A Multimodal Approach in Smart House for Fall Detection

Yves M. Galvão, Vinícius A. de Albuquerque, Bruno J. T. Fernandes
Polytechnic School of Pernambuco - POLI
University of Pernambuco - UPE
Email: {ymg,vaa,bjtf}@ecomp.poli.br

***Abstract*—The population of older adults is increasing quickly in the entire world. According to the U.S. Centers for Disease Control and Prevention, every 11 seconds an older adult is treated in the emergency room for a fall, and every 19 minutes an older adult dies from a fall. So it is important to identify and classify a fall correctly. Although health professionals do the monitoring of this people, the cost of this monitoring is expensive. A possibility for these people is a smart noninvasive system that monitors a daily behavior and emits alerts in case of detecting an abnormal event. To detect falls, sensors like accelerometers and gyroscopes can be used for the acquisition of data and in conjunction with the acquired image from cameras allow the creation of a multimodal approach to fall detection. This short paper focus on detecting falls using different classifiers in three data sets. The experiments realized show that it is possible to identify a fall with an accuracy of 99,70% in most scenarios.**

***Index Terms*—multimodal, fall detection, older adults, elderly people, k-nn, mlp, svm, accelerometers, kinect camera, noninvasive, smart house.**

I. INTRODUCTION

A Fall represents a significant health problem, especially for older adults [1]. The number of senior people will double until 2050, so it is essential the monitoring of these people [2]. Chew et al. [3] say that among the elderly, 55 percent of fall injuries occur inside their home. In this context, one of the methods to monitor these people is accomplished with a daily human presence of a professional health care. However, this process is expensive and invasive, eliminating the privacy of the individual [4], [5].

Another approach is a noninvasive monitoring. Smart houses, a relevant application to Internet of Things (IoT), is a noninvasive monitoring system [6]. So this type of application is advantageous when it comes to houses of individuals who need constant care, for example, older adults.

A smart home allows monitoring these people using connected devices, such as cameras, infrared sensors, and others. In the context of falls, the use of cameras in combination with techniques of computer vision and image processing allows the identification of falls [7], [8].

Specifically for falls, the use of accelerometers help to classify an event as a fall. Accelerometers correspond on being physical devices that need to be utilized daily by monitored elderly. These devices frequently are a belt, a necklace or a smartphone.

Combining the use of accelerometers and computer vision, we create a multimodal approach [9]. Srivastava and Salakhutdinov [10] define a multimodal setting as data that have multiple input modalities, where each modality has a different kind of representation and correlational structure.

This short paper realizes a comparative study between accelerometers' data, images' data and a multimodal approach combining accelerometers and pictures in the same dataset. The classifiers used the data from accelerometers and preprocessed images to determine if a fall occurs.

After the evaluation of the experiments, the results of classifiers achieve a high rate of accuracy, reaching an accuracy above 99%.

II. RELATED WORK

There are various works in the field of Fall Detection, most of them are related to wearable approach using accelerometers [9]. In this Section, we describe a few of these related works. Chen et al. [3] use a custom device to detect fall, this design consists of a board with two batteries and two accelerometers.

In another study, Ozdemir and Bachar [11] use machine learning to identify and classify a fall. They use different classifiers, and the data was acquired via six wearable motion sensors, where each unit has three tri-axial devices (accelerometer, gyroscope, and magnetometer/compass). K-Nearest Neighbor (K-NN) obtained the best result, getting 99% of accuracy.

In the context of computer vision, a depth camera was used by Bian et al. [12] to motion tracking a person and detect a fall. For the input of Support Vector Machine (SVM) classifier, they have used the feature extraction of images.

In this work, we use a combination of accelerometers and pictures obtained from a camera to create a model that detects a fall with a minimum number of false negatives (FN).

III. METHODOLOGY

A device with two accelerometers and one gyroscope and a Microsoft Kinect camera were used for the data acquisition.

The datasets have 30 falls registers and 40 of activities of Daily Living (ADLs). Each of falls and ADLs register has a sequence of images or accelerometer data. The number of data is variable according to the executed sample. For the preprocessing, we executed the following steps.

