

## 3層ニューラルネットワークと 変形テンプレート法による動画像の人物検出

館 俊太<sup>†</sup> 武藤佳恭<sup>††</sup>

動画像から人物を検知する画像処理システムを提案する。カメラを用いた自動検知システムはノイズの多い屋外の監視環境では誤判別を起こすことが多く、多様な監視環境下でも正確に作動するシステムの実現が望まれている。本手法では人物の見かけの大きさや姿勢、部分的な遮蔽、検知対象外の移動物体の混入、照明の急激な変動などに左右されない判別を実現するために、以下の手法を組み合わせて用いた。まず動物体の領域を抽出するためにフレーム間の差分を計算し、これを輪郭の画像に変換する。変形テンプレート法を用いて複数のテンプレートとこの輪郭とを比較する。比較した結果すべてを入力として3層ニューラルネットワークに与え、教師付き学習をさせて判別を行う。本手法は個々の移動物体を追跡して同定することはせず、画像中の侵入者の有無のみを輪郭形状を手がかりに判定する点を特徴とする。屋内と屋外の様々なシーンで実験を行い、手法の有効性を示した。

### Detecting Human Shapes in Video Images by Three-layered Neural Network and Deformable Templates

SHUNTA TATE<sup>†</sup> and YOSHIYASU TAKEFUJI<sup>††</sup>

We propose a visual surveillance system for detecting an intruding person in video images. In this system, a boundary contour image of moving objects is extracted from inter frame difference and it is compared with several templates of human body outline based on deformable template method. After the template matchings, all likelihood values of these matchings are given to a three-layered neural network and the network determines whether the video frame contains a person or not. The aim of the system is neither tracking nor identifying individual moving objects, but rather discriminating video frames that contain an outline of a person from those that don't. The experimental result shows that the system can distinguish an intruding person from other moving objects and can be tolerant of sudden illumination changes, occlusion and body posture variation.

#### 1. はじめに

動画像を用いた侵入者の自動検知システムは、監視の業務を行う人間の負担軽減が期待できるうえ、設置条件の自由度の高さやコストの面でも有利なため、市場の潜在的需要が大きい。関連研究も広範にわたり、様々な画像処理技法を用いたシステムが提案されている。使用するカメラも提案手法のように単一カメラを使う手法から、複数カメラを用いる手法<sup>1)~3)</sup>まで様々だが、ここでは設置が容易でコストの安い単一カメラを用いた研究<sup>4)~11)</sup>を中心に概説する。単一カメラによる屋内と屋外の動画像の例を図1に示す。

提案手法と従来手法について、各々がベースとしている画像処理技法と検出対象が複数か単一か、監視環境が屋内か屋外かなどを表1にまとめた。[ただし表1中で略されている記号はそれぞれ、(I) 屋内の監視、(O) 屋外の監視、(S) 単独の人物を検出可能、(M) 複数の移動物体が重なっていても人物の検出が可能、を意味する。]

オプティカルフローによる方法<sup>4),5)</sup>は、樹木など反射率の変動の大きい屋外シーンでの対象の追跡に効果的であるが、計算コストが大きいという特徴を持つ。また形の判定を別に行うなどして人物と人物以外の移動物体をなんらかの方法で分離する必要がある。

伊藤らのモデル<sup>6)</sup>はテンプレートを上下2つに分けることによって、見た目の大きさが変動する人物や車などの対象に対しても1組のテンプレートで検出できるような工夫をしている。ただしテンプレートの形状が固定されているため、ある程度以上の姿勢の変化には追従できない。

<sup>†</sup> 慶應義塾大学大学院政策・メディア研究科  
Graduate School of Media and Governance, Keio University

<sup>††</sup> 慶應義塾大学環境情報学部  
Faculty of Environmental Information, Keio University

表 1 従来手法と提案手法の特性の比較

Table 1 Classification of previous systems and proposed system according to their methods and advantages.

System	Method of motion detection	Method of human shape recognition	Operation area <sup>1</sup>	Recognition target <sup>2</sup>	Tolerance of occlusion and posture variation	Tolerance of drastic illumination change	Tracking
<b>Proposed model</b>	<b>Inter frame difference</b>	<b>Multiple templates &amp; Neural network</b>	I/O	S/M			-
Ito, et al. <sup>6)</sup>	Inter frame difference	Splitted single template	I/O	S			
Tabb, et al. <sup>8)</sup>	Inter frame difference	Active Contour model & Neural network	I/O	S			
Sakuma, et al. <sup>9)</sup>	Inter frame difference	Pixel labelling & Object size analysis	I	S			
W <sup>4</sup> <sup>10)</sup>	Background subtraction	Silhouette analysis & Histogram analysis	I/O	S/M			
Pfinder <sup>11)</sup>	Background subtraction	Body parts analysis	I	S			
Iketani, et al. <sup>5)</sup>	Optical flow	-	I/O	-	-		

<sup>1</sup> (I)...Indoor scene (O)...Outdoor scene

<sup>2</sup> (S)...Single isolated object (M)...Multiple overlapped objects



図 1 屋内と屋外の動画の例

Fig. 1 Examples of video images of an indoor and an outdoor scene.

