

Modelado computacional de habilidades lingüísticas y visuales

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)
Jesús Figueroa (Mexico)
Alexander Gelbukh (Russia)
Ioannis Kakadiaris (USA)
Serguei Levachkine (Russia)
Petros Maragos (Greece)
Julian Padget (UK)
Mateo Valero (Spain)

Editorial Coordination:

Socorro Méndez Lemus

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 74**, mayo 2014. 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 74**, May 2014. 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.

Volume 74

Modelado computacional de habilidades lingüísticas y visuales

Grigori Sidorov (Ed.)



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



Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2014

ISSN: 1870-4069

Copyright © Instituto Politécnico Nacional 2014

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 and Periodica / Indexada en LATINDEX y Periódica

Printing: 500 / Tiraje: 500

Printed in Mexico / Impreso en México

Prefacio

El propósito de este volumen es reflejar las nuevas direcciones de investigación y los avances en el área del modelado computacional de las habilidades humanas —en este caso, el manejo de lenguaje natural y la visión— que tradicionalmente se modelan en área de la computación llamada Inteligencia Artificial.

Los artículos de este volumen fueron seleccionados con base en un estricto proceso de revisión efectuada por los miembros del Comité de revisión (revisores), tomando en cuenta la originalidad, aportación y calidad técnica de los mismos. Cada artículo fue revisado por lo menos por dos miembros del Comité de revisión del volumen.

Este volumen contiene 13 artículos relacionados con varios aspectos de la lingüística computacional y la visión artificial.

Este volumen de la revista puede ser interesante para los investigadores y estudiantes de las ciencias de la computación, especialmente en áreas relacionadas con la lingüística computacional y la inteligencia artificial y su aplicación a los diferentes ámbitos de la vida cotidiana; así como, para el público en general interesado en estos fascinantes temas.

En este número especial de la revista RCS, a nombre de la comunidad del programa educativo de Ingeniero en Computación del CU UAEM Zumpango expresamos nuestro agradecimiento al Dr. Jorge Olvera García, Rector de la Universidad Autónoma del Estado de México (UAEM) y al M. en A. Rodolfo Téllez Cuevas, encargado del despacho de la dirección del Centro Universitario UAEM Zumpango, por apoyar de manera ingente la investigación, y el desarrollo de la ciencia y la tecnología, sustentado todo ello en un humanismo que transforma.

El proceso de revisión y selección de artículos se llevó a cabo usando el sistema libremente disponible EasyChair, www.EasyChair.org.

Grigori Sidorov
Mayo 2014

Table of Contents

Page

Análisis empírico de la dispersión del español mexicano.....	9
<i>Orlando Ramos, David Pinto, Belem Priego, Iván Olmos, Beatriz Beltrán</i>	
Detectando la prioridad de contenidos generados en Twitter por medio de <i>n</i> -gramas de palabras	21
<i>Gerardo García-Rodríguez, Gabriela Ramírez-de-la-Rosa, Esaú Villatoro-Tello, Héctor Jiménez-Salazar</i>	
Diseño de agentes para recuperar información para el enriquecimiento de ontologías dirigidas a epidemiología: el caso de la tuberculosis.....	33
<i>Joanna Alvarado-Uribe, Ari Yair Barrera-Animas, Miguel González-Mendoza, Ma. de los Ángeles Junco-Rey</i>	
Recomendaciones contextuales basadas en el enfoque de post-filtrado	47
<i>Xochilt Ramírez-García, Mario García-Valdéz</i>	
Sonificación de EEG para la clasificación de palabras no pronunciadas.....	61
<i>Erick Fernando González-Castañeda, Alejandro Antonio Torres-García, Carlos Alberto Reyes-García, Luis Villaseñor-Pineda</i>	
Software decodificador de texto a voz para ayuda a personas con capacidades diferentes	73
<i>Luis Javier Mona Peña, Joaquín Antonio Alvarado Bustos, Octavio Mendez Hernandez</i>	
Is there Hope for Interlingua methods? A CLIR Comparison Experiment between Interlingua and Query Translation.....	81
<i>Marta R. Costa-jussà, Rafael E. Banchs</i>	
Acoplamiento de micro multitudes para el desarrollo de videojuegos controlados por movimiento	89
<i>Iván Rivalcoba, Krelly Rodríguez, Oriam Degives, Isaac Rudomín</i>	
Procesos de decisión de Markov y microescenarios para navegación y evasión de colisiones para multitudes.....	103
<i>Sergio Ruiz, Benjamín Hernández</i>	

Sistema inteligente para procesamiento de imágenes en control de calidad basado en el modelo difuso singleton tipo 1	117
<i>Pascual Noradino Montes Dorantes, Juan Pablo Nieto González, Gerardo Maximiliano Méndez, Rolando Praga Alejo, Karla Lizette Guajardo Cosío</i>	
Segmentación automática en imágenes RGB aplicando la técnica Fuzzy C-means de la morfología matemática para la ayuda de la fotoidentificación de cetáceos	131
<i>Cesar Muñoz Pérez, Diego Cabrera Padilla, B.E. Carvajal-Gámez, F.J. Gallegos-Funes, Diane Gendron</i>	
Descriptores basados en combinaciones lineales de polinomios de Jacobi aplicados en la caracterización de melanoma, comparados estadísticamente con los momentos de Hu	143
<i>Damián A. Sánchez-Trasgallo, Martín Carpio, R. Montero, Luis E. Mancilla Espinoza, Miguel A. Casillas Araiza</i>	
Implementación de un sistema de reconocimiento de imágenes por contenido usando algoritmos genéticos	157
<i>Juan Villegas-Cortez, Yolanda Pérez-Pimentel, Ismael Osuna-Galán</i>	

Análisis empírico de la dispersión del español mexicano

Orlando Ramos¹, David Pinto¹, Belem Priego^{1,2},
Iván Olmos¹, Beatriz Beltrán¹

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

² Laboratoire de Lexiques, Dictionnaires et Informatique,
Université Paris 13, Francia

{orlandxrf, belemprs, ivanoprkl}@gmail.com, {dpinto, bbeltran}@cs.buap.mx

Resumen. En este artículo se presenta un sistema que pretende facilitar el análisis de la dispersión del español mexicano. Se presentan gráficas resultantes, así como los modelos del sistema. El objetivo es mostrar el avance del sistema y su posible aplicación en el cálculo de la dispersión del lenguaje para otros idiomas. Los experimentos fueron realizados sobre dos tipos de corpora: noticias y tweets.

Palabras clave: Dispersión del idioma, noticias, tweets.

1. Introducción

En el idioma español existen rasgos y características que distinguen a un pueblo de otro de una manera muy particular, ya sea el español de España en donde su pronunciación y significado varían en comparación al de Latinoamérica, especialmente al español mexicano. México es un pueblo rico en cultura y tradiciones, y esto precisamente es lo que hace, que en cada una de sus entidades federativas se encuentren frases o palabras que los distingan de una manera en particular, estas pequeñas diferencias son las que dan pie a esta investigación, para lograr identificar las regiones de nuestro país en donde se dan estas variaciones.

Como objetivo general nos hemos planteado analizar el uso del idioma español en la República Mexicana y su posible dispersión de acuerdo a la ubicación geográfica. Así, nuestros objetivos específicos son los siguientes:

1. Construir un corpus etiquetado geográficamente del español usado en la República Mexicana.
2. Estudiar diversos métodos para la identificación automática de la dispersión en el uso del lenguaje natural.
3. Construir un mecanismo de visualización para el uso del idioma, de acuerdo a la ubicación geográfica.
4. Evaluar los resultados obtenidos en base a métricas estándar tales como precisión y recall.

5. Comprobar si existe o no una marcada dispersión en el uso del español mexicano de acuerdo a la ubicación geográfica.

En particular, hemos definido la siguiente hipótesis que debe ser evaluada: Consideramos que el español mexicano ha sufrido un fenómeno de dispersión en cuanto a la agregación o modificación del vocabulario en función de la ubicación geográfica. Por tanto, se formulan la siguiente hipótesis: -existe una variabilidad significativa y consistente con la ubicación geográfica en el uso del español mexicano.

Para observar si la hipótesis es correcta, hemos diseñado un sistema basado en el web que permite observar la dispersión del idioma a través de toda la República Mexicana.

2. Trabajo relacionado

En esta sección se introducen algunos trabajos que se encuentran de alguna manera relacionados con el análisis del lenguaje. Si bien, algunos son análisis sintácticos, morfológicos o puramente léxicos, sirven como punto de partida para definir estrategias en el análisis de la dispersión del español en la República Mexicana.

En [7] se hace uso de los datos sobre diversidad de fonemas para estimar la fecha de origen del lenguaje. Esta diversidad fonética denota el número de unidades perceptualmente distintas de sonido (consonantes, vocales y tonos) en un lenguaje, dado que la diversidad de fonemas varía considerablemente entre las lenguas, y varios idiomas funcionan con un número limitado de fonemas. La diversidad de fonemas también se correlaciona positivamente con el número de idiomas que rodean, lo que sugiere que los fonemas, como otros rasgos culturales, se pueden pedir prestados. La diversidad de fonemas de un lenguaje depende del tamaño de la población de los hablantes, el área geográfica sobre la cual se habla el idioma, y la diversidad lingüística local.

En [6] se estudia la competencia entre los lenguajes o los rasgos culturales de dos poblaciones que se difunden en la misma zona geográfica combinando el modelo de competencia del lenguaje de Abrams-Strogatz (AS), y un modelo de dispersión humana en un sustrato no homogéneo. Por “competencia” se entiende que en cualquier momento hablantes pueden cambiar a otro lenguaje, como consecuencia de la interacción entre los hablantes de la lengua 1 y 2.

En [1] se explica cómo la diversidad lingüística y las bases biológicas del lenguaje han sido tradicionalmente tratados por separado. Algunos debates han generado una propuesta que argumenta a favor de un sistema biológico para usos específicos, por analogía como el sistema visual. Otra propuesta es que el lenguaje en cambio, confía en los mecanismos neuronales de dominio general evolucionando para otros fines. Sin embargo, existe un mayor acuerdo sobre el origen de la diversidad lingüística, que normalmente se atribuye a la evolución cultural divergente siguiendo la migración humana. En este trabajo se muestra un modelo teórico de la relación entre la diversidad lingüística y la base biológica para el lenguaje, que consiste en analizar el número de agentes hablantes de un

lenguaje y de cuantos principios cuenta, el conjunto de genes con que cada individuo está dotado y los tres alelos con los que cuenta cada gen.

En [2] se analiza cómo los estudios de lingüística, cultura y sociolingüística, geografía y dialectología es lo que permite un estudio espacial del lenguaje en su contexto geográfico, además de social y cultural. El estrecho vínculo de la dialectología con la necesidad del uso de mapas se ha centrado en la dimensión diatópica (fenómenos que se producen en una lengua en virtud de su extensión geográfica) de la lengua. La dialectología estudia las variaciones de una lengua según los lugares, y la geografía lingüística es uno de los métodos para espacializar y reconocer estas variaciones en cartografías y mapas. Según esto, en la lingüística, y en la elaboración de lo que se conoce como atlas lingüísticos, la dimensión espacial ha estado presente y ha sido reconocida para visualizar distribuciones espaciales. Sin embargo, la manera de enfocar los estudios espaciales en la lingüística antes de los sesenta fue la de representar aspectos lingüísticos en escenarios regionales o desarrollar técnicas cuantitativas y estadísticas de recolección de información en mapas, lo cual ha dado lugar a lo que en conjunto se conoce como geografía lingüística. Así, los estudios regionales clásicos desarrollan una recolección de información en campo presentando esto en cartas geográficas que permiten ver la distribución espacial de los hablantes regionales, que se integran en el hablar común de una nación, facilitando la delimitación de zonas dialectales.

Aunque existen varios estudios que realizan análisis morfológico del español, tal vez uno de los más completos ha sido realizado por Gelbukh y Sidorov [5,3,4]. Ellos plantean un sistema computacional que analiza el español usando un modelo denominado “de generación”. Se plantea en dicho artículo que este modelo es capaz de obtener mejores resultados para a lenguajes con alternaciones irregulares de raíz, como es el caso del español. El modelo consiste en un conjunto de reglas para obtener todas las raíces de una forma de palabra para cada lexema, su almacenamiento en el diccionario, la producción de todas las hipótesis posibles durante el análisis y su comprobación a través de la generación morfológica.

Quizás el trabajo más relacionado es el presentado en [8], en donde se estudian diferentes variedades del español. En particular para los siguientes países: México, Argentina, Perú y España. Se hace uso de textos noticieros también. Desde nuestro particular punto de vista, este trabajo puede servir como base para analizar la dispersión del español en México, teniendo en cuenta que de alguna manera esta última tarea (la nuestra) presenta una mayor grado de complejidad que aquella presentada en [8].

3. Análisis y Diseño del Sistema

El sistema desarrollado considera el cálculo de ngramas como un medio de determinación del grado de dispersión del idioma. Se usan n -gramas de letras y n -gramas de palabras. Estos n -gramas son calculados por regiones geográficas (estados de la República Mexicana) y comparados con respecto a la media na-

cional. En general, se calcula la frecuencia de cada n -grama (unigrama, bigrama o trigramas) y se ordenan los n -gramas en forma descendente. Se usa una porción de los n -gramas más frecuentes y se calcula el grado de intersección entre los n -gramas calculados a nivel nacional y aquellos calculados a nivel estatal. El grado de traslape indica la cercanía del idioma en el estado con respecto a la media nacional.

Un primer Corpus que se usó para este análisis está conformado por notas periodísticas publicadas digitalmente por la Organización Editorial Mexicana (OEM) para cada una de las entidades federativas de la República Mexicana, con más de medio millón de noticias. Si bien, este corpus muestra información perteneciente a cada estado, es muy probable que el modelado basado en n -gramas represente el estilo de escritura de un reportero, más que el estilo de escritura de la población en general. Por esta razón, hemos recopilado un segundo corpus basado en Tweets.

Este segundo Corpus se extrajo de Twitter, con aproximadamente medio millón de tweets. La forma de extracción fue a través de geo-coordenadas de las capitales de todas las entidades federativas de la República Mexicana. Se buscaron todos aquellos tweets que estuviesen localizados en un radio de 10 Kilómetros alrededor de cada capital. Consideramos que este corpus es mucho más representativo del lenguaje que el corpus de noticias, aunque tiene las particularidades asociadas a los Tweets: textos cortos y con un vocabulario ligeramente modificado (eliminación de vocales y errores ortográficos).

En la Figura 1 se muestran los casos de uso del sistema propuesto. Los diagramas de procesos y modelos de procesos se muestran en la Figuras 2 y 3, respectivamente.

En general, en el sistema se ha buscado poder mostrar la dispersión del idioma (español mexicano) usando diversos mecanismos de visualización. Se usan tablas de resultados, gráficas y mapas. De momento, los dos corpora usados son estáticos, pero como trabajo a mediano plazo consideramos implementar un módulo que permita subir documentos al sistema para que sean procesados automáticamente.

4. Resultados

El sistema desarrollado permite visualizar en base a tablas, gráficas y mapas el grado de dispersión del español en la República Mexicana. Hemos denominado a este sistema AEDEM por ser las siglas de Análisis Empírico de la Dispersión del español mexicano (ver Figura 4).

En la Figura 5 se muestra una tabla que indica la cantidad de trigramas de letras que se comparten entre cada estado y la media nacional. El cálculo es realizado usando el corpus de Tweets como datos de entrada. La tercera columna muestra la cantidad de Tweets por estado, y las columnas siguientes muestran la cantidad de n -gramas que comparten con la media nacional. Si el usuario coloca el cursor sobre un valor se mostrará el porcentaje de cobertura.

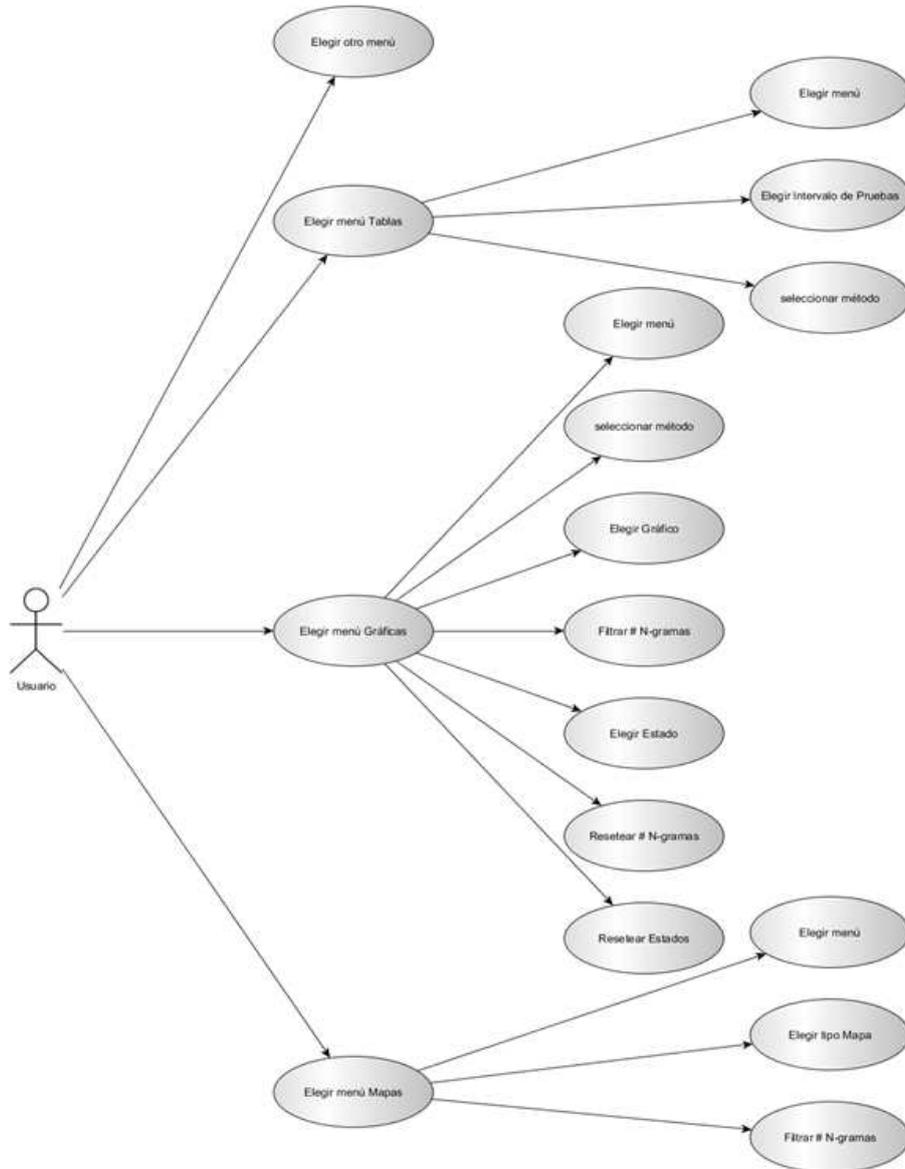


Fig. 1. Casos de uso del sistema AEDEM

En la Figura 6 se presenta una gráfica comparativa entre tres estados de la República Mexicana y su grado de cobertura con respecto a la media nacional. El sistema permite hacer este tipo de comparaciones entre cualquier número de estados.

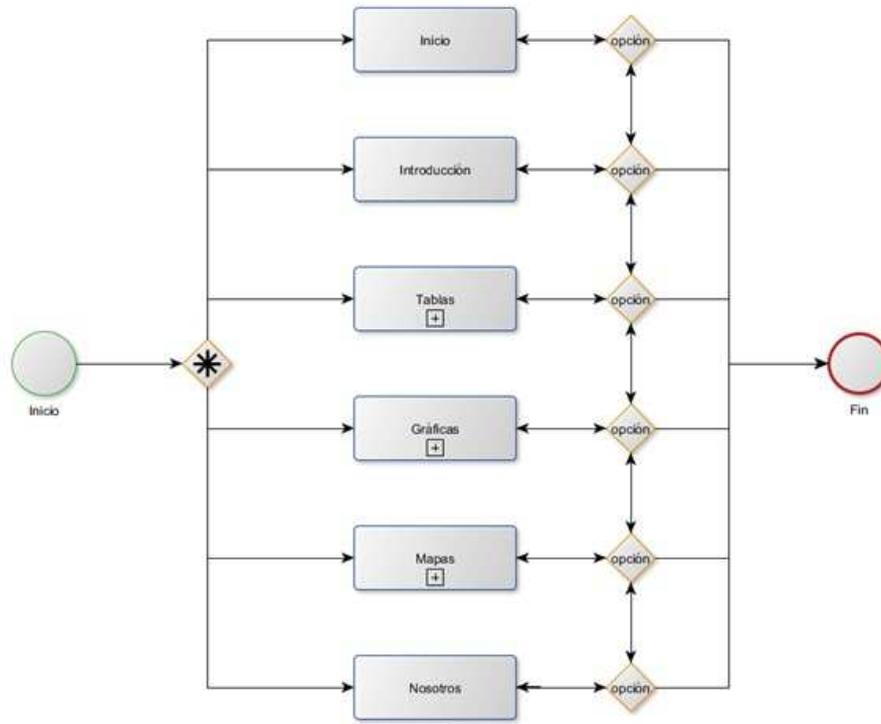


Fig. 2. Diagrama de procesos

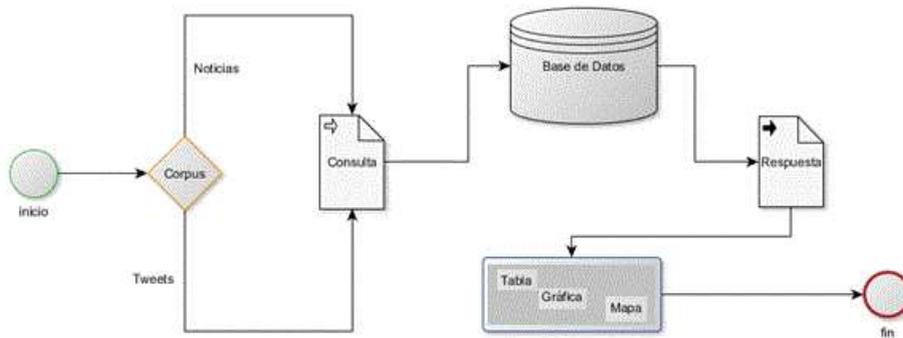


Fig. 3. Modelo de procesos

Observe en las Figuras 7 y 8, cómo ciertos estados tienen un color mucho más oscuro, lo que significa que se acercan mucho más a la media nacional. Mientras que otros estados muestran una coloración mucho más clara, lo cual significa que los modelos del lenguaje están más alejados de la media nacional.



Análisis Empírico de la Dispersión del Español Mexicano

AEDEM es un proyecto investigación para identificar la posible dispersión y similitudes del idioma español de México, con ayuda de una de las ramas de la Inteligencia Artificial, que es el Procesamiento de Lenguaje Natural. Para realizar esta tarea se utilizaron dos Corpus (de Noticias Periodísticas y de Tweets) con por lo menos medio millón de documentos.



Tablas

Después de realizar las pruebas con los métodos: trigramas de letras, uni-gramas de palabras, bi-gramas de palabras y uni-gramas de POS se muestran los resultados en tablas para analizar el comportamiento del idioma español mexicano.



Gráficas

Los resultados obtenidos también son presentados en gráficas para visualizar el comportamiento de los Estados de la República Mexicana, de forma individual, comparados unos con otros y todos a la vez.



Geo-gráficas

Además los resultados obtenidos se visualizan en geo-gráficas es decir, en mapas de la República Mexicana en donde se puede apreciar, el comportamiento del lenguaje en cada una de las entidades federativas de México.

Fig. 4. Pantalla principal del sistema AEDEM

Tablas - Trigramas de letras Tweets

Seleccione las opciones de filtrado

Visualizar Método: Trigramas de letras Unigramas de palabras Bigramas de palabras Unigramas de POS

Rango de Pruebas: 10 - 100 100 - 1000 1000 - 10000

No.	Estados	Tweets	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
1	Aguascalientes	14.426	933	1842	2676	3464	4184	4885	5596	6306	7016	7726
2	Baja California Norte	13.645	931	1816	2634	3446	86.60 % 477	4879	5666	6453	7240	8026
3	Baja California Sur	11.145	923	1799	2595	3376	4083	4797	5585	6372	7159	7946
4	Campeche	15.497	930	1828	2633	3413	4166	4853	5640	6427	7214	8001
5	Chihuahua	12.653	928	1830	2666	3454	4202	4879	5666	6453	7240	8026
6	Chiapas	12.440	935	1830	2660	3422	4129	4818	5605	6392	7179	7966
7	Coahuila	13.582	897	1759	2605	3403	4173	4897	5670	6443	7216	7989
8	Colima	14.498	916	1805	2653	3453	4192	4877	5664	6451	7238	8025
9	Distrito Federal	13.453	936	1852	2709	3522	4296	5005	5700	6493	7286	8079
10	Durango	15.838	914	1757	2609	3412	4145	4832	5527	6314	7101	7888
11	Guerrero	11.013	931	1821	2649	3420	4105	4846	5633	6420	7207	7994
12	Guanajuato	15.021	926	1826	2658	3431	4171	4874	5607	6340	7073	7806
13	Hidalgo	14.305	930	1832	2673	3440	4182	4891	5678	6465	7252	8039
14	Jalisco	14.202	924	1839	2711	3489	4246	4992	5689	6476	7263	8050
15	México	13.209	942	1850	2685	3478	4270	5007	5697	6488	7279	8066

Fig. 5. Tablas de análisis de n -gramas

En particular, para el caso del corpora de noticias, se puede observar que el uso del lenguaje en el centro de la República tiene un impacto importante sobre el resto de los estados. Es posible que esto se deba al hecho de que la OEM tiene

Gráfica - Trigramas de letras Tweets

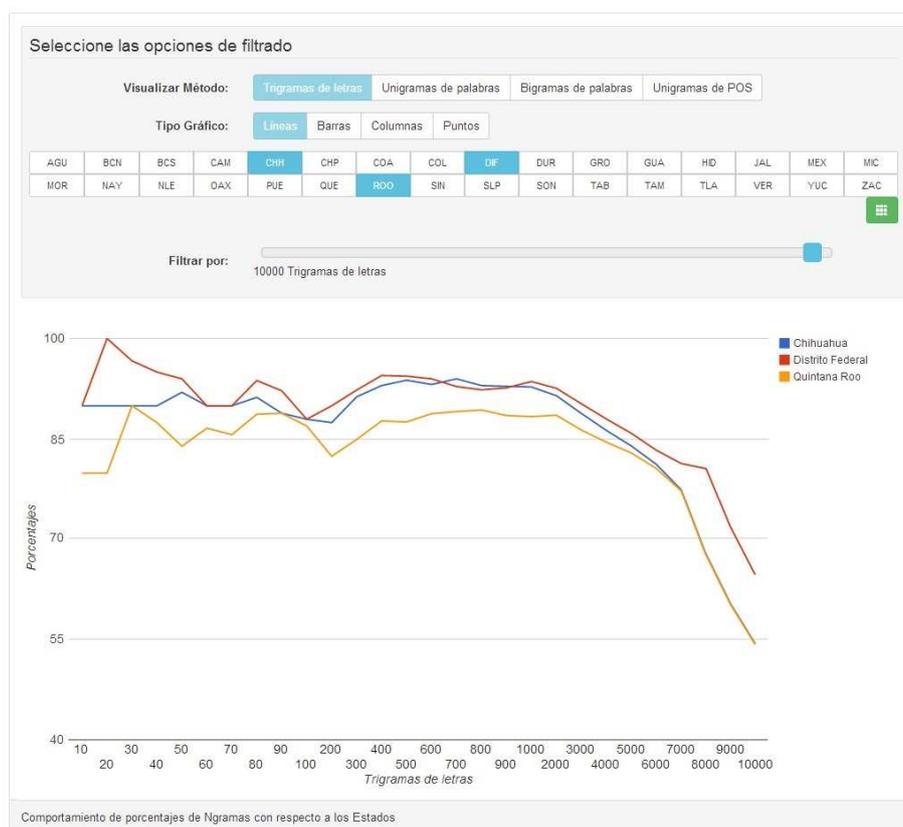


Fig. 6. Gráfica comparativa de n -gramas para tres estados

presencia en todo el país pero existe una gran cantidad de noticias que provienen precisamente del centro de México.

Cuando se toman en cuenta los textos escritos mediante Tweets, la historia es bastante diferente. Se observa una clara dispersión del idioma en estados como Oaxaca y Colima. Algunas zonas de la República como el sureste también muestran una dispersión con respecto al centro de México. Tal y como lo mencionamos anteriormente, consideramos que este mapa refleja mucho mejor el grado de dispersión del español mexicano de cada estado con respecto a la media nacional. Sin embargo, este hecho habrá que analizarlo con mayor detalle.

Con la finalidad de garantizar que la muestra es representativa, en la Tabla 1 se muestra el número de usuarios diferentes por estado considerados en este experimento. Se observa que el total de usuarios es de 196,288, sin embargo, dado que ciertos usuarios fueron localizados en más de un estado, es necesario calcular el número de usuarios diferentes en toda la colección, y este número es

Geo-gráfica - Trigramas de letras Noticias

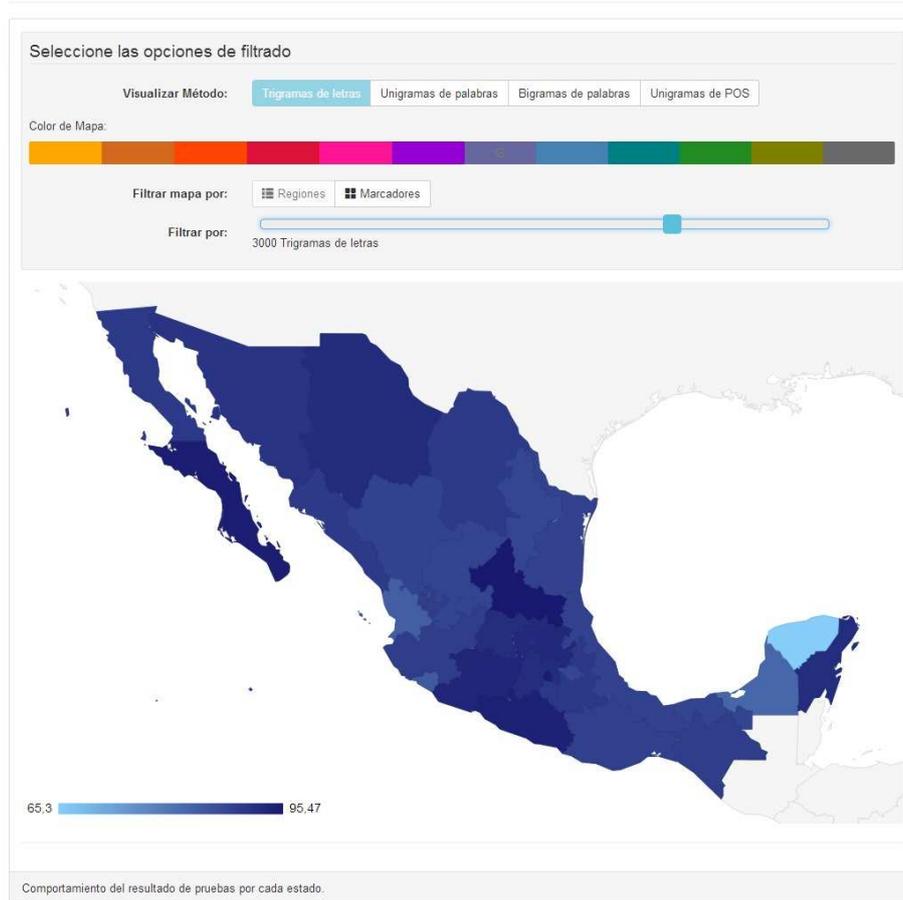


Fig. 7. Visualización de la dispersión del idioma en el mapa (n -gramas de letras) usando el corpus de noticias

177,753. De cualquier manera, consideramos que esta cantidad de usuarios puede ser suficiente para estimar el grado de dispersión del español en la República Mexicana.

5. Conclusiones

Se ha presentado un sistema para la visualización del grado de dispersión del español mexicano en México. El sistema es flexible en el sentido que permite visualizar los datos en formas de tablas, gráficas y mapas. De esta manera, el usuario puede analizar desde diversas perspectivas el grado de cobertura del

Geo-gráfica - Trigramas Tweets

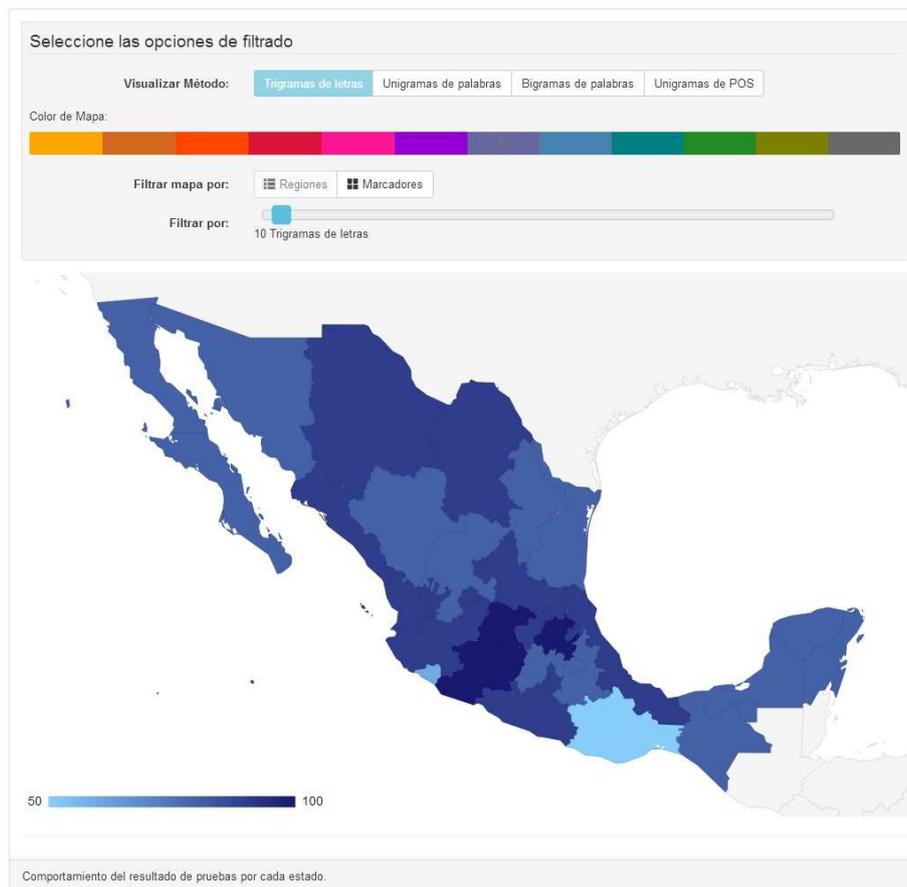


Fig. 8. Visualización de la dispersión del idioma en el mapa (n -gramas de letras) usando el corpus de tweets

idioma en los diferentes estados de la República Mexicana. Para el caso actual, se usan dos corpora, uno del dominio de noticias y el otro extraído desde Twitter. Como trabajo a futuro se implementarán métricas mucho más formales para medir el grado de entropía entre el modelo del lenguaje de un estado y la media nacional. También se pondrá a disposición el sistema con una opción para subir datos a discreción, lo cual permitirá una total flexibilidad en el usuario final.

Referencias

1. Baronchelli, A., Chater, N., Pastor-Satorras, R., Christiansen, M.H.: The biological origin of linguistic diversity. *PloS one* 7(10), e48029 (2012)

Tabla 1. Cantidad de usuarios diferentes por estado considerados en el corpus.

Estado	Usuarios	Estado	Usuarios
Aguascalientes	4,536	Morelos	7,882
Baja California Norte	5,743	Nayarit	2,351
Baja California Sur	2,158	Nuevo León	8,857
Campeche	7,684	Oaxaca	6,891
Chihuahua	6,178	Puebla	8,623
Chiapas	5,074	Querétaro	5,497
Coahuila	5,939	Quintana Roo	10,189
Colima	3,302	Sinaloa	5,491
Distrito Federal	11,367	San Luis Potosí	4,144
Durango	3,569	Sonora	6,710
Guerrero	2,226	Tabasco	5,993
Guanajuato	6,224	Tamaulipas	3,267
Hidalgo	5,557	Tlaxcala	9,111
Jalisco	8,373	Veracruz	4,863
México	11,426	Yucatán	8,668
Michoacán	4,824	Zacatecas	3,571

- Córdoba, H., Gloria, A.: Geografía, lingüística y geolingüística. una propuesta para comprender el contacto dialectal. *Forma y Función* 24, 47–60 (2011)
- Gelbukh, A., Sidorov, G.: Approach to Construction of Automatic Morphological Analysis Systems for Inflective Languages with Little Effort. In: *Computational Linguistics and Intelligent Text Processing*. pp. 157–162. Springer (2003)
- Gelbukh, A., Sidorov, G.: Analizador morfológico disponible: un recurso importante para pln en español. In: *Memorias de talleres del congreso internacional Iberamia-2004*. pp. 209–216 (2004)
- Gelbukh, A., Sidorov, G., Velázquez, F.: Análisis morfológico automático en español a través de generación. *Revista Escritos* 28, 9–26 (2003)
- Patriarca, M., Heinsalu, E.: Influence of geography on language competition. *Physica A: Statistical Mechanics and its Applications* 388(2), 174–186 (2009)
- Perreault, C., Mathew, S.: Dating the origin of language using phonemic diversity. *PloS one* 7(4), e35289 (2012)
- Zampieri, M., Gebre, B.G., Diwersy, S.: N-gram language models and pos distribution for the identification of spanish varieties. In: *Proceedings of TALN2013*. pp. 580–587 (2013)

Detectando la prioridad de contenidos generados en Twitter por medio de n -gramas de palabras

Gerardo García-Rodríguez, Gabriela Ramírez-de-la-Rosa,
Esaú Villatoro-Tello, Héctor Jiménez-Salazar

Departamento de Tecnologías de la Información,
División de Ciencias de la Comunicación y Diseño,
Universidad Autónoma Metropolitana, Unidad Cuajimalpa, DF, México

210368185@alumnos.cua.uam.mx,
{gramirez,evillatoro,hjimenez}@correo.cua.uam.mx

Resumen. En la actualidad, el uso de las redes sociales ha revolucionado la forma en que los usuarios intercambian ideas, opiniones e información. Gracias a este cambio de paradigma en la forma de interactuar entre los usuarios, las grandes compañías y personajes públicos han comenzado a prestar particular interés a la opinión generada respecto a sus productos y/o servicios, actos y/o eventos dentro de las redes sociales. A esta actividad se le conoce como el Análisis de Reputación en Línea (ARL), actividad desempeñada principalmente por usuarios expertos en el análisis de imagen y que al mismo tiempo son capaces de tomar decisiones estratégicas que ayuden a mejorar la reputación de una empresa o personaje público. La relevancia que ha tomado esta actividad en últimos años ha motivado a la comunidad científica a proponer métodos automáticos que apoyen las labores de un ARL. En este trabajo se propone un método automático para determinar cuándo un tuit es importante dentro de una categoría predefinida de mensajes. El método propuesto se basa en el uso de n -gramas para establecer la importancia de los contenidos generados en Twitter. Los experimentos realizados muestran que la ocurrencia de ciertos términos permite a un modelo de clasificación automática determinar de forma efectiva (*medida-F=0.7*) cuando un tuit es o no importante para un ARL.

Palabras clave: Bolsa de n -gramas, clasificación de contenidos, clasificación no temática, procesamiento de lenguaje natural, aprendizaje automático.

1. Introducción

Hoy en día, el amplio uso de las redes sociales ha generado una nueva forma de intercambiar ideas, opiniones e información en tiempo real y desde cualquier parte del planeta, incluso desde fuera de él¹. Estos intercambios de ideas e

¹ El Ingeniero astronauta T.J. Creamer envió, a través de Twitter, un mensaje desde la Estación Espacial Internacional en el 2010: <http://techland.time.com/2010/01/22/a-message-from-outer-space/>

información se llevan a cabo no solo entre usuarios comunes, sino que también involucra a usuarios que representan grandes empresas, artistas, figuras públicas, personajes políticos, por mencionar algunos.

A través del uso de Twitter, por ejemplo, un consumidor puede contactar de forma casi instantánea a la empresa de algún producto para reportar un fallo o para preguntar por un componente que necesita. De igual manera, la compañía puede saber de forma inmediata cómo está respondiendo un determinado grupo de usuarios ante la presencia de alguno de sus productos en el mercado; o si se piensa en algún personaje público, este puede conocer cómo sus seguidores están reaccionando ante una determinada campaña, etc.

Si bien las ventajas que las redes sociales ofrecen a las empresas son grandes, la cantidad de información producida en éstas es gigantesca. Consecuentemente, un analista de reputación enfrenta el gran problema de revisar y analizar toda la información producida en las redes sociales referentes a su producto/servicio o personaje público en cuestión, lo cual resulta en una actividad extremadamente abrumadora y que es realizada con poca eficiencia. Es importante mencionar que tan solo en Twitter se producen cerca de 500 millones de mensajes (tuits) al día [13]. El manejo manual de tal cantidad de información es prácticamente imposible. En este contexto, el proponer sistemas automáticos que permitan realizar un Análisis de Reputación en Línea (ARL) es necesario para procesar y entender las grandes cantidades de contenidos generados en Twitter asociados a una compañía o una entidad en particular [1].

Este análisis de reputación en línea se basa en el constante monitoreo de los mensajes producidos en las redes sociales. El monitoreo de reputación en línea se realiza en etapas. La primera etapa del proceso es filtrar sólo los mensajes que pertenecen a la entidad (compañía o persona); posteriormente, el analista debe decidir la polaridad de dichos mensajes, *i.e.* las implicaciones positivas o negativas que dichos mensajes podrían producir a la compañía o persona. Debido a que una compañía no solo maneja un único producto, así como los personajes públicos se desenvuelven en diferentes ámbitos, el siguiente paso del proceso de ARL consiste entonces en agrupar los mensajes de acuerdo a temáticas relacionadas a un determinado producto o servicios. Finalmente, una vez que se han identificado las temáticas abordadas en los mensajes, el analista deberá determinar cuáles de esos mensajes son los realmente importantes para la compañía o en cuáles debe prestarse mayor atención debido a las implicaciones que podría generar al interior de la compañía, o en el caso de una persona, implicaciones en la imagen pública de dicho personaje. Note que para determinar la importancia de estos mensajes es necesario conocer el giro de la empresa o ámbito en el que el personaje se desenvuelve, pues de ésto dependerá la forma en que el analista emita un juicio adecuado.

Cada uno de los procesos mencionados anteriormente tienen sus propios desafíos, *e.g.*, desambiguación de palabras, análisis de opinión, análisis de sentimientos, agrupamiento, etc. Dada la complejidad que implica realizar todas las tareas involucradas en el ARL, en este trabajo nos enfocaremos únicamente en la detección automática de la importancia de cada mensaje; es decir, una

vez que los tuits están agrupados por temáticas, nos interesa detectar de forma automática cuáles son los tuits de mayor prioridad que un analista deberá tener presente para su análisis. A este último proceso del monitoreo de reputación en línea también se le llama *detección de prioridad* [1].

El método propuesto en este artículo se basa en un enfoque de clasificación de textos tradicional. Nuestra hipótesis principal establece que la importancia de un tuit puede ser determinada a partir de las palabras contenidas en él y que de alguna forma reflejan la forma y uso de ciertos términos que son dependientes del giro o categoría de la compañía y/o personaje que se este analizando. Los resultados obtenidos en 4 diferentes categorías muestran que un enfoque tradicional de clasificación de textos aporta elementos suficientes para determinar cuándo un tuit es importante dentro de una determinada categoría.

El resto del artículo está organizado como sigue, en la Sección 2 se describirán algunos trabajos evaluados dentro de la competencia RepLab [9] que abordan el problema de detección de prioridad. El método propuesto se describirá en detalle en la Sección 3. En la Sección 4 se describe el marco experimental usado para evaluar el método propuesto y finalmente se presentarán conclusiones y trabajo futuro en la Sección 5.

2. Trabajo relacionado

Actualmente, sistemas automáticos para el monitoreo de reputación en línea pueden ser evaluados en el contexto de la competencia internacional RepLab [9]. En dicho espacio se presentan métodos que abordan cada uno de las etapas involucradas en monitoreo de reputación en línea descritas en la sección anterior. Debido a que en este artículo se propone un sistema automático para detectar la prioridad de los tuits, a continuación se analizarán de forma breve los sistemas dentro de la competencia RepLab que se enfocaron en este problema.

La mayoría de los sistemas que atacan el problema de detección de prioridad en tuits utilizan técnicas de análisis de sentimientos, bajo el supuesto de que tuits de connotación negativa hacia la compañía deberán ser más importantes que tuits de connotaciones positivas [3,7,12]. Por ejemplo, en [12] hacen uso de un diccionario de términos afectivos llamado *SentiSense*, el método que proponen genera un grafo para cada dominio con el objetivo de expandir este diccionario de palabras de afecto.

En [3] asignan la prioridad de cada tuit basándose en la polaridad y la intensidad de sentimiento en ellos, debido a que no contaban con un diccionario de sentimiento especial para redes sociales, en particular Twitter, utilizaron un diccionario inicialmente creado para noticias. Sin embargo, el uso de sentimiento (de noticias) no ayudó a la identificación de la prioridad de los tuits.

A diferencia de los trabajos anteriores, el método propuesto en este trabajo no busca utilizar técnicas de análisis de sentimiento para encontrar la prioridad de los tuits, ni hace uso de recursos externos para apoyar la clasificación. Por el contrario, nuestro método entrena un modelo para asignar prioridades por categoría utilizando un enfoque tradicional de clasificación de textos, como se

explicará en la Sección 3. Los experimentos realizados en cuatro diferentes categorías muestran que un enfoque de clasificación de textos tradicional aporta suficientes elementos para seleccionar tuits importantes.

3. Método propuesto

Como se ha venido mencionando en secciones anteriores, en este trabajo se aborda la problemática de detección de tuits importantes aplicando el paradigma de Clasificación de Textos². Bajo este paradigma un primer paso necesario es el *indexado* (*i.e.*, representación) de los documentos de entrenamiento (Tr), actividad que denota hacer el mapeo de un documento d_j en una forma compacta de su contenido. La representación más comúnmente utilizada para representar cada documento es un vector con términos³ ponderados como entradas, concepto tomado del modelo de espacio vectorial usado en recuperación de información [2]. Es decir, un texto d_j es representado como el vector $\vec{d}_j = \langle w_{kj}, \dots, w_{|\tau|j} \rangle$, donde τ es el *diccionario*, *i.e.*, el conjunto de términos que ocurren al menos una vez en algún documento de Tr , mientras que w_{kj} representa la importancia del término t_k dentro del contenido del documento d_j .

En ocasiones τ es el resultado de filtrar los términos del vocabulario, este filtrado normalmente es el resultado de un preprocesamiento (ver Sección 3.2 para más detalles). Una vez que hemos hecho los filtrados necesarios, el diccionario τ puede definirse de acuerdo a diferentes criterios, sin embargo el que se empleó en esta propuesta corresponde a la Bolsa de Palabras (BOW⁴).

La BOW es la forma tradicionalmente utilizada para representar los documentos [11]. Este método de representación utiliza a las palabras simples como los elementos del vector de términos. Dentro de este trabajo se evaluó la influencia tanto de palabras simples (unigramas de palabras) así como de secuencias de palabras de longitud dos (bigramas de palabras).

Respecto al peso w_{kj} , existen diferentes formas de calcularlo, entre las más usadas en la comunidad científica se pueden mencionar: el ponderado booleano, ponderado por frecuencia y ponderado por frecuencia relativa de términos. A continuación se describen de manera breve cada uno de estos esquemas de pesado.

3.1. Esquemas de pesado utilizados

El peso w_{kj} puede ser calculado utilizando diferentes enfoques, el más simple de estos enfoques es el *booleano* que consiste en asignar un valor de 1 al término si éste aparece en el documento, y 0 en caso contrario. Este esquema de pesado

² La Clasificación de Textos es la tarea de asociar automáticamente categorías predefinidas con documentos a partir del análisis de su contenido [11].

³ Entiéndase por términos ya sea palabras, n -gramas de palabras y/o caracteres, secuencias frecuentes, etc.

⁴ Bag Of Words por sus siglas en Inglés.

captura de forma intuitiva la presencia y/o ausencia de los términos dentro de un documento.

Otro esquema de pesado es conocido como *frecuencia relativa*. En este enfoque se determina el peso de un término t_k en el documento d_j en proporción directa al número de veces que el término aparece en el documento, e inversamente proporcional al número de documentos en los que aparece el término t_k el conjunto total de entrenamiento. En particular, el peso está dado por:

$$w_{kj} = TF(t_k) \times IDF(t_k) \quad (1)$$

donde $TF(t_k) = f_{kj}$, es decir, la frecuencia del término t_k en el documento d_j . Y la *frecuencia inversa*, o $IDF(t_k)$ es una manera de medir la *rareza* del término t_k . Para calcular el valor de IDF se utiliza la siguiente fórmula [10]:

$$IDF(t_k) = \log \frac{|Tr|}{\{d_j \in D : t_k \in d_j\}} \quad (2)$$

A esta frecuencia relativa le denominaremos *Tf-Idf* a lo largo de este documento.

Ambos esquemas han demostrado tener buenos resultados en métodos de clasificación de textos. El primer esquema es simple y sólo considera la aparición o no de términos en el documento, mientras que el segundo método es más complejo de calcular pero ayuda a discriminar términos que no apoyan la clasificación de una u otra clase. *Tf-Idf* prefiere términos que sean frecuentes en el documento pero que aparezcan poco en documentos del resto de las clases. En los experimentos evaluaremos estos dos esquemas por separado, nuestra intuición es que debido a lo corto de los mensajes de los tuits, el primer enfoque será suficiente.

3.2. Preprocesamiento

Un tuit es un texto de no más de 140 caracteres que puede contener componentes inherentes a esa red social como: palabras fuera de algún diccionario, texto informal, *hashtags*, menciones a otros usuarios, URLs, abreviaciones de palabras, repeticiones de letras, repeticiones de signos de puntuación, etc. La mayoría de estos componentes no son útiles en un enfoque de clasificación de textos tradicional [11,8] ni en la definición de la prioridad del mensaje [3,7,12]. Por lo anterior, cada uno de los mensajes han sido preprocesados, aplicando los siguientes pasos:

- Los tuits se convierten a minúsculas con la finalidad de normalizar el vocabulario.
- Cualquier secuencia de espacios en blanco se convierte en un solo espacio.
- Cualquier mención a usuarios encontrada es eliminada.
- Cualquier enlace (URL) encontrado es eliminado.
- Se eliminan los signos de puntuación. En consecuencia cualquier emoticono queda de igual forma eliminado.
- Se eliminan las palabras vacías y/o funcionales.

