

UNA BREVE REVISION SOBRE REDES NEURONALES ARTIFICIALES APLICADAS AL RECONOCIMIENTO GESTUAL

Ríos Ortiz, Brian Alejandro; Márquez Gutiérrez, Pedro Rafael
Tecnológico Nacional de México – Tecnológico de Chihuahua
División de Estudios de Posgrado e Investigación
Avenida Tecnológico 2909, Chihuahua, Chih.
Tel. 614-201-2000
{barrios,pmarquez}@itch.edu.mx

RESUMEN.

Hoy en día el reconocimiento gestual está presente en muchos ámbitos, desde manejar robots con el movimiento de la mano, hasta ámbitos más difíciles como lo es el reconocimiento gestual con propósito de traducir los diferentes lenguajes de señas que existen en el mundo. El uso de algoritmos de Machine Learning puede ayudar de gran manera al reconocimiento gestual para diversos propósitos.

En el presente trabajo se presenta una breve revisión de literatura acerca del uso de redes neuronales artificiales aplicadas al reconocimiento gestual.

Palabras Clave: Reconocimiento Gestual, Machine Learning, Neural Networks, Lenguaje de Señas.

ABSTRACT.

Nowaday, gestural recognition is present in many areas, ranging from handling robots with the movement of the hand, to more difficult areas such as gestural recognition in order to translate the different sign languages that exist in the world. The use of Machine Learning algorithms can greatly help gesture recognition for various challenges

Keywords: Gestural Recognition, Machine Learning, Neural Networks, Sign Language

1. INTRODUCCIÓN

El problema de reconocimiento gestual se aborda desde tres principales enfoques (Fig. 1). El primero basado en visión, generalmente consiste en cámaras para obtener datos acerca del gesto que se está realizando. El segundo basado en sensores, en los que con distintos sensores corporales se busca obtener los datos de algún gesto que se esté realizando en el momento. Y por último una combinación de las anteriores.

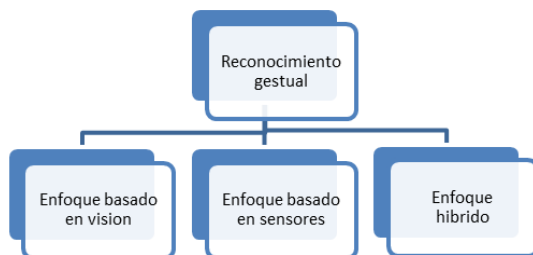


Fig. 1. Enfoques utilizados en el reconocimiento gestual.

El verdadero problema del reconocimiento gestual consiste en determinar que método se va a utilizar para hacer el procesamiento de los datos que se obtienen con cualquier enfoque que se tome.

En este trabajo se revisan trabajos representativos relacionados con el uso de algoritmos de Machine Learning para caracterizar y reconocer el gesto que se está realizando, y evaluar cualitativamente su aplicabilidad.

2. ANÁLISIS DE INVESTIGACIONES RELACIONADAS

2.1. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition [1]

En esta investigación se propone una arquitectura híbrida que usa redes neuronales convolucionales y redes neuronales recurrentes, para una mejor captura de las cualidades temporales de la señal electromiográfica.

1. Relación con el tema de investigación.

- Utiliza sensores electromiográficos de superficie.
- Se usan redes neuronales para reconocimiento gestual.

2. Funcionamiento

La arquitectura propuesta consiste en un modelo híbrido en que implementa CNN y RNN. El modelo CNN consiste de siete capas:

- Las primeras dos capas son capas convolucionales.
- Las siguientes dos capas son capas localmente conectadas.
- Las últimas tres capas son totalmente conectadas.

La arquitectura propuesta se muestra en la Fig. 2.

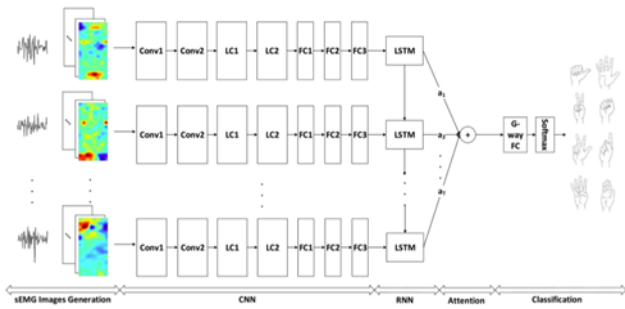


Fig. 2. Arquitectura CNN-RNN propuesta.

