Feature selection to detect fallen pose using depth images

Carolina Maldonado*,Homero Ríos*, Efrén Mezura-Montes*, Antonio Marin*,
*Centro de Investigación en
Inteligencia Artificial,
Universidad Veracruzana,
Sebastián Camacho No. 5
Xalapa, Ver. CP 91000

{cmaldonado,hrios,emezura,amarin}@uv.mx

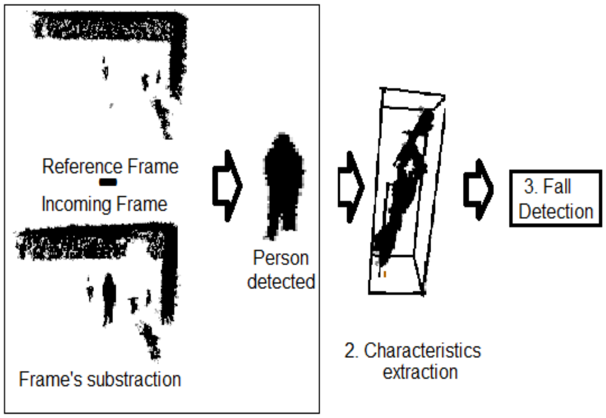
***Abstract*—In this paper we are interesting in knowing which features provide useful information for detecting a fall and how the set of selected characteristics impact the performance of detection. Then we define a large set of possible features, which are extracted from a cloud of points of a person by the kinect device, some of features were used in previous work, and we propose to add and evaluate the effect of using 3D moment invariants translation, scale an rotation, and other geometric characteristics. Two experiments are carried out to analyze the effect of using two different subset of features, one of them selected by a Genetic Algorithm and the second by Principal Component Analysis (PCA). The obtained results suggest that the success of detection of fall depends on the selected features, and the genetic algorithm is a good technique to select them, when compared with PCA.**

I. Introduction

The elderly population is a sector that can be considered vulnerable because they are more prone to falls, even when they are inside the home. According to [1] [2], 65% of falls occur at home and 26% on public roads. Several works have been developed for detecting falls events, in [3] it was defined that a fall is characterized by a large motion combined with a change in the human shape, for detecting the human shape and orientation change. The authors approximated an ellipse to a human and they obtained a fall detection rate with a sensitivity of 88% and specificity of 87.5%, by using a webcam. In [4] a depth sensor was used and a fall was detected based on two features: human centroid height relative to the ground and 3D person velocity, they reported the success of their method was 98.7%. In [5], a fallen pose in elderly persons was detected based on the combination of two features: aspects ratio and fall angle, while the accuracy rate of the system was 92.5%. In [6] a robust fall detection approach was proposed by the tracking of the head and hip joints using a single depth camera and the 3D trajectory of the head joint. The authors calculated the distance between the joint and the floor. They report an accuracy of 97.6%. In [7] the detection of the fall was done by collecting data from an accelerometer and a depth sensor; they extracted several characteristics like a ratio of width to height of the persons bounding box, a ratio expressing the height of the persons surrounding box in the current frame to the physical height of the person, the distance of the persons centroid to the floor, standard deviation from the centroid for the abscissa. They reported an accuracy of 90% using only data from the depth sensor, and 98.33%, with data from the depth sensor and the accelerometer.

From the above literature review, it can be noted that, not all cited works used the same set of features, and they reported different rates of accuracy. Therefore, the interest of this work is to evalute which characteristics are those leading to efficiently detect falls, and compare how different characteristics affect the results of detecting falls. For this purpose, the main contributions of this paper is proposing a set of twenty characteristics that give information about the human shape and orientation. Table I shows the complete set and a brief description of each one. Another contribution is including 3D invariant moments [8], which describe the objects by a set of scalar quantities that are insensitive to particular deformations (translation, rotation and scale- TRS) and that provide enough discrimination power to distinguish objects belonging to different classes, these, to the best of the authors' knowledge, have never been used in previous fall detecting work. For selecting the optimal subset of characteristics, we propose to use a genetic algorithm, because they have demonstrated to find good solutions in large search spaces, and the PCA method which measures the data variances and it is a well known method. The results of both methods let us to know the optimal subset of features for detecting a fallen pose. In this paper the detection of the fall is from a non-intrusive system for the observed person, then the person should not be placed sensors on his body, for which a solution based on computer vision system is used, we use a depht sensor.

