

時空間動き特徴に着目した屋外侵入者監視技術に関する研究

羽下 哲司[†] 八木 康史[‡]

[†]三菱電機 (株) 先端技術総合研究所

[‡]大阪大学 産業科学研究所

本論文では、屋外情景を撮影した監視カメラ映像を処理することによって侵入者を自動的に監視し、監視業務の省力化、低コスト化、効率化を実現するための画像処理手法を提案する。屋外侵入者監視で必要とされる、(1) 複雑に変動する背景下での誤報の少ない侵入者の検知、(2) 侵入者検出後の詳細な行動を見逃さずとらえる自動追尾、(3) 記録時間を向上させるための監視映像の高効率な符号化、のそれぞれの機能を、背景からの前景の抽出問題としてとらえ、侵入者領域内部の局所的な動きの時空間的な分布の特徴に着目することにより、従来手法の問題点を解決する画像処理アルゴリズムを提案し、その実用性を実シーンを用いた実験によって実証する。

In This paper, we propose

Intruder Surveillance Technology for Outdoor Scenes based on Spatio-temporal Motion Features

Tetsuji Haga[†], Yasushi Yagi[‡]

[†]Advanced Technology R&D Center, Mitsubishi Electric corp.,

[‡]The Institute of Science and Industrial Research, Osaka University

In this paper we propose an image processing method for video surveillance system in outdoor scenes which realizes automatic intruder detection, tracking and efficient video storage for labor and cost saving. We deal with the existing three subjects as the segmentation problem; 1. intruder detection with low false alarm rate under complicated and varying background, 2. intruder tracking with automatic pan-tilt-zoom active camera control and 3. high-compression rate surveillance video coding of for a long-term storage. With the experiments tested in the real scene, we proved the practicality of proposed algorithm which uses the motion features consists of spatio-temporal distribution of local flow.

1 はじめに

1.1 映像監視システムの概要

発電所、変電所、上下水道など、普段から無人の公共プラントや、空港周辺施設、港湾施設、工場、学校、駐車場など、夜間に無人になる屋外施設では、敷地内に侵入する人や車両をカメラによって遠隔から監視することにより、監視業務の省力化を行いたいというニーズが高い。

ここでは侵入者(広義に解釈して車両も侵入者とみなす)が出現したことをいち早く見つけ出して警報を発する侵入者の検知と、出現した侵入者が動いた場合、カメラのパン、チルト、ズームを制御して、見逃さずに視野内にとらえ続ける侵入者の追跡、そして侵入者の行動履歴や詳細な特徴をとらえた監視映像を記録する映像記録の三つが重要な課題である。

図1はカメラを用いた監視システムの基本構成である、パン、チルト、ズームカメラ(Pan-Tilt-Zoom Camera)、カメラから監視センタまでの映像伝送系、画像処理装置(Image Processing Unit)、監視センタ内の表示装置(Monitor)、映像記録レコーダー(Temporary Storage)、およびバックアップ装置(Long-term Storage)を示したものである。

画像処理に求められるタスクは、侵入者の自動検知により監視員の現場への派遣回数を減らす、検知後すぐに視野から外れるように動く侵入者をカメラのパン、チルト、ズームを制御して自動追尾を行う、監視映像の長期的なバックアップに要するデータ容量を削減してコストダウンを図る、などである。

本論文で扱う侵入者は、単数、あるいは極少人数(2~3人)の人を想定しており、雑踏や群集は対象外である。このとき、侵入者は、遠方から近方まで、画像中の任意位置に出現する、照明変化によ

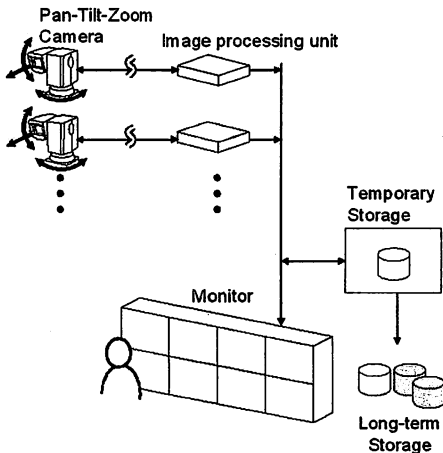


図 1: カメラを用いた監視システムの基本構成

り色やテクスチャが変化する、見える方向が変わる、任意の姿勢、スケールをとり得る、荷物を持つ、四肢が動くなど形状が変化する、といった特徴を有する。

1.2 従来の研究の概要

1.2.1 侵入者検出

侵入者などの移動物体を検出する技術としては、背景差分法が良く用いられる。文献 [1][2] にも詳しい解説がなされているが、単純なものは、中央値を用いたもの [3]、最頻値を用いたもの [4]。また時系列的な重みを入力画像に持たせたもの [5]、[6] が試されている。時間的に緩やかな照明変動には追従ができるが、急激な照明変動に対しては適応できないことがある。

画素値の最大値、最小値およびフレーム間差分の最大値を用いてモデル化を行なったもの [7]、[8]、[9]。もあるが、前景である移動物体やノイズの影響を大きく受けるため、誤検出が生じ易いという問題がある。

画素値の分布を正規分布でモデル化した [10]。や、部分画像毎にそのベクトル距離を正規分布でモデル化した [11]、[12]。また、各画素での移動ベクトルと画素値変化量をモデル化した [13] などは、背景の画素値の統計的な分布が単峰であれば精度の良いモデル化が可能である。しかし実際には草木のゆれや光のきらめきなど、一画素あたりに複数の対象が入り込むことがあるため、誤検出を生じる可能性がある。

これに対して、複数の正規分布からなる混合正規分布でモデル化を行なった手法 [15]、[14]、[16] は、屋外の監視映像においては比較的良好な結果を示す

ことが知られている。しかし全画素について 3~5 個の正規分布の演算を行わなければならないため、演算量の削減が課題である。また、混合正規分布でも表現できないような複雑な背景変動が生じた場合に誤検出が生じる。

これに対応するために、背景変動をヒストグラムによってモデル化した手法 [17]、[18] は、サンプル数が少ない場合はうまくモデル化できない。この点を改良し、カーネル密度関数を定義して、少ないサンプル数でも安定なモデル化を行なった手法 [19] は混合正規分布でも表現できないような複雑な背景変動をうまくモデル化できるが、演算時間がかかるという問題がある。

以上のように、背景差分処理によって抽出される領域は必ずしも侵入者だけとは限らず、背景の変動による誤検出を含むことがある。これに対して、背景差分によって抽出した領域を侵入者の候補領域としてフレーム間で関連付けを行ない、一定時間以上追跡が行なわれたものを侵入者とする手法が行なわれている [20]、[15]、[14]、[16]。領域の重心や面積といった単純な特徴量による領域のフレーム間での関連付けは誤対応が生じ易く、誤報や失報を生じることが多い。

これらに対して、背景のモデリングではなく、前景である侵入者すなわち人の全身や頭部の輪郭エッジなどのモデルを用いる手法として、Condensation [21]、[22]、[23]、[24] や、その考え方を複数人物の頭部追跡に応用した手法 [25] が検討されている。空間的にユニークな形状をパラメトリックな形式でモデル化し、画像中で移動対象の候補が存在する確率を、モデルのパラメータ空間における確信度の分布で表し、確信度分布の伝搬と再構成を複数フレームにわたって繰り返す手法である。しかし、例えば、追跡の途中から、対象の上半分、あるいは下半分が長時間にわたり背景中の他の対象によって隠されるなど、対象そのものの形状や見え方が途中で変わってしまう場合には、確信度のモデルをうまく適合させることができないため、追跡に失敗することがある。

画像中に現れる局所的な動きに着目して、例えば、画像中から得られるフローを、空間的に統合して動きセグメントとした後、一定時間の予測軌跡を時空間に投票し、投票値が高くなるものを抽出する手法 [26]、同様の考え方を小ブロックによる局所相関演算で実現した [27] では侵入者領域内部の局所的な動きが時間的にも空間的にも連続しているという特徴は重要であるが、単純に時空平均した動きのピークのみで判定する場合、例えば追跡時間内に瞬時的に生じた空間コントラストの強い領域の動きが平均値を引き上げてしまい値を越え、誤報が生じる場合がある。また、局所領域で求めたフローはエラーを含むため、各フローをボトムアップ的にまとまりのある領域に統合することは容易ではない。

1.2.2 侵入者追跡

移動する対象を追跡する画像処理技術に関しては、[28]にも詳しく記載されているが、ここでは一般的な幾つかの手法を紹介する。

最も一般的な、移動する物体を追跡する手法としては、ブロック相関による相関追尾 [32], [33], [27], [34], [35] がよく知られている。これは、発見した移動対象に対して複数個の相関ブロックをテンプレートとして配置し、各々のブロックの動きからブロック群全体の動きを求め、以後テンプレートの配置を更新しながら追跡を続けるもので、研究事例や製品としての実績も多い。しかし、人のような非剛体の移動対象では、常に全身でフローが求まるとは限らず、部分的にフローの出現と消滅が繰り返されて、重心位置が安定せず、追跡が不安定になるという問題がある。

