Complex human action recognition on daily living environments using wearable inertial sensors

Irvin Hussein López-Nava
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro 1, Tonantzintla
Puebla, Mexico
hussein@inaoep.mx

Angélica Muñoz-Meléndez
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro 1, Tonantzintla
Puebla, Mexico
munoz@inaoep.mx

ABSTRACT

The aim of this study is to evaluate how similar a set of human actions are when they are performed under controlled conditions versus the same set of actions when they are performed under uncontrolled conditions, namely in daily living environments such as the users' houses. This research is important for automatic recognition of human actions in daily living environments, mainly using wearable sensors, which is still an open research challenge of the field of pervasive computing. Action recognition is important for various applications of this field, such as, for instance, ambient intelligence, smart devices, and healthcare. In this work, we measure and analyze five human complex actions using wearable sensors in both, structured and daily living environments. Three wearable inertial sensor units were used in this study and they were worn by three healthy young subjects on three points of their upper limbs: the scapula, the upper arm and the forearm. The complex actions involved in this study are: grooming, cooking, eating, doing housework, and mouth care. Dynamic Time Warping algorithm was used to measure the intra and inter test variability of actions in both environments. Additionally, the results of the application of three supervised classification techniques, namely C4.5, Naïve Bayes and Logistic Regression, are compared in terms of true positive rate TPR, true negative rate TNR and F-measure metrics. The classification models were based on time-domain and frequency-domain features extracted from orientation signals. According to the analysis cooking and eating are the actions with highest and lowest variability, respectively. Concerning the classification results, Naïve Bayes and Logistic Regression obtain a TPR of 0.911 using relevant attributes. Our results provide valuable information to measure the similarity of a set of complex actions in daily living environments and to classify them.

Keywords

Action recognition; action classification; action variability analysis; inertial sensors; wearable sensors

.

1. INTRODUCTION

Automatic recognition of human actions in daily living environments, mainly using wearable sensors, is still an open research challenge of the field of pervasive computing [9]. Action recognition is important for various applications of this field, such as, for instance, ambient intelligence, smart devices, and healthcare [18, 8]. In effect, action recognition aims at providing information about behavior and intentions of users that enable computing systems to assist users proactively with their tasks [6].

However, there are number of reasons why human action recognition is a very challenging problem. Firstly, human body is non rigid and has many degrees of freedom [21]. Also, human body can perform infinite variations for every basic movement. Most existing work on action recognition is built upon simplified structured environments, normally focusing on single-user single-action recognition. In real world situations, human actions are often performed in complex manners. That means that a person performs interleaved and concurrent actions and he/she can interact with other person(s) to perform joint actions, such as cooking [9].

Recently, many research studies have been done to analyze both, simple and complex human actions based on wearable sensors. A large number of these studies focus on identifying which are the most informative features that can be extracted from the actions data as well as in searching which are the most effective machine learning algorithms for classifying these actions [25]. Sensors attached to human anatomical references, *i.e.*, wearable sensors often comprise inertial measurement units (accelerometers, gyroscopes and magnetometers), vital sign processing devices (heart rate, temperature) and RFID tags, can be used to gather information about the behavioral patterns of a person. Robustness to occlusion and to lighting variations, as well as portability are the major advantages of wearable sensors over visual motion-capture systems. Additionally, the visual motion-capture systems require very specific settings for properly operating [2].

When compared to approaches based on specialized systems, the wearable sensor based approach is effective and also relatively inexpensive for data acquisition and action recognition for certain types of human actions, mainly human movements involving the upper and lower limb motions [2]. Actions such as walking, running, sitting down/ up, climbing or physical exercises, are generally characterized by a distinct, often periodic, motion pattern [9].

PervasiveHealth 2016, May 16-19, Cancun, Mexico

DOI 10.4108/eai.16-5-2016.2263332

Nevertheless, action recognition using wearable sensors suffers from two drawbacks. First, most wearable sensors are not suitable for real daily living environments due to technical issues related to the sensors' size and weight, and battery life, in conjunction with the general issue of acceptability or willingness of wearing them during long periods that last, for instance, several hours. Second, many actions in real world situations involve complex physical motions and complex interactions that happen within daily living environment. Sensor readings from wearable sensors alone may not be able to distinguish complex actions involving simple interactions with the environment, *e.g.*, making tea or making coffee [9].

In this research an action is defined as the human movement voluntarily performed to achieve an aim. In this regard, a movement or motion can be described as the displacement of an object in space over time. In this research, an action differs from a gesture, since the latter is the meaning being expressed through an action or motion. We define two types of human actions of the upper limbs according to the degrees of freedom of the human segments that are involved and to the type of movement that is made, as explained below.

