

異常検知のための自己組織化モデルと その監視映像への適用

菅沼 雅徳^{1,2,a)} 長尾 智晴¹

受付日 2015年9月3日, 再受付日 2015年10月20日,
採録日 2015年11月16日

概要: 本論文では, 監視映像中における異常検知のための環境に適応的な自己組織化モデルを提案する. 提案するモデルでは, 監視映像からの入力刺激に応じてモデル内のノードを生成, 削除, またノードの更新を行うことで映像内の正常性を表現するモデルを構築する. そして構築されたモデル内のノードの状態に基づき, 映像内の正常度を算出することで異常検知を行う. 実験では, 固定カメラと水平方向に等速旋回する旋回カメラから撮影された2種類の監視映像を用いて, 映像内に現れる歩行者と車両を検出対象とした侵入物体検知問題を扱い, 提案するモデルの性能評価を行った.

キーワード: 異常検知, 侵入物体検知, 監視カメラ, 動画画像処理

A Self-Organising Model for Anomaly Detection and Its Application to Video Surveillance

MASANORI SUGANUMA^{1,2,a)} TOMOHARU NAGAO¹

Received: September 3, 2015, Revised: October 20, 2015,
Accepted: November 16, 2015

Abstract: In this paper, we propose a self-organising model for anomaly detection in surveillance videos. The model is used to model normal patterns in video scenes and detects deviations from them as abnormal patterns. The model can add nodes into its map and delete nodes from its map for representation of normal patterns more accurately. To demonstrate the effectiveness of our method, we have applied our method to intrusion detection tasks in surveillance videos obtained by a fixed camera and a rotating camera.

Keywords: anomaly detection, intrusion detection task, surveillance camera, video processing

1. はじめに

近年, 防犯の目的で監視カメラの普及が進んでいる. その一方で, 監視カメラ映像の多くは人間の目視によって確認されているため, 膨大な労力が必要となっている. そこで, 監視者の負担軽減のため, 計算機を用いた映像内の人物の異常動作や, 映像内への侵入物体などの異常検知の自

動化に関する研究が数多く行われている.

異常検知に関する先行研究では, 映像内での出現頻度が低いパターンや, 正常パターンと特徴が大きく異なるパターンを異常と定義して, それらを検出する手法が多い [1], [2], [3]. そのため, 多くの手法では事前に異常を含まない正常パターンから映像内の正常モデルを構築し, その正常モデルにおける生起確率が低いパターンや, モデルを逸脱するパターンを異常として検出している. 本論文でも多くの先行研究と同様に, 映像内での出現頻度が低いパターンや, 正常パターンから逸脱するパターンを異常と定義する.

南里らは立体高次局所自己相関特徴 (Cubic High-order

¹ 横浜国立大学大学院環境情報学府
Graduate School of Environment and Information Sciences,
Yokohama National University, Yokohama, Kanagawa 240-
8501, Japan

² 日本学術振興会特別研究員 DC
JSPS Research Fellow

^{a)} suganuma-masanori-hf@ynu.jp

Local Auto-Correlation; CHLAC) 特徴 [4] を用いて、人物の歩行動作を正常動作、走行動作と転倒動作を異常動作と定義した実験において、複数の人物が映る動画像から異常動作を検出することに成功している [5]. Xiang らは屋内と屋外で撮影された動画像内での人物の行動パターンについてのモデルを学習し、学習データにはなかった行動パターンを異常として検出する手法を提案している [6]. また、近年では人が多く存在する混雑シーンにおける異常検知の研究 [7], [8], [9], [10] もさかんに行われており、Mahadevan らは mixture of dynamic textures [11] を用いて時空間特徴に関する正常モデルを構築し、歩道の混雑シーンにおいて車両などの異常を検出することができることを示している [12]. これら多くの先行研究では事前に構築された正常モデルが固定であり、異常検知の適用中に更新されることはない。しかし、現実世界での運用を考えると、風によって樹木などが揺れることで正常パターンが変化したり、学習データには出現しなかった正常パターンが出現する場合など、事前に構築した固定の正常モデルだけでは正確に映像内の正常性を表現することが難しいと考えられる。したがって、異常検知をより頑健に行うには、適用中の環境に応じてモデルの更新が行われる環境に適応的な手法が望まれる。そこで本論文では、異常検知の適用中にモデルの更新が行われる手法を提案する。

環境に適応的な異常検知手法の例としては Masland らの Grow When Required (GWR) ネットワークがあげられる [13], [14], [15]. GWR ネットワークでは頻繁に観測される入力刺激に対して、馴化モデルである Stanley モデル [16] を用いてネットワークの出力を次第に減少させていくことで環境の正常性を表現する。GWR ネットワークを移動ロボットに搭載し、ソナーセンサ情報を入力刺激として扱った実験では、ロボットがそれまで観測していた環境とは異なる環境に置かれるとネットワークが強い出力を示すことが確認されている。また、Nehmzow らは GWR ネットワークの入力刺激に画像情報を用いた異常検知実験を行っている [17], [18], [19]. しかし、GWR ネットワークは単純な構造のため、環境に変動が含まれるような場合に対して適用することは難しいと考えられる。武田らは環境からの入力パターンに対する反応とそのパターンが生じる領域に対する反応の抑制によって、環境に変動を含む監視映像内の侵入物体を検知するネットワークモデルを提案している [20]. Stauffer らは画素ごとに混合ガウス分布を用いて環境の背景のモデル化を行っている [21]. この手法では、混合ガウス分布を用いることで背景の揺らぎなどを多峰性の分布で表現しつつ、パラメータ更新によって環境変化への適応も実現している。これらの手法は固定カメラからの映像を用いた侵入物体検知問題に対して優れた性能を示している [20]. しかし、武田らの手法では映像内の位置関係を環境の正常性を表現するために利用していること、

Stauffer らの手法では複雑な入力パターンを多峰性の分布で表現することが困難なことから、映像内の背景が周期的に変動する旋回カメラなどで撮影された映像に対して適用することは困難が予想される。旋回カメラによる異常検知が可能となれば、固定カメラより広範囲な監視を行うことができたり、決まったルートを巡回する警備ロボットへの応用が行えるなど、より多種類の場面で適用することが可能となりさらなる有用性が期待できる。