これに対して、色ヒストグラムの確率密度分布を用いて、追跡対称の次のフレームでの予測位置における相関を最大にする位置ずれ量 (Mean Shift Vector) を山登り法によって求め、追跡を行なう手法がある [29],[30]。Mean Shift と呼ばれるこの手法は演算量が少ないことと、変形に対してロバストであるという特徴を有する。しかし、安定な追跡のためには、追跡対象のサイズに適したサイズのカーネル関数と呼ばれる特徴量の重み付け積分を行う必要がある。この点を改良し、カーネル関数のサイズをオンラインで推定しながら追跡する手法 [31] もある。しかしいずれの場合も、夜間は有効なカラー情報が使えないため追跡が不安定になる。

また、輝度情報のみを用い、移動対象のシルエットの情報と、textural temporal template とよばれる、移動対象の出現頻度に応じて重み付け平均化された、テクスチャ情報のモデルを併用する追跡手法 [9] も提案されているが、対象が途中で向きを変えたり、する場合は別途モデルが必要になる。

1.2.3 監視映像の高効率記録

MPEG-2,4 では、画像を単一のプレーンと見なし、フレーム間の動き補償 (MC) と DCT 変換を用いて動画像符号化 (以後、本稿では単一 VOP 符号化と呼ぶ) を行なうが、屋外の監視映像においては木の揺れや水面の乱反射など、背景の変動領域に多くの符号化データが割り当てられるため、効率が良くない。これに対して、MPEG-4 に規格として採用されているオブジェクト符号化 [39, 40] は、映像を前景領域と背景領域とに分けて別々に符号化するため効率の良いデータ量の削減が期待できる [41]。また、同様の考え方として、文献 [42] では、映像中の各画素を背景、静止領域、移動領域の3種類にクラス分け、背景を定期的に静止画として圧縮保存し、静止領域の画素は前フレームの情報を用い、移動領域の情報のみを LZH 圧縮することに

より、高圧縮率な可逆符号化方式を実現する手法が提案されている。

しかし、いずれの方法においても、前景領域の自動抽出は容易ではなく、誤った領域の抽出が生じると、形状記述のオーバーヘッドにより符号化効率が低下する。また逆に、領域の一部の欠けや、フレーム単位での領域の検出もれが生じると、復号時に対象の一部が欠落するため、映像記録として使うことができない。

1.3 本論文の構成

本論文では、2以降を以下のように構成する。

2では、木の揺れや、水面の乱反射等の背景変動のある屋外情景から、侵入者のみを精度良く検出する画像処理アルゴリズムについて述べる。

3では、検出した侵入者を追跡し、カメラのパン、チルト、ズームを自動的に制御しながら、侵入者をカメラ視野内にとらえ続ける画像処理アルゴリズムについて述べる。

4では、屋外の監視カメラ映像のバックアップなど、長期蓄積のための高効率な動画像符号化技術について述べる。

5では、まず本研究で得られた成果を要約し、今後に残された課題について述べる。

2 屋外侵入者検知の誤報低減

2.1 提案アルゴリズムの概略

屋外の監視カメラ映像を処理して自動的に侵入者を検出する技術に求められるタスクは、昼夜、天候を問わず、近景から遠景にわたる監視領域内で動く侵入者を検出するものである。その際、見逃しによる失報が許されないと同時に、日照変動や木の揺れ、水面の光の乱反射等の環境の変動に起因する誤報を低減することが要求される。

まずはじめに、提案アルゴリズムの基となった、基本的なアイデアを図2に示す。

背景差分処理などによって抽出された侵入者の候補は、その領域内の動きの空間的な分布に着目すると、(1) 一様な分布、(2) 不均一な分布、(3) 全てゼロ、のいずれかに分類可能である。

次に、それらの動き (ベクトル) を加算して平均的な動きを求めると、それぞれ (1) 動きあり、(2) 動きゼロ、(3) 動きゼロ、に分類できる。

さらに、平均的な動きを代表移動ベクトルとして時間的に追跡を行なうと、(1) は (1a) 一方向に直線運動と (1b) ランダムに運動とに分類され、(2) 動き無し、(3) 動き無し、のいずれかに分類される。

このとき (A) が侵入者 (一部車両や小動物なども含まれる) であるが、実際にはこのような単純な分類で侵入者とそれ以外の背景変動とを区別することは不可能であり、さらに詳細な解析が必要となる。






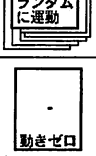



領域内の個々の動きの分布	領域を代表する平均的動き	平均的動きの時間的な分布	事象の分類
 一様な分布	 動きあり	 一方向に連続運動	A 侵入者 および 車両、小動物 虫・鳥等 の一部
 不均一な分布	 動きゼロ	 ランダムに運動	B 草木、影の揺れ、光の反射 ヘッドライトの一部
 全てゼロ	 動きゼロ	 動きゼロ	C 水面の強らぎ 西陽の直射光 光輪、ヘッド ライト、レン ズ面の水滴 D 日照変動 および 影の出現と 消滅

図 2: 領域内の動きの分布

そこで、提案するアルゴリズムでは、抽出した変化領域内部の局所的な動きを相関演算処理によって追跡し、次に示す 3 つの動きの特徴量、

- F_1 : 空間平均した動きの強さ
- F_2 : 時間平均した動きの強さ
- F_3 : 時間的な動きの一様性

により構成される特徴空間で人の動きと、その他の誤報要因とを弁別する [36], [37].

特長として、変化領域内部で、エラーを含む各局所領域の速度のアウトライヤ除去を行い、変化領域の代表速度を精度良く求めることができる。この結果、従来手法 [20], [15], [14], [16] で問題となった変化領域の関連付けの誤対応が減少する。

また、変化領域があらかじめ求まっているので、従来手法 [26][27] で問題となった、フローの統合の誤りによる誤報の発生を抑えることができる。

そして、特に身体の形状モデルを用いないので、従来手法 [21], [22], [23], [24] [25] で問題となった、体の一部が遮蔽されて、上半身のみ、あるいは下半身のみしか写っていない場合でも人の動きの検出が可能である。

さらに、動き解析に用いる特徴量として、従来手法 [27] では F_2 に相当する、時間平均した動きの強さという特徴のみが用いられたが、提案手法ではこれに空間平均した動きの強さ F_1 、および時間的な動きの一様性 F_3 の特徴が加えられ、より詳細に人の動きと背景変動を分類することが可能である。

2.2 動き解析に用いる特徴空間の定義

水面の光の乱反射と、歩行者を撮影したシーンで、動きの検出を行った処理結果をそれぞれ、図 3, 4 に示す。それぞれの図中 (a) には、原画像中で人として検出した領域を四角形でオーバーレイ表示している。図 3(a) 誤って人と認識している。

人と人以外の動きの違いを詳細に調べるために、図 3, 図 4 の図中 (b) に、領域内部で局所相関演算が行われたブロック位置を拡大し、相関値マップの値を輝度値に対応付けてオーバーレイ表示した。ここでは相関値マップは、相関が高い程明るい色で、逆に相関が低い程暗い色で表示されている。例えば中心よりも右上に明るい相関値のピークがある場合は、フレーム間で右上に移動したと判断することができる。なお、局所相関演算は領域内部の全ての場所で行われるのではなく、相関ブロック内の空間周波数が高いものを優先して選択的に実行される。

各図中 (c) に示す 8 つの相関値マップは、それぞれが追跡過程の各時刻における局所的な相関値マップを空間的に累積したものである。(c) の一番左の相関値マップは変化領域検出時刻における相関値マップ (b) の空間的な累積結果で、これから前述の空間平均した動きの強さ F_1 が得られる。

(d) の相関値マップは、(c) の 8 つの相関値マップを累積したもので、これから前述の時間平均した動きの強さ F_2 が得られる。

また、(c) の 8 つの相関値マップのうち、空間平均した動きの強さがしきい値を越えたものの数は、前述の時間的な動きの一様性 F_3 の特徴量を示している。

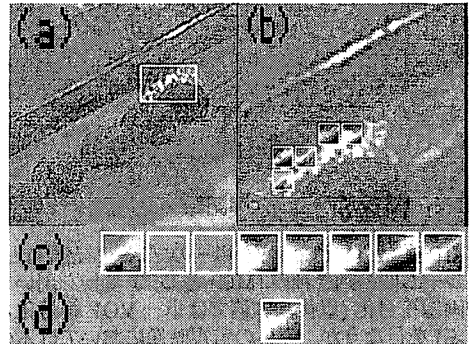


図 3: 誤検出領域 (a) と相関値マップ (b, c, d) (水面)

これらの結果から、以下の傾向が読み取れる。

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 (F_3) は

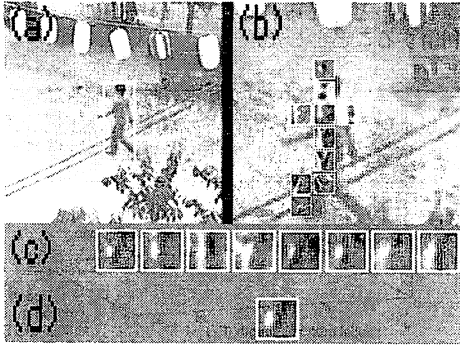


図 4: 正検出領域 (a) と相関値マップ (b,c,d) (人)

