

Advances in Computer Vision, Signal Processing and Virtual Environments

Research in Computing Science

Series Editorial Board

Editors-in-Chief:

Grigori Sidorov (Mexico)
Gerhard Ritter (USA)
Jean Serra (France)
Ulises Cortés (Spain)

Associate Editors:

Jesús Angulo (France)
Jihad El-Sana (Israel)
Alexander Gelbukh (Mexico)
Ioannis Kakadiaris (USA)
Petros Maragos (Greece)
Julian Padget (UK)
Mateo Valero (Spain)

Editorial Coordination:

Alejandra Ramos Porras

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 140**, noviembre 2017. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No.: 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 140**, November 2017. Printing 500. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

Advances in Computer Vision, Signal Processing and Virtual Environments

Oscar Herrera Alcántara (ed.)



Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"



Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2017

ISSN: 1870-4069

Copyright © Instituto Politécnico Nacional 2017

Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal
Unidad Profesional “Adolfo López Mateos”, Zacatenco
07738, México D.F., México

<http://www.rcs.cic.ipn.mx>

<http://www.ipn.mx>

<http://www.cic.ipn.mx>

The editors and the publisher of this journal have made their best effort in preparing this special issue, but make no warranty of any kind, expressed or implied, with regard to the information contained in this volume.

All rights reserved. No part of this publication may be reproduced, stored on a retrieval system or transmitted, in any form or by any means, including electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the Instituto Politécnico Nacional, except for personal or classroom use provided that copies bear the full citation notice provided on the first page of each paper.

Indexed in LATINDEX, DBLP and Periodica

Printing: 500

Printed in Mexico

Editorial

This volume of the journal “Research in Computing Science” contains selected papers related to Signal Processing, Vision and Virtual Environments. The papers were carefully chosen by the editorial board on the basis of the at least two reviews by the members of the reviewing committee or additional reviewers. The reviewers took into account the originality, scientific contribution to the field, soundness and technical quality of the papers. It is worth noting that various papers for this special issue were rejected.

As far as Signal Processing is concerned, the paper of this volume discuss the parameter selection to classify imagined speech on electroencephalograms, the classification of activity and linguistic inactivity from electroencephalogram signals, the development of a BCI using the P300 potential and the MindWave headset, the identification of breast cancer cell features by means of typical testors, the comparison of filters for the analysis of the two-dimensional Fourier space spectrum in the 3D digitization process, and a first step towards automatic beetle recognition.

As far as Vision is concerned, the papers of this volume describe the state of the art of automatic image recognition of the brain, a 3D object reconstruction proposal based on the Morlet wavelet best parameter selection, an age classifier for digital images using statistical methods, the facial expressions recognition based on facial landmarks dynamics, and a face recognition survey.

As far as Virtual Environments is concerned, the papers of this volume describe the body postures reproduction in virtual agents based on neuroscience evidence, a method to transfer learning in imagined speech EEG signals, a simulator ROV multiplayer for oil submarine scenarios, and the development of a collaborative application to design class diagrams in educational environments.

I would like to thank Mexican Society for Artificial Intelligence (Sociedad Mexicana de Inteligencia Artificial) for its support during preparation of this volume.

The entire submission, reviewing, and selection process, as well as preparation of the proceedings, were supported for free by the EasyChair system (www.easychair.org).

Oscar Herrera Alcántara
UAM, Mexico
Guest Editor

September 2017

Table of Contents

	Page
Reconocimiento de expresiones faciales con base en la dinámica de puntos de referencia faciales	9
<i>E. Morales-Vargas, C.A. Reyes-Garcia, Hayde Peregrina-Barreto</i>	
Estado del arte en reconocimiento facial	19
<i>Martín Adrián Garduño Santana, L. E. Díaz-Sánchez, Israel Tabarez Paz, Marcelo Romero Huertas</i>	
Generación de posturas corporales en agentes virtuales a partir de evidencia neurocientífica	29
<i>Diana G. Gómez-Martínez, Marco Ramos, Vianney Muñoz-Jiménez, J. Raymundo Marcial-Romero</i>	
Identificación de características de células de cáncer de mama por medio de testores típicos	43
<i>Alexis Gallegos, Dolores Torres, Francisco Álvarez, Aurora Torres</i>	
Propuesta de reconstrucción 3d de objetos basado en la selección de los mejores parámetros de la Wavelet Morlet	55
<i>Claudia Victoria López, María Eugenia Quintanar Pérez, Jesús Carlos Pedraza Ortega, Gonzalo Elías Blanco Silva, Juan Manuel Ramos Arreguín, Efrén Gorrostieta Hurtado</i>	
Simulador ROV multiplayer para escenarios petroleros submarinos	67
<i>Rodrigo Beltran-Labra, Miguel Pérez-Ramírez, Carlos Alberto Ochoa-Zezzatti, Norma J. Ontiveros-Hernández</i>	
Un primer paso hacia el reconocimiento automático de escarabajos	79
<i>Caleb Jiménez, Maya Carrillo, María de Lourdes Sandoval S., Luis Alfredo Moctezuma</i>	
Clasificador de edad en imágenes digitales usando métodos estadísticos	91
<i>Luis Enrique Colmenares Guillen, Roberto Guillermo Torres López, Maya Carrillo Ruiz, Francisco Javier Albores Velasco</i>	
Estado del arte y elementos del reconocimiento automático de imágenes del cerebro	105
<i>Keila Vasthi Cortés-Martínez, Manuel Mejía-Lavalle</i>	

Selección de parámetros en el enfoque de bolsa de características para clasificación de habla imaginada en electroencefalogramas	123
<i>Jesús S. García-Salinas, Luis Villaseñor-Pineda, Carlos A. Reyes-García, Alejandro A. Torres-García</i>	
Hacia la clasificación de actividad e inactividad lingüística a partir de señales de electroencefalogramas (EEG)	135
<i>Luis Alfredo Moctezuma, Maya Carrillo, Luis Villaseñor Pineda, Alejandro A. Torres García</i>	
Desarrollo de una BCI utilizando el potencial P300 y la diadema Mindwave®	151
<i>Sergio René Vivar Vera, María Antonieta Abud Figueroa, Gabriel Rodríguez Vásquez, Ulises Juárez Martínez, Gustavo Peláez Camarena, Jorge Airy Mercado Gutiérrez, Josefina Gutiérrez Martínez</i>	
Hacia un método de transferencia de aprendizaje en señales de EEG de habla imaginada	165
<i>Jessica Nayeli López Espejel, Maya Carrillo Ruíz, Luis Villaseñor Pineda, Alejandro Torres García</i>	
Comparación de filtros para el análisis del espectro espacial bidimensional de Fourier, en el proceso de digitalización 3D	175
<i>Juan Carlos Moya Morales, Jesús Carlos Pedraza Ortega, Juan Manuel Ramos Arreguín, José Emilio Vargas Soto</i>	
Hacia el desarrollo de una aplicación colaborativa para el diseño de diagramas de clases en ambientes educativos	189
<i>S. David Díaz-García, Luis G. Montané-Jiménez, Betania Hernández-Ocaña</i>	

Reconocimiento de expresiones faciales con base en la dinámica de puntos de referencia faciales

E. Morales-Vargas, C.A. Reyes-García, Hayde Peregrina-Barreto

Instituto Nacional de Astrofísica Óptica y Electrónica,
División de Ciencias Computacionales, Tonantzintla, Puebla,
México

emoralesv@inaoep.mx, kargaxxi@inaoep.mx, hperegrina@inaoep.mx

Resumen. Las expresiones faciales permiten a las personas comunicar emociones, y es prácticamente lo primero que observamos al interactuar con alguien. En el área de computación, el reconocimiento de expresiones faciales es importante debido a que su análisis tiene aplicación directa en áreas como psicología, medicina, educación, entre otras. En este artículo se presenta el proceso de diseño de un sistema para el reconocimiento de expresiones faciales utilizando la dinámica de puntos de referencia ubicados en el rostro, su implementación, experimentos realizados y algunos de los resultados obtenidos hasta el momento.

Palabras clave: Expresiones faciales, clasificación, máquinas de soporte vectorial, modelos activos de apariencia.

Facial Expressions Recognition Based on Facial Landmarks Dynamics

Abstract. Facial expressions allow people to communicate emotions, is practically the first thing that we observe when interacting with someone. In computer science, facial expressions recognition is important because their analysis has direct application in areas such psychology, medicine, education, among others. In this paper is presented the design process of a facial expressions recognition system which uses facial landmarks dynamics, its implementation, experiments performed and some of the results obtained until now.

Keywords: facial expressions, classification, support vector machines, active appearance models.

1. Introducción

Existen diversas áreas en las que las expresiones faciales son estudiadas, entre ellas se incluyen la psicología, neurociencia, educación, o sociología [15,19]. Existe

una fuerte evidencia que soporta el hecho de que existen siete emociones básicas que tienen asociadas una expresión facial, que pueden ser: enojo, desprecio, disgusto, miedo, felicidad, tristeza o sorpresa [17,8,2,16].



Fig. 1. Expresiones faciales básicas, de izquierda a derecha: enojo, disgusto, desprecio, felicidad, miedo, tristeza y sorpresa. Imágenes tomadas de la base de datos CK+ [12].

Se han propuesto varias metodologías para el reconocimiento de expresiones faciales con un enfoque estático. Los enfoques estáticos utilizan descriptores locales en un solo recuadro de una secuencia para extraer características. Un operador utilizado en varios trabajos es Patrones Binarios Locales (*LBP*) y sus mejoras, el cual se utiliza para describir la textura en una imagen y posteriormente utilizan como clasificador Máquinas de Soporte Vectorial (*SVM*) [18,14,20].

Los enfoques dinámicos para el reconocimiento de expresiones faciales toman como referencia la diferencia entre estado neutral y la representación de una expresión facial, extrayendo como característica discriminadora la dinámica del rostro. [12] utiliza el desplazamiento de puntos de referencia faciales obtenidos a través de Modelos Activos de Apariencia (*AAM*), y el trabajo propuesto por [9] utiliza los ángulos que se forman en el rostro para realizar el reconocimiento.

En este artículo nos enfocamos en la generación de un vector de características que reconozca expresiones faciales de manera robusta en secuencias de imágenes que sea lo más compacto y simple posible con respecto a trabajos presentados anteriormente, capturando la dinámica del rostro mediante el desplazamiento y la dirección de puntos de referencia faciales.

2. Sistema propuesto

El sistema propuesto utiliza la base de datos CK+ [12] la cual contiene un conjunto de secuencias de imágenes en las que se utilizó *AAM* para estimar la forma del rostro o puntos de referencia faciales, que se define como un conjunto de coordenadas. La forma del rostro del estado neutral y de la expresión facial

se alinean para reducir el ruido y posteriormente se extraen características que describen la dinámica de los puntos de referencia faciales. Se utilizó SVM para clasificar las expresiones faciales. En la Fig. 2 se puede observar el sistema completo y en las secciones siguientes se describe cada una de las etapas.

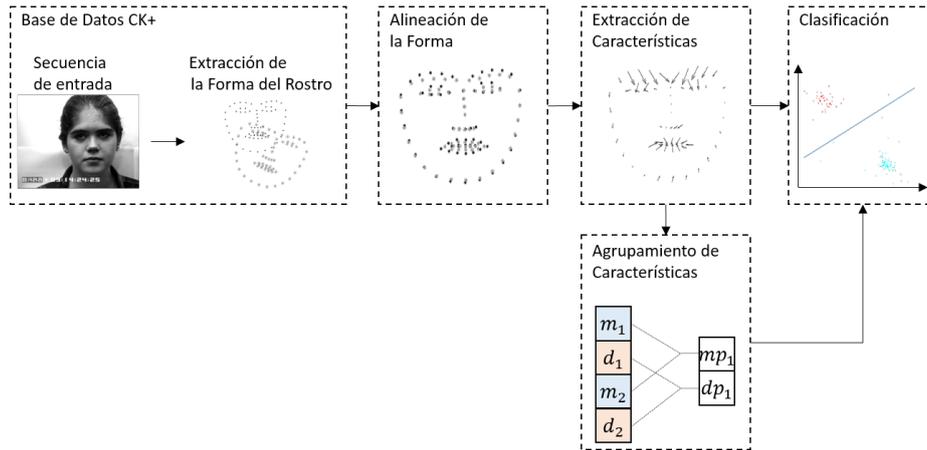


Fig. 2. Metodología general propuesta, base de datos tomada de [12].

2.1. Alineación de la forma del rostro

Al utilizar los puntos de referencia faciales existen factores que se deben tomar en cuenta, es necesario remover el efecto del tamaño, orientación y ubicación de las coordenadas para reducir el ruido que se introduce al sistema [3]. Uno de los métodos utilizados en la literatura para alinear la forma del rostro es el análisis de Procrustes [4,12], en nuestro estudio se utilizaron transformaciones afines para reducir la orientación y la variación espacial, y posteriormente los valores de las coordenadas se normalizan en un rango entre 0 y 1 en términos del estado neutral para evitar perder las deformaciones causadas por el movimiento del rostro.

2.2. Extracción de características

Una vez que se ha alineado la forma del rostro del estado neutral de una secuencia con su expresión facial es posible extraer información correspondiente a la dinámica de los puntos de referencia faciales de manera adecuada. Para hacer distinción entre las coordenadas en una secuencia entre el estado neutral y su expresión facial, re-definiremos la forma del rostro s de la siguiente manera:

- Estado neutral: $sn = [xn_1, yn_1, xn_2, yn_2, \dots, xn_n, yn_n]$.

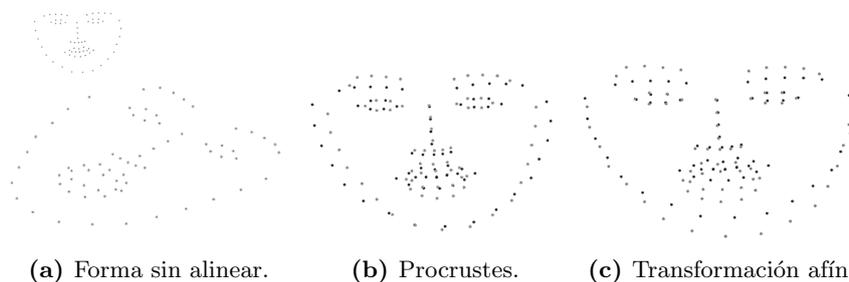


Fig. 3. En a(a) se puede observar que la forma del rostro en estado neutral y la representación facial pueden estar afectadas por ubicación, orientación y tamaño, en (b) se puede observar el resultado que se obtiene al realizar el análisis de Procrustes a ambas formas y en (c) se observa el resultado al realizar un conjunto de transformaciones afines para así preservar información del movimiento de las cejas y mandíbula que se puede perder con el análisis de Procrustes.

- Expresión facial: $se = [xe_1, ye_1, xe_2, ye_2, \dots, xe_n, ye_n]$.

La primera etapa de la caracterización consiste en obtener el desplazamiento horizontal y vertical del rostro, para esto se extrae el estado neutral a la expresión facial, dando como resultado el desplazamiento, lo cual se realiza con las ecuaciones 1 y 2:

$$\Delta x_i = xe_i - xn_i, \quad (1)$$

$$\Delta y_i = ye_i - yn_i. \quad (2)$$

Con el desplazamiento horizontal y vertical de los puntos de referencia es posible calcular la magnitud del movimiento con la ecuación 3 y la dirección con la ecuación 4.

$$m_i = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2}, \quad (3)$$

$$d_i = \tan^{-1}\left(\frac{\Delta x_i}{\Delta y_i}\right). \quad (4)$$

La caracterización base consiste en la concatenación de la intensidad y magnitud del movimiento de los puntos de referencia faciales, la cual se define de la siguiente manera: $c = [m_1, d_1, m_2, d_2, \dots, m_n, d_n]$.

2.3. Agrupación

Diversos trabajos utilizan el agrupamiento de características para combinar resultados de varios descriptores o para agrupar características de una región. Algunos métodos populares que utilizan la agrupación son la Transformada de Características Invariante a Escala (SIFT) [11], el Histograma de Gradientes Orientados (HOG) [5], entre otros. La agrupación ayuda a producir una representación más estable de un grupo de características inestables [10]. De manera general, el agrupamiento de características consiste en transformar la

representación de características en una nueva representación más útil y estable que preserva solo información relevante [1]. La dificultad al agrupar radica en identificar cuales características corresponden a cada grupo.

Se redujo el vector de características mediante operaciones de agrupación $f = \{f_a(x), f_m(x)\}$ donde $x \subset c$ a un nueva nueva representación con 22 valores. El criterio de agrupación consiste en seleccionar y agrupar las características

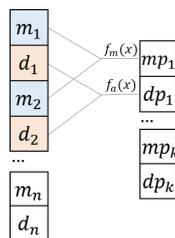


Fig. 4. Diagrama de agrupación, del lado izquierdo se puede observar el vector de características sin agrupar, varias características se agrupan en una sola mediante y se concatenan en una nueva representación mediante operaciones de agrupación.

correspondientes a las diferentes áreas del rostro: interior de cejas, exterior de cejas, parpados, nariz, labio superior, labio inferior, esquina derecha del labio, esquina izquierda del labio y mandíbula [7]:

$$f_a(x) = \frac{1}{|x|} \sum_{i=1}^{|x|} x_i, \tag{5}$$

$$f_m(x) = \max(x). \tag{6}$$

3. Datos

Se realizaron experimentos en la base de datos CK+ [12] la cual contiene 327 secuencias de imágenes. En la Tabla 1 se puede observar el número de secuencias para cada expresión facial. Cada secuencia comienza en estado neutral y termina con la representación de una expresión facial. Un juez experto, manualmente codificó las secuencias de imágenes mediante el Sistema de Codificación Facial (FACS) [6], asignó una etiqueta la cual indica qué expresión facial se percibe en la secuencia. Para cada imagen, los autores obtuvieron un conjunto de 68 coordenadas que describen la forma del rostro mediante AAM, un algoritmo basado en gradiente descendente propuesto en [13]. La forma del rostro es un conjunto de n coordenadas, donde cada coordenada pertenece a un vértice de la forma del rostro, la cual se encuentra definida por $s = [x_1, y_1, x_2, y_2, \dots, x_n, y_n]$.

Tabla 1. Frecuencia de las muestras para cada expresión facial de la base de datos CK+ [12].

Emoción	N
Enojo(En)	45
Disgusto(Dis)	59
Desprecio(Des)	18
Felicidad(Fel)	69
Miedo(Mi)	25
Tristeza(Tris)	28
Sorpresa(Sor)	83

4. Experimentos y resultados

Las pruebas del sistema propuesto fueron hechas utilizando la base de datos extendida de Cohn-Kanade (CK+) [12], la cual contiene secuencias de imágenes de personas actuando las siete expresiones faciales básicas (Fig. 1). Debido a las clases desbalanceadas y siguiendo el marco de referencia propuesto en [12], el sistema propuesto se validó utilizando la estrategia de dejar un sujeto fuera (*leave one out*) y se presentan la exactitud promedio y la exactitud ponderada por expresión facial para poder realizar un análisis de las consecuencias que esto conlleva. El clasificador utilizado fué Máquinas de Soporte Vectorial.

En la Tabla 2 se presenta la matriz de confusión que muestra los resultados correspondientes utilizando la dinámica de puntos de referencia faciales, sin agrupar características. El porcentaje de clasificación para esta configuración es de 93.5% y corresponde al promedio pesado de la diagonal. En la Tabla 3 se muestra la matriz de confusión con los resultados correspondientes a la clasificación cuando se agrupan las características, con esta configuración se obtuvo un 92.3% de exactitud.

Tabla 2. Matriz de confusión correspondiente al reconocimiento de expresiones faciales con un enfoque dinámico sin agrupación de características.

		Predicción						
		En	Des	Dis	Mi	Fel	Tris	Sor
Valor real	En	91.1	0	6.7	0	0	2.2	0
	Des	0	83.3	0	0	0	16.7	0
	Dis	3.4	1.7	94.8	0	0	0	0
	Mi	0	0	0	88.8	8	4	0
	Fel	0	1.4	0	1.4	98.6	0	0
	Tris	3.6	0	0	0	0	96.4	0
	Sor	0	1.2	0	4.8	0	1.2	92.8

Los porcentajes de clasificación individuales para cada expresión facial muestran que nuestro sistema en sus dos configuraciones presenta un desempeño

Tabla 3. Matriz de confusión correspondiente al reconocimiento de expresiones faciales con un enfoque dinámico y agrupación de características.

		Predicción						
		En	Des	Dis	Mi	Fel	Tris	Sor
Valor real	En	91.1	0	4.4	0	0	4.4	0
	Des	5.6	83.3	0	0	0	11.1	0
	Dis	3.4	0	96.6	0	0	0	0
	Mi	0	0	0	84	12	0	4
	Fel	0	1.4	0	1.4	97.1	0	0
	Tris	10.7	3.6	0	3.6	0	82.1	0
	Sor	0	1.2	0	3.6	0	1.2	94

aceptable con respecto a los trabajos relacionado con vectores de características que subjetivamente se han descrito como pequeños.

5. Discusión

En la tabla 4 se presenta una comparación entre los trabajos fuertemente relacionados con nuestra propuesta.[12] comienza con una etapa de normalización en la cual se extrae el ruido de la ubicación, escala y rotación de lo puntos de referencia faciales utilizando la superposición o análisis de Procrustes. Con estos datos se extrae la Forma Normalizada de Similitud (*SPTS*) la cual se refiere a la forma del rostro después de la normalización del estado neutral, obteniendo un vector con 136 valores que describe el desplazamiento de los puntos de referencia del rostro.

Por otra parte, [9] propone un descriptor basado en capturar los cambios de 560 ángulos obtenidos a partir de la combinación entre los 68 puntos de referencia en el rostro. Los autores mencionan que su enfoque es independiente de la pose del rostro debido a que solo se mide la variación del movimiento. El descriptor utilizado puede tener tres valores discretos dependiendo de la magnitud de la diferencia de los ángulos entre la expresión facial y el estado neutral. En el trabajo de [9] también se utilizó la base de datos de [12] por lo que es directamente comparable con la metodología propuesta.

La metodología propuesta hace uso de transformaciones afines en lugar de utilizar Procrustes[12], método de referencia para la normalización de los puntos que describen al rostro. Como se puede observar en la Tabla 4, el uso de dichas transformaciones afines permite conservar mejor la información del movimiento del rostro al comparar con el estado neutral. En los trabajos relacionados se requiere analizar un mayor número de direcciones [9] y el movimiento sólo tiene una descripción horizontal y vertical [12]. La metodología propuesta en este trabajo permite calcular directamente (entre puntos correspondientes) en qué dirección se produjo el movimiento de un determinado punto del rostro y brinda una mejor descripción de dicho movimiento. Lo anterior permite también mejorar la exactitud del reconocimiento de las expresiones.

Tabla 4. Comparación entre diferentes trabajos relacionados.

Trabajo	Tamaño del vector	Exactitud promedio	Exactitud pesada
SPTS+SVM [12]	136	50.4	66.7
DA+CRF [9]	560	78	86.9
Propuesto S.A.	136	92.1	93.6
Propuesto C.A.	22	89.7	92.3

Al realizar la presente investigación se buscó encontrar un vector de características que describa el movimiento del rostro con el menor número de valores posibles, es por eso que se optó por adoptar un esquema de agrupamiento. Si bien comparado con la configuración sin agrupamiento (S. A.), la metodología con agrupamiento (C. A.) no aumenta el porcentaje de clasificación, lo cual puede ser debido a la selección de los puntos de referencia escogidos para representar cada zona de interés del rostro, si reduce la dimensionalidad del vector de características, de 136 valores a 22. Esto permite generar una representación simple y compacta que reduce el procesamiento en comparación con los trabajos relacionados.

En la Fig. 5 se pueden observar la comparación entre nuestro método con sus dos configuraciones: con agrupación y sin agrupación, *SPTS+SVM* [12] y *DA+CRF* [9]. Las clases desbalanceadas afectan el desempeño del clasificador debido a que si en las clases con un número pequeño de muestras ocurre un acierto o error estos afectan la exactitud drásticamente, es por esto que también se presenta el promedio ponderado o pesado por cada expresión facial, el cual asigna un peso al resultado de cada expresión facial dependiendo del número de muestras que se tienen por clase, otra forma de calcular este promedio pesado es obteniendo la exactitud general del clasificador.

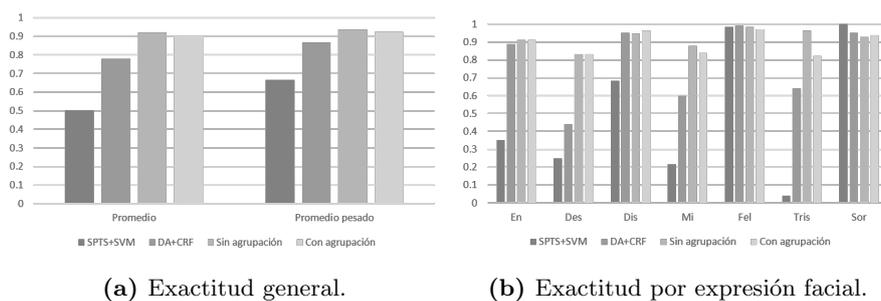


Fig. 5. En (a) se observa la comparación del promedio pesado del porcentaje de reconocimiento entre nuestro método con sus dos variantes, *SPTS+SVM*, y *DA+CRF*. En (b) se observa la comparación del porcentaje de reconocimiento por expresión facial entre las variantes de nuestro método y métodos de la literatura.

6. Conclusiones

En este artículo se presentó el proceso de diseño de un sistema simple y rápido para el reconocimiento de expresiones faciales el cual se basa en la dinámica de puntos de referencia faciales, con los cuales se calcula la magnitud y la dirección del movimiento en el rostro. Se agruparon los resultados del descriptor de movimiento en regiones de interés del rostro y se realizó una comparación entre los resultados obtenidos al clasificar antes y después de la agrupación.

Agrupar los valores de la representación de la dinámica del rostro y clasificar con SVM alcanzó un porcentaje de reconocimiento de 92.3 %, generar una representación mediante agrupación no aumentó el porcentaje de reconocimiento sin agrupar debido a la pérdida de información que conlleva el utilizar el enfoque de agrupación empleado pero es importante remarcar que existe una diferencia significativa, debido a que al se agrupan los valores se obtiene una nueva representación compacta y simple que logra discriminar entre las expresiones faciales de una manera similar a su contraparte.

Agradecimientos. El autor E. Morales-Vargas agradece al Consejo Nacional de Ciencia y Tecnología (CONACyT) por al apoyo a esta investigación a través de la beca #702647. Los autores agradecen al proyecto MX14MO06 "Técnicas de análisis y clasificación de voz y expresiones faciales: aplicación a las enfermedades neurológicas en recién nacidos y adultos" del programa ejecutivo de cooperación científica y tecnológica México-Italia financiado por AMEXID de la SRE y el Ministerio de Asuntos Exteriores de Italia.

Referencias

1. Boureau, Y.L., Ponce, J., Lecun, Y.: A Theoretical Analysis of Feature Pooling in Visual Recognition. In: 27th International Conference on Machine Learning, Haifa, Israel (2010)
2. Burrows, A.M., Waller, B.M., Parr, L.A., Bonar, C.J.: Muscles of facial expression in the chimpanzee (*Pan troglodytes*): descriptive, comparative and phylogenetic contexts. *Journal of Anatomy* 208(2), 153–167 (Feb 2006), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2100197/>
3. Cohn, J.F., Zlochower, A.J., Lien, J., Kanade, T.: Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology* 36(1), 35–43 (Jan 1999), <http://onlinelibrary.wiley.com/doi/10.1017/S0048577299971184/abstract>
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 681–685 (Jun 2001)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 1, pp. 886–893 vol. 1 (Jun 2005)
6. Ekman, P., Friesen, W.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press (1978)

7. Ekman, P.: Facial expression and emotion. *American Psychologist* 48(4), 384–392 (1993)
8. Galati, D., Miceli, R., Sini, B.: Judging and coding facial expression of emotions in congenitally blind children. *International Journal of Behavioral Development* 25(3), 268–278 (May 2001), <http://dx.doi.org/10.1080/01650250042000393>
9. Iglesias, F., Negri, P., Buemi, M.E., Acevedo, D., Mejail, M.: Facial expression recognition: a comparison between static and dynamic approaches. In: *International Conference on Pattern Recognition Systems (ICPRS-16)*. pp. 1–6 (Apr 2016)
10. Krig, S.: Feature Learning and Deep Learning Architecture Survey. In: *Computer Vision Metrics*, pp. 375–514. Springer International Publishing (2016), http://link.springer.com/chapter/10.1007/978-3-319-33762-3_10, doi: 10.1007/978-3-319-33762-3_10
11. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), 91–110 (Nov 2004), <https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>
12. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. pp. 94–101 (Jun 2010)
13. Matthews, I., Baker, S.: Active Appearance Models Revisited. *International Journal of Computer Vision* 60(2), 135–164 (Nov 2004), <http://link.springer.com/article/10.1023/B%3AVISI.0000029666.37597.d3>
14. Mohammadi, M.R., Fatemizadeh, E.: Fuzzy local binary patterns: A comparison between Min-Max and Dot-Sum operators in the application of facial expression recognition. In: *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*. pp. 315–319 (Sep 2013)
15. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1424–1445 (Dec 2000)
16. Park, S., Kim, D.: Spontaneous facial expression classification with facial motion vectors. In: *8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08*. pp. 1–6 (Sep 2008)
17. Peleg, G., Katzir, G., Peleg, O., Kamara, M., Brodsky, L., Hel-Or, H., Keren, D., Nevo, E.: Hereditary family signature of facial expression. *Proceedings of the National Academy of Sciences* 103(43), 15921–15926 (Oct 2006), <http://www.pnas.org/content/103/43/15921>
18. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing* 27(6), 803–816 (May 2009), <http://www.sciencedirect.com/science/article/pii/S0262885608001844>
19. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(1), 39–58 (Jan 2009)
20. Zhao, G., Pietikainen, M.: Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(6), 915–928 (Jun 2007)

Estado del arte en reconocimiento facial

Martín Adrián Garduño Santana, L. E. Díaz-Sánchez, Israel Tabarez Paz,
Marcelo Romero Huertas

Universidad Autónoma del Estado de México, Toluca, México

martingarduno.94@gmail.com, lediazs@uaemex.mx, israeltabarez@gmail.com,
mromeroh@uaemex.mx

Resumen. En este trabajo se resumen los métodos más utilizados para el reconocimiento facial, incluyendo las ventajas y desventajas de los sistemas desarrollados hasta ahora. También se describen las futuras líneas de investigación y se discute el rumbo del reconocimiento facial en los próximos años. Esta revisión es relevante pues se busca la implementación de un novedoso sistema de reconocimiento facial.

Palabras clave: reconocimiento facial, sistemas biométricos, ciudades inteligentes, imágenes 2D y 3D.

Face Recognition: a Survey

Abstract. In this paper, we summarize the most used methods for face recognition, including advantages and drawbacks of each system that has been developed until now. Also, we describe the future research lines and discuss the direction of face recognition. This survey is relevant for our research, as we are encouraged to develop a novel face recognition system as part of our future work.

Keywords: face recognition, biometrics, smart cities, 2D and 3D images.

1. Introducción

El reconocimiento facial (RF) se ha convertido un área importante para la investigación desde la década de los setentas del siglo XX, puesto que varios científicos e ingenieros alrededor del mundo se han enfocado para establecer algoritmos y métodos cada vez más robustos y precisos para este tipo de sistemas y su aplicación en la vida cotidiana. Además, recientemente, el objetivo de desarrollar aplicaciones biométricas, como el reconocimiento facial, es importante en las ciudades inteligentes.

Actualmente existen diversas áreas dentro del funcionamiento de una ciudad inteligente que van desde el entretenimiento hasta seguridad, en las cuales el RF está involucrado. La Figura 1 muestra las áreas y aplicaciones específicas para el RF.

De manera general, el RF se clasifica en dos tipos: verificación o autenticación e identificación o reconocimiento.

Áreas	Aplicaciones específicas
Entretenimiento	Videojuegos, realidad virtual, Interacción Humano-robot y humano-computadora.
Tarjetas Inteligentes	Licencias de conducir, ID's, pasaportes, inmigración.
Seguridad	Inicio de sesión o desbloqueo en dispositivos personales, encriptamiento de archivos, bases de datos.
Cumplimiento de la ley y vigilancia	Video-vigilancia avanzada, control de circuitos cerrados, control de accesos.

Fig. 1. Principales aplicaciones y áreas de desarrollo de sistemas de reconocimiento facial [34].

La verificación consta de comparar la cara de un sujeto, quien introduce su información personal, con una base de datos, el sistema dice si el sujeto es quien dice ser o no. En otras palabras, es una correspondencia uno a uno.

Por otro lado, la identificación es una búsqueda de uno en muchos, pues un rostro desconocido es comparado con toda la base de datos y el sistema establece si el rostro pertenece a algún sujeto dentro de la base de datos o si definitivamente es desconocido. Ciertos sistemas de identificación son capaces de aprender de manera automática, por lo que es posible identificar un rostro que antes era desconocido luego de que el sistema lo haya visto en previas ocasiones añadiéndolo a la base de datos [1, 2, 21, 32].

Ya sea para la verificación o la identificación, todos los métodos para el reconocimiento facial se clasifican en tres categorías:

- Basados en imágenes 2D,
- Basados en imágenes 3D,
- Híbridos (2D+3D).

En los últimos años, se han publicado varios artículos de reseña literaria [1, 6, 24, 34], teniendo en cuenta los métodos de RF más relevantes, algunos también han incluido enfoques multimodales. En esta revisión, damos una introducción y la aplicación potencial de sistemas RF para ciudades inteligentes.

2. Estado del arte

El reconocimiento facial es un trabajo complicado, incluso para los seres humanos, porque solo es posible recordar e identificar un número limitado de caras. La psicología y neurociencias han estudiado como los seres humanos pueden reconocer diferentes rostros [34], por su parte, investigadores en el área computacional e informática buscan que esta tarea sea realizada de manera eficiente mediante computadoras.

En los setenta, los investigadores de la interacción hombre-máquina intentaron automatizar este proceso. Harmon y Kanade fueron los primeros en implementar un sistema capaz de reconocer caras humanas por una computadora. En sus trabajos, [11, 12], Harmon identificó perfiles del rostro dibujados por un artista. En su tesis doctoral, Kanade desarrolló un sistema para reconocer caras con una alta precisión, detectando 608 caras de una base de datos de 670 imágenes [16].

2.1. Reconocimiento facial en 2D

Al principio el RF fue tratado como un reconocimiento de patrones de imágenes en 2D y con las diversas investigaciones se implementó una subcategoría para estos sistemas basados en 2D, misma que se muestra a continuación.

Los métodos basados en características extraen rasgos o marcas locales de cada cara, como ojos, nariz, boca, etc., y se colocan en un clasificador. Por geometría, los métodos genéricos utilizan bordes, líneas y curvas para reconocer caras diferentes. Otros utilizan plantillas o algunas restricciones en las características [34].

Manjunath et al., utilizaron un enfoque basado en características usando los datos de intensidad de la imagen para adquirir las características faciales. El modelo de detección está biológicamente motivado en la corteza visual de los mamíferos y para obtener la información de las características que utilizan una transformación matemática llamada Gabor wavelet [22].

En la aproximación holística el rostro completo del sujeto se toma como datos de entrada en el sistema de reconocimiento [34]. Uno de los métodos holísticos más utilizados hasta ahora, fue presentado por Kirby y Sirovich en 1990, que implementa un modelo matemático conocido como análisis de componentes principales (PCA por sus siglas en inglés), que toma los $N \times N$ píxeles de una imagen de un rostro y los linealiza para crear un vector de longitud N^2 . Después de eso, con una galería o conjunto de entrenamiento se obtiene un rostro medio y se calcula la matriz de covarianza para encontrar sus eigenvectores y eigenvalores. La principal ventaja del PCA es la reducción de la dimensionalidad, debido a que solo se pueden elegir únicamente los vectores propios más grandes y crear un espacio de dimensión $M < N$.

Para averiguar si un nuevo rostro pertenece a un sujeto de la base de datos, este se proyecta en el mismo espacio del conjunto de entrenamiento y la distancia de la imagen proyectada y las imágenes de la base de datos se calcula, si dicha distancia es menor que un parámetro ϵ_k , el rostro pertenece al k -ésimo sujeto de la base de datos [18][33].

Los vectores propios obtenidos fueron llamados por Turk y Pentland como eigenrostros (ver Figura 2). También presentaron una aproximación utilizando eigenfaces para la detección e identificación de caras, además de un sistema de reconocimiento que sigue a la persona y es capaz de identificar si está en una base de datos conocida [32].

Otro método que incluye eigenvectores fue propuesto por Etemad y Chellappa, a través de un modelo lineal conocido como análisis de discriminante lineal (LDA por sus siglas en inglés) que se basa en el análisis de las diferentes características de la cara en términos de su potencial discriminante, este método proporciona una mejor

discriminación entre clases, tratando de maximizar las diferencias entre clases y minimizarla dentro de la misma clase [8]. En 1998, Chao et al., desarrollaron el análisis discriminante de los componentes principales combinando los métodos PCA y LDA [35].



Fig. 2. Eigenrostros calculados por PCA [18].

En la misma línea, los investigadores habían tratado de resolver los problemas de usar PCA, tales como, que no es discriminante entre clases y problemas con las variaciones de pose y de iluminación. En 1997, Belhumer et. al., propusieron la aproximación de Fisherostros basado en el método de discriminante lineal de Fisher, y demostraron que esta técnica tiene tasas de error más bajas que el método de eigenrostros [4].

En [3], Bartlett presenta una generalización de PCA, conocida como análisis de componentes independientes (ICA por sus siglas en inglés) que da una base de imágenes más local, y autores como Déniz et al. combinan esta técnica y PCA con máquinas de soporte vectorial (SVM por sus siglas en inglés), con una mejor aproximación para el reconocimiento [7]. Una investigación reciente por Olivares hace modificaciones a un método similar usando eigenfaces y SVM [23].

Las redes neuronales también se han utilizado para RF, la idea básica es considerar una red con una neurona para cada píxel en la imagen [1]. Lin et al. proponen un sistema basado en una red neural probabilista basada en la toma de decisiones (PDBNN por sus siglas en inglés) y consiste en tres módulos, un detector de cara, que localiza el rostro humano en una imagen, un localizador de ojos y un reconocedor de rostro a partir de una base de datos [20].

El reto de reconocimiento facial del bueno, el malo y el feo (GBUFRC por sus siglas en inglés) es un trabajo reciente para alentar el desarrollo de algoritmos que son robustos a través de los cambios en imágenes fijas frontales 2D. Consiste en tres particiones: el bueno contiene pares de imágenes consideradas fáciles de reconocer con una tasa de verificación base (VR por sus siglas en inglés) de 98%. El malo contiene

pares de imágenes en promedio difíciles de reconocer con un VR de 80%. El feo son pares de imágenes difíciles de reconocer con VR de 15% [26].

2.2. Reconocimiento facial 3D

En las técnicas basadas en imágenes 2D algunas características se pierden debido a la estructura 3D de la cara. Las variaciones en la pose y la iluminación son los principales problemas de las imágenes 2D, por lo que los investigadores han elegido imágenes 3D para el reconocimiento facial y compensar esta desventaja. Hay dos maneras de representar una imagen en 3D, a través de imágenes de rango o imágenes 2.5D y 3D representadas por mallas poligonales. La imagen de rango es una representación de los puntos 3D establecidos (x, y, z) , donde los puntos z son un valor de profundidad y corresponden al valor de intensidad de cada píxel en el plano XY (ver Figura 3). Si una imagen 2.5D se toma desde un único punto de vista, solo una superficie facial podría ser modelada en lugar de la cabeza entera. Para un modelo de cabeza completo se requieren varias tomas desde diferentes puntos de vista. La malla poligonal 3D consiste en una lista de vértices conectados bordes [1].

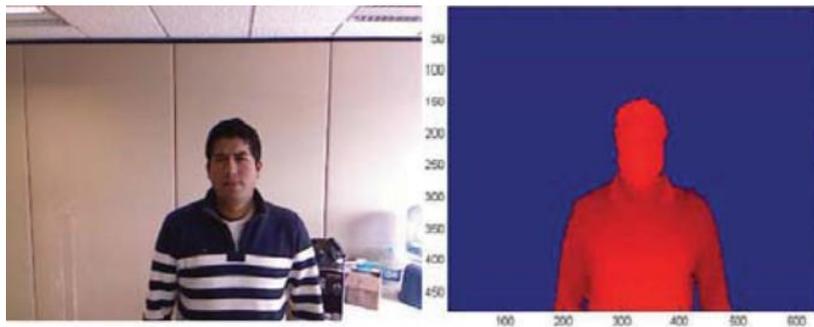


Fig. 3. Imagen de rango tomada mediante el sensor Kinect de Microsoft [28].

La adquisición estéreo, la luz estructural y el sensor láser son tres técnicas de adquisición diferentes para datos 3D. Para la adquisición estéreo, dos o más cámaras se posicionan y calibran para obtener imágenes simultáneas del sujeto.

Entonces los datos de profundidad se calculan a partir de modelos geométricos y con puntos de referencia resuelven el problema de correspondencia. Esta técnica es la más barata y más fácil de usar. La técnica de luz estructural implica un patrón de luz proyectado y la información de profundidad se adquiere a partir de la distorsión del patrón. Es relativamente rápido y barato. La técnica láser es más precisa pero la más cara y la más lenta también. Emplea un sensor láser para escanear.

Generalmente, los datos 3D requieren ser preprocesados después de la adquisición, dependiendo del sensor, hay agujeros, picos y ruido en los datos [2].

Blanz et al., proponen un método basado en el modelo deformable del rostro humano 3D. Su algoritmo simula el proceso de formación de imágenes en el espacio 3D y codifica la forma y la textura en términos de parámetros del modelo. Utilizaron una base de datos de escaneo láser y la identificación se basa en la comparación de

parámetros de coeficientes de una imagen de entrada con los coeficientes de la base de datos [5].

En 2006, Kakadiaris et al., desarrollaron un sistema completamente automático basado en digitalizadores faciales en 3D y reportan una de las más altas precisiones en la base de datos del Face Recognition Grand Challenge (FRGC) [15]. También en 2009, otro sistema 3D fue implementado por Khalid y Lili, el cual extrae rasgos faciales locales de imágenes de rango y calcula la distancia y el ángulo de doce puntos de referencia para la identificación [17]. El uso de Kinect de Microsoft en investigaciones recientes ha facilitado la adquisición de imágenes en profundidad [30]. En su tesis de licenciatura, seis estudiantes de la Universidad Tecnológica de Delft desarrollaron un sistema utilizando el dispositivo Kinect y MATLAB. Dividieron el proceso en tres etapas: adquisición de imágenes, procesamiento de datos y reconocimiento [9, 10, 14].

2.3. Métodos híbridos o multimodales

Los métodos híbridos o biométricos multimodales combinan imágenes en 2D y 3D para obtener un mejor rendimiento [6].

En una investigación reciente se propone el sistema DeepFace, y es posible debido a la combinación de redes neuronales e imágenes 3D, y prueban que este sistema tiene una precisión del 91,4 % [31].

En [6] se muestra una comparación entre sistemas multimodales, 3D y 2D, concluyendo que los multimodales tienen un mejor rendimiento, sin embargo, en estas comparaciones no se controla el número de muestras para cada experimento.

La primera evaluación biométrica multimodal se presentó en el Face Recognition Vendor Test 2006, midiendo los rendimientos en el reconocimiento facial y de iris.

La prueba se enfocó en el reconocimiento de imágenes de alta resolución 2D y 3D tomadas en condiciones controladas y no controladas. Además, por primera vez se comparó la capacidad de reconocimiento entre hombre y máquina, teniendo como resultados que varios algoritmos tienen un mejor rendimiento que las personas [29].

2.4. Base de datos de RF

Como parte importante del desarrollo de sistemas de RF, las bases de datos son utilizadas por los investigadores para evaluar el algoritmo empleado, entre los más utilizados son la Face Recognition Technology (FERET por sus siglas en inglés) [27] y Face Recognition Vendor Test (FRVT) [28, 29]. Face Recognition Grand Challenge es una continuación de FRVT, que introduce pruebas 3D con una base de 50,000 imágenes y aporta cinco conjeturas de diversos criterios de los autores acerca de que método tiene mejor rendimiento [25].

Cada base de datos contiene diferentes imágenes y varios escenarios para evaluar, desde variaciones de iluminación, pose, expresión, cambios de edad o añadir accesorios, como gafas o bufandas. Por otro lado, los investigadores crean sus propias bases de datos y el desempeño del sistema puede ser medido por dos curvas: la curva característica de coincidencia acumulativa (CMC por sus siglas en inglés) y la característica de operación del receptor (ROC por sus siglas en inglés). El primero es

una herramienta para resumir el porcentaje acumulativo de reconocimiento correcto y el segundo resume el espacio de posibles puntos de operación para un sistema de verificación a través de dos tipos de tasas de error [6, 30].

3. Discusión

Los métodos basados en características tienen un problema cuando se trata de detectar características faciales cuando éstas no se encuentran en condiciones normales, es decir, los ojos cerrados, la boca abierta, o con el uso de accesorios. Los métodos basados en imágenes 2D demuestran que son sensibles principalmente a los cambios en la iluminación y la postura. Bajo condiciones controladas, la mayoría de los métodos basados en la imagen 2D alcanzan una precisión superior al 90% pero caen cuando se introducen las condiciones reales.

El método de eigenrostros se ha vuelto una base natural para comenzar el estudio de reconocimiento facial, una vez que se conocen e identifican las deficiencias de este modelo puede pasarse a uno más complejo o buscar las soluciones a sus desventajas.

Las imágenes 3D podrían ser invariantes para los cambios en la pose y la iluminación. En condiciones reales, la mayoría de los sistemas tienen baja precisión, por lo que los investigadores están buscando nuevas técnicas y nuevas tecnologías.

En los próximos años el uso de cámaras de alta calidad y computación de alto rendimiento será una gran ventaja para los sistemas RF y su implementación en ciudades inteligentes.

4. Trabajo a futuro

Estamos interesados en el desarrollo de un sistema de RF por su posible aplicación en ciudades inteligentes. En un trabajo previo, [30], el sensor Kinect se implementó con las bibliotecas Matlab, y esto representa un inconveniente por la licencia requerida de Matlab, por lo que, queremos conectar este dispositivo con OpenSource.

Además, la colaboración con Fiware-México, nos brinda la oportunidad de usar su plataforma y es una manera de implementar un sistema RF en una ciudad inteligente.

Referencias

1. Abate, A., Nappi, M., Riccio, D., Sabatino, G.: 2d and 3d face recognition: A survey. *Pattern recognition letters*, 28(14), pp. 1885–1906 (2007)
2. Akarun, L., Gokberk, B., Salah, A. A.: 3d face recognition for biometric applications. In: *Signal Processing Conference, IEEE, 2005 13th European*, pp. 1–5 (2005)
3. Stewart-Bartlett, M., Movellan, J. R., Sejnowski, T. J.: Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6), pp. 1450–1464 (2002)
4. Belhumeur, P. N., Hespanha, J. P., Kriegman, D. J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), pp. 711–720 (1997)

5. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9), pp. 1063–1074 (2003)
6. Bowyer, K. W., Chang, K., Flynn, P.: A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. *Computer vision and image understanding*, 101(1), pp. 1–15 (2006)
7. Déniz, O., Castrillon, M., Hernández, M.: Face recognition using independent component analysis and support vector machines. *Pattern recognition letters*, 24(13), pp. 2153–2157 (2003)
8. Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. *JOSA A*, 14(8), pp. 1724–1733 (1997)
9. Fikke, F. S., Gardiner, B. K.: 3d face recognition (image acquisition). Bachelor of Science Thesis, Electrical Engineering (2013)
10. Gerlach, M. M. J., Rooijers, C. T.: 3d face recognition (data processing: Registration and deformation). Bachelor of Science Thesis, Electrical Engineering (2013)
11. Harmon, L. D., Kuo, S. C., Ramig, P. F., Raudkivi, U.: Identification of human face profiles by computer. *Pattern Recognition*, 10(5-6), pp. 301–312 (1978)
12. Harmon, L. D., Hunt, W. F.: Automatic recognition of human face profiles. *Computer Graphics and Image Processing*, 6(2), pp. 135–156 (1977)
13. Heshner, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: *Signal processing and its applications, Proceedings, Seventh international symposium on, IEEE, Vol. 2*, pp. 201–204 (2003)
14. Huijbregts, M. A., Stobbe, B.: 3d face recognition: How to make a fast and reliable database and compare the database with 2d+3d facial input? Bachelor of Science Thesis, Electrical Engineering (2013)
15. Kakadiaris, I. A., Passalis, G., Toderici, G., Murtuza, M. N., Theohari, T.: 3d face recognition. In: *BMVC*, pp. 869–878 (2006)
16. Kanade, T.: *Computer recognition of human faces*, Birkhäuser Basel, Vol. 47 (1977)
17. Khalid, F., Lili, N.: 3d face recognition using multiple features for local depth information. *IJCSNS*, 9(1), pp. 27–32 (2009)
18. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1), pp. 103–108 (1990)
19. Lee, Y., Yi, T.: 3d face recognition using multiple features for local depth information. In: *Video/Image Processing and Multimedia Communications, 4th EURASIP Conference focused on, IEEE, Vol. 1*, pp. 429–434 (2003)
20. Shang-Hung, L., Sun-Yuan, K., Long-Ji, L.: Face recognition/detection by probabilistic decision-based neural network. *IEEE transactions on neural networks*, 8(1), pp. 114–132 (1997)
21. Xiaoguang, L.: Image analysis for face recognition. *Personal notes*, pp. 36 (2003)
22. Manjunath, B. S., Chellappa, R., von der Malsburg, C.: A feature based approach to face recognition. In *Computer Vision and Pattern Recognition, 1992, Proceedings CVPR'92, 1992 IEEE Computer Society Conference on, IEEE*, pp. 373–378 (1992)
23. Olivares-Mercado, J., Toscano-Medina, K., Sánchez-Pérez, G., Nakano-Miyatake, M., Pérez-Meana, H.: Modificaciones del método de eigenphases para el reconocimiento de rostro basado en svm. *Ingeniería, investigación y tecnología*, 17(1), pp. 119–129 (2016)
24. Divyarajsinh, N. P., Brijesh, B. M.: Face recognition methods & applications. *arXiv preprint arXiv,1403.0485* (2014)
25. Phillips, P. J.: Face recognition grand challenge. In *Biometric Consortium Conference* (2004)

26. Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D. S., Dunlop, J., Lui, Y. M., Sahibzada, H., Weimer, S.: An introduction to the good, the bad, & the ugly face recognition challenge problem. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, pp. 346–353 (2011)
27. Phillips, P. J., Grother, P. M., Ross, B., Duane, M., Tabassi, E., Bone, M.: Face recognition vendor test 2002. In: Analysis and Modeling of Faces and Gestures, 2003, AMFG 2003, IEEE International Workshop on, IEEE, pp. 44 (2003)
28. Phillips, P. J., Moon, H., Rizvi, S. A., Rauss, P. J.: The Feret evaluation methodology for face-recognition algorithms. IEEE Transactions on pattern analysis and machine intelligence, 22(10), pp. 1090–1104 (2000)
29. Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K W., Schott, C. L., Sharpe, M.: Frvt 2006 and ice 2006 large-scale results. National Institute of Standards and Technology, NIS- TIR, 7408(1) (2007)
30. Romero, M., Flores, C., Muñoz, V., Altamirano, L. C.: Face recognition using eigensurface on kinect depth-maps. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (World- Comp), pp. 241 (2016)
31. Taigman, Y., Yang, M., Ranzato, M. C., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
32. Turk, M. A., Pentland, A. P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, proceedings CVPR'91., IEEE Computer Society Conference on, IEEE, pp. 586–591 (1991)
33. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), pp. 37–52 (1987)
34. Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A.: Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4), pp. 399–458 (2003)
35. Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D. L., Weng, J.: Discriminant analysis of principal components for face recognition. In: Face Recognition, Springer, pp. 73–85 (1998)

Generación de posturas corporales en agentes virtuales a partir de evidencia neurocientífica

Diana G. Gómez-Martínez, Marco Ramos, Vianney Muñoz-Jiménez,
J. Raymundo Marcial-Romero

Universidad Autónoma del Estado de México, Toluca, Estado de México
México

dgomez279@alumno.uaemex.mx, marco.corchado@gmail.com,
vmunozj@uaemex.mx, jrmarcialr@uaemex.mx

Resumen. La Inteligencia Artificial busca proponer nuevos mecanismos de interacción humano-computadora con un alto mimetismo en comportamientos humanos. Por otro lado, la Neurociencia estudia los procesos cognitivos que están presentes en la generación de agentes autónomos, y que representan un desafío para mostrar comportamientos similares a los humanos para una interacción más natural. Es así, que en este artículo se propone un modelo para generar posturas corporales con respecto a estados emocionales en agentes virtuales basado en la neurociencia, que posteriormente en la interacción los usuarios sean capaces de reconocer y en consecuencia pueda ajustar su comportamiento propio para el logro de tareas y objetivos.

Palabras clave: agente virtual, postura corporal, neurociencia, inteligencia artificial.

Body Postures Reproduction in Virtual Agents based on Neuroscience Evidence

Abstract. Artificial Intelligence aims to propose new mechanisms of interaction between humans and computers with a high level of mimicry regarding human behavior. In another hand, neuroscience studies the cognitive processes that can be used to develop autonomous agents, and that represent the challenge to create one interaction between human-computer the more natural possible. Hence in this paper proposes a model to reproduce body postures on emotional states on virtual agents based on the field Neuroscience, so that by the time interaction arrives users will be able to recognize and consequently adjust their behavior to achieve tasks and objectives.

Keywords: virtual agent, body posture, neuroscience, artificial intelligence.

1. Introducción

Actualmente el uso de agentes virtuales en diferentes áreas de investigación tales como: medicina [5], psicología [11], enseñanza [24], comercio [1], etc. está creciendo sustancialmente, debido a que los agentes virtuales están desempeñando diversos mecanismos que buscan lograr una interacción transparente entre el usuario (ser humano) y la computadora [4]. Este proceso de interacción deberá asemejar la interacción que existe entre los humanos y lograr esto requiere conocer los mecanismos basados en neurociencias para comprender cómo la parte motora del cerebro es regulada por la evaluación emocional, que nos permita movimiento corporal en un agente virtual.

Los agentes virtuales deberán tener la capacidad de moverse tanto corporalmente como en su ambiente en el cual se encuentran embebidos, el movimiento corporal tiene alto impacto en el proceso de comunicación [7], ya que funge como medio para expresar información específica de estados emocionales a partir de variaciones de movimientos y de posturas del cuerpo, además la Psicología ha demostrado que el movimiento y las posturas corporales desempeñan un papel importante en el proceso de regulación emocional [26,27].

Cabe añadir que los agentes [30] virtuales están siendo utilizados en sistemas que tienen como objetivo apoyar en las tareas del ser humano, por ejemplo, los asistentes virtuales utilizados convencionalmente en la telefonía hasta avatares que intentan ayudar a personas con problemas de relación social (p. ej. autismo, esquizofrenia, entre otras) [2]. Por lo que ha sido necesario desarrollar modelos para la generación de agentes virtuales que simulen un ser humano que sea capaz de moverse y comunicarse como si se tratase de nosotros mismos.

Uno de los objetivos que persigue esta propuesta es generar agentes virtuales capaces de reproducir posturas corporales con respecto al estado emocional que desea manifestar, para que de este modo se logre una empatía con el usuario además de facilitar la interacción. Para ello este proceso que encierra la relación de un estado emocional con una postura corporal determinada se respaldó por un modelo cognitivo, dado que éste se basa en explorar la capacidad de las mentes humanas para modificar y controlar la forma en que los estímulos afectan la conducta, sustenta al aprendizaje como un proceso en donde se modifican significados de manera interna. Un modelo cognitivo representa una aproximación de los procesos cognitivos de un ser vivo, lo cual se hace con base en las distintas aproximaciones, una de estas aproximaciones es el enfoque de la neurociencia a partir del uso de los resultados encontrados en la Neurociencia Aplicada [28].

2. Trabajos relacionados

2.1. Emociones

La Psicología ha demostrado que la expresión de las emociones es representada mediante una postura corporal además del habla. Una postura corporal es entendida como una configuración de las articulaciones del cuerpo, para el análisis de estas configuraciones se precisan determinados parámetros con la

finalidad de facilitar su descripción y distinguir una de otra. Por ejemplo, en el trabajo de Kleinsmith et al. [17] distinguen 24 características para definir una postura corporal, para la determinación de estas características se basaron en los tres planos ortogonales (sagital, horizontal y frontal), la medida de la extensión lateral, frontal y vertical del cuerpo, torsión del cuerpo, la inclinación de la cabeza y hombros, y en la orientación del cuerpo.

El trabajo de Kleinsmith et al. [15] partió de precisar las posturas corporales conforme las 6 emociones básicas (enojo, miedo, alegría, sorpresa, desagrado y tristeza) definidas por Ekman y Friesen [8], las posturas fueron mostradas a un conjunto de personas cuya cultura no era la misma con el propósito de identificar el impacto que tiene la cultura en la identificación de las emociones, de sus resultados derivaron que las posturas que tienden a confundirse son: el miedo con la sorpresa y el enojo con el desagrado, caso semejante pasa con las expresiones faciales [14].

Por otro lado, Shafir en sus estudios [26, 27] ha mostrado que a partir de un conjunto de características motoras tales como el cuerpo, esfuerzo, forma y espacio (mismas que son estudiadas en Análisis de Movimiento Laban) se puede realizar la identificación de emociones específicamente la felicidad, enojo, miedo y tristeza, de las cuatro características dos de ellas se basan en la configuración del cuerpo tales son el cuerpo y el espacio.

En dichos estudios obtuvieron posturas corporales peculiares que son interpretadas en el marco de una emoción, con respecto al enojo los movimientos son rápidos, fuertes y directos, la cabeza ligeramente inclinada al frente, con los brazos estrechados en frente; el miedo fue predicho con un cuerpo condensado y cerrado, que conforme el espacio tendía a moverse hacia atrás, con el torso inclinado hacia atrás (es decir, con postura de retirada) y hombros de la misma forma. Para la felicidad los movimientos eran brincos rítmicos, con los brazos levantados que describían movimientos libres juntamente con un cuerpo expandido horizontalmente. Finalmente, con la tristeza el cuerpo se mostraba hundido con pérdida de tono muscular, torso hundido y cabeza abajo.

Desde el punto de vista de Neurociencias las emociones son procesos complejos que tienen gran influencia en el comportamiento de las personas, que implican una respuesta motora tanto facial como todo el cuerpo, mismo que conlleva un proceso cognitivo que realiza la valoración de la situación [6] para generar una respuesta adecuada. El procesamiento emocional puede llevarse a nivel consciente e inconsciente, las conexiones que la amígdala mantiene con el hipotálamo producirán una respuesta emocional, sin que hasta el momento la información hubiera alcanzado la corteza y mediación consciente. En lo que concierne al aspecto consciente de las emociones, las estructuras involucradas incluyen la porción anterior del cíngulo, la corteza prefrontal orbital y ventromedial, el lóbulo temporal y la ínsula [21].

Cabe agregar que se han desarrollado modelos enfocados a generar el proceso emocional en agentes virtuales, tal es el caso de Rosales et al. [13] que proponen un modelo para la regulación emocional con memoria para agentes virtuales, en donde definen la *regulación emocional* como el proceso que valora el estado

emocional y calcula el comportamiento apropiado para el logro de las metas. El modelo propuesto se fundamenta en evidencia neurocientífica y psicológica, mismo que está conformado por tres fases:

- Respuesta emocional: genera el comportamiento emocional visceral y la información procesada es enviada a la amígdala para calcular la emoción de esta información.
- Regulación emocional: basada en las técnicas de regulación emocional propuestas en el modelo de Gross [12], se recibe la decisión tomada y acorde a ésta se encuentra un comportamiento emocional apropiado, si la emoción calculada por la amígdala no es la adecuada el proceso de revaloración es activado y se calcula de nuevo la emoción.
- Comportamiento de salida: se concibe a partir del comportamiento emocional y el comportamiento de respuesta, este proceso tiene lugar en la corteza cingulada anterior.

2.2. Sistema motor de control

En el campo de la Neurociencia Aplicada [6] con base en estudios realizados han demostrado que el centro de locomoción del sistema motor está localizado en el lóbulo frontal [18] mismo que se encarga de planificar los movimientos, organizar la secuencia de movimientos y finalmente su ejecución (ver Figura 1).

