

## 映像監視における人物位置行動の検出技術

林 健太郎<sup>†</sup> 羽 下 哲 司<sup>†</sup>  
関 真規人<sup>†</sup> 笹 川 耕 一<sup>†</sup>

本論文では様々な監視領域に存在する人物の位置と行動の検出技術について述べる。ここでは以下に示す3つの目的に分けて考える。まず位置検出が重要でない程度の領域を小領域と定義し、小領域の場合は行動検出に重点を置く。次に位置検出が必要となる領域、すなわち小領域よりも大きい領域を大領域と定義し、大領域における位置検出を行う。さらに、大領域において通常人物が小さく映るので、ポーズや見え方から行動検出するのは困難であるため、時系列の人物位置から行動検出する。上記3つの目的を達成する手法をそれぞれ“小域行動検出手法”、“大域位置検出手法”、“大域行動検出手法”と呼び、これらを実現すれば幅広い人物監視応用に対応できる。本論文ではそれぞれに対応する手法を述べる。その後、より広い応用範囲を実現するための将来展望について述べる。

## Detecting Human Position and Action in Surveillance Scenes

KENTARO HAYASHI,<sup>†</sup> TETSUJI HAGA,<sup>†</sup> MAKITO SEKI<sup>†</sup>  
and KOICHI SASAKAWA<sup>†</sup>

This paper presents our methods for detection, tracking, and action detection from a surveillance video camera. First we classify our methods by observation size. When observation size is small, we do not need to detect human position, but need to detect human action. When size is large, first we need to detect human position. Then, we detect human action using its positions. We describe our methods corresponding to above each classes. All methods have enough performance for real-use. Moreover, we describe other methods such as a tracking with a stereo vision, which mitigates performance limitation of above methods. At last we show the vision of our research.

### 1. はじめに

安全、安心な社会を築くために、人物の侵入や行動を自動監視する映像監視技術が求められており、近年その重要性が増している。

映像監視において、あらゆる人物の位置と行動を検出することを目的とすると、以下の条件を考慮する必要がある。

- (a) 照明条件の変化
- (b) 監視領域内の存在人数
- (c) 監視領域の大きさ

(a)の照明条件の変化に対処するための手法は数多く存在する。背景差分だけでも、エッジやテクスチャに着目して照明変化にロバストな方法<sup>12)</sup>から、あらかじめいくつかの照明条件での画像を記憶しておく方法<sup>1),5)</sup>などがあり幅広い。本論文では従来の手法を適

切に利用し、その詳細については省略する。また (b)の存在人数の幅も広く、無人状態に1人で侵入する場合から街の雑踏のような多人数までである。本論文では1~数名までの小人数であると仮定し、人の密度は一定以下とする。本論文では残りの(c)の課題に取り組む。

本論文でのアプローチを図1に示す。まずカメラ設置位置によって監視領域を分類するが、本論文では単眼カメラを想定しており、カメラ近傍の小領域か、遠方の大領域かの2つに単純に分類できる。その後それぞれの場合で人物の位置、行動を検出するが、小領域の場合はそもそも領域が限られているので位置を検出する必要が少ない。したがって行動認識に集中でき、結果として処理を単純にすることができる。我々はこれを利用し、閉空間での行動検出手法を提案している<sup>6)</sup>。逆に大領域、特に屋外を見る場合では位置検出が重要であり、屋外侵入監視手法を提案している<sup>8)</sup>。さらに屋外の人物の行動を自動認識するニーズが高いが、大領域の場合、画像上の人物のサイズが小さく姿

<sup>†</sup> 三菱電機株式会社先端技術総合研究所  
Advanced Technology R&D Center, Mitsubishi Electric Co.

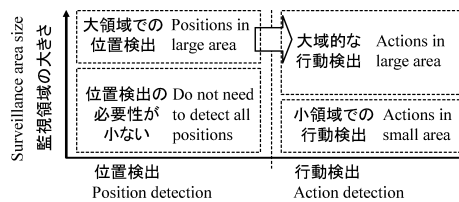


図1 監視領域の大きさと、位置行動検出アプローチの違い

Fig.1 Various approaches against surveillance area sizes.

勢や見え方から行動を認識するのは困難である。そのため、人物の検出位置を用いて特定の行動を検出する手法を提案している<sup>13)</sup>。上記のアプローチにより幅広い課題を解決することができる。以上の手法を“小域行動検出手法”、“大域位置検出手法”、“大域行動検出手法”と呼び、以下これらの内容について述べる。

小域行動検出手法ではエレベータのかごや、壁で仕切られた部屋などの閉空間を仮定している。これにより屋外のような大きな照明変化を扱わずに済む。また、上記閉空間は人物位置の検出が必要とされないほど十分小さいとする。以上により個々の人物を分離する必要がなくなり、マクロの手法で行動を検出することができる。実際の応用では人物が非定常動き状態にあるかないかの二値識別問題であったため、たとえば動画像の時空間エッジ特徴を DP マッチングする方法<sup>3)</sup>や、画像上のオプティカルフローにより動き特徴を抽出し、隠れマルコフモデル (HMM) で認識を行った研究<sup>11)</sup>のような多クラス識別手法は必要ない。我々はフローと物体領域 (複数人物を含む) の両方の情報から、領域中の動きベクトルの大きさと方向の分散を計算し、“暴れ度”と呼ぶ独自の評価値を用いて小さな計算量で二値識別問題を解決する。エレベータへの適用率を高めるために“暴れ度”を自動的に正規化する方法も述べる<sup>6)</sup>。

