

Evaluación de modelos para el reconocimiento de gestos en señales biométricas, para un usuario con movilidad reducida

Evaluation of Models for Gesture Recognition from Biometric Signals of a Person with Reduced Mobility

Holman S. Cabezas ¹ y Wilson J. Sarmiento  ²

Recibido: 25 de septiembre de 2019

Aceptado: 19 de noviembre de 2019

Cómo citar / How to cite

H. S. Cabezas, W. J. Sarmiento, “Evaluación de modelos para el reconocimiento de gestos en señales biométricas, para un usuario con movilidad reducida”, *TecnoLógicas*, vol. 22, pp. 33-47, 2019. <https://doi.org/10.22430/22565337.1513>



¹ Ingeniero en Multimedia, Grupo de Investigación en Multimedia -GIM, Facultad de Ingeniería, Universidad Militar Nueva Granada, Bogotá-Colombia, u1201569@unimilitar.edu.co

² PhD. en Ciencias de la Electrónica-Áreas Computación, Grupo de Investigación en Multimedia -GIM, Facultad de Ingeniería, Universidad Militar Nueva Granada, Bogotá-Colombia, wilson.sarmiento@unimilitar.edu.co

Resumen

Este trabajo presenta los resultados de una comparación de tres modelos computacionales (reconocimiento de patrones, modelos ocultos de Markov y bolsas de características), para el reconocimiento de gestos por medio del procesamiento de señales biométricas, para un usuario con movilidad reducida. La evaluación involucra ocho gestos diseñados de forma participativa con un usuario con problemas de movilidad y se desarrolló mediante un esquema de validación cruzada, en el que se calcularon métricas de sensibilidad y precisión, para un conjunto de datos formado por diez repeticiones de cada gesto. Los resultados obtenidos permitieron concluir que las bolsas de características son el modelo con mejor desempeño para las dos métricas evaluadas. El modelo de reconocimiento de patrones al usar máquinas de soporte vectorial mostró los resultados más estables y los modelos ocultos de Markov presentaron el desempeño más bajo.

Palabras clave

Reconocimiento de gestos, interacción hombre-máquina, procesamiento de señales, aprendizaje computacional, reconocimiento de patrones.

Abstract

This paper compares the results of three computational models (pattern recognition, hidden Markov models, and bag of features) for recognizing the hand gestures of a user with reduced mobility using biometric signal processing. The evaluation of the models included eight gestures co-designed with a person with reduced mobility. The models were evaluated using a cross-validation scheme, calculating sensitivity and precision metrics, and a data set of ten repetitions of each gesture. It can be concluded that the bag-of-features model achieved the best performance considering the two metrics under evaluation; the traditional pattern recognition model, using vector support machines, produced the most stable results; and the hidden Markov models had the lowest performance.

Keywords

Gesture recognition, Human computer interaction, Signal processing, Machine learning, Pattern recognition.

1. INTRODUCCIÓN

La interacción mediante comandos gestuales, particularmente de la mano o brazo, es uno de los objetivos de la interacción natural y se espera que permita y facilite procesos interactivos intuitivos al usuario [1]-[3]. Sin embargo, el reconocimiento de gestos presenta todavía grandes limitaciones que dificultan su implantación como una alternativa real de interacción para un usuario final.

Una de estas limitaciones se origina en el proceso de captura o adquisición de información, en el que existen principalmente dos opciones: 1) sensores ópticos y 2) sensores de señales biométricas.

El uso de sensores ópticos (cámaras) es la opción más usada y estudiada para el reconocimiento de gestos [3]-[9]; esta ha incentivado la fabricación de dispositivos comerciales ampliamente conocidos, como el Microsoft Kinect y el Leap Motion, y equipos especializados en los laboratorios de marcha y captura de movimiento para producción audiovisual. Los métodos de captura son muy variados e incluyen imágenes RGB, infrarrojas (IR), de profundidad (*depth*), cámaras estéreo y arreglos de cámaras, entre otros [3]-[9].

Las principales limitaciones del uso de sensores ópticos son la necesidad de tener una vista directa entre el usuario y los sensores y el requerimiento de un espacio adecuado de trabajo, donde los sensores puedan ser correctamente instalados.

En cuanto a la segunda opción, el uso de sensores biométricos, resaltan los sensores inerciales (acelerómetros y giroscopios) [10]-[12] y las señales de electromiografía (EMG) [13]-[20].

Estas últimas han cobrado importancia recientemente debido a que las señales de EMG del brazo/antebrazo contienen información relevante sobre la actividad eléctrica producida por el proceso neuromuscular, por lo tanto, es posible analizar esas señales para un

reconocimiento efectivo de los movimientos de contracción o relajación de dedos, mano, brazo y antebrazo [13]-[20]. Lo anterior ha dado lugar al diseño de diferentes sistemas comerciales de reconocimiento de gestos usando señales EMG, para fines de interacción [13]-[20]. La principal limitación del uso de sensores biométricos es su carácter “invasivo”, que implica vestir o usar un dispositivo.

Otra importante limitación del reconocimiento de gestos en interacción humano-computador es la efectividad de los métodos de análisis y reconocimiento [3]-[8]. Si bien es cierto que los dispositivos comerciales facilitan un proceso de reconocimiento de conjunto de gestos prediseñados, este es limitado debido a que, muchas veces, el conjunto de gestos no es suficiente, o a que las características particulares de los usuarios (anatómicas o funcionales) dificultan o impiden su uso.

Adicionalmente, un conjunto de gestos fijo impide cualquier personalización de la interacción; si se considera que la flexibilidad, la libertad y el control que tenga el usuario de la interfaz son aspectos importantes de la usabilidad [1], [2], [21], su ausencia es otra importante limitación.

Lo anterior motiva, en muchos casos, a adelantar un proceso de diseño del conjunto de gestos para la interacción y el sistema de reconocimiento. Esta causa es la principal razón de este trabajo: el diseño de un sistema de interacción para una persona con problemas de movilidad motora.

