

## 複雑背景における人の追跡

白井良明

大阪大学大学院工学研究科電子制御機械工学専攻  
〒565-0871 大阪府吹田市山田丘2-1

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伴う場合に関しては、オプティカルフロー、距離、明るさなどを組み合わせる方法、および複数カメラの協調による追跡を述べる。

## Human Tracking in Complex Environment

Yoshiaki Shirai

Dept. of Computer-Controlled Mechanical Systems, Osaka University  
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

This paper describes methods of human tracking in a variety of conditions. Using a stationary camera, the following tracking 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 persons in a background which contains random motions. In case of human tracking with an active camera systems, integration of optical flow, range data, and uniform brightness regions are described. Finally coordination of multiple cameras are described for tracking a person with heavy occlusions, and for identifying persons.

### 1 はじめに

人の追跡は、人のモニター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なわれているが、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、多数の研究がある。

人の追跡を一般の動物体追跡と比較したときの特徴は、

- 回転を伴い、見え方が変わる（自動車では緩やかな回転のみ、ボールは見え方が一定）
- 非剛体、関節物体である。

動画像を用いる人の追跡は、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非常に簡単で、すでに実用化しているものもある。しかし、人の切り出しが難しい場合には、以前として困難な問題が残されている。

その主な追跡の条件は以下のようなになる。

- 対象：複数か？人の重なりは？物体による隠れは？衣服の色の制限は？
- 動き方：直線運動か？面内運動か？関節は動くか？

- 背景：テクスチャは一樣か？ 距離は連続か？ 時間的に変化するか？
- カメラ：固定か、回転運動か、移動か？ 複数か？
- タスク：複数を同時追跡か？ オンラインか？ 姿勢も必要か？

これらの条件は必ずしも独立ではない。例えば、姿勢が必要な場合は、関節は動く。しかし、逆は真でなく、単に人を追跡するだけで、姿勢が不要でも、歩いている人の足や手の関節は動く。この場合、人を観測する分解能が高いと、関節の動きはかえって追跡の妨げになる。

多くの研究はいくつかの困難な条件を扱い、それ以外の条件は簡単化している。例えば、人や手指の姿勢を追跡する場合は、対象の切り出しが簡単にできるという条件である。また、オンライン追跡を行なう場合は、カメラが固定か、カメラモデルが簡単で背景差分が使えるか、人の重なりがないかなどの条件がつけられている。

ここでは、関節物体の姿勢を求める問題（例えば [1]）は除き、種々の条件での人の追跡を述べる。

## 2 雑踏における複数歩行者の追跡

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るまい（生成・移動・消滅など）の解析 [3] が行われている。都市の雑踏が数時間に渡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、歩行者追跡ツールの開発の必要である

このタスクの条件は、カメラ固定、背景ほぼ一定、オフラインであるが、不特定多数の人を同時に追跡しなければならない、

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三人以上が一組に重な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。



図 1: 歩行者天国の雑踏（鍛佳代子（東京工芸大学）の提供）

人の候補領域は処理が簡単な背景差分法で抽出する。雑踏では、画像中には常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）を背景とする。

全画像を通して人の領域から、大きさと形状によって単独に分離している歩行者のみを検出する。ここ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、どこかで分離していれば追跡できる。

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 [4]。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保持して追跡を続け、誤った候補を排除する。このようにして信頼性の高い追跡を実現している [2]。

## 3 背景が歩行者の影響を受ける場合の実時間追跡

普通の背景差分法は、影を人の一部として検出する。廊下では、人の長い影が床や壁に

投影され、影が人の一部として含まれると、人の形の推定が正確でなくなる。

もし高さ情報が得られれば、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 [5].

影の領域でも、色度に大きな変化がないと仮定する手法もある [6]. 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。

本研究では、学習段階で、対象物体や影の領域を標本にして、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。さらに、テンプレートの顔の領域から人の顔を探すことができる。

### 3.1 ピクセルの分類

各ピクセルを影のない背景、影、前景に分類したい。さらに、前景の領域において人の顔の領域を抽出する必要がある。そこで、ピクセルを次の4つのクラスに分類する。

背景  $\omega_b$

人が投じる影  $\omega_s$

人の肌色  $\omega_h$

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  $C = (R, G, B)$  を取るピクセルがクラス  $\omega$  に属する確率は、

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

事前確率  $P(\omega)$  は等確率とする。尤度  $P(C|\omega)$  は学習段階で求める。

### 3.2 物体の尤度

背景の尤度は、誰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  $R, G, B$  はガウス分布にしたがうとし、平均と分散  $\{\mu_R, \mu_G, \mu_B, \sigma_R^2, \sigma_G^2, \sigma_B^2\}$  を求め

