

**IDENTIFICACIÓN DE ACTIVIDAD HUMANA USANDO
APRENDIZAJE NO SUPERVISADO EN SISTEMAS
MULTIMODALES**

Ing. Santiago Morales García

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

Director

Ing. Jose Bestier Padilla Bejarano

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
PROGRAMA DE MAESTRIA EN INGENIERÍA ELÉCTRICA
LINEA DE AUTOMÁTICA
PEREIRA
2016**

Nota de Aceptación

Firma del Presidente del jurado

Firma del jurado 1 - Evaluador

Firma del jurado 2 - Evaluador

Firma del jurado 3 - Director

Pereira, 11 de Noviembre de 2016

Agradecimientos

Agradezco a todas las personas que compartieron conmigo en este proceso de maestría que tomó un largo mucho tiempo en completarlo. Primero le doy gracias a mi padres, Jhon Jairo Morales Calderon y Martha Azucena García Arias, por haberme dado la educación que sirvió para haber realizado este proceso de formación importante en mi vida. También le agradezco a mis otros familiares, y en especial a mi abuela Oliva Arias Parra, ya que ella es un ejemplo a seguir y ha enseñado a toda su familia que el estudio es lo más importante que una persona se pueda dar en su vida.

A mi compañero de maestría, Andrés Felipe Calvo Salcedo, le doy muchas gracias porque me ofreció mucho de su apoyo en la realización de este proceso. También agradezco a Andrés Álvarez Mesa por sus consejos y ayuda en la programación del trabajo, y doy gracias a otros grandes amigos como Cristian Daniel Molina, Arley, Leydi, porque con ellos también compartí muy buenos momentos que sirvieron a mi formación.

Por último quiero agradecer a la Universidad Tecnológica de Pereira por ofrecerme la oportunidad de realizar uno de los procesos más importantes de mi vida.

Muchas Gracias.

CONTENIDO

	pág.
TÍTULO	
CONTENIDO	
1. INTRODUCCIÓN	11
1.1. DEFINICIÓN DEL PROBLEMA	11
1.2. JUSTIFICACIÓN	14
2. OBJETIVOS	17
2.1. OBJETIVO GENERAL	17
2.2. OBJETIVOS ESPECÍFICOS	17
3. ESTADO DEL ARTE	19
4. MARCO CONCEPTUAL	23
4.1. Selección de Características	23
4.2. Clasificación	23
4.2.1. Clasificación supervisada y no supervisada	23
4.3. <i>Machine Learning</i>	24
4.4. <i>Clustering</i>	24
4.5. Desempeño de un clasificador	24
4.6. Dispositivos de captura de movimiento	26
4.6.1. Unidad de Masa Inercial (IMU)	26
4.6.2. Electromiografía (EMG)	26
4.6.3. Kinect®1.0	26
5. MARCO TEÓRICO	29
5.1. Técnicas de Selección de Características	29

5.1.1.	Análisis de Componentes Principales	29
5.1.2.	<i>ReliefF</i>	29
5.2.	Técnicas de Agrupamiento	30
5.2.1.	<i>Spectral Clustering (SC)</i>	30
5.2.2.	<i>K-Means Clustering (KmC)</i>	31
5.3.	Técnicas de Clasificación	31
5.3.1.	Modelos Ocultos de Marcov (HMM)	31
5.3.2.	K-Vecinos más Cercanos	33
6.	METODOLOGIA	35
6.1.	BASE DE DATOS	35
6.1.1.	Extracción de Características	36
6.2.	Selección de características	38
6.3.	RECONOCIMIENTO DE MOVIMIENTOS PRIMITIVOS	38
6.3.1.	Codificación de movimientos primitivos	39
6.3.2.	Asignación de etiquetas	39
6.3.3.	Clasificación de Movimientos	40
6.4.	RECONOCIMIENTO DE ACTIVIDAD	41
6.4.1.	Fusión de datos	41
6.4.2.	Clasificación con Modelos Ocultos de Markov (HMM)	42
6.4.3.	Entrenamiento y validación del HMM	42
7.	EXPERIMENTOS Y RESULTADOS	43
7.1.	Umbralización de movimientos primitivos y selección de sensores y puntos articulados.	44
7.2.	Análisis de movimientos primitivos	56
7.3.	Análisis de reconocimiento de actividad física humana . .	72
8.	CONCLUSIONES Y RECOMENDACIONES	83
8.1.	CONCLUSIONES	83

8.2. RECOMENDACIONES	84
BIBLIOGRAFÍA	87
9. ANÉXOS	93

LISTA DE TABLAS

1.	Tablas de Actividades	41
2.	Movimientos primitivos para cada etiqueta.	44
3.	SC-Kinect (4 <i>clusters</i> y 6 dimensiones)	57
4.	SC-Kinect+IMU (4 <i>clusters</i> y 9 dimensiones)	57
5.	SC-Kinect+IMU+EMG (4 <i>clusters</i> y 5 dimensiones)	57
6.	SC-Kinect (5 <i>clusters</i> y 6 dimensiones)	59
7.	SC-Kinect+IMU (5 <i>clusters</i> y 6 dimensiones)	59
8.	SC-Kinect+IMU+EMG (5 <i>clusters</i> y 10 dimensiones)	59
9.	KmC-Kinect (4 <i>clusters</i>)	61
10.	KmC-Kinect+IMU (4 <i>clusters</i>)	61
11.	KmC-Kinect+IMU+EMG (4 <i>clusters</i>)	61
12.	KmC-Kinect (5 <i>clusters</i>)	63
13.	KmC-Kinect+IMU (5 <i>clusters</i>)	63
14.	KmC-Kinect+IMU+EMG (5 <i>clusters</i>)	63
15.	SC-Kinect (8 <i>clusters</i> y 9 dimensiones)	65
16.	SC-Kinect+IMU (8 <i>clusters</i> y 11 dimensiones)	65
17.	SC-Kinect+IMU+EMG (8 <i>clusters</i> y 9 dimensiones)	65
18.	KmC-Kinect (8 <i>clusters</i>)	67
19.	KmC-Kinect+IMU (8 <i>clusters</i>)	67
20.	KmC-Kinect+IMU+EMG (8 <i>clusters</i>)	67
21.	HMM-Kinect (SC y cuatro movimientos primitivos)	72
22.	HMM-Kinect+IMU (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	73
23.	HMM-Kinect+IMU (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.	73
24.	HMM-Kinect+IMU+EMG (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	74

25.	HMM-Kinect+IMU+EMG (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.	74
26.	HMM-Kinect (SC y cinco movimientos primitivos)	75
27.	HMM-Kinect+IMU (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	75
28.	HMM-Kinect+IMU (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.	75
29.	HMM-Kinect+IMU+EMG (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	76
30.	HMM-Kinect+IMU+EMG (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.	76
31.	HMM-Kinect (KmC y cuatro movimientos primitivos)	77
32.	HMM-Kinect+IMU (KmC y cuatro movimientos primitivos). Fu- sión de sensores por medio de sus etiquetas.	77
33.	HMM-Kinect+IMU (KmC y cuatro movimientos primitivos). Fu- sión de sensores por medio de sus características.	77
34.	HMM-Kinect+IMU+EMG (KmC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	78
35.	HMM-Kinect+IMU+EMG (KmC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.	78
36.	HMM-Kinect (KmC y cinco movimientos primitivos).	79
37.	HMM-Kinect+IMU (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	79
38.	HMM-Kinect+IMU (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.	80
39.	HMM-Kinect+IMU+EMG (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.	80

40.	HMM-Kinect+IMU+EMG (KmC y cinco movimientos primitivos).	
	Fusión de sensores por medio de sus características.	80
41.	Medias y varianzas de las diagonales de las matrices de confusión	
	del análisis de reconocimiento de actividad física humana	81

LISTA DE FIGURAS

1.	Gráfico del proceso llevado a cabo para el reconocimiento de actividad	35
2.	Posición de los puntos articulados del Kinect y de los sensores IMU y EMG en el cuerpo de una persona.	37
3.	Gráfica de similaridad.	39
4.	Gráficas de similaridad para cada uno de los sensores de reconocimiento.	45
5.	Desempeño de cada uno de los puntos articulados del Kinect® y de cada sensor IMU y EMG utilizando el algoritmo <i>ReliefF</i>	46
6.	Desempeño de cada uno de los puntos articulados del Kinect® y de cada sensor IMU y EMG utilizando el algoritmo <i>PCA</i>	46
7.	Desempeño de los clusters con todos los puntos articulados del Kinect®.	47
8.	Desempeño de los clusters con 8 de los puntos articulados más relevantes (usando <i>PCA</i>) del Kinect®.	48
9.	Desempeño de los clusters con 6 de los puntos articulados más relevantes (usando <i>ReliefF</i>) del Kinect®.	49
10.	Desempeño de los clusters con todos los sensores IMU.	50
11.	Desempeño de los clusters con los 3 sensores IMU más relevantes (<i>PCA</i>).	51
12.	Desempeño de los clusters con los 3 sensores IMU más relevantes (<i>ReliefF</i>).	52
13.	Desempeño de los clusters con todos los sensores EMG.	53
14.	Desempeño de los clusters con los 3 sensores EMG más relevantes (<i>PCA</i>).	54
15.	Desempeño de los clusters con los 2 sensores EMG más relevantes (<i>ReliefF</i>).	55

16.	Desempeño de cada una de las combinaciones de sensores al variar las dimensiones utilizando <i>Spectral Clustering</i> con 4 <i>clusters</i>	58
17.	Desempeño de cada una de las combinaciones de sensores al variar las dimensiones utilizando <i>Spectral Clustering</i> con 5 <i>clusters</i>	60
18.	Desempeño de cada una de las combinaciones de sensores utilizando <i>K-means Clustering</i> con 4 <i>clusters</i>	62
19.	Desempeño de cada una de las combinaciones de sensores utilizando <i>K-means Clustering</i> con 5 <i>clusters</i>	64
20.	Desempeño de cada una de las combinaciones de sensores al variar las dimensiones utilizando <i>Spectral Clustering</i> con 8 <i>clusters</i>	66
21.	Desempeño de cada una de las combinaciones de sensores utilizando <i>K-means Clustering</i> con 8 <i>clusters</i>	68
22.	Media y varianza de cada una de las combinaciones de sensores al variar las dimensiones utilizando <i>Spectral Clustering</i> con 4, 5, 6, 7 y 8 <i>clusters</i>	70
23.	Media y varianza de cada una de las combinaciones de sensores utilizando <i>K-means Clustering</i> con 4, 5, 6, 7 y 8 <i>clusters</i>	71

1. INTRODUCCIÓN

El reconocimiento de actividad humana se ha vuelto un asunto importante en diferentes campos como deporte, rehabilitación y diagnóstico médico, sistema de vigilancia inteligente, entre otros [1, 2, 3], los cuales han llevado a que surjan investigaciones de metodologías que permitan la detección automática de un conjunto de actividades físicas. Muchas de estas metodologías de reconocimiento utilizan sensores como acelerómetros, cámaras de video, electrodos, etc, pero la mayoría utilizan aprendizaje supervisado y aunque el uso de este aprendizaje ha ofrecido buenos resultados, el problema es que para tareas más complejas no es eficiente, es sensible al ruido, y como sus valores de entrada son especificados por el usuario, puede presentar errores de umbralización entre muchos otros [1]. Por lo tanto se utiliza el aprendizaje no supervisado, el cual, ha visto un aumento en investigaciones basadas en dicha metodología en los últimos años, ya que suprime muchos de los problemas que acarrea el uso de aprendizaje supervisado[4].

Así que este trabajo propone una metodología de aprendizaje no supervisado para reconocer los movimientos para la actividad humana en un sistema multimodal. Para esto se utilizan técnicas de agrupamiento no supervisado (*Unsupervised Clustering*) que separa cada actividad física en movimientos primitivos. También se aplican técnicas de selección de características para cada sensor (Kinect[®], IMU, EMG). El enfoque propuesto se prueba en una base de datos de actividad física humana, donde se estima el número mínimo de movimientos primitivos para el conjunto de actividades después de encontrar el número de sensores y puntos articulados que son necesarios para obtener un rendimiento similar como si se utilizara todas las características del sistema de reconocimiento. Esto demuestra que el aprendizaje no supervisado es capaz de reconocer cada movimiento de forma satisfactoria y evitar ciertos inconvenientes que conlleva el uso de aprendizaje supervisado.

1.1. DEFINICIÓN DEL PROBLEMA

En los últimos tiempos se ha vuelto un factor importante el poder realizar seguimiento a las acciones físicas de las personas, debido a la cantidad de aplicaciones que se benefician de esto,

como son el deporte, la salud (rehabilitación, cuidado monitoreado y diagnóstico médicos) [1, 2, 3], sistema de vigilancia inteligente (monitoreo de personas en un centro comercial o de automóviles en una autopista), el análisis de acciones y la realidad virtual [5], entre otras. Esto ha provocado que surjan investigaciones que trabajen con sistemas que puedan detectar, reconocer y sintetizar las acciones humanas [3].

Los sistemas de detección, reconocimiento y sintetización de las acciones de los seres humanos, son aquellos que han surgido de los aportes de la ingeniería, algunos de estos son los sensores como cámaras de video, giroscopios, acelerómetros, goniómetros, sensores electromiográficos, etc. Aun cuando se puede aprovechar el uso de la tecnología, llevar esto a la práctica no es una labor sencilla, ya que estos sistemas de reconocimiento no tienen la capacidad de descomponer temporalmente las acciones tan fácilmente como lo haría un humano [5], debido a que existen una gran cantidad de parámetros que hay que tener en cuenta como son la larga variabilidad física de la persona, el amplio rango de escalas, la irregularidad de las acciones humanas, y la naturaleza potencial de las posibles combinaciones de movimientos [6].

Hay que agregar que cuando se trata de un sistema multimodal como el caso estudiado en este trabajo (Kinect[®], IMU, EMG), se improvisa el rendimiento de reconocimiento de las actividades [4], pero esto conlleva a un mayor número de sensores y dispositivos, generando que la dimensionalidad de los datos crezca y provocando que el uso de algoritmos computacionalmente costosos [1, 7].

Para esto es recomendable utilizar la menor cantidad de dispositivos por sensor o en casos extremos, utilizar menos tipos de sensores. Pero esto lleva que se pierda rendimiento de forma significativa si no se extrae los dispositivos con menos importancia o si se extraen demasiados.

En cuanto a las metodologías de aprendizaje para la detección y clasificación de movimientos, existen dos tipos de enfoques que son: el aprendizaje supervisado y el no supervisado. El primero se basa en el conocimiento previo de las características de la acción, es rápido y robusto para tareas simples [4]. El problema es que para tareas más complejas no es eficiente, es sensible al ruido y es deficiente para adaptarse a eventos fuera de lo común [1].

Mientras que la mayoría de las técnicas de aprendizaje no supervisado detectan estructuras gruesas y trabajan con pocas dimensiones [1], su fortaleza es que es capaz de agrupar y clasificar adecuadamente los datos provenientes de las acciones y los movimientos, a diferencia de como lo hace un humano. Por ejemplo, Andrés Calvo comenta en su trabajo, que aun con personal clasificado, un humano no separa de forma correcta los umbrales entre las acciones, además que en la etiquetación, un humano no tiene la capacidad de escoger el número indicado de etiquetas que determinan una actividad [4].

Llegando al punto en el que la clasificación está completa, es importante destacar que cada actividad está compuestas por movimientos los cuales en la literatura se conocen como movimientos primitivos. Estos son movimientos que el ser humano ha realizado en el proceso evolutivo, y el agruparlos en determinado orden, forman actividades como el caso de leer un libro, practicar un deporte, entre muchos otras.

Para ordenar los movimientos primitivos que definen una actividad, se utilizan metodologías inspiradas en los modelos de reconocimiento de la voz humana [4]. Algunas de las más comunes para el reconocimiento de movimientos primitivos son acerca de aproximaciones basadas en cadenas de coincidencias; o en estadísticas, donde el último caso posee técnicas como *Bag of Features* (BoF) el cual será fundamental en este trabajo. Alguno de los problemas que resultan al usar estas metodologías es determinar el tamaño de la ventana de tiempo o definir un umbral entre los movimientos. Muchas de estas son sensibles al ruido y su desempeño puede disminuir si hay presencia de variaciones intra-clase altas [8].

Dado un sistema multimodal (IMU's, EMG y Kinect®) en el que se adquirieron diferentes actividades físicas, se diseñará un sistemas de reconocimiento con aprendizaje no supervisado para la fusión de datos provenientes de las acciones primitivas, ya que según el estado del arte, la cantidad de trabajos relacionados con esta temática son pocos y más bien utilizan metodologías de aprendizaje supervisado, las cuales generan inconvenientes que se pueden solucionar si se usa los enfoques propuestos en este trabajo. Además se dará soluciones a los problemas de dimensionalidad que que está ligado a la cantidad de dispositivos por tipo de sensor del sistema de reconocimiento de actividad usado.

1.2. JUSTIFICACIÓN

El reconocimiento de actividad de las acciones que realiza el ser humano ha llevado a los investigadores a una era donde es importante realizar aplicaciones que monitoreen, detecten y clasifiquen los movimientos primitivos de las personas. Aún cuando se han usado múltiples metodologías de detección y clasificación [7, 9, 2, 3, 5, 10], las cuales usan sensores y sistemas de visión, se tiene el problema de usar solo un tipo de los sistemas mencionados. Esto causa que los datos que llegan, no sean los suficientes para lograr clasificar correctamente una acción o actividad humana [6, 4]. Lo que genera que al usar más de un tipo de sistema de detección (sensores o sistemas de video), se pueda ofrecer una mejor percepción de las actividades, y se obtengan datos significativos en el momento de procesarlos.

Para solucionar gran parte de la fusión de datos y de otros problemas que surgen en el reconocimiento de actividades y movimientos primitivos, como son la clasificación y ordenamiento de dichas acciones, se utilizan metodologías de aprendizaje no supervisado y técnicas para ordenar los movimientos primitivos como es el caso de la técnica conocida como *Bag of Features* [8].

Usar metodologías de aprendizaje no supervisado, ofrece herramientas para dar solución a los problemas que presenta al utilizar aprendizaje supervisado, ya que clasificar manualmente los datos, conlleva a la pérdida de información importante, debido a que una persona no es capaz de seleccionar correctamente el conjunto de datos que representa un movimiento primitivo [4]. También el uso de metodologías de aprendizaje no supervisado ayuda a dar soluciones a casos de extracción de características, ya que este trabajo se pretende utilizar varios sensores y de diferente tipo [6].

Por lo tanto, queda claro que usar metodologías de aprendizaje no supervisado para el reconocimiento de los movimientos primitivos de personas para sistemas captura de datos multimodales (Kinect®), IMU y EMG), eliminan las falencias al usar aprendizaje supervisado, ya que se da un mejor uso de los datos entregados para la detección y la clasificación de los movimientos primitivos, el cual se aprovechará para un mejor rendimiento en la aplicación que vayan a

requerir su uso. Esto con el fin de seguir avanzando en el estado del arte, debido a que existen pocos trabajos que combinen todos los elementos anteriormente mencionados.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Diseñar una metodología adecuada para la detección y clasificación de actividades usando lenguaje no supervisado para la fusión de datos de los sensores IMU, EMG y Kinect®.

2.2. OBJETIVOS ESPECÍFICOS

1. Utilizar una base de datos probada que contenga información provista por los tres sensores definidos en este proyecto.
2. Determinar una metodología de extracción de características y fusión de características, utilizando los datos provenientes de los sensores.
3. Determinar una metodología de clasificación de movimientos primitivos, usando aprendizaje no supervisado.
4. Implementar la metodología de clasificación de actividad usando HMM con información provista por la clasificación de movimientos primitivos.
5. Validar estadísticamente la metodología diseñada.

3. ESTADO DEL ARTE

El reconocimiento de actividad humana ha traído grandes aportes a la ciencia de la computación, debido a la gran cantidad de investigaciones que han sido realizadas en este tema. De forma breve, se dará una explicación de los últimos avances y aportes de las metodologías de reconocimiento automático de las actividades humanas en ambientes no estructurados.

Un enfoque común en las investigaciones sobre el reconocimiento de las acciones humanas, es el uso de la información proveniente de los acelerómetros o unidades de masa inercial (IMU). En un trabajo realizado por Takashi Aoki, Gentiane Venture y Dana Kulić en el 2013, recolectaron datos de los brazos al golpear de distintas formas un Taiko (tambor japonés) por medio de IMU's, para explorar las posibilidades del reconocimiento de patrón usando velocidades angulares [2]. En [11], se propone que los IMU's pueden ser potencialmente usados en sistemas de navegación de visión guiada, obteniendo resultados de hasta un 99 % precisión, cuando sólo un IMU era colocado en el pecho de la persona. Otros sensores muy usados son los electromiográficos, ya que además de ser útiles para el monitoreo de actividades musculares, sirven para interconectar interfaces humano-máquina, como es el caso de usar redes neuronales artificiales para distinguir entre un musculo tensionado o relajado, con promedio de éxito de más de 97 % [12]. En [13], utilizan algoritmos PDR (*Pedestrian Dead Reckoning*) en conjunto con los sensores EMG, para obtener los datos de la fuerza ejercida por la pierna dependiendo de la longitud abarcada por cada paso al andar, y al usar técnicas de clasificación como HMM y GMM, se obtienen resultados con una relación de error entre 0.03 % y 4.35 %. El inconveniente al usar sensores IMU's y electromiográficos es que no logran identificar objetos cuando se está clasificando actividades en donde esté involucrado un objeto [14].

El uso de sistemas de visión también también ha traído grandes aportes al reconocimiento de la actividad humana, debido a la gran diversidad de estos sistemas, como el caso del Kinect[®], donde en el 2015, Andrés Calvo determina que al usar la técnica SVM como método de clasificación, junto con el Kinect[®], consiguió unos resultados de hasta el 96.57 %, en la clasificación de ciertas actividades rutinarias [4]. Otras casos se centran en segmentar los

objetos que interactúan con la persona, ya que al tener en cuenta los objetos, discriminar ciertas actividades de otras resulta ser más eficiente de esta forma [9]. Aun así, usar el Kinect® puede presentar problemas si hay cambios constantes de iluminación y si existen oclusiones parciales sobre el objeto o persona, puede generar interferencias en el reconocimiento [2, 3, 14, 5].

Por lo tanto se ha decidido utilizar una base de datos multimodal que contenga la información conjunta de los sensores mencionados, y se aproveche las ventajas que ofrece cada uno de estos, como es el caso con el trabajo propuesto por [4], el cual fusiona los sensores Kinect®, IMU y EMG, y así lograr obtener resultados cercanos al 100 % de precisión, en comparación que cuando se realizan las mismas pruebas con cada uno de los sensores por separado.

Para encontrar las características mas relevantes de la base de datos (donde en este caso son los puntos articulados del Kinect® y los sensores IMU y EMG), se aplica técnicas de selección de características. Una de las técnicas mas utilizadas se llama *ReliefF*, donde [15], la utiliza para evaluar las medidas de cada etiqueta de 10 base de datos referenciales y así trabajar con menos características con una pérdida de rendimiento insignificante.

Otro enfoque útil es usar metodologías de segmentación. [8] utiliza aproximaciones basadas en cadenas de coincidencia conocidas en estadística como *Bag of Features* para descomponer las actividades en movimientos primitivos.