La revisión del estado del arte presenta una amplia gama de métodos y técnicas que pueden ser utilizados [3]-[8], pero pocos trabajos han comparado de forma experimental diferentes modelos computacionales. Generalmente, las propuestas reportan la efectividad del método y una comparación contra métodos del mismo modelo computacional.

Este trabajo ofrece un aporte en esa dirección y presenta una comparación de tres modelos de reconocimientos de gestos

presentes en el estado del arte, aplicados a un usuario con movilidad reducida. Dichos modelos han reportado un buen comportamiento por sus autores: 1) reconocimiento tradicional de patrones usando máquinas de soporte vectorial [19], [22]; 2) modelos ocultos de Markov [23], [24] y 3) bolsas de características [25], [26].

Se aclara que, en el estado del arte se han identificado trabajos que usan aprendizaje profundo; sin embargo, ese modelo no fue considerado en este trabajo, debido a que requiere una gran cantidad de muestras de entrenamiento. La principal motivación de este trabajo —una persona con problemas de movilidad motora— impide obtener la cantidad de muestras para usar aprendizaje profundo. Los tres modelos elegidos se aplicaron a un conjunto de gestos diseñados para el usuario. Se usó un dispositivo comercial de EMG, debido a que su diseño inalámbrico brindaba un adecuado confort al usuario y no requería un espacio preconfigurado con sensores.

Las siguientes secciones se presentan de la siguiente forma: en la sección 2 se ofrece una revisión de trabajos previos y se enfatiza en aquellos que han propuesto el uso de dispositivos comerciales para el reconocimiento de gestos por EMG.

En la sección 3 se describen los métodos usados, desde el diseño del conjunto de gestos, la captura de los datos y la comparación computacional de los métodos de reconocimiento. En la siguiente sección se expone un análisis de los resultados, en el que se presentan, principalmente, métricas de sensibilidad y especificidad de los métodos. Finalmente, en la última sección se plantean las conclusiones finales y algunas reflexiones sobre trabajos futuros.

2. TRABAJOS RELACIONADOS

El reconocimiento de patrones en señales biométricas, principalmente con

señales de EMG, es un método muy importante para la clasificación de movimientos de las extremidades superiores [5], [6], [8]. Sin embargo, hay muchos factores como los clasificadores, las longitudes de ventana y el número de canales que se pueden analizar, que influyen en la eficiencia del proceso.

Por tal razón, uno de los mayores desafíos para el reconocimiento de las tareas motoras ha sido crear sistemas de clasificación robustos, que empleen algoritmos que permitan procesar la información con el menor costo computacional y un alto porcentaje de acierto en la identificación de los movimientos.

Se han dedicado importantes esfuerzos al preprocesamiento, el filtrado, la extracción/reducción de características y la clasificación de patrones dirigidos al recogimiento de movimientos de los miembros superiores, usando señales EMG.

Tal es el caso de la propuesta de [27], quienes mostraron que es posible la identificación de cinco movimientos de la mano con un rango de exactitud entre el 94 % a 99 %. También se ha sugerido la detección de movimientos de la muñeca y del dedo anular, lo cual procesó el mismo tipo y alcanzó una precisión del 87,3 %, al usar máquinas de soporte vectorial (SVM) [22].

Dardas y Georganas presentan un sistema que en tiempo real permite la interacción con una aplicación o videojuego a través de gestos con las manos [25].

La propuesta usa bolsas de características (BoF) y SMV, en la que construye un vocabulario que genera comandos de gestos para controlar la aplicación [25].

Oskoei y Hu hacen uso de diferentes clasificadores, como el análisis de discriminante lineal, las redes neuronales artificiales y SMV, en la clasificación de nueve clases de movimientos, a partir de señales EMG [28]. Gracias a los resultados obtenidos, los autores concluyen que el

clasificador SMV, al usar características frecuenciales, tiene la mejor precisión y alcanza una tasa de errores de clasificación media de 1,53 %, con una desviación estándar de 1,08 [28]. Un trabajo similar alcanza el reconocimiento de ocho gestos de la mano con una exactitud entre 96,9 % y 99,65 % [29].

Trabajos más recientes han propuesto modelos que pueden llegar a procesar la clasificación de este tipo de señales con resultados sobresalientes y una menor complejidad computacional. Este es el caso de la propuesta de Rossi y col. [19], que emplea un enfoque mixto con modelos ocultos de Markov (HMM) y SMV. El trabajo mostró que, independientemente del rendimiento de SMV, se puede optimizar un buen clasificador con métodos típicos (HMM) y lograr una mejora del 12 % de precisión del reconocimiento [19].

Xu y col. [12] presentan tres (3) modelos diferentes para el reconocimiento de siete gestos de las manos (arriba, abajo, izquierda, derecha, tic, círculo y cruz), en función de las señales de entrada de los acelerómetros, los tres (3) ejes de movimiento. En este punto, se muestra que, para el mejor caso, se logra una precisión de reconocimiento general del 95,6 %.

Otros trabajos se han dirigido al diseño de nuevos sensores, que permitan la adquisición de señales EMG, para posterior clasificación. Es el caso del trabajo de Du y col. [14], que usa técnicas de aprendizaje profundo, con una precisión de reconocimiento del 85,0 %.

Nuevamente, se resalta que los trabajos publicados no contrastan de forma explícita y experimental diferentes modelos computacionales para el reconocimiento de gesto; los autores generalmente reportan únicamente la efectividad de su método y una comparación contra métodos del mismo modelo computacional.

3. MÉTODOS

3.1 Diseño de gestos

El proceso de definición de los gestos se inspiró en los grados de libertad de la muñeca y el antebrazo, para generar un movimiento adecuado y ergonómico al usuario. El proceso involucró un diseño participativo con un usuario con limitaciones motoras. Este sufrió un traumatismo vertebro-medular con afectación de las vértebras de la C3 a la C8, en las cuales presenta un déficit sensitivo motor completo en más de tres segmentos medulares y por debajo del nivel de la lesión cervical. Lo anterior le impide el movimiento completo de los miembros inferiores y presenta una restricción importante de movimiento en muñeca, mano y dedos. Se aclara que la atrofia y pérdida de masa muscular genera condiciones particulares de las señales de EMG.

