XXI IMEKO World Congress “Measurement in Research and Industry”
August 30 – September 4, 2015, Prague, Czech Republic

APPLICATION OF DECISION TREES TO THE FALL DETECTION OF ELDERLY PEOPLE USING DEPTH-BASED SENSORS

Piotr Bilski [1], *Paweł Mazurek* [1] *Jakub Wagner* [1] *Wiesław Winiecki* [1]

[1] Institute of Radioelectronics, Warsaw University of Technology, Warsaw, Poland, pbilski@ire.pw.edu.pl, pmazurek@ire.pw.edu.pl, j.wagner@ire.pw.edu.pl, W.Winiecki@ire.pw.edu.pl

Abstract – The paper presents application of the Decision Tree (DT) to the fall detection of elderly people monitored by the infrared depth sensors. The decision making system works on data acquired by the sensor, recording movement of the person and raising the alarm if his/her behaviour suggests that the accident occurred. From the measurement data so-called morphological features are extracted, further processed by the DT. Various configurations of the classifier have been verified, proving its usefulness to solve the presented task, but also revealing disadvantages.

Keywords: fall detection, decision tree, depth sensor, binary classification

1. INTRODUCTION

The medical care of elderly persons became important in most of developed countries, as their societies age quickly. The fraction of young people (in the range between 20 and 40 years) in the whole population steadily decreases. This causes new problems, amplified by the lengthening of the citizens’ life span. The current challenges lie in improving the quality of life for the increasing the numbers of elderly people, who become the dominant part of the society. Applications of computer technologies in this field were studied for some time. One of the broad group of tasks to solve is the on-line monitoring of the people’s physical condition [1]. To avoid the decrease of their comfort, such persons should be monitored discretely, any encroachment into their privacy (for example using cameras) is potentially risky [2]. This calls for non-intrusive methods of monitoring.

Modern measurement systems may be used for that purpose, as multiple infrared (IR) [3] and radar-based solutions are available. The on-line monitoring module may then be applied, including the distributed architecture, which involves advanced data acquisition (DAQ) and preprocessing, decision making and reporting services. In such applications, Artificial Intelligence (AI) methods are potentially useful.

The paper presents the novel methodology for online monitoring of the elderly person and detecting his/her fall based on the information collected by the IR depth sensors. Such a solution allows for the discreet observation of lodgers in the nursing homes. Two main problems arise here. The first one is the extraction of characteristic features from the measured signals bearing the maximum amount of information about the person’s actual state. The second is the selection of the appropriate decision making module, processing features and determining whether the person requires assistance of the medical staff, or not. This is the binary classification task.

The structure of the paper is as follows. In Section 2 the architecture of the monitoring system is presented. Section 3 contains the information about the measurement data acquisition and processing. In Section 4 the feature extraction from the measurement data is described. Section 5 briefly introduces the decision tree (DT) classifier, adjusted for the task. In Section 6 experimental results are provided. Section 7 contains conclusions and future prospects for the whole system and the DT in particular.

2. MONITORING SYSTEM ARCHITECTURE

The structure of the monitoring system is presented in Fig. 1. The IR sensors are located in the vicinity of the patient, constantly gathering information about her/his condition. If the event of the possible fall is detected (usually related to the abrupt increase in the persons’ position or velocity), the decision making module working as the intelligent algorithm in the computer system processes the data from sensors to decide whether the event that occurred is “the fall” (labelled as “1”) or not (labelled as “0”). In the first case, the alarm is raised and the medical personnel comes to help. Otherwise, the event is ignored and treated as the typical behaviour of the patient.

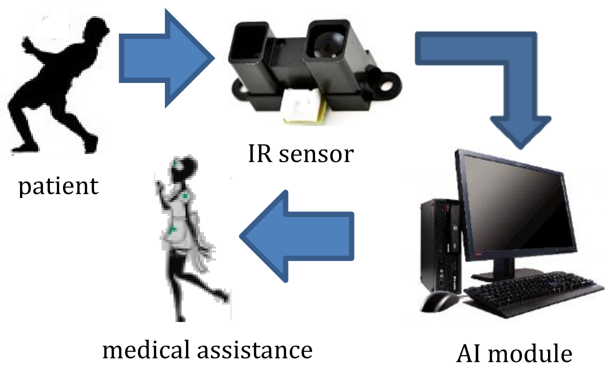

Fig. 1. Architecture of the intelligent system for the elderly people monitoring.