El modelo RNN contiene lazos de retroalimentación que permiten obtener información de una secuencia temporal. Dadas las secuencias de entrada $\{F_1, F_2, \dots, F_T\}$ (vectores de característica extraídos del modelo CNN), los estados ocultos h_t y las salidas y_t se calculan de la siguiente manera:

$$h_t = H(W_{ih} + W_{hh}h_{t-1} + b_h)$$

$$y_t = W_{ho}h_t + b_o$$

donde W_{ih}, W_{hh}, W_{ho} son las matrices de pesos entre entradas.

3. Resultados obtenidos

Los autores prueban la arquitectura que proponen con distintas bases de datos que contienen gestos ya definidos, además comparan su modelo con diferentes modelos ya implementados. La Fig. 3 muestra una tabla que presenta la efectividad de la clasificación de cada gesto comparada con otros trabajos.

	NinaProDB1			NinaProDB2		BioPatRec26MOV			CappMyo-DBa			csl-hdemg		
	150ms	200ms	Trial	200ms	Trial	50ms	150ms	Trial	40ms	150ms	Trial	150ms	170ms	Trial
Feature-LDA [15]	-	-	-	-	-	86.3%	92.9%	-	-	99.0%	-	-	-	-
Traditional-RF [55]	-	75.3%	-	-	-	-	-	-	-	-	-	-	-	-
AtzoriNet [25]	-	66.6%	-	-	-	-	-	-	-	-	-	-	-	-
GengNet [26]	-	77.8%	96.7%	-	-	-	-	-	99.0%	99.5%	-	89.3%	90.4%	96.8%
ZhaiNet [29]	-	-	-	78.71%	-	-	-	-	-	-	-	-	-	-
RNN Module with raw-signal	78.1%	79.8%	95.0%	Did not converge		76.4%	82.3%	92.6%	71.8%	80.4%	90.4%	65.3%	71.1%	75.8%
CNN Module with raw-image1	82.6%	83.5%	96.5%	73.4%	97.6%	82.1%	83.9%	92.2%	98.0%	97.7%	98.9%	92.0%	92.1%	95.2%
CNN Module with feature-signal-image1	85.4%	86.3%	97.2%	81.4%	97.5%	85.2%	90.0%	95.8%	-	-	-	-	-	-
Hybrid CNN-RNN with raw-image1	83.5%	84.7%	96.5%	74.6%	97.7%	88.5%	92.2%	96.8%	99.1%	99.6%	99.9%	94.3%	94.8%	96.1%
Hybrid CNN-RNN with feature-signal-image1	86.4%	86.7%	97.1%	82.0%	97.5%	89.9%	93.9%	97.5%	-	-	-	-	-	-
Attention-based hybrid CNN-RNN with raw-image1	83.7%	84.8%	96.5%	74.8%	97.6%	88.7%	92.5%	96.8%	99.3%	99.7%	99.9%	94.5%	94.9%	96.1%
Attention-based hybrid CNN-RNN with feature-signal-image1	86.8%	87.0%	97.3%	82.2%	97.6%	90.0%	94.1%	97.7%	-	-	-	-	-	-

Fig. 1. Tabla de comparación de resultados.

2.2 Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching [2]

Los autores proponen un lápiz basado en un acelerómetro con los cuales se pretende clasificar un total de 24 gestos que incluyen gestos básicos y complejos.

1. Relación con la investigación

- Utiliza un acelerómetro para reconocer los movimientos dinámicos
- El procesamiento de los datos se hace con una red neuronal

2. Funcionamiento

El algoritmo del sistema de reconocimiento gestual que se propone se puede observar en el diagrama de bloques que se presenta en la Fig. 2.

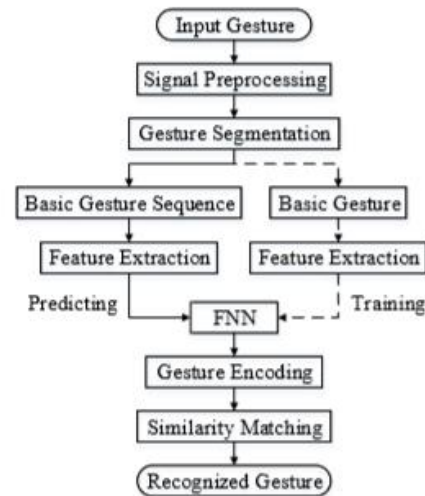


Fig. 2. Diagrama a bloques del sistema.