En la Fig. 1, se puede observar la posición final de los siete gestos para los que el usuario manifestó un adecuado grado de comodidad. Es importante anotar que los gestos fueron diseñados bajo la definición de gesto dinámico, es decir, como un movimiento de la mano y no como una pose estática[3]-[8]. Así, todos los gestos inician en la posición de relajación del brazo (h).

Adicionalmente, se incluyó el gesto de relajación, que se entiende como el mantenimiento del brazo y la mano en la posición de relajación por un periodo de aproximadamente dos (2) segundos.

3.2 Adquisición de señales

El proceso de adquisición se hizo mediante el uso de un brazalete *Myo gesture armban*, el cual estaba equipado con ocho (8) sensores de electromiografía (EMG) y seis (6) sensores inerciales (acelerómetros y giroscopios en los tres ejes).

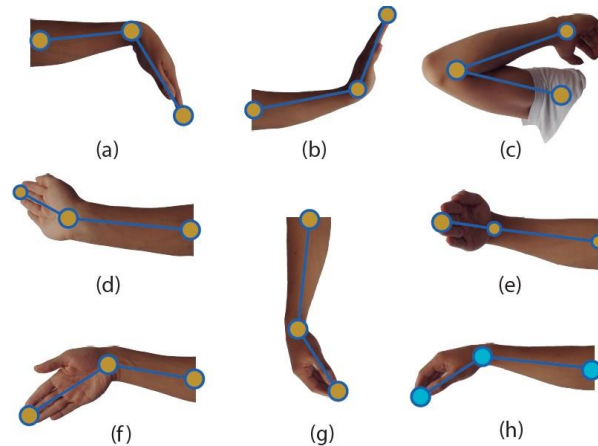


Fig. 1. Posición final del conjunto de gestos diseñados. Todos los gestos inician en la posición de relajación (h)
Fuente: elaboración propia.

Al comienzo de la sesión de adquisición de datos, se mostró al usuario la secuencia de movimientos que debía hacer para cada uno de los gestos. El dispositivo se ubicó en el brazo derecho del usuario y se siguieron las recomendaciones del laboratorio *Thalnic*, por lo que se ubicó el electrodo principal sobre el músculo braquiorradial y el extensor radial del carpo [30].

Se hizo un ajuste del ancho del brazalete con base en las indicaciones del fabricante, de forma que no generara presión sobre el brazo, pero permitiera una adecuada ubicación de los sensores sobre los músculos. El usuario hizo varias repeticiones de cada gesto, para lo cual se tuvo en cuenta que cada uno implica una duración del movimiento entre uno (1) y tres (3) segundos, dejando un (1) minuto de descanso entre cada toma. Para evitar la fatiga muscular, se dividió el proceso de captura en dos sesiones, con una diferencia de una (1) semana. Al sumar los datos adquiridos en las dos sesiones, se completó una adquisición de diez repeticiones de cada gesto (incluido el gesto de relajación), para un total de 1120 señales, así: ocho gestos, diez repeticiones, catorce señales (ocho de electromiografía y seis inerciales).

3.3 Extracción de características

Cada una de las catorce señales adquiridas fue procesada de igual forma,

para adelantar un proceso de extracción de características que permita diferenciar los múltiples gestos computacionalmente.

El primer paso fue una segmentación temporal de la señal, para lo cual se usó una ventana deslizante de W segundos de duración, con un solapamiento de 0,5 segundos. Cada segmento de señal fue caracterizado, en tiempo, usando la entropía, la mediana y los cuatro (4) primeros momentos estadísticos (media, varianza, kurtosis y sesgo), la entropía y la energía; y, en frecuencia, por las siete frecuencias más altas del espectro de potencia de la señal ($F_1 \dots F_7$), después de eliminar la componente DC mediante el procesamiento Offset-DC [31]. Dado que el proceso de extracción de características se aplica a cada una de las señales de forma independiente, se obtiene un vector \hat{f}_w de $13 \times 14 = 182$ dimensiones.

3.4 Clasificación

En este punto, cada uno de los modelos computacionales hace un procesamiento diferenciado, definido a la representación de un gesto en cada modelo y, por lo tanto, en el reconocimiento del mismo.

Las siguientes secciones describen cada uno de los modelos evaluados.

3.5 Reconocimiento de patrones

Si bien es cierto que todo modelo de inteligencia computacional es en esencia un reconocimiento de patrones, en el contexto de este trabajo, dicho término se refiere al modelo “tradicional” que propone un proceso de cuatro (4) pasos: 1) adquisición, 2) extracción de características, 3) selección de características y 4) clasificación —los dos primeros descritos en las secciones anteriores—. En este modelo, un gesto es un punto \hat{f}_w n dimensional en el espacio de características. El proceso de clasificación consiste en separar las clases (gestos diferentes) maximizando la distancia entre clase y minimizando la distancia intraclase. Cuando el espacio de características tiene un número alto de dimensiones, es aconsejable adelantar un proceso de selección de características, que permita identificar un subconjunto ($\hat{f}_w^* \in \hat{f}_w$) de aquellas que aportan mayor relevancia a la clasificación. En este caso, se usó una estrategia subóptima basada en un algoritmo de búsqueda egoísta paso a paso (*greedy stepwise*). Se seleccionó el subconjunto de características (\hat{f}_w^*) que

aportara, por lo menos, el 70% en la clasificación, lo que redujo la dimensión de espacio de característica de 182 a 18.

En la Tabla 1 se expone el detalle de las 18 características seleccionadas. Se resalta la presencia de las señales de los siete (7) primeros sensores de EMG. Diez de las características seleccionadas provienen de los sensores inerciales, lo cual es coherente con el movimiento amplio de extremidad superior al practicar los gestos.

