

The study of skeleton description reduction in the human fall-detection task

O. S. Seredin¹, A. V. Kopylov¹, E. E. Surkov¹

¹Tula State University, 300012, Tula, Russia, Lenin Ave. 92

Abstract

Accurate and reliable real-time fall detection is a key aspect of any intelligent elderly people care system. A lot of modern RGB-D cameras can provide a skeleton description of a human figure as a compact pose presentation. This makes it possible to use this description for further analysis without access to real video and, thus, to increase the privacy of the whole system. The skeleton description reduction based on the anthropometrical characteristics of a human body is proposed. The experimental study on the TST Fall Detection dataset v2 by the Leave-One-Person-Out method shows that the proposed skeleton description reduction technique provides better recognition quality and increases the overall performance of a Fall-Detection System.

Keywords: fall detection, human activity detection, skeleton description, RGB-D camera, elderly people care system.

Citation: Seredin OS, Kopylov AV, Surkov EE. The study of skeleton description reduction in the human fall-detection task. *Computer Optics* 2020; 44(6): 951-958. DOI: 10.18287/2412-6179-CO-753.

Acknowledgments: The work is supported by the Russian Fund for Basic Research, grants 18-07-00942, 18-07-01087, 20-07-00441. The results of the research project are published with the financial support of Tula State University within the framework of the scientific project 2019-21NIR. The part of the research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

Introduction

According to the World Health Organization [1], every year there are near 37.3 million falls with quite serious consequences, after which medical care is required. Also, about 650 thousand fatal falls occur annually, which makes such falls the second most important cause of death after traffic accident injuries. The type and severity of injuries sustained as a result of a fall can depend on age, gender, or person condition of health. Age is one of the main risk factors for falls [1]. The highest risk of death or serious injuries as a result of fall threatens elderly people and this risk increases with age. Besides, people after the fall, especially elderly people who did not get medical help in time are exposed to a significant risk of the need for follow-up care and placing them in a special health facility. In this regard, recently there has been a growing interest in on-line fall detection systems based on round-the-clock monitoring data.

The “fall” is defined as an event when a person unintentionally finds oneself lying on the floor or on the ground. Falls are a serious problem for social healthcare worldwide.

The fall detection system besides accuracy, specificity and reliability must have two important properties for everyday usage: unobtrusiveness and confidentiality. According to research [2] elderly people are more likely to adopt in-home surveillance technologies if they are private and unobtrusive, namely, if they do not bring discomfort in everyday life, do not require to wear and maintain any device and to attain new technical skills and, that is extremely important, they do not capture any video images.

The privacy aspect is especially important for the realization of such a system since it can reduce the anxiety of people under surveillance. Monitoring systems with high-resolution cameras demonstrate good enough recognition accuracy and allow to avoid the use of wearable devices. However, it is difficult to cover all areas of the room with a cameras field of view. There can be “blind spots” – places that will not be available for the camera, for example, due to large-sized furniture that can obstacle monitoring object. In addition, capturing and processing of high-resolution video does not allow to fully preserve confidentiality of received information. Alternatively, for example, infrared sensors can be used [3]. Nevertheless, some research indicates that the presentation of a human figure in the form of silhouettes [4] or skeletons [5] allows to preserve confidentiality and therefore can decrease people's concern regarding personal privacy in surveillance system based on image and video processing. Description of a human pose is based on the representation of the figure by a set of segments corresponding to the limbs and torso joined together at the points approximately corresponding to the joints. The spatial coordinates of these points, together with the straight lines that connect them, form a graph, which is usually called a skeleton description of human pose. Such skeleton model provides an ability to build depersonalized feature description for human activity monitoring.

Nowadays the market has gotten depth sensors that allow to obtain a skeleton description in real-time, such as Microsoft Kinect, Intel RealSense, Orbbec Astra Pro. Comparative characteristics of cameras are presented in the table 1.

Table 1. Comparison of RGB-D cameras with the ability to skeleton build

Device name	Microsoft Kinect v1	Microsoft Kinect v2	Intel RealSense 435i	Orbbec Astra Pro
RGB Sensor Resolution and Frame Rate	640×480 @30 fps	1920×1080 @30 fps	1920×1080 @30 fps	1280×720 @30 fps
Depth Sensor Resolution and Frame Rate	320×240 @30 fps	512×424 @30 fps	1280×720 @90 fps	640×480 @30 fps
Maximal operating distance	4.5 m	4.5 m	10 m	8 m
Horizontal Field of View	57 deg	70 deg	87 deg	60 deg
Vertical Field of View	43 deg	60 deg	58 deg	49.5 deg
Number of skeleton points	17	25	19	19

These technologies allow to obtain skeleton description and exclude other identifying features as well as source image itself. Devices perform analysis of image and build a skeleton model of a person without external computing facilities, therefore photo and video data are not sent to remote servers for processing.

