup>(16)</sup>。

そこで DS 理論を領域抽出と追跡に適用するため、まず Bayes 理論で問題となった式 (18) を文献<sup>(16)</sup> の包含関係をもとに定義する。以後、基本確率を  $P$  と表記する。

$$P(\omega_i|x) + P(\omega_0|x) = 1 \dots\dots\dots (21)$$

ここでは、得られた情報以外の確率を全集合  $\omega_0$  に対する基本確率に割り当てている。

また複数の情報源から、算出された基本確率は次式を用いて統合する。

$$P(\omega_k) = \sum_{\omega_i \cap \omega_j = \omega_k} P^\alpha(\omega_i) \cdot P^\beta(\omega_j) \dots\dots\dots (22)$$

これは、確率の正規化を行わず、積集合が空集合に割り当てられた値を無視する統合法<sup>(17)</sup> であり、このような割り振りによって曖昧な認識結果となる場合においても、複数の情報を統合して認識を行なうことで、より正しい認識結果を得ることができると考えられる。

3.2 情報の統合による肌領域抽出 ここでは、2. で得られた各情報を Dempster-Shafer の結合規則により統合し、人物の肌領域を抽出する方法について述べる。

まず、色による背景情報と肌情報の 2 つを統合する方法について述べる。ある画素  $x$  について、色による背景確率  $P_B(\omega_B|Cx)$  および色による肌確率  $P_S(\omega_S|Cx)$  が式 (5), (9) により与えられたとき、これらの確率に対して式 (21) より以下の関係式が得られる。

$$P_B(\omega_B|Cx) + P_B(\omega_0|Cx) = 1 \dots\dots\dots (23)$$

$$P_S(\omega_S|Cx) + P_S(\omega_0|Cx) = 1 \dots\dots\dots (24)$$

ここで、クラス  $\omega_0$  はどのクラスに属しているか分からない場合を表しており、背景、肌色、服、髪などの領域の内、どの領域に属しているかが分からない場合を表している。つまり、式 (23), 式 (24) は明確に与えられる確率以外の部分をあいまいな確率として残しておくことを示している。

これらの確率の統合の概念図を図 2 に示す。ここで、図に示すように、本手法では背景かつ肌のような矛盾する領域については無視している。そして、背景情報により分からないとされ、肌情報により肌領域とされている部分を人物の肌とする。このとき、式 (22) により人物の肌領域であるクラス  $\omega_S$  に  $Cx$  が属する確率は、次式で表される。

$$P(\omega_S|Cx) = P_B(\omega_0|Cx) \cdot P_S(\omega_S|Cx) \dots\dots (25)$$

これは、図 2 の肌となっている面積を確率として取り出したものとなっている。

以上は、2 つの情報の統合を考えたが、この考え方をもとに 3 つの情報 (背景・肌色・追跡) を統合し、人物の肌領域である確率  $P(\omega_S|Cx)$  を求めると次式のようになる。

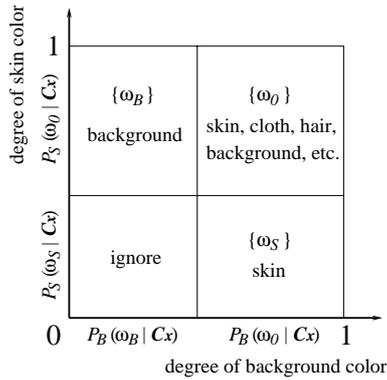


図2 背景確率と肌確率の統合

Fig.2. An integration of background and skin probabilities.

$$\begin{aligned}
 P(\omega_S | Cx) &= P_T(\omega_S | x) \cdot \{P_S(\omega_S | Cx) \\
 &\quad + P_B(\omega_0 | Cx) \cdot P_S(\omega_0 | Cx)\} \\
 &\quad + P_T(\omega_0 | x) \cdot P_S(\omega_S | Cx) \dots\dots (26)
 \end{aligned}$$

