

**RECONOCIMIENTO AUTOMÁTICO DE ACTIVIDADES FÍSICAS
HUMANAS EN SISTEMAS MULTIMODALES**

Ing. Andrés Felipe Calvo Salcedo

Proyecto de grado presentado como requisito parcial
para aspirar al título de Magíster en Ingeniería Eléctrica

Director

Ing. Germán A. Holguín L, M.Sc, Ph.D(C)

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
PROGRAMA DE MAESTRIA EN INGENIERÍA ELÉCTRICA
LINEA DE AUTOMÁTICA
PEREIRA
2015**

Nota de Aceptación

Firma del Presidente del jurado

Firma del jurado 1 - Evaluador

Firma del jurado 2 - Evaluador

Firma del jurado 3 - Director

Pereira, 09 de Octubre de 2015

Agradecimientos

Después de mucho trabajo y tiempo invertido en este proceso de maestría, es una tarea complicada agradecer a toda las personas que con su presencia a lo largo de este tiempo, me ha ayudado a escalar las diferentes etapas que conllevó este proceso. Inicialmente quiero agradecer a Dios por guiarme en esta etapa de mi vida y a darme la posibilidad de vivir esta porción de mi vida. Agradezco a mis padres Edgar Calvo y Patricia Salcedo por su apoyo condicional, incluso estas palabras son simples y cortas para expresar lo que siento, a demás quiero agradecer a mi novia Carolina Marín, por su amor y entrega durante estos años de maestría, la descomunial paciencia y madurez para entender la cantidad de tiempo que debía invertir para lograr triunfar en este proceso.

Agradecer a mi tutor de maestría German Holguín por su asesoría y acompañamiento durante estos años, en este momento de mi vida lo puedo considero un padre académico. Agradecer también a mi compañero de maestría Daniel Molina, por su apoyo condicional y leatad durante este proceso, *!amigos como el hay pocos!*. También quiero agradecer a mis demas compañeros, Rodrigo, Leidy, Arley y Rafael por gran aporte durante este tiempo, creanme que han sido una gran escuela para mi, he aprendido mucho observandolos a ustedes. Por supuesto tengo que agradecer al programa de Ingeniera Electrónica, el programa de Ingeniería Eléctrica y a la Universidad Tecnológica de Pereira, por su apoyo durante es proceso de maestría. Todas las personas de esta comunidad han sido de gran importancia para este proceso, estudiantes de pregrado, Clarita, Edwín Quintero, Germán López y Mauricio Holguín, no queda más que agradecer su noble intención en desearme progresar.

Por último quiero agradecer de todo corazón a mis amigos de siempre, con los que he compartido ya tantos años y experiencias que han llegado a convertirse en parte fundamental en mi vida, estos momentos serán importantes para mí. Gracias por todo.

Muchas Gracias.

CONTENIDO

pág.

TÍTULO

CONTENIDO

1. INTRODUCCIÓN	11
1.1. DEFINICIÓN DEL PROBLEMA	12
1.2. JUSTIFICACIÓN	14
1.3. OBJETIVOS	15
1.3.1. Objetivo General	15
1.3.2. Objetivos Específicos	16
2. ESTADO DEL ARTE	17
2.1. Cámaras de Video y cámaras de Profundidad	17
2.1.1. Procesamiento y extracción de características . . .	17
2.1.2. Clasificación	18
2.2. Acelerómetros y unidades de Masa Inercial	18
2.2.1. Procesamiento y extracción de características . . .	18
2.2.2. Clasificación	19
2.3. Sensores Electromiográficos	19
2.3.1. Procesamiento y extracción de características . . .	20
2.3.2. Clasificación	20
2.4. Fusion de Datos	20
3. MARCO TEÓRICO	21
3.1. Dispositivos para la captura de movimiento	21
3.1.1. Sensor	21
3.1.2. Unidad de masa inercial (IMU)	21

3.1.3.	Kinect	22
3.1.4.	Electromiografía (EMG)	22
3.2.	Extracción de características	22
3.2.1.	Transformada Wavelet Discreta (<i>DWT-Discrete Wavelet Transform</i>)	23
3.2.2.	Mel Frequency Cepstrum Coefficients (MFCC)	23
3.3.	Clasificación	24
3.3.1.	Clasificación no supervisada [1]	24
3.3.2.	Clasificación supervisada [1]	25
3.3.3.	Máquinas de vectores de soporte (SVM)	25
3.3.4.	Máquinas de vectores de soporte multiclase	27
3.3.5.	Estimación de probabilidades en SVM:	27
3.3.6.	Modelos Ocultos de Markov (HMM)	28
3.3.7.	Análisis de rendimiento de un clasificador	30
3.3.8.	Intervalo de confianza	31
4.	HARDWARE Y CAPTURA DE DATOS	33
4.1.	Captura sincronizada de la red de sensores IMU, red de sensores EMG y dispositivo Kinect® en LabVIEW.	34
4.1.1.	Adquisición de sensores EMG	34
4.1.2.	Adquisición de sensores IMU	35
4.1.3.	Adquisición de puntos Articulados Kinect®	37
4.2.	Almacenamiento de base de datos anotada de actividades humanas	39
4.3.	Almacenamiento de base de datos anotada de movimientos primitivos	40
5.	RECONOCIMIENTO DE MOVIMIENTOS PRIMITIVOS	43
5.1.	Codificación de movimientos primitivos	43

5.2.	Extracción de características	44
5.2.1.	Extracción de características del Kinect	44
5.2.2.	Extracción de características para red sensores IMU	47
5.2.3.	Extracción de características Red de sensores EMG	49
5.3.	Clasificación de Movimientos	49
6.	RECONOCIMIENTO DE ACTIVIDAD	51
6.1.	Fusión de datos	51
6.2.	Clasificación con Modelos Ocultos de Markov (HMM), Entrenamiento y validación del modelo	51
7.	EXPERIMENTOS Y RESULTADOS	53
7.1.	Reconocimiento de Movimientos Primitivos	53
7.2.	Reconocimiento de actividades físicas	56
8.	CONCLUSIONES Y RECOMENDACIONES	65
8.1.	CONCLUSIONES	65
8.2.	RECOMENDACIONES	66
	BIBLIOGRAFÍA	69
9.	ANÉXOS	79
9.1.	Anéxo Acondicionamiento de señal sensores EMG	79

LISTA DE TABLAS

1.	Lista de actividades	39
2.	Etiquetas del SVM para cada postura	43
3.	Descriptores estadísticos utilizados	46
4.	Características IMU	48
5.	Descriptores estadísticos utilizados	48
6.	Matriz de Confusión SVM-Kinect®	54
7.	Matriz de Confusión SVM-IMU todo el conjunto de sensores	54
8.	Matriz de Confusión SVM-IMU con sensor de mano derecha	55
9.	Matriz de Confusión SVM-EMG con todo el conjunto de sensores	56
10.	Matriz de Confusión SVM-EMG con sensores de mano derecha y pies	56
11.	Matriz de Confusión HMM-Kinect® con todos lo sensores	58
12.	Matriz de Confusión HMM-IMU con todos lo sensores	58
13.	Matriz de Confusión HMM-IMU con sensor de la mano derecha	59
14.	Matriz de Confusión HMM-EMG con todos los sensores	59
15.	Matriz de Confusión HMM-Kinect®+EMG	60
16.	Matriz de Confusión HMM-Kinect®+EMG con sensores mano de- recha y pie derecho	60
17.	Matriz de Confusión HMM-Kinect®+IMU con todos los sensores	60
18.	Matriz de Confusión HMM-Kinect®+IMU con sensor de la mano derecha	61
19.	Matriz de Confusión HMM-IMU+EMG con todos los sensores	61
20.	Matriz de Confusión HMM-IMU+EMG con sensores brazo derecho y pie derecho	61
21.	Matriz de Confusión HMM-Kinect®+IMU+EMG con todos los sensores	62
22.	Matriz de Confusión HMM-Kinect®+IMU+EMG con mano dere- cha y brazo derecho	62

23.	Comparación de desempeño de la metodología para todo el conjunto de sensores	63
24.	Comparación de desempeño de la metodología con el conjunto mínimo de sensores.	63
25.	Comportamiento promedio de la metodología	63
26.	Comportamiento promedio de la metodología para el conjunto mínimo de sensores	64

LISTA DE FIGURAS

1.	Etapas del sistema de captura sincronizado	34
2.	Adquisición de sensores EMG	36
3.	Distribución de los sensores (esquema)	36
5.	Adquisición de sensores IMU	37
4.	Distribución de los sensores (usuario real)	37
6.	Adquisición Kinect®	38
7.	Interfaz gráfica diseñada para captura de actividades	38
8.	Configuración y ejecución de actividades	40
9.	Estructura base de datos	40
10.	Umbrales para asignación de etiquetas	42
11.	Actividades 1, 2 y 3	44
12.	Actividad 4 y 5	44
13.	Diagrama de Segmentación de señales	45
14.	Puntos articulados del cuerpo	45
15.	Número de centroides Vs Eficiencia de Acierto	57
17.	Comportamiento global para las combinaciones con cantidad reducida de sensores	64
16.	Comportamiento global para las combinaciones con todo el conjunto de sensores	64
18.	Acondicionamiento de señal para un sensor EMG	79

1. INTRODUCCIÓN

La importancia del análisis de actividad física ha motivado a diferentes investigadores a desarrollar metodologías que permitan la detección automática de un conjunto de actividades. Este tipo de métodos permiten desarrollar aplicaciones en áreas como la salud, seguridad, deporte, entre otras [2, 3, 4, 5]. Los sensores más comunes para captar movimientos humanos son las cámaras de vídeo, los sensores de profundidad, los sensores electromiográficos (EMG), goniómetros, giroscopios, acelerómetros, unidades de masa inercial (IMU-*Inertial Measured Unit*), entre otros. Aunque cada sensor tiene ventajas significativas con respecto a otras modalidades, utilizar un solo tipo sensor para la detección de actividad, ha demostrado ser insuficiente para caracterizar una amplia variedad de actividades físicas humanas [6]. Entonces la combinación de diferentes fuentes sensoricas o fusión de datos es una opción para mejorar el factor discriminante de una observación aprovechando las ventajas individuales de cada sensor [7, 8, 9, 10].

En la literatura se han diseñado metodologías eficientes de fusión de datos [11, 12, 13, 14, 15, 16, 17], sin embargo, no se evidencia una metodología que realice la fusión de los sensores más utilizados para estas tareas, como lo son, el Kinect®¹, los IMU y los EMG [2, 8, 10]. En este trabajo, se propone una nueva metodología para el reconocimiento de actividades humanas utilizando fusión de características provistas por sensores IMU, EMG y Kinect®. Para la captura de los datos se utiliza una red de cuatro acelerómetros distribuidos en diferentes puntos del cuerpo, una red de cuatro sensores EMG utilizados para sensar la actividad eléctrica de músculos del cuerpo y un sensor Kinect encargado de extraer las coordenadas 3D de los principales puntos del cuerpo.

Estos datos son utilizados para el reconocimiento del conjunto de movimientos primitivos que codifican de forma única una actividad y la secuencia generada por esta codificación permitirá la clasificación durante una ventana de tiempo utilizando un modelo oculto de Markov (HMM-*Hidden Markov Model*). Esta metodología demuestra que la fusión multimodal de los sensores Kinect®, IMU y EMG, permite obtener mayor desempeño en el reconocimiento, que cualquier combinación de este grupo de sensores o incluso, el uso individual de alguno de ellos.

1.1. DEFINICIÓN DEL PROBLEMA

Actualmente, el análisis de la actividad física humana es un campo atractivo para distintas áreas de la investigación, esto se debe en gran parte a la posibilidad de diseñar aplicaciones en áreas como la salud, seguridad, domótica, deporte, entre otras [2, 3, 4, 5]. A través del análisis del movimiento humano es factible detectar un posible episodio de enfermedades como Parkinson o Epilepsia [13], realizar rehabilitación de pacientes [18], vigilar el estado del adulto mayor en el diario vivir [19], monitorear el rendimiento físico de deportistas e incluso detectar comportamientos erráticos en centros comerciales, aeropuertos y demás espacios donde hay una afluencia de público [13]. Desde el punto de vista de la ingeniería y las ciencias de la computación, el reconocimiento de actividad física humana tiene como objetivo la identificación automática de diferentes tipos de movimientos físicos a partir de un conjunto de observaciones. Estas observaciones pueden ser captadas por cámaras de vídeo, sensores de profundidad, sensores de ritmo cardíaco, goniómetros, giroscopios, acelerómetros, entre otros.

Para un ser humano, el reconocimiento de actividad física se considera una habilidad totalmente natural. Tratar de emular este tipo de habilidad en sistemas de detección automático se considera un gran desafío y ha motivado a varios investigadores a proponer diferentes tipos de soluciones [13, 11, 12, 20, 14]. Aunque la detección de actividad es una labor que se puede realizar de forma manual por personal capacitado, esto es costoso, dispendioso, requiere un ambiente estructurado y está sujeto a múltiples niveles de errores humanos. Entre dichas soluciones se encuentran técnicas como la construcción de diarios normalizados, donde los sujetos en prueba llevan un registro señalando el inicio y fin de cada actividad. Sin embargo este procedimiento es propenso a sesgos de memoria y subjetividad e incluso el desarrollo mismo de los diarios es un inconveniente para los sujetos en prueba, ya que interfiere con el desarrollo de la actividad [13]. En el desarrollo de metodologías automáticas, un enfoque común es el uso técnicas de procesamiento digital de imágenes (*DPI-Digital Image Processing*) y visión por computador (*CV-Computer Vision*). Estas técnicas utilizan procedimientos de segmentación de imagen, extracción de características y aprendizaje de máquina [11, 21, 22]. Una técnica popular es la substracción de fondo de la escena [23, 24], este método permite separar el fondo

de los objetos en movimiento, sin embargo este tipo de enfoques son sensibles a las condiciones fotométricas de la escena.

En la extracción de características generalmente se utilizan descriptores tales como forma, silueta, colores, poses y movimientos del cuerpo [22, 25, 26]. Un enfoque común, es el cálculo del volumen durante una ventana de tiempo (STV-Spatio Temporal Volume) [22, 27, 28], aunque el método puede capturar la continuidad de la acción humana, se limita a actividades que no son periódicas y es sensible a las oclusiones parciales del objetivo. La transformada discreta de Fourier (DFT-*Discrete Fourier Transform*) [25], ha sido ampliamente utilizada para representar información acerca de la estructura geométrica del objeto, sin embargo este enfoque también es muy sensible a la oclusión parcial o total del objetivo que se está analizando en la escena.

Otras técnicas han sido utilizadas para obtener robustez ante cambios de escala, rotación, apariencia y oclusiones parciales del objetivo en la escena [29, 16, 30, 31], entre los métodos más populares se encuentran la transformada característica invariante en escala (SIFT-*Scale Invariant Feature Transform*) [32, 33] y el descriptor de histogramas de gradientes orientados (HOG-*Histogram Oriented Gradient*) [26].

La actual proliferación de cámaras de profundidad ha permitido agregar información relevante a las metodologías de detección de actividad. El sensor más común para este tipo de aplicaciones es el Kinect®[®], ya que es popular, de fácil acceso y posee librerías que permiten estimar los datos de profundidad de la escena, información acústica e incluso la estimación de los puntos articulados del cuerpo. Enfoques como [34, 35, 10, 17] utilizan las coordenadas espaciales 3D de los principales puntos articulados, sin embargo este tipo de metodologías son susceptibles a oclusiones parciales del objetivo en la escena [36, 37].

Algunos enfoques utilizan sensores de movimiento como acelerómetros, giroscopios o unidades de masa inercial (IMU) [12, 38, 39]. M. Bocksch y J. Jahn proponen una metodología de clasificación utilizando IMU de 9 grados de libertad obteniendo resultados hasta del 100% de acierto para algunas actividades, sin embargo sus algoritmos pueden requerir altas capacidades de procesamiento y de memoria en el dispositivo [40]. Aunque los sensores de movimiento como

acelerómetros o IMU han permitido el desarrollo de metodologías con buenos resultados, requieren del uso de más de un sensor para observar la actividad. Además, se ha demostrado, que este tipo de sensores no son adecuados para el reconocimiento de actividades con movimientos similares, como el bajar escaleras o subir escaleras [3]. Otro sensor utilizado es el electrodo de registro electromiográfico (EMG), el cual miden la contracción y extensión de los músculos del cuerpo. Múltiples metodologías se han propuesto en la literatura [41, 5, 42, 43, 44], estos métodos usan técnicas del área de procesamiento digital de señal como Mean Absolute Value (MAV), LDF, transformadas Wavelets, Mel Frequency Ceptrum, entre otros [45, 46], logrando resultados hasta del 98 %. Aunque estas metodologías presentan buenos desempeños y son robustas a oclusiones parciales del objetivo, tienen el mismo problema de requerir más de un sensor para detectar un conjunto significativo de actividades.

