12th scientific and technical conference
“Intellectual processing of information - 2018” (ИОИ-2018)

08–12 October 2018, Italy, Gaeta

Microsoft Kinect v2 for fall detection

Rodionov D.S., **Kopylov A.V.,**
Seredin O.S.

den191097@gmail.com, andkopylov@gmail.com,
oseredin@yandex.ru

Tula State University
Laboratory of Data Analysis

Human Fall Detection Based on Data From Depth Cameras

Simple cameras, environmental sensors, on-body sensors and depth cameras can be used in clients/patients behavior remote monitoring.

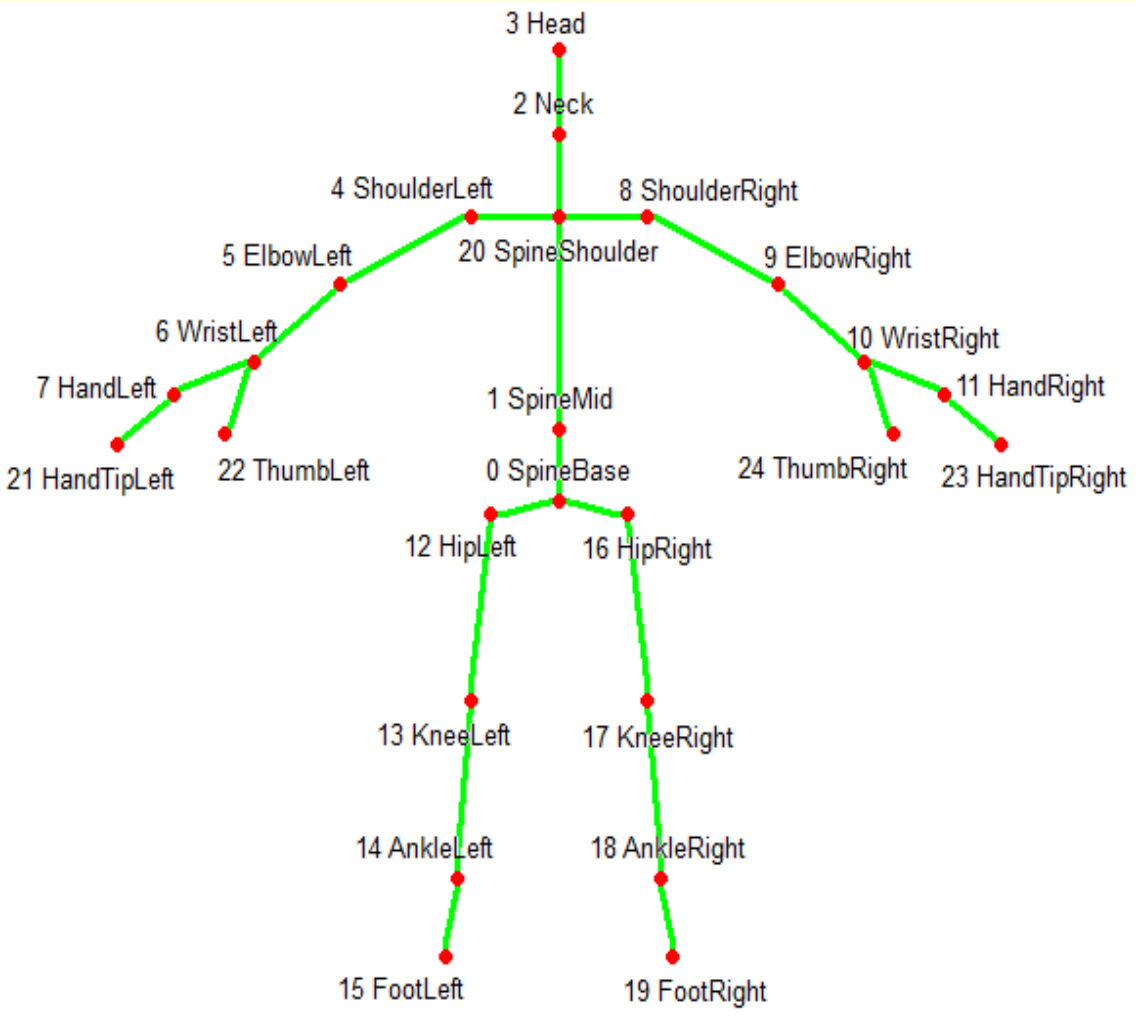
Currently Available Depth Cameras

RGB-D cameras: a - Microsoft Kinect v1, b - Microsoft Kinect v2, c - Asus Xtion PRO LIVE, d - Intel RealSense D435