Lo realmente interesante es el sistema de clasificación de gestos, el cual está basada en redes neuronales tipo feed-forward. Para hacer la clasificación de los gestos los autores desarrollaron una FNN de tres capas, que consiste de una capa de entrada, una capa oculta y por último una capa de salida. La FNN desarrollada por los autores puede observarse en la Fig. 5

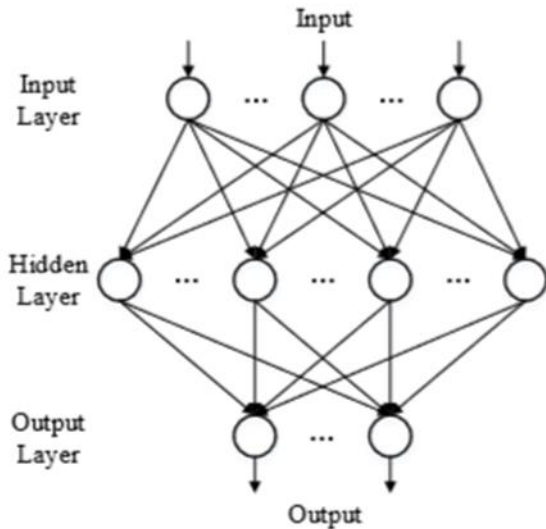


Fig. 5. FNN desarrollada.

Suponiendo que N_i , N_h y N_o son el número de nodos en las capas de entrada, oculta y salida respectivamente, la salida del k-ésimo nodo en la capa de salida está dado por:

$$y_k = f(0_k + \sum_{h=1}^{N_h} w_{kh} f(\lambda_h + \sum_{i=1}^{N_i} v_{hi} x_i)$$

En donde:

x_i = Es la entrada del i-ésimo nodo en la capa de entrada.

v_{hi} = peso de acoplamiento entre capas de entrada y oculta.

λ_h y 0_k = bias.

3. Resultados obtenidos

Los autores presentan experimentos que presentan una efectividad de hasta el 99.88 % lo cual es muy alta en comparación con otros sistemas propuestos.

El problema con el sistema que ellos proponen es que se dedica exclusivamente a identificar gestos ya definidos, además que los gestos que se hacen dependen exclusivamente del movimiento generado, por lo que, si se quisiera usar para hacer un reconocimiento gestual para el lenguaje de señas, se vería limitado.

Por otro lado, si se desea detectar movimiento el sistema propuesto sería una buena opción para detectar los gestos realizados.

2.2. Hand Body Language Gesture Recognition Based on Signals from Specialized Glove and Machine Learning Algorithms [3].

El sistema descrito busca crear una interface hombre-máquina. Por lo que se presenta un sistema para reconocer gestos de una manera rápida y efectiva basado en datos que se obtienen de un guante equipado con 10 sensores.

Los autores diseñan tres algoritmos de *machine learning* basados en clasificadores.

1. Relación con el tema de tesis

- Sistema basado en un guante diseñado específicamente para reconocimiento gestual.
- El desafío principal de los autores consiste en hacer un reconocimiento gestual exitoso que pretenden hacer con algoritmos de *machine learning*.

2. Funcionamiento

El algoritmo que los autores siguen se muestra en la **Error! Reference source not found.** Se puede observar que los autores se centran mucho en la parte de adquisición y pre-procesamiento de los datos.

El análisis se realizó como consecuencia del procesamiento de datos. Su objetivo era el reconocimiento de los tipos de gestos basados en muestras (mediciones). Este análisis se basó en el diseño y la selección de los parámetros apropiados.

Los autores aplican distintos métodos de *machine learning*:

- Redes neuronales probabilísticas
- Máquina de vector de soporte
- *k-nearestneighbor (kNN)*

Cabe mencionar que, aunque los autores mencionan que se iban a desarrollar los algoritmos anteriormente mencionados solo se limitan a implementar métodos que ya se habían implementado anteriormente.

3. Resultados obtenidos

De acuerdo a los experimentos realizados por los autores se demuestra que se encontró el menor número de errores de reconocimiento para los gestos que involucraron solo uno o dos dedos y el movimiento de la mano no fue complicado. En gestos que implicaban el movimiento de la mayoría de los dedos el número de aciertos disminuyó, sin embargo se menciona que el rango de efectividad es mayor al 99.5 %.

2.4 SkinGest: artificial skin for gesture recognition via filmy stretchable strain sensors [4]

Aquí se propone un novedoso método de reconocimiento gestual que consiste en diseñar una piel artificial con sensores elásticos y sensibles.

1. Relación con nuestra investigación

- Reconocimiento de lenguaje de señas.
- El reconocimiento de la seña se hace en base a algoritmos de *machine learning*.

2. Funcionamiento

Después de obtener los datos con el novedoso sistema que los autores pretenden implementar, se diseñaron diferentes gestos que luego son realizados para reconocerlos.