Es evidente que un solo tipo sensor no es suficiente para caracterizar una amplia variedad de actividades físicas humanas [6], por lo que la fusión multimodal es uno de los enfoques que permite mejorar el factor discriminante de una observación aprovechando las ventajas individuales de cada sensor [7, 8, 9, 10]. Aunque se han diseñado metodologías eficientes donde se combinan diferentes grupos de sensores, no se evidencia una metodología que realice la fusión de los sensores más utilizados para estas tareas, como lo son, el Kinect®[®], los IMU y los EMG [2, 8, 10]. Esto hace pensar que es necesario desarrollar un método que utilice las ventajas de cada sensor individual para obtener un mayor rendimiento que en cualquiera de sus combinaciones.

1.2. JUSTIFICACIÓN

Dada la importancia del reconocimiento de actividad física en las diferentes áreas de investigación y desarrollo, se puede pensar que hay un creciente interés en desarrollar aplicaciones que permitan el monitoreo y la detección de un conjunto amplio de actividades físicas humanas. En la literatura [11, 12, 13, 14, 15, 16, 17] se han desarrollado múltiples metodologías donde se utilizan diferentes modalidades de sensores, tales como sensores de profundidad, sensores electromiográficos, goniómetros, giroscópicos, acelerómetros, entre otros; para la detección au-

tomática de actividad. Sin embargo se considera que un solo tipo de sensor no brinda las características necesarias para lograr clasificar una amplia variedad de actividades [6]. Por lo que se piensa, que al agregar sensores con diferente naturaleza, se puede ampliar la percepción y la variedad de las actividades a ser reconocidas.

La combinación de diferentes fuentes sensoricas hace referencia a técnicas de fusión de datos o fusión de sensores. Este tipo de enfoque brinda la posibilidad de aumentar el desempeño de las técnicas aprovechando las ventajas individuales de cada sensor. Esto ha permitido aplicar con éxito la fusión de datos en diferentes campos de la ingeniería como la robótica, medicina e industria militar, entre otros [47]. Específicamente en el área del reconocimiento automático de actividad física humana, se han propuesto metodologías que fusionan la información proveniente de sensores IMU y Kinect obteniendo métodos robustos a oclusiones parciales [48] o métodos con mayor desempeño [49]. La fusión de sensores también ha sido aplicada a otros campos de la ciencia logrando resultados bastantes satisfactorios, por ejemplo, el diseño de interfaces hombre-computador (*HCI-Human Computer Interface*) [50, 9, 51, 52].

Aunque se han diseñado metodologías eficientes para el reconocimiento de actividad e incluso muchas de ellas con enfoque de fusión de datos, no existe ningún método en la literatura que realice la fusión de datos de los tres sensores más utilizados en la literatura, como lo son Kinect®[®], IMU y EMG. Esta conclusión, se puede considerar como una gran oportunidad para innovar en el campo de reconocimiento de actividad, haciendo usos de las técnicas ya exploradas y adaptándolas para realizar una fusión multimodal, que podría llevar al desarrollo de una metodología que aproveche las ventajas de cada sensor y aumente el desempeño en la detección.

1.3. OBJETIVOS

1.3.1. Objetivo General

Determinar una metodología para la detección automática de actividades físicas humanas fusionando la información provista por sensores de profundidad, de masa inercial y de registro

electromiográfico.

1.3.2. Objetivos Específicos

1. Desarrollar una metodología para la captura de datos, sincronización, ubicación de sensores, acondicionamiento de señal y creación de una base de datos local anotada que utilice las tres modalidades de sensores descritas.
2. Determinar una metodología para la identificación de movimientos utilizando independientemente cada uno de los sensores.
3. Determinar una metodología de fusión de datos para la clasificación de actividad utilizando la secuencia temporal de movimientos identificados por cada sensor.
4. Verificar estadísticamente el rendimiento del sistema y determinar el conjunto mínimo de sensores necesario para mantener la confiabilidad.

2. ESTADO DEL ARTE

En las ciencias de la computación, se han desarrollado diferentes investigaciones sobre el reconocimiento de la actividad física humana. A continuación, se desea realizar un recorrido por los avances y contribuciones hechas al desarrollo de metodologías, que realicen el reconocimiento automático de actividades físicas humanas en un ambiente no estructurado.

2.1. Cámaras de Video y cámaras de Profundidad

Algunos trabajos se han centrado en detectar automáticamente acciones humanas aplicando metodologías basadas en el análisis de secuencias de vídeo, estos enfoques usan técnicas de procesamiento digital de imágenes y visión por computador, donde la estructura general de método se constituye por etapas de segmentación, extracción de características, clasificación, seguimiento y reconocimiento de actividad [11].

2.1.1. Procesamiento y extracción de características

Aplicando análisis de vídeo, [22, 28, 27] se aplica el método STV logrando un factor discriminante en la clasificación de actividad. En [23, 24] utilizan técnicas de substracción de fondo de la escena utilizando modelos de mezclas de gaussianas para separar el fondo de los objetos en movimiento. El crecimiento acelerado de tecnologías concernientes a sensores de profundidad, ha permitido incrementar los desempeños de las metodologías y llevar a cabo un seguimiento en tiempo real. Los enfoques más comunes [2, 34, 35, 10], estiman las coordenadas espaciales de los puntos articulados del cuerpo, obteniendo un vector de características que permite obtener un hiperplano de separación entre las diferentes actividades a detectar. Para 2012, Kang Li y Yun Fu [20], aplican un método que utiliza la información de profundidad del dispositivo durante una ventana de tiempo, para así calcular el histograma de velocidad orientada proporcionando una representación continua de la actividad física humana. En 2014 Wu y Wei Pan [17], proponen utilizar coordenadas polares para la caracterización de los puntos articulados y así lograr un mayor desempeño en la clasificación. Métodos como histogramas de gradiente

orientado 3D [53] o SIFT 3D [54], utilizan la descripción de una nube de puntos 3D para el reconocimiento, aunque estas metodologías reportan desempeños eficientes, su implementación requiere de un gran costo computacional, lo que limita su implementación con algunos sistemas embebidos [55].

2.1.2. Clasificación

En 2011, K. Biswas y Saurav Kumar Basu, proponen un método utilizando Kinect® y SVM para clasificación movimientos, lo que permite el uso eficiente de la profundidad de la cámara para reconocer con éxito múltiples gestos humanos [56]. En 2013, Thi-Lan y Minh-Quoc exploran la capacidad de utilizar la información proporcionada por el esqueleto de Kinect para el reconocimiento postura humana con clasificador SVM multiclase en un contexto de un marco de vigilancia de salud. Se realizan la estimación de 7 experimentos diferentes con 4 tipos de características extraídas de humanos esqueleto, los resultados obtenidos muestran que este dispositivo puede detectar con gran precisión posturas como acostado, sentado, de pie, y flexión [57].

2.2. Acelerómetros y unidades de Masa Inercial

Otros enfoques utilizan la información de acelerómetros o unidades de masa inercial (IMU) para el reconocimiento de actividad, caracterizando los cambios velocidad y aceleración que se producen al ejecutar un movimiento [58].

2.2.1. Procesamiento y extracción de características

En el 2000, Foerster y Fahrenberg [15], recogieron datos de 31 sujetos utilizando cinco acelerómetros y construyeron un modelo de clasificación jerárquica para identificar diferentes posturas corporales y movimientos. En 2004 Bao, aplica la FFT como método de caracterización de señal y adiciona el cálculo de características estadísticas como media, desviación estándar, energía y correlación, logrando hasta un 90 % de acierto en la clasificación [12]. Posteriormente en

2011, Mi Zhang y Alexander A. Sawchuk, caracterizan la información de los sensores IMU, basándose en los parámetros físicos del movimiento humano, obteniendo resultados hasta del 90 % de acierto en la clasificación [39].

2.2.2. Clasificación

En 2011, Mannini y Sabatini proponen la clasificación de 7 actividades a partir de 5 acelerómetros biaxiales ubicados en diferentes partes del cuerpo, esta metodología usa Wavelets y un clasificador HMM para lograr resultados de un 98,4 % de eficiencia [59]. En 2012 David Martín de Castro, propone el uso de la información de los acelerómetros y el GPS de un *smartphone* para clasificar actividad utilizando como método de aprendizaje redes neuronales artificiales [38]. M. Bocksch y J. Jahn proponen una metodología de clasificación utilizando un sensor IMU de 9 grados de libertad obteniendo resultados hasta del 100 % de acierto para algunas actividades, sin embargo sus algoritmos son computacionalmente costosos y pueden requerir altas capacidades de procesamiento y de memoria en el dispositivo [40].

Otros enfoques se basan en la detección de movimientos primitivos, los cuales se inspiran en las técnicas de reconocimiento automático de voz. En estas metodologías, las frases se dividen en palabras aisladas, para luego dividirse en una secuencia de fonemas. De este modo se detectan los fonemas comunes a todas las palabras construyendo los bloques básicos para el reconocimiento de palabras y frases de manera jerárquica. En 2012 Mi Zhang y Alexander A. Sawchuk [60], utilizan esta idea para obtener un clasificador basado en movimientos primitivos aplicando técnicas de caracterización como Bag-of-Features (BOF) obteniendo un rendimiento significativo.

2.3. Sensores Electromiográficos

Los sensores electromiográficos (EMG) han sido ampliamente utilizados en este tipo de aplicaciones, debido a la capacidad de correlacionar el sensado de contracción y flexión de los músculos del cuerpo con el movimiento humano.

2.3.1. Procesamiento y extracción de características

Los primeros en clasificar estas señales EMG con razonable éxito (75 % de acierto) fueron Graupe y Cline en 1975 [41], ellos utilizaron técnicas de caracterización como mean absolute value (MAV) y linear discriminant function (LDF) para la clasificación de actividad. En 1983 Doerschuck utilizando las mismas técnicas desarrolladas por [41], pero aplicando técnicas de correlación entre la información de múltiples electrodos lograron un 83.5 % de acierto [43]. En 1995 W. Kang utilizó MFCC obteniendo un acierto del 85 % [42]. También se han aplicado descriptores tiempo-frecuencia como Wavelets [45, 51, 46], obteniendo resultados mayores al 90 % de acierto.

2.3.2. Clasificación

En [61] propone usar Modelos Ocultos de Markov (HMM) para el reconocimiento. En este enfoque se utilizan redes neuronales para realizar la segmentación de movimientos y un HMM para realizar fusión de datos y obtener el reconocimiento de la actividad corporal logrando una resultados entre el 80 % y el 100 % de acierto para algunas actividades, sin embargo otras actividades no registraron el rendimiento deseado.

2.4. Fusión de Datos

Algunos enfoques aplican la fusión de sensores multimodales, para así aprovechar las ventajas de cada sensor. En [7] se propone un método utilizando la fusión de una red de sensores IMU y un sensor Kinect®. Este enfoque utiliza un sistema MIMO para el mapeo de las señales de la red IMU y la fusiona con los datos de la nube de puntos provista por el Kinect® para ser clasificada por una SVM, la cual fue entrenada con una base de datos con más de 40000 posturas. Esta técnica permite obtener los datos que son visibles para la cámara y así evitar sesgos en la estimación de la postura.

3. MARCO TEÓRICO

Para el desarrollo de este proyecto, se utilizan técnicas aprendizaje de máquina y procesamiento digital de señales en 1D y 2D, el cálculo de transformadas y diferentes descriptores, por lo que es muy importante la revisión de algunas teorías fundamentales para el entendimiento de la metodología. En este capítulo se realizará un repaso de los conceptos y teorías básicos para el desarrollo de este trabajo.

3.1. Dispositivos para la captura de movimiento

Usualmente los dispositivos utilizados para el registro de actividades físicas son sensores de movimiento, estos son dispositivos electrónicos equipados para obtener una respuesta eléctrica cuando se ejecuta una acción física. Los sensores más comunes para este tipo de tareas son los acelerómetros, giroscopios, unidades de masa inercial (IMU), sensores electromiográficos (EMG) y dispositivos Kinect®. En esta sección se realizará un repaso de su principio de funcionamiento.

3.1.1. Sensor

Un sensor es un dispositivo que permite la transducción de cualquier magnitud física o química en una señal eléctrica aprovechando las propiedades eléctricas de la materia. Un sensor se diferencia de un transductor, ya que el sensor siempre transforma cualquier tipo de energía en energía eléctrica y un transductor la transforma en cualquier tipo de energía.

3.1.2. Unidad de masa inercial (IMU)

Una unidad de medición inercial, es un sensor electrónico que mide aceleración, orientación y fuerzas gravitacionales ejercidas sobre el dispositivo, usando la combinación de acelerómetros y giroscopios. Las unidades de medición inercial son normalmente usadas para aplicaciones de control e instrumentación, en el análisis de vibraciones y estimación de la posición y orientación de un objeto [58].

3.1.3. Kinect

El Kinect es un dispositivo diseñado para la detección de movimiento e interacción de un usuario con la consola de videojuegos Xbox 360. Es desarrollado por la empresa Microsoft® y permite a los usuarios controlar e interactuar con la consola sin necesidad de tener contacto físico. Este dispositivo puede proporcionar una variedad de datos basado en la información de una escena, entre algunas de sus características más importantes, se encuentra el mapa de profundidad de la escena, el cálculo de los puntos articulados de la silueta humana, entre otros [62].

3.1.4. Electromiografía (EMG)

La electromiografía es una técnica para medir la actividad eléctrica producida por los músculos esqueléticos. Esta se desarrolla utilizando electrodos de superficie, los cuales miden la actividad muscular esquelética transmitida a la superficie de la piel, obteniendo así un nivel de voltaje en el punto de medido. Este tipo de sensor sólo permite evidenciar la contracción o relajación de conjunto de músculos medidos, no recogiendo directamente la actividad de las diferentes unidades motoras. Este tipo de electrodo es el más utilizado para recolectar este tipo de medidas y consiste en un adhesivo que se adhiere a la piel permitiendo sensar los músculos donde el sensor fue ubicado [52].

3.2. Extracción de características

Consiste en la búsqueda de atributos o cualidades que facilitan la identificación de una entidad en particular, como fenómenos físicos, comportamientos, movimientos, entre otros. En el caso específico de las ciencias de computación e ingeniería, la extracción de características, se realiza aplicando descriptores matemáticos. Los descriptores matemáticos dependen específicamente de la naturaleza de los datos a analizar. Algunos de los descriptores más comunes son los: temporales, espaciales, representaciones en frecuencia, bases ortogonales, entre otros [63].

3.2.1. Transformada Wavelet Discreta (*DWT-Discrete Wavelet Transform*)

Esta técnica nace como una alternativa para superar los problemas de resolución de la transformada de Fourier de tiempo corto (*STFT-Short Time Fourier Transform*), haciendo posible una representación de una señal tanto en tiempo como en frecuencia de forma simultánea. Básicamente, la idea de la transformada Wavelet es filtrar una señal en el dominio del tiempo mediante filtros paso bajo y paso alto para eliminar ciertas componentes de alta o baja frecuencia de la señal, el procedimiento se repite para las señales resultantes del proceso de filtrado y a esta operación se le denomina descomposición [64].

Dada la función discreta $f[n]$, la cual se define como la señal a analizar, su transformada Wavelet $c[i, k]$, se define como:

$$c[i, k] = f[n] \Psi_{i,j}[n] \quad (1)$$

Donde, $\Psi_{i,j}$ es una wavelet discreta definida como:

$$\Psi_{i,j}[n] = 2^{-\frac{j}{2}} \cdot \Psi[2^{-j}n - k] \quad (2)$$

Los parámetros τ y s , están definidos según la escala diádica, de manera que:

$$\tau = 2^j \quad (3)$$

$$s = 2^j k \quad (4)$$

3.2.2. Mel Frequency Cepstrum Coefficients (MFCC)

Es una técnica de caracterización introducida por Davis y Mermelstein (D&M) en 1980 [15], la cual consiste en combinar filtros triangulares y perceptualmente distribuidos, con la transformada discreta del coseno del logaritmo de las energías de salida de los filtros. El procedimiento de cálculo de los MFCC se basa en el siguiente procedimiento [65, 66]: Segmentar la señal utilizando ventanas, luego a cada segmento, se le debe aplicar la transformada de Fourier discreta

(DFT-*Discrete Fourier Transform*), después se aplica el banco de filtros correspondientes a la escala Mel, al espectro obtenido en el paso anterior y sumar las energías en cada uno de ellos, por último se calcula el logaritmo de todas las energías de cada frecuencia Mel y se aplica la transformada de Coseno discreta a estos logaritmos.

