

PISTAS

Educativas

NUEVA ÉPOCA • No. 112 • NOVIEMBRE 2015 • ISSN 1405-1249

SENIE 2015

XI Semana de Ingeniería Electrónica



SEP
SECRETARÍA DE
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CELAYA



40
Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA



TECNOLÓGICO NACIONAL DE MÉXICO

Manuel Quintero Quintero
Director

Ignacio López Valdovinos
Encargado Secretaría Académica

DIRECTORIO ITC

M.E.D. Ignacio López Valdovinos
Director

M.C. Martha Carreño Juárez
Subdirectora Académica

M.F. Martín Campos Moreno
Subdirector de Servicios Administrativos

M.C. Julián Ferrer Guerra
Subdirector de Planeación y Vinculación

MDPH Teresita de las Nieves Armengol Rico
Jefa del Departamento de Desarrollo Académico



PISTAS EDUCATIVAS

pistaseducativas@itcelaya.edu.mx

COMITÉ EDITORIAL

Juan José Martínez Nolasco

Luis Alberto López González

COMITÉ TÉCNICO

Dra. Abascal Mena Ma. del Rocío - UAM-Cuajimalpa

Dr. Alfaro Ruiz Víctor M. - Universidad de Costa Rica

Dr. Anzures Marín Juan - Universidad Michoacana de San Nicolás de Hidalgo

Dr. Arechiga Martínez René - New Mexico Tech Electrical Engineering Department, U.S.A.

Dr. Astorga Zaragoza Carlos - CENIDET-Cuernavaca

Dr. Aviles Cruz Carlos - UAM-Azcapotzalco

M. C. Borrero Guerrero Henry - Universidad de Sao Paulo

Dr. Bonilla Gutiérrez Isela - Universidad Autónoma de San Luis Potosí

Dra. Bravo Contreras Maricela - UAM-Azcapotzalco

Dr. Campos Delgado Daniel Ulises - Universidad Autónoma de San Luis Potosí

PISTAS EDUCATIVAS

• No. 112 • Noviembre 2015

Contenido

| | |
|---|---------|
| Editorial..... | 3 |
| Presentación..... | 6 |
| Simulación de la sístole y diástole utilizando el modelo del autómatas híbrido Braulio Cruz Jiménez, Jannette Contreras Rivero, Ricardo Peón Escalante..... | 8-21 |
| Diseño, construcción y control de un sistema de levitación neumática Irving Ulises Hernández Miguel, Eufracia Mendoza Monjaraz, Jesús Pacheco Mendoza, José Alberto Cruz Tolentino, Alejandro Jarillo Silva..... | 22-42 |
| On model parameter estimation methods of DC electric motors José Jiménez González, Francisco Beltrán Carbajal, Felipe González Montañez, Irvin López García..... | 43-60 |
| Modelado y control de un generador de SAGS basado en un convertidor back to back Gabriel Calderón Zavala, Jesús Darío Mina Antonio, Adolfo Rafael López Núñez, Leodan Robisel Martínez Flores..... | 61-80 |
| Close loop step test used for tuning PID controller by genetic algorithms Rubén Lagunas Jiménez, Alonzo González Aguilar, Víctor Lanz Gutiérrez De Velasco..... | 81-98 |
| Control de ganancias variables para robots de n-grados de libertad Oscar Ramos Arroyo, Fernando Reyes Cortés, María Aurora Diczorca Vargas Treviño, Sergio Vergara Limón..... | 99-114 |
| Modelado de un sistema máquina bus infinito usando transformada de Laplace Irma Martínez Carrillo, Carlos Juárez Toledo..... | 115-117 |
| Diseño y simulación de un controlador analógico para el convertidor cd-cd reductor Alberto Martínez Barbosa, Esteban Osvaldo Guerrero Ramírez..... | 118-143 |
| Control del módulo de calentamiento e instrumentación de una máquina para ensayos de torsión en caliente Gustavo Aquino Mendieta, Miguel Magos Rivera, Juana Eloina Mancilla Tolama..... | 144-163 |
| Control de la velocidad de un motor de cd a través de un convertidor cd-cd reductor alimentado mediante paneles solares Alberto Martínez Barbosa, Esteban Osvaldo Guerrero Ramírez..... | 164-178 |
| Ocupación espectral de la banda ISM de 2.4 GHz para aplicaciones de radio cognoscitivo Miguel Ángel Díaz Ibarra, Marco Aurelio Cárdenas Juárez, Ulises Pineda Rico, Enrique Stevens Navarro..... | 179-194 |
| Sistema de comunicaciones para mini UGVs/UAVs en tareas de protección civil mediante COTS Dulce Mayra Janet Martínez García, Emmanuel Ruiz López, Marco Aurelio Cárdenas Juárez, Enrique Stevens Navarro, Ulises Pineda Rico..... | 198-234 |
| Red avanzada de comunicaciones mediante Raspberry Pi para aplicaciones en vehículos aéreos no tripulados Luis Fernando Hernández Medina, Marco Aurelio Cárdenas Juárez, Enrique Stevens Navarro, Armando Arce Casas, Ulises Pineda Rico..... | 215-234 |
| Caracterización de canal para sistemas de comunicación por luz visible en interiores A. M. Ramírez Aguilera, J. M. Luna Rivera, R. Pérez Jiménez..... | 235-251 |
| Diseño de radiocomunicación de CanSat didáctico Andrés Calvillo Téllez, José Cruz Núñez Pérez, Teodoro Álvarez Sánchez, Ricardo Cárdenas Valdez, Antonio Gómez Roa, Raúl Rascón Carmona..... | 252-262 |
| Design, characterization and modeling of a tunneling break-down photodiode integrated in a standard 0.5µm CMOS technology Juan J. Ocampo Hidalgo, J. Alex Zenteno Hernández, Rosnelly Cruz Cecilio, David Flores-Montoya, Olaf E. Duran-Nava, Armando Gómez-Vieyra..... | 263-281 |
| Implementación de un circuito caótico memresistivo Moisés Torres Rivera..... | 282-301 |
| Delta-Sigma Converter Processing: aplicación de herramienta de software libre para el análisis y caracterización de convertidores $\Sigma\Delta$. Javier Alducin Castillo, Juan Jesús Ocampo Hidalgo, Iván Vázquez Álvarez..... | 302-319 |

PISTAS EDUCATIVAS

• No. 112 • Noviembre 2015

Contenido

| | |
|---|---------|
| Filtro CMOS pasa bajas con parámetros variantes en el tiempo Edgar López Delgadillo, Luis Alejandro Flores Oropeza, Alfonso Vela Rivera..... | 320-333 |
| Configuración óptima para un enfriador termoelectrico de dos etapas usando BiTe y PbTe en cada etapa Karen Anaíd Rendón Cortés, Miguel Angel Olivares Robles..... | 334-350 |
| Análisis de la conductividad térmica efectiva en un generador termoelectrico segmentado usando Bi2Te3 y PbTe Ana Lilia Nieto Magdaleno, Miguel Ángel Olivares Robles, José Jorge Chanona Pérez..... | 351-367 |
| Análisis térmico de un enfriador termoelectrico para bebidas impulsado por rueda mecánica Gerardo Plata Medina, Miguel Ángel Olivares Robles..... | 368-387 |
| Influencia del dopante de In_2O_3 en las propiedades eléctricas de varistores a base de $SnO_2-Co_3O_4-Nb_2O_5-Cr_2O_3$ Julio Cesar Martínez Ramírez, A.N. Bondarchuk, A.B. Glot..... | 388-398 |
| Optimización geométrica de un MEMS tipo Hot-Arm María Teresa Orvañanos Guerrero, Jorge Varona Salazar, Ramiro Velázquez Guerrero..... | 399-421 |
| Virtualización tridimensional interactiva de un parque eólico con fines didácticos Karla del Rosario Trujillo González, A. Guadalupe Toledo Toledo, J. Jesús Arellano Pimentel, Miguel Ángel Hernández López, María de los Angeles Alonso Lavernia, Verónica Martínez Lazcano, Iliana Castillo Pérez, Yira Muñoz Sánchez..... | 422-441 |
| Desarrollando competencias digitales en los docentes María de los Angeles Alonso Lavernia, Verónica Martínez Lazcano, Iliana Castillo Pérez, Yira Muñoz Sánchez..... | 442-463 |
| Diseño y desarrollo de una red social de aprendizaje colaborativo para dispositivos móviles María Lucía Barrón Estrada, Ramón Zatarain Cabada, Rodrigo Beltrán Lugo..... | 464-485 |
| Videojuego didáctico empleando el kinect para la manipulación virtual de un aerogenerador Martín Hernández Reyna, Miguel Ángel Hernández López, J. Jesús Arellano Pimentel, A. Guadalupe Toledo Toledo..... | 486-507 |
| Metodología de diseño de osciladores LC con componentes discretos: una propuesta Edgar Omar Lara Alemán, F. Sandoval Ibarra, Esaú Villatoro-Tello, Carlos R. Jaimez-González..... | 508-532 |
| Diseño e implementación de Applets como material didáctico de apoyo para cursos de estructuras de datos Betzalet García-Mendoza, Pablo A. Ruiz-Mendoza, Gerardo Real-Flores, Carlos R. Jaimez-González, Esaú Villatoro-Tello..... | 533-553 |
| Experiencia de Didáctica Lúdica para Incentivar el Aprendizaje Sergio Zepeda-Hernández, Rocío Abascal-Mena, Erick López-Ornelas..... | 554-572 |
| Programación de mini robots para el desarrollo de aprendizaje significativo Margarita Elizabeth Gómez Mayorga, Luz Noé Oliva Moreno..... | 573-584 |
| Optimización de un prototipo de sistema fotovoltaico autónomo para iluminación de anuncios espectaculares Pedro Manuel Rodrigo Cruz, José Sebastián Gutiérrez Calderón, Arturo de la Torre Infante..... | 585-606 |
| Nuevo inversor fotovoltaico integrado sin transformador Jeziel Vázquez, Nimrod Vázquez, Claudia Hernández, Héctor López, Joaquín Vaquero..... | 607-627 |
| Evaluación del desempeño de un transformador monofásico mediante software Hiram Alberto Canseco García, José Jiménez González, Eduardo Campero Littlewood, Irvin López García, Juan Carlos Olivares Galván..... | 628-642 |
| Sistema de calentamiento por inducción electromagnética para pruebas experimentales en laboratorios de Físico-Química Julio Ortega Alejos, Leonel Estrada Rojo, Carlos A. Fuentes Hernández, Elías J. J. Rodríguez Segura, Israel Álvarez Ariza, José Jesús Arizaga Ibarra, Diego Díaz López..... | 643-665 |
| Sistemas de detección de modo isla en inversores conectados a la red eléctrica: análisis del estado del arte Carlos Javier Camacho Márquez, Nimrod Vázquez Nava, Claudia Verónica Hernández Gutiérrez, Héctor Juan Carlos López Tapia..... | 666-683 |
| Análisis del rendimiento de TEG's utilizando una teoría de circuitos térmicos y eléctricos Vargas Almeida Alexander, Olivares Robles Miguel Ángel..... | 684-696 |

COMITÉ TÉCNICO

Dr. Correa Gómez Javier - Instituto Tecnológico de Morelia
 M. C. Cruz Roa Ángel Alfonso - Universidad Nacional de Colombia
 Dra. Cuevas Rasgado Alma Delia - Universidad Autónoma del Estado de México-Texcoco
 Dr. Díaz Carmona José Javier - Instituto Tecnológico de Celaya
 Dr. Espinosa Calderón Alejandro - Instituto Tecnológico de Celaya
 Dr. Esquit Hernández Carlos A. - Universidad del Valle de Guatemala
 Dr. Fallad Chávez Jalil - Centro Universitario de la Costa Sur
 Dra.Félix Aviña Vanessa Guadalupe - Universidad Politécnica de Sinaloa
 M. C. Fuentes Covarrubias Ricardo - Universidad de Colima
 Dr. Gutiérrez Calderón Sebastián - Universidad Panamericana-Aguascalientes
 Dr. Gutiérrez Díaz de León Carlos Adrian - Universidad Autónoma de San Luis Potosí
 Dr. Herrera Alcántara Oscar - UAM-Azcapotzalco
 Dr. Incera Diéguez José Alberto D. - Instituto Tecnológico Autónomo de México
 Dr. Jiménez Alaniz Juan Ramón - UAM-Iztapalapa
 M. T. Jiménez Orozco Gabriel - ITESM Cd de México
 Dr. López Ornelas Erick - UAM-Cuajimalpa
 Dr. Magos Rivera Miguel - UAM-Azcapotzalco
 Dr. Martínez Pelaez Rafael - Universidad de la Sierra Sur
 Dr. Mena Camare Luis Javier - Universidad Politécnica de Sinaloa
 Dr. Mendoza Gutiérrez Marco Octavio - Universidad Autónoma de San Luis Potosí
 Dr. Monroy Borja Raúl - ITESM Edo de México
 Dr. Morales Alanis Eduardo - New England Baptist Hospital
 Dr. Parra Michel Ramón - CINVESTAV-GDL
 Dr. Peña Campos Fernando - CINVESTAV-GDL
 Dr. Potes Cristhian M. - Philips Research North America
 Dr. Ramírez Treviño Alberto - Instituto Tecnológico Superior de Cajeme
 M. C. Reider Burstin Jerry - Universidad Anáhuac – México Norte
 Dr. Reyes Ortiz José Alejandro - UAM-Azcapotzalco
 Dr. Rivas Araiza Edgar - Universidad Autónoma de Querétaro
 Dr. Rizo Domínguez Luis - Universidad del Caribe
 Dr. Rodrigo Cruz Pedro - Universidad Panamericana Aguascalientes
 Dr. Rodríguez García José G. - CINVESTAV-Zacatenco
 Dr. Rodríguez Segura Elías José J. - Instituto Tecnológico de Celaya
 M. C. Sanchez Gomez Claudia Nallely - Universidad Panamericana-Aguascalientes
 M. C. Santiago Espinosa Felipe - Sanviago Tecnológica de la Mixteca
 Dr. Tecpanecatí Xihuitl Jose Luis - Universidad Autónoma de San Luis Potosí
 Dr. Toral Cruz Homero - Universidad de Quintana Roo
 Dr. Torres Román Deni Librado - CINVESTAV-GDL
 Dr. Vargas Rubio Juan Gaspar - UAM-Azcapotzalco
 Dr. Vázquez Álvarez Iván - UAM-Azcapotzalco
 Dr. Vázquez Castillo Javier - Universidad de Quintana Roo
 Dr. Vázquez Cerón Ernesto R. - UAM-Azcapotzalco
 M.C. Velázquez Buendía Francisco - Universidad Panamericana Aguascalientes
 Dr. Velázquez Guerrero Ramiro - Universidad Panamericana Aguascalientes
 Dr. Villegas Cortez Juan - UAM-Azcapotzalco

Pistas Educativas es una publicación impresa del Departamento de Desarrollo Académico del Instituto Tecnológico de Celaya. Los artículos y ensayos son responsabilidad de sus autores y no expresan necesariamente la ideología de la institución ni de la revista. En caso de citar la información de los contenidos se agradecerá dar los créditos a los autores, sus obras, así como a esta fuente.

Dirección: Antonio García Cubas Pte. No. 600 Esq. Av. Tecnológico, C.P. 38010, Celaya, Gto. Tel. 01-(461)-61- 175-75. Fax. 01-(461)-61-179-79. Certificado de Licitud y Contenido 6216 y 4777, respectivamente. Registro de Derecho de Autor, Expediente de Reserva: 6 98 92. ISSN 1405-1249.

Editorial

En su edición 112 la revista Pistas Educativas publica con agrado la Memoria de la Décimo Primera Semana Nacional de Ingeniería Electrónica –SENIE 2015-, que se llevó a efecto entre el 7 y el 9 de octubre bajo la organización conjunta de la División de Ciencias Básicas e Ingeniería de la Universidad Autónoma Metropolitana Azcapotzalco así como la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí, fungiendo esta última institución como anfitriona del evento.

Cada uno de los artículos que aquí se publican fueron sometidos a la consideración de un grupo evaluador constituido por dos investigadores expertos en el tema y es del Comité Técnico de SENIE 2015 la responsabilidad en lo que se refiere a su calidad técnica. Por su parte, Pistas Educativas tuvo bajo su cuidado la edición de los textos de acuerdo, como siempre, con sus normas de publicación. A través de esta edición, el Instituto Tecnológico de Celaya, por medio de Pistas Educativas, avanza en su cometido de ser un vehículo para la difusión del conocimiento, albergando en esta ocasión las experiencias y logros de los que dan cuenta en sus artículos los estudiantes, profesores e investigadores de buena parte del sistema educativo nacional que se congregaron en SENIE 2015.

PISTAS EDUCATIVAS

• No. 112 • Noviembre 2015

Contenido

| | | |
|--|--|-----------|
| Sistema fotovoltaico multifuncional conectado a la red eléctrica controlado por medio de modos deslizantes | Iván Méndez, Nimrod Vázquez, Claudia Hernández, Héctor López, Joaquín Vaquero..... | 697-715 |
| Análisis del proceso de balanceo de voltaje en un arreglo de supercapacitores | Edgar Peralta Sánchez, Pedro Celestino Castellanos Morales, Nayeli Itzel Hernández Zárate, Gerardo Mino Aguilar, Beatriz E. Graniel García..... | 716-734 |
| Estudio comparativo del coeficiente de potencia y de torque presentados en diversas turbinas eólicas | Viviana Reyes Andrade, Oscar Carranza Castillo, Jaime José Rodríguez Rivas..... | 735-756 |
| Sistema para el control y prueba de motores síncronos con base en el procesador digital de señales TMS320 | Fermin Hugo Ramírez Leyva, Mónica Edith García García, Felipe Santiago Espinosa, Gerardo Cruz González..... | 757-774 |
| Medida experimental y modelado matemático de módulos fotovoltaicos | Raúl Castillo Meraz, Roberto Carlos Martínez Montejano, Isaac Campos Cantón, Misael Francisco Martínez Montejano..... | 775-796 |
| Modelado, control y simulación de elementos básicos que componen una micro red de CD | Arturo De los Santos Vázquez, Elías José Juan Rodríguez Segura, Nancy Visairo Cruz..... | 797-817 |
| Tarjeta didáctica de control para tiristores | Luis Alejandro Flores Oropeza, Eduardo Anguiano Mendieta, Edgar López Delgado..... | 818-838 |
| Medición de velocidad sin sensor en motores de inducción sumergibles conectados directamente a la línea | Antonio Zamarrón Ramírez, Salomón Rodríguez Carrera, Daniel Zamarrón Barrón, Francisco Valdivia..... | 839-852 |
| Convertidores balanceadores de voltaje: Estado del arte | Fidel Alejandro Castro Espinosa, Elías José Juan Rodríguez Segura, Ciro Alberto Núñez Gutiérrez..... | 853-872 |
| Desarrollo de un medidor portátil para la detección de metales pesados disueltos en medios acuosos utilizando principios de fluorescencia | Marco Antonio González Cantellano, Luis Manuel Montaña Zetina..... | 873-894 |
| Sistema integral de software y hardware para el aprendizaje del funcionamiento y manejo de los sensores | Javier Silvestre Zavala, José Nicolás Márquez Márquez..... | 895-917 |
| Desarrollo de un sistema de electro estimulación transcutánea de bajo costo | Miriam C. Reyes Fernández, Ruben Posada Gómez, Albino Martínez Sibaja, Oscar O. Sandoval González, Mario Alberto García Martínez, Luis H. Sánchez Medel..... | 918-934 |
| Método para la medición de la inductancia bajo condiciones reales de operación | Jerry Nathan Reider Burstin..... | 935-953 |
| Estimación de la humedad y resistencia a compresión final, de una mezcla de concreto fresco utilizando microondas. | Gerardo Calva Olmos, Rafael Prieto Meléndez, Alberto Herrera Becerra, Alejandro Padrón Godínez, Mario Pacchiano de la Garza..... | 954-976 |
| Plataforma de adquisición y monitoreo para el estudio de las condiciones ambientales externas y de confort en casas prototipo para la validación de un control térmico por medio de un sistema de muro verde y aislante térmico | Cesar Agustín Corona Patricio, Cesar Alberto Dueñas Sosa, Jesús Iván Castro Enríquez, José Luis Avendaño Juárez, Jorge Adán Sánchez Reséndiz, Eusebio JR. Ventura Ramos, Francesca Olivieri..... | 977-997 |
| Mejora de la resolución espectral de un monocromador usando un controlador de micropasos | Ricardo Castro García, Luis Felipe Lastras Martínez..... | 998-1019 |
| Efectos de holgura de una banda trapezoidal de acoplamiento de un sistema electromecánico sobre magnitudes mecánicas y eléctricas | Jorge Iván Ramírez López, Efraín Ramos Buenfil, Marco Antonio Rodríguez Blanco, Javier Torres Jurado..... | 1020-1035 |

PISTAS EDUCATIVAS

• No. 112 • Noviembre 2015

Contenido

| | |
|--|-----------|
| Registrador de energía fotovoltaica inyectada a la red eléctrica mediante LabView Luis Carlos Gallegos Hernández, Mariela Serna Serna, Eduardo Salinas Rodríguez, Antonio García Narvaez, Guillermo Reséndez Morales, Pedro Casillas Ríos..... | 1036-1049 |
| Monitoreo del consumo doméstico de agua potable utilizando tecnología ZigBee Pedro Alvarado Medellín, Job Daniel Rodríguez Hernández, Ruperto Ortiz Gómez, Blanca Esthela Solís Recendez..... | 1050-1064 |
| Sistema teleoperado para estimulación eléctrica transcorneal de señales múltiples Juan Jaime Salinas Lara, Eric Simancas Acevedo, Daniel Robles Camarillo, Luis Niño de Rivera y Oyarzabal..... | 1065-1086 |
| Interpolación polinomial para determinar el tiempo óptimo de tratamiento criogénico para acero AISI D2 en medio básico Irma Hernández Casco, Carlos Juárez Toledo, Irma Martínez Carrillo, Rubén Herrera Galicia, José David Arroyo Pérez, Manuel de Jesús Hernández Gutiérrez, Miriam Roxana Vázquez Flores, Tania Berenice Palacios Molina | 1087-1100 |
| Monitoreo y Control de Procesos Basado en Arduino y Raspberry Vía Internet Rubén Herrera Galicia, José David Arroyo Pérez, Manuel de Jesús Hernández Gutiérrez, Miriam Roxana Vázquez Flores, Tania Berenice Palacios Molina..... | 1101-1120 |
| Análisis para el diseño de circuitos osciladores de Colpitts con sensores de gas QCM Juan Jesús Jiménez Arellano, Severino Muñoz Aguirre, Georgina Beltrán Pérez, Juan Castillo Mixcoatl, José Lorenzo Muñoz Mata | 1121-1135 |
| Análisis cinemático y síntesis de un sistema de palancas para la sub-actuación de un dedo artificial con 3 articulaciones Jesús Alberto Ordaz Rivera, Emilio Miguel Soto García, José Eligio Moisés Gutiérrez Arias, José Eladio Flores Mena, María Monserrat Morín Castillo.... | 1136-1156 |
| Una propuesta de arquitectura para el control de un robot guía Karla Lourdes Luna Gallegos, Elvia Ruth Palacios Hernández, Antonio Marín Hernández..... | 1157-1173 |
| Fault detection in VSD-fed induction motors through Park's impedance and fuzzy systems Arturo Mejía Barron, Armando Guadalupe García Ramírez, Roque Alfredo Osornio Ríos, Rene de Jesus Romero Troncoso..... | 1174-1192 |
| Diseño e implementación de tele operación de un robot móvil diferencial programado con Python desde un dispositivo Android Josué Cirilo Cruz, Arturo Zúñiga López, Juan Villegas Cortez, Carlos Avilés Cruz..... | 1193-1209 |
| Control en tiempo real de un mecanismo XY virtual por medio de un joystick Gustavo Aquino Mendieta, Miguel Magos Rivera, Ricardo Godínez Bravo..... | 1210-1231 |
| Inicialización de un robot humanoide tipo BIOLOID para tareas de caminado Antonio Benitez Ruiz, María Auxilio Medina Nieto, Jorge de la Calleja Mora, Emmanuel Baleón Flores, Carmina Cruz Ochoa, Ivonne López Cuacuas..... | 1232-1253 |
| Diseño de un sistema de desarrollo para la emulación de tráfico en una intersección vial con vehículos autónomos Andrés Ferreyra Ramírez, Arturo Zúñiga López, Roberto A. Alcántara Ramírez..... | 1254-1274 |
| Renderizado háptico para el Novint Falcon Enrique Linares Ramírez, Jorge Gudiño Lau, Saida Miriam Charre Ibarra, Miguel Duran Fonseca..... | 1275-1287 |
| Caracterización de los sonidos implicados en el ciclo cardíaco mediante un fonocardiograma Braulio Cruz Jiménez, Jannette Contreras Rivero, Luis Ricalde Castellanos..... | 1288-1308 |
| El invisible y asombroso proceso de la comunicación oral: bases sobre reconocimiento de voz Ángel David Pedroza Ramírez, José Ismael de la Rosa Vargas..... | 1309-1329 |
| Deconvolución óptima de mediciones de fluorescencia Gerardo Ibarra Vázquez, Daniel U. campos Delgado, Edgar R. Arce Santana, Javier A. Jo..... | 1330-1349 |
| Descomposición de datos multi-espectrales: interfaz gráfica para Matlab Omar Gutiérrez Navarro, Daniel U. Campos Delgado, Edgar R. Arce Santana, Javier A. Jo..... | 1350-1368 |
| Estimación de erosión de suelos utilizando sensores remotos y programación genética Cesar Augusto Puente Montejano, Gustavo Olague Caballero..... | 1369-1388 |
| Hacia la clasificación de fonocardiogramas utilizando descriptores caóticos y estadísticos Carlos Antonio Osorio Maceda, Hugo G. González-Hernández..... | 1389-1408 |
| Activación sistema de frenos ABS usando visión por computadora para la detección de la superficie de rodamiento Gabriel García Ponds, Juan Villegas Cortez, Carlos Avilés Cruz, Iván Vázquez Álvarez, Ismael Osuna Galán, Yolanda Pérez Pimentel..... | 1409-1423 |
| Implementación de una estructura neuronal celular en hardware reconfigurable Luis F. Muñoz M., Juan José Raygoza P., J. Roberto R. Barón, Susana Ortega Cisneros..... | 1424-1444 |
| Identificación de rostros por técnica de puntos de interés SURF Cesar Benavides Alvarez, Graciela Román Alonso, Juan Villegas Cortez, Carlos Avilés Cruz..... | 1445-1465 |
| Segmentación de Iris con OpenCV en Android Oscar Fuentes Salome, Aldrin Barreto Flores, Verónica Edith Bautista López..... | 1466-1486 |
| Detección de Puntos Salientes para Medición de la Simetría Facial Ricardo Irving Barrón Martínez, Ruth Mariela Aguilar Ponce, José Luis Tecpanecatl Xihuitl..... | 1487-1507 |
| Virtualización mediante MetaRouter para la implementación de una red wireless de navegación anónima tipo TOR en equipos Mikrotik Roberto David Meneses Basantes, Darwín Leonidas Aguilar Salazar, Christian Nestor Vega Muñoz, Rita Paola León Pérez..... | 1508-1527 |

| | |
|---|-----------|
| La aplicabilidad del cómputo ubicuo para el monitoreo de animales de caza Adrian Castañeda Morfin, María Andrade-Aréchiga, Pedro Damián-Reyes..... | 1528-1545 |
| Tecnología móvil para evaluar la calidad en uso de Proyectos-Productos-Servicios en eventos de innovación e invención Laura Silvia Vargas Pérez, Agustín Francisco Gutiérrez Tornés, Edgardo Manuel Felipe Riverón, Vanessa Atenea Vargas Pérez, Jorge Peralta Escobar..... | 1546-1566 |
| Diseño y construcción de un sistema de supervisión para la evaluación de la calidad del agua en sistemas de cultivo de camarón José Juan Carbajal Hernández, Luis Pastor Sánchez Fernández, Luis Alfonso Aguilera Larrañaga..... | 1567-1587 |
| Diseño de un middleware tolerante a fallas basado en el protocolo Paxos Ricardo Adán Madrid Trejo, Ricardo Marcelín Jiménez, Orlando Muñoz Texzocotetla..... | 1588-1609 |
| Arquitectura de descubrimiento de servicios para entornos hospitalarios (módulo tiny application) José Luis Santiago López, Víctor Alberto Gómez Pérez, Adrián Josué Ramírez Díaz, Alejandro Jarillo Silva, Juan Carlos Santiago López..... | 1610-1632 |
| Detección y análisis de comunidades en redes sociales (#TodosSomosPolitécnico) Erick López Ornelas, Rocío Abascal Mena, Sergio Zepeda Hernández..... | 1633-1651 |
| Sistema de multas automático mediante la tecnología NFC para una biblioteca María Elena Acevedo Mosqueda, Marco Antonio Acevedo Mosqueda, Karen Ailed Neri Espinoza, Alexa Chávez Álvarez..... | 1652-1664 |
| Propuesta de tutorial interactivo en línea para apoyar la enseñanza del lenguaje de consulta estructurado Arturo Palma Hernández, Carlos R. Jaimez González..... | 1665-1684 |
| Identificación de influyentes en Twitter a través del análisis textual y la elaboración de grafos Dra. Rocío Abascal Mena, Dr. Erick López Ornelas, Dr. Sergio Zepeda Hernández..... | 1685-1702 |
| Proyección de los índices de criminalidad de los delitos del fuero común: una representación sobre un sistema de información geográfica para la Ciudad de México César Hernández Anaya, Héctor Rafael Orozco Aguirre..... | 1703-1720 |
| Aplicación web para la generación de material didáctico incorporando recursos educativos Reyna Carolina Medina Ramírez, Carlos Alberto López Pérez, Fausto Casco Sánchez..... | 1721-1739 |
| Repositorio de grafos para el conteo de conjuntos independientes Juan Antares Perdomo Flandez, Pedro Bello López, Meliza Contreras González, Brayan Chavez Benavides..... | 1740-1759 |
| Agrupamiento de servicios Web usando el algoritmo de colonia de hormigas Maricela Claudia Bravo Contreras, Román Anselmo Mora Gutiérrez, Roberto Alfonso Alcántara Ramírez..... | 1760-1778 |
| RespirAtorio, una aplicación de realidad aumentada para niños María Auxilio Medina Nieto, César Rafael Chío Plata, Brenda Susana Ponce Valencia, Antonio Benitez Ruiz, Jorge de la Calleja, Eduardo López Domínguez..... | 1779-1799 |
| La lógica difusa como un medio para identificar y valorar las alteraciones de conducta en niños a nivel de educación básica: primaria y secundaria Andrés Ferreyra Ramírez, David G. Maxinez, Arturo Zúñiga López, Roberto A. Alcántara Ramírez..... | 1800-1819 |
| Plataforma de entrenamiento para PIC Mauricio López Villaseñor, Miguel Ángel Ruiz Sánchez, Miguel López Guerrero..... | 1820-1836 |
| Sistema de acceso a áreas restringidas con detección de intrusos José Ignacio Vega Luna, Mario Alberto Lagos Acosta, Gerardo Salgado Guzmán, Víctor Noé Tapia Vargas..... | 1837-1852 |
| Uso del lenguaje VHDL como apoyo a la enseñanza de arquitectura de computadoras Carlos Federico Hernández Farfán, Diego Andrés Navarro Jiménez, Emmanuel Eduardo Hernández Rico, Jonathan Paris Vargas Mosqueda..... | 1873-1892 |
| Desarrollo e implementación de un algoritmo tolerante a fallas para una red ZigBee Stephany Bucio Herrera, Alberto Carlos Cañada Gómez, David Rodríguez Uribe, Ricardo Marcelín Jiménez, Mauricio López Villaseñor..... | 1853-1872 |
| Estructuras diseñadas a medida para adquisición de datos de un sensor de visión dinámica usando el protocolo AER en un FPGA Nicolás Maldonado Orozco, J. Roberto Reyes Barón, Juan José Raygoza P., Susana Ortega Cisneros, Juan Luis del Valle..... | 1893-1911 |
| Implementación y optimización del uso de DPS en FPGA en diseño de circuitos a medida para calcular determinantes de orden 4 Francisco J. Plascencia Jauregui, Juan José Raygoza P., Edwin C. Becerra A., Susana Ortega Cisneros..... | 1912-1933 |
| Sistema de alertas y control de luces de una casa, utilizando plataformas de internet y comunicación inalámbrica Roberto de Lima Hernández, Aldrin Barreto Flores, Verónica Edith Bautista López..... | 1934-1946 |
| Arquitectura genérica de una red en chip de enrutamiento unidireccional en FPGA Jorge Ernesto López Arce Delgado, Juan José Raygoza P., Susana Ortega Cisneros, Jorge Rivera D., Pablo Moreno Villalobos..... | 1947-1959 |
| Monitor de alerta de fugas de gas José Ignacio Vega Luna, Mario Alberto Lagos Acosta, Gerardo Salgado Guzmán, Víctor Noé Tapia Vargas..... | 1960-1973 |
| Desarrollo de un sistema embebido móvil de bajo costo utilizando la tarjeta beaglebone Edgar Alejandro Rivas Araiza, Alexander Rodríguez Rosales, Estefanía Desiree Avalos Rivera..... | 1974-1994 |
| Implementación de un multiplicador de punto flotante de doble precisión basado en el estándar IEEE 754-2008 José Itzcóatl Sandoval López, Juan José Raygoza Panduro, Susana Ortega Cisneros, Jorge Rivera Domínguez..... | 1995-2016 |
| Diseño y construcción de un limnógrafo electrónico Felipe Santiago Espinosa, Gabriela Álvarez Olguín, Fermín Hugo Ramírez Leyva..... | 2017-2038 |
| Desarrollo de un frecuencímetro de conteo recíproco con un FPGA para sensores basados en resonador de cuarzo José Lorenzo Muñoz Mata, Severino Muñoz Aguirre, Juan Jesús Jiménez Arellano, Georgina Beltrán Pérez, Juan Castillo Mixcóatl, Takamichi Nakamoto..... | 2039-2053 |

PRESENTACIÓN

El avance que la industria electrónica a nivel mundial ha tenido en las dos últimas décadas ha sido impresionante, en la actualidad los dispositivos y equipos electrónicos se encuentran en prácticamente todas las áreas del quehacer humano. Así mismo, las áreas de investigación y de aplicación se multiplican cada día dando paso a nuevas ramas afines a la electrónica y sus aplicaciones. Entre estas tenemos: robótica, biónica, telemática, nano-electrónica, domótica, automatización, telecomunicaciones, electrónica de potencia, sistemas computacionales, procesamiento de señales y educación, por mencionar sólo algunas.

Los resultados de los trabajos de investigación, así como de las aplicaciones tecnológicas, que realizan los grupos de profesores y de estudiantes de las instituciones educativas y de investigación son numerosos y variados.

La Semana Nacional de Ingeniería Electrónica (SENIE) es un congreso nacional que busca servir de foro en el cual los investigadores y estudiantes de las diversas áreas de la Ingeniería Electrónica puedan intercambiar experiencias, conocimientos y establecer vínculos de colaboración, mediante actividades tales como: presentación de ponencias, conferencias magistrales, conferencias técnicas y exposición industrial.

La décimo primera edición del evento se realizó del 7 al 9 de octubre en las instalaciones de la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí. Esta institución goza de un reconocido prestigio a nivel nacional, por la calidad de enseñanza que se imparte, la investigación y divulgación del quehacer científico y la alta habilitación de su planta de profesores-investigadores.

Se recibieron 183 trabajos para su evaluación provenientes de 91 distintas instituciones de educación superior y centros de investigación tanto públicos como privados de 26 estados del país. La revisión de los trabajos fue realizada por un Comité Técnico formado por más de 70 especialistas en las distintas temáticas, varios de ellos de adscritos a instituciones en el extranjero. Como resultado del proceso de revisión, fueron aceptados para su presentación 109 artículos.

En este número especial de *Pistas Educativas*, el cual recopila los artículos expuestos en el evento, el lector encontrará una muestra representativa de los trabajos realizados en el último año por investigadores, profesores y estudiantes de diversos Centros de Investigación y Educación Superior del país.

Es de esta forma que la División de Ciencias Básicas e Ingeniería de la Universidad Autónoma Metropolitana campus Azcapotzalco, la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí y el Instituto Tecnológico de Celaya, cumplen con el objetivo de proporcionar un foro para la divulgación de los trabajos de investigación relacionados con la Ingeniería Electrónica y sus diversas ramas.

Dr. Miguel Magos Rivera
Presidente del Comité Organizador
XI Semana Nacional de Ingeniería Electrónica 2015

Inicialización de un robot humanoide tipo BIOLOID para tareas de caminado

Antonio Benítez Ruiz

Universidad Politécnica de Puebla, Departamento de Posgrado, Tercer Carril del Ejido Serrano S/N Juan
C. Bonilla, Puebla, Teléfono: (01 222) 7 - 74 - 66 – 64
antonio.benitez@uppuebla.edu.mx

María Auxilio Medina Nieto

Universidad Politécnica de Puebla, Departamento de Posgrado, Tercer Carril del Ejido Serrano S/N Juan
C. Bonilla, Puebla, Teléfono: (01 222) 7 - 74 - 66 – 64
maria.medina@uppuebla.edu.mx

Jorge de la Calleja Mora

Universidad Politécnica de Puebla, Departamento de Posgrado, Tercer Carril del Ejido Serrano S/N Juan
C. Bonilla, Puebla, Teléfono: (01 222) 7 - 74 - 66 – 64
jorge.delacalleja@uppuebla.edu.mx

Emmanuel Baleón Flores

Universidad Politécnica de Puebla, Departamento de Posgrado, Tercer Carril del Ejido Serrano S/N Juan
C. Bonilla, Puebla, Teléfono: (01 222) 7 - 74 - 66 – 64
emmanuel.baleo@uppuebla.edu.mx

Carmina Cruz Ochoa

Universidad Politécnica de Puebla, Departamento de Posgrado, Tercer Carril del Ejido Serrano S/N Juan
C. Bonilla, Puebla, Teléfono: (01 222) 7 - 74 - 66 – 64
carmina.cruz@uppuebla.edu.mx

Ivonne López Cuacuas

Universidad Politécnica de Puebla, Departamento de Posgrado, Tercer Carril del Ejido Serrano S/N Juan
C. Bonilla, Puebla, Teléfono: (01 222) 7 - 74 - 66 - 64
Ivonne.lopez@uppuebla.edu.mx

Resumen

El presente trabajo describe el desarrollo de rutinas de comportamiento reactivo para un robot humanoide con 16 articulaciones. Específicamente, esta enfocado a llevar a un robot humanoide a una configuración inicial que le permita poder ejecutar cualquier tipo de rutina de locomoción. Todos estos comportamientos dependen de la programación de servomotores así como del uso de un sensor de giro o giroscopio. Con este sensor se determina a través de un evento si el robot perdió la vertical y ha caído al piso, en cuyo caso ejecuta una rutina para llevar al robot nuevamente a una configuración que le permita continuar con su operación. El trabajo esta soportado en un proceso experimental para llegar a obtener un conjunto de parámetros que aseguran el buen funcionamiento de las rutinas, como son las velocidades de operación y valores que arroja el sensor de giro. Los resultados son muy satisfactorios tomando en cuenta que para los robots bípedos o humanoides es muy fácil que pierdan el equilibrio dadas las distintas rutinas de locomoción que se implementan y que provocan la caída del robot. Así, esta propuesta permite estar censando el estado del robot para que en cuanto éste caiga al suelo ejecute alguna de las dos rutinas de reincorporación propuestas.

Palabra(s) Clave(s): robótica reactiva, robots bípedos, sensores.

1. Introducción

La robótica reactiva es una corriente de la robótica que surgió en la década de los 80s, que permitió el incremento en la capacidad de reacción en los robots con características de computo reducidas [1]. Se denomina reactiva porque el robot no tiene un modelo del entorno, sino reacciona ante los estímulos de éste [2]. La robótica reactiva se caracteriza por lo siguiente:

- Las tareas se descomponen en pequeñas sub tareas llamadas comportamientos.
- Los comportamientos se procesan en paralelo.
- Los comportamientos compiten por el control de los actuadores.

- Existe un arbitraje que controla esta competencia.
- En muchos casos los comportamientos se pueden implementar con facilidad como autómatas de estados finitos.

Desde hace muchos años, se ha venido trabajando en el desarrollo de sistemas robóticos con características humanas, esto es, robots que puedan ejecutar comportamientos parecidos a los que los seres humanos son capaces de hacer. Así, diferentes empresas dedicadas al desarrollo de estos robots han puesto a disposición de estudiantes e investigadores diferentes plataformas de robots humanoides o robots bípedos. Estas plataformas varían de manera significativa en sus funcionalidades y capacidades, las cuales dependen directamente de los sensores que se incorporan a ellas y el costo asociado.

Sin duda, uno de los retos principales de los investigadores dedicados a este tipo de robots consiste en contar con rutinas asociadas al caminado bípedo de estos robots. Sin embargo, para poder desempeñar esta actividad es necesario colocar al robot en cierta configuración inicial, esto es, que el robot este de pie y listo para caminar.

Hoy en día, los robots bípedos empiezan a jugar un papel muy importante en el desarrollo de aplicaciones capaces de interactuar con seres humanos. Por esta razón, una de las aplicaciones que empiezan a tomar fuerza son las relacionadas con rutinas de rehabilitación, específicamente rutinas que permitan la interacción entre robots y niños, buscando que el robot se convierta en un instrumento de ayuda en distintas actividades que apoyen la educación y rehabilitación de los niños.

El presente trabajo tiene como objetivo llevar al robot humanoide a una configuración inicial. Esta configuración consiste en colocar al robot de pie con la posición de las articulaciones de tal manera que pueda desempeñar la tarea de caminado o algún otro comportamiento. Para poder llevar al robot a esta configuración, es necesario considerar que el robot al ejecutar algún comportamiento, cualesquiera que este sea, el robot puede caer al suelo, ya sea en una posición boca arriba(decúbito prono o boca

abajo(decúbito dorsal). Así, parte del trabajo presenta como levantar al robot para poder llevarlo a la configuración inicial deseada.

Es importante decir que este proyecto involucra elementos de hardware como son la calibración y configuración de los motores y sensores que forman la estructura del robot [2]; y por otro lado, elementos de software que permiten el desarrollo y programación de rutinas que controlan el buen funcionamiento de los elementos electrónicos y mecánicos [7].

Este proyecto se desarrolló utilizando una plataforma de robots bípedos llamada BIOLOID – Premium Kit [9]. Esta arquitectura cuenta con 18 servomotores distribuidos en las articulaciones que asemejan un cuerpo humano como se muestra en la Figura 1.

2. Desarrollo

2.1. Preliminares

Un actuador es un dispositivo inherentemente mecánico cuya función es proporcionar fuerza para mover o “actuar” hacia otro dispositivo mecánico [3]. La fuerza que provoca el actuador proviene de tres fuentes posibles: presión neumática, presión hidráulica, y fuerza motriz eléctrica (motor eléctrico o solenoide). Dependiendo del origen de la fuerza del actuador, éste se denomina “actuador neumático”, “actuador hidráulico, o “actuador eléctrico”.

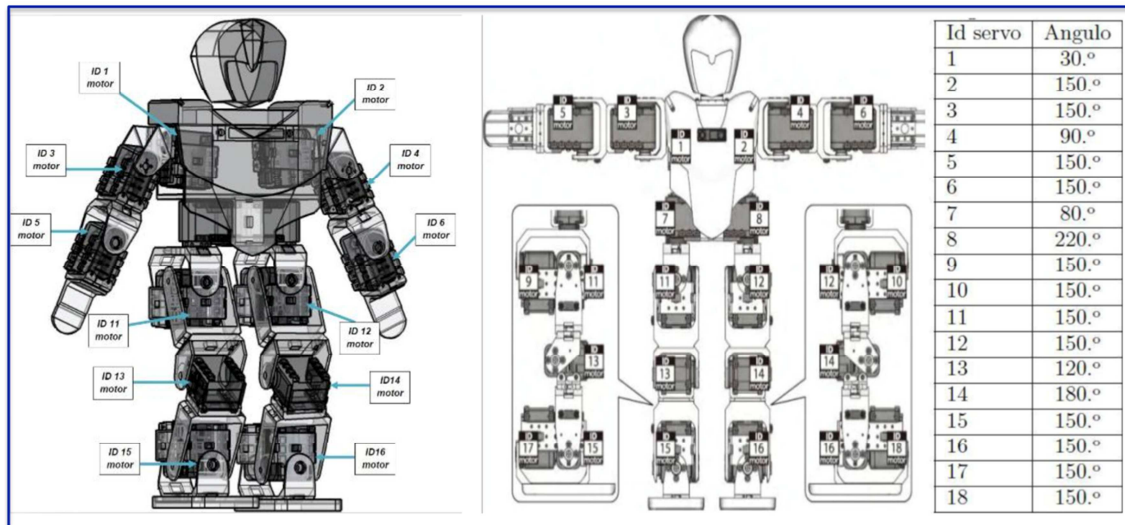


Fig. 1. Distribución de servomotores en robot humanoide con 16 articulaciones y su configuración inicial incluyendo los valores para cada servomotor.

2.1.1. Servo

Un servo cuenta con un encapsulado electrónico y se define como un motor de corriente directa, principalmente se utiliza para el modelado de aviones, carros y barcos. Un servo está compuesto por tres cables: alimentación, tierra y señal de control [8]. La señal utilizada por los servos siempre tiene una frecuencia de 50 Hz y sus pulsos son generados cada 20 ms. El tamaño de cada pulso especifica la posición deseada para mover el servo, tanto los servos y los motores a pasos implementan algunas soluciones para las tareas realizadas en la robótica.

2.1.2. Servomotor Dynamixel AX-12+

El servomotor Dynamixel AX-12+ es un actuador inteligente que tiene incorporado un reductor de engranes, un motor de precisión DC y circuitos de control con

funcionamiento en red. Pueden estar conectados hasta 254 Dynamixel AX-12+ en serie a la Dual-POB con un solo cable, también tienen la capacidad de ser conectados uno a otro para crear cadenas de servomotores (Daisy-Chain), el cable soporta una transmisión de 1M BPS. La velocidad y la posición del servomotor está controlada con una resolución de 1024 pasos. Posee pruebas de retroalimentación con capacidad de reconocer la posición angular, la velocidad angular y la carga de torque válido. En la figura 2, se muestran los rangos de movimiento para el dynamixel AX-12 [9,10,11].

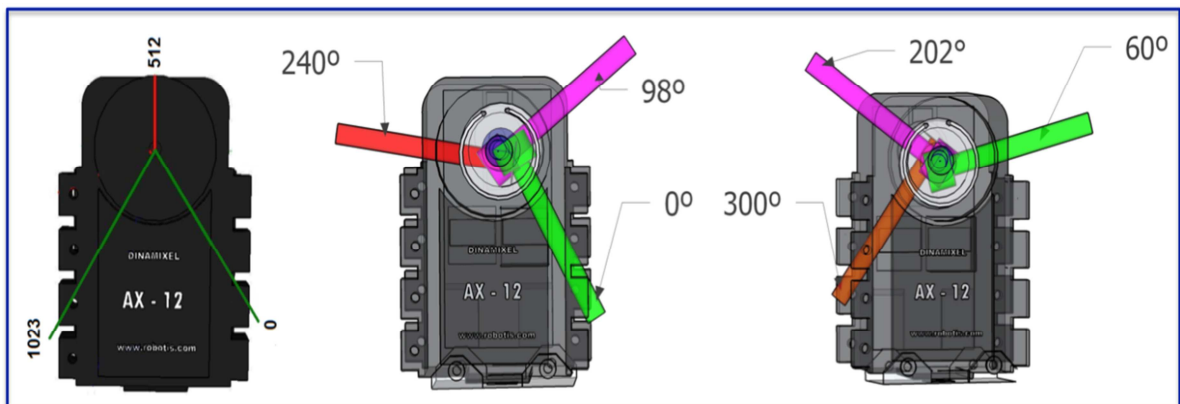


Fig. 2. Cada servomotor sigue la siguiente configuración: tiene un rango de movimiento de 0 a 300 grados ó de 0 a 1024 pasos (lado izq.); cada articulación se configura de manera diferente (indicador verde: rango mínimo de movilidad; indicador rojo: rango máximo e indicador lila; posición actual).

2.1.3. Sensores

Los sensores trasladan la información desde el mundo real al mundo abstracto de los micro controladores. En general, la mayoría de los sensores pueden ser divididos en dos grandes grupos: sensores analógicos y sensor digitales [5].

2.1.4. Sensor de giro o giroscopio

La definición de giroscopio fue impuesta en el año de 1851 por el historiador Michel Foucault, que intentó demostrar la rotación de la tierra [5,6]. “un giroscopio consiste en un rotor montado en un conjunto de anillos de cardan o balancines”, Un giroscopio

actúa con dos fuerzas las cuales son la fuerza gravitacional con dirección hacia arriba y la fuerza gravitacional con dirección hacia abajo con el centro de giro punto 0 [12].

El robot utilizado en este proyecto incorpora con un sensor de giro. Este sensor cuenta con 2 ejes (**X** y **Y**), el eje **X** determina la velocidad angular aplicada en el frente o en la parte trasera del robot; y el eje **Y** determina la velocidad angular aplicada en la parte izquierda o derecha del robot. Utilizando los ejes **X** y **Y** se puede identificar la inclinación del robot o la dirección que este consumiendo mayor potencia. Otra de las funcionalidades del sensor de giro es la implementación en el equilibrio para tareas de caminado.

2.2. Metodología

El proyecto se modeló a través de un diagrama de estados los cuales son: *conocer la velocidad angular, identificar posición* (decúbito prono o decúbito dorsal), *reincorporación a configuración inicial, caminar con dirección hacia atrás*. En la figura 3, se muestra un diagrama que ilustra cada uno de los eventos y estados con los que el robot reaccionará.

2.2.1. Censando de la posición angular

Para realizar esta actividad, es necesario utilizar el giroscopio en el rango de resultados que arroja el giroscopio cuando el robot se encuentra en la configuración inicial, la cual es en el eje **X** (250 - 240) y en el eje **Y** (250 - 240). En base a un proceso de experimentación, se desean establecer cuáles son los rangos de los ejes **X** y **Y** que permitan identificar si el robot está cayendo y hacia qué dirección. Se hicieron cincuenta mediciones de los ejes **X** y **Y** por cada movimientos que realice el robot y la rutina de identificación de posición dará inicio para pasar al estado de identificando posición.

2.2.2. Identificando la posición y reincorporación

En este estado, es necesario identificar cuál es la posición del robot, se tiene que identificar si el robot se encuentra en el piso boca abajo o boca arriba. Esto permite determinar qué comportamiento necesita ejecutar el robot para regresar a su configuración inicial.

La información obtenida por el giroscopio será comparada con los rangos que se establecerán para los ejes (X y Y). Si la información del giroscopio supera los rangos permitidos, entonces se establece que el robot tiene una inclinación para cierto lado (izquierda, derecha, enfrente o atrás). Conociendo la inclinación del robot, se determina en que posición cayó (boca arriba o boca abajo) y se activa la rutina de reincorporación necesaria para llevarlo a la configuración inicial.

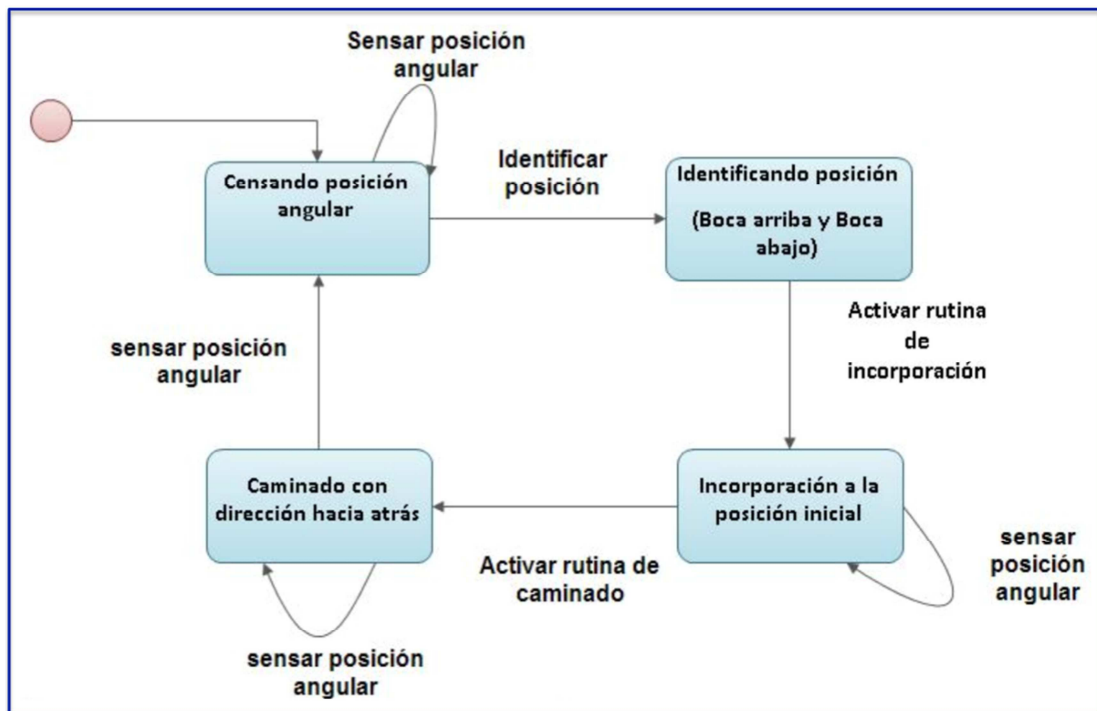


Fig. 3. Diagrama de estados para los distintos comportamientos que ejecutará el robot.

Para poder llevar el robot a la configuración inicial, es necesario la coordinación de los servomotores para ejecutar la rutina de reincorporación. Teniendo definida la posición en la que el robot cayó, se ejecuta la rutina de reincorporación.

2.2.3. Caminado hacia atrás

Se implemento un comportamiento de caminado hacia atrás con el objetivo de probar las rutinas de reincorporación. Mientras el robot ejecuta las secuencias necesarias para ejecutar el caminado hacia atrás, cada movimiento pasa por el estado de *censando posición angular* y a su vez, se activa el estado *identificando posición*. Si el movimiento realizado por el robot provoca una caída, entonces se ejecutará la rutina de reincorporación hacia la configuración inicial. En caso de que no se ejecute el estado de reincorporación, se determina que se tiene un caminado limpio, es decir, sin caídas o interrupciones.

2.3. Implementación de rutinas

Ya que uno de los principales problemas de los robots bípedos es mantenerse en equilibrio, el implementar comportamientos en el robot puede generar la pérdida del equilibrio. Por esta razón, es necesario identificar las posiciones en los servos que brinden mayor estabilidad en el robot, así como las velocidades a las que se ejecutará cada movimiento, para que cuando el robot pierda el equilibrio se pueda identificar hacia que dirección está cayendo (hacia el frente o hacia atrás) y qué rutinas se ejecutarán para reincorporarlo.

2.3.1. Proceso general de movimiento del robot humanoide

Cada movimiento del robot está sujeto a cuatro tareas, si alguna de estas tareas no es completada, el movimiento no se realizará. Las tareas a ejecutar son presentadas en la figura 4.

En el proceso de configuración de posiciones es donde se introducen los ángulos para cada servomotor. Estos ángulos se almacenan en una matriz de 18 columnas (dado que son 18 servos en el robot) y de un número indefinido de filas (tantas como secuencias se deseen enviar al robot). Cabe aclarar que cada fila es una postura que toma un servomotor y una secuencia (ó comportamiento) está formada por una o varias filas.

El proceso de administración de secuencia y velocidades, permite determinar cuántos movimientos se necesitan hacer para ejecutar una secuencia. En este proceso también se controla la velocidad global (aplicada a todos los servos) y la velocidad independiente (para cada servo). Por último, en esta parte se transforman los ángulos en pasos (recordar que los servomotores se manejan en pasos y no en ángulos).

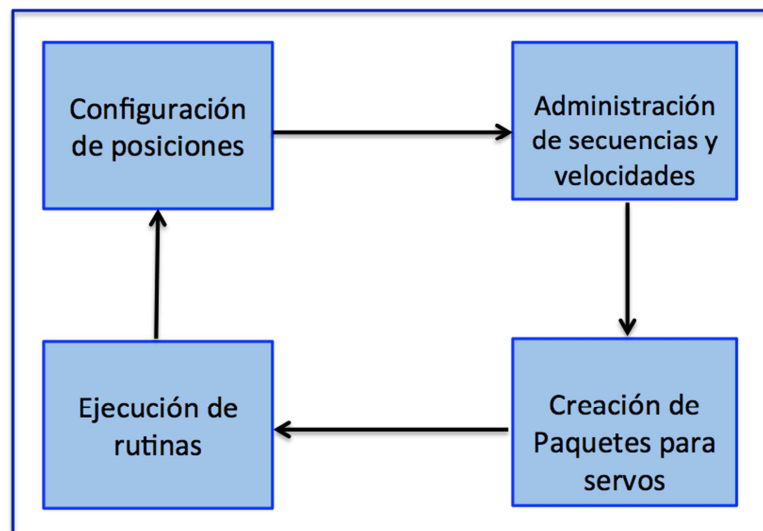


Fig. 4. Diagrama del proceso general para la generación de movimientos en el robot.

La creación de paquetes es un proceso que permite preparar un conjunto de instrucciones que integran la velocidad y la posición que se desea que tome cada servomotor. Este proceso es necesario para controlar los movimientos síncronos entre los servomotores.

En el proceso de *ejecución de rutinas*, se leen los paquetes recibidos y se ejecutan las instrucciones que contiene cada paquete. Este proceso lo ejecuta de forma interna cada dinamixel AX-12.

2.3.2. Control de las velocidades

El control de la velocidad depende del movimiento que se busca desde la posición actual hacia la posición deseada. El diagrama de la figura 5, muestra el proceso en el cuál se asigna la velocidad. Cada movimiento realizado por el robot pasa por este proceso. La velocidad permite controlar la transición de los movimientos de la posición inicial a la posición deseada de forma que no afecte el equilibrio del robot.

El intervalo permitido para la velocidad de cada servomotor esta en el rango de [0 a 1023] y cada unidad es proporcional a 0.111 rpm (revoluciones por minuto). Si el valor que ingresa es 0, indica que no existe un control de la velocidad y se utiliza la máxima. Si se ingresa un valor de 1023, éste es equivalente a 114 rpm. Las pruebas iniciales en este proyecto se hicieron con los valores de 120 como velocidad global y 60 como velocidad independiente, esto equivale a 13.32 rpm y 6.66 rpm. Es recomendable no utilizar la velocidad máxima, ya que esto provocaría que el servomotor se pueda quemar. De forma general, la velocidad máxima utilizada en este proyecto es de 200 (22.2 rpm) y la velocidad mínima es de 15 (1.665 rpm).□

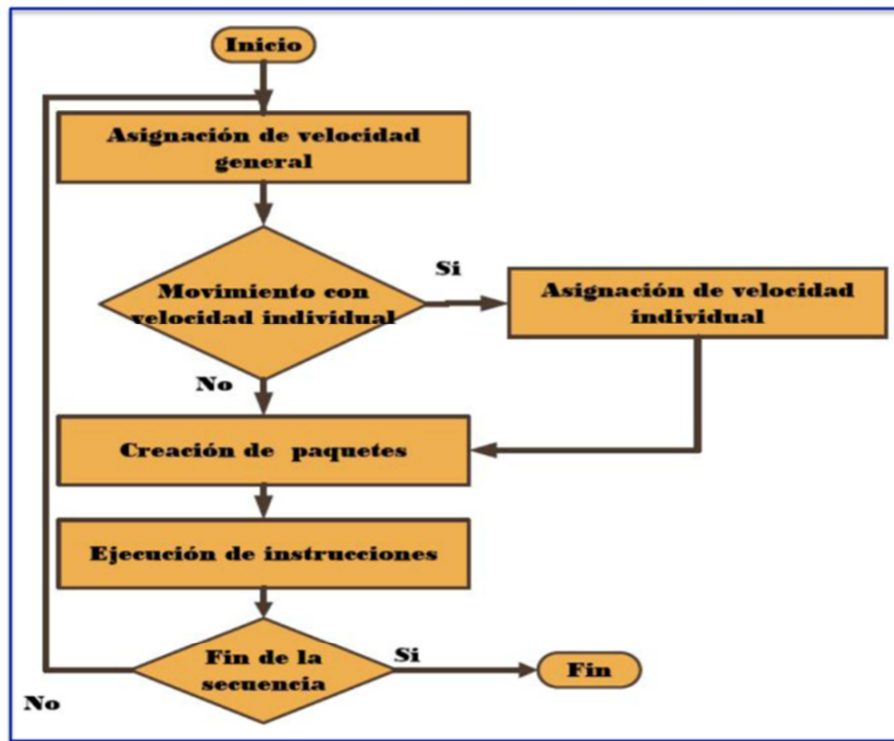


Fig. 5. Diagrama para la configuración de velocidad de operación de los servomotores.

2.3.3. Configuración de reincorporación

Las rutinas de reincorporación permiten al robot humanoide pasar de la posición decúbito prono o decúbito dorsal a la configuración inicial. En la figura 6, se presenta el diagrama del proceso de la reincorporación.

Después de un proceso de experimentación, se encontraron los siguientes valores permitidos de inclinación para el robot, mismo que se presentan en la Tabla 1.

Completada la ejecución de la rutina de reincorporación, el robot continua con la rutina de locomoción que estaba ejecutando hasta antes de perder el equilibrio y caer. Cada movimiento sucesivo seguirá siendo censado a través del giroscopio si es necesario ejecutar nuevamente la rutina de reincorporación.

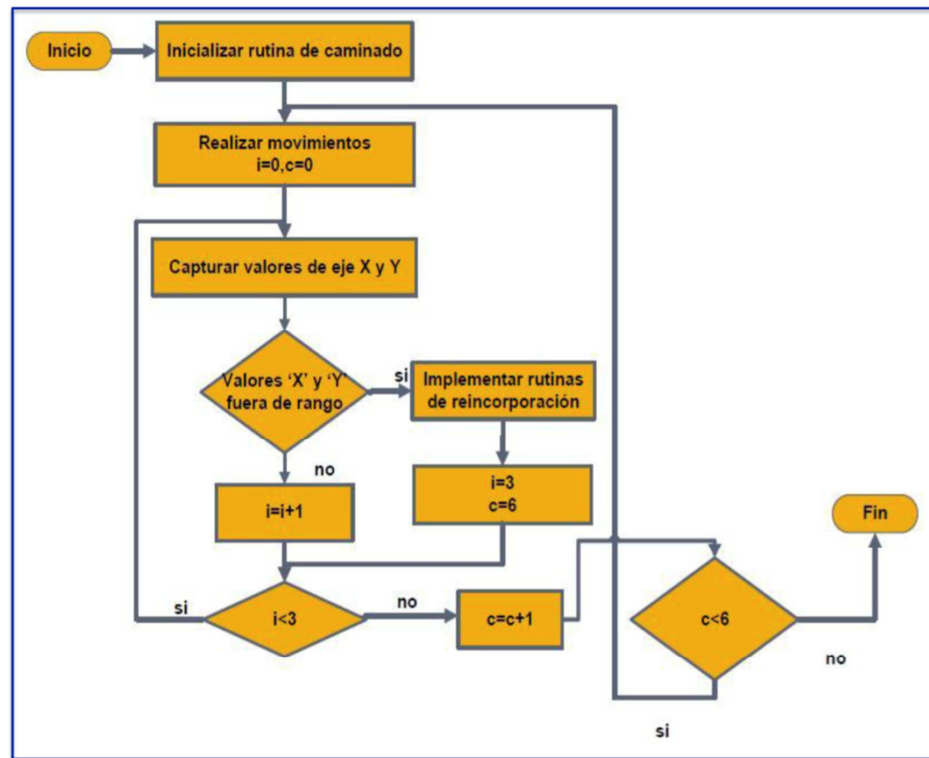


Fig. 6. Diagrama de flujo para la reincorporación del robot hacia la configuración inicial.

En este proyecto la rutina de locomoción utilizada consiste en desplazar al robot hacia atrás utilizando pasos pequeños (caminado con dirección hacia atrás) y por cada paso que el robot ejecuta se revisa la inclinación que registra el sensor de giro.

2.3.4. Reincorporación decúbito prono

La rutina de reincorporación decúbito prono está formada por 8 posiciones, mismas que se presentan en la tabla 2. Además, en la figura 7, se muestran las 8 posiciones necesarias para reincorporar al robot.

La velocidad general que maneja la rutina es de 200 (22.2 rpm) y las velocidades independientes se aplican en los dos últimos movimientos, los cuales llevan al robot a la configuración inicial. La Tabla 2, muestra la relación entre los servomotores y las velocidades independientes.

| Eje X | Eje Y | Posición |
|-------------------|-------------------|-----------------------|
| X < 215 | * | Decúbito Prono |
| X > 292 | * | Decúbito Dorsal |
| * | Y < 208 | Decúbito Dorsal |
| * | Y > 292 | Decúbito Dorsal |
| X > 208 & X < 292 | Y > 208 & Y < 292 | Configuración inicial |

Tabla 1. Valores permitidos de inclinación para el robot.

2.3.5. Reincorporación decúbito dorsal

El principal objetivo de esta rutina es girar al robot llevándolo de la posición decúbito dorsal a la posición decúbito prono, y utilizar parte de la rutina de reincorporación de cubito prono. La rutina de decúbito dorsal tiene una velocidad general de 100 (11.1 rpm) y la velocidades independientes de esta rutina son las mismas implementadas en la rutina decúbito prono, esta información se presenta en la tabla 2.

| Decúbito prono | | | | Decúbito dorsal | | | |
|----------------------|---------------|-----------|-------|-----------------|---------------|-----------|------|
| Posición | Id servomotor | Velocidad | RPM | Posición | Id servomotor | Velocidad | RPM |
| 6 | 11 | 70 | 7.77 | 1 | 9 | 100 | 11.1 |
| | 12 | 70 | 7.77 | | 10 | | |
| | 13 | 15 | 1.665 | | 11 | | |
| | 14 | 15 | 1.665 | | 12 | | |
| | 15 | 15 | 1.665 | | 13 | | |
| | 16 | 15 | 1.665 | | 14 | | |
| Cofiguración inicial | 1 | 70 | 7.77 | 2 | 15 | 70 | 7.77 |
| | 2 | 70 | 7.77 | | 16 | | |
| | 11 | 65 | 7.215 | | 17 | | |
| | 12 | 65 | 7.215 | | 18 | | |
| | 13 | 50 | 5.55 | | 9 | | |
| | 14 | 50 | 5.55 | | 10 | | |
| Cofiguración inicial | 15 | 35 | 3.885 | 3 | 11 | 100 | 11.1 |
| | 16 | 35 | 3.885 | | 12 | | |
| | | | | | 13 | | |
| | | | | | 14 | | |
| | | | | | 15 | | |
| | | | | | 16 | | |
| Cofiguración inicial | | | | 4 | 17 | 100 | 11.1 |
| | | | | | 18 | | |
| | | | | | 9 | | |
| | | | | | 10 | | |
| | | | | | 11 | | |
| | | | | | 12 | | |
| Cofiguración inicial | | | | 5 | 13 | 70 | 7.77 |
| | | | | | 14 | | |
| | | | | | 15 | | |
| | | | | | 16 | | |
| | | | | | 17 | | |
| | | | | | 18 | | |
| Cofiguración inicial | | | | 6 | 9 | 100 | 11.1 |
| | | | | | 10 | | |
| | | | | | 11 | | |
| | | | | | 12 | | |
| | | | | | 13 | | |
| | | | | | 14 | | |

Tabla 2. Velocidades y revoluciones por minuto independientes para las rutina de decúbito prono y decúbito dorsal.

2.3.6. Rutina de caminado con dirección hacia atrás

El caminado con dirección hacia atrás está formado con seis posiciones que son equivalente a 2 pasos (uno izquierdo y uno derecho). Para ejecutar más pasos, se introduce esta secuencia en un ciclo dependiendo de la distancia que se necesite recorrer. Esta rutina manejan una velocidad general de 100 (11.1 RPM) y se aplican velocidades independientes para todo los movimientos.

3. Resultados

De manera general, los resultados obtenidos durante el desarrollo del proyecto fueron satisfactorios y se presentan en las subsecciones siguientes.

3.1. Medición de inclinación de robot

Los valores calibrados del giroscopio son utilizados exclusivamente para la rutina de caminado con dirección hacia atrás, por lo que si se desea implementar para alguna otra rutina es necesario recalibrar los valores.

Por otra parte, cada vez que se rebasan los valores del eje “Y” la caída esperada es del lado izquierdo si su valor es menor de 210, o del lado derecho si su valor es mayor de 290. Por cada 20 caídas del lado izquierdo o derecho, 19 de ellas terminaron en la posición decúbito dorsal, por este motivo en el eje “Y” siempre que los valores salga de los rango presentados en la tabla 2, se ejecutará la rutina de reincorporación decúbito dorsal.

Para mejorar la velocidad y estabilidad de la rutina de caminado con dirección hacia atrás las capturas del giroscopio se redujeron de 5 a solo 3 capturas por movimiento. En total, por cada dos pasos que el robot ejecuta, el giroscopio captura 18 veces los valores de los ejes “X” y “Y”.

3.2. Rutinas de reincorporación de decúbito prono y decúbito dorsal

La rutina de reincorporación de decúbito prono presenta inestabilidad en la cancha de futbol y lo hace caer antes de reincorporarse, por este motivo la rutina fue modificada. De esta manera se le hicieron más de 50 modificaciones. La figura 7, muestra las posturas que toma el robot para la reincorporación con la rutina de decúbito prono.

La rutina de reincorporación de decúbito dorsal contiene parte de la rutina de decúbito prono, por este motivo fueron aplicadas las mismas correcciones para mejorar el desempeño de la misma. La velocidad general implementada en esta rutina esta en el rango de 100 (11.1 rpm) a 200(22.2 rpm). Las velocidades independientes implementadas son las mismas que las consideradas en la rutina de decúbito prono. La rutina ejecuta los 10 movimientos con la velocidad de 200 (22.2 rpm) en un tiempo aproximado de 18 segundos. La secuencia de estos movimiento se puede ver en la figura 8.

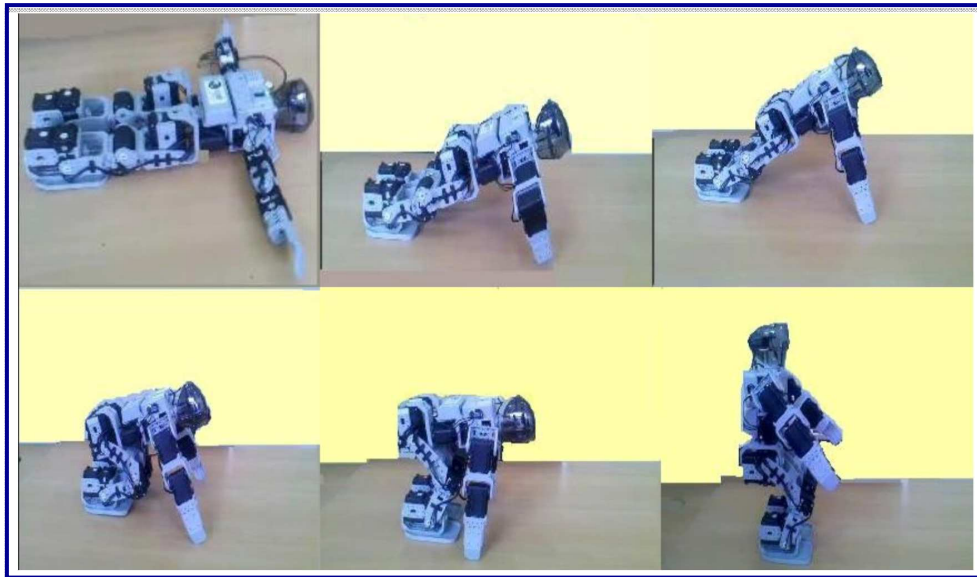


Fig. 7. Secuencia de movimientos para la rutina de reincorporación de decúbito prono



Fig. 8. Secuencia de 10 movimientos para la rutina de reincorporación de decúbito dorsal.

Por otra parte, la rutina presenta dos inconvenientes que afectan al robot y a la ejecución de la reincorporación. El primer inconveniente es el movimiento de la posición 3 a la posición 4, este movimiento implica una demanda excesiva en la fuerza de los servos 4 y 10 que cargan al robot para lograr la transición de la posición decúbito dorsal a la decúbito prono, por lo que el uso excesivo de esta rutina puede llegar a afectar a estos dos servomotores. El segundo inconveniente es el número excesivo de movimientos hace lenta la reincorporación. Por esta razón, se decidió generar una nueva rutina de reincorporación llamada decúbito dorsal-B, la cuál permitirá mejorar el tiempo de ejecución y evitar el desgaste de los servomotores.

La rutina de reincorporación de decúbito dorsal-B está formada por seis posiciones, tiene un desempeño estable utilizando una la velocidad general en un rango de 100 (11.1 rpm) a 200 (22.2 rpm).

En la figura 9, muestra las 6 posiciones con las que está conformada la rutina de reincorporación de cubito dorsal-B. Así mismo las velocidades independientes son implementadas para la posición 4,5.

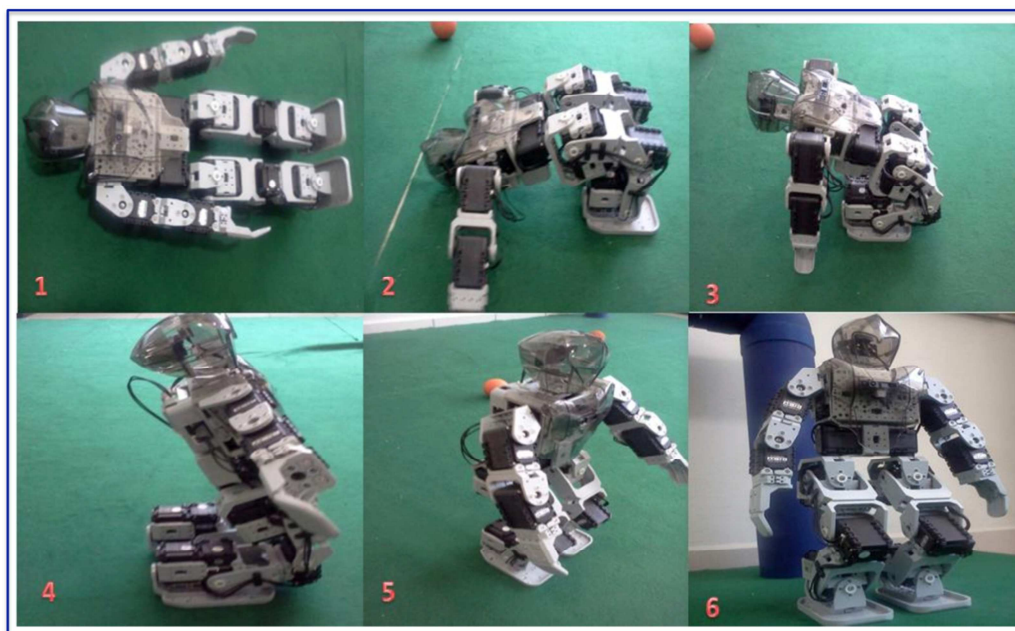


Fig. 9. Secuencia de movimientos para la rutina de reincorporación de decúbito dorsal-B.

Se debe agregar que la rutina de reincorporación decúbito dorsal-B ejecuta el total de sus movimientos en un tiempo aproximado de 11.045s, lo que indica que es un 62.36 % más rápida que la rutina de reincorporación decúbito dorsal. La efectividad que presenta esta rutina es de un 92%. De las 50 pruebas realizadas solo en cuatro ocasiones el robot no logro llegar a la configuración inicial.

La rutina de caminado con dirección hacia atrás, presenta inestabilidad durante su ejecución. En 20 pruebas desarrolladas el robot se cayó en 19 ocasiones, la distancia aproximada que recorre por cada 20 pasos es de 15 cm en un tiempo aproximado de 30 segundos. La rutina de caminado hacia atrás se sometió a más de 100 modificaciones que permitieron obtener una rutina con mayor estabilidad en el caminado y disminuir los tiempos de ejecución de la misma.

Con los cambios realizados de la rutina de caminado con dirección hacia atrás, por cada 20 pasos que el robot ejecuta la distancia que recorre aproximadamente es de 89cm en un tiempo de 1.14 minutos. En 20 pruebas realizadas de la rutina de caminado con dirección hacia atrás solo en una ocasión el robot tuvo una caída.

Las modificaciones implementadas para la rutina de caminado hacia atrás hicieron que el desempeño de la rutina mejorará. Antes de la modificaciones el riesgo de que el robot callera era de un 95%, con las modificaciones implementadas el porcentaje decremento a un 5 % y la distancias que recorre el robot a mejor-o en un 240.5 %.

4. Conclusiones

En este proyecto se desarrollaron rutinas que permiten la reincorporación de un robot humanoide utilizando un giroscopio para identificar la inclinación del robot. El objetivo principal es desarrollar rutinas de caminado hacia atrás, identificar si durante la rutina de caminado existe alguna caída, entonces reincorporar al robot y seguir con la rutina de caminado hacia atrás.

Las rutinas de reincorporación son parte esencial para la implementación de otras rutinas de movimiento, principalmente rutinas donde intervengan las piernas del robot. Es importante mencionar que cualquier movimiento implementado sobre las piernas del robot, afectará directamente al centro de masa del mismo. Además, el método implementado para medir el desequilibrio del robot es rápido, lo que permite tener un control constante sobre el equilibrio del robot.

Las rutinas de reincorporación son una opción para corregir los errores ocurridos en rutinas implementadas sobre las piernas, si bien no previenen la caída del robot, si permite que el robot regrese a su configuración inicial y pueda ejecutar alguna otra rutina. □

Por otra parte, el caminado es una de las rutinas que presenta más inestabilidad y siempre dependerá de la superficie donde el robot ejecute el caminado. La rutina de caminado desarrollada durante el proyecto consta de seis movimientos, si embargo, si esos seis movimientos se dividieran en mas sub movimientos y se disminuyera el tiempo entre las pausas, se podría aumentar la velocidad sin que esto afecte al centro de masa del robot. Cada movimiento que el robot ejecuta en la rutina de caminado es

monitoreado por el giroscopio en tres ocasiones, lo cual permite tener un rango aproximado de medición de 0.8 segundos. Sin embargo, la implementación del giroscopio tiene algunas limitantes que pueden ser eliminadas con la ayuda de otros sensores, tal como acelerómetros, sensores de presión, cámaras, sensores de proximidad, etc.

Este trabajo es el punto de partida para dotar al robot de comportamiento o rutinas de locomoción mas complejas y con cierto grado de inteligencia [4], por ejemplo: comportamientos propios de un jugador de fut-bol como portero, tiro a gol, seguimiento de una pelota, localización de porterías ó rutinas de rehabilitación motriz. Todos estos comportamientos requieren que el robot los ejecute estando de pie. Por otro lado, es importante decir que se han estado desarrollando diferentes comportamientos en paralelo a este trabajo y en todos ellos se están utilizando las rutinas de reincorporación del robot. Algunos videos de este trabajo se pueden ver en [13].

5. Referencias

- [1] Bruno Siciliano, Lorenzo Sciavicco, Luigi Villani, Giuseppe Oriolo, *Robotics Modelling*. 1° Edición. 2009. Planning and Control. Springer.
- [2] T. Brain, *Embedded Robotics Mobile Robot Design and Application with Embedde Systems*. Tercera Edición. 2008. Springer.
- [3] J. Roldan Vioria, *Arranque y protección de motores trifásicos*. Tercera Edición. 2005. Thomson.
- [4] Brooks A. Rodney, "Intelligence without representation". Vol. 545. No. 47. 1991. 139 – 159 pp.
- [5] Gonzales Arjona Davis, *Sensores*. Universidad Autónoma de Madrid.
- [6] Figueroa Medina Luis Enrique, "Simulación de comportamiento para robots humanoides en un juego de futbol". CINVESTAV, IPN. 2011.

- [7] Baleón Flores E, Benitez Ruiz A, Inicialización de un robot bípedo para tareas de caminado. Reporte Técnico "PII-51-08-12". Universidad Politécnica de Puebla. 2012.
- [8] Actuadores. Soltex Chile S.A.
- [9] ROBOTIS, API Reference. http://support.robotis.com/en/software/dynamixel_sdk/api_reference.htm/. Septiembre, 2013.
- [10] Dinamixel AX-12, ROBOTIS User Manual. <http://support.robotis.com/en/>. Septiembre, 2013.
- [11] Giro Sensor, ROBOTIS User Manual. <http://support.robotis.com/en/girosensor>. Septiembre, 2013.
- [12] Referencias de videos de reincorporación del robot. <https://www.youtube.com/watch?v=8vHo-Frveml>.

6. Autores

El Dr. Antonio Benitez Ruiz realizó el Doctorado en Ciencias de la Computación y la Maestría en Ciencias Computacionales en la Universidad de las Américas Puebla ([UDLAP](#)). Es egresado de la Licenciatura en Computación de la Facultad de Ciencias de la Computación ([FCC](#)) de la Benemérita Universidad Autónoma de Puebla (BUAP). Su trabajo de investigación está relacionado con las áreas de robótica reactiva, algoritmos de planificación de movimientos, graficación por computadora e interacción humano - robot. Cuenta con reconocimiento a Perfil Deseable del PRODEP desde 2007.

La Dra. María Auxilio Medina Nieto realizó el Doctorado en Ciencias de la Computación y la Maestría en Ingeniería en Sistemas Computacionales en la Universidad de las Américas Puebla (UDLAP). Es egresada de la Licenciatura en Computación de la Facultad de Ciencias de la Computación (FCC) de la Benemérita Universidad Autónoma de Puebla (BUAP). La Dra. Medina ha participado en proyectos relacionados con agentes, bibliotecas digitales, bases de datos y redes sociales; cuenta con el reconocimiento a perfil deseable de PRODEP desde 2005. Es miembro nivel candidato del Sistema Nacional de Investigadores (SNI). Actualmente, su investigación está dirigida a temas como web semántica, representación del conocimiento a través de

ontologías, tecnologías de la información y comunicación (TICs), e interacción humano – computadora.

Dr. Jorge de la Calleja Mora realizó el Doctorado en Ciencias Computacionales y la Maestría en Ciencias Computacionales en INAOE. Es egresado de la Licenciatura en Computación de la Facultad de Ciencias de la Computación (FCC) de la Benemérita Universidad Autónoma de Puebla (BUAP). Su trabajo de investigación está relacionado con las áreas de Aprendizaje Automático y Visión por Computadora. Cuenta con reconocimiento a Perfil Deseable del PRODEP desde 2007.

El Ing. Emmanuel Baleón Flores, la Ing. Carmina Cruz Ochoa y la Ing. Ivonne López Cuacuas. Son egresados de la Ingeniería en Informática de la Universidad Politécnica de Puebla (UPPue). Actualmente son estudiantes de la Maestría en Ingeniería en Sistemas y Cómputo Inteligente. Sus áreas de interés son Aprendizaje Automático, Visión por Computadora y Robótica Reactiva.

Diseño de un sistema de desarrollo para la emulación de tráfico en una intersección vial con vehículos autónomos

Andrés Ferreyra Ramírez

U.A.M.-Azcapotzalco, Av. San Pablo 180, C.P. 02200, Azcapotzalco, México D.F., Teléfono: 5318-9546

Ext. 1007

fra@correo.azc.uam.mx

Arturo Zúñiga López

U.A.M.-Azcapotzalco, Av. San Pablo 180, C.P. 02200, Azcapotzalco, México D.F., Teléfono: 5318-9546

Ext. 1007

azl@correo.azc.uam.mx

Roberto A. Alcántara Ramírez

U.A.M.-Azcapotzalco, Av. San Pablo 180, C.P. 02200, Azcapotzalco, México D.F., Teléfono: 5318-9546

Ext. 1000

raar@correo.azc.uam.mx

Resumen

El presente trabajo describe el desarrollo e implementación de un sistema para la emulación de tráfico en una intersección a partir de vehículos autónomos. Se desarrollan e implementan los protocolos de comunicación a través de los cuales los vehículos autónomos pueden transitar en una intersección de forma ordenada y sin colisiones. Emulando cadenas de vehículos mediante un grupo de robots móviles, los autores demuestran la factibilidad de las diferentes maniobras que se pueden realizar, utilizando diferentes estrategias de coordinación distribuidas que van desde ninguna

comunicación, a los intercambios unidireccionales o bidireccionales entre los vehículos, y la toma de decisiones totalmente centralizada del vehículo líder.

Palabra(s) Clave(s): intersección vial, pelotón, robot mOway, vehículos autónomos.