一種のテンプレートマッチング法である動的輪郭モデル<sup>7)</sup>を用いた Tabb らによる研究<sup>8)</sup>は、輪郭を動物体に収束させた後、外周の輪郭線の離心率で形状を数値化する。次にこの数値をニューラルネットワークに入力として与えて人物と人物以外の移動物体を判定している。この方法では複数の歩行者が重なった場合に、シルエットが単独の歩行者の一般的な形状から大きく変化してしまうために誤判別を起こしやすい。そのため表 1 では判別対象は単独の人物となっている。また、動物体の領域をある程度綺麗に切り出せるという前提で動的輪郭を収束させているので、急激な照明変動が起こると歩行者の周囲の変動領域が混入してしまうと判別できなくなる。そのため表中では急激な照明の変動についての耐性がやや劣るとしている。

提案手法と同じくフレーム間差分を用いた佐久間らの手法<sup>9)</sup>では、差分画像から画像中の動物体の外接矩形を推定し、時間的な前後の関係からパラメータの判定を行って人物か否かを判別する。ただし形状を考慮

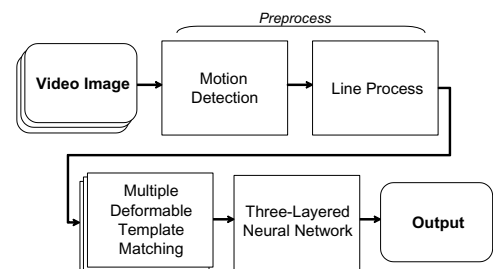


図 2 提案システムの構成

Fig. 2 Diagram of proposed system.

していないことと、人物と人物以外の小動物の判別に見かけの大きさを用いているため、動物体が重なって出現したり、照度変化などの外乱要因の大きな環境では正確に領域が同定できずに、物体の距離や大きさの判定を誤る可能性がある。

Haritaoglu らの W<sup>4</sup> と呼ばれる監視システム<sup>10)</sup>では背景の輝度の統計モデルを保持してこれを更新することで、前景の移動物体の領域を精度良く切り出している。また移動物体のシルエットを分析することで人物の姿勢を推定しているが、W<sup>4</sup> のように背景モデルを用いた手法は、急激な全体的な照明の変化が起こったときに、背景モデルのパラメータが更新されるまでの間動作が不安定になることがある。

Wren らの提案した Pfinder<sup>11)</sup>では、人体を複数の部分に分けて人物の姿勢を精度良く推定しているが、ノイズの多い屋外画像や複数人体には適さない。

ここで図 2 に提案手法の概略を示す。本手法ではまずフレーム間差分などの前処理をして移動物体の輪

郭を抽出し，変形テンプレート法<sup>12),13)</sup>により複数テンプレートとの比較を行い，ニューラルネットワークで結果を最終的に判別する．本手法の前提となった変形テンプレート法は，回転や形状の変形をともなった複数の対象図形を高い精度で検出する成果を出している．一方，侵入者検知システムは対象とする人間の姿勢や形状が多様なので1枚のテンプレートでは対応できない．そこで複数の変形テンプレートを用いるとすると，その結果をニューラルネットワークで総合的に判定するところに提案手法の新規性がある．また動物体の追跡(トラッキング)を行う他のいくつかの手法と異なり，本手法は正確な物体の追跡を行わない．本手法では，侵入者すべてを同定することは目的にせず，各フレーム画像中に侵入者が1人以上存在するかどうかを判定することを主眼としている．

また同じくニューラルネットワークを用いて形状を判定する Tabb らの手法と比較すると，Tabb らの手法では人物が別の動物体と重なったり周囲のノイズに接すると，外周の離心率が大きく変化するのに対して，提案手法が手がかりとする輪郭形状の線分はそれほど変化しないため，このような外乱の多い状況下では提案手法の方がより安定して判別できる特長がある．

## 2. 手法の概要

### 2.1 フレーム間差分と輪郭線の抽出

前処理としてフレーム間差分の計算と輪郭線の抽出を行う(図3)．

動画の各フレームを  $t_0, t_1, t_2, \dots, t_n$  とする．前後のフレーム間で輝度の差の絶対値をとって差分画像  $s_0, s_1, s_2, \dots, s_{n-1}$  を求める．ここで， $t_0$  と  $t_1$  の差分画像を  $s_0$ ， $t_1$  と  $t_2$  の差分画像を  $s_1$  とする． $s_0$  と  $s_1$  の画像間の積を求めることによって  $t_1$  の画像中の動部分のみを抽出した画像が得られる．

上の処理で得られた画像にラプラシアンガウスフィルタ  $\nabla^2 G$  をかけて平滑化と微分処理を行う．次に零交点(zero cross)を求めてエッジを抽出し，最終的に輪郭のみの二値画像を得る．

### 2.2 変形テンプレート法

変形テンプレート法<sup>12),13)</sup>は，局所変形する二次元のテンプレートを用いて緩和計算を行ってテンプレートマッチングを行う手法である．対象への近似の度合いと，収束した時点の変形の大きさによってテンプレートと対象画像が一致するかどうかの尤度(評価値)が決まる．図4にテンプレートマッチングが収束した例を示す．図中の数値  $L$  はテンプレートと画像の一致の評価値である．

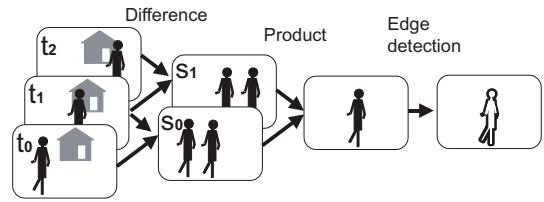


図3 フレーム間差分と輪郭線の抽出の模式図

Fig. 3 Diagram of computation of inter frame difference and edge detection.