大域位置検出において、従来から、背景差分法により抽出した変化領域を人の候補として一定時間以上追跡し、人の動きが時空間的に連続していることを用いて人とそれ以外の誤報要因とを区別する手法<sup>2),7)</sup>がよく用いられるが、重心や面積といった単純な特徴量による変化領域のフレーム間での関連付けは誤対応が生じやすく、誤報や失報を生じることが多い。これは背景差分だけでは時空間特徴量が正しく把握できないためと考えられる。したがって最初から時空間特徴量を用いれば性能を向上させることができるという考えがあり、複雑背景下で歩行者を抽出する様々な手法が提案されている<sup>4),10)</sup>。我々はさらに歩行者の持つフローの特徴を詳細に調べ、空間的な動きの平均、時間的な動きの平均、およびその分散からなる3つの特徴から

構成される特徴空間で、移動する人とそれ以外の背景の変動とを精度良く識別する手法を提案している<sup>8)</sup>。またアクティブカメラを用いた人物追跡手法<sup>9)</sup>では、人物が固定カメラの監視範囲を越えて移動した場合でも、それらを個別に精度良く追跡し、その位置を求めることが可能である。

さらに大域行動検出手法では、大域位置検出で得られた各時刻、各人物の位置をもとに行動を解析し、事前に与えられた条件を満たす行動を検出することができる。ここでいう行動とは、人物が画像上の特定の領域に侵入するような単純な場合から、まっすぐ歩くべきところをふらふらしながら歩くような場合までを含む。また検出した時刻と位置をトリガとして画像記録装置にイベントを同期記録させたり、リアルタイムに行動を検出して発報させたりすることもできる。

上記の手法は単眼固定カメラを用いたものであり、汎用性が高い手法といえる。しかし今後は単眼固定カメラにこだわることなく、様々な形態のカメラを用いて、より多様な条件下で人物の行動を認識・追跡できる手法を開発することが重要になると考えられる。最後に我々は、より高度な行動認識、より多人数を追跡、より広範囲での人物追跡の3つの視点から、我々の研究事例を交えながら将来展望について述べる。

以下では、小域行動検出手法、大域位置検出手法、大域行動検出手法、将来展望について述べる。

## 2. 小域行動検出手法

エレベータかご内での犯罪シーンを例題としてとりあげ、小さな閉空間における“暴れ度”に基づいた非定常な行動の検出技術について述べる。

図2はエレベータかご内における平常時と犯罪時の映像である。このとき画像上の人物領域で観測されるオプティカルフローは、その向き(8方向)と大きさ(4段階)に関し、図3に示すような分布を形成する。図を見ると、平常時と犯罪時ではばらつき度合いが異なることが分かる。

また、各画像におけるフローのばらつき度合いを次式により数値化し、複数枚分をプロットしたものを図4に示す。縦軸がフローの向きのばらつき度合い  $V_d$ 、横軸がフローの大きさのばらつき度合い  $V_m$  である。ここで  $N$  は観測点の数、 $D_i$  および  $M_i$  は各点でのフローの向きと大きさであり、 $\bar{D}$  と  $\bar{M}$  はその平均値を表す。図を見ると、犯罪時には、平常時に比べ、向きのばらつきと大きさのばらつきがともに大きい場合があることが分かる。このような傾向は他の犯罪映像に対しても見受けられる。

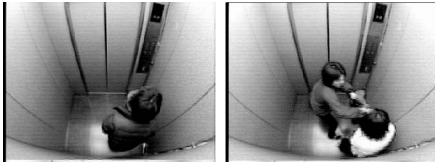


図 2 エレベータのかご内映像例 (左: 平常時, 右: 犯罪時)  
Fig. 2 Video image in an elevator (Left: usual, Right: at the time of a crime).

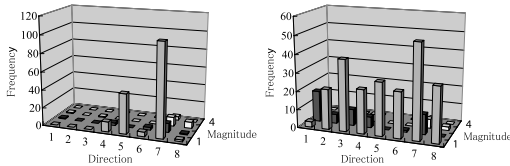


図 3 フローの分布例 (左: 平常時, 右: 犯罪時)  
Fig. 3 Distribution of flow vectors (Left: usual, Right: at the time of a crime).

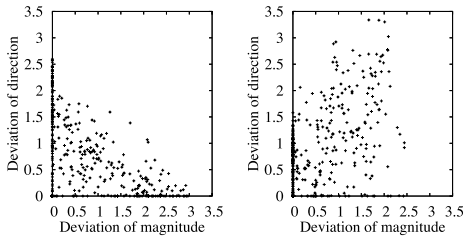


図 4 フローのばらつき (左: 平常時, 右: 犯罪時)  
Fig. 4 Deviation of flow vectors (Left: usual, Right: at the time of a crime).

$$V_d = \left\{ \frac{1}{N} \sum_{i=1}^N (D_i - \bar{D})^2 \right\}^{\frac{1}{2}} \quad (1)$$

$$V_m = \left\{ \frac{1}{N} \sum_{i=1}^N (M_i - \bar{M})^2 \right\}^{\frac{1}{2}} \quad (2)$$