1. Introducción

Imagina una ciudad en la que circulen vehículos autónomos que recorran las avenidas a 90 km/hr en horas pico, operando en su totalidad sin la guía de un humano, evitando obstáculos y respetando las leyes de tránsito; vehículos que determinen la ruta más corta para llegar a su destino ahorrando tiempo y combustible, interactúen con la infraestructura del camino para determinar las condiciones de tráfico, respeten la distancia mínima entre vehículos, soliciten y cedan el paso, intercambien información. Este sueño comenzó en 2007 con el DARPA Urban Challenge que se llevó a cabo en la base George de la Fuerza Aérea en Victorville, California [1]; en donde se construyeron y presentaron vehículos autónomos capaces de conducir en el tráfico, realizando maniobras complejas. Fue un evento, en donde por primera vez, los vehículos autónomos interactuaron con el tráfico de vehículos tanto tripulados como no tripulados en un entorno urbano [2].

La conducción autónoma está progresando rápidamente y es probable que se convierta en el centro del transporte urbano en el futuro, pero ¿Qué tan apropiada será la infraestructura de tráfico para los vehículos autónomos?. Actualmente las intersecciones viales están controladas por semáforos que fueron diseñados para gestionar el tráfico y aumentar la seguridad de los peatones. Hoy en día, las intersecciones viales son consideradas cuellos de botella del transporte urbano, lo que genera una creciente preocupación acerca de su eficiencia y seguridad; este problema representa una gran oportunidad para introducir nuevos métodos que sean adecuados para la conducción autónoma en las intersecciones viales.

En [3] se propone un nuevo sistema de control de tráfico sin semáforos en una intersección, se asume que el sistema cuenta por completo con vehículos totalmente autónomos y la infraestructura para evitar colisiones; cuando los vehículos se acercan a una intersección, se comunican con un punto de acceso que se encarga de coordinar el movimiento de los mismos, mediante dos modos de operación: votación y aleatorio. El algoritmo propuesto está diseñado para redes de carreteras urbanas para mejorar la congestión vehicular y hacerla más predecible al mismo tiempo. El algoritmo utiliza los mecanismos IEEE 802.11 DCG/PCF para controlar el flujo de tráfico de los vehículos y recurre al concepto de contención/periodo libre de contención del IEEE 802.11 para controlar el equilibrio entre la eficiencia del flujo de tráfico y la equidad entre los usuarios.

En [4] se plantea el diseño de una intersección inteligente sin semáforos, los vehículos negocian la intersección a través de una interacción de la toma de decisiones centralizadas y distribuidas. En esta propuesta, se combinan las tecnologías de posicionamiento global GPS, las comunicaciones inalámbricas intervehicular, la capacidad de cálculo en el vehículo y la detección, para potenciar mejoras notables en la seguridad y eficiencia de la intersección vial. El algoritmo distribuido propuesto, garantiza la seguridad y proporciona un buen rendimiento.

En [5] se expone un control de tráfico basado en prioridad (CBP), una especie de sistema multi-agente que considera a los vehículos como agentes inteligentes y controla el tráfico a través de las interacciones entre estos. CBP es un protocolo común a través del cual, los agentes de los vehículos negocian sus prioridades cuando sus líneas de conducción interfieren mutuamente, con el fin de determinar el cruce en una intersección dada.

En [6] se formulan protocolos de intersección de nueva generación, basados en la tecnología de automóvil "Vehículo a Vehículo (V2V)", que aumentan significativamente el rendimiento de las intersecciones y evitan colisiones. Los protocolos que se

diseñaron pueden gestionar el tráfico a través de las intersecciones viales y glorietas, manteniendo la seguridad y mejorando el rendimiento.

En los trabajos citados, se considera que se cuenta con vehículos totalmente autónomos, con la infraestructura para evitar colisiones y solo se concentran en el desarrollo y simulación de los algoritmos de control; sin embargo, la toma de decisiones basada en suposiciones puede conducir a equivocaciones.

En este trabajo, se propone el desarrollo e implementación de un sistema para la emulación del control de tráfico en una intersección vial, sin semáforos. En este sistema, los vehículos autónomos son emulados con robots móviles que aunque no pueden modelar la dinámica real del vehículo, como los simuladores, pueden lidiar con las limitaciones del mundo real, tales como: percepción limitada, acciones imprecisas, latencia, toma de decisiones en tiempo real, informática embebida, y acontecimientos imprecisos. Los robots móviles son utilizados como caravanas de vehículos (también llamadas cadenas o pelotones) que siguen de manera autónoma a un vehículo líder [7, 8, 9, 10], para presentar dos estrategias de control: incorporación o separación de un miembro en el pelotón; las cuales se probaron en intersecciones con un único carril y una sola dirección, en una pista a escala especialmente diseñada para la aplicación.

2. Desarrollo

2.1. Los vehículos y la pista

Los vehículos

Los robots mOway (ver Fig. 1) son herramientas educativas que cuentan con suficientes sensores e indicadores [11] que les permiten interactuar con el entorno, así como también cuentan con un módulo de radio frecuencia a través del cual pueden realizar trabajo colaborativo. Estas características les permiten funcionar como vehículos autónomos. Para este proyecto se utilizaron 4 robots que fueron sometidos a un análisis

de compatibilidad tanto para el alcance de los sensores como para la velocidad de avance.



Fig. 1. Robot mOway.

La pista

La pista fue trazada con cinta adherible negra y las señales de intersección con cinta adherible azul como se muestra en la Fig. 2.

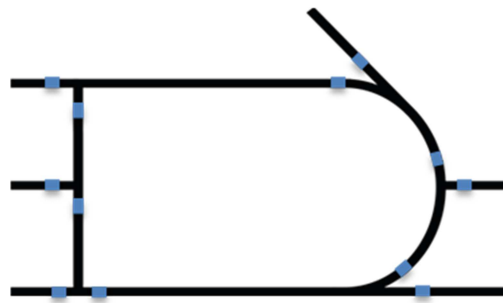


Fig. 2. Pista (primera versión).

Los sensores de línea de los robots, fueron sometidos a pruebas para ver su respuesta con respecto a los colores utilizados para la pista. La lectura del sensor de línea para el color blanco estuvo en un rango de 0 a 16, mientras que el color negro estuvo en un rango de 200 a 255 (que es el valor máximo). El sensor presenta un incremento de 16 a 200 que no es instantáneo, por lo que pasa por todos los valores intermedios. El color azul es medido en un rango de 28 a 120, por lo que cada vez que el sensor pasa de un

color blanco a un negro y viceversa, se reconoce el color azul sin estar presente en la pista. Para resolver este problema, la pista fue trazada con cinta azul y las señalizaciones con cinta negra.

Las esquinas de la pista fueron redondeadas en las intersecciones, para evitar que en el seguimiento de los robots en el pelotón, alguno de estos salga del campo de sensado como se muestra en la Fig. 3.

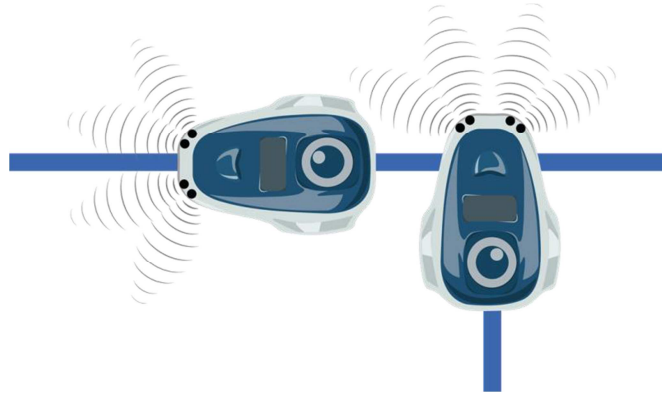


Fig. 3. Pelotón de robots recorriendo esquinas de 90°.

La versión final de la pista utilizada en el sistema, se muestra en la Fig. 4; en donde por uniformidad, todas las líneas inclinadas forman ángulos de 45°.

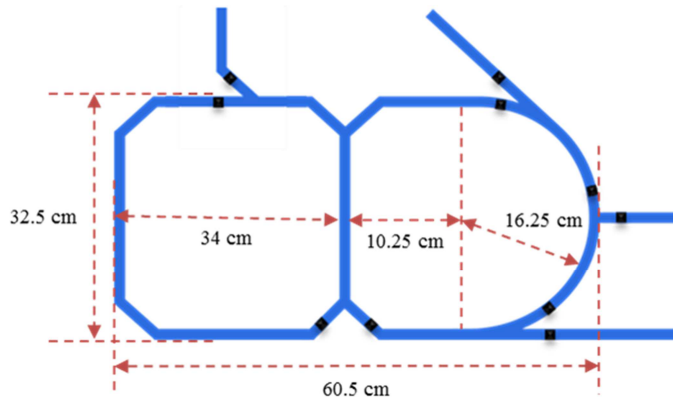


Fig. 4. Pista (versión final).

Dentro de la pista se definieron las secciones en donde se llevan a cabo las intersecciones y/o separaciones de los robots dentro del pelotón; estas se muestran en la Fig. 5.

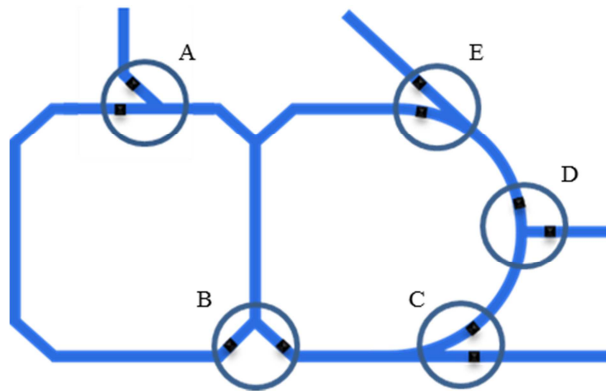


Fig. 5. Secciones de intersección y/o separación.

Las señalizaciones de intersección fueron iniciadas a cierta distancia antes del punto donde se unen los caminos, de tal forma que los robots tengan el tiempo suficiente para frenar y evitar colisiones, así como también dejar el suficiente espacio para permitir el libre tránsito del vehículo contrario. En la Fig. 5, las señalizaciones para las intersecciones A, B, y C, fueron colocadas aproximadamente a 6 cm; en el caso de las intersecciones C y E, estas fueron colocadas a una distancia aproximada de 8 cm.

2.2. Avance de los vehículos

El rango de velocidad de avance que se puede establecer a un robot varía entre 0 y 100; sin embargo, existe cierta diferencia entre los motores de cada uno de los robots que provoca que el avance de estos sea irregular aún cuando se les establezca la misma velocidad. Por lo anterior y debido a lo reducido de la pista que se diseñó, se decidió establecer una velocidad mínima de 10 para un vehículo que no se encuentre

siguiendo a otro, mientras que para un vehículo que sigue a otro la velocidad puede variar de 0 a 80.

Las intersecciones o separaciones de los vehículos que conforman el pelotón se llevo a cabo de la siguiente manera:

- Las intersecciones de dos caminos a uno.
- Las separaciones de uno a dos caminos.

Por lo que la ruta se guarda en un arreglo unidimensional, el cual sólo puede contener ceros que denotan vuelta a la izquierda o unos que denotan vuelta a la derecha.

En todo momento el avance de los vehículos está regido por los sensores de seguimiento de línea y los de reconocimiento de obstáculos. Los sensores utilizan la reflexión de la luz infrarroja que permite no sólo detectar contrastes fuertes sino también discernir entre diferentes tonos. Cuando se detecta una superficie clara, la superficie blanca hace que la luz infrarroja se refleje; por otro lado, cuando se detecta una superficie de color, sólo una parte de la luz emitida se refleja, de esta manera es como se realiza la identificación de colores.

Seguimiento de línea

El seguimiento de línea se realiza a través de los dos sensores de línea; esto es, cuando el vehículo debe avanzar por la izquierda, el sensor izquierdo debe de mantenerse en el color blanco, mientras que el sensor derecho debe permanecer en el color azul, tal como se ilustra en la Fig. 6.

Para lograr que el vehículo avance por la izquierda se establecieron las siguientes reglas:

- Si el sensor izquierdo esta sensando blanco y el sensor derecho esta sensando azul, el vehículo debe avanzar derecho.

- Si ambos sensores están sensando blanco, el vehículo debe girar a la derecha hasta que el sensor derecho sense el color azul (ver Fig. 7).
- Si ambos sensores están sensando azul, o si el sensor izquierdo está sensando azul y el derecho esta sensando blanco, el vehículo debe girar a la izquierda hasta que el sensor izquierdo sense el color blanco (ver Fig. 8).

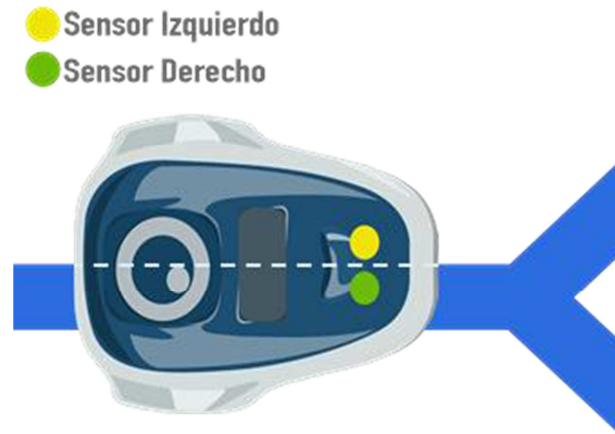


Fig. 6. Seguimiento por la izquierda.

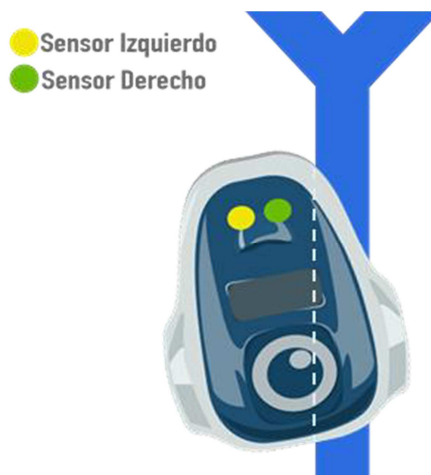


Fig. 7. Ambos sensores de línea sensando blanco.

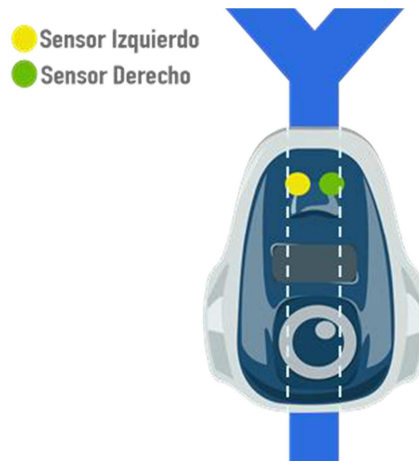


Fig. 8. Ambos sensores de línea sensando azul.

El rango numérico que se estableció para cada color (ver Tabla 1) se realizó con base al análisis del alcance de los sensores.

| Color | Sensor izquierdo | Sensor derecho |
|--------|------------------|----------------|
| Blanco | 0-14 | 0-16 |
| Azul | 15-120 | 17-140 |
| Negro | 121-255 | 141-255 |

Tabla 1. Rango numérico establecido para el sensado de colores.

Velocidad del vehículo

Los robots cuentan con 4 sensores de obstáculos que pueden ser utilizados para determinar la velocidad de un vehículo. Para poner estos sensores en función de la velocidad, es necesario obtener el promedio de los sensores que se encuentren activos; el rango de cada sensor es de 0 a 255, siendo 255 la lectura obtenida cuando un objeto se encuentra relativamente cerca, si el sensor está justo enfrente del objeto y el objeto no permite el paso de la luz, a pesar de que el sensor debería de entregar una lectura de 255, entrega una lectura cercana a 0 como se muestra en la Fig. 9.

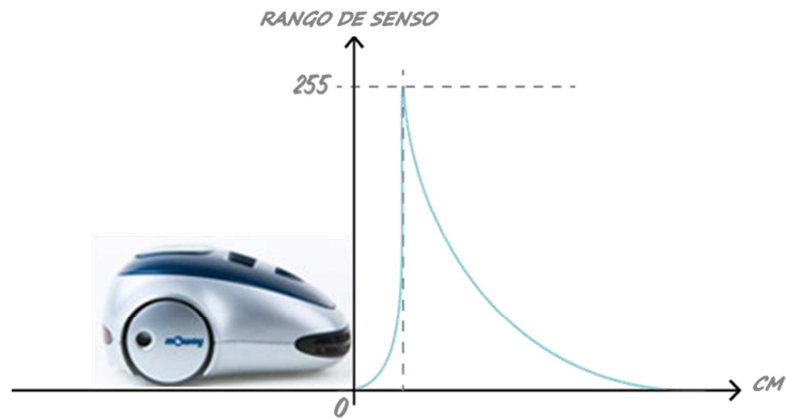


Fig. 9. Gráfica de sensado de obstáculos.

Algunos de los sensores de objetos son muy sensibles al entorno, por lo que a pesar de no existir un objeto, pueden entregar lecturas denotando lo contrario; estas lecturas erróneas tienen un rango aproximado de 0 a 3 por lo cual se acordó que las lecturas menores o iguales a 5 sean omitidas.

Para asegurar que los vehículos guarden cierta distancia entre ellos en todo momento, a partir de la lectura 210 la velocidad debe ser cero, por lo tanto un sensor de obstáculos se considera activo cuando este entrega lecturas intermedias entre 5 y 210.

2.3. Protocolo de Intersección

En el momento en que los sensores de línea reconocen la señalización de intersección (cinta negra), se activa una variable que funciona como temporizador y la lectura del sensor pasa de 0 a 120; la variable se va reduciendo de uno en uno y el tiempo que tarda en llegar a cero es el tiempo aproximado que tarda el vehículo en atravesar la intersección, este lapso de tiempo se reconoce como tiempo de intersección t_{int} .

Cuando el vehículo se encuentra en t_{int} este debe de abrir el canal de comunicación e inmediatamente comenzar a enviar y a recibir mensajes para determinar si hay otro vehículo en la intersección, de no ser así, el vehículo no necesita detenerse. El mensaje

que se envía durante el t_{int} contiene el valor de la variable que funciona como temporizador del t_{int} , ya que este valor denota la distancia de la intersección a la que se encuentra el vehículo; si un vehículo envía un valor aproximado de 120, entonces el vehículo acaba de entrar en la intersección; si un vehículo envía un valor cercano a 0, entonces el vehículo está por salir de la intersección.

Con base a lo anterior si un vehículo A envía un valor de 80 y recibe el valor de 120 de un vehículo B, el vehículo A debe dejar de enviar mensajes y cerrar el canal de comunicación puesto que llegó primero a la intersección y por lo tanto tiene preferencia de paso; en este caso el vehículo B debe detenerse inmediatamente y esperar a que el vehículo A recorra la intersección. Durante el tiempo de espera, la variable que funciona como temporizador del t_{int} es detenida, a pesar de esto el vehículo B debe de seguir enviando mensajes para averiguar si existe un tercer vehículo C siguiendo al vehículo A y de ser el caso volver a repetir el proceso que realizó con el vehículo A, de lo contrario cuando el tiempo de espera termina, el t_{int} continua disminuyéndose y el vehículo B debe comenzar a seguir al vehículo A. En el caso en que dos vehículos lleguen al mismo tiempo a la intersección, el vehículo que transita por el camino de la derecha tiene la preferencia de paso. En la Fig. 10, se muestra el diagrama de flujo que explica el algoritmo utilizado para la intersección.

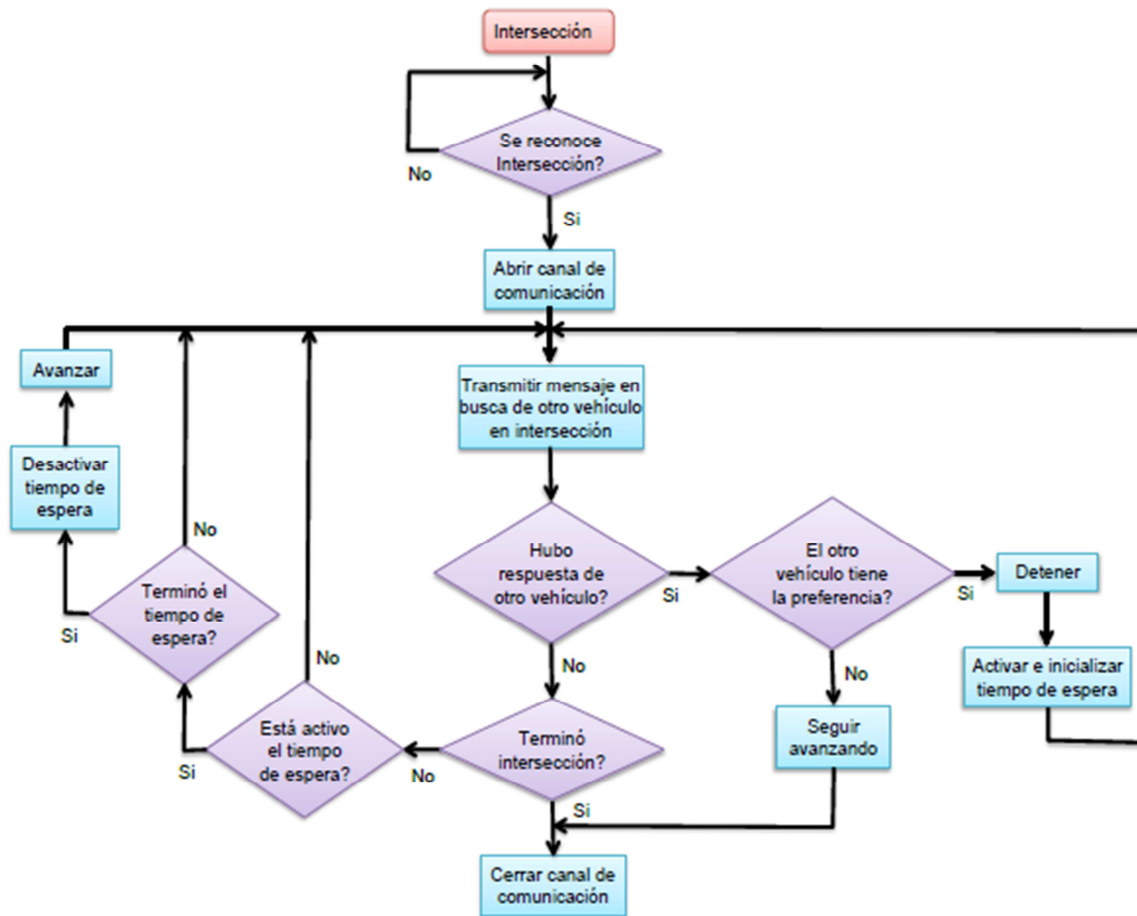


Fig. 10. Diagrama de flujo del algoritmo de intersección.

2.4. Protocolo de separación

Como ya se mencionó, en todo momento los vehículos se encuentran pendientes de las lecturas que reciben de los sensores de obstáculos, por lo tanto, cuando un vehículo A se encuentra siguiendo a un vehículo B y el vehículo B tomó una trayectoria distinta, en el momento en que el vehículo A deja de sensar al vehículo B, el vehículo A inmediatamente deja de obtener la velocidad del vehículo B y comienza a avanzar a la velocidad constante establecida (ver Fig. 11).



Fig. 11. Separación del pelotón.

Cuando se tiene un pelotón de tres vehículos, A, B y C, y el vehículo intermedio B abandona al pelotón, C tiene que avanzar a velocidad constante. En el momento en que C logre alcanzar a A, C debe dejar de avanzar a velocidad constante y comenzar a obtener la velocidad de su sensor de proximidad y formar un nuevo pelotón. En la Fig. 12 se muestra el diagrama de flujo que explica el algoritmo utilizado para la separación.

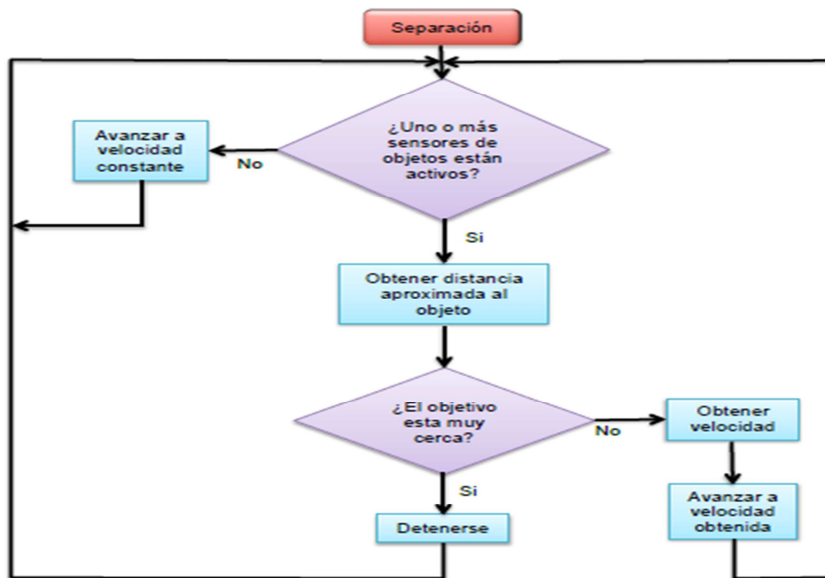


Fig. 12. Diagrama de flujo del algoritmo de separación.

3. Pruebas y Resultados

Cada uno de los sensores que conforman el robot, presentan diferentes alcances y estos también varían de robot a robot, por lo que es importante conocer el alcance de cada uno de ellos.

Cada robot cuenta con dos sensores de línea montados en la parte inferior delantera. Los sensores de línea fueron puestos a prueba en su modo analógico, usando tres colores diferentes; blanco (fondo de la pista), azul (color del camino) y negro (señal de intersección). Los resultados se muestran en la Tabla 2.

| Robot | Blanco | | Azul | | Negro | |
|-------|--------|-------|--------|--------|---------|---------|
| | Izq | Der | Izq | Der | Izq | Der |
| 1 | 13-14 | 14-16 | 15-100 | 17-100 | 101-235 | 101-240 |
| 2 | 13-14 | 14-16 | 15-85 | 17-120 | 86-240 | 121-245 |
| 3 | 12-14 | 12-14 | 15-80 | 15-70 | 81-240 | 71-235 |
| 4 | 11-12 | 12-13 | 13-30 | 14-55 | 31-235 | 56-235 |
| 5 | 15-22 | 13-14 | 23-130 | 15-95 | 131-245 | 96-240 |
| 6 | 13-14 | 15-16 | 15-85 | 17-115 | 86-245 | 116-240 |
| RA | 0-14 | 0-16 | 15-100 | 17-120 | 100-240 | 120-245 |

Tabla 2. Rango numérico de los sensores de línea para los colores Blanco, Azul y Negro; en donde RA representa el rango aproximado de cada sensor.

Sensores de obstáculos

Cada robot cuenta con 4 sensores de obstáculos en la parte frontal, el alcance de cada uno de ellos fue puesto a prueba en su modo analógico a distancias de 2.6 cm y 5.2 cm y en 5 diferentes ángulos, como se muestra en la Fig. 13; los resultados se muestran en las Tablas 3, 4, 5, 6 y 7.

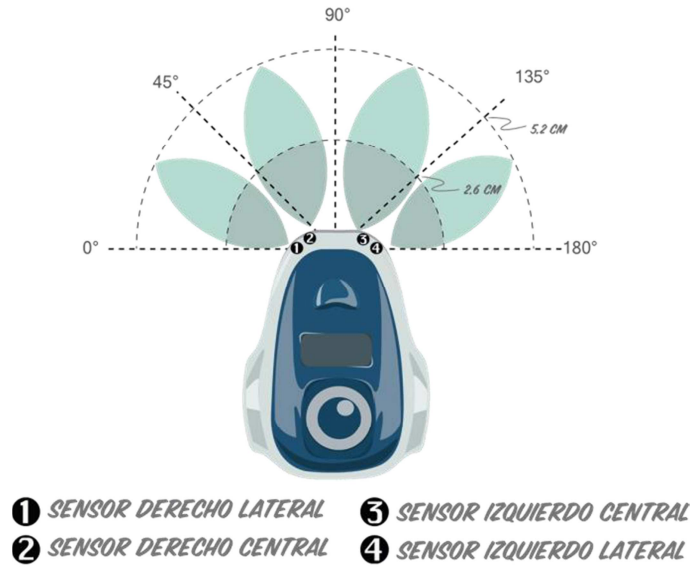


Fig. 13. Gráfica aproximada del alcance de los sensores de objetos.

| Robot | Sensores activos | 2.6 cm. (1/4 de vuelta) | 5.2 cm. (1/2 vuelta) |
|-------|-------------------|----------------------------|-------------------------|
| 1 | Izquierdo lateral | 255 | 50 - 75 |
| 2 | Izquierdo lateral | 195 - 255 | 0 - 5 |
| 3 | Izquierdo lateral | 255 | 0 - 10 |
| 4 | Izquierdo lateral | 255 | 65 - 80 |
| 5 | Izquierdo lateral | 255 | 0 - 20 |
| 6 | Izquierdo lateral | 255 | 25 - 40 |

Tabla 3. Sensado a 0°.

| Robot | Sensores activos | 2.6 cm. (1/4 de vuelta) | 5.2 cm. (1/2 vuelta) |
|-------|-------------------|----------------------------|-------------------------|
| 1 | Izquierdo lateral | 255 | 120 - 130 |
| | Izquierdo central | 255 | 45 - 60 |
| 2 | Izquierdo lateral | 255 | 65 - 75 |
| | Izquierdo central | 215 - 230 | 0 - 1 |
| 3 | Izquierdo lateral | 255 | 95 - 102 |
| | Izquierdo central | 215 - 230 | 30 - 40 |
| 4 | Izquierdo lateral | 255 | 95 - 110 |
| | Izquierdo central | 255 | 45 - 55 |
| 5 | Izquierdo lateral | 255 | 255 |
| | Izquierdo central | 255 | 185 - 205 |
| 6 | Izquierdo lateral | 255 | 80 - 90 |
| | Izquierdo central | 255 | 20 - 40 |

Tabla 4. Sensado a 45°.

| Robot | Sensores activos | 2.6 cm. (1/4 de vuelta) | 5.2 cm. (1/2 vuelta) |
|-------|-------------------|----------------------------|-------------------------|
| 1 | Izquierdo central | 255 | 115 - 125 |
| | Derecho central | 185 - 200 | 25 - 45 |
| 2 | Izquierdo central | 165 - 175 | 20 - 35 |
| | Derecho central | 225 - 235 | 80 - 90 |
| 3 | Izquierdo central | 150 - 155 | 45 - 50 |
| | Derecho central | 175 - 185 | 35 - 40 |
| 4 | Izquierdo central | 205 - 15 | 60 - 70 |
| | Derecho central | 255 | 130 - 145 |
| 5 | Izquierdo central | 255 | 100 - 115 |
| | Derecho central | 200 - 225 | 25 - 45 |
| 6 | Izquierdo central | 255 | 105 - 125 |
| | Derecho central | 205 - 230 | 30 - 55 |

Tabla 5. Sensado a 90°.

| Robot | Sensores activos | 2.6 cm. (1/4 de vuelta) | 5.2 cm. (1/2 vuelta) |
|-------|------------------|----------------------------|-------------------------|
| 1 | Derecho central | 255 | 60 - 70 |
| | Derecho lateral | 170 - 190 | 0 - 10 |
| 2 | Derecho central | 245 - 255 | 45 - 55 |
| | Derecho lateral | 225 | 20 - 30 |
| 3 | Derecho central | 230 - 250 | 50 - 70 |
| | Derecho lateral | 255 | 45 - 70 |
| 4 | Derecho central | 255 | 65 - 80 |
| | Derecho lateral | 255 | 65 - 80 |
| 5 | Derecho central | 255 | 20 - 40 |
| | Derecho lateral | 255 | 50 - 70 |
| 6 | Derecho central | 255 | 55 - 70 |
| | Derecho lateral | 255 | 95 - 110 |

Tabla 6. Sensado a 135°.

| Robot | Sensores activos | 2.6 cm. (1/4 de vuelta) | 5.2 cm. (1/2 vuelta) |
|-------|------------------|----------------------------|-------------------------|
| 1 | Derecho lateral | 200 - 255 | 55 - 65 |
| 2 | Derecho lateral | 255 | 115 - 135 |
| 3 | Derecho lateral | 255 | 125 - 140 |
| 4 | Derecho lateral | 255 | 70 - 90 |
| 5 | Derecho lateral | 255 | 30 - 65 |
| 6 | Derecho lateral | 255 | 45 - 80 |

Tabla 7. Sensado a 180°.

4. Discusión

En un futuro, los vehículos autónomos estarán circulando en las zonas urbanas de las principales ciudades en el mundo, por lo que las intersecciones viales y los cruces de caminos se convertirán en un segmento importante del transporte urbano. Las tecnologías actuales tales como los semáforos y las señales de alto, no son adecuadas para la conducción autónoma, ya que no son muy seguras en la gestión de tráfico y generan retrasos de tiempo debido a su ineficiencia. Se tendrán que desarrollar sistemas de control de tráfico más eficientes y seguros, implementados en los mismos vehículos para no hacer uso de semáforos en las intersecciones viales; sistemas en donde los vehículos sean los que controlen el tráfico con la ayuda de las comunicaciones vehículo a infraestructura y vehículo a vehículo. La prueba de estos sistemas requerirá de infraestructura costosa a gran escala, con medidas de seguridad importantes, que garanticen desarrollos confiables, seguros y con buen desempeño bajo cualquier ambiente; o simplemente tendrán que ser provados en los lugares de interés, lo que podría provocar grandes riesgos.

En este trabajo se presenta el diseño de un sistema de desarrollo económico y seguro, para la emulación de tráfico en intersecciones viales para conducción autónoma; el objetivo es proporcionar una herramienta de apoyo, para la prueba de protocolos de intersección de nueva generación, que puedan gestionar el tráfico a través de intersecciones viales y mejorar significativamente tanto el rendimiento como la seguridad.

5. Conclusiones

El sistema de desarrollo presentado, es adecuado para realizar la simulación en tiempo real del comportamiento de un pelotón vehicular en zonas de intersección vial. Es una alternativa que permite la realización de diferentes sesiones de prueba y simulaciones más extensas y exhaustivas para perfeccionar algoritmos de: control, comunicación y

toma de decisiones entre vehículos, sin provocar daño alguno por la falla en el sensado o por errores en el procesamiento de la información.

Los protocolos de intersección y separación tienen un desempeño aceptable ya que en lugar de asumir que el líder del pelotón monitorea globalmente el entorno y toma las decisiones, sólo los vehículos implicados en una maniobra particular están relacionados, distribuyendo las decisiones locales entre el pelotón.

Los protocolos de intersección y separación como enfoques de coordinación, demuestran su viabilidad con los resultados obtenidos. Estos revelan el desarrollo de un sistema robusto que involucra más de una modalidad de sensado y comunicación, garantizando los procedimientos de seguridad de posibles fallos en los diversos problemas que puedan producirse en el pelotón.

Como trabajo futuro, se desarrollará la programación y el código adecuado para cada uno de los robots, para corregir los problemas de similitud presentados por los sensores de obstáculos, los sensores de línea y los motores; lo cual requiere de un análisis más extenso de las características así como los alcances de cada robot.

6. Referencias

- [1] The DARPA urban challenge, 2007. <http://archive.darpa.mil/grandchallenge>.
- [2] M. Buehler, K. Iagnemma, S. Singh, *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*. 2009. Springer-Verlag. Berlin Heidelberg. 1-59 pp.
- [3] C. Park, J. Lee, "Intelligent Traffic Control Based on IEEE 802.11 DCF/PCF Mechanisms at Intersections". *IEEE Vehicular Technology Conference*. 5-8 September, 2011. 1-4 pp.

- [4] H. Kowshik, D. Caveney, P. R. Kumar, "Provable Systemwide Safety in Intelligent Intersections". IEEE Transactions on Vehicular Technology. Vol. 60. No. 3. March 2011. 804-818 pp.
- [5] Y. Senda, I. Tanev, K. Shimohara, "On the Possibility of Priority-Based Road Traffic Control". IEEE SICE Annual Conference. 20-22 August, 2008. 1510-1513 pp.
- [6] R. Azimi, G. Bhatia, R. Rajkumar, P. Mudalige, "STIP: Spatio-Temporal Intersection Protocols for Autonomous Vehicles". IEEE International Conference on Cyber-Physical Systems. 14-17 April, 2014. 1-12 pp.
- [7] SARTRE Project website. www.sartre-project.eu.
- [8] INVENT Project website www.invent-online.de.
- [9] R. Rajamani, S. E. Shladover, "An Experimental Comparative Study of Autonomous and Co-operative Vehicle-Follower Control Systems", Transportation Research Part C: Emerging Technologies, Volume 9, Issue 1, February 2001, Pages 15-31.
- [10] H. G. Tanner, G. J. Pappas, V. Kumar, "Leader-to-Formation Stability". IEEE Transactions on Robotics and Automation. Vol. 20. No. 3. 2004. 443-455 pp.
- [11] Tutorial 2, "Introduction to mOway robot Electronics". www.moway-robot.com.

7. Autores

Dr. Andrés Ferreyra Ramírez es Ingeniero Mecánico Electricista con Especialidad en Electrónica, por la Universidad Nacional Autónoma de México. Maestro en Ingeniería Biomédica, por la Universidad Autónoma Metropolitana, unidad Iztapalapa. Doctorado en Ciencias en Control Automático, por el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional CINVESTAV-IPN. Desde 1996 se desempeña como Profesor Investigador Titular "C" de tiempo completo en el Departamento de Electrónica de la UAM-Azcapotzalco, donde desarrolla trabajos de

Investigación en las áreas de Softcomputing, Machine Learning y Sistemas de Transporte Inteligentes.

M. en C. Arturo Zúñiga López es Profesor Investigador del Departamento de Electrónica, en el Área de Instrumentación, de la UAM Azcapotzalco. Actualmente realiza investigación en el Área de Redes de Computadoras, Sistemas Embebidos y Procesamiento Digital de Imágenes.

M. en C. Roberto A. Alcántara Ramírez es Ingeniero en Electrónica, por la Universidad Autónoma Metropolitana. Maestro en Ciencias e Ingeniería Mecatrónica, por el T.E.S.E. Desde 1987 se desempeña como Profesor Investigador de tiempo completo en el Departamento de Electrónica de la UAM-Azcapotzalco, donde desarrolla trabajos de Investigación en las áreas de Electrónica de Potencia y Control de Procesos.

Renderizado háptico para el Novint Falcon

Enrique Linares Ramírez

Facultad de Ingeniería Electromecánica, Universidad de Colima, km 20.5
carretera Manzanillo-Cihuatlán, Teléfono: 01 31 43 31 12 07
elinares@ucol.mx

Jorge Gudiño Lau

Facultad de Ingeniería Electromecánica, Universidad de Colima, km 20.5
carretera Manzanillo-Cihuatlán, Teléfono: 01 31 43 31 12 07
jglau@ucol.mx

Saida Miriam Charre Ibarra

Facultad de Ingeniería Electromecánica, Universidad de Colima, km 20.5
carretera Manzanillo-Cihuatlán, Teléfono: 01 31 43 31 12 07
scharre@ucol.mx

Miguel Duran Fonseca

Facultad de Ingeniería Electromecánica, Universidad de Colima, km 20.5
carretera Manzanillo-Cihuatlán, Teléfono: 01 31 43 31 12 07
mduran@ucol.mx

Resumen

Este trabajo presenta el diseño de una interfaz háptica empleando un dispositivo háptico (Novint Falcon) para fines de diagnóstico de rehabilitación. Se diseñó un algoritmo de computadora que envía al dispositivo háptico un vector de fuerza para que sean transmitidos al usuario, esto se logra al estimular el sentido del tacto del usuario y para que vea real se diseñó un renderizado háptico para estimular la vista. Se empleó la ley de Hook para un resorte lineal fuerza para calcular la fuerza. Por último se muestran los resultados experimentales.

Palabra(s) Clave(s): háptica, renderizado, virtual.

1. Introducción

Los algoritmos de renderizado háptico permiten que el usuario interactúe con un entorno virtual y pueda percibir de manera “realista” los objetos. Esta representación se logra por medio de algoritmos de computadora que envían al dispositivo háptico los vectores de fuerza necesarios para que sean transmitidos al usuario, esto se logra al estimular el sentido del tacto del usuario, más específicamente, el sistema háptico.

El sistema háptico presenta un canal bidireccional de comunicación entre el entorno y el cerebro [5], permitiendo de esta forma no solo transmitir un estímulo al tocar a otra persona por ejemplo, si no al mismo tiempo percibir un estímulo como puede ser la textura diferentes superficies, la dureza de la superficie, la deformación de la superficie, entre otros.

Con base en el carácter multidisciplinario de la háptica se ha creado la siguiente subdivisión [5]:

- la háptica humana, es el estudio de la percepción y manipulación de objetos a través del tacto,
- la háptica de máquina, se refiere al diseño, construcción y uso de las máquinas para reemplazar o mejorar el sentido del tacto, y
- la háptica de computadora, son un conjunto de algoritmos y software asociado con la generación del contacto y percepción de objetos virtuales.

Un dispositivo háptico muy utilizado por su bajo costo es el Novint Falcon, fue lanzado al mercado en el año 2007 como el primer dispositivo de consumo con retroalimentación de fuerzas en tres dimensiones (3D) destinado para el mundo del entretenimiento y los videojuegos [3]. Este dispositivo tiene una configuración de robot delta y su efector final tiene un área de trabajo de 10cmx10cmx10cm, aunque se han

presentado trabajos como el de Martin y Hiller que reportan que el espacio de trabajo no es uniforme [3].

Otro dispositivo háptico muy utilizado y que se prefiere más en la investigación es el Phantom Omni, el cual cuenta con seis grados de libertad. Este cuenta con un efector final en forma de pluma lo que permite una manipulación más natural [2].

Los dispositivos hápticos pueden ser encontrados en prácticamente cualquier área de investigación, tal es el caso de la medicina, en la que se han desarrollado dispositivos hápticos como apoyo a los cirujanos. Un ejemplo de este tipo de tecnología se encuentra en Da Vinci [1].

En general, un dispositivo háptico se conecta a una computadora por medio de un puerto (en el caso particular del Falcon es por el puerto USB), envía un vector posición a la computadora a través de sus encoders ópticos y el dispositivo recibe un vector de fuerza que es enviado por la computadora, es decir el dispositivo háptico tiene como señal de entrada un vector de fuerza (f_x , f_y y f_z) y como señal de salida un vector de posición cartesiano (x , y y z), por lo que internamente el dispositivo háptico tiene la cinemática inversa y no es proporcionada por el fabricante, mediante un algoritmo de control háptico permite al usuario esa sensación de estar tocando un objeto real [3].

2. Desarrollo

Los algoritmos de computadora responsables de procesar la información procedente del dispositivo háptico y enviar la información de fuerzas deben asegurar la sincronización entre la representación visual del entorno (gráficas por computadora) y la representación háptica para garantizar que los estímulos transmitidos al usuario se perciban lo más real posibles.

Para completar la interacción se tienen algoritmos de computadora que generan el entorno virtual y reaccionan a las interacciones entre el usuario (a través del dispositivo háptico) y el entorno.

La Fig. 1 presenta un esquema general del algoritmo de renderizado (es una “representación” de un objeto, en el texto se utilizarán indistintamente los términos renderizar y representar) en el que se puede observar el funcionamiento básico de representación objetos en un entorno virtual con un dispositivo de retroalimentación de fuerzas.

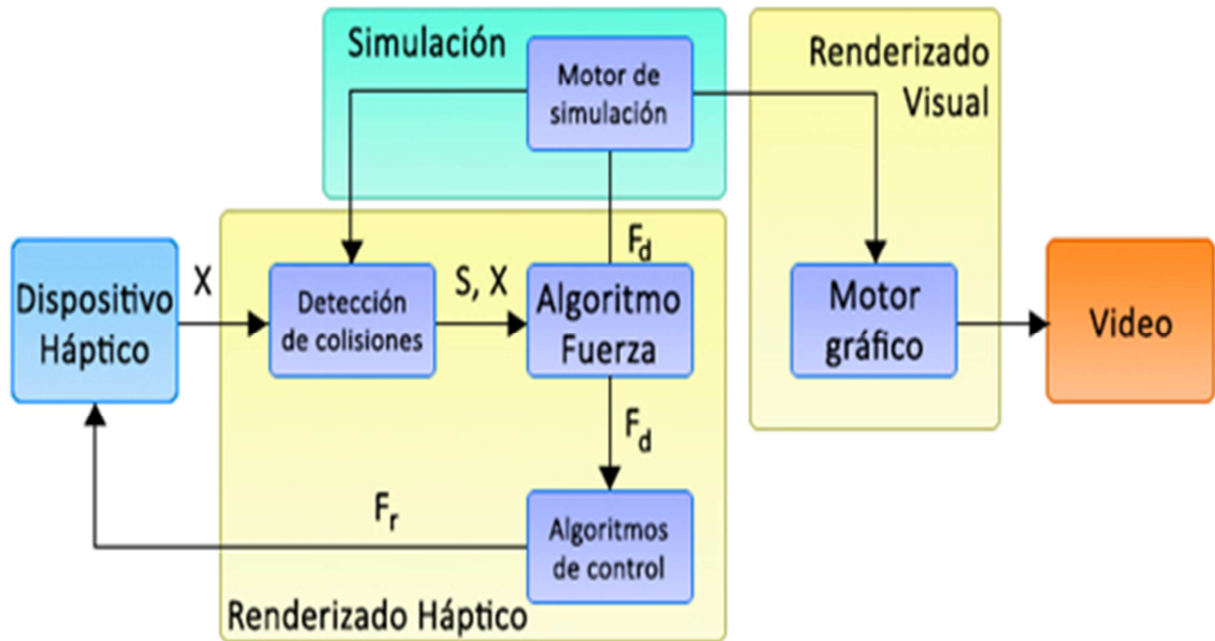


Fig. 1. Bloques funcionales de un algoritmo de renderizado háptico [4].

El funcionamiento completo de este tipo de sistemas inicia con la obtención de la posición del efector final del dispositivo háptico (algoritmo de control de bajo nivel), esta información es enviada a la computadora para su procesamiento por el entorno virtual el cual calcula los efectos que puede tener el movimiento y en caso de presentarse alguna colisión con un objeto calcular las fuerzas resultantes (algoritmo de detección de colisiones) y enviarlas de regreso al dispositivo háptico para que las transmita al usuario (algoritmo de fuerza) [4]. A este proceso se le llama ciclo háptico ya que es ejecutado constantemente aproximadamente 1000 veces por segundo (1Khz), de otro modo la interacción con la superficie no se sentirá real (más suave o solo vibración) [5].

Las interacciones para el renderizado háptico con retroalimentación de fuerzas se clasifican basados en la forma en la que se modela el avatar (representación virtual de la interfaz háptica por medio de la cual el usuario interactúa con el entorno virtual [2], en términos de háptica se suele llamar a esta representación como Haptic Interface Point (HIP por sus siglas en inglés) por lo que ambos términos se utilizarán indistintamente) del dispositivo háptico [6]:

- interacción basada en un punto, es cuando el avatar se modela como un punto como si se explorara y manipulara el entorno a través de la punta de una herramienta,
- interacción basada en rayos, se modela el avatar como uno o más segmentos de línea, e
- interacción basada en un objeto 3D, es cuando el avatar está formado por un grupo de puntos, segmentos de línea y/o polígonos.

El método utilizado para la interacción depende de las necesidades y complejidades de la aplicación [5].

Debido a las características de los entornos virtuales y a que los dispositivos hápticos solo pueden representar una cantidad de fuerza limitada el dispositivo háptico siempre “atravesará” los objetos virtuales no importando si son rígidos o no [4].

Para la interacción basada en un punto, al momento de detectar una colisión, se verifica si el HIP está dentro del objeto, en caso de ser así, se calcula la distancia entre el HIP Haptic Interface Point por sus siglas en inglés) y el punto más cercano a este sobre la superficie del objeto, esta distancia permite calcular la fuerza resultante que será aplicada al usuario [6].

La fuerza se calcula utilizando las leyes de Hook para un resorte lineal [5]:

$$F = Kx$$

donde K es el coeficiente del resorte, x es la profundidad de penetración y F es la fuerza de reacción (para interacciones sin fricción) que es normal a la cara poligonal con la que colisiona el HIP (ver fig. 2).

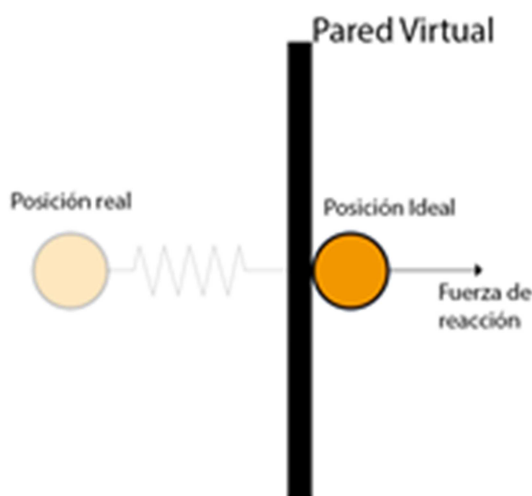


Fig. 2. Representación háptica de la pared virtual.

Para objetos rígidos el valor de K se establece al valor más alto posible limitado por las inestabilidades del dispositivo al contacto. Según algunos estudios el agregar un amortiguador a la dinámica de la interacción se mejora la estabilidad del sistema y la precepción háptica de la rigidez.

Los algoritmos de renderizado de fuerzas se clasifican por el número de grados de libertad necesarios para describir la fuerza de interacción a representar [4]:

- Interacción de un grado de libertad, se utiliza sólo una dimensión espacial, es el equivalente a crear una pared virtual con, las fuerzas de interacción serían las equivalentes a entrar en contacto con un objeto infinitamente rígido. Este caso

se representa mediante un resorte atado a una pared (ver fig. 2), un ejemplo es abrir una puerta al girando la manija [4].

El algoritmo más simple para renderizar una pared virtual viene dado por

$$F = \begin{cases} 0, & x > x_w \\ k(x_w - x), & x \leq x_w \end{cases}$$

donde F = Fuerza de reacción

K = dureza de la pared (idealmente es un valor muy grande)

X = posición del HIP

X_w = posición de la pared

- Interacción de dos grados de libertad, es el desplazamiento se tiene en dos dimensiones espaciales, como el movimiento de un mouse de computadora [4].
- Interacción de tres grados de libertad, se asume que se interactúa con el mundo virtual a través de un HIP y se requiere el cálculo de las 3 componentes de fuerza. Un algoritmo para este tipo de interacción es el proxy [4].
- Se tienen además clasificaciones para seis grados de libertad.

Por otra parte, el dispositivo se mueve por el entorno sin que exista colisión alguna, el HIP y el proxy se visualizan en la misma posición y eso hace que el usuario no perciba ninguna fuerza que restrinja su movimiento. Al momento de detectarse una colisión, el proxy se mantiene visible fuera del objeto (haciendo contacto con la superficie) en la posición más cercana al HIP. [4]. Relación entre las posiciones del HIP y del proxy se observa en la Fig. 3. En la Fig. 3a se representa la visualización en el espacio libre de contacto, el proxy y el HIP se sobreponen. La Fig. 3b muestra lo que sucede al presentarse una colisión, el proxy y el HIP se separan (el HIP no es visible). En la Fig. 3c se muestra el proxy sobre la superficie del objeto en la posición más cercana a la posición real del HIP [7]

La Fig. 3 muestra la relación entre las posiciones del proxy y el HIP.

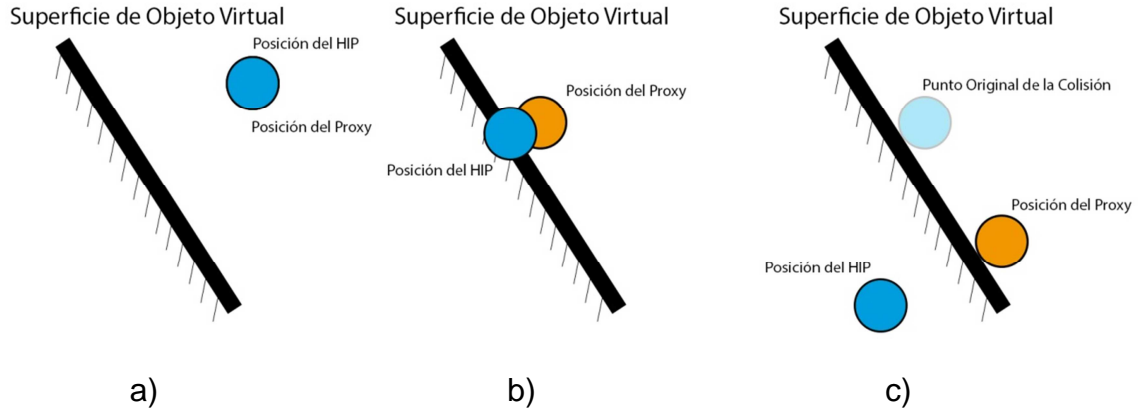


Fig. 3. Relación entre las posiciones del HIP y del proxy.

3. Resultados

El presente trabajo utiliza los conceptos mostrados en la sección 2, para el diseño de una aplicación de seguimiento de trayectoria. Dicha trayectoria es guiada por algoritmos de renderizado háptico para establecer una guía que le indique al usuario cual es el camino a seguir sin que esto evite el movimiento libre por el entorno. Para la creación del prototipo se utiliza Matlab por las facilidades que brinda para la graficación y análisis de la información recopilada, también permite un rápido diseño y desarrollo del entorno en 3 dimensiones (3D).

La Fig. 4 muestra la interfáz del prototipo, en ella se muestra una libreta virtual sobre una base de madera. En la libreta se muestra un círculo, el cual representa la trayectoria a seguir. Como proxy se utiliza una esfera pequeña.

Al ejecutar la simulación se activa la retroalimentación de fuerzas del dispositivo háptico. Mientras el usuario siga la trayectoria deseada (círculo) no se impone ninguna restricción al usuario, lo que significa que no se representan fuerzas que limiten el

movimiento. Si el usuario se desvía de su trayectoria se envía un vector de fuerzas al dispositivo con el esquema de resorte explicado anteriormente para indicar al usuario que la trayectoria se está desviando de la deseada, entre más se desvíe mayor será la fuerza representada en el dispositivo.

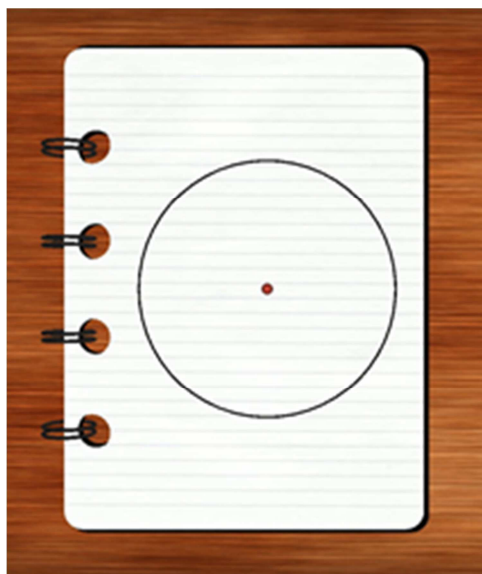


Fig. 4. Interfáz del prototipo.

También se permite el movimiento libre al “levantar” el proxy del papel se permite un comportamiento completamente libre, lo que representa moverse en el aire.

La interfaz fue creada en el mismo Matlab con la herramienta 3d Simulink Animation. El círculo que se muestra sobre la libreta sirve de referencia para el movimiento que se debe realizar por el usuario. Para un funcionamiento mas realista se dividió el círculo en cuadrantes y la posición del proxy cada uno se evalúa por separado.

La adaptación que se hace a este prototipo del algoritmo de renderizado de una pared háptica es tomar la circunferencia como si fuera la pared en cuestión, si el usuario se desvía de la ruta predefinida se genera una fuerza a partir del punto en el que se separó del círculo representando internamente en ese punto un resorte que sirve para calcular la fuerza a aplicar al dispositivo.

En cada movimiento del proxy se calcula una desviación del ideal (circunferencia) y esa desviación es utilizada en (1) para el cálculo de la fuerza.

La Fig. 5 muestra un ejemplo de seguimiento de la trayectoria circular por el usuario, se puede notar que no es circular completamente lo cual se hizo a propósito para representar las fuerzas que restringen en movimiento. En la Fig. 6 se puede observar que a mayor distancia de la trayectoria deseada se aplica más fuerza al dispositivo para indicar al usuario que se ha salido de la ruta preestablecida.

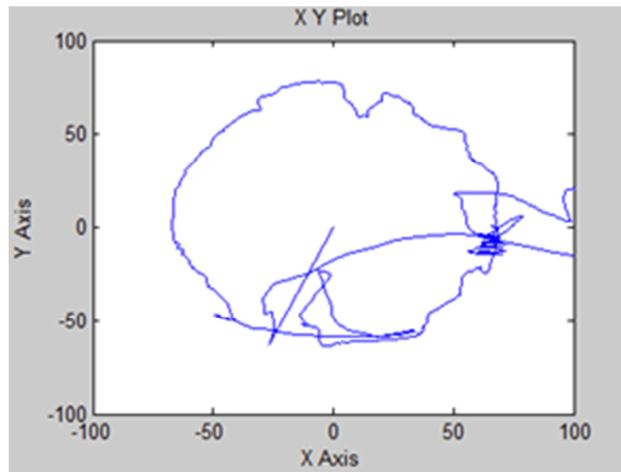


Fig. 5. Gráficas en el plano cartesiano del seguimiento.

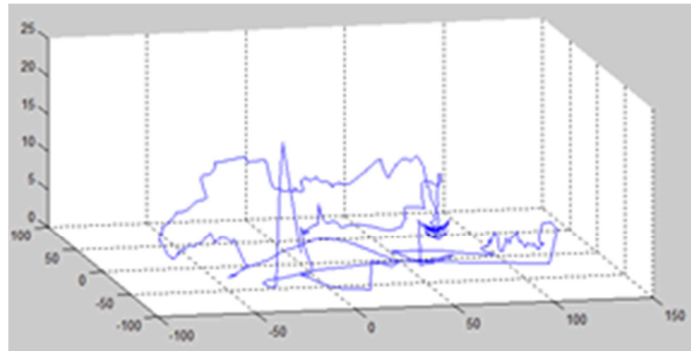


Fig. 6. Gráficas de la fuerza resultante durante el trayecto.

4. Discusión

Este artículo muestra los primeros pasos de investigación, referente al diagnóstico y evaluación de extremidades superiores (manos) de pacientes que han perdido capacidad sensorial y motriz, con el propósito de mejorar la salud humana empleando la tecnología de interfaces hápticas, que sirve para la simulación de entornos, sensaciones mediante realidad virtual y tecnología del tacto, desarrollo de entornos y ambientes inteligentes. El propósito de esta investigación, es hacer un interfaz háptica en busca de mejorar la calidad de vida los pacientes que tienen problemas en extremidades superiores, más específicamente niños. El seguimiento de trayectorias presenta una excelente vía para el diagnóstico y rehabilitación de pacientes ya que permite evaluar el grado de daño en la articulación evaluada.

5. Conclusiones

Se dieron los primeros pasos en el área de las interfaces hápticas, se diseñó un entorno virtual con el objetivo de diagnosticar y evaluar pacientes con algún daño en extremidades superiores, en manos específicamente. La información obtenida proporciona al especialista un panorama más preciso del grado de lesión evaluada, ya que en general la evaluación de pacientes se realiza por inspección sin ningún instrumento de precisión que proporcione información exacta. Se obtuvieron resultados satisfactorios al aplicarlo a un grupo de personas sanas para evaluar el desempeño normal al realizar este tipo de actividad, sin embargo todavía no se aplica a pacientes con problemas motores en el miembro superior, el siguiente paso es realizarlo con pacientes que presenten diferentes grados de lesión en la muñeca. A partir de este trabajo se puede deducir que la realidad virtual y la retroalimentación de fuerzas son muy útiles al momento de realizar seguimiento de trayectorias dirigidas, ya que ofrece un mejor control sobre la ruta que se plantea seguir, y se puede diagnosticar y evaluar a un paciente; ya que el algoritmo de control de fuerza utilizado es bastante sencillo permite representar fuerzas de una forma aceptable.

6. Referencias

- [1] El Saddik, Abdulmotaleb, Orozco, Mauricio, Eid, Mohamad; et al., "Haptics Technologies. Bringing Touch to Multimedia". 2011. Berlin, Heidelberg: Springer.
- [2] Kawai, Satoru, Faust Jr., H. Paul, L. Christine MacKenzie, Computer Graphic and PHANToM Haptic Displays: Powerful Tools in Understand How Human Perceive Heaviness. 2012. En: El Saddik, Abdulmotaleb ed., Haptics Rendering and Applications. Rijeka, Croatia: InTech. 25-46 pp.
- [3] J. L. Rodríguez, R. Velázquez, "Haptic Rendering of Virtual Shapes with the Novint Falcon". *Procedia Technology*. Vol. 3. 132-138 pp.
- [4] K. Salisbury, F. Conti, Francois, "Haptic Rendering: Introductory Concepts". *Computer Graphics and Applications*, IEEE. Vol. 2. No. 24. 2004. 24-32 pp.
- [5] What is Haptics?. <http://touchlab.mit.edu>. Agosto 2014.
- [6] M. A. Srinivasan. C. Basdogan, "Haptics in virtual environments: Taxonomy, research status, and challenges". *Comput. & Graphics*. Vol. 4. No. 21. 1997. 393-404 pp.
- [7] F. Rydén. H. Chizeck, Howard, "A Proxy Method for Real-Time 3-DOF Haptic Rendering of Streaming Point Cloud Data". *IEEE Transactions on Haptics*. Vol. 6. No. 3. 2013. 257-267 pp.

7. Autores

Dr. Jorge Gudiño Lau es Doctor en Ingeniería Eléctrica en el área de control con especialidad en Robótica por la Universidad Nacional Autónoma de México en el 2005. Actualmente es Profesor-Investigador de Tiempo Completo y Líder del Cuerpo Académico UCOL-CA 21 Sistemas Mecatrónicos y Eléctricos de la Facultad de Ingeniería Electromecánica de la Universidad de Colima.

Ing. Enrique Linares Ramírez es Ingeniero en Sistemas Computacionales por la Universidad de Colima. Laboró como desarrollador en la empresa Sistematización

donde desarrollo software en el lenguaje de programación Clipper, posteriormente se desempeñó como Administrador de Centro de Cómputo de un plantel de la Universidad de Colima. Actualmente es catedrático de la Facultad de Ingeniería Electromecánica de la Universidad de Colima impartiendo las materias Algorítmica, Estructura de Datos y Teoría de Autómatas, también cursa los estudios de posgrado en Maestría en Ingeniería en la misma facultad.

Dr. Miguel Ángel Duran Fonseca Doctor en Ingeniería Eléctrica en el Centro Nacional de Investigación y Desarrollo Tecnológico CENIDET en el 2010. Actualmente es Profesor-Investigador de Tiempo Completo, Coordinador Académico e integrante del Cuerpo Académico UCOL-CA 21 Sistemas Mecatrónicos y Eléctricos de la Facultad de Ingeniería Electromecánica de la Universidad de Colima.

M.I Saida Miriam Charre Ibarra es Maestra en Ingeniería en el área de automatización y control por el Instituto Tecnológico de Ciudad Guzmán en el 2004. Actualmente es Profesor-Investigador de Tiempo Completo y pertenece al Cuerpo Académico UCOL-CA 21 Sistemas Mecatrónicos y Eléctricos de la Facultad de Ingeniería Electromecánica de la Universidad de Colima.

Caracterización de los sonidos implicados en el ciclo cardíaco mediante un fonocardiograma

Braulio Cruz Jiménez

Universidad Autónoma de Yucatán – Facultad de Ingeniería, Av. Industrias no contaminantes por periférico norte, Teléfono: (999)-9300550
braulio.cruz@correo.uady.mx

Jannette Contreras Rivero

Universidad Autónoma de Yucatán – Facultad de Ingeniería, Av. Industrias no contaminantes por periférico norte, Teléfono: (999)-9300550

Luis Ricalde Castellanos

Universidad Autónoma de Yucatán – Facultad de Ingeniería, Av. Industrias no contaminantes por periférico norte, Teléfono: (999)-9300550

Resumen

En este trabajo se desarrollaron dos métodos para la extracción automática de las características temporales y espectrales de la señal acústica cardíaca basados en la transformada de Fourier de tiempo corto y en la transformada Wavelet. En ambos algoritmos se redujo de manera sustancial el ruido en la señal fonocardiográfica debido a la respiración y a los movimientos del paciente, mientras que la evaluación del ciclo cardíaco, ubicación del primer tono cardíaco y segmentación del fonocardiograma se realiza sin la necesidad de utilizar información del electrocardiograma.

Palabras clave: auscultación, ciclo cardíaco, fonocardiograma, segmentación.

1. Introducción

Muchos de los padecimientos del sistema cardiovascular causan murmullos y aberraciones en los sonidos cardiacos mucho antes de que se presenten otros síntomas, por ejemplo un cambio en la señal del electrocardiograma (ECG). Aunque el ECG y el examen ultrasónico (ecocardiografía) son ampliamente usados en el diagnóstico cardiaco, una técnica empleada tradicionalmente por los médicos, aún antes de proceder a estudios más sofisticados, es la auscultación. La auscultación es una técnica no invasiva que permite identificar los distintos sonidos que causan las arritmias por medio de un estetoscopio. Uno de los problemas más importantes de la auscultación es encontrar un sitio apropiado en la pared del pecho donde se puedan escuchar con mayor claridad los componentes de los sonidos cardiacos, además de esto el médico debe desarrollar cierta habilidad para detectar los sonidos anómalos. Lo anterior hace que esta prueba sea subjetiva, debilitando su poder como una herramienta de diagnóstico.

Existe actualmente una forma moderna de la auscultación, llamada fonocardiografía, en la cual se obtiene un registro gráfico de los sonidos cardiacos (fonocardiograma). El fonocardiograma consiste en la representación gráfica de las ondas sonoras del latido cardiaco, permite documentar la temporización, intensidad relativa, frecuencia, calidad, tono, timbre y localización precisa de los diferentes componentes de los sonidos cardiacos de forma objetiva y repetible. Es posible transformar los sonidos cardiacos de una vibración mecánica a una señal eléctrica mediante un micrófono; la amplitud de la señal eléctrica registrada de esta manera varía según las tensiones ejercidas por las ondas sonoras y puede amplificarse y reproducirse para su análisis. El dispositivo que realiza este procedimiento recibe el nombre de estetoscopio electrónico [1].

Los estetoscopios electrónicos disponibles en el mercado poseen una calidad de sonido y amplificación aceptable, al mismo tiempo que se ha reducido la interferencia por ruidos externos; sin embargo, una característica de la cual carecen estos dispositivos es

una interfaz entre el estetoscopio electrónico y una computadora para el procesamiento de los sonidos cardiacos, de modo tal que un correcto análisis matemático de los mismos permita la detección temprana no invasiva de fallas cardiacas [2].

Los avances en los conocimientos sobre la naturaleza de los sonidos y soplos cardiacos, y su significado clínico, han sido resultado del desarrollo de tecnologías como el cateterismo cardiaco, los estudios angiográficos, la fonocardiografía externa e intracardiaca y las correlaciones fonocardiográficas-ecocardiográficas [3].

El campo de la fonocardiografía ha experimentado un impulso debido a los potentes medios disponibles actualmente para la adquisición y procesamiento de señales de audio en tiempo real. En [4] desarrolló un sistema para la grabación, procesamiento y análisis de los sonidos cardiacos en un espectro de frecuencia de 20 Hz a 22 kHz para detectar los murmullos en niños [4].

La mayoría de los métodos existentes de detección de las señales acústicas del corazón aplican filtros digitales pasa bandas como método de cancelación de ruido. Aunque los filtros pasa banda eliminan una parte relevante del ruido, también afecta a la señal pues se mezcla con la frecuencia de la banda filtrada y aún puede contener ruido. La técnica empleada en este artículo permite obtener una mayor cancelación del ruido y puede ser usada como una etapa de preprocesado en los sistemas fonocardiográficos [5].

2. Desarrollo

El fonocardiograma (FCG) es la representación gráfica de los componentes acústicos de los latidos cardiacos. El ciclo cardiaco normal tiene dos sonidos principales, S1 y S2, que delimitan las dos fases del ciclo cardiaco: sístole (entre S1 y S2) y diástole (entre S2 y S1). En la Fig. 1 se muestran los diferentes sonidos componentes del ciclo cardiaco. La intensidad de cada sonido depende de la zona de auscultación. Se pueden distinguir cuatro tonos cardiacos cuyas características se mencionan a continuación [6]:

a. Primer tono cardíaco (S1). Al contraerse los ventrículos, se escucha un sonido de tono bajo y relativamente prolongado producido por el cierre de las válvulas AV. Esta vibración se conoce como primer tono cardíaco (S1), tiene una duración de 0.14 segundos y dos componentes: M1 (mitral) y T1 (tricúspide).

b. Segundo tono cardíaco (S2). Al final de la sístole se escucha un sonido relativamente rápido originado por el cierre de las válvulas semilunares. Este sonido se conoce como segundo tono cardíaco (S2), dura cerca de 0.11 segundos y tiene un componente auricular (A2) y un componente pulmonar (P2). La razón de la mayor brevedad de S2 respecto de S1 es que las válvulas semilunares son más rígidas que las válvulas AV, de forma que vibran durante un período más corto.

c. Tercer tono cardíaco (S3). Al inicio del tercer cuarto de la diástole es posible percibir un débil tercer tono cardíaco (S3). La frecuencia de este sonido suele ser tan baja que el oído no la percibe.

d. Cuarto tono cardíaco (S4). Su frecuencia es de aproximadamente 20 ciclos por segundo y es inaudible con el estetoscopio.

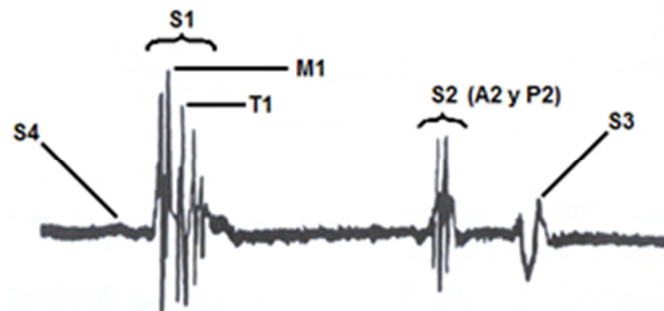


Fig. 1. Representación gráfica de los sonidos cardíacos [6].

En una persona sana no existe ningún sonido entre los intervalos S1-S2 y S2-S3, con la excepción del ruido del ambiente. Sin embargo, en determinadas condiciones es posible encontrar en estos intervalos ruidos de alta frecuencia denominados soplos. Los

soplos son causados por turbulencias del flujo sanguíneo cuando la sangre circula a alta velocidad; esto puede deberse a dos tipos de afecciones: estenosis valvular o insuficiencia cardiaca. En la Fig. 2 se presentan los diferentes componentes del FCG y su relación con los diferentes eventos que ocurren durante el ciclo cardiaco.

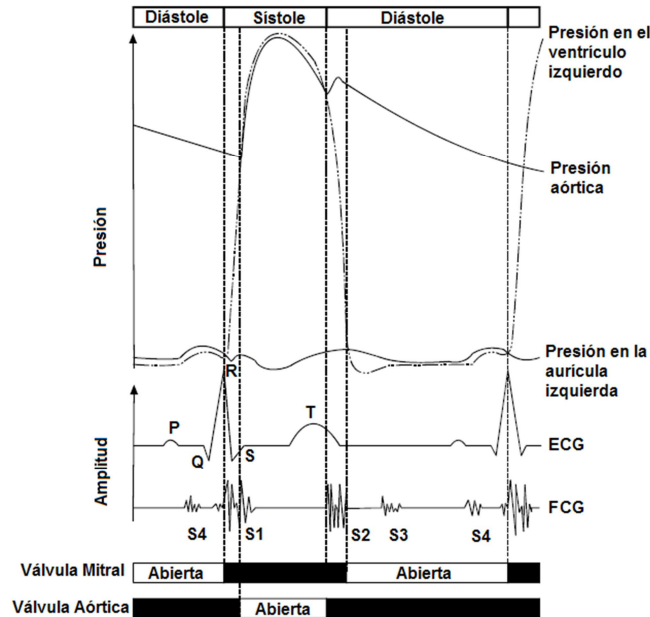


Fig. 2. Correlación de los cuatro sonidos cardiacos con los eventos eléctricos y mecánicos del ciclo cardiaco [6].

Dadas las relaciones que existen entre los sonidos cardiacos y los eventos mecánicos y eléctricos del ciclo cardiaco, el médico es capaz de emitir un diagnóstico referente a alguna cardiopatía valvular mediante el análisis de los diferentes componentes acústicos de un fonocardiograma. Esto realza el valor clínico de la fonocardiografía como método de diagnóstico, especialmente en el caso de enfermedades de las válvulas cardiacas.

2.1. Segmentación

La segmentación es el proceso por el cual la señal del FCG se divide en ciclos cardiacos, identificándose sus componentes acústicos más importantes para el

diagnóstico de cardiopatías. Se da en dos etapas: en la primera se divide al FCG en ciclos cardiacos independientes, comenzando cada uno de ellos con S1 y finalizando justo antes del siguiente S1; la segunda etapa consiste en la identificación de los eventos que forman el ciclo cardiaco, tanto los sonidos como los soplos, identificándolos correctamente y delimitando su comienzo y final. Los métodos de segmentación se basan en dos enfoques diferentes: el primero emplea únicamente la señal del FCG para determinar los eventos del ciclo mientras el segundo utiliza alguna otra señal como referencia (por ejemplo, el electrocardiograma).

Entre las técnicas disponibles para el análisis del FCG se encuentra la transformada de Fourier, la cual determina las componentes espectrales de la señal. Sin embargo, el análisis de la transformada de Fourier se aplica generalmente a señales no estacionarias pues al pasar la señal al dominio de la frecuencia se pierde la información temporal; esto hace que no sea útil en la detección del inicio y terminación de los diferentes eventos del ciclo cardiaco. Para corregir esta deficiencia se ha adaptado la transformada de Fourier para analizar solamente una pequeña sección de la señal, en un tiempo dado; esta técnica se llama transformada de Fourier en tiempo corto (STFT, por sus siglas en inglés) [7].

Otro método apropiado para emplearse en la segmentación del FCG es la transformada wavelet, la cual puede entenderse como una STFT con resolución de ventana variable a medida que varía la escala (frecuencia) de la señal; es decir, para escalas grandes (frecuencias bajas) la resolución temporal es baja y la resolución en frecuencia es alta, mientras que para escalas pequeñas (frecuencias altas) ocurre lo contrario. Así, la transformada wavelet se vislumbra como una herramienta mucho más adecuada para la segmentación del FCG que la STFT [8].

En [9] se realiza la adquisición y el análisis de tiempo-frecuencia de una señal de fonocardiograma. La distribución pseudo-suavizada Wigner-Ville proporcionó un rendimiento adecuado para las representaciones tiempo-frecuencia de estas señales no estacionarias.

2.2. Registro de los sonidos cardíacos

Para la selección de los pacientes se revisaron los expedientes clínicos de las personas hospitalizadas en la clínica Centro Médico Pensiones. Se determinó que pacientes ameritaban valoración de las condiciones cardiacas, incluyendo realización de electrocardiogramas de control, realización de ecocardiogramas o simple auscultación convencional. De 24 pacientes hospitalizados, 12 cumplían con los requisitos para tomar las lecturas; se tomaron 12 grabaciones de fonocardiogramas mediante software personalizado a los pacientes seleccionados. Se realizó en los pacientes seleccionados la auscultación cardiaca convencional y se enunciaron los hallazgos encontrados por el medico mediante la exploración física. Todas las grabaciones tuvieron aproximadamente 16 segundos de duración. De acuerdo a los fundamentos teóricos del ancho de banda de la señal cardiaca (30Hz-600Hz aproximadamente), al Teorema de muestreo y en la cantidad de datos para la graficación en el software de recepción, se eligió una frecuencia de muestreo de 16kHz. Tal frecuencia supera las necesidades mínimas del procesamiento, lo suficiente para evitar las señales alias. El sensor acústico que obtuvo la señal deseada se colocó en el punto del pecho donde la señal cardiaca es más fuerte y corresponde a la zona mitral, el segundo sensor se colocó en la zona pulmonar para servir como señal de referencia para el filtrado adaptativo.

Para la detección de la señal cardiaca se emplean dos micrófonos capacitivos AT-831cW de la marca Audio-Technica. Son de tipo lavalier, lo que los hace pequeños, y tienen un patrón polar del tipo cardioide. El patrón polar indica la sensibilidad del micrófono a los sonidos que llegan de los alrededores a diferentes ángulos respecto de su eje central. La respuesta en frecuencia del micrófono es de 40 Hz a 20 kHz, estas características se ajustan adecuadamente a las de la señal física a monitorear.

Para demostrar la aplicabilidad de la metodología desarrollada, se seleccionó un fonocardiograma típico (FCG3) para el análisis de segmentación. En la Fig. 3 se muestra el FCG3 en el rango de tiempo de 0 a 4.4 segundos; donde se aprecian

claramente los picos correspondientes a los sonidos cardiacos principales. En la Fig. 4 se observa el mismo FCG3 pero en el rango de 1.25 segundos a 3.24 segundos.

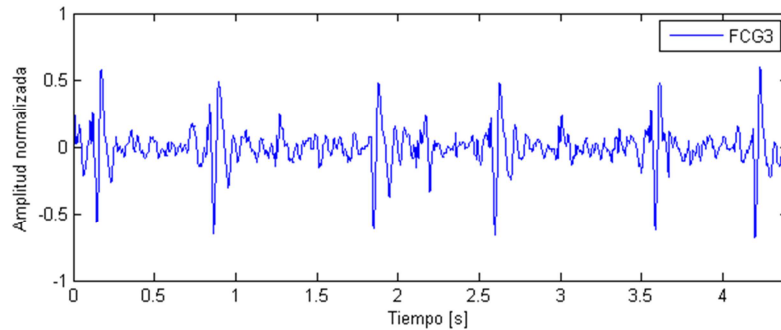


Fig. 3. Sección de FCG3, de 0 a 4.4 segundos.

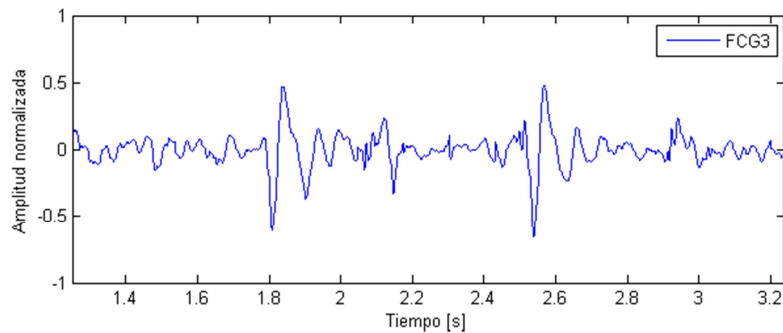


Fig. 4. Sección de FCG3, de 1.25 a 3.24 segundos.

2.3. Análisis espectral del FCG

El análisis espectral de una señal consiste en obtener los componentes característicos de la frecuencia de una señal. La transformada de Fourier definida en (1) proporciona información para hallar los componentes espectrales de la señal original, para señales continuas se representa como:

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) \exp(-j\omega t) dt \tag{1}$$

donde ω es el parámetro de la frecuencia; la información espectral se presenta generalmente como una gráfica de frecuencias: amplitud de los componentes vs.

número de componente o su equivalente en frecuencia. En la Fig. 5 se muestra el espectro de frecuencias correspondientes a 3 FCG.

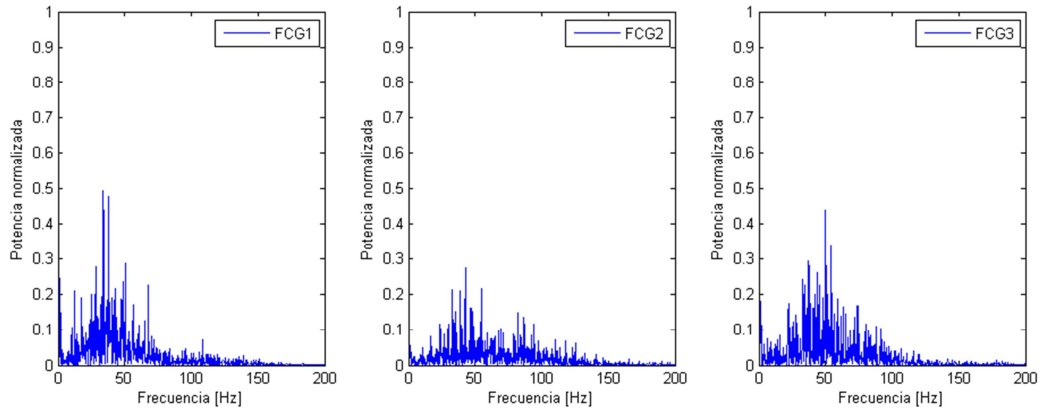


Fig. 5. Componentes espectrales de los fonocardiogramas FCG1 a FCG3.

La frecuencia máxima de las gráficas es de 200 Hz, originalmente se grafica hasta $fs/2$ que corresponde a 8 kHz, pero para mejor observación se limitó a 200 Hz pues no se obtuvieron valores de la potencia para frecuencias mayores a los 200 Hz. Las frecuencias de los componentes con mayor potencia se encuentran en la región cercana a los 50 Hz y a los 100 Hz; sin embargo, de este análisis no puede extraerse la información relativa a cuáles de los eventos del fonocardiograma corresponden estas frecuencias, ni el momento preciso en que ocurren estos eventos.

2.4. Análisis en el dominio tiempo-frecuencia

Para incluir al tiempo en el análisis y obtener la frecuencia de la señal de una mejor manera, se realiza el análisis empleando la STFT y la transformada wavelet. El objetivo de este análisis es encontrar las frecuencias características de los sonidos cardiacos principales, S1 y S2, así como identificar los intervalos de tiempos en que cada uno de estos sonidos ocurre y poder determinar así la duración del ciclo cardiaco. Para señales continuas, la STFT de la señal se representa por:

$$X(t, f) = \int_{-\infty}^{+\infty} x(\tau)w(t - \tau)\exp(-j\pi f\tau)d\tau \quad (2)$$

donde $w(t-\tau)$ es la función de ventaneo y τ es la variable que desliza esta ventana a través de toda la señal $x(t)$. Para obtener los espectrogramas se grafican la densidad espectral de potencia calculada con la STFT (ventana haming) definida en (2), en una gráfica 3D en función del tiempo y la frecuencia. La densidad espectral de potencia o espectro de energía de una señal se define en términos de la función de autocorrelación, para señales continuas se tiene:

$$PSD(f) = \int_0^T r_{xx}(\tau) \exp(-j2\pi f_T \tau) d\tau \quad (3)$$

La Fig. 6 representa las proyecciones de los espectrogramas del segmento de FCG3, las barras de color corresponden a la magnitud de la potencia. Se observa en la Fig. 6(a), que la frecuencia para los eventos con mayor potencia en este segmento de FCG3 es de 93.75 Hz. También aparecen en este espectrograma frecuencias de 164 Hz que corresponden a eventos con menor potencia. Los eventos con mayor potencia corresponden al primer tono cardiaco S1. Las dos frecuencias registradas para S1 corresponden a sus dos componentes, M1 (mitral) y T1 (tricúspide).

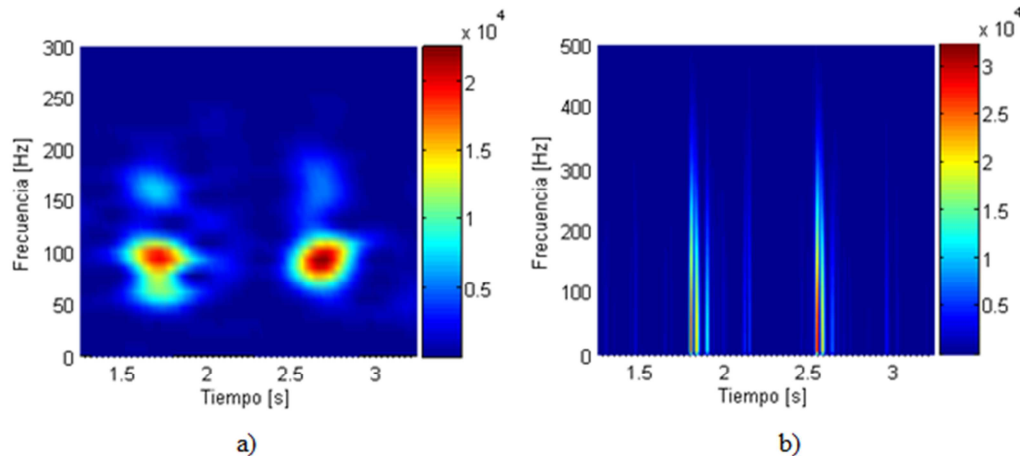


Fig. 6. Análisis en el dominio tiempo-frecuencia del segmento de FCG3, empleando una ventana de (a) 1024 puntos y otra de (b) 64 puntos para la STFT.