`Simple actions` are actions involving movement of the anatomical segments restricted to one degree of freedom, and whose trajectory of movement is predefined, *e.g.*, touching the opposite shoulder with a hand. Simple actions are usually related to therapeutic movements performed by patients who are told by rehabilitation specialists how to perform these movements with precise instructions. `Complex actions` are actions involving movement of the anatomical segments with two or more degrees of freedom, and whose trajectory of movement is not predefined explicitly, therefore these actions can be performed differently each time, *e.g.*, brushing teeth. Complex actions are related to the activities of daily living, activities in which people engage on a day-to-day basis.

For this study five complex actions were selected from the common daily activities of three test subjects: grooming, cooking (preparing a sandwich), eating, doing housework (cleaning a table), and mouth care (brushing teeth). The aim of this study was to measure and evaluate complex human actions performed by three test subjects in both structured environments under controlled conditions, and daily living environments with non-controlled conditions, using three wearable inertial sensors placed on the dominant upper limb of the test subjects. And conduct a variability analysis of these complex human actions, and study if they can be automatically classified.

The rest of the paper is organized as follows. Section 2 addresses related work. Section 3 presents the experimental setup and the computational techniques used in this study. Section 4 gives details about the variability analysis and classification results. Finally, Sections 5 and 6 close with concluding remarks about the results, challenges and opportunities of this study.

2. RELATED WORK

In recent years, there has been an increasing interest in designing and implementing methods, algorithms, and wearable inertial sensors for automatic action recognition [6].

The action recognition based on wearable inertial sensors depends mainly on the sensor signal to be processed and the number of sensor units fixed to the human anatomical references. In most studies analyzed in this research, linear acceleration [15] or a combination of linear acceleration and angular velocity [4, 23, 22, 10, 24, 11] are used to extract signals features for characterizing human movements. In other cases, the fusion of inertial sensor information with video-cameras was explored [7]. The number of sensor units fixed to a particular anatomical reference is primarily one [7, 4, 23, 10, 24, 11] or two [15, 22].

The action recognition process can be divided into two tasks: data preprocessing and data analysis. Data preprocessing consists in selecting and preparing data for further analysis. Data analysis can be made evaluating the time-series signals [22], or using representative time-domain [7, 4, 23, 10, 11] or a combination of time-domain and frequency-domain features [15, 24]. Then, unsupervised [4, 24], such as k-means clustering, and supervised [7, 4, 15, 23, 22, 10, 24, 11], such as support vector machines classifier, approaches are used for classification purposes.

Additionally, the experimental setting has two important elements to consider. The first one is related to the type of human action to recognize: simple [4] and complex [7, 15, 23, 22, 10, 10, 24, 11] actions. It is important to remark that most of the complex actions studied in the related work are associated with movements of both upper and lower limbs. The second important element of the experimental setting to highlight is related to the environment in which the actions are performed, structured environments under controlled conditions [7, 4, 15, 23, 22, 24, 11], daily living environments under controlled conditions [15, 10], and daily living environments with uncontrolled conditions.

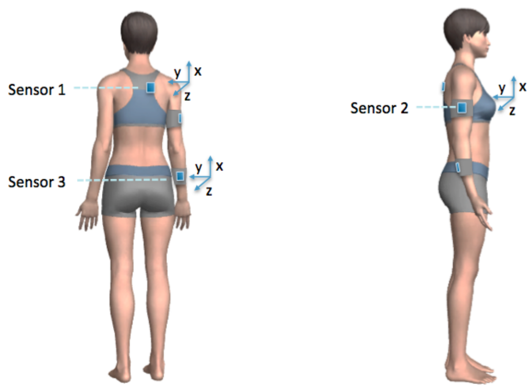
In contrast to previous work, three important aspects of our study can be highlighted: (1) two computational techniques are presented for variability evaluation and for action classifying; (2) orientation signals were obtained from three wearable inertial sensors placed on the upper limbs of test subjects; (3) test subjects were asked to perform five complex actions in their daily living environments with uncontrolled conditions. As a matter of fact, there is a lack of research about the third type of environment, presumably because of the difficulties to acquire data "in the wild".

3. METHODS

3.1 Setup

Three asymptomatic young adults (mean age 27.3 $\pm$ 3.5 years, from now on referred as $S1$, $S2$ and $S3$), were asked to wear three LPMS-B sensors (LP-research, Japan). They have signed an informed consent. LPMS-B sensor is a miniature wearable inertial measurement unit (IMU) / attitude and heading reference system (AHRS). This device includes three different MEMS sensors: 3-axis gyroscope, 3-axis accelerometer and 3-axis magnetometer. Its communication distance scope is $18m$, it has a lithium battery of $3.7V$ at $800mAh$, and it weights $34g$. This sensor provides Euler angles as orientation signals calculated using an extended Kalman filter.

