Novelties in Intelligent Digital Systems
C. Frasson et al. (Eds.)

doi:10.3233/FAIA210103

Unsupervised Man Overboard Detection Using Thermal Imagery and Spatiotemporal Autoencoders

Nikolaos Bakalos[a], Iason Katsamenis[a], Eleni Eirini Karolou[a] and Nikolaos Doulamis[a]
[a]*National Technical University of Athens*

Abstract. Man overboard incidents in a maritime vessel are serious accidents where the rapid detection of the even is crucial for the safe retrieval of the person. To this end, the use of deep learning models as automatic detectors of these scenarios has been tested and proven efficient, however, the use of correct capturing methods is imperative in order for the learning framework to operate well. Thermal data can be a suitable method of monitoring, as they are not affected by illumination changes and are able to operate in rough conditions, such as open sea travel. We investigate the use of a convolutional autoencoder trained over thermal data, as a mechanism for the automatic detection of man overboard scenarios. Morever, we present a dataset that was created to emulate such events and was used for training and testing the algorithm.

Keywords. Man overboard; Human detection; Deep learning Computer, thermal image processing

1. Introduction

Man overboard refers to an emergency scenario where a ship passenger or crew member has fallen off the vessel and into the sea. With a mortality rate of 79%, an estimate of 22 people annually lose their lives due to such incidents[1], with the majority of them being untrained passengers. The high mortality is caused by the low speed of detection and retrieval, combined by the usual low temperature and rough conditions of the waters that can quickly result in drowning or hypothermia. Thus, the use of intelligent systems is imperative, in order to continuously monitor for such incidents and raise timely alerts. To this end, models based on deep learning paradigms used for the analysis of video streams have displayed great performance.

However, even these approaches have some drawbacks, as they rely on the use of RGB video streams, i.e. data streams monitoring over the visible spectrum. While the use of such data is popular, due to the cost efficiency of installing normal video surveillance systems, and the high performance of algorithms for object detection and classification over such data, these streams are greatly affected by illumination changes, and poor visual conditions. This indicates that the use of additional or alternative data modalities is needed. A valid alternative is video streams using thermal capturing devices. These devices monitor the infrared spectrum and are not affected by the change of lighting.

1.1. Previous Work

In a universal maritime surveillance system, human detection is a key issue and must be completely independent of the environment as well as light and weather conditions. Several human detection methods have been presented in the literate and have emphasized the importance of real-time home surveillance systems ([2], [3]) that focus on fall detection through visual sensors, deep learning and computer vision applications (e.g [[4][5][6]), however, little work has been presented in the literature on the man overboard situation.

In essence though the incident can be modelled as an abnormal behavior detection problem, where the normal situation consists of a normal capturing a seafaring vessel, while the abnormality would be the capturing of a fall. To this end, the main approaches for abnormal event recognition involve either the use of supervised deep learning techniques to learn a dictionary of abnormal sub-events or unsupervised outlier detection techniques. in many applications [7]-[9]. Examples include surveillance in industrial environments [7] or critical infrastructures [9] for safety/security and quality assurance, traffic flow management [10] and intelligent monitoring of public places [11]

Regarding outlier detection, the works of [12], [[13], [14] learn dictionary of sub-events, through a training process, and then those events that do not lie in the partitioned sub-space are marked as abnormal ones.

Regarding deep learning, the work of [15] employs convolutional auto-encoders (ConvAE) to learn temporal regularity in videos, while auto-encoders are exploited in [16] to learn feature and reconstruct the input images. Then, one-class Support Vector Machines (SVMs) are used for detecting the abnormal events. The work of [17] introduces a hybrid scheme which aggregates ConvAE with Long Short-Term Memory (LSTM) encoder-decoder. Recently, deep generative models have been applied [15]-[17]. These models are trained to produce normal events while the abnormal ones are given as the difference between the original frames and the generated ones.

Computer-vision tools that operate outside of the visible spectrum (i.e., thermal sensors) are also gaining traction in this context, because they are not significantly affected by illumination changes [18]. However, such approaches do not capture texture or color information. Vision techniques focus on background and target modeling [8], object tracking [19], activity recognition [20], crowd dynamics, and identification of unusual and suspicious behavior [21]. These approaches seek to detect abnormalities in crowded environments by analyzing actions on both the spatial and temporal scales. Detailed surveys about video-based abnormal activity recognition have been published [23], [24].

