Journal of ELECTRICAL ENGINEERING, VOL 73(2022), NO5, 318–324

sciendo

PAPERS

Graph based anomaly detection in human action video sequence

Pranoti Shrikant Kavimandan[1,*], Rajiv Kapoor[2], Kalpana Yadav[3]

In our paper, we have proposed to use graphs to detect anomaly in human action video. Although the detection of anomaly is a widely researched topic, but very few researchers have detected anomaly in action video using graphs. in our proposed method we have represented the smaller section (sub-section) of our input video as a graph where vertices of the graph are the space time interest points in the sub-section video and the association between the space time interest points exists. Thus, graphs for each sub section are created to look for a repeated substructure. We believe most of the actions inherently are repeated in nature. Thus, we have tried to capture the repetitive sub-structure of the action represented as a graph and used this repetitive sub-structure to compress the graph. If the compressed graph has few elements that have not been compressed, we suspect them as anomaly. But the threshold value takes care not to make the proposed method very much sensitive towards the few uncompressed elements. Our proposed method has been implemented on locally created "extended KTH" and "extended Weizmann" datasets with good accuracy score. The proposed method can also be extended for few more applications such as training athletes and taking elderly care.

K e y w o r d s: anomaly detection, graphs interpretation, substructures, space-time interest points

1 Introduction

Anomalous behavior detection has gained tremendous importance in the field of computer vision. The exponential use of gadgets collecting images, videos and audios has aroused the need to analyze the data before we permanently store that huge amount of data. This analysis of data is important from several points of view. One of the major reasons the data is analyzed is to find out the presence of an anomaly in the data. To find out presence of rare occurrences in datasets using data mining approach is one among many definitions of anomaly as jotted down in [1]. It is very natural that rare occurrences catch the attention more than the general repetitive structures. Thus, the aim in anomaly detection is to look for the structure which occurs less frequently. Anomaly detection is one of the major applications in several fields such as human motion analysis, discovery of rare disease in medical field, discovery of rare molecular structure in chemistry, detection of fraud in a kind of transaction and the list contains many more such applications. In our proposed method we are detecting an anomaly while a human is performing some action. Action here means a repetitive pattern and an anomaly here would be anything that is not expected to happen when a human is performing same action repetitively. Whenever the pattern is not repeated our method flags it as on anomaly. The proposed methodology uses graph theory approach to detect an anomaly since graph are insensitive towards the change in background and illumination. Anomaly detection using graphs is also comparatively less explored area and thus we have chosen to work upon the same. This approach for anomaly detection in human action is one of its kind. Our assumption is that the events in a video are normal. The action is always getting performed in the way it is supposed to. But the moment something anomalous happens the pattern of graph that repeats is not detected that means something which is not expected has happened. We termed it as an anomaly. Many a times a simple diversion from regular action may happen, such as the jump action by the actor may be interrupted to tie up the shoelaces. Proposed methodology is not very sensitive to such rare diversion because we have a threshold value associated with the video. This threshold value decides to call an action an anomaly. The standard datasets for action recognition have only normal data, thus we have tried to create a local database called as extended KTH and extended Weizmann, very similar to the standard human action recognition datasets KTH, [2] and Weizmann [3] to detect the anomaly with good accuracy score. Our proposed method could also be extended to use for the safely of elderly people. Anything that happens different from their routine could be captured as an anomaly and a timely concern could be raised to reach out help for them. We propose a novel method of using graphs to detect anomaly in action. We deal with over-sensitiveness of a system to label an action anomalous and propose a robust method to changes in background and illumination.

[1] Department of Computer Science and Business Systems, Bharati Vidyapeeth Deemed to be University College of Engineering, Pune, India, Corresponding author: pranoti.sk@gmail.com, [2] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India, [3] Department of Information Technology, Indira Gandhi Delhi Technical University for Women, Delhi, India

https://doi.org/10.2478/jee-2022-0042, Print (till 2015) ISSN 1335-3632, On-line ISSN 1339-309X

