

## 肌色情報を用いた複数人物追跡

松村 朱里†      岩井 儀雄†      谷内田 正彦†

†大阪大学大学院 基礎工学研究科

あらまし      本研究は肌色情報を用いて画像中の人物領域を抽出し、追跡を行なうことを目的としたものである。肌色モデルを作成することによって学習されていない人物の抽出が可能となるうえ、人物に違和感を与えることなく追跡することができる。そして、その肌色情報を用いることで人物特徴を取得し、人物領域の前後関係を考慮することにより重なりを含めた複数人物の追跡が可能となる手法を提案する。また、複数人物を同時にかつ長時間撮影するため、撮影には全方位視覚センサを用いる。最後に実画像を用いて本手法の有効性を検討する。

### Tracking People by Using Skin-Color Information

Akari Matsumura†      Yoshio Iwai†      Masahiko Yachida†

†Graduate School of Engineering Science, Osaka University

**Abstract**      This paper proposes a method for finding and tracking people in input images by using skin color information. Making color models for skin color and background enables us to find people easily and to track people without any discomfort. Our proposed method can track overlapped people by considering their mutual positions and textures. We use a omnidirectional image sensor for taking images in order to capture many people for a long time in this paper. We conduct experiments for evaluation of the proposed method.

### 1 はじめに

人物の追跡は、画像処理の分野において重要なテーマである。例えば、侵入者の検知や人物の監視のような、セキュリティ面への応用や、人物のジェスチャー認識の前処理として利用されている。

人物の追跡処理は、人物領域抽出部と人物領域追跡部で構成される。人物領域抽出の手法として従来提案されている代表的なものは、背景差分法 [1]、カラー情報を用いる方法 [2]、人物領域モデルを用いる方法 [3] などである。これらの手法の中で、本研究で提案する手法はカラー情報を用いる手法に属する。

カラー情報を用いる手法として、本研究では人物の肌色情報を利用する。同じ人種の肌色は色空間上である程度まとまった分布をとることが分かっている [4] ので、このカラー分布を予め作成しそれを利用することで人物領域を抽出することができる。フレーム間差分とは違い、肌色情報を利用することで静止した人物を抽出したり影を排除したりすることが可能である。また、髪などの不必要な情報を含まない顔領域のみの抽出が可能であるため人物認識などへの発展が考えられる。

カラー情報が利用できるのは

1. 照明と物体とカメラが一定の関係にある
2. 照明が変動しない
3. 対象人物の撮像範囲が十分な大きさをもつ

などの制限条件が満足される必要があるが、上記の条件が満たされ、肌色情報が比較的安定に得ることができる室内で、固定カメラを用いて本研究を行なう。

次に、人物領域追跡の手法としては、各フレームで先ず追跡対象の特徴を全て抽出し、それらを記憶されている特徴群と比較し、識別していくという手法 [5] と、追跡対象の移動先を予測して追跡を行なう手法 [1,2,6,7] がある。前フレームの情報を用いて予測追跡を行なうことは、余分な領域の探索を行わない分計算時間が短縮される。このような予測追跡は、オプティカルフローを用いる方法 [2] やカルマンフィルタを用いる方法 [1,6,7] などがある。これらの方法の中でカルマンフィルタは線形ガウス型状態空間で行動モデルを作成するので、計算機での実現が容易であり、計算コストも低い。従って本研究ではカルマンフィルタを用いて予測追跡を行なう。

また、本研究では周囲 360 度の撮影が一度に可能である全方位視覚センサを用いて撮影を行なう。標準的な画角をもつカメラでの撮影は視野角が狭いため、複数人物を同時に撮影することが難しい。またそのようなカメラでは、人物がすぐに視野から外れてしまう恐れがある。そこで、この全方位視覚センサを用いれば広視野角の映像が得られるので、複数の人物を長時間に渡って撮影することが可能となる [8]。全方位視覚センサでの撮影画像は単位画角あたりの解像度が低いため、一般に色情報を用いての人物追跡が行なわれることは少ないが、本研究では上記のように室内で固定カメラを用いるという条件のもとで、どの程度肌色が抽出可能かを評価する。