そこで、我々は、この特性に着目し、フローのばらつきというマクロな指標により、犯罪のような非定常な人物行動を検出する。アルゴリズムを以下に示す。

- (1) 背景差分法<sup>12)</sup>により、観測画像中の変化領域を人物領域として抽出する。
- (2) 人物領域中の各点で、SAD ブロックマッチングによりフローベクトルを求める。その際、マッチングレベルの低いフローは破棄する。
- (3) 求めたフローベクトルに対し、向きのばらつき度合い  $V_d$ 、大きさのばらつき度合い  $V_m$  を求める。また、絶対的な動きの量に制限を設けるために、フローの個数  $Q$  を求める。
- (4) 3つの統計量の積  $G = V_d \times V_m \times Q$  (“暴れ度”と呼ぶ) に対し、閾値処理を行う。ここでは、

$G$  が高い場合に、非定常な行動状態と見なす。

検出環境をエレベータかごに限ったとしてもその意匠は様々であり、壁の色・模様、硝子張りや窓の有無、照明の明るさ、容量の違いがある。上記暴れ度は、こういった環境の違いに依存した値となっている。より汎用性の高い指標とするためには、この環境依存度を軽減させる必要がある。そこで、本論文では、過去の観測事例をもとに暴れ度を正規化する。

過去に得られた観測画像のうち、有人状態の画像  $N$  枚から暴れ度  $G(t)$  ( $t = 1, \dots, N$ ) を求め、その平均値  $\bar{G}$  と標準偏差  $\sigma_G$  を算出する。このとき、非定常な行動の発生確率が十分に小さいと仮定すれば、暴れ度は次式により正規化することが可能となる。なお、各画像における人物の有無は背景差分結果により判断する。

$$G' = \frac{G - \bar{G}}{\sigma_G} \quad (3)$$

実験では、意匠の異なる 5 種類のエレベータ映像に対し、本手法は、過剰検出 6% のもて 80% の非定常な行動 (暴れ) を正しく検出できている。

### 3. 大域位置検出手法

#### 3.1 固定カメラによる屋外環境での侵入者検出

屋外の侵入監視のように、広い監視範囲のどこから出現するか予測できない侵入者をリアルタイムで検出するシステムでは、監視視野を検出対象である人と比べて大きく設定するため、日照変動や木の揺れ、光の乱反射などの多くの外乱がカメラ映像に取り込まれる。このため、これら外乱と侵入者との識別が重要となる。

詳細は論文 8) を参照されたいが、屋外環境において、空間平均した動きの強さや、時間的な動きの一様性には人の動きと背景の変動の違いが見られることが分かっている。大まかな傾向として、人の動きでは他の背景変動と比較して、

- 空間平均した動きの強さ ( $F_1$ ) は高い値
- 時間平均した動きの強さ ( $F_2$ ) は高い値
- 各時刻の空間平均した動きの強さはいずれも高く、途中で動きゼロにならない。すなわち、時間的な動きの一様性 ( $F_3$ ) が高い値

という傾向があることが確認できた。これらの結果から、 $F_1, F_2, F_3$  の 3 種類の動きの特徴を用いて多次元的に判定することにより、より精度良く侵入者と背景変動を弁別して抽出することができるものと推測される。

空間平均した動きの強さ  $F_1$  は、変化領域内で求めた局所的な相関値マップのうち、ゼロでないものの

ピーク位置がどれだけまとまりが良いかを表す尺度である。

背景差分により検出した領域  $R_j$  ( $j = 0, 1, \dots, N_R - 1$ ) の中に割り当てた  $w_x \times w_y$  画素の各相関ブロック  $B_i$  ( $i = 0, 1, \dots, N_B - 1$ ) の位置において、輝度差分の絶対値総和に基づく相関演算を行う。

$F_1$  は、このとき得られる相関値マップ  $S_{B_i}(u, v)$  のうち、動きがゼロでないものを累積し、正規化した相関値マップ

$$S_{R_j}^*(u, v) = \frac{1}{N_{R_j}} \sum_{B_i \in R_j} \{W_{B_i} \cdot S_{B_i}(u, v)\} \quad (4)$$

のピーク形状の急峻さの尺度を用いて、以下のように定義することができる。

$$F_1 = S_{R_j}^*(0, 0) - S_{R_j}^*(u_{min}, v_{min}) \quad (5)$$

ただし  $(u_{min}, v_{min})$  は変化領域の代表速度 ( $= \arg \min_{(u, v)} S_{R_j}^*(u, v)$ ) である。

時間平均した動きの強さ  $F_2$  は、変化領域を追跡する過程で得られる、空間的に累積した相関値マップのうち、原点でない相関値マップのピーク位置がどれだけまとまりが良いかを表す尺度である。

求められた変化領域の代表速度  $(u_{min}, v_{min})$  を用いて、領域  $R_j$  に属する全ブロック  $B_i$  ( $i = 0, 1, \dots, N_B - 1$ ) を  $(u_{min}, v_{min})$  だけ移動させた位置において、時刻  $t - 1, t - 2$  の画像間で、相関値マップを計算する。以後同様に、各時刻の代表速度を用いて、次は時刻  $t - 2, t - 3$  の画像間、 $\dots$  という形で、順次時間軸を遡る方向に追跡処理を全  $L$  回繰り返す。

