

環境の通常状態を記憶するネットワークを用いた監視映像からの異常検知

Anomaly Detection for Video Surveillance using a Network Memorizes Normal States

菅沼雅徳*1 Masanori Suganuma
長尾智晴*1 Tomoharu Nagao

*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

We humans are quickly able to detect anomaly in a visual scene without prior knowledge. We consider that humans are adaptively able to memorize normality of a given environment, which enables humans to detect anomaly. This ability seems to be useful for video processing systems like anomaly detection from surveillance cameras. This report proposes a network model that memorizes normality of a given environment. This network model is constructed by repeatedly observing image sequences. The proposed model is applied to intrusion detection problems including cars and pedestrians, and shows how proposed model deals with these tasks.

1. はじめに

監視カメラ映像からの不審者、不審物検出や外観検査におけるキズ製品の検出などにみられるように異常検出システムは様々な分野で必要とされており、近年では省コスト化や異常検出支援のために計算機による自動化が求められている。しかし、対象となる環境によって検出すべき異常の定義が異なることやあらゆる異常を事前に定義することは困難であることが多いため、計算機による異常検出システムの構築は容易ではない。

一方、我々人間は異常に関する事前知識が無くても、普段見かけないものに注目することで結果的に不審者の検出やキズ製品の検出が可能である。人間がこのように異常を検知することができるのは、与えられた環境を繰り返し観測するだけで、環境の通常状態というものを自然に定義し、異常を非通常状態と定義し自然に注意を向けることができるためであると考えている。人間が異常を検知する能力を計算機によって自動化することができれば、異常検出の省コスト化や高精度化など多くの有用性が期待できる。そこで本研究ではこの考えに基づき、入力画像を繰り返し観測することで環境の通常状態を記憶し、異常に反応を示すネットワークモデルを提案する。実験では、歩行者や車両を検出すべき異常対象とする侵入物体検出問題に提案モデルを適用し、提案モデルの有効性を検証する。

本研究と同じく、環境の通常状態を利用し異常検出を行う先行研究として、Grow When Required (GWR) ネットワーク [Marsland 02] が挙げられる。GWR ネットワークでは入力頻度が多い入力ベクトルに対して次第にネットワークが反応を抑制させることで環境の通常状態を表現する。実験では学習期間にみられた入力ベクトルとは異なるベクトルに対して強い反応を示すといった結果が得られている。また、武田らは反復される入力に基づいて環境の通常状態を表現するネットワークモデルを提案し、屋外監視映像からの侵入物体検出を行っている [武田 11]。この手法では、入力画像からの入力パターンの種類に対する反応の抑制と入力パターンが生じる位置に対する反応の抑制によって環境の通常状態を表現している。

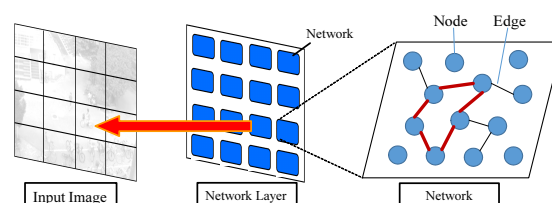


図 1: 提案ネットワークモデルの構造

2. 提案ネットワークモデル

2.1 概要

提案するネットワークモデルの構造を図 1 に示す。提案モデルでは動画中の 1 フレームを入力画像として扱い、同一構造のネットワークが入力画像中に格子状に整列している。例えば、入力画像サイズを $M \times N$ 画素、1つのネットワークが配置される格子領域サイズを $m \times n$ 画素とした場合、入力画像中には $\frac{M}{m} \times \frac{N}{n}$ 個のネットワークが配置される。各ネットワークには各格子領域から算出した特徴量が入力され、ネットワークの内部構造に応じた出力値が算出される。そして、各ネットワークの出力値を環境の通常状態らしさと定義し、その出力値の大小によって異常検出を行う。

