IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR349
DOI: https://doi.org/10.52731/liir.v006.349

How Computer Vision Complements Health Care/Nursing Environment -Anomaly Detection with Hierarchical Classification of Human Motion-

Kayako Miya [*], Atsuo Yoshitaka [†]

Abstract

Aging of the population revealed that people of younger generation are facing to take care of more of old generation. It may not be satisfactory to arrange enough number of nursing staffs even in late-evening in hospitals or nursing and personal care facility, or it is not possible to watch every old people continuously in daytime. Detecting anomaly of human motion such as falling down, while no personal care assistant or nurse is accompanied, is expected to complement the insufficiency or lack of care by such staff. However, simple detection of anomaly of human motion such as falling down or sudden stoppage of the staff should be ignored as 'false positive' but adaptive detection of such motion by the persons who need assistance is preferable, in order to improve the operability of a system. This paper presents human anomaly detection based on human posture recognition by means of two-stage classification for the solution of above-mentioned issues. Proposed method first classifies a person appeared in video into two classes, one corresponds to caregiver and another corresponds to aged people to be taken care of, and then detects anomaly motion only for the latters.

Keywords: Anomaly Detection, Classification, Health Care/Nursing, Human Motion

1 Introduction

A point of view for observing the maturity of country can be seen in the age structure of citizen: it is often changing from young generation centric to elderly generation centric. In recent years, birth rate is declining trend in some countries especially in Asia or Europe. Japan is also one of such countries, and it revealed that less younger generation of people need to sustain increasing cost of social welfare than before. In another word, the capacity of social infrastructure for elderly people such as elderly care facilities is considered to be less than it is demanded. It results insufficient number of caregiver in elderly care facilities, or the family of younger generation needs to take care of their parents of old age by themselves, balancing their work and care.

In these circumstances it is not easy to watch the elderly people who need assistance by accompanying caregiver or physically unimpaired family all the time. Detecting anomaly status of elderly person, while no person is available for assistance, is one of the urgent issues to be resolved. Applying deep neural network (DNN) for object detection or identification in image/video made a breakthrough to achieve remarkable improvement in accuracy and flexibility for various cases in real occasions. Anomaly detection is one of the objectives in surveillance

[*] Japan Advanced Institute of Science and Technology, Tokyo, Japan
[†] Japan Advanced Institute of Science and Technology, Ishikawa, Japan

video, and DNN based method is also applied in this domain as well [1]. In general, anomaly detection with DNN needs training process which requires a large number of training data prior to detection, and collecting a number of anomaly data for training is not easy in many cases. Not only in the above-mentioned example of anomaly detection but also in other studies on anomaly detection, main focuses are in achieving the improvement of accuracy or wider coverage of detection for various types of anomaly.

In general, anomaly detection of human motion is applied for every person appeared in video. In case where anomaly detection is targeted for specific type of person, there is another direction for improving usability or operability of a system in addition to pursuing higher precision, recall, or accuracy. Here, assume that a system is employed, which detects anomaly motion of "falling down" in a nursing and personal care facility. Even a caregiver may fall down occasionally, but most of the cases it does not results serious injury and he/she can recover by him/herself. In this scenario, detecting such anomaly motion is not considered to be the target of the operation of the system, and therefore such the anomaly motion may be ignored. As shown in this example, another direction for the improvement of operability of anomaly detection system is to classify the target person whether he/she is elderly person to be taken care of or caregiver prior to anomaly detection, then anomaly detection is applied only for the motion which is classified into "elderly person being taken care of". The class of a person whether he/she is caregiver or care-receiver is mentioned as attribute of a person hereafter.

Based on above mentioned consideration, we propose a method of anomaly detection of human motion integrated with human attribute classification, which is assumed to be operated to detect anomaly motion (e.g., falling down) of elderly person in hospitals or nursing and personal care facilities. The objective of the method is to detect anomaly motion of elderly person excluding to detect that of nursing staff or caregiver. There are possible solutions for this objective, i.e., supervised learning-based method, unsupervised learning-based, and so on. Our proposed method is unsupervised learning based, considering to adapt various circumstances or environments without preparation of large amount of data for training.