4. Evaluación experimental

A continuación se describen las medidas usadas en la evaluación del sistema, el conjunto de datos usados y los resultados obtenidos en los experimentos realizados.

4.1. Medidas de Evaluación

Para evaluar el método propuesto se utilizaron las medidas tradicionales para evaluación de sistemas de clasificación de textos, tales como *precisión*, *recuerdo* y *medida F*. La precisión (P) es la proporción de textos clasificados correctamente en una clase c_i con respecto a la cantidad de textos clasificados en esa misma clase. El recuerdo (R), la proporción de textos clasificados correctamente en una clase c_i con respecto a la cantidad de textos que realmente pertenecen a esa clase. Así, la precisión se puede ver como una medida de la corrección del sistema, mientras que el recuerdo da una medida de cobertura o completitud.

Normalmente se emplea la medida F para describir el comportamiento de la clasificación, la cual se define como:

$$F = \frac{(1 + \beta^2)P * R}{\beta^2(P + R)} \quad (3)$$

donde β representa la media armónica entre la precisión y el recuerdo. La función de β es la de controlar la importancia relativa entre las medidas de precisión y recuerdo. Es común asignar un valor de 1 indicando igual importancia a ambas medidas.

4.2. Conjunto de datos

En los experimentos presentados se usó la colección de tuits proporcionada por el RepLab 2013 [1,9]. La colección contiene tuits de cuatro grandes categorías, cada categoría busca cubrir un tipo diferente de evaluación de reputación. Las cuatro categorías son: *automotriz* que incluye entidades para las cuales la reputación se determina en base a sus productos; *banca* con entidades para las cuales la transparencia y la ética de sus actividades son factores a considerar en cuanto a la evaluación de su reputación; *universidades* donde su reputación depende de la variedad de productos intangibles; y *musical* en la que sus entidades basan su reputación tanto en la calidad de sus productos como en las cualidades personales de forma igualitaria.

Agregado a lo anterior, los tuits proporcionados por el RepLab 2013 están en dos idiomas distintos: Inglés y Español. Para la realización de nuestros experimentos se trabajó únicamente con el subconjunto de tuits en español (vea la Tabla 1).

En la Tabla 1 se muestra información sobre el subconjunto de datos usado en nuestros experimentos. Cabe mencionar que la colección de datos original menciona que los tuits están clasificados en tres categorías, a saber, *importante* (I),

medianamente importante (MI) y no importante (UN); sin embargo, es necesario mencionar que en los datos de entrenamiento liberados por los organizadores del RepLab 2013, no existe ningún ejemplo de tuit *importante* en ninguna de las cuatro categorías. Dada esta situación, nuestro problema se redujo a un problema de clasificación binaria, es decir, identificación de tuits *medianamente importantes y no importantes*.

Tabla 1. Número de tuits MI y UN existentes en cada una de las diferentes categorías en el subconjunto de datos considerado

	Automotriz	Banca	Musical	Universidades	Total
<i>MI</i>	952	540	1497	141	3130
<i>UN</i>	1256	1011	1462	79	3808
<i>Total</i>	2208	1551	2959	220	6938

Note que de forma global el conjunto de tuits MI y UN está balanceado. Sin embargo, al ver el número de tuits por clase en cada categoría, especialmente para las categorías Automotriz, Banca y Universidades, no dice que estas clases se encuentran muy desbalanceados, situación que representa ya un reto por sí misma.

4.3. Clasificadores

Dado que nuestra propuesta para identificar la importancia de los contenidos generados en Twitter no depende de ningún algoritmo de aprendizaje en particular, podemos emplear prácticamente cualquier clasificador para enfrentar el problema. Para los experimentos realizados seleccionamos dos diferentes algoritmos de aprendizaje, los cuales son algoritmos representativos dentro de la gran variedad de algoritmos de aprendizaje disponibles actualmente en el campo de aprendizaje computacional [4,6]. Específicamente, consideramos los siguientes:

- **Naïve Bayes(NB).** Método probabilístico que asume la independencia de los atributos entre las diferentes clases del conjunto de entrenamiento.
- **Máquinas de Vectores de Soporte (Support vector machine, SVM).** Método que hace uso de discriminantes lineales que intentan localizar de forma óptima el hiperplano que separa los elementos de las diferentes clases. Para nuestro trabajo se utilizó un kernel lineal.

En nuestros experimentos se empleó la implementación de Weka [5] de cada uno de estos algoritmos con los parámetros por defecto. Es importante mencionar que para todos los experimentos se empleó la estrategia de validación cruzada a diez pliegues.

4.4. Configuración experimental y resultados

Como se ha mencionado en secciones anteriores, el objetivo principal de este trabajo fue determinar hasta qué punto es posible construir un clasificador, mediante el enfoque tradicional de clasificación de textos, capaz de determinar cuándo un tuit es importante o no dentro de un dominio de interés particular para un analista de reputación en línea, con la utilización exclusiva de las palabras contenidas en el mismo tuit.

Para lograr dicho objetivo se propusieron dos conjuntos de experimentos, mismos que se describen a continuación:

Experimento 1 . Evaluar el impacto que tiene emplear palabras simples como forma de representación, *i.e.*, unigramas de palabras, en la clasificación de tuits importantes. La hipótesis principal de este experimento sugiere que existen algunos términos característicos del dominio de interés (Automotriz, Banca, Musical y Universidades) que permiten identificar de manera eficiente cuándo un tuit es importante.

Experimento 2 . Evaluar el impacto que tienen las secuencias de palabras de longitud dos, *i.e.*, bigramas de palabras, en la clasificación de tuits importantes. La hipótesis principal de este experimento sugiere que para determinados dominios existen secuencias de palabras, frecuentemente utilizadas por los usuarios, las cuáles permitirían a un ARL identificar los tuits importantes.

Finalmente, para ambos experimentos se emplearon dos diferentes esquemas de pesado: *booleano* y *Tf-Idf*. El objetivo de emplear ambos esquemas de pesado fue valorar el impacto que tiene la presencia de ciertos términos contra la frecuencia de aparición de los mismos.

Los resultados de los experimentos planteados pueden observarse en la Tabla 2 y en la Tabla 3. Observe que en ambas tablas se evaluó el aporte de la forma del esquema de pesado *boolean* y *Tf-Idf* tanto con un clasificador bayesiano (NB) como con las máquinas de vectores de soporte (SVM). Los resultados de los experimentos aparecen en términos de su *Precisión (P)*, *Recuerdo (R)* y de la medida *F* (vea Sección 4.1). Cada línea representa los resultados obtenidos en cada una de las categorías existentes dentro de la colección de tuits proporcionados por el RepLab 2013. Agregado a estos datos, cada tabla contiene en la última fila el promedio de la precisión, recuerdo y medida F obtenidos en las cuatro categorías de tuits empleados durante la experimentación.

Observe que en ambos experimentos, *i.e.*, empleando tanto unigramas (Tabla 2) como bigramas (Tabla 3) de palabras, es claro que la frecuencia de los términos no aporta elementos suficientes al clasificador para determinar cuándo un tuit es o no importante dentro de una determinada categoría. Los experimentos realizados muestran, hasta cierto punto, que la sola presencia de los términos (*i.e.*, ponderado booleano) permite a los algoritmos de clasificación automática contar con mejores elementos para discriminar tuits importantes de los que no lo son.

Note que los resultados obtenidos empleando el clasificador Naïve Bayes (NB) son muy cercanos a los resultados obtenidos al emplear Máquinas de

Tabla 2. Resultados de la clasificación empleando unigramas de palabras

unigramas												
Categoría de Tuits	Booleano						TF-IDF					
	NB			SVM			NB			SMV		
	P	R	F	P	R	F	P	R	F	P	R	F
<i>Automotriz</i>	0.61	0.62	0.61	0.53	0.54	0.54	0.63	0.58	0.58	0.54	0.54	0.54
<i>Banca</i>	0.76	0.75	0.75	0.81	0.81	0.81	0.29	0.72	0.69	0.79	0.80	0.79
<i>Musical</i>	0.71	0.70	0.70	0.70	0.70	0.70	0.65	0.64	0.64	0.67	0.66	0.66
<i>Universidades</i>	0.75	0.75	0.75	0.75	0.76	0.75	0.75	0.72	0.72	0.70	0.71	0.70
<i>Promedio</i>	<i>0.70</i>	<i>0.70</i>	<i>0.70</i>	<i>0.70</i>	<i>0.70</i>	<i>0.70</i>	<i>0.58</i>	<i>0.67</i>	<i>0.66</i>	<i>0.67</i>	<i>0.68</i>	<i>0.67</i>

Tabla 3. Resultados de la clasificación empleando bigramas de palabras

bigramas												
Categoría de Tuits	Booleano						TF-IDF					
	NB			SVM			NB			SMV		
	P	R	F	P	R	F	P	R	F	P	R	F
<i>Automotriz</i>	0.52	0.55	0.52	0.50	0.54	0.55	0.63	0.53	0.49	0.52	0.54	0.52
<i>Banca</i>	0.74	0.75	0.75	0.80	0.80	0.79	0.74	0.71	0.71	0.79	0.79	0.77
<i>Musical</i>	0.68	0.68	0.68	0.70	0.70	0.69	0.64	0.63	0.63	0.69	0.69	0.68
<i>Universidades</i>	0.72	0.72	0.72	0.72	0.72	0.69	0.76	0.65	0.65	0.69	0.69	0.64
<i>Promedio</i>	<i>0.67</i>	<i>0.68</i>	<i>0.68</i>	<i>0.68</i>	<i>0.69</i>	<i>0.68</i>	<i>0.69</i>	<i>0.63</i>	<i>0.62</i>	<i>0.68</i>	<i>0.66</i>	<i>0.65</i>

Vectores de Soporte (SVM), lo cual es un indicador de que los resultados del método propuesto no dependen del algoritmo de clasificación. Esta situación es en realidad un indicativo de que la forma de representación propuesta es apropiada para el problema planteado, es decir, el uso de palabras (unigramas y bigramas) aportan elementos suficientes para que un algoritmo de clasificación automática sea capaz de distinguir tuits importantes de aquellos de no lo son.

En la Figura 1 se muestra de forma gráfica y resumida el comportamiento de los diferentes experimentos realizados para cada una de las cuatro categorías del conjunto de datos empleado. Los resultados se muestran en términos de la medida F .

De la Figura 1 se pueden extraer varias conclusiones preliminares. En primer lugar parece ser claro que el uso de bigramas de palabras tienen un desempeño menor que con los unigramas de palabras. Este comportamiento queda reflejado también en las filas de *Promedio* en las Tablas 2 y 3 respectivamente, donde se puede notar, bajo una configuración con esquema de pesado *booleano*, un decremento en la medida F , de un valor de 0.70 usando unigramas a un valor de 0.68 usando bigramas. Sin embargo, a pesar de que el desempeño es relativamente menor, no es posible asegurar que los bigramas de palabras no aportan información importante debido a que las diferencias en el desempeño no parecieran ser significativas. Más experimentos necesitan ser realizados en este punto para descartar la hipótesis planteada en nuestro experimento 2.

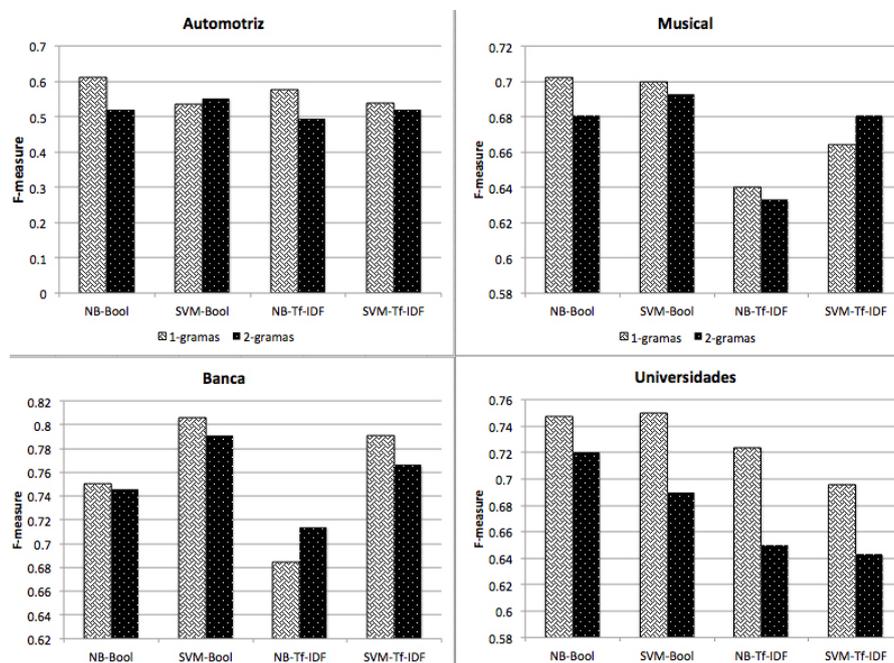


Fig. 1. Comparación de los resultados obtenidos en términos de la medida F

Agregado a esto, es importante resaltar que los mejores resultados de la clasificación se obtienen en la categoría *Banca*, lo cual indica hasta cierto punto, que los usuarios tienden a usar el mismo vocabulario cuando expresan opiniones sobre instituciones bancarias, permitiendo esto que un clasificador sea capaz de aprender cuando la ocurrencia de determinados términos está asociada a la importancia de los tuits. Por el contrario, el dominio *Automotriz* fue en el que los resultados de clasificación resultaron ser más bajos.

Los resultados obtenidos en el dominio *Automotriz* nos hace creer que el tipo de lenguaje empleado entre los usuarios es más variado en este dominio en específico al momento de referir sus opiniones, como puede verse en la Tabla 4 donde se presentan los bigramas con mayor ganancia de información para dos dominios, en particular, para el dominio *Musical* y *Automotriz*. Es claro que para el caso del dominio *Automotriz* las opiniones no mencionan directamente la entidad de la que se opina, algunos de sus términos bien pueden representar al dominio *Banca* o *Universidad*, lo que puede ser la causa del desempeño pobre en este dominio.

En general, de los experimentos podemos afirmar que la presencia de ciertos términos, en particular palabras, aporta elementos suficientes que permiten distinguir tuits importantes dentro de una determinada de categoría de tuits.

Tabla 4. Ejemplos de bigramas con mayor ganancia de información para los dominios *Musical* y *Automotriz*

<i>Musical</i>	jennifer lopez, ac dc, alicia key, a shakira, de jennifer, via youtube, shakira y, bon jovi, que shakira, de maroon, por cantar, millones de
<i>Automotriz</i>	anuncio gratis, download free, pdf download, manual pdf, gratis yamaha, de robar, favor rt, acaban de, de honda, con yamaha, de wolkswagen, service manuav

5. Conclusiones y trabajo futuro

Se presentó un método de clasificación de textos para la detección de prioridad de tuits. Se encontró que efectivamente existen palabras que son comúnmente utilizadas por los usuarios para referir una opinión, mismas que permiten a un sistema de clasificación automática determinar cuándo un tuit es o no importante para los propósitos de un Analista de Reputación en Línea.

Se realizaron varios experimentos los cuales fueron pensados de tal forma que nos permitieran validar varias hipótesis, las cuales se listan a continuación: *i*) se mostró que la presencia de los términos es suficiente para que un clasificador aprenda a distinguir tuits importantes de aquellos que no lo son; *ii*) se comprobó que la forma de representación propuesta es la que permite resolver el problema con cierta efectividad, y que los resultados no dependen del algoritmo de aprendizaje; *iii*) finalmente, se pudo observar que el uso de bigramas de palabras no son lo suficientemente descriptivos como para superar el desempeño obtenido por una representación basada en unigramas. Se concluyó que, debido a las diferencias poco significativas en el desempeño, más experimentos hacen falta para aceptar o descartar el aporte de los bigramas en la tarea propuesta.

Como trabajo futuro se planea realizar un análisis más detallado de la relevancia de los atributos propuestos en este trabajo, es decir, evaluar el aporte tanto de los unigramas como de los bigramas en el problema atacado. Para esto se pretende realizar más experimentos aplicando estrategias de selección de atributos, los cuales nos permitirán identificar al verdadero valor discriminativo de los atributos. Agregado a esto, se evaluará el método propuesto en el subconjunto de tuits en idioma Inglés del RepLab.

Los resultados obtenidos muestran que el uso de una representación basada en las palabras contenidas en los tuits puede ser exitosa, pues obtuvo resultados competitivos a los reportados en la competencia de RepLab 2013, al mismo tiempo queda claro que aún existe un amplio margen de trabajo que resta por hacer. Es importante mencionar que de entre los equipos participantes en la tarea del RepLab 2013, el desempeño del sistema mejor evaluado fue de un $F=0.63$ en promedio, mientras que el método base obtuvo un $F=0.60$ [1].

Agradecimientos. Agradecemos el apoyo otorgado por la Universidad Autónoma Metropolitana Unidad Cuajimalpa y el SNI-CONACyT.

Referencias

1. Amigó, E., de Albornoz, J.C., Chugur, I., Corujo, A., Gonzalo, J., Martín-Wanton, T., Meij, E., de Rijke, M., Spina, D.: Overview of replab 2013: Evaluating online reputation monitoring systems. In: Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B. (eds.) CLEF. Lecture Notes in Computer Science, vol. 8138, pp. 333–352. Springer (2013)
2. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison Wesley (1999)
3. Balahur, A., Tanev, H.: Detecting entity-related events and sentiments from tweets using multilingual resources. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF (Online Working Notes/Labs/Workshop) (2012)
4. Duda, R., Hart, P.: Pattern classification and scene analysis. Wiley (1996)
5. Garner, S.R.: Weka: The Waikato environment for knowledge analysis. In: Proc. of the New Zealand Computer Science Research Students Conference. pp. 57–64 (1995)
6. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer (2009)
7. Jean-Valère Cossu, Benjamin Bigot, L.B.M.M.X.B.G.S.R.D.V.B.J.M.T.M., El-Bèze, M.: Lia@replab 2013. In: CLEF 2013 Eval. Labs and Workshop Online Working Notes (2013)
8. Leon-Martagón, G., Villatoro-Tello, E., Jiménez-Salazar, H., Sánchez-Sánchez, C.: Análisis de polaridad en twitter. Journal of Research in Computing Science. 62, 69–78 (2013)
9. LiMoSINe: Replab 2013 – an evaluation campaign for online reputation management systems (2014), Última visita 15 de marzo de 2014
10. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manage. 24(5), 513–523 (Aug 1988)
11. Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1–47 (2002)
12. Spina, D., Carrillo de Albornoz, J., Martín, T., Amigó, E., Gonzalo, J., Giner, F.: UNED Online Reputation Monitoring Team at RepLab 2013. In: CLEF 2013 Eval. Labs and Workshop Online Working Notes. CLEF (2013)
13. Twitter, I.: About twitter (2014), Última visita 15 de marzo de 2014

Diseño de agentes para recuperar información para el enriquecimiento de ontologías dirigidas a epidemiología: el caso de la tuberculosis

Joanna Alvarado-Uribe, Ari Yair Barrera-Animas, Miguel González-Mendoza,
Ma. de los Ángeles Junco-Rey

Instituto Tecnológico y de Estudios Superiores de Monterrey,
Edo. de México, México

joanna.1890@gmail.com, gnpliquid@gmail.com, mgonza@itesm.mx,
ajunco@itesm.mx

Resumen. En este trabajo se presenta un Sistema Multiagente que busca información relevante en Sitios Web, que puede ser filtrada de tal manera que puede contribuir en el enriquecimiento de una ontología dirigida a la Tuberculosis en México. Esta ontología está basada, por un lado, en la estructura de la Base de Datos Relacional (BDR) del Sistema Nacional Mexicano de Vigilancia Epidemiológica (SINAVE), información estructurada; y, por otro lado, de un Sistema Multiagente (SMA) que recolecta información de la web en diferentes formas (información semiestructurada). Particularmente, los agentes recopilan información de distintas fuentes dinámicas: Facebook, Twitter y Blogs (páginas de Internet), generales y especializadas, que permiten la actualización de esta ontología. El análisis y diseño de los agentes se realiza utilizando la metodología GAIA y el protocolo de comunicación se lleva a cabo por medio del estándar FIPA-ACL; desarrollado en el entorno de programación JADE (Java) a través de NetBeans. Se realiza la conversión de la información relacional a semántica utilizando la herramienta D2R Server, para generar Ontologías por medio del método de mapeo directo de BDR.

Palabras clave: Sistema multiagente, ontología, GAIA, FIPA, JADE.

1. Introducción

Actualmente, cualquier búsqueda en la Web nos arroja todos los posibles resultados que estén relacionados con la petición que el usuario realizó, aunque no se asegura que todos los resultados sean únicamente los que se quieren conocer. Así, el usuario tiene que continuar su búsqueda manualmente en cada uno de los resultados, lo que consume más tiempo.

La incorporación del enfoque semántico en la Web nos asegura que todos los resultados de la búsqueda sean exactos [1], debido a que la información se encuentra mejor definida; esto es posible gracias a que la representación del conocimiento se realiza por medio de ontologías. Las ontologías permiten definir un vocabulario

común que describe en un dominio específico, los conceptos que involucra y las relaciones entre ellos; por tal motivo, las ontologías se convierten en una pieza fundamental de las tecnologías orientadas a la Web Semántica [12].

Con base en el párrafo anterior, se plantea una forma para enriquecer una ontología de acuerdo con el tema que se requiera incorporar a cierto Sistema Computacional o en el caso más general, a la Web Semántica. En este trabajo se propone el enriquecimiento de una ontología acerca del tema de Tuberculosis en México. Esta ontología se compone de dos partes: la primera se basa en la estructura de una Base de Datos Relacional proporcionada por el Sistema Nacional Mexicano de Vigilancia Epidemiológica (SINAVE) dependiente del Sector Salud [3]; y la segunda, de información obtenida de páginas Web. Por lo que, se busca recopilar información de distintas fuentes estáticas y/o dinámicas, generales y especializadas; que permitan la actualización de esta ontología.

La búsqueda de información útil para el enriquecimiento de la ontología se realiza por medio de un Sistema Multiagente; dicho sistema se compone de 3 agentes: RSATFB (Recolector del Sistema Multiagente en Facebook), RSATTW (Recolector del Sistema Mutiagente en Twitter) y RSATBLOG (Recolector del Sistema Multiagente en los BLOGs). Estos agentes recopilan la información que los usuarios proporcionan en estos sitios y la analizan para determinar qué información podría ser relevante para la actualización de la ontología de Tuberculosis. Los términos que sean extraídos de estos sitios Web deben tener una alineación con al menos uno de los términos almacenados en la BDR de Tuberculosis. Una vez que analizan la información, la clasifican de acuerdo con el área a la que pertenezca y la almacenan en una Base de Datos Relacional que es mapeada manualmente para obtener la ontología actualizada.

El proceso de convertir la información relacional a semántica en este trabajo se realiza con la herramienta D2R Server, con la que se generan dos tipos de archivos: uno en formato RDF, que es el utilizado convencionalmente por la Web Semántica y; otro con extensión .ttl, que permite la explotación y el enriquecimiento de los datos semánticos utilizando SPARQL con un razonador denominado Snorql para hacer inferencia sobre los datos [3].

El uso de un Sistema Multiagente permite que la ontología vaya creciendo con información obtenida de diversas fuentes y con un grado mayor de confianza al considerar la definición de un agente “Los agentes son componentes de software activos y persistentes que perciben, razonan, actúan y se comunican” Huns & Singh (1998) [2]. Estas características permiten a los agentes convertir una actividad humana a semiautomática.

A continuación, se describen los 5 apartados que conforman este artículo: Análisis y diseño, donde se realiza la descripción de la organización del sistema, los roles de los agentes, protocolos y relaciones; Prototipo, se describe el prototipo y la forma de usarlo; Experimentos y resultados, se describen las partes cubiertas hasta el momento de la propuesta; y Conclusiones.

2. Análisis y diseño

El sistema se compone de 3 agentes: RSATFB (Recolector del Sistema Multiagente en Facebook), RSATTW (Recolector del Sistema Multiagente en Twitter) y RSATBLOG (Recolector del Sistema Multiagente en los BLOGs) [7].

Estos agentes recopilan la información que los usuarios proporcionan en estos sitios, la analizan para determinar qué información es irrelevante o útil para la actualización de la ontología de Tuberculosis a partir de los términos que sean extraídos de estos sitios Web. Una vez que analizan la información, la clasifican de acuerdo con el área a la que pertenece, almacenando la información recolectada en una Base de Datos Relacional denominada SMATERMINOS (BDR SMATERMINOS). Posteriormente, los términos presentes en la BDR SMATERMINOS son comparados con los términos contenidos en la BDR de Tuberculosis (BDR), con el propósito de preservar exclusivamente términos nuevos. Finalizando con el mapeo manual de la BDR de Tuberculosis denominada “tuberculosis_uno”, en el caso de haber clasificado términos nuevos, con la finalidad de obtener la ontología actualizada en formato RDF y .ttl para su explotación. Como se puede ver en la Fig. 1.

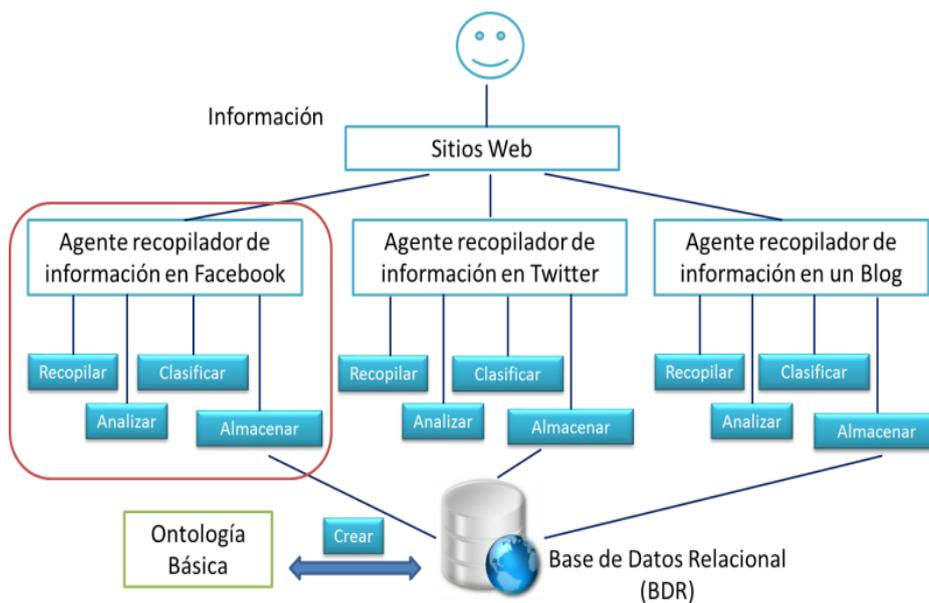


Fig. 1. Organización del sistema.

2.1. Modelo del agente

Se tienen tres modelos:

- Modelo del agente RSATFB (Recolector del Sistema Multiagente en Facebook), ver Fig. 2.

- Modelo del agente RSATTW (Recolector del Sistema Multiagente en Twitter).
- Modelo del agente RSATBLOG (Recolector del Sistema Multiagente en Blogs).

Cabe mencionar que todos los agentes presentan el mismo modelo (actividades que ejecuta el agente).

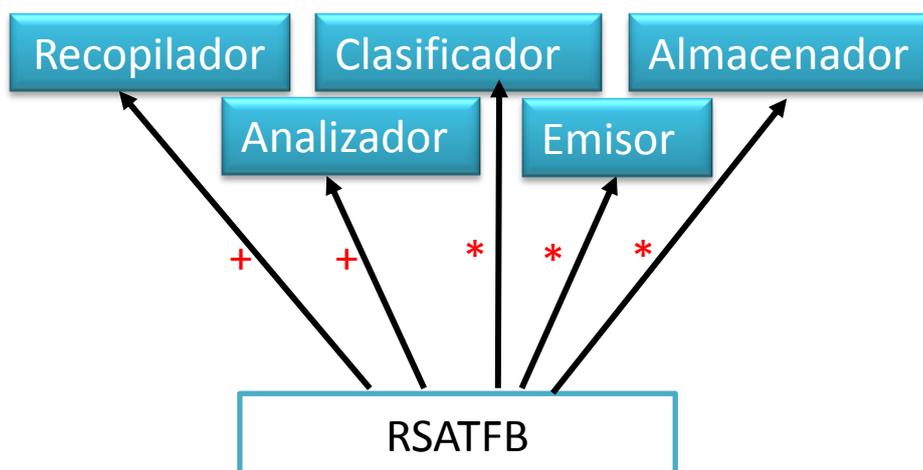


Fig. 2. Modelo del agente RSATFB.

2.2. Modelo de roles (a nivel de análisis)

Los agentes presentan, de manera específica y congruente a su función, un esquema de rol definido; en el que se establece el comportamiento del mismo a través del sistema. El esquema que se muestra en la Tabla 1 denota el rol para el agente RSATFB (Facebook), que funge como ejemplo demostrativo de los roles correspondientes a cada agente previsto.

Tabla 1. Esquema del rol del agente propuesto para Facebook.

Esquema del Rol: RSATFB	
1 Descripción:	Este rol involucra recopilar información de la red social Facebook en las publicaciones de los usuarios para analizarla, identificar términos nuevos de acuerdo con la BDR SMAterminos, y la tabla correspondiente a su Sitio Web y almacenar dichos términos en la misma BDR, en el caso de que otro agente no lo haya identificado.
Protocolos y Actividades:	Actividades:

<ul style="list-style-type: none"> • Recopilar: consiste en extraer toda la información que los usuarios de Facebook publiquen en esta red social. • Analizar: consiste en leer la BDR para determinar los términos que no se encuentran clasificados en la BDR SMAterminos. (Posteriormente, esta tabla será evaluada por un especialista en el área para introducir dichos términos a la Ontología de Tuberculosis). • Clasificar: identifica e indica los términos nuevos y los que ya se han clasificado en su momento. • Almacenar: consiste en guardar los nuevos términos en la BDR SMAterminos que se identifiquen.
<p>Protocolo:</p> <ul style="list-style-type: none"> • Emitir: consiste en transmitir a los agentes que se identificó un nuevo término, los cuales revisan si en ese momento ya lo han identificado y si esto es verdadero, le responden con la performativa INFORM “Término clasificado”; de lo contrario, su respuesta es un “Término NO clasificado”.
<p>Permiso:</p> <ul style="list-style-type: none"> • Modificar la BDR SMAterminos.
<p>Responsabilidades:</p> <ul style="list-style-type: none"> • <u>De vida:</u> RSATFB = (Recopilar.Analizar)^w.(Clasificar.Emitir)[*] Almacenar[*] • <u>De seguridad:</u> <ul style="list-style-type: none"> ➢ Nueva Publicación = "Sí". ➢ Prioridad máxima al protocolo "Emitir".

2.3. Modelo de interacción (a nivel de análisis)

Se tiene únicamente un protocolo "Emitir", que tiene incluidas las actividades Analizar, Clasificar y Almacenar, siendo el mismo protocolo para todos los agentes contenidos en el sistema, presentando cada uno las modificaciones correspondientes en la implementación del protocolo "Emitir" con la finalidad de establecer la comunicación entre los otros agentes y recibir sus peticiones. En la Tabla 2 se muestra el esquema del protocolo de interacción "Emitir", que define la descripción de la interacción entre agentes cuyo propósito principal es la emisión (propagación y recepción) de todo nuevo término capturado por cada uno de los agentes.

Tabla 2. Esquema del protocolo Emitir.

Propósito: Emitir término	
Iniciador: RSAT	Respondedor: RSAT diferente a Iniciador
Descripción:	

Transmitir a los agentes que se identificó un nuevo término, los cuales revisan si en ese momento ya lo han identificado y si esto es verdadero, le responden con la performativa INFORM “Término clasificado”; de lo contrario, su respuesta es un “Término NO clasificado” y lo introducen a la tabla termino_clasificado de acuerdo con el Sitio Web que exploran.

Entrada: Término Nuevo = “Termino NO clasificado”.

Salida: Clasificación = “Término clasificado” || “Término NO clasificado”.

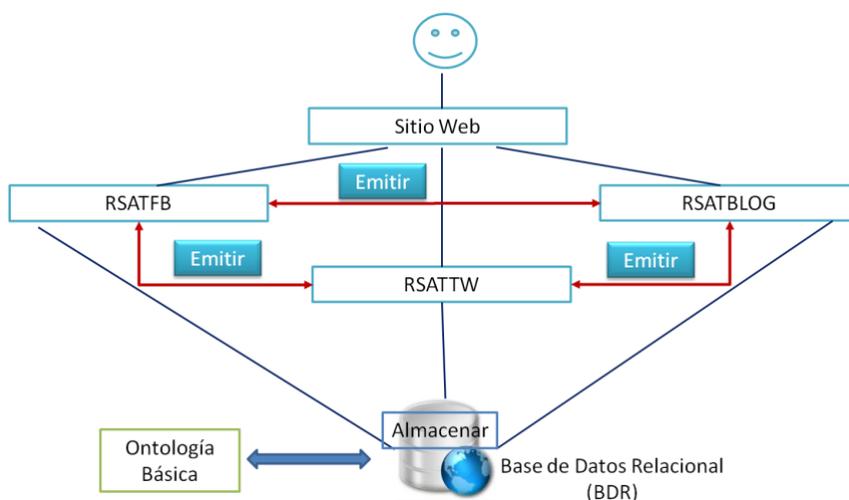


Fig. 3. Modelo de conocimiento.

2.4. Conocimiento de los agentes

Como conocimiento previo se tienen dos Bases de Datos Relacionales (BDR):

- 1) La primera BDR se denomina "tuberculosis_uno", esta Base de Datos contiene las tablas y los registros que se utilizan para construir la primera versión de la ontología. Se usa la ontología resultante como base para los agentes, con la finalidad de que todos los agentes presenten un vocabulario común y especializado en esta área.
- 2) La segunda BDR se denomina "SMATERMINOS", esta Base de Datos contiene tres tablas, una para los términos clasificados de Facebook (FB), otra para los términos clasificados de Twitter (TW) y la última para los Blogs (B). Estas tablas contienen los mismos campos: `cve_termFB`, `nombre_termFB` y `desc_tablaFB`; las siglas FB cambian respecto al Sitio Web de donde se obtienen los términos, ej. TW y B, donde
 - `cve_term` es la clave identificadora del término recopilado del sitio Web especificado.

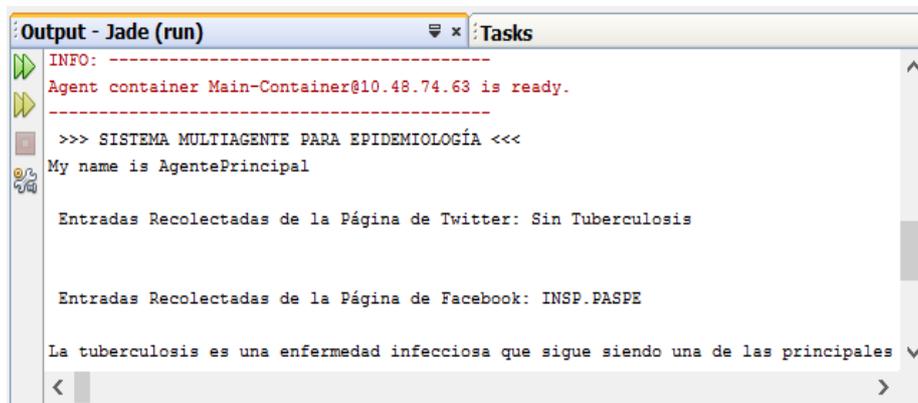
- nombre_term contiene los nombres de los términos nuevos recopilados que cada agente va clasificando.
- desc_tabla contiene el nombre de la tabla correspondiente a la BDR "tuberculosis_uno" que el agente establece como una posible pertenencia del término que se clasificó.

Como conocimiento adquirido, se propone incrementar el vocabulario de la ontología con los términos nuevos que cada agente clasifica en la BDR SMAterminos, de acuerdo con las entradas de los Sitios Web y la validación de los términos que realizará un especialista en el área; que, hasta el momento, es el encargado de introducir cada término manualmente a la BDR "tuberculosis_uno".

3. Prototipo

El SMA propuesto está codificado en el entorno de programación JADE versión 4.3.0 (Java) a través de NetBeans IDE 7.0.1; se ejecutó en una Laptop con las siguientes características: Procesador AMD Quad – Core a 3.5 GHz, 6 GB de memoria DDR3, 750 GB de Disco Duro, Sistema Operativo Windows 8 de 64 bits.

Al realizar la ejecución del Sistema Multiagente, hará que los tres Agentes (Agente Facebook = RSATFB, Agente Twitter = RSATTW, Agente Blog = RSATBLOG) comiencen a recopilar información y de esta manera, se clasifiquen los términos nuevos en la BDR.



```
Output - Jade (run) Tasks
INFO: -----
Agent container Main-Container@10.48.74.63 is ready.
-----
>>> SISTEMA MULTIAGENTE PARA EPIDEMIOLOGÍA <<<
My name is AgentePrincipal

Entradas Recolectadas de la Página de Twitter: Sin Tuberculosis

Entradas Recolectadas de la Página de Facebook: INSP.PASPE

La tuberculosis es una enfermedad infecciosa que sigue siendo una de las principales
```

Fig. 4. Ejemplo de recopilación de información tras la ejecución de todos los agentes del sistema.

En la Fig. 4 se muestra un ejemplo del funcionamiento de los agentes al realizar la ejecución del Sistema Multiagente. Donde se puede observar que se realiza la recopilación de entradas a través de dos páginas Web: la primera es del agente recopilador de Twitter en el perfil denominado “Sin Tuberculosis” [13], del que no se tienen resultados para nuestra área de estudio al momento de la recopilación de datos; y la segunda es de la página de Facebook denominada “INSP.PASPE” [14] (entrada

del agente recopilador de Facebook), en la que sí se visualiza una entrada para el área de Tuberculosis.

Es posible cambiar cada uno de los términos que “envía” o “busca” cada uno de los Agentes a través de sus respectivos Comportamientos. Para cambiar cada uno de los términos de cada Agente, se selecciona el Comportamiento EmitirBlog para el Agente Blog, el Comportamiento EmitirFacebook para el Agente Facebook y el Comportamiento EmitirTwitter para el Agente Twitter; una vez que se encuentra en el Agente que desea modificar, se localiza la variable “mensaje_a_enviar” y se modifica su valor por cualquier texto que se desee probar, ver Fig. 5.

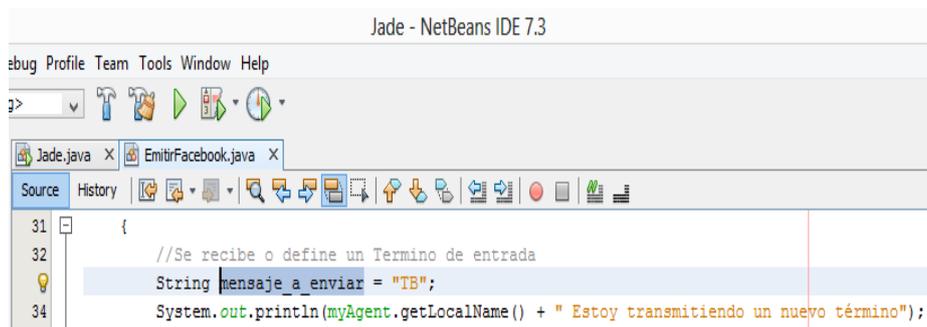


Fig. 5. Ejemplo de cambio de valor en la variable "mensaje_a_enviar" en EmitirFacebook.

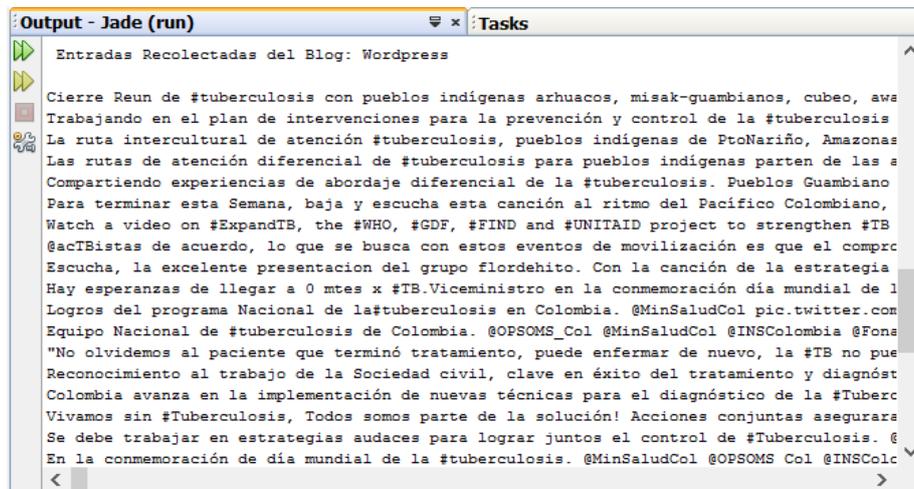


Fig. 6. Información obtenida del agente para BLOGS.

4. Experimentos y resultados

Los siguientes experimentos se llevaron a cabo.

- Extraer los términos de una publicación en un Sitio Web. La Fig. 6 muestra las entradas correspondientes a la página Wordpress [15]

indicada en la sección de BLOGs (correspondiente al agente recopilador de BLOGs).

- Clasificar los términos. La Fig. 7 muestra las tablas que conforman a la BDR SMAterminos, que contiene los términos clasificados por los agentes en cada Sitio Web y los nombres de las tablas de la BDR "tuberculosis_uno" a las que posiblemente pertenezcan los términos clasificados, éste se almacena en el campo desc_tabla. Estos dos valores son los que utiliza el especialista que valida los términos para verificar que efectivamente los agentes están realizando la clasificación correctamente y, de esta manera, el especialista al verificar que la actividad Clasificar está ejecutándose satisfactoriamente, los agentes podrán directamente vaciar los términos clasificados en la BDR de Tuberculosis sin tener que esperar la validación del especialista en el área.

```
mysql> use SMAterminos
Database changed
mysql> show tables;
+-----+
| Tables_in_smaterminos |
+-----+
| termino_clasificadobl |
| termino_clasificadofb |
| termino_clasificadotw |
+-----+
3 rows in set (0.19 sec)

mysql> describe termino_clasificadoFB;
+-----+-----+-----+-----+-----+-----+
| Field      | Type      | Null | Key | Default | Extra      |
+-----+-----+-----+-----+-----+-----+
| cue_termFB | int(11)   | NO   | PRI | NULL    | auto_increment |
| nombre_termFB | char(60) | YES  |     | NULL    |             |
| desc_tablaFB | char(30) | YES  |     | NULL    |             |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.27 sec)
```

Fig. 7. Estructura de la BDR SMAterminos.

- Comunicación entre los agentes. Se realiza por medio del protocolo Emitir mediante performativas que son enviadas entre los agentes; las performativas utilizadas son *REQUEST*, *INFORM* y *CONFIRM*. Tal como se indica en el código siguiente:

```
//Se establece la estructura para el envío de mensajes de tipo respuesta
con
    //la performativa INFORM.
    public void EnviarRespuesta(ACLMessage mensaje, String contenido1)
    {
        ACLMessage responder = mensaje.createReply();
        responder.setPerformative(ACLMessage.INFORM);
        responder.setContent(contenido1);
        myAgent.send(responder);
    }
    //Se establece la estructura para el envío de mensajes de tipo
    //agradecimiento con la performativa CONFIRM.
    public void EnviarAgradecimientos(ACLMessage mensajeG, String
    contenidoG)
    {
        ACLMessage responderG = mensajeG.createReply();
        responderG.setPerformative(ACLMessage.CONFIRM);
```

```
        responderG.setContent(contenidoG);
        myAgent.send(responderG);
    }

    //Se establece la estructura para el envio de mensajes de tipo
    pregunta con
    //la performativa REQUEST.
    public void EnviarMensaje(String receptor, String lenguaje, String
    contenido, String IdConversacion)
    {
        ACLMessage mensaje = new ACLMessage(ACLMessage.REQUEST);
        mensaje.setLanguage(lenguaje);
        mensaje.setContent(contenido);
        mensaje.setConversationId(IdConversacion);
        mensaje.addReceiver(new AID(receptor, AID.ISLOCALNAME));

        myAgent.send(mensaje);
    }
}
```

- Inteligencia colectiva al identificar los términos ya clasificados en las entradas. Un ejemplo de esto es cuando el agente Twitter le pregunta al agente Facebook si tiene clasificado el término TB y como éste ya lo tiene clasificado, le responde al agente Twitter que ya lo tiene clasificado, ver Fig. 8.

```
Request from RSATTW
Conexión a la BD
Resultado de la BD TB
Término del agente que envió: TB
TERMINO CLASIFICADO
```

Fig. 8. Ejemplo de un término clasificado por el agente Facebook.

- Planeación en las actividades. Para realizar la clasificación de los términos, las actividades deben seguir un orden; ya que no se puede indicar que un término está clasificado, si antes no se ha almacenado en la BDR SMAterminos. Además, no se puede obtener un término, si no se ha realizado el parseo correspondiente a las entradas de los usuarios.
- Se utilizan *Strings* que representan contenidos de mensajes (Ontología básica), tal como se muestra en la Fig. 9.

```
//Se recibe o define un Termino de entrada
String mensaje_a_enviar = "Poblacion contagiada";
System.out.println(myAgent.getLocalName() + " Estoy transmitiendo un nuevo término");
//Se crean los mensajes bajo el estandar FIPA a ser enviados
messageHandler.EnviarMensaje("RSATTW","Español", mensaje_a_enviar ,"NuevoTermino");
block(1000);
messageHandler.EnviarMensaje("RSATFB","Español",mensaje_a_enviar,"NuevoTermino");
block(1000);
```

Fig. 9. Ejemplo de código para la propagación de mensajes entre agentes.

- Enriquecer ontologías especializadas con los términos y relaciones previamente identificadas (manualmente); de igual manera, mantener las ontologías actualizadas (manualmente) [3]. Como resultado se tiene el mapeo manual de la BDR "tuberculosis_uno" para obtener la ontología

actualizada en formato RDF y .ttl para su explotación, Fig. 10 y 11, respectivamente.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:db="http://localhost/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:vocab="http://localhost/vocab/"
  xmlns:map="file:tmp#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
  <rdf:Description rdf:about="cat_municipio/15/1/1">
    <rdfs:label>cat_municipio #15/1/1</rdfs:label>
    <vocab:cat_municipio_cve_jur rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</vocab:cat_municipio_cve_jur>
    <vocab:cat_municipio_cve_edo rdf:datatype="http://www.w3.org/2001/XMLSchema#int">15</vocab:cat_municipio_cve_edo>
    <vocab:cat_municipio_cve_mpo rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</vocab:cat_municipio_cve_mpo>
    <vocab:cat_municipio_des_mpo>ACAMBAY</vocab:cat_municipio_des_mpo>
    <rdf:type rdf:resource="vocab/cat_municipio"/>
  </rdf:Description>
```

Fig. 10. Ontología obtenida en formato RDF.