El sistema motor está integrado jerárquicamente por varias estructuras que están interactuando llevando a cabo la transferencia de información aferente y eferente (ver Tabla 1) para llevar a cabo el movimiento voluntario, es decir, aquel que está dirigido al cumplimiento de un objetivo, con la estructura jerárquica los diferentes niveles se influyen unos a otros, por lo que el flujo de información es tanto en serie (comunicación entre diferentes niveles) como en paralelo (múltiples vías entre cada nivel). En la Figura 2 se pueden identificar las estructuras principales que participan en la función motora, estructuras que se explican enseguida.

Tabla 1. Jerarquía del control motor [3].

Nivel	Función	Estructuras
Alto	Estrategia	Áreas asociadas a la neocortex, ganglio basal
Medio	Tácticas	Corteza motora, cerebelo
Bajo	Ejecución	Tronco cerebral, médula espinal

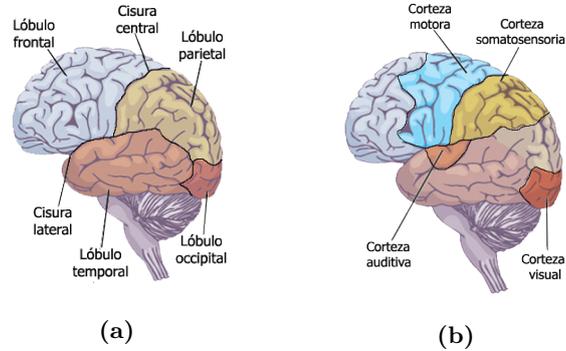


Fig. 1. a) Mapeo de los lóbulos en el cerebro [9]. b) Localización de la corteza motora en el lóbulo frontal [9].

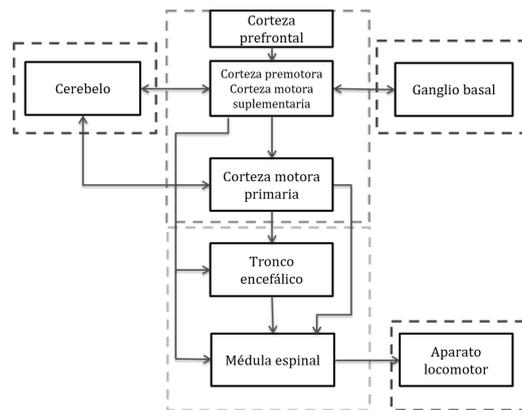


Fig. 2. Circuito del sistema motor [18].

- *Ganglio basal*. Facilita la selección e inhibición de movimientos y es la entrada para iniciar el movimiento.
- *Cerebelo*. Cumple la función de corregir errores en el movimiento comparando las ordenes motoras producidas con la retroalimentación sensorial sobre el movimiento que realmente se produjo, conjuntamente sincroniza los diferentes movimientos.
- *Corteza premotora y Corteza motora suplementaria*. La corteza premotora está involucrada en la selección de los planes motores apropiados para los movimientos voluntarios, también se encarga de seleccionar los programas motores basados en estímulos visuales, por otro parte la corteza motora suplementaria está asociada con escoger los movimientos basados en secuencias de movimientos recordados y en la práctica mental de estos movimientos.

- *Corteza motora primaria*. Abarca la ejecución de movimientos seleccionados por la corteza premotora.
- *Tronco encefálico y médula espinal*. Todos los movimientos son producidos por neuronas motoras de la médula y el tronco encefálico que inervan los músculos. Estas motoneuronas están controladas y coordinadas por el encéfalo, por neuronas de la corteza cerebral y el tronco del encéfalo. Asimismo, la médula espinal aporta al control postural.
- *Sistema locomotor*. Está integrado por los músculos que al contraerse generan el movimiento de los huesos.
- *Corteza parietal posterior*. Procesa las relaciones espaciales de los objetos en el mundo y construye una representación del espacio externo que es independiente de la posición del ojo del observador.
- *Corteza prefrontal*. Realiza la evaluación de una actividad en particular.

Específicamente la corteza prefrontal y la corteza parietal posterior no son áreas motoras en sentido estricto, sin embargo, estas áreas son necesarias para asegurar que los movimientos se adapten a las necesidades del organismo y sean apropiados al contexto conductual.

Existen modelos encaminados a generar movimiento, uno de ellos fue propuesto por Shadmehr y Krakauer [25] el cual aportan que existen 3 situaciones que envuelven la generación de movimiento dirigido a un objetivo, estas situaciones son:

1. *Sistema de identificación*: la necesidad de predecir qué consecuencias sensoriales tendrá nuestro movimiento o comando motor.
2. *Estado de estimación*: realizar una combinación de las predicciones anteriores con la actual retroalimentación sensorial para crear una nueva creencia acerca del estado del cuerpo y el ambiente.
3. *Control óptimo*: con la creencia nueva del estado del cuerpo y del ambiente, se tiene que ajustar de nuevo el ciclo sensomotor de manera que maximice su desarrollo.

En 2009 Shadmehr y Krakauer [25] clasifican las siguientes estructuras cerebrales dirigidas a producir movimiento voluntario en donde cada una desempeña una tarea específica.

- *Ganglio basal*: provee el costo del comando motor y de los beneficios de la predicción del estado sensorial.
- *Cerebelo*: predice las consecuencias sensoriales del comando motor, es decir, provee los cambios propioceptivos y retroalimentación visual.
- *Corteza parietal*: combina la actual retroalimentación sensorial con la retroalimentación sensorial esperada, calculando una nueva creencia.
- *Corteza motora y premotora*: apoyándose de los costos motores y beneficios del estado sensorial previstos asignan una retroalimentación para el estado visual y propioceptivo (sentido que informa al organismo de la posición de los músculos), resultando un cambio a la creencia interna acerca del comando de motor.

De manera semejante Madrigal et al. [19] en 2013, proponen un modelo para la generación de movimiento de una extremidad a partir de un músculo basado en una bio-arquitectura, cuyo objetivo consistió en replicar la forma en que la información eferente (información que viaja del cerebro al órgano receptor) y aferente (información que se transita del órgano receptor hacia el cerebro) es procesada entre las diversas regiones nerviosas produciendo el movimiento de una extremidad.

El modelo de Madrigal mantiene un enfoque jerárquico en el cual las estructuras superiores son responsables de las actividades complejas tales como la planeación y la coordinación de los movimientos sin tomar en cuenta el trabajo hecho por los músculos. Dicho modelo se integró por la corteza motora, ganglio basal, cerebelo, centros del tronco cerebral, neuronas del circuito local, *pools* de neuronas motoras, receptores sensoriales y músculo-esqueléticos. La implementación del modelo se centró en las estructuras inferiores, específicamente en el *pool* de neuronas motoras, con el fin de mostrar el movimiento continuo de una extremidad, mismo que es realizado por la activación y desactivación del *pool* de neuronas motoras. Las pruebas realizadas conducen en que la personificación del movimiento no todas las neuronas motoras pertenecientes al grupo son activadas, puesto que la activación depende del esfuerzo y la velocidad requerida para la realización del movimiento.

3. Propuesta

Esta propuesta está basada en un modelo cognitivo que permita la reproducción de posturas corporales a partir de la regulación emocional, en donde los agentes virtuales podrán adaptarse a las condiciones del ambiente. El modelo está fundamentado en evidencia neurocientífica y psicológica. De acuerdo con las definiciones en neurociencia se presenta el diagrama de conexión por el cual fluye la información como se ilustra en la Figura 3.

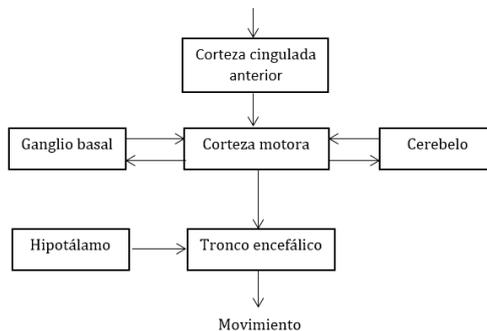


Fig. 3. Modelo cognitivo propuesto para generar posturas corporales con base en las emociones.

El modelo expone como es que cada una de las estructuras del cerebro se van activando para poder generar una postura corporal a partir de la definición de una respuesta emocional con base en un estímulo. Este modelo está orientado a las estructuras superiores del circuito biológico (Figura 2) el cual aportan a la planificación, selección y coordinación de los movimientos.

El modelo parte de un estímulo que ocasionará una respuesta emocional, respuesta emocional que estará asociada a una emoción y a su vez la emoción a un plan con el fin de generar la respuesta motora (función que realiza la corteza cingulada anterior), para llevar a cabo dicho plan es necesario definir el conjunto de movimientos que son requeridos para ejecutar el plan, en seguida se le asignará la dirección y el esfuerzo de cada movimiento. Esta tarea es efectuada por la corteza motora, una vez que ya se tienen mapeados los movimientos es necesario fijar la secuencia de su ejecución.

Posteriormente esta secuencia de movimientos serán producidos por las estructuras inferiores del circuito biológico, dichos movimientos son generados conscientemente, ahora bien, si nos encausamos a las respuestas emocionales viscerales los cuales son inconscientes, el flujo de la información comienza del hipotálamo mismo que tiene una conexión con el tronco encefálico el cual realiza la innervación de los músculos para generar el movimiento, sin embargo antes no existió un plan el cual fuera asociado a los movimientos.

a) *Planificación de la respuesta emocional*

Para la obtención de las señales que representan la entrada a nuestro modelo serán basadas en el modelo cognitivo de Rosales et al. [22], modelo cuyo objetivo fue la generación de comportamientos apropiados en creaturas virtuales a partir de una regulación emocional, en donde el significado del estímulo S es representado $M = \{E_1, E_2, E_3, E_4, E_5, E_6\}$ después de ser sometido a la regulación emocional, cada componente M simboliza cada una de las seis emociones clasificadas por Ekman [8], cuyo valor es de tipo real en un intervalo $0 \leq E_x \leq 1$, personificando de esta manera la intensidad de cada emoción.

De este tipo son las señales recibidas en la estructura de la corteza cingulada anterior que posteriormente se evaluarán con el objetivo de definir a cuál de las emociones básicas tienden el conjunto de señales, o bien, como es que impactan las otras emociones. Para ello se realizará un análisis discriminatorio de las emociones a partir del establecimiento de umbrales de cada emoción.

Después de la determinación de la emoción, ésta será asociada a una postura corporal que representará el objetivo a llegar (dichas configuraciones serán basadas en los estudios de Kleinsmith et al. [15, 16]), la asociación emoción-postura está apoyada por experiencias previas que van enriqueciendo el conocimiento y por consecuencia va moldeando la forma en que los seres humanos tendemos a expresarnos.

b) *Mapeo de movimientos*

Una vez que se tiene definido el objetivo es necesario definir qué partes del cuerpo participan, para ello se trabajará bajo un enfoque homuncular [29]. Es decir, nos apoyaremos en el homúnculo motor de Penfield que se refiere

a un “hombrecillo en el cerebro” que hace referencia a la representación de los mapas corporales en el cerebro, mismo que indica la ubicación y la cantidad de corteza dedicada a cada parte del cuerpo. Las partes con mayor tamaño representa que sus movimientos son más complejos y requieren mayor precisión (por ejemplo, la mano, la boca y los pies), entonces la cantidad de neuronas motoras que estén alojadas en cada parte del cuerpo para la generación del movimiento deberá ser proporcional a su tamaño.

c) *Codificación de la dirección del movimiento*

Ahora bien, con respecto a la definición de la dirección del movimiento de las diferentes partes del cuerpo participes para generar la postura, basándose en la literatura [3, 10, 20] en donde se menciona que la definición de la dirección de los movimientos está determinada por un conjunto de neuronas. Cada una de ellas responde durante el movimiento en cada dirección, la actividad de cada neurona es representada como un vector de dirección en donde cada neurona apunta en la dirección que mejor convenga en la celda (Figura 4a), y la longitud del vector representa el grado de actividad de esa neurona durante el movimiento.

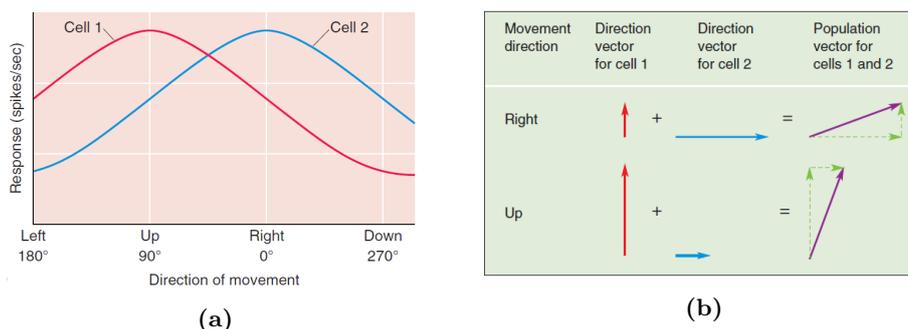


Fig. 4. a) Curvas de ajuste para 2 celdas en la corteza motora [3]. b) Respuesta de cada celda representada como un vector de dirección [3].

Los vectores que representan la actividad de cada neurona se trazarán juntos para cada dirección de movimiento (Figura 4b), luego se promediarán para producir un vector de población, que será representado por la ecuación 1, el cual estima la dirección del movimiento en un tiempo (t) de una población de vectores de dirección $P(t)$, cuyo rendimiento es en cada celda i en la dirección preferida C_i ponderada por la actividad neuronal de la celda V_i :

$$P(t) = \sum_i V_i(t)C_i. \quad (1)$$

La relación entre la tasa de descarga de la neurona y la dirección del movimiento se describirá por una función de ajuste de coseno dada por la ecuación

2, donde b_i y a_i son los coeficientes de regresión y $\theta - \theta_i$ es el ángulo formado por la dirección preferida de la celda y la dirección del movimiento actual M . Cada parte del cuerpo anteriormente mapeada tendrá un vector de población asociado:

$$V_i(M) = b_i + a_i * \cos(\theta - \theta_i). \quad (2)$$

d) *Selección de la secuencia de movimientos*

Una vez que se tienen las partes del cuerpo con su vector de población definido, se puntualizará la secuencia bajo la cual se ejecutarán los movimientos, para eso las diferentes partes del cuerpo estarán subagrupadas para generar movimiento en extremidades superiores e inferiores, tronco y cabeza [23]. A cada integrante del subgrupo se le ligará un peso el cual estará generado bajo la jerarquía de las articulaciones, con ello se poseerá un mapa de movimiento para ejecutarlo con las estructuras inferiores que participan en la función motora.

La Figura 5 muestra el esqueleto asociado al agente que deberá buscar la postura corporal adecuada mediante la respuesta emocional a partir de la percepción del ambiente.

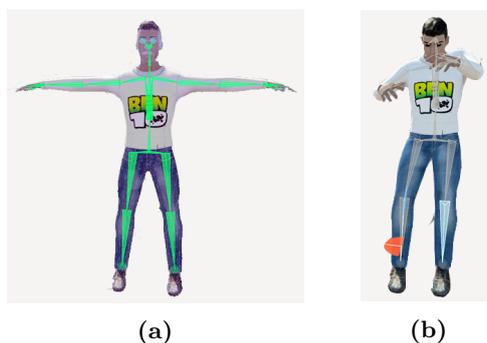


Fig. 5. a) Planificación de la respuesta emocional, mapeo de movimientos, codificación del esfuerzo y dirección del movimiento. b) Selección de la secuencia de movimientos.

3.1. Resultados preliminares

Los resultados obtenidos hasta este momento permiten observar el tratamiento de información aferente que se realiza en cada estructura cerebral, la cual está generando que el agente virtual adopte una postura corporal. Actualmente se trabajan con cuatro emociones: alegría, tristeza, miedo y enojo, se eligieron estas cuatro emociones ya que son la base de comparación con resultados de trabajos anteriormente citados, además la posibilidad de confundir una emoción

con respecto a otra a partir de la postura corporal es remota, en otras palabras, hay una mayor facilidad de discriminar una emoción con relación a otra pues el umbral para el discernimiento es considerable.

Es necesario validar que el agente virtual adquiera las posturas corporales adecuadas con respecto a su estado emocional actual, mismo que es inducido por una retroalimentación del ambiente y que permita transmitir este lenguaje al usuario y a otros agentes con los que comparte el ambiente, como se puede apreciar en las Figuras 6a, 6b, 6c y 6d. Para ello será necesario embeber al agente en un ambiente virtual, en donde el usuario tenga la experiencia de interactuar con el agente, el agente se apoyará de las posturas corporales para enfatizar el dialogo que se establezca. Es importante destacar que la postura corporal que exprese una emoción no debe ser siempre la misma pues puede existir un marco de variación de acuerdo a la secuencia de movimientos musculares y a la amplitud del ángulo entre las articulaciones, esta variación estará a juicio del usuario pues es quien discriminará la emoción que el agente pretende expresar.

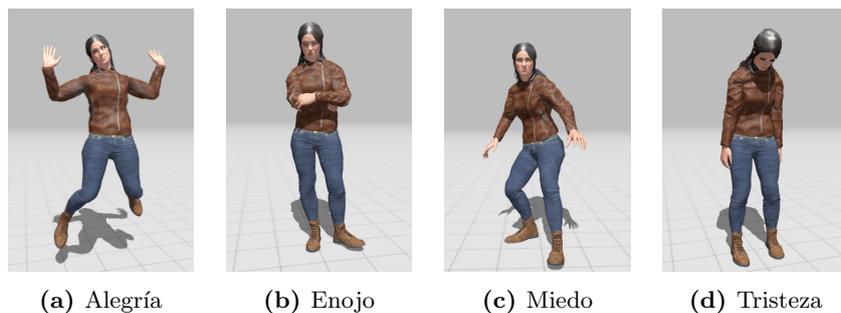


Fig. 6. Posturas corporales a partir de una respuesta emocional.

Para poder definir que una postura es semejante a otra nos apoyaremos de la definición de una distancia máxima que podrá variar con respecto a la postura objetivo, esta distancia se obtendrá de las articulaciones que enfatizan la postura corporal.

4. Conclusiones y trabajos futuros

La integración de posturas corporales en agentes virtuales brindará mayor credibilidad a la expresión de sus estados emocionales además de atraer por un mayor intervalo de tiempo la atención del usuario, de esta manera se estará reforzando el proceso de interacción que existe entre el usuario y la computadora.

De igual manera, la integración del modelo cognitivo fundamentado en evidencia neurocientífica y psicológica permite que el proceso interno que se lleva a cabo para producir una postura sea semejante al del ser humano, y no solo la respuesta externa que el usuario percibe, representando un respaldo para

el comportamiento emocional, dado que ayudará a entender el porqué de la respuesta motora.

Es necesario la integración de parámetros que describan la personalidad del agente, personalidad que influirá en la evaluación del estímulo para determinar qué emoción la está produciendo, ya que experiencias previas representarán el detonante para la formación de la personalidad. También la incorporación de contextos bajo los cuales los agentes virtuales estarán inmersos para efectuar la ampliación del marco de las emociones, apoyará al discernimiento de las emociones conforme la postura que adopte el agente, reduciendo de esta manera el porcentaje de confusión entre las emociones sorpresa-miedo y enojo-desagrado.

Referencias

1. AIVO: Agentbot (2016), <http://agentbot.net/es>, accedido el 16-01-2017
2. AlterEgo: Proyecto alterego (2012), <http://www.euromov.eu/alterego/>, accedido el 02-02-2017
3. Bear, M.F., Connors, B.W., Paradiso, M.A.: *Neuroscience exploring the Brain*. Lippincott Williams & Wilkins, United States of America, third edn. (2007)
4. Bevacqua, E., Richard, R., Soler, J., Loor, P.D.: *INGREDIBLE: A platform for full body interaction between human and virtual agent that improves co-presence*. Archive ouverte HAL (Jul 2016)
5. Caballero Lambert, A.O., Torres Montañez, C.H.: *Asistente virtual para el diagnóstico de trastornos alimentarios usando un motor cognoscitivo general: Nutmin*. In: *iaedp Symposium* (2016)
6. Cardinali, D.P.: *Neurociencia aplicada. Sus fundamentos*. Editorial medica pan-americana, Buenos Aires (2007)
7. Dael, N., Mortillaro, M., Scherer, K.R.: *Emotion Expression in Body Action and Posture*. *American Psychological Association* 12(5), 1085–1101 (2012)
8. Ekman, P., Friesen, W.V.: *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*. Malor Books (2015)
9. EMAZE: *Sistema nervioso* (Dec 2014), <https://www.emaze.com/@A0IILCFI/sistema-nervioso-pa>, accedido el 28-03-2017
10. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: *Neurona population coding of movement direction*. *Science* 233, 1416–1419 (Sep 1986)
11. Gómez Cancino, C.K.: *Psiquiatras virtuales para tratamiento remoto de pacientes con depresión* (Aug 2016), <http://www.conacytprensa.mx/index.php/ciencia/salud/9796-psiquiatras-virtuales-para-tratamiento-remoto-de-pacientes-con-depresion>, accedido el 07-01-2017
12. Gross, J.: *The emerging field of emotion regulation: An integrative review*. *Peer Reviewed Journal* 2(3), 271–299 (1998)
13. Hernando Rosales, J., Jaime, K., Ramos, F., Ramos, M.: *An Emotional Regulation Model with Memories for Virtual Agents*. In: *12th IEEE International Conference on Cognitive Informatics & Cognitive Computing* (2013)
14. Jack, R.E., Garrod, O.G.B., Schyns, P.G.: *Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time*. *Current Biology, Elsevier* 24, 187–192 (Jan 2014)

15. Kleinsmith, A., Bianchi-Berthouze, N., Silva, P.R.D.: Cross-Cultural Differences in Recognizing Affect from Body Posture. Elsevier Science 18(6), 1371–1389 (Dec 2006)
16. Kleinsmith, A., Bianchi-Berthouze, N., Steed, A.: Automatic Recognition of Non-Acted Affective Postures. IEEE Transactions on Systems, MAN, And Cybernetics-Part B: Cybernetics 41(4), 1027–1038 (Aug 2011)
17. Kleinsmith, A., De Silva, P.R., Bianchi-Berthouze, N.: Grounding Affective Dimensions into Posture Features. Springer 3784, 263–270 (2005)
18. Kolb, B., Whishaw, I.Q.: Neuropsicología humana. Editorial medica panamericana, Argentina, quinta edn. (2006)
19. Madrigal, D., Torres, G., Del Valle, J.L., Ramos, F.: A Bio-Inspired Architecture of a Motor Neuron System for Virtual Creatures: Movement of a Single Limb with a Single Muscle. Procedia Computer Science 22, 440–449 (2013)
20. The National Academy of Sciences of the USA: Mapping of the preferred direction in the motor cortex, vol. 104 (Jun 2007)
21. Ostrosky, F., Vélez, A.: Neurobiología de las emociones. Revista Neuropsicología, Neuropsiquiatría y Neurociencias 13(1), 1–13 (Jan 2013)
22. Rosales, J.H., Ramos, F., Ramos, M.: Search for an appropriate behaviors within the emotional regulation in virtual creatures using learning classifier (2017)
23. Rosell Puig, W., González Fano, B., Dovale Borjas, C., Domínguez Hernández, L.: División regional del cuerpo humano para facilitar su estudio. Diferencias entre las regiones superficiales y esqueléticas. Educación Médica Superior 20(3) (Jul 2006)
24. Róspide, C.G., Puente, C.: Agentes virtuales inteligentes aplicados a un entorno educativo. Pensamiento matemático 2(2), 195–208 (Sep 2012)
25. Shadmehr, R., Krakauer, J.W.: The cognitive neurosciences, chap. V. Motor Systems. Computational neuroanatomy of voluntary motor control, pp. 587–597. Massachusetts Institute of Technology, fourth edition edn. (2009)
26. Shafir, T.: Using movement to regulate emotion: Neurophysiological findings and their application in psychotherapy. Frontiers in Psychology (Sep 2016)
27. Shafir, T., Tsachor, R.P., Welch, K.B.: Emotion regulation through movement: Unique sets of movement characteristics are associated with and enhance basic emotions. Frontiers in Psychology (Jun 2016)
28. Torres Blanco, G.A.: Metodología para el desarrollo de arquitecturas cognitivas basada en neurociencias. Ph.D. thesis, Cinvestav del IPN Unidad Guadalajara, México (2013)
29. Won, A.S., Bailenson, J., Lee, J., Lanier, J.: Homuncular flexibility in virtual reality. Journal of Computer-Mediated Communication 20, 241–259 (May 2015)
30. Wooldridge, M.: An Introduction to MultiAgent Systems. UK: John Wiley & Sons, ltd (2009)

Identificación de características de células de cáncer de mama por medio de testores típicos

Alexis Gallegos, Dolores Torres, Francisco Álvarez, Aurora Torres

Universidad Autónoma de Aguascalientes, Departamento de Ciencias de la Computación,
Aguascalientes, Aguascalientes, México

alexisEdm@gmail.com, mdtorres@correo.uaa.mx, fjalvar@correo.uaa.mx, atorres@correo.uaa.mx

Resumen. Una de las preocupaciones más importantes del mundo ha sido la salud humana, especialmente en enfermedades como el cáncer. Por esta razón, este artículo se enfoca en la aplicación de las Ciencias Computacionales, específicamente, la Selección de Subconjuntos de Características y Testores típicos para mejorar el diagnóstico de cáncer. En este caso, se procesó una base de datos de cáncer de mama. Ésta base de datos fue publicada por la Universidad de California para aprendizaje máquina. Los datos describen características del núcleo de las células obtenido de una imagen digitalizada de un aspirado con aguja fina de masa mamaria clasificando cada célula como maligna o benigna. Finalmente, el método proveerá el peso informacional de cada característica. Ésta información permitirá saber si una característica realmente describe una célula y así, clasificar nuevas instancias con la información correcta.

Palabras clave: peso informacional, testor típico, selección de subconjuntos, cáncer de mama, lógica combinatorial.

Identification of Breast Cancer Cell Features by Means of Typical Testors

Abstract. One of the most significant concerns around the world has been human health, especially in diseases such as cancer. For this reason, this paper is focused on the application of Computer Science, specifically, Feature Subset Selection and Typical Testors to improve the diagnosis of cancer. In this case, a breast cancer cell database was processed. This database was published by the University of California for machine learning. The data describes features of the cell nuclei obtained from a digitized image of a fine needle aspirate of a breast mass classifying every cell as malignant or benign. Finally, the method will provide the informational weight of each feature. This information will let know if a feature actually describes a cell and then, classify new instances with the right data.

Keywords: informational weight, typical testor, subset selection, breast cancer, combinatorial logic.

1. Introducción

El cáncer de mama es el cáncer más común y la principal causa de muerte por enfermedad tumoral en mujeres alrededor en el mundo [1], lo que representa el 16% de los cánceres en mujeres [2]. Hoy en día, el cáncer se ha vuelto cada vez más difícil de ignorar. Cada día, nuevos estudios sobre las causas y tratamientos son publicados; sin embargo, todo coinciden que el punto crítico de estos estudios es la detección temprana [3].

De acuerdo con el Instituto Nacional de Cáncer [4], la detección de cáncer significa la comprobación de cáncer o de condiciones que pueden convertirse en cáncer en personas que no presentan síntomas.

La detección temprana es importante debido a que cuando un tejido anormal o cáncer es encontrado a tiempo, puede ser más fácil de tratar. Al momento que los síntomas aparecen, el cáncer ha comenzado a extenderse y es más difícil de tratar [4].

A pesar de la utilidad de la detección temprana, pueden existir algunos riesgos, así como los métodos utilizados. Por ejemplo, una prueba de detección puede presentar resultados falsos positivos; significa que la prueba indica la presencia de cáncer cuando no es verdad. Por otro lado, la prueba puede tener resultados falsos negativos indicando que el cáncer no está presente, aunque si lo este.

Por otra parte, el sobrediagnóstico es posible, el cual sucede cuando la prueba de detección muestra que una persona tiene cáncer, pero el cáncer es de crecimiento lento y no habría perjudicado a la persona en toda su vida [4]. Lo anterior justifica la necesidad de mejorar el diagnóstico de cáncer.

El diagnóstico clínico es un proceso cognitivo que parte del pensamiento concreto sensible. Está relacionado con la realidad objetiva; se desarrolla en el pensamiento abstracto y tiene el criterio de verdad en la práctica [5]. Involucra práctica, experiencia, reconocimiento de patrones y cálculo de probabilidad condicional, entre otros componentes. Sin embargo, el diagnóstico tiene tratamiento humano, por lo tanto, no está libre de errores que pueden causar enfermedad, daños, gastos extra e incluso la muerte, especialmente en enfermedades sensibles como el cáncer [6].

Los errores representan un estimado de 150 de cada 1000 pacientes con diagnóstico erróneo [7]. Por esta razón, el campo de la medicina es una de las áreas que pueden beneficiarse mejor de una interacción cercana con las Ciencias Computacionales y las Matemáticas para mejorar procesos como lo es el diagnóstico médico [6]. Siendo así, la razón por la que se decide aplicar métodos matemáticos integrales para apoyar el diagnóstico de enfermedades como el cáncer, en este caso, cáncer de mama.

Este artículo está dividido en tres secciones y organizado de la siguiente manera. La primera sección trata conceptos importantes en Selección de Subconjuntos de Características y testores típicos en Ciencias Computaciones, el cáncer de mama y su impacto en el mundo. La sección siguiente examina marco de trabajo del análisis, siendo una revisión de la metodología aplicada a las células de cáncer de mama. Finalmente, la tercera sección describe los resultados de la metodología y su revisión.

2. Conceptos importantes

2.1. Selección de subconjuntos de características

Normalmente, la Selección de Subconjuntos de Características (FSS, por sus siglas en inglés: Feature Subset Selection) [8] es usado para reducir la dimensionalidad [9], lo que significa que reduce el número de variables, atributos o características con las cuales se describen los objetos y encontrar su influencia en un problema. Este un método alternativo que inicia usando el conjunto de testores típicos, descartando características irrelevantes o redundantes [9, 10].

La importancia de la FSS recae en la reducción del número de características, el cual puede ayudar a disminuir el costo de adquisición de información y hacer que los modelos de clasificación sean más fáciles de entender [9, 11]. Además, el número de características podría afectar la precisión de la clasificación. Algunos autores también han estudiado la Selección de Subconjuntos de Características para el aprendizaje de clasificación [9].

Los problemas de FSS han sido estudiados con gran atención por estadísticos y comunidades de aprendizaje máquina durante años debido a la investigación entusiasta de la minería de datos [12]. Existen muchos beneficios potenciales de la selección de características, como lo son [13]:

- Facilita la visualización de información y su entendimiento,
- Reduce requerimientos de medición y almacenamiento,
- Reduce tiempos de capacitación y utilización,
- Reduce la dimensionalidad para mejorar el rendimiento de la predicción.

La selección de las variables más relevantes suele ser subóptima para construir un predictor, sobre todo si las variables son redundantes [13]. Por ejemplo, el método de selección por fuerza bruta evalúa exhaustivamente todas las posibles combinaciones de las características de entrada y así encontrar el mejor subconjunto [12]. Más adelante, en las secciones 3 y 4 describirán el método de fuerza bruta aplicado a la base de datos ya mencionada con el objetivo de evaluar si una célula es maligna o benigna y calcular el peso informacional de cada característica.

2.2. Testores típicos

La teoría de testores fue formulada como una dirección científica independiente de Cibernética Matemática en los años 60 en la formada Unión de Repúblicas Socialistas Soviéticas (USSR), cuyo origen está vinculado con el uso de lógica matemática para localizar fallas en circuitos electrónicos que realizan funciones booleanas [14].

Más tarde, los testores fueron utilizados para realizar clasificación supervisada y selección de variables en problemas de geología [14, 15]. El uso de datos a los testores y testores típicos para este artículo está relacionado con la Selección de Subconjuntos de Características, cuyos precursores son Dmitriev, Zhuravlev, y colegas [15].

De este modo, un testor es un subconjunto de características que distingue objetos de diferentes clases [15]. De acuerdo con Santiesteban y Pons [10], Shulcloper [14] y

Torres [15], un testor típico es un testor al que no es posible eliminar alguna característica sin perder su estado de testor. En otras palabras, un testor típico ya está formando por el conjunto mínimo de características necesarias para asegurar la identificación de la clase a la que pertenece un objeto específico.

Los testores típicos determinan cuestiones como la evaluación del peso informacional de los rasgos y la selección de variables. Pueden reducir el espacio de representación de los objetos [10] y pueden ser usados como un conjunto de soporte para la algoritmos de clasificación [16]. En consecuencia, el objetivo de este estudio es probar que el análisis de testores puede ayudar a clasificar las células basadas en un conjunto de datos real; esto se explicara en la sección 3.

2.3. Peso informacional

El uso del peso informacional para Selección de Subconjuntos de Características es una excelente herramienta que muestra resultados tangibles [9]. El peso informacional es una puntuación, es decir, es una medida de significancia para predecir si un objeto pertenece a un grupo o a otro (clasificación) [15, 17]. Más información en la sección 4.

2.4. Cáncer de mama

El cáncer es una colección de enfermedades relacionadas que causa que algunas células del cuerpo comiencen a dividirse sin detenerse y se extiendan a tejidos cercanos. El cáncer puede generarse en casi cualquier parte del cuerpo, el cual está hecho de millones de células [18]. Se trata del resultado de mutaciones o cambios anormales en los genes que regulan el crecimiento de la célula. Normalmente, una célula crece y divide para formar nuevas células según como el cuerpo lo necesite. Cuando las células envejecen y resultan dañadas, mueren, y nuevas células las reemplazan [18, 19].

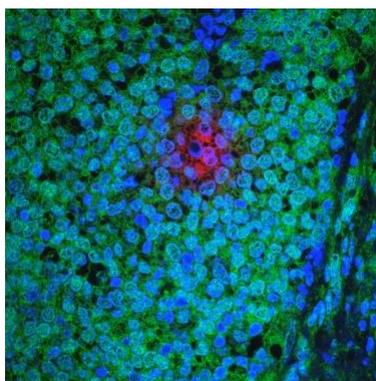


Fig. 1. Tumor invasivo de cáncer de mama [18].

Cuando el cáncer se desarrolla, el proceso celular se descompone. Las mutaciones pueden “activar” ciertos genes y “desactivar” otras en la célula. La célula modificada

adquiere la habilidad de dividirse sin ningún control u orden, lo que produce células idénticas y generan un tumor [18, 19].

En consecuencia, el cáncer de mama es un tumor maligno que se ha desarrollado de células de mama [20]. La mama está hecha de glándulas llamadas lóbulos que pueden producir leche y tubos delgados llamados ductos que llevan leche de los lóbulos al pezón, generalmente, el cáncer de mama se origina en las células de éstos lóbulos [19, 20].

El cáncer de mama tiene gran impacto en el mundo. Según la Organización Panamericana de Salud (PAHO), en América el cáncer de mama es el más común en mujeres con el 29% de los casos de cáncer. PAHO estima más de 596,000 casos nuevos y más de 142,100 muertes en la región para 2030, principalmente en Latinoamérica y el Caribe [21]. La siguiente figura muestra la incidencia de tumores malignos de mama en mujeres mayores a 20 años divididos por grupo de edad, en el año 2014:

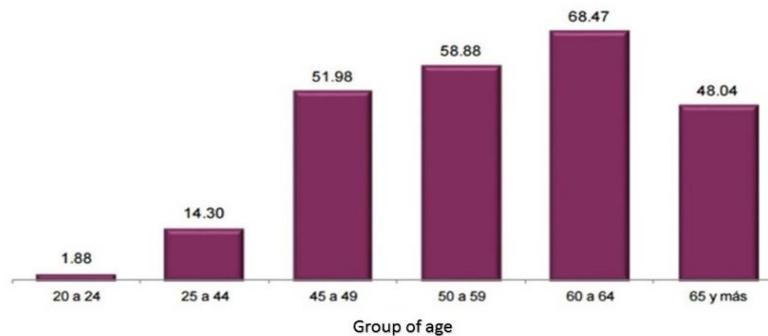


Fig. 2. Incidencia de tumores malignos de mama en mujeres mayores de 20 años dividido por grupo de edad. Por 100 mil mujeres por grupo de edad. INEGI [21].

En general, cualquier tipo de cáncer representa un impacto importante en el estado físico de la persona, su esfera emocional, un alto costo de tratamiento y puede incluso, socavar la economía de los países; así que la prevención y el diagnóstico temprano son críticos para abordar el problema [22]. Por lo tanto, este trabajo se concentra en la aplicación del Selección de Subconjuntos de Características y Testores Típicos para mejorar el diagnóstico de cáncer en células de cáncer. Las secciones siguientes se explicará el estudio realizado.

3. Marco de trabajo

La metodología general utilizada (ver Fig. 3) inicia con una Matriz de Aprendizaje (MA). La MA es la fuente de información que contiene la descripción de los objetos [14, 23]. Para éste trabajo, la MA viene de la Universidad de California y su Repositorio de Aprendizaje Máquina. La base de datos seleccionada es Diagnóstico de Cáncer de Mama de Wisconsin [24].

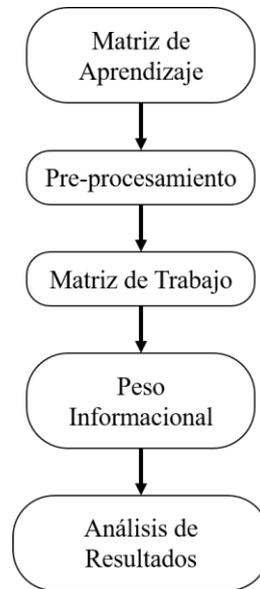


Fig. 3. Metodología general.

La base de datos contiene el diagnóstico y 10 características obtenidas de una imagen digitalizada de una aspiración de tejido de mama con aguja fina y describe las características del núcleo de la célula presentada en la imagen [25]. La imagen a continuación muestra un ejemplo de las imágenes descritas.

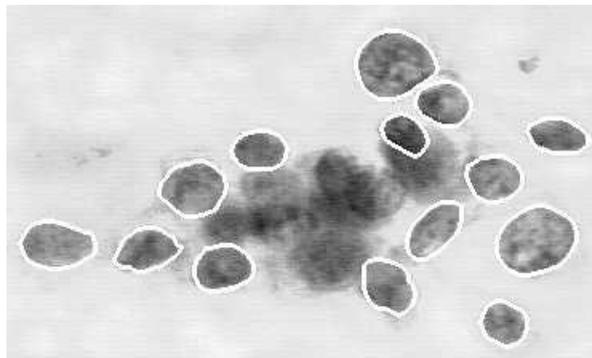


Fig. 4. Ejemplo de una imagen tomada por un sistema de visión por computadora y el contorno de la célula [26].

Las características evaluadas para cada núcleo de célula son [25, 26]:

1. Diagnóstico (M=maligno, B=benigno),
2. Radio,
3. Textura,
4. Perímetro,

5. Área,
6. Suavidad,
7. Compacidad,
8. Concavidad,
9. Puntos cóncavos,
10. Simetría,
11. Dimensión fractal.

El diagnóstico es el resultado final de la evaluación de las características de la célula con un sistema de diagnóstico de visión por computadora [26-28]. Cada célula en la base de datos tiene uno de dos posibles diagnósticos, puede ser célula maligna registrada con la letra M o benigna registrado con la letra B.

El radio de la célula fue medido promediando la longitud de los segmentos de líneas radiales definidos por el centroide de la célula y los puntos individuales en el límite de la célula. Las líneas radiales fueron definidas por Street, Wolberg y Magasarian en [26, 27] como se puede observar en la Fig. 5.

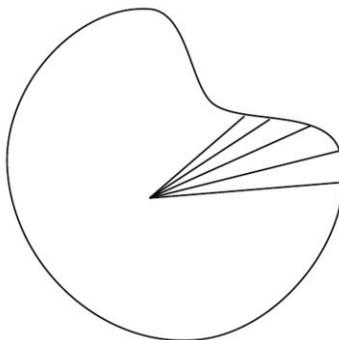


Fig. 5. Líneas radiales medidas en una célula [26].

Como se mencionó anteriormente, cada característica de la célula fue extraída por sistema de visión por computadora, por tanto, la textura fue medida encontrando la varianza en intensidades de escala de grises en los pixeles de la computadora [26, 27]. (Ver Fig. 4).

El perímetro es definido como la distancia total entre puntos individuales llamados puntos serpiente en [26]. Estos puntos individuales comprenden las líneas blancas en el perímetro de las células (ver Fig. 4).

El área es obtenida contando el número de pixeles en el interior de la línea blanca añadiendo la mitad de los pixeles en el perímetro [26].

Mientras tanto, la suavidad del núcleo de la célula se calcula midiendo la diferencia entre la longitud de una línea radial y la longitud principal que la rodea [26]. Básicamente, la suavidad es la variación local en las longitudes de radio [25].

El perímetro y el área son combinados para calcular la medida de compacidad; la cual es una medida de forma [26, 27]. La compacidad está dada por la fórmula:

$$\text{Compacidad} = \frac{\text{perímetro}^2}{\text{área}}$$

Este número es minimizado por un disco circular e incrementa con la irregularidad del perímetro y aumenta también para núcleos celulares alargados, lo que puede indicar mayor probabilidad de malignidad [26].

La concavidad analiza las irregularidades de forma en el núcleo de la célula. Street, Wolberg y Mangasarian miden el número y la severidad de las concavidades y hendiduras en el núcleo de la célula. Ellos dibujan cuerdas entre cada punto blanco no adyacente y miden hasta qué punto el límite real del núcleo se encuentra en el interior de cada cuerda (ver Fig. 6).



Fig. 6. Cuerdas usadas para calcular la concavidad [27].

Los puntos cóncavos usan una medida similar a la concavidad, pero ésta característica solo mide el número, más que la magnitud, de las concavidades del contorno [26].

La simetría se obtiene encontrando la línea más larga que pase por el centro. Entonces, de acuerdo con [26], se trazan líneas perpendiculares a dicha línea para medir la diferencia de longitudes en las dos direcciones de la lineal central (ver Fig. 7).

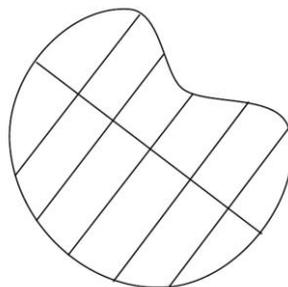


Fig. 7. Segmentos usados en el cálculo de la simetría [26].

Finalmente, la dimensión fractal es una característica de forma [27], es decir, a mayor valor corresponde a un menor contorno y por tanto a una mayor probabilidad malignidad [26]. La dimensión fractal se aproxima usando la aproximación de costa de Mandelbrot [26, 29]. El perímetro del núcleo es medido usando “reglas” cada vez más grande. Esto es, a medida que aumenta el tamaño de la regla, decrece la precisión de la medición, el perímetro observado disminuye. Ahora, trazando estos valores a una escala

logarítmica y medir la pendiente descendente de el negativo de una aproximación de la dimensión fractal [26] (ver Fig. 8).

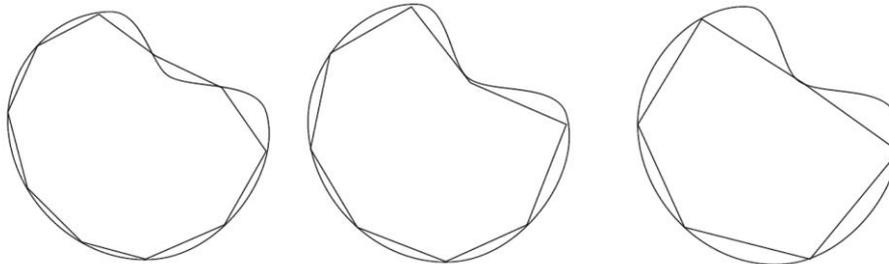


Fig. 8. Secuencia de medidas para calcular dimensión fractal [26].

La base de datos contiene un total de 569 instancias, 357 de ellas son instancias benignas y 212 instancias malignas. Esta base de datos fue pre-procesada (ver Fig. 3), esto significa que es necesario un análisis profundo de la base de datos buscando instancias duplicadas en cada clase y contradicciones (eliminar registros iguales pero diagnósticos diferentes). Consecuentemente, la base de datos debe contener instancias únicas.

El siguiente paso de la metodología requiere una matriz de trabajo, la cual se obtiene de matriz de aprendizaje pre-procesada. La matriz de trabajo se compone de datos discretizados. Cada característica es discretizada de acuerdo a la literatura del problema y el consejo de un experto, quién confirma los criterios de comparación [10].

Finalmente, el elemento principal de la metodología es el peso informacional. Para calcularlo es necesario aplicar la teoría de testores típicos mencionada en la sección 2.2. Como resumen, los testores típicos están formados por el conjunto mínimo necesario para asegurar la identificación de clases en la que un objeto específico pertenece. Para más información ver las referencias [10, 14].

4. Resultados y conclusiones

Al final del proceso, el peso informacional es calculado de acuerdo con los testores típicos encontrados. Como se puede observar en la Tabla 1, el radio y el área del núcleo de la célula tienen 50% de peso informacional. Esto significa que es posible clasificar una instancia de célula conociendo a menos una de las dos características, el radio o el área, pero el resto de las características debe conocerse debido a que obtuvieron un 100%. Por ejemplo, una instancia puede ser clasificada conociendo su textura, perímetro, suavidad, compacidad, puntos cóncavos, simetría y la dimensión fractal, pero si el radio es desconocido, el área debe conocerse. Por otro lado, si se desconoce el área, el radio debe conocerse. Finalmente, en el mejor caso se da cuando ambos valores se conocen mientras que no es posible clasificar una célula si ambos datos son desconocidos.

Table 1. Peso informacional de acuerdo con lo testores típicos.

Feature	Informational weight
Radius	50%
Texture	100%
Perimeter	100%
Area	50%
Smoothness	100%
Compactness	100%
Concave points	100%
Symmetry	100%
Fractal dimension	100%

El peso informacional se obtiene calculando un factor de porcentaje que indica la frecuencia de cada variable en el conjunto de testores típicos [30]. El valor del peso informacional representa el grado de importancia de cada característica analizada en un proceso de clasificación. Un valor de 100% indica que la característica es crítica no puede ser ignorada en ningún caso.

Además, es posible que una o más características obtengan 0% de peso informacional, lo que significa que no es necesaria. Por lo tanto, el número de características se reduce y hace el problema más sencillo. Recuerde que éste es uno de los objetivos del análisis.

El peso informacional puede ser validado por la teoría del problema o un especialista, de manera que la información final se apegue a la realidad. Para este experimento, un patólogo validó el peso informacional y el comportamiento de los datos.

Referencias

1. Guerra-Merino, I.: Factores pronóstico del cáncer de mama en 108 mujeres menores de 36 años. Universidad Complutense de Madrid (2000)
2. CEAMEG: Cancer de Mama. Vol. 1, No. Cancer de Mama, pp. 1 (2014)
3. Canceronline: Detección Precoz de Cáncer. Available: http://www.canceronline.cl/index.php?option=com_content&view=article&id=48&Itemid=57
4. NIH: Cancer Screening. Available: <http://www.cancer.gov/about-cancer/screening> (2015)
5. Pérez-Guirado, N. M.: El diagnóstico médico: algunas consideraciones filosóficas (2009)
6. Lugo-Reyes, S. O., Maldonado-Colín, G., Murata, C.: Inteligencia artificial para asistir el diagnóstico clínico en medicina. *Artificial Intelligence to Assist Clinical Diagnosis in Medicine*, Vol. 61, No. 2, pp. 110–120 (2014)
7. Reed, K.: HealthGrades Patient Safety in American Hospitals Study. Disponible en: <https://www.hospitals.healthgrades.com/>
8. Wang, G., Song, Q., Sun, H., Zhang, X., Xu, B., Zhou, Y.: A Feature Subset Selection Algorithm Automatic Recommendation Method. *China Journal of Artificial Intelligence Research* (2013)

9. Torres, D., Ponce de León, E., Torres, A., Ochoa, A., Díaz, E.: Hybridization of Evolutionary Mechanisms for Featured Subset Selection in Unsupervised Learning. In: MICAI 2009, Advances in Artificial Intelligence, pp. 610–621
10. Santiesteban-Algaza, Y., Pons-Porrata, A.: LEX: A New Algorithm for the Calculus of all Typical Testors. Vol. 1, pp. 85–95.
11. Pelikan, M., Sastry, K., Cantú-Paz, E.: Scalable Optimization vía Probabilistic Modeling: From Algorithms to Applications. Springer (2006)
12. Deng, K.: OMEGA: On-line Memory-Based General Purpose System Classifier. Doctor Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA (1998)
13. Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, Vol. 3 (2003)
14. Ruíz-Shucloper, J., Alba-Cabrera, E., Lazo-Cortés, M.: Introducción a la Teoría de Testores. Departamento de Ingeniería Eléctrica, CINVESTAV-IPN, pp. 197 (1995)
15. Torres-Soto, M. D., Torres-Soto, A., Torres-Soto, L., Bermudez-Rosales, L., Ponce de León-Sentí, E. E.: Factores Predisponentes en Relajación Residual Neuromuscular. Research in Computing Science, Vol. 93, pp. 163–174 (2015)
16. Lias-Rodríguez, A., Pons-Porrata, A.: Un nuevo Algoritmo de Escala Exterior para el Cálculo de los Testores Típicos. Research in Computing Science, Vol. 93 (2015)
17. Cotilla, M. O.: Un Recorrido por la Sismología de Cuba. 1 ed., Cuba, Editorial Complutense, S. A. (2006)
18. National Cancer Institute: What is Cancer? Disponible en: <http://www.cancer.gov/about-cancer/understanding/what-is-cancer> (2015)
19. Breastcancer.org: ¿Qué es el Cáncer de mama? Disponible en: <http://www.cancer.gov/about-cancer/understanding/what-is-cancer> (2014)
20. NIH: Breast Cancer - Patient Version. National Cancer Institute
21. INEGI: Estadísticas a Propósito del... Día Mundial de la Lucha contra el Cáncer de Mama. Estadísticas Nacionales, México: Instituto Nacional de Estadística y Geografía (2015)
22. INEGI: Estadísticas a Propósito del Día Mundial Contra el Cáncer. México, Instituto Nacional de Estadística y Geografía (2016)
23. Santiesteban, Y., Pons, A.: LEX: un nuevo algoritmo para el calculo de los testores tipicos. Revista Ciencias Matematicas, Vol. 21, No. 1, pp. 85–95 (2003)
24. Mangasarian, O. L., Street, W. N.: Breast cancer diagnosis and prognosis via linear programming. Operations Research, Vol. 43, No. 4, pp. 570 (1995)
25. Wolberg, W. H., Street, N., Mangasarian, O. L.: Wisconsin Diagnostic Breast Cancer (WDBC). California, Ed., USA (1995)
26. Street, W. N., Wolberg, W. H., Mangasarian, O. L.: Nuclear Feature Extraction for Breast Tumor Diagnosis. In: International Symposium on Electronic Imaging, Science and Technology, Vol. 1905, pp. 861–870
27. Wolberg, W. H., Street, W. N., Heisey, D. M., Mangasarian, O. L.: Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. Archives of surgery (Chicago, Ill.: 1960), Vol. 130, No. 5, pp. 511 (1995)
28. Wolberg, W. H., Street, W. N., Heisey, D. M., Mangasarian, O. L.: Computer-derived nuclear features distinguish malignant from benign breast cytology. Human Pathology, Vol. 26, No. 7, pp. 792–796 (1995)
29. Mandelbrot, B. B.: The fractal geometry of nature. New York, W.H. Freeman (1982)

Alexis Gallegos, Dolores Torres, Francisco Álvarez, Aurora Torres

30. Rodríguez-de León, P.: Heurística lógico combinatoria para la selección de subconjuntos de características en diabetes mellitus. Tesis (maestría en informática y tecnologías computacionales), Universidad Autónoma de Aguascalientes, Centro de Ciencias Básicas, Aguascalientes, Ags., México (2016)

Propuesta de reconstrucción 3d de objetos basado en la selección de los mejores parámetros de la Wavelet Morlet

Claudia Victoria López¹, María Eugenia Quintanar Pérez³,
Jesús Carlos Pedraza Ortega², Gonzalo Elías Blanco Silva¹,
Juan Manuel Ramos Arreguín², Efrén Gorrostieta Hurtado²

¹ Facultad de Informática, Universidad Autónoma de Querétaro, Querétaro, México

² Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro, México

³ Instituto Tecnológico de Querétaro, Querétaro, México

azul.cielo.2007@gmail.com, mequintanar@hotmail.com, caryoko@yahoo.com,
gblanco25@alumnos.uaq.mx, jramos@mecamex.net, efrengorrostieta@gmail.com

Resumen. Entre los métodos de proyección de luz estructurada para el proceso de reconstrucción 3D, se encuentran el de Perfilometría de Fourier y Perfilometría Wavelet, siendo el primero el más utilizado. Este trabajo muestra una comparativa entre el desempeño de la Perfilometría Wavelet y la de Fourier, implementando la selección de los mejores parámetros para la wavelet Morlet, considerando el análisis de la estimación de fase mediante la frecuencia espacial f_0 , la cual es establecida por el número de franjas y sirve como parámetro para llevar a cabo el filtrado en esa frecuencia dentro de la Wavelet Morlet. La metodología aplicada utiliza la selección de objetos reales y virtuales, así como la determinación de la frecuencia espacial en caso de no tenerla como parámetro de entrada. Se llevaron a cabo pruebas primero con objetos virtuales y se obtuvieron los parámetros en donde el error en la reconstrucción 3D es menor. Los resultados obtenidos con estos parámetros fueron implementados para la reconstrucción 3D en objetos reales.

Palabras clave: reconstrucción 3D, wavelet Morlet, parámetros, perfilometría de Fourier.

3D Object Reconstruction Proposal Based on the Morlet Wavelet best parameter selection

Abstract. Among the structured light projection methods for the 3D reconstruction process, there are the Fourier and Wavelet based Profilometry, being Fourier the most used. In this work we present a comparative between Fourier and Wavelet Profilometry performance by means of the selection of the best parameters of the Morlet Wavelet. Analyzing the phase estimation by using the f_0 spatial frequency, which is established by the fringe numbers and serves as

a parameter to perform the frequency filtering with the Morlet Wavelet. The proposed methodology is applied and it uses the selection of virtual and real objects, later, the spatial frequency is determined and in case of not having this frequency it is obtained and used later in the methodology. Some test were carried out firstly with virtual objects, here the parameters that lead to the minimum error in the 3D reconstruction are obtained. The results with the obtained parameters were implemented in the 3D reconstruction of real objects.

Keywords: 3D reconstruction, Morlet wavelet, parameter selection, Fourier profilometry.

1. Introducción

El avance tecnológico se ha extendido, abarcando áreas significativas que anteriormente no se habían considerado, debido a la falta del desarrollo de herramientas necesarias para el estudio de estos campos. Este avance ha generado demandas de nuevos métodos para mejorar la calidad de los procesos, los cuales, son cada vez más sofisticados y precisos, haciendo uso de herramientas que implementan una mejora continua. Dentro de este desarrollo, el uso de las técnicas de procesamiento digital de imágenes se ha incrementado en las últimas décadas, mostrando progresos significativos en el área de reconstrucción de objetos 3D. Existen dos clasificaciones dentro de la medición de objetos las cuales son técnicas de contacto o no contacto. Las técnicas de no contacto son las más utilizadas por sus características no invasivas y su velocidad de procesamiento. De esta técnica deriva el método de luz estructurada, del cual parte el de proyección de franjas, basado en la proyección de franjas emitido por un proyector sobre un objeto, donde posteriormente, el efecto de la distorsión de dichas franjas es capturado por una cámara. A. Asundi, Z. Wensen, 1998 [1].

Entre la variedad de métodos que consideran la proyección de patrones espaciales sinusoidales, existen técnicas que ofrecen mejores resultados en la extracción de distribuciones de fase de dichos patrones bi-dimensionales deformados: uno de ellos es llamado cambio de fase (phase-shifting) y el otro Perfilometría de Fourier (Fourier Transform Profilometry o FTP), éste último propuesto originalmente por Takeda [2] y posteriormente modificado y mejorado por Wu Lu [3] y Pedraza [4]. La diferencia entre estos métodos es que el primero requiere de al menos 3 imágenes para obtener la información de fase del objeto. Por otro lado, la FTP extrae la información del mapa de profundidad (el cual contiene la información de la altura del objeto) utilizando una sola imagen, lo que lo ha convertido en un método popular para el análisis de patrones de franjas. Sin embargo, independientemente de las modificaciones implementadas en este método, una de las desventajas que muestra es su localización espacial pobre. El uso del primer armónico es afectado por el ruido local y la discontinuidad de los puntos, los cuales al final arrojaran un error en la fase, este error se incrementa cuando este primer armónico es superpuesto por otros altos armónicos, imposibilitando la extracción del primero. Zhang [5], Con estas altas frecuencias existe un traslape en la etapa del desdoblamiento de fase, generando que la altura estimada del objeto se extienda de forma global, obteniendo una reconstrucción 3D de objetos no adecuada. [4, 6-8].

Debido a que la FTP trabaja solo con señales estacionarias existe una pérdida de información en la localización en el tiempo de las frecuencias de los componentes dados.

Otra técnica que obtiene el mapa de profundidad partiendo de una sola imagen es la Perfilometría Wavelet (WTP), la cual trabaja con Transformada Wavelet, sobresaliendo por su capacidad de procesamiento de señales no estacionarias, además de la localización de dominio de tiempo frecuencia. Este concepto fue introducido por Zhong et al en 2004 [9]. La extracción de fase de la WTP maneja dos enfoques, uno es el método de fase gradiente y el segundo es la estimación de fase. Este trabajo utiliza éste último debido a que este método supera al de gradiente de fase para un patrón con o sin ruido. Gdeisat [10].

El uso de la WTP está siendo muy investigado, con un gran número de trabajos que le preceden, Como es el caso de Xiang [11] y Qiang Zhang [12], donde ambos proponen el uso de las Wavelets Haar y Morlet para obtener la información de altura de los objetos. Abdulbasit Z, et al en 2008 [13] propone la transformada continua Morlet modificada en 1D para el análisis de patrón de franjas. Sergio F, et al, 2011 [14], utiliza el análisis de patrón de franjas haciendo una comparación de las técnicas principales de tiempo-frecuencia (Transformada de Fourier, ventaneo de Fourier y Transformada Wavelet: Shannon, B-Spline, Paul y Morlet), trabajando con objetos reales y virtuales mostrando resultados cualitativos y cuantitativos. Por ello, este trabajo propone el uso de la Perfilometría Wavelet haciendo una comparación con la Perfilometría de Fourier. La transformada Morlet es usada con parámetros específicos con la finalidad de obtener una mejor reconstrucción, tomando como ventaja que la transformada wavelet trabaja con señales no estacionarias, así como su análisis multiresolución; de esta forma el proceso de reconstrucción 3D es llevado a cabo con ambos métodos en 1D, para verificar cuál de los dos presenta un mejor desempeño en la reconstrucción de objetos virtuales, y así, reproducir ambos resultados en objetos reales.

2. Perfilometría de Fourier y Wavelet

El método de Perfilometría de Fourier (FTP) es uno de los más importantes dentro del conjunto de técnicas que existen para la adquisición de imágenes 3D. La idea básica del método consiste en proyectar un patrón de franjas sobre el objeto del que se quiere obtener su imagen, al grabar con una cámara la imagen de la escena obtenemos un patrón distorsionado que lleva consigo toda la información 3D del objeto particular.

Aplicando la transformada de Fourier al objeto con el patrón de franjas distorsionado, se obtiene un espectro de amplitud y uno de fase. Posteriormente se lleva a cabo un filtrado en la frecuencia espacial f_0 donde esta frecuencia corresponde al número de franjas proyectado sobre el objeto. Aplicando este filtro a ambos espectros y llevando a cabo una traslación de la imagen filtrada hacia el centro de la imagen, se obtiene la transformada inversa de Fourier. Entonces se aplica la transformada inversa de Fourier y se lleva a cabo el proceso de desdoblamiento de fase a la parte imaginaria para poder obtener así la forma o altura del objeto. Este método ha sido investigado ampliamente desde mediados de 1980, aunque tiene múltiples desventajas, ya que en ocasiones no es posible obtener buenos resultados debido a que no se obtiene de forma

apropiada la frecuencia f_0 o bien el filtro en esa frecuencia no es muy preciso y por el traslape de frecuencias con armónicos superiores o con la componente de DC, lo cual acarrea valores no deseados en la etapa de desdoblamiento de fase. Por último, en cuando a los algoritmos de desdoblamiento de fase, también son una limitante ya que se requiere que el mapa de fase tenga el menor número de transiciones de alta frecuencias para que no afecte al momento de obtener la forma o altura del objeto.