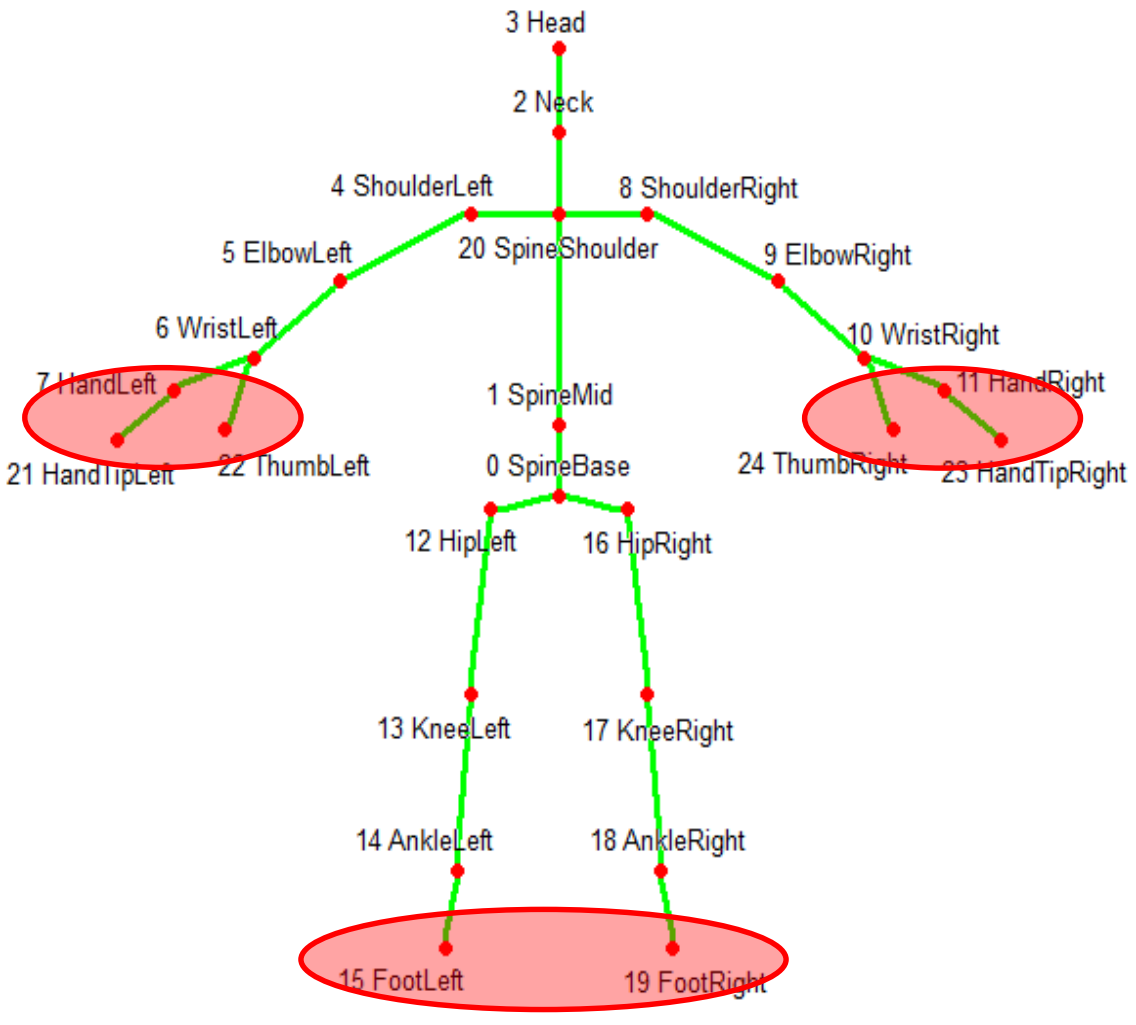
Parameters of Depth Cameras Available on Market

	Microsoft Kinect v1	Microsoft Kinect v2	Asus Xtion PRO LIVE	Intel RealSence D435
RGB Sensor Resolution and Frame Rate	640x480 @30 fps	1920x1080 @30 fps	1280x1024	1920x1080 @30 fps
Depth Sensor Resolution and Frame Rate	320x240 30 fps	512x424 30 fps	640x480 30 fps	up to 1280x720 up to 90 fps
Maximum distance of use	4.5 m	4.5 m	3.5 m	up to 10 m
Horizontal Field of View	57	70	58	86
Vertical Field of View	43	60	45	57
Skeletal tracking	Yes	Yes	NO	NO

