

# Human activity recognition using penalized support vector machines and hidden Markov models

Reconocimiento de actividades humanas utilizando máquinas de soporte vectorial penalizadas y modelos ocultos de Markov

Leidy Esperanza Pamplona Berón <sup>1</sup>, Carlos Alberto Henao Baena <sup>2\*</sup>, Andrés Felipe Calvo Salcedo <sup>3</sup>

<sup>1</sup>Facultad de Ingenierías, Universidad Santiago de Cali, Docente programa ingeniería electrónica. Calle 5 # 62 -00. C. P. 760035. Santiago de Cali, Colombia.

<sup>2</sup>Centro de Atención Sector Agropecuario, SENA - Tecnoparque- Nodo Pereira, Línea de Electrónica y Telecomunicaciones. Carrera 8 # 26-79, centro, carrera 8. C. P. 660002. Pereira, Colombia.

<sup>3</sup>Facultad de Ingenierías, Universidad Tecnológica de Pereira, Docente programa ingeniería electrónica. Carrera 27 # 10-02 Barrio Álamos. C. P. 660003. Risaralda, Colombia.



## CITE THIS ARTICLE AS:

L. E. Pamplona, C. A. Henao and A. F. Calvo. "Human activity recognition using penalized support vector machines and Hidden Markov Models in multimodal systems", *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 103, pp. 152-163, Apr-Jun 2022. [Online]. Available: <https://www.doi.org/10.17533/udea.redin.20210532>

## ARTICLE INFO:

Received: June 04, 2020

Accepted: May 17, 2021

Available online: May 24, 2021

## KEYWORDS:

Multisensor; data fusion; primitive motion; machine learning; overlapping databases

Multi sensor; fusion de datos; movimientos primitivos; aprendizaje de máquinas; datos traslapados

**ABSTRACT:** Human activity detection has evolved due to the advances and developments of machine learning techniques, which have enabled solutions to new challenges without ignoring prevalent difficulties that need to be addressed. One of the challenges is the learning model's sensitivity regarding the unbalanced, atypical, and overlapping information that directly affects the performance of the model. This article evaluates a methodology for the classification of human activities that penalizes defective information. The methodology is carried out through two redundant classifiers, a penalized support vector machine that detects the sub-movements (micro-movements) and the Markov Hidden Model that predicts the activity given the micro- movements sequence. The performance of the method was compared with state-of-the-art techniques, and the findings suggested significative advance in the detection of micro-movements compared to the data obtained with non-penalized paradigms. In this research, an adequate performance is found in the classification of primitive movements, with hit rates of 95.15% for the Kinect One®, 96.86% for the IMU sensor network, and 67.51% for the EMG sensor network.

**RESUMEN:** La detección de actividades humanas ha logrado evolucionar debido a los avances y desarrollos de técnicas del aprendizaje de máquinas, las cuales han permitido dar soluciones a nuevos desafíos sin ignorar las dificultades que aún persisten y abogan atención; uno de ellos concierne a la sensibilidad que presenta el modelo de aprendizaje ante información traslapada, desbalanceada y atípica que repercute propiamente en el desempeño del modelo. En este artículo se evalúa una metodología para la clasificación de actividades humanas que castiga información con imperfectos. El proceso metodológico se lleva cabo por medio de dos clasificadores redundantes, una Máquinas de Vectores de Soporte penalizada que detecta los sub movimientos (micromovimientos) y luego un Modelo Oculto de Markov que predice la actividad dada la secuencia de micro movimiento. El desempeño del método fue comparado con técnicas del estado de arte, los resultados sugieren un avance significativo en la detección de micromovimientos frente a los obtenidos con paradigmas no penalizados. En este trabajo se obtiene un adecuado desempeño en la clasificación de movimientos primitivos, con aciertos del 95,15% para el Kinect One®, del 96,86% para la red de sensores IMU y del 67,51% para la red de sensores EMG. Lo anterior impacta directamente la detección de actividades físicas con aciertos mayores al 95% de eficiencia.

## 1. Introduction

The recognition of physical activities aims to understand the actions carried out by people and how they interact

\* Corresponding author: Carlos Alberto Henao Baena

E-mail: [c\\_henao\\_86@hotmail.com](mailto:c_henao_86@hotmail.com)

ISSN 0120-6230

e-ISSN 2422-2844

with their physical environment. Some of the areas of application are videogames, robotics, rehabilitation, sports engineering, safety, among others [1–4]. In this way, the identification of activities seeks to track the human body's movements [1, 4, 5]. Different sensor methods such as depth cameras, inertial measurement unit sensors (IMU), and electromyography sensors (EMG) have been used to detect physical activity. Kinect One® is one of the most widely used non-invasive devices for motion tracking. This device records different types of data, such as articulated points, depth maps, and video. However, it poses practical problems with partial occlusions of the objective [6–9]. In other works, IMU sensors have been used, which measure the acceleration changes generated by movement. Although there have been advances using this sensor method, the advances have been invasive and require a device network to track activity [10, 11]. Other approaches have used electromyographic sensors (EMG) to measure muscle contraction and extension and detect movement [7]. Although these methods are robust to partial occlusions, both IMU's and EMG's require more than one sensor to completely register different activities; this is made computationally in the algorithm development process [2–7]. Modern approaches suggest that an activity can be represented as a continuous sequence or actions, known as primitive movements. In other words, the segmentation of activities into simpler components that are configured in a specific order can be classified using a learning machine approach. Although this approach produces satisfactory findings, a large volume of information is presented when labeling the data, making this process extremely expensive. Furthermore, determining a window size is a subject of study, and there is no satisfactory solution for it yet [1, 6, 7, 12, 13]. In other case studies, a codebook has been built with the key positions for each of the activities. Unfortunately, this process can only be applied with the articulated points of the skeleton, which limits the usage of other sensor methods [13]. The combination of different types of sensor devices has been an emerging topic of study in the activities recognition. However, there are a few studies where the fusion of more than two types of multimodal sensor system is performed [14–16]. In state of the art, combinations such as Kinect One® IMU, Kinect® EMG, and EMG IMU and Kinect One® IMU EMG are reported [6–16]. This method has been found to attain a more satisfactory performance with respect to the methods that mix one or two types of devices [17–21].

The work from [16] carries out the fusion of the Kinect One®, IMU, and EMG for the characterization of the movements through two redundant classifiers (SVM and a hidden Markov chain (HMM)) where efficiency greater than 95% in the detection of activities is achieved. Although satisfactory findings are computed, their computational

cost is high, and the number of activities under study is limited, which does not allow evaluating the real scope of the method. On the other hand, the state-of-the-art reports difficulties in classifying primitive movements claiming issues in the labeling and overlapping the data. Therefore, a penalty paradigm is required in order to assign less weight to data that biases the model. For instance, some papers have achieved satisfactory findings in classification problems where overlapping conflicts regarding the data reported.