ただし、確率  $P_T(\omega_S | x)$  は式 (11) の領域の追跡確率とし、式 (21) より次式が成り立つものとする。

$$P_T(\omega_S | x) + P_T(\omega_0 | x) = 1 \dots\dots\dots (27)$$

このように 3 つの情報を統合することで、ある情報では分からないとされている部分を他の情報をもとに積極的に該当する領域であると判定することにより抽出することができ、各情報をより有効に扱うことが可能となる。

最後に、各画素  $x$  について、人物の肌領域である確率  $P(\omega_S | Cx)$  がある閾値  $\theta$  以上であれば、その画素を肌領域とし、 $\theta$  未満であれば肌領域ではないとする。即ち、

$$x \begin{cases} \in \omega_S & ; \text{ if } P(\omega_S | Cx) \geq \theta \\ \notin \omega_S & ; \text{ others} \end{cases} \dots (28)$$

なお本手法では、3 つの情報を統合して得られた肌領域に対し雑音の影響を考慮して孤立点除去を行っている。その後、石山らのラベリングアルゴリズム<sup>(18)</sup>により一つの肌領域を形成している。さらに、形成された肌領域の面積が大きいものを残し、それ以外の肌領域や面積が一定以下のものは除去している。

#### 4. 実験および考察

本手法の有効性を検証するために実環境において実験を行なった。まず、背景の動的変化や背景に存在する肌色などといった人物の肌領域の抽出が困難であると考えられる環境に対し、本手法と他手法で肌領域を抽出し、その比較を行なった。次に、上記の抽出結果と追跡情報をもとに顔と手の軌跡を抽出した。また、これらの実験結果を検討・考察した。

ここでは、Pentium 266MHz×2 により動作するシステムを構築し、解像度 160×120 における処理速度は約 6.6 フ

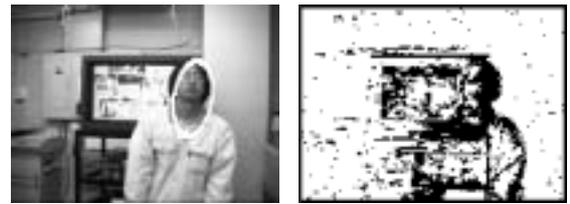


図3 テレビのある背景での抽出結果と背景確率画像

Fig.3. An extracting result and its background probabilities image in background with TV screen.



図4 ダンボールのある背景での抽出結果と肌確率画像

Fig.4. An extracting result and its skin probabilities image in background with corrugated cardboards.

レーム/秒であった。また、人物の服装は肌色以外の服装で顔と手以外は肌部分がでないものとして、室内環境で実験を行なった。実験での各パラメータは  $\theta = 0.5$ ,  $P(\omega_B) = 0.5$ ,  $P(\omega_S) = 0.1$ ,  $p(x) = 3/(160 \times 120)$  とし、背景画像は 50 フレーム撮影し、肌色サンプルは 10 人から採取した。

4.1 抽出結果 従来の手法では、テレビ画面などのように動的に変化するものを含む場合や背景に肌色と類似する色を含む場合は人物の肌領域を抽出することは困難であるとされている。これは、図 3 で示すように背景が動的に変化する場合 (テレビが置かれた環境など)、背景から得られる情報が減少し、図 3 の右の画像のように背景確率が不安定となるためである。また同様に、図 4 のように背景に肌色と類似する物体 (ダンボールなど) が置かれた場合においても、色による肌情報が減少し、図 4 の右の画像のように肌確率が不安定となる。そこで、次のような 6 つのシーンを設定し、認識実験を行なうとともに他手法との比較を行なった。

- (1) 背景にダンボールやテレビを含まない場合
- (2) 背景にテレビを含む場合
- (3) 背景にダンボールを含む場合
- (4) (1) で速く動く場合
- (5) 顔や手が交差する場合
- (6) (2) (3) (5) を含む場合

シーン (1) は情報の欠落がない場合であり、シーン (2)(3)(4) はそれぞれ色による背景情報、色による肌情報、追跡による肌情報が欠落した場合を考慮して設定した。シーン (5) はジェスチャ認識には欠かせないシーンであり、シー

表 1 各シーンに対する抽出率

Table 1. An extracting rate for each scenes.