La wavelet es una herramienta utilizada para el análisis de señales no estacionarias y variantes en el tiempo. La forma de onda de las wavelets es lo que le permite no limitarse a las funciones seno y coseno. Por otro lado tienen la capacidad de multi-resolución, trasladando la señal sobre el eje x y alargando o contrayendo dicha señal en el eje y, llamando a este evento escalamiento de la señal, Cortes et al, 2007 [12]. De forma general, la Transformada Wavelet WT de una señal f es un conjunto de coeficientes $W(a, b)$, cuyos índices se asocian con la escala y la posición de la señal. Considerando el caso unidimensional, los coeficientes se obtienen como:

$$W_{(a,b)} = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) dt, \quad (1)$$

donde, $a \in \mathbb{R}^+ \setminus \{0\}$, y $b \in \mathbb{R}$.

La transformada continua wavelet CWT se considera como la suma sobre todo el intervalo de tiempo de la señal, multiplicada por las versiones de escala y traslación de la función ψ . Arellano [16] La señal original es obtenida cuando se multiplica cada uno de los coeficientes por la wavelet escalada y trasladada. Sin embargo, es casi imposible llevar a cabo el cálculo de todas las combinaciones posibles entre los factores de escala, coeficientes y traslaciones.

$$C(\text{escala}, \text{posición}) = \int_{-\infty}^{\infty} f(t) \psi(\text{escala}, \text{posición}, t) dt \quad (2)$$

Al trabajar con regiones de escala-tiempo el análisis wavelet permite el uso de intervalos largos en el tiempo brindando información precisa de baja frecuencia e información de alta frecuencia utilizando regiones pequeñas. Esta capacidad de convergencia y adaptabilidad se debe a su tendencia asimétrica e irregular, es decir, que la forma de onda de la wavelet pende de la variación de los coeficientes, determinando cómo la señal original será reconstruida.

El cambio de escala está representado por:

$$\frac{1}{\sqrt{a}} \psi\left(\frac{x}{a}\right), \quad (3)$$

donde $a > 0$.

Por otro lado, la traslación se define por:

$$\psi(x - b), \quad (4)$$

entonces la traslación y el cambio de escala en una dimensión está definida por:

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right), \quad (5)$$

donde, $a > 0$, $b \in \mathbb{R}$.

Por lo tanto, el análisis continuo está dado por:

$$C_{(a,b)} = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) dt, \quad (6)$$

donde, $a \in \mathbb{R}^+ \setminus \{0\}$, y $b \in \mathbb{R}$.

De aquí tenemos que:

$$W_{(a,b)} = \int f(t) \psi_{a,b}^*(t) dt. \quad (7)$$

Debido a estas variaciones de coeficientes surge la familia de wavelets, entre las que se encuentran, B-Spline, Paul, Fan, Morlet, Mexican Hat, Haar, entre otras. Además de la familia de wavelets, éstas también se clasifican en complejas y no complejas, esta investigación usará la wavelet compleja, debido a que ésta trabaja con la parte real e imaginaria, mismas que se utilizan para obtener la amplitud y la información de fase del objeto a reconstruir.

3. Metodología propuesta

La metodología propuesta para el proceso de reconstrucción 3D, se muestra en la Fig. 1. Haciendo uso de los métodos de Perfilometría de Fourier y Perfilometría Wavelet para realizar la comparativa de dicha reconstrucción entre ambos métodos. Aunque la literatura muestra que las reconstrucciones se validan comparando el objeto reconstruido con las mediciones del objeto real, lo que hace que este proceso pueda resultar costoso en su momento. Debido a esto se propone hacer esta comparativa con ambos métodos al mismo objeto. La metodología, la cual consta de 7 pasos, también incluye la identificación de si el objeto capturado es real o virtual, permitiendo con esto mejorar el proceso.

En el primer paso del proceso se adquiere la imagen a reconstruir. En el segundo paso se determina si la imagen adquirida es virtual o real; para ello, se realiza un conteo de la cantidad de cambios de color que haya en el primer renglón de cada imagen, si hay más de dos cambios de color en ella, entonces, se concluye que la imagen es real y que tiene franjas proyectadas. En caso de que la imagen adquirida sea virtual, se le suma el patrón de distorsión de franjas seleccionado. Para ambos casos, la imagen adquirida es redimensionada a una de 512 x 512 píxeles, con la finalidad de tener consistencia durante todo el proceso. El tercer paso para el caso de la imagen real es calcular el número de franjas proyectadas sobre el objeto, debido a que existe una relación entre el número de franjas proyectadas y la frecuencia f_0 .

Para el caso del objeto virtual es determinar la frecuencia f_0 . La f_0 brinda la información correspondiente a la altura del objeto en cada píxel y contribuye a su

reconstrucción 3D. Una vez identificado el tipo de objeto, el número de franjas y su frecuencia f_0 . El cuarto paso es aplicar al objeto ambos métodos: Perfilometría de Fourier y Perfilometría Wavelet, para la wavelet se seleccionaron los mejores parámetros. En el quinto paso se obtiene la fase envuelta, donde se encuentra, la información de la altura del objeto. Dicha información está envuelta entre los valores de $-\pi$ a π , por lo que es necesario implementar un análisis de desdoblamiento de fase. La fase desenvuelta es obtenida en el sexto paso, es en este punto donde el análisis es aplicado. Se considera el algoritmo de desdoblamiento de fase propuesto por Itoh, 1982 [8], el cual permite obtener la profundidad del objeto.

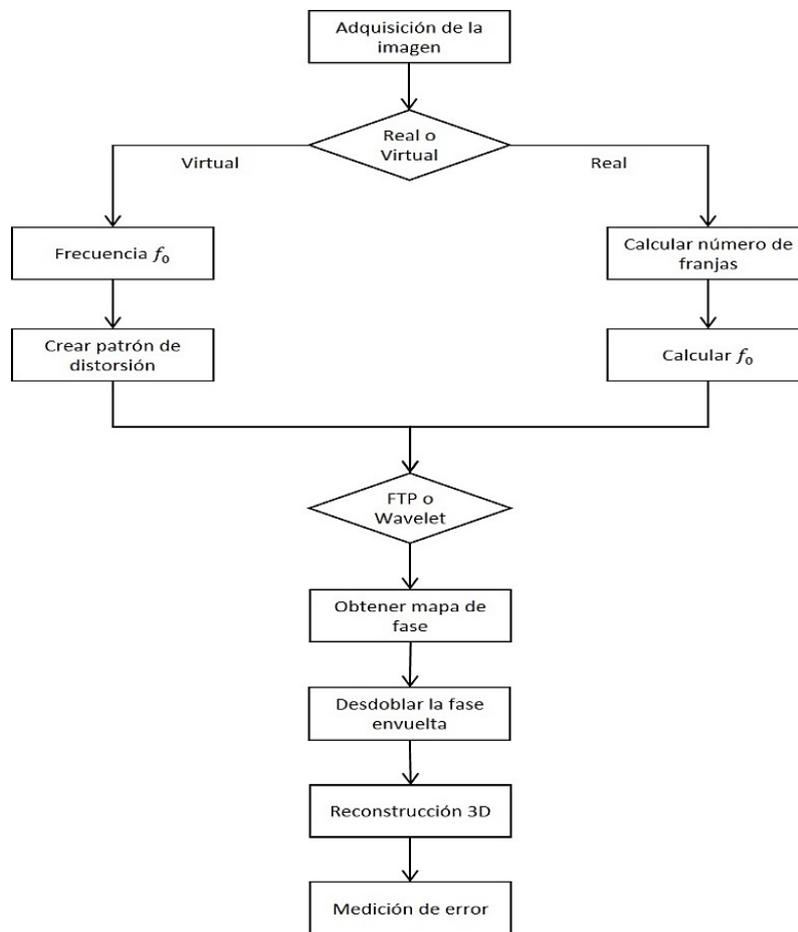


Fig. 1. Metodología propuesta para la reconstrucción 3D y la comparativa entre la FTP y la WTP.

Para hacer la comparación de los métodos, la medición del error es implementada en el último paso, esta medición se llevará a cabo solo con el objeto virtual, es decir, al

objeto virtual original se le restará el objeto obtenido de la reconstrucción para ambos métodos, obteniendo así el error haciendo así, la validación de los métodos aplicados.

La Fig. 1 describe el método propuesto para enmarcar los resultados del presente trabajo.

En la metodología hay un paso (4), que aplica Perfilometría Wavelet a una determinada imagen para su procesamiento, para llevar a cabo esta etapa es necesario formalizar la Transformada Wavelet, que para el caso del presente trabajo es la Transformada Morlet Compleja, debido a que ésta tiene una mejor respuesta comparativamente con respecto a otras transformadas Wavelets Complejas, definidas para el mismo fin López, 2012 [17]. Esta función está definida mediante

$$\Psi(x) = \pi^{-1/4} \exp(i2 \pi w_0 x) \exp(-x^2/2), \quad (8)$$

donde:

$$morl(x) = e^{-x/2} \cdot \cos(5x), \quad (9)$$

donde, $i = \sqrt{-1}$, y $w_0 = \text{Frecuencia de la Wavelet Morlet}$.

La importancia de esta transformada radica en que es una función continua y compleja, y esto permite hacer la descomposición de la señal (la imagen) en su parte real e imaginaria, de aquí se tiene un conocimiento mayor de la imagen. Esta función será la función madre que alimentará a la Transformada Wavelet, y llevar a cabo el proceso llamado Perfilometría Wavelet.

4. Resultados experimentales

Para la metodología propuesta fueron seleccionados dos objetos virtuales, los cuales se muestran en la Fig. 2.

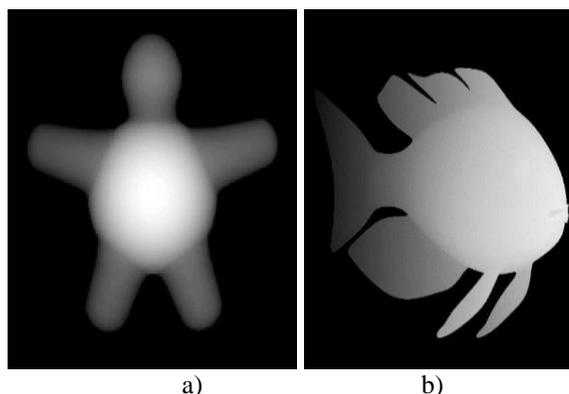


Fig. 2. Objetos virtuales usados en este trabajo: a) Tortuga y b) Pez.

Con la selección de los parámetros apropiados es factible obtener una buena reconstrucción 3D de objetos mediante el uso de algoritmos simples de desdoblamiento de fase, como se muestra en la Figura 3.

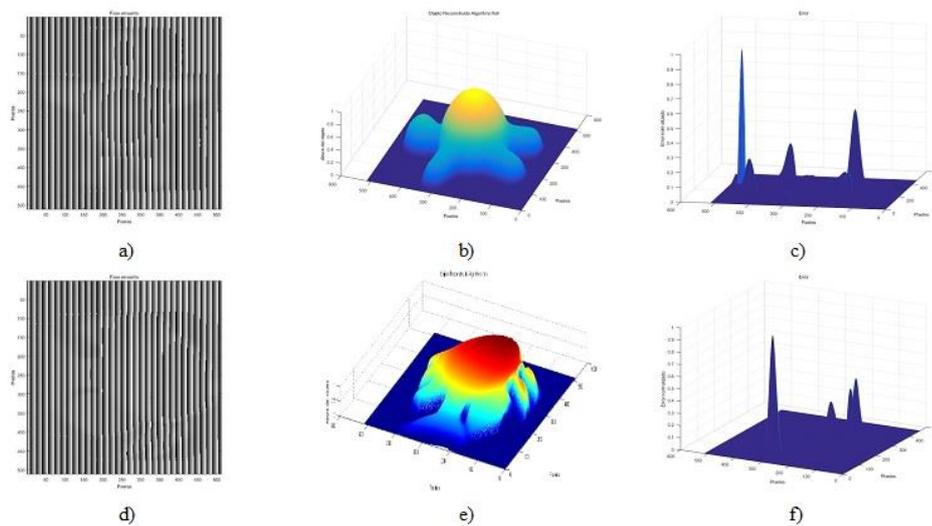


Fig. 3. a) y d) Muestran la fase envuelta de los objetos, b) y e) muestran el objeto reconstruido y c) y f) muestran el error, el cual se obtiene restando el objeto original al objeto reconstruido.

Después de realizadas las pruebas con los objetos virtuales, el proceso es implementado a un objeto real, mostrado en la Fig. 4.

La Figura 5(a) muestra el mapa de fase obtenido, es el mapa de fase o fase envuelta la que contiene la información de la altura del objeto, para posteriormente obtener la fase desenvuelta, que es la reconstrucción 3D del objeto, el cual se muestra en 5 (b). Con objetos reales, el error no se maneja, dado que no se tiene la facilidad de hacer la diferencia entre el objeto original y el reconstruido.

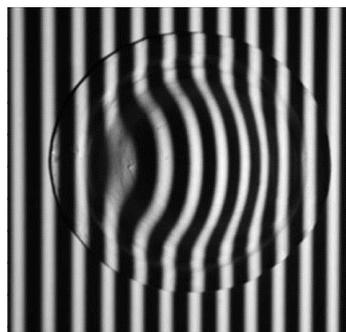


Fig. 4. Objeto Real con 12 franjas proyectadas.

La Fig. 5 muestra la reconstrucción 3D del objeto real utilizando la Wavelet Morlet.

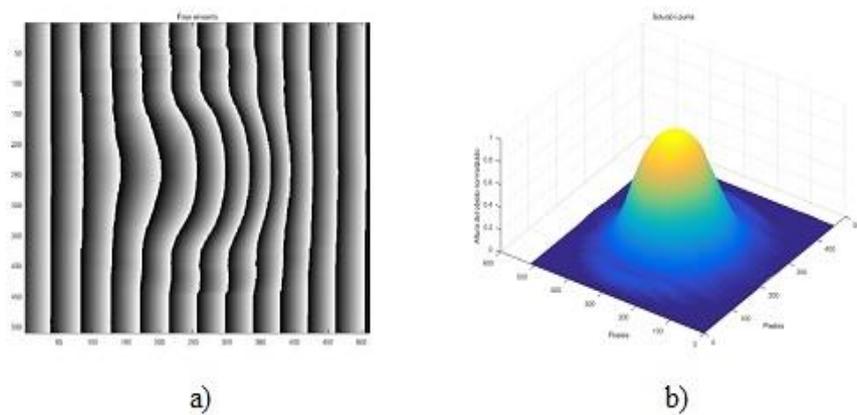


Fig. 5. a) Fase envuelta y b) Reconstrucción 3D del objeto con wavelet Morlet.

En la Fig. 6 y 7 se muestran la reconstrucción de los objetos virtual y real respectivamente, pero utilizando la perfilometría de Fourier. En el caso de la Fig. 6, en la reconstrucción 3D de los objetos se presentaron ciertas ondulaciones, tanto en la base, así como en algunas partes de los objetos.

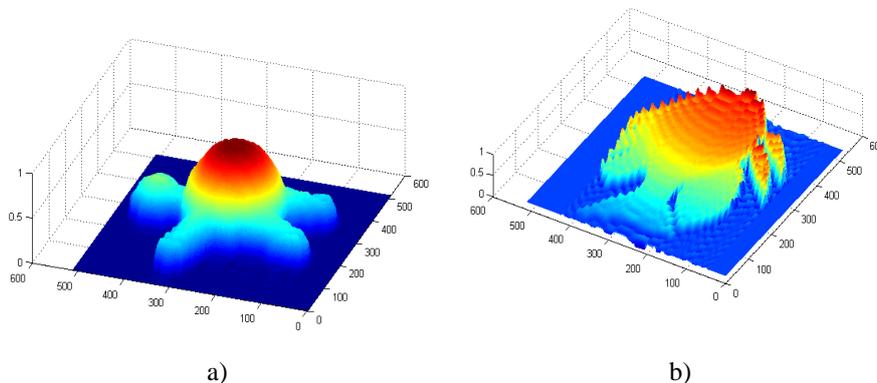


Fig. 6. Reconstrucción 3D de objetos virtuales utilizando Perfilometría de Fourier: a) Tortuga, b) Pez.

Nótese como el mapa de fase, mostrado en la Fig. 7 a) tiene algunas zonas en donde se hace más gruesa una de las líneas proyectadas sobre el objeto, afectando este mapa de fase y posteriormente cuando se lleva a cabo la reconstrucción 3D del objeto completo, se obtiene una zona donde el objeto reconstruido tiene una pequeña deformación, la cual no existe en el objeto original. Esta deformación no está presente en la reconstrucción utilizando la Wavelet Morlet.

Adicionalmente, en la base del objeto se llegan a notar algunas ondulaciones en el objeto reconstruido utilizando la Perfilometría de Fourier como se observa en la Fig. 6 b), mientras que en la Fig. 5 b) la base del objeto reconstruido no tiene ondulaciones.

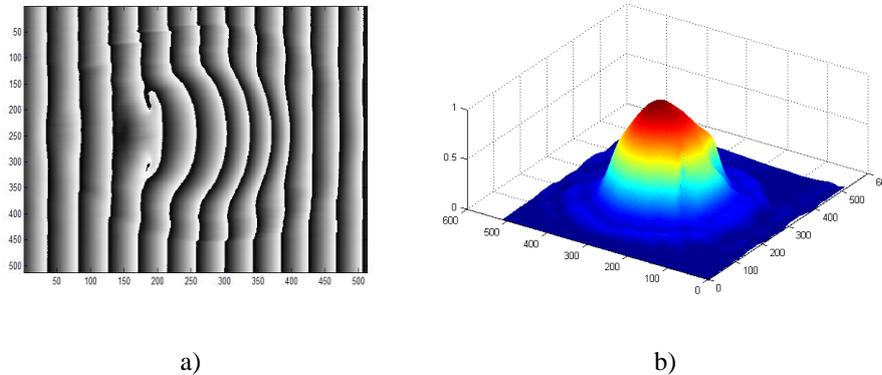


Fig. 7. a) Fase envuelta y b) Reconstrucción 3D del objeto con Perfilometría de Fourier.

5. Conclusiones y trabajo futuro

En el presente artículo se propuso el uso de los métodos de Perfilometría de Fourier y Perfilometría Wavelet para la reconstrucción 3D de objetos. Se presenta una comparativa entre el desempeño de ambos métodos, implementando la selección de los mejores parámetros para la wavelet Morlet en la Perfilometría Wavelet, los cuales siendo seleccionados de forma adecuada, reducen el ruido en el mapa de fase de los objetos a reconstruir. Dentro de estos parámetros se considera el análisis de la estimación de fase mediante la frecuencia espacial f_0 determinada por el número de franjas proyectada sobre el objeto. Este análisis es utilizado para llevar a cabo el filtrado en esa frecuencia dentro de la Wavelet Morlet.

La metodología propuesta hace una distinción entre objetos reales y virtuales, así como la determinación de la frecuencia espacial f_0 si el caso de entrada es un objeto virtual. Los mejores resultados obtenidos de las pruebas llevadas a cabo en los objetos virtuales, fueron implementados en objetos reales, obteniendo los mejores resultados con la Perfilometría Wavelet. Con esto se puede apreciar que el uso de la Perfilometría Wavelet muestra un mejor desempeño que la Perfilometría de Fourier en este caso.

Como trabajo futuro se tiene la implementación y/o uso de algoritmos más robustos de desdoblamiento de fase, con el propósito de reducir el error en la reconstrucción 3D de objetos, independientemente del método a utilizar (Fourier o Wavelet).

Referencias

1. Asundi, A., Wensen, Z.: Fast phase-unwrapping algorithm based on a gray-scale mask and flood fill. *Applied optics*, Vol. 37, No. 23, pp. 5416–5420 (1998)
2. Mitsuo, T., Hideki, I., Seiji, K.: Fourier-Transform method of fringe pattern analysis for computed-based topography and interferometry. *J.Opt. Soc.Am.*, Vol. 72, No. 1, pp. 156–160 (1982)
3. Wu, L.: Research and development of fringe projection-based methods in 3D shape reconstruction. *Journal of Zhejiang University SCIENCE A*, pp. 1026–1036 (2006)

4. Pedraza, J. C., Rodriguez, W., Barriga, L.: Image Processing for 3D Reconstruction using a Modified Fourier Transform Profilometry Method. MICAI 2007, LNAI 4827, Springer-Verlag, Berlin Heidelberg, pp. 705–712 (2007)
5. Zhang, Z., Jing, Z., Wang, Z., Kuang, D.: Comparison of Fourier transform, windowed Fourier transform and wavelet transform methods for phase calculation at discontinuities in fringe profilometry. *Optical and Lasers Engineering*, Vol. 50, pp. 1152–1160 (2012)
6. Berryman, F., Pynsent, P., Cubillo, J.: A theoretical Comparison of three fringe analysis methods for determining the three-dimensional shape of an object in the presence of noise. *Optics and Lasers in Engineering*, Vol. 39, pp. 35–50 (2003)
7. Pramod, R.: *Digital Speckle Pattern Interferometry and related Techniques*. Edit. Wiley (2001)
8. Itoh, K.: Analysis of the phase unwrapping algorithm. *Applied Optics*, Vol. 21, No. 14, pp. 2470–2486 (1982)
9. Zhong, J., Weng, J.: Spatial carrier-fringe pattern analysis by means of wavelet transform: wavelet transform profilometry. *Applied optics*. Vol. 43, No. 26, pp. 4993–4998 (2004)
10. Gdeisat, M. A., Abid, A., Burton, D. R., Lalor, M. J., Lilley, F., Moore, C., Qudeisat, M.: Spatial and temporal carrier fringe pattern demodulation using the one-dimensional continuous wavelet transform: recent progress, challenges, and suggested developments. *Optics and Lasers in Engineering*, Vol. 47, No. 12, pp. 1348–1361 (2009)
11. Xiang, Z., Hong, Z.: Three-Dimensional Shape Measurement Using Maar Wavelet Transform. In: *Proceedings of ICIRA 2008, Part I LNAI 5314*, pp. 753–760 (2008)
12. Qiang, Z., Wenjing, C., Yan, T.: Method of choosing the adaptive level of discrete wavelet decomposition to eliminate zero component. *Optics Communication*, pp. 778–785 (2008)
13. Abid, A. Z., Gdeisat, M. A., Burton, D. R., Lalor, M. J., Abdul-Rahman, H. S., Lilley, F.: Fringe pattern analysis using a one-dimensional modified Morlet continuous wavelet transform In *Photonics Europe*. International Society for Optics and Photonics, pp. 70000Q–70000Q (2008)
14. Fernandez, S., Gdeisat, M. A., Salvi, J., Burton, D. R.: Automatic window size selection in windowed Fourier transform for 3D reconstruction using adapted mother wavelets. *Optics Communications*, Vol. 284, No. 12, pp. 2797–2807 (2011)
15. Cortés, J. A., Garzón, H. B. C., Chaves, J. A.: *Del Análisis de Fourier a las Wavelets-Transformada Continua Wavelet (CWT)*. *Scientia et Technica*, Vol. 1, No. 37 (2007)
16. Arellano, P.: *Missing information in remote sensing: wavelet approach to detect and remove clouds and their shadows*. International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands (2003)
17. López, T.: *Estudio comparativo entre tipos de transformada wavelet para su uso en reconstrucción tridimensional*. Doctoral dissertation (2012)

Simulador ROV multiplayer para escenarios petroleros submarinos

Rodrigo Beltran-Labra¹, Miguel Pérez-Ramírez¹, Carlos Alberto Ochoa-Zezzatti²,
Norma J. Ontiveros-Hernández³

¹ Instituto Nacional de Electricidad y Energías Limpias, Cuernavaca, Morelos, México

² Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua, México

³ Instituto Tecnológico de Zacatepec, Zacatepec, Morelos, México

blro130524@upemor.edu.mx, mperez@iee.org.mx, alberto.ochoa@uacj.mx,
njoh_314@yahoo.com.mx

Resumen. Se presenta la versión multijugador de SimROV, un simulador basado en Realidad Virtual e Inteligencia Artificial. Incluye un escenario submarino donde un operador de un ROV puede realizar algunas maniobras a las instalaciones de varios pozos de petróleo. El sistema incluye consolas para un instructor y para uno o más estudiantes, todos comparten un escenario en común, cada participante tiene control sobre su propio ROV aunque pueden observar e interactuar con los otros. El instructor puede introducir corrientes marinas y fugas de combustible, los estudiantes necesitan solucionar el problema, pero sus acciones son detectadas para que puedan ser guiados por el sistema mediante el uso de algoritmos de inteligencia artificial.

Palabras clave: realidad virtual, inteligencia artificial, ROV, multiplayer, simulador.

Simulator ROV Multiplayer for Oil Submarine Scenarios

Abstract. A multiplayer version of SimROV is presented based on Virtual Reality and Artificial Intelligence. It includes a submarine scenario where an operator of an ROV can analyze some maneuvers to the installations of several oil wells. The system includes consoles for an instructor and for one or more students, and all of them share a common scenario, each participant has control over its own ROV although can observe and interact with others. The instructor can introduce marine currents and fuel leaks, the students need to solve the problem, but their actions are detected to be guided by the system through the application of artificial intelligence algorithms.

Keywords: virtual reality, artificial intelligence, ROV, multiplayer, simulator.

1. Introducción

El objetivo de este trabajo es presentar la versión multijugador de SimROV, un simulador de un ROV (Vehículo de Operación Remota por sus siglas en inglés). Además de permitir compartir el mismo escenario entre un instructor y diferentes estudiantes a través de una red, el sistema incluye algoritmos de inteligencia artificial. Por el momento estos algoritmos perciben las acciones de los estudiantes dentro del entorno virtual y proporcionan ayuda para que los estudiantes sean guiados para lograr satisfactoriamente la meta asignada por el instructor. Entre otras cosas se verifica la trayectoria de los movimientos de cada ROV, de esta manera cuando se trabaja en una maniobra específica como el cierre de una válvula para detener una fuga de combustible, se compara la trayectoria de los alumnos con la trayectoria que se espera que siga el estudiante, en caso de que el camino real esté fuera de un rango permitido, los sistemas enviarán al alumno algún mensaje para ayudarlo a volver al camino correcto.

Además de lo anterior, el ROV para trabajos submarinos está equipado con dos brazos y 2 cámaras de video. No está demás agregar que aun sin pensar en la realización de acciones para responder a una maniobra de contingencia, SimROV también puede ser utilizado para que un estudiante aprenda a utilizar y controlar un ROV. Esto incluye no solo el guiar o dirigir el ROV hacia un punto determinado sino saber controlarlo aun en la presencia de corrientes marinas, las cuales dificultan su maniobrabilidad. El control de los brazos también requiere de cierta pericia que puede ser obtenida con la ayuda de SimROV, la dificultad que se presenta aquí, es que hay que aprender a usar los controles para realizar una acción, lo cual implícitamente demanda la realización de actividades en paralelo: mover los diferentes controles de cada brazo, además de guiar al ROV dentro del ambiente virtual.

Detectar la posición de cada ROV y los movimientos de los brazos permitirá identificar posibles áreas de mejora en el proceso de aprendizaje del estudiante. Esto también alimenta al modelo de los estudiantes para que SimROV siempre conozca el progreso de los estudiantes y su estado actual.

El resto del artículo está organizado de la siguiente manera. La sección 2 incluye algunos conceptos de simuladores. La sección 3 presenta la arquitectura de SimROV. La sección 4 discute los principales logros del desarrollo de esta versión de SimROV. La sección 5 incluye algunas conclusiones que son seguidas por una lista de referencias.

2. Simuladores

Un simulador para entrenamiento es un sistema, que permite adquirir algún tipo específico de habilidad [1]. La literatura sobre el tema de los simuladores distingue dos tipos de simuladores [2]:

- a) Los simuladores orientados al diseño de instalaciones (SODI),
- b) Los simuladores de entrenamiento (SE).

A su vez dentro de los SE se distinguen dos tipos:

1. Los simuladores de alcance total (SAT), que reproducen completamente el entorno de operación. Estos simuladores están asociados a una réplica de la sala de control (SC), tanto en apariencia como en funciones.

2. Los simuladores que reproducen la conducta del proceso (SRCP), pero no reproducen la SC, pero que sí permiten actuar sobre la simulación en tiempo de ejecución.

En lo que respecta a SimROV, este incluye un ambiente virtual y todos los controles para operar un ROV, es decir, incluye una representación del entorno de operación (pozos petroleros en el fondo del mar), y desde su creación fue concebido como un simulador de entrenamiento (SE). Por otra parte, también es un SRCP, lo único que le falta para poder ser considerado un simulador de alcance total (SAT), es la cabina de control. Los controles de un ROV real están basados en joysticks y otros controles, mientras que SimROV por el momento está basado en el uso de teclado y mouse, y un control de videojuegos.

Sin embargo, a falta de la cabina de control, SimROV contiene el modelo del estudiante a fin de personalizar la instrucción, además de estar equipado con algoritmos de inteligencia artificial para monitorear y guiar al estudiante en su actuar durante su respuesta a situaciones de contingencia, que son introducidas por el instructor.

3. Arquitectura de SimROV

La arquitectura del sistema SimROV obedece en gran medida a una serie de requisitos incluidos en la IMCA-International Marine Contractors Association, la cual define un simulador como la creación de ciertas condiciones a través de un modelo, para simular condiciones dentro de una esfera de operación apropiada [3] y clasifica los simuladores en dos clases:

Clase A) Utilizan una máquina de físicos realista y crea una representación visual adecuada para un propósito.

Clase B) Utilizan una máquina de físicos casi realista y crea una representación visual apropiada.

SimROV cumple con la mayoría de los requisitos de un simulador de clase A [4]. La Figura 1 muestra la representación gráfica de la arquitectura de SimROV multiplayer, en dicha figura se puede observar que tanto estudiante como instructor tienen su propia interfaz y son diferentes dado que ambos usuarios tienen funciones diferentes.

A continuación, se muestran los componentes de la arquitectura:

1) *Interfaces*: El sistema permite el acceso a un ambiente virtual en la red. Los usuarios pueden acceder al ambiente virtual de dos formas: como estudiante (operador) y como instructor. La Figura 2 muestra la interfaz principal.

Dentro de la interfaz del instructor se tiene control sobre el ambiente virtual, permite ingresar algunas maniobras y observar las acciones del estudiante a través de sus cámaras en primera persona, además el instructor tendrá acceso a la cámara global del

ambiente para observar las acciones que siguen todos los alumnos en tiempo real. La cámara global sólo está disponible para que el instructor. Posteriormente el avance de las maniobras y de la capacitación en general será guardado dentro de una base de datos, a la cual podrá acceder el instructor.

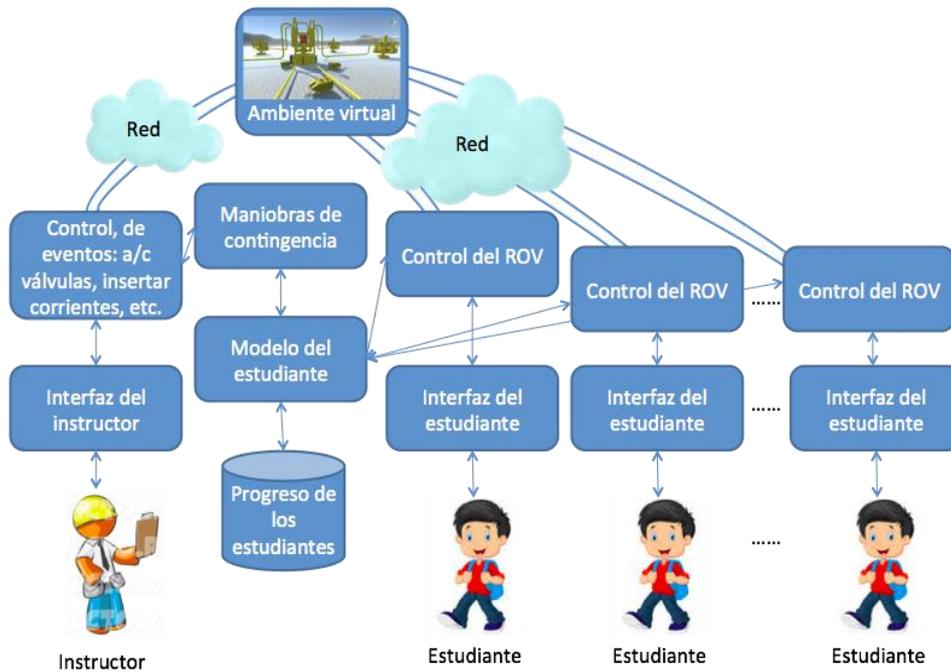


Fig. 1. Arquitectura de SimROV multiplayer.

La interfaz del estudiante le permite controlar el ROV, esto es; controlar su movimiento para trasladarse de un lugar a otro, controlar el brazo izquierdo de apoyo así como su brazo derecho para realizar las maniobras encomendadas por el instructor.

Al implementar el algoritmo de Inteligencia Artificial para calcular la ruta más corta y la base de datos con los recorridos de todas las maniobras, el sistema podrá ir guiando al estudiante durante todo el tiempo de la maniobra para que pueda concluir exitosamente la maniobra asignada.

2) *Ambiente virtual común*: Ambas interfaces pueden ver el mismo ambiente virtual ya que se implementa el método “NetworkManager” de Unity [5] para multijugador y el uso de la red ya sea intranet o internet. Las animaciones de las corrientes marinas y de los movimientos del ROV no están previamente definidas, los brazos y la translación del ROV serán originadas a voluntad por el estudiante a través de los controles definidos para este propósito.

3) *Control del ROV*: El estudiante tiene el control total del ROV, de tal forma que puede moverlo en la dirección que desee a través de las flechas del teclado o un control de videojuegos. Además permite el control sobre los brazos del ROV, de tal manera

que con el brazo derecho pueda sujetarse a algún gancho para evitar el arrastre de las corrientes y de esta manera tomar y manipular objetos con el brazo izquierdo. Con esta funcionalidad el estudiante puede realizar las maniobras asignadas por el instructor.

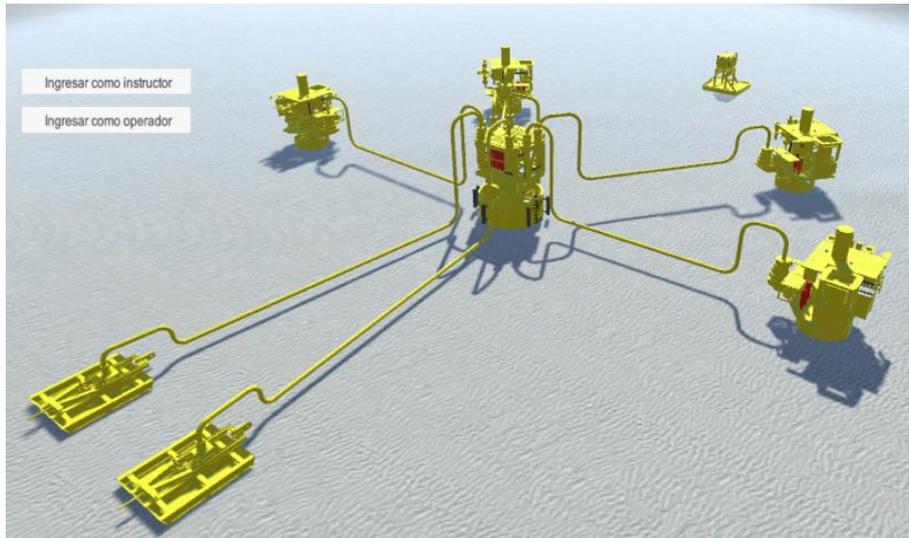


Fig. 2. Interfaz de SimROV multiplayer: Ambiente Virtual.

4) *Control de eventos*: El instructor podrá introducir en el ambiente virtual situaciones de contingencia tales como abrir y cerrar válvulas, inyectar corrientes marinas que puedan afectar al ROV, introducir fugas, etc.

5) *Maniobras de contingencia*: Este módulo contiene información de las maniobras de contingencia que un estudiante puede aprender y a su vez puede registrar su progreso en el aprendizaje en una Base de Datos. Una maniobra de contingencia podría darse por ejemplo por una fuga, que demandaría el cierre de válvulas o incluso el cambio de una sección de tubería, además en este mismo módulo se encuentran alojadas en la base de datos las posiciones de los objetos más importantes del escenario y donde los alumnos pueden llegar a realizar algunas maniobras y es con esto que el sistema puede tomar decisiones para guiar al alumno.

6) *Modelo del estudiante*: A la fecha con la maniobra de contingencia de la introducción de fugas, lo que se puede monitorear son dos cosas principalmente: a) que el usuario sabe controlar al ROV para desplazarse a la válvula que desea cerrar y b) que sabe cómo operar los dos brazos del ROV, de tal modo que puede sostenerse de algún gancho para evitar que la corriente lo arrastre o lo vuelque y de que puede tomar la válvula correspondiente, para cerrarla sin causar daños a otros equipos o al mismo ROV. Así el modelo del estudiante registra en una base de datos la actuación de los estudiantes en estas dos actividades globales. Sin embargo, debido al accionar concurrente de los controles que se requiere para operar los brazos y la navegación del

ROV, también es posible monitorear dichas actividades a fin de obtener una idea precisa de si está aprendiendo a controlar el ROV o no.

4. Desarrollo

A. Ambiente Virtual.

Los objetos del escenario fueron creados como modelos 3D y se diseñaron para ser importados a Unity, para darles un aspecto físico se le asignan propiedades tales como RigidBody y Mesh Collider. RigidBody permite a los objetos actuar bajo el control de la física, puede recibir fuerza y torque para que el objeto se mueva de forma realista [6], además es un componente obligatorio para que los objetos sean afectados por la gravedad y de esta manera interactuar con otros objetos a través de Physx que es un motor de física de NVIDIA [7; 8]. Mesh Collider construye una presentación de colisión para el objeto [9], esto es que al no ocupar figuras primitivas se crea una malla con la forma física del objeto para colisionar con los demás objetos en cualquier parte de su estructura sin importar que no se trate de alguna figura primitiva.

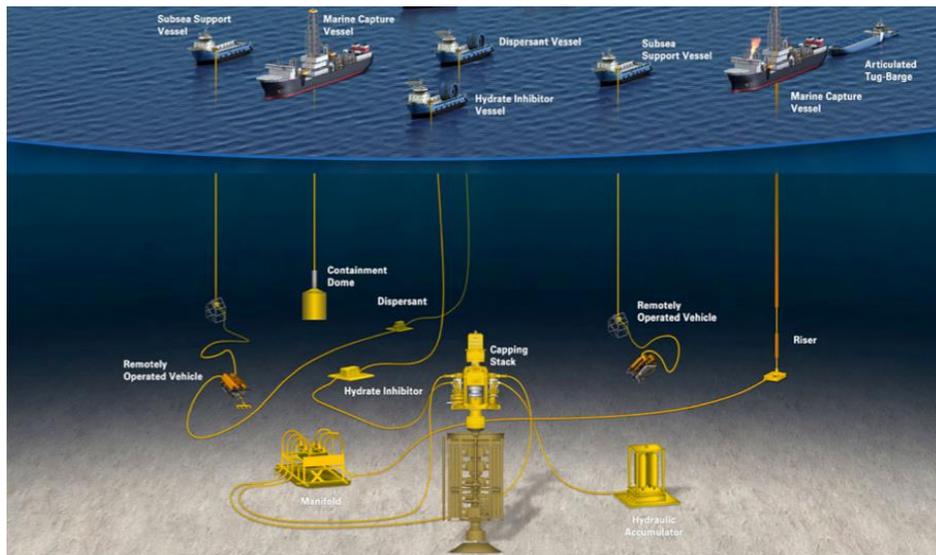


Fig. 3. Escenario submarino: Marine Well (Tomado de [10]).

El escenario es amplio con el objetivo de que todos los usuarios puedan interactuar entre sí, dicho escenario está basado en un ambiente submarino según la compañía Marine Well mostrado en la Figura 3.

B. Multiplayer.

Tanto el estudiante como el instructor comparten el mismo ambiente virtual pero con distintos privilegios, por tal motivo se crean dos diferentes consolas; la consola del

estudiante le permite controlar el ROV y observar las maniobras asignadas, la consola del instructor le permite asignar tareas a cada estudiante así como observar qué es lo que está realizando el estudiante en tiempo real a través de su cámara en primera persona.

En el sistema, el instructor está configurado como servidor y los estudiantes como clientes tal y como lo muestra la Figura 4. Dado que el instructor debe también estar presente en la simulación, se configura al instructor como su propio cliente ya que funge como servidor o Host. El anfitrión es un servidor y un cliente en el mismo proceso. El anfitrión utiliza un tipo especial de cliente llamado LocalClient (cliente local), mientras otros clientes son RemoteClientes (clientes remotos) [5].

Todos los estudiantes ejecutan la misma consola, para poder distinguirse de entre todos lo demás se agrega el componente "NetworkIdentity" el cual controla la identidad de un objeto en red, de esta forma se logra que cada estudiante controle su propio ROV.

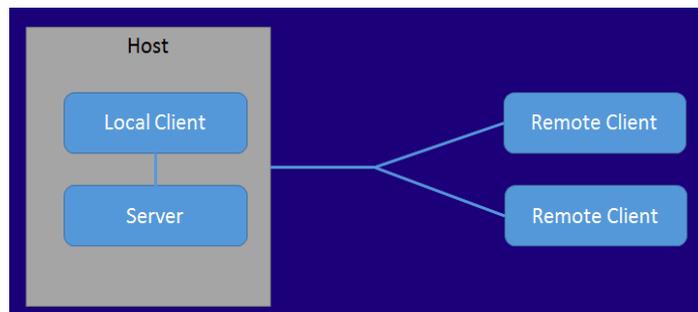


Fig. 4. Configuración de la red (Tomado de [5]).

C. Algoritmo.

La estructura tiene puntos de referencia que permiten al sistema identificar la posición de los diferentes pozos o lugares donde el estudiante puede aplicar alguna maniobra. Todas las rutas de los diversos puntos del escenario se encuentran guardadas en la base de datos, cuando el instructor asigna una tarea al alumno, el sistema automáticamente construye la ruta más corta desde la posición del ROV hasta el lugar donde deberá hacer la maniobra, posteriormente y a lo largo de todo el recorrido que haga el alumno el sistema le irá guiando, de esta manera evita confusiones y permite el el alumno complete la maniobra satisfactoriamente.

El algoritmo de Dijkstra encuentra la ruta de menor coste dentro de un conjunto de caminos previamente definidos y su funcionamiento se puede observar en la Figura 5, el algoritmo es implementado y codificado como script en lenguaje C#.

Conforme se va moviendo el ROV, el sistema detecta si va en la ruta adecuada o se va alejando de su destino, esto es posible gracias al rango de posicionamiento de los puntos de la ruta óptima calculada por Dijkstra. La Figura 6 muestra los puntos de referencia con sus respectivos rangos.

$G = (V, E)$ donde V es el conjunto de vértices y E el de arcos.
 S es el conjunto de vértices cuyos caminos más cortos al origen han sido ya determinados.
 $V-S$ es el resto de vértices.
 d : ARRAY de estimaciones de caminos más cortos a dichos vértices.
 pr : ARRAY de predecesores para cada vértice.
<<Inicializar d y pr >>
<<Poner $S = \emptyset$ >> // aún no hemos estudiado ningún vértice
While $\{V-S\} \neq \emptyset$ // mientras queden nodos sin determinar su camino mínimo al origen
 <<Ordenar los vértices en $V-S$ y analizar de acuerdo a la menor distancia al origen>>
 <<Añadir u , el vértice más cercano en $V-S$, a S >> // $S = S + \{u\}$
 <<Recalcular la distancia a todos los vértices todavía en $V-S$ adyacentes a u >>

Fig. 5. Pseudocódigo del algoritmo de Dijkstra para la búsqueda de la ruta de menor costo (Tomado de [11]).

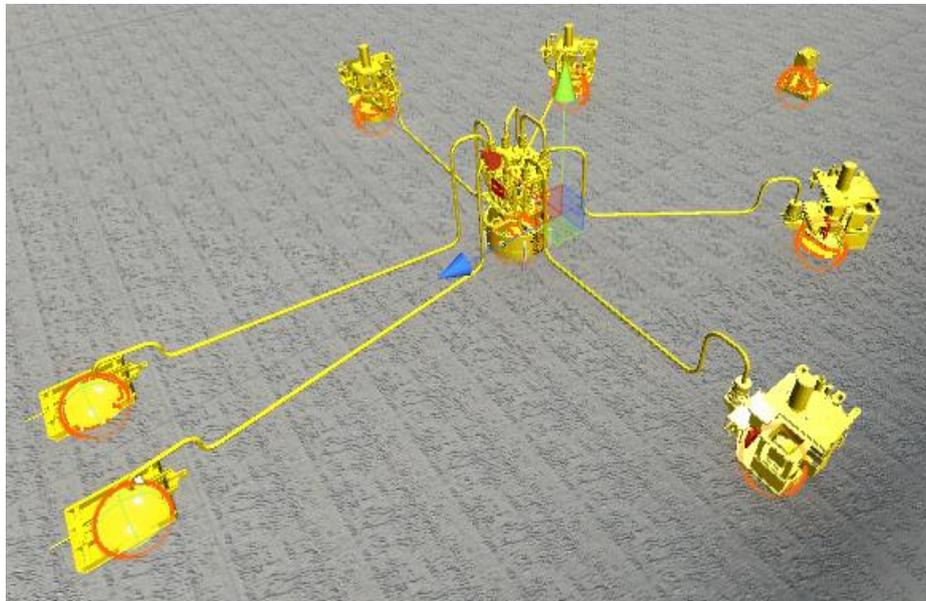


Fig. 6. Puntos de referencia que sigue el algoritmo de Dijkstra.

D. Desarrollo del modelo del estudiante.

El modelo del estudiante se programó del mismo modo que el instructor sólo restringiendo algunas funciones que tiene éste último como es el acceso a las cámaras de los demás estudiantes, la asignación de maniobras, la creación de corrientes marinas y sobre todo la capacidad de ser el Host de la red. Con lo que respecta a la adaptación de red en Unity se hace uso del componente NetworkManager proporcionado por Unity el cual controla el estado de red del sistema multi-jugador, es usado sin scripting al tener controles que permiten la configuración de sus funciones [5].

E. Software utilizado.

Para el diseño de los modelos 3D se utilizó el software de modelado, animación y renderización “3DS Max”. Para el desarrollo de la aplicación y codificación de los scripts se utilizó el motor de videojuegos “Unity3D”.

Finalmente, para poder generar las animaciones dinámicas de acuerdo a la interacción de usuario, se hizo uso de Ageia PhysX [12; 13], es el motor de física de UNITY.

5. Discusión de resultados

En SimROV convergen diferentes tecnologías en una sola herramienta de apoyo al aprendizaje. Este sistema es un simulador casi de alcance total, por otra parte, de acuerdo a la clasificación de IMCA, el sistema es de tipo A. No conforme con ser un simulador, se le trató de integrar la tecnología de Tutores Inteligentes, ya que es una herramienta de aprendizaje, además todo esto, haciendo uso de la tecnología de Realidad Virtual. Esto último evitará la restricción de contar con un ROV real para poder entrenar al personal que va a operar un ROV. En la tecnología de RV se hacen uso de la animación dinámica, es decir, los movimientos no están predefinidos, más bien serán producidos en tiempo real de acuerdo a la interacción de los usuarios y de la operación inherente a los equipos involucrados en el ambiente virtual.

El modelo del estudiante almacena datos sobre si el estudiante logró la meta de la tarea asignada o no. Por ejemplo si logró cerrar la válvula o no, a fin de evitar que continúe algún tipo de derrame, así como el tiempo que le llevó lograr la meta. También se almacena información sobre la presencia e intensidad de corrientes marinas y la diferencia promedio de la ruta de navegación seguida por el estudiante, por ejemplo para alcanzar una válvula que debe cerrar, contra la ruta óptima. Esto será un indicador de si el estudiante es capaz de controlar el ROV en presencia de corrientes marinas o simplemente si es capaz de operarlo en caso de no haber corrientes.

En el primer caso se introducen corrientes de menor intensidad a fin de que el usuario aprenda a operar el ROV dentro de ese contexto, por ejemplo recurriendo a uno de los brazos del ROV para sujetarse, mientras opera con el otro brazo, y en el segundo se despliegan sugerencias de las combinaciones de comandos, que le ayuden a aprender a operar el ROV.

En la arquitectura de la Fig. 1 puede observarse cómo el modelo del estudiante cuenta con un repositorio donde almacena la información descrita arriba, la cual es generada por el módulo del Control del ROV que monitorea las trayectorias seguidas por el estudiante al tratar de lograr una meta dictada por el instructor a través del módulo de Maniobras de Contingencia. El registro de información se realiza al momento de alcanzar la meta establecida o cuando el instructor cancela la misión, quizá debido a que ha transcurrido mucho tiempo sin haberse logrado la meta o para cancelar corrientes e introducir otras de menor o mayor intensidad.

La capacidad multiplayer de SimROV, por el momento quedará reservada para la integración de maniobras colaborativas, donde dos o más operadores puedan unir esfuerzos para realizar determinadas maniobras.

La intención es contar con una herramienta sólida que en un futuro cercano sea de utilidad en el sector petrolero de aguas profundas del país, es decir, en el ámbito real de aplicación. Entre otras posibles aplicaciones para un futuro cercano de SimROV se pueden mencionar las siguientes: a) maniobras de instalación de equipos relacionados con los pozos petroleros; b) maniobras de contingencia tales como la reparación de fugas en tuberías; c) limpieza de petróleo derramado en el mar, a fin de evitar daños a los ecosistemas marinos; d) inspección; e) mantenimiento; y f) reparación.

Usualmente esta tecnología de ROVs viene del exterior, así, este trabajo constituye un primer paso hacia el desarrollo propio en el INEEL dentro de este ámbito.

6. Conclusiones

Se ha presentado un simulador multiplayer basado en Realidad Virtual e inteligencia artificial aplicado a la industria petrolera para escenarios submarinos donde se sitúan los pozos petroleros. El escenario incluye las estructuras metálicas representativas de los pozos petroleros y uno o más ROVs (Remotely Operated Vehicle), para realizar actividades submarinas, cada ROV representan a cada uno de los estudiantes conectados en una red LAN (Local Area Network), cada estudiante tiene control solamente sobre su propio ROV, para poder realizar tareas que le encomiende el instructor.

El simulador ayuda al usuario a realizar con éxito las tareas asignadas ya que lo va guiando a lo largo de todo el proceso y le da recomendaciones en tiempo real para cada una de las maniobras, de esta manera se optimiza el tiempo y permite a los operadores aprender a controlar un ROV real ya que el comportamiento de SimROV es casi de alcance total como se mencionó anteriormente.

Una vez que ya se cuenta con la infraestructura básica sobre la operación del ROV en ambiente de simulación multiusuarios (multiplayer), en un futuro cercano se desea revisar e incluir un conjunto de situaciones de contingencia y otras actividades propias de la operación en un ambiente petrolero en el fondo del mar. Por otra parte, el modelo del estudiante es solo parte del inicio de la integración de algoritmos de inteligencia artificial dentro de SimROV, sin embargo, se desea integrar toda la tecnología de los Sistemas Tutores Inteligentes (STI), a fin de que el sistema pueda ofrecer mayores facilidades como herramienta de apoyo al entrenamiento.

Referencias

1. Gopinath, C., Sawyer, J.: Exploring the Learning from an Enterprise Simulation. En: Journal of Management Development (1999)
2. Merino, A., Pelayo, S., Rueda, A., Alves, R., García, A., Acebes, F., de Prada, C., Gutiérrez, G., García, M.: Un simulador de alcance total para la formación de los operarios de sala de control de factorías azucareras. En: 2ª Reunión de Usuarios de EcosimPro, UNED, Madrid 24–25 Febrero (2003)
3. IMCA: Guidance on the Use of Simulators. International Marine Contractors Association, Sitio web: <https://www.imca-int.com/?s=simulator> (2016)
4. Olguín-Tolentino, N. D., Ingram-Ramírez, J. W., Pérez-Ramírez, M., Nava-Ayala, E. R., Hernández-Pérez, M. Y.: Prototype of a ROV simulator using dynamic animation and Virtual Reality. En: CIIDET-Congreso Internacional sobre Innovación y Desarrollo Tecnológico (2016)
5. Unity: Documentation Network Manager. Unity Technologies. Sitio web: <https://docs.unity3d.com/es/current/Manual/UNetManager.html> (2016)
6. Unity: 3D physics reference. Unity Technologies. Sitio web: <https://docs.unity3d.com/es/current/Manual/class-Rigidbody.html> (2016)
7. NVIDIA: Tecnología PHYSX. NVIDIA Corporation. Sitio web: <http://www.nvidia.es/object/nvidia-physx-es.html> PhysX (2016)
8. Craighead, J., Murphy, R., Burke, J., Goldiez, B.: A survey of commercial & open source unmanned vehicle simulators. In: Robotics and Automation, 2007 IEEE International Conference on, IEEE, pp. 852–857 (2007)
9. Unity: Mesh Collider. Unity Technologies. Sitio web: <https://docs.unity3d.com/es/current/Manual/class-MeshCollider.html> (2016)
10. Marine Well Containment Company: Containment System Overview. MWCC's Marine Well Containment System, 1, pp. 1 (2007)
11. Torrubia, G. S., Terrazas, V. L.: Algoritmo de Dijkstra. Un tutorial interactivo. En: VII Jornadas de Enseñanza Universitaria de la Informática (2012)
12. Tumult: Scene Physics Gravity. Tumult Inc. Sitio web: <http://tumult.com/hype/documentation/3.0/#physics> (2016)
13. Craighead, J., Murphy, R., Burke, J., Goldiez, B.: A survey of commercial & open source unmanned vehicle simulators. In: Robotics and Automation, 2007 IEEE International Conference on IEEE, pp. 852–857 (2007)

Un primer paso hacia el reconocimiento automático de escarabajos

Caleb Jiménez¹, Maya Carrillo¹, Hortensia Carrillo-Ruiz²,
María de Lourdes Sandoval S.¹, Luis Alfredo Moctezuma¹

¹ Benemérita Universidad Autónoma de Puebla, Facultad de ciencias de la computación,
Puebla, México

² Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Biológicas, Puebla,
México

calebji@hotmail.com, cmaya@cs.buap.mx,
{hortensia.carrillo, malos1753, luisalfredomoctezuma}@gmail.com

Resumen. Los escarabajos pertenecen al orden *Coleoptera*, y son organismos importantes para el funcionamiento de los ecosistemas en los que habitan. Los taxónomos, clasifican a estos organismos para facilitar su identificación y estudio. Para ello emplean claves de identificación compuestas por dilemas, mismos que aumentan en número al acercarse a la categoría menos inclusiva. En este trabajo se presenta un método para identificar escarabajos de manera automática empleando como atributos, medidas estadísticas y la transformada elíptica de Fourier. Se utilizaron 48 imágenes de 8 especies diferentes recolectadas en la región de Jolalpan en el estado de Puebla. Después de procesar las imágenes y extraer los atributos, se obtiene en la clasificación una exactitud del 89.58% utilizando *Random Forest*, con validación cruzada a 10 pliegues.

Palabras clave: identificación taxonómica, escarabajos, transformada elíptica de Fourier, medidas estadísticas.

A First Step towards Automatic Beetle Recognition

Abstract. Beetles belong to the order *Coleoptera*, and are important organisms for the ecosystems functioning in which they inhabit. Taxonomists classify these organisms to facilitate their identification and study. For this purpose, they use identification keys, composed by dilemmas, which increase in number when classification approaches to the least inclusive category. This work presents a method to automatically identify beetles using as attributes, statistical measures and the Elliptic Fourier Transform. 48 images of 8 different species collected in the Jolalpan region in Puebla state were used. After processing the images and extract the attributes, a classification accuracy of 89.58% was obtained using Random Forest, and 10-folds cross-validation.

Keywords: taxonomic identification, beetles, Fourier elliptic transform, statistical measures.

1. Introducción

Dentro de la clase Insecta, el orden Coleoptera es uno de los grupos biológicos más diversos. Los escarabajos pertenecen a este orden y son organismos clave para el funcionamiento de los ecosistemas en los que habitan, ya que entre sus funciones se encuentra la degradación de la materia orgánica y su incorporación al suelo, favoreciendo así la circulación de la energía almacenada en los desechos orgánicos, esto los convierte en recicladores naturales. Además, estos insectos ocupan los primeros eslabones de la cadena trófica, siendo fuente importante de alimento para aves y mamíferos [1].

Debido a su morfología, colores llamativos e importancia ecológica, han sido estudiados, desde distintas áreas de la Biología, entre las que se encuentra la Sistemática. Los taxónomos elaboran esquemas de clasificación con el propósito de organizar a los diferentes organismos y de esta forma facilitar su estudio. Los sistemas de Clasificación consisten en categorías jerarquizadas, son una estructura de grupos inclusivos en donde cada grupo o categoría está definida por características compartidas entre los organismos, siendo en estos momentos la categoría más inclusiva reconocida en la Clasificación Biológica el Dominio y la menos inclusiva la Especie [2].

Con los sistemas de clasificación, los biólogos establecen relaciones de identidad entre un organismo particular y una categoría a la cual pertenece, es decir, un ejemplar puede identificarse taxonómicamente con un nombre (taxón) dentro de cada categoría. Para lograr esta relación de identidad, se deben construir claves de identificación para cada categoría a partir de una secuencia ordenada de dilemas o disyuntivas, donde cada dilema sucesivo va planteando dilemas cada vez más restringidos y finalmente se llega al nombre del grupo o taxón al que pertenece [2]. Por lo tanto, las claves de identificación resultan útiles al momento de determinar el nombre de alguna especie, sin embargo, la tarea de identificar no es fácil, ya que en las claves, conforme la tarea de clasificación se acerca a la categoría menos inclusiva (especie), los dilemas aumentan en número y complejidad, pudiendo ser más de 400 dilemas a resolver antes de poder determinar el nombre específico. Además, sólo pueden ser empleadas por especialistas, ya que se requiere de conocer los términos propios de la biología y de equipo especializado como microscopios estereoscópicos con un zoom mínimo de 0.5x.

Por lo anteriormente expuesto, se plantea este trabajo como un primer intento para generar una herramienta automática que facilite el trabajo de identificación de estos organismos, a partir de la transformada elíptica de Fourier para extraer características del borde del cuerpo del escarabajo y medidas estadísticas.

El presente artículo está organizado de la siguiente manera en la sección 2 se describen algunos trabajos relacionados, en la sección 3 se describe el método propuesto. En la sección 4 los experimentos y resultados y finalmente en la sección 5 las conclusiones y trabajo futuro.

2. Trabajos relacionados

Adams, et al. en [3] plantean un enfoque denominado Morfología geométrica. Este enfoque se basa en la digitalización de contornos o puntos clave, cuyas conformaciones

espaciales son analizadas matemáticamente. Dentro de este grupo de técnicas, los métodos de contornos se basan en la digitalización de los puntos alrededor de un contorno para ajustarlos a una función matemática, generalmente derivada del análisis de Fourier. Posteriormente los coeficientes de éste, pueden emplearse en estudios comparativos.

Tabla 1. Ejemplares de escarabajos empleados en el estudio, se muestran los nombres que les corresponden dentro de cada una de las categorías taxonómicas a partir de familia.

Familia	Subfamilia	Tribu	Género, subgénero y especie	No. ejemplares	
Scarabaeidae	Scarabaeinae	Scarabaeini	Canthon (Canthon) humectus	6	
			Canthon (Canthon) indigaceus	6	
			Deltochilum gibbosum sublaeve	6	
			Deltochilum tumidum	5	
			Coprini	Copris incertus	2
				Dichotomius colonicus	6
				Dichotomius amplicollis	1
			Phanaeini	Ateuchus rodriguezi	6
				Phanaeus mexicanus	7
			Onthophagini	Coprophanaeus (Coprophanaeus) pluto	3
		Onthophagus lecontei		1	
		Onthophagus mextexus		2	
		Aphodiinae	Aphodiini	Digitonthophagus gazella	6
				Labarrus pseudolividus	2
				Eupariini	Ataenius castaniellus
Hybosoridae	Hybosorinae		Hybosorus illigeri	5	

Los descriptores elípticos de Fourier, propuestos por Kuhl y Giardina [4], pueden delinear cualquier forma con un contorno cerrado bidimensional. Estos han sido aplicados al análisis de varias formas biológicas.

Así, Rohlf y Archie en [5] presentan un análisis del contorno de las alas de 127 especies de mosquito del norte de México y comparan diferentes descriptores de Fourier, sin normalizar, normalizado, con coordenadas polares, entre otros y menciona que los descriptores de la transformada elíptica de Fourier prometen resultados satisfactorios para la diferenciación entre especies.