Para la etapa de clasificación, se usó un algoritmo de máquina de soporte vectorial (SMV, del inglés *support vector machine*).

La selección del kernel se hizo de forma experimental, comparando el error de entrenamiento al usar una función polinomial, una función polinomial normalizada, una función radial y una distribución de Pearson tipo VII.

En este caso, el mejor desempeño se logró con el kernel polinomial normalizado.

Para la evaluación del métodos, se tomaron como parámetros de estudio el tamaño de la ventana deslizante (W) variando de $W = \{0.5, 1.0, \dots, 2.5, 3.0\}$ segundos, y el exponente del kernel polinomial normalizado, $E = \{1.0, 1.1, \dots, 9.1, 10.0\}$.

Tabla 1. Vector de 18 característica seleccionadas (\hat{f}_w^*).
En la primera columna la señal de origen. Fuente: elaboración propia.

Señal	Características
EMG 1	kurtosis
EMG 2	F_2
EMG 3	kurtosis, F_1
EMG 4	kurtosis
EMG 5	kurtosis
EMG 6	kurtosis
EMG 7	media
Ac. X	media, F_1
Ac. Y	media, F_2
Ac. Z	media, varianza
Gr. Pitch	media, varianza, kurtosis
Gr. Roll	media

3.6 Modelos ocultos de Markov

Una cadena de Markov es un proceso estocástico de estados discretos, en el que la probabilidad de ocurrencia de un evento depende únicamente del evento anterior (propiedad markoviana). Una cadena de Markov se puede definir como la dupla $M = (Q, A)$, donde Q es el conjunto discreto finito de S estados y A una matriz cuadrada de $S \times S$ que especifica las probabilidades de transición entre los estados. Los modelos ocultos de Markov (HMD, del inglés *hidden Markov model*) son, en esencia, una cadena de Markov desconocida, oculta, que puede ser inferida a partir de una secuencia de observaciones de variables estocásticas (Y'_t) que cumplen con la condición markoviana, donde el número de estados S se asume de forma a priori. El uso de los HMD como modelo de inteligencia computacional fue propuesto por Rabiner para el reconocimiento de voz [32], [33], donde un vocable o sílaba es una instancia de una cadena de Markov.

En reconocimiento de gestos se ha usado de forma directa la propuesta de Rabiner [23], [24], lo que implica que, en el momento de entrenamiento, es necesario inferir tantos modelos de Markov (M_g) como gestos se deben reconocer.

Así, en este modelo computacional, un gesto corresponde a una secuencia de observación (Y'_t). Durante la etapa de clasificación, a un gesto particular se le calcula la probabilidad de ser una observación de un modelo M_g y se le asigna la etiqueta del modelo más probable [32].

Sin embargo, dada la complejidad algorítmica de la inferencia de un HMM, este no usa directamente el vector de características (\hat{f}_w) o (\hat{f}_w^*) como secuencia observada del modelo (Y'_t), sino que se desarrolla un procedimiento de cuantización vectorial (*vector quantization*) del vector de características [34].

De esa forma, un algoritmo de agrupación (*K-means*) calcula las K

agrupaciones de cuantización sobre un conjunto global de características.

Posteriormente, un gesto es cuantizado mediante un histograma de membrecía (VQ_g), donde cada vector de características de una ventana deslizante es asignado a la agrupación más cercana, dada una distancia, en este caso, la euclidiana.

La Fig. 2 ilustra gráficamente el proceso. Se puede observar que la señal del gesto es procesada con tres ventanas deslizantes (W_i). Para fines ilustrativos las ventanas no están sobrepuestas.

El vector de características f_i es calculado a cada ventana. En el ejemplo todo el espacio de características es cuantizado en tres (3) clúster. Así, a cada vector de características se le asigna una membrecía. En este caso, para las tres (3) ventanas deslizantes, se establecieron dos (2) membrecías a la clase naranja, una (1) a la clase azul y ninguna (0) a la clase verde. Entonces, se puede ver que el histograma de membrecías normalizado es el vector de cuantización (VQ_g).

En consecuencia, se toma el histograma (VQ_g), calculado como la secuencia observada (Y'_t) de un HMM. Pruebas preliminares permitieron establecer un mejor rendimiento del algoritmo al usar el subconjunto de características seleccionadas \hat{f}_w^* para el cálculo de los VQ_g .

Para la evaluación se estudiaron tanto el número de estados del HMM, $S = \{3, 4, 5, 6\}$ como el número de agrupaciones de la cuantización vectorial, $K = \{6, 7, \dots, 19, 20\}$.

3.7 Bolsa de características

Las bolsas de palabras (BoW del inglés *bag-of-words*) son un modelo que permite simplificar el procesamiento de documentos al hacer un conteo de frecuencia de aparición de un conjunto finito de palabras claves (diccionario) para cada documento [35].

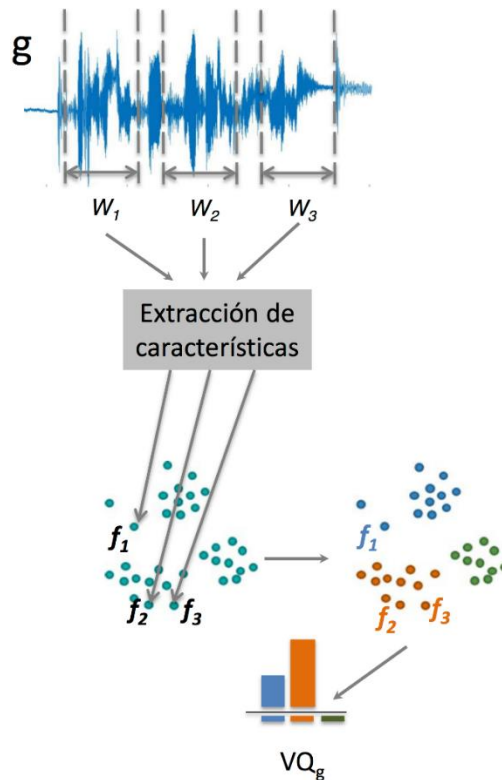


Fig. 2. Proceso de cálculo de un vector de cuantización (VQ_g), para tres (3) clusters de cuantización
Fuente: elaboración propia.