Los valores de voltaje que se obtienen tras las mediciones de cada gesto transforman la resistencia del sensor como una característica debido a que la resistencia tiene una relación relativamente lineal con el estiramiento del sensor, el cual está directamente relacionado con el movimiento del dedo. Para reconocer el gesto se empleó *linear discriminant analysis* (LDA), *K nearest neighbor* (KNN) y máquina de vectores de soporte (SVM). Todos los algoritmos se realizaron con MATLAB.

Los gestos que se proponían utilizar son gestos intuitivos y de uso común. Sin embargo, no se menciona cuáles fueron los gestos a reconocer.

Aunque se implementaron algoritmos de *machine learning*, los autores no proporcionan información de cómo realizaron la implementación de dichos algoritmos.

El algoritmo general del sistema se muestra en la Fig. 3.

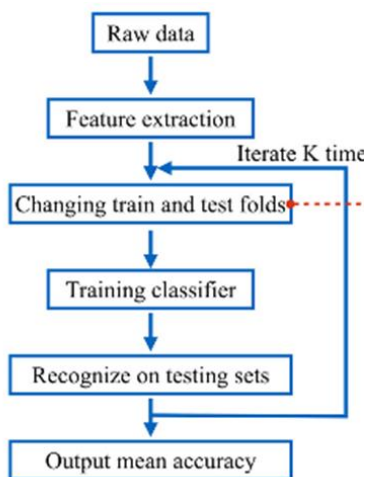


Fig. 3. Diagrama de flujo general del sistema.

3. Resultados obtenidos

Se proporciona una gráfica que muestra los porcentajes de efectividad de los diferentes clasificadores usados. La gráfica en cuestión se muestra en la **Error! Reference source not found.**

Si bien los porcentajes de acierto son muy altos, no se menciona cuantas pruebas se hicieron, ni cuantos gestos el sistema puede reconocer.

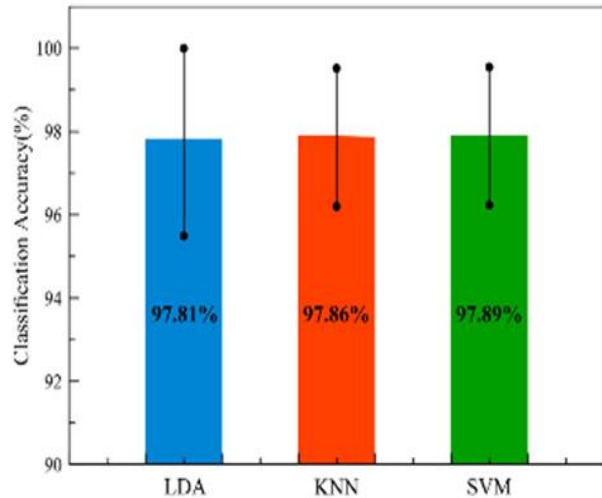


Fig. 8 Porcentaje de acierto de los diferentes clasificadores utilizados.

2.5 Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker [5].

Los autores de este artículo pretendían hacer reconocimiento gestual del lenguaje de señas Americano, y los gestos reconocidos correspondían a palabras que posteriormente eran convertidas a voz.

1. Relación con el tema de investigación

- Los datos se obtienen a partir de un guante con diferentes sensores.
- El reconocimiento de los gestos se hace con una red neuronal artificial.

2. Funcionamiento

La Fig. 9 muestra el diagrama de bloques del sistema.