This article studies the usage of SVM penalized classifiers as methodological tools to increase performance detecting primitive movements. In summary, it follows the procedure shown in [13–16] with a penalized classification model in the recognition of sub-activities using the SVM sanctioning strategy explained in Section 2. The findings are validated and compared with different sensor methods and state-of-the-art models. The main contributions of this article are:

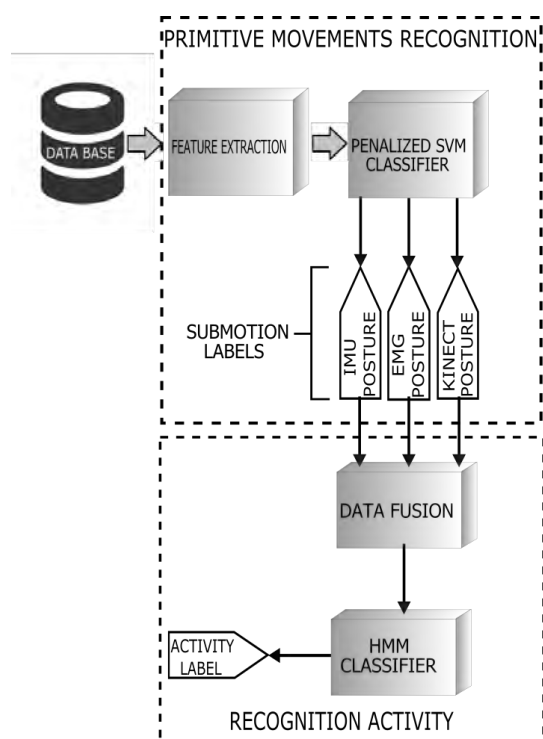
- Construction of an annotated database with the synchronized registration of three sensor methods (Kinect One®, IMU, and EMG). The same configuration proposed by [16] is used to construct the database extending it to 10 physical activities.
- A procedure that improves the performance of the primitive movement classifier through a paradigm of data penalization that can compare its own performance with similar methodologies that do not consider penalization [13–16], evidencing the benefits of the proposed technique.

This article is organized as follows. Firstly, a methodological section shows the summary of the procedure used to recognize physical activities, emphasizing the methodological instruments utilized. Secondly, a section goes over the findings where the performance of the method is evaluated and quantified. Lastly, a section describes the conclusions and further discussions of this research.

## 2. Methodology

Figure 1 shows the methodology used for human physical activity recognition, mainly based on the work shown in [16]. The method starts with the identification of the primitive movements that later determine the activity performed according to the coding or sequence of the sub-movements. In turn, the model controls a task based on the execution of a sequence of simpler sub-tasks through a hybrid model between HMM and SVM, as is pointed out by [22]. This means that the SVMs segment the input data, and the HMMs use the output of the SVMs to determine the most probable sequence regarding

time; thus, the methodology allows the execution of new actions according to a set of known movements. Moreover, it is possible due to the characteristics of the support vector machine to classify multidimensional data and to identify the HMMs ability when they operate with temporary sequential data [12, 22, 23]. It is important to point out that the HMM allows decoding the labels computed by the penalized classifier of micro-movements (sub-activities) and then reconstructing and recognizing the physical activity. Modern deep learning methods have been introduced to identify human physical activities achieving good findings when compared with classical techniques. However, they have been considered under a procedure that classifies the movement first and predicts the activity later, giving context to the movement. The foregoing disagrees with the approach to classifying human activities by primitive movements since the merged micromovement space is small, discrete, and inhomogeneous [24].



**Figure 1** Methodology for human physical activity recognition

Unlike the work by [16], a penalized SVM is introduced in order to punish atypical and spurious data. Consequently, the performance and efficiency in the detection of the sub-movements through a sanctioning paradigm can be evaluated.

## 2.1 Data capture system

A database consisting of 10 activities with some challenges in classifying activities such as walking and jogging, sitting, and getting up, among others, was created. The

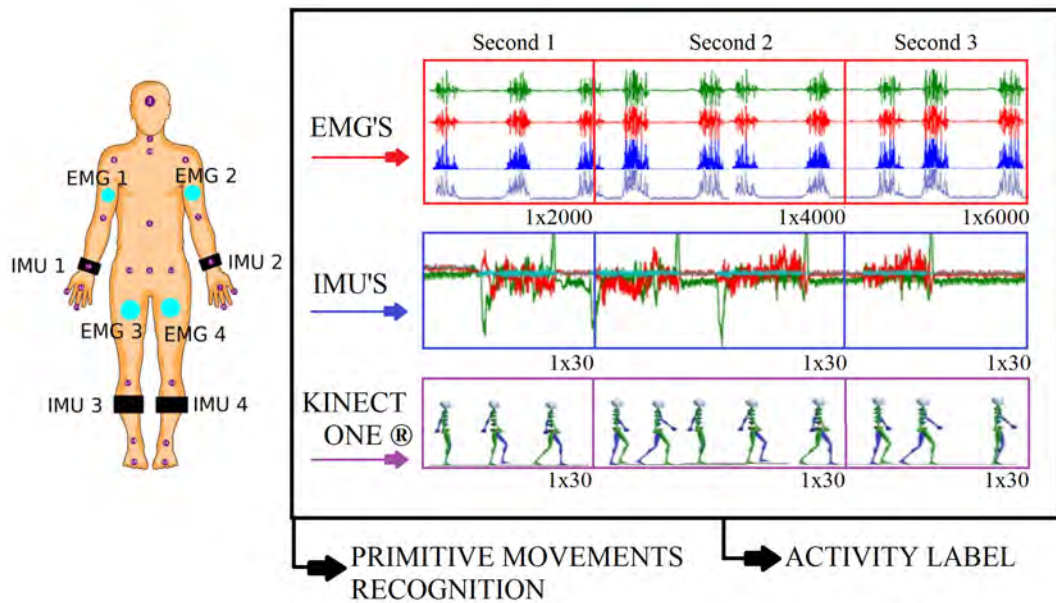
construction is divided into two tasks. Firstly, there is a storage of the synchronized record of the sensors (Kinect One®, IMU, EMG) in a binary structure. The data acquisition was carried out through the LabView Software following the suggestion expressed by [16, 25]. Figure 2 shows a graph that summarizes this stage. It is important to say that there are different articulated points for each sensing modality, where the articulated points of the Kinect One® are represented with  $\Psi$  using Cartesian coordinates system, the acceleration vector  $l_k$  is delivered by the IMU sensor, and EMG is an information vector which takes information by electromyographic sensors.

The second stage consists of labeling the activities as well as the primitive movements of the executed sequences. The first sequence was determined according to the ones presented in [1, 15, 16, 24–26]. Table 1 lists the 10 activities with their respective labels provided in this work. The data was retrieved from 12 participants (8 men and 5 women with 5 repetitions).

The primitive movement labels were constructed by segmenting the signal for each submovement sequence for each of the actions. Therefore, the data collected in three seconds (see Figure 2) is divided into  $N$  windows. These signals are stored in a file encoded according to the following structure: Base {Example} {Second} {Sensor} {Segment}.

On the other hand, the labeling of the database was carried out manually. In this way, it was possible to observe the spatial distribution of the different postures provided by Kinect One® during the recording time, establishing the separation between each of the submovements. The selection of each primitive movement is achieved by analyzing the execution of each activity. In addition, the phases introduced in the study of the kinematic analysis of the upper and lower limb of the human body used in [27–30] were considered. Table 1 summarizes the primitive movements that were determined for each activity. After recording the execution of different physical activities, these data are segmented into small windows to obtain the primitive movements, producing a new database, which presents unbalance and overlapping issues. This is probably because during the developments of the activity, there are micro-movements that appear more frequently. For instance, positions like standing still have more samples than movements such as raising the left hand to  $\frac{1}{4}$ . There are also overlap issues with the EMG sensor network, a problem also reported in [16].