Por otra parte, Furuta et al. en [6] analiza la forma de las hojas del frijol de soya empleando los coeficientes de los armónicos de la transformada elíptica de Fourier, emplea 20 armónicos normalizados y por lo tanto 77 descriptores por imagen de hoja. Con ellos realiza análisis de componentes principales y encuentra que la contribución acumulativa del quinto componente es del 96%. Concluye que los coeficientes de

Fourier proporcionan una medida cuantitativa poderosa para evaluar la forma de las hojas del frijol de soya.

Zhan y Wang en [7] utiliza la transformada elíptica de Fourier para analizar la forma del contorno del ala de cinco especies de Antlion (hormiga león parecida a la libélula). Ellos analizaron variaciones de la forma del perfil de alas de 5 especies del género: Considerando un total, 98 alas anteriores y 98 alas posteriores. Utiliza los veinte primeros armónicos de Fourier que resume a través de un análisis de componente principal y considera los primeros 8 componentes principales de la variación de forma para realizar pruebas estadísticas (análisis de varianza multivariable, análisis de variables canónicas y análisis de conglomerados). Concluye que sus resultados del análisis de Fourier de contorno de alas de las 5 especies están de acuerdo con el sistema taxonómico actual.

Recientemente los descriptores elípticos de Fourier, han sido empleados como para proponer la construcción de herramientas para la identificación automatizada de especies como lo mencionan Singh, et al. en [8] quienes han desarrollado un sistema automatizado de reconocimiento de especies de bambú basado en las características de forma de la vaina de Culm de bambú usando momentos de Fourier y de Legendre. Concluyen que el momento de Fourier tiene resultados significativamente mejores que el momento de Legendre, obteniendo un 100% de exactitud en la clasificación. El autor comenta que su sistema puede eliminar la necesidad de métodos laboriosos de reconocimiento humano que requieran un taxónomo de plantas. Los resultados obtenidos muestran una considerable precisión de reconocimiento, demostrando que las técnicas utilizadas son adecuadas para ser implementadas con fines comerciales.

Yang et al. en [9] presentan un trabajo para la identificación de insectos basado en el borde de sus alas reportando una exactitud media para la identificación de especies que varía entre el 90% y 98% . Utilizan 120 ejemplares de 7 especies distintas. Emplean la transformada elíptica de Fourier con 30 armónicos y consideran 117 coeficientes para obtener las características del borde de las alas de estos insectos.

Para la clasificación emplean máquinas de soporte vectorial. Finalmente los descriptores elípticos de Fourier invariantes, simétricos, asimétricos y estandarizados, también se ha utilizados para clasificar granos en [10] Mebatsion et al., evalúan la forma de cuatro diferentes tipos de grano: cebada, avena centeno y trigo. Utilizan 100 imágenes por cada grano, los autores mencionan que los descriptores elípticos de Fourier ofrecen buenos resultados considerando la variabilidad de la forma de los granos y que podrían utilizarse para realizar clasificación no supervisada.

Hasta el momento no se tiene información de la utilización de los descriptores elípticos de Fourier para el estudio de las formas de los escarabajos.

3. Metodología

Los pasos para la identificación automática de escarabajos son: (1) adquisición de imágenes, (2) preprocesamiento de imágenes, (3) extracción de características y (4) clasificación de los datos. Pasos que se describen en las secciones subsecuentes.

3.1. Adquisición de imágenes

Las imágenes fueron obtenidas de un grupo de ejemplares de colección provenientes de la región de Jolalpan en el estado de Puebla, estos ejemplares fueron recolectados durante un año en esta zona, su captura no es sencilla por lo que son pocos los representantes empleados en este estudio [11]. En total se emplearon 67 especímenes conservados en seco, los cuales fueron previamente identificados de manera tradicional empleando las claves de identificación para las especies de escarabajos del estado de Puebla [12] y pertenecen a dos familias, tres subfamilias, seis tribus, 12 géneros y 16 especies (Tabla 1).



Fig. 1. Escarabajo en posición dorsal y escala de referencia de diez milímetros.



Fig. 2. Obtención de imágenes de los ejemplares estudiados.

Para obtener las imágenes, cada uno de los ejemplares se reblandecieron con vapor de agua para poder retirar el alfiler entomológico que se emplea para conservarlos. Posteriormente fueron colocados en posición dorsal sobre una placa de unicel forrada con papel blanco, sobre la cual se dibujó una escala de referencia en milímetros (Fig. 1). Finalmente fueron fotografiados con una cámara canon 1068C001AA montada sobre un tripie (Fig. 2).

3.2. Pre-procesamiento de imágenes

Para extraer las características de las imágenes, estas deben pre-procesarse previamente, para ello: Se recortaron los escarabajos rodeados por un margen de 15

pixeles, figura 3a; se convirtieron a escala de grises y se bina-rizaron figura 3b. Posteriormente, se eliminaron las regiones menores de 100 pixeles, figura 3c. Estos 100 pixeles se determinaron en función de la línea de referencia, figura 1, misma que tiene 104 pixeles. Mantener dicha línea será de utilidad en futuros experimentos.

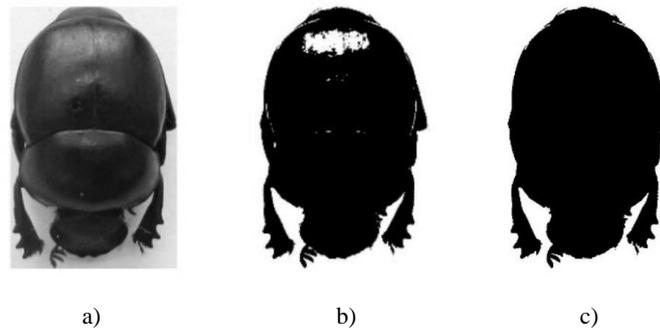


Fig. 3. Pre-procesamiento de las imágenes de un ejemplar de la especie *Canthon (Canthon) humectus*: a) imagen recortada, b) imagen bina-rizada, c) imagen después de eliminar regiones menores de 100 pixeles.

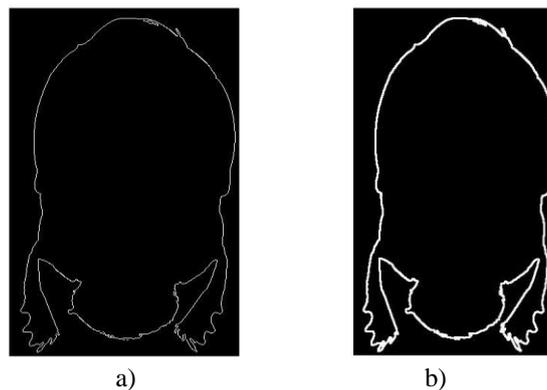


Fig. 4. Imagen de un ejemplar de la especie *Canthon (Canthon) humectus* a) aplicación de mascara sobel b) aplicación de la operación de dilatación.

Posteriormente se procedió a detectar el borde de las imágenes aplicando el algoritmo Sobel de detección de borde, definido de manera muy simple debido a que solo utilizamos ocho puntos de la imagen alrededor del punto analizar ya que buscamos un cambio abrupto en nuestra imagen bina-rizada que representa el borde de nuestra imagen figura 4.

Finalmente se aplicó la operación morfológica de dilatación que consiste en adicionar pixeles a la frontera de la imagen, el procedimiento consiste en analizar la imagen por punto considerando sus ocho vecinos más cercanos y tener un contorno completamente delimitado, utilizamos la función de Matlab *imdilate*, figura 4b.

Las coordenadas (x, y), puntos del contorno, se emplearon para obtener las características para el proceso de clasificación.

3.3. Extracción de características

Las coordenadas (x, y) obtenidas en el paso anterior se utilizaron para calcular descriptores elípticos de Fourier. Dichos descriptores fueron calculados empleando códigos de cadena para describir los contornos de las imágenes como se describe en [4]. El algoritmo no requiere que los puntos sean equidistantes, y puede describir cualquier contorno cerrado siempre y cuando el número de armónicos sea el adecuado. Los coeficientes, para el n-ésimo armónico de la expansión elíptica de Fourier de la secuencia de coordenadas x, son:

$$A_n = \frac{T}{2n^2\pi^2} \sum_{p=2}^k \frac{\Delta x_p}{\Delta t_p} (\cos[2n\pi t_p/T] - \cos[2n\pi t_{p-1}/T]), \quad (1)$$

$$B_n = \frac{T}{2n^2\pi^2} \sum_{p=2}^k \frac{\Delta x_p}{\Delta t_p} (\sin[2n\pi t_p/T] - \sin[2n\pi t_{p-1}/T]), \quad (2)$$

donde k es el número de pasos (segmentos) en el trazo del contorno, Δx_p es el desplazamiento en el eje x entre los pasos p-1 y p, Δt_p es la longitud del segmento entre los pasos p-1 y p, t_p es la longitud acumulada de los segmentos y $T=t_k$ es la longitud total del contorno aproximado por el trazo de un polígono. Los coeficientes para la coordenada y, C_n y D_n se calculan de igual manera empleando los datos para el eje y.

Los coeficientes calculados se normalizaron para hacerlos invariantes al tamaño, a la rotación, y punto de inicio, como se describe en [4]. Después de la normalización los primeros tres coeficientes del primer armónico son constantes y por lo tanto no considerados como atributos. El resto de coeficientes se utilizaron como atributos en el proceso de clasificación.

Por otra parte también se calcularon medidas estadísticas, considerando las coordenadas x e y, estas fueron: a) media, b) mediana, c) máximo, d) mínimo, e) suma, f) desviación estándar, g) varianza, h) curtosis, i) oblicuidad.

3.4. Clasificación de datos

Después de extraer los atributos para cada imagen, estos fueron los datos de entrada para tres clasificadores: a) Máquinas de soporte vectorial (Support vector machine, SVM) que es de aprendizaje supervisado, Bayes ingenuo (Naive Byes) donde el aprendizaje es supervisado y Selvas aleatorias (Random Forest) que es de aprendizaje no supervisado.

Estos clasificadores fueron ejecutados primero considerando sólo los atributos estadísticos, es decir 18 atributos por imagen. Después los coeficientes de los armónicos de la transformada elíptica de Fourier y finalmente ambos conjuntos. Se utilizó validación cruzada a 10 pliegues. En la siguiente sección se muestran los resultados obtenidos.

4. Experimentos y resultados

Para realizar los experimentos se consideraron únicamente las especies de la familia Scarabaeidae, que tienen 5 ejemplares o más para un total de 48 imágenes, para este trabajo la orientación de la imagen es importante pues en la extracción de coordenadas para las medidas estadísticas se consideran las imágenes en una misma posición que es vertical con la masa antenal en la parte inferior izquierda. Para estas imágenes se calcularon las medidas estadísticas mencionadas en la sección 3.3. Los resultados se muestran en la Tabla 2.

Tabla 2. Resultados de la exactitud obtenidos para las imágenes de la familia Scarabaeidae caracterizadas con medidas estadísticas.

No. puntos	Random Forest	SVM	Naive Bayes
30	83.34	77.08	83.34
50	83.34	83.34	75.00
100	83.34	83.34	85.42
250	81.25	79.17	79.17
500	79.17	83.34	83.34

Como puede observarse en la Tabla 2, ocho resultados alcanzan una exactitud del 83.33 %, dos de ellos empleando 18 atributos.

Tabla 3. Resultados de la exactitud obtenidos para las imágenes de la familia Scarabaeidae caracterizadas con los coeficientes de los armónicos de la transformada elíptica de Fourier.

No.armonico	No.puntos	Random Forest	SVM	Naive Bayes
2	30	47.92	12.50	47.92
2	50	47.92	12.50	47.92
2	100	47.92	12.50	47.92
4	30	41.67	22.92	52.08
4	50	41.67	22.92	52.08
4	100	41.67	22.92	52.08
8	30	35.42	39.58	41.67
8	50	35.42	39.58	41.67
8	100	35.42	39.58	41.67
16	30	18.75	29.17	29.17
16	50	18.75	29.17	29.17
16	100	18.75	29.17	29.17

Posteriormente se clasificaron los mismos ejemplares pero ahora calculando los coeficientes para los armónicos de la transformada elíptica de Fourier, variando el número de puntos y armónicos, los resultados obtenidos se muestran en la Tabla 3.

En la Tabla 3, la exactitud de la clasificación no se altera variando el número de puntos. Pero si al variar el número de armónicos. La mejor exactitud se obtiene con 4 armónicos, lo que nos genera 13 atributos, con el clasificador Naive Bayes y 30 puntos, lo implican menos tiempo de procesamiento según nuestra tabla.

Nuestras tablas muestran que estadísticamente tenemos mejores resultados comparados con los descriptores de Fourier y esto se debe a que la forma de nuestros escarabajos no es totalmente elíptica pues en la parte de la cabeza cuenta con la maza antenal, la tibia y el tarsus.

Finalmente se combinaron los atributos estadísticos con los descriptores de Fourier, considerando los resultados con 30 puntos pues aportan la misma exactitud y menor tiempo de procesamiento. Los resultados están en la Tabla 4.

Tabla 4. Resultados de la exactitud obtenidos para las imágenes de la familia Scarabaeidae combinando atributos estadísticos y descriptores de Fourier.

No. puntos Estadística	No. armónicos	No.puntos descriptor	Random Forest	SVM	Naive Bayes
30	2	30	85.42	81.25	79.17
30	4	30	87.50	83.34	87.50
30	8	30	89.58	81.25	83.34
30	16	30	81.25	75.00	79.17

Como puede observarse en la Tabla 4, la exactitud de la clasificación mejora cuando se combinan los atributos, obteniendo una mayor exactitud igual a 89.58% con Random Forest, 30 puntos para calcular las medidas estadísticas y los descriptores de Fourier. En la Tabla 5 se muestra la matriz de confusión para estos resultados donde puede observarse que los errores se dan entre las especies *Deltochilum gibbosum sublaeve* y *Deltochilum tumidum* que de acuerdo a la Tabla 1 pertenecen a la misma tribu. En la Tabla 6 se presentan los resultados para la precisión, recuerdo y medida F donde puede observarse que el promedio de la precisión es de 0.90 y de medida F de 0.89.

Tabla 5. Matriz de confusión para la clasificación obtenida.

```

a b c d e f g h  <-- classified as
6 0 0 0 0 0 0 0 | a = AteuchusRodriguezi
0 6 0 0 0 0 0 0 | b = Canthon (canthon) humectus
0 0 6 0 0 0 0 0 | c = Canthon (canthon) Indigaceus
0 0 0 5 1 0 0 0 | d = Deltochilumgibbosumsublaeve
0 0 0 3 2 0 0 0 | e = Deltochilumtumidum
0 0 0 0 0 6 0 0 | f = DichoTomiusColonicus
0 0 0 0 0 0 6 0 | g = Digitonthophagusgazella
0 0 0 0 0 1 0 6 | h = Phanaeusmexicanus
    
```

Tabla 6. Precisión, recuerdo y medida F obtenidos.

Especie	Precisión	Recuerdo	Medida F
Ateuchus rodriguezi	1.00	1.00	1.00
Canthon (Canthon) humectus	1.00	1.00	1.00
Canthon (Canthon) indigaceus	1.00	1.00	1.00
Deltochilum gibbosum sublaeve	0.63	0.83	0.71
Deltochilum tumidum	0.67	0.40	0.50
Dichotomius colonicus	0.86	1.00	0.92
Digitonthophagus gazella	1.00	1.00	1.00
Phanaeus mexicanus	1.00	0.86	0.92
Phanaeus mexicanus	1.00	0.86	0.92
Promedio	0.90	0.89	0.89

En la Figura 5 se muestra la aproximación al contorno de un ejemplar con 30 puntos y 8 armónicos es decir 29 atributos, puede observarse que aún hay imprecisiones considerables.

5. Conclusiones y trabajo futuro

Las medidas estadísticas proporcionan mejor exactitud al clasificar los escarabajos empleando 18 atributos contra 29 empleados con la transformada elíptica de Fourier con 8 armónicos. Si bien los descriptores de Fourier alcanzan un máximo de 52 % de exactitud pensamos que esto se debe al escaso número de imágenes con que contamos, únicamente 48, mientras que en el trabajo relacionado se cuenta con más de 100 imágenes

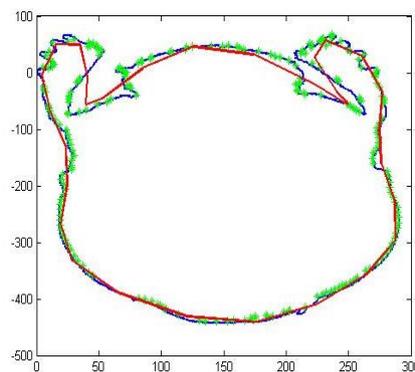


Fig. 5. Aproximación al contorno de una ejemplar de la especie *Canthon (Canthon) humectus* empleando la transformada elíptica de Fourier.

Los errores en la clasificación se dan entre ejemplares de la misma tribu, lo que implica que de manera natural tienen ciertos atributos en común, así que estos errores deben disminuir al contar con más ejemplares. En cinco especies la precisión es igual a 1, con el 100 por ciento de ejemplares clasificados correctamente. Como trabajo futuro se experimentará agregando nuevos ejemplares y otros atributos como el color e identificación de líneas o manchas específicas de los ejemplares.

Bibliografía

1. Castañeda-Osorio, R., Carrillo-Ruiz, H., Rivas-Arancibia, S. P., Sánchez- Carrillo, M.: Melolonthidae y Cetoniidae (Coleoptera: Scarabaeoidea) en el Rancho El Salado, Jolalpan, Puebla, México. *Dugesiana*, 22(2), pp. 227–241 (2015)
2. Morrone, J. J.: *Sistemática. Fundamentos, métodos, aplicaciones*. Facultad de Ciencias, UNAM, Ciudad de México (2013)
3. Adams, D. C., Rohlf, F. J., Slice, D. E.: Geometric morphometrics: ten years of progress following the ‘revolution’. *Italian Journal of Zoology*, 71, pp. 5–16 (2004)
4. Kuhl, F. P., Giardina, C. R.: Elliptic Fourier features of a closed contour. *Computer graphics and image processing*, 18, pp. 236–258 (1982)
5. Rohlf, F. J., Archie, J. W.: A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Systematic Zoology*, 33, pp. 302–317 (1984)
6. Furuta, N., Ninomiya, S., Takahashi, S., Ohmori, H., Ukai, Y.: Quantitative evaluation of soybean (*Glycine max* L. Merr.) leaflet shape by principal component scores based on elliptic Fourier descriptor. *Breeding Science*, 45, pp. 315–320 (1995)
7. Zhan, Q. B., Wang, X. L.: Elliptic Fourier Analysis of the Wing Outline Shape of Five Species of Ant- lion (Neuroptera: Myrmeleontidae: Myrmeleontini). *Zoological Studies*, 51(3), pp. 399–405 (2012)
8. Singh, K., Gupta, I., Gupta, S.: Classification of Bamboo Species by Fourier and Legendre Moment. *International Journal of Advanced Science and Technology*, 50, pp. 61–70 (2013)
9. Yang, H. P., Ma, C. S., Wen, H., Zhan, Q. B., Wang, X. L.: A tool for developing an automatic insect identification system based on wing outlines. *Scientific reports* (2015)
10. Mebatsion, H. K., Paliwal, J., Jayas, D. S.: Evaluation of variations in the shape of grain types using principal components analysis of the elliptic Fourier descriptors. *Computers and electronics in agriculture*, 80 (2012)
11. Sánchez-Velázquez, B., Carrillo-Ruiz, H., Morón, M. Á., Rivas-Arancibia, S. P.: Especies de Scarabaeidae e Hybosoridae (Coleoptera: Scarabaeoidea) que habitan en la comunidad del Rancho El Salado, Jolalpan, Puebla, México. *Dugesiana*, 18(2), pp. 207–215
12. Morón, M. A., Aragón-García, A., Carrillo-Ruiz, H.: Fauna de escarabajos del estado de Puebla. Publicado por MA Morón, Coatepec, Veracruz, México (2013)

Clasificador de edad en imágenes digitales usando métodos estadísticos

Luis Enrique Colmenares Guillen¹, Roberto Guillermo Torres López¹,
Maya Carrillo Ruiz¹, Francisco Javier Albores Velasco²

¹ Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla,
Puebla, México

² Facultad de Ingeniería y Tecnología, Universidad Autónoma de Tlaxcala, Tlaxcala, México

{lecolme, robert.torres.lopez, crllrzmy, javier.albores} @gmail.com

Resumen. La estimación automática de edad tiene considerables aplicaciones en áreas como el marketing, como por ejemplo, al generar y mostrar contenido específico para cualquier grupo etario, y en la seguridad informática donde se permitiría proteger a menores de edad de contenidos no aptos para su edad. El objetivo general de este trabajo es mostrar la metodología desarrollada para generar un sistema clasificador basado en apariencia utilizando los modelos de representación: análisis de componentes principales y análisis discriminante lineal; y exponer los resultados obtenidos sobre las bases de imágenes FG-Net y IMDB-Wiki con las cuales se clasificaron en dos grupos, mayores (+18) y menores (-18) de edad donde se obtuvo hasta un 89% de efectividad.

Palabras clave: FisherFaces, estimación de edad, análisis de componentes principales, análisis discriminante lineal, imágenes faciales, diferenciador de minoría de edad.

Age Classifier in Digital Images Using Statistical Methods

Abstract. The automatic age estimation has considerable applications in areas such as marketing, for example, while generating and displaying specific content for some age group, and in computing security, where protecting minors from contents not suitable for their age would be allowed. The general objective of this work is to show the methodology developed to generate a classifier system based on appearance, using representation models: analysis of main components and a linear discriminant analysis; and to present the results so obtained on the FG-Net and IMDB-Wiki images basis, with which were classified into two groups; major (+18) and minor (-18) of age, where up to an 89% of effectiveness was obtained.

Keywords: FisherFaces, age estimation, principal component analysis, linear discriminant analysis, facial images, age minority differentiator.

1. Introducción

El reconocimiento de diferentes variaciones faciales como la expresión, el género o la identidad de una persona han sido extensamente estudiadas, en comparación con la estimación automática de edad que impacta directamente en áreas de marketing como indican Santos, Sánchez, Carmonet, Guerra y Casanova en [1], la seguridad informática de menores de edad sostiene Suárez Ricardo en [2] y en caso concreto permitirá al laboratorio de análisis forense de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla implementar un detector de sitios web que almacenen y distribuyan contenido relacionado con pedofilia por ello se define el criterio de los grupos etarios de mayoría y minoría de edad.

Si bien, para los humanos es relativamente fácil clasificar a las personas en grupos de edad mediante rasgos faciales, para un computador no lo es debido a que el rostro humano es un objeto dinámico que tiene un alto grado de variabilidad y el proceso de envejecimiento de las caras es incontrolable y personalizado explican Nabil, Olwan, Tubeel, El-sar & Sultan en [10].

1.1. Modelos de representación de características

En 1994 Kwon Young propuso por primera vez en [3] un modelo antropométrico donde la idea principal es consultar los estudios biológicos del desarrollo del cráneo que permite describir el crecimiento de la cabeza de un individuo desde su infancia hasta su edad adulta.

Así mismo, los puntos fiduciales permiten extraer de forma automática las características de un rostro que son registradas en un vector de 16 componentes que contiene relaciones entre distancias y áreas obtenidas entre los puntos marcados como demuestran Cox, Ghosn y Yianilos en [4].

Por último, Hayashi, Yasumoto e Ito en [5] utilizan los modelos de apariencia en los cuales se considera la textura de la piel y de la forma del rostro para caracterizar de forma adecuada cada una de las imágenes faciales. En esta publicación se realiza una estimación de edad por medio de un esquema de clasificación grupal usando una descripción semántica del rostro para la caracterización de los rasgos faciales.

1.2. Algoritmos de estimación de edad

La estimación de edad es un proceso de reconocimiento de patrones donde cada uno de los descriptores de edad es considerado una clase, por lo tanto, puede ser resuelta por medio de clasificadores por ejemplo el vecino más cercano como lo utiliza Khaung Tin en [11] o con redes neuronales demostrado por Hewahi, Olwan, Tubeel en [12] o puede abordarse como un problema de regresión donde la edad es un conjunto de valores secuenciales, se han desarrollado tres formulaciones para la función de envejecimiento: lineal, cuadrática y cubica. El modelo óptimo se obtuvo con imágenes faciales de diferentes edades basado en un algoritmo genético demostrado por Lanitis, Taylor, Cootes en [6].

1.3. Métodos estadísticos como modelos de representación

Los métodos estadísticos: análisis de componentes principales y análisis discriminante lineal son representados a través de los modelos de representación Eigenfaces y Fisherfaces.

La técnica de Eigenfaces fue propuesta por Turk y Pentland en [7] la cual consiste en encontrar los componentes principales de la distribución de rostros o los eigenvectores de la matriz de covarianza del conjunto de imágenes faciales, estos eigenvectores pueden ser utilizados como el conjunto de características que representan la variación entre las imágenes faciales.

Por otro lado, Belhumeur, Hespanha y Kriegman sostienen en [8] que las fisherfaces aumentan la relación entre la distribución de las distintas clases y la distribución intra-clases.

1.4. Análisis de componentes principales

Es una técnica que toma una proyección lineal y maximiza la dispersión de todas las imágenes proyectadas. En primer lugar, se considera el conjunto de N imágenes con valores en el espacio de imágenes n -dimensional. Como lo indica la formula (1):

$$\{x_i\} \quad i = 1, 2, \dots, N. \quad (1)$$

Se asume además que cada una de las imágenes pertenece a una de las c clases $\{x_1, x_2, \dots, x_c\}$ Asimismo, se considera una transformación lineal que lleva el espacio de imágenes original de n dimensiones al espacio de características de dimensión m , donde $m < n$. Los nuevos vectores de características $Y_k \in \mathfrak{R}_m$ son definidos por la siguiente transformación lineal, formula (2):

$$Y_k = W^T X_k \quad K = 1, 2, \dots, N, \quad (2)$$

donde $W \in \mathfrak{R}^{n \times m}$ es una matriz con columnas ortonormales. Se define además la matriz de distribución total S_T como la formula (3):

$$S_T = \sum_{K=1}^N (x_k - \mu)(x_k - \mu)^T. \quad (3)$$

Donde $\mu \in \mathfrak{R}_n$ es la media de todas las imágenes de (1). Luego de aplicar la transformación lineal W_T , la distribución de los vectores de características $\{y_1, y_2, \dots, y_N\}$ es $W_T S_{T_w}$. Se toma aquella proyección W_{opt} que maximiza el determinante de la distribución total de la matriz de las N imágenes. Como lo indica la formula (4):

$$W_{opt} = \arg \max W |W^T S_T W| = [w_1, w_2 \dots w_m]. \quad (4)$$

1.5. Análisis discriminante lineal

Este método selecciona el W de la formula (2), de manera que el cociente entre la distribución de las diferentes clases y la distribución intra-clases sea máxima. Para esto, se define la matriz S_B de distribución entre clases como lo indica la formula (5):

$$S_T = \sum_{i=1}^c N_i (\mu_i - \mu) (\mu_i - \mu)^T. \quad (5)$$

Y la matriz S_W de distribución intra-clases, como se puede apreciar en la formula (6).

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} N_i (\mu_i - \mu) (\mu_i - \mu)^T, \quad (6)$$

donde μ_i es la imagen media de la clase X_i , y N_i es el número de imágenes en la clase X_i . Si la matriz S_W es no singular, la proyección W_{opt} se elige como la matriz con columnas ortonormales que aumenta el cociente del determinante de la matriz de distribución entre clases de las imágenes proyectadas y el determinante de la matriz de la distribución intra-clases de las N imágenes, de la formula (7):

$$W_{opt} = \arg \max W \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1, w_2 \dots w_m], \quad (7)$$

donde $\{W_i \mid i = 1, 2, \dots, m\}$ es el conjunto de valores propios de S_B y S_W correspondiente a los m mayores valores propios $\{\lambda_i \mid i = 1, 2, \dots, m\}$, entonces, se muestra de la formula (8).

$$S_B W_i = \lambda_i S_W W_i \quad i = 1, 2, \dots, m. \quad (8)$$

Por lo tanto, se obtienen $c - 1$ valores propios distintos de cero, y por consiguiente el límite superior de m es $c - 1$, donde c es el número de clases, se muestra formula (9):

$$W_{pca}^T = W_{fld}^T W_{pca}^T, \quad (9)$$

donde se derivan las formulas (10) y (11).

$$W_{pca} = \arg \max W |W^T S_T W|, \quad (10)$$

$$W_{fld} = \arg \max W \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|}. \quad (11)$$

2. Trabajos relacionados

Kwon y Lobo proponen en [9] un modelo antropométrico donde la idea principal es consultar los estudios biológicos del desarrollo craneo facial existentes en la literatura. Esta teoría utiliza un modelo matemático de transformación craneal para describir el crecimiento de la cabeza de un individuo desde su infancia hasta su edad adulta.

El modelo de apariencia considera que la textura de la piel y de la forma del rostro pueden ser registrados en un vector de 16 o 71 componentes que contienen relaciones entre distancias y áreas obtenidas entre los puntos marcados como explican Cox et al. En [4] en ambas investigaciones el mayor problema es la incapacidad de estimar la edad en imágenes faciales que no estén completamente de frente.

Finalmente, el modelo de apariencia considera que la textura de la piel y de la forma del rostro permiten caracterizar de forma adecuada cada una de las imágenes, usando una descripción semántica del rostro para caracterizar los rasgos faciales. Hayashi, Yasumoto e Ito en [5], usa este modelo con un esquema de clasificación grupal. Sanhueza Riveros en [10], genera un modelo clasificador con local binary pattern, buscando los elementos más significativos del rostro como son boca y ojos, en este trabajo se genera un clasificador de 4 diferentes grupos (niños, adolescentes, adultos y ancianos).

En [11] Khaung Tin, indica que la estimación de edad es vista como un proceso de reconocimiento de patrones donde cada uno de los descriptores actúa como una clase. De esta manera la estimación de edad se aprecia como un problema de clasificación resuelta con el algoritmo del vecino más cercano, así mismo, se hace referencia al análisis de componentes principales, como técnica estadística, y se hace un acercamiento a los Eigenfaces.

Las redes neuronales las utiliza Hewahi en [12], para clasificar en 4 grupos principales (niño (1-12) , joven (13-25), adulto (26-45), viejo (46-63)) usando como conjuntos de entrenamiento las bases de datos de FG-NET la cual está compuesta por 1002 imágenes de diferentes sujetos de pruebas, con diferente iluminación y con un rango de edad de 0 a 62 años y empleando 4 capas de redes neuronales lograron obtener hasta un 78% de efectividad total con un porcentaje 87.6%, 67%, 86.4% y 70.9% respectivo a cada grupo principal.

Por otro lado, las edades son un conjunto de números con valores secuenciales, por lo tanto, la estimación de edad puede ser vista como un problema de regresión como lo explica Lanitis et al en [6], que propone un algoritmo genético usando funciones lineales, cuadráticas y cúbicas para representar el envejecimiento en las imágenes faciales.

En el tema de los modelos de apariencia se analizaron diversos trabajos que se relacionan entre sí siendo el punto de intersección la creación de los modelos con los métodos estadísticos ACP y ADL.

Terradez Gurrea indica en [13] que ACP es una técnica estadística de síntesis de la información o reducción del número de variables, es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número, perdiendo la menor cantidad de información posible.

Por otro lado, ADL maximiza la razón de dispersión entre clases a dispersión dentro de las clases. La idea es que las mismas clases deben agruparse firmemente cerca, mientras que las clases diferentes están tan lejos como es posible en la representación de baja dimensión, como lo indica Hernandez y Mendez en [14].

Los métodos ACP y ADL, son estadísticos que funcionan para sintetizar la información y para maximizar la diferenciación entre clases respectivamente. En el ámbito de la visión por computadora ACP y ADL ofrecen una técnica llamada Fisherfaces, método propuesto por Belhumeur, Hespanha y Kriegman en [15].

De esta manera el modelo de apariencia permite obtener las características de las imágenes faciales generando los clasificadores con un conjunto de imágenes de entrenamiento con una descripción previa.

En el artículo documental [16], Kyle Patterson presenta una propuesta para crear exhibidores interactivos para museos, los cuales mostrarán información variada para diferentes tipos de visitantes, en él se utilizan las Fisherfaces como método para obtener las características de las imágenes haciendo énfasis en la separación entre clases que genera este método.

Como resultado dio un porcentaje de efectividad del 72% en hombres y 62% en mujeres esto sin realizar algún preprocesamiento sobre las imágenes de entrenamiento o de prueba.

3. Bases de datos y preprocesamiento

3.1. IMDB-Wiki data set

Rothe, R. en [17] explica que IMBD-Wiki data set es una base de datos que extrae fotos de la página web IMDB y de Wikipedia, consiguiendo un total de 524,230 imágenes de rostros con información para obtener su edad. Se obtuvieron fotos de un rango de 0 a 100 años con variación de expresión, pose y luminosidad, ver figura 1.



Fig. 1. Base de datos IMDB-WIKI.

3.2. FG-Net

FG-net base de datos de envejecimiento disponible en [18], es una base de imágenes pública que podemos utilizar para los experimentos. La base de datos contiene 1002 rostros a color y escala de grises de alta resolución con alta variación de luminosidad, pose expresión y tamaño. Contiene en total 82 sujetos de múltiples razas con un rango de edad de 0 a 69 años explicado por Guo, Fu y Dyer en [19] y es ampliamente usada como en [20] por Guo, Mu, Fu y Huang y por Luu, Ricanek, Bui y Suen en [21]. Ejemplo de imagen en FG-Net, figura 2.



Fig. 2. Sujeto 13 Base de datos FG-Net.

3.3. Preprocesamiento

Dado que la base de datos IMDB-wiki, data set no indica la edad de la persona que aparece en la foto se tiene que realizar un procedimiento para calcular la edad, esto es restar el año de nacimiento al año en que fue tomada la fotografía. Después de obtener las edades de cada imagen de la base de datos, enseguida se procede con el algoritmo para obtener el conjunto de imágenes con el que se entrenará el clasificador.

El siguiente paso de la metodología permite obtener los rostros de la base de imágenes con la única condición que la imagen sea mayor a 400x400 píxeles y eliminar la mayor cantidad de características de las imágenes que no aportan información al modelo, esto con el fin de optimizar y liberar carga computacional. Para ello se aplica el filtro escala de grises, se ecualiza el histograma y se extrae el área de interés para la estimación de edad, el resultado es una lista de imágenes de 200x200 píxeles con rostros ecualizados.

En la figura 3, se puede observar el preprocesamiento aplicado paso a paso a un elemento de la base de imágenes FG-net.

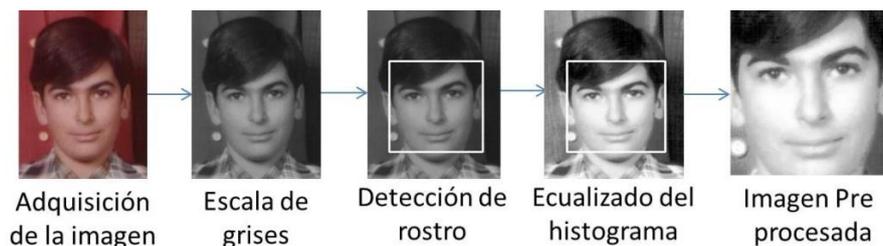


Fig. 3. Preprocesamiento de las imágenes de entrenamiento.

4. Propuesta

En este trabajo se propone el desarrollo de un clasificador basado en apariencia, de acuerdo en lo documentado por Suárez Ricardo en [2], Kyle Patterson en [16] y Sanhueza Riveros en [10] debido que este tipo de clasificadores obtienen buenos resultados sin importar que los rostros varíen en pose y luminosidad siendo este un problema recurrente cuando se emplean puntos fiduciales como lo demuestran Cox et al. en [4] o clasificadores basados en modelos antropométricos trabajados por Kwon y Lobo en [9] donde se tiene que preseleccionar las imágenes ocupando solamente las que contienen rostros completamente frontales.

En los trabajos de investigación el clasificador funciona mediante dos partes específicas, la creación de los modelos de representación de características y la clasificación de la nueva imagen. En la figura 4 se observa la propuesta del clasificador.

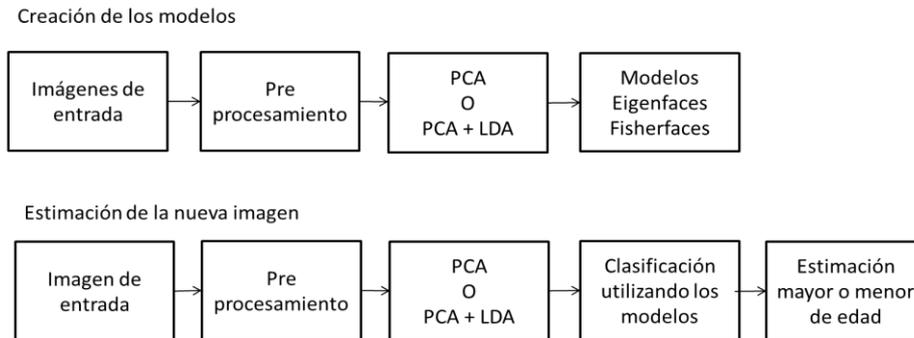


Fig. 4. Propuesta del clasificador.

5. Metodología

Como se indica la figura 4, el clasificador está conformado por dos módulos separados, el primero está orientado a la creación de los modelos que describan las características de los conjuntos de imágenes de cada clase haciendo énfasis en sus diferencias, para generar los modelos se utiliza el conjunto de imágenes preprocesadas resultado del procedimiento anterior y se asigna una cantidad similar de imágenes para cada clase después se aplica sobre ellas los métodos estadísticos: Análisis de Componentes Principales (Eigenfaces) usado por Turk et al. en [7] (figura 5) o la combinación de Análisis de Componentes Principales y Análisis Discriminante Lineal (Fisherfaces) introducido por Belhumeur et al. en [8] (figura 6).

La segunda parte corresponde a la clasificación de las nuevas imágenes, para poder clasificarlas se debe usar el mismo preprocesamiento, filtro de escala de grises, ecualizado del histograma y obtener el área de interés. Una vez que se obtiene la imagen preprocesada se extraen las características con el mismo método estadístico (ACP o la combinación ACP y ADL) con el que se haya entrenado el clasificador. Por último, se utiliza el clasificar K-nn para verificar con que clase se tiene un mayor número de características semejantes, como se desarrolla en la figura 7.

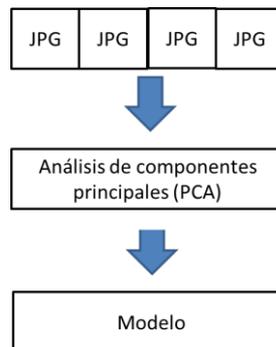


Fig. 5. Creación de modelos Eigenfaces.

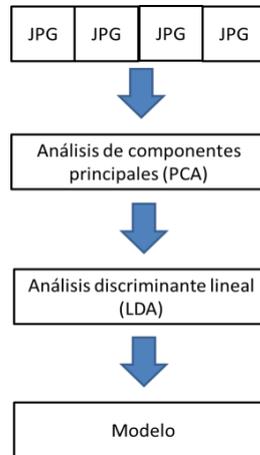


Fig. 6. Creación de modelos Fisherfaces.

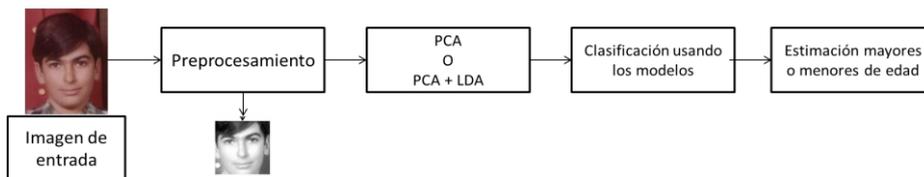


Fig. 7. Proceso de clasificación de una nueva imagen.

El modelo general del clasificador se compone de tres modelos independientes que son adaptados como se describe en la figura 8, en los siguientes pasos.

1.- Preprocesamiento de las imágenes; Permite cargar un conjunto de imágenes y mediante un proceso que incluye aplicación de filtros (Escala de grises y ecualizado del histograma) y detección y extracción de rostros mediante la técnica Haar-like features se obtiene una lista de imágenes faciales depurada y estandarizada.

2- Modelo de representación de características; Toma las imágenes preprocesadas y realiza un Análisis de Componentes Principales o una combinación de la misma técnica seguida de un Análisis Discriminante Lineal, este proceso genera los modelos de representación estadística de las imágenes haciendo énfasis en las diferencias entre clases.

3.- Clasificación de la nueva imagen; Una vez que se crearon y cargaron los modelos de representación, se comparan con una nueva imagen de prueba la cual es preprocesada con los mismos filtros y se extraen sus características usando los mismos filtros y el mismo modelo de representación que en las imágenes de entrenamiento con el fin de obtener la estimación aproximada del grupo etario al que pertenece el nuevo rostro.

La implementación de la metodología se realizó en una computadora personal acer aspire modelo E5-411 con un procesador Intel Celeron N2830, 4GB DDR3 Memoria Ram y 1000GB de Disco Duro.

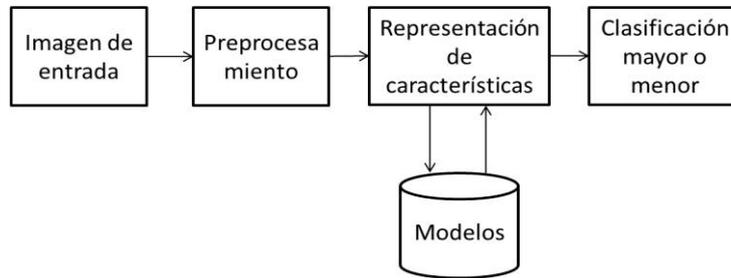


Fig. 8. Modelo general.

6. Resultados

La exactitud que tiene un clasificador se mide mediante el número de aciertos que realiza en función del número de pruebas que se llevan a cabo. Sin embargo, un sistema que acierte siempre puede asignar gran cantidad de falsos positivos, y por el contrario, un clasificador estricto que minimice los falsos positivos puede llegar a aumentar el número de falsos negativos. Es por ello que la evaluación de un clasificador deberá de considerar todos los aspectos anteriores, y proporcionar una medida que evalúe simultáneamente la capacidad de discernir a qué clase pertenece cada dato. Por lo tanto, se utilizan tres criterios basados en la matriz de confusión (Figura 9) que son precisión, recall y F-measure, Factor F (F - Measure), que es la media armónica entre Precisión (P) y Recall (R).

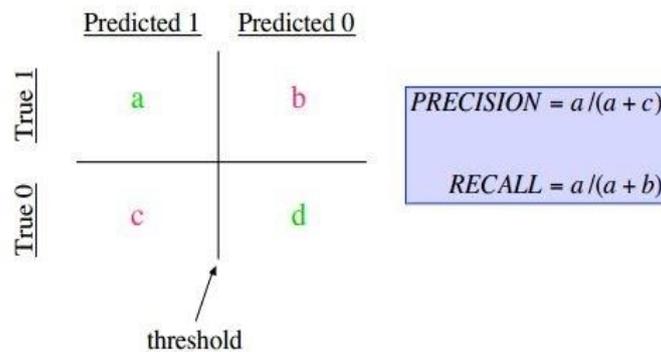


Fig. 9. Matriz de confusión (Performance Measures for Machine Learning [22]).

Para la primera parte del experimento se utilizaron las imágenes de entrenamiento de la base de datos FG-Net generando dos clases representativas, mayores (+18) y menores (-18) de edad de las cuales se obtuvieron los siguientes resultados, que se muestran en la tabla 1 y en la tabla 2.

Tabla 1. Resultados con la base de datos Fg-Net y Eigenfaces.

Imágenes	Efectividad	Precisión	Recall	F
10	40%	0.70	0.56	0.68
50	55%	0.86	0.56	0.68
90	62%	0.90	0.62	0.74
120	64%	0.91	0.63	0.75
150	78%	0.91	0.65	0.77

Tabla 2. Resultados con la base de datos Fg-Net y Fisherfaces.

Imágenes	Efectividad	Precisión	Recall	F
10	43%	0.79	0.46	0.58
50	60%	0.88	0.62	0.71
90	64%	0.91	0.63	0.75
120	70%	0.87	0.76	0.81
150	71%	0.89	0.73	0.81

Para la segunda parte del experimento se utilizó un total de 2000 imágenes seleccionadas de IMDB-Wiki data set y se aplicó sobre ellas el mismo preprocesamiento con el objetivo de construir los modelos de las mismas clases representativas del experimento anterior, mayores (+18) y menores (-18) de edad obteniendo los resultados mostrados en la tabla 3 y en la tabla 4.

Tabla 3. Resultados con la base de datos IMDB-Wiki y Eigenfaces.

Imágenes	Efectividad	Precisión	Recall	F
2000	77%	0.92	0.80	0.83

Tabla 4. Resultados con la base de datos IMDB-Wiki y Fisherfaces.

Imágenes	Efectividad	Precisión	Recall	F
2000	89%	0.97	0.89	0.86

En la tabla 5, se muestran los porcentajes en cuanto a efectividad en la clasificación de edad, resultado del presente trabajo junto con los obtenidos por otros autores que ocupan diferentes metodologías. En la figura 10, se muestra, un resultado positivo del clasificador con un menor y un mayor de edad.



Fig. 10. Resultado del clasificador (rojo menor de edad, azul mayor de edad).

Tabla 5. Comparación de mejores resultados en efectividad con trabajos relacionados.

Eigenfaces propio	Fisherfaces propio	Modelo antropométrico [9]	Redes Neuronales [12]	Eigenfaces [16]
77%	89%	92%	86%	72%

7. Discusión

La estimación de edad automática de edad más específicamente la clasificación de mayores y menores de edad es un problema que impacta en distintas áreas con múltiples aplicaciones potenciales, por ello se desarrolló un clasificador que puede realizar esta tarea. Según los resultados encontrados en esta investigación y en comparación con los previamente investigados en el estado del arte podemos concluir que nuestro clasificador obtiene buenos resultados con imágenes provenientes de entornos reales (cámara web, fotogramas de video y fotografías en entornos no controlados) a diferencia de los clasificadores basados en modelos antropométricos y puntos fiduciales los cuales obtienen mejores resultados con imágenes donde el rostro se encuentra completamente de frente, aunque no se puede clasificar imágenes con variación de pose, que está demostrado en el trabajo por Kwon et al. en [9], por Cox et al. en [4] y Hewahi et al. en [12].

También se observó que existen investigaciones de clasificadores basados en apariencia que obtienen mejores resultados debido a que realizan un preprocesamiento más exhaustivo como Kyle Patterson en [16], que realiza un etiquetado de las imágenes por grupo etario y sexo y Kwon Young en [3] el cual obtiene todos los elementos característicos del rostro lo cual representa un alto costo computacional. Otro punto a resaltar del presente trabajo es la demostración de la precisión del clasificador que aumenta a medida del número de imágenes de entrenamiento como lo afirman Santos et al. en [1] y Suárez Ricardo en [2]. Por último se recomienda el uso de la base de datos FG-net la cual es ampliamente usada en investigaciones como las realizadas por

Luu et al. en [21], Guo et al. en [20] y Hewahi et al. en [12], debido a que proporciona un conjunto de imágenes de entrenamiento y de prueba con su respectiva edad.

8. Conclusiones y trabajo a futuro

En este trabajo se desarrolló un clasificador basado en apariencia utilizando dos modelos diferentes de representación, Eigenfaces y Fisherfaces. Se demostró que las pruebas realizadas con cada modelo, el porcentaje de aciertos es proporcional al número de imágenes utilizadas para su creación, debido a que se extraen y clasifican más características en cada clase. Esto indica que si se tiene mayor número de imágenes se obtienen mejores resultados de estimación. Esto se demuestra al verificar que los mejores resultados se obtienen cuando se utilizaron 2000 imágenes para la creación de los modelos donde se obtuvo una efectividad del 89% en la diferenciación de mayoría o minoría de edad.

También se determinó que el clasificador desarrollado con Fisherfaces, presenta mejores resultados comparado al clasificador basado en Eigenfaces. Por último, se muestra que el número de imágenes incide directamente sobre el tamaño de los modelos estadísticos.

El trabajo a futuro corresponde a la recolección de nuevas bases de imágenes faciales que brinden la edad de los individuos. Así mismo el generar una nueva metodología basada en multiprocesamiento y procesamiento con GPU que permita trabajar con grandes volúmenes de imágenes sin que esto conlleve a un problema de rendimiento.

Referencias

1. Santos, S., Sanchez, A., Carmonet, B., Guerra, C., Santos, S.: Control de edad en redes sociales mediante biometría facial (2012)
2. Suárez-Fuentes, R. A.: Diferenciación visual de niños y adultos a partir de imágenes faciales (2013)
3. Kwon, Y.: Age classification from facial images in Computer Vision and Pattern Recognition. In: IEEE Computer Society Conference, pp. 762–767 (1994)
4. Cox, I., Ghosn, J., Yianilos, P.: Features based face recognition using mixture-distance in Computer Vision and Pattern Recognition. In: IEEE computer Society Conference, pp. 209–216 (1996)
5. Hayashi, J., Yasumoto, M., Ito, H., Koshimizu, H.: Method for estimating and modeling age and gender using facial image processing. In: Virtual Systems and Multimedia, IEEE Proceedings, Seventh International Conference, pp. 439–448 (2001)
6. Lanitis, A., Taylor, C., Cootes, T.: Toward automatic simulation of aging effects on face images. In: Transactions on Pattern Analysis and Machine Intelligence. 24(4), pp. 442–455 (2002)
7. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience (1991)

8. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs fisherfaces: Recognition using class specific linear projection. In: Bernard F. Boston and Roberto Cipolla. Springer, Vol. 1064, pp. 45–58 (1996)
9. Kwon, Y., Vitoria, L.: Age classification from facial images. *Computer Vision and Image Understanding*, 74(1), pp. 1–21 (1999)
10. Sanhueza, R.: Análisis y detección de características faciales usando aprendizaje estadístico. Tesis del Universidad de Chile, Facultad de ciencias Físicas y Matemáticas (2008)
11. Tin, H.: Gender and Age Estimation Based on Facial Images. *Act technical Napocensis*, 52(3), pp. 37 (2011)
12. Hewahi, N., Olwan, A., Tubeel, N., El-Asar, S., Abusultan, Z.: Age estimation based on neural networks using face features. *Journal of Emerging Trends in Computing and Information Sciences*, 1(2), pp. 61–67 (2010)
13. Terradez, G.: Análisis de componentes principales. Universidad Oberta de Cataluña, 11 (2002)
14. Hernández, J., Méndez, N.: Reconocimiento de expresiones faciales para interacción con el computador. *Jóvenes en la ciencia*, 2(1), pp. 62–66 (2016)
15. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *Transactions on pattern analysis and machine intelligence*, 19(7), pp. 711–720 (1997)
16. Patterson, K.: Automatic Age Estimation and Interactive Museum Exhibits. *Experimental Aging Research*, 35(2), pp. 268–275 (2014)
17. Rothe, R., Timofte, R., Van, G.: Dex: Deep proceeding. In: *Proceedings of the IEEE. International Conference on Computer Vision Workshops*, pp. 10–15 (2015)
18. FG-Net. Disponible en: <http://www.fgnet.rsunit.com>
19. Gou, G., Gu, Y., Dyer, C., Huang, T.: Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7), pp. 1178–1188 (2008)
20. Gou, G., Gu, Y., Huang, T.: Human age estimation using bio-inspired features. In: *Computer vision and Pattern Recognition. IEEE conference*, pp. 112–119 (2009)
21. Luu, K., Ricanek, K., Bui, T., Suen, C.: Age estimation using active appearance models and support vector machine regression. In: *Biometrics: Theory, Applications and Systems. IEEE 3rd International Conference*, pp. 1–6 (2009)
22. Performance Measures for Machine Learning. Disponible en: https://www.cs.cornell.edu/courses/cs578/2003fa/performance_measures.pdf

Estado del arte y elementos del reconocimiento automático de imágenes del cerebro

Keila Vasthi Cortés-Martínez, Manuel Mejía-Lavalle

Centro Nacional de Investigación y Desarrollo Tecnológico, Departamento de Ciencias Computacionales, Cuernavaca, Morelos, México

{keila.cortes, mlavalle}@cenidet.edu.mx

Resumen. El área de la Inteligencia Artificial ha incursionado en el ámbito médico con el propósito de apoyar en el reconocimiento automático de imágenes del cerebro. Una de las enfermedades que ha ido en aumento en la última década es la Enfermedad Cerebro Vascular, cuyo conjunto de enfermedades, afectan los vasos sanguíneos cerebrales. Dados los efectos que puede causar este tipo de enfermedad, es importante caracterizar las fuentes que generan imágenes médicas y los procedimientos empleados para mejorarlas e interpretarlas. El documento describe los avances reportados en la literatura en relación a las fuentes que obtienen las imágenes y las etapas seguidas para su reconocimiento. El trabajo considera que existe la oportunidad de mejorar o desarrollar más algoritmos avanzados, que ofrezca un nivel más alto en la detección de enfermedades en imágenes del cerebro.

Palabras clave: enfermedad cerebro vascular, reconocimiento de imágenes médicas, inteligencia artificial, técnicas de visión artificial.

State of the Art and Elements of Automatic Image Recognition of the Brain

Abstract. Artificial Intelligence area has penetrated the medical field with the purpose of supporting the automatic recognition of brain imaging. One of the diseases that has been increasing in the last decade are Brain Stroke, whose set of diseases, affect the cerebral blood vessels. Given that effects that can cause this type of disease, it is important to characterize the sources that generate medical images and the methods used to improve and interpret them. This paper describes the advances reported in the literature in relation to the sources that obtain the images and the stages followed for their recognition. Moreover considers that there is an opportunity to improve and development of advanced algorithms, which will offer a higher level in the detection of diseases in brain images.

Keywords: brain stroke, medical image recognition, artificial intelligence, artificial vision techniques.

1. Introducción

En el año 2001, la Secretaría de Salud de México [1], ubicó a las Enfermedades Cerebro-Vasculares (ECV), en el sexto lugar dentro de las diez principales causas de muerte en México. En el año 2014, la Organización Panamericana de la Salud [2], ubicó a las ECV en el segundo lugar, del Índice de enfermedades que incapacita a los adultos, lo cual indica que en el lapso de una década, la ECV se convirtió en un serio problema de salud.

Las ECV ocurren cuando hay afectaciones en los vasos sanguíneos del cerebro, provocando la disminución repentina de flujo sanguíneo o el rompimiento de ellos, haciendo que la sangre se esparza entre los espacios que rodean a las células cerebrales, dando lugar a dos tipos principales de términos: isquémico y hemorrágico [3]; se caracterizan por presentarse sin aviso y causan secuelas desde leves hasta severas: como parálisis, problemas de raciocinio, derrames, coordinación motora e incluso la muerte [4].

Las ECV pueden diagnosticarse mediante el uso de imágenes médicas obtenidas por medio de tomografías computarizadas, ultrasonidos (como el Doppler), electroencefalogramas, imágenes de resonancia magnética, etc. [3] El diagnóstico por interpretación de imágenes es invaluable, ya que se trata de un método no invasivo para la detección de la enfermedad. El reconocimiento de imágenes médicas de forma automatizada reduce el tiempo en la evaluación y minimiza errores en el diagnóstico y tratamiento de enfermedades [5].

Actualmente la inteligencia artificial aborda el tratamiento y análisis de imágenes médicas, desarrollando técnicas o modelos de aprendizaje automático y de reconocimiento de patrones, tales como la lógica difusa, red neuronal artificial (RNA) [6], k-means, entre otros [7], teniendo un gran éxito en la medicina como apoyo en la toma de decisiones para un diagnóstico más certero, representando uno de los avances más importantes de los últimos años, puesto que hoy en día se pueden examinar distintas actividades del cerebro al mismo tiempo (su función metabólica, eléctrica o cognitiva) [7].

Por la importancia del tema se hace necesario establecer una sinergia entre la disciplina médica y la disciplina de las ciencias de la computación; por lo tanto en este trabajo se presentan los estudios enfocados en el uso de herramientas computacionales para el tratamiento y diagnóstico de imágenes médicas. La literatura consultada corresponde a 79 artículos (Tabla 1).

Tabla 1. Literatura de países citados en el uso de imágenes médicas del cerebro (elaboración propia).

PAÍSES	TC	MRI	EEG	Doppler	En general	TOTAL
Alemania		[60]				1
Argentina				[14]		1
Bélgica		[52]				1
Canadá					[47] [67]	2

PAÍSES	TC	MRI	EEG	Doppler	En general	TOTAL
China		[70]	[31] [36]			3
Colombia			[28]			1
Cuba				[16]		1
España		[10]	[4] [3] [9] [59] [77]	[18]	[85]	8
Estados Unidos	[34] [81]	[11] [34] [43] [45]	[20] [22] [26] [29] [32] [71]		[23] [83]	14
India	[12] [13] [30] [40]	[25] [33] [38] [39] [62] [40] [48] [49]	[21]		[27][58] [74] [76]	17
Irán					[53] [64][65]	3
Iraq	[35]				[74]	2
Israel	[66]					1
Corea	[44]					1
Malasia	[41] [56] [80]		[24]		[78]	5
México	[55]		[8] [72] [82]	[17]		5
Nueva Zelanda			[69]			1
Polonia	[57] [75]					2
Reino Unido					[61]	1
Rumania					[6] [79]	2
Sri Lanka	[37]					1
Sudán	[46]					1
Suiza		[5]				1
Taiwán	[36]					1
Túnez					3	1
Turquía		[42]	[73]			2
TOTAL	18	18	21	4	18	79

2. Fuentes de imágenes médicas cerebrales

Para diagnosticar las ECV es necesario el uso de imágenes médicas. Una imagen médica es aquella que resulta del conjunto de técnicas y procesos usados para crear imágenes del cuerpo humano, o partes de él, con propósitos clínicos, es decir imágenes que buscan revelar, diagnosticar o examinar enfermedades, o bien, servir sólo para propósitos científicos médicos [9]. El uso de tecnologías generadoras de imágenes

cerebrales es uno de los avances más significativos en la medicina [9]. A continuación se describen las ventajas y desventajas de cada una de ellas.

2.1. Ultrasonido Doppler

Es una herramienta ultrasonográfica rápida, fácil de usar, de bajo costo, indolora y no utiliza radiación ionizante; ayuda a visualizar el funcionamiento de los vasos cerebrales (dirección y velocidad del flujo sanguíneo); detecta si existe sangrado intracraneal u obstrucción así como tejidos blandos [14,15,16]. El ultrasonido Doppler ha sido utilizado durante procedimientos quirúrgicos como un método auxiliar en la valoración del flujo vascular en tiempo real [17] (Fig. 1).

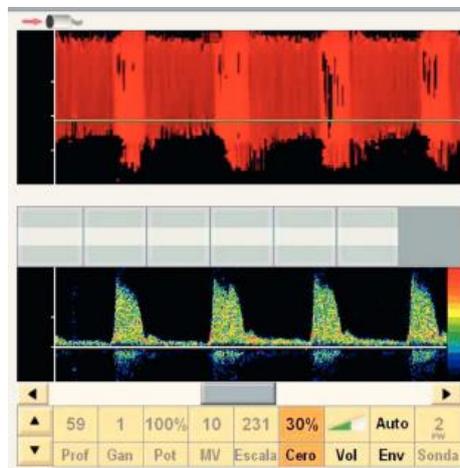


Fig. 1. Arteria cerebral media con alta presión intracraneal captada por ultrasonido Doppler [18].

2.2. Imágenes de resonancia magnética (MRI)

Las imágenes de resonancia magnética (MRI) contribuyen a la detección, diagnóstico y tratamiento de diversas patologías y enfermedades; una de sus ventajas es que tienen gran capacidad de distinguir entre materia gris y blanca del cerebro (Fig. 2). La MRI se realiza dentro de un campo magnético (sin radiación), y junto con la radio frecuencia altera el alineamiento natural de los átomos de hidrógeno en el cuerpo [10], por lo que resulta no invasiva para el paciente. Durante la obtención de las MRI, existe ruido y variación paramétrica que presentan problemas especiales en los procesos de automatización, por lo que es importante evaluar cada imagen y calcular cada uno de sus píxeles con el fin de suavizar los datos [11].

2.3. Tomografía computarizada

La tomografía computarizada (TC) crea una serie de imágenes transversales del cerebro, que precisan la detección de anomalías pequeñas; permite encontrar tumores

y otras enfermedades con mayor facilidad, reduciendo tiempo y costo en el diagnóstico [12]. Se realiza mediante la aplicación de rayos-X, obteniendo múltiples imágenes detalladas en distintos ángulos, lo cual permite diversas vistas del cerebro (Fig. 2).