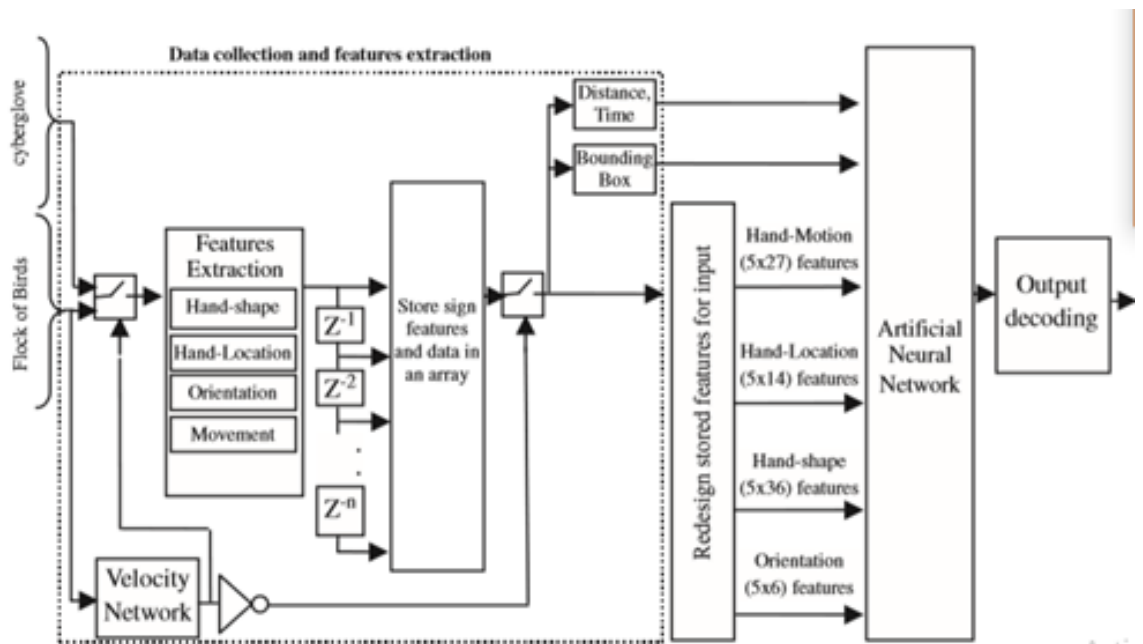


Fig. 9. Diagrama de bloques del sistema de reconocimiento de palabras.

Se puede observar que se extraen distintas características del gesto como lo son la posición manual, la localización de la mano y la orientación de la misma. Para cada una de las características extraídas se desarrolla una ANN. El número de neuronas de entrada depende de los sensores que intervienen en cada característica y el número de neuronas de salida depende del número de particiones que se desea obtener.

Para la posición manual se tiene una ANN que consiste de 15 neuronas de entrada (una neurona por cada sensor del guante utilizado), 20 neuronas ocultas y 36 neuronas de salida cada una representando a un gesto específico.

Los pesos iniciales y los *bias* se seleccionaron aleatoriamente. La red neuronal tenía salidas de tipo sigmoide, generando que la salida se mantuviera entre 0 y 1.

La red neuronal artificial desarrollada genera un vector de 36 elementos. De estos 36 elementos todos los valores eran 0 excepto uno, el cual correspondía a uno de los gestos que se quería identificar (Fig. 10).



Fig. 10. Gestos a reconocer del sistema implementado.

3. Resultados obtenidos

Para la parte específica en la que se aplica la red neuronal artificial los autores reportan un porcentaje de efectividad de 98%. Además los autores hacen el comentario de que el sistema reconocerá los gestos de personas que tenga la mano del mismo tamaño que las manos que fueron utilizadas para el entrenamiento, por lo que dan al usuario la facilidad de entrenar al sistema de acuerdo a lo que ellos requieran.

2.3. Conclusiones.

Las redes neuronales y los algoritmos de *Machine Learning* se aplican de gran manera en el reconocimiento de gestos, sin importar de que naturaleza sean los datos obtenidos (obtenidos por visión, sensores, o ambos).

Generalmente en toda la literatura revisada los temas vistos se orientan a hacer la clasificación de los gestos, por lo que eso indica que en nuestra investigación se está tomando un buen

camino, ya que se pretende utilizar para poder hacer el reconocimiento de los gestos del Lenguaje Mexicano de Señas, que posteriormente se utilizaran para generar lenguaje hablado.

Con base en la literatura revisada se amplía el panorama de como la inteligencia artificial se puede aplicar para poder reconocer los gestos, i.e., interpretarlos adecuadamente.

Los temas vistos también pueden servir para clasificar las características de acuerdo al modelo que se está manejando en nuestra investigación. Por ejemplo, se puede clasificar la posición digital, además de la orientación de la mano, en base a criterios ya establecidos por el modelo que se está manejando.

2.5 Referencias

- [1] Y. Hu, Y. Wong, W. Wentao, Y. Du, M. Kankanhalli y W. Geng, «A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition,» Plos ONE, vol. 13, n° 10, 2018.
- [2] R. Xie y J. Cao, «Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching,» IEEE Sensors Journal, vol. 11, n° 4, 2012.
- [3] P. Plawiak, T. Sosnicki, M. Niedzwiecki, Z. Tabor y K. Rzecki, «Hand Body Language Gesture Recognition Based on Signals From Specialized Glove and Machine Learning Algorithms,» IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, 2015.
- [4] «SkinGest: artificial skin for gesture recognition via filmy stretchable strain sensors,» Advanced Robotics, vol. 32, n° 21, pp. 1112-1121, 2018.
- [5] C. Oz y M. Leu, «Linguistic properties based on American Sign Language isolated word Recognition with artificial neural networks using a sensory glove and motion tracker,» Neurocomputing, n° 70, 2007.