```
# Table cat_estado
map:cat_estado a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "cat_estado/@@cat_estado.cve_edo@@";
  d2rq:class vocab:cat_estado;
  d2rq:classDefinitionLabel "cat_estado";
.
map:cat_estado__label a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cat_estado;
  d2rq:property rdfs:label;
  d2rq:pattern "cat_estado #@@cat_estado.cve_edo@@";
.
map:cat_estado_cve_edo a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cat_estado;
  d2rq:property vocab:cat_estado_cve_edo;
  d2rq:propertyDefinitionLabel "cat_estado cve_edo";
  d2rq:column "cat_estado.cve_edo";
  d2rq:datatype xsd:int;
.
map:cat_estado_des_edo a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cat_estado;
  d2rq:property vocab:cat_estado_des_edo;
  d2rq:propertyDefinitionLabel "cat_estado des_edo";
  d2rq:column "cat_estado.des_edo";
.
```

Fig. 11. Ontología obtenida en formato .ttl.

5. Conclusiones y trabajo futuro

En este artículo se presenta un Sistema Multiagente que propone un marco (framework) para realizar el enriquecimiento de una ontología al relacionar información proveniente de Bases de Datos relacionales (información estructurada) con información proveniente de la Web (información semiestructurada). Este marco utiliza la comunicación a través del protocolo FIPA proporcionando un marco común entre los mensajes que deben conocer los agentes para un mejor desempeño en el

SMA. Proporcionando de esta manera, una herramienta de apoyo a los expertos en el área de Tuberculosis para la recopilación de información en Sitios Web que pudiera ser de relevancia para la actualización de la ontología de Tuberculosis.

Este tipo de propuestas pueden ayudar a extender el conocimiento en áreas que necesitan manejo de datos y experimentación con información diversa para su implementación. El caso de la Epidemiología, y en particular de la Tuberculosis, muestran un campo fértil para poder compaginar todo tipo de información y ayudar a que pequeños focos de infecciones no crezcan demasiado y que la misma sociedad pueda incluso ayudar a combatir las enfermedades.

Como trabajo futuro de esta propuesta, puede concentrarse en la creación de una ontología de uso automática y su actualización cuando ya se tiene implementada en los agentes. De esta manera, se pueden considerar los siguientes puntos:

- La inclusión de un parser que extraiga cada término de las entradas de los Sitios Web considerando la semántica de su contexto.
- La implementación de la ontología de Tuberculosis en los agentes, que se encuentre actualizada con los términos nuevos que se vayan clasificando, cada periodo de tiempo determinado.
- Un agente que se encuentre monitoreando las páginas de Internet que se hayan elegido para la extracción de sus entradas, con la finalidad de verificar la disponibilidad de los mismos en la Web; así como para determinar nuevas páginas que se podrían incorporar para obtener información de acuerdo con la ontología especificada.

Referencias

1. WORLD WIDE WEB consortium Oficina Española, <http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica#>.
2. Junco R., María de los A. Introducción Sistemas Multiagente. <https://sites.google.com/a/itesm.mx/ajunco/home/sma>.
3. González, M. D.: Marco Ontológico para la Interoperabilidad en la Explotación de Información en Bases De Datos Heterogéneas. Ph.D. Tesis Doctoral, ITESM Campus Estado de México (2012)
4. Martínez, V. M. de L., González M. M., y Hernández G. N.: Towards a Ubiquitous User Model for Profile Sharing and Reuse. In: Sensors (2012)
5. Julian V., Carrascosa C. y Soler J.: Una arquitectura de sistema multi-agente para la recuperación y presentación de información. Universidad Politécnica de Valencia, Departamento de Sistemas Informáticos y Computación. Valencia (1999)
6. Muñoz N., Cobos C., Rivera W., López J., Mendoza M.: Uso de la metodología GAIA para modelar el comportamiento de personajes en un juego de estrategia en tiempo real. Univ. Antioquia. 53, 214–224 (2010)
7. Junco R., María de los A. Sistemas Multiagente, <https://sites.google.com/a/itesm.mx/ajunco/home/sma>
8. Jade Tutorial and Primer, <http://www.iro.umontreal.ca/~vaucher/Agents/Jade/JadePrimer.html>
9. Programación Jade, <http://www.programacionjade.wikispaces.com/Agentes#BookBuyerAgent>

10. Jade Tutorial, Jade Programming for Beginners,
<http://jade.tilab.com/doc/tutorials/JADEProgramming-Tutorial-for-beginners.pdf>
11. Agents step by step,
<https://docs.google.com/presentation/d/1XgzVSKDxcTrYMzMdvvI9IGk5011IuifAQHH04BP7zys/edit#slide=id.i0>
12. Barrera M., Núñez H., y Ramos E.: Ingeniería Ontológica, Lecturas en Ciencias de la Computación. Universidad Central de Venezuela, Facultad de Ciencias, Escuela de Computación. Caracas (2012)
13. Sin Tuberculosis, <https://twitter.com/SinTuberculosis>
14. INSP-Programa de Actualización en Salud Pública y Epidemiología,
<https://www.facebook.com/INSP.PASPE>
15. WordPress.com, <http://es.wordpress.com/tag/tuberculosis/>

Recomendaciones contextuales basadas en el enfoque de post-filtrado

Xochilt Ramírez-García, Mario García-Valdéz

Instituto Tecnológico de Tijuana, Tijuana, B.C.,
México

xochilt.ramirez@gmail.com, mariosky@gmail.com

Resumen. Hoy en día, los sistemas de recomendación son utilizados en diferentes tipos de aplicaciones. Un sistema de recomendación de restaurantes proporciona recomendaciones considerando los gustos y preferencias del usuario con el propósito de facilitar al usuario la selección de restaurantes que posteriormente visitará. El objetivo de este trabajo es desarrollar una aplicación para hacer recomendaciones de restaurantes que permitan difundir información sobre productos y servicios ofrecidos en Tijuana a través de una página Web. El algoritmo de recomendación propuesto se basa en el enfoque de post-filtrado, utiliza el algoritmo de filtrado colaborativo y la información contextual de la situación actual del usuario para obtener recomendaciones contextuales. El conjunto de datos fue adquirido a través de cuestionarios contestados por 50 personas, se hicieron recomendaciones con una matriz de valoraciones que contiene 1,422 votos contextuales, 50 usuarios y 40 restaurantes. Se utilizó el Error Absoluto Medio para evaluar el desempeño del sistema de recomendación.

Palabras clave: Sistemas de recomendación, filtrado colaborativo, contexto, post-filtrado.

1. Introducción

Los seres humanos tienen un proceso de comunicación para transmitir ideas y reaccionan apropiadamente, esto se debe a factores como la riqueza del lenguaje y la comprensión de las palabras en diferentes situaciones. Cuando las personas conversan usan implícitamente información situacional (o contextual) logrando aumentar el rango de la conversación. Utilizando la información contextual se incrementa la comunicación en la interacción humano-computadora, esto contribuye a mejorar el desarrollo de aplicaciones inteligentes. Este proceso también es aplicado en los sistemas de recomendación sensibles al contexto.

Actualmente los sistemas de recomendación pueden ser usados en aplicaciones modernas que muestran al usuario una amplia variedad de ítems. Tales sistemas, generalmente proporcionan al usuario una lista de ítems que prefiere o predicen el

grado de cuánto podría gustarle un cierto ítem al usuario, el objetivo es ayudarlo a seleccionar un ítem dentro una gran cantidad de estos.

La mayoría de sistemas de recomendación se centran en los ítems más relevantes para el usuario y no consideran información contextual como tiempo, lugar o acompañantes, por ejemplo para ver una película [9] o salir a cenar [22]. En otras palabras, los sistemas de recomendación trabajan con las aplicaciones considerando solo dos tipos de entidades: usuarios e ítems, y no utilizan el contexto cuando proveen recomendaciones al usuario.

En aplicaciones como recomendación de paquetes turísticos [6, 7], recomendación de contenidos personalizados en una página Web [18], o recomendación de una película [16], puede ser insuficiente solo considerar información acerca de los usuarios e ítems, es importante considerar información contextual en el proceso de recomendación con el fin de sugerir ciertos ítems en ciertas circunstancias [28]. Recientemente, existe una gran cantidad de investigaciones en el área de sistemas de recomendación y la mayoría se enfoca en diseñar algoritmos para recomendaciones contextuales [6, 19, 15]. Esta investigación propone el uso del enfoque de post-filtrado que utiliza el algoritmo de filtrado colaborativo y la información contextual del usuario en el dominio de restaurantes.

Los temas en el trabajo se presentan como sigue: en la sección 2, se describen los conceptos relacionados; en la sección 3, se explica la arquitectura del sistema de recomendación propuesto; en la sección 4, se define la configuración del ambiente, el conjunto de datos obtenido, las métricas y los experimentos para evaluar el algoritmo; en la sección 5, se explican los resultados de la evaluación; y en la sección 6, se explican las conclusiones del estudio, los puntos más importantes del enfoque, las lecciones aprendidas y el trabajo futuro.

2. Sistemas de recomendación

Los sistemas de recomendación forman parte de la solución al problema de la sobrecarga de información que sufren los consumidores en Internet. Estos sistemas han sido probados para ser una de las técnicas más utilizadas que ayudan a los usuarios a encontrar ítems relevantes para ellos [2].

2.1. Algoritmo de filtrado colaborativo

Un algoritmo de filtrado colaborativo [2] hace recomendaciones analizando la correlación entre usuarios. El algoritmo calcula la similaridad entre los usuarios para hacer predicciones. Por ejemplo, suponiendo que el usuario A es similar al usuario B, si el usuario A está interesado en el producto X, entonces se puede inferir que el usuario B también estará interesado en el producto X, entonces el producto X será recomendado al usuario B. El enfoque de filtrado colaborativo necesita recolectar información de los votos de los usuarios sobre los ítems, e identificar vecindarios tomando en cuenta la similaridad entre estos. Este enfoque se basa en el supuesto de

que cuando las personas seleccionan productos o información son influenciadas por las opiniones de otras personas que han tenido alguna experiencia con el producto o la información. Además, en el filtrado colaborativo, el análisis del contenido de los ítems es ignorado y solo usa las opiniones sobre los ítems que son considerados más relevantes. Esto es importante para recomendar ítems para los cuales el análisis de contenido es escaso [26].

2.2. Contexto

En [4] definen el contexto como el estado físico, social, emocional o formativo en el del usuario. El contexto es toda la información acerca de la situación pertinente a una aplicación y su conjunto de usuarios, aunque en ocasiones, no toda la información es considerada relevante para la aplicación. Es imposible enumerar cuáles son los aspectos importantes de todas las situaciones, ya que esto cambia de una situación a otra, por ejemplo, en algunos casos el entorno físico puede ser importante mientras que en otros puede ser completamente irrelevante.

Definición. El contexto es cualquier información que se puede utilizar para caracterizar la situación de una entidad. Una entidad es una persona, lugar u objeto que se considera relevante para la interacción entre un usuario y una aplicación, incluyendo el usuario y aplicaciones en sí [4].

La definición anterior facilita la enumeración de los aspectos del contexto de una aplicación dada. Si una pieza de información puede ser utilizada para caracterizar la situación de un participante en una interacción, entonces esta información será el contexto [13].

2.3. Enfoque de post-filtrado

El enfoque post-filtrado inicialmente ignora la información contextual en la primera etapa de las recomendaciones, es decir, cuando se genera la lista ordenada de los N ítems seleccionados para la lista de recomendaciones. Posteriormente, el enfoque post-filtrado ajusta la lista de recomendaciones para cada usuario utilizando la información contextual. Los ajustes de la lista de recomendación pueden ser hechas de dos maneras [3]:

- Filtrando las recomendaciones que son irrelevantes (en un contexto dado), o
- Ajustando la clasificación de las recomendaciones de la lista (en base al contexto dado).

En la figura 1 se representa el esquema del proceso de post-filtrado contextual. En este proceso se analizan los datos de preferencia de contexto para un usuario determinado en un contexto dado para encontrar los patrones de uso específicos de los ítems y luego usar estos patrones para ajustar la lista de elementos, dando lugar a las recomendaciones contextuales. En otras palabras, el post-filtrado desecha los ítems recomendados que son menos relevantes en el contexto específico y mantiene los elementos más relevantes. La ventaja más importante del post-filtrado contextual es

que permite el uso de cualquiera de las numerosas técnicas de recomendación tradicionales previamente propuestas en la literatura [2].

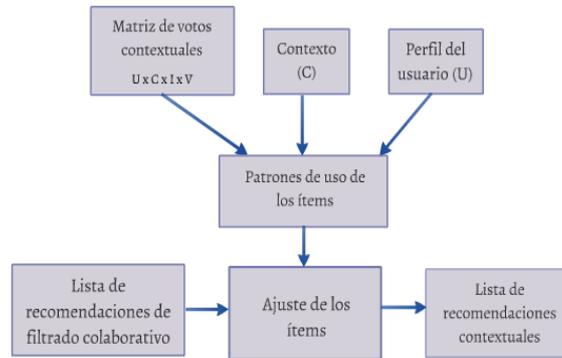


Fig. 1. Esquema del proceso de post-filtrado para obtener recomendaciones contextuales [3].

3. Sistema de recomendación de restaurantes

El sistema de recomendación de restaurantes utiliza el algoritmo de filtrado colaborativo para encontrar restaurantes relevantes para el usuario [22]. Las votaciones en el perfil de usuario (representado como un vector de enteros) son utilizadas para determinar la similitud entre estos, es decir, qué tan parecidos son unos con otros considerando el perfil de cada usuario. El vecindario para el usuario es obtenido del conjunto de usuarios con mayor similitud. La similitud es calculada con la correlación de Pearson. Posteriormente, el algoritmo obtiene una lista de los 10 mejores restaurantes recomendados para el usuario. La salida del algoritmo de filtrado colaborativo suministra la siguiente etapa de post-filtrado y en la lista de recomendaciones el algoritmo agrega el promedio de los votos para cada ítem de la lista.

La etapa de filtrado de restaurantes por contexto consiste clasificar todos los votos en el contexto que corresponde, es decir, de acuerdo a la fecha en que fueron votados el sistema los clasifica en el contexto “entre semana” o “fin de semana”.

El contexto actual es la información acerca del momento en que el usuario está interactuando con el sistema. En la etapa de selección de restaurantes en el contexto actual, el sistema analiza cuáles son los restaurantes votados en el mismo contexto del usuario actual y guarda esos datos para empezar a compararlos con los ítems que están en el mismo contexto, finalmente obtiene una lista de ítems recomendados contextualizada.

En la etapa del cálculo de la predicción por promedio, el sistema obtiene el promedio de los votos que están en el mismo contexto para cada ítem de la lista contextualizada. Enseguida, para calcular la predicción por promedio, se utiliza: 1) la predicción del algoritmo de filtrado colaborativo del ítem y, 2) el promedio del ítem en el contexto actual, es decir, el promedio de las votaciones de los usuarios

previamente calculado. La lista de recomendaciones contextuales obtenida en la última etapa contiene los restaurantes con altas predicciones (basada en el umbral de 3.5). El esquema del proceso se representa en la figura 2.

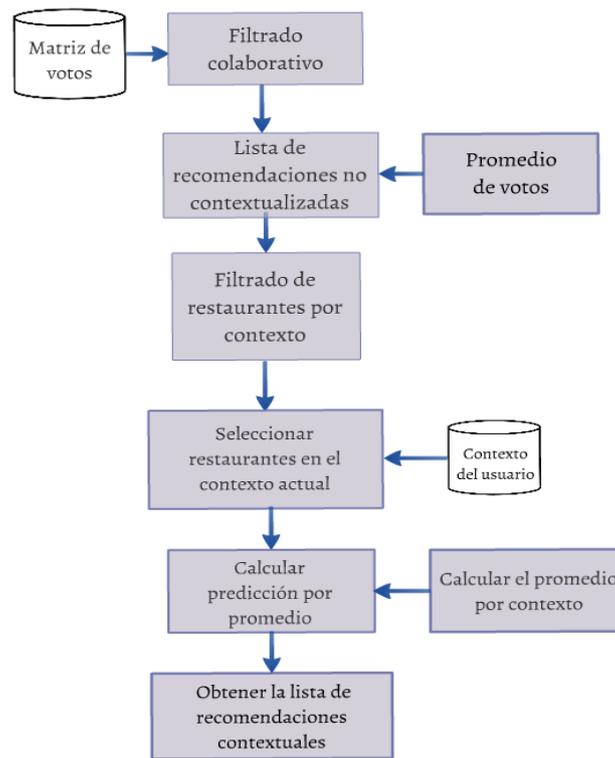


Fig. 2. Metodología de post-filtrado en el sistema de recomendación de restaurantes.

4. Configuración

4.1. Conjunto de datos

Para recolectar el conjunto de datos se aplicó un cuestionario a los estudiantes de Ingeniería en Sistemas, estudiantes de Maestría en Ciencias Computacionales y profesores del Instituto Tecnológico de Tijuana. Un total de 50 personas respondieron el cuestionario, las preguntas solicitaban información principalmente acerca de sus gustos y preferencias sobre los restaurantes cercanos al Instituto Tecnológico de Tijuana y la tecnología o recursos utilizados para buscar opciones de restaurantes. El cuestionario consta de 8 preguntas donde las personas, asignan votos a los restaurantes que han visitado de una lista de 40. Cada restaurante elegido fue calificado 6 veces, una para cada contexto, sin embargo, hubo quienes votaban menos contextos y quienes no votaban ningún restaurante. Al final de la encuesta se

obtuvieron un total de 1,422 votos. Las preguntas del cuestionario se enlistan en la tabla 1 al igual que las posibles respuestas.

Tabla 1. Cuestionario para obtener la retroalimentación del usuario.

Pregunta	Respuesta
1. ¿Cuál es tu ocupación?	1. Estudiante 2. Empleado
2. De acuerdo a su prioridad, ordene por importancia las características que consideras cuando escoges un restaurante.	1. Instalacion 4. Platos 2. Precio 5. Ambiente 3. Servicios 6. Ubicación
3. ¿Qué tecnología utilizas con más frecuencia?	1. Teléfono inteligente 3. Laptop 2. Tableta 4. PC
4. ¿Qué sistemas operativos utilizas?	1. Android 4. Symbian 2. Windows 5. BlackberryOS 3. iOS 6. Otro
5. ¿Has utilizado alguna aplicación para buscar restaurantes en Tijuana?	1. Si 3. ¿Cuál? 2. No
6. ¿Te gustaría usar una aplicación de recomendación de restaurantes de Tijuana?	1. Si 2. No
7. Asigne una calificación a los restaurantes que prefieras sin considerar situaciones contextuales.	Lista de Restaurantes de Tijuana
8. Asigne una calificación a los restaurantes que prefieras considerando situaciones contextuales.	Lista de Restaurantes de Tijuana

Posteriormente, el total de las respuestas a las preguntas 1-6 se muestran en la Tabla 2. Las respuestas de las preguntas 2, 3 y 4 demuestran las preferencias de los usuarios. La respuesta de la pregunta 2 demuestra que tanta importancia tiene cada uno de los elementos propuestos, es decir, cuál es el factor que determina la toma de decisión al elegir visitar un restaurante. En la tabla 2 está la cantidad de usuarios que prefieren como su favorito cada uno de los factores propuestos. En la respuesta de la pregunta 3, los usuarios revelan cuál es el dispositivo que utilizan comúnmente para buscar opciones de restaurantes. En su mayoría los usuarios utilizan laptops, las prefieren por su portabilidad y fácil acceso a las redes y enseguida los teléfonos celulares inteligentes que proporcionan los mismos servicios y accesos. La respuesta de la pregunta 3, demuestra que Windows y Android son los sistemas operativos que utilizan la mayoría, siendo estos los más fáciles de usar y obtener para estudiantes y maestros. Las respuestas de las preguntas 5 y 6 demuestran lo inusual que es el uso de un sistema de recomendación de restaurantes entre los estudiantes y maestros, siendo esta una motivación más para desarrollar la aplicación.

Tabla 2. Total de respuestas de los usuarios obtenidas en el cuestionario.

Pregunta	Respuestas					
	1	2	3	4	5	6
1	4	46				
2	6	9	8	13	2	10
3	30	8	35	17		
4	38	39	20	4	6	5
5	6	42				
6	46	3				

La respuesta a las preguntas 7 y 8 se representan con las graficas de la figura 3, donde se observa la cantidad de votos obtenidos para cada restaurante, siendo estos los 10 restaurantes favoritos de los usuarios sin considerar información contextual (figura a) y los 10 restaurantes favoritos considerando información contextual (figura b).

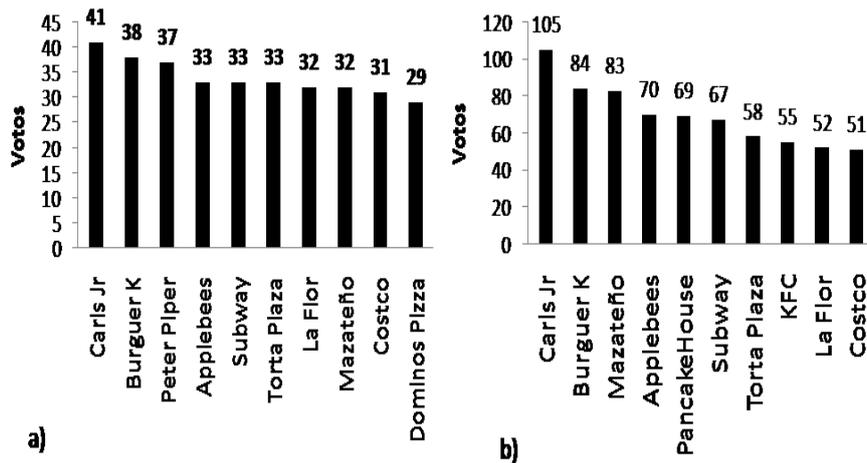


Fig. 3. Gráficas de las preferencias de los usuarios sobre los restaurantes propuestos en el sistema de recomendación.

En la gráfica a), los restaurantes favoritos de acuerdo a los votos de los usuarios fueron “Carls Jr” y “Burger King” con mayor cantidad de votos, en la gráfica b), se aprecia la diferencia a partir del tercer restaurante, donde los restaurantes cambian de posición o simplemente no aparecen en alguna lista. Las recomendaciones contextuales del enfoque de post-filtrado depende del factor contextual "Día de la semana", que significa el día en que se votaron los restaurantes.

Posteriormente, el resultado de la consulta es refinada de acuerdo al contexto del usuario actual; los 6 contextos mencionados que corresponden a combinaciones de factores contextuales se describen en la Tabla 3.

Tabla 3. Factores contextuales considerados en el post-filtrado.

Factor contextual	Dominio
Día de la semana	<ul style="list-style-type: none"> – Entre semana (lunes, martes, miércoles y jueves) – Fin de semana (viernes, sábado y domingo)
Lugar	<ul style="list-style-type: none"> – Escuela – Casa – Trabajo

4.2. Métricas

Sistemas de Recomendación son ampliamente utilizados en diferentes dominios, por lo tanto, los objetivos pueden ser diferentes en cada aplicación, las métricas utilizadas dependen del objetivo del sistema. Para algunas aplicaciones la precisión de la predicción es fundamental y la evaluación del sistema se centra en la precisión de las recomendaciones. Existen varios indicadores para medir la precisión: Error Cuadrático Medio (RMSE por sus siglas en inglés) [9, 10], el Error Absoluto Medio (MAE por sus siglas en inglés) [9] y la Precisión y Relevancia [9, 10], son algunos ejemplos. Por otro lado, si el sistema intenta medir aspectos cualitativos como la satisfacción del usuario, calidad de las recomendaciones o utilidad de sistema para un grupo de usuarios o comunidad, se consideran otras métricas como se menciona en [14].

Precisión en la predicción. Precisión de la predicción es la propiedad más discutida en la literatura sistema de recomendación. El centro de la gran mayoría de los sistemas de recomendación son motores de predicción. Este motor puede predecir las opiniones de los usuarios sobre los elementos (por ejemplo, las calificaciones de las películas) [9] o la probabilidad de uso (por ejemplo en compras de consumidores) [24]. Un supuesto básico en los sistemas de recomendación es que el usuario va a preferir un sistema que proporciona predicciones más precisas, a raíz de esto muchos investigadores se han propuesto encontrar algoritmos que proporcionen mejores predicciones. La precisión de la predicción es normalmente independiente de la interfaz del usuario, por lo que se puede medir en un experimento off-line. La medición de precisión en un estudio de usuarios significa medir la exactitud dada una recomendación. Este es un concepto diferente de la predicción del comportamiento de los usuarios sin recomendaciones, y está más cerca de la verdadera exactitud en el sistema real [14].

Error Absoluto Medio (MAE). El Error Absoluto Medio es una métrica popular usada en la evaluación de la precisión de las predicciones, el sistema genera predicciones \hat{r}_{ui} para un conjunto de datos de prueba \mathcal{T} de los pares usuario-item (ui) para el cual el verdadero rating r_{ui} es conocido. Típicamente, r_{ui} es conocido

porque está oculto en un experimento off-line, o porque fue obtenido a través de un estudio del usuario o experimento on-line [14]. El error entre la predicción y el voto real es dado por la siguiente ecuación (1):

$$\text{MAE} = \sqrt{\frac{1}{|\mathcal{F}|} \sum_{(u,i) \in \mathcal{F}} |\hat{r}_{ui} - r_{ui}|} \quad (1)$$

5. Resultados del experimento

El algoritmo fue desarrollado en el lenguaje Python y probado con un gestor de base de datos PostgreSQL, el conjunto de datos se recolectó explícitamente de 50 usuarios que contestaron el cuestionario detallado en la sección 4.1. La figura 5 muestra la distribución de los votos en una matriz tridimensional obtenida con los votos (en una escala de 1 a 5), 40 restaurantes y 6 diferentes situaciones contextuales, en total la matriz que contiene 1,422 votos. Se realizó un total de 172 predicciones para usuarios tomados aleatoriamente y el error fue de 0.58. Las pruebas se hicieron con diferentes usuarios y en diferentes contextos, el error calculado se obtiene de todas las predicciones hechas aleatoriamente. El resultado está limitado al tamaño del conjunto de datos, que en este caso particular restringe el rendimiento del algoritmo. Por otro lado, tener el “día de la semana” como principal factor contextual no mejora la precisión de las recomendaciones en este dominio.

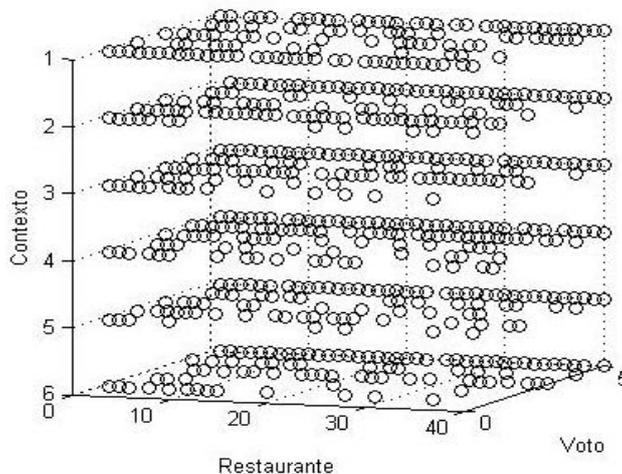


Fig. 4. Distribución de los votos en los 6 contextos propuestos.

En un sistema de recomendación que utiliza algoritmos sin información contextual, las predicciones pueden ser menos precisas con un error de 0.77, como se ha hecho anteriormente en trabajo previo [22], sin embargo estas predicciones están basadas solamente en el perfil de usuario (usuario-ítem) y no consideran información diferente de los votos (filtrado colaborativo) o diferente a la descripción de los restaurantes

(basado en contenido). Además el conjunto de datos utilizado en el trabajo previo no era más grande que el que proponemos en este trabajo, el tamaño del conjunto de datos afecta el desempeño del algoritmo.

Al agregar información contextual la complejidad del proceso de recomendación en el post-filtrado aumenta. La matriz de votos contextuales en cualquier caso tendrá al menos 3 dimensiones (usuario-ítem-contexto), cada factor contextual agregado representa una dimensión más en la misma. Por lo tanto, hacer la selección de la información contextual para el usuario es una tarea compleja que aumenta cuando los conjuntos de datos son robustos pero a la vez permite obtener recomendaciones más precisas y apegadas a la realidad del usuario.

6. Conclusiones y trabajo futuro

La cantidad de información que se suministra para el sistema es fundamental para lograr recomendaciones más exactas [26]. El problema de la escasez de información contextual [20] conduce a malas recomendaciones y es por lo tanto, un problema latente en este trabajo debido al tamaño del conjunto de datos con poca información contextual. Otro problema encontrado al analizar los resultados del cuestionario, es que los usuarios no cambian realmente sus calificaciones cuando se le preguntó a considerar un contexto particular. Una posible explicación de este comportamiento la carga cognitiva adicional que los usuarios tienen al pensar en una diferencia en la preferencia en cada escenario, es posible que su preferencia pueda ser determinada independientemente del contexto y no es hasta que se enfrenta realmente con el problema de tomar la decisión de elegir un restaurante entre distintos competidores en un contexto particular, cuando los usuarios asignan un rating de utilidad a cada opción elegida.

Los factores contextuales utilizados en otras investigaciones [6], [19], [11] son basados en el dominio del sistema de recomendación, estos factores se pueden incluir desde las consultas a la base de datos y se utilizan como referencia para la clasificación de los ítems [19], en base a las preferencias del usuario [22], o a los modelos de probabilidad para predecir el comportamiento de los usuarios [23].

Sin embargo, en el experimento se identificó que los usuarios tenían opiniones diferentes cuando asumen situaciones contextuales y cuando ocurre la situación en la vida real. La selección de un restaurante podría verse afectada no sólo por los gustos y preferencias de los usuarios, sino también por otros factores que no son percibidos en esas suposiciones. El usuario no toma decisiones teniendo en cuenta los factores contextuales, como los días de la semana, el dispositivo móvil o la distancia, de hecho, sería interesante preguntarse cómo influye la distancia cuando el usuario es estudiante y cómo cambia la decisión cuando el usuario no es un estudiante, o cómo la distancia influye en la decisión cuando el usuario interactúa con diferentes dispositivos. Otros factores (los que podrían ser subjetivos) son cada vez más relevantes: 1) las relaciones personales (amigos, familiares y compañeros de trabajo), 2) el dinero con que cuenta en ese momento, 3) el tipo de alimentos que el usuario prefiere consumir o que se le antoja debido al clima ó a su estado de ánimo, ó 4)

simplemente por el hábito de ir al mismo lugar. Somos conscientes de que hay una diferencia en la selección del restaurante, los deseos del usuario (lo que quiere el usuario en ese momento) son más importantes que sus preferencias ó gustos, y también es prioritario considerar la influencia del razonamiento lógico para decidir qué restaurante visitará.

Por otra parte, la información contextual obtenida de los usuarios a través del cuestionario no es totalmente fiable, la razón es que la mayoría de los estudiantes no mostraron una diferencia entre visitar un lugar en fin de semana o entre semana, el voto fue el mismo en los diferentes contextos. En el peor de los casos, los estudiantes solo ignoraron la pregunta dejando votos vacíos que el sistema no puede inferir si la información es escasa o nula.

En este estudio se observó que para los usuarios hacer suposiciones al votar fue extenuante [8], esta razón no facilita obtener información relevante, probablemente, las respuestas de los estudiantes no fueron las correctas en comparación con una situación de la vida real donde el usuario antes de seleccionar piensa acerca de los criterios que deben cumplirse para visitar un restaurante. Entonces, el "día de la semana" y "el lugar" no se consideran factores primordiales para estas recomendaciones contextuales, el experimento demostró que para los estudiantes no es importante que la situación contextual cambie.

De acuerdo a las lecciones aprendidas en este estudio, tomando en cuenta el desempeño del algoritmo y con el objetivo de mejorarlo, las propuestas para trabajo futuro son las siguientes:

- El utilizar en el factor contextual "día de la semana" un mayor rango del dominio, que permita al usuario identificar claramente las diferencias cuando el contexto cambia.
- Utilizar la localización geográfica para identificar los puntos donde se encuentra el usuario y las ubicaciones de los restaurantes, esto no limitaría al usuario a consumir en restaurantes cercanos al Instituto Tecnológico de Tijuana.
- Utilizar otras maneras de obtener información relevante del contexto con un esfuerzo mínimo de usuario, puede hacerse de forma implícita, como se menciona en [17].
- Implementar otro enfoque sistema de recomendación (por ejemplo, pre-filtrado) y compararlo con las recomendaciones de post-filtrado, esto con el fin de evaluar la precisión de la lista de recomendaciones en ambos casos.
- Probar y evaluar el algoritmo en un dominio diferente para observar el rendimiento con diferente información contextual.
- Para mejorar los resultados se deben aplicar más pruebas al sistema de recomendación con grandes conjuntos de datos y en otros dominios (por ejemplo, cursos de aprendizaje en línea), donde la información contextual es fundamental para lograr recomendaciones más precisas.

Referencias

1. Abbar, S., Bouzeghoub, M., Lopez, S.: Context-Aware Recommender Systems: A service oriented approach. In: VLDB PersDB Workshop (2009)

2. Adomavicius, G.; Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *Transactions on Knowledge and Data Engineering*, vol. 17, pp. 734–749 (2005)
3. Adomavicius, G., Tuzhilin, A.: Context-Aware Recommender Systems. *Recommender Systems Handbook*. pp. 217–253 (2011)
4. Dey, A., Abowd, K., Gregory, D.: Towards a better understanding of context and context-awareness. *Graphics, Visualization and Usability Center and College of Computing, Georgia Institute of Technology, Atlanta, GA.* (1999)
5. Baltrunas, L., Ricci, F.: Context-Based Splitting of Item Ratings in Collaborative Filtering. In: *Proceedings of the ACM Conference Series on Recommender Systems* (2009)
6. Baltrunas, L., Ludwig, B., Peer, S., Ricci, F.: Context-Aware Places of Interest Recommendations and Explanations. *Free University of Bozen-Bolzano, Bolzano, Italy* (2011)
7. Baltrunas, L., Ludwig, B., Peer, S., Ricci, F.: Context relevance assesment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing, Free University of Bolzano, Bolzano, Italy* (2011)
8. Baltrunas, L., Ludwig, B., Ricci, F.: Context Relevance Assessment for Recommender Systems. In: *Proceedings of the International Conference on Intelligent User Interfaces*. pp. 287–290 (2011)
9. Campochiaro, E., Cassata, R., Cremonesi, P., Turrin, R.: Do Metrics Make Recommender Algorithms? In: *Proceedings of International Conference on Advanced Information Networking and Applications Workshops, Milano, Italy* (2009)
10. Caraciolo, M.: Artificial Intelligence in Motion. <http://aimotion.blogspot.mx/2011/05/evaluating-recommender-systems.html>. Fecha de consulta: noviembre del 2013.
11. Chung-hua, C., Se-Hsien, W.: A chinese Restaurant Recommendation System based on Mobile context-Aware Services. In: *Proceedings of 14th International Conference on Mobile Data Management, IEEE, National Taichung University of Science and Technology* (2013)
12. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: *Recommender Systems: An Introduction*. Cambridge University Press (2010)
13. Fischer, G.: *Context-Aware Systems – The Right Information, at the Right Time, in the Right Place, in the Right Way, to the Righth Person*. ACM, University of Colorado, Boulder, USA (2012)
14. Guy, S., Gunawardana, A.: *Evaluating Recommendation Systems*. Microsoft Research (2009)
15. Hussein, T., Linder, T., Gaulke, W., Ziegler, J.: *Context-Aware recommendations on rails*. (2009)
16. Kyung-Rog, K., Ju-Ho, L., Jae-Hee, B.: Recommender System Using the Movie Genre similarity in Mobile Service. *Proceedings of 4th Internartional Conference on Multimedia and Ubiquitous Engineering (MUE)* (2010)
17. Lakshmith, R. et al.: CAESAR: A context-aware, social recommender system for low-end mobile devices. *Proceedings of International Conference on mobile Data Management: systems, and services and middleware, IEEE* (2009)
18. Martinez, M.L., Calles, G.J., Martin, B.E.: Ontology-based Web service to recommend spare time activities. *International Workshop on Information Heterogeneity and Fusion in Recommender Systems(HetRec)*. Barcelona, Spain (2010)
19. Minsukn, Kahng, Sagkeun, Lee, Sang-goo, Lee: Ranking in Context-Aware Recommender System. *School of Computer Science and Engineering, ACM, Hyderabad, India* (2011)

20. Kavitha Devi, M.K., Thirumalai, S.R., Vinoth, K.S., Venkatesh, P.: Probabilistic Neural Network approach to Alleviate sparsity and cold start problems in Collaborative Recommender Systems. In: Proceedings of International Conference on Computational Intelligence and Computing Research (ICIC), pp. 1–4 (2010)
21. Palmisano, C., Tuzhilin, A., Gorgoglione, M.: Using context to improve predictive modeling of customers in personalization applications. *IEEE Transactions on Knowledge and Data Engineering*, 20(11):1535–1549 (2008)
22. Ramirez-Garcia, X., Garcia-Valdez, M.: Restaurant Recommendations based on a Domain Model and Fuzzy Rules. In: Proceedings of International Seminary in Computer Intelligence (ISCI). Tijuana Institute of Technology. Tijuana B.C., Mexico (2012)
23. Ramzi, H.M., Baazaoui, H., Ziou, D., Ben Ghézala, H.: Towards a New Model for Context-Aware Recommendation. Department of Infoematic, University of Sherbrooke. Québec, Canada (2012)
24. Romadhony, A., Al Faraby, S., Pudjoatmodjo, B.: Online Shopping Recommender System Using Hybrid Method. In: Proceedings of International conference of Information and Communication Technology (ICoICT) (2013)
25. Wen, Wu, Liang, He, Jing, Yang.: Evaluating Recommender Systems. Department of Computer Science and Technology. East China Normal University (2012)
26. Yanlin, Zheng, Luyi, Li, Fanglin, Zheng: Context-awareness Support for Content Recommendation in e-learning Environments. In: Proceedings of International Conference on Information Management, Innovation Management and Industrial Engineering (2009)
27. Zied, Z., Godin, R., Faucher, L.: Evaluating Recommender Systems. In: Proceedings of International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution. IEEE (2008)
28. Zimmermann, M., Specht, L. A.: Personalization and context management. *User Modeling and User-Adapted Interaction*, vol. 15, no.3-4, pp. 275–302 (2005)

Sonificación de EEG para la clasificación de palabras no pronunciadas

Erick Fernando González-Castañeda, Alejandro Antonio Torres-García, Carlos Alberto Reyes-García, Luis Villaseñor-Pineda

Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE),
Tonantzintla, Puebla, México

{erick.gonzalezc,alejandro.torres,kargaxxi,villasen}@inaoep.mx

Resumen. Las interfaces cerebro-computadora (BCI) basadas en electroencefalograma (EEG) son una alternativa que pretende integrar a las personas con discapacidad motriz severa a su entorno. Sin embargo, éstas aún no son utilizadas en la vida cotidiana por lo poco intuitivas que son las fuentes electrofisiológicas para controlarlas. Para tratar este problema, se han realizado trabajos con el objetivo de clasificar las señales de EEG registradas durante el habla imaginada. En este trabajo se utilizó la técnica de sonificación de señales de EEG, la cual nos permite caracterizar la señal de EEG como una señal de audio. El objetivo es analizar si al aplicar el proceso de sonificación de la señal de EEG se puede discriminar o resaltar patrones que mejoren los resultados de clasificación de palabras no pronunciadas. Para ello se procesó la señal con y sin sonificación. Se obtuvieron los resultados de los 4 canales más cercanos a las áreas de lenguaje de Broca y Wernicke. Los porcentajes de exactitud promedio para las señales sin aplicar sonificación y aplicando sonificación son 48.1 % y 55.88 %, respectivamente, por lo que se pudo observar que el método empleado de sonificación de EEG mejora ligeramente los porcentajes de clasificación.

Palabras claves: Electroencefalogramas (EEG), interfaces cerebro-computadora (BCI), sonificación (sonification), habla imaginada (imagined speech/unspoken speech), random forest.

1. Introducción

En 2001 se presentó la “Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud”¹. En ésta se establece que las personas con discapacidad son aquellas que tienen una o más deficiencias físicas, mentales, intelectuales o sensoriales y que al interactuar con distintos ambientes del entorno social pueden impedir su participación plena y efectiva en igualdad de condiciones a

¹ Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud. Consultado el 10 de febrero de 2014. Disponible en http://www.conadis.salud.gob.mx/descargas/pdf/CIF_OMS_abreviada.pdf

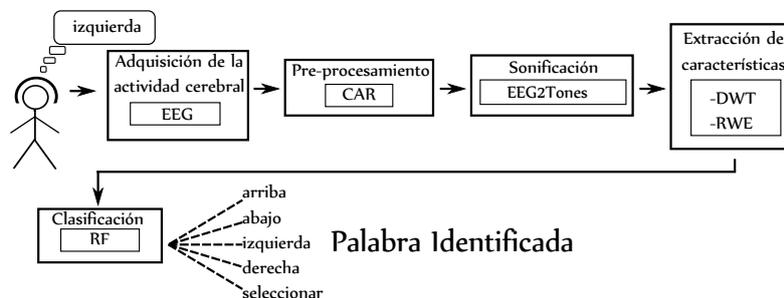


Fig. 1: Diagrama de la metodología seguida para clasificar señales de EEG registradas durante el habla imaginada. En los bloques internos se detalla el método propuesto. Además se ilustra a un individuo realizando la pronunciación imaginada de la palabra “izquierda” ante lo cual la tarea del método es inferir, con base a las señales de EEG registradas, cuál de las cinco palabras del vocabulario propuesto fue imaginada.

las demás. De acuerdo con el Informe mundial sobre la discapacidad 2011² de la Organización de las Naciones Unidas (ONU), en esta situación se estima que viven más de mil millones de personas en todo el mundo, esta cifra representa aproximadamente el 15% de la población mundial (según las estimaciones de la población mundial en 2010). Dentro de este porcentaje, se estima que casi 200 millones de personas experimentan dificultades considerables en su funcionamiento. Por ejemplo, discapacidades motrices severas como: la tetraplejía (cuadraplejía), la esclerosis lateral amiotrófica (ELA), la embolia (ictus cerebral), las lesiones de médula espinal o cerebral, la parálisis cerebral, las distrofias musculares, la esclerosis múltiple entre otros padecimientos. Estas discapacidades frecuentemente provocan que la persona no pueda controlar voluntariamente sus movimientos, incluyendo aquellos relacionados directa o indirectamente con la articulación del habla [6,28]. En consecuencia, una persona en estas condiciones está prácticamente aislada de su entorno.

En la búsqueda de un medio para integrar a la sociedad a las personas con discapacidad motriz severa, se ha explorado el uso de la actividad cerebral registrada mediante el electroencefalograma (EEG) para controlar dispositivos e interfaces. De forma general, una BCI puede ser vista como un sistema de reconocimiento de patrones donde el EEG es usado como la fuente primaria de información, un algoritmo de aprendizaje es usado para aprender una función de inferencia a partir del EEG, y por último, de acuerdo con la salida predicha por el algoritmo se ejecuta la acción deseada en el dispositivo a utilizar.

1.1. Sonificación

El concepto de sonificación (en inglés 'Sonification' o 'Auditory display'), se refiere al uso de sonido no hablado (non-speech audio) para transmitir informa-

² Informe mundial sobre la Discapacidad 2011. Consultado el 10 de febrero de 2014. Disponible en http://www.who.int/disabilities/world_report/2011/summary_es.pdf

ción [13]. La sonificación es la contraparte de la visualización, es un método que en lugar de asignar posiciones específicas a los valores abstractos de acuerdo a las reglas uniformes en un diagrama, asigna sonidos a ellos, también siguiendo normas uniformes. En la sonificación se realiza la transformación de relaciones de datos en relaciones acústicas con el propósito de facilitar la comunicación y la interpretación. Podemos plasmar en sonido imágenes 2D, datos sismológicos e incluso información de datos estadísticos.

La sonificación de EEG ha sido utilizada para hacer análisis exploratorio de las señales de EEG, [12], para hacer composiciones musicales [5] o para hacer diagnóstico temprano de enfermedades neurológicas como la enfermedad de Alzheimer mediante la retroalimentación audible [24,10]. Existen diversas técnicas para hacer sonificación de señales de EEG como: audificación, asignación por mapeo de parámetros, Sonificación basada en modelos [14], sonificación basada en modelado de bumps [21], entre otras.

2. Problemática

Una parte medular de la definición de BCI de Wolpaw [28] es cómo enviar los mensajes y comandos usando el EEG. La respuesta radica en los mecanismos neurológicos o procesos empleados por el usuario para generar las señales de control, denominadas fuentes electrofisiológicas. Las más utilizadas son los potenciales corticales lentos (SCP, por sus siglas en inglés), los potenciales P300, las imágenes motoras (ritmos sensoriales motrices mu y beta) y los potenciales evocados visuales (VEP, por sus siglas en inglés) [2,6,28].

Las fuentes electrofisiológicas descritas anteriormente (SCP, P300, imágenes motoras, y VEP) tiene dos grandes inconvenientes que repercuten en el posible uso de una BCI. El primero es el largo periodo de entrenamiento (algunas semanas o hasta meses) requerido para que un usuario pueda utilizar una BCI. Esto se debe que, estas fuentes son generadas por el usuario de forma poco consciente[17]. El segundo son las bajas tasas de comunicación (una sola palabra procesada, o menos, por minuto) que resultan insuficientes para permitir una interacción natural. Este último problema se debe a que cada una de estas fuentes requieren un “mapeo” o traducción al dominio del habla [6].

Recientemente, producto de los problemas descritos previamente, algunos trabajos tratan de utilizar los potenciales relacionados con la producción del habla, con diversos grados de éxito [6]. En estos trabajos, la fuente electrofisiológica es el habla imaginada (*imagined speech*), también referida como habla interna o habla no pronunciada (*unspoken speech*). De acuerdo con Wester [27], el término habla imaginada se refiere a la pronunciación interna, o imaginada, de palabras pero sin emitir sonidos ni articular gestos para ello. Es importante mencionar que, Denby [8] incluye a estos trabajos dentro de un área de investigación denominada interfaces de habla silente (SSI, por *Silent Speech Interfaces*) cuya finalidad es desarrollar sistemas capaces de permitir la comunicación “hablada” que tienen lugar cuando la emisión de una señal acústica entendible es imposible. Es importante remarcar que los trabajos que utilizan *habla imaginada* pueden

dividirse, por la unidad de habla utilizada, en dos enfoques: palabras y sílabas. El primer enfoque es seguido en [19,25,27,22,23,26]. Mientras que, en [4,7,9] únicamente se tratan sílabas. La presente investigación tiene como objetivo aplicar un método de sonificación de EEG que permita obtener una retroalimentación audible de la señal cerebral durante la pronunciación imaginada de palabras y que además permita resaltar patrones que ayuden a un clasificador automático a mejorar los porcentajes de exactitud reportados en el trabajo de [23].

3. Metodología

El presente trabajo está compuesto de las siguientes etapas: Adquisición de la actividad cerebral, Preprocesamiento, Sonificación, Extracción de características y Clasificación. La metodología seguida en el trabajo se muestra de mejor manera en la Figura 1. Es importante mencionar que también se procesarán las señales de EEG sin la etapa de sonificación, de tal manera que pueda existir un marco de comparación con el trabajo descrito en [23].

3.1. Adquisición de la actividad cerebral

En esta etapa se utiliza el conjunto de datos de EEG registrados durante el habla imaginada utilizados en [23]. Este conjunto de datos se compone de las señales de EEG de 27 individuos cuyo idioma nativo es el español. Las señales de EEG se registraron utilizando el kit EPOC de EMOTIV[®]. Este kit es inalámbrico y consta de 14 canales (electrodos) de alta resolución (más las referencias CMS/DRL en las posiciones P3/P4, respectivamente) cuya frecuencia de muestreo es de 128 Hz. Los nombres de los canales, de acuerdo con el sistema internacional 10-20, son: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 (ver Figura 2).

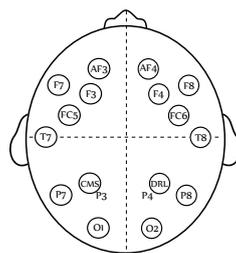


Fig. 2: Localización de los electrodos en el kit EPOC de Emotiv

Este conjunto de datos se registró por medio de un protocolo básico para adquirir las señales de EEG de cada individuo. El protocolo consistió en colocar a la persona cómodamente sentada con los ojos abiertos cerca de un escritorio, y con la mano derecha sobre el mouse de una computadora. Con un clic al mouse,

el usuario delimitaba tanto el inicio como el fin de la pronunciación imaginada de alguna de las palabras del vocabulario reducido, compuesto de las siguientes cinco palabras en español: “arriba”, “abajo”, “izquierda”, “derecha”, y “seleccionar” (ver Figura 3). El objetivo detrás de este protocolo de adquisición es saber a priori en qué parte de la señal de EEG se deben buscar los patrones asociados con la pronunciación imaginada de la palabra indicada.

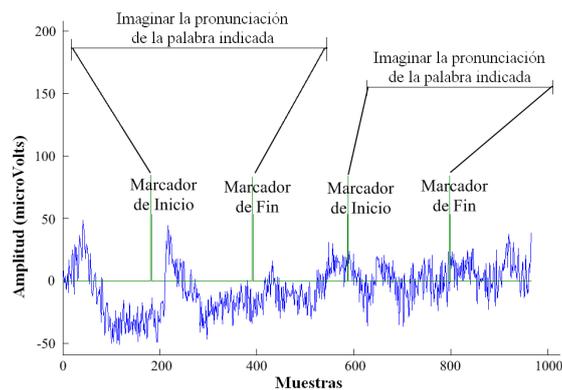


Fig. 3: Señal de EEG del canal F7 del individuo S1 mientras imagina la dicción de la palabra “Abajo” siguiendo el protocolo de adquisición de datos