Recently, unsupervised learning models are utilized for abnormal event detection. In [25], the anomalies in videos are scored independently of temporal ordering and without any training by simply discriminating between abnormal frames and the normal ones. Other approaches exploit on-line incremental coding [26], deep cascading neural networks, and unmasking (a technique previously used for authorship verification in text documents) [27]. Recently, the works of [28] and [29] incorporate autoencoders and supervised learning for abnormal event detection. Other approaches employ tracking algorithms to extract salient motion information which is then classified either as normal or abnormal [30], [31]. However, tracking fails in complex visual scenes where multiple humans are present.

In this paper we present the use of an unsupervised fall detection method for man overboard scenarios. Our approach is based on the use of convolutional spatiotemporal

autoencoders trained using a thermal imagery dataset that simulates man overboard incidents.

2. Proposed System Architecture

The presented system using only thermal. video streams to identify overboard falls. Each image property was fed into the spatiotemporal autoencoder. Autoencoders are a type of Neural Network that manage to learn efficient data encodings by training the network to ignore signal noise. Their usefulness comes from the fact that they are trained in an unsupervised manner. They are essentially composed from two main components that are trained in parallel. The dimensionality reduction component aims at extracting an efficient encoding of the input signal, while the reconstruction side tries to generate from the reduced encoding a representation as close as possible to the original input. To identify the abnormalities, the reconstruction error of each autoencoder was monitored, and when the error was bigger than a predefined threshold, an alert was raised. The selection of the threshold took place during the training, to identify the exact value that maximized detection performance.

The autoencoders used for each image property had the structure presented in **Figure 1**. Each thermal frame was reduced to a grayscale image with a resolution of 227x227x1. A 10 frame batch was used for the analysis.

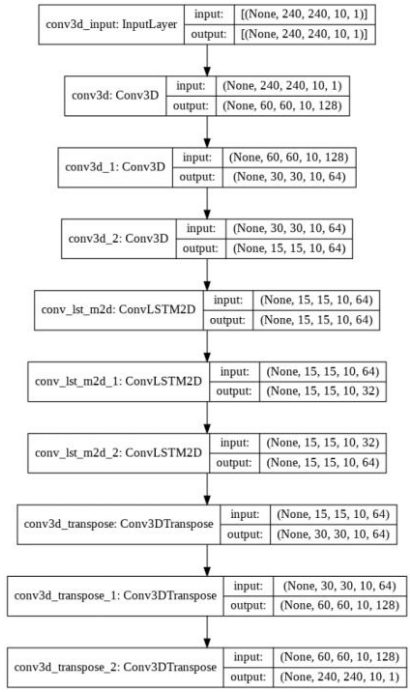

Figure 1. Proposed Model Structure

3. Dataset Description

In order to train and evaluate the proposed methodology, a mock man-overboard event was conducted that concerned the fall of a human-sized dummy from the balcony of a high-rise building. In particular, the human dummy, weighting 30 Kg, was thrown from an approximate height of 20 meters, which is roughly equivalent to two seconds of free-falling. For the needs of the experiment, we made 320 test throws of the dummy, to simulate a man-overboard event (see Figure 2(a)-(d)). Additionally, we recorded several videos without dropping the dummy as well as numerous throws of various objects, such as plastic bags and bottles (see Figure 2(e)). This way we can implement deep learning models that are not prone to false-positive alarms, triggered by non-human-related events.