Sensors were placed on three anatomical references of the right upper limbs of the test subjects, in all cases of the dominant limb. The first anatomical reference was located in the center of the subscapular fossa. A sensor, $Sen1$, was settled over this anatomical reference using an orthopedic vest. The second anatomical reference was located 10 cm up



(a) Posterior view in the coronal plane

(b) Right view in the sagittal plane

Figure 1: Configuration of the three wearable inertial sensors on the upper limb of subjects. Subfigure 1(a) shows the posterior view of a person and illustrate the coordinated system of sensors 1 and 3, Subfigure 1(b) shows a lateral view of a person and illustrate the coordinated system of sensor 2.

to the right elbow joint, in the lateral side of the upper arm, aligning the plane formed by the x-axis and z-axis of $Sen1$ with the sagittal plane of the subject. The third anatomical reference was located 10 cm up to the right wrist joint, in the posterior side of the forearm, aligning the plane formed by the x-axis and z-axis of $Sen1$ with the coronal plane of the subject. The sensors $Sen2$ and $Sen3$ were firmly attached to the second and third anatomical references using elastic velcro straps. The configuration and coordinated systems of the sensors are shown in Figure 1.

The movement of the segments is limited to the range of motion (ROM) allowed by the adjacent joints [16]. The scapula segment rotates around the sternoclavicular joint, and the upper arm segment rotates around the glenohumeral joint, both are joints of the complex shoulder joint. The forearm segment rotates around the elbow joint. The execution of the studied complex actions depended on the ROM anatomically permitted for these upper limb joints of each person.

3.2 Environments

Two type of environments were arranged to study the five complex actions described previously. In the first environment, referred to as structured environment, the conditions were controlled. This environment contains a bathroom, and an area that was divided into a dining room and a kitchen. The furniture, kitchen tools, and objects in the environment were located in the same starting position for each test. The second type of environment, referred to as daily living environment, was each one of the houses of the test subjects and the conditions in there were not controlled. The daily living environments used in our experiments consisted of 2 and 3 rooms. The position of furniture, kitchen tools and other objects in the environment was not predefined.

The five complex actions were executed in the structured environment in the order presented in Figure 2 and they were repeated three times. However, in the daily living environments the actions were freely performed by the subjects according to their daily habits without any particular order and until the actions were repeated three times. Before the recording of a test in the structured environment, test subjects received clear instructions about the utensils and furniture that could be used to perform each action. On the other side, during test recording in daily living environments, test subjects were able to behave arbitrarily to complete the actions.Three video cameras were installed in the environments in order to simultaneously record the actions performed by the subjects. Video recordings were used as basis for manually segmentating the sensor signals and for labeling the complex actions. In Figure 2 a subject performing the five actions in both the structured environment 2(a) and his house 2(b) is shown. The structured environment was adapted with utensils and furniture needed for performing the actions within the three rooms, on the contrary, in the everyday environment the number and type of utensils, furniture and rooms were not controlled at all.

3.3 Multidimensional Dynamic Time Warping

A distance was measured from orientation signals using Dynamic Time Warping (DTW) algorithm in order to compare the variability of the studied actions. DTW is an algorithm capable of measuring similarity between two temporal sequences that vary over time [17]. DTW matches two time signals by computing a temporal transformation aligning the signals. This alignment is optimal in the sense that a cumulative distance measure between the aligned samples is minimized [13]. For each pair of segmented signals, X and Y, of length m and n, respectively, where $X = \{x_1, x_2, ..., x_i, ..., x_m\}$ and $Y = \{y_1, y_2, ..., y_j, ..., y_n\}$ the calculation of a matching cost $DTW(X, Y)$ is required. The matching cost was computed based on dynamic programming using Formula 1.

$$D(i,j) = d(x_i, y_i) + min \begin{Bmatrix} (i-1, j-1) \\ (i-1, j) \\ (i, j-1) \end{Bmatrix} \tag{1}$$

where the distance function $d(\cdot,\cdot)$ is the Manhattan distance (L_1-$norm$) used as cost measure as defined in Formula 2.

$$d(x_i, y_j) = |x_i - y_j| \tag{2}$$

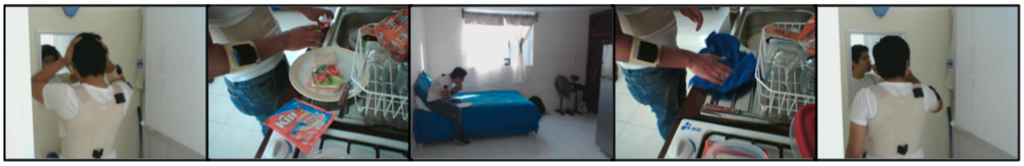
and consequently, the dynamic time warping distance between a pair of signals is $DTW(X, Y) = D(m, n)$.