To correctly execute the reasoning process during the decision making, the AI module must possess knowledge about "the fall" scenarios and all other events. The most popular method for collecting such information is to record multiple situations for both types of events, measuring signals, from which features are extracted. Afterwards, the machine learning algorithm may be used to generate knowledge required for distinguishing between events. The generic procedure is presented in Fig. 2. Firstly, the training stage is commenced, during which the measured data is preprocessed (to extract relevant features) and knowledge extracted to establish dependencies between the features and the category of the particular case (further called the example). This knowledge may be then used to classify the actual person's condition to one of two categories. To obtain the system generic enough to react correctly on the subsequent events, multiple scenarios must be considered. Because databases with the desired data are rarely available, recordings are performed with the help of actors, who play the role of the elderly patients in the laboratory environment. This way the training $\boldsymbol{L}$ and testing $\boldsymbol{T}$ data sets are created, containing n examples $\boldsymbol{e}$ (vectors of m characteristic features s_{ij}). They are supplemented with the actual category c_i ("0" or "1"), allowing for implementation of the supervised learning. Both $\boldsymbol{L}$and $\boldsymbol{T}$ have identical structure and their cardinality n depends on the applied validation procedure (see Section 5).

$$\boldsymbol{L} = \boldsymbol{T} = \begin{bmatrix} s_{11} & \cdots & s_{1m} & c_1 \\ \vdots & \ddots & \vdots & \vdots \\ s_{n1} & \cdots & s_{nm} & c_n \end{bmatrix} \quad (1)$$

The remaining question is the accuracy of this "simulated" approach, which may be inaccurate when confronted with the behaviour of the actual person. In the presented research the training examples were acquired by two PhD students (see Section 3) playing their roles in the laboratory environment.

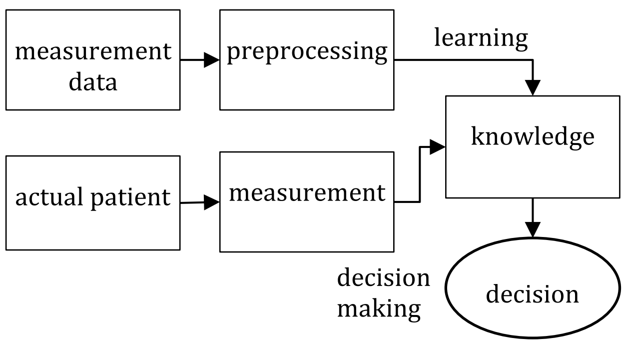

Fig. 2. Scheme of the AI-based fall detection system.

The next problem is making the correct decision about the patient's state. The classifier's efficiency will be evaluated using the sample error e_s, i.e. the number of the incorrectly classified examples from the testing set $\boldsymbol{T}$.

$$e_s = \frac{|\boldsymbol{e} : d(\boldsymbol{e}) \neq c(\boldsymbol{e})|}{|\boldsymbol{T}|} = \frac{FP + FN}{|\boldsymbol{T}|} \quad (2)$$

Usefulness of this approach is limited in this case, as two types of possible errors must be considered. The first one is the false alarm (so-called "False Positive" - FP), i.e. the situation when the fall (event "1") is detected, while it actually did not happen. The consequence of this error is the unnecessary dispatch of the medical assistance to the patient. The second mistake is missing the actual fall (i.e. classifying the event "1" as "0", so-called "False Negative" - FN). This situation is more serious, as such defect of the system may lead to the death of the monitored person. The correct decisions belong to the "True Positive" (correct fall detection) and "True Negative" (correct detection of normal behaviour) class. During the evaluation of the intelligent method used for the task, the TP and TN ratios should be maximized, the FP ratio should be minimized and the FN ratio must be suppressed to zero at the same time. The primary aim of the implemented system is then to detect all falls, even if some false alarms will be raised.