La ecuación que permite el cálculo de estos coeficiente esta dada por:

$$MFCC_i = \sum_{k=1}^{n_f} X_k \cos \left[i \left(k - \frac{1}{2} \right) \frac{\pi}{n_f} \right] \quad (5)$$

Donde:

- i es el número de coeficientes a calcular.
- X_k es el logaritmo de la energía de salida del $k - \text{ésimo}$ filtro.
- n_f es el número de filtros a utilizar en la transformada.

3.3. Clasificación

Es un proceso de estimación matemático que tiene como objetivo encontrar un modelo matemático que permita separar o identificar una entidad en particular utilizando un conjunto de ejemplos o información suministrada. El proceso de clasificación se puede realizar utilizando clasificación supervisada, clasificación no supervisada, aprendizaje por esfuerzo, transducción o aprendizaje multi-tarea [1, 67]. Algunos de los clasificadores más comunes en este tipo de tareas son: los vecinos más cercanos (KNN-*K nearest neighbors*), las máquinas de vectores de soporte (SVM), los modelos ocultos de Markov (HMM), las redes neuronales artificiales (ANN-*Artificial Neural Network*), clasificador Bayesiano, entre otros [68].

3.3.1. Clasificación no supervisada [1]

Son metodologías de agrupamiento que tienen como objetivo asignar una etiqueta a una observación, ajustando las fronteras entre las diferentes clases a identificar. Se distingue de la

clasificación supervisada, porque no hay un conocimiento a priori del comportamiento de las observaciones. En síntesis su objetivo, es agrupar entidades que pertenezcan a un mismo conjunto.

3.3.2. Clasificación supervisada [1]

Son técnicas de aprendizaje que parten de un conocimiento a priori sobre el comportamiento de la observación a ser etiquetada. Estas técnicas en general buscan la estimación de los parámetros de un hiperplano de separación que permita establecer un límite entre las etiquetas asignar.

3.3.3. Máquinas de vectores de soporte (SVM)

Las máquinas de soporte vectorial son una metodología de clasificación asistida basada en la teoría de Vapnik [69]. Este método tiene como objetivo la estimación de un modelo de aprendizaje que permita la clasificación de un evento binario, su formulación matemática depende de la naturaleza de los datos; es decir que existen diferentes formulaciones dependiendo de la linealidad o de la distribución de sus datos. En su forma más sencilla una SVM tiene como objetivo encontrar el hiperplano que separe y maximice el margen entre las clases en este espacio [70]. La SVM se considera un clasificador binario, el cual calcula una etiqueta $y \in \{-1, 1\}$ utilizando el siguiente modelo:

$$y = w^T \phi(x) + b \tag{6}$$

Donde, $\phi(x) : \mathbb{R} \rightarrow H$ es una transformación del conjunto de entrada a un conjunto con igual o mayor dimensión, lo que se supone que facilita la separación de las clases. El vector w define el hiperplano de separación en en espacio $\phi(X)$. b representa el bias o sesgo con respecto al origen de coordenadas.

La SMV aborda el problema de clasificación relajando el concepto de margen, para lo que se introduce variables de sesgo $\xi_i \geq 0$, donde la función de costo o función fitness se define como:

$$c(y, f(x, w)) = \begin{cases} 0 & \text{si } y = \text{sng}(w^T \phi(x) + b) \\ 1 & \text{si } y \neq \text{sng}(w^T \phi(x) + b) \end{cases} \quad (7)$$

Donde existe el riesgo empírico $R_{emp} = \sum_{i=1}^n \xi_i^\sigma$ [71], siendo generalmente $\sigma = 1$ en problemas de clasificación. Por lo tanto el problema de estimación del hiperplano de separación, se define como la solución al siguiente problema de optimización cuadrático (QP-*Quadratic Problem*):

$$\min_{w, b, \xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (8)$$

$$\text{Sujeto a } y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \quad \forall_i = 1, \dots, n \quad (9)$$

$$\xi_i \geq 0 \quad \forall_i = 1, \dots, n \quad (10)$$

Donde:

- x_i es una muestra del conjunto de entrenamiento con su respectiva etiqueta.
- C se define como una constante de regularización.

Buscando el hiperplano óptimo en (9), se requiere resolver un problema QP, que puede ser resuelto construyendo un Lagrangiano y transformándolo en un problema dual del problema de optimización. Por lo que el problema de optimización se define como:

$$\max_{\alpha_i} L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \phi^T(x_i) \phi(x_j) \quad (11)$$

$$\text{sujeto a } \sum_{j=1}^n \alpha_j y_j = 0 \quad (12)$$

$$0 \leq \alpha_i \leq C \quad \forall_i = 1, \dots, n \quad (13)$$

3.3.4. Máquinas de vectores de soporte multiclase

Las SVM son originalmente son clasificadores formulados para problemas binarios, donde solo existen 2 clases, no obstante, en la práctica es común encontrar problemas en los que se debe clasificar $k > 2$ distintas, donde k es el número de clases a clasificar. Por lo que existen diferentes metodologías para lograr una extensión de las SVM a la clasificación multiclase.

Entre las metodologías más comunes se encuentran los siguientes enfoques [72]:

Uno contra todos: Esta estrategia consiste en la estimación de una SVM por clase, la cual está en capacidad para distinguir el conjunto de datos que pertenece a una clase en particular del conjunto de datos de todas las clases restantes. Por lo general, cuando hay un ejemplo de una clase en particular, una sola SVM debe detectar positivo un ejemplo y por medio de codificación se debe asignar la etiqueta. Cuando más de un modelo SVM clasifica un conjunto de datos positivo, se calcula la probabilidad de pertenencia a la clase y se escoge la clase con máxima probabilidad.

Uno contra uno: Esta estrategia de clasificación consiste en construir un clasificador por par de clases, la pertenencia a la clase se escoge buscando la clase que recibió la mayoría de votos, en caso de un empate, entre dos clases con el mismo número de votos, se calcula la probabilidad de pertenencia a la clase y se escoge la clase con máxima probabilidad. Para implementar esta metodología se requiere de m clasificadores, este método es generalmente más lento que uno contra todos. El cálculo de m se define como:

$$m = \frac{n_{clases}(n_{clases} - 1)}{2} \quad (14)$$

3.3.5. Estimación de probabilidades en SVM:

Dependiendo de la solución multiclase tomada, las máquinas de soporte vectorial pueden estimar la probabilidad a posteriori entre dos clases o la pertenencia de una clase con respecto al resto de las clases. El procedimiento del cálculo de probabilidades a posteriori es [73]:

1. Obtener la probabilidad de que la observación evaluada sea perteneciente a cada clase.
2. Convertir estas probabilidades binarias a probabilidades multiclase.

En el método uno contra todos, el paso número dos puede ser reducido a una normalización simple de modo que la suma de probabilidades a posteriori sea igual a uno, ya que el número de SVM binarias es el mismo del número de clases. El modo que se usa usualmente para transformar la salida de una SVM en una probabilidad binaria de pertenencia a la clase, esto es posible utilizando la función sigmoide. Para asignar las funciones de densidad de probabilidad condicional de la SVM $p(f|y = 1)$ y $p(f|y = -1)$ se aplica el teorema de Bayes:

$$P(y = 1|f) = \frac{p(f|y = 1)P(y = 1)}{\sum_{i=\pm 1} p(f|y = i)P(y = i)} \quad (15)$$

Al sustituir para las funciones de densidad de probabilidad condicional, se obtiene [74]:

$$P(y = 1|f) = \frac{1}{1 + e^{(af^2+bf+c)}} \quad (16)$$

En [74] se plantea un modelo paramétrico para la probabilidad binaria a posteriori. Se diferencia entonces de (16), en que los parámetros a y b son estimados maximizando la verosimilitud [75]. Esta ecuación concede directamente probabilidades multiclase en el caso del método de uno contra todos. Si existen k clases diferentes, la probabilidad a posteriori de la clase i -ésima se calcula como:

$$P(y = i|x) = \frac{1}{1 + e^{(A_i f_i(x) + B_i)}} \quad (17)$$

Donde $f_i(x)$ es la salida de la SVM binaria que clasifica la clase i contra el resto.

3.3.6. Modelos Ocultos de Markov (HMM)

Los HMM son autómatas de estados finitos que permiten modelar procesos estocásticos, donde la transición a un estado en particular depende de una distribución de probabilidad. Los estados no son visibles en general y su ocurrencia depende del estado anterior. Esta metodología, asume

que el sistema a modelar es un proceso de Markov de parámetros desconocidos y su objetivo es determinar estos parámetros a partir de diferentes observaciones [76]. Un HMM se define como:

$$\lambda = f(O, Q, A, B, \pi)$$

Donde, $O = \{O_1, O_2, \dots, O_M\}$ son los estados visibles, $Q = \{1, 2, \dots, N\}$ son los estados ocultos. $A = \{a_{ij}\}$ es la matriz de transición de estados, donde el elemento a_{ij} es la probabilidad de que genere la transición desde el estado i al estado j , $B = \{b_i(k)\}$ se define como la matriz de probabilidad de salida, donde $b_i(k)$ es la probabilidad de emitir el símbolo O_k en el estado i , $\pi = \{\pi_i\}$ es el vector de probabilidades del estado inicial, donde π_i es la probabilidad de iniciar en el estado i .

Los tres problemas básicos de los HMM

Cuando se utilizan Modelos ocultos de Markov es necesario plantear tres problemas básicos, donde la solución de ellos permite, que esta metodología sea adecuada para representar la señal [76]:

El problema de la evaluación: Calcular eficientemente $P(O|\lambda)$ la probabilidad de la secuencia de observación O dado el modelo $\lambda = (A, B, \pi)$ y la secuencia de observación $O = \{O_1, O_2, \dots, O_T\}$. Entonces dado un modelo y una secuencia de observaciones de la señal, la evaluación se reduce a calcular la probabilidad de que dicha secuencia de observaciones haya sido generada por el modelo. Es posible solucionar este problema utilizando el algoritmo de avance-retroceso [77].

El problema de la decodificación: Encontrar la trayectoria $q = \{q_1, q_2, \dots, q_T\}$ más probable dado el modelo λ y la secuencia de observación $O = \{O_1, O_2, \dots, O_T\}$. Dado un modelo y una secuencia de observaciones, ¿Cuál es número de estados óptimo que se debe seguir para generar una salida igual o similar a la secuencia de observaciones?. Para solucionar este problema se calcula la máxima verosimilitud [78].

El problema del aprendizaje: Ajustar los parámetros A, B, π con el fin de maximizar $P(O|\lambda)$. Para la solución de este problema, se utiliza un procedimiento iterativo llamado

Método de Baum-Welch [79].

Libro de Código (codebook): Es el proceso de transformar el contenido del conjunto de entrenamiento en una representación codificada, donde se analiza cierta cantidad de muestras y se busca elementos que pueda ser agrupados y representados como un solo código. Este procedimiento es llevado a cabo reuniendo los patrones típicos de la señal y designando un código a cada patrón. Esta cuantificación trae consigo la idea de clasificar los patrones y luego buscar, los mejores representantes para reducir el tamaño del conjunto de entrenamiento. Para realizar el agrupamiento en el libro de códigos, se utilizan algoritmos basados en clustering, la técnica de K-means, permite ubicar y calcular los centroides de los vectores de observación, agrupando estos en torno a K valores medios arrojando como resultado un libro de códigos o codebook [76].

3.3.7. Análisis de rendimiento de un clasificador

El rendimiento de un clasificador, se define como un indicador de la capacidad de etiquetar una observación correctamente. Estos indicadores se definen como [68]:

Verdaderos Positivos (v_p): Se definen como los ejemplos que pertenecen a la clase y el clasificador estimó que pertenecen a esa clase.

Verdadero Negativos (v_n): Se definen como los ejemplos que no pertenecen a la clase y el clasificador estimó que no pertenecen a esa clase.

Falsos Negativos (f_n): Son los casos que el clasificador no los clasificó en la clase a la que eran pertenecientes.

Falsos Positivos (f_p): Son los casos que el clasificador los clasificó en una clase a la que no eran pertenecientes.

Especificidad: Es el cociente entre el número de observaciones positivas incorrectamente clasificadas y el número total de observaciones. Este parámetro se puede calcular como:

$$Especificidad = \frac{f_n}{v_p + f_n} \quad (18)$$

Sensibilidad: Cociente entre el número de observaciones positivas correctamente clasificados y el número total de observaciones. Este parámetro se puede calcular como:

$$\text{Sensibilidad} = \frac{v_p}{v_p + f_n} \quad (19)$$

3.3.8. Intervalo de confianza

Cuando se tiene una variable aleatoria con distribución normal, el intervalo de confianza se define como la posible desviación que una medida normalizada puede presentar dada su valor probable de acierto. La probabilidad de éxito en la estimación se representa con $1 - \alpha$ y se denomina nivel de confianza. En estas circunstancias, α se define como error aleatorio y es una medida de la posibilidad de fallar en la estimación mediante tal intervalo [80].

Entonces el error aleatorio para una confianza para del 95 % se define como:

$$e = 1,96 \sqrt{\frac{p(1-p)}{n}}$$

Donde

- p es la probabilidad de acierto normalizada.
- n es la muestra poblacional o el número de datos a tener en cuenta.

El error para una confianza del 99 % se define como:

$$e = 2,575 \sqrt{\frac{p(1-p)}{n}}$$

4. HARDWARE Y CAPTURA DE DATOS

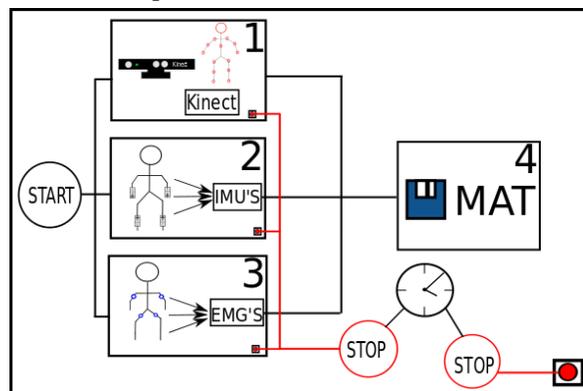
En diferentes áreas de investigación, es necesario el análisis de acciones cotidianas que involucren el desarrollo de actividades físicas. El análisis de estas acciones requiere de una base de datos anotada que permita evaluar, en igualdad de condiciones, los comportamientos estáticos y dinámicos de las señales registradas por las diferentes modalidades de sensores. Usualmente, los sensores utilizados para el registro de actividades físicas humanas son sensores de movimiento, tales como acelerómetros, giroscopios, unidades de masa inercial (IMU), sensores electromiográficos (EMG) y dispositivos Kinect.

En el estado del arte existen diferentes bases de datos anotadas donde se han registrado la ejecución de actividades físicas humanas capturando datos provistos por este tipos de sensores; algunas de la bases de datos son: Human Activity Database – CHAD, KTH, WEIZMANN [81], UCF-Sports [82], Hollywood [83], base de datos UCiI-Smartphones Data Set [84] y RGB-D Human Activity Recognition and Video Database [85]. Aunque existen bases de datos estandarizadas que facilitan el análisis de reconocimiento de actividad, no existe una base de datos anotada que realice el registro sincronizado de sensores IMU, sensores EMG y dispositivos Kinect®), lo que fuerza a construir una base de datos que cumpla con el requerimiento anterior. Este hecho conlleva al diseño de un sistema de captura sincronizado que garantice el correcto registro de las señales obtenidas por los diferentes sensores.

Para lograr un registro sincronizado de los datos, se debe implementar la captura de los sensores con programación multi-hilo, ya que este tipo de programación permite ejecutar códigos en paralelo permitiendo iniciar todas las capturas al mismo tiempo e ir capturando información de cada sensor por separado. Labview es un entorno multi-hilo por naturaleza, además ofrece la versatilidad y robustez para trabajar con sistemas de adquisición de datos (DAQ), en especial con dispositivos Kinect®, Wii-motion y tarjeta DAQ de la *National Instrument*. Por las razones anteriores el sistema fue diseñado en Labview y se puede observar en la figura 1, donde se evidencian cuatro etapas fundamentales para el registro sincronizado de estas señales; estas etapas constan de:

1. La captura sincronizada de la red de sensores IMU.
2. La captura sincronizada de la red de sensores EMG.
3. La captura de los puntos articulados del cuerpo utilizando el Skeleton Tracking provisto por el Kinect.
4. El almacenamiento de los datos en la estructura binaria .mat utilizada por entornos como Matlab y Octave para el almacenamiento de datos.

Figura 1. Etapas del sistema de captura sincronizado



4.1. Captura sincronizada de la red de sensores IMU, red de sensores EMG y dispositivo Kinect® en LabVIEW.

En esta etapa se diseñó un sistema de adquisición de datos que consta de la captura sincronizada de cuatro acelerómetros tri-axiales ADLX330, ocho electrodos de superficie en conexión diferencial (hecho que permite obtener cuatro medidas EMG) y un dispositivo Kinect®.