DTW is applicable to only one-dimensional time signals, for multidimensional alignment it is necessary to extend DTW algorithm. For this purpose, DTW_3 distance represents the cumulative distances of the three axis independently measured using DTW as indicated in Formula 1. If $DTW(Q_m, C_m)$ is defined as the DTW distance of the m^{th} axis of Q and the m^{th} axis of C, written as DTW_3, it can be calculated using Formula 3.

$$DTW_3(Q, C) = \sum_{m=1}^{M} DTW(Q_m, C_m) \tag{3}$$

where Q and C are 3D-time signals and correspond to Euler angles, for $M = 3$.

(a) Structured environment



(b) Daily living environment

Figure 2: Complex actions performed by a subject in the structured environment 2(a) and in his daily living environment 2(b). The actions from left to right are: grooming, cooking (preparing a sandwich), eating, doing housework (cleaning the table) and mouth care.

3.4 Feature extraction

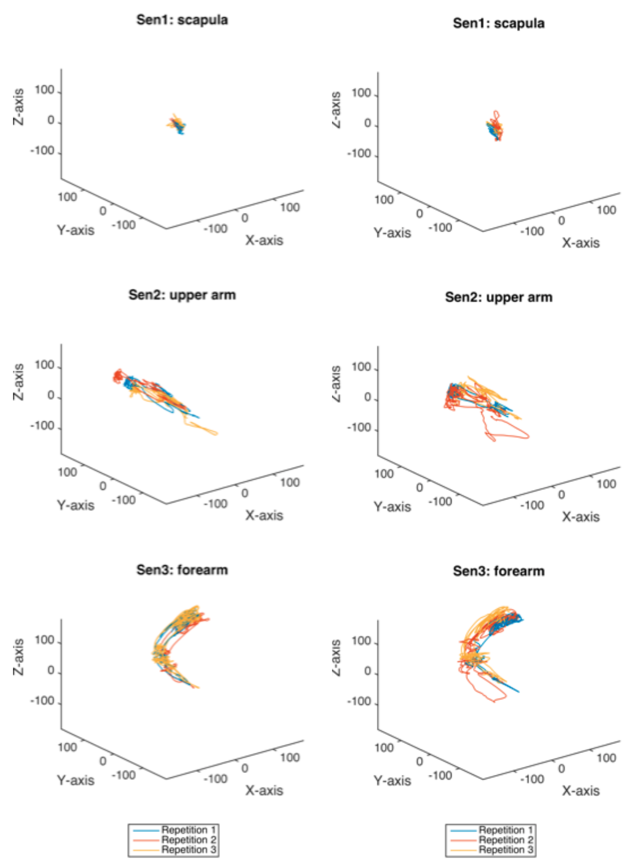
Time-domain and frequency domain features were calculated for each signal in order to classify the studied actions. Time-domain features selected were: action completion time (act), arithmetic mean, standard deviation, range of motion, zero crossing rate, and root mean square. Frequency-domain features selected were power of the signal in frequencies: 0-2 Hz, 2-4 Hz, and 4-6 Hz. Each attribute was calculated for each Euler angle signal captured from the three sensors, except act feature that was calculated from values captured from all sensors and signals. These features were selected according to their relevance in the area [5, 14, 3]. For a detailed description of these features see Appendix A.

4. RESULTS

In Figure 3 the trajectories of the Euler angles of the three sensors for the grooming action performed by a subject in both environments are shown. The trajectories of the three sensors during all the repetitions used for intra and inter environment data have the same characteristics, but a difference of magnitude. As can be noted, the magnitude of the trajectories recorded with the sensor placed on the scapula have the lowest values in comparison to the recordings of the sensors placed on the upper arm and the forearm. Graphics of the forearm sensor present the same visual patterns in both environments. In the next Subsections, a detailed analysis to quantify the difference or variability among actions performed in both environments using DTW distances is presented. Also, a classification of the actions using the time-domain and frequency-domain features is given.