3. TRAINING DATA COLLECTION

This section contains the information about the procedure of collecting the measurement data and preparing it for the training by the DT.

3.1. Data acquisition

The data for testing fall detection algorithms have been acquired using two synchronized IR depth sensors, further called S1 and S2, respectively. The measurement devices are elements of two Kinect module (model 1473). Their configuration, relative to the observed area, is presented in Fig. 3. The distance between them was set to about 3 m. The monitored person was moving at the distance of about 1.5 m to5 m from each device. The subsequent experiments lasted for 10 seconds, and consisted in recording data from both sensors simultaneously, with the frame frequency of 30 fps. They resulted in a sequence of 300 depth images, *i.e.* 480×640 matrices of integer numbers representative of the distance from the sensors.

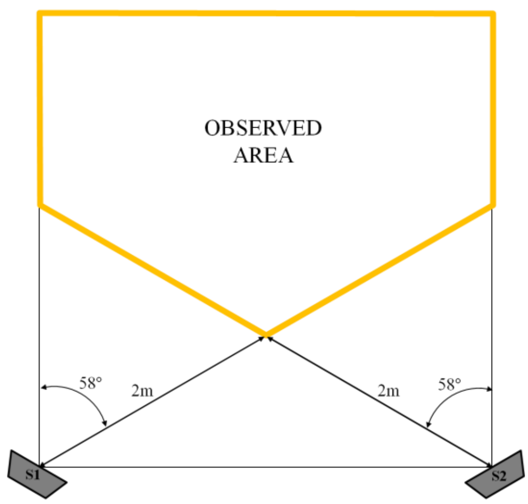

Fig. 3. Configuration of two depth sensors (S1 and S2) relative to the observed area.

A set $\boldsymbol{L}$ of 18 fall scenarios and 18 scenarios corresponding to other actions of a monitored person, called *non-falls*, has been designed. All scenarios have been realized by two actors. Therefore whole set of collected data contains 144 sequences of depth images: each of 36 scenarios was repeated by two actors, and recorded by two S1 and S2.

3.2. Data preprocessing

Every pixel of an image, mapped by S1 or S2, is represented by a triplet of integer numbers (i,k,d), where $i \in \{1, 2, ..., I\}$ is the column index, $k \in \{1, 2, ..., K\}$ is the row number, and $d \in \{1, 2, ..., 5000\}$ is the distance to the sensor (in mm). Since such a relative representation of the image may be a source of ambiguity – the same person may appear as larger or smaller, depending on the distance from the device – it was transformed to the absolute representation based on the global space coordinates (x, y, z). The corresponding mathematical procedure, consists of two operations [4]:

- identification of a set P of pairs (i,k), being the representative of the silhouette of a monitored person (Fig. 4);
- transformation of the triplets (i,k,d), corresponding to P, into the triplets (x, y, z).

Fig. 4. Extraction of the silhouette of a monitored person: a) the original image recorded by the sensor S1; b) the result of extraction.

The absolute coordinates $(x_{i,k}, y_{i,k}, z_{i,k})$, calculated for all $(i,k) \in \mathrm{P}$, were used for computation of the coordinates of the silhouette "mass centre":

$$(x_C, y_C, z_C) = \frac{1}{|\mathrm{P}|} \sum_{(i,k)\in \mathrm{P}} (x_{i,k}, y_{i,k}, z_{i,k}), \tag{3}$$

and its magnitude, i.e. the effective reflection area:

$$M = \frac{1}{|\mathrm{P}|}\left(\sum_{(i,k)\in \mathrm{P}} d_{i,k}\right)^2 . \tag{4}$$

The application of this procedure to a sequence of depth images, acquired at the predefined time instants t_n ($n = 1, 2, ...$), resulted in four-dimensional trajectories:

$$\{x_C(t_n), y_C(t_n), z_C(t_n), M(t_n)\} \tag{5}$$

which may be used for classification of events recorded by S1 or S2.