Proposed method consists of human posture inference, feature representation and time series data pooling, and 2-stage clustering. After human posture is inferred, a sequence of features that corresponds to human motion is represented based on distances between joints. Then 1st stage of clustering is applied in order to classify the motion by caregiver (nursing staff) or care-receiver (elderly person). After that, anomaly motion is detected by applying 2nd stage of clustering for the cluster that corresponds to the motion of elderly person, i.e., care-receiver.

The organization of this paper is as follows: Section 2 describes related work on anomaly detection. Proposed method that consists of 2-stage clustering is described in section 3. Experimental setup and the result of experiment is shown in section 4. Concluding remarks and remaining issues are mentioned in section 5.

2 Related Work

Anomaly detection is one of the major topic in computer vision since the last decades, where various recent frameworks are employed for pursuing higher detection accuracy or adapting the variety of visual features. Features such as HOG (Histogram of Gradient), HOF (Histogram of Optical Flow), MBH (Motion Boundary Histogram) are often applied for human action/motion recognition not only with conventional analytic methods but with DNN [2]. As well as other domains of applications, DNN based methods are recent main stream for action/motion recognition. 3D-CNN is the extension of CNN so that it processes the sequence of images in time domain, and anomaly detection with the 3D-CNN was also proposed [3]. Another promising method for handling sequential data is to employ LSTM which handles not only short term temporal relationship but also long term for anomaly detection [5]. GNN (Graph Neural Network) is employed for representing features found in image sequence as nodes, and relationship between nodes as links. Detecting anomaly based on GNN is also proposed in [4]. In recent years, anomaly detection of human motion employs a human posture inference model to acquire positions of joints in a human body as motion features, and human motion is represented with GNN [6]. In [6], it is reported that anomaly detection with GNN outperforms straightforward method of feeding video data directly to CNN based model. Note that above mentioned method for anomaly motion detection of human concentrates on improving the performance (e.g., accuracy) of detection.

Focusing on tasks of human attribute detection, the definition of attribute differs in each study, but an attribute is associated with an entity or a feature which is not commonly observed in all the persons appeared in video. An example of attribute is an object which a person wearing, holding or carrying [7]. This kind of attribute is referred to for characterizing a person. Human attribute is referred to as visual feature of human body parts, which is used for human re-identification [8]. Human re-identification is a task to identify a person appeared in different video streams (thus taken by different cameras placed at different positions), where the common visual features are observable in human body parts. In addition to the type of clothing or objects he/she is holding, Gender and/or age are also the scope of human attribute recognition. Deep neural network based model is also popular for the solution of this task [10] as well as for object detection or recognition. Not only image features but motion information such as gait energy is also referred to for recognizing human attribute of gender or age [11].

However, as far as we know, there is no study which focus on improving the practicability or operability of anomaly detection of human motion by excluding the anomaly motion of persons who are out of the target since they are not the persons to be watched (i.e., caregivers in the scenario of anomaly detection in nursing and personal care facility). Handling this matter is achieved by human classification by attribute prior to anomaly motion detection.

3 Anomaly Motion Detection

3.1 Assumption and Scenario

We assume that the proposing method is employed in hospital or nursing and personal health care facility in order to detect anomaly motion of elderly person who is taken care of. Here, we assume that nursing staff or caregiver is in younger-generation than patient/care-receiver and physically unimpaired. Footpace between nursing staff/caregiver and care-receiver of elderly person is considered to be different because of the difference of physical motor function. We consider this assumption is reasonable in most of the cases observed in hospitals or nursing and personal health care facilities. The objective of this study is to detect anomaly motion of “falling down” while a patient/care-receiver is walking down the hallway. The reason why we focus on the above-mentioned occasion is that there are much less alarm equipment such as emergency-call buttons or vital sensors while patient or care-receiver is away from his/her room and walking down hallway. Therefore, anomaly motion during he/she is lying or sitting on a bed is out of the scope of this study.