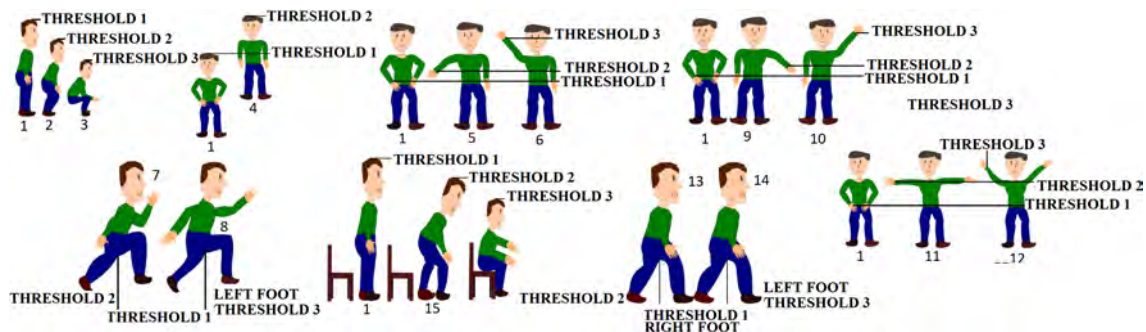
Figure 3 shows the thresholds of the labels established for each of the activities, considering the primitive movements defined in Table 1. Although threshold management could be considered a way to solve the issue, this approach does



**Figure 2** Sensors distribution diagram and synchronized capture system

**Table 1** Physical activities and primitive movements list with their corresponding labels

Activity (label)	Primitive Movement (label)
Hold Still (1)	Repose (1)
Squat and stand up (2)	Repose (1), Half squat (2), Full squat (3)
Jump (3)	Repose (1), Suspended in the air $\frac{1}{4}$ (4)
Rise right hand (4)	Repose (1), Rise right hand to $\frac{1}{4}$ (5), Rise right hand to $\frac{3}{4}$ (6)
Jogging (5)	Repose (1), Left leg forward with Flexed knee (7), Right leg forward with Flexed knee (8)
Rise left hand (6)	Repose (1), Rise left hand to $\frac{1}{4}$ (9), Rise left hand to $\frac{3}{4}$ (10)
Rise both hands (7)	Repose (1), Rise both hands to $\frac{1}{4}$ (11), Rise both hands to $\frac{3}{4}$ (12)
Walking (8)	Repose (1), Step forward left leg (13), Step forward right leg (14)
Sit down (9)	Repose (1), Halfway sit down (15), full sit down (16)
Stand up (10)	Full sit down (16), Halfway sit down (15), Repose (1)



**Figure 3** Threshold for label assignment of every primitive movement



not accurately describe the activities and, requires human intervention for different cases [16].

## 2.2 Primitive movement recognition

In this section, a dictionary describes the set of sub-movements as designed. Figure 3 summarizes the process for primitive movement detection for each sensing modality.

### Kinect one® feature extraction

For all 25 joint points provided by the Kinect One® (see Figure 1), the following descriptors were computed (polar and statistical) at a frequency of 15 Hertz [31, 32]:

**Polar Features [33]** Each jointpoint was transformed  $\psi = \{x, y, z\}$  to the polar plane by using the center of mass as the origin of coordinates  $\{CM_x, CM_y, CM_z\}$  of the Kinect One® joint points (see Figure 1), which allows obtaining the vector Equation 1.

$$P_i = [r_1 \theta_1 r_2 \dots r_{25} \theta_{25}] \quad (1)$$

Where  $P_i$  is the feature polar vector in the  $i$ -th sampling window for  $i = \{1, 2, 3\}$ ,  $r_j$  and  $\theta_j$  are the radial and angular components, respectively, of the jointpoint, with  $j = \{1, \dots, 25\}$ . Figure 4 shows a graphical diagram primitive movements detection.

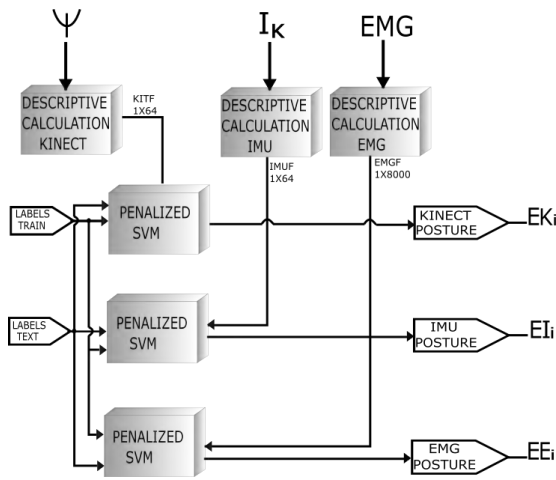


Figure 4 Primitive movement detection

**Statistical descriptors** The arithmetic mean ( $m_x, m_y, m_z$ ) and the variances ( $v_x, v_y, v_z$ ) of the spatial coordinates of the Kinect One® joint points  $\{x, y, z\}$  and their polar equivalent  $\{r, \theta\}$  were calculated with respect to the centers of mass  $\{CM_x, CM_y, CM_z\}$ , obtaining the following feature vector Equation 2.

$$MD_j = [m_x m_y m_z m_r m_\theta v_x v_y v_z v_r v_\theta] \quad (2)$$

Where ( $m_r, m_\theta$ ) are the mean of the polar coordinates and ( $v_r, v_\theta$ ) their variances. Finally, the total feature vector for each joint point  $KITF_j$  Equation 3 of the Kinect One® was calculated by concatenating  $P_i$  and  $MD$ . For more detail of the process of these descriptors refer to [7].

$$KITF_j = [P_1 P_2 P_3 MD] \quad (3)$$

### IMU Sensor network feature extraction

The IMU network data (see Figure 2) focuses on measuring the tri-axial components supplied by this network; the vector ( $a_x, a_y, a_z$ ) is obtained, where the variables are the rectangular acceleration components. Then, we calculated the Roll and Pitch orientations by performing the spherical coordinates conversion. This way, vector  $I_k$  Equation 4 is reached at each sampling time and for the  $k$ -th IMU of the network,  $k$  is defined as  $k = \{1, 2, 3, 4\}$ . The sampling frequency for each  $k$  sensor was 30Hz.

$$I_k = [a_x a_y a_z Pitch Roll] \quad (4)$$

The characterization of the vector  $I_k$  is performed by calculating the characteristics based on the physical parameters of human movement [34] (measurement of the  $AI$ , variance of  $AI$  ( $VI$ ), area of magnitude of normalized signal  $SMA$ , dominant direction eigenvalues  $EVA$ , average acceleration energy  $AAE$  and average rotation energy  $ARE$ ). Additionally, statistical measures to  $I_k$ , which are the arithmetic mean and the variance of the rectangular and spherical components of the accelerations were computed, obtaining the following vector Equation 5:

$$IMUF = [IMH Im_a Iv_a]_{1 \times 15} \quad (5)$$

With  $IMH = [AI VI SMA EVA AAE ARE]$ ,  $Im_a = [m_{ax} m_{ay} m_{az} m_{ar} m_{ap}]$  and  $Iv_a = [v_{ax} v_{ay} v_{az} v_{ar} v_{ap}]$ . Where  $m_{aw}$  and  $v_{aw}$  is the arithmetic mean and variance of the rectangular and spherical components of  $I_k$ , with  $w = \{x, y, z, Pitch, Roll\}$ . For more detail of the process of these descriptors, refer to [7].