Con la Fig. 6(a) se pueden obtener las frecuencias representativas en FCG3 pero no el tiempo en el que ocurren, para esto empleamos la Fig. 6(b). Se observa que la potencia se disipa a lo largo del eje de la frecuencia, por lo que se los eventos están definidos en tiempo pero no en frecuencia. Los eventos con mayor potencia corresponden al primer tono cardiaco S1 y ocurren a los 1.8 y 2.6 segundos; los eventos con menor potencia que se encuentran entre los S1 consecutivos corresponden al segundo tono cardiaco S2 y ocurren aproximadamente a los 2.1 y 2.9 segundos. La densidad de potencia definida en (3) para estos espectrogramas se muestra en la Fig. 7, en (a) se grafica la densidad de potencia en el espacio de la frecuencia para una ventana de 1024 puntos y en (b) la densidad de potencia en el espacio del tiempo para una ventana de 64 puntos.

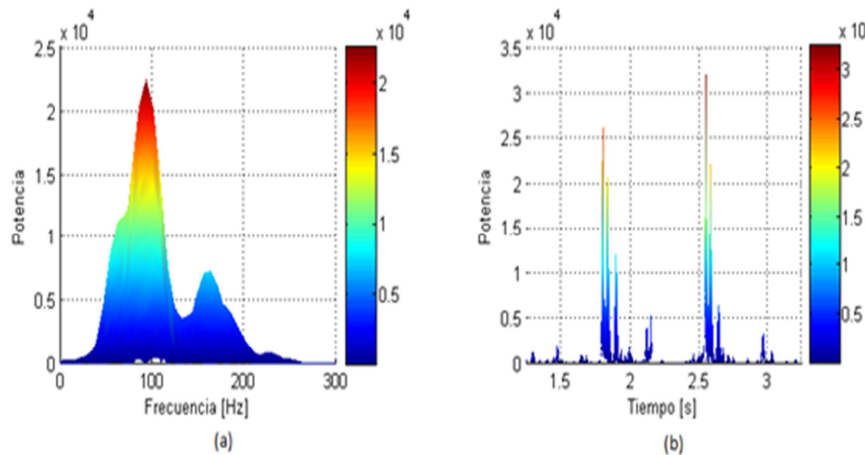


Fig. 7. Densidad de potencia en (a) el espacio de la frecuencia y (b) el espacio del tiempo.

De la figura 7(b) se observa que S1 y S2 no son eventos únicos, sino que tienen diferentes componentes. La correlación de la densidad de potencia temporal de la figura 7(b) con los eventos del segmento de FCG3 se muestra en la Fig. 8. En ella se observa que los picos con mayor potencia corresponden a S1, tiene diferentes componentes y éstos son detectables con el sistema implementado. Se detectan también los picos correspondientes a S2, los cuales son de menor potencia que los de S1 y también tienen componentes detectables.

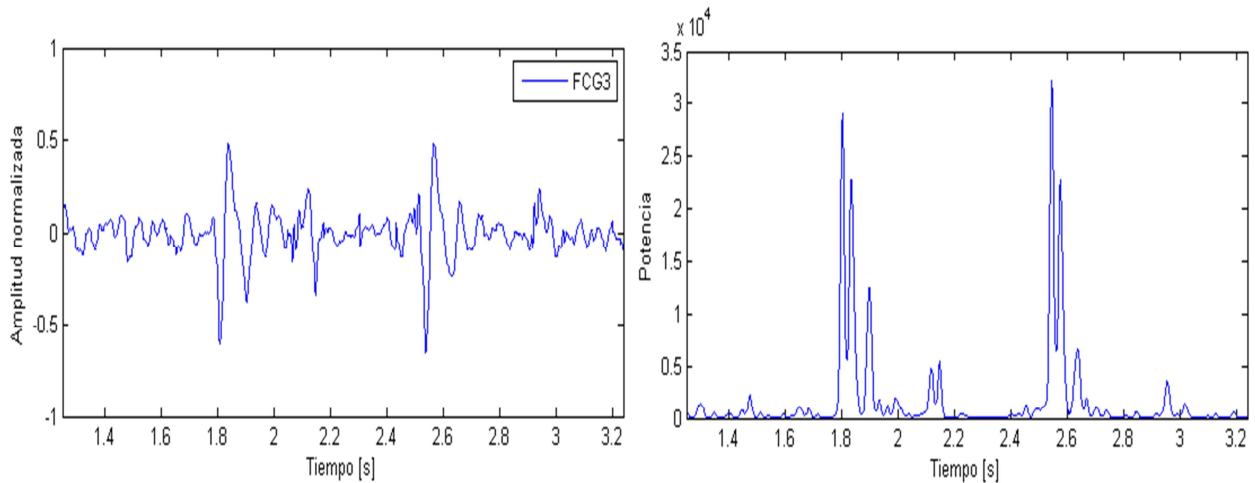


Fig. 8. Comparación de un segmento de FCG3 con su densidad de potencia obtenida mediante su STFT.

2.5. Escalogramas

Para no depender del tamaño de la ventana con el método de la STFT, se emplea el método de la transformada wavelet. La función wavelet que proporciona la mejor localización tiempo-frecuencia para el análisis de señales biomédicas es la wavelet Morlet definida en (4).

$$\psi(x) = \pi^{-1/4} \left(\exp(j2\pi f_0 x) - \exp(-(2\pi f_0)^2 / 2) \right) \exp(-x^2 / 2) \quad (4)$$

donde f_0 representa su frecuencia central. Para valores de $f_0 \gg 0$, se puede simplificar la expresión anterior de modo tal que:

$$\psi(x) = \pi^{-1/4} \exp(j2\pi f_0 x) \exp(-x^2 / 2) \quad (5)$$

Se obtiene el escalograma del segmento de FCG3 y se muestra en la Fig. 9; en el eje vertical se tiene a los coeficientes de la transformada, en los ejes horizontales se tiene al número de muestras del segmento de FCG3, que corresponde a 3978, y a la escala empleada al calcular la transformada wavelet.

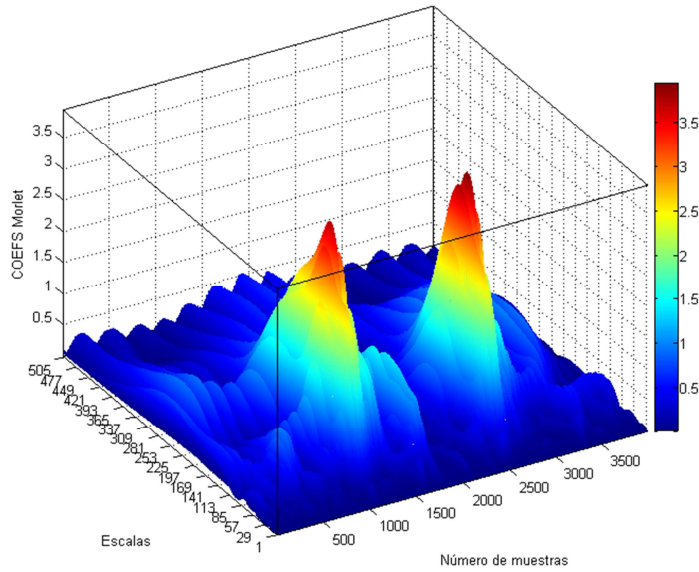


Fig. 9. Escalograma de FCG3 empleando la wavelet Morlet.

La densidad de potencia tiene su análogo para la frecuencia en la Fig. 10; el perfil de potencia servirá para la identificación en el tiempo de los eventos de FCG3.

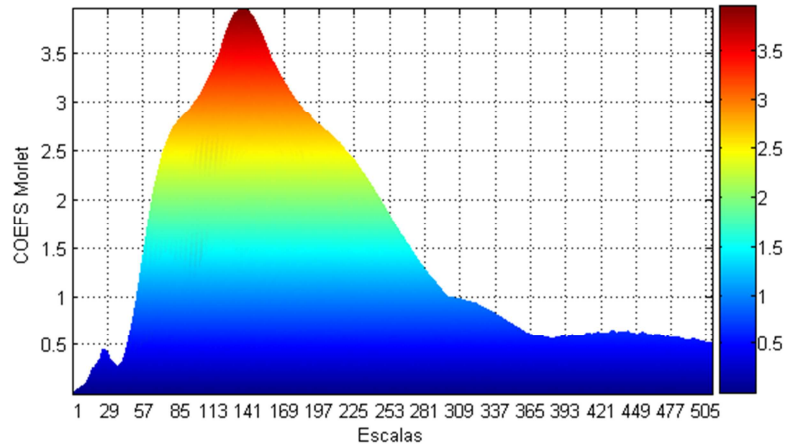


Fig. 10. Densidad de potencia de FCG3 en el dominio de la frecuencia (escala).

Al igual que en la Fig. 8, en la Fig. 11 se observa la correspondencia entre la densidad de potencia en el dominio del tiempo, obtenido mediante la transformada wavelet, y los eventos que ocurren en este segmento de FCG3. Los eventos con mayor potencia corresponden a S1 de nuevo, aunque mediante este análisis de la transformada

wavelet se aprecian mejor los picos para S2 que con la STFT. Se determinan los tiempos de los picos de mayor potencia, tanto para el análisis de la STFT como de la transformada wavelet; esto es para localizar a S1 y proceder a la segmentación del fonocardiograma.

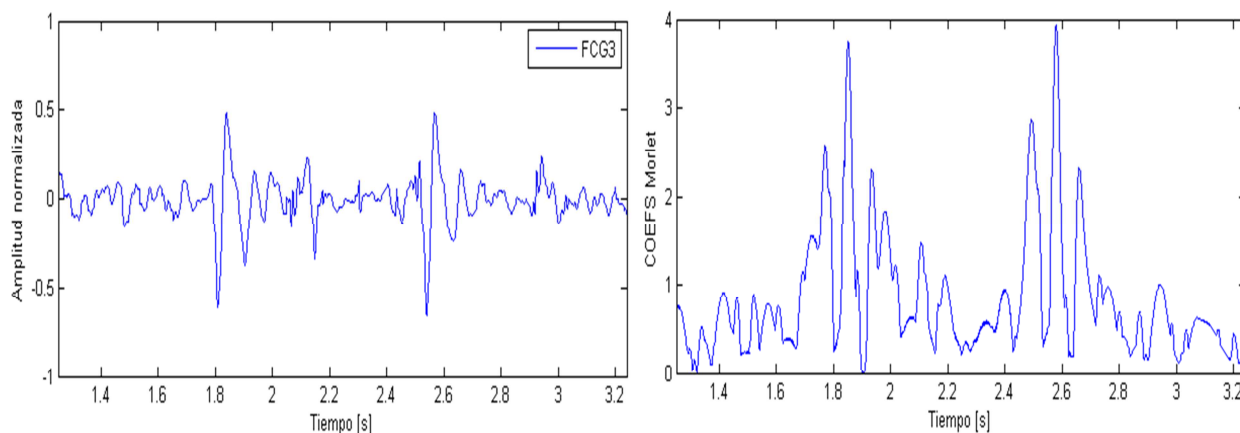


Fig. 11. Comparación de un segmento de FCG3 con su densidad de potencia obtenida mediante su transformada wavelet.

3. Resultados

La representación de los eventos acústicos del ciclo cardiaco en un fonocardiograma incluye a las actividades fundamentales, sonidos S1 y S2, así como a otras actividades no fundamentales (soplos, ruidos, etc.). Todos los eventos del ciclo cardiaco pueden variar su duración, amplitud o frecuencia de un latido a otro, en un mismo fonocardiograma incluso para el mismo sujeto. Se hacen ciertas consideraciones para la segmentación del FCG [6]: las actividades fundamentales del ciclo cardiaco, S1 y S2, son los eventos con mayor energía; la duración del intervalo de tiempo entre S1 y S2 (sístole) es mayor que la duración del intervalo de tiempo entre S2 y S1 (diástole).

La adquisición de la señal cardiaca se implementa en el procesador de señales digitales (DSP) del DSK C6713; la herramienta de programación del DSP es el Code Composer Studio (CCS), el cual se basa en el lenguaje C para la programación. El despliegue de

los datos en tiempo real se realiza mediante el programa MATLAB. Se desarrolló un proyecto en el CCS nombrado como *rtdx_matlab.pjt*, el archivo principal de este proyecto es *IIR ANC.c*, en él se encuentran los comandos para las tareas de adquisición de la señal cardiaca por medio de los dos sensores acústicos, los algoritmos del filtro IIR y el filtro adaptativo para la cancelación de ruido (ANC), así como los comandos necesarios para establecer la comunicación del CCS con la plataforma MATLAB para el despliegue de los datos en tiempo real.

Para el intercambio de datos en tiempo real (RTDX) con MATLAB, se crea un canal en el CCS que transfiera los datos del DSP hacia MATLAB mediante el comando *RTDX_CreateOutputChannel*. Ya que las muestras están listas, los datos se mandan a través de este canal hacia el MATLAB, donde se definen los parámetros adicionales para el RTD; es necesario crear un programa mediante el cual MATLAB registre esos datos, los despliegue de manera gráfica en tiempo real y los almacene para su análisis posterior. Este archivo de datos se nombra como *FCG.m* y se encarga de precisar la información acerca del DSP conectado a la PC. Los datos registrados se guardan en el archivo *fcg.mat*.

Se desarrolló un programa en MATLAB que se encarga de realizar la segmentación del archivo de datos *fcg.mat* de los fonocardiogramas resultantes de la adquisición. La densidad de potencia del espectrograma y el escalograma sirven de base para la segmentación. La primera tarea que realiza el programa consiste en hallar la densidad de potencia para ese segmento de FCG3, empleando la STFT con una ventana de 64 puntos para obtener una buena resolución temporal. La siguiente tarea consiste en la detección de los picos de la densidad de potencia, pues estos picos corresponden al primer tono cardiaco S1. Posteriormente encuentra los índices de los datos a los cuales ocurren esos picos y los tiempos en que se dan. Esto proporciona la localización temporal de S1. Ya que se conoce la ubicación en tiempo de los eventos de mayor potencia se hace una estimación de la duración del ciclo cardiaco del fonocardiograma.

En la Fig. 13 y Fig. 14 se muestran, respectivamente, los resultados finales de la segmentación del fonocardiograma en ciclos cardiacos individuales para el FCG3 desde 0 hasta 4.09 segundos empleando la STFT y la transformada wavelet. En este segmento de FCG3 se identificaron 4 ciclos cardiacos, con duraciones promedio de 0.753 ± 0.15 segundos empleando la STFT y de 0.769 ± 0.13 segundos con la transformada wavelet. Estos resultados corresponden a los valores estándares del ciclo cardiaco [7].

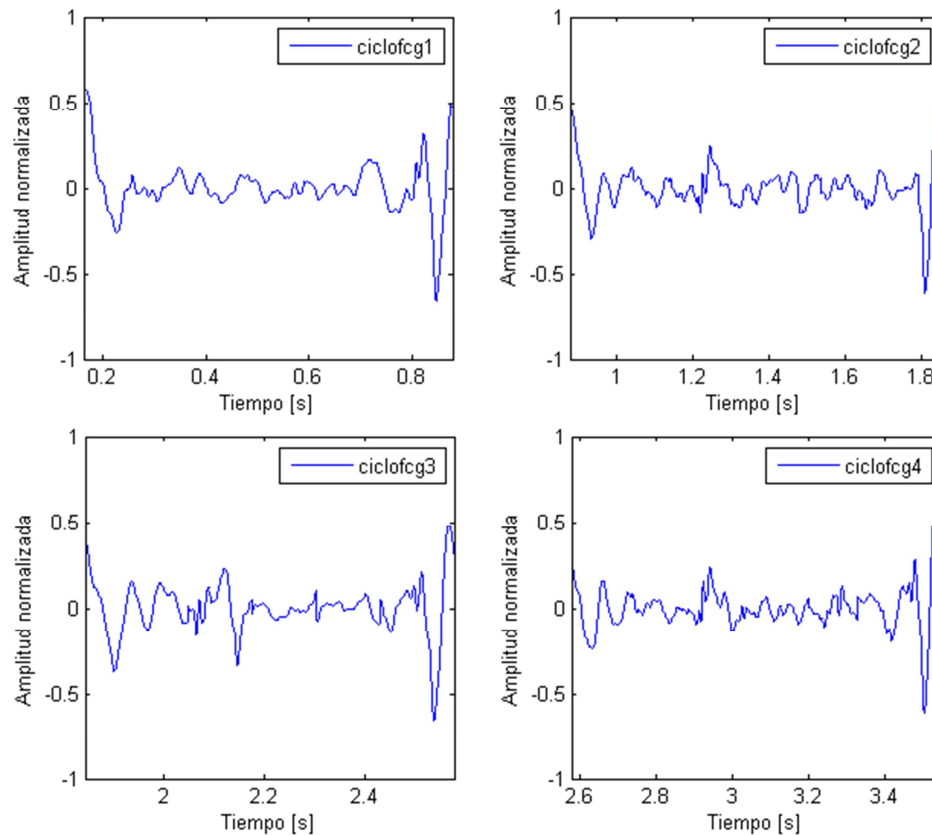


Fig. 12. Segmentación de FCG3 empleando la densidad de potencia obtenida de su STFT.

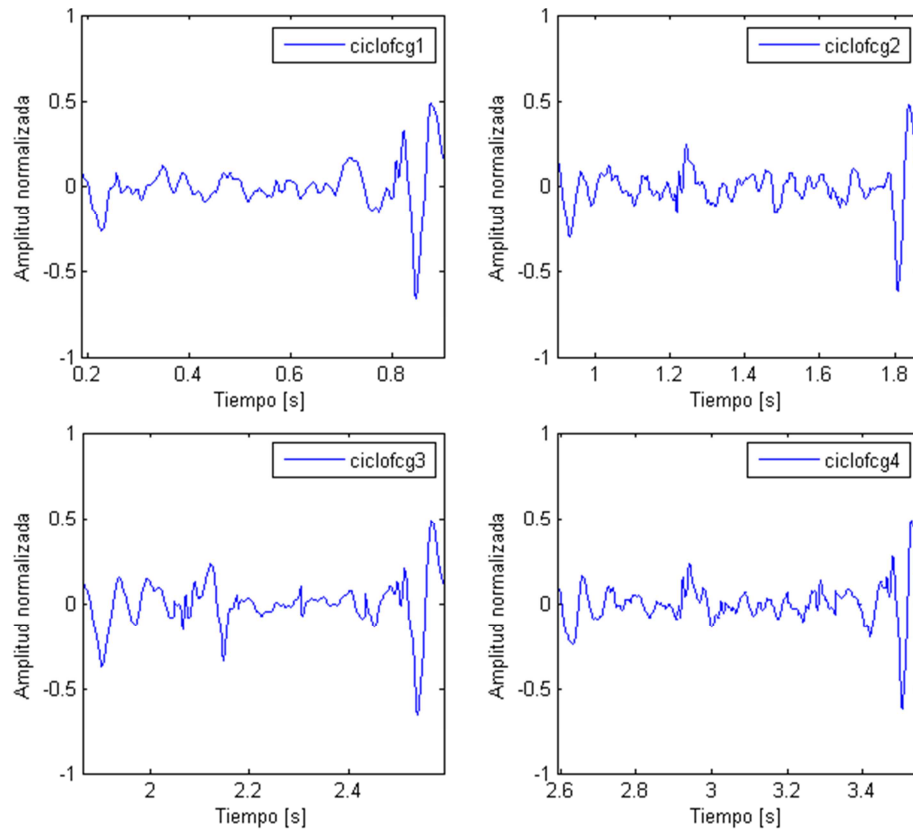


Fig. 13. Segmentación de FCG3 empleando la densidad de potencia obtenida de su transformada wavelet.

4. Discusión

En este apartado se presenta el resumen de la información obtenida acerca de los sonidos cardiacos principales, S1 y S2. Se realizó el análisis de los fonocardiogramas obtenidos con el sistema de monitoreo implementado tanto en el dominio del tiempo como en el de la frecuencia, mediante el análisis de las transformadas corta de Fourier y wavelet de la señal cardiaca. Se comparan los resultados con las mediciones obtenidas por [7] mediante el análisis de la STFT de un fonocardiograma de una persona sana.

| Característica | Análisis de la STT | Análisis de la transformada wavelet | Análisis de la STFT en [7] |
|---|--------------------|-------------------------------------|----------------------------|
| Frecuencia de S1 (componente con mayor energía) | 93.75 Hz | 92.2 Hz | 100 – 150 Hz |
| Frecuencia de S1 (componente con menor energía) | 164 Hz | 160.5 Hz | |
| Rango de frecuencias de S2 | No detectable | 78.8 – 144.5 Hz | 50 – 200 Hz |
| Duración del ciclo cardíaco | 0.753 ± 0.15 s | 0.769 ± 0.13 s | 0.75946 segundos |

Tabla 1. Resumen de la caracterización de los sonidos cardíacos.

Los valores obtenidos para las frecuencias de los sonidos principales S1 y S2 concuerdan con los valores que aparecen en la literatura (ver Tabla 1). La determinación de la duración del ciclo cardíaco permite estimar la variabilidad del ritmo cardíaco. La variabilidad es un índice para identificar cardiopatías como la arritmia, por lo que determinando la variación de la duración de cada ciclo es posible emplear este análisis como un método eficaz de diagnóstico. Para ampliar el alcance como herramienta de diagnóstico del sistema de monitoreo implementado es necesario desarrollar algoritmos de análisis mucho más finos para la identificación de S2 y de sus componentes para realizar la detección de soplos y sonidos anormales en el fonocardiograma.

5. Conclusiones

Los resultados obtenidos al identificar los principales eventos del ciclo cardiaco en personas sanas demuestran, de manera cuantitativa, la factibilidad tecnológica y el potencial médico del sistema de monitoreo implementado en este artículo, destacándose como un método no invasivo para la evaluación de la funcionalidad cardiovascular.

El procesamiento y extracción automática de las características temporales y espectrales de la señal acústica cardiaca resulta ser más efectivo mediante el análisis de la transformada wavelet, empleando la wavelet Morlet, que mediante el análisis de la transformada de Fourier en tiempo corto. Se obtuvieron valores estándares para la duración del ciclo cardiaco y la ubicación del primer tono cardiaco, lográndose una correcta segmentación del fonocardiograma sin necesidad de métodos adicionales como la sincronización de la señal acústica con el electrocardiograma.

El trabajo a futuro consiste en obtener los datos de fonocardiogramas de personas sanas y personas con arritmia, esto para evaluar la funcionalidad del sistema como un método de diagnóstico basado en la variabilidad del ritmo cardiaco. Se podrían integrar algoritmos para extracción de parámetros para el diagnóstico de enfermedades valvulares, identificando principalmente a los componentes de S1, M1 y T1, y a S2.

6. Referencias

- [1] H. Naseri, M. R. Homaeinezhad, "Computerized quality assessment of phonocardiogram signal measurement-acquisition parameters". *Journal of Medical Engineering & Technology*. Vol. 36. No. 6. 2012. 308-318 pp.

- [2] R. Watrous, D. Grove, D. Bowen, "Methods and results in characterizing electronic stethoscopes". IEEE, Computers in cardiology. Vol. 29. 2002. 653-656 pp.
- [3] H. Sava, L. G. Durand, "Automatic detection of cardiac cycle based on an adaptive time-frequency analysis of the phonocardiogram". Proceedings of the 19th IEEE-EMBS International Conference. Oct 1997. 1316-1319 pp.
- [4] S. Lukkarinen, K. Sikio, A. Nojonen, A. Angerla, R. Sepponen, "Novel software for real-time processing of phonocardiographic signal". Proceedings of the 19th IEEE-EMBS International Conference. Vol. 4. Oct. 1997. 1455-1457 pp.
- [5] J. Johnson, D. Hermann, M. Witter, E. Cornu, R. Brennan, A. Dufaux, "An ultra-low power subband-based electronic stethoscope". Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. Vol. 3. May 2006. 1156-1159 pp.
- [6] S. B. Emmanuel, "A review of signal processing techniques for heart sound analysis in clinical diagnosis". Journal of Medical Engineering & Technology. Vol. 36. No. 6. August 2012. 303-307 pp.
- [7] A. Djebbari, F. Bereksi-Reguig, "Short-time Fourier transform analysis of the phonocardiographic signal". Proceedings of the International Conference on Electronics, Circuits and Systems. Vol. 2. Dec 2000. 844-847 pp.
- [8] F. Meziani, S. M. Debbal, A. Atbi, "Analysis of phonocardiogram signals using wavelet transform". Journal of Medical Engineering & Technology. Vol. 36. No. 6. 2012. 283-302 pp.
- [9] A. Djebbari, F. Bereksi-Reguig, "Smoothed-pseudo Wigner-Ville distribution of normal and aortic stenosis heart sounds". Journal of Mechanics in Medicine and Biology. Vol. 5. No. 3. 2005. 415-428 pp.

7. Autores

M. en C. Braulio Cruz Jiménez obtuvo su grado de Maestría en Ciencias con especialidad en Automatización por el Tecnológico de Monterrey, su área de investigación son los sistemas de control y procesamiento de señales.

M. en C. Jannette Contreras Rivero obtuvo su grado de Maestría en Ciencias con especialidad en Sistemas Ambientales por el Tecnológico de Monterrey, su área de investigación es matemáticas aplicadas.

Dr. Luis Ricalde Catellanos obtuvo su grado de Doctor en Ingeniería Eléctrica por el Centro de Investigación y de Estudios Avanzados del IPN Unidad Guadalajara, su área de investigación son los sistemas de control utilizando redes neuronales.

El invisible y asombroso proceso de la comunicación oral: bases sobre reconocimiento de voz

Ángel David Pedroza Ramírez

Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica, Carretera a la Bufa No.1
Col. Centro Zacatecas, Zac., Teléfono: (492) 92 296 99
P.A.D_16@hotmail.com

José Ismael de la Rosa Vargas

Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica, Av. López Velarde No.
801 Col. Centro Zacatecas, Zac., Teléfono: (492) 925 66 90, ext. 3956
joseismaelrv@gmail.com, ismaelrv@ieee.org

Resumen

La comunicación, cuyo fin primario es la transmisión de información, forma parte fundamental de las necesidades básicas del ser humano. El proceso de la generación del habla y más aún el de la comunicación, es mucho más complejo de lo que se podría llegar a creer dado el nivel de coordinación que se requiere para producir, transmitir y decodificar un mensaje. El reconocimiento de voz se basa en el estudio sobre el proceso del habla y la comunicación, y la forma en que este conocimiento puede ser aplicado.

El presente documento resume y brinda una revisión sobre el estado del arte y las bases para entender el reconocimiento de voz desde el punto de vista fisiológico y como una rama de la ciencia ampliamente utilizada en la tecnología de uso cada vez más común hoy en día.

Palabra(s) Clave(s): comunicación hombre-máquina, comunicación oral, habla, voz.

1. Introducción

La comunicación oral es una de las herramientas de expresión y comunicación más importante que en el pasado dio paso hacia la evolución y que en el mundo actual, forma parte de una necesidad básica.

El proceso de comunicación oral (entre dos personas) está formado por tres elementos principales: Locutor, medio y oyente. El locutor, mediante una serie de procesos de coordinación de órganos, articula las palabras (previamente procesadas por el cerebro) para formar un mensaje. Dicho mensaje se transmite por un medio (generalmente aire) para llegar al oyente (receptor) quien lo decodifica y así cierra el ciclo o cadena del habla.

El reconocimiento de voz es una rama de la investigación multidisciplinaria y que ha logrado dar solución a diversas problemáticas del mundo real que van desde la ayuda a personas con deficiencias hasta los complejos sistemas de navegación con tecnología GPS. Específicamente el reconocimiento de voz es de nuestro especial interés dado que mediante este podemos lograr, una vez extraída la información por algún método, reconocer palabras con algún margen de error para después utilizarla en algún proceso (como es el caso de la realización de pruebas de dicción).

En las siguientes secciones se dará una breve reseña sobre los aspectos más importantes que fundamentan al reconocimiento de voz aunado a aspectos relacionados con el estado del arte del área de investigación, pasando por una reseña histórica en la sección 2, luego presentando algunos aspectos formales sobre modelos de producción de voz en la sección 3, hasta llegar a la sección 4 en donde se presentan los aspectos más relevantes sobre el reconocimiento de voz y cómo se implementa para la comunicación hombre-máquina. Finalmente, en la sección 5 se presentan

algunas conclusiones sobre el diseño de sistemas de reconocimiento y algunos comentarios sobre el trabajo que actualmente realizamos.

2. Reseña histórica

Durante millones de años los seres vivos han utilizado la comunicación para diversas necesidades y de dicha forma transmitir información para lograr interactuar con el exterior.

El ser humano ha tenido la necesidad de transmitir ideas, sentimientos y pensamientos dada la necesidad de vivir en sociedad. En un inicio la comunicación se hacía solo por medio de sonidos o de forma pictográfica, sin embargo, no fue hasta que se logró tener una comunicación con sonidos articulados que tomó la importancia que tiene hoy en día.

La información que se transmite a través de la voz posee características que pueden ser analizadas por medio de sistemas de reconocimiento de voz. Dicha rama de la ciencia se subdivide en reconocimiento del habla y reconocimiento del locutor.

Entre las aplicaciones donde se encuentra el reconocimiento de voz en la actualidad se encuentran telefonía, Seguridad, Manos libres, etc.

Al principio los sistemas de reconocimiento de voz se enfocaron en la producción sintética de voz, como fue el caso de la máquina creada por Von Kempelen como lo describe [1], la cual, mediante un arreglo mecánico, producía sonidos parecidos a los producidos por el ser humano, sin embargo este sistema requería de horas de entrenamiento. Este novedoso pero burdo invento daría más tarde lugar a los sintetizadores musicales y a los órganos que utilizaban tarjetas perforadas para producir ciertas melodías.

Poco a poco los sistemas fueron haciéndose más sofisticados hasta que años más tarde Thaddeus Cahill descubrió que cualquier sonido podía ser sintetizado como una

sumatoria de pequeñas ondas sinusoidales el cual revolucionó el concepto sobre la síntesis de música. Diversos inventos fueron creados, como el caso del *telharmonium* el cual utilizaba este principio de adición de ondas.

Sin embargo, el primer invento enfocado al reconocimiento de voz fue un juguete llamado *Radio Rex*, como lo indica [2], manufacturado en 1926. Este funcionaba mediante un mecanismo de activación por comando de voz. Al ser nombrada la palabra "Rex", el mecanismo hacía que la energía contenida en la palabra disparara el dispositivo y liberara al perro. Sin embargo, este dispositivo comercial tenía la desventaja que cualquier palabra con la misma energía que la palabra "Rex", hacía que el sistema funcionara, es decir, liberara al perro.

No fue sino hasta 1950 cuando los laboratorios Bell crearon el primer reconocedor más sofisticado hasta esa fecha; logrando reconocer dígitos de un solo hablante. Sin embargo aún era un poco burdo para las necesidades tecnológicas.

El ámbito del reconocimiento de voz siguió desarrollándose y en 1960 aparecen tres técnicas que fueron la base para el reconocimiento de voz (indicadas en [3]):

- Transformada Rápida de Fourier (FFT): Forma eficiente de la transformada de Fourier, interpretada como un banco de filtros.
- Análisis Cepstral: Utilizado como una aproximación espectral.
- Codificación lineal predictiva (LPC): Modelos de auto-regresión para representar la generación de voz.

Adicionalmente se crearon nuevos métodos de aproximaciones:

1. Alineamiento temporal dinámico (DTW por sus siglas en inglés como lo indica [7]): Esquema de optimización, utilizado como método de normalización (puede ser utilizado para corregir variaciones por diferentes pronunciaciones de la misma palabra con diferente duración).

2. Modelos ocultos de Markov (MOM o HMM por sus siglas en inglés como lo indica [11]): Modela una secuencia observada como la generada por una secuencia desconocida de variables.

Específicamente, los modelos ocultos de Markov fueron desarrollados por primera vez por un grupo de IBM en 1970 los cuales, en primera instancia, se utilizaron para el reconocimiento continuo del habla.

Para 1976 se incorpora información semántica y de sintaxis con el fin de lograr más allá del reconocimiento del habla la comprensión del lenguaje. Para este fin, se incorporaron y desarrollaron investigaciones con redes neuronales, DTW y HMM.

Los Modelos Ocultos de Markov tuvieron gran impacto hasta mediados de 1980, siendo implementados hasta la fecha en algunos sistemas comerciales (Sistema de reconocimiento de voz de la marca Apple llamado Siri como lo indica en [4]).

A finales de 1980 se utilizó información acústico-fonética para desarrollar reglas de clasificación para sonidos del habla. Para ese mismo año se utilizaron redes neuronales para delimitar secciones de frontera (dentro de la señal de voz, ¿qué parte de la señal posee información y cuál no?). Luego de una serie de avances en materia de bases de entrenamiento y subsistemas de extracción de información de la señal de voz, para 1998 algunos sistemas podían reconocer cerca de 60,000 palabras de vocabulario en tiempo real con menos del 10 por ciento de error.

Hoy en día, y luego de un proceso que como se explicó con anterioridad comenzó con invenciones burdas hasta llegar a los complejos sistemas híbridos, el reconocimiento de voz es un área de investigación multidisciplinaria que tiene muchísimas aplicaciones aún por seguir explorando.

3. Teoría fundamental sobre el proceso de producción de voz

3.1. El sonido

El sonido, como se define en [5], es una vibración (o movimiento ondulatorio) que se transmite en el aire (generalmente), hasta alcanzar el oído de quien lo percibe. Dicha vibración tiene cierta amplitud, frecuencia, duración y forma. Las características de la vibración caracterizan al sonido producido.

Siendo así, la frecuencia corresponde al tono del sonido (también llamado altura). Las frecuencias que son altas son denominadas "frecuencias en tono agudo" y las frecuencias bajas se denominan "frecuencias en tono grave". Es interesante mencionar que la cualidad de percepción la proporciona el tono mientras que el dolor auditivo experimentado en la audición lo da la energía y no, por tanto, el tono que se produce.

Al respecto, el sonido se puede clasificar de la siguiente manera:

- Infrasonidos: Frecuencias menores a 20 Hz.
- Percepción humana: Entre 20 y 20,000 Hz.
- Ultrasonidos: Frecuencias mayores a 20,000 Hz.

La amplitud se corresponde directamente a la intensidad del sonido (volumen) clasificado como "alto" o "bajo". Esta característica se refiere a la distancia entre el extremo superior e inferior de la onda de sonido. Dicho parámetro está relacionado con la intensidad y por tanto con la energía.

| Características de la vibración. | | Características del sonido |
|---|---------|-----------------------------------|
| Frecuencia | produce | Tono |
| Amplitud | " | Intensidad |
| Duración | " | Duración |
| Forma | " | Timbre |

Tabla 1. Correspondencia vibración a sonido.

La duración de la vibración es también la duración del sonido. Por otro lado, las vibraciones parásitas permiten identificar a la fuente de sonido (a cada una corresponden diversos armónicos) dado que son producidas por diversas formas de vibración. Esa cualidad es llamada en la teoría del sonido como timbre. Así, el sonido de una flauta y un violín con la misma intensidad, tono y duración se diferencian por el timbre (ver Tabla 1).

Las vibraciones que puede captar el ser humano como el sonido están limitadas por la frecuencia. Inferiores a 30 ciclos por segundo se percibe como movimiento mientras que, si sobrepasan los 18,000 ciclos por segundo rebasa las posibilidades de audición.

Si la amplitud es muy leve no se produce sensación auditiva pero si es muy grande, puede dañar de forma permanente el órgano de la audición. En cuanto a la duración, si es menor a 100/10000 de segundo solo se percibe como un "click" no identificable. Si la forma de vibración y frecuencia permanecen regulares se percibe como sonido; si alguno de los dos es irregular, el resultado es un ruido.

3.2. ¿Cómo se produce la voz?

La voz humana es conocida como el flujo de ondas (de tipo sonoras o silencios) que se propaga por medio del aire (presión de moléculas). El proceso de producción de voz conlleva la combinación de órganos, huesos, músculos y algunos sistemas funcionales (como lo explica [18]). La combinación de estos elementos en conjunto con la postura, la respiración y el estado emocional influyen en dicho proceso de producción (como se puede ver en [6]). Este proceso se resume de la siguiente forma: Previa inhalación de aire en los pulmones, la voz se produce cuando el aire es exhalado; al expandirse el diafragma, los órganos de respiración proporcionan un flujo de aire a los órganos de la fonación, y es ahí donde el sonido adquiere sus características primarias en donde las cuerdas vocales juegan un rol importante. Luego, se agregan otras características impuestas por los órganos articulatorios, brindando al final una señal acústico-fonética.

El ser humano posee un sistema productor de sonido con ciertas características anatómicas (ver Fig. 1). En el sistema fonador interviene no solo la laringe, sino también los pulmones (para proporcionar el aire) y los modificadores de sonido (resonadores faringo-buco-nasales). Es por ello que la calidad de sonidos producidos está directamente determinada por la posición de la lengua, los labios y la mandíbula.

La fonación funciona mediante dos sistemas: Sistemas directos y sistemas indirectos. Los sistemas directos están compuestos por el sistema respiratorio, órganos articulatorios, órganos de resonancia y el aparato fonador. Por otro lado, los sistemas indirectos están conformados por el sistema muscular, el aparato auditivo y el sistema óseo.

La fuente de aire en el sistema de producción de voz se da gracias a los pulmones en la espiración. Según la cantidad de aire que se expulse por los pulmones es la intensidad que se tiene en la voz; mientras que gracias al diafragma se hace posible el variar su duración.

Específicamente el aparato fonador es el encargado de reproducir el sonido y está conformado por la laringe y las cuerdas vocales.

La laringe, además de proteger las vías respiratorias, es una caja sonora encargada de la fonación (donde se origina el tono fundamental de la voz). Entre las funciones más importantes de la laringe, como lo indica [6], se encuentran la respiratoria (permite entrada y salida del aire al respirar), esfinteriana (permite realizar esfuerzos), deglutatoria (impide el paso de comida a los pulmones) y de fonación (produce la voz gracias a las cuerdas vocales).

Las cuerdas vocales se encuentran dentro de la laringe y son membranas de tejido de la laringe. Estas producen el sonido mediante la vibración al paso del aire y la aproximación o separación entre si producen voz o silencio respectivamente. Su forma, aunado a las de la garganta, nariz y boca, es decir, las cavidades resonantes, determinan el sonido de la voz en cada persona.

El tono de voz se relaciona con la frecuencia a la que vibran las cuerdas vocales. Específicamente, en los adultos masculinos se presenta una frecuencia baja (120 ciclos por segundo) por lo que poseen una voz más grave. En el caso de las mujeres (250 ciclos por segundo) se presenta una voz aguda; en los niños y jóvenes es aún más aguda (400 ciclos por segundo). Es necesario aclarar que, aunado a esto, la longitud, tensión y masa de las cuerdas vocales influyen en el tono de voz. Por ejemplo, las cuerdas vocales más gruesas vibran más despacio por lo que producen un sonido más grave.

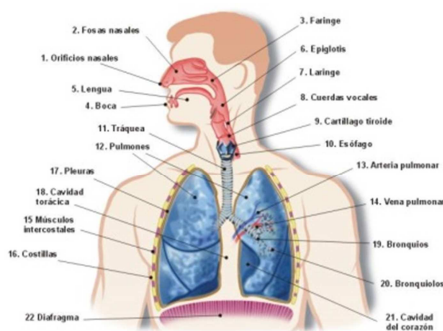


Fig. 1. Sistema Fonador.

En relación al aspecto psicológico y la voz, este parámetro es notorio en la respiración. La respiración varía en un ritmo particular en proporción al estado de ánimo (acelerado ante la excitación y limitado ante la tristeza).

3.3. Clasificación de los fonemas

En cuanto a la forma en que se clasifica cada sonido, cada lengua o lenguaje posee un conjunto propio de sonidos que el tracto vocal permite producir con naturalidad. A las unidades sonoras propias de un idioma se les llama "fonemas" (representados entre diagonales) y varían según las características impuestas por cada idioma. La combinación entre fonemas genera sílabas y a su vez, un conjunto de sílabas forma una palabra.

Una de las características más importantes de los fonemas es que, a un enunciado hablado, se le puede descomponer en estas unidades sonoras y si se cambia o sustituye por otro se cambia el contenido lingüístico. Se debe distinguir entre fonemas y letras ya que los fonemas son elementos sonoros de un idioma; mientras que las letras son las representaciones gráficas de los fonemas. En algunas ocasiones un mismo fonema es representado de diversas formas como en el caso de la representación de k, c y q para el mismo sonido.

Las frecuencias formantes permiten clasificar los diferentes sonidos o fonemas. Dichas formantes dependen de la dimensión y forma interna de la boca o tracto vocal (cada configuración provee unas formantes particulares como lo indica [7]).

En cuestiones de modelado, solo es necesario tomar en cuenta las formantes F1, F2 y F3. Las dos primeras formantes las determina la posición de la lengua; cuanto más baja esté, F1 posee una frecuencia más alta. Para el caso de F2, su frecuencia es mayor cuanto más hacia adelante esté la lengua.

Por otro lado, la frecuencia fundamental (pitch) o F0 corresponde a la frecuencia de oscilación de las cuerdas vocales en los sonidos sonoros. Si la frecuencia fundamental es mayor que la de la frecuencia de las formantes se hace difícil de distinguir y por ende difícil al reconocer.

Para el caso del idioma español la subdivisión de los fonemas es la siguiente:

1. Timbres Básicos,
2. Sonidos auxiliares.
3. Ataques:
 - Fuertes,
 - Suaves,
 - Sonidos Mixtos.

Timbres Básicos: Producidas con la intervención de las cuerdas vocales. Se denominan como timbres básicos debido a que son variaciones del sonido producido naturalmente por las cuerdas vocales y la columna de aire que se produce mediante la respiración. En esta clasificación se encuentran las vocales.

Para el caso del idioma español las vocales son: /a/, /e/, /i/, /o/ y /u/. De entre ellas la que se produce con naturalidad es la /a/. En cualquier otra, la posición de la lengua se modifica y por ello se crea otra clasificación como lo indica [8]:

- i. **Vocales Palatales:** La lengua se expone gradualmente y se eleva hacia el paladar. En esta clasificación se encuentran las vocales /e/ e /i/.
- ii. **Vocales velares:** La lengua se contrae y se eleva al velo del paladar. En esta clasificación se encuentran las vocales /o/ y /u/.

De modo general, las vocales se pueden clasificar de acuerdo al grado de abertura de la boca, posición de la lengua y grado de sonoridad (ver Fig. 2). La correcta producción de ellos (sin esfuerzo y con naturalidad) es el material de estudio en la impostación de la voz.

Sonidos auxiliares: Se producen por el aparato resonador sin intervención de las cuerdas vocales. La vibración no se produce en la laringe sino en el paladar blando, la lengua, etc. Corresponden a las consonantes sonoras: /m/, /n/, /l/, /s/, /j/, /r/.

Ataques: No son propiamente un sonido dado que no posee una vibración periódica. Son modos de iniciar un timbre básico o un sonido auxiliar y representan las distintas formas de la oclusión que impide fluir al aire. Una vez liberada, la columna de aire vibra produciendo un timbre básico o un sonido auxiliar; en ese momento el ataque deja de existir. Su naturaleza es momentánea y requiere de su identificación perfecta.



Fig. 2. Triángulo vocálico del español (HELWAG).

Pueden ser suaves o fuertes según la brusquedad de la liberación de aire. Ellos son: /b/, /p/, /d/, /t/, /k/. Los sonidos mixtos, como caso particular de ataques, se refiere a la sucesión de sonidos auxiliares /l/ o /n/ y el timbre básico /i/, los cuales producen los fonemas /ll/ y /ñ/; o las combinaciones del ataque /k/ con el sonido auxiliar /s/ produciendo /x/.

La calidad de la dicción se fundamenta en la correcta articulación de los ataques. Por otro lado, algo que debe ser tomado en cuenta para poder hablar, es respetar las reglas que posee cada idioma para formar combinaciones de fonemas. Por ejemplo, para el caso del castellano, los sonidos básicos y los sonidos auxiliares son combinables con cualquier fonema. Los ataques casi nunca son combinables entre si y no pueden producirse aislados; requieren de la combinación con un timbre básico o un sonido auxiliar.

4. ¿Qué es el reconocimiento de voz?

El reconocimiento de voz se basa en el conocimiento sobre el proceso del habla y la estructura que conlleva el lenguaje en el ser humano. Desde esta perspectiva se tiene que el fin último del reconocimiento de voz es hacer que las computadoras (máquinas) logren un nivel suficiente para expresar y comprender la comunicación como lo hace en su naturaleza el ser humano. Actualmente las empresas más importantes en desarrollo de aplicaciones en sistemas de reconocimiento de voz, como en [9], son Philips,

Lernout & Hauspie, Sensory Circuits, Dragon Systems, Speechworks, Vocalis, Dialogic, Novell, Microsoft, NEC, Siemens, Intel.

El reconocimiento automático de voz (RAV) se divide en dos áreas, como se indica en [7], enfocadas cada una en una tarea específica:

- I. Reconocimiento automático del habla (RAH).
- II. Reconocimiento automático de locutor (RAL), el cual se subdivide a su vez en:
 1. Identificación automática de locutor (IAL).
 2. Verificación automática del locutor (VAL).

Los sistemas de reconocimiento se pueden además clasificar como dependientes e independientes del locutor. Para el caso de los sistemas dependientes del locutor (cooperativo), el mensaje que se identifica o articula es una palabra o mensaje fijo (frase pre-establecida); mientras que para el caso en el que son independientes del locutor (no cooperativo), se posee la libertad para articular cualquier mensaje para ser reconocido.

Debe aclararse que para este último caso, la librería es mucho más extensa en relación al vocabulario que el locutor posee por lo que se puede decir que tiene la "libertad" para pronunciar cualquier mensaje. La eficiencia de los sistemas dependientes es mayor que la de los sistemas independientes.

El estudio del reconocimiento de voz se basa en tres principios:

1. La información de la señal de voz se puede representar por el espectro en amplitud a corto plazo de la forma de onda de la voz.
2. El contenido de la voz se puede expresar en forma escrita (una secuencia de símbolos fonéticos o caracteres del alfabeto).
3. Es un proceso cognoscitivo, es decir, la comprensión está ligada a la gramática, semántica y estructura del lenguaje.

La dificultad que se presenta para poder llevar a cabo reconocimiento de voz (en cualquiera de sus áreas) es la variabilidad presente en las señales de voz, es decir, las variaciones que se presentan para cada locutor, las condiciones acústicas, entre otras. Sin embargo, dentro de estas variaciones se debe determinar cuáles de ellas son relevantes y cuáles no, es decir, para cada aplicación determinar los parámetros a tomar en cuenta y los que se pueden despreciar.

En general, entre los problemas y dificultades que se encuentran en las tareas de reconocimiento de voz se consideran:

- El tamaño del vocabulario,
- ambigüedad acústica y grado de confusión,
- calidad ambiental.

Entre las más importantes (existen algunas adicionales).

Existen cuatro enfoques principales, como lo indica [10], que se pueden utilizar para realizar el reconocimiento del habla:

1. Contraste de patrones

- Supone al habla como una secuencia de palabras cada una con un patrón o conjunto de patrones.
- Se compara la unidad de habla entrante con los patrones de referencia almacenados.
- Utilizado en reconocimiento de palabras aisladas o conectadas.
- Esta técnica no permite generalizar los patrones por lo que es necesario crear una biblioteca para cada hablante lo cual reduce su nivel de aplicación en tareas de reconocimiento avanzadas.

-

2. Sistemas basados en el conocimiento:

- Emula y aplica los conocimientos sobre el habla en tareas de reconocimiento que utiliza el ser humano.
- Usa técnicas con reglas y sistemas expertos, desde el nivel acústico-fonético hasta niveles más complejos.

3. Modelos estocásticos:

- Hace uso de modelos estocásticos (Modelos Ocultos de Markov) en lugar de modelos determinísticos.
- Utilizado comúnmente para el reconocimiento de palabras continuas.

4. Modelos neuronales o conexionistas:

- No posee tantas restricciones como los modelos estocásticos.
- Sus tiempos de entrenamiento son razonables.
- Se han utilizado también como parte de fusiones entre varios enfoques.

Es necesario aclarar que el uso de cada uno o combinación entre ellos, dependerá de cuál es el objetivo que el reconocimiento pretende alcanzar.

Para llevar a cabo el reconocimiento de voz es necesario obtener los parámetros que representen la información espectral contenida en la señal de voz. Para ello, es necesario aplicar un procesamiento matemático con el fin de adecuar la señal para poder analizarla posteriormente.

En principio, es necesario aplicar un pre-procesamiento para después extraer los parámetros propios de la señal. Algunos de los principales métodos de análisis que se aplican en el reconocimiento de voz son: Bancos de Filtros, LPC, Cepstrum, etc.

Por otro lado, existen diversas técnicas para llevar a cabo el reconocimiento de voz. Algunas de ellas, están enfocadas al reconocimiento de palabras aisladas y otras más son comúnmente utilizadas en reconocimiento de habla continua. Sin embargo es necesario tomar en cuenta que la elección de ellas dependerá del objetivo final que se

pretende alcanzar. De los enfoques anteriores, técnicas como Cuantificación Vectorial (CV) (como lo indica [12]), Alineamiento Dinámico en el tiempo (ADT o DTW), Modelos Ocultos Markov (MOM o HMM) (como se indica en [13]) y redes neuronales son algunas de las principales utilizadas en tareas de reconocimiento de Voz.

Cada una tiene algunas ventajas y desventajas propias del enfoque que posee cada uno. Una de las técnicas que actualmente se utiliza para llevar a cabo el reconocimiento de voz con un alto grado de eficiencia (por encima del 90%) son los Modelos Ocultos de Markov (posibilitando la creación de sistemas más complejos de reconocimiento como el descrito en [22]).

La ventaja de utilizar esta técnica respecto de otras es que el reconocimiento, al contrario de los métodos de comparación no estadísticos, es que el reconocimiento se hace respecto de un modelo que caracteriza de forma específica a cada palabra. Estos modelos pueden ser creados con características específicas (modelos en base a fonemas, con una dicción específica, en un idioma concreto, etc.) lo que le da la flexibilidad necesaria para ser aplicado en varias áreas. Ejemplo de ello es la investigación realizada sobre, una vez realizado el reconocimiento, cómo adecuar los Modelos de Markov para identificar la dicción presente en muestras de voz y en base a ello poder realizar ejercicios de dicción de diversas palabras en un idioma específico.

La idea básica de los Modelos Ocultos de Markov aplicados en reconocimiento de voz consiste en: i) crear un modelo para cada una de las palabras que se desean reconocer (proceso de entrenamiento) y ii) almacenarlas con sus respectivas etiquetas. Una vez finalizado este proceso, cuando se desea reconocer una nueva palabra, se calcula la probabilidad de que la repetición a reconocer haya sido generada por un modelo específico (proceso de prueba) y en base a esto se identifica la palabra (como lo explica a detalle [19]).

Actualmente se trabajan en conjunto con otras técnicas como Redes Neuronales (base del asistente comercial Siri), Alineamiento Dinámico en el Tiempo (como en [14]), etc.; con la finalidad de mejorar la eficiencia de reconocimiento.

Por otro lado, las redes neuronales (Neural Networks - NN) consisten en unidades de cálculo simple interconectadas (como lo indica [17]). Estas pretenden interconectar un conjunto de unidades de proceso (o neuronas) en paralelo de forma similar a como lo hace el ser humano (como lo indica [16]) (obteniendo también prestaciones similares en reconocimiento tanto en tiempo de respuesta como en tasa de error). Este método es útil cuando se desean evaluar varias hipótesis en paralelo (como lo indica [15] y [23]) o como herramienta de ayuda en algunas aplicaciones (como la descrita en [21]).

La idea básica detrás de una red neuronal es: dados una serie de parámetros, combinarlos con la finalidad de predecir un cierto resultado. Las unidades de proceso son de varios tipos pero la más utilizada dispone de varias entradas y la salida es el resultado de una transformación no lineal a la combinación lineal de las entradas.

En cuanto al tipo de red, son definidas en cuanto a la forma en que se conectan las neuronas, el tipo de neurona que lo conforma y la forma de entrenamiento de la red (como se explica en [20]).

5. Conclusiones

De la presente investigación se pueden obtener las siguientes conclusiones:

- El reconocimiento de voz es un área de investigación con bases bien fundamentadas y que ha dado la pauta para la solución de problemas en el mundo actual y para la comunicación hombre-máquina.

- Aún con las dificultades que se presentan para lograr un reconocimiento óptimo, es necesario saber notar cuales de éstas características son relevantes y cuáles no, dependiendo de la aplicación a la que se enfoque el reconocimiento de voz.
- Las escuelas de reconocimiento del habla existentes hasta hoy en día han respondido a una necesidad tecnológica presente en cada una de sus épocas de desarrollo, sin embargo, la elección de cada uno de los métodos que se debe utilizar responde al tipo de aplicación que se desea realizar.
- De entre las técnicas para reconocimiento de voz, los Modelos Ocultos de Markov conforman una buena herramienta capaz de ser aplicada en diversas áreas gracias a la flexibilidad con que se pueden diseñar, particularmente en tareas de mejora de dicción que es uno de nuestros objetivos.
- Independientemente de la técnica que se utilice, los métodos de reconocimiento siempre tendrán errores en mayor o menor proporción debido a que se basan en aproximaciones de la señal a identificar.

6. Referencias

- [1] B. Plínio, "On the Defense of von Kempelen as the Predecessor of Experimental Phonetics and Speech Synthesis Research". The Ninth International Conference on the History of the Language Sciences. 2007. 101-106 pp.
- [2] E. David, O. Selfridge, "Eyes and Ears for Computers". Proceedings of the IEEE. Vol.50. Mayo, 1962. 1093-1101 pp.
- [3] B. Gold, N. Morgan, D. Ellis, Speech and audio signal processing: Processing and Perception of Speech and Music. 2da Edición. 2011. Editorial WILEY. 688 pp.
- [4] R. Esparza, " Cómo funciona Siri". Como funciona: Edición México. No. 4. 2014. 45 p.
- [5] C. Cristían, La voz hablada y cantada. 8va. Edición. 1994. Editorial EDAMEX. 257 pp.
- [6] G. de las Heras, L. Rodríguez, Materiales para cuidar mi voz. Fundación MAPFRE-UCLM. 44 pp.

- [7] H. Silva, Reconocimiento Automático de locutor y realización de un sistema experimental. Tesis de Maestría. Centro de Investigación Científica y de Educación Superior de Ensenada. 1994.
- [8] L. Beltrán, Simulación de modelos ocultos de Markov aplicados al reconocimiento de palabras aisladas, utilizando el programa Matlab. Tesis de Licenciatura. Escuela Politécnica Nacional: Escuela de Ingeniería. Quito. 2003.
- [9] J. Flores, Técnicas para el reconocimiento de voz en palabras aisladas en la lengua náhuatl, Tesis de Maestría. Centro de Investigación en Computación. México. D.F. 2009.
- [10] A. Ramírez, Reconocimiento automático del locutor mediante técnicas dependientes e independientes del vocabulario para un sistema acotado por el ancho de banda telefónico y realización de un sistema experimental. Tesis de Maestría. Centro de Investigación Científica y de Educación Superior de Ensenada. 1996.
- [11] L. Rabiner, B. Juang, Fundamentals of speech recognition. 1993. Editorial Prentice Hall. 507 pp.
- [12] A. Buzó, A. Gray, R. Gray, J. Markel, "Speech Coding Based Upon Vector Quantization". IEEE Transactions on Acoustics, Speech, and Signal Processing. Vol. assp-28. No. 5. Octubre de 1980. 562-574 pp.
- [13] L. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceedings of the IEEE. Vol. 77. No. 2. Febrero de 1989. 257-286 pp.
- [14] S. Prasad, T. Kishore, "Hybrid HMM/DTW based Speech Recognition with Kernel Adaptive Filtering Method". International Journal on Computational Sciences & Applications. Vol.1. No.4. Febrero de 2014. 11-21 pp.
- [15] J. Varela, J. Loaiza, Reconocimiento de palabras aisladas mediante redes neuronales sobre FPGA. Tesis de licenciatura. Facultad de Ingenierías: Eléctrica, Electrónica, Física y de Sistemas. Universidad Tecnológica de Pereira. 2008.

- [16] S. Rascón, Reconocimiento de voz para un control de acceso mediante una red neuronal de retropropagación. Tesis de licenciatura. Escuela superior de ingeniería mecánica y eléctrica Unidad Culhuacan. México. D.F. 2009.
- [17] J. Pech, Desarrollo de un sistema de reconocimiento de voz para el control de dispositivos utilizando mixturas gaussianas. Tesis de Maestría. Instituto Politécnico Nacional. Centro de Investigación en Computación. México. D.F. 2006.
- [18] J. Rodríguez, Sistema de reconocimiento del locutor basado en modelado no paramétrico. Tesis de Maestría. Instituto Politécnico Nacional. Escuela Superior de Ingeniería Mecánica y Eléctrica. México. D.F. 2008.
- [19] G. Pérez, Herramientas de Segmentación y Evaluación de Series Temporales Basadas en Modelos Ocultos de Markov. Tesis de Licenciatura. Universidad Carlos III de Madrid. España. Madrid. 2010.
- [20] J. Schmidhuber, "Deep learning in neural networks: An overview". *Neural Networks*. No. 61. Enero de 2015. 85-117 pp.
- [21] A. Abad et al., "Automatic word naming recognition for an on-line aphasia treatment system". *Computer Speech and Language*. No. 27. Septiembre de 2013. 1235-1248 pp.
- [22] A. Hussen, S. Zeiler, D. Kolossa, "Learning Dynamic Stream Weights For Coupled-HMM-Based Audio-Visual Speech Recognitio ". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. Vol. 23. No. 5. Mayo 2015. 863-876 pp.
- [23] P. Cardinal, P. Dumouchel, G. Boulianne, "Large Vocabulary Speech Recognition on Parallel Architectures". *IEE Transactions on audio, speech, and lenguaje processing*. Vol.21. No. 11. Noviembre de 2013. 2290-2300 pp.

7. Autores

Ph.D. José Ismael de la Rosa Vargas es ingeniero en Comunicaciones y Electrónica egresado de la Universidad Autónoma de Zacatecas en el año de 1995. Obtuvo el grado de Maestro en Ciencias con especialidad en Sistemas Digitales en el área de

Procesamiento Digital de Señales (PDS) en mayo de 1998 por parte del Centro de Investigación y Desarrollo de Tecnología Digital (CITEDI) del Instituto Politécnico Nacional situado en Tijuana, Baja California. Posteriormente obtiene el grado de Doctor en Ciencias con especialidad en Procesamiento de Señales y Control (noviembre de 2002), por parte de la Universidad Paris Sud (XI) y de la Escuela Superior de Electricidad (SUPELEC) al sur de Paris (Gif-sur-Yvette), Francia. Trabaja actualmente en procesamiento de imágenes y voz, métodos estocásticos en problemas inversos e instrumentación.

Ing. Ángel David Pedroza Ramírez es Ingeniero en Comunicaciones y Electrónica por la Universidad Autónoma de Zacatecas. Actualmente se encuentra cursando el último semestre de la Maestría en Ciencias de la Ingeniería con especialidad en Procesamiento de Señales y Mecatrónica en la misma institución. Su línea de investigación actual es en reconocimiento de voz enfocada en pruebas de dicción mediante Modelos Ocultos de Markov.

M.C. Ernesto García Domínguez es ingeniero en Comunicaciones y Electrónica egresado de la Universidad Autónoma de Zacatecas en el año de 1989. Obtuvo el grado de Maestro en Ciencias con especialidad en Electrónica y Telecomunicaciones (área de Instrumentación) en julio de 1993 por parte del Centro de Investigación y de Educación Superior de Ensenada (CICESE), en Baja California.

Deconvolución óptima de mediciones de fluorescencia

Gerardo Ibarra Vázquez

UASLP, Facultad de Ciencias, Av. Salvador Nava s/n., C.P. 78290, SLP, Teléfono: 52 (444) 8262491
gerardo.ibarra.v@gmail.com

Daniel u. campos Delgado

UASLP, Facultad de Ciencias, Av. Salvador Nava s/n., C.P. 78290, SLP, Teléfono: 52 (444) 8262491
ducd@fciencias.uaslp.mx

Edgar R. Arce Santana

UASLP, Facultad de Ciencias, Av. Salvador Nava s/n., C.P. 78290, SLP, Teléfono: 52 (444) 8262491
arce@fciencias.uaslp.mx

Javier A. Jo

Texas A&M University, Biomedical Engineering Department, College Station, TX, EUA
javierjo@bme.tamu.edu

Resumen

El análisis de la respuesta al impulso de fluorescencia usando componentes exponenciales provee de información que permite caracterizar cuantitativamente sistemas biológicos Sin embargo, usando métodos tradicionales de deconvolución se tiene la problemática de que la respuesta del instrumento o entrada del sistema puede no estar disponible o tiene que ser medida fuera de línea, lo que provoca un problema de sincronización. Por lo que en este trabajo, se propone una metodología iterativa para resolver el problema de deconvolución de forma ciega, es decir estimando simultáneamente la respuesta del instrumento y las respuestas al impulso fluorescentes de forma simultánea, lo cual realizamos partiendo de un conjunto de mediciones de

imagenología microscópica de fluorescencia de tiempo de vida. En nuestra formulación se emplea una base conformada por funciones de Laguerre para expandir la respuesta al impulso fluorescente. El método de deconvolución ciega propuesto es formulado como un problema de aproximación cuadrática. Además, debido a la naturaleza no lineal en el proceso de estimación, se opta por un esquema de mínimos cuadrados alternados que resuelve iterativamente el problema de aproximación. En este trabajo se presenta la paralelización del algoritmo y los resultados de validación con bases de datos de placas de arterosclerosis, células mamarias con cancer y tejido oral.

Palabra(s) Clave(s): deconvolución, base de Laguerre, fluorescencia, mínimos cuadrados no negativos, optimización.

1. Introducción

Durante los últimos 20 años ha habido un crecimiento notable en el uso de la fluorescencia en las ciencias biológicas. La fluorescencia es una herramienta muy utilizada en las ciencias biológicas y clínicas para la caracterización de muestras de tejido. Una de las técnicas más recientes son las imágenes de microscopía de tiempo de vida de fluorescencia multi-espectrales (m-FLIM, por sus siglas en inglés) [1]. Esta tecnología captura la respuesta fluorescente a la excitación de un láser en diferentes canales espectrales (diferentes longitudes de onda), con el fin de medir la emisión óptica de fluoróforos sintéticos o endógenos. A fin de proporcionar evaluaciones cuantitativas, las imágenes m-FLIM son procesadas para extraer la respuesta de fluorescencia intrínseca (respuesta al impulso) de la muestra del tejido, en la cual el tiempo de vida es utilizado para proveer una propiedad cuantitativa usada para la caracterización de tejidos [2]. Por lo tanto las imágenes m-FLIM, requieren de la recuperación fiable (deconvolución) de la respuesta al impulso fluorescente de las mediciones, por lo que numerosas herramientas matemáticas se han diseñado para la recuperación de esta respuesta [2,3,4,5]. El potencial de las imágenes de fluorescencia se basa en que proveen de información temprana y no invasiva para el diagnóstico de

diferentes patologías como enfermedades cardiovasculares y dermatológicas, pre-cáncer oral y displasia, entre otras [6,7,8,9].

En la literatura se pueden observar dos tendencias principales en los algoritmos de deconvolución que dependen de la estructura de la respuesta al impulso fluorescente. Para la primera se asume una combinación lineal de funciones exponenciales que caracteriza a cada uno de los fluoróforos en la muestra [1,10,11]. Entonces, para esta aproximación tenemos que nuestros parámetros a buscar son los tiempos característicos de las funciones exponenciales y sus coeficientes de escalamiento. Por otro lado, en la segunda tendencia, tenemos que nuestra aproximación de la deconvolución se considera como una combinación lineal de la base discreta de Laguerre para la respuesta al impulso fluorescente [4].

Por una parte cuando modelamos la respuesta al impulso como una combinación lineal de funciones exponenciales, tenemos que los coeficientes de escalamiento son variables lineales pero los tiempos característicos de los exponenciales tienen una dependencia no lineal [12]. Por lo tanto, una aproximación no lineal se formula para calcular estas variables y minimizar el error de aproximación, que puede ser resuelto iterativamente mediante mínimos cuadrados no lineales para cada punto espacial [13]. En el otro caso, en donde se usa la base de las funciones de Laguerre, tenemos que los coeficientes de escalamiento de la base son variables lineales, por lo que una estimación usando mínimos cuadrados puede utilizarse. Una desventaja de las bases de Laguerre es que para algunos casos, la respuesta al impulso fluorescente tal vez no tenga un decaimiento monotónico. Entonces, durante la optimización cuadrática se deben incluir restricciones sobre la segunda o tercer derivada de la respuesta al impulso fluorescente resultante [3].

Una de las desventajas de los algoritmos de deconvolución que se han revisado en la literatura es que requieren de la medición de la respuesta del instrumento (láser de excitación) para desarrollar el cálculo. Además, para poder desarrollarse el proceso de deconvolución esta respuesta tiene que estar cuidadosamente alineada con la medición

de fluorescencia a fin de evitar tiempos muertos, y también puede sesgar las estimaciones de los tiempos de vida. Dada esta desventaja en los algoritmos de deconvolución tradicionales, en este trabajo se propone un algoritmo el cual no requiere del conocimiento de la respuesta del instrumento. Al mismo tiempo, este algoritmo propone estimar la respuesta al impulso fluorescente mediante funciones base para cada punto espacial, y la respuesta del instrumento con una perspectiva global, con lo que asumimos que la respuesta al instrumento es la misma para cada punto. Debido a que la función de costo de nuestra aproximación tiene una dependencia bilineal en nuestras variables de decisión, usamos una metodología de mínimos cuadrados alternados que iterativamente resuelve los coeficientes de escalamiento para nuestras funciones base y las muestras de la respuesta del instrumento.

2. Desarrollo

Los datos de las imágenes m-FLIM se modelan asumiendo que las mediciones de los decaimientos de fluorescencia son resultado de la convolución de la respuesta del instrumento y la respuesta al impulso fluorescente en cada punto espacial de la muestra. Además tenemos que los datos son discretos en el tiempo, y son muestreados con un periodo T sobre el dominio espacial de K puntos en el conjunto de datos [10,11]. Entonces, considerando que tenemos una ventana de tiempo de L muestras y una respuesta causal [14], el modelo de observación para la l -ésima muestra de tiempo y para el k -ésimo punto espacial está dado por la siguiente ecuación:

$$y_k[l] = u[l] \star h_k[l] + v_k[l] \quad (1)$$

$$y_k[l] = \sum_{j=0}^{L-1} u[l-j]h_k[j] + v_k[l] \quad \forall l \in [0, L-1], k \in [0, K-1] \quad (2)$$

donde $y_k[l]$, $u[l]$ y $h_k[l]$ denotan las muestras de las mediciones del decaimiento de fluorescencia, la respuesta del instrumento y la respuesta al impulso fluorescente, respectivamente, \star indica la operación de convolución, y $v_k[l]$ representa el ruido aleatorio relacionado a la instrumentación electrónica/óptica o en la medición. En

términos generales, la deconvolución es un problema inverso que puede tener soluciones múltiples [10,11]. Entonces, para restringir el espacio de búsqueda y tener un problema que se pueda manejar, en nuestra formulación, tenemos 2 hipótesis clave:

1. La respuesta al instrumento $u[l]$ es común para todos los K puntos espaciales en el conjunto de datos, y sus muestras son no-negativas y la suma de estas está normalizada a uno:

$$\sum_{l=0}^{L-1} u[l] = 1 \quad \& \quad u[l] \geq 0 \quad \forall l \in [0, L - 1] \quad (3)$$

2. La respuesta al impulso fluorescente $h_k[l]$ para cada punto espacial k e instante de tiempo l puede ser representada por una combinación lineal de N funciones bases en tiempo discreto $\{b_n[l]\}_{n=0}^{N-1}$, también común para todos los puntos espaciales.

Estas suposiciones nos permitirán tanto restringir la búsqueda de las posibles respuestas al instrumento y respuestas al impulso fluorescentes, y evitar problemas en los escalamientos numéricos, así como, formular las estimaciones como problemas de optimización cuadrática con restricciones que pueden ser resueltos eficientemente por algoritmos numéricos de optimización (punto interior, conjunto activo, etc.) [13]. Para comenzar con la explicación escribiremos nuestro modelo de observación de la convolución en notación vectorial asumiendo condiciones iniciales cero:

$$\begin{bmatrix} y_k[0] \\ y_k[1] \\ \vdots \\ y_k[L-1] \end{bmatrix} = \begin{bmatrix} u[0] & 0 & \cdots & 0 \\ u[1] & u[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u[L-1] & u[L-2] & \cdots & u[0] \end{bmatrix} \begin{bmatrix} h_k[0] \\ h_k[1] \\ \vdots \\ h_k[L-1] \end{bmatrix} + \begin{bmatrix} v_k[0] \\ v_k[1] \\ \vdots \\ v_k[L-1] \end{bmatrix}, \quad (4)$$

$$\mathbf{y}_k = \mathbf{U} \mathbf{h}_k + \mathbf{v}_k$$

donde tenemos como resultado que la matriz de entrada \mathbf{U} tiene una estructura toeplitz [15]. Enseguida se considera que la respuesta al impulso fluorescente en el punto espacial k y para cada instante l es modelada como una combinación lineal de las funciones base:

$$h_k[l] = \sum_{n=0}^{N-1} c_{k,n} b_n[l] \quad \forall l \in [0, L - 1] \quad (5)$$

donde los coeficientes de escalamiento cumplen $c_{k,n} \in \mathfrak{R}$. Las funciones base que se utilizaron en este trabajo fueron las funciones base de Laguerre [2,3]. La ec. (5) se puede expresar en notación vectorial como:

$$\mathbf{h}_k = \mathbf{B} \mathbf{c}_k \quad \forall k \in [0, K - 1] \quad (6)$$

donde

$$\mathbf{B} = \begin{bmatrix} b_0[0] & b_1[0] & \cdots & b_{N-1}[0] \\ b_0[1] & b_1[1] & \cdots & b_{N-1}[1] \\ \vdots & \vdots & \ddots & \vdots \\ b_0[L-1] & b_1[L-1] & \cdots & b_{N-1}[L-1] \end{bmatrix} \in \mathfrak{R}^{L \times N} \quad (7)$$

$$\mathbf{c}_k = [c_{k,0} \quad \cdots \quad c_{k,N-1}]^T \in \mathfrak{R}^N \quad (8)$$

con las funciones base de Laguerre definidas como [2,3,4]:

$$b_n[l] = \alpha^{\frac{1}{2}(l-n)} \sqrt{1-\alpha} \sum_{i=0}^n (-1)^i \binom{l}{i} \binom{n}{i} \alpha^{n-i} (1-\alpha)^i \quad \forall n \in [0, N-1], l \in [0, L-1] \quad (9)$$

donde el parámetro libre $0 < \alpha < 1$ en las funciones de Laguerre determina la tasa de decaimiento exponencial. Cuanto mayor sea el orden n y/o cuanto mayor sea el parámetro α , más tiempo tendrá la propagación en el tiempo de las funciones base y mayor será la separación de tiempo entre el cruce por el cero. Entonces, considerando el ruido en la medición, el modelo de observación puede ser escrito en forma compacta como:

$$\mathbf{y}_k = \mathbf{U} \mathbf{B} \mathbf{c}_k + \mathbf{v}_k \quad \forall k \in [0, K - 1] \quad (10)$$

y aplicado para todas los puntos espaciales lo podemos escribir en notación matricial de la siguiente forma:

$$\mathbf{Y} = \mathbf{U} \mathbf{B} \mathbf{C} + \mathbf{V} \quad (11)$$

donde se definen las siguientes matrices

$$\mathbf{Y} = [\mathbf{y}_0 \quad \cdots \quad \mathbf{y}_{K-1}] \in \mathfrak{R}^{L \times K} \quad (12)$$

$$\mathbf{C} = [\mathbf{c}_0 \quad \cdots \quad \mathbf{c}_{N-1}] \in \mathfrak{R}^{N \times K} \quad (13)$$

$$\mathbf{V} = [\mathbf{v}_0 \quad \cdots \quad \mathbf{v}_{K-1}] \in \mathfrak{R}^{L \times K} \quad (14)$$

Entonces, tenemos que nuestro problema de deconvolución ciega se aborda como un problema de optimización [13]:

$$\min_{\mathbf{U}, \mathbf{C}} \frac{1}{2} \underbrace{\|\mathbf{Y} - \mathbf{U}\mathbf{B}\mathbf{C}\|_F^2}_J \quad (15)$$

$$\text{tal que } \|\mathbf{U}\|_\infty = 1 \ \& \ \mathbf{U} \geq 0$$

donde $\|\cdot\|_F$ y $\|\cdot\|_\infty$ nos indican la norma Frobenius y la norma infinito, respectivamente. Una ventaja importante de la formulación anterior es que la matriz de entrada \mathbf{U} tiene una estructura toeplitz [15]. Dado que la respuesta del instrumento es un pulso muy estrecho, no es necesario estimar todos los términos L , ya que muchos serán cero. Por lo tanto consideramos sólo los primeros \hat{L} términos ($\hat{L} < L$) para representar la respuesta del instrumento. De esta manera, la matriz de entrada \mathbf{U} puede parametrizarse por las variables $\{\theta_l\}$ como una combinación de \hat{L} matrices toeplitz $\mathbf{U}_l \in \mathfrak{R}^{L \times L}$

$$\mathbf{U} = \sum_{l=0}^{\hat{L}-1} \theta_l \mathbf{U}_l \quad (16)$$

por lo que el esquema de deconvolución ciega se puede volver a escribir como

$$\min_{\{\theta_l\}_{l=0}^{\hat{L}-1}, \mathbf{C}} \frac{1}{2} \left\| \mathbf{Y} - \sum_{l=0}^{\hat{L}-1} \theta_l \mathbf{U}_l \mathbf{B}\mathbf{C} \right\|_F^2 \quad (17)$$

$$\text{tal que } \sum_l \theta_l = 1 \ \& \ \theta_l \geq 0 \ \forall l \in [0, \hat{L} - 1]$$

2.1. Restricciones en la respuesta al impulso fluorescente

La estimación de la respuesta al impulso fluorescente necesita considerar condiciones de suavidad para representar una muestra biológica, por lo que se restringe nuestro modelo lineal en (5) [10,11]. Por lo tanto, la respuesta estimada necesita mostrar un patrón monótonamente decreciente en cualquier punto espacial [3]. Como resultado, la derivada temporal de la respuesta al impulso fluorescente tiene que ser negativa definida ($h'_k < 0 \forall k$), pero sin puntos de inflexión o cambios de curvatura. Así que, una restricción alternativa es considerar que la segunda derivada temporal debe ser siempre positiva definida ($h''_k > 0$) [3]. Sin embargo, resulta ser más fácil incorporar en el proceso de optimización que la tercera derivada en tiempo debe ser negativa semi-definida ($h'''_k \leq 0$). Con esto, aseguramos que la función al impulso fluorescente sea monótonamente decreciente y estrictamente convexa. Para este propósito, utilizaremos una aproximación numérica para la derivada temporal de tercer orden basada en diferencias centradas para el l -ésimo instante de tiempo y el k -ésimo punto espacial [12]:

$$h'''_k[l] = \frac{-h_k[l+3] + 8h_k[l+2] - 13h_k[l+1] + 13h_k[l-1] - 8h_k[l-2] + h_k[l-3]}{8T^3} \quad (18)$$

Entonces, usando notación vectorial, la aproximación discreta de la derivada temporal de tercer orden de la respuesta al impulso fluorescente en el k -ésimo punto espacial es:

$$\mathbf{h}'''_k = \mathbf{A} \mathbf{c}_k \in \mathfrak{R}^{L-6} \quad (19)$$

donde

$$\mathbf{A} \triangleq \frac{1}{8T^3} (-\mathbf{B}_{7:L} + 8\mathbf{B}_{6:L-1} - 13\mathbf{B}_{5:L-2} + 13\mathbf{B}_{3:L-4} - 8\mathbf{B}_{2:L-5} + \mathbf{B}_{1:L-6}) \in \mathfrak{R}^{(L-6) \times N} \quad (20)$$

y la notación $\mathbf{B}_{p:q}$ representa la matriz resultante de extraer los renglones p a q en la matriz \mathbf{B} . De esta manera siguiendo la notación vectorial, la siguiente desigualdad puede representar la restricción en la respuesta al impulso fluorescente para los coeficientes de escalamiento \mathbf{c}_k en el punto espacial k :

$$\mathbf{A}c_k \leq \mathbf{0} \quad \forall k \in [0, K - 1] \quad (21)$$

Por lo que nuestro esquema de optimización cuadrática para nuestro algoritmo de deconvolución ciega, que incluye las restricciones en las muestras de la respuesta del instrumento como en los coeficientes de escalamiento, queda formulado como:

$$\min_{\{\theta_l\}_{l=0}^{\hat{L}-1}, c} \frac{1}{2} \left\| \mathbf{Y} - \sum_{l=0}^{\hat{L}-1} \theta_l \mathbf{U}_l \mathbf{B} \mathbf{C} \right\|_F^2 \quad (22)$$

$$\text{tal que } \sum_l \theta_l = 1 \ \& \ \theta_l \geq 0 \ \forall l \in [0, \hat{L} - 1], \mathbf{A}c_k \leq \mathbf{0} \quad \forall k \in [0, K - 1].$$

2.2. Estimación de los coeficientes de escalamiento

Asumiendo que los parámetros $\{\theta_l\}_{l=0}^{\hat{L}-1}$ son dados, entonces la matriz \mathbf{U} puede ser calculada. Dado que los coeficientes de escalamiento son estimados para cada punto espacial $k \in [0, K - 1]$, una estimación local de mínimos cuadrados con restricción se formula como [5]:

$$\min_{c_k} \frac{1}{2} \|\mathbf{y}_k - \mathbf{U} \mathbf{B} c_k\|_2^2 = \min_{c_k} \frac{1}{2} c_k^T \mathbf{H} c_k - f_k^T c_k \quad (23)$$

donde $\mathbf{H} = \mathbf{B}^T \mathbf{U}^T \mathbf{U} \mathbf{B}$, $f_k = \mathbf{B}^T \mathbf{U}^T \mathbf{y}_k$ y \mathbf{A} dada por (20). Además como otra alternativa, el problema de optimización planteado se puede resolver de una manera más eficiente con su formulación dual, a través de una aproximación usando mínimos cuadrados no negativos. La dualidad del problema nos indica que un problema de optimización puede ser visto de dos perspectivas, el problema primario o el problema dual [13,6]. La idea básica de usar la función dual de Lagrange es tomar las restricciones en cuenta, mediante el aumento de la función objetivo con una suma ponderada de las funciones de restricciones, utilizando el multiplicador de Lagrange como el peso asociado para cada restricción. Dado que la función dual de Lagrange es el supremo puntual de una familia de funciones afines al multiplicador, esta es cóncava, incluso cuando el problema no es convexo. Por lo que puede ser resuelta con respecto al multiplicador de

Lagrange y enseguida se calculan los coeficientes de nuestro problema primario como se muestra a continuación [3]:

$$\lambda^* = \arg \min_{\lambda \geq 0} \frac{1}{2} \|\mathbf{R}(\mathbf{A}^\top \lambda - \mathbf{f}_k)\|_2^2 \quad (24)$$

donde $\mathbf{R}^\top \mathbf{R} = \mathbf{H}^{-1}$ se obtiene de la factorización de Cholesky y los coeficientes de escalamiento óptimos son:

$$\mathbf{c}_k = \mathbf{H}^{-1}(-\mathbf{A}^\top \lambda^* + \mathbf{f}_k). \quad (25)$$

Ya que las matrices \mathbf{H} y \mathbf{A} son constantes para cada punto espacial en el conjunto de datos, la matriz \mathbf{R} puede ser calculada una vez para toda la imagen FLIM. Sin embargo, este paso en el algoritmo de deconvolución ciega puede ser también paralelizado, a fin de reducir el tiempo de computo. Además se propone una disminución en las muestras espaciales equidistante en tamaño $D \in \mathbb{Z}$ ($D > 1$) para la matriz de mediciones de FLIM \mathbf{Y} , durante la estimación de la respuesta del instrumento y así obtener una versión reducida (en columnas) $\hat{\mathbf{Y}}$ donde $\hat{\mathbf{K}} = \lfloor K/D \rfloor$ y $\lfloor \cdot \rfloor$ denota la función piso.

2.3. Estimación de la respuesta del instrumento

Para la estimación de la respuesta del instrumento, suponemos que la matriz de coeficientes de escalamiento \mathbf{C} es conocida en la función de costo J , y entonces la optimización se realizará para los parámetros $\{\theta_l\}_{l=0}^{\hat{L}-1}$ que representan las primeras \hat{L} muestras de la respuesta del instrumento. Para este caso, se puede calcular una solución cerrada para toda la base de datos FLIM. Primero, la función de costo es aumentada para incluir la condición de normalización $\sum_l \theta_l = 1$ con un multiplicador de Lagrange [13,16]:

$$\hat{f} = \frac{1}{2} Tr \left\{ \left(\mathbf{Y} - \sum_{l=0}^{\hat{L}-1} \theta_l \mathbf{U}_l \mathbf{B} \mathbf{C} \right)^\top \left(\mathbf{Y} - \sum_{l=0}^{\hat{L}-1} \theta_l \mathbf{U}_l \mathbf{B} \mathbf{C} \right) \right\} + \mu \left(\sum_l \theta_l - 1 \right) \quad (26)$$

donde $\mu > 0$ es el multiplicador de Lagrange que está relacionado a la condición de igualdad y $Tr\{\cdot\}$ denota la traza de la matriz. Por lo tanto, las condiciones de optimalidad estacionaria son:

$$\frac{\partial \hat{f}}{\partial \theta_m} = 0 \quad \forall m \in [0, L - 1], \quad \& \quad \frac{\partial \hat{f}}{\partial \mu} = 0 \quad (27)$$

Al aplicar la función de costo en la ec. (26) a las condiciones en (27), se obtiene el siguiente par de ecuaciones:

$$\sum_{l=0}^{\hat{L}-1} \underbrace{Tr\{\mathbf{C}^T \mathbf{B}^T \mathbf{U}_m^T \mathbf{U}_l \mathbf{B} \mathbf{C}\}}_{\delta_{m,l}} + \mu = \underbrace{Tr\{\mathbf{Y}^T \mathbf{U}_m \mathbf{B} \mathbf{C}\}}_{b_m} \quad \forall m \in [0, \hat{L} - 1] \quad (28)$$

$$\sum_{l=0}^{\hat{L}-1} \theta_l = 1 \quad (29)$$

Por consiguiente, obtenemos un sistema de $\hat{L} + 1$ ecuaciones lineales y $\hat{L} + 1$ variables desconocidas:

$$\begin{bmatrix} \delta_{0,0} & \cdots & \delta_{1,\hat{L}-1} & 1 \\ \vdots & \ddots & \cdots & \vdots \\ \delta_{0,\hat{L}-1} & \cdots & \delta_{\hat{L}-1,\hat{L}-1} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_{\hat{L}-1} \\ \mu \end{bmatrix} = \begin{bmatrix} b_0 \\ \vdots \\ b_{\hat{L}-1} \\ 1 \end{bmatrix}, \quad (30)$$

la cual provee la solución óptima a los parámetros $\{\theta_l\}_{l=0}^{\hat{L}-1}$. Si al resolver la ec. (3), un parámetro viola la restricción de no negatividad, el algoritmo lo obliga a ser cero ($\theta_l = 0$) y reescala los parámetros para conservar la condición de normalización ($\sum_{l=0}^{\hat{L}-1} \theta_l = 1$).

2.4. Estimación de la deconvolución ciega

Utilizando la hipótesis planteada en la sección anterior, podemos construir el algoritmo que deconvolucione de forma ciega los datos m-FLIM, estimando las respuestas al impulso fluorescente, así como la respuesta al instrumento. Con lo que se presenta el algoritmo en forma general:

Algoritmo de deconvolución ciega

1. Inicialización y selección de un conjunto reducido de datos: Se provee la matriz de mediciones de FLIM (Y) y basado en la selección de la disminución de muestras espaciales D , se crea el conjunto de datos \hat{Y} reducido. Se define el número de muestras a ser estimadas para la respuesta del instrumento \hat{L} , su condición inicial θ^0 , el número máximo de iteraciones t_{max} , el umbral de convergencia para el error estimado κ y las funciones bases $\{b_n[l]\}_{n=0}^{N-1}$ para construir la matriz B en (7). Se pone el contador de iteraciones $t=0$ y $J^0 = 10^6$. Con el punto de inicio θ^0 , enseguida se calcula la matriz de entrada U^0 .
2. Estimación de los coeficientes de escalamiento de la base de Laguerre: Se aumenta el contador en 1, y para el conjunto reducido de puntos espaciales $k \in [0, \hat{K} - 1]$ con U^{t-1} , se calcula el algoritmo de mínimos cuadrados no negativos para c_k^t en (24) and (25). Se construye la matriz resultante de coeficientes de escalamiento C^t .
3. Estimación de la respuesta del instrumento: Se fija C^t , se evalúan los parámetros óptimos de entrada θ^t según la ec. (30) y se obtiene la matriz de entrada U^t .
4. Prueba de convergencia: Se estima el error en la iteración t como $J^t = \|Y - U^t B C^t\|_F$, se evalúa la mejora en el error con respecto de su iteración anterior $\Delta J = |J^t - J^{t-1}|$. Si $\Delta J > \kappa$ & $t < t_{max}$ regresar al paso 2, y en caso contrario continuar.
5. Estimación de los coeficientes de escalamiento para el conjunto de datos completo: Para todo el conjunto de puntos espaciales $k \in [0, K - 1]$ con U^t se calcula el algoritmo de mínimos cuadrados no negativos para c_k^t en (24) and (25).
6. Computo final de las estimaciones de la respuesta al impulso fluorescente y los decaimientos medidos de fluorescencia: Si asumimos que el algoritmo calcula

en la \hat{t} -ésima iteración las salidas $(\theta^{\hat{t}}, \mathbf{C}^{\hat{t}})$, las estimaciones finales están dadas por:

$$\hat{u}[l] = \begin{cases} \theta_l^{\hat{t}} & 0 \leq l \leq \hat{L} - 1 \\ 0 & \hat{L} \leq l \leq L - 1 \end{cases} \quad (31)$$

$$\hat{\mathbf{h}}_k = \mathbf{BC}^{\hat{t}} \Rightarrow \hat{h}_k[l] = \sum_{n=0}^{N-1} c_{k,n}^{\hat{t}} b_n[l] \quad \forall l \in [0, L - 1], k \in [0, K - 1] \quad (32)$$

$$\hat{y}_k = \mathbf{U}^{\hat{t}} \mathbf{BC}^{\hat{t}} \Rightarrow \hat{y}_k[l] = \sum_{n=0}^{N-1} c_{k,n}^{\hat{t}} \hat{u}[l] \star b_n[l] \quad (34)$$

donde $\mathbf{U}^{\hat{t}}$ se construye con $\{\theta_l^{\hat{t}}\}$.