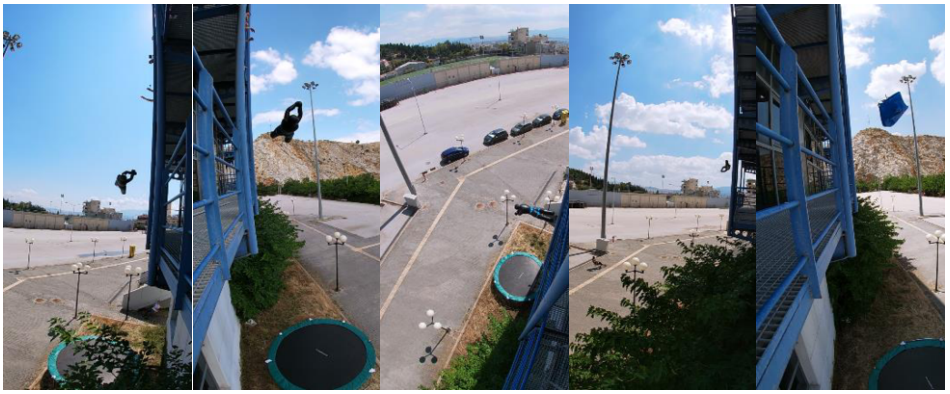

Figure 2. Test throws during the data collection experiments. The free fall (a)-(d) of the human dummy from different shooting angles (positive event), and (e) of a plastic bag (negative event).

The experiments took place in the surrounding area of Nikaia Olympic Weightlifting Hall, and lasted five days. Due to the fact that the test throws were carried out throughout the whole day, from 9:00 AM to 5:00 PM, the acquired videos vary in terms of illumination conditions (e.g. underexposure, overexposure). Additionally, we shot under various weather conditions (e.g. sunny, cloudy, rainy, windy, hot, cold), thus providing further variations in the background of the event.

In this paper, we are using a dataset consisted of RGB videos featuring the free falls of the dummy (see Figure 2(a)-(d)). For the dataset collection, which contains video sequences with a resolution of 1080×1920 pixels, we used a GoPro Hero 7 Silver. The camera was set to shoot at a high frame rate, at 50 frames per second, to ensure sufficient acquisition of data that concerns the critical event.

It is underlined that to avoid training bias and guarantee replicability of the results to other datasets, we placed the sensor in four different locations of the building, in order to obtain data that vary in terms of background, illumination, shooting angle, and distance (see Figure 2(a)-(d)). In particular, as depicted in Figure 3, we placed the RGB camera (i) on the left of the fall at a close distance of 7m (see Figure 2((a)), (ii) on the right of the fall at a close distance of 5m (see Figure 2((b)), (iii) on the top left of the fall at an angle of roughly 45° (see Figure 2((c)), and (iv) to the left of the fall at a long distance of 13m (see Figure 2((d)). It is emphasized that to further generalize the learning procedure, we augmented the training data by horizontally flipping the corresponding videos.

Figure 3. The four locations of the building that the optical sensor was placed, during the data acquisition experiments.

4. Experimental Evaluation

The proposed method was implemented in Python, using the Tensorflow and Keras libraries. The implementation used Python 3.6 and the Keras (1.08) and Tensorflow (2.1.0) machine learning libraries, in combination with a number of other scientific and data management libraries. The model was trained using the Google Collab Platform. The Area Under Curve (AUC) metric was employed in assessing the performance of the proposed method and the compared ones. The AUC is computed with regard to ground-truth annotations at the frame-level and it is a common metric for many abnormal event detection methods. It measures the ability of the learning algorithm to correctly distinguish normal from abnormal events and summarises the ROC curve of the system, i.e. the probability curve that plots the raising a true alert (true positive rate) and a false alarm (false positive rate) at various thresholds. Our algorithm achieves an AUC score of 88.. Due to the fact that there are no similar publications for fall detection in man overboard scenarios, at least to the authors knowledge, a comparative analysis of the performance is hard to achieve. The performance of our system using these metrics can be viewed in **Figure 4**.