### EMG Sensor network feature extraction

As can be seen in Figure 1, during this stage, four muscles of the body are sampled at a frequency of 2KHz, obtaining an analogous signal in the  $p$ -th EMG sensor being  $p = \{1, 2, 3, 4\}$ . Then, sampling window  $V_q$  was characterized by a Waveleth transform, acquiring a feature vector  $EMG_p$  Equation 6, whose dimensions are  $1 \times 2000$ . For each window  $q$ , the following feature vector reached.

$$EMGF_q = [EMG_1 \dots EMG_4]_{1 \times 8000} \quad (6)$$

**C-SVM algorithm****Input**

$D_T$ : Training data  $\{x_{train}, y_{train}\}$  with  $x_{test} \rightarrow [KITF \ IMUF \ EMGF]_{train}$

$D_V$ : Test data  $\{x_{test}, y_{test}\}$  with  $x_{test} \rightarrow [KITF \ IMUF \ EMGF]_{test}$

class: [1,2,3 ... n]

**Output**

$C_{SVM}\{u\}$ : Support Vectors

$\gamma_{best}$ : hyperparameters ( $C$  y  $\xi$ )

**Start**

**For**  $j = 1, \dots, n$

$Index(x, y) < \begin{cases} 1, & \text{if } y_i == class(j) \\ 0, & \text{In another case} \end{cases}$

$X_T \leftarrow -D_T(x_{train}, Index)$

$\gamma_{best} \rightarrow OP(X_T, D_V, L)$

$N \rightarrow length(find(index == 1))$

$C_{SVM}\{j\} = SVM\left\{X_T, \frac{C}{N}, \xi, OVA, RBF\right\}$

**End For**

**End**

(a)

**SVM\_BP algorithm****Input**

$D_T$ : Training data  $\{x_{train}, y_{train}\}$  with  $x_{test} \rightarrow [KITF \ IMUF \ EMGF]_{train}$

Class: [1,2,3 ... 19]

$N_{kx1}$ : [ $N_1 \ N_2 \dots N_{19}$ ] Data size set for each class

$\gamma_{best_{xx1}}$ : hyperparameters ( $C$  y  $\xi$ )

**Output**

$C_{SVM(i)}$ : Support Vectors

**Start**

$numclas \rightarrow length(clases)$

$comb \rightarrow nCr(numclas, 2)$

$ncomb \rightarrow length(comb)$

**For**  $j = 1, \dots, ncomb$

$N1 \rightarrow N(j, 1)$

$N2 \rightarrow N(comb(j, 2))$

$C_1 \rightarrow \gamma_{best}(1)$

$C_2 \rightarrow \frac{N_{C1}}{N_{C2}} * C_1$

$C \rightarrow [C_1 \ C_2]$

$X_T \rightarrow D_T(Comb(j, :) \text{ class data, } Comb(j, :) \text{ class labels})$

$SVM_{BP(i)} \rightarrow SVM\{X_T, C, \xi, OvO, RBF\}$

**End For**

**End**

**Definitions**

*RBF*: Radial Basic Function

*OvO*: One vs One

*OvA*: One vs All

*SVM*: Support Vector Machine

$nCr$ : Combinations number  $= \frac{n!}{(n-r)!r!}$

(b)

**Figure 5** Algorithm penalized by: (a) C-SVM method; (b) algorithm of binary combined SVM

## 2.3 Primitive movement classification

Three models of multiclass support vector machines with several classification strategies were used. Firstly, the C-SVC penalty method was used (see Figure 5a) [35]; secondly, the Weighted Binary SVM was implemented [36] (see Figure 5b); and lastly, a classic SVM method was used. For all the models, a Gaussian kernel with a radius of  $1 \times 10^{-4}$  was established by coupling the database in Section 2.2. In this case, it is preferred to penalize the data in the database because they present issues such as class overlapping and unbalance, in addition to the Kinect One®

partial occlusions or auto-occlusions, or loss of connection in the acquisition systems of the signals from IMUs or EMGs, which affect the performance and efficiency of the classifier in the identification of the primitive movements [16–37]. For more details, refer to [22]. On the other hand, the specified value of  $\tau$  corresponds to an initial value [6–16], which is refined in a search grid through a Monte Carlo experiment.

On the other hand, the penalized models seek to punish the regularization SVM parameter ( $C$  and  $\xi$ ) considering the size of the data in the search grid. For this specific

case, the Bayesian Optimization Procedure (OP) was implemented. This algorithm allows computing the regularization parameters of the penalized strategies ( $C$  and  $\xi$ ), given the training vector  $D_t$  and the validation vector  $D_v$ .

The value of this parameter can be very small (close to zero) or very large (tend to infinity), allowing the data penalization that is on the wrong side of the margin limit and thus minimizing the training error. However, these methods are sensitive to atypical information present in the data, the error is increasing in a linear process. Therefore, it is important to choose a suitable initial value for  $C$  [38]. Some authors recommend performing an observation grid by finding a value for  $C$  within the range in order to maximize the margin while penalizing the data located on the wrong side [36, 39]. In this work, an observation grid of  $[1 \times 10^{-3}, 1 \times 10^3]$  was selected to compute a  $C$  value for each class (see Table 1). On the other hand, these models consider the size of the training data set for the penalization of the regularization parameter  $C$ , which operates differently for each case. Assuming that SVM is a bi-class approach, with C-SVC there is a value of  $C$  with respect to the size of the data set to be classified, while with binary-weighted SVM, two values of  $C$  ( $C_2, C_1$ ) for each one of the classes are obtained.

### 3. Activity recognition

Based on the list of activities listed in Table 2 and the response of each of the SVMs, a posterior merge is developed, and the HMM is applied to identify the physical activity. Therefore, a vector of characteristics  $EF$  is created that linearly concatenates and no weighting the labels generated by the classifiers of each sensor during the observation window of three seconds, as shown in Figure 6. Equation 7 describes the vector structure  $EF$ .

$$EF = \begin{bmatrix} EK_1 & EK_2 & \dots & EK_{15} \\ EI_1 & EI_2 & \dots & EI_{30} \\ EE_1 & EE_2 & \dots & EE_{60} \end{bmatrix}_{105 \times 1} \quad (7)$$

Where:

- $EK$ : Feature vector of the SVM provided by the Kinect One®.
- $EI$ : Feature vector of the SVM provided by the IMU sensors network.
- $EE$ : Feature vector of the SVM provided by the EMG sensors network.