Los segmentos de interés de las señales de EEG son aquellos que se encuentran entre los marcadores de inicio y fin, estos segmentos son denominados épocas (ventanas). Cada época tiene una duración variable como el habla normal (pronunciada). Además, la pronunciación imaginada de cada una de las cinco palabras fue repetida 33 veces consecutivas durante el registro del EEG, es decir, cinco bloques de 33 repeticiones por palabra. Antes de cada bloque se le indicó al individuo cuál es la palabra que debía pronunciar internamente. Todos las épocas de un mismo individuo fueron registradas en una única sesión (mismo día). Además, todas las sesiones se registraron en un laboratorio alejado de ruido audible externo y de ruido visual.

Es importante mencionar que al inicio del registro de las señales de EEG, se le indicó a la persona que evitara parpadear o realizar movimientos corporales mientras imaginaba la pronunciación de la palabra indicada, ya que después de cada marcador de fin podía tomarse un tiempo de descanso para dichos movimientos. Además, con la finalidad de que el individuo no supiera cuántas veces se repetiría una palabra, en la sala de experimentos, otra persona alejada del campo visual y guardando el debido silencio, se encargaba de realizar el conteo de repeticiones e indicaba cuando el individuo debía concluir. Esto con la finalidad de que el individuo no se distrajera contando el número de repeticiones ni se predispusiera a saber que le falta poco o mucho para concluir el experimento.

3.2. Preprocesamiento

Las señales de EEG obtenidas son preprocesadas con el método de referencia promedio común (CAR, por sus siglas en inglés). Este método tiene como fin mejorar la relación señal a ruido de la señal de EEG. Básicamente, se busca quitar todo aquello que es común en todas las lecturas simultáneas de los electrodos. La CAR puede ser calculada mediante la resta del potencial entre cada electrodo y la referencia (el potencial promedio de todos los canales), se repite esto para cada instante de tiempo en la frecuencia de muestreo.

3.3. Sonificación de la señal EEG

Algoritmo 1 Sonificación de EEG

Requerir: *EEG* (señal EEG), *NT* (Número de tonos), *LowF* (frec. mínima de la señal EEG), *HiF* (frec. máxima de la señal EEG), *W* (tamaño de la ventana del espectrograma), *Shf* (Traslape entre ventanas del espectrograma), *Dur* (Duración de los tonos), *LowFAu* (frec. mínima del audio), *HiFAu* (frec. máxima del audio), *Fs* (frec. de muestreo del audio)

Asegurar: *Audio* (Sonificación de la señal EEG)

Spec \leftarrow Construir Espectrograma de la señal EEG, usando *EEG, LowF, HiF, W, Shf*)

Escalar amplitudes de frecuencias de *Spec* dividiendo entre la máxima amplitud.

for *i* \leftarrow 1 hasta tamaño horizontal de *Spec* **do**

 Ordenar descendentemente *Spec.columna_i* de acuerdo con las amplitudes.

Winners.columna_i \leftarrow tomar de *Spec.columna_i* las primeras *NT* frecuencias y sus amplitudes.

end for

TonesF \leftarrow Escalar las frecuencias unicas de *Winners* al rango audible usando *LowFAu* y *HiFAu*.

Tones \leftarrow Crear la onda sinusoidal para cada frecuencia en *TonesF*, usando *Fs* y *Dur*.

for *j* \leftarrow 1 hasta tamaño horizontal de *Spec* **do**

Audio \leftarrow Unir el valor anterior de *Audio* con la Suma de las señales en tonos que corresponden a las frecuencias ganadoras en *Winners.columna_i* de acuerdo con *Tones* y *TonesF*.

end for

return *Audio*

En el presente trabajo se utilizó la técnica de Sonificación 'EEG to tones' [1]. Esta técnica es un proceso de audificación, en la cual se escalan las frecuencias de EEG a frecuencias audibles. Esta técnica se basa en el espectrograma de la señal EEG de entrada, el cual es calculado usando la transformada rápida de Fourier (*FFT*). De cada ventana del espectrograma se calculan las frecuencias dominantes de la señal de EEG. Las frecuencias dominantes son escaladas a tonos, los cuales representan una frecuencia en el rango audible. Al final cada

conjunto de tonos dominante por ventana es unido para formar el audio de salida. El pseudocódigo que explica con más detalle el método utilizado es mostrado en el algoritmo 1.

3.4. Extracción de características

Transformada wavelet discreta. En [15] se menciona que las características utilizadas en las BCI son no estacionarias ya que las señales de EEG pueden rápidamente variar con el tiempo. Además, estas características deben contener información del tiempo debido a que los patrones de actividad cerebral están generalmente relacionados a variaciones específicas del EEG en el tiempo. Lo anterior, hace necesaria una representación que considere eso.

Una técnica que permite modelar dichas variaciones, en el dominio tiempo-escala, es la transformada wavelet discreta (DWT, por su siglas en inglés). La DWT provee una representación wavelet altamente eficiente mediante la restricción de la variación en la traslación y la escala, usualmente a potencias de dos. En ese caso, la DWT es algunas veces llamada transformada wavelet diádica.

El análisis DWT puede ser realizado usando un algoritmo piramidal rápido descrito en términos de bancos de filtros multi-tasa, es decir, aquellos donde se tiene más de una tasa de muestreo realizando conversiones mediante las operaciones de decimación e interpolación. La DWT puede ser vista como un banco de filtros con espacio de una octava entre ellos. Cada sub-banda contiene la mitad de las muestras de la frecuencia de la sub-banda vecina más alta. En el algoritmo piramidal la señal es analizada en diferentes bandas de frecuencias con diferentes resoluciones mediante la descomposición de la señal en una aproximación burda (coeficientes de aproximación) e información detallada (coeficientes de detalle). La aproximación burda es entonces adicionalmente descompuesta usando el mismo paso de descomposición wavelet. Esto se logra mediante un filtrado sucesivo de pasa-bajas y pasa-altas de la señal de tiempo, y un sub-muestreo. En [18] se muestra la información detallada de lo antes mencionado.

En el presente trabajo se aplica la transformada wavelet discreta a los archivos de audio resultantes de la sonificación. Se calcularon 6 niveles de descomposición usando la wavelet madre Daubechies de orden 20 (db20). Asimismo, para el caso de las señales de EEG que no fueron sonificadas, la DWT se calculó con 5 niveles de descomposición utilizando la Daubechies de segundo orden (db2), como se describe en [23].

Como es evidente, el número de coeficientes wavelet en cada uno de los niveles variará dependiendo del tamaño de la señal de EEG delimitada entre los marcadores. Esto debido a que, de manera similar al habla convencional, la duración de las ventanas de pronunciación imaginada de una palabra es variable tanto en ventanas de un sólo individuo como en ventanas de individuos distintos. Para tratar con este problema, los coeficientes wavelets son normalizados mediante la energía relativa wavelet que se describe a continuación.

Energía wavelet relativa. Tal como se explica en [23], una vez aplicada la DWT sobre la señal se obtienen coeficientes de aproximación y de detalle, desde los cuales es posible calcular la energía relativa wavelet. La energía relativa wavelet representa la energía que algún nivel de descomposición aporta al total de la energía wavelet de la señal. La energía relativa provee información para caracterizar la distribución de energía de la señal en diferentes bandas de frecuencia, con lo que se obtiene una independencia del tamaño de la ventana de señal de EEG o de audio, según sea el caso.

A partir de la descripción anterior, en las señales EEG sonificadas se determinó usar 10 valores que representan la energía wavelet en todos los niveles de descomposición y el último de aproximación (D1-D9 y A9). Mientras que, cada ventana de habla imaginada de las señales de EEG no sonificadas se representa mediante un conjunto de 5 valores de energía wavelet, 4 de los niveles de descomposición y uno de aproximación (D2-D5 y A5) con respecto a la energía wavelet total. Tal como se realizó en [23], el valor asociado con D1 es descartado.

3.5. Clasificación

De acuerdo con Michie *et al.* [16], la clasificación cubre cualquier contexto en el que alguna decisión o pronóstico es hecho sobre la base de información histórica disponible. La base de información disponible está dada por vectores de información multi-dimensionales, cuyos valores (llamados atributos o características) pueden ser reales, discretos o nominales. El objetivo de la clasificación es inferir una relación entre un vector de datos y una posible clase (o categoría), para ello se crea un modelo que automáticamente encuentra dichas relaciones. El modelo es creado basado en una partición de entrenamiento del vector de datos. Los modelos aprendidos de los datos de entrenamiento son, entonces, evaluados con un conjunto de prueba distinto para determinar si los modelos pueden ser generalizados a nuevos casos. En el presente trabajo se entrena y prueba el clasificador Random Forest (RF) bajo un enfoque de validación cruzada usando 10 pliegues (folds).

Random forest (RF). RF es una combinación de árboles predictores tal que cada uno de los árboles depende de los valores de un vector aleatorio muestreado independientemente y con la misma distribución para todo los árboles en el bosque. Cada árbol arroja un único voto para la clase más popular para una entrada x dada, y al final la salida de RF se realiza usando voto mayoritario. En [3,20] se describe a detalle el algoritmo de clasificación de RF incluyendo el proceso para construir los arboles individuales.

En el presente trabajo se utilizaron los siguientes hiper-parámetros para la implementación del clasificador en Weka 3.6.8: el número de árboles es 50 y el número de atributos considerados en cada nodo es $\log_2(\text{numeroCaracteristicas}) + 1$.

4. Experimentación y resultados

A pesar de que el EPOC de Emotiv ofrece la posibilidad de registrar 14 canales, sólo serán de interés para los experimentos los canales F7, FC5, T7 y P7. Estos canales, de acuerdo con el modelo Geschwind-Wernicke, son los más relacionadas con la producción del habla en el hemisferio izquierdo del cerebro (a excepción de algunas personas zurdas) [11].

4.1. Selección de parámetros de la sonificación

Dado que existen diversos parámetros en el proceso de sonificación, se tenía que elegir valores que favorecieran los porcentajes de exactitud de la clasificación, por ello para la selección de parámetros se realizó un proceso empírico de elección iterativo para cada parámetro. Es decir sólo variando un parámetro se evaluó su comportamiento de acuerdo a la exactitud de clasificación y se eligió el mejor valor, después usando ese valor se varió el siguiente parámetro y se eligió también su mejor valor, así hasta tener los mejores valores de todos los parámetros. Los valores seleccionados de los parámetros fueron: número de tonos (14), frecuencia mínima de la señal EEG (1 Hz), frecuencia máxima de la señal EEG (60 Hz), tamaño de la ventana del espectrograma (26 muestras), traslape entre ventanas del espectrograma (1 muestra), duración de los tonos (0.6 secs), frecuencia mínima del audio (50 Hz), frecuencia máxima del audio (5000 Hz), frecuencia de muestreo del audio (8000 Hz). En la figura 4 se muestran ejemplos de espectrogramas obtenidos usando los parámetros mencionados anteriormente.

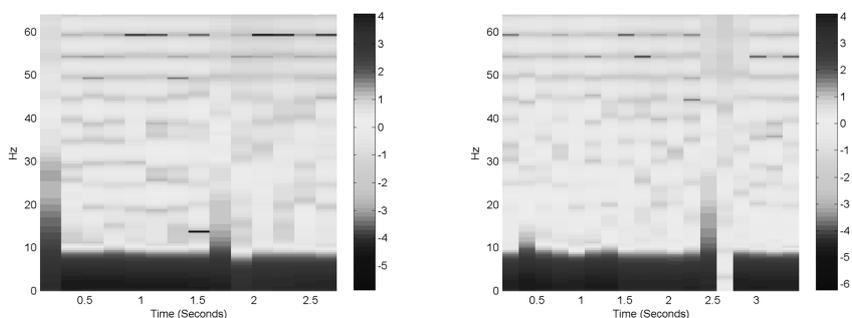


Fig. 4: Ejemplos de espectrogramas del sujeto 11, durante la pronunciación imaginada de dos distintas repeticiones de la palabra *arriba*, en el canal F7.

4.2. Selección de wavelet

En el proceso de extracción de características se siguió el mismo enfoque que en la selección de parámetros para la sonificación, se experimentó con distintos niveles de diferentes Wavelets Daubechies (db2, db6 y db20) probando una variación a la vez. La wavelet Daubechis 20 con 6 niveles fue la que mejores resultados obtuvo.

4.3. Resultados comparativos

Los porcentajes de exactitud promedio al clasificar usando Random Forest para los 27 sujetos en los dos enfoques son mostrados en la figura 5.

Al analizar la tabla podemos observar que el método que sonifica la señal de EEG mejora los porcentajes de exactitud promedio en 24 de los 27 sujetos analizados, inclusive existen diferencias arriba de 15% en algunos casos. De manera general se puede concluir que el método para sonificar la señal de EEG usando el algoritmo *EEG to tones* permite mejorar en promedio 7.72% al conjunto de los 27 sujetos.

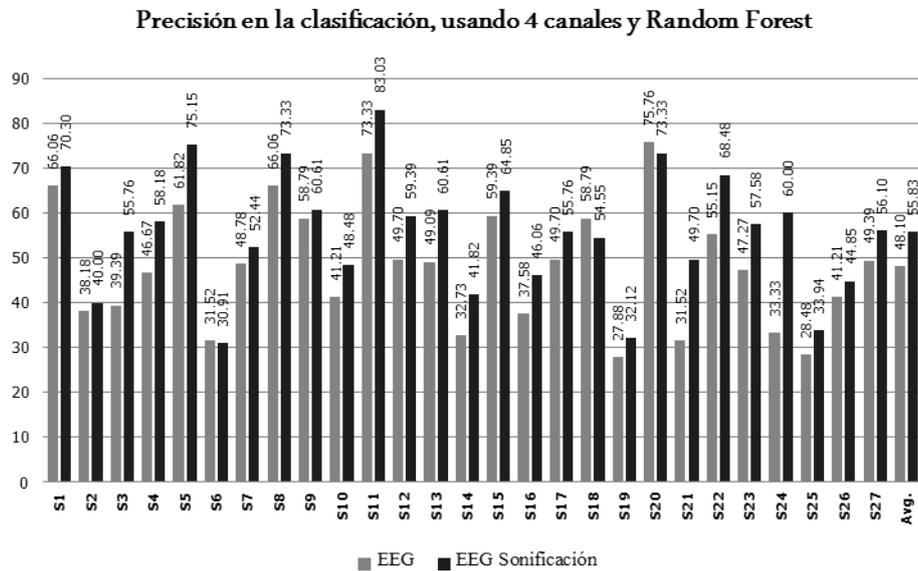


Fig. 5: Gráfica comparativa de los dos métodos EEG y Sonificación de EEG. Se muestran los porcentajes de exactitud promedio en la clasificación con Random Forest para los 27 sujetos usando 4 canales. En la última columna se muestran los porcentajes de clasificación promedio: EEG 48.10% y EEG usando Sonificación 55.83%.

5. Conclusiones

En este trabajo se realizó la clasificación de la señal de EEG aplicando una transformación de la señal original de EEG a una señal de audio, con lo cual se mejoró la exactitud promedio de clasificación en 7.72%. Mediante la elección de las frecuencias dominantes del espectrograma de la señal de EEG y el mapeo de las frecuencias de EEG a frecuencias del audio, se logró resaltar patrones de la señal que ayudaron a mejorar la exactitud de la clasificación, esto ocurrió incluso

usando la misma metodología de extracción de características y el mismo método de clasificación, con respecto con el cual se compara. Los resultados obtenidos dan pie a realizar pruebas con otros métodos de sonificación, extracción de características y clasificación, que permitan mejorar los porcentajes de exactitud alcanzados. Un aspecto a considerar, en el trabajo por realizar, es la elección de parámetros para cualquier forma de sonificación, la cual podría ser realizada por un algoritmo de búsqueda automática, como por ejemplo un algoritmo genético. También se tendrá que calcular y comparar el comportamiento de los dos enfoques usando los 14 canales disponibles.

Referencias

1. Andersonl, C.: Eeg to tones, dept. of commputer science, colorado state university (2005)
2. Bashashati, A., Fatourehchi, M., Ward, R., Birch, G.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering* 4, R32–R57 (2007)
3. Breiman, L.: Random forests. *Machine learning* 45(1), 5–32 (2001)
4. Brigham, K., Kumar, B.: 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*. pp. 1–4. IEEE (2010)
5. Brouse, A., Filatriau, J.J., Gaitanis, K., Lehembre, R., Macq, B., Miranda, E., Zénon, A.: An instrument of sound and visual creation driven by biological signals. *Proceedings of ENTERFACE06, Dubrovnik (Croatia)*. (Not peer-reviewed report.) (2006)
6. Brumberg, J.S., Nieto-Castanon, A., Kennedy, P.R., Guenther, F.H.: Brain-computer interfaces for speech communication. *Speech Communication* (52), 367–379 (2010)
7. DaSalla, C.S., Kambara, H., Koike, Y., Sato, M.: Spatial filtering and single-trial classification of EEG during vowel speech imagery. In: *i-CREAtE '09: Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*. pp. 1–4. ACM, New York, NY, USA (2009)
8. Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J., Brumberg, J.: Silent speech interfaces. *Speech Communication* 52(4), 270–287 (2010)
9. D'Zmura, M., Deng, S., Lappas, T., Thorpe, S., Srinivasan, R.: Toward EEG sensing of imagined speech. *Human-Computer Interaction. New Trends* pp. 40–48 (2009)
10. Elgendi, M., Rebsamen, B., Cichocki, A., Vialatte, F., Dauwels, J.: Real-time wireless sonification of brain signals. In: *Advances in Cognitive Neurodynamics (III)*, pp. 175–181. Springer (2013)
11. Geschwind, N.: Language and the brain. *Scientific American* (1972)
12. Hermann, T., Meinicke, P., Bekel, H., Ritter, H., Müller, H.M., Weiss, S.: Sonification for eeg data analysis. In: *Proceedings of the 2002 International Conference on Auditory Display* (2002)
13. Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, J., Bargar, R., Barrass, S., Berger, J., et al.: The sonification report: Status of the field and research agenda. report prepared for the national science foundation

- by members of the international community for auditory display. International Community for Auditory Display (ICAD), Santa Fe, NM (1999)
14. Kramer, G.: Auditory display: Sonification, audification, and auditory interfaces. Addison-Wesley Reading, MA (1994)
 15. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnald, B.: A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering* 4, r1–r13 (2007)
 16. Michie, D., Spiegelhalter, D., Taylor, C.: *Machine Learning, Neural and Statistical Classification*. Overseas Press (2009)
 17. Pfurtscheller, G.: Brain-computer interfaces: State of the art and future prospects. In: *Proceedings of the 12th European Signal Processing Conference: EUROSIPCO 04*. pp. 509–510 (2004)
 18. Pinsky, M.: *Introduction to Fourier analysis and wavelets*, vol. 102. Amer Mathematical Society (2002)
 19. Porbadnigk, A., Schultz, T.: EEG-based Speech Recognition: Impact of Experimental Design on Performance. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany (2008)
 20. Rokach, L.: *Pattern Classification Using Ensemble Methods*. World Scientific (2009)
 21. Rutkowski, T.M., Vialatte, F., Cichocki, A., Mandic, D.P., Barros, A.K.: Auditory feedback for brain computer interface management—an eeg data sonification approach. In: *Knowledge-Based Intelligent Information and Engineering Systems*. pp. 1232–1239. Springer (2006)
 22. Torres-García, A.A., Reyes-García, C.A., Villaseñor-Pineda, L.: Toward a silent speech interface based on unspoken speech. In: Huffel, S.V., Correia, C.M.B.A., Fred, A.L.N., Gamboa, H. (eds.) *BIOSTEC - BIOSIGNALS*. pp. 370–373. SciTePress (2012)
 23. 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* 34(1), 23–39 (2013)
 24. Vialatte, F., Cichocki, A., Dreyfus, G., Musha, T., Rutkowski, T.M., Gervais, R.: Blind source separation and sparse bump modelling of time frequency representation of eeg signals: New tools for early detection of alzheimer's disease. In: *Machine Learning for Signal Processing, 2005 IEEE Workshop on*. pp. 27–32. IEEE (2005)
 25. Wand, M.: *Wavelet-based Preprocessing of Electroencephalographic and Electromyographic Signals for Speech Recognition*. Studienarbeit Lehrstuhl Prof. Waibel Interactive Systems Laboratories Carnegie Mellon University, Pittsburgh, PA, USA and Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany (2007)
 26. Wang, L., Zhang, X., Zhong, X., Zhang, Y.: Analysis and classification of speech imagery eeg for bci. *Biomedical Signal Processing and Control* 8(6), 901–908 (2013)
 27. Wester, M., Schultz, T.: *Unspoken Speech - Speech Recognition Based On Electroencephalography*. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany (2006)
 28. Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., Vaughan, T.: Brain-computer interfaces for communication and control. *Clinical neurophysiology* 113(6), 767–791 (2002)

Software decodificador de texto a voz para ayuda a personas con capacidades diferentes

Luis Javier Mona Peña, Joaquin Antonio Alvarado Bustos,
Octavio Mendez Hernandez

Instituto Tecnológico de Saltillo, Departamento de Ingeniería Eléctrica-Electrónica,
Saltillo, Coahuila, México

lmona@isicnc.net,jalvarado@its.mx,omenh@hotmail.com

Resumen. El desarrollo de la tecnología históricamente ha ido de la mano de la resolución de problemas o necesidades que contribuyen al bienestar de la Humanidad. Cuando la tecnología se aplica al mejoramiento de la calidad de vida de personas con alguna discapacidad física y/o mental es mucho más evidente el propósito de la misma. El presente artículo se enfoca al uso del software LabVIEW (*Laboratory Virtual Instrumentation Engineering Workbench*) en el desarrollo de una herramienta capaz de auxiliar a personas con capacidades diferentes, en específico débiles visuales y/o mudos, aclarando que también es útil al resto de la población. Dicha herramienta es un convertidor de Texto a Voz (TTS) con la capacidad de leer archivos de texto en código ASCII (.txt), imágenes con extensión .jpg y texto introducido directamente desde la misma interfaz gráfica de la aplicación. Además es capaz de convertir el texto interpretado en documentos codificados en Braille y desplegarlos en un archivo del procesador de texto más común que es Microsoft Word. Como punto final, la aplicación es capaz de guardar la lectura computarizada en archivos de audio tipo WAV (*WAVE form audio file format*) y así crear audiolibros de las fuentes de texto ya descritas.

Palabras clave: ASCII, Braille, JPG, LabVIEW, TTS, WAV.

1. Introducción

La poca disponibilidad y accesibilidad de fuentes bibliográficas en Código Braille para personas con debilidad visual y el poco entendimiento de la mayor parte de la población del lenguaje a señas en el caso de personas con dificultad en el lenguaje hablado, merman su calidad de vida, y trae como consecuencias:

- Falta de interés en conocer la actualidad social.
- Problemas de interacción social.
- Rechazo desde/hacia el entorno.
- Sentimiento de frustración, etc.
- Poco o nulo desarrollo educativo profesional, etc.

Existen diferentes opciones tecnológicas para apoyar a las personas con discapacidad tales como los convertidores de Texto a Voz (TTS) los cuales son una

herramienta esencial en el apoyo de personas con capacidades diferentes. Existen varios tipos de los mismos, algunos de ellos requieren equipo especial tal como escáneres de alta velocidad, cámaras de alta resolución, teclados y sistemas de cómputo dedicados, los mismos que pueden llegar a tener costos prohibitivos para la mayoría de la población. Además de que su arquitectura es cerrada siendo muy complicado o imposible que pueda adaptarse a cambios personalizados requeridos por el usuario.

El reto principal era desarrollar una herramienta por software que no dependiera de hardware y software especial, es decir que pueda funcionar en cualquier computadora y utilizar cualquier cámara o escáner para la adquisición de la imagen, aunado a que fuera flexible al momento de poder escoger entre diferentes idiomas y voces para la lectura de los datos, además de convertir los documentos en texto convencional ASCII a documentos en Código Braille, dejando dichos documentos listos en un archivo en Microsoft Word para imprimirse y hacerlos disponibles para las personas que requieren leer en dicho código.

Para mayor comodidad del usuario, la aplicación es capaz de guardar los textos leídos en archivos de audio, en específico en formato WAV. Esto abre la cantidad de literatura disponible en formato Braille además de crear una audioteca digital.

2. Metodología

La implementación de la aplicación se puede realizar por medio de diversas herramientas, tales como Sistemas de visión con el uso de algoritmos de extracción morfológica, aprendizaje de patrones y su posterior entrenamiento, Algoritmos matemáticos para Reconocimiento Óptico de Caracteres (OCR) y por medio de la utilización de Referencias ActiveX.

En este caso se optó por el uso de la tecnología ActiveX con el propósito de reutilizar código ya compilado tanto del sistema operativo Windows (XP, Vista, 7, 8) y de Microsoft Office (2003, 2007, 2010), en específico el **Microsoft Office Document Imaging**.

El sistema operativo Windows trae la opción de Texto a Voz para lo cual utiliza la Referencia ActiveX **Microsoft Speech SDK 5.1**, para la obtención de la lectura computarizada de texto plano, la cual, sin embargo es incapaz de obtener texto desde imágenes. Microsoft Office por medio de la herramienta del Módulo Office Document Imaging **MODI.IDocument** permite la extracción de texto de imágenes para su posterior uso o edición. La aplicación por lo tanto hizo uso de dichas referencias ActiveX (**Microsoft Speech SDK 5.1** y **MODI.IDocument**) además del generador del módulo de Microsoft Office para LabVIEW 2013 para la integración de la misma.

3. Implementación de la aplicación

La Secuencia de operación en programación, involucra el uso de lo que el sistema operativo tiene instalado para herramientas TTS, es decir, voces computarizadas y

dispositivos de audio principalmente. Por lo que lo primero es llamar a la referencia Microsoft Speech SDK 5.1 para saber cuáles voces y que idiomas hay disponibles en el Sistema Operativo. Dicha información se muestra en la interfaz de usuario en forma de lista.

Cabe aclarar que existen diferentes ejemplos en la web para el uso del ActiveX de Microsoft Speech SDK lo que facilita la implementación de la aplicación. La figura 1 muestra el código necesario para dicha acción.

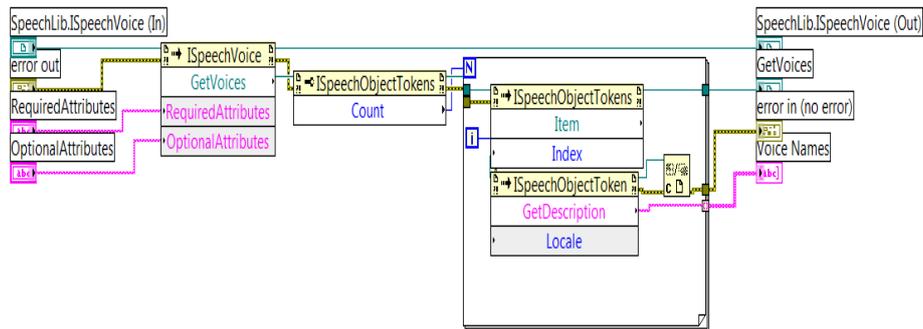


Fig. 1. Obtención de las voces del sistema.

Una vez inicializadas las voces del sistema se procede a obtener información del tipo de dispositivos de audio que están disponibles para escuchar la lectura. Esto se realiza haciendo la requisición directamente al sistema operativo por medio de la librería ya mencionada. El código de la figura 2 muestra el procedimiento realizado.

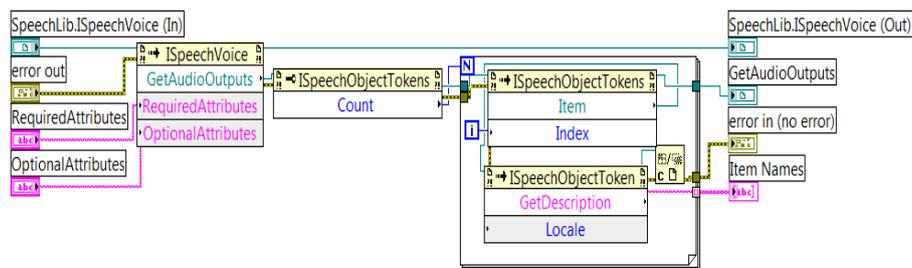


Fig. 2. Obtención de dispositivos de audio disponibles.

Una vez que se han determinado las voces y los dispositivos de audio con que cuenta la PC, se procede a la inicialización de los parámetros de velocidad, volumen y fuente de texto de la lectura. Esto se realiza con una estructura tipo case donde dependiendo de las acciones del usuario será el código ejecutado. Hay tres posibles opciones:

1. En el caso de que la fuente del texto sea un archivo .txt la interfaz despliega los elementos necesarios para que el usuario busque el archivo dentro de la estructura

del sistema operativo. Es importante señalar que esta versión del software solo acepta archivos de texto plano (.txt), por lo que archivos con otras extensiones (.pdf, Word, etc.) deben convertirse previamente a texto plano.

2. Si la opción seleccionada es por texto directamente en la interfaz, entonces el usuario tiene una caja de texto donde puede teclear todo el texto que requiera ser leído. No existe restricción en cuanto a la cantidad de caracteres.

3. La tercera opción es Texto a partir de una imagen. Para esta opción se utilizó la referencia ActiveX MODI.IDocument, la cual permite obtener diversa información sobre una imagen incluyendo el texto que está en ella, por lo que ya no es necesario realizar algoritmos OCR sino que solo utilizar las propiedades y métodos de dicho ActiveX. La figura 3 muestra parte del código necesario para obtener el texto en forma de oración.

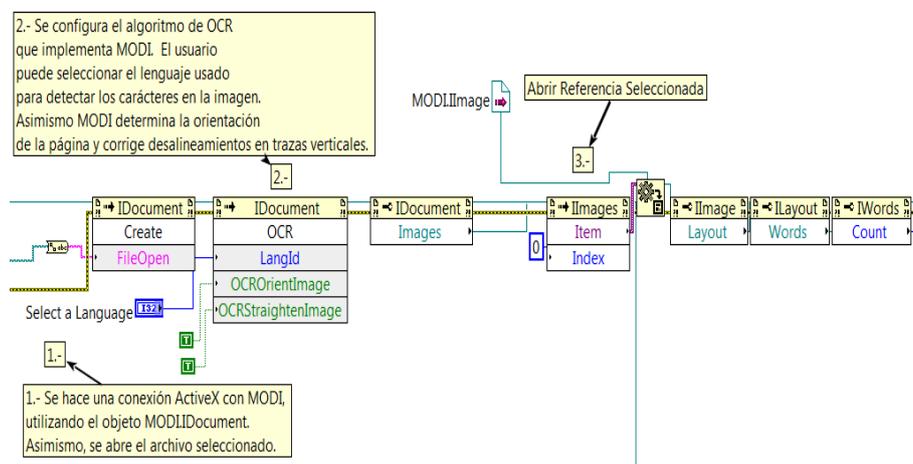


Fig. 3. Obtención de caracteres a partir de imagen.

Una vez inicializado el sistema se procede a realizar la acción de lectura. Dicha acción ejecuta código del ActiveX SpeechLib.ISpeechVoice el cual es parte del Microsoft Speech SDK. A dicho ActiveX es necesario pasar los datos previos de Voz, Volumen, Velocidad y Dispositivo de audio, dicho código se muestra en la figura 4.

El usuario escuchará en la lectura de acuerdo a la inicialización que haya configurado. Es importante recalcar que existen voces para lectura en español, inglés, ruso, etc. Estas voces son independientes de la aplicación y se instalan al sistema operativo para hacerlas disponibles. La conversión a Código Braille se llevó a cabo utilizando el Toolkit de Microsoft Office para LabVIEW el cual permite dar formato a un documento en Word, esto incluye la capacidad de cambiar el tipo de fuente que es la clave para dicha conversión. Únicamente se cambia el tipo de fuente a una tipo Braille, en este caso a la fuente tipo “XBraille”, con lo cual se obtiene cualquier documento codificado en Braille. El código que realiza dicha conversión se muestra en la figura 5.

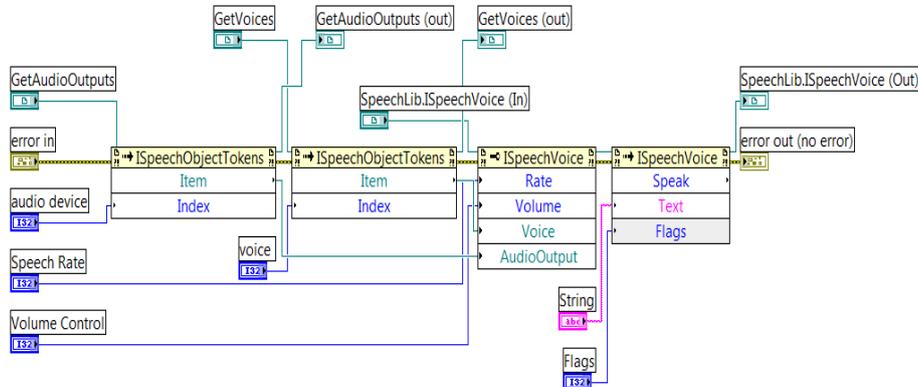


Fig. 4. Código para la acción de lectura.

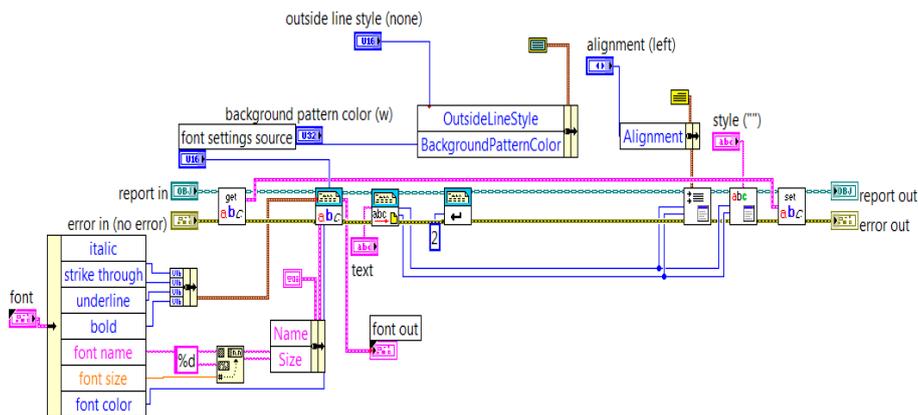


Fig. 5. Conversión a Braille.

Para la generación de los archivos de audio .wav se utilizaron las herramientas de LabVIEW de la paleta de Graphics and Sound y solo requiere una reconfiguración a la tarjeta de sonido para que permita grabar el audio que sale por las bocinas, muy parecido a utilizar el grabador de sonido de Windows.

Para el almacenamiento de los archivos en el disco duro, la aplicación crea una carpeta específica dentro del directorio raíz llamada Audio Texto a Voz (C:\Audio Texto a Voz\), donde se guardan los archivos de manera secuencial tomando el nombre del archivo de texto o de imágenes original. La aplicación verifica que no se sobrescriba el archivo y genera un número consecutivo que le anexa al nombre base en caso de que ya exista un archivo con el mismo nombre base, por lo que permite crear archivos secuenciales por ejemplo de los capítulos o temas de libros.

La interfaz de la aplicación con la que el usuario tiene interacción es parte fundamental ya que tiene las medidas necesarias para su rápida inicialización además

de impedir que se puedan cometer errores al no seguir el orden antes descrito. Dicha interfaz se muestra en la figura 6.

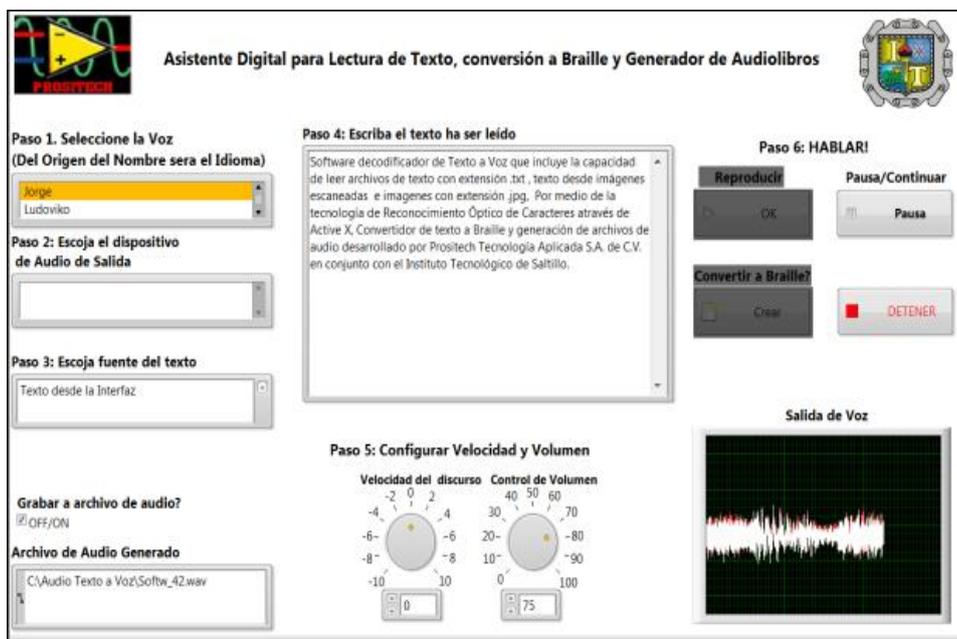


Fig. 6. Interfaz de Usuario

4. Resultados

El resultado es una herramienta sumamente efectiva para leer documentos por medio de software, ya sea escaneados (imágenes) o directamente en texto, lo que permite que cualquier documento se convierta en un audiolibro con las ventajas que eso representa, en especial para las personas con debilidad visual, además es una aplicación poderosa para las personas con algún problema de habla ya que les permite comunicarse escribiendo en el teclado lo que quieren decir y la aplicación hablarlo por ellos, lo que permite que puedan comunicarse con personas que no entiendan lenguaje de señas que son la mayoría. La conversión a Código Braille amplía el acervo de documentos disponibles para personas con debilidad visual siendo prácticamente ilimitada la cantidad de información que tendrían disponible en dicho Código.

Esta herramienta informática ha sido probada con una muestra de diez personas con discapacidad visual y tres con discapacidad de lenguaje hablado. En todos los casos las personas han experimentado una mejora en su calidad de vida debido a que han podido comunicarse de una manera asertiva, a la vez de que han tenido acceso a grandes cantidades de información lo que les ha permitido una mayor integración con el resto de la población mejorando su calidad de vida. Asimismo, el software se

pondrá al alcance de La Asociación Mexicana de Enfermedades Visuales y Estudios de la Retina (**AMEVER**) en la ciudad de Saltillo, Coahuila para su uso generalizado y poder obtener información cualificada y cuantificada del impacto de esta herramienta en dicha población.

5. Conclusiones

La implementación del desarrollo de Software para la solución de problemas específicos; en este caso, para ayuda a personas con problemas de visión o de habla ha resultado en una herramienta sumamente útil que colabora con el aumento en su calidad de vida.

Se desarrolló una herramienta de bajo costo que no requiere hardware especial para su funcionamiento, ya que funciona en cualquier PC bajo ambiente Windows y los textos pueden ser leídos desde diferentes fuentes incluyendo aquellos provenientes de cualquier escáner comercial. El impacto social que puede llegar a tener es muy alto ya que permite una comunicación más efectiva además de mayor interacción en el entorno de los usuarios.

Referencias

1. National Instruments, "LabVIEW 2012 User Manual", February 2012 Edition.
2. National Instruments, "LabVIEW 2012 Connectivity User Manual", February 2012 Edition.
3. Mona Peña, Luis Javier, "Introducción a la Instrumentación Virtual", Manual de Curso, Instituto Tecnológico de Saltillo (2009)
4. Microsoft Corporation, "Microsoft Windows Software Development Kit (SDK)", Technical Notes.
5. Microsoft Corporation, "Microsoft Windows Microsoft Office Document Imaging (MODI) Technical Notes".

Is there Hope for Interlingua methods? A CLIR Comparison Experiment between Interlingua and Query Translation

Marta R. Costa-jussà¹, Rafael E. Banchs²

¹ Instituto Politécnico Nacional, Mexico

² Institute for Infocomm Research, Singapore

marta@nlp.cic.ipn.mx, rembanchs@i2r.a-star.edu.sg

Abstract. A comparison of interlingua and query translation is proposed in a particular cross-language information retrieval (CLIR) application which consists on retrieving a book from the collection by using one of its chapters in a different language as a query. The experiments are performed in three languages (English, Chinese and Spanish) and all the possible combinations. It is shown that interlingua is able to outperform the query translation approach in some cross-language tasks. Results are further analysed and it is found that, for this particular task, the quality of translation (in terms of BLEU and PER) is not directly correlated with the query translation performance.

Keywords: Interlingua, machine translation, information retrieval.

1 Introduction

Cross-language information retrieval (CLIR) allows users for accessing to documents or information in a different language from their queries. CLIR is becoming more popular as the availability of information in languages different from English increases in the Internet [12]. This paper complements our previous works [3, 4] and compares the performance of different CLIR methodologies.

Research in CLIR has been significantly encouraged by three well-known evaluation campaigns: a cross-language information retrieval track at TREC ¹, the Cross-Language Evaluation Forum (CLEF) ² and the NTCIR ³ Asian Language Evaluation. There are CLIR applications available such as the cross-language search by Google ⁴ and the Europe Media Monitor [2]. Additionally, there have been recent projects such as Buceador that performs research on integration of all of them in a multilingual and multimodal information retrieval system [1] or XLike that develops technology to monitor and aggregate knowledge that is currently spread across mainstream and

¹ <http://trec.nist.gov>

² <http://www.clef-initiative.eu>

³ <http://research.nii.ac.jp/ntcir/>

⁴ http://translate.google.com/translate_s

social media, and enables cross-lingual services for publishers, media monitoring and business intelligence [15].

Given a query in a source language, the aim of CLIR is retrieving most similar and related documents in a target language. [13] identified four types of strategies for matching a query with a set of documents in the context of CLIR by: cognate matching, document translation, query translation or interlingua techniques. From these techniques the most commonly used is query translation.

Query translation is the approach where queries are translated into the document language. It is the most popular approach due to its tractability. Query translation methods translate user queries to the language of the document collection. Query translation has been applied by most CLIR experimental systems because of its convenience and the translation has been mainly addressed by using dictionary-based (i.e. using machine-readable dictionaries, MRD), machine translation (MT) and/or parallel texts techniques [5].

On the other hand, interlingua methods transforms both documents and queries into a language-independent representation. An interlingua method aims at associating related textual contents among different languages by means of language-independent semantic representations. The conventional interlingua-based CLIR approach uses latent semantic indexing (LSI) for constructing a multilingual vector-space representation [8, 9, 6] of a given parallel document collection. Vector-space representations are known to be noisy and sparse. That is why in order to obtain more efficient representations, space reduction techniques such as latent semantic indexing and probabilistic latent semantic indexing [10] are applied. The new reduced-space dimensions are supposed to capture semantic relations between words and documents in the collection. Recent approaches have achieved interesting results by using regression canonical correlation analysis (an extension of canonical correlation analysis) where one of the dimensions is fixed and demonstrate how it can be solved efficiently [14]. Also the use of nonlinear semantic mapping techniques have been proposed in our previous works [4].

The query translation approach has been considered as the only state-of-the-art approach for CLIR applications. However, in a N -lingual environment the number of required systems reaches the $\frac{N(N-1)}{2}$. The main advantage of the interlingua-based strategy in a highly multilingual environment is that, compared to the query translation strategy, it reduces cross-language information retrieval number of systems to N . That is why, we propose to test interlingua versus query translation for one particular application which consists on retrieving a book from the collection by using one of its chapters in a different language as a query.

The rest of the paper is structured as follows. Next section describes the methodologies that are compared in this paper. Section 3 reports several CLIR experiments performed on a trilingual document collection. The LSI methodology is compared with that of a standard IR system and the query translation CLIR approach showing that, in the case of cross-language tasks, the proposed approach is able to outperform the conventional one. Results are analysed in order to find out if the quality of translation can predict the quality of the query translation approach. Finally, Section 4 includes the most relevant conclusions derived from the experimental results are presented and some future research actions for continuing the present work are depicted.

2 CLIR methodologies

This section briefly describes the LSI-interlingua and the query translation methodologies. The LSI-interlingua methodology basically uses the singular value decomposition (SVD) of a tf-idf (term frequency - inverse document frequency) matrix, which considers that a rectangular matrix X of dimensions $M \times N$ can be factorized:

$$X = U \Sigma V^T \quad (1)$$

U and V are unitary matrices of dimensions $M \times M$ and $N \times N$, respectively, and Σ is a $M \times N$ diagonal matrix containing the singular values associated to the decomposition. Consider M the number of terms and N the number of documents. According to [9], a low-dimensional representation of a given document vector x can be obtained as follows:

$$y^T = x^T U_{M \times L} \quad (2)$$

y is the L -dimensional document vector corresponding to the projection of an M -dimensional document vector x , and $U_{M \times L}$, is a matrix containing the L first column vectors of the unitary matrix U that is obtained from (1) given X . This rank reduction has been proven to preserve most important semantic information in the collection of documents while reducing noise. This LSI methodology can be extended to the cross-language case [9], where the main difference is that X is a term-document matrix constructed with parallel documents in two languages:

$$[X_a; X_b] = U_{ab} \sum_{ab} V_{ab}^T \quad (3)$$

$[X_a; X_b]$ is a bilingual term-document matrix obtained by concatenating monolingual term-document matrices for a parallel document collection. In this case, low-dimensional representations for given document vectors x_a and x_b in languages a and b , can be obtained:

$$y_a^T = [x_a; 0]^T U_{abM \times L} \quad (4)$$

$$y_b^T = [0; x_b]^T U_{abM \times L} \quad (5)$$

If we assume that similar terms in multiple languages have approximately the same occurrence patterns, then, we could find a close representation in the multilingual reduced space for semantically related terms and documents. In this case, documents across languages could be compared in the reduced space. Finally, a measure of similarity could be used to compute the similarity among documents.

In contrast, the query translation methodology simply translates the query using a standard machine translation system (MT) and it uses a monolingual information retrieval system (IR). Notice that in this case, errors from the first step (MT) are concatenated with errors from the second step (IR). In this work, we are not considering query expansions and k-bases translations of queries that are more sophisticated ways of performing query translation.

3 Experiments

The data has been extracted from the Chinese, Spanish and English versions of the *Holy Bible*, which has been proved to be a good resource for CLIR experimenting [7]. The basic characteristics of the collection are described in Table 1.

Table 1. Basic characteristics of the experimental dataset.

Language	Chapters	Books	Vocabulary	Non singletons
Chinese	1189	66	12,670	6,286
Spanish	1189	66	26,251	13,632
English	1189	66	13,216	7,265

The task consists in retrieving a book from the collection by using one of its chapters in the same or different language as a query. We have randomly selected 200 chapters from the 1189 total chapters to be used as test set.

For evaluation purposes, the LSI-interlingua method is compared with standard CLIR approach. In the former, the training set is variable (5 to 300 chapters) and the test set are 60 books. The retrieval space dimensionality is equal training size. The number of performed runs is 100. The query translation was implemented concatenating the Google translation API and the monolingual information retrieval system, which was implemented by using Solr. Solr is an XML-based open-source search server based on the Apache-Lucene search library⁵.

In order to find out the correlation between why one system is better than the other, Table 2 shows the quality of translation using standard measures such as BLEU (i.e. Bilingual Evaluation Under Study) and PER (i.e. Position Error Rate).

Table 2. Translation quality. BLEU and PER metrics.

	Chinese		Spanish		English	
	BLEU	PER	BLEU	PER	BLEU	PER
Chinese	-	-	12.01	63.5	15.68	57.72
Spanish	12.77	64.86	-	-	26.47	44.69
English	17.94	63.13	28.07	45.72	-	-

Finally, we performed analysis of the correlation between the results and the translation metrics, see Figure 2. From the aforementioned figures one can see:

1. LSI (interlingua) outperforms query translation in five situations out of twelve and in one situation both techniques obtain equal results;

⁵ <http://lucene.apache.org/solr/>

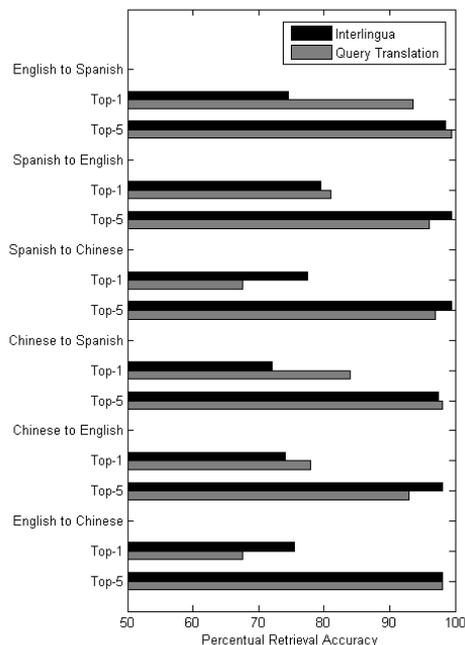


Fig. 1. Results for the cross-language information retrieval task for all pairs of languages. For each language pair, the first two bars correspond to the Top-1 results and the second two bars correspond the Top-5 results.

2. Translation metrics do not correlate well with any metric of CLIR. This means that the quality of translation evaluated with standard MT metrics does not provide information in a CLIR system.

These results differ from previous shown in [11]. We suspect that the main reason for contradicting the Kettunen conclusions is that we are working in a different CLIR task and the query is specially large; (3) the Top-1 and Top-5 metrics are correlated when using the interlingua approach but not in the query translation approach. Finally, we analysed the errors from both systems. Among the total errors only between 10% and 30% of the cases were the same errors, which indicates that a system combination could improve the task performance.

4 Conclusions and Future Work

This paper presents one particular application which consists on retrieving a book from the collection by using one of its chapters in a different language as a query. In this

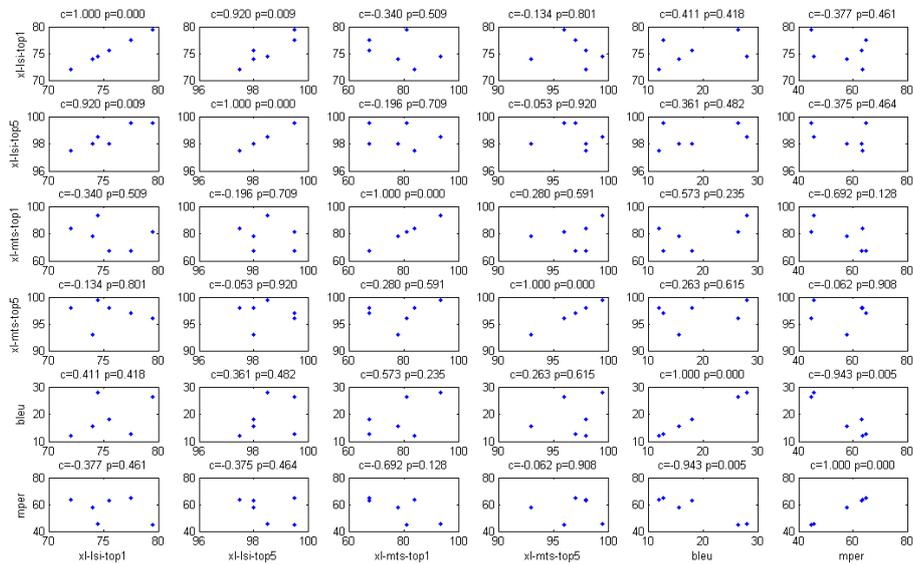


Fig. 2. Correlations (c) and p-value (p) between CLIR quality and MT quality (BLEU and PER). LSI stands for Latent Semantic Indexing and MTS stands for query translation.

framework, the LSI (or interlingua) method is compared with state-of-the-art CLIR approach. Evaluation results show that the proposed method is able to outperform the reference system in the case of cross-language information retrieval (specially when considering the Top-5 results). Additionally, comparing the errors of both system outputs, there is a maximum of 30% matching. This result was somehow expected given the different nature of both systems. Taking advantage of this information, we could try to perform system combination methods. Finally, we found that for this particular task, the translation quality (in terms of BLEU and PER) is not correlated with the CLIR quality. This may be explained because, for our task, the query is specially large. As future research in this area, we will focus on finding a correlation between the quality of translation and the interlingua and query translation performances.

With these results, we showed that there is hope for interlingua methods. Specially taking into account that, if we get similar results to a query translation approach, in a N -lingual environment, interlingua approaches reduce the number of systems from $\frac{N(N-1)}{2}$ to N .

Acknowledgements. This work has been supported in part by Spanish Ministerio de Economía y Competitividad, contract TEC2012-38939-C03-02 as well as from the European Regional Development Fund (ERDF/FEDER).

References

1. Adell, J., Bonafonte, A., Cardenal, A., Costa-Jussà, M.R., Fonollosa, J.A.R., Moreno, A., Navas, E., Banga, E.R.: Buceador, a multi-language search engine for digital libraries. In: Chair, N.C.C., Choukri, K., Declerck, T., Doğan, M.U., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S. (eds.) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). European Language Resources Association (ELRA), Istanbul, Turkey (may 2012)
2. Atkinson, M., der Goot, E.V.: Near real time information mining in multilingual news. In: Proceedings of the 18th international conference on World wide web. pp. 1153–1154 (2009)
3. Banchs, R.E., Costa-jussà, M.R.: A non-linear semantic mapping technique for cross-language sentence matching. In: Proceedings of the 7th International Conference on Advances in Natural Language Processing. pp. 57–66. IceTAL'10, Springer-Verlag, Berlin, Heidelberg (2010)
4. Banchs, R.E., Costa-jussà, M.R.: Cross-language document retrieval by using non-linear semantic mapping. Applied Artificial Intelligence Journal 27(9), 781–802 (2013)
5. Chen, J., Bao, Y.: Cross-language search: The case of google language tools. First Monday 14(3) (2009)
6. Chew, P., Abdelali, A.: Benefits of the passively parallel Rosetta stone? Cross-Language information retrieval with over 30 languages. In: Proc of the 45th Annual Meeting of the Association for Computational Linguistics. vol. 45, pp. 872–879 (2007)
7. Chew, P.A., Verzi, S.J., Bauer, T.L., McClain, J.T.: Evaluation of the bible as a resource for cross-language information retrieval. In: Proceedings of the Workshop on Multilingual Language Resources and Interoperability. pp. 68–74 (2006)
