

# Diseño de un Sistema Clasificador de las Posturas que Toma un Presentador, Utilizando Kinect y Redes Neuronales

Juan José García<sup>1</sup> Alex Ferrín<sup>2</sup> Edgar Villaceca<sup>3</sup> Daniel Izquierdo<sup>4</sup>

**Abstract**—This research aims to classify the postures of a presenter. To do so, kinect is used to measure the postures, and artificial neural networks to classify them. The posture is represented as a ten dimensional feature vector, comprised of the angle formed by the articulations of the presenter. This accounts for the lack of normalization found in previous work. Finally, a statistical analysis assures that the neural network will classify the postures of a presenter with a probability greater than 61%.

## I. INTRODUCCIÓN

La habilidad de hacer presentaciones a un grupo de personas es una de las competencias profesionales demandadas actualmente por empresas, negocios y organizaciones profesionales. Para poder evaluar esta habilidad es necesario definir variables que caractericen una buena presentación. [15] considera la postura durante la presentación, una métrica representativa de la calidad del presentador. Siendo esta una combinación específica de posición, orientación y flexión [6]. En [1-16] se determinan con una sola imagen o un set de imágenes en un intervalo de tiempo.

Kinect es un sensor de movimiento capaz de detectar la posición espacial de 20 articulaciones en un instante de tiempo. Una red neuronal es un sistema capaz de aprender a clasificar información desconocida en patrones conocidos. Las redes neuronales son útiles para definir reglas de clasificación que no son aparentes para el programador (la relación entre puntos en el espacio y posturas es un ejemplo).

Estas herramientas han sido utilizadas para clasificar posturas en varios contextos. En [1] las posturas están asociadas a fisioterapia, en [2] a problemas musculares, en [4] y [7] a las posturas tomadas al ver televisión, en [8] se implementa un Sistema que detecta intrusos y en [10] se utilizó para analizar posturas de un profesor impartiendo clases. Sin embargo, [15] es el único estudio aplicado al contexto de las presentaciones con un porcentaje de clasificación del 61%; relativamente bajo en comparación con [1, 2, 4, 11, 16], donde el promedio es del 90%.

\*This work was not supported by any organization

<sup>1</sup>J. García es de la Facultad de Ingeniería Eléctrica y de Computación de la Escuela Superior Politécnica del Litoral, Guayaquil, Guayas, Ecuador [juajogar@espol.edu.ec](mailto:juajogar@espol.edu.ec)

<sup>2</sup>A. Ferrín es de la Facultad de Ingeniería Eléctrica y de Computación de la Escuela Superior Politécnica del Litoral, Guayaquil, Guayas, Ecuador [aferrin@espol.edu.ec](mailto:aferrin@espol.edu.ec)

<sup>3</sup>E. Villaceca es de la Facultad de Ingeniería Eléctrica y de Computación de la Escuela Superior Politécnica del Litoral, Guayaquil, Guayas, Ecuador [esvillac@espol.edu.ec](mailto:esvillac@espol.edu.ec)

<sup>4</sup>D. Izquierdo es de la Facultad de Ingeniería Eléctrica y de Computación de la Escuela Superior Politécnica del Litoral, Guayaquil, Guayas, Ecuador [djizquie@espol.edu.ec](mailto:djizquie@espol.edu.ec)

Esta investigación tiene como objetivo clasificar las posturas de un presentador con un porcentaje de clasificación mayor al 61%. Para esto es necesario realizar la detección de la postura y la catalogación de la misma, empleando Kinect y redes neuronales, respectivamente.

Este documento está estructurado de la siguiente manera: La sección II presenta los trabajos relacionados con clasificación de posturas en otros ámbitos, utilizando cualquier técnica de machine learning para clasificar; que fueron útiles en este trabajo. En la sección III se describe la clasificación manual que se hizo, la representación de los datos para entrenar la red neuronal y el entrenamiento de la misma. La sección IV expone los resultados del entrenamiento de las redes neuronales y sus respectivas estadísticas. En la sección V se analizan los resultados de la sección anterior y en la sección VI se plantean ideas para trabajos futuros.