る。背景の尤度  $P(C|\omega_b)$  を以下によって求める。

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

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

for  $c = R, G, B$

肌色の尤度も実画像から求める。画像の肌領域において、画素値  $C$  のヒストグラム  $H_h(C)$  を作成する。

明るさ方向の標本は撮影を行なった環境の照明条件に左右されるので、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  $P(I|\omega_h)$  は一様分布で与える。

一方、ある明度における色の確率は

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

影の確率分布を作成するため、類似の色の領域を一つのグループとして扱う。各グループについて、様々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、 $P(C|\omega_s, I)$  をヒストグラムから求める。 $P(I|\omega_s)$  は、影が出来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基にして人らしさの度合いを評価する。テンプレートの評価は画像の手前に映る歩行者から行う。

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。例えば、

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

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、 $f_t = f_0 \times f_1 \times f_2$

種々の衣服の色の人の実時間追跡実験を行った。図2に例を示す。

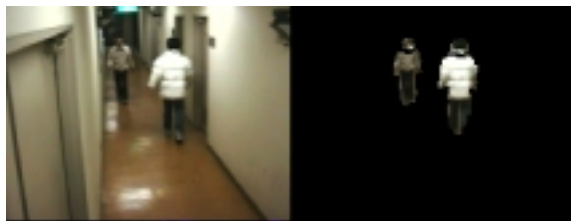


図 2: 人の追跡例

#### 4 背景に動きがある場合の人の追跡

背景に動物体がある場合に移動する人の検出を行なう場合、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。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の区別をする [7].

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、均一なフローを持つ領域を移動物体の候補領域として抽出する。そのために、位置  $x-y$ 、時間  $t$ 、移動方向  $\theta$  の 4 次元からなる投票空間を用いる。各時間の均一なフロー領域に対し、等速直線運動をすると仮定して、未来の時間に対して投票空間に投票する。これを続けると、ほぼ等速直線運動する領域の投票値が高くなる。ある程度の大きさの速度場領域が一定しきい値を越えるのたら人が検出されたとする。図3に時系列画像の1フレームを示す。



図 3: 原画像と検出結果の例

#### 5 カメラが動く場合

カメラが動く場合は背景差分が使えないので、それ以外の手がかりが試みられている。

まず、目標物体の見え方が一定ならば、画像中の領域の輝度分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明るさや形状が変化する場合には、追跡が困難となる。

距離情報を用いる手法 [8] 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。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。

速度情報に基づく手法 [9] も数多く提案されている。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を行なうことができる。しかし、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。

そこで、複数の情報を用いて目標物体領域を抽出し、追跡を行う手法も提案されている。例えば、速度情報と色情報、速度情報と距離情報 [10] 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、追跡をおこなうことができない。

追跡されている物体同士の、画面上での領域の重なりによって隠蔽を検出する手法 [11] が提案されている。目標物体を隠蔽する物体も追跡されている必要があるので、追跡されていない物体が目標物体を隠蔽する場合に

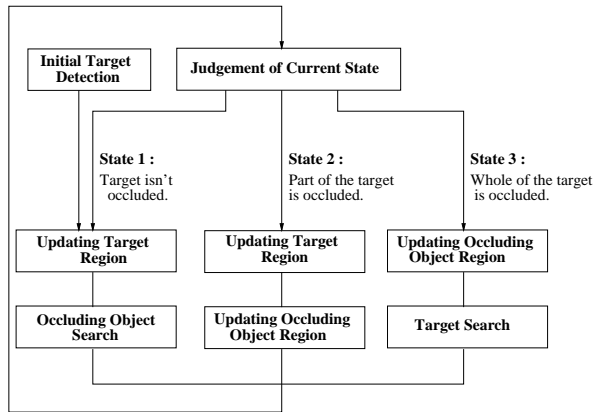


図 4: 追跡の概要

は、隠蔽を検出できない。

### 5.1 オプティカルフローと距離を統合する方法

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 [12].

各フレームでは、まず現在の右画像と 1 フレーム前の右画像を用いてオプティカルフローを、左右画像から視差画像を抽出する。目標物体の追跡処理は、目標物体のがどのように隠蔽されているかによって 3 つの状態に分かれている。各フレームで現在の状態の判断を行ない、それぞれの状態によって異なる処理が行われる (図 4).