8. Deerwester, S., Dumais, S., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6), 391–407 (1990)
9. Dumais, S.T., Landauer, T.K., Littman, M.L.: Automatic cross-linguistic information retrieval using latent semantic indexing. In: SIGIR96 Workshop on Cross-Linguistic Information Retrieval. pp. 16–23 (1996)
10. Hofmann, T.: Probabilistic latent semantic analysis. In: Proceedings of Uncertainty in Artificial Intelligence, UAI99. pp. 289–296 (1999)
11. Kettunen, K.: Choosing the best MT programs for CLIR Purposes: Can MT metrics be helpful? In: Proc. of the 31th European Conference on IR Research on Advances in Information Retrieval. pp. 706–712 (2009)
12. Nie, J.: Cross-Language Information Retrieval. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2010)
13. Oard, D.W., Diekema, A.R.: Cross-Language information retrieval. Annual Review of Information Science and Technology (ARIST) 33, 223–256 (1998)
14. Rupnik, J., J., S.T.: Multiview canonical correlation analysis and cross-lingual information retrieval. In: http://videolectures.net/lms08_rupnik_rcca/ (2008)
15. Wang, Z., Li, J., Zhao, Y., Setchi, R., Tang, J.: A unified approach to matching semantic data on the web. Knowl.-Based Syst. 39, 173–184 (2013)

Acoplamiento de micro multitudes para el desarrollo de videojuegos controlados por movimiento

Iván Rivalcoba¹, Krely Rodríguez², Oriam Degives¹, Isaac Rudomín³

¹ Tecnológico de Monterrey, Campus Estado de México,
México

² Tecnológico de Minatitlán,
Minatitlán, Veracruz, México

³ Barcelona Supercomputing Center
Barcelona, España

Resumen. La simulación de multitudes en tiempo real y los juegos controlados por movimiento se han vuelto muy populares en los últimos años. En conjunto estas dos tecnologías proporcionan una mejor experiencia de juego en entornos virtuales logrando escenas más realistas y vibrantes. Sin embargo, hasta ahora no se ha explotado la interacción de múltiples jugadores con una gran multitud bajo un entorno virtual. En este trabajo presentamos un sistema no intrusivo capaz de simular multitudes virtuales acopladas en tiempo real con varios usuarios, sentando con ello las bases para la creación de juegos donde interactúen muchos jugadores con muchas personajes, para ello se realiza una detección de personas en una secuencia de vídeo, nuestra contribución consiste en utilizar patrones binarios semánticos locales (S-LBP) como descriptores, creando con ello un sistema inteligente capaz de segmentar detectar y seguir personas en vistas aéreas, así mismo para aumentar la velocidad del dibujado y acoplamiento se aprovecha la unidad de procesamiento gráfico (GPU), almacenando la información en *texture shaders* delineando una arquitectura de procesamiento concurrente. Los resultados experimentales muestran un sistema viable para la interacción de multitudes virtuales contra más de una docena de personas reales simulados en hardware de consumo y en tiempo real (30 FPS), así mismo el uso de los patrones binarios semánticos locales como descriptores permitieron reducir los falsos positivos que normalmente se producen al emplear descriptores HOG (Histograma de gradientes orientados) usados frecuentemente en el reconocimiento de personas.

Palabras clave: Micro multitudes, videojuegos controlados por movimiento.

1. Introducción

La simulación de multitudes en tiempo real se ha convertido en una de las principales líneas de investigación en los juegos de computadora, películas y realidad virtual [27]. Sin embargo, solo unos pocos juegos utilizan las multitudes

como una forma de mejorar la experiencia de juego, esto se debe en gran medida a la capacidad limitada en el hardware de los usuarios finales, especialmente de las consolas de juegos. Los vídeo juegos han utilizado la simulación de multitudes para lograr escenas más realistas y vibrantes [8, 23].

Películas como la trilogía del Señor de los Anillos, Avatar y Guerra Mundial Z, hacen un amplio uso de las técnicas desarrolladas para la simulación de multitudes. Miles de personajes complejos interactúan entre sí y con su ambiente, alcanzando un sentido de inmersión más profundo en los espectadores. Sin embargo, debido a la naturaleza de la industria cinematográfica, en la cual se da prioridad al realismo, el empleo de simulación fuera de línea es la opción típica, por lo que la exigencia en el tiempo no es crucial; por otro lado los videojuegos son un caso aparte, debido a que el tiempo de dibujado por cada cuadro es crucial para no experimentar retardos en la animación, por lo que esta industria requiere de algoritmos más optimizados para lograr escenas en milésimas de segundo [7].

Los juegos controlados por movimiento se han vuelto muy populares en los últimos años con la introducción del hardware dedicado a la detección de movimiento, como el *Kinect* de Microsoft y el *Playstation Eye* de Sony. Sin embargo, en la medida de nuestro conocimiento, no existe juego que combine ambas características, el permitir a varios usuarios interactuar con una gran multitud virtual. El objetivo principal del presente trabajo es llevar a cabo el acoplamiento entre un grupo de personas reales y la simulación de una multitud de humanos virtuales. Para ello, proponemos un sistema interactivo de dos etapas. En la primera etapa un sistema de visión detecta personas desde una secuencia de vídeo todo ello en tiempo real. Esta etapa no requiere de una cámara especial como los otros sistemas de detección mencionados, una cámara web de hasta 3MB de resolución ofrecerá la calidad necesaria para permitir el funcionamiento óptimo del sistema. La posición de los usuarios capturados es utilizada en la segunda etapa, dicha etapa consiste en realizar el acoplamiento entre multitudes virtuales y usuarios. Con ello se logra que los personajes virtuales reaccionen ante la presencia de los humanos reales.

1.1. Trabajo relacionado

Una parte crucial del sistema interactivo presentado en este artículo, es la cantidad de tiempo requerido para realizar las tareas de simulación y presentación. Para obtener una comprensión más profunda del sistema, esta sección describe los desafíos y el estado del arte relacionado con la visión artificial y la simulación de multitudes.

Thalmann et al. presenta un sistema interactivo de navegación en el cual el usuario tiene la capacidad de controlar un personaje en la multitud a través de una interface natural, pero la interacción sigue siendo entre un único usuario y personajes virtuales [27]. Pelechano propuso un experimento para estudiar de cerca el comportamiento de las personas, en una fiesta de cocktail, interactuando con personajes virtuales [21].

Helbing et al. presentó el modelo de fuerzas sociales, en el cual, cada agente interactúa con el medio ambiente dependiendo de sus motivaciones internas [10]. Helbing impulsa su investigación y estudios, en el comportamiento de los peatones en condiciones normales y en situaciones de evacuación [9]. El modelo de las fuerzas sociales se ha ampliado para apoyar a los grupos de peatones relacionados entre sí [16].

Treuille et al. presentó el modelo de Multitudes *Continuum*, el cual unifica las etapas de planificación y ruta de evasión de colisión mediante el uso de la velocidad y campos potenciales. [24]. Millan et al. utilizó las fuerzas de atracción y repulsión codificadas en texturas para dirigir los agentes a través del medio ambiente [15]. Van den Berg et al. Propuso *Reciprocal Velocity Obstacles*, un conjunto de velocidades que conducirán a una colisión entre agentes [4]. Los agentes deben elegir velocidades que no pertenezcan a este conjunto para moverse a través de rutas libres de colisiones. Los mismos autores redujeron el problema a la solución de un sistema lineal de baja dimensión [3]. Un enfoque de prevención de colisiones basado en visión es presentada por Ondřej et al., usando visión sintética para detectar obstáculos en el campo de vista de cada agente [19].

En cuanto al reconocimiento de humanos, existen varias obras en la comunidad de investigación que se ocupan de este problema. Viola et al. Presento una técnica para la detección de peatones en una secuencia de vídeo, mediante el aprovechamiento de la apariencia del caminar de una persona [26]. Dalal et al. realizó un estudio completo de los Histogramas de gradientes orientados (*HOG*) aplicado a la representación de humanos [6] para detectar peatones con éxito. Tuzel et al. detectó peatones mediante la representación de una región de la imagen como matrices de covarianza de localidades espaciales, intensidades, y derivadas, por nombrar algunas [25].

En la mayoría de los casos, los sistemas de detección incluyen una etapa de preprocesamiento, la cual consiste en la eliminación del fondo de la escena. Banerjee et al. Convino un modelo de fondo adaptable con características descriptores *HOG*, aprendiendo solo las partes fijas de la escena e ignorando el primer plano en movimiento [2]. Otro enfoque es presentado por Bhuvaneshwari et al. Llamados rasgos *edgelet*, los cuales son segmentos cortos que pertenecen a una línea o curva [5].

Los trabajos mencionados anteriormente se ocupan de escenas en perspectiva lateral, las cuales son deseables para aplicaciones de video vigilancia, sin embargo, no son recomendados cuando se trata con el análisis de multitudes, debido a la oclusión que se presenta cuando el tamaño de la multitud se incrementa, resultando en una decaimiento de la precisión en los sistemas de reconocimiento. Lengvenis et al. Utilizó una perspectiva con vista aérea (ojo de pájaro) para la detección de los peatones en el transporte público de la ciudad de Kaunas [13]. Ozturk et al. Propuso un sistema para determinar la orientación de la cabeza y cuerpo de los humanos en las escenas de visión aérea [20]. Su sistema utiliza una aproximación de contextos de forma para detectar la orientación básica del cuerpo y propone un flujo óptico basado en las propiedades de la transformación característica invariante a la escala (*SIFT*), sin embargo, las oclusiones se dejan

como trabajo futuro, haciendo a este algoritmo inutilizable para escenas con multitudes.

2. Arquitectura del sistema

El sistema que describimos en este trabajo está diseñado para ser interactivo incluso en un hardware de consumo, siendo capaz al mismo tiempo de capturar peatones desde las imágenes de vídeo, simulando miles de agentes virtuales. Hemos implementados las técnicas descritas usando C++ para el código en CPU, y OpenGL moderno para la simulación y visualización, usando un sistema operativo Windows 7 x64 con un procesador Intel Core i7-2630 QM de cuatro núcleos a 2.00 GHz, 8GB de memoria y tarjeta gráfica NVIDIA GT 540M. El GT 540M es una tarjeta gráfica de la familia Fermi con la capacidad para manejar la versión de OpenGL 4.3; es considerada una GPU de gama baja-media. Para las etapas de dibujado y simulación hemos empleado el API de OpenGL moderno específicamente en su versión 4.3.

OpenGL introdujo los *Computer Shaders* a su API, el cual habilita su lenguaje de *shaders* para llevar a cabo cómputo de propósito general en hardware gráfico. Los *Compute Shaders* proporcionan un enfoque alternativo al cómputo de propósito general en unidades de procesamiento gráfico (GPGPU), además de Nvidia CUDA y OpenCL del Grupo Khronos. OpenGL es más accesible que Cuda, ya que a diferencia de Cuda y OpenCL, los *Compute Shaders* son similares a otras etapas de sombreado en la pipeline de gráficos y son capaces de leer desde texturas, imágenes y buffers así como escribir imágenes y buffers sin mapeo a otras estructuras de datos o cambiar entre los gráficos y contexto de cálculo. Estas características proporcionan una clara ventaja para las aplicaciones que requieren interoperabilidad de gráficos, como los videojuegos. Una visión general del sistema puede observarse en la figura 1.

Para lidiar con los problemas de oclusión presentados en escenas con una gran cantidad de personas, se decidió utilizar una perspectiva aérea también conocida como vista de pájaro. Por otra parte también se tuvo cuidado en los procesos que involucraban la segmentación y eliminación de fondo, de modo que el sistema pudiera enfocar sus esfuerzos computacionales en las tareas de clasificación, así mismo se investigó la mejor forma de representar humanos para las etapas de reconocimiento de personas en filmaciones desde vistas aéreas, para eso se consideró la deformación que se presenta cuando un humano se aleja de la cámara, para lo cual se tomó la decisión de segmentar la acción de los clasificadores, haciendo dependiente el tipo de clasificador empleado con respecto a la distancia entre el usuario y la cámara.

El sistema que realiza la detección humana consiste en 6 etapas mostradas en la Figura 2. La primera etapa consiste en la captura y digitalización del vídeo, posteriormente se realiza la sustracción del fondo, esta etapa es responsable de la división de la escena en el fondo de las regiones de primer plano. Las regiones de primer plano son etiquetadas como Regiones de interés (ROI), la separación

permite al sistema concentrar los esfuerzos computacionales en las tareas de detección de humanos (objetos en movimiento).

La sustracción del fondo se lleva a cabo utilizando una versión modificada del algoritmo codebook presentado por [12]. Para hacer el algoritmo más robusto contra los cambios de iluminación, empleamos un segundo codebook que es guardado en cada cuadro; este codebook auxiliar genera unas palabras claves considerando solo las regiones sin movimiento. Una vez que ha transcurrido cierto lapso de tiempo, este codebook es limpiado y es adherido a un codebook principal. Una vez removido el fondo, las ROIs, son etiquetadas y convertidas de *Red-Green-Blue* (RGB) a una escala de grises.

Una vez que se han obtenido el conjunto de regiones ROI, en una tercera etapa las ROI son pasadas una a una al módulo de extracción de características, en este caso los descriptores seleccionados para modelar la parte superior del cuerpo humano son una versión modificada de patrones binarios locales (LBP) presentados por [18], estos son llamados patrones binarios locales semánticos, utilizando estos descriptores se genera un vector de características que modela el objeto detectado como un histograma, Cada uno de estos vectores es enviado a una máquina de vectores de soporte (*SVM*) entrenada previamente misma que realiza la clasificación etiquetando la ROI como humano o no humano. Posteriormente en la cuarta etapa se realiza el seguimiento de cada personaje utilizando el algoritmo de *Lucas-Kanade*, cada uno de estos puntos es almacenado en la quinta etapa en *texture shaders* siendo enviados a la *GPU* para ser dibujados con los personajes virtuales.

Se tomó la decisión de seguir el área correspondiente a la cabeza y hombros del cuerpo humano debido al hecho de que la cabeza humana permanece casi constante sobre toda la escena [17], a diferencia de otras partes del cuerpo. La sección de cabeza y hombros presentan una forma característica en forma de Ω , ver figura 3. Esta propiedad es la que hace que la cabeza y hombros sean las partes más estables del cuerpo para ser detectadas y rastreadas.

2.1. Sistema de reconocimiento usando *S-LBP*

Nuestra principal aportación recae en la etapa de detección de personas la cuál es implementada de la siguiente forma: Un operador LBP asigna una etiqueta a cada píxel de una imagen por un umbral a la zona de 3x3 de cada píxel con su propio valor y considerando el resultado como un número binario, por ejemplo dado un píxel (x_c, y_c) , el LBP de ese píxel es calculado usando la ecuación 1, donde P corresponde al número de puntos de muestreo, R es el radio de la zona a muestrear g_p representa cada punto de muestreo y g_c representa el píxel (x_c, y_c) ; el histograma de los números obtenidos a lo largo de toda la imagen es usada como el descriptor.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (1)$$

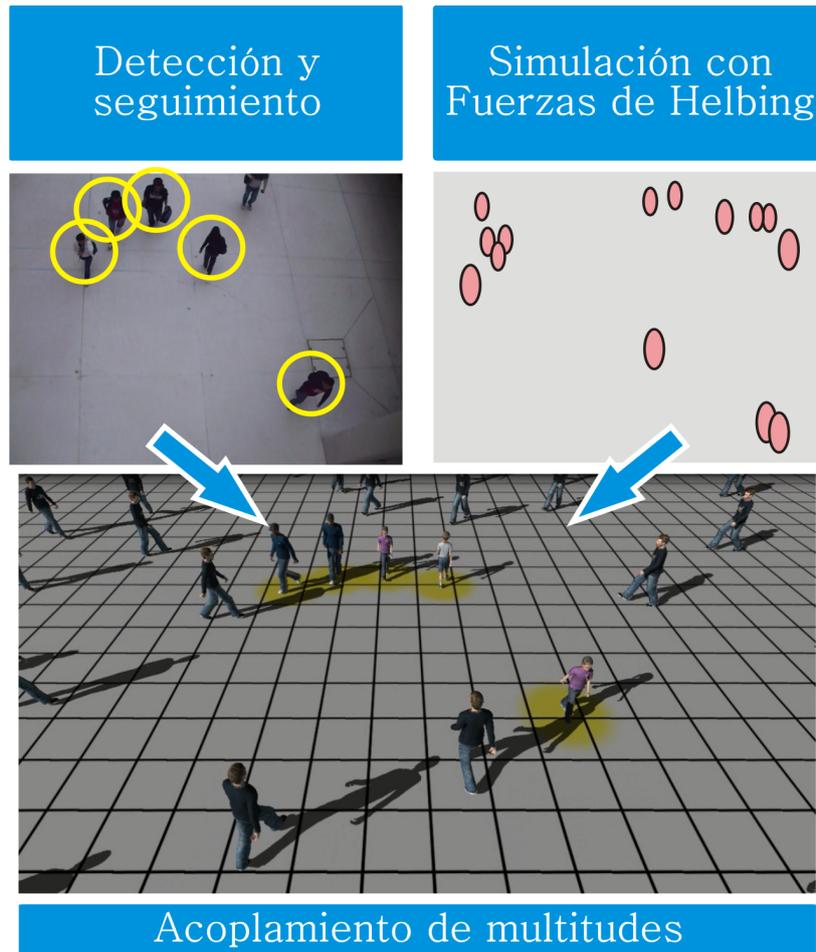


Fig. 1. Flujo del sistema. Los datos de los peatones capturados por la etapa de detección se fusionan con la simulación virtual produciendo un acoplamiento entre la parte virtual y real, con esto se consigue que la animación en el ambiente virtual sea influenciada por los peatones capturados en la secuencia de vídeo

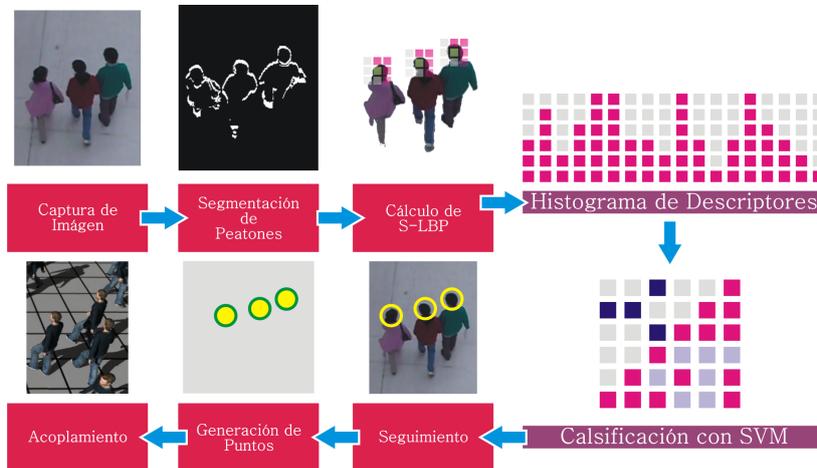


Fig. 2. Etapas del sistema de acoplamiento. Se muestra cada una de las seis etapas que componen el sistema de acoplamiento



Forma Omega Cabeza-Hombros

Fig. 3. Forma omega del cuerpo humano. Desde una vista aérea un ser humano muestra una silueta tipo omega característica, muy útil en los procesos de reconocimiento y segmentación, la silueta se preserva a pesar de los cambios sutiles de perspectiva

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \quad (2)$$

Aunque el operador LBP tiene muchas ventajas sobre otros descriptores basados en gradiente, la LBP no es suficientemente robusta para ser usada como descriptor humano, sin embargo sus puntos fuertes pueden ser explotados usando una variante llamada *LBP semántica* (S-LBP) propuesta por [11]. S-LBP está formado de LBP uniformes. Un LBP se llama uniforme si se presentan como máximo dos transiciones de bit a bit entre 0 y 1, en otras palabras, varios bits continuos “1” forman un arco en el círculo de muestreo; estos arcos puede ser representados con su dirección principal y la longitud del arco (ver figura 4); mediante el uso de esta estrategia, la dimensión del vector característico es reducido.

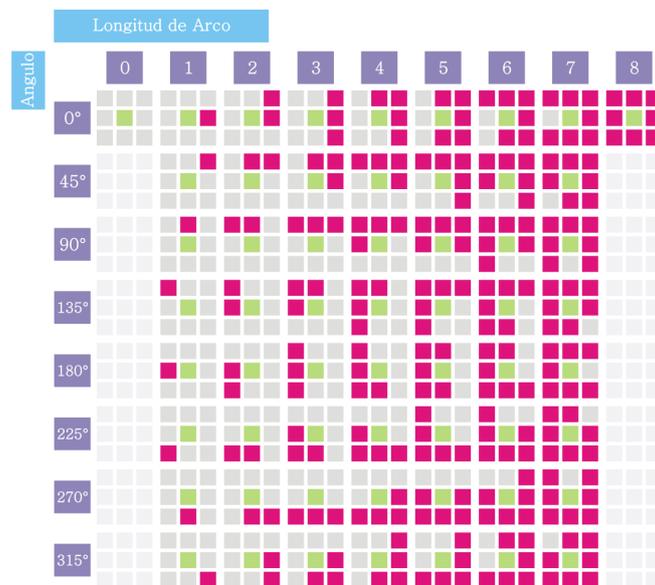


Fig. 4. Patrones Locales Binarios Semánticos. La figura muestra las 58 combinaciones semánticas de los LBP (Patrones Locales binarios - Local Binary Patterns), cada columna representa a la longitud del arco, por otra parte las filas representan el ángulo principal del LBP

En nuestro trabajo, utilizamos un núcleo de detección de 32x32 pixeles de tamaño, el cual produce un vector 1160D; todos los S-LBPs son usados para calcular un histograma, cada histograma representa el vector característico de cada región. Para acelerar el proceso de cálculo de SLBPs, proponemos el uso de una tabla de búsqueda para la clasificar el tipo de S-LBP. Se ha demostrado por

[25] que el conjunto de sub-ventanas de tamaño variable puede favorecer en gran medida la eficiencia de detección. Debido a esto, hemos adoptado un enfoque de varias resoluciones similares a [1], dividiendo cada núcleo de detección en ventanas locales de diferentes tamaños para capturar mejor la estructura principal de la cabeza y hombro de los seres humanos. El histograma de descriptores S-LBPs es ahora extraído para cada ventana de forma independiente (formando un total de 20 subventanas) y luego concatenada para formar una descripción global de la región. Esta estrategia de extracción de características es utilizada tanto para entrenar el sistema y para clasificar la imagen en tiempo de ejecución. Utilizamos la librería OpenCV SVM para la generación del vector de apoyo para la tarea de clasificación, el conjunto de datos utilizado es similar [14], al que utiliza la vista superior de tiros de cabeza y los hombros humanos. Una observación importante es que los peatones muestran una deformación debido a la perspectiva de la cámara, haciendo difícil la sintonización hasta el clasificador [22].

Para superar este problema, y así mejorar los resultados de la detección, dividimos la vista de la cámara en segmentos en función de la distancia entre la cámara y el peatón; un clasificador diferente es usado dependiendo de la región donde el usuario es detectado. Una vez que la cabeza humana es detectada, un punto es asociada a ella; cada punto detectado es rastreado sobre todos los marcos usando la versión piramidal de rastreo de Lucas Kanade. Para nuestro sistema de simulación, toda lógica relacionada con el control del agente se realiza en el hardware de gráficos, el cual nos permite realizar simulaciones a gran escala con miles de agentes en tiempo real. Para la interoperabilidad entre la representación y la simulación, almacenamos todos los datos requeridos para la simulación en la memoria de textura. Los datos registrados por el sistema de visión son almacenados en las mismas texturas que los utilizados por los agentes simulados, y la representación visual del arreglo entre los peatones y los agentes simulados puede observarse en la Figura 5.

2.2. Comportamiento de los agentes virtuales

Para dirigir los agentes simulados, utilizamos una implementación paralela del trabajo realizado por Moussaid et al [16]. Se requieren tres texturas de datos para la simulación: texturas de posición, velocidad y destino.

Cada elemento de textura (*texel*) i contiene los datos del agente o usuario i . Nosotros utilizamos la ecuación (3) para calcular la nueva posición de cada agente simulado:

$$\mathbf{f}_i = \mathbf{f}_i^d + \mathbf{f}_i^n + \mathbf{f}_i^o + \mathbf{f}_i^g \quad (3)$$

Donde \mathbf{f}_i representa la aceleración del agente i en el cuadro actual, y es afectado por motivaciones internas \mathbf{f}_i^d , sus vecinos \mathbf{f}_i^n , los obstáculos más próximos \mathbf{f}_i^o , y su grupo dinámico \mathbf{f}_i^g . La nueva velocidad $\mathbf{v}_{i_{new}}$ se obtiene mediante la adición de la aceleración a la velocidad anterior del agente \mathbf{v}_i . Finalmente, la posición del agente es actualizada en sus componentes de aceleración y velocidad.

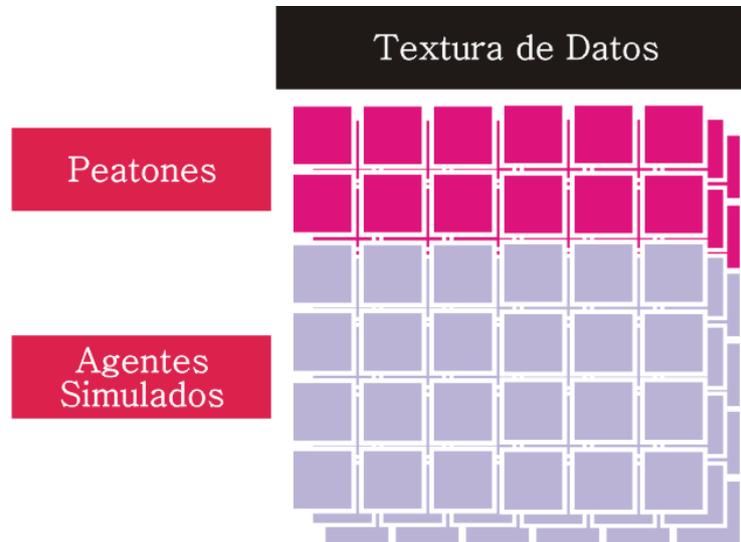


Fig. 5. Almacenamiento de la información en texturas. Los datos de los peatones capturados en las secuencias de vídeo son almacenadas en las mismas texturas que los agentes simulados, permitiendo un cómputo más eficiente así mismo se aprovechan las capacidades de cómputo de propósito general de las unidades de procesamiento gráfico

Detectamos vecinos basados en sus posiciones mediante un algoritmo *hashlike* espacial. El algoritmo busca los vecinos más cercanos utilizando las posiciones de textura. Debido a que tanto los agentes simulados y los peatones capturados almacenan sus posiciones en esta textura, los agentes simulados son conscientes de los peatones reales capturados, justo como si fueran otros agentes virtuales. Al ser conscientes de los peatones capturados, los agentes virtuales pueden reproducir tareas de evasión y realizar cualquier otra forma de interacción de la misma manera que con los agentes virtuales vecinos. Solo existen unas pocas obras en el estado del arte que tienen que ver con la interacción entre un grupo de personas reales y una multitud virtual, sin embargo, estos sistemas son típicamente muy limitados en el número de usuarios capturados o en el tamaño de la multitud. Por otro lado, la segmentación de peatones se realiza generalmente utilizando procesos manuales o procesos semi-automáticos, o en formas más invasivas como confiar en el uso de marcadores.

3. Resultados

Las vistas frontales dedicadas al uso de hardware, como el Kinect y PlayStation Eye, limitan el área y el número de personas que pueden ser detectados. Los resultados muestran que nuestro sistema puede realizar una interacción en tiempo real entre un grupo de usuarios y un grupo simulado de personajes virtuales. Con el uso de un codebook auxiliar para el modelo de fondo, el sistema

es capaz de mantener su funcionalidad frente a los cambios de iluminación. La combinación de LBP semánticas y la subdivisión de la escena en segmentos permitieron reducir la complejidad de la etapa de clasificación, aún a pesar de las deformaciones sufridas por los usuarios a medida que se alejan de la cámara. El uso de cámaras especiales no es necesario, cámaras web estándares cumplen plenamente las especificaciones técnicas mínimas para hacer el sistema confiable. La velocidad de fotogramas alcanzado fue por encima de 30 cuadros por segundo, suficiente para ser considerado un sistema en tiempo real. En el sistema de simulación, somos capaces de simular miles de agentes autónomos. El uso del modelo de fuerzas sociales en el lado de la simulación, produce un efecto plausible de interacción entre los dos mundos.

4. Conclusión y trabajo futuro

Hemos presentado un sistema interactivo que permite la interacción entre los usuarios peatonales y una multitud virtual, usando solo el hardware de consumo de grado. Este sistema es capaz de detectar y rastrear más de una docena de peatones en una escena, simulando miles de agentes virtuales y permitir la interacción entre los peatones y los agentes, manteniendo velocidades de escenas en tiempo real.

El sistema presentado en este trabajo puede ser útil para probar nuevos comportamientos de agentes virtuales, ya sean basadas en datos o basados en normas, ya que permite una comparación visual con peatones reales. También puede permitir simulaciones de planificación urbana y de seguridades más precisas, ya que una parte de la población en la multitud está constituida por personas reales.

En la industria del entretenimiento, este sistema proporciona una contribución al área de trabajo de aplicaciones que buscan la interacción entre el mundo real y un mundo virtual. Sobre la base de nuestro conocimiento, no hay juego en el mercado que utilice tanto la simulación de multitudes y la detección de multitudes para crear mundos y experiencias más inmersivas.

Como trabajo futuro se propone el empleo de dispositivos *wearables* como el *Google-Glass* para llevar el acoplamiento a la interacción permitiendo que las personas o usuarios reales puedan reaccionar a los personajes virtuales.

Referencias

1. Ahonen, T.: Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence* 28(12), 2037–41 (Dec 2006)
2. Banerjee, P., Sengupta, S.: Human motion detection and tracking for video surveillance. *Proceedings of the national Conference of tracking and video surveillance activity analysis* pp. 88–92 (2008)
3. van den Berg, J., Guy, S.J., Lin, M., Manocha, D.: Reciprocal n-body collision avoidance. *Robotics Research* 70, 3–19 (2011)

4. van den Berg, J., Manocha, D.: Reciprocal Velocity Obstacles for real-time multi-agent navigation. In: 2008 IEEE International Conference on Robotics and Automation. pp. 1928–1935. IEEE (May 2008)
5. Bhuvaneshwari, K., Rauf, H.A.: Edgelet based human detection and tracking by combined segmentation and soft decision. *Control, Automation, Communication and Energy Conservation* (June), 4–9 (2009)
6. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 1, 886–893 (2005)
7. De Gyves, O., Toledo, L., Rudomín, I.: Comportamientos en simulación de multitudes : revisión del estado del arte. *Research in Computer Science*. 62(Special Issue: Avances en Inteligencia Artificial.), 319–334 (2013)
8. Groenewegen, S.: Improving crowd behaviour for games and virtual worlds. In: *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*. pp. 256–258. ACM Press, New York, New York, USA (2010)
9. Helbing, D., Farkas, I.J., Molnár, P., Vicsek, T.: Simulation of Pedestrian Crowds in Normal and Evacuation Situations. *Pedestrian and evacuation dynamics* 21 (2002)
10. Helbing, D., Molnár, P.: Social force model for pedestrian dynamics. *Physical Review E* 51(5), 4282–4286 (May 1995)
11. Huang, T.: Discriminative local binary patterns for human detection in personal album. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (Jun 2008)
12. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. *Real-Time Imaging* 11(3), 172–185 (Jun 2005)
13. Lengvenis, P., Simutis, R., Vaitkus, V., Maskeliunas, R.: Application Of Computer Vision Systems For Passenger Counting In Public Transport. *Electronics and Electrical Engineering* 19(3), 69–72 (Mar 2013)
14. Li, M., Zhang, Z., Huang, K., Tan, T.: Rapid and robust human detection and tracking based on omega-shape features. In: 2009 16th IEEE International Conference on Image Processing (ICIP). pp. 2545–2548. IEEE (Nov 2009)
15. Millan, E., Hernandez, B., Rudomin, I.: Large Crowds of Autonomous Animated Characters Using Fragment Shaders and Level of Detail. In: Wolfgang Engel (ed.) *ShaderX5: Advanced Rendering Techniques*, chap. Beyond Pix, pp. 501—510. Charles River Media (2006)
16. Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz, G.: The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS one* 5(4), e10047 (Jan 2010)
17. Mukherjee, S., Das, K.: Omega Model for Human Detection and Counting for application in Smart Surveillance System. *arXiv preprint arXiv:1303.0633* 4(2), 167–172 (2013)
18. Ojala, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
19. Ondřej, J., Pettré, J., Olivier, A.H., Donikian, S.: A synthetic-vision based steering approach for crowd simulation. *ACM Transactions on Graphics* 29(4), 1 (Jul 2010)
20. Ozturk, O., Yamasaki, T., Aizawa, K.: Tracking of humans and estimation of body/head orientation from top-view single camera for visual focus of attention analysis. *Computer Vision Workshops (ICCV Workshops)* pp. 1020–1027 (Sep 2009)

21. Pelechano, N., Stocker, C.: Being a part of the crowd: towards validating VR crowds using presence. Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems (Aamas), 12–16 (2008)
22. Rivalcoba, I.J., Rudomín, I.: Segmentación de peatones a partir de vistas aéreas. In: Research in Computing Science. vol. 62, pp. 129–230 (2013)
23. Szymanczyk, O., Dickinson, P., Duckett, T.: From Individual Characters to Large Crowds : Augmenting the Believability of Open-World Games through Exploring Social Emotion in Pedestrian Groups. In: Think Design Play: DiGRA Conference (2011)
24. Treuille, A., Cooper, S., Popović, Z.: Continuum crowds. ACM Transactions on Graphics 25(3), 1160 (Jul 2006)
25. Tuzel, O., Porikli, F., Meer, P.: Human Detection via Classification on Riemannian Manifolds. 2007 IEEE Conference on Computer Vision and Pattern Recognition pp. 1–8 (Jun 2007)
26. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. International Conference on Computer Vision 63(2), 153–161 (2003)
27. Wang, Y., Dubey, R., Magnenat-Thalmann, N., Thalmann, D.: An immersive multi-agent system for interactive applications. The Visual Computer 29(5), 323–332 (Jun 2012)

Procesos de decisión de Markov y microescenarios para navegación y evasión de colisiones para multitudes

Sergio Ruiz¹, Benjamín Hernández²

¹ Tecnológico de Monterrey,
Campus Ciudad de México, México

² Barcelona Supercomputing Center,
Barcelona, España

sergio.ruiz.loza@itesm.mx, benjamin.hernandez@bsc.es

Resumen. La simulación en tiempo real de multitudes ha sido relevante en aplicaciones relacionadas a la planificación urbana, evacuaciones, entrenamiento para el control de multitudes y entretenimiento. Problemas fundamentales como la navegación de agentes y evasión de colisiones deben ser resueltos eficientemente para obtener una simulación interactiva. Este trabajo propone un método integrado que hace uso de Procesos de Decisión de Markov para el cálculo de trayectorias óptimas que los agentes deberán seguir y de Microescenarios, técnica geométrica basada en curvas Bézier, que permite la evasión de colisiones entre los agentes y objetos dinámicos. Se reportan resultados usando estas técnicas para la simulación interactiva de multitudes en ambientes virtuales.

Palabras clave: Procesos de decisión de Markov, simulación de multitudes.

1. Introducción

Los seres humanos constituyen una de las especies más complejas en términos físicos y conductuales [12,33], en consecuencia la simulación de multitudes representa un reto considerable, y aunque los primeros métodos³ para lograr dicha simulación surgieron a mediados de los noventa [33], la alta demanda de recursos computacionales necesarios para la simulación de multitudes en tiempo real, sigue motivando el desarrollo de algoritmos novedosos de alto rendimiento.

Por otro lado, la imposibilidad de aplicar el método científico a ciertos subconjuntos del fenómeno de las multitudes, aunado a los costos prohibitivos de logística y económicos relacionados a la dirección de eventos experimentales, hace inminente el uso de una alternativa virtual: una simulación.

La componente principal de toda simulación de multitudes en tiempo real es la síntesis de comportamiento; para que una simulación de multitudes sea precisa es necesario que reproduzca comportamientos humanos individuales y grupales, además los algoritmos que sintetizan estos comportamientos deben

³ Mientras que el primer análisis conductual realizado por Le Bon [14] data de 1896.

estar optimizados de tal forma que funcionen en tiempo real. Existen dos problemas fundamentales a resolver para reproducir comportamientos individuales y grupales de forma exitosa: evasión de colisiones y navegación. Este trabajo presenta tres contribuciones: primero, el uso de Procesos de decisión de Markov (MDP, por sus siglas en inglés) para la navegación de multitudes en un ambiente virtual, segundo el uso de microescenarios, una técnica geométrica basada en curvas Bézier, para la evasión de colisiones y finalmente se presentan resultados de rendimiento de los algoritmos que han sido desarrollados usando programación paralela en el procesador gráfico (GPU por sus siglas en inglés) para obtener rendimiento en tiempo real.

2. Trabajo previo

2.1. Evasión de colisiones

La evasión de colisiones se refiere a los movimientos anticipados de un agente para evitar colisionar con otros agentes presentes. Los métodos principales para la evasión de colisiones son bandada [25], fuerzas sociales [10] y obstáculos con velocidades recíprocas [4] con sus respectivas extensiones.

Bajo condiciones no óptimas, conforme el número de agentes (denotado por N) crece, la complejidad del algoritmo es $O(N^2)$, es decir, la posición de cada agente debe compararse con cada una de las demás. Los trabajos que se citan a continuación abordan el problema de reducir esta complejidad mediante tres enfoques, según Azma et al.[21]: el primero es adaptar la técnica al modelo de programación del GPU, el segundo enfoque es reducir el orden de complejidad del algoritmo de evasión de colisiones y finalmente el tercero es adoptar una técnica de aprendizaje máquina.

Para reducir la complejidad de los algoritmos de evasión de colisiones, Bonner y Kelley [7] proponen un método jerárquico basado en aproximaciones sucesivas esféricas (SSA por sus siglas en inglés), en el cual una jerarquía de reglas es usada heurísticamente para obtener rutas libres de colisiones en un ambiente estructurado. Por otro lado, Li et al. [16] proponen la transformación del problema de evasión de colisiones de 3D a 2D mediante la proyección de secciones de la escena a distancias predefinidas, para después hacer pruebas de colisiones en cada una de las sub-secciones. En este sentido, Bing et al. [5] propone un algoritmo en GPU basado en la división del espacio y utiliza elipses acotantes para reducir la complejidad de las pruebas de colisión. En otro método para reducir la complejidad inherente en los algoritmos de evasión de colisiones, Reynolds [25,24] implementa una cuadrícula que particiona el espacio navegable, de esta forma reduce el espacio de búsqueda de colisiones a nivel cuadrícula y lo simplifica aun más definiendo un radio de búsqueda por cada agente.

Aunque no es adecuada para multitudes a gran escala, debido a sus altos requerimientos computacionales y su enfoque en evasión de colisiones del tipo vehículo-peaton, Fernández et al. [19] proponen una base de conocimiento (lo cual implica una etapa de entrenamiento y depuración), una etapa de percepción, planeación y ejecución de una arquitectura de control para la evasión de

colisiones. Otra contribución relevante en el área de Aprendizaje Máquina es de Li et al. [15]. Los autores proponen una base dinámica de reglas, un motor de inferencia y una base de conocimiento para la evasión de colisiones entre agentes que representan embarcaciones.

Van den Berg et al. [3] introducen el concepto de “evasión de colisiones óptimas y recíprocas” (ORCA por sus siglas en inglés), extendiendo su método “obstáculos con velocidades recíprocas” (RVO por sus siglas en inglés) [4]. Mientras que Guy et al. [9] proponen una implementación paralela en CPU del método RVO original.

2.2. Navegación

La navegación es la habilidad de los agentes para evitar colisiones efectivamente contra los objetos de una escena mientras se dirigen hacia una meta. Estos obstáculos pueden ser dinámicos o estáticos, pero el criterio que los califica como tales es el hecho de que no evitan colisiones, o en otras palabras, no presentan comportamiento reactivo.

Los algoritmos clásicos de planeación de rutas como A^* o Dijkstra han sido utilizados intensivamente en video juegos cuando los obstáculos son fijos y el número de agentes permite al CPU efectuar la simulación en tiempo real. Se han propuesto alternativas a A^* para reducir el tiempo de planeación de la ruta mediante la búsqueda incremental y heurística [31,13], o mediante la búsqueda rápida de una solución subóptima usando restricciones imprecisas al inicio y conforme avanza el tiempo, se ajustan estas restricciones para encontrar una ruta probablemente óptima para el periodo de tiempo dado [17,18].

Para simulaciones complejas en ambientes dinámicos, varios métodos modelan y resuelven la navegación de multitudes: partición basada en celdas, mallas de navegación o áreas navegables.

Los métodos de partición consisten en dividir el espacio en celdas y mediante el uso diferentes reglas o modelos matemáticos, los agentes pueden encontrar su ruta hacia una meta en la presencia de obstáculos. Además de A^* , otro ejemplo lo encontramos en los autómatas celulares [6,35,36]. Los autómatas celulares resuelven el problema de evasión de colisiones y navegación en el mismo algoritmo, sin embargo carecen de control y separación de los individuos pues todos deben seguir el flujo y dirección mayoritarios hacia una meta. Sarmady et al. [29] proponen el uso de un autómata celular con una partición fina para mejorar el realismo, sin embargo rutas poco realistas siguen estando presentes.

Otra alternativa a la navegación es la definición de áreas navegables como alternativa a las particiones estructuradas basadas en celdas. Pettré et al. [22] usa diagramas de Voronoi para descomponer las áreas navegables en celdas cilíndricas empalmadas. Dichas celdas generan un grafo de navegación que permite la navegación de peatones, sin embargo no soporta escenarios dinámicos. Otra alternativa a las celdas cilíndricas es el uso de mapas para codificar áreas navegables [26], mientras que Treuille et al. [34] extienden el uso de imágenes para codificar información de planeación y comportamiento.

3. Procesos de decisión de Markov para navegación

Batty [2] define la navegación como un producto de la aleatoriedad, geometría, intenciones económicas y preferencias sociales, aun así concuerda en que la navegación humana puede ser modelada con un MDP y una componente aleatoria, además las restricciones geométricas del ambiente otorgan estructura a los movimientos. El uso de MDPs para calcular trayectorias libres de colisiones no es un concepto nuevo, se ha usado esencialmente en robótica donde el problema involucra un sólo agente o robot moviéndose en un ambiente total o parcialmente observable [8,20,32]. Sin embargo para el caso de la simulación de multitudes en tiempo real no hay mucho trabajo desarrollado. Banerjee et al. [1] ha propuesto un método de navegación de multitudes usando procesos de decisión semi-markovianos; en contraste nuestra solución propone:

- Un método de una sola capa, de esta forma se reduce el uso de memoria.
- Propone un enfoque reactivo para cambios inesperados en el ambiente, dicho enfoque simplifica las estructuras involucradas en el cálculo de adaptación de rutas y mejora la velocidad de cálculo.
- Nuestro método no está limitado por el número de objetos dinámicos presentes en la simulación, pues no agrega capas por cada obstáculo presente.
- Nuestro método calcula el MDP y la evasión de colisiones usando cómputo en el procesador gráfico (GPU por sus siglas en inglés).
- Estas mejoras permiten el uso del algoritmo en ambientes virtuales habitados con personajes detallados.

3.1. Modelado del problema

Un MDP es una tupla $M = \langle S, A, T, R \rangle$, donde S es un conjunto finito de estados, A es un conjunto finito de acciones, T es una función de transición $T(s, a, s')$ y R es la función de recompensa $R(s)$. Una solución que indica lo que un agente debe hacer dado un estado es una política $\Pi(s)$.

Considerando un escenario para el cual las posiciones de los obstáculos estáticos son determinadas previo a la simulación, un MDP acotado y totalmente observable puede ser evaluado para determinar las direcciones óptimas de navegación de grupos de agentes enmarcados en celdas, dado que la representación 2D del escenario está regularmente particionada en estas celdas, en este caso, el conjunto de estados finitos S del MDP. El conjunto de dichas direcciones óptimas es la política óptima (Π^*) (Fig. 1(c)).

Nuestro escenario base —arbitrario—consiste de una cuadrícula de 3x4, que representa una distribución recompensa-penalidad para un agente (Fig. 1(a)).

Se han definido ocho posibles movimientos como el conjunto de acciones que cada agente podrá realizar a cada paso de tiempo para llegar a su meta (Fig. 1(b)).

Establecer que un agente se moverá a lo largo de la mejor ruta hacia la meta, implica el uso de una función de utilidad. En este caso, $U_h([s_0, s_1, \dots, s_n]) = R(s_0 + \gamma R(s_1) + \gamma^2 R(s_2)) + \dots + \gamma^n R(s_n)$ representa la preferencia de un agente para

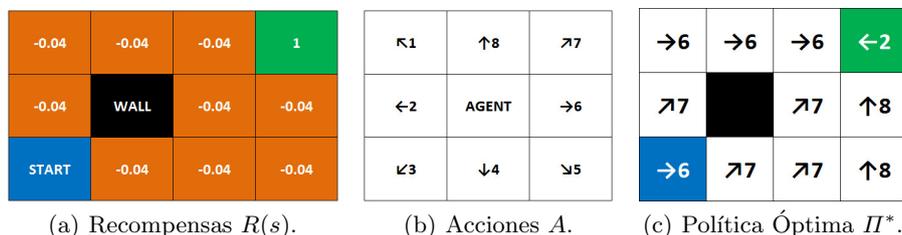


Fig. 1. Modelo Base de MDP.

las recompensas actuales y futuras que corresponden a los estados del conjunto $\{s_0, \dots, s_n\}$, donde el factor de descuento γ es un número en el rango $[0, 1]$. Para el modelo base de la Figura 1, a la meta se le asignó una recompensa de 1.0, a la pared una penalidad de -10.0 y a la celda inicial la misma penalidad que a las otras celdas: -0.04.

La observación de este modelo revela que varias soluciones o políticas son posibles para alcanzar la celda con la recompensa máxima (1.0), la cual representa una salida y es una meta para el agente.

Una política óptima Π^* , aquella que maximiza la función de recompensa R dada una utilidad óptima esperada de acciones futuras en un horizonte infinito, será calculada en un tiempo $t > 0$, usando la Ecuación 1.

$$\begin{aligned}
 \Pi_t^*(s) &= \operatorname{argmax}_a Q_t(s, a) \\
 Q_t(s, a) &= R(s, a) + \gamma \sum_{j=0}^5 T_{sj}^a V_{t-1}(j) \\
 V_t(s) &= Q_t(s, \Pi^*(s)) \\
 V_0(s) &= 0
 \end{aligned}
 \tag{1}$$

Donde $Q_t(s, a)$ representa el valor de elegir la acción a —en este caso direcciones de movimiento—de la celda s ; $V_t(s)$ representa el valor de la celda s en el tiempo t ; $\gamma \in [0, 1]$ es el factor de descuento de la recompensa futura y T_{sj}^a es la función de transición, tomando en cuenta la probabilidad con la que un agente elige el estado j desde el estado s mediante la acción a .

La evaluación de la ecuación 1 para $t = 1$, modificará los valores de $V_0(s) = 0$ a $V_1(s), \forall s \in \{0, \dots, n\}$, y cada evaluación sucesiva modificará marginalmente estos valores. Si después de la iteración x resulta que $V_{x-1}(s) = V_x(s) \forall s \in \{0, \dots, n\}$ entonces la política ha convergido a su valor óptimo.

El cálculo de la política óptima (Ecuación 1) se lleva a cabo mediante la técnica iteración de valor en el GPU como se describe en un artículo previo [28].

Un escenario de prueba se muestra en la Figura 2. Note que como resultado se obtiene un arreglo de direcciones (Fig. 2(a)) que de ser usado en la etapa de visualización, los agentes se moverían siguiendo posiciones discretas de una forma poca realista. Para evitar lo anterior, se utiliza el algoritmo recursivo de Shao y Zhou [30] para obtener curvas de Bézier y suavizar las trayectorias. La Figura 2(b) muestra dos rutas suavizadas con dicha técnica.

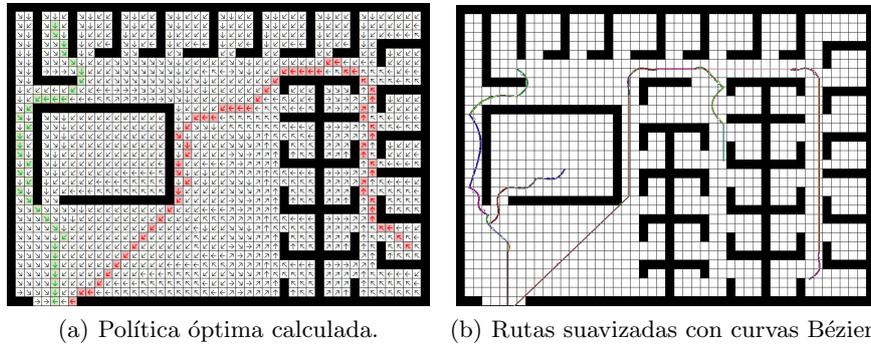


Fig. 2. Solución de un escenario de prueba.

4. Microescenarios

Una vez que a un agente se le ha asignado una ruta estática, libre de colisiones y suavizada, llegará a su meta si no se registran cambios en el escenario (Fig. 3(a)), sin embargo este caso es poco probable por dos razones:

- i) La simulación de multitudes implica que los agentes compitan para ocupar el espacio navegable al momento de cada uno de ellos intenta llegar a su meta.
- ii) No todos los escenarios son estáticos, por ejemplo, el caso de una simulación de evacuación en un terremoto.

Cuando los escenarios no son estáticos o es necesario que los agentes eviten colisiones con otros (Fig. 3(b)), proponemos el uso de micro escenarios para adaptar interactivamente las trayectorias previamente calculadas de los agentes.

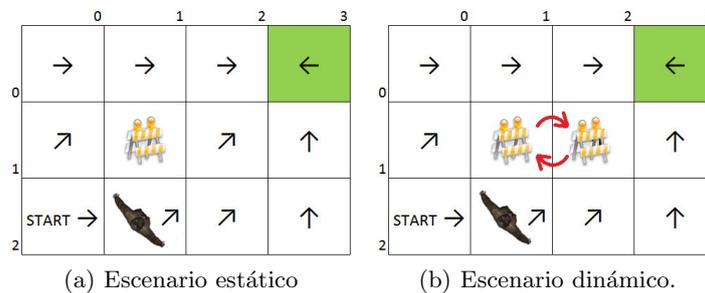


Fig. 3. Problema de Evasión de colisiones.

Un micro escenario describe las direcciones posibles que un agente puede tomar para evitar cuerpos en movimiento (otros agentes u obstáculos). Dependiendo del número y posición de los objetos en movimiento, se puede seleccionar un micro escenario específico utilizado para recalcularse en tiempo de ejecución, un segmento pequeño de la trayectoria previamente pre-calculada con el MDP. La Figura 4 muestra un microescenario de 3x3 celdas, i.e. con radio 1, los primeros dos microescenarios representan el caso cuando no hay cuerpos en movimiento cerca del agente, con dos diferentes posibilidades para una meta dada. El último microescenario, representa el caso cuando un agente es rodeado por cuerpos en movimiento.

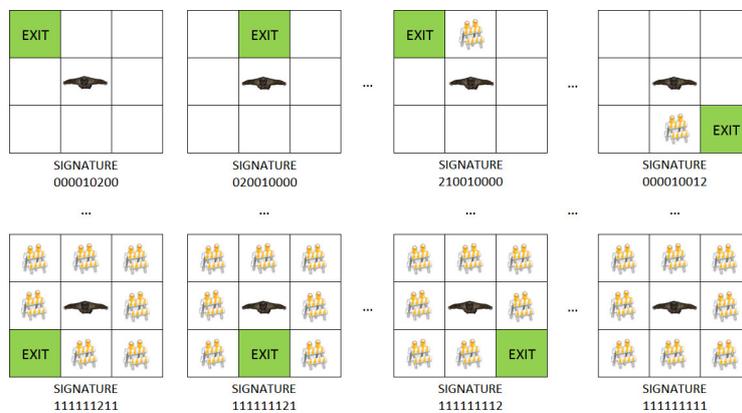


Fig. 4. Microescenarios posibles de radio 1 y sus respectivas firmas.

De esta forma el algoritmo para utilizar los microescenarios se describe a continuación:

1. En tiempo de pre procesamiento, se obtienen las trayectorias libres de colisiones suavizadas a partir del cálculo de MDP y curvas de Bézier.
2. También en etapa de pre procesamiento, se resuelve cada microescenario posible considerando que el agente está localizado en la celda del centro. Como resultado de este pre proceso se obtienen dos arreglos de información:
 - i) Un arreglo único de curvas Bézier (etiquetadas como “Bézier.” en la Figura 5) para ser empleado en la simulación.
 - ii) Un arreglo de “firmas” (etiquetado como “Signatures.” en la Figura 5) que relaciona cada micro escenario al conjunto de curvas Bézier (“Bézier-Signature Indirect Index.” en la Figura 5). Cabe mencionar que para estas firmas, es indiferente si los cuerpos son obstáculos o son otros agentes.
3. En tiempo de ejecución, el algoritmo 1 es ejecutado en el GPU, entonces los agentes se desplazarán en sus trayectorias originalmente asignadas o ajustarán su ruta utilizando el microescenario si la próxima celda está ocupada.