2.5. Paralelización del algoritmo

La implementación de este algoritmo se realizó en la plataforma de MATLAB, la cual tiene capacidad de poder paralelizar el procesamiento de los programas implementados en ella. Por lo que se estudió la posibilidad de paralelizar el algoritmo, para poder realizar la estimación más eficiente y en un menor tiempo. MATLAB contiene un herramienta (*Parallel Computing Toolbox*) la cual permite al usuario resolver problemas computacionalmente intensos usando los procesadores con múltiples núcleos, GPU's y grupos de computadoras. Usando esta herramienta en el algoritmo descrito en la Sección 2.3, se paralelizó el calculo de los coeficientes de la base de Laguerre para cada punto espacial en la ec. (24), ya que inicialmente se utilizaban los ciclos *for*, los cuales fueron sustituidos por los ciclos *parfor*. Al realizar este cambio lo que se hace es correr cada una de las iteraciones del ciclo en uno de los núcleos del procesador de la computadora. También se configuraron las preferencias de la función del método de mínimos cuadrados no negativos para que se ejecutase en paralelo. No se optó por utilizar GPU's dado a que la herramienta tiene soporte a un número reducido de modelos.

3. Resultados

Para la validación del algoritmo se probó con imágenes m-FLIM de 3 tipos: tejido de placas coronarias con arterosclerosis, células de cáncer de mama y tejido oral. Para cuantificar el desempeño del método de deconvolución ciega, se comparó la respuesta al impulso estimada de este algoritmo con nuestra referencia generada a partir de la deconvolución usando la respuesta del instrumento medida. Se analizaron para los tejidos de placas coronarias con arterosclerosis 60 imágenes, con 170 muestras temporales, 60x60 muestras espaciales y un tiempo de muestreo de 250 ns; para las imágenes de células de cáncer de mama se analizaron 13 imágenes, con 170 muestras temporales, 256x256 muestras espaciales y un tiempo de muestreo de 48.867 ns; y para los datos de tejido oral fueron 13 imágenes, con 200 muestras temporales, 160x160 muestras espaciales y un tiempo de muestreo de 160 ns. También para todas las pruebas, se obtuvieron los resultados del algoritmo de deconvolución ciega con el factor de disminución de muestras de $D=8$ y $D=36$.

Los resultados de la validación y de los tiempos de cómputo se muestran a continuación en las Figs. 1 a 3 y Tabla 1. Para poder mostrar la respuesta de fluorescencia intrínseca se calcula el tiempo de vida promedio para cada una de las muestras espaciales y se muestra la imagen. Este tiempo de vida promedio puede ser calculado con la siguiente formula:

$$\tau_k = \frac{\mathbf{t}^T \hat{\mathbf{h}}_k}{\mathbf{1}_L^T \hat{\mathbf{h}}_k} \quad \forall k \in [0, K - 1] \quad (35)$$

donde τ_k denota el tiempo de vida promedio para la k -ésima muestra, $\hat{\mathbf{h}}_k \in \mathbb{R}^L$ representa la aproximación de respuesta al impulso fluorescente, y $\mathbf{t} \in \mathbb{R}^L$ el vector de muestras de tiempo. Se escogió una imagen dentro del juego de datos analizados para cada método y así poder observar gráficamente los resultados. A continuación se muestran estas imágenes en las Figs. 1, 2 y 3, en las cuales se observa el método de deconvolución con la respuesta del instrumento medida (DEMIR)

y el método de deconvolución ciega (BDE) con factor de disminución de muestras (D=8 y D=36).

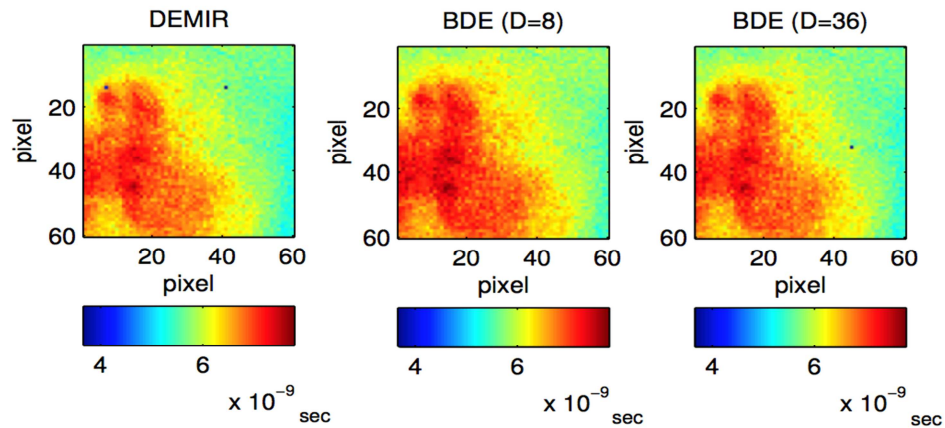


Fig. 1. Comparativo de resultados en los tiempos de vida promedio de la respuesta al impulso para tejido de placas coronarias con arterosclerosis en la imagen 16.

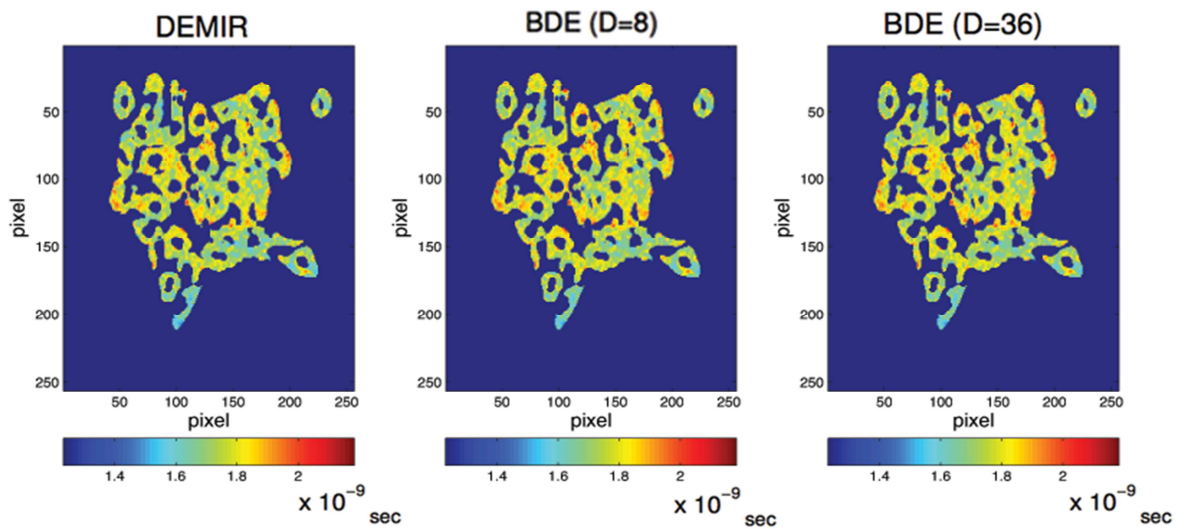


Fig. 2. Comparativo de resultados en los tiempos de vida promedio de la respuesta al impulso para tejido con células de cáncer de mama en la imagen 4.

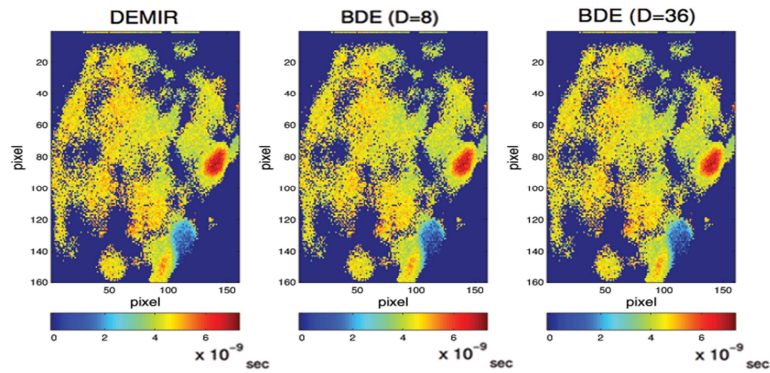


Fig. 3. Comparativo de resultados en los tiempos de vida promedio de la respuesta al impulso para tejido oral en la imagen 19.

Para poder evaluar la exactitud de la metodología se uso una métrica sobre todo el conjunto de datos, y considerando como el tiempo de vida promedio por DEMIR como nuestra referencia, el porcentaje de error en la estimación por el algoritmo de deconvolución ciega para toda la imagen se definió como:

$$E_l^{BDE} = \frac{1}{K} \sum_{k=0}^{K-1} \frac{|\tau_k^{DEMIR} - \tau_k^{BDE}|}{\tau_k^{DEMIR}} \tag{36}$$

En cuanto el tiempo de ejecución para cada uno de estos algoritmos se obtuvieron los siguientes valores que se muestran en la Tabla 1.

| | | Tiempo de Procesamiento (Segundos) | | | | | |
|--|---------------------|------------------------------------|---------|---------|--------------|---------|---------|
| | | Normal | | | Paralelizado | | |
| | | DEMIR | BDE | | DEMIR | BDE | |
| | | | D=8 | D=36 | | D=8 | D=36 |
| Tejido de placas coronarias con arteriosclerosis | Media | 1.4151 | 14.0877 | 12.1621 | 0.7203 | 4.566 | 4.214 |
| | Desviación estándar | 0.2465 | 6.5013 | 4.7898 | 0.1604 | 1.172 | 1.3493 |
| Células cancerígenas de mama | Media | 38.28 | 53.52 | 48.8687 | 13.9566 | 18.1065 | 14.5015 |
| | Desviación estándar | 7.5377 | 10.27 | 7.9506 | 2.1186 | 2.6843 | 2.2352 |
| Tejido oral | Media | 42.4477 | 57.4056 | 44.7883 | 13.594 | 18.4836 | 14.7639 |
| | Desviación estándar | 12.7222 | 17.6769 | 13.9839 | 3.2362 | 4.9334 | 4.2446 |

Tabla 1. Comparativo de resultados en los tiempos de procesamiento del algoritmo de deconvolución ciega.

4. Discusión

Como se podrá visualizar en las imágenes mostradas en las Figs. 1 a 3, se resalta la similitud entre nuestra referencia, la deconvolución usando la respuesta del instrumento medida (DEMIR) y el método de deconvolución ciega (BDE). Además teniendo en cuenta los resultados de la métrica del error en el cálculo del tiempo de vida promedio en (36) se observaron valores promedio para el tejido de placas coronarias con arterosclerosis de 3.4% y 4.0% para el BDE ($D=8$) y BDE ($D=36$), respectivamente. Para los datos de tejido con células de cáncer de mama se observó un error de 4.2% y 4.6% y para tejido oral 4.5% y 4.9%. Confirmando estos números que nuestros tiempos de vida promedio calculados con el algoritmo de deconvolución ciega (BDE) se encuentran apegados a los valores estimados con la medición de la respuesta del instrumento.

Mediante la paralelización del algoritmo se pudo reducir el tiempo de procesamiento del algoritmo en promedio para la deconvolución usando la respuesta del instrumento medida (DEMIR) aproximadamente en 40% del tiempo original. Mientras para los dos casos del algoritmo de deconvolución ciega (BDE) con el factor de disminución de muestras de 8 y 36 se redujo en promedio a aproximadamente 33% y 32% del tiempo original, respectivamente. Todas las mediciones del tiempo de ejecución fueron tomadas en la misma computadora con características de procesador Intel 2.3 GHz Quad-core I7-3820QM y 8 Gb de RAM.

5. Conclusiones

La propuesta del método de deconvolución ciega que se presenta en este trabajo, ayuda a resolver el problema de deconvolución para los datos FLIM de una manera novedosa, donde se modela la respuesta al impulso fluorescente como una combinación lineal de funciones base para cada punto espacial. Este algoritmo se encarga de estimar tanto las muestras de la respuesta del instrumento con una perspectiva global, así como los coeficientes de escalamiento de nuestras funciones

base localmente en cada punto espacial. La base matemática de este algoritmo se centra en la metodología de mínimos cuadrados alternados que iterativamente estima nuestras variables de decisión. Para el proceso de estimación de ambas variables se utilizan métodos como optimización cuadrática con restricciones para los coeficientes de escalamiento, y mínimos cuadrados para la respuesta del instrumento.

Para acelerar el procesamiento de la información se reduce el conjunto de mediciones hasta alcanzar la convergencia. Después, el método de deconvolución ciega provee la estimación final de la respuesta al instrumento y calcula las respuestas al impulso fluorescentes en todos los puntos espaciales. Al final en el análisis de los resultados, se puede observar que el método ofrece buenos resultados en la precisión de las estimaciones, así como la paralelización del algoritmo que ayuda a reducir el tiempo distribuyendo las operaciones en los núcleos del procesador sin modificar su respuesta.

6. Referencias

- [1] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. 3rd ed. 2006. Springer Press. New York.
- [2] P. Pande, J. A. Jo. "Automated analysis of fluorescence lifetime imaging microscopy (FLIM) data based on the Laguerre deconvolution method". *IEEE Transactions on Biomedical Engineering*. Vol. 58. No. 1. 2011. 172—181 pp.
- [3] J. Liu, Y. Sun, J. Qi, L. Marcu, "A novel method for fast and robust estimation of fluorescence decay dynamics using constrained least-squares deconvolution with Laguerre expansion". *Phys. Med. Biol.* Vol. 57. 2012. 843-865 pp.
- [4] J. A. Jo, Q. Fang, T. Papaioannou, L. Marcu, "Fast model-free deconvolution of fluorescence decay for analysis of biological systems". *Journal of Biomedical Optics*. Vol. 9. No. 4. 2004. 743-752 pp.

- [5] W. R. Ware, L. J. Doemeny, T. L. Nemzek, "Deconvolution of fluorescence and phosphorescence decay curves. A least square method". *J. Phys. Chem.* Vol. 77, 1973. 2038-2048 pp.
- [6] I. B. Berlman, *Handbook of fluorescence spectra of aromatic molecules*. 2nd ed. 1971. Academic Press. New York.
- [7] M. O'Donnell, E. R. McVeigh, H. W. Strauss, A. Tanaka, B. E. Bouma, G. J. Tearney, M. A. Guttman, E. V. Garcia, "Multimodality cardiovascular molecular imaging technology". *Journal of Nuclear Medicine*. Vol. 51. No. 1. 2010. 38S–50S pp.
- [8] V. De Giorgi, D. Massi, S. Sestini, R. Cicchi, F. S. Pavone, T. Lotti, "Combined non-linear laser imaging (two photon excitation fluorescence microscopy, fluorescence lifetime imaging microscopy, multispectral multiphoton microscopy) in cutaneous tumours: first experiences". *Journal of the European Academy of Dermatology and Venereology*. Vol. 23. No. 3. 2009. 314–316 pp.
- [9] N. Galletly, J. McGinty, C. Dunsby, F. Teixeira, J. Requejo-Isidro, I. Munro, D. Elson, M. Neil, A. Chu, P. French, G. Stamp, "Fluorescence lifetime imaging distinguishes basal cell carcinoma from surrounding uninvolved skin". *British Journal of Dermatology*. Vol. 159. No. 1. 2008. 152–161 pp.
- [10] A. Periasamy, R.M. Clegg, *FLIM Microscopy in Biology and Medicine*. 2009. Chapman and Hall. CRC.
- [11] L. Marcu, P.M.W. French, D.S. Elson, *Fluorescence Lifetime Spectroscopy and Imaging: Principles and Applications in Biomedical Diagnostics*. 2014. CRC Press.
- [12] J. H. Mathews, K. K. Fink, *Numerical Methods Using Matlab*. 4th Edition. 2004. Prentice-Hall. Upper Saddle River.

- [13] J. Nocedal, S. J. Wright, Numerical Optimization. 2nd Edition. 2006. Springer-Verlag.
- [14] J. G. Proakis, D. G. Manolakis, Digital Signal Processing. 4th Edition. 2006. Prentice Hall.
- [15] R. A. Horn, C. R. Johnson, Matrix Analysis. 1985. Cambridge University Press.
- [16] S. Boyd, L. Vandenberghe, Convex Optimization. 2004. Cambridge University Press.

7. Autores

Gerardo Ibarra Vazquez es actualmente alumno de la maestría en Ingeniería Electrónica de la Universidad Autónoma de San Luis Potosí (UASLP), y obtuvo el grado de Ingeniero Electrónico también de la UASLP. Su trabajo de investigación se centra en métodos numéricos para la caracterización de tejido utilizando imágenes microscópicas de tiempo de vida de fluorescencia.

Daniel U. Campos Delgado es profesor de tiempo completo en la UASLP desde 2001. Recibió su título de Ingeniero en Electrónica en la UASLP en 1996. Sus estudios de Maestría y Doctorado los realizó en Louisiana State University (LSU), Baton Rouge, LA, EUA, en 1999 y 2001, respectivamente. Es autor de más de 150 artículos en revistas científicas y congresos. Sus áreas de interés incluyen métodos de optimización, modelado de sistemas dinámicos y procesamiento de señales.

Edgar R. Arce Santana es profesor en la UASLP en México. Recibió su grado de licenciatura en ciencias de la computación en el Instituto Tecnológico de San Luis Potosí en 1987. Obtuvo sus grados de Maestría y Doctorado en el CIMAT en Guanajuato en 2000 y 2004, respectivamente. Es autor de más de 50 artículos publicados en revistas y congresos científicos. Sus áreas de interés incluyen la visión computacional, procesamiento de señales y reconocimiento de patrones.

Javier Jo es profesor en el Biomedical Engineering Department en la Universidad de Texas A&M. Recibió su grado de Ingeniero Electrónico de la Pontificia Universidad Católica del Perú en 1996. Obtuvo sus grados de Maestría y Doctorado de la University of Southern California, LA, EUA en 2000 y 2002, respectivamente. Sus principales áreas de interés incluyen análisis de sistemas y procesamiento de imágenes, así como instrumentación biomédica con aplicaciones en biofotónica, fisiología y medicina.

Descomposición de datos multi-espectrales: interfaz gráfica para Matlab

Omar Gutiérrez Navarro

Universidad Autónoma de San Luis Potosí, Facultad de Ciencias, S.L.P., México
omargn@fc.uaslp.mx

Daniel U. Campos Delgado

Universidad Autónoma de San Luis Potosí, Facultad de Ciencias, S.L.P., México
ducd@fciencias.uaslp.mx

Edgar R. Arce Santana

Universidad Autónoma de San Luis Potosí, Facultad de Ciencias, S.L.P., México
arce@fciencias.uaslp.mx

Javier A. Jo

Texas A&M University, College Station, TX, USA
javierjo@bme.tamu.edu

Resumen

Avances recientes han permitido el desarrollo de dispositivos capaces de capturar información en múltiples longitudes de onda. Estos datos tienen diversas aplicaciones con el problema en común de cómo interpretarlos. Una de las técnicas utilizadas con este fin es la descomposición espectral, que separa los datos de una muestra en sus componentes básicos y concentraciones proporcionales. Nuestro trabajo previo ha estado enfocado en la descomposición espectral de datos de fluorescencia multi-espectral, donde se han desarrollado métodos que proporcionan una solución cuantitativa, robusta y rápida, la cual no está limitada por el número de componentes

que se pueden caracterizar. En este trabajo, presentamos una interface desarrollada en Matlab que puede estimar los perfiles característicos de los componentes constituyentes de una muestra y sus abundancias. En caso de que no se tenga información alguna sobre la muestra, nos permite obtener además el número de componentes en ella. El artículo hace una descripción del software y sus herramientas. Además, se ejemplifica su uso en la caracterización de muestras ex-vivo de arterias coronarias. El programa se encuentra disponible de manera gratuita y provee al usuario de una herramienta fácil de usar para el análisis de datos multi o hiper-espectrales.

Palabra(s) Clave(s): descomposición ciega, fluorescencia endógena, interfaz gráfica, optimización cuadrática, quimiometría.

1. Introducción

Los datos multi-espectrales proveen gran cantidad de información en diferentes aplicaciones, desde teledetección remota [1] hasta aplicaciones médicas [2]. Una aplicación reciente es el uso de datos multi-espectrales de fluorescencia para facilitar la descripción bioquímica de tejido vivo [3], en función de sus fluoróforos endógenos y de manera mínimamente invasiva. Sin embargo, el cómo interpretar los datos multi-espectrales es un problema abierto. Un acercamiento consiste en considerar cada medición fluorescente como una mezcla de los fluoróforos contenidos dentro de la muestra [4]. De esta manera, se deben resolver tres problemas para poder caracterizar una muestra: estimar el número de componentes constituyentes, sus perfiles característicos y sus concentraciones proporcionales. Existen varias propuestas en el estado del arte, sin embargo casi todas tienen limitantes considerables como la necesidad de información a priori, sensibilidad a datos con gran similaridad o un límite sobre el máximo número de componentes que pueden caracterizar. Pocos métodos están diseñados para trabajar con datos de fluorescencia multi-espectral, pues la mayoría están hechos pensando en datos de satélite para teledetección remota [5] y

son adaptados sin aprovechar las características particulares de las emisiones fluorescentes.

Nuestro trabajo, previamente publicado en [6,7], ha estado enfocado en el desarrollo de métodos de descomposición lineal para imágenes microscópicas de tiempo de vida de fluorescencia multi-espectral, o m-FLIM por sus siglas en inglés. Dichos métodos han probado ser eficientes para calcular los perfiles característicos de los fluoróforos endógenos de una muestra y su concentración relativa; incluso es posible estimar cuantos componentes se encuentran dentro de la muestra [7], lo cual permite obtener resultados más precisos.

En este contexto, el objetivo de este trabajo es presentar el desarrollo de una interfaz gráfica para los métodos de descomposición lineal cuadrática ciega. Concretamente, se trata de un programa que facilita el acceso a los métodos en [6,7] para ser aplicados en diferentes datos multi-espectrales, no necesariamente de tipo m-FLIM. Esta primera versión del programa está hecha en inglés dentro de Matlab, y ha sido validada con la caracterización de datos ex-vivo. El programa y algunos datos de prueba se encuentran disponibles de manera gratuita en: <http://galia.fc.uaslp.mx/~omargn/mflim.html>.

El resto del trabajo se organiza de la siguiente manera. El método de la mezcla lineal es descrito en la Sección 2. En la Sección 3 se da una breve descripción de los métodos utilizados por el programa. Después, se detalla el funcionamiento general de la interface en la Sección 4. En la Sección 5 se describe la validación con datos m-FLIM de arterias coronarias. Finalmente, los comentarios finales se encuentran en la Sección 6.

2. Modelo de la mezcla lineal

De manera general, se considera que cada vector de medición $\mathbf{y}_k \in \mathbb{R}^L$ contiene información multi-espectral en la posición k en diferentes longitudes de onda. El objetivo de la descomposición es poder describir cada medición en función de sus componentes

constituyentes. El modelo utilizado nos dice que cada medición \mathbf{y}_k es resultado de una combinación lineal

$$\mathbf{y}_k = \sum_{n=1}^N \alpha_{k,n} \mathbf{p}_n \quad (1)$$

donde N es el número de perfiles característicos $(\mathbf{p}_1, \dots, \mathbf{p}_N)$. Los coeficientes $(\alpha_{k,1}, \dots, \alpha_{k,N})$ representan las abundancias o concentraciones proporcionales, las cuales deben cumplir con $\alpha_{k,n} \geq 0$ y $\sum_{n=1}^N \alpha_{k,n} = 1$ para poder obtener resultados con significado físico. En el caso de mediciones de fluorescencia multi-espectral, los perfiles característicos $(\mathbf{p}_1, \dots, \mathbf{p}_N)$ también están sujetos a restricciones similares [7]. Los métodos utilizados asumen que los perfiles característicos, así como el número de estos, son constantes en toda la base de datos. Mientras que las abundancias reflejan la contribución proporcional de cada perfil característico en diferentes posiciones.

Para considerar múltiples mediciones $k \in [1, \dots, K]$ correspondientes a una muestra de tejido, agrupamos las muestras dentro de una matriz $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K] \in \mathbb{R}^{L \times K}$. De igual manera, se agrupan los perfiles característicos en $\mathbf{P}^N = [\mathbf{p}_1, \dots, \mathbf{p}_N] \in \mathbb{R}^{L \times N}$ así como las abundancias dentro de la matriz $\mathcal{A}^N = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K] \in \mathbb{R}^{N \times K}$, donde cada vector de concentraciones está expresado por $\boldsymbol{\alpha}_k = [\alpha_{k,1}, \dots, \alpha_{k,N}]^T$ con $k \in [1, K]$ y $(\cdot)^T$ denota el operador transpuesto. Considerando todas las muestras disponibles, el modelo de la mezcla lineal a resolver es:

$$\mathbf{Y} = \mathbf{P}^N \mathcal{A}^N \quad (2)$$

Además, se deben tomar en cuenta las restricciones sobre los perfiles y abundancias:

$$\mathbf{1}_N^T \mathcal{A}^N = \mathbf{1}_N^T \quad (3)$$

$$\mathcal{A}^N \geq 0 \quad (4)$$

$$\mathbf{1}_N^T \mathbf{P}^N = \mathbf{1}_N^T \quad (5)$$

$$\mathbf{P}^N \geq 0 \quad (6)$$

donde $\mathbf{1}$ representan vectores de elementos unitarios de longitud N , y \succeq es el operador de desigualdad componente a componente. Para poder obtener una descripción cuantitativa se debe resolver el problema inverso, donde solo se conocen los datos de entrada \mathbf{Y} , mientras se busca estimar cuantos componentes N están presentes en la muestra, así como sus perfiles \mathbf{P} y abundancias \mathbf{A} .

3.- Métodos de descomposición

Para poder obtener una solución bajo el modelo de la mezcla lineal en (2) sujeta a las restricciones (3) a (6), el programa utiliza los métodos BEAE [6] y una modificación del método publicado en [7], al cual haremos referencia como QBLU (descomposición lineal cuadrática ciega, por sus siglas en inglés).

3.1. Extracción ciega de perfiles y abundancias (BEAE)

Este método se emplea para poder descomponer los datos de entrada en sus perfiles característicos \mathbf{P} y abundancias \mathbf{A} , para lo cual es necesario proveer el número de componentes N así como un conjunto inicial de perfiles característicos $(\mathbf{P})^0$. La estimación de las abundancias y perfiles se hace de manera alternada. Las abundancias se calculan resolviendo un problema de optimización cuadrática [8]:

$$\min_{\mathbf{A}^N \succeq 0, \mathbf{1}_N^\top \mathbf{A}^N = \mathbf{1}_N^\top} \frac{1}{2} \|\mathbf{Y} - \mathbf{P}^N \mathbf{A}^N\|_F^2 \quad (7)$$

donde el operador $\|\cdot\|_F$ denota la norma de Frobenius para matrices. Para estimar los perfiles, se utiliza la aproximación de las abundancias y la estrategia propuesta en [1,5], la cual consiste en buscar los perfiles característicos como los vértices del simplejo de menor volumen que contiene a todas las muestras observadas. Para ello, se busca minimizar la siguiente función de costo cuadrática, sujeta a las restricciones (5,6):

$$J(\mathbf{P}^N, \mathbf{A}^N) = \frac{1}{2} \|\mathbf{Y} - \mathbf{P}^N \mathbf{A}^N\|_F^2 + \rho \sum_{m=1}^{N-1} \sum_{n=m+1}^N \|\mathbf{p}_m - \mathbf{p}_n\|_2^2 \quad (8)$$

donde $\rho > 0$ representa un parámetro de regularización. El funcionamiento general del algoritmo BEAE se describe en el diagrama de la Figura 1.

3.2. Descomposición lineal cuadrática ciega (QBLU)

La estimación del número de componentes se basa en el algoritmo publicado en [7]. Tiene la ventaja de poder obtener una descomposición totalmente ciega, pues puede calcular el número de componentes presentes en la muestra. El método comienza con un conjunto inicial de perfiles candidatos $(\mathbf{P}^N)^0$ en la iteración $t = 0$ y se calcula $\{\mathbf{P}^N, \mathbf{A}^N\}$ utilizando BEAE para el número actual de componentes N . La solución es evaluada bajo dos criterios por medio de detectores de Neyman Pearson [9]. El primero verifica la independencia lineal entre los perfiles candidatos en \mathbf{P}^N , mientras que el segundo evalúa que las abundancias \mathbf{A}^N tengan una contribución significativa dentro de la muestra. Cuando un perfil candidato no pasa alguna de las pruebas, es identificado como un componente falso, lo cual significa que la estimación del número de componentes ha superado el número verdadero. Cuando no se detectan componentes falsos, en la siguiente iteración $t = t + 1$ se selecciona un nuevo perfil candidato de entre los datos de entrada \mathbf{Y} y se agrega al conjunto \mathbf{P}^N . Enseguida se incrementa el orden del modelo $N = N + 1$ y se repite el proceso de descomposición para refinar la solución $\{\mathbf{P}^N, \mathbf{A}^N\}$. El proceso iterativo es similar al que realiza BEAE pero con algunos pasos adicionales, justo como se ilustra en la Figura 1.

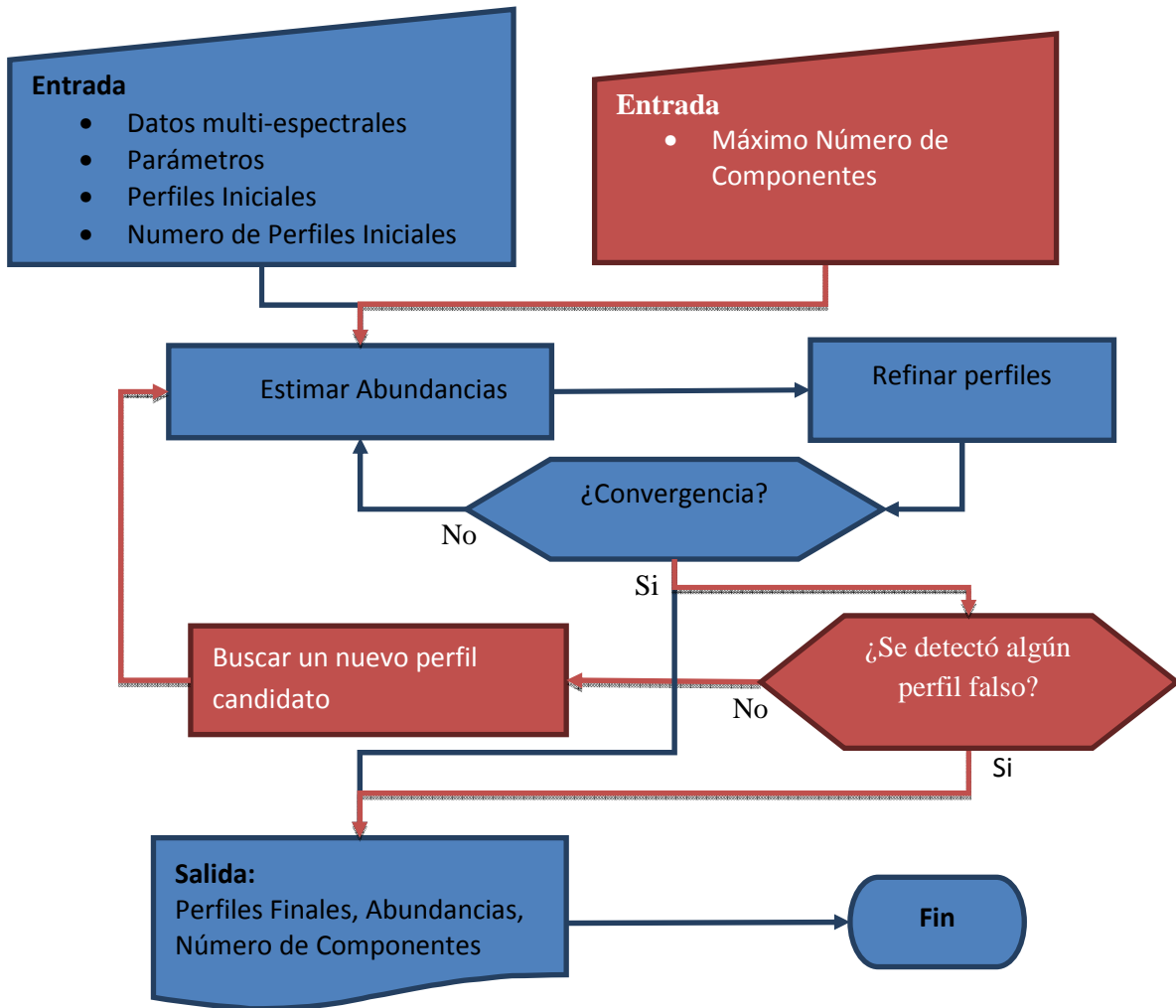


Fig. 1. Diagrama de flujo de los algoritmos de descomposición lineal cuadráticos. Los pasos de BEAE se muestran en azul. El algoritmo QBLU adicionalmente implementa los pasos en rojo.

4. La interfaz gráfica y sus herramientas

El programa fue desarrollado en Matlab con la intención de facilitar la aplicación de los algoritmos de descomposición descritos en la Sección 3. El software se encuentra disponible en <http://galia.fc.uaslp.mx/~omargn/mflim.html>. El programa tiene el formato de Matlab App, el cual es compatible con Matlab 2013a y versiones más recientes. El aspecto de la ventana principal se puede observar en la Figura 2.

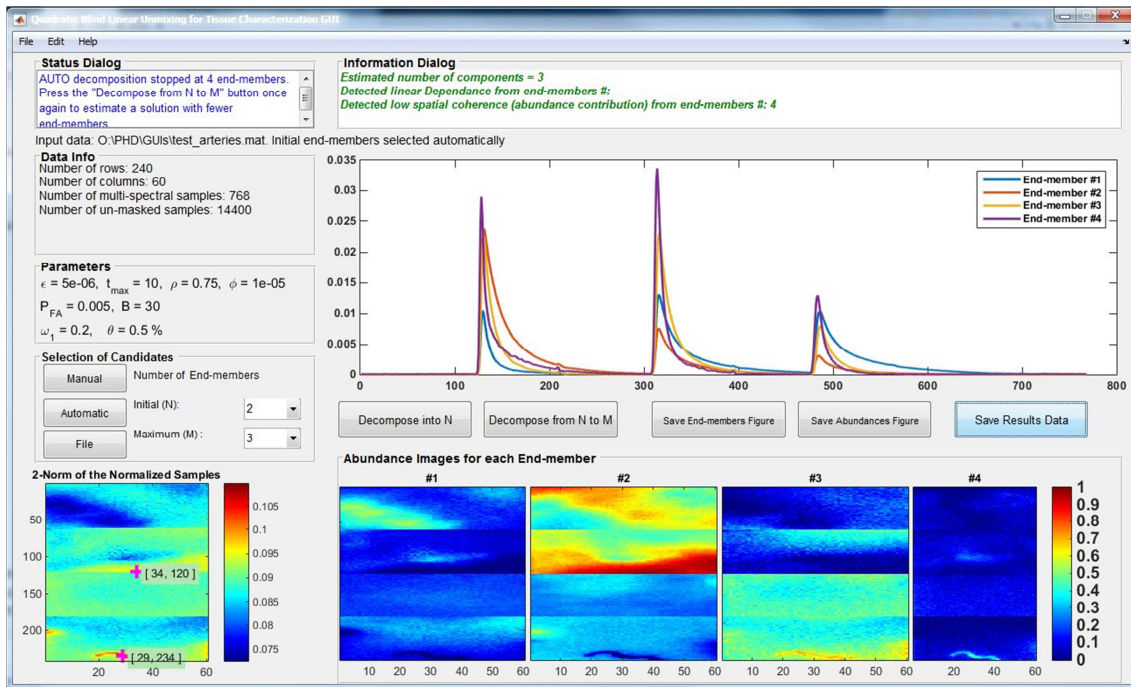


Fig. 2.- Ventana principal de la interfaz gráfica para descomposición lineal de datos multi-espectrales.

Los datos de entrada deben ser suministrados en un arreglo tridimensional por el usuario, donde las primeras dos dimensiones corresponden al plano espacial o superficie de la imagen, y la tercera debe contener los datos multi-espectrales. Esta primera versión ha sido desarrollada en inglés y permite al usuario hacer la carga de datos y pre-procesarlos fácilmente. Una vez que el usuario carga los datos por medio de la opción “*Load Dataset*” en el menú “*File*”, un dialogo permite al usuario seleccionar el arreglo que desea procesar. La interfaz gráfica provee dos opciones para descomponer los datos, cada una corresponde a los algoritmos BEAE o QBLU. Si se conoce el número de componentes presentes en la muestra, el usuario puede proporcionar esta información al programa y seleccionar la opción “*Decomposition into N*” para utilizar el método BEAE. Alternativamente, la opción “*Decomposition from N to M*” permite al usuario ejecutar el algoritmo QBLU para estimar adicionalmente el número de componentes. Esta opción requiere que el usuario suministre un número mínimo y máximo de esta variable.

La interfaz muestra dos ventanas de dialogo con las cuales comunica al usuario el estado del proceso de descomposición. Si en algún momento se detecta un componente falso, o el proceso ha terminado, el programa muestra un aviso en el cuadro de texto "*Information Dialog*". El cuadro "*Status Dialog*" notifica al usuario que proceso se ejecuta actualmente, así como información sobre el proceso de descomposición. Mientras se ejecuta, el programa va mostrando cómo evoluciona la estimación de los perfiles característicos y sus abundancias. La Figura 3 muestra un diagrama general del funcionamiento y opciones que presenta el programa.

4.1. Parámetros principales

El comportamiento de los métodos de optimización puede ser modificado por el usuario al personalizar los parámetros principales. Los valores por defecto de los parámetros de BEAE $(\epsilon, t_{max}, \rho, \phi)$ y del algoritmo QBLU $(P_{FA}, B, \omega_1, \theta)$ son mostrados en la Tabla 1 junto con una breve descripción de su función o interpretación.

4.2. Pre-procesamiento

El programa está pensado para la descomposición de datos multi-espectrales de acuerdo al modelo (2), siempre y cuando sean factible aplicar las restricciones (3) a (6). Para poder soportar datos de diferentes fuentes, con diferentes condiciones de ruido o campos de visión, la interfaz incorpora algunas herramientas de pre-procesado para lidiar con algunos de estos artefactos. El programa incluye herramientas para promediar y/o reducir el número de muestras, así como para sub-muestrear las señales de entrada en el eje multi-espectral. Además se incluye un filtro de mascara para poder eliminar muestras con baja relación señal a ruido. Es necesario mencionar, que una vez que los datos multi-espectrales son cargados por el programa, éste se encarga de procesarlos para que cumplan las restricciones de suma a uno y positividad.

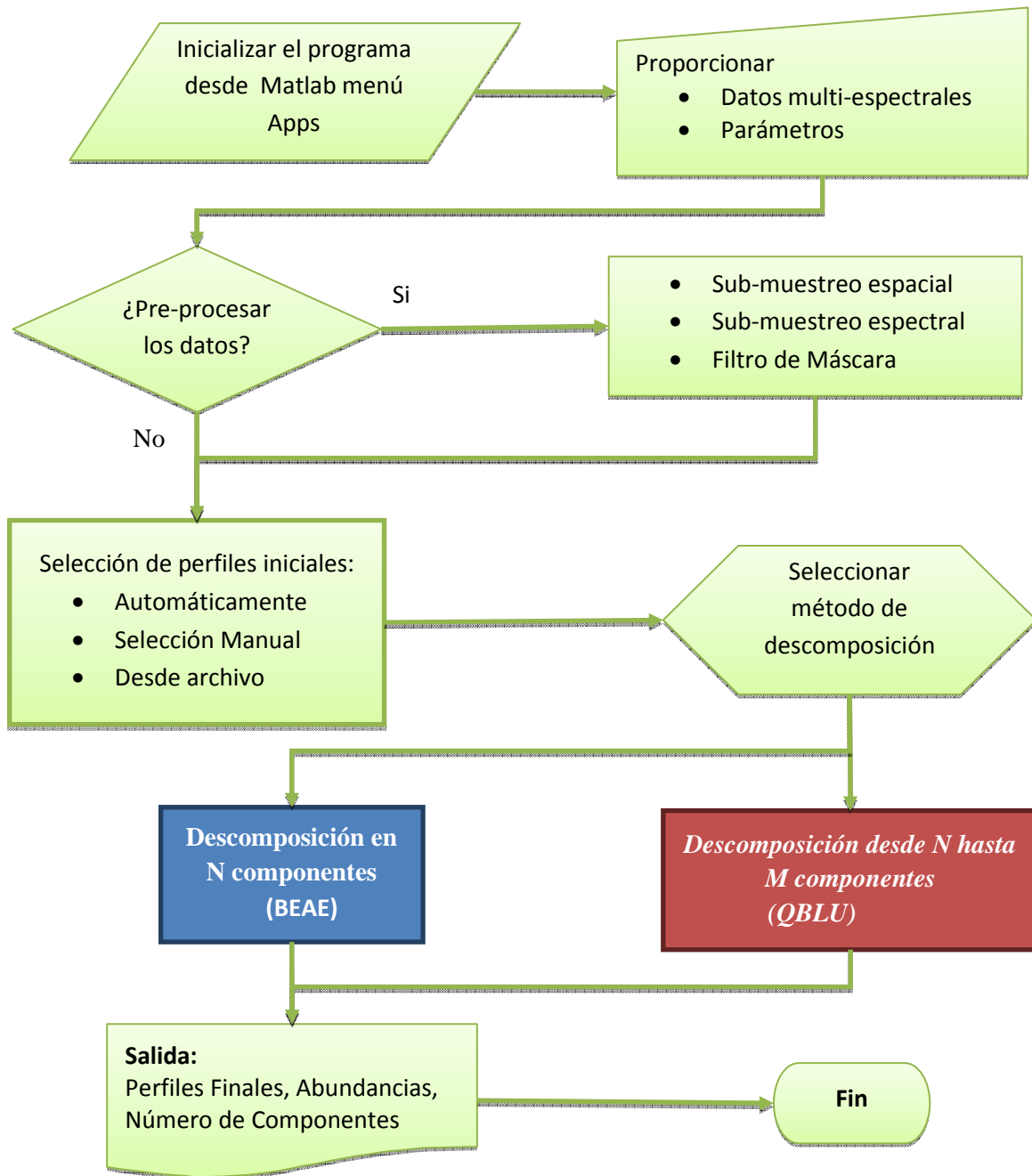


Fig. 3. Diagrama de flujo de la operación realizada por la interfaz gráfica para descomposición de datos multi-espectrales.

| Parámetro | Valor por defecto | Descripción |
|------------|--------------------|---|
| ϵ | 5×10^{-6} | Condición para la convergencia de BEAE |
| t_{max} | 10 | Máximo número de iteraciones de BEAE |
| ρ | 0.75 | Peso de regularización en la función de costo (8) de BEAE |
| ϕ | 1×10^{-5} | Condición de similaridad entre los perfiles iniciales |
| P_{FA} | 1×10^{-3} | Probabilidad de falsa alarma para los detectores NP |
| B | 30 | Numero de casillas utilizadas por el detector de coherencia espacial |
| ω_1 | 0.2 | Límite superior de la distribución uniforme en el descriptor de coherencia espacial |
| θ | 5×10^{-3} | Límite para evaluar abundancias significativas |

Tabla 1.- Valores por defecto de los parámetros principales del programa. Estos valores pueden ser modificados por el usuario en la opción “Parameters” del menú “Edit”.

4.3. Selección de perfiles iniciales

Los algoritmos empleados por la interface necesitan un conjunto de perfiles característicos iniciales (P)⁰. Estos candidatos, pueden ser elegidos por el usuario manualmente o automáticamente por el programa, dentro de los mismos datos de entrada. La selección automática de perfiles iniciales está basada en el método de detección de muestras puras en [10]. También es posible cargar los perfiles iniciales desde un archivo si el usuario los tiene disponibles.

4.4. Salida del Programa

Una vez que los métodos de descomposición (BEAE o QBLU) han terminado, el programa arroja tres grupos de resultados: los perfiles característicos, sus mapas de abundancias y el número de componentes o perfiles. Los perfiles característicos y sus mapas de abundancias pueden ser guardados como archivos “.jpg” o “.eps”. Esto puede hacerse por medio de los diálogos en los botones “*Save End-members Figure*” o “*Save Abundances Figure*”, para guardar los perfiles o los mapas de abundancias, respectivamente. El programa facilita el salvar estos datos en arreglos de Matlab mediante el botón “*Save Output Data*”.

5. Validación con datos m-FLIM de arterias humanas ex-vivo

Una de las posibles aplicaciones del programa es el obtener una interpretación cuantitativa de datos m-FLIM, lo cual permite describir el entorno bioquímico de una muestra. Para ilustrarlo, utilizamos 4 arreglos de datos m-FLIM obtenidos del lumen de placas ateroscleróticas humanas ex-vivo en [3]. El dispositivo utilizado [11] graba la respuesta fluorescente emitida por las muestras en tres longitudes de onda: 390 ± 20 nm, 450 ± 40 nm y 550 ± 80 nm. Cada arreglo contiene la información captada en un área de 2×2 mm² con una resolución temporal de 250 ps. La información de los 3 canales es almacenada en un solo arreglo tridimensional de 240x60x768 datos. El objetivo de este experimento es poder obtener una caracterización cuantitativa en función de los fluoróforos contenidos dentro de la muestra. Se sabe que las muestras de placas ateroscleróticas contienen principalmente tres fluoróforos: colágeno, elastina y lípidos de baja densidad (LDL, por sus siglas en inglés). Sin embargo, para probar la funcionalidad del programa, se analizaron las muestras utilizando el método QBLU para estimar de manera ciega el número de componentes. Los parámetros utilizados son aquellos mostrados en la tabla, además se seleccionó una descomposición desde 2 componentes hasta un máximo de 5. La descomposición se realizó utilizando la

selección automática de perfiles iniciales. Los perfiles seleccionados por el programa y sus posiciones con respecto a los datos, son mostrados en la Figura 4.

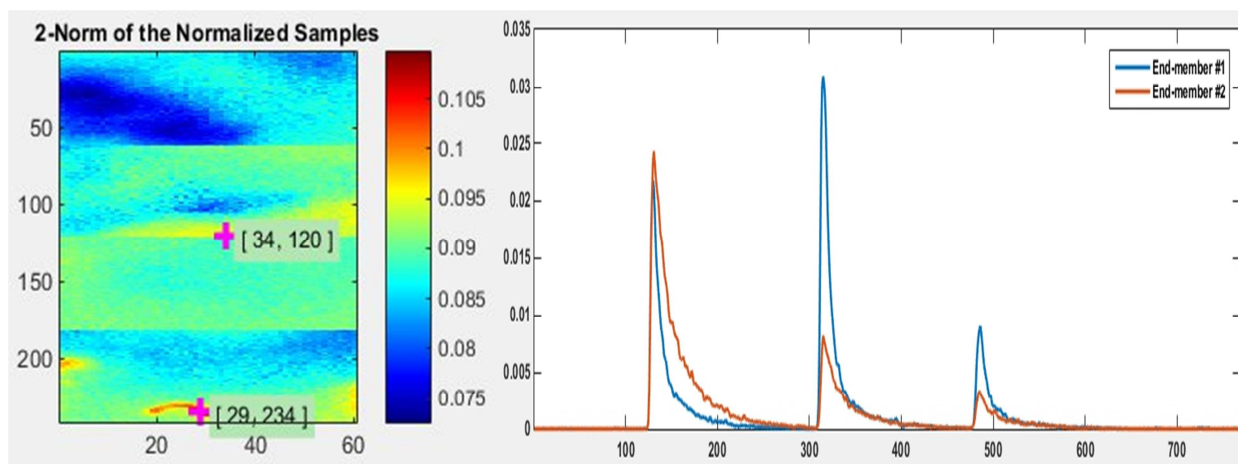


Fig. 4. La imagen de la izquierda muestra la posición de los perfiles dentro del mapa de la norma L_2 de los datos multi-espectrales. A la derecha se grafican los perfiles iniciales, correspondientes a las posiciones previamente seleccionadas, para la descomposición de los datos m-FLIM de arterias.

De acuerdo a los resultados del algoritmo QBLU, el programa se detuvo al realizar la descomposición para cuatro componentes. Se detectó que el cuarto componente es falso dado que la contribución de sus abundancias es muy bajo [7] a lo largo de la muestra. Los perfiles y abundancias calculados considerando la existencia de cuatro componentes pueden ser observados en la Figura 5. De esta manera, el programa aconseja al usuario que probablemente el número de componentes presente en la muestra es de tres y aconseja recalculer la solución con BEAE fijando el número correcto de componentes. La descomposición fue reiniciada utilizando estas condiciones, los perfiles y mapas de abundancia finales pueden ser apreciados en la Figura 6.

Para verificar la validez de los resultados se deben identificar los perfiles característicos. Una manera de hacerlo es por medio del cálculo de los tiempos de vida de cada perfil de decaimiento de fluorescencia [4]. Esto se realizó por medio del método en [12], los

tiempos de vida calculados se muestran en la Tabla 2. De acuerdo a estos resultados, los perfiles extraídos fueron efectivamente identificados como LDL, colágeno y elastina.

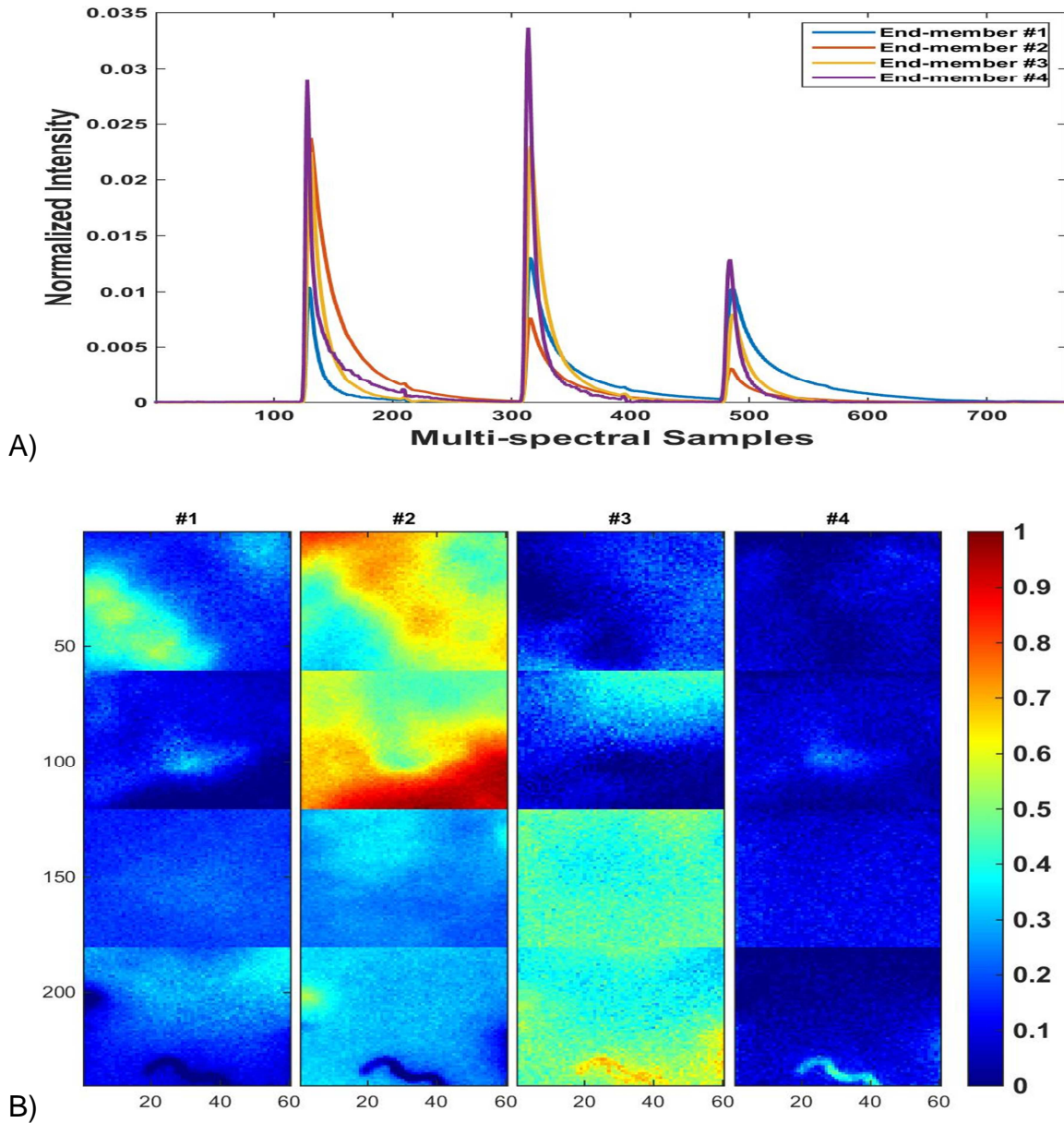


Fig. 5. Resultados obtenidos de la descomposición de las muestras de arterias tomando en cuenta la existencia de 4 componentes. A) Gráfica de los cuatro perfiles obtenidos. B) Mapa de abundancias resultante, cada columna representa la concentración.

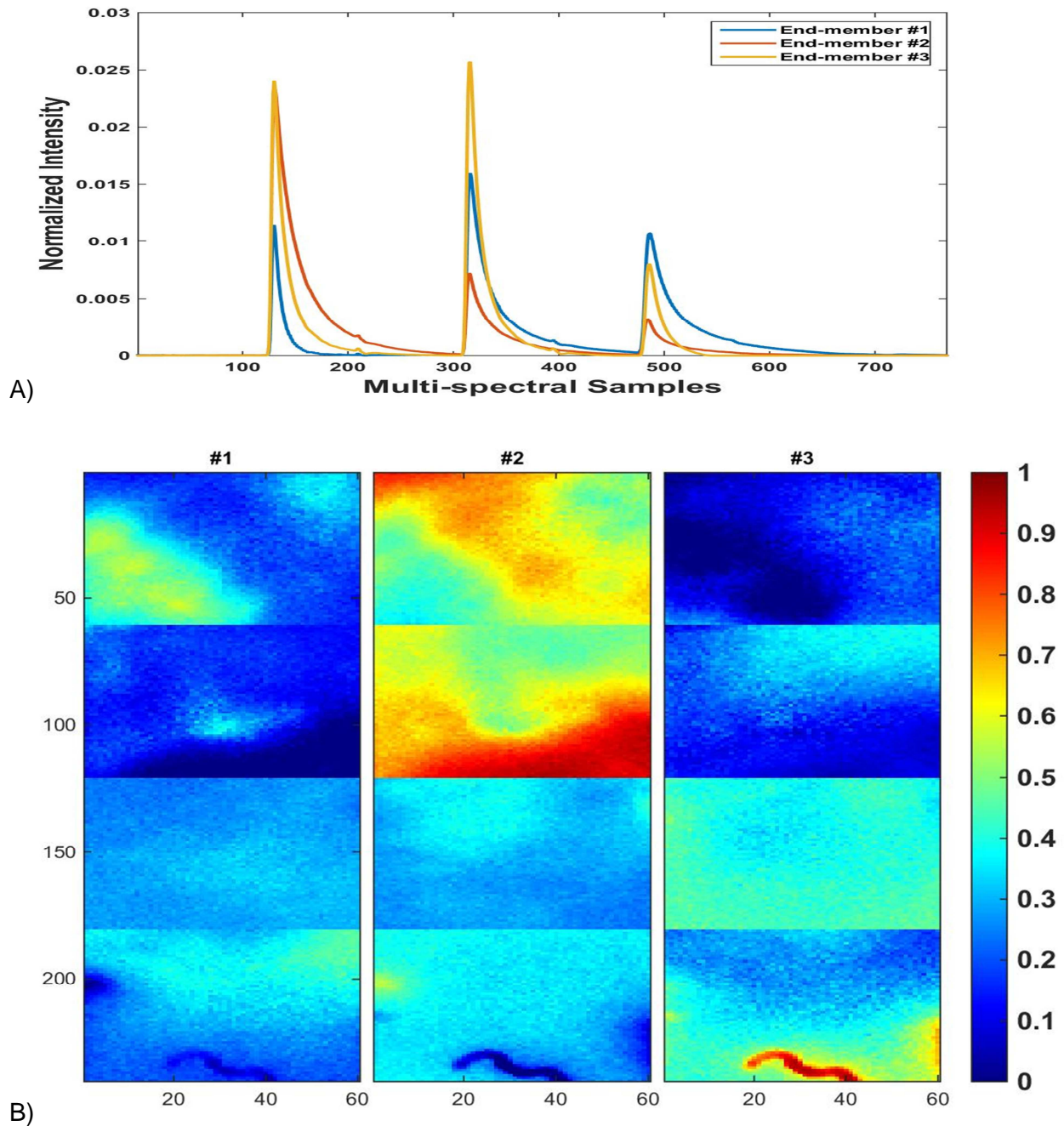


Fig. 6. Descomposición final de los datos de placas ateroscleróticas considerando tres componentes de acuerdo a la estimación del programa. A) Gráfica de los perfiles finales. B) Mapas de abundancia, cada columna representa la concentración de un componente. Los tres componentes fueron posteriormente identificados como LDL, colágeno y elastina.

| Componente | Tiempo de vida | | | Clasificación |
|------------|----------------|-----------|-----------|---------------|
| | Banda 390 | Banda 450 | Banda 550 | |
| 1 | 2.1108 ns | 7.0871 ns | 8.3642 ns | LDL |
| 2 | 6.2074 ns | 6.8167 ns | 7.1107 ns | Colágeno |
| 3 | 3.2079 ns | 3.0351 ns | 1.6998 ns | Elastina |

Tabla 2. Tiempos de vida estimados para los tres perfiles extraídos de las muestras de placas ateroscleróticas.

6. Comentarios Finales

En este trabajo se presenta una nueva herramienta de software para la caracterización de datos multi-espectrales. El programa incluye una interfaz **gráfica** y un conjunto de herramientas las cuales son validadas en la descomposición de datos m-FLIM. El software puede estimar el número de componentes constituyentes dentro de un conjunto de muestras, además de calcular una descripción cuantitativa mediante los perfiles característicos y sus concentraciones proporcionales. En cuanto a su eficiencia en tiempo de computo, el programa puede realizar la descomposición de un arreglo m-FLIM de tamaño 60x60x710 en alrededor de 60 segundos. Este tiempo fue calculado en un equipo con CPU Intel Core i7 con 12 GB en RAM. Los métodos numéricos empleados por el programa, se basan en aproximaciones cuadráticas del modelo de la mezcla lineal, lo cual permite obtener resultados robustos. La importancia de estos métodos numéricos radica en que ofrecen una alternativa para analizar de manera cuantitativa datos multi-espectrales, sin basarse en un modelo pre-definido para los perfiles característicos, o limitado en cuanto al máximo número de componentes posibles. El programa puede utilizarse para la descomposición de diferentes datos multi-espectrales y no es exclusivo para el procesamiento de datos m-FLIM. Esto es posible siempre y cuando el modelo lineal en (2) y las restricciones en (3) a (6) sean aplicables. Una limitante para la aplicación clínica de este programa es la

descomposición simultánea de una gran cantidad de muestra multi-espectrales. Es necesario desarrollar métodos que puedan caracterizar datos correspondientes a grandes regiones de tejido, lo cual es parte del trabajo de investigación actual.

7. Referencias

- [1] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Vol. 5. No. 2. 2012. 354–379 pp.
- [2] B. Green, A. R. Cobb, P. A. Brennan, C. Hopper, "Optical diagnostic techniques for use in lesions of the head and neck: review of the latest developments". *British Journal of Oral and Maxillofacial Surgery*. Vol. 52. No. 8. 2014. 675– 680 pp.
- [3] P. Thomas, P. Pande, F. Clubb, J. Adame, J. A. Jo, "Biochemical imaging of human atherosclerotic plaques with fluorescence lifetime angioscopy". *Photochemistry and Photobiology*. Vol. 86. No. 3. 2010. 727–731 pp.
- [4] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Springer. 2006.
- [5] W. K. Ma, J. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. Plaza, Ambikapathi, C.Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing". *IEEE Signal Processing Magazine*. Vol. 31. No. 1. 2014. 67–81 pp.
- [6] O. Gutierrez-Navarro, D. Campos Delgado, E. Arce-Santana, M. Mendez, J. Jo, "Blind end-member and abundance extraction for multi-spectral fluorescence lifetime imaging microscopy data". *IEEE Journal of Biomedical and Health Informatics*. Vol. 18. No. 2. 2014. 606–617 pp.

- [7] O. Gutierrez-Navarro, D. U. Campos-Delgado, E. R. Arce-Santana, K. C. Maitland, S. Cheng, J. Jabbour, B. Malik, R. Cuenca, J. A. Jo, "Estimation of the number of fluorescent end-members for quantitative analysis of multispectral flim data". *Optics Express*. Vol. 22. No. 10. 2014. 12255–12272 pp.
- [8] O. Gutierrez-Navarro, D. Campos Delgado, E. Arce-Santana, M. Mendez, J. Jo, "A fully constrained optimization method for time-resolved multispectral fluorescence lifetime imaging microscopy data unmixing". *IEEE Transactions on Biomedical Engineering*. Vol. 6. No. 6. 2013. 1711–1720 pp.
- [9] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. 1st Edition. Springer Publishing Company Incorporated. 2008.
- [10] T.-H. Chan, C.-Y. Chi, K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using ℓ_1 -norm-based pure pixel identification algorithm". *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 51. No. 5. 2013. 2753– 2769 No.
- [11] S. Shrestha, B. E. Applegate, J. Park, X. Xiao, P. Pande, J. A. Jo, "Highspeed multispectral fluorescence lifetime imaging implementation for in vivo applications". *Optics Letters*. Vol. 35. No. 15. 2010. 2558–2560 pp.
- [12] P. Pande, J. A. Jo, "Automated analysis of fluorescence lifetime imaging microscopy (flim) data based on the laguerre deconvolution method". *IEEE Transactions on Biomedical Engineering*. Vol. 58. No. 1. 2011. 172–181 pp.

8. Autores

Omar Gutiérrez Navarro obtuvo su Doctorado en Ingeniería Electrónica en la Universidad Autónoma de San Luis Potosí (UASLP) en 2014. Realizo sus estudios de Maestría en el Centro de Investigación en Matemáticas (CIMAT) en Guanajuato en 2010 y el título de Ingeniero en Electrónica por la UASLP en 2007. Su trabajo de investigación se centra en métodos numéricos para la caracterización de tejido utilizando imágenes microscópicas de tiempo de vida de fluorescencia.

Daniel U. Campos Delgado es profesor de tiempo completo en la UASLP desde 2001. Recibió su título de Ingeniero en electrónica en la UASLP en 1996. Sus estudios de Maestría y Doctorado los realizó en Louisiana State University (LSU), Baton Rouge, LA, EUA, en 1999 y 2001, respectivamente. Es autor de más de 150 artículos en revistas científicas y congresos. Sus áreas de interés incluyen métodos de optimización, modelado de sistemas dinámicos y procesamiento de señales.

Edgar R. Arce Santana es profesor en la UASLP en México. Recibió su grado de licenciatura en ciencias de la computación en el Instituto Tecnológico de San Luis Potosí en 1987. Obtuvo sus grados de Maestría y Doctorado en el CIMAT en Guanajuato en 2000 y 2004, respectivamente. Es autor de más de 50 artículos publicados en revistas y congresos científicos. Sus áreas de interés incluyen la visión computacional, procesamiento de señales y reconocimiento de patrones.

Javier Jo es profesor en el Biomedical Engineering Department en la universidad Texas A&M. Recibió su grado de Ingeniero Electrónico de la Pontificia Universidad Católica del Perú en 1996. Obtuvo sus grados de Maestría y Doctorado de la University of Southern California, LA, EUA en 2000 y 2002, respectivamente. Sus principales áreas de interés incluyen análisis de sistemas y procesamiento de imágenes, así como instrumentación biomédica con aplicaciones en biofotónica, fisiología y medicina.

Estimación de erosión de suelos utilizando sensores remotos y programación genética

Cesar Augusto Puente Montejano

UASLP, Facultad de Ingeniería, Dr. Manuel Nava # 8, Zona Universitaria poniente, C.P. 78290,
San Luis Potosí, S.L.P, México, Teléfono: 01 (444) 826-2330 ext. 2112
cesar.puente@uaslp.mx

Gustavo Olague Caballero

Carretera Ensenada-Tijuana No. 3918, Zona Playitas, C.P. 22860, Ensenada, B.C. México,
Teléfono: 01(646)175-05-00
olague@cicese.mx

Resumen

Los índices de vegetación (IVs) son ampliamente utilizados para extraer información de la vegetación a partir de imágenes satelitales. Los modelos de erosión, como la "Ecuación Universal Revisada de la Pérdida de Suelo" (RUSLE) usan IVs como insumo para estimar el factor de cobertura vegetal (C). El factor C es uno de los más importantes porque cuantifica la cobertura que actúa como capa protectora entre el suelo y los elementos atmosféricos. Sin embargo los IVs encontrados en el estado-del-arte arrojan pobres resultados, ya que la mayoría de éstos están diseñados para detectar vegetación verde y no vegetación seca; la cual es también un importante factor que contribuye al desempeño del factor C. El propósito de esta investigación es desarrollar un método basado en programación genética para sintetizar IVs que estén mejor correlacionados con el factor C. Los resultados experimentales ilustran la eficiencia de este método y su efecto en el cálculo de erosión en una zona geográfica real. Los índices sintetizados obtienen una mejor aproximación al factor C obtenido en campo que cuando se utilizan los índices reportados en el estado-del-arte.

Palabra(s) Clave(s): erosión por agua, índices de vegetación, programación genética, percepción remota, RUSLE.

1. Introducción

Cada vez existe un mayor interés por el aprovechamiento sostenible de los recursos y crece la atención por los cambios ambientales. Uno de estos cambios ambientales es la erosión del suelo. La erosión del suelo se define como el desplazamiento de suelo fértil, debido a la acción de un agente llamado erosivo [1]. La erosión causada por agua se debe a la precipitación de lluvia en zonas donde las pendientes del terreno son relativamente empinadas, la litología es suave o no consolidada y la vegetación está dispersa. Para mejorar la planeación ambiental y conservación de suelos, es importante la detección y monitoreo de las áreas sujetas a fenómenos erosivos.

Se han formulado distintos modelos para estimar erosión tanto empíricos [2, 3], como los basados en propiedades físicas formales [4, 5]. El método de estimación de erosión más utilizado es el de la Ecuación Universal de Pérdida de Suelo Revisada [3] (RUSLE, por sus siglas en inglés), el cual es un modelo empírico que calcula promedios anuales de erosión. RUSLE se compone por seis factores:

$$A = R \times K \times L \times S \times C \times P \quad (1)$$

donde A es el promedio de pérdida de suelo en $\text{Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$; R es el factor de lluvia-escurrimiento que representa la energía erosiva medida en $\text{MJ} \cdot \text{mm} \cdot \text{km}^{-2} \cdot \text{h}^{-1} \cdot \text{año}^{-1}$; K es la tendencia del suelo a erosionarse ($\text{t} \cdot \text{h} \cdot \text{MJ}^{-1} \cdot \text{mm}^{-1}$). Los demás factores son factores de escalamiento adimensionales. L es el factor de longitud de la pendiente; S es el factor de inclinación de la pendiente; C es el factor de cobertura; y P es el factor de soporte a la conservación de suelos.

Uno de los parámetros más importantes del modelo RUSLE es el que estima la cobertura del suelo o factor C. La cobertura del suelo se define como las diferentes cubiertas naturales que protegen al suelo del impacto directo de los elementos

atmosféricos. Por ejemplo, en un bosque el follaje vegetal absorbe la mayor parte de la energía de impacto que posee una gota de agua al caer durante un período de lluvia; lo que evita la erosión masiva del suelo que sostiene dicho bosque.

Resulta sumamente costoso medir directamente en campo los datos necesarios para alimentar el modelo RUSLE a gran escala, por lo que se han utilizado técnicas de percepción remota que ayudan a extrapolar medidas de campo representativas sobre extensas áreas geográficas [6]. La *Percepción Remota* es la disciplina encargada de desarrollar herramientas para extraer la información con la que se estiman algunas de las propiedades biofísicas de todo aquello que se encuentre sobre la superficie terrestre. En particular, se han desarrollado varias metodologías para identificar la cobertura vegetal; como los métodos de clasificación espectral, por ejemplo en [7] o los llamados índices de vegetación, por ejemplo en [8]. Sin embargo, de acuerdo con [2], la información que requieren los modelos de erosión para medir la cobertura difiere de la información que proveen estos métodos; los cuales, generalmente han sido diseñados para otras tareas, e.g. clasificación de uso de suelo, productividad agrícola, etc. Por lo tanto, existe una importante necesidad para desarrollar IVs que estén específicamente diseñados para la estimación de la cobertura vegetal en el contexto de erosión.

En este trabajo de investigación se propone una nueva metodología basada en la *Programación Genética* (PG) con el objetivo de obtener una estimación más precisa de uno de los parámetros más importantes en el modelo de erosión RUSLE: el *factor C*. En este trabajo, se plantea el problema de calcular el factor de cobertura como un problema de optimización; donde el objetivo es encontrar el índice de vegetación que muestre una mejor correlación con datos de campo del factor C. De esta forma, el algoritmo basado en PG construye nuevos índices de vegetación mediante la recombinación iterativa de un conjunto de operadores numéricos y bandas espectrales pertenecientes a imágenes de satélite.

2. Desarrollo

En la Figura 1 se muestra un diagrama con la metodología desarrollada durante esta investigación. El objetivo de esta metodología es desarrollar un enfoque automático para crear índices de vegetación que estén bien correlacionados con el factor C de RUSLE. En esta sección se describen los pasos que conforman dicha metodología.

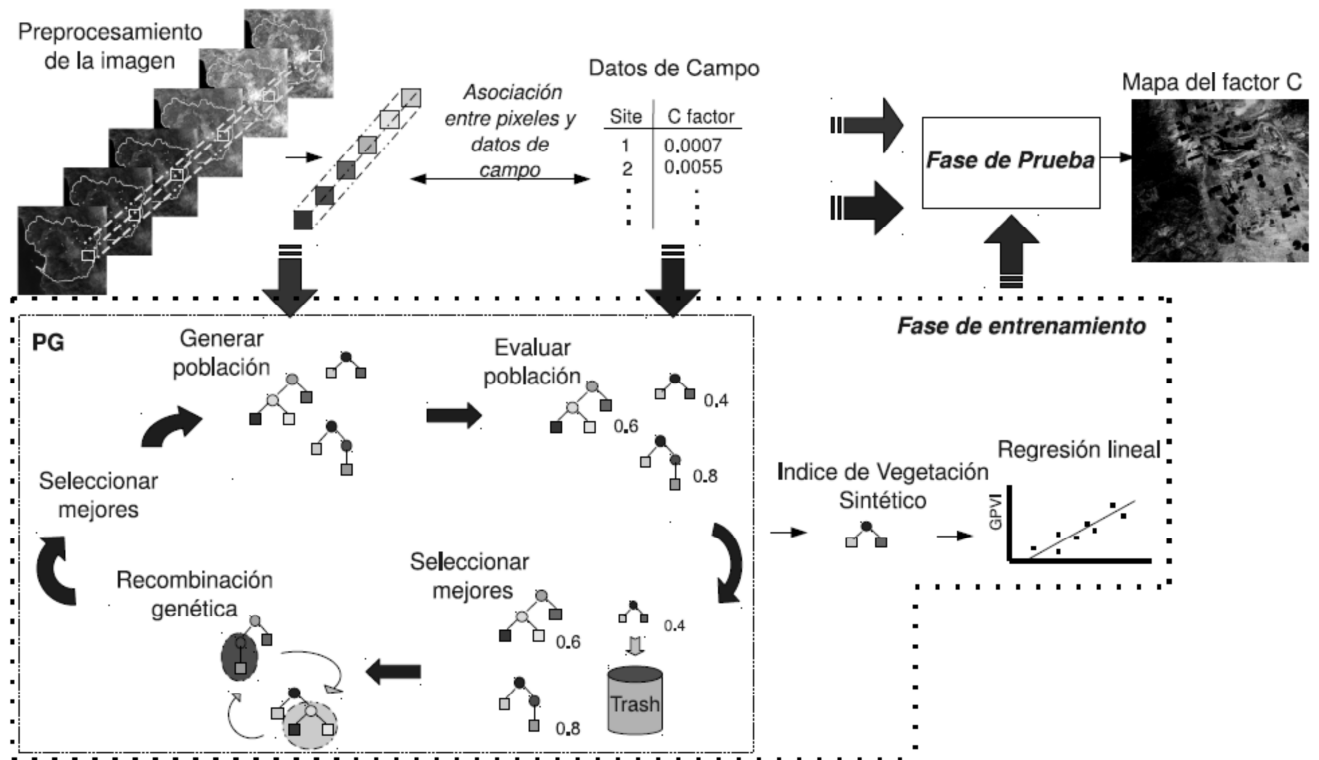


Fig. 1. Diagrama de flujo general de la metodología para estimar el factor C a partir de los IVs sintetizados por la PG.

2.1. Descripción de área de estudio

La cuenca Todos Santos se localiza al noroeste de la península de Baja California, México (ver Figura 2). Posee una topografía compleja cuya elevación oscila entre el nivel del mar y los 1876 metros sobre el nivel del mar. El promedio de pendientes es de 10°, aunque ~27% del área posee una entre los 10° y 20°; y el ~3% presenta una

inclinación de 30° o más. Las áreas aluviales (de 2°), ocupan aproximadamente el 19% de la cuenca. Existen dos grandes valles aluviales, conocidos como los valles agrícolas de Guadalupe y de Ojos Negros, En la planicie costera se encuentra asentada la mayor parte de la ciudad de Ensenada. Asimismo, destaca la presencia del valle agrícola de Maneadero al sur de dicha ciudad.

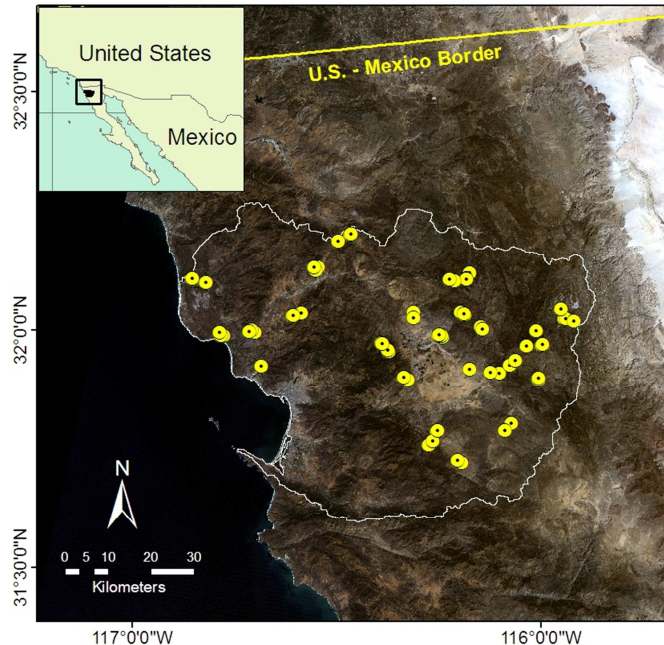


Fig. 2. Mapa de Baja California, México que muestra la localización de la cuenca Todos Santos. Los círculos amarillos representan los 67 sitios de muestreo de campo para obtener el factor C.

2.2. Muestreo de campo

El muestreo de campo implica las mediciones de diferentes parámetros para la estimación del factor C. Los autores de [9] identificaron los tres sub-factores principales que determinan la efectividad de la vegetación para limitar la erosión de suelo en ambientes áridos; a saber: *cobertura superficial del suelo* (vegetación a ras de piso: g), *cobertura vegetal aérea* (ramas suspendidas sobre el terreno: $p =$ porcentaje y $h =$ altura) y *cobertura residual* (rugosidad del suelo: r) y *subterránea* (raíces: b). Para la

campaña de muestreo de campo se realizó un análisis detallado del área de estudio con la ayuda de imágenes satelitales y mapas topográficos para identificar las zonas representativas de muestreo. La campaña se llevó a cabo entre Febrero y Mayo de 2007, adquiriendo un total de 67 sitios de muestreo (ver Figura 2). En cada uno de estos sitios se realizó un protocolo de muestreo basado en el método del transecto lineal descrito por [10, 11]. A partir de estas mediciones, el factor C para cada uno de los sitios de muestreo se define utilizando la siguiente ecuación, la cual fue derivada en el trabajo de [12] a partir de la Tabla 10 del manual del modelo RUSLE [9]:

$$C = 0.45(e^{-0.012b}) (1 - p e^{-0.328h}) (e^{-0.039g \frac{0.24}{r} 0.08}) \quad (2)$$

2.3. Pre-procesamiento de la imagen satelital

La imagen satelital utilizada para generalizar los sitios de campo en toda la cuenca pertenece al satélite Landsat 5 - TM. Cada imagen que el dispositivo Landsat TM entrega cubre aproximadamente 185 Km por lado y consta de siete bandas. Seis de ellas cubren la parte visible/infra-rojo cercano del espectro electromagnético y poseen una resolución espacial de 30 x 30 m por pixel. La banda restante cubre la parte del infra-rojo termal y tiene una resolución de 60 x 60 m (ver Tabla 1). Esta última banda no fue utilizada en este trabajo.

| Banda | Rango espectral (µm) | Región espectral | Utilidad principal |
|-------|----------------------|------------------------------------|--|
| 1 | 0.45 - 0.52 | Visible azul (B) | Penetración de cuerpos de agua |
| 2 | 0.52 - 0.60 | Visible verde (G) | Vigor de la vegetación sana |
| 3 | 0.63 - 0.69 | Visible rojo (R) | Absorción de clorofila |
| 4 | 0.76 - 0.90 | Infrarrojo cercano (NIR) | Contenido de biomasa |
| 5 | 1.55 - 1.75 | Térmico lejano (TIR) | Estrés de la vegetación |
| 6 | 10.4 - 12.5 | Infrarrojo de onda corta 1 (SWIR1) | Discriminación de rocas |
| 7 | 2.08 - 2.35 | Infrarrojo de onda corta 2 (SWIR2) | Absorción de agua y cantidad de hidroxilos |

Tabla 1. Características de las bandas del sensor Landsat-TM.

La fecha de la imagen fue lo más cercana posible a la fecha del muestreo: el 13 de Abril de 2007. Para la aplicación que se le quiere dar en este trabajo, es necesario darle un pre-procesamiento a la imagen, el cual consiste en dos tipos de correcciones: geométricas y radiométricas. Las primeras tienen el objetivo de empatar un sistema de coordenadas de referencia con la imagen satelital; mientras que las segundas se llevan a cabo para disminuir los defectos por desgaste en el sensor y para eliminar influencias atmosféricas (ver Figura 3). El resultado de este pre-procesamiento es una nueva imagen, donde el valor numérico de cada pixel representa una magnitud física conocida como *reflectancia* (R); la cual se define como la relación entre la potencia electromagnética de la radiación incidente en una superficie con respecto a la potencia reflejada en la misma con un rango de $R: 0 \leq R \leq 1$

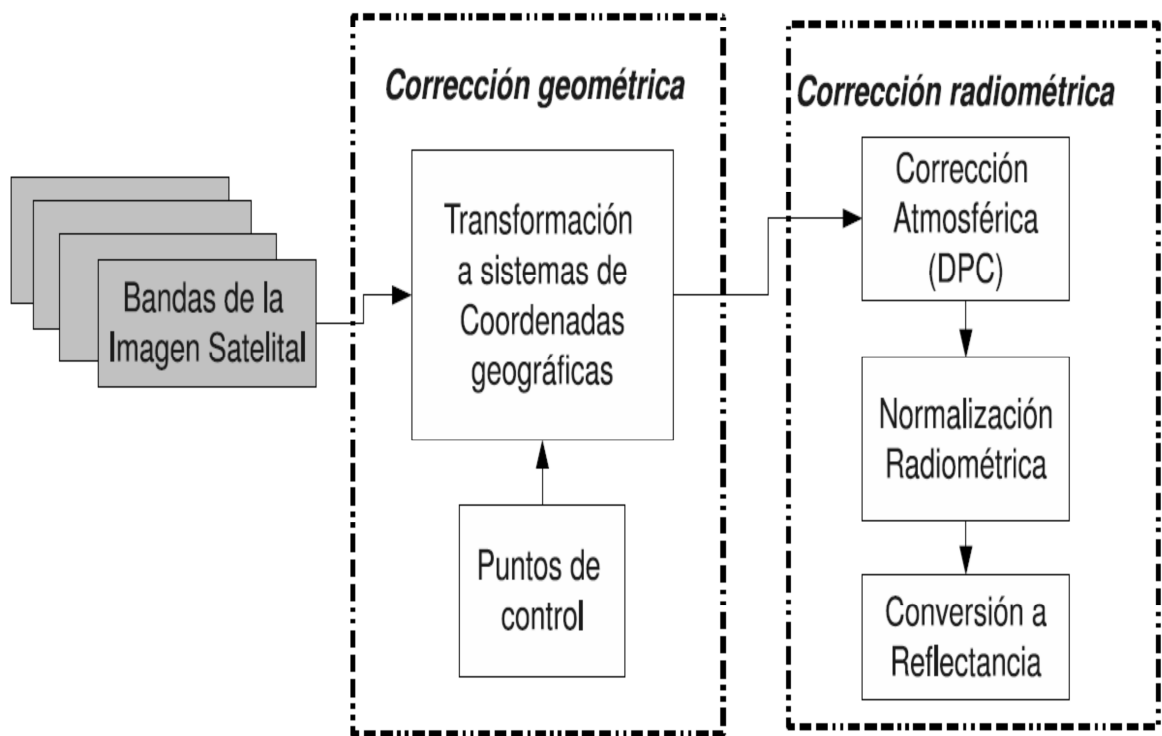


Fig. 3. Diagrama de flujo del proceso de pre-procesamiento de las imágenes satelitales utilizadas durante este trabajo.

2.4. Relación entre la imagen y los datos de campo

Una vez que se corrigieron las distorsiones en la imagen, se procedió a relacionar los datos de campo con los píxeles correspondientes en cada banda de la imagen. Se localizaron los 67 sitios de muestra en la imagen satelital. Se definió una ventana de 3×3 píxeles alrededor de cada punto identificado, ya que dichos puntos fueron escogidos pensando que cada sitio tenía una cobertura uniforme de al menos una hectárea (100×100 metros). De esta manera, por cada ventana de 9 píxeles, se extrajo la mediana, y se convirtió en el valor representativo de cada sitio de muestreo. Al final se ordenaron los datos obtenidos en una tabla de 67×6 valores de mediana. Por último, el conjunto completo de datos resultante se dividió en dos partes. El primer conjunto consta de 47×6 valores. Este grupo corresponde a los *datos de entrenamiento*. El segundo conjunto se compone de los 20×6 valores restantes y corresponde a los *datos de prueba*.

