

## 複雑背景における人の追跡

白井良明<sup>†</sup> 三浦純<sup>†</sup>

屋内や屋外で歩く複数の人を追跡する研究をいくつか述べる。まず、カメラが静止している場合に関しては、背景がほぼ一定の場合の多人数の追跡、人の動きによって背景の色や明るさが変化する場合、および背景に動きがある場合の追跡を述べる。カメラのパン、ティルトをともなう場合に関しては、オプティカルフローを用いる方法、距離、明るさなどと組み合わせる方法について述べる。最後に、複数カメラの協調による追跡を述べる。

### Human Tracking in Complex Environment

YOSHIAKI SHIRAI<sup>†</sup> and JUN MIURA<sup>†</sup>

This paper describes several research results on tracking humans in indoor or outdoor scenes. Concerning the tracking by stationary cameras, tracking methods for the following conditions are described: tracking of many persons in a city street, that of a few persons in a corridor whose color is influenced by the walking persons, and that of a person in a background which contains random motions. Concerning the tracking by active cameras, tracking methods are described which uses optical flow or integrates optical flow, range data, and uniform brightness regions. Finally, a coordination method of multiple cameras is described for tracking a person with heavy occlusions.

#### 1. はじめに

人の追跡は、人のモニタや監視などの直接的な応用以外にも、人のジェスチャや姿勢の認識や顔認識のための前処理としても使われる。この研究は古くから行われているが、最近では、オンライン処理が可能な範囲が広がっているため、研究や開発がさかんになっている。

人の追跡のためには、人の発見と発見した人の追跡が必要である。

人の発見の初期の研究は、静止画像を対象としていた。たとえば、濃淡画像から顔を発見する試みがあったが<sup>1)</sup>、色が使えないので、ほぼ正面を向いている人に限定されていた。カラー画像や距離や動き情報が得られるようになると、以下の手がかりが使えるようになる。

- 肌色
- 画像の背景差分
- 距離画像の背景差分
- 画像の時間差分

#### ● オプティカルフロー

信頼性を上げるためには、以上の項目を組み合わせることも多い。

人の追跡は、人の発見を繰り返すことによっても可能であるが、普通は以前の画像で得られている情報を利用する。事前情報の利用により、よりロバストで効率良く追跡を行えるので、ここではこの場合の追跡に限定する。さらに、人の発見は事前情報が乏しい場合であり、追跡の特殊例と見なすこととし、ここでは事前情報を利用した追跡を中心に述べる。

人の追跡を一般の動物体追跡と比較したときの特徴は、

- その場回転をともない、見え方が変わる。自動車では移動しながら緩やかに回転するのみであり、ボールは回転しても見え方が一定である。
- 非剛体、関節物体である。人の領域の動きが均一でない。このような対象だけを扱ったワークショップも開かれている<sup>2),3)</sup>。

一般に、人の追跡は難しいが、人の動きの制約を用いることができるため、1枚の静止画像から人を見つけることが困難な場合でも前後関係から人を切り出せる可能性がある。追跡の条件によっては、すでに実用化しているものもあるが、依然として困難な問題も多

<sup>†</sup> 大阪大学大学院工学研究科

Graduate School of Engineering, Osaka University

数残されている。

その追跡の条件としては以下のような項目がある。

- (1) 対象：人は1人か複数か。人の重なりがあるかないか。他の物体に隠されることがあるかないか。衣服の色に制限があるかないか。
- (2) 動き方：直線運動か、面内曲線運動か、あるいは一般3次元運動か。関節は動くか。
- (3) 背景：背景のテクスチャは一樣かそうでないか。距離は連続であるか不連続のところがあるか。時間的に一定か変化するか。
- (4) カメラ：固定か、回転運動か、並行移動か、一般運動か。カメラは1台か複数か。
- (5) タスク：1人だけを追跡するか、複数を同時に追跡するか。いったん動画像を蓄えてから追跡すればいいか、オンラインで追跡する必要があるか。人の姿勢も求める必要があるか。

これらの条件は必ずしも独立ではない。たとえば、姿勢が必要な場合は、関節は動くことが前提である。しかし、逆は真でなく、単に人を追跡するだけで、姿勢が不要である場合も、歩いている人の足や手の関節は動く。この場合、関節の動きは追跡の妨げになることがある。

多くの研究はいくつかの困難な条件を扱い、それ以外の条件を単純化している。たとえば、人の姿勢を求める場合は、人の輪郭が簡単に切り出せるという条件をつけることが多い<sup>4)</sup>。また、オンライン追跡を行う場合は、カメラが固定である、カメラモデルが簡単で背景差分が使える、あるいは人の重なりがない、などの条件がつけられていることが多い。

本論文では、人の姿勢を求める問題と、人の向いている方向を求める問題<sup>5)</sup>を除き、種々の条件での人の追跡を述べる。

## 2. カメラ固定、背景一定の場合

屋外環境で人を追跡する場合は、固定カメラで、背景がほぼ一定という条件が成立することがある。典型的な例は、スポーツにおける選手の追跡や、道路などでの人の追跡である。前者は、登場人物が決まっている、服装の制限も使えるので、選手の抽出が比較的簡単である<sup>6)</sup>。道路などでは、人の種類や服装が多様であり、人数が多いことが問題となる。本章では後者の例として、雑踏における複数歩行者の追跡<sup>7)</sup>を述べる。

都市デザインの分野では、快適な都市づくりのために歩行者の群の振舞い(生成・移動・消滅など)の解析<sup>8)</sup>が行われている。都市の雑踏が数時間にわたって撮影されたビデオテープを調べなければならないので、



図1 歩行者天国の雑踏(鍛佳代子氏(東京工芸大学)の提供)  
Fig.1 A crowd in a vehicle-free promenade.