いずれも高く、途中で動きゼロにならない

このような傾向は、その他の誤報要因である、木の揺れ、ヘッドライトによる光や影の動き、街中の旗や揺れる装飾品、雨粒等について比較した場合にも確認できた。

以下では、これら F_1, F_2, F_3 の特徴をそれぞれ定式化することを試みる。

2.2.1 空間平均した動きの強さ (F_1) の定式化

空間平均した動きの強さとは、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。

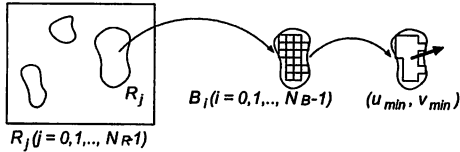


図 5: 変化領域内のブロックの配置と代表速度の算出

図 5 に示すように、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 $w \times wy$ 画素の各相関ブロック B_i ($i = 0, 1, \dots, N_B - 1$) の位置から、現在 (時刻 t) の輝度パターンをテンプレートとして切り出し、時刻 $(t-1)$ の画像を探索画像として、以下の式で示すように、輝度差分の絶対値総和を計算することにより相関値マップ $S_{B_i}(u, v)$ を求める。

$$S_{B_i}(u, v) = \sum_{(x,y) \in B_i} |I_{t-1}(x+u, y+v) - I_t(x, y)| \quad (1)$$

ここでは領域 R_j 内の各ブロック B_i の相関値マップ $S_{B_i}(u, v)$ 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 $S_{R_j}^*(u, v)$ のピークを判定することにより調べる。各局所相関値マップのピーク位置が空間的に一樣な場合、各ピークが強め合い、累積した相関値マップに明瞭なピークが現れる。反対に空間的に不均一な場合は、各ピークが弱め合い、累積した相関値マップのピークは平坦となる。 $S_{R_j}^*(u, v)$ は以下の式で表される。

$$S_{R_j}^*(u, v) = \frac{1}{N_{R_j}} \sum_{B_i \in R_j} \{W_{B_i} \cdot S_{B_i}(u, v)\} \quad (2)$$

ただし、重み付け関数 W_{B_i} 、正規化のための分母の値 N_{R_j} を以下の式で定める。

$$W_{B_i} = \begin{cases} 1 & (\arg \min_{(u,v)} S_{B_i}(u, v) \neq (0, 0)) \\ 0 & (\text{otherwise}) \end{cases}$$

$$N_{R_j} = \sum_{B_i \in R_j} W_{B_i} \quad (3)$$

空間的な動きの強さ F_1 は、変化領域の代表速度 (u_{min}, v_{min}) を用いて、以下の式で定義される。ただし $(u_{min}, v_{min}) = \arg \min_{(u,v)} S_{R_j}^*(u, v)$ である。

$$F_1 = S_{R_j}^*(0, 0) - S_{R_j}^*(u_{min}, v_{min}) \quad (4)$$

2.2.2 時間平均した動きの強さ (F_2) の定式化

時間平均した動きの強さとは、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。

前節で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 $t-1$ の画像から輝度パターンをテンプレートとして取り出し、時刻 $t-2$ の画像を探索画像として、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 $t-2$ の画像から時刻 $t-3$ の画像への相関演算、... 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。

以後、 $S_{R_j}^*(u, v)$ に関しては、追跡回数 k を示す添字を付けて、 $S_{R_j}^{*(k)}(u, v)$ と記載する。

空間的な動きの強さと同様に、時間的な動きの強さに関しても、累積した相関値マップ $S_{R_j}^{*(k)}(u, v)$ 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 $S_{R_j}^{**}(u, v)$ のピークを判定することにより調べる。 $S_{R_j}^{**}(u, v)$ は以下の式で表される。

$$S_{R_j}^{**}(u, v) = \frac{1}{N} \sum_{k=0}^{L-1} \{W_{R_j}^{(k)} \cdot N_{R_j}^{(k)} \cdot S_{R_j}^{*(k)}(u, v)\} \quad (5)$$

ただし、重み付け関数 $W_{R_j}^{(k)}$ 、正規化のための分母の値 N を以下の式で求める。

$$W_{R_j}^{(k)} = \begin{cases} 1 & (\arg \min_{(u,v)} S_{R_j}^{*(k)}(u, v) \neq (0, 0)) \\ 0 & (\text{otherwise}) \end{cases}$$

$$N = \sum_{k=0}^{L-1} (W_{R_j}^{(k)} \cdot N_{R_j}^{(k)}) \quad (6)$$

空間的な動きの強さ F_2 は、以下の式で定義される。ただし $(u_{min2}, v_{min2}) = \arg \min_{(u,v)} S_{R_j}^{**}(u, v)$

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

2.2.3 時間的な動きの一様性 (F_3) の定式化

時間的な動きの一様性とは、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、ここでは先に求めた $W_{R_j}^{(k)}$ を用いて、以下の式 F_3 で定義する。

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

2.3 動きの特徴空間におけるクラス判別

人の動きと、人以外の誤報要因を集めた合計 13,820 種類の変化領域をとらえたシーンを用いて、前節で定式化した F_1, F_2, F_3 の 3 つの特徴空間における人の動きと人以外の誤報要因のクラス分離性能を検証した。誤報サンプルには、屋外環境で誤報となる例が多い、風に揺れる装飾品、木の揺れ、ヘッドライトによる光と影の揺れ、水面の乱反射の 4 種類が含まれている。

従来手法として使われている特徴軸 F_2 のみによる識別ではエラーが多く不十分なため、他の特徴軸 F_1, F_3 も取り入れたが、それぞれの特徴を個別に使っても、いずれも識別性能が不十分であることが確認できた。そこで、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。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。

最適判別平面 $0.213F_1 + 0.204F_2 + 0.956F_3 = 7.5$ と平行な視線方向から見た、特徴空間における人と人以外の分布を図 6 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。

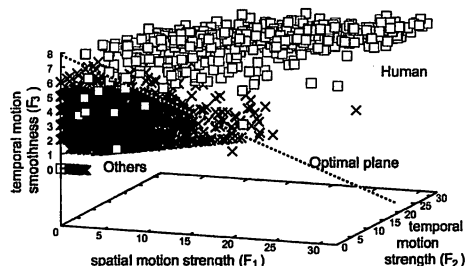


図 6: 最適判別平面 (点線で表示) と平行な面から見た、特徴空間 F_1, F_2, F_3 における人と誤報要因の分布

2.4 実データを用いた性能検証実験

前述のクラス判別平面を求めるために用いた 13,820 種類のシーンとは異なる、従来法 1 (背景差分処理のみ) では誤報が発生するシーンを集めた 3,364 フレームの屋外環境のテストシーケンスを用い、従来法 2 (背景差分処理と動きの特徴 F_2 による判定)、および提案手法 (背景差分処理と動きの特徴 F_{opt} による判定) とで、発生する誤報数を比較する実験を行った。評価は目視判定により実施した。

性能評価に用いた背景差分処理は文献 [15] の手法を採用した。

表 1: 提案アルゴリズム導入による誤報数の低減

評価シーン	総フレーム数	手法 1	手法 2	提案法
急激な日照変動	450	1	0	0
ヘッドライト	653	17	4	2
草木の揺れ	215	8	5	2
影の揺れ	449	13	5	1
水面の乱反射	512	14	8	2
西陽、直接光	512	1	0	0
レンズの水滴	375	28	12	3
犬猫等の小動物	99	11	8	3
虫、鳥	99	11	8	3
合計	3364	104	50	16
1 シーンの平均		11.5	5.6	1.8

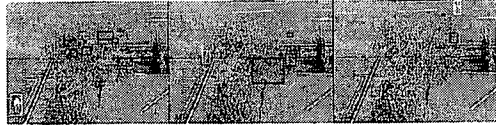
結果を表 1 に示す。従来法 2 は従来法 1 の誤報を平均的に約 1/2 に低減しているが、提案手法は、従来法 2 に対して誤報をさらに約 1/3 に低減することが確かめられた。なお、カメラレンズの水滴、犬猫等の小動物、虫、鳥に関しても誤報の低減が見られたが、これらは速度が一定でないか、フレー

ム間の移動量が極端に大きく想定した探索範囲を越えたために発報とならなかったもので、条件によっては誤報となる可能性がある。

画像処理結果の一部を図 7(a)~(d) に示す。



(a) 人, 風に揺れるカーテン, ちょうちん, 樹木



(b) 人, 樹木の揺れ



(c) ヘッドライトの光と影の移動



(d) 人(自転車), 水面の乱反射

図 7: 侵入者検出の処理結果

図 7(a)は風に揺れるカーテン, ちょうちん, 樹木が存在する背景中を移動する侵入者を, (b)は樹木の揺れが存在する背景中を移動する侵入者を, (c)はヘッドライトを点灯した車の移動によるの光と影の移動を, (d)は水面の乱反射が存在する背景中を移動する侵入者(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。

図中で侵入者として認識した領域は白の四角形を, 侵入者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。いずれも侵入者と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。

また, 市販の PC にアルゴリズムを実装し, オンラインで性能評価を行うシステムを構築した。

実験後に誤報を確認するために, 侵入者を検出した前後の映像をハードディスクに記録することとし, また, 定められた時刻に故意に監視領域内に侵入することにより, 失報率のチェックを実施した。

14 日間にわたり, 昼夜連続で行った実験により, 得られた結果ログ(合計 312 時間, 8,400,000 フレーム)を目視により解析した結果, 一日当りの平均誤報数 = $15/14 = 1.1$ (件/日), 平均失報率 = 失報 3/総侵入回数 390 = 0.77% 以下の性能が確認できた。