4.1.1. Adquisición de sensores EMG

Se utilizaron electrodos de superficie adheridos a la piel generando un mínimo de riesgo sobre el usuario; a estos sensores se les añade gel conductor con el fin de generar un efecto químico que mejore la captura de la señal. Este gel está compuesto de plata - cloruro de plata

(Ag – AgCl) y es el producto más común para este tipo de tareas [52]. El sistema consta de un equipamiento encargado de almacenar el DAQ para los sensores EMG; todos los cables se adhieren al cuerpo de la persona y son conducidos hacia el equipamiento de almacenamiento mediante comunicación USB con el dispositivo de procesamiento y almacenamiento. Debido al uso de electrodos, este sistema cumple las normas de aislamiento para protección del usuario EN60601-1 (*Medical Electrical Equipment, General Requirements*) y EN60601-2-40 (*Specification for electromyographs and evoked response equipments*).

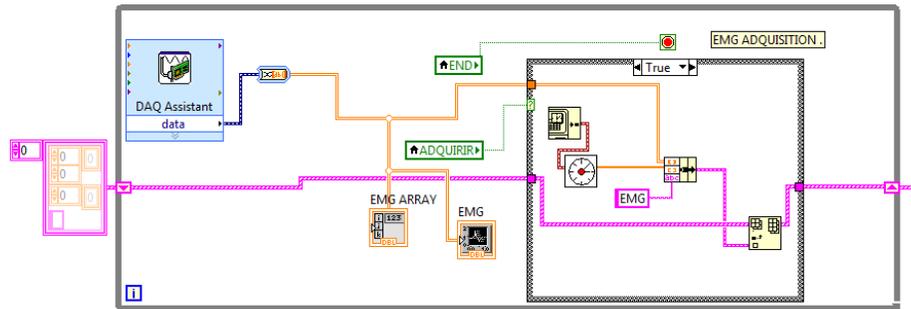
La señal de los sensores EMG es obtenida por medio de los drivers de National Instrument para las DAQ, obteniendo cuatro arreglos unidimensionales correspondientes a cada sensor registrado; esta información es almacenada durante cada instante de muestreo. El acondicionamiento de este equipo de adquisición consta de un amplificador de instrumentación, el cual se encarga de referenciar la señal diferencial proveniente de los electrodos medidos y del electrodo de referencia; posteriormente esta señal pasa por una etapa de amplificación, luego por un filtro pasa-banda Sallen Key y por último se realiza el respectivo aislamiento eléctrico utilizando un ISOOPAM-AD2010JN. El diseño circuital se puede observar en el anexo 9.1. Este acondicionamiento se conecta a una tarjeta de adquisición de datos NI6008, con resolución de 12 bits, rango de entrada de [0-5] [V] y frecuencia de muestreo de 2 [kHz], permitiendo al software realizar el procesamiento respectivo. El tamaño de los arreglos de cada señal es ampliamente descrito en la sección 5.2.

En la figura 2 se puede observar, el digrama de bloques diseñado en LabVIEW para esta etapa de la adquisición, en esta figura se puede observar los VI correspondientes a la adquisición de señal con la DAQ y su almacenamiento sincronizado por medio de una estructura de selección como lo es un *case* y una bandera de adquisición global.

4.1.2. Adquisición de sensores IMU

Los sensores IMU utilizados son acelerómetros ADXL330 que permiten sensar los cambios de velocidad de los puntos articulados configurados en el experimento. Este sensor es un acelerómetro acondicionado de 3 grados de libertad (a_x, a_y, a_z) con salidas de voltaje para medir

Figura 2. Adquisición de sensores EMG



en un rango de $\pm 3g$. Para lograr obtener un sensor IMU, y debido a que este dispositivo sólo posee tres grados de libertad, se realizó el cálculo de las orientaciones *Roll* y *Picth* utilizando la conversión a coordenadas esféricas, transformación que se explica detalladamente en la sección 5.2.2. Para la instrumentación y comunicación inalámbrica de este acelerómetro, se decidió utilizar el control de mando de la consola de video Wii, que facilita toda la instrumentación y garantiza la sincronización bluetooth con el PC, la señal de cada sensor es obtenida por medio del driver WiimotLib_1.7 y estos datos son digitalizados con LabView para obtener los estados de las aceleraciones durante cada instante de muestreo. La ubicación de los sensores garantiza sensar las partes de cuerpo que son usadas en las actividades propuestas. En la figuras 3 y 4 se puede observar la ubicación y distribución del sistema de captura.

Figura 3. Distribución de los sensores (esquema)

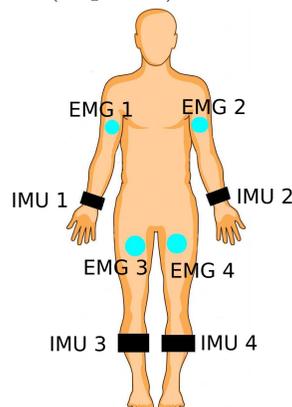


Figura 5. Adquisición de sensores IMU

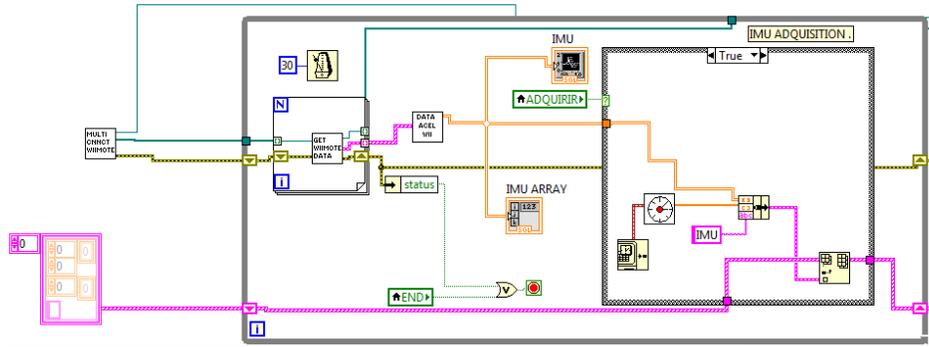


Figura 4. Distribución de los sensores (usuario real)

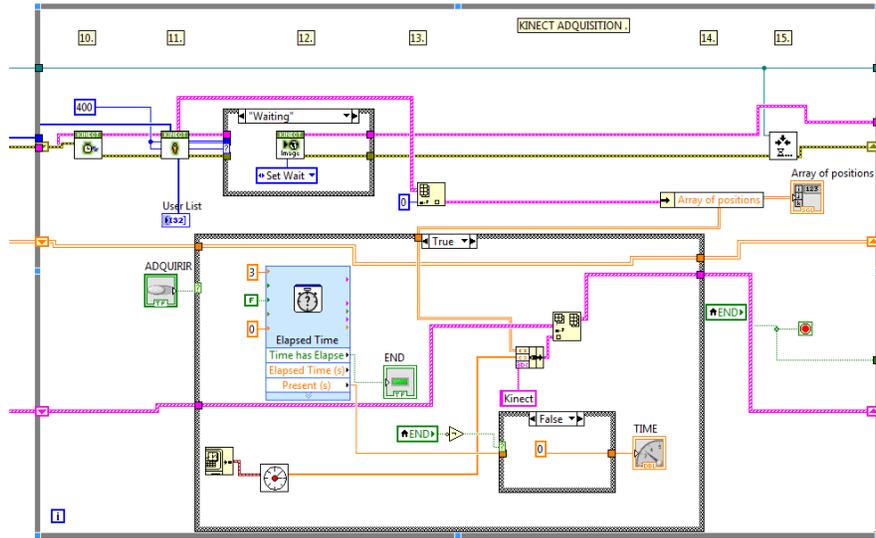


En la figura 5 se puede observar, el digrama de bloques diseñado en LabVIEW para esta etapa de la adquisición, en esta figura se puede observar los VI correspondientes a la adquisición y sincronización de datos bluetooth de los sensores acelerómetros y su almacenamiento sincronizado por medio de una estructura de selección como lo es un *case* y una bandera de adquisición global.

4.1.3. Adquisición de puntos Articulados Kinect®

El dispositivo Kinect® realiza el cálculo de 14 puntos articulados del cuerpo utilizando una cámara de profundidad, que provee una nube de puntos 3D, dicha información es utilizada por diferentes metodologías de asignación de puntos articulados [86] y [87]. Para esta aplicación se utilizó la metodología de estimación del esqueleto de [87]. En la figura 5 se puede observar, el digrama de bloques diseñado en LabVIEW para esta etapa de la adquisición, en esta figura se

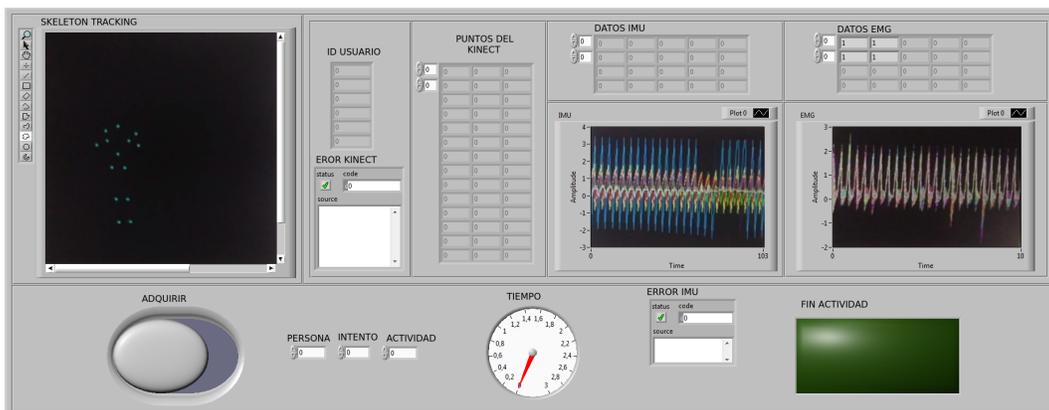
Figura 6. Adquisición Kinect®



puede observar los VI correspondientes a la adquisición y cálculo de los 14 puntos articulados del cuerpo. El almacenamiento sincronizado se realiza por medio de una estructura de selección y un bandera de global de adquisición.

La sincronización de las tres modalidades de sensores se realiza a través de una bandera booleana activando la grabación simultánea (muti-hilo) de los datos durante una ventana de observación de tres segundos para cada actividad registrada. En la figura 7 se puede observar la interfaz gráfica diseñada para la captura de actividades físicas humanas; ésta permite codificar cada acción ejecutada, almacenando los datos con la nomenclatura *PersonaN_intentoNu.mat*.

Figura 7. Interfaz gráfica diseñada para captura de actividades



4.2. Almacenamiento de base de datos anotada de actividades humanas

Para realizar la construcción de la base de datos anotada se utilizaron 16 voluntarios, los cuales ejecutaron 3 veces las actividades de la tabla 1. En la ejecución de estas acciones se logró almacenar un total de 48 ejemplos por clase. Las señales capturadas son almacenadas en archivos con extensión .mat y agrupadas de acuerdo a la actividad ejecutada. Las orientaciones de los sensores y distribución espacial de cada red en el cuerpo fueron cuidadosamente organizados en cada individuo con el fin de garantizar las misma configuración en todos los ejemplos almacenados.

Tabla 1. Lista de actividades

Actividad	Etiqueta
Quedarse quieto	1
Agacharse y levantarse	2
Saltar	3
Levantar mano derecha	4
Trotar	5

Para garantizar la integridad de los datos, cada experimento de captura fue cuidadosamente configurado de la misma manera para garantizar que la ubicación de los sensores y pose de cada sensor acelerómetro fuera la misma. Esta configuración permitió asumir que las observaciones grabadas tenían comportamientos similares y podían ser agrupadas por una metodología de aprendizaje de máquina.

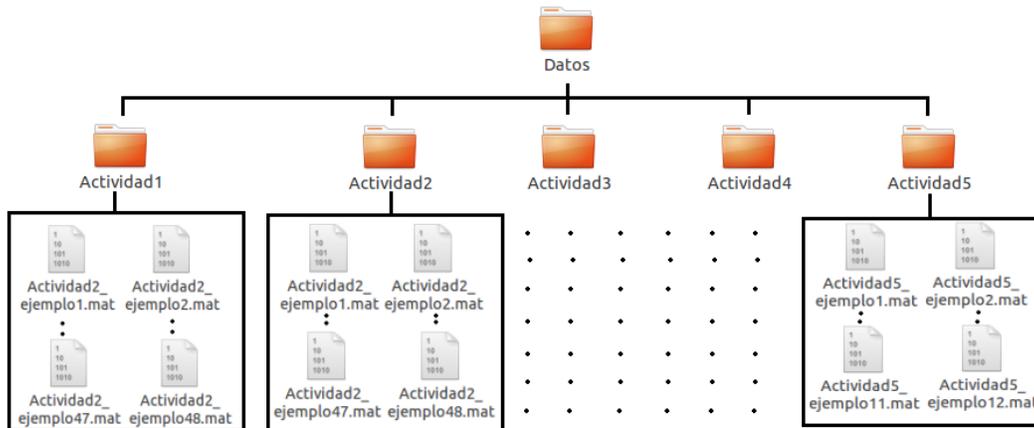
En la figura 8 se puede observar la configuración y ejecución de algunas actividades.

Figura 8. Configuración y ejecución de actividades



En la figura 9 se puede observar la distribución de los archivos en la base de datos. Cada estructura .mat contiene una celda de dimensión 1×3 , donde es almacenada la información de cada segundo de registro de la información; cada posición de esta celda almacena otra celda, con dimensión 1×3 correspondiente a la información de cada sensor. Los tamaños de las señales obtenidas se especifican en la sección 5.2.

Figura 9. Estructura base de datos



4.3. Almacenamiento de base de datos anotada de movimientos primitivos

Esta base de datos es construida utilizando la base de datos de actividades humanas. De cada actividad se segmentan los fragmentos de señal correspondientes a la ejecución de movimientos

primitivos de la tabla 2, dividiendo la ventana de observación de tres segundos en N ventanas, donde el tamaño depende de las características de muestreo de cada grupo de sensores. Estas señales son almacenadas en un archivo con extensión `.mat`, el cual se codifica de la siguiente manera:

$$\text{Base}\{\text{Ejemplo}\}\{\text{Segundo}\}\{\text{Sensor}\}\{\text{Segmento}\}$$

Donde:

- *Ejemplo* corresponde a una celda 1×240 , con la información de cada ejemplo ejecutado en la base de datos.
- *Segundo* corresponde a una celda con dimensión 1×3 , correspondiente a la información de los tres sensores durante cada segundo.
- *Sensor* corresponde a una celda con dimensión 1×3 , correspondiente a la información de cada modalidad de sensor.
- *Segmento* almacena los N arreglos correspondiente a cada submovimiento almacenado.

Los tamaños de las ventanas de observación son descritos en la sección 5.2.

Para etiquetar esta base de datos se realizó un proceso manual que consistió en observar la distribución espacial de las diferentes posturas provistas por el Kinect® durante cada ventana de observación, para así obtener la coordenada espacial que permitía facilitar la separación. En el caso de los movimientos primitivos de la actividad 2 (tabla 1), se planteó utilizar el punto de la cabeza y tomar la coordenada del eje y , esta coordenada fue normalizada con respecto al mayor valor y se plantearon tres umbrales que permitían obtener el movimiento primitivo del instante evaluado, tal y como se puede observar en la figura 10. En la actividad 3 también se sigue utilizando la metodología de la actividad 1, sin embargo se utilizan dos umbrales para asignar la etiqueta al movimiento. Para la actividad 4 se cambia el punto articulado; para este caso se utiliza el punto de la muñeca de la mano derecha y se vuelve a normalizar la coordenada y con respecto a la coordenada mayor y son planteados tres umbrales de separación. Finalmente,

para la actividad 5 se utilizan los puntos articulados de las muñecas y los tobillos y se establece como umbral el punto central del torso con la siguiente lógica:

$$if (a \leq umbral \ \&\& \ b \geq umbral) \{movimiento \ primitivo \ 7\}$$

$$if (a > umbral \ \&\& \ b < umbral) \{movimiento \ primitivo \ 8\}$$

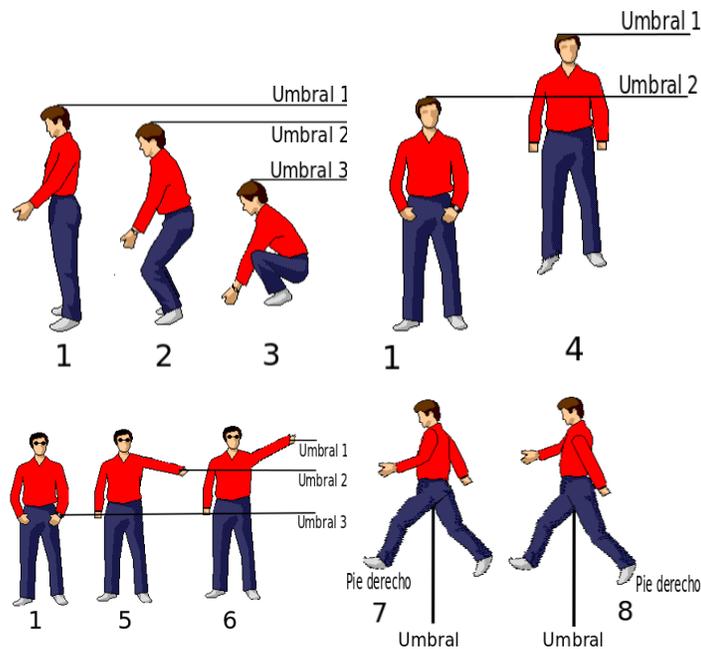
Donde

$$\blacksquare \ a = \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \text{Altura} \\ \text{Profundidad} \end{bmatrix} \text{ de la muñeca derecha.}$$

$$\blacksquare \ b = \begin{bmatrix} y \\ z \end{bmatrix} \text{ del tobillo derecho.}$$

En la figura 10, se pueden observar los umbrales establecidos para cada actividad física humana propuesta.