Skeleton Points Used in Microsoft Kinect 2

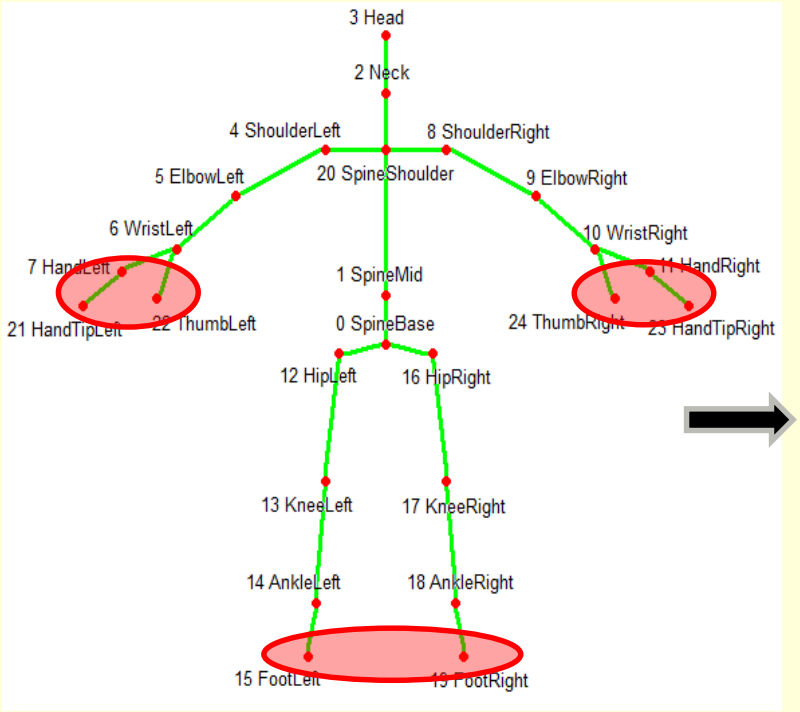


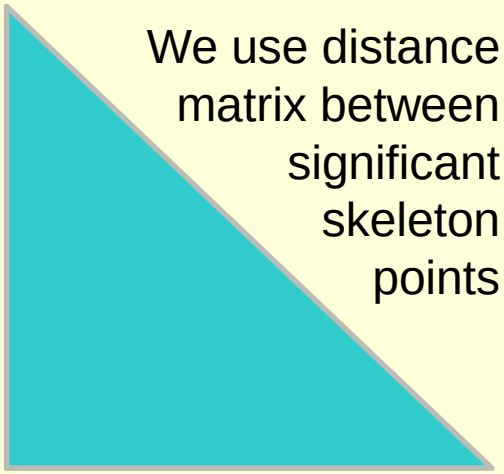
Skeleton Points Used in Microsoft Kinect 2



Not all points are used!

Feature Based Description of the Figure on the First Frame





We use distance matrix between significant skeleton points

Heights of all points are also used

136

17

Feature Based Description of the Figure on the First Frame

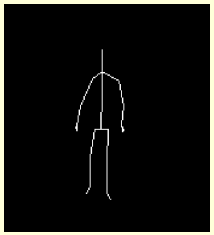
Feature Based Description of the Figure on the Second Frame

Additionally computing speed
regarding first frame

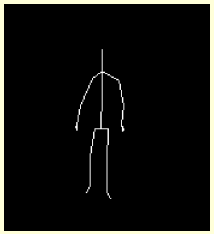
Feature Based Description of the Figure on the Third Frame

Additionally computing speed
regarding second frame

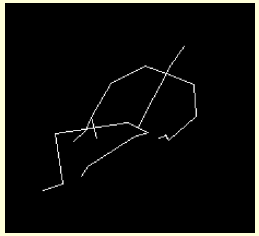
Feature Based Description of the Figure in a Record Frames



136 17 136 17 136 17



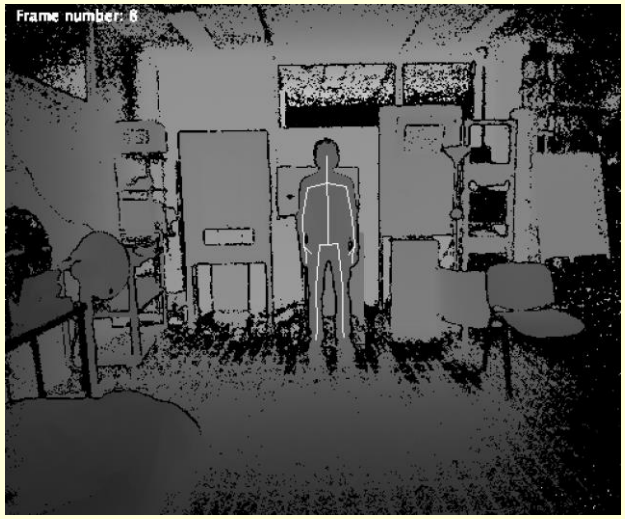
136 17 136 17 136 17

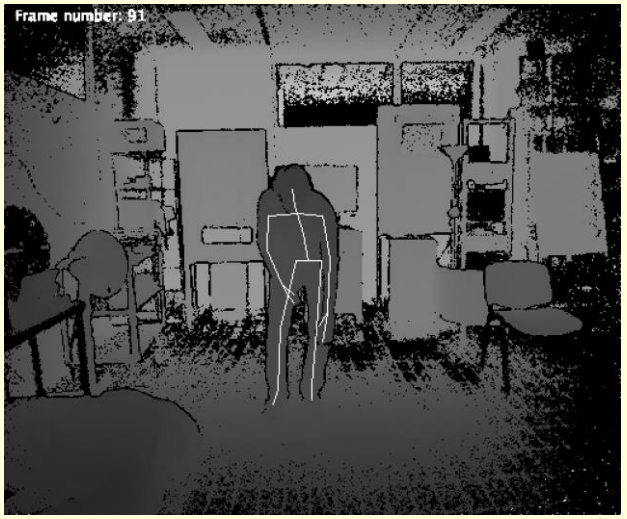


136 17 136 17 136 17

One Class Classifier Outlier Removal

One class classifier is used to determine frames with skeleton tracking errors.