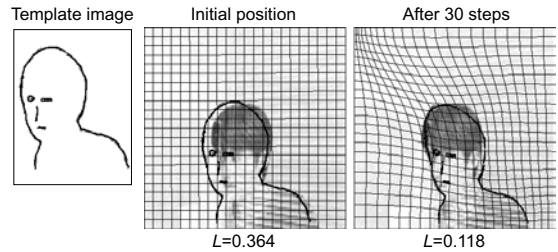


図4 変形テンプレートが収束した例

Fig. 4 Example of template matching process, which displays the deformable template successfully fitting the given image.

テンプレート画像を二次元空間  $[0, 1]^2$  上の輝度値の集合とする．この二次元空間上に変形のための関数  $D_\xi(x, y) : R^2 \rightarrow R^2$  を次のように定める．

$$D_\xi(x, y) = \sum_{m=1}^M \sum_{n=1}^N \frac{\xi_{mn}^x \cdot e_{mn}^x + \xi_{mn}^y \cdot e_{mn}^y}{\lambda_{mn}}, \quad (1)$$

$$e_{mn}^x(x, y) = (2 \sin(\pi n x) \cos(\pi m y), 0),$$

$$e_{mn}^y(x, y) = (0, 2 \cos(\pi m x) \sin(\pi n y)). \quad (2)$$

ただし， $\xi_{mn}^x, \xi_{mn}^y (m, n = 1, 2, \dots)$  は正規分布  $N(0, \sigma^2)$  に従う互いに独立なパラメータ， $\lambda_{mn}$  は  $\lambda_{mn} = \alpha \pi^2 (n^2 + m^2) (m, n = 1, 2, \dots)$  となる正規化のための定数である． $e_{mn}^x, e_{mn}^y$  は直交基底の関数で，三角関数を用いてスムーズな変形を可能にしている． $M$  と  $N$  は変形の直交基底関数の最大の次数である． $m, n$  は  $m = n = 1, 2, 3, \dots$  をとる添数で，変形の各段階に対応している．

パラメータ  $\xi_{mn}^x, \xi_{mn}^y (m, n = 1, 2, \dots)$  の集合を  $\xi$  と表す． $\xi$  の各要素に適当な実数を代入すれば，関数  $D_\xi(x, y)$  は二次元空間上で任意の連続な変形を実現することができる．なお，あるパラメータセット  $\xi' = \{(\xi_{mn}^x, \xi_{mn}^y), m, n = 1, 2, \dots\}$  の生起確率  $Pr(\xi')$  は下式で与えられる．

$$Pr(\xi') = \prod_{m,n=1}^{M,N} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{\xi_{mn}^x{}^2 + \xi_{mn}^y{}^2}{2\sigma^2}\right\}. \quad (3)$$

ここで、 $\sigma$  は  $\xi$  の確率分布の分散の大きさである。

テンプレート  $T$  を変形して得られる  $T_\xi$  と入力画像  $\Upsilon$  の類似度の評価値  $E$  を以下の式で与える。

$$E(T_\xi, \Upsilon) = \frac{1}{n_T} \sum (1 + \Phi(x, y) | \cos(\beta(x, y)) |). \quad (4)$$

ただし  $\Phi(x, y)$  は画像上の  $(x, y)$  の点におけるエッジのポテンシャルを表し、

$$\Phi(x, y) = -\exp\{-\rho(\delta_x^2 + \delta_y^2)^{\frac{1}{2}}\} \quad (5)$$

である。なお  $(\delta_x, \delta_y)$  は点  $(x, y)$  から最近傍のエッジまでの距離の成分であり、 $\rho$  はエッジどうしの対応をスムーズにするための正の係数である。 $\beta(x, y)$  はテンプレート上の  $(x, y)$  のエッジの傾きと入力画像上の最近傍のエッジの傾きとがなす角度、 $n_T$  はテンプレートの線分を構成するピクセルの数である。評価値  $E$  は、テンプレートが完全に画像に一致したときを 0 として 0~1 の間で得られる。

入力画像  $\Upsilon$  が与えられたときの、 $\xi$  の事後確率  $Pr(\xi | \Upsilon)$  はベイズ則に従って以下のように変形される。ただし  $\kappa$  は正規化のための係数である。

$$Pr(\xi | \Upsilon) = \frac{Pr(\Upsilon | \xi)Pr(\xi)}{Pr(\Upsilon)}, \quad (6)$$

$$Pr(\Upsilon | \xi) = \kappa \exp\{-E(T_\xi, \Upsilon)\}. \quad (7)$$

あるテンプレート  $T$  に最適な変形を施したときに得られる最大の類似度を求める問題は上式 (7) の最大事後確率推定である。式の変形の結果、さらに

$$L(T_\xi, \Upsilon) = E(T_\xi, \Upsilon) + \gamma \sum_{m=1}^M \sum_{n=1}^N (\xi_{mn}^x{}^2 + \xi_{mn}^y{}^2) \quad (8)$$