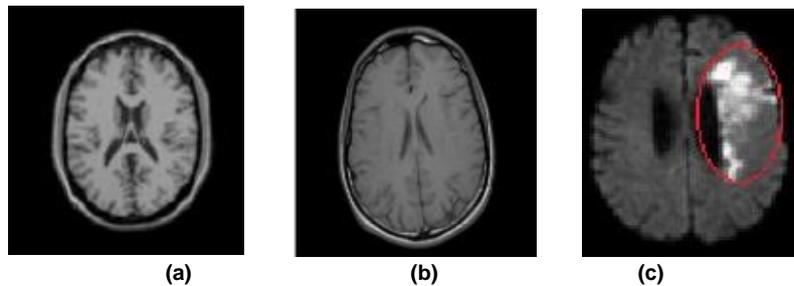


Fig. 2. Imágenes cerebrales: (a) MRI normal, (b) TC normal, (c) TC anormal [13].

2.4. Electroencefalogramas

El electroencefalograma (EEG) representa las actividades eléctricas del cerebro llamadas impulsos durante un corto período de tiempo, las cuáles se registran a partir de electrodos colocados en el cuero cabelludo obteniendo una representación de líneas en múltiples canales a lo largo del tiempo [19, 20]. Esta técnica de diagnóstico requiere de profesionales capacitados para su interpretación lo que resulta tediosa y lenta, no obstante es una importante herramienta no invasiva que diagnostica, evalúa, monitorea y predice una ECV [21, 22] (Fig. 3); cuando percibe una ECV del tipo isquémico, el EEG es muy sensible ya que se observan diversos cambios anormales en los patrones debido al bajo nivel de oxígeno que presenta el cerebro [23].

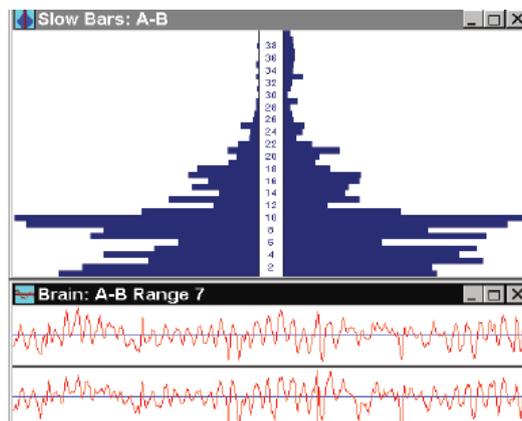


Fig. 3. Señal de ondas cerebrales con EGG [24].

Las imágenes obtenidas mediante herramientas computacionales pasan por varias etapas para llegar a la interpretación y la obtención de un diagnóstico de una enfermedad cerebro vascular (Fig. 4).

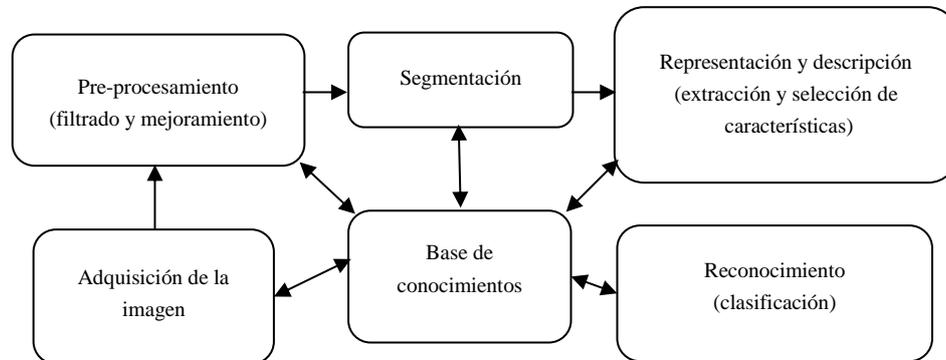


Fig. 4. Etapas comunes (no aplicables en algunos casos) en la interpretación de imágenes médicas [8].

3. Etapas en la interpretación de imágenes médicas

En esta sección se describen brevemente algunas técnicas o métodos para el pre-procesamiento, segmentación, descripción y reconocimiento de características de imágenes médicas cerebrales, presentados en artículos científicos orientados a la neurociencia con base en la inteligencia artificial.

3.1. Pre-procesamiento de los datos

En la literatura se identificaron 31 referencias; en estas se indica que el pre-procesamiento de una imagen consiste en aplicar técnicas para la búsqueda de datos a partir de una imagen original, mejorando sus características en otra imagen, con la finalidad de ser analizadas y que permitan realizar operaciones a partir de información hallada; aplicando esto a imágenes médicas, su papel es crucial para el tratamiento y diagnóstico de enfermedades. Las imágenes originales contienen “detalles ajenos” (o artefactos) cuando presentan algún tipo de ruido o moteado, por lo que el resultado del diagnóstico del médico al paciente puede errar. Uno de los objetivos del pre-procesamiento es el de normalizar o estandarizar los datos de una imagen previa. En algunos casos [11] no es necesaria esta etapa; en otros, por ejemplo [25], los artefactos presentes en las imágenes médicas suelen eliminarse manualmente, algunos de estos son derivados de los aparatos con los que se toman las muestras [21,26]. Muchos autores coinciden en que una de las técnicas más importantes en esta etapa es el filtrado de imágenes [27], que ayuda en el mejoramiento de una imagen con el fin de adquirir información; dentro de la literatura citada se dividen en 2 grupos:

1) De acuerdo a su dominio de frecuencia: Este tipo de filtrado es usado comúnmente en los EEG, por ejemplo usando la transformada de Fourier [21,24,28,29], se elimina la presencia de ruido ocasionados por parpadeos oculares, actividad muscular, etc.[30]. Para artefactos ocasionados por los aparatos empleados, se usa el filtro paso de banda [31], para estandarizar datos se utiliza el filtrado pasa-bajo [32].

2) De acuerdo a su dominio espacial: Se identifican frecuentemente en el pre-procesamiento de MRI y TC, los filtros destacados son: medio [27], Laplaciano [27], máximo [33], Gaussiano [13,27,36] y el filtrado mediano, los cuáles ayudan en la eliminación de ruido sin afectar la nitidez de la imagen médica [25,27,33,34,35],

En diversos artículos, se mencionan otras técnicas recurrentes usadas en esta etapa como: los métodos de binarización [13,35,37,38,39,40], eliminación de ruido [40], ecualización(o mejoramiento del contraste) [27,38,39,40,46], curva cúbica aplicada al contraste [36], corrección de ángulo inclinado [13], detección simétrica del eje [41], redimensión(como normalización en el tamaño de las imágenes) [13,37,42,43,44], intensidad de la imagen [40,45] y transformada wavelet [27]. Alhawaimil y otros investigadores [35,45, 48] describen un método de pre-procesamiento de cuatro pasos para tratar la naturaleza patológica del tejido cerebral en imágenes de TC:

1) Se analiza la imagen RGB (modelo de color) de una tomografía computarizada y se convierte a una imagen en escalas de grises [10,42,45];

2) Se aísla la sección del tejido cerebral escaneado, mediante el cálculo del área de la materia blanca del cráneo [35,40], por medio de algún software [43] o mediante la técnica de crecimiento de regiones [35,47]. Aquí también se puede hacer una detección de bordes usando Canny [48].

3) Se enmascara la imagen mediante un generador lógico OR para eliminar ruido con un filtro mediano de ventana [35].

4) Por último se utilizan operaciones morfológicas de erosión, dilatación y expansión para "limpiar" la imagen eliminando imperfecciones de la parte que corresponde al cráneo [13,35,36,39,42]; o mediante la técnica de Otsu [46].

En el caso de los electroencefalogramas, al pre-procesamiento de las señales se agregan dos métodos adicionales: la técnica Welch [24] y la ventana de Haming. Dentro de esta etapa las redes neuronales han tenido buenos resultados durante esta etapa, las más comunes son: las redes de función de base radial (RBF) [28], de retropropagación (FeedFoward) [42,49] y las multi-layer perceptron (MLP) [28,32,50].

3.2. Segmentación de los datos

En la literatura consultada se identificaron 26 referencias. Después de "limpiar" o pre-procesar las imágenes se prosigue a hacer una segmentación, que es el proceso mediante el cual se divide la imagen en regiones con propiedades similares, tales como color, textura, brillo, contraste, niveles de grises; dentro de las imágenes médicas, la segmentación se define como la partición de una imagen en regiones constituyentes sin traslape, las cuáles son homogéneas con respecto a alguna característica como una intensidad o una textura [51,52,53]. Su objetivo radica en el estudio estructural e identificación de regiones de interés (tumores, lesiones y otras anomalías), siendo esto una tarea difícil, ya que este tipo de imágenes son de naturaleza compleja, ocasionando que seleccionar un algoritmo de segmentación dependa de diferentes aspectos a considerar ya que cada sistema de imágenes médicas tiene sus propias limitaciones; sin embargo, diversos métodos de segmentación se pueden aplicar a una amplia gama de

datos [12,52,54]. Por ejemplo, Diyana et al. [56] segmenta el área intracraneal con técnicas ROI en las imágenes de tomografía computarizada usando fuzzy C-Means y el método de umbral de Otsu para definir importantes estructuras cerebrales y otras regiones que pueden contener anomalías; para Alhawaimil [34], los mejores métodos de segmentación son: dilatación y erosión; Sangeetha y Shahin [58] alimenta una RNA con el algoritmo K-NN; Yeu-Sheng et al. [36] aplica un algoritmo de crecimiento de regiones con filtro Gaussiano y la técnica Canny para obtener un mapa de bordes de las áreas importantes dentro de una tomografía computarizada original, separando la materia gris y resaltando el área de una ECV. A manera introductoria, investigadores [12,40,54], describen que los métodos de segmentación se dividen en cinco grandes grupos (Tabla 2).

En una revisión de Ramírez y Chacón [8] se divide a las técnicas de segmentación de la siguiente manera, aunque se enfocan especialmente en las RNAs:

Tabla 2. Métodos de segmentación.

Basados en:	Descripción
Umbrales	-Umbralización (thresholding) [12,34,35,39,40,55] - Textura y por niveles de grises (subdivisión propuesta por Aggarwal et al.[54]) - Manuales, semi-automatizadas o automatizadas(sub-categoría propuesta en [40]) - Método de umbralización óptimo para región de interés (ROI) [56,57] -Otsu [56]
Regiones [3,14,48]	-Crecimiento de regiones: [12,36,40,55] -Crecimiento de cuencas (Watershed): [37,40]
Bordes[48]	-Canny: [12,36,55] -RNA Convolutacional (CNN), Red Neuronal Pulso-Acoplada (PCNN), Mapa Auto-organizado de Kohonen (SOM), Redes Neuronales Oscilatorias (ONN) [8]
Pixeles	-Fuzzy C-Means: [6,11,40] -Campo Aleatorio de Markov: [40] -RNA: [12,34,40] -RNA Multi-Layer Perceptron: [33]
Agrupación	-Modelos Paramétricos Deformables: [40] -Modelos Geométricos Deformables: [40] -Clustering: [12]

- De bordes y formas: Se crea una imagen binarizada, extrayendo las regiones más representativas a partir de contornos de objetos en la imagen, la RNA más versátil fue la red neuronal convolutacional (CNN) usada en imágenes con formas irregulares [8].
- Segmentación de Color y textura: Se hace una partición de la imagen en regiones que agrupan colores similares e intensidades de grises. Las variantes de las RNAs con mejores resultados son: mapa auto-organizado de características (SOFM), variante de la SOM paramétrica (PSOM), variación de SOM con un algoritmo genético (SOFT-NN), SOM de estructura adaptiva (SASOM), así como algoritmos de mezclas gaussianas (GMM) y oscilaciones sincronizadas con CNN (SON) [53].

- Movimiento: A partir de una secuencia de imágenes se detectan objetos, según la literatura analizada los más eficientes son: self organization background subtraction (SOBS), fuzzy clustering neural network (FCNN), CNN de tiempo discreto (DTCNN) y time delay neural network (TDNN). Anusha et al. y Ceballos [21,28] utilizan el método de “ventana deslizante” (sliding window segmentation, en inglés) dentro de encefalogramas, para dividir las señales producidas en épocas y segmentarlas en canales [31,59].

Otras técnicas de esta etapa son: operaciones morfológicas (de erosión y dilatación) [34,37,38], binarización [60], filtro mediano [35,36], euclidiana [62], inferencia bayesiana [61], así como las RNAs [53] descritas como técnicas rápidas y precisas en la segmentación de una imagen cerebral [11]; Cárdenas [63] utiliza redes artificiales pulso-acopladas (PCNN, por sus siglas en inglés) con metodologías clásicas, como entropía cruzada y la métrica de correspondencia de píxeles (PCM, por sus siglas en inglés) para detección de bordes. Shanbehzadeh y Rastgarpour [64,65] dividen las técnicas de segmentación de acuerdo a su aplicación en cuatro grupos:

1) Métodos Híbridos. Son técnicas de Inteligencia Artificial combinadas con algoritmos de aprendizaje automático y operadores de procesamiento de imágenes, lo que hace que sean semi-automáticos.

2) Sistemas expertos. Aquí el conocimiento de los expertos se utiliza en forma de reglas, modelos o atlas [56].

3) Métodos de procesamiento de imágenes. Se utilizan características múltiples de diferentes imágenes para aumentar la precisión en cada una de las categorías a analizar de una imagen [66].

4) Métodos Multiespectrales o multimodales. Se utilizan múltiples imágenes de un órgano para tener características de variantes en la imagen, y realizar una fusión para la obtención de plantillas [67.]

La comparación de la literatura indica que la segmentación basada en RNAs y por umbralización son los métodos más frecuentes debido al aislamiento de píxeles que realizan; partiendo de esto, Gordillo et al. [68] realiza un resumen de las ventajas y desventajas que poseen estos métodos en la aplicación de imágenes de resonancia magnética con presencia de tumores cerebrales.

3.3. Representación y descripción (extracción y selección de características)

De un total de 79 artículos, 26 referencias contienen esta etapa y se mencionan 45 técnicas. Para representar y describir de manera adecuada una imagen es necesario representar la naturaleza de la misma en forma numérica, para ello se extraen y se seleccionan atributos o rasgos más relevantes y con menor redundancia, que ayuden a facilitar la visualización y comprensión de los datos. Por ejemplo, el diseño utilizado para la alimentación de árboles de decisión en [46] se formó de un conjunto de 14 atributos obtenidos de las características sugeridas por especialistas, formando un conjunto de referencia que ayudan en la descripción de ECV isquémica; algo parecido se plantea en [27], donde usando el algoritmo K-NN y clasificadores bayesianos, para extraer características de una imagen médica se toman decisiones con respecto a la

patología de una estructura o tejido ateromatosa. Para el diagnóstico de la enfermedad de Alzheimer [47], se hace un etiquetado de voxels a partir la segmentación hipo campal dimensional o mediante una red neuronal pulsante (SNN, por sus siglas en inglés) [69]. Para representar la actividad normal y anormal del cerebro, se usan diversas técnicas como: características binarias, análisis de componentes principales, análisis discriminante lineal, código de cadena, transformación de características invariante a escala, filtrado Gabor, distancia euclidiana, red neuronal artificial, máquinas de soporte vectorial [13], matriz de co-ocurrencia de niveles de grises [39,40], descriptores de textura Haralick [44], umbralización (cambio en los niveles de intensidad); Stier et al. [43] utiliza una red neuronal convolucional (CNN) para generar un modelo predictivo. Tyan et al. [36], Balasooriya y Perera [37] usan los métodos de Otsu, filtro anisotrópico, máscara gaussiana, métodos morfológicos (erosión y dilatación) para determinar el área cerebral izquierda o derecha, en la que sucedió la ECV y eliminar ruido de fondo; Haiyan Zhang [70] usa un modelo de contorno geométrico activo mejorado. En imágenes de electroencefalogramas, se captura la duración y el tamaño de la amplitud de las ondas mediante la técnica de ventana deslizante [28,32], la descomposición wavelet[44], el dominio de frecuencias características, transformación de wavelet discreta, la longitud de la transformada de Fourier rápida [21,24,59,71,72,73], y la transformada de ondulada discreta de superposición máxima (MODWT) [72]; Ceballos [28] con una RNA multi-layer perceptron crea una matriz de características y patrones del electroencefalograma. Otras técnicas mencionadas en general, para realzar el contraste de la imagen en búsqueda de tumores son: técnicas de media, valores de histograma [44], operaciones morfológicas (dilatación y erosión) [11,25,38,42,62]; curtosis [28,33], matriz de co-ocurrencia de niveles de gris[74,75], desviación estándar, asimetría [57], entropía de Shanon [32], algoritmo Gabor [44], el algoritmo k-means [33]; máscara binaria basada en regiones(max-min)[38], así como diversas técnicas de minería de datos [44]. Vishal et al. [76] menciona a la red neuronal convolucional (CNN) como la mejor técnica para extraer características, comparada con máquinas de soporte vectorial y árboles aleatorios; Boubilil et al. [66] extrae características de diferentes versiones de una misma imagen de tomografía computarizada mediante una RNA, para después fusionarlas y mejorar su resolución; algunos métodos de fusión se mencionan en [67]. En esta sección generalmente se presentan algunos problemas al escoger características importantes de la imagen, debido a que estas deben de presentar el menor número de invariancia respecto a la imagen original.

3.4. Interpretación y reconocimiento

De un total de 79 artículos, 15 tienen referencia de esta etapa, mencionándose 11 técnicas y 5 algoritmos. La interpretación y reconocimiento de las características de una imagen médica sirven para dar un significado a los datos obtenidos de la imagen. En esta etapa se utilizan algunos algoritmos de agrupamiento y métodos de clasificación, los cuales unen espacios característicos de un conjunto de datos, mediante intensidades de la imagen. En ocasiones, esta etapa se realiza de manera “manual”, con la ayuda de especialistas [3,77], que clasifican y diagnostican una ECV mediante un algoritmo de diagnóstico etológico del ictus cerebral, o [35] a partir de una comprobación visual de

la región con lesiones. Lashari y Ibrahim[78] relatan que la clasificación de imágenes médicas se centra principalmente en el uso de algoritmos eficientes de minería de datos y técnicas de visualización, en este caso, se propone un algoritmo de clasificación basado en las nociones de teoría de conjuntos blandos, y se utilizan técnicas como la wavelet y Gabor, entre otras. Para tener una mejor interpretación de una imagen médica, se suelen afinar detalles obtenidos en etapas anteriores, por ejemplo Lee et al. [44] realizan un método para suavizar imágenes de tomografía computarizada mediante características obtenidas de una distancia ponderada gaussiana, reconstruyendo el cerebro humano en 3D y encontrando la ECV. Hussein Ali et al. [35], obtuvieron buenos resultados en el reconocimiento de cerebros enfermos, haciendo que los valores de gris de la parte anormal brillaran más. Karthikeyan y Dr.Ezhilarasi [25] interpretan lesiones mediante el crecimiento de regiones. RosemehahWan et al. [24] predice la actividad normal y anormal del cerebro en un encefalograma con K-NN (k-nearest neighbors); por medio de un análisis usando RNAs y máquinas de soporte vectorial, Menaka y Kanchana [13] desarrollan una base de datos con lesiones cerebrales; Yahiya et al. [46] emplea árboles de decisión y algoritmos de aprendizaje (k-means) para presentar un modelo de clasificación de ECV isquémica. En [61,78,79] se realizan algunas métricas de similitud con el fin de reconocer y clasificar anomalías dentro de las tomografías computarizadas, así como detección de anomalías basadas en reglas, en [80] se hace una colorización en base a un mapa de colores estándar. Chawla et al. [81] usa una metodología de tres pasos: primero mejora la imagen, después hace una detección de la simetría cerebral y al último hay una clasificación de cortes anormales, similar a la realizada por Bhaiya y Verma [49] donde se utilizan técnicas como la transformación wavelet, el análisis de componentes principales y métodos de aprendizaje supervisado como el algoritmo de retropropagación, red neuronal de funciones radiales y un vector de aprendizaje de cuantización. Toprak [73], Menaka y Kanchana [13] usan al final de la clasificación RNAs con algoritmos basados en en Levenberg-Marquardt, BFGS Quasi-Newton (Broyden-Fletcher-Goldfarb-Shanno with Quasi-Newton, en inglés) y máquinas de soporte vectorial. Otras técnicas descritas en esta etapa son: la técnica de características binarizadas, el algoritmo de código de cadena y la técnica de transformada de características invariantes, usando una distancia euclidiana [13]. Así mismo, las RNAs [55] se mencionan en distintos artículos de la literatura consultada, principalmente:

- Backpropagation: Útil en reconocimiento de tumores cerebrales o anomalías pequeñas [42], considerada como una gran clasificadora [28].
- FeedForward: [59,72].
- RNA sin especificar: [37,55] red neuronal con método de cuenca [21,26] como modelo de predicción en la detección de una ECV recurrente.
- Multi-Layer Perceptron (MLP): útil al clasificar los tejidos a dos clases de normales y anormales [39].

En la revisión realizada por Ramírez y Chacón [8] las técnicas de clasificación usando RNA más relevantes son: teoría de resonancia adaptiva (ART), multi-layer perceptron, red neuronal con funciones de base radial y red neuronal probabilística. De acuerdo a Ramírez-Quintana et al. [82], la RNA de teoría de resonancia adaptativa tuvo

muchos resultados favorables en el reconocimiento al igual que la red neuronal convolucional(CNN)[83], red neuronal probabilística y mapa auto-organizado de Kohonen (SOM, por sus siglas en inglés), debido a su flexibilidad en el procesamiento de color, análisis de movimiento y detección de formas.; la red neuronal con funciones de base radial fue la más exitosa en la segmentación de MRI; así mismo preveen que el uso de éstas RNAs sean una tendencia, especialmente en las SOM, CNN y Pulsantes; por lo que experimentar con redes neuronales pulsantes, augura buenos resultados (en especial en la clasificación). Actualmente existen algunas herramientas de software de código abierto utilizadas en la interpretación de imágenes médicas en general, en [84] se realiza una comparación de distintas aplicaciones, de las cuáles sólo SPM, FSL y DTI-TK ayudan en imágenes relacionadas con las ECV, con una tendencia por las que analizan Imágenes de Resonancia Magnética, como se describe en la Tabla 3.

Tabla 3. Fragmento de [84], sobre herramientas de software utilizadas en la interpretación de imágenes médicas.

Herramientas de Procesamiento de Imágenes		SPM ¹	FSL ²	ELASTIX ³	MITK ⁴	DTI-TK ⁵
Última Versión		12	5.0	4.7	3	3
Publicados		2014	2014	2014	2014	2011
Imágenes Sopotadas	MRI	X	X	X	X	X
	TC	X		X	X	
	EEG	X		X	X	X
	Ultrasonido imágenes de 3D			X	X	X
Funciones	Segmentación	X		X	X	
	Visualización			X	X	
	Reconstrucción	X	X	X		X
Lenguajes del Sistema	C	X			X	
	C++			X		
	C#		X			X
	Java					X
	Código Abierto	X	X	X	X	X
Plataformas	Windows	X	X	X	X	X
	Linux	X	X	X	X	X
	Mac OS X			X	X	X

¹Statistical Parametric Mapping; ²FMRIB Software Library; ³Elastix es un software de código abierto, basado en el conocido Insight Segmentation and Registration Toolkit (ITK); ⁴MITK Medical Imaging Interaction Toolkit; ⁵DTI-TK Conjunto de herramientas de normalización espacial que utiliza datos de neuro-imágenes.

4. Discusión

En la literatura revisada se identificó que las imágenes de resonancia magnética son herramientas eficientes para detectar de manera temprana una ECV isquémica, al

basarse en campos magnéticos logran distinguir mejor la materia blanca de la gris y puede predecir una mayor variedad de ECV.

La tomografía computarizada es una técnica relativamente rápida en la creación de la imagen cerebral y tiene una buena precisión en la detección de ECV hemorrágicas; sin embargo, presenta una exposición a la radiación dado que utiliza rayos-X. El electroencefalograma es una fuente lenta en la obtención de imágenes cerebrales, sin embargo es utilizada principalmente para rectificar la presencia de alguna ECV, también se utiliza para detectar Alzheimer y la epilepsia.

El ultrasonido Doppler es una técnica que evalúa la dirección como la velocidad del flujo sanguíneo generando datos espectrales, en este estudio se encontró poca literatura, sin embargo debido a su naturaleza algunos autores han especulado que se pueden estudiar mediante redes neuronales. En la literatura consultada se encontró una mayor cantidad de estudios cuyo origen es el país de la India, seguido de Estados Unidos y España (Tabla 1).

También se muestra que el electroencefalograma se utiliza con mayor frecuencia, este dato debe ser interpretado con cautela, puesto que según lo descrito depende de las ECV que se desean buscar; en el estudio actual, las tomografías computarizadas y las imágenes de resonancia magnética obtuvieron la misma relevancia en cuanto a su uso, sin embargo las imágenes de resonancia magnética son menos invasivas. Las técnicas más utilizadas por etapa fueron:

- Pre-procesamiento: el Filtrado, las Operaciones morfológicas y la Binarización.
- Segmentación: las técnicas basadas en píxeles (Umbralización) y RNAs.
- Representación y Descripción: hubo variedad en la extracción de características dependiendo del tipo de imagen a analizar, así como de la selección de parámetros a estudiar, sobresaliendo las RNAs y filtros.
- Interpretación: las RNAs fueron las más utilizadas y con mejores resultados en comparación a otras técnicas.

5. Conclusiones

La literatura indica que con cada herramienta computacional y con cada enfermedad se definen métricas o parámetros, descriptores, técnicas, algoritmos y herramientas de software diferenciados para el tratamiento de imágenes. En la literatura consultada se describen 120 técnicas, 5 algoritmos y 5 herramientas de software.

Las técnicas de Umbralización, Filtrado y Redes Neuronales Artificiales aparecieron en todas las etapas. En los 79 estudios se identificaron 17 tipos de RNAs pero pocos estudios sobre redes pulsantes. Consideramos que este tipo de red puede tener una buena efectividad en la extracción y selección de características de imágenes del cerebro.

Agradecimientos. Se agradece al Tecnológico Nacional de México por el financiamiento al proyecto 5683.16-P.

Referencias

1. Secretaría de Salud: Programa de Acción: Enfermedades Cardiovasculares e Hipertensión Arterial. Secretaría de Prevención y protección de la Salud, México, D.F. (2001)
2. MÉXICO: Perfil de Enfermedades Cardiovasculares. Organización Panamericana de la Salud. Revista de la Facultad de Medicina de la UNAM, pp. 11 (2014)
3. Blanco-González, M., Arias-Rivas, S., Castillo-Sánchez, J.: Diagnóstico del accidente cerebrovascular isquémico. Unidad de Ictus, Área de Neurociencias, Servicio de Neurología, Hospital Clínico Universitario, Santiago de Compostela, España, pp. 4919–4923 (2011)
4. Monge-Pereira, E., Molina-Rueda, F.: Electroencefalografía como método de evaluación tras un ictus. Una revisión actualizada. Neurología (2014)
5. Rihardi, J.: Machine Learning with Brain Graphs. IEEE Signal Processing Magazine, Vol. 30, pp. 58–70 (2013)
6. Rotariu, C., Costin, H.: Medical Image Processing by means of some Artificial Intelligence methods. Buletinul Institutului Politehnic Din Iasi (2013)
7. Fundación Española para la Ciencia y la Tecnología(FECYT): Viaje al Universo Neuronal. Unidad Didáctica (2007)
8. García, I.: Aportaciones a la Segmentación y Caracterización de Imágenes Médicas 3D. Capítulo II. Imágenes Médicas, Madrid (2010)
9. López, A.: Desarrollo de un sistema de análisis automático de perfusión en RMI de cerebro. Departamento de Tecnología Electrónica y de las Comunicaciones, Madrid, España, (UAM) (2015)
10. Pukish, M. S., Wang, S., Wilamowski, B. M.: Segmentation of cerebral cortex MRI images with artificial neural network (ANN) training. Biomedical Engineering (2013)
11. Menze, B. H., Koen, V. L., Lashkari, D.: A Generative Probabilistic Model and Discriminative Extensions for Brain Lesion Segmentation With Application to Tumor and Stroke. IEEE Transactions On Medical Imaging, Vol. 35 (2016)
12. Fernández, J., Martínez-Sánchez, P., García-Rodríguez, R.: El eco-doppler color transcraneal en el estudio. Argentina, Neurología Argentina (2012)
13. Despotović, I., Goossens, B., Philips, W.: MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, pp. 23 (2015)
14. Lai, M.: Deep Learning for Medical Image Segmentation (2015)
15. Pappachen, A., Dasarathy, B.: A Review of Feature and Data Fusion with Medical Images, Canadá (2014)
16. Haiyan, Z., Jiafeng, L., Zixin, Z., Haiyun, L.: An automated and simple method for brain MR image extraction, BioMedical Engineering OnLine (2011)
17. Wei-Long, Z., Bao-Liang, Lu.: Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks2015, IEEE Transactions On Autonomous Mental Development, Vol. 7 (2015)
18. Yeu-Sheng, T., Ming-Chi, W., Chiun-Li, C., Yu-Liang, K., Ming-Sian, L., Hao-Yan, C.: Ischemic Stroke Detection System with a Computer-Aided Diagnostic Ability Using an Unsupervised Feature Perception Enhancement Method. Taiwan (2014)
19. Ceballos-Herrera, A.: Detección de epilepsia en señales de electroencefalografía, mediante el uso de RNA y MSV. Facultad De Ingeniería, Programa De Ingeniería Biomédica, Universidad Autónoma De Occidente, Santiago de Cali (2010)

20. Scherle, C. E., Pérez-Nellar, J.: Utilidad del ultrasonido Doppler transcraneal en Neurología. Matamoros, La Habana, Cuba, Hospital Clínicoquirúrgico Hermanos Ameijeiras, Unidad de Ictus (2009)
21. Martínez-Vila, E., Fernández, M., Pagola, I., Irimia, P.: Enfermedades Cerebrovasculares (2011)
22. Ling, G., Rivero, D., Seoane, J. A., Pazos, A.: Classification of EEG Signals Using Relative Wavelet Energy and Artificial Neural Networks. GEC, pp. 177–183 (2009)
23. Escudero, D.: Doppler transcraneal en el diagnóstico de la muerte encefálica. ¿Es útil retrasar el diagnóstico? Medicina Intensiva (2015)
24. Peek, N., Combi, C., Marin, R., Bellazzi, R.: Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes (2015)
25. Alhawaimil, A.: Segmentation of Brain Stroke Image. International Journal of Advanced Research in Computer and Communication Engineering, Issue 9, Vol. 4, pp. 375–378 (2015)
26. Chawla, M., Sharma, S., Sivaswamy, J.: A method for automatic detection and classification of stroke from brain CT images. Engineering in Medicine and Biology Society, pp. 3581–3584 (2009)
27. Stier, N., Vincent, N., Liebeskind, D., Scalzo, F.: Deep Learning of Tissue Fate Features in Acute Ischemic Stroke. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BTBM), California, pp. 1316–1321 (2015)
28. Younis, A., Ibrahim, M., Kabuka, M., John, N.: An Artificial Immune-Activated Neural Network Applied to Brain 3D MRI Segmentation. Journal of Digital Imaging (2008)
29. Carey, H. J.: EEG Interictal Spike Detection Using Artificial Neural Networks. III Department of Computer Science, Virginia Commonwealth University. Richmond, Virginia (2016)
30. Faught, E.: Current Role of Electroencephalography in Cerebral Ischemia. Dallas, TX, USA, Comments, Opinions, and Reviews, Vol. 24, pp. 609–613 (2016)
31. Zhenzhong, W., Lilin, G., Adjouadi, M.: Wavelet decomposition and phase encoding of temporal signals using spiking neurons. Neurocomputing, 173, pp. 1203–1210 (2016)
32. Abend, N. S., Dlugos, D. J., Clancy, R. R.: A review of long-term EEG monitoring in critically ill children with hypoxic-ischemic encephalopathy, congenital heart disease, ECMO, and stroke. 2, Pennsylvania, USA, J Clin Neurophysiol, Vol. 30, pp. 134–142 (2014)
33. Amarasinghe, K., Wijayasekara, D., Manic, M.: EEG Based Brain Activity Monitoring using Artificial Neural Networks. Idaho Falls, Idaho, USA, pp. 61–66 (2013)
34. Benayoun, M., Kohrman, M., Cowan, J., van Drongelen, W.: EEG, Temporal Correlations and Avalanches. Journal of Clinical Neurophysiology, Vol. 27, pp. 458–464 (2010)
35. Tajbakhsh, N., Jae-Shin, Y.: Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? (2016)
36. Claassen, J., Foreman, B.: Quantitative EEG for the detection of brain ischemia (2012)
37. Jyoti, G.: A Survey on Brain Stroke Detection in CT Images Using Image Segmentation. ISSUE 1, India, IJRECE, Vol. 4, pp. 41–43 (2016)
38. Menaka, R., Kanchana, R.: Computer Reinforced Analysis for Ischemic Stroke Recognition: A review. Indian Journal of Science and Technology, 8(35) (2015)
39. Gupta, N., Mittal, A.: Brain Ischemic Stroke Segmentation: A Survey (2014)
40. Mani-Maran, A., Saravanan, S.: Artificial Neural Networks (ANNs) for EEG Purging using Wavelet Analysis. International Journal of Electronics and Communication Engineering, Vol. 4, pp. 563–570 (2011)

41. Karthikeyan, S., Ezhilarasi, M.: Automatic Stroke Lesion Segmentation From Diffusion Weighted Mri Images. Issue II, International Journal of Advanced Engineering Technology, Vol. VII, pp. 111–115 (2016)
42. Artificial Neural Network based Lesion Segmentation of Brain MRI. Tapas Si, Arunava De, Anup Kumar Bhattacharjee. 5, New York, USA : Foundation of Computer Science FCS, February de 2016, Communications on Applied Electronics, Vol. 4.
43. Sudharani, K., Sarma, T. C., Prasad, K.: Brain Stroke detection using K-Nearest neighbor and Minimum Mean Distance Technique. 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICT) (2015)
44. Venkateswara-Reddy, B., Bhaskara-Reddy, P., Satish-Kumar, P., Siva-Reddy, S.: Developing an Approach to Brain MRI Image Preprocessing for Tumor Detection. International Journal of Research, Vol. 1, pp. 725–731 (2014)
45. Syed, A. Q., Narayanan, K.: Detection of Tumor in MRI Images Using Artificial Neural Networks. 9, September de 2014, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3
46. Tijare, P. A., Vrushali, D. D.: CYST detection in MRI Brain Image. IMPACT: International Journal of Research in Engineering & Technology, Vol. 2, pp. 129–138 (2014)
47. Bhaiya, L. P., Verma, V. K.: Classification of MRI Brain Images Using Neural Network. International Journal of Engineering Research and Applications, 5, Vol. 2, pp. 751–756 (2012)
48. Anusha, K S., Mathews, M. T., Puthankattil, S. D.: Classification of Normal and Epileptic EEG Signal using Time & Frequency Domain Features through Artificial Neural Network. IEEE 2012 International Conference on Advances in Computing and Communications, pp. 98–101 (2012)
49. Sangeetha, C., Shahin, A.: Brain tumor segmentation using Artificial Neural Network. International Research Journal of Engineering and Technology, Vol. 2 (2015)
50. Nados, W. L., Kumar, R. D., Naser, M. A., Hussein, A. A.: Medical Images Classification by using Artificial Intelligence Techniques. Internal Journal of Scientific Engineering and Technology Research, Vol. 3, No. 34, pp. 6812–6816 (2014)
51. Priyanjana, S., Shagun, M., Surbhi, S., Jyotika, P.: Computer Aided Diagnosis Based on Medical Image Processing and Artificial Intelligence Methods. International Journal of Information and Computation Technology, Vol. 3, No. 9, pp. 887–892 (2013)
52. Vishal, T. V., Srinidhi, S., Srividhya, S., Sri-Vishnu, K., Swathika, R.: A Survey and Comparison of Artificial Intelligence Techniques for Image Classification and Their Applications. International Journal of Science and Research (IJSR), Vol. 5, No. 4 (2016)
53. Rastgarpour, M., Shanbehzadeh, J.: The Status Quo of Artificial Intelligence Methods in Automatic Medical Image Segmentation. International Journal of Computer Theory and Engineering, Vol. 5 (2013)
54. Soltanian-Zadeh, H., Jabarouti-Moghaddam, M.: Medical Image Segmentation Using Artificial Neural Networks. Artificial Neural Networks - Methodological Advances and Biomedical Applications, China, InTech, pp. 121–138 (2011)
55. Alyaa-Hussein, A., Shahad-Imad, A., Ihssan-Subhi, N.: Detection and Segmentation of Ischemic Stroke Using Textural Analysis on Brain CT Images. International Journal of Scientific & Engineering Research, Vol. 6, No. 2 (2015)
56. Boubllil, D., Elad, M., Shtok, J., Zibulevsky, M.: Spatially-Adaptive Reconstruction in Computed Tomography using Neural Networks. European Research Council under European Union's Seventh Framework Program (2014)

57. Lee, H., Lee, J., Kim, N., Kim, S. J., Shin, Y. G.: Robust feature-based registration using a Gaussian-weighted distance map and brain feature points for brain PET/CT images. *Computers in Biology and Medicine*, Vol. 38, No. 9, pp. 945–961 (2008)
58. Diyana, W. M., Zaki, W., Faizal, M., Fauzi, A., Besar, R., Munirah, W. S. H., Ahmad, W.: Multi-level Segmentation Method for Serial Computed Tomography Brain Images. 2009 IEEE International Conference on Signal and Image Processing Applications, pp. 107–112 (2009)
59. Diyana, W. M., Zaki, W., Kong, C. R.: Identifying Abnormalities in Computed Tomography Brain Images Using Symmetrical Features. Selangor, Malasia, 2009 International Conference on Electrical Engineering and Informatics, Vol. 5, pp. 88–92 (2009)
60. Tan, T. L., Sim, K. S., Tan, C. K., Chong, A. K.: CT Image Enhancement by Colorization for Brain Infarct detection (2009)
61. Wan-RosemehahWan, O., Norfaiza, F., Mohd-Nasir, T., Rozita, J., Roshakimah-Mohd, I., Zunuwanas, M., Zaiton, S.: Brainwave Classification for Acute Ischemic Stroke Group Level Using k-NN Technique. 2014 Fifth International Conference on Intelligent Systems, Modelling and Simulation, pp. 117–120 (2014)
62. Lashari, S. A., Ibrahim, R.: A Framework for Medical Images Classification Using Soft Set. The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013), pp. 548–556 (2013)
63. Franco-Martínez, E. A.: Análisis digital de imágenes tomográficas sin contraste para la búsqueda de tumores cerebrales. Departamento de Computación, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, ciudad de México, Unidad Zacatenco, Diciembre (2011)
64. Juárez-Guerra, E., Alarcon-Aquino, V., Gómez-Gil, P.: Epilepsy Seizure Detection in EEG Signals Using Wavelet Transforms and Neural Networks. *Conferences on computer, Information and Systems Sciences and Engineering (CCISSE)*, pp. 1–6 (2013)
65. Ramírez, J. A., Chacón, M. I.: Redes Neuronales Artificiales para el procesamiento de imágenes, una revisión de la última década. *Revista De Ingeniería Eléctrica, Electrónica Y Computación*, Vol. 9 (2011)
66. Ramírez-Quintana, J. A., Chacon-Murguia, M. I., Chacon-Hinojosa, J. F.: Artificial Neural Image Processing Applications: A Survey (2012)
67. Valderrama, J. A., Zárate, A., Martínez, B.: Utilidad Doppler En El Tratamiento De Aneurismas Cerebrales. México, *Revista de especialidades médico quirúrgicas*, Vol. 15, No. 4, pp. 189–195 (2010)
68. Kasabov, N., Capecchi, E.: Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes, *Information Sciences*, Vol. 294, pp. 535–575 (2015)
69. Ostrek, G., Nowakowski, A., Jasionowaka, M., Przelaskowski, A., Szopinski, K.: Stroke Tissue Pattern recognition based on CT texture analysis (2014)
70. Hachaj, T.: Artificial intelligence methods for understanding dynamic computer tomography perfusion maps (2010)
71. Jiang, J., Trundle, P., Ren, J.: Medical Image Analysis with Artificial Neural Networks (2011)
72. Gal, N., Stoicu-Tivadar, V.: Simulation of Medical Image Interpretation. INES 2011 15th International Conference on Intelligent Engineering Systems, pp. 33–37 (2011)
73. Balasooriya, U., Perera, M. U. S.: Intelligent Brain Hemorrhage Diagnosis System. IEEE (2011)

74. Adam, S .Y., Yousif, A., Bakri, M.: Classification of Ischemic Stroke using Machine Learning Algorithms (2016)
75. Dimililer, K., İlhan, A.: Effect of image enhancement on MRI brain images with neural networks. Vienna, Austria, 12th International Conference on Application of Fuzzy Systems and Soft Computing (ICAFS 2016), pp. 39–44 (2016)
76. Burcu-Toprak, I., Fatih-Caglar, M., Merdan, M.: Automatic Recognition of Epilepsy from EEG using Artificial Neural Network and Discrete Wavelet Transform (2007)
77. RadiologyInfo.org para pacientes. Disponible en: <http://www.radiologyinfo.org/sp/info.cfm?pg=ultrasound-cranial>
78. Zhanyu, M., Zheng-Hua, T., Jun, G.: Feature selection for neutral vector in EEG signal classification. *Neurocomputing*, pp. 937–945 (2016)
79. Dridi, M., Bouallegue, B., Ali-Hajjaji, M., Mtibaa, A.: An Enhancement Medical Image Compression Algorithm Based on Neural Network. *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 5 (2016)
80. García, C.: Análisis de la deformación craneal por aplicación de esfuerzos. Instituto Politécnico Nacional, Centro de Investigación en Computación (2013)
81. Sharma, N., Lalit, M.: Automated medical image segmentation techniques. India, *Journal of Medical Physic*, Vol. 35, No. 1, pp. 3–14 (2010)
82. Cárdenas, S. Y.: Implementación y Evaluación de Redes Neuronales Artificiales tipo “Pulse-Coupled Neural Networks” (PCNN) aplicadas a Visión Artificial. CENIDET, Cuernavaca, Morelos, México (2015)
83. Rastgarpour, M., Shanbehzadeh, J.: Application of AI Techniques in Medical Image Segmentation and Novel Categorization of Available Methods and Tools. In: *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. I, Iran, (2011)
84. Gordillo, N., Montseny, E., Sobrevilla, P.: State of the art survey on MRI brain tumor segmentation. México, *Magnetic Resonance Imaging*, pp. 1426–1438 (2013)
85. Lay-Khoo, L., Siau-Chuin, L.: A Survey of Medical Image Processing Tools. In: *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*, pp. 171–176 (2015)
86. Rodrigues, P. M., Teixeira, J. P.: Classification of Electroencephalogram Signals Using Artificial Neural Networks. In: *2010 3rd International Conference on Biomedical Engineering and Informatics* (2010)

Selección de parámetros en el enfoque de bolsa de características para clasificación de habla imaginada en electroencefalogramas

Jesús S. García-Salinas, Luis Villaseñor-Pineda, Carlos A. Reyes-García,
Alejandro A. Torres-García

Instituto Nacional de Astrofísica Óptica y Electrónica,
Laboratorio de Bioseñales y Cómputo Médico, México

{jss.garcia,villasen,kargaxxi,alejandro.torres}@inaoep.mx

Resumen. El uso de las interfaces cerebro-computadora como un nuevo canal de comunicación ha despertado un gran interés, no obstante aún hay muchos retos por resolver para alcanzar una comunicación natural. En el caso particular de las BCIs basadas en habla imaginada aún se presentan dificultades para analizar las señales cerebrales. En este trabajo se explora una representación de bolsa de características para identificar patrones en la señal cerebral adquirida a través de electroencefalogramas. Esta técnica ha demostrado ser útil en tareas relacionadas. Sin embargo, determinar la bolsa de características más adecuada al problema depende de diversos parámetros. El presente trabajo describe el uso de algoritmos genéticos para encontrar la configuración más adecuada para la clasificación de habla imaginada. Los resultados, calculados en una base de datos de habla imaginada de 27 sujetos, muestran la pertinencia del método alcanzando resultados comparables con el estado del arte.

Palabras clave: interfaces cerebro-computadora, electroencefalogramas (EEG), habla imaginada, bolsa de características, algoritmos genéticos.

Parameter Selection in Bag of Features to Classify Imagined Speech on Electroencephalograms

Abstract. The use of brain-computer interfaces as a new communication channel has become an interest topic, nevertheless, there are many challenges to achieve a natural communication. In the particular case of imagined speech BCIs, there are difficulties to analyze the brain signals. In this work, a bag of features representation to identify patterns in electroencephalograms is explored. This method has shown to be useful in similar tasks. Nevertheless, to determine the most adequate bag of features to the problem depends on many parameters. The present work describes the use of genetic algorithms to find an adequate configuration to imagined speech classification. The results, obtained from an imagined speech database of 27 subjects, show the relevance of the method achieving results which are comparable with related work.

Keywords: brain-computer interfaces, electroencephalograms, imagined speech, bag of features, genetic algorithms.

1. Introducción

Cada vez existe un mayor interés en las interfaces cerebro computadora (BCI). Inicialmente se despertó este interés al brindar un nuevo canal de comunicación a personas con discapacidad. No obstante, debido a la reducción de costos de los dispositivos para lectura de la señal cerebral, esta nueva forma de comunicación puede ser utilizada por cualquier persona.

Para controlar un dispositivo a través de una BCI, el usuario debe producir un patrón de actividad cerebral, ya sea evocado internamente o por un estímulo externo, el cual será identificado por el sistema y transformado en comandos para dicho dispositivo. La lectura de la actividad cerebral puede realizarse con diferentes instrumentos, en particular en este trabajo se utiliza el electroencefalograma (EEG) para registrar la señal electrofisiológica. Además, este trabajo se orienta a analizar la señal cerebral evocada por el habla imaginada, es decir, cuando el sujeto imagina la dicción de la palabra sin emitir ni articular sonidos.

A pesar de contar con diferentes métodos computacionales para el procesamiento, caracterización y clasificación de las señales cerebrales en EEG [10] el análisis de las señales cerebrales resultado del habla imaginada, presentan propiedades que complican su análisis [8]. Distintas soluciones se han propuesto para realizar esta tarea, sin embargo, aún existen retos importantes para lograr una comunicación natural y fluida mediante una BCI basada en habla imaginada.

El presente trabajo muestra la generación de una representación del habla imaginada a partir del método de bolsa de características. Las bolsas de características tienen como objetivo lograr una caracterización automática al obtener, en primera instancia, *unidades características* de la señal, y posteriormente generar un patrón representativo a partir de ellas. De esta forma, el método encuentra un conjunto de unidades características representativo de cada clase (i.e. cada palabra imaginada), y estos conjuntos son usados para el reconocimiento y clasificación del vocabulario imaginado por el sujeto. El descubrimiento de las unidades características se realiza a partir de la señal utilizando algún método de *clustering*. Los prototipos calculados se consideran las unidades características, comúnmente llamadas *codewords*. Cada *codeword* es una entrada del diccionario general, llamado *codebook*.

Para la generación de las *codewords* es necesario tener en cuenta el método de extracción de características de la señal cerebral, además de definir *a priori* el tamaño del *codebook*. El objetivo de los experimentos mostrados en este trabajo consiste en encontrar la configuración más apropiada entre las combinaciones posibles de estos parámetros. Para ello, se utilizó un algoritmo genético evaluado sobre una base de datos de habla imaginada (5 palabras) de 27 sujetos. En un intento por generalizar los parámetros, el método es aplicado a todos los sujetos utilizando los mismos parámetros, y la evaluación se realizó obteniendo la exactitud promedio de entre todos ellos. Los resultados obtenidos, a pesar de

ser preliminares, muestran la factibilidad del método, al discriminar segmentos representativos de la señal para las distintas palabras en habla imaginada.

2. Trabajo relacionado

2.1. Primeros enfoques

La comunicación a través de señales cerebrales comenzó con la detección de potenciales eléctricos indirectamente relacionados al proceso cognitivo del habla. Uno de los primeros trabajos reportados, presentado en [3], hace uso de la activación o bloqueo de los ritmos alfa, un rango de frecuencias de la señal cerebral, para generar código Morse. Esto requiere entrenamiento previo que consiste en manipular la configuración oculomotora para lograr el control de los ritmos alfa.

En [4] se utilizan señales P300 para detectar respuestas visuales sobre caracteres mediante un alfabeto en una pantalla, este sistema detecta la respuesta del cerebro a un estímulo visual que se mueve a través de esta pantalla, la respuesta cerebral permite determinar el carácter que el sujeto desea comunicar.

2.2. Habla imaginada

Métodos anteriormente usados en BCIs involucran que el usuario aprenda a generar señales cerebrales específicas, o aprovechan la respuesta cerebral natural a cierto estímulo externo. Una BCI basada en habla imaginada, trabaja con la señal evocada por el proceso cognitivo del habla, con la gran ventaja que el sujeto no tiene que aprender a generar señales cerebrales específicas.

A continuación se presentan diversos trabajos relacionados al habla imaginada. Cabe hacer notar que cada uno de ellos difieren no sólo en el método usado, sino también en la evaluación, donde se experimentó con diferentes conjuntos de sujetos y distintos vocabularios de palabras imaginadas.

La clasificación de palabras imaginadas fue presentada inicialmente en [14], donde se analizan las señales EEG y EMG para clasificar siete palabras (*first, second, third, yes, no, right, left*). La caracterización se basó en la Transformada Rápida de Fourier (FFT) y filtros pasa banda, para después aplicar la Transformada Inversa de Fourier (IFFT). La clasificación se realizó mediante mínimos cuadrados para comparar la señal con un prototipo creado a partir de la media de las muestras, obteniendo una exactitud promedio de 52.57 ± 20 para cinco sujetos.

En [2] se propone la clasificación de dos vocales imaginadas a través de patrones espaciales comunes (CSP), Support Vector Machines (SVM) y filtros pasa banda, logrando una exactitud promedio de 62.6 ± 8.3 para tres sujetos.

En [16] se utilizó un vocabulario en español de cinco palabras (*arriba, abajo, izquierda, derecha, seleccionar*), se utilizaron canales de EEG cercanos al área de lenguaje y se aplicó un filtro pasa banda entre 4 y 25 Hz. La caracterización se realizó mediante la Transformada Wavelet Discreta (DWT) y se entrenaron

cuatro clasificadores: Naive Bayes, Random Forest, Support Vector Machines y Bagging Random Forest, y se obtuvo una exactitud de 41.96 ± 3 para tres sujetos.

Un esquema más simple, mostrado en [13], clasifica entre dos palabras en árabe (*Si*, *No*). Se analizaron los ritmos alfa y beta de un EEG de un solo canal mediante dos métodos, el primero obtiene datos estadísticos de la señal (mínimo, máximo y media) y el segundo aplica la DWT con seis niveles de descomposición. La clasificación se llevó a cabo mediante SVM, Linear Discriminant Analysis (LDA), Self-Organized Maps (SOM), Multilayer Perceptron y ensambles de ellos, la exactitud promedio obtenida fue de 56 para un conjunto de siete sujetos.

Recientemente, en [17] se exploraron diversas familias wavelet y clasificadores para clasificación multiclase de cinco palabras (*arriba*, *abajo*, *izquierda*, *derecha*, *seleccionar*). Se implementó una selección de canales automática basada en inferencia difusa para reducir el conjunto de datos y obtuvo una exactitud de 68.18 ± 16 .

2.3. Bolsa de características

Este método está basado en el enfoque tradicional de *cuantificación vectorial* (Vector Quantization) cuyo objetivo es lograr una caracterización automática de la señal, discretizando su representación. En el área de análisis de señales se han presentado diferentes variantes y adaptaciones; y recibe diferentes nombres de acuerdo a su área de aplicación como, por ejemplo, bolsa de patrones, bolsa de imágenes o bolsa de segmentos.

En breve, la señal es segmentada y las unidades representativas son generadas mediante una técnica de *clustering*, los prototipos de cada clúster reciben el nombre de *codewords* y en conjunto forman el *codebook*. Una vez generado el *codebook*, se toma la señal segmentada y por cada segmento se identifica aquella *codeword* más similar. Una vez codificada la señal se calcula un histograma de las *codewords* presentes en la señal (véase la siguiente sección para su descripción formal).

En el caso específico de señales de EEG, la bolsa de características fue usada por [18], con el uso de señales de EEG (y EKG) para detección de epilepsia. Las señales fueron obtenidas mediante un EEG de un solo canal y se extrajeron características mediante la DWT, posteriormente fueron agrupadas mediante *k-means*. Los histogramas se crearon mediante *1-Nearest Neighbor* y finalmente se clasificó con un clasificador *1-Nearest Neighbor*, la exactitud obtenida fue de 87.8 ± 2.3 .

En [12], se presenta una modificación al método de *clustering*, al que llaman *Bag of Super-Features* o Bolsa de Super-Características. El método consiste en generar clústers para cada una de las clases y posteriormente unirlos. Realizar esta división y calcular de esta forma el *codebook* reduce la pérdida de información de las clases [9].

3. Método

3.1. Bolsa de características

Una serie de tiempo está definida por $x^i = (x_1^i, x_2^i, \dots, x_p^i)$, para p muestras. Cada instancia x^i esta asociada con una clase y^i para $i = 1, 2, \dots, n$ y $y^i \in \{1, 2, \dots, C\}$ donde n es el número de instancias y C es el número de clases. Para extraer patrones locales, es necesario deslizar una ventana w sobre las series de tiempo, el deslizamiento m no puede ser mayor al tamaño de la ventana $m \leq w$. Las subsecuencias extraídas serán $\lceil \frac{p-w+1}{m} \rceil$, por lo tanto, el conjunto de datos tendrá $n(\lceil \frac{p-w+1}{m} \rceil)$ subsecuencias. A continuación se aplica un algoritmo de *clustering* con k centroides, que serán los *codewords* de nuestro *codebook* $K \in \mathbb{R}^{(w \times d)}$ [7].

El método propuesto está basado en el trabajo presentado en [18] al que le fueron realizadas adaptaciones, el esquema general se muestra en la fig. 1, y se aplica individualmente a cada uno de los sujetos en la base de datos, y finalmente los resultados son promediados en un intento por observar de forma general los resultados alcanzados por el modelo para todos los sujetos.

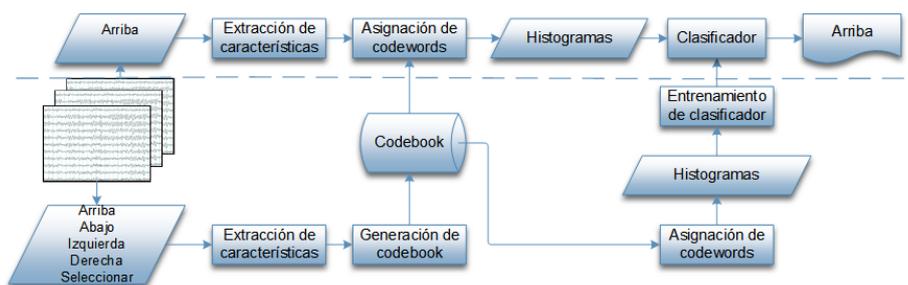


Fig. 1. Diagrama del método.

El resultado es la exactitud total de clasificación de las palabras imaginadas en el conjunto de datos. En nuestro caso, es la exactitud la que se toma como función objetivo en el algoritmo genético, y con ello determinar los valores más adecuados para el conjunto de parámetros en cada una de sus etapas.

Datos. El conjunto de datos fue obtenido en [6], se registraron los EEG de 27 sujetos hablantes nativos de español a través de un kit EPOC de la compañía Emotiv que cuenta con 14 canales (AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, O2) y una frecuencia de muestreo de 128 Hz. Los datos consisten en 5 palabras imaginadas ("arriba", "abajo", "izquierda", "derecha" y "seleccionar") repetidas 33 veces cada una, con un tiempo de descanso entre cada repetición.

Los datos fueron procesados mediante la Referencia Promedio Común (CAR), por sus siglas en inglés, para reducir el ruido obtenido en la toma de muestras.

El CAR es una técnica computacionalmente simple y ha demostrado superar algunas técnicas de referenciado eléctrico [11]. También se aplicó un filtro pasa bajas Butterworth con corte en 50 Hz, para eliminar algunos artefactos.

Ya que cada conjunto de datos está conformado de 14 canales, se procesó cada señal de forma independiente, es decir, el procesamiento fue realizado a cada una de ellas por separado. Cabe señalar que para el proceso de creación del clasificador se tomaron, para cada sujeto, 25 repeticiones aleatoriamente para entrenamiento y generación del *codebook*, y las 8 restantes fueron utilizadas para pruebas de clasificación.

Extracción de características. La extracción de características de la señal se llevó a cabo mediante una ventana en movimiento, con un tamaño W y un desplazamiento M que podía tener sobreposición pero no saltos, es decir, $M \leq W$ donde $M > 0, W \geq 8$, además se tiene un límite superior de 128, que es la frecuencia de muestreo de la señal.

En este caso se utilizaron únicamente los coeficientes obtenidos por la transformada rápida de Fourier [5] y la energía relativa de la transformada Wavelet, $C = FFT, DWT$. La transformada Wavelet es la integral de una señal multiplicada por versiones escaladas y desplazadas de una función wavelet. Una vez obtenidos los coeficientes de la DWT, se obtiene la energía relativa mediante el método mostrado en [15].

Clustering. El algoritmo de *clustering* utilizado fue k-means++ [1], a partir de cada una de las señales se obtuvieron K prototipos o *codewords* que forman el *codebook*.

Cabe señalar que para conformar el *codebook* se aplicó el *clustering* por separado a cada una de las clases (i. e. las palabras a reconocer), el número de clústers se divide entre el número de clases para tener una distribución equitativa de los clústers, una vez generados los prototipos de cada una de las clases son concatenados para formar el *codebook*.

Generación de histogramas. Una vez generado el *codebook*, el siguiente paso es codificar los segmentos de las señales, esto es, decidir a cual de los elementos del *codebook* son más parecidos. Esta asignación se realiza utilizando una búsqueda del vecino mas cercano (1-NN) tomando como medida la distancia Euclidiana.

Una vez asignados los *codewords*, se genera un histograma por cada señal. Los histogramas fueron etiquetados según la palabra imaginada a la que corresponden, y esta información fue utilizada para crear el clasificador. En nuestro caso se utilizó el método de *Naive Bayes Multinomial*.

3.2. Algoritmo genético

Dada la combinatoria entre los parámetros para calcular la bolsa de características, se aplicó un algoritmo genético para determinar la configuración más

adecuada. Los parámetros a definir en la generación de la bolsa de características son entonces: el tipo de extracción de características ($C = \text{FFT}, \text{DWT}$), el tamaño de ventana para la extracción de características ($8 \leq W \leq 128$), el desplazamiento de la ventana ($8 \leq M \leq 128$), y el número de clústers ($K \leq 1000$); dando un total de 4 parámetros.

C = FFT	K = 100	W = 128	M = 32
Tipo de caracterización	Tamaño de <i>codebook</i>	Tamaño de ventana	Movimiento de la ventana

Fig. 2. Ejemplo de un individuo del algoritmo genético.

La función objetivo del algoritmo se orientó a reducir el error de clasificación promedio de todos los sujetos.

El tamaño de población y las generaciones se fijaron en 100, la fracción de cruce de individuos es de 80 % y se lleva a cabo tomando una mezcla de promedios pesados de los padres. El proceso de selección se lleva a cabo mediante una función estocástica uniforme que se mueve a través del conjunto de individuos con un número de pasos fijo. Además, se considera un elitismo de 2 %, es decir, 2 individuos pasan sin alteraciones a la siguiente generación. Las probabilidades de mutación de un sujeto en cada generación es de 1 %. Este sujeto será alterado aleatoriamente para diversificar los parámetros obtenidos.

No se definieron límites de tiempo o de aptitud que detuvieran la ejecución del algoritmo, es decir, el algoritmo podrá continuar hasta completar las generaciones ignorando el tiempo y la exactitud obtenida en cada generación. El único criterio de paro es no encontrar un cambio en el promedio de aptitud después de 20 generaciones, en este caso el algoritmo habrá llegado a un estancamiento y detendrá su ejecución, tomando el mejor sujeto hasta ese instante.

4. Experimentos y resultados

Los resultados de clasificación se obtuvieron separando aleatoriamente el 75 % de las instancias para entrenamiento y 25 % para pruebas. Se clasificaron cinco clases que corresponden a cada una de las palabras imaginadas mediante un clasificador *Naive Bayes Multinomial*.