```

for agent = 1 → N do
  bezierIndex = signature[agent]
  if bezParam[agent] == 0 then
    if occupancy[agentCell[agent]] == 1 then
      aSignature = readMicroScenario( agentCell[agent] )
      bezierIndex = BezierSignature[aSignature]
      signature[agent] = bezierIndex
    end if
  end if
  position[agent] = evalBezier( bezParam[agent], bezierIndex )
  bezParam[agent] += bezInc[agent]
  if bezParam[agent] > 1 then
    bezParam[agent] = 0
  end if
  occupancy[agentCell[agent]] = 0
  agentCell[agent] = positionToCell( position[agent] )
  occupancy[agentCell[agent]] = 1
end for

```

Algorithm 1: Algoritmo para modificar trayectorias a partir de los microescenarios.

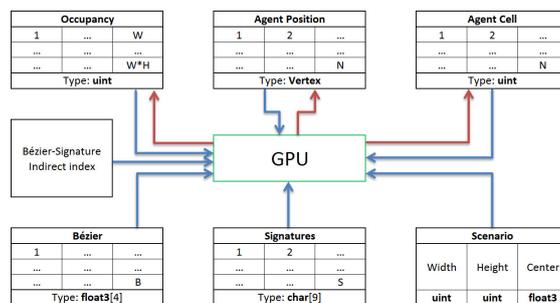


Fig. 5. Arreglos utilizados en el algoritmo propuesto.

5. Resultados

Se han diseñado tres casos para evaluar nuestro método, el primero de ellos consiste en evaluar la rapidez de cálculo de la política óptima. Se hicieron comparaciones entre la implementación de MDP en un CPU Intel móvil i7 a 1.87 GHz usando un solo hilo de ejecución y un GPU Geforce 445 móvil para el escenario base de la Figura 2. Los resultados muestran que la política óptima se obtiene en 184 iteraciones para las cuales el GPU tarda 4.1 segundos (Tabla 1) mientras que el CPU para 10 iteraciones tarda 1500 segundos (Tabla 2) si el promedio de tiempo de cálculo se mantiene constante, la versión de CPU tardaría alrededor de 7.6 horas en obtener la política óptima.

Adicionalmente la Figura 8 muestra diferentes visualizaciones de una multitud en el escenario base. La descripción del algoritmo utilizado para visualizar la multitud puede encontrarse en [27].

Tabla 1. Rendimiento en el GPU

GPU	NVIDIA® GeForce™ GT445M
columnas	48
filas	33
Total de celdas	1,584
Iteraciones	184
Política óptima en	4.1s

Tabla 2. Rendimiento en el CPU

CPU	Intel® Core™ i7@1.87GHz
columnas	48
filas	33
Total de celdas	1,584
Iteraciones	10
Política óptima en	1500s

Para la segunda prueba (Fig. 6) presentamos la solución de un escenario de 400 celdas (20 filas y 20 columnas) de las cuales 164 representan obstáculos, con el fin de comparar el desempeño y requerimientos de memoria de los algoritmos A^* (Figura 6(a) y Tabla 3, utilizando la implementación de Heyes [11]) y MDP (Figura 6(b) y Tabla 4) al encontrar rutas óptimas de navegación. Encontramos que el tiempo de ejecución de A^* depende del número de agentes a procesar, mientras que la solución con MDPs se realiza en tiempo constante independientemente del número de agentes. Para procesar las 236 celdas libres en el escenario de prueba, el método basado en MDPs tarda 0.02079 segundos en promedio por agente, mientras que A^* toma 0.078 segundos. Notamos también que para procesar estas 236 celdas, A^* requiere 6,580 bytes en promedio por agente para almacenar listas de nodos, mientras que el método basado en MDPs requiere 1,803 bytes en promedio por agente para almacenar Π_t^* , Q_t y V_t en vectores [28].

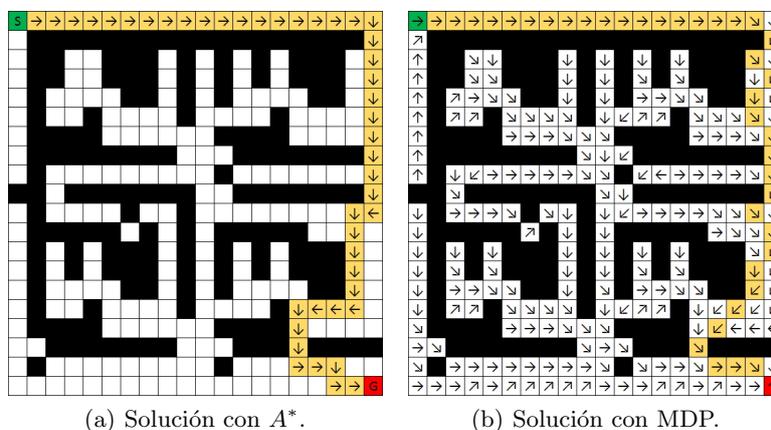


Fig. 6. Escenario de prueba para comparar A^* y MDP.

Cuando un escenario se modela mediante un MDP para navegación, definiremos i como el número de iteraciones, a el número de acciones, c el número total de celdas y t como el número máximo de transiciones. Entonces la complejidad

Tabla 3. Desempeño de A^* .

Método	A^*
CPU	Intel i7-4800MQ@2.7GHz
Opciones	4
Salidas procesables	1
Tiempo de solución	0.078s
Tiempo 236 agentes	18.408s
Pasos de solución	46
Pasos en búsqueda	235
Agentes procesados	1
Memoria total	6.42 KB
Memoria por agente	6.42 KB

Tabla 4. Desempeño de MDP.

Método	MDP
GPU	NVIDIA GeForce GTX780M
Opciones	8
Salidas procesables	235
Tiempo de solución	4.907s
Tiempo 236 agentes	4.907s
Pasos de solución	39
Iteraciones	167
Agentes procesados	236
Memoria total	415.63 KB
Memoria por agente	1.76 KB

para estimar las rutas navegables es $O(iact)$. Como se muestra en la Figura 7, este proceso se lleva a cabo eficientemente en el GPU.

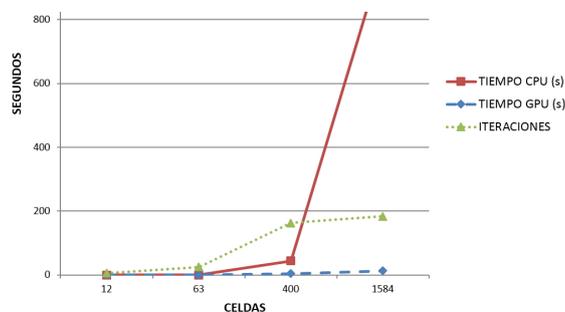


Fig. 7. Tiempo de solución de MDP para GPU y CPU.

La tercera prueba consistió en simular una multitud masiva de Lemmings y de seres humanos. Este caso fue útil para probar los microescenarios, i.e. la evasión de colisiones. De esta forma se comprobó que los agentes efectivamente evadían colisiones entre ellos, además presentaban comportamiento reactivo al momento de agregar interactivamente obstáculos. En este caso se usó un GPU Geforce GTX 560M.

Resultados numéricos de esta simulación se muestran en la figura 9(a). Cuando los personajes solamente siguen sus trayectorias, sin experimentar obstáculos inesperados, el método se ejecuta en MIN TIME (entre 1 y 2 milisegundos para una multitud de entre 128 y 8192 personajes). Por otro lado, para escenarios congestionados la simulación toma MAX TIME (de 6 a 13 milisegundos).

A partir de estas mediciones, otra ventaja de usar MDP y microescenarios es que en tiempo de ejecución, dado que la solución es independiente del tamaño y complejidad del escenario, el método es razonablemente predecible (como lo implica el número de agentes y celdas) y estable (Fig. 9(b)). A partir del rendimiento obtenido y del comportamiento de MAX TIME, se puede predecir que si se tiene memoria y poder de cómputo suficientes en un GPU, se lograrían simular

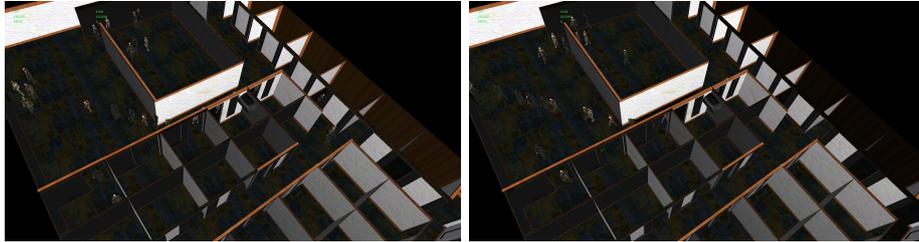


Fig. 8. Visualización de una evacuación.

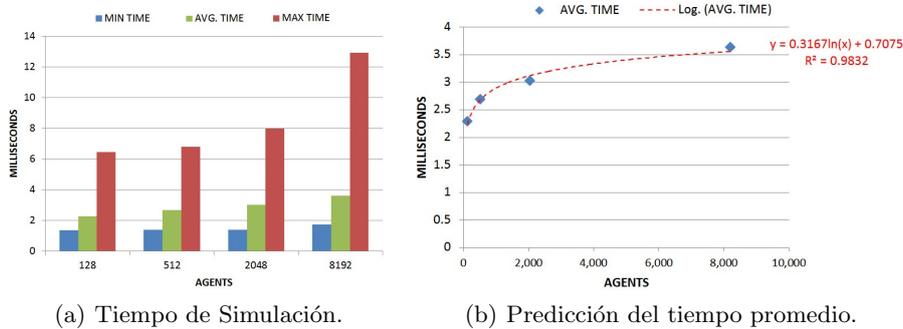


Fig. 9. Rendimiento general del algoritmo MDP con microescenarios.

1,000,000 de personajes en 806.37 milisegundos en escenarios congestionados, lo cual es un punto a destacar de nuestro algoritmo.

Cuando los microescenarios resueltos durante el preprocesamiento se utilizan para evadir colisiones, la complejidad del algoritmo resulta en la de la lectura de las celdas vecinas aunado a la complejidad de la búsqueda en una tabla de firmas con la correspondiente obtención de la curva Bézier para el movimiento continuo para cada agente. Esto es, para un número N de agentes: $O(10N)$, un costo lineal como confirma la Figura 9(a).

La Figura 10 muestra diferentes visualizaciones de las multitudes de Lemmings (Fig. 10(a)) y de personas (Fig. 10(c)) de este experimento. Note que en ambas multitudes los personajes cambian de dirección para evadir colisiones con obstáculos (Fig. 10(b)) o con otros agentes (Fig. 10(d)).

Finalmente invitamos al lector a que revise el video de la simulación que se encuentra en la siguiente liga: <http://tinyurl.com/p3aqxm6>

6. Conclusiones y trabajo futuro

El algoritmo propuesto permite adaptar un MDP no sólo para resolver el problema de navegación sino también el problema de evasión de colisiones. Además

resaltando que Reyes y Sucar [23] notaron que “...el problema del formalismo de los MDPs es que el espacio de estados crece exponencialmente con el número de variables, y los métodos de inferencia crecen con el número de acciones, entonces, en problemas grandes, los MDPs se tornan impracticos e ineficientes...”, nuestra implementación en el procesador gráfico o GPU permite aplicar el algoritmo en simulaciones de multitudes en tiempo real con rendimiento suficiente para llevar a cabo la visualización tridimensional de los resultados. Además, conforme se disponga de mejor hardware, problemas previamente inabordables serán reexaminados para verificar si su complejidad puede ser sujeto de análisis y solución como lo es el caso de simulación de evacuaciones masivas. Otro punto relevante que deberá explorarse es la modificación de las variables de los MDPs, una modificación correcta nos permitirá representar simulaciones más complejas. Por otro lado, es deseable desacoplar la etapa de simulación con la etapa de visualización mediante la utilización de dos GPUs, de esta forma se mejoraría aun más el rendimiento del método. Finalmente la integración y comparación del algoritmo propuesto con otros modelos conductuales permitirían otras áreas de desarrollo como por ejemplo simulaciones sociales.

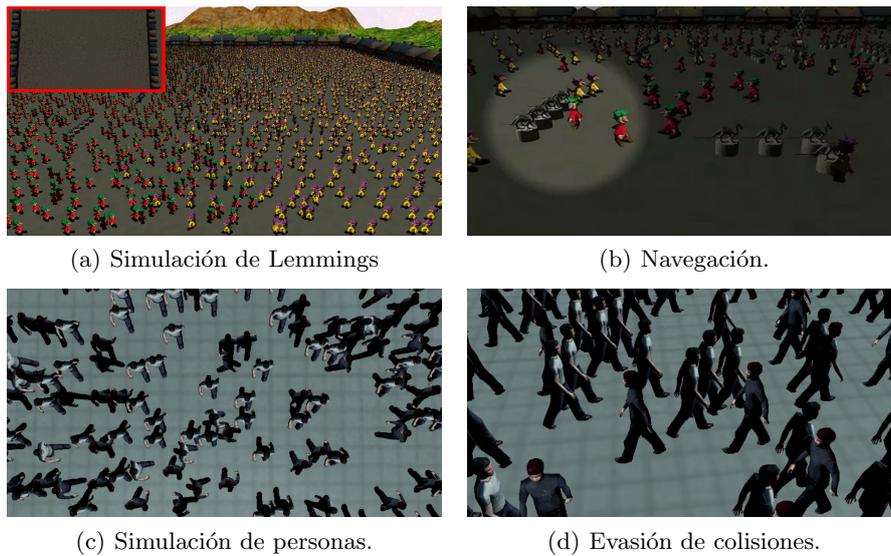


Fig. 10. Multitudes de Lemmings y personas.

Referencias

1. Banerjee, B., Abukmail, A., Kraemer, L.: Advancing the layered approach to agent-based crowd simulation. In: Proceedings of the 22nd Workshop on Principles of

- Advanced and Distributed Simulation. pp. 185–192. IEEE, IEEE Computer Society (2008)
2. Batty, M.: Agent-based pedestrian modelling. In: Centre for Advanced Spatial Analysis. Working Paper Series, University College London (2003)
 3. van den Berg, J., Guy, S.J., Lin, M.C., Manocha, D.: Reciprocal n-body collision avoidance. In: International Symposium on Robotics Research (ISRR). vol. 70, pp. 3–19. IFRR (May 2011)
 4. van den Berg, J., Lin, M., Manocha, D.: Reciprocal velocity obstacles for real-time multi-agent navigation. In: International Conference on Robotics and Automation (ICRA). pp. 1928–1935. IEEE (May 2008)
 5. Bing, H., Yangzihao, W., Jia, Z.: An improved method of continuous collision detection using ellipsoids. In: Computer-Aided Industrial Design & Conceptual Design. IEEE (Nov 2009)
 6. Blue, V.J., Embrechts, M.J., Adler, J.L.: Cellular automata modeling of pedestrian movements. In: Systems, Man, and Cybernetics. vol. 3, pp. 2320–2323. IEEE (Oct 1997)
 7. Bonner, S., Kelley, R.B.: A novel representation for planning 3-d collision-free paths. In: Systems, Man, and Cybernetics. vol. 20, pp. 1337–1351. IEEE (Nov 1990)
 8. Foka, A.F., Trahanias, P.E.: Real-time hierarchical pomdps for autonomous robot navigation. *Robotics and Autonomous Systems* 55(7), 561–571 (2007)
 9. Guy, S.J., Chhugani, J., Kim, C., Satish, N., Lin, M., Manocha, D., Dubey, P.: Clear path: highly parallel collision avoidance for multi-agent simulation. In: Symposium on Computer Animation. pp. 177–187. ACM (Aug 2009)
 10. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. *PHYSICAL REVIEW E* 51, 4282 (1995)
 11. Heyes-Jones, J.: Implementation of the a* algorithm in c++. Web Page (Mar 2014), [justinhj/astar-algorithm-cpp](https://github.com/justinhj/astar-algorithm-cpp). GitHub. Retrieved May 6, 2014, from <https://github.com/justinhj/astar-algorithm-cpp>
 12. Initiative, T.H.O.: Brains - the smithsonian institution's human origins program (2011), <http://humanorigins.si.edu/human-characteristics/brains>
 13. Koenig, S., Likhachev, M.: D*lite. In: Eighteenth national conference on Artificial intelligence. pp. 476–483. American Association for Artificial Intelligence, Menlo Park, CA, USA (2002)
 14. Le Bon, G.: *The Crowd: A Study of the Popular Mind*. The Criminology Series, The Macmillan Co. (1896)
 15. Li, L., Yang, S., Zhou, W., Chen, G.: Mechanism for constructing the dynamic collision avoidance knowledge-base by machine learning. In: Proceedings of the 2010 International Conference on Manufacturing Automation. pp. 279–285. ICMA '10, IEEE Computer Society, Washington, DC, USA (2010)
 16. Li, X., Zhong, Z., Lu, Z.: Collision detection algorithm based on slice projection. In: International Conference on Mechatronics and Automation. IEEE (Aug 2009)
 17. Likhachev, M., Ferguson, D., Gordon, G., Stentz, A.T., Thrun, S.: Anytime dynamic a*: An anytime, replanning algorithm. In: Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS) (June 2005)
 18. Likhachev, M., Gordon, G.J., Thrun, S.: Ara*: Anytime a* with provable bounds on sub-optimality. In: NIPS (2003)
 19. Llorca, D.F., Milanés, V., Alonso, I.P., Gavilán, M., Daza, I.G., Pérez, J., Sotelo, M.A.: Autonomous pedestrian collision avoidance using a fuzzy steering controller. In: Intelligent Transportation Systems. IEEE (Jun 2011)

20. Mausam, Weld, D.S.: Solving concurrent markov decision processes. In: Proceedings of the 19th National Conference on Artificial Intelligence. pp. 716–722. AAAI'04, AAAI Press (2004)
21. Mustapha, N.A.B., Bade, A.B., Kari, S.: A review of collision avoidance technique for crowd simulation. In: International Conference on Information and Multimedia Technology. IEEE (2009)
22. Pettré, J.: Motion planning and autonomy for virtual humans: Part 4. case study 2. part i. design and simulation of virtual crowds. In: SIGGRAPH. ACM (2008)
23. Reyes, A., Sucar, L.: An operation auxiliary system for power plants based on decision-theoretic planning. In: International Conference on Intelligent Systems Application to Power Systems (2005)
24. Reynolds, C.: Big fast crowds on ps3. In: Sandbox Symposium. ACM (Jul 2006)
25. Reynolds, C.W.: Flocks, herds, and schools: A distributed behavioral model. In: SIGGRAPH (1987)
26. Rudomin, I., Perez, F., Millán, E.: Groups and crowds with behaviors specified in the environment. In: Intelligent Virtual Environments and Virtual Agents Proceedings. ITESM Campus Ciudad de Mexico (2004)
27. Ruiz, S., Hernández, B., Alvarado, A., Rudomín, I.: Reducing memory requirements for diverse animated crowds. In: Proceedings of Motion on Games. pp. 55:77–55:86. MIG '13, ACM, New York, NY, USA (2013)
28. Ruiz, S., Hernández, B., Rudomín, I.: A gpu implementation of markov decision process for path planning. In: The 5th International Supercomputing Conference in Mexico (ISUM 2014) (2014)
29. Sarmady, S., Haron, F., Talib, A.Z.: Simulating crowd movements using fine grid cellular automata. In: International Conference on Computer Modelling and Simulation. pp. 428–433. IEEE (2010)
30. Shao, L., Zhou, H.: Curve fitting with bézier cubics. In: Graphical Models and Image Processing. vol. 58, pp. 223–232 (1996)
31. Stentz, A.: The focussed d* algorithm for real-time replanning. In: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. pp. 1652–1659. IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995)
32. Sucar, L.: Parallel markov decision processes. In: Lucas, P., Gmez, J., Salmern, A. (eds.) Advances in Probabilistic Graphical Models. Studies in Fuzziness and Soft Computing, vol. 214, pp. 295–309. Springer Berlin Heidelberg (2007)
33. Thalmann, D., Musse, S.R.: Crowd Simulation. Springer (2007)
34. Treuille, A., Cooper, S., Popović, Z.: Continuum crowds. ACM Trans. Graph. 25(3), 1160–1168 (Jul 2006)
35. Zhang, S., Li, M., Li, F., Liu, A., Cai, D.: A simulation model of pedestrian flow based on geographical cellular automata. In: Geoinformatics. IEEE (Jun 2011)
36. Zhiqiang, K., Chongchong, Y., Li, T., Jingyan, W.: Simulation of evacuation based on multi-agent and cellular automaton. In: Mechatronic Science, Electric Engineering and Computer. IEEE (Aug 2011)

Sistema inteligente para procesamiento de imágenes en control de calidad basado en el modelo difuso *singleton* tipo 1

Pascual Noradino Montes Dorantes¹, Juan Pablo Nieto González¹,
Gerardo Maximiliano Méndez², Rolando Praga Alejo¹, Karla Lizette Guajardo Cosío¹

¹ Corporación Mexicana de Investigación en Materiales S.A. de C.V.,
Saltillo, Coahuila, México

² Posgrado en Ingeniería Mecatrónica, Instituto Tecnológico de Nuevo León,
Guadalupe, Nuevo León, México

pascualresearch@gmail.com; juan.nieto@comimsa.com;
gmm_paper@yahoo.com.mx; rolandopraga@comimsa.com; Karla.guajardo@comimsa.com

Resumen. En este artículo se presenta una metodología híbrida basada en un modelo difuso del tipo 1 en versión singleton usando un diseño factorial 2^k que optimiza el modelo del sistema experto y sirve para realizar inspección en línea. El método de diseño factorial proporciona la base de datos necesaria para realizar la creación de la base de reglas para el modelo difuso y también genera la base de datos para entrenar el sistema experto. El método propuesto ha sido validado en el proceso de verificación de parámetros dimensionales por medio de imágenes comparándolo con los modelos ANFIS y RBFN los cuales muestran mayores márgenes de error en la aproximación de la función que representa el sistema comparada con el modelo propuesto. Los resultados obtenidos muestran que el modelo tiene un excelente desempeño en la predicción y el control de la calidad del proceso industrial estudiado cuando se compara con técnicas de sistemas expertos similares (ANFIS, RBFN).

Palabras clave: T1-SFSL, sistema difuso, diseño factorial, reducción de reglas, inspección de calidad, procesamiento de imágenes, visión artificial.

1 Introducción

La inspección de calidad hoy en día es aún un proceso que depende de las capacidades humanas. Junto al desarrollo de la ciencia y tecnología de la computación los procesos de inspección comenzaron a cambiar de la visión humana a la visión artificial. Un sistema de visión artificial se puede definir como el ensamble de una fuente de luz, un dispositivo que captura la imagen y un software que se dedique a procesarla [1]. Por otro lado, la minería de datos es muy importante para estos sistemas ya que, se requiere procesar una gran cantidad información proveniente de diferentes bases de datos, las cuales contienen múltiples campos de información que

no es necesaria para el modelo. Para optimizar los sistemas es necesario elaborar una búsqueda de patrones para simplificar la complejidad [2], esta búsqueda reduce las demandas computacionales del sistema particularmente en la fase de adquisición del conocimiento.

Los sistemas de computación suave, sistemas inteligentes o sistemas expertos usan diferentes algoritmos para el manejo del conocimiento en el entrenamiento y la aproximación. Algunos de estos son: los algoritmos de Karnik-Mendel (KM), este algoritmo presenta complicaciones para generar una aproximación de la función que representa el proceso; debido a su naturaleza iterativa generalmente presenta retraso en el tiempo de respuesta [3]. Otros algoritmos comúnmente usados para el entrenamiento de los sistemas inteligentes son los siguientes: modelo de paso descendente (steepest descend) [4], mínimos cuadrados [4-5], modelo de gradiente descendente [5], agrupamiento (clustering) [5], estos presentan naturaleza iterativa que provee el ajuste necesario para reducir el error. Adicionalmente, las herramientas que se han usado principalmente para unir el Diseño de experimentos (DOE) con la computación suave y control de calidad son: superficie de respuesta [6], Metodología de Taguchi, diseño factorial (DF) [7] y el diseño central compuesto (DCC) [8].

Los modelos de computación suave, particularmente la lógica difusa y las redes neuronales o modelos híbridos, brindan la posibilidad de evaluar dentro del sistema una gama de posibilidades prácticamente infinita; estos sistemas realizan su evaluación en un universo continuo (difuso) en lugar de un universo discreto (crisp) [9]. Debido a que no son restrictivos, los sistemas difusos se pueden simplificar por medio de una reducción en el tamaño de su base de reglas. No se requiere probar todos los estados del sistema ya que la dinámica de este tipo de sistemas expertos es capaz de aproximar comportamientos a partir de un patrón, en este caso en particular, esto se realiza por medio de una interpolación para obtener una conversión de los valores de la especificación a píxeles y obtener el valor de los límites en píxeles para poder generar las reglas difusas, esto sucede debido a que no en todas las ocasiones se dispone de los recursos necesarios para la realización de pruebas físicas (en los límites de control) ya que el producir estos ajustes a las máquinas para volver al control del proceso. Los resultados utilizados para los consecuentes se pueden simular a partir de datos históricos por medio de la interpolación mencionada. Es suficiente contar con las observaciones de los límites de la especificación: límite inferior de control (LIC) $\forall x_i \in X \ i = 1, 2, \dots, n$, límite superior de control (LSC) $\forall x_i \in X \ i = 1, 2, \dots, n$. en caso de no conocer todos los datos, estos se pueden aproximar por medio de alguna técnica matemática de aproximación a partir de los datos de la especificación que serán convertidos a píxeles. e.g. mínimos cuadrados.

Los sistemas de visión artificial usados industrialmente, particularmente los sistemas de adquisición, supervisión y control (SCADA) necesitan evolucionar adaptarse para ser más confiables. Además se debe lidiar con múltiples factores que afectan su desempeño. Estos factores generan variaciones en la magnitud de los datos de las entradas [10]. Esta incertidumbre puede originarse en diferentes fuentes como las características de las materias primas, producción, maquinaria, programación de la producción, los sistemas de adquisición de datos (sensores) y del ambiente principalmente [11]. Los sistemas la adquisición de los datos constituyen una parte

fundamental para la evaluación, el monitoreo y el control de los procesos de manufactura, el cual es afectado continuamente por la incertidumbre [12]. Por otro lado el modelo de Taguchi toma en cuenta la incertidumbre y la clasifica en tres formas: interna, externa y de producción [13]. En base a estas dos perspectivas se puede definir la siguiente clasificación:

- Incertidumbre interna, este procede de las materias primas y la principal causa de esta clase de incertidumbre consta de dos partes: a) propiedades físicas de los materiales, b) propiedades químicas. Este apartado se refiere a como está constituida la formulación del material.
- Incertidumbre de producción, la cual se debe a la operación del proceso teniendo como componentes principales a los operadores y métodos de trabajo, maquinaria, mantenimiento, entre otras.
- Incertidumbre externa generalmente procede de la naturaleza entre los factores más comunes figuran: ambiente, temperatura, iluminación, humedad, radiación.

Los factores ambientales antes mencionados, contribuyen a que el desempeño de los operadores no sea adecuado, puesto que alteran el área de confort donde ellos se desenvuelven y provocan que ese desempeño no sea óptimo, ocasionando que los juicios que emiten y las decisiones que toman no sean acertadas. Los cambios (de diseño de ingeniería, ambientales, entre otros), también provocan incertidumbre en las tres modalidades presentadas en el párrafo anterior [10]. Algunos ejemplos pueden ser encontrados en la literatura acerca del tema de sistemas de procesamiento industrial de imágenes; estos sistemas son usados en diversos campos de la industria y la ciencia que con el paso del tiempo ha evolucionado de la mano de diferentes técnicas a tareas más complicadas pero aún siguen teniendo una serie de restricciones. e.g. identificación de placas, el sistema está restringido a un punto de referencia, se eliminó el número “0” (cero) y la vocal “o” para evitar confusión [14]; clasificación de hojas de tabaco [15] que en este caso la clasificación automática usa solo el 20 % del total de las clases ya que en estas se evalúan varias características, e.g. color, tamaño, forma, superficie y posición (esto es la evaluación de los tallos presentes en las hojas), adicionalmente estas características principales se sub-clasifican e.g. posición se divide en: salientes (X), cortos (C) y hoja sin tallo (B); color: limón (L), naranja (F) y rojo-café (R). Cada grupo de estos es a su vez subdividido en cuatro o cinco sub-grupos debido al tono de la hoja. Esto produce una combinatoria de $(3^2)^5 = 59049$ posibilidades debido a los subgrupos generados por el tono, el sistema se restringe a evaluar únicamente 24 de las 40 categorías principales debido a que es muy costoso (computacionalmente, matemáticamente y económicamente) para desarrollar un modelo completo.

En las aplicaciones de visión y procesamiento de imágenes existe la restricción de evaluar un pequeño intervalo de las características más comunes debido a que algunas son muy escasas o difíciles de presentarse, por otra parte, si se evaluaran todas las características presentes generarían un crecimiento exponencial en su tamaño, ya que cuando se usan modelos factoriales n^k se requiere tener toda la combinatoria del sistema, es decir, tener todos los posibles resultados. Por lo cual se utilizan técnicas

para reducir el modelo tales como el análisis de componentes principales (PCA). Se puede reducir el tamaño del sistema eliminando las combinaciones que puedan producir confusión tal como en el caso de la identificación de placas con el número 0 y la vocal o. Las limitaciones dentro del modelo de inspección visual provocan que se deba conservar parte del personal altamente entrenado para estudiar los casos que no puedan ser evaluados por el sistema [15]. En este caso se tiene un ejemplo de una muestra que se ubica justo en la frontera de la división entre dos grupos por lo tanto el sistema de clasificación automático reconoce dicha muestra como 50% perteneciente al grupo 1 y 50% perteneciente al grupo 2 quedando sin definición particular debido a que no existe esta clase particular dentro del sistema puesto que se eliminó y deberá ser estudiada por personal experto por fuera del sistema de clasificación automática.

El caso de estudio de este artículo se basa en un proceso de inspección para dimensionamiento basado en imágenes y se propone generar un modelo híbrido con diseño factorial y lógica difusa *singleton* del tipo 1 (T1 SFLS) el cual será validado contra modelos similares de sistemas expertos tales como: Red adaptativa de inferencia neurodifusa (ANFIS) designado en este caso como T1 ANFIS y Red de función de base radial (RBFN) para verificar su funcionamiento y su precisión. La sección 2 muestra un panorama general de la lógica difusa. En la sección 3 se presenta el modelo del diseño factorial. La sección 4 describe la metodología para crear el modelo híbrido. La sección 5 presenta el caso de aplicación. En la sección 6 se muestran los resultados y finalmente la sección 7 muestra las conclusiones.

2 Lógica Difusa

La lógica difusa se creó para estudiar las propiedades de los sujetos pertenecientes a un conjunto en el que cada sujeto debe pertenecer a algún conjunto en cierto grado. Para evaluar esta pertenencia [16] presenta un algoritmo matemático para crear aproximaciones por medio de iteraciones vectoriales, en este tipo de evaluación se establece un universo continuo con un número infinito de valores en lugar de uno discreto [17].

Para el diseño de aplicaciones con lógica difusa, se pueden utilizar datos lingüísticos en vez de números o datos crisp que son completamente deterministas [18-20] se basa en reglas condicionadas de la forma (1). Estas reglas son evaluadas por medio de la norma-t o la conorma-t para establecer una evaluación normalizada que se conoce como nivel de activación para cada regla, esta proporcionará una fracción de la evaluación final (2). Posteriormente se ensamblará la función de base difusa (FBF) representada por (3) la cual arrojará un valor difuso de salida que deberá transformarse a un dato crisp o a un valor lingüístico para obtener la aproximación final de la valoración del objeto de estudio,

$$\text{Regla } n: \text{ Si } X_i \text{ es } A_i \text{ y } \dots \text{ y } X_n \text{ es } A_n \text{ entonces } Y_n = B_n \quad (1)$$
$$i = 1, 2, \dots, n$$

donde: X_n representa las variables difusas de entrada, A_n representa la entrada crisp, Y_n representa la salida difusa de la regla y B_n representa la salida crisp,

$$\mu_{R \circ S}(x, z) = \sup_{y \in V} [\mu_R(x, y) \star \mu_S(y, z)] \quad (2)$$

donde: \star denota el producto y μ_R, μ_S representan las funciones de membresía de cada variable, y $\mu_{R \circ S}$ representa el grado de activación de la regla.

$$FBF = \frac{\sum_{j=1}^M \bar{z}^j (\prod_{i=1}^n \mu_i^j(x_i))}{\sum_{j=1}^M (\prod_{i=1}^n \mu_i^j(x_i))} \quad (3)$$

donde: \bar{z}^j es la salida de una regla en el espacio R , $\mu_i^j(x_i)$ representa la activación de la regla.

3. Diseño factorial

El diseño factorial [21] es un técnica del DOE, que fue desarrollada para el análisis de factores correlacionados, estos factores presentan limites dentro un universo y son denominados niveles, los cuales se diferencian en alto y bajo (Tabla. 1). Para el establecimiento del modelo general se produce una combinatoria por medio de permutaciones de los factores evaluados en ambos niveles para el caso del modelo 2^k , solo se produce una combinatoria de 4 posibilidades (vértices de la Fig. 1).

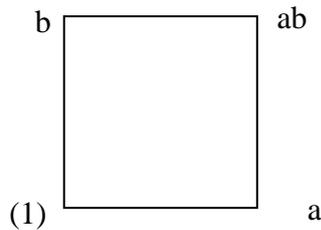


Fig. 1. Representación gráfica del DF 2^k , adaptado de [21].

Por otra parte, dentro de los límites de la especificación del producto se presentan los máximos y mínimos los cuales se establecerán como los niveles alto y bajo dentro de los tratamientos del DF, estos puntos se interpolan para crear las medias de los conjuntos difusos para las entradas del sistema. Los factores (para el caso del sistema difuso se usan las variables x_1 y x_2) se denominan **A** y **B** en el nivel alto de la especificación, para los niveles bajos de la especificación se omiten. Las combinaciones son descritas con letras, signos, símbolos o números. Los niveles bajos son omitidos con propósitos gráficos. E. g. **ab** representa las dos variables en nivel alto, **b** es usada cuando es combinado el nivel alto de **B** con el nivel bajo de **A**, **a** es usada cuando se combinan el nivel alto de **A** y el nivel bajo de **B**, **1** es usado cuando las dos variables se encuentran en el nivel bajo (Fig.1).

La simpleza del modelo de experimento factorial (2^k) se puede expresar matemáticamente como una función lineal multivariada de primer orden (4). Por ejemplo, la función multivariada dada por (5) es llamada el efecto principal debido a que representa las variables de control principales en el proceso, estas son llamadas **A**

y **B**. Los efectos principales mencionados se generan por medio de la combinatoria para crear una matriz que representa todas las combinaciones del modelo para el caso del DF/sistema experto solo se producen 4 permutaciones marcando las interacciones de las variables de control del proceso. Las combinaciones de los efectos son llamadas tratamientos y la matriz producida por la combinatoria y sus tratamientos son presentados en la tabla 2. Esta reducción ayuda a que el sistema trabaje en línea sin disminuir su precisión.

$$a_1x_1@a_2x_2@ \dots @a_nx_n = y \tag{4}$$

$$a_1x_1@a_2x_2 = y \tag{5}$$

donde: a_n , es un escalar (coeficiente de la variable independiente) y @ es un operador matemático, e.g. (suma, resta, multiplicación, división, entre otros).

Tabla 1. Representación simbólica del DF.

Variable		Efecto	
A	B		
-	-	Bajo	Bajo
+	-	Alto	Bajo
-	+	Bajo	Alto
+	+	Alto	Alto

Tabla 2. Entradas del DF, efectos y sus tratamientos.

Variable		Efecto		Tratamiento
A	B	Bajo	Bajo	A Bajo, B Alto
A	B	Alto	Bajo	A Alto, B Bajo
A	B	Bajo	Alto	A Bajo, B Bajo
A	B	Alto	Alto	A Alto, B Alto

4. Metodología

El proceso metodológico que requiere el sistema de inspección de imágenes necesita de un proceso de adaptación para generar la evaluación en una sola capa debido a que la mayoría las cámaras actuales realizan la adquisición a color se requiere eliminar las capas adicionales. En el caso de matrices RGB (Rojo, verde y azul) 2 capas, 3 capas para matrices YUV (amarillo y ultravioleta) para generar una sola matriz en lugar de un arreglo de matrices, esta se convertirá en una matriz en escala de grises de 8 bits que deberá tratarse para generar una matriz binaria necesaria para la segmentación y la evaluación. El proceso requiere de los siguientes pasos:

1. Adquirir una serie de imágenes de n muestras.
2. Filtrar cada imagen. Por medio de un umbral convertir la imagen a una matriz binaria.

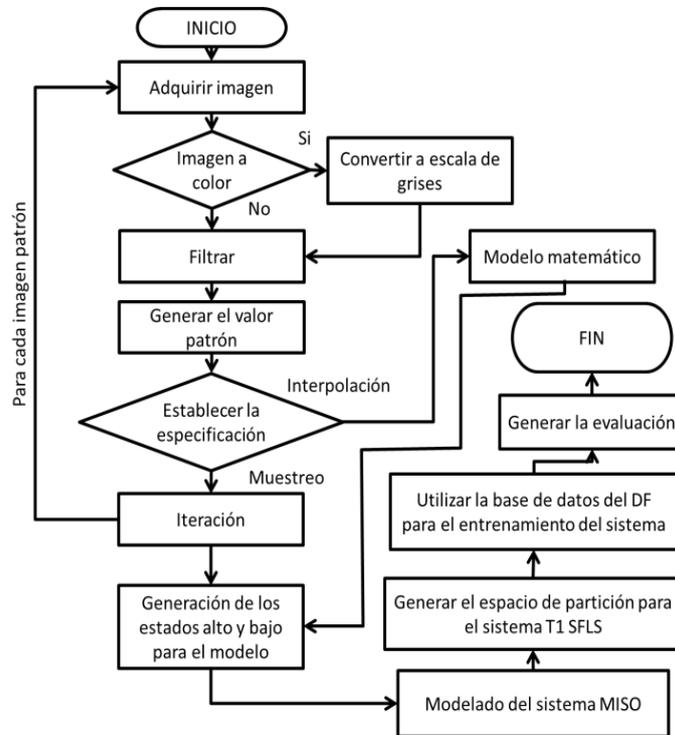


Fig. 2. Diagrama metodológico.

3. Establecer un contador para evaluar la cantidad presente en la muestra en la matriz segmentada.
4. Obtener los límites de la especificación.
5. Por medio de una interpolación del valor obtenido por el contador (generado en el paso 3) y obtener los límites de la especificación en función de lo obtenido en el paso 3.
6. Obtener los límites de la especificación por muestreo de imágenes en los límites de especificación. (repetiendo la metodología varias ocasiones. Una para cada muestra.)
7. Obtener los escalares de los límites para cada variable de los datos obtenidos de los pasos 5 y 6.
8. Con los valores obtenidos del paso 7 generar un sistema de entradas múltiples y una salida (MISO) que servirá para crear la base de reglas del sistema experto. Estas se obtienen por medio de la ecuación (4) obteniendo los resultados para cada combinación de estados de la tabla 1.
9. Generar un espacio de partición uniforme interpolando los estados alto y bajo de cada variable (Fig. 1), es decir, establecer los vértices del DF como las medias para los conjuntos difusos y su dispersión.

10. Utilizar la base de datos del DF como datos de entrenamiento para el sistema experto.
11. Generar la evaluación por aproximación.

Con lo puntos anteriores desarrollados se obtiene el sistema experto con la aproximación (T1 SFLS, T1 ANFIS y RFBN) de base de reglas reducida por medio de la metodología denominada sistema experto (SE) SE/DF/2^k. (Fig. 2), cuando la especificación se establece por muestreo, se modelaría el sistema por medio de una base de inferencia individual la cual requiere de múltiples pruebas o iteraciones y posteriormente se hará la reducción de la base de reglas; cuando se establece la especificación por medio de un modelo matemático se deberá generar una interpolación entre la especificación de diseño (sistema CGS, MKS o sistema ingles) para generar unidades en pixeles generando de forma directa un modelo de base de inferencia compuesta con la interpolación de los límites de especificación. La reducción de reglas se da cuando se compara este modelo con un sistema de reglas basado en inferencia individual [5, 8]. El sistema de base de inferencia compuesta o base de reglas reducida se genera por medio del modelo SE/DF/2^k. Para realizar la evaluación se repiten los pasos 1 al 3 y 11 para cada muestra a ser evaluada.

5. Aplicación del sistema experto (T1 SFLS, T1 ANFIS, RFBN)

La aplicación establecida en el experimento consiste en generar dimensionamiento por medio de imágenes, requiriendo calcular dos parámetros de control dentro de un proceso de producción de placas de base o soporte, la cuales consisten en placas de

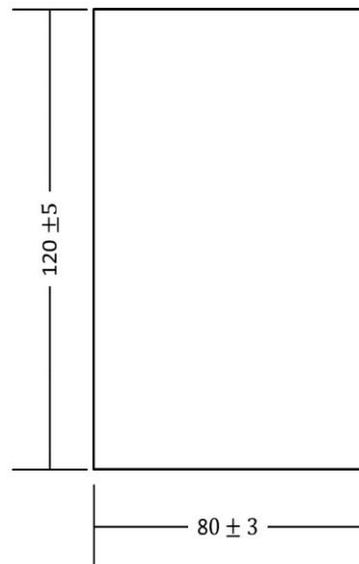


Fig.3. Especificación de la muestra.

aluminio de 1/16" de espesor con las siguiente especificación (Fig. 3). La imagen obtenida por medio del sistema de visión produce una imagen afectada por la reflexión la cual, ocasiona brillo que provoca que la dimensión se altere, esto se puede observar en la Fig. 4 en la cual aparecen zonas en diferentes tonos de gris en el fondo de la imagen para eliminar esta característica es necesario realizar un pre-procesamiento por medio de un filtro el cual funciona por medio de un umbral que realiza el procesado necesario para realizar la conversión a una matriz binaria (Fig. 5). Posteriormente se genera una interpolación de la imagen, y esta se convierte en un vector del cual se obtendrán los valores representativos de cada característica (tabla 3). E.g. para generar la interpolación de la matriz numérica que representa la imagen es necesaria (6), la cual representa la característica que dimensiona el ancho de la muestra.

La fase experimental para esta aplicación se ensambla de la siguiente manera: Todos los experimentos fueron elaborados en una cabina con forma de prisma rectangular de dimensiones 35 X 35 cm. de base y 55 cm. alto, el equipo de adquisición de imágenes es una cámara web Marca Logitech modelo C110, y el análisis se realizó por medio del software MatLab versión R2009a en una laptop con Windows 7 profesional y un procesador Intel Core i-5 @ 2.5 GHz. La iluminación provista al sistema es la siguiente:

Las pruebas del sistema experto (T1 SFLS, T1 ANFIS y RBFN), utilizan un juego de 2 luminarias marca General Electric (GE) colocadas a 10 cm. de la cámara 1 a cada costado, siendo estas focos incandescentes de 75 w con intensidad lumínica de fábrica de 1070 lúmenes por pieza y el objetivo a 43 cm. de la cámara.

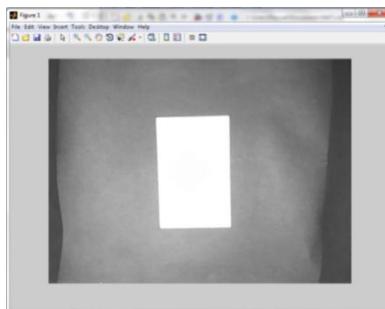


Fig.4. imagen muestra para la inspección.

$$\sum_{i=1}^n \sum_{j=1}^n x_{ij} \quad \forall x \in X \quad (6)$$

A partir de la especificación se obtienen los tratamientos tomando el límite superior de la especificación (muestra 8) y el límite inferior de la especificación (muestra 7 para x_1 , muestra 3 para x_2) para calcular los parámetros **A** alto y **A** bajo, así como **B** alto y **B** bajo (es necesario ajustar los valores del intervalo de evaluación debido a que generalmente no se pueden derrochar recursos realizando todas combinaciones que marcan las diferentes reglas difusas en pruebas físicas para genera un espacio de partición uniforme), e.g. la diferencia entre la media (muestra 1 para x_1 y x_2) y el límite superior de especificación (muestra 8 para ambas variables) produce

un intervalo de 10 unidades para x_1 y 5 unidades para x_2 , igualmente se deben establecer los valores para los límites inferiores que crearan la matriz de reglas del sistema, esta matriz de reglas basadas en conocimiento funcionan de igual manera para los sistemas: T1 SFLS, T1 ANFIS y RBFN teniendo aquí 3 aproximaciones diferentes de sistemas expertos para los procesos de inspección. Estos límites de la especificación son equivalentes a los tratamientos. Los valores usados para la salida de cada regla difusa son determinados mediante conocimiento experto de evaluación ya sea por un procedimiento o una interpretación de los datos del departamento de calidad. La salida del modelo se generará con el modelo híbrido (SE/DF/2^k). El DF es usado para crear la base de reglas reducida del sistema experto generando una forma simplificada y teniendo un modelo metodológico específico para la creación de una base de inferencia compuesta por medio del diseño factorial.

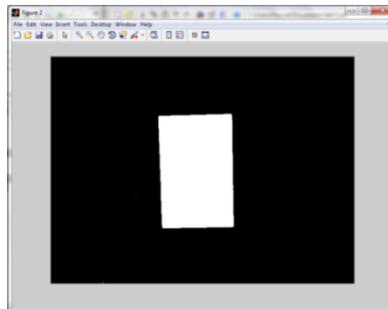


Fig. 5. Imagen filtrada.

La base de datos que aparece en la tabla 3 puede ser usada para modelar el sistema difuso basado en una base de inferencia individual que crearía una base de 13 reglas donde cada prueba o par de datos representaría una regla. Esta se reduce usando los límites mostrados en la Tabla 4. Esta base de inferencia representa al proceso sin la necesidad del conocimiento incierto de una o más personas [22]. El sistema difuso creado utiliza: un modelo tipo Mamdani (T1 SFLS), con un fusificador gaussiano y una implicación de producto, con un defusificador de centro promedio y 4 reglas para el modelo reducido.

Tabla 3. Diseño factorial.

Muestras	1	2	3	4	5	6	7	8	9	10	11	12	13
X_1	208	214	218	218	210	209	205	218	206	211	211	208	207
X_2	139	141	135	142	142	141	137	143	141	146	147	143	145
Objetivo	0.6	3.6	1.2	5.4	3	2.1	-1.5	6	1.2	5.7	6.3	3	3.9

Tabla 4. Base para la aplicaciones DF/ 2^k/ del sistema expeto (T1 SFLS, T1 ANFIS y RBFN).

X_1	133	133	143	143
X_2	198	218	198	218
Y	-6	0	0	6

6. Resultados

Los diferentes experimentos se basan en comparar el desempeño de diferentes sistemas inteligentes como T1 SFLS, T1 ANFIS y el RBFN muestran que el desempeño del modelo híbrido propuesto es viable para su uso con el beneficio que solo requiere el establecimiento de cuatro reglas derivadas de los puntos axiales del diseño factorial, que a su vez, se obtiene de las especificaciones de fabricación. Las etiquetas para el establecimiento de la partición del universo de propagación de los datos se pueden obtener de los diferentes tratamientos del DF. El modelo híbrido ofrece la ventaja de operar con muy pocas reglas. Se obtienen resultados con una precisión aceptable basada en el diseño factorial. Los resultados de los diferentes experimentos se presentan en la Tabla 5.

Las muestras 3, 4, presentan el caso de una variable cercana al límite de la especificación, debido a esta condición los valores de salida son suavizados o pronunciados, es decir los picos de la función se eliminan o se hacen más pronunciados dependiendo del punto donde estén ubicados en la distribución de los datos, como consecuencia la aproximación generada presentara variaciones contra el objetivo provocando una tasa de error del modelo (Fig. 6).

Tabla 5. Muestras evaluadas por medio de la aplicación del sistema experto.

Muestra	X ₁	X ₂	Objetivo	T1 SFLS	ANFIS	RBFN
1	208	139	0.6	0.2115	0.4369	0.52707681
2	214	141	3.6	3.3454	2.7652	1.52429782
3	218	135	1.2	1.4086	-1.2013	-1.52429782
4	218	142	5.4	4.78	4.1033	1.96821401
5	210	142	3	3.5536	2.9742	1.96821401
6	209	141	2.1	1.7998	2.0809	1.52429782
7	205	137	-1.5	-1.0835	-1.3567	-0.52707681
8	218	143	6	4.7123	4.8777	2.36081604
9	206	141	1.2	0.8261	1.6753	1.52429782
10	211	146	5.7	5.088	5.9014	3.15864074
11	211	147	6.3	6.6451	6.5113	3.28683149
12	208	143	3	3.8143	3.4153	2.36081604
13	207	145	3.9	4.2509	4.6287	2.9606785

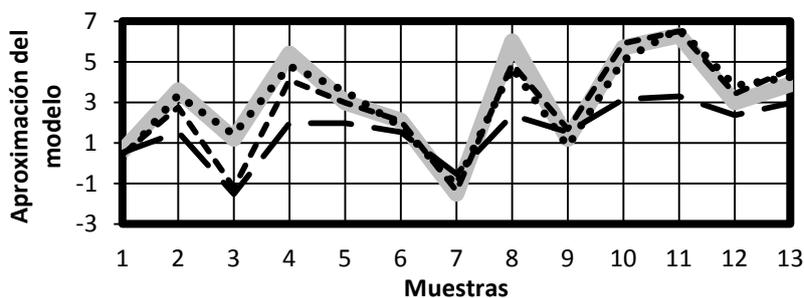


Fig. 6. Resultados de la aproximación generados por el sistema experto. (—) Objetivo, (···) RBFN, (---) T1 SFLS, (- - -) T1 ANFIS.

Las muestras que se encuentran en el intervalo 0.5 a 2.5 desviaciones estándar muestran un desplazamiento proporcional al valor real debido a que la distribución normal en este intervalo presenta un comportamiento lineal, estas variaciones producen pequeños cambios en la salida cuando los cambios en la entrada son significativos. Las muestras encontradas entre la media y 0.5 desviaciones estándar y el intervalo entre 2.5 y 3 desviaciones estándar muestran cambios significativos cuando se producen pequeñas variaciones en la entrada, estas variaciones son producidas debido a la forma de la curva después de los puntos de inflexión, los cuales presentan un comportamiento exponencial. (Fig. 7).

El método desarrollado mejora desempeño de modelos similares tales como: sistema difuso con DCC [4] el cual consta de 11 reglas y base de inferencia individual para la versión de un modelo difuso; por otra parte el modelo (ANFIS/Diseño central compuesto) con base de inferencia compuesta solo opera con 4 reglas obtenidas de los tratamientos del diseño central compuesto [23] y optimiza el modelado del sistema difuso reduciéndolo en más del 60% debido a que se eliminan 7 de las 11 reglas (63.63 %).

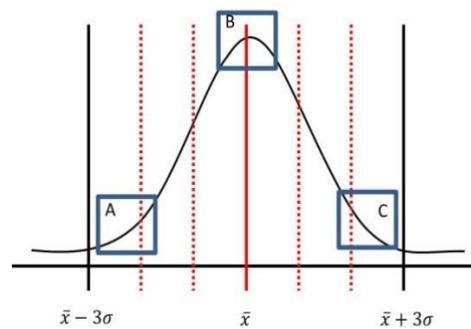


Fig. 7. Puntos de inflexión en una distribución normal (Conjunto difuso).

7. Conclusiones

Las propuestas presentadas como modelos híbridos basados en DOE y computación suave operan con una base de reglas obtenidas de múltiples estados basados en la combinatoria de las pruebas y los datos arrojados por el propio DOE.

El modelo propuesto puede reducir la cantidad de reglas, permitiendo que el sistema de inspección trabaje on-line. Debido a que la base de reglas únicamente requiere de 4 reglas para este caso particular el sistema se puede delimitar dentro de un universo específico sin las restricciones de un sistema con límites internos de partición sin las restricciones de un sistema con límites internos de partición dentro del espacio difuso y la aproximación se puede generar en cualquier punto dentro de este universo sin la necesidad de activar múltiples reglas que brindan variaciones por el ajuste y el redondeo que pudiera provocar la función base difusa teniendo la limitante de realizar múltiples cálculos para cada regla y tener que esperar un tiempo de retraso por la aproximación.

El poder ubicar una muestra en cualquier punto dentro del universo difuso brinda la posibilidad de generar una aproximación de manera rápida solo teniendo que generar el producto por escalar para generar un comportamiento dado en la muestra para cualquier estado del modelo.

Debido a que esta es una investigación en proceso se presentan solo resultados parciales son metodologías basadas en el modelo *singleton* con sus variantes difusa, neurodifusa y red de base radial puesto que se trabaja en otros modelos capaces de procesar la incertidumbre procedente de las mediciones y esta propuesta como trabajo futuro.

Referencias

1. Santos, J.F., Rodrigues, L.F.: Applications of a computer vision techniques in the agriculture and food industry: a review. *European Food Research and Technology*. DOI 10.1007/s00217-012-1844-2 (2012)