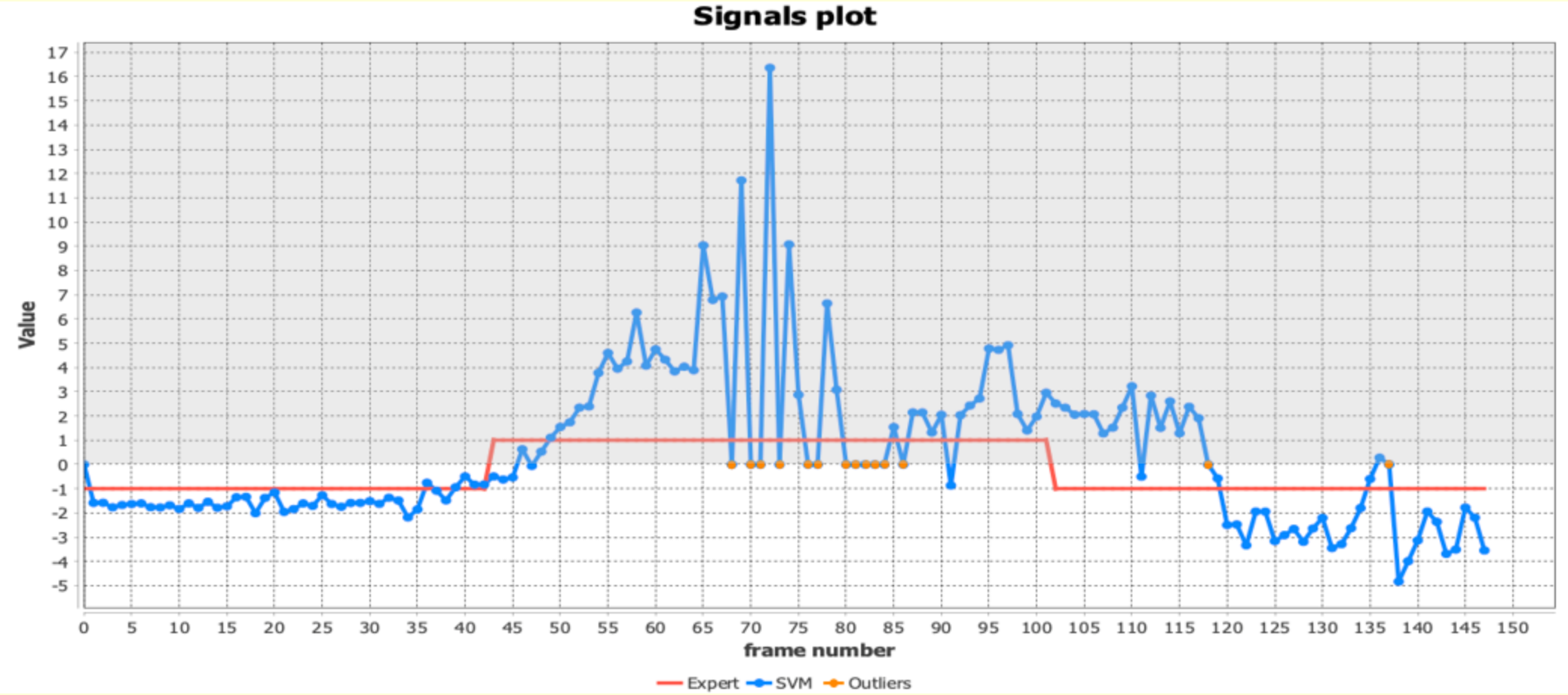
Good frames

Bad frame (outlier)

Outliers do not participate in two-class training procedure. Statistic for one class classifier trained on all dataset applied for records only with falls: number of outliers in Fall class – 803, number of outliers in ADL class – 151. For all records number of outliers in ADL class – 391.

B. Schölkopf, J. C. Platt, J. Shawe-Taylor, a J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution.," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

Two-Class SVM Fall Detection



Combining the Individual Decisions on the Consecutive Frames Based on Cumulative Sum Procedure