3 カメラ制御による侵入者追跡

3.1 背景

カメラを用いた遠隔監視システムでは, 監視員は常時監視カメラ映像を注視しているのではなく, 侵入者などの警報発生時に実際の映像を確認し, 必要な警備行動をとるとともに, 記録された映像を再生して侵入者の行動を事後分析する。このため, 死角を減らして侵入者の行動履歴を見逃さなく記録する必要がある。しかし, 屋外環境に監視カメラを設置する場合, カメラだけでなく, ポール建設や配線工事などの費用が発生するため, できるだけ少ないカメラ台数で広い範囲を監視したいという要求がある。

このため一台のカメラで, パン, チルト, ズームを制御して侵入者を追跡する必要がある。また, 侵入者個人を特定することよりも, むしろ侵入者が監視範囲全体の中で, どこを通過し, どこに滞り, どのように行動したかを, そのときの周囲の状況と同時に記録し, 事後に確認できることが重要である。このため, 侵入者を極端に大きくズームアップするのではなく, 例えばカメラ視野内に, 画像の縦サイズのおよそ 1/5 程度の大きさになるように維持して, 発見から約 30 秒間は見失わない, といった追跡が必要となる。また, 夜間は, 白色光で照明されることは少なく, 単色放電ランプによる照明や近赤外の照明が使われたり, フレーム蓄積により感度を上げて撮像することも多い。そのため, カラー処理や, フレームレートの高い映像を利用することを前提とした手法は避けなければならない。また, 侵入者自身が持つ荷物や, 周囲の物体によって体の一部が隠れることも多いため, 対象の形状や見え方に関する事前知識の利用を前提とした手法も避ける必要がある。更に, カメラが動くと, それまでに学習した背景情報も使えないので, 固定視野での背景差分を前提とした手法も避けなければならない。

なお, 移動対象は, 通常は無人の場所に侵入した単数の人を想定しており, 移動, 停止, 方向転換を行うが, 複数人物の同時追跡や, 追跡中の人どうしの接触, 相互遮へいは取り扱わない。

近年提案されている各種追跡アルゴリズムは, 個々の性能は高いものの, 色情報を用いる手法は, 照明条件の変動や, 侵入者の画像中でのサイズが小さい場合に追跡が不安定になり易い。また, 形状モデルを使う手法は, 対象そのものの形状や見え方が途中で変わってしまう場合に追跡が不安定になる。また, 移動対象の輪郭の輝度こう配を特徴量として追跡を行なう方法は, 夜間などフレーム蓄積を行なってカメラ感度を高めた場合に追跡が不安定になる。また, 移動対象のテクスチャ情報のモデルを併用する方法は, 侵入者のカメラに対する向きが変化する場合に追跡が不安定になる, などの問題があり, 本論文が想定する適応対象においては, いずれもその要求事項と制約条件を同時に満たす

ことができず、適応不可能であることが分かった。そして、むしろ古典的なブロック相関を用いた相関追尾が、本論文の適応対象に最も適しているということが分かった。

提案するアルゴリズムは、相関追尾において、従来より問題となっている、人の追跡時に生じる、部分的なフローの欠落による重心位置の不安定性を解決し、安定な侵入者の追跡を可能にするものである。時間平均シルエットと呼ばれる、追跡対象の存在頻度の高い領域を表す領域情報を表現したモデルを用いることにより、相関追尾の安定性を向上させる [38]。

3.2 提案する侵入者追尾システムの概要

提案する侵入者追尾システムは、カメラ静止状態で、移動対象である侵入者を自動的に発見した後、常にカメラ映像の中心付近で、ほぼ一定した大きさに写るように、カメラのパン、チルト、ズームの制御を行うものである。

侵入者の発見に関しては、2で述べた方法で、背景差分により抽出した領域内部の動きベクトルの時間的な振舞いに基づき、移動する人と背景の変動とを信頼度高く弁別して、人のみを精度良く抽出する。そして、侵入者が発見されると、次のフレームからただちに処理内容を追跡処理に切り替える。

追跡処理の過程では、カメラ自身が動くため、輝度に基づく背景差分ではなく、移動対象と背景とのフローの違いを用いる。これらの違いにより抽出した移動対象に属する相関ブロック群を用い、移動対象の重心の x 座標 $g_x(t)$ 、 y 座標 $g_y(t)$ 、幅 $w(t)$ 、高さ $h(t)$ の、それぞれの次フレームにおける予測値を推定しながら追跡を行う。

監視員により侵入者に対する警報がリセットされる、あるいは移動対象を見失ってから、あらかじめ設定した一定の時間が経過すると、カメラは初期監視位置に戻り、背景差分による監視を行う。

カメラの動特性は未知であり、しかも制御の遅延や残差が発生することが多い。また、カメラから得られるパン、チルト、ズームの位置情報 c_p 、 c_t 、 c_z にも誤差が含まれるため、制御目標値はカメラパラメータではない。ここでは、画像処理により求めた重心の予測位置 $(g_x(t), g_y(t))$ と画像の中心との差がゼロになるようにカメラのパン速度 c_p 、チルト速度 c_t を制御する。また、対象の高さ $h(t)$ が画像の縦サイズ l_y に対して、ほぼ $1/5$ の大きさになるようにズーム比 c_z を制御する。

このようなカメラの閉ループ制御を実現するために、移動対象の重心位置 $(g_x(t), g_y(t))$ とサイズ $h(t)$ を常に安定に抽出することが本論文の課題であるが、同じ結果が得られるのであれば、必要な演算量はできるだけ少なくしたい。ここでは、必要な演算量を、一つのブロックにおける相関演算に要する演算量と、ブロック数との積にほぼ比例

するとして、背景の動きを計測するために配置する局所相関ブロック数 (N_b) と、移動対象の動きを計測するために配置する局所相関ブロック数 (N_o) の合計値 $(N_b + N_o)$ によって評価する。

3.3 時間平均シルエットを用いた追尾アルゴリズム

一般に、画像全体で、常に信頼度の高いフローが検出されることは稀であり、毎フレーム画像全体で密にフローを計算することは演算量的に効率が悪い。ここで、移動対象内部及びその周囲の時空間的な高周波成分を多く含む、固定サイズの小ブロックを有限個数選択してフローを計算することにより、演算量を大幅に削減することができるが、単純にフローの計測点数を選択的に削減した場合、特に人のような非剛体の移動対象では、形状やアスペクトの変形による一時的なフローの欠落や、障害物による遮へい、高コントラスト背景による偽フローの出現、等のシーンにおいては対象の重心位置が変動し、追跡が不安定になることがある。

そこで、提案手法では、背景とのフローの違いにより抽出した移動対象の領域を、過去数処理サイクルに渡る追跡過程において画素単位で位置合わせしながら、移動対象域内の画素値を 1、その他を 0 として画像上に投票し、最終的に得られる度数分布、すなわち移動対象の検出頻度が十分に高い画素により構成される領域を生成する。このようにして得られた領域を時間平均シルエットと呼び、このシルエットの重心に基づいて追跡を行う。たとえ各々のフレームではフローの欠落があったとしても、時間平均シルエットでは時間的な累積の効果によって、フローの部分的な欠落が補われているため、人のような非剛体の移動対象であっても検出される重心の変動が抑えられ、結果として安定な追跡処理が可能となる。

これは、フローの時間的な連続性に着目して、空間的に密に相関演算を行う代わりに、時間的に長く観測することにより、追跡の安定性を向上させる手法と考えることができる。

以後、提案する時間平均シルエットを用いた追尾処理の各ステップの原理を、以下の処理の流れに従って示す。

- 移動対象領域内外の相関演算ブロックの配置
- 背景との動きの違いによる移動対象領域の抽出
- 時間平均シルエットの生成
- 移動対象の重心、サイズ、速度の算出

3.3.1 相関演算ブロックの配置

演算コストの制約から、できるだけ少ない数の相関ブロックを効率良く配置し、背景の速度、移動

対象の速度を推定するとともに、移動対象を背景から切り出すことを考える。図 8(a) に示すように、移動対象の重心の移動先 (初期状態では、背景差分処理により発見した移動対象領域の重心) 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 よりも相関ブロック数個分大きなサイズ $w' \times h'$ の広がりをもった矩形の領域を設定し、その外側 (背景側) と内側 (移動対象側) に複数の相関ブロックを設置する。ここでは、画像全体に密にブロックを配置するのではなく、図 8(b) に示すように、位置に優先順位をつけて、サイズ $w' \times h'$ の矩形領域の外側と内側にそれぞれ最大 N_b 個と N_o 個づつ有限個のブロックを配置する。

画像中で横方向の隣の画素どうしの輝度差分演算、縦方向の輝度差分演算、現在の画像と 1 フレーム前の画像との間の同一画素どうしの輝度差分演算によって得られる各微分値 D_x, D_y, D_t のブロック内でのヒストグラムを求め、時空間的な高周波成分が多く含まれるブロック位置から順に配置する。この優先順位の計算は毎フレーム更新される。

背景側の第 k 番目の相関ブロック B_{bk} の中心位置を (cx_{bk}, cy_{bk}) ($k = 0, 1, \dots, N_b - 1$)、移動対象側の第 k 番目の相関ブロック B_{ok} の中心位置を (cx_{ok}, cy_{ok}) ($k = 0, \dots, N_o - 1$) とする。ブロックサイズはいずれも $b_x \times b_y$ とする。