3.2 Processing Flow of Anomaly Detection and Pre-processing

Processing flow of the proposed method is illustrated in Fig. 1. Proposed method is based on clustering of motion feature which is defined based on the distances between specific keypoints of human body. First, key points of human body are inferred with human posture inference model from every frame in a video stream at 30fps, which is fed from video camera placed at ceiling in a corridor in the actual operation. In this study, we employed MediaPipe [12] for this component, which infers and results the position of keypoints on the human body. Though a number of keypoints are available as the result of MediaPipe, five keypoints, i.e. nose, right and left wrist, right and left ankle are referred to for calculating the distance between keypoints. Distances between nose and right/left wrist/ankle are calculated for motion feature representation. Figure 2 shows an example showing the transition of distances between keypoints. Horizontal axis represents frame number of a video stream (i.e., time, 30 frames correspond to 1 second in this study), and the vertical axis denotes Euclidean distance between keypoints. Physical configuration of the environment is as follows: a camera is placed at upper corner in a corridor, and the optical axis of the camera is nearly parallel to the direction of walking. Fluctuation of distance will be smaller as shown in Fig. 2(a) where he/she walks away from the camera, since the size of a person appeared will be gradually smaller. In case where he/she is walking toward the camera, the fluctuation is observed as shown in Fig 2(b). After calculating keypoint distances, they are normalized by dividing the distances between the keypoints of nose and hip in order to reduce the influence of the change of size appeared in a video frame, since the size of person will be larger when he/she walks toward camera, and vise versa.

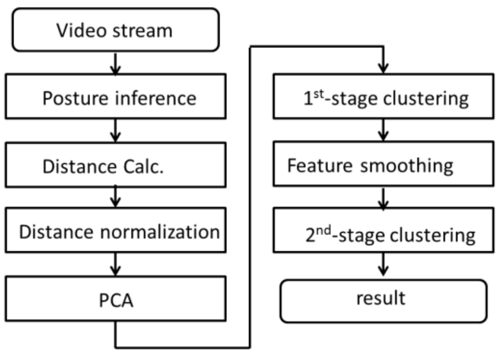
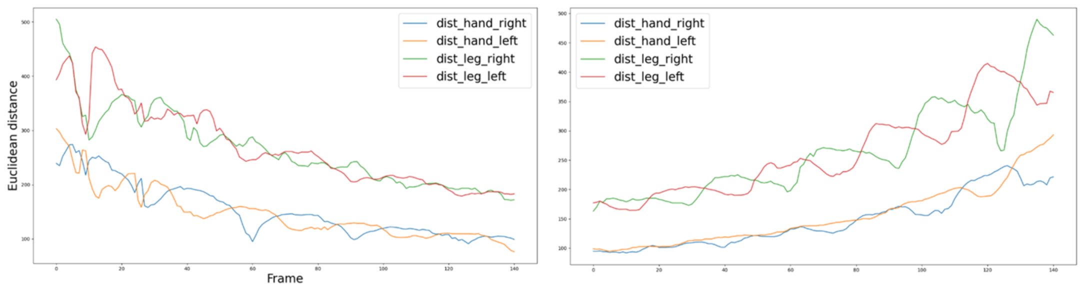


Figure 1: Processing Flow

After that, PCA (Principal Component Analysis) is applied for the four series of temporal transition of the normalized distance between keypoints in order to obtain primary feature of human motion, which is to be represented as one dimensional data. Until obtaining one dimensional data of human motion by PCA is pre-processing step for anomaly detection. When a person is detected as resulting inferred posture by the posture inference module, a sequence of data resulted through the process of distance calculation, distance normalization, and PCA, is stored into the pool of motion data. The sequence of motion data is terminated when tracking of a person (i.e., detection of human posture of the same person) is lost by frame out or at the predefined amount of interval.



(a) Walking away from camera (b) Walking toward camera

Figure 2: An example of transition of distances between keypoints

3.3 Anomaly Detection based on 2-stage clustering and feature smoothing

After pre-processing, 1st-stage clustering is carried out in order to separate sequences of human motion of nursing staff/care giver from those of care-receiver of elderly person. This clustering process is expected to separate motion data of caregiver from that of care-receiver, that is based on the prospective where the feature of footpace of the former differs from the latter. Based on preliminary evaluation, we employed time series K-means clustering by Euclidean distance and that by DTW (Dynamic Time Warping) with regard to the performance of separability on human attribute and anomaly. The preliminary evaluation also revealed that separability of anomaly from normal motion is degraded for a set of motion sequences that consists of normal walking and walking accompanied with falling down by the same person. Therefore, the process of feature smoothing is introduced prior to 2nd stage clustering. The feature smoothing is to calculate simple moving average of the one dimensional time-series feature obtained by PCA, by following the equation below.

$$A_t = \frac{1}{k+1}(x_{t-k} + x_{t-(k+1)} + \cdots + x_t + \cdots x_{t+k})$$

In this equation, A_t denotes moving average at time t, k corresponds to the number of samples to refer, and x_t is the feature value at time t obtained by applying PCA. Strength of smoothness is adjusted by k: short term rapid change is suppressed when k is increased.