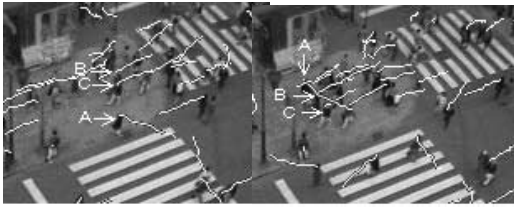
歩行者追跡ツールの開発が必要である。

都市の雑踏では歩行者が他の歩行者や電柱などの陰に入って隠蔽されることがあり、歩行者同士の隠蔽では3人以上が重なって見えることも珍しくない。図1では、隠蔽が上部歩道だけでも10個近く見られ、歩行者の検出と追跡を困難にしている。

人の検出は、処理が簡単な背景差分法を用いて行う。雑踏では、画像中にはつねに歩行者が行き交うので、背景を撮影することができない。そこで、一連の動画像から各ピクセルのR, G, B成分(各8ビット)それぞれについてヒストグラムを作り、モード(最頻値)を背景のピクセルの成分とする。

背景差分では、重なって見える人を分離したり、背景と似ている色の人を検出したりすることが困難である。そこで、全画像系列を通して人の検出を行い、得られた候補領域の大きさと形状によって、単独に分離していることが明らかな歩行者のみを検出する。いったん検出されれば、その時点から未来および過去へ追跡する。これにより、歩行者が重なることがあっても、どこかで検出できれば、追跡することができる。追跡は、最初に検出された人の領域をテンプレートとし、相関が最も高い位置を探索する。相関は、SAD(Sum of Absolute Difference)によって計算する。探索範囲は、過去の位置から求めた予測位置の周りとする。

追跡で問題になるのは隠蔽前後における対応付けである。従来は、隠蔽の前後で見かけの類似性から対応付けを決めていた<sup>9)</sup>。しかし、類似の歩行者同士が重なる場合には誤対応を生ずることがある。そこで、対応付けにあいまいさがある場合は対応の可能性のある候補をすべて保持したまま追跡を続け、明らかに候補でないことが分かったときにそれを除去する。この処理を、あいまい性がなくなるまで続ける。このようにして信頼性の高い追跡を実現している。図1を含む動



35th frame 55th frame

図2 歩行者追跡の例

Fig. 2 Experimental result of pedestrian tracking.

画像に対して本手法を適用し、その中の2フレームだけについて、歩行者の頭部の軌跡を図2に示す。歩行者Aが、右前方から来る歩行者BおよびCを左に迂回して避けている様子や、この間、歩行者Cが歩行者Aにじゃまされてあまり前へ進めなかった様子が分かる。

### 3. 背景が人の動きの影響を受ける場合

普通の背景差分法には以下の問題点がある。

- (1) 影を人の一部として検出する。上から照明され、カメラも人を見下ろしているという条件では、影が小さく見えるので、その影響は少ない。しかし、この条件が満たされないと、影が大きく検出されてしまう。
- (2) 屋内では、影が壁などに投影されると、影と人の区別がむずかしい。
- (3) 床での反射が正反射成分を持つと、人が映り込んでいる領域が検出される。

廊下を歩く人を追跡する場合は、以上の問題が発生するので、人の形の推定が正確でなくなる。

もし3次元情報が得られれば、影が既知平面に投影されると仮定して、影を除くことができる<sup>10)</sup>が、1台のカメラだけでは、3次元情報が得られない。一方、影の領域の色度は大きく変化しないと仮定する手法もある<sup>11)</sup>。しかし、実際の影の領域では、色度は明るさの変化に影響を受けて変化する。また、映り込みへの一般的な対処は困難である。

本章では、与えられた環境の特性を学習するアプローチ<sup>12)</sup>を述べる。すなわち、学習段階で、種々の対象物体や影の領域を用いて色度の分布を作成する。検出段階では、これらの色度の分布を用いて各ピクセルが各物体に属する確率を求める。そして、典型的な人の形のモデル(テンプレート)を用意しておき、人との対応の尺度を求め、人の領域を探す。

#### 3.1 ピクセルの分類

まず、カラー画像の各ピクセルを、影のない背景、影、前景に分類し、次に、前景の領域の中から人の顔

の領域を抽出する。そこで、ピクセルを次の4つのクラスに分類する。

背景： $\omega_b$

人の肌色： $\omega_h$

人が投じる影： $\omega_s$

その他の前景(衣類や髪など)： $\omega_o$

検出段階において、画素値  $C = (R, G, B)$  をとるピクセルがクラス  $\omega$  に属する確率は、

$$P(\omega|C) = \frac{P(C|\omega)P(\omega)}{P(C)}$$

である。事前確率  $P(\omega)$  は、知識として与えられていなければ、等確率とする。尤度  $P(C|\omega)$  は学習段階で求める。

#### 3.2 ピクセルに関する各クラスの尤度

背景の尤度は、だれもいない間に撮影した背景画像100枚から求める。各ピクセルの  $R, G, B$  はガウス分布に従うとし、平均と分散  $\{\mu_R, \mu_G, \mu_B, \sigma_R^2, \sigma_G^2, \sigma_B^2\}$  を求める。 $R, G, B$  を0から255までの整数値とし、背景の尤度  $P(C|\omega_b)$  を以下によって求める。

$$P(C|\omega_b) = P(R|\omega_b) \cdot P(G|\omega_b) \cdot P(B|\omega_b)$$

ここで、 $c = R, G, B$  に対して、 $P(c|\omega_b)$  は以下の式で求める。

$$P(c|\omega_b) = \int_{c-0.5}^{c+0.5} \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left\{-\frac{(x-\mu_c)^2}{2\sigma_c^2}\right\} dx$$