El aprendizaje no supervisado se ha presentado hoy en día como una herramienta útil, ya que aun cuando el enfoque supervisado es simple de implementar y es bastante utilizado, trae problemas como el de no generar umbrales para segmentar adecuadamente cada movimiento primitivo en un reconocimiento de actividad [4]. También se tiene el caso que hay actividades que no se diferencian unas de otras debido a su corta o larga duración de tiempo, y que dentro de estas hay acciones que co-ocurren y tienen una relación temporal similar entre ellas. Por lo tanto se utiliza aprendizaje no supervisado para percibir esas pequeñas diferencias [9]. Otro caso en que se aplica este tipo de aprendizaje es el que propone [5], que utiliza la técnica HACA (*Hierarchical Aligned Cluster Analysis*), para demostrar que estas técnicas superan en ciertos casos las técnicas de aprendizaje supervisado como son la HDP-SLDS (*Hierarchical Dirichlet Process Prior Switching Linear Dynamical System*) y la PS-SLDS (*Parametric*

Segmental Switching Linear Dynamical System).

4. MARCO CONCEPTUAL

4.1. Selección de Características

La Selección de características es importante para reconocer los factores mas relevantes de una base de datos. En sí se puede tener una que posee alta dimensionalidad, la cual puede tener características que son irrelevantes, engañosas o redundantes, provocando un incremento en el tamaño del espacio de búsqueda y dificultando el proceso de aprendizaje. Por lo tanto, la selección de características se puede utilizar para disminuir la complejidad y dar una representación simple de los datos, con solo escoger y trabajar con las características más relevantes de la base de datos [16].

4.2. Clasificación

Metodología usada en el aprendizaje de máquinas, donde su función es estimar un modelo matemático que separa o identifica un cuerpo o entidad usando información previamente suministrada. Existen muchos procesos de clasificación, como es el caso de clasificación supervisada o no supervisada, aprendizaje por esfuerzo o aprendizaje multi-tarea. Las técnicas de clasificación más comunes son *Maximun Likelihood Classification* (MLC), *Gaussian Mixture Model* (GMM), *Infnite Gaussian Mixture Model (IGMM)*, *Linear Discriminant Analysis* (LDA), *K-Nearest Neighbor* (KNN), *Markov Machine Model* (HMM), Redes Neuronales Artificiales (RNA), *Support Vector Machine* (SVM), entre otros [4, 17].

4.2.1. Clasificación supervisada y no supervisada

Metodología que se encarga de etiquetar las características especiales de una escena, formando un umbral entre diferentes clases a identificar. Para realizar este procedimiento existen dos tipos de clasificación que son la no supervisada y la supervisada. La clasificación no supervisada se basa en criterios donde no existe un a “priori” durante las etapas de entrenamiento, por lo tanto, su codificación es compleja [9]. Mientras que por otra parte está la clasificación

supervisada, donde se tiene un conocimiento previo de los datos de inicialización, por lo cual, el tiempo que toma realizar todo este proceso tiende a ser más rápido que en la clasificación no supervisada en muchos casos, pero también puede causar errores de clasificación.

4.3. *Machine Learning*

El aprendizaje automático de maquina o *Machine Learning*, es una rama de la ciencia de la computación, la cual, desarrolla algoritmos que sean capaces de aprender automáticamente a través de experiencias o ejemplos. El sistema en sí, aprende del cambio del ambiente y se adapta a este [18]. Cualquier metodología que aplique *machine learning*, debe seguir los siguientes pasos, en donde el primero es seleccionar un modelo que sea candidato, y el segundo, se estiman los parámetros del modelo usando los datos disponibles y los algoritmos de aprendizaje. Por lo tanto, se puede decir que el usuario escoge un modelo empírico, y luego, este emplea un algoritmo de aprendizaje para estimar los parámetros del modelo [19].

4.4. *Clustering*

Uno de los problemas de investigación más clásicos de reconocimiento de patrones y de extracción de datos se conoce como agrupamiento (en ingles se conoce como *clustering*). Hace parte del núcleo principal de las metodologías de aprendizaje no supervisado, y es una de las técnicas fundamentales en el aprendizaje de máquina. Al tratar con datos, un algoritmo de agrupamiento puede dividir un conjunto de datos en algunas particiones grupales y estas particiones son elementos de de una división grupal para el conjunto de datos generados por el algoritmo. Algunos de los algoritmos más famosos que emplean la técnica de clustering son K-means, HMM, GMM [20, 21, 22].

4.5. *Desempeño de un clasificador*

Es la forma de indicar la capacidad de etiquetar una observación correctamente. Algunos indicadores se definen como [23]

True Positives (tp): Los ejemplos que pertenecen a la clase fueron reconocidos correctamente.

True negatives (tn): El número de ejemplos reconocidos correctamente que no pertenecen a la clase.

False positives (fp): Los ejemplos fueron incorrectamente asignados a la clase.

False negatives (fn): Los ejemplos no fueron reconocidos como ejemplos que pertenecían a la clase.

Accuracy: Eficacia en general de un clasificador.

$$\frac{tp + tn}{tp + fn + fp + tn} \quad (1)$$

Precision: El acuerdo de clase entre las etiquetas de datos con las etiquetas positivas dadas por el clasificador.

$$\frac{tp}{tp + fp} \quad (2)$$

Recall (sensibilidad): Eficacia de un clasificador para identificar etiquetas positivas.

$$\frac{tp}{tp + fn} \quad (3)$$

Fscore: Relacion entre las etiquetas positivas de los datos y las que son dadas por el clasificador.

$$\frac{(\beta^2 + 1)tp}{(\beta^2 + 1)tp + \beta^2 fn + fp} \quad (4)$$

Specifity: Que tan efectivo un clasificador identificando las etiquetas negativas.

$$\frac{tn}{fp + tn} \quad (5)$$

AUC: La habilidad del clasificador para prevenir clasificaciones negativas.

$$\frac{1}{2} \left(\frac{tp}{tp + fn} + \frac{tn}{tn + fp} \right) \quad (6)$$

4.6. Dispositivos de captura de movimiento

4.6.1. Unidad de Masa Inercial (IMU)

Sensor electrónico que realiza la medida integrada de parámetros de movimiento en seis grados de libertad (DOF). Así se puede obtener la posición de un objeto en movimiento dentro del espacio, el cual incluye tres coordenadas (X, Y, y Z), y tres ángulos de rotación (*pitch*, *yaw* y *roll*) [24]. Los IMU tradicionales usan giroscopios para medir la relación de ángulo, y acelerómetros para proveer la magnitud de la aceleración [25].

4.6.2. Electromiografía (EMG)

La electromiografía es un método para monitorear las actividades musculares y las neuronas motoras (células nerviosas que controlan los músculos), por medio de ondas oscilantes las cuales, son perturbaciones provocadas por los músculos que se repiten constantemente y aumentan su amplitud durante la activación de estos. Utilizando tecnologías que aplican EMG, se puede capturar el movimiento humano directamente al medir y decodificar la actividad muscular, y de esta manera, usar esta información en aplicación como diagnostico medicina, rehabilitación, e interacción humano-computadora [12, 13].

4.6.3. Kinect[®]1.0

Dispositivo creado por Microsoft para su plataforma de juegos Xbox 360 (luego adaptado a computadores), fue diseñado para captar movimiento y audio, donde el usuario puede interactuar usando comandos de gesto y de voz. Una de las características del Kinect[®], es poseer un sensor que usa una técnica de luz estructurada para generar un mapa de profundidad de la escena, donde un patrón de puntos son proyectados por una proyector de luz IR y capturados

por una cámara IR desplazada. Al calcular las disparidades de los puntos proyectados entre el proyector IR y la cámara, la profundidad de la escena es obtenida [26].

5. MARCO TEÓRICO

5.1. Técnicas de Selección de Características

5.1.1. Análisis de Componentes Principales

La técnica conocida como Análisis de Componentes Principales (PCA; *Principal Components Analysis*), es una técnica no paramétrica que tiene como tarea, el extraer información relevante de un conjunto complejo de datos al reducir la dimensionalidad de un conjunto de datos de alta dimensionalidad en un subespacio de menor dimensión. Para esto, esta técnica transforma un número posible de variables correlacionadas en un número menor de variable no correlacionadas, las cuales son llamadas componentes principales (estos componentes poseen un alto grado de varianza).

Se tiene una matriz X de $N \times D$, la cual representa un conjunto de datos, donde las filas N son las muestras y las columnas D son las variables. El objetivo es obtener los valores mas relevantes de X , para esto es necesario encontrar los valores propios de la matriz de covarianza de X , donde esta última se obtiene de la siguiente forma [27]:

$$C_x = \frac{1}{(D - 1)} X X^T \quad (7)$$

5.1.2. *ReliefF*

Algoritmo de selección de características utilizado en clasificación binaria. Este algoritmo es una versión mejorada del algoritmo *Relief*, ya que improvisa la confiabilidad de la aproximación de probabilidad, haciendola robusta a datos incompletos y generalizandola a problemas multiclases.

Para encontrar las características y vectores más relevantes de X , se toma la instancia *same-class* más cercana (*near-hit*) y la instancia *different-class* más cercana (*near-miss*), y las aplicamos a la siguiente función [15]:

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \quad (8)$$

Donde W es el peso de cada carecterística.

5.2. Técnicas de Agrupamiento

5.2.1. *Spectral Clustering (SC)*

Spectral Clustering es una técnica de clasificación basada en encontrar los datos de los clusters usando vectores propios de la matriz de similaridad de X . Este tipo de metodología permite construir una gráfica de pesos en el que los nodos corresponden a los puntos de datos y las esquinas están relacionadas con la distancia que existe entre los puntos.

Matemáticamente dado un conjunto de puntos $X = \{x_1, \dots, x_n\}$ en R^L , los cuales se agrupan en k subconjuntos:

1. De la matriz de similaridad $A \in \mathbb{R}^{n \times n}$ definida por $A_{ij} = \exp(-\|x_i - x_j\|^2 / 2^{2\sigma})$ si $i \neq j$, y $A_{ii} = 0$.
2. Definir D como la matriz diagonal en donde su (i, i) -elemento es la suma de fila i -ésima de A , y así construir la matriz $L = D^{-1/2} A D^{-1/2}$.
3. Encontrar s_1, s_2, \dots, s_k , para los *eigenvector* más grandes de k de L (en el caso de que los valores propios repetidos, estos serán ortogonales), y formar la matriz $S = [s_1, s_2, \dots, s_k] \in \mathbb{R}^{n \times k}$ al apilar los vectores propios en columnas.
4. Formar la matriz Y de S al re-normalizar cada fila de S para tener una unidad de control (i.e. $Y_{ij} = S_{ij} / (\sum_j S_{ij}^2)^{1/2}$).
5. Tratar cada fila de Y como un punto en \mathbb{R}^k , clasificarlos en k *clusters* con *K-means* o con otro algoritmo ya que esto sirve para minimizar la distorsión.

6. Finalmente, asignar los puntos originales x_i al cluster j si y solo si la fila i de la matriz Y fue asignada al cluster j .

El parámetro de escala σ^2 , controla que tan rápido la afinidad A_{ij} decae con la distancia x_i y x_j [28].

5.2.2. *K-Means Clustering (Kmc)*

Es una de las técnicas más utilizadas de agrupamiento de datos, debido a su simplicidad en comparación a otras técnicas de clustering, ya que requiere que el número de clustering sea pre-definido, convirtiéndose simplemente una tarea de ensayo y error.

El objetivo de *K-means* es dividir las muestras en $k(\leq N)$ conjuntos $S = \{S_1, S_2, \dots, S_k\}$, para así minimizar la suma de cuadrados entre clusters (WCSS; *Within-Cluster Sum of Squares*) [29].

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (9)$$

donde μ_i es la media de los puntos en S_i . En sí *K-means* consta de dos pasos. El primer paso se conoce como *asignación*, en el que cada punto de n datos es asignado a la media más cercana. El siguiente paso se conoce como *actualización*, en el que las medias son ajustadas para acomodar las medias de la muestra de los puntos de datos que son responsables.

5.3. Técnicas de Clasificación

5.3.1. Modelos Ocultos de Marcov (HMM)

Un HMM (*Hidden Markov Model*) es un proceso estocástico en el que existe otro proceso del mismo tipo por debajo, el cual se encuentra oculto, y este solamente es visto a través de otro conjunto de procesos estocásticos, los cuales producen la secuencia de los símbolos observados.

Una definición más formal de HMM es la siguiente [30, 31]:

$$\lambda = f(O, Q, A, B, \pi) \tag{10}$$

Donde:

$O = \{o_1, o_2, \dots, o_M\}$ son las observaciones (estados visibles).

$Q = \{q_1, q_2, \dots, q_N\}$ son los estados ocultos.

N es el número de estados del modelo.

M es el número de símbolos de observación O_k .

$A = \{a_{ij}\}$ es un arreglo de transición de los estados, y a_{ij} es la distribución de probabilidad de transición de estados desde el estado i hasta el siguiente estado j .

$B = \{b_i(k)\}$ es un arreglo de observación, y $b_i(k)$ es una distribución de probabilidad del símbolo de observación en el estado i .

$\pi = \{\pi_i\}$ es el arreglo de probabilidad inicial, y π_i es la probabilidad del estado inicial i .

Los tres problemas para HMM: Estos problemas se han separado y se han clasificado en tres grupos distintos para mayor simplicidad en su entendimiento y solución. Los tres problemas que se enfrenta HMM son los siguientes [31]:

El primer problema es sobre la evaluación. Se tiene la secuencia de observación $O = O_1, O_2, \dots, O_T$, y el modelo $\lambda = (A, B, \pi)$, que se puede hacer para computar $Pr(O|\lambda)$ (probabilidad de la secuencia de observación). En otras palabras, dado un modelo y una secuencia de observaciones, se puede computar la probabilidad en la cual la secuencia observada fuera producida por el modelo. El procedimiento más directo para dar solución es enumerando toda secuencia de estados posibles de longitud T (número de observaciones).

El segundo problema se relaciona en descubrir las partes ocultas del modelo. Se tiene la secuencia de observación $O = O_1, O_2, \dots, O_T$, donde el problema es elegir la secuencia de estado $I = i_1, i_2, \dots, i_T$. En sí, es un típico problema de estimación. A menudo se utiliza un criterio de optimización para resolver este tipo de problemas. Lastimosamente hay mucho

de este tipo de criterios y causa que escoger uno de estos el cual sea adecuado no sea sencillo para la secuencia de estado descubierta. Un uso que se le puede dar a la secuencia de estados descubierta, es aprender acerca de la estructura del modelo, y obtener una estadística promedia como un caso de los estados individuales.

El tercer problema se relaciona en optimizar los parámetros del modelo los cuales describen el origen de la secuencia observada. Lo que se trata es maximizar $Pr(O|\lambda)$ ajustando los parámetros del modelo $\lambda = (A, B, \pi)$. Lo que se hace, es llamar esto, una secuencia de entrenamiento, debido a que es usada para entrenar el modelo. El problema de entrenamiento es crucial en las aplicaciones que usan HMM, porque se busca adaptar el modelo a los datos de entrenamiento observados. Para dar solución a este problema se usan métodos como el Baum-Welch, o técnicas de gradiente para optimización.

5.3.2. K-Vecinos más Cercanos

Conocido en inglés como *k-Nearest Neighbors (KNN)*, es una metodología que clasifica cada dato al escoger la mayoría de la clase entre las k muestras más cercanas en los datos de entrenamiento. Para el caso de este trabajo, se utiliza una transformación gaussiana para normalizar los valores de las características entre un valor de $[0,1]$, y usar una distancia de similaridad de datos, para medir la separación entre muestras [32].

6. METODOLOGIA

La metodología de este proyecto consta de cuatro etapas fundamentales las cuales son:

- Diseño de la base de datos.
- Selección de Características.
- El reconocimiento de Movimientos Primitivos.
- El reconocimiento de Actividad.

La figura 1, muestra resumidamente el proceso de como se diseñó la metodología de reconocimiento de actividad por medio de aprendizaje no supervisado en un sistema multimodal.

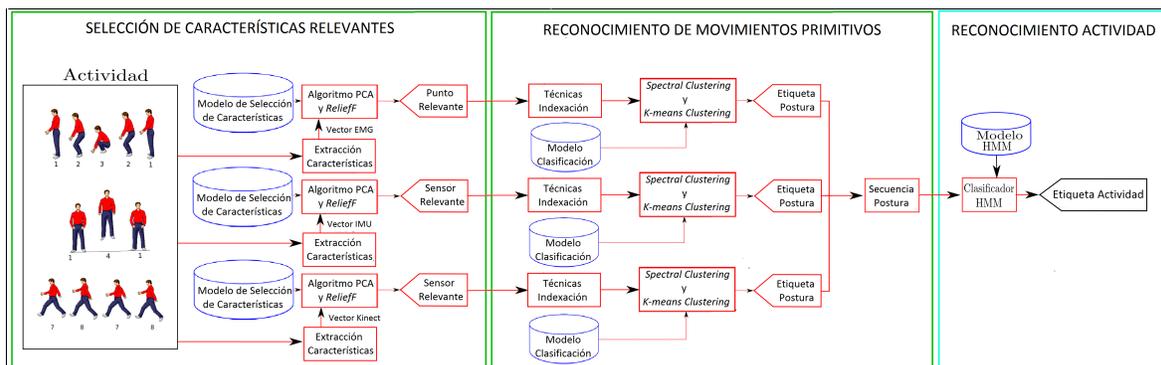


Figura 1. Gráfico del proceso llevado a cabo para el reconocimiento de actividad

6.1. BASE DE DATOS

La base de datos fue construida usando la información de 16 personas de diferente constitución física y de diferente sexo, las cuales ejecutaban cada una de las cinco actividades tres veces (tabla 1). Esta información se extrajo al emplear simultáneamente tres dispositivos de medición de movimiento (Kinect®), EMG, IMU). La base de datos es privada y fue construida por el grupo de investigación en Ingeniería Electrónica de la Universidad Tecnológica de Pereira.

6.1.1. Extracción de Características

El proceso de extracción de características fue dividido en tres módulos, en donde un módulo pertenece a la extracción de características del Kinect®), en el otro módulo es para la extracción de características de la red de los sensores IMU y el último módulo para la extracción de características de la red de los sensores EMG. En la siguiente sección se describirá como es el proceso de extracción y que métodos se aplican a cada módulo.

Extracción de características de los sensores IMU

La extracción de características de la red de sensores IMU usa la información proveniente de cuatro acelerómetros tri-axiales (a_x, a_y, a_z) que están localizados en diferentes partes del cuerpo (ver figura 2). Como estos dispositivos tienen tres grados de libertad, es posible calcular las orientaciones *Roll* y *Pitch* utilizando la conversión a coordenadas esféricas. Gracias a esto, es posible obtener el vector $I_k = [a_x \ a_y \ a_z \ pitch \ roll]_{1 \times 15}$ para cada instante de tiempo, donde k es el k -ésimo acelerómetro, $k = \{1,2,3,4\}$. Estos sensores proveen datos a una frecuencia de 30 *Hz*, sin embargo, los datos se agrupan en una ventana de tres muestras, generando que los datos estén a una frecuencia de 10 *Hz*. El vector de características para cada sensor es el siguiente:

$$IMU_i = [m_{ax}, m_{ay}, m_{az}, m_{ar}, m_{ap}, v_{ax}, v_{ay}, v_{az}, v_{ar}, v_{ap}, AI, VI, SMA, EVA_{3 \times 1}, AAE_x, AAE_y, AAE_z, AAE_r, AAE_p, MI_{1 \times 3}]_{1 \times 24}$$

$$i = 1, \dots, K$$

donde,

$m_{ax}, m_{ay}, m_{az}, m_{ar}, m_{ap}, v_{ax}, v_{ay}, v_{az}, v_{ar}, v_{ap}$, son la media y la varianza de la aceleración de las coordenadas $x, y, z, roll$ y $pitch$. La intensidad del movimiento para x, y y z se representa como MI . AI y VI es la media y varianza de MI . SMA es la señal de la magnitud del área normalizada. EVA son los valores propios de las direcciones dominantes. $AAE_x, AAE_y, AAE_z, AAE_r, AAE_p$, son la energía de la aceleración promedia de las coordenadas $x, y, z, roll$ y $pitch$. ARE es la energía de rotación promedia [33].

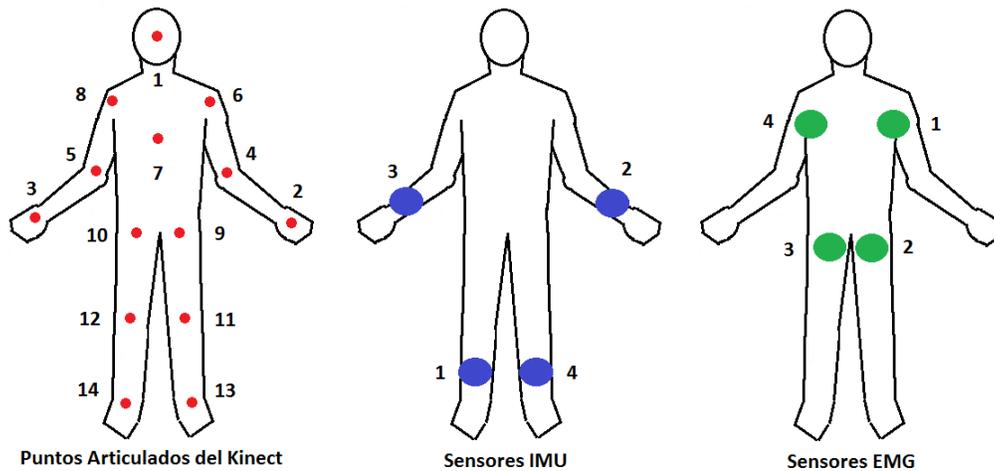


Figura 2. Posición de los puntos articulados del Kinect y de los sensores IMU y EMG en el cuerpo de una persona.

Extracción de características del Kinect®

El Kinect® usa catorce puntos articulados sobre el cuerpo humano (ver figura 2) a una frecuencia de 30 *Hz* para obtener información. Sin embargo, los datos están agrupados en una ventana de tres muestras, generando que los datos estén a una frecuencia de 10 *Hz*. El vector de característica para cada punto articulado se muestra a continuación: $KI = [p_{09x1}; m_x; m_y; m_z; m_r; m_t; v_x; v_y; v_z; v_r; v_t; velP]_{20 \times 1}$

Donde las coordenadas polares yacen en p_0 . Las siguientes características son la media y la varianza de las coordenadas x , y , z , r y θ , y $velP$ es la velocidad de cada punto [34].

Extracción de características de los sensores EMG

La información de cuatro grupos musculares (ver figura 2) se obtiene usando ocho electrodos de superficie en conexión diferencial (esto permite obtener cuatro medidas de EMG) con una frecuencia de muestreo de 2 *KHz*. Al usar una descomposición Wavelet deabauchie 7 de orden 4, y un filtro de banda con una frecuencia de corte de 70 *Hz* y 500 *Hz*, se obtiene el siguiente vector de características:

$$EMG = [IEMG, MAV, MMAV_{1 \times 2}, SSI, VAR, RMS, WL, ZC, WAMP, LOG, V3, ACC,$$

DASDV, MNP, TTP, MDF, MNF, PKF, SM1, SM2, SM3, PSR, SCC, TM3, TM4, TM5] 27x1

Donde el nombre de cada característica es el siguiente:

Integrated EMG, Mean Absolute Value, Modified Mean Absolute Value, Simple Square Integral, Variance of EMG, Root Mean Square, Waveform Length, Zero Crossing, Willison Amplitude, Log Detector, v-Order 3, Average Amplitude Change, Difference Absolute Standard Deviation Value, Mean Power, Total Power, Median Frequency, Mean Frequency, Peak Frequency, 1st, 2nd, and 3rd Spectral Moments, Power Spectrum Ratio [35].