$F_2$  は、 $k$  回目の追跡で得られる相関値マップ  $S_{R_j}^{*(k)}(u, v)$  のうち動きがゼロでないものを累積投票し、正規化した相関値マップ

$$S_{R_j}^{**}(u, v) = \frac{1}{N} \sum_{k=0}^{L-1} \{W_{R_j}^{(k)} \cdot N_{R_j}^{(k)} \cdot S_{R_j}^{*(k)}(u, v)\} \quad (6)$$

のピーク形状の急峻さの尺度を用いて、以下のように定義することができる。

$$F_2 = S_{R_j}^{**}(0, 0) - S_{R_j}^{**}(u_{min2}, v_{min2}) \quad (7)$$

ただし、 $W_{R_j}^{(k)}$  は  $S_{R_j}^{*(k)}(u, v)$  が  $(0, 0)$  で最小となる場合は 0、それ以外は 1 とする重みで、 $(u_{min2}, v_{min2}) = \arg \min_{(u, v)} S_{R_j}^{**}(u, v)$  である。

時間的な動きの一様性  $F_3$  は、追跡過程で得られる変化領域の代表速度全体のうち、ゼロでない動きベクトルが占める割合を表す尺度であり、追跡過程における、前述の空間平均した動きの強さがしきい値を超えたもの数で以下のように定義することができる。

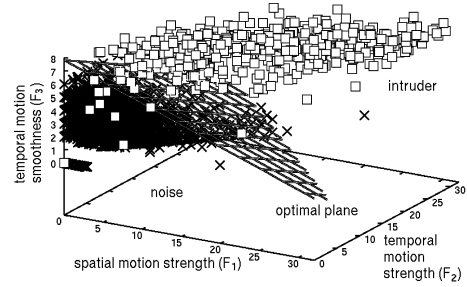


図 5 特徴空間  $F_1, F_2, F_3$  における人と誤報要因の分布、最適判別平面 (格子で表示)

Fig. 5 Feature distribution of human and others in  $F_1$ - $F_2$ - $F_3$  space seen from the optimal discriminant plane plotted by dotted line.

$$F_3 = \sum_{k=0}^{L-1} W_{R_j}^{(k)} \quad (8)$$

人の動きと、人以外の誤報要因を含めた合計 13,820 種類の変化領域をとらえたシーンを用いて、3 つの特徴軸  $F_1, F_2, F_3$  により構成させる特徴空間で、人と人以外の誤報要因の分離度を最も大きくする特徴軸をフィッシャーの線形判別法により求めた結果、最適判別軸は  $F_{opt} = 0.213F_1 + 0.204F_2 + 0.956F_3$  で、エラー率が最も低くなるしきい値は 7.5 となった。

$F_1, F_2, F_3$  の特徴空間における人と人以外の分布と、上述の最適判別平面とを図 5 に示す。この分布の傾向から判断して、これ以上高次の判別面を用いても効果が少ないと考えられるため、最終的に線形判別を利用することとした。

単純な差分処理では困難が予想されるシーンにおける、人物抽出処理結果の一部を図 6 (a) ~ (b) に示す。図 6 (a) は、樹木の揺れが存在する背景中を移動する人、(b) は、水面の乱反射が存在する背景中を移動する人 (自転車) をとらえたシーンに関する処理結果をそれぞれ示している。

図中で人として認識した領域は白の四角形を、人以外の背景変動と認識した領域は黒の四角形をオーバーレイ表示した。いずれも人和其他の背景変動とが精度良く弁別できていることが確認できる。これにより大領域において人物の位置を正しく検出することができる。

### 3.2 アクティブカメラによる個别人物の高精度追跡

大域的な位置検出で想定した監視範囲で人の顔や手足の動きなどの詳細な動きを知りたい場合、カメラのパン、チルト、ズームを移動させながら移動対象を追跡する必要がある。この場合、初期カメラポジションでは背景差分処理による抽出処理が可能であるが、いったんカメラが動き出すと背景差分以外の方法で、移動

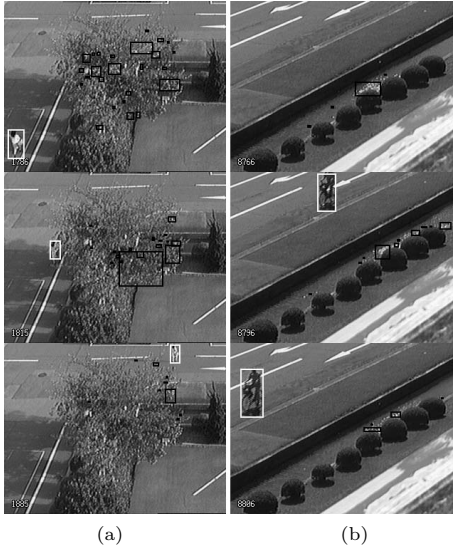


図 6 単純な差分処理では困難が予想されるシーンにおける人の抽出結果（人として認識した領域を白い長方形、背景変動と認識した領域を黒い長方形で表示）．(a) 激しい樹木の揺れが存在するシーン，(b) 水面の乱反射が存在するシーン

Fig. 6 Results of human detection in the changeful background, where the regions recognized as humans are framed with white, and non-human changes, with black, respectively. (a) Scene with unsettled trees, (b) Scene with diffused reflection of water surface.