2. Czarnowski, I.: Distributed learning with data reduction. *Lecture Notes in Computer Science, Transactions on Computational Collective Intelligence IV*. 6660, pp. 3–124 (2012)
3. Mendel, J.M. On KM Algorithms for Solving Type-2 Fuzzy Set Problems. *IEEE Transactions on Fuzzy Systems*, 21 (3), pp. 426–446 (2013)
4. Mendel, J.M.: *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. pp. 130–185, Upper Saddle River, NJ: Prentice-Hall (2001)
5. Wang, L.: *A Course in Fuzzy Systems and control*, Prentice Hall, 1st. edition (1996)
6. Makadia, A.J., Nanavati, J.I.: Optimisation of machining parameters for turning operations based on response surface methodology. *Measurement* 46, pp. 1521–1529 (2013)
7. Buragohain, M., Mahanta C.: A novel approach for ANFIS modelling based on full factorial design. *Applied Soft Computing*, 8, pp. 609–625 (2008)
8. Praga-Alejo, R., González, G. D., Pérez, V.P., Cantú, S.M., Flores, H.B.: Modeling a Fuzzy Logic System Using Central Composite Design. *Proceedings of 1st annual world Conference of the Society for Industrial and Systems Engineering*. Washington D.C., USA (2012)
9. Martínez, R., Castillo, O., Aguilar, L.T.: Optimization of interval type-2 fuzzy logic controllers for a perturbed autonomous wheeled mobile robot using genetic algorithms. *Information Sciences*. 179 (13), pp. 2158–2174 (2009)
10. Vale, Z., Morais, H., Faria, P., Ramos, C.: Distribution system operation supported by contextual energy resource management based on intelligent SCADA. *Renewable Energy* 52, pp. 143–153 (2013)
11. Méndez, G. M., Hernández, M.A.: Hybrid learning mechanism for interval A2-C1 type-2 non-singleton type-2 Takagi–Sugeno–Kang fuzzy logic systems. *Information Sciences*, 220, pp. 149–169 (2013)
12. Tomiyama, T. , Gu, P., Jin, Y., Lutters, D., Kind, Ch., Kimura F.: Design methodologies: Industrial and educational applications. *CIRP Annals - Manufacturing Technology*, 58. pp. 543–565 (2009)
13. NCS TIB 04-1: National Communications System, Technical Information Bulletin 04-1, Supervisory Control and Data Acquisition (SCADA) Systems, October 2004, Office of the Manager National Communications Systems. Arlington, Virginia. pp. 4–9 (1994)

14. Ghasemi, M. J., Tajozakerin, H. R., Omidian, A.R.: An Iranian National Number Plate Localization and Recognition System for Private vehicles. *International Journal of Academic Research*. 2(6), pp. 13–19 (2010)
15. Zhang, F., Zhang, X.: Classification and Quality Evaluation of Tobacco leaves Based in Image Processing and Fuzzy Comprehensive Evaluation. *Sensors*. 11(3), pp. 2369–2384 (2011)
16. Zadeh, L.A.: Fuzzy Sets. *Information and control*, 8, pp.338–353 (1965)
17. Morales-Lunak G.: *Introducción a la lógica difusa*, Centro de Investigación y Estudios Avanzados del IPN, (CINVESTAV-IPN), [Internet, Consultado el 8 de Octubre de 2011]. (2002)
18. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-I, *Information Sciences*, 8(3), pp. 199–249 (1975)
19. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-II, *Information Sciences*, 8(4), pp. 301–357 (1975)
20. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-III, *Information Sciences*, 9(1) , pp. 43–80 (1975)
21. Montgomery, D.C.: *Diseño y Análisis de experimentos*. 2ª Edición. pp. 218-279. Limusa-Wiley (2004)
22. Jian-Bo Yang, Jun Liu, Jin Wang, How-Sing Sii, Hong-Wei Wang: Belief Rule-Base Inference Methodology Using the Evidential Reasoning Approach—RIMER, *IEEE Transactions on systems, Man and Cybernetics, Part A: Systems and Humans*, 36(2), pp. 266–285 (2006)
23. Montes Dorantes, P.N., Praga-Alejo, R., Nieto González, J.P., Méndez, G.M.: Modelado de Sistemas Adaptativos de Inferencia Neuro-Difusa Usando Diseño Central Compuesto. *Research in Computing Science*, 62, pp. 259–269 (2013)

Segmentación automática en imágenes RGB aplicando la técnica *Fuzzy C-means* de la morfología matemática para la ayuda de la fotoidentificación de cetáceos

Cesar Muñoz Pérez¹, Diego Cabrera Padilla¹, B. E. Carvajal-Gómez¹,
F.J. Gallegos-Funes², Diane Gendron³

¹Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria de Ingeniería y Tecnologías Avanzadas, México

²Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica, México

³Instituto Politécnico Nacional,
Centro Interdisciplinario de Ciencias Marinas, México

becarvajal@ipn.mx

Resumen. Los cetáceos son mamíferos que son un componente importante de los ecosistemas marinos, por lo que es importante para aumentar el interés y el conocimiento de estos animales. La identificación de los cetáceos se puede realizar mediante la observación de sus patrones a través de su forma aleta caudal. En este trabajo se presenta el algoritmo de segmentación Fuzzy C-means (FCM) para imágenes en el espacio de color RGB de la ballena azul en el mar de Cortés, México. El FCM es uno de los algoritmos de agrupamiento utilizados con mayor frecuencia debido a su eficacia, y en conjunto con morfología matemática se comprueba que las imágenes se segmentan aun teniendo como ruido de fondo el cielo y la mar..

Palabras clave: Patrones de aleta dorsal, FCM, morfología matemática, ballena azul.

1 Introducción

Para identificar algunas especies se realiza a través de marcas artificiales; como etiquetas, tatuajes, o alguna marca otra realizada a propósito como la eliminación de una sección de piel [9], haciendo de esto una marca permanente. Las marcas naturales son únicas para cada animal, siendo un método muy fiable, a menos que el animal pierde sus marcas [9]. La aplicación de marcas artificiales, sin embargo; es invasiva porque el animal normalmente debe ser capturado y manejado al ser marcado y, a menudo recapturado para poder ser identificado [9]. En consecuencia, este método podría ser difícil y costoso de usar, posiblemente perjudicial para la salud o el comportamiento del animal [9]. La fotoidentificación (photo-id) es un método el cual mediante el uso de fotografías adquiridas en el hábitat natural del animal, son

utilizadas para localizar las marcas naturales de los animales, estas marcas suelen ser: cicatrices, patrones de pigmentación en la piel, en el pelo, en el hocico, patrones de aletas dorsales o caudales para el caso de los cetáceos. El uso de marcas naturales para identificar a los individuos de una especie tiene muchas ventajas, por ejemplo; este método es no invasivo y de bajo costo, dos cualidades que pueden explicar y justificar su amplia utilización en el estudio del comportamiento de los cetáceos [9], agregando también que el individuo puede ser identificado desde una distancia sin llegar a interferir directamente en la iteración del cetáceo con su hábitat natural. En la mayoría de las veces la fotoidentificación es hecha manualmente por una persona que puede tener poca experiencia o ser nuevo en el área esto puede tomar mayor tiempo en la realización de la fotoidentificación realizada por un experto, generando además un error adicional al ya esperado. Por ejemplo, en la fotoidentificación manual de las imágenes de los delfines, cuando se encuentra los contornos de las muescas de la aleta del delfín, algunas de estas características son omitidas ocasionando un bajo rendimiento en la clasificación e identificación de los delfines correctamente. Por esta razón, existe una dependencia directa con el experto causando que los resultados varíen. En muchos sistemas de identificación asistida por ordenador semiautomáticas cada nueva imagen requiere un pre-procesamiento manual (por ejemplo, recortar, mejorar, el brillo y contraste) aumentando el tiempo, también puede requerir la formación, la introducción de la subjetividad y error en el resultado [11].

Este trabajo es motivado debido a los pocos sistemas de segmentación automática de cetáceos para agilizar el proceso de fotoidentificación. El primer paso en cualquier proceso de identificación de individuos u objetos, es el paso de la localización del individuo de interés dentro de la imagen, este proceso es conocido como segmentación de la imagen. Los algoritmos de segmentación de imágenes están basados en propiedades básicas de los niveles de gris de una imagen como son: discontinuidad y semejanza [13]. En este trabajo se presenta un método automático de segmentación de imágenes de cetáceos, el cual consiste en la técnica de agrupación o mejor conocido como *Fuzzy C-means* (FCM), integrando simultáneamente la erosión y la dilatación técnicas de morfología matemática. La técnica de agrupación se aplica a imágenes reales de la ballena azul en el mar de Cortés, México. El proceso de agrupación de objetos es un proceso, dentro del cual se van separando dichos objetos en una clase o clúster, de manera que, los objetos con un grupo o clase tienen una alta similitud, pero son muy diferentes a otros objetos. Esto significa que el límite entre grupos debe estar claramente definido. Sin embargo, en la práctica, los límites entre los grupos pueden no estar claramente definidos. Algunos objetos pueden pertenecer a más de una agrupación [17]. Un sistema de fotoidentificación debe de contar previamente con un pre-procesamiento con las siguientes características:

Precisión. ¿Con qué frecuencia el programa devuelve el contorno de aleta correcto, o incluye la forma de la aleta correcta en la parte superior de una lista de posibles candidatos?, y ¿qué tan fiable puede el programa sugerir que no hay coincidencias correctas en el catálogo actual?

Velocidad. ¿Cuánto tiempo se tarda en ingresar los datos de una fotografía a la computadora?, y, más importante, ¿cuánto tiempo se tarda el equipo para que coincida con una fotografía de entrada frente a un catálogo de un tamaño determinado?

Costo. ¿Cuál es el costo del hardware y software necesarios para implementar el sistema?

Por esta razón se propone el uso del algoritmo FCM para la segmentación de imágenes, el cual se comprobó durante la implementación y ejecución de la segmentación de la imagen del cetáceo, en menos de 2 segundos, además; de que este es de fácil implementación en hardware. El algoritmo FCM depende de la medida de la distancia entre los objetos. Comúnmente FCM utiliza la distancia euclidiana que supone cada objeto tiene la misma importancia [17]. El algoritmo FCM asigna a cada objeto de una imagen un valor de pertenencia a cada grupo y, por lo tanto, un conjunto de datos específicos que puede pertenecer a más de una clase. Este algoritmo realiza una suave partición de la información proporcionada por la imagen, por lo que, los objetos tienen ningún valor de pertenencia en todos los grupos. Agregando la dilatación y erosión, se puede delimitar de una mejor manera los bordes del cetáceo con el fondo de la imagen, debido a que se verificó que utilizando solamente FCM, quedaban algunas secciones menos delimitadas, es decir; toma secciones del fondo de la imagen. Generalmente, la tonalidad de la piel de la ballena azul tiende a confundirse con el fondo de la imagen compuesta por secciones de cielo, superficie del mar, brillo debido al reflejo del agua con el sol, características que suelen complicar el proceso de segmentación de la imagen del cetáceo para la ayuda en la fotoidentificación. El objetivo de este proyecto es la extracción del cuerpo del cetáceo de la imagen y eliminar todo aquello que no es de interés (el cielo y el mar) con el fin de hacer un análisis, que es para detectar el tipo de aleta dorsal con el mamífero para su identificación, comparación y clasificación.

2. Identificación de la ballena azul

Las imágenes usadas para el desarrollo de este proyecto fueron proporcionadas por la Dra. Diane Gendron (CICIMAR-IPN). Esta base de datos cuenta con un catálogo de 621 imágenes de la ballena azul en su hábitat natural. Una importante característica es el tipo de su aleta dorsal, la cual es de mucha ayuda para la realización de la fotoidentificación debido a que esta se encuentra la mayor parte del tiempo fuera del agua y se puede fotografiar rápidamente. Los diferentes tipos de aleta dorsal de la ballena son mostrados en la Figura 1.

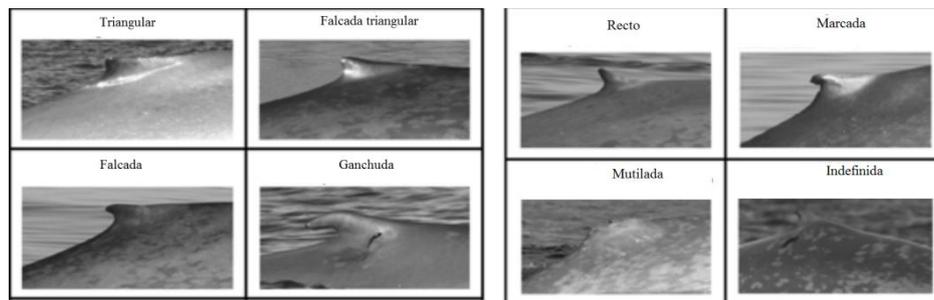


Fig 1. Diferentes tipos de aleta dorsal.

3. Imágenes digitales

Una imagen puede ser considerada como una función $f(x,y)$ la cual se ha discretizado tanto en coordenadas espaciales y como en el brillo. Una imagen digital es una matriz con índices de filas y columnas los cuales representan un punto en la imagen y su valor identifica el nivel de color en ese punto. Los elementos de estas matrices digitales son conocidos como píxeles [7]. Las imágenes digitales pueden ser presentadas en diferentes espacios de color, por ejemplo: RGB (rojo, verde, azul), HIS (matiz, saturación, intensidad), HSV (matiz, saturación, valor), CMYK (cyan, magenta, amarillo y negro), YCbCr (luminancia, cromancia en azul, cromancia en rojo) [7]. El desarrollo de la metodología propuesta en este trabajo se llevó a cabo inicialmente con imágenes en el espacio RGB y en escala de grises. Sin embargo, el mismo método se puede aplicar a otros espacios de color. Los resultados se analizaron con el fin de encontrar el mejor espacio de color para la segmentación de imágenes.

4. Método propuesto

El método propuesto se muestra en la Figura 2. Los pasos a seguir para la segmentación se describen más adelante. Este se verifica en la sección de prueba que el algoritmo propuesto es válido para cualquier tipo de imágenes obtenidas a partir de pruebas experimentales.

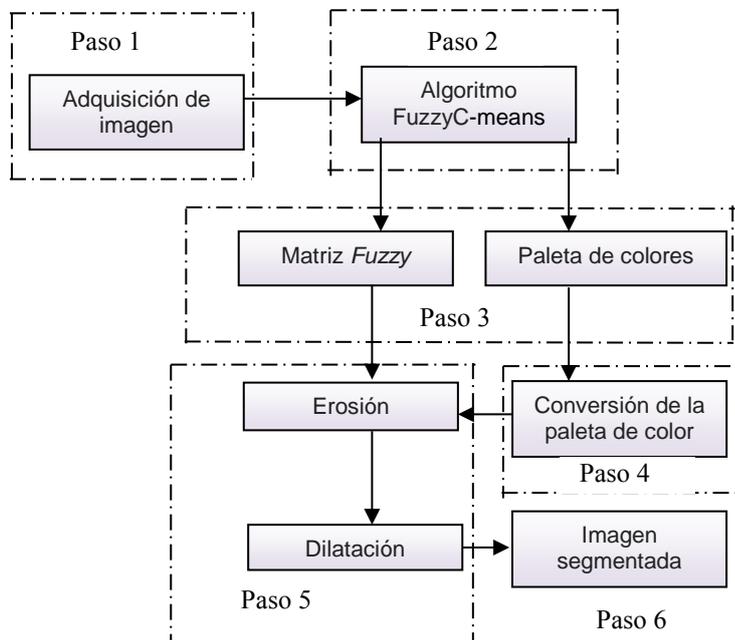


Fig 2. Diagrama a bloques del método propuesto.

- Paso 1. Adquisición de imágenes de los cetáceos.
- Paso 2. Se realiza el algoritmo FCM a la imagen original.
- Paso 3. Se obtiene de manera simultánea la matriz *fuzzy* y la obtención de la paleta de colores.
- Paso 4. Se conserva la paleta de color.
- Paso 5. Se realiza la erosión y la dilatación a la matriz *fuzzy*.
- Paso 6. Segmentación. En esta etapa la imagen se separa en cada uno de los objetos que lo hacen [13]. Las imágenes se toman en el formato RGB a la que se aplica un proceso digital que comienza con el algoritmo FCM para la segmentación y las operaciones de morfología matemática se aplican para eliminar completamente el ruido de la imagen.

4.1 Algoritmo Fuzzy C-Means

El algoritmo FCM es uno de las más populares técnicas de agrupación, este algoritmo fue desarrollado por Dunn [4] y, eventualmente, modificado por Bezdek [2] en 1981. El FCM es un algoritmo difuso e iterativo que forma parte de las técnicas de agrupamiento no supervisado y su objetivo es encontrar patrones o grupos en un determinado conjunto de datos, por lo que este tipo de patrones sirven para la clasificación de la información [16]. El principio de funcionamiento básico de este algoritmo es agrupar o separar los datos proporcionados en grupos llamados *clusters*. La ventaja de este algoritmo a otros que también trabajan en el principio de agrupación o *clustering*, es que los datos pueden estar suficientemente cerca de dos grupos de tal manera que es difícil etiquetar cualquiera de estos. El FCM es un algoritmo desarrollado para resolver estos problemas, por lo que un conjunto de datos de partición suave y restringido, es decir; con estas particiones en que todos los datos tienen un grado de pertenencia para todos los grupos, y la suma de estos grados de pertenencia es igual a 1. Para el desarrollo de esta investigación se presentan 2 grupos: el primero correspondiente al cuerpo de la ballena azul y el segundo contiene la información del mar [9]. Supongamos que X es un conjunto de datos y un elemento perteneciente a X es x_i , se dice que una partición $P = \{C_1, C_2, \dots, C_C\}$ donde c indica el número de clusters. Se decide que P es una partición suave y restringido de X cuando se cumplan las siguientes condiciones [6]:

$$\forall x_i \in X \quad \forall C_j \in P \quad 0 \leq \mu_{C_j}(x_i) \leq 1 \quad (1)$$

$$\forall x_i \in X \quad \exists C_j \in P \quad \text{tal que } \mu_{C_j}(x_i) > 0 \quad (2)$$

donde: $\mu_{C_j}(x_i)$ es el grado de pertenencia para la cual x_i pertenece al clúster C_j .

$$\sum_j \mu_{C_j}(x_i) = 1 \quad \forall x_i \in X \quad (3)$$

Al tener una partición suave y restringida, la función objetivo se da como se muestra a continuación:

$$J_m(P, V) = \sum_{i=1}^k \sum_{x_k \in X} (\mu_{C_i}(x_k))^m \|x_k - v_i\|^2 \quad (4)$$

Donde, el parámetro m es un peso que determina el grado en que los miembros pertenecen al clúster, y V_i se llama prototipo o centroide que minimiza la función objetivo J_m . Una vez que se ha determinado el número de grupos, el valor de m debe ser calculado utilizando las funciones de pertenencia de la ecuación 5 y posteriormente actualizar los centroides utilizando la ecuación 6.

$$\mu_{C_i}(x) = \frac{1}{\sum_{j=1}^k \left(\frac{\|x-v_i\|^2}{\|x-v_j\|^2} \right)^2} \quad 1 \leq i \leq k, x \in X \quad (5)$$

$$v_i = \frac{\sum_{x \in X} (\mu_{C_i}(x))^m x}{\sum_{x \in X} (\mu_{C_i}(x))^m} \quad 1 \leq i \leq k \quad (6)$$

Los pasos descritos anteriormente se repiten iterativamente hasta alcanzar el criterio de paro. Una vez aplicado el algoritmo FCM para las imágenes, se aplican las operaciones de morfología matemática para la eliminación de las regiones excedentarias que no son de interés.

4.2 Morfología matemática

Morfología matemática es una técnica bien establecida para el análisis de imágenes, con fundamentos matemáticos sólidos [16] que se encuentran con enormes aplicaciones en muchas áreas, principalmente el análisis de imágenes [14]. La mayor parte de las aplicaciones prácticas se basan en una combinación de un pequeño conjunto de operaciones llamadas: erosión y la dilatación. La morfología matemática se desarrolló inicialmente para imágenes binarias y más tarde en las imágenes generalizadas con valores de gris [15]. El instrumento fundamental en la morfología matemática es el elemento estructurante. Un elemento estructurante es simplemente como una configuración de píxeles en las que se define a partir de un origen [12].

Erosión. Es el resultado de la comprobación de si el elemento estructurante Y está totalmente incluido en el conjunto X . Cuando esto no sucede, el resultado de la erosión es el conjunto vacío. Se define de la siguiente manera [3]:

$$\varepsilon_B(Y) = X \ominus Y = \{x | Y_x \subseteq X\} \quad (7)$$

Cuando los objetos de la escena son más pequeños que el elemento de estructurante, estos desaparecen de la imagen.

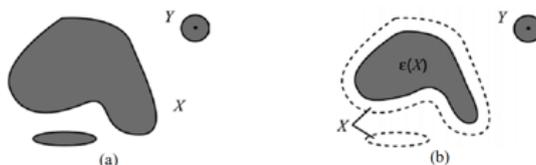


Fig 3. Efectos de la erosión.

Dilatación. La dilatación es la transformación dual a la erosión. El resultado de la dilatación es el conjunto de elementos de manera que al menos algún elemento de estructurante del conjunto B está contenido en el conjunto X , cuando B se mueve sobre el conjunto X :

$$\varepsilon_B(X) = X \oplus B = \{x | X \cap B_x \neq \emptyset\} \quad (8)$$

Esta operación representa un crecimiento progresivo del elemento estructurante llamado *passing* sobre el conjunto X , en el conjunto, esto no va a cambiar. Sin embargo, la frontera del conjunto X , pasando a B , el conjunto de resultados se

expandirá. La aplicación repetida de este operador podría degradar la imagen, haciendo coincidir la imagen completa a la dilatada [10].

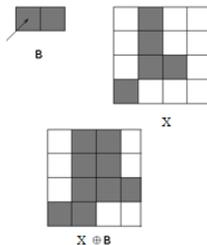


Fig 4. Efectos de dilatación.

5. Resultados

El algoritmo de FCM se puso a prueba para ayudar a realizar la fotoidentificación a través de la aleta dorsal de la ballena azul, esta aleta es importante debido a que se puede identificar casi a cualquier tipo de mamíferos que llegan hasta el mar de Cortés y ayudar a tener un registro de ellos cada año. Para el FCM todas las fotos que se utilizaron tienen diferentes condiciones de brillo y el tamaño de la imagen. El algoritmo también se demostró en diferentes categorías de imágenes, estas categorías se han propuesto y avalado por la investigación de la Dra. Diane Gendron (CICIMAR-IPN). En la Tabla 1 se muestran los promedios de los resultados obtenidos a partir del algoritmo FCM, además; se visualizan los resultados cualitativos y cuantitativos de este algoritmo en un primer momento se muestra la segmentación manual, en segundo lugar se muestra la imagen obtenida con el algoritmo FCM desarrollado a partir de la foto original. La implementación de este algoritmo se hizo con Matlab™ y finalmente se muestra una comparación entre la imagen de algoritmo FCM y la segmentación manual, la obtención de cuatro parámetros que ayudaron a mostrar cómo eran iguales ambas imágenes. Los resultados de rendimiento se calculan a partir de las segmentaciones manuales del primer observador como una realidad de la imagen. Para ver el rendimiento del método propuesto, se consideran cuatro casos: dos clasificaciones correctas y dos errores de clasificación. Las clasificaciones son: verdadero positivo (TP), falso positivo (FP), falso negativo (FN), y verdadero negativo (TN). Mediante el uso de estas métricas se obtienen diferentes medidas de rendimiento como el siguiente [15, 16]:

$$\text{Especificidad} = TN / (TN + FP) \quad (9)$$

$$\text{Sensitividad} = TP / (TP + FN) \quad (10)$$

$$\text{Precisión} = (TP + TN) / \text{píxeles correspondientes al cetáceo} \quad (11)$$

la especificidad (Sp) es la capacidad de detectar los píxeles no de cetáceos, la sensibilidad (Se) refleja la capacidad de un algoritmo para detectar el borde de cetáceos, la precisión (ACC), se mide por la relación entre el número total de píxeles clasificados correctamente (suma de los verdaderos positivos y verdaderos negativos)

por el número de píxeles de la imagen de los cetáceos , y el valor predictivo positivo o tasa de precisión da la proporción de píxeles de cetáceos identificados y que son verdaderas [1, 8], es la probabilidad de que un pixel de cetáceos identificado es un positivo verdadero. La similitud (S_i) en el contexto de este proyecto, se ha definido como una función de la precisión y de sensibilidad, con el fin de comparar los resultados obtenidos de manera más directa. La similitud de una imagen segmentada con la realidad (de la imagen segmentada manualmente) se define en la ecuación 12 [5],

$$Similitud = 1 - \frac{\sqrt{(1 - precision)^2 + (1 - sensibilidad)^2}}{\sqrt{2}} \quad (12)$$

La presentación de la máxima similitud con la realidad sobre la segmentada manualmente, el valor comprendido está entre 1 y 0, cuando se presenta la similitud mínima con la segmentada manualmente.

Tabla 1. Resultados obtenidos a partir de la técnica propuesta de segmentación.

<i>Tipo de aleta</i>	<i>Tiempo de procesamiento (seg)</i>	<i>Si(%)</i>	<i>Fuzzy C-Mean</i>		
			<i>Sp(%)</i>	<i>Se(%)</i>	<i>ACC(%)</i>
Ganchuda-ND	1.633	89.4813	96.077	86.8471	93.6305
Ganchuda clara	1.15	83.3155	90.477	80.7410	86.4039
Ganchuda clara moteada	1.5	78.7697	80.3055	78.5373	80.5325
Ganchuda rayada	2.0	70.2405	73.3529	69.0546	71.4756

Tabla 2. Resultados obtenidos a partir de la técnica K-means.

<i>Tipo de aleta</i>	<i>Tiempo de procesamiento (seg)</i>	<i>Si(%)</i>	<i>K-Mean</i>		
			<i>Sp(%)</i>	<i>Se(%)</i>	<i>ACC(%)</i>
Ganchuda-ND	5.3745	88.8453	96.2618	85.5975	93.5643
Ganchuda clara	4.9442	85.6447	98.6170	83.8277	90.2452
Ganchuda clara moteada	4.8854	80.7697	81.5578	83.2071	81.3250
Ganchuda rayada	4.9442	73.0257	78.0527	70.1536	76.4721

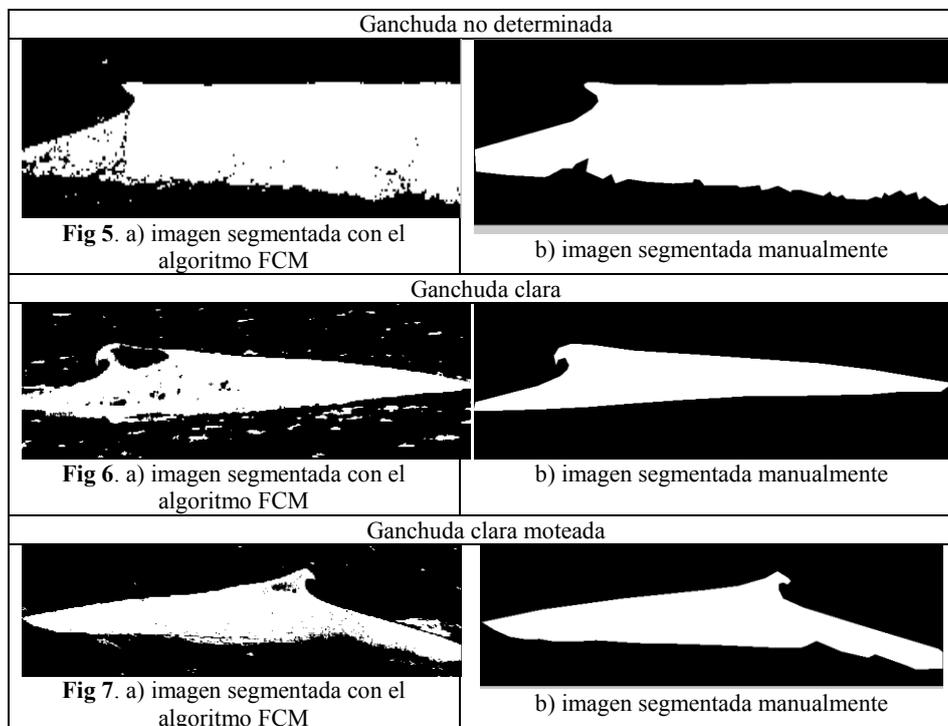
Del catálogo compuesto por 400 imágenes, se observaron los siguientes resultados para la categoría de "ganchuda clara" se obtuvieron que las 100 imágenes de prueba los resultados son superiores a 86% en la prueba de precisión, debido a que el color de la piel de la ballena, es uniforme y más claro que el color del agua del mar de Cortés, México.

En la categoría de "ganchuda clara moteada", se obtuvo a partir de las 100 imágenes de prueba para esta categoría que el 33% de las imágenes tuvo un resultado superior al 80% en la precisión, esta prueba tuvo un porcentaje menor debido a que la

coloración de la piel de la ballena tienen a camuflajearse más con el azul oscuro del mar de Cortés, México.

Tabla 3. Resultados obtenidos a partir de la técnica de morfología matemática.

Tipo de aleta	Tiempo de procesamiento (seg)	Si(%)	K-Mean		
			Sp(%)	Se(%)	ACC(%)
Ganchuda-ND	1.9442	88.0012	86.3678	85.0901	83.6921
Ganchuda clara	2.3745	81.4197	85.1702	80.4279	90.0451
Ganchuda clara moteada	2.8854	90.6997	81.5377	83.7221	91.3751
Ganchuda rayada	2.9442	79.7612	79.5326	70.6754	76.0761



En la categoría de "ganchuda rayada" se obtuvo que el 0% de las 100 imágenes tomadas como prueba tuvo un resultado superior al 80% en la precisión, debido a que esta se camuflajea mejor con el color del agua del mar de Cortés.

En la categoría de "no determinada" se obtuvo que el 100% de las imágenes tuvo un resultado superior al 93% en la precisión, porque el cuerpo de la ballena estaba se encontraba a contra luz con el fondo azul del mar de Cortés. Los resultados mostrados en la Tabla 1 con la técnica propuesta se encuentran en algunos casos superior a la técnica de K-means que se muestran en la Tabla 2, la técnica de K-means muestra en

algunos casos un mejor rendimiento en sus resultados, sin embargo; la técnica de K-means tiene un consumo de procesamiento mayor tanto para la técnica propuesta Tabla 1, como los resultados mostrados en la Tabla 3 con morfología matemática. Siendo así la técnica propuesta una técnica con rendimiento intermedio con un tiempo de procesamiento inferior a los 2 segundos. A continuación se muestran los resultados visuales obtenidos con el algoritmo propuesto en este trabajo de investigación.

6. Discusión

El algoritmo aquí propuesto para la segmentación de imágenes reales, dio resultados satisfactorios. Estos servirán para poder identificar el tipo de aleta de la ballena azul con el fondo de la imagen. Este algoritmo puede implementarse en cualquier espacio de color para ir localizando los mejores resultados obtenidos. De los resultados experimentales expuestos en este trabajo, se puede ver que el método propuesto provee una mejor calidad en la segmentación de imágenes reales, explícitamente de los cetáceos en su hábitat natural, presentando resultados en promedio superiores al 80% en cetáceos clasificados como “clara” y “no determinada”. En los resultados visuales subjetivos mostrados en las Figuras 5, 6 y 7, se puede observar que las imágenes segmentadas conservan las características principales de la forma de su aleta, siendo esto de utilidad para finalmente realizar la fotoidentificación de estas en un sistema autónomo. La técnica de segmentación propuesta en este trabajo muestra resultados superiores en algunos casos con los tipos de aleta de Ganchuda No Determinada y Ganchuda rayada, lo que cumple con el compromiso de costo de consumo en el procesamiento de las imágenes, a diferencia de las técnicas mostradas en la Tabla 2 y 3.

7. Conclusión

Después de probar diferentes imágenes del catálogo de la base de datos del CICIMAR-IPN, el algoritmo FCM junto con morfología matemática conservara el contorno de la aleta dorsal y la forma del lomo de la ballena. Con se encontró el algoritmo FCM propuesto que se requiere necesita al menos cuatro iteraciones para tener una separación entre los diferentes grupos en las capas RGB, porque durante estas cuatro iteraciones el resultado de los grupos de comparación estabilizado, por lo que es posible utilizar recursos mínimos de hardware y así llevarlo a un dispositivo móvil. También se encontró en la prueba de precisión que si el resultado fue inferior al 80%, las imágenes obtenidas muestran el ruido que no permite distinguir entre el ambiente de la ballena y la ballena. Para obtener los mejores resultados, el algoritmo funcionó mejor con la categoría de la "clara" y "no determinada". Obteniendo la imagen aislada del cetáceo con el fondo de la imagen al momento que se realiza la identificación visual, el biólogo puede agilizar la clasificación de las fotos en sus catálogos e ir automatizando poco a poco la técnica de fotoidentificación. Conservando además la sección más importante al momento de realizar la fotoidentificación manual por parte de los biólogos expertos, que es la forma del

contorno de la aleta dorsal. Se comprobó además que el tiempo de ejecución de este algoritmo es menor a los 2 segundos. Los resultados mostrados en la Tabla 1 con la técnica propuesta se encuentran en algunos casos superior a la técnica de K-means que se muestran en la Tabla 2, la técnica de K-means muestra en algunos casos un mejor rendimiento en sus resultados, sin embargo; la técnica de K-means tiene un consumo de procesamiento mayor tanto para la técnica propuesta Tabla 1, como los resultados mostrados en la Tabla 3 con morfología matemática.

Agradecimientos. El equipo de trabajo agradece al IPN por las facilidades para el desarrollo de este trabajo en especial al M. en C. Arodi Carvallo Dominguez, de igual modo agradecemos a CONACYT en general por el soporte de esta investigación.

Referencias

1. Abdolhossein, F., Naghsh-Nilchi, A. R.: Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation. *Biomedical Signal Processing and Control* 8, pp. 71–80 (2013)
2. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA (1981)
3. Carvajal-Gámez, B.E., Gallegos-Funes, F.J., Casas-Pérez, C., Muñoz- Pérez, C., Reyes-Morales, A.A., Gendron, D., Hernández-Pérez, A.: Algoritmo de segmentación de imágenes reales de la ballena azul en el mar de Cortes.
4. Chattopadhyay, S., Pratihar, D.K., De Sarkar, S.C.: Performance Studies of Some Similarity-Based Fuzzy Clustering Algorithm. *International Journal of Performability Engineering*, Vol. 2, No. 2, pp. 191–200 (2006)
5. Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P.: Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering and Remote Sensing*, pp. 289–299 (2010)
6. De la Rosa Flores, R.: *Procesamiento de imágenes digitales*. Semana de la informática. Instituto Tecnológico de Puebla (2007)
7. Escalante-Ramírez, B.: *Procesamiento Digital de Imágenes*. (2006)
8. Fraz, M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G. and Barman, S.A.: Blood vessel segmentation methodologies in retinal images. A survey, computer methods and programs in biomedicine 108, pp. 407–433 (2012)
9. Gomez-Salazar, C., Trujillo, F., and Whitehead, H.: Photo-Identification: A Reliable and Noninvasive Tool for Studying Pink River Dolphins (*Inia geoffrensis*). *Aquatic Mammals*, pp. 472–485 (2011)
10. González, R.: *Digital Image Processing*. 2da edición, Ed. Prentice-Hall. (2001)
11. Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters. International Whaling Commission. Report of the International Whaling Commission, Special Issue 12.
12. Laganière, R.: *OpenCV 2 Computer Vision Application Programming Cookbook*. (2011)
13. Martín, M.: *Técnicas clásicas de segmentación de Imágenes*. (2012)
14. Serra, J.: *Image Analysis and Mathematical Morphology*. Academic Press. London, UK. (1984)
15. Sternberg, S. R.: “Grayscale morphology”. *Computer Vision, Graphics, and Image Processing*, Vol. 35, No. 3, pp. 333–355 (1986)
16. Yen, J., Reza, L., *Fuzzy Logic intelligence control and information*. Prentice Hall, pp. 351-362. New Jersey (1999)

Cesar Muñoz Pérez, Diego Cabrera Padilla, B.E. Carvajal-Gómez, F.J. Gallegos-Funes, et al.

17. Yujie, L., Huimin, L., Lifeng, Z., Junwu, Z., Shiyuan, Y., Xuelong, H., Xiaobin, Z., Yun, L., Bin, L. and Seiichi, S.: An Automatic Image Segmentation Algorithm Based on Weighting Fuzzy C-Means Clustering. Springer Link (2012)

Descriptores basados en combinaciones lineales de polinomios de Jacobi aplicados en la caracterización de melanoma, comparados estadísticamente con los momentos de Hu

Damián A. Sánchez-Trasgallo, Martín Carpio, R. Montero, Luis E. Mancilla Espinoza, Miguel A. Casillas Araiza

División de Estudios de Posgrado e Investigación, Instituto Tecnológico de León, León, Guanajuato, México

damian.trasgallo@gmail.com, jmcarpio61@hotmail.com, raul.santiago@itleon.edu.mx, lmancilla01@hotmail.com, mcasillasa@prodigy.net.mx

Resumen. El presente trabajo describe una metodología encaminada a la caracterización de las lesiones cutáneas según su contorno, mediante combinaciones lineales de polinomios de base radial. Esta metodología es contrastada con los Momentos de Hu mediante un proceso de clasificación donde se emplean los clasificadores Distancia Mínima, KNN, Naive Bayes y CAP. Posteriormente estos porcentajes de clasificación son comparados mediante la prueba estadística no paramétrica de suma de rangos con signo de Wilcoxon.

Palabras clave: Metodología de caracterización, polinomios de base radial, melanomas, clasificación, momentos de Hu, prueba estadística no paramétrica de suma de rangos con signo de Wilcoxon.

1. Introducción

Es importante detectar lesiones cutáneas del tipo melanoma en una etapa temprana, ya que son una clase de tumor maligno cancerígeno altamente letal en fases avanzadas que produce metástasis y son muy resistentes a la quimioterapia. Los melanomas suelen ser confundidos con verrugas, hongos, hematomas, entre otros, lo cual dificulta su correcto diagnóstico y son altamente curables en fases tempranas. Existe la metodología no invasiva ABCDE (A = Asimetría, B = Borde irregular, C = Coloración irregular, D = Diámetro mayor a 6 mm. y Evolución) para la detección de melanomas. Aunque por estas características se sospecha que el melanoma es maligno, ninguna de ellas, ni la combinación de ellas, puede fortalecer la sospecha de que el melanoma es maligno y es necesario hacer una revisión del tejido cutáneo extirpado en el laboratorio [1].

Para diagnosticar un melanoma de entre una serie de lesiones cutáneas, se representa cada una de las muestras como un *vector de características* [5] como se muestra en la expresión 1 :

$$(x_1, x_2, x_3, \dots, x_n) \quad (1)$$

Donde cada característica del vector es obtenida por un descriptor. Éstos patrones posteriormente son clasificados y asignados a una clase.

las **características** son todos los atributos que describen un objeto y un **descriptor** es una medida que nos permite cuantificar ciertas características de los objetos. Un **patrón** es un conjunto de entidades que comparten características que las diferencian de otras, se agrupan en **clases** y es representado por un **vector de características**. Un **clasificador** es un subsistema que utiliza un vector de características de la entidad cuantificable y posteriormente lo agrupa en una de las clases [4]. En el caso de una lesión cutánea la circularidad y compacidad son ejemplos de descriptores [3] [5].

Existen descriptores basados en los momentos invariantes de Hu [26], momentos de Zernike [2], momentos de chebyshev [2] [27], entre otros. Estos descriptores mencionados se calculan empleando la **región de interés** que ocupa una lesión cutánea en una imagen digital, esto en un espacio discreto. Mientras que la metodología propuesta trabaja con el contorno de la región y posteriormente se trabaja en un espacio continuo. Esta elección del espacio continuo, busca caracterizar el contorno de cada muestra cutánea.

Las formaciones cutáneas a clasificar son representadas mediante polinomios de base radial, entre los que se incluyen los de Zernike, Jacobi, Chebyshev [6]. Esta representación es posible utilizando el método de Gram-Schmidt dual [7]. A partir de esta idea, es posible generar descripciones de las muestras cutáneas y posteriormente realizar el proceso de clasificación.

La importancia de ésta representación radica en los siguientes puntos: **(a)** Los polinomios de base radial son de naturaleza ortogonal. **(b)** La ortogonalidad proporciona una combinación lineal de polinomios no redundantes, ya que su construcción no es dependiente de otros términos calculados [2].

Las aportaciones de la metodología propuesta son: **(a)** Caracterizar contornos mediante Combinaciones Polinomiales de Base Radial empleando la metodología de Gram-Schmidt dual de los cuales es posible extraer características tales como área, perímetro, compacidad, combinación polinomial y grado polinomial de ajuste. **(b)** La metodología aquí expuesta no solo es aplicable en la caracterización de melanomas, es aplicable en la caracterización de contornos cerrados. **(c)** Presentar un análisis estadístico de la metodología propuesta evaluándola contra los momentos de Hu empleando la prueba estadística no paramétrica de suma de rangos con signo de Wilcoxon.

Los alcances y limitaciones del presente trabajo son: **(a)** Se busca caracterizar una lesión cutánea mediante su contorno y obtener una descripción de B = Borde irregular del ABCDE en el diagnóstico de un melanoma. **(b)** No se consideran en este trabajo las descripciones de A = Asimetría, C = Coloración, D = Diámetro y E = evolución. **(c)** El proceso de clasificación evalúa los descriptores obtenidos mediante los momentos de Hu y la metodología propuesta solamente en el caso biclase, donde se clasifica cada patrón en dos posibles clases: melanoma ó lesión cutánea.

El presente texto está organizado de la siguiente manera: En la sección 2 se habla de la metodología propuesta para obtener los descriptores, en la sección 3 se mencionan los momentos invariantes de Hu, en la sección 4 se comenta sobre la etapa de clasificación y sobre los parámetros necesarios para cada clasificador, en la sección 5 se muestran los experimentos y resultados obtenidos y en la sección 6 se muestran las conclusiones del escrito.

1.1. Estado del arte

Por el lado de los descriptores de forma básicos, son conocidos los de área perímetro, circularidad y compacidad [3][5], los cuales son extraídos de las regiones de interés en imágenes digitales previamente procesadas. Otro tipo de descripciones empleadas para la extracción de características son los códigos de cadenas, mediante los cuales es común implementar los descriptores de Fourier de manera más eficiente que una representación matricial extraída de una imagen digital [8]. Otras técnicas ampliamente conocidas son las basadas en Momentos, de los cuales algunos de ellos son los Momentos Invariantes de Hu, Momentos de Zernike y Momentos de Chebyshev [27], y en [2] se muestran las fortalezas e implicaciones de los 3 tipos de Momentos mencionados. También en [9] se muestran un análisis sobre la invariancia de los Momentos de Zernike aplicado como descriptores de forma. En los trabajos donde se aborda el reconocimiento de melanomas se encuentra [10] donde aborda el tema de la segmentación de imágenes y muestra un comparativo de clasificaciones. En [11] se hace uso de redes neuronales y usando descriptores para el índice de irregularidad, asimetría y color. En [17] se presenta El uso de momentos de Zernike de orden 32 para caracterizar y clasificar mamografías.

2. Metodología

La metodología propuesta para el cálculo de los descriptores se muestra en la figura 1, la descripción de los procesos generales se mencionan a continuación.

2.1. Imágenes

Es la base de datos que consta de imágenes de lesiones cutáneas. La metodología requiere que cada *muestra* de lesiones cutáneas sea previamente binarizada, simplificando las regiones de interés en color negro sobre un fondo blanco. Cada muestra será sometida a la metodología que se describe en esta sección

2.2. Extracción del contorno

De cada muestra binaria se extrae el contorno de la región de interés representada por píxeles negros. Este Proceso es realizado bajo la métrica de conectividad 4 [19]. En la figura 2 se muestra en (a) la región binaria de una lesión cutánea, y posteriormente en (b) se muestra el contorno extraído de la región.

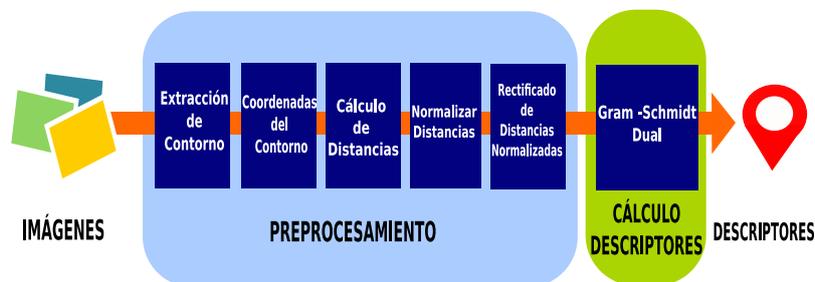


Fig. 1. Fases para obtener los Descriptores según la metodología propuesta.

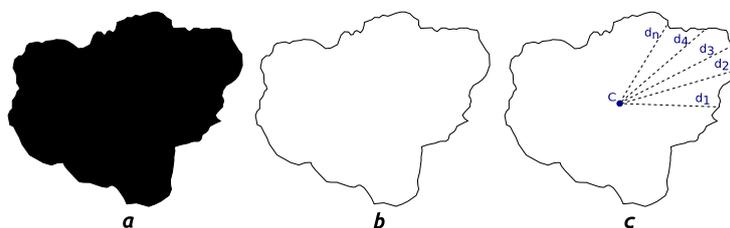


Fig. 2. Proceso de obtención del contorno de una imagen digital binarizada. (a) Región de lesión cutánea. (b) Contorno de Lesión cutánea. (c) Cálculo de distancias del centroide a cada píxel que conforma el contorno.

2.3. Creación de coordenadas a partir del contorno

Por cada muestra se interpreta cada píxel $pix(x, y)$ como una coordenada (x, y) . Una imagen se representa como una matriz de píxeles discretos con posición (x, y) cuyos valores x y y son enteros positivos. Para esta fase se procedió de la siguiente manera:

1. Se localiza aproximadamente el píxel central de la imagen digital.
2. Se busca el primer píxel $pix_1(x, y)$ del contorno que se encuentra a la derecha del píxel central previamente encontrado y se crea la coordenada (x, y) correspondiente.
3. Se identifican sucesivamente todos los píxeles $pix_n(x, y)$ en sentido antihorario a partir del primer píxel identificado del contorno y se genera para cada píxel la coordenada (x, y) correspondiente. Dos posibles alternativas para lograr esta fase, pueden ser el algoritmo de Papert [15] o alguna técnica de búsqueda Heurística [16].

Al final de este proceso se tiene una serie de coordenadas (x, y) que representan el contorno de la lesión cutánea.

2.4. Cálculo de distancias

Por muestra se calcula el **centroide** $C(c_x, c_y)$ de las coordenadas del contorno como se observa en las fórmulas descritas en 2:

$$c_x = \sum_{i=1}^N \frac{x_i}{N} \quad c_y = \sum_{i=1}^N \frac{y_i}{N} \quad (2)$$

Donde x_i y y_i son respectivamente los valores x y y de las coordenadas del contorno, y N el número de coordenadas (x, y) . En la figura 2 en (c), se ilustra cómo se calcula la distancia del centroide a cada píxel del contorno. Posteriormente se calcula la Distancia euclidiana desde el centroide $C(c_x, c_y)$ a cada coordenada $P(p_x, p_y)$ duplicando la primera coordenada al final de las mismas, para que las coordenadas representen un contorno cerrado. La fórmula 3 corresponde a la distancia euclidiana d .

$$d(C_{x,y}, P_{x,y}) = \sqrt{(c_x - p_x)^2 + (c_y - p_y)^2} \quad (3)$$

La representación final de estas distancias d son el contorno observado desde un plano x, y . Esta representación es conocida como **signatura ó perfil centroidal** [12], y es equivalente a cortar, en sentido figurado, el borde de la lesión cutánea y extenderlo en el plano x, y pero conservando sus irregularidades. Esta idea puede apreciarse en la figura 3 donde en (a) se muestran graficadas las alturas de las distancias d obtenidas y se aprecia la forma de la representación de perfil centroidal.

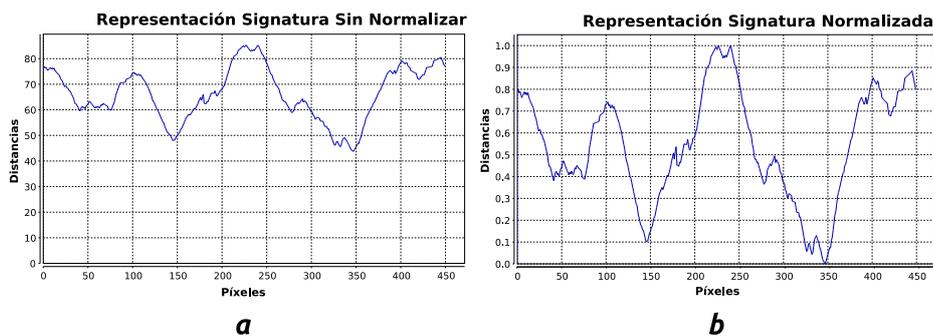


Fig. 3. Representación de *perfil centroidal* ó *signatura* de una lesión cutánea. (a) Representación sin normalizar alturas. (b) Representación con alturas normalizadas en el rango $[0, 1]$.

2.5. Normalización de distancias

Es necesario normalizar las alturas d de cada muestra en el rango $[0, 1]$ para obtener la invariancia de escala entre las muestras de lesiones cutáneas a realizar. El cálculo para esta normalización está en la fórmula 4.

$$d = \frac{d_i - d_{min}}{d_{max} - d_{min}} \quad (4)$$

Donde d es la distancia normalizada, d_i es la distancia a normalizar, d_{max} y d_{min} son respectivamente las distancias máxima y mínima de la serie de distancias d analizadas. En la figura 3 se muestra en (b) las distancias de la representación signatura normalizadas.

2.6. Rectificación de las distancias d de la representación del perfil centroidal

Si las distancias normalizadas de cada muestra son operadas con la metodología de Gram-Schmidt dual, se presentara el fenómeno de Gibbs ocasionando que los grados de ajuste obtenidos con la metodología sean elevados y costosos de calcular computacionalmente. Para evitar este problema se propone la rectificación de la representación del perfil centroidal.

Para hacer equitativa esta rectificación en todas las muestras, se aplica el rectificado determinando una ventana de filtrado, donde el tamaño de esta ventana es proporcional al número de distancias d que representan el contorno de una lesión cutánea, esta proporción ésta expresada en 5.

$$p = \frac{r}{h} \quad (5)$$

