Vol.2019-CG-175 No.1
2019/9/28

動画中の異常検出技術の実装と分析

岡部 誠 [1]

概要：動画中の異常を自動検出する技術の研究を行っている．入力は定点カメラによって常時撮影されている映像である．この問題への一般的なアプローチでは，異常状態を「正常状態から外れる状態」と定義し，正常状態をコンピュータに学習させる．これは一般に異常状態の映像を入手することが困難なためである．我々は機械学習技術を用いて正常状態を学習することで異常検出器を構築する．今回の発表では，Xu らの denoising autoencoder に基づく手法と Lu らのスパースモデリングに基づく手法を実装し，実験結果を報告する．

Implementation and Analysis of Video-based Anomaly Event Detectors

MAKOTO OKABE[†1]

1. はじめに

街の治安や工場等での機器の正常な動作を保証するために多くのカメラが導入されている．同時に動画を見張って異常を検出する作業が必要とされている．しかし，人間がその作業を担うとすると，体力的な負担，もしくは，人件費の高騰といった問題が生じる．そのため，我々は動画中の異常を自動検出する技術を開発したい．

この問題への一般的なアプローチでは，異常状態を「正常状態から外れる状態」と定義し，正常状態をコンピュータに学習させる．これは一般に異常状態の映像を入手することが困難なためである．正常状態の学習には機械学習技術が用いられる．正常状態として学習した以外の状態が起これば，それらを全て異常として検出させることが目標となる．ただし，今回の研究においてはカメラの状態は変化しないことを想定している．

今回の発表では，Xu らの denoising autoencoder に基づく手法[7]と Lu らのスパースモデリングに基づく手法[5]を実装し，実験結果を報告する．

2. 映像からの異常検出技術に関する調査

映像解析による異常検出は，コンピュータビジョンやパターン認識の分野で活発に研究が行われてきた問題である．特に深層学習に基づいた技術については，Kiran らが 2018 年に発表したサーベイ論文[1]にまとまっている．深層学習が世に出る以前の研究では，研究者が手作業で特徴量をデザインし，それらに基づいて動画を解析することで，正常／異常の判断を行っていた．Li らは動画中の局所的な 3 次元パッチに注目し，それを dynamic texture に変換した[2]．動画から抽出される多くの dynamic texture をクラスタリングすることで異常検出器を構築した．Kim らは動画中の速度場から得られる特徴量を用い，それらのパターンをマルコフランダム場（MRF）を用いて学習することで異常検出を行った[3]．また，Cong らは速度場のヒストグラムを特徴量に用い，それらにスパースモデリングを適用することで，異常検出器を構築した[4]．Lu らは勾配情報の 3 次元パッチを主成分分析で次元削減したものを特徴量とし，それらに対してスパースモデリングを適用することで異常検出器を構築した[5]．

深層学習を用いた手法では，手作業ではデザイン不可能な複雑な特徴量が自動生成できることから，異常検出の精度が向上している．Xu らは深層学習器の中から stacked denoising autoencoder (SDAE)を選択し，学習データセット毎に特徴量を学習するアプローチをとっている．抽出された特徴量に 1 クラスサポートベクトルマシン(SVM)[6]を適用することで異常検出器を構築している[7]．Ravanbakhsh らは学習済み CNN を特徴抽出器として利用し，高精度な異常検出器を構築している[8]．更に，Ravanbakhsh らは敵対的生成ネットワーク (GAN)[9]を用いて異常検出器を構築するとともに，学習済み CNN を用いず，Xu らと同様に学習データセット毎に特徴抽出器を学習した[10]．

3. SDAE に基づくアプローチ

Xu らの SDAE に基づく手法[7]について実装を行った．この手法の入力はグレースケールの動画である．この動画に異常があるかどうかは，以下の処理によって検出される．

① 動画の各フレームから，15×15 のパッチを切り出す．例えば，入力動画の解像度が 150×150 であった場合，10×10＝100 個のパッチを切り出すことができる．

② ①で切り出した各パッチを，15×15＝225 次元ベクトルとして Encoder に入力する．Encoder は特徴量を抽出する．

③ ②で抽出した特徴量を 1 クラス SVM に入力する．1 クラス SVM は正常なら 0，異常なら 1 を返す

1 静岡大学 大学院総合科学技術研究科
工学専攻 数理システム工学コース
Department of Mathematical and Systems Engineering,
Graduate School of Engineering, Shizuoka University