2.5. Revisión de la literatura sobre IVs existentes

Se tomaron treinta índices de vegetación extraídos de la literatura y se aplicaron sobre las tablas de píxeles de las zonas de estudio con el fin de obtener su factor de correlación con los datos de campo. De esta manera, se obtuvieron treinta valores asociados con cada valor de C de cada sitio de muestreo. Esto dio lugar a 30 matrices de 47×2 cada una. Posteriormente, por cada matriz se calculó el coeficiente de correlación ($\rho_{x,y}$) entre las dos columnas, obteniendo así 30 diferentes valores de correlación. Cada uno representa la correlación entre el factor C de los datos de campo, con uno de los índices. Es necesario hacer notar que de acuerdo a la definición del coeficiente de correlación, es posible obtener valores negativos. En este trabajo se optó por utilizar el valor absoluto $|\rho_{x,y}|$ ya que un valor de -1 indica un grado de correlación fuerte, lo mismo que un valor de 1. Por cuestiones de espacio aquí se muestra el desempeño de los mejores 10 índices (ver Tabla 2).

| Índice | Descripción | Autor | $ \rho_{x,y} $ |
|--------|--|--------------------|----------------|
| RVI4 | $\frac{SWIR1}{SWIR2}$ | Modificado de [13] | 0.512 |
| RVI5 | $\frac{SWIR1}{R}$ | Modificado de [13] | 0.383 |
| RVI1 | $\frac{NIR}{R}$ | Modificado de [13] | 0.342 |
| GEMI | $\mu(1 - 0.25 \mu) - \frac{(R - 0.125)}{1 - R}$ <p>donde $\mu = \frac{2(NIR^2 - R^2) + 1.5NIR + 0.5R}{NIR + R + 0.5}$</p> | [14] | 0.329 |
| RVI2 | $\frac{NIR}{G}$ | Modificado de [13] | 0.329 |
| NDVI | $\frac{NIR - R}{NIR + R}$ | [8] | 0.314 |
| IPVI | $\frac{NIR}{NIR + R}$ | [15] | 0.314 |
| SAVI2 | $\frac{NIR}{R + b/a}$ Donde a y b son la pendiente y la abcisa al origen de la línea del suelo | [16] | 0.309 |
| SASI | $\beta_{SWIR1}(NIR - SWIR2)$ Donde $\beta_{SWIR1} = \cos^{-1}\left(\frac{a_2 + b_2 - c_2}{2ab}\right)$ y a, b, c son las distancias euclidianas entre NIR, SWIR1 y SWIR2 | [17] | 0.254 |
| GVI | $-0.3344B - 0.3544G - 0.4556R + 0.6966NIR + 0.0242SWIR1 - 0.2630SWIR2$ | [18] | 0.228 |

Tabla 2. Los diez índices de vegetación del estado-del-arte que tuvieron un mejor rendimiento para la zona de estudio. Para más detalle sobre cada elemento que compone estos índices ver sección 2.6.1.

2.6. Desarrollo y aplicación de algoritmo basado en programación genética

En este trabajo, las soluciones candidatas que se someten al proceso de PG se codifican como árboles sintácticos, los cuales representan la fórmula matemática que define a un índice de vegetación. Por ejemplo, un IV muy utilizado en la literatura de la

Percepción Remota es el NDVI, cuya fórmula matemática se encuentra en la Tabla 2. Las variables y constantes en la fórmula: las bandas NIR y Red, son las hojas del árbol. Dentro de la nomenclatura de la PG, se les conoce como *Terminales*; mientras que a los operadores aritméticos (+, -, ÷), que son nodos internos, se les conoce como *Funciones*. El conjunto de Terminales y Funciones se define como el *conjunto de primitivas* de la PG; el cual representa el espacio de búsqueda del problema.

2.6.1. Conjunto de primitivas

- **Bandas espectrales (B, G, R, NIR, SWIR1, SWIR2).** Son los valores de reflectancias de los pixeles que representan los sitios de muestreo de cada una de las bandas de la imagen.
- **Ángulos espectrales (β_G , β_R , β_{NIR} , β_{SWIR1}).** Para conocer a detalle la manera de extraer los ángulos espectrales, sírvase revisar [17].
- **Parámetros de la línea del suelo (a, b).** La línea del suelo es la relación que existe entre las bandas R y NIR; la cuales, al ser graficadas una contra otra, agrupan los pixeles por encima de un umbral de valores. A este se le conoce como línea del suelo. El proceso que se presenta aquí para obtener los factores de la línea del suelo es el que propone [19].
- **Mejores índices convencionales (RVI4, RVI5, RVI1, GEMI, RVI2).** Se consideraron los cinco índices convencionales de mejor desempeño en la etapa anterior de esta metodología descrita en la sección 2.5.
- **Operadores aritméticos.** El conjunto de Funciones se compone primeramente por los operadores aritméticos básicos (+, -, y \times) ya que estos operadores son ampliamente utilizados para el diseño de índices de vegetación.
- **Operadores compuestos.** Representan estructuras aritméticas completas. Aquí se consideró la estructura de los índices más utilizados en la literatura de la Percepción Remota: el NDVI y el RVI. A la primera estructura se le conoce como índice espectral de diferencia normalizada (del inglés *Normalized Difference Spectral Index - NDSI*), mientras que a la segunda se le conoce como índice

espectral del cociente (del inglés *Ratio Spectral Index - RSI*), los cuales, se definen de la siguiente manera:

$$NDSI_{i,j} = \frac{R_i + R_j}{R_i - R_j}, \quad RSI_{i,j} = \frac{R_i}{R_j}, \quad (3)$$

donde R_k representa el valor de reflectancia de un pixel de la banda k .

2.6.2. Función de aptitud

La función de aptitud para este trabajo se basa en el coeficiente de correlación de Pearson $\rho_{x,y}$, el cual indica la fuerza y dirección de la relación lineal entre el factor C y cada uno de los índices sintetizados por la metodología. Se utilizó el valor absoluto del coeficiente de correlación como medida de desempeño porque mientras más cercano sea este valor a 1 ó -1, indistintamente la correlación entre las dos variables es más fuerte:

$$Q = \max(|\rho_{x,y}|) \quad (4)$$

donde x representa el factor C de RUSLE tomado en campo, mientras que y es el índice de vegetación generado por la metodología.

2.6.3. Ejecución del algoritmo de PG

Después del pre-procesamiento de la imagen (ver Figura 1), se genera la población inicial de índices de manera aleatoria, combinando los elementos de los conjuntos de Terminales y Funciones. Después, cada individuo de la población se evalúa con ayuda de la función de aptitud. El siguiente paso consiste en ordenar los índices de acuerdo a su desempeño con la función de aptitud, a fin de seleccionar los mejores candidatos para sobrevivir y descartar a los peores. Es entonces que sobreviene la recombinación genética con los candidatos elegidos. Finalmente, la siguiente generación resulta de la selección de los mejores entre los índices padres e hijos. Estos pasos son iterados sucesivamente hasta que se alcanza el máximo número de iteraciones. El índice con la

mayor correlación obtenido hasta ese momento es considerado el mejor, y constituye el nuevo *índice sintético*. A este nuevo índice se le llama GPVI_j (*Genetic Programming for Vegetation Index*), donde *j* es el número de la corrida en la que se generó dicho índice. En cuanto a los parámetros para correr el algoritmo de PG, estos serán descritos en la sección de resultados.

2.7. Evaluación de la solución

La evaluación de los resultados se lleva a cabo de acuerdo a la teoría del aprendizaje de máquina, es decir, a través de dos etapas llamadas entrenamiento y prueba. Para obtener la evaluación del desempeño de cada GPVI, se calculó la diferencia absoluta entre el coeficiente de correlación obtenido durante la etapa de entrenamiento y la etapa de prueba:

$$D = |\rho_{x,y}(\text{entren.}) - \rho_{x,y}(\text{prueba})| \quad (5)$$

Un desempeño similar en ambos conjuntos de datos implicaría una mayor estabilidad, ya que el comportamiento del índice ante diferentes valores de cobertura vegetal se mantiene consistente; obteniendo así el mejor mapa regional del factor C a partir de dicho índice.

2.8. Mapa del factor C y mapa de erosión

Finalmente, se lleva a cabo un mapeo lineal para convertir los valores numéricos de los mejores GPVI a los valores numéricos del factor C. De esta manera, los valores de C derivados de cada índice quedan en la misma escala que los valores de C derivados del muestreo de campo. Asimismo, para validar el desempeño de los nuevos índices, se aplican los nuevos mapas de C en la ecuación del modelo RUSLE para calcular la tasa de erosión. Todos los factores que componen el modelo RUSLE, excepto por C son conservados a lo largo de esta prueba. Es necesario decir que dichos factores (R, K, L, S, y P) fueron determinados usando un programa de Sistemas de Información Geográfica ArcGIS®, siguiendo la metodología descrita por [6].

3. Resultados

En esta sección se presentan los resultados de la metodología descrita en la sección anterior. La metodología se implementó utilizando MatLab®, con el módulo de programación genética conocido como GPLab, el cual está disponible con licencia libre en (<http://gplab.sourceforge.net/index.html>). La generación de los mapas de factor C y los mapas de erosión se llevaron a cabo con ArcGIS®. La Tabla 3 muestra el valor de los parámetros utilizados para el algoritmo de GP durante las pruebas experimentales. Se utilizaron valores canónicos recomendados por la literatura especializada en GP, los cuales se reportan con un buen desempeño en una gran variedad de aplicaciones. Para mayores detalles sobre los ajustes de los parámetros, sírvase revisar los trabajos de [20, 21].

| | |
|------------------------------------|--|
| Población inicial | Mitad árboles completos Mitad árboles de distribución aleatoria |
| Generaciones | 50 |
| Tamaño de la población | 50 individuos |
| Probabilidad de cruzamiento | 0.7 |
| Probabilidad de mutación | 0.3 |
| Tipo de selección | Torneo de presión lexicográfica |
| Tipo de elitismo | Conservar el mejor de la generación |
| Profundidad de los árboles | Dinámica, 4 niveles |

Tabla 3. Parámetros utilizados para el entrenamiento del algoritmo basado en PG.

Se llevaron a cabo treinta corridas del algoritmo de PG conservando el mejor individuo de cada una de estas corridas y fueron etiquetados como $TGPVI_j$ (*Todos santos Genetic Programming for Vegetation Index*), donde J representa el número de la corrida. Cada una de estos individuos es considerado como una solución aceptable. La Tabla 4 resume el desempeño de las soluciones obtenidas. Por razones de espacio, en

este artículo se publican los 10 mejores de esas corridas. El criterio para definir los mejores se encuentra en la sección 2.7. En general, los índices encontrados por el algoritmo basado en PG muestra un mejor rendimiento comparado con el de los índices convencionales, para ello se pueden comparar la Tabla 2 con la Tabla 4.

| Índice | Descripción | ρ_{entren} | ρ_{prueba} | Diferencia |
|---------|--|-----------------|-----------------|------------|
| TGPVI17 | $RSI(NIR, RSI(\beta_R, \beta_G)) + RSI(RSI(RV14, a), GEMI - RV14)$ | 0.648 | 0.645 | 0.002 |
| TGPVI27 | $RSI(GEMI, RV14 - NIR) - NIR$ | 0.642 | 0.639 | 0.003 |
| TGPVI19 | $GEMI \times \beta_{SWIR1} - GEMI - RV14$ | 0.622 | 0.629 | 0.008 |
| TGPVI28 | $\frac{GEMI}{RV14 - GEMI} - GEMI$ | 0.643 | 0.655 | 0.013 |
| TGPVI8 | $NDSI(NDSI(RV14, \beta_G), RV14) \times GEMI$ | 0.643 | 0.673 | 0.030 |
| TGPVI13 | $RV14 - GEMI - GEMI - G$ | 0.626 | 0.674 | 0.048 |
| TGPVI16 | $-RV14 - NIR$ | 0.626 | 0.679 | 0.053 |
| TGPVI21 | $\frac{NDSI(\beta_{SWIR1}, GEMI)}{NDSI(RV14, GEMI)}$ | 0.640 | 0.698 | 0.059 |
| TGPVI23 | $RSI(B - GEMI, RV14^2 \times b)$ | 0.648 | 0.710 | 0.062 |
| TGPVI3 | $NDSI(RV14, GEMI) \times RV14 \times \beta_{NIR}$ | 0.619 | 0.681 | 0.062 |

Tabla 4. Los diez mejores índices sintetizados por la metodología basada en PG.

Para obtener la tasa de erosión en la cuenca Todos Santos, es necesario construir mapas de C a partir de uno de los índices de vegetación, ya sea de los convencionales o los sintéticos. Generalmente los IVs poseen una escala numérica diferente a la escala numérica del factor C. Por ejemplo, la escala numérica para NDVI es de -1 a 1; mientras que el del factor C es de 0 a 1. Por ello es necesario aplicar una transformación lineal para hacer la conversión entre una escala y la otra. Esta conversión se hace utilizando la función *cftool* de MatLab, la cual trata de ajustar iterativamente una línea recta al conjunto de datos. Cuando se obtiene esta línea recta, se puede transformar el rango numérico del índice al rango numérico de C y viceversa. Esta transformación lineal se

utilizó para generar mapas de C a partir de los 3 mejores índices sintetizados por la PG y a partir del mejor índice del estado del arte. La definición del mejor índice sintético se encuentra en la sección 2.7.

En seguida se procede a obtener un mapa de erosión a partir del mejor índice. Se obtuvieron mapas de erosión usando el mapa de C a partir del índice TGPVI17 y del RVI4 (ver Figura 4). Ya que la comparación entre mapas sólo puede ser cualitativa, se ha diseñado un ejercicio con el que se puede comparar cuantitativamente la tasa de erosión punto por punto de los sitios de muestreo. Para calcular la tasa de erosión de la cuenca Todos Santos, se utilizó el modelo RUSLE. Actualmente RUSLE puede ser fácilmente implementado utilizando Sistemas de Información Geográfica (SIG), pues cada uno de los coeficientes se puede manejar como una capa de datos para SIG. Además, ya que RUSLE es un modelo multiplicativo, sencillamente basta con multiplicar cada una de estas capas pixel a pixel, y el resultado es un nuevo pixel, es decir, una nueva capa que contendrá la pérdida promedio de suelo por erosión: A . Para cada uno de los mapas de erosión que se presentan a continuación, los factores R, L, S, K y P de RUSLE se calcularon como se describe en el trabajo de [6]. En cambio, el factor C va cambiando de acuerdo al mapa de C del índice en turno.

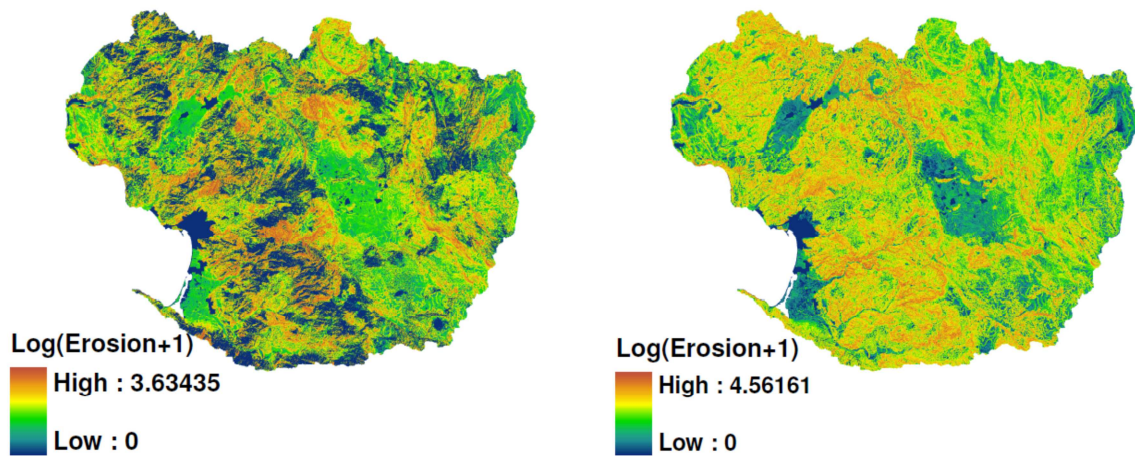


Fig. 4. Mapas del logaritmo de la tasa de erosión de la cuenca Todos Santos. a) Usado el factor C basado en el índice TGPVI17. b) Usando el factor C basado en RVI4.

Para un análisis cuantitativo es necesario observar la Tabla 5; la cual, en el primer renglón muestra el promedio de la tasa de erosión de los 67 sitios de muestreo de campo. Para este análisis, dicho renglón funge como punto de referencia o *ground truth*. De igual manera, en la misma tabla se establecen los promedios de la tasa de erosión calculados a partir de los valores de C derivados de los diferentes índices de vegetación. Nótese que las mejores aproximaciones a las mediciones de campo son las obtenidas de los índices de sintéticos. Por ejemplo, la tasa derivada del *TGPVI19* es de $76 \pm 104 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$ y muestra la mayor aproximación a la tasa de erosión obtenida usando el factor C calculado de los datos de campo; que es de $77 \pm 154 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$. Además, *TGPVI17* y *TGPVI27* también arrojan una tasa de erosión con una aproximación aceptable: 67 ± 97 y 69 ± 97 respectivamente. Finalmente, los índices tomados del estado del arte arrojan tasas de erosión promedio sobre-estimadas con respecto a los índices sintéticos y a los puntos de referencia. Por ejemplo, la tasa obtenida a partir del *RVI4* muestra una tasa de erosión aproximadamente un 50% más alta que la real: 120 ± 184 .

| Método utilizado para obtener C | Erosión ($\text{Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$) |
|--|--|
| Datos de campo (<i>ground truth</i>) | 76.6 ± 153.6 |
| TGPVI17 | 67.1 ± 97 |
| TGPVI19 | 76.5 ± 103.5 |
| TGPVI27 | 68.6 ± 97.4 |
| RVI4 | 119.7 ± 184.4 |
| RVI5 | 193.7 ± 393.3 |

Tabla 5. Media y desviación estándar de la tasa de erosión obtenida para los sitios de muestreo. Cada renglón representa una mapa del factor C diferente.

4. Conclusiones

En este documento se ha descrito una nueva metodología basada en programación genética, la cual sintetiza automáticamente índices de vegetación que tienen una buena correlación con el factor C del modelo de erosión RUSLE. Como resultado, los índices

sintetizados obtuvieron un desempeño mejor al de los índices convencionales reportados en la literatura, tales como el NDVI, GVI y un índice alternativo al RVI, llamado en este trabajo como RVI4. De entre los índices sintéticos obtenidos sobresalen los índices *TGPVI17*, *TGPVI27* y *TGPVI19*. Además, destaca el hecho de que la tasa de erosión a partir del *TGPVI19* es de $76 \pm 104 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$ y fue la que más se acercó al promedio de erosión obtenido usando el factor C medido en campo: $77 \pm 154 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$.

Basado en los resultados experimentales, es posible decir que el uso de programación genética es una buena alternativa para implementar análisis multi-espectral y otras aplicaciones de la percepción remota; tales como clasificadores e índices matemáticos que reconozcan y realcen la señal de diferentes rasgos en la superficie terrestre. Además, la ventaja de la PG es su propiedad de *caja blanca*; la cual consiste en que las soluciones generadas por la PG pueden ser analizadas, simplificadas e interpretadas por un científico o ingeniero experto en el área de aplicación.

El trabajo futuro está relacionado con la generalidad de la metodología desarrollada para identificar un buen índice para estimar C. En este trabajo se utilizaron datos de campo de una sola cuenca hidrográfica con características climatológicas específicas. Se pretende continuar aplicando la metodología bajo diferentes condiciones que permitan discernir sobre la generalidad de los resultados obtenidos.

Finalmente, nosotros pensamos que los resultados de este trabajo arrojan una novedosa metodología que contribuye al estado del arte en diferentes disciplinas científicas, tales como las geo-ciencias ambientales, la percepción remota y la inteligencia computacional. Esta contribución mejora la manera de obtener el factor C del modelo RUSLE y en consecuencia produce mapas de erosión más precisos, sin la necesidad de llevar a cabo extensas campañas de recolección de datos de campo; lo cual es un aliciente para continuar explorando el uso de la programación genética en aplicaciones de la percepción remota.

5. Referencias

- [1] A. Vrieling, Mapping Erosion from Space. Tesis de doctorado. 2007. Wageningen University. Wageningen, Netherlands. 152 pp.
- [2] S. de Jong, Applications of Reflective Remote Sensing for Land Degradation Studies in a Mediterranean Environment. Tesis de doctorado. 1994. Universiteit Utrecht, Netherlands. 237 pp.
- [3] K. Renard, G. Foster, G. Weesies, "Predicting Soil Erosion by Water: A Guide to Conservation Planning With the Revised Universal Soil Loss Equation". U.S. Department of Agriculture. Vol. 703. 1996.
- [4] J. Montoya, F. Francés, J. I. Vélez, P. Julien, "Desarrollo de un modelo distribuido de producción, transporte y depositación de sedimentos. Aplicación en una cuenca experimental". En XXII Congreso Latinoamericano de Hidráulica. 2006. Ciudad Guyana. Venezuela.
- [5] D. C. Flanagan, M. A. Nearing, "Water Erosion Prediction Project. Hillslope Profile And Watershed Model Documentation". USDA-ARS National Soil Erosion Research Laboratory. Vol. 10. 1995.
- [6] S. V. Smith, S. H. Bullock, A. Hinojosa-Corona, E. Franco-Viscaíno, M. Escoto.Rodríguez, T. G. Kretschmar, L. M. Farfán, J. M. Ceseña, "Soil erosion and significance for carbon fluxes in a mountainous mediterranean climate watershed". Ecological Applications. Vol. 17. No. 5. 2007. 1379-1387 pp.
- [7] R. G. Congalton, M. Balogh, C. Bell, K. Green, J. A. Milliken, R. Toman, "Mapping and monitoring agricultural crops and other land cover in the lower colorado river basin". Photogrammetric Engineering & Remote Sensing. Vol. 64. No. 11. 1998. 1107–1113 pp.

- [8] W. Rouse, R. H. Haas, J. A. Schell, D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS". Third ERTS Symposium, NASA SP-351. Vol. 1. 1973. 309-317 pp.
- [9] W. Wishmeier, D. Smith, Predicting Rainfall Erosion Losses: A Guide to Conservation Planning. U. S. Department of Agriculture. Vol. 537. 1978.
- [10] H. Bauer, "The statistical analysis of chaparral and other plant communities by means of transect samples". *Ecology*. Vol. 24. 1943. 45-60 pp.
- [11] B. Zippin, J. M. Vanderwier, Scrub community descriptions of the Baja California peninsula, Mexico. *Madroño*. 1994. 85–119 pp.
- [12] M. A. González Botello, Estimaciones de la cobertura vegetal y del suelo en el noroeste de Baja California y su aplicación a la modelación de la erosión. 2010. Tesis de maestría. Universidad Autónoma de Baja California. Facultad de Ciencias. Ensenada, Baja California.
- [13] C. F. Jordan. "Derivation of leaf area index from quality of light on the forest floor". *Ecology*. Vol. 50. 1969. 663-666 pp.
- [14] B. Pinty, M. M. Verstraete, "GEMI: A Non-Linear Index to Monitor Global Vegetation from Satellites". *Vegetatio*. Vol. 101. 1991. 15-20 pp.
- [15] R. E. Crippen, "Calculating the Vegetation Index Faster". *Remote Sensing of Environment*". Vol 34. 1990. 71-73 pp.
- [16] J. Qi, Y. Kerr, A. Chehbouni, "External Factor Consideration in Vegetation Index Development". En *Proc. of Physical Measurements and Signatures in Remote Sensing ISPRS*. 1994. 723-730 pp.
- [17] S. Khana, A. Palacios-Orueta, M. L. Whiting, S. L. Ustin, D. Riaño, J. Litago, "Development of angle indexes for soil moisture estimation, dry matter detection

and land-cover discrimination". *Remote Sensing of Environment*. No. 109. 2007. 154–165 pp.

- [18] E. P. Crist, R. C. Cicone, "Application of the tasseled cap concept to simulated thematic mapper data". *Photogrammetric Engineering and Remote Sensing*. Vol. 50. 1994. 343-352 pp.
- [19] F. F. Sabins, *Remote sensing: principles and interpretation*. 3° edición. 1997. Freeman, New York. U.S.A.
- [20] J. R. Koza. *Genetic programming: On the programming of computers by means of natural selection*. 1992. MIT press. Cambridge, MA. USA.
- [21] R. Poli, W. B. Langdon, N. F. McPhee, *A field guide to genetic programming*. Lulu. U.K. 2008.

7. Autores

Dr. Cesar Augusto Puente Montejano es profesor-investigador de tiempo completo de la Facultad de Ingeniería de la Universidad Autónoma de San Luis Potosí. Obtuvo el grado de Doctorado en Ciencias, en Ciencias de la Tierra por el Centro de Investigación Científica y de Educación Superior de Ensenada, B.C. (CICESE) en Junio de 2011. Sus intereses de investigación radican en aplicaciones del cómputo evolutivo en problemas de las ciencias de la Tierra.

Dr. Gustavo Olague Caballero es investigador científico en el Centro de Investigación Científica y de Educación Superior de Ensenada, B.C. (CICESE), en la División de Física Aplicada. Obtuvo su grado de *Diplôme de Doctorat en Imagerie, Vision et Robotique* por el Institut National Polytechnique de Grenoble, Francia.

Hacia la clasificación de fonocardiogramas utilizando descriptores caóticos y estadísticos

Carlos Antonio Osorio Maceda

Tecnológico de Monterrey, Campus Puebla,
Vía Atlixcáyotl 2301, Reserva Territorial Atlixcáyotl, C.P. 72431, Puebla, México

Hugo G. González-Hernández

Tecnológico de Monterrey, Campus Puebla,
Vía Atlixcáyotl 2301, Reserva Territorial Atlixcáyotl C.P. 72431, Puebla, México
hgonz@itesm.mx

Resumen

En este trabajo, tres descriptores: la dimensión de correlación, el exponente de Lyapunov y la entropía aproximada, son calculados a partir de series de tiempo procedentes de mediciones de fonocardiogramas, con el propósito de entrenar una red neuronal auto-organizada para clasificar afecciones cardíacas. El uso de la red neuronal, en conjunto con descriptores caóticos y estadísticos, muestran un buen desempeño y capacidad para lograr una segmentación entre clases.

Palabra(s) Clave(s): análisis de series de tiempo, entropía aproximada, fonocardiograma, reconstrucción del atractor, red neuronal auto-organizada, teoría de caos.

1. Introducción

En la actualidad el ser humano se ha enfocado en desarrollar áreas científico-técnicas que le permita generar procesos multidisciplinarios en donde las bases del conocimiento no permanezcan afines a una sola área. El tratar a los problemas con

este enfoque, permite contar con un mayor número de herramientas para el tratamiento o análisis de sistemas, tal es el caso del estudio de las señales biológicas, las cuales a lo largo de los últimos años han recibido un gran interés debido a su comportamiento irregular en el tiempo, lo que conlleva a que se trate de un sistema dinámico altamente impredecible.

El interés en su estudio ha derivado en la generación de conocimiento y herramientas que tratan de caracterizar y cuantificar cada una de las propiedades que conforman a las señales. De esta manera surgen algoritmos de análisis que buscan darle solución a ello, permitiendo así contar con un estudio en tiempo real de las mismas. Trabajos como los reportados en [1], [2], y [3] describen resultados en el análisis de señales del corazón, pulmón y cerebro en donde han sido empleadas técnicas estocásticas, Teoría de Caos y análisis en tiempo y frecuencia para modelar y extraer características de las mismas.

Dentro de los resultados encontrados en el análisis de señales biológicas se encuentran herramientas derivadas del estudio del Caos que, en conjunto con paradigmas de aprendizaje, han demostrado un porcentaje de efectividad elevado en la obtención y clasificación de parámetros para la caracterización de señales. Trabajos como los reportados en [4] y [5] han expuesto una comparación entre distintos métodos para la clasificación de cardiopatías y cuyos resultados arrojan una tendencia a lograr un mejor resultado al utilizar técnicas como máquinas de soporte vectorial y redes neuronales en conjunto con elementos caóticos. Estos hechos abren la oportunidad para la aplicación de análisis como [6] y [7], en la implementación de un algoritmo capaz de clasificar señales biológicas, específicamente del corazón, para la prevención o tratamiento oportuno de enfermedades cardíacas.

Este trabajo aborda el estudio de señales del corazón, particularmente fonocardiogramas, para su clasificación, mediante la extracción de descriptores caóticos y de uso de una red neuronal auto-organizada, con la finalidad de identificar afecciones cardíacas.

Este documento se encuentra organizado de la siguiente forma: la sección número dos presenta información técnica acerca de la adquisición de las señales del corazón y las afecciones que son el objeto de estudio. Así mismo, se introduce los algoritmos de Grassberger-Procaccia [8] y Rosenstein [9] para el cálculo de la dimensión de correlación y el exponente de Lyapunov respectivamente, se describe la Entropía Aproximada [10] y se detalla cómo se utilizan estos descriptores para entrenar la red neuronal. Los resultados obtenidos de la aplicación de la red son analizados en la sección tres. Finalmente se presentan algunas conclusiones.

2. Desarrollo

2.1. Adquisición de señales

Para el objeto de estudio se han seleccionado dos bases de datos provenientes de [11], las cuales se encuentran conformadas por archivos de audio cuya longitud varía entre 1 a 30 segundos y proceden de un estudio clínico en hospitales en el que se empleó un estetoscopio digital llamado DigiScope. A su vez, estos archivos se encuentran segmentados en dos categorías: señales de corazón normales y con murmullo.

La primera categoría se trata de sonidos provenientes de corazones sanos y se conforma por 200 muestras, mientras que la segunda categoría contiene 92 modelos de sonidos inusuales en el ciclo del latido del corazón, cuyas características se explican en [12]. Ambas categorías han sido obtenidas de niños y adultos en diferentes estados de relajación y excitación mediante la técnica de auscultación, descrita en [13], y pueden contener ruidos derivados de la respiración o del ambiente en el que se realizan las pruebas. Algunos ejemplos de las señales a analizar son mostrados en (Fig. 1).

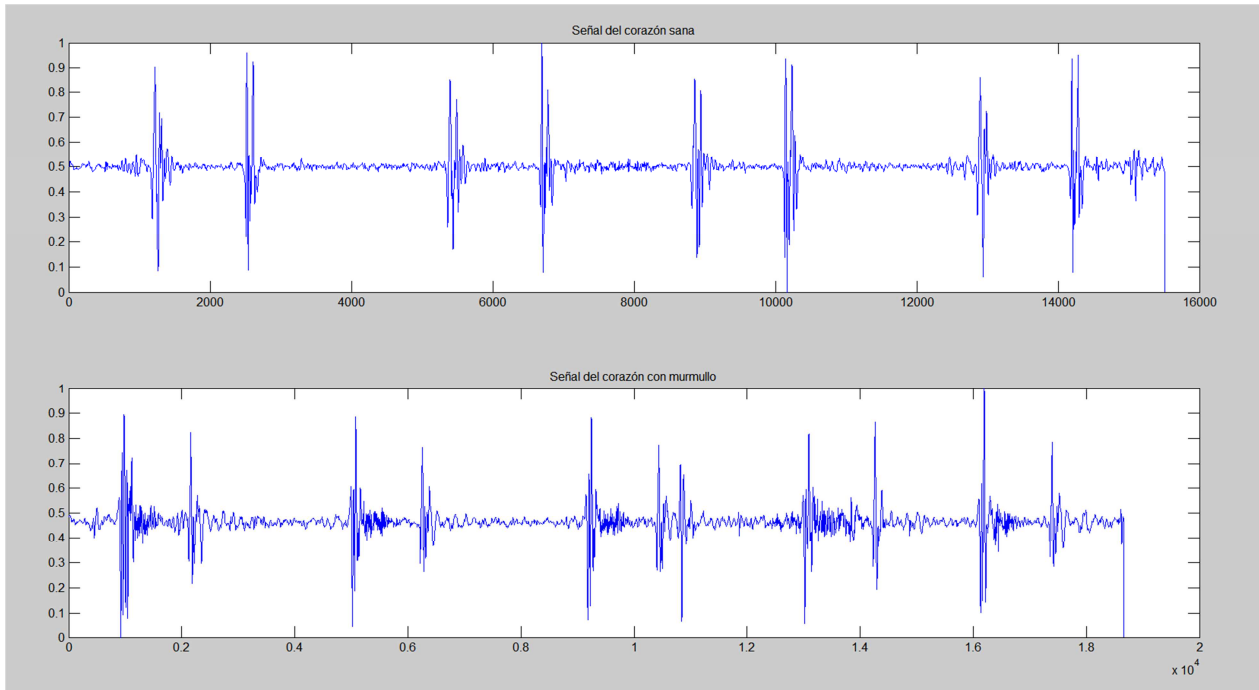


Fig. 1. Señal de corazón sana y con murmullo.

2.2. Reconstrucción del atractor

Antes de aplicar el algoritmo de Grassberger-Procaccia [8] para calcular el exponente de Lyapunov, es necesario reconstruir la dinámica del espacio de estados a partir de la serie de tiempo. Para ello se hace uso del método Coordenadas Retrasadas [14], en donde se crean vectores D -dimensionales con retrasos de la señal original:

$$y(n) = [x(n), x(n + \tau) \dots, x(n + (D - 1)T)] \quad (1)$$

En donde T describe la separación en múltiplos del tiempo de muestreo entre los valores de la serie de tiempo que son utilizados como coordenadas para cada vector y que se denota como el tiempo de retraso, y D es la dimensionalidad de los vectores reconstruidos, llamada dimensión de empotramiento. Para calcular estos parámetros en (1), se hace empleo de los algoritmos de Falsos Vecinos Cercanos y Promedio de Información Mutua [14]. Obteniendo de esta manera para la categoría de señales de corazón sanas $D = 4$ y para la categoría de señales con murmullo $D = 3$. Mientras que

el tiempo de retraso en la primera categoría varía entre $T = 19$ a $T = 21$ y para la segunda se obtienen valores entre $T = 14$ a $T = 20$.

2.3. Dimensión de correlación

Otorga un significado físico a las regiones de un atractor al resaltar las órbitas que son visitadas con mayor frecuencia, obteniendo así una cuantificación precisa de la naturaleza fractal del atractor [15]. Para ello, se construye una función $C(r)$ que denotará la probabilidad de que dos puntos, seleccionados arbitrariamente de la trayectoria de la reconstrucción del atractor, se encuentren en el vecindario de un elemento volumétrico D -dimensional con radio r [8]. Esta función es entonces normalizada de tal forma que cuando r es lo suficientemente grande, se obtiene $C(r) = 1$. Al asumir que se toman N puntos del espacio de estados de la trayectoria, se tiene que:

$$C(r) = \frac{2}{N(N-1)} \sum_{j=1}^N \sum_{i=j+1}^N \theta(r - r_{ij}) \quad (2)$$

En la ecuación (2) θ representa la función Heaviside y r es la separación entre los puntos i y j que se calcula a través de la norma Euclidiana. Al graficar $\log(C(r))$ contra $\log(r)$ se produce una línea recta cuya pendiente en los límites de valores pequeños de r y valores grandes de N es la dimensión de correlación:

$$D_2 = \lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} \frac{d \log(C(r))}{d \log(r)} \quad (3)$$

2.4. Exponente de Lyapunov

Representa una medida cuantitativa de la divergencia exponencial de las trayectorias que indica la sensibilidad o dependencia de los estados futuros de un sistema bajo ciertas condiciones iniciales [9]. Para un sistema determinista, si su valor es positivo entonces, dos trayectorias que en un inicio se encontraban muy próximas, se separan en el tiempo, indicando comportamiento caótico en el sistema.

De esta manera una perturbación infinitesimal inicial crecerá exponencialmente y, a la razón de crecimiento que tenga, se le conoce como exponente de Lyapunov:

$$\lambda_i = \frac{1}{KT} = \sum_{k=1}^K \ln \|\delta y_i^k\| \quad i = 1, \dots, n \quad (4)$$

Donde λ_i es el i -ésimo exponente de Lyapunov, T es el tiempo entre muestras, y_i es la i -ésima componente de y y δy es la desviación en la trayectoria de dos condiciones iniciales arbitrariamente cercanas.

Rosenstein [9] desarrolló un algoritmo para el cálculo del máximo exponente de Lyapunov a partir de una serie de tiempo. El primer paso consiste en la reconstrucción del atractor. Después de esto se calcula el vecino más cercano para cada uno de los vectores que representa la trayectoria del atractor y el máximo exponente de Lyapunov es estimado como la tasa media de las separaciones de estos vecinos. Al asumir una separación determinada por el máximo exponente de Lyapunov (λ), entonces al tiempo t , la divergencia promedio será:

$$d(t) \sim C e^{\lambda(i\Delta t)} \quad (5)$$

En donde $d(t)$ es el promedio de la divergencia de la trayectoria en el tiempo t , C es una constante de normalización que multiplica a la separación inicial entre trayectorias del atractor. Partiendo de (5), se llega a que la distancia obtenida de los vecinos cercanos será igual a la distancia mínima inicial.

$$d_j \approx C_j e^{\lambda(i\Delta t)} \quad (6)$$

Por lo que realizando un ajuste por mínimos cuadrados se busca la línea p que mejor se ajuste, obteniendo así el máximo exponente de Lyapunov.

$$p(i) = \frac{1}{\Delta t} \ln d_j(i) \quad (7)$$

2.5. Entropía aproximada

Es un indicador no lineal que permite cuantificar el grado de complejidad de un sistema a partir de una serie de tiempo. Permite discriminar señales en función de su regularidad sin tener en cuenta el modelo de sistema que las generó, en consecuencia, es independiente de la naturaleza de la señal [10]

La Entropía Aproximada se calcula al medir la semejanza logarítmica entre patrones consecutivos de longitud m que pertenecen a una secuencia $x(n)$ de N muestras y que tengan una diferencia menor que r veces la desviación estándar de $x(n)$.

De esta manera, al tener N muestras de la serie $x(n)$, para calcular la entropía aproximada primero se generan $N-m+1$ patrones denominados $X(1) \dots X(N-m+1)$, con $X(i) = [x(i), x(i+1), \dots, x(i+m-1)]$ donde $i = 1, 2, \dots, N-m+1$, representando m valores consecutivos de $x(n)$ y se obtiene la frecuencia de patrones similares de longitud m con tolerancia r a través de:

$$C_r^m(i) = \frac{d[X(i), X(j)] \leq r}{N-m+1}, j = 1, 2, \dots (N - m + 1), \forall j \neq i \quad (8)$$

Donde $d[X(i), X(j)]$ cuantifica la cantidad de patrones similares separados j veces un máximo de m muestras que tienen una diferencia menor a la tolerancia r , así:

$$d[X(i), X(j)] = \max_{k=1, 2, \dots, m} (x(i+k-1) - x(j+k-1)) \quad (9)$$

Se calcula el logaritmo de cada $C_r^m(i)$ y se promedia para todas las muestras N :

$$C_r^m = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_r^m(i) \quad (10)$$

Finalmente, se aumenta la dimensión en $m+1$ y se calcula nuevamente (10). Al ser un número finito, la entropía aproximada se obtiene mediante:

$$ApEn(x(n), m, r, N) = C_r^m - C_r^{m+1} \quad (11)$$

2.6. Clasificación de la información

En esta sección se aborda el problema de la clasificación de las dos clases de señales del corazón con el uso de una red neuronal auto-organizada. El objetivo es dividir el espacio de estados que contiene las características en regiones separadas para cada una de las diferentes clases de señales del corazón a través de un vector de clases para el entrenamiento de la red. Para ello se crea un vector $\mathbf{g} = [g_1, g_2, g_3]$ con tres componentes:

g_1 : Dimensión de Correlación

g_2 : Máximo exponente de Lyapunov

g_3 : Entropía aproximada

Este vector es empleado como entrada para el entrenamiento de la red neuronal de la (Fig. 2) El funcionamiento consiste en inicializar los pesos entre capas con valores pequeños y a través de un aprendizaje no supervisado y competitivo se ajustan los pesos para cada uno de los patrones de entrada, siendo que la neurona con el menor valor de la función discriminante será la que determine la ubicación espacial de la topología del vecindario que contiene a las neurona excitadas, proporcionando así la base para la cooperación entre las neuronas vecinas. Finalmente, las neuronas excitadas se adaptan al ajustar sus pesos de conexión en relación a la neurona ganadora, de tal forma que la respuesta a un patrón de entradas similares es alcanzado.

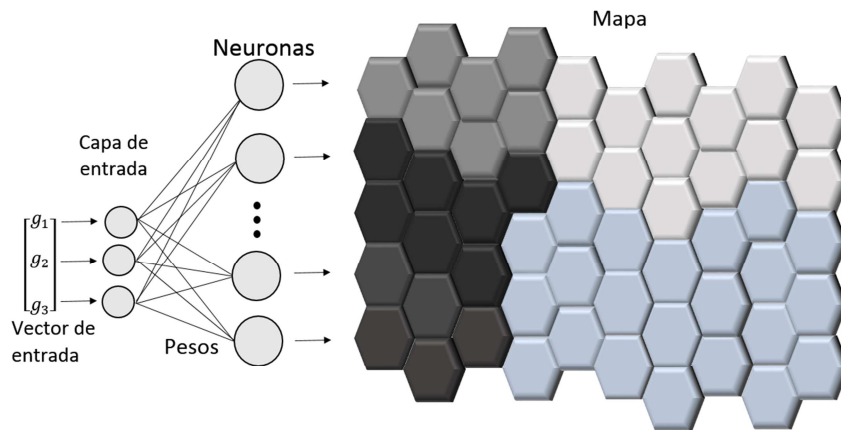


Fig. 2. Esquema de red neuronal auto-organizada.

El esquema de la red neuronal en la (Fig. 2) fue implementado utilizando Matlab® R13b y su Toolbox para redes neuronales. Se seleccionó un total de 25 neuronas para la capa oculta y a continuación se presentan los resultados obtenidos.

3. Resultados

Los resultados de la aplicación de los algoritmos de la sección III y clasificación IV son presentados a continuación. Primero, se muestran los resultados en la extracción de los descriptores, comenzando por la reconstrucción de Atractor para la obtención de la dimensión de correlación y máximo exponente de Lyapunov, ejemplo de esta reconstrucción se pueden observar de la (Fig. 3) a la (Fig. 8), seguido a esto, se prosigue con un análisis de los resultados del parámetro conocido como Entropía Aproximada. Finalmente, en la a (Fig. 11) y (Fig. 12) se presentan los resultados de la clasificación.

3.1. Extracción de descriptores

La reconstrucción del atractor para la obtención de la dimensión de correlación arrojó que para la señal de corazón sana los valores oscilan entre 1.4215 a 1.8081, mientras

que para la señal con murmullo en el corazón se obtuvo un rango entre 1.5547 a 1.8785.

El análisis para la obtención del máximo exponente de Lyapunov muestra que para las señales de corazón sano se tiene un rango de valores mínimo y máximo entre 0.01367 a 0.05041; y para señales con murmullo se tiene que el valor máximo es 0.04664 y el valor mínimo es de 0.00666.

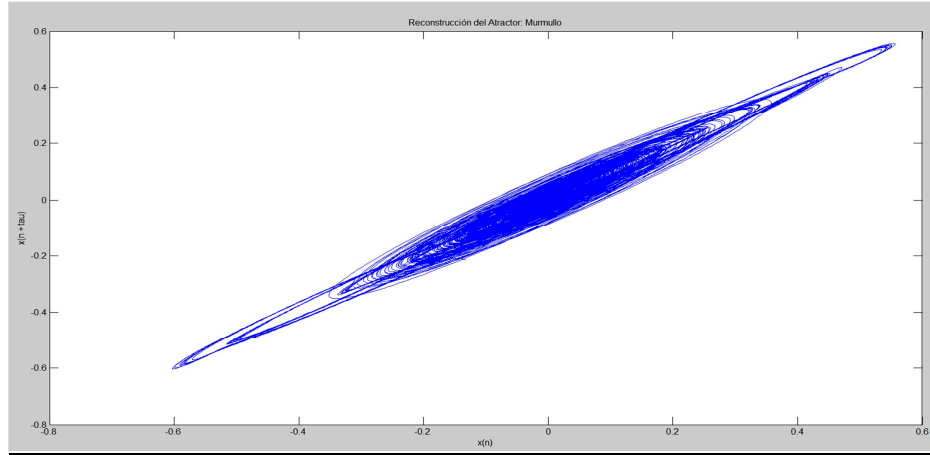


Fig. 3. Reconstrucción del atractor, señal del corazón con murmullo:

Proyección plano X-Y.

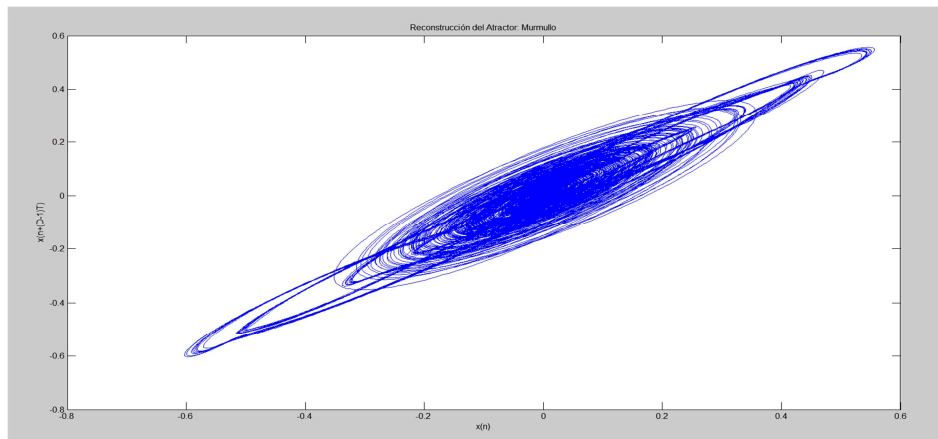


Fig. 4. Reconstrucción del atractor, señal del corazón con murmullo: Proyección plano X-Z.

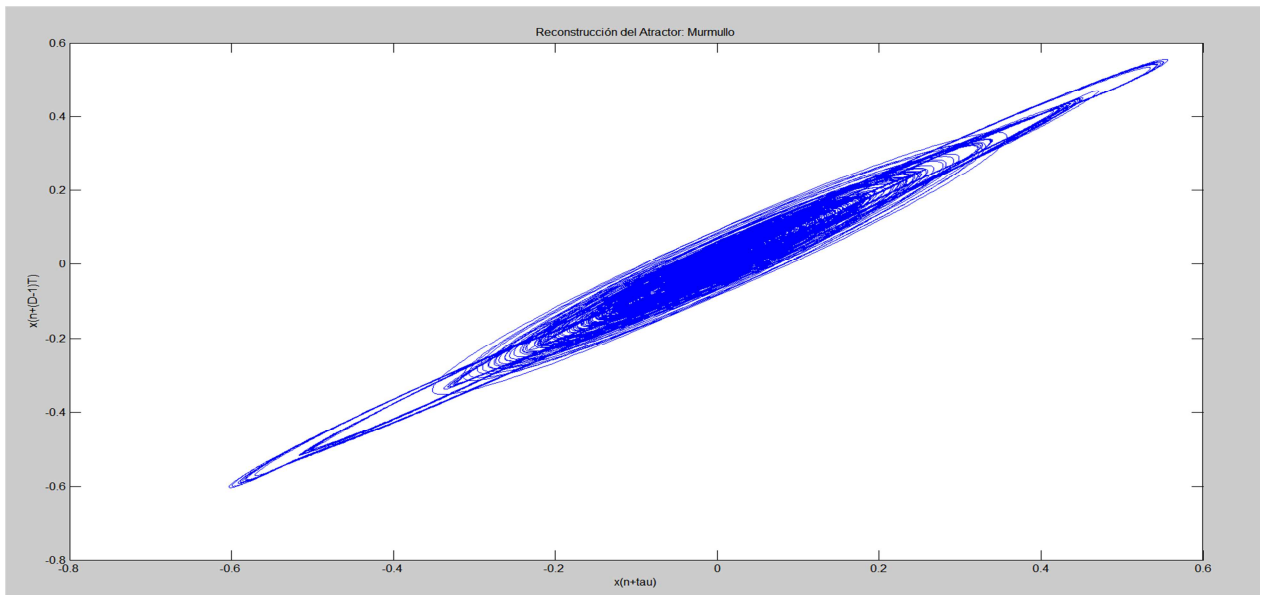


Fig. 5. Reconstrucción del atractor, señal del corazón con murmullo: Proyección plano Y-Z.

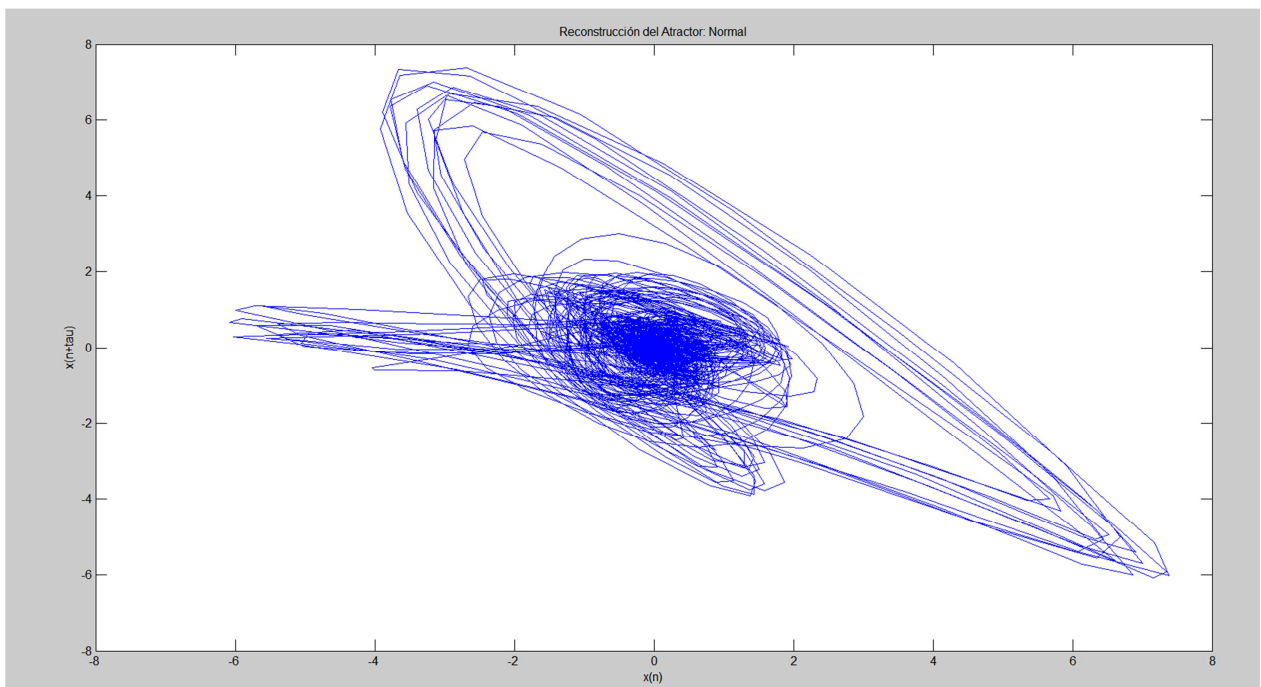


Fig. 6. Reconstrucción del atractor, señal del corazón sano: Proyección plano X-Y.

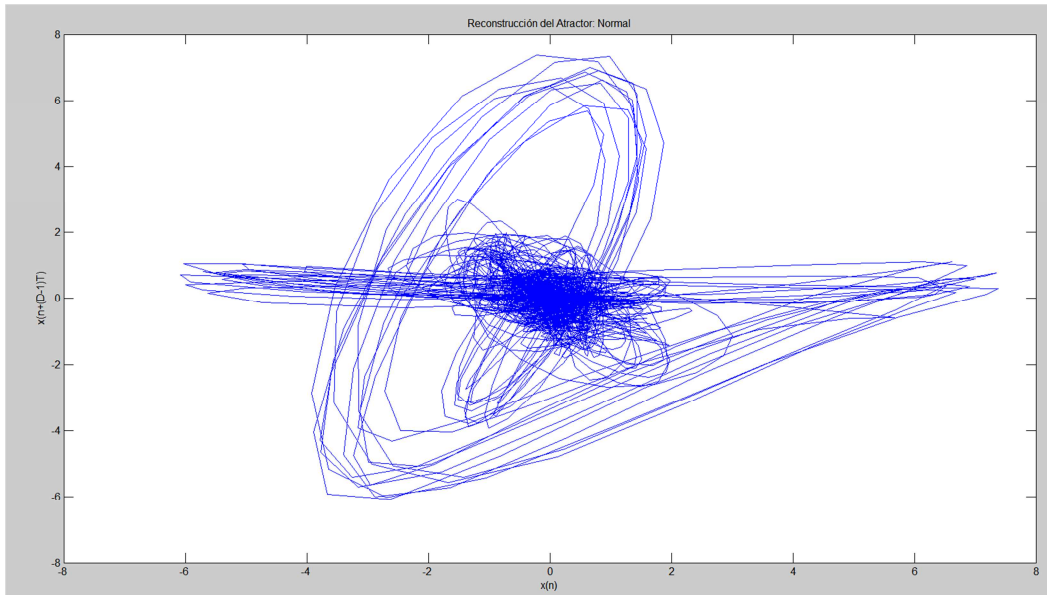


Fig. 7. Reconstrucción del atractor, señal del corazón sano: Proyección plano X-Z.

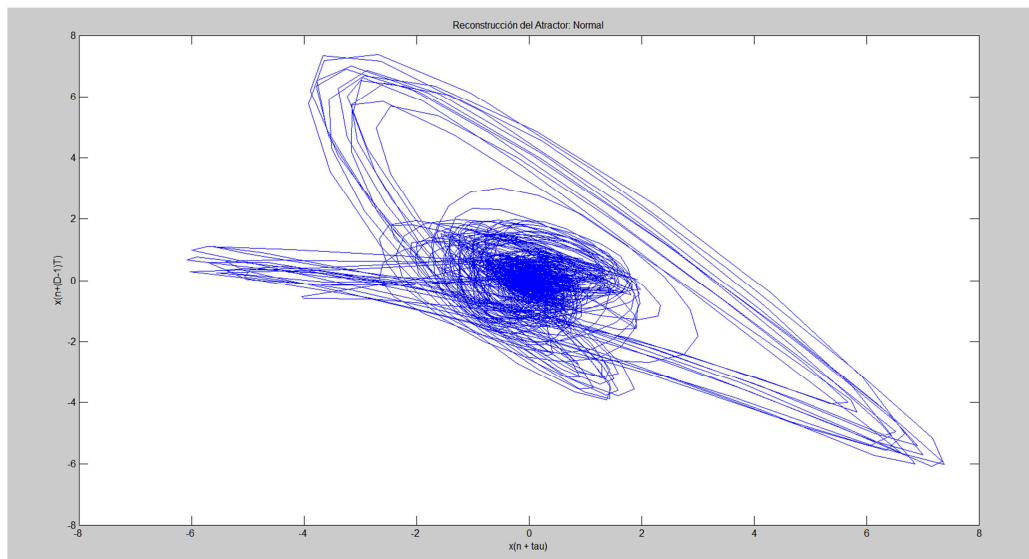


Fig. 8. Reconstrucción del atractor, señal del corazón sano: Proyección plano Y-Z.

El método de entropía aproximada muestra una variación para señales de corazón sano entre 0.32489 a 0.92903 y para señales con murmullo se tiene un rango entre 0.38198

a 0.95668. En la (Fig. 9) y (Fig. 10) se muestra la distribución de estos descriptores para todas las muestras analizadas.

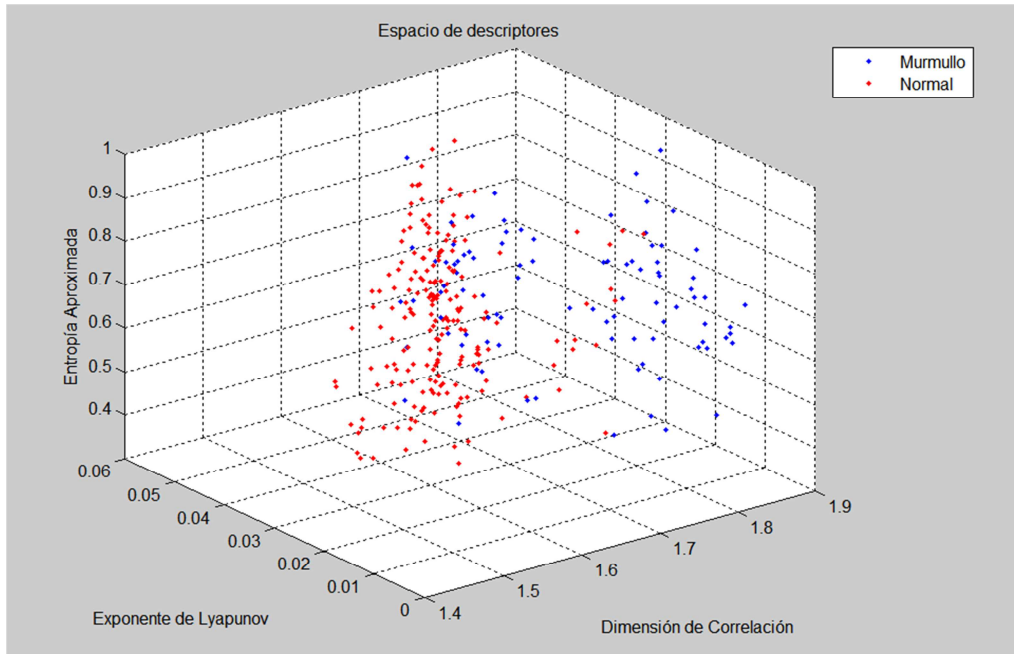


Fig. 9. Espacio de descriptores.

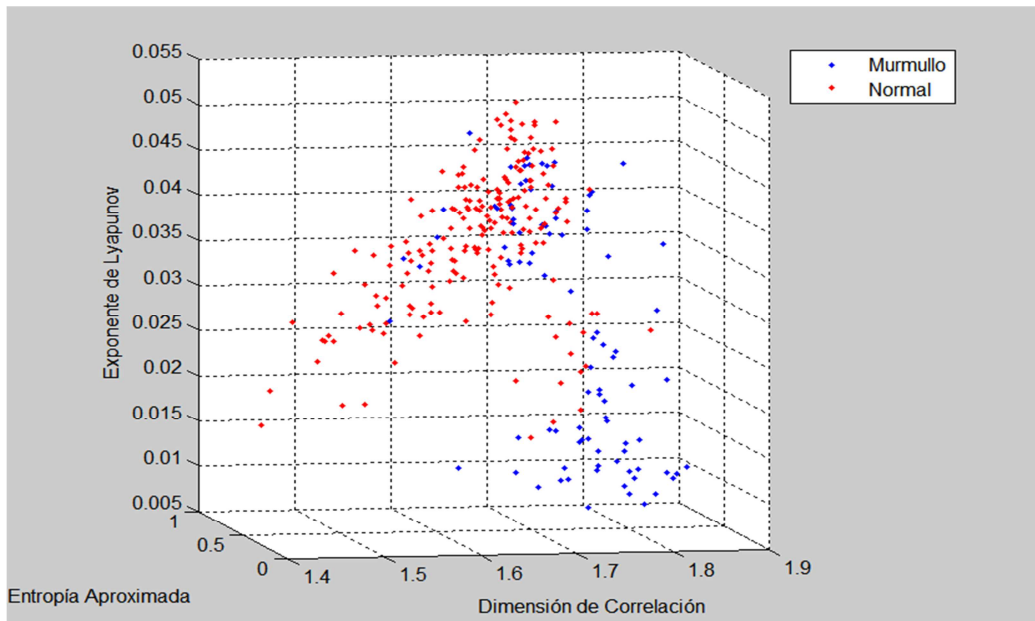


Fig. 10. Espacio de descriptores.

3.2. Clasificación

Los resultados del entrenamiento de la red neuronal se muestran en la (Fig. 11) y (Fig. 12). En la (Fig. 11) se aprecia el plano de los pesos para cada una de las entradas del vector \mathbf{g} , esta figura es una representación de la forma en que se distribuyen cada una de los pesos que conecta a cada una de las entradas con las neuronas que conforma la red neuronal, si el patrón de entrada es muy similar, entonces las entradas se encuentran altamente correlacionadas. De esta forma, se puede apreciar que para la entrada g_1 y g_2 se logra encontrar una similitud entre entradas que denota una segmentación entre las clases de señal del corazón sano y con murmullo, mientras que para la entrada g_3 , las entradas no generan una aglomeración que permita denotar una correlación entre clases, lo que ocasiona errores en el proceso de segmentación. Así mismo, en la (Fig. 12) se visualizan las distancias entre las conexiones de los pesos con las neuronas que generan la red neuronal, algunas de estas muestras se encuentran demasiado cercanas una de otras, lo que contribuye a obtener errores en la clasificación de las clases.

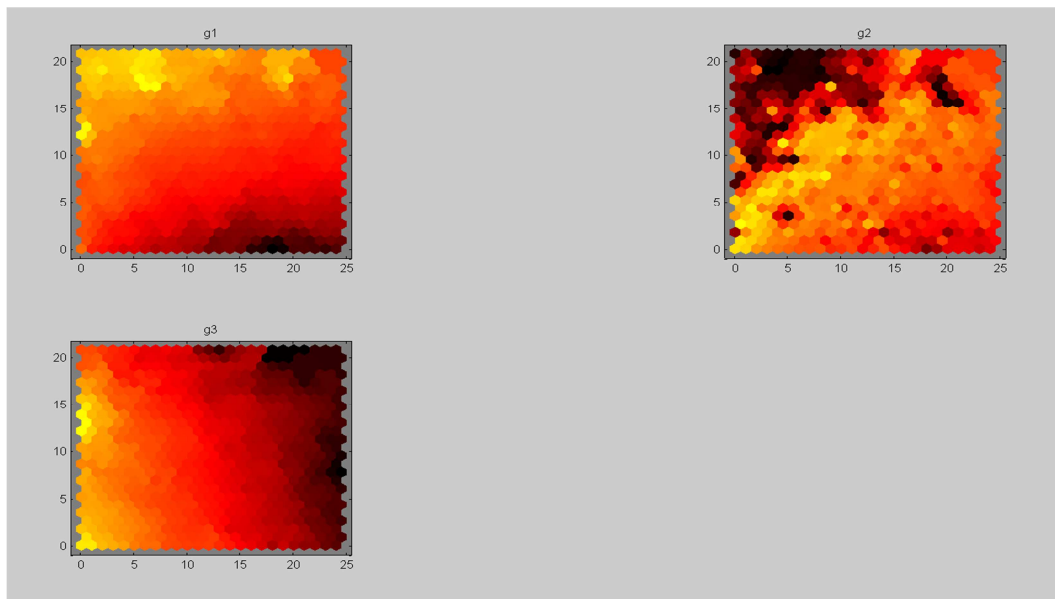


Fig. 11. Distribución de los pesos.

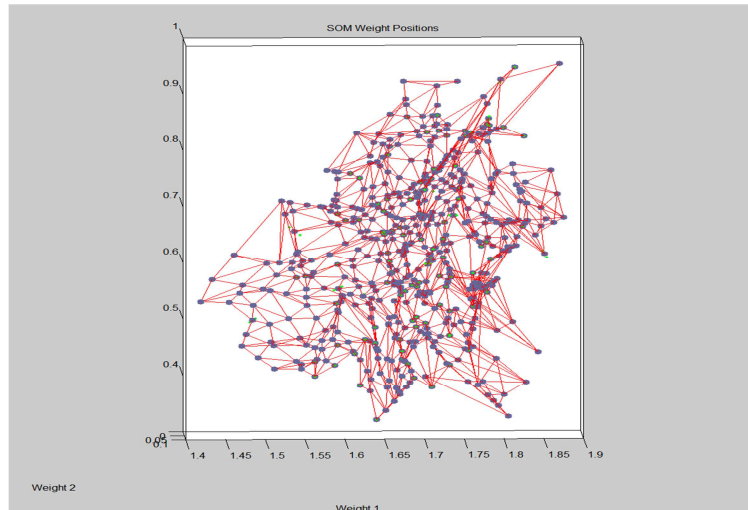


Fig. 12. Posición de los pesos.

A fin de mostrar el desempeño del clasificador, en la (Fig. 13) se presenta la gráfica Característica Operativa del Receptor (**ROC**), calculando el área bajo la curva (**AUC**). Se observa que para la clase normal se tiene un valor igual a 0.8334 el cual indica que la precisión en la clasificación es aceptable para esta clase, sin embargo, al obtener el segundo valor **AUC** se tiene un valor de 0.6053 que denota una precisión regular en la identificación de esta clase.

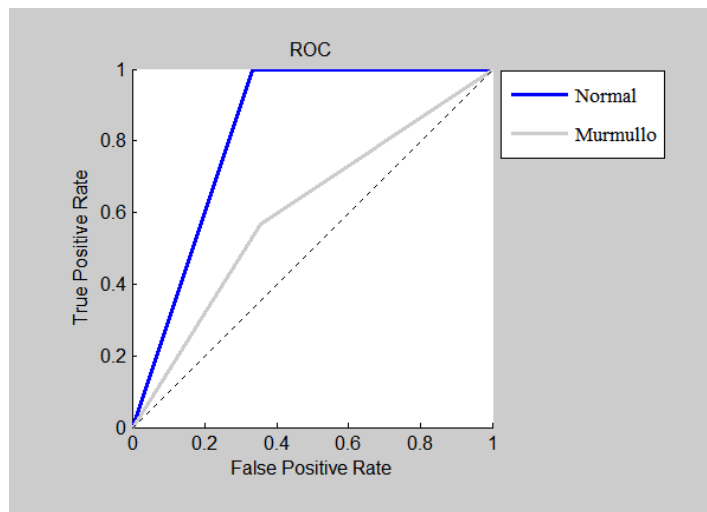


Fig. 13. Característica Operativa del Receptor.

4. Conclusiones

En este trabajo preliminar se han obtenido tres descriptores para la caracterización de dos clases de señales provenientes del corazón. La dimensión de correlación, el exponente de Lyapunov y la entropía aproximada fueron calculados para probar un sistema de clasificación basado en estas características.

Aunque los resultados muestran una segmentación en el entrenamiento de la red neuronal, la eficiencia en los resultados de clasificación demuestra que existen aún problemas de discriminación entre los descriptores que caracterizan las señales del corazón, por lo que el introducir y/o cambiar un descriptor podría mejorar la eficiencia y precisión en la clasificación. Así mismo, se abre la oportunidad de emplear esta técnica en aplicaciones en tiempo real para el reconocimiento de afecciones cardíacas, no solamente murmullo en fonocardiogramas.

5. Referencias

- [1] S. Schulz, K. Bär, A. Voss. Analyses of Heart Rate, “Respiration and Cardiorespiratory Coupling in Patients with Schizophrenia”. *Entropy*. Vol. 17. 2015. 483-501 pp..
- [2] C. Hoang, P. Stein, “Time frequency analysis of heart rate variability with chaos theory”. 13th International Conference on Computational Science and Its Applications. 2013. 174-177 pp.
- [3] B. Raghavendra. D. Dutt, “Multiscale fractal dimension technique for characterization and analysis of biomedical signals”. *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*. IEE. 2011. 370 – 374 pp.

- [4] N. Bhaskar, "Performance Analysis of Support Vector Machine and Neural Network in Detection of Myocardial Infarction". *Procedia Computer Science*. Vol. 46. 2015. 20-30 pp.
- [5] G. Redlarski, D. Gradolewski y A. Palkowski, "A System for Heart Sounds Classification". US National Library of Medicine National Institutes of Health. Ed. Alena Talkachova. 2014.
- [6] K. Balasubramanian, S. Nair y N. Nagaraj, "Classification of periodic, chaotic and random sequences using approximate entropy and Lempel-Ziv complexity measures". *Pramana, journal of physics*. Vol. 84. 2015. 365-372 pp.
- [7] D. Baoxiang, X. Wei, S. Bingbing, D. Qun y C. Shu-chuan, "Prediction of Chaotic Time Series of RBF Neural Network Based on Particle Swarm Optimization". *Intelligent Data Analysis and Its applications*. Vol. 2. 2014. 489-497 pp.
- [8] P. Grassberger y I. Procaccia, "Characterization of strange attractors". *Physical Review Lett*. Vol. 50. No. 346. 1983.
- [9] M. Rosenstein, J. Collins y C. Luca, "A practical method for calculating largest lyapunov exponents from small data sets". *Physica D*. Vol. 65. 1993. 117-34 pp.
- [10] M. Steven, Pincus y A. Goldberger, "Modelling in physiology. The american Physiological Society". *Cadiovascular Division*. No. 266. 1994.
- [11] The Pascal Classifying Heart Sounds Challenge. Disponible en: <http://www.peterjbentley.com/heartchallenge/index.html>. 2011.
- [12] P. O'Gara y J. Loscalzo, "Approach to the Patient with Heart Murmur". *Cardinal Manifestations and Presentation of Diseases*. 2012. McGraw-Hill. 13-10 pp.
- [13] T. Poari, H. Owaitha, "Heart and Lung Auscultation. CDHB Clinical Skills Unit. Canterbury". 2013. 1-30 pp.

- [14] H. Abarbanel, "Analysis of observed chaotic data". Springer-Verlag. 1995. 13-65 pp.
- [15] J. Sprott, "Chaos and Time-Series Analysis". 2003. Oxford University Press.
- [16] A. Casiano, H. González, "Automatic classification of EEG epileptic patterns using chaotic descriptors". Congreso Nacional de Control Automático. 2005.

6. Autores

Ing. Carlos Antonio Osorio Maceda es Ingeniero en Tecnologías electrónicas con concentración en Robótica por el Tecnológico de Monterrey, Campus Puebla. Actual estudiante de la Maestría en Sistemas Inteligentes en la misma institución, sus intereses de investigación involucran análisis de bioseñales y reconocimiento de patrones.

Dr. Hugo G. González-Hernández obtuvo el grado de Doctor en Ciencias con especialidad en Mecatrónica por el CINVESTAV-IPN. Su trabajo de investigación se enfoca en el análisis y control de sistemas no lineales, robótica y visión artificial. Actualmente es profesor del Tecnológico de Monterrey, Campus Puebla.

Activación sistema de frenos ABS usando visión por computadora para la detección de la superficie de rodamiento

Gabriel García Ponds

Universidad Autónoma Metropolitana-Azcapotzalco, Departamento de Electrónica, San Pablo Xalpa
No.180, Col. Reynosa Tamaulipas, C.P. 02200, México, D.F.
gabrielgponds@gmail.com

Juan Villegas Cortez

Universidad Autónoma Metropolitana-Azcapotzalco, Departamento de Electrónica, San Pablo Xalpa
No.180, Col. Reynosa Tamaulipas, C.P. 02200, México, D.F.
juanvc@azc.uam.mx

Carlos Avilés Cruz

Universidad Autónoma Metropolitana-Azcapotzalco, Departamento de Electrónica, San Pablo Xalpa
No.180, Col. Reynosa Tamaulipas, C.P. 02200, México, D.F.
caviles@azc.uam.mx

Iván Vázquez Álvarez

Universidad Autónoma Metropolitana-Azcapotzalco, Departamento de Electrónica, San Pablo Xalpa
No.180, Col. Reynosa Tamaulipas, C.P. 02200, México, D.F.

Ismael Osuna Galán

Universidad Politécnica de Chiapas, Tuxtla Gtz., Chiapas, México
iosuna@upchiapas.edu.mx

Yolanda Pérez Pimentel

Universidad Politécnica de Chiapas, Tuxtla Gtz., Chiapas, México
ypimentel@upchiapas.edu.mx

Resumen

Los frenos tipo ABS fueron desarrollados en aviones, debido a las exigentes condiciones de frenado que requieren, posteriormente fueron adoptados por la industria automotriz, hasta llegar a convertirse en equipo de seguridad obligatorio en muchos países. Actualmente los frenos ABS trabajan en conjunto con otros sistemas de seguridad activa, pero su función es permitir que el conductor mantenga el control del vehículo durante una frenada de emergencia. La idea del sistema presentado es detectar una superficie de rodamiento que implique riesgo en caso de una frenada de emergencia de forma que el sistema ABS se active antes de que exista una condición de bloqueo de alguna de las ruedas. En este artículo se presenta un algoritmo que permite la activación del sistema de frenos ABS, usando visión por computadora para analizar la superficie de rodamiento por medio de la velocidad, entropía y luminosidad; así como de un conjunto de reglas difusas, para detectar una situación de riesgo antes de que se detecte de forma convencional. Los resultados alcanzados de las simulaciones realizadas en laboratorio son muy satisfactorias y prometedoras.

Palabra(s) Clave(s): ABS, control difuso, visión por computadora.

1. Introducción

Un sistema típico ABS consiste en los frenos propios del vehículo, sensores de velocidad en cada rueda, una bomba hidráulica de frenado y una unidad de control, algunos sistemas avanzados incluyen acelerómetros para proveer mayor información a esta unidad. Las investigaciones actuales están dirigidas a mejorar la tecnología ABS apoyando la respuesta del sistema [1][2]. Se han propuesto diversas técnicas de control para mejorar el rendimiento del frenado ABS, entre las que se encuentran controladores óptimos [2], controladores difusos [6], o controladores de modo deslizante [2]. En todos estos trabajos se tiene como preocupación el control y deslizamiento de las ruedas o la reducción de los daños materiales en los componentes de los frenos ABS. Sin embargo, en pocos de estos trabajos se ha considerado un sistema por visión artificial, y se han utilizado sensores electrónicos.

Las técnicas de visión artificial o por computadora, son un conjunto de elementos de hardware y software para la adquisición y procesamiento de imágenes, con el fin de

resolver aplicaciones y realizar pruebas automatizadas y de monitoreo inteligente. Un sistema de visión por computadora, captura la imagen de un objeto particular y realiza un procesamiento digital para extraer las características que pueden proveer información relevante en función de la aplicación para el cual se destina (i.e. color, textuta, forma, etc.) [1]. Entre las aplicaciones de la visión artificial como apoyo vial se encuentran:

- Análisis secuencial de imágenes en intersecciones
- Conflictos vehículo-peatón
- Seguimiento de trayectoria de vehículos y peatones
- Predicción de hipotéticas colisiones
- Señalización
- Semáforos inteligentes
- Asistencia al conductor

El sistema de frenos ABS que se propone puede considerarse como asistencia al conductor, ya que de alguna forma servirá de apoyo para asegurar el control del vehículo en distancias mínimas ante una frenada brusca o de emergencia, derivada de una distracción del conductor, o de las condiciones ambientales.

En la figura 2 se muestra la implementación del sistema en el vehículo, indicando el campo de visión sobre la superficie de rodamiento [3], y sobre ésta la sub región de interés considerada para únicamente aplicar el análisis sobre ella, i.e. la ROI (Region Of Interest [Región de interés]) [4][5].

La implementación que se propone consiste de tres escenarios, acorde a los tipos de terrenos de desplazamiento del vehículo en la ciudad: (i) suelo seco y luz de día, (ii) suelo húmedo y luz de día, y (iii) suelo seco y luz baja.

La metodología del sistema de visión diseñado para mejorar los frenos ABS, consiste en cuatro etapas como se aprecia en la figura 1. En primer lugar, se realiza la

adquisición de video, la cual se logra mediante el uso de una cámara situada del tal forma que el campo visual frontal coincida con la del conductor, para que la información que pueda proporcionar sea complementaria para el mismo. Después, se procede a la extracción de características, en éste caso entropía, velocidad y luminosidad. La información de las variables mencionadas, pasa a la siguiente etapa, el sistema difuso el cuál la evaluará de acuerdo al conjunto de reglas establecidas y calcula el coeficiente de fricción asociado a la superficie de rodamiento [5]. Por último, el resultado del sistema difuso envía la señal de activación al sistema de frenos ABS, es decir, si se activa o no.

En la sección 2 se detalla la propuesta, los módulos que la componen y las condiciones de la implementación. En la sección 3 se proporcionan las especificaciones técnicas necesarias y, los resultados, conclusiones y trabajo futuro se presentan en la sección 4.

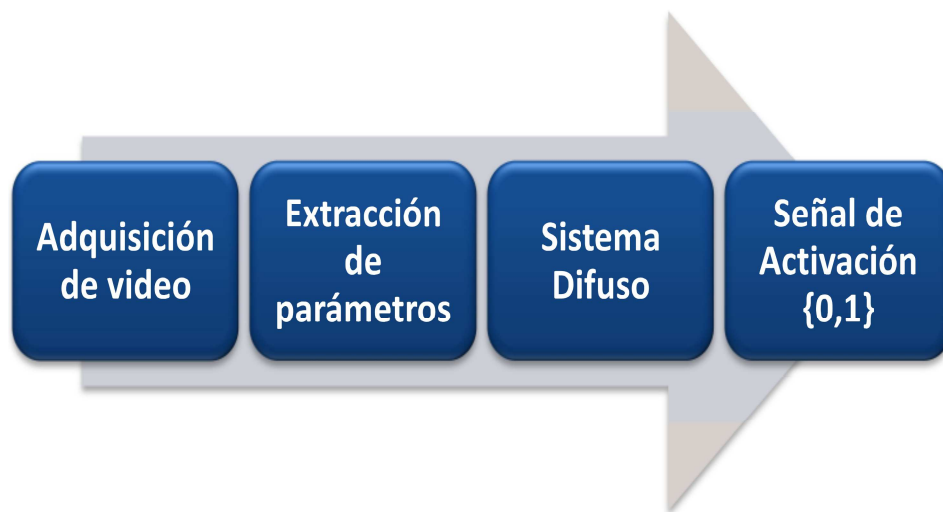


Fig. 1. Metodología propuesta del sistema de visión.

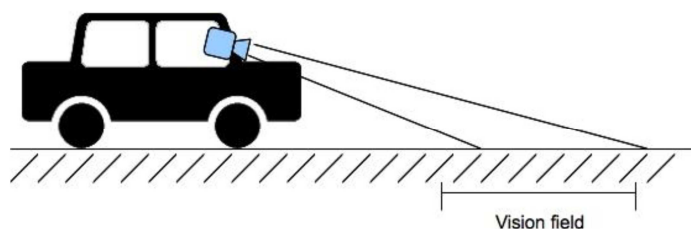


Fig. 2. Area de visión desde la cámara empotrada en el automóvil y la ROI.

2. Desarrollo

En una etapa inicial, la propuesta se desarrolló mediante simulaciones, las variables de entrada seleccionadas fueron la velocidad (v), la entropía (e) y la luminosidad (l); con éstas variables se configuró un sistema MISO (Multiple Input Single Output) en la plataforma de software LabVIEW, estableciendo los conjuntos difusos y reglas de inferencia que permitieran su implementación real. Corroborado el funcionamiento del sistema difuso, se desarrolló un programa para la extracción automática de características-variables consideradas (v, e, l), utilizando NI Vision Assistant, con lo que se extrajeron cuadro por cuadro las imágenes que conforman cada video y fueron procesadas para identificar las tres condiciones de superficie de rodamiento consideradas, (i, ii, iii). Del procesamiento de cada imagen se obtienen los valores de las tres variables (v, e, l), que se ingresan al sistema difuso, y de acuerdo a la serie de

reglas programadas se obtiene una decisión. La figura 3, describe el procedimiento de la implementación final.

La figura 4 muestra el diagrama de bloques implementado para el sistema completo, desde la extracción de características-variables hasta su evaluación por el sistema difuso; dados todos los detalles del gráfico, la figura muestra la generalidad de las interconexiones del sistema, ya que la plataforma de desarrollo no permite la generación de gráficos vectorizados, es que entre las figuras 3 y 4 se puede visualizar el desarrollo del sistema completo, y poder reproducirlo.

El programa fue desarrollado en módulos realizados para cada variable en NI Vision Assistant. Este asistente, permite realizar una enorme gama de procesamiento sobre una imagen y ofrece una visualización previa de cada fase y cada operación que se realiza. Por ello, los módulos de luminosidad y entropía fueron desarrollados primeramente como un script que luego fue exportado a LabVIEW para integrarlo en un sólo VI que incluyera cada parte del procesamiento como la aplicación del sistema difuso.

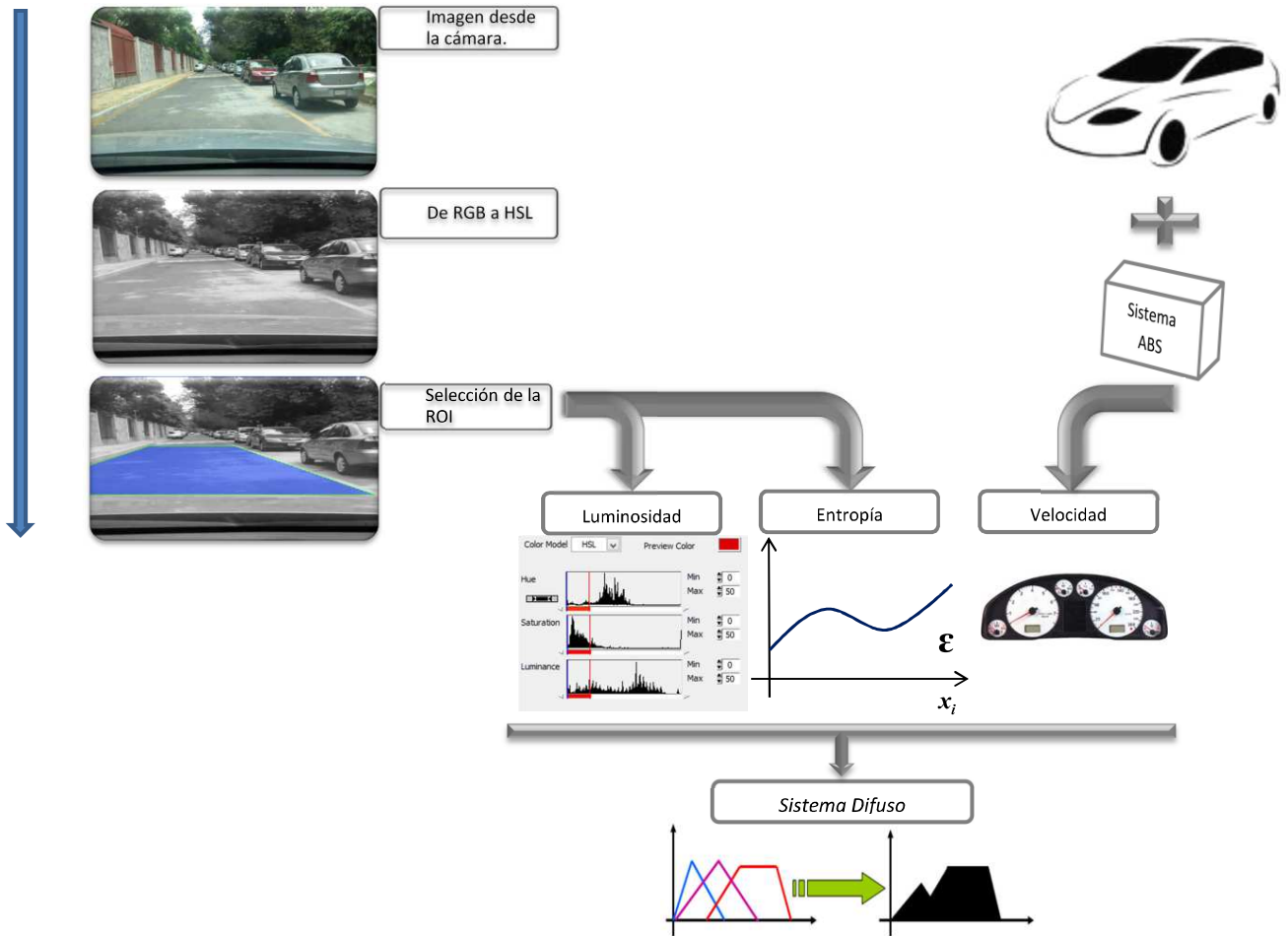


Fig. 3. Sistema desarrollado usando NI Vision Assistant y NI Fuzzy Logic.

2.1. Luminosidad

En la primera etapa, cada imagen obtenida del video tomado por la cámara, se exporta al espacio de color HSL (Hue, Saturation, Lightness); la segunda etapa transforma la capa de luminosidad (lightness) al tamaño kernel [5], con el propósito se resaltar los bordes y detalles de la imagen a fin de mejorar su resolución; la tercera etapa marca la región de interés (ROI), y sobre ella es que se realiza el cálculo de la luminosidad, obteniendo una mejor estimación al realizar el cálculo solo la porción de la imagen de nuestro interés, que es el área donde el vehículo se desplaza.

2.2. Entropía

Este módulo consiste en 4 etapas. Similar al proceso previamente mencionado, en la primera etapa la imagen ahora se exporta al espacio de color HSI (Hue, Saturation, Intensity), en la segunda etapa se aplica una convolución sobre la imagen con un tamaño 3 de kernel para resaltar detalles de la imagen; en la tercera etapa se delimita la ROI, y finalmente en la 4ta etapa calcula el histograma de la imagen. Para calcular la entropía sobre la ROI se uso la ecuación 1 sobre los pixeles de la imagen con valores normalizados [5].

$$E = \sum_{i=0}^n p(x_i) \log p(x_i) \quad (1)$$

2.3. Velocidad

El cálculo de la velocidad fue concebido primero como estimar la razón de cambio de los puntos de interés detectados en la ROI, aplicando un algoritmo de extracción automática de ellos, pero se terminó por tomar la señal de estimación de la velocidad proporcionada por la propia computadora del sistema ABS, que ya hace esta estimación de mejor forma, por medio de una adquisición de una variable analógica, ahorrando ciclos de procesamiento necesarios para optimizar la ejecución final del sistema.