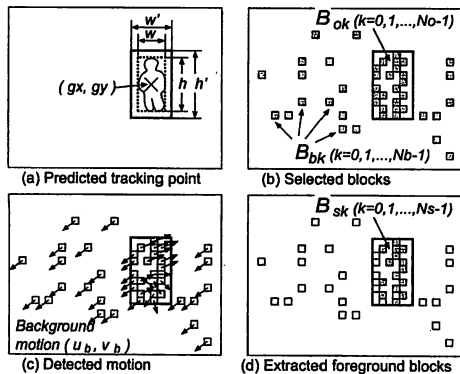


図 8: 背景用、移動対象用の各相関ブロックの配置、背景との速度の違いによる移動対象の抽出

3.3.2 背景との動きの違いによる移動対象領域の抽出

図 8(c) に示すように、選択された全相関ブロックで、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。

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様な並進運動であると仮定している。

次に、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。

この処理は、サイズ $w' \times h'$ の矩形領域内では、移動対象に属する相関ブロックの動きベクトルは (u_b, v_b) と異なり、背景に属する相関ブロックの動きベクトルは (u_b, v_b) と等しいという仮定に基づいている。しかし、ここで得られた、移動対象ブロック B_{sk} 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ために、抜け落ちることがある。しかし、次に述べる時間平均シルエットの生成処理により、このような部分的に生じる過剰抽出や欠けの影響は、相対的に弱められて無視できるものとなる。

また、ズーム中は背景のフローが一様な並進運動であるという仮定が崩れるため、強制的に $(u_b = 0, v_b = 0)$ とし、それまでと同じ追跡処理を行い続ける。ただし、ズーム値を連続して大きく変動させるような制御は行わず、微小な変化率 (例えば 5~10%以内) で、断続的に制御するものとする。このとき、サイズ $w' \times h'$ の矩形領域内では本来は背景に属するブロックの多くが、誤って移動対象ブロック B_{sk} として判断されることとなる。しかしそれら過剰抽出の影響は、ズームを短時間で、しかも非連続的に行うことと、次に述べる時間平均シルエットの生成処理により、相対的に弱められて無視できるものとなる。

3.3.3 時間平均シルエットの生成

時刻 t における追跡の中心を $(g_x^{(t)}, g_y^{(t)})$ 、背景との速度の違いにより抽出された、移動対象に属する相関ブロックを $B_{sk}^{(t)}$ ($k = 0, 1, \dots, N_s^{(t)} - 1$)、その中心位置を $(cx_{sk}^{(t)}, cy_{sk}^{(t)})$ とする。

ここで、ブロック $B_{sk}^{(t)}$ の内部の画素を 1、その他を 0 とし、すべてのブロックの画素情報を $w' \times h'$ のサイズの画像 $J^{(t)}(x, y)$ 上にビットマップとして

展開する。すなわち、

$$J^{(t)}(x, y) = \begin{cases} 1 & (|x + g_x^{(t)} - cx_{sk}^{(t)}| \leq \frac{b_x}{2} \\ & \cap |y + g_y^{(t)} - cy_{sk}^{(t)}| \leq \frac{b_y}{2}; \\ & \exists k (k = 0, 1, \dots, N_s^{(t)} - 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

ただし、ここで x, y は入力画像の左上の点を原点とする座標系ではなく、時刻 t における追跡の中心 $(g_x^{(t)}, g_y^{(t)})$ を原点とする座標系を用いる。

先にも述べたように、ある時刻 $t = t_0$ の瞬間的な相関ブロックの分布だけでは、人のような非剛体の移動対象ではその一部が欠けることが多く、安定して重心位置を求めることができない。

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 $J^{(t)}(x, y)$ で定義される、各時刻のビットマップを、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うことにより、多値の2次元画像 $J_{acc}^{(t_0)}(x, y)$ を得る。ただし、ここでも x, y は時刻 $t = t_0$ における追跡の中心 $(g_x^{(t_0)}, g_y^{(t_0)})$ を原点とする座標系を用いる。

$$J_{acc}^{(t_0)}(x, y) = \sum_{t=t_0-T+1}^{t_0} J^{(t)}(x - g_x^{(t)}, y - g_y^{(t)}) \quad (10)$$

$J_{acc}^{(t_0)}(x, y)$ は、各画素の度数が移動対象の検出頻度を表している。 $J_{acc}^{(t_0)}(x, y)$ を以下の式、

$$\begin{cases} i = x + g_x^{(t_0)} \\ j = y + g_y^{(t_0)} \end{cases} \quad (11)$$

で示されるような、入力画像の原点を基準とする座標系に変換した後、例えば最大度数の T の半分の値 $T/2$ をしきい値として2値化処理を行うことにより、原画像中の移動対象がシルエットとして切り出された2値画像 $I_s^{(t_0)}(i, j)$ を得ることができる。

画像 $I_s^{(t_0)}(i, j)$ における画素1の領域を時刻 $t = t_0$ における移動対象の時間的平均シルエットと呼ぶ。

図9に、各フレームで抽出される移動対象に割り当てられたブロック(上段)、と生成された時間的平均シルエット(下段)の概念図を示す。

時間平均シルエットからは、追跡処理に必要な、移動対象の重心位置やサイズを得ることができる。

3.4 実シーン映像を用いた性能検証実験

一般的な相関追尾では、人の安定な追跡が困難になると予想される以下の(a)~(d)シーンを含む

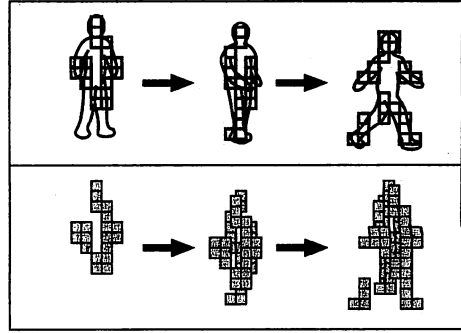


図9: 各フレームで抽出される、移動対象に割り当てられたブロック(上段)と生成された時間的平均シルエット(下段)

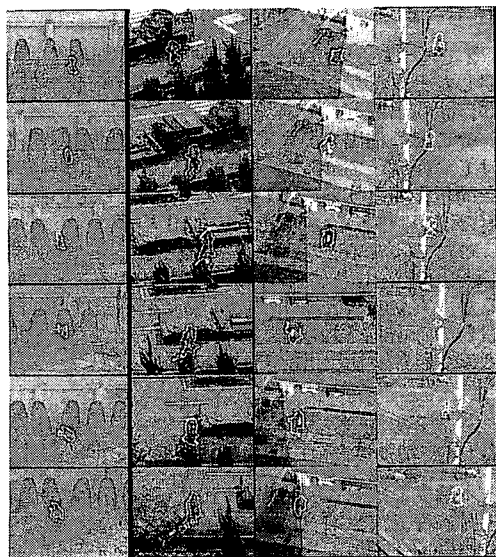
約5分間づつのテストシーケンス20回分に対して追跡評価実験を行った。評価基準は、監視領域内に侵入した人を発見してから30秒以上追跡を失敗することなく、その行動履歴を記録できれば追跡成功とした。

- (a) 移動対象の形状やスペクトルの変化
- (b) 障害物による部分遮へい
- (c) 高コントラスト背景の通過
- (d) 障害物による一時的な全遮へい

変形や遮蔽を含むシーンでの性能検証の結果、全シーンにおいて30秒以上人の追跡を維持することに成功した。評価に用いた原画像に、追跡処理に用いた時間平均シルエットの輪郭をオーバーレイした処理結果の一部を図10(a)~(d)に示す。

これら、一般的な相関追尾では、人の安定な追跡が困難になるシーンに対しても、提案手法が少ない演算量で、安定な追跡が可能であることが確かめられた。

提案手法による移動対象の追跡実験を行った結果、移動対象、背景領域ともに平均で32個、合計で64個のブロックが採用され、追跡処理が行われることが分かった。なお、対象の移動方向が極端に大きく変化する場合には、時間平均シルエットを用いたことによって、動きの慣性が増大し、重心の追跡が不安定になる可能性がある。このような場合は T の値をあまり大きくできない。しかし、本論文で想定するような、カメラの画角、移動対象の速度、サイズを有する、今回の実験に用いたシーンでは、時間平均シルエットの累積フレーム数を $T = 16$ 前後の値にしたとき、いずれも、良好な追跡結果を示すことが分かった。処理レートが15frame/secであるので、累積時間は約1秒となる。



(a) (b) (c) (d)

図 10: 相関追尾で困難が予想されるシーンの追跡処理結果 (a)形状やアスペクトの変化, (b)樹木による対象の一部遮へい, (c)高コントラスト背景の通過, (d)障害物による一時的な全遮へい)

カメラの制御は、横方向のずれに対してはパン、縦方向のずれに対してはチルトと、それぞれ独立した制御を行っている。パン軸の上に、チルト軸が搭載されたカメラの構造上、本来これらの値は独立ではないが、設定した撮像範囲、対象の大きさ、速度、画像処理周期では問題なく制御可能であることを確かめた。

4 監視映像の高効率記録

4.1 基本的な考え方

屋外の監視映像においては、侵入者や車両など、監視にとって重要な前景領域と背景領域とを分けて別々に符号化することにより、効率の良いデータ量の削減が期待できる [41]。例えば、従来の MPEG-2,4 の単一 VOP 符号化を用いて、1 次記録を行い、過去の映像を 2 次記録として長期記憶する際に、提案するオブジェクト符号化による高効率データ圧縮を実行すれば、データ容量の問題から、現状では廃棄されている過去の映像を、長期蓄積することが可能となる [43]。