El modelo de bolsa de características (BoF del inglés *bag-of-features*) es una adaptación del concepto BoW, aplicado a la visión por computador.

Mientras que en el modelo BoW el diccionario puede ser establecido de forma previa, dado el dominio del problema, en el modelo BoF es necesario inferir y refinar el diccionario visual. Una de las primeras propuestas propone dividir las imágenes (documentos) en una rejilla regular [25], para lo cual se usan procedimientos de extracción de características típicas de procesamiento de imágenes.

Posteriormente, mediante una estrategia de cuantización vectorial (*vector quantization*) [34], se segmenta el espacio de características. Es importante resaltar que en la interpretación de este modelo cada agrupación o cluster calculado (K) se entiende como una palabra visual y al conjunto de agrupaciones como el diccionario visual. De esta forma, la cuantización VQ de una imagen (o un

segmento de ella) es la frecuencia de aparición de las palabras del diccionario construido. Ese mismo concepto se ha extendido al procesamiento de señales, en el que una señal es dividida en ventanas deslizantes (*frames*) que pueden o no solaparse; asimismo, la construcción de las palabras y el diccionario se hace usando una estrategia de cuantización vectorial (*vector quantization*) [34], como se ilustra en la Fig. 2. En este caso, cada cluster o agrupación se entiende como una palabra, y al conjunto de todas las palabras como diccionario. Posteriormente, los VQ calculados son clasificados por algoritmos tradicionales de aprendizaje computacional [25]. Pruebas preliminares mostraron que las BoF presentaron mejor comportamiento al usar el conjunto completo de características (\hat{f}_w) en la construcción del diccionario de palabras. También se pudo establecer que la alternativa más adecuada para la clasificación era usar un algoritmo SMV con un kernel polinomial normalizado.

Para la evaluación, se estudiaron el número de palabras del diccionario, $K = \{6, 7, \dots, 19, 20\}$, y el exponente del kernel polinomial normalizado, $E = \{1.0, 1.1, \dots, 19.1, 20.0\}$.

3.8 Evaluación de los modelos

Los modelos fueron evaluados usando un esquema de validación cruzada de diez paquetes (*10-folding cross validation*), lo que garantiza una distribución aleatoria de la muestra de gestos tanto en el conjunto de entrenamiento (*train*) como en el conjunto de prueba (*test*). Con este fin, se calcularon como métricas de interés la precisión (*precision*) y la sensibilidad (*recall*). Para los modelos ocultos de Markov y las bolsas de características, se usó una ventana deslizante (W) de dos segundos.

4 ANÁLISIS DE RESULTADOS

La Tabla 2 presenta los resultados de la comparación de los tres modelos computacionales: reconocimiento de patrones con SMV (ML-SMV), modelos ocultos de Markov (HMM) y bolsas de características (Bow).

Las columnas muestran los valores máximos y mínimos de sensibilidad y

precisión de cada método. Se puede observar que los modelos HMM presentaron el rendimiento más bajo en las dos métricas (precisión 0,38% y sensibilidad 0,4 %) y, en general, el rendimiento promedio menos destacado. El modelo de BoF mostró el mejor comportamiento para las dos métricas, con un valor de precisión de 0,99 % para $K = 19$ palabras y un $E = 0,8$ el exponente del kernel polinomial normalizado. Los valores mínimos obtenidos el modelo BoF son superiores al del modelo ML-SMV.

La Fig. 3 muestra las matrices de confusión obtenidas en cada caso, mínimo y máximo para precisión y sensibilidad.

Destaca el hecho de que, para los HMM, se alcanzan los valores más altos de las dos métricas con los mismos parámetros ($S=5$ $K=11$). La escala de color de las matrices de confusión varía entre azul (1) y rojo (1).

Se observa que, para los modelos de BoF y ML-SMV, tanto para los valores máximos y mínimos, el gesto de reposo (H) es el gesto con mejor tasa de reconocimiento. También es posible afirmar que el gesto G es el de mayor dificultad de reconocimiento para estos dos modelos. Es importante anotar que, para los HMM, el gesto de reposo (H) es el que presenta mayor confusión.

Tabla 2. Resumen de comparación de modelos (las), con respecto a las métricas de precisión y sensibilidad (columnas). Fuente: elaboración propia.

	Precisión		Sensibilidad	
	Mín.	Máx.	Mín.	Máx.
ML-SMV	0.771 %	0.980 %	0.771 %	0.976 %
	$E=2.2$	$E=6$	$E=2$	$E=6$
	$W=0.5$	$W=3$	$W=0.5$	$W=3$
HMM	0.380 %	0.690%	0.400%	0,700%
	$S=4$	$S=5$	$S=4$	$S=5$
	$K=14$	$K=11$	$K=16$	$K=11$
BoF	0.380 %	0.690 %	0.400 %	0,700 %
	$E=1.8$	$E=0.8$	$E=15$	$E=6.7$
	$K=8$	$K=19$	$K=18$	$K=16$

Evaluación de modelos para el reconocimiento de gestos en señales biométricas, para un usuario con movilidad reducida