そこで本論文では、固定カメラと水平方向に等速旋回する旋回カメラから撮影された2種類の監視映像を対象に異常検知を行う、より環境への適応性が高い自己組織化モデルを提案する。提案するモデルは、環境からの入力刺激に応じてノードの生成、削除、またノード状態の更新を行うことで環境の正常性を表現する。実験では固定カメラと水平方向に等速旋回する旋回カメラから撮影された2種類の監視映像を用いて、映像内に現れる歩行者と車両を検出対象とした侵入物体検知問題に提案手法を適用し、従来手法と比較することで性能の検証を行う。

2. 自己組織化モデルによる異常検知

2.1 概要

提案する異常検知モデルの構造を図 1 に示す。提案手法では、同一構造の自己組織化モデルが入力画像中に格子状に整列している。本論文では $m \times n$ 画素に対して1つのモデルを配置しており、たとえば入力画像サイズが $M \times N$ 画素の場合、入力画像中には $L = \frac{M}{m} \times \frac{N}{n}$ のモデルが配置される。

各モデルは複数のノードによって構成される。ノード i は、重みベクトル v_i 、時刻 t における馴化係数 $h_i(t)$ 、年齢 $age_i(t)$ を持ち、重みベクトルは環境からの入力刺激と同次元数で各要素が $[0.0, 1.0]$ の実数値、馴化係数はノードが持つ重みベクトルと類似した入力刺激の出現頻度を表す $[0.0, 1.0]$ の実数値、年齢はノードの年齢を表す非負の整数値である。

提案手法では、入力画像中の各格子領域から算出した画像統計量を入力刺激として対応する各モデルに入力する。そして、モデル内でその入力刺激と類似した重みベクトルを持つノードの選択を行い、選択されたノードの年齢と馴化係数の更新を行う。さらに選択されたノードの入力刺激

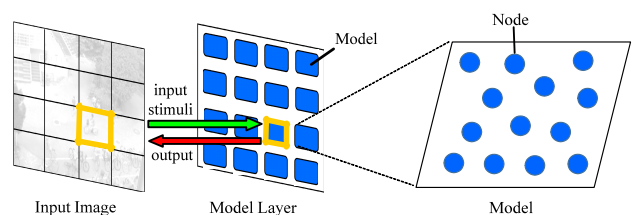


図 1 提案手法の構造

Fig. 1 The structure of the proposed method.

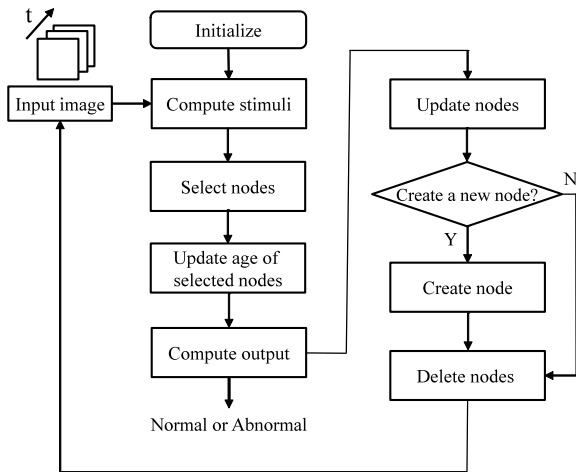


図 2 提案手法の処理の流れ

Fig. 2 Flowchart of the proposed method.

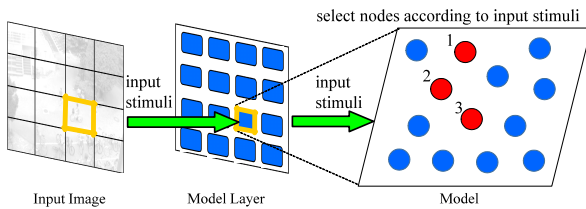


図 3 ノード選択の例 ($S = 3$ の場合). ノードの左上の数字は入力刺激に対する類似度の大きさの順位である

Fig. 3 An example of node selection ($S = 3$). A figure indicates rank of similarity of input stimuli.

との類似度と馴化係数に応じてモデル内へのノード追加や、ノードの年齢に応じてノード削除を行うことで環境の正常性を表現する構造を構築していく。そして、ノードの類似度を用いて環境内の刺激に対する応答値をモデルごとに非負の実数値で出力することで異常検知を行う。

2.2 処理の流れ

図 2 に提案手法の処理の流れを示す。処理の詳細については次のとおりである。

- 1 すべてのモデルを次の手順で初期化する。
 - (a) モデルに I 個のノードを生成する。
 - (b) 各ノードの重みベクトルをあらかじめ算出した入力刺激からランダムに選択する。
 - (c) 各ノードの馴化係数 $h_i(0)$ を 1.0 にする。
 - (d) 各ノードの年齢 $age_i(0)$ を 0 にする。
- 2 すべてのモデルについて、入力動画の 1 フレームごとに以下の処理 (a) から (h) を繰り返す。
 - (a) モデルが配置された格子領域から入力刺激を算出し、モデルに入力する。入力刺激は格子領域内の画素値から算出した平均、中央値などの画像統計量を用いる。すべての入力刺激は $[0.0, 1.0]$ に正規化される。
 - (b) モデル内のすべてのノードについて、図 3 に示すように、入力刺激との類似度が大きい上位 S 個のノ

ドを選択する。ノード i における類似度 D_i の算出には式 (1) を用いる。

$$D_i = \exp(-\|\mathbf{x} - \mathbf{v}_i\|) \quad (1)$$

ここで、 \mathbf{x} は入力刺激、 \mathbf{v}_i はノード i が持つ重みベクトル、 $\|\cdot\|$ はノルムである。式 (1) は、類似度 D_i が大きいノードほど入力刺激と類似した重みベクトルを持っていることを示す。

- (c) 選択された S 個のノードの年齢 $age_i(t)$ を 0 にする。式 (2) における i は選択された S 個のノードのインデックスを表す。

$$age_i(t) = 0 \quad (2)$$

- (d) 各格子領域に対するモデル l の出力値 O^l は選択された S 個のノードの類似度を用いて、式 (3) によって算出される。

$$O^l = \begin{cases} D_{c(1)} & (S = 1) \\ \sum_{i=1}^{S-1} D_{c(i)} D_{c(i+1)} & (\text{otherwise}) \end{cases} \quad (3)$$