Furthermore, there are several methods to build a skeleton with additional software on the depth map. Recently, there have been ways to build skeletons using 2D images from a conventional RGB camera, for example, PoseNet technology [10].

In the literature on fall detection there are three main groups of human figure representation based on skeleton description using RGB-D sensors.

The first group of methods uses General skeleton geometric characteristics, such as bounding rectangle, geometric moments and their invariants, position and distances from specific skeleton point, for example, the distance from the point, corresponding to the head or mass center, to the floor, etc. These methods are less sensitive to the skeleton estimation defects but do not have enough flexibility to operate well in the complex or changing environment. For example, the method [3] uses an RGB-D camera fixed on tripod and skeleton geometric characteristics are used for fall detection. This is based on speed measurement of width, height and depth changing of bounding rectangle and makes it possible to detect fall in real-time with sufficient accuracy and exclude false triggering during normal people activity (for example, lying on the bed or the floor). This method does not require knowledge about the scene, such as the equation of the floor plane. The algorithm [11] uses an RGB-D camera located under the ceiling and directed downward. This location allows to practically avoid blocking a person with objects in the room. The fall is determined by skeleton point coordinates according to the simple threshold rule as 0.4 m relative to the floor distance. This algorithm uses only 3D information from the camera, that makes it invariant to ambient light, and it also increases the confidence of the system, because it is impossible to see the face by the 3D camera from such position. The disadvantages of this method are limited field of view with this camera location and difficulty of detection such daily activity like sitting and lying.

Method [12] also uses a bounding rectangle together with the first derivative in height and the first derivative in the width-depth composition. But these parameters are

subject to noise because of the low accuracy of the sensor. The method involves the use of the Kalman filter for a more accurate estimate of the rate of change of height and the composition of width and depth. The next idea is used for excluding false triggering during usual action: y-coordinate of the upper left corner of tracking bounding rectangle, because it is close to y-coordinate of the head center. Fall is identified when this coordinate is below the required threshold.

Method [6] uses distances from skeleton to floor. This method requires the presence of the floor in the monitoring zone. Fall is detected, when skeleton points (head, shoulder center, the center of pelvic, left and right ankle) located below the required threshold. Nine types of geometric features are presented in [13]. A concatenation of all of them is used as pose and motion representations. Work [14] also utilizes that set of relational geometric features.

The second group of methods utilizes the correspondence between the skeleton and human body parts. Human body representation like hinge system of rigid segments connected by points (joints) and person movement considered as a continuous change of these segments locations. Therefore, if it is possible to reliably extract and track the human skeleton, then movement recognition can be performed by identification of real-time rebuilding a skeletal model method [8]. Other methods take into account that the human body moves in accordance with the shapes, lengths and locations of the bones, which are more obvious and stable to observe [15]. Work [16] considers the relations of neighboring parts of a human body (two arms, two legs and torso).

The third group is based on the position of skeleton vertices in three-dimensional space, approximately correspond to the joint locations. The pairwise relative position of skeleton vertices [7] or skeletal covariance matrices [9] are often used to pose description. However, the relative positions themselves are not enough for exact fall detection and additional space-time characteristics could be applied [17].

In the work [5] the matrix of pairwise Euclidean distances between skeleton vertexes, the speed of changing of these distances between neighboring frames and inter-frame speed of changing of these characteristics are suggested to use as a data feature description in a fall detection system. In addition, the height of some selected points is used. Distances matrix is built on the skeleton

points provided by the RGB-D sensor Microsoft Kinect V2. The transition from explicit 3D coordinates of skeleton points to generalized representation in the form of a pairwise distances matrix at least makes it difficult for malicious attempts to pose restoration of the human being monitored.

In this work, we propose the method for reducing skeleton features description, which is used in [5], based on anthropometric features of the human body what allows to exclude twenty-four redundant features. Experimental research, conducted on the TST Fall Detection v2 dataset [18], showed that the proposed method of reducing features description doesn't cause deterioration of quality and allows to improve performance by decreasing overfitting and at the same time to raise the speed of calculation. Results were received by the procedure of quality assessment based on the Leave-One-Person-Out method [19].