肌色の尤度も実画像の肌領域から求める。肌色の場合は、対象の肌色、照明条件を変数とする確率を求めなければならない。標本の数を限定して各変数に対する十分な確率を得るために、色と明るさの分布が独立であると仮定し、尤度を明るさと色に分解して分布を求める。

$$P(C|\omega_h) = P(C|\omega_h, I) \cdot P(I|\omega_h)$$

ここで明度  $I$  は  $C$  の明度 ( $R + G + B$ )。明度の分布  $P(I|\omega_h)$  は、標本の明るさの最小値から最大値までの範囲での一様分布とする。一方、ある明度における色の確率は、肌色領域の画素値  $C$  のヒストグラム  $H_h(C)$  を用いて次式で与える。

$$P(C_k|\omega_h, I_k) = \frac{H_h(C_k)}{\sum_{I(C)=I_k} H_h(C)}$$

人の影は、すべての背景に対して影ができた場合の色の確率分布を各ピクセルについて求めなければならない。すると、必要な記憶量と確率分布作成の手間がかかるので、背景を類似の色の領域に分割し、各領域を1つのグループとして扱う。各グループについて、さまざまな照明下で背景の色を撮影してヒストグラムを作成する。肌色の場合と同様に、 $P(C|\omega_s)$  を  $P(C|\omega_s, I)$  と  $P(I|\omega_s)$  に分解して、 $P(C|\omega_s, I)$  を



図3 廊下での人の追跡 (原画像と人の抽出結果)

Fig.3 Human tracking in a corridor (input image and extraction result).

ヒストグラムから求める.  $P(I|\omega_s)$  は、影ができた場合に明るさは暗くなるという仮定を適用して、背景の明度より暗い範囲で一様分布を与える.

その他の前景は、衣服などを含み、あらかじめ色を予測できないので、その尤度  $P(C|\omega_o)$  は一様分布とする.

### 3.3 人のテンプレートを用いた検出

学習段階において、頭、上半身、下半身で構成される人のテンプレートを作る. 検出段階において、これらのテンプレートについて、事後確率の計算結果をもとにして人らしさの度合を評価する.

テンプレートの人らしさの評価は、頭、上半身、下半身のそれぞれについての評価値  $f_0, f_1, f_2$  を計算する. たとえば、

$$f_0 = \frac{P(\omega_h|C) + 0.5P(\omega_o|C)}{\alpha}$$

ここで、オーバーラインは頭テンプレート内の加重 ( $\alpha$ ) 平均を示す. 人のテンプレートの総合評価値は  $f_t = f_0 \times f_1 \times f_2$  で与える.

種々の衣服の色の人の実時間追跡実験を行った. 図3に例を示す.

## 4. 背景に動きがある場合の人の追跡

背景の変動が大きい場合には差分による人物検出は難しい. 背景の変動をモデル化することによりある程度の変動に対処できるが<sup>13),14)</sup>, 屋外環境で変動が大きい場合にはモデルの作成が困難である.

画像上の動き情報から移動する人物を検出する場合には、観測された動きが背景の変動によるものなのか、背景上を移動する物体によるものなのかを区別する必要がある. そこで、背景はランダムに変動するのに対し、移動物体は画像内で滑らかな運動をすると仮定して、両者を区別する<sup>15)</sup>.

まず、入力画像から画面内での速度場であるオプティカルフローを計算し、ある程度の面積を持つ均一フロー領域を移動物体の候補領域として抽出する. 次に、各時間で得られた候補領域が等速直線運動すると

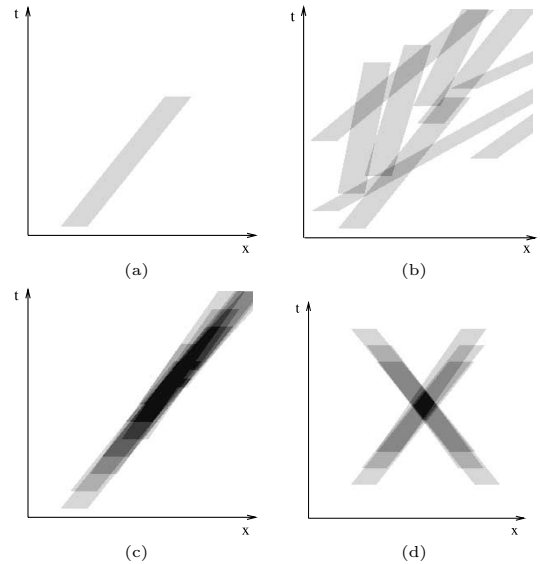


図4 投票空間 ( $x, t$  の2次元のみ表示)

Fig.4 Voting space (showing only 2 dimensions  $x, t$ ).

いう仮定の下に、各時点で得られたデータを投票により統合する.