を導くことにより、最大事後確率推定は最終的に上式 (8) の右辺を最小化する  $\xi$  を求める問題として与えられる。ただし、 $L(T_\xi, \Upsilon)$  は変形の尤度も含めたテンプレートマッチングの評価値であり、最大の類似度を 0 とする。また  $\gamma = 1/2\sigma^2$  である。式 (8) の右辺を最小化するような  $\xi$  は解析的に求められないので、 $\xi$  による  $L$  の偏微分値を数値的に求め、下式の勾配法に従って  $\xi$  を更新する。

$$\Delta\xi = -\eta \frac{\partial L(T_\xi, \Upsilon)}{\partial \xi}. \quad (9)$$

ただし、 $\eta$  は正の微小な係数である。

### 2.3 テンプレート集合

複数のテンプレートをあらかじめ用意しておき、テンプレート 1 つ 1 つについて変形テンプレート法でテンプレートマッチングを行う。侵入者の判別に有効な

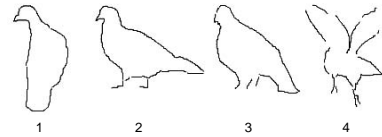


図5 クラス III) のテンプレートの例  
Fig. 5 Examples of class III) template.

テンプレートは主に

- I) 侵入者の輪郭形状
- II) 侵入者の部分的形状
- III) 侵入者以外の物体の形状

の 3 つのクラスに分類できる。クラス III) のテンプレートの形状としては、あらかじめ監視環境に混入する可能性が高いことが分かっているような小動物などがあげられる。クラス III) の具体的な例としてここでは鳩の形状のテンプレートの例を図 5 に示す。

### 2.4 3層ニューラルネットワーク

入力画像とすべてのテンプレートとの類似度をニューラルネットワークで教師付き学習させ、侵入者の有無を判別する。本手法では 3 層で全結合型の誤差逆伝搬法のニューラルネットワーク<sup>14)</sup>を用いた。

なお、テンプレートのセットにクラス III) のテンプレートを含める場合も、テンプレート自体を正事例と負事例に分けて学習させる必要はない。入力画像のデータについて教師値を正しく与えてさえいれば、後はニューラルネットワークがテンプレートマッチングの結果を自動的に学習し、侵入者を判別してくれる。

## 3. 実験の条件

### 3.1 前処理

CCD カメラによって得た動画を 500 msec 間隔でサンプリングし、16 シーンの合計 438 フレームについて、 $160 \times 120$  画素、256 階調のグレースケールのフレームとして得た。続いてフレーム間差分の計算および輪郭抽出の前処理を行った。図 10 の中段の画像が前処理の後得られた画像の例である。

### 3.2 テンプレートマッチング

今回の実験ではテンプレート画像として図 6 のように 17 種類のビットマップ画像を用意した。初めの 12 個がクラス I)、残り 5 個がクラス II) のテンプレートにあたる。今回はクラス III) のテンプレートは用いていない。テンプレートを作成するにあたっては、実験データとは別に撮影された人物の画像を任意に選び、輪郭線の中から高いコントラストの線分を手書きで書き出す方法を用いた。この生成方法は直感的で不確実性が強く、テンプレート間の冗長性も大きいので、判

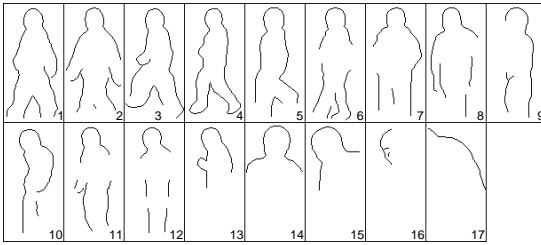


図6 作成したテンプレートのセット(各升目は  $67 \times 133$  画素)  
Fig. 6 Prepared template set (The size of each template's cell is  $67 \times 133$  pixels).

別に必要と思われる数をカバーするために17個と多少多めに用意している。なお、クラスI)のテンプレートの数を12個から多少変化させても精度に大きく影響することはないが、クラスII)のテンプレートの数を少なくすると遮蔽された人物の判別精度が大きく下がることが分かっている。

テンプレートマッチングは段階的に解像度を変えて行った。アルゴリズムは以下である。

- (1) テンプレートを画面上の  $1/8$  単位でシフトさせ、最も類似度の評価値の高い位置を探索する。これを3方向の回転(30度, 0度, -30度)と左右反転像の計6つの回転について行い、その中で最も評価値の高かった回転と位置を初期値とする。
- (2) 勾配法でテンプレートを変形する。
- (3) 開始時は  $M = N = 1$  の段階であり、以降処理が進むごとに  $M = N = 2$ ,  $M = N = 3$  とする。入力画像を3段階にぼやけさせるために式(5)の  $\rho$  の値を段階的に変更する。
- (4)  $M = N = 1$  の段階のみ勾配法でテンプレートの移動を行う。
- (5) 類似度の評価値  $L$  が閾値を超えれば次の段階の処理に進む。その前に繰返し数の上限を超えたらそこでテンプレートマッチングを終了する。
- (6) 収束したら終了。していなければ(2)に戻って計算を繰り返す。

なお、実験では人物のサイズの大小をカバーするために、縮小率を変えて1つのテンプレートにつき4つのサイズを用いた。結果的に  $17 \times 4$  で計68個の評価値  $L$  の値が得られる。これは1つの入力画像が  $R^{68}$  の特徴空間上の1点に変換されることを意味する。テンプレートマッチングの各パラメータを表2に記す。

なお、文献13)では12方向の回転画像を用い、初期位置を複数定めて入力画像中の複数の対象を同時に検出させているのに対して、提案手法では計算コストの理由で初期位置を1つしかとっていない。初期位置

表2 テンプレートマッチングのパラメータ  
Table 2 Parameter set of template matching.