Figura 10. Umbrales para asignación de etiquetas



5. RECONOCIMIENTO DE MOVIMIENTOS PRIMITIVOS

Los modelos basados en movimientos primitivos se inspiran en la similitud de las señales de voz humana con el movimiento humano [88]. En el reconocimiento de voz, las frases se dividen en palabras aisladas o fonemas generando una secuencia o codificación de cada frase; lo que permite construir un modelo de clasificación de N fonemas compartidos para el conjunto de palabras a reconocer. Estos modelos se utilizan como los bloques básicos para construir palabras y frases de una manera jerárquica [89]. Siguiendo la misma idea, cada actividad se representa como una secuencia de sub-movimientos generando un código único que permite obtener un factor discriminante en la clasificación de actividad. En esta etapa se aplicó un enfoque de aprendizaje supervisado, utilizando tres máquinas de vectores de soporte durante la ventana de observación para obtener la secuencia de etiquetas que describen de forma única la ejecución de una actividad. Para efectos de su descripción, este proceso de reconocimiento consta de tres etapas fundamentales: codificación de movimientos primitivos, extracción de características y clasificación de movimientos.

5.1. Codificación de movimientos primitivos

Esta etapa consiste en el diseño de un vocabulario o diccionario, que permita describir el conjunto de sub-movimientos que son comunes para el conjunto total de actividades. Como este trabajo tiene un enfoque de aprendizaje supervisado, para la detección de estos sub-movimientos se realizó la codificación de las actividades de la tabla 1, con los movimientos primitivos que aparecen en las figuras 11 y 12. La lista de de etiquetas se puede observar en la tabla 1.

Tabla 2. Etiquetas del SVM para cada postura

Sub-actividad	Etiqueta	Sub-actividad	Etiqueta
Reposo	1	Mano levantada a $1/4$	5
Medio Agachado	2	Mano levantada a $3/4$	6
Totalmente agachado	3	Paso adelante pie derecho	7
Suspendido en el aire $1/4$	4	Paso adelante pie izquierdo	8

Figura 11. Actividades 1, 2 y 3

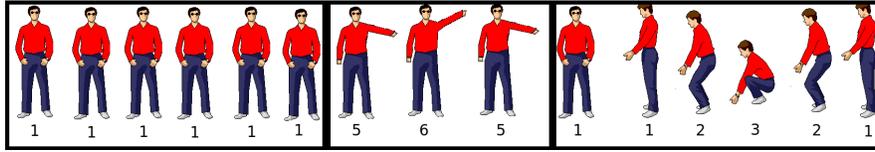
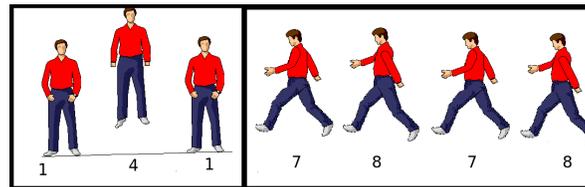


Figura 12. Actividad 4 y 5



5.2. Extracción de características

El proceso de extracción de características se divide en tres módulos fundamentales, que son: la extracción de características del Kinect, la extracción de características del Kinect para la red sensores IMU y la extracción de características para la red de sensores EMG. En esta sección se describirá el proceso de extracción, como los métodos y modelos computacionales aplicados en estos módulos. Para realizar la extracción de características es necesario segmentar las señales de cada modalidad de sensor en ventanas más pequeñas pudiendo describir un movimiento primitivo. Los tamaños de estas ventanas se calcularon experimentalmente obteniendo un tamaño de 3 muestras para los datos provenientes del Kinect® y las red de IMU, para la red de sensores EMG el tamaño que fue determinado fue de 500 muestras por movimiento. En la figura 13 se puede observar un diagrama de las señales capturadas y la respectiva segmentación para ventana de observación de 3 segundos.

5.2.1. Extracción de características del Kinect

Este módulo utiliza la información de 14 puntos articulados del cuerpo provistos por el Kinect® los cuales pueden ser observados en la figura 14. El sensor Kinect® puede proveer

Figura 13. Diagrama de Segmentación de señales

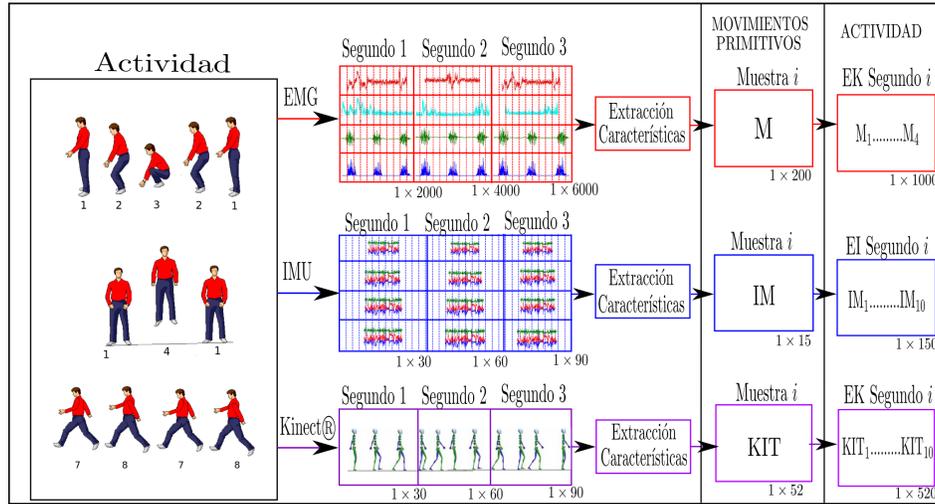
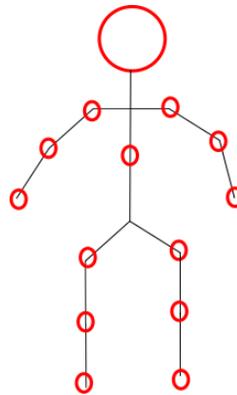


Figura 14. Puntos articulados del cuerpo



estos datos a una frecuencia de 30 muestras por segundo; sin embargo, se decide estructurar los datos en grupos de 3 muestras obteniendo información a una frecuencia de 10 muestras por segundo.

La caracterización de estos datos consiste en el cálculo de los siguientes descriptores:

- Características Polares [17].
- Descriptores estadísticos.

El cálculo de las características polares se basa en la transformación de los puntos articulados de coordenadas cartesianas J a coordenadas polares cilíndricas P . El procedimiento para

calcular el descriptor es el siguiente:

Dado:

$$J = [X_1 Y_1 Z_1 \dots X_{14} Y_{14} Z_{14}] \quad (20)$$

1. Se calcula el centro de masa del objetivo evaluado.
2. Se transforma cada articulación a coordenadas polares utilizando como origen de coordenadas el centro de masa, lo que permite obtener el vector

$$P_i = [r_1 \theta_1 r_2 \theta_2 \dots r_{14} \theta_{14}] \quad (21)$$

Donde i es la i -ésima muestra de la ventana de tres muestras, $i = \{1, 2, 3\}$.

Para el cálculo de los descriptores estadísticos, se utilizan las fórmulas de la tabla 3

Tabla 3. Descriptores estadísticos utilizados

Nombre	Descripción	Nombre	Descripción
Media de x	$m_x = \frac{1}{3} \sum_{i=1}^3 x_i$	Varianza de x	$v_x = \frac{1}{3} \sum_{i=1}^3 (x_i - \bar{x})^2$
Media de y	$m_y = \frac{1}{3} \sum_{i=1}^3 y_i$	Varianza de y	$v_y = \frac{1}{3} \sum_{i=1}^3 (y_i - \bar{y})^2$
Media de z	$m_z = \frac{1}{3} \sum_{i=1}^3 z_i$	Varianza de z	$v_z = \frac{1}{3} \sum_{i=1}^3 (z_i - \bar{z})^2$
Media de r	$m_r = \frac{1}{3} \sum_{i=1}^3 r_i$	Varianza de r	$v_r = \frac{1}{3} \sum_{i=1}^3 (r_i - \bar{r})^2$
Media de θ	$m_\theta = \frac{1}{3} \sum_{i=1}^3 \theta_i$	Varianza de θ	$v_\theta = \frac{1}{3} \sum_{i=1}^3 (\theta_i - \bar{\theta})^2$

Entonces para un ventana de observación de 3 muestras, se obtendrá el vector de características KIT, que se define como:

$$\text{KIT} = [P_1 P_2 P_3 m_x m_y m_z m_r m_\theta v_x v_y v_z v_r v_\theta]_{1 \times 52} \quad (22)$$

5.2.2. Extracción de características para red sensores IMU

Esta etapa de caracterización utiliza la información de 4 acelerómetros tri-axiales (a_x, a_y, a_z) , distribuidos en diferentes partes del cuerpo. Considerando que este dispositivo posee tres grados de libertad, se realizó el cálculo de las orientaciones Roll y Pitch utilizando la conversión a coordenadas esféricas basado en las siguientes ecuaciones:

Fuerza gravitacional G .

$$G = Raw_{accel} \left[\frac{R}{2^n - 1} \right] \quad (23)$$

Donde:

- n es la resolución en bits del conversor análogo digital.
- R es el rango de del voltaje de entrada del conversor A/D.
- Raw_{accel} es la señal de aceleración del sensor.

$$Pitch = \arctan \left(\frac{G_y}{\sqrt{G_x^2 + G_z^2}} \right) \quad (24)$$

$$roll = \arctan \left(\frac{-G_x}{G_z} \right) \quad (25)$$

Anexando estas dos componentes se obtiene el vector $I_k = \left[a_x \ a_y \ a_z \ Pitch \ Roll \right]_{1 \times 5}$ para cada instante de tiempo, donde k es el k -ésimo acelerómetro, $k = \{1, 2, 3, 4\}$. Estos sensores fueron instrumentados para proveer datos a una frecuencia de 30 Hz, sin embargo, se elige agrupar los datos con una ventana de tres, muestras obteniendo datos a frecuencia de 10 Hz. La caracterización del vector I consiste en el cálculo de los siguientes descriptores:

- Características basadas en los parámetros físicos de movimiento humano [39].
- Descriptores estadísticos

Para el cálculo de las características basadas en los parámetros físicos de movimiento humano se utilizaron las fórmulas de la tabla 4.

Tabla 4. Características IMU

Nombre	Definición
<i>Mean of Movement Intensity (AI)</i>	$AI = \frac{1}{3} \sum_{i=1}^3 MI(i),$ $MI = \ A_{xyz}\ _2$ $A_{xyz} = [a_x \quad a_y \quad a_z]^T$
<i>Variance of MI (VI)</i>	$VI = \frac{1}{3} \sum_{i=1}^3 (MI(i) - AI)^2$
<i>Normalized Signal Magnitude Area (SMA)</i>	$SMA = \frac{1}{3} \left(\sum_{i=1}^3 a_x(i) + \sum_{i=1}^3 a_y(i) + \sum_{i=1}^3 a_z(i) \right)$
<i>Eigenvalues of Dominant Directions (EVA)</i>	Vectores propios de la matriz de covarianza de la aceleración a lo largo de los ejes x,y,z en cada ventana de observación
<i>Averaged Acceleration Energy (AAE)</i>	Valor medio de la energía a lo largo de tres ejes de aceleración donde la energía es la suma de las magnitudes de los componentes discretos FFT al cuadrado de la señal de cada eje del sensor, normalizado por la longitud de la ventana
<i>Averaged Rotation Energy (ARE)</i>	Valor medio de la energía a lo largo de los ángulos de rotación, para este caso solo Roll y Pitch

Para el cálculo de los descriptores estadísticos, se utilizan las fórmulas de la tabla 5

Tabla 5. Descriptores estadísticos utilizados

Nombre	Descripción	Nombre	Descripción
Media de a_x	$m_{ax} = \frac{1}{N} \sum_{i=1}^N ax_i$	Varianza de x	$v_{ax} = \frac{1}{N} \sum_{i=1}^N (ax_i - \bar{a}_x)^2$
Media de a_y	$m_{ay} = \frac{1}{N} \sum_{i=1}^N ay_i$	Varianza de y	$v_{ay} = \frac{1}{N} \sum_{i=1}^N (ay_i - \bar{a}_y)^2$
Media de a_z	$m_{az} = \frac{1}{N} \sum_{i=1}^N az_i$	Varianza de z	$v_{az} = \frac{1}{N} \sum_{i=1}^N (az_i - \bar{a}_z)^2$
Media de a_r	$m_{ar} = \frac{1}{N} \sum_{i=1}^N ar_i$	Varianza de r	$v_{ar} = \frac{1}{N} \sum_{i=1}^N (ar_i - \bar{a}_r)^2$
Media de a_p	$m_{a\theta} = \frac{1}{N} \sum_{i=1}^N ap_i$	Varianza de θ	$v_{ap} = \frac{1}{N} \sum_{i=1}^N (ap_i - \bar{a}_p)^2$

Aplicados los cálculos de las tablas 4 y 5, se obtiene el vector de características *IM* en cada

ventana de 3 muestras.

$$IM = [AIVISMAEVA AAE ARE m_{ax} m_{ay} m_{az} m_{ar} m_{ap} v_{ax} v_{ay} v_{ar} v_{ap}]_{1 \times 15} \quad (26)$$

5.2.3. Extracción de características Red de sensores EMG

En esta etapa se sensaron cuatro músculos del cuerpo con frecuencia de muestreo de 2kHz como se pueden observar en el figura 3, obteniendo una señal análoga E_i , donde i es el i -ésimo sensor EMG. Para esta metodología, E_i es segmentado utilizando ventanas V_j de 500 muestras donde j es la j -ésima ventana, esto permite obtener observaciones a una frecuencia de 4 Hz.

Cada ventana V_j es caracterizada utilizando una transformada Waveleth discreta tipo haar de 20 niveles, obteniendo un vector de características M_i de 1×2000 , donde i es la i -ésimo sensor, entonces para cada ventana j se obtendrá el vector de características.

$$EMG_j = \begin{bmatrix} M_1 & M_2 & M_3 & M_4 \end{bmatrix}_{1 \times 2000}$$

También se probó caraterizar la señal con MFCC, sin embargo no fue posible obtener una detección de movimiento primitivos mayor al 50 % de acierto, razón por la cual, no se tuvo en cuenta en este trabajo.

5.3. Clasificación de Movimientos

Para esta etapa se utilizaron tres máquinas de vectores de soporte multiclase con estrategia de clasificación “*All-pairs*”; este método de aprendizaje utiliza un kernel gaussiano con radio 1×10^4 para el Kinect®. Este proceso se repite con los datos de los sensores IMU y con la información de los sensores EMG.

6. RECONOCIMIENTO DE ACTIVIDAD

El proceso de reconocimiento de actividad consiste en la clasificación de las actividades establecidas en la tabla 1, utilizando las secuencias generadas en la etapa 5. Para lograr este objetivo se utiliza un módulo de fusión de datos encargado de centralizar la información de las máquinas de vectores de soporte, y posteriormente se realiza la clasificación utilizando Modelos Ocultos de Markov (HMM).

6.1. Fusión de datos

En la sección 5.3 cada clasificador genera una etiqueta: EI corresponde a la salida de la SVM con información provista por la red de sensores IMU, EK corresponde a la salida de la SVM con información provista por el dispositivo Kinect® y EE corresponde a la salida de la SVM con información provista por la red de sensores EMG. Estos sensores proveen la información de sensado a diferentes velocidades de muestreo, por lo que es necesario diseñar un módulo de fusión de datos que centralice la información y la haga apta para la etapa de clasificación con HMM.