しかし、前景領域の自動抽出は容易ではなく、誤った領域の抽出が生じると、形状記述のオーバーヘッドにより符号化効率が低下する。また逆に、領

域の一部の欠けや、フレーム単位での領域の検出もれが生じると、復号時に対象の一部が欠落するため、映像記録として使うことができない。

そこで、従来監視のために開発された移動物体検出手法に対して、(1) 前景としての確信度の高い領域のみを厳選して抽出する (2) 抽出した前景領域に対して、動き情報の時空間的な連続性を用いた平滑化を行い、フレーム内での空間的な領域の欠けを補う (3) 前景としての確信度の高い領域の存在するフレームを起点として、時間軸方向の前後に追跡を行い、検出もれが生じたフレームの前景領域を補間する、といった拡張を行うことにより、映像記録に適用可能な、高効率なオブジェクト符号化技術を提案する。

4.2 課題の定式化

ここでは、監視映像におけるオブジェクト符号化の課題を定式化する。現実の監視映像が、背景 $B^{(t)}$ と、前景領域 $F^{(t)}$ から成り、 $F^{(t)}$ はさらに形状を表す情報 $A^{(t)}$ とテクスチャを表す情報 $T^{(t)}$ から構成されるモデルを考える。ここで、 (t) は観測された時刻を表す添字である。

$$\left. \begin{array}{l} A^{(t)} \\ T^{(t)} \end{array} \right\} \left. \begin{array}{l} B^{(t)} \\ F^{(t)} \end{array} \right\} \rightarrow Scene^{(t)} \quad (12)$$

カメラから得られる映像は、これらが合成されて、各画素の座標が (x, y) で表される単一のプレーン上の情報 $I^{(t)}(x, y)$ として得られるが、オブジェクト符号化では、高効率な符号化を行うために、これを再度、仮想的な背景 $\hat{B}^{(t)}(x, y)$ と、前景 $\hat{F}^{(t)}(x, y)$ (すなわち $\hat{A}^{(t)}(x, y)$ と $\hat{T}^{(t)}(x, y)$) とに分離する。

$$I^{(t)}(x, y) \rightarrow \left\{ \begin{array}{l} \hat{B}^{(t)}(x, y) \\ \hat{F}^{(t)}(x, y) \left\{ \begin{array}{l} \hat{A}^{(t)}(x, y) \\ \hat{T}^{(t)}(x, y) \end{array} \right. \end{array} \right. \quad (13)$$

そして、高効率な符号化を行うために、背景 $\hat{B}^{(t)}(x, y)$ を、その冗長な成分を取り除いて $\tilde{B}^{(t)}(x, y)$ に変換し、 $\hat{A}^{(t)}(x, y)$ 、 $\hat{T}^{(t)}(x, y)$ と共に符号化する。

$$\left. \begin{array}{l} \hat{B}^{(t)}(x, y) \rightarrow \tilde{B}^{(t)}(x, y) \\ \hat{A}^{(t)}(x, y) \rightarrow \hat{A}^{(t)}(x, y) \\ \hat{T}^{(t)}(x, y) \rightarrow \hat{T}^{(t)}(x, y) \end{array} \right\} \rightarrow code \quad (14)$$

このように、オブジェクト符号化を行うためには、以下の 2 つの手続きが必要となる。

- (P1) $I^{(t)}(x, y)$ からの $\hat{A}^{(t)}(x, y)$ の抽出
- (P2) $\hat{B}^{(t)}(x, y)$ からの、 $\tilde{B}^{(t)}(x, y)$ への変換

手続き (P1) に関しては、単純な移動体抽出手段に加えて、以下の技術が必要となることは既に述べた。

- (T1) 前景領域の誤抽出の低減
- (T2) 動き特徴を用いた前景領域の整形
- (T3) 双方向追跡による前景領域の補間

一方、手続き (P2) に関しては、例えば、背景の木の揺れなどの変動を無視して、一定期間同じ背景画像を更新せずに使い続けることにより、データ量を大幅に削減することができる。ただし背景画像の更新を長時間行わなければ、例えば侵入者によって物が動かされた、あるいは取り除かれた等の情景変化を記録することができない。記録映像を見て、人の行為と情景変化を関連付けて判断するために、少なくとも十数秒に1回は背景画像の更新を行う必要がある。

ここで、監視映像に適應する際の制約と、高効率な符号化という観点から、オブジェクト符号化に求められる条件をまとめると、以下のようになる。

- 監視における制約条件 (a),(b)
 - (a) 抽出した前景領域の形状 $\hat{A}^{(t)}(x,y)$ が、真の前景領域 $A^{(t)}(x,y)$ の内部を削り取らない
 - (b) 前景領域のテクスチャ $T^{(t)}(x,y)$ の品質を高く保って符号化する
- 高効率符号化のための条件 (c),(d),(e)
 - (c) 背景 $\hat{B}^{(t)}(x,y)$ は、状況の概略が判読できれば良い (誤った前景への分類を極力減らす必要がある)
 - (d) 背景 $\hat{B}^{(t)}(x,y)$ において、移動物体の侵入や行為によって生じた変化は、その情報を保つ
 - (e) 抽出した前景領域の形状 $\hat{A}^{(t)}(x,y)$ は、真の前景領域 $A^{(t)}(x,y)$ の内部を削り取らないかぎり、形状は問わない (構造が単純なほど、形状符号化に要する符号量は少ない)。

4.3 動き特徴の時空間双方向追跡によるオブジェクト符号化

本論文で提案する、監視映像のオブジェクト符号化手法は、4.2で定義した2つの手続き (P1), (P2) に基づいている。ここでは、これらのうち前景領域のセグメンテーションに係わる手続き (P1) を実現するための (T1)~(T3) の技術と、背景領域の符号量の削減に係わる手続き (P2) を実現するオブジェクト符号化手法について、順に詳しく述べる。

4.3.1 前景領域の誤抽出の低減

2で述べた、変化領域内の動きの時空間特徴に着目したセグメンテーション手法が利用可能であるが、誤検出はゼロではない。

そこで、抽出された領域をフレーム間で関連付けて、連続して4フレーム以上出現した領域のみを前景領域とすることにより、前景としての確信度の高い領域のみを厳選して抽出する。

ここでは7.5fpsの処理レートを想定しており、約1秒以上存在した移動対象のみを抽出することになるが、実験の結果、サイズが 10×20 画素以上、背景との輝度レベルが10 (256階調) 以上の対象は、ほぼもれなく検出可能であることを確認した。

4.3.2 動き特徴を用いた前景領域の整形

本論文では、この時間平均シルエットを拡張し、追跡過程で、4.3.1の手法により得られた、確信度の高い前景領域を用いて、動き特徴の時空間的な平滑化の際に、得られるシルエット領域を下方向、および左右方向に膨張させることにより、足元および手の欠落が生じないようにしている。確信度の高い前景領域が得られないフレームでは、前フレームと同サイズの前景領域が得られたものとして、領域を膨張させて追跡を行う。なお、侵入者を前景領域がもれなくカバーするのに必要な膨張率は10%を越えないことを実験により確認した。

このようにして抽出された前景領域の形状 $\hat{A}^{(t)}(x,y)$ は、真の前景領域の形状 $A^{(t)}(x,y)$ よりも多少広がった形状となるが、真の領域内部を削らない限り問題とはならない。

4.3.3 双方向追跡による前景領域の補間

4.3.1, 4.3.2の手法を用いると、追跡過程において、前景領域が全く得られないフレームが生じて、それを補間することが可能である。しかし、これだけでは、初期状態で確信度の高い前景領域が確定されるまでのフレームで生じる前景領域の抽出もれ (ディレイ) を補間することはできない。

本論文では、初期状態から始まる追跡の過程で、前景領域が確定すると、図11に示したように、時間軸の正の方向の追跡処理をそのまま継続しながら、その一方で、画像バッファに記憶した過去の映像を用いて、全く同様の、小領域の相関演算による追跡処理を、時間軸の負の方向に対して開始し、前景領域が画像中に出現する時刻までさかのぼりながら、時空間的な平滑化を行って前景領域を補間することにより、初期状態で確信度の高い前景領域が確定されるまでのディレイの問題を解決する。なお、図では追跡対象として人を示したが、車両であっても同様の処理が可能である。

以上の4.3.1 ~ 4.3.3の技術は、監視における制約条件 (a) と、符号化のための条件 (e) を満たすよ

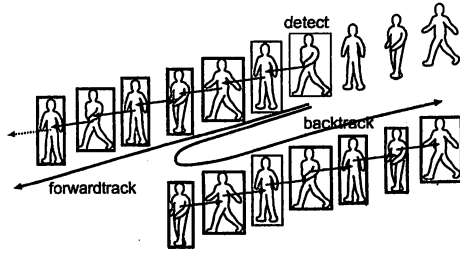


図 11: 双方向時空間追跡による前景領域の補間

うに実施した。具体的な方法と結果については、4.4 における説明と、それぞれ False Negative Rate の比較、符号化データ量の比較、の実験結果で示す。

4.3.4 オブジェクト符号化

ここでは P2 を実現する手法である、オブジェクト符号化について述べる

MPEG-4 の任意形状オブジェクト符号化では、物体の形状と透過率を記述する、アルファ画像と呼ばれるマップを用いて、シーン中の対象を、その前後関係に従って、前景、背景に対応する各 VOP (ビデオオブジェクトプレーン) に分解する。