各段階の繰返し計算の上限回数: (10, 15)
各段階の $L$ の閾値: (0.70, 0.50)
$\alpha = 0.18$ $\gamma = 0.2$ $\eta = 0.008$ $\rho = (12, 8, 3)$
縮小率: (1.0, 0.86, 0.7, 0.5)

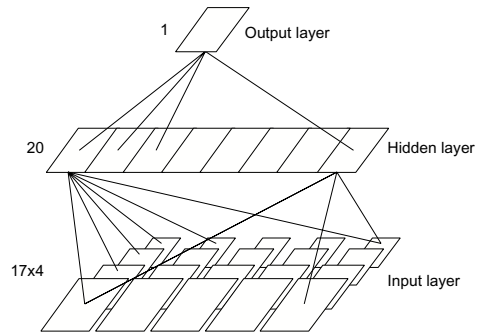


図7 ニューラルネットワークモジュールの構成  
Fig. 7 Diagram of the neural network component.

が1つのみだと結果が局所解に陥りやすくなるが、提案手法は複数のテンプレートでテンプレートマッチングを行うことと、侵入者をすべて同定するのではなく、侵入者が1人以上存在するかどうかを最終的に判別することが目的なので、このように簡略した手法でも十分目的を達成できる。

### 3.3 3層ニューラルネットワークによる学習

ニューラルネットワークを図7のように構成する。ニューロンの数はそれぞれ、入力層68、中間層20、出力層1とする。1つの入力画像につきテンプレートマッチングの評価値68個を入力値とし、人物が含まれているならば1、含まれていなければ0の教師値を与えて学習を行う<sup>14)</sup>。入力画像438フレーム中100フレームを選んでトレーニング用のデータセットとし、残り338フレームを検定用のデータセットとした。なおトレーニングデータ中50フレームに侵入者が含まれ、残り50フレームは侵入者が含まれない画像である。トレーニングデータは屋内と屋外の環境の両方を含むが、屋内では部屋の照明を明滅させることによって急激な照明の変化を学習させ、屋外では外乱要因となる樹木と天候変化による照明の変化のあるシーンを用いた。検定用のデータにも同様の外乱要因のある画像を用いたが、加えて、遮蔽の影響をテストするために検定用のデータのみ歩行者が昇降する階段付近や、歩行者の体が半分以上隠れるような家屋の門などのシーンを含めた。

図8を見ると、中間層の数は  $\approx 8 \sim 20$  程度の範囲なら局所解に陥らず、いずれも問題なく学習が収束す

表3 屋内画像と屋外画像における判別の精度

Table 3 Detection accuracy in terms of indoor scenes and outdoor scenes.

	Operating on indoor images		Operateing on outdoor images	
	Positive example data	Negative example data	Positive example data	Negative example data
Detection success / Trial	101 / 108	88 / 92	49 / 72	62 / 66
Detection accuracy	94.50%		80.43%	

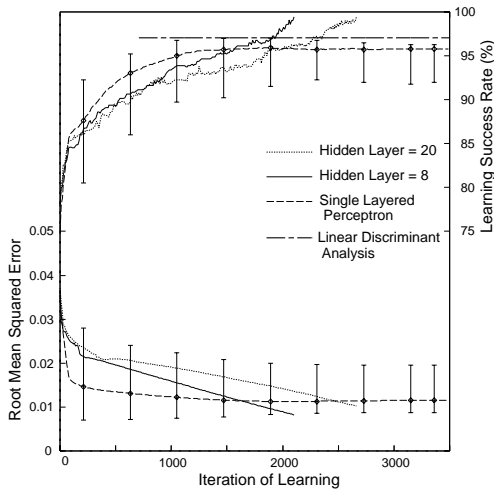


図8 誤差逆伝搬法と単層パーセプトロンによる学習曲線

Fig.8 Learning curve and training error of back-propagation methods and a single-layer Perceptron.

ることが分かる．今回は今後の事例の追加などを考えて多めに  $m = 20$  とした．同図より単層パーセプトロンや線形判別分析手法では実験データを完全には学習できていないため，学習には非線形の識別面が必要であることが分かる．

## 4. 実験結果

### 4.1 判別の精度

学習終了後，検定用のデータセットを用いて侵入者を判別させた．判別の閾値は 0.5 とした．表 3 に屋内画像と屋外画像それぞれの判別精度を示す．屋内画像では 94%，屋外画像全体では 80%程度が正しく判別できた．なお，表中の Positive example data と Negative example data はそれぞれ正事例（侵入者あり）と負事例（侵入者なし）のデータを意味する．

図 9 に判別の閾値を動かしたときの判別精度の変化を示す．なお，図 9 中の Detection accuracy は 338 の検定データのうち，侵入者の有無を正確に判別できたデータの割合，False positive rate は侵入者のないデータのうち，誤ってありと判定されたデータの割合，False negative rate は侵入者のあるデータのうち，誤ってなしと判定されたデータの割合，をそれぞ

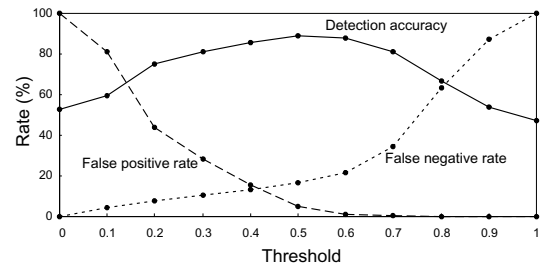


図9 閾値の変化による判別の精度

Fig.9 Detection accuracy with respect to threshold.