6.2. Selección de características

En esta etapa se selecciona las características más relevantes de cada uno de los dispositivos de reconocimiento o en este caso, cuales sensores o puntos articulados son los más importantes, con el fin de reducir la complejidad del sistema y mantener resultados óptimos.

Para realizar esto se utilizan dos metodologías conocidas como *PCA* y *ReliefF*, las cuales determina N los sensores o puntos articulados más importantes al realizar 30 iteraciones cada una. Ya al momento de escoger cuál de las dos metodologías es más determinante y cuál debe ser el número mínimo de sensores o de puntos articulados para que la pérdida de rendimiento sea insignificante, depende de los resultados de las técnicas de clustering.

6.3. RECONOCIMIENTO DE MOVIMIENTOS PRIMITIVOS

Las metodologías enfocadas en movimientos primitivos se basan en las metodologías de reconocimiento de voz humana, debido a que en el reconocimiento de voz, las frases se dividen en palabras aisladas o fonemas generando una secuencia que permite construir un modelo de clasificación de N fonemas compartidos para el conjunto de palabras a identificar. Por lo tanto se forman bloques básicos para construir palabras y frases de manera jerárquica [8]. Así que cada actividad se puede representar mediante una secuencia de movimientos, los cuales generan un código único que permite obtener un factor discriminante en la clasificación de actividad. El

proceso de reconocimiento de movimientos primitivos consta de tres etapas fundamentales, las cuales son: codificación de movimientos primitivos, asignación de etiquetas y clasificación de movimientos.

6.3.1. Codificación de movimientos primitivos

Para determinar el número exacto de movimientos primitivos para el conjunto total de actividades, es esencial escoger una metodología que indique un umbral de cuantos agrupamientos son necesarios. Para lograr esto, se saca una gráfica de similaridad binaria para cada uno de los tipos de sensores (Kinect[®], IMU y EMG), como se observa en la figura 3. Esta gráfica se obtiene al determinar la similaridad de los datos, donde los datos que pertenecen a la misma clase, se asigna el valor 1, y si pertenecen a diferentes clases, entonces se asigna el valor 0. Las clases se asignan por medio de técnicas de agrupamiento o clustering, las cuales se mencionan en la sección 6.3.2.

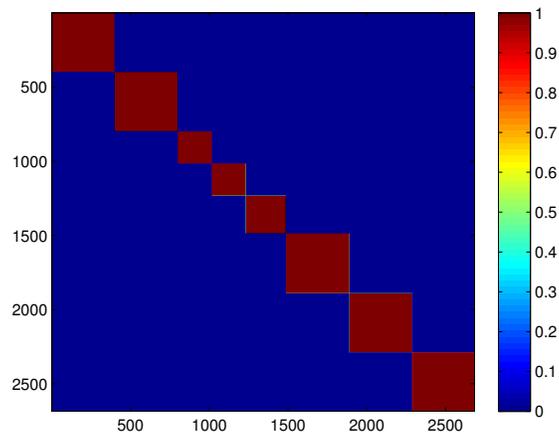


Figura 3. Gráfica de similaridad.

6.3.2. Asignación de etiquetas

Para esta etapa se utilizó dos técnicas de clasificación no supervisada, donde la primera se conoce como *K-means* y la segunda es *Spectral Clustering*. Estas metodologías de clustering fueron aplicadas a cada uno de los sensores. Para el proceso de evaluación se divide en dos

partes. La primera utiliza cuatro técnicas de indexación, las cuales son: *Adjusted Rand Index*, *Purity Index*, *Jaccard Index* y *Accuracy Index*, donde cada una evalúa el rendimiento de cada técnica de agrupamiento por medio de las etiquetas entregadas por estos procesos. Estas cuatro técnicas se utilizan para tener mayor confiabilidad en los resultados entregados, ya que cada una tiene un funcionamiento característico que las diferencia de las demás, así que lo ideal es comparar los resultados y elegir el número de agrupaciones con mejor rendimiento. La segunda parte, es una extensión de la primera parte, ya que vuelve a comprobar el rendimiento de los clusters, pero esta vez lo hace por medio de una matriz de confusión. Este procedimiento se hace con el fin de dar más validez a los resultados expuestos por las técnicas de indexación. Además que se comprueban las combinaciones de cada uno de los sensores (Kinect+IMU, Kinect+EMG, EMG+IMU, Kinect+IMU+EMG). En esta parte se realiza un cambio del funcionamiento del *Spectral Clustering*, el cual se crea una variable d cuyo funcionamiento es el de almacenar los elementos de la diagonal de la matriz D . El tamaño de d varía dependiendo del número de clusters más la cantidad de dimensiones extras que el usuario desea agregar, aunque se determina que el máximo valor para esto sea de +5, ya que como se evidencia en la sección 7.2, la media y la varianza de los resultados comienzan a decrecer y a crecer respectivamente, cuando se agrega un valor de +4 y +5 en la mayoría de los casos.

A cada procedimiento de etiquetación se aplica una estrategia de validación cruzada, la cual fracciona la base de datos en un 70 % para el entrenamiento y 30 % para la evaluación; estos procesos utilizan 30 iteraciones de Montecarlo.

6.3.3. Clasificación de Movimientos

Esta etapa utiliza el algoritmo k-vecinos más cercanos para el proceso de evaluación de datos. Este algoritmo usa algoritmos de distancia para comprobar la similaridad entre muestras y así escoger la clase a la que corresponden.

6.4. RECONOCIMIENTO DE ACTIVIDAD

El proceso de clasificación de actividad consiste en la clasificación de las actividades establecidas por [4] que se muestran en la tabla 1. Para realizar esto se utiliza un módulo de fusión de datos que se encarga de centralizar la información de las máquinas de vectores de soporte, y luego se realiza la clasificación usando Modelos Ocultos de Markov (HMM; *Hidden Markov Model*)

Tabla 1. Tablas de Actividades

ACTIVIDAD	ETIQUETA
Quedarse Quieto	1
Agacharse y levantarse	2
Saltar	3
Levantar brazo derecho	4
Trotar	5

6.4.1. Fusión de datos

Para la fusión de datos se utilizar dos tipos de metodologías muy diferentes. La primera de ellas es la fusión por medio de las etiquetas que entrega los clasificadores, y la otra es la fusión por medio de las características de las muestras.

Fusión de datos por medio de etiquetas

En la etapa 6.3.2, cada clasificador genera una etiqueta, EI corresponde a la salida del clasificador (Spectral Clustering o K-means) con información provista de la red de sensores IMU, EK para el Kinect® y EE para la red de sensores EMG. Estos sensores proveen información de sensado a diferentes velocidades de muestreo, así que es necesario el diseño de un módulo de fusión de datos que centralice la información y la haga apta para la etapa de clasificación con el HMM. El proceso de fusión de datos consiste en generar un vector de características EF con las etiquetas de los clasificadores durante una ventana de observación de tres segundos. El EF se forma al concatenar linealmente cada etiqueta de los clasificadores durante la ventana

de observación, logrando un vector de características de la ecuación 11

$$EF = [[EK1 EK2 \dots EK20][EI1 EI2 \dots EI24][EE1 EE2 \dots EE27]]_{71 \times 1} \quad (11)$$

Fusión de datos por medio de características

Para esta fusión de datos, solo es necesario tomar las características de las muestras del Kinect® (*EK*), las de los sensores IMU (*EI*) y la de los sensores EMG (*EE*), y concatenarlas para crear un conjunto de datos de mayor dimensionalidad. Generando una ecuación similar a 11

6.4.2. Clasificación con Modelos Ocultos de Markov (HMM)

Este modelo de clasificación parte de un modelo entrenado, eligiendo las etiquetas de cada clase para así formar una secuencia que determine las actividades.

6.4.3. Entrenamiento y validación del HMM

El entrenamiento de los HMM se usa 24 estados y 32 centroides en la construcción del libro de códigos; este proceso escoge el modelo de mejor desempeño después de realizar 100 iteraciones. Para realizar el procedimiento de evaluación del modelo entrenado se utiliza una estrategia de validación cruzada, la cual fracciona la base de datos en un 70 % para el entrenamiento y 30 % para la evaluación; este proceso utiliza 100 iteraciones Montecarlo. Las estadísticas obtenidas consisten en el cálculo de la matriz de confusión y cálculos de indexación.

7. EXPERIMENTOS Y RESULTADOS

En esta etapa se presenta el proceso de adquisición de las características más relevantes (en este caso son los sensores y los puntos articulados) y los resultados obtenidos al validar el desempeño de la clasificación de movimientos primitivos al aplicar Montecarlo para la estimación del error. Se demuestra que usar aprendizaje no supervisado en un sistema multimodal compuesto por sensores IMU, sensores EMG y Kinect®[®], donde se escogen las características más relevantes para el reconocimiento de actividad, se obtiene un buen desempeño en la clasificación, además que se puede escalar fácilmente si en casos futuros se agregan más actividades.

Para probar dicha afirmación, todo el proceso se divide en tres partes:

1. Adquirir un umbral del número de movimientos primitivos requeridos para el reconocimiento del conjunto de actividades y se seleccionan los sensores y puntos articulados más relevantes.
2. Determinar el número de movimientos primitivos necesarios del conjunto de actividades y se realiza el análisis del desempeño de los clasificadores propuestos para cada una de las combinaciones (Kinect, IMU, EMG, Kinect+IMU, Kinect+EMG, IMU+EMG, Kinect+IMU+EMG), donde se seleccionan las combinaciones con mayor eficacia para uno y dos sensores (como solamente hay una combinación de tres sensores la cual es Kinect+IMU+EMG, pues esta ya se encuentra seleccionada por defecto).
3. Validar la clasificación de actividades humanas con el Modelo Oculto de Markov utilizando la base de datos anotada en la sección 6.4.

Antes de realizar todos estos procesos, es correcto comentar que los datos de cada uno de los sensores son normalizados con una media de valor 0 y una varianza de valor 1, para cada una de las características de los sensores. Además se equilibra el número de datos por clase debido a la descompensación que existe entre ellos. Por lo tanto se escoge un número limitado de datos por clase, que en este caso es 300. Si la cantidad de datos por cada clase sobrepasa el límite dado, se utiliza la técnica Bolsa de Palabras (*Bag of Words*) junto con *K-means Clustering*,

para evitar escoger datos que estén muy dispersos y así tener una región del espacio más concentrada y evitar lo más posible, traslapes con otros agrupamientos.

7.1. Umbralización de movimientos primitivos y selección de sensores y puntos articulados.

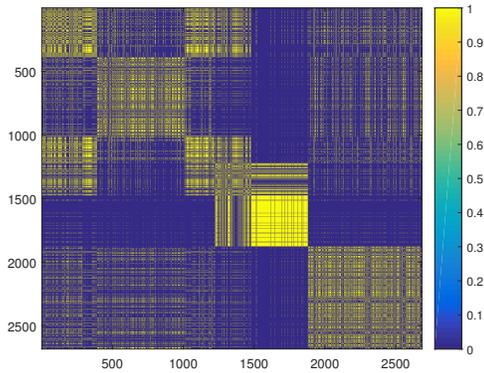
Para definir el umbral del número de movimientos primitivos necesarios para la clasificación de actividades, se realiza un cálculo de similitud binaria de clases de las muestras del Kinect®[®], de los sensores IMU y de los sensores EMG, utilizando las dos técnicas de agrupamiento. Estos resultados se muestran en la figura 4, donde el mínimo número de movimientos primitivos puede ser 4. Para el máximo número se toma como referencia el trabajo de [4], donde define 8 movimientos primitivos para clasificar las actividades.

En la tabla 2, se muestra la etiquetación de cada uno de los movimientos primitivos usando como eje principal para la etiquetación las clases definidas por [4].

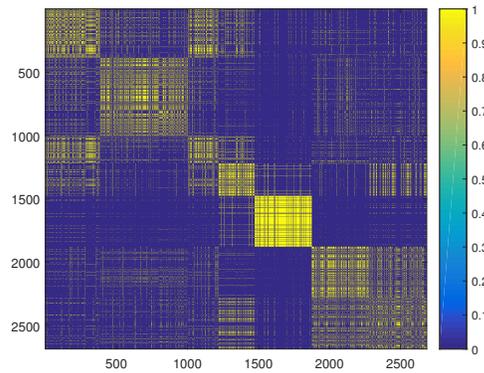
	Número de clusters				
Movimientos primitivos	4	5	6	7	8
Reposo	1	1	1	1	1
Medio agachado	2	2	2	2	2
Agachado	2	2	2	2	3
Suspendido en el aire	1	4	4	4	4
Mano levantada 1/4	5	5	5	5	5
Mano levantada 3/4	5	5	5	6	6
Paso adelante (pie derecho)	7	7	7	7	7
Paso adelante (pie izquierdo)	7	7	8	8	8

Tabla 2. Movimientos primitivos para cada etiqueta.

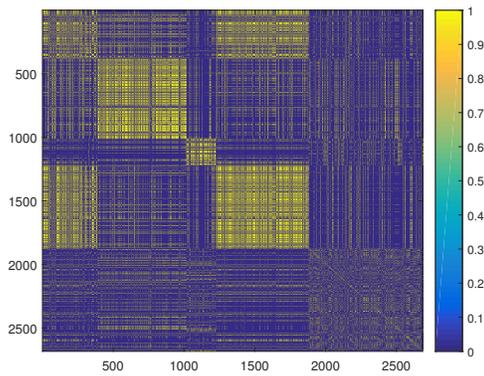
Para hallar los puntos articulados más relevantes del Kinect®[®] y los sensores IMU y EMG más relevantes se utiliza los algoritmos *PCA* y *ReliefF*. Los resultados provenientes al usar los algoritmos de selección de características, se muestran en la figura 5 y 6, donde la relevancia de cada puntos articulado y de cada sensor se determina por la barra de colores que se encuentra en dichas figuras (entre más cercano al color amarillo, más relevante es el punto articulado o sensor). Al realizar este proceso de selección de características, se revela la distribución de



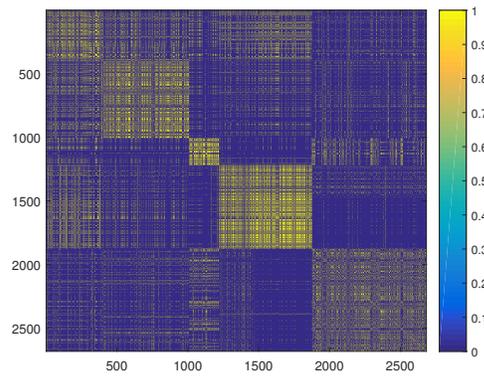
Gráfica de similaridad del Kinect (*K-means Clustering*)



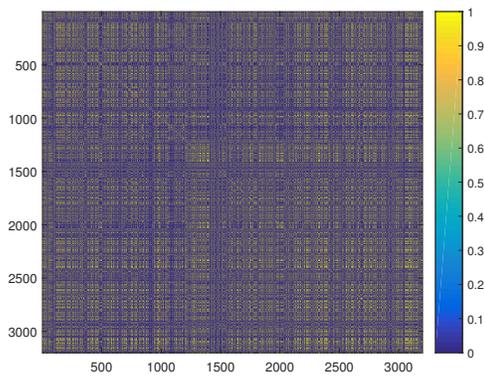
Gráfica de similaridad del Kinect (*Spectral Clustering*)



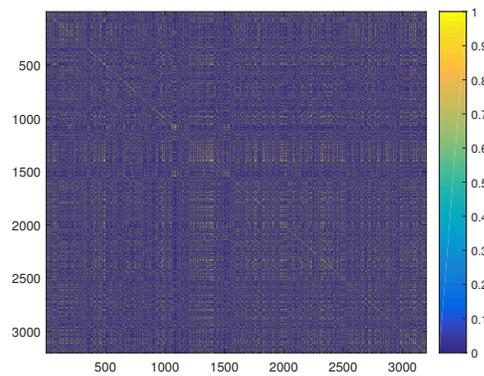
Gráfica de similaridad de los IMU (*K-means Clustering*)



Gráfica de similaridad de los IMU (*Spectral Clustering*)



Gráfica de similaridad de los EMG (*K-means Clustering*)



Gráfica de similaridad de los EMG (*Spectral Clustering*)

Figura 4. Gráficas de similaridad para cada uno de los sensores de reconocimiento.

pesos es diferente al usar ambos algoritmos, ya que en el caso de *PCA*, este le da mayor relevancia a los puntos articulados y sensores ubicados en los pies, mientras que la *ReliefF*, se da más relevancia a los brazos.

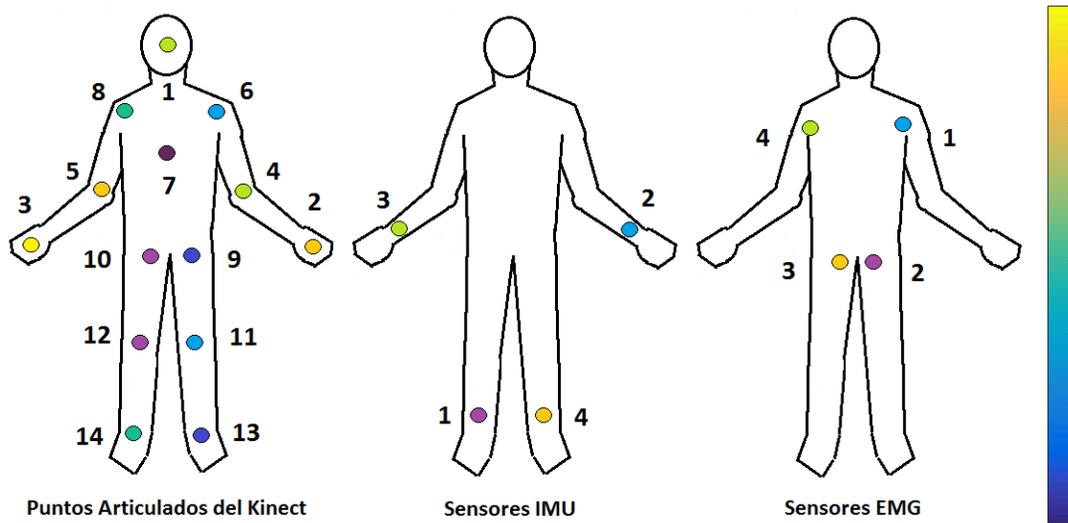


Figura 5. Desempeño de cada uno de los puntos articulados del Kinect® y de cada sensor IMU y EMG utilizando el algoritmo *ReliefF*

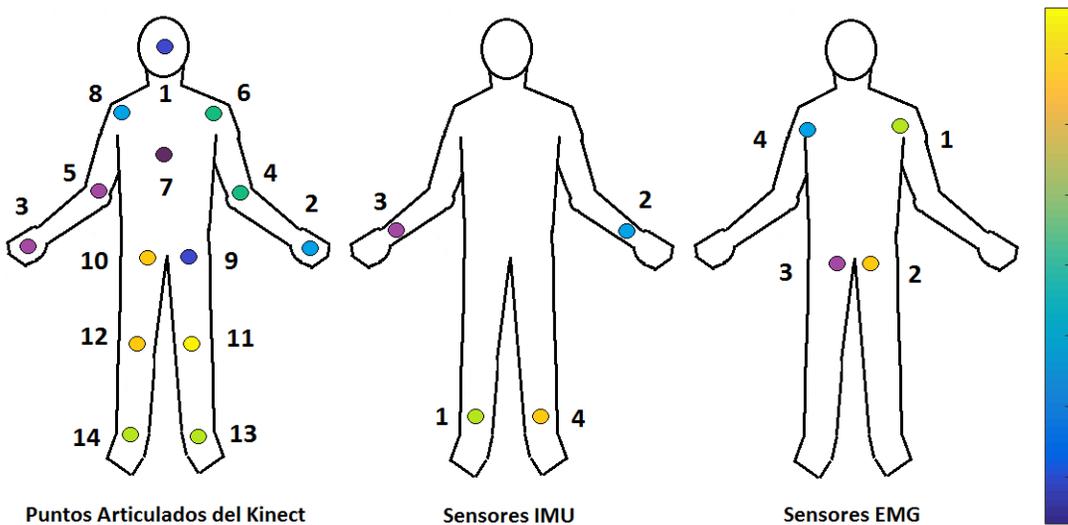


Figura 6. Desempeño de cada uno de los puntos articulados del Kinect® y de cada sensor IMU y EMG utilizando el algoritmo *PCA*

Al determinar el rendimiento de las técnicas de indexación para el caso del Kinect®, la

figura 7 muestra que *Spectral Clustering* y *K-means Clustering* con 5 *clusters* ofrecen la mejor eficiencia dado que en la mayoría de los casos su media es la más alta y su varianza es de las más pequeñas. Mientras si se utilizan 4 *clusters*, el rendimiento en general es muy pobre ofreciendo poca precisión debido a su alto valor en la varianza.

Nota: Para las siguientes figuras, realizar un acercamiento para apreciar adecuadamente el comportamiento y características de cada gráfica.

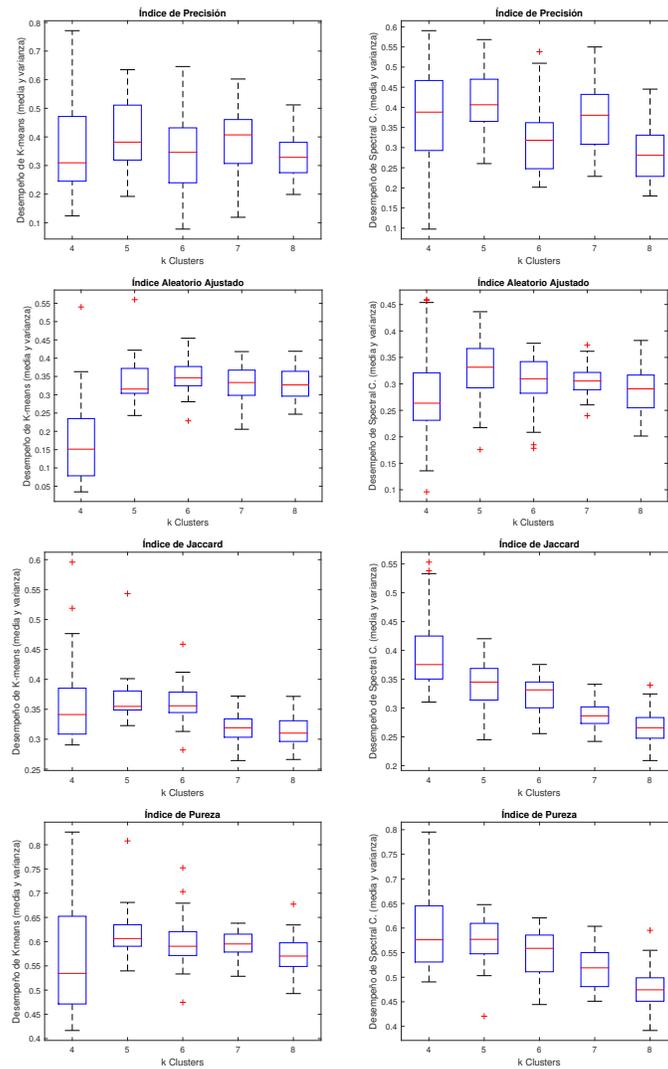


Figura 7. Desempeño de los clusters con todos los puntos articulados del Kinect®.