2.2 提案モデルの処理の流れ

ネットワークは複数のノードとエッジによって構成されている。ノード i には入力特徴量 f と同次元の重みベクトル v_i と馴化係数 $h(t)$ 、ノード i とノード j 間のエッジには結合荷重 $w_{ij}(t)$ が存在する。各ノードはノード自身もつ重みベクトルと環境からの入力特徴量を比較することで、そのノードが環境の入力をどの程度表現しているかを示す。エッジは 2つのノード間に生成され、環境からの入力に応じて結合荷重が更新される。提案モデルでは、環境からの入力に応じてノードの追加やエッジの生成、結合荷重の更新を行うことで環境の通常状態を記憶するネットワーク構造を動的に構築していく。

次に提案モデルによる処理の詳細について述べる。

1. 全てのネットワークを次の手順で初期化する。
 - (a). ネットワークに I 個のノードを生成する。

連絡先: 菅沼雅徳, 横浜国立大学大学院環境情報学府, 240-8501 横浜市保土ヶ谷区常盤台 79-7, email: suganuma-masanori-hf@ynu.jp

- (b). 各ノードの重みベクトルを入力特徴量からランダムに生成する.
- (c). 各ノードの馴化係数 $h(0)$ を 1.0 にする. なお, 馴化係数は $[0.0, 1.0]$ の値をとるようにしている.
2. 全てのネットワークについて, 毎フレーム以下の処理を T 回繰り返す.

(a). 入力
ネットワークが配置された格子領域から特徴量を算出し, ネットワークに入力する. 入力特徴量は格子領域の画素値から算出した平均, 中央値などの基本統計量を用いる. 全ての特徴量は $[0.0, 1.0]$ に正規化される.

(b). ノードの選択
ネットワーク内のすべてのノードから入力特徴量との類似度が大きい上位 S 個のノードを選択する. ノード i における類似度 D_i の算出には次式 (1) を用いる.

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

ここで, D_i はノード i の入力特徴量 \mathbf{f} との類似度, \mathbf{v}_i はノード i がもつ重みベクトルである. 類似度 D_i が大きいノードほど入力特徴量と類似した重みベクトルをもっていることになる.

(c). エッジの生成と結合荷重の更新
選択された上位 S 個のノードについて, 類似度 D が大きい順に連続されるようにノード間にエッジを生成する. すでにエッジが存在している場合は次式 (2) によってノード i, j 間の結合荷重を更新する. α は定数である.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \quad (2)$$

(d). ネットワークの出力
ネットワークの出力値 O は選択された上位 S 個のノードとエッジの結合荷重を用いて, 次式 (3) によって算出される.

$$O = \sum_{i=1}^S D_{c(i)} D_{c(i+1)} w_{c(i), c(i+1)} \quad (3)$$

$c(i)$ 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を返す関数である. 環境で頻出する特徴量に近い重みベクトルをもつノード間の結合荷重が大きくなるように結合荷重の更新を行っているため, 環境で頻りに観測される通常状態に対してネットワークの出力値は大きくなるように構築されていく. そのため, 本論文では出力値 O があるしきい値 O_{th} より小さい場合, その領域を異常であると判定する.

(e). エッジの更新
現フレームで選択された上位 S 個のノード間以外のエッジの結合荷重を次式 (4) によって更新する. β は定数である.

$$w_{ij}(t+1) = w_{ij}(t) - \beta \quad (4)$$

(f). 馴化係数の更新
ネットワーク内で最大の類似度 D_{max} であるノード i_{max} の馴化係数 $h(t)$ を次式 (5) によって更新する. γ は定数である.

$$h(t+1) = h(t) - \gamma D_{max} \quad (5)$$

ここで, 馴化係数 $h(t)$ はそのノードが環境に対してどの程度馴化しているかを表す指標であり, 次の処理ステップ (g) において新たなノードをネットワークに追加するか否かの指標となる. 馴化係数の更新式 (5) には入力特徴量との類似度を用いており, 環境からの入力特徴量と類似した重みベクトルをもつノードの馴化係数は 0 に近づいていく.