対象を背景から切り出す必要がある．詳細は論文 9) を参照されたいが，ここでは局所的な動きベクトルを用いて，背景との動きの違いにより移動対象領域の抽出を行う．

図 7 (a) に示すように，移動対象の重心の移動先である  $(g_x, g_y)$  を中心として，移動対象領域の幅  $w$ ，高さ  $h$  よりも  $w' \times h'$  のサイズの相関ブロック数個分大きなサイズ  $w' \times h'$  の広がりを持った矩形の領域を設定し，図 7 (b) に示すように，その外側と内側にそれぞれ最大  $N_b$  個と  $N_o$  個ずつの相関ブロックを配置して，次フレームとの間の相関演算を実行する．相関演算のスコアは輝度差の絶対値の総和を用いた．

図 7 (c) に示すように，背景側の相関ブロック  $B_{bk}$  で求めた相関演算スコア  $S_{bk}(dx, dy)$  ( $k = 0, \dots, N_b - 1$ ) を最小にする動き量  $(u_{bk}, v_{bk})$  を求めて，2次元の速度空間に投票し，得られるヒストグラムの最頻値を与える動き量  $(u_b, v_b)$  を背景の代表動きベクトルとする．ここではカメラのパン，チルト時に生じる，移動対象近傍の背景領域のフローは一律な並進運動であると仮定している．

次に，図 7 (d) に示すように，移動対象側の相関ブロック  $B_{ok}$  で求めた相関演算スコア  $S_{ok}(dx, dy)$  ( $k = 0, \dots, N_o - 1$ ) を最小にする動き量  $(u_{ok}, v_{ok})$  を

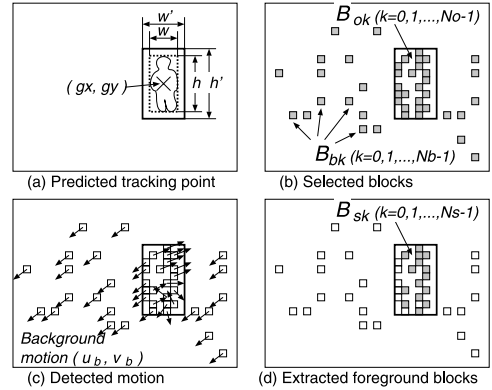


図 7 背景用，移動対象用の各相関ブロックの配置，背景との速度の違いによる移動対象の抽出

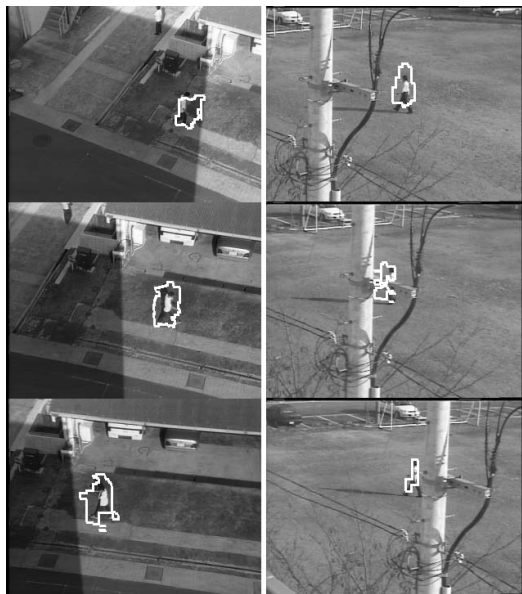
Fig. 7 Selected template blocks for correlation, and outline of object detection through flow-based background subtraction.

求めて，先程求めた背景の代表動きベクトル  $(u_b, v_b)$  と比較し， $(u_{ok} \neq u_b \cup v_{ok} \neq v_b)$  となる相関ブロックを移動対象の領域に割り当てられたブロックとしてグループ化する．このようにして， $N_o$  個の移動対象側相関ブロック  $B_{ok}$  ( $k = 0, \dots, N_o - 1$ ) から， $N_s$  個 ( $N_s \leq N_o$ ) の移動対象相関ブロック  $B_{sk}$  ( $k = 0, \dots, N_s - 1$ ) が選択抽出される．

ここで得られた，移動対象ブロック  $B_{sk}$  の内部の画素を 1，その他を 0 とし，ビットマップとして展開したものは，理想的には，背景との速度の違いにより抽出した移動対象領域となる．しかし一般的には，背景に生じた不均一な動きなど，本来は背景に属するブロックが含まれていたり，また逆に，本来は移動対象に属すべきブロックが，偶然に背景と同じ動きであったりするために，抜け落ちることがある．

そこで，ある時定数  $T$  を定義し，過去の時刻 ( $t = t_0 - T + 1$ ) から現在の時刻 ( $t = t_0$ ) に至るまでの  $T$  フレーム間の一連の追跡過程において，上述の移動対象ブロックから得られるビットマップを，追跡の中心，すなわちセグメントの重心を基準として画素単位で位置合わせしながら累積投票を行い，たとえば最大度数の  $T$  の半分の値  $T/2$  をしきい値として 2 値化処理を行うと，時間的な平滑化的作用により，おおむね原画像中の移動対象がシルエットとして切り出された二値画像を得ることができる．これを移動対象の時間的平均シルエットと呼ぶ．

移動対象を追跡しながら，カメラのパンチルトズームを自動的に制御する際に，困難が予想されるシナリオにおける，人物追跡処理結果の一部を図 8 (a) ~ (b) に示す．図 8 (a) は，影などの高コントラストな背景



(a) (b)

図8 移動対象を追跡しながら、カメラのパンチルトズームを自動的に制御する際に、困難が予想されるシーンにおける、人物追跡処理結果（追跡対象の時間平均シルエットの輪郭を白線でオーバーレイ表示）。(a) 影などの高コントラストな背景の通過、(b) 障害物による一時的な遮へい、のある各シナリオ

Fig.8 Tracking results of critical scenes for correlation tracking (Contour of the Temporal Averaging Silhouette is drawn over the original image sequence). (a) Passing through high contrasted shadow, (b) Temporal full target occlusion by obstacle.