Principles of the fall detection system development

Using human figure skeleton description received by the RGB-D sensor allows to build a fall detection system without wearable sensors, what makes the system more comfortable for elderly people, and also without direct usage of video and photo information during analysis that reduces the risk of unauthorized distribution of confidential information. The existence of different RGB-D cameras allows to choose a device that is least affected by lighting conditions and provides a more accurate skeletal description of the pose and therefore makes the task of human figure detection and segmentation easier.

In the work [5] a system of fall detection is proposed. The general architecture of the system is shown in fig. 1.

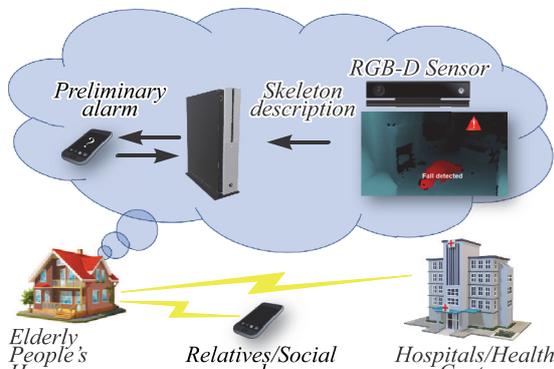


Fig. 1. The general architecture of the proposed system

The system is a software and hardware complex that includes an RGB-D sensor and client-server application. The system provides a device connected to a cellular network as an operational alert. It should be noted that in the case of a registered fall an alarm sent firstly to the person being monitored (to a mobile phone or special device) for confirmation and if he or she is able to react to the message, then the subsequent distribution of the warning message should be stopped for preventing false alarm. Otherwise, relatives and/or social worker will receive an emergency notification. Several levels of private access to event information are offered. At the lowest level, only

information contained directly in the message itself is available. On the second level a real RGB-video of a person fall is replaced with a reliable animated avatar based on skeleton description. The third level is intended for a detailed analysis of the situation using the actual fall detection video record [5].

As a result of statement [7], that pairwise relative positions of the joints provide more distinguishing features than 3D coordinates of skeleton points, the Euclidian matrix of the distance between joints, normalized by the height of the monitored figure, is used as feature description of a human pose.

In addition, the feature description is expanded by the dynamics of the pose in terms of the characteristics of speed and acceleration between frames. The data matrix is built based on the skeleton, which is obtained from points, provided by the Microsoft Kinect v2 RGB-D sensor.

In the work [5] points, that correspond to the fingers of the hands and feet coordinates in space, were excluded because these elements of skeleton are rather flexible and they do not carry useful information. Eventually, only 17 from 25 points of skeleton provided Microsoft Kinect v2 are considered, as shown in fig. 2.

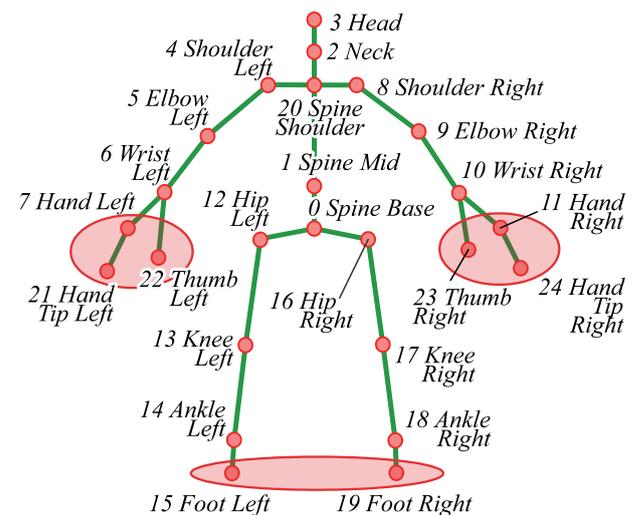


Fig. 2. Skeleton description of human, provided by Microsoft Kinect v2, and ovals indicate points that are excluded from the skeletal description

Thus, in the work [5] a total of 459 features reflecting the distance, the rate of change of distance and the change in the rate of change of distance on each frame starting from the second are considered. Then Cumulative SUM procedure [20] is applied for combining single solutions on consecutive frames.

Because Microsoft Kinect v2 does not always provide a stable skeleton representation it is necessary to detect frames with incorrectly constructed description and exclude them. It is difficult to mark such a large number of frames with mistakes manually, therefore one-class SVM classifier [5] is applied to find and exclude these frames (outliers) from the training set based on skeleton feature representation. At the test stage, the output of the two-class

classifier for such frames was replaced by a zero value, which means that the object lies directly on the separating hyperplane. The final decision is based on adjacent frames using the Cumulative SUM procedure [5].