## 2 システム概要

本システムの処理概要を図 1 に示す。図 1 のように、本システムでは予め作成された肌色モデルと背景モデルを用いて肌色領域の切り出しを行い、色情報を用いてその人物を追跡する。また、各肌色情報を格納しておくためのクラスタを複数用意し、切り出された肌色領域の情報を順次格納する。そして、このクラスタの状態に応じてカルマンフィルタを使用することで予測追跡を行い、人物の識別はテンプレートマッチングとそのクラスタの状態により行う。追跡対象は複数人物であり、各人物が自由に動き回ることによる隠れ、重なりも考慮する。

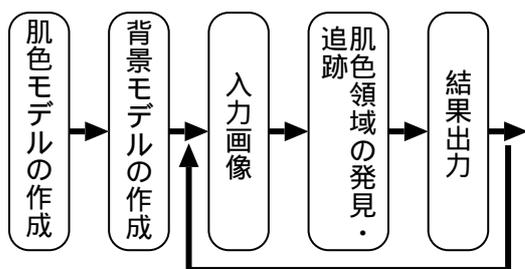


図 1: システムの概要

## 3 全方位視覚センサ [8]

本研究では周囲 360 度の撮影が一度に可能な全方位視覚センサを用いて撮影を行う。全方位視覚センサは視野角が広いため同時に複数の人物を視野から外れることなく撮影することが可能である。

全方位視覚センサは鉛直下向きの双曲面ミラーと上向きの CCD カメラからなる。3 次元世界座標系とカメラ座標系の構成を図 2 に示す。双曲面ミラー

は 3 次元座標系において式 1 のように表される。

$$\frac{X^2 + Y^2}{a^2} + \frac{Z^2}{b^2} = -1 \quad (Z > 0) \quad (1)$$

ミラーの焦点及びカメラのレンズ中心は  $(0, 0, c)$ 、 $(0, 0, -c)$  (ただし  $c = \sqrt{a^2 + b^2}$ ) に位置し、全方位画像面は XY 平面に平行でカメラのレンズ中心からカメラの焦点距離  $f$  だけ離れた位置に得られる。

3 次元空間内の点  $(X, Y, Z)$  に対応する全方位画像上の点  $(x, y)$  は以上より式 2、式 3 のように求められる。

$$x = X \times f \times \frac{(b^2 - c^2)}{(b^2 + c^2)Z - 2bc\sqrt{X^2 + Y^2 + Z^2}} \quad (2)$$

$$y = Y \times f \times \frac{(b^2 - c^2)}{(b^2 + c^2)Z - 2bc\sqrt{X^2 + Y^2 + Z^2}} \quad (3)$$

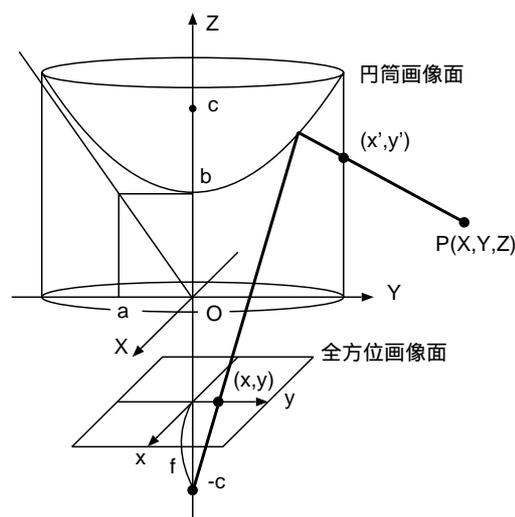


図 2: 座標系の構成

## 4 人物領域モデル

肌色領域の抽出は肌色モデルと背景モデルを用いて行なう。肌色モデルの他に背景モデルを利用することで、より正確に人物の肌色領域を抽出する。それぞれでモデルを作成し、最終的には両者を満たす領域を肌色領域とする。