The paper is organized as follows: Section II details the proposed system to detect the fall. Section III shows how to extract features using a genetic algorithm and the PCA, as well as the selected features for each method. After that, in Section IV the experimental design and the obtained results are presented and discussed. Finally, Section V summarizes the findings of this research and sets the future work.



1. Person Detection

Fig. 1. The computer vision system used in this work consists of the following phases: Person Detection, Characteristics extraction and Fall detection.

II. Proposed System

The computer vision system used in this work consists of the following phases (Fig. 1):

1) Person detection: Our fall detection system is based on the usage of the Microsoft Kinect depth sensor, it lets us to capture depth frames from a scene. A reference frame with no people is generated in order to help better identification of human subjects. The incoming depth frame is compared with the reference image for getting the difference cloud of points, which assumes to represent a person. The kinect [9] overcome some of the limitations of images captured by conventional cameras, for example the background subtraction, it easier to do with depth sensor and the data can be captured with different lighting conditions, without change in the captured information.
2) Feature extraction: from the cloud of points representing a person, a set of characteristics is obtained, we calculate the feature vector norm to determine if a person is falling. The features we use are the result of an extraction process using a genetic algorithm(see section III A). It is important to mention, that in real time only the selected features in Table V are extracted, in contrast, during the training stage(off-line), the twenty features in Table I are calculated in order to select some of them.
3) Fall detection: to detect falling, the system calculates the feature vector change velocity over time, the feature vector was extracted in the previous block, if the change is greater than a defined threshold then the fall is occurring, otherwise the person is performing another activity such as walking, sitting, crouching down and lying, among others.

III. Characteristics extraction and Fall detection

We are interesting in evaluating which features are selected in order to detect a fallen pose efficiently. Then we start

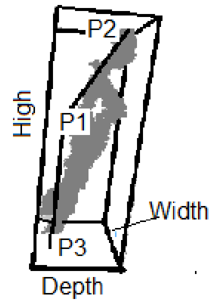


Fig. 2. Extracted characteristics from the cloud of points of the detected person.

defining a set of twenty characteristics, some of them are used in other previous works [3] [5] [7]. Table I shows the complete set and a brief description of each one. As we mentioned before, another contribution is including 3D invariant moments [8]. In our case we use three second order invariant moments to translation, scale an rotation. The second order moments describe the distribution of points in the cloud of points with respect to the coordinate axes. We consider important the inclusion of these characteristics in order to evaluate if the distribution of mass and its shape. The 3D TRS invariants ψ_1, ψ_2 and ψ_3 were computed according to the following three equations:

$$\psi_1 = \upsilon_{200} + \upsilon_{020} + \upsilon_{002}, \tag{1}$$

$$\psi_2 = \upsilon_{020}\upsilon_{002} + \upsilon_{200}\upsilon_{002} + \upsilon_{200}\upsilon_{020} - (\upsilon_{011}^2 + \upsilon_{101}^2 + \upsilon_{110}^2), \tag{2}$$

and

$$\psi_3 = \upsilon_{200}\upsilon_{020}\upsilon_{002} + 2\upsilon_{110}\upsilon_{101}\upsilon_{011} - \upsilon_{200}\upsilon_{011}^2 - \upsilon_{020}\upsilon_{101}^2 - \upsilon_{002}\upsilon_{110}^2. \tag{3}$$

Invariance to scaling and translation is achieved by the next equations:

$$\upsilon_{pqr} = \frac{\mu_{pqr}}{\mu_{000}^{\omega}} \tag{4}$$

where

$$\omega = \frac{p+q+r}{3} + 1, \tag{5}$$

and

$$\mu_{pqr} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-x_c)^p (y-y_c)^q (z-z_c)^r f(x,y,x)\,dxdydz \tag{6}$$

where *(xc, yc, zc)* is the centroid of the cloud of points.