El algoritmo genético obtuvo como parámetros óptimos: una caracterización C mediante FFT, un tamaño de *codebook* K de 75, una ventana W de tamaño 40 con un desplazamiento M de 8. Tomando en cuenta estos parámetros se obtuvieron los resultados de clasificación mostrados en la figura 4.

Como es de esperar, el algoritmo genético minimiza la aptitud del método en cada generación, en este caso se reduce el error de clasificación. En la figura 3 se puede observar la aptitud obtenida en cada generación del algoritmo genético,

se llegó a un estancamiento de la aptitud en la generación número setenta y uno, es decir, la variación de la exactitud obtenida en cada generación fue menor a 1×10^{-6} .

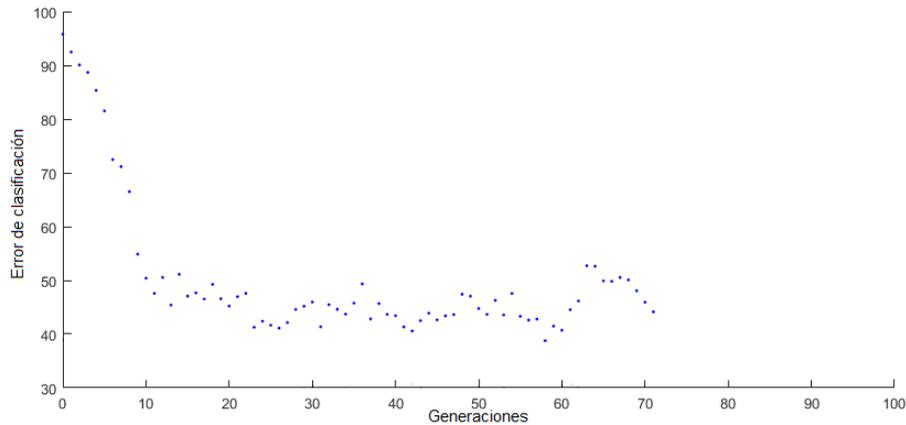


Fig. 3. Comportamiento del algoritmo genético.

El experimento con los parámetros obtenidos fue repetido diez veces, la exactitud promedio alcanzada se muestra en la figura 4. La exactitud se obtiene del total obtenido por las cinco clases. La desviación del promedio representa la desviación de las exactitudes de los veintisiete sujetos.

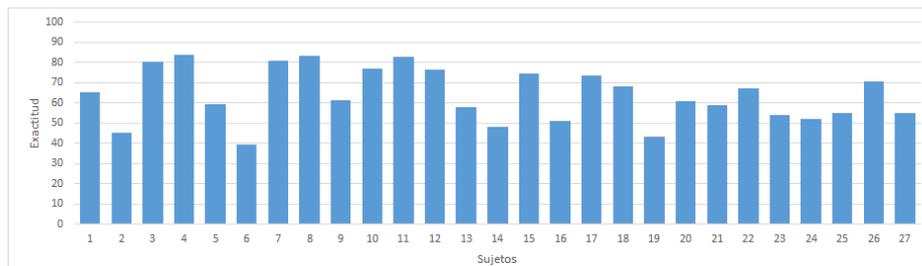


Fig. 4. Exactitud promedio para 10 iteraciones del método al aplicarlo a cada uno de los 27 sujetos de nuestra base de datos.

Para tener una idea de como se comporta la clasificación a nivel de palabras, se presenta una matriz de confusión promedio en la tabla 2. La matriz se calculó al promediar y normalizar con los resultados obtenidos de todos los sujetos considerando las 10 iteraciones del método.

Tabla 1. Resultados promedio de exactitud para 10 iteraciones por sujetos con su desviación estándar, comparado con los resultados alcanzados en [17].

Sujeto	Método propuesto	Torres et al. 2016
1	65.3 ± 5.5	88.38
2	45.3 ± 5.5	50.77
3	80.3 ± 4.5	69.74
4	83.8 ± 7.6	77.46
5	59.3 ± 6.5	73.34
6	39.5 ± 6.3	40
7	80.8 ± 7.7	70.62
8	83.5 ± 6.8	89.7
9	61.5 ± 3.8	81.21
10	77.0 ± 6.2	70.92
11	83.0 ± 6.9	90.33
12	76.5 ± 5.6	75.11
13	58.0 ± 6.9	66.65
14	48.3 ± 8.1	48.01
15	74.5 ± 9.3	85.44
16	51.3 ± 6.0	63.53
17	73.8 ± 7.8	67.87
18	68.3 ± 4.7	81.8
19	43.5 ± 5.2	46.76
20	61.0 ± 6.6	83.09
21	59.0 ± 6.7	48.42
22	67.0 ± 2.8	70.33
23	54.3 ± 4.6	63.2
24	52.0 ± 9.1	57.46
25	55.0 ± 6.2	29.78
26	70.5 ± 7.2	67.24
27	55.0 ± 7.0	83.64
Prom.	64.0 ± 13.24	68.18 ± 15.9

Tabla 2. Matriz de confusión promedio considerando todos los sujetos.

Arriba	Abajo	Izquierda	Derecha	Seleccionar	
0.73	0.12	0.05	0.06	0.06	Arriba
0.11	0.62	0.08	0.14	0.06	Abajo
0.04	0.08	0.60	0.14	0.15	Izquierda
0.06	0.16	0.14	0.58	0.08	Derecha
0.05	0.08	0.15	0.06	0.67	Seleccionar

Como puede observarse la palabra con menor confusión es *arriba* y la de mayor confusión es *derecha*. Cabe notar que esta última palabra se confunde principalmente con las palabras *abajo* e *izquierda*.

5. Conclusiones

Los parámetros obtenidos con el algoritmo genético lograron una configuración que permitió crear representaciones basadas en bolsa de características cuya clasificación fue comparable con lo reportado en el estado del arte. La exactitud promedio alcanzada fue de 64 ± 13.24 para 5 palabras imaginadas por 27 sujetos.

Como se muestra en la figura 3, el algoritmo logró converger a una solución óptima en menos generaciones de las esperadas, esto abre la posibilidad de aumentar el número de parámetros. Entre los parámetros a añadir pueden considerarse los parámetros propios de las técnicas de extracción de características. Por ejemplo, la familia Wavelet o el número de niveles de descomposición. Por otro lado, aún falta observar el resultado al aplicar el método de forma individual, es decir, ajustar los parámetros a cada sujeto. Esto podría mejorar los resultados individuales de clasificación pero incidir sobre las conclusiones generales sobre los parámetros del método.

Como trabajo futuro, se desea explorar con distintos métodos en algunas de las etapas en el cálculo del *codebook*. Por ejemplo, utilizar un método de *clustering*, como *Expectation Maximisation*, esto permitiría eliminar el parámetro de tamaño de clúster del algoritmo genético. Así como la modificación en la generación de histogramas, donde es posible considerar secuencias de *codewords*, incluyendo de esta forma información temporal de la señal.

Agradecimientos. Los autores agradecen al Consejo Nacional de Ciencia y Tecnología (CONACyT) por al apoyo a esta investigación a través de la beca # 702603 y del proyecto 2016-01-2228. Asimismo, agradecen al apoyo del Ministerio de Asuntos Exteriores y de Cooperación Internacional de Italia y la Agencia Mexicana de Cooperación Internacional para el Desarrollo (AMEXCID) bajo el proyecto MX14MO06 INAOE-Universidad de Florencia.

Referencias

1. Arthur, D., Vassilvitskii, S.: K-means++: The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1027–1035. SODA '07, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2007), <http://dl.acm.org/citation.cfm?id=1283383.1283494>
2. DaSalla, C.S., Kambara, H., Sato, M., Koike, Y.: Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks* 22(9), 1334 – 1339 (2009), <http://www.sciencedirect.com/science/article/pii/S0893608009000999>, brain-Machine Interface

3. Dewan, E.M.: Occipital alpha rhythm eye position and lens accommodation. *Nature* 214, 975–977 (1967)
4. Farwell, L., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology* 70(6), 510 – 523 (1988), <http://www.sciencedirect.com/science/article/pii/0013469488901496>
5. Frigo, M., Johnson, S.G.: The design and implementation of fftw3. *Proceedings of the IEEE* 93(2), 216–231 (2005), <http://ieeexplore.ieee.org/abstract/document/1386650/>
6. González Castañeda, E., Torres-García, A., Reyes-García, C., Villaseñor-Pineda, L.: Applying Brain Signals Sonification for Automatic Classification. *Revista Mexicana de Ingeniería Biomédica* 36(3), 233–248 (Sep 2015)
7. Gui, Z.W., Yeh, Y.R.: Time Series Classification with Temporal Bag-of-Words Model, pp. 145–153. Springer International Publishing, Cham (2014)
8. Klonowski, W.: Everything you wanted to ask about eeg but were afraid to get the right answer. *Nonlinear Biomedical Physics* 3(1), 1 – 5 (2009), <http://dx.doi.org/10.1186/1753-4631-3-2>
9. Lazebnik, S., Raginsky, M.: Supervised learning of quantizer codebooks by information loss minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(7), 1294–1309 (July 2009)
10. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering* 4(2), R1 (2007), <http://stacks.iop.org/1741-2552/4/i=2/a=R01>
11. Ludwig, K.A., Miriani, R.M., Langhals, N.B., Joseph, M.D., Anderson, D.J., Kipke, D.R.: Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of Neurophysiology* 101(3), 1679–1689 (2009), <http://jn.physiology.org/content/101/3/1679>
12. Plinge, A., Grzeszick, R., Fink, G.A.: A bag-of-features approach to acoustic event detection. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3704–3708 (May 2014)
13. Salama, M., Lashin, H., Gamal, T.: Recognition of unspoken words using electrode electroencephalographic signals. *COGNITIVE 2014 : The Sixth International Conference on Advanced Cognitive Technologies and Applications* pp. 51–55 (2014)
14. Suppes, P., Lin, L.Z., Bing, H.: Brain wave recognition of words. *Proceedings of the National Academy of Sciences* 94(26), 14965 – 14969 (1997), <http://www.pnas.org/content/94/26/14965.abstract>
15. Torres-García, A., Reyes-García, C., Villaseñor-Pineda, L., Ramírez, J.: Análisis de señales electroencefalográficas para la clasificación de habla imaginada. *Revista mexicana de ingeniería biomédica* 34(1), 23–39 (2013)
16. Torres-García, A.A., Reyes-García, C.A., Villaseñor-Pineda, L.: Toward a silent speech interface based on unspoken speech. *Proceedings of biosignals* pp. 370 – 373 (2012)
17. Torres-García, A.A., Reyes-García, C.A., Villaseñor-Pineda, L., García-Aguilar, G.: Implementing a fuzzy inference system in a multi-objective {EEG} channel selection model for imagined speech classification. *Expert Systems with Applications* 59, 1 – 12 (2016), <http://www.sciencedirect.com/science/article/pii/S0957417416301774>
18. Wang, J., Liu, P., She, M.F., Nahavandi, S., Kouzani, A.: Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control* 8(6), 634 – 644 (2013), <http://www.sciencedirect.com/science/article/pii/S174680941300089X>

Hacia la clasificación de actividad e inactividad lingüística a partir de señales de electroencefalogramas (EEG)

Luis Alfredo Moctezuma¹, Maya Carrillo¹, Luis Villaseñor Pineda²,
Alejandro A. Torres García²

¹Benemérita Universidad Autónoma de Puebla,
Puebla, México

²Instituto Nacional de Astrofísica Óptica y Electrónica,
Tonantzintla, Puebla, México

luisalfredomoctezuma@gmail.com, cmaya@cs.buap.mx, {villaseñor,
alejandro.torres}@ccc.inaoep.mx

Resumen. En este artículo se evalúan dos formas de caracterización de señales de EEG para clasificación de actividad e inactividad lingüística, preproceso indispensable para el desarrollo de interfaces cerebro-computadora (BCI) basadas en habla imaginada. La primera caracterización utilizó el cálculo de la energía *teager* con base en la transformada discreta wavelet (DWT). La segunda caracterización se basó en el cálculo de valores estadísticas directamente sobre la señal. Los experimentos se realizaron sobre dos bases de datos de habla imaginada, una con 27 sujetos y otra con 20. Los resultados muestran que al usar los valores estadísticos, la actividad e inactividad lingüística se distingue mejor que al usar la DWT; sin embargo, cabe señalar que con ambas caracterizaciones se alcanzan altas tasas de exactitud.

Palabras clave: habla imaginada, interfaces cerebro computadora, electroencefalogramas (EEG), transformada discreta wavelet (DWT), características estadísticas.

Towards Classification of Activity and Linguistic Inactivity from Electroencephalograms (EEG) Signals

Abstract. This article evaluates two ways of characterization of EEG signals for classification of activity and linguistic inactivity, essential preprocess for the development of brain-computer interfaces (BCI) based on imagined speech. The first characterization used the calculation of the *teager* energy based on the discrete wavelet transform (DWT). The second characterization was based on the calculation of statistical

values directly from the signal. The experiments were performed on two imagined speech databases, the first one with 27 subjects and the other one with 20. The results show that when using statistical values, linguistic inactivity and activity are better distinguished than when using DWT; however, it should be noted that with both characterizations high accuracy rates are reached.

Keywords: imagined speech, brain computer interfaces, electroencephalograms (EEG), discrete wavelet transform (DWT), statistical characteristics.

1. Introducción

La organización mundial de la salud (OMS) define el término de discapacidad como todas las deficiencias, limitaciones y las restricciones de la participación en situaciones vitales, por ejemplo, la comunicación hablada. Datos del Instituto Nacional de Estadística y Geografía¹ (INEGI) del 2015 muestran que el 6 % de la población mexicana sufre alguna discapacidad, de las cuales el 18 % sufre alguna discapacidad para hablar o comunicarse; entre las más destacadas se encuentran la esclerosis lateral amiotrófica, esclerosis múltiple, lesiones de médula espinal o cerebral, entre otras. Las personas que sufren dichas discapacidades tienen muy diversos retos al convivir día a día con las personas que los rodean, no pueden desempeñar algunas tareas y no pueden comunicar lo que piensan o sienten. Una solución a esta situación es el uso de interfaces cerebro computadora (BCI, por sus siglas en inglés) que brindan un nuevo canal de comunicación a personas con este tipo de discapacidades.

Una BCI es un sistema de comunicación que monitorea la actividad cerebral y traduce determinadas características, correspondientes a las intenciones del usuario, en comandos de control de un dispositivo. Las BCI se pueden dividir en invasivas y no invasivas. Las primeras requieren de procedimientos quirúrgicos y se obtienen señales cerebrales muy nítidas gracias a que las medidas no son atenuadas por el cráneo y el cuero cabelludo. Las segundas, las BCI no invasivas, no requieren ningún tipo de cirugía, sin embargo, las señales que se obtiene son más débiles. Dentro de las no invasivas las basadas en electroencefalogramas son las más utilizadas debido a su costo relativamente bajo.

Las fuentes electrofisiológicas se refieren a los mecanismos o procesos neurológicos usadas por un usuario de una BCI para generar señales de control[2]. Wolpaw[3] separó estas fuentes en 5 categorías basadas en los mecanismos neuronales y la tecnología de grabación que utilizan. Estas categorías son: actividad sensoriomotora, potenciales P300, potenciales evocados visuales(VEP), potenciales corticales lentos(SCP) y actividad de la célula neural(ANC). Luego Bashashati[4] agregó 2 categorías más: respuesta a tareas mentales y múltiples neuromecanismos.

¹ Estadísticas a propósito del día internacional de las personas con discapacidad realizado por el INEGI en México (2015)

Sin embargo, dichas fuentes electrofisiológicas presentan dos problemas: (i) el largo periodo de entrenamiento, pues las fuentes electrofisiológicas son generadas por el usuario de manera poco consiente; y (ii) las bajas tasas de comunicación, una sola palabra o menos procesada por minuto, lo que hace imposible la comunicación de manera natural. Es por esto que Desain[5] dice que aunque prometedor, todavía se está lejos de lo que se necesita para un control rápido y fiable de las interfaces y que en lugar de tratar de mejorar los métodos que usan las fuentes electrofisiológicas anteriores, se deben aprovechar otras tareas mentales. Entre ellas menciona el habla imaginada o interna; que se refiere a la pronunciación interna, o imaginada, de palabras pero sin emitir sonidos ni articular gestos.

El habla imaginada como fuente electrofisiológica tiene ventajas respecto a las otras, pues necesita poco entrenamiento; sin embargo, es una tarea compleja, pues está ligada a múltiples artefactos tales como movimiento de músculos y pulsos de arterias en el cuero cabelludo. En este trabajo se explora lo que sería un primer paso de una BCI utilizando EEG para habla imaginada : distinguir la actividad cerebral generada por el sujeto al imaginar una palabra (i.e. emprende una actividad lingüística) de cualquier otra actividad cerebral. Los experimentos realizados buscan evaluar diferentes formas de representación de la señal para realizar dicha distinción. Las señales de EEG fueron separadas en 2 clases, un conjunto de palabras imaginadas consideradas la clase de *actividad lingüística*; y los estados de reposo o pausa como ejemplos de otra actividad cerebral. Estos últimos denominados por simplicidad *inactividad lingüística*.

2. Trabajos relacionados

A continuación se presenta una revisión de los trabajos realizados en la tarea de reconocimiento de habla no pronunciada o imaginada. En algunos de ellos se han reportado experimentos enfocados a reconocer sólo sílabas o fonemas para el control de algún dispositivo, y otros consideran un enfoque orientado a reconocer palabras imaginadas. Cabe señalar que no todos estos trabajos intentan detectar el inicio del habla imaginada.

Respecto al reconocimiento de palabras imaginadas, en [11] se grabaron 20 repeticiones de las palabras imaginadas: *alpha, bravo, charlie, delta, echo* de 18 sujetos. Para extraer sus características se utilizó *Double-Tree Complex Wavelet Transform (DTCWT)*, y para la etapa de clasificación *offline* se utilizó Janus recognition toolkit(Jrtk). La tasa de reconocimiento promedio alcanzada fue de 45.50%. En la discusión presentada por los autores, se destacó que ciertos inconvenientes afectan en general a los métodos de caracterización y clasificación, el primero es que cuanto mayor sea el tiempo de reposo, la tasa de reconocimiento disminuirá; y el segundo es que sólo los bloques de palabras imaginadas que fueron grabadas en la misma sesión obtendrán tasas de reconocimiento significativas. Estos puntos deben tenerse en cuenta para crear un dispositivo que pueda reconocer palabras imaginadas en tiempo real.

En [9] se hizo una aproximación hacia detectar el inicio del habla imaginada vs. estados de inactividad, en un estado de relajación y con el menor ruido posible en el área. En dicho trabajo se realizaron 3 tareas con 4 personas (3 hombres y una mujer) para obtener información lingüística. La primera tarea, llamada *overt*, consiste en obtener la señal cerebral cuando el sujeto imaginaba la palabra. La segunda tarea, llamada *I_overt*, se exhalaba y tensaban las cuerdas vocales sin pronunciar el sonido; y la tercera tarea, llamada *covert*, consistió en imaginar y escuchar el sonido. La única palabra que imaginaron los sujetos fue la palabra *um* y se hicieron 6 registros con cada uno de los participantes.

Cada registro es una secuencia aleatoria de entre las tres tareas, por ejemplo: *over-covert-covert-I_overt...* etc. Cada tarea fue de 5 segundos, y después de procesar los datos se tomaron ventanas de tamaño 1,1.5, 2 y 3 segundos para extraer características usando un modelo de regresión lineal (Método de Burg). Lo anterior fue realizado con el fin de comprobar si las tareas planteadas se encontraban en una ventana menor a los 5 segundos de cada instancia. Para la clasificación se usó análisis de discriminación lineal que busca una línea que maximiza la separación entre clases y minimiza el traslape entre ellas. Los resultados muestran que la tarea 3 (*covert*) es la que mejor se pudo discriminar alcanzando para cada individuo 82.4 % 81.20 % 85.12 % y 74.72 % de precisión respectivamente.

En [10] se utilizaron las vocales *a* y *u* para mostrar la factibilidad de controlar una prótesis utilizando la imaginación de dichas vocales, y se obtuvieron resultados de 68 % y 78 % de exactitud para cada vocal. Este trabajo consistió de un framework que realiza el procedimiento de clasificación de las vocales en cuestión de segundos. Cabe mencionar, que se aprovechó la activación de las vocales seleccionadas *a* y *u* con diferentes músculos. Para este trabajo se usó Biosemi ActiView capturando 64 canales EEG con una frecuencia de muestreo de 2048 Hz. Se hicieron ventanas de 500 ms y se obtuvieron características con patrones espaciales comunes (*common spatial pattern* CSP) para clasificar con máquinas de soporte vectorial con una función de base radial.

En [8] se utilizó la DWT para extraer características y éstas usarlas como entrada para 4 clasificadores: Naive Bayes, Máquinas de soporte vectorial, Random Forest, y Perceptrón multicapa para una base de datos de 120 ejemplos de las palabras *arriba* y *abajo* obtenidas de 3 personas. Se usó el kit emotiv EPOC para obtener los datos, en este trabajo se estudiaron más a profundidad los canales cercanos al área de Wernicke, que corresponden a los canales F7, FC5, T7 y P7 en el sistema internacional 10-20.

Se discute que la clasificación de las palabras imaginadas es independiente del clasificador ya que no se muestran cambios significativos entre los clasificadores usados, sin embargo, los mejores resultados son con Random Forest. Además se da una aproximación de cómo se podrían mejorar los resultados, aplicando análisis de componentes independientes y filtrando los componentes mediante el exponente de Hurst.

Los trabajos antes mencionados hacen uso de habla imaginada, sin embargo, en [7] se incluyen expresiones faciales y audio. En dicho trabajo

participaron 4 mujeres y ocho hombres que pronunciaron e imaginaron las sílabas /iy//uw//piy//tiy//diy//m//n//pat//pot//knew//gnaw/. Se realizaron 132 ensayos, después de cada 40 ensayos los participantes descansaban. En esta investigación se calcularon medidas estadísticas para cada ventana: media, mediana, desviación estándar, varianza, máximo, mínimo, entropía espectral, energía, curtosis y oblicuidad. Para la clasificación emplearon máquinas de soporte vectorial y se combinó información tanto de EEG, como expresiones faciales y audio.

En este trabajo presentamos 2 métodos para la caracterización de la señal de EEG y con ello distinguir entre actividad e inactividad lingüística, sin importar la palabra imaginada. De esta forma, una vez identificado un segmento de actividad lingüística un proceso posterior identificará la palabra imaginada.

3. Método

Nuestro método consiste en 3 etapas: preprocesar los datos, extraer las características, y crear el clasificador. Dichas etapas se detallan en las secciones a continuación.

Cabe señalar que para realizar la clasificación se utilizó Random Forest pues ha mostrado lo mejores resultados [8]. Las características fueron calculadas usando la DWT bior 2.2 con diferentes niveles de descomposición y a cada uno de los niveles se le calcularon varios tipos de energía con el fin de compararlos. También se extrajeron otro tipo de características, basadas en información estadística sobre la señal cruda, retomando algunas de las presentadas en [7].

En general, el procedimiento seguido se muestra en la figura 1.



Fig. 1. Procedimiento general.

3.1. Pre-procesamiento

Esta etapa se refiere a la reducción y transformación de las bases de datos con el fin de tener datos adecuados con los cuales trabajar. Para pre-procesar las señales EEG se usa el método de referencia promedio común (CAR). Este método mejora la relación señal-ruido de la señal de EEG. El CAR se calcula

mediante la siguiente formula:

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n} \sum_{j=1}^n V_j^{ER}, \quad (1)$$

donde V_i^{ER} es el potencial entre el i -ésimo electrodo y la referencia, y n es el número de electrodos.

3.2. Extracción de características

Para obtener características se calcularon coeficientes de energía con DWT o medidas estadísticas. En ambos casos fueron calculadas las características para cada uno de los canales y posteriormente concatenados, formando así un sólo vector de características para cada instancia o época.

DWT Las señales de EEG generalmente son no estacionarias ya que cambian rápidamente con el tiempo. Además, estas características contienen información del tiempo debido a que los patrones de actividad cerebral están generalmente relacionados a variaciones específicas del EEG en el tiempo. Por esa razón se debe usar una representación que considere este comportamiento.

Como primera opción se aplicó la DWT con la función madre *bior 2.2* a cada uno de los canales, con diferentes niveles de descomposición. Al aplicar dicha transformada con un nivel de descomposición j se obtiene una estructura con vectores de coeficientes de aproximación (CA_j) y detalle (CD_j): $[CA_j, CD_j, CD_{j-1}, \dots, CD_1]$. Sin embargo, el número de coeficientes wavelet en cada nivel de descomposición variará dependiendo la duración de la ventana de la pronunciación de la palabra imaginada (tanto entre palabras de un mismo sujeto como entre palabras de distintos sujetos). Para tratar dicho problema se calculó la distribución de varios tipos de energía (Instantánea, Jerárquica, Relativa y Teager)[4] en cada nivel, de esta manera se logra normalizar todas las instancias de EEG. A continuación se muestran las formulas de los diferentes tipos de energía.

Sea $w_j(r)$ el coeficiente wavelet en el tiempo r y banda de frecuencia j . Si N es la longitud de la ventana de análisis, $w_j(r)$ tiene $N_j = N/2^j$ muestras en el nivel de descomposición j :

- Energía Instantánea: este tipo de energía refleja la amplitud de la señal.

$$f_j = \log_{10} \left(\frac{1}{N_j} \sum_{r=1}^{N_j} (w_j(r))^2 \right). \quad (2)$$

- Energía Jerárquica: provee una resolución jerárquica en tiempo y le da más importancia al centro de la ventana de análisis.

$$f_j = \log_{10} \left(\frac{1}{N_j} \sum_{r=(N_j-N_j)/2}^{(N_j+N_j)/2} (w_j(r))^2 \right), \quad (3)$$

donde además N_j es el número de muestras en el último nivel de descomposición.

- Energía Teager: refleja las variaciones tanto en amplitud como en frecuencia de la señal y es un parámetro robusto para el reconocimiento de habla ya que atenúa el ruido auditivo.

$$f_j = \log_{10} \left(\frac{1}{N_j} \sum_{r=1}^{N_j-1} |(w_j(r))^2 - w_j(r-1) * w_j(r+1)| \right). \quad (4)$$

- Energía Relativa: la distribución de energía puede ser considerada como una densidad tiempo-escala y provee información para caracterizar la distribución de energía de la señal en diferentes bandas de frecuencia.

$$f_j = \frac{E_j}{E_{total}}, \quad (5)$$

donde E_j representa la energía en el j -ésimo nivel de descomposición y E_{total} representa la energía total de los coeficientes wavelet de una señal dada. La energía en el j -ésimo nivel de descomposición E_j se define como:

$$E_j = \begin{cases} \sum_k |d_{j,k}|^2 & \text{Si } j \leq N, \\ \sum_k |a_k|^2 & \text{en caso contrario.} \end{cases} \quad (6)$$

Y la energía total E_{total} como sigue:

$$E_{total} = \sum_{j=1}^{N+1} E_j. \quad (7)$$

Estos coeficientes de energía fueron calculados para cada electrodo y concatenados para formar un vector de características que representan a la señal de EEG.

Estadísticas Otra opción para la caracterización de la señal fue motivada por lo planteado en [7]. En dicho trabajo se caracterizó la señal al calcular un conjunto de valores estadísticos. Inicialmente tomamos 15 características estadísticas por canal. Es decir, para cada uno de los canales se calcularon las 15 características y luego se concatenaron, para tener un vector de características por cada instancia. Como se verá más adelante se usaron dos conjuntos de valores estadísticos.

3.3. Creación del clasificador

Una vez obtenidos los vectores de características para cada muestra de EEG se utilizó un algoritmo de clasificación automática para calcular un modelo que distinga entre actividad e inactividad lingüística. Para ello se utilizó el algoritmo de *Random Forest*. Este algoritmo fue seleccionado debido a los resultados reportados en trabajos relacionados. Para evaluar el rendimiento del clasificador se usó la exactitud.

4. Experimentos y resultados

El objetivo de esta sección es mostrar los resultados al experimentar con los dos conjuntos de características diferentes para clasificar en actividad e inactividad. Cabe recordar que la clase de actividad lingüística está conformada por las instancias de las palabras imaginadas y la inactividad esta formada por pausas o reposos. A continuación se detallan las bases de datos usadas, para posteriormente mostrar los resultados alcanzados al utilizar diferentes tipos de caracterizaciones.

4.1. Datos

Para los experimentos se utilizaron dos bases de datos. La primera base de datos (BD1) contiene las señales de EEG de 27 sujetos hablantes nativos del idioma español registradas usando EMOTIV EPOC mientras imaginaban 33 repeticiones de 5 palabras (*arriba*, *abajo*, *izquierda*, *derecha* y *seleccionar*) en idioma español, separadas por un estado de reposo, tal como se muestra en la figura 2 y se describe en [1]. Para cada palabra imaginada se obtuvo y guardo información de 14 canales de EEG con una frecuencia de muestreo de 128 Hz organizados de acuerdo al sistema internacional 10-20.

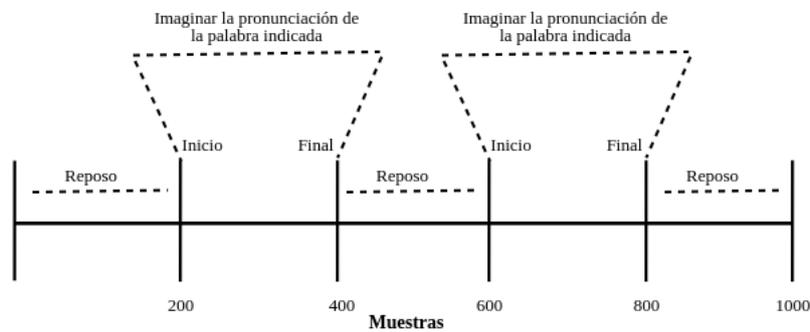


Fig. 2. Diseño del protocolo para adquisición de la señal de EEG para la BD1 usando EMOTIV EPOC.

Para la BD1 la clase actividad contiene 165 instancias que corresponden a las 33 repeticiones de cada una de las 5 palabras. La clase inactividad contiene 165 instancias que corresponden a los estados de reposo que existen entre una repetición de una palabra y otra. Eso para cada uno de los 27 sujetos.

La segunda base de datos (BD2) contiene las señales de EEG de 20 sujetos hablantes nativos del idioma español que fueron grabados mientras imaginaban 40 repeticiones de 4 palabras (*arriba*, *abajo*, *derecha*, *izquierda*) separadas por una pausa, así como se describe en [1].

El protocolo de estimulación y adquisición de las señales de EEG de la BD2 se muestra en la figura 3, donde RN es el marcador final para $N = 1, 2, 3, 4$ correspondientes a las 4 palabras imaginadas: $R1$ para la palabra imaginada *arriba*, $R2$ para *abajo*, $R3$ para *derecha* y $R4$ para *izquierda*. El marcador $\{R13$ o $R14$ o $R15\}$ es el final de la respuesta por teclado que indica el momento en que el sujeto presionaba una flecha en el teclado y dicha flecha debería coincidir con la palabra que se solicitó imaginar. El tiempo máximo de espera para imaginar la palabra fue de $1,5s$, a menos que el sujeto presionara la tecla antes. Además $\{R13$ o $R14$ o $R15\}$ es el inicio de una pausa terminando en el marcador $S4$.

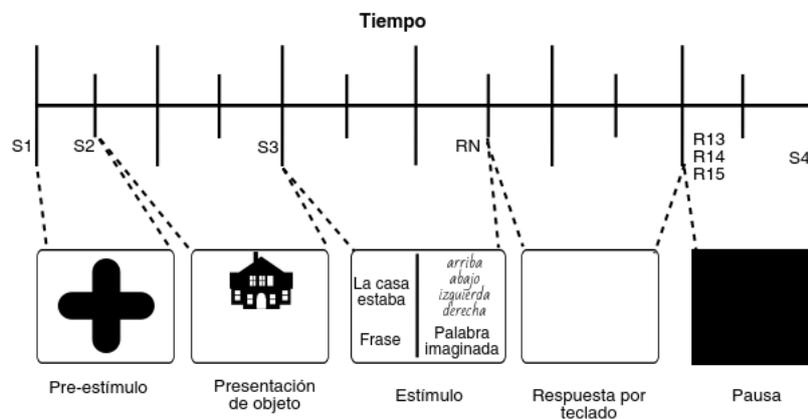


Fig. 3. Diseño del protocolo para estimulación y adquisición de la señal de EEG para la BD2 usando Brain Vision.

Para cada palabra imaginada se guardó información de 64 canales de EEG con una frecuencia de muestreo de 500 Hz. Para la BD2 la clase actividad contiene 160 instancias que corresponden a las 40 repeticiones de cada una de las 4 palabras imaginadas (la frase que precede a la palabra imaginada también se consideró como actividad) entre los marcadores $S3$ a RN según la figura 3 y la clase inactividad contiene 160 muestras que corresponden a 40 instancias de el pre-estímulo, 40 instancias de la presentación del objeto, 40 instancias de la respuesta por teclado y finalmente 40 instancias de la pausa (ver figura 3).

La tabla 1 presenta una descripción de las principales características de las bases de datos utilizadas.

4.2. Características usando DWT

Las características calculadas con DWT bior2.2 fueron realizadas con distintos niveles de descomposición. Para describir la descomposición se parte de conocer la frecuencia de muestreo (fN) que es de 128 Hz para la BD1 y 500 Hz

Tabla 1. Comparación de las características de las bases de datos usadas.

	BD1	BD2
Num. sujetos	27	20
Num. palabras imaginadas	5	4
Num. repeticiones por palabra	33	40
Canales de EEG	14	64
Frecuencia de muestreo	128 Hz	500 Hz

para la BD2. Luego, el teorema de Nyquist dice que la frecuencia de muestreo debe ser por lo menos 2 veces mayor a la frecuencia de la señal a muestrear, es decir: $f_N \geq 2f_s$, donde f_s es la frecuencia de la señal muestreada. Despejando f_s , se obtiene: $f_s \approx \frac{f_N}{2}$. Por lo tanto para la BD1 se tiene que $f_s \approx 64Hz$. y para la BD2 $f_s \approx 250Hz$.

Por otro lado, también es necesario observar el tamaño de las muestras para determinar el máximo número de niveles de descomposición que pueden ser calculados. La tabla 2 muestra los posibles niveles de descomposición para cada base de datos.

Tabla 2. Rangos de frecuencias(Hz) para el nivel de descomposición 4 de DWT para la BD1 y los niveles de descomposición 5 y 6 para la BD2.

BD1		BD2			
Nivel	Rango	Nivel	Rango	Nivel	Rango
D1	32-64	D1	125-250	D1	125-250
D2	16-32	D2	62.5-125	D2	62.5-125
D3	8-16	D3	31.25-62.5	D3	31.25-62.5
D4	4-8	D4	15.62-31.25	D4	15.62-31.25
A4	0-4	D5	7.81-15.62	D5	7.81-15.62
		A5	0-7.81	D6	3.9-7.81
				A6	0-3.9

Para seleccionar el conjunto de coeficientes DWT más adecuado para la señal de EEG se consideró lo planteado en [6,11] donde menciona que las frecuencias de EEG llegan hasta 60 Hz y más allá de dichas frecuencias no tienen interés para nuestro análisis. Tomando en cuenta lo anterior se realizó la caracterización para los 27 sujetos de la BD1 y los 20 sujetos de la BD2 con la DWT bior 2.2 con 4 niveles de descomposición para la BD1 y con 5 y 6 para la BD2 para luego eliminar los coeficientes de detalle correspondientes a los rangos mayores a 60 Hz.

Los resultados promedio se muestran en la tabla 3. Para la BD1 sólo se pueden calcular 4 niveles de descomposición pues la frecuencia de muestreo es de 128 Hz. Para la BD2 la frecuencia de muestreo es de 500 Hz, por lo tanto

Tabla 3. Exactitud y desviación estándar del clasificador Random Forest a 10 pliegues para DWT bior2.2 con 4 niveles de descomposición usando la BD1 al calcular diferentes tipos de energía.

Nivel	Num. Características	Instantánea	Jerárquica	Relativa	Teager
4	70, 5 por canal	0.75 ±0.15	0.73 ±0.14	0.73 ±0.16	0.78 ±0.16

se puede tener un mayor número de niveles de descomposición. Si se calculan niveles de descomposición menores a 5 sólo se obtienen 2 características por canal después de eliminar los rangos más allá de 60 Hz, es por eso que se deben calcular las características con niveles de descomposición mayores a 5. Además se debe considerar que el máximo nivel de descomposición de esta base de datos es 6. A continuación se muestran los resultados promedio de exactitud para los niveles de descomposición 5 y 6 para la BD2.

Tabla 4. Exactitud y desviación estándar del clasificador Random Forest a 10 pliegues para DWT bior2.2 con 5 y 6 niveles de descomposición usando la BD2 al calcular diferentes tipos de energía. El sub-índice en el Nivel indica los niveles de detalle eliminados.

Nivel	Num. Características	Instantánea	Jerárquica	Relativa	Teager
5 ₂	256, 4 por canal	0.90 ±0.02	0.90 ±0.02	0.89 ±0.03	0.91 ±0.03
6 ₂	320, 5 por canal	0.90 ±0.02	0.90 ±0.01	0.89 ±0.03	0.92 ±0.03

En la tabla 4, los sub-índices $n = 2$ para los niveles de descomposición indican que se han eliminado los primeros n coeficientes de detalle para evitar las frecuencias más allá de 60 Hz. Como se puede observar en la tabla 3 y 4 la caracterización usando la energía Teager es la que presenta mejores resultados.

4.3. Características estadísticas

Las características utilizadas para este experimento constan de 2 conjuntos: el primero es de 15 valores estadísticos y el segundo de 9. El primer conjunto(C1) son las características usadas por [7] que consta de 9 valores estadísticos y 6 combinaciones de algunos de ellos. En el segundo(C2) conjunto sólo usamos los 9 valores sin considerar las combinaciones, como se muestra a continuación.

- C1: media, media⁶, absoluto de la media, máximo, absoluto del máximo, mínimo, absoluto del mínimo, máximo+mínimo, máximo-mínimo, desviación estándar, varianza, curtoris, oblicuidad, suma y mediana.
- C2: media, máximo, mínimo, desviación estándar, varianza, curtoris, oblicuidad, suma y mediana.

En la tabla 5 se muestran los resultados de las 2 base de datos al usar los valores estadísticos como características. En los 2 casos se obtienen los mejores resultados al usar 9 características por canal.

Tabla 5. Exactitud y desviación estándar del clasificador random forest a 10 pliegues para características estadísticas usando las 2 bases de datos.

BD	Características	Num. Características	Exactitud
BD1	C1	210, 15 por canal	0.82 ±0.18
BD1	C2	126, 9 por canal	0.83±0.17
BD2	C1	960, 15 por canal	0.91 ±0.05
BD2	C2	576, 9 por canal	0.91±0.05

Como puede observarse en los resultados reportados, la caracterización usando el conjunto de 9 características(C2) son mejores o iguales que los obtenidos al usar la DWT. En la tabla 6 se muestra el comparativo de los resultados alcanzados para los 27 sujetos de la BD1 con los dos tipos de caracterización (utilizando DWT y características estadísticas). Utilizando DWT con la función madre bior 2.2, los mejores resultados se obtuvieron con 4 niveles de descomposición y calculando la energía Teager a cada uno de los niveles. Para las características estadísticas los mejores resultados se obtuvieron al usar el conjunto de 9 características(C2) por canal.

Tabla 6. Exactitud, media y desviación estándar (Std) del clasificador random forest a 10 pliegues para características DWT calculando la energía Teager y estadísticas(EST) para la BD1.

Sujeto	DWT	EST	Sujeto	DWT	EST	Sujeto	DWT	EST
1	0.76	0.83	10	0.98	1.00	19	0.99	1.00
2	0.90	0.99	11	0.64	0.74	20	0.81	0.93
3	0.67	0.66	12	0.77	0.84	21	0.88	0.94
4	0.74	0.79	13	0.46	0.57	22	0.93	0.98
5	0.97	1.00	14	0.57	0.63	23	0.57	0.53
6	0.90	0.95	15	0.90	0.96	24	0.86	0.92
7	0.92	0.91	16	0.75	0.76	25	0.98	1.00
8	0.90	0.93	17	0.52	0.52	26	0.77	0.87
9	0.66	0.77	18	0.48	0.42	27	0.85	0.97
Media							0.78	0.83
Std							± 0.16	± 0.17

En la tabla 7 se muestran los mejores resultados para los 20 sujetos de la BD2 tanto usando DWT como características estadísticas. Para el caso de DWT se usaron 6 niveles de descomposición, se eliminaron los primeros 2 coeficientes

de detalle y se calculó la energía Teager para los restantes 5 coeficientes (D3, D4, D5, D6 y A6). Esto se repitió por cada canal. Con las características estadísticas se utilizó el conjunto de características C2.

Tabla 7. Exactitud, media y desviación estándar (Std) del clasificador Random Forest a 10 pliegues para características DWT calculando la energía Teager y estadísticas(EST) para la BD2.

Sujeto	DWT	EST	Sujeto	DWT	EST	Sujeto	DWT	EST
1	0.99	1.00	8	0.95	0.98	15	0.91	0.88
2	0.87	0.86	9	0.92	0.92	16	0.89	0.88
3	0.88	0.88	10	0.92	0.92	17	0.94	0.94
4	0.90	0.84	11	0.93	0.93	18	0.94	0.96
5	0.88	0.88	12	0.89	0.89	19	0.93	0.96
6	0.89	0.84	13	0.92	0.93	20	0.96	0.94
7	0.88	0.88	14	0.94	0.96			
Media							0.92	0.91
Std							± 0.03	± 0.05

5. Conclusiones y trabajo futuro

El reconocimiento de habla imaginada, a partir de señales de EEG, ha sido estudiado en sistemas fuera de línea como una tarea de clasificación de distintas palabras de un vocabulario propuesto. En este trabajo se presentó un estudio para identificar la actividad e inactividad lingüística con el fin de dar el primer paso hacia el reconocimiento en línea. Los experimentos realizados a pesar de también llevarse a cabo fuera de línea dan evidencia de la factibilidad de la tarea. Se analizó si, independientemente de la palabra imaginada (actividad lingüística) es posible diferenciarse de otras actividades cerebrales (inactividad lingüística).

El método propuesto fue evaluado con dos bases de datos, con diferente protocolo de registro y diferentes instrumentos. Los resultados son muy alentadores, indicando que es posible distinguir entre actividad o inactividad lingüística. Los resultados alcanzados están muy por arriba de la cota del azar, lo que permite conjeturar que es posible crear un proceso en línea que logre clasificar una señal como actividad o inactividad y posteriormente con otro clasificador identificar la palabra imaginada que finalmente se convertirá en un comando para fines de comunicación o control de dispositivos.

Por otra parte, este trabajo evaluó si el tipo de caracterización de las señales de EEG tiene algún efecto en la discriminación de actividad e inactividad lingüística. Las caracterizaciones estudiadas fueron características estadísticas obtenidas desde las señales en tiempo y energías calculadas a partir de la DWT. Usando los vectores estadísticos se obtuvieron resultados iguales o superiores

que con las energías wavelet; con la ventaja de que éstos no requieren de una transformación a otro dominio, ni definición de parámetros adicionales a la señal.

Al comparar los resultados de las tablas 6 y 7 se puede notar que siguiendo el mismo procedimiento en ambas bases de datos se obtiene mayor exactitud en la BD2. Son varios factores por los cuales la exactitud es mayor para la BD2. Como puede verse en la tabla 1 la BD2 comprende un número mayor de repeticiones de cada palabra, una frecuencia de muestreo mayor y se cuenta con información de un mayor número de canales (64 canales).

Para comprobar los resultados obtenidos aquí, como trabajo futuro se planea probar otras bases de datos con protocolos de adquisición y frecuencia de muestreo diferentes. Además, se plantea dividir la señal de EEG y comprobar si aún existe información suficiente para distinguir entre actividad e inactividad lingüística. De esta manera se podrá identificar en un menor tiempo el inicio y fin de la actividad lingüística. Es así que será posible orientar los trabajos de reconocimiento de habla imaginada a tiempo real.

Agradecimientos. Los autores agradecen al Consejo Nacional de Ciencia y Tecnología (CONACyT) por al apoyo a esta investigación, el cual se realizó a través de la beca No. 591475 y del proyecto No. Ref. 2016-01-2228. Los autores también desean agradecer el apoyo de la Red Temática en Tecnologías del Lenguaje/CONACYT, No. Ref. 260178, 271622, por los mecanismos de colaboración brindados para realizar este trabajo.

Referencias

1. Torres-García, A. A.: Análisis y clasificación de electroencefalogramas (EEG) registrados durante el habla imaginada. Tesis de doctorado, Instituto Nacional de Astrofísica, Óptica y Electrónica, México (2016)
2. Bashashati, A., Fatourehchi, M., Ward, R. K., Birch, G. E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, vol. 4 (2007)
3. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T. M.: Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, vol. 113, pp. 767–791 (2002)
4. Didiot, E., Illina, I., Fohr, D., Mella, O.: A Wavelet-Based Parameterization for Speech/Music Discrimination. vol. 24, pp. 341, Vandoeuvre-les-Nancy, France (2009)
5. Desain, P., Farquhar, J., Haselager, P., Hesse, C., Schaefer, C.: What BCI research needs. In: *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, vol. 55, pp. 181–184, Florence, Italy (2008)
6. Brigham, K., Vijaya Kumar, B.V.K.: Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy. In: *Bioinformatics and Biomedical Engineering (iCBBE)*, 2010 4th International Conference on, IEEE, pp. 1–4. Pittsburgh PA, USA (2010)
7. Zhao, S., Rudzicz, F.: Classifying Phonological Categories in Imagined and Articulated Speech. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 22, pp. 992–996, Toronto, Canada (2015)

8. Torres-García, A. A., Reyes-García, C. A., Villaseñor-Pineda, L.: Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada. *Revista Mexicana de Ingeniería Biomédica*, vol. 34, pp. 23–29 (2013)
9. Song, Y., Sepulveda, F.: Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state. In: *IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, vol. 55, pp. 568–571, Essex, United Kingdom (2014)
10. Kambara, H., Sato, M., DaSalla, C. S., Koike, Y.: Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, vol. 22, pp. 1334–1339 (2009)
11. Porbadnigk, A., Schultz, T.: Eegbased speech recognition: Impact of experimental design on performance. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe, Germany (2008)

Desarrollo de una BCI utilizando el potencial P300 y la diadema Mindwave®

Sergio René Vivar Vera¹, María Antonieta Abud Figueroa¹,
Gabriel Rodríguez Vásquez², Ulises Juárez Martínez¹, Gustavo Peláez Camarena¹,
Jorge Airy Mercado Gutiérrez³, Josefina Gutiérrez Martínez³

¹ Instituto Tecnológico de Orizaba, Veracruz, |
México

² Asti Consultoría de Software S.A. de C.V., Veracruz,
México

³ Instituto Nacional de Rehabilitación, Ciudad de México,
México

svivar@acm.org, {mabud, ujuarez, gpelaez}@ito-depi.edu.mxdireccion@astiservices.com,
jmercado@inr.gob.mx, jgutierrez@inr.gob.mx

Resumen. En este trabajo se presenta un esquema de registro, procesamiento y clasificación para una Interfaz Cerebro-Computadora de deletreo, basada en el potencial P300 y el dispositivo de electroencefalografía (EEG) Mindwave-NeuroSky®; este dispositivo cuenta con un solo canal de registro de EEG y es el más económico en el mercado en la actualidad. El registro de la señal se llevó a cabo con el software libre OpenVibe, para el análisis de la señal se crearon *scripts* en Matlab y se utilizó el algoritmo de clasificación Análisis Discriminante Lineal (*Linear Discriminant Analysis*, LDA). Cinco sujetos de prueba realizaron 3 sesiones de deletreo cada uno, donde se analizaron dos cantidades de repeticiones: 8 y 12. En cuanto a los tipos de secuencia de caracteres a deletrear, se establecieron dos: al azar y predefinida; y se analizaron dos zonas de colocación del electrodo: Oz y Pz. Los sujetos lograron, en total, el 45 % de exactitud de reconocimientos de renglones y columnas objetivo. En un trabajo similar, con registros de 64 canales, se obtuvo el 90.4 % de exactitud contra el 41.9 % obtenido en el presente, tomando en cuenta sólo la secuencia predefinida. La diadema MindWave-NeuroSky® no fue diseñada para utilizarse en BCI; a pesar de ello y en vista de los resultados obtenidos, su utilización en esta tecnología es factible si se combina con algún esquema de predicción de palabras basándose en el contexto de la frase deletreada u otros que ayuden en la construcción de frases correctas.

Palabras clave: interfaz cerebro-computadora, OpenViBE; MindWave NeuroSky, dispositivo de bajo costo, clasificación, deletreo.

Development of a BCI Using P300 Potential and MindWave® Headset

Abstract. In this paper we propose a recording, processing and classification scheme for a spelling Brain-Computer Interface application based on the P300 potential and the electroencephalography (EEG) device Mindwave-NeuroSky®; this device has only one EEG recording channel and is currently the most affordable in the market for this purpose. Open access software Openvibe was chosen for signal recording, and Matlab scripts were written for signal analysis and classification, the latter being based on Linear Discriminant Analysis (LDA). Five test subjects were exposed to several spelling sessions each, in which different acquisition and stimulation parameters were varied to evaluate the effects in character identification. Those parameters were the number of repetitions of the stimulation sequence of the speller character matrix ($r=8, 12$), the type of character sequence (chosen randomly or beforehand), and the EEG recording site (Pz or Oz). The subjects achieved, an average performance of 45% correct row and column identifications. In comparison, a very recent work in which 64 EEG channels were acquired, and a complex classification and feature extraction scheme was used, they achieved an average of 90.4% correct character identifications. The MindWave-Neurosky® headset was not designed for application in BCI systems. However, by looking at the results in this work, its application for BCI it seems feasible when thinking of combining the proposed approach with some word prediction scheme based on the context of the spelled phrase.

Keywords: brain-computer interface, OpenViBE; MindWave NeuroSky, low cost device, classification, spelling.

1. Introducción

Una Interfaz Cerebro-Computadora (BCI, del inglés Brain-Computer Interface) es un canal de comunicación entre el cerebro de una persona y un dispositivo de procesamiento y retroalimentación, tal como una computadora, un sistema embebido o un dispositivo mecatrónico [1]; estos sistemas son muy útiles para personas que son incapaces de mover gran parte de su cuerpo[2]; tienen utilidad en otras áreas como la navegación, la domótica, entre otros [1, 3].

La comunicación con el cerebro se logra registrando las señales eléctricas en el rango de los microvolts (μV) generadas por la actividad cerebral; con estos datos, es posible el desarrollo de aplicaciones y dispositivos que sirvan para interactuar con el usuario de una manera práctica [1].

EEG es una técnica que permite el registro de la actividad cerebral mediante electrodos y es adecuada para detectar la respuesta cerebral a los estímulos [4-6].

El hardware de EEG, tiene un alto costo monetario, entre US \$1,000 y US \$25,000; los trabajos relacionados a BCI requieren un gran presupuesto. Por otro lado, existen

lectores EEG de bajo costo, como la diadema MindWave-NeuroSky® [7]; sin embargo, son pocos los estudios que abordan el análisis con este dispositivo.

El potencial P300 se utiliza ampliamente en BCI[3,4]. El P300 se genera cuando el individuo recibe un estímulo de interés poco frecuente entre una serie de estímulos irrelevantes, pero más frecuentes (paradigma evento raro, *oddball*); lo que permite averiguar el objeto de atención del sujeto.

El *software* OpenViBE [8] es una plataforma que sirve para el desarrollo de experimentos y aplicaciones BCI; es libre y de código abierto y tiene integrada una aplicación (similar al utilizado por Farwell y Donchin [9]) que muestra una matriz de caracteres de 6 renglones x 6 columnas que estimula secuencialmente (estimulación es la iluminación o intensificación simultánea de los 6 caracteres contenidos en una misma fila o columna) de acuerdo con el paradigma de evento raro (Figura 1).

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	0

Fig. 1. Interfaz gráfica de deletreador para la etapa de adquisición.

El análisis de la señal EEG se desarrolló mediante algoritmos en Matlab®, una plataforma de software que proporciona un Entorno de Desarrollo Integrado (IDE) y un lenguaje de programación propio para el análisis matemático

En este trabajo se nombrará *reconocimiento* a la determinación correcta del renglón o columna en donde se encuentra ubicado el carácter en el que el usuario centra su atención (carácter objetivo) y *deletreo correcto* a la determinación correcta de un carácter objetivo, habiendo reconocido su renglón y columna de ubicación.

En la Sección 2 se mencionan trabajos previos relacionados en el área de las BCI; en la Sección 3 se detallan los dispositivos utilizados, se mencionan las características generales de los sujetos de prueba, se especifica cómo será la evaluación de los resultados y se describen las sesiones de deletreo y los experimentos para llevar a cabo los reconocimientos. En la Sección 4 se presentan algunas conclusiones. Finalmente, en la Sección 5 se proponen actividades para darle continuidad al presente trabajo.

2. Estado del arte

A continuación, se presenta la revisión y comparación de algunos trabajos con BCI.

En 2012, Pires y cols. [10] desarrollaron una BCI para el deletreo basado en renglones y columnas, y otra que presenta los caracteres individualmente. La exactitud de deletreo fue ligeramente superior en la aplicación basada en renglones y columnas.

En 2013, Poli y cols. [11] experimentaron con un apuntador que consta de 8 círculos en el centro de la pantalla que se iluminan secuencialmente, el cual utilizaron para controlar un apuntador en la pantalla de un simulador. Los usuarios obtuvieron el 66 % de exactitud de movimientos.

En 2015, Wang y cols. [12] utilizaron un deletreador de P300 y Potenciales Evocados Visuales de Estado Estable (Steady-State Visual Evoked Potentials, SSVEP) al cambiar la forma y el color de iluminación de los objetivos. Concluyeron que este paradigma híbrido logra mejores resultados que otros paradigmas híbridos basados únicamente en el cambio de color de los elementos objetivo.

En 2015, Suryotrisongko y cols. [13] llevaron a cabo un análisis de la utilización de electrodos OpenBCI Spiderclaw VI [14] en una aplicación BCI de deletreo basada en Ritmos Sensorial Motores (SMR, Sensory Motor Rhythms). Los autores concluyeron que estos electrodos no son óptimos para aplicaciones basadas en SMR.

Un nuevo enfoque, llamado diseño centrado en el usuario, se presentó en 2015 por Schreuder y cols. [15], en el cual se tomaron en cuenta las deficiencias propias del paciente para llevar a cabo dos sistemas BCI, uno basado en estímulos auditivos y otro en estímulos visuales, ambos mediante el paradigma P300.

Otra aplicación BCI de deletreo basado en un paradigma híbrido SSVEP y P300 se presentó en 2016 por Chang y cols. [16] con el cual se obtuvo mayor exactitud que con los paradigmas SSVEP y P300 por separado y utilizando estímulos que se iluminan a una frecuencia menor que los necesarios para la técnica SSVEP.

En 2017, Yoon y cols. [17] utilizaron, en un deletreador BCI P300, un nuevo algoritmo de clasificación, el Multiple Kernel Learning (MKL) con tres tipos de características discriminantes (señal en crudo raw, amplitud y parte negativa de la señal), Los registros de señal utilizados en esta BCI fueron tomados del *Dataset II of BCI competition III*, que consiste en registros EEG de 64 canales. El estudio demostró que la característica discriminante raw tiene un mayor peso que las características amplitud y parte negativa.

De la revisión bibliográfica se observó que, con la finalidad de mejorar el desempeño de los deletreadores BCI, se han utilizado diferentes algoritmos e interfaces de retroalimentación y se han combinado diferentes paradigmas de BCI.

3. Propuesta

Se presenta el desarrollo de un deletreador BCI basado en el ERP (Event-Related Potential, Potencial Relacionado a Eventos) P300. Se utilizó la diadema MindWave NeuroSky® como hardware de EEG; OpenViBE como *software* de adquisición e interfaz de retroalimentación al usuario; y una implementación, desarrollada en Matlab® para este trabajo, como software de análisis y reconocimiento.

Se utilizó solamente un algoritmo de clasificación ya que se pretende conocer una línea base de desempeño con un solo electrodo, más que una comparativa entre clasificadores. La solución presentada en este trabajo no representa por sí misma una solución de bajo costo, adjetivo que se asigna únicamente a la diadema utilizada, la cual es el dispositivo comercial de EEG de menor precio en la actualidad.

3.1. Dispositivos utilizados

La diadema MindWave MW001 NeuroSky® cuenta con un único electrodo, el cual es posible fijar en la posición de interés con ligera presión sobre el cráneo por la flexibilidad del brazo que lo soporta (Figura 2).



Fig. 2. Diadema MindWave MW001 NeuroSky®.

En el clip de la oreja se encuentran la referencia y la parte negativa del electrodo, de esta manera se logra una colocación rápida del dispositivo para su uso.

El equipo de cómputo utilizado es portátil, con la pantalla a una distancia de 60 cm. del sujeto. La computadora se desconectó de la corriente eléctrica funcionando únicamente con batería, para reducir la contaminación de la señal por ruido de línea.

3.2. Posiciones del electrodo

El diseño de la diadema determina presupone la zona frontal del cráneo como el lugar de la colocación del electrodo; sin embargo, como los potenciales P300 se registran con mayor amplitud en la zona parietal y occipital, la diadema se utilizó en una posición invertida para alcanzar las posiciones Oz y Pz del Sistema Internacional 10-20 (Figura 3).

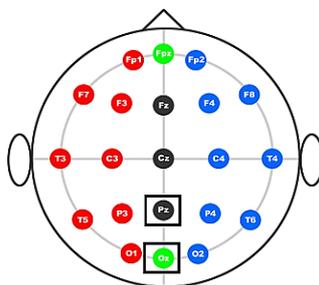


Fig. 3. Sistema Internacional de posicionamiento de electrodos 10-20. Se encuentran remarcados los puntos Pz y Oz que se utilizaron en los experimentos.

3.3. Valoración de los resultados

Para deletrear correctamente un carácter se requiere reconocer el renglón y la columna de la interfaz gráfica en los que se encuentra ubicada, por lo tanto, se asignó un puntaje de 1 acierto por cada renglón o columna reconocidos. Cada sesión constó de 8 caracteres, lo que indica un total de aciertos posibles de 16.

Sujetos de prueba. Las personas que participaron en las sesiones de este trabajo fueron:

- Sujeto 1: Masculino de 14 años de edad.
- Sujeto 2: Femenino de 15 años de edad.
- Sujeto 3: Femenino de 22 años de edad.
- Sujeto 4: Femenino de 40 años de edad.
- Sujeto 5: Masculino de 41 años de edad.

Todos ellos son sujetos sanos sin limitación cognitiva alguna. Para que los experimentos no se vieran afectados por el cansancio, se programaron para llevarse a cabo después de sus horas de descanso.

3.4. Sesiones de deletreo y registro de la señal

Las sesiones de deletreo tuvieron por objetivo la generación del potencial P300 en los sujetos de prueba mediante el deletreador, así como el almacenamiento del registro EEG en archivos CSV.

En la Figura 4 se muestra el escenario de OpenViBE utilizado, que consta de dos partes: el deletreador y el escritor del registro EEG, que se ejecutan simultáneamente.

Los parámetros de operación del deletreador se establecieron como muestra la Tabla 1. Con el retardo en la primera estimulación se trató de lograr una señal con línea base estable. El espacio de tiempo entre caracteres fue para que la persona descansara sus ojos y reacomodara su postura.

Tabla 1. Valores de configuración del deletreador.

Parámetro	Valor
Retardo de la primera estimulación	70 s
No. de repeticiones	{8, 12}
Cantidad de caracteres	8
Duración de la estimulación	0.18 s
Intervalo inter-estímulo	0.07 s
Intervalo entre repeticiones	1 s
Intervalo entre caracteres	6 s

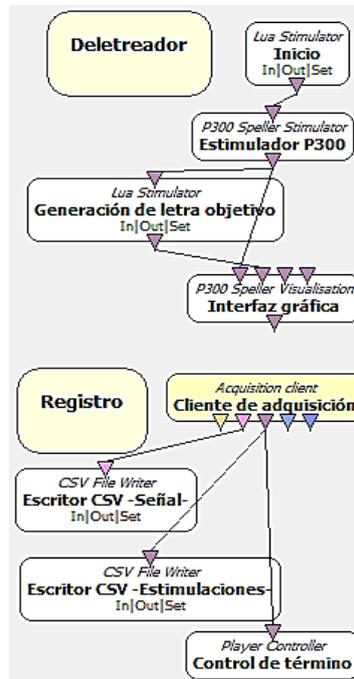


Fig. 4. Escenario de OpenViBE utilizado para la adquisición de la señal.