El proceso de fusión de datos consiste en generar un vector de características EF con las etiquetas de los clasificadores SVM durante una ventana de observación de tres segundos. EF se construye concatenando linealmente y sin ponderación cada etiqueta de los clasificadores durante la ventana de observación, obteniendo el vector de características de la ecuación 27.

$$EF = [[EK_1 EK_2 \dots EK_{30}] [EI_1 EI_2 \dots EI_{30}] [EE_1 EE_2 \dots EE_{12}]]_{72 \times 1} \quad (27)$$

6.2. Clasificación con Modelos Ocultos de Markov (HMM), Entrenamiento y validación del modelo

Este proceso de clasificación parte de un modelo entrenado que proporciona una separación entre cada clase. Este modelo permite asignar una etiqueta de pertenencia evaluando dicho modelo. El entrenamiento de los modelos SVM se realiza utilizando un kernel gaussiano y una

estrategia de clasificación multiclase “*All-pairs*”. Básicamente este proceso estima un modelo de separación entre las clases utilizando el método de minimización sequential minimal optimization (SMO). Para el entrenamiento de la HMM se utilizaron 24 estados y 32 centroides en la construcción del libro de códigos; este proceso es iterado hasta que se encuentre un desempeño mayor al 95 % o se cumplan 100 iteraciones y se escoje el modelo con mejor desempeño. Se escogieron 32 centroides en el HMM, debido al análisis de centroides hecho, este análisis se muestra con mayor detalle en la sección de resultados.

Para realizar el procedimiento de evaluación del modelo entrenado se utilizó una estrategia de validación cruzada, fraccionando la base de datos con el 70 % para el entrenamiento y 30 % para evaluación; este proceso realiza iteraciones de Montecarlo con criterio de paro $\|\text{diag}(M_k) - \text{diag}(M_{k-1})\| < 0,001$, donde $\text{diag}(M_k)$ es la el vector generado por la diagonal de la matriz de confusión y k es la iteración promedio actual de Montecarlo.

Este procedimiento de Montecarlo, separa de forma aleatoria los datos de entrenamiento y los datos, permitiendo en cada irteración obtener modelos con diferentes datos y observar el comportamiento en la clasificación de actividad. Las estadísticas obtenidas en el experimento consisten en el cálculo de la matriz de confusión y el comportamiento promedio del porcentaje de acierto para cada clasificador aplicado. Para entrenar las SVM se utilizó la base de datos anotada explicada amplimente en la sección 4.3 y para el HMM se utilizó la base de datos de la sección 4.2.

7. EXPERIMENTOS Y RESULTADOS

En esta etapa, se presentan los resultados obtenidos al validar el desempeño de la clasificación de actividad y movimientos primitivos aplicando Montecarlo para la estimación del error. Se logra demostrar que la fusión multimodal de los sensores Kinect® , IMU y EMG, permite obtener mayor desempeño en el reconocimiento de actividad que cualquier combinación de este grupo de sensores, o incluso, el uso individual de alguno de ellos. Para probar dicha afirmación se decidió aplicar la metodología propuesta y evaluar el desempeño para las siguientes combinaciones de sensores: Kinect® + IMU + EMG, Kinect® + IMU, Kinect® + EMG, EMG + IMU, solo EMG, solo IMU y solo EMG.

La evaluación consta de dos partes fundamentales que se denominan Análisis de movimiento primitivo y Análisis de reconocimiento de actividad. En la primera parte se realizó el análisis de desempeño de los clasificadores SVM para los sub-movimientos propuestos en la tabla 2. Para la segunda parte se validó la clasificación de actividades humanas con el modelo oculto de Markov utilizando la base de datos anotada de la sección 6.2. También se realiza el cálculo del conjunto mínimo de sensores que permite garantizar una clasificación confiable de actividades físicas humanas y movimientos primitivos. Además de los análisis anterior se realizó una comparación de la eficiencia de todos los métodos, donde se calcula el valor probable y su desviación para un intervalo de confianza del 99 %.

7.1. Reconocimiento de Movimientos Primitivos

En esta etapa se aplicó la metodología de validación descrita en la sesión 6.2 obteniendo las matrices de confusión para los datos provistos con el Kinect® , IMU y EMG. Este análisis permite observar la variabilidad en la detección de movimientos primitivos, ya que es fundamental saber cuanto se equivoca esta etapa ya que es el insumo para detección de actividad. Los resultados del análisis de reconocimiento de movimientos primitivos utilizando los datos provistos por el Kinect® pueden ser observados en la tabla 6, donde se evidencia el porcentaje de acierto con un rendimiento aceptable para todas las clases a reconocer. El análisis con los

Tabla 6. Matriz de Confusión SVM-Kinect®

	1	2	3	4	5	6	7	8
1	92,78	1,5	0,8	3,23	0,63	0	1	0,78
2	8,53	73,13	12,66	0,39	0,19	0,17	2,34	2,58
3	0,65	22,30	75,87	0,11	0	0,12	0,58	0,37
4	15,11	0	0	84,89	0	0	0	0
5	4,26	0	0	0,01	88,04	7,27	0,42	0
6	0	0	0	0	3,43	96,57	0	0
7	5,45	0	0,01	0,03	0	0	74,77	19,74
8	4,13	0,6	0,27	0	0	0	19,17	76,38

datos provistos por la red de sensores IMU evidencia un comportamiento aceptable para el reconocimiento de movimientos primitivos, tal y como se puede observar en la tabla 7. Además de realizar el proceso anterior, también se llevó a cabo un análisis de rendimiento del sistema mediante la eliminación de la mayor cantidad posible de sensores en la fusión de datos, esto permite encontrar el número mínimo de sensores que garantiza un rendimiento confiable en la detección. De este análisis se puede concluir que solo utilizando el sensor IMU 1 (mano derecha) es posible obtener un rendimiento confiable para los movimientos de la base de datos (sección 4.2). En la tabla 8 se puede observar la matriz de confusión obtenida de aplicar análisis de Montecarlo, logrando la convergencia de la clasificación con 74 iteraciones, para un error de estado estacionario de 0.0021.

Tabla 7. Matriz de Confusión SVM-IMU todo el conjunto de sensores

	1	2	3	4	5	6	7	8
1	89,08	1,92	0,18	1,48	1,51	0,32	3,04	2,48
2	7,75	77,61	13,80	0	0	0	0,59	0,24
3	0,67	26,65	71,24	0	0	0	1,22	0,21
4	12,22	0	0	87,07	0	0	0,59	0,13
5	14,36	0	0	0	74,87	13,30	0,47	0
6	0,47	0	0	0	6,47	93,07	0	0
7	8,67	0,16	0,01	0,33	0,03	0	62,28	28,52
8	9,56	0,11	0	0,27	0	0	27,74	62,28

Tabla 8. Matriz de Confusión SVM-IMU con sensor de mano derecha

	1	2	3	4	5	6	7	8
1	81.33	2.15	0.31	1.81	3.00	0.0241	4.97	4.01
2	7.89	75.51	14.27	0.02	0	0.0001	1	1.30
3	1.41	27.21	69.29	0.87	0	0.0001	0.69	0.52
4	11.64	0.02	0.80	83.87	0	0	1.01	2.67
5	32.16	0.39	0	0	56.85	0.0777	1.18	1.66
6	9.02	0.33	0.04	0	6.33	0.8394	0.28	0.07
7	17.43	0.13	0.45	1.09	0	0	54.68	26.22
8	13.17	0.19	0.02	0.15	0	0	30.83	55.64

El análisis de desempeño con los datos proporcionados por la red de sensores EMG evidencia un comportamiento con baja confiabilidad, ya que se obtienen aciertos de clasificación en el rango de 43.23.41 % - 82.98 %, tal como se observa en la tabla 9. Dado el rango anterior, es evidente que la clasificación con sensores EMG presenta baja confiabilidad, debido a que la clase 5 tiene una probabilidad menor al 50 %. Aunque el rendimiento de los sensores EMG presenta detecciones con baja confiabilidad, se realizó el análisis del conjunto mínimo de sensores, y se concluye que es necesaria la información de sólo tres sensores (mano derecha y pies) para garantizar una confiabilidad similar, tal y como puede verse en la tabla 10.

Como se puede observar en las tablas 6, 7 y 9, los clasificadores lograron obtener un porcentaje de acierto aceptable para la mayoría de clases. Sin embargo, algunos movimientos primitivos no son reconocidos de forma confiable, debido en gran parte a la incapacidad que tiene la red de sensores por sí misma para brindar información discriminante en el aprendizaje. Un ejemplo de esto puede observarse en la tabla 7, donde los sensores IMU no pueden distinguir entre las clases 7 y 8, ya que los cambios de aceleración para estos movimientos primitivos son similares. Este mismo fenómeno sucede con las señales EMG, ver tabla 9, donde no es posible obtener una confiabilidad en el acierto mayor al 80 % para las clases 1, 2, 5, 7 y 8.

Tabla 9. Matriz de Confusión SVM-EMG con todo el conjunto de sensores

	1	2	3	4	5	6	7	8
1	62.41	7.72	2.86	2.86	3.38	5.60	8.41	6.77
2	15.62	62.23	21.19	0	0.39	0.058	0	0
3	0	26.24	76.31	0	0	0	0	0
4	15.97	0	0	82.98	0	0.70	0	0
5	22.92	2.46	0	0	43.23	1.05	1.14	0
6	9.89	0	0	0	10.09	78.89	0.07	1.06
7	7.64	2.72	1.09	0	0.40	4.06	56.19	27.90
8	10.50	0	0	0	0	0	0.2117	62.34

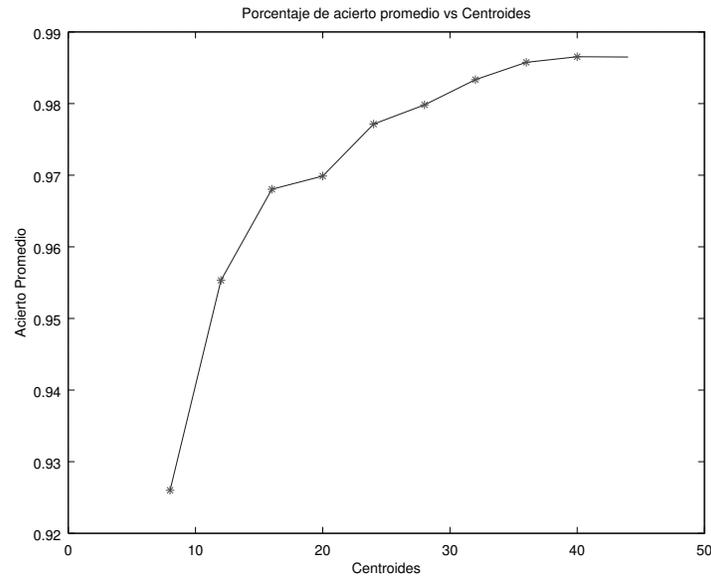
Tabla 10. Matriz de Confusión SVM-EMG con sensores de mano derecha y pies

	1	2	3	4	5	6	7	8
1	62.52	7.72	2.92	2.84	3.37	5.59	8.32	6.72
2	10.84	67.81	20.82	0	0.03	0.37	0	0.13
3	0.17	26.24	73.57	0.02	0	0	0	0
4	19.92	0.04	0	79.34	0	0.70	0	0
5	25.58	1.38	0	0	45.44	26.05	0.85	0.70
6	8.11	0.01	0.02	0.08	10.08	79.03	1.11	1.56
7	5.78	1.89	0.55	0	0.27	3.31	51.98	36.21
8	9.37	0.08	0	0	0.25	0.83	33.76	55.71

7.2. Reconocimiento de actividades físicas

En esta etapa se calcularon las matrices de confusión correspondientes al reconocimiento de las actividades propuestas en la tabla 1. Estos indicadores de desempeño son calculados para las distintas combinaciones de sensores, y posteriormente se realiza un análisis comparativo del desempeño generado por cada clasificador.

Figura 15. Número de centroides Vs Eficiencia de Acierto



En la figura 15 se puede observar un previo análisis hecho al número de centroides con el que se debe realizar el entrenamiento de la HMM, este análisis consiste en calcular la porcentaje de acierto promedio para diagonal de la matriz de confusión variando el número de centroides en cada experimento. En esta figura se observa que cuando se aumenta el número de centroides la eficiencia del método aumenta, sin embargo este aumento viene acompañado de un gran costo computacional, por lo que se decidió utilizar 32 centroides con el fin de garantizar una alta eficiencia mayor al 98 % del acierto y costo computacional menor. Este experimento fue hecho para la fusión de los tres tipos de sensores.

En las tablas 11, 12, 15, 17, 14, 19, y 21, se puede observar la matriz de confusión del reconocimiento de actividades físicas humanas para cada posible combinación de sensores formuladas en este trabajo.

Tabla 11. Matriz de Confusión HMM-Kinect® con todos lo sensores

	1	2	3	4	5
1	82	0	18	0	0
2	0	100	0	0	0
3	0	0.81	98.73	0	0.46
4	0	0	0	98.74	1.51
5	0	0.10	0	0.46	99.44

En la tabla 11 se puede observar un alto desempeño de clasificación para las 181 iteraciones de Montecarlo, garantizando un error de estado estable menor a 0.03 al utilizar datos provistos por el Kinect®. Para la clase 1 se obtiene un porcentaje de acierto del 82.50% debido a que los datos se normalizan con respecto al centro de masa de la persona y la distribución espacial de los puntos articulados es similar a la distribución de la actividad 3.

Tabla 12. Matriz de Confusión HMM-IMU con todos lo sensores

	1	2	3	4	5
1	100	0	0	0	0
2	0	98.37	0.14	0	1.49
3	0	1.89	96.14	0.36	1.61
4	0	0	0	100	0
5	0	0.24	0.13	0.13	97.80

La clasificación de actividad utilizando sensores IMU evidencia un alto desempeño, como se puede observar en la tabla 12, donde la clase con menor desempeño (3) presenta una probabilidad de acierto del 96.14%. Este buen resultado se debe en gran parte a la etapa de detección de movimientos primitivos, ya que la probabilidad de estimar etiquetas diferentes a las que describe cada actividad es mínima, permitiendo así obtener una codificación discriminante para la HMM, tal y como se puede observar en la tabla 7. El análisis de sensores mínimos también se aplicó a la metodología de reconocimiento de actividad mediante 30 iteraciones de Montecarlo para garantizar un error de estado estable menor a 0.03. De este análisis se puede observar que solo es necesario el sensor de la mano derecha para reconocer el conjunto de actividades de forma confiable; sin embargo, en la actividad 5 se reduce considerablemente el desempeño, pero se garantiza un desempeño mayor al 85% de acierto.

Tabla 13. Matriz de Confusión HMM-IMU con sensor de la mano derecha

	1	2	3	4	5
1	100	0	0	0	0
2	0	99.49	0	0	0.51
3	0	4.48	92.30	0.69	2.54
4	0	0	0	99.63	0.37
5	0	5.35	6.36	0.94	87.34

Tabla 14. Matriz de Confusión HMM-EMG con todos los sensores

	1	2	3	4	5
1	38.78	13.44	19.72	16.76	11.30
2	0	90.94	2.11	1.03	5.93
3	0	1.89	96.16	1.95	0
4	0	0.49	1.07	94.08	4.36
5	0	3.58	0	6.73	89.69

En la tabla 14 se puede observar el desempeño de la red de sensores EMG en el reconocimiento de actividad física humana, evidenciando un alto desempeño de clasificación para las clases 2, 3, 4 y 5. La clase 1, que reporta el menor desempeño, obtuvo una probabilidad de acierto del 38.78 %. Sin embargo, se piensa que para reconocer todas las actividades de forma confiable se debe aumentar el número de sensores EMG, ya que con esta configuración no es posible proporcionar las características suficientes para reconocer la clase 1. El análisis de Montecarlo para la fusión de datos de los tres tipos de sensores se puede observar en las tablas 15, 17, 19 y 21, reportando un aumento en el rendimiento de la metodología de detección.

En la combinación Kinect® y EMG, se observa un aumento considerable de la metodología, logrando reconocer la clase 1 con probabilidad de acierto del 91 % sin disminuir el desempeño en las otras clases, lo cual no era posible de lograr con los sensores de forma individual. Estos resultados se observan en la tabla 15. En el análisis del conjunto de sensores mínimo se puede concluir que es necesario utilizar 2 sensores (Pie derecho y Mano derecha) EMG y el Kinect para obtener un rendimiento confiable del clasificador, esto se puede observar en la tabla 16.