The following results were obtained on the entire TST Fall Detection dataset v2: number of outliers in frames containing falls – 803, number of outliers in frames not containing falls – 391. Examples of frames with and without outliers are given in fig. 3.

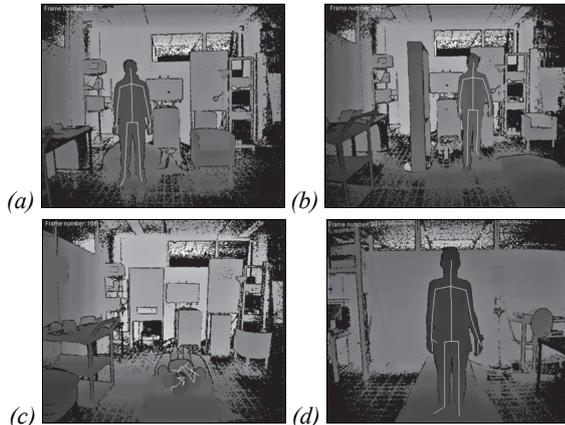


Fig. 3. Examples of frames with outliers (c) and without outliers (a, b, d)

After excluding outliers from the training set two-class classifier was trained to separate classes of fall and normal human activity. In the method [5] accuracy 0.917 was attained. The quality measurement of classification

was based on the Leave-One-Person-Out method with excluding records of the particular person from the database entirely.

Reduction of the skeleton based feature description

In this work we propose to reduce the skeleton description by taking into account human anthropometric features. Keep in mind that the recognition quality depends on careful selection of features that maximally eliminate redundant information that can lead to the use of complex models and over-fitting. One can come to a simpler description of the skeleton by reducing the size of the source distances matrix. It is proposed to exclude the skeleton points distances, which do not change during any movement of a human, that makes such information about human dynamics and the change of these distances redundant. From this point of view, distances between joints: shoulder and elbow, elbow and wrist, hip and knee, knee and ankle should be considered as redundant information. Also, the data of speed and the acceleration of changes in these distances should be excluded. The Diagram of the variance of distances across the frames of the entire data set is shown in fig. 4. Distances that were excluded are highlighted in red. Quite a high variance of redundant data relative to the other distances shows that these distances introduce distorted information to the pose description. In accordance with the nature of distance changes in video sequences, six features with low dispersion were found which are marked in blue in fig. 4.

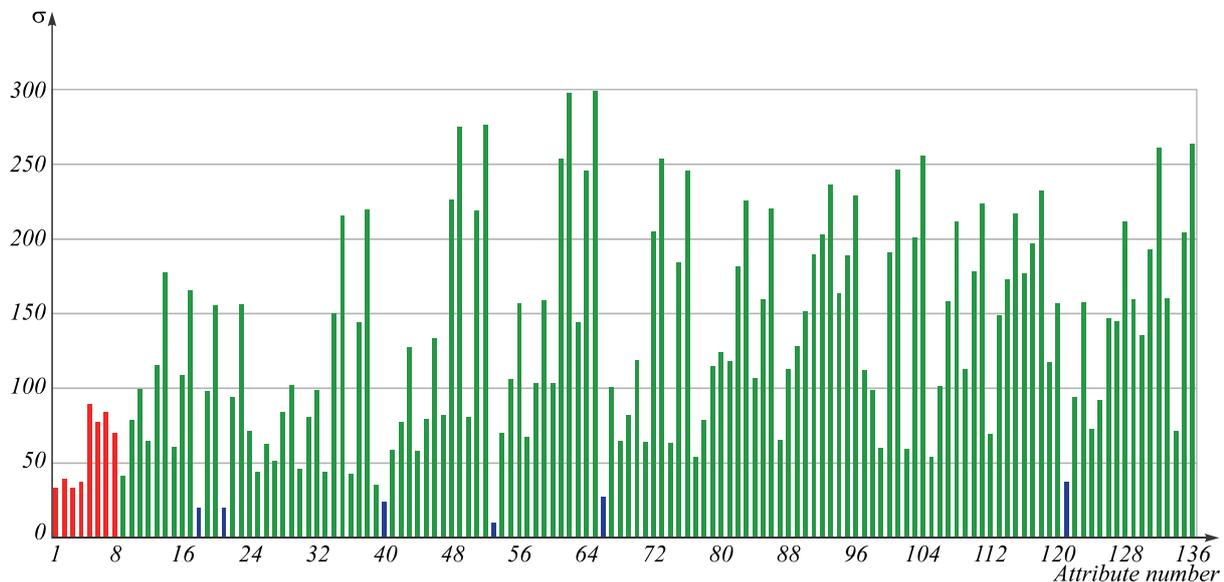


Fig. 4. Diagram of the average square deviation of distances between points of the skeleton. Red color indicates distances excluded based on anthropometric data, and blue color indicates distances that are proposed to be excluded based on the experimental data

These features also correspond to anthropometric characteristics. This is the distance from base of spine to left and right hip joints, the distance between these joints themselves, the length of the cervical spine and head size (fig. 5). Such distances are also excluded from the feature description.