Al disminuir el número de puntos articulados utilizando *PCA*, el rendimiento sigue siendo muy similar como si se utilizaran todos los puntos articulados del Kinect, como se muestra en la figura 8, solo que los cambios mas notoros son en las varianzas donde algunos casos como al utilizar *K-means clustering* con 4 clusters, la varianza crece más al utilizar menos puntos articulados.

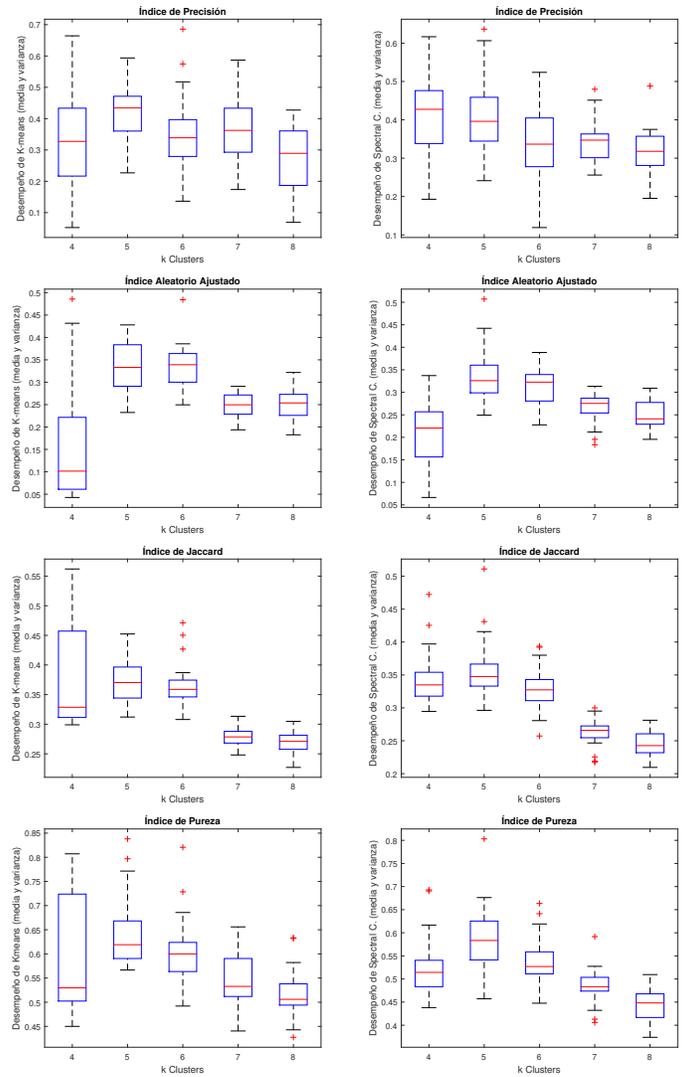


Figura 8. Desempeño de los clusters con 8 de los puntos articulados más relevantes (usando *PCA*) del Kinect®.

Con respecto a la técnica *ReliefF*, la figura 9 muestra que solo es necesario 6 puntos de los 14 puntos articulados del Kinect [®] para tener un rendimiento similar a si se utilizaran todos los puntos articulados, llegando el caso a utilizar 2 puntos articulados menos que con *PCA*, aunque se sacrifica un poco la precisión, pero la ventaja es que mas rapido computacionalmente trabajar con 6 puntos articulados que con 8 puntos articulados.

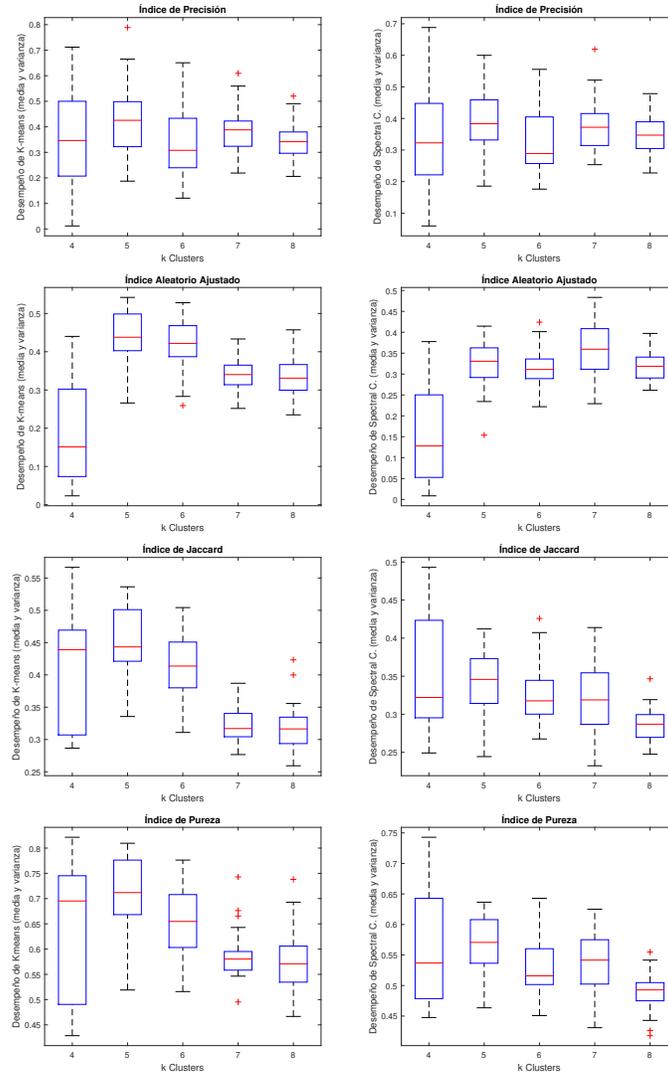


Figura 9. Desempeño de los clusters con 6 de los puntos articulados más relevantes (usando *ReliefF*) del Kinect[®].

Al utilizar todos los sensores IMU, el rendimiento cambia del *Spectral Clustering* y del *K-means Clustering* cambia dependiendo de la técnica de indexación utilizada como se aprecia en la figura 10. Por ejemplo, al utilizar índice de precisión, el uso de 5 *clusters* demuestra tener al mejor eficiencia. Mientras que si se utiliza índice aleatorio ajustado, la mejor eficiencia la tiene al usar 8 *clusters*. Con índice de Jaccard, la mejor eficiencia pasa a ser con 4 *clusters* y finalmente con índice de pureza, 5 *clusters* ofrece nuevamente el mejor rendimiento.

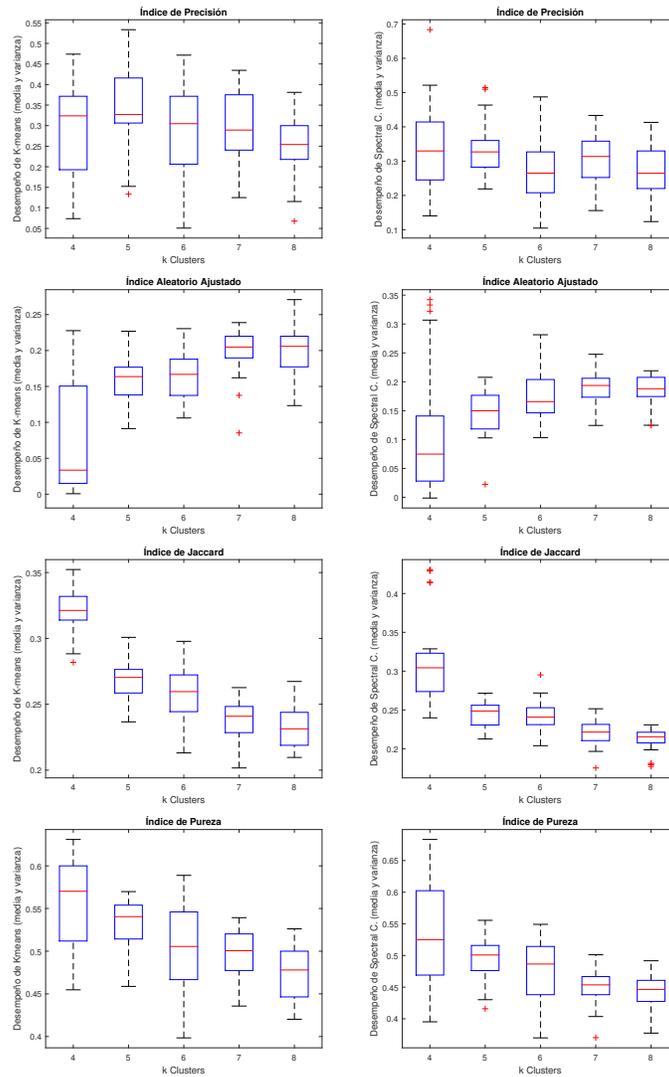


Figura 10. Desempeño de los clusters con todos los sensores IMU.

Con *PCA* se logro rebajar el número de sensores IMU a tres para tener resultados similares a si se utilizara cuatro sensores IMU (ver figura 11). La única desventaja con este es que la eficiencia tiende a bajar un poco debido al aumento en el valor de la varianza en cada uno de los resultados.

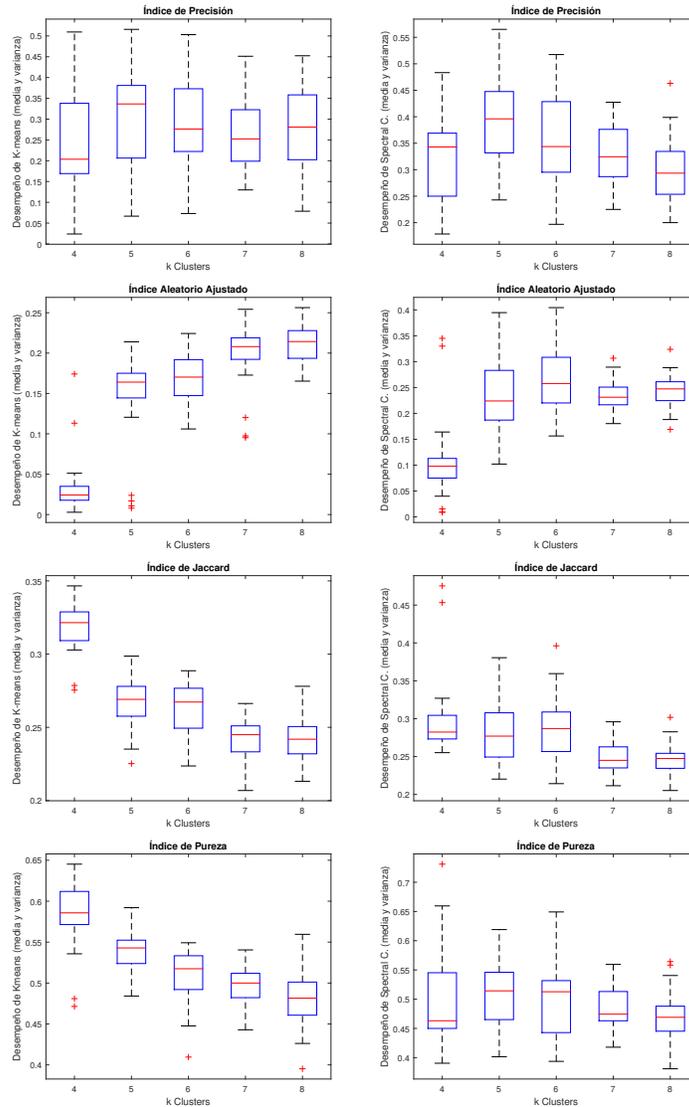


Figura 11. Desempeño de los clusters con los 3 sensores IMU más relevantes (*PCA*).

Utilizando la técnica *ReliefF* se logro disminuir el número de sensores IMU a tres como ocurre con *PCA*, pero se distingue que los resultados que se muestran en la figura 12, tienen varianzas de menor valor mejorando la eficiencia.

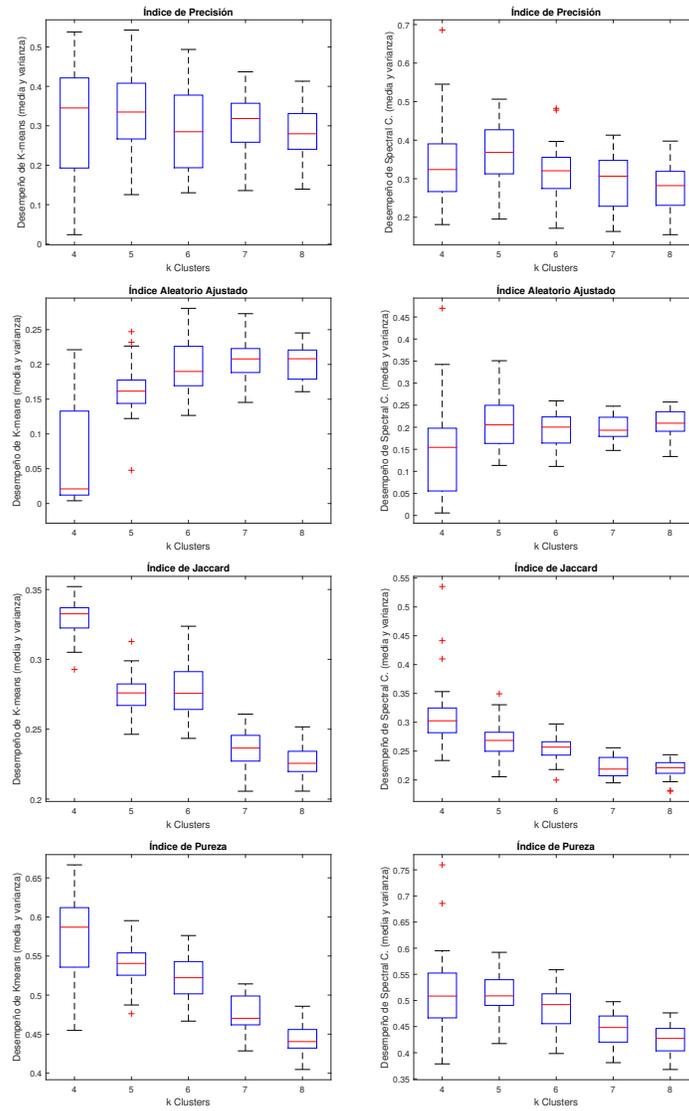


Figura 12. Desempeño de los clusters con los 3 sensores IMU más relevantes (*ReliefF*).

La figura 13 muestra que al utilizar los cuatro sensores EMG, el número de clusters con mejor desempeño varía entre 4 y 5 clusters, y al usar índice aleatorio ajustado se observa que la clasificación tiene un rendimiento muy pobre, ya que la eficiencia al 5 %.

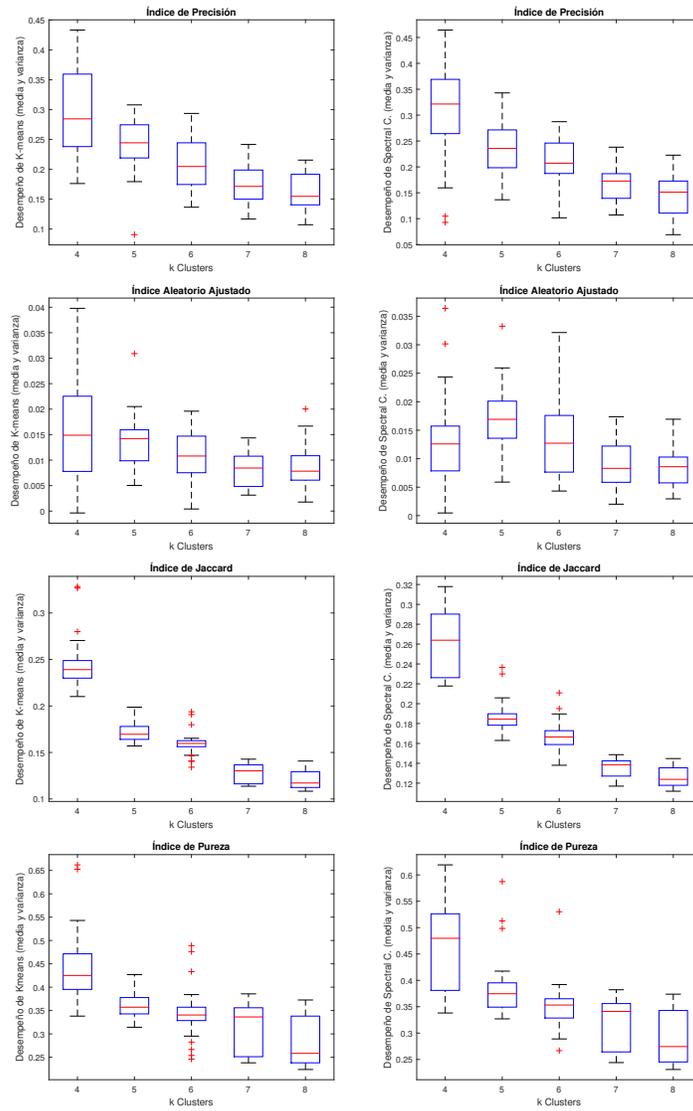


Figura 13. Desempeño de los clusters con todos los sensores EMG.

Al disminuir los sensores EMG utilizando *PCA*, se determina que con tres sensores EMG el resultado es similar a si se utilizaran todos los sensores, pero la diferencia es que la eficiencia con 4 clusters aumenta, logrando dar mejores resultados que con 5 clusters como se observa en la figura 14.

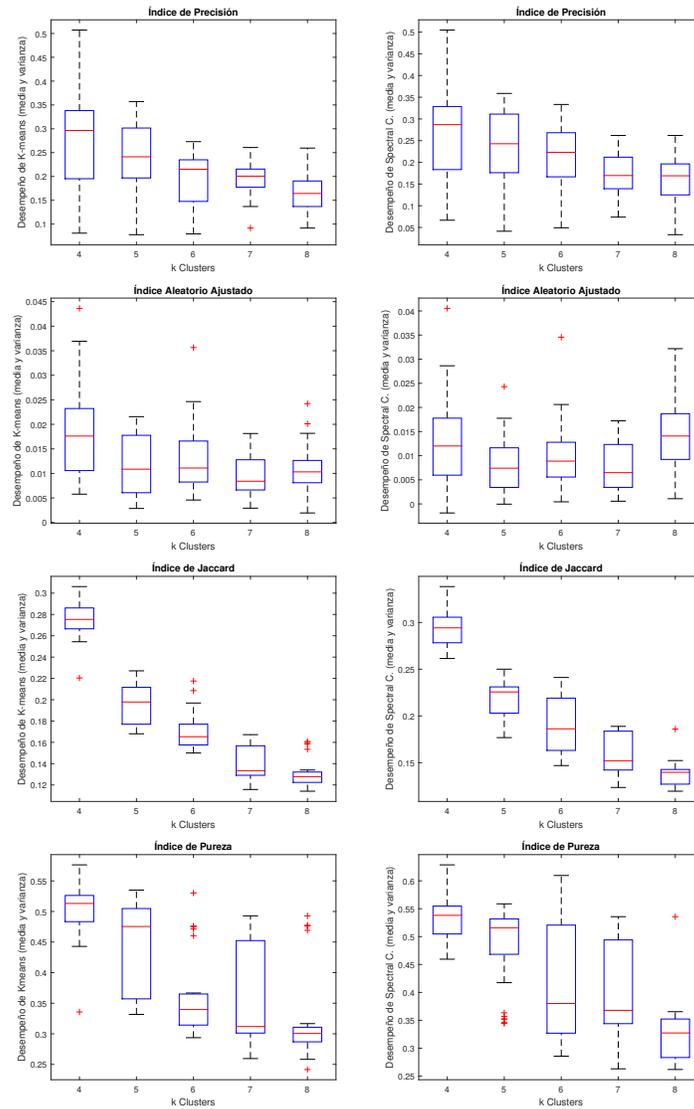


Figura 14. Desempeño de los clusters con los 3 sensores EMG más relevantes (*PCA*).

Para el caso con *ReliefF*, solamente es necesario dos sensores EMG para tener un desempeño similar si se utilizaran cuatro sensores o si se utilizaran los tres sensores elegidos por *PCA* (ver figura 15). La eficiencia al utilizar 4 *clusters* disminuye un poco en comparación con *PCA*, aun así, en la mayoría de los casos 4 *clusters* sigue dando los mejores resultados, en comparación a las demás agrupaciones, donde mejoraron su desempeño. Pero como ocurre con el caso del Kinect, utilizar los sensores EMG seleccionados por *ReliefF* es la mejor opción debido a que solamente son dos, mientras que con *PCA*, se necesita tres sensores y al usar un sensor de

más, el número de características es más grande, aumentando el tiempo de clasificación.

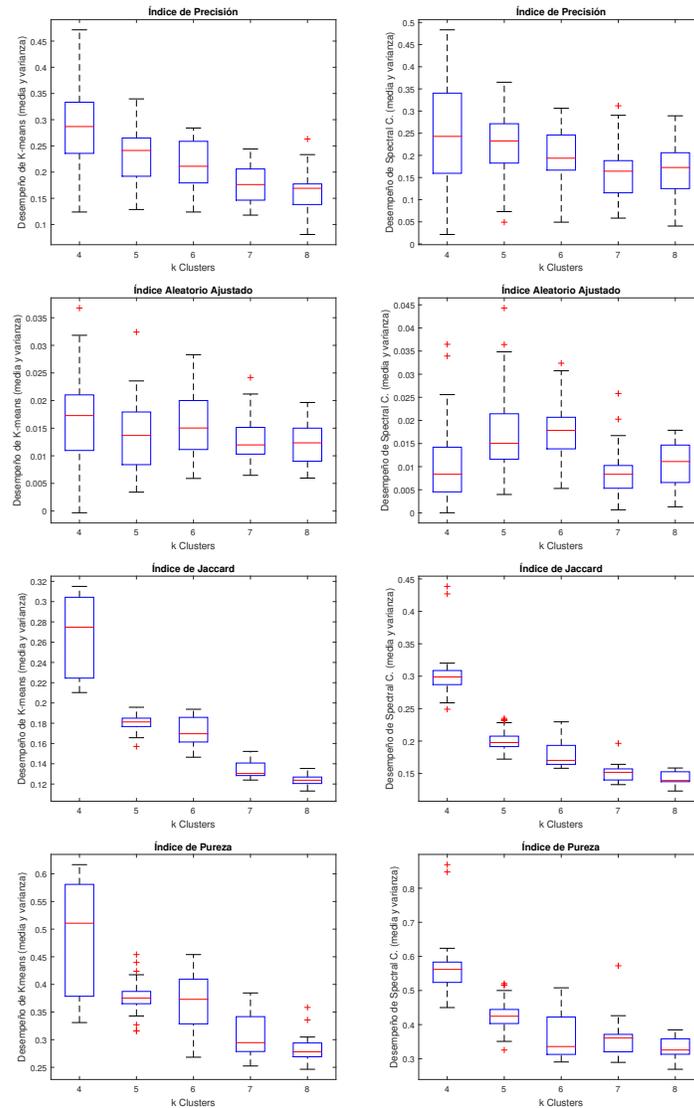


Figura 15. Desempeño de los clusters con los 2 sensores EMG más relevantes (*ReliefF*).