The information to calculate all the proposed characteristics are extracted from the cloud of points of the detected person (Fig. 2), e.g., to calculate C3, we need to calculate the centroid (P1) and the highest point (P2) later to calculate the angle between P1-P2.

Once we define the set of characteristics that allows to detect a fallen pose, we want to extract the minimal subset of them in order to increase predictive accuracy of a classifier, removing irrelevant characteristics, reducing computational

TABLE I
FEATURES PROPOSALS.

Feature	Description
C1	The angle between P3 and P1 in plane X-Y
C2	The angle between P3 and P1 in plane X-Z
C3	The angle between P1 and P2 in plane X-Y
C4	The angle between P1 and P2 in plane X-Y
C5	The Angle from person respect to floor, plane XY
C6	The Angle from person respect to floor, plane YZ
C7	Ratio High/Width of the persons bounding box
C8	Ratio High/Depth of the persons bounding box
C9	Ratio Width/Depth of the persons bounding box
C10	The distance P3 to P1
C11	The distance P2 to P1
C12	The distance of the persons centroid to the floor
C13	The distance of the persons highest point to the floor
C14	The width length of the persons bounding box
C15	The height length of the persons bounding box
C16	The depth length of the persons bounding box
C17	The change the centroid position from the initial person's highest point in the sequence
C18	ψ_1 3D Invariant to TRS 1
C19	, ψ_2 3D Invariant to TRS 2
C20	ψ_3 3D Invariant to TRS 3

TABLE II
TRAINIG DATASET.

Number of frame	Pose
95	walking
55	sitting
118	falling

and storage requirements, making measurements on only those variables relevant for discrimination, providing an improved understanding of the data and the model [10] [11].

We use two methods to select the most important characteristics and we compare the results:

- Genetic Algorithm (GA): consist in a adaptive heuristic search which simulates the processes of natural selection, where competition among individuals for resources results in the fittest individuals dominating over the weaker ones. As in nature, they use selection mechanisms for mating, recombination and mutation of genetic material to evolve solutions to a given problem [12] [13]. This technique is useful when the search space is big and traditional methods fail to provide competitive solutions.
- Principal component analysis (PCA): [14] this method captures the most variable data components of samples, generates a new set of variables, called principal components, and eliminate redundant information. The main characteristic of PCA is that the first principal component contains the maximum variance of the first axis, and it is common that sum of the variances of the first few principal components to exceed 80% of the total variance of the original data.

To select the subset of characteristics to distinguish a fallen pose with respect to a non-fallen pose, we use a sequence of 268 frames, which contain several poses as we decribe in the Table II.

TABLE III
SOLUTION ENCODING FOR THE GA.

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I1	1	1	0	0	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	1
I2	1	1	0	0	1	1	0	0	0	1	0	0	0	1	1	0	0	0	1	1

Before selecting the characteristics' subset using GA or PCA, we extract the set of twenty features from every frame from the training dataset, and we classify them manually in two groups, a group which contains the frame of falls and the second one containing non-fall frames.

A. Genetic Algorithm

The GA performs the following steps

1) Create a random population with ten individuals, every individual contain twenty binary aleles, if the alele gets a 1 value the characteristic must be considered to detect fallen pose, and if gets a 0 value the feature is not present. Table III present 2 possible individuals,for example in I1 the characteristics C1, C2, C5, C6, C10, C11, C12, C13, C14, C15 and C19 are included in the subset of optimal characteristics, and in the case of I2 we only include C1, C2, C5, C6, C10, C14, C15, C19 and C20.
2) Calculate the fitness for each individual in the population. For each individual, the K-means algorithm is used to classify training frames in two groups, for which only the characteristics that have value 1 in the individual are used, Then the confusion matrix is obtained, and we calculate the percentage of correctly classified frames, and this percentage represents the individual fitness.
3) Select the best individual for elitism
4) Select individuals in order to create new population using two-point crossover, and simple mutation.
5) Apply elitism: transfer the best individual to the population for the next generation.
6) Go to step 2, repeat for 100 generations