の通過、(b) は、障害物による一時的な遮へい、の各シナリオにおける追跡の処理結果をそれぞれ示している。

図中で追跡対象の時間平均シルエットの輪郭を白線でオーバーレイ表示した。いずれの場合も、対象の追跡に成功していることが確認できる。

#### 4. 大域行動検出手法

大域行動検出で想定した監視映像では各人物の移動軌跡や速度などの、体全体を特徴とする行動の検知に対するニーズが多く存在する。たとえば図9は、人の移動軌跡を特徴とした人物行動の例である。ここでは、

- ビル、マンションなどの玄関口ピーで立ち止まりエレベータを待つ、またはビル内の奥側に直進する、
- 駐車場で何台もの車を物色しながらうろろする、あるいは通常に通るすぎる、
- 建物に入場する、あるいは建物から退出する、庭など玄関以外の場所に立ち入る、
- ゆっくりと動く、あるいは速く移動する、

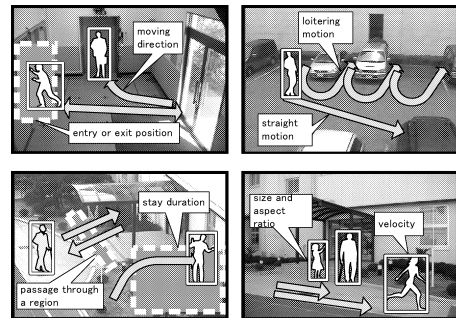


図9 大領域で検出される人物行動の例

Fig.9 Example of human activities to be detected.

などの特定行動の検知を想定している。

画像処理の方式としては、(1) 前述の大領域での位置検出により、映像中から人物領域を抽出し（これら各フレーム単位で抽出した人物の特徴量を1次メタデータと呼ぶ）、(2) その後、抽出した領域の位置情報を前後のフレームにわたって追跡し、1つの人物オブジェクトとして関連付ける（これら人物オブジェクト単位でまとめられた特徴量を2次メタデータと呼ぶ）。

追跡処理は、各フレームで抽出される人物領域について、前章で示した小ブロックの相関演算などにより動きベクトルを求め、前フレームの追跡対象の、現フレームにおける予測位置を求め、対象の候補との距離が最も小さいものどうしを関連付けることにより実現する。ただし、複数の対象候補が重なり合った場合は、それぞれが分離して正しい重心位置が求まるまでの間、直前の対象のサイズ、速度、および、テンプレートパターンを用いた予測追跡を行っている。

行動検知処理は、2次メタデータとユーザが指定したパラメータとを比較することにより行われる。以下は2次メタデータの例で、検索時に複雑な計算を必要とせず、単純な数値比較で済むような構造とした。これにより特定の領域行動を効率良く検出することができる。

- 1つの移動オブジェクトに関する特徴  
（出現・消失位置、最終的な進行方向、総移動量、存在時間、平均速度、軌跡の直線性など）
- 画面を格子状に分割した各小領域単位における特徴  
（特定位置における通過方向、特定位置における滞留時間）

さらにこのような行動情報を用いれば、たとえば、リアルタイムで取り込まれる監視映像からあらかじめ設定した条件に合致した行動をとる人物のみを、検出する「オンライン検索」だけでなく、映像が取り込ま



図 10 大域的な行動のオフライン検索結果の例  
Fig. 10 Results of human action retrieval.

れる時点で、これら 2 次メタデータを映像と関連付けて記録しておき、後からユーザが検索条件をインタラクティブに変更して目的の行動を行った人物を絞り込みながら見つけ出す「オフライン検索」を実現することも可能である。

実験の結果、たとえば、4 時間分の映像データ (JPEG) 約 12 GB に対して、我々の提案する 2 次メタデータのサイズは、およそ 1/1000 の 6 MB となった。Pentium-4/3.2 GHz の PC 上で動作検証を行った結果、この 4 時間で約 900 人が映っている映像に対して、検索条件を複数指定してオフライン検索を行った結果、結果表示を含めて 1 秒以内で検索が可能ながことが確かめられた<sup>13)</sup>。

図 10 に、検索処理結果の例を示す。ここでは、建物の玄関前で長時間うろうろと滞留する、あるいは玄関横の庭に侵入するといった行動を行う人のみを検出する条件で検索を行った。

## 5. 将来展望

以上で述べた手法は単眼固定カメラを用いたものであり、汎用性が高い手法といえる。しかし今後は単眼固定カメラにこだわることなく、様々な形態のカメラを用いて、より多くの条件下で、人物の位置を追跡し、行動を認識できる手法を開発することが重要と考えている。この章では我々の考える将来展望について簡潔に述べる。方向性として (a) より高度な行動の認識、(b) より多数を追跡、(c) より広範囲な追跡の 3 点をあげる。