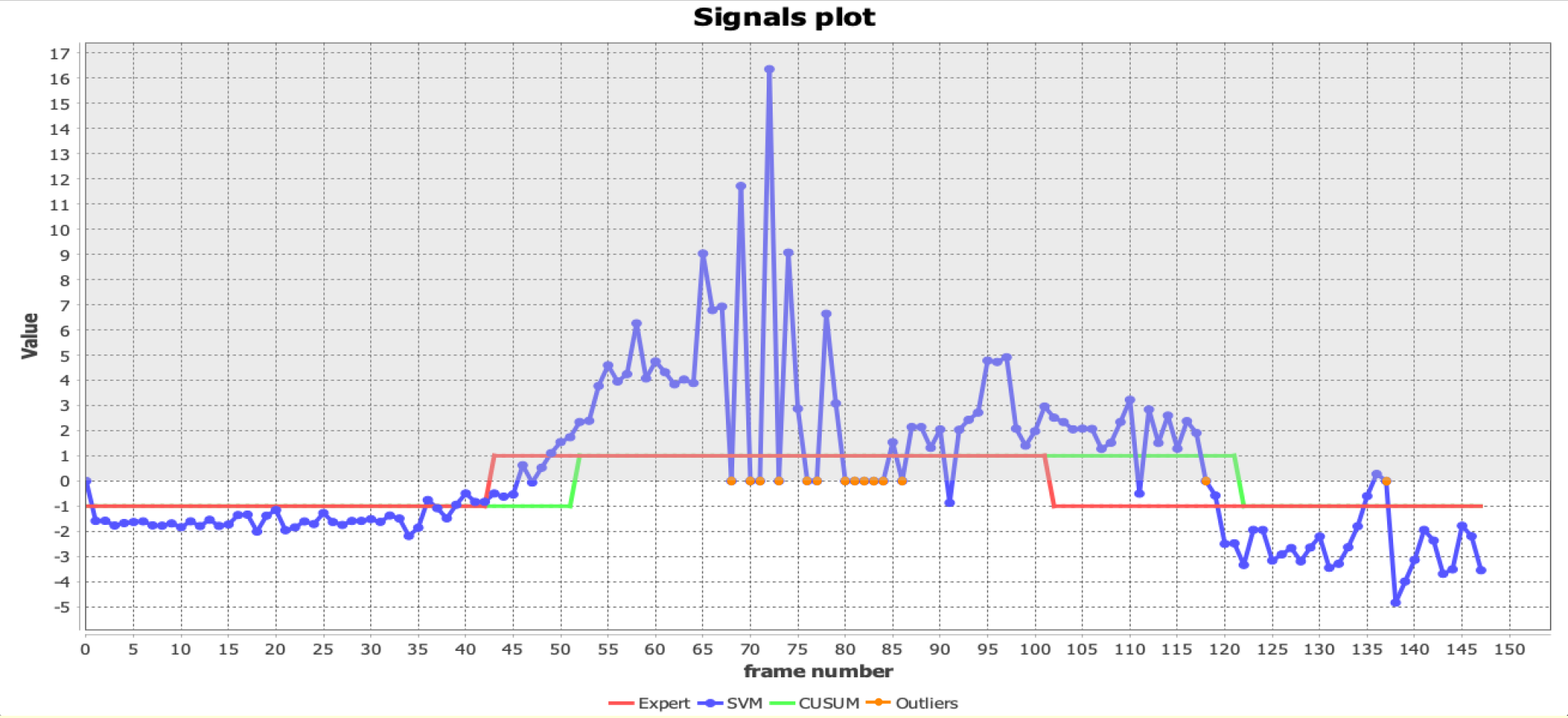
The nature of the actions like fall leads to the relatively long frame sequences with a constant stage.

In combination with online character of the fall-detection system, this allows to apply a sequential analysis technique developed by E. S. Page.

$$L(t) = \begin{cases} b_1, L(t-1) + d(t) > b_1, \\ L(t-1) + d(t),\ b_2 < L(t-1) + d(t) < b_1, \\ b_2,\ L(t-1) + d(t) < b_2. \end{cases}$$