After reduction, just 122 features remain in the spatial feature description of the object. In fig. 5, all excluded from feature description distances are marked.

Besides, heights of 17 points are used, thus the number of pose features equal to 139 instead of the original 153 before excluding distances [5]. As in [5], human ac-

tivity dynamics is considered as information of skeleton description on several neighbors' frames. Differences in posture features on neighbor frames give an additional 139 features (interframe speed). Finally, 139 attributes are added to reflect the change of interframe speed characteristics between neighbor frames (acceleration). Thus, the pose description has 417 features instead of the original 459. The general flowchart of the proposed system is shown in fig. 6.

Experimental results

In the literature, relatively few databases that contain the Microsoft Kinect v2 data of the human activity, including falls, are described. A fairly broad survey can be found in the work [14]. In the work [5], the TST Fall Detection v2 database was used for the experimental study of the fall detection algorithm [18]. It contains depth frames and skeleton description points in the space data which are collected by Microsoft Kinect v2 and presented as records of various durations. A dataset consists of records that reflect normal activity and fall records modeled by 11 actors. The dataset contains activity of daily living (ADL) in the following categories: sit, bend down and grasp something, walk, lie down and action related to

falls category (FALL): fall forward, fall back, fall sideways, fall back, and end up in a sitting position.

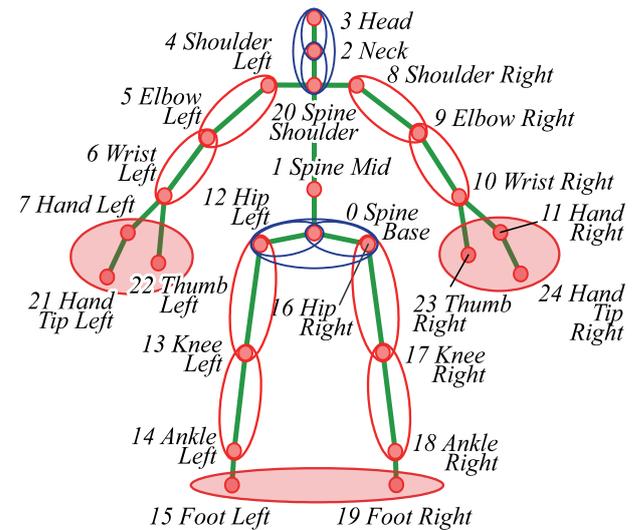


Fig. 5. Reduced skeleton description of human, red-shaded ovals indicate excluded points in [5] and red ovals – distances that are excluded based on anthropometrical features and blue ovals – distances that are excluded based on analysis of variance