$c(i)$ は現フレームで選択された上位 S 個のノードのインデックスのうち、 i 番目に類似度が大きいノードのインデックスである。本論文では映像内における出現頻度が低い刺激や特徴が大きく異なる刺激を異常として定義しているため、出力値 O^l がしきい値 O_{thr} より小さい場合、その格子領域を異常であると判定する。

- (e) モデル内で最大の類似度 $D_{c(1)}$ であるノードの馴化係数 $h_{c(1)}(t)$ を式 (4) によって更新する。

$$h_{c(1)}(t+1) = h_{c(1)}(t) - \gamma D_{c(1)} \quad (4)$$

馴化係数の更新式 (4) には入力刺激との類似度を用いており、環境からの入力刺激と類似した重みベクトルを持つノードの馴化係数の値は減少していく。なお、馴化係数の値が 0 を下回った場合は、馴化係数の値は 0 とする。

- (f) 現フレームで選択された上位 S 個以外のノードの $age_i(t)$ を式 (5) によって更新する。

$$age_i(t+1) = age_i(t) + 1 \quad (5)$$

- (g) 式 (6) を満たす場合、新たなノード k をモデルに追加する。

$$\begin{cases} D_{c(1)} < D_{thr} \\ h_{c(1)} < H_{thr} \end{cases} \quad (6)$$

このとき、ノード k の重みベクトルは入力刺激 \mathbf{x} とノード $c(1)$ の重みベクトル $\mathbf{v}_{c(1)}$ との平均値とし、馴化係数の初期値 $h_k(0)$ は 1.0、年齢の初期値 $age_k(0)$ は 0 とする。モデルが式 (6) を満たすことは、ある

Algorithm 1: Anomaly detection algorithm

Initialize: initialize node set $\{\{N_i^k\}_{i=1}^L\}_{k=1}^L$, weight vector v_i are randomly chosen from input stimuli, habituation coefficient $h_i(0) \leftarrow 1.0$, $age_i(0) \leftarrow 0$.

Input: a set of frames $\{t_i\}_{i=1}^T$, input stimuli $\{\{x_i^k\}_{k=1}^L\}_{i=1}^T$.

Output: anomaly labels for each location of each frame $\{\{y_i^k\}_{k=1}^L\}_{i=1}^T$.

```

foreach frame  $t_i$  do
  foreach model  $l$  do
    compute input stimuli  $x_i^l$ ;
    foreach node  $i^l$  do
      compute similarity  $D_i^l$  using (1);
    end
    select the top  $S$  nodes in  $D_i^l$  and update age of them using (2);
    compute normal value  $O^l$  using (3);
    if  $O^l < O_{thr}$  then
       $\{y_{t_i}^l\} \leftarrow anomaly$ ;
    else
       $\{y_{t_i}^l\} \leftarrow normal$ ;
    end
    update habituation coefficient  $h_{c(1)}^l(t_i)$  using (4) and age of nodes using (5);
    if  $D_{c(1)}^l < D_{thr}$  and  $h_{c(1)}^l < H_{thr}$  then
      insert a new node  $k$  with weight vector  $v_k^l$ :
       $v_k^l = \frac{1}{2}(v_{c(1)}^l + x_i^l)$ ;
    end
    foreach node  $i^l$  do
      if  $age_{i^l}^l(t_i) > age_{thr}$  then
        delete node  $i^l$ ;
      end
    end
  end
end
  
```

程度モデルの更新が行われているにもかかわらず環境からの入力刺激を表現するノードがモデル内に存在しないことを示している。このことから、環境の刺激をより正確に表現するために、環境の入力刺激を利用して新たなノードをモデルに追加する操作を行っている。

- (h) $age_i(t)$ の値がしきい値 age_{thr} より大きい場合、ノード i をモデルから削除する。

以上の提案手法の処理についてまとめたものを Algorithm 1 に示す。

3. 固定カメラと旋回カメラによる監視映像からの侵入物体検知実験

3.1 概要

提案する自己組織化モデルの有効性を検証するため、屋

表 1 提案手法に関するパラメータ

Table 1 Parameter settings for the proposed method.

パラメータ	固定	旋回
初期化時のノード数 I		20
ノードの選択数 S		1
馴化係数の更新係数 γ		0.1
ノード追加の馴化係数しきい値 H_{thr}		0.1
最大年齢 age_{thr}		800
ノード追加の類似度しきい値 (モデル適用時) D_{thr}		0.1
ノード追加の類似度しきい値 (モデル構築時) D_{thr}	0.6	0.9
出力値のしきい値 O_{thr}	0.76	0.87

表 2 比較手法に関するパラメータ

Table 2 Parameter settings for GWR and MOG.

比較手法	パラメータ	固定	旋回
GWR	発火係数のしきい値 h_T	0.1	
	発火係数の更新係数 α_b, α_n	1.05	
	重みベクトルの更新係数 ϵ_b	0.05	0.01
	重みベクトルの更新係数 ϵ_n	0.01	0.001
	活性度のしきい値 a_T	0.9	0.8
	発火係数の更新係数 τ_b	0.1	0.4
	発火係数の更新係数 τ_n	0.05	0.4
MOG	分布数	5	
	背景を決定するしきい値	0.6	0.9
	更新に利用するフレーム数	300	200
	分布一致を決定するしきい値	6	30

外監視映像内に現れる歩行者と車両を検出対象とした侵入物体検知問題に提案手法を適用し、先行研究との比較を行う。対象とした屋外監視映像は、固定カメラから撮影された監視映像と水平方向に等速旋回する旋回カメラから撮影された2種類の監視映像である。比較手法は、文献 [13] で提案された GWR ネットワーク (以下, GWR) と, Stauffer らの手法 [21] に対して直近の入力フレームを重視した更新を行う混合ガウス分布を用いた背景モデルの手法 [22] (以下, MOG) を用いた。GWR は [13] の筆者の Web ページ*1をもとに筆者らが作成したものを用い、MOG は OpenCV に実装されている関数*2を使用した。