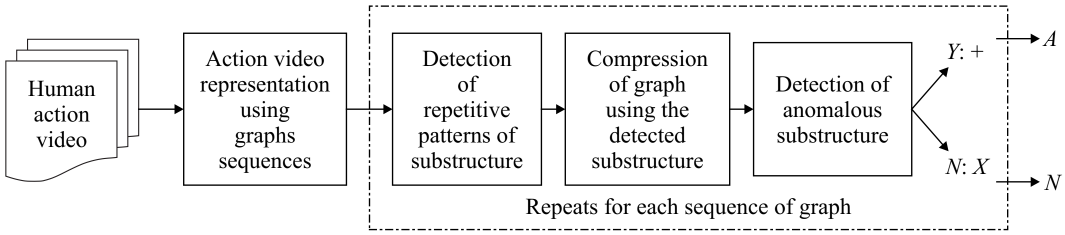

Fig. 1. Proposed methodology to detect anomaly in an action

2 Literature survey

Graphs have been studied and widely implemented in the field of computer science. But their use in detection of anomaly is comparatively less. Human actions in the video are represented using skeleton graphs of foreground objects that deform in [5]. Deforming skeletons are matched to recognize the human action. The dynamism in the method makes it consume more time. Authors of [6] have used autoencoder long short-term memory network for the reconstruction of video from raw data. The errors during the reconstruction are detected as the anomalies present in the data. They have used convolutional Autoencoders to encode spatial variations whereas temporal variations are modeled by a long short-term memory network. The background impact on detection is well overcome by weighted Euclidean loss that concentrates only on moving foreground. Assigning labels in supervised learning is not only costly but also is susceptible to human errors. Thus, authors of [7] and [8] have used unsupervised learning to detect anomalies in the video. In [7] the motion and intensity data are represented at multiple levels. Authors have used denoising autoencoders for learning representations. The conditional generative adversarial networks are used to generate level-wise representation. An anomaly is detected based on consolidated anomalous regions in all levels of representations. In [8] an unsupervised learning to detect an anomaly in the video was used. They have proposed using ensemble random projection-based reconstruction loss neural network that is a novel three staged framework to detect anomaly in a video. In [9] authors have used Schrödinger equation to detect anomaly in a video during run time. They proposed to use change in the kinetic energy at the scene as a criterion to classify if the video belongs to the normal or abnormal class. In [10], authors have used the best possible combination of joint and bone data by representing it as directed acyclic graph according to the kinematic dependency that exists between them in a human body. The two-stream framework proposed by the authors, have very well used both the information gathered due to the motion and the spatial data that has relatively enhanced the accuracy of their proposed model. Authors of [11] have assigned a combined score of two networks namely RGB and flow two stream networks and thus have overcome the issue of complementary information that often remains unexplored. In [12] authors have represented an activity as a graph where the nodes of the graph are spatial-temporal interest points. If appearance and dynamics between the interest points are related, then the edge is present in the activity graph. They have used support vector machine to classify normal and abnormal activities.

Use of a graph in detection of an anomaly is not just limited to human action videos, but also is extensively applied in different streams such as networking, fraud detection, detection of suspicious patterns in video and many more. In [25], authors have used latent features that are present in graphs to reduce the false positives while detecting anomalies. The network traffic is first converted into a first-order graph from the local point of view and later converted into a second-order graph from the global point of view. This feature makes the system independent of manual interference. Moreover, their proposed method is also capable of discovering unknown network attacks. Authors of [13] have implemented three different algorithms to deal with three different types of graph anomalies caused by change in label, insertion and deletion of any node or edge. The main application of their algorithms is in detection of any fraud/anomaly in graph-based data. According to [14], graph-based anomaly detection methodologies have been instrumental in detection of fraud activity in network and have been considered as strong and reliable filed that has attracted severe attention. In [15], authors have used graphs to deal with the anomaly hidden between suspicious patterns in the video. They have used the data gathered from physical access control system to discover the suspicious patterns observed in the video.