Scene	Proposed Method			Bayes		
	EoE	EoA	Rate	EoE	EoA	Rate
(1)	1	0	99.8	9	0	98.2
(2)	23	0	95.4	47	0	90.6
(3)	28	0	94.4	68	1	86.2
(4)	18	4	95.6	107	9	76.8
(5)	28	9	92.6	103	11	77.2
(6)	54	1	89.0	116	4	76.0

表 2 誤抽出の内容

Table 2. Details of error extracting.

Scene	Proposed Method				Bayes			
	EoE	UE	TE	EE	EoE	UE	TE	EE
(1)	1	0	0	1	9	9	0	0
(2)	23	12	3	8	47	31	10	6
(3)	28	9	19	0	68	28	37	3
(4)	18	6	4	8	107	107	0	0
(5)	28	23	1	4	103	63	30	7
(6)	54	24	16	14	116	43	28	46

ン (6) はこれら全ての場合を含むシーンである。

また、本手法で用いた式 (21) の有効性を考えるために、式 (18) を用いて、背景でない確率  $P_B(\bar{\omega}_B|C_x)$  と色により肌である確率  $P_S(\omega_S|C_x)$ 、追跡により肌である確率  $P_T(\omega_S|x)$  の積をとったものを Bayes 手法とし、本手法と比較を行なった。ただし、追跡による肌確率は一度肌領域が抽出されないと 0 のままであるため、Bayes 手法では領域が現れない場合には常に確率が 0 となってしまう。そのため、Bayes 手法では領域が 3 つより少ない場合は全画素に一定の確率を加えるようにした。

評価方法として、各シーンに対し 500 フレームの映像から肌領域を抽出し、フレーム毎に正確に抽出できているかを目視により判定した。これらの実験結果を表 1 および 2 に示す。表 1 に、各シーンに対する誤抽出 (EoE)、誤対応 (EoA) のフレーム数と正確に抽出した割合 (Rate(%)) を示す。また、表 2 は誤抽出の内容として人物の肌領域を抽出できなかった未抽出 (UE)、余分に抽出した過抽出 (TE)、またこれ以外で人物の肌領域ではない領域を抽出した場合 (EE) に分け、そのフレーム数を示している。

まず、シーン (1) は情報の欠落がないため本手法、Bayes 手法とも高い抽出率が得られた。次に、シーン (2), (3), (4) はそれぞれ各情報の欠落がある場合であり、本手法ではシーン (1) に対し誤抽出が若干増えているが全体として高い抽出率を得ている。また、Bayes 手法と比較しても本手法は非常に安定して肌領域を抽出していることが分かり、情報の欠落に対し他の情報で補うことによりロバストに肌領域を抽出できることを確認した。

シーン (5) の顔や手が交差する場合では、未抽出は肌領域が少し速く動いた場合によるものである。また、Bayes 手法では領域どうしが重なった時に他の領域を抽出しようとして過抽出が多くなっている。シーン中、顔や手の交差

回数は 26 回であったがその内どちらの手法でも約 1/3 が追跡に失敗しており、この点については他の追跡アルゴリズムを導入する必要がある。シーン (6) は、交差回数 8 回を含み色による背景情報、肌情報が欠落している場合であるが、Bayes 手法の 76% に対し、本手法では約 90% の抽出率であり、本手法の方がより確実に抽出できており、有効性を確認することができた。

これらの結果から、人物の肌領域の抽出が困難と考えられる様々な環境において、本手法ではほぼ 90% 以上の抽出率であり、多様な環境に対してロバストに人物の肌領域を抽出できることを確認した。

4.2 人物の顔と手の追跡 本手法では、人物の肌領域の抽出と同時に追跡も行なっている。ここでは、先のシーン (1) において「バイバイ」というジェスチャをした場合とシーン (5) において手が交差した場合についての軌跡を調べた。その結果を図 5 および図 6 に示す。

まず「バイバイ」のジェスチャの様子を図 5 (a) に示す。特に整備されていない背景の環境であるにも関わらず、正確に抽出できている。また、図 5 (b) で示すように「バイバイ」の軌跡も抽出できていることが分かる。