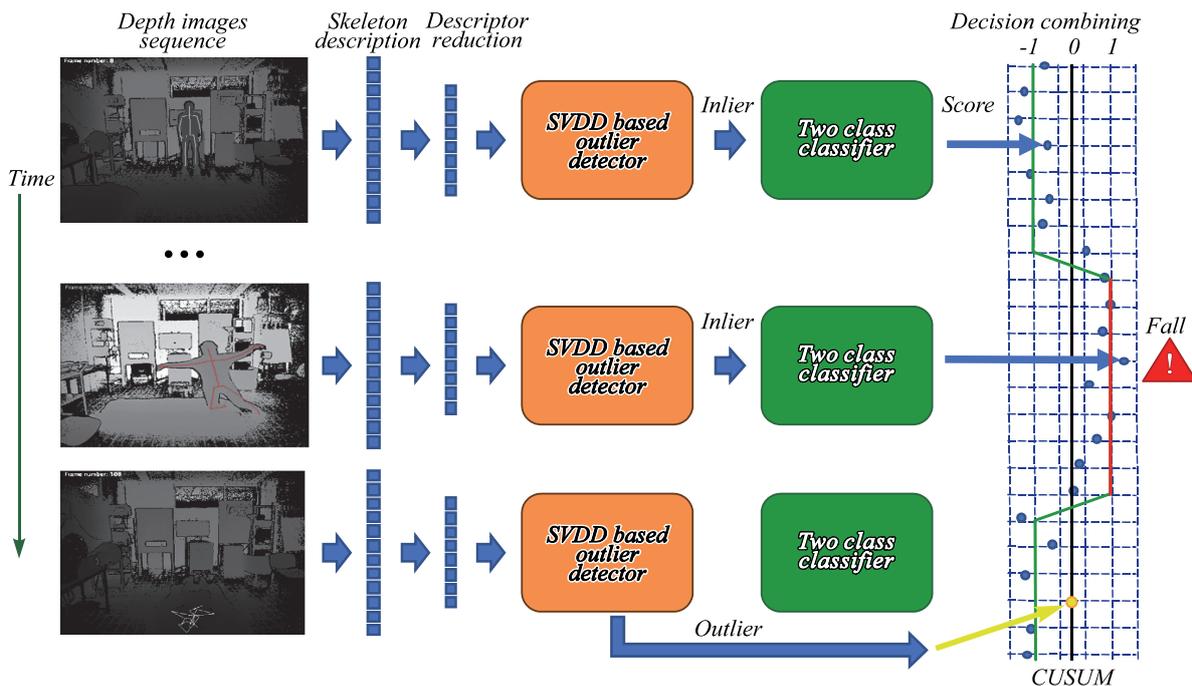


Fig. 6. The general flowchart of the proposed fall detection system

The total number of records on the TST Fall Detection v2 dataset is 264 and the general number of frames – 46 418. The frame rate of records is 30 frames per second. The shortest record lasts 2.5 s (75 frames), and the longest one lasts 15.4 (463 frames).

This set is one of the most recent datasets which have quite a large number of videos with various content, so right this dataset was selected for experiments.

All records in the TST Fall Detection v2 database are marked as containing and non-containing frames with

falls. However, in addition to frames with a fall, there are frames in the records, which are marked as fall-containing, that show the normal activity of people. For example, frames on which the actor walks or lies. Obviously, such frames do not belong to the fall. Therefore, every frame in the database must be marked as the FALL category frame or as the ADL category frame.

In previous work [5] authors made the frames markup for all records in the fall category. The number of frames related to falls is 8 306. The shortest, average and longest

time of falling is 0.9 s (27 frames), 2 s (62 frames) and 3.2 s (97 frames) respectively. In fig. 7 the positions of fall category fragment in the record are shown. Dark grey color corresponds to the frame with the fall and light grey color – without fall. Since the question of the exact definition of the beginning and the end is quite disputable, the decision to use the consistent solution of several experts was accepted.

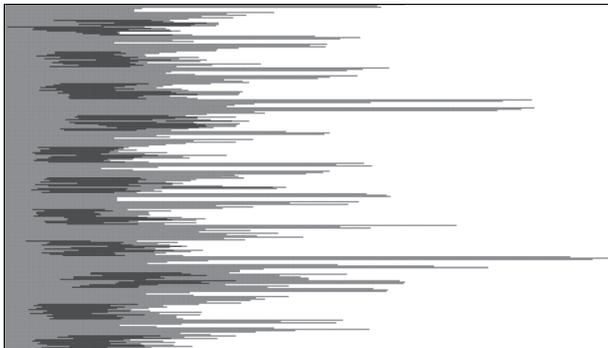


Fig. 7. Representation of markup of database records in the form of segmented lines, where dark grey marks falls, each line corresponds to the single video record

The advantage of such markup is the ability to more accurately evaluate the quality of the fall detection algorithm in certain time positions of the fall beginning and

end. Taking into account the proposed method of reduction of the skeleton description the preliminary data matrix contains 45809 objects of two classes – FALL (8306) and ADL (37 589) in the space of 417 features.

For quality estimation of classifier in a reduced feature space the modified version of the cross-validation procedure was used. To simulate the actual conditions of applying the classifier to an unknown scene with a new person, for each experiment information about a particular person is completely removed from the database, and the classifier is trained only on the remaining ten persons.

Recognition quality (the ratio of correctly recognized records to their total number) is evaluated for the person which was previously deleted from the database.

This procedure is repeated 11 times and eventually the average for all results is found. Table 2 represents the evaluated quality, accuracy and delay in frames for each actor.

A complete study with the consecutive exclusion of each actor shows 0.936 classification accuracy for all records. Accuracy of coincidence in position and duration of fall segments estimated by the proposed method and marking by experts is 0.876. The average delay of starting position obtained by the classifier on test records is 9.52 frames. The accuracy of the developed algorithm in comparison with other state-of-the-art algorithms is shown in table 3.