3 Proposed methodology

Figure 1 shows the working of the proposed methodology. The input video is divided into smaller sections with n frames. Each smaller video section is processed further to find out the space-time interest points (STIP). Using these STIPs we constructed graphs $G(V, E)$ for each video section. The use of graphs gives the advantage of being insensitive to local changes such as background and illumination. The proposed method further deals with each graph individually. Each graph is processed to find the best repetitive sub-structure [16]. This sub-structure is a part of a graph that repeats frequently

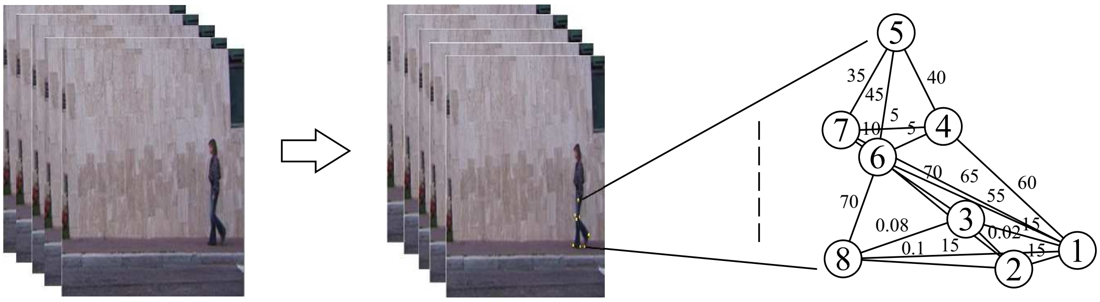

Fig. 2. Sub-section of action "walk" from the standard dataset weizmann, respective STIP images and graph constructed from STIP (actor Daria)

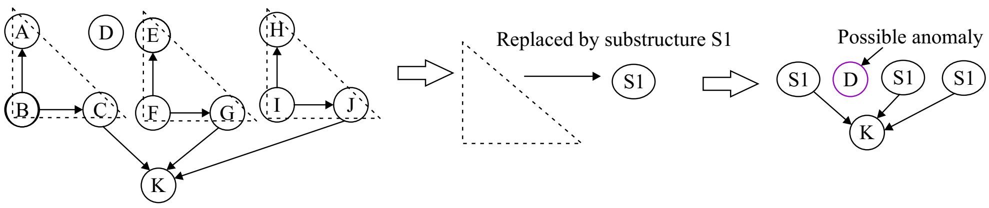

Fig. 3. Diagram showing the sub-structure used to compress the graph and the detection of the possible anomaly

and can be used to compress the graph. Since we are considering the human actions that are repetitive, our intuition here is if there lies some part in a graph that is not the part of repetitive sub-structure, it could be a possible anomaly as discussed earlier in the introduction. But an anomaly detected in one graph may or may not be the true anomaly. For an instance, while performing jump" action if the actor realizes that the shoelaces are not tied properly, the actor may stop the jump" action and tie the shoelaces to avoid any mishap. This kind of action would not be a repetitive one but should also be not called an anomaly. Our proposed method takes care of such a situation that happens for a small period. The anomaly is detected individually/locally in each graph and the associated anomaly score is incremented. If the overall score crosses the set threshold value, we can say that the video has an anomaly present in it. The proposed methodology is further discussed in detail in subsequent sub sections.

3.1 Representation of video sections as graphs

As discussed above in the proposed method, the input video is divided into smaller video sections with n frames in each section. The space time interest points (STIP) in each subsection are found out using the method described in [18, 19]. The STIP plays a very important role because the data analysis done by the traditional methods such as optical flow does not capture the sudden change very well. But interest points focus on the sudden changes in the movement and thus can offer important information in the video related to the motion. Several researchers have done remarkable work to find STIP and take good advantage of it for further applications [20, 21]. Authors of [22] have modified the STIP version of [18] to increase the number of interest points. [23] enhanced the detection of interest points with the use of Gabor filter. In [24] authors have proposed the method to detect interest points that are scale invariant. Although method by [18] has a limitation of sparse points, but the method works good for graph application. As mentioned in [18], Harris corner function (H) is modified to find out the interest points in the temporal domain and spatial domain both using equation 1. It takes into consideration the temporal domain and modify the determinant as well as trace of windowed second moment matrix (μ). The eigen value $e_1, e_2$and e_3 are kept high. The value of constant c is experimentally decided. Once the STIP are detected for the sub-section of the graph, a graph $G(V,E)$ is constructed using the STIPs as the vertices and edges are represented based on the association of vertices. Figure 2 shows sub-section of the action "walk" from the standard dataset Weizmann, its respective STIP images and the graph constructed from the STIP"s of the sub-section of video. Each STIP corresponds to a vertex in graph as shown in figure. As can be seen from the graph there is no association between vertex 1 and 5, because the two vertices are too far on the timeline and the proposed method takes care that the association is not included in the corresponding graph. Similarly, graph $G(V,E)$ for each subsection of the video is created

$$\begin{aligned} H &= \det(\mu) - c\,\mathrm{trace}(\mu)^3 = \\ &= \big(e_1 e_2 e_3 - c(e_1 + e_2 + e_3)^3\big). \end{aligned} \tag{1}$$