各実験に共通して用いた提案手法と比較手法のパラメータをそれぞれ表 1, 表 2 に示す。これらのパラメータは次節 3.2 で説明する F 値が最大となるように、事前に行った実験によって決定した。各パラメータの詳細は文献 [13], [22] を参照されたい。提案手法と GWR の入力刺激には、格子領域内の画素値の平均, 最大値, 最小値, レンジ, 中央値, 第一四分位数, 第三四分位数の7種類の統計量を RGB カラー画像, RGB エッジ画像のそれぞれから算出した計 42 次元の特徴量を用いた。レンジは格子領域内の最大画素値と最小画素値の差分値である。本論文では入力刺激が 42

*1 <https://seat.massey.ac.nz/personal/s.r.marsland/gwr.html>
 *2 http://docs.opencv.org/2.4.9/modules/video/doc/motion_analysis_and_object_tracking.html

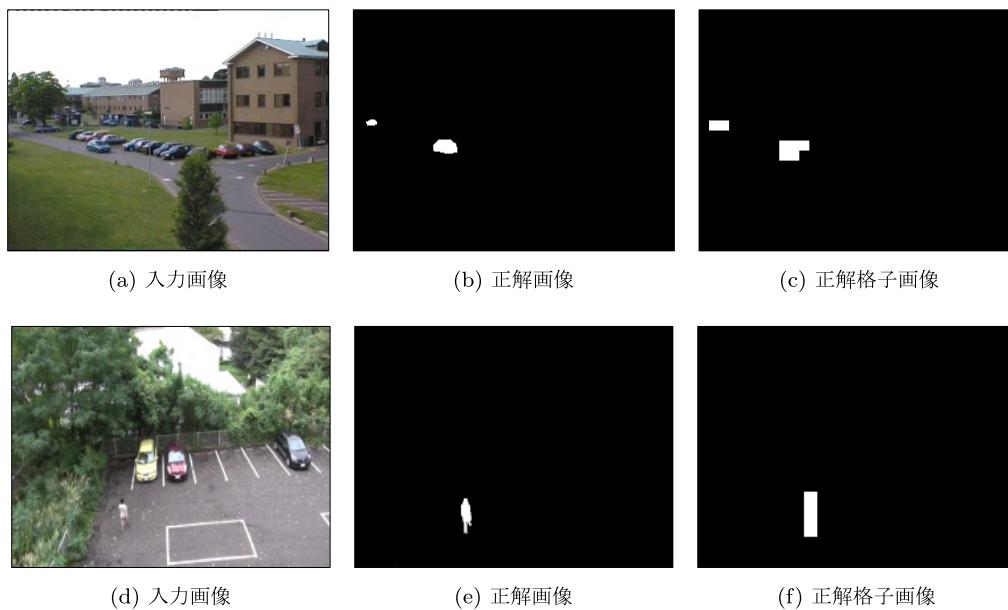


図 4 入力画像と正解画像例. 上段は固定カメラから撮影された映像例, 下段は旋回カメラから撮影された映像例

Fig. 4 Examples of input images, ground truth and block ground truth. First row: images are obtained by a fixed camera; second row: images are obtained by a rotating camera.

次元ベクトルであるため, 提案モデルの各ノードは 42 次元空間の 1 点を表す.

3.2 評価方法

まず, 異常検知性能の定量評価を行うために, 監視映像内に現れる歩行者と車両を対象に正解画像を作成した. 正解画像の例を図 4(b), 図 4(e) に示す. さらに, 提案手法と GWR では画像内の格子領域単位の出力であるため, 作成した正解画像から正解格子画像を作成した. 正解格子画像は, 正解画像を提案手法および GWR と同様のサイズの格子領域に分割し, 格子領域内に占める正解画素の割合が 0.1 以上の格子領域を正解格子領域とすることで作成した. 1 つの格子領域サイズが 10×10 画素とした場合の正解格子画像の例を図 4(c), 図 4(f) に示す. また, 画素単位で出力する MOG と評価方法を揃えるために, MOG による検出画像を同じサイズの格子領域で分割し, 格子領域内に占める検出画素の割合が同じく 0.1 以上の格子領域を MOG による検出格子領域と定義する.

本論文では作成した正解格子画像をもとに, 以下の式で示される再現率 (R), 適合率 (P), F 値 (F) の指標を用いて定量評価を行う. なお, 過剰な過検出結果をノイズとして除外するために, 物体の侵入によって生じた過検出格子領域は検出格子領域から除外した.

$$R = \frac{C}{C_a} \quad (7)$$

$$P = \frac{C}{A} \quad (8)$$

$$F = \frac{2 \cdot R \cdot P}{R + P} \quad (9)$$

C は正しく検出した格子領域数, C_a は正解格子領域数, A は検出した格子領域数である.

3.3 固定カメラからの監視映像による侵入物体検知

3.3.1 実験設定

固定カメラからの監視映像には, PETS2001 データセット中の DATASET2 TESTING CAMERA1 を使用した. 用いた画像は 320×240 画素のカラー画像である. 提案手法と GWR では 10×10 画素ごとくに 1 つのモデルを配置したため, 画像サイズからモデル数は $768 (= \frac{320}{10} \times \frac{240}{10})$ である. この動画像では手前の樹木が風によって揺れており, また背景にちらつきなどのノイズがみられる. 固定カメラ映像の例を図 4(a) に示す.

歩行者と車両が現れない開始から 300 フレームまでを提案手法と比較手法のモデル構築期間とし, 301 フレームから 1,800 フレームまでをモデル適用期間とする. なお, 正解格子画像は 301 フレームから 1,800 フレームまで 10 フレームおきに作成し, 計 150 フレーム分を用意した. また, 提案手法のノード追加の類似度しきい値 D_{thr} は, 表 1 に示すようにモデル構築時と適用時では異なる値を用いた.