- Image transformation to gray scale (256).
- Image resizing to 40 x 30 pixels.

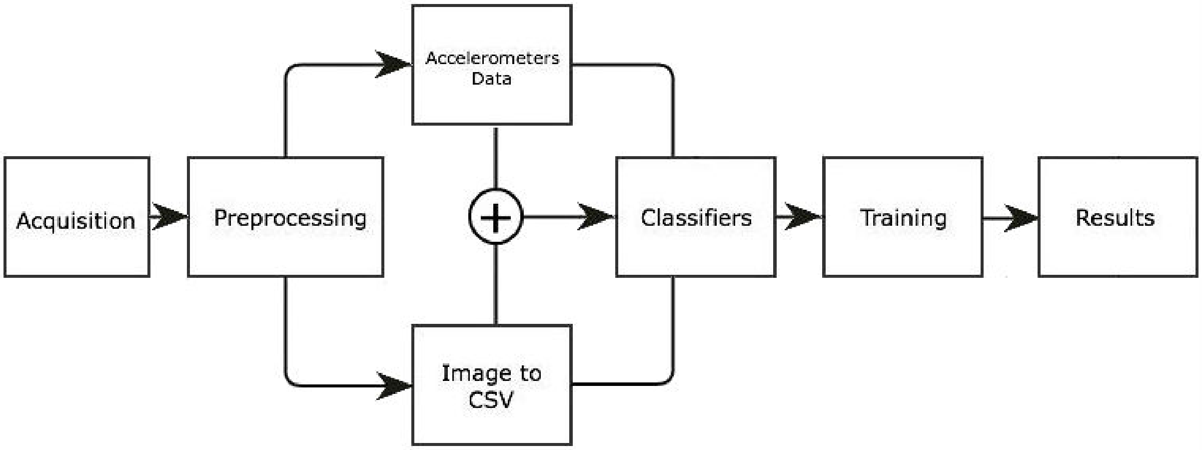

Fig. 1. Multimodal proposed model.

- All pixels were added to a file and labeled as fall or no fall

The image downsizing was necessary for the allow that training phase occurs faster. For the experiments realized, we have used the following classifiers: Multi Layer Perceptron (MLP), K-NN and SVM. Section IV describes the best combination of parameters for these classifiers.

These classifiers are used to predict falls using the image and accelerometers data, or both, in a multimodal experiment proposed in this work.

The intention to use a multimodal approach is to improve the results obtained without this method. The model for the proposed method can be seen in the Figure 1.

IV. Experiments

This section explains the utilized dataset, the experiments and results obtained.

A. Dataset

The following public dataset was utilized in the experiments: UR Fall Detection Dataset [13]. We have split the dataset into three different parts:

- Dataset 1: contains the captured image frames with the ADL and fall registers.
- Dataset 2: contains the data of the accelerometers with the ADL and fall records.
- Dataset 3: multimodal data, combining dataset 1 and dataset 2.

For the images of the dataset 1 and 3, We have chosen the camera 0 with the RGB images. The dataset 1 and dataset 2 have the image frame and accelerometer synchronized. So it is easy to identify when a fall occurs using the accelerometers.

After the preprocessing, the number of falls was defined in 921 and no fall is 10623, totalizing 11544 samples.

The images of dataset 1 were acquired from a Kinect camera. On the dataset 2, all data were acquired from a device that contains two accelerometers and one gyroscope. Dataset 2 has raw accelerometer data defined in gravity units (g).

To create the dataset 3 the acquired images were concatenated with the accelerometers data in the same file. This process is represented by the (+) symbol in the flow described in Figure 1. To synchronize the data of falls and no falls between dataset 1 and dataset 2, the label of dataset 1 was utilized and the data of accelerometer dataset was synchronized according to the timestamp of the event.

In table I, it is possible to verify the number of input of all datasets utilized in this work.