## II. TRABAJOS RELACIONADOS

El diseño de un clasificador de posturas consta de teorías desarrolladas, y nuestro diseño buscará adaptar una de estas teorías al contexto de las presentaciones orales. [1, 2, 4, 5, 7, 10] adaptan diseños de clasificadores previos a nuevos contextos.

Los clasificadores revisados hacen uso de Kinect y un sistema de clasificación. [1, 2, 3, 7, 8, 11, 16] utilizan el Kinect para la captura de los datos. Debido al bajo costo del kinect en comparación a las cámaras en 3D, la disponibilidad de kits de desarrollo de software, y la facilidad de programación e integración, hacen que Kinect sea ideal como instrumento de medición. [3, 8]. Existen varios sistemas de clasificación como SVM, kNN [2, 15], LB, LGT y J48 [4], sin embargo, en [2, 4], las redes neuronales presentan la mejor probabilidad de clasificación en comparación a estos.

El diseño del clasificador consta de dos partes: el vector de entrada o vector característico de las posturas, y el diseño de la red neuronal. Las posturas a clasificar deben ser en primera instancia, definidas. En el caso de las redes neuronales, esta definición se conoce como feature vector o vector de entrada, y es la información que la red neuronal clasificará en una postura. Es muy importante que este vector sea representativo de la postura, y diferente a vectores de otras posturas.

En [1, 2, 6-8, 10, 12] se definen vectores de entrada que en promedio pertenecen a R12. El planteamiento del vector característico de [1, 2, 11, 15] es análogo a nuestro planteamiento. Estas investigaciones proponen esquemas comunes para representar una postura. Entre las variables están: Distancias entre las juntas de las extremidades, y ángulos que forman las juntas respecto a una referencia.

Para el diseño de la red neuronal, [1, 2, 5, 6] utilizan el algoritmo de Levenberg-Marquard como algoritmo de aprendizaje; [1] utiliza la función de activación log-sigmoid; y en [1],[2], [5], [9], [12] y [14] la red neuronal consta de tres capas.

### III. METODOLOGÍA

#### A. Clasificación manual de los datos

El set de datos usados para el entrenamiento y prueba de las redes neuronales se obtuvieron de los datos usados en [15], que representan los datos del Kinect las posturas por frame durante la presentación de estudiantes. Se clasificó aproximadamente 16 estudiantes. Estos datos fueron después procesados por un script hecho en Python, en el cual se hacía una gráfica de puntos de la postura, la cual se procedía a clasificar manualmente de acuerdo a un subgrupo de 9 posturas que se nos proporcionó para la clasificación. Luego, los datos eran pasados por otro script en Python, que dividía los datos, tomando un frame por timestamp y preparaba los datos para la red neuronal.

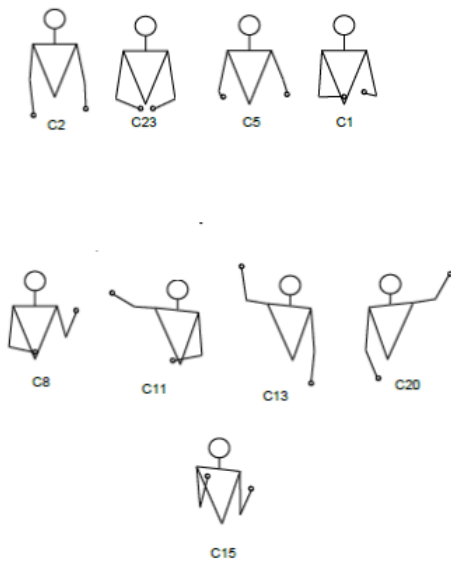


Fig. 1. Subconjuntos de posturas clasificadas.

