For the Sake of Privacy: Skeleton-based Salient Behavior Recognition

[1] Fraunhofer Institute for Optronics, System Technologies and Image Exploitation IOSB
[2] Vision and Fusion Lab, Karlsruhe Institute of Technology KIT

Thomas Golda[1,2], Johanna Thiemich[1], Mickael Cormier[1,2], and Jürgen Beyerer[1,2]

Figure 1. Exemplary result for the GOPR0854 video from the VFP290k test set. The non-salient behaving person is drawn in white, the anoma-lous one in red. The red curve at the bottom shows the presence of a fallen person (ground truth) over on frame level over time, whereas the red shaded areas indicate when the method rates the situation as salient. Furthermore, the blue and green line above indicate the anomaly scoring on instance level for the two persons over time.

Table 1. AUC ROC values on three different datasets

Dataset	Method	Frame	Instance
ShanghaiTech campus	MPED-RNN*	0.7703	0.6673
	GEPC	0.6630	0.6192
VFP290k	MPED-RNN*	0.6934	0.5698
	GEPC	0.6403	0.6167
	Ours*	0.9602	0.9440
Internal	MPED-RNN*	0.5957	0.5243
	GEPC	0.5893	0.5504
	Ours*	0.7213	0.6840

Table 2. Contribution of different architecture parts.

Dataset	Method	Frame	Instance
VFP290k	Normal	0.6216	0.6081
	+ Abnormal	0.9259	0.9164
	+ Memory	0.9602	0.9440
Internal	Normal	0.6826	0.6285
	+ Abnormal	0.7231	0.6830
	+ Memory	0.7213	0.6840

Problem Statement

- Increasing interest in smart video surveillance for public safety and in urban scenarios
- Various terrorist attacks and tragic accidents during festivals in Europe
- But: Bad reputation of video surveillance systems in society, especially due to data privacy issues and misuse
- Upcoming interest in identity agnostic assisting systems for behavioral analysis from authorities
- In particular: anomaleous and salient behavior recognition

Skeleton-based Anomaly Detection

- MPED-RNN [1] and GEPC [2] are the only existing approaches addressing skeleton-based anomaly detection
- Input:
  - Temporal sequence of human skeletons
- Output:
  - Saliency or anomaly score
- Experiments on two datasets:
  - VFP290k [3] (falling people)
  - Internal real world dataset (fighting and aggressive behavior)

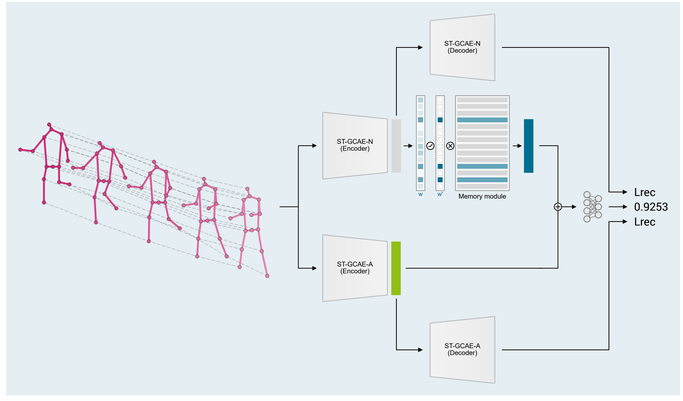

Figure 2. Two-branched model architecture with an memory unit for normal behavior which is based on the approach proposed by Gong et al. [4] and applied to skeleton-based data.

Base Architecture

- Two branches generating representations for salient and non-salient behavior
- Skeleton sequence as a spatio-temporal graph
- Graph Convolutional Layers als feature extractors
- Memory unit for normal behavior branch

Conclusion

- Încreasing interest in data privacy friendly approaches for behavior analysis
- Certain scenarios described by public datasets like ShanghaiTech campus or VFP290k are quite „simple“
- Real-world scenarios like those captured in our internal dataset are still very challenging
- Getting high quality human skeletal representations is the key and still needs special focus when dealing with video surveillance setups

1 Romero Morais, Vuong Le, Truyen Tran, Budhaditya Saha, Moussa Mansour, and Svetha Venkatesh, “Learning regularity in skeleton trajectories for anomaly detection in videos,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11988–11996.
2 Amir Markovitz, Gilad Sharir, Itamar Friedman, Lihi Zelnik-Manor, and Shai Avidan, “Graph embedded pose clustering for anomaly detection,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10536–10544.
3 Jaeju An, Jeongho Kim, Hanbeen Lee, Jinbeom Kim, Junhyung Kang, Minha Kim, Saebyeol Shin, Minha Kim, Donghee Hong, and Simon S. Woo, “VFP290k: A large-scale benchmark dataset for vision-based fallen person detection,” in Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), 2021.
4 Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel, “Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection,” in IEEE International Conference on Computer Vision (ICCV), 2019.