### 5.1 高度な行動認識

2 章および 4 章で述べた行動認識は、小領域での行動認識と大領域での行動認識手法が別々になっている。将来たとえば小領域での行動認識アルゴリズムを、大

領域に適用し、より詳細な行動認識を行うことも考えられる。ただし以下の課題が残されている。

- (1) 大領域に含まれる小領域は、その場所や観測の時間、状況によって照明条件や見え方が大きく変化する。したがって、これらの変化に依存しない手法が必要となる。
- (2) 大領域中の小領域候補は複数存在するため、特に複数人物に対応するためには計算量が増加する可能性がある。

### 5.2 追跡の多人数化

本論文で述べた手法は 1 から数名程度の小人数を対象としている。しかし実際には監視システムの適用範囲を広げるために多人数を同時に追跡しなければならない場合がある。ただし対象が増加すると、人物同士の隠蔽が多くなり、追跡途中で見逃しや入れ替わりが頻発する。

たとえば Zhao ら<sup>14)</sup> は、背景差分の結果から人物のシルエットモデルをもとに多人数を追跡する手法を提案している。ただし、エッジ情報に依存しているので人物の境界上に明らかなエッジが存在しない場合は分離に失敗する場合がある。我々はステレオカメラから得られる 3 次元情報をもとに複数人物を追跡する手法を提案している。詳細は論文<sup>15)</sup>を参照されたいが、本手法により境界上に明確なエッジがなくても、距離情報をもとに複数人物を分離できるのでよりロバストな追跡が可能となる。複数人物の追跡例を図 11 に示す。画像上では人物同士の隠蔽が発生しているが、3 次元情報をもとに複数人物を追跡できていることが分かる。

ただしステレオカメラは計測範囲に制約があり、特に遠方での計測精度が著しく低下する性質がある。今後は Zhao らの手法のような「見え方」を用いた追跡手法と 3 次元情報による追跡手法を融合し、より広範囲でロバストな追跡手法も考えられる。

### 5.3 追跡の広域化

映像監視システムが大規模化するに従い、対象を追跡すべき領域が増大していくと考えられる。本論文での大域とは、カメラ視野がカバーする範囲が大域であるという意味であるが、ここで述べる広域とは複数カメラ (センサ) が存在し、かつ視野外の領域を含む範囲を想定しており、大域より大きな領域を議論する。

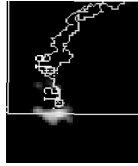
広域内には本論文の大域追跡などが含まれるが、それ以外にも以下の課題が存在する。

- (1) 視野外となる領域が存在するため、広域に分散する離散的な観測から視野外での人物位置と行動を推定する必要がある。



追従する歩行パターン．隠蔽が多いにもかかわらず，後方人物を追跡している

'Follow' walking pattern. Large and long occlusion occurs in this pattern.



自由歩行．複数人物を正しく追跡している

'Free-flow' walking pattern.

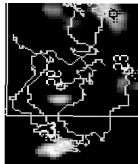


図 11 追跡実験の結果例．左列は各人物の現在の追跡位置をワイヤボックスで示した画像，右列は現在位置とそれまでの軌跡を体積素ヒストグラム上に示した画像

Fig. 11 Example of result. Left images show current positions by each wired-box. Right shows trajectories on voxel histogram.

## (2) 離散的な別々の観測を接続し，1つの追跡，行動を生成する必要がある．

上記(1)と(2)は表裏一体の関係にある．つまり視野外での位置，行動が推定できれば，異なる観測を一連の観測であると結論付けられるし，別々の観測が一連の観測と推定されれば，観測間の位置と行動もある程度推定可能である．たとえば Rahimi ら<sup>16)</sup>は，人物のダイナミクスを採り入れ各観測から MAP 推定を用いて人物の軌跡を推定する．しかし複数人物には未対応であり，またカメラ設置条件が水平設置のみと制約が大きい．また平澤ら<sup>17)</sup>は分散配置された複数の顔認証システムを用いて，まず顔認証間の接続関係の候補を推定(上記(2))し，時空間の無矛盾性を拘束条件として人物の移動パスを推定する(上記(1))．ただし顔認証システム間でしか機能しない．

今後はより柔軟なカメラ配置や，カメラ以外のセンサとの連携などが必要になっていくと考えられる．

## 6. おわりに

監視映像からの人物位置，行動検出技術について述べた．映像監視においては，照明条件，人物の状況，監視領域の大きさの違いにより様々なアプローチが考えられる．本論文では監視領域の大きさと人物位置，行動を検出する目的の組合せから人物監視応用に必要となる小域行動検出手法，大域位置検出手法，大域行動検出手法について述べた．小域行動検出ではエレベ-

タのような小閉空間から非定常動き(暴れ)を検出する手法，大域位置検出では木の揺れがあるような環境において人物の位置を検出する手法，さらに大域行動検出では大域位置を用いて汎用的に行動を検出する手法について述べた．今後は様々な手法の利点を融合した新しい手法や，ステレオ，アクティブカメラを含む多カメラ，多センサ連携による大規模，高信頼，高精度監視システムの研究を進める．