2.4. Sistema difuso

El sistema difuso es el módulo donde se realiza la toma de decisión de todo el proceso, a fin de liberar la señal indicadora de activación correspondiente. Éste recibe las tres señales (v,e,l), con los valores calculados en los módulos anteriormente descritos, cuadro a cuatro de video analizado, y con base a valores de referencia es que se asignaron límites para los valores de la señal: Bajo, Medio y Alto [6][7]. Las gráficas con los rangos tomados para cada variable se muestran en la figura 5. Es así que el sistema difuso toma la decisión de activar o desactivar (on/off, 0 o 1) el freno ABS [8], la

decisión es tomada con base al conjunto de reglas establecidas, las cuales se muestran en la tabla 1.

| IF | THEN |
|---------------------|------|
| V=B AND L=B AND E=M | 1 |
| V=M AND L=B AND E=M | 1 |
| V=M AND L=B AND E=B | 1 |
| V=A AND L=B AND E=M | 1 |
| V=A AND L=B AND E=B | 1 |
| V=A AND L=M AND E=A | 1 |
| V=A AND L=A AND E=A | 1 |

Tabla 1. Reglas del sistema difuso propuesto.

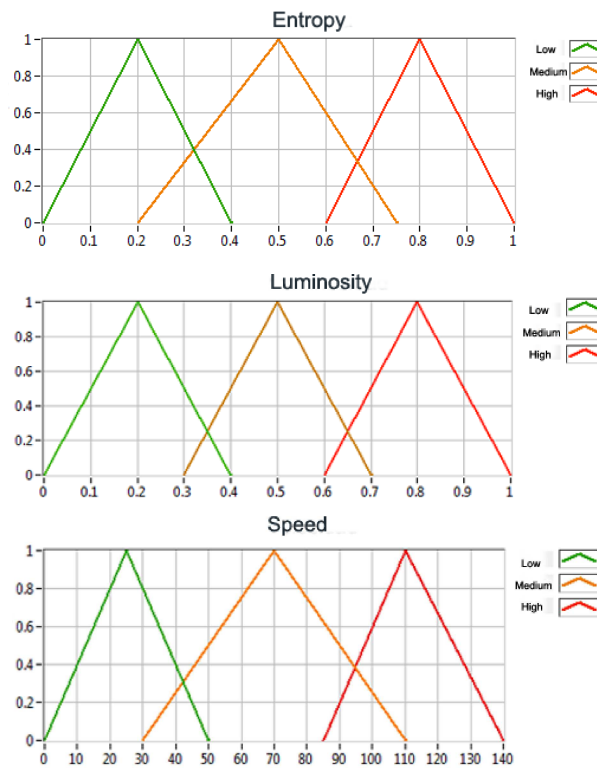


Fig. 5. Valores de la variable entropía, luminosidad y velocidad respectivamente en el sistema difuso.

3. Resultados

El uso del software LabVIEW nos ha proporcionado rapidez para la implementación de nuestra propuesta, y adicionalmente nos brinda eficiencia y funcionalidad para realizar las simulaciones con la integración del hardware.

En el sistema final sólo requiere el video como entrada, luego se calculan las variables y finalmente tras el procesado, se tiene como salida dos posibles valores {0, 1}, como coeficiente de asociación de las condiciones de la superficie de rodamiento. El sistema fue implementado con video digital de 1080 × 720 pixeles de resolución, en una computadora tipo WorkStation con procesador Intel Xeon y sistema operativo MS-Windows ver 7 / 64 bits; logrando con estas condiciones un máximo de procesamiento de 16 cuadros por segundo.

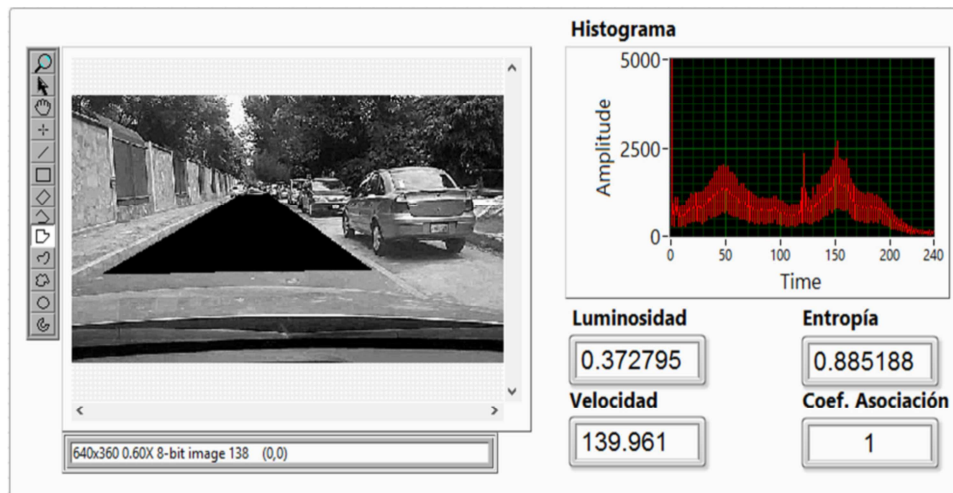


Fig. 6. Simulación de escenario de superficie seca y luz de día.

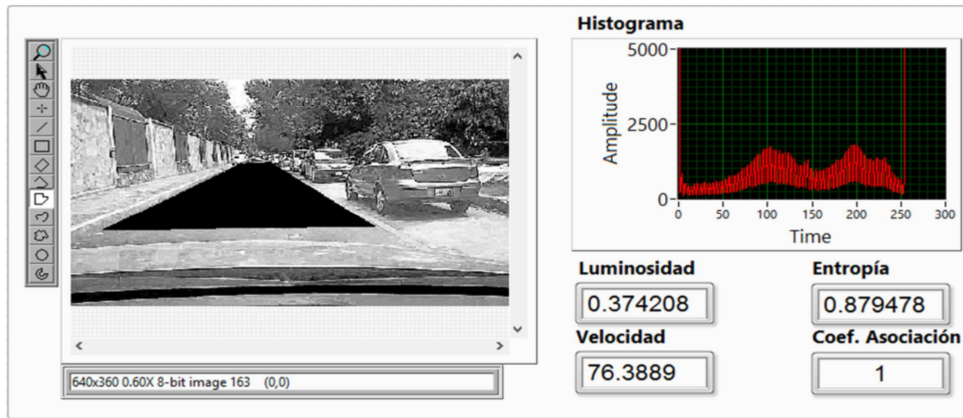


Fig. 7. Simulación de escenario de superficie húmeda y luz de día.

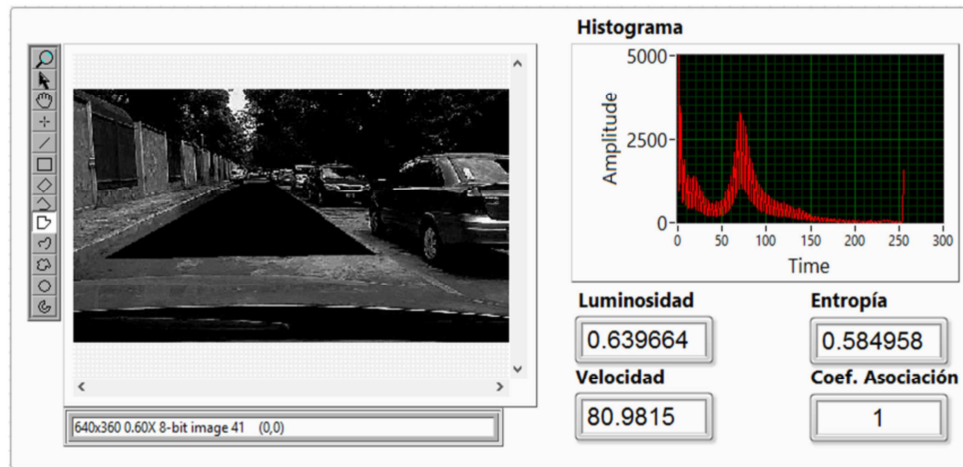


Fig. 8. Simulación de escenario de superficie seca y luz baja.

Las figuras 6, 7 y 8, presentan los resultados de las simulaciones en el sistema, a través de la imagen resultante del procesamiento y un gráfico con el histograma de la misma. Las imágenes usadas corresponden a los tres tipos de escenarios considerados ((i) suelo seco y luz de día, (ii) suelo húmedo y luz de día, y (iii) suelo seco y luz baja).

El control del frenado ABS es un problema de control no lineal debido a la la complicada relación entre sus componentes y parámetros. La investigación llevada a cabo ha cubierto una amplia gama de cuestiones y desafíos. Muchos métodos de control se han desarrollado y la mayoría de estos enfoques requieren aproximaciones al

modelo del sistema, y algunos de ellos no pueden alcanzar satisfactoriamente un buen rendimiento en los cambios de diversas carreteras y condiciones ambientales. Si bien los métodos de control inteligente no tiene la necesidad del modelo matemático, requiere otra fuente de información adicional a los propios de un sistema ABS.

La ventaja convincente de la lógica difusa es la capacidad de modificarse y ajustarse ciertas partes de la superficie de rodamiento. Se ha presentado una propuesta que puede ser un apoyo para la activación de un sistema de frenos ABS usando visión por computadora. La solución fue modelada y realizada sobre el sistema LabVIEW con la perspectiva futura de implementarse en un sistema embebido en tiempo real, y a partir del procesamiento de los videos obtenidos se tuvieron buenos resultados con los diferentes parámetros considerados, siendo así que éstos reflejan el comportamiento concebido con la idea planteada, i.e. el hecho de poder supervisar el campo de visión del conductor, enfocándonos al análisis de la superficie de rodamiento.

4. Conclusiones

Se presentó un algoritmo capaz de proporcionar información adicional que permita la activación de un sistema ABS, antes de que un sistema convencional detecte una situación de riesgo usando visión por computadora, y un sistema difuso. La solución fue modelada y realizada sobre el sistema LabVIEW con la perspectiva futura de implementarse en un sistema embebido en tiempo real, y a partir de las simulaciones con los videos obtenidos se tuvieron buenos resultados con los diferentes parámetros considerados, siendo así que éstos reflejan el comportamiento concebido con la idea planteada, i.e. el hecho de poder supervisar el campo de visión del conductor, enfocándonos al análisis de la superficie de rodamiento.

El trabajo a futuro se considera hacia tres vertientes: (a) la implementación de la solución en un vehículo real, usando una CompactRIO y una cámara de propósito específico del tipo NI 1772C; (b) realizar pruebas de la posible implementación de este sistema en un dispositivo móvil de fácil acceso, e.g. un teléfono móvil o tableta, que sea

de bajo costo y portable para el conductor; y (c) agregar un módulo de seguimiento de conducta de manejo, o de seguimiento de movimientos de manejo de un conductor, a fin de realizar un seguimiento estadístico del conductor ante diversas situaciones con un número de horas acumuladas de manejo .

5. Referencias

- [1] BOSCH: Automotive Handbook. Bentley Publishers. 2007.
- [2] K. R. Castleman, Frenos ABS. 1996. Prentice Hall.
- [3] Y. Ishino, H. Saji, Extraction of road markings from aerial images. In: SICE Annual Conference. 2008. 2180–2183 pp.
- [4] G. Chen, Texture based road surface detection. Master's thesis, Case Western Reserve University. 2008.
- [5] R. C. Gonzalez, R. E. Woods, Digital Image Processing. 1992. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.
- [6] K. Tanaka, An Introduction to Fuzzy Logic for Practical Applications. 1996. Springer-Verlag New York.
- [7] E. Aguirre, A. González, “A fuzzy perceptual model for ultrasound sensors applied to intelligent navigation of mobile robots”. Vol. 19. 2003. 171–187 pp.
- [8] A. S. M. B. Martín del Brío, Redes Neuronales y Sistemas Borrosos. 2007. Edit. Alfa-Omega.

6. Autores

C. Gabriel García-Ponds es estudiante de Ingeniería en Computación en la Universidad Autónoma Metropolitana Unidad Azcapotzalco. Actualmente trabaja en su Proyecto Terminal orientado al diseño e implementación de algoritmos de visión por computadora para dispositivos móviles.

Dr. Juan Villegas Cortezes Profesor–Investigador del Departamento de Electrónica, en el área de Instrumentación, de la UAM Azcapotzalco. Actualmente realiza investigación en el área de Memorias Asociativas, Visión por computadora y Procesamiento digital de imágenes.

Dr. Carlos Avilés Cruz es Profesor–Investigador del Departamento de Electrónica, en el área de Instrumentación de la UAM Azcapotzalco. Actualmente trabaja en las áreas de investigación del Reconocimiento de Patrones, Visión por computadora y análisis de señales.

Dr. Iván Vázquez Álvarez es Profesor–Investigador del Departamento de Electrónica, en el área de Instrumentación de la UAM Azcapotzalco. Actualmente trabaja en las áreas de investigación de Control y Análisis de señales.

Dr. Ismael Osuna Galán es Profesor–Investigador de la Universidad Politécnica de Chiapas, unidad Tuxtla. Actualmente trabaja en las áreas de investigación del Reconocimiento de Patrones, Visión por computadora y Aplicaciones mecatrónicas.

Dra. Yolanda Pérez Pimentel es Profesora–Investigadora de la Universidad Politécnica de Chiapas, unidad Tuxtla. Actualmente trabaja en las áreas de investigación de Redes neuronales artificiales, Visión por computadora y Aplicaciones mecatrónicas.

Implementación de una estructura neuronal celular en hardware reconfigurable

Luis F. Muñoz M.

Universidad de Guadalajara, Centro Universitario de Ciencias Exactas e Ingenierías,
Blvd. Marcelino García Barragán #1421, C.P. 44430, Guadalajara, Jalisco, México

Teléfono: +52 (33) 1378 5900

luisfmm@yahoo.com.mx

Juan José Raygoza P.

Universidad de Guadalajara, Centro Universitario de Ciencias Exactas e Ingenierías,
Blvd. Marcelino García Barragán #1421, C.P. 44430, Guadalajara, Jalisco, México

Teléfono: +52 (33) 1378 5900

juan.raygoza@cucei.udg.mx

J. Roberto R. Barón

Universidad de Guadalajara, Centro Universitario de Ciencias Exactas e Ingenierías,
Blvd. Marcelino García Barragán #1421, C.P. 44430, Guadalajara, Jalisco, México

Teléfono: +52 (33) 1378 5900

jose.reyes@red.cucei.udg.mx

Susana Ortega Cisneros

Centro de Investigación y de Estudios Avanzados del I.P.N, CINVESTAV, Unidad Guadalajara
Av. del Bosque 1145, colonia el Bajío, C.P. 45019, Zapopan, Jalisco, México

susana.ortega@gdl.cinvestav.mx

Resumen

El siguiente artículo presenta el diseño e implementación de una red neuronal celular (CNN) desarrollado en dispositivos reconfigurables FPGA con aplicaciones para procesamiento digital de imágenes en escala de grises a ocho bits y dimensión M x N.

La CNN es capaz de hacer el procesamiento mediante una exploración sobre todo el patrón de entrada, desplazando por bloques la matriz para procesar la imagen. La red propuesta consiste en una matriz de 24 elementos de procesamiento, constituidos por unidades aritméticas independientes que determinan su valor, colocados en una malla rectangular de cuatro filas por seis columnas. Se describe un conjunto de enlaces programables que permiten modificar la salida mediante el tipo de conectividad definido. El diseño fue desarrollado utilizando lenguaje de descripción de hardware VHDL. La implementación se realizó en un dispositivo FPGA Xilinx® de la familia Virtex-6. Se muestra la simulación, así como los resultados en área de ocupación y latencia.

Palabra(s) Clave(s): CNN, FPGA, procesamiento digital de imágenes.

1. Introducción

Las redes neuronales celulares (CNN del inglés *Cellular Neural Network*) son un paradigma de procesamiento en paralelo que permite la comunicación entre unidades vecinas, fueron desarrolladas en la universidad de Berkeley por Leon O. Chua y Yang en 1988 [1]. Este concepto involucra características de autómatas celulares y redes neuronales artificiales, de los primeros toma su organización y estructura a manera de rejillas regulares con interacción entre elementos locales, mientras que de las segundas su capacidad de procesamiento asíncrono paralelo e interacción global de sus elementos en la red [2].

Su importancia radica en el desempeño computacional, principalmente su estructura de cálculo concurrente, la cual presenta una mejoría para la solución a problemas de velocidad encontrados en métodos de computo convencional por software [3], además, un campo de investigación significativo es el procesamiento digital de imágenes, fundamental en la inteligencia artificial, robótica, visión artificial y control, áreas donde se requiere procesar una gran cantidad de datos en alta velocidad [4].

Este trabajo plantea el diseño, desarrollo e implementación sobre hardware reconfigurable de una estructura basada en CNN, con la capacidad de procesar imágenes mediante diferentes tipos de conectividad.

1.1. Arquitectura estándar

La arquitectura de la red neuronal celular permite únicamente la interconexión entre células “vecinas”, generando un arreglo el cual puede tomar diferentes formas geométricas, siempre y cuando se sigan respetando las reglas de interconexión de sus elementos, lo que cual conlleva a presentar un gran número de arquitecturas y conectividades diferentes [5,6]. En el espacio bidimensional, las células de la red pueden ser mapeadas sobre diferentes esquemas físicos, como por ejemplo, los mostrados en la Fig. 1. Con independencia de la estructura empleada, la CNN puede estar formada por células idénticas entre sí o por un conjunto reducido de células diferentes, como ocurre en el caso de la redes NUP-CNN [7]. (*Non Uniform Processor CNN*), Fig. 1d y 1f.

La vecindad de una célula es definida por el conjunto de células que se sitúan alrededor de ésta, sobre las cuales se ejerce una influencia de forma directa. A partir de lo anterior, es posible encontrar una función que determine exactamente la región de vecindad de cada célula. En general, esta región suele venir representada mediante una esfera, centrada en la propia célula, y cuyo radio r define el tamaño de la vecindad. En la práctica, sin embargo, lo habitual es que las redes formen estructuras bidimensionales de una sola capa y radios de vecindad reducidos. Un ejemplo de este tipo de redes, constituida por células con vecindad de radio $r = 1$ y con vecindad de radio $r = 2$, se muestra en la Fig. 1e.

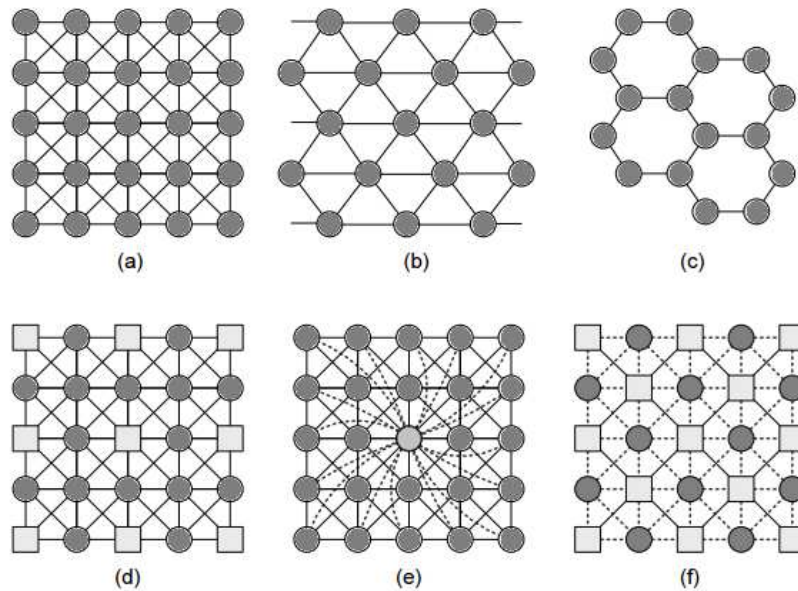


Fig. 1. Ejemplos de CNN con diferentes estructuras (a) rectangular, (b) triangular, (c) hexagonal (d) NUP-CNN, (e) diferente tamaño de vecindad y (f) conexiones distintas [7].

1.2. CNN en tiempo discreto

Las CNN de tiempo discreto (TDCNN) fueron propuestas por Hubert Harren, se definen como una clase especial de CNNs. Dichas redes, pueden ser descritas por un algoritmo recursivo cuya dinámica permite actuar en periodos definidos de tiempo [8]. La DTCNN se establece mediante las ecuaciones 1 y 2.

$$x_{ij}(k) = -x_{ij}(k) + A * y_{ij}(k) + B * u_{ij} + I \quad (1)$$

$$y_{ij}(k + 1) = f(x_{ij}(k)) = \begin{cases} 1 & \text{si } x_{ij}(k) > 0 \\ 0 & \text{si } x_{ij}(k) < 0 \end{cases} \quad (2)$$

Dónde: u_{ij}, x_{ij}, y_{ij} son las variables de entrada, estado y salida respectivamente, A (matriz de retroalimentación) y B (matriz de control) son matrices del tamaño de vecindad definida, e I es un valor de *offset*. Este conjunto de parámetros es llamado

cloning template. La ventaja primordial de las DTCNN se debe a que la descripción de la salida para el siguiente estado puede ser descrito por un conjunto de desigualdades lineales [8]. Además, la simulación en una computadora digital es más sencilla, debido a que no requiere ningún algoritmo de integración, como en el modelo original el cual está basado en ecuaciones diferenciales en tiempo continuo.

De manera general el funcionamiento de una red celular neuronal de cualquier tipo se representa mediante la Fig. 2 [9].

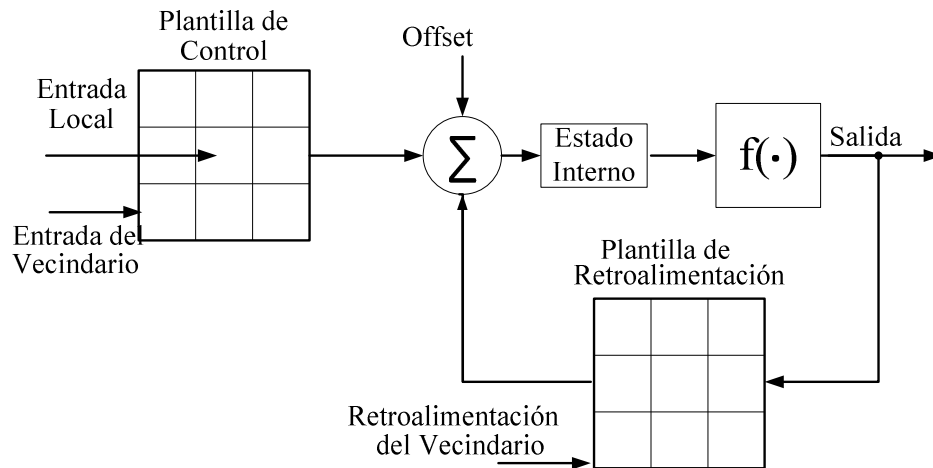


Fig. 2. Modelo funcional de una CNN.

1.3. Procesamiento digital de Imágenes y la CNN

Una imagen digital es la representación bidimensional de una imagen real, expresado como una función de la intensidad de luz $f(x,y)$, donde el valor de la amplitud f en el punto definido por las coordenadas espaciales (x,y) , está determinado por el brillo de la imagen original en ese punto, denominado píxel. Un píxel p con coordenadas (x,y) tiene dos vecinos horizontales, dos verticales y cuatro diagonales, cada uno de estos píxeles con una distancia 1 de p y en los bordes algunos de estos elementos quedan fuera de la imagen [10], lo anterior es mostrado en la Fig. 3.

Entre las primeras aportaciones de las CNNs al procesamiento digital de imágenes cabe destacar los diseños de plantillas (*templates*) que han marcado una tendencia sobre el desarrollo de algoritmos utilizados en distintos campos de la inteligencia artificial, entre otros. Estas redes definen procesos que explotan la conectividad en retroalimentación del modelo original de la CNN [1], caracterizando la red como un procesador de imágenes binarias, capaz de realizar procesamientos lineales y no lineales, operaciones morfológicas como el rellenado automático de huecos, la codificación de caracteres y la detección de elementos simples como líneas y bordes, entre otros [1,2 y 4]. Las transformaciones morfológicas, que parten de la teoría de conjuntos, son herramientas matemáticas que se encargan de modificar formas y estructuras de objetos, además de obtener características especiales de estos. Dichas transformaciones se realizan en imágenes binarias, aunque también existen casos para su aplicación en imágenes con escalas de grises.

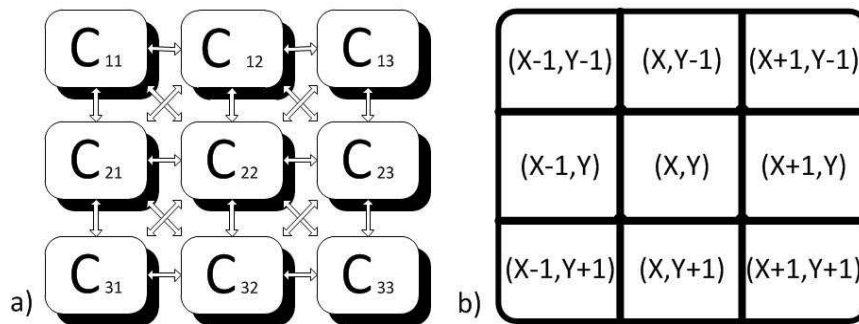


Fig. 3. Ejemplo de arreglos bidimensionales, (a) CNN 3x3 (b) imagen 3x3.

2. Desarrollo

En este trabajo se propone una estructura de una red neuronal celular para procesamiento de imágenes a escala de grises por medio de la utilización de Hardware Reconfigurable FPGA. La red consta de un conjunto de veinticuatro elementos, conectados totalmente por vecindarios de radio $r = 1$ en arreglos rectangulares, es decir, cada neurona se puede comunicar hasta con ocho vecinos, como se ilustra en la Fig. 4.

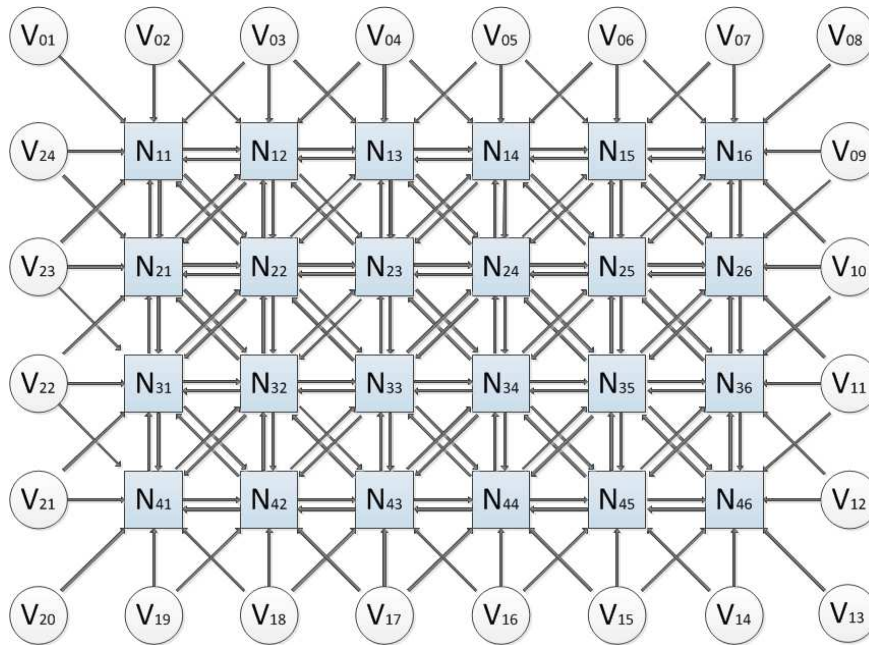


Fig. 4. Diagrama general de la red.

2.1. Registros de configuración

La comunicación entre los elementos de la red se lleva a cabo a nivel local, con la posibilidad de alterar su salida mediante la variación de los siguientes parámetros, buscando obtener operaciones que puedan ser utilizadas para el procesamiento de imágenes en escala de grises:

- Peso o valor a incrementar en la entrada original.
- Tipo de conectividad.
- Retroalimentación consigo misma y/o un valor adicional (offset).

Para ello se debe tener conocimiento por medio de registros específicos, los cuales son enviados desde una unidad de control global a toda la red (decodificador) y son recibidos en cada célula que los reconoce y actúa de la manera en que se le indique. Los cuatro tipos de registros de configuración generados se muestran en la Tabla 1.

| Nombre del registro | Función del registro | Tipo | Número de bits |
|------------------------|--|-------|----------------|
| Registro_control_peso | Modifica directamente la entrada (peso), ya que es multiplicado por valor original. | Valor | 4 |
| Registro_control_in | Asigna los enlaces entre elementos, habilita de acuerdo a un orden en sentido de las manecillas del reloj. | Bus | 9 |
| Registro_control_SR | Define el tipo de operación entre los enlaces, suma o resta los valores vecinos, lleva el mismo orden que <i>Registro_control_in</i> | Bus | 9 |
| Registro_control_retro | Modifica directamente la salida sumando un <i>offset</i> , al resultado. | Valor | 8 |

Tabla 1. Tipos de registros de configuración.

La Fig. 5 muestra los registros de configuración y el tipo de conectividad que genera en caso de requerir influencia con los vecinos laterales o los diagonales.

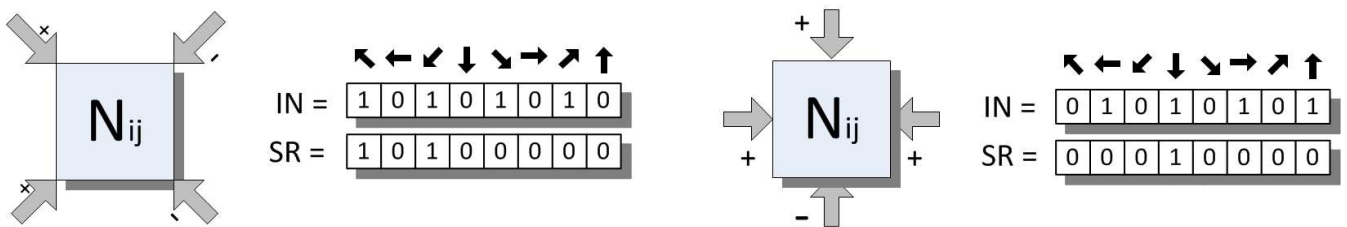


Fig. 5. Ejemplo de configuración para influencia lateral.

Cualquier alteración en algún bit de los registro representa un cambio considerable en el resultado a la salida de la red, por lo que se deben identificar las estructuras de conexión necesarias para realizar procesamientos definidos.

Con esta metodología es posible asignar conexiones tanto de manera global como local, configurar las neuronas individualmente buscando operaciones específicas a nivel pixel, o realizar conexiones homogéneas en toda la red para obtener filtrados espaciales genéricos, de igual manera realizar configuraciones híbridas, conectando una parte de la red de manera similar y el resto de manera específica.

2.2. Elementos de procesamiento

La Fig. 6 muestra el diagrama a bloques de la estructura interna de un elemento de procesamiento, donde destaca un módulo que calcula el peso asignado a la entrada original más el valor del vecindario, de acuerdo al tipo de conectividad que se desee, y otro módulo para añadir una posible retroalimentación más un valor determinado (*offset*), además de un registro de salida que almacena el dato después de cierta cantidad de flancos.

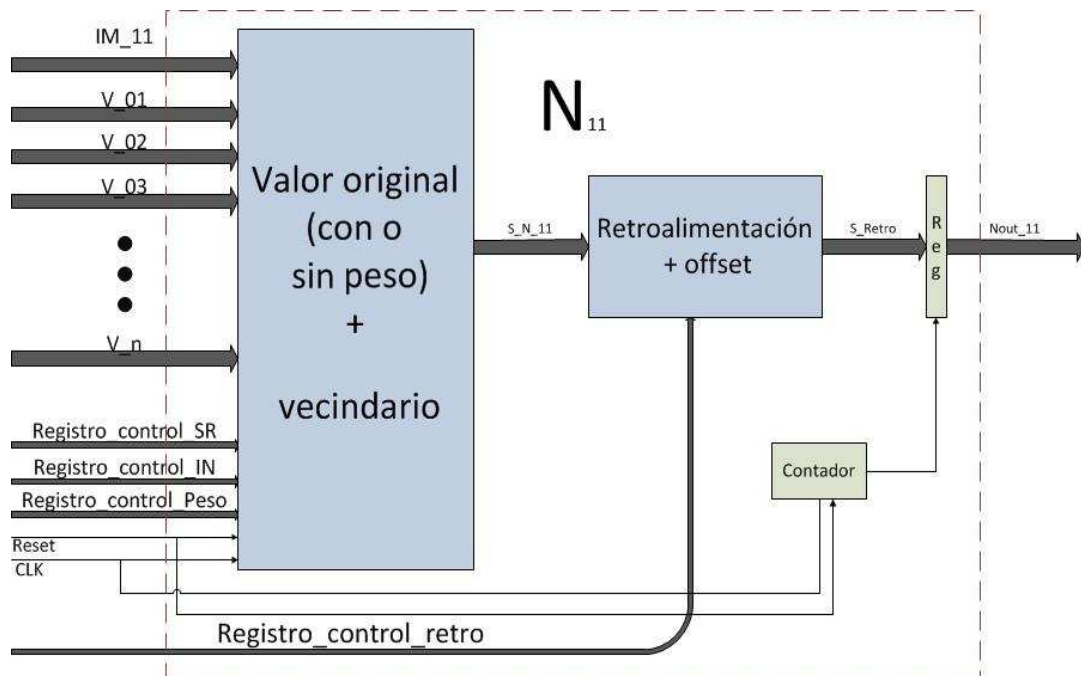


Fig. 6. Diagrama a bloques de un elemento de la red.

2.2.1. Vecindario

Este bloque recibe datos de la entrada original, de las neuronas vecinas y señales de control. El dato de entrada es almacenado en un primer registro síncrono que trabaja en flancos de subida, posteriormente es modificado por un peso predefinido, seguido de esto, se añade el valor del vecindario, el cual se obtiene mediante un arreglo de entidades aritméticas conectadas entre sí.

Los datos de los elementos vecinos se adquieren por medio de una unidad de control interna, que valida su habilitación de acuerdo a los registros de configuración recibidos a la entrada, siendo un bit del tipo *registro_control_in* para cada vecino posible, y de igual manera respetando el mismo orden se indica el tipo de operación local con *registro_control_sr*. Finalmente cuenta con otro registro síncrono de almacenamiento que trabaja en flancos de bajada, útil para evitar problemas de inestabilidad al presentarse cambios en las entradas produciendo variaciones no deseadas a la salida. El diagrama de este bloque se observa en la Fig. 7.

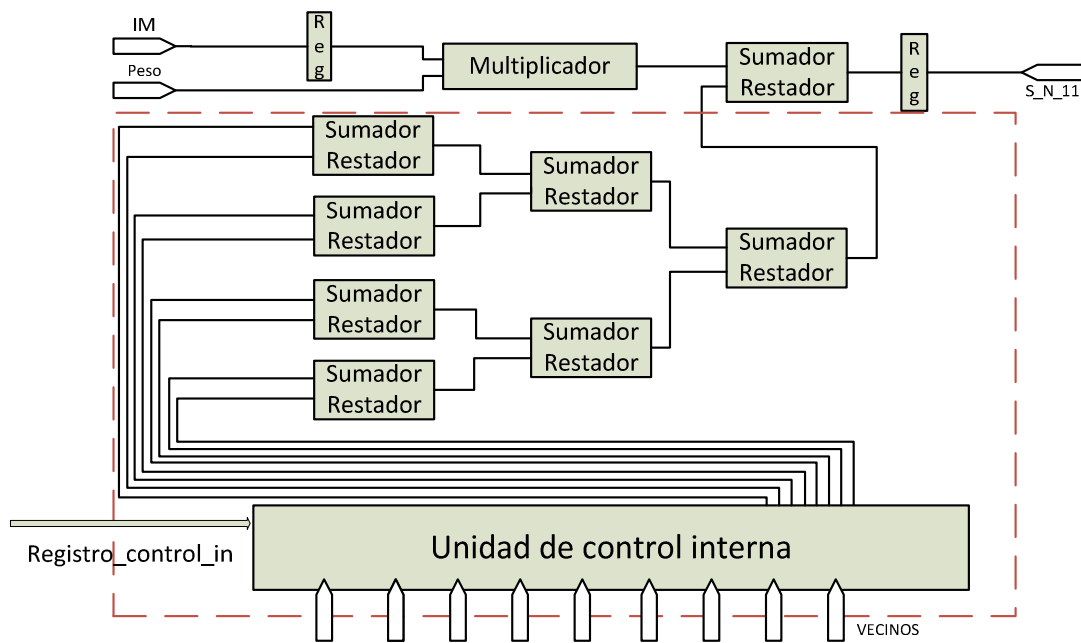


Fig. 7. Estructura de cálculo de vecindario neurona.

2.2.2. Módulo de retroalimentación y offset

En este módulo, ilustrado en la Fig. 8, se le agrega un nuevo término a la salida del cálculo de vecindario y el valor original modificado, dicho término corresponde al valor de retroalimentación, más un offset, el cual representa una cantidad predefinida que permite una mayor gama de posibilidades a la salida final de la red. Este efecto es habilitado mediante el último bit del *registro_control_in*, que indica al multiplexor si es necesario aplicarlo.

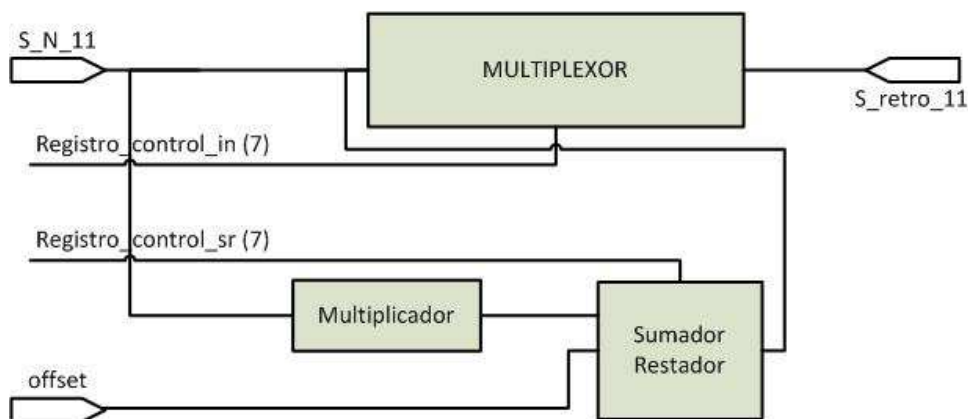


Fig. 8. Módulo de retroalimentación y offset.

Con esta estructura es posible generalizar el comportamiento de cada elemento de la red en base a los siguientes puntos:

- Modificar el valor de entrada en relación a un peso. (*registro_control_peso* > 1) o utilizar el valor original (*registro_control_peso* = 1).
- Hacer un procesamiento en base a la conexión de cada elemento con su vecindario, ya sea que todos los elementos se comporten de manera semejante (*registro_control_in*, *registro_control_sr* iguales en todos), o efectuar el procesamiento con conexiones particulares en deferentes elementos (*registro_control_in*, *registro_control_sr* independientes para cada elemento).

2.3. Funcionamiento

Para fines prácticos y poder corroborar el funcionamiento de esta metodología se acota las salidas de la red a 16 casos específicos, para lo cual se desarrolla un decodificador.

En la Tabla 2 se muestran como ejemplo los casos: 0000, 1000, 1111, donde existe una modificación de los registros de configuración para la red.

| No. de caso | Código | Operaciones | Conexiones necesarias | Parámetros |
|-------------|--------|-------------------------------------|---|--|
| 1 | 0000 | Dilatación bilateral (Homogénea) | Habilitación de vecinos izquierdo y derecho | Peso = 1 Retro = 0 offset = 0 |
| 8 | 0111 | Detección de Contornos | Habilitación de todos los vecinos | Peso = 8 Retro = 1 offset = 0 |
| 16 | 1111 | Conectividad específica por neurona | Cada neurona con conexión independiente | Peso = 11 Retro = 1 offset = - 2 |

Tabla 2. Casos específicos implementados en este trabajo.

Para realizar la demostración del funcionamiento del módulo de la red, se toma como valores de entrada los mostrados en la matriz A, y los resultados esperados para cada caso se muestran de manera independiente en relación a los pesos y las conexiones asignados por los registros de configuración.

ENTRADA

A =

| | | | | | |
|---|---|---|---|---|---|
| 1 | 3 | 1 | 5 | 1 | 1 |
| 3 | 1 | 5 | 1 | 3 | 5 |
| 1 | 3 | 1 | 5 | 1 | 1 |
| 3 | 1 | 5 | 1 | 3 | 3 |

Matriz de pesos

Registros de configuración

A₀₀₀₀

| | | | | | |
|---|---|---|---|---|---|
| 4 | 5 | 9 | 7 | 7 | 2 |
| 4 | 9 | 7 | 9 | 9 | 8 |
| 4 | 5 | 9 | 7 | 7 | 2 |
| 4 | 9 | 7 | 9 | 7 | 6 |

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 1 | 1 |
| 0 | 0 | 0 |

SR
IN
peso

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | | |
| 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | | |
| | | | | | | | 0 | 0 | 0 | 1 |

A₁₀₀₀

| | | | | | |
|---|----|----|----|----|---|
| 0 | 0 | 7 | 0 | 7 | 1 |
| 0 | 10 | 0 | 14 | 0 | 0 |
| 3 | 0 | 14 | 0 | 14 | 7 |
| 0 | 5 | 0 | 7 | 0 | 0 |

| | | |
|---|----|---|
| 1 | 1 | 1 |
| 1 | -8 | 1 |
| 1 | 1 | 1 |

SR
IN
peso

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| | | | | | | | 1 | 0 | 0 | 0 |

A₁₁₁₁

| | | | | | |
|----|----|----|----|----|----|
| 3 | 28 | 0 | 0 | 6 | 0 |
| 28 | 0 | 51 | 2 | 10 | 0 |
| 0 | 27 | 0 | 49 | 0 | 0 |
| 29 | 0 | 52 | 0 | 28 | 29 |

| | | |
|----|----|----|
| 0 | -1 | 0 |
| -1 | 11 | -1 |
| 0 | -1 | 0 |

SR
IN
peso

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | | |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | | |
| | | | | | | | 1 | 0 | 1 | 1 |

offset = -2

| | | |
|---|----|---|
| 1 | 0 | 1 |
| 0 | -4 | 0 |
| 1 | 0 | 1 |

SR
IN
peso

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | | |
| | | | | | | | 0 | 1 | 0 | 0 |

2.4. Procesamiento de una imagen

Para desarrollar el procesamiento de una imagen de tamaño M x N en escala de grises, se realiza el recorrido del módulo CNN por todo el patrón de entrada, debido a que el proceso se realiza solo con 24 pixeles de la imagen (tamaño de la red) a la vez, este trabajo propone un sistema que efectúe dicho recorrido.

La imagen es una matriz y cada pixel se almacena en una localidad de memoria, la cual envía datos hacia el módulo de la red mediante un bus de direccionamiento, manejado por una entidad de control encargada de la correcta distribución de los datos, entre una memoria de entrada y otra de salida, donde se guardan los datos procesados, a razón de ser requeridos para su posible visualización. El diagrama a bloques del sistema generado por esta metodología se muestra en la Fig. 9(a), y la Fig. 9(b) es un ejemplo del desplazamiento a través de la matriz de entrada, haciendo un barrido horizontal, es decir de izquierda a derecha y de arriba hacia abajo.

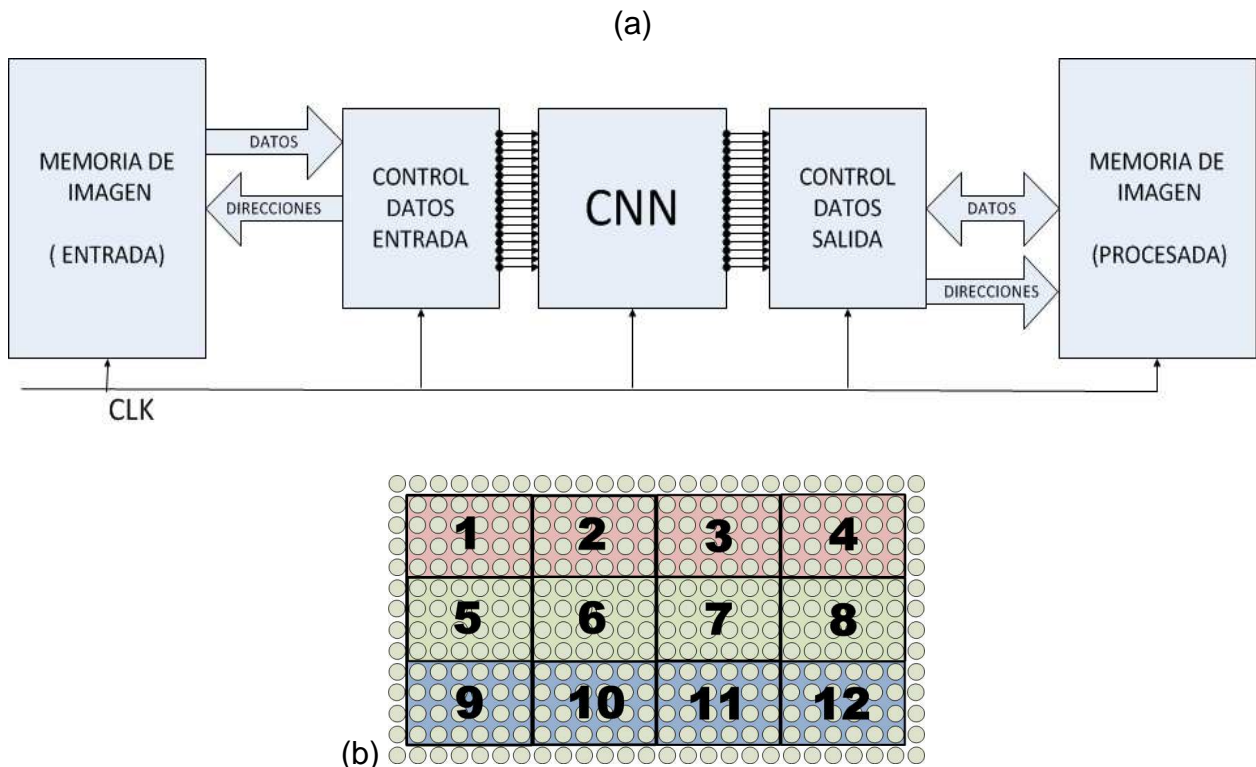


Fig. 9. (a) Diagrama a bloques del sistema completo, (b) ejemplo del barrido de la imagen.

En esta etapa del diseño, se utilizó la herramienta de software MATLAB®, para simulación y validación, como medio de lectura de imagen y generador datos para almacenar en memorias, empleando un algoritmo iterativo para colocar pixeles enteros en la memoria del FPGA, y esta pueda proporcionar los datos de manera más eficiente,

colocando anchos de palabra múltiplos de 8 bits, y localidades de memoria acorde al tamaño de la imagen a procesar.

3. Resultados

La implementación de este trabajo se realiza en una plataforma reconfigurable de la familia Virtex-6 de Xilinx® [11-13], haciendo uso de las herramientas de diseño ISE 14.6. Los resultados de la simulación para el módulo de la red se muestran en la Fig. 10. El primer caso, cuando el bus control = “0000” se realiza una dilatación lateral, posteriormente se cambia el bus control = “1000” y por último se muestra la salida cuando el bus control = “1111”.

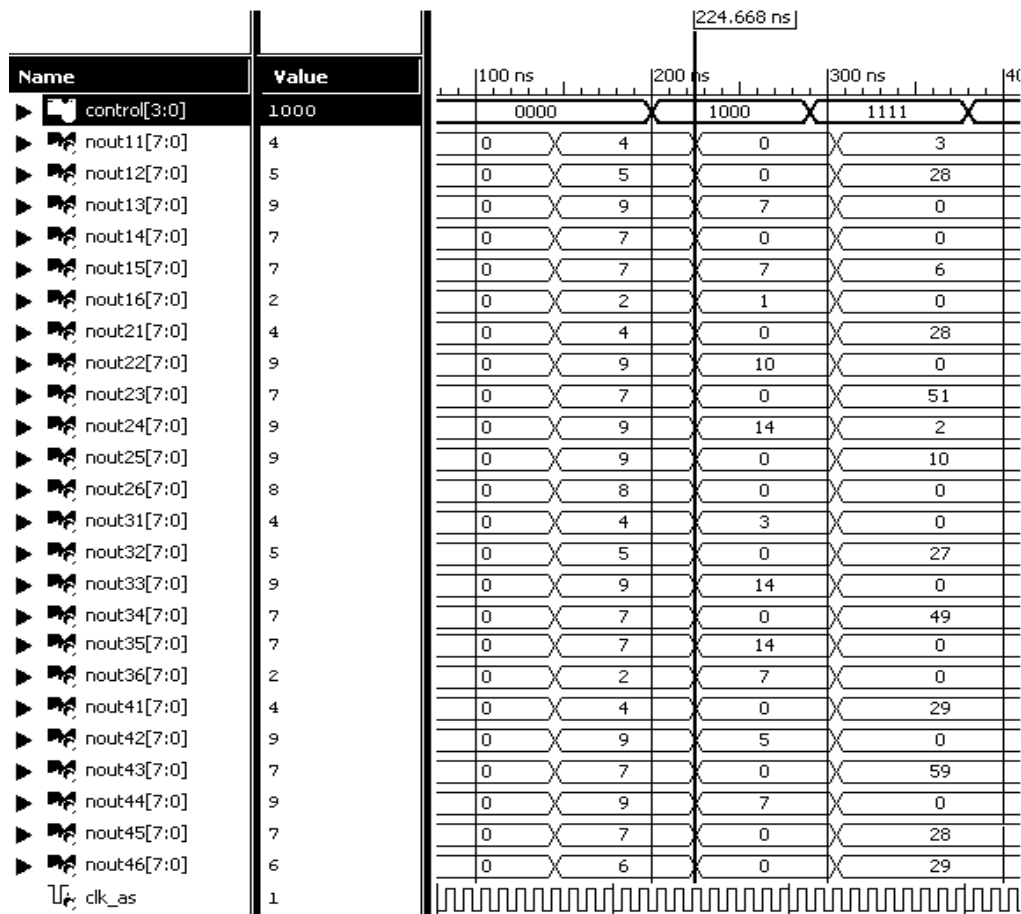


Fig. 10. Simulación del módulo CNN.

Cada tres pulsos de la señal de sincronización se produce el resultado, esto es debido a que el flujo de datos se presenta de la siguiente manera:

- En el primer flanco de subida se carga el dato a la entrada.
- En el primer flanco de bajada aún no hay actividad de los vecinos ya que se inicializan en cero por lo que solo se transfiere el valor original afectado por el peso.
- En el segundo flanco de subida ya se presentan datos del vecindario y se agregan al valor actual, es decir se realiza el procesamiento de las conexiones definidas.
- En el segundo flanco de bajada ya se cuenta con el dato más el vecindario y se envía a la retroalimentación si es que se este se habilita, de lo contrario se queda el valor actual.
- En el tercer flanco de subida se activa el último registro, perteneciente a la salida de la entidad y el valor final es mostrado.

Teniendo en cuenta lo anterior, se pueden considerar que la red tiene una latencia de alrededor de 100 ns, si se consideran pulsos de una duración no superior a los 20ns, aun cuando los cambios sean muy drásticos a la entrada, la estructura de segmentación de la red deberá poder mantener resultados estables en este lapso. Una vez obtenido el dato a la salida es necesario realizar un *reset* para poder cargar un dato nuevo y realizar otra operación, esto para evitar resultados erróneos causados por estados anteriores en los elementos de la red.

En la Tabla 3 se puede apreciar un breve resumen de los niveles de ocupación, tras la implementación del módulo CNN en diferentes FPGAs, siendo la Virtex-6 aquella con menor porcentaje de recursos utilizados.

| | Usados | Totales | % | Usados | Totales | % |
|-----------|--------|----------|----------|--------|-----------|----------|
| | | Virtex-6 | | | Spartan-6 | |
| Registros | 403 | 301,440 | 0.13369 | 403 | 54,576 | 0.73842 |
| LUTs | 3,525 | 150,720 | 2.33877 | 2536 | 27,288 | 4.67 |
| Slices | 1,419 | 37,680 | 3.76592 | 445 | 2,278 | 19.5347 |
| | | Artix-7 | | | Zynq | |
| Registros | 403 | 82,000 | 0.49146 | 403 | 35,200 | 1.14489 |
| LUTs | 3,472 | 63,400 | 5.47634 | 4,107 | 17,800 | 23.07307 |
| Slices | 1184 | 10250 | 11.55122 | 532 | 17,600 | 3.02273 |

Tabla 3. Niveles de ocupación del módulo CNN en diferentes FPGAs.

Por otro lado, al utilizar los módulos externos a la CNN, se pueden hacer pruebas para casos de imágenes completas, realizando los efectos de acuerdo a la configuración que se le asigne. En la Tabla 4 se muestran algunos ejemplos del procesamiento obtenido de acuerdo al tipo de conectividad entre elementos de la red y su respectiva matriz de pesos, todos los resultados teniendo en consideración la imagen de entrada ilustrada en Fig. 11, con una resolución 480x640. El tiempo total que el sistema obtuvo, es alrededor de 1.3 ms.



Fig. 11. Imagen de prueba.









| Matriz de pesos | Registros de conexión | Resultado | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|-----------------------|-----------|----|----|----|----|----|----|----|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <table border="1" style="width: 100%; text-align: center;"> <tr><td>-1</td><td>0</td><td>-1</td></tr> <tr><td>0</td><td>5</td><td>0</td></tr> <tr><td>-1</td><td>0</td><td>-1</td></tr> </table> | -1 | 0 | -1 | 0 | 5 | 0 | -1 | 0 | -1 | <p style="text-align: center;">Realce</p> <p style="text-align: center;">  </p> <table border="1" style="width: 100%; text-align: center;"> <tr><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr> </table> | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |  |
| -1 | 0 | -1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 5 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -1 | 0 | -1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | | | | | | | | | | | | | | | | | | | | | |
| <table border="1" style="width: 100%; text-align: center;"> <tr><td>0</td><td>-1</td><td>0</td></tr> <tr><td>-1</td><td>11</td><td>-1</td></tr> <tr><td>0</td><td>-1</td><td>0</td></tr> </table> | 0 | -1 | 0 | -1 | 11 | -1 | 0 | -1 | 0 | <p style="text-align: center;">Aumento de brillo</p> <p style="text-align: center;">  </p> <table border="1" style="width: 100%; text-align: center;"> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> </table> | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |  |
| 0 | -1 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -1 | 11 | -1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | -1 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | |
| <table border="1" style="width: 100%; text-align: center;"> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>-8</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table> | 1 | 1 | 1 | 1 | -8 | 1 | 1 | 1 | 1 | <p style="text-align: center;">Detección de bordes</p> <p style="text-align: center;">  </p> <table border="1" style="width: 100%; text-align: center;"> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  |
| 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | -8 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | | | | | | | | | | | | | | |
| <table border="1" style="width: 100%; text-align: center;"> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>-1</td><td>11</td><td>-1</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> </table> | -1 | -1 | -1 | -1 | 11 | -1 | -1 | -1 | -1 | <p style="text-align: center;">Aumento de contraste</p> <p style="text-align: center;">  </p> <table border="1" style="width: 100%; text-align: center;"> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> </table> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |
| -1 | -1 | -1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -1 | 11 | -1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -1 | -1 | -1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | |

Tabla 4. Resultados del procesamiento utilizando diferentes configuraciones.

4. Conclusiones

En este trabajo se presentó la descripción y los resultados de la implementación de una red neuronal celular totalmente programable en dispositivos reconfigurables FPGAs. En base a esta propuesta, es posible realizar cualquier configuración de conectividad en la CNN mediante cambios en registros específicos, diseñados para otorgar una gran flexibilidad y realizar diferentes operaciones, generando una amplia gama de posibilidades a su salida. Los resultados expuestos en la Tabla 4, así como la latencia del módulo mostrada en la Fig. 10, demuestra que el sistema de procesamiento de imágenes por medio de una red neuronal celular programable, es una alternativa viable para realizar operaciones morfológicas, y que la opción de poder configurar su conectividad, tanto en enlaces como en pesos, resulta en una estructura de procesamiento en Hardware capaz de adaptarse para ser usada en conjunto con cualquier sistema o microprocesador.

5. Referencias

- [1] L. O. Chua, T. Roska, T. Kozek, A. Sarandi, "The CNN paradigm short tutorial book of Cellular Neural Networks". John Wiley. 1993. 1- 14.
- [2] L. O. Chua, "Cellular Neural Networks: Theory". IEEE Transactions on Circuits and Systems. Vol. 35. No. 9. 1988. 1257-1272.
- [3] A. Flores, E. Gómez, "Tutorial sobre Redes Neuronales Celulares: Aplicación al Procesamiento de Imágenes". LIDETEA. Universidad La Salle. 2004.
- [4] T. B. Roska, P. Thira, L.O. Chua, "Detecting Simple Motion Using Cellular Neural Networks". Proc. First IEEE Int. Workshop on Cellular Neural Networks and their Applications. 1990. Budapest. 127-138 pp.

- [5] T. Roska, L. Chua, "Cellular Neural Networks with Non-Linear and Delay-Type Template Elements and Non-Uniform Grids". *Int'l Journal of Circuit Theory and Applications*. Vol. 20. 1992. 469-481 pp.
- [6] J. J. Raygoza-Panduro, J.J., Ortega-Cisneros S., Salazar Santos, J. C. Ibarra, De la Mora A., Cárdenas-Rodríguez R, "Prototipado de una Red Neuronal Celular Digital en FPGAs". Segundo congreso CONCIBE y CACIF. Vol.1. 2006. 1-6 pp.
- [7] A.G. Radvanyi, "On the rectangular grid representation of general CNN networks, Cellular Neural Networks and Their Applications". *Proceedings of the 2000 6th IEEE International Workshop, Digital Object Identifier No. 10.1109/CNNA.2000.877360*. May 2000. 387 – 393, 23-25 pp.
- [8] H. Harrer, "Multiple layer discrete-time cellular neural networks using time-variant templates". *IEEE Transactions on Circuits and Systems part II*. Vol. 40. 1993. 191–199 pp.
- [9] L. O. Chua and T. Roska, *Cellular neural networks and visual computing-Foundation and applications*. 2002. Cambridge University Press. 29-30 pp.
- [10] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*. 3rd ed. 2008. Prentice Hall. Upper Saddle River, NJ.
- [11] Xilinx Inc. *Getting Started with the Xilinx Virtex-6 FPGA ML605 Evaluation Kit, manual XPN 0402771-01 UG533 (v1.5)*. www.xilinx.com. Oct. 20 2011.
- [12] Xilinx Inc., *ML605 Hardware User Guide UG534 (v1.8)*. www.xilinx.com. Oct. 2 2012.
- [13] "Virtex-6 Family Overview. Xilinx. 2012.

Identificación de rostros por técnica de puntos de interés SURF

Cesar Benavides Alvarez

Universidad Autónoma Metropolitana-Iztapalapa, Departamento de Ing. Eléctrica, San Rafael Atlixco 186, Vicentina, C.P. 09340, México, D.F, Teléfono: 5804-4600
cesarba@xanum.uam.mx

Graciela Román Alonso

Universidad Autónoma Metropolitana-Iztapalapa, Departamento de Ing. Eléctrica, San Rafael Atlixco 186, Vicentina, C.P. 09340, México, D.F, Teléfono: 5804-4600
grac@xanum.uam.mx

Juan Villegas Cortez

Universidad Autónoma Metropolitana-Azcapotzalco, Departamento de Electrónica, Av. San Pablo Xalpa No.180, Col Reynosa Tamaulipas, C.P. 02200, México, D.F, Teléfono: 01 55 5318 9000
juanvc@azc.ua.mx

Carlos Avilés Cruz

Universidad Autónoma Metropolitana-Azcapotzalco, Departamento de Electrónica, Av. San Pablo Xalpa No.180, Col Reynosa Tamaulipas, C.P. 02200, México, D.F, Teléfono: 01 55 5318 9000
caviles@azc.uam.mx

Resumen

En este trabajo se muestra un sistema de identificación de rostros a través de la técnica de extracción y análisis de los puntos de interés SURF (Speeded Up Robust Features). En esta metodología se hace uso de los puntos extraídos a partir de aplicar una partición al rostro en rejillas para obtener ventanas de la imagen, por medio de una ventana que se desliza a través de la imagen a analizar; el sistema obtiene los puntos de interés y el descriptor correspondiente a estos, para cada una de las ventanas

(rejillas) de la imagen y son guardados en la base de datos de entrenamiento. Este problema requiere un alto costo computacional ya que para extraer 225 rejillas de una imagen y generar los descriptores, de manera secuencial tiene un alto costo computacional, por lo que también se muestra la implementación con técnicas de programación en paralelo, logrando una reducción significativa del tiempo de cómputo para hacer factible su aplicación. El sistema se probó con cuatro bases de datos de rostros y se ha alcanzado un resultado fiable.

Palabra(s) Clave(s): Matching, reconocimiento de rostros, reconocimiento de patrones, puntos de interés SURF.

1. Introducción

Considerando al rostro humano como la primera impresión de identidad entre personas, su análisis para lograr una clasificación y posterior reconocimiento es importante, con base a desarrollar aplicaciones que automaticen las tareas de reconocimiento de personas en gran cantidad o bien de forma repetitiva a fin de tener un mínimo de error en la tarea. Al hablar de la biometría, ésta se encarga del reconocimiento de personas de manera automática por medio de sus características físicas (retina, iris, huellas dactilares, rostro, etc.). El rostro (también cara y faz) es la parte de la cabeza que se tiene al frente, en los humanos la parte anteroinferior, incluye cejas, ojos, nariz, mejilla, boca, labios, dientes, piel y barbilla; es así que se considera también a la clasificación y reconocimiento del rostro una parte de aplicación del reconocimiento de patrones [1]; por otra parte, también puede entrar en el reconocimiento de objetos, si vemos al rostro como un objeto tridimensional que puede tener forma, iluminación, etc. Los estudios de reconocimiento facial han determinado que, los ojos, la boca y la nariz se encuentran entre las partes más importantes para recordar rostros; esto quiere decir que presentan características distinguibles que no se pueden encontrar en otras partes como la frente o las mejillas [2].

Dentro de los diferentes tipos de métodos para la extracción de características del rostro, los más destacados son los basados en apariencia, basados en plantillas y en geometría. En los basados en geometría [3,4], las características son obtenidas a través de la información geométrica, posición, tamaño y distancias de los rasgos de la cara; en estos métodos es importante seleccionar de manera cuidadosa los umbrales de las características. Los basados en apariencia [5,6], utilizan la técnica de segmentación por color, usando el color de la piel para separar la región del rostro; cualquier región que no tenga el color de la piel puede ser aceptada como zona candidata para la ubicación de la boca o los ojos. La efectividad de esta técnica es limitada a causa de los diferentes tonos de piel de las personas. Los basados en plantillas [7,8,9], se realizan por medio de la correspondencia, con alguna plantilla de puntos de referencia que se realiza previamente, la mayor correspondencia produce la energía mínima. El costo computacional en estos métodos es muy elevado y presenta grandes dificultades cuando las características cambian de manera significativa, e.g. en el caso de las oclusiones, expresiones, ojos cerrados, boca abierta, etc.

El problema del reconocimiento de rostros se torna difícil por la gran variabilidad que existe: en la rotación de la cabeza, su inclinación, la iluminación, el ángulo, la expresión facial, envejecimiento, etc. [10,11,12]. Los aspectos a considerar aquí son: la variabilidad interclases, la similitud entre clases y las propiedades particulares de cada base de datos de imágenes.

Uno de los métodos más utilizados para el reconocimiento de rostros mencionado en la literatura, es el de "Eigenfaces" descrito en [13], el cual toma un enfoque holístico para el reconocimiento facial. Utiliza los componentes principales de una distribución de rostros, los vectores propios de la matriz de covarianza del conjunto de imágenes del rostro, donde cada imagen con N píxeles se considera un punto del espacio *N-Dimensional*.

En este trabajo presentamos la propuesta de usar una técnica utilizando los puntos de interés SURF [14], para la clasificación y reconocimiento de rostros, ya que ha probado

funcionar muy bien para el reconocimiento de objetos y se plantea una variante aplicada al reconocimiento de rostros así como el uso de computo paralelo, para la extracción de las características. Se aplicara la metodología a fotos de rostros en condiciones ideales (rostro de frente, sin oclusiones).

2. Método

Los descriptores de características locales describen un pixel de una imagen a través del contenido local en una vecindad. Estos deben ser distintivos y al mismo tiempo robustos a cambios de las condiciones de visualización. En este tiempo se han propuesto diversos descriptores y detectores dentro de la literatura, se ha encontrado que el rendimiento del descriptor depende a menudo del detector del punto de interés [15]. Recientemente se ha encontrado en un estudio comparativo en [16] que el rendimiento es superior en las características locales para el reconocimiento de rostros en ambientes sin restricciones. Los descriptores SURF se han utilizado en combinación con una maquina de soporte vectorial (SVM) únicamente para los componentes del rostro [17]. Sin embargo existen pocos análisis detallados del uso de SURF aplicado al reconocimiento facial hasta el momento.

Los factores principales que se deben tener en consideración en la detección de rostros son: *(i)* iluminación, *(ii)* orientación, *(iii)* escala y *(iv)* textura. El análisis automatizado del contenido de una imagen se refiere a la detección y estudio de su contenido e.g. colores, formas, texturas o cualquier otra información que pueda derivarse de la propia imagen. Específicamente para la detección de rostros la metodología tradicional marca 4 etapas: *(i)* detección de la cara, *(ii)* alineación de la cara, *(iii)* extracción de las características y *(iv)* reconocimiento. A continuación tenemos las etapas secuenciales a partir cada una de las imágenes a analizar ver Fig. 1.

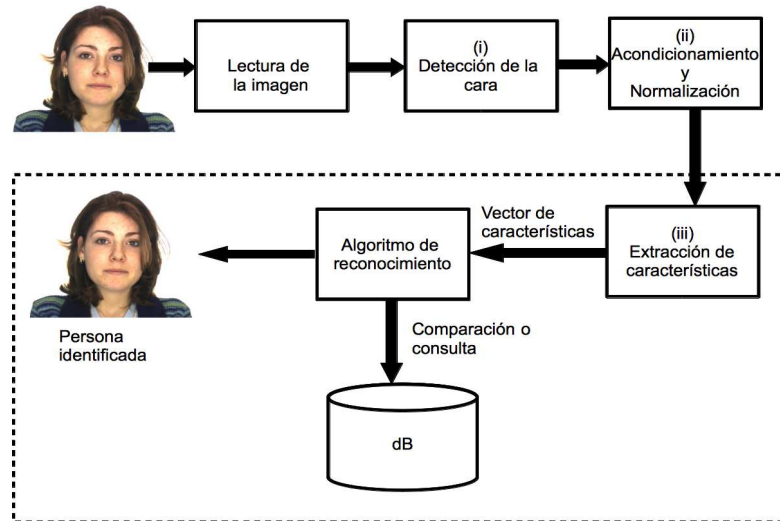


Fig. 1. Metodología clásica del reconocimiento de rostros.

En el presente proyecto se utiliza la técnica SURF la cual considera que todos estos factores no afectan el resultado del análisis ya que son invariantes a iluminación, rotación, escala, y no se utiliza el análisis de textura. Para la extracción de características solo es necesario el análisis de la región de interés, compuesta por una de las rejillas ver Fig. 2 considerado en [18].

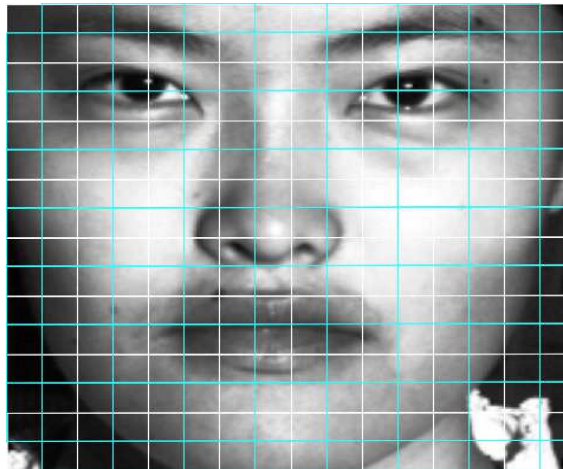


Fig. 2. Imagen del rostro descompuesta en regiones de análisis.

El uso e implementación de los atributos SURF en la metodología que planteamos, consta de una entrada, y una salida al sistema. En nuestra propuesta el sistema de

identificación y clasificación consta de dos etapas: entrenamiento y recuperación. El sistema propuesto ver Fig. 3, junto con su operación.

Entradas al sistema:

- Entrenamiento del sistema SURF propuesto con imágenes de rostros: Bases de imágenes de rostros en condiciones ideales.

Salidas del sistema:

- Recuperación: Las tres o cinco imágenes de las bases de datos que más se aproximen o parezcan a la imagen de consulta.

A. Entrenamiento

La arquitectura de entrenamiento consta de las siguientes partes:

- dB: Contiene las imágenes de rostros en condiciones ideales

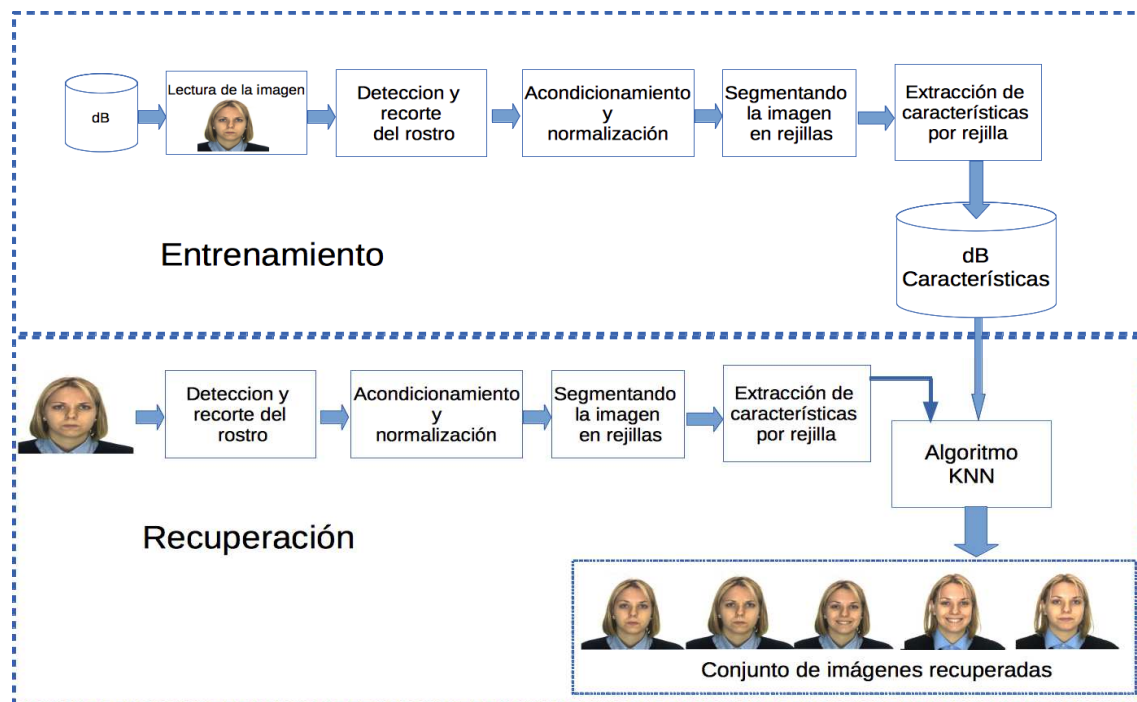


Fig. 3. Arquitectura del sistema SURF.

- Lectura de Imágenes: Lee las imágenes que serán procesadas.
- Detección y recorte del rostro: Utiliza el algoritmo de detección de rostro de la biblioteca OpenCV, una vez detectado el rostro se recorta de la imagen y se realiza un ajuste de tamaño las imágenes.
- Acondicionamiento y normalización: Realiza el acondicionamiento de las imágenes, tratando de eliminar impurezas que contengan las imágenes.
- Segmentando la imagen en rejillas: Se realiza una segmentación de la imagen en rejillas del mismo tamaño pasando una ventana deslizante de tamaño 40x40 pixeles, en total se generan 225 rejillas para cada imagen.
- Extracción de características por rejilla: A cada una de las rejillas de la imagen se le extraen los puntos de interés y el descriptor SURF para cada uno de esos puntos. Para cada imagen se obtendrán 225 contenedores de puntos de interés y 225 descriptores.
- dB Características: contiene la información de los puntos de interés y los descriptores de cada una de las imágenes, por lo que al final se tendrá un arreglo de puntos de interés y un arreglo de descriptores SURF: $VectorKeyPoins[Nimags*225]$ y $VectorDescriptors[Nimags*225]$

B. Recuperación

La arquitectura del sistema en la recuperación consta de una esquema similar al de entrenamiento, el cual se describe de la siguiente manera:

- Los bloques de las primeras cinco etapas se ejecutan de la misma manera que en la etapa de entrenamiento, hasta la parte de la extracción de características por rejilla.

- Aplicar algoritmo k -NN con $N=5$. Los 255 descriptores generados, son comparados uno a uno con el descriptor correspondiente en cada una de las imágenes en la base de datos de características y al final se extraen las imágenes más cercanas a la imagen de consulta que tuvieron el máximo número de puntos que realizaron el mejor "matching" en cada una de las rejillas. El resultado de los puntos que hacen "matching" se les aplica una operación de distancia Euclidiana para los puntos de interés involucrados, por medio de un radio mínimo para tener una tolerancia y asegurarse que la comparación realizada corresponde con la región analizada y no esta en una región más alejada ver Fig. 4, donde se observa un punto que esta dentro de la zona de búsqueda.

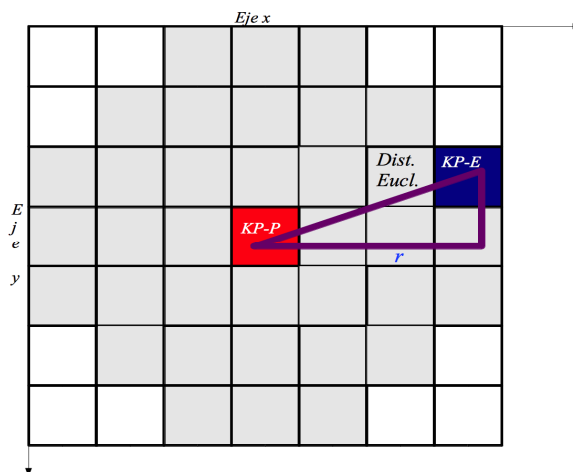


Fig. 4. "Matching" entre punto de interés de prueba y puntos de interés de entrenamiento.

C. Aceleración de la construcción de datos de entrenamiento.

El tiempo de construcción de la base de datos de entrenamiento depende principalmente del procesamiento de las N imágenes de una base de datos específica. Para reducir este tiempo, desarrollamos un algoritmo paralelo que utiliza las herramientas de OpenMP y OpenCV con las cuales el proceso de la extracción de características se paraleliza, por lo que podemos acelerar el

análisis de las imágenes ya que se procesan n-rejillas en paralelo, en lugar de hacerlo de la forma serial, es decir una tras de otra.

El algoritmo paralelo consta esencialmente de los siguientes pasos.

1. El proceso principal lee una de las imágenes a procesar.
2. Se generan 15 hilos los cuales procesan 15 rejillas de 40x40 pixeles cada uno, se extraen los puntos de interés y descriptores de cada rejilla y guardan esa información en el arreglo correspondiente.
3. Al final del proceso de entrenamiento se tienen dos arreglos: uno contiene los puntos de interés de todas las imágenes $VectorPuntosInterés[Nimag*225]$ y el otro contiene todos los descriptores $VectorDescriptores[Nimag*225]$.

D. Pruebas

Esta metodología se probó con las bases de datos de imágenes de rostros: ArDatabase [19], BioID [20], ExtendedYaleFaces [21] y YaleFaces [22]. Algunos ejemplos de los rostros de las bases de datos ver Fig. 5 se muestran respectivamente. Estas bases de datos no tienen el mismo tipo de resolución, así como el formato de la imagen, por lo que los programas se ajustaron para cada una de ellas. La cantidad de imágenes por cada base de datos, así como sus características se describe de manera detallada, para su completo entendimiento ver Tabla 1.

Para la versión de programación secuencial y paralela se implementó sobre una computadora tipo WorkStation en lenguaje C++, en una computadora tipo WorkStation, con sistema operativo Linux de 64 bits, con 16Gb de ram, procesador intel Xeon E5-2620 con 12 núcleos a una velocidad de 2Ghz, así como el uso de las herramientas OpenMPI, OpenMP y OpenCV, para el manejo de hilos en memoria compartida y procesamiento digital de imágenes respectivamente.



Fig. 5. Ejemplo de imágenes de las bases de datos: (a) Yale Faces, (b) Extended Yale, (c) AR, (d) BioID.

Para cada base de datos se extraerán dos mediciones en base a los porcentajes de recuperación:

1. Precisión

$$\text{precisión} = \frac{\text{No_imagenes_recuperadas}}{\text{total_imagenes_por_persona}} \times 100\% \quad (1)$$

2. Recuperación

$$\text{recuperación} = \frac{\text{No_imagenes_recuperadas}}{\text{total_img_en_la_dB}} \times 100\% \quad (2)$$

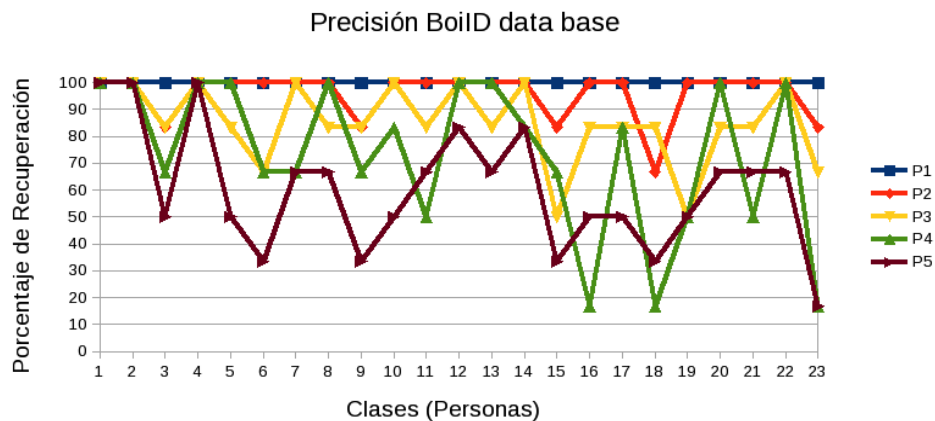
| dB | Número de imágenes | Tipo de imagen | Formato | Resolución |
|-------------------------|---------------------------|-----------------------|----------------|-------------------|
| AR data base | 4000 | Color | Jpg | 768x576 |
| BioID | 1521 | Gray Scale | Pgm | 648x286 |
| Extended Yale data base | 16128 | Gray Scale | Pgm | 640x480 |
| Yale data base | 165 | Gray Scale | Gif | 320x243 |

Tabla 1. Descripción de las bases de datos.

3. Resultados

En esta sección se presentan los resultados generados por la metodología propuesta anteriormente. Se trabajó con las bases de datos BioID, AR data base, Yale Faces data base y Extended Yale Faces data base. La programación se realizó en C, C++ y OpenCV.

Para cada una de las bases de datos, se seleccionaron las imágenes que cumplen con la descripción de condiciones ideales mencionadas anteriormente, a cada una de estas bases de datos por el momento solo se seleccionaron 6 imágenes por clase (persona), se extrajeron las características descriptivas de cada una de las imágenes. A partir de este análisis se obtuvieron las k imágenes más cercanas a la imagen de consulta para el algoritmo k - MN , usando la métrica Euclidiana en las pruebas. Usando la matriz de confusión podemos analizar el porcentaje de clasificación para los experimentos realizados, como se muestra en los resultados que se obtuvieron de clasificación para las bases de datos ver Fig. 6; teniendo que para el método de validación por resustitución el 100% de recuperación para la 1er imagen más parecida y los porcentajes de recuperación para las siguientes imágenes más parecidas va en disminución como se observa en las gráficas obtenidas para cada una de las bases de datos.



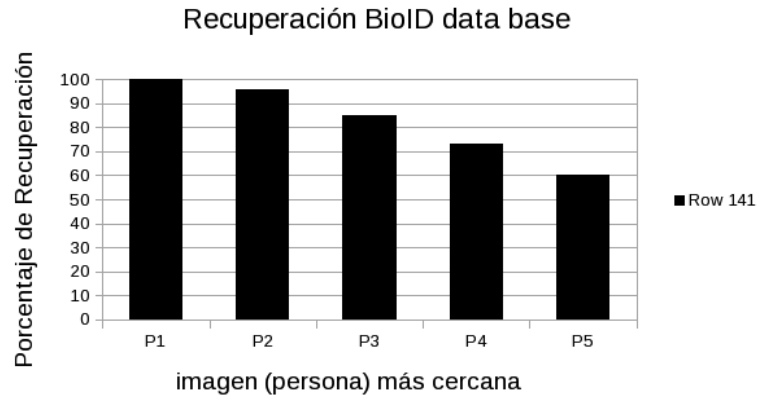


Fig. 6. (a) Gráficas de recuperación y precisión para la base de datos Biold.

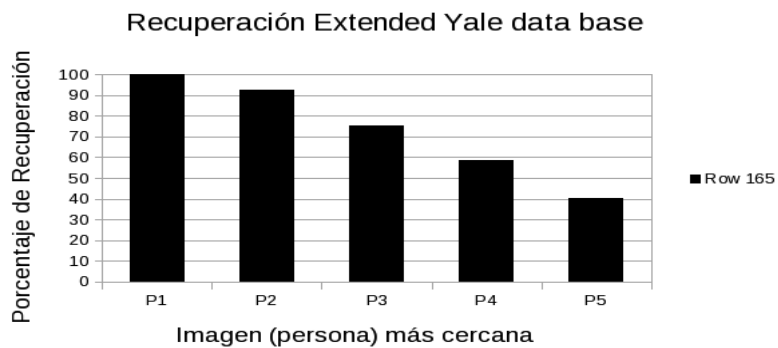
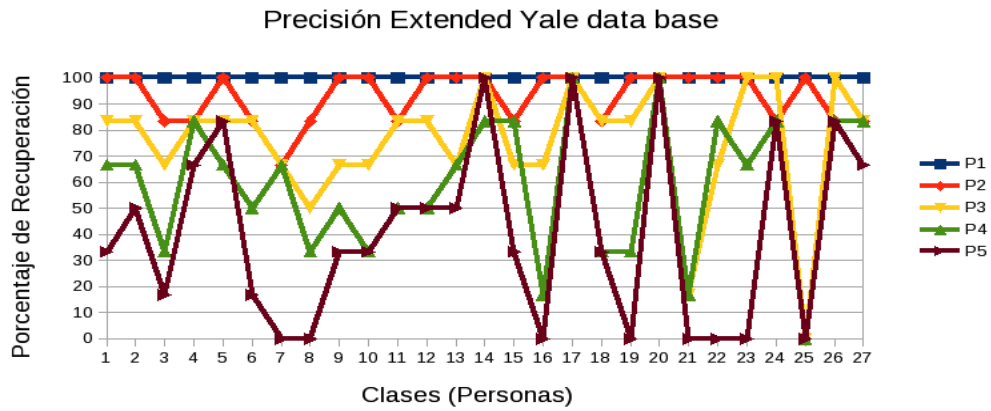


Fig. 6. (b) Gráficas de recuperación y precisión para la base de datos Extended Yale.