After the process of feature smoothing, 2nd stage clustering is carried out for detecting anomaly motion. The 2nd stage clustering is applied for one of the clusters, which corresponds to the cluster of motion data of care-receiver (i.e. elderly persons) obtained by the 1st stage clustering (but each time-series data is smoothed prior to 2nd stage). The reason why we take this strategy is as follows. The 1st stage clustering results two clusters: one is expected to consist of motion data of caregiver and another is expected to consist of care-receiver. One of these cluster, whose dominant fluctuation of motion feature is smaller, is considered to be the cluster that corresponds to the motion feature of care receiver (i.e., elderly persons). Since the variance of the distance between keypoints is considered to be larger in caregiver at fixed amount of time, we think this criterion is reasonable. Therefore, 2nd stage clustering for the cluster of care-receiver is expected to separate motion data of normal walking from walking accompanied with anomaly motion of falling down, and thus anomaly motion is expected to be detected. Finally, one of the clusters whose member is less than another is identified to the cluster that corresponds to anomaly motion. Moreover, if 2nd stage clustering is applied for the cluster of motion by care-giver, it is also expected to separate normal walking from anomaly motion. However, this is out of the objective of this study since a caregiver is able to recover oneself or calling others for help.

4 Experimental Results

4.1 Test Scenario and Data

In order to evaluate the proposed method, test data is prepared as follows. A camera is placed around the upper corner above a corridor, whose position, direction and angle is similar to that of security camera in a building. Three persons, who consisted of two males and one female and they are in their 20s to 50s, are instructed to walk toward or away from the camera (optical axis of the camera and his/her direction of walking is nearly parallel). One sequence of video data consists of a scene where one of the persons is walking normally (i.e., walking at average speed), walking slowly as if he/she is elderly person, or falling down after walking slowly. In total, 56 normal walking data, 22 normal slow walking data, and 3 motion of falling down accompanied with slow walking are used for the experiment. In the actual operation of the proposed method, motion feature data of a person is segmented at fixed amount of interval (e.g., in every five seconds), and stored into a stream data pool for clustering.

In the experiment, 7 cases of stream data mixture are prepared. More specifically, each of them is prepared as follows. Here, NW stands for normal walking, SW for slow walking, and FD for slow walking accompanied with falling down as anomaly motion.

Case 1: 56 NW, 22 SW, SW+FD by person 1

Case 2: 56 NW, 22 SW, SW+FD by person 2

Case 3: 56 NW, 22 SW, SW+FD by person 3

Case 4: 56 NW, 22 SW, SW+FD by person 1 and 2

Case 5: 56 NW, 22 SW, SW+FD by person 1 and 3

Case 6: 56 NW, 22 SW, SW+FD by person 2 and 3

Case 7: 56 NW, 22 SW, SW+FD by person 1, 2, 3

Note that one anomaly motion is captured for each of the person 1, 2, and 3. That is, case 1 to 3 includes one anomaly motion data, case 4 to 6 includes two anomaly motion data, and Case 7 includes three anomaly motion data.

Figure 3 shows sequences of motion feature after applying PCA. Fig. 3(a) shows motion feature of normal walking toward a camera, while Fig. 3(b) shows that of slow walking. We can observe that the fluctuation of 1 dimensional motion feature is larger in normal walking (Fig. 3(a)) than that in slow walking (Fig. 3(b)). Moreover, steep slope as seen in Fig. 3(a) is not observed in Fig. 3(b), which is considered to be due to the lower motion velocity than normal walking. Figure 4 shows motion feature of three cases of anomaly motion, each of which consists of falling down accompanied with slow walking. These sequences of motion feature in Fig. 4 correspond to anomaly motions to be detected, since it corresponds to the situation where a care-receiver of elderly person accidentally falls down. Fluctuation of motion feature observed in Figure 4 is more drastic than those observed in Fig. 3(b), which correspond to slow walking without falling down. Note that the preliminary experiment of applying time series K-means by

Euclidean distance where K=3, expecting three clusters consisting of normal walking, slow walking, and falling down accompanied by slow walking, revealed that it was not able to classify anomaly motion into a cluster without the other two kinds of instances.