The genetic algorithm was run 35 times. Table IV shows the statisticals results, the best fit is 99.62%, then the GA did not find a global optimum. The best set with 6 characteristics (C2, C3, C5, C7, C8, C19) is enough for discriminate a fallen pose from others. With the feature vector, we calculate its norm for each frame. In Fig. 3 the norm of fall frames are plotted by a cross '+', and all of them have a value less than 2.5. Next for all fall frames, the threshold for detecting them was calculated by averaging the feature vectors norm, it resulted in 1.7.

B. Principal Component analysis

With the 268 frames and using Matlab we perfomed the PCA analysis. The PCA reported that the first component describes 82% of the total information, and the characteristics that contribute with the maximum variance information are C7, C8, C10, C11, C13, C15 and C18. Using these characteristics and K means classify with the training frames and we obtained

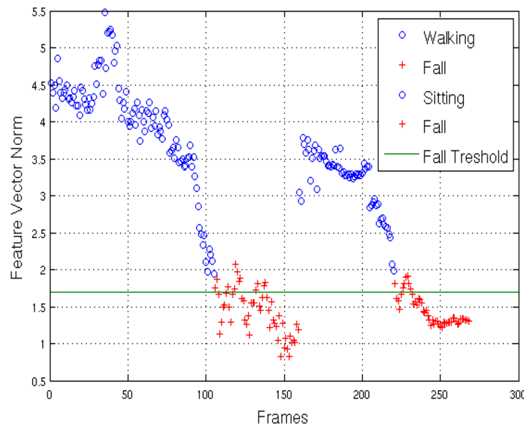


Fig. 3. For the selected features by GA, the norm is plotted for each frame, the frames that include a fallen pose are represented by a cross '+', and all of them have a value less than 2.5, then it is possible to set a threshold to differentiate a fallen pose from non-fallen pose.

TABLE IV
GA STATISTICALS RESULTS.

Metric	Value	Indivdual
Mean	98.91	
Median	98.88	
Standard Deviation	0.29	
Best Fit	99.62	0 1 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0
Worst Fit	98.13	0 0 0 0 1 0 1 1 1 0 0 1 0 0 1 1 1 0 1 1

a 78.87% classification rate. With the feature vector, we calculated its norm for each frame, in (Fig. 4) the norm of fall frames are plotted by a cross '+'. Such results indicate that there does not exist a threshold for discriminating a fallen pose from non-fallen poses.

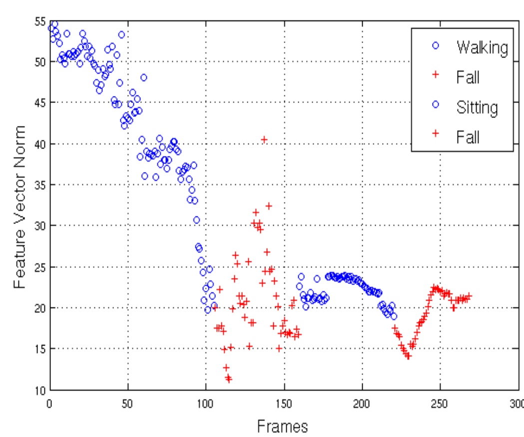


Fig. 4. For the selected features by PCA, the norm is plotted for each frame, the frames that include a fallen pose are represented by a cross '+', but it is not possible to set a threshold to differentiate a fallen pose from non-fallen pose.

TABLE V
VECTOR OF SELECTED FEATURES.

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
GA	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0
PCA	0	0	0	0	0	0	1	1	0	1	1	0	1	0	1	0	0	1	0	0

TABLE VI
GA RESULTS.