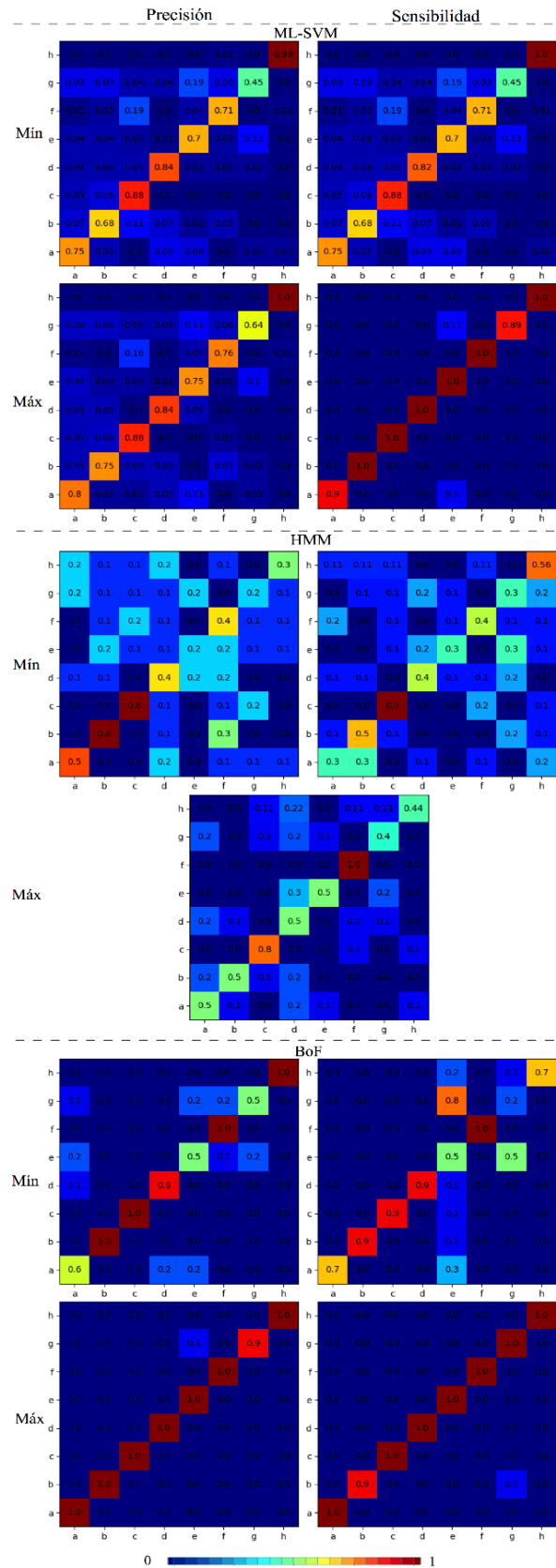


Fig. 3. Matrices de confusión para los vales máximos y mínimos de precisión y sensibilidad
Fuente: elaboración propia.

La Fig. 4 muestra las gráficas obtenidas para cada una de las métricas (curvas de nivel a color) frente a los parámetros evaluados (ejes). La escala de color varía entre 40 % (azul) y 100 % (rojo).

Se observa que el modelo que presentó mayor variación de las métricas, dados los parámetros, fue HMM, mientras que el más estable fue ML-SMV.

Las BoF presentaron una variabilidad mayor que ML-SMV, donde se observan zonas de picos de alto rendimiento y valles en las que el rendimiento desciende

principalmente en la métrica de sensibilidad (zonas de color naranja).

En los dos casos, se evidencian curvas de estabilización del modelo hacia los parámetros adecuados, lo que permite descartar fenómenos de sobreentrenamiento.

Por último, se resalta que los modelos ML-SMV y BoF mostraron resultados similares e incluso superiores a los reportados en el estado del arte, presentados en la sección 2 de este documento.