3.2 Repetitive pattern detection in video section and compression of the graph

The advantage of the action databases used in proposed methodology is that the videos in the databases can be represented as graph that is basically a well-formed

Fig. 4. Dataset frames showing "walk" action from Weizmann dataset and its STIP (actor Daria)

Fig. 5. Dataset frames showing "run" action from Weizmann Dataset and its STIP (actor Daria)

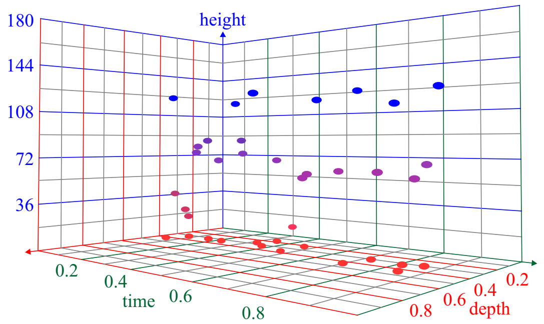

Fig. 6. STIP for subsection of "walk" action represented in the temporal domain

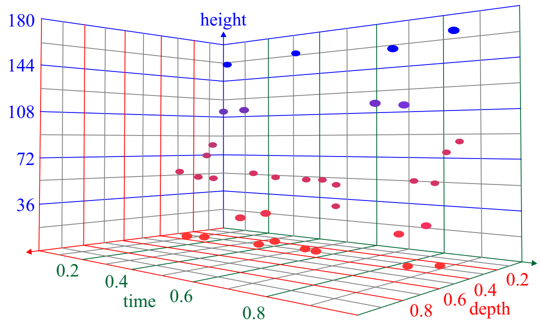

Fig. 7. STIP for subsection of "run" action represented in the temporal domain

structure just like the actions. And thus, we have an intuition that there could be a presence of repetitive structures with important information in a video that represents a well-defined action". These repetitive structures can also be used to compress the data represented in the form of graphs. Moreover, the repetitive structure (sub-structure) also helps to determine if the action being performed is getting done in an expected way or if any unexpected change has happened. The unexpected change happened could indicate the possible presence of an anomaly. Thus, our goal is to find out the repetitive pattern in our data that is represented as graph. Once the repetitive substructures are detected the data can very well be compressed using those sub- structures. Figure 3 shows how the sub-structures are compressed to determine the possible anomaly that is the most infrequent part of the graph. The repetitive structure shown in dashed triangle is replaced by new vertex S1 and the graph is compressed replacing all dashed triangles with this sub-structure node S1. No structure is matched with node D and thus is left unchanged, resulting in being a suspicious node.

3.3 Detection of anomaly locally and globally

The anomaly that is detected as discussed above, is purposefully termed as possible anomaly, since the presence of just one vertex which is not the part of sub-structure can not be called as an anomaly. As mentioned above, we term an action in our database as the thing that occurs repeatedly. Repeatedly occurring thing is bound to create similar graph structure and similar structures will create a sub- structure that can be compressed in a graph. But if some structure is not getting repeated, we consider it as a possible anomaly. We termed it as "local anomaly", basically that means it is some unexpected thing happened at some place in sub-section of video. But if such substructure occurs more than some preset threshold times in the whole video, we finally flag it as an anomaly. This anomaly is termed as "global anomaly". The presence of global anomaly is the final indication that the video has anomaly present in it. Later, it could be checked to decide if the detected anomaly is true positive or false positive. The next section explains the results we got for the proposed method in detail.