### 4.1 肌色モデル

本研究では色空間上での肌色分布モデルを予め作成し、入力画像の画素値をその分布と比較することによって肌色領域の切り出しを行う。肌色分布モデ

ルは肌色分布が rgb 空間上で正規分布 (式 (4)) に従うと仮定し, 作成する. 予め数枚のサンプル画像より数人の肌色画素の値を取得し, rgb 空間上へ変換後, その平均・分散を求め正規分布を作成する. ただし, 輝度値が一定値  $T_R$  以下である場合は, その値は信頼できないものとしてたとえ肌色分布モデル上にあったとしても肌色とは認めないこととする. 入力画像の各画素における正規分布の密度関数の値が閾値  $T_d$  以上であればその画素を肌色画素とし, 肌色領域を抽出する.

$$\phi(z) = \exp \left\{ -\frac{1}{2} [z - \mu]^T \Sigma^{-1} [z - \mu] \right\} \quad (4)$$

$$z = \begin{bmatrix} r \\ g \\ b \end{bmatrix}, \quad \mu = \begin{bmatrix} \bar{r} \\ \bar{g} \\ \bar{b} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_r^2 & \sigma_r \sigma_g & \sigma_r \sigma_b \\ \sigma_r \sigma_g & \sigma_g^2 & \sigma_g \sigma_b \\ \sigma_r \sigma_b & \sigma_g \sigma_b & \sigma_b^2 \end{bmatrix}$$

ここで,  $\sigma_x^2$  はそれぞれ  $x = r, g, b$  の分散を示す.

## 4.2 背景モデル

背景にダンボールのような肌色に似通ったものがあると誤って肌色領域として抽出される恐れがあるので, 背景情報を加えることで誤抽出を軽減する. 背景モデル作成のためにまず背景画像を作成する. そして, 背景モデルは入力画像上の画素が背景であるならば, 画素値は背景画像での各 RGB 値を平均とした正規分布に従うと仮定し, 作成する. この正規分布による各画素での背景らしさの尤度は式 (5) のように求める. ここで, RGB の各分散は画像中で一様とする.

$$B(X) = \exp \left\{ -\frac{1}{2} \frac{(R - \bar{R})^2}{\sigma_R^2} - \frac{(G - \bar{G})^2}{\sigma_G^2} - \frac{(B - \bar{B})^2}{\sigma_B^2} \right\} \quad (5)$$

ここで,  $\sigma_x^2$  はそれぞれ  $x = R, G, B$  の分散を示す.

各画素の肌色モデルを  $S(X) (= \phi(z))$  とすると, 背景モデルも加えて最終的な肌色領域らしさは

$$S(X) \{1 - B(X)\} \quad (6)$$

となる. この値は肌色領域の重心を求める際の重みパラメータとして利用する.

## 5 追跡対象と状態の定義

それぞれの肌色領域を各フレームにおいて連続に追跡するために, 本システムでは認識した肌色領域の情報を格納しておくためのクラスタを予め設定された数だけ保持している. 各肌色領域は1つのクラスタに対応する. 各クラスタは以下のような情報を保持している.

1. 重心位置
2. 速度
3. クラスタのテンプレート
4. クラスタの状態

クラスタの状態とは absence, appear, exist, lap, missing の5つの状態のいずれかに当てはまる.

absence はクラスタ未配当の初期状態, appear は肌色領域発見状態, exist は安定に肌色領域を追跡している状態, lap, missing はいずれも追跡中に肌色領域を見失ってしまった状態を示す. lap, missing の違いは, lap が他の肌色領域によって隠されたことによって見失ってしまった状態を示し, missing は肌色でない物体によって隠され見失ってしまった状態や肌色領域が発見できてもテンプレートマッチングの差が大きく前フレームと同じ肌色領域とみなされなかった状態を示すことである. 見失ってしまった状態を lap と missing の2状態を示すことで, 人物同士の間を重なりを含めた追跡を可能とする. 追跡時に肌色領域を見失ってしまったり, 再発見したりすれば図3のように状態が遷移する. このような情報を持ったクラスタの更新を各フレームでそれぞれ独立に行う.