Donde p es el **tamaño de clase** y representa la cantidad de datos que componen el tamaño de la ventana de filtrado. r es el intervalo de variación de las distancias d que componen un *perfil centroidal* y lo llamaremos **rango**, donde $r = (d_{max} - d_{min})$. h es el **número de clase** y hace referencia a la cantidad de segmentos en que puede ser dividido el total de las distancias d del contorno. Esta forma de nombrar a las variables de la expresión 5 tienen sentido en un contexto estadístico, donde empleando la regla de Scott [14] es posible determinar el tamaño de ventana buscado. Esto queda formulado en la expresión 6.

$$p = \text{valor entero} \left[\frac{d_{max} - d_{min}}{3,5 s} \right] \sqrt[3]{n} \quad (6)$$

Donde d_{max} y d_{min} son respectivamente el valor máximo y mínimo de los datos a analizar, y estas dos variables representan el rango de las distancias d . S es la desviación estándar de las distancias d y n es la cantidad de distancias d que modela un contorno.

La expresión 6 proporciona el número de distancias d a operar en cada *perfil centroidal*. Se cuida que el tamaño de ventana sea impar para poder calcular la

mediana entre los números que se encuentran en la ventana previamente fija y sustituir el valor central por la *mediana*. Éste paso se aplica a todos los datos desde el primero hasta el último. Estas operaciones se realizan en un espacio Toroidal, para garantizar la rectificación a todos los datos.

2.7. Cálculo de descriptores mediante Gram-Schmidt dual

Cada conjunto de distancias rectificadas que representan el perfil centrolidal de cada muestra, son ingresados uno por uno en la metodología propuesta para obtener los descriptores empleados en este escrito. La metodología del Gram Schmidt dual funciona a grandes rasgos, como una técnica basada en ajustes de datos [20]. Esta metodología trabaja mediante el proceso de ortogonalización de Gram-Schmidt, el cual es empleado para crear una base ortogonal sobre los perfiles centroidales ingresados a la metodología y después desde esa base creada, intenta regresar y ajustarse a los datos ingresados inicialmente. Esta técnica es capaz de operar polinomios ortogonales, tales como Jacobi, Zernike, Chebyshev, entre otros. Durante el proceso de esta técnica, se va proponiendo una combinación polinomial de grado n que se ajuste a los datos ingresados al cálculo, esto de maneja iterativa. Durante cada iteración, el grado polinomial propuesto inicia en $g = 0$, el grado propuesto va incrementándose durante cada iteración y se detiene hasta que cumple un error de paro. El instrumento para determinar el paro es el error cuadrático medio (ECM) y se muestra en la expresión 7. El valor sugerido mediante experimentos preliminares para el error de paro fue seleccionado como $error = 0,00999$. en [20] se muestra la metodología y el modelado formal de la técnica de Gram-Schmidt dual.

$$ECM(g) = \sqrt{\frac{\sum_{i=0}^n w(d_i)[f(d_i) - f'(d_i)]^2}{n - g}} \quad (7)$$

Donde $w(d_i)$ es el peso w calculado para cada distancia, a estos pesos pueden asignarseles el valor de 1 para una evaluación donde se desprecia el peso w . $f(d_i)$ es la combinación polinomial de ajuste propuesto por la metodología en cada grado, $f'(d_i)$ son las distancias que representan un contorno y se buscan ajustar mediante la combinación polinomial propuesta por $f(d_i)$ en cada grado, n es el número distancias que componen el contorno a ajustar y g es el grado polinomial $g = 0, 1, 2, \dots, grado$.

2.8. Descriptores obtenidos del proceso

Para obtener los descriptores mediante la metodología de Gram-Schmidt dual es necesario asignar valores a los parámetros α , β en los polinomios de Jacobi [6]. En el presente trabajo se emplearon los siguientes: $\alpha = -0,5$, $\beta = -0,5$, $error\ de\ paro = 0,00999$. Los parámetros se emplean en la expresión 8.

$$f(d) = \sum_{j=0}^g a_j P_j^{(\alpha, \beta)}(d), \quad \alpha, \beta > -1 \quad (8)$$

En la expresión 8 se muestra la combinación polinomial de los coeficientes a_j y los polinomios de Jacobi $P_j^{(\alpha, \beta)}(d)$, donde $j = 0$ hasta $g = \text{grado del ajuste}$.

Los descriptores obtenidos para cada muestra de lesiones cutáneas mediante Gram-Schmidt dual son los siguientes:

1. **Combinación polinomial.** Están descritas por la expresión 8 que se desarrolla hasta el número del mejor grado de ajuste. El número de términos de un ajuste de grado g sera de $g + 1$, esto porque el grado 0 también se toma en cuenta.
2. **Grado polinomial.** Es el grado que alcanza el ajuste mediante la metodología.
3. **Área.** Es obtenido mediante la integración de las combinaciones lineales polinomiales obtenidas de la expresión 8. En la expresión 9 se muestra la fórmula A a integrar. Esto puede realizarse de manera numérica con métodos de integración como los son el método compuesto ó Romberg [21].
4. **Perímetro.** Se calcula con la integral de línea donde la función a integrar es la combinación polinomial obtenida con el método. En la expresión 9 se muestra la fórmula de la integral de línea P [31].
5. **Compacidad clásica.** Se encuentra en [3] y es calculada conociendo los valores de área y perímetro. Esta descrita como C en la expresión 9.

$$A = \int_a^b f(d) dd, \quad P = \int_a^b \sqrt{1 + f'(d)^2} dd, \quad C = P^2/A \quad (9)$$

3. Comparativa entre descripción propuesta y momentos de Hu

La metodología propuesta para caracterizar los contornos de una lesión cutánea, es comparada con los momentos de Hu, esto porque ambas técnicas son similares en cuanto a la invariancia de traslación y escala.

3.1. Momentos de Hu

Los momentos de Hu [19] se emplean para contrastar la metodología desarrollada. Estos momentos son calculados directamente de la región de píxeles que es el área de interés. Para poder calcular estos momentos es necesario primeramente calcular los momentos invariantes [25], posteriormente calcular los momentos centrales [19] o trasladar la región de interés al origen, que es equivalente al cálculo de los momentos centrales. Después se calculan los momentos centrales normalizados [30]. Tras haber obtenido estos cálculos, con los momentos de orden 2 y 3 son calculados los siete momentos invariantes a traslación, rotación y escala propuestos por Hu [26].

4. Etapa de clasificación

En esta fase, se emplean los clasificadores supervisados de Distancia Mínima [22], KNN [22], Naive Bayes [23] y CAP [24] para poder evaluar el índice de aciertos obtenidos. Los clasificadores supervisados pueden conceptualizarse en dos etapas:

1. **Entrenamiento.** Se seleccionan aleatoriamente un número proporcional de patrones por clase para el entrenamiento de cada clasificador. Para la generación aleatoria de los números se emplea la técnica descrita en [29].
2. **Clasificación.** Cada patrón no seleccionado para el entrenamiento es sometido a evaluación y se determina la clase de pertenencia. Los aciertos o fallos de la clasificación se contabilizan mediante la matriz de error [13].

Sobre los clasificadores empleados se hacen las siguientes observaciones:

1. En el caso de los clasificadores Distancia mínima y CAP, son de naturaleza binaria donde se evalúan dos clases.
2. En el caso de KNN y Naive Bayes, son clasificadores multiclase, es decir, admiten en sus evaluaciones más de dos clases.
3. El clasificador Naive Bayes es de naturaleza probabilística ya que evalúa la probabilidad de pertenencia de un patrón a clasificar a una clase.
4. El clasificador KNN se evalúa asignando un valor a k entero, mayor a cero e impar. KNN agrupará un patrón entrante a una clase cuando se contabilicen los k vecinos más cercanos a una clase y se emplea con los valores $k = 1, 3, 5$.

5. Experimentos y resultados

En esta sección se presentan los experimentos y resultados obtenidos, así como el análisis estadístico de la metodología propuesta y los momentos de Hu.

5.1. Base de datos empleada

La base de Datos consta de 30 muestras de tipo Lesión Cutánea y 40 muestras diagnosticadas como melanoma, y fue tomada de [18] donde las imágenes ya se encontraban segmentadas y binarizadas aplicando los siguientes filtros: filtro Gaussiano, conversión a escala de grises y segmentación por algoritmo Otsu.

5.2. Definición de experimentos

La descripción de la formulación del experimento se muestra a continuación.

1. Son calculados los descriptores para todas las muestras de imágenes de lesiones cutáneas, para la metodología del Gram-Schmidt dual y los momentos de Hu. Los vectores característicos de la metodología contienen cuatro características: grado polinomial, área, perímetro, compacidad clásica. Los vectores de los momentos de Hu contienen siete características.

- Ya que se tienen los vectores de características para cada imagen como se muestra en la expresión 1, se emplean los vectores en la fase de clasificación donde son evaluados. En la tabla 1 se pueden observar los parámetros empleados en cada clasificador.

Tabla 1. Parámetros para los clasificadores empleados.

Clasificador	% Patrones entrenamiento	Repeticiones	k
D-Min	40 %	100	N/A
KNN	40 %	100	1, 3, 5
Navie Bayes	40 %	100	N/A
CAP	40 %	100	N/A

- En la tabla 2 se reportan los porcentajes de la precisión total de aciertos de las clasificaciones. Éste cálculo es obtenido de la matriz de confusión [13].
- Se aplica el análisis estadístico no paramétrico de suma de rangos con signo de Wilcoxon, donde se comparan los porcentajes generales de aciertos para los resultados de las clasificaciones en la metodología propuesta y los momentos de Hu.

5.3. Resultados

Los resultados de evaluar la metodología propuesta y los momentos de Hu se presentan en la tabla 2. Los parámetros empleados para cada clasificador esta descrito en la sección 5.2.

Los resultados aquí mostrados se obtuvieron de la siguiente manera: **(a)** Se calcularon los descriptores y se crearon los vectores de características para los momentos Hu y la metodología propuesta empleando cada una de las 70 muestras de la base de datos de imágenes. **(b)** Por cada clasificador empleado se tomo aleatoriamente el 40 % de los patrones de las dos clases (melanoma y lesión cutánea) para realizar la fase de entrenamiento de cada clasificador. Los patrones de las dos clases que no se usaron en la fase de entrenamiento se emplearon para la fase de clasificación correspondiente a cada clasificador empleado. Este proceso se repitió 100 veces por clasificador para los patrones calculados con los descriptores de los momentos de Hu y la metodología propuesta. **(c)** En cada una de las 100 veces que se operan los clasificadores empleados para los patrones obtenido mediante la metodología propuesta y los momentos de Hu, se reporta el porcentaje total de clasificaciones correctas. Dicho cálculo se obtiene de la matriz de confusión [13], en la cual se reporta por clase los patrones clasificados correctamente ó erróneamente. En la tabla 3 se muestra la estructura para una matriz de confusión donde se evalúan 2 clases, donde: A = verdadero positivo de clase 1, B = falso positivo clase 1, C = verdadero positivo clase 2, D = falso positivo clase 2. El porcentaje total de aciertos (PA) de una matriz de confusión de dos clases se obtiene de la siguiente manera: $PA = \left[\frac{A+D}{(A+B+C+D)} * 100 \right]$. **(d)**

Ya que se obtuvieron los 100 porcentajes de acierto totales por clasificador tanto para los patrones obtenidos de los momentos de Hu y la metodología propuesta, en la tabla 2 se reportan para cada clasificador los siguientes datos: porcentaje máximo, porcentaje mínimo, porcentaje promedio y mediana de porcentajes. (e) Se realiza la prueba estadística de suma de rangos con signos de Wilcoxon empleando los valores de las medianas calculadas para cada clasificador.

Tabla 2. Clasificaciones de la metodología propuesta y momentos de Hu.

Descriptores	Clasificador	% Máximo	% Mínimo	% Promedio	Mediana	σ
Metodología Propuesta	D-Min	69.04	33.33	54.90	54.76	6.94
	1NN	73.80	42.85	59.97	60.71	6.92
	3NN	69.04	35.71	56.30	57.14	6.97
	5NN	69.04	35.71	55.80	57.14	7.26
	Navie Bayes	75.00	39.28	55.60	53.57	9.13
	CAP	69.04	33.33	54.57	54.76	6.81
Momentos HU	D-Min	64.28	42.85	48.97	45.23	5.72
	1NN	78.57	52.38	68.59	69.04	6.47
	3NN	80.95	45.23	68.76	69.04	7.18
	5NN	83.33	47.61	64.16	64.28	7.76
	Navie Bayes	75.00	50.00	61.25	60.71	5.40
	CAP	69.04	42.85	47.85	45.23	5.40

Tabla 3. Ejemplo de una matriz de confusión para dos clases.

	clase 1	clase 2
clase 1	A	B
clase 2	C	D

5.4. Prueba de suma de rangos con signos de Wilcoxon

La prueba de Wilcoxon [28] consta de tener n observaciones aparejadas de la forma (x_i, y_i) y que $D_i = (x_i - y_i)$. Se busca probar la hipótesis (h_0) de que las x y las y tienen la misma distribución, frente la alternativa de que las distribuciones difieren de ubicación (h_a). Se realiza una prueba de 2 colas, donde se busca obtener $W = \min(W^+, W^-)$, donde W^+ es la suma de los rangos de las diferencias positivas y W^- es la suma de los rangos de las diferencias negativas. Esta prueba se empleó para evaluar si existe o no diferencias entre las distribuciones de las n observaciones de los descriptores a evaluar. Se emplean las medianas de las 100 repeticiones por clasificador para ambos descriptores, ver tabla 2. El nivel de significancia empleado fue de 0,05.

Comparando la metodología propuesta y los momentos de Hu, para la prueba Wilcoxon de dos colas, se tiene que $W^+ = 9, W^- = 11$ $W = \min(W^+, W^-) = (9, 11) = 9$. El número de instancias evaluadas es 6 y el valor correspondiente según las tablas de Wilcoxon para $W_0 = 1$, posteriormente se verifica $W \leq W_0$, quedando como $9 \leq 1$. En este caso se acepta la Hipótesis H_0 por lo que se concluye de ésta prueba estadística que: "Las distribuciones de frecuencias coinciden en posición".

5.5. Discusión

El resultado de la prueba de suma de rangos con signos de Wilcoxon, da evidencia estadística de que la metodología propuesta y los momentos de Hu, son similares.

Se destaca que la metodología propuesta sólo es invariante a traslación y escala, y a pesar de esto, logra obtener resultados de clasificación comparables a los obtenidos con los momentos de Hu, que son invariantes a rotación, traslación y escala. La ventaja que posee la metodología propuesta es que solo emplea vectores con cuatro características (grado polinomial, área, perímetro y compacidad clásica) en comparación con los momentos invariantes de Hu, que emplea siete características.

6. Conclusiones y trabajo futuro

De la prueba estadística de Wilcoxon se obtiene que la distribución de frecuencias coinciden en posición, por tanto la prueba no puede hacer distinción entre los valores de las medias obtenidas. En este caso los descriptores propuestos compiten en desempeño con los descriptores de Hu, basándonos en la conclusión estadística.

La metodología propuesta presenta oportunidades de mejora: al realizar la interpretación del *perfil centroidal* de cada imagen, dependiendo de donde se inicie la selección del contorno de una lesión cutánea, se pueden obtener diversos perfiles para una misma figura, y por ende diferentes valores de los coeficientes de ajuste a_j en la ecuación 8, esta se puede subsanar realizando una fase previa para hacer invariante ante rotación a cada muestra. Otra área de oportunidad se encuentra en la sintonización de los coeficientes α, β que se emplean en el cálculo de los polinomios ortogonales. Las mejoras y el trabajo futuro incluyen:

1. Buscar la invariancia de rotación para cada muestra analizada antes de calcular los descriptores mediante la metodología Gram-Schmidt dual.
2. Plantear alguna técnica heurística ó metaheurística para sintonizar los valores α, β necesarios para calcular los descriptores de base radial dentro de la metodología.
3. Complementar los descriptores empleados en la metodología propuesta con descriptores que incluyan la asimetría, coloración irregular, diámetro y evolución con el objetivo de completar la metodología del ABCDE para el diagnostico de melanomas.

4. Realizar un comparativo en la calidad de las clasificaciones de la metodología propuesta variando los porcentajes de patrones para la fase de entrenamiento en los clasificadores.
5. Evaluar la metodología propuesta en un entorno multiclase para poder determinar las fortalezas y debilidades de la misma.

Agradecimientos. Agradecemos al Consejo de Ciencia y Tecnología (CONACYT) por el apoyo para realizar la presente investigación y a la Dirección General de Educación Superior Tecnológica (DGEST).

Referencias

1. González, N., Flores, A.: El melanoma en México. *R. Esp. Méd.* 15, 161–164 (2010)
2. Celebi, M.E., Aslandogan, T.A.: A comparative study of three Moment-Based Shape Descriptors. In: IEEE Computer society, Proceeding of the international conference on information technology: Coding and computing (ITCC'05). pp. 788–793. IEEE Press, USA (2005)
3. Montero, R., Bribiesca, R.: State of the Art of Compactness and Circularity Measures. In: International Mathematical Forum, pp.1305–1335. Hikari ltd. Press, USA (2009)
4. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press, Canada (2009)
5. Sossa, J.: Rasgos descriptores para el reconocimiento de objetos. Centro de investigación en computación Instituto Politécnico Nacional, México (2006)
6. Abramowitz, M., Stegun, I.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, USA (1965)
7. Carpio, Valadez, J., M., Malacara, Hernández, D.: Ajuste de datos interferométricos con polinomios de Zernike. Tesis de Maestría del Centro de Investigaciones en Óptica A. C. Tesis Maestría, CIO Press (1987)
8. Nixon, M., Aguado, A.: Feature Extraction And Image Processing. Elsevier, London (2008)
9. Bin, Y., Peng, J., X.: Improvement and invariance analysis of Zernike moments using as a region-based shape descriptor. In: XV Brazilian Symposium on Computer Graphics and Image Processing. Proceedings. pp.120–127. IEEE Press, Brazil (2002)
10. La Torre, E., Caputo, B, Tommasi, T.: Learning methods for melanoma recognition. *Int. J. Im. Sys. Tech.* 20, 316–322 (2010)
11. Ercal, F., Chawla, A., Stoecker, W.: Neural network diagnosis of malignant melanoma from color images. *IEEE Trans. Bio. Eng.* 41, 837–845 (1994)
12. Davis, E., R.: Computer and machine vision: Theory, algorithms, practicalities. Elsevier Academic Press, USA (2012)
13. Kohavi, R., Provost, F.: On Applied Research in Machine Learning. Machine Learning, Kluwer Academic Publishers (1998)
14. Scott, D., W.: On optimal and data-based histograms. *Bio.* 66, 60–610 (1979)
15. Pitas, L.: Digital Image Processing: Algorithms and applications. Wiley, Canada (2000)
16. Russel, S., Norving, P.: Artificial Intelligence. A modern approach. Prentice Hall, USA (2010)

17. Tahmasbi, A., Saki, F., Shokouhi, S. B.: Clasificación de benign and malignant masses based on Zernike moments. *Comp. Bio. Med.* 41, 726–735 (2011)
18. Barajas Saavedra, L.N., et al.: Reducción del vector de características para la clasificación de vectores mediante imágenes binarias. *CIINDET 2014 (En prensa)*, IEEE Press (2014)
19. Sossa, J.: *Visión Artificial, Rasgos Descriptores para el Reconocimiento de Objetos*. RA-MA. España (2013)
20. Malacara, D., Carpio, J., M., Sánchez, J.: Wavefront fitting with discrete orthogonal polynomials in a unit radius circle. *SPIE Proc. Op. Eng.* 29, 672–675 (1990)
21. Chapra, R., C., Canale, R., P.: *Numerical Methods for engineers*. McGraw Hill, USA (2009)
22. Friedman, M., Kandel, A.: *Introduction to pattern recognition: Statistical, structural, neural and fuzzy logic approaches*. World Cientific Publishing Co., London (1999)
23. Teodoridis, S., Koutroumbas, K.: *Pattern Recognition*. Academic Press, USA (2006)
24. Montero, R.: *Clasificador híbrido de patrones basado en la Lernmatrix de Steinbuch y el línea asociator de Anderson-Kohonen*. Tesis de Maestría del Instituto Politécnico Nacional. Centro de investigación en computación. Tesis Maestria, IPN Press (2003)
25. Pajares, G., De la Cruz, J., M.: *Visión por computador: Imágenes digitales y aplicaciones*. Alfaomega RA-MA, España (2002)
26. Hu, M., K.: Visual Pattern recognition by Moment Invariants. *IEEE Trans. inf. theo.* 8, 179–187 (1962)
27. Xiao, B., Ma, J., F., Cui J., T.: Radial Tchebichef moment invariants for image recognition. *J. Vis. Commun. Imagen R.* 23, 381–386 (2012)
28. Wilcoxon, F.: Individual Comparisons by Ranking Methods. *Bio.* 1, 80–83 (1945)
29. Carpio, M., Soria-Alcaraz J. et al.: Variable length number chains generation without repetitions. *Soft Comp. Rec. Based Bio. Studies Comp. Int.* 312, 349–364 (2010)
30. Nixon, M., Aguado, A.: *Feature Extraction & Image Processing*. Academic Press, Ungaria (2008)
31. Bers, L., Karal, F.: *Cálculo*. Interamericana, México (1978)

Implementación de un sistema de reconocimiento de imágenes por contenido usando algoritmos genéticos

Juan Villegas-Cortez¹, Yolanda Pérez-Pimentel², Ismael Osuna-Galán²

¹ Departamento de Electrónica, Universidad Autónoma Metropolitana, Azcapotzalco, D.F., México

² Ingeniería en Mecatrónica, Universidad Politécnica de Chiapas, Tuxtla Gutiérrez, Chiapas, México

jvillegas@gmail.com, {ypimentel, iosuna}@upchiapas.edu.mx

Resumen. La recuperación de imágenes basada en su contenido propio (CBIR: Content-Based Image Retrieval), describe una serie de técnicas y métodos que utilizan contenidos visuales extraídos de las propias imágenes a estudiar, para buscar o clasificar imágenes de acuerdo con los intereses de un usuario o sistema; ésta ha sido un área de investigación con notables progresos en la investigación teórica y ya ha sido aplicada en sistemas de análisis de bases de datos de imágenes via software. En la actualidad, conviven gran cantidad de métodos y se han integrado técnicas más complejas con el fin de hacer sistemas CBIR más eficientes. En este artículo se presenta el desarrollo y la implementación de un sistema CBIR de análisis estadístico local de textura sobre LabVIEW, orientado a implementarse en hardware, dicho sistema emplea algoritmos genéticos para realizar la clasificación de las imágenes.

Palabras clave: Recuperación de imágenes basada en contenido, CBIR, algoritmos genéticos, K-Means, LabVIEW.

1. Introducción

La recuperación de información es el proceso de convertir una solicitud de información en un conjunto significativo de referencias. Los primeros trabajos sobre la recuperación de información aplicada a imágenes se puede remontar a finales de 1980. Desde entonces, el potencial de aplicación de las técnicas en bases de datos ha atraído la atención de los investigadores[6]. Las primeras técnicas no eran, en general, sobre la base de las características visuales, sino en metadatos extraídos en las propiedades de las imágenes tales como anotación al margen o el nombre de las imágenes, a estas técnicas se le llaman “de etiquetado”. En otras palabras, las imágenes eran localizadas y caracterizadas utilizando un enfoque basado en texto con sistemas de gestión de etiquetas organizadas en bases de datos tradicionales.

A principios de 1990, como resultado de los avances en la Internet y las nuevas tecnologías en cámaras digitales, la creación de bases de datos de imágenes

digitales de gran volumen ha crecido enormemente, de la mano de la masificación de uso de las cámaras digitales de bolsillo. Las dificultades que enfrentó la recuperación basada en texto se hizo más difícil y producía resultados insatisfactorios al principio, pero fue perfeccionándose hasta estar prácticamente agotada al día de hoy, obteniendo muchas veces resultados irrelevantes dado que a sólo usa el etiquetado, que no siempre contempla el contenido total de la imagen, o bien no contempla detalles finos, por ello la necesidad de clasificar o realizar la búsqueda de imágenes fue la fuerza motriz detrás de la aparición de las técnicas de recuperación de imágenes basadas en contenido o CBIR (*Content-Based Image Retrieval*) por sus siglas en inglés. La investigación sobre los métodos de extracción, organización e indexación de información visual han aumentado del mismo modo que una creciente demanda para fines comerciales [6].

Actualmente los sistemas CBIR se apoyan en diversas disciplinas para realizar la clasificación. Por un lado se requiere una forma “inteligente” para identificar las imágenes, y también una forma para realizar la clasificación en grupos o clases afines. Para realizar la tarea de identificación, una técnica es la de algoritmos genéticos (AG), hay varias aplicaciones de AG para la recuperación de información con especial énfasis en la de retroalimentación proveniente del usuario. En este caso el AG aplicado es un algoritmo matemático que transforma un conjunto de propiedades de las imágenes de forma individual, considerándola como una propiedad dinámica con respecto al tiempo y usando operaciones modeladas de acuerdo al principio Darwiniano de evolución, contemplando la reproducción y supervivencia del individuo, como una posible solución al problema, más apto. Cada uno de estos objetos matemáticos suele ser una matriz de propiedades que se ajusta al modelo de las cadenas de cromosomas, y se les asocia con una cierta función que refleja su desempeño o aptitud. Estos individuos con las características ideales se almacenan para la comparación de los datos extraídos de las imágenes y tomar una decisión.

En este artículo mostramos el desarrollo e implementación de un sistema CBIR, orientado a imágenes de escenarios naturales. Dicha implantación se realizó en LabVIEW y la misma está orientada a trabajar sobre hardware (FPGA). El trabajo se desarrolla proporcionando el Estado del arte, para posteriormente en la sección de Desarrollo explicar la metodología de la extracción de las características sobre las imágenes digitales, a partir de descriptores estadísticos; posteriormente se analiza el proceso de agrupación de las imágenes por medio del algoritmo *K*-means, con ello se proporcionan los resultados experimentales obtenidos y finalmente se comparten las conclusiones y el trabajo futuro de este trabajo.

2. Estado del arte

A principios de los años 90's, aparece por primera vez mencionada la técnica CBIR en un trabajo de T. Kato, el término CBIR fue utilizado para describir a un sistema que recuperaba imágenes de una base de datos basándose en el color y la forma, pero parece que hasta inicios de la década del 2000 se tiene precisión en

el concepto orientado hacia la extracción automática de características, así como a la representación de los datos. Las técnicas CBIR usan características de bajo nivel, e.g. texturas, color y forma para representar a las imágenes.

Por otro lado, el gran reto actualmente es el reconocimiento y la clasificación de imágenes en grandes volúmenes, y el mejor ejemplo de esto es la Internet, que se ha convertido en un lugar de confluencia de diversos tipos de imágenes (rostros, coches, dispositivos electrónicos, mapas, paisajes, flora, fauna, etc.). Cada tipo de imagen tiene sus retos a resolver en las dos tareas de fondo, el reconocimiento y la clasificación, más allá del propósito final del usuario de la consulta, de ahí que generalmente las aplicaciones de CBIR se aterrizan hacia un tipo de imágenes, siendo las de escenarios naturales un tipo muy complicado por la mezcla de colores y formas no regulares. La forma como tradicionalmente de atacó el problema de la organización de imágenes fue a partir de la generación de “etiquetas” asociadas a cada una. Es así que se generaban bases de datos de registros de etiquetas ligadas a cada imagen, pudiendo una imagen tener tantas etiquetas como conceptos se le puedan asociar. Esta técnica se ha automatizado en motores de búsqueda usando autómatas, tal que al analizar una página web con imágenes, se hace un análisis estadístico de las palabras contenidas en texto de la página, y de forma automática se asocian las palabras con mayores frecuencias, como etiquetas de las imágenes de esa página. Esta técnica ha sido muy probada y ampliamente usada, pero se tienen resultados irrelevantes, i.e. que al buscar imágenes de un concepto o palabra el buscador nos regresa imágenes no tienen relación con lo buscado. Otra desventaja directa de la técnica de etiquetado es el idioma de las etiquetas, teniendo que usar un proceso adicional de traducción de las mismas a otros idiomas.

Uno de los pioneros en aplicación de CBIR fue IBM, que patentó el sistema Query By Image Content en el año 1995. Con el paso de los años, fueron apareciendo más sistemas con diferentes variantes, entre los cuales se destacan Photobook realizado en el MIT, Blobworld desarrollado en UC Berkeley, SIMBA, FIRE entre otros [6]. Con el auge y masificación de las imágenes digitales en la Internet aparecieron buscadores de imágenes como Webseer en el año 1996 y Webseek en el año 1997. A su vez, hubo desarrollos importantes en motores de búsqueda de bases de datos relacionales como IBM DB2 y Oracle, donde se incluyeron herramientas a sus productos para facilitar la recuperación de imágenes por contenido visual, acercando el área de CBIR al ámbito de la industria.

En este trabajo presentamos un sistema CBIR, basado en la extracción de características de bajo nivel desarrollado por [6], que realiza de forma automática la detección y clasificación de las imágenes sin la necesidad de un etiquetado, y brindando un enfoque estadístico de análisis de la textura local de ventanas que se abren a partir de puntos de la imagen a estudiar. Otros enfoques de estudio de la imagen, local y global, se han realizado y se pueden analizar su diferentes niveles de desempeño a partir del tipo de imagen, grado de iluminación, rotación, etc. Una detallada revisión de los avances en los últimos años sobre CBIR puede ser consultada en [2], así como en [6].

3. Desarrollo

Para poder desarrollar un sistema de reconocimiento por contenido de imágenes, es necesario contar con bases de datos de imágenes de un tipo o propósito para el entrenamiento del sistema. En ésta fase pueden utilizarse bases de datos estándar y de libre uso como la de Torralba y Vogel. Realizado el entrenamiento se puede proceder con las pruebas, primero con las mismas imágenes con que fue entrenado el sistema, y posteriormente probando con nuevas imágenes del tipo o propósito deseado.

3.1. Extracción de características

El proceso de extracción de características desarrollado en el presente trabajo se divide en cuatro subprocesos, todo visto como una metodología: (i) extracción de Información de la imagen en espacio de color RGB, (ii) conversión de los datos al espacio de color HSI (para tener más información a bajo nivel), (iii) selección de muestras en ventanas de 10×10 píxeles, y (iv) cálculo valores estadísticos asociados a las texturas seleccionadas (la media (μ), desviación estándar (σ) y *homogeneidad*); tal como se muestra en la Fig. 1.

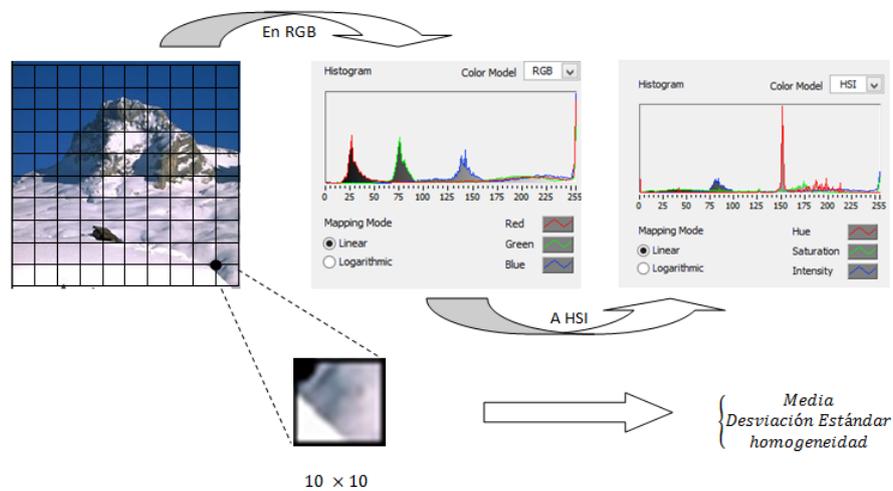


Fig. 1. Extracción de características de la imagen.

Proporcionando más detalle de la metodología, una vez determinada una base de datos con la cual trabajar, se realiza la extracción de información de cada imagen en valores de espacio RGB. Las imágenes de las bases de datos procesadas tienen un tamaño de 256×256 píxeles. La información que se obtiene de cada imagen se trabaja en matriz-vector, teniendo un vector unidimensional

de 196608 elementos. En éste vector, se encuentran tres datos por cada pixel, uno para el canal R, uno para el G y otro para el B, iniciando con el $pixel_0$ hasta el $pixel_n$ de la imagen. Para facilitar el manejo de los datos, se redimensiona dicho vector a una matriz de 65536×3 elementos, teniendo en cada fila los datos de los canales RGB correspondientes a cada $pixel_i$.

En esta parte, los valores de la imagen aún están expresados en RGB, pero para obtener mejores resultados, deben ser convertidos a valores HSI (*Hue, Saturation, Intensity*).

El primer paso para la conversión consiste en normalizar los valores RGB,

$$r = \frac{R}{R + G + B}, g = \frac{G}{R + G + B}, b = \frac{B}{R + G + B}$$

Realizada la normalización, los componentes HSI se obtienen mediante:

$$h = \cos^{-1} \left\{ \frac{0.5[(r - g) + (r - b)]}{[(r - g)^2 + (r - b)(g - b)]^{1/2}} \right\} h \in [0, \pi] \text{ for } b \leq g \quad (1)$$

$$h = 2\pi - \cos^{-1} \left\{ \frac{0.5[(r - g) + (r - b)]}{[(r - g)^2 + (r - b)(g - b)]^{1/2}} \right\} h \in [\pi, 2\pi] \text{ for } b > g \quad (2)$$

$$s = 1 - 3 \cdot \min(r, g, b) \quad s \in [0, 1] \quad (3)$$

$$i = \frac{R + G + B}{3 \cdot 255} \quad i \in [0, 1] \quad (4)$$

Por conveniencia para valores de colores neutros tales como negro, blanco y gris, en dónde los valores para R, G y B son iguales, se considera $H = 0$. De igual manera, los valores H , S e I se convierten en los siguientes rangos:

$$H = \frac{h \times 180}{\pi} \quad (5)$$

$$S = s \times 100 \quad (6)$$

$$I = i \times 255 \quad (7)$$

El siguiente paso, es la toma de muestras. Para el procesamiento de la imagen no se requiere utilizar el total de los datos extraídos de cada imagen. En vez de eso, se decide tomar 100 muestras con una distribución uniforme de la imagen. Por ello, la imagen se trata como si fuera una cuadrícula uniforme de 10×10 . En el origen de cada cuadro, se abre una ventana de 10×10 pixeles que conforma la muestra y se integra a una nueva matriz cuyo tamaño final es de 10000×3 , la cual se encuentra ya expresada en valores HSI. Para cada ventana se obtienen, la media, desviación estándar y la homogeneidad (de la matriz de co-ocurrencia), la cuál está dada por:

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (8)$$

Ahora, se tiene de cada imagen, una matriz de 100×9 , en los que cada renglón corresponde a los rasgos característicos de cada pixel correspondiente a una muestra de de 10×10 pixeles, conformado de la siguiente manera,

$$\{H_{\mu}, H_{\sigma}, H_{homogeneidad}, S_{\mu}, S_{\sigma}, S_{homogeneidad}, I_{\mu}, I_{\sigma}, I_{homogeneidad}\}$$

3.2. Agrupación

Habiendo obtenido la matriz de características de cada imagen se procede a la elección de los centroides. El centroe no es sino un representante que por sus características es el adecuado para cada clase, es decir para cada grupo. Para la definición de los centroides se utiliza un AG estándar de estado estacionario con una estrategia de selección elitista, en donde los elementos con mejores candidatos de cada iteración se mantienen a la siguiente iteración y es invariante bajo permutaciones. La operación de mutación se aplica como es habitual en los candidatos elegidos al azar, como se muestra en la Fig. 2.



Fig. 2. Selección de Centroides usando AG

3.3. Algoritmo genético

La primera generación se elige aleatoriamente, los K elementos elegidos de los grupos $\{C_1, C_2, \dots, C_K\}$ respectivamente, definidos en la base de datos de entrenamiento. Consideremos la j -ésima iteración $\mathcal{P}_j = C_1, C_2, \dots, C_K$ donde

C_i es un centroide que corresponde a la matriz de características de una imagen de la base de datos. Se toma una matriz B correspondiente a una imagen en la base de datos de entrenamiento, se calcula la matriz $D_i^B = |C_i - B|$ para $i = 1, 2, \dots, K$ y se define la distancia

$$d(C_i, B) = \frac{\sum d_{ij}}{m \times n}$$

donde $D_i^n = (d_{ij})$ con tamaño $m \times n$.

Se considera al índice

$$i^* = \{ i \mid \text{el valor } d(C_i, B) \text{ sea mínimo para } i = 1, 2, \dots, K \}$$

Este valor indica que la imagen con matriz de características B pertenece al grupo C_{i^*} .

Una vez asignados todos los elementos de la base de datos de entrenamiento a sus respectivos grupos, se verifica que se cumpla la condición

$$\#(C_i) \in [m - d, m + d]$$

donde $m = \frac{N}{K}$ y $d = \frac{3N}{K^2}$ para $i = 1, 2, \dots, K$. En caso de no cumplir con esta condición, se procede a modificar el conjunto de centroides para la siguiente iteración del AG.

En el momento en el que el sistema deja de iterar se toman el vector de centroides para ser usados en la fase de pruebas.

Para el trabajo actual $N = \text{No. de elementos en la base de datos}$ y $K = 10$. Cada elemento del vector de centroides es una matriz de características de 100×9 .

Algoritmo Genético

Begin AG

Obtener población inicial $\mathcal{P}_0 = \{C_1, C_2, \dots, C_K\}$ aleatoriamente.

Mientras (PARO) no se cumpla:

Para $i = 1, 2, \dots, K$

Calcular $d(C_i, B)$ para cada B en la base de datos.

Fin

Calcular i^* .

Asignar B al grupo C_{i^*} .

Para $j = 1, 2, \dots, K$.

Si $\#(C_j) \in [m - d, m + d]$ entonces se conserva el elemento C_j .
caso contrario mutar el elemento C_j .

Fin.

Si no hubo mutación entonces PARO.

Fin.

4. Resultados experimentales

El sistema fue implementado usando el lenguaje de programación LabVIEW. Se realizaron diversas subrutinas para integrar dos módulos principales: Entrenamiento y Pruebas. Sólo en este último se tiene la visualización de imágenes por cuestiones de rendimiento y velocidad de procesamiento.

Prueba 1 Se realizaron diversas corridas de pruebas, iniciando con una pequeña base de 36 imágenes de bosques de la Base de Torralba, aunque las imágenes correspondían a escenarios de bosques específicamente, se seleccionaron aquellas que presentaran mayores diferencias entre ellas. Los resultados obtenidos fueron bastante alentadores ya que el sistema devuelve imágenes bastante similares a la solicitada, como se aprecia en la Fig. 3. En caso de que la imagen que se presenta al sistema corresponde a un centroide, el sistema responde con un 100 % de exactitud, es decir, devuelve la misma imagen.

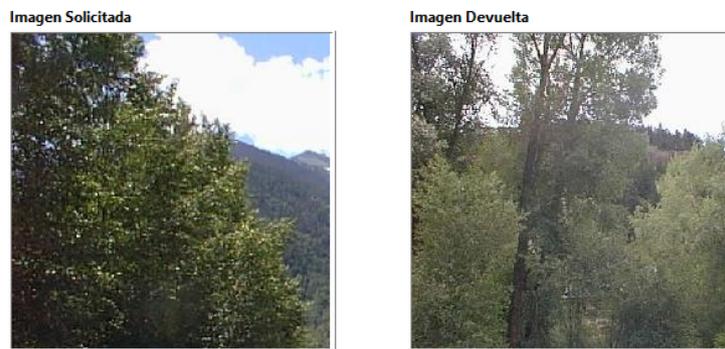


Fig. 3. Resultados con base de 36 imágenes.

Prueba 2 Para la segunda prueba, se utilizaron 328 imágenes de bosques de la misma base de imágenes de la Base Torralba para lograr un mejor entrenamiento que se reflejara en una mayor similitud en el momento de devolver la imagen, lo que efectivamente se consiguió.

Una de las principales interrogantes fue: ¿Qué imagen devolvería el sistema si se le presenta una imagen que no corresponda a un bosque? Así que entre las imágenes de entrada, se le presentaron imágenes pertenecientes a otras categorías como montañas, costas y praderas, en los que se observó un par de cosas interesantes,

1. En primer lugar, el tiempo que el sistema se tomaba para devolver una imagen fue hasta diez veces más que cuando se le presentaron imágenes de bosques solamente.
2. En segundo lugar, tal como se esperaba, la imagen devuelta tiene un alto grado de similitud en textura con la imagen solicitada como lo muestra la Fig. 4, aunque corresponde un bosque, lo anterior demuestra que los clusters del sistema han sido formados de manera correcta.

Prueba 3 Partiendo de los resultados de la prueba anterior, se integró una base de datos mixtas con imágenes extraídas de la base de datos de Torralba, pero en ésta ocasión mixta, usando 25 imágenes de costas, 25 de bosques, 25 de praderas y 25 de montañas.

La Fig. 5, muestra una de las imágenes con la que se probó el sistema con la base de datos mixta.

Prueba 4 Teniendo los resultados preliminares, se realizaron pruebas utilizando imágenes que no se encontraban en las bases de datos de entrenamiento. En ésta cuarta prueba, se tomaron imágenes escenarios naturales del Ejido Emiliano Zapata, municipio de Jiquipilas, Chiapas.

Una de las principales diferencias que se observó fue el tamaño, se hizo necesario escalarlas para que la información obtenida al procesarlas fuera representativa. El tamaño de éstas imágenes quedó en (256×192) , y al presentarlas al sistema los resultados fueron similares a los obtenidos usando las imágenes de la misma base de datos usada en el entrenamiento, como se ve en la Fig. 6.

Finalmente, se realizaron pruebas con las diferentes clases de imágenes por el contenido asociado–detectado, y estas se muestran en la Tabla 1.

Tabla 1. Resultados obtenidos con matriz mixta

No	<i>Vector</i> ₁		<i>Vector</i> ₂		<i>Vector</i> ₃	
	Presentada	Devuelta	Presentada	Devuelta	Presentada	Devuelta
1	costa	costa	costa	costa	costa	montaña
2	bosque	bosque	costa	costa	costa	montaña
3	bosque	bosque	costa	costa	costa	costa
4	montaña	bosque	bosque	bosque	bosque	bosque
5	pradera	montaña	bosque	bosque	bosque	bosque
6	pradera	pradera	bosque	pradera	montaña	montaña
7	costa	costa	montaña	montaña	montaña	montaña
8	costa	costa	pradera	bosque	montaña	montaña
9	montaña	montaña	pradera	pradera	montaña	pradera
10	montaña	montaña	pradera	pradera	pradera	pradera
Exactitud	80 %		90 %		70 %	

5. Conclusiones

Con respecto a las diferentes técnicas de CBIR, se eligió estudiar la implementación de los métodos de K-means y AG's, partiendo del hecho que se desea un equilibrio, entre el tiempo de procesamiento y eficiencia en la tarea de caracterización de imágenes en base de imágenes de gran volumen. El software creado fue comparado para medir la calidad de recuperación contra los módulos de MATLAB. Con el fin de mostrar avances en cuanto al software, se realizó la

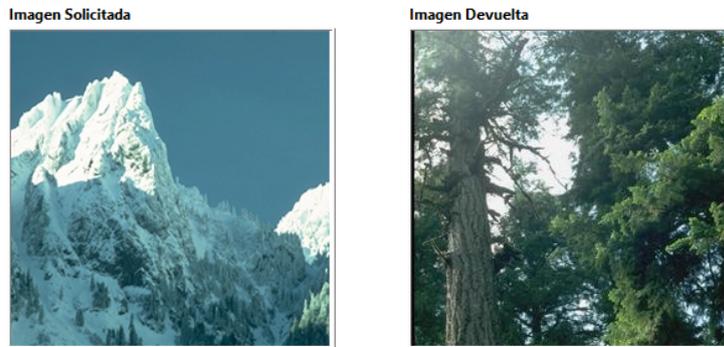


Fig. 4. Resultado solicitando imagen de montaña con base de datos de bosques.

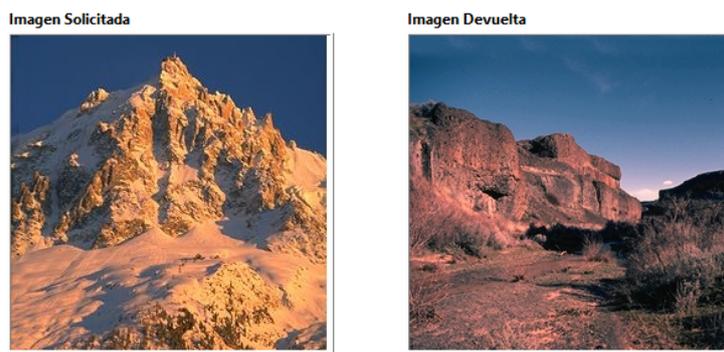


Fig. 5. Resultado solicitando imagen en Base de Imágenes Mixta



Fig. 6. Resultado solicitando imagen de paisaje de Chiapas.

programación completa de rutinas de K-means y AG. Es de destacar dichos módulos no se encuentran disponibles en LabVIEW a diferencia de MATLAB. Se usó el software LabVIEW porque en etapas futuras se desea la implementación del software en hardware especializado para tener un procesamiento en tiempo real. Este aspecto se considera importante, si se toma en cuenta que fueron usadas fotos personales ajenas a cualquier base de datos y se usan computadoras comerciales. Se observaron tiempos de entrenamiento lineales con respecto a la cantidad de elementos en la base de datos de entrenamiento, de igual forma, se obtuvieron diferencias en tiempos de procesamiento, mostrando una reducción considerable en los tiempos de entrenamiento y prueba.

Los resultados fueron satisfactorios como se observa en la Tabla 1, logrando niveles de exactitud en recuperación hasta de un 90 %

Como trabajo futuro planteamos dos puntos inmediatos: *(i)* ampliar el número de imágenes de prueba, y *(ii)* realizar la implementación en hardware FPGA o dispositivo móvil del sistema CBIR optimizado, de tal manera que mediante una cámara del propio dispositivo se adquieran y procesen las imágenes, y sea capaz de clasificar el tipo de imagen por medio de sus características usando esta metodología aquí dispuesta.

Referencias

1. Bosch, A., Zisserman, A., Muñoz, X. Scene classification using a hybrid generative/discriminative approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30, 712–727 (2008)
2. T. Dharani and I. Aroquiaraj. A survey on content based image retrieval. In: *International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME)*, pp. 485–490 (Feb 2013)
3. Fukunaga, K. *Introduction to statistical pattern recognition*. (2nd ed.) San Diego, CA, USA: Academic Press Professional, Inc. (1990)
4. Li, J., Wang, J. Z. Real-time computerized annotation of pictures. *Proceedings of the 14th annual ACM international conference on multimedia*. New York, NY, USA: ACM. (pp. 911–920) (2006)
5. Liu, Y., Zhang, D., et al. A survey of content-based image retrieval with high level semantics. *Pattern Recognition*, 40, 262–282 (2007)
6. Serrano-Talamantes J.F. , Avilés-Cruz C., Villegas-Cortez J., and Sossa-Azuela J. H., Self organizing natural scene image retrieval, *Expert Systems with Applications*, vol. 40, no. 7, pp. 2398–2409 (2013)
7. Vogel, J., Schiele, B. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2), 133–157 (2007)

Reviewing Committee

Noé Alejandro Castro-Sánchez
Jair Cervantes
William De La Cruz De Los Santos
Anilu Franco-Arcega
Sofía N. Galicia-Haro
Alexander Gelbukh
David Gonzalez
Miguel Gonzalez-Mendoza
Oscar Herrera
Hector Jimenez Salazar
Asdrúbal López

Lourdes Martínez
Sabino Miranda
Raul Monroy
Antonio Neme
Obdulia Pichardo-Lagunas
Rafael Rojas-Hernández
Grigori Sidorov
Israel Tabarez
Valentín Trujillo Mora
Edgar Vallejo
Nestor Velasco Bermeo

Additional reviewers

Jorge Rodriguez Ruiz
Roberto Alonso
Soujanya Poria

José Camiña
Victor Ferman
Gabriela Ramírez-De-La-Rosa

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
mayo de 2014
Printing 500 / Edición 500 ejemplares