TABLE I
The number of the datasets' inputs

Dataset	Number of inputs
Dataset 1 (Images)	1200
Dataset 2 (Accelerometers)	4
Dataset 3 (Images + Accelerometers)	1204

Different types of ADLs and falls are presented in this dataset, such as:

- ADLs
  - Walking slowly
  - Praying
  - Sit on a chair
- Falls
  - Fall forward while walking
  - Fall forward when trying to get up of a chair

An important fact is that older adults group did not perform the ADLs and fall activities because of medical recommendation.

B. Experimental settings

In this short paper, we realized different tests. For all datasets the following parameters were used: For MLP, the activation functions tested were sigmoid, hyperbolic tan and linear. The number of hidden layer neurons also was modified between 1 and 5.

SVM had one eps value changed in the tests and the kernel values. Finally, for the K-NN classifier, the number of K was tested for different values, varying from 1 to 15. The tests were executed with cross-validation with 3 folds.

The best combination of parameters for the dataset 1, 2 and 3 is described in Table II.

TABLE II
BEST COMBINATION OF PARAMETERS

Classifier	Parameters	Dataset 1	Dataset 2	Dataset 3
MLP	Activation Function	Sigmoid	Sigmoid	Sigmoid
	Learning Rate	0.3	0.5	0.3
	Hidden Layers	1	1	1
	Hidden Neurons	4	4	6
K-NN	K	3	5	3
	Cross validate	False	False	True
	Search Algorithm	Linear	Linear	Linear
SVM	Kernel	Poly	Linear	Poly
	Eps	0.001	0.6	0.4
	Cost	1	1	1

C. Results

In this Section, we describe the obtained results for the tests.

1) Dataset 1: The results for the classifiers K-NN, MLP and SVM are respectively in the Table III. The table IV show the confusion matrices for all results.

TABLE III
RESULTS FOR CLASSIFIERS ON DATASET 1 (IMAGES)

Classifier	Accuracy	Cost for classification
K-NN	99.67%	0.24s
MLP	99.50%	0.01s
SVM	99.48%	0.01s

TABLE IV
CONFUSION MATRICES FOR DATASET 1 (IMAGES)

	MLP		K-NN		SVM	
	F	N	F	N	F	N
F = FALL	888	33	899	22	889	32
N = NO_FALL	24	10599	16	10607	28	10595

The K-NN classifier obtained the highest accuracy, but the time for classification is 24x greater than the other classifiers. That represents a possible issue in a real application.

2) Dataset 2: For accelerometer data, the results for the classifiers are in Table V. The Table VI show the confusion matrices for all results.

TABLE V
RESULTS FOR CLASSIFIERS ON DATASET 2 (ACCELEROMETERS)

Classifier	Accuracy	Cost for classification
K-NN	97.40%	0.10s
MLP	95.03%	0.03s
SVM	94.24%	0.01s

TABLE VI
CONFUSION MATRICES FOR DATASET 2 (ACCELEROMETERS)

	MLP		K-NN		SVM	
	F	N	F	N	F	N
F = FALL	2309	64	2251	122	2086	287
N = NO_FALL	509	8662	178	8993	377	8794

In this phase, the classifier that has obtained the highest accuracy was the K-NN, but the number of False Positives of the MLP was the lowest. The classification time for the SVM and MLP classifiers were approximately 10x faster than the K-NN.

3) Dataset 3: At the final phase, the best results were obtained by the K-NN classifier. All results can be seen in the Table VII. The table VIII show the confusion matrices for final results.

TABLE VII
RESULTS FOR CLASSIFIERS ON DATASET 3 (IMAGES + ACCELEROMETERS)

Classifier	Accuracy	Cost for classification
K-NN	99.70%	1.01s
MLP	99.69%	0.03s
SVM	99.48%	0.44s

TABLE VIII
CONFUSION MATRICES FOR DATASET 3 (IMAGES + ACCELEROMETERS)

	MLP		K-NN		SVM	
	F	N	F	N	F	N
F = FALL	900	21	907	14	890	31
N = NO_FALL	14	10609	20	10603	29	10594

To verify the misclassifications, the Figure 2 show some the False Positives (FP) and False Negatives (FN) identified during the classification process. The pictures labeled with 1 and 2, are commons mistakes for each dataset tested.