(g). ノードの追加
ネットワーク内での最大の類似度 D_{max} がしきい値 D_{th} より小さい, かつノード i_{max} のもつ馴化係数 $h(t)$ がしきい値 H_{th} より小さい場合, 新たなノード k をネットワークに追加する. このとき, ノード k の重みベクトルは入力特徴量 \mathbf{f} とノード i_{max} の重みベクトル \mathbf{v}_{max} との平均値とし, 馴化係数の初期値 $h(0)$ は 1.0 とする. ネットワークがこの条件を満たすことは, ある程度ネットワークの更新が行われているにも関わらず環境からの入力を表現するノードがネットワークに存在しないことを示している. そこで, 環境の状態をより正確に表現するために, 環境の入力特徴量を利用して新たなノードをネットワークに追加する操作を行う.

3. 侵入物体検出実験

3.1 実験設定

本稿では屋外監視映像を対象に, 映像中に現れる歩行者と車両を検出対象とする侵入物体検出問題を扱う. 対象とする屋外監視映像として, PETS2001 を用いた. 対象とした環境では手前の樹木が風によって揺れていたり, 背景にちらつきなどのノイズがみられる. 画像サイズは 320×240 画素であり, 10×10 画素の格子領域毎に 1 つのネットワークを配置した.

提案モデルと比較手法による性能検証を行うために人手によって正解画像を作成した. 正解画像の例を図 2(b) に示す. さらに, 提案モデルは矩形領域ごとの出力となるため, 正解画像から矩形領域単位で評価する正解矩形画像を用意した. 正解画像を 10×10 画素の矩形領域に分割し, 矩形領域内に占める正解画素の割合が 0.1 以上の矩形領域を正解矩形領域とした.

提案モデルのパラメータは, $I = 20$, $S = 3$, $\alpha = 0.01$, $\beta = 0.001$, $\gamma = 0.05$, $O_{th} = 1.5$, $D_{th} = 0.7$, $H_{th} = 0.1$ を用いた. また, 提案モデルの入力特徴量には矩形領域内の画素値から平均, 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 画像, RGB エッジ画像それぞれから算出した計 42 次元を用いた.

提案モデルに対する比較手法として, 武田らの刺激の選択性を用いた領域検出ネットワーク [武田 11] (以下, EAN), KaewTraKulPong らの混合ガウス分布を用いた適応的背景モデル [KaewTraKulPong 02] (以下, MOG) を用いた. MOG は OpenCV のライブラリ*1を使用した. また, 画素毎に出力する MOG の評価方法を揃えるため, MOG の出力画像の矩形領域内に占める検出画素の割合が 0.1 以上の矩形領域を検出

*1 <http://opencv.willowgarage.com/wiki>

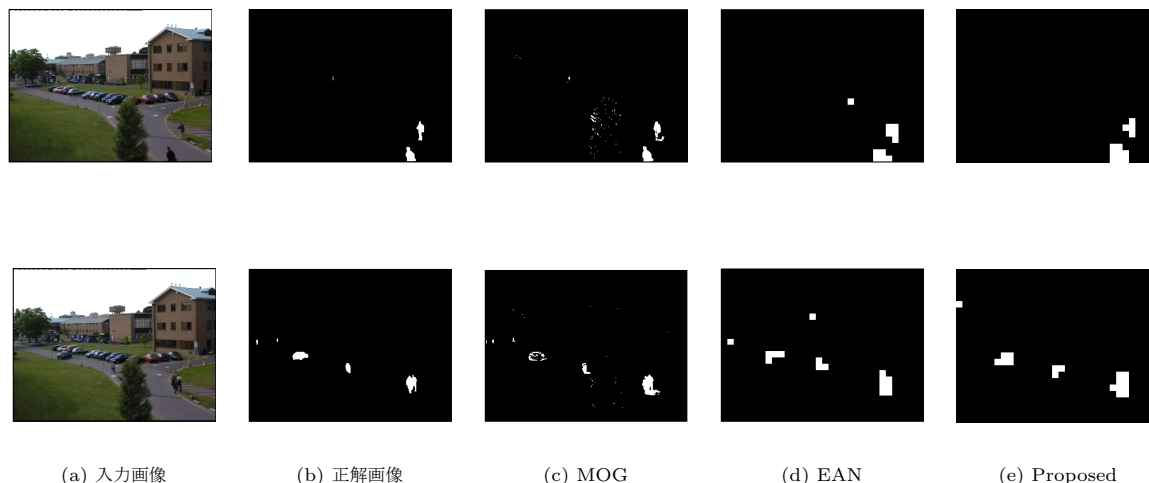
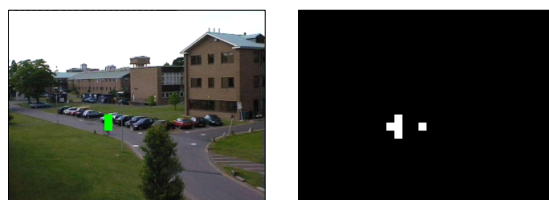