E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

Example of Combining the Individual Decisions

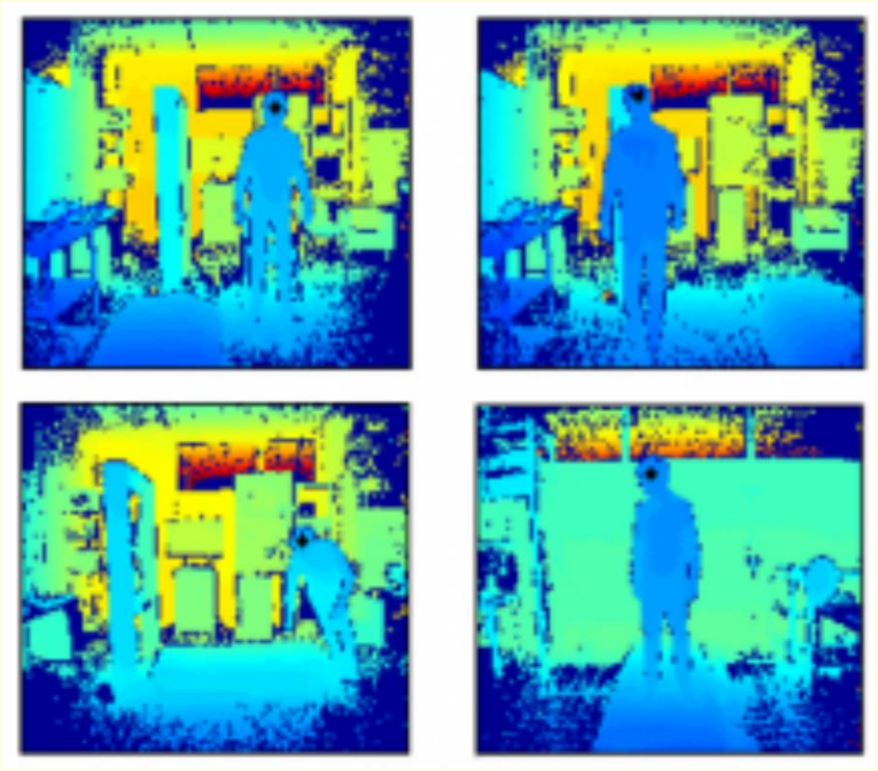


Fall Detection Procedure Steps

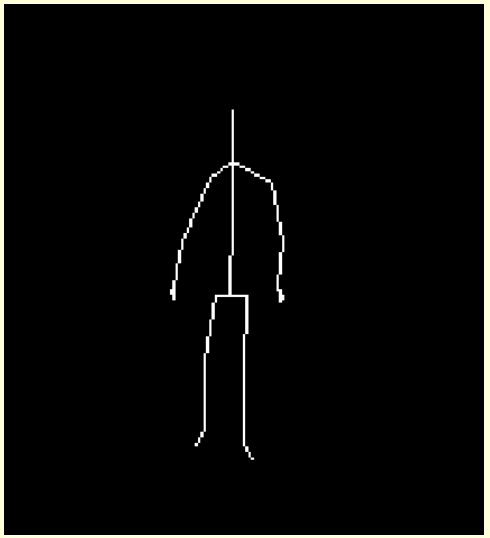
Our fall detection procedure consists of the following steps:

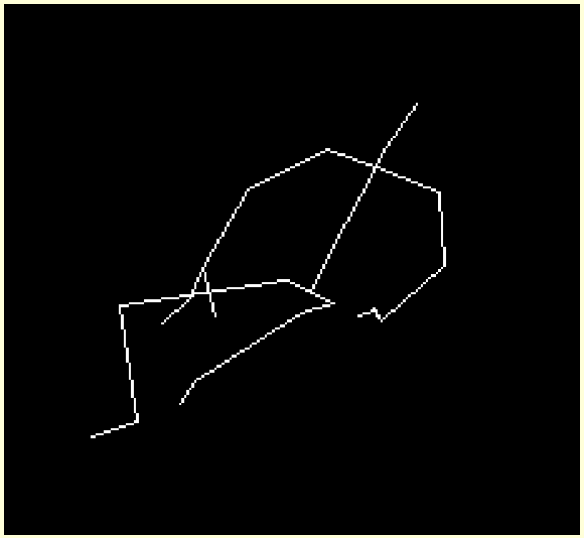
1. Cumulative sum initialization.
2. For each frame in a sequence perform the following operations:
    1. acquiring skeleton descriptions by Microsoft Kinect v2;
    2. calculation of EDM and dynamic features;
    3. outlier removal based on one-class classifier;
    4. calculation of the distance from a hyperplane, obtained by pretrained two-class SVM classifier.