次に、シーン (6) において手が交差する場合の軌跡を抽出した。その様子を図 6 (a) に示している。この場合も人物の肌領域が正確に抽出できている。また、手が交差している間は二つの手が一つの肌領域として抽出されているが、この領域はどちらか一方の手と見なされ、もう一方の手は追跡情報により現在の位置を予測している。そのため、図 6 (b) のように手が交差していてもその軌跡を抽出できていることが分かる。

これらの結果から、様々なジェスチャに対しても軌跡を得ることができ、これらの軌跡からジェスチャ認識の前処理として有効であることが期待できる。また、シーン (5) のように肌色と類似した色を持つダンボールやビデオが再生されているテレビが置かれている環境においても手が交差する場合の軌跡が正しく得られた。さらに追跡情報により数フレーム間、手が交差している場合や手が隠れた場合などもその軌跡が予測できることを確認している。したがって「バイバイ」のような簡単なジェスチャに加え、手が交差するようなジェスチャや歩くといった動作などの認識に利用することができる。

4.3 考察 様々なシーンにおける人物の肌領域の抽出結果とその軌跡を示し、本手法の有効性を確認した。ここでは特に抽出が難しかった例をもとに本手法について考察する。また、本手法の応用についても検討する。

4.3.1 色情報の誤抽出に対する考察 まず、シーン (2) においてテレビ領域中に顔があるときに顔領域が図 7 (a) のように分裂することがあった。これは、背景撮影時にテレビ画面が動的な変化をしており、テレビ領域での色の共分散行列が大きくなり、式 (5) により得られる背景確率が全体的に小さくなるため、テレビ領域での背景確率が不安定となる。これにより、背景情報以外の情報が減少し

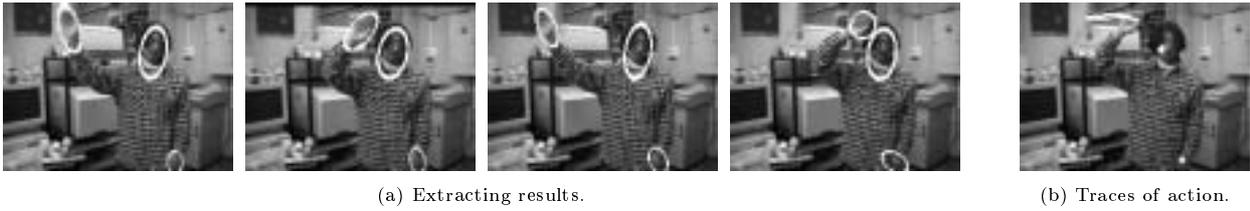


図5 「バイバイ」の抽出結果と軌跡

Fig. 5. Extracting results and traces of action "bye".



図6 手が交差する場合の抽出結果と軌跡

Fig. 6. Extracting results and traces of crossing hands.



図7 誤抽出の例

Fig. 7. An example of EoE.

た場合に顔の分裂などが起こると考えられる。

また、シーン (3) においてダンボールを数フレームの間、誤抽出することがあった。図 7 (b) がその様子である。これは、図 4 でも分かるようにダンボールは人物の肌色と非常によく類似しており、色による肌確率は高くなる。そのため、一度追跡領域がダンボールをとらえるとそのままダンボールを追跡しようとすることがある。また、顔や手の領域がダンボールの近傍にあるときに一つの領域になることがあった。

これらの誤抽出に対してはともに、色や追跡による肌情報をさらに精度のよいものにすること、また得られる情報源を増やすことによって回避できると考えられる。

4.3.2 追跡方法に対する考察 シーン (5) では、特に両手が交差する場合に、その前後で手の対応をとれないことがあった。本手法では、追跡中の領域が他の領域と重なった場合には等速直線運動すると仮定している。しかし、

交差が開始するフレームで両手が近い場合には、重なるまでのフレーム数が少ないため、重なっている最中の移動速度を十分に予測することが困難であり、移動速度を実際より遅く予測してしまうことがある。そのため、交差後の移動距離が十分に得られず、対応が入れ替わってしまったと考えられる。これに対しては、追跡方法を交差を考慮に入れたものにするか、処理速度をあげることで回避できると考えている。