3.3.2 実験結果

図 5 に各手法による検出結果例を示す. なお, 図 5 の各画像は原画像を一部拡大した画像を示している. また, MOG は格子領域単位に変換した結果を示している. 図 5 の上図の結果から, 提案手法では手前の揺れている樹木領

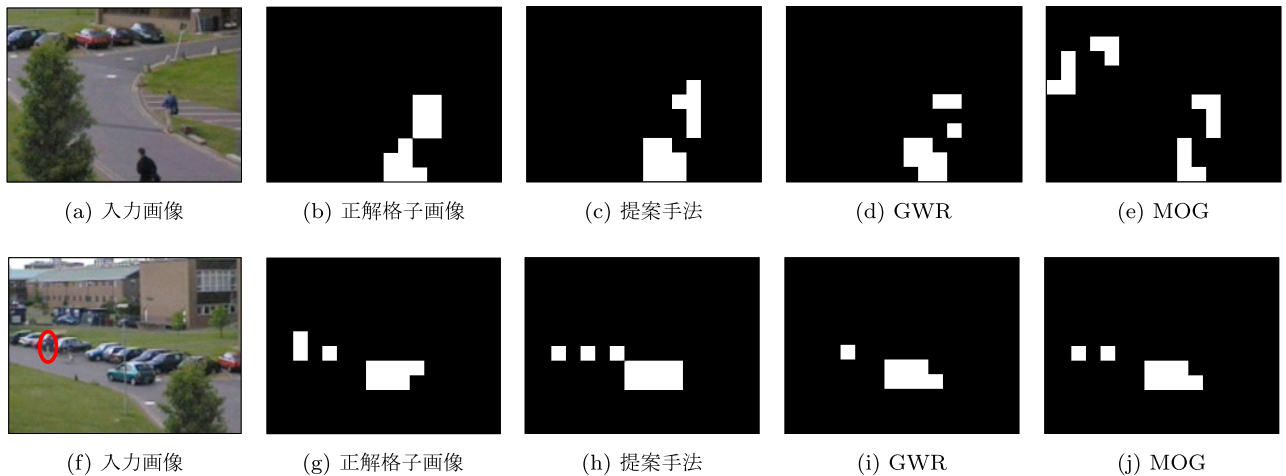


図 5 各手法の検出結果例 (固定カメラ映像)

Fig. 5 Detection results of each method in the fixed camera.

表 3 検出結果に対する定量評価 (固定カメラ映像)

Table 3 Recall, Precision and F-measure on the fixed camera.

	Proposed	GWR	MOG
再現率	0.838	0.656	0.897
適合率	0.868	0.758	0.874
F 値	0.853	0.703	0.885

域への過検出を抑制しつつ、歩行者を検出できていることが分かる。MOG では樹木領域で生じる多様な入力パターンを適切にモデル化できなかったため、樹木領域に対して過検出してしまっている。GWR は提案手法と同様に樹木領域に対して過検出を抑制できているが、図 5 (i) に示すように、歩行者に対して検出漏れをしてしまう場合が多くみられた。

再現率、適合率、F 値による定量評価結果を表 3 に示す。なお、正解格子領域数は正解画像 150 フレーム分に対して 1,732 である。表 3 から提案手法は GWR と比べて再現率、適合率において優れており、MOG と比べると適合率で同等程度の性能を示していることが分かる。しかし、再現率において提案手法は MOG より低い値となっている。これは図 5 (f) 内の赤枠で示すように、背景と類似した色情報を持つ歩行者領域に対して検出漏れをしまっていることが原因だと考えられる。提案手法では映像内の色情報をもとに映像内の正常性を記述するため、背景の色情報と類似している侵入物体に対しては検出が抑制されてしまう。同様の理由で GWR の再現率の値も低い結果となっている。これは今後、入力刺激に時空間特徴を使用し、入力刺激の時間的な変化も考慮することで解決が可能であると考える。

3.4 旋回カメラからの監視映像による侵入物体検知

3.4.1 実験設定

本論文では、水平方向に 45° の範囲で等速旋回する旋回

カメラから撮影された監視映像を対象に侵入物体検知を行う。11~13 秒で監視範囲をカバー可能なように、背景画像が 1 フレームで約 0.4~0.6 画素の速度で動く旋回カメラとなっている。動画のフレームレートは 30 [fps] である。図 4 (d) に旋回カメラ映像の例を示す。対象とした環境では映像内の左右上下から歩行者が映像内に出現する。動画は 240 × 160 画素のカラー画像を使用し、3.3 節の実験と同様に提案手法と GWR では 10 × 10 画素に 1 つのモデルを配置したため、画像サイズからモデル数は 384 である。

歩行者が出現しない開始から 3,000 フレームまでを各手法のモデル構築期間とし、歩行者が出現する 3,001 フレームから 5,000 フレームまでをモデル適用期間とする。正解格子画像は 3,001 フレームから 5,000 フレームまで 10 フレームおきに作成し、計 200 フレーム分を用意した。提案手法のパラメータ D_{thr} は、3.3 節の実験と同様にモデル構築時と適用時ではそれぞれ表 1 に示す値を用いた。

旋回カメラ映像への対処法として、旋回による移動量をあらかじめ算出しておき、その移動量を用いて入力画像列を基準となるフレームの画像に補正し、固定カメラを想定した手法を適用する方法が考えられる。本論文では、固定カメラ映像での異常検知精度が高い MOG をこの補正画像列に適用し、比較手法として用いる。まず、映像内の白線などのエッジ強度が高い特徴点に対してフレーム間で画像の対応付けを行うことで、それら特徴点の 1 フレームにおける移動量を算出する。そして、これら移動量の平均値を旋回による 1 フレームでの移動量とする。次に、事前に定めた基準となるフレームから現在のフレームまでの経過時間と先に求めた旋回による移動量を用いて、現在の画像を移動させることで基準となるフレームの画像に変換を行う。すべての入力画像に対してこの補正処理を行うことで、旋回カメラによる画像列から擬似的な固定カメラ画像列を生成する。そして、生成された擬似的な固定カメラ画像列に対して MOG を適用することで評価を行う。なお、