#### B. Preparación de los datos para la red neuronal.

Uno de los problemas observados en [15] es que las mediciones no fueron normalizadas. Según [2] la normalización permite que las predicciones sean independientes al sujeto medido. Por esta razón, utilizamos el vector característico de [1] como guía para entrenar la red neuronal. El vector característico propuesto es de dimensión diez, compuesto de los ángulos que forman los vectores entre juntas, respecto a una base particular. Cada una de estas bases es independiente de la otra, asegurando independencia entre ángulos de vectores distintos.

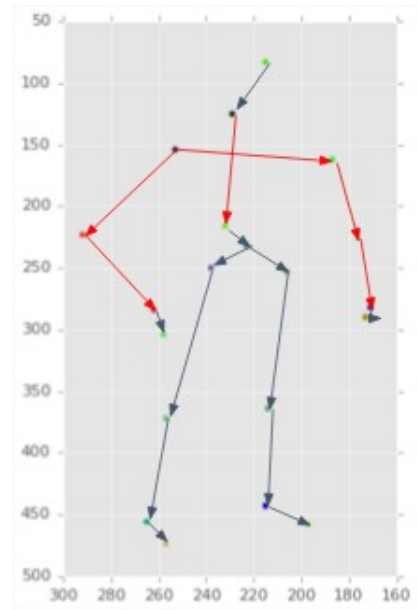


Fig. 2. Representación de las posturas mediante vectores, para ser convertidos a ángulos.

Un script en python realiza la conversión de los puntos, a vectores y finalmente a ángulos. En total trabajamos con una muestra de 2164 posturas.

#### C. Entrenamiento y prueba de la red neuronal

Utilizamos el paquete "neuralnets" en R para entrenar la red neuronal. Decidimos crear 9 redes neuronales, una para cada postura, y entrenarlas en distintas modalidades. Cada modalidad varía: la muestra con la que se entrenaba a la red neuronal, la muestra de prueba de la red neuronal, la cantidad de neuronas en el "hidden layer", y la predicción de la red.

La muestra de entrenamiento se compone de dos grupos: El primero contiene muestras donde aparece la postura que clasifica la red, y el segundo grupo contiene muestras donde no aparece la postura que clasifica la red. El objetivo de esto es facilitar la clasificación de muestras que pertenezcan o no pertenezcan a la postura.

La muestra de prueba es un subconjunto de todas las mediciones, escogido aleatoriamente.

La cantidad de neuronas en el hidden layer afecta: Tiempo de entrenamiento de la red neuronal, y error de la red neuronal.

La predicción de la red neuronal está asociado a la función de activación de las neuronas de la salida. Para evitar los saltos repentinos de una función de activación lineal, utilizamos la función de activación logística. Esto implica que la red predice un número real entre uno y cero. Para asociar este número a nuestro esquema de clasificación se establece un límite; todos los reales mayores a este límite son uno, y el resto son cero. Este límite es variado para controlar la predicción de la red.

#### IV. RESULTADOS

La figura tres indica: Las características de la modalidad de una red neuronal, y el porcentaje de clasificación medido.

##### A. Resultados de la Redes Neuronales

correct_size	varied_size	hidden	test_size	threshold	Average
40	30	9	30	1	68.15%
25	25	9	50	0.85	67.00%
40	50	11	30	0.85	65.00%
20	20	3	50	0.7	61.50%
35	20	7	50	1	60.67%
35	35	9	35	0.85	72.38%
30	30	9	20	0.85	63.33%
35	35	9	35	0.85	69.20%
40	35	9	40	0.9	64.72%
40	40	11	45	0.85	72.34%
40	60	10	50	0.85	73.11%
40	70	10	60	0.85	72.03%
40	20	9	50	0.85	58.20%
40	150	11	50	0.8	58.20%

Fig. 3. Porcentajes de clasificación promedio, de diferentes modalidades de redes neuronales

Podemos observar que el porcentaje de clasificación varía en gran medida debido a: el número de neuronas en el "hidden layer"(hidden); el límite para decidir cuando una postura es clasificada o no (threshold); y el diseño de la muestra para entrenar la red neuronal (correct size, varied size).