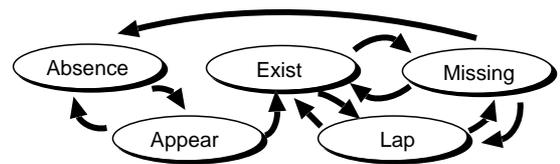


図3: クラスタの状態遷移

## 6 肌色領域の発見と追跡

### 6.1 領域発見

肌色領域は画像中をランダムサンプリングすることによって発見する. ランダムサンプリングによって画像中を全走査するのと比較して計算時間が短縮される. ランダムサンプリングのサンプリング数  $N$  は, 肌色領域を人物肌色領域と認める最小の面積  $T_a$

から決定される．肌色領域の縦，横の長さを  $W, H$  とすると  $N$  は以下ようになる．

$$N = \frac{W \cdot H}{T_a} \quad (7)$$

肌色領域が発見されれば肌色らしさの重みパラメータを用いて重みづけ重心を求め，クラスタに情報を格納する．このときに肌色領域の面積が閾値  $T_a$  以下であればノイズであるとみなし人物顔領域と認めず，クラスタに情報を格納しない．面積が閾値  $T_a$  以上である場合は，初期状態のクラスタに情報を格納し，クラスタ状態を `appear` に移行する．状態 `appear` でカルマンフィルタに必要な情報を集めた後状態 `exist` に移行する．

## 6.2 カルマンフィルタによる予測追跡

クラスタの状態 `exist` での肌色領域の追跡はカルマンフィルタを用いて予測しながら行う．各フレーム間の時間差が微小であるという前提と奥行情報の考慮を行っていないことから，人物は画像上で等速直線運動を行なうと仮定する．ただし，本研究での入力画像は全方位画像であり，全方位画像上では等速直線運動とならない．そこで，図2にあるように全方位画像上の点  $(x, y)$  を円筒画像上の点  $(x', y')$  へ変換した値を状態変数として用いる．カルマンフィルタによる状態変数の予測は以下ようになる[9]．

状態変数を  $X_n = [x', y', \dot{x}', \dot{y}']^T$  とおく．状態更新式は

$$X_{n+1} = \mathbf{A}X_n + \mathbf{B}\omega_n \quad (8)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}$$

ただし， $\omega_n$ : 白色雑音  $E[\omega_i \omega_j^T] = \mathbf{Q}\delta_{ij}$   
( $i, j = 1, 2, 3, 4$ )

観測方程式は

$$Y_n = \mathbf{C}X_n + \varepsilon_n \quad (9)$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}$$

ただし， $\varepsilon_n$ : 白色雑音  $E[\varepsilon_i \varepsilon_j^T] = \mathbf{R}\delta_{ij}$   
( $i, j = 1, 2, 3, 4$ )

推定値  $\hat{X}_n$ ，予測値を  $\bar{X}_n$  とすると

$$\hat{X}_n = \bar{X}_n + \mathbf{K}_n(Y_n - \mathbf{C}\bar{X}_n) \quad (10)$$

となり，現在値  $Y_n$  より推定値  $\hat{X}_n$  を修正する．この修正後の推定値  $\hat{X}_n$  より，次フレームの予測値  $\bar{X}_{n+1}$  を求める．

$$\bar{X}_{n+1} = \mathbf{A}\hat{X}_n \quad (11)$$

ここで， $\mathbf{K}_n$  はカルマンゲインであり，推定誤差分散を最小するように以下のように設定される．

推定誤差  $\epsilon_n = \hat{X}_n - X_n$  に対し推定誤差分散行列  $\mathbf{P}_n = E[\epsilon_n \epsilon_n^T]$  を最小にするカルマンゲインは，予測誤差分散行列を  $\mathbf{W}_n$  とすると

$$\mathbf{W}_{n+1} = \mathbf{Q} + \mathbf{A}\mathbf{P}_n\mathbf{A}^T \quad (12)$$

$$\mathbf{K}_n = \mathbf{W}_n\mathbf{C}^T[\mathbf{R} + \mathbf{C}\mathbf{W}_n\mathbf{C}^T]^{-1} \quad (13)$$