Seq	Non-Fall Frames	Fall Frames	FP	TP	FN	TN
3	166	34	3	34	0	163
4	34	62	5	62	0	29
5	95	25	5	25	0	90
6	29	71	0	62	9	29
7	97	59	1	59	0	96
8	49	42	3	42	0	46
9	131	53	0	50	3	131
10	58	72	13	72	0	45
TOTAL	659	418	30	406	12	629

IV. EXPERIMENTS AND RESULTS

Table V, shows that the GA selected 6 features, three of them are related with angles in plane X-Y and X-Z, the other two with ratio of aspects of bounding box, and the last one is the second moment invariant. On the other hand, the PCA selected 7 characteristics: the same two ratio aspects selected features by the GA, others 4 are distances related with P1, P2 and P3, and the last one is the first moment invariant. As we can see, each method selected a different set of features, which allow us to assess what impact have different sets in the accuracy of the classifier of falls. For this purpose, three experiments were designed, in which we used several sequences from a public fall's datasets in [7]. This repository contains sequences of 30 falls, fall events were recorded with 2 Microsoft Kinect cameras. In our case, we only used the dataset recorded with camera 0, which is parallel to the floor, and 40 sequences of Activities of Daily Living (ADL) recorded with camera 0 parallel to the floor. Depth data is stored in PNG16 format, then we converted them to PCD format using PCL libraries. After that, we segmented the person manually, because we didn't find a frame with a scene without person to use it as a reference frame.

A. *First experiment*

In the first experiment, we used 8 sequences from fall, two types of falls were performed, namely from standing position and from sitting on the chair, we selected 4 for each type. We used these sequence for measuring the performance of the classifier.

To measure the performance we calculated the following metrics [15]:

- True Positive (TP): Fallen pose was detected correctly
- False Positive (FP): Fallen pose was detected incorrectly
- True Negative (TN): A non-fallen pose was detected correctly
- False Negative (FN): A non-fallen pose was detected as fallen pose

TABLE VII
PCA RESULTS.

Seq	Non-Fall Frames	Fall Frames	FP	TP	FN	TN
1	66	34	0	34	92	74
3	166	34	92	34	0	74
4	34	62	2	62	0	32
5	95	25	3	25	0	92
6	29	71	0	59	12	29
7	97	59	0	59	0	97
8	49	42	0	39	3	49
9	131	53	0	50	3	131
10	58	72	2	72	0	56
TOTAL	659	418	99	400	18	560

TABLE VIII
PERFORMANCE MEASURES.

	GA	PCA
Sensitivity	97.13	95.69
Specificity	95.45	84.98
Accuracy	96.10	89.14
Error rate	3.90	10.86

- Sensitivity (Se): the capacity to detect falls correctly

$$Se = TP/(TP + FN) \tag{7}$$

- Specificity (Sp): the capacity to detect non-falls correctly

$$Sp = TN/(TN + FP) \tag{8}$$

- Accuracy (Ac): the correct classification rate

$$Ac = (TP + TN)/(TP + TN + FP + FN) \tag{9}$$

- Error rate (ER): the incorrect classification rate

$$ER = (FP + FN)/(TP + TN + FP + FN) \tag{10}$$

In tables VI and VII, the results for detecting a fallen pose by using the features suggested by the GA and PCA are presented, respectively. Each table is organized by seven columns, the first one represents the number of fall sequence, it was taken from the repository in [7], for our purposes, in every sequence are identified two kind of frames: non-fall frame and fall frame; and the quantity of each one of them in every sequence is showed in the columns two and three, finally, the last four columns represent the results of evaluating the metrics: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), which let us to calculate the Sensitivity, Specificity, Accuracy and Error rate using the selected feature by GA and PCA, the obtained results are in table VIII. The results using the selected characteristics by GA are better than those obtained by PCA.