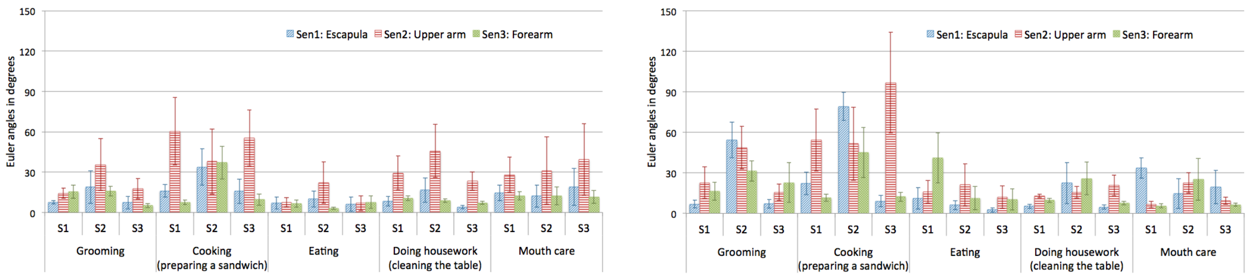
4.1 Variability analysis

DTW_3 distances were calculated from Euler angle signals to obtain the variability of the performed complex actions by test subjects in both environments. In Figure 4 individual results of the variability analysis are shown. According to this Figure, subject $S2$ presents the highest variability in both environments for all complex actions in general. In both environments, the action with the greatest variability is cooking, and the action with the greatest similarity is eating. The sensor with the greatest variability is the sensor placed on the upper arm.



(a) Structured environment (b) Daily living environment

Figure 3: Spatial trajectories of the grooming action performed by a subject in the structured environment 3(a) and in his daily living environment 3(b) from Euler angle signals.



(a) Structured environment

(b) Daily living environments

Figure 4: Individual DTW_3 distances of the five complex actions performed by the three subjects in the structured environment 4(a) and their daily living environments 4(b).

Tables 1 and 2 summarize the mean variability of the three subjects by sensor for the structured environment and the daily living environments, respectively. As can be noted, the sensor placed on the upper arm presents the greatest mean variability in both environments, it has also the highest dispersion. Concerning the actions, cooking presents the greatest variability in both environments, and this variability is even higher in daily living environments. The second action with greater variability in the structured environment was mouth care, whereas grooming was the action with the greatest variability in daily living environments. Finally, the actions with less variability are cooking in the structured environment and doing housework in daily living environments, additionally the last action has the lowest inter subject dispersion.

4.2 Complex action classification

With the goal of recognizing complex actions, the time-domain and frequency-domain features presented in Subsection 3.4 were extracted from the orientation signals. Then three classification algorithms were selected for building models based on previous features: J48 classifier, Naïve Bayes and Logistic regression models. Naïve Bayes, a statistical classifier, has proven to be successful and simple for classification purposes[20]. This algorithm is based on the Bayes' theorem that uses the joint probabilities of sample observations to estimate the conditional probabilities of classes given an observation [12]. J48 is an implementation of the C4.5 decision tree classifier [19], that builds a binary classification tree. Determined by a splitting criterion, attributes are selected as branching points that separate the two classes in the training dataset. Finally, Logistic Regression is a generalized linear model to apply regression to categorical variables. Generalizations of logistic regression apply to multi category responses and assume a multinomial distribution [1]. These algorithms are commonly used in the field of machine learning and for this reason they were selected for initiating this research.

The classification was performed dividing data into datasets registered in structured and daily living environments. In order to select more relevant attributes, gain information algorithm was applied for ranking attributes. This algorithm evaluates the worth of an attribute by measuring the information gain with respect to the class, in our case an action. True positive rate TPR and true negative rate TNR metrics were obtained from the classification results.

Sensitivity is the True Positive Rate (TPR), also called Recall, and measures the proportion of positives which are correctly identified as indicated in Formula 4.

$$TPR = \frac{True\ positives}{True\ positives + False\ negatives} \tag{4}$$

Specificity is the True Negative Rate (TNR) and measures the proportion of negatives which are correctly identified as indicated in Formula 5.

$$TNR = \frac{True\ negatives}{True\ negatives + False\ positives} \tag{5}$$

In Table 3 the classification results are shown. Columns describe, from left to right, the data treatment, the number of instances, the number of attributes, and the calculated value of the metrics. Except for the treatment in which a model is trained and evaluated with different data partitions, in all treatments a 10 cross-fold validation was used.

According to the classification results, for the three classifiers the rate of true negatives is greater than the rate of true positives in all treatments. In general J48 classifier has the poorest classification results, except for data of the daily living environments. It is also worth to remark that classification based on a combination of data of both environments, structured and daily living environments, is better than classification based on any separate dataset. For Naïve Bayes and Logistic regression, which consider all attributes to build their classification models, selecting relevant attributes improves the rate of true positives 2.4% and 9.8%, respectively.