投票空間は位置  $x-y$ , 時間  $t$ , 移動方向  $\theta$  の4次元からなる.  $t$  は予測する未来の時間分だけ確保し、つねに  $t=0$  が現在になるように、1単位時間ずつ過去方向へシフトさせていく.  $\theta$  は候補領域の進行方向であり、8方向程度に粗く量子化する.

投票方法は以下のものである. まず  $t=0$  における位置は現時点で得られた候補領域の重心位置とし、 $t>0$  では、候補領域の平均フローをもとに等速直線運動するものとして予測した重心位置を用いる (図4(a)参照). 背景変動により発生した候補領域からの投票値は分散してしまう (図4(b)参照) が、実際に移動領域により発生した候補領域からの投票値は、次第に加算され大きな値になってくる (図4(c)参照). これをしきい値処理によって抽出し、移動物体と見なす.  $\theta$  に関しては、候補領域の平均フローが相当する  $\theta$  にだけ投票する.  $\theta$  軸は、複数領域の予測移動パスが交差するとき (図4(d)参照), 投票値が突然高くなってしまいうのを避けるために必要となる.

提案する処理を9個のDSPを用いて実時間システムとして実装した<sup>16)</sup>. 図5に人物の抽出結果 (1フレームのみ) を示す.

## 5. カメラが動く場合の追跡

カメラが動く場合には背景差分や時間差分が使えないので、それ以外の手がかりを用いる必要がある. 目標物体の見え方が一定ならば、画像中の領域の輝度分



図5 複雑変動背景下での人物の発見  
Fig. 5 Person detection in changing background.

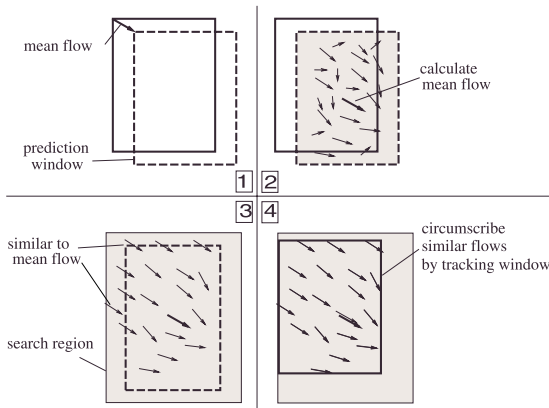


図6 オプティカルフローに基づく追跡  
Fig. 6 Optical flow-based tracking.

布の相関に基づく手法が有効である。しかし、画像間で目標物体の見かけの明るさや形状が変化する場合に、追跡が困難となる。本章では、動き情報による追跡、および動き情報と他の情報を組み合わせた追跡について述べる。

### 5.1 オプティカルフローを用いた追跡

人間が撮像系の光軸とある程度の角度を持って移動する場合、手足の先が移動の方向と違う方向に動くことがあるが、普通の歩行では手足の移動は体全体の移動の割合からすれば小さい部分なので、観測される画像中の人間の領域内のフローベクトルはほぼ一樣になる（そのような領域を目標フロー領域と呼ぶ）。このことを利用し、予測される目標人物の位置の周囲で目標フロー領域を探索することにより、目標人物を検出する<sup>17)</sup>。

追跡処理の概要は以下のものである（図6参照）：

- (1) 前フレームでの目標人物の位置と大きさ（矩形の追跡ウインドウ）および平均フローから、現フレームでの位置を予測（予測ウインドウ）。
- (2) 予測ウインドウ内でフローとその平均を計算。
- (3) 予測ウインドウおよびその近傍で平均フローと似た速度を持つ領域を探し、物体領域とする。



図7 アクティブカメラによる追跡  
Fig. 7 Tracking by active camera.

### (4) 物体領域を囲むように追跡ウインドウを更新。

図7に追跡結果を示す。黒いウインドウがフローから得られた領域（追跡ウインドウ）である。ここではさらに、画像中の人間の形状がほとんど変化しないという仮定を加えて、白いウインドウを人間領域としている。

この手法をもとに、複数の物体の同時追跡と隠蔽検出を行う手法も提案している<sup>18)</sup>。

### 5.2 オプティカルフローと距離を統合する方法

前節の方法は、物体の形状や輝度の分布パターンが滑らかに変化する場合にも追跡できるが、目標物体の周辺に似た速度の物体がある場合には目標物体だけを検出できない。距離情報を用いる手法<sup>19)</sup>では、前フレームの追跡結果から目標物体の距離を予測し、その予測に近い距離を持つ領域を目標物体領域とするが、目標物体の周辺に似た距離の物体と区別することができない。

そこで、複数の情報を用いて目標物体領域を抽出し、追跡を行う手法が提案されている。たとえば、速度情報と色情報、速度情報と距離情報<sup>20)</sup>などがある。ところが、これらの手法では、片方の情報で目標物体を他の物体と区別できなくなると、追跡を行うことができない。

したがって、複数の情報を同時に検出し統合することが必要になる。ここでは、オプティカルフローと距離情報の統合による追跡<sup>21)</sup>について述べる。なお、本手法では距離情報として、ステレオカメラの左右画像間の見え方の違いによって生じる視差を用いる。

#### 5.2.1 追跡処理の概要

各フレームでは、現在の右画像と1フレーム前の右画像を用いてオプティカルフローを、左右画像から視差を抽出する。目標物体の状態によって以下の3つの場合に分け、それぞれについて異なった追跡処理を行う。