Fig. 5 and Fig. 6 show the results of sequence 3 using the selected characteristics by the GA and by PCA respectively, using GA, it can see the fall can detected using the threshold 1.7, which was calculated at the final of section A. In contrast with PCA it is not possible to identify a threshold, i.e.,the fall occurs in a value close to 15, but in the next frames it value increases in average 20.

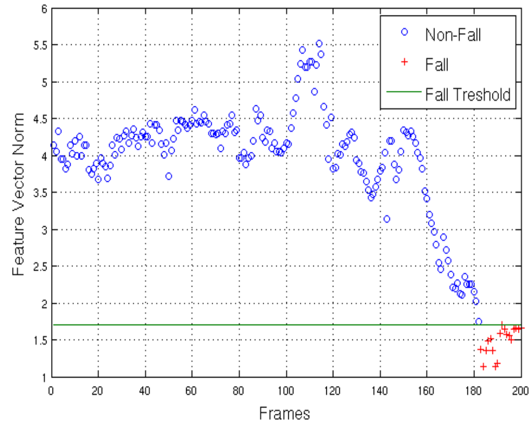


Fig. 5. GA chosen features results in sequence 3, the person is walking and fall, it can be seen for all fall's frame the value of norm is less than 1.7.

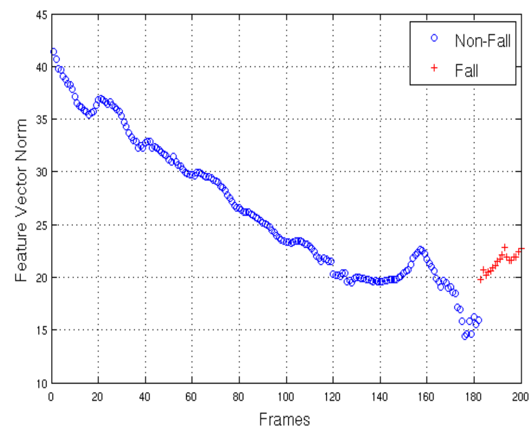


Fig. 6. PCA chosen features results in sequence 3, the person is walking and fall, using the vector norm it is no possible to distinguish fallen pose from others.

TABLE IX
ADL SEQUENCES.

Sequence	Description
1	The person is walking and crouch down
4	The person is walking and stops to pick up an object
14	The person is walking and bends down to pick up an object
32	The person is walking and lie down on the floor
40	The person is walking and lie down on the floor

B. Second experiment

In the second experiment, we selected 5 sequences from ”Activity daily living” (ADL), that include postures like crouching down, picking up an object, lying on the floor, in order to see if it possible to recognize these postures as non-fall, although they were not considered in the training set. The detail is in Table IX, the firts column represents the number of ADL sequence, it was taken from the repository [7], and the second column is a description of the action, which was performed in the sequence.

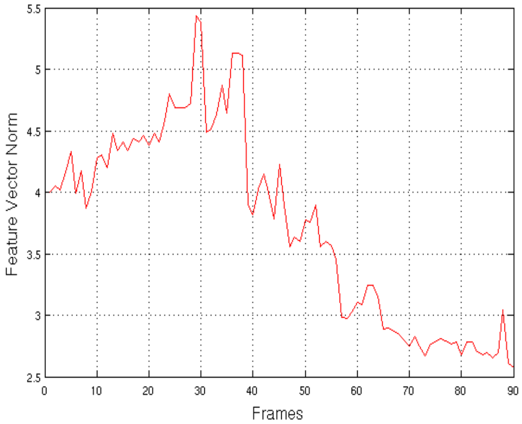


Fig. 7. In this sequence the person is walking and crouch down, using the selected features by the GA, a fall never happens, because the norm vector is higher than 1.7.