In Tables 4-6 the confusion matrices for the treatment that includes relevant attributes and data of structured and daily living environments are shown. J48 algorithm misclassified 4 instances of grooming as doing housework and 3 instances as mouth care; 5 instances of cooking as mouth care and 3 instances of mouth care as cooking, were misclassified too. Naïve Bayes misclassified 2 instances of cooking as grooming and 2 instances of eating as cooking. Logistic regression misclassified 2 instances of cooking as eating.

5. DISCUSSION

According to variability study of complex actions, we can make this remark: the sensor that was placed on the upper arm, $Sen2$, presents greater variability of the orientation recordings for the all actions that were performed, than the sensor placed on the forearm, $Sen3$, 18.9 ± 10.7 deg of difference in the structured environment, and 11.4 ± 18.6 deg of difference in daily living environments; except for eating action in daily living environments in which the variability

Table 1: Means of DTW_3 distances of the three subjects performing the complex actions in structured environment expressed in degree units, sd means standard deviation.

Action	Sen1: scapula	Sen2: upper arm	Sen3: forearm	Mean of action (sd)
Grooming	11.3	22.6	12.1	15.3 (6.3)
Cooking	21.9	51.2	18.1	30.4 (18.1)
Eating	7.8	12.2	5.8	8.6 (3.3)
Doing housework	9.7	32.8	8.9	17.1 (13.6)
Mouth care	15.3	32.9	12.1	20.1 (11.2)
Mean of sensor (sd)	13.2 (5.6)	30.3 (14.5)	11.4 (4.6)	

Table 2: Means of DTW_3 distances of the three subjects performing the complex actions in daily living environments expressed in degree units, sd means standard deviation.

Action	Sen1: scapula	Sen2: upper arm	Sen3: forearm	Mean of action (sd)
Grooming	22.8	29.0	23.5	25.1 (3.4)
Cooking	36.8	67.5	23.0	42.4 (22.8)
Eating	6.5	16.4	20.9	14.6 (7.3)
Doing housework	10.8	16.4	14.3	13.8 (2.8)
Mouth care	22.6	12.9	12.3	15.9 (5.8)
Mean for sensor (sd)	19.9 (11.9)	28.4 (22.7)	18.8 (5.1)	

Table 4: Confusion matrix for J48 classification performance using relevant attributes and instances of both environments.

		Action classified as:				
Num.	Action	1	2	3	4	5
1	Grooming	11	0	0	4	3
2	Cooking	1	10	2	0	5
3	Eating	0	1	17	0	0
4	Doing housework	0	0	0	18	0
5	Mouth care	0	3	0	0	15

Table 5: Confusion matrix for Naïve Bayes classification performance using relevant attributes and instances of both environments.

		Action classified as:				
Num.	Action	1	2	3	4	5
1	Grooming	17	1	0	0	0
2	Cooking	2	16	0	0	0
3	Eating	0	2	16	0	0
4	Doing housework	1	0	0	16	1
5	Mouth care	1	0	0	0	17

Table 6: Confusion matrix for Logistic regression classification performance using relevant attributes and instances of both environments.

		Action classified as:				
Num.	Action	1	2	3	4	5
1	Grooming	16	1	0	0	1
2	Cooking	1	14	2	1	0
3	Eating	0	1	17	0	0
4	Doing housework	1	0	0	17	0
5	Mouth care	0	0	0	0	18

of $Sen3$ is greater than $Sen2$. That indicates that the movements of the upper arm around the glenohumeral joint, in particular the movements related with the flexion/extension, abduction/adduction, and internal and external rotation of the upper arm enable the execution of the studied actions in a greater number of configurations than the movements involving sternoclavicular and the elbow joints.

The orientation variability of the sensors placed on the forearm and on the scapula are equivalent for all the actions, $Sen3$, 2.1 ± 1.3 *deg* in the structured environment, and $Sen3$ and 8.5 ± 6.2 *deg* in daily living environments between both sensor recordings.

Concerning to the actions, eating is the action with lowest variability in the structured environment, 8.6 *deg*, and it has the lowest standard deviation among the three sensors with 3.3 *deg*. Eating, doing housework (cleaning the table) and mouth care present the greatest similarity in daily living environments, 14.8 *deg* on average, indicating that these actions have motion characteristics that are usually repeated each time they are performed; cleaning the table has the lowest standard deviation of sensors in both environments with 2.8 *deg*. Cooking is the action with highest variability in both environments, 2 and 2.5 times higher in proportion than the average of the other four actions. Cooking is presumably the action less repeatable, in this study this action was preparing a sandwich. Finally, doing housework and mouth care present lower variability in daily living environments than in the structured environment, 3.3 *deg* and 4.2 *deg* less, respectively; however the standard deviation is greater for the structured environment. The mean difference of the other three actions, that have greater variability in the structured environment than in daily living environments, is 9.3 ± 3 *deg*.