$$\mathbf{P}_n = [\mathbf{I} - \mathbf{K}_n\mathbf{C}]\mathbf{W}_n \quad (14)$$

で求まる．ここで， $\mathbf{P}_n$  はリカッチ方程式の解となる．

以上のようにカルマンフィルタによって各顔領域の移動先  $(x', y')$  と移動予測分散  $(Px'_v, Py'_v)$  が得られるので，その予測域内  $(Px'_v + W) \times (Py'_v + H)$  で肌色の重心を推定する．もし，予測域内に肌色が検出されなければクラスタの状態は `missing` に移行するが，カルマンフィルタでの予測追跡を続行することで再び肌色画素の発見を目指す．ある閾値  $T_F$  フレーム以上に渡ってこの `missing` 状態が続けば，カルマンフィルタの予測値も信頼できない値となるので，そのクラスタの肌色領域は消滅したとして追跡を終了する．肌色領域を見失っているもう一つの状態 `lap` でも `missing` 状態と同様にカルマンフィルタでの追跡を続行する．また，本研究ではフレーム間での各クラスタの識別評価にテンプレートマッチングを用いている．予測域内で肌色画素が発見されても，フレーム間でテンプレートとの差が閾値  $T_T$  より大きければ，この場合も `missing` 状態に移行する．

## 6.3 フレーム間のクラスタ識別

本研究ではフレーム間での各クラスタの識別評価にテンプレートマッチングを用いている．テンプレートは顔の一部で作成しクラスタごとに保持している．各テンプレートは安定に肌色領域を追跡している限り(状態 `exist`)，各フレーム毎に随時更新する．隠れの状態に遷移した場合(状態 `lap, missing`)は隠れ状態前でテンプレートの更新を停止し，次フレームからは再び合致する肌色領域を探索する．合致する肌色領域が発見されれば，テンプレートを更新し再び追跡を行う．

ここで，人物同士が重なった場合の処理について述べる．なんらかの肌色でない物体によって肌色領域が隠されている場合はテンプレートマッチングの

差により missing 状態に移行することで対応しているが、他の人物などの肌色の物体に隠れてしまった場合は人物の識別がうまくいかない。これは人物識別のためのテンプレート内は肌色画素が多く、他の人物のテンプレートとの差がそれ程大きくなならない場合があるためである。従って新たに lap という状態を用意し、他のクラスタ領域に重なった場合はこの状態に移行することで重なりに対応する。

重なり処理方法として、まず各クラスタ番号によって画像上のカルマン予測域内のラベル付けを行う。予測域に重なりが生じる場合は以下のような条件からクラスタ番号を優先的に振り当てていく。

条件 1) 前フレームで lap または missing 状態にない。

条件 2) テンプレートマッチングの差が小さい。まず、両クラスタで条件 1) について調べどちらのクラスタも条件を満たせば条件 2) について調べる。どちらかのクラスタのみ条件 1) を満たす場合は満たしたクラスタ番号でラベル付ける。どちらのクラスタも条件 1) を満たさない場合は存在しないことを表すクラスタ番号でラベル付けしておき、その領域はどのクラスタにも属さないようにする。後は、それぞれのテンプレートマッチングの結果によって状態の移行を行う。この処理によって、予測域面積の一定以上の割合  $R_a$  だけ他のクラスタ番号に占領されてしまった場合は lap 状態に移行する。このようにして隠れ状態を missing と lap とで区別する。

## 7 実験

デジタルビデオカメラに全方位視覚センサを取り付け撮影を行った。複数人物に自由に移動してもらいその様子を 10 分間撮影した。人物はそれぞれ肌色モデル作成時に利用された人物と、モデル作成には用いられていない人物とが混在する。

### 7.1 実験結果

実験結果を図 4, 5 に示す。肌色領域の周りを囲む四角形の大きい方がカルマン予測域で、小さい方が実際に発見した肌色領域を示す。それぞれの肌色領域の軌跡を線で示す。