4 Results and discussion

The proposed method has been implemented in different way on two standard action datasets Weizmann and KTH. The standard datasets consist of normal videos. The proposed method aims to find out anomaly in an action video. The standard dataset videos have no inherent anomaly in them and thus due to lack of such database

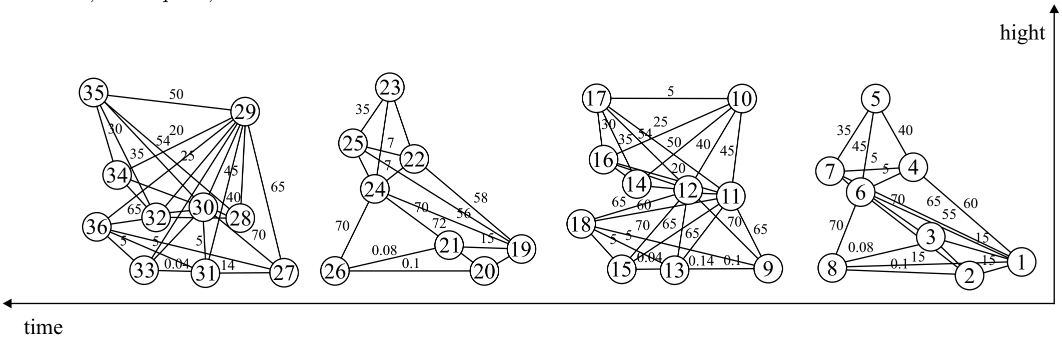

Fig. 8. The sub-section of the video of action "walk" represented as a graph $G(V, E)$

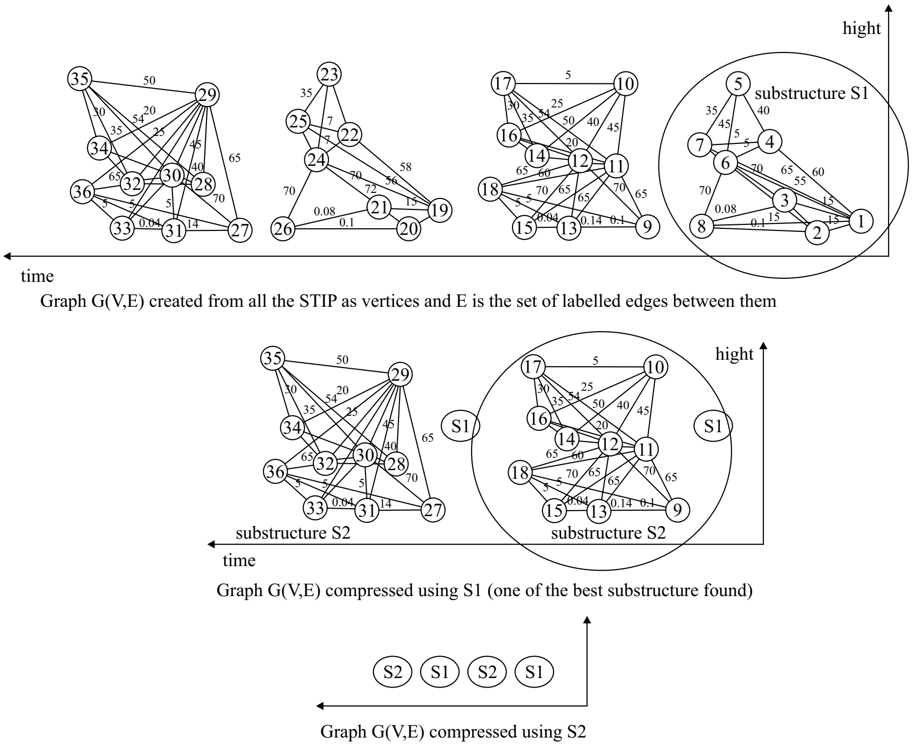

Fig. 9. The process of compression using the best substructure on the subsection of "walk" video

we have created extended Weizmann and extended KTH datasets locally. These datasets are created by combining a small part of any action video with other action at several intervals. For example, we merged a small part of videos of actions run", "bend", "jump in place" *etc* with the video of action "walk". Thus, a small part of video of action "run" in video of action "walk" is an unexpected happening and is termed as "local anomaly". If same kind of action other than "walk" is observed in a video of action "walk" it will also be termed as local anomaly. When the preset threshold value of local anomalies is reached, the proposed method flags it as "global anomaly". The intuition was that the action "walk" when performed by same actor under similar background condition generates graphs with repetitive structures. The repetitive sub structures during the whole graph indicates that the action is being carried out well, but if the action does not generate repetitive substructure there is a probability that anomaly is present in the video. Thus, taking this intuition forward we generated the results of the implementation of our proposed method. Figure 4 shows the frames from the sub-section of video for "walk" action and its corresponding STIP from Weizmann dataset performed by actor "Daria". Figure 5 shows the same kind of structure for "run" action. The STIP of sub-section of both the actions represented in the temporal domain are shown in Fig. 6 and Fig. 7.