れ意味する．学習時に 0.5 を閾値として学習させたため，0.5 をピークに最も判別の精度が高いという結果になった．

### 4.2 照明光源の変化に対する耐性

図 10 に実際の判別の例を示す．画像の上段は入力画像，中段は差分画像から抽出した輪郭線の画像，下段はそれぞれの画像に  $L < 0.3$  の類似度で一致したテンプレートのうち，最大の類似度のものを示した．ただし，6 番目のフレームに関してのみマッチングしたテンプレートを複数示している．各列の下の数値はニューラルネットワークの出力値であり，この値が 0 に近ければ侵入者がいないと，1 に近ければ侵入者がいるとニューラルネットワークが判定したことを意味する．

この例では監視中に室内の電灯を消して，照明条件を急激に変化させた．そのため 6 番目のフレームで背景の線分のすべてが変動領域として抽出されている．しかしニューラルネットワークはあらかじめこのような画像を学習しているので，0.055 と低い値を出力しており，人物がいないことを正しく判定できている．差分画像を用いた従来手法では照明の急激な変動時に誤検出を起こすことが多かったため，これを学習によって正しく判別できるのは提案手法の特長といえる．

### 4.3 代表的な判別の成功例と失敗例

図 11 に代表的な判別の成功例と失敗例を示す．ニューラルネットワークが侵入者ありと判定した画像については最大の類似度で適合したテンプレートの位置を四角の枠で示した．

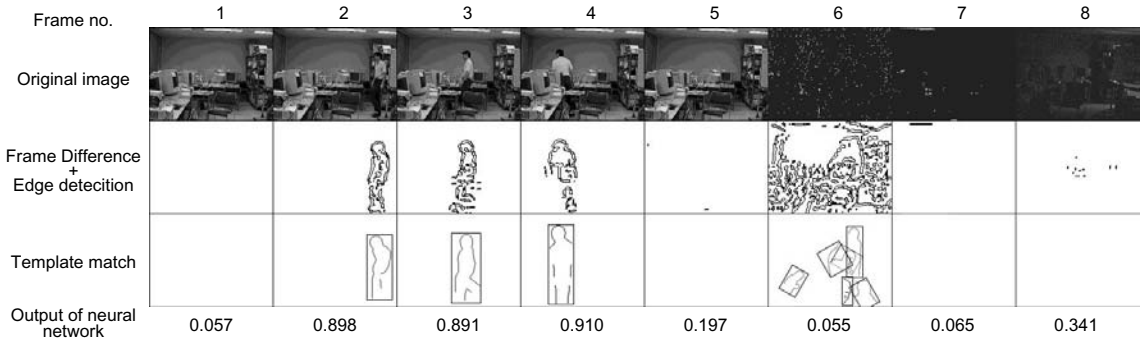


図 10 実験例．画像の上段は入力画像，中段は抽出された輪郭線の画像，下段はそれぞれの画像に適合したテンプレート

Fig. 10 Experimental result. The top row of the images shows the original video frames, the middle row shows boundary contour images, the bottom row shows deformable templates fitting respective frames.