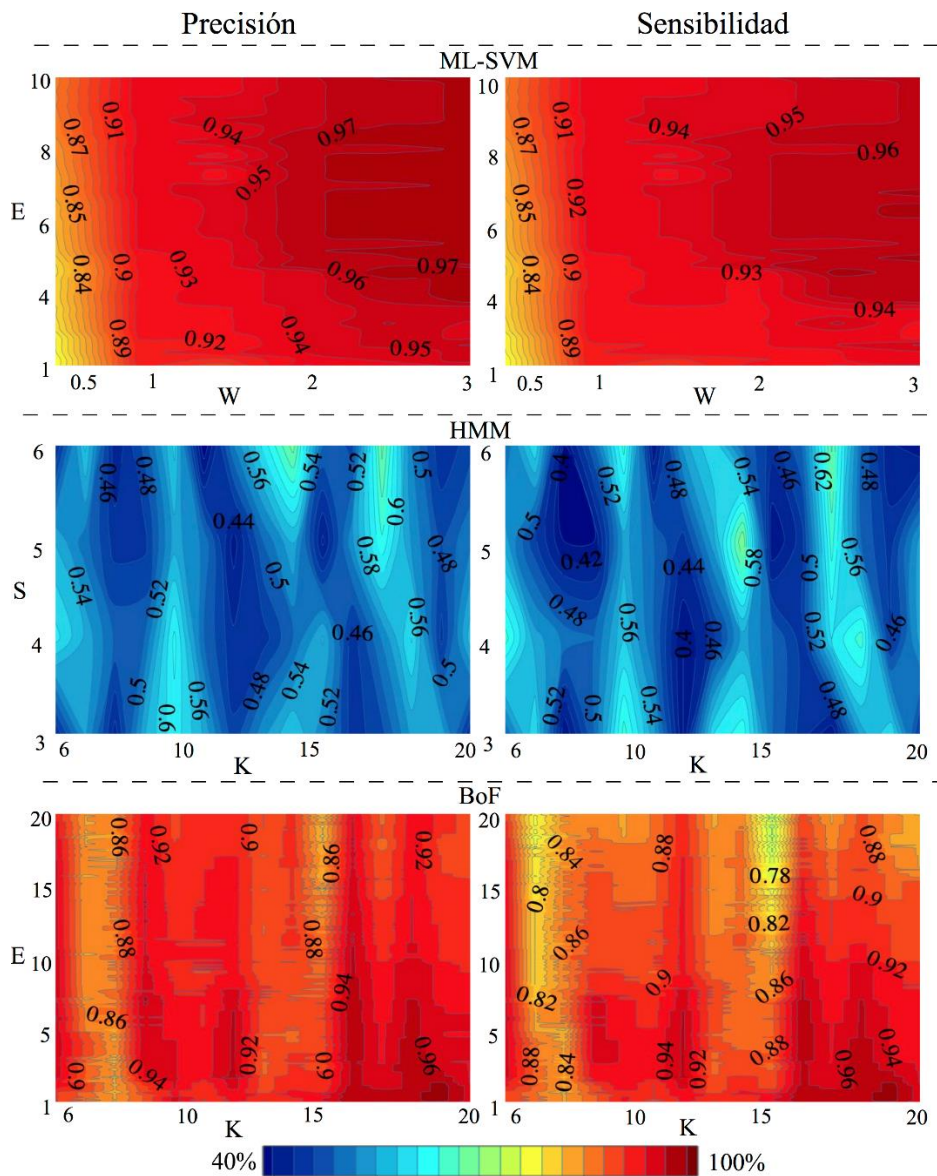


Fig. 4. Gráficas de sensibilidad y precisión para los tres modelos computacionales, en curvas de nivel. La escala de color varía entre 0% (azul) y 100% (rojo). Fuente: elaboración propia.

5. CONCLUSIONES Y TRABAJO FUTURO

Este artículo presentó una comparación de tres modelos computacionales para el reconocimiento de gestos: el reconocimiento de patrones usando máquinas de soporte vectorial (ML-SMV), los modelos ocultos de Markov (HMM) y las bolsas de características (BoF). La evaluación se hizo sobre un conjunto de ocho gestos para un usuario con limitaciones motoras, definido mediante diseño participativo.

Como dispositivo de adquisición se usó un brazalete MyO gesture, con el cual se adelantó un proceso de adquisición de 10 repeticiones para cada gesto.

Tras una evaluación de los modelos mediante una validación cruzada de diez paquetes (*10-folding Cross validation*), los resultados mostraron un mejor rendimiento de las BoF, respecto a las métricas de sensibilidad y precisión.

El modelo de ML-SMV expuso una mayor estabilidad respecto a los parámetros de estudio, mientras que los MHH generaron los resultados más discretos.

En general, los resultados de los tres modelos computacionales fueron adecuados para un conjunto de datos de 80 muestras (ocho gestos por diez repeticiones).

Las bolsas de características muestran resultados prometedores, de los cuales se puede colegir que, a futuro, es posible evaluar métodos de ajuste del diccionario de palabras, para seleccionar solo las más representativas para conjunto de gestos.

Adicionalmente, es necesario evaluar la usabilidad del conjunto de gestos, ya que, si bien es cierto que se hizo mediante un diseño participativo con el usuario interesado.

6. AGRADECIMIENTOS

Producto derivado del proyecto INV-ING-2996 financiado por Vicerrectoría de Investigaciones de la Universidad Militar Nueva Granada-Vigencia 2019.

7. REFERENCIAS

- [1] D. A. Norman, "The way I see it: Natural user interfaces are not natural," *interactions*, vol. 17, no. 3, pp. 6-10, May. 2010. <https://doi.org/10.1145/1744161.1744163>
- [2] A. Jaimes y N. Sebe, "Multimodal human-computer interaction: A survey," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 116-134, Oct. 2007. <https://doi.org/10.1016/j.cviu.2006.10.019>
- [3] C. Manresa, J. Varona, R. Mas, y F. J. Perales, "Hand Tracking and Gesture Recognition for Human-Computer Interaction", *ELCVIA Electron. Lett. Comput. Vis. Image Anal.*, vol. 5, no. 3, pp. 96-104, Nov. 2005. <https://doi.org/10.5565/rev/elcvia.109>
- [4] S. S. Rautaray y A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1-54, Jan. 2015. <https://doi.org/10.1007/s10462-012-9356-9>
- [5] P. K. Pisharady y M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Computer Vision and Image Understanding*, vol. 141, pp. 152-165, Dec. 2015. <https://doi.org/10.1016/j.cviu.2015.08.004>
- [6] M. J. Cheok, Z. Omar, y M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131-153, Jan. 2019. <https://doi.org/10.1007/s13042-017-0705-5>
- [7] H. Liu y L. Wang, "Gesture recognition for human-robot collaboration: A review", *Int. J. Ind. Ergon.*, vol. 68, pp. 355-367, Nov. 2018. <https://doi.org/10.1016/J.ERGON.2017.02.004>
- [8] H. Cheng, L. Yang, y Z. Liu, "Survey on 3D Hand Gesture Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659-1673, Sep. 2016. <https://doi.org/10.1109/TCSVT.2015.2469551>
- [9] D. J. Botina-Monsalve, M. A. Domínguez-Vásquez, C. A. Madrigal-González, y A. E. Castro-Ospina, "Automatic classification of vowels in Colombian sign language", *TecnoLógicas*, vol. 21, no. 41, pp. 103-114, Jan. 2018. <https://doi.org/10.22430/22565337.730>
- [10] Kui Liu, C. Chen, R. Jafari, y N. Kehtarnavaz, "Fusion of Inertial and Depth Sensor Data for Robust Hand Gesture Recognition," *IEEE Sensors Journal*, vol. 14,