## 参 考 文 献

- 1) Oliver, N.M., Rosario, B. and Pentland, A.P.: A bayesian computer vision system for modeling human interactions, *IEEE Trans. PAMI*, Vol.22, No.8, pp.831-843 (2000).
- 2) Stauffer, C. and Grimson, W.E.L.: Adaptive background mixture models for real-time tracking, *Proc. Conf. Computer Vision and Pattern Recognition* (1999).
- 3) 高橋，関，小島，岡：ジェスチャー動画像のスポットティング認識，*信学論*，Vol.J77-D-II, No.8, pp.1552-1561 (1994).
- 4) 森田：局所相関演算による動きの検出と追跡，*信学論*，Vol.J84-D-II, No.2, pp.299-309 (2001).
- 5) 土田，川西，村瀬，高木：背景差分法による物体検出を目的とした逐次モンテカルロ法による背景推定，*信学論*，Vol.J87-D-II, No.5, pp.1062-1070 (2004).
- 6) 関，林，谷口，橋本，笹川：リアルタイム人物暴れ検出システム，*画像センシングシンポジウム*，pp.273-278 (June 2004).
- 7) 中井，福井，久野：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*，Vol.J77-D-II, No.7, pp.1209-1218 (1994).
- 8) 羽下，鷺見，八木：変化領域内の動きの時空間特徴に着目した屋外情景における歩行者の検出，*信学論*，Vol.J87-D-II, No.5, pp.1104-1111 (2004).
- 9) 羽下，鷺見，八木：時間平均シルエットを用いた能動カメラによる人の追跡，*信学論 D-II*，Vol.J88-D-II, No.2, pp.291-301 (2005).
- 10) 長井，久野，白井：複雑変動背景下における移動物体の検出，*信学論 D-II*，Vol.J80-D-II, No.5, pp.1086-1095 (1997).
- 11) 畠，岩井，谷内田：動き情報と情報圧縮を用いたロバストなジェスチャ認識手法，*信学論*，Vol.J81-D-II, No.9, pp.1983-1992 (1998).
- 12) 松山，和田，波部，棚橋：照明変化に頑健な背景差分，*信学論*，Vol.J84-D-II, No.10, pp.2201-2211 (2001).
- 13) 佐藤，羽下，野沢：人物映像の特定行動検索方式に関する検討，2005 年度信学会春季全国大会講演論文集，Vol.D-11, p.132 (2005).
- 14) Zhao, T. and Nevatia, R.: Stochastic human



segmentation from a static camera, *Workshop on Motion and Video Computing*, pp.9-14 (2002).

- 15) Hayashi, K., Hirai, T., Sumi, K. and Sasakawa, K.: Multiple-person tracking using a plan-view map with error estimation, *Proc. Computer Vision — ACCV 2006*, Vol.3851 of LNCS, pp.359-368 (2006).
- 16) Rahimi, A., Dunagan, B. and Darrell, T.: Tracking people with a sparse network of bearing sensors, *Proc. ECCV 2004*, Vol.3024 of LNCS, pp.507-518 (2004).
- 17) 平澤, 鹿毛, 三輪, 橋本: 時空間的無矛盾性を考慮したマルチポイント顔認証による動線認識, 第11回画像センシングシンポジウム, pp.357-360 (2005).

(平成 17 年 8 月 31 日受付)

(平成 18 年 3 月 17 日採録)

(担当編集委員 村上 和人)



林 健太郎

平成 7 年大阪大学工学部電子制御機械工学科卒業。平成 11 年同大学大学院工学研究科博士後期課程修了。同年三菱電機(株)入社。産業システム研究所を経て現在同先端技術総合研究所所属。コンピュータビジョン, 3次元認識等の研究に従事し, 人物の行動計測手法やセキュリティ用映像監視装置等を研究開発。博士(工学)。平成 18 年電機工業技術功績者発達賞受賞。電子情報通信学会会員。



羽下 哲司(正会員)

平成元年大阪大学基礎工学部制御工学科卒業。平成 3 年同大学大学院基礎工学研究科修士課程修了。同年三菱電機(株)入社。産業システム研究所を経て, 平成 14 年より先端技術総合研究所にてセキュリティ監視用画像処理システムの研究開発に従事。平成 18 年大阪大学大学院情報科学研究科後期博士課程修了。博士(情報科学)。電子情報通信学会, 画像電子学会各会員。



関 真規人

平成 4 年岡山大学工学部情報工学科卒業。平成 6 年同大学大学院工学研究科修士課程修了。同年三菱電機(株)入社。産業システム研究所を経て, 現在先端技術総合研究所にて監視用画像処理システムの研究開発に従事。電子情報通信学会会員。



笹川 耕一(正会員)

昭和 57 年京都大学工学部電気工学科卒業。昭和 59 年同大学大学院修士課程修了。同年三菱電機(株)入社。以来, 指紋照合や映像監視セキュリティ等画像処理応用システムの研究開発を経て, ビル・セキュリティ・エレベータシステムの開発に従事。現在, 同社先端技術総合研究所・物理セキュリティシステム開発プロジェクトグループマネージャ。平成 6 年工学博士(京都大学)取得。平成 10 年近畿地方発明表彰受賞。平成 15 年神戸大学連携大学院客員教授兼任。電子情報通信学会会員。