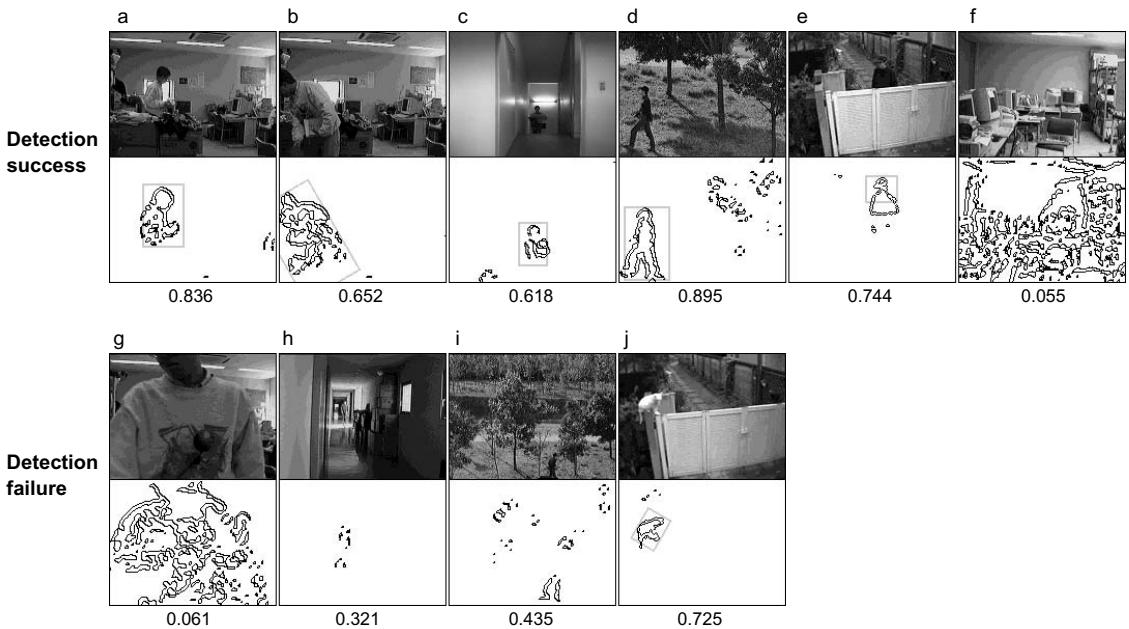


図 11 代表的な判別の成功例と失敗例

Fig. 11 Examples of detection success and detection failure.

成功例では図中 a, c, e の遮蔽された人物, b の姿勢の変動のある人物, d の樹木の反射の変動によるノイズを含む屋外画像中の人物, f の室内照明を消灯・点灯したときに生じる急激な変動, をそれぞれ正しく判別している。

失敗例では g は人物が近すぎたため, h は人物が遠くで見かけの大きさがどのサイズのテンプレートにも適合しなかったため, i は人物の見かけの大きさが小さいことと樹木の反射の変動によって頭部が切れ, 樹木と区別がつかなくなってしまっているため, それぞれ人物を検出できなかった。また j は動物(猫)の輪

郭形状が人物の遮蔽された上半身の形状に似ているため, 侵入者と誤って判定してしまった例である。

#### 4.4 変形テンプレート法の効果

表 4 に変形テンプレートを用いた場合と, テンプレートをいっさい変形させずに回転と位置だけを変えてテンプレートマッチングして判別させた場合の結果を比較して示す。判別には屋内と屋外の両方の画像を用いた。変形テンプレート法を用いた場合, 判別の精度が 10% 以上良いことが分かる。

表4 変形テンプレートを使った場合と使わなかった場合の判別の精度

Table 4 Detection accuracy in terms of deformable templates and normal templates.

	Using deformable templates		Using normal templates	
	Positive example data	Negative example data	Positive example data	Negative example data
Detection success / Trial	150 / 180	150 / 158	114 / 180	146 / 158
Detection accuracy	88.76%		76.92%	

## 5. 考 察

画像による侵入者検知システムの外乱要因として、主に表5のような項目が知られている<sup>9)</sup>。提案手法で用いたフレーム間差分による方法は、背景モデルを保持する手法<sup>10),11)</sup>に比べると、表5の3),4)の外乱要因によるノイズが混入しやすい<sup>9)</sup>。また、表5の1)の外乱が起きるとフレーム間で大きな差異が生じるため、一時的に監視領域の全域で変動部分が抽出されることがある。

検知対象である人物が画像中、どのような姿勢と位置と見かけの大きさで出現するかをあらかじめ決定することはできないので、このような外乱要因が増せば増すほど、また検知対象の形状の自由度が増せば増すほど、両者の区別が難しくなる。従来手法はこの困難を避けるために、特定の対象のみを検知対象とするか、動作環境を限定することで精度を上げることが多かった。たとえば図1上段の通路のように歩行者の見かけの大きさが変化しない角度と位置にカメラの設置場所を制限したり、Pfinder<sup>11)</sup>のように人間以外の動物が監視領域に侵入しない前提で設計するような方法がこれにあたる。本研究では、テンプレートマッチングとニューラルネットワークの形状認識の処理によって、侵入人物が画像中に起こす変動とこれらの外乱とを区別することができるので、特定の制約を行わずに外乱要因の多い環境でも侵入者を検出できた。

また提案手法は他の手法と異なり、動物体の正確な追跡を行わない。代わりに、変動の起こった領域のあらゆる場所で人物らしき輪郭形状がないかどうかパターンマッチングすることに主眼を置いている。そのため結果的に、部分的に遮蔽された人物、重なった複数物体の移動、前処理で排除できなかった樹木のノイズなどがある状況下でも侵入者を検出することができる。対して、 $W^4$ などのシステムでは動物体があるかを正しく判別するためには、背景から動物体のシルエットが正確に切り出せていることが前提となっている。この点に相違がある。

また、監視システムの運用時には、設置環境の特徴として図11jの例のように猫や鳥といった既知の特

表5 侵入者検知システムの外乱要因の分類

Table 5 Classification of external noises which deteriorate the accuracy of surveillance system.

1)	照明光源の変化	雲の動きや日周変化などによって起こる照度の変化、室内監視時に照明を点灯/消灯したときに起こる画像全体に生じる変化
2)	移動光源による光の入射	巡回する警備員の懐中電灯や窓の外の車のヘッドライト
3)	影や反射の変動	侵入物体やそれ以外の物体の影や反射率の変動
4)	監視領域中の既存物の動き	風になびく樹木やカーテン、切れかかった蛍光灯の明滅など 1)よりも狭い範囲に現れる変動
5)	監視中の背景の変化	駐車車両が動き去った場合などに起こる背景の変化
6)	監視対象以外の移動物体の混入	小動物や風に飛ばすゴミ袋やカメラの傍を飛ばす昆虫等

表6 計算時間の平均

Table 6 Average of the runtime.