- no. 6, pp. 1898–1903, Jun. 2014.
<https://doi.org/10.1109/JSEN.2014.2306094>
- [11] C. Chen, R. Jafari, y N. Kehtarnavaz, “A survey of depth and inertial sensor fusion for human action recognition,” *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 4405–4425, Feb. 2017.
<https://doi.org/10.1007/s11042-015-3177-1>
- [12] R. Xu, S. Zhou, y W. J. Li, “MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition,” *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1166–1173, May. 2012.
<https://doi.org/10.1109/JSEN.2011.2166953>
- [13] M. B. I. Raez, M. S. Hussain, y F. Mohd-Yasin, “Techniques of EMG signal analysis: detection, processing, classification and applications”, *Biological procedures online*, vol. 8, pp. 11–35, 2006.
<https://doi.org/10.1251/bpo115>
- [14] Y. Du, W. Jin, W. Wei, Y. Hu, y W. Geng, “Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation,” *Sensors*, vol. 17, no. 3, pp. 458, Feb. 2017.
<https://doi.org/10.3390/s17030458>
- [15] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, y A. Skodras, “Hand Gesture Recognition Based on EMG Data: A Convolutional Neural Network Approach,” in *International Conference on Physiological Computing Systems*, Sevilla, 2019, pp. 180–197.
https://doi.org/10.1007/978-3-030-27950-9_10
- [16] W.-T. Shi, Z.-J. Lyu, S.-T. Tang, T.-L. Chia, y C.-Y. Yang, “A bionic hand controlled by hand gesture recognition based on surface EMG signals: A preliminary study,” *Biocybernetics and Biomedical Engineering*, vol. 38, no. 1, pp. 126–135, Nov. 2018.
<https://doi.org/10.1016/J.BBE.2017.11.001>
- [17] A. S. Kundu, O. Mazumder, P. K. Lenka, y S. Bhaumik, “Hand Gesture Recognition Based Omnidirectional Wheelchair Control Using IMU and EMG Sensors,” *Journal of Intelligent & Robotic Systems*, vol. 91, no. 3–4, pp. 529–541, Sep. 2018.
<https://doi.org/10.1007/s10846-017-0725-0>
- [18] S. Benatti *et al.*, “A sub-10mW real-time implementation for EMG hand gesture recognition based on a multi-core biomedical SoC,” in *2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*, Vieste 2017, pp. 139–144.
<https://doi.org/10.1109/IWASI.2017.7974234>
- [19] M. Rossi, S. Benatti, E. Farella, y L. Benini, “Hybrid EMG classifier based on HMM and SVM for hand gesture recognition in prosthetics,” en *2015 IEEE International Conference on Industrial Technology (ICIT)*, Seville, 2015, pp. 1700–1705.
<https://doi.org/10.1109/ICIT.2015.7125342>
- [20] A. Rahimi, S. Benatti, P. Kanerva, L. Benini, y J. M. Rabaey, “Hyperdimensional biosignal processing: A case study for EMG-based hand gesture recognition,” en *2016 IEEE International Conference on Rebooting Computing (ICRC)*, San Diego, CA, 2016, pp. 1–8.
<https://doi.org/10.1109/ICRC.2016.7738683>
- [21] J. Nielsen, “Finding usability problems through heuristic evaluation,” en *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, California, 1992, pp. 373–380.
<https://doi.org/10.1145/142750.142834>
- [22] G. Naik, D. Kumar y Jayadeva, “Twin SMV for Gesture Classification Using the Surface Electromyogram”, *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 301–308, Mar. 2010.
<https://doi.org/10.1109/TITB.2009.2037752>
- [23] F.-S. Chen, C.-M. Fu, y C.-L. Huang, “Hand gesture recognition using a real-time tracking method and hidden Markov models,” *Image and Vision Computing*, vol. 21, no. 8, pp. 745–758, Aug. 2003.
[https://doi.org/10.1016/S0262-8856\(03\)00070-2](https://doi.org/10.1016/S0262-8856(03)00070-2)
- [24] I.-J. Ding y Y.-J. Chang, “HMM with improved feature extraction-based feature parameters for identity recognition of gesture command operators by using a sensed Kinect-data stream,” *Neurocomputing*, vol. 262, pp. 108–119, Nov. 2017.
<https://doi.org/10.1016/J.NEUCOM.2016.11.089>
- [25] N. H. Dardas y N. D. Georganas, “Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques,” *IEEE Trans. Instrum. Measurement*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.
<https://doi.org/10.1109/TIM.2011.2161140>
- [26] N. Dardas, Q. Chen, N. D. Georganas, y E. M. Petriu, “Hand gesture recognition using Bag-of-features and multi-class Support Vector Machine,” en *2010 IEEE International Symposium on Haptic Audio Visual Environments and Games*, Phoenix, 2010, pp. 1–5.
<https://doi.org/10.1109/HAVE.2010.5623982>
- [27] M. A. Oskoei y H. Hu, “Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1956–1965, Aug. 2008.
<https://doi.org/10.1109/TBME.2008.919734>

- [28] M. Leon, J. M. Gutierrez, L. Leija, y R. Munoz, "EMG pattern recognition using Support Vector Machines classifier for myoelectric control purposes," en *2011 Pan American Health Care Exchanges*, Rio de Janeiro, 2011, pp. 175–178.
<https://doi.org/10.1109/PAHCE.2011.5871873>
- [29] N. Wang, Y. Chen, y X. Zhang, "The recognition of multi-finger prehensile postures using LDA," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 706–712, Nov. 2013.
<https://doi.org/10.1016/j.bspc.2013.06.006>
- [30] G. D. F. Ulloa, N. Sreenivasan, P. Bifulco, M. Cesarelli, G. Gargiulo, y U. Gunawardana, "Cost effective electro — Resistive band based myo activated prosthetic upper limb for amputees in the developing world," en *2017 IEEE Life Sciences Conference (LSC)*, Sydney, 2017, pp. 250–253.
<https://doi.org/10.1109/LSC.2017.8268190>
- [31] C. Savur y F. Sahin, "American Sign Language Recognition system by using surface EMG signal," *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, 2016, pp.1-6.
<https://doi.org/10.1109/SMC.2016.7844675>
- [32] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
<https://doi.org/10.1109/5.18626>
- [33] A. Cohen, "Hidden Markov models in biomedical signal processing," en *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, Hong Kong, 1998, pp. 1145-1150.
<https://doi.org/10.1109/IEMBS.1998.747073>
- [34] D. Burton, J. Shore, y J. Buck, "A generalization of isolated word recognition using vector quantization," *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Boston, 1983, pp. 1021–1024.
<https://doi.org/10.1109/ICASSP.1983.1171915>
- [35] E. Nowak, F. Jurie, y B. Triggs, "Sampling Strategies for Bag-of-Features Image Classification," en *European Conference on Computer Vision— ECCV 2006*, Berlin-Heidelberg, 2006, pp. 490–503.
https://doi.org/10.1007/11744085_38