Al determinar los puntos articulados y sensores relevantes, para el caso del Kinect® (ver figuras 7, 8 y 9) se determina que el algoritmo *ReliefF* es la mejor opción debido a que con los 6 puntos articulados más relevantes, se obtiene un rendimiento similar si se usara todos los 14 puntos, mientras que para tener este mismo rendimiento con *PCA*, es necesario usar 8 puntos.

Para el caso con los sensores IMU, se decidió que 3 sensores para cada técnica de selección de características son suficientes (ver figura 11 y 12), ya que con solo estos, se puede tener

un desempeño similar si todos los sensores se usaran, como se muestra en la figura 10. La diferencia entre ambas técnicas es muy pequeña, pero es más recomendable utilizar los tres sensores elegidos por la técnica *ReliefF*, ya que son un poco más precisos los resultados.

En el caso de los sensores EMG, en las figuras 13, 14 y 15, la técnica *ReliefF* ofrece el mejor desempeño en general, ya que son necesarios 2 sensores relevantes para tener un desempeño similar si todos los sensores fueran utilizados, mientras que para *PCA*, son necesarios 3 sensores.

Al observar los resultados, la razón por la que la técnica *ReliefF* ofrece un mejor desempeño en general que la técnica *PCA*, es debido a que el punto fuerte de *PCA* es la reducción de dimensiones, pero no ofrece ese mismo desempeño en la selección de características y es débil contra ruido y outliers [36], algo donde *ReliefF* es fuerte en ambas áreas [15]. También se puede destacar que desde la figura 7 hasta la figura 15, que 4 y 5 clusters, son los clusters con mejor desempeño en general en todas las pruebas para el reconocimiento de actividad y demuestra que no son requeridos tantos movimientos primitivos para clasificar las 5 actividades.

7.2. Análisis de movimientos primitivos

Gracias a los datos obtenidos por la prueba anterior, se determinó que el número de clusters con mejor desempeño es con 4 y 5 clusters, por lo tanto, se determinará por medio de matrices de confusión, la eficiencia (media y varianza) de cada agrupamiento al utilizar *Spectral Clustering* y *K-means Clustering*. La eficiencia de la técnica de agrupamiento se determina por medio de la diagonal principal de la matrix de confusión. Al usar *Spectral Clustering*, se realizarán una serie de pruebas con esta técnica de agrupamiento en la cual se cambiaba el número de dimensiones a utilizar como se menciona en 6.3.2. Así que el número de dimensiones equivale al número de agrupamientos + un valor de dimensión que va desde 1 hasta 5. Las tablas de todas las pruebas realizadas se encuentran en la Sección 9 (Anéxos).

Al realizar las pruebas con *Spectral Clustering* para 4 clusters, se determina que el sensor Kinect es el que ofrece el mejor desempeño para el reconocimiento de movimientos primitivos cuando se usa un solo tipo de sensor como se observa en la tabla 3.

	1	2	3	4
1	86.66±16.98	7.88±12.37	0.06±0.11	8.40±5.90
2	30.29±17.07	60.70±25.24	0.27±1.49	8.74±12.97
3	15.47±6.61	0.80±1.24	80.24±3.96	3.48±3.84
4	33.37±6.89	4.17±3.19	0.03±0.17	62.43±5.06
Media y Varianza de la diagonal: 71.75±12.80				

Tabla 3. SC-Kinect (4 *clusters* y 6 dimensiones)

Para la combinación de 2 tipos de sensores, la que ofrece mejor resultado es la Kinect+IMU (Tabla 4), la cual mejora el desempeño de reconocimiento en comparación contra el Kinect®), ya que cada sensor disminuye las falencias del otro sensor al reconocer ciertos movimientos primitivos.

	1	2	3	4
1	76.20±26.07	15.95±25.70	0.22±0.23	7.63±6.31
2	17.08±18.24	81.99±19.50	0.29±1.57	0.64±0.55
3	20.38±9.39	3.17±7.02	75.91±4.70	0.55±2.99
4	22.76±7.38	9.31±9.50	0.01±0.06	67.92±4.20
Media y Varianza de la diagonal: 75.50±13.61				

Tabla 4. SC-Kinect+IMU (4 *clusters* y 9 dimensiones)

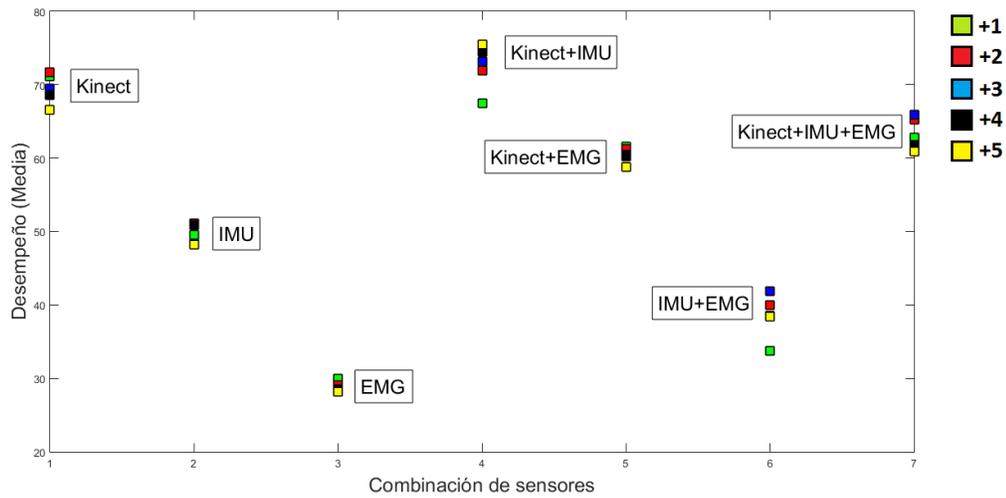
Con la combinación de los 3 tipos de sensores (Tabla 5), el desempeño cae con respecto al uso de un solo tipo de sensor y a la combinación de dos sensores. Esto se debe a que los sensores EMG tienen un desempeño muy bajo y esto afecta al reconocimiento de movimientos primitivos. Pero se resalta que la desviación estandar disminuye, aun a pesar de la disminución de la media.

	1	2	3	4
1	22.89±9.43	68.66±14.06	0.30±0.12	8.15±5.51
2	18.08±5.69	79.61±15.53	0.00±0.00	2.30±10.92
3	5.76±1.94	8.82±2.46	84.18±2.46	1.24±1.24
4	12.33±4.02	23.17±5.46	0.00±0.00	64.50±3.81
Media y Varianza de la diagonal: 62.79±7.807				

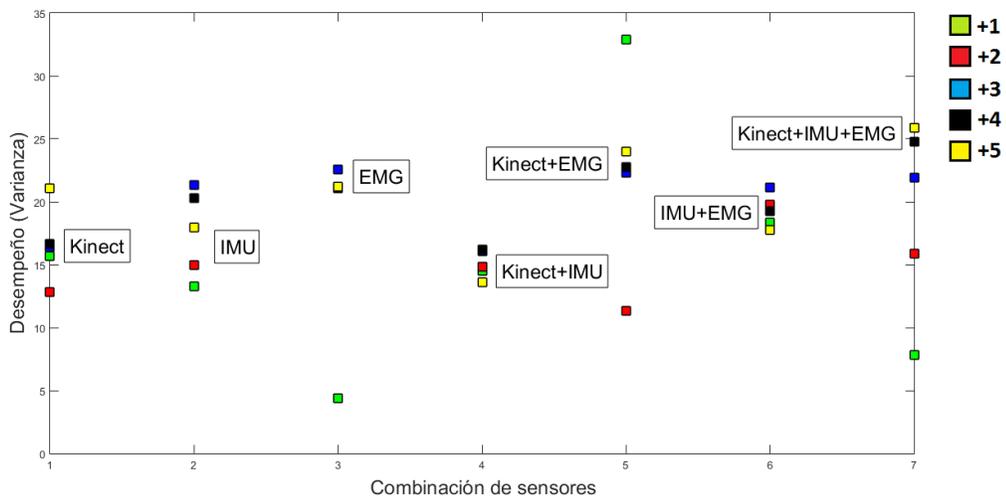
Tabla 5. SC-Kinect+IMU+EMG (4 *clusters* y 5 dimensiones)

La figura 16 muestra la media y la varianza de todas las combinación de sensores para 4 clusters utilizando *Spectral Clustering*. Los cuadros de colores representan cada una de las

cinco pruebas en el que d varía, donde el cuadro verde es +1 dimensión, el cuadro rojo es +2, el cuadro azul es +3, el cuadro negro es +4 y el cuadro amarillo es +5. De esta figura se observa que la mejor combinación es claramente la Kinect+IMU, ya que tiene la media mas grande y su varainza esta entre la de menor valor. También se destaca que los sensores EMG ofrecer poco rendimiento y eso afecta a las demas combinaciones en la que los sensores estan fusionados.



Media de la diagonal de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

Figura 16. Desempeño de cada una de las combinaciones de sensores al variar las dimensiones utilizando *Spectral Clustering* con 4 clusters.

En el caso de 5 clusters utilizando *Spectral Clustering*, la eficiencia en todos los casos disminuye en comparacion con 4 clusters utilizando la misma tecnica de agrupamiento. Por ejemplo, al observar las tablas 6, 7 y 8, se determina que la etiqueta 3 es la que presenta problemas en el reconocimiento, ya que los sensores la confunde como si dicha etiqueta fuera la etiqueta 1.

	1	2	3	4	5
1	82.20±13.44	6.67± 8.29	1.15±6.30	0.04±0.10	9.94±9.08
2	26.11±5.98	67.17±6.13	0.66±3.60	0.00±0.00	6.07±3.52
3	57.29±23.44	12.92±13.53	1.93±10.56	0.42±0.70	27.45±21.00
4	15.12±6.66	0.94±1.16	0.15±0.83	80.73±4.24	3.06±4.30
5	34.08±7.52	4.50±3.75	0.49±2.67	0.01±0.06	60.92±6.40
Media y Varianza de la diagonal: 58.59±8.154					

Tabla 6. SC-Kinect (5 clusters y 6 dimensiones)

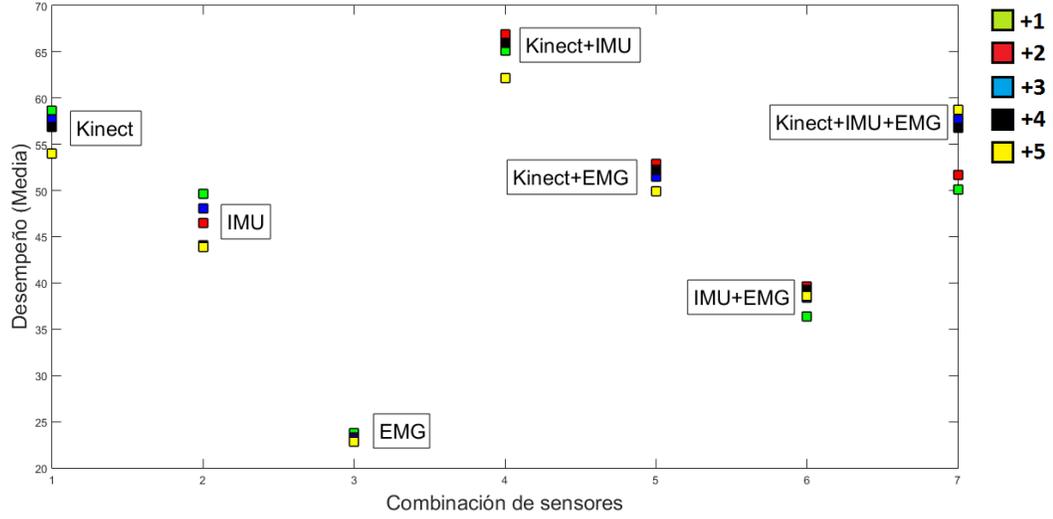
	1	2	3	4	5
1	66.51±25.57	23.91±26.10	2.67±9.99	0.22±0.13	6.69±6.48
2	35.45±18.88	61.69±17.70	1.73±6.88	0.00±0.00	1.13±1.56
3	15.21±11.71	15.63±22.17	57.81±26.75	0.00±0.00	11.35±11.29
4	11.89±4.32	3.91±5.04	0.27±1.49	82.45±2.52	1.47±2.86
5	27.06±3.57	9.61±9.76	6.21±5.53	0.00±0.00	57.11±7.74
Media y Varianza de la diagonal: 65.09±16.05					

Tabla 7. SC-Kinect+IMU (5 clusters y 6 dimensiones)

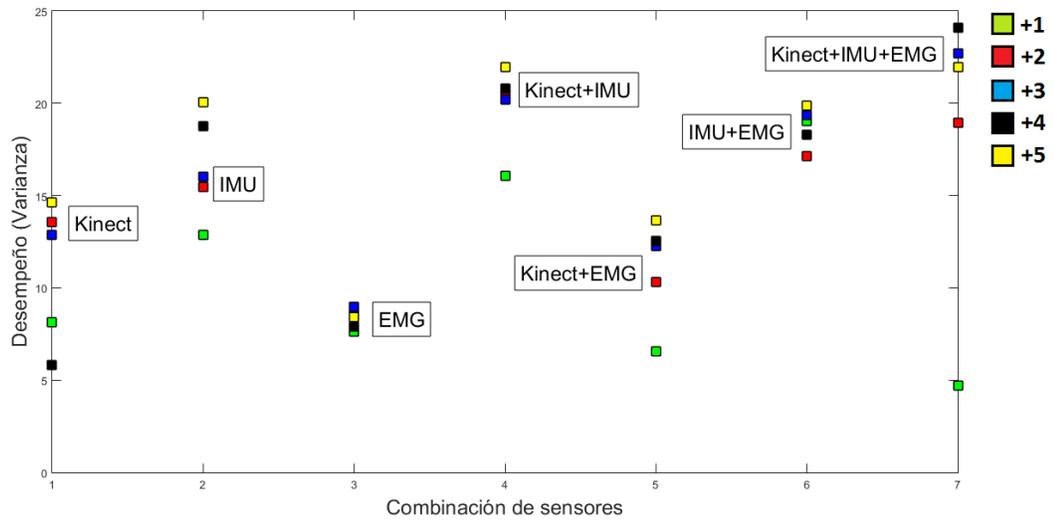
	1	2	3	4	5
1	66.52±35.77	19.04±29.24	0.62±1.23	0.17±0.15	13.65±24.67
2	31.75±21.18	64.18±20.98	0.46±2.43	0.01±0.08	3.61±7.55
3	30.73±26.33	23.33±27.31	30.47±33.54	0.99±4.01	14.48±13.21
4	16.89±9.64	3.29±6.74	0.03±0.17	77.41±4.02	2.38±5.63
5	29.60±18.18	11.30±10.59	3.79±4.40	0.00±0.00	55.30±15.55
Media y Varianza de la diagonal: 56.79±24.11					

Tabla 8. SC-Kinect+IMU+EMG (5 clusters y 10 dimensiones)

La figura 17 ofrece una vista general de cómo se comporta cada combinación de sensores, donde el máximo valor de la media de la diagonal de cada tabla de confusión para 5 clusters proviene de la fusión Kinect+IMU, y el menor valor de varianza proviene de la fusión Kinect+IMU+EMG.



Media de la diagonal de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

Figura 17. Desempeño de cada una de las combinaciones de sensores al variar las dimensiones utilizando *Spectral Clustering* con 5 *clusters*.

Cuando se tiene 4 *clusters*, pero esta vez utilizando *K-means Clustering*, el desempeño de dicho clasificador disminuye en todas las combinaciones de sensores en comparación (ver tablas 9, 10 y 11) con *Spectral Clustering*, esto se debe a que a *K-means* hace uso de la distancia euclidiana para determinar que muestras pertenecen a cierto agrupamiento, y es sabido que entre más características tenga una muestra, la dimensionalidad del problema crece, y la clasificación

por *K-means* disminuye. Este problema de dimensionalidad afecta en menor medida a *Spectral Clustering*, el cual ofrece mejor resultados.

	1	2	3	4
1	13.01±33.74	72.14±41.71	0.02±0.04	14.83±30.98
2	8.60±22.15	86.49±21.86	0.00±0.00	4.91±7.18
3	2.70±7.01	13.21±8.08	81.76±2.33	2.33±6.08
4	10.70±28.27	45.23±29.96	0.00±0.00	44.07±26.87
Media y Varianza de la diagonal: 56.33±21.20				

Tabla 9. KmC-Kinect (4 clusters)

	1	2	3	4
1	9.85±28.59	85.31±28.76	0.26±0.11	4.58±1.15
2	6.81±20.55	93.03±20.50	0.00±0.00	0.16±0.31
3	1.77±5.46	12.68±4.79	85.55±2.50	0.00±0.00
4	5.73±15.38	46.42±15.80	0.00±0.00	47.85±4.46
Media y Varianza de la diagonal: 59.07±14.01				

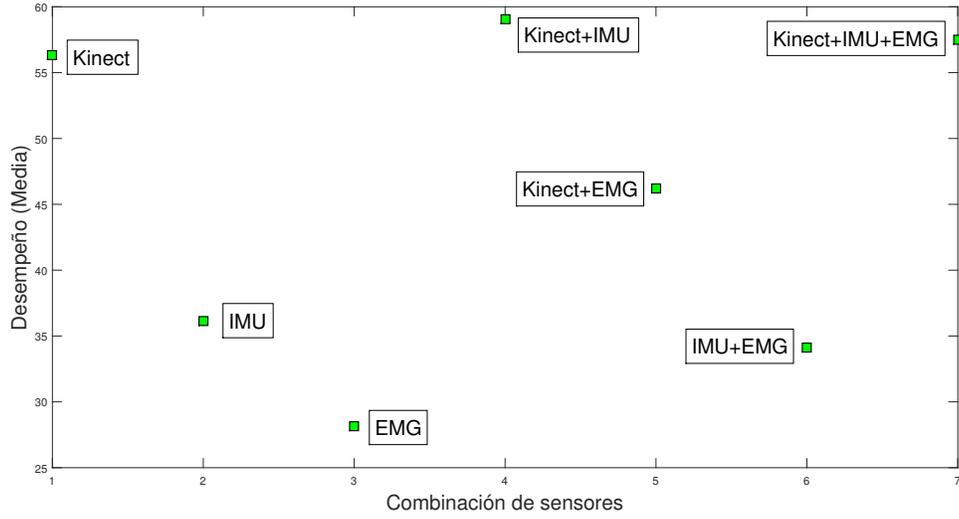
Tabla 10. KmC-Kinect+IMU (4 clusters)

	1	2	3	4
1	10.20±23.12	83.87±22.85	0.24±0.12	5.69±2.13
2	9.48±16.95	89.47±16.60	0.07±0.16	0.97±2.15
3	2.65±4.69	13.33±4.11	83.77±2.77	0.24±0.79
4	8.13±11.96	45.18±14.57	0.00±0.00	46.69±7.78
Media y Varianza de la diagonal: 57.53±12.56				

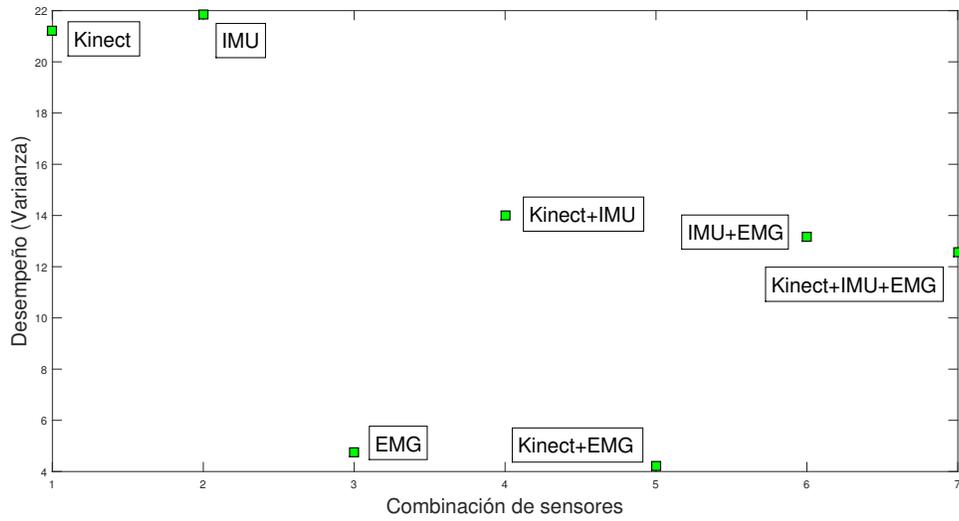
Tabla 11. KmC-Kinect+IMU+EMG (4 clusters)

La figura 18 muestra la media y varianza de cada combinación de sensores para 4 clusters y utilizando *K-means Clustering*, en donde se resalta que la media más alta la ofrece nuevamente Kinect+IMU, pero esta vez la varianza más baja proviene de Kinect+EMG. También existen cambios más sutiles, ya que en este caso la combinación Kinect+IMU+EMG tiene una mejor eficiencia comparándola con el sensor Kinect®), algo que en *Spectral Clustering* era lo contrario.

Con 5 clusters, se tiene el caso en el que disminuye la eficiencia en el reconocimiento de movimientos primitivos, ya que en comparación contra *Spectral Clustering*, *K-means Clustering*



Media de la diagonal de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

Figura 18. Desempeño de cada una de las combinaciones de sensores utilizando *K-means Clustering* con 4 *clusters*.

tiene problemas en reconocer más de un movimiento primitivo como se muestra en las tablas 12, 13 y 14, y el mejor desempeño el cual proviene del Kinect+IMU, no supera el 50% en la media. La combinación que ofrece la mejor medida de varianza que es la EMG, tiene una media inferior al 25%. Lo anterior se muestra en la figura 19.