(a) normal walking (b) slow walking

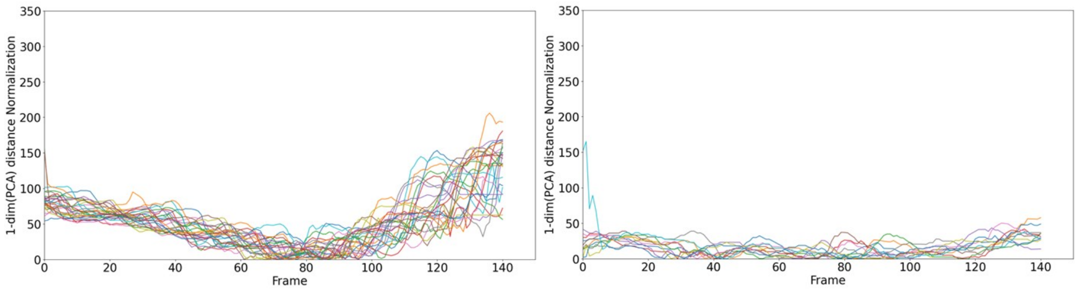


Figure 3: Motion Feature after applying PCA

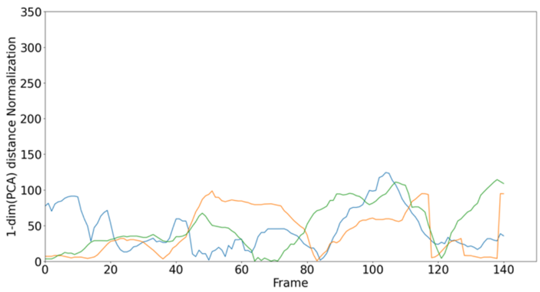
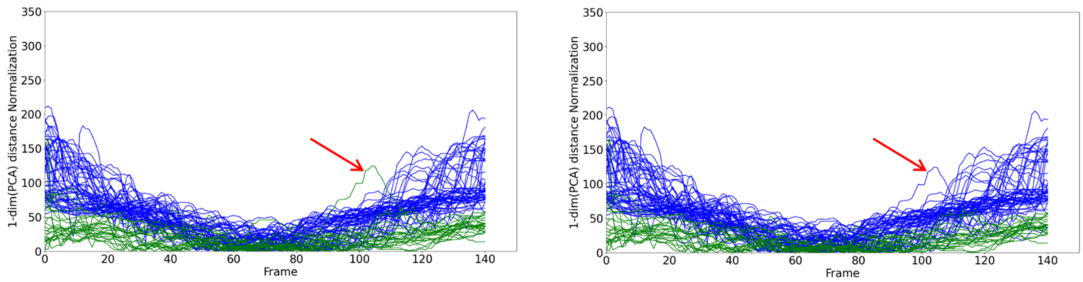


Figure 4: Motion Feature of slow walking with falling down

4.2 Result of 1st stage clustering

The 1st stage clustering, aiming at separating one cluster consisting of normal walking from another consisting of slow walking as well as falling down accompanied with slow walking, was evaluated by applying time series K-means by the metric of Euclidean distance and DTW. We tested 7 cases of mixture of normal and anomaly motion. As a result, both K-means by Euclidean distance and DTW showed good separation of normal and slow walking. More precisely, clustering by DTW failed to make an anomaly motion accompanied with slow walking be included in the cluster consisting of normal walking in Case 1, while K-means by Euclidean distance correctly clustered all the seven cases into the cluster consisting of only normal walking and another consisting of slow walking and slow walking accompanied with falling down. Figure 5(a) shows the result of 1st stage clustering by Euclidean distance and Fig. 5(b) is that by DTW in the Case 1, where the blue curves correspond to normal walking whereas green curves correspond to slow walking except one indicated with red arrow. Only the difference is the curve indicated with red arrow, which corresponds to falling down accompanied with slow walking, and therefore it should be classified into the cluster of curves colored with green.