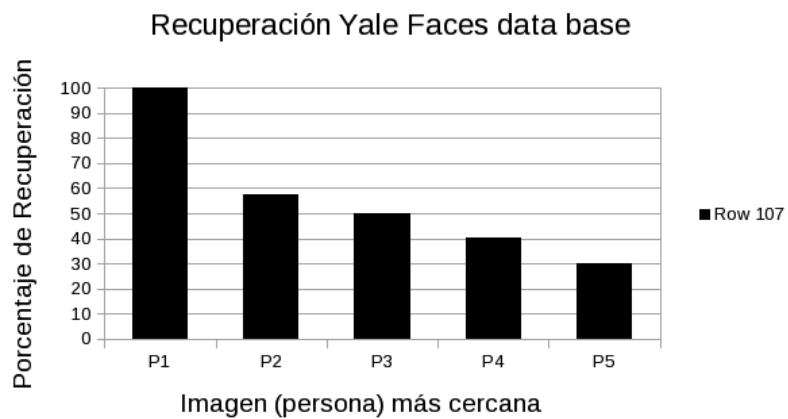
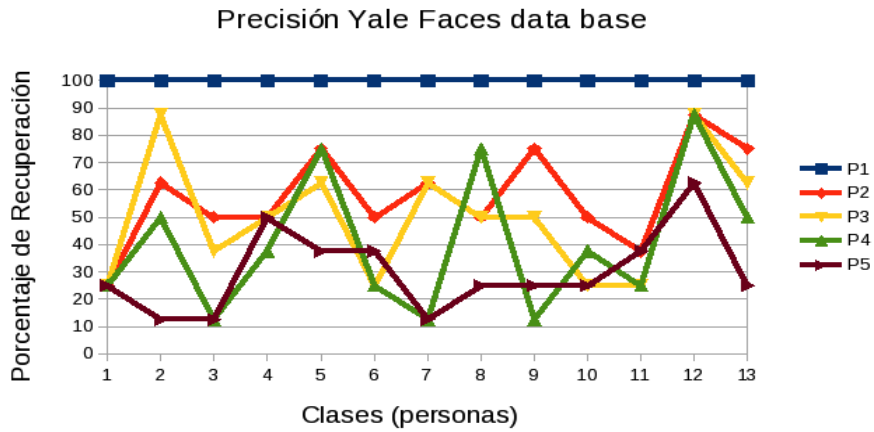
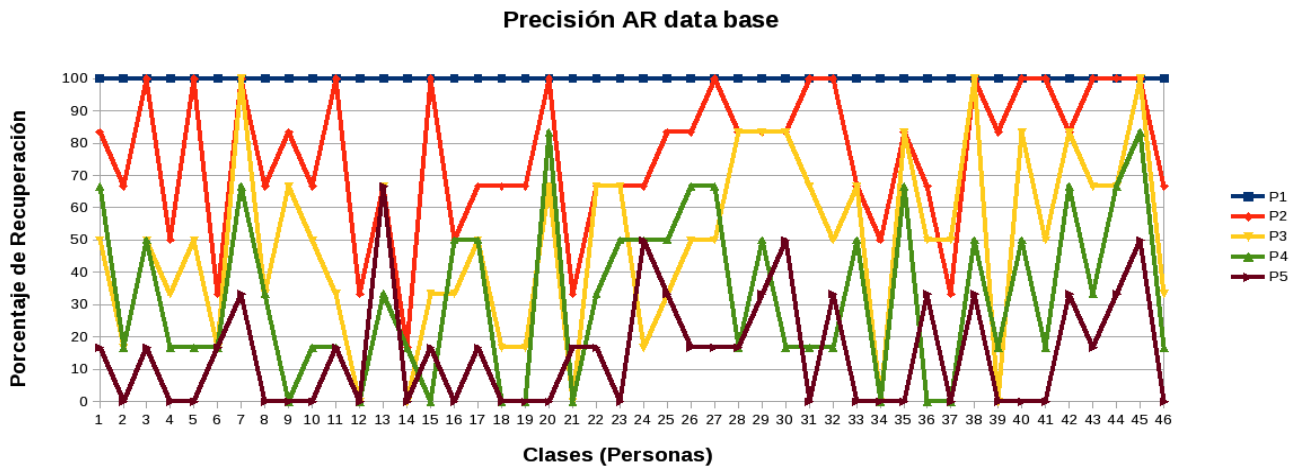


Fig. 6 (c) Gráficas de recuperación y precisión para la base de datos Yale Faces.



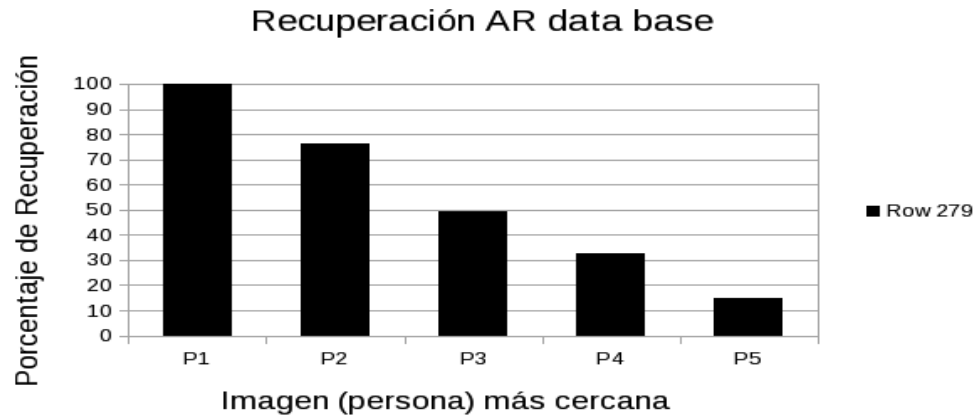


Fig. 6 (d) Gráficas de recuperación y precisión para la base de datos AR.

4. Discusión

En las gráficas de recuperación vemos en el eje y el porcentaje de recuperación y en el eje x , la imagen recuperada más cercana p_1 es la primera más parecida, p_2 la segunda y así sucesivamente hasta p_5 , para el valor de $k=5$ en $k-NN$.

En las gráficas de precisión vemos en el eje y el porcentaje de recuperación y en el eje x el número de clases (personas) que contiene la base de datos, y las series de la gráfica representan los porcentajes generados para cada una de las clases, ya que cada base de datos contiene por persona el mismo número de imágenes, en este caso 6 imágenes por persona. Por lo que para la 1er imagen recuperada se tiene de nuevo el 100% de recuperación para las 6 imágenes esperadas, y el porcentaje varía para las imágenes siguientes {2, 3, 4, 5} de cada persona como se observa en las gráficas de precisión de cada base de datos.

En la gráfica ver Fig. 6 (a) se observan los porcentajes de recuperación obtenidos de las 5 personas identificadas más cercanas a una de las imágenes de consulta, cada una de las líneas representa el porcentaje de clasificación, al modificar el valor de k en el algoritmo de $k-NN$, donde $k=\{1,2,3,4,5\}$, lo mismo pasa en la en las gráficas de las

demás bases de datos ver Fig. 6 (b), ver Fig. 6 (c) , ver Fig. 6 (d), para el calculo de la recuperación.

En la gráfica de barras ver Fig. 6 (a), se observa el porcentaje total promedio (precisión), que se obtiene para las k imágenes más cercanas ala imagen de consulta, donde se observa que con forme se piden más imágenes recuperadas el porcentaje de clasificación disminuye considerablemente, lo mismo en las gráficas de las demás bases de datos ver Fig. 6 (b), ver Fig. 6 (c) , ver Fig. 6 (d), para el calculo de precisión.

Ejemplos de recuperación de imágenes por consulta de la propia imagen ver Fig. 7, para cada una de las bases trabajadas respectivamente.

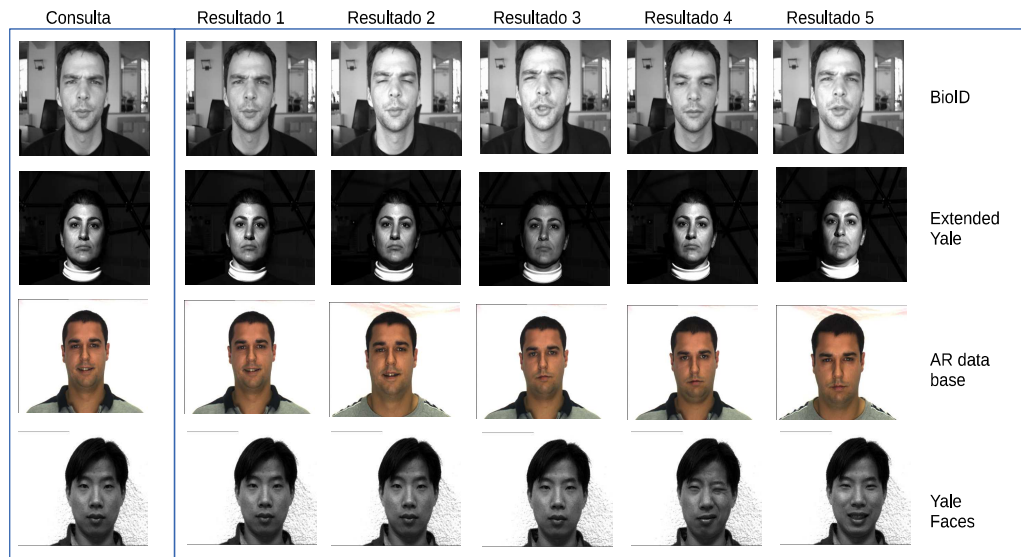
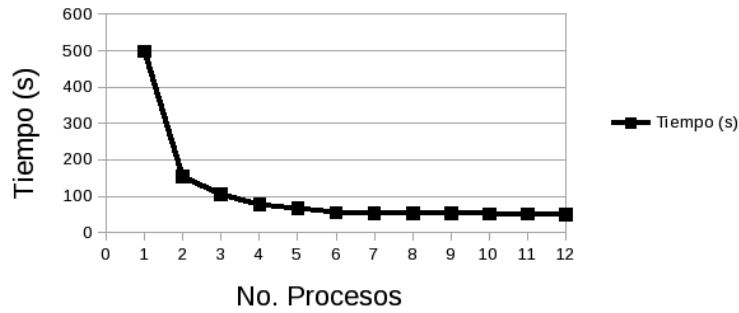


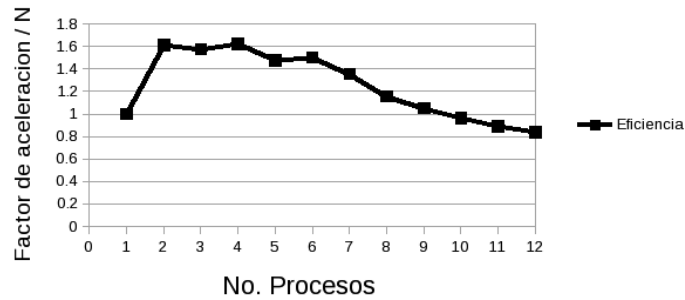
Fig. 7. Resultado para cada una de las bases de datos.

Tiempo Total Vs No. Procesos



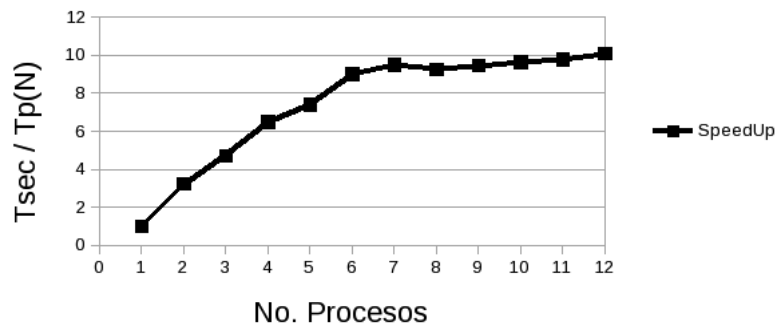
(a)

Eficiencia



(b)

Factor de Aceleración



(c)

Fig. 8. Graficas de análisis computacional: (a) Tiempo paralelo vs procesos, (b) Eficiencia, (c) Factor de aceleración.

Se observó que el tiempo de ejecución tuvo una reducción considerable mientras aumentaba el número de procesos ver Fig. 8 (a). El factor de aceleración (FA) ver Fig. 8 (b) de la ecuación (3), el cual va aumentado de manera lineal hasta el punto 6 donde se observa una caída y se vuelve a incrementar a partir del proceso 9.

$$FA = \frac{T_{sec\ uencial}}{Tp(N)} \quad (3)$$

Por último eficiencia ver Fig. 8 (c) la cual es calculada como se ve en la ecuación (4):

$$Eficiencia = \frac{FA}{Tp(N)} \quad (4)$$

En donde se observa que la eficiencia decrece de manera considerable a partir del proceso 6 y después del proceso 9 esta se mantiene estable.

Aunque en el análisis del desempeño se observa que a partir del proceso 6 el rendimiento decrece un poco, tanto en el factor de aceleración como en la eficiencia, por lo que parece que la máquina simula tener 12 núcleos y que se trata de un sistema Hyper Threading, por lo que tendría 6 microprocesadores reales y por este motivo, se observa un decremento en la evaluación de nuestro sistema dentro de esta máquina.

5. Conclusiones

Se propuso una nueva implementación a la técnica SURF, utilizada ahora para el reconocimiento de rostros. Se tuvo que esta técnica genera muy buenos resultados cuando las imágenes con las que se trabaja cumplen con las condiciones ideales, como lo es la iluminación, la resolución y que las imágenes son de la misma sesión o tiempo, como en el caso de las bases de datos BioID y MUCT, no así en AR data base puesto que algunas imágenes son tomadas en diferentes temporadas y las resoluciones de ellas hacen que el análisis de puntos surf tenga demasiadas variaciones.

Como trabajo futuro se planean dos tareas inmediatas: (i) aumentar el número de imágenes para poder medir el desempeño de la metodología propuesta en grandes volúmenes de información, y (ii) complementar esta técnica enriqueciéndola con el análisis de textura local del rostro a partir de los puntos de interés, usando la técnica CBIR, la cual se ha estudiado que funciona muy bien para la recuperación de escenarios naturales, y así generar un patrón descriptivo más robusto el cual disminuya los errores de recuperación y al mismo tiempo aumente los porcentajes de clasificación.

Con base a las estimaciones experimentales con las bases de datos hasta ahora probadas, será interesante reportar el desempeño de reducción de tiempo de cómputo con bases de datos mucho mayores, sobre la versión de programación paralela, y estudiar el nivel de recuperación sobre un número mucho mayor de imágenes.

6. Bibliografía

- [1] T de la Luz-Raul, "System de reconocimiento facial por medio de eigenfaces y redes neuronales". M.S. thesis. Esime IPN. Junio 2009.
- [2] A-González and F-Prieto, "Extracción de puntos característicos del rostro para medidas antropométricas". *Revista Ingenierías Universidad de Medellín*. Vol. 9. No. 17. Julio-diciembre 2010. 139–150 pp.
- [3] K-A-Kim. Facial feature extraction using pca and wavelet multi-resolution images. sixth IEEE International Conference on Automatic Face and Gesture Recognition. Mayo 2004. 17–19 pp.
- [4] Y-Zaho, Part-based pca for facial feature extraction and classification. International Workshop in Haptic Audio visual Environments and Games. Noviembre 2009. 99–104 pp.
- [5] B-Amarapur, N-Patil, The facial texture extraction for face recognition based on geometrical approach. IEEE CCECE/CCGEI. Mayo 2006. 1936–1939 pp.

- [6] G-G-Yen, N-Nithianandan, Facial feaature extraction using genetic algorithm. Congress on Evolutionary Computation, 1895–1900. Mayo 2002.
- [7] M-H-Mahoor, M-Abdel-Mottaleb, Facial features extraction in color images using enhanced active shape model. 7th International Conference on Automatic Face and Gesture Recognition. 2006. 10–12 pp.
- [8] T-F-Cootes, “Active appearance models”. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 23. No. 6. 2001. 681–685 pp.
- [9] Y-Yokogawa, “A proposal of improved lip contour extraction method using deformable template matching and its application to dental treatment”. Sys-tems and Computers in Japan. Vol. 38. No. 5. 80–89 pp.
- [10] H. K. Ekenel, R. Stiefelhagen, Why is facial occlusion a challenging problem? In International Conference on Biometrics. Sassari, Italy. June 2009.
- [11] A. M. Martinez, “Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class”. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 24. No. 6. 2002. 748–763 pp.
- [12] L. Wolf, T. Hassner, Y. Taigman. Descriptor based methods in the wild. In ECCV, 2008.
- [13] M. Turk, A. Pentland, “Eigenfaces for recognition”. Journal of Cognitive Neuroscience. Vol. 3. 1991. 71–86 pp.
- [14] Surf: Speeded up robust features. Computer Vision and Image Understanding (CVIU). <http://www.vision.ee.ethz.ch/~surf/>.
- [15] A performance evaluation of local de- scriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence. <http://dx.doi.org/10.1109/TPAMI.2005.188>.

- [16] Javier Ruiz del Solar, Rodrigo Verschae, Mauricio Correa, Face recognition in unconstrained environments: A comparative study. In ECCV Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France, October 2008.
- [17] Donghoon Kim, Rozenn Dahyot, Face components detection using surf descriptors and svms. In International Machine Vision and Image Processing Conference. Portrush, Northern Ireland. September 2008. 51–56 pp.
- [18] Philippe Dreuw, Pascal Steingrube, Harald Hanselmann, Hermann Ney, "SURF-Face: Face Recognition Under Viewpoint Consistency Constraints". 2009. In A. Cavallaro, S. Prince and D. Alexander, editors. Proceedings of the British Machine Conference. September.
- [19] A.M. Martinez, R. Benavente, "The AR Face Database". CVC Technical Report No. 24. June 1998.
- [20] O. Jesorsky, K. Kirchberg, R. Frischholz, In J. Bigun, F. Smeraldi, editors, Audio and Video based Person Authentication - AVBPA. Springer, 2001.
- [21] Athinodoros Georghiades, Peter Belhumeur, David Kriegman's paper, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose". PAMI. 2001.
- [22] Yale University, Yale Face Database. URL:<http://vision.ucsd.edu/content/yale-face-database>. 5/06/2015.

7. Autores

Ing. Cesar Benavides Alvarez recibió su título de Ingeniero en Electrónica en 2012 en la Universidad Autónoma Metropolitana Campus Iztapalapa (UAM-Iztapalapa). Actualmente está estudiando una Maestría en Ciencias y Tecnologías de la Información en la Universidad Autónoma Metropolitana. Sus intereses de investigación son: Visión por Computadora, Reconocimiento de Patrones y Sistemas Distribuidos.

Dr. Juan Villegas Cortez es profesor-investigador en el Departamento de Electrónica de la Universidad Autónoma Metropolitana desde 2010 su intereses actuales de investigación incluyen reconocimiento de Patrones, Visión por Computadora, Computación Evolutiva, Memorias Asociativas y procesamiento digital de señales.

Dra. Graciela Román Alonso es profesora-investigadora en el Departamento de Ingeniería Eléctrica en la UAM Iztapalapa. Sus áreas de investigación son sistemas distribuidos, cómputo paralelo y programación avanzada.

Dr. Carlos Avilés Cruz Actualmente se desempeña como profesor-Investigador Titular de tiempo completo en el departamento de Electrónica de la UAM-Azcapotzalco, donde desarrolla trabajos de investigación en las áreas de Reconocimiento de Patrones, visión artificial e Interfaces Cerebro Computadora BCI.

Segmentación de Iris con OpenCV en Android

Oscar Fuentes Salome

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ciudad Universitaria, Av. San Claudio y 18 sur, Col. San Manuel, C.P. 72570, Puebla, Puebla, México
ofuentessalome@gmail.com

Aldrin Barreto Flores

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Electrónica, Ciudad Universitaria, Av. San Claudio y 18 sur, Col. San Manuel, C.P. 72570, Puebla, Puebla, México, Teléfono: (222) 2295500
abarreto@ece.buap.mx

Verónica Edith Bautista López

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ciudad Universitaria, Av. San Claudio y 18 sur, Col. San Manuel, C.P. 72570, Puebla, Puebla, México, Teléfono: (222) 2295500
vbautista@cs.buap.mx

Resumen

En este trabajo presentamos la implementación de un algoritmo para la segmentación del iris contenido en imágenes oculares con el propósito de contribuir a la realización de un sistema de diagnóstico automático que pueda ser usado por la Iridología (detección de lesiones en el iris). Para realizar la segmentación, primeramente se localiza la pupila, posteriormente se detecta el iris y finalmente se crea una máscara para dejar únicamente el área correspondiente al iris. Se emplea la transformada circular de Hough para obtener la pupila e iris. El algoritmo fue implementado en una aplicación

móvil (app) para dispositivos con sistema operativo Android y se usaron diversos algoritmos de procesamiento digital de imágenes implementados en OpenCV.

Palabras Claves: Android, OpenCV, procesamiento digital de imágenes, segmentación de Iris, transformada circular de Hough.

1. Introducción

En el campo de la Iridología se considera que existe una estrecha conexión entre el iris y cada uno de los órganos y tejidos del cuerpo, la cual se establece mediante el cerebro y el sistema nervioso. Debido a esta conexión, el estado general de salud de un ser humano, así como la presencia de residuos tóxicos y su acumulación en el cuerpo podrían manifestarse en el iris a través de ciertas lesiones (signos, marcas o coloraciones)(véase [1], p. 6,7).

La Iridología como medicina alternativa proporciona grandes ventajas: el equipo necesario es mínimo, únicamente se requiere de una cámara fotográfica para capturar una imagen del iris, pues si bien es cierto que el diagnóstico se puede realizar directamente observando el iris del paciente, se recomienda el uso de una cámara para obtener imágenes de los ojos a fin de que este proceso sea más fácil y efectivo; es un método de diagnóstico inofensivo, es decir, no produce ningún daño, molestia o efecto secundario, debido a que sólo es necesario enfocar con una fuente luminosa los ojos del paciente y posiblemente capturar fotografías de estos (véase [2], p. 327).

El diagnóstico de padecimientos usando la Iridología se realiza analizando visualmente las características de las lesiones, por ejemplo: forma, textura, coloración, entre otras, y se asocian con un padecimiento en un determinado órgano del cuerpo de acuerdo a su localización en el iris relacionándola a su vez con alguna de las cartas iridológicas (véase [1], p. 8, 9,10, [2], p. 332). Estas cartas iridológicas son mapas de los iris (izquierdo y derecho) que muestran las áreas del iris correspondientes a cada órgano del cuerpo.

Se han creado diversas herramientas de software con el fin de apoyar al iridólogo (persona que tiene conocimientos en Iridología) en su diagnóstico, las cuales permiten principalmente: gestionar los datos del paciente y las imágenes de sus iris, aplicar operaciones a la imagen (zoom, brillo, contraste) y realizar marcas a esta. Por otra parte, también se han realizado trabajos con el propósito de automatizar el diagnóstico de algunas de las lesiones más importantes en Iridología usando algoritmos de Procesamiento Digital de Imágenes (PDI) y Visión Artificial (VA), ejemplos de esto se muestran en [3, 4, 5, 6]. En estos trabajos se identifica algún tipo de lesión, sin embargo, el reconocimiento de patrones (lesiones) en una imagen requiere de etapas previas. Una de estas etapas previas es la segmentación, la cual se refiere a dividir la imagen en las regiones u objetos que la componen y es decisiva en la identificación correcta de los objetos de interés (véase [7], p. 49).

Las técnicas que se han usado para afrontar la segmentación de iris son diversas, en [8] se utiliza difusión anisotrópica para segmentar el iris en imágenes no idealmente capturadas. En [9] se propone una metodología basada en texturas para realizar la segmentación. Con el mismo propósito, en [10] también se emplea la transformada de Hough. En [11] se utilizan contornos activos geodésicos para extraer el iris. Debido a la importancia que adquiere la segmentación de iris en el desarrollo de un sistema de diagnóstico automatizado basado en técnicas de Iridología hemos realizado este trabajo, en el cual presentamos un algoritmo para la segmentación de iris basado principalmente en la transformada de Hough.

Además usamos algunos de los algoritmos de OpenCV (véase [14], p. 158) dado que es una librería de código abierto que contiene una gran cantidad de algoritmos de VA. OpenCV está disponible para usarse en Android, por lo que la implementación del algoritmo propuesto la realizamos en este sistema operativo.

La organización de este documento es la siguiente: en la sección 2 describimos las etapas del algoritmo propuesto, así como las técnicas de PDI empleadas en cada una

de estas. Posteriormente, en la sección 3 describimos los resultados de las pruebas experimentales.

Finalmente, se presentan las conclusiones y el trabajo a futuro.

2. Desarrollo

El algoritmo propuesto en este trabajo se compone de dos etapas. En la primera etapa se identifica la pupila y en la segunda, el iris. Este algoritmo lo hemos desarrollado para su funcionamiento con imágenes capturadas bajo un ambiente no controlado y con características de captura (como la iluminación, ángulo y distancia de captura, cámara, resolución) diferentes, ya que parte del objetivo del trabajo desarrollado fue implementarlo en una aplicación móvil para que sea utilizado por cualquier persona.

2.1. Identificación de la pupila

La primera etapa se inicia aplicando un filtro de suavizado a la imagen ocular. Específicamente se usó la media aritmética y determinamos experimentalmente que era necesario aplicar dos iteraciones. Este filtro reduce las transiciones fuertes, lo cual ayuda para evitar la detección de falsos bordes (véase [12], p. 269). La aplicación de este filtro se realiza mediante convolución con el kernel que se muestra en la Fig. 1. Una vez suavizada la imagen original, se transforma a escala de grises.

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Fig. 1. Matriz de convolución para suavizar la imagen

Posteriormente, se debe calcular el histograma de la imagen. Siguiendo una idea similar a la que se muestra en [13], se analizó el histograma de varias imágenes oculares en escala de grises y se pudo observar que la pupila aparecía como la montaña más cercana al origen, debido a que generalmente, corresponde a los píxeles con tonos más oscuros, un ejemplo de esto se puede apreciar en la Fig. 2. En base al patrón que representa la pupila en el histograma, se dedujo que la identificación del límite interior del iris en la imagen, es decir, la pupila, se podía segmentar mediante umbralización (véase [20], p. 108).

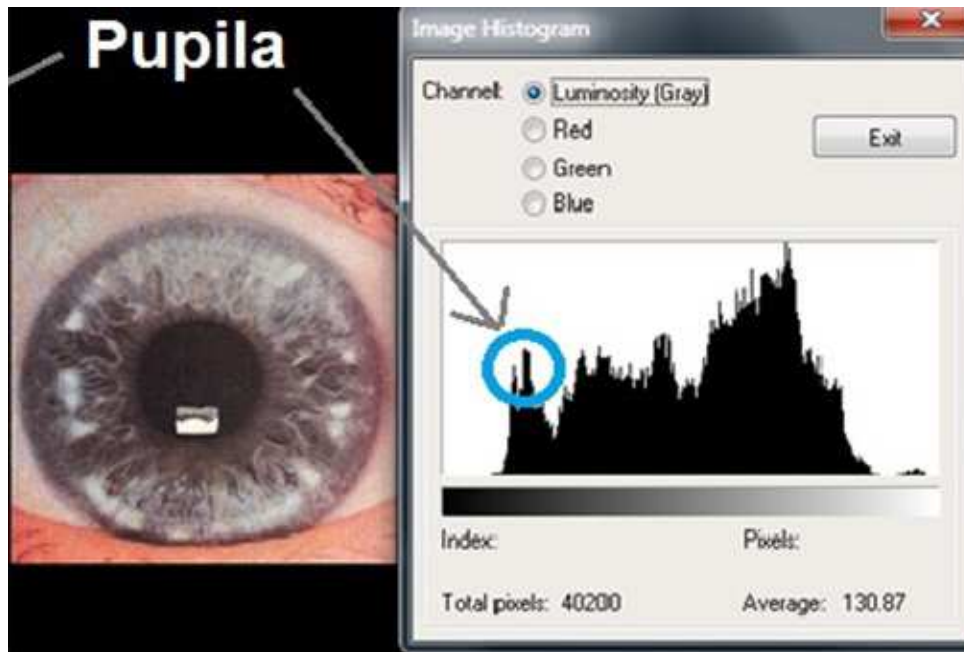


Fig. 2. La primera montaña en el histograma representa la pupila

La umbralización, sin embargo, requiere de un umbral (u), es decir, un valor mínimo para discriminar entre dos clases de píxeles, los que pertenecen y los que no pertenecen a la pupila. Visualmente, mediante la observación del histograma, podía aproximarse el valor del umbral, pero establecer un valor específico para todas las imágenes no era posible debido a que, como se mencionó en párrafos anteriores, las condiciones de captura variaban drásticamente entre una imagen y otra. En general, el valor de u es el valor máximo del histograma de la imagen más cercano al origen, pero

antes de determinar dicho valor es necesario suavizar el histograma, esto a fin de evitar falsos máximos que pudieran aparecer.

Para suavizar el histograma de una imagen se desarrolló y empleó el algoritmo 1, Fig. 3. Este algoritmo requiere que el histograma de la imagen se haya previamente calculado (*histograma [i]*), al cual se le aplica una mediana con una ventana de 5 elementos (líneas de la 3 a la 9) y a continuación una media entre dos valores seguidos (línea 11, 12 y 13).

De la línea 4 a la 6 se copian del histograma 5 elementos consecutivos a *subarreglo[i]*, los cuales se ordenan en la línea 7 y en la línea 8 se obtiene la mediana del subarreglo para asignarla al histograma en la posición *i*. Este proceso se repite hasta la posición 253 del histograma (líneas de la 3 a la 9).

De la línea 11 a la 12 se suaviza el histograma parcial mediante la media entre dos elementos consecutivos y se redondea la división con el propósito de obtener un valor entero para el histograma suavizado. Este segundo proceso se repite hasta la posición 255 del histograma.

En la Fig. 4 se puede observar una imagen, su histograma original y su histograma suavizado. Tras suavizar el histograma se puede observar cómo queda dividido en regiones y los falsos máximos y mínimos desaparecen. Habiendo suavizado el histograma se puede localizar el valor de *u*.

Algoritmo 1 Suaviza el histograma de una imagen

Entrada: Histograma []

Salida: HistogramaSuavizado []

1: *subarreglo* [5]

2: //Se aplica la mediana

3: **para** $i \leftarrow 2$ **hasta** $i \leq \text{histograma.tamaño} - 2$ **hacer**

4: **para** $s \leftarrow -2$ **hasta** $s \leq 2$ **hacer**

5: $\text{subarreglo}[s+2] \leftarrow \text{histograma}[i+s]$

6: **fin para**

7: **Ordenar** (*subarreglo*)

8: $\text{histograma}[i] \leftarrow \text{subarreglo}[2]$

9: **fin para**

10: // Se aplica la media

11: **para** $i \leftarrow 1$ **hasta** $i \leq \text{histograma.tamaño} - 1$ **hacer**

12: $\text{histogramaSuavizado}[i] \leftarrow \text{Redondear} ((\text{histograma}[i - 1] + \text{histograma}[i + 1]) / 2)$

13: **fin para**

Fig. 3. Algoritmo para suavizar el histograma de una imagen

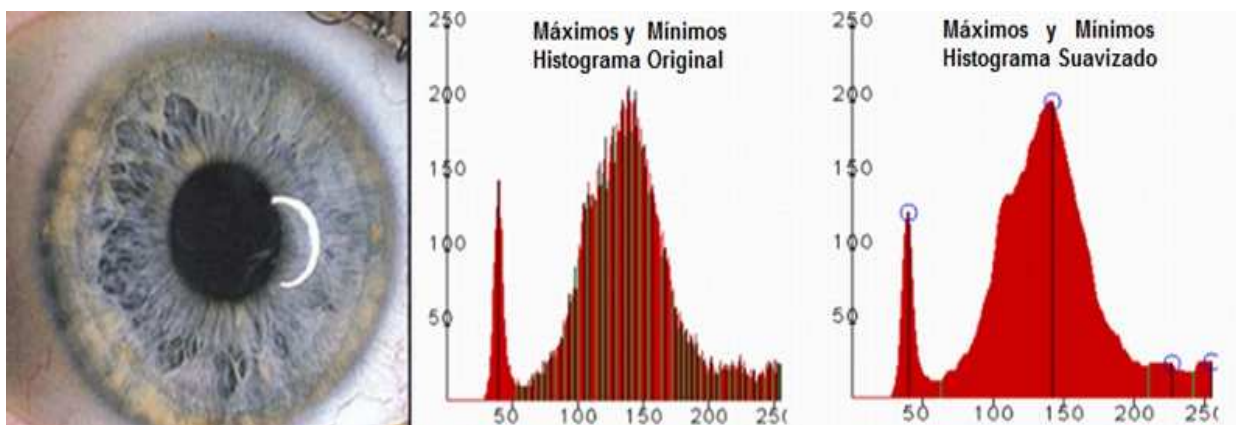


Fig. 4. Histograma original y suavizado

Con el objetivo de determinar u se desarrolló y empleó el algoritmo 2, Fig. 5, el cual asigna a un arreglo (MaxMin, inicializado con ceros) de 256 elementos, los siguientes valores: 1 si el tono correspondiente al índice en el histograma suavizado representa un máximo y -1 si corresponde a un mínimo. Se usan dos banderas DIRARRIBA que toma

el valor de 1 y DIRABAJO que toma el valor de -1 (línea 1 y 2) para que durante el recorrido del histograma suavizado en busca de los máximos y mínimos se pueda saber si se asciende o desciende, y dependiendo de esto, la variable sentido tomara el valor de DIRARRIBA o DIRABAJO respectivamente, de esta manera se sabe cuándo hay un cambio de sentido y se puede determinar si se trata de un máximo o un mínimo (línea 7 y 11).

El valor de umbralización se estableció a partir del primer máximo y mínimo empleando la ecuación 1.

$$u = (max + min) / 2 \quad (1)$$

Algoritmo 2 Obtiene los máximos y mínimos de un histograma

Entrada: histogramaSuavizado []

Salida: MaxMin [] // Arreglo con máximos y mínimos

```

1: DIRARRIBA ← 1
2: DIRABAJO ← -1
3: sentido ← DIRARRIBA
4: para i ← 2 hasta i ≤ histogramaSuavizado.tamaño - 1 hacer
5:   si sentido == DIRARRIBA & histogramaSuavizado [i] > histogramaSuavizado [i+1] entonces
6:     MaxMin[i] ← 1
7:     sentido ← DIRABAJO
8:     i++
9:   de lo contrario si sentido == DIRABAJO & histogramaSuavizado [i] < histogramaSuavizado [i+1] entonces
10:    MaxMin[i] ← -1
11:    sentido ← DIRARRIBA
12:    i++ fin si
13:  fin para
14: fin para
15: si histogramaSuavizado [histograma.tamaño - 1] < histogramaSuavizado [histogramaSuavizado.tamaño - 2] entonces
16:  MaxMin [histogramaSuavizado.tamaño - 1] ← -1
17: de lo contrario
18: MaxMin [histogramaSuavizado.tamaño - 1] ← 1
19: fin si

```

Fig. 5. Algoritmo para obtener los máximos y mínimos de un histograma

Una vez obtenido u se realiza una umbralización con ese valor, es decir se binariza la imagen mediante la ecuación 2.

$$I(x, y) = \begin{cases} 0, & \text{si } I(x, y) < u \\ 1, & \text{si } I(x, y) \geq u \end{cases} \quad (2)$$

Después de binarizar la imagen se obtiene segmentada la pupila, sin embargo, en algunos casos también permanecen puntos o regiones innecesarias (artefactos), este problema se puede apreciar en la Fig. 6. Dichos artefactos podrían afectar en la detección correcta de la pupila. Para evitar este problema se utiliza la transformada circular de Hough (véase [14], p. 158), ya que podemos considerar que la pupila es muy semejante a esta forma geométrica.

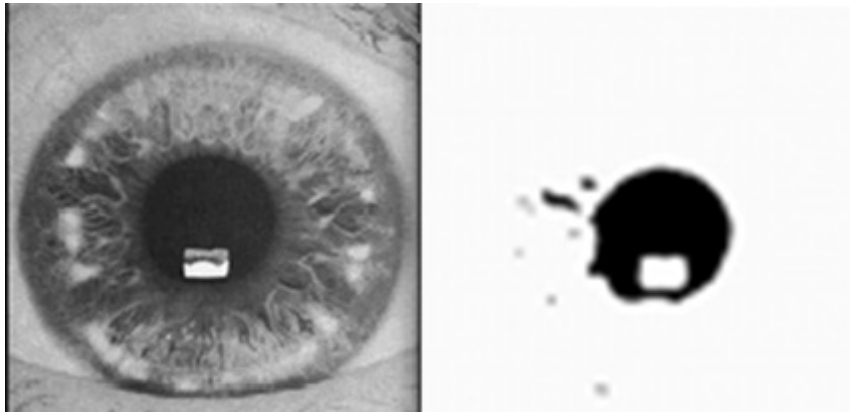


Fig. 6. Pupila segmentada y artefactos

La Transformada de Hough es una técnica que permite el reconocimiento de patrones globales en el espacio de la imagen, reconociendo patrones locales (idealmente un punto) en un espacio de parámetros transformado, en otras palabras permite reconocer formas paramétricas simples en una imagen usando para ello un espacio diferente, llamado espacio de Hough. Este método se usa principalmente para la detección de rectas, la cual es la forma más básica de implementarla, sin embargo, también se puede utilizar para reconocer polinomios, círculos, etc. Utilizando la transformada de Hough se puede detectar la pupila sin importar que posiblemente existan artefactos en la imagen, ver Fig. 7.

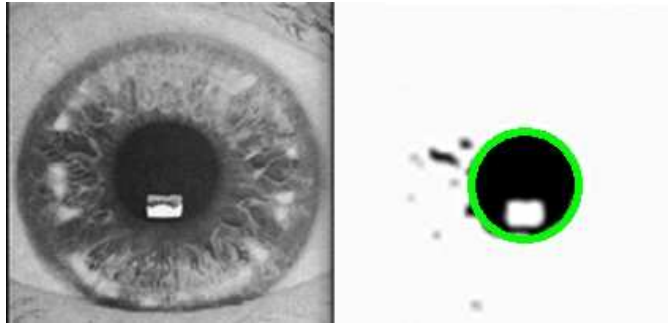


Fig. 7. Pupila detectada con la transformada circular de Hough

Para usar la transformada circular de Hough en nuestro trabajo, no hemos implementado el algoritmo como tal, sino utilizamos el método `HoughCircles` de la clase `Imgproc` de OpenCV, el cual tiene la siguiente sintaxis:

```
Imgproc.HoughCircles(Mat imag, Mat circulos, int metodo, int dp, int minDist, int param1, int param2, int minRadio, int maxRadio)
```

Dónde:

imag.- es la imagen en la que se desean detectar los círculos.

circulos.- es el vector donde se almacenaran los valores de las coordenadas del centro y el radio de cada círculo encontrado.

metodo.-indica el método que será usado para la detección. Por ahora, OpenCV únicamente utiliza el método `21HT`, el cual se indica mediante la constante `Imgproc.CV_HOUGH_GRADIENT`.

dp.- indica la proporción inversa de la resolución del acumulador respecto a la resolución de la imagen. Si $dp = 1$ el acumulador tendrá la misma resolución que la imagen, si $dp = 2$ el acumulador tendrá la mitad de la resolución de la imagen.

minDist.- es la distancia mínima entre círculos detectados. Si el parámetro es demasiado pequeño podrían ser detectados múltiples círculos falsos. Si es demasiado grande, algunos podrían ser omitidos.

param1.- este parámetro es un umbral usado por el detector de bordes Canny, el cual es usado internamente en este método (HoughCircles).

param2.- se refiere al valor del acumulador. Los círculos con valor de acumulador más altos serán retornados primero.

minRadio.- radio mínimo de los círculos.

maxRadio.- radio máximo de los círculos.

La llamada al método `Imgproc.HoughCircles` se realizó con los siguientes valores:

`dp = 1, minDist = 50, param1 = 150, param2 = 13, minRadio = mImag.height()/8, maxRadio = mImag.height()/5.`

Los parámetros del método `HoughCircles` más decisivos para obtener una segmentación satisfactoria son el valor del acumulador, y el tamaño mínimo y máximo de los radios buscados en la imagen. El valor del acumulador (`param2`) se estableció en 13, lo que significa que un conjunto de puntos equidistantes a un punto en común se debía considerar como círculo si por lo menos tenía 13 elementos, esto debido a que en ocasiones al umbralizar la mayor parte de la periferia de la pupila se elimina o queda demasiado discontinua. El radio mínimo y máximo se determinó mediante la comparación del radio de la pupila con la altura de la imagen que la contenía. Para determinar el valor del radio mínimo se consideró el caso en el que la pupila se encontraba contraída; y para el del radio máximo, el caso en el que la pupila se encontraba dilatada. En la Fig.8, en el lado izquierdo se muestra el modelo de una imagen ocular donde la pupila se encuentra dilatada (de una forma un poco exagerada). Se puede observar que el radio de esta pupila es aproximadamente una quinta parte de la altura de imagen. De manera similar, en el lado derecho de la figura, se muestra un modelo de una imagen ocular donde la pupila se encuentra contraída (de una forma también un poco exagerada) la cual tiene un radio aproximadamente igual a una octava parte de la altura de dicha imagen. Tomando en consideración lo anterior, se estableció

que los valores óptimos para el radio mínimo y máximo eran $\text{mlmag.height()}/8$ y $\text{mlmag.height()}/5$ respectivamente.

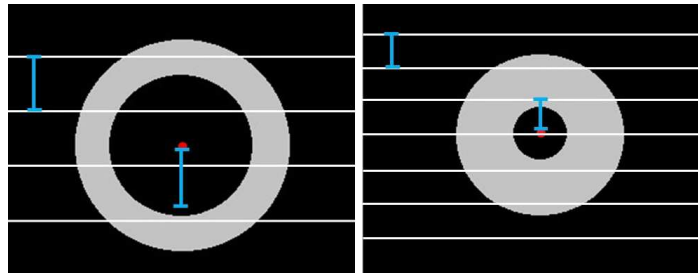


Fig. 8. Modelo de una pupila dilatada (izquierda) y una contraída (derecha)

Una vez que se ejecuta el método `HoughCircles`, se obtiene en el vector que se pasa como segundo parámetro los valores de las coordenadas del centro y el radio del círculo encontrado, es decir la pupila.

2.2. Identificación del borde exterior del iris

Para la detección del borde exterior del iris, el procedimiento que se utilizó es similar al utilizado en la detección de la pupila. Esta etapa inicia aplicando el filtro de la mediana a la imagen original para reducir el ruido de la imagen ocular. El siguiente paso es calcular su histograma y suavizarlo mediante el algoritmo 1. Después se calculan todos los máximos y mínimos empleando el algoritmo 2. A partir de que se obtuvo el histograma dividido en intervalos, se busca el intervalo más grande, el cual se pudo observar, analizando el histograma de varias imágenes, es el rango de tonos correspondientes al iris. Para averiguar el intervalo se usó el algoritmo 3, ver Fig. 9.

El algoritmo 3 obtiene el punto inicial (*intervalo [0]*) y el punto final (*intervalo[1]*) de la región con mayor área en el histograma. Se definió como región al intervalo que existe entre dos puntos mínimos consecutivos. La búsqueda de la primera región inicia estableciendo como punto inicial (punto1) la posición 0 del histograma (línea 8) y después la posición inicial de la búsqueda es donde se encuentre un mínimo (-1). Mientras la bandera punto2 encontrado se mantenga en falso la búsqueda del punto

final de la región continuará (línea 7 a la 16). Una vez encontrado el intervalo de la región (punto1 y punto2), se calcula el área de la región y se almacena en la variable *suma*. Posteriormente si el número de intervalos (*numInter*) es mayor que uno (línea 23), es decir, ya se ha calculado otra área, se compara el área recién calculada con la anterior (línea 24), si esto también es cierto *sumaAnterior* tomará el valor de *suma* que es mayor al área y se establece el intervalo en el que se encuentra. Si es la primera región que se encuentra entonces *suma Anterior* toma el valor de *suma* (línea 30). Este proceso se repite dependiendo del número de mínimos en el histograma suavizado.

Con el valor de $u = (p1 + p2) / 2$ se umbraliza la imagen previamente suavizada con la matriz de convolución mostrada en la Fig.1. Después de haber aplicado la umbralización, se utiliza nuevamente el método HoughCircles de OpenCV para detectar el círculo que representa el iris en la imagen ocular, es decir, con este método se obtienen los valores de las coordenadas del centro y de su radio, Fig.10.

Algoritmo 3 Obtiene el intervalo de los tonos correspondientes al iris

Entrada: histogramaSuavizado [], valoresMaxMin []
Salida: intervaloIris []

- 1: $p2Encontrado \leftarrow$ falso
- 2: $sumaAnterior \leftarrow 0$
- 3: $numInter \leftarrow 0$
- 4: $i \leftarrow 0$
- 5: $punto2 \leftarrow 0$
- 6: //Se obtiene el intervalo
- 7: **mientras** $i < valoresMaxMin.tamaño$ **hacer**
- 8: $punto1 \leftarrow i$
- 9: **mientras** $punto2Encontrado \neq true$ **hacer**
- 10: **si** $valoresMaxMin [i] == -1$ o $i == valoresMaxMin.tamaño - 1$
 entonces
- 11: $punto2 \leftarrow i$
- 12: $punto2Encontrado \leftarrow verdadero$
- 13: $numInter ++$
- 14: **fin si**
- 15: $i ++$
- 16: **fin mientras**
- 17: //Se calcula el área del histograma en el intervalo
- 18: $suma \leftarrow 0$
- 19: **para** $j \leftarrow punto1$ hasta $j < punto2$ **hacer**
- 20: $suma \leftarrow suma + histogramaSuavizado[j]$
- 21: **fin para**
- 22: // Comparación de áreas, se decide qué área es la más grande
- 23: **si** $numInter > 1$ **entonces**
- 24: **si** $suma > sumaAnterior$ **entonces**
- 25: $sumaAnterior \leftarrow suma$
- 26: $intervalo [0] \leftarrow punto1$
- 27: $intervalo [1] \leftarrow p2$
- 28: **fin si**
- 29: **de lo contrario**
- 30: $sumaAnterior \leftarrow suma$
- 31: **fin si**
- 32: $p2encontrado \leftarrow false$
- 33: **fin mientras**

Fig. 9. Algoritmo para obtener el intervalo de los tonos correspondientes al iris

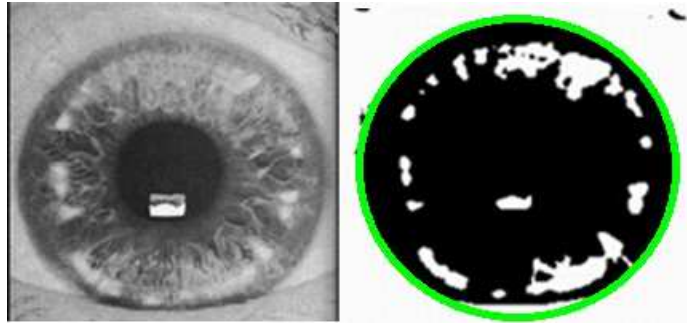


Fig. 10. Iris detectado con la transformada circular de Hough

2.3. Obtención del iris segmentado

Finalmente, habiendo obtenido los centros y radios que representan la pupila y el iris, se crea una máscara con estos datos y los métodos `bitwise_not` y `bitwise_and` de la clase `Core` de `OpenCV` para obtener el iris totalmente segmentado tal como se muestra en la Fig. 11.

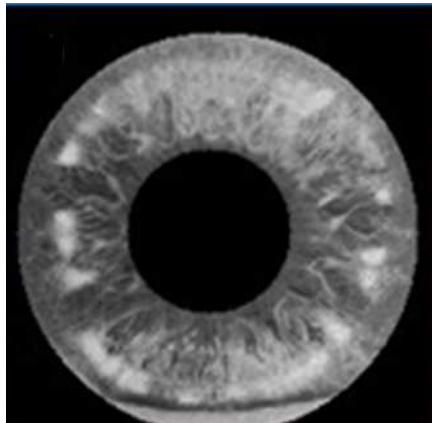


Fig. 11. Iris Segmentado

3. Resultados

El algoritmo propuesto se evaluó tanto en un smartphone como en un emulador, ver Fig. 12, y las características técnicas de ambos se pueden observar en la tabla 1. Para realizar las pruebas se usaron 20 imágenes de la base de datos CASIAIRISV3-Interval

[21]. Además, también se usaron 25 imágenes de dos tipos de lesiones: rosario linfático y anillo de colesterol, estas imágenes se obtuvieron de sitios web dedicados a la Iridología [15],[16],[17],[18],[19]. Los resultados de las pruebas en el emulador y el celular mostraron algunas diferencias con las mismas imágenes. A pesar de las diferencias imprevistas, los resultados alcanzados en la segmentación del iris fueron satisfactorios en la mayoría de los casos, el criterio para considerar una segmentación satisfactoria ha sido si se obtiene al menos 90% del iris y a lo más el 10% de otras partes, en caso contrario se consideró como segmentación no satisfactoria, este criterio se comprobó visualmente. Algunos ejemplos de casos considerados como segmentación satisfactoria se pueden observar en la Fig. 13 y de casos no satisfactorios en la Fig. 14.

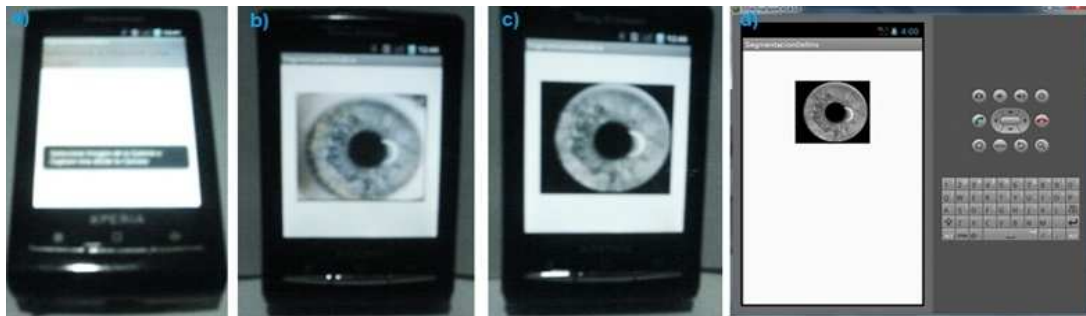


Fig. 12. App para segmentar el iris funcionando en dispositivo móvil y emulador

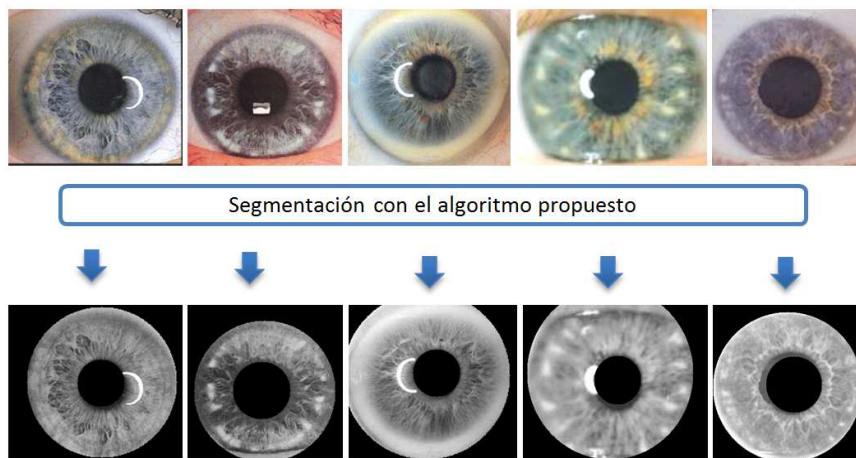


Fig. 13. Ejemplos de segmentación satisfactoria

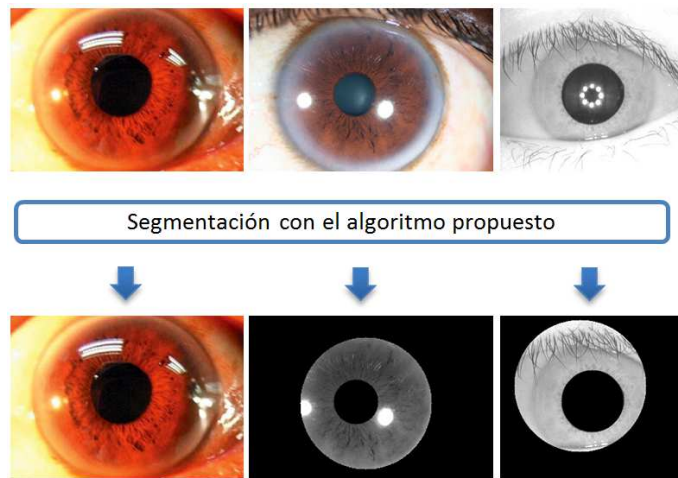


Fig. 14. Ejemplos de segmentación no satisfactoria

| | Smartphone | Emulador |
|---------------------------|--------------------------|---------------------|
| Versión de Android | 2.1 | 4.1.2 |
| Procesador | 600 MHz Qualcomm MSM7227 | ARM (armeabi - v7a) |
| Resolución | 240x320 pixeles (QVGA) | WVGA800 |
| Memoria interna | 120 MB | 256 MB |
| Memoria externa | 2GB | 2GB |

Tabla 1.- Características del Smartphone y emulador usados para evaluar el algoritmo.

El porcentaje de aciertos obtenidos en las pruebas que se ejecutaron en el celular, así como, en el emulador, se muestran en la tabla 2.

| | Smartphone | Emulador |
|--|------------|----------|
| Detección de la pupila | 84.4 % | 95.5 % |
| Detección del borde exterior del iris | 80.0 % | 88.8 % |
| Segmentación completa | 75.5 % | 84.4 % |

Tabla 2.- Tasa de aciertos en la segmentación del iris

4. Discusión

Se detectó que las diferencias en los resultados del algoritmo obtenidos en el emulador y el smartphone aparecían desde el momento en que se obtenía el histograma de la imagen, afectándose de manera inmediata las demás etapas del proceso de segmentación. El problema probablemente se debió a la forma de representación interna de la imagen.

Como se puede apreciar, los porcentajes de aciertos más altos se consiguieron en el emulador. Cabe mencionar que los resultados que se esperaban obtener como salida de los algoritmos en cada etapa del proceso de segmentación fueron los obtenidos en este.

Por otra parte, se puede notar que la detección de la pupila tiene una tasa de acierto más alta a la de la detección del borde exterior del iris, esto se debe principalmente a que la pupila generalmente se encuentra bien definida mientras que por el contrario el borde exterior del iris puede confundirse con la esclera, especialmente en lesiones de colesterol ya que esta se presenta como un anillo en tonos muy claros.

5. Conclusiones

En este trabajo se ha presentado un algoritmo eficaz para la segmentación de iris, el cual se ha implementado aprovechando los diversos métodos de procesamiento de imágenes de OpenCV.

Por otra parte, se experimentó la posibilidad de realizar procesamiento de imágenes oculares digitales en dispositivos móviles con sistema operativo Android, demostrando de esta manera, que esto es factible en cuanto a las capacidades técnicas de estos dispositivos, que día a día rápidamente se incrementan.

Finalmente, como trabajo a futuro, se pretende utilizar el algoritmo de segmentación aquí presentado como etapa inicial en el desarrollo del sistema completo de diagnóstico basado en Iridología y que este a su vez se implemente como App, con el objetivo de

que cualquier persona que cuente con un dispositivo con este sistema operativo pueda emplearlo.

6. Referencias

- [1] B. Jensen, *Iridology Simplified: An Introduction to the Science of Iridology and its Relation to Nutrition*. 1980. Iridologist International. California.
- [2] D. W. Morrison, *How We Heal: Understanding the Mind-Body-Spirit Connection*. 2006. North Atlantic Books. California.
- [3] L. Yu, K. Wang, D. Zhang, "Extracting the autonomic nerve wreath of iris based on an improved snake approach". *Neurocomputing*. Vol. 70. 2007. 743-748 pp.
- [4] R. A. Ramlee, S. Ranjit, "Using iris recognition algorithm, detecting cholesterol presence". *International Conference on Information Management and Engineering*. Vol. 61. 2009. 714-717 pp.
- [5] J. Wang, L. Ma, K. Wang, N. Li, "Radii solaris extraction through primitive modelling" *ICBM*. Vol. 2. 2010. 94-103 pp.
- [6] L. Ma, K. Wang, D. Zhang, "A universal texture segmentation and representation scheme based on ant colony optimization for iris image processing". *Computers and Mathematics with Applications*. Vol. 57. 2009. 1862-1868 pp.
- [7] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*. 2007. Prentice Hall, New Jersey.
- [8] W. Hong-Lin, L. Zhi-Cheng, Q. Jian-Ping, L. Bao-Sheng, "Non-ideal iris segmentation using anisotropic diffusion". *IET Image Process*. Vol. 7. 2013. 111-120 pp.

- [9] J. Zuo, N. A. Schmid, "On a methodology for robust segmentation of nonideal iris images". *IEEE Transactions on Systems, Man, and Cybernetics - part B: cybernetics*. Vol. 40. 2010. 703-718 pp.
- [10] S. Chawla, A. Oberoi, "A robust segmentation method for iris recognition". *International Journal of Advanced Research in Computer Science*. Vol. 2. 2011. 340-343 pp.
- [11] S. Shah, A. Ross, "Iris segmentation using geodesic active contours". *IEEE Transactions on Information Forensics and Security*. Vol. 4. 2009.
- [12] *Digital Image Processing: PIKS Scientific inside*. John Wiley & Sons, New Jersey. 2007.
- [13] C. Diaz, Y. Torres, O. Torres, "Extracción digital de características biométricas basadas en patrones del iris humano". *Revista Colombiana de Física*. Vol. 43. 2011. 244-249 pp.
- [14] *Learning OpenCV: Computer Vision with the OpenCV*. O'Reilly Media, Inc. Sebastopol. 2008.
- [15] Iridologist. Disponible en: <http://www.betterhealththruresearch.com/Iridologist.htm>
Consultada en: Julio 2013.
- [16] Net irisdiagnosis. Disponible en: <http://www.bionordic.dk/page.asp?id=374&sprog=se>. Consultada en: Julio 2013.
- [17] Iridología. Disponible en: <http://canaelsalvador.wordpress.com/category/saludnatural/iridiologia/>. Consultada en: Julio 2013.
- [18] Curso de iridología digital. Disponible en: <http://saludbio.com/curso-de-iridolog%C3%ADa/curso-de-iridolog%C3%ADa>. Consultada en: Julio 2013.

- [19] Iridología. Disponible en: <http://iridologiaprofesorrodriguez.blogspot.mx/>. Consultada en: Julio 2013.
- [20] L. E. Sucar, G. Gómez, *Visión Computacional*. Instituto Nacional de Astrofísica, Óptica y Electrónica. Puebla, México.
- [21] CASIAIRISV3-Interval, Biometric Ideal Test. Disponible en: <http://biometrics.idealtest.org/findDownloadDbByMode.do?mode=Iris>. Consultada en: Julio 2013.

7. Autores

Oscar Fuentes Salome egresado de la Ingeniería en Ciencias de la Computación en la Benemérita Universidad Autónoma de Puebla. Áreas de interés: procesamiento digital de imágenes, reconocimiento de patrones y aprendizaje automático.

Dr. Aldrin Barreto Flores Estudió en el Instituto Tecnológico de Veracruz obteniendo el grado de Ingeniero en Electrónica en el año de 1998. Realizó estudios de Maestría en Ciencias de la Electrónica en el Instituto Nacional de Astrofísica, Óptica y Electrónica en el año 2000, posteriormente en la misma Institución obtuvo el grado de Doctor en Ciencias en el área de Ciencias de la Computación en el año 2005. Sus áreas de interés son visión por computadora, sistemas embebidos y aplicaciones en la industria.

M.C. Verónica Edith Bautista López estudió en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla, obteniendo el grado de Lic. en Ciencias de la computación en 2001, con la especialidad de programación de sistemas. En 2007 obtuvo el grado de Maestra en Ciencias de la Computación, con especialidad en sistemas distribuidos en la misma Institución. Actualmente es candidata a Dra. en Educación de las Ciencias, Ingenierías y Tecnologías por la Universidad de las Américas, Puebla y sus áreas de interés son el desarrollo de aplicaciones móviles y criptográficas, sistemas embebidos, así como la evaluación del aprendizaje e innovación educativa.

Detección de puntos salientes para medición de la simetría facial

Ricardo Irving Barrón Martínez

Universidad Autónoma de San Luis Potosí,
Av. Salvador Nava Martínez s/n, C.P.78290, San Luis Potosí, Teléfono: (444) 826-2316
ricardo_9216@live.com.mx

Ruth Mariela Aguilar Ponce

Universidad Autónoma de San Luis Potosí,
Av. Salvador Nava Martínez s/n, C.P.78290, San Luis Potosí, Teléfono: (444) 826-2316

José Luis TecpanecatI Xihuitl

Universidad Autónoma de San Luis Potosí,
Av. Salvador Nava Martínez s/n, C.P.78290, San Luis Potosí, Teléfono: (444) 826-2316

Resumen

El análisis y estudio del rostro se ha convertido recientemente en un gran foco de interés debido a sus múltiples aplicaciones entre las que se encuentra identificación de expresiones faciales, identificación de personas de interés, y el cálculo de la simetría facial entre otras. Este trabajo presenta la detección de 14 puntos característicos faciales. Estos puntos característicos nos ayudan a obtener la simetría facial siguiendo el método de Mondragón. Para su detección se utilizaron múltiples algoritmos. Los detectores de esquinas se utilizaron para localizar las comisuras de los ojos, por lo que se hizo una comparación de los tres algoritmos más populares en la literatura correspondientes a SUSAN, Harris y Shi-Tomasi. Se determinó que el detector Shi-Tomasi obtiene mejores resultados para esta aplicación en particular. Las comisuras de los labios se localizaron por medio de la segmentación utilizando el algoritmo K-medias

y la detección de bordes. Mientras que el detector de bordes de Canny fue utilizado para localizar los puntos que delimitan el ancho facial.

Palabras Clave: detección de esquinas, procesamiento digital de imágenes, puntos característicos faciales, segmentación.

1. Introducción

El estudio del rostro como método de identificación ha sido ampliamente utilizado involucrando la detección de puntos característicos o salientes. Un punto saliente en el rostro es aquel que permanece estable ante cambios globales o locales debido a cambios en la expresión facial. Adicionalmente, la posición de estos puntos está claramente definida y puede ser establecida por un experto en antropometría. [1].

Entre los puntos salientes más importantes en el rostro se encuentran las comisuras de los ojos, las fosas nasales y la punta de la nariz entre otros. Estos puntos característicos nos ayudan a determinar la expresión facial que presenta un rostro, así como a establecer la simetría facial.

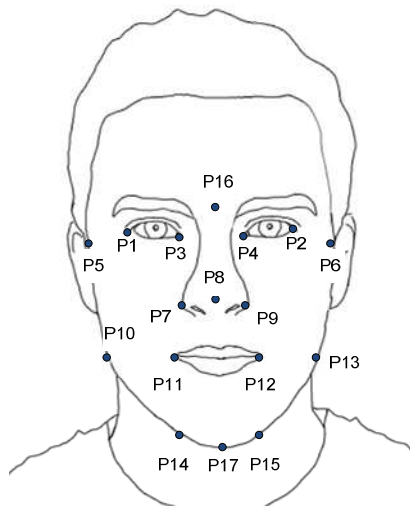


Fig. 1. Puntos característicos faciales para el método de Simmons.

En particular, la simetría facial basada en la antropometría ha sido propuesta como un método de seguimiento en la evolución de pacientes de parálisis facial periférica [2]. Existen diferentes métodos para obtener la simetría facial. Usando los puntos característicos faciales se toman medidas del rostro y se comparan para calcular la simetría. Las medidas y los puntos usados varían de acuerdo al método utilizado. Algunos de los métodos analizados incluyen los métodos propuestos por Mondragón, et. al. [2] y Simmons et. al. [3].

El método de Simmons mide la simetría encontrando una línea media facial compuesta por el punto medio de la boca y los puntos medios entre (P1, P2), (P3, P4), (P5, P6), (P7, P9), (P10, P13), (P11, P12) y (P14, P15) los cuales se muestran en la Fig. 1. Si la cara es simétrica, entonces cada punto medio debe estar ubicado en la línea media determinada.

El método propuesto por Mondragón et. al., determina la línea media facial que divide al rostro en dos. A continuación se miden las distancias desde diferentes puntos característicos faciales a la línea media sagital para ambos lados de la cara. La distancia bipupilar se determina del punto medio de cada ojo hacia la línea sagital y se denomina AP. La distancia de las comisuras de la boca a la línea media es denominada AN. La distancia de las ventanas alares a la línea sagital es AC. El ancho facial vertical o Afv, se mide desde la glabella hasta el subnasal y del subnasal al gnation. Finalmente el ancho horizontal del rostro o Afh, se mide de los puntos P5 a P6. Estas distancias están representadas en la Fig. 2. Posteriormente se encuentra el máximo y mínimo de las distancias AP, AN, AC, Afh y Afv. La Simetría Proporcional Radial (AsPrR) es entonces calculada con la siguiente ecuación:

$$\frac{AP_{min}}{AP_{max}} + \frac{Afh_{min}}{Afh_{max}} + \frac{AN_{min}}{AN_{max}} + \frac{AC_{min}}{AC_{max}} + \frac{Afv_{min}}{Afv_{max}} = AsPrR \quad (1)$$

Un rostro perfectamente simétrico alcanza un valor de 5. Entre menor sea el resultado, mayor será la asimetría. Este trabajo localiza los puntos característicos faciales

correspondientes a P1, P2, P3, P4, P5, P6, P7, P8, P9, P11, P12, P16 y P17 de la Fig. 1. Usando estos puntos es posible determinar la simetría proporcional radial propuesta por Mondragón.

En seguida se describen los métodos utilizados para el desarrollo de la detección de estos puntos y la línea media sagital del rostro. Posteriormente, se introducen los resultados alcanzados en esta implementación y finalmente se presentan las conclusiones del trabajo.

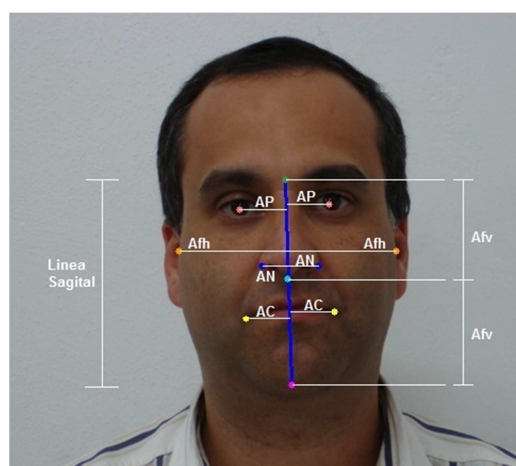


Fig. 2. Puntos característicos faciales para el método de Mondragón.

2. Desarrollo

Los puntos salientes del rostro tienen diferentes características por lo que no es posible utilizar un único algoritmo para todos ellos. Las características antropométricas del rostro nos ayudan a la localización de cada punto. Por lo que es necesario localizar la región del rostro y posteriormente localizar las áreas correspondientes a los ojos, nariz, cejas y boca para su análisis.

Para la localización del rostro se utilizó el algoritmo de Viola-Jones. Este algoritmo es usado por su alto porcentaje de éxito y velocidad de procesamiento [4]. Además no

solamente se utiliza para la detección del rostro, sino también para localizar el área de los ojos, nariz, cejas y boca. El algoritmo se divide en las siguientes etapas: Obtención de la imagen integral, aplicación de las ventanas de Haar y finalmente la cascada de clasificadores.

La imagen integral nos ayuda a calcular las ventanas de Haar con un menor número de operaciones. La imagen integral $I'(x,y)$ de la imagen original $I(x,y)$ se calcula mediante la siguiente ecuación:

$$I'(x,y) = \sum_{x_0 \leq x, y_0 \leq y} I(x_0, y_0) \quad (2)$$

En base a esta imagen se utilizan un conjunto de características rectangulares simples como las mostradas en la Fig. 3.

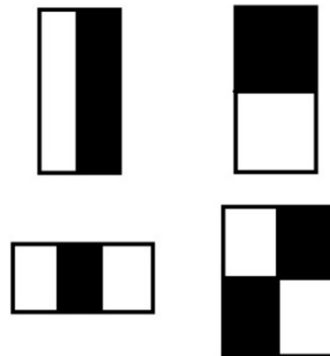


Fig. 3. Características rectangulares simples.

Para evaluar una característica rectangular simple, se colocan los rectángulos descritos en la imagen como se ilustra en la Fig. 4. Todos los valores ubicados en los rectángulos blancos son sumados mientras que los valores localizados en los rectángulos negros son restados.



Fig. 4. Aplicación de características rectangulares.

La evaluación de la característica rectangular simple nos indica si existe un objeto de interés. Este valor es alimentado a una cascada de clasificadores simples usando el algoritmo de Adaboost. La imagen es recorrida iterativamente moviendo en cada ciclo una sub-ventana. Cada sub-ventana es alimentada a la cascada de clasificadores como se muestra en la Fig. 5.

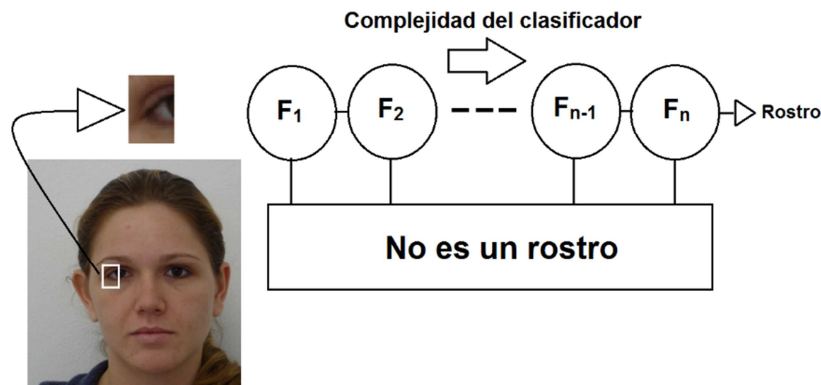


Fig. 5. Ejemplo de cascada de clasificadores.

El algoritmo localiza el objeto de interés si la sub-ventana aprobó toda la cascada de clasificadores. La cascada de clasificadores cambia para cada región que se desea localizar. Por lo que se usan cuatro cascadas de clasificadores correspondientes al rostro, el área de los ojos, la nariz y la boca. Con las áreas de interés delimitadas, se puede proceder a localizar los puntos de interés faciales.

2.1 Comisuras de los ojos

Las comisuras de los ojos pueden ser descritas como esquinas. Por lo tanto su búsqueda se realizó mediante un detector de esquinas. Para seleccionar el más adecuado, se compararon los tres algoritmos de detección de esquinas más populares en la literatura, los cuales son SUSAN, detector Harris y el detector Shi-Tomasi. A continuación se describen brevemente dichos métodos.

El algoritmo de detección de esquinas SUSAN, asocia cada pixel a otro con niveles de brillo similar en una región llamada Univalve Segment Assimilating Nucleus (USAN) [6]. Este algoritmo no requiere derivadas, por lo cual no es necesario hacer una reducción de ruido previo al procesamiento. La detección de esquinas está basada en la minimización de USAN, por ello toma el nombre de Smallest Univalve Segment Assimilating Nucleus (SUSAN). Una máscara circular es utilizada en cada pixel de la imagen. La intensidad de cada pixel dentro de esta máscara es comparada con la intensidad del pixel central. Al final todos los resultados son sumados para obtener el área de USAN.

$$n(x_0, y_0) = \sum_{x,y} \exp\left(-\left[\frac{I(x,y) - I(x_0, y_0)}{t}\right]^6\right) \quad (3)$$

Donde $I(x, y)$ corresponde a los pixeles dentro de la máscara, (x_0, y_0) es el pixel bajo consideración, t corresponde al umbral del ruido y $n(x_0, y_0)$ es el área de USAN.

$$R(x_0, y_0) = \begin{cases} g - n(x_0, y_0) & \text{if } n(x_0, y_0) < g \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Una vez obtenido el valor $n(x_0, y_0)$ se compara con un umbral geométrico fijo g que está en el intervalo de (0,1) y se obtiene la respuesta inicial $R(x_0, y_0)$. El valor $R(x_0, y_0)$ debe ser menor al 50% del área de la ventana para ser considerado borde, mientras que si es una esquina este valor debe ser menor al 25% del área de la ventana. Los

falsos positivos son descartados mediante la localización del centroide, dado que una esquina debe tener el centroide alejado del núcleo de la máscara.

El detector Harris está basado en el detector de Moravec [6]. Harris implementa mejoras al cambiar la ventana rectangular por una ventana Gaussiana. Además considera todos los posibles movimientos de dicha ventana. Estas mejoras son descritas en la siguiente ecuación:

$$S(x, y) \approx \sum_u \sum_v w(u, v) (I_x(u, v)x + I_y(u, v)y)^2 \quad (5)$$

Donde I_x es la imagen diferencial en x , mientras I_y es la imagen diferencial en y . La ecuación puede ser escrita en forma matricial de la siguiente manera:

$$S(x, y) \approx \begin{pmatrix} x & y \end{pmatrix} M \begin{pmatrix} x \\ y \end{pmatrix} \quad (6)$$

La matriz M tiene 2 eigenvalores para un punto de interés. Si ambos eigenvalores son 0 el píxel no es un punto de interés. Si uno de los eigenvalores es positivo, entonces es un píxel borde. Mientras que si ambos eigenvalores son positivos el píxel es una esquina.

El Detector Shi-Tomasi se basa en el Detector Harris y modifica el cálculo de los eigenvalores [7]. Estos se obtienen de la siguiente manera en el detector de Harris:

$$\lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 = \det(M) - k[\text{trace}(A)]^2 \quad (7)$$

Donde k es un parámetro para la sensibilidad modificable en el rango de 0.04 a 0.15. En Shi-Tomasi una esquina es detectada si el mínimo de los dos eigenvalores es mayor que un umbral predefinido:

$$\min(\lambda_1, \lambda_2) > \lambda \quad (8)$$

Para examinar la precisión de los algoritmos, se usó la base de imágenes del Centro Universitario Da FEI [8]. Las imágenes cuentan con una resolución de 640X480 píxeles

representadas en el espacio de color RGB. En todas se usa un fondo blanco homogéneo. Las personas fotografiadas tienen entre 19 y 40 años, las cuales tienen características faciales diferentes. Los algoritmos fueron implementados en C++ utilizando OpenCV [8].

Se emplearon 100 personas para tomar 100 recortes del área de los ojos derechos. Usando Matlab se identificaron manualmente las coordenadas de los puntos característicos faciales. Posteriormente se compararon con las coordenadas resultantes de los algoritmos. Las Fig. 6, 7, y 8 muestran el error promedio en píxeles de los detectores de esquinas con relación a la variación de los parámetros de cada detector.

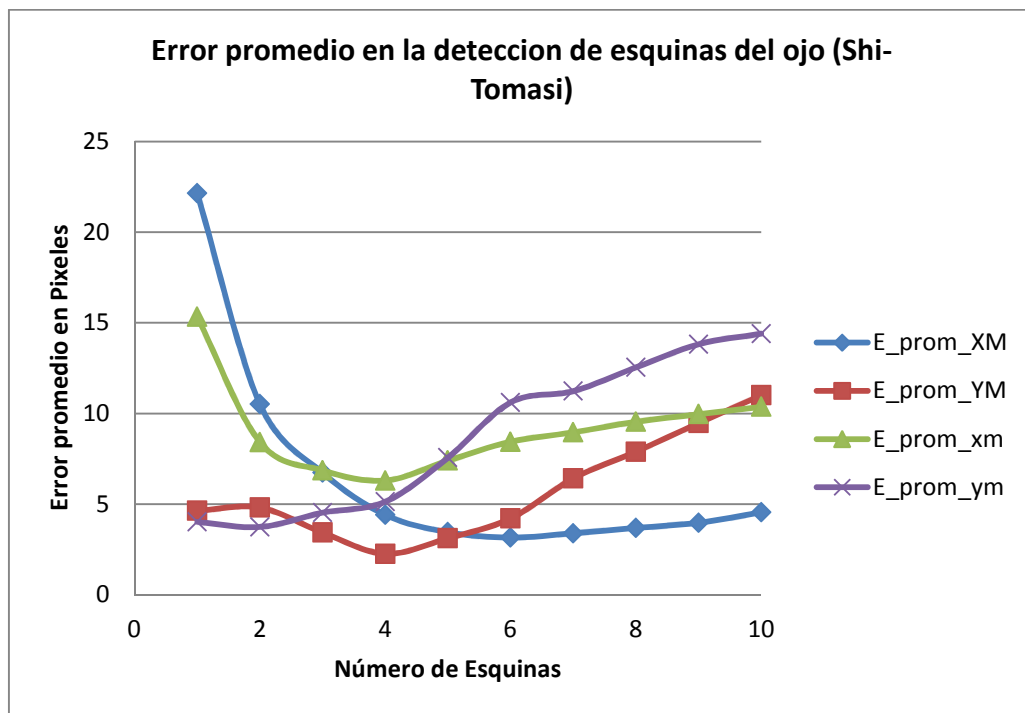


Fig. 6. Resultados de Shi-Tomasi para el ojo derecho.

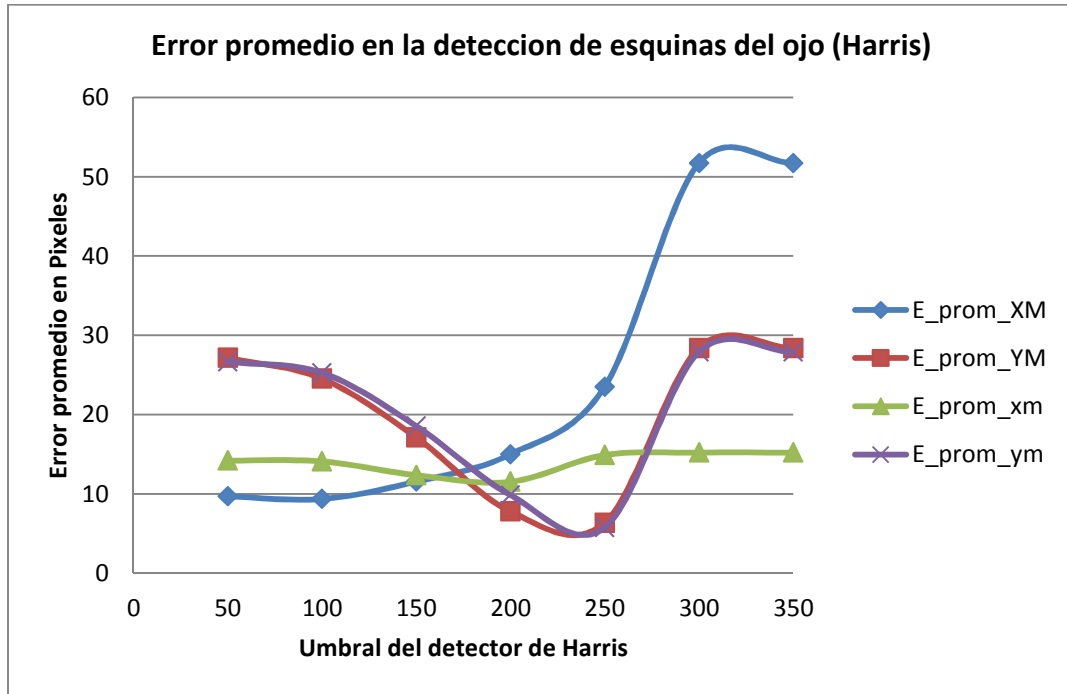


Fig. 7. Resultados de Harris para los ojos.

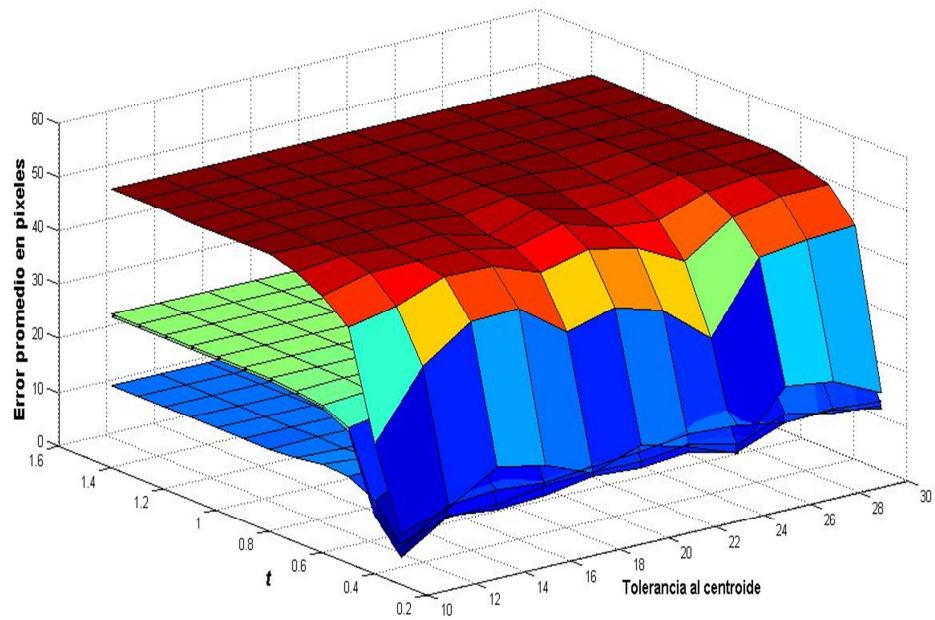


Fig. 8. Resultados de SUSAN para los ojos.

2.2 Comisuras de la Boca

Las comisuras de la boca se localizaron con la detección de bordes debido a que la línea que divide a los labios es un borde muy bien delimitado. La detección de bordes se hace mediante el operador Sobel [10]. Este es un operador diferencial discreto que calcula una aproximación del gradiente de una función. Los bordes son pixeles donde la derivada es mayor que un umbral predefinido. La derivada horizontal se calcula mediante la convolución de la imagen I con la siguiente máscara:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (9)$$

Debido a que las derivadas son susceptibles al ruido, las imágenes deben ser suavizadas mediante un filtro pasabajas para obtener mejores resultados. Para localizar las comisuras de los labios, se aplica el detector de Sobel en el área de la boca como se ilustra en la Fig. 9. El primer y último punto de la línea resultante corresponde a las comisuras de la boca.

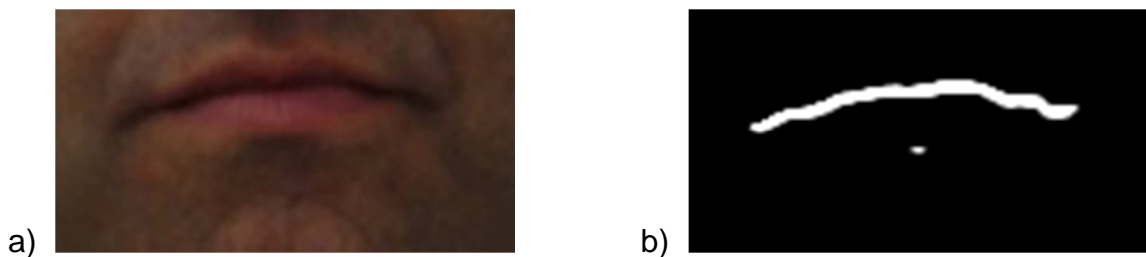


Fig. 9. a) Imagen de la región de la boca. b) Bordes localizados por el detector de Sobel.

2.3 Punta de la nariz, fosas nasales y punto subnasal

El primer paso para encontrar los puntos de la nariz es segmentar su área para diferenciar la zona más brillante y las más oscuras. La primera contiene la punta de la nariz, mientras que las últimas son las fosas nasales. Existen muchos métodos y algoritmos para la segmentación de imágenes. El algoritmo de K-medias se destaca por ser uno de los más simples y ampliamente usado [11].

Considere n puntos $x_i, i = 1 \dots n$ que tienen que ser asignados a k grupos; el objetivo es asociar cada punto a un grupo. K-medias busca las posiciones $\mu_i, i = 1 \dots k$ de los grupos, que minimizan la distancia a los puntos. El agrupamiento por K-medias se puede representar con la siguiente ecuación:

$$\arg \min_c \sum_{i=1}^k \sum_{x \in c_i} d(x, \mu_i) = \arg \min_c \sum_{i=1}^k \sum_{x \in c_i} \|x - \mu_i\|_2^2 \quad (10)$$

Donde c_i es el conjunto de puntos que pertenecen al grupo i . El algoritmo establece el número k de grupos en los que se clasificarán los n puntos. Para entonces inicializar aleatoriamente el centro μ_i de los grupos. Los puntos se asocian al grupo que este más cercano a ellos:

$$c_i = \{j: d(x_j, \mu_i) \leq d(x_j, \mu_l), l \neq i, j = 1, \dots, n\} \quad (11)$$

Posteriormente, todos los centros μ_i son calculados nuevamente. Los nuevos centros corresponden al promedio de todos los puntos de cada grupo:

$$\mu_i = \frac{1}{|c_i|} \sum_{j \in c_i} x_j, \forall i \quad (12)$$

La asociación de puntos y el cálculo de los nuevos centros son realizados de manera iterativa. El ciclo se detiene, hasta que ya no exista diferencia entre el centro μ_i actual y el anterior [11].

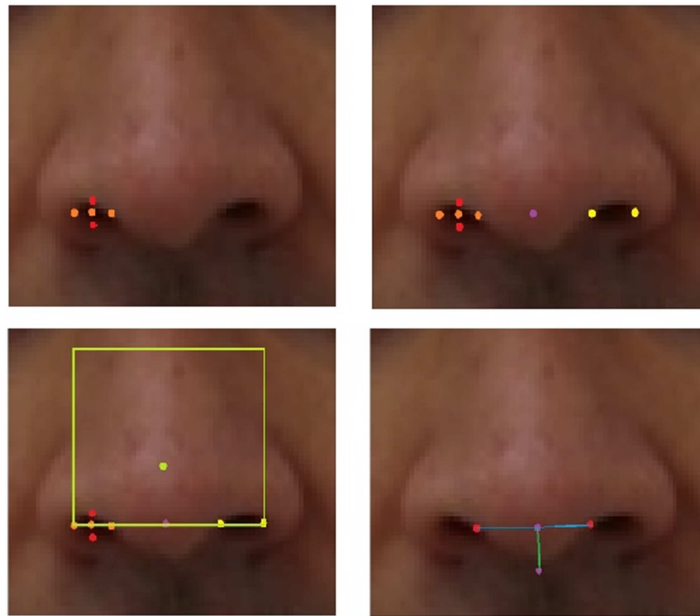


Fig. 10. a) Detección de la primer fosa nasal. b) Detección de la segunda fosa nasal y cálculo de punto medio. c) Localización de la punta de la nariz. d) Ubicación del punto subnasal.