	1	2	3	4	5
1	80.92±36.77	16.75±36.21	0.00±0.00	0.02±0.07	2.31±1.10
2	50.49±24.04	46.88±24.24	0.00±0.00	0.03±0.11	2.60±3.26
3	75.52±34.04	17.81±32.81	0.00±0.00	0.05±0.29	6.61±4.56
4	15.89±7.45	2.29±5.22	0.00±0.00	81.80±2.96	0.02±0.08
5	54.53±30.61	9.20±17.44	0.00±0.00	0.00±0.00	36.27±19.96
Media y Varianza de la diagonal: 49.04±16.78					

Tabla 12. KmC-Kinect (5 clusters)

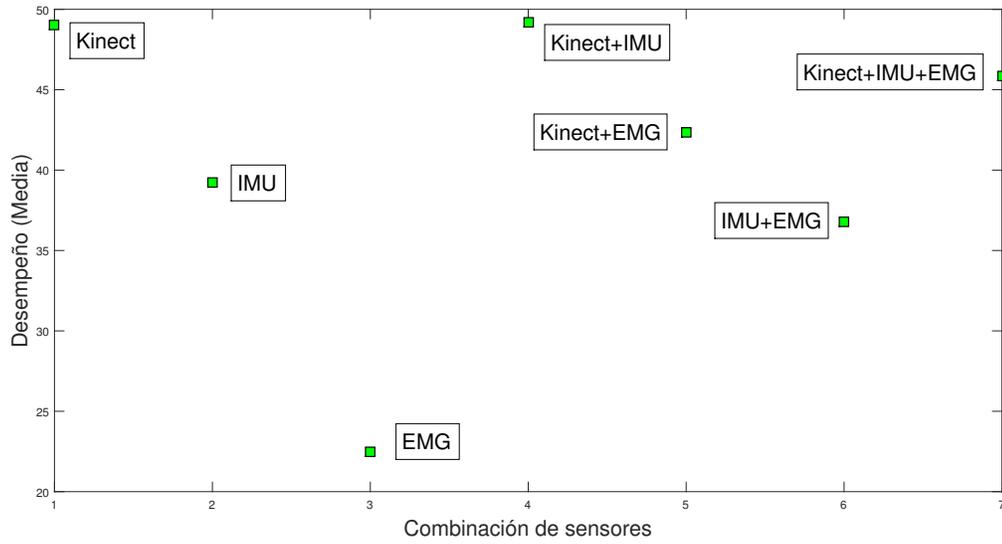
	1	2	3	4	5
1	24.05±39.94	72.16±40.18	0.08±0.44	0.21±0.13	3.50±1.95
2	15.59±28.64	82.79±27.96	0.00±0.00	0.03±0.11	1.59±3.16
3	22.08±34.34	54.38±35.90	2.29±12.55	0.05±0.29	21.20±19.99
4	3.58±6.67	11.00±6.38	0.00±0.00	85.42±1.95	0.00±0.00
5	15.38±20.44	32.13±18.92	1.18±6.46	0.00±0.00	51.31±11.36
Media y Varianza de la diagonal: 49.17±18.75					

Tabla 13. KmC-Kinect+IMU (5 clusters)

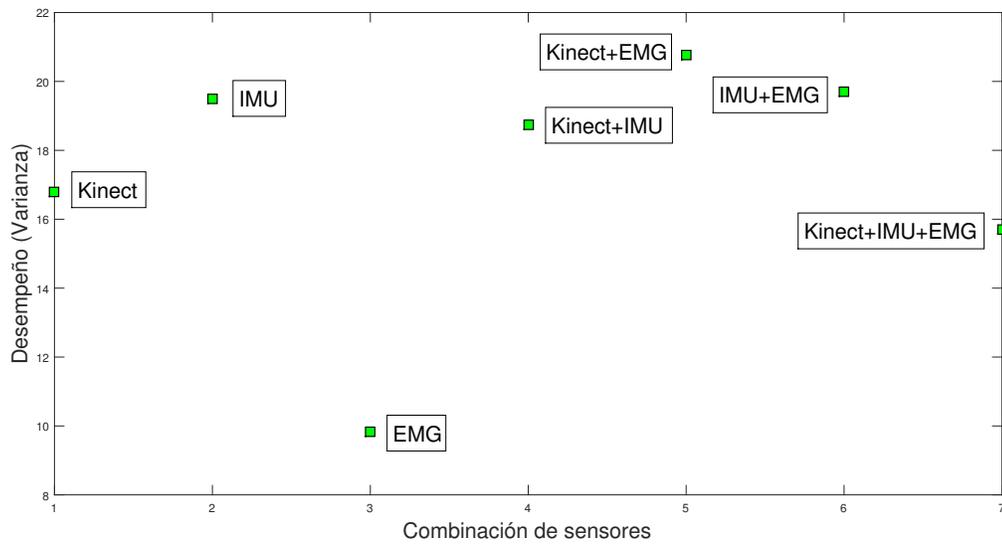
	1	2	3	4	5
1	8.30±16.68	85.94±16.25	0.00±0.00	0.23±0.11	5.53±3.05
2	8.35±12.04	89.63±11.63	0.00±0.00	0.06±0.15	1.96±2.59
3	8.80±15.82	73.18±17.68	0.00±0.00	0.05±0.29	17.97±11.00
4	2.05±2.85	13.97±3.38	0.00±0.00	83.70±2.57	0.29±0.58
5	8.12±8.56	44.24±10.29	0.00±0.00	0.00±0.00	47.64±6.58
Media y Varianza de la diagonal: 45.85±15.70					

Tabla 14. KmC-Kinect+IMU+EMG (5 clusters)

Al realizar una comparación de los anteriores resultados con 8 clusters usando ambas técnicas de agrupamiento, se demuestra que el rendimiento de dichas técnicas para reconocer movimientos primitivos decae. Esto se observa en la diagonal principal de cada matriz de confusión, la cual existen tres etiquetas que no se reconocen adecuadamente porque su porcentaje de acierto es menor del 50% y en ciertos casos dicho porcentaje es de 0% para *Spectral Clustering*, como se muestra en las tablas 15, 16 y 17.



Media de la diagonal de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

Figura 19. Desempeño de cada una de las combinaciones de sensores utilizando *K-means Clustering* con 5 clusters.

	1	2	3	4	5	6	7	8
1	87.92±11.13	5.03±5.82	0.00±0.01	0.00±0.00	2.91±9.80	0.03±0.04	1.94±0.61	2.17±0.83
2	32.56±8.92	65.00±8.83	0.00±0.00	0.00±0.00	0.44±1.09	0.00±0.00	0.95±0.66	1.05±0.93
3	26.56±7.76	62.82±7.79	0.10±0.39	0.00±0.00	0.10±0.56	0.00±0.00	0.82±2.24	9.59±4.99
4	68.07±20.34	15.47±16.44	0.00±0.00	0.00±0.00	3.75±14.28	0.63±0.78	7.19±3.57	4.90±3.52
5	28.84±5.99	0.89±2.11	0.00±0.00	0.00±0.00	68.22±4.49	1.73±4.17	0.00±0.00	0.31±0.67
6	0.32±0.40	3.09±3.03	0.04±0.14	0.00±0.00	14.26±4.21	85.42±4.06	0.00±0.00	0.00±0.00
7	26.52±6.79	3.09±3.03	0.04±0.14	0.00±0.00	1.19±4.07	0.00±0.00	51.59±3.63	17.57±6.76
8	33.45±7.32	4.94±4.79	0.00±0.00	0.00±0.00	1.38±5.19	0.00±0.00	25.08±8.85	35.15±12.89
Media y Varianza de la diagonal: 47.83±5.677								

Tabla 15. SC-Kinect (8 *clusters* y 9 dimensiones)

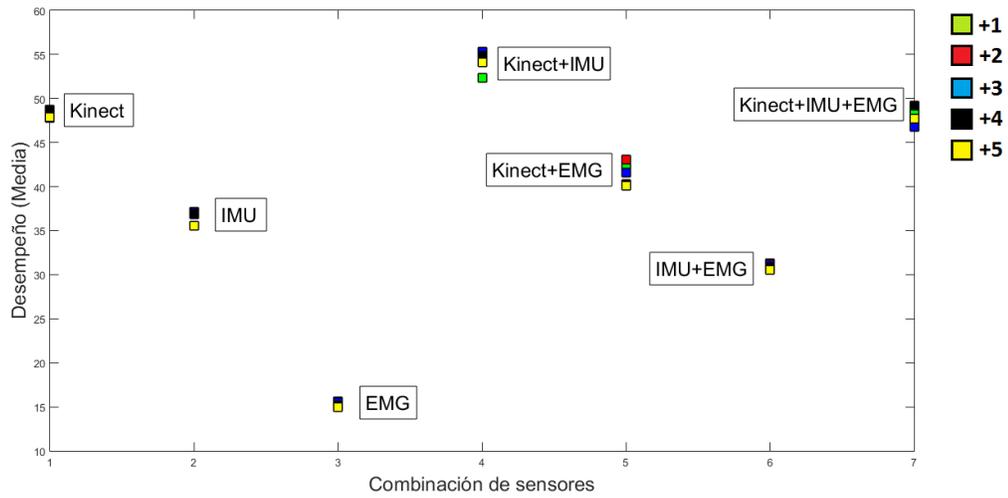
	1	2	3	4	5	6	7	8
1	84.58±14.22	5.44±9.51	0.01±0.02	0.91±0.38	3.10±9.87	0.22±0.12	3.94±5.41	1.80±2.50
2	23.23±6.46	74.07±8.58	0.00±0.00	0.00±0.00	0.67±2.16	0.00±0.00	0.67±2.49	1.35±3.35
3	19.95±9.27	74.10±11.52	0.05±0.28	0.00±0.00	0.00±0.00	0.00±0.00	1.69±2.26	4.21±8.83
4	24.74±9.05	6.15±15.17	0.00±0.00	61.51±21.34	1.15±4.36	0.00±0.00	4.58±4.00	1.88±2.70
5	23.69±9.80	0.04±0.24	0.00±0.00	0.00±0.00	72.53±15.24	3.20±7.45	0.53±2.92	0.00±0.00
6	0.12±0.26	0.00±0.00	0.00±0.00	0.00±0.00	13.08±5.02	86.78±4.98	0.02±0.13	0.00±0.00
7	26.37±9.97	5.96±4.43	0.04±0.14	7.78±3.66	0.48±1.55	0.00±0.00	45.22±19.82	14.15±18.42
8	23.72±8.23	7.71±5.90	0.00±0.00	7.35±3.63	0.75±2.71	0.00±0.00	42.34±21.03	18.14±21.29
Media y Varianza de la diagonal: 55.36±13.21								

Tabla 16. SC-Kinect+IMU (8 *clusters* y 11 dimensiones)

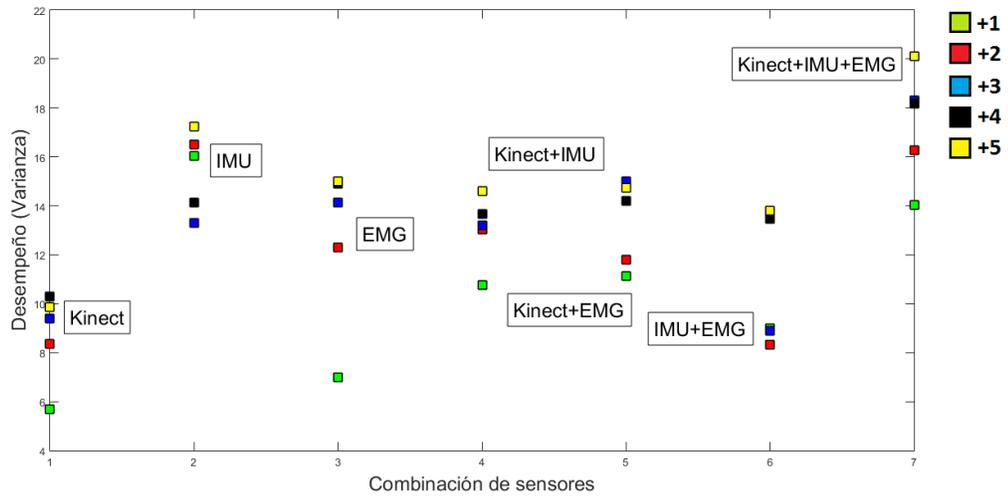
	1	2	3	4	5	6	7	8
1	57.47±20.99	25.11±18.30	0.00±0.00	2.73±4.72	10.89±12.60	0.12±0.08	0.57±1.63	3.10±1.24
2	27.54±13.57	68.13±13.72	0.00±0.00	2.54±5.51	1.19±2.84	0.00±0.00	0.54±2.93	0.06±0.18
3	24.41±14.55	71.23±15.71	0.00±0.00	2.05±5.94	0.21±0.67	0.00±0.00	0.36±1.19	1.74±1.44
4	12.92±12.75	18.28±14.65	0.00±0.00	59.06±17.93	3.28±8.07	0.05±0.29	0.62±1.77	5.78±3.46
5	28.89±17.99	2.62±1.38	0.00±0.00	0.09±0.49	63.07±25.90	5.29±8.22	0.04±0.24	0.00±0.00
6	1.18±1.13	0.28±0.35	0.00±0.00	0.00±0.00	8.19±5.11	90.25±4.92	0.09±0.51	0.00±0.00
7	27.09±5.06	14.63±5.71	0.00±0.00	11.56±7.46	1.98±2.35	0.00±0.00	4.43±12.62	40.31±14.72
8	24.38±6.66	11.72±7.76	0.00±0.00	10.77±6.66	2.22±3.81	0.00±0.00	5.22±15.31	45.69±16.27
Media y Varianza de la diagonal: 48.51±14.04								

Tabla 17. SC-Kinect+IMU+EMG (8 *clusters* y 9 dimensiones)

La figura 20 sigue demostrando que entre menos cantidad de dimensiones como entradas se utilicen, el desempeño del *Spectral Clustering* será mejor. Solo es con observar que las medias para los tres primeros valores del número de dimensiones son altas y las varianzas son pequeñas.



Media de la diagonal de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

Figura 20. Desempeño de cada una de las combinaciones de sensores al variar las dimensiones utilizando *Spectral Clustering* con 8 *clusters*.

Utilizando *K-means Clustering*, el reconocimiento de etiquetas decae debido a que aumenta el número de estas, donde su acierto está por debajo del 50 %, como se muestra en las tablas 18,

19 y 20. Aunque hay más posibilidad de que el acierto de ciertas etiquetas sea superior al 90 %, en comparación con *Spectral Clustering*, el cual, con la combinación Kinect+IMU+EMG, tiene una etiqueta con un acierto superior a dicho valor de porcentaje.

	1	2	3	4	5	6	7	8
1	76.12±24.91	16.79±21.37	0.02±0.04	0.01±0.07	0.23±0.25	0.00±0.001	4.54±9.03	2.28±3.30
2	33.57±19.22	63.19±19.01	0.00±0.00	0.00±0.00	0.24±0.34	0.00±0.00	1.87±2.54	1.13±1.79
3	29.03±14.44	60.56±14.79	1.28±1.35	0.00±0.00	0.00±0.00	0.00±0.00	4.00±4.33	5.13±4.75
4	54.06±26.92	28.91±28.05	0.00±0.00	0.21±1.14	0.00±0.00	0.36±0.67	9.11±5.88	7.34±9.98
5	31.64±26.92	5.73±11.97	0.00±0.00	0.00±0.00	46.09±35.83	14.04±18.58	0.93±3.25	1.56±7.04
6	0.14±0.28	0.00±0.00	0.00±0.00	0.00±0.00	9.79±7.85	90.07±7.81	0.00±0.00	0.00±0.00
7	31.85±9.42	7.80±7.81	0.15±0.25	0.00±0.00	0.13±0.32	0.00±0.00	39.61±22.59	20.46±19.57
8	37.39±10.68	9.61±8.84	0.11±0.26	0.00±0.00	0.00±0.00	0.00±0.00	27.98±20.01	24.90±21.19
Media y Varianza de la diagonal: 42.68±16.72								

Tabla 18. KmC-Kinect (8 clusters)

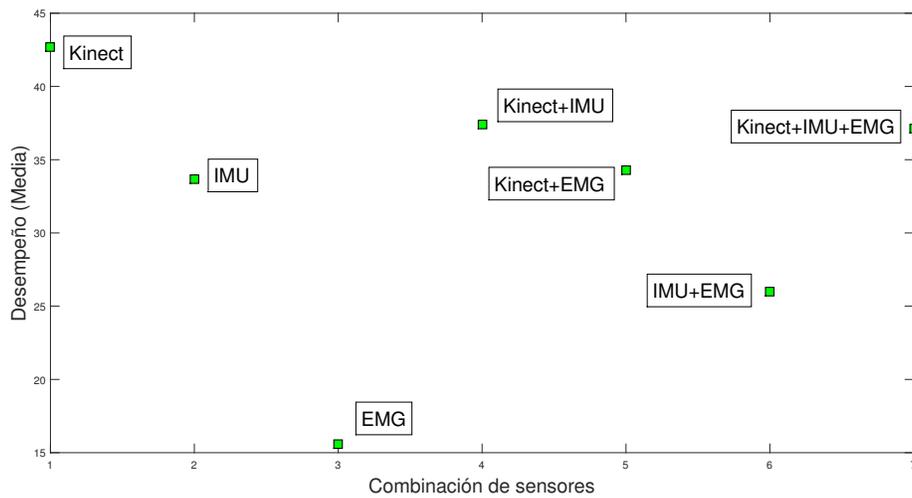
	1	2	3	4	5	6	7	8
1	17.37±26.75	77.20±26.23	0.03±0.05	0.24±0.67	0.25±0.56	0.11±0.11	4.24±5.15	0.56±0.98
2	2.92±6.00	95.52±5.88	0.00±0.00	0.00±0.00	0.06±0.18	0.00±0.00	1.41±1.78	0.10±0.44
3	3.90±5.99	93.38±6.09	1.49±1.31	0.10±0.39	0.00±0.00	0.00±0.00	0.87±1.04	0.26±0.58
4	7.40±7.36	27.60±10.13	0.00±0.00	8.80±22.85	0.00±0.00	0.00±0.00	49.48±27.32	6.72±18.31
5	9.02±18.10	37.42±19.96	0.00±0.00	0.04±0.24	14.04±31.97	37.82±17.33	1.56±5.44	0.09±0.34
6	0.02±0.13	0.00±0.00	0.00±0.00	0.00±0.00	2.41±5.51	97.57±5.50	0.00±0.00	0.00±0.00
7	15.72±9.66	19.06±6.72	0.11±0.23	2.35±8.13	0.06±0.17	0.00±0.00	50.06±17.76	12.65±14.41
8	18.57±9.53	21.16±6.97	0.11±0.26	2.22±7.51	0.05±0.17	0.00±0.00	43.67±15.65	14.22±15.71
Media y Varianza de la diagonal: 37.38±15.96								

Tabla 19. KmC-Kinect+IMU (8 clusters)

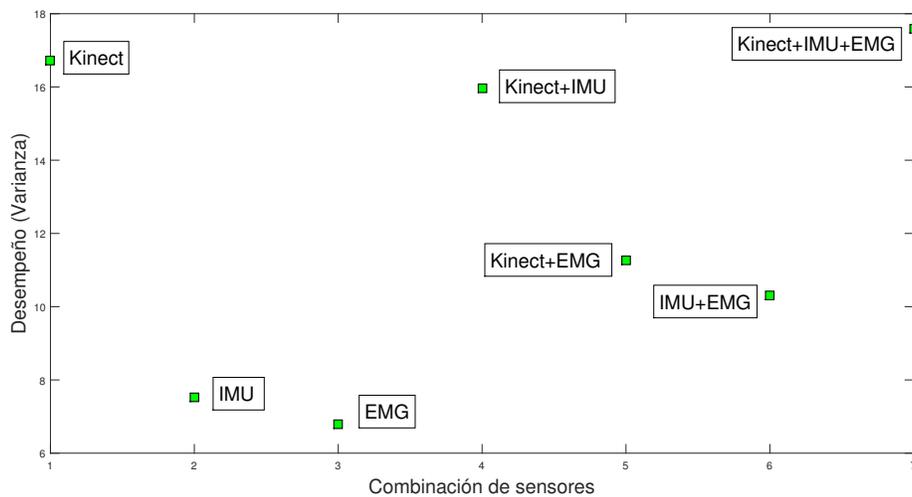
	1	2	3	4	5	6	7	8
1	65.29±30.57	27.62±29.86	0.03±0.05	0.15±0.43	0.24±0.82	0.12±0.10	4.03±2.63	2.53±1.51
2	44.15±25.44	51.31±24.14	0.00±0.00	0.06±0.18	0.00±0.00	0.00±0.00	3.87±3.98	0.62±2.06
3	39.90±26.24	56.05±25.26	0.92±1.18	0.00±0.00	0.00±0.00	0.00±0.00	2.21±3.17	0.92±3.09
4	27.71±16.93	19.32±18.73	0.00±0.00	8.28±21.80	0.05±0.29	0.05±0.29	36.82±25.79	7.76±12.16
5	41.64±22.01	13.42±16.98	0.00±0.00	0.00±0.00	8.36±22.99	35.78±12.41	0.76±1.43	0.04±0.24
6	0.21±0.37	0.39±0.39	0.00±0.00	0.00±0.00	2.04±5.50	97.04±5.86	0.30±0.99	0.02±0.13
7	21.81±9.24	14.31±8.56	0.11±0.23	2.85±9.21	0.06±0.22	0.00±0.00	38.63±17.91	22.22±16.74
8	22.56±9.95	12.36±9.75	0.07±0.21	2.40±8.39	0.00±0.00	0.00±0.00	35.37±17.54	27.23±16.90
Media y Varianza de la diagonal: 37.13±17.58								

Tabla 20. KmC-Kinect+IMU+EMG (8 clusters)

Utilizando 8 *clusters* con *K-means Clustering* entrega resultados donde el uso de un solo tipo de sensor (Kinect®o IMU), tiene un desempeño superior a la combinación Kinect+IMU, como se observa en la figura 21. Esto es debido que al mezclar una cantidad alta de clusters con el aumento de características, hace que ciertas combinaciones de sensores que en anteriores pruebas ofrecían los mejores desempeños, tengan peores resultados.



Media de la diagonal de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

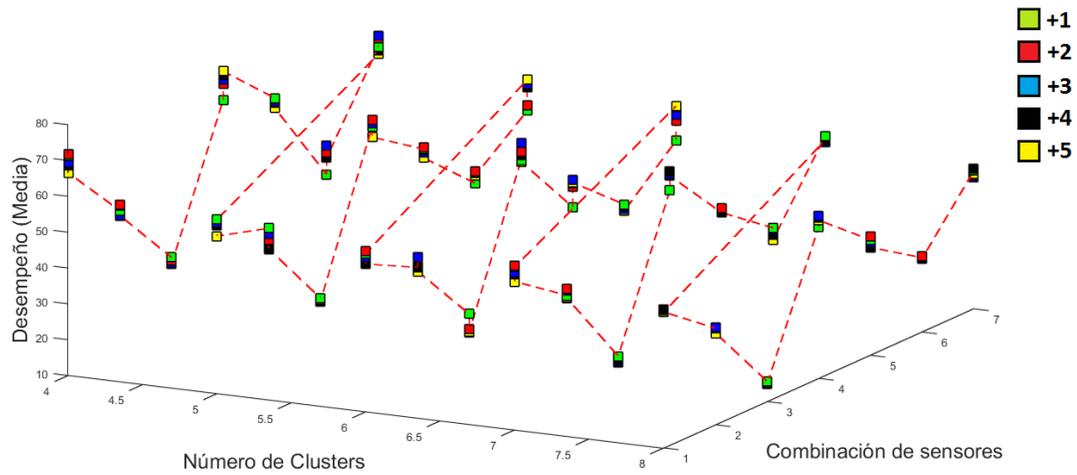
Figura 21. Desempeño de cada una de las combinaciones de sensores utilizando *K-means Clustering* con 8 *clusters*.

Al observar los resultados entregados por *Spectral Clustering* y *K-means Clustering* para 4, 5, 6, 7 y 8 *clusters* con cada combinación de sensores (ver figuras 22 y 23), se determina que *Spectral Clustering* ofrece mejores desempeño, además de que esta técnica de agrupamiento funciona con d pequeño, ya que para estos casos, se demuestra que entre las tres primeras pruebas con esta variable (donde se suma +1, +2 y +3 dimensiones) se tiene las mejores medias y varianzas del todo el conjunto de pruebas en el cual d cambia. También se destaca que al aumentar el número de clusters para el reconocimiento de movimientos primitivos disminuye la eficiencia de ambas técnicas de agrupamiento.