Identifying the cluster of normal walking and slow walking is performed based on the variance of 1 dimensional feature data. Since larger variance corresponds to rapid motion, the cluster whose variance is smaller is identified to be the cluster of slow walking: it corresponds to motion data of elderly person. Thus, the 2nd stage of clustering is applied only for this cluster.

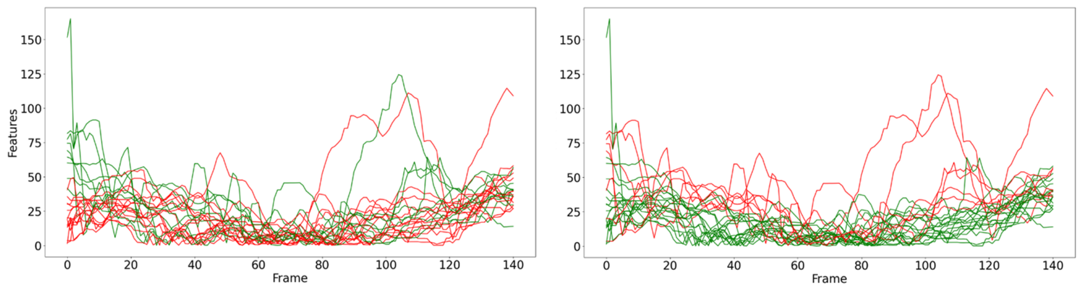


(a) Clusters by Euclidean distance (b) Clusters by DTW

Figure 5: Result of 1st stage clustering by K-means by Euclidean distance and DTW

4.3 Effect of feature smoothing and the result of 2nd stage clustering

After 1st stage clustering, each of the sequences of motion feature data is smoothed in order to emphasize the difference between slow walking and falling down accompanied with slow walking. In order to show the effect of feature smoothing, result of applying K-means clustering by Euclidean distance and DTW is shown in Figure 6. Figure 6(a) shows the result of applying K-means by Euclidean distance and Fig. 6(b) show that of DTW for Case 5 that includes two anomaly motions as an example. Color of curve corresponds to cluster index: curves of the same color belong to the same cluster. As shown, both of these failed to achieve clear separation of slow walking with falling down from those without falling down: one of the clusters consists of the mixture of slow walking and slow walking with falling down, therefore, clustering without feature smoothing failed to detect anomaly.



(a) Clusters by Euclidean distance (b) Clusters by DTW

Figure 6: Referential result of 2nd stage clustering without feature smoothing

Figure 7 shows the effect of feature smoothing that is obtained by applying moving average. Each of the sequences of motion features without falling down are more flattened than those without applying moving average, whereas those with anomaly of falling down shows steep slopes (curves with arrow).

The 2nd stage clustering is evaluated by K-means with Euclidean distance and DTW for sequences of data after feature smoothing. K-means by Euclidean distance is carried out for all the seven cases prior to 2nd stage clustering, which resulted correct separation of slow walking with/without falling down from normal walking for all the cases. In case of applying

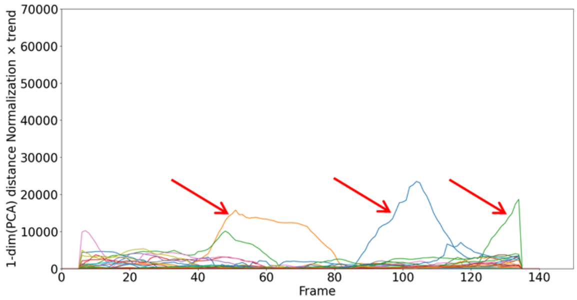


Figure 7: Motion feature after feature smoothing

K-means by Euclidean distance for 2nd stage clustering, successful classification separating slow walking with falling down (i.e., anomaly motion) from slow walking succeeded only in Case 1 and 2. That is, some slow walking without falling down and slow walking with falling down are clustered into the same cluster: thus it failed detecting anomaly motion without error. As for the result of applying K-means by DTW, only the anomaly motion(s) is separated into a cluster without any slow walking for all the cases of Case 1 through Case 7.