図 2: 検出対象に対する各手法の出力結果例 (EAN の結果画像は文献 [武田 11] から引用)

表 1: 検出結果に対する定量評価

	Proposed	EAN	MOG
再現率	0.800	0.801	0.902
適合率	0.904	0.821	0.825
F 値	0.849	0.811	0.862



(a) 提案モデルの出力 (b) 正解矩形画像

図 3: 提案モデルによる検出漏れの例

矩形領域とする。EAN は本稿と実験設定が同じであるため、EAN による実験は行わずに実験結果は文献 [武田 11] から引用した。

3.2 実験結果

提案モデルと比較手法による検出結果例を図 2 に示す。比較手法である EAN と MOG が手前の揺れている樹木領域に対して過検出しているのに対して、提案モデルは過検出を抑制しつつ歩行者や車両に反応を示せていることがわかる。

次に、表 1 に再現率、適合率、F 値による定量評価を行った結果を示す。再現率、適合率はそれぞれ式 (6)、式 (7) によって算出した。

$$\text{再現率} = \frac{\text{正しい検出矩形領域数}}{\text{正解矩形領域数}} \quad (6)$$

$$\text{適合率} = \frac{\text{正しい検出矩形領域数}}{\text{検出矩形領域数}} \quad (7)$$

表 1 から提案モデルは適合率では比較手法より高い値を示していることがわかる。F 値で比較すると、EAN より高精度であり、MOG と同程度の性能を示していることがわかる。これは提案モデルでは複数のノード間の共起関係を考慮することやネットワークにノードを適応的に追加することで、手前の樹木領域や背景のノイズへの過検出を抑制することができたためであると考えられる。一方、提案モデルと比較手法を再現率で比較すると提案モデルの方が低い値となっている。これは図 3 に見られるように、背景の色と類似した歩行者に対して検出漏れが生じてしまったことが原因であると考えられる。提案モデルでは環境の通常状態を記憶するための入力情報として色情報だけをを用いているため、背景と類似した色情報をもつ歩行者には反応が抑制されてしまう結果となった。

4. 結論

本研究では、環境を繰り返し観測することで環境の通常状態を記憶し、通常状態とは異なる異常に対して反応を示すネットワークモデルの提案を行った。屋外監視映像を対象に、環境中に現れる歩行者や車両を異常対象とした侵入物体検出問題に提案モデルを適用した結果、背景の変動領域への過検出を抑制しつつ、対象を検出することができた。今後は多くの環境における実験を行い、提案モデルの精度向上のため内部パラメータの最適化、入力情報の決定などを行っていく。

参考文献

- [KaewTraKulPong 02] KaewTraKulPong, P. and Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection, in Video-Based Surveillance Systems, pp.135–144 (2002)
- [Marsland 02] Marsland, S., Shapiro, J. and Nehmzow, U.: A self-organising network that grows when required, in Neural Networks, Vol. 15(8), pp.1041–1058 (2002)
- [武田 11] 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)