- 状態1 目標物体が隠蔽されていない状態。
- 状態2 目標物体の一部が隠蔽された状態。
- 状態3 目標物体が完全に隠蔽された状態。

各フレームで現在の状態の判断を行い、それぞれの

状態によって異なった処理を行う。

初期フレームでは、まず目標物体の初期領域を抽出し、この初期領域をもとに、目標物体が隠蔽されていない状態（状態1）であるとして、追跡を開始する。

状態1の場合、目標物体の領域を更新し、目標物体領域周辺で目標物体を隠蔽する物体を探索する。このとき、十分な隠蔽領域が検出されれば、物体の一部が隠蔽された状態（状態2）であるので、目標物体領域の更新とともに隠蔽物体領域を更新する。隠蔽物体と目標物体が近接していなければ、隠蔽物体の追跡を行う必要はないので、状態1に戻る。

状態2の次のフレームで、目標物体領域がなくなると、目標物体の全体が隠蔽された状態（状態3）とし、隠蔽物体領域を更新し、隠蔽物体領域周辺で目標物体を探索する。目標物体領域が検出されれば、状態1に戻る。

### 5.2.2 目標物体が隠蔽されていない状態での追跡

右画像上の点  $p_i = (x, y)$  で観測  $o_i = (u_i, v_i, d_i)$  が得られたときにその点が目標物体 ( $T$ ) に属する確率を、目標物体の領域予測と目標物体上の点の速度および視差の分布の推定に基づいて計算し、その確率が高い点を目標物体領域とする。この確率を  $P(p_i \in T | o_i)$  と表す。

まず、領域の予測は、基本的には前フレームでの目標物体領域とその速度から線形予測で行うが、その際前フレームでの速度・視差の推定の誤差分布と加速度の影響を考慮する。これにより、現フレームで各点が領域に属する事前確率  $P(p_i \in T)$  を求める。

次に、速度、視差の分布がそれぞれ独立な正規分布に従うとして、予測領域内の速度、視差についてそれぞれ平均と分散を推定する。このとき、速度、視差の信頼性を考慮した重みづけを行う。これにより、ある点が物体に属するとしたときの観測値の確率  $P(o_i | p_i \in T)$  を求める。

以上から、目標物体上の点で  $o_i$  が観測された場合の確率は、ベイズの定理により次式で計算できる。

$$P(p_i \in T | o_i) = \frac{P(o_i | p_i \in T) P(p_i \in T)}{P(o_i)} \quad (1)$$

### 5.2.3 隠蔽の処理

目標物体領域の周辺では目標物体を隠蔽する物体を探索する。目標物体領域の周辺で、視差が目標物体の視差よりもある程度大きい領域（隠蔽物体領域）は、目標物体を隠蔽する物体と判断する。各フレームで観測される視差は雑音が多く信頼性がないので、複数のフレームにわたって隠蔽物体と判断された点を隠蔽物体領域とする。目標物体が完全に隠蔽された後は、隠

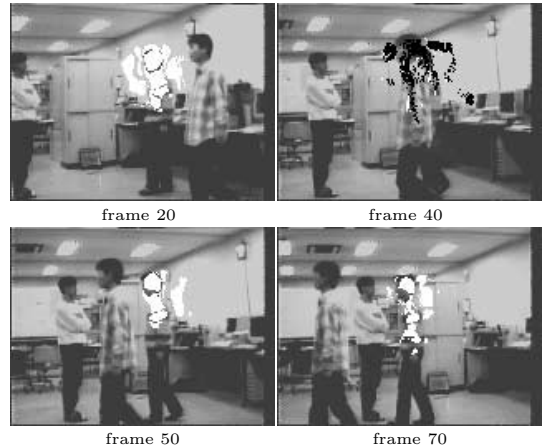


図8 追跡結果（白：目標物体領域，黒：目標物体を隠蔽する物体領域）

Fig. 8 Tracking result (white: target object region, black: occluding object region).

蔽物体の周囲に注目し、隠蔽前の目標物体の視差に近い領域が得られれば、目標物体が再び現れたと判断する。

### 5.2.4 追跡例

本手法で追跡を行った結果を図8に示す。第40フレームでは目標物体は完全に隠蔽されて見えなくなっているが、第50フレームでは目標物体を再び検出し、追跡している。隠蔽物体は目標物体と似た速度を持つが距離が異なっている。また、最後方の人物は目標人物と似た距離を持っているが、速度が異なっている。このように目標物体に似た速度、視差を持つ物体や、目標物体を隠蔽する物体が存在するシーンでも、追跡に成功している。

### 5.3 オプティカルフローと明度一様領域を統合した人間の追跡

オプティカルフローと距離情報は、コントラストがない場合には得られない。よって、前節の方法は目標物体が明度一様領域である場合には追跡に失敗する。そこで、目標物体が明度一様領域の場合も追跡できるように、オプティカルフローと明度一様領域を組み合わせる。目標物体の領域内ではコントラストのある部分でフローが得られ、コントラストのない部分で明度一様領域が得られる。得られた目標のフローまたは明度一様領域のうち、少なくともどちらか一方が背景や他の物体の領域と区別できる場合に目標物体を追跡できる。