ここでは、時刻 t ($= 0, 1, \dots$) のアルファ画像 $A^{(t)}(x, y)$ は、それぞれ対応する時刻の、前景領域のセグメンテーション結果を用いて、以下のように透過率が 0 または 255 の 2 値に設定する。

$$A^{(t)}(x, y) = \begin{cases} 255 & (\text{前景領域}) \\ 0 & (\text{otherwise}) \end{cases} \quad (15)$$

そして、この $A^{(t)}(x, y)$ に基づいて、前景側、背景側の 2 種類の VOP が生成される。

背景画像は、4.2 で述べた、手続き (P2) に従い、ある時刻に取り込んだ、原画像を基準背景画像として符号化し、次に更新するまでの間、同じものを使い続けることとした。背景の動的な更新は重要な研究課題であるが、ここでは、前景セグメントの抽出手法に焦点を絞って、オブジェクト符号化の可能性を検討したため未検討である。背景の更新レートは、4.2 に従って、400 フレーム (約 13 秒) の固定とした。

4.4 実シーンを用いた性能検証

4.4.1 性能検証試験の仕様

提案するオブジェクト符号化 (以後 O/C) 手法の有効性を確認するために、以下の 5 種類の動画画像符号化の比較実験を行った。

1. S-VOP : 単一プレーンでの MC+DCT
2. Typical : 一般的な領域抽出手法+O/C
3. L-FPR : 技術 (T1) の誤報低減+O/C
4. Proposed : (T1)~(T3)+O/C (提案手法)
5. Manual : 手動での領域抽出+O/C

いずれも、動画画像標準化委員会の MPEG-4 エンコードプログラム [39] を用いて実行、手法 (1) についてはアルファ画像を "None" に、手法 (2)~(5) に関しては "Binary" に設定、前景、背景ともに、量子化パラメータの値は $Q = 10$ 固定とした。

また、オブジェクト符号化では、各シーン (400 フレーム) の先頭フレームの画像を背景画像として用い、同一シーケンス内は常に同じ背景画像を用いた。手法 (2)~(4) の背景差分処理は、いずれも文献 [15] の手法を利用した。

比較実験に用いた映像シーケンスは、以下の A~F の 6 種類である。いずれも解像度 352×288 , YUV420, サンプリングレート 7.5fps で、全シーケンスとも背景に木、布、装飾物、電線の揺れを含む。

- A: 激しい背景変動、端から出現し、端へ消失
- B: 端から出現し、樹に半遮へいされて消失
- C: 比較的静かな背景、端から出現、端へ消失
- D: 樹の半遮へい部分から出現、画像端へ消失
- E: 走って移動、端から出現、端へ消失
- F: 三名が樹の半遮へい部分から出現、端で消失

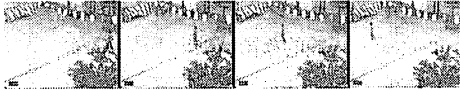
画像端からの出現と消失だけでなく、樹木による半遮へい領域からの出現や、半遮へい領域への消失、背景が静かなシーン、対象が走る高速移動、複数人の同時出現など、本論文で取り扱う監視映像において想定される、監視映像の特徴を網羅するように選定した。図 12 に実験に用いた映像の例を示す。

本実験では、提案手法 (4) が、以下の 5 つの条件を満たすことを確認することにより、その有効性を示す。

- 手法 (1) よりも符号化量が圧倒的に少ない。
- 手法 (2) よりも符号化コード量が少ない。
- 手法 (3) よりも、領域の欠け割合 (False Negative Ratio) が少ない。
- 手法 (5) と比較して符号化効率的が同等である。
- 前景領域での、原画像と復号化画像との間のパワー SNR が 30dB 程度の値を有する。

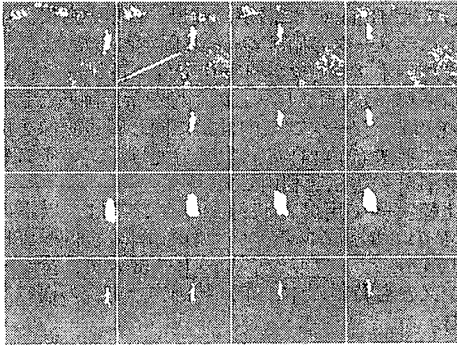
4.4.2 性能検証試験結果

図 13 に、図 12 の時刻に対応した、手法 (2)~(5) の各オブジェクト符号化で用いられたアルファ画像の前景領域 (白く表示された部分) を示す。



frame0030 frame0040 frame0050 frame0060

図 12: 実験に用いた屋外監視映像の例 (Scene A : frame0030 ~ frame0060)



frame0030 frame0040 frame0050 frame0060

図 13: 手法 (2)~(5) の各オブジェクト符号化に用いた, 図 12 の原画像に対応する, 前景オブジェクトのセグメンテーション結果 (アルファ画像) 上から順に, 手法 2(Typical), 手法 3(L-FPR), 手法 4(Proposed), 手法 5(Manual)

手法 (2) では, 背景領域における変動の多くを前景領域として誤抽出している。また, 手法 (3) では, 出現時のディレイにより, 人の領域もれが生じている。一方, (4) の提案手法では, 人の領域をきれいに前景領域として抽出している。手法 (5) は比較のために示した, 人手により抽出した真値データである。

各シーン毎, 各手法の符号化データ量の比較結果を図 14 に, 前景領域の検出精度の評価結果を図 15 に示す。ここでは, 手法 (5) のアルファ画像を真値データとして, 手法 (2)~(4) の各手法のセグメンテーション結果において, 真の前景領域の内部で削られた画素数の割合 (False Negative Ratio) の最悪値を示した。

図 14 に示したように, A~F の全ての評価シーケンスにおいて, 提案手法 (4) は, 単一 VOP 符号化 (1) と比較して符号化量が圧倒的に少なく, 最大で 10.3% (背景変動の最も激しいシーケンス A), 最小でも 26.6% (背景変動の最も少ないシーケンス C), 平均で 15.1% となった。抽出された前景領域の面積 (画素数) が, 真値と比較して増大しているため, テクスチャの記述に要する符号量が増す一方で, 単純化された形状の記述に要する符号量が減少する

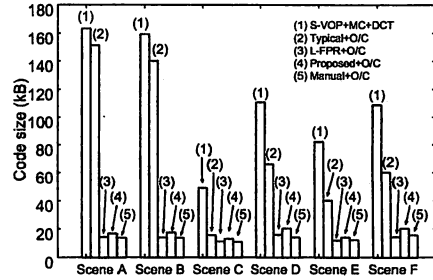


図 14: 動画像符号化方式によるデータ量の比較

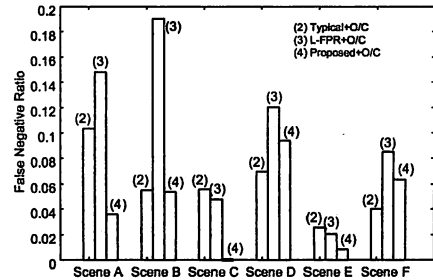


図 15: 前景領域抽出精度の False Negative Ratio による比較 (真値に対する面積比の最悪値をプロット)

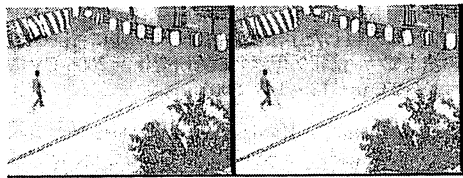
ため, トータルとして符号量の増加は僅かな値に抑えられる。この結果により, 符号化のための条件 (e) が達成されたことが確認できる。

また, 図 15 に示したように, A~F の全ての評価シーケンスにおいて, 領域の欠けの割合 (False Negative Ratio) は提案手法 (4) がもっとも少なく, 最悪値で 0.1, 平均で 0.01 となった (真値に対する面積比)。この結果により, 監視における制約条件 (a) が達成されたことが確認できる。

また, (2)~(5) の各オブジェクト符号化手法で符号化したデータをデコードして得られた複号化画像と原画像との間で, 真の前景領域 (マニュアルで求めたセグメンテーション結果の領域) 内のパワー SNR を計算した結果, いずれも 30dB 前後の値を維持しており, 監視用途の画質として問題の無いレベルであると言える。この結果により, 監視における制約条件 (b) が達成されたことが確認できる。

図 16 に, 手法 (5) でオブジェクト符号化した画像を, デコードした複号化画像を, 原画像と並べて示す。

前景領域である, 左から侵入する人の形状とテクスチャの情報が保存されていることが確認できる。この結果により, 符号化のための条件 (c) が達成されたことが確認できる一方で, 背景は過去のものを使い続けているため, 原画像と異なる部分



Original Image Decode Image

図 16: 原画像と提案手法の復号化画像との画像の比較 (Scene B : frame0015)

が見られる。しかし、十数秒後に背景が更新された後には、このような背景の不整合は生じない。この結果により、符号化のための条件 (d) が達成されたことが確認できる。

このように、監視にとって重要な、侵入者の領域の情報は欠落せず保存し、周辺部の重要度の低い領域の情報を削減することで、高効率の動画画像圧縮を実現することができる。

5 結論

監視映像における、侵入者の検出における誤報の低減、およびカメラのパン、チルト、ズームを制御しながら侵入者を視野内にとらえ続ける自動追尾、そして監視映像の長期的な保存を可能にする高効率な画像圧縮に関する新手法を提案し、これらが従来手法の問題点を解決し、屋外侵入者監視システムの省力化、低コスト化、効率化を実現できることを示した。