Table 2. Recognition indicators for eleven actors

Actor Parameter	1	2	3	4	5	6	7	8	9	10	11
Classification accuracy	0.958	0.917	0.914	1	0.833	0.917	0.958	0.958	0.917	0.958	0.958
Accuracy of coincidences in position and duration	0.891	0.894	0.838	0.871	0.856	0.868	0.844	0.876	0.861	0.854	0.901
Average delay of fall start position, in frames	4.916	9.583	11.330	2.917	10.500	10.330	22.250	4.250	3.750	20.417	4.500

Table 3. The accuracy of fall detection algorithms on TST Fall Detection Dataset v2

Method	Source data	Classifier	Evaluation Scheme	Accuracy
Gasparrini et al., 2016 [21]	Skeleton joint position; accelerometer data	Empirical thresholding rule	Not described	0.990
Fakhrulddin et al., 2018 [22]	Two accelerometers time-series data	CNN	Separate data on 90% and 10%, then averaging	0.923
Hwang et al., 2017 [23]	Depth map	3D-CNN	5 random trials from 240 and 24 records splitting and averaging	0.942
Min et al., 2018 [24]	Skeleton joints information	SVM	Not described	0.920
Seredin et al., 2019 [5]	Skeleton joints information	SVM + One-class classifier + CUSUM	Leave-One-Person-Out	0.917
Our	Reduction of skeleton joints information	SVM + One-class classifier + CUSUM	Leave-One-Person-Out	0.936

Conclusion

We propose a method of reduction of the skeleton feature description, used in [5]. The method allows to exclude 42 redundant features based on the anthropometric characteristics of a human body.

Research on the TST Fall Detection v2 database shows that a proposed method of features description reduction does not lead to a deterioration in the falls recognition quality, but on the contrary, it allows to improve the characteristics of the human activity monitoring system by reducing overfitting.

As a result, the accuracy of fall detection recognition was increased from 0.917 to 0.936 and the accuracy of the coincidence of fall moment on the test dataset was changed from 0.879 to 0.877. Delay of starting position was increased by 3 frames, that is not significant losing. General calculation speed with new recognition models was increased by 15–20%. These results were received by a quality assessment procedure based on the method of consistently excluding one person from the training set.

References

- [1] Falls. World Health Organization. Source: <https://www.who.int/en/news-room/fact-sheets/detail/falls>.
- [2] Wild K, et al. Unobtrusive in-home monitoring of cognitive and physical health: Reactions and perceptions of older adults. *J Appl Gerontol* 2008; 27: 181-200.
- [3] Mastorakis G, Makris D. Fall detection system using Kinect's infrared sensor. *J Real-Time Image Process* 2012; 9(4): 635-646.
- [4] Demiris G, et al. Older adults' privacy considerations for vision based recognition methods of eldercare applications. *Technol Heal Care* 2009; 17(1): 41-48.
- [5] Seredin OS, Kopylov AV, Huang S-C, Rodionov DS. A skeleton features-based fall detection using Microsoft Kinect v2 with one class-classifier outlier removal. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2019; 42(2:W12): 189-195.
- [6] Mundher Z, Zhong J. A real-time fall detection system in elderly care using mobile robot and Kinect sensor. *Int J Mater Mech Manuf* 2014; 2(2): 133-138.
- [7] Wang J, et al. Mining actionlet ensemble for action recognition with depth cameras. *Proc IEEE Comp Soc Conf Comp Vis Pattern Recognit* 2012: 1290-1297.
- [8] Vemulapalli R, Arrate F, Chellappa R. Human action recognition by representing 3D skeletons as points in a lie group. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2014: 588-595.
- [9] Hussein ME, et al. Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations. *IJCAI Int Jt Conf Artif Intell* 2013: 2466-2472.
- [10] Papandreou G, et al. Towards accurate multi-person pose estimation in the wild. *Proc 30th IEEE Conf Comput Vis Pattern Recognit (CVPR)* 2017; 2017: 3711-3719.
- [11] Pathak D, Bhosale VK. Fall detection for elderly people in homes using Kinect Sensor. *Int J Innov Res Comput Commun Eng* 2017; 5(2): 1468-1474.
- [12] Bevilacqua V, et al. Fall detection in indoor environment with Kinect sensor. *2014 IEEE Int Symp Innov Intell Syst Appl Proc* 2014: 319-324.
- [13] Chen C, et al. Learning a 3D human pose distance metric from geometric pose descriptor. *IEEE Trans Vis Comput Graph* 2011; 17(11): 1676-1689.
- [14] Zhang S, Liu X, Xiao J. On geometric features for skeleton-based action recognition using multilayer LSTM networks. *Proc 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* 2017: 148-157.
- [15] Zhang X, Xu C, Tao D. Graph edge convolutional neural networks for skeleton based action recognition. 2018. P. 1–22.
- [16] Du Y, Wang W, Wang L. Hierarchical recurrent neural network for skeleton based action recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2015; 07-12-June: 1110-1118.
- [17] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. *arXiv Preprint* 2018. Source: <https://arxiv.org/abs/1801.07455>.
- [18] TST Fall detection dataset v2. IEEE DataPort. Source: <https://ieee-dataport.org/documents/tst-fall-detection-dataset-v2>.
- [19] Sung J, et al. Unstructured human activity detection from RGBD images. *Proc IEEE Int Conf on Robotics and Automation* 2012: 842-849.
- [20] Page ES. Continuous inspection schemes. *Biometrika* 1954; 41(1/2): 100-115.
- [21] Gasparrini S, Cippitelli E, Gambi E, Spinsante S, Wählén J, Orhan I, Lindh T. Proposal and experimental evaluation of fall detection solution based on wearable and depth data fusion. In Book: Loshkovska SSK, ed. *ICT Innovations 2015: Advances in intelligent systems and computing*. Cham: Springer; 2016: 99-108.
- [22] Fakhruddin AH, Fei X, Li H. Convolutional neural networks (CNN) based human fall detection on Body Sensor Networks (BSN) sensor data. *Proc 2017 4th Int Conf Syst Informatics (ICSAI)* 2018: 1461-1465.
- [23] Hwang S, Ahn D, Park H, Park T. Maximizing accuracy of fall detection and alert systems based on 3D convolutional neural network. *Proc Second Int Conf Internet-of-Things Des Implement (IoTDI'17)* 2017: 343-344.
- [24] Min W, Yao L, Lin Z, Liu L. Support vector machine approach to fall recognition based on simplified expression of human skeleton action and fast detection of start key frame using torso angle. *IET Comput Vis* 2018; 12(8): 1133-1140.