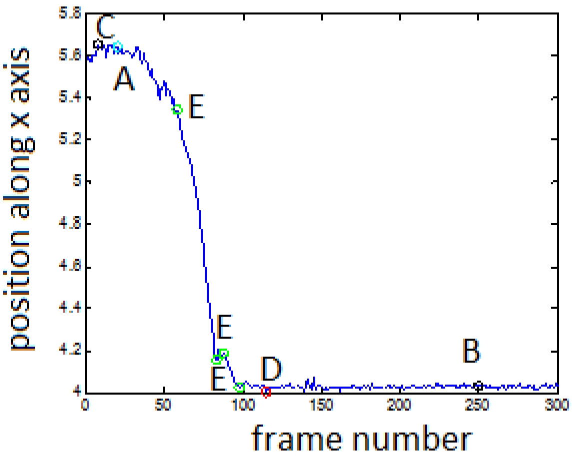

Fig. 5. Extraction of the features from the position pattern.

4. FEATURE EXTRACTION

From the recorded trajectories (3), six patterns were analysed for the single scenario. As opposed to other researchers [5], in this project the characteristic points (further called "morphological") were extracted from each pattern (Fig. 5). They include:

- the value of the pattern (position or velocity in every of three dimensions) at the beginning of the event (in the 20th frame - the black point A)
- the value of the pattern at the end of the event (in the 50th frame from the end – the black point B).
- the maximum and minimum value of the pattern and the number of corresponding frames (points C and D, respectively).
- coordinates of two the greatest differences between the neighbouring minimum and maximum values (green points E in Fig. 5).

The features were extracted from the original patterns without any prior preprocessing, such as the elimination of the background noise. The applied approach was sufficient to ensure the acceptable accuracy. Alternatively, the de-noising procedures (including the polynomial and spline approximation) were considered.

The overall number of features from each of six patterns was 14, which gives 84 features for the particular example. This way 144 training examples were created for the DT to process.

Note, that the feature extraction in the actual case require the prior event detection (i.e. suspicious behaviour of the patient, which might be his fall) based on the on-line analysis of the IR signals. This stage is omitted in the presented research, but must be required in the practical implementation of the system.

5. DECISION TREES

The DT [6] (Fig. 6) is the hierarchical structure of nodes, starting from the initial root at the first level and ending with leaves, which contain the particular category. Each parent node is connected to two child nodes (or leaves) belonging to the lower level. The edges connecting nodes of lower level are related to the test stored in the parent node. It is the threshold value θ of the selected feature s_i, to which the corresponding value of the example is compared. The sequence of tests enables the tree to make the decision about the event category. The example "travels" from the root to one of the leaves. At each intermediate node the test is executed for the example. The comparison between the threshold θ and the value of the corresponding attribute s_i in the example causes its relocation to one of two child nodes. This process is repeated until the example reaches the leaf, which points at the particular category.

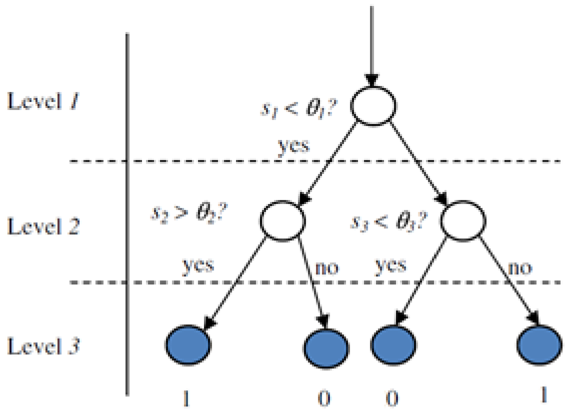

Fig. 6. Structure of the decision tree in the binary classification problem.