Tabla 15. Matriz de Confusión HMM-Kinect®+EMG

	1	2	3	4	5
1	91	0	9	0	0
2	0	100	0	0	0
3	0	0	98.8	0	1.2
4	0	0	0	97.68	2.31
5	0	1.86	0	2.63	95.51

Tabla 16. Matriz de Confusión HMM-Kinect®+EMG con sensores mano derecha y pie derecho

	1	2	3	4	5
1	94.49	0	5.51	0	0
2	0	100	0	0	0
3	0	4.16	88.07	3.17	4.59
4	0	0	4.79	94.14	1.07
5	0	1.01	0	0	98.99

La combinación Kinect® e IMU, presenta un aumento del desempeño de acierto obteniendo resultados del 100% para la clases 1, 2 y resultados cercanos al 100% para las clases 3, 4 y 5. Este comportamiento se debe a que la clasificación individual de los sensores IMU ya presentaba un alto desempeño y combinando estos datos con la red EMG aumenta tal desempeño. En el análisis del conjunto de sensores mínimo se puede concluir que es necesario utilizar 1 sensor (Mano derecha) IMU y el Kinect para obtener un rendimiento confiable del clasificador, esto puede observarse en la tabla 18.

Tabla 17. Matriz de Confusión HMM-Kinect®+IMU con todos los sensores

	1	2	3	4	5
1	100	0	0	0	0
2	0	100	0	0	0
3	0	2.29	97.71	0	0
4	0	0	0	97.95	2.05
5	0	0	1.61	0	98.39

Tabla 18. Matriz de Confusión HMM-Kinect®+IMU con sensor de la mano derecha

	1	2	3	4	5
1	100	0	0	0	0
2	0	98.32	0	0	1.68
3	0	0	99.41	0	0.59
4	0	0	0	100	0
5	0	1.06	3.06	3.47	92.41

La combinación IMU y EMG presenta un aumento del desempeño, obteniendo resultados del 100% para la clase 1 y resultados cercanos al 100% para las otras clases. En el análisis del conjunto de sensores mínimo se puede concluir que es necesario utilizar 2 sensores (Pie derecho y Mano derecha) EMG y un sensor IMU (Mano derecha) para obtener un rendimiento confiable del clasificador, tal como puede observarse en la tabla 20.

Tabla 19. Matriz de Confusión HMM-IMU+EMG con todos los sensores

	1	2	3	4	5
1	100	0	0	0	0
2	0	98.14	0.31	0	1.55
3	0	0.16	99.46	0	0.38
4	0	0	0.33	99.67	0
5	0	0.95	2.14	0.26	96.65

Tabla 20. Matriz de Confusión HMM-IMU+EMG con sensores brazo derecho y pie derecho

	1	2	3	4	5
1	100	0	0	0	0
2	0	99.49	0.51	0	0
3	0	0.35	95.11	3.75	0.79
4	0	0	1.92	97.71	0.37
5	0	1.69	2.71	0.31	95.29

La combinación de los tres tipos sensores Kinect® + IMU + EMG, puede ser observada en la tabla 21, donde se reportan resultados de clasificación con el 100% de acierto para la clase 1 y resultados cercanos al 100% para las clases restantes; también se puede observar que es la combinación que mejor desempeño presenta ya que se cuenta con la información de todos los sensores. En el análisis del conjunto de sensores mínimo se puede concluir que es necesario

utilizar 1 sensor (Mano derecha) EMG, un sensor IMU (Mano derecha) y el sensor Kinect® para obtener un rendimiento confiable del clasificador, esto puede ser observarse en la tabla 22.

Tabla 21. Matriz de Confusión HMM-Kinect®+IMU+EMG con todos los sensores

	1	2	3	4	5
1	100	0	0	0	0
2	0	99.60	0.40	0	0
3	0	0.14	99.62	0.15	0.09
4	0	0	0	99.09	0.91
5	0	0.68	1.31	2.25	95.76

Tabla 22. Matriz de Confusión HMM-Kinect®+IMU+EMG con mano derecha y brazo derecho

	1	2	3	4	5
1	100	0	0	0	0
2	0	99.68	0.21	0	0.11
3	0	3.01	91.98	4.04	0.96
4	0	0	0.86	98.99	0.14
5	0	1.19	1.67	1.12	95.93

Realizando el análisis comparativo, se puede observar en la tabla 23 que la mayor eficiencia de la metodología se da cuando se realiza la fusión de datos con los tres tipos de sensores, lo que implica que combinar todos los sensores agrega características importantes para la clasificación. En las tablas 23 y 24 se puede observar que la combinación Kinect®+IMU+EMG presenta el mejor desempeño para la clasificación de todas las clases, sin embargo existen algunas combinaciones que pueden detectar una clase en particular mejor que otras combinaciones, e incluso mejor que la combinación de los tres sensores; un ejemplo de esto puede ser observado en la tabla 23, donde se evidencia que la combinación Kinect® + IMU detecta con el $100 \pm 0\%$ de acierto las clases 1, 2, con un intervalo de confianza del 99% y presenta un acierto del $98,39 \pm 2,10\%$ para la clase 5 con un intervalo de confianza del 99% . Otro ejemplo de este comportamiento se puede observar con el Kinect®, esta combinación es la que detecta con mejor desempeño la clase 5.

Tabla 23. Comparación de desempeño de la metodología para todo el conjunto de sensores

Grupo de Sensores	1	2	3	4	5
Kinect® + IMU +EMG	100 ± 0	99,60 ± 1,53	99,62 ± 1,03	99,09 ± 1,59	95,76 ± 3,36
Kinect® + IMU	100 ± 0	100 ± 0	97,71 ± 2,50	97,95 ± 2,36	98,39 ± 2,10
Kinect® +EMG	91 ± 4,78	100 ± 0	98,8 ± 1,82	97,68 ± 2,51	95,51 ± 3,45
IMU+EMG	100 ± 0	98,14 ± 0,023	99,46 ± 1,23	99,67 ± 0,96	96,65 ± 3,00
Kinect®	82 ± 6,40	100 ± 0	98,73 ± 1,87	98,74 ± 1,86	99,44 ± 1,25
IMU	100 ± 0	98,37 ± 2,11	96,14 ± 3,21	100 ± 0	97,80 ± 2,45
EMG	38,78 ± 8,12	90,94 ± 4,81	96,16 ± 3,21	94,08 ± 3,94	89,69 ± 5,07

Tabla 24. Comparación de desempeño de la metodología con el conjunto mínimo de sensores.

Sensor Group	1	2	3	4	5
Kinect® + IMU +EMG	100 ± 0	99,68 ± 0,95	91,98 ± 4,53	98,99 ± 1,97	95,93 ± 3,40
Kinect® + IMU	100 ± 0	98,32 ± 2,15	99,41 ± 1,28	100 ± 0	92,41 ± 4,42
Kinect® +EMG	94,49 ± 3,80	100 ± 0	88,07 ± 5,40	94,14 ± 3,92	98,99 ± 1,67
IMU+EMG	100 ± 0	94,49 ± 3,80	95,11 ± 3,60	97,71 ± 2,50	95,29 ± 3,53
IMU	100 ± 0	94,49 ± 3,80	92,30 ± 4,44	99,63 ± 1,02	87,34 ± 5,54

Para obtener una métrica del comportamiento global de cada metodología, se decidió calcular el valor promedio de la diagonal de cada matriz de confusión y hacer una gráfica de barras con el fin de comparar el rendimiento de todos los métodos. En las figuras 16 y 17, se puede observar el rendimiento promedio para cada combinación de sensores propuesta en este trabajo.

Tabla 25. Comportamiento promedio de la metodología

KI+EMG+IMU	KI+EMG	KI+IMU	EMG+IMU	KI	IMU	EMG
98,81 ± 1,81	96,81 ± 2,93	98,81 ± 1,81	98,84 ± 1,78	95,73 ± 3,37	98,46 ± 2,06	81,93 ± 6,41

En la figura 16 se observa que las combinaciones KI+IMU y EMG+IMU presentan un comportamiento similar a la combinación total de sensores, sin embargo, la combinación KI+EMG no tiene un rendimiento comparable con esta combinación y su rango de incertidumbre es mayor. Se puede observar que la combinación KI+IMU generan el mejor desempeño al tener un acierto 98,81 ± 1,81 %. En la figura 17, se puede observar que el rendimiento al eliminar sensores no disminuye considerablemente, esto garantiza obtener un sistema confiable con un mínimo de sensores.

Figura 17. Comportamiento global para las combinaciones con cantidad reducida de sensores

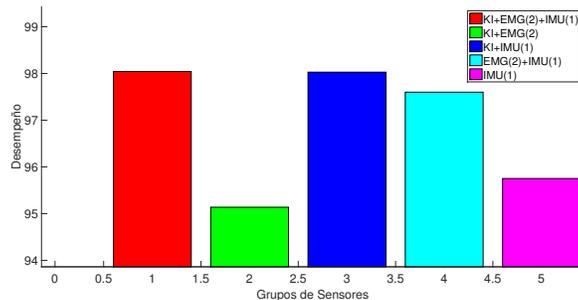


Figura 16. Comportamiento global para las combinaciones con todo el conjunto de sensores

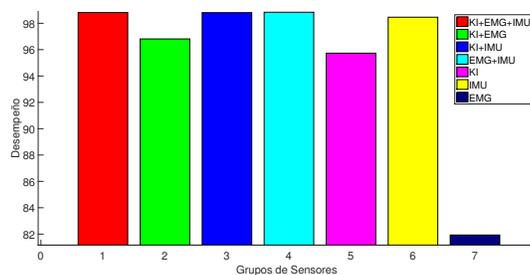


Tabla 26. Comportamiento promedio de la metodología para el conjunto mínimo de sensores

KI+EMG(2)+IMU(1)	KI+EMG(2)	KI+IMU(1)	EMG(2)+IMU(1)	IMU(1)
98,04 ± 2,31	95,14 ± 3,59	98,03 ± 2,32	97,60 ± 2,55	95,75 ± 3,36

8. CONCLUSIONES Y RECOMENDACIONES

8.1. CONCLUSIONES

Se desarrolló una metodología de reconocimiento automático de actividad física humana utilizando la fusión de datos multimodal de una red de sensores IMU, una red de sensores EMG y un sensor de profundidad (Kinect®). Este método implementa la detección de movimientos primitivos utilizando máquinas de vectores de soporte multiclase y posteriormente clasifica la actividad entrenando las secuencias de movimientos provistas por la salidas de las SVM en una ventana de 3 segundos con HMM.

En este trabajo, se realiza el estudio comparativo del aporte de cada sensor en el reconocimiento de actividades físicas humanas, probando las diferentes configuraciones de sensores y evaluando su desempeño del método. Para este trabajo se probaron las combinaciones IMU+EMG+Kinect®, IMU+Kinect®, EMG+Kinect®, IMU+EMG, IMU, EMG y Kinect®, donde se evidencia que las combinaciones con mejor desempeño son las que involucran la red de sensores IMU. Además, se evidencia que al utilizar un conjunto mínimo de sensores, se logra garantizar la confiabilidad en la detección, presentando un rendimiento comparable al uso de todos los sensores. Cuando el número de sensores es mínimo, la combinación IMU (Mano derecha)+ EMG (Mano derecha)+ Kinect®, permite obtener la mejor confiabilidad, reportando un desempeño del $98,04 \pm 2,31$ %, esto debido a que es la detección con mayor exactitud y su rango de incertidumbre es menor.

Para realizar el reconocimiento de actividad, se hace necesario el uso de todos los sensores disponibles debido a las fallas que estos dispositivos podrían presentar durante el proceso. Algunas de estas fallas son las oclusiones parciales o auto-oclusiones del sensor Kinect®, la pérdida de conexión en los sistemas de comunicación inalámbrica donde generalmente son conectados los sensores IMU o EMG. Este tipo de errores hacen pensar que este sistema debe contar con la información multimodal de todos los sensores, para así utilizar fusión de datos en cualquiera de las posibles combinaciones y evitar dichas fallas cuando algún sensor deje de ser confiable.

Se probó la metodología propuesta en una base de datos anotada, creada específicamente para este trabajo, esto debido a que no existía una base de datos con el registro sincronizado de los tres tipos de sensores utilizados. Sin embargo, la base de datos se debe expandir para validar la metodología en un conjunto de actividades más amplio, permitiendo así evaluar actividades donde la separación entre clases en acciones presente un mayor desafío. Estas actividades deben ser similares en la combinación de movimientos primitivos, para forzar el entrenamiento del HMM. Ejemplo de esto, es la captura de actividades como caminar, trotar y correr, o agacharse, sentarse o recoger algún objeto de suelo.

Se demostró el poder de caracterización que tienen los descriptores utilizados para cada sensor en el reconocimiento de movimientos primitivos, logrando así recopilar en un solo documento descriptores efectivos para este tipo de tareas.

8.2. RECOMENDACIONES

El núcleo de este trabajo está centrado en el reconocimiento de actividades físicas humanas utilizando diferentes modalidades de sensores; evaluando el desempeño cuando se combinan diferentes grupos de sensores. Durante el desarrollo del mismo, se observó, que el conjunto de actividades de la base de datos anotada es mínimo, generando un interrogante en cuanto a la capacidad de separación de la metodología para otro tipo de actividades. Por lo que es necesario re-diseñar el equipo de captura, ya que su instrumentación presenta dificultades para la ejecución de actividades donde se requieren distancias considerables; por lo que se propone un sistema de comunicaciones con protocolo inalámbrico.

Otro factor importante en la metodología, es la etiquetación manual de movimientos primitivos; la cual está sujeta a múltiples errores humanos, generando como consecuencia, no encontrar un entrenamiento óptimo del sistema con la base de datos anotada. En esta etapa, se propone explorar otros métodos de entrenamiento probando técnicas de aprendizaje de máquina no supervisado como: mezcla de Gaussianas, mezclas de Gaussianas infinitas, *hierarchical clustering*, modelos ocultos de Markov, entre otros.

Aunque en el documento no se realizó un estudio de complejidad computacional temporal para la metodología, se aclara que en las pruebas efectuadas para las etapas de entrenamiento y clasificación, la metodología se volvía sumamente costosa; cuando se trataban con datos de la red de sensores EMG. Este factor causa que la metodología sea poco atractiva para algunos sistemas de procesamiento. Con el fin de solucionar este problema, se propone realizar un análisis de componentes principales o un análisis de relevancia en la etapa de extracción de características, esto con el fin de reducir la dimensionalidad de estos vectores y así disminuir el costo computacional. También se debe explorar otro conjunto de características para evaluar su capacidad de separación y considerar si deben agregarse en la metodología. Ejemplos de estos descriptores son: HOG3D y SURF3D para la nube de puntos generada por el Kinect® y transformadas como FFT, SFFT, STFT, Hilbert, para los sensores EMG e IMU.

En la etapa de entrenamiento de reconocimiento de actividad con HMM, se evidencia que al incluir algunas características en la fusión de datos se disminuye el rendimiento del método, provocando que no se dese realizar la fusión de sensores para algunos casos. Una propuesta para evitar el problema anterior, es explorar técnicas de fusión de datos más robustas, que permitan concluir la etapa del proceso donde se deben fusionar todas las fuentes de información.