初期には、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 (State 1) であるとして、追跡を開始する。

State 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。

State 1 において、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 (State 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、State 1 に戻

る。

State 2 の次のフレームで、目標物体領域がなくなった場合、目標物体の全体が隠蔽された状態 (State 3) とし、隠蔽物体領域の更新と、隠蔽物体領域周辺で目標物体の探索を行う。目標物体領域が検出されれば、State 1 に戻る。

目標物体が隠蔽されていない状態での領域の更新は以下のように行なう。

右画像上の点  $\mathbf{p}_i$  で観測  $\mathbf{o}_i = (u_i, v_i, d_i)$  があった時その点が目標物体 ( $T$ ) に属する確率を計算して、その確率が高い点を目標物体領域とする。目標物体上の点で  $\mathbf{o}_i$  が観測された場合の確率は、

$$P(\mathbf{p}_i \in T | \mathbf{o}_i) = \frac{P(\mathbf{o}_i | \mathbf{p}_i \in T)P(\mathbf{p}_i \in T)}{P(\mathbf{o}_i)} \quad (1)$$

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が観測された点を隠蔽物体領域とする。

本手法で追跡を行なった結果を図 5 に示す。1 フレーム目の黒い長方形は、初期フレームで物体領域を初期化するとき与えた、物体領域を囲む四角形である。27 フレーム目で隠蔽物体が検出され、追跡を開始し、37 フレーム目で目標物体は完全に隠蔽されて見えなくなっている。42 フレーム目で目標物体が再び検出され、追跡を再開している。

なお、姿勢を考慮した追跡は、文献 [13] を参照。



No. 20

No. 40



No. 50

No. 70

図 5: 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)

## 5.2 オプティカルフローと明度一様領域を統合した人間の長時間追跡

### 5.2.1 概要

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、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。本手法は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することができる。

### 5.2.2 追跡

人間が撮像系の光軸方向とある程度の角度をもって移動する場合、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。このようなフローがほぼ一樣な領域を「目標フロー領域」と呼ぶ。

撮像系の光軸方向とほぼ平行に移動する場合、画像中の人間の領域はあまり移動せず拡大又は縮小するだけなので、フローベクトルが一樣になることはほとんどない。この場合、目標フロー領域は得られない。

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様領域が存在する。この領域を「目標明度一様領域」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様領域は得られない。

目標フロー領域か目標明度一様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。

追跡の手順は以下のものである。

(1) 初めに画像を入力し、フローを手掛かりに初期の目標フロー領域と目標明度一様領域を抽出し、初期の目標領域を決定する。目標フロー領域と目標明度一様領域の少なくとも一方が得られていれば、得られた領域の次フレームでの領域を予測する。

(2) 新しく画像を入力して、予測された領域を基に目標フロー領域と目標明度一様領域の候補領域を決定する。目標フロー領域と目標明度一様領域のうち少なくとも一方が得られれば、得られた領域から目標領域を決定する。両方とも得られない場合、目標領域は得られない。

(3) A. 目標領域が得られていれば再び予測を行ない、(2)へ戻る。

B. 目標領域が得られなかった場合、

B-1. 得られないまま一定数のフレームを経過していれば、(1)へ戻って、初期の領域の抽出から追跡をやり直す。

B-2. まだ数フレーム経過していない場合、予測されている目標フロー領域もしくは目標明度一様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

明るさ一様領域とし、(2)へ戻る。

なお、詳細は[14]参照。

### 5.2.3 実時間追跡例

16個のDSPを用いた実時間追跡の例を図6に示す。同図(a)は目標に似た明るさを持つ

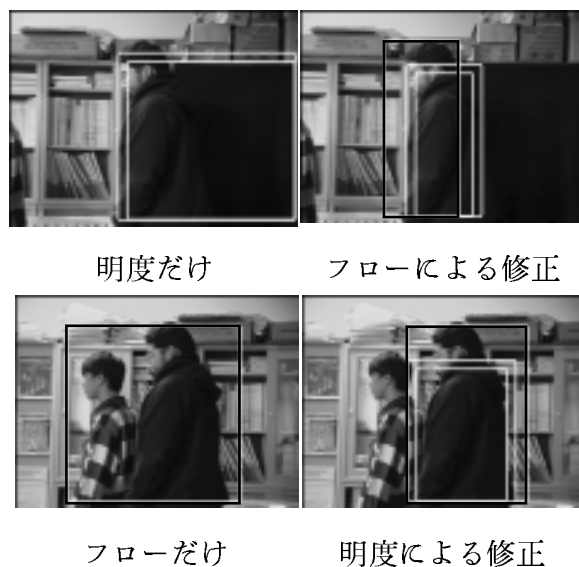


図 6: 背景の明るさあるいはフローが類似の場合の追跡例

物体が背景にあり、明るさ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めている。また、同図(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)では目標明るさ一様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