オプティカルフローによる目標物体抽出では、5.1節で述べた手法により目標物体領域（目標フロー領域と呼ぶ）を抽出する。しかし、撮像系の光軸方向とほぼ

平行に移動する場合には、フローベクトルは一樣にはならず、目標フロー領域は得られない。

一方、人間に明るさが一樣な部分がある場合、画像中の人間の領域内には明度一樣領域が存在する。この領域を目標明度一樣領域と呼ぶ。人間に明るさが一樣な部分がない場合には目標明度一樣領域は得られない。

目標フロー領域が目標明度一樣領域の少なくとも一方が得られれば人間の領域を推定する。この領域を目標領域と呼ぶ。

### 5.3.1 追跡処理の手順

(1) 初期にはカメラを静止させて、最初に動いた領域をフローで抽出し、その周囲で初期目標領域を検出する。

(2) 前フレームの処理結果と平均速度から、目標フロー領域と目標明度一樣領域それぞれについて現フレームでの位置を予測し、目標フロー領域と目標明度一樣領域の候補領域を決定する。その際、それぞれの領域の前フレームからの大きさの変化を調べ、変化があまり大きくないときには正しく候補領域が得られたとする。また急激に大きくなったときには、似た速度あるいは明度一樣領域を持つ他の物体との重なりがあると考え、候補領域は得られないとする。少なくともどちらか一方の候補領域が得られれば、得られたものすべてを囲む領域として目標領域を設定する。

(3) 前項の処理で目標領域が得られれば、(2)へ戻る。目標領域が得られなかった場合は以下の処理を行う。

- (a) 得られないまま一定数のフレームが経過していれば、(1)へ戻って初期領域の抽出からやり直す。
- (b) 経過していなければ、予測されている目標フロー領域もしくは目標明度一樣領域が得られたものとして、それぞれの予測領域をそのまま目標フロー領域、目標明度一樣領域とし、(2)へ戻る。

### 5.3.2 実時間追跡例

16個のDSPを用いた実時間追跡システムを構築した。図9にDSPボード<sup>23)</sup>の構成を示す。DSP1-9およびDSP13-14はオプティカルフローの計算を、DSP10-12は明度一樣領域の抽出と目標明度一樣領域の決定を、DSP15は目標フロー領域の決定と目標領域の計算を、それぞれ行う。DSP16は表示用画像を生成する。160×120画素の画像を毎秒15フレーム処理することができる。

図10は追跡結果の一例である。図10(a)は目標に似た明度を持つ物体が背景にあり、明度からは区別できないが、(b)では目標フロー領域から人を求

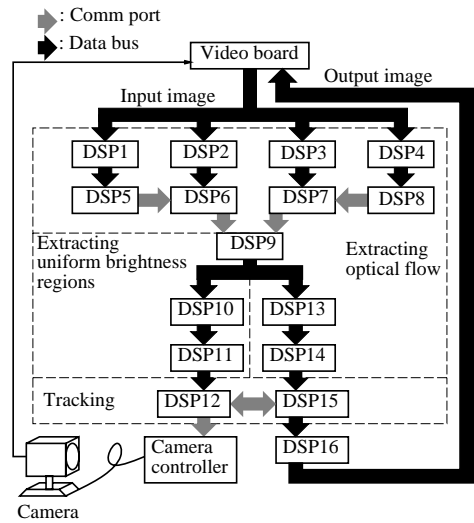


図9 複数のDSPによる並列処理  
Fig. 9 Parallel processing by multiple DSPs.



(a) another object with similar brightness (b) correction by target motion region



(c) another object with similar flow (d) correction by target brightness region

図10 背景の明度あるいはフローが類似の場合の追跡例  
Fig. 10 Another object with similar visual cue exists near the target.

めている。また、図10(c)では目標に似たフローを持つ物体が目標の近傍に存在しており、フローだけでは区別できないが、(d)では目標明度一樣領域を用いて人間の領域を得ている。

## 6. 複数カメラによる追跡

他の人物や物体による目標人物の隠蔽に対処するために、複数カメラを用いた追跡が有効であり、さまざまな手法が提案されている。多くの手法は背景差分によって人間領域を抽出しているが<sup>(24),25)</sup>、背景の変化に対応することが難しく、また画面上で複数の人物が

重なるとそれらを区別することができない。背景差分によって得られた領域から世界座標系での位置を求め、複数のカメラからの投影直線の交点に対する残差が大きい場合に追跡が失敗していると判断する手法もあるが<sup>26)</sup>、半数のカメラが誤って他の物体を追跡すると両方の物体に対する残差が大きくなり、どちらが目標物体か分からなくなる。

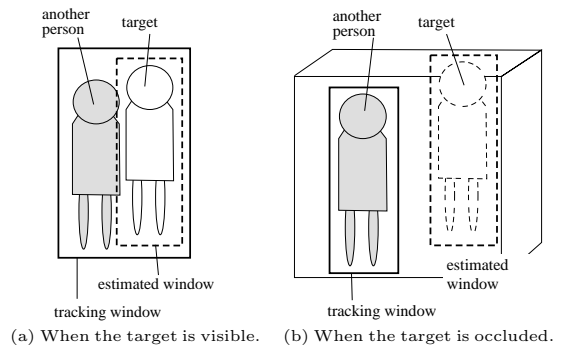
そこで、オプティカルフローに基づいた複数カメラによる人間追跡手法を提案している<sup>27)</sup>。各カメラは5.1節で述べた手法に基づいて独立に目標人物の追跡を行い、必要に応じて複数カメラの情報を統合・伝播させることにより、特定人物の追跡を高信頼で行う。