BIBLIOGRAFÍA

- [1] BISHOP, Christopher. Pattern Recognition and Machine Learning. Springer, 2008. (document), 3.3, 3.3.1, 3.3.2
- [2] A.J. BERNHEIM BRUSH, John Krumm y SCOTT, James. Activity Recognition Research: The Good, the Bad, and the Future. En: The Eighth International Conference on Pervasive Computing, 2010. 1, 1.1, 2.1.1
- [3] AKRAM BAYAT, Marc Pomplun y TRAN, Duc A. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. En: The 11th International Conference on Mobile Systems and Pervasive Computing, 2014. 1, 1.1
- [4] BAO, Ling y INTILLE, Stephen S. Activity Recognition from User-Annotated Acceleration Data. En: in Proceedings of PERVASIVE, 2004, págs. PP 1–17. 1, 1.1
- [5] DOS SANTOS, Diliana Maria Barradas Rebelo. Human Activity Recognition for an Intelligent Knee Orthosis. Tesis de Maestría, Facultad de Ciencias e Tecnologia, Universidade Nova de Lisboa, 2012. 1, 1.1
- [6] MEDIA INFORMATICS GROUP UNIVERSITY OF MUNICH, GERMANY. Multi-Sensor Activity Context Detection for Wearable Computing. Perceptual Computing and Computer Vision ETH Zurich, Switzerland. 1, 1.1, 1.2
- [7] BAÑOS, O., *et al.* Kinect=IMU? Learning MIMO Signal Mappings to Automatically Translate Activity Recognition Systems across Sensor Modalities. En: Wearable Computers (ISWC), 2012 16th International Symposium on, 2012. ISSN 1550-4816, págs. 92–99. 1, 1.1, 2.4
- [8] LIU, Kui, *et al.* Multi-HMM classification for hand gesture recognition using two differing modality sensors. En: Circuits and Systems Conference (DCAS), 2014 IEEE Dallas, 2014, págs. 1–4. 1, 1.1

- [9] BURMEISTER, O.P., *et al.* Synchronous stereo-video and biosignal recording - a basic setup for Human-Computer-Interface applications. En: Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on, 2005, págs. 501–505. 1, 1.1, 1.2
- [10] DESTELLE, F., *et al.* Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors. En: Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, 2014, págs. 371–375. 1, 1.1, 2.1.1
- [11] JINGEN LIU, Saad Ali y SHAH, Mubarak. Recognizing Human Actions Using Multiple Features. En: Computer Vision and Pattern Recognition, 2008. 1, 1.1, 1.2, 2.1
- [12] BAO, L. y INTILLE, S. S. Activity recognition from user-annotated acceleration data. En: in Proceedings of the 2nd International Conference on Pervasive Computing, 2004, págs. 1–17. 1, 1.1, 1.2, 2.2.1
- [13] BAÑOS, Oresti; POMARES, Héctor y ROJAS, Ignacio. Hibridación de paradigmas de clasificación en el contexto del reconocimiento automático de actividades físicas cotidianas. En: ESTYLF, Congreso Español sobre Tecnologías y Lógica Fuzzy, 2010. 1, 1.1, 1.2
- [14] ZHANG, Zequn, *et al.* A novel method for user-defined human posture recognition using Kinect. En: Image and Signal Processing (CISP), 2014 7th International Congress on, 2014, págs. 736–740. 1, 1.1, 1.2
- [15] FOERSTER, Friedrich y FAHRENBERG. Motion pattern and posture: Correctly assessed by calibrated accelerometers. En: Behavior Research Methods, Instruments, & Computers, 2000. 1, 1.2, 2.2.1
- [16] LU, X, Liu Q y S, Oe. Recognizing Non-rigid Human Actions using Joints Tracking in Space-Time. En: Conference on Information Technology: Coding and Computing, 2004. 1, 1.1, 1.2

- [17] WU, Haitao, *et al.* Human activity recognition based on the combined SVM and HMM. En: Information and Automation (ICIA), 2014 IEEE International Conference on, 2014, págs. 219–224. 1, 1.1, 1.2, 2.1.1, 5.2.1
- [18] D, Leightley, *et al.* Human Activity Recognition for Physical Rehabilitation. En: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, 2013, págs. 261–266. 1.1
- [19] GYLLENSTEN, Ilpha Cuba. Physical Activity Recognition in Daily Life using a Triaxial Accelerometer. Tesis de Maestría, School of Computer Science and Engineering Royal Institute of Technology, 2010. 1.1
- [20] LI, Kang y FU, Yun. ARMA-HMM: A New Approach for Early Recognition of Human Activity. 21, 2012, págs. 11–15. 1.1, 2.1.1
- [21] DUONG, T.V., Bui H.H. Phung D.Q. Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model. En: Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005. 1.1
- [22] BLANK, M. Gorelick, L. Shechtman E. Irani M. y BASRI, R. Actions as Space-time Shapes. En: Tenth IEEE International Conference on Computer Vision, 2005. 1.1, 2.1.1
- [23] A, Dragan M y I., Mocanu. Human Activity Recognition in Smart Environments. En: Control Systems and Computer Science (CSCS), 2013 19th International Conference on, 2013, págs. 495–502. 1.1, 2.1.1
- [24] PANVALKAR, Rujuta; VASHI, Ami y KARANI, Ruhina. Human Posture Recognition: Background Subtraction and Skeletonization. En: IOSR Journal of Computer Engineering, tomo 16, 2014, págs. 67–70. 1.1, 2.1.1
- [25] KUMARI y S.K, Mitra. Human Action Recognition Using DFT. En: Third IEEE National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, 2011. 1.1

- [26] DALAL, N., Triggs B. Histograms of Oriented Gradients for Human Detection. En: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. 1.1
- [27] KE, Y., Sukthankar R. y M, Hebert. Spatio-temporal Shape and Flow Correlation for Action Recognition. En: Conference on Computer Vision and Pattern Recognition, 2007. 1.1, 2.1.1
- [28] DOLLÁR, P, Rabaud V Cottrell G y S, Belongie. Behavior Recognition via Sparse Spatio-Temporal Features. En: 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 1.1, 2.1.1
- [29] LU, W y LITTLE, J.J. Simultaneous tracking and action recognition using the PCA-HOG descriptor. En: 3rd Canadian Conference on Computer and Robot Vision, 2006. 1.1
- [30] SCOVANNER, P Ali, S y M, Shah. A 3-dimensional SIFT Descriptor and Its Application to Action Recognition. En: 15th International Conference on Multimedia, ACM,, 2007. 1.1
- [31] VEERARAGHAVAN, A.; A.K., Roy-Chowdhury y CHELLAPPA, R. Matching shape sequences in video with applications in human movement analysis. En: Pattern Analysis and Machine Intelligence, IEEE Transactions on, tomo 27, nº 12, 2005, págs. 1896–1909. ISSN 0162-8828. 1.1
- [32] LOWE, D.G. Object Recognition from Local Scale-invariant Features. En: Seventh IEEE International Conference on Computer Vision, Kerkyra, 1999. 1.1
- [33] LOWE, D.G. Distinctive image features from scale-invariant keypoints. En: Int J Comput Vis, 2004. 1.1
- [34] LI, W., Zhang Z. y LIU, Z. Action recognition based on a bag of 3d points. En: In Workshop on CVPR for Human Communicative Behavior Analysis, 2010. 1.1, 2.1.1
- [35] J. SUNG, C. Ponce, B. Selman y SAXENA, A. Human activity detection from rgbd images. En: in Plan, Activity, and Intent Recognition, tomo WS- 11-16, 2011. 1.1, 2.1.1

- [36] XIA, Lu; CHEN, Chia-Chih y AGGARWAL, J.K. Human detection using depth information by Kinect. En: Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on, 2011. ISSN 2160-7508, págs. 15–22. 1.1
- [37] TAO, G.; ARCHAMBAULT, P.S. y LEVIN, M.F. Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis. En: Virtual Rehabilitation (ICVR), 2013 International Conference on, 2013, págs. 164–165. 1.1
- [38] DE CASTRO, David Martín. Aplicación Android para el reconocimiento automático de actividades físicas en tiempo real. Tesis de Maestría, Universidad Carlos III de Madrid Departamento de Informática, 2012. 1.1, 2.2.2
- [39] ZHANG, Mi y SAWCHUK, Alexander A. A feature selection-based framework for human activity recognition using wearable multimodal sensors. En: In International Conference on Body Area Networks (BodyNets), Beijing, China, 2011. 1.1, 2.2.1, 5.2.2
- [40] M. BOCKSCH, J. Seitz y JAHN, J. Pedestrian Activity Classification to Improve Human Tracking and Localization. En: Forth International Conference on Indoor Positioning and Indoor Navigation, 2013. 1.1, 2.2.2
- [41] GRAUPE, D. EMG pattern analysis for patient-responsive control of FES in paraplegics for walker-supported walking. En: Biomedical Engineering, IEEE Transactions on, tomo 36, nº 7, 1989, págs. 711–719. ISSN 0018-9294. 1.1, 2.3.1
- [42] KANG, Wen-Juh, *et al.* The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition [movements classification]. En: Biomedical Engineering, IEEE Transactions on, tomo 42, nº 8, 1995, págs. 777–785. ISSN 0018-9294. 1.1, 2.3.1
- [43] K. FARRY, I. Walker y BARANIUK, R. Myoelectric teleoperation of a complex robotic hand. En: IEEE Transactions On Robotic and Automation, tomo 12, nº 5, 1996. 1.1, 2.3.1

- [44] KUNDU, A.S., *et al.* Trajectory generation for myoelectrically controlled lower limb active knee exoskeleton. En: Contemporary Computing (IC3), 2014 Seventh International Conference on, 2014, págs. 230–235. 1.1
- [45] FERGUSON, S. y DUNLOP, R. G. Grasp recognition from myoelectric signals. En: Australasian Conference on Robotics and Automation, 2002. 1.1, 2.3.1
- [46] DAISUKE, N. Studies on electromyogram to motion classifier. Tesis Doctoral, Hokkaido University, 2001. 1.1, 2.3.1
- [47] GAVIRIA, Sergio y OSORIO, Marisol. Fusión de Sensores con Lógica Difusa para la Detección de Objetos Enterrados. En: Ingeniar UPB, 2015. 1.2
- [48] HELTEN, T., *et al.* Real-Time Body Tracking with One Depth Camera and Inertial Sensors. En: Computer Vision (ICCV), 2013 IEEE International Conference on, 2013. ISSN 1550-5499, págs. 1105–1112. 1.2
- [49] FENG, Shimin, Murray-Smith y RODERICK. Fusing Kinect Sensor and Inertial Sensors with Multi-rate Kalman Filter. En: Data Fusion & Target Tracking 2014: Algorithms and Applications (DF&TT 2014), 2014, págs. 1–8. 1.2
- [50] WANG, Baocheng; YANG, Chenguang y XIE, Qing. Human-machine interfaces based on EMG and Kinect applied to teleoperation of a mobile humanoid robot. En: Intelligent Control and Automation (WCICA), 2012 10th World Congress on, 2012, págs. 3903–3908. 1.2
- [51] C. JORGENSEN, K. Wheeler y STEPNIIEWSKI, S. Bioelectric control of a 757 class high fidelity aircraft simulation. En: World Automation Congress, Wailea, Maui, Hawaii, 2000. 1.2, 2.3.1
- [52] GIRALDO S, Eduardo y OROZCO GUTIÉRREZ, Alvaro A. Sistema automatizado efectivo y apropiado de caracterización y clasificación de señales electromiográficas para el control de prótesis y brazos robóticos. Inf. téc., Colciencias Colombia., 2008. 1.2, 3.1.4, 4.1.1

- [53] XIAO, Qian y CHENG, Jun. Human action recognition framework by fusing multiple features. En: Information and Automation (ICIA), 2013 IEEE International Conference on, 2013, págs. 985–990. 2.1.1
- [54] PAUL SCOVANNER, Saad Ali y SHAH, Mubarak. A 3-Dimensional SIFT Descriptor and its Application to Action Recognition. En: Proceedings of the 15th international conference on Multimedia, 2007. 2.1.1
- [55] CABRAL, Joao Carlos Paula Santos. Kinect-Based Algorithms for Motion Analysis. Tesis de Maestría, Técnico Lisboa, Master of Science Degree in Electrical and Computer Engineering, 2013. 2.1.1
- [56] BISWAS, K. K. y BASU, Saurav Kumar. Gesture Recognition using Microsoft Kinect. En: Proceedings of the 5th International Conference on Automation, Robotics and Applications, 2011. 2.1.2
- [57] LE, Thi-Lan; NGUYEN, Minh-Quoc y NGUYEN, Thi-Thanh-Mai. Human posture recognition using human skeleton provided by Kinect. En: Computing, Management and Telecommunications (ComManTel), 2013 International Conference on, 2013, págs. 340–345. 2.1.2
- [58] EETIMES. GPS system with IMUs tracks first responders. R. Colin Johnson. 2.2, 3.1.2
- [59] MANNINI, A. y SABATINI, A.M. Machine Learning Methods for Classifying Human Physical Activities from on-body sensors. En: Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA, 2011. 2.2.2
- [60] ZHANG, Mi; SAWCHUK y A., Alexander. Motion Primitive-based Human Activity Recognition Using a Bag-of-features Approach. En: Proceedings of the 2Nd ACM SIGHT International Health Informatics Symposium, IHI '12. ACM, New York, NY, USA, 2012. ISBN 978-1-4503-0781-9, págs. 631–640. URL <http://doi.acm.org/10.1145/2110363.2110433>. 2.2.2

- [61] KORBINIAN FRANK, Maria Josefa Vera, Patrick Robertson y ANGERMAN, Michael. Reliable Real-Time Recognition of Motion Related Human Activities Using MEMS Inertial Sensors. En: Proceedings of the 23rd International Technical Meeting of The Satellite Division of the Institute of Navigation, 2010. 2.3.2
- [62] MICROSOFT. Kinect for Windows Technical Datasheet. Partner Network. 3.1.3
- [63] GUYON, Isabelle y ELISSEEFF, André. Foundations and Applications Studies in Fuzziness and Soft Computing: An Introduction to Feature Extraction. Springer Verlag, 2006. 3.2
- [64] AKANSU, Ali N Haddad y A., Richard. Multiresolution signal decomposition: transforms, subbands, and wavelets. Boston, MA: Academic Press, 1992. 3.2.1
- [65] TYAGI, V. y WELLEKENS, C. "Shape Activity": a continuous-state HMM for moving/deforming shapes with application to abnormal activity detection. En: IEEE International Conference In Acoustics, Speech, and Signal Processing,, 2005. 3.2.2
- [66] FANG ZHENG, Guoliang Zhang y SONG, Zhanjiang. Comparison of Different Implementations of MFCC. En: Computer Science and Technology, tomo 16, nº 6, 2001, págs. 582–589. 3.2.2
- [67] FLACH, Peter. Machine Learning: The Art and Science of Algorithm that Make Sense of Data. Cambridge University Press, 2012. 3.3
- [68] MEHRYAR MOHRI, Afshin Rostamizadeh y TALWALKAR, Ameet. Foundations of Machine Learning. The MIT Press, 2012. 3.3, 3.3.7
- [69] VAPNIK, V. Statistical Learning Theory. John Wiley, 1998. 3.3.3
- [70] SCHÖLKOPF, Bernhard, Smola y J, Alexander. Learning with Kernels. MIT Press, 2002. 3.3.3

- [71] TREJO, Juan Angel Resendiz. Las máquinas de vectores de soporte para identificación en lÃnea. Tesis de Maestría, Centro de Investigación y de estudios avanzados del Instituto Polotécnico Nacional México, 2006. 3.3.3
- [72] CRAMMER, Koby y SINGER, Yoram. On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines. En: J of Machine Learning Research, 2001. 3.3.4
- [73] WU, Deihui. Reconstruction Strategy for Multi-Class SVM Based on Posterior Probability. En: Pattern Recognition, 2008. CCPR '08. Chinese Conference on, 2008, págs. 1–6. 3.3.5
- [74] NATIONAL TAIWAN UNIVERSITY. A Note on Platt's Probabilistic Outputs for Support Vector Machines. Department of Computer Science and Information Engineering. 3.3.5, 3.3.5
- [75] FESTA, Robert. Opthimun Inductive Methods, tomo 232. Springer, 1993. 3.3.5
- [76] RABINER, Lawrence R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. En: Proceedings of the IEEE, tomo 77, n^o 2, 1989. 3.3.6
- [77] RABINER, L. A tutorial on hidden Markov models and selected applications in speech recognition. En: Proceedings of the IEEE, tomo 77, n^o 2, 1989, págs. 257–286. ISSN 0018-9219. 3.3.6
- [78] ALDRICH, John. Fisher and the making of maximum likelihood. En: Statistical Science, 1997. 3.3.6
- [79] RABINER, Lawrence. First Hand: The Hidden Markov Model. En: IEEE Global History Network, 2013. 3.3.6
- [80] FISHER, Ronald. Statistical Methods and Scientific Inference. Edinburgh, 1956. 3.3.8
- [81] EPA UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. Consolidated Human Activity Database (CHAD) for use in human exposure and health studies and predictive models. EPA. 4

- [82] UNIVERSITY OF CENTRAL FLORIDA. UCF Sports Action Data Set. Center for Research in Computer Vision. 4
- [83] IVAN LAPTEV. Learning Human Actions from Movies. IRISA INRIA Rennes France. 4
- [84] UCI. Human Activity Recognition Using Smartphones Data Set. Center for Machine Learning and Intelligent Systems. 4
- [85] ILLINOIS. RGB-D Human Activity Recognition and Video Database. Advanced Digital Sciences Center. 4
- [86] FENG, Shimin Murray-Smith, Roderick. Fusing Kinect Sensor and Inertial Sensors with Multi-rate Kalman Filter. En: Data Fusion & Target Tracking 2014: Algorithms and Applications (DF&TT 2014), IET Conference Publications, 2014, págs. 1–8. 4.1.3
- [87] CHANJIRA SINTHANAYOTHIN, Nonlapas Wongwaen, Wisarut Bholsithi. Skeleton Tracking using Kinect Sensor & Displaying in 3D Virtual Scene. En: International Journal of Advancements in Computing Technology(IJACT), tomo 4, 2012. 4.1.3
- [88] GHASEMZADEH, H., *et al.* A phonological expression for physical movement monitoring in body sensor networks. En: Mobile Ad Hoc and Sensor Systems, 2008. MASS 2008. 5th IEEE International Conference on, 2008, págs. 58–68. 5
- [89] HUANG, Xuedong; ACERO, Alex y HON, Hsiao-Wuen. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. 1^a ed^{ón}. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001. ISBN 0130226165. 5

9. ANÉXOS

9.1. Anéxo Acondicionamiento de señal sensores EMG

Figura 18. Acondicionamiento de señal para un sensor EMG