    5. Updating the cumulative sum and threshold touch checking.
    6. Fall or ADL detection.

TST Fall Detection Dataset v2



The dataset contains depth frames and skeleton joints collected using Microsoft Kinect v2 and acceleration samples provided by an IMU during the simulation of ADLs and falls.





IEEE DataPort TST Fall Detection Dataset v2 URL:
https://ieee-dataport.org/documents/tst-fall-detection-dataset-v2

TST Fall Detection Dataset v2

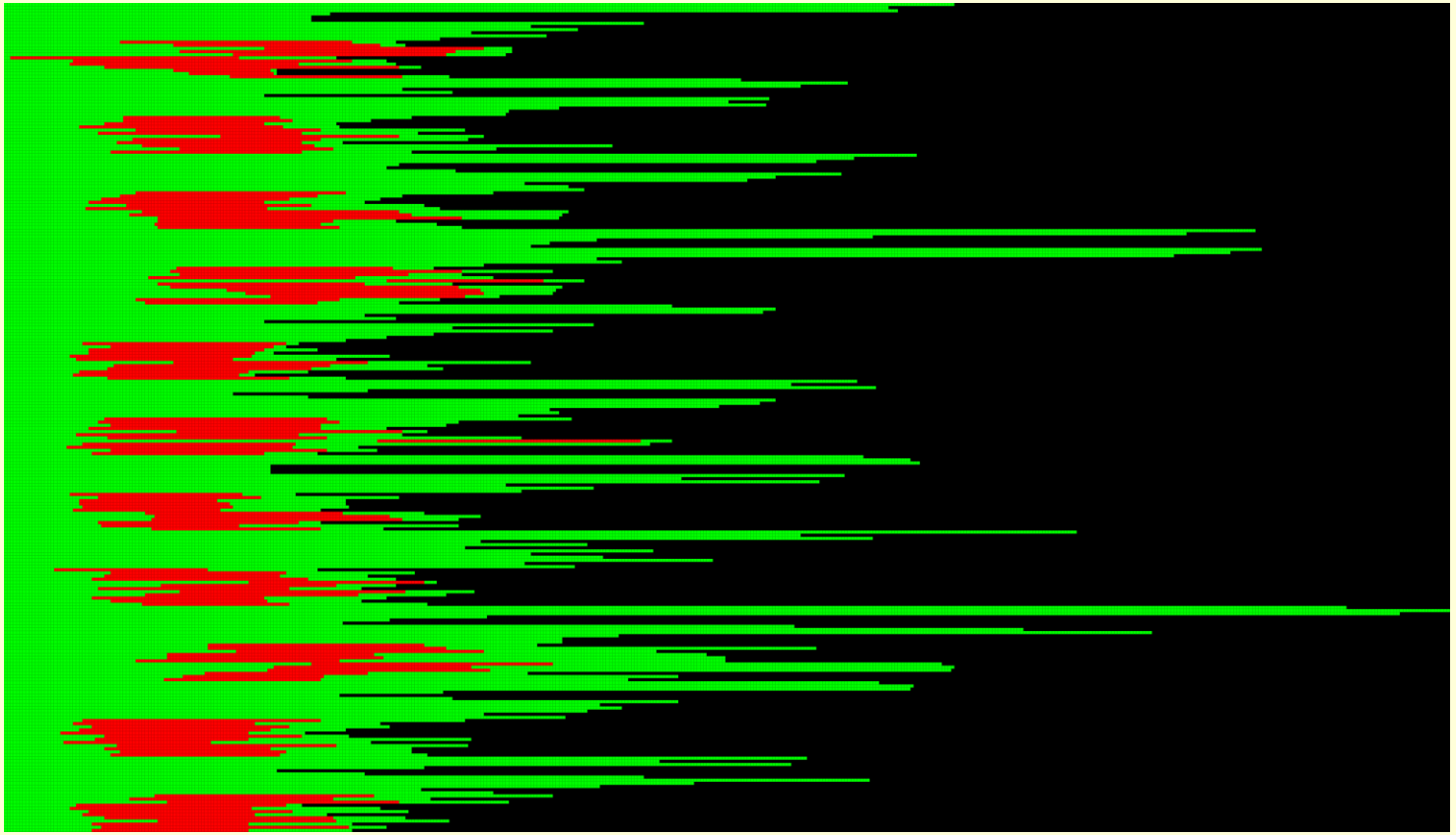
The dataset contains depth frames and skeleton joints collected using Microsoft Kinect v2 and acceleration samples provided by an IMU during the simulation of ADLs and falls.