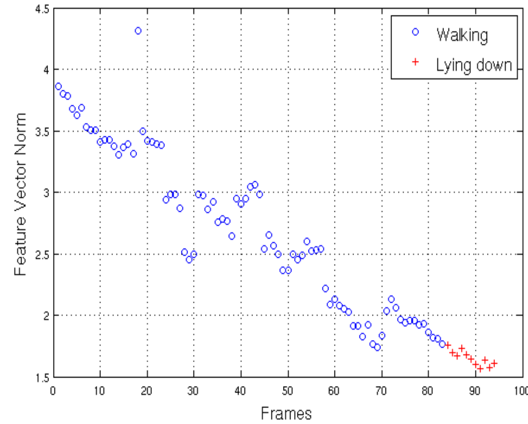


Fig. 9. The person is walking and lie down on the floor, using the selected features by GA, a fall is detected, because the norm vector is less than 1.7, in this case we have to consider another parameter to differentiate a fall event that occur involuntarily from another similar that occur in control of our movements.

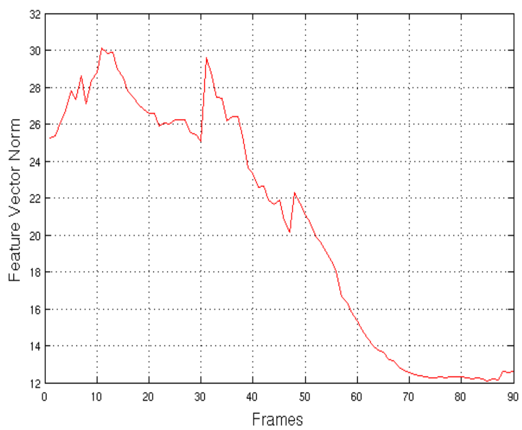


Fig. 8. In this sequence the person is walking and crouch down, using the selected features by PCA, it is possible that a false fall is detected, which is incorrect.

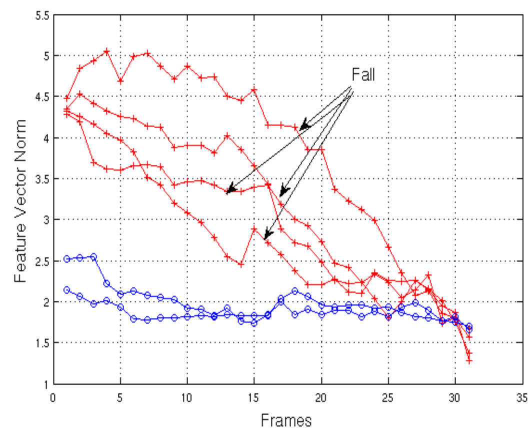


Fig. 10. Feature vector norm for fall events is shown in the line with crosses, and in the line with circles when a person is lying down on the floor

For the first sequence, in Fig. 7 and Fig. 8 the results using feature vector found by the GA and PCA are shown, respectively. In Fig. 7 it can be seen the norm vector value is higher than 1.7, then a fallen pose is never detected, the result is correct, but in Fig. 8 the norm vector gets values close to 12 and if it is compared with Fig. 6 then it could be detected as a false negative.

C. *Third experiment*

The third experiment was designed to test, if the change of the feature vector over time allows to detect a fall.

With the sequence 40, the selected features by the GA were adopted, when a person is lying down on the floor is plotted by a cross '+' in Fig. 9. In order to differentiate a fallen pose of another similar event, it is important remember a fall event happens in a involuntary way, however when a person is lying down on the floor is in control of his movements. In Fig. 10 the norm for fall events is shown in the line with crosses, and in the line with circles when a person is lying down on the floor. In the former, it can see the vector change velocity is higher from the second. Therefore it was measured the feature vector change velocity over time, considering that the fall happens in 1 or 2 seconds and the adquisicion of images was 30 fps, we experimentally selected if the norm vector change rate is less -1.0 every 30 frames and the feature vector norm is less than 1.7 a fall event is happen. All 13 sequences in tables VI and IX were tested using the thresholds selected, and only in the sequences in table VI a fall event was detected.