2つ以上のカメラから目標人物が追跡できているとき、その世界座標系での位置・速度を求めることができる。そして、目標人物が隠蔽されるなどして追跡を行っていないカメラについては、求めた人間の位置においた円筒の人物モデルを画面上に投影した領域を物体の探索領域とし、世界座標系速度を画面上に投影したものを探索速度として、目標人物の検出を試みる。

複数の人物が画面上で重なり合っているとき、それらの見かけの速度が異なればオプティカルフローを用いて区別することは容易である。しかし、類似した速度を持っている場合には目標人物だけを切り出すことは難しく、そのまま追跡を続けると誤った人物領域を切り出してしまい、結果として目標人物の世界座標系での位置・速度の誤差が大きくなる可能性がある。そこで、あるカメラでそのような重なりが起こっている間は、そのカメラによるトラッキングは信頼性が低いとして、世界座標系での位置・速度算出に用いないものとする。そのために、重なり開始と終了を正しく判別する必要がある。

ここでは、他の人物は目標人物あるいは目標人物の探索領域に徐々に近づき、重なり後は徐々に離れていくという仮定をおく。重なり開始は、目標人物(あるいは探索領域)の両側に他の人物を検出するためのウィンドウを設け、そこに類似した速度の物体が入ってきたときに重なりが始まったとする。

重なり終了判定について次の2つの条件を考える。目標人物が見えている場合には、追跡ウィンドウ(tracking window)が2人分の領域を囲むことにより、その大きさが予想される人物領域(estimated window)の大きさよりかなり大きくなる(図11(a)参照)。目標人物が他の物体に隠蔽されている場合には、重なりが終わった時点で追跡ウィンドウと予想人物領域の位置が大きくずれる(図11(b)参照)。これら2つの条件のどちらかが成り立ったときに重なりが



(a) When the target is visible. (b) When the target is occluded.

図11 重なり終了判定

Fig. 11 Termination condition of occlusion.

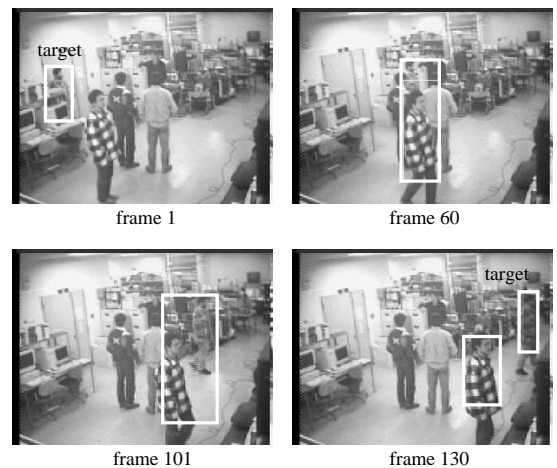


図12 複数カメラによる追跡結果

Fig. 12 Tracking result by multiple cameras.

終了したと判断する。

図12に追跡の実験結果を示す。実験では対象領域を囲むようにほぼ90度ずつ角度を変えた位置に4つのカメラを配置した。図では1つのカメラの画像だけを示している。初期フレームでは目標人物は画像の左上にいる。第60フレームでは目標人物と他の人物が重なり、追跡ウィンドウは両方を囲んでいる。重なり終わりに近づくにつれ追跡ウィンドウの大きさが大きくなり、第101フレームで重なりが終わったと判断した(図11(a)参照)。第130フレームでは対象物体だけを再び追跡している。

## 7. おわりに

さまざまな条件での人の追跡があるが、ここでは著者らが行ってきたいくつかの典型的な研究について述べた。使える情報としては、色、オプティカルフロー、距離、および複数カメラからの情報などがあり、困難な条件ほど多くの情報を必要とする。今後の課題とし



ては以下のようなものがある。

- (1) より多くのセンサ情報の統合(独立でない情報の扱いなどを含む)。
- (2) 複数カメラによる複数の人の追跡のための最適なパン、ティルト、ズームの制御。人の動きは確定的でないので、動きの予測による確率的な最適化となる。
- (3) 多数カメラによる多数の中の特定の人の追跡。あらかじめその人の特徴を抽出し、隠れても可能性のある範囲を探索し、一部現れたときに、その人を識別することにより、追跡を続行する。
- (4) 追跡をしながら追跡対象の特徴の学習

また、処理が複雑になると、現在のコンピュータではオンライン追跡ができない。PC クラスタを用いると、ある程度の処理までは可能である。しかし、複数カメラの場合には全体の装置が大きくなりすぎるのが問題である。