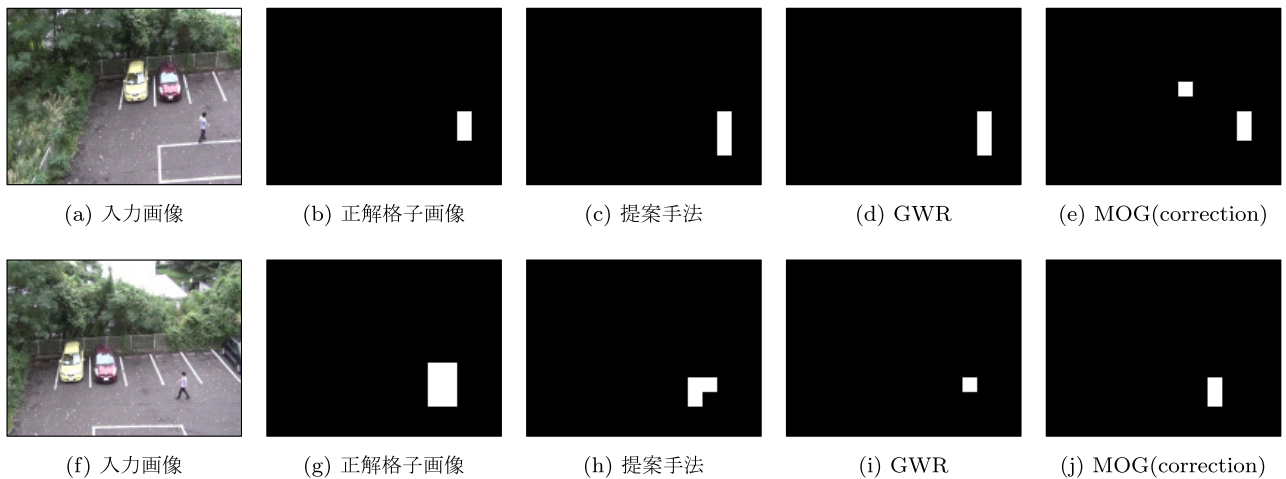


図 6 各手法の検出結果例 (旋回カメラ映像)

Fig. 6 Detection results of each method in the rotating camera.

表 4 検出結果に対する定量評価 (旋回カメラ映像)

Table 4 Recall, Precision and F-measure on the rotating camera.

	Proposed	GWR	MOG	MOG (correction)
再現率	0.753	0.478	0.340	0.725
適合率	0.845	0.831	0.081	0.512
F 値	0.797	0.607	0.131	0.600

旋回による動き補正を行った際に、補正後の画像内に画像情報が欠落してしまう領域が存在するが、情報が欠落している領域に対しては MOG の更新および異常検知判定は行わないようにした。補正処理を行う MOG のパラメータは分布数を 5、背景を決定するしきい値を 0.9、更新に利用するフレーム数を 200、分布一致を決定するしきい値を 27.5 とした。

3.4.2 実験結果

再現率、適合率、F 値による定量評価結果を表 4 に示す。補正処理を行った MOG の結果は MOG (correction) として示している。また、正解格子領域数は正解画像 200 フレーム分に対して 247 である。表 4 から提案手法は再現率、適合率、F 値において他の 3 手法と比べて高い値を示していることが分かる。F 値は比較手法と比べて約 19% 以上優れている。補正処理を行う前の MOG では、変動する背景領域の多様な状態を適切にモデル化できなかったため、再現率、適合率ともに低い数値となっているが、補正処理を行うことで再現率、適合率ともに大幅な改善がみられた。

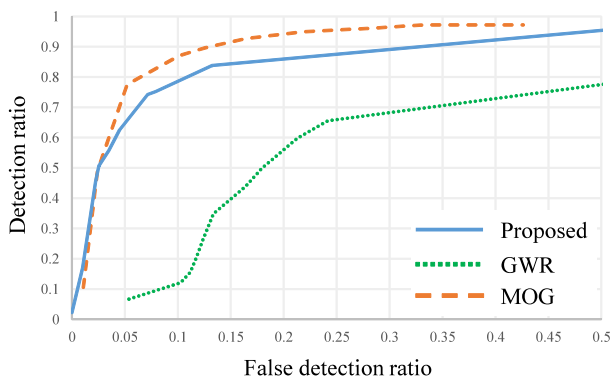
図 6 に各手法における検出結果例を示す。図 6 の各画像は原画像を一部拡大した画像を示している。なお、図 6 には補正処理を行った MOG の結果を示している。図 6 の結果から、提案手法では変動する背景領域への過検出を抑制しつつ、侵入物体である歩行者を検出できていることが分かる。図 6 (h) の歩行者領域の一部に対して検出漏れが

みられるが、前後の評価対象外のフレームをみると、この検出漏れの領域に対して正しく検出できていることが確認されている。GWR は提案手法と同様に変動する背景領域への過検出を抑制できているが、歩行者領域に対して検出漏れをしてしまう場合が多くみられた。補正処理を行った MOG では歩行者領域に対して良好に検出できているが、変動する背景領域に対して過検出する場合がみられた。これは旋回装置の旋回精度や振動、補正時の補正誤差が原因で適切に背景領域のモデル化が行えなかったためであると考えられる。

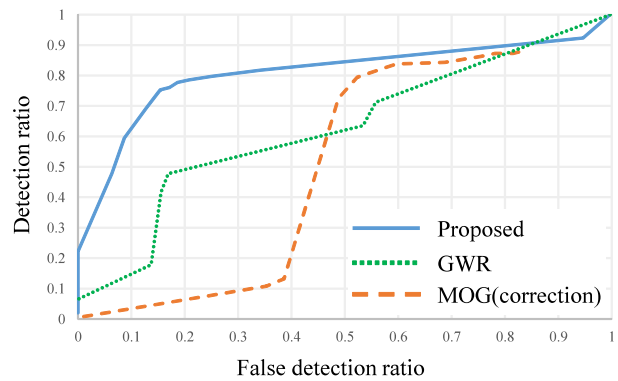
3.5 考察

各手法で異常検知のしきい値を変えたときの異常検知率と誤検知率を図 7 の ROC 曲線に示す。図 7 (a) は固定カメラ映像に対する結果、図 7 (b) は旋回カメラ映像に対する結果を示している。なお、図 7 (b) には補正処理を行った MOG の結果を示している。提案手法では出力値のしきい値、GWR では出力値のしきい値と活性度のしきい値、MOG では背景を決定するしきい値と分布一致を決定するしきい値を変えて異常検知率と誤検知率を求めた。図 7 は横軸に誤検知率、縦軸に異常検知率をとり、グラフが左上にあるほど性能が優れていることを表している。

図 7 (a) の固定カメラ映像の結果から、提案手法と MOG が GWR と比べて優れた結果を示しており、MOG が提案手法より良好な結果となっている。しかし、提案手法では図 5 の上図に示したように風によって揺れている樹木領域への過検出を抑制しつつ、侵入物体を検出できている。このことを定量的に評価するため、図 8 内の赤枠で示した樹木領域における過検出数を 301 フレームから 1,800 フレームの 1,500 フレーム分について調べた。その結果、過検出数が提案手法では 179、GWR では 301、MOG では 402 となった。このことから、提案手法は他の 2 手法と比べて変動が生じる領域への過検出を抑制できていることが分かる。



(a) 固定カメラ映像



(b) 旋回カメラ映像

図 7 各手法の ROC 曲線

Fig. 7 ROC curves of each method.