Authors' information

Oleg Sergeevich Seredin received the Ph.D. Degree in Theoretical Foundations of Informatics from Computing Center of the Russian Academy of Sciences, Moscow, Ph.D. Thesis: "Methods and Algorithms of Featureless Pattern Recognition" (2001). Now he is Associate Professor at the Institute of Applied Mathematics and Computer Science at Tula State University. His scientific interests are data mining, pattern recognition, machine learning, signal and image analysis, bioinformatics, visual representation of multidimensional data, statistical methods of decision making. He is a member of program committee at several conferences (CloudCom, PSBB, PRIB, AIST, GraphiCon, VISAPP). Prof. Seredin is principal investigator of several grants of the Russian Fund for Basic Research, including international. He worked as visiting scientist at Rutgers University and National Taipei University of Technology. He has published more

than 100 scientific papers in refereed journals, handbooks, and conference proceedings in the areas of machine learning, pattern recognition and computer vision. Prof. Seredin is a member of The International Association for Pattern Recognition (IAPR). E-mail: oseredin@yandex.ru.

Andrei Valerievich Kopylov received the Ph.D. degree from the Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia, in 1997. In 1997, he joined the Department of Automation and Remote Control, Tula State University, as an Assistant Professor and became an Associate Professor in 2005. Currently he is an Associate Professor with the Institute of Applied Mathematics and Computer Science, Tula State University. He worked as visiting researcher at the Dorodnicyn Computing Centre of Russian Academy of Sciences and National Taipei University of Technology. His scientific interests are signal and image analysis, data mining, machine learning. Prof. Kopylov was the principal investigator of several grants of the Russian Fund for Basic Research, including international. He is a member of program committee at several conferences (CloudCom, PSBB, SoICT, AIST, GraphiCon, VISAPP), reviewer of scientific journals Sensing and Imaging (SSTA), Computer Optics, Machine Learning and Data Analysis (JMLDA). He has published more than 70 scientific papers in refereed journals, handbooks, and conference proceedings in the areas of machine learning, pattern recognition and computer vision. Prof. Kopylov is a member of The International Association for Pattern Recognition (IAPR). E-mail: and.kopylov@gmail.com.

Egor Eduardovich Surkov, born in 1997. Since 2017, a student of the Tula State University, majoring in Information System and Technology. His research interests are machine learning, data mining, pattern recognition. He participated in several scientific conferences and has publication in the book of abstracts of the 18th Conference «Mathematical Methods for Pattern Recognition». E-mail: eg-su@mail.ru.

GRNTI: 28.23.15.

Received May 17, 2020. The final version – July 17, 2020.