Una vez localizadas las zonas más oscuras, éstas nos ayudan a determinar la altura de las fosas nasales así como el ancho. El punto medio de las fosas nasales se utiliza como posición vertical de la punta de la nariz. Mientras que la coordenada horizontal de la punta de la nariz corresponde al punto más brillante del área de la misma. Finalmente el punto subnasal tiene la misma posición horizontal que la punta de la nariz (ver Fig. 10). La posición vertical del punto subnasal se localiza con la distancia de las fosas nasales al punto medio entre estas.

2.4 Ancho facial

Los puntos que delimitan el ancho facial son localizados ubicando los bordes de la cara. Primeramente se emplea un pre-procesamiento sobre el área del rostro para destacar los bordes del rostro. Para determinar el ancho facial, el detector de bordes de Canny es utilizado [12]. Esto debido a que se considera óptimo por tres características: su tasa

de error baja, su buena localización de bordes y la capacidad de disminuir los falsos positivos.

El detector aplica un filtro gaussiano para reducir el ruido en la imagen. Posteriormente, obtiene las derivadas horizontal y vertical G_x y G_y , con las que estima la aproximación de la magnitud del gradiente y su dirección como se muestra en la siguiente ecuación:

$$G = \sqrt{G_x^2 + G_y^2}, \quad \theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (13)$$

Si el gradiente de un pixel es mayor que un umbral superior predefinido, el pixel es aceptado como borde. En caso de que el gradiente esté por debajo de un umbral inferior predefinido, se rechaza. Finalmente si el gradiente está entre los dos umbrales, el pixel es aceptado si y solo si está conectado a un borde. Por consiguiente se evitan los falsos positivos.

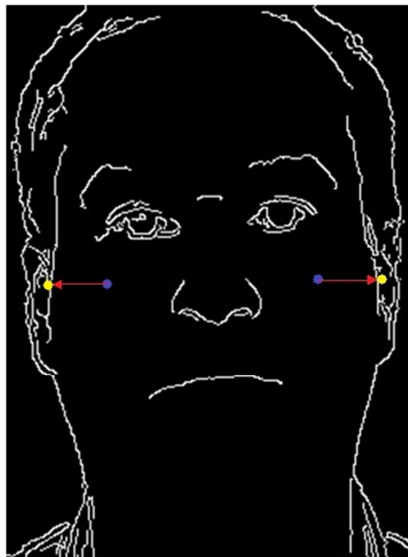


Fig. 11. Recorrido realizado para encontrar los puntos del ancho facial.

Finalmente se recorre la imagen horizontalmente en ambas direcciones (derecha e izquierda). Cuando se encuentre el límite del rostro representado por los bordes, se asignan los puntos del ancho facial. Para hacer más rápido el procedimiento, se inicia el

recorrido en la coordenada horizontal de las comisuras de los ojos y en la coordenada vertical de la punta de la nariz. El procedimiento se ilustra en la Fig. 11.

2.5 Glabela y Gnation

La glabela se define como el punto más prominente de la frente. Resulta difícil localizar este punto de manera directa, pues no se conoce la profundidad de la imagen. Sin embargo, la glabela se encuentra en el punto medio de ambas cejas. Por lo que se detectan las cejas por medio de segmentar el área de las cejas en 2 grupos utilizando K-medias. Uno representa la piel y el otro la ceja. El punto medio entre ambas cejas corresponde a la glabela.

El gnation es el límite inferior de la cara. La antropometría facial indica que la distancia del punto subnasal a la glabela es igual a la distancia ente el punto subnasal y el gnation. Por lo que se midió la distancia vertical del punto subnasal a la glabela. Este valor se sumó a la posición del punto subnasal. El resultado es la posición aproximada del gnation como se ve en la Fig. 12.

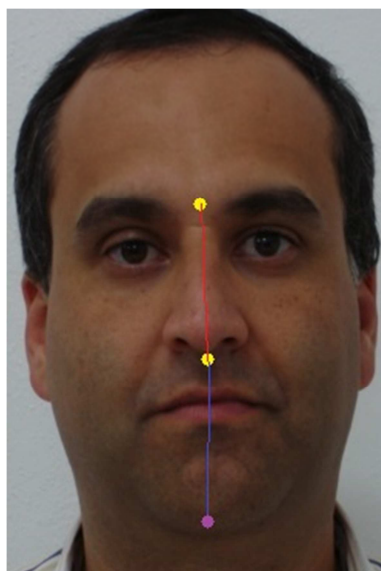


Fig. 12. Aproximación del gnation (punto morado), con la distancia de la glabela al punto subnasal (línea roja).

3. Resultados

Todos los algoritmos y métodos descritos fueron implementados en C++. Para el procesamiento de imágenes se usaron las librerías de OpenCV [8], utilizando un procesador Intel Core i7 4790K con 16GB de RAM DDR 3.

Los resultados mostrados se obtuvieron utilizando 100 imágenes de la base de datos FEI. A este conjunto de imágenes se les aplicó el algoritmo de Viola-Jones. Los porcentajes de éxito en la localización de las regiones de interés se muestran en la Tabla 1.

| Región de interés | Éxito | Fallo | Porcentaje de éxito |
|-------------------|-------|-------|---------------------|
| Rostro | 100 | 0 | 100% |
| Ojo derecho | 100 | 0 | 100% |
| Ojo izquierdo | 100 | 0 | 100% |
| Nariz | 100 | 0 | 100% |
| Boca | 99 | 1 | 99% |
| Total | 499 | 1 | 99.80% |

Tabla 1. Resultados de Viola-Jones.

El porcentaje de éxito alcanzado en nuestra implementación es del 99.80% el cual concuerda con lo reportado por Viola-Jones el cual consiguió un porcentaje de éxito de 92.1% [4].

Para evaluar el desempeño de la detección de los 14 puntos característicos se tomaron 30 imágenes de la base de datos FEI. Se obtuvieron manualmente las coordenadas de los puntos característicos faciales y la distancia entre pupilas utilizando Matlab.

La precisión de la localización de los puntos se obtuvo mediante la diferencia entre las coordenadas de los puntos localizados por el algoritmo y los puntos manualmente localizados utilizando Matlab. Esta diferencia se normaliza con referencia a la distancia inter ocular, la cual corresponde a la distancia entre los centros de las pupilas. Esta medida de precisión fue propuesta por Celiktutan [14].

En la Tabla 2 se muestra el promedio, la mediana y la varianza de la precisión del algoritmo para cada punto característico facial.

| Puntos característicos faciales | Precisión promedio | Mediana de la precisión | Varianza de la precisión |
|--------------------------------------|--------------------|-------------------------|--------------------------|
| Comisura izquierda del ojo izquierdo | 0.03345509 | 0.03115932 | 0.00037537 |
| Comisura derecha del ojo izquierdo | 0.03596308 | 0.03189575 | 0.00049146 |
| Comisura izquierda del ojo derecho | 0.02981611 | 0.02496936 | 0.00044621 |
| Comisura derecha del ojo derecho | 0.03092287 | 0.03030277 | 0.00041508 |
| Fosa nasal izquierda | 0.01179244 | 0.01161423 | 0.00011098 |
| Fosa nasal derecha | 0.03442929 | 0.01907465 | 0.00755722 |
| Punta de la nariz | 0.07209481 | 0.06363589 | 0.00351399 |
| Comisura izquierda de la boca | 0.03703805 | 0.03533928 | 0.00042953 |
| Comisura derecha de la boca | 0.0514752 | 0.04652244 | 0.00120362 |
| Subnasal | 0.04710002 | 0.0367657 | 0.00216887 |
| Gnation | 0.13245924 | 0.12591352 | 0.00485325 |
| Glabela | 0.05515784 | 0.03954127 | 0.00533827 |
| Límite de la cara izquierdo | 0.07296596 | 0.07013022 | 0.00171937 |
| Límite de la cara derecho | 0.07431787 | 0.07353806 | 0.0021793 |

Tabla 2. Resultados de la precisión del algoritmo propuesto.

Las Fig. 13 y Fig. 14 muestran visualmente el desempeño del algoritmo.

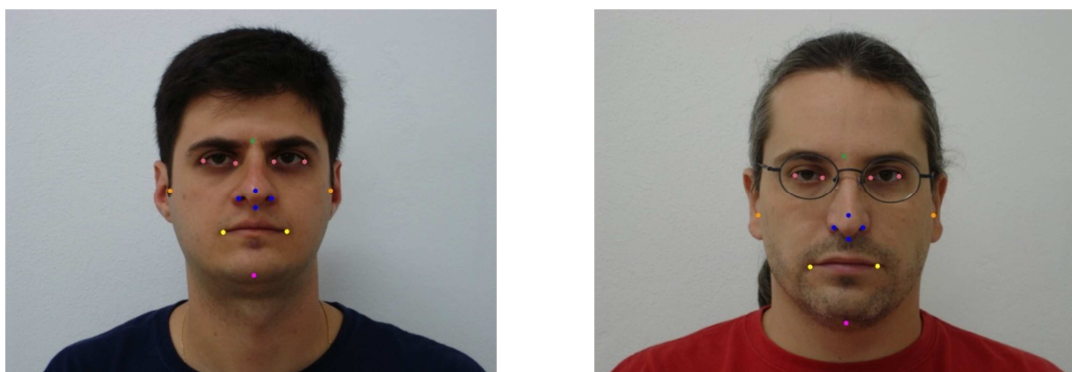


Fig. 13. Localizaciones correctas.

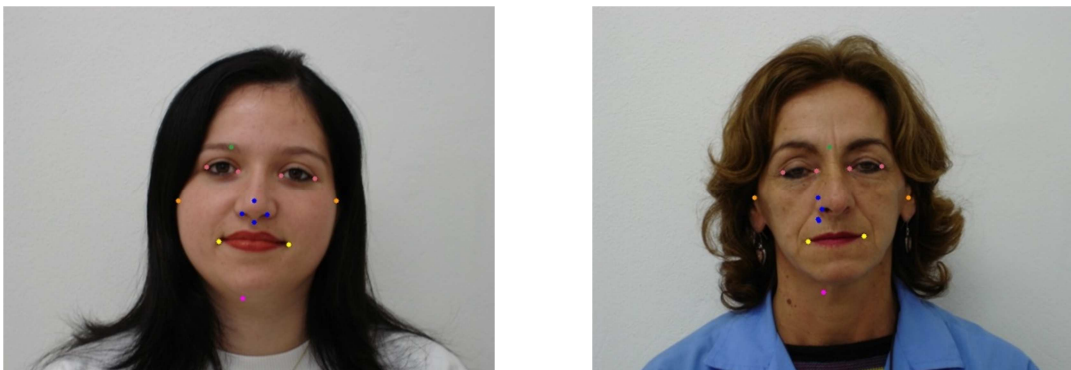


Fig. 14. Errores en las localizaciones.

4. Discusión

Celiktutan establece que una detección exitosa corresponde a una precisión inferior de 0.05 [14]. Mientras que una detección aceptable corresponde a un valor de precisión superior a 0.05 e inferior a 0.1. La detección de las comisuras de los ojos y las fosas nasales alcanzan una clasificación de detección exitosa como se muestra en la Tabla 1. Mientras que el resto de los puntos alcanza una detección aceptable. Los resultados muestran que, la mayoría de los puntos de interés tienen una precisión promedio por debajo de 0.05, a pesar de que las personas tienen vello facial y diferentes accesorios. Sin embargo, el gnation tiene un error en precisión superior a 0.1 debido a que este punto solo es aproximado en base a la distancia del punto subnasal a la glabella. Un problema para la localización de este punto es el vello facial y la papada. Por tanto, la aplicación creada permite la interacción con los puntos detectados para corregir errores que el experto determine que están incorrectos. Las posiciones de los puntos característicos pueden ser corregidas por medio del ajuste manual utilizando dispositivos apuntadores. La interfaz gráfica es mostrada en la Fig. 15.

Esta aplicación funciona exclusivamente con rostros frontales en el plano Frankfurt, debido a que es un requisito para poder establecer la simetría facial correctamente. La rotación o inclinación de la cara provocara errores en la búsqueda. Algunas imágenes de la base de datos FEI presentan estas características. Aun así el algoritmo y la

interfaz gráfica pueden conseguir las posiciones exactas de los puntos de interés faciales.



Fig. 15. Punta de la nariz seleccionada con la interfaz gráfica.

5. Conclusiones

Los puntos salientes en el rostro ayudan a la identificación de personas, así como a determinar la expresión facial del rostro y su simetría. Los puntos salientes son aquellos que no son modificados debido a la expresión facial y la posición de estos puntos está claramente definida y puede ser establecida por un experto en antropometría. En este trabajo los puntos salientes localizados pertenecen a las comisuras de los ojos, comisuras de los labios, ventanas alares, punto subnasal, punta de la nariz, límites del ancho horizontal facial, glabella y gnation. Se compararon los detectores de esquina SUSAN, Harris y Shi-Tomasi, resultando Shi-Tomasi el detector que alcanza el error más pequeño.

La detección de las comisuras de los ojos así como las fosas nasales obtuvieron una precisión exitosa según la clasificación mostrada por Celiktutan [14]. Mientras que el resto de los puntos tiene una precisión aceptable. El único punto que no se logró detectar correctamente es el gnation debido a los problemas que se presentan con el

vello facial y la papada. Sin embargo, este problema es solventado mediante la implementación de una interfaz gráfica que permite la interacción con los puntos detectados para su corrección manual.

6. Referencias

- [1] A. Willis, S. Yunfeng, "An algebraic model for fast corner detection". In IEEE 12th International Conference on Computer Vision. Kyoto. Japan. 2009. 2296-2302 pp.
- [2] G. F. Mondragón, M. P. Espinoza, N. H. Campos, L. S. Chapul, R. P. Pérez, C. M. Canseco, S. R. León, V. A. Monsalvo, J. P. Paz, A. F. Jacinto, "Facial anthropometry: a tool for quantitative evaluation in patients with peripheral facial paralysis". International Journal of Scientific & Engineering Research. Vol. 6. No. 5. 2015. 1657-1660 pp.
- [3] L. W. Simmons, G. Rhodes, M. Peters, N. Koehler. "Are human preferences for facial symmetry focused on signals of developmental instability?". Behavioral Ecology. Vol. 15. 2004. 864-871 pp.
- [4] P. Viola, M. J. Jones, "Robust Real-Time Detection". International Journal of Computer Vision. Vol. 57. No. 2. May 2004. 137- 154 pp.
- [5] L. Zou, J. Chen, J. Zhang, L. Dou, "The Comparison of Two Typical Corner Detection Algorithms". In the Second International Symposium on Intelligent Information Technology Application. Shanghai, China. 2008. 211, 215 pp.
- [6] J. Shi, C. Tomasi, "Good Features to Track". In IEEE 9th Conference on Computer Vision and Pattern Recognition. Seattle. WA. 1994. 8 pp.
- [7] FEI Face Dataset. <http://fei.edu.br/~cet/facedatabase.html>. Agosto 2014.
- [8] Open Source Computer Vision. <http://opencv.org/>. Julio 2014.

- [9] R. C. González, R. E. Woods, *Digital Image Processing*. 3th ed. 2008. Pearson. United States of America. 706-713 pp.
- [10] T. Kanugo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, A.Y. Wu, "An Efficient K-Means Clustering Algorithm: Analysis and Implementation". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 24. No. 7. 2002. 881-892 pp.
- [11] R. C. González, R. E. Woods, *Digital Image Processing*. 3th ed. 2008. Pearson. United States of America. 162-165, 719-725 pp.
- [12] M. Flores Méndez, I. Hernández, G. Rossano, "Estructuración y estandarización de la antropometría facial en función de proporciones". *COSME NEWS*. Vol. 6. Mar. 2004. 10-14 pp.
- [13] O. Celiktutan, S. Ulukaya, B. Sankur, "A Comparative Study of Face Landmarking Techniques". *EURASIP Journal on Image and Video Processing*. 2013. 27 pp.

7. Autores

Ing. Irving R. Barrón-Martínez es Ingeniero Electrónico por la Universidad Autónoma de San Luis Potosí, actualmente se encuentra estudiando la Maestría en Ing. Electrónica en la misma universidad.

Dra. Ruth M. Aguilar-Ponce es Doctora en Ingeniería en Computación por la Universidad de Louisiana en Lafayette.

Dr. J. Luis Tecpanecatli-Xihuitl es Doctor en Ingeniería en Computación por la Universidad de Louisiana en Lafayette.

Virtualización mediante MetaRouter para la implementación de una red wireless de navegación anónima tipo TOR en equipos Mikrotik

Roberto David Meneses Basantes

Universidad de Fuerzas Armadas, Av. Progreso s/n, Quito-Ecuador, Teléfono: 59323989400
rdmeneses@espe.edu.ec

Darwin Leonidas Aguilar Salazar

Universidad de Fuerzas Armadas, Av. Progreso s/n, Quito-Ecuador, Teléfono: 59323989400 ext. 1869
dlaguilar@espe.edu.ec

Christian Nestor Vega Muñoz

Universidad de Fuerzas Armadas, Av. Progreso s/n, Quito-Ecuador, Teléfono: 59323989400 ext. 1872
cnvega@espe.edu.ec

Rita Paola León Pérez

Universidad de Fuerzas Armadas, Av. Progreso s/n, Quito-Ecuador, Teléfono: 59323989400 ext. 1872
rpleon1@espe.edu.ec

Resumen

Hoy en día la virtualización se ha convertido en la primera opción a la hora de considerar servidores y como tal presenta muchas ventajas sobre soluciones propietarias. Pero que tal si un administrador de red pudiera tener ambas soluciones en un dispositivo embebido el cual delegará responsabilidades, tanto a la parte virtual como a la parte física, manejando así una solución mixta.

La marca europea de equipos Mikrotik, mediante el software de sus dispositivos de capa 3 RouterOS, permite desde su versión 3.30, virtualizar mediante MetaRouter, una versión de Linux comprimido, llamado OpenWRT el mismo que proporciona un sistema de archivos totalmente modificable con gestión de paquetes, proporcionando al usuario la posibilidad de personalizar su firmware de acuerdo a su necesidad, lo que lo convierte en un sistema ideal para crear una aplicación sin que el firmware sea el limitante. OpenWRT tiene sus orígenes con la línea Wireless de cisco Lynksys la cual es una plataforma fácil de familiarizarse, con el manejo de servidores Linux basados en Debian [1].

En este documento se detallará la implementación y discutirá los resultados de un servidor de navegación anónima de tipo TOR instalado en OpenWRT y virtualizado sobre un equipo Mikrotik RB450G.

Palabra(s) Clave(s): MetaRouter, Mikrotik, OpenWRT, TOR, virtualización.

1. Introducción

The Onion Router (TOR), es un proyecto que fue lanzado el 20 de Septiembre del año 2002. Se trata de una red de comunicaciones con baja latencia, funcionando y operativa en Internet y cuya característica principal es permitir el envío de mensajes entre usuarios sin revelar su identidad (su dirección IP, anonimato desde un punto de vista de red). La Red TOR es el caso más conocido de red oscura que existe hoy en día. La Red Oscura, o darknet, se suele referir a la parte de Internet que no quiere ser procesada ni encontrada fácilmente y que suele alojar actividades ilegales [2].

La red TOR, es la mayor red de enrutamiento anónimo de tercera generación que está diseñado para realizar el intercambio de direccionamiento, escondiendo la ubicación física del usuario y tiene como uno de sus beneficios la defensa de los ataques por “análisis de tráfico”, que se realiza mediante el análisis de tramas que el host o terminal

envía y recibe desde la red.

A diferencia del enrutamiento común, que establece que se debe conocer su camino completo hasta llegar al destino antes de establecer las comunicaciones, las redes tipo TOR, utilizan para su comunicación, rutas sinuosas difíciles de seguir y que periódicamente cambian usando diferentes caminos, cada vez que se establece una nueva conexión. Los paquetes de datos de la red TOR, siguen caminos aleatorios de tres saltos encriptados, a través de varios repetidores teniendo en cuenta que cada repetidor conoce únicamente al repetidor anterior y al próximo repetidor, mas no la ruta completa. Por eficiencia TOR, utiliza la misma ruta de intercambio de datos, para las conexiones que se producen dentro de un tiempo determinado, para peticiones posteriores se traza una nueva ruta [3].

MIKROTIK es una marca proveniente de Latvia dedicada a fabricar dispositivos de networking y equipos de comunicaciones, cuenta con su sistema operativo RouterOS, que tiene funcionalidades como: manejar direccionamiento IP, Manejo de acceso al medio en capas dos y tres, redes privadas NAT, Firewall, Módulo Wireless, para motivos del presente proyecto una de las mayores virtudes de la marca, es que permite la Virtualización de Sistemas Operativos para Routers mediante códigos fuente abiertos de tipo OpenSource, entre otras ventajas, presenta las condiciones ideales para el desarrollo del presente estudio.

En el presente documento se describe y analiza los resultados de implementar una topología de red que permita navegar de forma segura sobre redes anónimas TOR. Se ha considerado, que los usuarios se encuentren detrás de una red NAT, protegiendo así su identidad, detrás de un direccionamiento privado, considerando además que el MetaRouter ejecutará el servicio TOR, enrutando así todo el tráfico web, por el MetaRouter implementado sobre un dispositivo de capa 3 con módulo Wireless, que permitirá una validación de portal cautivo de tipo Hotspot en un punto de acceso Inalámbrico (Wireless), mediante una instancia virtualizada, aprovechando el recurso de MetaRouter de Mikrotik a nivel de software. En cambio, a nivel de hardware se utilizará

placas (mainboard's) de la serie RB4XXAH con arquitectura MIPSBE, que soportan imágenes OpenWRT MetaRouter y cuentan con las características físicas necesarias para la implementación que se propone y muestra el desarrollo a continuación.

2. Desarrollo

a) Topología de la red

La topología de red considerada en la figura 1, permite la implementación de un servidor de navegación anónima tipo TOR virtualizado mediante la herramienta MetaRouter de Mikrotik, donde se obtendrá una red NAT de tráfico anónimo y tráfico tradicional. Los únicos puertos de transmisión de datos que estarán abiertos a los usuarios finales son 80 TCP, 53 UDP, 8118 TCP y 9050 TCP, el puerto 8118 TCP es el proxy Privoxy que actúa como un proxy HTTP estándar al proxy TOR Socks mientras que el puerto 9050 de TCP, es el proxy Socks disponible para el enrutamiento del tráfico a través de la red TOR.

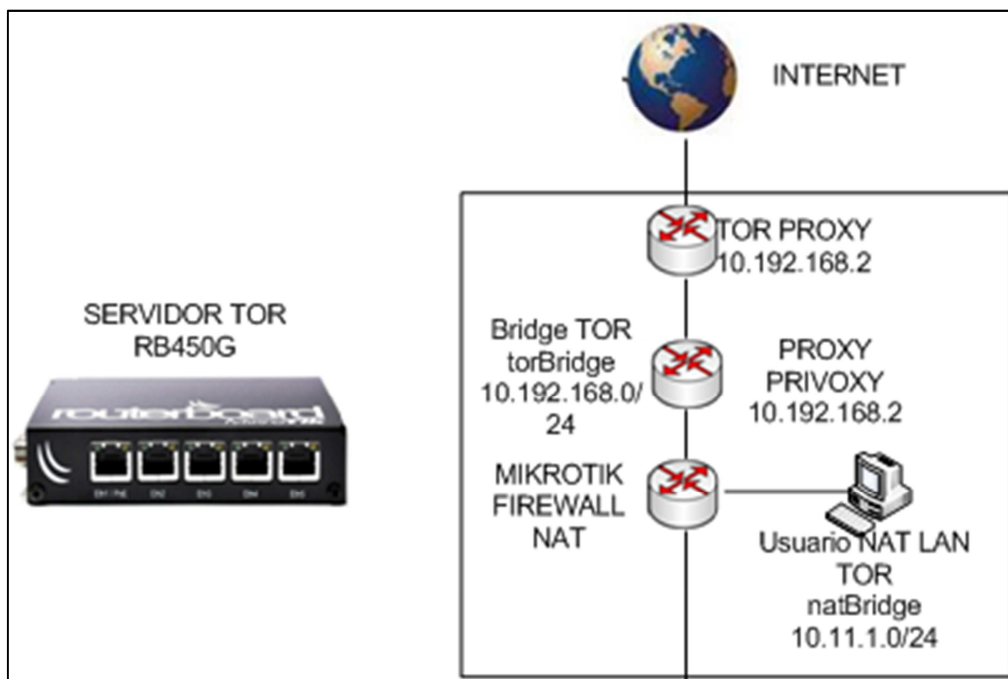


Fig. 1. Topología de la red.

b) Direccionamiento IP.

El enrutamiento considerado para las diferentes interfaces tanto del dispositivo físico como del dispositivo virtual se encuentra descrito en la Tabla 1.

| Dispositivo | Interfaz | Dirección IP | Máscara de subred |
|-------------|-------------|--------------|-------------------|
| RB450G | ether1 | 192.168.1.10 | 255.255.255.192 |
| | natBridge | 10.11.1.1 | 255.255.255.0 |
| | torBridge | 10.192.168.1 | 255.255.255.0 |
| | DHCP server | 10.11.1.0/24 | S/N |
| | DNS Server | 10.11.1.1 | S/N |
| OpenWRT | Ether0 | 10.192.168.2 | 255.255.255.0 |
| | Gateway | 10.192.168.1 | S/N |

Tabla 1. Direccionamiento IP del servidor Mikrotik.

Es importante tener en cuenta que las interfaces de OpenWRT, son virtuales y deben ser modificadas la primera vez desde MetaRouter, como se indicará posteriormente.

c) Configuraciones.

A) Configuración sobre Mikrotik.

1. Ingreso al dispositivo

El ingreso al dispositivo se lo realizará a través del software de administración WinBox.

2. Creación de las interfaces Bridge de comunicación

La secuencia de comandos de la figura 3, configura las interfaces bridges y añaden los puertos necesarios para el bridge encargado de realizar el NAT de la red.

```
> interface bridge
/interface bridge> add name=torBridge
/interface bridge> add name=natBridge
/interface bridge> port
/interface bridge port> add interface=ether4 bridge=natBridge
/interface bridge port> add interface=ether5 bridge=natBridge
/interface bridge port> █
```

Fig. 3. Configuración de interfaces Bridge.

3. Configuración de direccionamiento IP

La configuración de direcciones IP, se deberá realizar en base a la Tabla 1 y el resultado de este deberá proporcionar la información como la mostrada en la Figura 4.

```
> ip address add interface=ether1 address=192.168.1.10/26 disabled=no
> ip address add interface=natBridge address=10.11.1.1/24 disabled=no
> ip address add interface=torBridge address=10.192.168.1/24 disabled=no
> █
```

Fig. 4. Configuración de direccionamiento IP.

4. Configuración DNS

En afán de ocultar toda la información que pueda revelar la identidad del cliente, se utilizará los servidores DNS de google, acorde a lo indicado en la figura 5.

```
> /ip dns set servers=8.8.8.8 allow-remote-requests=yes
> █
```

Fig. 5. Configuración del servidor DNS.

5. Configuración de firewall NAT Rules

Las reglas de Firewall NAT son las encargadas de redireccionar todo el tráfico que ingresa o sale a través de la interfaz WAN (ether 1) hacia el servidor virtual por su dirección IP 10.192.168.2. (Ver Figura 6)

```
> /ip firewall nat
/ip firewall nat> add chain=srcnat out-interface=ether1 action=masquerade
/ip firewall nat> add chain=dstnat in-interface=ether1 action=dst-nat to-addresses=10.192.168.2
/ip firewall nat> █
```

Fig. 6. Configuración de reglas de Firewall.

6. Configuración de la ruta por defecto

La ruta por defecto establece que todo el tráfico de red será enviado al internet a través del gateway del proveedor de internet. (Ver Figura 7)

```
> /ip route add dst-address=0.0.0.0/0 gateway=192.168.1.1
>
```

Fig. 7. Configuración de ruta por defecto.

7. Configuración del servidor DHCP sobre natBridge

El servidor DHCP, se encuentra en el pool 10.11.1.0/24, gateway 10.11.1.1 y DNS 8.8.8.8., esto se muestra en la Figura 8.

```
> /ip pool add name="nat-DHCP" ranges="10.11.1.10-10.11.1.250"
> /ip dhcp-server network add address=10.11.1.0/24 gateway=10.11.1.1 dns-server=8.8.8.8
> /ip dhcp-server add interface="natBridge" lease-time="3d" name="natDHCPserver" address-pool="nat-DHCP" authoritative=yes disabled=no
>
```

Fig. 8. Configuración del servidor DHCP.

B) Configuración sobre MetaRouter.

9. Descarga del firmware OpenWRT

Se puede descargar de manera gratuita el compilado de OpenWRT para dispositivos Mikrotik de arquitectura MIPSBE, así como todos los paquetes instalables del repositorio de Mikrotik <http://openwrt.wk.cz/>. (Ver Figura 9)

Location: /attitude_adjustment/mr-mips/

| Name | Last modified | Size |
|---|-------------------|------|
| Parent Directory | | - |
| packages/ | | - |
| OpenWrt-ImageBuilder-mr-mips-for-linux-i686.tar.bz2 | 11-Mar-2014 13:36 | - |
| OpenWrt-SDK-mr-mips-for-linux-i686-gcc-4.6-linaro_uClibc-0.9.33.2.tar.bz2 | 11-Mar-2014 13:37 | 317M |
| OpenWrt-Toolchain-mr-mips-for-mips-gcc-4.6-linaro_uClibc-0.9.33.2.tar.bz2 | 11-Mar-2014 13:40 | 591M |
| config-mips-aa-36088-20140311 | 11-Mar-2014 13:38 | 40M |
| kernel-debug.tar.bz2 | 11-Mar-2014 13:28 | 162K |
| kernel-debug.tar.bz2 | 11-Mar-2014 13:40 | 5.6M |
| mD5sums | 11-Mar-2014 13:40 | 227 |
| openwrt-mr-mips-rootfs-36088-basic.tar.gz | 08-Mar-2014 22:28 | 2.2M |
| openwrt-mr-mips-rootfs.tar.gz | 11-Mar-2014 13:28 | 2.3M |

Fig. 9. Repositorio de Mikrotik para OpenWRT.

Una vez descargado el firmware OpenWRT Attitude Adjustment 12.09, se deberá importar el archivo .tar.gz a la carpeta files ubicada en el software WinBox. (Ver Figura 10)

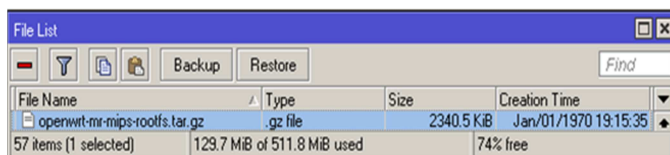


Fig. 10. Importación de la imagen de OpenWRT.

10. Creación del dispositivo virtual

El dispositivo virtual se crea a través de la herramienta MetaRouter importando la imagen descargada y asignando la memoria del dispositivo (se deberá asignar al menos 100MB de memoria para el correcto funcionamiento del servidor de navegación anónima). (Ver Figura 11)

```
admin@MikroTik] > /metarouter import-image memory-size=100 file-name=openwrt-mr-mips-rootfs.tar.gz
imported: 100%
```

Fig. 11. Creación de la máquina virtual.

11. Creación de la interfaz virtual de MetaRouter

La interfaz de OpenWRT es una interfaz de tipo dinámico y se la deberá vincular a la interfaz bridge torBridge (el bridge torBridge es un túnel de conexión directa entre el dispositivo RB450G y el dispositivo virtual), la configuración se muestra en la Figura 12.

```
> /metarouter set 0 name=attitude_tor
> metarouter interface add type=dynamic dynamic-bridge=torBridge virtual-machine=attitude_tor
```

Fig. 12. Configuración de la interfaz de OpenWRT.

12. Configuración de direccionamiento IP sobre OpenWRT

El primer paso que se deberá realizar es la configuración de la dirección IP del dispositivo virtual, de acuerdo a la Tabla 1.

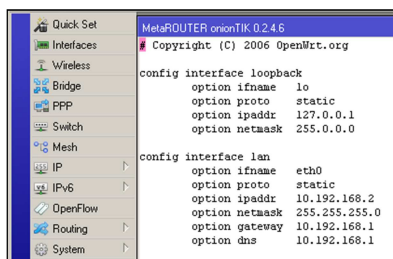


Fig. 13. Configuración de la interfaz ether0.

Para que los cambios tengan efecto se debe reiniciar el servicio de red con el comando “/etc/init.d/network restart”.

13. Habilitación de administración mediante SSH

El comando “passwd” permite habilitar el protocolo de comunicaciones SSH y automáticamente deshabilita TELNET. (Ver figura 14)

```
root@metarouter:/# passwd
Changing password for root
New password:
Bad password: too weak
Retype password:
Password for root changed by root
root@metarouter:/# passwd
Changing password for root
New password:
Bad password: too weak
Retype password:
Password for root changed by root
```

Fig. 14. Configuración de SSH.

C) Configuración sobre OpenWRT.

14. Ingreso al dispositivo virtual

El software mediante el cual se configurará y administrará OpenWRT es el demonio SSH Putty, luego se ejecutará e ingresará a OpenWRT.

15. Configuración de firewall

El compilado de OpenWRT de Mikrotik provee un pequeño Firewall el cual se deberá modificar para que acepte todo tipo de conexiones, como lo muestra la Figura 15 y después reiniciarlo “/etc/init.d/firewall restart”.

-

```
10.192.168.2 - PuTTY
# config defaults
option syn_flood      1
option input          ACCEPT
option output         ACCEPT
option forward        ACCEPT
# Uncomment this line to disable ipv6 rules
option disable_ipv6   1

config zone
option name           lan
option network        'lan'
option input          ACCEPT
option output         ACCEPT
option forward        ACCEPT

config zone
option name           wan
option network        'wan'
option input          ACCEPT
option output         ACCEPT
option forward        ACCEPT
option masq           1
option mtu_fix        1

config forwarding
option src            lan
option dest           wan
```

Fig. 15. Configuración del Firewall.

16. Configuración del archivo de resolución DNS

El servidor DNS deberá ser configurado desde el archivo “resolve.conf” con el siguiente comando “vi /etc/resolve.conf” (Ver Figura 16)

```
10.192.168.2 - PuTTY
# search lan
nameserver 127.0.0.1
nameserver 10.192.168.1
```

Fig. 16. Configuración del DNS de OpenWRT.

17. Configuración del módulo de actualización de librerías opkg

OpenWRT cuenta con el gestor de paquetes OPKG, que deberá ser configurado mediante “vi /etc/opkg.conf” con la dirección web del repositorio de Mikrotik visto anteriormente (Ver Figura 17).

```
10.192.168.2 - PuTTY
# rc/gz packages http://openwrt.wk.cz/attitude_adjustment/mr-mips/packages
dest root /
dest ram /tmp
lists_dir ext /var/opkg-lists
option overlay_root /overlay
```

Fig. 17. Configuración de paquete opkg.

18. Instalación/configuración módulo administración gráfico Luci

La instalación del paquete Luci se la realiza como muestra se muestra en la figura 18.

```
10.192.168.2 - PuTTY
root@Tesis:~# opkg install luci
```

Fig. 18. Instalación del paquete Luci.

Una vez instalado el paquete Luci, se deberá habilitar el paquete web uhttpd (figura 19).

```
root@Tesis:~# /etc/init.d/uhttpd enable
```

Fig. 19. Habilitación del paquete uhttpd.

El ingreso a la interfaz web, se lo realiza con el mismo usuario y clave configuradas para SSH (Figura 20).

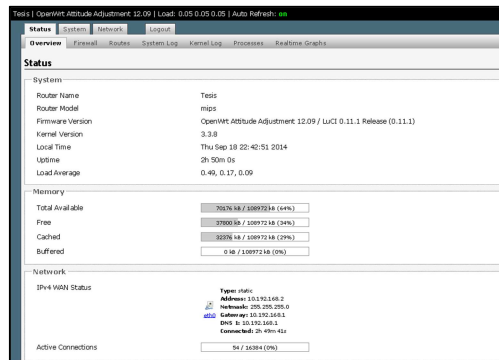


Fig. 20. Ventana del paquete Luci.

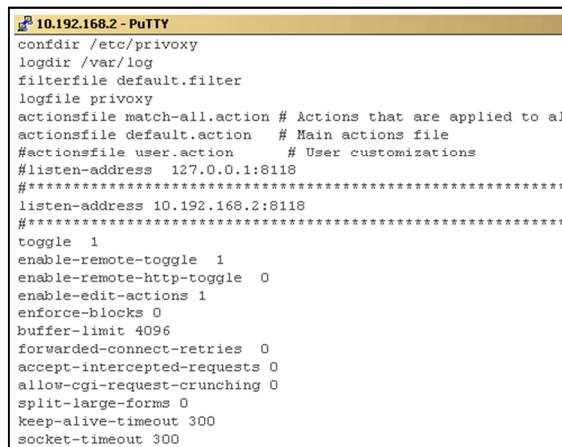
19. Instalación y configuración de Privoxy [4, 5]

El proxy Privoxy se encargará de ser el nexo entre el cliente solicitante de navegación y la red anónima TOR (Figura 21).

```
10.192.168.2 - PuTTY
root@Tesis:~# opkg install privoxy
```

Fig. 21. Instalación de Privoxy.

No existe una configuración estándar de Privoxy por lo cual esta deberá ser configurada de acuerdo a las condiciones de cada red. Privoxy deberá ser configurado para el presente escenario como lo muestra la Figura 22, en el archivo “vi /etc/privoxy/config”.



```
10.192.168.2 - PuTTY
confdir /etc/privoxy
logdir /var/log
filterfile default.filter
logfile privoxy
actionsfile match-all.action # Actions that are applied to al
actionsfile default.action # Main actions file
#actionsfile user.action # User customizations
#listen-address 127.0.0.1:8118
#*****
listen-address 10.192.168.2:8118
#*****
toggle 1
enable-remote-toggle 1
enable-remote-http-toggle 0
enable-edit-actions 1
enforce-blocks 0
buffer-limit 4096
forwarded-connect-retries 0
accept-intercepted-requests 0
allow-cgi-request-crunching 0
split-large-forms 0
keep-alive-timeout 300
socket-timeout 300
```

Fig. 22. Configuración de Privoxy.

Terminada la edición del archivo “config”, se deberá reiniciar el servicio.

20. Instalación y configuración del paquete TOR

Debido a que el repositorio de Mikrotik ofrece varias versiones del paquete de navegaciones anónimas TOR, es recomendable instalar una versión que ya ha sido probada previamente “tor-alpha_0.2.3.22-rc-1_mr-mips.ipk” (Figura 23)



```
10.192.168.2 - PuTTY
root@metarouter:~# opkg install http://openwrt.wk.cz/attitude_adjustment
/mr-mips/packages/tor-alpha_0.2.3.22-rc-1_mr-mips.ipk
```

Fig. 23. Instalación del paquete TOR.

Debido a que la memoria es el principal inconveniente en dispositivos embebidos, se creará una variable flotante que se sobrescriba cada vez que la red TOR, emita cualquier mensaje hacia el servidor (Ver Figura 24).



```
10.192.168.2 - PuTTY
root@Tesis:~# cat /dev/null >> /etc/tor/torrc
root@Tesis:~# cat /dev/null > /etc/tor/torrc
```

Fig. 24. Variable de sobre escritura torrc.

El archivo “vi /etc/tor/torrc” será configurado como lo muestra la figura 25. En este archivo está considerado que se podrá únicamente variar los parámetros de las líneas 9, 10 y 11. El resto de la configuración ha sido adecuada a la topología y

direccionamiento planteados en la Figura 1 y Tabla 1 respectivamente.

```
10.192.168.2 - PuTTY
# Archivo de configuracion y autenticacion
SocksPort 9050
SocksListenAddress 10.192.168.2:9050
SocksPolicy accept 10.11.1.0/24
Log notice file /var/log/tor/notices.log
RunAsDaemon 1
ORPort 9001
DirPort 9030
ExitPolicy reject *:
Nickname TesisTor
RelayBandwidthRate 100 KB
RelayBandwidthBurst 200 KB
DNSPort 53
DNSListenAddress 10.192.168.2
AutomapHostsOnResolve 1
```

Fig. 25. Configuración de TOR.

Para que se produzca el intercambio de información con la red TOR el archivo TOR debe ser reiniciado, como se indica en la figura 26.

```
root@Tesis:~# /etc/init.d/tor restart
```

Fig. 26. Reinicio del servicio de TOR.

Al finalizar la configuración se podrá visualizar un mensaje de escucha de los puertos con los que opera la red TOR (ver figura 27).

```
Oct 21 02:06:29.455 [notice] Opening Socks listener on 10.192.168.2:9050
Oct 21 02:06:29.459 [notice] Opening DNS listener on 10.192.168.2:53
Oct 21 02:06:29.464 [notice] Opening OR listener on 0.0.0.0:9001
Oct 21 02:06:29.467 [notice] Opening Directory listener on 0.0.0.0:9030
root@metarouter:~#
```

Fig. 27. Notificación de puertos de TOR.

21. Configuración de navegación mixta en el cliente

Al ser una solución de navegación mixta, se deberá configurar el proxy en el navegador que se utilizará para acceder de forma anónima, como lo muestra la Figura 28.

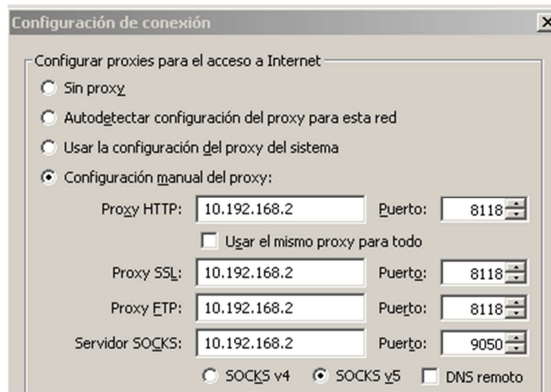


Fig. 28. Configuración del proxy en el navegador.

22. Navegación anónima

Una vez finalizada la configuración, se podrá verificar si está o no navegando en la red TOR mediante la página web www.check.torproject.org. (Ver Figura 29)

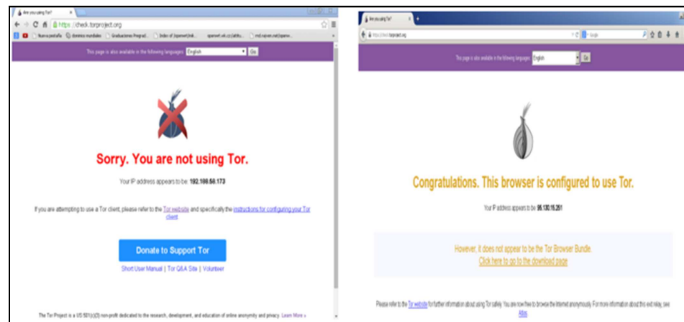


Fig. 29. Página de verificación de navegación.

Se podrán observar, dos resultados dependiendo del navegador utilizado, si el navegador es en el que se configuró previamente el proxy, se podrá navegar en la red TOR, si se utiliza otro navegador se podrá hacer únicamente uso de internet de la manera tradicional.

3. Resultados

A) Medición de ancho de banda.

Para poder realizar una comparación de Ancho de Banda, se utilizó un enlace de datos de 3 Mbps con una compartición 8:1 (ADSL dato del proveedor CNT - ECU). (Figura 30)

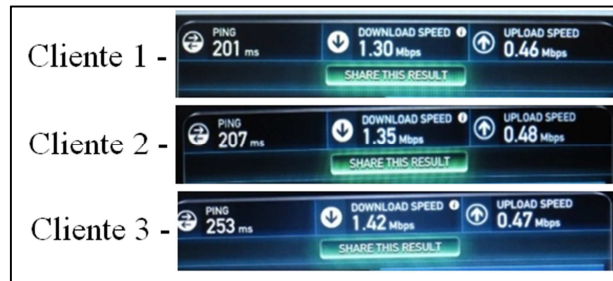


Fig. 30. Medición de ancho de banda para 3 clientes.

En la Figura 30, se realizó la medición de ancho de Banda con tres clientes simultáneos a través del mismo nodo de salida en la red TOR.

Las gráficas de resultados fueron tamizadas a partir de tres escenarios de evaluación, descritos en la Tabla 2.

| Escenario | Dispositivos | |
|-----------|--------------|-----------------------|
| | LAN | Wireless |
| 1 | 1 | 1 Tor – 1 Tradicional |
| 2 | 2 | 1 Tor – 1 Tradicional |
| 3 | 3 | 1 Tor – 1 Tradicional |

Tabla 2. Escenarios de evaluación del servidor.

B) Comportamiento del dispositivo físico Mikrotik.

Se puede verificar que tanto las características físicas del servidor como las características de los servicios implementados se encuentran dentro de un rango estable de operación, esto puede visualizarse en las figuras 31 y 32 respectivamente.

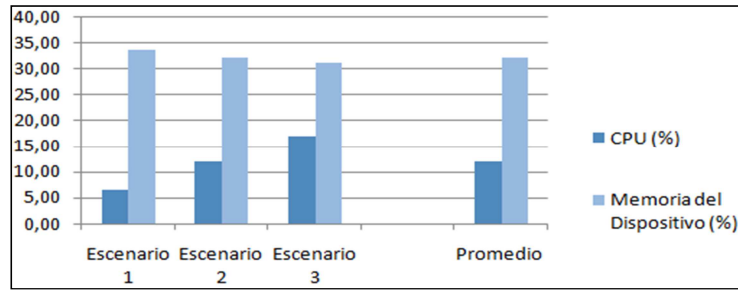


Fig. 31. Uso del CPU y memoria de Mikrotik.

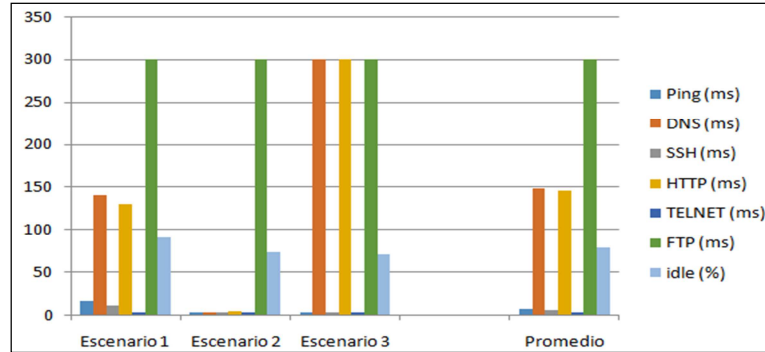


Fig. 32. Tiempos de conexión a servicios de Mikrotik.

C) Comportamiento del dispositivo virtual OpenWRT.

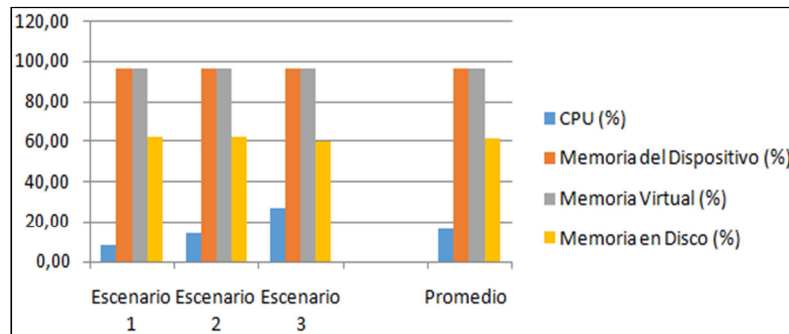


Fig. 33. Uso del CPU y memoria de OpenWRT.

El consumo de recursos generado por el firmware virtual, no representa una carga de procesos superior a la carga del dispositivo físico, permitiendo que tanto el dispositivo físico como el virtual puedan coexistir en un mismo entorno de red, esto es plenamente concluyente a partir de la figura 33.

D) Medición de ancho de banda.

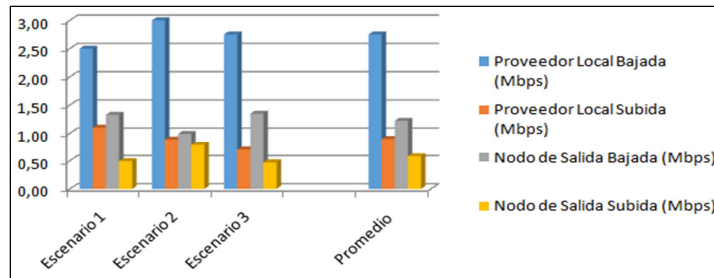


Fig. 38. Medición de ancho de banda en TOR.

De acuerdo a la prueba de ancho de banda se verifica que la navegación anónima, provee menor velocidad de navegación, sin embargo el tener una solución de navegación mixta permitirá aprovechar la velocidad de la navegación tradicional y el anonimato de la red TOR, compensando de alguna manera la pérdida de velocidad en la red TOR.

4. Discusión

Mediante las pruebas realizadas a la topología de red implementada, se evidencia que el servidor de navegación virtualizado mediante MetaRouter, provee navegación anónima sobre el pseudo-dominio .onion de la red TOR, sin comprometer el funcionamiento del servidor Mikrotik, sobre el cual se realizó la virtualización ni tampoco la interacción con los demás dispositivos de comunicación que operan sobre la red.

El aplicar este tipo de herramientas (configuraciones) puede potencialmente permitir que los usuarios evadan ciertos sistemas de control o autenticación ya que su aplicación involucra el desconocer la identidad de los usuarios, potencialmente podrían ser utilizadas para aplicaciones de accesos no autorizados a redes de terceros.

Es muy importante, el grado de seguridad que se quiere implementar, sobre una red de comunicaciones, especialmente en afán de prevenir, problemas tan frecuentes como ladrones de identidad, suplantadores, vendedores de información y demás estafas a las

que se está expuesto en internet. Uno de los problemas, más básicos de privacidad, se produce cuando, un atacante ajeno a la transmisión de datos, se infiltra en la comunicación mediante medios físicos, teniendo acceso, a la información y con ello a las cabeceras y datos embebidos en ellos. Pero también existen formas mucho más potentes, para infiltrarse en la comunicación desde fuera de la red local y realizar un análisis de tráfico, muchos atacantes espían múltiples puntos de internet, rastreando patrones de comunicaciones comunes entre emisores y receptores, en esos casos la protección usando encriptación, no es suficiente en la protección de la información, ya que únicamente esconde los datos del tráfico de internet, mas no las cabeceras. Los campos donde se puede aplicar estos tipos de bloqueo son sumamente amplios, ya que van desde la protección económica, al prevenir ataques bancarios y uso de información indebida, hasta la protección personal, al no revelar la ubicación geográfica de cada usuario.

No menos importante sería el tener en cuenta que con fines de realizar delitos informáticos, sería muy conveniente y útil el usar estas técnicas de “disfraz”. Sin embargo, la investigación actual de la ciencia debe tender o proponer la generación de códigos y herramientas que puedan manejar técnicas de predicción “aprendidas” de este tipo de software y hardware a fin de que estas potenciales aplicaciones no sean utilizadas con fines delictivos.

5. Conclusiones

Los resultados de las pruebas muestran que la navegación anónima sobre la red TOR es aproximadamente dos tercios más lenta que la navegación tradicional, esto se debe al algoritmo de funcionamiento de TOR el cual es independiente del entorno de red sobre el cual esta implementado el servidor de navegación anónima.

Debido a que la red TOR maneja únicamente tráfico TCP, la conversión mediante socks de los demás protocolos a TCP aumenta el tiempo de petición de comunicación a la red

TOR, lo que a su vez se ve reflejado en el aumento de los tiempos de respuesta de la navegación web.

El nivel de exigencia al que está sometido el servidor RB450G, cuando la red ya está en producción no sobrepasa el 65% del uso del procesador, lo que lo convierte en una opción válida y de bajo costo a la hora de implementar servicios virtualizados en pequeñas y medianas redes de comunicaciones.

El firmware OpenWRT permite añadir características y herramientas mediante paquetes de instalación que firmwares propietarios no proveen, permitiendo potenciar un equipo, lo que a su vez brinda la posibilidad de implementar un firmware personalizado a medida de las necesidades de la red.

Independientemente del uso comercial que pueda representar esta solución de firmware mixta, el aprovechar los recursos de un firmware virtual sobre un firmware propietario en un dispositivo de bajo costo, representa una nueva opción de aprendizaje y entrenamiento en sistemas de OpenSource en entornos académicos, aunque se deberá tener especial atención con el número de usuarios que accederán a través del proxy Privoxy ya que el procesamiento del dispositivo Mikrotik se incrementa y esto involucraría la afectación a la velocidad de acceso a la red y la disminución del ancho de banda ofertada.

6. Referencias

- [1] MikroTik RouterOS. MikroTik. Letonia. 2010.
- [2] J. I. Cerón Bergantiños, "Onion routing y Red Tor". Universidad Politécnica de Madrid Escuela Técnica Superior de Sistemas Informáticos. 2014.
- [3] Tor:Overview. Disponible: <https://www.torproject.org/about/overview.html>. Julio 2015.

[4] Tor Stable Manual. Disponible: <https://www.torproject.org/docs/tor-manual.html>. Julio 2015.

[5] RouterBOARD 450G Series User's Manual. MikroTik. Letonia. 2009.

7. Autores

Ing. Roberto Meneses Basantes, Ingeniero Electrónico en Redes y Comunicación de Datos, graduado el año 2014 en la Universidad de las Fuerzas Armadas en Quito – Ecuador. Sus trabajos principalmente están relacionados con seguridad basada en aplicaciones Open Source.

Ing. Darwin Aguilar Salazar, Ingeniero Electrónico en Telecomunicaciones y Master en Redes de Comunicaciones. Docente principal de la Universidad de las Fuerzas Armadas. Coordinador del área de Redes de Información. Sus estudios de investigación están alineados con las vulnerabilidades y análisis de rendimiento eficiencia en redes de datos.

Ing. Christian Vega Muñoz, Ingeniero Electrónico en Telecomunicaciones. Docente principal de la Universidad de las Fuerzas Armadas. Jefe del Laboratorio de Networking, su investigación está relacionada en la optimización de la comunicación de datos sobre las Redes de Transporte.

Ing. Paola León Pérez, Ingeniera Electrónica en Automatización y Control y Master en Diseño, Producción y Automatización Industrial. Docente principal de la Universidad de las Fuerzas Armadas. Sus estudios e investigaciones están alineadas con dispositivos FPGA y Microelectrónica.

La aplicabilidad del cómputo ubicuo para el monitoreo de animales de caza

Adrian Castañeda Morfin

Facultad de Telemática, Universidad de Colima,
Av. Universidad N° 333, CP 28040, Colima, México
acmorfin1@gmail.com

María Andrade-Aréchiga

Facultad de Telemática, Universidad de Colima,
Av. Universidad N° 333, CP 28040, Colima, México
mandrad@ucol.mx

Pedro Damián-Reyes

Facultad de Telemática, Universidad de Colima,
Av. Universidad N° 333, CP 28040, Colima, México
damian@ucol.mx

Resumen

En este documento se exponen varios trabajos de interacción animal computadora (ACI), dado que recientes investigaciones reportan el uso de la tecnología en aplicaciones para una mejor interacción del hombre con los animales. No obstante, en la práctica de la cacería con perros, hay ocasiones que se desconoce su ubicación y trayectoria así como su estado físico, además los perros se desplazan largas distancias, lo que imposibilita poderle transmitir una orden. Por consiguiente, en este trabajo se muestra el desarrollo, implementación y pruebas que se realizaron a una herramienta de software con tecnología móvil para la interacción con perros de caza. La aplicación está enfocada principalmente en conocer la localización y el estado físico del perro durante una actividad de caza, así como transmitir sonidos y vibraciones al dispositivo

del perro para darle alguna instrucción, con el propósito de tener una mejor relación y control sobre estos animales. El software desarrollado en esta investigación tiene como ventaja el que la comunicación de datos se hace a través de telefonía celular, resolviendo el problema que tienen los dispositivos comerciales que utilizan señal de frecuencia muy alta (VHF), y el costo de implementación está por debajo de los dispositivos comerciales.

Palabra(s) Clave(s): caza, interacción animal-computadora, perro, telefonía celular.

1. Introducción

Actualmente es común el uso de las tecnologías de información en muchas de las actividades del ser humano. Su utilización le permite interactuar de manera transparente con dispositivos que perciben lo que rodea al usuario para proporcionarle información o servicios en el lugar y en el momento en que los necesita [1].

Hoy en día se pueden ver perros de trabajo en diferentes actividades en apoyo al hombre, por ejemplo: la cacería, búsqueda y rescate, seguridad, etc. y la tecnología ha abierto un área de oportunidad ofreciendo a los animales dispositivos electrónicos y aplicaciones para que interactúen de mejor manera con sus dueños y darles una mejor calidad de vida proporcionándole los estímulos que necesita [2].

Sin embargo, aún con esa tecnología, en las actividades de caza se presentan situaciones particulares que es necesario atender, por ejemplo: al practicar la cacería con perros, con frecuencia éstos se separan largas distancias de sus dueños perdiéndose de vista y generalmente en terrenos irregulares, lo que origina el desconocimiento de su ubicación, trayectoria y estado físico. Por lo anterior, en el presente trabajo se describe el desarrollo e implementación de una herramienta de software para el apoyo a las actividades de caza, utilizando tecnología móvil.

2. Antecedentes

La tecnología de la comunicación que ha conectado a las personas ha sido una parte integral de nuestra vida cotidiana y recientemente se ha ampliado para ofrecer aplicaciones para los dueños de perros. La tecnología de comunicación móvil se ha utilizado por un par de décadas para la comunicación entre personas y recientemente para monitorizar eventos remotos, salud y diferentes actividades. Actualmente, la tendencia es hacia la investigación del uso de la tecnología para la interacción entre personas y perros [2]. El número de investigadores interesados en la ACI gradualmente ha aumentado, sin embargo, sigue siendo poco con relación a la comunidad de Interacción Humano Computadora (HCI), y por sus contribuciones, la ACI como disciplina es relevante y pertenece a la comunidad de HCI [3].

Uno de los pioneros en interfaz humano-animal fue Resner (2001), quien diseñó un sistema remoto de interacción humano-perro llamándolo Rover@Home. El objetivo fue construir un dispositivo que permitiera a los animales y a sus propietarios tener una interacción mientras se encuentran separados geográficamente, de la misma manera que un teléfono permite a dos seres humanos distantes comunicarse, desarrollaron un dispositivo que realiza una función similar entre animales domésticos y seres humanos. Este sistema permite enseñar a sus perros nuevas tareas, usando internet, una computadora y técnicas de entrenamiento, aplicando principios de diseño de HCI. En su investigación, explica que hay una serie de razones de por qué los perros representan un nuevo e interesante reto de diseño. En primer lugar, que los perros son muy sociables y por lo tanto se benefician de una interacción bien diseñada con un compañero remoto humano. En segundo lugar, los humanos están familiarizados con los perros y esa interacción proporciona un valioso conjunto de conocimiento que sirve de guía para el desarrollo. Y tercero, menciona que el proceso de pensar a través de lo que es una interacción bien diseñada en el contexto de una computadora entre un perro y un ser humano aclara la comprensión para la creación de dispositivos para humanos

y animales por igual. Argumenta que los puntos clave de diseño centrado en el usuario son también muy útiles guiando el proceso de diseño de dispositivos para animales [4].

Años más tarde, Demi Mankoff (2005) desarrolló un sistema de vigilancia a distancia para la interacción humano-perro, ayudando a los perros a mantener un sentido de conexión con sus dueños mediante un juego de lanzarle pelotas de tenis al perro, mitigando el aislamiento social, ya que los humanos se encuentran muchas horas del día fuera de casa o trabajando. El sistema consiste en una Interfaz Usuario-Humano (HUI) y una Interfaz Usuario-Perro (DUI). Demostraron que los sonidos y videos son más efectivos para atraer el interés del perro. El sistema provee a los seres humanos herramientas para proporcionar a los animales una mejor calidad de vida [5].

Actualmente, en el Instituto de Georgia de Tecnología están trabajando en un proyecto llamado FIDO (Facilitating Interactions for Dogs with Occupations), el cual investiga aspectos fundamentales en la tecnología portátil para apoyar la comunicación entre perros de trabajo y sus dueños. Utiliza un chaleco para perros con diferentes sensores (morder, tirar y gestos de nariz) y muestra la creación de sensores portátiles que los perros pueden activar. El principal objetivo de este proyecto es investigar los aspectos fundamentales de tecnología portátil de apoyo en la comunicación de los perros de trabajo a sus dueños. Los perros proporcionan información a través de los sensores integrados en el chaleco del perro, creando interfaces de usuario para perro, los cuales fueron probados con tres diferentes perros de asistencia entrenados para evaluar la facilidad de la interacción. A partir de un estudio piloto, determinaron qué tipo de sensores para perro puede ser más apto para favorecer la interacción. Con la colocación de múltiples sensores en el chaleco del perro para comunicar una variedad de mensajes, probaron cada sensor individualmente en la misma ubicación del lado izquierdo del chaleco del perro, y para cada sensor midieron las lecturas mediante una serie de interacciones con el perro. Los resultados del estudio piloto fueron extremadamente alentadores, y demostraron que es posible crear sensores portátiles que los perros pueden activar para comunicarse con sus dueños [6].

Uno de los dispositivos disponibles en el mercado utilizado en la actividad de caza es el Garmin Astro con un precio aproximado de \$ 9,000 pesos, para el cual su principal problema es la pérdida de conexión entre el collar y el dispositivo de mano, debido a la utilización de señal VHF [7].

Los sistemas basados en frecuencia muy alta (VHF) tienen un rango máximo de 3 a 10 km dependiendo del entorno, por lo que la comunicación se anula cuando no hay línea de vista, es decir, cuando existen obstáculos de por medio [2].

3. Enfoque

Se desarrolló una aplicación móvil para Smartphone utilizando datos a través del servicio de un proveedor celular. Las principales funciones de la aplicación son: localización, sonidos, vibraciones y ritmo cardiaco.

En la Fig. 1 se muestra el esquema arquitectónico de la aplicación desarrollada. Se utilizaron dos Smartphone, uno es portado por el cazador y el otro se integra en un chaleco especialmente diseñado para el perro. También se utiliza un sensor de ritmo cardiaco ubicado por debajo del chaleco del perro. El sensor transmite el ritmo cardiaco mediante comunicación Bluetooth al Smartphone del perro. Estos datos junto con la localización es transmitida al teléfono del cazador por medio de la red celular. El cazador transmite instrucciones al teléfono del perro, en donde se reproducen sonidos y vibraciones que representan órdenes que el cazador le indica al perro.

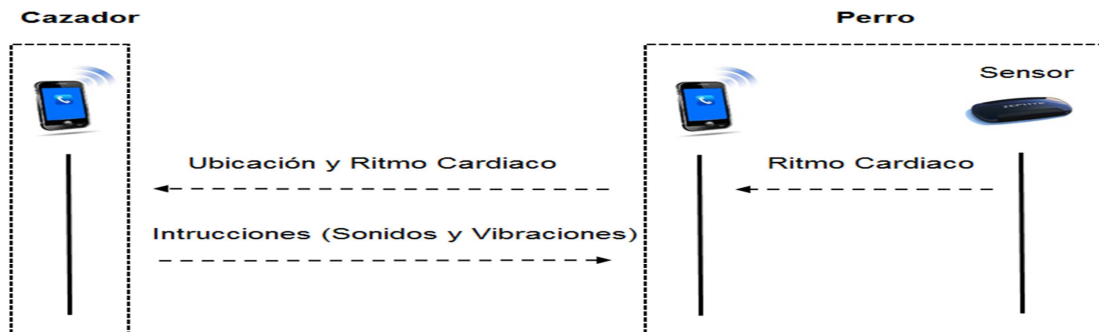


Fig. 1. Diagrama arquitectónico de la aplicación.

4. Diseño y desarrollo

Se analizaron una gran variedad de lenguajes y entornos de programación que pueden ser utilizados para crear aplicaciones para dispositivos móviles. Después de esto, se optó por utilizar la plataforma Microsoft .Net Framework por ser una de las más actuales y eficaces. Específicamente se utilizó el entorno de desarrollo integrado Microsoft Visual Studio 2013 y el lenguaje de programación C# [8]. Los teléfonos utilizados tienen el sistema operativo Windows Phone 8.1.

La comunicación entre los dispositivos se hizo a través de un servicio web, que realiza la interacción entre las aplicaciones instaladas en los dispositivos, como se presenta en el diagrama de secuencia de la Fig. 2.

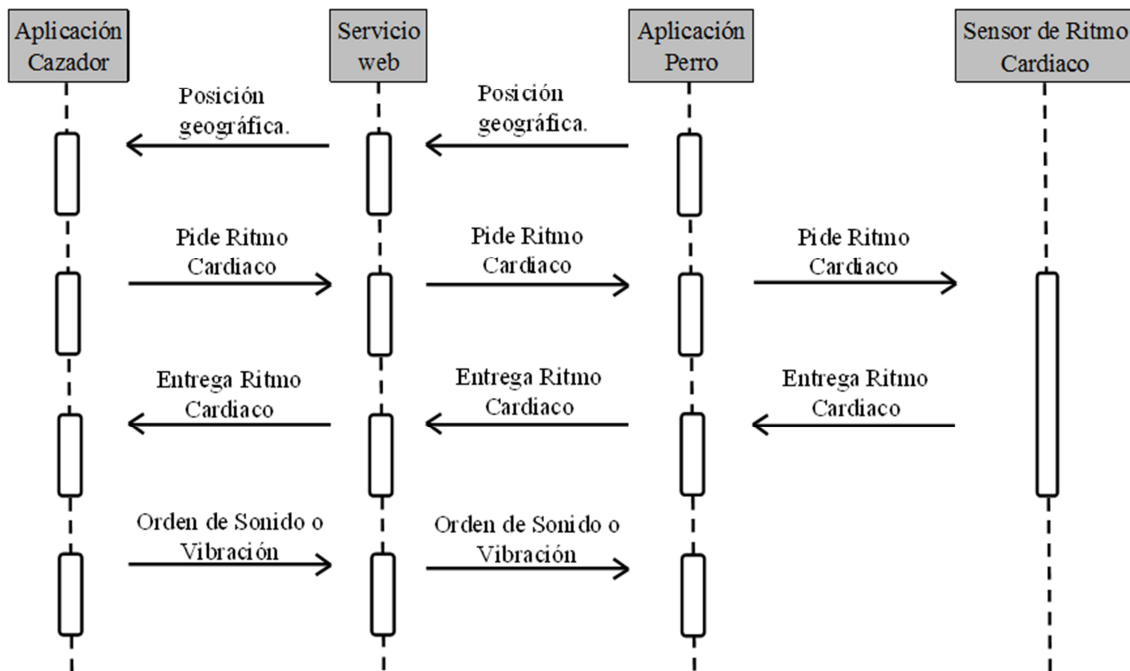


Fig. 2. Diagrama de secuencia.

Las aplicaciones en los dispositivos utilizan mecanismos de localización, reproducción de sonidos y vibraciones. La aplicación instalada en el dispositivo del cazador transmite órdenes y la del perro transmite ubicación geográfica y ritmo cardíaco,

4.1. Servicio Web

El servicio web hace uso de una base de datos con el fin de tener identificado a cada dispositivo por medio de su URI (Identificador de Recursos Uniforme). Además, se crearon cuatro métodos: dos de ellos permiten estar actualizando la URI, uno para cada dispositivo, y los otros dos permiten intercambiar mensajes entre los dispositivos.

Los procedimientos de actualización de la URI tienen la función de recibir un mensaje de tipo cadena y mediante una consulta SQL actualizar el URI en la base de datos.

Los procedimientos que permiten enviar mensajes, reciben el mensaje de un dispositivo y lo mandan al otro dispositivo utilizando la tecnología de notificaciones Push, la cual es un tipo de comunicación en la que el servidor inicia una petición al cliente, cuando tiene nueva información o una notificación. Las notificaciones Push, crean un canal para enviar y recibir notificaciones. El tipo de notificaciones que se utilizan son Raw, las cuales permiten enviar información a la aplicación cuando se encuentra en ejecución [9].

4.2. Aplicación para el dispositivo del cazador

La aplicación del dispositivo del cazador contiene dos secciones que pueden ser desplazadas deslizando la pantalla hacia la izquierda o derecha, la primer sección muestra el mapa con la ubicación de los dos dispositivos, en la parte inferior contiene una barra para poder cambiar el tipo de vista del mapa (Aéreo, Híbrido, Camino y Terreno), así como activar-desactivar y limpiar la ruta, como se muestra en la Fig. 3(a), y la segunda sección contiene 6 botones (Busca, Vámonos, Quieto, Vibración, RC-Bat Sensor y Bat. Del Celular) para la comunicación con el dispositivo del perro, como se muestra en la Fig. 3(b).



Fig. 3(a). Sección de lo localización.



Fig. 3(b). Sección de órdenes.

4.3. Aplicación para el dispositivo del perro

La aplicación del dispositivo del perro contiene 3 secciones. La primera muestra la ubicación del dispositivo con la opción de poder cambiar la vista del mapa así como poder activar el rastreo, como se muestra en la Fig. 4(a). La segunda sección se utiliza para activar el sensor de ritmo cardiaco Fig. 4(b). Y la tercera sección muestra si está activado el rastreo, junto con la última posición registrada con fecha y hora Fig. 4(c).

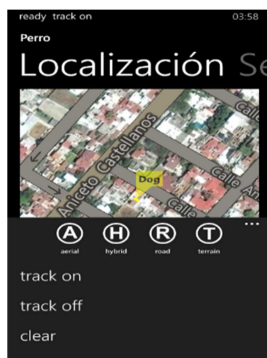


Fig. 4(a). Sección de localización.



Fig. 4(b). Sección para el ritmo cardiaco.

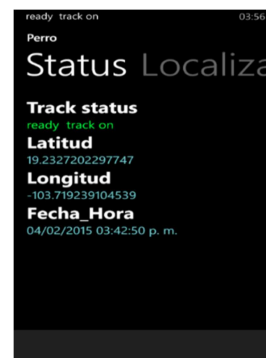


Fig. 4(c). Sección de status de la aplicación.

La interfaz gráfica en la aplicación del dispositivo del perro se utiliza para activar el sensor de ritmo cardiaco y la función de rastreo, además el software registra la trayectoria en caso de que el Smartphone pierda cobertura por parte del proveedor de servicio de datos, esto con el fin de poder conocer la trayectoria que recorrió.

Cabe señalar que el fabricante del sensor Zephyr HxM BT utilizado en la aplicación del dispositivo del perro para obtener el ritmo cardiaco no proporciona códigos para el sistema operativo de Windows Phone, por lo cual se requirió investigar exhaustivamente cómo establecer conexión del Bluetooth con el sensor y la lectura de los paquetes de datos. Luego se hicieron múltiples pruebas hasta lograrlo.

Otro elemento que se utilizó en los dispositivos fue una aplicación gratuita para Windows Phone llamada “Easy Recorder” con la cual se grabaron los archivos de audio con las órdenes del cazador o dueño de los perros.

La programación de cada una de los módulos no se exhibe debido a que es bastante amplia, pero se cuenta con el código respectivo para una correcta revisión en su funcionamiento.

5. Implementación y pruebas

Para la implementación física en los perros se confeccionaron dos chalecos, uno grande y uno pequeño de acuerdo al tamaño de la raza a utilizar. El chaleco contiene un compartimento para portar el celular como se muestra en la Fig. 5.



Fig. 5. Chaleco para perro.

Los dispositivos que se utilizaron fueron dos Smartphone Nokia Lumia (520 y 625) y un sensor de ritmo cardiaco marca Zephyr HxM BT, cuyas características básicas se describen en la Tabla 1.

| | Peso | Tamaño | Batería Aprox. |
|------------------------|-------------|-----------------------|-----------------------|
| Nokia Lumia 520 | 124 g | 119.9 x 64 x 9.9 mm | 10 Hrs. |
| Nokia Lumia 625 | 159 g | 133.3 x 72.3 x 9.2 mm | De 15 a 24 Hrs. |
| Zephyr (Sensor) | 85 g | 70 x 35 mm | 26 Hrs. |

Tabla 1. Descripción de los dispositivos.

Para realizar las pruebas, se utilizaron dos perros que se identifican en la Fig. 6(a) y 6(b), con las características que se describen en la Tabla 2.

| Nombre | Raza | Edad | Peso |
|---------------|-----------------|-------------|-------------|
| Muñeca | Beagle | 6 Años | 10 Kgs. |
| Max | Azul de Gascuña | 3 Años | 20 Kgs. |

Tabla 2. Descripción de los perros.



Fig. 6(a). Muñeca y 6(b). Max, ambos portando chaleco.

Se observó el uso del chaleco en ambos perros el cual se adapta perfectamente al cuerpo por su diseño, permitiendo su movilidad para caminar como para correr de manera natural. Además no mostraron algún cambio en su comportamiento, ni les pareció extraño su uso.

El ritmo cardiaco se obtiene mediante el uso de un sensor que se muestra en la Fig. 7(a), el cual está adherido a una banda elástica por medio de unos broches. Además se investigó con un veterinario la mejor posición del sensor en el perro, el cual es el lado izquierdo del tórax y se les cortó el pelo para una mejor lectura del ritmo cardiaco, Fig. 7(b).



Fig. 7(a). Sensor de ritmo cardiaco.



Fig. 7(b). Perro con corte de pelo.

Se les colocó el sensor de ritmo cardiaco y el celular por turnos a cada perro.

Las pruebas de localización, ritmo cardiaco y órdenes se realizaron en el Parque del Rodeo, en la colonia La Estancia al oriente de la ciudad de Colima, como se muestra en la Fig. 8.

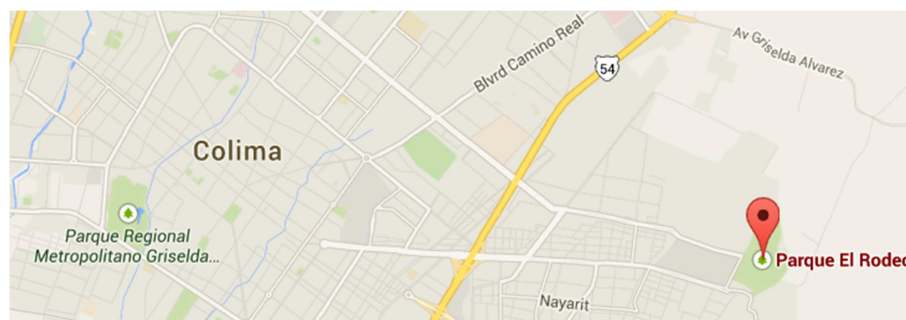


Fig. 8. Ubicación del Parque El Rodeo

5.1. Prueba de rastreo y localización

Se le colocó a uno de los perros el equipo y se soltaron libremente reconociendo el área, lo cual se pudo apreciar desde el equipo del cazador su trayectoria en tiempo real

ya que los perros se perdieron de la vista, porque la zona se encuentra en desnivel con una barranca y un arroyo.

En las Fig. 9(a) a la 9(d) se muestran diversas pantallas tomadas en diferentes momentos que muestran la trayectoria que el perro realizó, iniciando a las 04:14:12 p.m. y terminando a las 04:41:32 p.m. Durante este periodo de tiempo se dio seguimiento en todo momento de la ubicación y trayectoria del perro.

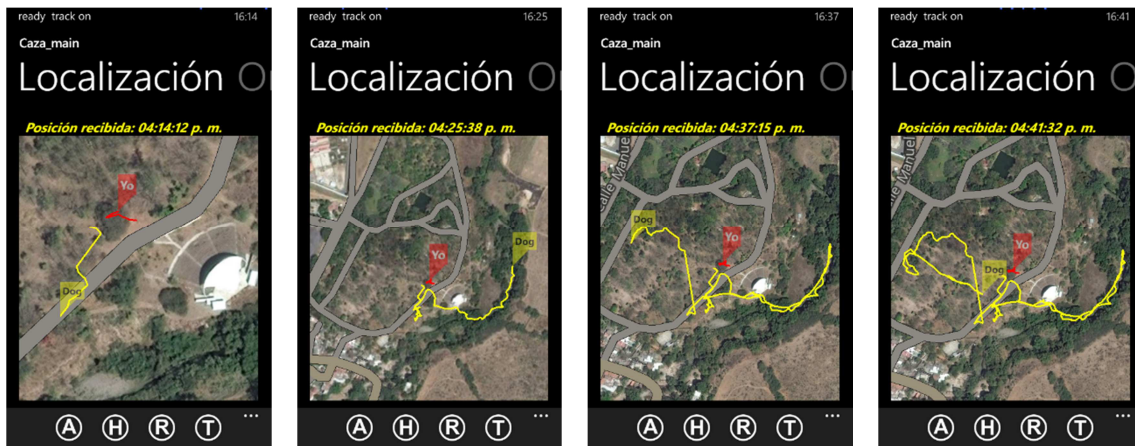


Fig. 9(a).

Fig. 9(b).

Fig. 9(c).

Fig. 9(d).

5.2. Prueba de ritmo cardiaco

Para probar la funcionalidad del sensado del ritmo cardiaco es importante hacer la siguiente consideración. En esta parte intervienen dos aspectos: a) función de la aplicación para obtener el ritmo cardiaco del sensor y b) función del sensor colocado en el perro. En relación a la función de la aplicación para obtener la información del ritmo cardiaco del sensor, pudo comprobarse que en todo momento que el cazador solicitó obtener dicha información, la aplicación funcionó adecuadamente, pues en el dispositivo del cazador se desplegó de manera inmediata la información. En relación a la funcionalidad del propio sensor colocado en el perro, se pudo observar lo siguiente:

- Por la forma en que está colocado en el perro es muy fácil de moverse de lugar lo cual presenta problemas para la lectura del ritmo cardiaco, y en estos casos, en el dispositivo del cazador no se despliega información alguna. Como solución momentánea, se determinó apretar el elástico que sujeta al sensor para tratar de mantenerlo lo más ajustado posible y mantener con ello su posición adecuada para la lectura del ritmo. Ya que los perros tienen movimientos bruscos, se tiene contemplado como trabajo futuro, trabajar en un mejor mecanismo de colocación.
- Se consideró conveniente activar el sensor antes de colocársele al perro ya que la primera vez que se obtiene el ritmo cardiaco puede tardar hasta 15 segundos en visualizarse la información sensada en el dispositivo del cazador. Posteriormente, la información se despliega inmediatamente de que el cazador lo solicita.

5.3. Prueba de órdenes

En cuanto a las órdenes de sonidos y vibraciones que se envían al perro se detectó que en ese momento la velocidad de transmisión de datos estaba un poco lenta ya que en ocasiones la aplicación reflejaba un retraso de varios segundos, pero en lo general funcionó bien.

Además se observó que al reproducir sonidos y vibraciones, los perros en ocasiones mostraron sorpresa, a causa de que la ejecución de una orden por un perro depende del entrenamiento constante por parte de su dueño.

Al momento que se envía una orden al dispositivo del perro, este regresa un mensaje de confirmación, esto con el fin de asegurar que la aplicación del perro está activa y recibió la orden.

Además de poder obtener el ritmo cardiaco del sensor, también se recupera el porcentaje de batería del celular del perro, con la finalidad de conocer el tiempo de carga restante.

Se observó que la herramienta no tiene límite en cuanto a distancia entre los dispositivos siempre y cuando se encuentre en el rango de cobertura del proveedor de datos.

El consumo de datos que realiza la aplicación por acciones o tareas se presenta en la Tabla 3. Como se observa, el consumo de datos no es significativo, ya que se utiliza menos del 1% de 1 Megabyte disponible.

| Acción o tarea | Bytes |
|---|--------------|
| Actualizar URI | 203 |
| Transmitir/Recibir orden | 2 |
| Transmitir/Recibir posición | 41 |
| Status de notificación del servicio web | 25 |
| Transmitir/Recibir confirmación de orden recibida | 3 |
| Transmitir/Recibir Ritmo Cardíaco | 15 |
| Transmitir/Recibir el porcentaje de batería del dispositivo del perro | 3 |

Tabla 3. Consumo de datos por la aplicación.

6. Trabajo futuro

Con base a lo observado en las pruebas de campo con la aplicación, se detectaron los siguientes aspectos a desarrollar:

- Por el momento, con la aplicación se conoce la ubicación, trayectoria de movimiento y ritmo cardíaco de un solo perro, pero habría que considerarse el monitoreo de una cantidad mayor de perros de manera simultánea.
- Permitir que el cazador personalice sus instrucciones/órdenes de caza, ya que por el momento todas se encuentran predefinidas y grabadas en los dispositivos.
- Encontrar un mejor mecanismo para la colocación del sensor que evite que se mueva de lugar con los movimientos bruscos del perro.

- Que a la aplicación se le programen diferentes categorías de acuerdo al registro de niveles de ritmo cardiaco, que indiquen problemas en el estado de salud del perro o distintas situaciones en las que éste se encuentra (sin actividad, reposo, cansancio, entre otras), a modo de que la aplicación envíe alertas al cazador cuando se rebasen estos niveles.

7. Conclusiones

En el artículo se presenta información sobre la aplicabilidad del cómputo ubicuo para apoyar a un cazador en el monitoreo de sus perros de caza. A través de pruebas realizadas en campo, se mostró que con el software desarrollado, el cazador utilizando un dispositivo móvil, puede dar visualizar la localización, trayectoria y dirección que el perro sigue, así como el estado de salud del mismo (a través de la información de ritmo cardiaco) en tiempo real. Además, el cazador puede dar indicaciones pertinentes a su perro, aún cuando no esté cerca de él.

Si bien, como se mencionó, existen dispositivos comerciales cuyo objetivo es apoyar en la actividad de la caza, tienen también sus limitaciones ya que al utilizar señal VHF para enviar y recibir información, ésta se pierde al no tener línea de vista. Por lo anterior, la aplicación desarrollada en esta investigación presentó las siguientes ventajas:

- La transmisión y recepción de datos se hace a través de telefonía celular resolviéndose el problema de los dispositivos comerciales, sin embargo, es requisito indispensable la existencia de cobertura de una red celular.
- El costo de implementación de la aplicación comparada con los dispositivos comerciales es bastante bajo, ya que los teléfonos celulares de gama media están por debajo de lo que cuesta un dispositivo comercial. Se pueden reusar dispositivos celulares. Incluso el consumo de datos, no es significativo de acuerdo a lo encontrado en el experimento, por ejemplo, la tarea que con mayor frecuencia realiza la herramienta corresponde al cambio de posición, lo cual requiere 41 bytes (un byte como identificador, 16 para la latitud, 16 para la

longitud y 8 para la hora), lo que significa que con 1 Megabyte se pueden enviar o recibir hasta 25,575 cambios de posición.

Por lo anterior, se aprecia que la aplicación desarrollada en esta investigación puede apoyar en el control canino para actividades de caza. Además la manera de aplicar el cómputo ubicuo ha quedado demostrada al combinar tecnologías que pueden ser utilizadas en todo lugar y en todo momento de una manera transparente para el usuario. En este caso, los teléfonos celulares son un instrumento de uso cotidiano que no resulta intrusivo para el cazador y con la utilización del chaleco en los perros se observó que no afectó su comportamiento. La implementación y pruebas fueron exitosas.

8. Referencias

- [1] M. Weiser, "The Computer for the Twenty-First Century". *Scientific American*. Vol. 265. No. 3. September 1991. 94-104 pp.
- [2] M. Paldanius, T. Kärkkäinen, Communication technology for human-dog interaction: exploration of dog owners' experiences and expectations. *Proc. CHI'11, ACM*. 2011. 2641-2650 pp.
- [3] Animal-computer interaction (ACI): changing perspective on HCI, participation and sustainability. Paper presented at the CHI '13 Extended Abstracts on Human Factors in Computing Systems. Paris, France. 2013.
- [4] B. I. Resner. *Rover@home: Computer Mediated Remote Interaction Between Humans and Dogs*. Master's Thesis. Massachusetts Institute of Technology. 2001.
- [5] Supporting interspecies social awareness: using peripheral displays for distributed pack awareness. Paper presented at the Proceedings of the 18th

annual ACM symposium on User interface software and technology. Seattle, WA, USA. 2005.

- [6] FIDO - facilitating interactions for dogs with occupations: wearable dog-activated interfaces. Paper presented at the Proceedings of the International Symposium on Wearable Computers. Zurich, Switzerland. 2013.
- [7] DOG TRACKING SYSTEM. <http://sites.garmin.com/astro/>. Abril, 2015.
- [8] B. Johnson, Profesional Visual Studio 2013. 2014. John Wiley & Sons, Inc. Indianapolis, Indiana.
- [9] Whitechapel, S. McKenna, Windows Phone 8 Development Internals Professional. 2013. O'Reilly Media, Inc. Sebastopol, California.

9. Autores

L.I. Adrian Castañeda Morfin. Graduado de licenciado en informática por la Universidad de Colima, actualmente se encuentra estudiando la Maestría en Computación en ésta misma institución.

La dra. María Andrade-Aréchiga obtuvo su grado de Maestría en Educación con especialidad en Matemáticas en el ITESM y el doctorado en Ciencias de la Computación en el Centro de Investigación Científica y de Educación Superior de Ensenada. Se desempeña como profesora-investigadora de tiempo completo en la Universidad de Colima.

D. en C. Pedro Damián-Reyes es Ingeniero en Computación por el Instituto Tecnológico de Colima, Maestro en Telemática por la Universidad de Colima y Doctor en Ciencias en Ciencias Computacionales por el Centro de Investigación Científica y de Educación Superior de Ensenada. Actualmente es profesor investigador de tiempo completo en la Universidad de Colima.