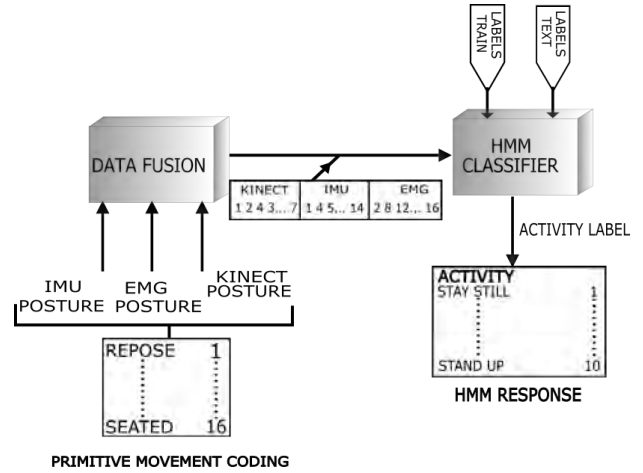


Figure 6 Data fusion

### 3.1 Model classification, training, and validation

The evaluation and validation of the model by using the cross-validation strategy through iterations in a Monte Carlo experiment with the stop criteria  $\|diag(M_k) - diag(M_k - 1)\| < 0.001$  were performed, where  $diag(M_k)$  is the vector generated by the diagonal of the confusion matrix, and  $k$  is the current average Monte Carlo iteration. The training data set (according to the database) of the model established in this work was 70%, and the remaining 30% was used for validation; the fragmentation of these percentages was performed randomly for each iteration. This allows observing the behavior of the HMM in the activity classification with different input data. In this last process, 24 states and 32 centroids for the construction of the codebook were used. To evaluate the classifier's performance, the confusion matrix, and the total number of hits per iteration were calculated and evaluated, respectively, determining the average behavior of the findings.

### 4. Experiments and findings

This section shows the experimental findings that validate the proposed methodology. These are divided into two sections. Firstly, the performance findings in the primitive movement classification stage are documented, in such a way that the returns are recorded for both the SVM penalized cases (C-SVM and weighted binary SVM) with respect to the model proposed in [16]. Secondly, the results are taught in the physical activities classification stage; in addition, the efficiency of the HMM model is evaluated for different types of fusion of the sensor modalities. Given the large number of experiments carried out, the confusion arrays are omitted. It is decided to specify the classes with low and high performance,

as well as the average behavior of the findings. For the identification of each finding, the coding of the experiments is presented based on Table 2:

The physical activity recognition for the following sensor combinations was performed: Kinect One®, IMUs, EMGs, Kinect One® + IMUs, EMGs, Kinect One® + IMUs, Kinect One® + EMGs, IMUs + EMGs, Kinect One® + IMUs + EMGs

The metric chosen to show the findings in this work is the mean value of correctness and standard deviation.

#### 4.1 Primitive movement analysis

Figure 7 shows the results obtained in the identification of primitive movements using the Kinect One® modality. A high performance of the experiments is evidenced, reaching efficiencies higher than 90%. In general, it should be noted that the SVM-BP method presents better performance than the one obtained with SVM\_CyH. However, class 16 (axis of the abscissa) presents a low yield close to 20%. On the other hand, the C-SVM strategy shows a more stable behavior with a success rate greater than 90%. Figure 8 shows the results obtained for the primitive movements classification using the IMU sensor network. The penalized strategies (C-SVM and SVM\_BP) have a performance value higher than 90% of accuracy. On the other hand, it is evident that these have a better performance than the one computed by SVM\_CyH. It is highlighted that the SVM\_BP method presents the highest efficiency with a success rate of  $96.86\% \pm 0.01\%$ . Figure 9 shows the performances of the classifiers by using the EMG sensor mode, where a low performance in comparison with those obtained in Figures 7 and 8 was observed. In summary, the SVM\_BP method presents a better performance with  $67.51 \pm 0.01\%$ . However, class 8 showed a low percentage of success of  $11.85 \pm 0.01\%$ , which suggests the inability of the classifier to detect it. Although this result is not adequate, the other methods also show a similar trend of low efficiencies. This could suggest that the extraction of characteristics under this modality may not be representative. Although it is not possible to obtain a competitive detection with the EMG sensor modality, it is inferred that the penalized strategies manage to improve the performance of the classification with respect to the non-penalized model SVM\_CyH. On the other hand, Figures 7, 8, and 9 show better performance under the SVM\_BP paradigm.

Figure 10 shows an average result that summarizes the finding shown in Figures 7, 8, and 9. It is important to highlight that the three sensor modalities under study present a better performance in the classification of micro-movements when the penalized strategy is coupled

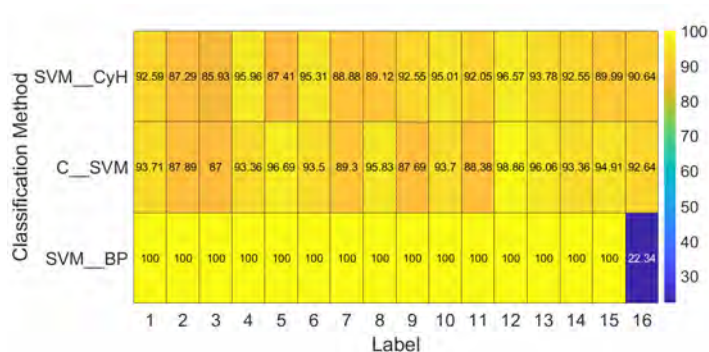


Figure 7 Primitive movement detection results for the kinect one® modality

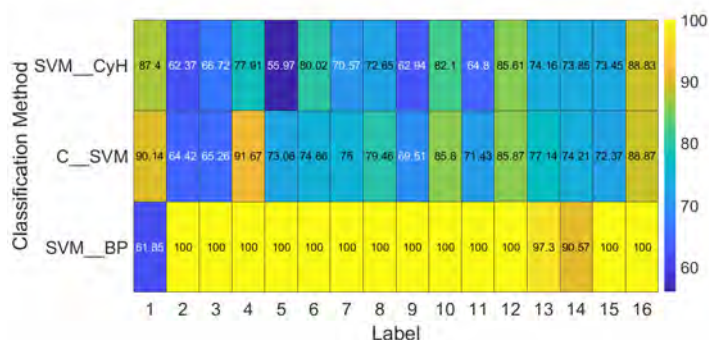


Figure 8 Primitive movement detection results for the IMU sensors network

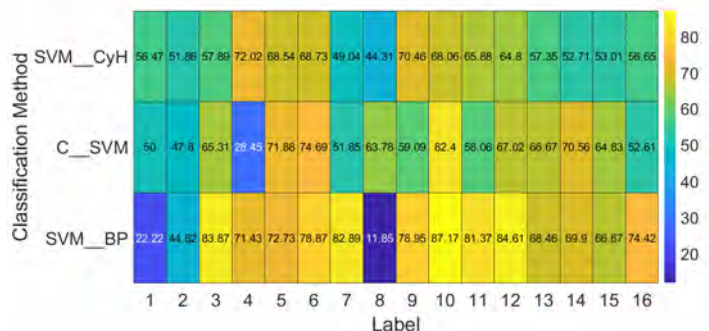


Figure 9 Primitive movement detection results for the EMG sensors network

through SVM\_BP and C-SVM. On the other hand, it is interesting to observe how the identification of primitive movements is more stable under this paradigm.