Fig. 2. The frames 1 and 2 show two FP, 3 and 4 show two FN

Important information in this results is the total cost of time. This value represents the time necessary for an image frame to be detected as a fall.

The proposed model needs to classify a fall quickly to the daily behavior real time monitor. So, although the accuracy obtained by the K-NN classifier is higher than the other classifiers, MLP have got similar results and also can be utilized for this goal with success.

V. Conclusion

This short paper shows possible solutions to detect falls. We tested different classifiers on the UR Fall Detection dataset for the reached proposed goal.

The obtained results showed that the multimodal approach reached an accuracy score above 99%. In comparison with the single method (only image or accelerometers data), the multimodal represents a gain of 0.1%. Furthermore, the number of FP has reduced from 22 to 14. It is suitable for the problem described.

For the proposed work, we had a limitation on the dataset of not having data specific of older adults. So the accuracy obtained can be lower when tested with the senior people.

For the future work:

- Include cross validation test with data for senior people x current data set.
- Perform tests with different public data sets.

References

[1] A. F. Ambrose, G. Paul, and J. M. Hausdorff, "Risk factors for falls among older adults: a review of the literature," vol. 75, no. 1. Elsevier, 2013, pp. 51–61.

[2] U. Bakar, H. Ghayvat, S. Hasanm, and S. Mukhopadhyay, *Activity and anomaly detection in smart home: A survey*. Springer, 2016.

[3] J. Chen, K. Kwong, D. Chang, J. Luk, and R. Bajcsy, "Wearable sensors for reliable fall detection," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2006, pp. 3551–3554.

[4] V. Jakkula, D. J. Cook *et al.*, "Anomaly detection using temporal data mining in a smart home environment," *Methods of information in medicine*, vol. 47, no. 1, pp. 70–75, 2008.

[5] S. Simoens, M. Villeneuve, and J. Hurst, "Tackling nurse shortages in oecd countries: Oecd health working paper number 19," *Cedex, France: Directorate for Employment, Labour and Social Affairs*, 2005.

[6] A. Agarwal, S. Dawson, D. McKee, P. Eugster, M. Tancreti, and V. Sundaram, "Detecting abnormalities in iot program executions through control-flow-based features," in *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*. ACM, 2017, pp. 339–340.

[7] E. Hoque, R. F. Dickerson, S. M. Preum, M. Hanson, A. Barth, and J. A. Stankovic, "Holmes: A comprehensive anomaly detection system for daily in-home activities," in *Distributed Computing in Sensor Systems (DCOSS), 2015 International Conference on*. IEEE, 2015, pp. 40–51.

[8] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and V. Bulgari, "From lab to life: Fine-grained behavior monitoring in the elderly's home," in *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*. IEEE, 2015, pp. 342–347.

[9] F. Feldwieser, M. Gietzelt, M. Goevercin, M. Marschollek, M. Meis, S. Winkelbach, K. H. Wolf, J. Spehr, and E. Steinhagen-Thiessen, "Multimodal sensor-based fall detection within the domestic environment of elderly people," *Zeitschrift für Gerontologie und Geriatrie*, vol. 47, no. 8, pp. 661–665, 2014.

[10] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 2222–2230. [Online]. Available: http://papers.nips.cc/paper/4683-multimodal-learning-with-deep-boltzmann-machines.pdf

[11] A. T. Özdemir and B. Barshan, "Detecting falls with wearable sensors using machine learning techniques," *Sensors*, vol. 14, no. 6, pp. 10 691–10 708, 2014.

[12] Z.-P. Bian, J. Hou, L.-P. Chau, and N. Magnenat-Thalmann, "Fall detection based on body part tracking using a depth camera," *IEEE journal of biomedical and health informatics*, vol. 19, no. 2, pp. 430–439, 2015.

[13] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," vol. 117, no. 3. Elsevier, 2014, pp. 489–501.