2 では、動きの特徴量として、時間平均した動きの強さ、空間平均した動きの強さ、時間的な動きの一意性という三つの特徴量を定義し、これらにより構成される特徴空間で歩行者と他の背景変動とを識別することにより、従来手法と比べて誤検出率、未検出率ともに約 1/3 に低減できることを示し、実環境での評価試験でも失報率 0.01 未満、誤報発生 1 日平均 3 件以下の性能を得ることが確認できた。

本手法の課題は侵入者と背景とを識別する識別器のさらなる性能向上と自動的な学習である。

3 では、カメラ自身の動きにより生じる背景の動き(フロー)と、侵入者領域のフローの違いにより得られる小領域を、位置合わせを行いながら時間的に累積することにより得られる、時間平均シルエットと呼ばれる領域セグメントを定義し、このシルエットの位置とサイズに基づいて追跡を行なうことにより、部分的、一時的なフローの消失にロバストで安定な追跡結果を得ることが確認できた。

本手法の課題は侵入者を追跡するモデル(時間平均シルエット)の精度である。

4 では、提案アルゴリズムでは、(1) 誤検出の低

減、(2) 抽出した前景領域の整形、(3) 検出領域の時間軸方向の補間、により、高効率かつ必要な情報の欠落が少ないオブジェクト符号化技術が実現でき、従来の MPEG-2 と比較してとして MC+DCT 動画画像符号化を行った結果に対して平均で約 15 %にまでデータ量を削減でき、しかも前景領域の真値データに対して、画素の欠けが 1 %未満に抑えられることが確認できた。

本手法の課題は長時間の監視映像に対応するための、背景画像の動的な更新アルゴリズムを検討である。

参考文献

- [1] Toyama K., et.al, "Wallflower: Principles and Practice of Background Maintenance," ICCV 99, pp.255-261, 1999.
- [2] 鷺見, 関, 波部, "物体検出-背景と検出対象のモデリング," 情報処理学会技術研究報告 2005-CVIM-150-11, pp. 79-98, 2005.
- [3] Lo B. and Velastin S.A., "Automatic congestion detection system for underground platforms," ISIMVSP, pp. 158-161, 2001.
- [4] 佐藤, 土川, 伴野, 石井, "歩行者計数のための照明変化にロバストな背景画像更新法," 信学春秋全大, no. D-408, 1994.
- [5] 川端, 谷藤, 諸岡, "移動物体像の抽出技術," 情報論文誌 vol. 28, no.4, pp.395-402, 1997.
- [6] 谷岸, 上田, 池谷, 堀場, "背景画像更新処理を用いた路面濡潤状況の検出," 電子情報通信学会誌 D-II, Vol.J80-D-II, No. 9, pp.2270-2277, 1997.
- [7] Haritaoglu I. Harwood D. and Davis L.S., "A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance," 15th ICPR, vol.4, pp.179-183, 2000.
- [8] Haritaoglu I. Harwood D. and Davis L.S., "W4s: A Real-Time System for Detecting and Tracking People in 2 1/2 D," ECCV '98, pp.877-892, 1998.
- [9] Haritaoglu I. Harwood D. and Davis L.S., "W⁴:Real-Time Surveillance of People and Their Activities," IEEE Tran. on PAMI, vol. 22, no. 8, pp.809-830, 2000.
- [10] Wren C. R., Azarbayejani A., Darrell T. and Pentland A.P., "Pfinder: Real-Time Tracking of the Human Body," IEEE Trans. on PAMI, vol. 19, no. 7, pp. 780-785, 1997.
- [11] Makito Seki, Hideto Fujiwara and Kazuhiko Sumi, "A Robust Background Subtraction Method for Changing Background," WACV 2000, pp. 207-213, 2000.

- [12] 関, 藤原, 鷺見, “背景変動に頑健な背景差分法,” 画像の認識・理解シンポジウム MIRU2000, vol. II, pp. 403-408, 2000.
- [13] 和田, 松山, “動的背景モデルを用いた移動領域の抽出,” 情処全国大会, vol 2, pp. 141-142, 1994.
- [14] Grimson W.E.L., Stauffer C., Romano R. and Lee L., “Using Adaptive Tracking to Classify and Monitor Activities in a Site,” IEEE CVPR’98, pp. 22-31, 1998.
- [15] Stauffer C. and Grimson W.E.L., “Adaptive Background Mixture Models for Real-time Tracking,” IEEE CVPR’99, vol. II, pp.256-252, 1999.
- [16] Stauffer C. and Grimson W.E.L., “Learning patterns of activity using real-time tracking,” Trans. on PAMI, vol. 22, no. 8, pp. 747-757, 2000.
- [17] Nakai H., “Non-Parameterized Bayes Decision Method for Moving Object Detection,” ACCV95, pp. 447-451, 1995.
- [18] 中井, “事後確率を用いた移動物体検出手法,” 情処研報 CV90-1, pp.1-8, (1994)
- [19] Elgammal A., Harwood D. and Davis L. S., “Non-Parametric Model for Background Subtraction,” ECCV00, vol. II, pp. 751-767, 2000.
- [20] 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 D-II, Vol. J77-D-II No.7 pp.1209-1218, 1994.
- [21] Isard M. and Blake A., “Visual tracking by stochastic propagation of conditional density,” Proceedings on ECCV1996, pp.343-356, 1996.
- [22] Isard M. and Blake A., “CONDENSATION - Conditional Density Propagation for Visual Tracking,” International Journal of Computer Vision, vol. 29, no.1, pp.5-29, 1998.
- [23] Isard M. and Blake A., “ICondensation: Unifying low-level and high-level tracking in a stochastic framework,” Proceedings on ECCV1998, pp.893-908, 1998.
- [24] Isard M. and Blake A., “A smoothing filter for CONDENSATION,” Proceedings on ECCV1998, pp.767-781, 1998.
- [25] 杉本, 谷内, 松山, “確信度付き仮説群の相互作用に基づく複数対象追跡,” 情報処理学会論文誌: コンピュータビジョンとイメージメディア, Vol.43, No.SIG 4(CVIM 4), pp.69-84, 2002.
- [26] 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.
- [27] 森田, “局所相関演算による動きの検出と追跡,” 信学論 (D-II) Vol. J84-D-II No.2, pp.299-309, 2001.
- [28] 加藤, 深尾, 羽下, “対象追跡-フレーム間差分の類似度に着目した手法から動きのモデルに着目した手法まで-,” 情報処理学会技術研究報告 2005-CVIM-150-23, pp. 185-198, 2005.
- [29] Comaniciu D., Ramesh V. and Meer P., “Real-Time Tracking of Non-Rigid Objects using Mean Shift,” IEEE CVPR2000, pp. 142-149, 2000.
- [30] Comaniciu D., Ramesh V. and Meer P., “Kernel-Based Object Tracking,” Trans. PAMI, pp.564-577, 2003.
- [31] Collins R. T., “Mean-shift Blob Tracking through Scale Space,” IEEE CVPR2003, pp.234-240, 2003.
- [32] 稲葉, “局所相関プロセッサを用いたロボットビジョン,” 日本ロボット学会誌, vol.13 no.3, pp.327-330, 1995.
- [33] 伊藤, 上田, “画像認識ハードウェア内蔵カメラを用いた自律移動体追尾監視カメラ,” SSII’99, pp.13-18, 1999.
- [34] 金子, 堀, “小領域のブロックマッチングを複数用いたロバストなオブジェクト追跡法,” 信学論 (D-II), vol.J85-D-II, no.7, pp.1188-1200, Sep. 2002.
- [35] 菅谷, 金谷, “運動物体分離のためのカメラモデルの自動選択,” 情報処理学会 CVIM 研究会 2002-CVIM-134-2, pp.9-16, 2002.
- [36] 羽下, 鷺見, 八木, “変化領域内部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 D-II Vol.J87-D-II, No.5, pp.1104-1111, 2004.
- [37] Haga T., Sumi K., Yagi Y., “Human Detection in Outdoor Scene Using Spatio-Temporal Motion Analysis,” 17th ICPR, vol.4, pp.331-334, 2004.
- [38] 羽下, 鷺見, 八木, “時間平均シルエットを用いた能動カメラのための人物追跡,” 信学論 D-II Vol.J88-D-II, No.2, pp.291-301, 2005.
- [39] ISO/IEC 14496-2:2001, “Coding of Audio-Visual Objects - Part 2 : Visual,” 2nd Edition, 2001.
- [40] 山口, “MPEG-4の任意形状オブジェクト符号化技術,” 東芝レビュー Vol.57, No.6, pp.10-13, 2002.
- [41] Vetro A., Haga T., Sumi K., and Sun H., “Object-Based Coding for Long-Term Archive of Surveillance Video,” IEEE ICME, Vol. 2, pp. 417-420, July 2003
- [42] 西, 藤吉, 梅崎, “ピクセル状態分析による固定監視映像の圧縮,” SSII2004 講演論文集, pp.397-402, 2004.
- [43] 羽下, 鷺見, 八木, “動き情報の時空間双方向追跡による監視映像のオブジェクト符号化,” 画像電子学会論文誌 Vol.34, No.4, pp.379-386, 2005.