##### B. Resultados Estadísticos

Se tomaron 14 modelos de redes neuronales para el análisis estadístico. Haciendo un análisis descriptivo, obtenemos que la mediana de clasificación es del 66%, lo que implica que la mitad de las clasificaciones está por encima del mínimo de clasificación esperado. Lo anterior va de acuerdo con el promedio de clasificación, que es del 66,13%, desviados en promedio de ese dato 5.288 unidades de la escala según la desviación estándar.

La hipótesis a ser puesta a prueba es: Las redes neuronales en general presentan un mayor porcentaje de clasificación que kNN para clasificación de posturas tomadas por estudiantes al momento de realizar presentaciones

Para el análisis estadístico, se usó la prueba t de student, t-simple para ser específicos, donde el valor a contrastar con nuestra hipótesis será el porcentaje de clasificación mínimo obtenido por kNN en [15], siendo este 61%. Antes de entrar a la prueba de hipótesis, se ejecutó una prueba Shapiro-Wilk de normalidad, para usar la asunción que los los porcentajes de clasificación estaban distribuidos normalmente, obteniendo un valor del estadístico de  $w = 0.92396$  y un valor p de 0.2506. Al ser el valor p mayor a 0.05, no se rechaza la hipótesis nula y podemos asumir que los porcentajes están normalmente distribuidos. Adicional a esta prueba, se realizó una prueba de Kolmogorov-Smirnov para justificar aún más nuestra asunción de normalidad, donde se obtuvo un valor del estadístico de  $D = 0.15342$  y un valor p de 0.8967, haciendo más sólido nuestro supuesto de normalidad. Adicional a ambos análisis se procedió a trazar una gráfica Q-Q Normal, para contrastar la muestra con una distribución

normal estándar, se puede apreciar que los datos presentan en buena medida una distribución normal.

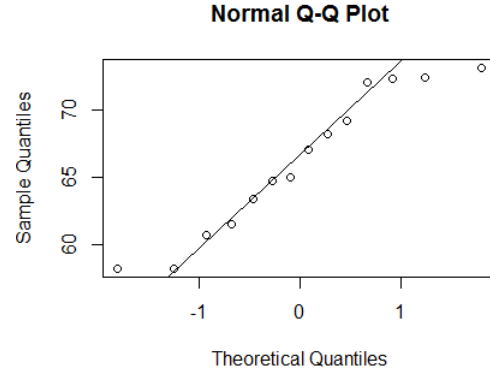


Fig. 4. Gráfica Q-Q Normal.

Habiendo ejecutado la prueba t simple, se observaron los siguientes resultados: Un valor del estadístico t de 3.6302, con 13 grados de libertad, un valor p del 0.001525 y un intervalo del 95% de confianza de [63.6277 - infinito]. De estos últimos resultados, se puede concluir que se rechaza la hipótesis nula, lo que nos sitúa a favor de nuestra hipótesis de investigación, las redes neuronales en general presentan un mayor porcentaje de clasificación que kNN para clasificación de posturas tomadas por estudiantes al momento de realizar presentaciones.

#### V. DISCUSIÓN

Para nuestra investigación consideramos la sugerencia de [2], es decir, normalizar los datos y los resultados fueron comparados con [15], cuyos datos no fueron normalizados. La diferencia en el porcentaje de clasificación la atribuimos a la normalización de los datos con los que se entrena el clasificador.

En este trabajo hubo una clasificación en 9 subconjuntos de posturas, mayor que en [15], donde, en base a 6 subconjuntos de posturas, se clasificaban como buena postura y mala postura.

Se observó además que la clasificación de ciertos subconjuntos de posturas producía más problemas que las demás, siendo estas los subconjuntos C1, C8, C11 y C20. Esta dificultad al clasificar es debido a que algunos de estos subconjuntos son muy similares en forma, por lo que la interpretación de los mismos varía por cada individuo.

En conclusión, las redes neuronales pueden predecir la postura de un presentador con una probabilidad mayor al 61%, el resultado obtenido en [15] utilizando kNN, esto está de acuerdo con [2], donde se observó que las redes neuronales superan a kNN al momento de clasificar.