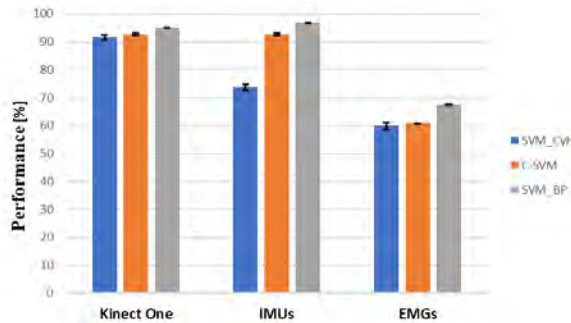
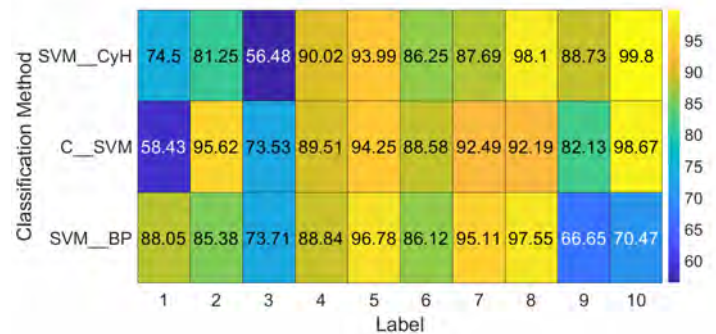
#### 4.2 Physical activity recognition

Figures 11, 12, and 13 show each sensor modality results. In summary, the performance of the activity classifier under the Kinect One® modality or with the IMU sensor network in most classes is greater than 90%, except for activities 2 and 3 with the HMM\_CyH and HMM\_C

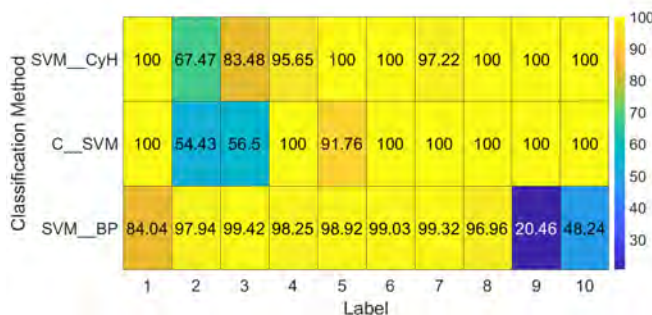


**Table 2** Experiment codification initials

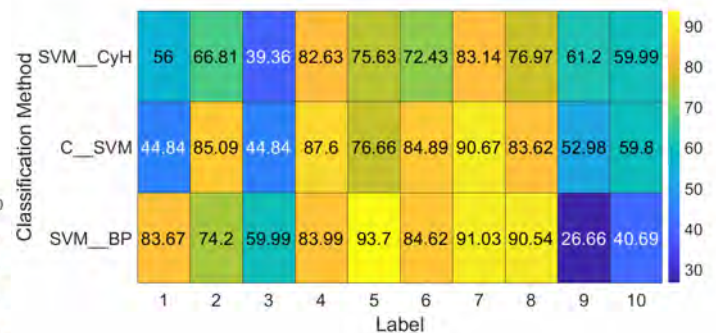
Primitive Movements Analysis (first column initial, second column description)	
SVM_CyH	Non-penalized SVM proposed by [16].
C-SVM	SVM with optimization of the regularized parameter C
SVM_BP	Weighted binary SVM
Activity Analysis	
HMM_CyH	HMM classifier with the labels delivered by SVM_CyH.
HMM_C	HMM classifier with the labels delivered by C-SVM.
HMM_BP	HMM classifier with the labels delivered by SVM_BP.

**Figure 10** Primitive movement detection results with the one Kinect One®, IMUs, and EMGs sensor networks**Figure 12** Physical activity recognition results by using IMU sensor network

strategies, and activities 1, 9, and 10 with the HMM\_BP technique in the Kinect One® modality. With the IMU sensor network, there are some issues with labels 3 and 5 by applying HMM\_CyH and HMM\_C algorithms, where efficiencies are under 80% of accuracy.

**Figure 11** Physical activity recognition results by using kinect one®

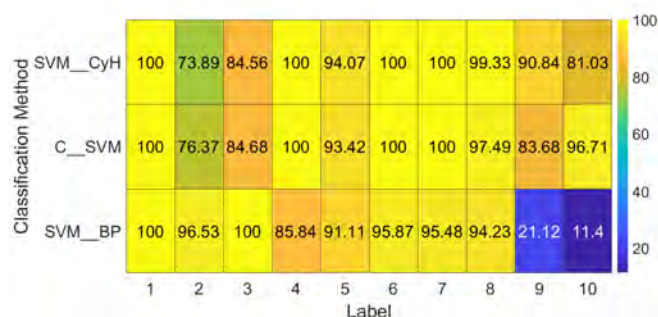
The same outcome is observed with classes 9 and 10 of the HMM\_BP model. On the other hand, with the EMG sensor network, an acceptable performance was achieved because label 7 (see Figure 13) shows the highest percentage of success compared to other activities, exceeding 83% of accuracy. The results suggest that only one sensor modality is sufficient for physical activity recognition. Although acceptable detection was managed with the EMG sensor modality, strategies to

**Figure 13** Physical activity recognition results by using EMG sensor network

improve its performance should be explored in further studies. Meanwhile, Figure 14 shows the performances for the three human physical activity identification techniques are higher than 90%, under a structure of IMU sensors. On the other hand, the results do not show a significant statistical gap between them. At the same time, the three classification methods show an activity detection less than 70% of accuracy due to the sequence of labels generated by the SVM. Comparing Figures 11, 12, and 13, the sensor modality with the best performance for human physical activity recognition is the IMU sensor.

### Kinect one® + IMUs experiment

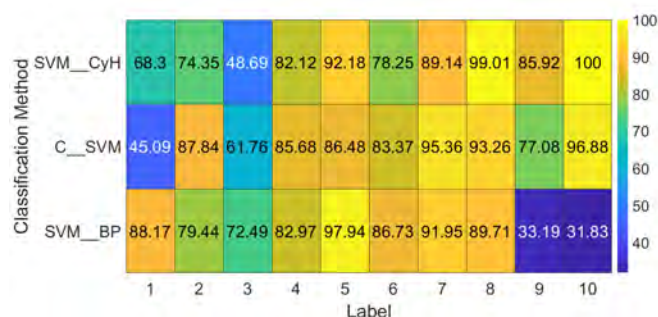
Figure 14 shows the results of the fusion of two sensor modalities, reaching a performance greater than 82% with the HMM\_CyH and HMM\_C methods. It is important to highlight that the identification of activity 1 presents a performance of 100.00% for the three models under study. However, the lowest performance is shown by the HMM\_BP method with 11% contrasting with those computed by HMM\_CyH and HMM\_C, which were 70%, respectively.



**Figure 14** Physical activity recognition results by using Kinect one® + IMUs

### Kinect + EMGs experiment

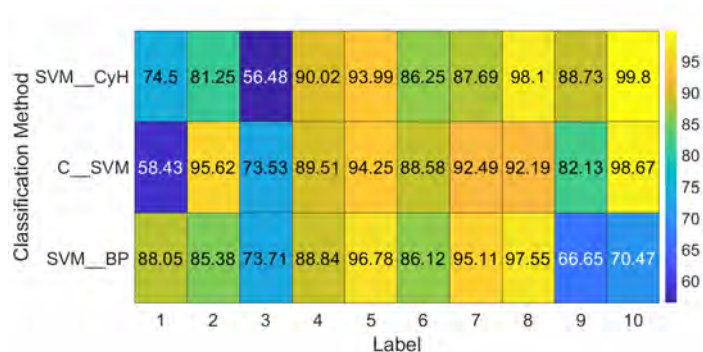
Figure 15 shows that this combination of sensors performs poorly for classes 1, 2, and 3, making it for motion detection.