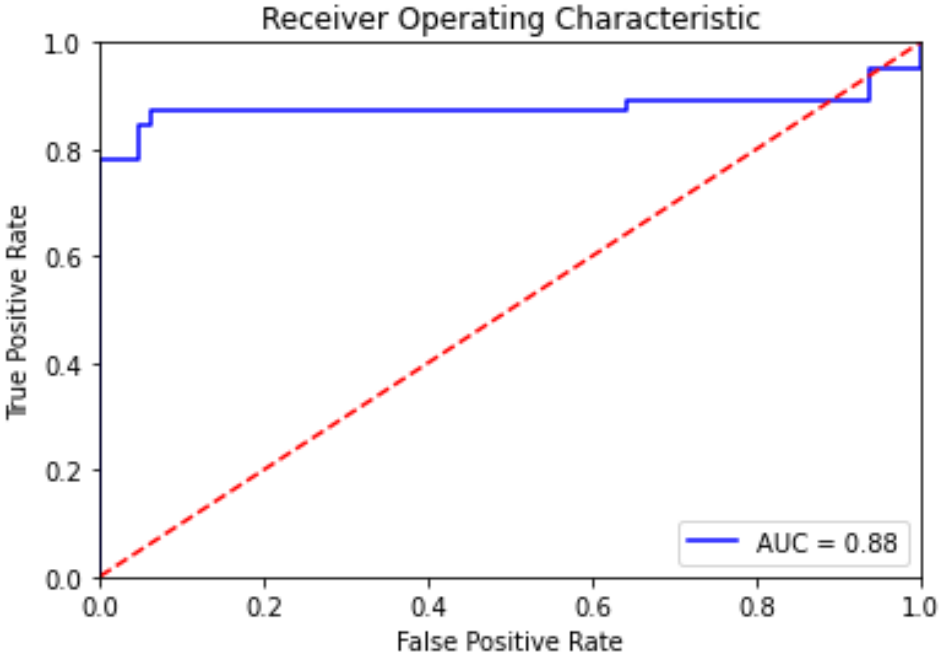

Figure 4: ROC Curve of spatiotemporal autoencoder

5. Conclusions

In this paper, we presented and evaluated a learning algorithm for man overboard detection over thermal data frames. The employed techniques use a deep machine learning framework, modelling a man overboard incident as an abnormal action recognition one. We then proceed in identifying falls by the autoencoders' success or failure to reconstruct a scene due to the presence of abnormal events.

Future work should include the fusion with additional imaging modules, such as normal RGB frames, and the studying of additional ways for inter and intra property encoding of all the available modalities to maximize the detection capabilities.

Acknowledgments

This paper is supported by the Greek Funded Project MHTIS No. 01169.

References

[1] E. Örtlund and M. Larsson, "Man Overboard detecting systems based on wireless technology," 2018.