In summary, K-means clustering by Euclidean distance is suitable for classification of human motion, which is considered to be associated with the attribute of nursing staff/caregiver or care-receiver as the 1st stage clustering. After that, it is reasonable to apply K-means by DTW for detecting anomaly motion as the 2nd stage clustering. In addition, we confirmed that the feature smoothing by moving average prior to the 2nd stage clustering contributes to emphasize the difference between slow walking and falling down (i.e., anomaly motion) accompanied with slow walking, and thus achieves better detection of anomaly motion.

5 Conclusion

In this paper, we proposed an unsupervised method of anomaly detection aiming at detecting anomaly motion of falling down caused by elderly person. Detection of this type of anomaly is important in hospitals or nursing and personal care facilities especially while nursing staff/caregiver is temporarily unable to watch with care. The proposed method consists of 2 stage clustering with motion feature smoothing, where the 1st stage clustering is employed for classifying human attribute by motion feature, and the 2nd stage clustering is for detecting anomaly. Experimental result showed that K-means clustering by Euclidean distance is suitable for 1st stage, and K-means by DTW for the 2nd stage, with the intermediate process of feature smoothing by moving average. Exploring the applicability of the proposed method for various facilities is one of the future issues to be investigated.

Acknowledgement

Authors would like to acknowledge G. Matsukawa for his valuable suggestions and comments.

References

[1] Anala M. R, Malika Makker, Aakanksha Ashok, Ambika Sana, "Anomaly Detection in Surveillance Videos", 2019 26th International Conference on High Performance Computing, Data and Analytics Workshop, 2019, pp. 93-98.

[2] Limin Wang, Yu Qiao, Xiaoou Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors", IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4305-4314.

[3] Amin Ullah, Khan Muhammad, Killichbek Haydarov, "One-Shot Learning for Surveillance Anomaly Recognition using Siamese 3D CNN", 2020 International Joint Conference on Neural Networks, 2020, 8 pp.

[4] J. Lin, Q. Cai, M. Lin, "Multi-label classification of fundus images with graph convolutional network and self-supervised learning", IEEE Signal Process, Vol.28, 2021, pp.454-458.

[5] Devashree R. Patrikar, Mayur Rajaram Parate, "Anomaly Detection by Predicting Future Frames using Convolutional LSTM in Video Surveillance", 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing, 2023, 6 pp.

[6] Xianlin Zeng, Yalong Jiang, Wenrui Ding, Honggua, "A Hierarchical Spatio-Temporal Graph Convolutional Neural Network for Anomaly Detection in Videos", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30, No. 1, 2023, pp. 200-212.

[7] Jian Jia, Houjing Huang, Xiaotang Chen, Kaiqi Huang, "Rethinking of Pedestrian Attribute Recognition: A Reliable Evaluation under Zero-Shot Pedestrian Identity Setting", arXiv:2107.03576v2, 2021, 13pp.

[8] H. Huang, W. Yang, J. Lin, G. Huang, J. Xu. G. Wang, X. Chen, K. Huang, "Improve person re-identification with part awareness learning", IEEE Transactions on Image Processing, Vol. 29, 2020, pp. 7468–7481.

[9] Mingzhe Chen, Xi. Xiao, Bin Zhang, Xinyu Liu, Runiu Lu, "Neural Architecture Searching for Facial Attributes-based Depression Recognition", Proc. International Conference on Pattern Recognition, 2022, pp. 877-884.

[10] Dangwei Li, Xiaotang Chen, Zhang Zhang, Kaiqi Huang, "Pose guided deep model for pedestrian attribute recognition in surveillance scenarios", IEEE International Conference on Multimedia and Expo., 2018, 6pp.

[11] S. Zhang, Y. Wang, "Gait Energy Image-Based Human Attribute Recognition Using Two-Branch Deep Convolutional Neural Network", IEEE Transactions on Biometrics, Behavior, and Identity Science, Vol 5., No. 1, 2023, pp. 53-63.

[12] Camillo Lugaresi, et al., "MediaPipe: A Framework for Building Perception Pipelines", arXiv:1906.08172v1, 2019, 9pp.