The several such graphs of sub-sections as shown in Fig. 8 together become the graph of whole video. It is very clear from these graphs that they have several repetitive structures that are used to compress the graph. The whole process of compression of graph using repetitive structure and anomaly detection is shown in Fif. 9. At first the sub-structure S1 is detected which is like other graphical structures and thus could be compressed. Later the sub structure S2 is also detected and compressed. the resultant graph is shown in Fig. 9. Since there has not been any sub-structure observed that has not been compressed, our proposed method concludes that there is not any anomaly in the video.

The proposed method has followed the semi supervised learning paradigm. The method has been imple-

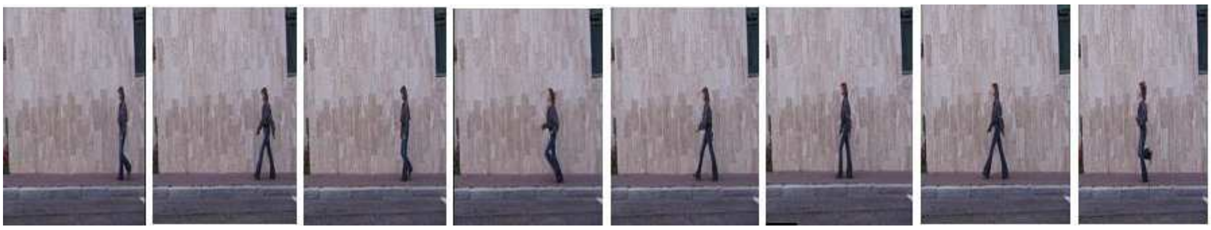

Fig. 10. The glimpse from locally created "Extended Weizmann" where the performing actor is Daria

Table 1. The graph for action "run" is shown in comparison to other graphs for action "walk"

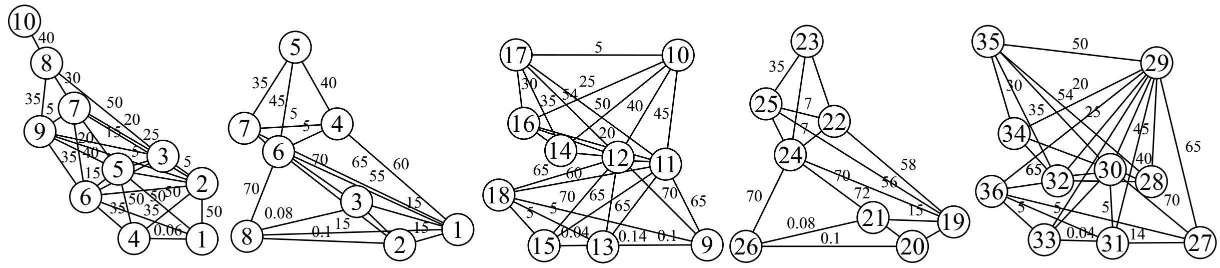

Table 2. Accuracy score (in %) extended Weizmann and extended KTH" datasets at various threshold values

Action from extended Weizmann dataset

Threshold value:	2	3	4
Walk + Run	85	86	78
Walk + Jump in place	96	95	89
Walk + Bend	91	90	84
Walk + One hand wave	98	96	91
Walk + Jumping Jack	97	95	90