#### VI. TRABAJOS FUTUROS

Como trabajos futuros se podría implementar el mismo análisis usando las junturas de todo el cuerpo, por tanto, se necesitaría definir una mayor cantidad de posturas para entrenar la red neuronal. Otra forma de mejora sería observar

y analizar como incide en el porcentaje clasificación el uso de otros modelos de redes neuronales.

Además, para trabajos adicionales se podría analizar el impacto que tienen las características de las redes neuronales en el porcentaje de clasificación de las mismas.

El porcentaje de clasificación obtenido está asociado al tiempo invertido en el entrenamiento de las redes neuronales. En este trabajo se generan nueve redes neuronales independientes, que clasifican una de las nueve posturas. Se sugiere que se utilicen los pesos encontrados para entrenar una nueva red que clasifique las posturas de forma conjunta.

## REFERENCES

- [1] F. Cary, O. Postolache, and P. Silva Girao, Kinect based system and Artificial Neural Networks classifiers for physiotherapy assessment, in Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on, 2014, pp. 16.
- [2] S. Saha, M. Pal, A. Konar, and R. Janarthanan, Neural network based gesture recognition for elderly health care using kinect sensor, in Swarm, Evolutionary, and Memetic Computing, Springer, 2013, pp. 376386.
- [3] S. Sorce, V. Gentile, and A. Gentile, Real-Time Hand Pose Recognition Based on a Neural Network Using Microsoft Kinect, 2013, pp. 344350.
- [4] T. Visutarrom, P. Mongkolnam, and J. H. Chan, Multiple-Stage Classification of Human Poses while Watching Television, 2014, pp. 1016
- [5] C. Zhao, Y. Gao, J. He, and J. Lian, Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier, Engineering Applications of Artificial Intelligence, vol. 25, no. 8, pp. 16771686, Dec. 2012.
- [6] H. Hasan and S. Abdul-Kareem, Static hand gesture recognition using neural networks, Artificial Intelligence Review, vol. 41, no. 2, pp. 147181, Feb. 2014.
- [7] T. Visutarrom, P. Mongkolnam, and J. H. Chan, Postural classification using Kinect, in Computer Science and Engineering Conference (ICSEC), 2014 International, 2014, pp. 403408.
- [8] Z. Wei and G. Chakraborty, Construction of intelligent intrusion detection system based on KINECT, in Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on, 2013, pp. 8187.
- [9] K. Rimkus, A. Bukis, A. Lipnickas, y S. Sinkevicius, 3D human hand motion recognition system, en iHuman System Interaction (HSI), 2013 The 6th International Conference on, 2013, pp. 180183.
- [10] J. R. Zhang, K. Guo, C. Herwana, y J. R. Kender, Annotation and taxonomy of gestures in lecture videos, en iComputer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, 2010, pp. 18.
- [11] S. Monir, S. Rubya, y H. S. Ferdous, Rotation and scale invariant posture recognition using Microsoft Kinect skeletal tracking feature, en iIntelligent Systems Design and Applications (ISDA), 2012 12th International Conference on, 2012, pp. 404409.
- [12] Trujillo-Romero, F.; Caballero-Morales, S.-O., "3D data sensing for hand pose recognition," in Electronics, Communications and Computing (CONIELECOMP), 2013 International Conference on , vol., no., pp.109-113, 11-13
- [13] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. Commun. ACM 56, 1 (January 2013), 116-124.
- [14] P. Young and B. Sayer, Body Pixel Classification by Neural Network, 6th ed. Fort Worth: Dryden Press; Harcourt Brace College Publishers, 2012.
- [15] K. Chiluiza, V. Echeverra, A. Avendao, A. Vsquez, and X. Ochoa, Presentation Skills Estimation Based on Video and Kinect Data Analysis. Presentation Skills Estimation Based on Video and Kinect Data Analysis, ACM, 2014.
- [16] Y. Liu, Z. Zhang, A. Li, and M. Wang, View independent human posture identification using Kinect, in Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on, 2012, pp. 15901593.