CPU: Pentium III 667 Mhz		
	Average time	Standard deviation
Preprocess	0.10 sec	0.034 sec
Template matching	0.90 sec	0.54 sec
Neural network	0.0091 sec	0.00036 sec
Total	1.01 sec	0.54 sec

定の外乱要因が高い確率で出現することがあらかじめ分かっているような場合がある。その場合、猫や鳥の形状のビットマップのテンプレートを複数書き出してつけ加え、ニューラルネットワークを再学習させることにより、判別の精度を向上させることができる。このように検出対象以外の負事例の対象についてもテンプレートを用意できる点に提案手法の利点がある。

## 6. おわりに

侵入者検知のための手法を提案した。提案手法は輪郭形状を手がかりとしてテンプレートマッチングと学習を行うため、人間の姿勢の変化や身体の一部の遮蔽、または照明の消灯・点灯によって起こる背景の変化に対処できることが示された。

また提案手法は監視対象の形状特徴をテンプレート



として可視化して保持している点に特徴があり，監視対象以外の負事例の物体についてもテンプレートを用意して判別の精度を上げることができる．

また，1フレームにかかる処理時間を示した表6を見ると，システム運用時の計算時間の大部分をテンプレートマッチングの計算が占めている．今後，並列処理による高速化や，冗長性を小さくするために統計的な手法でテンプレートを取捨選択するなどの改善が必要である．

### 参考文献

- 1) Collins, R., Lipton, A., Fujiyoshi, H. and Kanade, T.: Algorithms for cooperative multi-sensor surveillance, *Proc. IEEE*, Vol.89, No.10, pp.1456-1477 (2001).
- 2) Stauffer, C. and Tieu, K.: Automated multi-camera planar tracking correspondence modeling, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Vol.1, pp.259-266 (2003).
- 3) Khan, S., Javed, O., Rasheed, Z. and Shah, M.: Human tracking in multiple cameras, *Proc. 8th IEEE International Conference on Computer Vision (ICCV'01)*, Vol.1, pp.331-336 (2001).
- 4) 谷内田正彦：ロボットビジョン，8章，昭晃堂，東京(1990).
- 5) Iketani, A., Nagai, A., Kuno, Y. and Shirai, Y.: Real-Time Surveillance System Detecting Persons in Complex Scenes, *Real-Time Imaging*, Vol.7, No.5, pp.433-446 (2001).
- 6) 伊藤 渡，上田博唯：物体の見かけの大きさ変動に追従できるテンプレートマッチング法，信学技報 (PRMU 96-43)，Vol.96, No.141, pp.45-50 (1996).
- 7) Kass, M., Witkin, A. and Terzopoulos, D.: Snakes: active contour models, *Int. J. of Computer Vision*, Vol.1, No.4, pp.321-331 (1987).
- 8) Tabb, K., Davey, N., Adams, R. and George, S.: The Recognition and Analysis of Animate Objects using Neural Networks and Active Contour Models, *Neurocomputing*, Vol.43, pp.145-172 (2002).
- 9) 佐久間喜朗，伊東 潔，増田 功：フレーム間差分を用いた侵入物体検出法，テレビ学技報，Vol.14, No.49, pp.1-6 (1990) .
- 10) Haritaoglu, I., Harwood, D. and Davis, L.S.: W4: real-time surveillance of people and their activities, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No.8, pp.809-830 (2000).
- 11) Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A.: Pfunder: Real-time tracking of the human body, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.780-785 (1997).
- 12) Amit, Y. Grenander, U. and Piccioni, M.: Structural image restoration through deformable template, *J. of the American Statistical Association*, Vol.86, No.414, pp.376-387 (1991).
- 13) Jain, A.K., Zhong, Y. and Lakshmanan, S.: Object matching using deformable templates, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.18, No.3, pp.267-277 (1996).
- 14) Rumelhart, D.E. and McClelland, J.L.: *Parallel Distributed Processing*, The MIT Press, University of California, San Diego (1986).

(平成 15 年 2 月 5 日受付)

(平成 15 年 3 月 31 日採録)



館 俊太 (正会員)

1997年慶應義塾大学環境情報学部卒業。1999年同大学院修士課程了(政策・メディア)。2003年同大学院博士課程了(学術)。現在、大阪大学生命機能研究科研究員(COE)。ニューラルネットワーク、視覚脳機能のモデルとデータ解析の研究に従事。



武藤 佳恭

慶應義塾大学工学部電気工学科卒業(1978)，同大学院修士課程修了(1980)，同博士課程修了(1983)。工学博士(1983)。南フロリダ大学コンピュータ学科助教授(1983~1985)，南カロライナ大学コンピュータ工学助教授(1985~1988)，ケースウエスタンリザーブ大学電気工学科准教授(1988~1996)。tenured受賞(1992)。慶應義塾大学環境情報学部助教授(1992~1997)，同大学教授(1997~現在)。研究：ニューラルコンピューティング，セキュリティ。NSF-RIA賞(1989)，IEEE Trans. on NN功労賞(1992)，IPSJ論文賞(1980)，TEPCO賞(1993)，KAST賞(1993)，高柳賞(1995)，KDD賞(1997)，NTT tele-education courseware賞(1999)。政府顧問：NCC(フィリピン)，VITTI(ベトナム)，CTTISC(ジョルダン)，タイ，スリランカ，マルチメディア大学(マレーシア)。22冊の本と200編以上の論文。