また今回用いた追跡情報は、速度と加速度から次の位置を予測するものであり、十分な予測結果が得られたとは言えない。処理速度を上げることで精度が上がると考えられるが、同時にカルマンフィルタを用いた追跡アルゴリズム<sup>(19)</sup>などを用いることも考えられる。

4.3.3 服装が肌色の場合 本手法では、人物領域中の肌色部分を顔あるいは手の領域と仮定しており、服装が肌色である場合に図 7 (c) のような誤抽出が起こり得る。これは、本手法で抽出した人物の顔と手に関する情報が色情報に基づいているためである。しかし、本手法では複数の情報源から得られた情報を統合し、認識する枠組を提案しており、例えば人物領域の形状から構造的に人物の顔と手の領域を判断する<sup>(20)</sup>などといった新たな情報を統合することでこのような誤抽出を回避できると考えている。

4.3.4 応用に関する検討 本研究室では、人物をとらえた動画からその行動・状態を頭部の位置・姿勢の変化に基づいて推論し、テキストで出力を行なうシステムを提案している<sup>(21)</sup>が、手の軌跡を用いることでより高度な行動・状態を推論することが可能となる。

また、人物の顔や手の動きからジェスチャを認識する研究も行なわれており<sup>(22)</sup>、提案手法を用いることで、より安定したジェスチャ認識が行なえると考えている。

## 5. あとがき

本論文では、人物をとらえた動画から色による背景情報、色による肌情報、追跡による肌情報を抽出し、これらを統合することで人物の顔と手の領域を抽出し、追跡する手法を提案した。これにより、従来の問題点であった背景部分の肌色領域や動的に変化する領域に対してロバストに顔と手の領域を得ることができた。今後は、本手法により得られた顔と手の軌跡の変化からジェスチャ認識することを考えている。

謝辞 本研究を進めるに当たり、情報の統合方法について御討論頂いた本研究室の奥井英樹君(当時)、また筆者の一人が日頃御指導頂いている本学総合情報センターの田村武志教授に感謝致します。

(平成 11 年 2 月 26 日受付, 平成 11 年 7 月 12 日再受付)

## 文 献

- (1) Matthew B., Nuria O. and Alex P., "Coupled hidden Markov models for complex action recognition," Proc. of Computer Vision and Pattern Recognition, pp.994-999, Jun. 1997.
- (2) Ming-Hsuan Y. and Narendra A., "Extraction and Classification of Visual Motion Patterns for Hand Gesture Recognition," Proc. of Computer Vision and Pattern Recognition, pp.892-897, Jun. 1998.
- (3) Andrew D. W. and Aaron F. B., "Recognition and Interpretation of Parametric Gesture," Proc. of Inter. Conf. in Computer Vision, pp.329-336, Jan. 1998.
- (4) 呉海元, 陳謙, 谷内田正彦, "ファジィパターン照合を用いた色彩画像からの顔検出システム," 信学論(D-II), vol.J80-D-II, no.7, pp.1774-1784, Jul. 1997.
- (5) 浅田稔, 谷内正彦, 辻三郎, "運動物体の検出と追跡," 信学論(D), J62-D, No.6, pp.395-402, Jun. 1979.
- (6) 川端敦, 谷藤真也, 諸岡泰男, "移動物体像の抽出技術", 情処論, Vol.28, No.4, Apr. 1987.
- (7) C. R. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp.780-785, Jul. 1997.
- (8) 中井宏章, "事後確率を用いた移動物体検出手法," 情報処理学会研究報告, 94-CV-90, pp.1-8, 1994.
- (9) 伊東渡, 山田浩正, 上田博唯, 千葉隆広, "映像監視システムにおける差分用背景画像の照度変化追従性向上の検討," 情報処理学会研究報告, 98-CVIM-112, pp.1-7, 1998.
- (10) 波部斉, 大矢崇, 松山隆司, "動的環境における頑強な背景差分の実現法," MIRU'98, Vol.1, pp.467-472, Jul. 1998.
- (11) 中井宏章, 福井和広, 久野義徳, "3段階連続処理モジュールによる運動物体の検出," 信学論(D-II), J62-D-II, No.7, pp.1209-1218, Jul. 1974.
- (12) S. Iwasawa, K. Ebihara, J. Ohya and S. Morishima, "Real-Time Estimation of Human Body Posture from Monocular Thermal Images," IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.15-20, Jun. 1997.
- (13) G. Shafer, "A Mathematical Theory of Evidence," Princeton University Press, 1976.
- (14) S. A. Hutchinson, A. C. Kak, "Planning Sensing Strategies in a Robot Work Cell with Multi-Sensor Capabilities," IEEE Trans. on Robotics and Automation, vol.5, no.6, pp.765-783, Dec. 1989.
- (15) 山根定章, 青木啓志, 泉正夫, 福永邦雄, "基本確率割当てに基づくモデルベース物体認識," 信学論(D-II), vol.J77-D-II, no.11, pp.2247-2254, Nov. 1994.
- (16) 松山隆司, 栗田充隆, "Dempster-Shafer の確率モデルに基づくパターン分類—観測情報からの信念の形成と仮想信念空間を用いた信念の統合—," 信学論(D-II), vol.J76-D-II, no.4, pp.843-853,