[2] A. S. Voulodimos, D. I. Kosmopoulos, N. D. Doulamis, and T. A. Varvarigou, “A top-down event-driven approach for concurrent activity recognition,” Multimed. Tools Appl., vol. 69, no. 2, pp. 293–311, Mar. 2014, doi: 10.1007/s11042-012-0993-4.
[3] N. D. Doulamis, A. S. Voulodimos, D. I. Kosmopoulos, and T. A. Varvarigou, “Enhanced Human Behavior Recognition Using HMM and Evaluative Rectification,” in Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, New York, NY, USA, 2010, pp. 39–44, doi: 10.1145/1877868.1877880.
[4] K. Makantasis, E. Protopapadakis, A. Doulamis, N. Doulamis, and N. Matsatsinis, “3D measures exploitation for a monocular semi-supervised fall detection system,” Multimed. Tools Appl., vol. 75, no. 22, pp. 15017–15049, Nov. 2016, doi: 10.1007/s11042-015-2513-9.
[5] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, “Robust video surveillance for fall detection based on human shape deformation,” IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 5, pp. 611–622, 2011.
[6] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. Chambers, “A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment,” IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 6, pp. 1274–1286, 2012.
[7] S. Lee, H. G. Kim and Y. M. Ro, "BMAN: Bidirectional Multi-Scale Aggregation Networks for Abnormal Event Detection," IEEE Trans. on Image Proc., vol. 29, pp. 2395-2408, 2020.
[8] A.S. Voulodimos, N.D. Doulamis, D.I. Kosmopoulos, and T.A. Varvarigou, “Improving multi-camera activity recognition by employing neural network based readjustment,” Applied Artificial Intelligence, 26(1-2), 97-118, 2012.
[9] N. Bakalos, et al. "Protecting water infrastructure from cyber and physical threats: Using multimodal data fusion and adaptive deep learning to monitor critical systems." IEEE Signal Processing Magazine, 36.2, pp. 36-48, 2019.
[10] S. Wan, X. Xu, T. Wang and Z. Gu, "An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems," IEEE Trans. on Intell. Transportation Systems, (to be published)
[11] R. Leyva, V. Sanchez and C. Li, "Fast Detection of Abnormal Events in Videos with Binary Features," IEEE ICASSP, Calgary, AB, pp. 1318-1322, 2018.
[12] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, “Video anomaly detection and localization using hierarchical feature repre- sentation and Gaussian process regression,” IEEE CVPR, pp. 2909–2917, 2015.
[13] C. Lu, J. Shi, and J. Jia, “Abnormal Event Detection at 150 FPS in MATLAB,” IEEE ICCV, pages 2720–2727, 2013.
[14] H. Ren, W. Liu, S. I. Olsen, S. Escalera, and T. B. Moes- lund, “Unsupervised Behavior-Specific Dictionary Learning for Abnormal Event Detection,” Proc. of BMVC, pp. 28.1–28.13, 2015.
[15] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” IEEE CVPR, pages 733–742, 2016.
[16] Xu Dan, Ricci Elisa, Yan Yan, Song Jingkuan and Sebe Nicu, "Learning Deep Representations of Appearance and Motion for Anomalous Event Detection", BMVC, 2015.
[17] L. Wang, F. Zhou, Z. Li, W. Zuo and H. Tan, "Abnormal Event Detection in Videos Using Hybrid Spatio-Temporal Autoencoder," 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 2276-2280, 2018.
[18] K. Makantasis, A. Nikitakis, A. Doulamis, N. Doulamis, and Y. Papaefstathiou, “Data-driven background subtraction algorithm for in-camera acceleration in thermal imagery,” IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 9, pp. 2090–2104, Sept. 2018.
[19] S. Herrero and J. Bescs, “Background subtraction techniques: Systematic evaluation and comparative analysis,” in Advanced Concepts for Intelligent Vision Systems (Lecture Notes in Computer Science, vol. 5807), J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, Eds. Berlin: Springer-Verlag, 2009.
[20] D. S. Yeo, “Superpixel-based tracking-by-segmentation using Markov chains,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017, pp: 511–520.
[21] D. Kosmopoulos, A. Voulodimos, and A. Doulamis, “A system for multicamera task recognition and summarization for structured environments,” IEEE Trans. Ind. Informat., vol. 9, no. 1, pp. 161–171, 2013.
[22] H. M. Mousavi, “Analyzing tracklets for the detection of abnormal crowd behavior,” in Proc. IEEE Winter Conf. Applications Computer Vision, 2015, pp. 148–155.
[23] S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, “Trajectory-based surveillance analysis: A survey,” IEEE Trans. Circuits Syst. Video Technol., 2018. doi: 10.1109/TCSVT.2018.2857489
[24] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep learning for computer vision: A brief review,” Computational Intell. Neurosci.,2018, article ID 7068349. [Online]. Available: https://doi.org/10.1155/2018/7068349

[25] A. Del Giorno, J. Bagnell, and M. Hebert, "A Discrimina- tive Framework for Anomaly Detection in Large Videos," Proc. of ECCV, pp. 334–349, 2016.
[26] J. K. Dutta and B. Banerjee, "Online Detection of Abnormal Events Using Incremental Coding Length," In Proceedings of AAAI, pages 3755–3761, 2015.
[27] R.T. Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, "Un-masking the abnormal events in video," IEEE ICCV, pp. 2895–2903, 2017.
[28] S. Smeureanu, R. T. Ionescu, M. Popescu, and B. Alexe, "Deep Appearance Features for Abnormal Behavior Detection in Video," In Proceedings of ICIAP, Volume 10485, pages 779–789, 2017.
[29] Y. Liu, C.-L. Li, and B. Poczos, "Classifier Two-Sample Test for Video Anomaly Detections," In Proceedings of BMVC, 2018.
[30] X. Mo, V. Monga, R. Bala, and Z. Fan, "Adaptive sparse representations for video anomaly detection," IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 4, pp. 631–645, Apr. 2014.
[31] F. Jiang, Y. Wu, and A. K. Katsaggelos, "A dynamic hierarchical clustering method for trajectory-based unusual video event detection," IEEE Trans. Image Process., vol. 18, no. 4, pp. 907–913, Apr. 2009.