**Figure 15** Physical activity recognition results by using Kinect one® + EMGs

### IMUs + EMGs experiment

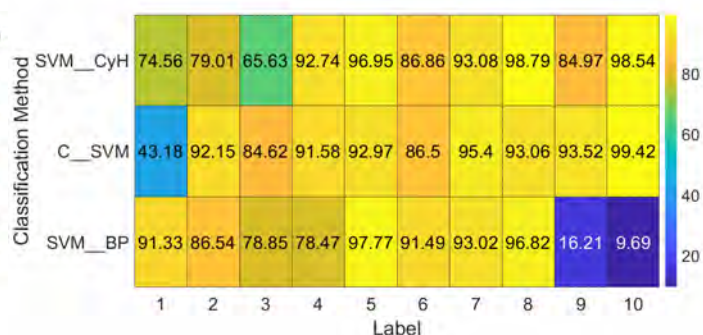
Figure 16 shows that the fusion of these two sensor modalities allows attaining a performance higher than 84%, where with activity 10 under the HMM\_CyH and HMM\_C methods, efficiencies greater than 98% are achieved, similarly with the HMM\_BP strategy it stands out a yield close to 100% with the label 8.



**Figure 16** Physical activity recognition results by using IMUs + EMGs

### Kinect one® + IMUs + EMGs experiment

Figure 17 shows the results obtained under the fusion of the three sensor modalities, where the methods HMM\_CyH and HMM\_C present similar performances of approximately 87%. On the other hand, class 10 with the HMM\_C method shows a 100% hit rate, unlike the HMM\_BP method, where this class shows the lowest hit rate with 9%. Similar to what is shown in Figure 11 and Figure 18, the results (means  $\pm$  standard deviation) of Figures 11 to 17 are condensed. It is observed that in the detection of the activities, the penalized strategy is competitive with respect to the non-penalized one (HMM\_CyH). This suggests similar performances of human physical activities identification regardless of the penalty.



**Figure 17** Physical activity recognition results by using Kinect one® + IMUs + EMGs

## 5. Conclusions and recommendation

This research work carried out a comparative study by articulating different learning models for primitive movement identification. These models [16] were compared against a punished paradigm that uses

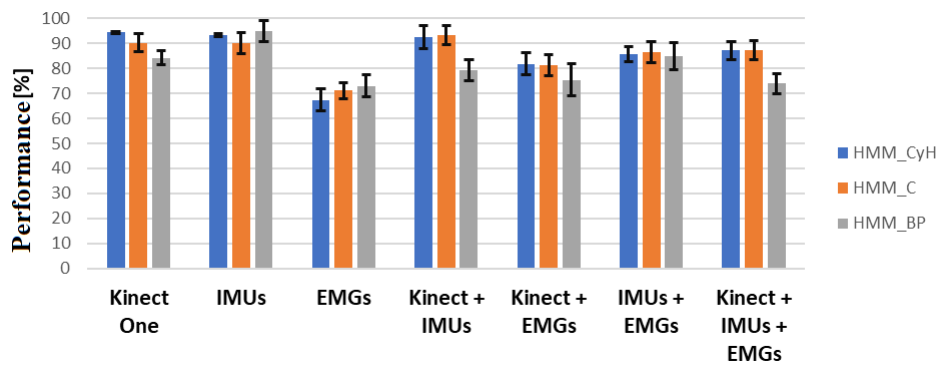


Figure 18 Physical activity detection results

penalized SVM. The research found out that either of the two penalty methods increases the classifiers' performance for the detection of primitive movements [16]. For the Kinect One®, the best result is achieved using the weighted binary SVM, which has an efficiency of 95.15%. For the IMU sensor network, the weighted binary SVM generates the best results with 96.86% accuracy. This method generates these results for the two sensor modalities due to the consideration of an existing imbalance between the classes, which improves the separation boundary. On the other hand, in the activity detection stage using HMM, it is possible to show that the Kinect One® sensor generates detections with greater efficiency, with a 93.23% performance. It was also found that merging different modalities does not always improve detection performance. This can be observed in Figures 15, 16, and 17, where these are reduced in the combinations that add information from different sensory sources. This data contradicts the results obtained in the work by [16] because by extending the database, the complexity of the data increases, and the sensors can deliver information that biases the model.

It is important to highlight that in this work, the database of physical activities developed by [16] was extended in 10 activities with the synchronized recording of the Kinect One®, IMUs, and EMGs, where more join points of the human body and 16 primitive movements were included. This finding means a significant contribution to the study of these methodologies. Indeed, it evaluates the performance based on the variations of activities and sub-movements to be identified in order to determine the scope or restrictions they present.

## 6. Declaracion of competing interest

We declare that we have no significant competing interests, including financial or non-financial, professional, or personal interests interfering with the full and objective

presentation of the work described in this manuscript.

## 7. Acknowledgements

The authors are grateful to the Centro de Atención al Sector Agropecuario - SENA; to Tecnoparque Nodo Pereira, línea de Electrónica y Telecomunicaciones; and Universidad Tecnológica de Pereira for the support received during the development of this research.

## References

- [1] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, "Gchar: An efficient group-based context aware human activity recognition on smartphone," *Journal of Parallel and Distributed Computing*, vol. 118, Part 1, Aug. 2018. [Online]. Available: <https://doi.org/10.1016/j.jpdc.2017.05.007>
- [2] A. Khan, N. Hammerla, S. Mellor, and T. Plötz, "Optimising sampling rates for accelerometer-based human activity recognition," *Pattern Recognition Letters*, vol. 73, Apr. 1, 2016. [Online]. Available: <https://doi.org/10.1016/j.patrec.2016.01.001>
- [3] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Information Fusion*, vol. 35, May. 2017. [Online]. Available: <https://doi.org/10.1016/j.inffus.2016.09.005>
- [4] Y. L. Chen and *et al.*, "Dimensionality reduction of data sequences for human activity recognition," *Neurocomputing*, vol. 210, Oct. 19, 2016. [Online]. Available: <https://doi.org/10.1016/j.neucom.2015.11.126>
- [5] W. Takano, H. Imagawa, and Y. Nakamura, "Spatio-temporal structure of human motion primitives and its application to motion prediction," *Robotics and Autonomous Systems*, vol. 75, Part B, Jan. 2016. [Online]. Available: <https://doi.org/10.1016/j.robot.2015.09.017>
- [6] S. Morales, "Identificación de actividad humana usando aprendizaje no supervisado en sistemas multimodales," M.S. thesis, Facultad de Ingenierías, Universidad Tecnológica de Pereira, Pereira, CO, 2016.
- [7] A. F. Calvo, "Reconocimiento automático de actividades físicas humanas en sistemas multimodales," M.S. thesis, Facultad de Ingenierías, Universidad Tecnológica de Pereira, Pereira, CO, 2015.
- [8] M. Jiang, J. Kong, G. Bebis, and H. Huo, "Informative joints based human action recognition using skeleton contexts," *Signal Processing: Image Communication*, vol. 33, Apr. 2015. [Online]. Available: <https://doi.org/10.1016/j.image.2015.02.004>
- [9] R. Qiao, L. Liu, C. Shen, and A. V. Den, "Learning discriminative trajectorylet detector sets for accurate skeleton-based action