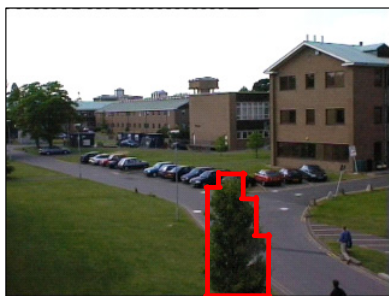


図 8 樹木領域

Fig. 8 The tree area.

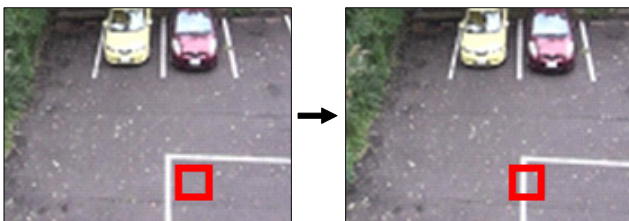


図 9 ノード追加シーンの例

Fig. 9 An example case of node addition.

次に、図 7(b) の旋回カメラ映像の結果から、提案手法は比較手法と比べて優れた結果を示していることが分かる。固定カメラ映像における樹木領域への過検出の抑制結果と旋回カメラ映像における評価結果から、提案手法は比較手法と比べてより環境の変化に頑健に異常検知を行うことができることが確認された。

このように提案手法が環境の変化に対して頑健に異常検知を行えるのは、環境からの入力刺激との類似度を考慮して環境の正常性を表現している点にある。まず、提案手法では入力刺激と類似しているノードがモデル内に存在しない場合、入力刺激と類似したノードをモデル内に追加することで環境の変化に適応する。たとえば、図 9 内の赤枠で示した位置のモデルでは、旋回カメラの旋回によって移動した図 9 右のフレームで白線を観測したときにノード追加が行われる。これは現在までのフレームで白線を表現

表 5 ノード選択数 S に対する提案手法の性能の変化

Table 5 Performance of the proposed method with changing S .

S	1	2	3	4	5
再現率	0.753	0.680	0.628	0.607	0.425
適合率	0.845	0.824	0.752	0.630	0.587
F 値	0.797	0.745	0.684	0.619	0.493

するノードがモデル内に存在せず、式 (6) を満たしたためである。また、ノードの削除も行われており、環境内で出現頻度が低い入力刺激を表現するノードはモデルから削除される。固定カメラ映像における樹木領域や、旋回カメラ映像における背景領域のように様々な入力刺激が観測される環境では、このノード追加や削除の操作によって環境の状態を表現するモデルを構築していく。次に、提案手法ではこの構築されたモデル内のノードと入力刺激との類似度を用いて出力値を算出するため、環境内で特徴の異なる入力刺激に対して出力値の正常度を抑制することができる。GWR では入力刺激との類似度ではなく、出現頻度を考慮して異常判定を行うため、モデルの入力刺激に対する表現力が弱いといえる。また MOG については、変動領域で観測される様々な入力パターンを背景として適切にモデル化できなかったことが過検出結果の原因だと考えられる。

3.6 パラメータによる影響

提案手法の各パラメータが出力に及ぼす影響について考察する。ここでは、特に出力に影響を及ぼすと考えられる、ノード選択数 S 、モデル構築時のノード追加の類似度しきい値 D_{thr} 、モデル適用時のノード追加の類似度しきい値 D_{thr} の 3 つのパラメータについて検討を行う。出力値のしきい値 O_{thr} については、図 7 に示したとおりである。これらの 3 つのパラメータを 1 つずつ独立に変化させながら、3.4 節と同様の評価実験を行った結果を表 5、表 6、表 7 に示す。なお、変更していないパラメータについては

表 6 モデル構築時の類似度しきい値 D_{thr} に対する提案手法の性能の変化Table 6 Performance of the proposed method with changing D_{thr} during training.

D_{thr}	0.825	0.850	0.875	0.900	0.925	0.950
再現率	0.773	0.765	0.761	0.753	0.709	0.700
適合率	0.136	0.376	0.793	0.845	0.862	0.869
F 値	0.231	0.505	0.777	0.797	0.778	0.776

表 7 モデル適用時の類似度しきい値 D_{thr} に対する提案手法の性能の変化Table 7 Performance of the proposed method with changing D_{thr} during detection.

D_{thr}	0.1	0.3	0.5	0.7	0.8
再現率	0.753	0.749	0.737	0.518	0.340
適合率	0.845	0.845	0.843	0.810	0.800
F 値	0.797	0.794	0.786	0.632	0.477

表 1 と同様の値である。

まず選択ノード数 S について、提案手法では選択したノード数分の類似度を使用して出力値を算出することができる。表 5 の結果から選択するノード数 S の数は少ない方がより優れた性能を示していることが分かる。巡回カメラ映像では背景が大きく変動するため、それぞれ異なる重みベクトルの値を持ったノードが多く存在する。そのため選択数 S を増加させると出力値が抑制されてしまい、結果として再現率、適合率が下がる結果となった。

次に、モデル構築時のノード追加の類似度しきい値 D_{thr} について、表 6 の結果から D_{thr} の値が高くなるとより優れた性能を示すことが分かる。これはモデル構築時に D_{thr} の値を低くしてノード追加の制限を厳しくすると、変動する背景領域を表現するノードが十分にモデルに追加されず、モデル適用時に背景領域に対して過検出を多く引き起こしてしまうためである。そのため結果として、 D_{thr} の値が低いと適合率が大幅に減少してしまったと考えられる。反対にしきい値 D_{thr} を高くすると、環境内の背景領域を表現するノードが追加されやすくなるため、過検出が抑制され適合率が上昇する結果となった。