The application of the decision tree requires dividing the available data $\boldsymbol{D}$ into the training and testing subsets: $\boldsymbol{L}$ and $\boldsymbol{T}$, respectively (Fig. 2). The Leave-One-Out (LOO) and Repeated Random Subsampling (RRSS) Cross-Validation (CV) were used to check the generalization abilities of the classifier. In the former, $\boldsymbol{D}$ is divided into k subsets (each of the size of $\boldsymbol{D}/k$). The operation consists in selecting one subset as $\boldsymbol{T}$, while remaining ones become $\boldsymbol{L}$. The assignment operation is repeated k times, resulting in k classification errors. The overall performance of DT is measured as the mean value of e_s for k trials:

$$\bar{e}_s = \frac{\sum_{i=1}^{k} e_s(i)}{k} \qquad (6)$$

Because the division of $\boldsymbol{D}$ into $\boldsymbol{L}$ and $\boldsymbol{T}$ may be conducted in multiple ways, RSSSCV should be repeated a couple of times to verify, if the random configurations of $\boldsymbol{L}$ and $\boldsymbol{T}$ content change the value of e_s and the threat of missing the fall event. Therefore the mean value of e_s with the standard deviations may be used to supplement the information about the accuracy (6) with its repeatability.

In LOOCV, $\boldsymbol{D}$ is systematically divided into one-example testing subsets, while all remaining data belong to the $\boldsymbol{L}$. It is the most exhaustive verification ($k=|\boldsymbol{D}|$), evaluating the influence of all possible training and testing data combinations on the value of $\boldsymbol{e_s}$.

The tree induction [7] is applied to extract knowledge from the measurement data. In each step of this recurrent process, the node with the threshold θ, or the leaf is created. The node divides the currently processed set $\boldsymbol{L}$ into two subsets of similar cardinality in such a way that each category is present only in one of them. This way examples are sorted according to their categories using the minimal number of nodes. If the current subset contains only examples belonging to one category, the further division is not needed. Instead, the leaf is created, representing this category. The training process is susceptible to overlearning. All examples from the training set may be separated correctly, but the performance on the testing set will be low. This means the DT has limited generalization abilities, which are evaluated with CV.

The entropy was used to select the best candidates for the threshold. This is the measure of disorder in data, with high value for the threshold θ creating two subsets, which contain the same categories, distributed evenly (in this case, both would contain examples labelled with "0" and "1"). The low entropy value is for the thresholds dividing examples into two subsets with mutually excluding categories. This is preferred, as it partitions data into separate categories in the shortest amount of time, leading to the simplest tree structure. Candidates for thresholds are calculated as the middle points between neighbouring values of sorted symptom values. Next, one of them with the minimum entropy is selected. Because there are multiple candidates for the threshold in each node with equal, minimum entropy value, various strategies of the selection may be applied. In the presented research, the following methods were implemented:

a) selection as the threshold the attribute with the largest distance from the neighbouring values
b) selection as the threshold the attribute with the smallest distance from the neighbouring values
c) selection as the threshold the attribute occurring in the tree nodes the greatest number of times
d) selection as the threshold the features occurring in the tree nodes the smallest number of times
e) random selection of the feature from the subset of the ones with the smallest entropy.

Influence of the selected strategy on the obtained classification accuracy is discussed in the experimental results section.

6. EXPERIMENTAL RESULTS

The conducted experiments included verification of the decision tree parameters and its ability to suppress the *FN* ratio (missing actual falls). Comparison between the classification outcomes for various strategies is in Table 1, where "1-e_s" is the mean percentage of the correct classifications. Columns "*FP*" and "*FN*" present the percentages of false alarms and missed falls, respectively. Three CV methods (RRSSCV for k=*5* and *10* and LOOCV) are considered. Criteria (a) and (d) are the most promising, ensuring the greatest separation in the training data and providing the greatest variability in features. They not only provide the smallest sample error, but also ensure the minimal number of falls missed. The percentage of false alarms is also small, although this error is of secondary importance.

The detailed results for the threshold selection strategy (a) are in Table 2, while for the criterion (c) – in Table 3. They show the contrast in accuracy between the best and the worst option and their relation with the tree structure.