本論文では人の追跡に限定したが、追跡だけでは十分でない。追跡をしながら、顔の認識や人の行動の分析などを行うことも今後の課題である。

#### 参 考 文 献

- 1) 坂井, 長尾, 木戸出: 電子計算機による濃淡図形の処理—顔写真の場合, 電子通信学会論文誌 C, Vol.54C, No.6, pp.445–452 (1971).
- 2) *Proc. IEEE Workshop on Motion of Non-Rigid Articulated Objects* (1994).
- 3) *Proc. IEEE Workshop on Human Motion* (2000).
- 4) 亀田, 美濃, 池田: シルエット画像からの関節物体の姿勢推定法, 電子情報通信学会論文誌, Vol.J79-D-II, No.1, pp.26–35 (1996).
- 5) 岡田, 白井, 三浦, 久野: オプティカルフローと距離情報に基づく3次元運動する人間の追跡, 電子情報通信学会論文誌, Vol.J82-D-II, No.8, pp.1252–1261 (1999).
- 6) Ohno, Y., Miura, J. and Shirai, Y.: Tracking Players and Estimation of the 3D Position of a Ball in Soccer Games, *Proc. 15th ICPR*, pp.145–148 (2000).
- 7) 内田, 三浦, 白井: 雑踏における複数歩行者の追跡, システム制御情報学会論文誌, Vol.14, No.4, pp.180–185 (2001).
- 8) Kitai, K.: Space Activity — Quantitative analysis and evaluation with experiments on the street, *Proc. S&T/SPIE Conf. on Videometrics VI*, pp.268–275 (1999).
- 9) Intille, S.S., Davis, J.W. and Bocick, A.F.: Real-Time Closed-World Tracking, *Proc. CVPR-97*, pp.697–703 (1997).
- 10) Onoguchi, K.: Shadow Elimination Method for Moving Object Detection, *Proc. 14th ICPR*, pp.583–587 (1998).
- 11) McKenna, S.J., Jabri, S., Duric, Z., Rosenfeld, A. and Wechsler, H.: Tracking Groups of People, *Computer Vision and Image Understanding*, Vol.80, No.1, pp.42–56 (2000).
- 12) Uchida, K., Shirai, Y. and Shimada, Y.: Probabilistic Method of Real-Time Person Detection Using Color Image Sequences, *Proc. IEEE/RSJ IROS2001*, pp.1983–1988 (2001).
- 13) 和田, 松山: 動的背景モデルを用いた移動領域の抽出, 情報処理学会第49回全国大会講演論文集(2), pp.141–142 (1994).
- 14) Nakai, H.: Non-parameterized bayes decision method for moving object detection, *Proc. ACCV-95*, pp.447–451 (1995).
- 15) 長井, 久野, 白井: 複雑変動背景下における移動物体の検出, 電子情報通信学会論文誌, Vol.J80-D-II, No.5, pp.1086–1095 (1997).
- 16) Iketani, A., Nagai, A., Kuno, Y. and Shirai, Y.: Detecting persons on changing background, *Proc. 14th ICPR*, pp.74–76 (1998).
- 17) Mae, Y., Yamamoto, S., Shirai, Y. and Miura, J.: Optical Flow Based Object Tracking by Active Camera, *Proc. 2nd Japan-France Congress on Mechatronics*, pp.545–548 (1994).
- 18) Yamamoto, S., Mae, Y., Shirai, Y. and Miura, J.: Realtime Multiple Object Tracking Based on Optical Flows, *Proc. ICRA-95*, pp.2328–2333 (1995).
- 19) Coombs, D. and Brown, C.: Real-time Smooth Pursuit Tracking for a Moving Binocular Robot, *Proc. CVPR'92*, pp.23–28 (1992).
- 20) Uhlin, T., Nordlund, P., Maki, A. and Eklundh, J.-O.: Towards an Active Visual Observer, *Proc. 5th ICCV*, pp.679–686 (1995).
- 21) 岡田, 白井, 三浦, 久野: オプティカルフローと距離情報に基づく動物体追跡, 電子情報通信学会論文誌, Vol.J80-D-II, No.6, pp.1530–1538 (1997).
- 22) 山根, 白井, 三浦, 久野, 島田: オプティカルフローと明度一様領域を統合した人間の実時間追跡, 日本ロボット学会誌, Vol.18, No.4, pp.521–528 (2000).
- 23) 白井良明: 画像処理ボード, 日本ロボット学会誌, Vol.16, No.5, pp.44–46 (1998).
- 24) Rao, B.S. and Durrant-Whyte, H.: A Decentralized Bayesian Algorithm for Identification of Tracked Targets, *IEEE Trans. on SMC*, Vol.SMC-23, No.6 (1993).
- 25) Utsumi, A., Mori, H., Ohya, J. and Yachida, M.: Multiple Human Tracking using Multiple Cameras, *Proc. 3rd IEEE Int. Conf.*

on *Automatic Face and Gesture Recognition*, pp.498-503 (1998).

- 26) Matsuyama, T.: Cooperative Distributed Vision - Dynamic Integration of Visual Perception, Action, and Communication, *Proc. Int. Workshop on Cooperative Distributed Vision*, pp.1-39 (1998).
- 27) Tsutsui, H., Miura, J. and Shirai, Y.: Optical Flow-Based Person Tracking by Multiple Cameras, *Proc. MFI-2001*, pp.91-96 (2001).

(平成 14 年 1 月 17 日受付)

(平成 14 年 3 月 8 日採録)

(担当編集委員 八木 康史)



白井 良明(正会員)

1964 年名古屋大学工学部機械工学科卒業．1969 年東京大学大学院工学系博士課程修了，工学博士．同年電子技術総合研究所入所．1985 年制御部部长．コンピュータビジョン，ロボティクスの研究に従事．1971～1972 年，MIT AI ラボ客員研究員．1988 年大阪大学工学部電子制御機械工学科教授．



三浦 純(正会員)

1984 年東京大学工学部機械工学科卒業．1989 年同大学院工学系研究科情報工学専攻博士課程修了，工学博士．同年大阪大学助手．現在同大学院工学研究科電子制御機械工学専攻助教授．知能ロボット，人工知能，コンピュータビジョンの研究に従事．1994～1995 年，CMU 客員研究員．1997 年日本ロボット学会論文賞受賞．