フレーム間に渡り複数人物の同時追跡が可能となっている。また、人物同士が重なった場合も重なり後の人物が隠れ状態に遷移することで重なりを含めた追跡が可能であることが分かる。重なり後の人物が再び現れた場合は重なり前のテンプレートと合致することで重なり後も重なり前と連続して同じクラスタで追跡している。

## 8 まとめ

本報告では、全方位視覚センサにより得られる全方位画像から肌色情報を用いることにより人物の顔領域の切り出しを行ない、追跡する手法について報告した。全方位画像は単位画角あたりの解像度が低く色情報が適切に得られない場合があるが、前フレームの情報によって補間することで人物の切り出しが適切に行なわれることが確認できた。また、テンプレートを用いることで各人物の識別が可能になることが確認できた。

## 9 参考文献

- [1] 波田 尚哉, 三宅 哲夫「アクティブビジョンシステムによる遮へいを伴う移動物体の追跡」, 信学会 (D-II), Vol. J84-D-II, No. 1, pp. 93-101 (2001)
- [2] 小渡 悟, 星野 聖「オプティカルフローと色情報に基づく拳の検出と追跡によるジェスチャー認識」, 情報処理学会, CVIM, No. 125, pp. 47-54 (2001)
- [3] 白井 良明「複雑背景における人の追跡」, 情報処理学会, CVIM, No. 128, pp. 81-88 (2001)
- [4] 谷内田 正彦『ロボットビジョン』, 昭晃堂, 1990
- [5] 菅 幹生, 田谷 基教, 湊小 太郎, 山村 可奈子, 友久 久雄, 小森 優「全方位カメラを用いた発達障害児の無拘束行動測定システム」, 信学論 (D-II), Vol. J83-D-II, No. 10, pp. 2320-2327 (2001)
- [6] T. Broida, R. Chellappa, "Estimation of Object Motion Parameters from Noisy Images", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 8, No. 1, pp. 90-99 (1986)
- [7] 富山 忠文, 松山 隆司「ダイナミックメモリを用いた実時間対象追跡」, 情報処理学会, CVIM, No. 121, pp. 49-56 (2000)
- [8] 山澤 一誠, 八木 康史, 谷内田 正彦「移動ロボットのナビゲーションのための全方位視覚系 HyperOmni Vision の提案」, 信学論 (D-II), Vol. J79-D-II, No. 5, pp. 698-707 (1996)
- [9] 片山 徹『応用カルマンフィルタ』, 朝倉書店, 1983



frame : 1



frame : 2



frame : 3



frame : 4



frame : 5



frame : 6



frame : 7



frame : 8



frame : 9



frame : 10

図 4: 人物の追跡実験結果



frame : 1



frame : 2



frame : 3



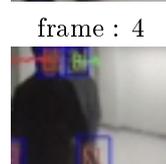
frame : 4



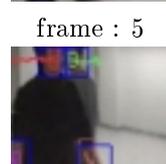
frame : 5



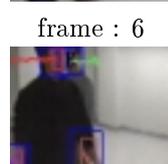
frame : 6



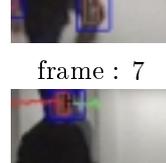
frame : 7



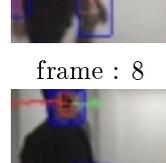
frame : 8



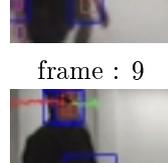
frame : 9



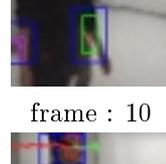
frame : 10



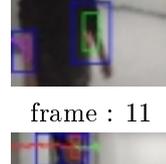
frame : 11



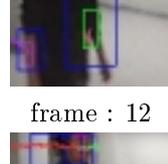
frame : 12



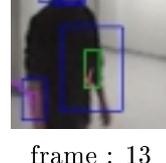
frame : 13



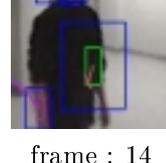
frame : 14



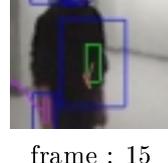
frame : 15



frame : 16



frame : 17



frame : 18

図 5: 重なりの処理の結果