The comparison between the CV approaches shows the LOOCV method is the most reliable, as it does not depend on the random process of selecting examples to training and testing subsets. Although the most time-consuming, it also shows the performance of the classifier working on the largest ***L*** available. On the other hand, variants of the RRSS verification depend on the partitioning of examples into ***L*** and ***T*** and will be different each time. To suppress the randomness between trials, every RRSSCV experiment was repeated five times, resulting not only in the mean classification accuracy, but also standard deviation, representing variability in particular trials. Mean classification accuracies for the single experiment regarding the strategy (c) are in Fig. 7, while standard deviations of these results are in Fig. 8. The fall detection experiment was repeated five times. The standard deviation is relatively high, proving that LOOCV is the most reliable validation approach. Also, there are no significant differences between the different values of *k*. The greatest impact on the obtained accuracy has the selected node generation strategy. Also, the problem of separating between two categories of different weights (as missing the fall has more serious consequences than the false alarm) exists for the DT, which does not have any intrinsic mechanism for distinguishing between these two types of classifier errors. To solve the problem, the additional method of training data preprocessing must be proposed and will be investigated in the future.

Table 1. Comparative analysis of the fall detection accuracy regarding various node construction strategies (LOOCV).

Strategy	1-e_s [%]	*FP* [%]	*FN* [%]
(a)	93,75	3,41	2,72
(b)	87,58	7,28	4,62
(c)	75,69	9,02	15,27
(d)	89,33	8,16	2,13
(e)	88,19	9,02	2,77

Table 2. Cross-validation results (mean values for RSSSCV) for the decision tree with thresholds selected using the maximum distance from separated attribute values (criterion (a)).

CV type	1-e_s [%]	*FP* [%]	*FN* [%]
LOOV	93,75	3,41	2,72
RSSCV-5	92,58	4,22	3,25
RSSCV-10	89,33	7,3	3,3

Table 3. Cross-validation results (mean values for RSSSCV) for the decision tree with thresholds selected using the maximum number of occurrences of this threshold in the tree (criterion (c)).

	1-e_s [%]	FP [%]	FN [%]
LOOV	75,69	9,02	15,27
RSSCV-5	85,51	8,6	5,55
RSSCV-10	87,33	6	6,66

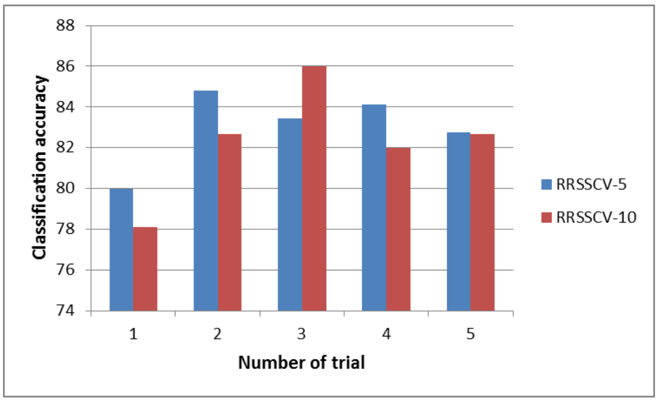

Fig. 7. Classification outcomes for subsequent trials in two RRSSCV schemes (strategy (c)).

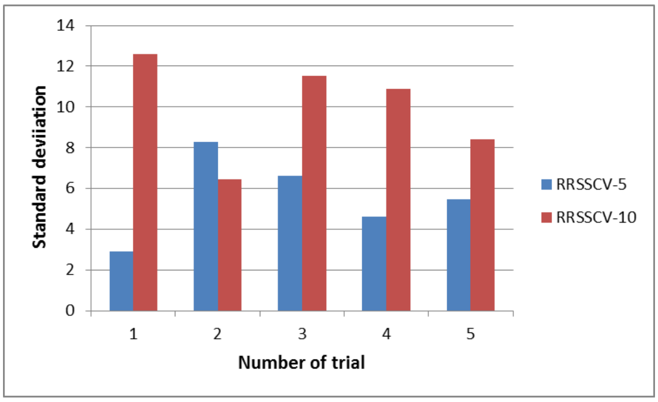

Fig. 8. Standard deviations of classification outcomes for subsequent trials in two RRSSCV schemes (strategy (c)).