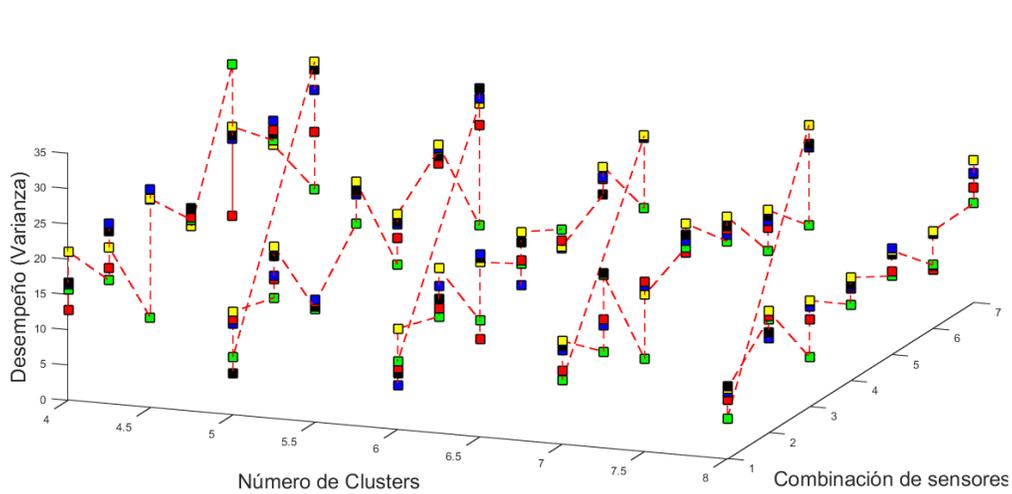
Observando el caso del Kinect®¹, se muestra que existe un problema al reconocer la etiqueta 3, debido a que esta la confunde como si fuera la etiqueta 1. Esto tal vez se deba a que [4], normaliza el cuerpo de la persona teniendo de referencia el punto articulado que se encuentra en el pecho (punto articulado número 7), y este proceso hace que los movimientos primitivos “reposo” y “suspendido en el aire” ofrezcan datos con características similares, provocando traslapes que dificultan el reconocimiento de ambos movimientos primitivos para estas técnicas de aprendizaje no supervisado. Algo similar puede ocurrir con los sensores IMU, pero se determina que la combinación Kinect+IMU suprime significativamente estos problemas, debido a que si un sensor no reconoce un movimiento primitivo, el otro si lo va a poder reconocer.

Otro detalle a destacar es que los sensores EMG entregan mucha información errónea, como si fuera un ruido cuando estos están fusionado con los sensores IMU y Kinect®. Aunque las medidas de varianzas para las combinaciones que incluyen EMG son buenas, concluyendo que los resultados en estos casos tienen muy poco error aun cuando los resultados no son lo esperado.

Gracias a los resultados de esta sección y de la sección 7.1, se permite identificar que solo con 4 y 5 *clusters* sea suficiente para reconocer las cinco actividades para la siguiente sección, debido a que se conoce que con 6, 7 y en especial con 8 *clusters* no va haber un buen nivel de desempeño en comparación con los valores de clusters anteriormente nombrados.

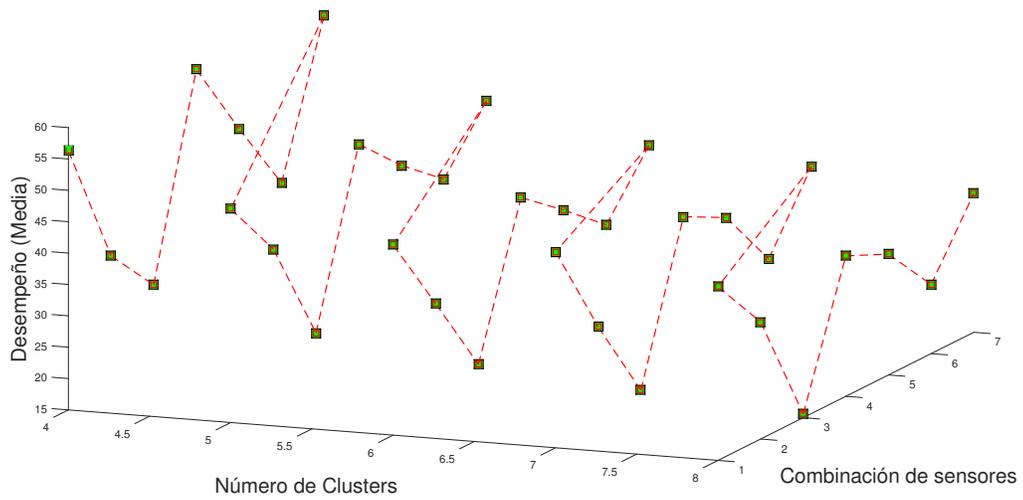


Media de cada una de las siete combinaciones

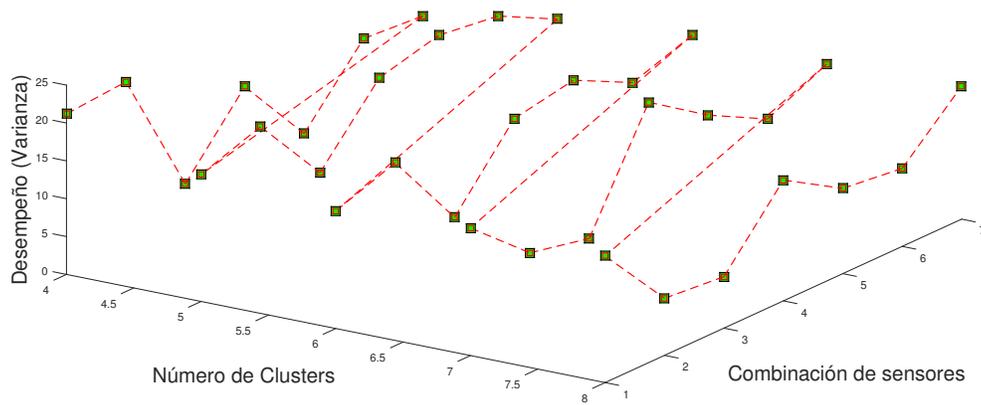


Varianza de la diagonal de cada una de las siete combinaciones

Figura 22. Media y varianza de cada una de las combinaciones de sensores al variar las dimensiones utilizando *Spectral Clustering* con 4, 5, 6, 7 y 8 clusters.



Media de cada una de las siete combinaciones



Varianza de la diagonal de cada una de las siete combinaciones

Figura 23. Media y varianza de cada una de las combinaciones de sensores utilizando *K-means Clustering* con 4, 5, 6, 7 y 8 clusters.

7.3. Análisis de reconocimiento de actividad física humana

Esta etapa presenta las matrices de confusión correspondientes al reconocimiento de actividades propuestas en la tabla 1. Estos indicadores de desempeño se calculan para las combinaciones de sensores con los mejores resultados obtenidos por *Spectral Clustering* y *K-means Clustering* para 4 y 5 *clusters*, debido a que en la sección 7.2, se identificó que Kinect, Kinect+IMU y Kinect+IMU+EMG, son las combinaciones con mejores desempeños en el reconocimiento de movimientos primitivos y que al utilizar 6, 7 y 8 *clusters*, había una pérdida en rendimiento con respecto a 4 y 5 *clusters*. La fusión de datos se realiza por medio etiquetas y características de cada uno de los sensores, esto es con el fin de realizar un análisis comparativo para determinar que técnica de fusión de datos es más determinística para la clasificación de actividades cuando existe una combinación de sensores.

El reconocimiento de actividades con los datos provenientes de *Spectral Clustering* con 4 *clusters* y usando solamente el sensor Kinect®[®], muestra que hay un conflicto para reconocer las actividades 2 y 4, como se aprecia en la tabla 21. Tal vez se deba a que la cantidad de movimientos primitivos para clasificar las cinco actividades sea muy poca, por lo tanto, gran parte de la actividad 2 sea etiquetada por el clasificador como si fuera la actividad 4, y la actividad 3 sea visualizada como si fuera la actividad 2. Aun así la varianza de la diagonal de la matriz de confusión es menos de 10 % y exista una actividad que tenga un acierto superior al 99 %.

	1	2	3	4	5
1	67.38±18.57	5.75±0.79	25.81±16.66	0.00±0.00	1.07±1.13
2	0.00±0.00	28.25±8.70	14.90±26.46	48.04±8.47	8.81±9.29
3	3.16±4.20	44.35±17.25	34.54±13.82	5.97±6.29	11.99±1.34
4	0.00±0.00	0.00±0.00	0.85±0.90	99.15±0.90	0.00±0.00
5	0.00±0.00	3.78±3.98	9.95±2.96	0.74±0.78	85.53±7.73
Media y Varianza de la diagonal: 62.97±9.94					

Tabla 21. HMM-Kinect (SC y cuatro movimientos primitivos)

Para la combinación Kinect+IMU aplicando fusión de datos por medio de etiquetas (tabla 22) y características (tabla 23), se tiene que la fusión por etiquetas logra un mejor rendimiento que la fusión por características, ya que esta última tiene una actividad demás, con un porcentaje

de acierto menor al 50 %. También se destaca que esta combinación tiene un desempeño menor que al usar solo el Kinect®), aun cuando la fusión por medio de etiquetas tiende a clasificar adecuadamente más movimientos primitivos.

	1	2	3	4	5
1	69.44±20.50	5.56±5.86	0.00±0.00	25.00±26.35	0.00±0.00
2	31.04±13.40	35.88±1.70	14.49±15.28	10.79±1.81	0.00±1.63
3	14.79±11.69	18.44±7.89	42.31±6.15	5.43±5.72	19.02±3.37
4	15.15±15.97	1.85±1.95	8.13±3.84	74.87±14.09	0.00±0.00
5	1.04±1.10	8.74±1.33	18.10±12.55	9.85±10.38	62.28±2.41
Media y Varianza de la diagonal: 56.96±8.97					

Tabla 22. HMM-Kinect+IMU (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	28.82±22.30	23.53±19.71	23.90±22.64	20.78±15.53	2.97±3.60
2	36.01±.98	15.89±12.67	2.32±4.49	45.78±26.56	0.00±0.00
3	14.67±12.55	8.29±10.07	73.45±14.34	2.62±4.44	0.97±1.50
4	20.50±0.09	49.56±30.73	0.33±1.01	29.61±15.67	0.00±0.00
5	0.00±0.00	2.73±5.38	0.00±0.00	1.21±1.03	96.06±26.52
Media y Varianza de la diagonal: 48.77±18.30					

Tabla 23. HMM-Kinect+IMU (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.

La combinación Kinect+IMU+EMG (ver tablas 24 y 25) tiene una insignificante disminución en el rendimiento en comparación con la combinación Kinect+IMU, ya que aunque la media general aumenta, también lo hizo la desviación estandar con ambas técnicas de fusión de datos. Por otra parte, el rendimiento es muy similar en ambas combinaciones de sensores (Kinect+IMU y Kinect+IMU+EMG), ya que se tiene los mismos movimientos primitivos que son difíciles de reconocer (etiquetas 2 y 4 para fusión por etiquetas, y etiquetas 1, 2 y 4 para fusión por características) . Así que los datos provenientes de *Spectral Clustering* con 4 clusters del reconocimiento de movimientos primitivos no son los suficientemente adecuados para el reconocimiento de las 5 actividades de este trabajo.

	1	2	3	4	5
1	70.29±4.96	6.21±2.24	0.79±0.84	19.40±11.66	3.31±5.30
2	16.36±5.35	39.48±11.09	17.76±11.20	6.37±6.71	20.04±1.47
3	10.77±2.92	19.68±4.56	33.59±1.67	13.90±10.43	22.07±10.47
4	5.90±6.96	7.08±7.46	16.82±8.63	66.96±17.88	3.24±9.76
5	0.00±0.00	3.70±3.90	15.46±16.30	2.78±2.93	78.06±23.13
Media y Varianza de la diagonal: 57.68±11.75					

Tabla 24. HMM-Kinect+IMU+EMG (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	31.58±21.59	18.94±22.04	25.46±20.66	24.02±11.18	0.00±0.00
2	25.68±24.00	20.22±18.50	4.23±6.98	48.86±32.29	1.01±2.61
3	18.18±20.79	11.04±13.91	69.01±11.50	0.42±0.88	1.35±5.80
4	24.56±14.38	47.81±30.07	1.30±5.91	26.33±23.15	0.00±0.00
5	0.00±0.00	1.99±3.39	0.00±0.00	0.37±2.02	97.64±17.71
Media y Varianza de la diagonal: 48.96±18.49					

Tabla 25. HMM-Kinect+IMU+EMG (SC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.

De los datos provenientes del reconocimiento de movimientos primitivos utilizando *Spectral Clustering* con 5 *clusters* utilizando el sensor Kinect® (tabla 26), el HMM reconoce adecuadamente 4 actividades, mejorando el rendimiento del clasificador con respecto a las anteriores pruebas con 4 clusters, donde este último solo reconoce adecuadamente 3 actividades. La actividad que aun ofrece dificultades en su reconocimiento es la tercera, ya que su acierto es menor del 50 %. Esto se debe a la normalización que se hace al cuerpo humano, donde el punto articulado de referencia es el 7, y por tal razón, el HMM confunde la actividad “Quedarse quieto” con “Saltar”. Este problema se origina en los movimientos primitivos, como se menciona al final de la etapa 7.2.

	1	2	3	4	5
1	75.24±0.26	3.76±3.96	21.00±4.22	0.00±0.00	0.00±0.00
2	0.00±0.00	64.58±2.20	27.08±28.55	0.00±0.00	8.33±26.35
3	2.25±0.20	31.72±2.56	28.81±9.80	28.73±13.43	8.50±6.39
4	0.00±0.00	0.00±0.00	1.67±1.76	98.33±1.76	0.00±0.00
5	0.00±0.00	3.96±5.41	7.91±10.82	0.00±0.00	88.13±16.24
Media y Varianza de la diagonal: 71.02±6.05					

Tabla 26. HMM-Kinect (SC y cinco movimientos primitivos)

Para la fusión Kinect+IMU, el reconocimiento de la tercera actividad mejora para ambos casos de fusión de datos, donde la fusión por etiquetas (tabla 27) realiza el mejor trabajo, ya que el porcentaje de acierto para todas las actividades es superior al 50 %. Aunque la fusión por medio de características (tabla 28) reconoce tres actividades con un acierto de mas del 90 %

	1	2	3	4	5
1	95.31±4.94	4.69±4.94	0.00±0.00	0.00±0.00	0.00±0.00
2	4.73±0.03	57.92±15.88	0.00±0.00	18.68±20.47	18.67±4.62
3	2.65±2.79	10.05±0.51	65.46±8.67	1.25±4.23	20.59±10.61
4	31.06±34.17	3.29±3.47	0.00±0.00	62.54±27.59	0.00±0.00
5	4.81±5.07	1.41±1.49	19.58±20.64	1.04±1.10	73.15±28.30
Media y Varianza de la diagonal: 70.876±17.08					

Tabla 27. HMM-Kinect+IMU (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	93.83±5.90	6.17±5.90	0.00±0.00	0.00±0.00	0.00±0.00
2	3.33±3.51	95.07±5.19	0.00±0.00	1.60±1.69	0.00±0.00
3	0.00±0.00	35.00±68.52	40.00±42.16	0.00±0.00	5.00±5.27
4	0.00±0.00	2.34±2.47	0.00±0.00	97.66±2.47	0.00±0.00
5	0.32±0.34	5.77±6.09	45.52±16.02	0.00±0.00	48.39±9.59
Media y Varianza de la diagonal: 74.99±13.06					

Tabla 28. HMM-Kinect+IMU (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.

De los datos provenientes de la combinación Kinect+IMU+EMG, el rendimiento del HMM baja con respecto a la anterior combinación de dos sensores, aunque la desviación estandar de la diagonal principal de ambas matrices de confusión (tabla 29 y 30) improvisa y se tiene que

la fusión por medio de características obtenga dicha desviación estandar menor a 10 %.

	1	2	3	4	5
1	76.86±0.42	9.39±2.50	0.00±0.00	13.75±2.09	0.00±0.00
2	18.62±14.08	49.30±7.59	3.62±12.83	12.35±1.92	16.11±10.76
3	6.32±2.92	18.11±9.50	42.21±13.01	15.72±16.57	17.65±10.14
4	4.08±4.30	1.79±11.29	1.79±11.29	88.52±27.43	3.83±9.14
5	1.59±10.04	5.63±5.93	24.38±9.44	4.22±4.45	64.18±9.10
Media y Varianza de la diagonal: 64.21±11.51					

Tabla 29. HMM-Kinect+IMU+EMG (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	90.79±12.07	0.00±0.00	4.35±4.44	3.14±4.81	1.72±2.90
2	0.00±0.00	94.84±5.120	3.03±6.34	0.00±0.00	2.13±1.69
3	3.39±5.01	1.80±2.74	38.65±13.51	6.86±7.95	49.30±28.03
4	3.83±5.16	0.00±0.00	4.08±4.76	85.32±4.780	6.77±8.12
5	1.99±1.87	3.36±3.09	49.89±30.38	4.68±2.62	40.08±14.07
Media y Varianza de la diagonal: 69.94±9.91					

Tabla 30. HMM-Kinect+IMU+EMG (SC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.

Al observar estos resultados se tiene que los datos provenientes al usar *Spectral Clustering* con 5 *clusters*, mejoran en gran medida y sigue hasta este momento la lógica al mostrar que se requieren como mínimo 5 movimientos primitivos para representar 5 actividades.

Al tener en cuenta la información que proviene de *K-means Clustering* usando 4 *clusters*, el HMM tiene un desempeño decente en detectar 4 actividades con una precisión superior al 60 % como se observa en la tabla 31, hasta se da el caso, de que la precisión en reconocer la actividad 4 es de un 100 %. Como anteriores pruebas, la actividad 3 aun presenta problemas para que el clasificador la reconozca.

Al observar las tablas 32 y 33, la combinación Kinect+IMU muestra un aumento en el desempeño al tener las 5 actividades con una tasa de precisión superior al 50 % al usar fusión de datos por medio de etiquetas. Mientras que la fusión de datos por medio de características tiene una disminución en el desempeño con respecto al usar el sensor Kinect® y la fusión del

	1	2	3	4	5
1	61.10±6.91	7.34±2.95	29.83±11.68	0.00±0.00	1.73±1.82
2	2.06±2.18	64.26±52.68	13.96±14.72	15.00±59.48	4.71±10.10
3	11.88±5.93	24.80±13.39	28.61±3.80	24.36±25.68	10.35±22.03
4	0.00±0.00	0.00±0.00	0.00±0.00	100±0.00	0.00±0.00
5	0.00±0.00	6.49±7.21	19.39±6.38	0.00±0.00	74.12±0.83
Media y Varianza de la diagonal: 65.62±12.84					

Tabla 31. HMM-Kinect (KmC y cuatro movimientos primitivos)

Kinect+IMU por medio de etiquetas, ya que para las actividades 1, 2 y 3, el HMM no las detecta adecuadamente y por lo tanto, la precisión al reconocer estas actividades es menor al 50 %.

	1	2	3	4	5
1	71.88±16.47	23.96±12.08	4.17±4.39	0.00±0.00	0.00±0.00
2	38.68±20.27	54.23±22.03	2.31±2.44	0.00±49.78	4.78±5.04
3	3.94±4.15	7.75±8.17	67.36±25.62	11.43±12.05	9.52±1.25
4	0±20.46	0.00±0.00	0.00±0.00	100±0.00	0.00±0.00
5	0.00±0.00	3.35±3.53	13.78±11.83	0.00±0.00	82.87±8.30
Media y Varianza de la diagonal: 75.27±14.48					

Tabla 32. HMM-Kinect+IMU (KmC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	29.86±31.47	28.44±29.98	1.70±1.79	0.00±0.00	0.00±0.00
2	12.03±8.88	17.82±31.52	17.37±13.51	32.78±5.81	0.00±0.00
3	0.00±0.00	3.41±3.43	75.69±18.59	5.03±5.30	15.86±16.72
4	0.00±0.00	2.51±2.64	7.77±0.81	36.73±38.71	0.00±0.00
5	0.00±0.00	1.48±1.56	4.41±3.46	0.00±0.00	94.11±1.90
Media y Varianza de la diagonal: 50.84±24.44					

Tabla 33. HMM-Kinect+IMU (KmC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.

Para la combinación Kinect+IMU+EMG, la fusión por medio de etiquetas hace que el HMM baje de rendimiento al no detectar con precisión las actividades 2 y 3, mientras que la fusión por carecterísticas hace que el HMM mantenga un rendimiento similar que en el caso con la combinación Kinect+IMU. Pero la desviación estandar para estas pruebas (tablas 34 y 35),

mejora considerablemente, así que en el caso de generar outliers es mas difícil que ocurra cuando se agrega los sensores EMG.

	1	2	3	4	5
1	87.94±12.72	10.40±10.96	1.67±1.76	0.00±0.00	0.00±0.00
2	22.95±17.98	46.69±10.21	6.62±6.97	15.34±4.92	5.42±5.71
3	6.34±1.66	22.62±8.79	39.66±6.67	14.83±4.45	16.55±12.67
4	21.06±7.14	9.53±5.01	2.00±2.11	64.20±7.62	3.21±3.38
5	3.34±3.53	5.23±0.34	26.88±4.91	0.00±0.00	64.54±8.10
Media y Varianza de la diagonal: 65.21±9.06					

Tabla 34. HMM-Kinect+IMU+EMG (KmC y cuatro movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	34.09±15.60	26.36±20.04	12.90±18.07	26.65±14.94	0.00±0.00
2	41.98±34.80	18.11±16.75	8.60±10.38	26.79±20.02	4.52±4.18
3	16.24±18.46	6.18±8.13	62.93±12.19	10.16±9.19	4.49±4.53
4	7.69±7.47	45.91±35.80	10.00±8.73	36.40±13.81	0.00±0.00
5	0.00±0.00	3.44±4.67	5.57±4.82	0.00±0.00	90.99±8.050
Media y Varianza de la diagonal: 48.50±13.28					

Tabla 35. HMM-Kinect+IMU+EMG (KmC y cuatro movimientos primitivos). Fusión de sensores por medio de sus características.

En sí, la información de *K-means Clustering* usando 4 *clusters*, permite una mejor clasificación por medio del HMM, comparando estas pruebas con las que surgieron al usar los datos de *Spectral Clustering* con 4 *clusters*. Tal vez, el motivo de esto es que el máximo número de características el cual posee la combinación Kinect+IMU+EMG no supera las trecientas, esto se debe gracias las técnicas que identifican las características relevantes y permiten disminuir la dimensionalidad del problema. Por tal razón es que *K-mean Clustering* se desempeña adecuadamente.