Regarding to the complex action classification, the results obtained by Naïve Bayes classifier are more consistent than the classification results using the other two classifiers. It is important to remark that J48 classifier, that is an implementation of C4.5, builds a decision tree measuring the

Table 3: **Classification results using J48, Naïve Bayes and Logistic regression classifiers according to data treatments. StDa: instances of the structured environment, DLDa: instances of daily living environments.**

Data treatment	Instances	Attributes	J48		Naïve Bayes		Logistic reg.	
			TPR	TNR	TPR	TNR	TPR	TNR
Structured data StDa	45	73	0.778	0.994	0.822	0.956	0.844	0.961
Daily living data DLDa	45	73	0.800	0.950	0.778	0.944	0.644	0.911
Training with StDa, testing with DLDa	90	73	0.667	0.917	0.773	0.933	0.778	0.944
StDa and DLDa	90	73	0.789	0.947	0.889	0.972	0.822	0.956
StDa and DLDa with relevant attributes	90	44	0.789	0.947	0.911	0.978	0.911	0.978

information entropy of the attributes and uses the more relevant ones for the model. For the proposed treatments, J48 classifier uses only between 4 and 6 attributes for building its classification models. The information gain algorithm selects the most relevant attributes to reduce the number of attributes and this fact impacts more to Naïve Bayes and Logistic regression classifiers. When ranking attributes, the sensor that provides more information for classifying is the sensor placed on the forearm; the more relevant attributes are: action completion time, the power of the frequency 2-4Hz, the power of the frequency 4-6Hz and the standard deviation. The feature mean does not provide information; in particular, the $axis-Z$ of the sensor placed on the scapula, the $axis-Y$ and the $axis-Z$ of the sensor placed on the upper arm do not provide information either. These axes correspond to elevation/depression, abduction/adduction and flexion/extension of the shoulder. If in the last treatment proposed for classification, the attribute action completion time is not considered, the true positive rate for J48 decreases from 0.789 to 0.644; for Naïve Bayes from 0.911 to 0.889; and for Logistic regression models from 0.911 to 0.878, indicating that this attribute is crucial for the models based on the decision trees.

Our results suggest that selected complex actions can be classified with good performance using data recorded in the structured environment under controlled conditions, even using data recorded in daily living environments of the three subjects under non-controlled conditions. Based on these findings, we claim that it is feasible to recognize complex actions in daily living environments using computational techniques applied to orientation signals obtained from wearable inertial sensors.

6. CONCLUSIONS

Our results provide valuable information to measure the similarity of a set of complex actions in daily living environments. The studied actions can be classified with good precision using Naïve Bayes and Logistic regression models. Also, given the low complexity for training, building and evaluating models using Naïve Bayes, it is feasible to recognize the complex actions efficiently. Thereby, it is feasible to combine *DTW* distances and feature-based classification in order to recognize complex actions in real-time using a sliding window approach. Concerning to the complex actions, it is needed to consider a greater set of daily living actions than the actual one, as well as to extending the number of possible configurations of an action, *e.g.*, cooking. Finally, with the aim to distinguish very similar actions, merging the information obtained from wearable inertial sensors with data acquired by environmental sensors to combine spatial and contextual information in richer evidences is considered a possible avenue of research.

7. ACKNOWLEDGMENTS

The first author is supported by the Mexican National Council for Science and Technology, CONACYT, under the grant number 271539/224405. The authors would like to acknowledge the anonymous reviewers of the conference Pervasive Health 2016 for their valuable remarks.

8. REFERENCES

[1] A. Agresti and M. Kateri. *Categorical data analysis.* Berlin Heidelberg, 2011.

[2] K. Altun and B. Barshan. Human activity recognition using inertial/magnetic sensor units. In A. Salah, T. Gevers, N. Sebe, and A. Vinciarelli, editors, *Human Behavior Understanding*, volume 6219 of *Lecture Notes in Computer Science*, pages 38–51. Springer Berlin Heidelberg, 2010.

[3] H. Banaee, M. U. Ahmed, and A. Loutfi. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, 13(12):17472–17500, 2013.

[4] D. Biswas, A. Cranny, N. Gupta, K. Maharatna, J. Achner, J. Klemke, M. Jöbges, and S. Ortmann. Recognizing upper limb movements with wrist worn inertial sensors using k-means clustering classification. *Human movement science*, 40:59–76, 2015.