El bloque escritor de archivos CSV creó dos archivos por sesión, uno de ellos contiene las muestras (valores de amplitud en unidades de μV entregados por la diadema), que componen la señal de EEG así como los tiempos de ocurrencia de cada una y el valor de la frecuencia de muestreo del hardware. Otro archivo generado contiene información relativa a los eventos de estimulación ocurridos durante la sesión, tal como el tiempo de ocurrencia, la indicación del renglón o la columna estimulados, el renglón y la columna en los que se encuentra el carácter objetivo (renglón y columna objetivos). La estimulación de un renglón o columna objetivos se le llama *estimulación objetivo*.

Descripción de las sesiones de deletreo. Se mostró al sujeto la interfaz del deletreador en la pantalla, centrando el sujeto su atención en un carácter individual del deletreador (carácter objetivo). Posteriormente, las columnas y filas se estimularon aleatoria y alternadamente hasta que cada una (12) se estimuló una vez formando un ciclo denominado *repetición*; este ciclo se repitió r veces (cantidad de repeticiones). Posteriormente, el sujeto coloca su atención en el siguiente carácter de la secuencia de caracteres, repitiéndose la secuencia anterior. Las sesiones constaron de 8 caracteres. Se analizaron valores de $r = 12$ y $r = 8$.

Para cada posición del electrodo (Oz y Pz), se llevaron a cabo tres sesiones de 8 repeticiones y tres de 12 repeticiones, en total 12 sesiones por cada sujeto.

La secuencia de caracteres de las dos primeras sesiones se eligió al azar por el deletreador; la secuencia de la tercera sesión fue la cadena "P E R S O N A 9". Se eligió

la cadena mencionada ya que contiene ocho caracteres, no contiene caracteres repetidos, contiene al menos un número y, exceptuando ese número, la cadena es una palabra común.

El sujeto permaneció pendiente de las estimulaciones del carácter objetivo y al ocurrir tal evento, repitió mentalmente la palabra “ya”; esto tuvo como finalidad evitar la distracción de la atención del sujeto durante el experimento.

En las sesiones de deletreo predefinido, no se le proporcionó al sujeto ninguna ayuda para recordar el carácter que continúa dentro de la cadena mencionada, la persona controló por sí misma el orden de los caracteres.

3.5. Análisis de la señal y reconocimientos

Los registros generados en las tres sesiones se utilizaron para llevar a cabo cuatro experimentos de entrenamiento-reconocimiento mediante algoritmos implementados en Matlab® con el objetivo de determinar, a partir de señales EEG, los deletreos correctos y obtener parámetros de desempeño relacionados:

1. Entrenamiento con registro 1 y reconocimiento con registro 2 (E1 - R2).
2. Entrenamiento con registro 2 y reconocimiento con registro 1 (E2 - R1).
3. Entrenamiento con registro 1 y reconocimiento con registro 3 (E1 - R3).
4. Entrenamiento con registro 2 y reconocimiento con registro 3 (E2 - R3).

El registro 3 no contiene la identificación carácter objetivo ya que no son determinados por el deletreador, por lo tanto, no es posible utilizar este registro como entrenamiento.

Para poder evaluar el efecto de los distintos valores de r en la cantidad de reconocimientos, en cada experimento se utilizaron registros con el mismo valor r , es decir, los registros de entrenamiento de 12 repeticiones se utilizaron únicamente con los registros de reconocimiento de 12 repeticiones y de modo equivalente para aquellos de $r = 8$.

El procesamiento de los registros de EEG se dividió en dos etapas: entrenamiento y reconocimiento. En la primera se entrenó un clasificador binario de épocas individuales de EEG, las clases que se generan fueron “objetivo” y “no-objetivo”. En la segunda etapa se utilizó el clasificador generado en la primera para clasificar los datos de otra sesión de registro distinta (datos no observados), para determinar el renglón y columna objetivos. En la primera etapa los fragmentos de registro (épocas) se generaron basándose en estimulaciones objetivo y no-objetivo y en el segundo tratamiento, las épocas de generaron por estimulaciones de renglones y columnas (Figura 6).

El proceso de extracción de épocas se explica detalladamente al final de esta subsección.

Entrenamiento. Este proceso tiene como finalidad el entrenamiento del clasificador, esto es, la generación de pesos y sesgos que maximicen la clasificación de épocas individuales de EEG (Figura 6a).

Registro de entrenamiento. Los registros utilizados para el entrenamiento del clasificador en los experimentos 1 y 3 fueron los de la sesión 1; para los experimentos 2 y 4 fue el registro de la sesión 2.

Filtrado. Se aplicó un filtro Butterworth de orden 4, con banda de paso de .15 a 15 Hz, ya que esta configuración ofrece balance entre reducción de ruido y conservación de la información para un delectador P300 [18].

Decimación. El factor de decimación (reducción de la frecuencia de muestreo del dispositivo) se estableció en 8, ya que la frecuencia de muestreo de la diadema es de 512 Hz, quedando la frecuencia efectiva en 64 Hz. Esta frecuencia final de muestreo es mayor que dos veces la frecuencia máxima que se recibe de la señal cerebral (15 Hz), por lo que cumple con el criterio del muestreo de Nyquist.

Extracción de épocas objetivo y no-objetivo. Las épocas se generaron con base en la estimulación: estimulación objetivo, o no-objetivo.

Creación de las matrices de clase. Los dos tipos de épocas se colocaron en matrices diferentes (matriz objetivo y matriz no-objetivo).

Balanceo de clases. Las matrices no-objetivo generadas contuvieron cinco veces más épocas que las matrices objetivo, debido a que por cada estimulación de un renglón o columna objetivo existieron cinco estimulaciones no-objetivo. Es necesario entrenar al clasificador con el mismo número de épocas de ambas clases (balanceadas), para que el clasificador no genere una frontera de decisión con sesgo hacia la clase con mayor prevalencia; para ello, la matriz objetivo final consistió en la replicación, por 5 veces, de las épocas de la matriz objetivo original.

Creación de los vectores de etiquetas de clase. Por cada matriz se creó un vector que contiene la etiqueta de cada época de la matriz correspondiente; la etiqueta de clase 1 para la matriz objetivo, y la etiqueta de clase 0 para la no-objetivo.

Creación del conjunto de entrenamiento. En este paso se unieron verticalmente las dos matrices de épocas, de forma que se duplicó el número de renglones y se conservó el número de columnas. La matriz resultante de esta unión constituyó el *conjunto de entrenamiento* para el clasificador, y este conjunto contiene todas las épocas objetivo y las no-objetivo relacionadas con una sesión de registro.

Aleatorización del conjunto de entrenamiento. Se reordenaron aleatoriamente las épocas del conjunto de entrenamiento para eliminar el orden obtenido al momento del registro y en el paso anterior (épocas objetivo - épocas no-objetivo).

Clasificación. Se entrenó un clasificador a partir del conjunto de entrenamiento, esto es: se obtuvieron los pesos y sesgos que determinan la ecuación de la frontera de decisión que servirá para realizar la clasificación en la etapa de reconocimiento. El algoritmo de clasificación utilizado es LDA elegido por su alta estabilidad [19].

Reconocimiento. Esta etapa tiene como finalidad conocer la etiqueta de clase de un conjunto de datos no observados, distinto al utilizado para entrenar el clasificador (Figura 6b). Los registros utilizados para el reconocimiento en el experimento 1 fue el correspondiente a la sesión 2; para el experimento 2 se utilizó el registro de la sesión 1 y finalmente para los experimentos 3 y 4 fue el registro de la sesión 3.

En esta etapa las épocas se extrajeron basándose en el tipo de estimulación de renglón y columna para clasificarlas como clase 1 o clase 0, por lo tanto, se generaron

12 matrices de datos para reconocimiento, una por cada renglón y cada columna del deletreador.

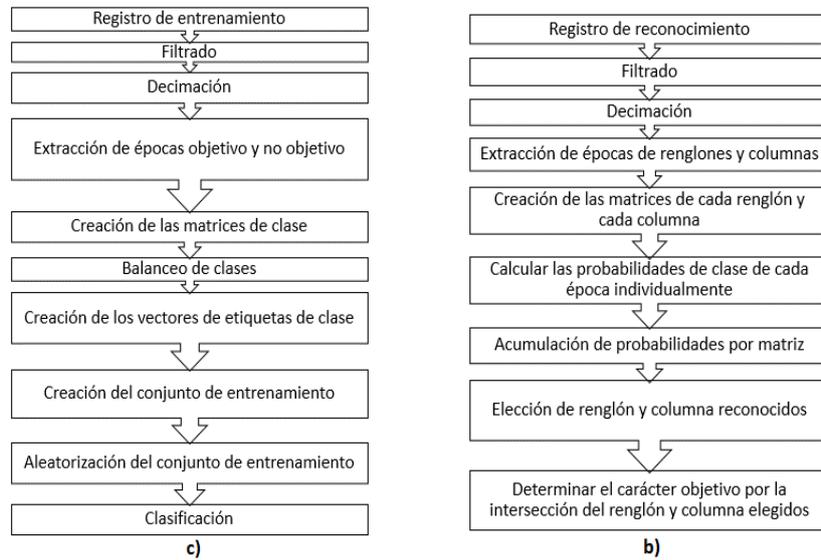


Fig. 5. Preparación de la señal para el entrenamiento (a) y para reconocimiento (b).

Una vez que se obtuvieron las matrices de datos correspondientes a cada renglón y cada columna, se proyectó cada época de cada matriz sobre la ecuación de la frontera de decisión; en otras palabras, para cada época, se multiplicaron los valores de amplitud de cada una de sus muestras con los coeficientes discriminantes para obtener una probabilidad a posteriori de pertenencia a una clase. Las probabilidades de cada matriz de datos de renglones se acumularon para conocer cuál acumuló la mayor probabilidad de pertenecer a la clase 1 y el renglón relacionado se denominó *renglón reconocido*; el mismo procedimiento se realizó para las columnas y obtener las *columnas reconocidas*. El deletreo correcto, entonces, se determinó por la intersección del renglón y columna reconocidos.

Proceso de extracción de épocas. Se obtuvo la información del tiempo en el que ocurrió una estimulación del tipo buscado (objetivo, no-objetivo, renglón n o columna n , donde $n = 1, 2, \dots, 6$). Posteriormente se localizó en el registro EEG la muestra correspondiente a ese momento y se tomaron, -a partir de ella, un número de muestras que correspondan a $t = 600$ ms (el potencial P300 se produce aproximadamente a 300 ms post-estímulo esperado, por lo que se tomó un espacio de tiempo suficiente para registrarlo); se obtuvieron 38 muestras en cada época.

La época correspondiente al estímulo actual contiene todas las muestras registradas durante t , y se colocó en la matriz que corresponde a su tipo; de esta manera, cada época constituye un renglón de la matriz a la cual pertenece y cada muestra que existe en el lapso de tiempo t , constituye una columna o rasgo de la señal EEG en la matriz. La figura 5 ejemplifica el proceso de extracción de épocas basándose en un estímulo de

carácter objetivo y uno de carácter no-objetivo. Las flechas representan los instantes de presentación de los estímulos.

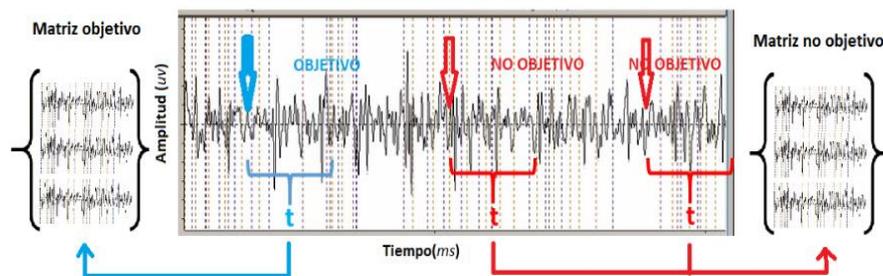


Fig. 6. Extracción de épocas basada en estimulaciones objetivo y no-objetivo.

3.6. Resultados

La tabla 2 muestra la cantidad de reconocimientos logrados mediante la BCI para cada sujeto y en cada posición del electrodo. Los valores de los renglones de cada sujeto indican el número de reconocimientos logrados, de un máximo de 16.

La Tabla 3 muestra el número de aciertos promediados de cada sujeto, las dos primeras columnas muestran promedios por tipo de secuencia, las columnas 3 y 4 promedian según el número de repeticiones, las columnas 5 y 6 por zona de colocación de los electrodos y la última columna muestra el promedio total por sujeto. La última fila muestra promedios de todos los sujetos para los dos tipos de secuencia, para los dos valores de repeticiones, para las dos zonas y el total.

Tabla 2. Cantidad de reconocimientos, de un total de 16, logrados por cada sujeto.

Tipo de secuencia		Al azar				Predefinida			
		E1 - R2		E2 - R1		E1 - R3		E2 - R3	
Experimentos		12	8	12	8	12	8	12	8
Repeticiones		12	8	12	8	12	8	12	8
Oz	Sujeto 1	12	11	11	9	9	5	9	8
	Sujeto 2	9	6	8	7	8	5	9	7
	Sujeto 3	11	11	11	10	5	4	3	2
	Sujeto 4	7	7	5	6	4	3	3	4
	Sujeto 5	7	7	11	7	14	8	13	10
Pz	Sujeto 1	5	6	7	6	6	4	7	3
	Sujeto 2	8	2	3	3	8	6	8	6
	Sujeto 3	13	14	11	11	12	11	12	11
	Sujeto 4	3	3	6	3	2	0	4	6
	Sujeto 5	9	6	10	8	7	7	10	5

Para tener una mejor perspectiva del alcance de estos resultados, se compararon con los obtenidos por Yoon y cols. en 2017 [17], donde utilizaron el algoritmo MKL con

tres tipos de rasgos discriminantes, 5 caracteres, dos sujetos de prueba y número de repeticiones $r = \{1, 2, 3, \dots, 15\}$, utilizando los registros del *Dataset II of BCI competition III*; en este estudio se promediaron los resultados de los dos sujetos con el conjunto de prueba completo, por lo que se obtuvieron los porcentajes de exactitud de deletreos (reconocimiento de carácter) de 85.1% para $r = 8$ y de 95.7 para $r = 12$.

Tabla 3. Cantidad de reconocimientos, de un total de 16, promediados por tipo de secuencia de caracteres, por cantidad de repeticiones, por zonas de colocación del electrodo y totales.

Sujeto	Secuencia		Repeticiones		Zonas		Total
	Azar	Pred	12	8	Oz	Pz	
1	8.3	6.3	8.2	6.5	9.2	5.5	7.3
2	5.7	7.1	7.6	5.2	7.3	5.5	6.4
3	11.5	7.5	9.7	9.2	7.1	11.8	9.5
4	5	3.2	4.2	4	4.8	3.3	4.1
5	8.1	9.2	10.1	7.2	9.6	7.7	8.6
Total	7.7	6.7	8	6.4	7.6	6.8	7.2

Para el trabajo actual se tomaron de la Tabla 2 únicamente lo referente a los resultados para la secuencia predefinida; se obtuvieron 5.75 de reconocimiento para $r = 8$ y 7.65 reconocimientos para $r = 12$ (reconocimientos de renglones o columnas); por lo tanto, los porcentajes de aciertos (exactitud de reconocimientos) correspondiente a cada valor de r : $r = 8 \Rightarrow 35.94\%$; $r = 12 \Rightarrow 47.81\%$. La comparativa resultante se observa con mayor detalle en la Tabla 5.

Tabla 4. Comparativa con los resultados obtenidos con el conjunto de datos *Dataset II of BCI competition III*.

Fuente de los datos	Repeticiones			Cantidad de electrodos utilizados	Número de caracteres utilizados
	8	12	Prom		
Yoon y cols (exactitud de reconocimiento de caracteres)	85.1	95.7	90.4	64	5
Presente trabajo (exactitud de reconocimiento de renglones y columnas)	35.94	47.81	41.875	1	8

4. Conclusiones

La secuencia de caracteres al azar permitió mayor cantidad reconocimientos que una predefinida; esto muy probablemente se debió a que, en el segundo caso, el usuario tuvo dificultades para mantener la atención en un solo carácter, o bien, olvidó, momentáneamente en algunas ocasiones, el siguiente carácter de la secuencia. Así, se considera que los deletreos con secuencia predefinida fueron más representativos del desempeño real del enfoque propuesto.

El número de repeticiones r influyó directamente en el desempeño de reconocimientos, descendió este un 20% (de 8 a 6.4) con una disminución del 33.33%

en el número de repeticiones (de 12 a 8); en otras palabras, una disminución del 33.33% de repeticiones provocó una disminución del 20% de reconocimientos. Un valor de r aún mayor mejoraría los mismos, pero también provocaría un cansancio excesivo en las sesiones para el sujeto.

La diadema MindWave, a pesar de contar con un sólo electrodo, permitió obtener, en este grupo de sujetos, un promedio general de 7.2 reconocimientos (renglón o columna) de un total de 16, es decir, el 45% del total. Con el trabajo comparado en la Sección 4, con un conjunto de datos de 64 canales, se obtuvo un 90.4 de exactitud de reconocimientos de caracteres; esta comparación se hace considerando únicamente los resultados del presente trabajo en los experimentos 3 y 4, en los cuales el tipo de secuencia a deletrear es predefinida.

Cabe mencionar que, dado el número de caracteres de la matriz del deletreador (36), si se operara con señales aleatorias y no relacionadas con la atención del sujeto, el porcentaje de aciertos estaría apenas por encima del 2.7 % (1/36). En perspectiva, un porcentaje de reconocimientos del 45% denota que tanto la señal de EEG registrada, como el enfoque propuesto de procesamiento y clasificación, son informativos de la presencia del potencial P300 y de los caracteres de interés del sujeto, a pesar de contar con un solo canal de registro.

5. Trabajo a futuro

Se recomienda continuar experimentando con la diadema MindWave en aplicaciones BCI con otros algoritmos de clasificación como Máquinas de Soporte Vectorial, Perceptrón Multicapa y otros enfoques novedosos que puedan surgir. También puede ser viable explorar el uso de la diadema con otros paradigmas de BCI, como los Potenciales Visuales de Estado Estable o los Ritmos Sensorial Motores.

Agradecimientos. Al Consejo Nacional de Ciencia y Tecnología CONACYT, por la beca de estudios de maestría de Sergio Vivar y al Instituto Nacional de Rehabilitación, por las facilidades prestadas durante la estancia de investigación.

Referencias

1. Abdulkader, S. N., Artia, A., Mostafa, S., Mostafa, M.: Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal*, Vol. 16, pp. 213–230 (2015)
2. Mochizuki, Y., Hayashi, K., Nakayama, Y., Shimizu, T., Kamide, M., Ogino, M., Komori, T., Hasegawa, M., Isozaki, E., Nakano, I.: ALS patients with ability to communicate after long-term mechanical ventilation have confined degeneration to the motor neuron system. *Journal of the Neurological Sciences*, Vol. 363, pp. 245–248 (2016)
3. Ceres, R., Mañanas, M. Á., Azorín, J. M.: Interfaces y Sistemas de Rehabilitación y Compensación Funcional para la Autonomía Personal y la Terapia clínica. *Revista Iberoamericana de Automática e Informática Industrial*, Vol. 8, No. 2, pp. 5–15 (2011)

4. García, P. A., Spinelli, E. M., Toccaceli, G. M., Haberman, M. A.: Interfaz Cerebro Computador basada en potenciales evocados visuales de estado estacionario: ensayos preliminares. In: XVIII Congreso Argentino de Bioingeniería SABI 2011 - VII Jornadas de Ingeniería Clínica, Mar de Plata (2011)
5. Hornero, R., Corralejo, R., Álvarez, D.: Brain-Computer Interface (BCI) aplicado al entregamiento cognitivo y control domótico para prevenir los efectos del envejecimiento. *Lychnos*, No. 8, pp. 29–34 (2012)
6. Strehl, U., Leins, U., Goth, G., Klinger, C., Hinterberger, T., Birbaumer, N.: Self-regulation of Slow Cortical Potentials: A New Treatment for Children With Attention-Deficit/Hyperactivity Disorder. *Pediatrics*, Vol. 118, No. 5 (2006)
7. NeuroSky. Disponible en: <http://neurosky.com>. Último acceso: 10 Abril 2017
8. Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., Lécuyer, A.: OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments. *Presence*, Vol. 19, No. 1, pp. 35–53 (2010)
9. Farwell, L. A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, Vol. 70, No. 6, pp. 510–523 (1988)
10. Pires, G., Nunes, U., Castelo-Branco, M.: Evaluation of Brain-Computer Interfaces in Accessing Computer and other Devices by People with Severe Motor impairments. *Procedia Computer Science ELSEVIER*, Vol. 14, pp. 283–292 (2012)
11. Poli, R., Cinel, C., Matran-Fernandez, A., Sepúlveda, F., Stoica, A.: Towards cooperative brain-computer interfaces for space navigation. *De IUI '13: Proc. of the 2013 international conf. on Intelligent, ACM* (2013)
12. Wang, M., Daly, I., Allison, B. Z., Jin, J., Zhang, Y., Chen, L., Wang, X.: A new hybrid BCI paradigm based on P300 and SSVEP. *Journal of Neuroscience Methods*, Vol. 244, pp. 16–25 (2015)
13. Suryotrisongko, H., Samopa, F.: Evaluating OpenBCI Spiderclaw V1 Headwear's Electrodes Placements for Brain-Computer Interface (BCI) Motor Imagery Application. *Procedia Computer Science*, Vol. 72, pp. 398–405 (2015)
14. OpenBCI.: Disponible en: <http://openbci.com>. Último acceso: Mayo 2017
15. Schreuder, M., Riccio, A., Riseti, M., Dähne, S., Ramsay, A., Williamson, J. D., Mattia-Tangermann, M.: User-centered design in brain-computer interfaces-A case study. *Artificial Intelligence in Medicine (Elsevier B.V.)*, Vol. 59, No. 2, pp. 71–80 (2013)
16. Chang, M. H., Lee, J. S., Heo, J., Park, K. S.: Eliciting dual-frequency SSVEP using a hybrid SSVEP-P300 BCI. *Journal of Neuroscience Methods*, Vol. 258, pp. 104–113 (2016)
17. Yoon, K., Kim, K.: Multiple kernel learning based on three discriminant features for a P300 speller BCI. *Neurocomputing*, Vol. 237, No. 12, pp. 133–144 (2017)
18. Bougrain, L., Saavedra, C., Ranta, R.: Finally, what is the best filter for P300 detection? *Tools for Brain-Computer Interaction* (2012)
19. Breiman, L.: Arcing classifiers. *The Annals of Statistics*, Vol. 26, No. 3, pp. 801–849 (1998)

Hacia un método de transferencia de aprendizaje en señales de EEG de habla imaginada

Jessica Nayeli López Espejel¹, Maya Carrillo Ruíz¹, Luis Villaseñor Pineda²,
Alejandro Torres García²

¹ Benemérita Universidad Autónoma de Puebla, Puebla, México

² Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México

acissejol@hotmail.com, crrllrzmy@gmail.com, luis.villasenor.pineda@gmail.com,
alejandro.torres@ccc.inaoep.mx

Resumen. El fin último de este trabajo es disminuir o eliminar la etapa de entrenamiento, para cuando un nuevo sujeto utilice una BCI (brain-computer interface) basada en habla imaginada. La etapa de entrenamiento se debe a que es necesario adquirir suficiente información para identificar los patrones que permitan distinguir lo que el sujeto imagina pronunciar. Ahora bien, en habla imaginada, como cualquier otro potencial evocado, el proceso de entrenamiento es tardado y tedioso. En este artículo se presenta una serie de experimentos que busca comprobar si su información de entrenamiento ya disponible para un grupo de sujetos puede aprovecharse para un nuevo sujeto. El método empleado usa mapas auto-organizados para seleccionar la información ya disponible con la cual se generan clasificadores binarios para identificar las palabras presentes en los EEG del nuevo sujeto. Los resultados alcanzados son alentadores y dan pauta para el diseño de un método apropiado para transferencia de aprendizaje sujeto-a-sujeto en habla imaginada.

Palabras clave: transferencia de aprendizaje, habla imaginada, EEG, random forest.

Towards a Method of Transfer Learning in Imagined Speech EEG Signals

Abstract. The target of this research is decrease or eliminate the training stage, when a new subject uses a brain-computer interface (BCI) based on imagined speech. In the training phase it is necessary to acquire enough information to identify the patterns that allow to distinguish what the subject imagines to pronounce. Then, in imagined speech, like any other evoked potential, the training process is tedious. This article presents experiments that seek to verify if the training information and the available information for a group of subjects can be used for a new person. The proposed method uses self-organized maps to select information in order to generate binary classifiers to identify the words presented in a new EEG subject. The results are encouraged and give the

guideline to design a suitable method for a subject to subject transfer learning method for imagined speech.

Keywords: transfer learning, imagined speech, EEG, random forest.

1. Introducción

Durante las últimas dos décadas, una variedad de métodos han evaluado la posibilidad de transformar las señales cerebrales obtenidas a través del cráneo y el cuero cabelludo para activar diferentes dispositivos. De esta forma surge una nueva interfaz que no requiere ser manipulada a través de actividades musculares. Este tipo de interfaz recibe el nombre de BCI (Brain-Computer Interface, en español Interfaz Cerebro-Computadora). En breve, una BCI es un sistema cuyo objetivo es establecer la comunicación entre un humano y un sistema automático, a través de las señales cerebrales, las cuales pueden ser detectadas a través de diferentes instrumentos [1].

Las BCI utilizan diferentes fuentes electrofisiológicas, tales como: los 2 ritmos sensoriales motrices (μ y β) y los potenciales evocados. Bajo estos últimos cae el caso del habla imaginada, también referida como habla interna o habla no pronunciada. Este término se refiere a imaginar la pronunciación de una sílaba o palabra sin emitir sonidos ni articular gestos para ello [2].

Las BCIs en general presentan dos grandes problemas: (i) el usuario debe pasar por una fase de entrenamiento que en ocasiones puede llevar días o incluso meses, y (ii) las bajas tasas de comunicación, es decir, comunicar un comando puede llevarse mucho tiempo, por ejemplo, en el caso de las BCI basadas en P300 comunicar una sola palabra procesada, puede llevarse un minuto.

Una BCI basada en habla imaginada disminuye de forma importante las tasas de comunicación, y aunque la fase de entrenamiento es menor ésta sigue siendo importante. De ahí el interés en aplicar transferencia de aprendizaje en esta tarea, para disminuir el tiempo dedicado por un usuario en la etapa de entrenamiento. La transferencia de aprendizaje se describe como el procedimiento de usar los datos de una tarea para aumentar el rendimiento de una tarea relacionada. [3]

La transferencia de aprendizaje en BCI puede aplicarse en diferentes escenarios, dos de ellos son: de sujeto-a-sujeto y de sesión-a-sesión. En ambas situaciones existen variaciones de muy diversa naturaleza. Éstas van desde una leve desigualdad en colocación de los electrodos por la forma de la cabeza, el contacto del cuero cabelludo con los electrodos es distinto, además de diferencias debidas al estado anímico de las personas, lo cual puede cambiar de un día a otro, su concentración puede verse alterada, etc.

En el presente trabajo se muestra los primeros pasos hacia un método transferencia de aprendizaje sujeto-a-sujeto. En este caso, observamos si los datos de entrenamiento adquiridos para otros sujetos pueden ser utilizados como datos de entrenamiento para un nuevo sujeto. La tarea no es simple pues además de las variaciones mencionadas, cuando se trabaja con habla imaginada, no se tiene la certeza de las regiones cerebrales que intervienen, y es claro que cada sujeto produce distintas señales cerebrales, aunque se trate de la misma palabra.

El resto del trabajo está organizado de la siguiente manera: en la Sección 2 se presenta el trabajo relacionado, en la Sección 3 se describe el método propuesto, en la Sección 4 los experimentos y resultados obtenidos, y finalmente en la Sección 5 las conclusiones y el trabajo a futuro.

2. Trabajo relacionado

La transferencia de aprendizaje en la problemática de las BCI basadas en EEG puede aplicarse desde diferentes perspectivas. El escenario más frecuente es el tratar de adaptar una BCI a un nuevo sujeto a partir de datos previos registrados con otros sujetos. Sin embargo, la transferencia de aprendizaje también puede aplicarse a un mismo sujeto para abordar el problema de la habituación (i.e. la variabilidad de la señal cerebral entre diferentes sesiones) o para incrementar el vocabulario de palabras imaginadas a reconocer para el caso del habla imaginada. A continuación se presentan algunos trabajos que han aplicado transferencia de aprendizaje en BCI. Cabe señalar que hasta donde sabemos esta técnica no ha sido aplicada a habla imaginada.

En [4] la base de datos que contiene los EEG es de tareas de movimientos imaginarios de la mano derecha y pie derecho, y está formada por 5 sujetos. En este artículo presentaron una modificación del algoritmo de Patrones Espaciales Comunes CSP (Common Spatial Pattern) para la transferencia de aprendizaje de sujeto a sujeto, dicha modificación consistió en una combinación lineal de matrices de covarianza de los individuos a los que se les tomó el EEG. Además, también desarrollaron dos métodos para determinar una matriz de covarianza compuesta, que es una suma ponderada de las matrices de covarianza. En el primer método se ajusta el parámetro λ [0,1] para ajustar la matriz de covarianza. En el segundo método se enfatiza en las matrices de covarianza para los sujetos con características similares, por lo que se calcula la divergencia entre las distribuciones de datos para dos o más sujetos que se estén comparando. La precisión de clasificación media para los sujetos en el primer método es de 75.23 y para el segundo método propuesto es de 82.5653, cuyos resultados son mejores que en el método tradicional, puesto que para dicho método la precisión de clasificación media en los sujetos es de 74.16.

En [5] presentaron un estudio de las señales EEG de niños con déficit de atención e hiperactividad con el propósito de apoyar un sistema de diagnóstico asistido por computadora. Las señales EEG se tomaron a 13 sujetos, y de ellos la mitad estaba diagnosticado médicamente con TDAH (por sus siglas, Trastorno de Déficit de Atención e Hiperactividad) y la otra mitad estaban clínicamente sanos. Los EEG se analizaron con la Transformada Discreta de Wavelet para extraer las características del espectro de potencia de la señal cerebral, y una técnica de umbral fue utilizada para mejorar aún más la calidad de los datos debido a que tenían ruido de artefactos. Para discriminar el nivel de atención en la actividad eléctrica cerebral de los niños con TDAH, se utilizaron mapas auto-organizados cuya entrada eran los coeficientes obtenidos de la TDW. La transferencia de aprendizaje se utiliza para discernir entre los niños que tienen TDAH y los que no, a través de las mejores señales, agrupadas por la SOM.

En [6] propusieron un algoritmo para reducir el tiempo de adaptación (o calibración) para un sujeto nuevo. Debido a la variabilidad de sujetos, por lo general un nuevo individuo tiene que someterse a una sesión de 30 minutos de calibración para poder recuperar datos suficientes para la formación de un modelo BCI basado en sus patrones/cerebro. El algoritmo propuesto consta de dos pasos: en el primer paso se adaptan los datos del nuevo sujeto a los datos de los EEG's que ya se tienen en la base de datos; en el segundo paso se selecciona el EEG del individuo de la base de datos que tenga más coincidencias con los datos del EEG del nuevo sujeto.

La base de datos utilizada contiene EEG's de 9 sujetos, y para cada uno se emplearon 22 electrodos. La base de datos tiene EEG de cuatro actividades imaginarias: mano derecha, mano izquierda, lengua y pie; sin embargo, para este trabajo únicamente se utilizaron dos tipos de movimiento imaginado: mano derecha y mano izquierda. En la etapa experimental de los nuevos sujetos, fueron 8 los individuos con los que se trabajó. De acuerdo a los resultados obtenidos, se puede apreciar que el algoritmo propuesto en este trabajo de investigación reduce el tiempo de calibración en un 85%.

En [7] están conscientes de la problemática que existe entre la variabilidad de las señales cuando se toman de sujeto-a-sujeto y de sesión-a-sesión, por lo cual propusieron lo siguiente: 1) un método para la extracción de diccionarios compartido por múltiples sujetos, mediante el empleo de una técnica de diccionario de procesamiento digital de señales, modificada para compensar las variaciones entre sujetos y sesiones; 2) un enfoque de decodificación de transferencia, en donde los sujetos se encuentran en estado de reposo, y de esta forma eliminan la necesidad de una calibración estándar. La aplicación de la metodología que los autores proponen a un conjunto de datos de EEG's durante una tarea de atención visual-espacial, muestran que la compensación de variabilidad se logra estando los sujetos en estado de reposo, y de esa manera se reduce la redundancia en los diccionarios.

La aplicación de la decodificación de transferencia de conocimiento en los sujetos logró un mejor desempeño frente a los métodos existentes de decodificación. Por lo tanto, los resultados revelan que el análisis de las actividades cerebrales con el método propuesto permite reducir el tiempo de calibración, y es muy útil para el uso práctico de BCI en ambientes variables.

En [3] desarrollaron un framework para la transferencia de aprendizaje utilizando una BCI. Este trabajo consiste de dos experimentos, el primer experimento se realizó tomando el EEG de 10 sujetos sanos (2 mujeres y 8 hombres de entre 25.5 y 25.6 años), el procedimiento constó de un total de 150 ensayos por clase (derecha/ izquierda) y por persona. La extracción de características se llevó a cabo a través de filtros espaciales usando una configuración de superficie de Laplace y también empleando la Transformada Discreta de Fourier.

Para la clasificación de las actividades mentales se utilizaron cinco algoritmos (dos de aprendizaje multitarea denominados: MT_FD y MT por sus siglas en inglés "multitask learning", un algoritmo de cresta de regresión estándar, otro de cresta de regresión pero usando FD, y por último máquinas de soporte vectorial), y entre ellos, el que tuvo la mayor precisión de clasificación aplicando la transferencia de aprendizaje fue el algoritmo MT_FD con el 65% de precisión.

3. Método propuesto

El método para realizar nuestros primeros experimentos hacia la transferencia de aprendizaje consiste de cinco pasos: pre- procesamiento de la señal, caracterización de las señales, selección del conjunto de entrenamiento, entrenamiento de clasificadores binarios y evaluación con nuevos sujetos. Las subsecciones a continuación detallan cada uno de estos pasos.

3.1. Pre-procesamiento

El pre-procesamiento de la señal de los EEG se llevó a cabo con el método de referencia promedio (CAR, por sus siglas en inglés). El propósito de este método es mejorar la señal aprovechando que se tienen lecturas simultáneas de los electrodos. La fórmula que lo describe, es la siguiente:

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n} \sum_{j=1}^n V_j^{ER}, \quad (1)$$

donde V_j^{ER} es el potencial entre el i -ésimo electrodo y la referencia, y n es el número de electrodos.

3.2. Caracterización de las señales

En este paso se extraen características de las señales utilizando la Transformada Discreta de Wavelet (DWT). La Transformada Discreta Wavelet se utilizó con 4 niveles de descomposición ya que sólo son de interés los rangos de frecuencia menores a 32 Hz. Se empleó la wavelet biortogonal 2.2 de la toolbox de Matlab R2013b ya que de acuerdo a [2] es la familia de wavelet que genera los mejores resultados. De estos 4 niveles, se obtienen coeficientes de aproximación y coeficientes de detalle, con las siglas cA y cD , tal como se muestra en la Figura 1. Posteriormente, se calculó la energía instantánea entre los 4 niveles.

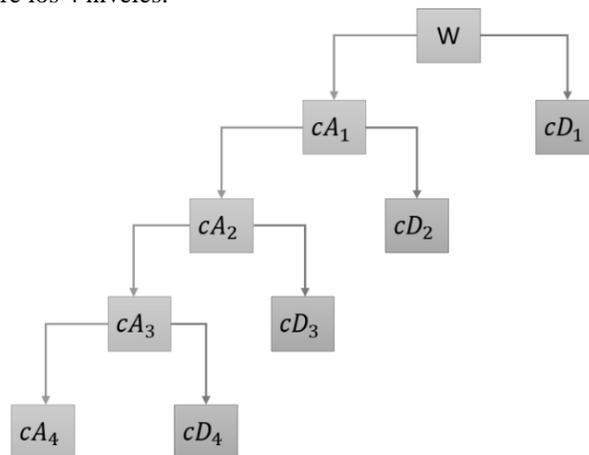


Fig. 1. Descomposición wavelet en 4 niveles.

La energía instantánea se obtiene a partir de los coeficientes calculados con la DWT. De esta forma, cada repetición de cada una de las palabras imaginadas se representó mediante 5 valores que representan la diferencia de energía de cada uno de los niveles de descomposición (D1-D4 y A4). Gracias a este procedimiento se obtiene el mismo número de coeficientes para todas las ventanas de la señal. Este proceso de extracción se realiza para cada electrodo.

Una vez caracterizadas las señales mediante la energía instantánea, se procedió a entrenar clasificadores binarios como primer paso para el proceso de transferencia de aprendizaje.

3.3. Entrenamiento de clasificadores binarios

En este paso se crearon clasificadores binarios para las combinaciones de las palabras imaginadas arriba, abajo, derecha, izquierda. Se decidió hacer clasificadores binarios para observar la tasa de transferencia para cada par de palabras. A pesar de que se sabe que existen diferencias entre sujetos a pesar de imaginar la misma palabra, deben existir ciertas semejanzas pues se evoca el movimiento de aparato fonador para producir sonidos que corresponden a cierta norma. Así para la palabra arriba deben existir semejanzas en las señales producidas por diferentes sujetos.

Sin embargo, el grado de semejanza puede variar mucho de un sujeto a otro, de ahí que se crearon clasificadores binarios considerando los datos de entrenamiento de 10 sujetos. Para crear estos clasificadores base se seleccionaron aquellas señales más semejantes entre los 10 sujetos. Para ello se utilizó un algoritmo de *clustering* en nuestro caso se utilizaron mapas auto-organizados (SOM por sus siglas en inglés). Los elementos agrupados de manera correcta por la SOM, se tomaron como datos de entrenamiento para generar cada uno de los clasificadores base.

Los clasificadores binarios consideran todas las posibles combinaciones de las palabras como en [6], entrenados con las instancias agrupadas de manera correcta por la SOM.

3.4. Evaluación en nuevos sujetos

En este paso se evaluaron los clasificadores binarios con los datos de las cuatro palabras imaginadas de un conjunto de sujetos diferente. De esta forma se calculó el porcentaje de instancias clasificadas correctamente para estos nuevos sujetos. Es decir, el porcentaje de transferencia realizado de manera correcta.

4. Experimentos y resultados

Base de datos

La base de datos empleada es la descrita en [2]. Para la elaboración de esta base de datos se utilizó el kit de EMOTIV. Las muestras de EEG se le practicaron a 27 sujetos, cada individuo repitió 33 veces (en algunos casos más repeticiones) 5 palabras (arriba, abajo, derecha, izquierda, seleccionar). De manera gráfica, la base de datos se representa como muestra la Figura 2.

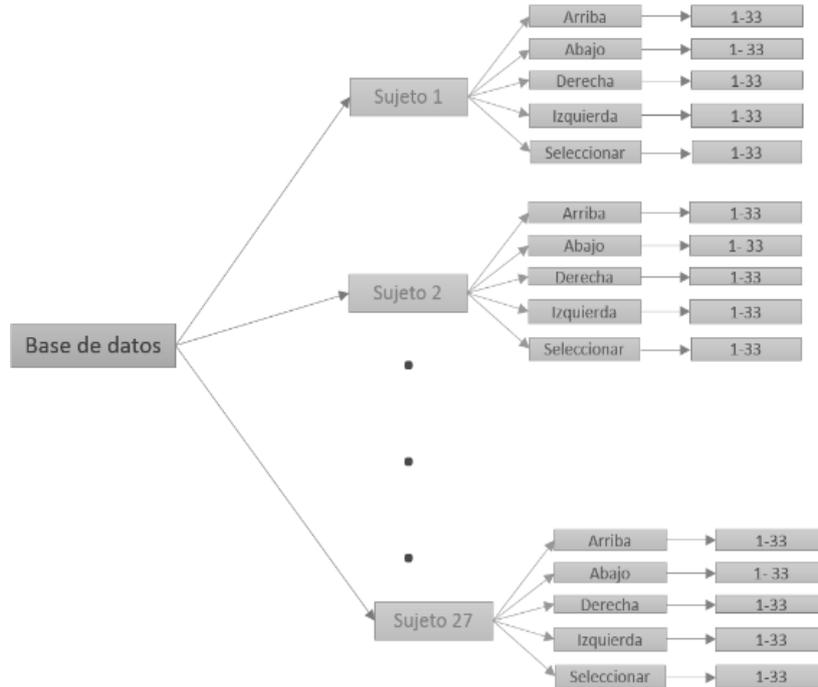


Fig. 2. Base de datos.

Cabe mencionar que la palabra “seleccionar” presente en la base de datos se excluyó para este trabajo. Las SOMs fueron implementadas en Matlab 2015a utilizando la toolbox y los clasificadores binarios se implementaron en Weka empleando Random Forest, con las instancias agrupadas de manera correcta por la SOM. Cabe señalar que para construir los clasificadores binarios los datos fueron balanceados usando undersampling.

A partir de estos clasificadores, se evaluó la transferencia de aprendizaje al clasificar las señales de nuevos sujetos. Los resultados obtenidos de cada uno de los clasificadores binarios se presentan en las tablas 3 y 4.

Como primera observación de las tablas 3 y 4, nuestros resultados muestran tasas de exactitud promedio entre 53% y 60%. Estos resultados son muy alentadores pues hay que recordar que estos resultados se han obtenido sin hacer ningún tipo de adaptación hacia los nuevos sujetos. Lo que es más, las clases más diferenciables alcanzan exactitudes promedio de 60.07% y 59.18% (clasificadores arriba-abajo y derecha-izquierda), alcanzando rendimientos similares a los reportados en el estado del arte [5].

Observando a detalle los resultados, algunos sujetos alcanzan resultados muy altos, a pesar de que para otros los resultados no son favorables. Por ejemplo, en el caso del sujeto número 22 en todas las combinaciones que incluyen la clase arriba presenta altas exactitudes (arriba-abajo = 90.91%, arriba-derecha = 89.39% y arriba-izquierda = 83.33%). De igual manera, el sujeto 26 para todas las combinaciones su exactitud va de 57.58% a 89.39%, seguido del sujeto 3, cuyo porcentaje de exactitud en la

transferencia esta entre 59.09% y 66.67% en cinco de seis combinaciones entre las clases.

Tabla 1. Resultado de tres combinaciones para 17 sujetos nuevos.

Sujeto	ARRIBA-ABAJO			ARRIBA-DERECHA			ARRIBA-IZQUIERDA		
	EXAC.	Precisión arriba	Precisión abajo	EXAC.	Precisión arriba	Precisión derecha	EXAC.	Precisión arriba	Precisión izquierda
2	50.00%	0.50	0.00	50.00%	0.500	0.000	48.48%	0.492	0.400
3	63.64%	0.636	0.636	62.12%	0.645	0.580	66.67%	0.900	0.620
6	48.48%	0.400	0.492	48.48%	0.333	0.492	50.00%	0.500	0.500
7	56.06%	0.534	0.750	45.45%	0.471	0.400	53.03%	0.531	0.529
10	63.64%	0.636	0.636	62.12%	0.633	0.611	48.48%	0.486	0.484
12	62.12%	0.912	0.645	51.52%	0.700	0.512	56.06%	0.900	0.652
13	50.00%	0.500	0.500	51.52%	1.000	0.508	51.52%	1.000	0.508
14	45.45%	0.455	0.455	53.03%	0.589	0.519	48.48%	0.455	0.491
16	63.64%	0.588	0.714	63.64%	0.700	0.610	51.52%	0.714	0.645
17	62.12%	0.571	0.900	63.64%	0.585	0.846	66.67%	0.604	0.923
19	48.48%	0.492	0.429	53.03%	0.519	0.583	50.00%	0.500	0.500
21	51.52%	0.788	0.182	51.52%	0.509	0.538	56.06%	0.540	0.625
22	90.91%	0.846	1.000	89.39%	0.842	0.964	83.33%	0.750	1.000
23	63.64%	0.712	0.636	43.94%	0.500	0.468	48.48%	0.400	0.492
24	62.12%	0.700	0.841	63.64%	0.580	0.840	50.00%	0.500	0.500
25	50.00%	0.500	0.500	57.58%	0.667	0.549	83.33%	1.000	0.502
26	89.39%	0.842	0.964	57.58%	0.600	0.610	66.67%	0.600	0.814
	60.07 %	0.624	0.605	56.95 %	0.610	0.566	57.58 %	0.640	0.599

En cuestión de precisiones, las mejores se presentan en la clase de arriba porque en cada una de las combinaciones donde está presente, su precisión está por encima de 0.60. La transferencia de aprendizaje con mayor deficiencia se observa en el sujeto 2 con algunas precisiones iguales a cero. En lo que concierne a la combinación de las clases, las de mayor precisión son arriba-abajo y arriba-izquierda. Por otro lado, se tiene en caso de las clases arriba-izquierda en donde se encuentra la precisión promedio más baja (izquierda = 0.492).

Tabla 2. Resultado de 3 combinaciones entre clases.

Sujetos	DERECHA-IZQUIERDA			ABAJO-DERECHA			ABAJO-IZQUIERDA		
	EXAC	Precisión derecha	Precisión izquierda	EXAC.	Precisión abajo	Precisión derecha	EXAC.	Precisión abajo	Precisión izquierda
2	50.00%	0.500	0.000	50.00%	0.500	0.000	50.00%	0.500	0.500
3	59.09%	0.684	0.645	63.64%	0.615	0.575	50.00%	0.500	0.500
6	59.09%	0.750	0.556	46.97%	0.333	0.483	43.94%	0.500	0.250
7	54.55%	0.552	0.541	48.48%	0.490	0.471	37.88%	0.357	0.578
10	42.42%	0.381	0.444	50.00%	0.500	0.500	60.61%	0.585	0.640
12	89.39%	0.574	0.812	59.09%	0.526	0.667	60.62%	0.585	0.756
13	63.64%	0.570	0.700	50.00%	0.500	0.500	100.00%	1.000	0.000
14	63.64%	0.800	0.588	42.42%	0.381	0.444	46.97%	0.480	0.438
16	59.09%	0.571	0.751	59.09%	0.845	0.610	60.62%	0.612	0.712
17	59.09%	0.570	0.611	59.09%	0.575	0.615	46.97%	0.462	0.475
19	45.45%	0.472	0.385	54.55%	0.526	0.667	48.48%	0.444	0.491
21	63.64%	0.528	0.615	62.62%	0.754	0.700	48.48%	0.474	0.489
22	59.09%	0.588	0.594	40.91%	0.842	0.446	57.58%	0.551	0.647
23	51.52%	0.571	0.508	50.00%	0.500	0.500	39.39%	0.431	0.267
24	63.63%	0.800	0.712	63.64%	0.512	0.600	57.58%	0.550	0.641
25	59.09%	0.512	0.754	42.42%	0.459	0.450	43.94%	0.463	0.333
26	63.64%	0.514	0.700	59.09%	0.526	0.662	60.62%	0.585	0.640
	59.18%	0.585	0.583	53.06%	0.552	0.523	53.75%	0.534	0.492

5. Conclusiones y trabajo futuro

El presente trabajo describe un método que abarcaría los primeros pasos hacia un método transferencia de aprendizaje sujeto-a-sujeto en señales de EEG en habla imaginada. Los experimentos realizados se orientaron a observar si los datos de entrenamiento adquiridos para otros sujetos pueden ser utilizados como datos de entrenamiento para un nuevo sujeto, sin incluir ningún tipo de información o calibración.

De los resultados obtenidos y mostrados en la sección anterior, se observa que el porcentaje de exactitud entre dos clases es diferente en cada combinación, y eso nos indica que hay clases que son más diferenciables de otras, por ejemplo, la clase con mayor exactitud promedio es arriba-abajo con el 60.07%, seguida de las clases derecha-

izquierda con 59.18%. Por otro lado, tenemos que las clases menos diferenciables y en la que las instancias se confunden más son las clases abajo-izquierda y abajo-derecha.

No obstante, las tasas de exactitud promedio están entre 53% y 60%, que aunque lejos de resolver el problema, son resultados muy alentadores. Sobre todo porque en estos primeros experimentos no se recurrió a ningún proceso de adaptación hacia los nuevos sujetos. De hecho, se llegó a tener casos con exactitudes promedio similares a los reportados en el estado del arte [5].

Por otro lado, existe una fuerte variación de los clasificadores entre los sujetos. Se tiene el caso de los sujetos 22 y 26 que presentan altas exactitudes en todos los clasificadores; pero también se presenta el caso del sujeto 2 con muy bajos resultados.

Como trabajo futuro se desean explorar varias ideas. Por un lado, utilizar otro método para seleccionar las instancias de entrenamiento, al menos para aquellas clases en que no se tienen información para discriminar correctamente entre ellas. Por otro lado, incluir un módulo de adaptación o calibración que permita incluir información del nuevo sujeto. Para ello se pretende seguir lo propuesto en [6]. Donde como primer paso para adaptar los datos de cada EEG de un sujeto a una base de datos base de EEG, se minimiza la distribución entre los dos conjuntos de EEG.

Agradecimientos. Los autores agradecen al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo a esta investigación, el cual se realizó a través de la beca 701206 y del proyecto No. Ref. 2016-01-2228. Los autores también desean agradecer el apoyo de la Red Temática en Tecnologías del Lenguaje/CONACYT, No. Ref. 260178, 271622, por los mecanismos de colaboración brindados para realizar este trabajo. De igual manera, agradecemos a la Vicerrectoría de Investigación y Estudios de Posgrado (VIEP) por el apoyo brindado para la publicación del presente trabajo.

Bibliografía

1. Morales-Flores, E., Ramírez-Cortés, J. M., Gómez-Gil, P., Alarcón-Aquino, V.: Brain Computer Interface Development Based on Recurrent Neural Networks and ANFIS Systems. Springer, pp. 215–236 (2013)
2. Torres-García, A., Reyes-García, C. A., Villaseñor-Pineda, L., Ramírez-Cortés, J. M.: Análisis de Señales Electroencefalográficas para la Clasificación de Habla Imaginada. Revista Mexicana Ingeniería Biomédica, pp. 23–29 (2013)
3. Jayaram, V., Alamgir, M., Altun, Y., Schököpf, B.: Wentrup G.: Transfer Learning in Brain-Computer Interfaces. IEEE, pp. 20–31 (2016)
4. Kang, H., Nam, Y., Choi, S.: Composite Common Spatial Pattern for Subject-to-Subject Transfer. IEEE Signal Processing Letters, pp. 683–686 (2009)
5. Hyun-Lee, S., Abibullaev, B., Kang, W. S., Shin, Y., An, J.: Analysis of Attention Deficit Hyperactivity Disorder in EEG Using Wavelet Transform and Self Organizing Maps. In: International Conference on Control, Automation and Systems, pp. 2439–2442 (2010)
6. Arvaneh, M., Robertson, I., Ward, T. E.: Subject-to-Subject Adaptation to Reduce Calibration Time in Motor Imagery-based Brain-Computer Interface. IEEE, pp. 6501–6504 (2014)
7. Marioka, H., Kanemura, A., Hiramaya, J. I., Shikauchi, M., Ogawa, T., Ikeda, S., Kawanabe, M., Ishii, S.: Learning a common dictionary for subject-transfer decoding with resting calibration. NeuroImage, pp. 167–178 (2015)

Comparación de filtros para el análisis del espectro espacial bidimensional de Fourier, en el proceso de digitalización 3D

Juan Carlos Moya Morales, Jesús Carlos Pedraza Ortega,
Juan Manuel Ramos Arreguín, José Emilio Vargas Soto

Universidad Autónoma de Querétaro, Querétaro, Querétaro, México

moyajc@gmail.com, caryoko@yahoo.com, jsistdig@yahoo.com.mx, emilio@mecatronica.net

Resumen. En el presente trabajo se muestra una comparación entre tres filtros diferentes, dentro del proceso de digitalización 3D por Transformada bidimensional de Fourier. Se lleva a cabo la implementación y comparación de un filtro por umbral, un filtro morfológico rectangular y uno elíptico, estos filtros son generados y representados mediante una máscara binaria, el cual se multiplica por el espectro espacial bidimensional de Fourier, eliminando la mayoría de las frecuencias y obteniendo una escena con franjas deformadas disminuyendo la mayor parte del ruido inicial. Para validar esta comparación, se utilizan dos objetos; un maniquí y un murciélago, se aplican los diferentes filtros a ambos objetos y se realiza la comparación de los resultados. En la etapa de filtrado también se realiza un filtro sobre la frecuencia espacial f_0 , que permite obtener un patrón de franjas uniforme.

Palabras clave: Fourier bidimensional, filtro morfológico, umbral, digitalización 3D.

Comparison of Filters for the Analysis of the Two-dimensional Fourier Space Spectrum in the 3D Digitization Process

Abstract. In this paper, a comparison between three different filters in the 3D scanning processes by Fourier Transform two-dimensional. The implementation and comparison about the filter by threshold, morphological filtering like rectangle and elliptic, the representation about the filters are in a binary mask and it is multiplied by the spectrum of Fourier Transform two-dimensional, with this process the frequencies are eliminated and the new scene shows less noise than the original scene. To compare we used two objects a manikin and a bat figure, the different filters are applied and the results are compared. In the filtering step also we filter the one pixel about f_0 to get a uniform scene.

Keywords: two-dimensional Fourier, morphological filter, threshold, scanning 3D.

1. Introducción

1.1. Perfilometría por transformada de Fourier

La Perfilometría por Transformada de Fourier (FTP) se encuentra dentro de las técnicas de digitalización de objetos, el cual es un proceso por el cual se pueden almacenar, las características físicas de un objeto en la memoria de una computadora. La FTP ha sido objeto de análisis y estudio por muchos investigadores a lo largo de las últimas décadas, desde que sus autores Takeda, et al. lo propusieron en los 80's donde presentaron su metodología basada en la Transformada de Fourier para topografía, mostrando que era posible la identificación de valores que se encontraban en elevación y depresión de la superficie analizada, a partir de la deformación de una franja con perfil senoidal, la cual es proyectada y deformada por las características morfológicas del objeto analizado, tal como se menciona en [1]. En el siguiente año Takeda et al, [2] presentó una nueva técnica para mediciones automáticas basadas en la perfilometría por medio de la Transformada de Fourier.

A partir de estas publicaciones los trabajos presentados por muchos investigadores han tratado de mejorar el método, en el 97 Jiang Yi et al. proponen realizar cambios en la fase con lo cual analizan 4 imágenes para posteriormente realizar la obtención de la fase, siguiendo el proceso de Takeda como se menciona en [3]. La eliminación de DC, para filtrar la escena fue propuesto por Tavares, el proceso que realiza requiere de 2 escenas una con franjas deformadas verticales y otra con franjas horizontales, al obtener los espectros espaciales y sumarlos se puede identificar los límites de la componente DC para su posterior extracción, tal como lo menciona en [4]. Pedraza, et al y Abdul-Rahman presentan trabajos separados en el 2007 donde el primero realiza modificaciones mediante el análisis de discontinuidad local [5] y el segundo enfoca su tesis doctoral al análisis de franjas de Fourier y el desdoblamiento de fase [6]. En el 2008 Hussein S, et al en el 2008, analizan el tiempo de procesamiento con el algoritmo 2D-FFA y 3D-FFA que proponen.

En años recientes investigadores como Feng S, Feng L, Miguel A. Hai L, han realizado propuestas para mejorar los algoritmos de digitalización 3D por Perfilometría de Fourier, tanto del proceso general de digitalización 3D, como en [7, 8, 9, 10], así también proponen mejoras en el proceso de desdoblamiento de fase tal como se establece en [11, 12, 13, 14].

La Perfilometría por Transformada de Fourier en 2D (FTP2D) proviene de la FTP que consiste en generar y proyectar un patrón de franjas uniformes sobre una forma que será digitalizada, las franjas provocarán una escena de franjas distorsionadas, con la cual se realizara el proceso de digitalización, esta escena lleva consigo toda la información 3D del objeto.

El proceso de digitalización 3D en la FTP utiliza la configuración experimental mostrada en la figura 1, también conocida como geometría de ejes ópticos cruzados, donde los componentes básicos son un proyector, un dispositivo de captura, un equipo de cómputo para procesar los datos y un objeto a digitalizar.

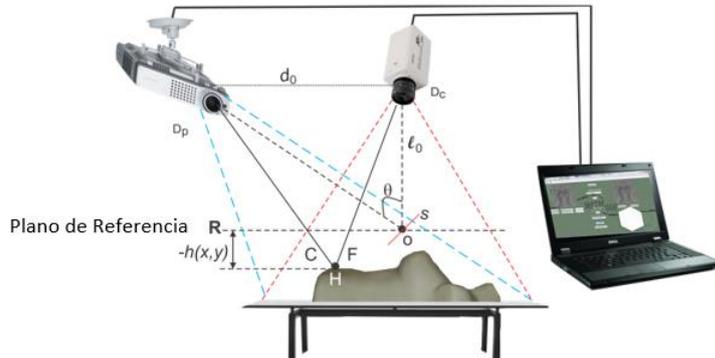


Fig. 1. Geometría de ejes ópticos cruzados usado en la FTP.

El plano de ficticio **R** mostrado en la figura 1, es utilizado de referencia para medir la altura $h(x,y)$ del objeto en particular, la lente de la cámara captura el plano deformado (escena) generado por la lente del proyector que forma su imagen conjugada en el plano R. Cuando el objeto es uniforme, la altura expresado por $h(x,y)$ es cero. La escena capturada por la cámara es expresada mediante la ecuación 1:

$$g(x,y) = a(x,y) + b(x,y) * \cos[2 * \pi f_0 x + \varphi(x,y)]. \quad (1)$$

Acorde la ecuación 1 $g(x,y)$ representa la intensidad de la imagen en el punto (x,y) , $a(x,y)$ es la iluminación del fondo, $b(x,y)$ el contraste entre las franjas claras y oscuras proyectadas, la frecuencia espacial del patrón de franjas está representada por f_0 , y $\varphi(x,y)$ corresponde a la fase que se genera por la distorsión.

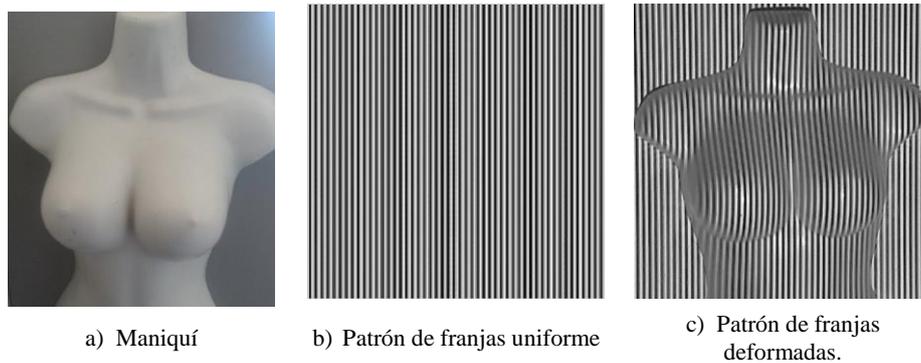


Fig. 2. Proceso de proyección de franjas.

El eje x es la línea que representa el plano R en la figura 1, siendo el eje y perpendicular a éste; Si consideramos un objeto genérico con variaciones de altura descritas por la función $h(x,y)$, el patrón de franjas capturado será un patrón de franjas distorsionado que contendrá la información 3D del objeto considerado; este

efecto se muestra en la figura 2, donde un objeto real en este caso un maniquí se observa en la figura 2 a), el patrón de franjas uniforme que se proyectara se muestra en la figura 2 b), una vez que se ha capturado la escena se obtendrá una imagen como la mostrada en la 2 c).

La ecuación 1 también puede ser expresada como la ecuación 2.

$$g(x,y) = a(x,y) + c(x,y)e^{(2\pi if_0x)} + c^*(x,y)e^{(-2\pi if_0x)}, \quad (2)$$

donde

$$c(x,y) = \frac{1}{2} b(x,y)e^{[i\varphi(x,y)]}. \quad (3)$$

De la ecuación 3 el símbolo * denota el complejo conjugado.