V. Conclusion and Future Work

In our work we evaluated which of those feature combined with others not considered in previous works like 3D TRS moments invariants and others geometric characteristics like angles or distances were important to distinguish a fallen pose, and how different subsets of features affected the sensitivity, specificity of the classification. We tested 2 methods for selecting the optimal features: GA and PCA, each of them selected a different subset of characteristics. The result was that the GA gave us better results, the selected feature vector let us to establish a threshold to differentiate fallen pose from non-fallen pose. With PCA we obtained less specificity, i.e., more false positive, which could activate an alarm when a fall does not happen. Based on the obtained results we showed that the selected features affect the sensitivity, specificity, accuracy and error of the classification. With the GA we obtained better performance, which suggest it as a good option for selecting an optimal subset of features in classification aplications. Furthermore PCA and GA selected at least one 3D TRS moment invariant for detecting fallen pose. Thus they proved to capture important information for the classification. The future work consist in evaluating the selected feautures by the GA in real time and extending our approach for action recognition, using skeletal joints and cloud of points information like invariants.

References

[1] J. R. S. Fhon, S. C. C. Fabrício-Wehbe, T. R. P. Vendruscolo, R. Stackfleth, S. Marques, and R. A. P. Rodrigues, “Caídas en el adulto mayor y su relación con la capacidad funcional,” *Rev. Latino-Am. Enfermagem [Interne]*, vol. 20, no. 5, Oct. 2012.

[2] M. del Carmen and J. Manuel, “Guía de práctica clínica para la prevención de caídas en el adulto mayor,” *Rev Med Inst Mex Seguro Soc*, vol. 43, no. 5, pp. 425–441, 2005.

[3] C. Rougier, A. St-Arnaud, J. Rousseau, and J. Meunier, *Video surveillance for fall detection*.

[4] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, and J. Meunier, “Fall detection from depth map video sequences,” in *Toward useful services for elderly and people with disabilities*. Springer, 2011, pp. 121–128.

[5] V. D. Nguyen, M. T. Le, A. D. Do, H. H. Duong, T. D. Thai, and D. H. Tran, “An efficient camera-based surveillance for fall detection of elderly people,” in *Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on*, June 2014, pp. 994–997.

[6] Z.-P. Bian, J. Hou, L.-P. Chau, and N. Magnenat-Thalmann, “Fall detection based on body part tracking using a depth camera,” *Biomedical and Health Informatics, IEEE Journal of*, vol. 19, no. 2, pp. 430–439, 2015.

[7] B. Kwolek and M. Kepski, “Human fall detection on embedded platform using depth maps and wireless accelerometer,” *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 489–501, 2014. [Online]. Available: http://fenix.univ.rzeszow.pl/ mkepski/ds/uf.html

[8] F. Jan, S. Tomáš, and Z. Barbara, “Moments and moment invariants in pattern recognition,” *Chippenham, UK: Wiley & Sons Ltd*, 2009.

[9] L. Chen, H. Wei, and J. Ferryman, “A survey of human motion analysis using depth imagery,” *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1995–2006, 2013.

[10] A. R. Webb, *Statistical pattern recognition*. John Wiley & Sons, 2003.

[11] L. Yu and H. Liu, “Efficient feature selection via analysis of relevance and redundancy,” *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, Dec. 2004. [Online]. Available: http://dl.acm.org/citation.cfm?id=1005332.1044700

[12] C. A. C. Coello and C. S. P. Zacatenco, “Introducción a la computación evolutiva,” *Ediciones CINVESTAV-IPN. Instituto Politécnico Nacional, México*, 2003.

[13] D. E. Goldberg, *The design of innovation: Lessons from and for competent genetic algorithms*. Springer Science & Business Media, 2013, vol. 7.

[14] J. A. Richards and J. Richards, *Remote sensing digital image analysis*. Springer, 1999, vol. 3.

[15] A. G. Lalkhen and A. McCluskey, “Clinical tests: sensitivity and specificity,” *Continuing Education in Anaesthesia, Critical Care & Pain*, vol. 8, no. 6, pp. 221–223, 2008.