Action from KTH dataset

Threshold Value:	2	3	4
Box + Clap	95	95	88
Box + Wave	98	97	91
Box + Jog	89	88	84
Box + Run	90	90	83
Box + Walk	92	91	86

mented on the standard datasets Weizmann, KTH and their extended versions "extended Weizmann" and "extended KTH" as well. This kind of anomaly detection on action recognition is nave and thus lack standard dataset. Figure 10 shows the glimpse of locally created database in which the small part of video of action "run" is merged with the video of action "walk" at several locations. The corresponding graph for one of the frames of action "run" is shown in comparison to other graphs for the action "walk" in Table 1. The only similarity is that all of them are graphs, but no graph of action "walk" is same as the graph of action "run". Thus, the presence of such a graph corresponding to action "run" in other graphs for action "walk" is "local anomaly". If the presence of this kind of anomalous graph is detected more than the preset threshold value, it is called as global anomaly and the video would be called as anomalous. Other actions such as "bend", "skip", "one hand wave" and "jumping jack" are also merged and checked for detection of anomaly. Same kind of local dataset is also created for KTH action database. The threshold value parameter is tuned well to get a very good accuracy score for both the datasets.

The accuracy of the proposed method is calculated based on true and false positive and true and false negative. The accuracy with different threshold values is also calculated for "walk action" of Weizmann dataset merged with other actions of Weizmann dataset and "box" action of KTH dataset merged with other actions of KTH dataset and is shown in Tab. 2. It is clear from the table that the significantly high value for threshold doesn't give good accuracy since the number of true positive goes down. The relatively low value of threshold makes the system very sensitive to small changes in the action video resulting in increase in false positive. Thus, we concluded that the threshold value of 2 or 3 is a relatively good choice.

5 Conclusions and future work

In our proposed method we implemented a graph-based anomaly detection in human action sequences. The input video that is divided into smaller sub-sections is represented as graphs $G(V, E)$ where, vertices of graphs are the STIPs and edges of the graph are the association between two STIPs. This graph is looked upon for repetitive sub-structures with the intuition that most of the actions are repetitive in nature. The graph is then compressed using these repetitive sub-structures. If the compressed graph still has some section that is not part of repetitive sub-structure and thus hasn"t get compressed, we call them as suspicious sub-section. If such suspicious sub-sections increase in number than the preset threshold our proposed method tags the video as anomalous. We experimented the proposed method on standard datasets Weizmann and KTH and locally created versions of both the datasets called as "extended Weizmann" and "extended KTH". Since there are no such action datasets for anomaly detection, we created these local databases with a unique technique. Our proposed method has shown very good results on this dataset. The proposed method has a good future scope in various fields. This kind of anomaly detection could be very helpful for sports trainers or athletes to practice or to find how to improve specific postures that could help them achieve their goals. Also, it could be very helpful to take care of elderly people.

References

[1] L. Akoglu, H. Tong and D. Koutra, *Data MKnowl Disc*, vol. 29, 626-688, 2015, https://doi.org/10.1007/s10618-014-0365-y.

[2] C. Schüldt, I. Laptev and B. Caputo, "Recognizing human actions: a local SVM approach", *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, 32-36, 2004, doi: 10.1109/ICPR.2004.1334462.

[3] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes", *Transactions on Pattern Analysis and Machine Intelligence*, www.wisdom.weizmann.ac.il/∼ vision /SpaceTimeActions.html.

[4] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes", *10th IEEE International Conference on Computer Vision*, 1395-1402, 2005, www.wisdom.weizmann.ac.il /∼vision/SpaceTimeActions.html.

[5] A. Abdulmunem, L. Y-Kun, A. Hassan, and S. Xianfang, "Human action recognition using graph matching", *7th International Conference on Applied Science and Technology*, Proc. 2144. 050003-1-050003-10, 2019, https://doi.org/10.1063/1.5123119.

[6] B. Yang, C. Jinmeng, N. Rongrong, and L. Zou, "Anomaly Detection in Noving Crowds through Spatiotemporal Autoencoding and Additional Attention", *Advances Multimedia*, https://doi/org/10.1155/2018/2087574.

[7] H. Vu, T. D. Nguyen, T. Le, W. Luo, and D. Phung, "Robust anomaly detection in videos using multilevel representations", *Proceedings of AAAI19-Thirty-Third AAAI conference on Artificial Intelligence*, 5216-5223, 2019, https://doi.org/10.1609/aaai.v33i01.33015216.

[8] H. Jingtao, E. Zhu, S. Wang, X. Liu, X. Guo and J. Jin, "An Efficient and Robust Unsupervised Anomaly Detection Method Using Ensemble Random Projection in Surveillance Videos", *Sensors*, 4145, 2019, https://doi.org/10.3390/s19 194145.