The dataset is composed by ADLs (Activity of Daily Living) and falls simulated by 11 young actors. The following actions are part of **ADL category**:
- sit, the actor sits on a chair;
- grasp, the actor walks and grasps an object from the floor;
- walk, the actor walks back and forth;
- lay, the actor lies down on the floor;

The following actions are part of the **fall category**:
- front, the actor falls from the floor and ends up lying;
- back, the actor falls backward and ends up lying;
- side, the actor falls to the side and ends up lying;
- endUpSit, the actor falls backward and ends up sitting.

S. Gasparrini, E. Cippitelli, E. Gambi, S. Spinsante, J. Wahslen, I. Orhan and T. Lindh, “Proposal and Experimental Evaluation of Fall Detection Solution Based on Wearable and Depth Data Fusion”, ICT Innovations 2015, Springer International Publishing, 2016. 99-108, doi: 10.1007/978-3-319-25733-4_11.

TST Fall Detection Dataset v2 Labeling



Representation of fall action segments (red), obtained as a result of manual labeling of records of 11 persons in TST Fall Detection dataset v2

Experimental Results

SVM Parameters Tuning

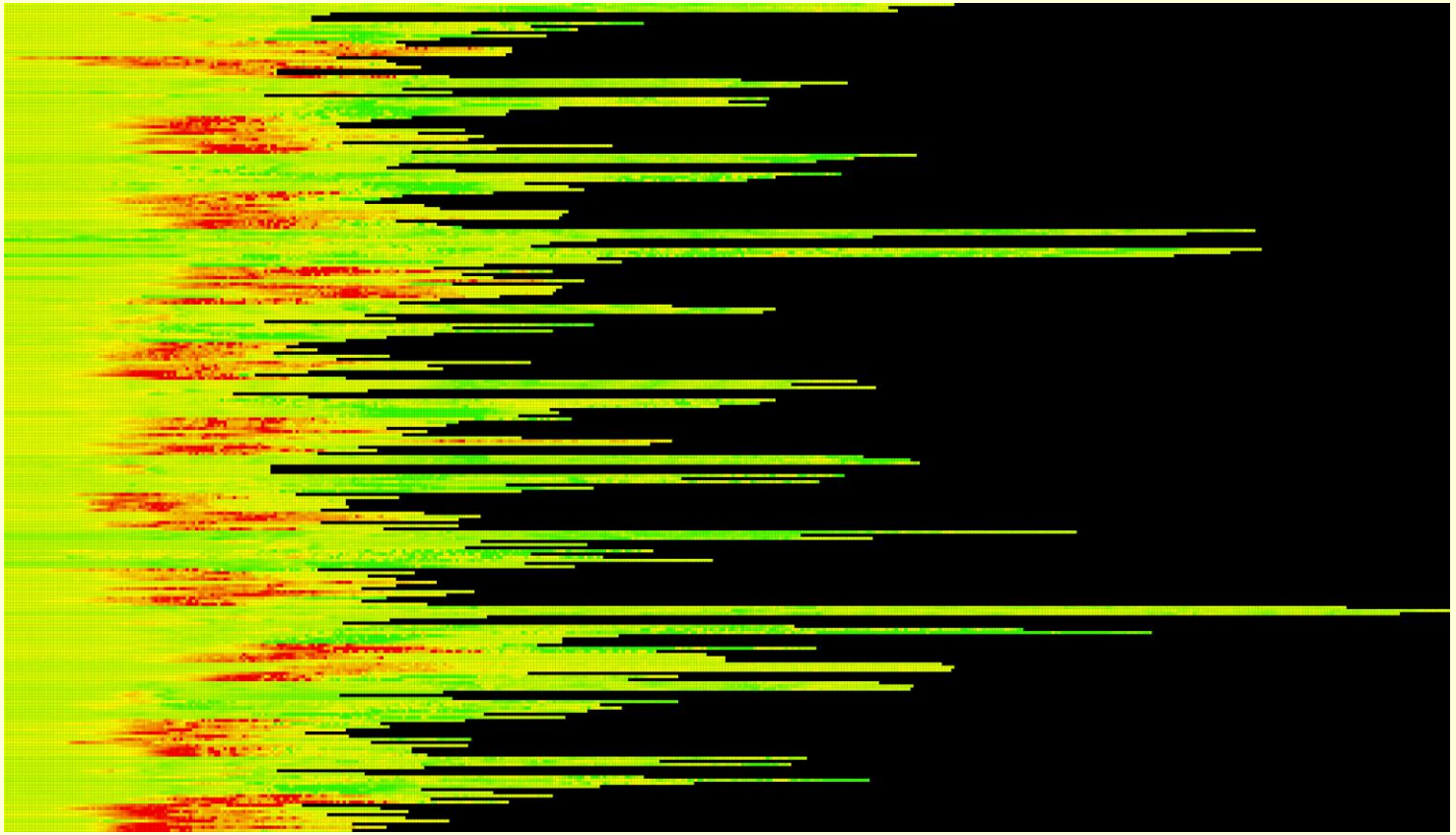
gamma\c	1000	10000	100000
0.001	0.922	0.912	0.905
0.0001	0.930	0.930	0.924
0.00001	0.909	0.922	0.931
0.000001	0.891	0.903	0.913

Persons 1-10 SVM parameters tuning in cross validation procedure on frames

gamma\c	1000	10000	100000
0.001	0.863	0.839	0.837
0.0001	0.889	0.884	0.819
0.00001	0.903	0.896	0.885
0.000001	0.892	0.903	0.896

Person ID=11 SVM recognition accuracy

SVM Parameters Tuning Results



Representation of fall action output (red), obtained as a result of SVM classification of records of 11 persons in TST Fall Detection dataset v2

Results of Recognition

Coincidence in position and duration of fall segments determined by the procedure and labeled by experts for person, not used during training.

lower/ upper	0	1	2	3	4	5
0	0.981	0.987	0.987	0.987	0.987	0.987
-1	0.987	0.987	0.987	0.987	0.987	0.986
-2	0.987	0.987	0.987	0.987	0.986	0.986
-3	0.987	0.987	0.987	0.986	0.986	0.986
-4	0.987	0.987	0.986	0.986	0.986	0.985
-5	0.987	0.986	0.986	0.986	0.985	0.985

Persons 1-10 cumulative sums procedure parameters tuning for coincidence in position and duration of fall segments determined by the procedure and labeled by experts

lower/ upper	0	1	2	3	4	5
0	0.855	0.881	0.878	0.882	0.878	0.878
-1	0.881	0.878	0.882	0.878	0.878	0.880
-2	0.878	0.882	0.878	0.878	0.880	0.878
-3	0.882	0.878	0.878	0.880	0.878	0.870
-4	0.878	0.878	0.880	0.878	0.870	0.870
-5	0.878	0.880	0.878	0.870	0.870	0.868

Coincidence in position and duration of fall segments determined by the procedure and labeled by experts for Person ID=11

Accuracy of fall detection algorithms on TST Fall Detection Dataset v2

Method	Source data	Classifier	Evaluation Scheme	Accuracy
Proposal and Experimental Evaluation of Fall Detection Solution Based on Wearable and Depth Data Fusion	Skeleton joint position; accelerometer data	Empirical thresholding rule	Not described	0.99
Convolutional neural networks (CNN) based human fall detection on Body Sensor Networks (BSN) sensor data	Two accelerometers time series data	CNN	Random 90% and 10% splitting and averaging	0.923
Maximizing Accuracy of Fall Detection and Alert Systems Based on 3D Convolutional Neural Network	Depth map	3D-CNN + data augmentation	5 random trials from 240 and 24 records splitting and averaging	0.942
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	Skeleton joints information	SVM		0.9205
Our	Skeleton joints information	SVM + One-class classifier + CUSUM	Leave-One-Person_Out	0.958

Conclusions

The accuracy of classification for whole record (is fall present inside record or not) is 0.958. The accuracy of coincidence in position and duration of fall segments determined by the procedure and labeled by experts is 0.882. Average delay of fall start position determined by suggested classifier on test records is 6.52 frames. Program realization provides real time fall detection.

Thank you for attention!