Apr. 1993.

- (17) P. Smets, "About Updating," Proc. of 7th Conf. on Uncertainty in Artificial Intelligence, pp.378-385, 1991.
- (18) 石山豊, 久保文雄, 高橋裕信, 富田文明, "境界追跡型ラベリングボード," 信学論(D-II), vol.J78-D-II, no.1, pp.69-75, Jan. 1995.
- (19) 今川和幸, 呂山, 猪木誠二, 松尾英明, "肌色領域により隠れて見える場合を考慮した手話画像からの手の実時間追跡," 信学論(D-II), vol.J81-D-II, no.8, pp.1787-1795, Aug. 1998.
- (20) I. Haritaoglu, D. Harwood, L. S. Davis, "Ghost: A Human Body Part Labeling System Using Silhouettes," Proc. of 14th Inter. Conf. on Pattern Recognition, pp.77-82, Aug. 1998.
- (21) 小島篤博, 田原典枝, 田村武志, 福永邦雄, "動画における人物行動の自然言語による説明の生成," 信学論(D-II), J81-D-II, No.8, pp.1867-1875, Aug. 1998.
- (22) 間瀬健二, "顔とジェスチャの検出および認識," 日本ロボット学会誌, Vol.16, No.6, pp.745-748, Jun. 1998.

浅 沼 克 紀 (非会員) 1974 年生。1998 年 3 月大阪府立大学工学部情報工学科卒業。現在同大学院博士前期課程在学中。コンピュータビジョンの研究に従事。電子情報通信学会学生会員。

大 西 正 輝 (学生会員) 1973 年生。1997 年 3 月大阪府立大学工学部情報工学科卒業。1999 年 3 月同大学院博士前期課程修了。現在同大学院博士後期課程在学中。ロボットビジョンに関する研究に従事。電子情報通信学会, 日本ロボット学会, 映像情報メディア学会各学生会員。

小 島 篤 博 (非会員) 1967 年生。1990 年 3 月大阪府立大学工学部電気工学科卒業。1992 年 3 月同大学院博士前期課程修了。同年富士通(株)入社。1996 年 9 月より大阪府立大学・総合情報センター助手。画像処理, 人工知能, およびそれらの境界領域に関する研究に従事。情報処理学会, 人工知能学会, 電子情報通信学会各会員。

福 永 邦 雄 (正員) 1945 年生。1967 年 3 月大阪府立大学工学部電気工学科卒業。1969 年 3 月同大学院修士課程修了。同年同大学工学部電気工学科助手。現在同大学工学部情報工学科教授。コンピュータビジョン, グラフ理論とその応用などの研究に従事。情報処理学会, システム制御情報学会, IEEE 各会員。工博。