ⓒ2019 Information Processing Society of Japan

④ ③で１が返った場合，入力動画のパッチに対応する部分を赤く塗って表示する

以上の処理を入力動画の全てのフレームに対して繰り返すことで，動画全体に対する異常検出処理が完了する．

この手法では動画の見た目と速度場の情報に基づいて異常検出を行っている．UCSD データセット[2]に対し，見た目の情報のみを用いて実験を行った結果を図１に示す．はっきりと見た目が違うカートやトラックなどは，速度場の情報がなくても異常として検出できている．ただし，見た目だけでは判断できないような異常，例えば，スケートボードに乗る人や，走る人，自転車に乗る人のような，見た目的には異常といい難いケースでは検出が難しい．

図１）見た目の情報のみを用いた異常検出

速度場を併用した場合の結果を図２に示す．今回は，見た目に基づく検出器と速度場に基づく検出器の結果を，１：１の重みで混ぜ合わせて検出を行っている．図１の異常検出に加え，速く走る物体（自転車，スケートボード，走る人など）が検出されるようになった．

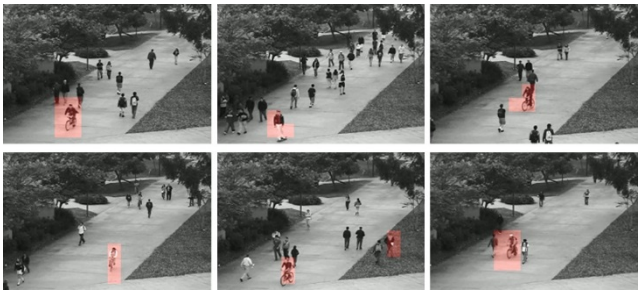

図２）速度場を併用した場合の異常検出

4. スパースモデリングに基づくアプローチ

Lu らのスパースモデリングに基づく手法[5]では，動画を時空間ボリュームとみなす．ボリュームの 3 次元的な勾配を計算し，そこから 10×10×5 のパッチを切り出す．例えば，入力動画のフレーム数が 100 であった場合，各 x-y 座標につき，96 個のパッチを切り出すことができる．切り出した各パッチは 1500 次元の特徴ベクトルとみなせる．主成分分析を用いて次元削減し，100 次元の特徴ベクトルに変換する．得られた特徴ベクトル集合に対し，これらをなるべく少ない数の基底ベクトルの線形結合で表現することを考える．Lu らは単一の巨大な辞書（基底ベクトルの集合）ではなく，小さな辞書を複数作成するようなアルゴリズムを採用し，非常に高速な異常検出技術を提案した．研究発表会にて実装と実験結果を報告する．

5. 参考文献

[1]. B Ravi Kiran, Dilip Mathew Thomas, Ranjith Parakkal, “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos”, arXiv:1801.03149 [cs.CV].

[2]. V. Mahadevan, W.-X. LI, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 1975–1981.

[3]. Kim, J.，Grauman, K.，“Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20‑25 June 2009, pp. 2921–2928.

[4]. Y. Cong, J. Yuan, and J. Liu, “Sparse reconstruction cost for abnormal event detection”, in proceedings of IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR), pages 3449-3456, 2011.

[5]. Cewu Lu, Jianping Shi, Jiaya Jia, “Abnormal Event Detection at 150 FPS in MATLAB,” ICCV 2013, pages 2720-2727.

[6]. B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. “Estimating the support of a high-dimensional distribution”, Neural Computation, 13(7), pages 1443–1471, 2001.

[7]. Dan Xu, Yan Yan, Elisa Ricci, Nicu Sebe, “Learning Deep Representations of Appearance and Motion for Anomalous Event Detection”, Computer Vision and Image Understanding, Volume 156, pages 117-127, 2017.

[8]. M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, “Plug-and-play cnn for crowd motion analysis: An application in abnormal event detection”, arXiv:1610.00307, 2016.

[9]. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, “Generative adversarial nets”, Advances in Neural Information Processing Systems (NIPS) 27, pages 2672--2680, 2014.

[10]. Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., Sebe, N., “Abnormal Event Detection in Videos using Generative Adversarial Nets”, In Proceedings of the IEEE International Conference on Image Processing (ICIP) 2017, Beijing, China, pages 17–20, September 2017.

ⓒ2019 Information Processing Society of Japan