複数のカメラを用いた追跡についても様々な手法が提案されている。その一つに複数のカメラを使い、背景差分により追跡された人間の位置を求める手法がある[15]。これは背景差分を用いているので、背景の変化に対応することが難しい。また複数のカメラで、背景差分によって抽出した物体領域から世界座標系位置を求めることによる追跡[16]もあ

る。この時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ものは、追跡を失敗していると判断している。半数のカメラが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がわからなくなる。

そこで、オプティカルフローに基づき、複数のカメラを用いて人間を追跡し、1つのカメラの画面上で、目標物体と類似した速度をもつ他の物体が近くにある場合、他のカメラからの情報と、抽出される物体領域の大きさ・位置を考慮する方式をとった。

2つ以上のカメラから目標物体が追跡できている時、人間の世界座標系位置・速度を求めることができる。人間が障害物に隠されて観測できないカメラに対して、求めた人間の位置に置いたモデルの円筒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する。

あるカメラにおいて人間が正しく観測されていないと判断する基準を3つ設ける。1つはオプティカルフローが観測されない時で、他の人や段ボール、動かない障害物に目標物体が隠れた考えられる。

2つ目は探索領域と観測される物体の大きさが大きく異なる場合である。これは目標物体の速度と似た速度をもつ他の物体が近傍にいるため、それらを区別できなかったと考えられる。

3つ目は探索領域と観測される物体の位置が大きく異なる場合である。これは目標物体を追跡している他のカメラと異なる物体を追跡していると考えられる。

これらの場合目標物体の追跡に失敗していると判断し、他のカメラからの情報を基に求めた画面上での探索範囲において探索速度に類似したフローを持つ領域を探す。

詳細は[17]を参照。

## 7 おわりに

今後、以下のような課題がある。

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. 追跡をしながら対象の特徴の学習
3. 複数カメラの注目方向の最適計画（例えば， [18] の試みは未完成）

## 参考文献

- [1] 島田他： 確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定，信学論，D-II Vol.J79-D-II,No.7,p.1210-1217,1996
- [2] 内田他： 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
- [3] K. Kitai: Space Activity - Quantitative analysis and evaluation with experiments on the street, Proc. S&T/SPIE Conf. on Videometrics VI, pp. 268-275, 1999
- [4] S. S. Intille et al: Real-Time Closed-World Tracking, Proc. CVPR, pp. 697-703, 1997
- [5] K.Onoguchi: Shadow Elimination Method for Moving Object Detection, Proc. ICPR, pp.583-587, 1998
- [6] S.J. McKenna et al: Tracking Groups of People, Computer Vision and Image Understanding, Vol. 80, No. 1, pp. 42-56, 2000
- [7] A. Iketani et al: Detecting persons on changing background, Proc. 14th ICPR, pp.74-76, 1998
- [8] D. Coombs et al: Real-time Smooth Pursuit Tracking for a Moving Binocular Robot, Proc. CVPR'92, pp. 23-28, 1992
- [9] Y. Shirai et al: Moving Object Perception and Tracking by Use of DSP, Proc. of CAMP, pp. 251-256, 1993
- [10] T. Uhlin et al: Towards an Active Visual Observer, Fifth ICCV, pp. 679-686, 1995
- [11] S. Yamamoto et al: Realtime Multiple Object Tracking Based on Optical Flows, Proc. ICRA, pp.2328-2333, 1995
- [12] 岡田他： オプティカルフローと距離情報に基づく動物体追跡，信学論， Vol. J80-D-II, No. 6, pp. 1530-1583, 1997
- [13] R. Okada et al: Tracking a Person by Integrating Optical Flow and Depth, Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 336-341, 2000.
- [14] 山根他： オプティカルフローと明度一様領域を統合した人間の長時間追跡，ロボット学会誌， Vol.18, No.4, pp.521-528, 2000
- [15] B.S. Rao et al: A Decentralized Bayesian Algorithm for Identification of Tracked Targets, IEEE Trans. Vol. SMC-23, No. 6, 1993
- [16] T. Matsuyama et al: Cooperative Distributed Vision, Proc. Int. Workshop on Cooperative Distributed Vision, pp. 1-39, 1998
- [17] H. Tsutsui et al: Optical Flow-Based Person Tracking by Multiple Cameras, Proc. MVA, pp. 418-421 1998
- [18] 山根 他： 複数カメラを用いた人物の追跡と認識，ROBOMECH'00, p. 90, 2000年