La Transformada de Fourier en 2 Dimensiones se puede expresar como:

$$F_2(u,v) = F\{f(x,y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)} dx dy. \quad (4)$$

Al aplicar la transformada de Fourier en dos dimensiones se obtiene una función compleja expresada por:

$$F_2(u,v) = R_2(u,v) + iI_2(u,v). \quad (5)$$

El espectro espacial de amplitud de Fourier se puede visualizar en forma de una imagen bidimensional, donde al centro se encuentra la frecuencia fundamental o componente de DC, los armónicos de la transformada de Fourier se visualizan posteriormente, tanto a la izquierda como a la derecha de la componente de DC, este fenómeno se puede apreciar en la figura 3 donde se observa una representación en 3D, del espectro bidimensional de Fourier, de igual forma en la figura 4 se muestra el espectro espacial en dos dimensiones desde una vista superior en escala de grises.

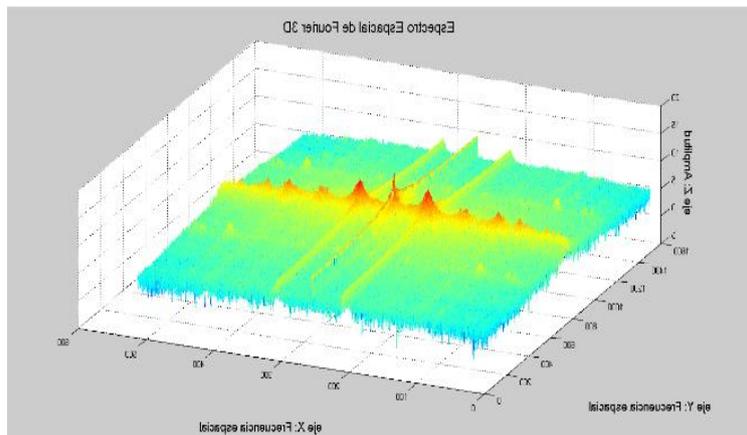


Fig. 3. Espectro espacial de Fourier vista 3D.

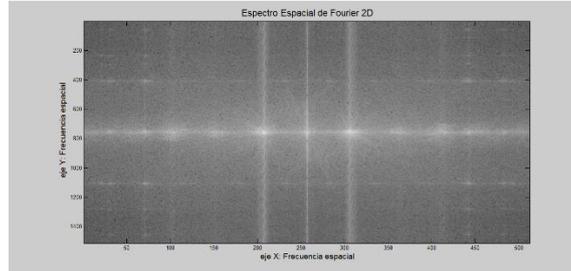


Fig. 4. Espectro espacial de Fourier vista 2D.

En este espectro se realiza el filtro 2D de la frecuencia espacial f_0 , la aplicación del filtro en esta etapa es fundamental ya que la información 3D se encuentra ubicada en esta zona. Este filtro es realizando aplicando una ventana binaria, tal como se verá más adelante. Aplicando la transformada inversa de Fourier en 2D obtenemos la función en el dominio del tiempo por lo que podemos obtener de la parte imaginaria (fase), la cual nos proporcionara los valores ortogonales, mediante la ecuación 6:

$$\log[c(x, y)] = \log\left[\left(\frac{1}{2}\right)b(x, y)\right] + i\varphi(x, y). \quad (6)$$

De la ecuación 6, se obtiene la fase $\varphi(x, y)$ en la parte imaginaria, los valores obtenidos en la fase se encuentran entre $-\pi$ a π , lo cual es conocido como valores envueltos, que no permiten visualizar adecuadamente los valores 3D del objeto por lo que es necesario realizar un proceso denominado desdoblamiento de fase, que permita recuperar la información correctamente, la fase envuelta obtenida de la ecuación 6 se puede expresar mediante la ecuación 7:

$$\varphi(x, y) = \varphi_0(x, y) + \varphi_z(x, y), \quad (7)$$

donde $\varphi_0(x, y)$ es la fase producida por el ángulo de proyección correspondiente al plano de referencia, y $\varphi_z(x, y)$ es generada por la distribución de la altura del objeto analizado. φ_z es utilizada dentro de la ecuación 8 para determinar la altura $h(x, y)$ del objeto:

$$h(x, y) = \frac{L\varphi_z(x, y)}{\varphi_z(x, y) - 2\pi f_0 d}. \quad (8)$$

1.2. Algoritmos de desdoblamiento de fase

Los algoritmos de desdoblamiento de fase han sido utilizados en las últimas décadas dentro de la metodología de la Perfilometría por Transformada de Fourier, ya que los valores de las fases obtenidos se encuentran envueltos, como ya se mencionó anteriormente, por consecuencia no proporcionan las alturas correctas. Acorde a lo anterior los algoritmos de desdoblamiento de fase tratan de encontrar los valores finales, que representarían al objeto 3D, sin embargo no son objeto de análisis del presente trabajo.

2. Desarrollo

2.1. Filtros

Los filtros en la frecuencia espacial de Fourier y particularmente en la frecuencia espacial f_0 es muy importante dentro del proceso de digitalización 3D ya que en esta zona se encuentra la información 3D del objeto, así mismo el utilizar un filtro de dimensiones pequeñas o muy grandes puede provocar que se procesen datos basura, que no corresponden con la información del objeto analizado, por ello se proponen la aplicación de los filtros incluidos en este trabajo.



Fig. 5. Filtrado del espectro bidimensional de Fourier con el Filtro Rectangular.

Para la generación de los filtros se emplean, una matriz de filtro de las mismas dimensiones que la imagen del espectro de Fourier, de tal manera que se ubica la frecuencia espacial f_0 , para aplicar sobre este punto una multiplicación de matrices, filtrando de esta forma la información adecuada; Un ejemplo de filtro rectangular se puede observar en la figura 5.

Cabe destacar que el ancho de los filtros corresponde exactamente al número de franjas proyectadas sobre el objeto, ubicando el valor de la f_0 en el centro del filtro.

El filtro elíptico se logra utilizando una máscara binaria de forma elíptica aplicada al espectro de Fourier, donde el centro de la elipse se encuentra en la posición de la frecuencia espacial f_0 , el eje menor de la elipse está determinada por el número de franjas y el eje mayor tiene sus vértices en el renglón 1 y el último renglón de la imagen analizada. La figura 6 muestra un ejemplo de aplicación del filtro elíptico.

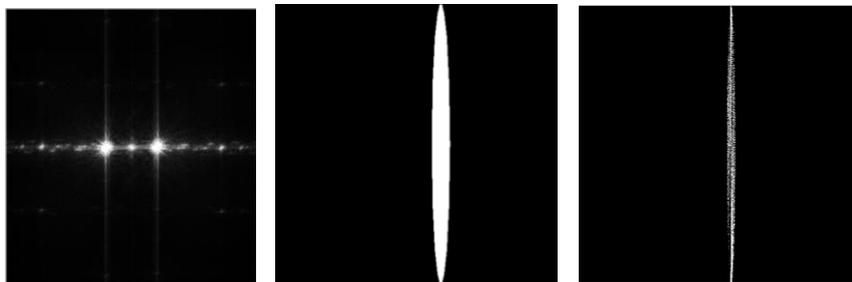


Fig. 6. Filtrado del espectro bidimensional de Fourier con el Filtro Elíptico.

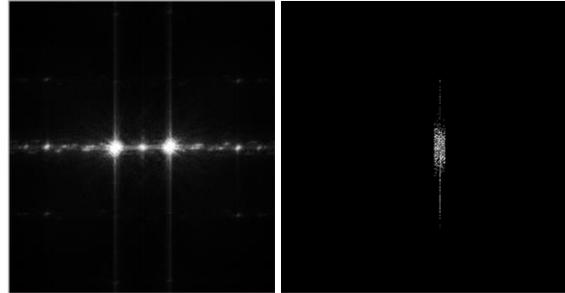


Fig. 7. Filtrado del espectro bidimensional de Fourier utilizando filtro por umbral.

El filtro por umbral se genera utilizando los márgenes del filtro rectangular, adicionalmente a este proceso se establece un valor de umbral para su filtrado, el cual está representado por la ecuación 9, donde se considera el valor de la frecuencia espacial f_0 así como el número de franjas proyectadas nf , a partir de estos datos se genera una máscara binaria, que será utilizada como filtro al multiplicarse por el espectro espacial de Fourier 2D. El resultado de aplicar el filtro por umbral se aprecia en la figura 7:

$$\text{Umbral} = \frac{f_0}{nf^2} \quad (9)$$

2.2. Algoritmo

El algoritmo utilizado, para la digitalización de objetos utilizando Transformada de Fourier Bidimensional y la aplicación de los filtros propuestos se muestra en el Algoritmo 1, los requisitos para utilizar el algoritmo es una escena de franjas deformadas, la cual puede ser obtenida desde un archivo almacenado en la memoria o capturada por medio de un dispositivo de captura, como una webcam o cualquier dispositivo de adquisición de imágenes.

Algoritmo 1. Algoritmo para la aplicación de filtros

Se requiere capturar o ingresar la escena de franjas deformadas

- 1 Capturar o ingresar la escena de franjas deformadas
- 2 Transformada de Fourier Bidimensional de la escena
- Análisis del espectro especial.**
- 3 Encontrar f_0 y $-f_0$
- 4 Determinar el número de franjas
- 5 Generar máscara de filtro
- 6 Umbral:
 - Determinar el umbral mediante la ecuación 9
- 7 Rectangular:
 - Determinar ancho de la máscara
- 8 Elíptica:
 - Determinar puntos focales de la elipse
- 9 Generar filtro para patrón de referencia en la ubicación de f_0
- 10 Aplicar Inversa de Fourier en 2 dimensiones tanto a la escena como al filtro

- 11 para el patrón de referencia.
- 12 Determinar el ángulo de fase en ambas imágenes.
- 13 Encontrar diferencia de Fases.
- 14 Desdoblamiento de fase sencillo.
- 15 Desdoblamiento de fase Graph Cut.
Mostrar resultado

3. Metodología

La metodología propuesta se observa en la figura 8 donde como primer etapa se requiere de la proyección de las franjas sobre el objeto, el cual es capturado y se almacena la escena de la figura con las franjas deformadas, una vez que se tiene la escena en la memoria, se aplica la transformada de Fourier 2D, de este proceso se obtiene el espectro espacial de Fourier en dos dimensiones y se procede a la etapa de análisis, donde se determina el número de franjas que presenta la escena, se aplica el filtro al espectro y así mismo se adquiere la posición de la frecuencia espacial f_0 y se genera una imagen de igual dimensión a la escena original.

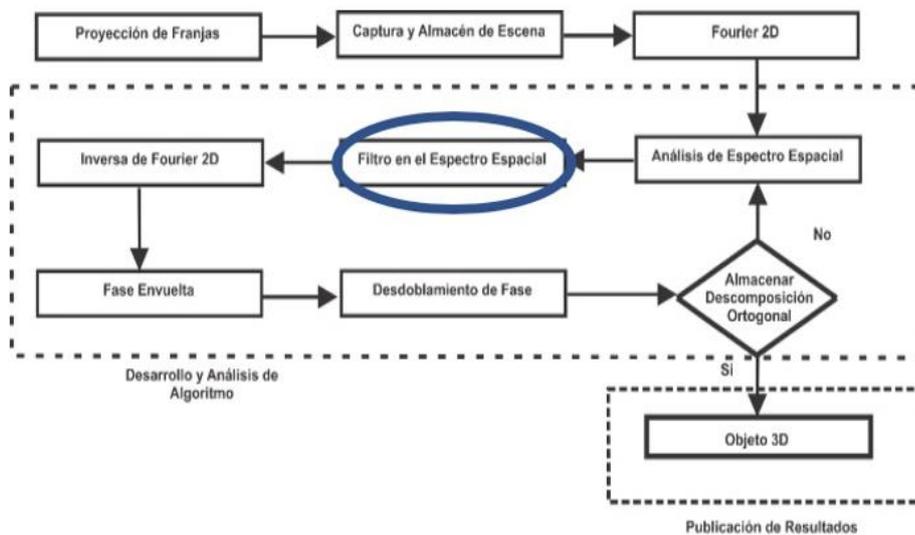


Fig. 8. Metodología propuesta.

En este trabajo se utiliza el filtro rectangular, elíptico y un filtro a base de un umbral, ya que es fundamental la aplicación de un filtro para encontrar en esta zona la información 3D del objeto. una vez que el filtro ha sido aplicado se procede a utilizar la Transformada Inversa de Fourier 2D, a ambas imágenes con lo que se obtendrá, un patrón de referencia uniforme y un patrón de franjas deformadas filtrado, posteriormente se obtiene la fase envuelta de ambas escenas y se aplica una diferencia de fases, posteriormente se utiliza el algoritmo de desdoblamiento de fase para

verificar si el resultado es satisfactorio, ya que se mostrara la descomposición ortogonal en el espacio del objeto, de ser un resultado satisfactorio se muestra el resultado en 3D y de lo contrario se procede a realizar un nuevo análisis del espectro espacial de Fourier.

4. Pruebas y resultados

En la figura 9 se muestra la escena de franjas deformadas por la forma de un maniquí, el cual es sometido al filtro por umbral, rectangular, elíptico. En la figura 10 se muestra la aplicación de estos filtros; en 10 a) se observa el espectro espacial y la zona donde será aplicado cada uno de los filtros, en 10 b) se muestra el resultado del filtro aplicado a la izquierda se muestra el filtro por umbral, al centro el rectangular y a la derecha el elíptico; En 10 c) se muestra el resultado de la inversa de la transformada de Fourier, en 10 d) se observa una vista del objeto en dos dimensiones y en 10 e) se muestra el resultado del objeto en 3D.

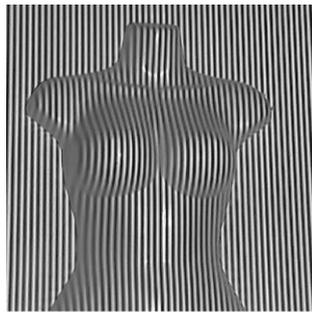
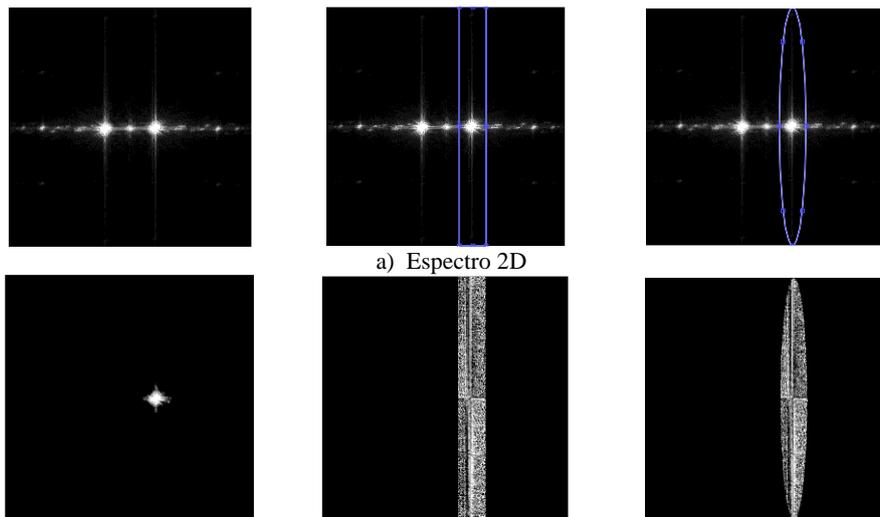


Fig. 9. Escena analizada.



a) Espectro 2D

b) Filtros (Umbral, Rectangular y elíptico)

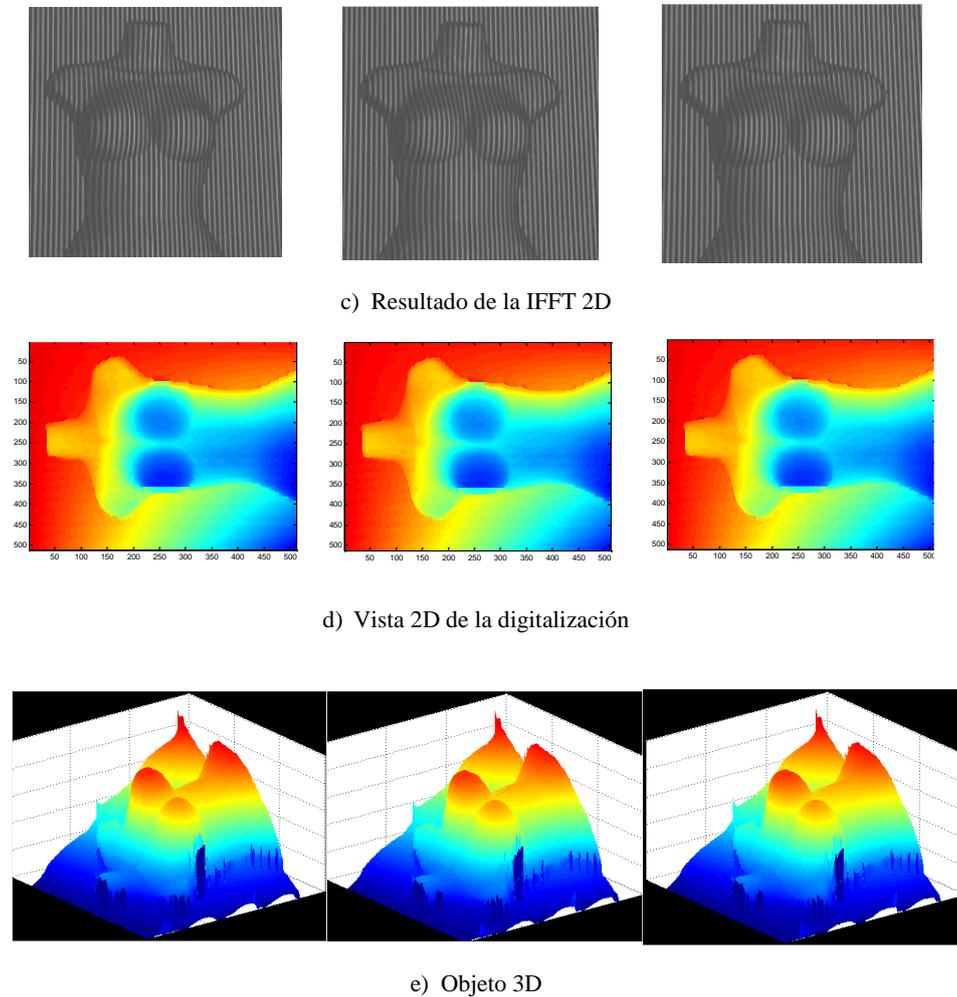


Fig. 10. Digitalización de maniquí utilizando los filtros.

En la figura 11 se muestra la escena de franjas deformadas por la forma de un murciélago, el cual es sometido al filtro por umbral, rectangular, elíptico.

En la figura 12 se muestra la aplicación de estos filtros; en 12 a) se observa el espectro espacial y la zona donde será aplicado cada uno de los filtros, en 12 b) se muestra el resultado del filtro aplicado a la izquierda se muestra el filtro por umbral, al centro el rectangular y a la derecha el elíptico.

En 12 c) se muestra el resultado de la inversa de la transformada de Fourier, en 12 d) se observa una vista del objeto en dos dimensiones y en 12 e) se muestra el resultado del objeto en 3D.

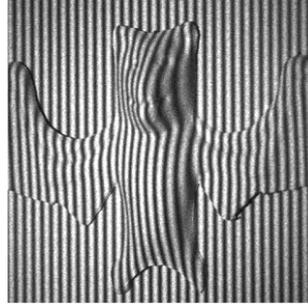
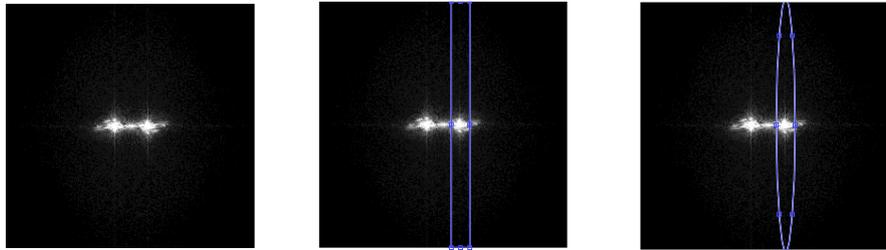
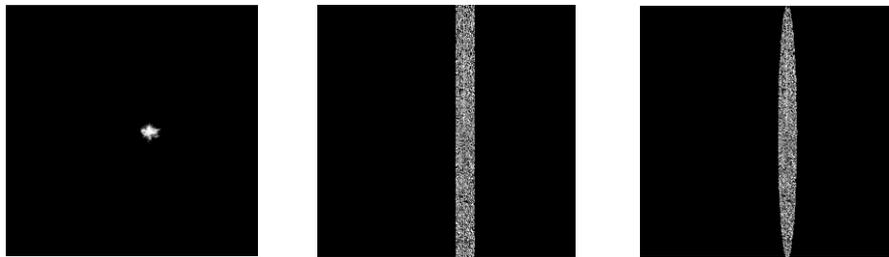


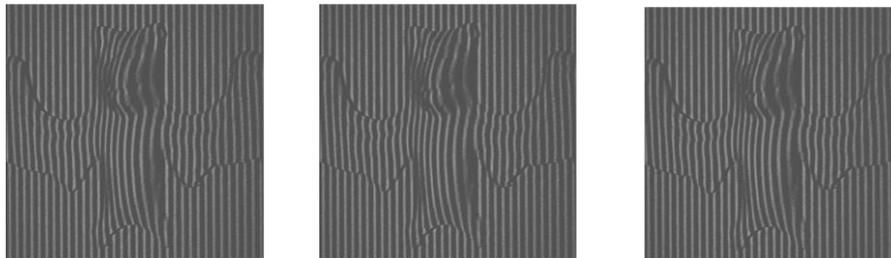
Fig. 11. Escena murcielago.



a) Espectro 2D



b) Filtros (Umbral, Rectangular y elíptico)



c) IFFT 2D

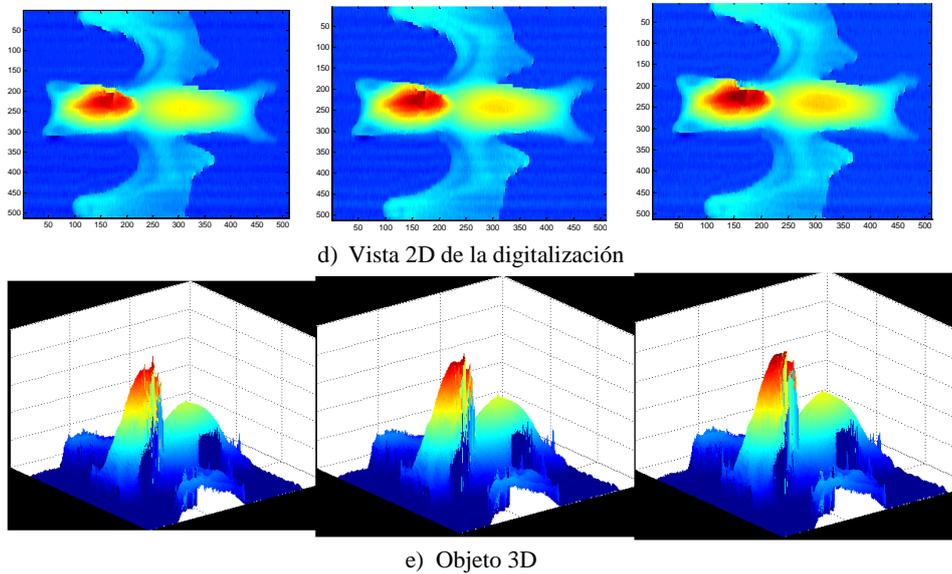


Fig. 12. Digitalización de murciélago utilizando los filtros.

5. Conclusiones

Se presentó un comparativo de tres filtros, donde los tres al terminar el proceso de digitalización 3D presentan buenos resultados ya que se aprecia correctamente las figuras a digitalizadas, aunque presentan algunas inconsistencias mínimas en ciertas zonas del objeto tales como en los bordes de los objetos y en las zonas centrales, en la zona del rostro del murciélago.

El filtro por umbral presenta la menor cantidad de inconsistencias, y valores más uniformes, el filtro rectangular muestra más inconsistencias que los resultados obtenidos con el filtro por umbral pero menos que los obtenidos por el filtro elíptico, los resultados obtenidos con el filtro elíptico presentan la mayor cantidad de inconsistencias y con valores más dispersos, se aplicó la metodología y los diferentes filtros a 10 objetos obteniendo resultados similares, por lo cual podemos concluir que cualquiera de los filtros presentados puede ser utilizado para la digitalización de objetos, sin embargo se sugiere utilizar el filtro por umbral ya que con este filtro se obtuvieron la menor cantidad de inconsistencias.

Referencias

1. Takeda, M., Hideki, I., Seiji, K.: Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J. Opt. Soc. Am.*, 72, pp. 156–160 (1982)

2. Takeda, M., Mutoh, K.: Fourier transform profilometry for the automatic measurement of 3-D object shapes. *Appl. Opt.*, 22, pp. 3977–3982 (1983)
3. Jiang, Y., Shanglian, H.: Modified fourier transform profilometry for the measurement of 3-D steep shapes. *Optics and Lasers in Engineering*, 27, pp. 493–505 (1997)
4. Paulo, J., Tavares, M. A. V.: Orthogonal projection technique for resolution enhancement of the Fourier transform fringe analysis method. *Optics Communications*, 266, pp. 465–468 (2006)
5. Pedraza, J. C., Rodriguez, J. W., Barriga, L., Gorrostieta, E., Salgado, T., Ramos, J. M., Rivas, A.: Image Processing for 3D Reconstruction Using a Modified Fourier Transform Profilometry Method. Springer Berlin / Heidelberg, MICAI: Advances in Artificial Intelligence, 705-712 ISBN 3-540-76630-8 978-3-540-76630-8 (2007)
6. Hussein, A. R.: Three-Dimensional Fourier Fringe Analysis and Phase Unwrapping. Tesis Doctoral General Engineering Research Institute (GERI), Liverpool John Moores University (2007)
7. Feng, S., Qian, C., Chao, Z., Jiasong, S., Tianyang, T.: A carrier removal technique for Fourier transform profilometry based on principal component analysis. *Optics and Lasers in Engineering*, 74, pp. 80–86 (2015)
8. Arevalillo, M., Gdeisat, M., Lilley, F., Burton, D.: A spatial algorithm to reduce phase wraps from two dimensional signals in fringe projection profilometry. *Optics and Lasers in Engineering*, 82, pp. 70–78 (2016)
9. Feng, L., Wenjing, C., Xianyu, S.: Eliminating zero spectra in Fourier transform profilometry by application of Hilbert transform. *Optics Communications*, 365, pp. 76–85 (2016)
10. Hussein, A. R., Gdeisat, M. A., Burton, D. R., Lalor, M. J., Lilley, F., Abdulbasit, A.: Three-dimensional Fourier Fringe Analysis. *Optics and Lasers in Engineering*, 46, ISSN: 0143-8166, pp. 446–455 (2008)
11. Guangliang, D., Chaorui, Z., Canlin, Z., Shuchun, S., Hui, L., Yanjie, L.: Iterative two-step temporal phase-unwrapping applied to high sensitivity three-dimensional profilometry. *Optics and Lasers in Engineering*, 79, pp. 22–28 (2016)
12. Hai, L., Xin-yu, C., Fei, W., Xiao-Tang, H., Xiao-Dong, H.: A novel algorithm based on histogram processing of reliability for two-dimensional phase unwrapping. *Optik*, 126, pp. 1640–1644 (2015)
13. Mariano, R., Francisco, J. H., Adonai, G.: Phase unwrapping by accumulation of residual maps. *Optics and Lasers in Engineering*, 64, pp. 51–58 (2015)
14. Yuan, G., Xiaotian, C., Tao, Z.: Robust phase unwrapping algorithm based on least squares. *Optics and Lasers in Engineering*, 63, pp. 25–29 (2014)

Hacia el desarrollo de una aplicación colaborativa para el diseño de diagramas de clases en ambientes educativos

S. David Díaz-García¹, Luis G. Montané-Jiménez², Betania Hernández-Ocaña³

¹ Laboratorio Nacional de Informática Avanzada,
Xalapa, México

² Universidad Veracruzana,
Facultad de Estadística e Informática, Xalapa, México

³ Universidad Juárez Autónoma de Tabasco,
Villahermosa, Tabasco, México

sdiaz.mca15@lania.edu.mx, lmontane@uv.mx, betania.hernandez@ujat.mx

Resumen. Cada vez más compañías utilizan tecnologías colaborativas para apoyar el trabajo de equipo mediante computadoras, mejorando la comunicación e interacción de manera centralizada o distribuida. En el trabajo de equipo, tener fácil acceso a información del resto del equipo, puede determinar el cumplimiento exitoso de metas. En el trabajo colaborativo, un elemento de diseño que permite compartir información es el *awareness* o consciencia de grupo. Con *awareness* es posible compartir información que ayude a entender cómo se desarrolla la actividad de equipo. En actividades colaborativas desarrolladas en ambientes educativos tradicionales, por ejemplo, diseño de diagramas de clases en un pizarrón, se dificulta identificar tareas realizadas y el grado de contribución de los miembros de equipo. Este trabajo presenta de manera conceptual una arquitectura funcional y el diseño de una aplicación colaborativa para diseñar diagramas de clases en un ambiente educativo, contemplando el uso de agentes para el mejoramiento del *awareness*.

Palabras clave: consciencia, trabajo colaborativo asistido por computadora, groupware, agentes.

Towards the Development of a Collaborative Application for Design of Class Diagrams in Educational Environments

Abstract. Nowadays, the companies use collaborative technologies to support teamwork through computers, improving communication and interaction in a centralized or distributed way. In teamwork, having easy access to information from the rest of the team, can determine the

successful accomplishment of goals. In collaborative work, a mechanism for sharing information is awareness. With awareness, it is possible to share information that helps to understand how the team activity is developed. In collaborative activities developed in traditional educational environments, e.g., designing class diagrams on a blackboard, it is difficult to identify tasks performed and the degree of contribution of team members. Therefore, in this paper we present a functional architecture and the design of a collaborative application in order to design class diagrams in an educational environment, contemplating the use of agents for the improvement of awareness.

Keywords: awareness, computer-supported cooperative work, groupware, agents.

1. Introducción

El trabajo colaborativo asistido por computadora (*Computer-Supported Cooperative Work*) es una área que se enfoca en estudiar ambientes donde personas que trabajan de forma conjunta son apoyadas con tecnología computacional, con la finalidad de desarrollar una actividad de equipo que define metas y objetivos comunes [1]. Los sistemas que apoyan el desarrollo de actividades colaborativas son denominados sistemas Groupware o Colaborativos. En la mayoría del trabajo colaborativo se requiere de las personas un grado de comunicación y cooperación con los demás, a eso se le llama trabajo de equipo. Sin embargo, todavía se han observado dificultades de coordinación y comunicación en equipos y organizaciones [2]. Estas dificultades tienden a contrarrestar los beneficios del trabajo colaborativo y la toma de decisiones [3]. En ambientes tradicionales educativos, como un escenario llevado a aulas de clases, un maestro solicita a sus alumnos realizar una tarea de equipo en un pizarrón, por ejemplo, diseñar un diagrama de clases. Realizar este tipo de tareas en un pizarrón tradicional, es decir, que no es apoyado por un sistema de información, dificulta a los participantes y al evaluador identificar la participación de todos los integrantes de equipo, además de impedir trabajar de manera fluida al no poder escribir en el pizarrón todos los participantes al mismo tiempo debido al espacio de trabajo. Para apoyar la identificación de estos niveles de contribución y ayudar en la organización de tareas de equipo en el espacio de trabajo se puede hacer uso de agentes [4,5,6].

Con la finalidad de mejorar el trabajo de equipo para la actividad mencionada y ayudar a formar una opinión social más informada de los participantes durante la actividad, se propone el desarrollo de una aplicación colaborativa para el diseño de diagramas de clases en ambientes educativos, que pueda ser implementado posteriormente en ambientes de escritorio, y finalmente adaptado con pizarras interactivas. Por lo tanto, esta propuesta inicial de aplicación permitiría trasladar una actividad educativa desarrollada en ambientes tradicionales a un ambiente digital a través de un sistema Groupware. El uso de este Groupware mantendría un contexto y consciencia sobre lo realizado por cada uno de los miembros de un equipo.

Este documento está estructurado de la siguiente forma: la Sección 2 presenta sistemas Groupware, *awareness* y sus aplicaciones, la Sección 3 describe el diseño de la aplicación colaborativa propuesta. Finalmente, la Sección 4 presenta las conclusiones y trabajo futuro.

2. Sistemas groupware y *awareness*

2.1. Sistemas groupware

Los requerimientos y problemas que enfrentan las empresas demandan esfuerzos organizados de equipo [7], para facilitar cumplir con estos requerimientos y solucionar los problemas existen los sistemas Groupware. En [8] Groupware se define como un sistema basado en computadoras que apoya el trabajo de equipo de personas que trabajan en una tarea o meta común. Los sistemas Groupware permiten a un equipo trabajar en una problemática uniendo esfuerzos apoyados por software.

Algunas áreas de aplicación de los sistemas Groupware son los ambientes de conocimiento colaborativo (p.ej. las Wikis [9]), en entretenimiento están los videojuegos *online* como *AssaultCube* [10], herramientas de desarrollo de software como CALICO [11], es un Groupware basado en pizarras interactivas que permiten diseñar *sketches* en un medio fluido, flexible y colaborativo. CALICO permite trabajar *sketches* de manera colaborativa síncrona y asíncrona, retomar trabajo, trabajar con múltiples lienzos y navegar entre ellos. El problema con estas aplicaciones es que es difuso identificar la aportación de trabajo de los participantes de equipo, además de carecer de un modelo de actividad que permita modelar los elementos que intervienen en la actividad colaborativa (p. ej. tareas, actores y la meta del equipo) para ser contabilizados.

2.2. Escenarios de aplicación

Durante el trabajo colaborativo en las interfaces de usuario pueden ocurrir 3 tipos de escenarios [12]: i) Misma tarea - Misma vista. Este escenario ocurre cuando dos o más colaboradores trabajan en la misma actividad al mismo tiempo, ellos ven lo mismo. ii) Misma tarea - Diferente vista. Sucede cuando dos o más colaboradores realizan una tarea que involucra acciones coordinadas en diferentes áreas de trabajo. iii) Misma tarea - Enfoque Mixto. Involucra actividades individuales y de equipo, y los usuarios desplazan su atención entre vistas separadas y compartidas en el espacio de trabajo.

En el trabajo colaborativo el conocimiento compartido ayuda a mejorar el trabajo de equipo [13]. Los usuarios que colaboran en equipo requieren conocer su situación de trabajo y la de sus compañeros con el fin de completar sus actividades satisfactoriamente. Por medio de *awareness* podemos representar información de la actividad del equipo, de lo que ocurre en el sistema, y compartirla entre el equipo.

2.3. *Awareness* y sus aplicaciones

Para [14] “*awareness* es información que el sistema provee sobre el estado de la colaboración en relación a la presencia de los usuarios, su ubicación dentro del sistema, nivel de actividad, acciones, cambios, objetos que se utilizan, intención etc.”. Mientras que en [15] los autores lo definen como “la percepción de los elementos del sistema con respecto al espacio y tiempo.”. Las dos definiciones plantean conceptos importantes, en este documento *awareness* está basado en la primera definición, y donde se propone un mecanismo para intercambiar información que proporcione una idea del estado de actividad de los demás compañeros de equipo y del sistema en términos de desempeño.

Usar *awareness* permite informar al equipo de la actividad que sus integrantes realizan y lo que debería hacer el equipo. Esto proporciona beneficios como mejorar la coordinación sin necesidad de comunicarse abiertamente, capacidad para pronosticar el resultado del curso de una acción, y también proporciona un entendimiento de contexto [16].

Utilizar *awareness* tiene aplicaciones en diversas áreas, p. ej. la salud, el entretenimiento o el conocimiento compartido. Un ejemplo de *awareness* se presenta en la Figura 1, esta vez en el área del entretenimiento (videojuegos *online*), en la Figura 1 (a) [17] el *awareness* muestra los compañeros y enemigos que pueden continuar jugando en la partida. En [18] se presenta un juego serio llamado “Escape from Wilson Island”, es sobre un juego colaborativo que incluye habilidades sociales como trabajo de equipo, coordinación y comunicación. En la interfaz de usuario se muestran elementos como chat, un minimapa, y atributos relevantes de los demás compañeros (ver Figura 1 (b)).



(a) Gears of war



(b) Escape from Wilson Island

Fig. 1. Ejemplo de *awareness* en el entretenimiento (videojuegos *online*).

Tele-Board [19] es un sistema colaborativo que utiliza pizarras digitales basadas en Web. En Tele-Board, el *awareness* son notas adhesivas que sirven como medio para comunicarse entre el equipo de trabajo (ver Figura 2).

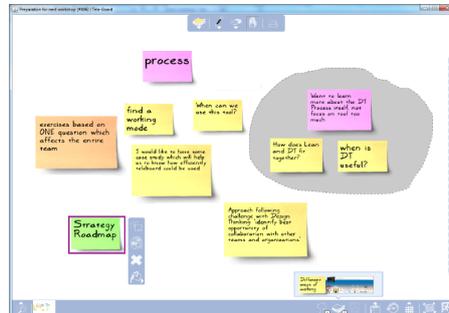


Fig. 2. Notas adhesivas que se utilizan como *awareness* en Tele-Board.

Para mejorar la interactividad de los Groupware, se plantea el uso de mecanismos de razonamiento que utilicen información de *awareness* para generar comportamientos reactivos dependiendo el estado de una actividad o de los usuarios. Esto apoyaría la creación de sistemas Groupware de siguiente generación.

2.4. Tipos de *awareness*

Los sistemas Groupware ofrecen mecanismos de *awareness*, basados en tiempo, lugar y espacio se conocen seis categorías de *awareness* en [20]: de colaboración, de ubicación, de contexto, sociales, de espacio de trabajo o de situación. De colaboración se refiere a la manera en la cual miembros de un equipo interactúan: síncronamente, asíncronamente o una mezcla de ambas. Un *awareness* de ubicación está basado en las relaciones geográficas entre colaboradores como ubicación, distancia, orientación y rango de atención. Por otra parte, *awareness* de contexto es fundamental para permitir a un equipo de colaboradores mantener un sentido de qué es lo está sucediendo en el espacio virtual. Social *awareness* comprende los roles y actividades de los demás miembros de equipo y como los miembros están contribuyendo a las tareas. Mientras de espacio de trabajo se enfatiza en un lugar que sirve para organizar tareas de acuerdo a un conjunto lógico por ejemplo un editor de equipo que sirve para organizar tareas como escribir y revisar mientras se mantiene el trabajo en la actividad que se desarrolla. Finalmente, un *awareness* de situación emerge de los esfuerzos individuales, de coordinación y colaboración. Esta clasificación es centrada en compartir información que permita conocer lo que ocurre en el ambiente de trabajo p. ej. las actividades que se realizan, el tipo de colaboración (centralizado o distribuido) y los objetos que intervienen.

En [12] se propone clasificar *awareness* en aspectos: i) sociales, contempla si una persona presta atención, su estado emocional y nivel de interés, ii) de tareas, considera el conocimiento para contemplar una tarea, herramientas necesarias, tiempo necesario y disponible, iii) de conceptos, atiende cómo una tarea se relaciona en el equipo, iv) de espacios de trabajo, relacionado a la actividad de los

demás miembros del equipo y que han hecho y donde, v) de construcción del conocimiento individual, con información de la propia actividad, la cual contempla la utilidad de las aportaciones para resolución de problemas, el tiempo necesario y disponible, vi) de construcción del conocimiento de equipo, correspondiente a información utilizada para mantener al tanto a los miembros de equipo, vii) de actividad, acerca de cómo van las cosas, cambio de planes y modificación de roles, viii) de acción, especifica avances en las tareas, que sucede, que cambios, el nivel de actividad y frecuencia en el uso de un recurso, y ix) de situación, describe minuto a minuto los cambios en el sistema. Esta clasificación se enfoca en la información que comparte el equipo y la frecuencia en el uso de los recursos. Los tipos de *awareness* analizados se presentan en la Tabla 1.

Tabla 1. Tipos de *awareness* de acuerdo a [20,12].

<i>Awareness</i>	Descripción
De colaboración	Se especializa en el trabajo centralizado o distribuido, síncrono o asíncrono.
De ubicación	Representa información de la relación geográfica entre colaboradores y su rango de atención.
De contexto	Contempla eventos en un espacio virtual y los objetos que intervienen.
Social	Proporciona información sobre roles y tareas que realizan los usuarios.
De espacio de trabajo	Centrado en el tipo de actividades (paralelas o coordinadas).
De situación	Enfocado a entender eventos, acciones, recursos y escenarios futuros.
De tareas	Aporta información sobre cómo se completarán las tareas.
De conceptos	Define cómo una actividad se relaciona con el conocimiento del equipo.
De construcción del conocimiento individual	Comparte información de la propia actividad.
De construcción del conocimiento de equipo	Se utiliza con el fin de estar al tanto de los conocimientos de otros miembros de equipo.
De actividad	Aborda aspectos de planeación para la actividad y cambios que surgen.
De acción	Describe avances en las tareas, que sucede, que cambios, el nivel de actividad y frecuencia en el uso de un recurso.

A partir de esta clasificación, se propone incluir *awareness* de tipo social y de tareas para ayudar al equipo en términos de organización y permitir al evaluador y los participantes formar una opinión más precisa de la actividad de los integrantes de equipo. Posteriormente, esto se puede implementar en diferentes escenarios colaborativos como realización de diagramas en el dominio de bases de datos relacionales, editores de texto y videojuegos. Para ayudar al equipo se podría utilizar agentes de manera similar a sistemas de tutores inteligentes [21,22,23].

3. Diseño de la aplicación

3.1. Análisis

Retomando el problema de trabajar en ambientes tradicionales educativos, se necesita una aplicación donde un usuario pueda colaborar en equipo para

construir diagramas de clases de manera colaborativa. El sistema debe permitir trabajar de manera fluida y organizada, además de aportar a los participantes y al evaluador una idea clara de la actividad de los usuarios en el equipo, esto para preparar a estudiantes de licenciatura en cursos de programación, particularmente diseño orientado a objetos.

Durante el trabajo de equipo asistido por computadoras, la comunicación y organización del equipo es mediante métodos directos (p. ej. chat, email, foro [6]), distrayendo en ocasiones a los participantes que realizan una tarea. Utilizar *awareness* en un Groupware permite mostrar información de manera indirecta al equipo y podría evitar el uso de un método directo, por ejemplo, para formar una opinión social durante el desarrollo de la actividad. Aprovechar la información proporcionada por *awareness* evita tener que escribir a otros integrantes de equipo. Hacer uso de un mecanismo de *awareness* puede evitar este tipo de retrasos permitiendo realizar una tarea en menor tiempo, facilitando de esta manera, la evaluación del producto construido colaborativamente y la participación del equipo por el uso de información de tipo social y de tareas.

La identificación de elementos necesarios para modelar aspectos de colaboración en la aplicación colaborativa, se realizó un experimento exploratorio donde se observa y analiza la actividad de estudiantes de licenciatura mientras resuelven ejercicios de diseño de software. Con esta observación se determinaron los elementos requeridos para modelar e identificar comportamientos individuales en un equipo de trabajo. Esto con la finalidad de obtener las dimensiones necesarias a incluir en un sistema Groupware de esta índole (p. ej. comunicación, cooperación, coordinación), particularmente una aplicación Groupware en diseño de software.

Para realizar este experimento exploratorio primero se indicaron las instrucciones del experimento a los usuarios (estudiantes). Después se presentó un problema de modelado de software. Se obtuvieron videgrabaciones de estudiantes de licenciatura mientras trabajaban de manera colaborativa en actividades de diseño de software (ver Figura 3). Posteriormente se analizó el comportamiento de los usuarios con el objetivo de identificar aspectos que apoyan el trabajo de equipo.



Fig. 3. Actividad colaborativa de diseño de software en un ambiente educativo.

Derivado de este análisis se propone una arquitectura funcional para la aplicación colaborativa de diseño de diagramas de clases y un prototipo que sirva como guía.

3.2. Arquitectura funcional

Para implementar la aplicación de diseño de software se diseñó una arquitectura funcional cliente-servidor basada en [24], [25], [26]. La Figura 4 muestra los componentes del cliente y del servidor de la aplicación.

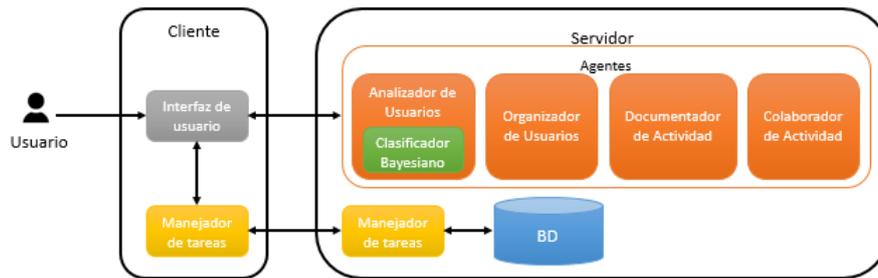


Fig. 4. Arquitectura de aplicación colaborativa de diseño de software para ambientes educativos.

A continuación, se describen los agentes propuestos que intervienen en el sistema:

- *Analizador de Usuarios*. Encargado de contabilizar las tareas de los usuarios durante la actividad de diseño de software, proporcionando información relacionada a la disponibilidad de usuarios para saber si están trabajando y cómo están trabajando. Esto con la finalidad de identificar comportamientos irregulares, por ejemplo, un usuario que construye y elimina clases para hacer creer a los compañeros de equipo que está trabajando, sin embargo, esa actividad pospone el cumplimiento de la meta que es construir el diagrama de clases completo. Entonces, el agente utiliza un historial de actividad para evaluar si el usuario está trabajando eficientemente. Además, clasifica a los usuarios de acuerdo a sus habilidades (construcción de clases, métodos, atributos, conectores), utiliza una escala de 1-5 donde 1 significa desempeño deficiente y 5 representa un desempeño competente. Para esta clasificación se podría utilizar un *clasificador bayesiano* que analice el historial de actividad de los usuarios.
- *Organizador de Usuarios*. Sugiere a los participantes de la actividad trabajar por subequipos según sus habilidades identificadas por el *Agente Analizador de Usuarios*, esto es, si el *usuario₁* es competente creando clases pero tiene habilidades deficientes en agregar atributos y el *usuario₂* por el contrario es bueno agregando atributos pero es deficiente en habilidades para crear clases, el *Agente Organizador de Usuarios* sugerirá a estos usuarios apoyarse durante la actividad, estas decisiones las toma basado en el historial de actividad de los trabajos de los usuarios realizados anteriormente.

- *Documentador de Actividad.* Proporciona un informe de la actividad al evaluador, este informe contiene el desempeño individual de los usuarios durante la actividad y una sugerencia de cómo reorganizar los equipos de trabajo. Las sugerencias para reorganizar los equipos las ofrece basándose en la información recopilada por el *Agente Analizador de Usuarios* (habilidades demostradas por los usuarios).
- *Colaborador de Actividad.* Ayuda a los usuarios mediante el uso de sugerencias o notas en la construcción del diagrama de clases. El agente colabora de acuerdo al historial de trabajo del equipo y al contexto de la actividad que se realiza [27], por lo que se requeriría un entrenamiento previo con otros usuarios.

El lado del Cliente contiene la i) Interfaz de Usuario y el ii) Manejador de Tareas. La i) Interfaz de Usuario es la ventana donde los usuarios construyen los diagramas de clases y en el componente ii) Manejador de Tareas se procesan las tareas realizadas por los usuarios, por ejemplo, crear clase, atributo, método o conector. Por otra parte, el Servidor contiene, el i) Agente Analizador de Usuarios, ii) Agente Organizador de Usuarios, iii) Agente Documentador de Actividad, iv) Agente Colaborador de Actividad, v) Manejador de Tareas, y vi) una Base de Datos. El v) El Manejador de Tareas recibe las tareas procesadas en el lado del servidor y las envía a la Base de Datos. La Base de Datos procesa peticiones de tipo *insert*, *delete*, *update* y *query*.

Con la observación realizada en el experimento exploratorio, se identificó la importancia de proponer e incluir en la arquitectura agentes para la contabilización de tareas de los participantes en el equipo, para analizar y sugerir información para apoyar la coordinación de trabajo de equipo y para emitir reportes de actividad de las personas que colaboran en la actividad [28], [29].

3.3. Interfaz de usuario cliente

Flujo de actividad

Para utilizar la aplicación colaborativa se sigue el Flujo de Actividad que presenta la Figura 5. Seguir este Flujo permite al evaluador y a los participantes de equipo aprender sobre la actividad de cada integrante del equipo además de obtener sugerencias para mejorar la interacción del equipo en términos de coordinación y optimizar tiempos de trabajo.

La aplicación cuenta con un Login (ver Figura 6) para que se identifique el usuario. Posterior de iniciar sesión, se muestra una ventana para continuar trabajando en un diagrama de clases (ver Figura 7 a)). Primero se selecciona el equipo al que pertenece el usuario y se muestran los archivos relacionados a el equipo seleccionado. Después, se selecciona un archivo para continuar trabajando en el diagrama de clases. Otra opción es crear un nuevo archivo (ver Figura 7 b)) para trabajar en un diagrama de clases, se selecciona el equipo de trabajo, se escribe el nombre del archivo de diagrama de clases y se presiona el botón "Crear".

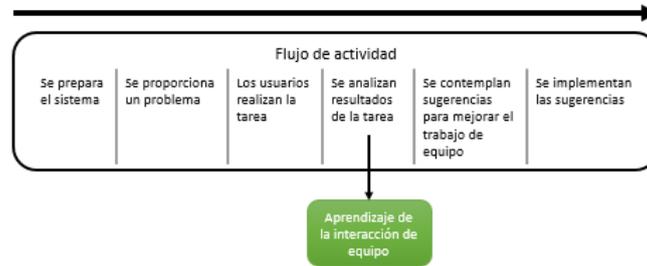


Fig. 5. Proceso para el uso de la aplicación colaborativa.

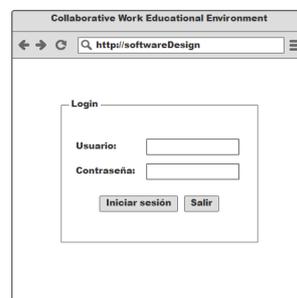


Fig. 6. Ventana para iniciar sesión en la aplicación de diseño de software.

Posteriormente de abrir un archivo existente o crear uno nuevo se muestra el entorno donde los usuarios colaboran en equipo para diseñar un diagrama de clases (ver Figura 8). En esta ventana los usuarios pueden agregar, modificar o eliminar una clase, un atributo, un método o un conector entre clases para construir un diagrama de clases. Además, pueden ver los usuarios que se encuentran colaborando en la actividad, y las tareas que cada uno realiza.

Un ejemplo de sugerencia de colaboración del *Agente Organizador de Usuarios* se muestra en el rectángulo 1) de la Figura 8, en esta sugerencia el *Agente Organizador de Usuarios* invita al *usuario₂* a ayudar al *usuario₁*.

Por último, la Figura 9 presenta un informe generado por el *Agente Documentador de Actividad*. Este informe tiene el objetivo de ayudar al evaluador a comprender como se han comportado los usuarios durante la actividad, utilizando una escala de 1-5 donde 1 significa desempeño deficiente y 5 representa un desempeño competente.

Para generar este informe, los agentes utilizarían información de las tareas realizadas en comparación con los participantes en el equipo. Apoyando de esta manera el *awareness* de tipo Social.

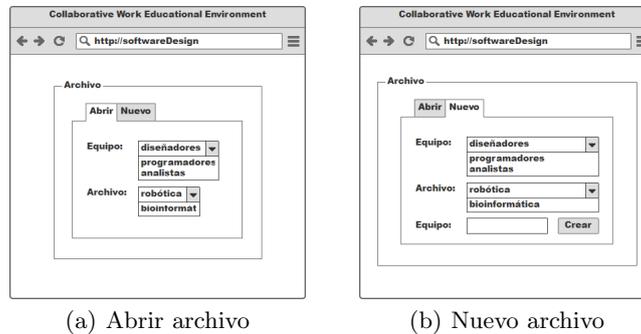


Fig. 7. a) Ventana para continuar trabajando en un diagrama de clases. b) Ventana para trabajar en un nuevo diagrama de clases.

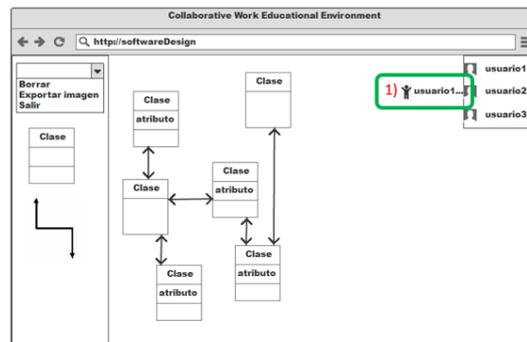


Fig. 8. Ventana para construir diagramas de clases de manera colaborativa en ambiente educativo.

4. Conclusiones y trabajo futuro

En este trabajo, se ha presentado un diseño inicial de una aplicación colaborativa de diseño de software en un ambiente educativo de manera conceptual. La finalidad de la aplicación es facilitar a los usuarios y a la persona que evalúa la actividad identificar como colaboran en equipo los usuarios cuando realizan una tarea. Con la arquitectura propuesta será posible llevar un registro de las tareas y acciones realizadas por cada integrante de equipo.

En etapas futuras de este trabajo, se espera finalizar la implementación de la arquitectura y aplicación propuesta en este artículo. Con la finalidad de incluir un mecanismo que mida el desempeño y contribución de los usuarios en relación a las actividades que desarrollan, esta medida es conocida como presencia social [3]. Con esta medición, se incluirán agentes, un agente que presente la presencia social de acuerdo al estilo de trabajo de los participantes (mediante una gráfica o conjunto de objetos/símbolos relacionados al contexto de la actividad). Este

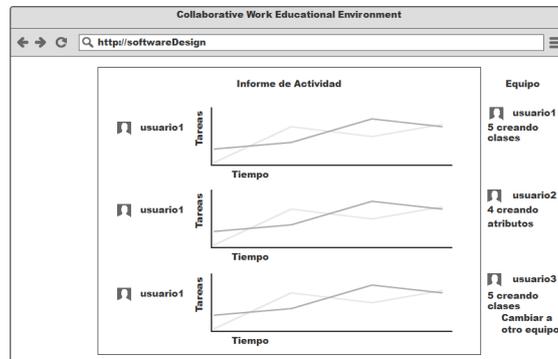


Fig. 9. Informe que muestra el historial de tareas en el tiempo durante la actividad.

agente podría ser implementado con la finalidad de proporcionar información que facilite a los usuarios completar la actividad en menor tiempo. Finalmente, incorporar otro agente para aportar sugerencias sobre como presentar la presencia social (ubicación de la información, tiempo para mostrar la información y cantidad de información a mostrar). Por lo tanto, la perspectiva final de nuestro trabajo es ofrecer información que facilite la actividad colaborativa y que ayude a formar una opinión social en los integrantes de equipo y el evaluador.

Agradecimientos. Los autores agradecen a los revisores de este documento sus útiles comentarios. El primer autor agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca otorgada para realizar estudios de posgrado (No. de beca: 413500).

Referencias

1. Palmer, T.D., Fields, N.A.: Computer supported cooperative work. *Computer*, 27 (5):15–17 (1994)
2. Galegher, J., Kraut, R.E., Egido, C.: *Intellectual teamwork: Social and technological foundations of cooperative work*. Psychology Press (2014)
3. Montané-Jiménez, L. G., Benítez-Guerrero, E., Mezura-Godoy, C., Pino, J. A.: Measuring Social Presence in Groupware Systems. In: *Proceedings of the International Conference on CSCW in Design (CSCWD'2015)*, Calabria, pp. 200-205 (2015)
4. Presentación de la Monografía: Técnicas de la inteligencia artificial aplicadas a la educación. *IA, Revista Iberoamericana de Inteligencia Artificial*, (33):7–12 (2007)
5. Flores Quisbert, E.: Agentes inteligentes: El siguiente paso en la inteligencia artificial. *Revista de Información, Tecnología y Sociedad*, pp. 106 (2008)
6. Costaguta, R.: Una revisión de desarrollos inteligentes para aprendizaje colaborativo soportado por computadora. *Revista Ingeniería Informática*, 13:2013–2015 (2006)

7. Chaffey, D.: Groupware, workflow, and intranets: reengineering the enterprise with collaborative software. Gulf Professional Publishing (1998)
8. Ellis, C.A., Gibbs, S.J., Rein, G.: Groupware: some issues and experiences. *Communications of the ACM*, 34(1):39–58 (1991)
9. Xiao, W.P., Chi, C.Y., Yang, M.: On-line collaborative software development via wiki. In: *Proceedings of the 2007 international symposium on Wikis*, pp. 177–183, ACM (2007)
10. AssaultCube: What is assaultcube? <http://assault.cubers.net/> (2013)
11. Mangano, N., LaToza, T.D., Petre, M., van der-Hoek, A.: Supporting informal design with interactive whiteboards. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 331–340, ACM (2014)
12. Herrera, A., Rodríguez, D., García-Martínez, R.: Taxonomía de mecanismos de awareness. In: *XVIII Congreso Argentino de Ciencias de la Computación* (2013)
13. Belkadi, F., Bonjour, E., Camargo, M., Troussier, N., Eynard, B.: A situation model to support awareness in collaborative design. *International Journal of Human-Computer Studies*, 71(1):110–129 (2013)
14. Bibbó, L. M.: Modelado de sistemas colaborativos. PhD thesis, Facultad de Informática (2009)
15. Galdámez, N., Sanz, C.V., De-Giusti, A. E.: Diseño de un entorno web colaborativo orientado al ámbito educativo para desarrollar la técnica de brainstorming. In: *XVII Congreso Argentino de Ciencias de la Computación* (2011)
16. Robert, L. McLeod, A., Davis, A. R. Erp configuration: Does situation awareness impact team performance? In: *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, pp. 1–8, IEEE (2011)
17. Epic Games: Gears of War. <https://gearsofwar.com/> (2006)
18. Wendel, V., Gutjahr, M., Göbel, S., Steinmetz, R.: Designing collaborative multiplayer serious games. *Education and Information Technologies*, 18(2):287–308 (2013)
19. Gumienny, R., Gericke, L., Wenzel, M., Meinel, C.: Supporting creative collaboration in globally distributed companies. In: *Proceedings of the 2013 conference on CSCW*, pp. 995–1007, ACM (2013)
20. Antunes, P. Herskovic, V., Ochoa, S. F., Pino, J. A.: Reviewing the quality of awareness support in collaborative applications. *Journal of Systems and Software*, 89:146–169 (2014)
21. Cataldi, Z., Salgueiro, F.A., Lage, F.J., Martínez, R. G.: Sistemas tutores inteligentes: los estilos del estudiante para selección del tutorizado. In: *VII Workshop de Investigadores en Ciencias de la Computación* (2005)
22. Peña, C. I., Marzo, J. L., De la Rosa, J. L., Fabregat, R.: Un sistema de tutoría inteligente adaptativo considerando estilos de aprendizaje. *Revista UIS ingenierías*, 1(2) (2012)
23. Cataldi, Z., Lage, F. J.: Modelado del estudiante en sistemas tutores inteligentes. *TE & ET* (2010)
24. Benítez-Guerrero, E., Mezura-Godoy, C., Montané-Jiménez, L. G.: Context-aware mobile collaborative systems: Conceptual modeling and case study. *Sensors*, 12(10):13491–13507 (2012)
25. Carrascosa, C., Julián, V., Soler, J.: Una arquitectura de sistema multi-agente para la recuperación y presentación de la información. In: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: actas del IV Congreso ISKO-España EOCONSID'99*, 22-24 de abril de 1999, Granada, pp. 291–296, Universidad de Granada (1999)

26. Gil, A., García, F.: Un sistema multiagente de recuperación de objetos de aprendizaje con atributos de contexto. In: Workshop on Agents and Multi-agent Systems for Enterprise Integration (ZOCO 2007) at CAEPIA, pp. 1–11 (2007)
27. Baylor, A. L., Kim, Y.: Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15(2):95–115 (2005)
28. Hurtatiz, Y. E. M., Rengifo, Y. S. P., Rojas, E. E. M.: Sistemas tutores inteligentes como apoyo en el proceso de aprendizaje. *Redes de Ingeniería*, 6(1):25–44 (2015)
29. Hernández, J. A. D., Rengifo, Y. S. P.: Los sistemas tutores inteligentes y su aplicabilidad en la educación. *Revista horizontes pedagógicos*, 17(2):104–116 (2016)

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
noviembre de 2017
Printing 500 / Edición 500 ejemplares