[9] R. Kapoor O. Mishra and M. Tripathi, "Anomaly detection in group activities based on fuzzy lattices using Schrdinger equation", *Iran Journal of Computer Science*, https://doi.org/10.1007/s42044- 019-00045-y.

[10] L. Shi, Y. Zhang, J. Cheng, and H. Lu "Skeleton-Based Action Recognition With Directed Graph Neural Networks", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, doi: 10.1109/CVPR.2019.00810.

[11] W. Hao, R. Zhang, S. Li, J. Li, F. Zhao, and Z. Zhang, "Anomaly Event Detection in Security Surveillance Using Two-Stream Based Model", *Security and Communication Networks*, 2020, Article ID 8876056, 2020, https://doi.org/10.1155/2020/8876056.

[12] D. Singh and K. Mohan, "Graph formulation of video activities for abnormal activity recognition", *Pattern Recognition*, 65, 265-272, 2017, https://doi.org/10.1016/j.patcog.2017.01.001.

[13] W. Eberle and L. Holder "Discovering Structural Anomalies in Graph-Based Data", *Proceedings - IEEE International Conference on Data Mining*, 393-398, 2007, 10.1109/ICDMW.2007.91.

[14] T. Pourhabibi, K. L. Ong, B. H. Kam, and Y. L. Boo, "Fraud detection: A systematic literature review of graph- based anomaly detection approaches", *Decision Support Systems*, vol. 133,2020, https://doi.org/10.1016/j.dss.2020.113303.

[15] L. Davis, W. Liu, P. Miller, and G. Redpath, "Detecting anomalies in graphs with numeric labels", *Proceedings of the 20th ACM international conference on Information and knowledge management*, Association for Computing Machinery, New York, NY, USA, 1197–1202, 2011, doi:https://doi.org/10.1145/2063576.2063749.

[16] C. Noble and D. Cook, "Graph-based anomaly detection", *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, https://doi.org/10.1145/956750.956831.

[17] D. Cook and L. Holder, "Graph-Based Data Mining", *IEEE Intelligent Systems*.

[18] I. Laptev, "On Space-Time Interest Points", *Int J Comput Vision*, 107–123, 2005, https://doi.org/10.1007/s11263-005-1838-7.

[19] I. Laptev and T. Lindberg, "Space-time interest points", *International Journal of Computer Vision - IJCV*, vol. 1, (2003), https://doi.org/10.1109/ICCV.2003.1238378.

[20] R. Kapoor, O Mishra, and M. Tripathi, "Human action recognition using descriptor based on selective finite element analysis", *Journal of Electrical Engineering*, 443-453, 2019, https://doi.org/10.2478/jee- 2019-0077.

[21] O. Mishra, R. Kapoor, and M.Tripathi, "Human Action Recognition Using Modified Bag of Visual Word based on Spectral Perception",*International Journal of Image, Graphics and Signal Processing*, 11, 34-43, 2019, https://doi.org/10.5815/ijigsp.2019.09.04.

[22] I. Bellamine and H. Tairi, "Motion Detection using the space-time interest points", *Journal of Computer Science*, 828-839, 2014, https://doi.org/10.3822/jcssp.2014.828.839.

[23] P. Dollar, V. Rabaut, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features", *Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 15-16, IEEE Xplore Press, 65-72, 2005, doi: 10.1109/VSPETS.2005.15 70899.

[24] G. Willems, T. Tuytelaars, and L. V. Gool, "An efficient dense and scale-invariant spatio-temporal interest point detector", *Proceedings 10th European Conference on Computer Vision*, 12-18, Springer Berlin Heidelberg, Marseille, France. 650-663 (2008). doi: 10.1007/978-3-540-88688-4_48.

[25] Q. Xiao, J. Liu, Q. Wang, Z. Jiang, X. Wang, and Y. Yao, "Towards Network Anomaly Detection Using Graph Embedding", *In: Krzhizhanovskaya V.V. et al (eds) Computational Science*, vol. 12140, Springer, 2020, https://doi.org/10.1007/978-3-030-5023-6_12.

Received 15 August 2022