最後に、モデル適用時のノード追加の類似度しきい値 D_{thr} に対する結果を表 7 に示す。類似度しきい値は値が高くなると前述したように、環境内の刺激を表現するノードがモデルに追加されやすくなる。そのため、適用時の類似度しきい値を高く設定すると、背景領域に加えて歩行者領域に対する検出も抑制されるようになる。表 7 から、しきい値が高くなると歩行者領域に対する再現率が低下していることが分かる。また、しきい値 D_{thr} の値が 0.1, 0.3, 0.5, 0.7, 0.8 のときの背景領域に対する過検出数はそれぞれ 34, 34, 34, 30, 21 であり、しきい値が高くなると過検出数が抑制されることが確認された。したがって、適用

時の類似度しきい値 D_{thr} を高く設定すると、再現率は低下するが、過検出数は抑制される結果となった。反対にモデル適用時の D_{thr} を低く設定すると、歩行者領域を表現するノードがモデルに追加されにくくなり、結果として再現率が上昇する結果を示した。

4. まとめ

本論文では、環境から入力される刺激に基づいて環境の正常性を表現し、異常検知を行う自己組織化モデルを提案した。提案手法では、環境からの入力刺激との類似度に応じてノードの追加、削除、状態の更新、また出力値を算出することで環境の正常性を表現する。提案手法を固定カメラと巡回カメラから撮影された 2 種類の監視映像からの侵入物体検知問題に適用し、先行研究と比較を行った結果、風による樹木の揺れや、巡回による変動がある領域への過検出を抑制しつつ、異常である侵入物体を検知できることが確認された。

今後の課題として、長時間の監視映像への適用実験と、その実験を通じての精度向上およびパラメータ設定に関する明確な基準を定めることがあげられる。また、本論文では画像内の領域ごとに異なる正常性を表現しているが、近傍の領域や画像全体で正常性を共有することで、より効率的な正常性の表現を行う必要があると考えている。さらに、提案手法の拡張として人物の異常動作検知や、混雑シーンにおける適用などを検討したい。

謝辞 本研究は、株式会社半導体理工学研究センター (STARC) との共同研究によるものである。

参考文献

- [1] Adam, A., Rivlin, E., Shimshoni, I. and Reinitz, D.: Robust real-time unusual event detection using multiple fixed-location monitors, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30, No.3, pp.555–560 (2008).
- [2] Basharat, A., Gritai, A. and Shah, M.: Learning object motion patterns for anomaly detection and improved object detection, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp.1–8 (2008).
- [3] Zhong, H., Shi, J. and Visontai, M.: Detecting unusual activity in video, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.819–826 (2004).
- [4] Kobayashi, T. and Otsu, N.: Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation, *Proc. 17th International Conference on Pattern Recognition*, Vol.4, pp.741–744 (2004).
- [5] 南里卓也, 大津展之: 複数人動画からの異常動作検出, 情報学論, コンピュータビジョンとイメージメディア, Vol.46, No.15, pp.43–50 (2005).
- [6] Xiang, T. and Gong, S.: Video Behavior Profiling for Anomaly Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30, No.5, pp.893–908 (2008).
- [7] Andrade, E.L., Blunsden, S. and Fisher, R.B.: Modelling

- crowd scenes for event detection, *Proc. International Conference on Pattern Recognition*, Vol.1, pp.175–178 (2006).
- [8] Mehran, R., Oyama, A. and Shah, M.: Abnormal crowd behavior detection using social force model, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp.935–942 (2009).
- [9] Saxena, S., Brémond, F., Thonnat, M. and Ma, R.: Crowd behavior recognition for video surveillance, *Advanced Concepts for Intelligent Vision Systems*, pp.970–981 (2008).
- [10] Saligrama, V. and Chen, Z.: Video anomaly detection based on local statistical aggregates, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.2112–2119 (2012).
- [11] Chan, A.B. and Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30, No.5, pp.909–926 (2008).
- [12] Mahadevan, V., Li, W., Bhalodia, V. and Vasconcelos, N.: Anomaly detection in crowded scenes, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp.1975–1981 (2010).
- [13] Marsland, S., Shapiro, J. and Nehmzow, U.: A self-organising network that grows when required, *Neural Networks*, Vol.15, No.8-9, pp.1041–1058 (2002).
- [14] Marsland, S., Nehmzow, U. and Shapiro, J.: Detecting novel features of an environment using habituation, *Proc. Simulation of Adaptive Behavior*, pp.189–198 (2000).
- [15] Marsland, S., Nehmzow, U. and Shapiro, J.: Environment-specific novelty detection, *From Animals to Animats, the 7th International Conference on Simulation of Adaptive Behaviour*, pp.36–45 (2002).
- [16] Stanley, J.C.: Computer simulation of a model of habituation, *Nature*, Vol.261, pp.146–147 (1976).
- [17] Neto, H.V. and Nehmzow, U.: Visual novelty detection for inspection tasks using mobile robots, *Proc. 8th Brazilian Symposium on Neural Networks* (2004).
- [18] Nehmzow, U. and Neto, H.V.: Novelty-based visual inspection using mobile robots, *Proc. Towards Autonomous Robotic Systems* (2004).
- [19] Nehmzow, U. and Neto, H.V.: Visual attention and novelty detection: Experiments with automatic scale selection, *Proc. Towards Autonomous Robotic Systems*, pp.139–146 (2006).
- [20] 武田真人, 矢田紀子, 長尾智晴: 映像監視のための環境に適応的な異常検知ネットワーク, *電子情報通信学会論文誌*, Vol.J94-D, No.10, pp.1631–1639 (2011).
- [21] Stauffer, C. and Grimson, W.E.L.: Adaptive background mixture models for real-time tracking, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.246–252 (1999).
- [22] KadewTraKuPong, P. and Bowden, R.: An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection, *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems* (2001).



菅沼 雅徳 (学生会員)

2013年横浜国立大学工学部電子情報工学科を飛び級のため中退。現在、同大学大学院環境情報学府情報メディア環境学専攻博士課程後期在学中。画像処理, パターン認識を含む知能情報学の研究に従事。人工知能学会, 日本医

用画像工学会会員。



長尾 智晴 (正会員)

1985年東京工業大学大学院総合理工学研究科博士課程後期中退。同年同大学助手。同大学助教授を経て2001年横浜国立大学大学院環境情報研究院教授。工学博士。画像処理, 進化計算法等の知能情報学の研究に従事。電子情

報通信学会, 人工知能学会, 電気学会, 進化計算学会, IEEE等会員。