7. CONCLUSIONS

The decision tree is the fast and accurate method for the fall detection based on the IR sensors measurement data. To maximize its performance, the method of the threshold for the node selection must be carefully designed. Also, the training data preprocessing should be done before training the tree to decrease the number of false alarms and missing actual falls. The considered task is complex and requires applying additional classification methods and considering various sets of features.

As the selected features have a significant impact on the behaviour of the decision making module, various attributes should be checked and compared (including, for example, spectral or cepstral information). The second problem to solve regarding the DT application is the introduction of the category weights, allowing for detection of all actual falls (i.e. *FN=0*). The tree itself is unable to identify the varying importance of categories, therefore the separating areas between the categories should be dislocated (by changing the original labels of some examples belonging to category “0” to “1”). This should prevent from selecting thresholds correctly separating only examples from ***L***.

The future research should also include the comparative analysis between the DT and other classifiers, including random forest, artificial neural network or k Nearest Neighbours. The classifier fusion (including various methods working in parallel) must be considered as well.

ACKNOWLEDGMENTS

This work was supported by the Norway Grant entitled “Care support for elderly and disabled people by radar-sensor technology (RadCare)”.

REFERENCES

[1] Z. Zhou, E.E. Stone, M. Skubic, J. Keller and Z. He, "Nighttime In-Home Action Monitoring For Eldercare," *33rd Annual International Conference of the IEEE EMBS*, pp. 5299-5302, Boston, Massachusetts USA, August 30 - September 3, 2011.

[2] M. Kepski, B. Kwolek, "Fall Detection Using Ceiling-Mounted 3D Depth Camera," *Int. Conf. on Computer Vision Theory and Applications VISAPP 2014*, vol. 2 pp. 640-647, SciTePress 2014.

[3] G. Mastorakis and D. Makris, "Fall detection system using Kinect’s infrared sensor," *Journal of Real-Time Image Processing*, March 2012, DOI 10.1007/s11554-012-0246-9.

[4] P. Mazurek, J. Wagner and R.Z. Morawski, "Acquisition and preprocessing of data from IR depth sensors, developed for a system of patients monitoring," submitted to the *8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Warsaw, Poland, 2015.

[5] A. Davari, T. Aydin and T. Erdem, "Automatic fall detection for elderly by using features extracted from skeletal data," *International Conference on Electronics, Computer and Computation (ICECCO)*, pp. 127-130, Ankara , Turkey, 7-9 Nov. 2013.

[6] N.-T. Nguyen and H.-H. Lee, "Decision Tree with Optimal Feature Selection for Bearing Fault Detection," *Journal of Power Electronics*, vol. 8, nº. 1, pp. 101-107, January 2008.

[7] A. Jamehbozorg and S.M. Shahrtash, "A Decision-Tree-Based Method for Fault Classification in Single-Circuit Transmission Lines," *IEEE Transactions on Power Delivery*, vol. 25, nº. 4, pp. 2190-2196, October 2010.

[8] H. Foroughi, B. Shakeri Aski, and H. Pourreza, "Intelligent Video Surveillance for Monitoring Fall Detection of Elderly in Home Environments," Proc. 11th International Conference on Computer and Information Technology (ICCIT 2008), 25-27 December, 2008, Khulna, Bangladesh, pp. 219-224.

[9] L.F. Yeung, K. C. Cheng, C.H. Fong, W. C.C. Lee, K. Y. Tong, "Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway," Gait & Posture 40 (2014), pp. 532–538.

[10] M. Mubashir, L. Shao, L. Seed, "A survey on fall detection: Principles and approaches," Neurocomputing 100 (2013) pp. 144–152.

[11] M. Kepski, B. Kwolek, and I. Austvoll, "Fuzzy Inference-Based Reliable Fall Detection Using Kinect and Accelerometer," ICAISC 2012, Part I, LNCS 7267, pp. 266–273, 2012.

[12] S. K. Saidulu, B. Karunaiah, K. V. Murali Mohan, "Sensors based Wearable and Systems for Detection of Human Movement and Fall," International Journal of Engineering Trends and Technology (IJETT) – Vol. 8 No. 3, Feb 2014, pp. 109-112.