[5] A. Bulling, U. Blanke, and B. Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[6] A. Bulling, U. Blanke, D. Tan, J. Rekimoto, and G. Abowd. Introduction to the special issue on activity recognition for interaction. *ACM Trans. Interact. Intell. Syst.*, 4(4):16e:1–16e:3, Jan. 2015.

[7] C. Chen, R. Jafari, and N. Kehtarnavaz. A real-time human action recognition system using depth and inertial sensor fusion. *Sensors Journal, IEEE*, 16(3):773–781, Feb 2016.

[8] L. Chen, J. Hoey, C. Nugent, D. Cook, and Z. Yu. Sensor-based activity recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(6):790–808, Nov 2012.

[9] L. Chen and I. Khalil. Activity recognition: Approaches, practices and trends. In *Activity Recognition in Pervasive Intelligent Environments*, pages 1–31. Springer, 2011.

[10] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook. Simple and complex activity recognition through smart phones. In *Intelligent*

Environments (IE), 2012 8th International Conference on, pages 214–221. IEEE, 2012.

[11] K. Frank, M. J. V. Nadales, P. Robertson, and M. Angermann. Reliable real-time recognition of motion related human activities using mems inertial sensors. In *Proceedings of the 23rd International Technical Meeting of the Satellite Division of the Institute of Navigation*, pages 2906–2912, 2010.

[12] J. Han, M. Kamber, and J. Pei. *Data mining: concepts and techniques: concepts and techniques*. Elsevier, 2011.

[13] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, 2005.

[14] O. D. Lara and M. A. Labrador. A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE*, 15(3):1192–1209, 2013.

[15] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise*, 45(11):2193–2203, 2013.

[16] K. L. Moore, A. F. Dalley, and A. M. R. Agur. *Clinically oriented anatomy*. Lippincott Williams and Wilkins, a Wolters Kluwer business, 7th edition, 2014.

[17] L. Muda, M. Begam, and I. Elamvazuthi. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *arXiv preprint arXiv:1003.4083*, 2010.

[18] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato. A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology. *Proceedings of the IEEE*, 98(3):450–461, 2010.

[19] J. R. Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.

[20] C. R. Rao. *Handbook of Statistics: Data Mining and Data Visualization*, volume 24. Elsevier, 2005.

[21] S. Sempena, N. Maulidevi, and P. Aryan. Human action recognition using dynamic time warping. In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*, pages 1–5, July 2011.

[22] J. P. Varkey, D. Pompili, and T. A. Walls. Human motion recognition using a wireless sensor-based wearable system. *Personal and Ubiquitous Computing*, 16(7):897–910, 2012.

[23] W. Xu, M. Zhang, A. Sawchuk, and M. Sarrafzadeh. Co-recognition of human activity and sensor location via compressed sensing in wearable body sensor networks. In *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*, pages 124–129, May 2012.

[24] M. Zhang and A. A. Sawchuk. A feature selection-based framework for human activity recognition using wearable multimodal sensors. In *Proceedings of the 6th International Conference on Body Area Networks*, pages 92–98. ICST (Institute for Computer Sciences, SocialInformatics and Telecommunications Engineering), 2011.

[25] M. Zhang and A. A. Sawchuk. Human daily activity recognition with sparse representation using wearable sensors. *Biomedical and Health Informatics, IEEE Journal of*, 17(3):553–560, 2013.

APPENDIX

A. TIME-DOMAIN AND FREQUENCY-DOMAIN FEATURES

The time-domain features are calculated using Formulas 6 - 8. Range of motion, zero crossing rate, and action completion time are time-domain features. The frequency-domain feature is calculated using Formula 9.

In all the formulas detailed below: n is the size of the signal X, x is a value of X, th indicates the th-term of the ordered time serie X, and $Q1$ and $Q3$ are the lower and upper quartile of X.

- Arithmetic mean ($\bar{x}$):

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} \tag{6}$$

- Standard deviation (σ):

$$\sigma = \sqrt{\frac{1}{n}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2\right]} \tag{7}$$

- Root mean square (rms):

$$rms = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + ... + x_n^2\right)} \tag{8}$$

- The range of motion (rom) is calculated by subtracting the minimum value from the maximal value of each Euler angle. It measures the angular distance of a segment around a joint.
- Zero crossing rate (zcr) is the total number of times the signal X changes from below to above or vice versa.
- Action completion time (act) is the total time of the signal X.
- Power of a discrete signal (P_X):

$$P_X = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |x(i)|^2 \tag{9}$$