- recognition," *Pattern Recognition*, vol. 66, Jun. 2017. [Online]. Available: <https://doi.org/10.1016/j.patcog.2017.01.015>
- [10] A. Bayat, M. Pomplun, and D. Tran, "A study on human activity recognition using accelerometer data from smartphones," *Procedia Computer Science*, vol. 34, 2014. [Online]. Available: <https://doi.org/10.1016/j.procs.2014.07.009>
- [11] A. Akbari, X. Thomas, and R. Jafari, "Automatic noise estimation and context-enhanced data fusion of imu and kinect for human motion measurement," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Eindhoven, NL, 2017.
- [12] I. Serrano, V. Kyrki, D. Kragic, and M. Larsson, "Action recognition and understanding through motor primitives," *Advanced Robotics*, vol. 21, no. 15, Nov. 2007. [Online]. Available: <https://doi.org/10.1163/156855307782506156>
- [13] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-D posture data," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, Dec. 18, 2014. [Online]. Available: <https://doi.org/10.1109/THMS.2014.2377111>
- [14] M. Zhang and A. Sawchuk, "Motion primitive-based human activity recognition using a bag-of-features approach," in *IHI '12: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 2012, pp. 631–640.
- [15] J. F. S. Lin, M. Karg, and D. Kulić, "Movement primitive segmentation for human motion modeling: A framework for analysis," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, Jun. 2016. [Online]. Available: <https://doi.org/10.1109/THMS.2015.2493536>
- [16] A. F. Calvo, G. A. Holguin, and H. Medeiros, "Human activity recognition using multi-modal data fusion," in *CIARP 2018: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 2018, pp. 946–953.
- [17] B. Wang, C. Yang, and Q. Xie, "Human-machine interfaces based on emg and kinect applied to teleoperation of a mobile humanoid robot," in *Proceedings of the 10th World Congress on Intelligent Control and Automation*, Beijing, CN, 2012, pp. 3903–3908.
- [18] S. Feng and R. Murray, "Fusing kinect sensor and inertial sensors with multi-rate kalman filter," in *IET Conference on Data Fusion & Target Tracking 2014: Algorithms and Applications (DF&TT 2014)*, Liverpool, UK, 2014, pp. 1–8.
- [19] C. Xiang, H. H. Hsu, W. Y. Hwang, and J. Ma, "Comparing real-time human motion capture system using inertial sensors with microsoft kinect," in *2014 7th International Conference on Ubi-Media Computing and Workshops*, Ulaanbaatar, MN, 2014, pp. 53–58.
- [20] M. Caon and *et al.*, "Kinesiological electromyography for activity recognition," in *PETRA '13: Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*, New York, USA, 2013, pp. 1–7.
- [21] H. Koskimaki and P. Siirtola, "Accelerometer vs. electromyogram in activity recognition," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 5, no. 3, Nov. 2016. [Online]. Available: <https://doi.org/10.14201/ADCAIJ2016533142>
- [22] A. Castellani, D. Botturi, M. Bicego, and P. Fiorini, "Hybrid hmm/svm model for the analysis and segmentation of teleoperation tasks," in *IEEE International Conference on Robotics and Automation, 2004*, New Orleans, LA, USA, 2004, pp. 2918–2923.
- [23] B. Hannaford and P. Lee, "Hidden markov model analysis of force/torque information in telemanipulation," *The International journal of robotics research*, vol. 10, no. 5, Oct. 1, 1991. [Online]. Available: <https://doi.org/10.1177/027836499101000508>
- [24] K. Chen and *et al.*, "Deep learning for sensor-based human activity recognition: Overview, challenges and opportunities," *J. ACM*, vol. 37, no. 4, Aug. 2018. [Online]. Available: <https://doi.org/10.1145/1122445.1122456>
- [25] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley mhad: A comprehensive multimodal human action database," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, Tampa, FL, USA, 2013, pp. 53–60.
- [26] H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, "Exploiting deep residual networks for human action recognition from skeletal data," *Computer Vision and Image Understanding*, vol. 170, May. 2018. [Online]. Available: <https://doi.org/10.1016/j.cviu.2018.03.003>
- [27] C. Roldán, "Estudio de la cinemática del miembro superior e inferior mediante sensores inerciales," Ph. D. dissertation, Facultad de Ciencias de la Salud, Universidad de Málaga, Málaga, ES, 2017.
- [28] D. A. Winter, "Moments of force and mechanical power in jogging," *Journal of biomechanics*, vol. 16, no. 1, 1983. [Online]. Available: [https://doi.org/10.1016/0021-9290\(83\)90050-7](https://doi.org/10.1016/0021-9290(83)90050-7)
- [29] S. A. Dugan and K. P. Bhat, "Biomechanics and analysis of running gait," *Physical Medicine and Rehabilitation Clinics*, vol. 16, no. 3, Aug. 2005. [Online]. Available: <https://doi.org/10.1016/j.pmr.2005.02.007>
- [30] A. Gomez, R. Becerro, and M. E. Losa, "Reliability of the optogait portable photoelectric cell system for the quantification of spatial-temporal parameters of gait in young adults," *Gait & posture*, vol. 50, Oct. 2016. [Online]. Available: <https://doi.org/10.1016/j.gaitpost.2016.08.035>
- [31] S. Zennaro and *et al.*, "Performance evaluation of the 1st and 2nd generation kinect for multimedia applications," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, Turin, IT, 2015, pp. 1–6.
- [32] D. Pagliari and L. Pinto, "Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors," *Sensors*, vol. 15, no. 11, Oct. 2015. [Online]. Available: <https://doi.org/10.3390/s151127569>
- [33] H. Wu, W. Pan, X. Xiong, and S. Xu, "Human activity recognition based on the combined SVM&HMM," in *2014 IEEE International Conference on Information and Automation (ICIA)*, Hailar, CN, 2014, pp. 219–224.
- [34] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *BodyNets '11: Proceedings of the 6th International Conference on Body Area Networks*, Zadar, HR, 2011, pp. 92–98.
- [35] B. Schölkopf and A. J. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*, 1st ed. Cambridge, MA, USA: MIT Press, 2001.
- [36] S. Shao, K. Shen, C. J. Ong, E. P. V. Wilder, and X. Li, "Automatic EEG artifact removal: A weighted support vector machine approach with error correction," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, Feb. 2009. [Online]. Available: <https://doi.org/10.1109/TBME.2008.2005969>
- [37] J. F. Gallego and D. F. Rengifo, "Comparación de técnicas de reducción de dimensionalidad para la clasificación de actividades físicas humanas utilizando métodos estadísticos," Undergraduate degree, Facultad de Ingeniería, Universidad Tecnológica de Pereira, Pereira, CO, 2016.
- [38] L. Jiang and R. Yao, "Modelling personal thermal sensations using c-support vector classification (c-svc) algorithm," *Building and Environment*, vol. 99, Apr. 2016. [Online]. Available: <https://doi.org/10.1016/j.buildenv.2016.01.022>
- [39] N. Becker, W. Werft, G. Toedt, P. Lichter, and A. Benner, "penalizedsvm: a r-package for feature selection svm classification," *Bioinformatics*, vol. 25, no. 13, Jul. 1, 2009. [Online]. Available: <https://doi.org/10.1093/bioinformatics/btp286>