Por finalizar, se utilizan los datos provenientes de *K-means Clustering* con 5 *clusters* de la etapa de reconocimiento de movimientos primitivos (sección 7.2). Para las pruebas eficiencia realizadas para el sensor kinect®[®], que se muestran en la tabla 36, el HMM tiene un acierto superior al 50 % al reconocer 4 actividades de las 5 actividades propuestas en este trabajo, presentando una insignificante reducción en el rendimiento si se compara con su homólogo,

el cual utiliza *Spectral Clustering*. Pero, si se observa dicha tabla, existe una actividad que el HMM detecta con un 100% de precisión, la cual es la cuarta actividad. Esta actividad ha demostrado ser de las que el HMM mejor clasifica en las pruebas donde solo se utiliza el sensor Kinect® para el reconocimiento, debido a que el Kinect® es capaz de segmentar adecuadamente una acción en donde una extremidad del cuerpo esté extendida.

	1	2	3	4	5
1	58.49±8.95	11.22±5.74	30.28±3.21	0.00±0.00	0.00±0.00
2	0.00±0.00	81.95±5.30	13.53±1.95	0.69±0.73	3.82±4.08
3	1.97±2.07	25.35±20.52	27.99±13.90	24.96±4.69	19.73±4.00
4	0.00±0.00	0.00±0.00	0.00±0.00	100±0.00	0.00±0.00
5	0.00±0.00	4.21±8.74	12.92±0.44	1.01±1.06	81.86±7.23
Media y Varianza de la diagonal: 70.58±7.08					

Tabla 36. HMM-Kinect (KmC y cinco movimientos primitivos).

Cuando la información proviene de la combinación Kinect+IMU (ver tablas 37 y 38), el HMM pierde rendimiento para la fusión por medio de etiquetas, ya que la precisión para detectar la actividad 1 y 2 es menor que el 50%. Mientras que para la fusión por medio de características, el HMM detecta todas las actividades con una tasa superior al 50%, algo que HMM no es capaz de lograr cuando se tiene en cuenta los datos del Kinect®.

	1	2	3	4	5
1	15.02±15.07	26.04±8.48	22.37±5.89	19.53±20.43	17.04±1.47
2	39.14±8.34	43.68±3.56	0.00±0.00	14.68±9.27	2.50±2.63
3	6.06±6.39	14.80±3.00	56.96±8.16	4.07±4.29	18.11±0.49
4	36.62±20.00	4.66±4.92	0.00±0.00	58.72±24.91	0.00±0.00
5	3.16±1.68	10.82±6.38	20.67±6.73	3.00±1.86	62.35±9.57
Media y Varianza de la diagonal: 47.35±12.26					

Tabla 37. HMM-Kinect+IMU (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	52.85±7.53	47.15±7.53	0.00±0.00	0.00±0.00	0.00±0.00
2	5.29±5.58	54.18±4.41	13.50±20.91	18.36±19.35	8.67±8.43
3	0.00±0.00	3.78±3.99	81.42±19.59	10.67±11.24	4.14±4.36
4	0.00±0.00	2.69±2.43	0.00±0.00	77.31±18.65	0.00±0.00
5	0.00±0.00	1.68±1.77	22.86±2.25	0.00±0.00	75.46±0.48
Media y Varianza de la diagonal: 68.24±10.13					

Tabla 38. HMM-Kinect+IMU (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.

Al determinar el desempeño del HMM al detectar las cinco actividades con los datos provistos por la combinación Kinect+IMU+EMG (tablas 39 y 40), hay una mejora de rendimiento para ambas técnicas de fusión en comparación con las anteriores pruebas, ya que se detectan los datos con una precisión superior al 50 %, además de que las varianzas mejoran, más que todo con la fusión por características donde la varianza de la diagonal principal de dicha matriz de confusión es menor a 10 %.

	1	2	3	4	5
1	73.09±18.48	11.15±0.05	0.00±0.00	15.76±18.53	0.00±0.00
2	10.05±10.60	69.16±22.71	2.12±17.85	13.16±1.19	5.51±14.27
3	10.56±3.93	6.42±6.77	56.08±1.12	16.12±1.94	10.82±3.65
4	28.53±2.36	11.09±35.9	0.74±0.78	59.63±2.01	0.00±0.00
5	1.18±1.24	16.39±1.89	19.81±11.30	0.67±0.71	61.95±11.36
Media y Varianza de la diagonal: 63.98±11.14					

Tabla 39. HMM-Kinect+IMU+EMG (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus etiquetas.

	1	2	3	4	5
1	55.06±17.04	36.34±24.77	0.00±0.00	8.60±9.73	0.00±0.00
2	18.13±17.67	55.87±20.11	5.30±10.91	4.84±4.97	15.86±14.04
3	0.00±0.00	2.80±3.35	82.77±3.65	5.18±7.39	9.25±8.77
4	21.16±10.95	1.44±6.48	0.00±0.00	77.40±2.060	0.00±0.00
5	5.65±4.00	3.55±5.96	11.93±5.49	3.98±1.50	74.89±6.89
Media y Varianza de la diagonal: 69.20±9.95					

Tabla 40. HMM-Kinect+IMU+EMG (KmC y cinco movimientos primitivos). Fusión de sensores por medio de sus características.

En general se puede concluir que para el reconocimiento de las 5 actividades propuestas en este trabajo, solo son necesarios 5 movimientos primitivos, ya que al contar con menos o más *clusters*, se pierde desempeño al detectar dichas actividades, como se puede observar en la tabla 41.

		SPECTRAL CLUSTERING		K-MEANS CLUSTERING	
		4 Clusters	5 Clusters	4 Clusters	5 Clusters
Kinect		62.97±9.940	71.02±6.050	65.62±12.84	70.06±7.080
Kinect+IMU	Fusión por Etiquetas	56.96±8.970	70.87±17.08	75.27±14.48	47.35±12.26
	Fusión por Características	48.77±18.30	74.99±13.06	50.84±24.44	68.24±10.13
Kinect+IMU+EMG	Fusión por Etiquetas	57.68±11.75	64.21±11.51	65.21±9.060	63.98±11.14
	Fusión por Características	48.96±18.49	69.94±9.910	48.50±13.28	69.20±9.950
PROMEDIO		55.07±13.49	70.21±11.52	61.09±14.82	63.77±10.11

Tabla 41. Medias y varianzas de las diagonales de las matrices de confusión del análisis de reconocimiento de actividad física humana

En cuanto al desempeño del HMM dado la información provista por las técnicas de agrupamiento, *Spectral Clustering* con 5 *clusters* ofrece el mejor desempeño debido a que posee el sensor que detecta con una precisión del 71.02% y con la varianza mas pequeña, la cual es de 6.05%. También se agrega que esta técnica logra un promedio de precisión del 70.21%±11.52% para todas las combinaciones de sensores con mejores resultados, y tiene otras características especiales que cuenta con una fusión de sensores (tabla 28) en el que el HMM detecta tres actividades con una precisión superior al 90%, y otra fusión de sensores (tabla 27) la cual detecta las 5 actividades con una tasa superior al 50%. Pero *Spectral Clustering* con 4 *clusters*, ofrece el rendimiento mas bajo, ya que los resultados de sus pruebas no destacan y dichos valores son muy pobres para tenerlos en cuenta en la clasificación de actividades físicas.

En cambio la información proveniente por *K-means Clustering* con 4 y 5 *clusters*, ya que en el caso con 5 *clusters*, el HMM ofrece un gran desempeño generando la varianza general mas pequeña de los datos provenientes de las combinaciones de sensores, además de que tres de dichas combinaciones (tablas 38, 39 y 40) aportan una precisión del mas del 50% en la detección de cada una de las actividades, y cuando solo se tiene en sensor Kinect® (ver tabla

36), este ofrece la información suficiente para que el HMM detecte una actividad con una precisión del 100 %. Para el caso de 4 *clusters*, existen 2 combinaciones que ofrecen datos al HMM para que esta técnica detecte una actividad con una tasa de precisión del 100 %, en el cual, una de estas combinaciones también tiene la capacidad de detectar las 5 actividades con un porcentaje superior al 50 %.

Al observar cada una de las tablas de confusión de esta etapa, se determina que la actividad “Levantar mano derecha”, la cual tiene la etiqueta número 4 como se muestra en la tabla 1, es fácilmente detectable por el HMM, ya que en muchas pruebas, el HMM tiene a reconocer la tabla con un porcentaje alto de precisión, llegando el caso de que dicho porcentaje sea del 100 %. Este caso especial se debe a que se tomaron las características más importantes por la técnica *ReliefF*, en el cual, la mayor parte de dichas características se encuentran en el brazo y mano derecha, como se observa en la figura 5. Para la actividad “Saltar” se tiene una situación diferente, en que el HMM se le dificulta detectar esta actividad. Como se explicó anteriormente, esto sucede porque hay una interplación entre los datos de dicha actividad con la actividad “Quedarse” obtenidos por el Kinect®[®], y esta interpolación se origina porque se normalizaron todos los datos de cada punto articulado con el punto articulado número 7 (ver figuras 5 y 6). Así que *Spectral Clustering* y *K-means Clustering*, clasifican el movimiento primitivo “Reposo” como si fuera el movimiento primitivo “Suspendido en el aire”.

8. CONCLUSIONES Y RECOMENDACIONES

8.1. CONCLUSIONES

Se desarrolló una metodología para el reconocimiento de actividad humana utilizando lenguaje no supervisado en un sistema multimodal compuesto por una red de sensores IMU y EMG, y un sensor de profundidad conocido como Kinect®. Para este proceso se utilizan técnicas de agrupamiento como *Spectral Clustering* y *K-means Clustering* para asignar el número adecuado de movimientos primitivos los cuales son necesarios para el reconocimiento de actividad que se logra al aplicar *Hidden Markov Model* (HMM). Gracias a este procedimiento se logra determinar que solo es necesario 5 movimientos primitivos para detectar todo el conjunto de actividades físicas, reduciendo la complejidad que provoca al utilizar más *clusters* como sucede en [4].

Para agilizar este tipo de proceso con una pérdida insignificante de rendimiento, se hace uso de técnicas de selección de características relevantes, como *PCA* y *ReliefF*, donde este último logra aportar los mejores resultados para el sistema de reconocimiento, ya que determina cuáles son los sensores IMU y EMG, y los puntos articulados del Kinect® que aportan más información al sistema. De esta manera se eliminan los sensores o puntos articulados menos significantes y se reduce la dimensionalidad de los datos, provocando que los algoritmos corran eficientemente y que la pérdida de información sea poca para no afectar las técnicas de clasificación. Además, para este caso, los movimientos de las extremidades superiores del cuerpo (brazo derecho e izquierdo) son las que más datos relevantes aportan para reconocer las actividades.

En cuanto a las técnicas de agrupamiento, *Spectral Clustering* muestra una ligera ventaja en comparación con *K-means Clustering*, debido a su estructura más compleja que le beneficia realizar mejores agrupaciones. Aun así *K-means Clustering* entrega resultados satisfactorios, debido a la reducción de características cuando se utiliza la técnica *ReliefF*, por lo tanto, *K-means Clustering* podría ser utilizado más adecuadamente en grandes grupos de personas debido a que opera más rápido, mientras que *Spectral Clustering* puede usarse para ocasiones más especiales donde la extrema precisión sea un objetivo de vital importancia.

Se comprueba que la fusión de dos sensores como es el caso de Kinect+IMU aporta los mejores resultados para el reconocimiento de actividades al utilizar *Spectral Clustering* o *K-means Clustering*, inclusive al usar solamente el Kinect se tiene resultados satisfactorios, teniendo en cuenta los problemas de normalización que hay. Pero puede existir fenomenos como oclusiones parciales o auto-occlusiones para el sensor Kinect® , o la pérdida de conexión en los sistemas de comunicación inalámbrica donde generalmente están conectados los sensores IMU o EMG, por tal motivo la combinación Kinect+IMU+EMG, puede ser esencial, ya que ofrece una mayor robustez a este tipo de fallos, y sus resultados siguen siendo buenos sin importar el ruido que genera los sensores EMG como se muestra en la tablas 39 y 40.

El uso de esta metodología a dado buenos en el momento de aplicar metodos de aprendizaje no supervisado al reconocimiento de actividades físicas humana en sistemas multimodales. Sin embargo aún se requiere realizar un estudio más profundo de las técnicas usadas con el fin de optimizar la metodología y su debido proceso de validación. Esto con el fin de que esta metodología no esté compometida por la dispersión e interpolación de las muestras que entrega esta base de datos u otra diferente. Además se añade que esta metodología no es totalmente no supervisada, ya que se requiere de técnicas de aprendizaje supervisado como HMM y K-Vecinos más cercanos para la clasificación de movimientos primitivos y actividades físicas, así que en futuros trabajos se tratará lo más posible de convertir esta metodología en una verdadera metodología no supervisada.

8.2. RECOMENDACIONES

El nucleo de este trabajo se centra en una metodología de aprendizaje no supervisado para el entrenamiento y validación de clasificación para el reconocimiento de actividades físicas humanas. Sin embargo se requiere un estudio mas exhaustivo de cada una de las etapas de este proyecto, con el fin de optimizar la metodología y su proceso de validación. En este trabajo se observa que existe un error de clasificación con la actividad 3, el cual es provocado por la normalización se realiza a todo el cuerpo entero tomando como referencia el punto articulado 7. No solamente con el Kinect® se presenta este tipo de problemas, los sensores

IMU y EMG ofrecen datos que el clasificador tiende a similar ciertos movimientos primitivos y ciertas actividad con otras, más aún con los sensores EMG que presentan un ruido que afecta el reconocimiento de actividad. Para dar solución a esto, se puede optimizar las técnicas de clasificación usadas o utilizar otras diferentes como es el caso de Mezclas Gaussianas Infinitas (IGMM; *Infinite Gaussian Mixture Model*), *Density-Based Spatial Clustering of applications with noise* (DBSCAN), entre otras. Otra solución es hacer un re-ajuste a la base de datos, como variar la amplitud de la ventana de tiempo para el Kinect® o los sensores IMU o utilizar otras técnicas de procesamiento de señales para obtener mejores muestras con los sensores EMG.

Por otro lado se espera que la metodología utilizada tenga menos métodos o técnicas de aprendizaje supervisado, ya que por el momento, hace mucho uso intensivo de estas técnicas como es el caso de K-vecinos más cercanos o HMM para el reconocimiento de actividades. Como opciones adecuadas se puede utilizar el algoritmo *Expectation-Maximization* (EM), Redes Neuronales Artificiales (*Neural Networks*), entre otras.

También se espera asignar este tipo de trabajos a actividades físicas más complejas donde su ejecución requiere más tiempo en comparación con las actividades físicas tratadas en este trabajo, y así poder expandir esta metodología a nuevas fronteras y avanzar más en el estado del arte.

BIBLIOGRAFÍA

- [1] UNIVERSITY OF ROSTOCK. Unsupervised Detection of Motion Primitives in very High Dimensional Sensor Data. Karlsruhe, Dept. of Computer Science, 2010. 1, 1.1
- [2] AOKI, T.; VENTURE, G. y KULIC, D. Segmentation of Human Body Movement Using Inertial Measurement Unit. En: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, 2013, págs. 1181–1186. 1, 1.1, 1.2, 3
- [3] MANNINI, Andrea y SABATINI, Angelo Maria. Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. En: OPEN ACCESS SENSORS, 2010. 1, 1.1, 1.2, 3
- [4] SALCEDO, ANDRÉS FELIPE CALVO. RECONOCIMIENTO AUTOMÁTICO DE ACTIVIDADES FÍSICAS HUMANAS EN SISTEMAS MULTIMODALES. Tesis de Maestría, UNIVERSIDAD TECNOLÓGICA DE PEREIRA, 2015. 1, 1.1, 1.2, 3, 4.2, 6.4, 7.1, 7.1, 7.2, 8.1
- [5] ZHOU, Feng; DE LA TORRE, F. y HODGINS, J.K. Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion. En: Pattern Analysis and Machine Intelligence, IEEE Transactions on, tomo 35, nº 3, 2013, págs. 582–596. ISSN 0162-8828. 1.1, 1.2, 3
- [6] SPRIGGS, E.H.; DE LA TORRE, F. y HEBERT, M. Temporal segmentation and activity classification from first-person sensing. En: Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, 2009. ISSN 2160-7508, págs. 17–24. 1.1, 1.2
- [7] YANG, Yang; SALEEMI, I. y SHAH, M. Discovering Motion Primitives for Unsupervised Grouping and One-Shot Learning of Human Actions, Gestures, and Expressions. En: Pattern Analysis and Machine Intelligence, IEEE Transactions on, tomo 35, nº 7, 2013, págs. 1635–1648. ISSN 0162-8828. 1.1, 1.2

- [8] SIGNAL AND IMAGE PROCESSING INSTITUTE. Motion Primitive-Based Human Activity Recognition Using a Bag-of-Features Approach. University of Southern California, 2012. 1.1, 1.2, 3, 6.3
- [9] WU, Chenxia, *et al.* Watch-n-patch: Unsupervised understanding of actions and relations. En: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 2015, págs. 4362–4370. 1.2, 3, 4.2.1
- [10] CHRISTIAN VOLLMER, Julian P. Eggert y GROSS, Horst-Michael. Modeling Human Motion Trajectories by SparseActivation of Motion Primitives Learned from Unpartitioned Data. En: in: Proc. 35th German Conference on Artificial Intelligence, LNCS 7526, 2012, págs. pp. 168–179. 1.2
- [11] PANAHANDEH, G., *et al.* Continuous Hidden Markov Model for Pedestrian Activity Classification and Gait Analysis. En: Instrumentation and Measurement, IEEE Transactions on, tomo 62, nº 5, 2013, págs. 1073–1083. ISSN 0018-9456. 3
- [12] HORINEK, F., *et al.* Design of Electromyography classification system using Artificial Neural Network. En: Radioelektronika (RADIOELEKTRONIKA), 2014 24th International Conference, 2014, págs. 1–4. 3, 4.6.2
- [13] WANG, Qian, *et al.* A novel pedestrian dead reckoning algorithm using wearable EMG sensors to measure walking strides. En: Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010, 2010, págs. 1–8. 3, 4.6.2
- [14] DOMINIK ENDRES, Yaron Meirovitch, Tamar Flash y GIESE, Martin A. Segmenting sign language into motor primitives with Bayesian binning. En: Frontiers in COMPUTATIONAL NEUROSCIENCE, 2013. 3
- [15] KIRA, Kenji y RENDELL, Larry A. A practical approach to feature selection. En: Proceedings of the ninth international workshop on Machine learning, 1992, págs. 249–256. 3, 5.1.2, 7.1

- [16] KHALID, S.; KHALIL, T. y NASREEN, S. A survey of feature selection and feature extraction techniques in machine learning. En: Science and Information Conference (SAI), 2014, 2014, págs. 372–378. 4.1
- [17] SHANG, X. y CHISHOLM, L.A. Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms. En: Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, tomo 7, nº 6, 2014, págs. 2481–2489. ISSN 1939-1404. 4.2
- [18] VIDHATE, D. y KULKARNI, P. Cooperative Machine Learning with Information Fusion for Dynamic Decision Making in Diagnostic Applications. En: Advances in Mobile Network, Communication and its Applications (MNCAPPS), 2012 International Conference on, 2012, págs. 70–74. 4.3
- [19] JIN, Yaochu y SENDHOFF, B. Pareto-Based Multiobjective Machine Learning: An Overview and Case Studies. En: Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, tomo 38, nº 3, 2008, págs. 397–415. ISSN 1094-6977. 4.3
- [20] YANG, Yang, *et al.* Multitask Spectral Clustering by Exploring Intertask Correlation. En: Cybernetics, IEEE Transactions on, tomo 45, nº 5, 2015, págs. 1083–1094. ISSN 2168-2267. 4.4
- [21] WANG, Jing, *et al.* A novel clustering algorithm for Unsupervised Relation Extraction. En: Digital Information Management (ICDIM), 2012 Seventh International Conference on, 2012. ISSN pending, págs. 16–21. 4.4
- [22] ZHANG, Zhenya, *et al.* Clustering aggregation based on genetic algorithm for documents clustering. En: Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence). IEEE Congress on, 2008, págs. 3156–3161. 4.4
- [23] MARINA SOKOLOVA, Guy Lapalme. A systematic analysis of performance measures for classification tasks. En: Information Processing and Management, tomo 45, 2009, págs. 427–437. 4.5

- [24] CHEN, Tsung-Lin. Design and Analysis of a Fault-Tolerant Coplanar Gyro-Free Inertial Measurement Unit. En: Microelectromechanical Systems, Journal of, tomo 17, nº 1, 2008, págs. 201–212. ISSN 1057-7157. 4.6.1
- [25] SHU-ZHI, Mu; XIONG ZHU, Bu y YONG XIN, Li. Optimization Design and Calibration of Installation Error Coefficients for Gyroscope-Free Strapdown Inertial Measurement Unit. En: Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on, 2006, págs. 238–242. 4.6.1
- [26] DENG, Teng, *et al.* Kinect Shadow Detection and Classification. En: Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on, 2013, págs. 708–713. 4.6.3
- [27] SYMS, C. Principal components analysis. En: , 2008. 5.1.1
- [28] NG, Andrew Y, *et al.* On spectral clustering: Analysis and an algorithm. En: Advances in neural information processing systems, tomo 2, 2002, págs. 849–856. 5.2.1
- [29] KANUNGO, Tapas, *et al.* An efficient k-means clustering algorithm: analysis and implementation. En: Pattern Analysis and Machine Intelligence, IEEE Transactions on, tomo 24, nº 7, 2002, págs. 881–892. ISSN 0162-8828. 5.2.2
- [30] RABINER, L. y JUANG, B.H. An introduction to hidden Markov models. En: ASSP Magazine, IEEE, tomo 3, nº 1, 1986, págs. 4–16. ISSN 0740-7467. 5.3.1
- [31] BLUNSOM, Phil. Hidden Markov Models, 2004. 5.3.1, 5.3.1
- [32] HORTON, Paul y NAKAI, Kenta. Better Prediction of Protein Cellular Localization Sites with the it k Nearest Neighbors Classifier. En: Ismb, tomo 5, 1997, págs. 147–152. 5.3.2
- [33] ZHANG, Mi y SAWCHUK, Alexander A. A feature selection-based framework for human activity recognition using wearable multimodal sensors. En: Proceedings of the 6th International Conference on Body Area Networks. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011, págs. 92–98. 6.1.1

- [34] XIA, Lu; CHEN, Chia-Chih y AGGARWAL, Jake K. Human detection using depth information by kinect. En: CVPR 2011 WORKSHOPS. IEEE, 2011, págs. 15–22. 6.1.1
- [35] PHINYOMARK, A, *et al.* Feature extraction and reduction of wavelet transform coefficients for EMG pattern classification. En: Elektronika ir Elektrotechnika, tomo 122, nº 6, 2012, págs. 27–32. 6.1.1
- [36] QIAN, Z, *et al.* Robust principle component analysis with complex noise. En: International Conference on Machine Learning, 2014. 7.1

9. ANÉXOS

Las tablas de confusión de las demás pruebas realizadas para el reconocimiento de movimientos primitivos se encuentran en los archivos conjuntos a este trabajo. Las carpetas que indican el número de clusters tienen en su interior cada prueba realizada con las diferentes combinaciones de sensores utilizando *K-means* y *Spectral Clustering*, esto escrito en un “.txt”. En el caso de *Spectral Clustering* también están las pruebas al cambiar las dimensiones.